

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**

March 18, 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	48
3.4.3	Jets	49
3.4.4	Leptons	51

3.4.5	Missing Transverse Momentum	53
4	Tracking and flavour tagging	54
4.1	<i>b</i> -hadron Reconstruction	54
4.1.1	Decay Topology	54
4.2	<i>b</i> -tagging Algorithms	57
4.3	Challenges Facing <i>b</i> -hadron Reconstruction	58
4.4	Investigations into High p_T <i>b</i> -hadron Tracking	61
4.4.1	Shared Hits	61
4.4.2	Global χ^2 Fitter Outlier Removal	62
4.5	Conclusion	65
5	Track Classification MVA	67
5.1	Machine Learning Background	67
5.1.1	Neural Networks	69
5.1.2	Training with Gradient Descent	71
5.2	Graph Neural Network Theory	71
5.3	Track Truth Origin Labelling	73
5.4	Fake Track Identification Tool	74
5.4.1	Datasets	76
5.4.2	Model Inputs	77
5.4.3	Model Hyperparameters	78
5.4.4	Results	78
5.5	<i>b</i> -hadron Track Identification	80
5.6	Combined Approach	81
5.7	Conclusion	82
6	Graph Neural Network Flavour Tagger	84
6.1	Motivation	84
6.2	Experiemental Setup	87
6.2.1	Datasets	87
6.3	Model Architecture	87
6.3.1	Model Inputs	87
6.3.2	Auxiliary Training Objectives	88
6.3.3	Architecture	91
6.3.4	Training	95

6.4	Results	96
6.4.1	b -tagging Performance	97
6.4.2	c -tagging Performance	99
6.4.3	Ablations	102
6.4.4	Inclusion of Low-Level Vertexing Algorithms	104
6.4.5	Jet Display Diagrams	104
6.4.6	Vertexing Performance	106
6.4.7	Track Classification Performance	112
6.4.8	Looser Track Selection	114
6.5	Other Implementations of GN1	115
6.6	Conclusion	117
7	Boosted VH$b\bar{b}$ Analysis	121
7.1	Analysis Overview	122
7.1.1	Data & Simulated Samples	124
7.1.2	Object Reconstruction	124
7.1.3	Selection Criteria	126
7.1.4	Control Regions	127
7.1.5	Background Composition	127
7.2	Systematic Uncertainties & Background Modelling	130
7.2.1	Implementation of Variations	131
7.2.2	Sources of Systematic Modelling Uncertainties	132
7.2.3	Sources of Experimental Uncertainties	134
7.2.4	Vector Boson + Jets Modelling	135
7.2.5	Diboson Modelling	139
7.2.6	$t\bar{t}$ and single-top Modelling	144
7.2.7	Signal Modelling	145
7.3	Statistical Treatment	146
7.3.1	Likelihood Function	146
7.3.2	Background Normalisations	149
7.3.3	Asimov Dataset & Expected Results	150
7.4	Results	150
7.4.1	Post-fit Distributions	151
7.4.2	Post-fit Yields	151
7.4.3	Observed Signal Strength & Significance	151

7.4.4	Impact of Systematics	157
7.4.5	STXS Interpretation	161
7.5	Conclusion	161
8	Conclusion	163
8.1	Summary	163
8.2	Future Work	164
Bibliography		166
List of figures		181
List of tables		193

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- $\frac{1}{2}$ 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	$\frac{2}{3}$
	ν_e	$< 1.1 \times 10^{-6}$	0	d	4.67	$-\frac{1}{3}$
Second	μ	105.7	-1	c	1.27×10^3	$\frac{2}{3}$
	ν_μ	< 0.19	0	s	93.4	$-\frac{1}{3}$
Third	τ	1776.9	-1	t	173×10^3	$\frac{2}{3}$
	ν_τ	< 18.2	0	b	4.18×10^3	$-\frac{1}{3}$

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 interactions 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 a fermionic spin- $\frac{1}{2}$ 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 the ψ
105 and A_μ fields to Eq. (2.2). For completeness, the kinetic term for the the new field
106 A_μ is also added in terms of $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, which is trivially invariant under
107 the 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 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 interac-
117 tions. 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 scales
¹³⁶ (and corresponding larger distance scales) the interaction is strong. This leads to
¹³⁷ quark confinement, whereby an attempt to isolate individual colour-charged quarks
¹³⁸ requires so much energy that additional quark-antiquark pairs 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 the “Higgs mechanism”) is the
165 process through which the fundamental particles of the SM acquire mass [11–13].
166 Experimentally it was known that the weak force had a low effective strength, which
167 was suggestive of a massive mediating gauge particle. However, directly adding
168 mass to the weak gauge bosons violates the non-Abelian symmetry of the SM.
169 Instead, the gauge bosons gain mass through the interaction with a scalar Higgs
170 field which results from the spontaneous breakdown of symmetry as discussed in
171 Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as discussed
172 in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the Higgs
173 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.

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

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

¹⁸⁵ where the covariant derivative is given by

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

¹⁸⁶ and T^a are the generators of $SU(2)$. The potential term $V(\phi)$ is made up of a
¹⁸⁷ quadratic and quartic term in the scalar field ϕ , which each contain an arbitrary
¹⁸⁸ parameter, respectively λ and μ . The quartic term gives the field self-interaction, and
¹⁸⁹ cannot be negative as this would lead to a potential that was unbounded from below.
¹⁹⁰ The quadratic term can be positive or negative. In the case where the quadratic term
¹⁹¹ is positive, it is interpreted as a mass term for the scalar field. By choosing $\mu^2 < 0$
¹⁹² the field becomes unphysical due to its negative mass. The shape of the potential in
¹⁹³ this case is shown in Fig. 2.1. Note that in the case of the Standard Model, the scalar
¹⁹⁴ field ϕ is a complex doublet, and so the corresponding potential is 5-dimensional.
¹⁹⁵ In order to obtain a physical interpretation of the Lagrangain in Eq. (2.14) for the
¹⁹⁶ case where $\mu^2 < 0$, the field ϕ is expanded around the vacuum state. The vacuum
¹⁹⁷ expectation value (VEV) is the expected value of the field ϕ which minimises the
¹⁹⁸ potential $V(\phi)$ (equivalently the expected value of the field operator ϕ when the
¹⁹⁹ 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

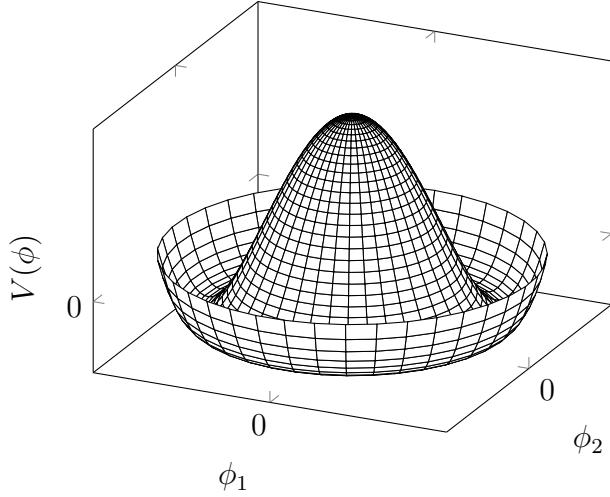


Figure 2.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin.

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

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

where θ_W is the weak mixing angle defined by

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

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

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

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

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

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

236 Spontaneous symmetry breaking has modified the original symmetry group of the SM
237 $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$. Three broken generators from the symmetry
238 group $SU(2)_L \times U(1)_Y$ have been absorbed into the definition of the physical weak
239 vector bosons, giving them mass. The same methodology can be used to generate
240 the fermion masses, as shown in the next section.

241 2.2.2 Fermionic Yukawa Coupling

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

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

251 Charged Lepton Masses

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

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

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

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

258 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
259 the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore

260 be identified as the electron mass term. An interaction term between the electron
 261 field e and the true vacuum Higgs field H is also present. Mass is generated for the
 262 other charged lepton generations in the same way.

