

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

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

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

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

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

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 innovative 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. The training of such highly skilled individuals, has a sustained and significant positive economic impact.

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^AT_EX 2 _{ε} using the “heptesis” class [1].

Contents

1	Introduction	12
2	Theoretical Framework	14
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
3	The Large Hadron Collider and the ATLAS Detector	28
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
4	Tracking and flavour tagging	53
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 χ^2 Fitter Outlier Removal	61
4.3	Conclusion	63
5	Track Classification MVA	65
5.1	Machine Learning Background	65
5.1.1	Neural Networks	66
5.1.2	Training with Gradient Descent	68
5.2	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
6	Graph Neural Network Flavour Tagger	82
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	b -tagging Performance	95
6.4.2	c -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
7	Boosted VH$b\bar{b}$ Analysis	119
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 Modelling Uncertainties	130
7.2.3	Sources of Experimental Uncertainties	132
7.2.4	Vector Boson + Jets Modelling	133
7.2.5	Diboson Modelling	137
7.2.6	$t\bar{t}$ and single-top Modelling	142
7.2.7	Signal Modelling	143
7.3	Statistical Treatment	144
7.3.1	Likelihood Function	144
7.3.2	Background Normalisations	147
7.3.3	Asimov Dataset & Expected Results	148
7.4	Results	148
7.4.1	Post-fit Distributions	149
7.4.2	Observed Signal Strength & Significance	149
7.4.3	Impact of Systematics	153

7.5 Conclusion	157
8 Conclusion	158
8.1 Summary	158
8.2 Future Work	159
Bibliography	161
List of figures	176
List of tables	188

² Chapter 1

³ Introduction

- ⁴ This thesis describes various efforts in improving the understanding of the Higgs
⁵ boson and its coupling to heavy flavour quarks, primarily through the improvement
⁶ of the algorithms used to reconstruct and analyse jets. The thesis is structured in
⁷ the following manner:
- ⁸ Chapter 2 describes the theoretical foundations of the work presented in the rest of
⁹ the thesis.
- ¹⁰ Chapter 3 describes the ATLAS detector and the CERN accelerator complex. Details
¹¹ of reconstructed physics objects are also provided.
- ¹² Chapter 4 provides an overview of the reconstruction of charged particle tracks
¹³ (tracking) and identification of jets containing b -hadrons (b -tagging) at ATLAS, and
¹⁴ studies into the challenges of high transverse momentum b -tagging.
- ¹⁵ Chapter 5 describes the development of an algorithm to predict the origins of tracks.
¹⁶ The tool is used to improve b -tagging performance by the identification and removal
¹⁷ of fake tracks before their input to the b -tagging algorithms.
- ¹⁸ Chapter 6 introduces a novel monolithic approach to b -tagging using graph neural
¹⁹ networks and auxiliary training objectives.
- ²⁰ Chapter 7 describes the measurement of the associated production of a Higgs boson
²¹ decaying into a pair of b -quarks at high transverse momentum.
- ²² Chapter 8 contains some concluding remarks.

²³ The author's contribution to the work presented in this thesis is as follows.

²⁴ **Tracking:** The author was an active member of the Cluster and Tracking in Dense
²⁵ Environments group throughout their PhD, starting with their qualification task
²⁶ on the understanding of tracking performance at high transverse momentum. The
²⁷ author played a key role in software r22 validation studies for the tracking group,
²⁸ including the validation of the quasi-stable particle interaction simulation and the
²⁹ radiation damage Monte-Carlo simulation. The author helped design and improve
³⁰ several tracking software frameworks, and contributed to heavy flavour tracking
³¹ efficiency studies in dense environments. The author developed a tool to identify
³² and reject fake-tracks, which is being investigated for use in the upcoming tracking
³³ paper.

³⁴ ***b*-tagging:** The author has been an active member of the Flavour Tagging group
³⁵ since September 2014. The author played a key role in investigating the performance
³⁶ of the low level taggers at high transverse momentum and led studies into the
³⁷ labelling and classification of track origins. Based on work by Jonathan Shlomi [2],
³⁸ the author helped develop a new flavour tagging algorithm which offers a large
³⁹ performance improvement with respect to the current state of the art. The author
⁴⁰ was the primary editor of a public note associated with this work [3], which will
⁴¹ also be further developed in an upcoming paper. The author also contributed to
⁴² the proliferation of the new algorithm to the trigger, High Luminosity LHC, and
⁴³ $X \rightarrow bb$ use cases. The author also played a key role in software r22 validation
⁴⁴ studies for the Flavour Tagging group, including the validation of the quasi-stable
⁴⁵ particle interaction simulation. The author maintains and contributes to various
⁴⁶ software frameworks used in the Flavour Tagging group, including as lead developer
⁴⁷ of three packages, to create training datasets, pre-process samples for algorithm
⁴⁸ studies and a framework for training graph neural networks, and contributes to group
⁴⁹ documentation.

⁵⁰ **Higgs:** The author was an active member of the Boosted VHbb analysis group. The
⁵¹ author performed various studies deriving systematic uncertainties for the $V+jets$
⁵² and diboson backgrounds. The author also produced and maintained samples, ran fit
⁵³ studies and cross checks, and gave the diboson unblinding approval talk to the Higgs
⁵⁴ group. The author also contributed to the development of the analysis software.

55 **Chapter 2**

56 **Theoretical Framework**

57 The Standard Model (SM) of particle physics is the theory describing all known
58 elementary particles and their interactions via three of the four fundamental forces.

59 Developed by merging the successful theories of quantum mechanics and relativity
60 in the second half of the 20th century, the SM's position today at the centre of our
61 understanding of the nature of the Universe is firmly established by an unparalleled
62 level of agreement between the model predictions and experimental results [4, 5].

63 The SM has predicted the discovery of the top and bottom quarks [6–8], the W
64 and Z bosons [9], and the tau neutrino [10]. The last missing piece of the SM to be
65 discovered was the Higgs boson, first theorised in the 1960s [11–13], and eventually
66 observed at the LHC in 2012 [14, 15]. After its discovery, much ongoing work has
67 been carried out performing detailed measurements of its mass and interactions with
68 other particles.

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

71 **2.1 The Standard Model**

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

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

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

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

Generation	Leptons			Quarks		
	Flavour	Mass [MeV]	Charge [e]	Flavour	Mass [MeV]	Charge [e]
First	e	0.511	-1	u	2.16	$2/3$
	ν_e	$< 1.1 \times 10^{-6}$	0	d	4.67	$-1/3$
Second	μ	105.7	-1	c	1.27×10^3	$2/3$
	ν_μ	< 0.19	0	s	93.4	$-1/3$
Third	τ	1776.9	-1	t	173×10^3	$2/3$
	ν_τ	< 18.2	0	b	4.18×10^3	$-1/3$

Table 2.1: The fermions of the SM [16]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

86

2.1.1 Quantum Electrodynamics

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

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

Table 2.2: The bosons of the SM [16]. 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.

⁹² Dirac Lagrangian density is

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

⁹³ where $\cancel{\partial} = \gamma^\mu \partial_\mu$ denotes the contraction with the Dirac gamma matrices γ^μ (summation over up-down pairs of indices is assumed). Application of the Euler-Lagrange equation on Eq. (2.2) yields the Dirac equation

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

⁹⁶ Suppose some fundamental symmetry that requires invariance under a local $U(1)$
⁹⁷ gauge transformation

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

⁹⁸ where α varies over every spacetime point x . Under this transformation, the Dirac
⁹⁹ equation transforms as

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

¹⁰⁰ For the Dirac equation to remain invariant under the transformation in Eq. (2.4),
¹⁰¹ a new field A_μ which transforms as $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$ must be added. The
¹⁰² transformed interaction term

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

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

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

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

115 2.1.2 Quantum Chromodynamics

116 Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interactions.
117 Quarks and gluons carry colour charge, which comes in three kinds, called
118 red, green and blue. While the $U(1)$ symmetry group in Section 2.1.1 was Abelian,
119 the QCD Lagrangian is specified by requiring invariance under transformations from
120 the non-Abelian $SU(3)$ group, making it a Yang-Mills theory [17] which requires the
121 addition of self-interacting gauge fields. The infinitesimal $SU(3)$ group generators
122 are given by $T_a = \lambda_a/2$, where λ_a are the eight Gell-Mann matrices. These span the
123 space of infinitesimal group transformations and do not commute with each other,
124 instead satisfying the commutation relation

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

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

¹²⁹ strength tensors, one for each generator T_a , defined as

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

¹³⁰ where A_μ^a are the gluon fields and g_s is the strong coupling constant. The covariant
¹³¹ derivative is written as

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

¹³² The full QCD Lagrangian is then given by

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

¹³³ Cubic and quartic terms of the gauge fields A_μ^a appear in the Lagrangian, leading to
¹³⁴ the gluon's self interaction.

¹³⁵ The QCD coupling constant g_s varies, or “runs”, with energy. At lower energy
¹³⁶ scales (and corresponding larger distance scales) the interaction is strong. This
¹³⁷ leads to quark confinement, whereby an attempt to isolate individual colour-charged
¹³⁸ quarks requires so much energy that additional quark-antiquark are produced. At
¹³⁹ higher energy scales (and corresponding smaller distance scales), asymptotic freedom
¹⁴⁰ occurs as the interactions become weaker, allowing perturbative calculations to be
¹⁴¹ performed. Hadrons are bound states of quarks. They are invariant under $SU(3)$
¹⁴² gauge transformations (i.e. are colour-charge neutral, or *colourless*).

¹⁴³ 2.1.3 The Electroweak Sector

¹⁴⁴ The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam
¹⁴⁵ (GWS) model of electroweak interaction [18–20]. The Lagrangian is specified by
¹⁴⁶ requiring invariance under the symmetry group $SU(2)_L \otimes U(1)_Y$, as motivated by a
¹⁴⁷ large amount of experimental data. Here, $SU(2)_L$ is referred to as weak isospin and
¹⁴⁸ $U(1)_Y$ as weak hypercharge.

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

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

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

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

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

159 The weak force violates parity conservation [21–23], i.e. invariance under parity
160 transformations (mirror reflections). Only left handed fermions participate in the
161 weak interaction. Since there is no other force through which neutrinos interact with
162 other particles, there are no right handed neutrinos in the standard model.

163 2.2 The Higgs Mechanism

164 The Brout-Englert-Higgs mechanism (henceforth just “Higgs mechanism”) is the
165 mechanism through which the fundamental particles of the SM acquire mass [11–13].
166 Experimentally it was known that the weak force had a weak effective strength,
167 which was suggestive of a massive mediating gauge particle. However, directly
168 adding mass to the weak gauge bosons violates the non-Abelian symmetry of the
169 SM. Instead, the gauge bosons can gain mass through the interaction with a scalar
170 Higgs field which results from the spontaneous breakdown of symmetry as discussed
171 in Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as
172 discussed in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the
173 Higgs particle relevant to hadron colliders.

¹⁷⁴ 2.2.1 Electroweak Symmetry Breaking

¹⁷⁵ Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It
¹⁷⁶ is the transition of a physical system from a state of manifest symmetry to a state
¹⁷⁷ of hidden, or *broken*, symmetry. In particular, this applies to physical systems
¹⁷⁸ where the Lagrangian observes some symmetry, but the lowest energy vacuum states
¹⁷⁹ do not exhibit that same symmetry. In other words, the symmetry is broken for
¹⁸⁰ perturbations around the vacuum state.

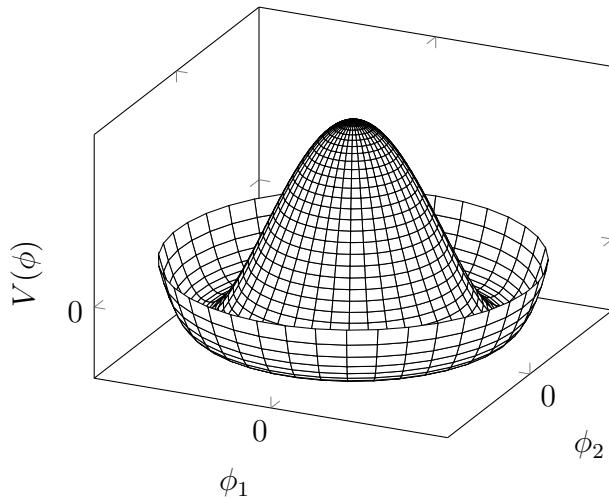


Figure 2.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a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

¹⁸¹ Consider gauge fields from the local $SU(2)_L \otimes U(1)_Y$ symmetry group discussed in
¹⁸² Section 2.1.3 coupled to a complex scalar field $\phi = \phi(x)$. The scalar field ϕ transforms
¹⁸³ as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing
¹⁸⁴ $\phi^2 \equiv \phi^\dagger \phi$, the Lagrangian is

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

¹⁸⁵ where the covariant derivative is given by

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

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

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

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

Adding the particle content back to the theory by expanding the field around the vacuum state, and making a transformation to the unitary gauge to remove unphysical Nambu-Goldstone modes (which arise in the context of global symmetries [24, 25]), yields

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

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

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

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

213 where θ_W is the weak mixing angle defined by

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

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

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

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

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

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

233 Spontaneous symmetry breaking has modified the original symmetry group of the SM
234 $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$. Three broken generators from the symmetry
235 group $SU(2)_L \times U(1)_Y$ have been absorbed into the definition of the physical weak

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

238 2.2.2 Fermionic Yukawa Coupling

239 Adding the masses of the fermions by hand breaks the gauge invariance of the
240 theory. Instead, we can use a Yukawa coupling between the fermion fields and the
241 Higgs field in order to generate mass terms after spontaneous electroweak symmetry
242 breakdown [19]. In this way, the fermion masses are determined by both the respective
243 couplings to the Higgs field and the VEV of the Higgs field itself, which sets the
244 basic mass scale of the theory.

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

248 Lepton Masses

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

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

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

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

255 using $\bar{e} e = (\bar{e}_L + \bar{e}_R)(e_L + e_R) = \bar{e}_L e_R + \bar{e}_R e_L$. The VEV component of ϕ provides
256 the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore
257 be identified as the electron mass term. An interaction term between the electron
258 field e and the true vacuum Higgs field H is also present. Mass is generated for the
259 other lepton generations in the same way.

260 **Quark Masses**

261 The down-type quarks acquire their mass analogous to the leptons, with $\psi_L = (u_L, d_L)$
262 and $\psi_R = d_R$ for the first quark generation. Mass is generated for the up-type quarks
263 using the conjugate field to ϕ which transforms under $SU(2)$ as a doublet with
264 $Y = -1/2$. The conjugate field $\tilde{\phi}$ is constructed as

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

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

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

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

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

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

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

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

2.2.3 Higgs Sector Phenomenology

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

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

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

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

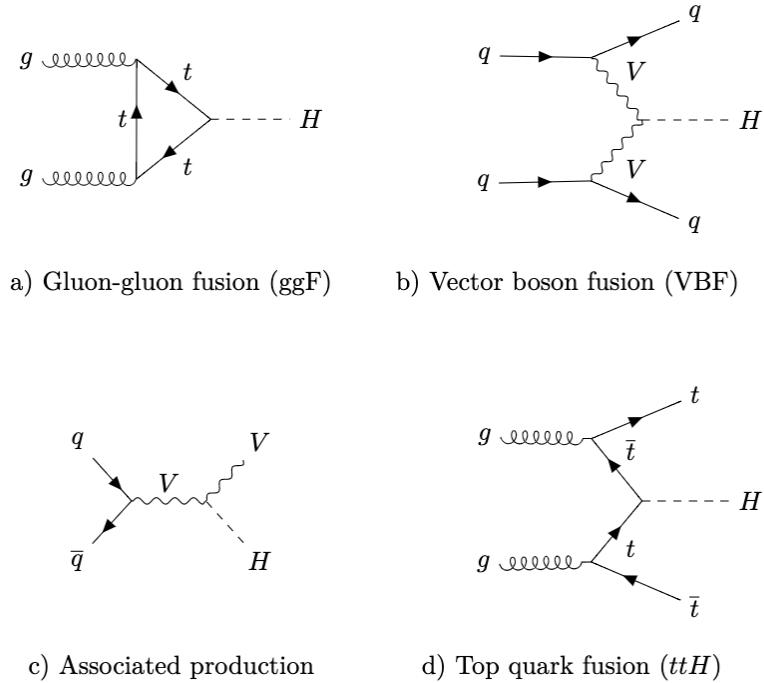


Figure 2.2: Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.

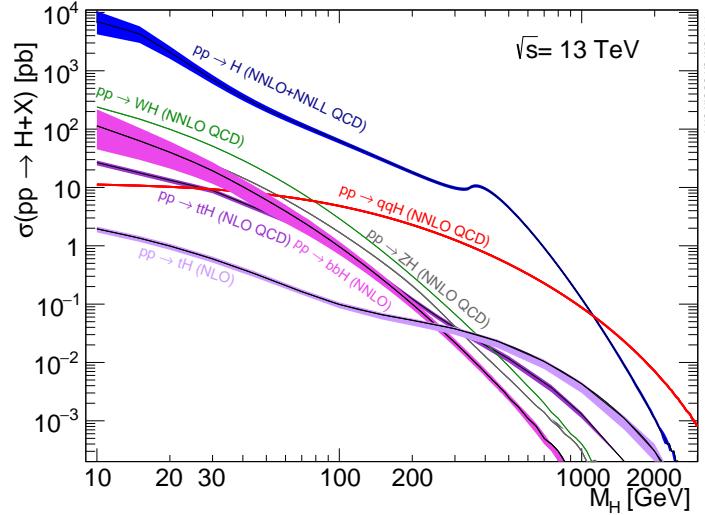


Figure 2.3: Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fussion, vector boson fusion, associated production with vector bosons, and top quark fusion.

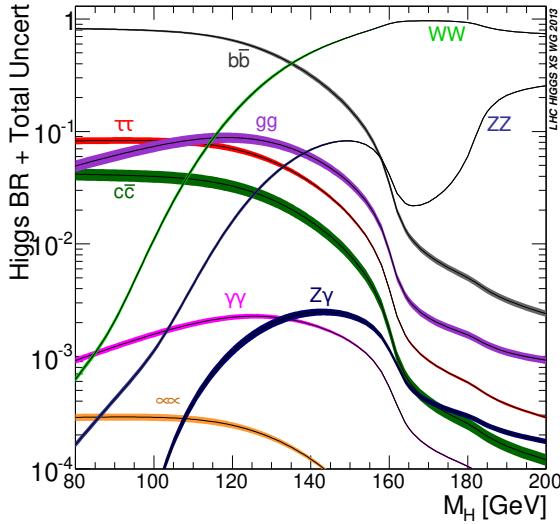


Figure 2.4: Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13 \text{ TeV}$ [26]. Uncertainties are shown in the shaded bands. At $m_H = 125 \text{ GeV}$, the Higgs predominantly decays to a pair of b -quarks, around 58% of the time. The leading subdominant decay mode is 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 [27, 28]. 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.

318 Chapter 3

319 The Large Hadron Collider and the
320 ATLAS Detector

321 Since the completion of its construction in 2008, the Large Hadron Collider (LHC) [29]
322 at CERN has extended the frontiers of particle physics through its unprecedented
323 energy and luminosity. The LHC accelerates protons around a 27 km ring until they
324 are travelling just 3 m s^{-1} slower than the speed of light, at which point they
325 are made to collide. The protons travel round the ring 11,000 times per second in
326 two concentric beams, which are guided by superconducting magnets cooled using
327 liquid helium to -271.3°C (1.9 K). The beams travel in opposite directions around
328 the ring and are crossed at four locations so that collisions between protons can
329 take place. Around these collision points four specialised detectors, ALICE [30],
330 CMS [31], LHCb [32] and ATLAS [33], are located to capture information about the
331 products of the collisions.

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

³³⁷ 3.1 The Large Hadron Collider

³³⁸ The LHC is operated in multi-year *runs* during which beams of protons are circulated
³³⁹ and collided. Between runs there are periods of shutdown while the accelerator and
³⁴⁰ detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC
³⁴¹ collided proton bunches, each containing more than 10^{11} particles, 20 million times
³⁴² per second, providing 7 TeV proton-proton collisions at instantaneous luminosities
³⁴³ of up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The centre-of-mass energy was increased to 8 TeV
³⁴⁴ towards the end of Run 1 in 2012. Run 2, which spanned in 2015–2018, further
³⁴⁵ increased the the proton-proton collision energy to 13 TeV. During Run 2 the bunch
³⁴⁶ spacing was reduced, leading to a collisison rate of 40 MHz. Over the course of
³⁴⁷ Run 2 a total usable integrated luminosity of 139 fb^{-1} was recorded. 2022 marked the
³⁴⁸ beginning of Run 3 which, with a higher center of mass energy and peak luminosity,
³⁴⁹ is expected to culminate in the approximate tripling of the dataset size. A summary
³⁵⁰ of key information about each run is listed in Table 3.1.

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

Table 3.1: Overview of the different LHC runs [34,35]. 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.

³⁵¹ An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is
³⁵² at the final stage of a chain of accelerators which incrementally step-up the energy
³⁵³ of incoming protons. The first accelerator is Linac4, a linear accelerator which
³⁵⁴ accelerates negative hydrogen ions to an energy of 160 MeV. Upon leaving Linac4,
³⁵⁵ the ions are stripped of both electrons and the resulting protons are fed into the
³⁵⁶ Proton Synchrotron Booster (PSB), which increases the energy of the protons to
³⁵⁷ 2 GeV. The protons leaving the PSB are passed to the Proton Synchrotron (PS),
³⁵⁸ which increases the energy to 26 GeV, and then from the PS to the Super Proton
³⁵⁹ Synchrotron (SPS) which further increases the energy to 450 GeV. Finally, the proton

³⁶⁰ beams are injected in the LHC where they are accelerated to their final energy of
³⁶¹ 6.5 TeV (for Run 2).

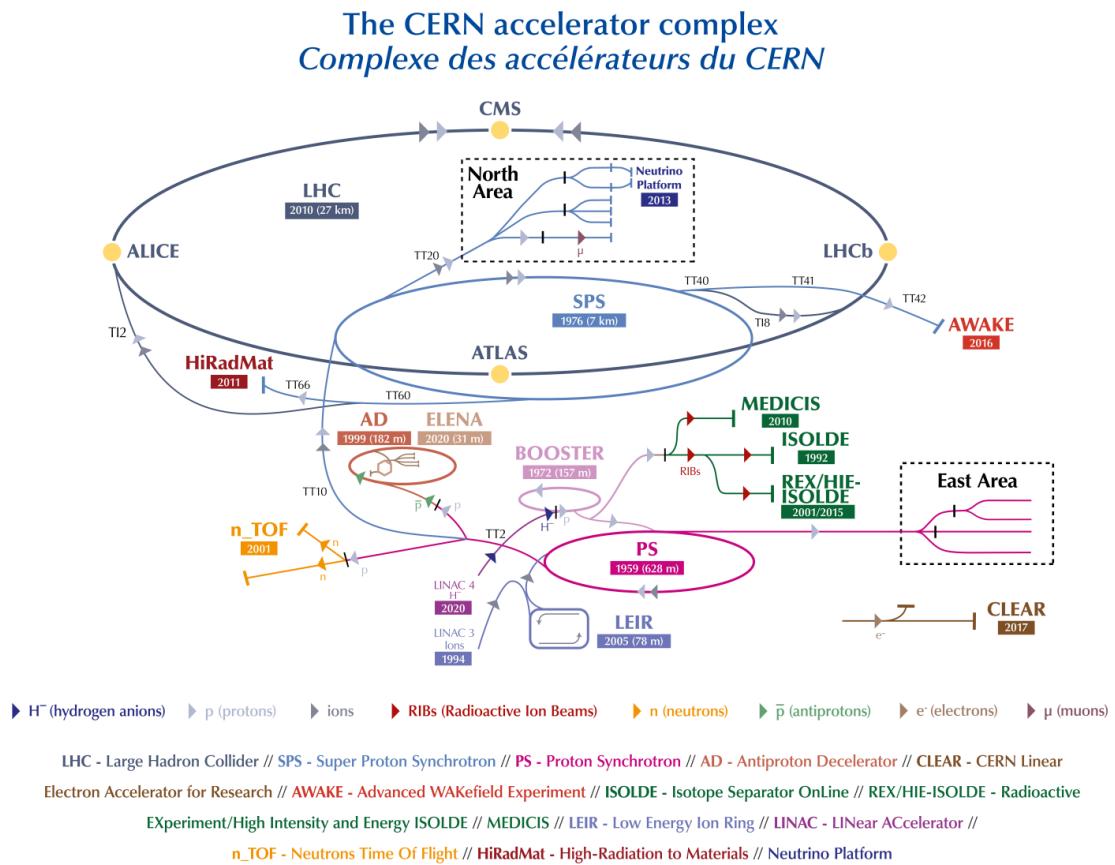


Figure 3.1: An overview of the CERN accelerator complex [36]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

³⁶² 3.2 Coordinate System & Collider Definitions

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

367 3.2.1 ATLAS Coordinate System

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

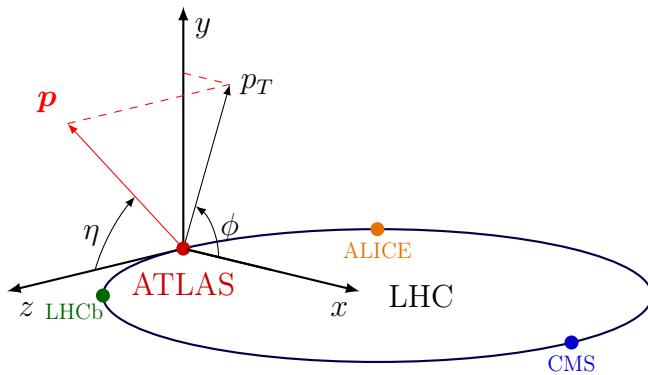


Figure 3.2: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [37].

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

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

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

379 The transverse momentum p_T of an object is the sum in quadrature of the momenta
380 in the transverse plane

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

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

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

383 3.2.2 Track Parameterisation

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

392 Impact parameter significances are defined as the IP divided by its corresponding
393 uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. When used in flavour tagging
394 (see Chapter 4), track IP significances are lifetime signed according to the track's
395 direction with respect to the jet axis and the primary vertex [39]. The sign IP
396 significances is positive if the track crosses the jet axis in front of the primary vertex
397 and negative if the crossing is behind the primary vertex.

398 3.2.3 Hadron Collider Definitions

399 Cross Section

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

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

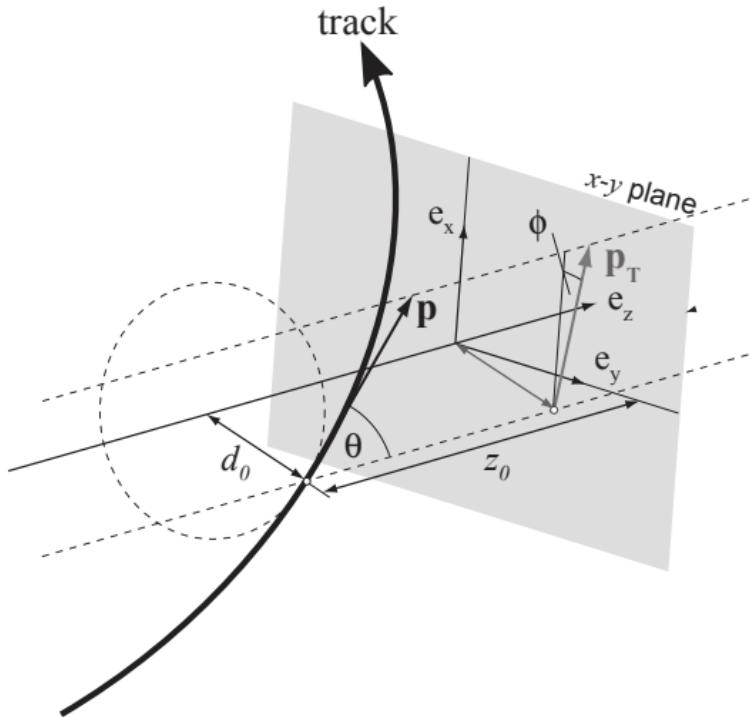


Figure 3.3: The track parameterisation used at the ATLAS detector. Five coordinates $(d_0, z_0, \phi, \theta, q/p)$ are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum \mathbf{p} and the transverse momentum p_T (defined in Eq. (3.2)). The basis vectors e_x , e_y and e_z are also shown. Reproduced from Ref. [38].

⁴⁰⁵ section can be factorised as

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

⁴⁰⁶ The partonic cross section can be calculated at sufficiently high energies such as
⁴⁰⁷ those found at the LHC, while the parton distribution functions (PDFs) have to be
⁴⁰⁸ extracted from experimental results.

⁴⁰⁹ Luminosity

⁴¹⁰ The total number of proton-proton collisions N is related to the total pp cross σ
⁴¹¹ section by the integrated luminosity L , as in

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

⁴¹² The instantaneous luminosity \mathcal{L} relates the cross section to the number of collisions
⁴¹³ per unit time. For two colliding bunched proton beams, it is defined as

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

⁴¹⁴ where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
⁴¹⁵ crossing frequency, and σ_x and σ_y are the rms width of the beam in the horizontal
⁴¹⁶ and vertical directions.

⁴¹⁷ The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In
⁴¹⁸ total, 139 fb^{-1} of usable physics data was collected over the three-year run. The
⁴¹⁹ uncertainty on the total integrated luminosity is 1.7% [40].

⁴²⁰ Pile-up

⁴²¹ At the centre of the ATLAS detector, bunches of more than 10^{11} protons meet at a
⁴²² small crossing angle. Each bunch-crossing is called an *event*. There is generally at
⁴²³ most one hard proton-proton scatter per event. Additional interactions are typically
⁴²⁴ relatively soft and are known as *pile-up*. Pile-up from interactions within the same
⁴²⁵ bunch-crossing is known as *in-time* pile-up while residual signatures from previous
⁴²⁶ bunch-crossings is known as *out-of-time* pile-up. The number of pile-up interactions

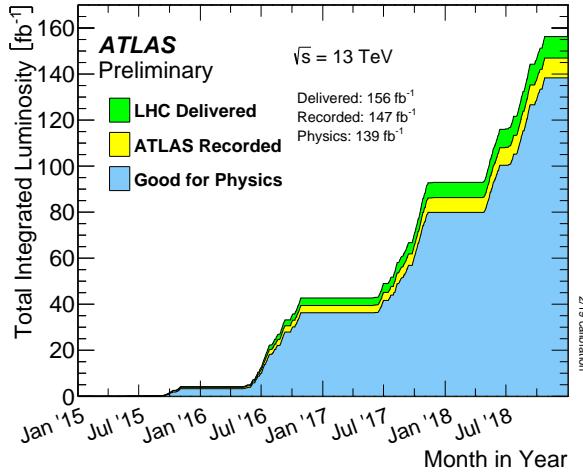


Figure 3.4: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [35]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.

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

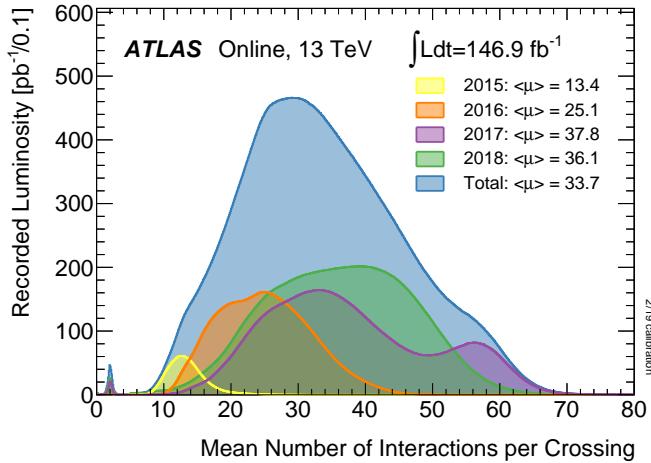


Figure 3.5: Average pile-up profiles measured by ATLAS during Run 2 [35]. Higher levels of pile-up are planned for Run 3.

⁴²⁹ 3.3 The ATLAS Detector

⁴³⁰ The ATLAS² detector is made up of several specialised sub-detectors which are
⁴³¹ arranged concentrically around the nominal interaction point at the centre of the
⁴³² detector. The detector is designed to cover nearly the entire solid angle around the
⁴³³ collision point. In this section a condensed overview of each sub-detector is given, in
⁴³⁴ order of increasing radial distance from the point of collision. The inner tracking
⁴³⁵ detector is described in Section 3.3.1, the electromagnetic and hadronic calorimeters
⁴³⁶ in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the trigger is
⁴³⁷ described in Section 3.3.4. More complete information on the detector can be found
⁴³⁸ in Ref. [33], while an overview of physics performance is given in [41].

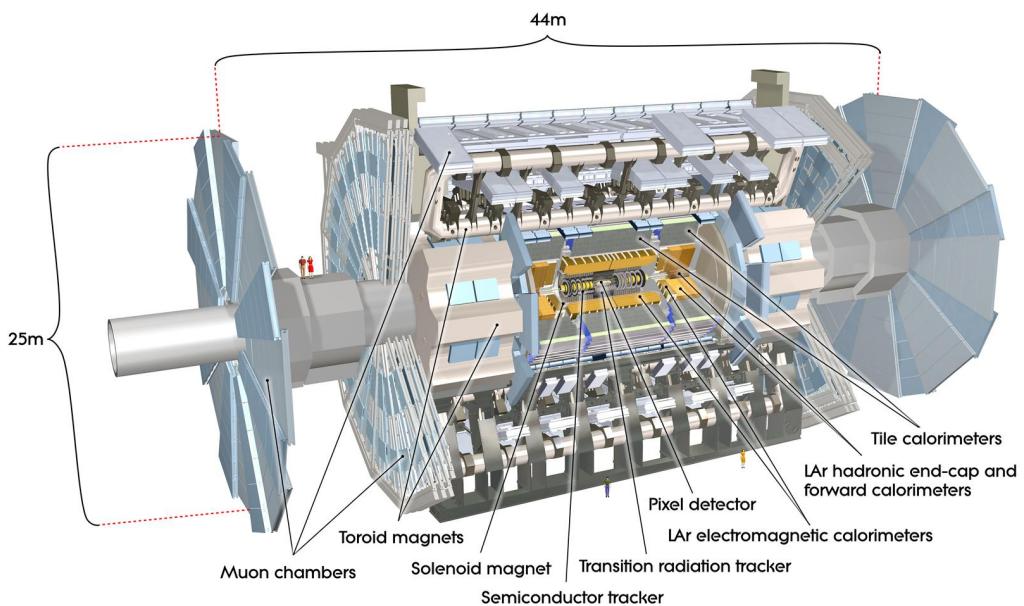


Figure 3.6: A 3D model of the entire ATLAS detector [42]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.

²A Toroidal LHC ApparatuS.

⁴³⁹ 3.3.1 Inner Detector

⁴⁴⁰ The inner-detector system (ID) provides high-resolution charged particle trajectory
⁴⁴¹ tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field,
⁴⁴² produced by a superconducting solenoidal magnet, which enables the measurement
⁴⁴³ of particle momentum and charge. After Run 3, the ID will be replaced by the
⁴⁴⁴ ITk [43, 44].

⁴⁴⁵ The inner detector is made up of several sub-systems, shown in Figs. 3.7 and 3.8. The
⁴⁴⁶ high-granularity silicon pixel detector covers the vertex region and typically provides
⁴⁴⁷ four spacepoint measurements per track. It is followed by the silicon microstrip
⁴⁴⁸ tracker (SCT), which usually provides a further four spacepoint measurements per
⁴⁴⁹ track. These silicon detectors are complemented by the Transition Radiation Tracker
⁴⁵⁰ (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$.

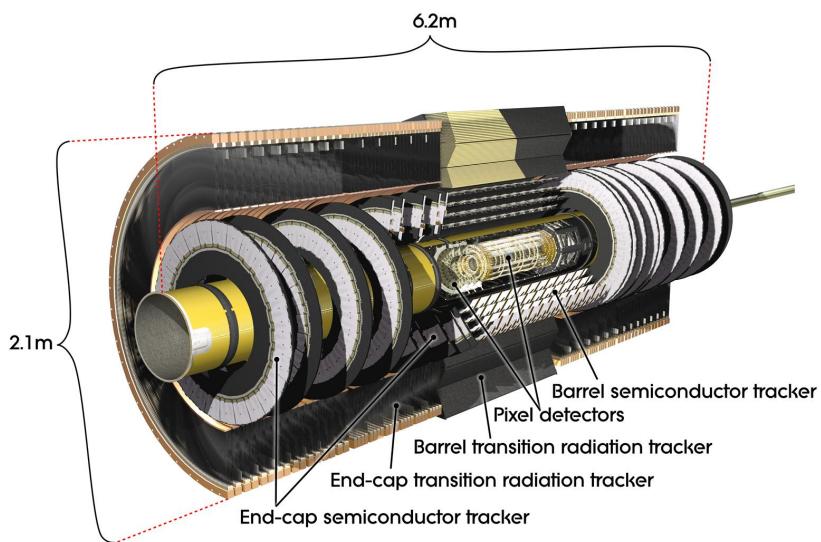


Figure 3.7: A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [45].

⁴⁵¹ The target inverse momentum resolution for the combined ID measurement is
⁴⁵² parameterised as a function of the track transverse momentum and polar angle [41].
⁴⁵³ The parameterisation is given by

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

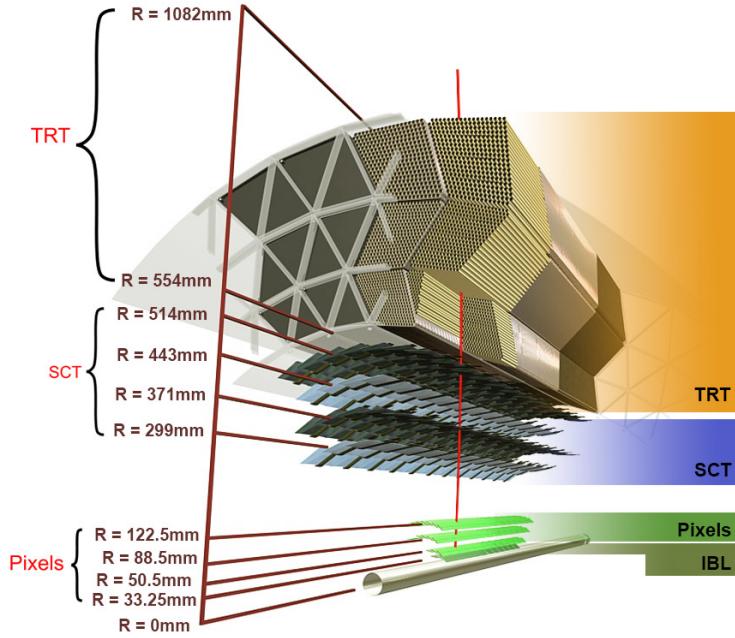


Figure 3.8: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [38].

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

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

459 Pixel Detector

460 The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
 461 from the beamline, and four disks on each side. The innermost barrel layer is
 462 the insertable B-layer (IBL), which was installed before Run 2 [46, 47] and lies
 463 approximately just 33 mm from the beam axis. The second-to-innermost layer is
 464 often referred to as the B-layer. The specification of the pixel detector determines the
 465 impact parameter resolution and the ability to reconstruct primary and secondary

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

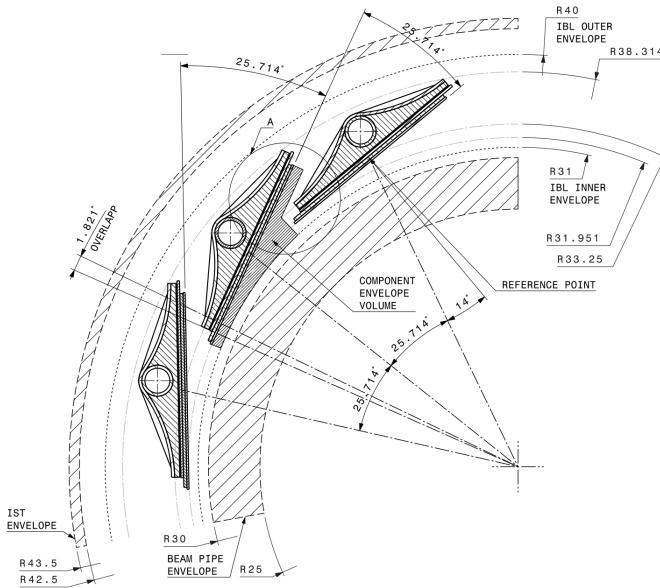


Figure 3.9: A schematic cross-sectional view of the ATLAS IBL [46].

471 Semi-Conductor Tracker (SCT)

472 The SCT is made up of four concentric barrel layers in the central region, and nine
 473 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
 474 with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate
 475 to be measured from a pair of strip measurements. The SCT typically provides four
 476 precision spacepoint measurements (eight strip measurements) per track in the barrel
 477 region. These have intrinsic uncertainties of 17 μm in the transverse direction $R\phi$, and
 478 580 μm in the longitudinal direction z [48]. The measurements provide a contribution
 479 to the measurement of charged particle momentum and impact parameter, along
 480 with vertex position. Charge-particle tracks can be distinguished if separated by
 481 more than $\sim 200 \mu\text{m}$.

482 Transition Radiation Tracker (TRT)

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

496 3.3.2 Calorimeters

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

507 Liquid Argon (LAr) Electromagnetic Calorimeter

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

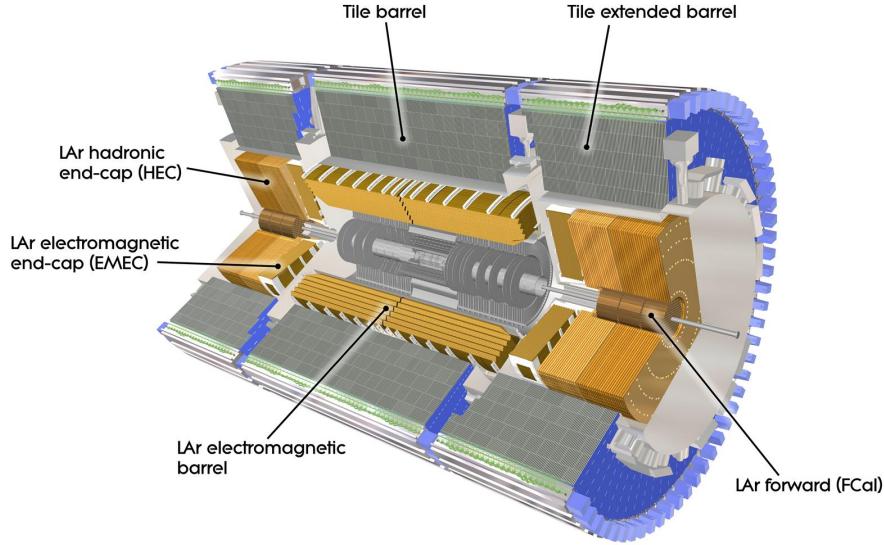


Figure 3.10: The ATLAS calorimeters [49]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic endcaps.

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

516 The energy resolution of the LAr calorimeter is made up of a sampling and a constant
 517 term, which are summed in quadrature to produce the overall energy resolution. The
 518 sampling term contributes approximately $10\%/\sqrt{E}$, while the constant term adds an
 519 additional 0.7%. Photons with moderate transverse energy $E_T \approx 50 \text{ GeV}$ have an
 520 overall energy resolution of at most 1.6% over most of the pseudorapidity range. At
 521 lower $E_T \approx 10 \text{ GeV}$, the resolution is degraded to approximately 5%. The resolution
 522 measurements are obtained from test beam data [41].

523 Hadronic Tile Calorimeter

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

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

530 The jet energy resolution of the HCal is parameterised as a function of the jet
 531 transverse energy

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

532 corresponding to a jet energy resolution of 10% at a jet p_T of approximately 100 GeV
 533 [50].

534 3.3.3 Muon Spectrometer

535 Due to their higher mass, muons easily pass unimpeded through the ID and calorime-
 536 ters and therefore require specialised detectors for their measurement. The Muon
 537 Spectrometer (MS) is made up of dedicated tracking and triggering hardware. The
 538 precision tracking system uses three layers of monitored drift tubes with a barrel
 539 region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$. The inner layers of
 540 the end-caps use cathode strip chambers to better cope with the high occupancy
 541 in the forward region. Precision tracking resolution is approximately 50 μm . The
 542 trigger system is comprised of resistive plate chambers in the barrel region covering
 543 $|\eta| < 1.0$ and thin gap chambers in the end-cap regions covering $1 < |\eta| < 2.4$. A set
 544 of three superconducting air-core toroidal magnets, each made up of eight coils, is
 545 used in each of the barrel and end-caps to deflect the muons as they pass through
 546 the MS, allowing their momentum and charge to be measured from the direction
 547 and magnitude of curvature. The toroidal magnets generate a field which is largely
 548 orthogonal to the muon trajectories which allows for maximum deflection. The
 549 transverse momentum resolution has been measured to be approximately 1.7% in the
 550 central region for low- p_T muons, increasing to 4% for high- p_T muons in the forward
 551 regions [51].

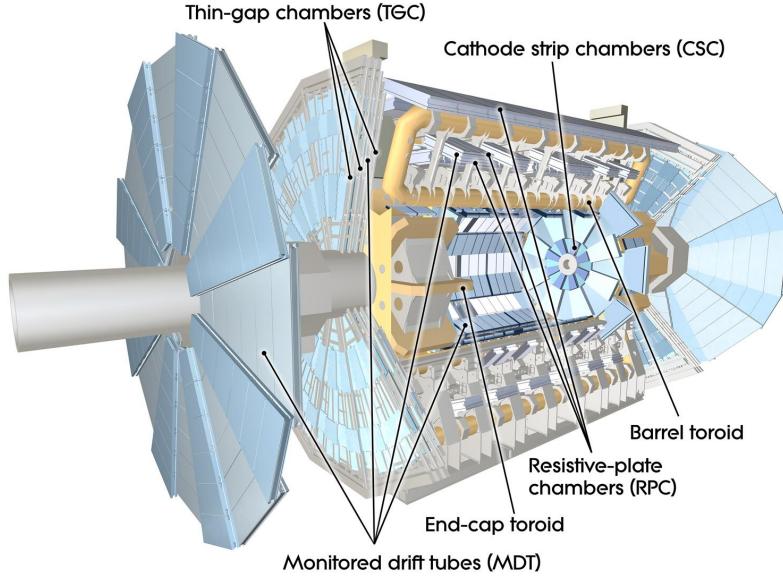


Figure 3.11: The ATLAS muon spectrometer [52].

552 3.3.4 The Trigger

553 The 25 ns bunch spacing used over the course of Run 2 corresponds to a bunch-
 554 crossing or event rate of 40 MHz (see Table 3.1). If the full information for the
 555 detector was written out for each event, this would correspond to the generation
 556 of 60 TB of data each second. This is more than can be feasibly read out from
 557 the hardware, processed and stored, requiring the use of a trigger system which
 558 quickly makes a decision about whether or not an event is potentially interesting
 559 and should be kept for further analysis. The trigger system is comprised of two
 560 levels which search for signs of electrons, muons, taus, photons, and jets, as well as
 561 events with large total or missing transverse energy. The hardware-based Level-1
 562 (L1) trigger uses coarse information from the calorimeters and MS to accept events at
 563 an average rate of 100 kHz approximately 2.5 μ s after the event. After the L1 trigger,
 564 the software-based High Level Trigger (HLT) makes use of 40 000 CPU cores to make
 565 a final selection on surviving events in approximately a few hundred milliseconds.
 566 The final event read-out rate is approximately 1.2 kHz, corresponding to 1.2 GB s^{-1}
 567 of permanent data storage. More information is provided in [53].

568 3.4 Reconstructed Physics Objects

569 Event reconstruction is the process of analysing the output from the detector to
570 determine the type and properties of particles present in an event. The reconstructed
571 event provides information about the underlying physics process that led to these
572 observable final state particles. Events passing the trigger selection (described in
573 Section 3.3.4) undergo offline reconstruction, which makes use of the full information
574 from the detector. Reconstruction and analysis of events relies on the extensive
575 ATLAS software stack, see Ref. [54] for more information.

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

578 3.4.1 Tracks

579 The reconstructed trajectories of charged particles are referred to as *tracks*. Track are
580 reconstructed from the energy depositions (called *hits*) left by the particles as they
581 traverse the the inner detector. Tracks are widely used for a variety of downstream
582 applications, including vertexing and jet tagging, so their accurate reconstruction
583 is a critical task. A comprehensive introduction to ATLAS tracking is available
584 in Ref. [55], while specific optimisations for dense environments are detailed in
585 Refs. [56, 57]. An overview of track reconstruction is given below.

586 Space-point Formation (Clustering)

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

593 Track Finding

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

610 Ambiguity Solving

611 Ambiguity solving was introduced as part of the ATLAS New Tracking effort [55],
612 which was intended to improve track reconstruction performance in dense envi-
613 ronments. In the ambiguity solver, track candidates are processed individually in
614 descending order of a track score. The track score quantifies the likelihood of the
615 track corresponding to the trajectory of a real particle. Scoring uses a number of
616 variables, including the number and positions of hits (preferring hits in more precise
617 regions of the detector), the transverse momentum of the track and the track fit
618 quality. The track fit quality describes the quality of the track as the χ^2 divided
619 by the number degrees of freedom on the track. A preference for high transverse
620 momentum tracks promotes the successful reconstruction of the more physically
621 interesting energetic particles, and suppresses the large number of wrong hits assigned
622 to low momentum tracks. The ambiguity solver also penalises tracks with missing
623 hits on the innermost detector layers.

624 During the processing of a given highest-scoring track candidate, the track is cleaned
625 (whereby problematic hits are removed), and, if the resulting track satisfies the quality
626 selection criteria, a high precision fit of the track parameters using the surviving hits
627 is performed. The high precision fit makes full use of all available information, and
628 uses an updated position and uncertainty estimate for each cluster obtained from
629 a Neural Network (NN) [58]. If the track has reached this stage without rejection
630 by passing various quality regiments, it is re-scored and returned to the list of track
631 candidates. If the same track is then processed again without requiring modification,
632 it is added to the final track collection. Track candidates that fall below a certain
633 quality cut are rejected. This selection does allow for the possibility of a track having
634 small number of shared hits.

list shared
hit cut?

635 Neural Network Cluster Splitting

636 As part of track cleaning, shared hits are classified by a NN to determine if they are
637 compatible with the characteristic features of a merged cluster [56, 58]. A merged
638 cluster is one made up of a combination of energy deposits from more than one
639 particle, which have become merged due to the closeness of the associated particles
640 and the limited resolution of the detector. While in general this event is rare, it
641 is common for clusters to become merged in dense environments, as discussed in
642 Section 4.1. If the cluster is predicted to be merged it is labelled as being freely
643 shareable, or *split*. Hits not compatible with the merged hypothesis can still be
644 shared by a limited number of tracks, but come with a penalty for the track which
645 may hinder its acceptance into the final track collection.

646 Pseudotracking

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

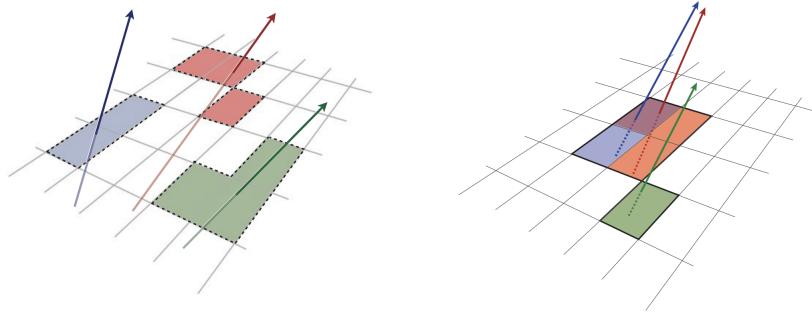


Figure 3.12: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [57].

654 ATLAS detector, with perfect hit-to-track association and and track reconstruction
 655 efficiency. The approach was introduced in Ref. [59] as a way to obtain a fast
 656 approximation of tracking reconstruction for simulated data, however the technique
 657 has become a useful tool for studying tracking performance in general [56].

658 3.4.2 Vertices

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

667 Primary Vertices

668 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices
 669 are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [60].
 670 In Run 3, the IVF will be replaced with an adaptive multi-vertex finder (AMVF) [61].

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

673 **Secondary Vertices**

674 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient
675 distance from the primary vertex to be resolved from the primary vertex (see
676 Section 4.1.1). Two widely used secondary vertexing tools are used within ATLAS:
677 SV1 and JetFitter [62]. Each attempts to reconstruct secondary vertices inside a jet
678 using the tracks associated to that jet (see Section 3.4.3 for more information about
679 track association). SV1 by design attempts to reconstruct only a single inclusive
680 vertex per jet. This inclusive vertex groups all b -hadron decay products, including
681 tracks from the b -hadron decay itself and tracks from $b \rightarrow c$ decays. The second tool,
682 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary
683 vertices from b -hadron decays are reconstructed separately to tertiary vertices from
684 $b \rightarrow c$ decay chains.

685 **3.4.3 Jets**

686 Jets are an aggregate reconstructed object corresponding to a collection of collimated
687 stable particles which results from a decay chain of an quark or gluon progenitor. Jets
688 are built by clustering constituent objects (e.g. tracks or calorimeter clusters) using
689 a jet finding algorithm, for example the anti- k_t algorithm [63], which is implemented
690 in FASTJET [64].

691 Objects can be associated to jets in one of two ways. The first is via a geometrical
692 matching in $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, where $\Delta\eta$ and $\Delta\phi$ are the differences in
693 pseudorapidity and azimuthal angle between the jet and the object. The second is
694 via a ghost association [65], where the object is assigned a negligible momentum and
695 re-clustered into the jet after its formation.

696 EMTopo Jets

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

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

700 where E_{cell} is the energy measured in a given calorimeter cell, and $\sigma_{\text{noise, cell}}$ is the
 701 expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are
 702 formed from a seed cell with a large S_{cell} , and expanded by iteratively adding neigh-
 703 bouring cells with a sufficiently large energy significance. Collections of topoclusters
 704 are then clustered into a jet using the anti- k_t algorithm with a radius parameter of
 705 0.4 (small- R jets) or 1.0 (large- R jets). More information is available in Ref. [66].

706 Particle Flow Jets

707 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [67] using
 708 the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate
 709 information from both the ID and the calorimeters, improving the energy resolution
 710 at high transverse momenta and reducing pile-up contamination. The PFlow jet
 711 energy scale is calibrated according to Ref. [68].

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

717 Jet flavour labels are assigned according to the presence of a truth hadron within
 718 $\Delta R(\text{hadron}, \text{jet}) < 0.3$ of the jet axis. If a b -hadron is found the jet is labelled a b -jet.
 719 In the absence of a b -hadron, if a c -hadron is found the jet is called a c -jet. If no b -
 720 or c -hadrons are found, but a τ is found in the jet, it is labelled as a τ -jet, else it is
 721 labelled as a light-jet.

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

723 Large- R Jets

724 Large- R jets have a radius parameter $R = 1.0$ and are built by clustering topological
 725 calorimeter clusters using the anti- k_t algorithm [69]. The large radius parameter
 726 is especially useful for containing the decay products of a boosted Higgs boson, as
 727 discussed in Chapter 7. Due to their large size, large- R jets benefit from a grooming
 728 procedure called trimming which remove soft contaminants inside the jet [70, 71].
 729 Trimming aims to remove jet constituents from pile-up and the underlying event,
 730 which helps to improve the jet mass resolution and its robustness to varying levels
 731 of pile-up. The jet mass is computed using a combination of information from the
 732 calorimeters and ID, and a calibration to data is applied [72].

733 Track-jets

734 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm and
 735 are used in the analysis described in Chapter 7. The radius parameter is allowed
 736 to vary with transverse momentum such that a broader cone (up to $R = 0.4$) is used
 737 for low- p_T track-jets and a narrower cone (down to $R = 0.02$) for high- p_T
 738 track-jets [73, 74]. The narrower cone is better suited to clustering highly collimated
 739 jet constituents at high- p_T . Truth flavour labels for track-jets are derived using the
 740 same $\Delta R(\text{hadron}, \text{jet}) < 0.3$ matching scheme as used for PFlow jets.

741 3.4.4 Leptons

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

745 Electrons

746 Electrons candidates are reconstructed by matching PV-compatible³ inner detector
 747 tracks to topological calorimeter clusters. The track-cluster matching criteria takes
 748 into account the significant energy loss of the electron due to bremsstrahlung. If a

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

749 match is found, a refit of the track is performed using the Gaussian Sum Filter (GSF)
 750 [75], which better handles trajectory reconstruction in the presence of bremsstrahlung.
 751 Various identification criteria are then applied to the candidates using a likelihood-
 752 based (LH) method to improve purity. These include requirements on the track
 753 quality and cluster matching, the shape of electromagnetic shower in the ECal,
 754 leakage into the HCal, and the amount of transition radiation detected in the TRT.
 755 Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may
 also be applied. A full description can be obtained from Ref. [76].

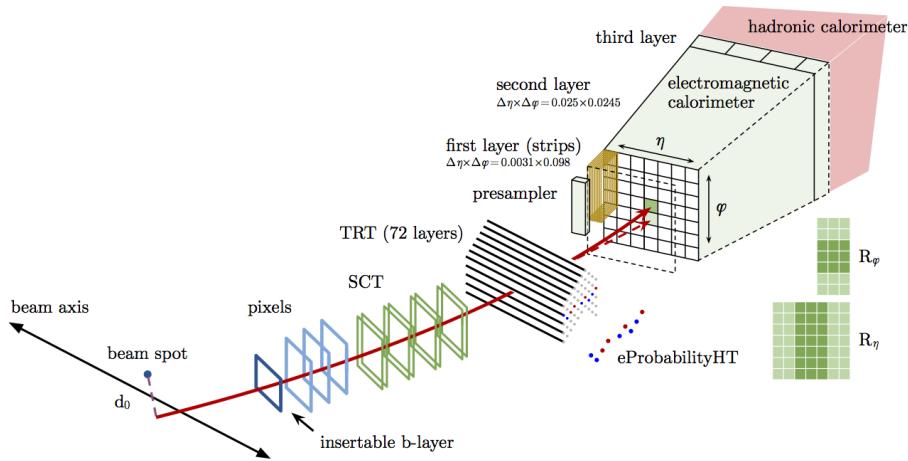


Figure 3.13: A sketch of electron reconstruction using the ATLAS detector [76]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

756

757 Muons

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

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

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

779 More information on muon reconstruction, identification and isolation can be found
780 in Ref. [77].

781 3.4.5 Missing Transverse Momentum

782 An imbalance in the final state transverse momentum can occur as a result of
783 incomplete measurement of the final state particles. In particular, neutrinos are
784 not measured by the detector and contribute to the missing transverse momentum
785 $\mathbf{E}_T^{\text{miss}}$. Incomplete detector acceptance and inaccuracies in the reconstruction of the
786 final state can also contribute to the missing transverse momentum of an event. In
787 order to calculate the missing transverse momentum, the negative vector sum of
788 the momentum of all photons, leptons and small- R jets with $p_T > 20 \text{ GeV}$ is taken.
789 The momenta of tracks associated to the primary vertex are also taken into account.
790 The magnitude of $\mathbf{E}_T^{\text{miss}}$ is written E_T^{miss} . More information about missing transverse
791 momentum reconstruction is provided in [78].

792

Chapter 4

793

Tracking and flavour tagging

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

803 The current ATLAS flavour tagger, DL1r [79], is a deep neural network which
804 accepts as inputs the outputs of a number of independently optimised *low-level*
805 algorithms [62]. Correspondingly, DL1r is referred to as a *high-level* tagger (i.e. one
806 that uses a multivariate approach to combine the outputs of the low-level taggers).
807 Each of these low-level algorithms reconstructs a distinct feature of the experimental
808 signature of heavy flavour jets using the tracks associated to the jet. The low-level
809 algorithms are a combination of manually optimised reconstruction algorithms, for
810 example the SV1 and JetFitter algorithms that reconstruct displaced decay vertices,
811 and trained taggers such as RNNIP and DIPS that use the IP and hit information
812 from a variable number of tracks to identify the flavour of the jet [62, 80–82].

813 In addition to DL1r, another widely used high-level tagger is the MV2c10 algorithm
814 [62, 79, 83]. This tagger is used in the analysis described in Chapter 7. Similar to
815 DL1r the MV2c10 algorithm takes inputs from the outputs of a number of low-level
816 algorithms (IPxD, SV1 and JetFitter). The outputs of the low-level algorithms

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

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

826 4.1 b -hadron Reconstruction

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

829 4.1.1 Decay Topology

830 b -hadrons are quasi-stable bound states of a bottom quark and one or more lighter
 831 quarks. Collectively, these are the B -mesons (e.g. $B^+ = u\bar{b}$, $B^0 = d\bar{b}$) and baryons
 832 (e.g. $\Lambda_b^0 = udb$). After a b -quark is produced as the result of a proton-proton collision,
 833 they quickly hadronise. The hadronisation process is hard – around 70-80% of
 834 the b -quark’s momentum is passed to the b -hadron, with the rest being radiated
 835 as prompt hadronisation or fragmentation particles. See Ref. [84] for a more in
 836 depth discussion on hadronisation and the closely related process of fragmentation.
 837 Henceforth the combined hadronisation and fragmentation products will be referred
 838 to collectively as fragmentation.

839 b -hadrons are interesting objects of study due to their relatively long proper lifetimes
 840 $\tau \approx 1.5$ ps [85]. This lifetime corresponds to a proper decay length $c\tau \approx 450$ μm . In
 841 the rest frame of the detector, the typical b -hadron travels a distance

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

842 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.

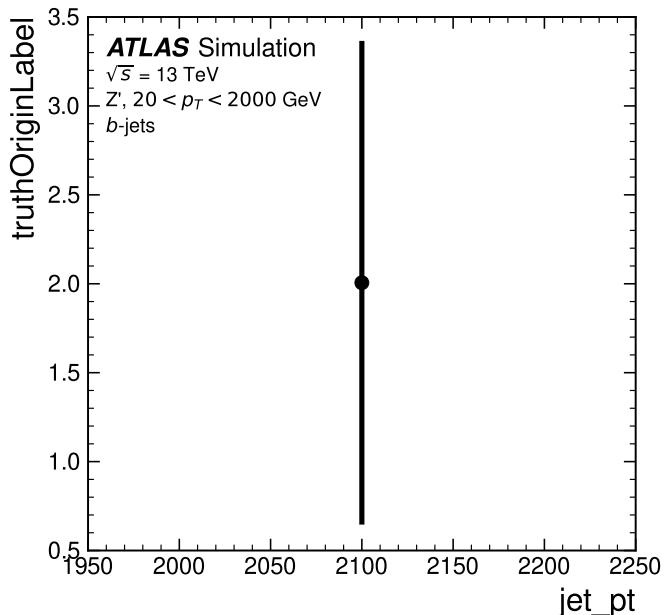


Figure 4.1: The truth b -hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed b -jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

b -hadrons decay weakly to on average four or five collimated stable particles [86]. 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 occur due to its lifetime. The primary feature is the presence of a high mass secondary vertex that is significantly displaced from the primary vertex. Reconstruction of these vertices from tracks with common points of spatial origin is a common approach used in the identification of b -jets.

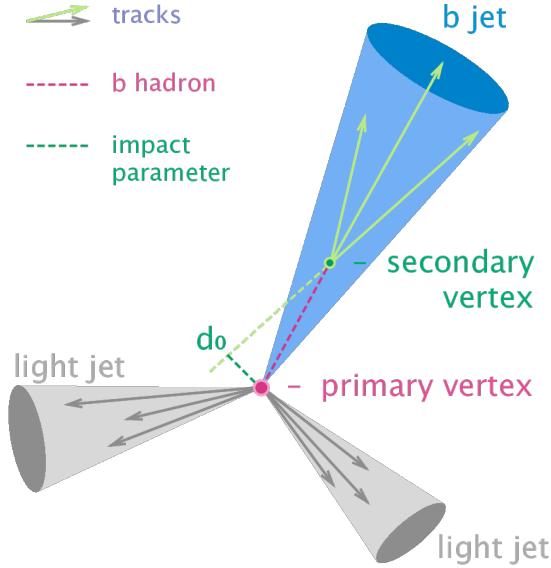


Figure 4.2: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [87]. The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

Additional signatures of b -hadrons are as follows. Associated tracks and SVs can have a large transverse impact parameter d_0 as a result of the b -hadron displacement (as shown in Fig. 4.2). Since it is common for the b -hadron to decay to a c -hadron with non-negligible lifetime, tertiary vertices can be found within b -jets resulting from $b \rightarrow c$ decay chains. The b -hadron also decays semileptonically in approximately 23% of cases [16]. The presence of a reconstructed electron or muon inside a jet can also be a key indicator that the jet was instantiated by a b -hadron.

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

4.1.2 Challenges Facing b -hadron Reconstruction

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

($p_T > 200$ GeV) this task becomes difficult due to a combination of effects. As the 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.

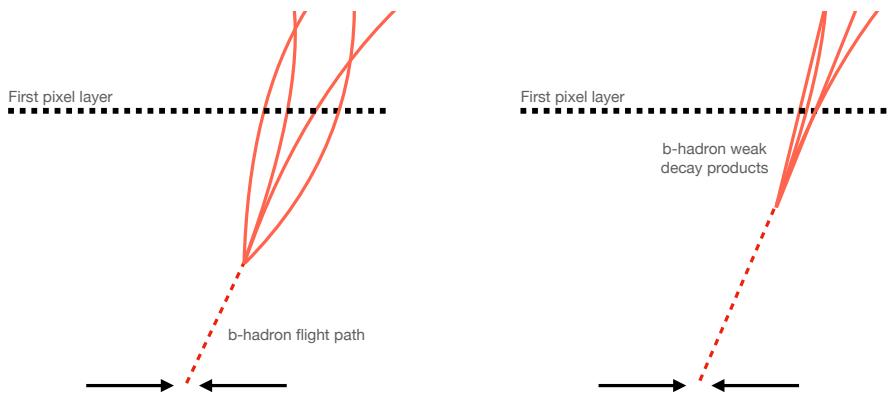


Figure 4.3: At lower p_T (left) the decay length of the b -hadron is reduced, and the resulting decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching 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

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

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

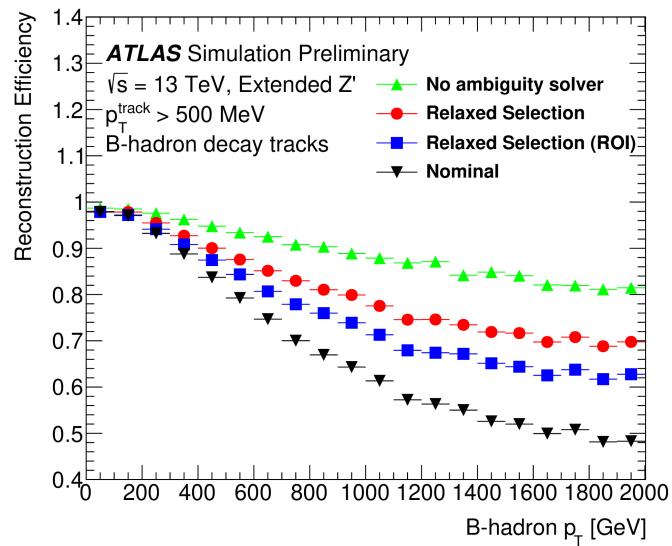


Figure 4.4: b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [88]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T b -hadrons, 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.

903 The second part of the problem is that, due to the high multiplicity of clusters available
 904 for assignment in the vicinity of the typical high energy b -hadron decay track, and
 905 also given the strong positive bias of the ambiguity solver towards those tracks
 906 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.

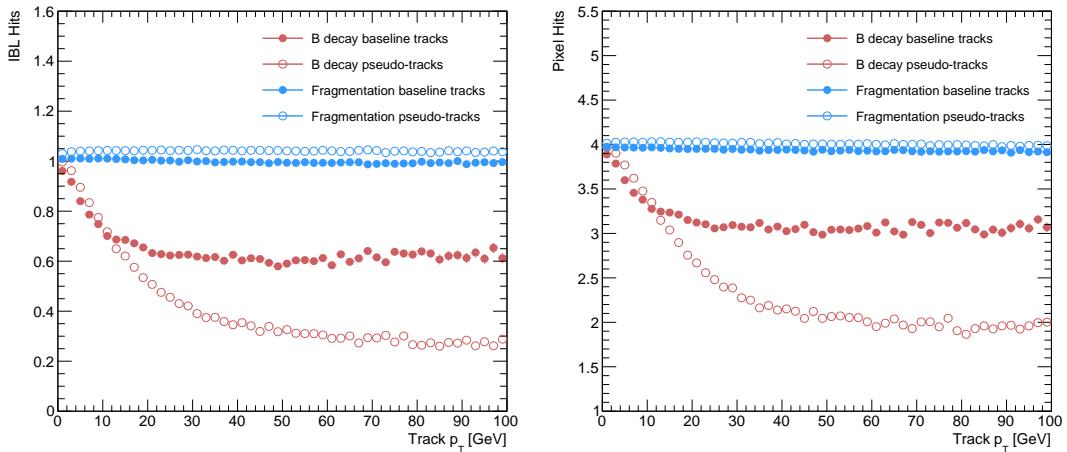


Figure 4.5: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. Baseline tracks are those produced in the standard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

cause of the observed drop in b -tagging efficiency at high energies, however further 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

4.2 Investigations into High p_T b -hadron Tracking

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

4.2.1 Shared Hits

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

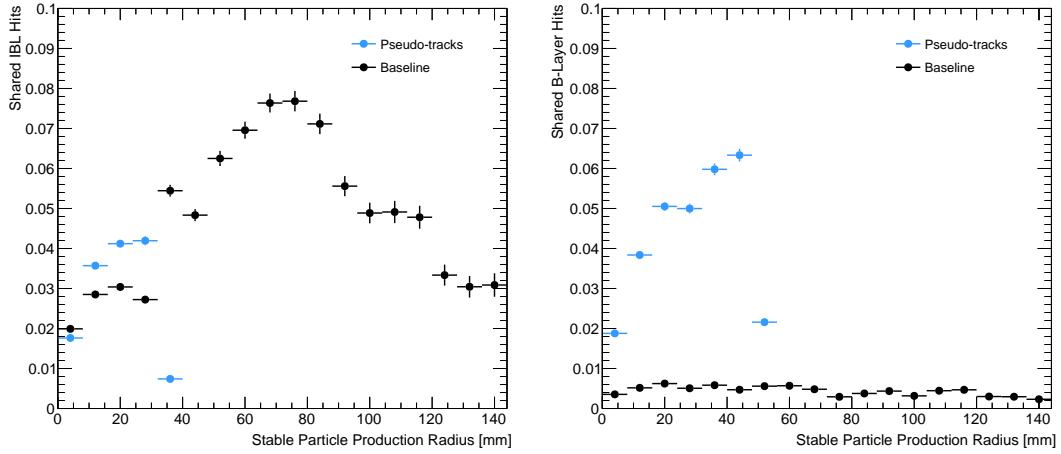


Figure 4.6: The fraction of hits which are shared on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudotrack represent the ideal performance given the ATLAS detector, see Section 3.4.1.

950 4.2.2 Global χ^2 Fitter Outlier Removal

951 This section documents ongoing progress into improvement of hit-to-track assignment
 952 by using the Global χ^2 Fitter (GX2F) to identify and prevent incorrect hits from
 953 being assigned to tracks during the track fit. This is in contrast to a previously
 954 investigated approach [89] which attempted to identify and remove wrong hits after
 955 the reconstruction of the track. As part of the track fit, an outlier removal procedure
 956 is run, in which suspicious hits are identified and removed.