263 **Quark Masses**

264 The down-type quarks acquire their mass analogous to the leptons, with $\psi_L = (u_L, d_L)$
 265 and $\psi_R = d_R$ for the first quark generation. Mass is generated for the up-type quarks
 266 using the conjugate field to ϕ which transforms under $SU(2)$ as a doublet with
 267 $Y = -1/2$. The conjugate field $\tilde{\phi}$ is constructed as

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

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

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

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

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

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

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

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

282 2.2.3 Higgs Sector Phenomenology

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

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

303 Since the Higgs boson couples proportional to mass, decays to heavier particles are
304 favoured. The branching ratios of different Higgs boson decay modes are shown
305 as a function of m_H in Fig. 2.4. Approximately 58% of the time the Higgs boson
306 decays to a pair of b -quarks, the dominant decay mode. The next most likely decay
307 mode is to a pair of W bosons, which occurs approximately 20% of the time. After
308 the b -quark, the next heaviest fermions are the tau lepton and the c -quark, decays
309 to pairs of these particles happen approximately an order of magnitude less often.
310 Decays to pairs of vector bosons are via a virtual off shell Higgs boson only. While
311 the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$ branching ratios are small compared with fermionic decay

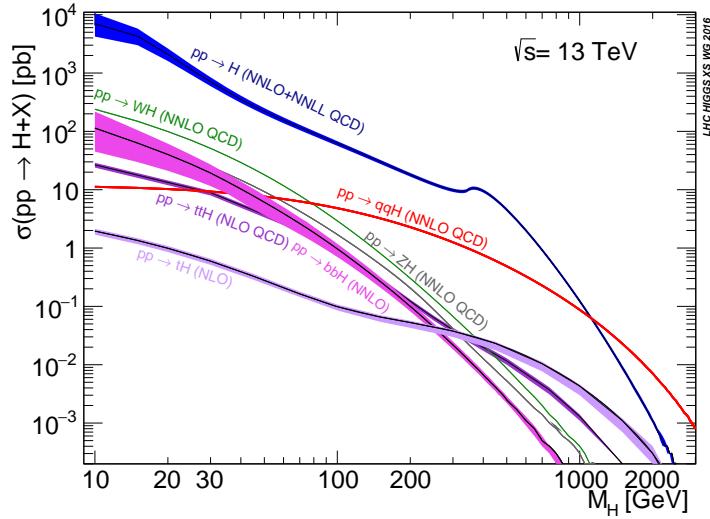


Figure 2.2: Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13 \text{ TeV}$ [26]. Uncertainties are shown in the shaded bands. At $m_H = 125 \text{ GeV}$, Higgs boson production is dominated by gluon-gluon fusion, vector boson fusion, associated production with vector bosons, and top quark fusion.

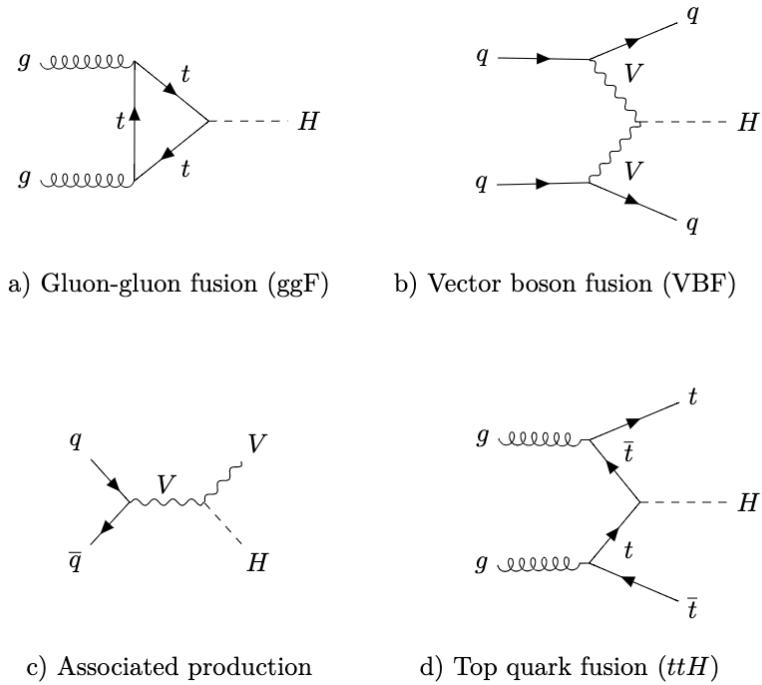


Figure 2.3: Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125 \text{ GeV}$ at a centre of mass energy $\sqrt{s} = 13 \text{ TeV}$.

312 modes (around 0.2% for $H \rightarrow \gamma\gamma$), these decay channels were instrumental in the
313 initial discovery of the Higgs due to the low level of background processes which
314 mimic the final state [14, 15].

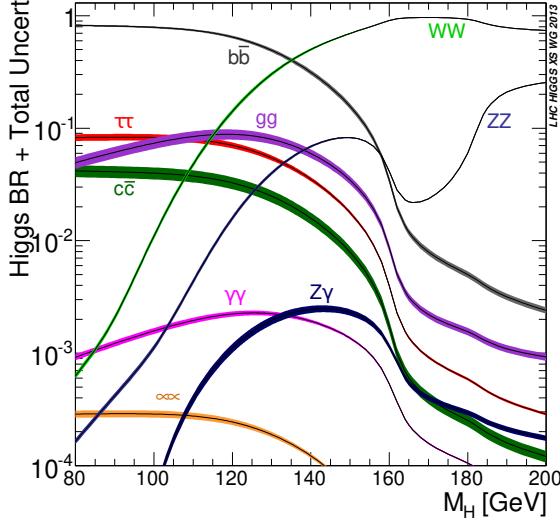


Figure 2.4: Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, the Higgs predominantly decays to a pair of b -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of W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

315 This thesis presents a measurement of the Higgs bosons production rate using events
316 with a Higgs boson produced in association with vector boson and decaying to a pair
317 of b -quarks, i.e. $pp \rightarrow VH(b\bar{b})$. The $H \rightarrow b\bar{b}$ decay mode directly probes the Higgs
318 coupling to fermions, and more specifically to the bottom quark. This coupling was
319 first observed in 2018 [27, 28]. Ongoing work measuring the coupling strengths, in
320 particular in the high energy regime, is the focus of the analysis presented in this
321 thesis in Chapter 7.

322 Chapter 3

323 The Large Hadron Collider and the
324 ATLAS Detector

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

337 In this chapter, a brief overview of the LHC and the accelerator complex at CERN
338 is given in Section 3.1. The coordinate system used at the ATLAS detector and
339 other common definitions are introduced in Section 3.2. An overview of the different
340 detector systems is provided in Section 3.3, and finally descriptions of various
341 commonly used reconstructed objects is given in Section 3.4.

³⁴² 3.1 The Large Hadron Collider

³⁴³ The LHC is operated in multi-year *runs* during which beams of protons are circulated
³⁴⁴ and collided. Between runs there are periods of shutdown while the accelerator and
³⁴⁵ detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC
³⁴⁶ collided proton bunches, each containing more than 10^{11} particles, 20 million times
³⁴⁷ per second, providing 7 TeV proton-proton collisions at instantaneous luminosities
³⁴⁸ of up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The centre-of-mass energy was increased to 8 TeV
³⁴⁹ in 2012. Over the course of Run 1, 26.4 fb^{-1} of usable integrated luminosity was
³⁵⁰ recorded. Run 2, which spanned 2015–2018, further increased the the proton-proton
³⁵¹ collision energy to 13 TeV. During Run 2 the bunch spacing was reduced, leading
³⁵² to a collision rate of 40 MHz. Over the course of Run 2 a total usable integrated
³⁵³ luminosity of 139 fb^{-1} was recorded. 2022 marked the beginning of Run 3 which,
³⁵⁴ with a higher center of mass energy and peak luminosity, is expected to culminate in
³⁵⁵ an approximate tripling of the dataset size. A summary of key information about
³⁵⁶ 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	50–150	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. The luminosity is the peak instantaneous luminosity. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.

³⁵⁷ An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is
³⁵⁸ at the final stage of a chain of accelerators which incrementally step-up the energy
³⁵⁹ of incoming protons. The first accelerator is Linac2 (which has been replaced by
³⁶⁰ Linac4 in 2020), a linear accelerator which accelerates negative hydrogen ions to an
³⁶¹ energy of 160 MeV. Upon leaving Linac4, the ions are stripped of both electrons
³⁶² and the resulting protons are fed into the Proton Synchrotron Booster (PSB), which
³⁶³ increases the energy of the protons to 2 GeV. The protons leaving the PSB are passed
³⁶⁴ to the Proton Synchrotron (PS), which increases the energy to 26 GeV, and then
³⁶⁵ from the PS to the Super Proton Synchrotron (SPS) which further increases the

366 energy to 450 GeV. Finally, the proton beams are injected in the LHC where they
 367 are accelerated to their final energy of 6.5 TeV (for Run 2).