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

967 Implementation

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

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

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

982 Cut Optimisation

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

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

995 1 TeV track, the rate to assignment good hits to the track increases by approximately
 996 10%, while the rate to assign incorrect hits decreases by approximately 16%. The
 997 improvements are also observed when looking inclusively in all tracks, which avoids
 998 the need for a specific b -jet region-of-interest selection.

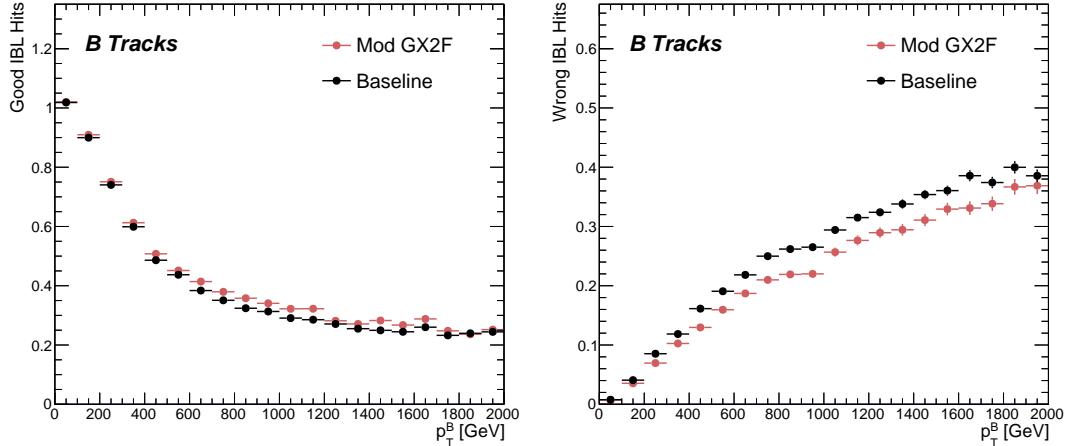


Figure 4.7: The rate to assign good (left) and wrong (right) IBL hits as a function of b -hadron p_T for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red). For each track, the corresponding p_T bin is filled with the number of good or wrong hits and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

999 An improvement, though modest, of all track parameter resolutions and pulls is
 1000 observed. The improvement for the transverse impact parameter pull is shown in
 1001 Fig. 4.8. The results demonstrate an improvement in hit assignment, unchanged
 1002 reconstruction efficiency, and modest improvement in track parameter resolutions
 1003 and pulls. In addition, the truth match probability of track is unchanged, suggesting
 1004 that there is no increase in fake track rates. The changes are expected to have a
 1005 negligible impact on computational resources.

1006 4.3 Conclusion

1007 In this section, the difficulties facing efficient and accurate track reconstruction,
 1008 and hence performant b -tagging, have been outlined. The ambiguity solver, which
 1009 attempts to clean or reject tracks which have an excessive number of shared hits,
 1010 is shown to be overly aggressive in the removal of b -hadron decay product track
 1011 candidates. The ambiguity solving process relies on a complicated pre-defined

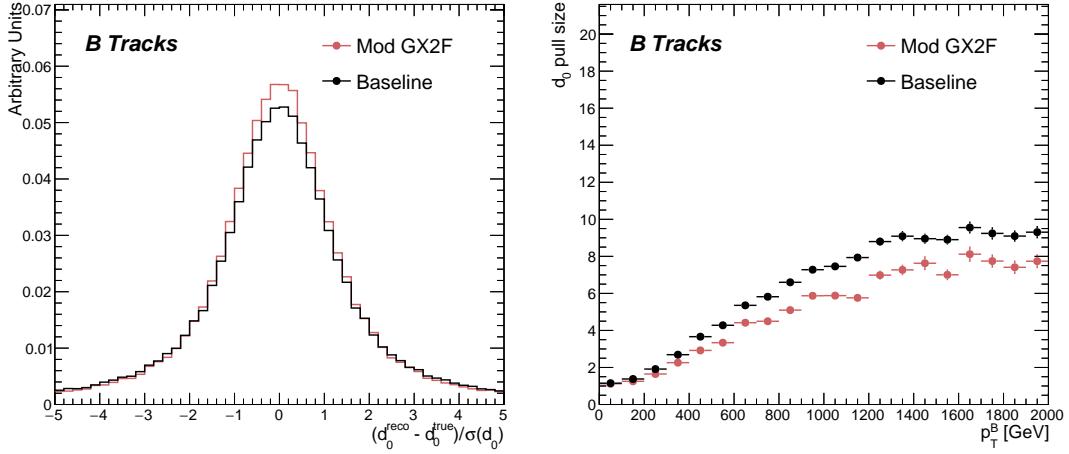


Figure 4.8: (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified GX2F tracks. (right) The absolute value of the d_0 pull as a function of b -hadron transverse momentum.

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. [88].

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

1027 **Chapter 5**

1028 **Track Classification MVA**

1029 The chapter details work on implementing a multivariate algorithm (MVA) to predict
1030 the truth origin of reconstructed tracks. An introduction to formalisms of machine
1031 learning is given in Section 5.1. In Section 5.3, the truth origin label is defined,
1032 and in Section 5.4 these labels are used to train a machine learning model that can
1033 effectively discriminate between good and fake tracks. Several studies motivated this
1034 work by demonstrating that at high p_T , b -tagging performance was degraded by the
1035 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
1036 could be trained to detect fake tracks, these could be removed before their input to
1037 the b -tagging algorithms with the aim of improving performance.

1038 **5.1 Machine Learning Background**

1039 Over the past few decades, machine learning (ML) techniques have become increasing-
1040 ly popular in high energy physics experiments due the increased volumes of
1041 high-dimensional data and improvements in the techniques used (in particular deep
1042 learning). Machine learning is the process by which a computer program uses data
1043 to learn suitable parameters for a predictive model. This is opposed to explicitly
1044 providing instructions on how to perform a task. A subfield known as *supervised*
1045 *learning* is used in this work, and consists of exposing a model to a large number of
1046 labelled examples in order to extract relationships between the input data and their
1047 labels. These relationships are often complex, and explicitly programmed rules can
1048 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 [90, 91]. 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 θ , the model outputs a prediction $\hat{y}_i \in (0, 1)$ for the true label y_i , as in

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

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

5.1.1 Neural Networks

Neural networks (NNs) are a common choice for the machine learning model f since they have the ability to approximate any function [92] and are easy to train via backpropagation [93].

1078 Artificial Neurons

1079 The basic functional component of a NN is the *artificial neuron* or node, which is
 1080 loosely inspired by a mathematical model of a biological neuron [94, 95]. A diagram
 1081 of an artificial neuron is shown in Fig. 5.1 Each neuron is defined by its parameters
 1082 or *weights* θ and a choice of activation function. Each neuron takes a fixed number
 1083 of inputs and computes the dot product of the input and weight vectors $x^T \theta$ and
 1084 additionally adds a constant bias term θ_0 . This term plays the role of a trainable
 1085 constant value that is independent of the inputs.

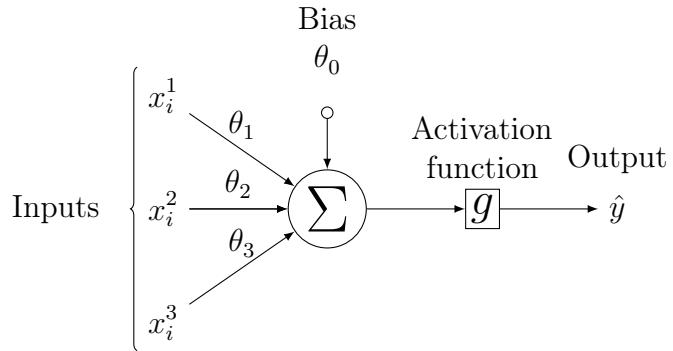


Figure 5.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression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

1086 The output of the dot product and bias term z is fed into an activation function
 1087 g . The activation function has several uses, most notably acting as a source of
 1088 non-linearity and bounding the output of the neuron. Some common activation
 1089 functions (sigmod, tanh, ReLU and SiLU) are shown in Fig. 5.2. The choice of
 1090 activation function can have implications for the performance and convergence of
 1091 the network, since the gradient of g is used to compute the weight updates during
 1092 training. This is also why input data is typically normalised to have zero mean and
 1093 unity variance [96].

1094 Networks

1095 Several neurons are linked together in layers to form a neural network. The inputs
 1096 are propagated layer-by-layer through the network until reaching the final output

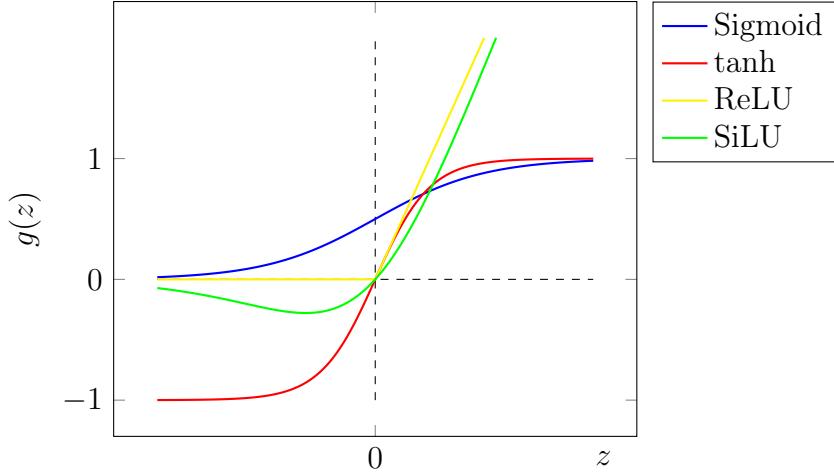


Figure 5.2: The output of several common choices for the activation function $g(z)$ of an artificial neuron. The input z is the output of the dot product between the activation and the weights, plus a bias term.

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

1112 problems, the binary cross entropy given in Eq. (5.3) loss if often used.

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

1113 Since the model f is differentiable, the error for each parameter θ_i can be computed by
1114 taking the partial derivative of L with respect to the parameter. Updated parameters
1115 θ'_i are calculated by updating the original parameter in the direction which reduces
1116 the loss.

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

1117 The hyperparameter α is known as the *learning rate* and dictates the size of the
1118 step taken in the direction of the slope. The errors for each parameter are efficiently
1119 calculated using the backpropagation algorithm [93]. The process of updating weights
1120 is repeated until the weights converge, which means the network is trained. In practice,
1121 small batches of the input data are shown to the network at a time. For each batch
1122 the average loss is calculated and the network's weights are updated. There are many
1123 extensions and variations of the gradient descent algorithm. This work uses the Adam
1124 optimiser which adds momentum to the weight updates (dampening oscillations)
1125 and an adaptive per-parameter learning rate [97].

1126 5.2 Graph Neural Network Theory

1127 Graph neural networks are a more sophisticated neural network model (see Sec-
1128 tion 5.1.1) that are designed to operate on graph structured data. A brief introduction
1129 to GNNs is provided in this section following the formalism in Ref. [98].

1130 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 $\mathcal{E} =$
1131 $\{e_i\}_{i=1:N^e}$, and a global representation u . Each node represents an individual object,
1132 and edges are directed connections between two nodes, called the *sender* and *receiver*
1133 nodes. The connectivity of the graph therefore encodes information about the
1134 relationships between objects that exist in the graph.

1135 A single graph network layer consists of three separate update functions ϕ^e , ϕ^h and
1136 ϕ^u one for each of the nodes, edges, and global graph representation, and similarly
1137 three aggregation functions $\rho^{e \rightarrow h}$, $\rho^{e \rightarrow u}$ and $\rho^{h \rightarrow u}$. The aggregation functions combine

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

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

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

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

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 ϕ^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

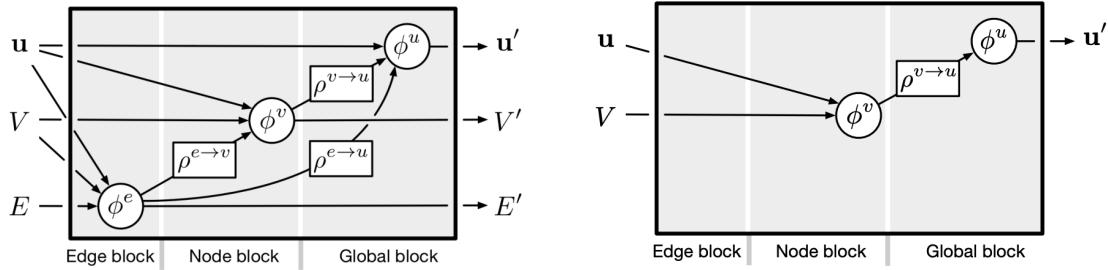


Figure 5.3: The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [99]. Reproduced from Ref. [98].

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 $\text{TMP} < 0.75$ ¹ 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.

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

Truth Origin	Description
Pile-up	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 [57]. A truth-matching probability of less than 0.5 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions, K_S^0 and Λ^0 decays, and hadronic interactions.

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

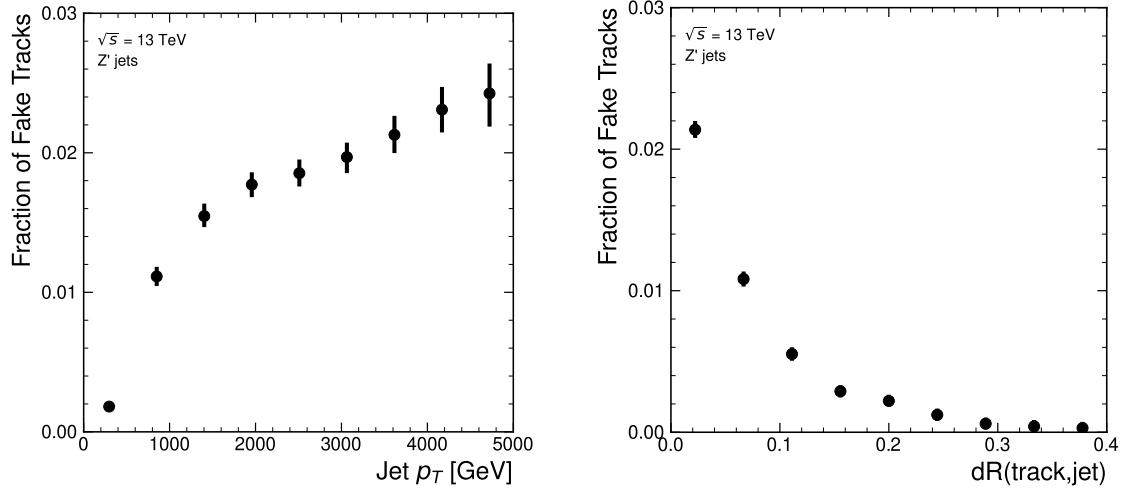


Figure 5.4: Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track},\text{jet})$ (right). The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases.

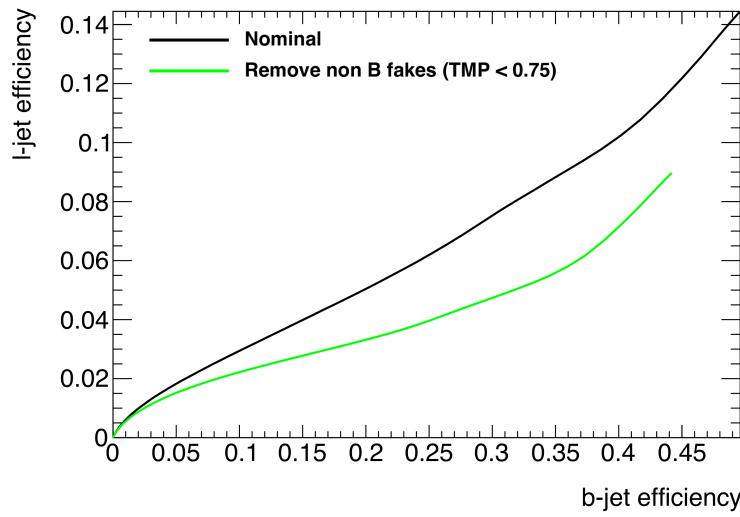


Figure 5.5: The light-jet efficiency of the low level tagger SV1 for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$, as a function of b -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a b -hadron are removed.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

1197 interactions per bunch crossing with an average pile-up of $\langle \mu \rangle = 40$, which includes
1198 the effect on the detector response due to interactions from bunch crossings before
1199 or after the one containing the hard interaction.

1200 The $t\bar{t}$ events are generated using the POWHEGBOX [100–103] v2 generator at
1201 next-to-leading order with the NNPDF3.0NLO [104] set of parton distribution
1202 functions (PDFs). The h_{damp} parameter² is set to 1.5 times the mass of the top-quark
1203 (m_{top}) [105], with $m_{\text{top}} = 172.5$ GeV. The events are interfaced to PYTHIA 8.230 [106]
1204 to model the parton shower, hadronisation, and underlying event, with parameters
1205 set according to the A14 tune [107] and using the NNPDF2.3LO set of PDFs [108].
1206 Z' events are generated with PYTHIA 8.2.12 with the same tune and PDF set. The
1207 decays of b - and c -hadrons are performed by EVTGEN v1.6.0 [109]. Particles are
1208 passed through the ATLAS detector simulation [110] based on GEANT4 [111].

1209 Jets are required to have a pseudorapidity $|\eta| < 2.5$ and $p_{\text{T}} > 20$ GeV. Additionally, a
1210 standard selection using the Jet Vertex Tagger (JVT) algorithm at the tight working
1211 point is applied to jets with $p_{\text{T}} < 60$ GeV and $|\eta| < 2.4$ in order to suppress pile-up
1212 contamination [112].

1213 5.4.2 Model Inputs

1214 The fake track MVA is given two jet variables and 20 tracking related variables
1215 for each track fed into the network. The jet transverse momentum and signed
1216 pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in
1217 Table 5.2. The track parameters and hit pattern are key indicators of whether or
1218 not a track is fake. The FracRank variable is the ordered index of the tracks that
1219 pass the ambiguity solver’s selection divided by the total number of successfully
1220 reconstructed tracks in the event. The ambiguity solver processes track candidates
1221 iteratively in order of an internal score (see Section 3.4.1), and the order in which
1222 tracks are accepted is preserved. Since tracks with shared hits have lower scores,
1223 tracks which do not require the removal of shared hits are likely to be processed
1224 and accepted earlier on, whereas tracks with shared hits will be processed later and

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

1225 potentially have their shared hits removed. Hence the FracRank variable gives an
1226 indication of the of how easy it was for the track to be reconstructed.

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	Description
p_T	Track transverse momentum
ΔR	Angular distance between the track and jet
d_0	Closest distance from the track to the PV in the longitudinal plane
z_0	Closest distance from the track to the PV in the transverse plane
nIBLHits	Number of IBL hits
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nTRTHits	Number of TRT hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
r_{first}	Radius of first hit
nDOF	Number of degrees of freedom on the track
FracRank	Ambiguity solver ordering variable

Table 5.2: Input features to the fake track classification NN. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [57], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.

1227 Track selection follows the loose selection described in Ref. [82] and outlined in
1228 Table 5.3, which was found to improve the performance compared to previous tighter
1229 selections, whilst ensuring good resolution of tracks and a low fake rate [57]. Inputs
1230 are scaled to have a central value of zero and a variance of unity before training and
1231 evaluation.

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

Table 5.3: Quality selections applied to tracks, where d_0 is the transverse IP of the track, z_0 is the longitudinal IP with respect to the PV and θ is the track polar angle (see Section 3.2.2 for the IP definitions). Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [57]. 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 $1e-3$. Optimisation of the networks architecture was carried out to ensure optimal performance with a relatively small number of learnable parameters – 54 thousand. The model was trained using 40 million tracks with a 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.

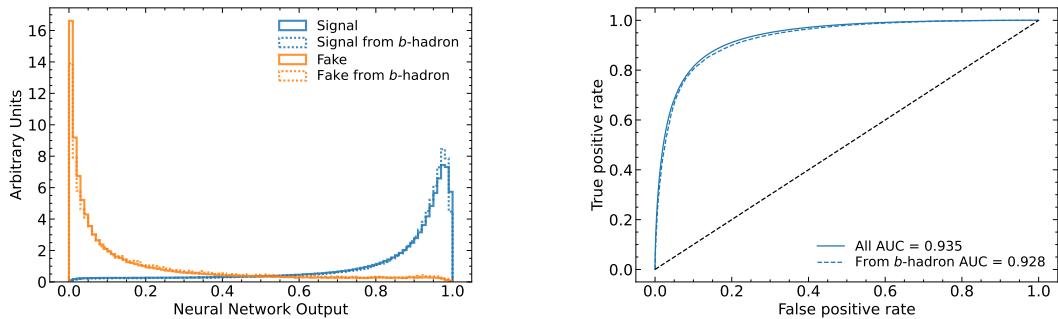


Figure 5.6: (left) Normalised histogram of the model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line). The model is able 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)

1258 45.6% of fake tracks. Table 5.5 also shows that a significant amount of tracks which
1259 are labelled as both fake and from the decay of a b -hadron are also removed. This can
1260 happen because fake tracks with $\text{TMP} < 0.75$ are still matched to a truth particle,
1261 which can be the decay product of a b -hadron.

MVA Output Cut	Signal Track Efficiency		Fake Track Efficiency	
	All	From b	All	From b
0.06	98.8%	98.9%	45.6%	39.8%
0.12	97.3%	97.5%	59.4%	53.6%

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

1262 5.5 b -hadron Track Identification

1263 After initial tests and investigation, it was found that fake tracks which were the
1264 result of b -hadron decays actually aided b -tagging performance. The application of
1265 a single tool which removed all fake tracks was therefore not optimal. A second
1266 tool was therefore trained in the same manner as the first, this one was designed to
1267 distinguish between those tracks which were from the decay of a b -hadron (FromB
1268 and FromBC in Table 5.1) and those which were not (all other truth origins). The
1269 b -hadron decay track MVA was trained using the same setup as described above,
1270 with the same tracks, input variables, and training procedure. The performance of
1271 the model to separate b -hadron decay tracks from other tracks is shown in Fig. 5.7.
1272 Using a selection WP of 0.1, the model can retain 98.5% of b -hadron tracks and
1273 reject 46.2% of tracks not from the decay of a b -hadron. In Section 5.6, this model is
1274 used in conjunction with the fake track identification MVA described in Section 5.4.4
1275 to identify and remove fake tracks which are not from the decay of a b -hadron.

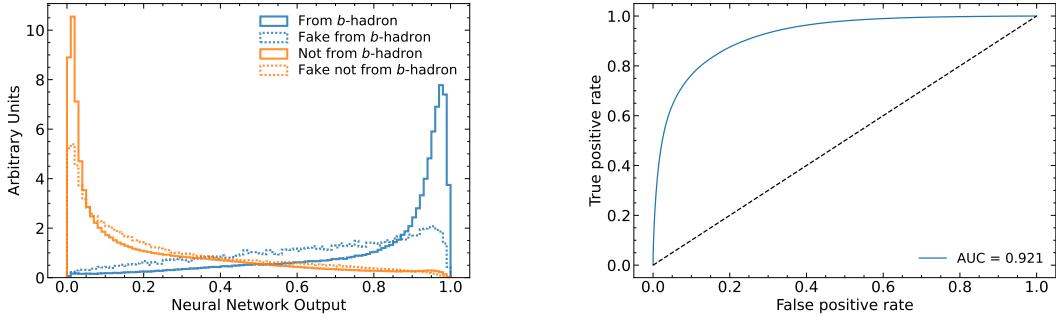


Figure 5.7: (left) Normalised histogram of the model output separated for tracks from the decay of a b -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line).

1276 5.6 Combined Approach

1277 A 2-dimensional cut was then used to only reject those tracks that had a high
 1278 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).

1279 The light-jet efficiency of SV1 is successfully reduced when using the combined tools
 1280 to remove fake tracks that are not from a b -hadron decay, as shown in Fig. 5.8. At a
 1281 b -jet efficiency of 70%, the light-jet mistag rate for jets with $250 < p_T < 400 \text{ GeV}$
 1282 is reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For
 1283 jets with $400 < p_T < 1000 \text{ GeV}$ the mistage rate drops from 0.1 to 0.08 for a similar
 1284 relative improvement of 20%. The performance of the fake track removal approach
 1285 was also tested for the other low level vertexing algorithm: JetFitter. A similar level
 1286 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.

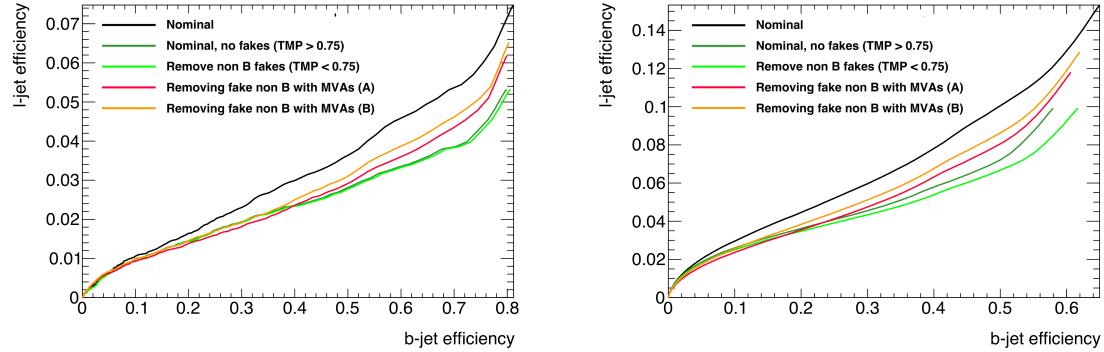


Figure 5.8: The effect of applying the fake track identification algorithm together with the b -hadron decay track identification on the jet tagging performance of SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). 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.

1303 Future Work

1304 While removing tracks prior to their input to the low level tagging algorithms is
1305 shown here to be beneficial, a more performant alternative might be to keep these
1306 tracks but label them as being fake (for example using the output of the classification
1307 tool), and allow the tagging algorithms to take this into consideration, potentially
1308 making use of this information. This is not straightforward with manually optimised
1309 taggers such as SV1 and JetFitter, but is possible with more advanced taggers as
1310 described in Chapter 6.

1311 Tools which identify the origin of a given track have other potential uses. One
1312 application is to isolate a relatively pure sample of fake tracks which can be used
1313 to estimate the fake track rate in data, which would be useful for estimating the
1314 uncertainty on fake track modelling. Another application would be to use the
1315 *b*-hadron track identification tool to improve the track-to-jet association. Both
1316 applications are currently under investigation.

1317 The approach here works on a track-by-track basis, but a more sophisticated approach
1318 would consider the correlations between the tracks inside a jet, as shown in Chapter 6.

1319 Also left for future work is to simultaneously train a single tool which discriminates
1320 between all the truth origins listed in Table 5.1. Such a tool would be useful as a
1321 general purpose multiclass classifier.

₁₃₂₂ **Chapter 6**

₁₃₂₃ **Graph Neural Network Flavour
Tagger**

₁₃₂₅ As discussed in Chapter 4, flavour tagging is the identification of jets instantiated
₁₃₂₆ from b - and c -hadrons. Flavour tagging is a critical component of the physics
₁₃₂₇ programme of the ATLAS experiment. It is of crucial importance for the study of the
₁₃₂₈ Standard Model (SM) Higgs boson and the top quark, which decay preferentially to
₁₃₂₉ b -quarks [113, 114], and additionally for several Beyond the Standard Model (BSM)
₁₃₃₀ resonances that readily decay to heavy flavour quarks [115].

₁₃₃₁ This chapter introduces GN1, a novel ML-based flavour tagging algorithm based
₁₃₃₂ on graph neural networks (GNNs). In Section 6.1, an overview of the approach
₁₃₃₃ used for GN1 is provided. An introduction to the theory of GNNs is provided in
₁₃₃₄ Section 5.2. Details of the experimental setup are provided in Section 6.2, while
₁₃₃₅ the architecture of GN1 is specified in Section 6.3.3. In Section 6.3.4, the training
₁₃₃₆ procedure is described, and in Section 6.4 the results are shown.

₁₃₃₇ **6.1 Motivation**

₁₃₃₈ GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. As opposed
₁₃₃₉ to the existing approach to flavour tagging described in Chapter 4, which relies
₁₃₄₀ on a two tiered approach requiring the use of both low- and high-level algorithms,
₁₃₄₁ GN1 takes as inputs information directly from an unordered variable number of
₁₃₄₂ 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 [2, 116].

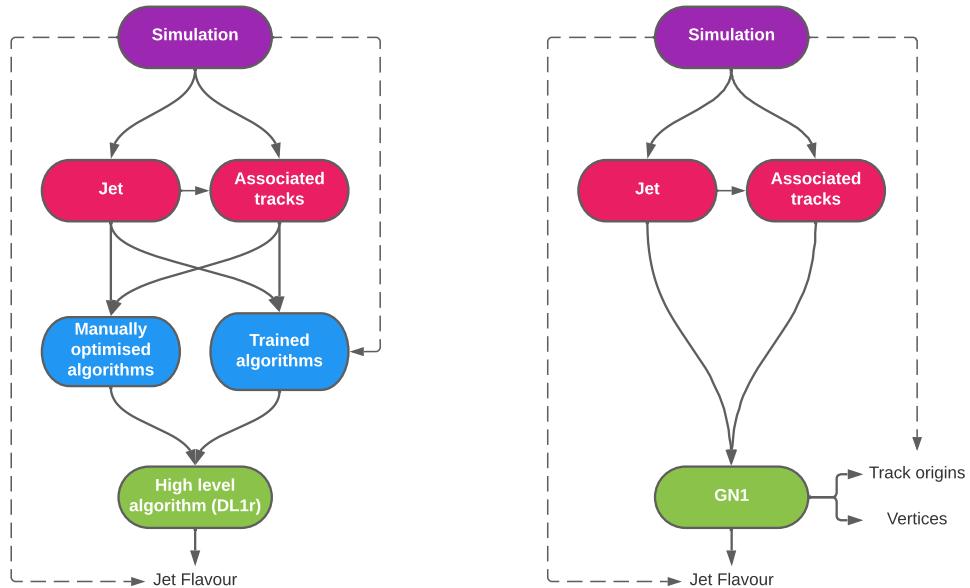


Figure 6.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right) [3]. The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, 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

1357 DL1r. This is particularly important when optimising the tagger for new regions of
1358 phase space (e.g. c -tagging or high- p_T b -tagging), or when the detector is upgraded
1359 or the charged particle reconstruction or selection algorithms are re-optimised.

1360 GN1 is trained to learn about the internal structure of the jet through the use of two
1361 auxiliary training objectives: the prediction of the underlying physics process from
1362 which each track originated, and the grouping of tracks originating from a common
1363 spatial position (i.e. a common vertex). These auxiliary objectives are meant to
1364 guide the neural network towards a more complete understanding of the underlying
1365 physics inside the jet, thereby removing the need for the low-level algorithms, which
1366 previously contained information about the underlying physics in their design. The
1367 training targets for the primary and auxiliary objectives are extracted from truth
1368 information, i.e. information that is only available in simulation, as opposed to
1369 reconstructed quantities available in both collision data and simulation.

1370 In this chapter, the following advantages of the GN1 approach will be demonstrated:

- 1371 1. GN1 boasts improved performance with respect to the current ATLAS flavour
1372 tagging algorithms, with significantly larger background rejection rates for a
1373 given signal efficiency. Alternatively the rejection rates can be kept fixed for a
1374 substantial increase in signal efficiency, in particular at high- p_T .
- 1375 2. The same network architecture can be easily optimised for a wider variety of
1376 use cases (e.g. c -jet tagging and high- p_T jet tagging) since there are no low-level
1377 algorithms to retune.
- 1378 3. There are fewer algorithms to maintain.
- 1379 4. Alongside the network's prediction of the jet flavour, the auxiliary vertex and
1380 track origin predictions provide more information on why a jet was (mis)tagged
1381 or not. This information can also have uses in other applications, for instance
1382 to explicitly reconstruct displaced decay vertices or to remove fake tracks.¹

¹A fake track is defined as a track with a truth-matching probability less than 0.5, where the truth-matching probability is defined in Ref. [57].

1383 6.2 Experimental Setup

1384 6.2.1 Datasets

1385 The datasets used to train the GN1 tagger are the same as described in Section 5.4.1.
1386 The training dataset contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40%
1387 of which are Z' jets. In order to evaluate the performance of the model during, a
1388 statistically independent set of 500k testing jets from both the $t\bar{t}$ and Z' samples are
1389 used. For the final testing of the model and the creation of the performance plots,
1390 a further 1 million independent testing jets from each of the $t\bar{t}$ and Z' samples are
1391 used. Before being fed into the model, the track- and jet-level inputs are normalised
1392 to have a mean of zero and a variance of unity. The jet flavour labels are assigned
1393 as described in Section 3.4.3. Truth labelled b -, c - and light-jets are kinematically
1394 re-sampled in p_T and η to ensure identical distributions in these variables.

1395 6.3 Model Architecture

1396 6.3.1 Model Inputs

1397 As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40
1398 tracks which have been associated to the jet. Each track consists of 21 variables. The
1399 kinematic jet variables are the jet transverse momentum and signed pseudorapidity.
1400 The input variables which are provided for each track are listed in Table 6.1. For
1401 each track, variables containing the track parameters and uncertainties, and detailed
1402 information on the hit content are provided as inputs to the model.

1403 In cores of high- p_T jets, track density is high due to the increased multiplicity and
1404 collimation of tracks (see Chapter 4). As a result, the separation between tracks
1405 can be of the same order as the active sensor dimensions, resulting in an increase
1406 in merged clusters and tracks which share hits [57]. Due to the relatively long
1407 lifetimes of b -hadrons and c -hadrons, which can traverse several layers of the ID
1408 before decaying and have highly collimated decay products, the presence of shared
1409 or missing hits is a critical signature of heavy flavour jets.

1410 Dependence of the model on the absolute value of the azimuthal jet angle ϕ is
1411 explicitly removed by providing only the azimuthal angle of tracks relative to the jet
1412 axis. The track pseudorapidity is also provided relative to the jet axis.

1413 Since heavy flavour hadrons can decay semileptonically approximately 20% of the time,
1414 the presence of a reconstructed lepton in the jet carries discriminating information
1415 about the jet flavour. To exploit this, a variant of GN1 called GN1Lep is trained in
1416 addition to the baseline model. The GN1Lep variant is identical to the baseline model,
1417 except for the inclusion an additional track-level input, leptonID, which indicates
1418 if the track was used in the reconstruction of an electron, a muon or neither. The
1419 variable is signed by the charge of the reconstructed lepton. The leptons used in the
1420 definition of the leptonID variable are required to satisfy basic quality requirements.
1421 The muons are required to be combined [117], and the electrons are required to pass
1422 the *VeryLoose* likelihood-based identification working point [118].

1423 The selections applied to the tracks is the same as that used for the fake track
1424 classification MVA described in Chapter 5. The full set of track selections is listed
1425 in Table 5.3. This selection was found to improve the flavour tagging performance
1426 compared to previous tighter selections, whilst ensuring good resolution of tracks
1427 and a low fake rate [57]. However, Section 6.4.8 demonstrates that further relaxation
1428 of the track selection requirements may be warranted.

1429 If more than 40 tracks are associated to a given jet, only the first 40 tracks with the
1430 largest transverse IP significance² $s(d_0)$ are fed into the model as inputs.

1431 6.3.2 Auxiliary Training Objectives

1432 In addition to the jet flavour classification, two auxiliary training objectives are
1433 defined. The first auxiliary objective is the prediction of the physical process that
1434 gave rise to each track within the jet (i.e. the track origin), while the second is the
1435 prediction of track-pair vertex compatibility. Each auxiliary training objective comes
1436 with a training target which, similar to the jet flavour label, is a truth labels derived

²Impact parameter significances are defined as the IP divided by its corresponding uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. Track IP significances are lifetime signed according to the track's direction with respect to the jet axis and the primary vertex [39].

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

Table 6.1: Input features to the GN1 model [3]. 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 [57], 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.

1437 from the simulation. The presence of the auxiliary training objectives improves the
1438 jet classification performance as demonstrated in Section 6.4.3.

1439 For the track origin prediction objective, each track is labelled with one of the
1440 exclusive categories defined in Table 5.1 of Section 5.3 after analysing the particle
1441 interaction (or lack thereof) which led to its formation. Since the presence of different
1442 track origins is strongly related to the flavour of the jet, training GN1 to recognise
1443 the origin of the tracks provides an additional handle on the classification of the
1444 jet flavour. This task may also aid the jet flavour prediction by acting as a form of
1445 supervised attention [119] - in detecting tracks from heavy flavour decays the model
1446 may learn to pay more attention to these tracks.

1447 The vertexing auxiliary objective makes use of the fact that displaced decays of b -
1448 and c -hadrons lead to secondary and tertiary vertices inside the jet, as described in
1449 Section 4.1.1. The presence of displaced secondary vertices is not a completely clean
1450 signal of a heavy flavour jet, as displaced secondary vertices can also occur in light-jets
1451 as a result of material interactions, conversions, and long-lived particle decays (e.g.
1452 K_S^0 and Λ^0). For the auxiliary object, GN1 predicts a binary label for each pair of
1453 tracks in the jet. The label has a value of 1 if the truth particles associated with the
1454 two tracks in the pair originated from the same spatial point, and 0 otherwise. To
1455 derive the corresponding truth labels for training, truth production vertices within 0.1
1456 mm are merged. Track-pairs where one or both of the tracks in the pair have an origin
1457 label of either Pile-up or Fake are given a label of 0. Using the pairwise predictions
1458 from the model, groups of tracks that have common compatibility can be formed,
1459 resulting in the finding of vertices. Two existing low-level tagging algorithms, SV1
1460 and JetFitter (introduced in Section 3.4.2), are currently used to find and reconstruct
1461 vertices inside jets and are used as inputs to the existing jet flavour tagger DL1r.
1462 The addition of this auxiliary training objective removes the need for inputs from a
1463 dedicated secondary vertexing algorithm.

1464 Both of the auxiliary training objectives described here can be considered as “stepping
1465 stones” on the way to classifying the flavour of the jet. By requiring the model to
1466 predict the truth origin of each track and the vertex compatibility of each track-pair,
1467 the model is guided to learn representations of the jet which are connected to the
1468 underlying physics and therefore relevant for classifying the jet flavour.

1469 6.3.3 Architecture

1470 As discussed in the previous sections, GN1 is a graph neural network which makes
1471 use of auxiliary training objectives in order to determine the jet flavour. A coarse
1472 optimisation of the network architecture hyperparameters (for example number of
1473 layers and number of neurons per layer) has been carried out in order to maximise
1474 the flavour tagging performance, but it is likely that further dedicated optimisation
1475 studies could lead to further performance improvements.

1476 The model architecture builds on a previous implementation of a GNN-based jet
1477 tagger [116]. The previous approach was comprised of two separate graph neural
1478 networks with the auxiliary tasks being performed at an intermediate stage after the
1479 first and before the second. This two stage approach was found to be unnecessary and
1480 as such GN1 simplifies the architecture into a single graph neural network with the
1481 auxiliary tasks being performed at the end, alongside the primary jet classification
1482 task. GN1 makes use of a more sophisticated graph neural network layer [120],
1483 which is described in more detail below. The changes significantly improved tagging
1484 performance and also led to a significant reduction in training time.

1485 As inputs, the model takes information about the jet and a number of associated
1486 tracks, as detailed in Section 6.3.1. The jet variables are concatenated with the
1487 variables for each track as shown in Fig. 6.2. The combined jet-track input vectors
1488 are then fed into a per-track initialisation network with three hidden layers, each
1489 containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.3. The
1490 track initialisation network is similar to a deep sets model [99], but does not include
1491 a reduction operation (mean or summation) over the output track representations.
1492 The initialisation network allows for initial per-track input processing without the
1493 associated parameter count cost of the graph convolutional layers described below.

1494 The outputs of the track initialisation network are used to populate the nodes of a
1495 fully connected graph, such that each node in the graph neighbours every other node.
1496 Each node h_i in the graph corresponds to a single track in the jet, and is characterised
1497 by a feature vector, also called a representation. The per-track output representations
1498 from the initialisation networks are used as the initial feature vectors of each node
1499 in the graph. In each layer of the graph network, output node representations h'_i
1500 are computed by aggregating the features of h_i and neighbouring nodes \mathcal{N}_i using
1501 a multi-head attention mechanism ($n = 2$) as described in Ref. [120, 121]. First,

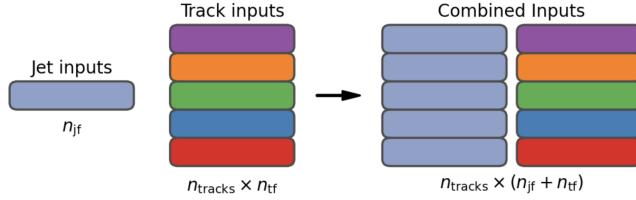


Figure 6.2: The inputs to GN1 are the two jet features ($n_{jf} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{tf} = 21$) [3]. The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1.

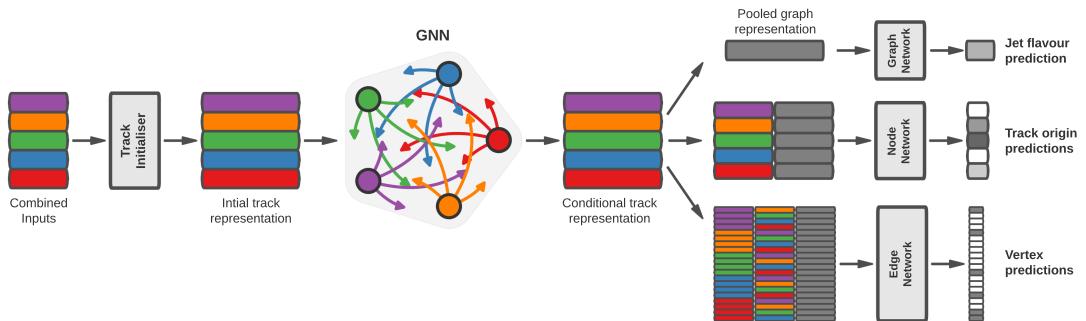


Figure 6.3: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting node representations are used to predict the jet flavour, the track origins, and the track-pair vertex compatibility.

1502 the feature vectors of receiver and sender nodes are fed into two fully connected
 1503 linear layers \mathbf{W}_r and \mathbf{W}_s , to produce an updated representation for each sender and
 1504 receiver node $\mathbf{W}_r h_i$ and $\mathbf{W}_s h_j$. These updated feature vectors are used to compute
 1505 edge scores $e(h_i, h_j)$ for each node pair, as in

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

1506 where, θ is a non-linear activation function, and \mathbf{a} is a learned vector. These edge
 1507 scores are then used to calculate attention weights a_{ij} for each pair of nodes using
 1508 the softmax function over the edge scores

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

1509 Finally, the updated representations for the receiver nodes h'_i are computed by taking
 1510 the weighted sum over each updated node representation $\mathbf{W}_r h_i$, with weights a_{ij}

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

1511 The set of operations described above constitute a single graph network layer. Three
 1512 such layers are stacked to construct the graph network, representing a balance
 1513 between achieving good performance in a reasonable time and avoiding overtraining
 1514 due to inflation of the parameter count of the model. The final output from the graph
 1515 neural network is a set of per-node (i.e. per-track) feature vectors that are conditional
 1516 representations of each track given the other tracks in the jet. In order to perform
 1517 the jet flavour prediction, a flattened global representation of the jet is needed. To
 1518 produce this, the output track representations are combined using a weighted sum,
 1519 where the weights are learned during training and therefore act as a form of attention
 1520 over the different tracks. The flattened outputs from the sum are then fed into a
 1521 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. [3]. 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.

1545 6.3.4 Training

1546 The full GN1 training procedure minimises the total loss function L_{total} , defined
 1547 in Eq. (6.4). This loss is composed of three terms: L_{jet} , the categorical cross
 1548 entropy loss over the different jet flavours; L_{vertex} , the binary track-pair compatibility
 1549 cross entropy loss; and L_{track} , the categorical cross entropy loss for the track origin
 1550 prediction. L_{vertex} is computed via a weighted average over all intra-jet track-pairs in
 1551 the batch, and L_{track} is computed by a weighted average over all tracks in the batch,
 1552 where the weights are described below.

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

1553 The different losses converge to different values during training, reflecting differences
 1554 in the relative difficulty of the various training objectives. The values of L_{vertex} and
 1555 L_{track} are weighted by $\alpha = 1.5$ and $\beta = 0.5$ respectively to ensure they converge to
 1556 similar values, giving them an equal weighting towards L_{total} . The values of α and β
 1557 are chosen to ensure that L_{jet} converges to a larger value than either L_{vertex} and L_{track} ,
 1558 which reflects the primary importance of the jet classification objective. It was found
 1559 that in practice the overall performance of the model was not sensitive to modest
 1560 changes in the loss weights α and β . Pre-training using L_{total} (i.e. on all tasks) and
 1561 fine tuning on only the jet classification task also did not improve performance versus
 1562 the described standard setup, indicating that the auxiliary tasks are not in direct
 1563 competition with the jet classification task. As there was a large variation in the
 1564 relative abundance of tracks of the different origins, the contribution of each origin to
 1565 L_{track} was weighted by the inverse of the frequency of their occurrence. In vertexing
 1566 loss L_{vertex} , the class weight for track-pairs where both tracks are from either a b - or
 1567 c -hadron was increased by a factor of two as compared with other track-pairs, to
 1568 encourage the network to focus on correctly classifying heavy flavour vertices.

1569 GN1 can be trained with either the node or edge networks (and their corresponding
 1570 auxiliary tasks), or both, removed, as discussed in Section 6.4.3. In such cases,
 1571 the corresponding losses L_{vertex} and L_{track} are also removed from the calculation
 1572 of the overall loss L_{total} . The performance of the resulting models provides a

1573 useful indication of the benefit of including the auxiliary tasks to the primary jet
1574 classification objective.

1575 GN1 was trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes
1576 approximately 25 mins to complete over the training sample of 30 million jets
1577 described in Section 6.2.1. The Adam optimiser [122] with an initial learning rate of
1578 $1e-3$, and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically
1579 the validation loss, calculated on 500k jets, became stable after around 60 epochs.
1580 The epoch that minimized the validation loss was used for evaluation. GN1 has
1581 been integrated into the ATLAS software [54] using ONNX [123]. The test sample
1582 jet flavour predictions scores are computed using the ATLAS software stack as a
1583 verification of this process.

1584 6.4 Results

1585 The GN1 tagger is evaluated both as a b -tagging and c -tagging algorithm in Sec-
1586 tion 6.4.1 and Section 6.4.2 respectively. Evaluation is performed separately on
1587 both jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV and jets in the Z' sample with
1588 $250 < p_T < 5000$ GeV. The performance of the model is compared to the DL1r
1589 tagger [79, 124], which has been retrained on 75 million jets from the same samples
1590 as GN1. The input RNNIP tagger [81] to DL1r has not been retrained. As discussed,
1591 each tagger predicts the probability that a jet belongs to the b -, c - and light-classes.
1592 To use the model for b -tagging, these probabilities are combined into a single score
1593 D_b , which is defined as

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

1594 where f_c is a free parameter that determines the relative weight of p_c to p_l in the
1595 score D_b , controlling the trade-off between c - and light-jet rejection performance.
1596 The choice of f_c is arbitrary, and is optimised based upon the desired light- vs c -jet
1597 rejection performance. This parameter is set to a value of $f_c = 0.018$ for the DL1r
1598 model, obtained through an optimisation procedure described in Ref. [79]. Based on
1599 a similar optimisation procedure, a value of $f_c = 0.05$ is used for the GN1 models.

1600 A fixed-cut working point (WP) defines the corresponding selection applied to the
 1601 tagging discriminant D_b in order to achieve a given efficiency on the inclusive $t\bar{t}$
 1602 sample.

1603 A comparison of the b -tagging discriminant D_b between DL1r and GN1 is shown in
 1604 Fig. 6.4. The shapes of the D_b distributions are generally similar for b -, c - and light-
 1605 jets between both models, however, GN1 shifts the b -jet distribution to higher values
 1606 of D_b in the regions with the greatest discrimination. The GN1 c -jet distribution is
 1607 also shifted to lower values of D_b when compared with DL1r, enhancing the separation
 1608 and indicating that GN1 is improving c -jet rejection when compared with DL1r.

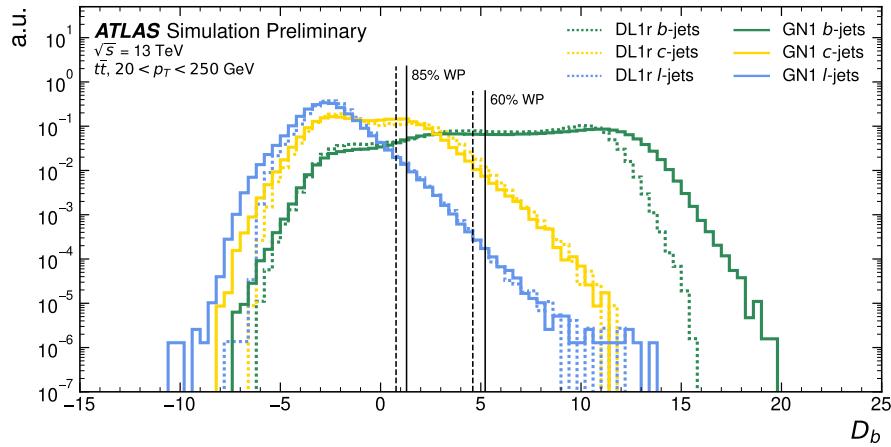


Figure 6.4: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample [3]. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of $f_c = 0.018$ 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.

1609 6.4.1 b -tagging Performance

1610 The performance of b -tagging algorithms is quantified by their ability to reject c - and
 1611 light-jets for a given b -jet selection efficiency WP. In order to compare the b -tagging
 1612 performance of the different taggers for the b -jet tagging efficiencies in the range
 1613 typically used by analyses, the corresponding c - and light-jet rejection rates are
 1614 displayed in Figs. 6.5 and 6.6 for jets in the $t\bar{t}$ and Z' samples respectively. Four
 1615 standard WPs are defined with b -jet tagging efficiencies of 60%, 70%, 77% and 85%

1616 respectively. These WPs are commonly used by physics analyses depending on their
1617 specific signal and background requirements. The WPs are defined based on jets
1618 in the $t\bar{t}$ sample only. Due to the much higher jet p_T range in the Z' sample, and
1619 the increased difficulty in tagging jets at high- p_T (see Chapter 4), the b -jet tagging
1620 efficiencies for jets in the Z' sample are lower than the corresponding WPs calculated
1621 in the $t\bar{t}$ sample. For instance the WP cut value computed to provide a 70% b -jet
1622 tagging efficiency on the $t\bar{t}$ sample results in a b -jet tagging efficiency of just \sim 30%
1623 on the Z' sample. In order to account for this, the range of b -jet tagging efficiencies
1624 displayed for plots showing the performance for jets in the Z' sample (for example
1625 Fig. 6.6) is chosen to span the lower efficiencies achieved in the Z' sample at high- p_T .

1626 For jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV, GN1 demonstrates considerably
1627 better c - and light-jet rejection when compared with DL1r across the full range of
1628 b -jet tagging efficiencies studied. The relative improvement is strongly dependent
1629 on the b -jet tagging efficiency under study. The largest improvements are found at
1630 lower b -jet tagging efficiencies. At a b -jet tagging efficiency of 70%, the c -jet rejection
1631 improves by a factor of \sim 2.1 while the light-jet rejection improves by a factor of \sim 1.8
1632 with respect to DL1r. For high- p_T jets in the Z' sample with $250 < p_T < 5000$ GeV,
1633 GN1 also brings a significant performance improvement with respect to DL1r across
1634 the range of b -jet tagging efficiencies studied. Again, the largest relative improvement
1635 in performance comes at the lower b -jet tagging efficiencies. At a b -jet efficiency of
1636 30%, GN1 improves the c -jet rejection with respect to DL1r by a factor of \sim 2.8 and
1637 the light-jet rejection by a factor of \sim 6. The performance comparison at lower b -jet
1638 tagging efficiencies is made more difficult due to the increased statistical uncertainties
1639 which result from the high rejection of background. Given that GN1 exploits the
1640 low-level detector information in a more complete and sophisticated way than DL1r,
1641 further studies are needed to confirm that the performance gain observed in these
1642 simulated samples is also observed in experimental data.

1643 The GN1Lep variant of GN1 demonstrates further improved performance with respect
1644 to the baseline model. This demonstrates the additional jet flavour discrimination
1645 power provided by the leptonID track input. For jets in the $t\bar{t}$ sample, the relative c -
1646 jet rejection improvement with respect to GN1 at the 70% b -jet WP is approximately
1647 25%. The improvement in light-jet rejection also increases by 40% at the same WP.
1648 For jets in the Z' sample, the relative c -jet rejection (light-jet rejection) performance

1649 with respect to GN1 improves by approximately 10% (25%) at a b -jet tagging
1650 efficiency of 30%.

1651 In general, the performance of all the taggers is strongly dependent on the jet p_T .
1652 This is due to the increased multiplicity and collimation of tracks, and the displaced
1653 decays that result from within the heavy flavour jets (see Chapter 4). Together,
1654 they contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a
1655 general degradation in quality of tracks inside the core of a jet, which in turn reduces
1656 the jet tagging performance.

1657 In order to study how the tagging performance changes as a function of the jet p_T ,
1658 the b -jet tagging efficiency as a function of p_T for a fixed light-jet rejection of 100 in
1659 each bin is shown in Fig. 6.7. For jets in the $t\bar{t}$ sample, at a fixed light-jet rejection
1660 of 100, GN1 improves the b -jet tagging efficiency by approximately 4% across all the
1661 jet p_T bins. Meanwhile, GN1Lep again demonstrates improved performance with
1662 respect to GN1, in particular at lower p_T . The relative increase in the b -jet tagging
1663 efficiency increases from 4% to 8% with respect to DL1r. For jets in the Z' sample,
1664 GN1 again outperforms DL1r across the entire jet p_T range studied. The largest
1665 relative improvement in performance is found at the highest transverse momenta
1666 of jet $p_T > 2 \text{ TeV}$, and corresponds to an approximate factor of 2 improvement in
1667 efficiency with respect to DL1r.

1668 The performance of the model was also evaluated as a function of the average
1669 number of pile-up interactions in the event. No significant dependence of the tagging
1670 performance was observed.

1671 6.4.2 c -tagging Performance

1672 As discussed previously, GN1 does not rely on any inputs from manually optimised
1673 low-level tagging algorithms. Since these algorithms were originally designed and
1674 tuned with the aim of b -tagging, and not c -tagging, the low level tagging algorithms
1675 may perform suboptimally for c -tagging purposes. The tagging of c -jets therefore
1676 presents a compelling use case for GN1. As each of the the models is trained with
1677 three output classes, using it as a c -tagging algorithm is trivially analogous to the
1678 approach used for b -tagging. The model output probabilities are combined into a
1679 single score D_c , which is defined similarly to Eq. (6.5) as

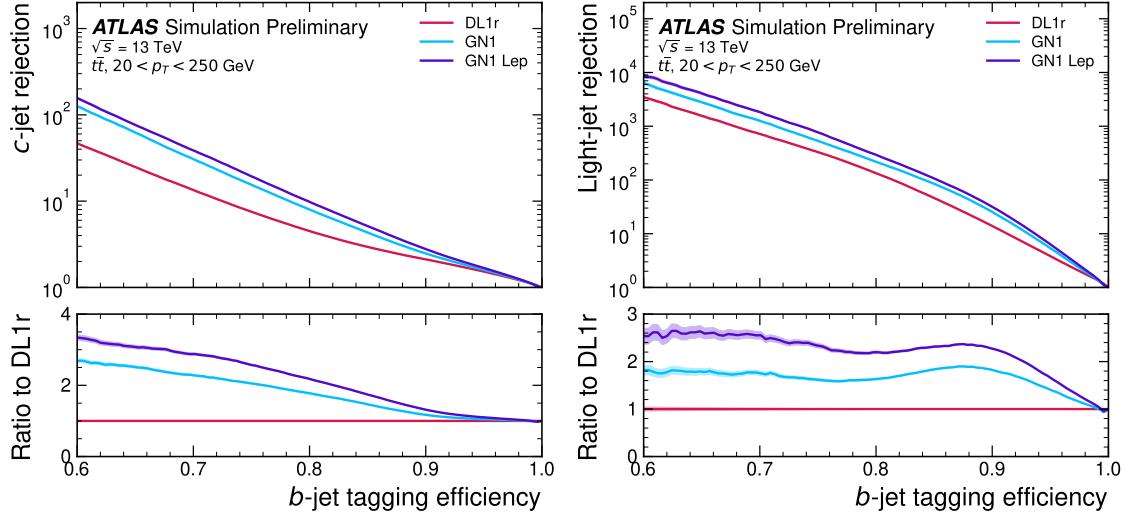


Figure 6.5: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ [3]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ 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.

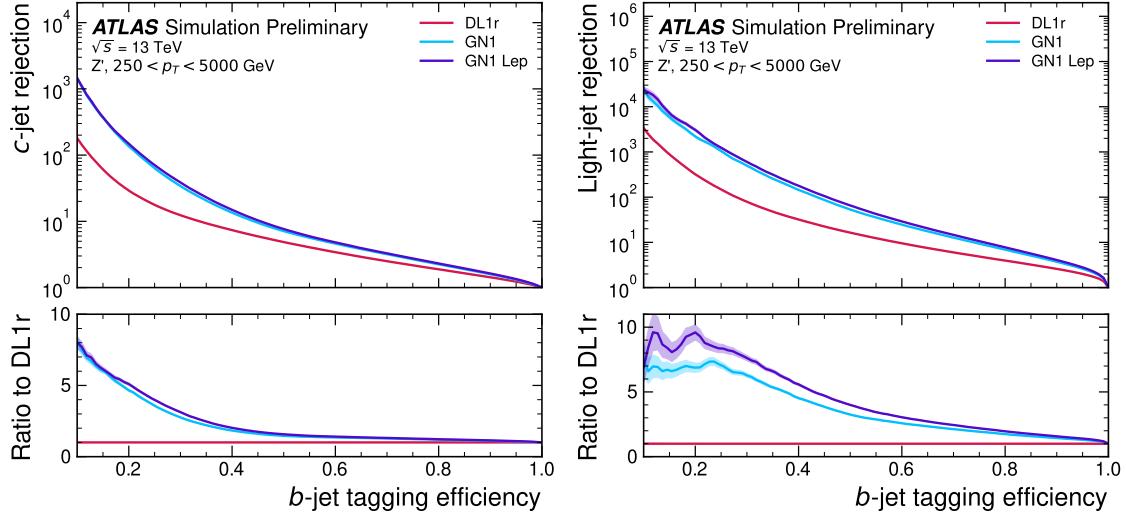


Figure 6.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$ [3]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used 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.

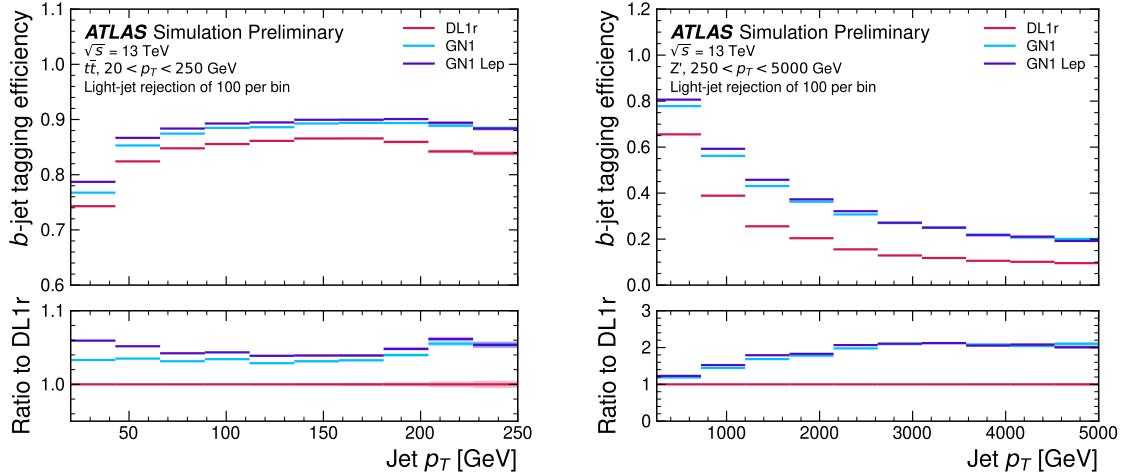


Figure 6.7: The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin [3]. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and 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

1694 case, with the both the b -jet rejection (light-jet rejection) improving with respect to
1695 GN1 by approximately 10% (20%) at the 25% c -jet WP. Fig. 6.9 shows the c -tagging
1696 performance on the jets in the Z' sample with $250 < p_T < 5000$ GeV. Both GN1 and
1697 GN1Lep perform similarly, improving the b -jet rejection by 60% and the light-jet
1698 rejection by a factor of 2 at the 25% c -jet WP.

1699 6.4.3 Ablations

1700 Ablation studies (the removal of certain components of a given model in order to
1701 study the impact of that component) are carried are carried out to determine the
1702 importance of the auxiliary training objectives of GN1 to the overall performance.
1703 The “GN1 No Aux” variant retains the primary jet classification objective, but
1704 removes both track classification and vertexing auxiliary objectives (see Section 6.3.2)
1705 and correspondingly only minimises the jet classification loss. The “GN1 TC” variant
1706 includes track classification objective but not the vertexing objective. Finally, the
1707 “GN1 Vert” includes the vertexing objective, but not the track classification objective.

1708 For jets in both the $t\bar{t}$ and Z' samples, a general trend is observed that the models
1709 trained without one or both of the auxiliary objectives results in significantly reduced
1710 c - and light-jet rejection when compared with the baseline GN1 model. This result
1711 is shown clearly in Figs. 6.10 and 6.11. For jets in the $t\bar{t}$ sample, the performance of
1712 GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert perform similarly to
1713 each other. For jets in the Z' sample meanwhile, the GN1 No Aux model already
1714 shows a clear improvement in c - and light-jet rejection when compared with DL1r at
1715 lower b -jet tagging efficiencies. Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1
1716 Vert perform similarly, and bring large gains in background rejection when compared
1717 with GN1 No Aux, but the combination of both auxiliary objectives yields the best
1718 performance.

1719 It is notable that the GN1 No Aux model matches or exceeds the performance of
1720 DL1r without the need for inputs from the low-level algorithms. This indicates that
1721 the performance improvements enabled by the improved neural network architecture
1722 used in GN1 appear to be able to compensate for the removal of the low-level
1723 algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform
1724 DL1r, demonstrating that both contribute to the overall high performance of the
1725 baseline model. The overall best performing model is the full version of GN1 trained

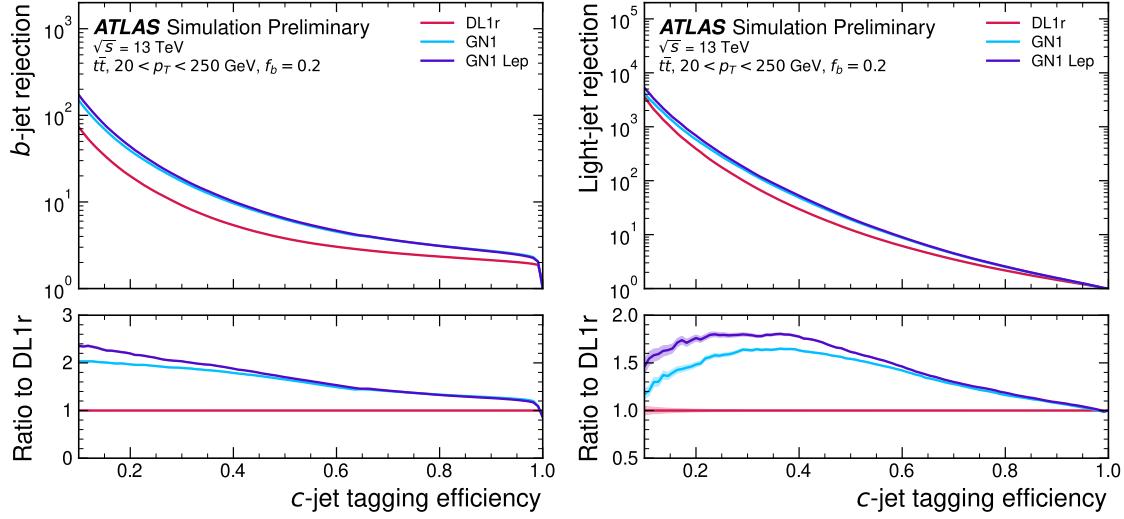


Figure 6.8: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$ [3]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

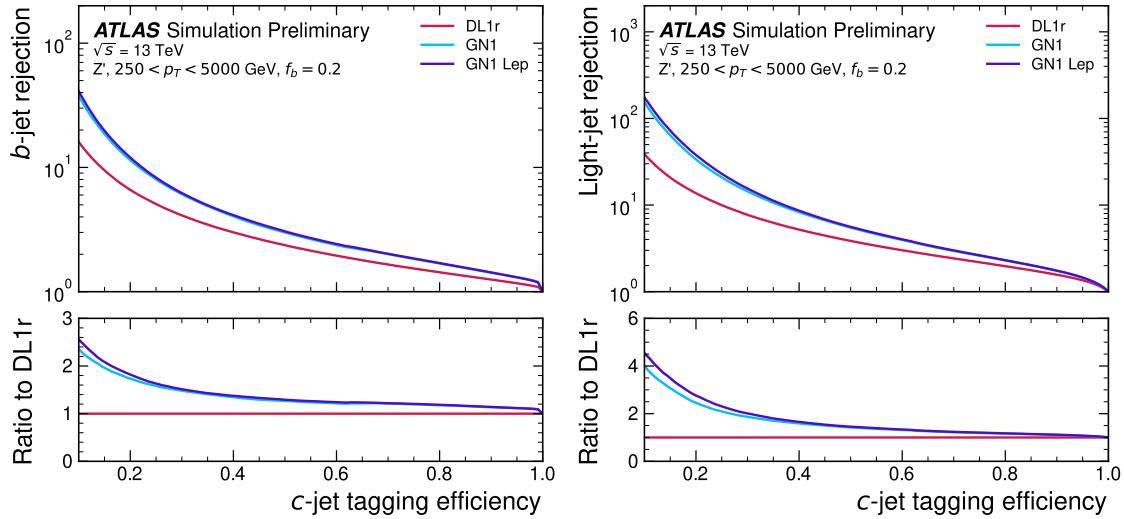


Figure 6.9: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000 \text{ GeV}$ [3]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

1726 with both auxiliary objective, demonstrating that the two auxiliary objectives are
1727 complementary.

1728 6.4.4 Inclusion of Low-Level Vertexing Algorithms

1729 As already mentioned, GN1 does not include any inputs from the low-level tagging
1730 algorithms, including the vertexing algorithms SV1 and JetFitter [62]. Since these
1731 algorithms are known to play a key role in contributing to the performance of DL1r, it
1732 was necessary to study whether their inclusion in GN1 resulted in further performance
1733 improvements. In a dedicated training of GN1 the SV1 and JetFitter tagger outputs
1734 were added to the GN1 jet classification network as an input, similar to how they
1735 are used in DL1r. These outputs include information on the reconstructed vertices,
1736 including the number of vertices, and the properties of the highest ranking recon-
1737 structed vertex (in the case of JetFitter). In addition, the index of the reconstructed
1738 SV1 or JetFitter vertices were included as two track-level inputs to GN1. These
1739 indices were also used to construct an input feature for the edge classification
1740 network used to identify vertices, which was given a value of one if the track-pair
1741 were from a common reconstructed SV1 or JetFitter vertex, and zero otherwise. The
1742 jet classification performance of this GN1 model was not significantly different to the
1743 baseline model, and in some cases the performance was slightly reduced. GN1 does
1744 not benefit from the inclusion of information from SV1 and JetFitter, indicating
1745 that the model is able to reconstruct the relevant information provided by these
1746 low-level algorithms. The study also demonstrates that the model can function as a
1747 highly performant standalone tagger that does not require (beyond retraining) any
1748 manual optimisation to achieve good performance in a wide range of phase spaces.
1749 A dedicated look at the vertexing performance of GN1 with some comparisons to
1750 SV1 and JetFitter is found in Section 6.4.6

determine ranking

1751 6.4.5 Jet Display Diagrams

1752 The auxiliary training objectives of GN1 allow for improved model interpretability,
1753 which is especially important for a monolithic approach as the low level taggers,
1754 which provide useful physical insight, are no longer present. Figs. 6.12 and 6.13
1755 provide example comparisons of the true origin and vertexing information compared

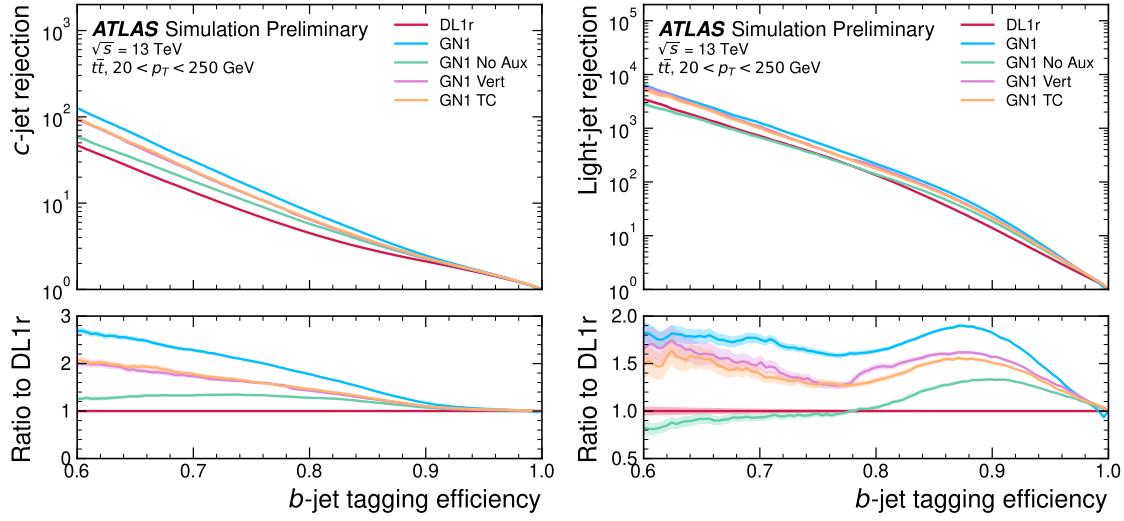


Figure 6.10: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed [3]. 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.

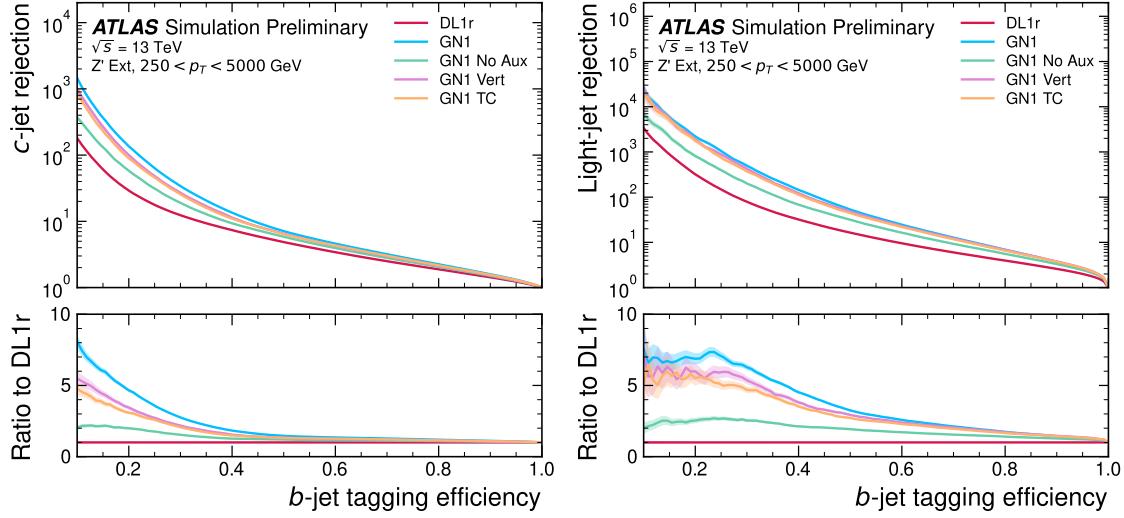


Figure 6.11: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed [3]. 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.