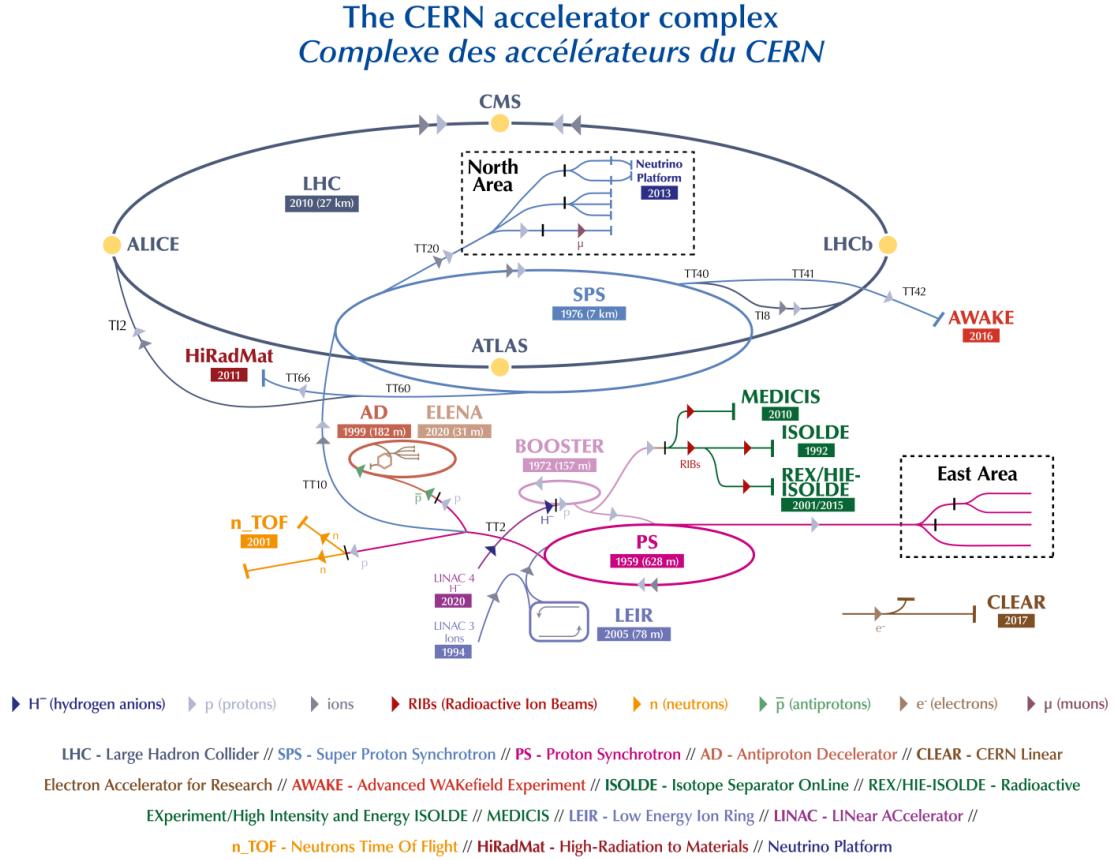


Figure 3.1: An overview of the CERN accelerator complex [36]. The LHC is fed by a series of accelerators starting with Linac2 (or Linac4 from 2020). Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

368 3.2 Coordinate System & Collider Definitions

369 In Section 3.2.1, the coordinate system used at ATLAS is introduced. The parame-
 370 terisation used for specifying the trajectory of charged particle tracks is described in
 371 Section 3.2.2, and definitions for some frequently occurring concepts and quantities
 372 is provided in Section 3.2.3.

³⁷³ 3.2.1 ATLAS Coordinate System

³⁷⁴ The origin of the coordinate system used by ATLAS is the nominal interaction point
³⁷⁵ in the centre of the detector. As shown in Fig. 3.2, the z -axis points along the
³⁷⁶ direction of the beam pipe, while the x -axis points from the interaction point to the
³⁷⁷ centre of the LHC ring, and the y -axis points upwards. The transverse plane lies
³⁷⁸ in x - y while the longitudinal plane lies along the z -axis. A cylindrical coordinate
³⁷⁹ system with coordinates (r, ϕ) is used in the transverse plane, where r is the radius
³⁸⁰ from the origin and ϕ is the azimuthal angle around the z -axis.

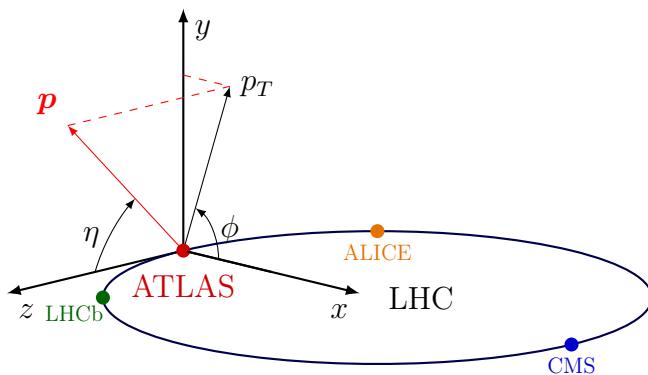


Figure 3.2: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [37].

³⁸¹ The polar angle θ is commonly specified in terms of the pseudorapidity η , defined as

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

³⁸² The pseudorapidity is a convenient quantity to work with as differences in η are
³⁸³ invariant under Lorentz boosts.

³⁸⁴ The transverse momentum p_T of an object is the sum in quadrature of the momenta
³⁸⁵ in the transverse plane

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

386 Angular distance between two objects is measured in units of ΔR and is defined as

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

387 where $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and azimuthal angle between
388 the two objects.

389 3.2.2 Track Parameterisation

390 The trajectories of charged particle tracks are parameterised as a helix which is fully
391 specified using five parameters: $(d_0, z_0, \phi, \theta, q/p)$. The transverse and longitudinal
392 impact parameters (IP) d_0 and z_0 specify the closest approach of the trajectory of a
393 particle to an given origin, where the hard scatter primary vertex (see Section 3.4.2)
394 is used in this thesis. ϕ and θ are the azimuthal and polar angles respectively, and
395 q/p is the measured charge on the track¹ divided by the scalar 3-momentum. Fig. 3.3
396 shows each of these parameters diagrammatically.

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

403 3.2.3 Hadron Collider Definitions

404 Cross Section

405 The cross section σ is closely related to the probability of an interaction between
406 two colliding particles, and is analogous to an effective cross-sectional area of the
407 particles. The cross section of a process depends on the transition matrix element
408 and a phase space integral. At hadron colliders such as the LHC, the proton-proton

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

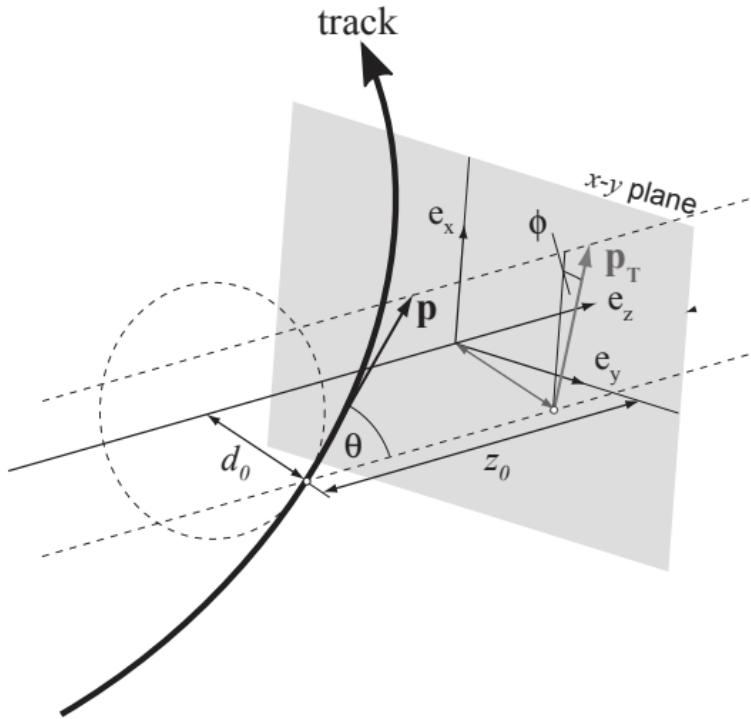


Figure 3.3: The track parameterisation used at the ATLAS detector. Five coordinates ($d_0, z_0, \phi, \theta, q/p$) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum \mathbf{p} and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

⁴⁰⁹ cross section can be factorised as

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

⁴¹⁰ The partonic cross section can be calculated at high energies such as those found at
⁴¹¹ the LHC, while the parton distribution functions (PDFs) have to be extracted from
⁴¹² experimental results.

⁴¹³ Luminosity

⁴¹⁴ The total number of proton-proton collisions N is related to the total pp cross σ
⁴¹⁵ section by the integrated luminosity L , as in

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

⁴¹⁶ The instantaneous luminosity \mathcal{L} relates the cross section to the number of collisions
⁴¹⁷ per unit time. For two colliding bunched proton beams, it is defined as

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

⁴¹⁸ where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
⁴¹⁹ crossing frequency, and σ_x and σ_y are the rms width of the beam in the horizontal
⁴²⁰ and vertical directions.