1756 with the predicted values from GN1, SV1 and JetFitter. Such comparisons can be
1757 used to provide an indication that GN1 reconstructs the correct representation of the
1758 jet structure, and may also help to identify limitations of the model. In the figures,
1759 the tracks in the jet are indexed twice on each of the x - and y -axes, and tracks are
1760 grouped into vertices along with other tracks as indicated by common markings in
1761 the relevant rows and columns.

1762 In Fig. 6.12, GN1 correctly groups the three primary tracks as having come from the
1763 primary vertex. The b -hadron and $b \rightarrow c$ -hadron decay vertices are also correctly
1764 predicted, and the origin of the tracks in each is correct. There is a single OtherSec-
1765 ondary track which GN1 incorrectly predicts as having come pile-up. Meanwhile SV1
1766 (by design) merges the two heavy flavour decay vertices, but incorrectly includes a
1767 track from the primary vertex. JetFitter reconstructs two vertices, one which is a
1768 combination of two tracks from different truth vertices and two other single track
1769 vertices in each of the heavy flavour vertices. GN1 also predicts the flavour of the jet
1770 with a high degree of certainty.

1771 Similarly, Fig. 6.12 shows that GN1 is able to relatively accurately predict the origin
1772 and vertex information of tracks inside a jet. The pile-up tracks and primary vertex
1773 tracks are correctly identified, and the heavy flavour decay tracks are also correctly
1774 identified with the exception of one of the b -hadron decay tracks. Again, SV1 merges
1775 the two heavy flavour decay vertices along with a track from pile-up, while JetFitter
1776 shows signs of being underconstrained by reconstructing two single track vertices,
1777 one with a pile-up track and one with a track from a $b \rightarrow c$ -hadron decay.

1778 6.4.6 Vertexing Performance

1779 From the track-pair vertex prediction described in Section 6.3.2, tracks can be
1780 partitioned into compatible groups representing vertices (see [116]). As such, GN1
1781 can perform vertex “finding”, but not vertex “fitting”, i.e. the reconstruction of
1782 a vertex’s properties, which currently still requires the use of a dedicated vertex
1783 fitter. In order to study the performance of the different vertexing tools, the truth
1784 vertex label of the tracks, discussed in Section 6.3.2, are used. To estimate the
1785 efficiency with which GN1 manages to find vertices inclusively, vertices containing
1786 tracks identified as coming from a b -hadron are merged together and compared to the
1787 inclusive truth decay vertices that result from a b -hadron decay (where if there are

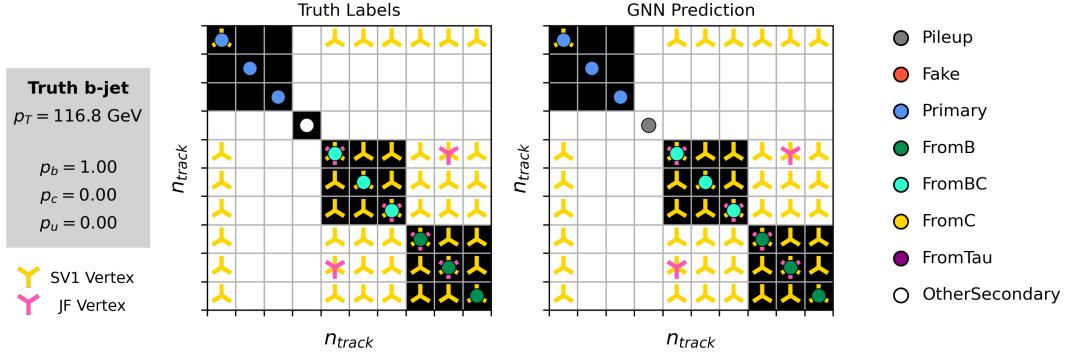


Figure 6.12: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed 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 pile-up.

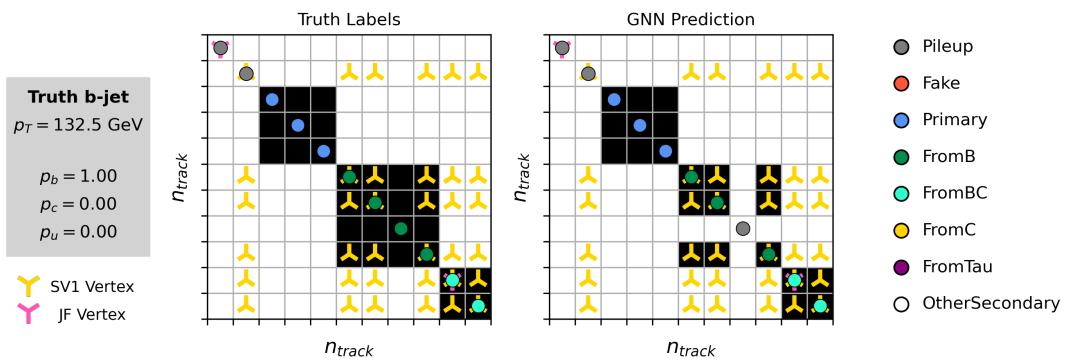


Figure 6.13: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed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

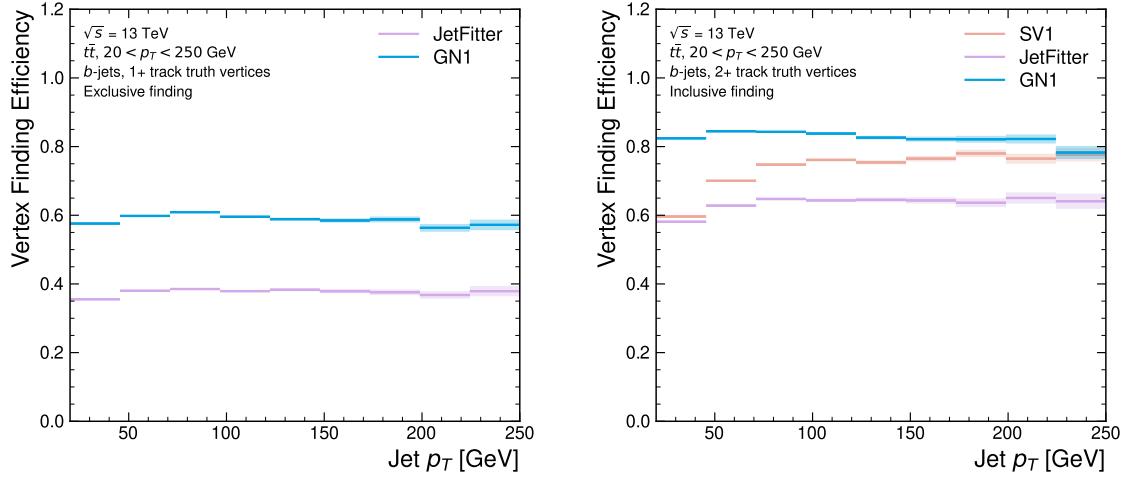


Figure 6.14: Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices 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.

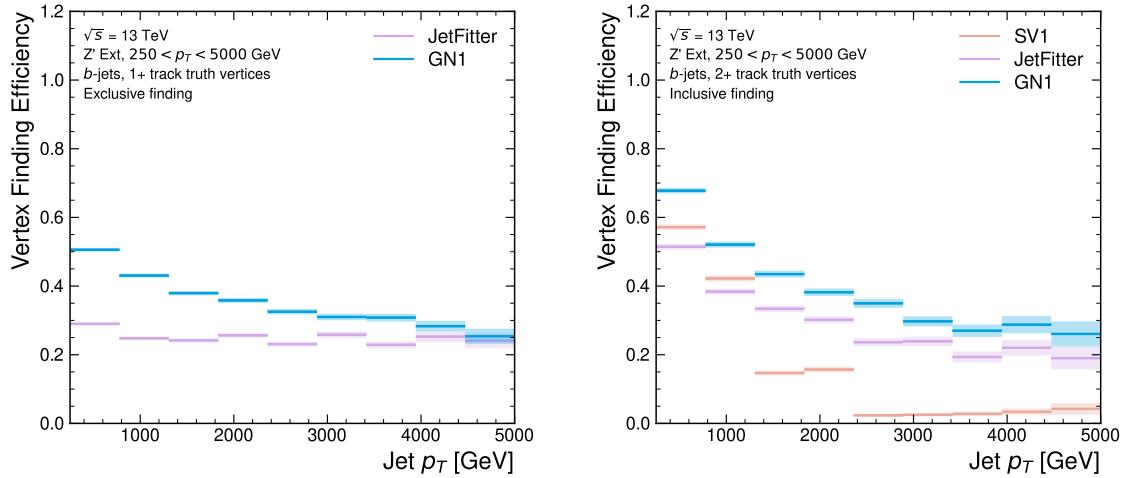


Figure 6.15: Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in the Z' sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain 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.

1853 While Figs. 6.14 and 6.15 indicate that GN1 is able to successfully find displaced heavy
 1854 flavour vertices in b -jets, it is also important to consider the vertexing performance
 1855 inside light-jets. light-jets may also contain displaced vertices due to long lived
 1856 secondary particles and material interactions. These tracks have a truth origin of
 1857 OtherSecondary in the truth labelling scheme enumerated in see Table 5.1. The
 1858 efficiency to reconstruct vertices comprised of OtherSecondary tracks can be computed
 1859 in an analogous way to the heavy flavour vertexing efficiency, which is described above.
 1860 Figs. 6.16 and 6.17 show the efficiency to reconstruct displaced OtherSecondary
 1861 vertices in light-jets as a function of p_T for jets in the $t\bar{t}$ sample and jets in the Z'
 1862 sample respectively. The figures demonstrate that GN1 is able to more effectively find
 1863 such vertices in light-jets as compared with SV1 and JetFitter. Since the properties
 1864 of the displaced vertices in light-jets are likely to be significantly different to heavy
 1865 flavour vertices found in heavy flavour jets, the improved reconstruction of such
 1866 vertices may help to differentiate between different flavour of jet.

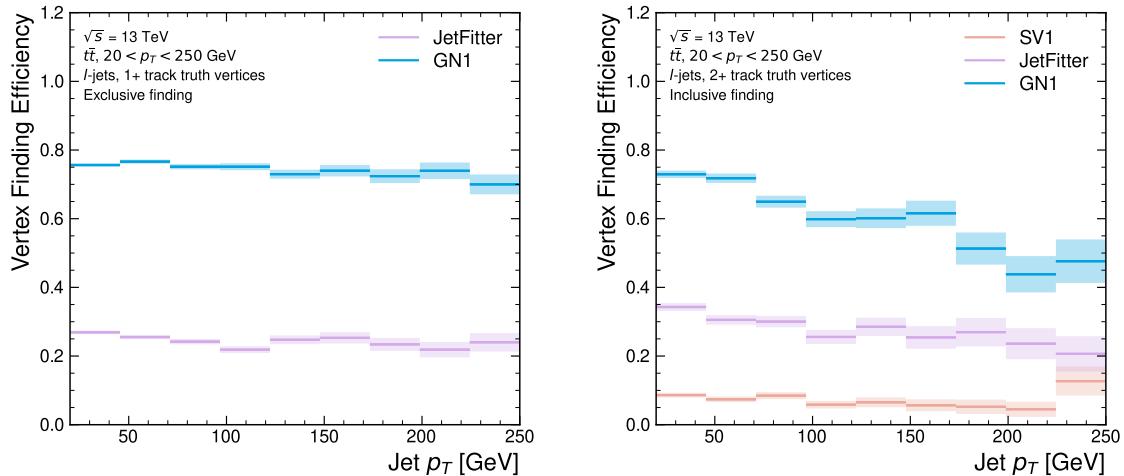


Figure 6.16: Vertex finding efficiency as a function of jet p_T for light-jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain 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.

1867 Collectively, the results in this section demonstrate that GN1 is able to accurately
 1868 group tracks by their spatial origin in both b -jets and light-jets. The purity of
 1869 the found vertices was also investigated and was found to be comparable or better
 1870 than that of SV1 and JetFitter. Using a vertex fitting algorithm to compare the

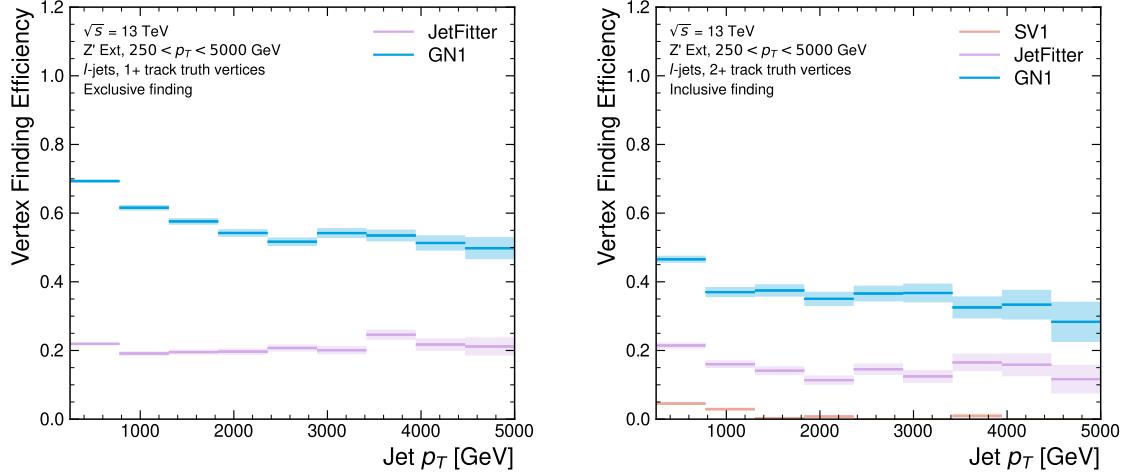


Figure 6.17: Vertex finding efficiency as a function of jet p_T for light-jets in the Z' sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least 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.

1886 To measure the track classification performance, the area under the curve (AUC)
 1887 of the receiver operating characteristic (ROC) curve is computed for each origin
 1888 class, using a one-versus-all classification approach. The AUCs for the different
 1889 truth origins are averaged using both an unweighted and a weighted mean. The
 1890 unweighted mean treats the performance of each class equally, while the weighted
 1891 mean uses as a weight the relative abundance of tracks of each class. Table 6.3
 1892 demonstrates clearly that GN1 outperforms the MLP both at $20 < p_T < 250 \text{ GeV}$
 1893 for jets in the $t\bar{t}$ sample and at $250 < p_T < 5000 \text{ GeV}$ for jets in the Z' sample. For
 1894 example, GN1 can reject 65% of fake tracks in jets in the $t\bar{t}$ sample, while retaining
 1895 more than 99% of good tracks (i.e. those tracks which are not fake). The GN1 model
 1896 has two advantages over the MLP which can explain the performance improvement.
 1897 Firstly, the graph neural network architecture enables the sharing of information
 1898 between tracks as discussed in Section 6.3.3. This is likely to be beneficial since the
 1899 origins of different tracks within a jet are correlated. Secondly, the jet classification
 1900 and vertexing objectives may be complementary to the track classification objective,
 1901 and so the track classification performance is improved by the combined training of
 1902 complementary objectives.

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

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

1903 Fig. 6.18 shows the track origin classification ROC curves for the different track
 1904 origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve visual readability
 1905 of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and
 1906 FromC) have been combined (weighted by their relative abundance), as have the
 1907 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 pile-up 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.

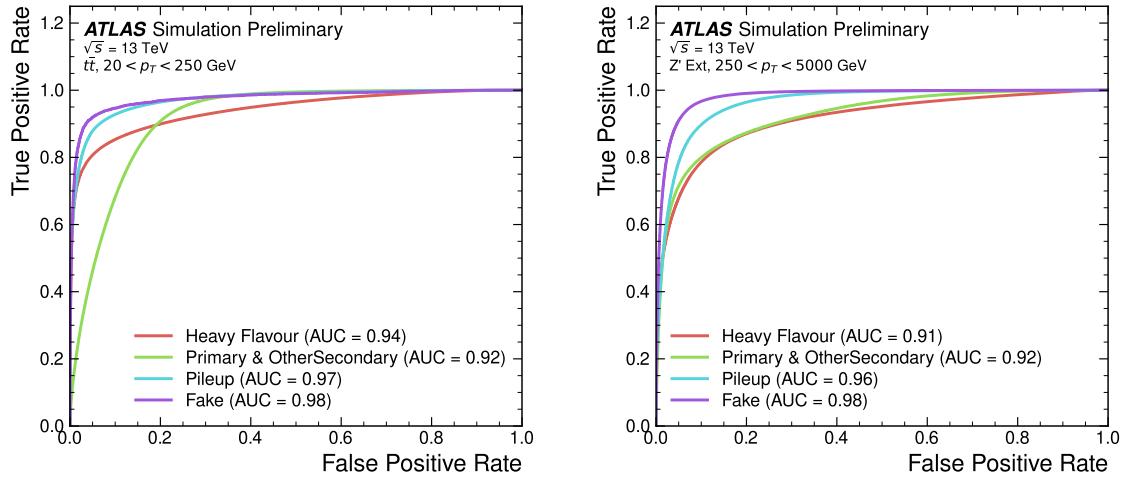


Figure 6.18: ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) [3]. The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and 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

1920 other tracks, especially at high- p_T , this cut rejects a significant proportion of these
1921 tracks.

1922 Figs. 6.19 and 6.20 show the result of training the GN1 tagger with the full relaxation
1923 of this cut, i.e. allowing tracks with any number of shared hits. The shared hit
1924 requirements applied by the ambiguity solver as part of track reconstruction (see
1925 Section 3.4.1) are still applied. In addition, the maximum allowed value of d_0 is
1926 increased from 3.5 mm to 5.0 mm. The results show that optimisation of the input
1927 track selection can lead to significant improvements in performance over the default
1928 selection. For the jets in the $t\bar{t}$ sample shown in Fig. 6.19, the effect of loosening
1929 the track selection is limited. This is expected due to the lower prevalence of shared
1930 hits at highly displaced tracks at lower transverse momenta. However for jets in the
1931 Z' sample as shown in Fig. 6.20, the light-jet rejection improves with respect to the
1932 baseline GN1 model by 30%, while the light-jet rejection improves by 70% at the
1933 50% b -jet WP.

1934 Although the results demonstrate a significant performance improvement at high- p_T ,
1935 it is also possible that additional studies on further loosening the selection could yield
1936 further improved results. For example the selections on the number of number of
1937 holes and the longitudinal impact parameter could be further relaxed. The maximum
1938 number of tracks provided as input to the model could also be increased from the
1939 default value of 40. In order to change the default tracking setup, studies investigating
1940 the modelling uncertainties of the additional tracks need to be carried out.

1941 6.5 Other Implementations of GN1

1942 The implementation of GN1 described in this chapter has been re-used in several
1943 other contexts, demonstrating its flexibility to easily provide good jet flavour tagging
1944 performance with minimal overhead. The model has been implemented as a b -jet
1945 tagger in the High Level Trigger (HLT) (see Section 3.3.4). The inputs to the
1946 model are the running on precision tracks and jet level quantities reconstructed after
1947 primary vertexing. Fig. 6.21 shows the performance of GN1 versus a comparable
1948 DL1d model [124], and two versions of DIPS [82], with EMTopo and PFlow jets (see
1949 Section 3.4.3) based on a low-precision region-of-interest based tracking pass. The

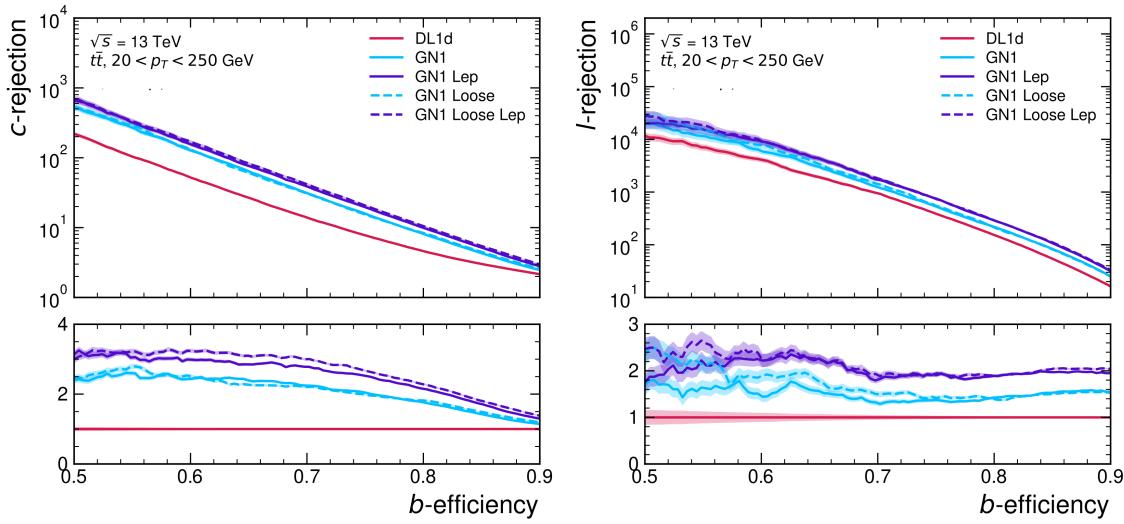


Figure 6.19: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. A value 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.

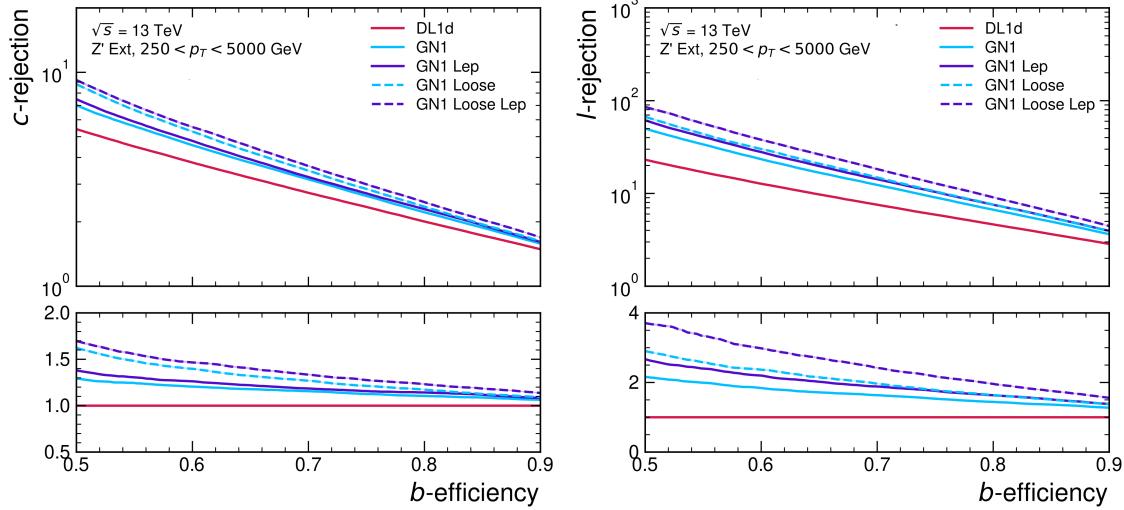


Figure 6.20: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. A value of $f_c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

1950 trigger implementation of GN1 improves upon the light-jet rejection of DL1d by 50%
1951 at the 60% b -jet WP for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV.

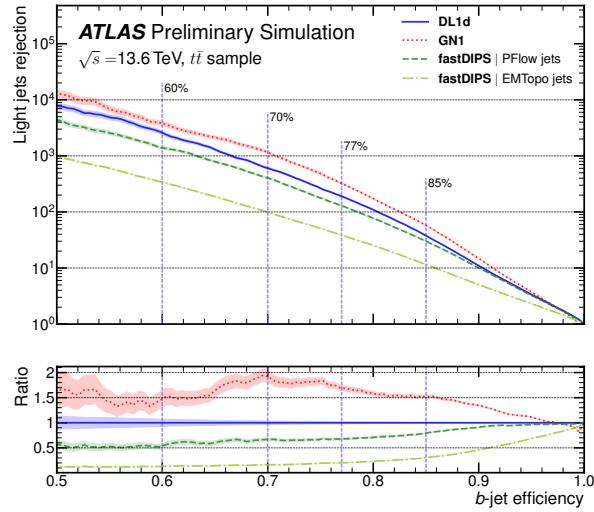


Figure 6.21: The light-jet rejection as a function of the b -jet efficiency jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 13.6$ TeV [125]. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. Binomial error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

1952 The model also demonstrates strong performance for the High Luminosity LHC
1953 (HL-LHC), as documented in Ref. [126]. Figs. 6.22 and 6.23 are reproduced from
1954 Ref. [126]. The results show that GN1 outperforms other existing flavour tagging
1955 algorithms when trained on an entirely different detector geometry. When compared
1956 with DL1d [124], GN1 improves on the c -jet rejection (light-jet rejection) by a factor
1957 of ~ 2 (~ 2.5) for jets in the $t\bar{t}$ sample at the 60% b -jet WP. Significant improvements
1958 in rejections are also observed for jets in the Z' sample.

1959 6.6 Conclusion

1960 In this chapter a novel jet tagger, GN1, is presented. The model has a graph neural
1961 network architecture and is trained with auxiliary training objectives, which are
1962 shown to improve the performance of the basic model. GN1 significantly improves

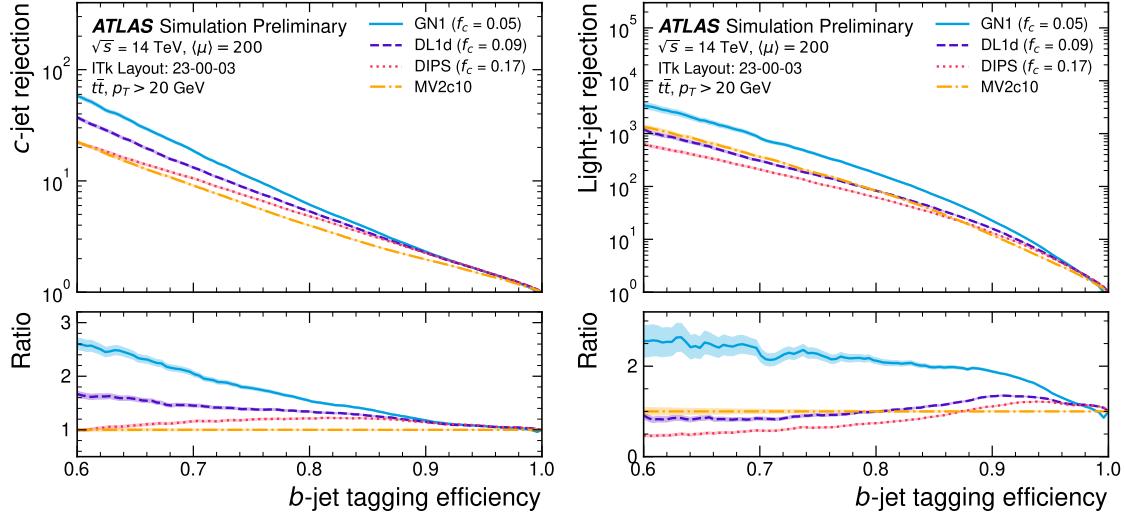


Figure 6.22: The c -jet rejection (left) and light-jet rejection (right) as a function of the b -jet efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ for events with a centre of mass energy $\sqrt{s} = 14 \text{ TeV}$ [126]. The ratio to the performance of the DL1d algorithm 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.

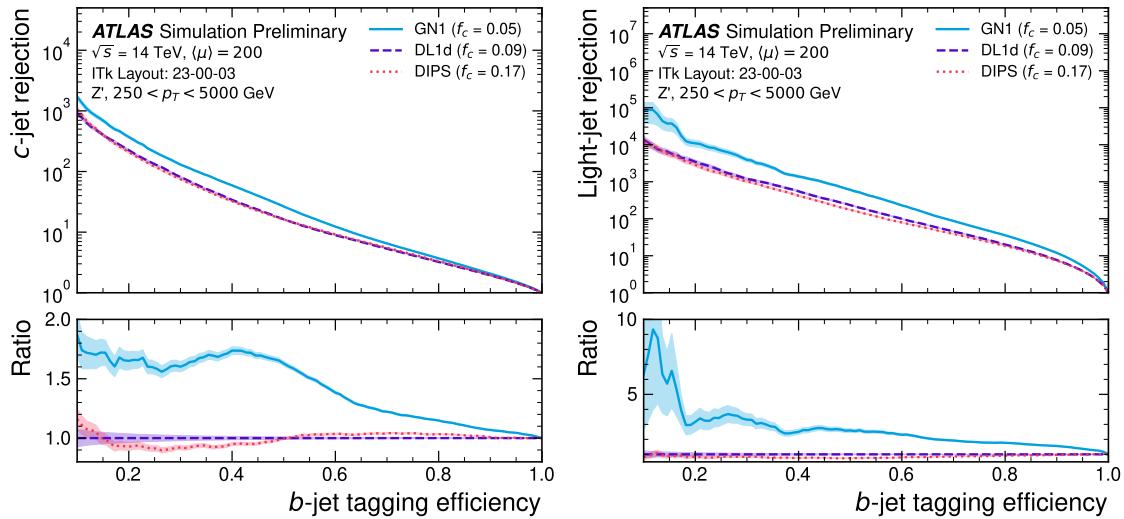


Figure 6.23: The c -jet rejection (left) and light-jet rejection (right) as a function of the b -jet efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$ for events with a centre of mass energy $\sqrt{s} = 14 \text{ TeV}$ [126]. The ratio to the performance of the DL1d algorithm is 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 ~ 2.1 and ~ 1.8 respectively at a b -jet tagging efficiency of 70% when compared with DL1r. For jets in the Z' sample with $250 < p_T < 5000$ GeV, GN1 improves the c -jet rejection by a factor of ~ 2.8 and light-jet rejection by a factor of ~ 6 for a comparative b -jet efficiency of 30%.

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

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

- ¹⁹⁹⁵ The model could also be repurposed as a pile-up jet tagger, or general primary and
¹⁹⁹⁶ secondary vertexing tool.

₁₉₉₇ **Chapter 7**

₁₉₉₈ **Boosted VHbb Analysis**

₁₉₉₉ The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [14, 15],
₂₀₀₀ is predicted by the standard model to decay primarily to a pair of b -quarks, with
₂₀₀₁ a branching fraction of 0.582 ± 0.007 for $m_H = 125$ GeV [26]. Observation of this
₂₀₀₂ decay mode was reported by ATLAS [113] and CMS [28] in 2018, establishing the
₂₀₀₃ first direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks
₂₀₀₄ (see Section 2.2.2). The $H \rightarrow b\bar{b}$ process is also important for constraining the total
₂₀₀₅ decay width of the Higgs [127].

₂₀₀₆ Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion
₂₀₀₇ as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet
₂₀₀₈ background and so overall sensitivity to the Higgs is low. The QCD multijet
₂₀₀₉ background refers to events containing one or more strongly produced jets which
₂₀₁₀ are not the decay product of heavy resonances, for example $g \rightarrow q\bar{q}$. The gluon-
₂₀₁₁ gluon fusion channel contains to leading order only jets in the final state, and
₂₀₁₂ therefore it is extremely difficult to distinguish signal events from the overwhelming
₂₀₁₃ multijet background. The $H \rightarrow b\bar{b}$ observation therefore searched for Higgs bosons
₂₀₁₄ produced in association with a vector boson V (where V can be a W or Z boson)
₂₀₁₅ which subsequently decays leptonically. The leptonic final states allow for leptonic
₂₀₁₆ triggering whilst at the same time significantly reducing the multijet background.

₂₀₁₇ Two full Run 2 dataset analyses were carried out as a follow-up to the $H \rightarrow$
₂₀₁₈ $b\bar{b}$ observation [113]. Similar to the observation, both measured the associated
₂₀₁₉ production of a Higgs with a vector boson, with the Higgs boson decaying to a pair
₂₀₂₀ of b -quarks. The first analysis [128] was focussed on the resolved phase-space, where
₂₀₂₁ the Higgs-jet candidate is reconstructed as two distinct jets with radius parameter

2022 $R = 0.4$. The second analysis [129] was focussed on the boosted phase-space, where
2023 the Higgs-jet candidate has a sufficiently large transverse momenta such that it can
2024 be reconstructed as a single jet with a radius parameter of $R = 1.0$. This chapter
2025 will focus on the latter analysis. The analysis is outlined in Section 7.1. Modelling
2026 studies performed are detailed in Section 7.2, and the results of the analysis are
2027 presented in Section 7.4. This analysis has been published in Ref. [129]. Figures and
2028 tables from Ref. [129] are reproduced here.

2029 7.1 Analysis Overview

2030 The boosted VH , $H \rightarrow b\bar{b}$ analysis is focused on the high transverse momentum
2031 regime, which has the benefit of being more sensitive to physics beyond the Standard
2032 Model [130], but the disadvantage of being more challenging due to the increased
2033 difficulty in the accurate reconstruction of high transverse momentum physics objects
2034 (discussed in Chapter 4). In order to focus on the high- p_T regime, the reconstructed
2035 vector boson p_T^V is required to be $p_T^V > 250$ GeV (see Section 7.1.2). Events are
2036 also split into two p_T^V bins with the first bin covering $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and
2037 the second covering $p_T^V > 400 \text{ GeV}$, which allows the analysis to benefit from the
2038 improved signal-to-background in the high- p_T regime.

2039 The previous ATLAS analysis in Ref. [113] was primarily sensitive to vector bosons
2040 with a more modest p_T^V boost in the region of 100–300 GeV. In this regime, the
2041 Higgs candidate was reconstructed using a pair of jets with radius parameter of
2042 $R = 0.4$, called small- R jets. However in the high- p_T regime, the decay products
2043 of the Higgs boson become increasingly collimated and the small- R jets may
2044 not be individually resolved. In order to enhance the reconstruction of the Higgs
2045 boson candidate, this analysis uses a large- R jet with radius parameter $R = 1.0$ to
2046 reconstruct the Higgs boson candidate (see Section 3.4.3). The Higgs candidate is
2047 required to have exactly two ghost-assciated (see Section 3.4.3) and b -tagged variable-
2048 radius track-jets. The candidate large- R jet is reconstructed using jet substructure
2049 techniques, in particular it is trimmed by removing soft and wide-angle components,
2050 which helps to remove particles from the underlying event and pile-up collisions [131].
2051 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 ℓ is an electron or muon and ν is a neutrino. Each channel has a dedicated set of selections which are listed in more detail in Section 7.1.3. Events in the 0- and 1-lepton channels are further split depending on the number of additional small- R jets not matched to the Higgs-jet candidate. The high-purity signal region (HP SR) has zero such jets, while the low-purity signal region (LP SR) has one or more, and therefore absorbs a larger number of background $t\bar{t}$ events. Maintaining a high purity signal region is important for the extraction of the signal yield. 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 described in Section 7.1.4. A complete overview of the different analysis regions is given in Table 7.1.

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

Table 7.1: Summary of the definitions of the different analysis regions [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

2071 model (described henceforth only as “the fit”) is described in more detail in described
2072 in Section 7.3.

2073 7.1.1 Data & Simulated Samples

2074 Data from centre-of-mass energy $\sqrt{s} = 13$ TeV proton-proton collisions at the LHC
2075 recorded over the course of Run 2 (between 2015 and 2018) were used for the analysis.
2076 The resulting dataset corresponds to a total integrated luminosity of 139 fb^{-1} (see
2077 Fig. 3.4).

2078 An overview of the MC simulated samples used in the analysis is given in Table 7.2.
2079 These samples are used to model the signal and background processes relevant to the
2080 analysis, with the exception of the multijet background which is modelled using a
2081 data-driven technique. Data and simulated events are reconstructed using the same
2082 algorithms, and a reweighting is applied to the simulated events in order to match
2083 the pile-up distribution observed in the data.

2084 7.1.2 Object Reconstruction

2085 The presence of neutrinos in the $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \ell\ell b\bar{b}$ signatures can be
2086 inferred from a momentum imbalance in the transverse plane Section 3.4.5. The
2087 vector boson transverse momentum p_T^V is reconstructed as the missing transverse
2088 energy E_T^{miss} in the 0-lepton channel, as the magnitude of the summed $\mathbf{E}_T^{\text{miss}}$ and
2089 charged-lepton momentum in the 1-lepton channel, and as the transverse momentum
2090 of the 2-lepton system in the 2-lepton channel (see Section 3.4.5).

2091 Electrons and muons are reconstructed as outlined in Section 3.4.4, and following the
2092 approach described in Ref. [113]. Leptons are required to satisfy the selections listed
2093 in Table 7.3. *Baseline* electrons are required to pass the likelihood-based method
2094 described in Section 3.4.4, and *Signal* electron additionally are required to satisfy
2095 a tighter likelihood identification selection. *Baseline* muons are required to pass
2096 the ‘loose’ identification described in Ref. [77], while *signal* muons are required to
2097 pass the ‘medium’ identification working point. All signal leptons are required to
2098 additionally satisfy a $p_T > 27$ GeV selection criteria, except for muons in the 1-lepton
2099 channel where a cut of 25 GeV is used. The number of baseline leptons is used to

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 (*) [104]	PyTHIA 8.212 [106]	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) ^(†) + 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 [107]	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 (α_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].

2100 categorise the event into the 0-, 1- or 2-lepton channels. The 1- and 2-lepton channels
2101 additionally require one signal lepton to be present.

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.

2102 The analysis makes use of large- R and variable-radius small- R track-jets, which
2103 are described in Section 3.4.3. The large- R jets are used to reconstruct the Higgs
2104 boson candidate, while the small- R jets are used for b -tagging and for selection of
2105 the analysis region. The track-jets matched to the Higgs candidate are b -tagged
2106 using the MV2c10 b -tagging algorithm (see Chapter 4). The efficiency of the tagging
2107 algorithm is calibrated to events in data [166–168]. The jet tagging strategy relies
2108 on extensive studies into track-jet b -tagging in boosted topologies [169, 170].

2109 7.1.3 Selection Criteria

2110 An extensive list of selection cuts are applied to each event in order to reject
2111 background events whilst retaining as many signal events as possible. A full list of
2112 selection cuts applied to the different analysis regions is given in Table 7.4, while
2113 some key selections are listed below.

2114 All channels require events with at least one large- R jet with $p_T > 250 \text{ GeV}$
2115 and $|\eta| < 2.0$. The vector boson transverse momentum is also required to satisfy
2116 $p_T^V > 250 \text{ GeV}$. The Higgs candidate is chosen as the highest p_T large- R jet satisfying
2117 these requirements. As mentioned, the candidate large- R jet is required to have
2118 two ghost-assciated and b -tagged variable-radius track-jets. These track-jets are
2119 required to have at least two constituent tracks with $p_T > 500 \text{ MeV}$ and $|\eta| < 2.5$.
2120 The track-jets themselves must satisfy $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$.

In the 0-lepton channel, trigger selections are applied using an E_T^{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^{miss} trigger as the 0-lepton channel is used. Since muons are not used for the E_T^{miss} trigger calculations, this is in effect a p_T requirement on the muon-neutrino system, which in the analysis phase space is more efficient than a single-muon trigger. The 2-lepton channel uses the same triggering strategy as the 1-lepton channel. In all channels, the trigger selections applied are fully efficient for events selected using the full requirements in Table 7.4.

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

7.1.4 Control Regions

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

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

7.1.5 Background Composition

After the selections described in Section 7.1.3 the number of background events mimicking the VH , $H \rightarrow b\bar{b}$ signal is greatly reduced. However, the number of background events still greatly outnumbers that of signal events. The background processes are channel dependent. In the 0-lepton channel the dominant sources of

Selection	0 lepton channel	1 lepton channel	2 leptons channel	
Trigger	E_T^{miss}	e sub-channel	μ sub-channel	e sub-channel
Leptons	0 baseline leptons	Single electron	E_T^{miss}	Single electron
		1 signal lepton		2 baseline leptons among which
		$p_T > 27 \text{ GeV}$	$p_T > 25 \text{ GeV}$	≥ 1 signal lepton, $p_T > 27 \text{ GeV}$
		no second baseline lepton		both leptons of the same flavour
				-
				opposite sign muons
E_T^{miss}	$> 250 \text{ GeV}$	$> 50 \text{ GeV}$	-	-
p_T^V			$p_T^V > 250 \text{ GeV}$	-
Large- R jets		at least one large- R jet, $p_T > 250 \text{ GeV}, \eta < 2.0$		
Track-jets		at least two track-jets, $p_T > 10 \text{ GeV}, \eta < 2.5$, matched to the leading large- R jet		
b -tagged jets		leading two track-jets matched to the leading large- R must be b -tagged (MV2c10, 70%)		
m_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.

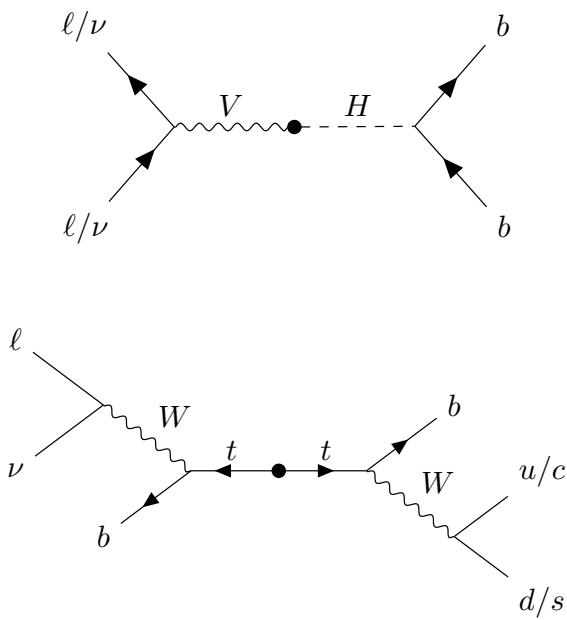


Figure 7.1: Diagrams of the signal process (top) and $t\bar{t}$ background (bottom). Object to the right of centre are reconstructed within the large- R jet. For the $t\bar{t}$ background, the large- R jet contains a mis-tagged c -jet or (less often) a mis-tagged light-jet. The contribution in the 0-lepton channel results from hadronically decaying τ lepton, or a electron or muon which is out of the analysis acceptance.

2150 backgrounds are $Z + \text{jets}$ ($Z \rightarrow \nu\nu$) and $t\bar{t}$, with $W + \text{jets}$ and diboson events being
2151 subdominant. In the event of $W \rightarrow \tau\nu$, and subsequent hadronic decay of the τ or
2152 lack of successful reconstruction/selection of the leptonic decay products, $W + \text{jets}$
2153 can also contribute to the 0-lepton channel. $t\bar{t}$ and $W + \text{jets}$ (with a leptonic decay
2154 of the W as in $W \rightarrow \ell\nu$) are dominant in the 1-lepton channel, while single-top is
2155 subdominant. In the 2-lepton channel, $Z + \text{jets}$ ($Z \rightarrow \ell\ell$) is again dominant followed
2156 by $Z Z$ diboson events.

2157 The diboson background VV consists primarily of WZ and ZZ events in which the
2158 Z decays to a pair of b -quarks. This process very closely matches the signal, with
2159 a resonant peak occurring at $m_Z = 91$ GeV and so is considered as an irreducible
2160 background ($V+b$ -jets is also irreducible).

2161 The $t\bar{t}V$, $t\bar{t}H$ and multijet backgrounds are negligible in the analysis phase space
2162 after the selections have been applied, with the exception of the 1-lepton electron
2163 sub-channel, in which multijet background is not negligible. The multijet background
2164 is made up of events where the isolated leptonic signature has been mimicked by
2165 either a jet or a muon or electron from a semi-leptonic heavy flavour decay, where
2166 the lepton has escaped the jet.

2167 The contributions from the different backgrounds are modelled using Monte Carlo
2168 event generators and the uncertainties associated with these samples are studied in
2169 Section 7.2. The multijet background is modelled using a data-driven technique.

2170 7.2 Systematic Uncertainties & Background 2171 Modelling

2172 Systemic uncertainties are extensively employed to give the fit model enough flexibility
2173 to account for inaccuracies in the various inputs. Two main types of systematic
2174 uncertainty are considered: experimental and modelling. Experimental uncertainties
2175 arise due to the imperfect reconstruction algorithms (in particular the jet recon-
2176 struction and b -tagging algorithms), and due to the imperfect modelling of pile-up
2177 and other effects, as described in Section 7.2.3. Modelling uncertainties arise due
2178 to the imperfections in the Monte-Carlo simulations used to model the signal and
2179 background events. In order to observe a certain process, for example VH , $H \rightarrow b\bar{b}$,

2180 an increase in the number of observed events with respect to the background-only
2181 hypothesis is looked for. The excess is often relatively small against the total number
2182 of background events, and hence accurate modelling of the expected number of
2183 background and signal events is crucial for successfully performing the analysis.
2184 Particular care is paid to the uncertainties on the modelling predictions as discussed
2185 in this section.

2186 Modelling uncertainties are described in detail in the following sections. *Nominal*
2187 samples are used as a reference to which different variations can be compared.
2188 The nominal samples are chosen as the best possible representation of the underlying
2189 physical process. *Alternative* samples are used to understand inaccuracies that may
2190 be present in the nominal samples. Some aspect of the nominal model is varied, and
2191 the discrepancy with respect to the nominal model is quantified. The discrepancy is
2192 used to estimate a systematic uncertainty associated with the model parameter which
2193 was varied. The alternative samples are sometimes obtained via internal weight
2194 variations or parameterisation methods, rather than by re-running the simulation.
2195 This is discussed in more detail in Section 7.2.1.

2196 Modelling studies involving c - and light-jets is hampered by the low number of events
2197 available after the analysis selection is applied, due to the high rejection rates of the
2198 b -tagging algorithm MV2c10. For modelling studies, truth tagging (TT) is therefore
2199 employed to ensure sufficient numbers of jets are available to calculate uncertainties.
2200 TT works by computing a 2-dimensional efficiency map using the jet p_T and η . The
2201 two leading track-jets associated to the large- R jet are weighted based on their p_T
2202 and η using the pre-calculated efficiency map, rather than being required to explicitly
2203 pass the b -tagging requirement.

2204 7.2.1 Implementation of Variations

2205 Modelling variations are implemented in different ways, depending on the associ-
2206 ated uncertainty. Table 7.5 lists the different sources of uncertainty described in
2207 Section 7.2.2 and for each lists the implementation. As production of high-stastic
2208 MC samples is computationally expensive, a technique in state of the art simulation
2209 packages is to store some sources of variation as internal weights, which can be
2210 generated alongside the nominal samples, saving computation time. The nominal

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

Table 7.5: 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.

2211 sample then effectively contains information about an ensemble of different samples,
2212 corresponding to different model parameters, which are accessible via reweightings.
2213 When filling histograms for the variations, bins are incremented by the internal
2214 weight of the event associated with the variation in question.

2215 While the inclusion of internal weight variation in MC event generators has decreased
2216 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently
2217 some sources of systematic uncertainty that are unable to be stored as internal weight
2218 variations due to technical limitations. Two examples are the choice of resummation
2219 and merging scales. A method to parameterise the systematic variation using
2220 one sample, and to then apply this parameterisation to another sample, has been
2221 developed by ATLAS [171]. This method was used to derive resummation and
2222 merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous (lower
2223 statistic) SHERPA 2.1 alternative sample. The resulting uncertainties were studied
2224 and found to be negligible in comparison with systematics from other sources.

2225 7.2.2 Sources of Systematic Modelling Uncertainties

2226 This section briefly describes the different sources of uncertainty in the analysis, and
2227 how each is implemented. For each source of uncertainty, acceptance and shape
2228 uncertainties are derived. Acceptance uncertainties account for the uncertainty in
2229 the overall number of events in each channel, and for the migration of events between

2230 different analysis regions. Meanwhile, shape uncertainties account for differences in
2231 the shapes, but not overall normalisations, of the large- R jet mass.

2232 QCD Scales

2233 The $V + \text{jets}$ matrix element calculations contains infrared and ultraviolet divergences.
2234 These are handled by introducing arbitrary parameters corresponding to the renor-
2235 malisation scale (μ_R) and factorisation scale (μ_F). Physical observables are not
2236 dependent on these parameters when using the infinite perturbation series expansion,
2237 however at some fixed order in QCD a limited dependence is present. To assess the
2238 impact of this, both μ_R and μ_F are independently varied from their nominal values
2239 by factors of 0.5 and 2 to account for higher order corrections to the calculation of
2240 the matrix element used to simulate the process.

2241 PDF Sets

2242 Parton distribution functions (PDFs) specify the probability of finding a parton with
2243 a given momentum inside a hadron (in this case, colliding protons). PDFs have
2244 to be derived from data and are a significant source of uncertainty in analyses of
2245 hadronic collision data. There are three sources of PDF uncertainties: the statistical
2246 and systematic errors on the underlying data used to derive the PDFs, the theory
2247 which is used to describe them (which is based on some fixed order perturbative
2248 QCD expansion), and finally the procedure which is used to extract the PDFs from
2249 the data. PDF-related uncertainties were derived following Ref. [163]. This involves
2250 considering 100 PDF replicas which, when combined, form a central value and
2251 associated uncertainty, and also in parallel direct changes to the central values of
2252 PDFs using the MMHT2014 [172] and CT14NLO [173] PDF sets.

2253 Event Generator

2254 The choice of parton shower (PS) and underlying event (UE) generators can affect
2255 the analysis outcome. Changing these models modifies several aspects of the event
2256 generation at the same time, such as the accuracy of matrix element predictions and

2257 different approaches to parton showering. This change tends to lead to the largest
2258 discrepancy with respect to the nominal samples.

2259 Resummation and Merging Scales

2260 Resummation is a technique used in QCD to help cope with calculations involving
2261 disparate energy scales, and involves the introduction of an associated resummation
2262 scale, the choice of which introduces some systematic uncertainty into the model.
2263 Parton showering models are accurate when simulating low- p_T radiation, however
2264 inaccuracies start to arrive when simulating hard emissions. To combat this, par-
2265 ton showering models utilise more precise matrix element calculations above some
2266 momentum threshold. The choice of threshold, or *merging scale* introduces some
2267 uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale
2268 variations are available for a subset of the SHERPA samples. The number of available
2269 events is significantly lower than the number of events in the nominal sample, and no
2270 statistically significant discrepancy with respect to the nominal samples is observed.
2271 The corresponding uncertainties and therefore neglected.

2272 7.2.3 Sources of Experimental Uncertainties

2273 The main experimental uncertainties in the analysis are due to the following sources:

- 2274 • The small- R jet energy scale and resolution, which are informed by in situ
2275 calibration studies and the dependence of the jet energy on the level of pile-
2276 up [68].
- 2277 • The large- R jet energy and mass scales and resolutions. The scales are calibrated
2278 as described in Ref. [72], and an uncertainty of 2% and 20% is applied for the
2279 jet energy and mass resolutions, respectively.
- 2280 • b -tagging uncertainties, which are computed separately for b -, c - and light-flavour
2281 jets as described in the calibration studies in Refs. [166–168]. An additional
2282 extrapolation uncertainty is added to account for jets with transverse momenta
2283 above that which is accessible in the calibration analyses.

what does
in situ mean
in this con-
text

- 2284 • Uncertainties associated with the lepton energy and momentum scales, and
2285 reconstruction and identification efficiencies.
- 2286 • Uncertainty on the pile-up models which are used in the simulated samples,
2287 described in Ref. [174].
- 2288 • Uncertainties associated with the reconstruction of the missing transverse energy
2289 E_T^{miss} , which have various sources as described in Ref. [78].

2290 7.2.4 Vector Boson + Jets Modelling

2291 After event selection, the $V+jets$ background is a dominant background in all three
2292 analysis channels as described in Section 7.1.5. The $V+jets$ samples are split into
2293 categories depending on the truth flavour of the track-jets which are ghost-associated
2294 to the large- R jet Higgs candidate. The categories are $V+bb$, $V+bc$, $V+bl$, $V+cc$,
2295 $V+cl$, $V+ll$, and $V+hf$ refers collectively to the categories containing at least one b -
2296 or c -jet. $V+bb$ is dominant generally accounting for 80% of the events, while $V+hf$
2297 accounts for around 90% of the events. The full flavour composition breakdown for
2298 each channel and analaysis region are given in Tables 7.6, 7.8 and 7.9.

2299 In order to access uncertainties associated with the use of MC generators, variations
2300 of the simulation are produced using alternative generators or variation of nominal
2301 generator parameters as described in Section 7.2.1. As described in Section 7.1.1,
2302 the nominal MC event generator used for $V+jets$ events is SHERPA 2.2.1, while
2303 MADGRAPH5_AMC@NLO+PYTHIA8 (which uses a different parton showering
2304 model) is used as an alternative generator.

2305 Modelling systematics can have several impacts, including affecting the overall
2306 normalisation for different processes, the relative acceptances between different
2307 analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR,
2308 and between p_T^V bins), and the shapes of the m_J distributions. Since the fit model
2309 fits only the large- R jet mass m_J to data, all shape uncertainties are estimated with
2310 respect to this observable. Several sources of uncertainty, summarised in Section 7.2.2,
2311 have been assessed.

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

Table 7.6: 0-lepton $W+jets$ nominal sample flavour composition and total event yield [175]. 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
Wbb	$77.2\% \pm 2.6\%$	$72.4\% \pm 4.3\%$	$77.8\% \pm 1.8\%$	$69.3\% \pm 2.5\%$	$64.6\% \pm 4.9\%$	$53.5\% \pm 6.3\%$
Wbc	$7.4\% \pm 0.7\%$	$7.3\% \pm 1.1\%$	$6.6\% \pm 0.4\%$	$6.3\% \pm 0.6\%$	$13.7\% \pm 1.9\%$	$16.4\% \pm 3.7\%$
Wbl	$4.0\% \pm 0.5\%$	$6.7\% \pm 1.1\%$	$5.1\% \pm 0.3\%$	$8.7\% \pm 0.8\%$	$10.3\% \pm 1.7\%$	$14.6\% \pm 3.0\%$
Wcc	$6.2\% \pm 1.1\%$	$5.5\% \pm 1.7\%$	$6.6\% \pm 0.6\%$	$6.4\% \pm 0.7\%$	$4.5\% \pm 1.7\%$	$9.5\% \pm 3.0\%$
Wcl	$3.6\% \pm 0.8\%$	$4.2\% \pm 1.8\%$	$2.8\% \pm 0.5\%$	$6.2\% \pm 0.8\%$	$4.6\% \pm 1.2\%$	$4.4\% \pm 1.5\%$
Wl	$1.5\% \pm 0.5\%$	$3.9\% \pm 1.3\%$	$1.1\% \pm 0.2\%$	$3.1\% \pm 0.5\%$	$2.3\% \pm 1.2\%$	$1.6\% \pm 0.6\%$
Events	477.1 ± 11.7	147.5 ± 6.4	784.7 ± 12.3	301.8 ± 7.2	68.7 ± 3.5	26.9 ± 2.0

Table 7.7: 1-lepton $W+jets$ nominal sample flavour composition and total event yield [175]. 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
Zbb	84.56%	81.84%	82.37%	76.06%	66.12%	63.18%
Zbc	6.03%	6.98%	5.80%	7.46%	15.04%	14.30%
Zbl	4.06%	6.55%	3.83%	6.59%	12.66%	12.81%
Zcc	3.68%	3.40%	5.82%	3.75%	3.36%	3.38%
Zcl	1.23%	0.44%	1.47%	3.97%	1.82%	4.95%
Zl	0.44%	0.78%	0.70%	2.16%	1.00%	1.38%
Events	259.91 ± 4.86	66.12 ± 2.04	420.45 ± 5.73	141.97 ± 2.50	43.49 ± 1.73	16.07 ± 0.83

Table 7.8: 0-lepton $Z+jets$ nominal sample flavour composition and total event yield [175]. 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
Zbb	80.80%	76.95%	79.76%
Zbc	8.10%	6.26%	7.60%
Zbl	4.95%	7.06%	5.52%
Zcc	3.97%	4.46%	4.10%
Zcl	1.61%	3.60%	2.14%
Zll	0.57%	1.68%	0.87%
Events	115.49 ± 2.42	42.42 ± 1.27	157.92 ± 2.73

Table 7.9: 2-lepton $Z + \text{jets}$ nominal sample flavour composition and total event yield [175]. $M p_T^V$ 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).

2312 Acceptance Uncertainties

2313 Several different types of acceptance uncertainties have been calculated and im-
 2314 plemented as nuisance parameters in the fit. These account for the uncertainty
 2315 in the overall number of events in each channel, and for the migration of events
 2316 between different analysis regions. The acceptance uncertainties relevant to the
 2317 $V + \text{jets}$ processes are summarised below.

- 2318 • **Overall normalisation:** only relevant where normalisation cannot be left
 2319 unconstrained (or “floating”, i.e. determined as part of the fit). The $V + \text{hf}$
 2320 component is left floating in the fit, with independent normalisations used for
 2321 $W + \text{hf}$ and $Z + \text{hf}$. The normalisations are mainly determined by the 1-lepton
 2322 (for $W + \text{hf}$) and 2-lepton (for $Z + \text{hf}$) regions respectively and then extrapolated
 2323 to the 0-lepton channel. The negligible $V + \text{jets}$ backgrounds were constrained to
 2324 their cross-sections in the fit. were constrained to their cross-sections
- 2325 • **SR-to-CR relative acceptance:** the uncertainty on the relative number of
 2326 $V + \text{jets}$ events in the signal and control regions.
- 2327 • **HP-to-LP relative acceptance:** the uncertainty on the relative number of
 2328 $V + \text{jets}$ events in the HP and LP SRs.
- 2329 • **Medium-to-high p_T^V relative acceptance:** the uncertainty on the relative
 2330 number of $V + \text{jets}$ events in the medium and high p_T^V bins.
- 2331 • **Flavour relative acceptance:** for each flavour $V + xx$, where $xx \in \{bc, bl, cc\}$
 2332 the ratio of $V + xx/V + bb$ events is calculated. This corresponds to the uncer-
 2333 tainty on the heavy flavour composition of the $V + \text{hf}$ background.

- 2334 • **Channel relative acceptance:** the uncertainty on the relative number of
 2335 $V+jets$ events between the channels.

2336 The uncertainties arising from the different sources described in Section 7.2.2 are
 2337 summed in quadrature to give a total uncertainty on each region. A summary of
 2338 the different acceptance uncertainties that were derived and subsequently applied
 2339 in the fit are given in Table 7.10. An effort has been made, wherever possible, to
 2340 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% [†]	40% [†]	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 [175]. $W+jets$ 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.

Shape Uncertainties

2342 In order to derive shape uncertainties for a given background or signal process,
 2343 normalised distributions of the reconstructed large- R Higgs candidate jet mass m_J
 2344 are compared for the nominal sample and variations. For each variation, the ratio of
 2345 the variation to nominal is calculated, the up and down variations are symmetrised,
 2346 and an analytic function is used to parameterise the ratio. If different analysis regions
 2347 or channels show the same pattern of variation, a common uncertainty is assigned.

2348 An example of a significant source of uncertainty, arising from choice of factorisation
 2349 scale μ_R is shown in Fig. 7.2. The HP SRs in the medium and high p_T^V bins are
 2350 shown for the 0-lepton channel for the $W+hf$ and $Z+hf$ jets. The 0- and 1-lepton
 2351 channels for the $W+hf$ contribution and the 0- and 2-lepton channels for the $Z+jets$
 2352 contribution were found to have compatible shapes in m_J across channels, and so
 2353 were jointly measured. An exponential function $e^{p_0 + p_1 x} + p_2$ has been fitted to the
 2354 ratio of the normalised distributions. The magnitude of the variation is p_T^V dependent,
 2355 and so separate uncertainties are implemented in the fit for each p_T^V region.

2356 The shape uncertainties for μ_R were derived on the SRs but are also applied to the
 2357 CRs, as the low statistics in the CRs make it difficult to derive dedicated shape
 2358 uncertainties. All the shape uncertainties are fully correlated accross regions.

2359 A comparison of the m_J shapes between SHERPA and MADGRAPH is shown in
 2360 Fig. 7.3. The plots are split by process and channel, but merged in SR purity and p_T^V
 2361 bins reflecting similarities between the m_J shapes and variations across these regions.
 2362 Due to the low statistics available for the alternate MADGRAPH sample, and the
 2363 lack of statistically significant variation between the samples, no additional shape
 2364 uncertainty was added to the fit in this case.

2365 The impacts of variations in the factorisation scale μ_F and the choice of PDF set on
 2366 m_J shape were also found to be negligible in comparison with μ_R and are hence no
 2367 additional uncertainty was added to the fit.

2368 7.2.5 Diboson Modelling

2369 The procedure to derive the uncertainties for the diboson background generally
 2370 follows that of $V+jets$.

2371 The nominal diboson samples are generated using SHERPA 2.2.1 (except for $gg \rightarrow VV$
 2372 which uses SHERPA 2.2.2) with the NNPDF3.0NNLO tune. Alternative samples were
 2373 generated using POWHEG interfaced with PYTHIA8, using the AZNLO shower tune
 2374 with the CTEQ6L1 PDFs [176]. Unlike SHERPA, POWHEG models the off-shell Z
 2375 contribution at NLO.

2376 Acceptance and shape uncertainties are derived in an analogous fashion to $V+jets$
 2377 as described below.

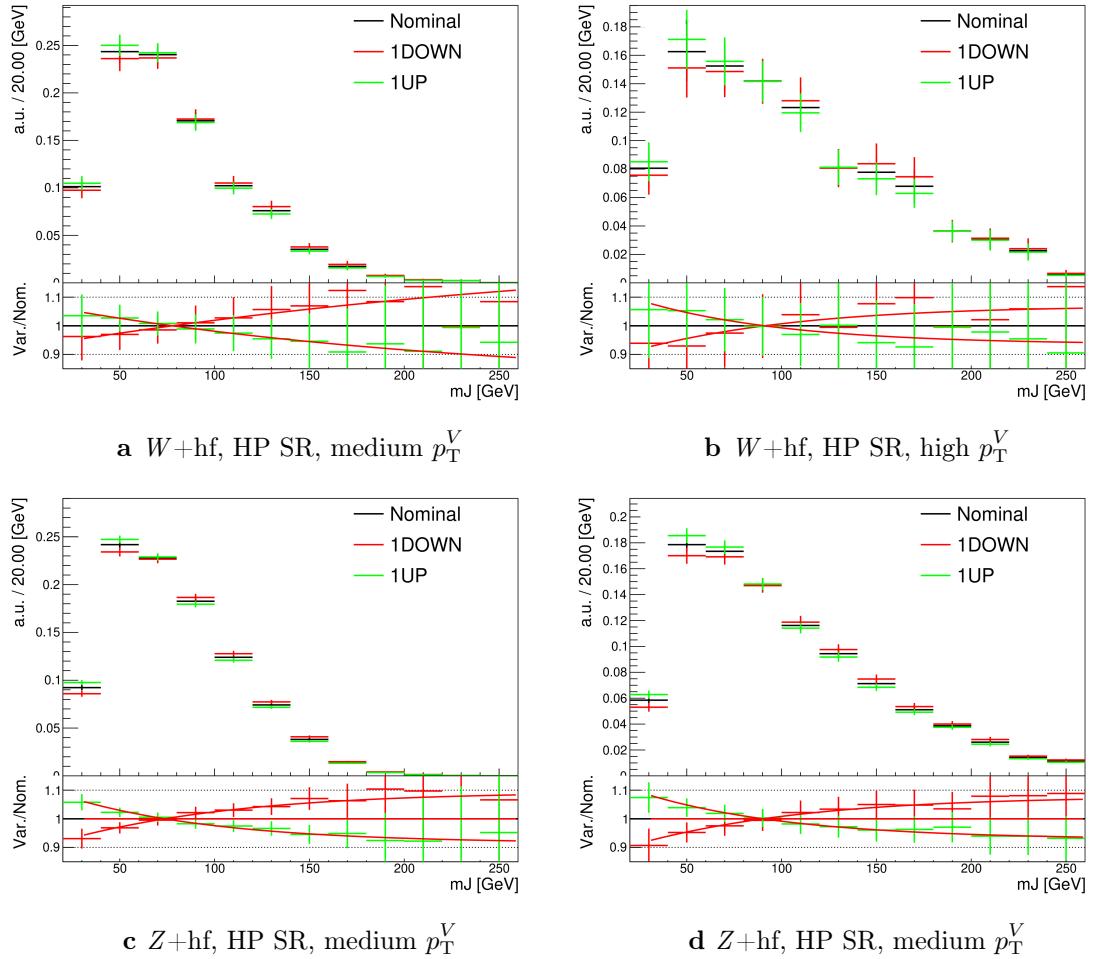


Figure 7.2: Leading large- R jet mass for the Z and $W+\text{hf}$ processes in the HP SR of the 0-lepton channel [175]. The renormalisation scale μ_r has been varied by a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to the ratio between the nominal and alternate samples.

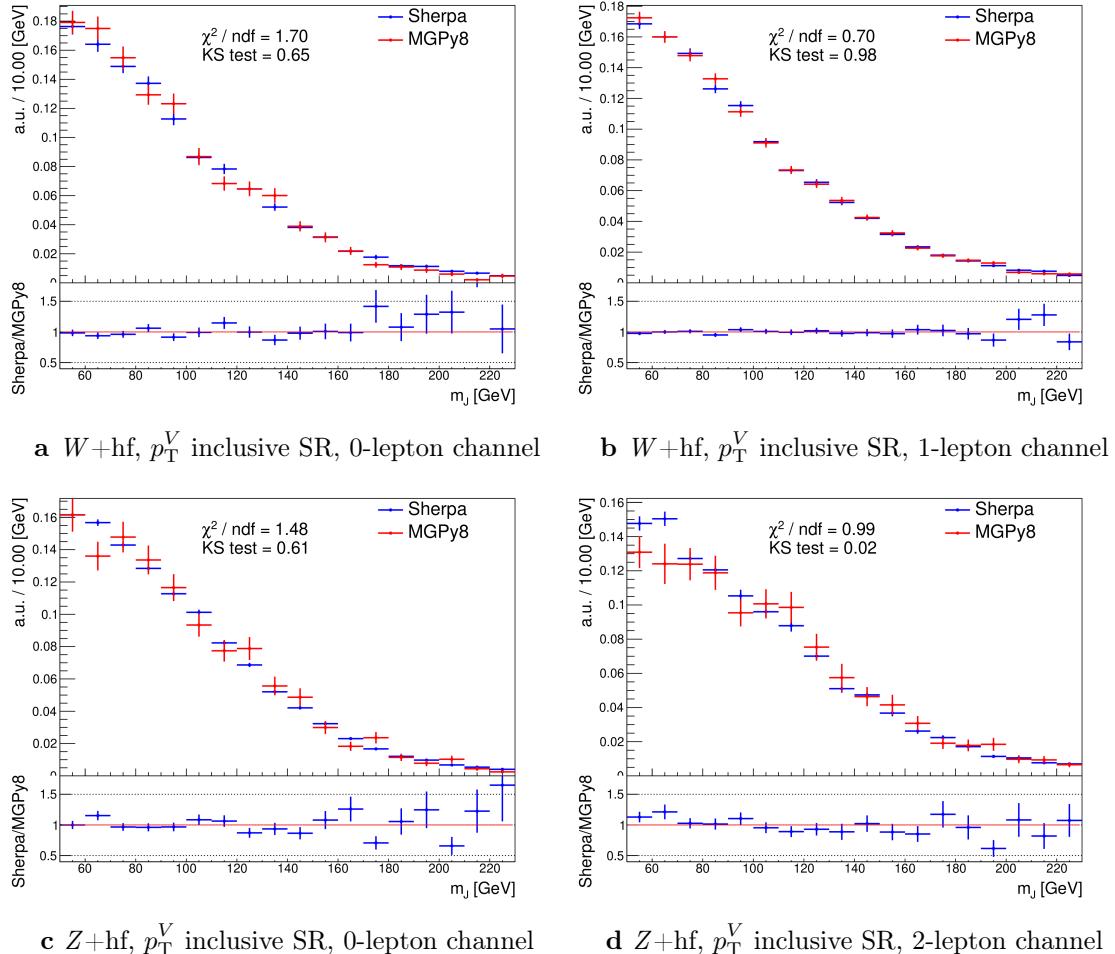


Figure 7.3: Leading large- R jet m_J inclusive in p_T^V for the $V + \text{hf}$ process modelled using both the SHERPA (blue) and MADGRAPH (red) samples [175]. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.

2378 Acceptance Uncertainties

2379 Diboson acceptance uncertainties are summarised in Table 7.11. Variations from μ_R ,
2380 μ_F , PDF choice and an alternative generator are considered and are combined via
2381 a sum in quadrature as described in Section 7.2.4. The largest modification to the
2382 nominal acceptance results from the POWHEG+PYTHIA8 alternate sample. Since
2383 the diboson contribution to the $t\bar{t}$ control region is negligible, no SR-to-CR relative
2384 acceptance uncertainty is necessary.

2385 For the WZ contribution, uncertainties are derived using the 1-lepton channel
2386 and applied to all three channels. The 1-lepton channel was used as it has the
2387 largest amount of available statistics. As far as was possible given the limited
2388 statistics available in the other channels, compatibility was checked between the
2389 derived uncertainties and the other channels. An additional 8% channel migration
2390 uncertainty is applied on the WZ 0-lepton channel. For the ZZ contribution, the
2391 normalisation uncertainty is calculated using the 2-lepton channel and applied to all
2392 three channels. The 0- and 1-lepton channels were found to have a similar HP-to-LP
2393 relative acceptance uncertainty of 18%. The 1-lepton medium-to-high p_T^V relative
2394 acceptance is based off the value obtained from the 2-lepton channel, since the
2395 1-lepton channel had an insufficient number of events to estimate the uncertainty
2396 directly. 30% and 18% channel migration uncertainties are applied in the 0- and
2397 1-lepton channels respectively.

2398 Since the contribution from WW is negligible, dedicated studies are not performed,
2399 but a 25% normalisation uncertainty is applied in all the three channels which is
2400 based on the modelling studies performed for the previous analysis [113].

2401 Shape Uncertainties

2402 Diboson shape uncertainties are derived in a similar fashion to $V+jets$. Only the
2403 uncertainties associated with systematic variation of μ_R and the alternate event
2404 generator have a non-negligible impact on the m_J shape. Variation of μ_R produces
2405 consistent m_J shape changes across all regions and channels, and hence only a single
2406 associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted
2407 to the symmetrised ratio.

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

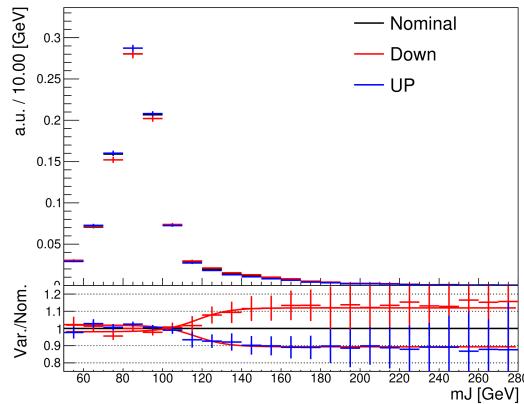


Figure 7.4: Leading large- R jet mass distribution for the WZ and ZZ processes, inclusive across all signal regions and lepton channels [175]. The renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.