⁴²¹ The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In
⁴²² total, 139 fb^{-1} of usable physics data was collected over the three-year run. The
⁴²³ uncertainty on the total integrated luminosity is 1.7% [40].

⁴²⁴ Pile-up

⁴²⁵ At the centre of the ATLAS detector, bunches of more than 10^{11} protons are collided.
⁴²⁶ Each bunch-crossing is called an *event*. There is generally one hard proton-proton
⁴²⁷ scatter per event. Additional interactions are typically relatively soft ($low-p_T$) and
⁴²⁸ are known as *pile-up*. Pile-up from interactions within the same bunch-crossing is
⁴²⁹ known as *in-time* pile-up while residual signatures from previous bunch-crossings
⁴³⁰ is known as *out-of-time* pile-up. The number of pile-up interactions is denoted μ ,

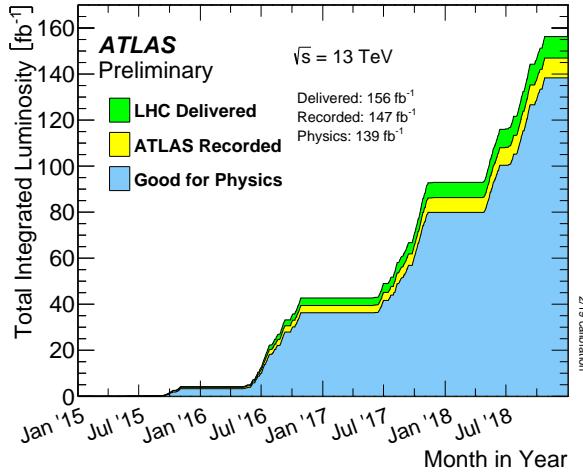


Figure 3.4: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [35]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.

431 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.

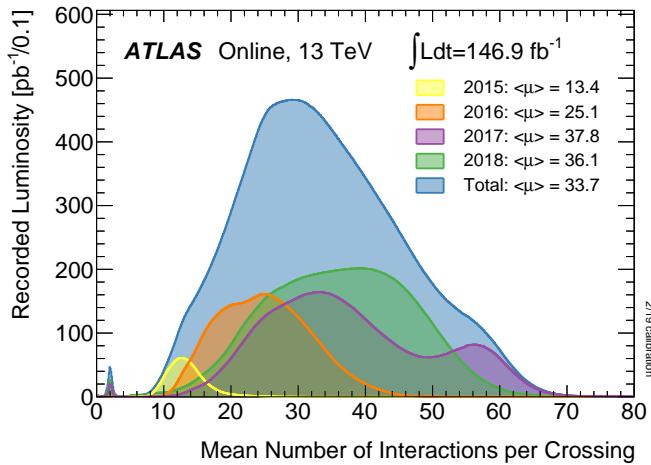


Figure 3.5: Average pile-up profiles measured by ATLAS during Run 2 [35].

⁴³³ 3.3 The ATLAS Detector

⁴³⁴ The ATLAS² detector is made up of several specialised sub-detectors which are
⁴³⁵ arranged concentrically around the nominal interaction point at the centre of the
⁴³⁶ detector, as shown in Fig. 3.6. The detector is designed to cover nearly the entire solid
⁴³⁷ angle around the collision point. In this section a brief overview of each sub-detector
⁴³⁸ is given, in order of increasing radial distance from the point of collision. The inner
⁴³⁹ tracking detector is described in Section 3.3.1, the electromagnetic and hadronic
⁴⁴⁰ calorimeters in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the
⁴⁴¹ trigger is described in Section 3.3.4. More complete information on the detector can
⁴⁴² be found in Ref. [33], while an overview of physics performance is given in [41].

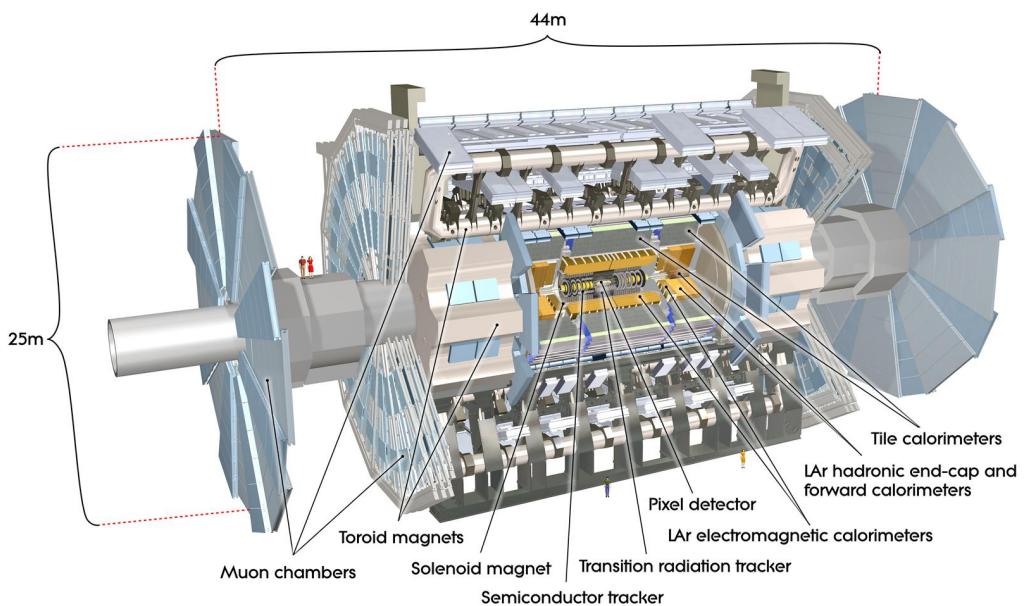


Figure 3.6: A 3D model of the entire ATLAS detector [42]. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.

²A Toroidal LHC ApparatuS.

⁴⁴³ 3.3.1 Inner Detector

⁴⁴⁴ The inner-detector system (ID) provides high-resolution charged particle trajectory
⁴⁴⁵ tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field,
⁴⁴⁶ produced by a superconducting solenoidal magnet, which enables the measurement
⁴⁴⁷ of particle momentum and charge. After Run 3, the ID will be replaced by the
⁴⁴⁸ ITk [43, 44].

⁴⁴⁹ The inner detector is made up of several sub-systems, shown in Figs. 3.7 and 3.8. The
⁴⁵⁰ high-granularity silicon pixel detector covers the innermost region and typically pro-
⁴⁵¹ vides four spacepoint measurements per track. It is followed by the silicon microstrip
⁴⁵² tracker (SCT), which usually provides a further four spacepoint measurements (8
⁴⁵³ hits) per track. These silicon detectors are complemented by the Transition Radiation
⁴⁵⁴ Tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$
⁴⁵⁵ and typically provides 33 (38) additional hits in the barrel (endcap).

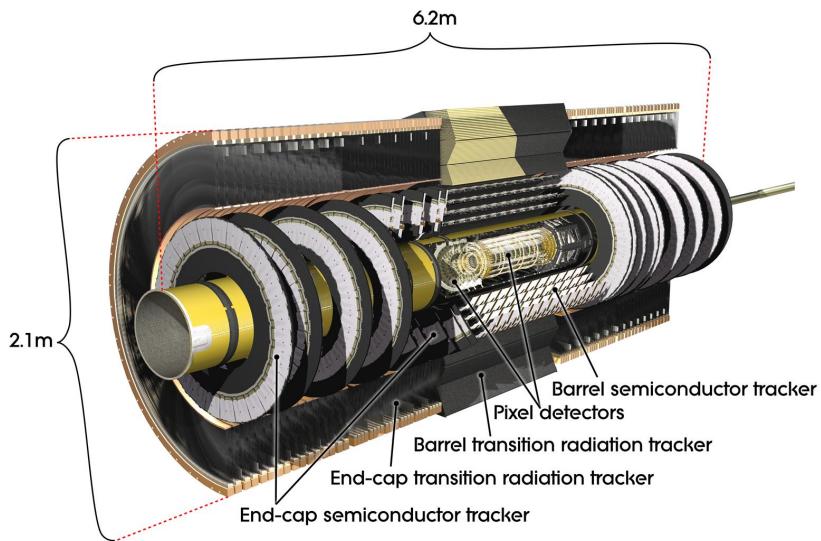


Figure 3.7: A 3D model of the ATLAS ID showing the pixel, SCT and TRT subdetectors [45].

⁴⁵⁶ The target inverse momentum resolution for the combined ID measurement is
⁴⁵⁷ parameterised as a function of the track transverse momentum and polar angle [41].

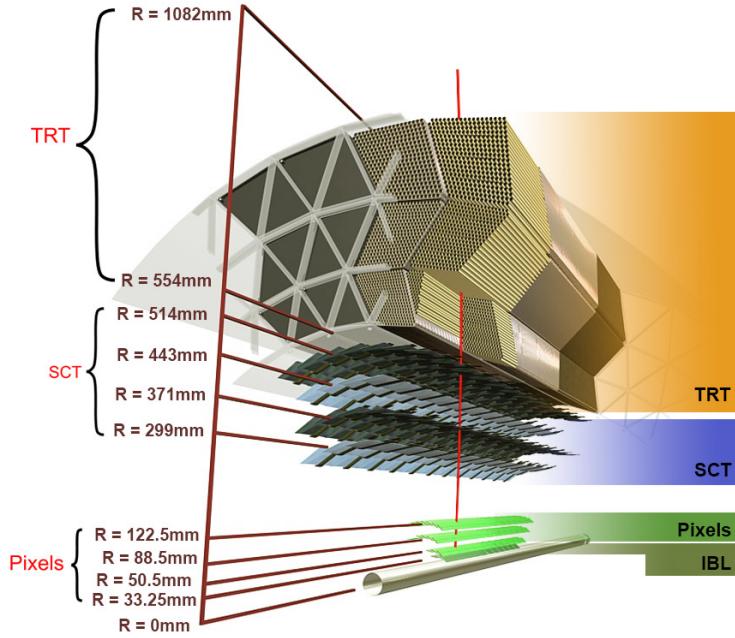


Figure 3.8: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [38].

⁴⁵⁸ The parameterisation is given by

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

⁴⁵⁹ where \oplus denotes a sum in quadrature. For low- p_T tracks (e.g. $p_T \approx 500 \text{ MeV}$) in the
⁴⁶⁰ central region this corresponds to a relative error of approximately 0.01%. Meanwhile
⁴⁶¹ for high- p_T tracks (e.g. $p_T \approx 100 \text{ GeV}$) in the central region this corresponds to
⁴⁶² a relative error of approximately 4%. The momentum resolution generally good
⁴⁶³ enough to correctly identify the sign of the charge on particles up to the highest
⁴⁶⁴ energies expected at the LHC. The transverse impact parameter resolution $\sigma(d_0)$ is
⁴⁶⁵ parameterised similarly as

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

⁴⁶⁶ Typical uncertainties for the transverse IP resolution are 230 μm and 11 μm for low
⁴⁶⁷ and high- p_T tracks in the central region, respectively.

468 Pixel Detector

469 The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
 470 from the beamline, and four disks on each side. The innermost barrel layer is the
 471 insertable B-layer (IBL), shown in Fig. 3.9. The IBL was installed before Run 2 [46,47]
 472 and lies approximately just 33 mm from the beam axis. The second-to-innermost layer
 473 is often referred to as the B-layer. The specification of the pixel detector determines
 474 the impact parameter resolution and the ability to reconstruct primary and secondary
 475 vertices. The detector is required to have a high granularity (i.e. resolution) to
 476 maintain the low occupancy required to resolve nearby particles. Individual pixels
 477 are $50\text{ }\mu\text{m}$ in the transverse direction $R\phi$ and $400\text{ }\mu\text{m}$ in the longitudinal z direction
 478 ($250\text{ }\mu\text{m}$ for the IBL). Cluster positions have a resolution of approximately $10\text{ }\mu\text{m}$ in
 479 $R\phi$ and $100\text{ }\mu\text{m}$ in z .

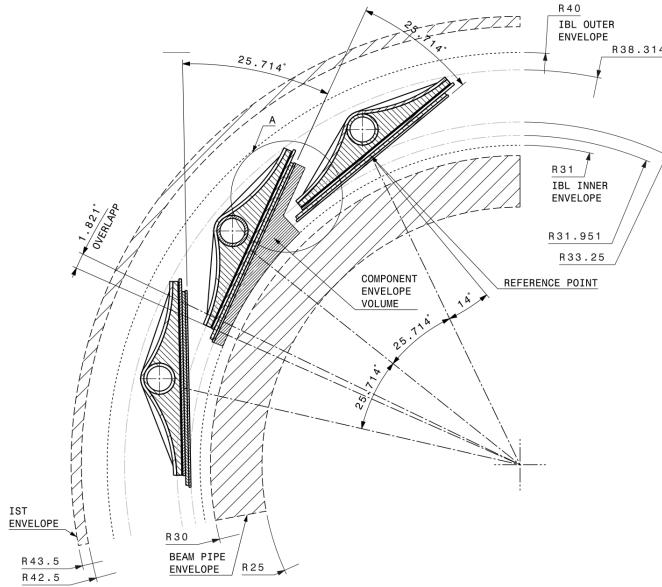


Figure 3.9: A schematic cross-sectional view of the ATLAS IBL [46].

480 Semi-Conductor Tracker (SCT)

481 The SCT is made up of four concentric barrel layers in the central region, and nine
 482 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
 483 with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate

484 to be measured from a pair of strip measurements. The SCT typically provides four
485 precision spacepoint measurements (eight strip measurements) per track in the barrel
486 region. These have intrinsic uncertainties of $17\text{ }\mu\text{m}$ in the transverse direction $R\phi$, and
487 $580\text{ }\mu\text{m}$ in the longitudinal direction z [48]. The measurements provide a contribution
488 to the measurement of charged particle momentum and impact parameter.

489 **Transition Radiation Tracker (TRT)**

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

503 **3.3.2 Calorimeters**

504 The calorimeter system measures the energy of incident particles over the range
505 $|\eta| < 4.9$. There are two main sub-systems: the electromagnetic calorimeter (ECal),
506 which focuses on the measurement of electrons and photons, and the hadronic
507 calorimeter (HCal), which measures the energy of hadrons. A schematic view of the
508 calorimeter system is shown in Fig. 3.10. Upon entering the calorimeter, incident
509 particles will interact with the detector material to produce a shower of secondary
510 particles with reduced energies. The charge deposited in this process is measured to
511 reconstruct the energy of the initial incident particle. The two calorimeter sub-systems
512 must provide strong containment of showering particles to prevent punch-through of
513 EM and hadronic particles to the HCal and muon systems respectively.

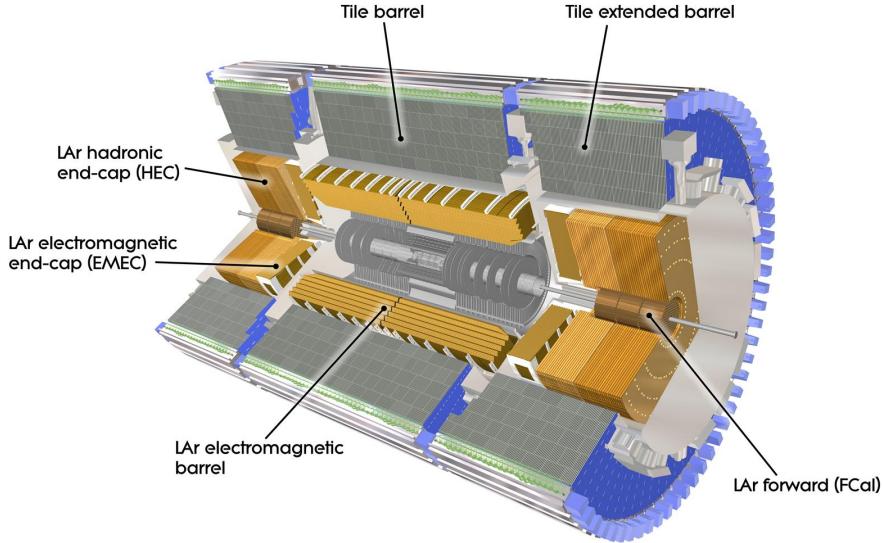


Figure 3.10: The ATLAS calorimeters [49]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic endcaps.

514 Liquid Argon (LAr) Electromagnetic Calorimeter

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

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

530 Hadronic Tile Calorimeter

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

537 The hadronic energy resolution of the HCal is parameterised as a function of the
 538 hadron's transverse energy

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

539 corresponding to a energy resolution of 11% (6.5%) for a hadron with E_T of approxi-
 540 mately 10 GeV (50 GeV) [50].