The comparison between the nominal SHERPA and alternative POWHEG+PYTHIA8 samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both the WZ and ZZ processes. For these channels, the shape of m_J varies in opposite directions in the LP and HP signal regions. Shapes are similar between p_T^V bins, the 0- and 1-lepton channels and for WZ and ZZ . In order to reduce the effects of statistical fluctuations on the fit, these regions are merged before deriving the shape uncertainty. A third order polynomial is fitted to the ratio, and this function transitions to a constant piecewise function in the high mass region to accurately represent the shape. Dependence on the event generator was found to be negligible within statistical uncertainty in the 2-lepton channel, and so no uncertainty was applied. All diboson shape uncertainties are fully correlated in the fit.

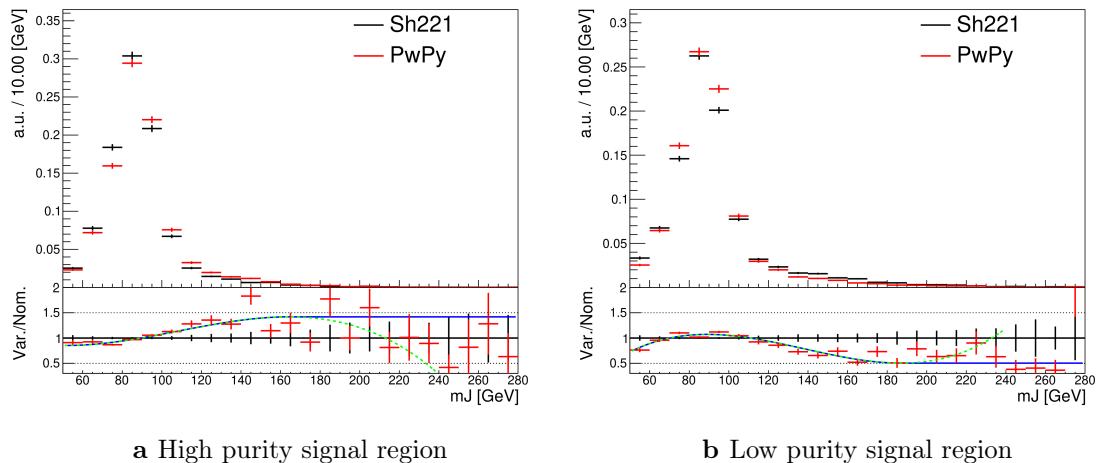


Figure 7.5: The comparison of the shapes of the large- R jet mass m_J between SHERPA (black) and POWHEG+PYTHIA8 (red) samples from WZ and ZZ process in high and low purity signal regions, integrated over p_T^V regions and the 0- and 1-lepton channels [175]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after a protection is added in the high mass region.

2419 7.2.6 $t\bar{t}$ and single-top Modelling

- 2420 The main features of the systematic uncertainties on the remaining two modelled
2421 backgrounds, $t\bar{t}$ and single-top, are described below.
2422 The modelling of the $t\bar{t}$ background uses a POWHEG+PYTHIA8 nominal sample.
2423 Two alternate samples were considered: POWHEG+HERWIG7 (providing a alternate

parton shower model) and MADGRAPH5+PYTHIA8.2 (providing an alternate hard scatter model). Effects of initial and final state radiation (ISR and FSR, respectively) were assessed using internal weight variations in the nominal sample. Acceptance and shape uncertainties were derived for each of the variations. Of the acceptance uncertainties, the largest contribution is due to the matrix element calculation, with the parton showering model being second. The ISR and FSR acceptance uncertainties were found to be subdominant. For the shape uncertainties, only the ISR and parton showering variations have non-negligible impacts on the m_J shape.

The dominant process contributing to the single-top background is Wt production, relevant for the 0- and 1-lepton channels. The same nominal and alternative samples are used as for the $t\bar{t}$ background. Again, ISR and FSR variations are obtained from internal weight variations in the nominal sample. At higher orders in QCD, diagrams contributing to the Wt production process can also be found in leading-order $t\bar{t}$ production processes. To account for the arising interference effects, the diagram removal (DR) scheme in Ref. [177] was employed for the nominal sample. The uncertainty on the DR scheme was assessed using an alternate sample using a diagram subtraction (DS) method which removes interference at the generator level. The largest sources of acceptance and shape uncertainties were due to this DS-DR variation.

7.2.7 Signal Modelling

The modelling of the systematic uncertainties affecting the signal processes follows the procedure described in Refs. [26, 178, 179]. The $qq \rightarrow VH$ signal samples are generated with POWHEG BOX v2 +GoSAM at next-to-leading order (NLO) accuracy in QCD. An additional $gg \rightarrow ZH$ sample is generated using POWHEG BOX v2 at leading order (LO) in QCD. In both cases, the generated events are interfaced with PYTHIA 8 for the parton showering modelling. An alternate HERWIG7 sample is used to assess the uncertainty on the parton showering model. Recommended systematic uncertainties on the signal production cross-sections and $H \rightarrow b\bar{b}$ branching ratio from the LHC Higgs Cross Section Working Group are applied [180, 181]. Acceptance and shape uncertainties arising from missing higher-order QCD and electroweak corrections, PDF uncertainties, renormalisation and factorisation scales, and alternate parton showering model are applied.

2456 7.3 Statistical Treatment

2457 A binned global maximum-profile-likelihood fit of the m_J distribution is performed
 2458 to extract information on the signal, combining all the analysis regions defined
 2459 in Table 7.1. The signal strength $\mu = \sigma/\sigma_{\text{SM}}$ is defined as the ratio between the
 2460 observed and predicted cross-sections, where $\mu = 0$ corresponds to the background-
 2461 only hypothesis and $\mu = 1$ corresponds to the SM prediction. This is a parameter of
 2462 interest (POI) which acts to scale the total number of signal events, and is determined
 2463 during the fit procedure.

2464 The present analysis makes use of two POIs. The first, μ_{VH}^{bb} , is the signal strength
 2465 for the VH , $H \rightarrow b\bar{b}$ process, the primary process under investigation. The diboson
 2466 production strength μ_{VZ}^{bb} for the VZ , $Z \rightarrow b\bar{b}$ process is measured simultaneously and
 2467 provides a validation of the analysis apparatus used for the primary $H \rightarrow b\bar{b}$ measure-
 2468 ment. Alongside the two POIs, the predictive model depends on several parameters
 2469 which are not the primary target of measurement, and represent the systematic
 2470 uncertainties discussed previously. These parameters are called nuisance parameters
 2471 (NPs), collectively referred to as θ . Freely floating background normalisations are
 2472 implemented as NPs and are also extracted during the fitting processes.

2473 7.3.1 Likelihood Function

2474 The statistical setup treats each bin as a Poisson counting experiment and is based on
 2475 the ROOSTATS framework [182]. The combined likelihood over N bins is constructed
 2476 as the product of Poisson probabilities in each bin. Considering the simplified case
 2477 of a single signal strength parameter μ , and neglecting sources of systematic or
 2478 statistical uncertainty, this is given by

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

2479 where s_i (b_i) is the expected number of signal (background) events in bin i , and n_i is
 2480 the number of observed data events in bin i .

2481 Treatment of Uncertainties

2482 Systematic uncertainties can modify the predicted signal and background yields s_i
 2483 and b_i . Each source of systematic uncertainty is taken into account by adding an
 2484 additional NP θ_j to the likelihood in the form of a Gaussian cost function. The
 2485 combined effect of the NPs is then

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

2486 where N_θ is the number of NPs, θ_j is the nominal value of the j th NP, $\hat{\theta}_j$ is the fitted
 2487 value, and σ_j is the corresponding associated prior uncertainty on the NP. As the
 2488 fitted value $\hat{\theta}_j$ deviates from its nominal value, a cost is introduced. The presence of
 2489 NPs modifies the likelihood as

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

2490 The predicted signal and background yields are also modified by the presence of the
 2491 NPs with

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

2492 For NPs which are left freely floating in the fit, no corresponding Gaussian constraints
 2493 are added to the likelihood.

2494 The pull of a NP is defined as the difference between the fitted value $\hat{\theta}_j$ and the
 2495 nominal value θ_j , divided by the uncertainty on the NP σ_j . To obtain the uncertainty
 2496 on the pull of a NP, the following procedure is used. The Hessian matrix \mathbf{H} is
 2497 calculated as

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_n} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} & \dots & \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_n \partial \theta_1} & \frac{\partial^2 \mathcal{L}}{\partial \theta_n \partial \theta_2} & \dots & \frac{\partial^2 \mathcal{L}}{\partial \theta_n^2} \end{bmatrix}. \quad (7.5)$$

2498 Taking the inverse of the Hessian matrix \mathbf{H}^{-1} yields the covariance matrix, from
2499 which the post-fit uncertainties on the different NPs can be extracted. If the post-fit
2500 uncertainty is smaller than the nominal uncertainty, additional information about
2501 the NP has been extracted by the fit, and NP is said to be *constrained*.

2502 The statistical uncertainty on the simulated events is implemented using a dedicated
2503 NP for each bin which can scale the background yield in that bin. Statistical NPs
2504 are also implemented using a Gaussian constraint.

2505 Smoothing and Pruning

2506 To simplify the fit to reduce and improve its robustness, systematic uncertainties
2507 are smoothed and pruned. Smoothing accounts for the large statistical uncertainty
2508 present in some samples that can lead to unphysical fluctuations in the shape
2509 systematics. The smoothing procedure relies on the assumption that the impact of
2510 systematics should be approximately monotonic and correlated between neighbouring
2511 bins.

2512 In addition to smoothing, pruning is the process of removing from the fit those
2513 systematics which only have a very small effect. This improves the stability of the
2514 fit by reducing the number of degrees of freedom. Acceptance uncertainties are
2515 pruned in a given region if they have a variation of less than 0.5%, or if the up and
2516 down variations have the same sign in that region. Shape uncertainties are pruned
2517 in a given region if the deviation in each bin is less than 0.5% in that region. In
2518 addition, acceptance and shape uncertainties are neglected in a given region for any
2519 background which makes up less than 2% of the total background in a given region.

2520 Fit Procedure and Statistical Tests

2521 The best-fit value of μ , denoted $\hat{\mu}$, is obtained via an unconditional maximisation of
 2522 the likelihood. The likelihood is also used to construct a statistical test which can
 2523 confirm or reject the background-only hypothesis. The test statistic q_μ is constructed
 2524 from the profile likelihood ratio,

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (7.6)$$

2525 where $\hat{\mu}$ and $\hat{\theta}$ are chosen to maximise the likelihood \mathcal{L} , and the profile value $\hat{\theta}_\mu$ is
 2526 obtained from a conditional maximisation of the likelihood for a specific choice of
 2527 $\mu = 0$ corresponding to the background-only hypothesis.

2528 The test statistic is used to construct a p -value which is used to probe the background-
 2529 only hypothesis. The p -value is typically reported in terms of the significance Z ,
 2530 defined as the number of standard deviations for a Gaussian Normal distribution
 2531 which will produce a one-sided tail integral equal to the p -value, as in

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z). \quad (7.7)$$

2532 Typically a value of $Z = 3$ constitutes *evidence* of a processes, while $Z = 5$ is required
 2533 for a *discovery*, or observation. Alongside the p -value, the best-fit value of the signal
 2534 strength $\hat{\mu}$ and its corresponding uncertainty are quoted, and compared to their
 2535 expected values.

2536 7.3.2 Background Normalisations

2537 The backgrounds which can be constrained by the fit are left freely floating and
 2538 the corresponding normalisation factors are extracted. Normalisation factors (NF),
 2539 represent the value by which the predicted normalisations are scaled, and are im-
 2540 plemented for the dominant backgrounds ($t\bar{t}$, $Z+hf$, $W+hf$). The NFs are also
 2541 subdivided into different regions of phase-space for $t\bar{t}$, given it is possible to obtain
 2542 a strong constraint in the individual channels. This also removes the need for an
 2543 extrapolation uncertainty.

2544 The normalisations and shapes of all other backgrounds, with the exception of the
2545 multijet background which is estimated using a data driven technique, are initialised
2546 using the nominal samples and the state-of-the art process normalisations, as outlined
2547 in Table 7.2.

2548 **7.3.3 Asimov Dataset & Expected Results**

2549 The Asimov dataset is constructed by replacing the data with the sum of the signal
2550 and background predictions $n_i = s_i + b_i$. A fit to this dataset using the nominal
2551 values of the NPs from the simulation will recover the input values and is useful for
2552 studying the expected result, in addition to constraints on and correlations between
2553 the NPs.

2554 Alternatively, a conditional fit to the Asimov dataset can be performed using values
2555 of the background NPs which are determined from an unconditional fit to data. The
2556 signal NPs and POIs are fixed at their nominal values from the SM simulation. The
2557 result of this fit can be used to calculate expected (median) significances given a
2558 more realistic background model, which can be compared to their observed values,
2559 as is done in Section 7.4.2.

2560 **7.4 Results**

2561 In the present analysis, the two signal strength parameters μ_{VH}^{bb} and μ_{VZ}^{bb} are extracted
2562 from a simultaneous maximisation of the likelihood described in Section 7.3. The
2563 results of the analysis are summarised in this section. The corresponding post-fit
2564 background normalisations are listed in Table 7.12. Post-fit m_J distributions are
2565 shown in Section 7.4.1. The observed signal strengths are given in Section 7.4.2,
2566 along with observed and expected significances. Finally in Section 7.4.3 the impact
2567 of systematic uncertainties on the results is examined.

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

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

2568 7.4.1 Post-fit Distributions

2569 In addition to the observed significance and signal strength, it is also necessary to
 2570 study the post-fit m_J distributions to compare level of the agreement between the
 2571 simulation (using the best-fit values of the signal strength $\hat{\mu}$ and the NP $\hat{\theta}$) and the
 2572 data. The best-fit values $\hat{\mu}$ and $\hat{\theta}$ are obtained from an unconditional fit to data over
 2573 all analysis regions. Post-fit m_J distributions are given for the signal regions in the
 2574 0-, 1- and 2-lepton channels in Fig. 7.6. The LP and HP regions are merged for the
 2575 0- and 1-lepton channels for the sake of simplicity. In general there is a good level of
 2576 agreement between the simulation and data, indicating the fit model is performing as
 2577 expected. Fig. 7.7 shows the post-fit plots for the $t\bar{t}$ control regions. Again, a good
 2578 level of agreement is observed given the statistical uncertainties on the distributions.

2579 7.4.2 Observed Signal Strength & Significance

2580 The measured signal strength is computed as the ratio between the measured signal
 2581 yield to the prediction from the SM. The combed result for all three lepton channels
 2582 and all analysis regions is given for μ_{VH}^{bb} in Eq. (7.8), and for μ_{VZ}^{bb} in Eq. (7.9). Both
 2583 results include a full breakdown of the systematic and statistical uncertainties.

$$\mu_{VH}^{bb} = 0.72_{-0.36}^{+0.39} = 0.72_{-0.28}^{+0.29}(\text{stat.})_{-0.22}^{+0.26}(\text{syst.}) \quad (7.8)$$

2584

$$\mu_{VZ}^{bb} = 0.91_{-0.23}^{+0.29} = 0.91 \pm 0.15(\text{stat.})_{-0.17}^{+0.24}(\text{syst.}) \quad (7.9)$$

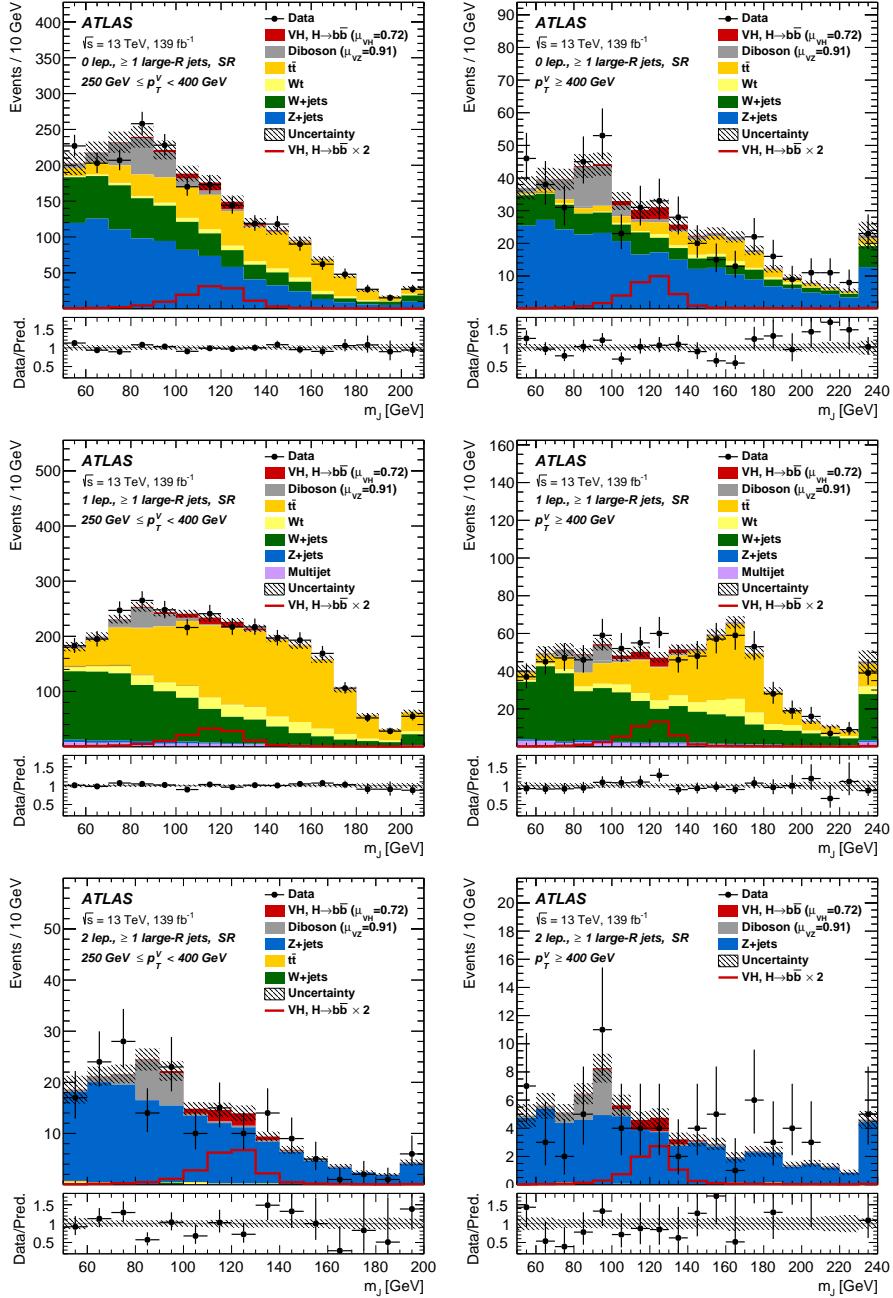


Figure 7.6: The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions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

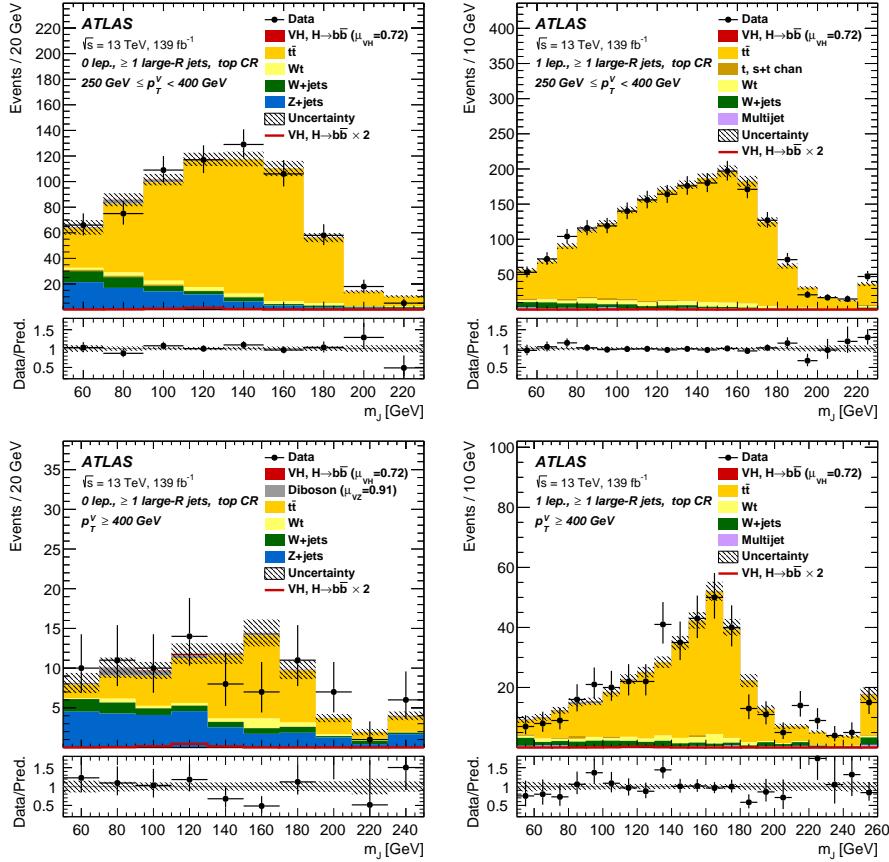


Figure 7.7: The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. The background contributions after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

2585 The results for μ_{VH}^{bb} and μ_{VZ}^{bb} are consistent with the expectation from the SM. The
 2586 μ_{VH}^{bb} measurement is dominated by statistical uncertainty, while the μ_{VZ}^{bb} measurement
 2587 is dominated by systematic sources of uncertainty. The measured signal strength
 2588 for μ_{VZ}^{bb} corresponds to an observed significance of 2.1 standard deviations, with an
 2589 expected (median) significance given the SM prediction of 2.7 standard deviations.
 2590 The diboson observed (expected) signal strength significance is 5.4 (5.7). These
 2591 results are summarised in Fig. 7.8, which shows the background-subtracted m_J
 2592 distribution. A clear signal excess is visible around the Higgs mass of $m_H = 125$ GeV.

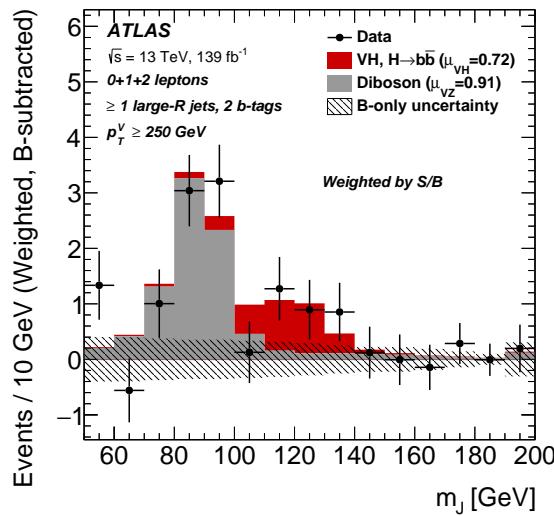


Figure 7.8: m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected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

2593 Compatability Studies

2594 Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting μ_{VH}^{bb}
 2595 into three separate POIs, one for each channel. A simultaneous fit to the channel
 2596 specific signal strengths can then be performed, which allows a comparison of the

2597 contributions from each channel. Fig. 7.9 compares the best-fit signal strengths.
 2598 The 0- and 1-lepton channels show a signal strength which is consistent with the
 2599 SM prediction, while the 2-lepton channel shows a small deviation within the 1σ
 2600 uncertainty. Overall, good compatibility is observed via a χ^2 test with a corresponding
 2601 p -value of 49%.

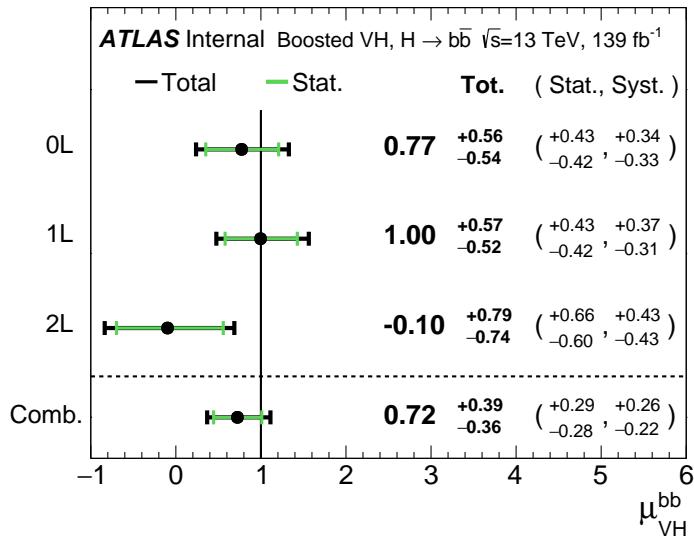


Figure 7.9: Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels splitted) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a χ^2 difference test and results in a p-value of 49% [129].

2602 7.4.3 Impact of Systematics

2603 The impact of systematic uncertainties on the final fitted value $\hat{\mu}^{bb}$ can be studied
 2604 using the NP rankings, and the uncertainty breakdown.

2605 Fig. 7.10 shows the NP ranking, which is used to visualise which out of the many
 2606 NPs involved in the fit have the largest impact on the sensitivity to the fitted
 2607 POI. To obtain the ranking, a likelihood scan is performed for each NP θ_j . First,
 2608 an unconditional fit is used to determine $\hat{\theta}_j$. The NP is then fixed to its post-fit
 2609 value varied by $\pm 1\sigma$, the fit is repeated and the best-fit value of the POI, $\Delta\hat{\mu}_{VH}^{bb}$, is
 2610 calculated, and used to rank the NPs. Also shown in Fig. 7.10 are the pulls and
 2611 constraints for the highest ranked NPs.

2612 The experimental uncertainty on the signal large- R jet mass resolution (JMR) has
 2613 the largest impact. JMR and jet energy scale (JES) uncertainties also have impacts
 2614 for the V +jets background and for the diboson background. The freely-floating
 2615 Z +hf normalisation is the second highest ranked NP, and is heavily constrained by
 2616 the fit. The VZ POI μ_{VZ}^{bb} is also a significant NP when considering the primary μ_{VH}^{bb}
 2617 measurement.

2618 The NP ranking highlights individual NPs which have a large impact on the POI
 2619 measurement sensitivity. Complementary information is provided at a higher level
 2620 by considering the overall impact of different groups of systematics. The groups
 2621 are constructed from NPs which have similar physical origin. The impact of each
 2622 group is calculated by running a fit with all the NPs in the given group fixed to their
 2623 nominal values. The uncertainty on the POI extracted from this fit is subtracted
 2624 in quadrature from the the uncertainty on the POI from the nominal fit, and the
 2625 resulting values are provided as the impact for each group. The full breakdown
 2626 for the observed impact of uncertainties on the μ_{VH}^{bb} signal strength is provided in
 2627 Table 7.13. The total systematic impact is the difference in quadrature between
 2628 the nominal uncertainty on μ_{VH}^{bb} and the combined statistical impact. The “data
 2629 stat only” group fixes all NPs at their nominal value, while the total statistical
 2630 impact fixes all NPs except floating normalisations. The floating normalisations
 2631 group fixes only the NPs associated with normalisation which are left floating in
 2632 the fit. The uncertainty on μ_{VH}^{bb} is dominated by combined statistcal effects (0.28),
 2633 although the combined impact of systematics (0.24) is of a comparable size. The
 2634 signal largest group is the data stat uncertainty (0.25), demonstrating that the
 2635 analysis would benefit from an increased integrated luminosity or improved efficiency
 2636 to select signal events (recall from Section 7.1.3 the signal efficiency is in the range of
 2637 10%). Of the experimental systematic sources of uncertainty, the dominant impact
 2638 is the experimental uncertainties associated with the reconstruction of large- R jets
 2639 (0.13). Other experimental sources of uncertainty are small in comparison. Modelling
 2640 uncertainties also have a large contribution to the overall systematic uncertainty. The
 2641 biggest contribution to the overall uncertainty is the combined statistical uncertainty
 2642 on the simulated samples (0.09). Out of the backgrounds, the W +jets and Z +jets
 2643 have the highest (0.06) and second-highest (0.05) impact respectively.
 2644

Perhaps add the STXS measurement (figure and table), which is a key result from this week

I would also consider adding the post-fit yield tables from the

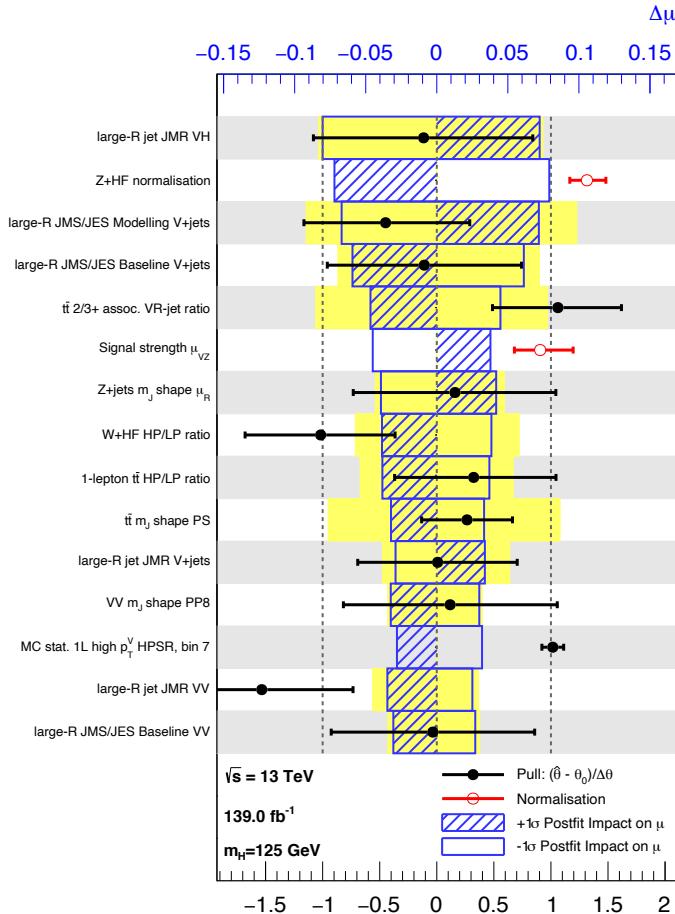


Figure 7.10: Impact of systematic uncertainties on the fitted VH signal-strength parameter $\hat{\mu}_{VH}^{\text{bb}}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value 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 ± 1 . The filled circles show the deviation of the fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value θ_0 expressed in standard deviations with respect to their nominal uncertainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show the fitted values and uncertainties of the normalization parameters that are freely floating in the fit. Pre-fit, these parameters have a value of one [175].

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

Table 7.13: Breakdown of the absolute contributions to the uncertainty on the signal strength μ_{VH}^{bb} obtained from the (1+1)-POI fit. The average impact represents the average between the positive and negative uncertainties on μ_{VH}^{bb} . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations [175].

2646 7.5 Conclusion

- 2647 The analysis of the associated production of vector bosons with boosted Higgs bosons
2648 decaying to a pair of b -quarks using large- R jets is presented. The Higgs candidate is
2649 reconstructed as a large- R jet in order to improve sensitivity in the boosted regime
2650 in which the Higgs decay products are significantly collimated. The analysis is
2651 performed using 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13\text{ TeV}$ collected
2652 throughout the duration of Run 2 of the LHC.
- 2653 In comparison with the null hypothesis, the Standard Model (SM) VH , $H \rightarrow b\bar{b}$
2654 process is found to have an observed significance of 2.1 standard deviations, whereas
2655 the corresponding expected significance is 2.7 standard deviations. The VH , $H \rightarrow b\bar{b}$
2656 process is measured simultaneously with the diboson VZ , $Z \rightarrow b\bar{b}$ process, which
2657 provide a cross-check for the main analysis. The observed (expected) significance for
2658 the diboson process is 5.4 (5.7).
- 2659 The statistical and systematic sources of uncertainty contribute a similar amount
2660 to the overall uncertainty on the result. This analysis would therefore likely benefit
2661 greatly from the improved b -tagging efficiency at high- p_{T} enabled by GN1 as discussed
2662 in Chapter 6, due to the associated reduction in statistical uncertainty provided by
2663 the increased number of events used in the analysis.
- 2664 The large- R jet mass resolution is found to be the dominant source of systematic
2665 uncertainty on the μ_{VH}^{bb} measurement. An improved method of reconstructing the
2666 large- R jet mass, for example by using a machine learning based regression approach,
2667 possibly as an additional auxiliary task to GN1 (see Chapter 6), would therefore be
2668 expected to reduce the systematic uncertainty on the μ_{VH}^{bb} measurement.

2669 Chapter 8

2670 Conclusion

2671 8.1 Summary

2672 The current understanding of particle physics contains many unanswered questions,
2673 and improving our understanding of the Standard Model is a promising way to
2674 attempt to answer some of them. One of the key objects which may enhance this
2675 understanding is the Higgs Boson, which was first observed only a decade ago and
2676 remains under intense scrutiny at the LHC. Given it's propensity to decay to heavy
2677 flavour b -quarks, reconstructing and identifying b -jets is of crucial importance to
2678 improving our understanding in this area. As discussed in Chapter 4, this task
2679 becomes increasingly difficult at high transverse momenta.

2680 One of the effects that hampered tracking and b -tagging performance at high- p_T was
2681 identified to be the increased rate of fake tracks. To address this issue, a tools was
2682 developed which was able to successfully identify fake tracks within jets 45% of the
2683 time, with a minimal loss of signal tracks of 1.2%. Removal of such tracks was found
2684 to improve the light-jet mistagging rate of the SV1 and JetFitter algorithms by up
2685 to 20% at high transverse momentum.

2686 A novel approach to b -tagging, GN1 was also developed using a Graph Neural Network
2687 (GNN) architecture. The model is encouraged to learn the topology of the jet through
2688 vertexing and track classification auxiliary tasks. As a single end-to-end trained
2689 model, GN1 simplifies the complexity of the flavour tagging pipeline and is able to
2690 achieve superior performance to the current state-of-the-art algorithms, which rely
2691 on a two-tiered approach. Compared with DL1r, GN1 improves the light-jet rejection

2692 by a factor of ~ 1.8 for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ at the 70%
2693 b -jet WP and by a factor of ~ 6 for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$
2694 for a corresponding b -jet efficiency of 30%. GN1 also demonstrates a significant
2695 improvement in the discrimination between b - and c -jets.

2696 The work in this thesis demonstrates that even with suboptimal track reconstruction
2697 in this regime, it is possible to make algorithmic advancements to the flavour tagging
2698 pipeline to improve the identification of b -jets. This work has impacts for any analysis
2699 which relies on the identification of b -jets, including those which are sensitive to the
2700 Higgs Boson.

2701 Analysis of VH , $H \rightarrow b\bar{b}$ events was also carried out with 139 fb^{-1} of Run 2
2702 ATLAS at $\sqrt{s} = 13 \text{ TeV}$. Various background modelling uncertainties were derived
2703 and investigations into the fit model were carried out. The analysis observed a
2704 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
2705 observed (expected) significance of 2.1σ (2.7σ). The result was validated using a
2706 simultaneous fit to the VZ , $Z \rightarrow b\bar{b}$ process, which acts as a cross check to validate
2708 the primary analysis.

add a bit more detail about the importance of this result and why it is interesting due to its sensitivity to new physics

2709 8.2 Future Work

2710 Additional algorithmic improvements are likely to yield further improved flavour
2711 tagging performance. Aside from these, large improvements to the flavour tagging
2712 performance will likely be possible if improvements are made to the b -hadron decay
2713 track reconstruction efficiency and accuracy.

2714 At the moment only the tracks from the Inner Detector and kinematic information
2715 about the jet are provided as inputs to the tagging algorithms. In Chapter 6 it
2716 was shown that the addition of a simple track-level variable corresponding to the
2717 ID of the reconstruction lepton to the model improved the performance. However
2718 there is still untapped potential in the form of additional information from the full
2719 parameters of the reconstructed leptons (making full use of the Calorimeters and
2720 Muon Spectrometer), the calorimeter clusters, and even the low level hits. Providing
2721 such additional inputs to the model is likely to complement the information provided
2722 by the tracks and further aid in the improvement of performance.

2723 On the output side, additional auxiliary training objectives may yield improved
2724 performance and also help to add to the explainability of the model. Regression of
2725 jet-level quantities such as the transverse momentum and mass, in addition to the
2726 truth b -hadron decay length are promising regression targets.

2727 The GN1 architecture can also be readily optimised for new use cases and topologies,
2728 as demonstrated by the studies described in Section 6.5. For example, a model
2729 with only hit-level information provided as inputs could be used for a fast trigger
2730 preselection on jets without the need to run the computationally expensive tracking
2731 algorithms. The model could also be repurposed for primary vertexing, or a pile-up
2732 jet tagger. Finally, the tagging of large- R jets would benefit those analysis that rely
2733 on it.

2734 Ultimately analysis which rely on the identification of heavy flavour jets will likely
2735 benefit immensely from the improved performance of the flavour tagging algorithms.
2736 For example, the $HH \rightarrow bbbb \dots$

make some
claim about
improved
selection
efficiency?

₂₇₃₇

Bibliography

- ₂₇₃₈ [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- ₂₇₃₉
- ₂₇₄₀ [2] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
₂₇₄₁ *Secondary vertex finding in jets with neural networks*, *The European Physical Journal C* **81** (2021) .
- ₂₇₄₂
- ₂₇₄₃ [3] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, **ATL-PHYS-PUB-2022-027** (2022).
- ₂₇₄₄
- ₂₇₄₅ [4] 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)
₂₇₄₆ pp. 61–65.
- ₂₇₄₇
- ₂₇₄₈ [5] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al.,
₂₇₄₉ *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606**
₂₇₅₀ (2022) pp. 479–483.
- ₂₇₅₁
- ₂₇₅₂ [6] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- ₂₇₅₃
- ₂₇₅₄ [7] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995)
pp. 2632–2637 [[hep-ex/9503003](#)].
- ₂₇₅₅
- ₂₇₅₆ [8] 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.
- ₂₇₅₇
- ₂₇₅₈ [9] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122**
₂₇₅₉ (1983) pp. 103–116.
- ₂₇₆₀
- ₂₇₆₁ [10] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504**
pp. 218–224 [[hep-ex/0012035](#)].

- 2762 [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2763 *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- 2764 [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.*
2765 **13** (1964) pp. 508–509.
- 2766 [13] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*
2767 *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 2768 [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
2769 *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)
2770 p. 1 [[1207.7214](#)].
- 2771 [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*
2772 *experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 2773 [16] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)
2774 p. 083C01.
- 2775 [17] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*
2776 *invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 2777 [18] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)
2778 pp. 579–588.
- 2779 [19] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 2780 [20] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*
2781 *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**
2782 (1968) pp. 367–377.
- 2783 [21] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
2784 *Phys. Rev.* **104** (1956) pp. 254–258.
- 2785 [22] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hopper and R. P. Hudson, *Experimental*
2786 *test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 2787 [23] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*
2788 *conservation of parity and charge conjugation in meson decays: the magnetic moment*
2789 *of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 2790 [24] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
2791 *Phys. Rev.* **117** (1960) pp. 648–663.

- 2792 [25] J. Goldstone, *Field theories with « superconductor » solutions*, *Il Nuovo Cimento*
2793 (1955-1965) **19** (1961) pp. 154–164.
- 2794 [26] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*
2795 *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](https://arxiv.org/abs/1610.07922).
- 2796 [27] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2797 *ATLAS detector*, [ATLAS-CONF-2018-036](https://cds.cern.ch/record/2620036) (2018).
- 2798 [28] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys.*
2799 *Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](https://arxiv.org/abs/1808.08242)].
- 2800 [29] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2801 [30] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of*
2802 *Instrumentation* **3** (2008) pp. S08002–S08002.
- 2803 [31] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008)
2804 p. S08004.
- 2805 [32] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of*
2806 *Instrumentation* **3** (2008) pp. S08005–S08005.
- 2807 [33] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
2808 *JINST* **3** (2008) p. S08003.
- 2809 [34] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data*
2810 *taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2811 [35] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*,
2812 (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2813 [36] CERN, “The accelerator complex.” (2012).
- 2814 [37] G. C. Strong, *On the impact of selected modern deep-learning techniques to the*
2815 *performance and celerity of classification models in an experimental high-energy*
2816 *physics use case*, [2002.01427](https://arxiv.org/abs/2002.01427).
- 2817 [38] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022).
2818 <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoverview/>.
- 2819 [39] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*,

- 2822 *JINST* **11** (2016) p. P04008 [[1512.01094](#)].
- 2823 [40] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ using the ATLAS detector at the LHC*, **ATLAS-CONF-2019-021** (2019).
- 2825 [41] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*, .
- 2827 [42] for the ATLAS Collaboration collaboration, *Status of the ATLAS Experiment*, Tech. Rep., Geneva (2010). 10.3204/DESY-PROC-2010-04/1.
- 2829 [43] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 2831 [44] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 2833 [45] ATLAS Collaboration, *The inner detector*, (2022).
<https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2835 [46] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 2837 [47] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) p. T05008 [[1803.00844](#)].
- 2839 [48] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, *JINST* **9** (2014) p. P08009 [[1404.7473](#)].
- 2841 [49] ATLAS Collaboration, *Calorimeter*, (2022).
<https://atlas.cern/Discover/Detector/Calorimeter>.
- 2843 [50] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, .
- 2844 [51] G. Artoni, *Muon momentum scale and resolution in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ in ATLAS*, Tech. Rep., Geneva (2016). 10.1016/j.nuclphysbps.2015.09.453.
- 2846 [52] ATLAS Collaboration, *Muon spectrometer*, (2022).
<https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2848 [53] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) p. 317 [[1611.09661](#)].
- 2850 [54] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, **ATL-SOFT-PUB-2021-001** (2021).

- 2852 [55] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts,*
2853 *Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2854 [56] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense*
2855 *Environments*, **ATL-PHYS-PUB-2015-006** (2015).
- 2856 [57] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in*
2857 *dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].
- 2858 [58] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon*
2859 *pixel detector*, *JINST* **9** (2014) p. P09009 [[1406.7690](#)].
- 2860 [59] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*
2861 *Experiment*, .
- 2862 [60] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment*
2863 *in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332
2864 [[1611.10235](#)].
- 2865 [61] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for*
2866 *LHC Run 3*, **ATL-PHYS-PUB-2019-015** (2019).
- 2867 [62] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency*
2868 *measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79**
2869 (2019) p. 970 [[1907.05120](#)].
- 2870 [63] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP*
2871 **04** (2008) p. 063 [[0802.1189](#)].
- 2872 [64] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur. Phys. J. C* **72** (2012)
2873 p. 1896 [[1111.6097](#)].
- 2874 [65] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04**
2875 (2008) p. 005 [[0802.1188](#)].
- 2876 [66] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*
2877 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 2878 [67] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
2879 *the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 2880 [68] ATLAS Collaboration, *Jet energy scale measurements and their systematic*
2881 *uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
2882 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].

- 2883 [69] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*
2884 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)
2885 p. 242001 [[0802.2470](#)].
- 2886 [70] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*
2887 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
2888 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2889 p. 136204 [[2008.02508](#)].
- 2890 [71] ATLAS collaboration, *Performance of jet substructure techniques for large-R jets in*
2891 *proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
2892 p. 076 [[1306.4945](#)].
- 2893 [72] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*
2894 *13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019)
2895 p. 135 [[1807.09477](#)].
- 2896 [73] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) p. 059
2897 [[0903.0392](#)].
- 2898 [74] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
2899 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010
2900 (2017).
- 2901 [75] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*
2902 *Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).
- 2903 [76] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
2904 *using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024 (2016).
- 2905 [77] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*
2906 *proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) p. 292
2907 [[1603.05598](#)].
- 2908 [78] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
2909 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys.*
2910 *J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 2911 [79] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
2912 *b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).
- 2913 [80] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*
2914 *ATLAS detector*, ATL-PHYS-PUB-2017-011 (2017).

- 2915 [81] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
2916 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003 (2017).
- 2917 [82] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*
2918 *Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- 2919 [83] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*
2920 *Run-2*, ATL-PHYS-PUB-2015-022 (2015).
- 2921 [84] B. R. Webber, *Fragmentation and hadronization*, Int. J. Mod. Phys. A **15S1** (2000)
2922 pp. 577–606 [[hep-ph/9912292](#)].
- 2923 [85] Particle Data Group collaboration, *Review of particle physics*, Phys. Rev. D **98**
2924 (2018) p. 030001.
- 2925 [86] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom*
2926 *and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets*,
2927 ATL-PHYS-PUB-2014-008 (2014).
- 2928 [87] Nazar Bartosik, *Diagram showing the common principle of identification of jets*
2929 *initiated by b-hadron decays*, (2022).
2930 https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png.
- 2931 [88] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).
2932 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2022-03/>.
- 2933 [89] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for*
2934 *New Physics in the Fully Hadronic Final State*, Ph.D. thesis, Geneva U., 2021.
2935 10.13097/archive-ouverte/unige:158558.
- 2936 [90] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*
2937 *Process Interface for Event Generators*, ArXiv High Energy Physics - Phenomenology
2938 *e-prints* (2001) .
- 2939 [91] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*
2940 *standard format for Les Houches Event Files*, Computer Physics Communications
2941 **176** (2007) pp. 300–304.
- 2942 [92] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*
2943 *universal approximators*, Neural Networks **2** (1989) pp. 359–366.
- 2944 [93] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*
2945 *back-propagating errors*, nature **323** (1986) pp. 533–536.

- 2946 [94] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*
2947 *activity, The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 2948 [95] J. Hopfield, *Neural networks and physical systems with emergent collective*
2949 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*
2950 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 2951 [96] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*
2952 *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 2953 [97] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*
2954 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2955 [98] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,
2956 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,
2957 *arXiv e-prints* (2018) p. arXiv:1806.01261 [[1806.01261](#)].
- 2958 [99] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
2959 *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 2960 [100] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,
2961 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 2962 [101] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*
2963 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007)
2964 p. 126–126.
- 2965 [102] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*
2966 *shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007)
2967 p. 070–070.
- 2968 [103] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*
2969 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*
2970 *Physics* **2010** (2010) .
- 2971 [104] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)
2972 p. 040 [[1410.8849](#)].
- 2973 [105] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
2974 *ATL-PHYS-PUB-2016-020* (2016).
- 2975 [106] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*
2976 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159

- 2977 [1410.3012].
- 2978 [107] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
2979 [ATL-PHYS-PUB-2014-021](#) (2014).
- 2980 [108] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)
2981 p. 244 [[1207.1303](#)].
- 2982 [109] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*
2983 **462** (2001) p. 152.
- 2984 [110] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**
2985 (2010) p. 823 [[1005.4568](#)].
- 2986 [111] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*
2987 *Instrum. Meth. A* **506** (2003) p. 250.
- 2988 [112] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*
2989 *detector*, [ATLAS-CONF-2014-018](#) (2014).
- 2990 [113] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2991 *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 2992 [114] ATLAS Collaboration, *Observation of Higgs boson production in association with a*
2993 *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173
2994 [[1806.00425](#)].
- 2995 [115] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*
2996 *using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **03**
2997 (2020) p. 145 [[1910.08447](#)].
- 2998 [116] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
2999 *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772
3000 [[2002.08772](#)].
- 3001 [117] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13\text{ TeV}$ data*,
3002 [ATL-PHYS-PUB-2015-037](#) (2015).
- 3003 [118] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*
3004 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
3005 *$\sqrt{s} = 13\text{ TeV}$* , *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 3006 [119] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
3007 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, *arXiv e-prints* (2020)

- 3008 p. arXiv:2007.08294 [[2007.08294](#)].
- 3009 [120] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,
3010 *arXiv e-prints* (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 3011 [121] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al.,
3012 *Attention Is All You Need*, *arXiv e-prints* (2017) p. arXiv:1706.03762 [[1706.03762](#)].
- 3013 [122] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv
e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 3015 [123] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
3016 <https://github.com/onnx/onnx>.
- 3017 [124] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp
3018 collision dataset*, [2211.16345](#).
- 3019 [125] ATLAS Collaboration, *Performance of Run 3 HLT b-tagging with fast tracking,
3020 Public b-Jet Trigger Plots for Collision Data*, (2022).
3021 [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/
3022 BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi).
- 3023 [126] ATLAS Collaboration, *Neural Network Jet Flavour Tagging with the Upgraded
3024 ATLAS Inner Tracker Detector at the High-Luminosity LHC,
3025 ATL-PHYS-PUB-2022-047* (2022).
- 3026 [127] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs
3027 Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- 3028 [128] ATLAS Collaboration, *Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$
3029 decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C*
3030 **81** (2021) p. 178 [[2007.02873](#)].
- 3031 [129] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson
3032 decaying into b-quarks with a vector boson at high transverse momentum in pp
3033 collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
3034 p. 136204 [[2008.02508](#)].
- 3035 [130] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated
3036 Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039
3037 [[1512.02572](#)].
- 3038 [131] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in*

- 3039 *proton–proton collisions at $\sqrt{s} = 7 \text{ TeV}$ using the ATLAS detector,* *JHEP* **09** (2013)
3040 p. 076 [[1306.4945](#)].
- 3041 [132] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO*
3042 *calculations in shower Monte Carlo programs: the POWHEG BOX,* *JHEP* **06** (2010)
3043 p. 043 [[1002.2581](#)].
- 3044 [133] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum*
3045 *distribution in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ with the ATLAS detector,* *JHEP* **09**
3046 (2014) p. 145 [[1406.3660](#)].
- 3047 [134] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al.,
3048 *Automated one-loop calculations with GoSam,* *Eur. Phys. J. C* **72** (2012) p. 1889
3049 [[1111.2034](#)].
- 3050 [135] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO,*
3051 *JHEP* **10** (2012) p. 155 [[1206.3572](#)].
- 3052 [136] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at*
3053 *NLO with the POWHEG BOX interfaced to GoSam and their merging within*
3054 *MinLO,* *JHEP* **10** (2013) p. 083 [[1306.2542](#)].
- 3055 [137] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to*
3056 *associated WH and ZH production at hadron colliders,* *Phys. Rev. D* **68** (2003)
3057 p. 073003 [[hep-ph/0306234](#)].
- 3058 [138] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the*
3059 *Higgs-strahlung processes at hadron colliders,* *Phys. Lett. B* **579** (2004) pp. 149–156
3060 [[hep-ph/0307206](#)].
- 3061 [139] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production*
3062 *at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO,* *Phys. Rev. Lett.*
3063 **107** (2011) p. 152003 [[1107.1164](#)].
- 3064 [140] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects*
3065 *in hadronic Higgs-Strahlung,* *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- 3066 [141] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated*
3067 *WH production and decay at the LHC,* *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- 3068 [142] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron*
3069 *colliders: The fully differential NNLO QCD calculation,* *Phys. Lett. B* **740** (2015)
3070 pp. 51–55 [[1407.4747](#)].

- 3071 [143] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson*
3072 *at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].
- 3073 [144] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
3074 *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078
3075 [[1211.5015](#)].
- 3076 [145] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via*
3077 *gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- 3078 [146] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for*
3079 *gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- 3080 [147] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron*
3081 *Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- 3082 [148] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron*
3083 *colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- 3084 [149] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte*
3085 *Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].
- 3086 [150] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair*
3087 *cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) p. 2930
3088 [[1112.5675](#)].
- 3089 [151] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with*
3090 *shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) p. 111
3091 [[0907.4076](#)].
- 3092 [152] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top*
3093 *quark production*, *Phys. Rev. D* **81** (2010) p. 054028 [[1001.5034](#)].
- 3094 [153] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for*
3095 *t-channel single top quark production*, *Phys. Rev. D* **83** (2011) p. 091503 [[1103.2792](#)].
- 3096 [154] E. Re, *Single-top Wt-channel production matched with parton showers using the*
3097 *POWHEG method*, *Eur. Phys. J. C* **71** (2011) p. 1547 [[1009.2450](#)].
- 3098 [155] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated*
3099 *production with a W^- or H^-* , *Phys. Rev. D* **82** (2010) p. 054018 [[1005.4451](#)].
- 3100 [156] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al.,
3101 *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) p. 007 [[0811.4622](#)].

- 3102 [157] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) p. 034
3103 [[1905.09127](#)].
- 3104 [158] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
3105 *Phys. Rev. Lett.* **108** (2012) p. 111601 [[1111.5206](#)].
- 3106 [159] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008)
3107 p. 039 [[0808.3674](#)].
- 3108 [160] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour*
3109 *dipole factorisation*, *JHEP* **03** (2008) p. 038 [[0709.1027](#)].
- 3110 [161] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton*
3111 *showers. The NLO case*, *JHEP* **04** (2013) p. 027 [[1207.5030](#)].
- 3112 [162] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson*
3113 *Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.*
3114 *Rev. Lett.* **103** (2009) p. 082001 [[0903.2120](#)].
- 3115 [163] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**
3116 (2016) p. 023001 [[1510.03865](#)].
- 3117 [164] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to*
3118 *Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* **03**
3119 (2012) p. 075 [[1112.5142](#)].
- 3120 [165] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo*
3121 *program for Higgs production in vector-boson fusion and Higgs strahlung at hadron*
3122 *colliders*, *Comput. Phys. Commun.* **195** (2015) pp. 161–171 [[1412.5390](#)].
- 3123 [166] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS*
3124 *detector using t̄t events at √s = 13 TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 3125 [167] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using*
3126 *ATLAS proton–proton collision data at √s = 13 TeV*, *ATLAS-CONF-2018-006*
3127 (2018).
- 3128 [168] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in t̄t events using*
3129 *a likelihood approach with the ATLAS detector*, *ATLAS-CONF-2018-001* (2018).
- 3130 [169] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*
3131 *ATLAS Detector*, *ATL-PHYS-PUB-2014-013* (2014).
- 3132 [170] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark*

- 3133 pairs with the ATLAS detector at 13 TeV, *Eur. Phys. J. C* **79** (2019) p. 836
3134 [[1906.11005](#)].
- 3135 [171] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a*
3136 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 3137 [172] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*
3138 *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) p. 204
3139 [[1412.3989](#)].
- 3140 [173] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*
3141 *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*
3142 **D93** (2016) p. 033006 [[1506.07443](#)].
- 3143 [174] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at*
3144 $\sqrt{s} = 13 \text{ TeV}$ *with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016)
3145 p. 182002 [[1606.02625](#)].
- 3146 [175] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in*
3147 *association with a vector boson and decaying to a pair of b-quarks using large-R jets*,
3148 ATL-COM-PHYS-2019-1125 (2019).
- 3149 [176] J. Pumplin et al., *New generation of parton distributions with uncertainties from*
3150 *global qcd analysis*, *JHEP* **07** (2002) p. 012 [[0201195](#)].
- 3151 [177] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top*
3152 *hadroproduction in association with a W boson*, *JHEP* **07** (2008) p. 029 [[0805.3067](#)].
- 3153 [178] ATLAS Collaboration, *Measurement of VH, $H \rightarrow b\bar{b}$ production as a function of the*
3154 *vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*,
3155 *JHEP* **05** (2019) p. 141 [[1903.04618](#)].
- 3156 [179] ATLAS Collaboration, *Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector*,
3157 *JHEP* **12** (2017) p. 024 [[1708.03299](#)].
- 3158 [180] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino
3159 and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential*
3160 *Distributions*, *CERN-2012-002* (2012) [[1201.3084](#)].
- 3161 [181] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3.*
3162 *Higgs Properties*, *CERN-2013-004* (2013) [[1307.1347](#)].
- 3163 [182] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*

³¹⁶⁴ *roostats project, arXiv preprint arXiv:1009.1003* (2010) .

₃₁₆₅ List of figures

₃₁₆₆ 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.	20
₃₁₇₁ 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.	26
₃₁₇₃ 2.3	Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. 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.	26
₃₁₇₈ 2.4	Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. 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.	27
₃₁₈₂ 3.1	An overview of the CERN accelerator complex [36]. 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.	30
₃₁₈₆ 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. [37].	31

3190	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. [38].	33
3196	3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [35]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.	35
3200	3.5 Average pile-up profiles measured by ATLAS during Run 2 [35]. Higher levels of pile-up are planned for Run 3.	35
3202	3.6 A 3D model of the entire ATLAS detector [42]. 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
3206	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 [45].	37
3208	3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [38].	38
3210	3.9 A schematic cross-sectional view of the ATLAS IBL [46].	39
3211	3.10 The ATLAS calorimeters [49]. 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
3214	3.11 The ATLAS muon spectrometer [52].	43
3215	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 [57].	47
3219	3.13 A sketch of electron reconstruction using the ATLAS detector [76]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.	51

3222	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.	55
3223		
3224		
3225	4.2 Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [87]. 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.	56
3226		
3227		
3228		
3229		
3230	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.	57
3231		
3232		
3233		
3234		
3235		
3236	4.4 b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [88]. 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.	58
3237		
3238		
3239		
3240		
3241		
3242		
3243	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.	59
3244		
3245		
3246		
3247		
3248		
3249		
3250		
3251		
3252		
3253	4.6 The fraction of hits which are shared on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudotracks represent the ideal performance given the ATLAS detector, see Section 3.4.1.	61
3254		
3255		
3256		

3257	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 corre- sponding p_T bin is filled with the number of good or wrong hits and this value is averaged to show the overall rate.	63
3262	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.	64
3265	5.1	A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression model.	67
3269	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.	68
3272	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 [99]. Reproduced from Ref. [98].	71
3275	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.	73
3278	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.	73

3284	5.6 (left) Normalised histogram of the model output separated for signal and 3285 fake tracks, and further separated by those tracks which are from the 3286 decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and 3287 tracks from the decay of a b -hadron (dashed line). The model is able to 3288 successfully discriminate between signal and fake tracks, and shows a very 3289 similar performance when looking specifically at tracks from the decay of a 3290 b -hadron.	77
3291	5.7 (left) Normalised histogram of the model output separated for tracks from 3292 the decay of a b -hadron and tracks from other sources. The two groups are 3293 further separated into those tracks which are fake. (right) The ROC curve 3294 for all tracks (solid line).	79
3295	5.8 The effect of applying the fake track identification algorithm together with 3296 the b -hadron decay track identification on the jet tagging performance of 3297 SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and 3298 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) 3299 is compared to two working points of fake track removal, labelled “A” (red) 3300 and “B” (orange), which represent two different 2D working points of the 3301 track classification tools. Removal of fake tracks based on truth information 3302 is shown by the green curves.	80
3303	6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right) [3]. 3304 The existing approach utilises low-level algorithms (shown in blue), the 3305 outputs of which are fed into a high-level algorithm (DL1r). Instead 3306 of being used to guide the design of the manually optimised algorithms, 3307 additional truth information from the simulation is now being used as 3308 auxiliary training targets for GN1. The solid lines represent reconstructed 3309 information, whereas the dashed lines represent truth information.	83
3310	6.2 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of n_{tracks} , 3311 where each track is described by 21 track features ($n_{\text{tf}} = 21$) [3]. The jet 3312 features are copied for each of the tracks, and the combined jet-track vectors 3313 of length 23 form the inputs of GN1.	90

3314	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.	
3315	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
3316		
3317		
3318		
3319		
3320	6.4 Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample [3]. 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
3321		
3322		
3323		
3324		
3325		
3326	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 [3]. 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
3327		
3328		
3329		
3330		
3331		
3332		
3333	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 [3]. 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
3334		
3335		
3336		
3337		
3338		
3339		
3340	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 [3]. 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
3341		
3342		
3343		
3344		
3345		

3346	6.8 The b -jet (left) and light-jet (right) rejections as a function of the c -jet 3347 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV [3]. The ratio to the 3348 performance of the DL1r algorithm is shown in the bottom panels. Binomial 3349 error bands are denoted by the shaded regions. The x -axis range is chosen 3350 to display the c -jet tagging efficiencies usually probed in these regions of 3351 phase space.	101
3352	6.9 The b -jet (left) and light-jet (right) rejections as a function of the c -jet 3353 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV [3]. The ratio to the 3354 performance of the DL1r algorithm is shown in the bottom panels. 3355 Binomial error bands are denoted by the shaded regions. The x -axis range 3356 is chosen to display the c -jet tagging efficiencies usually probed in these 3357 regions of phase space.	101
3358	6.10 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 3359 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal 3360 GN1, in addition to configurations where no (GN1 No Aux), only the 3361 track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary 3362 objectives are deployed [3]. The ratio to the performance of the DL1r 3363 algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in 3364 the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1.	103
3365	6.11 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 3366 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal 3367 GN1, in addition to configurations where no (GN1 No Aux), only the 3368 track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary 3369 objectives are deployed [3]. The ratio to the performance of the DL1r 3370 algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in 3371 the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1.	103
3372	6.12 A schematic showing the truth track origin and vertex information (left) 3373 compared with the predictions from GN1 (right). The diagrams show the 3374 truth (left) and predicted (right) structure of a b -jet. The shaded black boxes 3375 show the groupings of tracks into different vertices. Vertices reconstructed 3376 by SV1 and JetFitter are also marked. Class probabilities p_b , p_c and p_u are 3377 rounded to three significant figures. GN1 improves over SV1 and JetFitter 3378 by successfully determining the origins and vertex compatibility of all the 3379 tracks in the jet with the exception of the truth OtherSecondary track, 3380 which is misidentified as pile-up.	105

3381	6.13 A schematic showing the truth track origin and vertex information (left)	
3382	compared with the predictions from GN1 (right). The diagrams show the	
3383	truth (left) and predicted (right) structure of a b -jet. The shaded black boxes	
3384	show the groupings of tracks into different vertices. Vertices reconstructed	
3385	by SV1 and JetFitter are also marked. Class probabilities p_b , p_c and p_u are	
3386	rounded to three significant figures. GN1 improves over SV1 and JetFitter	
3387	by successfully determining the origins and vertex compatibility of all but	
3388	one tracks in the jet.	105
3389	6.14 Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in	
3390	the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding	
3391	approaches. The exclusive vertexing approach includes single track vertices	
3392	while the inclusive approach requires vertices to contain at least two tracks.	
3393	Efficient vertex finding requires the recall of at least 65% of the tracks in	
3394	the truth vertex, and allows no more than 50% of the tracks to be included	
3395	incorrectly. Binomial error bands are denoted by the shaded regions.	108
3396	6.15 Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in	
3397	the Z' sample using the exclusive (left) and inclusive (right) vertex finding	
3398	approaches. The exclusive vertexing approach includes single track vertices	
3399	while the inclusive approach requires vertices to contain at least two tracks.	
3400	Efficient vertex finding requires the recall of at least 65% of the tracks in	
3401	the truth vertex, and allows no more than 50% of the tracks to be included	
3402	incorrectly. Binomial error bands are denoted by the shaded regions.	108
3403	6.16 Vertex finding efficiency as a function of jet p_T for light-jets in the $t\bar{t}$ sample	
3404	using the exclusive (left) and inclusive (right) vertex finding approaches.	
3405	The exclusive vertexing approach includes single track vertices while the	
3406	inclusive approach requires vertices to contain at least two tracks. Efficient	
3407	vertex finding requires the recall of at least 65% of the tracks in the truth	
3408	vertex, and allows no more than 50% of the tracks to be included incorrectly.	
3409	Binomial error bands are denoted by the shaded regions.	109

3410	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
3411		
3412		
3413		
3414		
3415		
3416		
3417	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) [3]. 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
3418		
3419		
3420		
3421		
3422		
3423		
3424	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
3425		
3426		
3427		
3428		
3429		
3430	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
3431		
3432		
3433		
3434		
3435		
3436	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
3437		
3438		
3439		
3440		
3441		

3442	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
3448	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
3453	7.1 Diagrams of the signal process (top) and $t\bar{t}$ background (bottom). Object to the right of centre are reconstructed within the large- R jet. For the $t\bar{t}$ background, the large- R jet contains a mis-tagged c -jet or (less often) a mis-tagged light-jet. The contribution in the 0-lepton channel results from hadronically decaying τ lepton, or a electron or muon which is out of the analysis acceptance.	127
3459	7.2 Leading large- R jet mass for the Z and $W+\text{hf}$ processes in the HP SR of the 0-lepton channel [175]. The renormalisation scale μ_r has been varied by a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to the ratio between the nominal and alternate samples.	138
3463	7.3 Leading large- R jet m_J inclusive in p_T^V for the $V+\text{hf}$ process modelled using both the SHERPA (blue) and MADGRAPH (red) samples [175]. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.	139
3466	7.4 Leading large- R jet mass distribution for the WZ and ZZ processes, inclusive across all signal regions and lepton channels [175]. The renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.	141
3470	7.5 The comparison of the shapes of the large- R jet mass m_J between SHERPA (black) and POWHEG+PYTHIA8 (red) 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 [175]. 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.	142

3476	7.6 The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 3477 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. 3478 The LP and HP regions are merged for the 0-lepton and 1-lepton channels. 3479 The fitted background contributions are shown as filled histograms. The 3480 Higgs boson signal ($m_H = 125 \text{ GeV}$) is shown as a filled histogram and 3481 is normalised to the signal yield extracted from data ($\mu_{VH}^{bb} = 0.72$), and 3482 as an unstacked unfilled histogram, scaled by the SM prediction times a 3483 factor of two. The size of the combined on the sum of the fitted signal and 3484 background is shown in the hatched band. The highest bin contains the 3485 overflow [129].	150
3486	7.7 The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton 3487 channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bot- 3488 tom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. 3489 The background contributions after the likelihood fit are shown as filled 3490 histograms. The Higgs boson signal ($m_h = 125 \text{ GeV}$) is shown as a filled 3491 histogram on top of the fitted backgrounds normalised to the signal yield 3492 extracted from data ($\mu_{VH}^{bb} = 0.72$), and unstacked as an unfilled histogram, 3493 scaled by the SM prediction times a factor of 2. The size of the combined 3494 statistical and systematic uncertainty for the sum of the fitted signal and 3495 background is indicated by the hatched band. The highest bin in the 3496 distributions contains the overflow [129].	151
3497	7.8 m_J distribution in data after subtraction of all backgrounds except for the 3498 WZ and ZZ diboson processes. The contributions from all lepton channels 3499 and signal regions are summed and weighted by their respective values of 3500 the ratio of fitted Higgs boson signal and background yields. The expected 3501 contribution of the associated WH and ZH production of a SM Higgs 3502 boson with $m_H = 125 \text{ GeV}$ is shown scaled by the measured combined 3503 signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to 3504 its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and 3505 systematic uncertainty is indicated by the hatched band. This error band is 3506 computed from a full signal-plus-background fit including all the systematic 3507 uncertainties defined in Section 7.2, except for the VH/VZ experimental 3508 and theory uncertainties [129].	152

3509	7.9 Signal strength compatibility test between the (3+1)-POI fit (with the three	
3510	lepton channels splitted) and the default (1+1)-POI fit. The compatibility	
3511	of the three channels is evaluated via a χ^2 difference test and results in a	
3512	p-value of 49% [129].	153
3513	7.10 Impact of systematic uncertainties on the fitted VH signal-strength pa-	
3514	rameter μ_{VH}^{bb} sorted in decreasing order. The boxes show the variations	
3515	of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual	
3516	nuisance parameter to its post-fit value modified upwards or downwards by	
3517	its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of	
3518	up- and down-variations can be distinguished via the dashed and plane box	
3519	fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying	
3520	each nuisance parameter by ± 1 . The filled circles show the deviation of the	
3521	fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value	
3522	θ_0 expressed in standard deviations with respect to their nominal uncer-	
3523	tainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties	
3524	on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show	
3525	the fitted values and uncertainties of the normalization parameters that are	
3526	freely floating in the fit. Pre-fit, these parameters have a value of one [175].	155

3527 List of tables

<small>3528</small>	2.1 The fermions of the SM [16]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.	<small>3529</small>	<small>3530</small>	15				
<small>3531</small>	2.2 The bosons of the SM [16]. The photon, weak bosons and gluons are gauge bosons arising from gauge symmetries, and carry the four fundamental forces of the SM. The recently discovered Higgs boson is the only fundamental scalar particle in the SM.	<small>3532</small>	<small>3533</small>	<small>3534</small>	16			
<small>3535</small>	3.1 Overview of the different LHC runs [34, 35]. The average number of interac- tions per bunch-crossing is denoted as $\langle \mu \rangle$ (see Section 3.2.3), and is here averaged over the entire run. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.	<small>3536</small>	<small>3537</small>	<small>3538</small>	29			
<small>3539</small>	5.1 Truth origins which are used to categorise the physics process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [57]. A truth-matching probability of less than 0.5 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions, K_S^0 and Λ^0 decays, and hadronic interactions.	<small>3540</small>	<small>3541</small>	<small>3542</small>	<small>3543</small>	<small>3544</small>	<small>3545</small>	72
<small>3546</small>	5.2 Input features to the fake track classification NN. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [57], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.	<small>3547</small>	<small>3548</small>	<small>3549</small>	<small>3550</small>	<small>3551</small>	75	

3552	5.3 Quality selections applied to tracks, where d_0 is the transverse IP of the track, z_0 is the longitudinal IP with respect to the PV and θ is the track polar angle (see Section 3.2.2 for the IP definitions). Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [57]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.	76
3558	5.4 Hyperparameter for the track classification model	77
3559	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
3563	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
3567	6.1 Input features to the GN1 model [3]. 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 [57], 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
3575	6.2 A summary of GN1’s different classification networks used for the various training objectives, adapted from Ref. [3]. The hidden layers column contains a list specifying the number of neurons in each layer.	92
3578	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

3585	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.	121
3592	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 (α_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 $qq \rightarrow ZH$ process is normalised using the cross-section for the $pp \rightarrow ZH$ process, after subtracting the $gg \rightarrow ZH$ contribution. An additional scale factor is applied to the $qq \rightarrow VH$ processes as a function of the transverse momentum of the vector boson, to account for electroweak (EW) corrections at NLO. This makes use of the VH differential cross-section computed with HAWK [164, 165].	123
3604	7.3 Selections applied to baseline and signal electrons and muons.	124
3605	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.	126
3615	7.5 Different sources of uncertainty (i.e. variations in the model) considered for the $V+\text{jets}$ background, and the corresponding implementation. For each uncertainty, acceptance and shape uncertainties are derived.	130

3618	7.6 0-lepton W +jets nominal sample flavour composition and total event yield [175]. 3619 Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see 3620 Table 7.1).	134
3621	7.7 1-lepton W +jets nominal sample flavour composition and total event yield [175]. 3622 Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see 3623 Table 7.1).	134
3624	7.8 0-lepton Z +jets nominal sample flavour composition and total event yield [175]. 3625 Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see 3626 Table 7.1).	134
3627	7.9 2-lepton Z +jets nominal sample flavour composition and total event yield [175]. 3628 Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see 3629 Table 7.1).	135
3630	7.10 V +jets acceptance uncertainties [175]. W +jets SR and CR uncertainties 3631 marked with a superscript † are correlated. The 1L W +jets H/M uncertainty 3632 marked by * is applied as independent and uncorrelated NPs in both HP 3633 and LP signal regions.	136
3634	7.11 Diboson acceptance uncertainties [175]. All uncertainties except channel ex- 3635 trapolation uncertainties are fully correlated between ZZ and WZ processes 3636 and channels.	141
3637	7.12 Factors applied to the nominal normalisations of the $t\bar{t}$, W +hf, and Z +hf 3638 backgrounds, as obtained from the likelihood fit [129]. The errors represent 3639 the combined statistical and systematic uncertainties.	149
3640	7.13 Breakdown of the absolute contributions to the uncertainty on the signal 3641 strength μ_{VH}^{bb} obtained from the (1+1)-POI fit. The average impact repre- 3642 sents the average between the positive and negative uncertainties on μ_{VH}^{bb} . 3643 The sum in quadrature of the systematic uncertainties attached to the cate- 3644 gories differs from the total systematic uncertainty due to correlations [175]. 3645	156