541 3.3.3 Muon Spectrometer

542 Due to their higher mass, muons easily pass unimpeded through the ID and calorime-
 543 ters and therefore require specialised detectors for their measurement. The Muon
 544 Spectrometer (MS) is made up of dedicated tracking and triggering hardware, as
 545 shown in Fig. 3.11. The precision tracking system uses three layers of monitored drift
 546 tubes with a barrel region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$.
 547 The inner layers of the end-caps use cathode strip chambers to better cope with the
 548 high occupancy in the forward region. The trigger system is comprised of resistive
 549 plate chambers in the barrel region covering $|\eta| < 1.0$ and thin gap chambers in
 550 the end-cap regions covering $1 < |\eta| < 2.4$. A set of three superconducting air-core
 551 toroidal magnets, each made up of eight coils, is used in each of the barrel and
 552 end-caps to deflect the muons as they pass through the MS, allowing their momentum
 553 and charge to be measured from the direction and magnitude of curvature. The
 554 toroidal magnets generate a field which is largely orthogonal to the muon trajec-
 555 tories which allows for maximum deflection. The transverse momentum resolution
 556 (measured for combined ID and muon tracks, see Section 3.4.4) has been measured

557 to be approximately 1.7% in the central region for low- p_T muons, increasing to 4%
 558 for high- p_T muons in the forward regions [51].

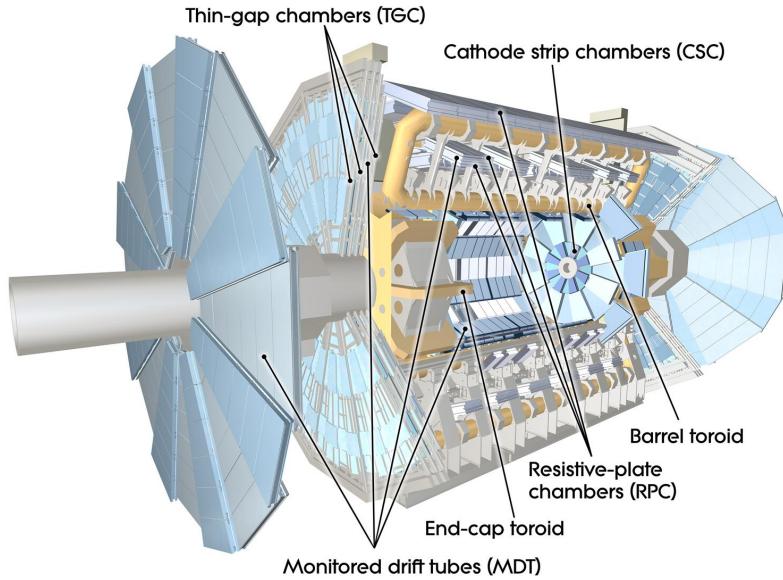


Figure 3.11: The ATLAS muon spectrometer [52].

558

559 3.3.4 The Trigger

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

574 corresponding to 1.2 GB s^{-1} of permanent data storage. More information is provided
575 in [53].

576 3.4 Reconstructed Physics Objects

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

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

586 3.4.1 Tracks

587 The reconstructed trajectories of charged particles are referred to as *tracks*. Tracks
588 are reconstructed from the energy depositions (called *hits*) left by the particles as
589 they traverse the inner detector. Tracks are used in the reconstruction of other
590 objects, including vertices and jets, so their accurate reconstruction is a critical
591 task. A comprehensive introduction to ATLAS tracking is available in Ref. [55],
592 while specific optimisations for dense environments are detailed in Refs. [56, 57]. An
593 overview of track reconstruction is given below.

594 Space-point Formation (Clustering)

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

601 Track Finding

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

618 Ambiguity Solving

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

During the processing of a track candidate, the track is cleaned (whereby problematic hits are removed), and, if the resulting track satisfies the quality selection criteria, a high precision fit of the track parameters using the surviving hits is performed. The high precision fit makes full use of all available information, and uses an updated position and uncertainty estimate for each cluster obtained from a Neural Network (NN) [58]. If the track has reached this stage without being rejected by passing various quality requirements, it is re-scored and returned to the list of track candidates. If the same track is then processed again without requiring modification, it is added to the final track collection. Track candidates that fall below certain quality threshold are rejected. This selection does allow for the possibility of a track having small number of shared hits, as detailed in Table 3.2.

Parameter	Selection
p_T	$> 500 \text{ MeV}$
$ \eta $	< 2.5
$ d_0 $	$< 3.5 \text{ mm}$
$ z_0 \sin \theta $	$< 5 \text{ mm}$
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

Table 3.2: Quality selections applied to tracks, where d_0 is the transverse IP of the track, z_0 is the longitudinal IP with respect to the PV and θ is the track polar angle (see Section 3.2.2 for the IP definitions). Silicon hits are hits on the pixel and SCT layers. 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.

643 Neural Network Cluster Splitting

As part of track cleaning, shared hits are classified by a NN to determine if they are compatible with the characteristic features of a merged cluster [56, 58]. A merged cluster is one made up of a combination of energy deposits from more than one particle, which have become merged due to the closeness of the associated particles and the limited resolution of the detector. It is common for clusters to become

649 merged in dense environments, as discussed in Section 4.1. If the cluster is predicted
 650 to be merged it is labelled as being freely shareable, or *split*. Hits not compatible
 651 with the merged hypothesis can still be shared by a limited number of tracks, but
 652 come with a penalty for the track which may hinder its acceptance into the final
 track collection.

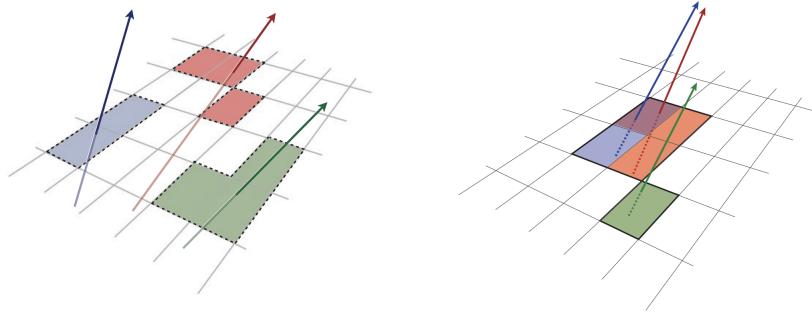


Figure 3.12: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [57].

653

654 Pseudotracking

655 Pseudotracking uses Monte Carlo truth information to group together all the hits
 656 left by each truth particle. Each collection of hits which, as a unit, satisfies basic
 657 quality requirements is directly used in a full resolution track fit. If the track fit is
 658 successful, a “pseudotrack” track is created and stored. If the track fit fails, or the
 659 collection of hits does not pass the basic quality requirements (for example because
 660 of a lack of hits) then the particle is said to be un-reconstructable. In this way,
 661 pseudotracking performance represents the ideal reconstruction performance given the
 662 ATLAS detector, with perfect hit-to-track association and track reconstruction
 663 efficiency. The approach was introduced in Ref. [59] as a way to obtain a fast
 664 approximation of tracking reconstruction for simulated data, however the technique
 665 has become a useful tool for studying tracking performance in general [56].

666 3.4.2 Vertices

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

675 Primary Vertices

676 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices
677 are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [60].
678 The *hard scatter vertex* of an event is chosen as the primary vertex whose associated
679 tracks have the largest sum of transverse momentum squared, $\Sigma(p_T^2)$.

680 Secondary Vertices

681 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient
682 distance from the primary vertex to be resolved from the primary vertex (see
683 Section 4.1.1). Two widely used secondary vertexing tools are used within ATLAS:
684 SV1 and JetFitter [61]. Each attempts to reconstruct secondary vertices inside a jet
685 using the tracks associated to that jet (see Section 3.4.3 for more information about
686 track association). SV1 by design attempts to reconstruct only a single inclusive
687 vertex per jet. This inclusive vertex groups all b -hadron decay products, including
688 tracks from the b -hadron decay itself and tracks from $b \rightarrow c$ decays. The second tool,
689 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary
690 vertices from b -hadron decays are reconstructed separately to tertiary vertices from
691 $b \rightarrow c$ decay chains.

692 3.4.3 Jets

693 Jets are an aggregate reconstructed object corresponding to a collection of collimated
 694 stable particles which results from the presence of a quark or gluon. Jets are built
 695 by clustering constituent objects (e.g. tracks or calorimeter clusters) using a jet
 696 finding algorithm, for example the anti- k_t algorithm [62], which is implemented in
 697 FASTJET [63].

698 Objects can be associated to jets in one of two ways. The first is via a geometrical
 699 matching in ΔR (see) The second is via a ghost association [64], where the object is
 700 assigned a negligible momentum and re-clustered into the jet after its formation.

701 Jets from pile-up interactions are suppressed using [65]

702 EMTopo Jets

703 EMTopo jets are reconstructed from noise-suppressed topological clusters (topoclus-
 704 ters) of calorimeter energy depositions [66]. The clustering uses the energy significance
 705 of each cell, defined as

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

706 where E_{cell} is the energy measured in a given calorimeter cell, and $\sigma_{\text{noise, cell}}$ is the
 707 expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are
 708 formed from a seed cell with a large S_{cell} , and expanded by iteratively adding neigh-
 709 bouring cells with a sufficiently large energy significance. Collections of topoclusters
 710 are then clustered into a jet using the anti- k_t algorithm with a radius parameter of
 711 0.4 (small- R jets) or 1.0 (large- R jets). More information, including information on
 712 the calibration of the topocluster jet energy scale, is available in Ref. [66].

713 Particle Flow Jets

714 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [67] using
 715 the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate
 716 information from both the ID and the calorimeters, improving the energy resolution

717 at high transverse momenta and reducing pile-up contamination. The PFlow jet
718 energy scale is calibrated according to Ref. [68].

719 **Large- R Jets**

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

729 **Track-jets**

730 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm. They
731 are associated to large- R jets as sub-jets and used to identify large- R jets containing
732 b -hadrons. The radius parameter is allowed to vary with transverse momentum such
733 that a broader cone (up to $R = 0.4$) is used for low- p_T track-jets and a narrower
734 cone (down to $R = 0.02$) for high- p_T track-jets [73, 74]. The narrower cone is better
735 suited to clustering highly collimated jet constituents at high- p_T .

736 **Jet Flavour Labels**

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

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

743 Jet Track Association

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

749 3.4.4 Leptons

750 Electrons and muons leave characteristic signatures that are picked up in the ECal
 751 and MS respectively. The reconstruction of both types of charged lepton is briefly
 752 outlined below.

753 Electrons

754 A diagrammatic view of electron reconstruction is shown in Fig. 3.13. Electrons
 755 candidates are reconstructed by matching PV-compatible³ inner detector tracks
 756 to topological calorimeter clusters. The track-cluster matching criteria takes into
 757 account the significant energy loss of the electron due to bremsstrahlung. If a match
 758 is found, a refit of the track is performed using the Gaussian Sum Filter (GSF) [75],
 759 which better handles trajectory reconstruction in the presence of bremsstrahlung.
 760 Various identification criteria are then applied to the candidates using a likelihood-
 761 based (LH) method to improve purity. These include requirements on the track
 762 quality and cluster matching, the shape of electromagnetic shower in the ECal,
 763 leakage into the HCal, and the amount of transition radiation detected in the TRT.
 764 Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may
 765 also be applied. A full description can be obtained from Ref. [76].

766 Muons

767 Muon reconstruction makes use of the dedicated MS (see Section 3.3.3), the tracks
 768 from the ID, and the presence of characteristic signatures in the calorimeters. Muon

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.

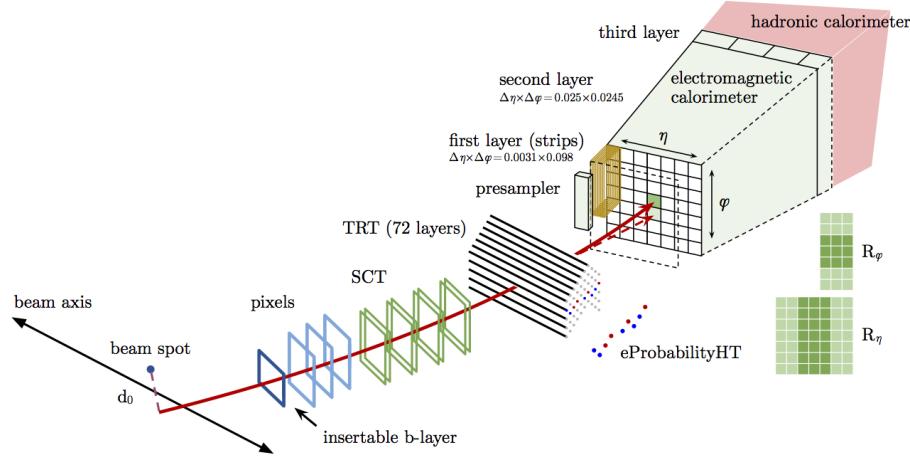


Figure 3.13: A sketch of electron reconstruction using the ATLAS detector [76]. Electron reconstruction makes use of the entire ID and the calorimeters.

769 tracks (i.e. a track reconstructed in the MS) are reconstructed by connecting straight-
 770 line track segments, which are identified via a Hough transform, and combined into
 771 a approximately parabolic trajectory. Finally, a global χ^2 fit is performed, taking
 772 into account possible interactions between the muon and the detector material. A
 773 reconstructed muon is called *combined* if it can be matched successfully to an to
 774 an ID track. Combined muons undergo a further fit with the combined ID and MS
 775 hits, with the energy loss due to the traversal of the calorimeters being taking into
 776 account.

777 After reconstruction, candidate undergo an identification processes which helps to
 778 efficiently identify prompt muons whilst rejecting background signals (e.g. non-
 779 prompt muons from pion and kaon decays, the punch-through of a hadron from the
 780 calorimeter, or the semi-leptonic decay of a heavy flavour hadron). Combined muon
 781 identification takes into account discrepancies in the p_T and charge measurements
 782 in the MS and ID, and the χ^2 of the combined track fit. Selections on the number
 783 of hits in the ID and MS are also applied. At the medium identification working
 784 point, approximately 96% of prompt muons with $20 \text{ GeV} < p_T < 100 \text{ GeV}$ are
 785 successfully identified. On top of the identification requirements, a number of
 786 isolation requirements can also be applied to further suppress background signals.

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

789 3.4.5 Missing Transverse Momentum

790 An imbalance in the final state transverse momentum can occur as a result of
791 incomplete measurement of the final state particles. In particular, neutrinos are
792 not measured by the detector and contribute to the missing transverse momentum
793 $\mathbf{E}_T^{\text{miss}}$. Incomplete detector acceptance and inaccuracies in the reconstruction of the
794 final state can also contribute to the missing transverse momentum of an event. In
795 order to calculate the missing transverse momentum, the negative vector sum of
796 the momentum of all photons, leptons and small- R jets with $p_T > 20 \text{ GeV}$ is taken.
797 The momenta of tracks associated to the primary vertex are also taken into account.
798 The magnitude of $\mathbf{E}_T^{\text{miss}}$ is written E_T^{miss} . More information about missing transverse
799 momentum reconstruction is provided in [78].

800 **Chapter 4**

801 **Tracking and flavour tagging**

802 Many ATLAS analyses rely on flavour tagging, which is the identification of jets
803 containing heavy-flavour hadrons (b -hadrons and c -hadrons) as opposed to those
804 containing only light-flavour hadrons or gluons. In particular, b -tagging is the identifi-
805 cation of jets originating only from b -hadrons (i.e. b -jets). The b -jet identification
806 algorithms (also called *taggers*) work by identifying the unique signatures of b -jets,
807 which are outlined in Section 4.1. The various b -tagging algorithms ultimately take
808 as input information about the reconstructed jet and its associated tracks. Successful
809 b -tagging relies therefore on the efficient and accurate reconstruction of tracks, and
810 especially those tracks corresponding to the products of b -hadron decays.

811 **4.1 b -hadron Reconstruction**

812 This section outlines the typical detector signature of a b -hadron in Section 4.1.1
813 and discusses some associated reconstruction difficulties in Section 4.3.

814 **4.1.1 Decay Topology**

815 b -hadrons are quasi-stable bound states of a bottom quark and one or more lighter
816 quarks. Collectively, these are the B -mesons (e.g. $B^+ = u\bar{b}$, $B^0 = d\bar{b}$) and baryons
817 (e.g. $\Lambda_b^0 = udb$). After a b -quark is produced as the result of a proton-proton collision,
818 they quickly hadronise. The hadronisation process is hard – around 70-80% of
819 the b -quark’s momentum is passed to the b -hadron, with the rest being radiated

as prompt hadronisation or fragmentation particles. See Ref. [79] for a more in depth discussion on hadronisation and the closely related process of fragmentation. Henceforth the combined hadronisation and fragmentation products will be referred to collectively as fragmentation.

b -hadrons are interesting objects of study due to their relatively long proper lifetimes $\tau \approx 1.5$ ps [80]. This lifetime corresponds to a proper decay length $c\tau \approx 450$ μm . In the rest frame of the detector, the typical b -hadron travels a distance

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

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.

b -hadrons decay weakly to on average four or five collimated stable particles [81]. 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. The primary feature is the presence of a high mass secondary vertex that is significantly displaced from the primary vertex. Reconstruction of these vertices from tracks with common points of spatial origin is a common approach used in the identification of b -jets.

Additional signatures of b -hadrons are as follows. Associated tracks and SVs can have a large transverse impact parameter d_0 as a result of the b -hadron displacement (as shown in Fig. 4.2). Since it is common for the b -hadron to decay to a c -hadron with non-negligible lifetime, tertiary vertices can be found within b -jets resulting from $b \rightarrow c$ decay chains. The b - or c -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.

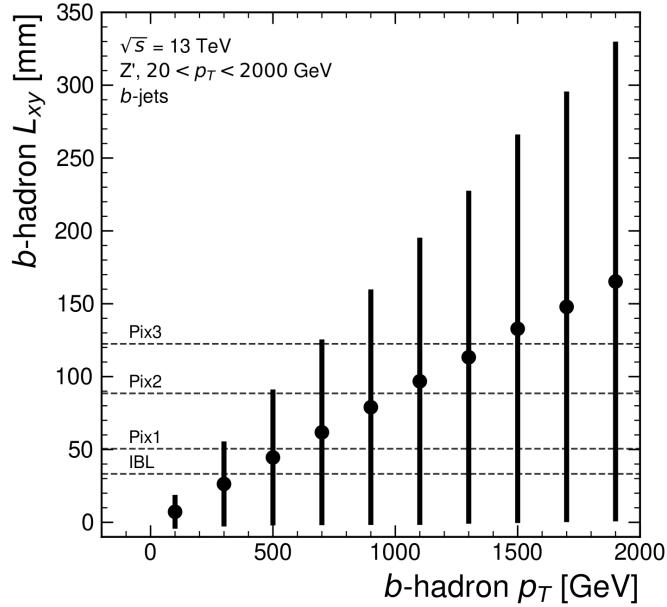


Figure 4.1: The truth b -hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed b -jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

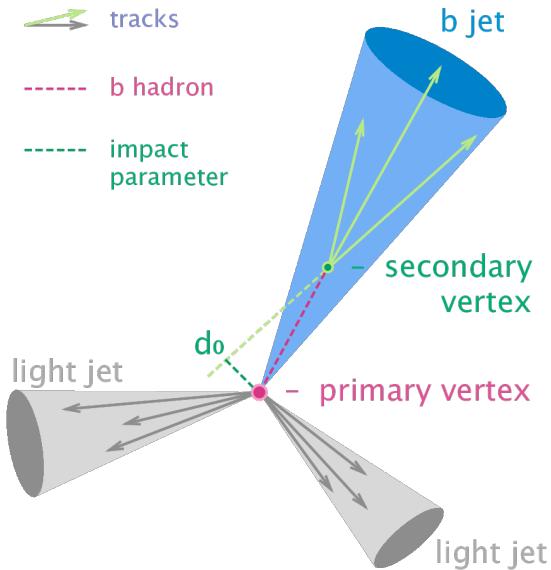


Figure 4.2: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [82]. The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

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

853 4.2 *b*-tagging Algorithms

854 The current ATLAS flavour tagger, DL1r [83], is a deep neural network which
855 accepts as inputs the outputs of a number of independently optimised *low-level*
856 algorithms [61]. Correspondingly, DL1r is referred to as a *high-level* tagger (i.e.
857 one that uses a multivariate approach to combine outputs of the low-level taggers).
858 Each of these low-level algorithms reconstructs distinct features of the experimental
859 signature of heavy flavour jets using the tracks associated to the jet. The low-level
860 algorithms are a combination of manually optimised reconstruction algorithms, for
861 example the SV1 and JetFitter algorithms reconstruct displaced decay vertices, and
862 trained taggers such as RNNIP and DIPS use the IP and hit information from a
863 variable number of tracks to identify the flavour of the jet [61, 84–86].

864 In addition to DL1r, another widely used high-level tagger is the MV2c10 algorithm
865 [61, 83, 87]. This tagger is used in the analysis described in Chapter 7. Similar to
866 DL1r the MV2c10 algorithm takes inputs from the outputs of a number of low-level
867 algorithms (IPxD, SV1 and JetFitter). The outputs of the low-level algorithms
868 are provided as inputs to a boosted decision tree. The working point is tuned to
869 achieve an average *b*-jet efficiency of 70% on simulated $t\bar{t}$ events. At this efficiency
870 working point, rejection factors for *c*-jets and light-jets are approximately 9 and 304
871 respectively.

872 As the different *b*-tagging algorithms ultimately rely on tracks, accurate and efficient
873 track reconstruction is essential. This chapter summarises the challenges facing
874 tracking and *b*-tagging at high transverse momentum with an investigation into track
875 reconstruction performance in Section 4.1. Some preliminary investigations into
876 improving tracking in this regime are investigated in Section 4.4.

877 4.3 Challenges Facing b -hadron Reconstruction

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

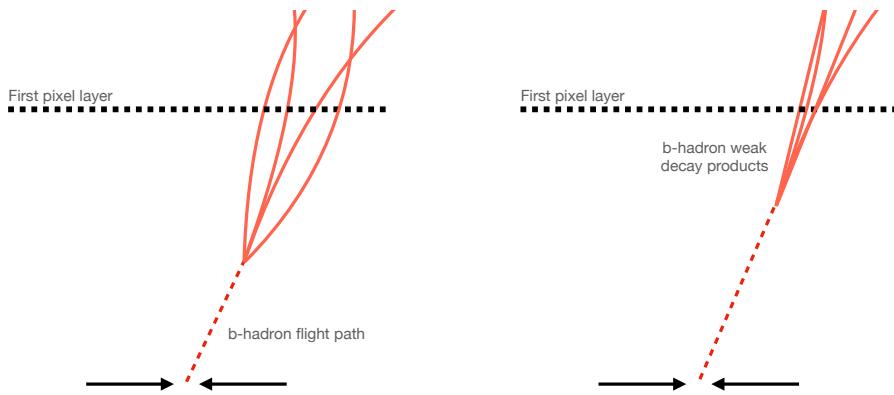


Figure 4.3: At lower p_T (left) the decay length of the b -hadron is on average reduced, and the decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching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

892 As discussed in Section 3.4.1, merged clusters are generally rare, and so shared
 893 hits generally predict bad tracks and are correspondingly penalised during track
 894 reconstruction. However, in the core of high p_T b -jets the density of particles is high
 895 enough that the probability of cluster merging increases dramatically. Successful

896 reconstruction of such tracks requires the presence of shared hits to be effectively
897 dealt with but in the standard reconstruction the presence of these can end up
898 impairing the successfully reconstruction of the track. Furthermore, decays may also
899 take place inside the tracking detectors themselves, which at best leads to missing
900 measurements on the most sensitive detector layers, and at worst can lead to wrong
901 inner layer hits being added to displaced tracks, since the reconstruction process
902 penalises tracks without inner layer hits.

903 The above effects create two related, but distinct problems for b -tagging. The first
904 part is a drop in track reconstruction efficiency. The presence of shared and missing
905 hits reduces a track's score in the ambiguity solver meaning that higher ranking, but
906 potentially worse, track candidates are processed first and take ownership of the hits.
907 This can make it difficult for otherwise reasonable b -hadron decay tracks to meet
908 the ambiguity solver's stringent track quality requirements, leading to their rejection
909 at this stage and an overall decrease in the b -hadron decay track reconstruction
910 efficiency. As shown in Fig. 4.4, this can result in a large drop in reconstruction
911 efficiency for b -hadron decay products of up to 50% for at $p_T = 2 \text{ TeV}$.

912 The second part of the problem is that, due to the high multiplicity of clusters
913 available for assignment in the vicinity of the typical high- p_T b -hadron decay track,
914 and also given the strong positive bias of the ambiguity solver towards those tracks
915 with pixel measurements in each layer (especially the innermost IBL measurement),
916 many b -hadron decay tracks are assigned incorrect inner layer hits. This is only a
917 problem for those decay products which were produced within the pixel detector
918 as a result of a significantly displaced b -hadron decay, and so do not have a correct
919 hit available for assignment. Fig. 4.5 shows the number of hits as a function of the
920 reconstructed track p_T for fragmentation tracks and tracks from the weak decay
921 of the b -hadron. The baseline tracks represent the standard reconstruction setup,
922 while the pseudotracks represent the ideal tracking setup as outlined in Section 3.4.1.
923 Hit multiplicities on the pseudotracks decrease at high p_T due to the flight of the
924 b -hadron before its decay. The baseline tracks have more hits than the pseudotracks,
925 indicating that they are being incorrectly assigned additional hits on the inner layers
926 of the detector.

927 These incorrect hits may skew the parameters of the track, which can in turn lower
928 the performance of the downstream b -tagging algorithms. In particular, b -tagging
929 algorithms rely heavily on the transverse impact parameter significance $s(d_0)$ of the

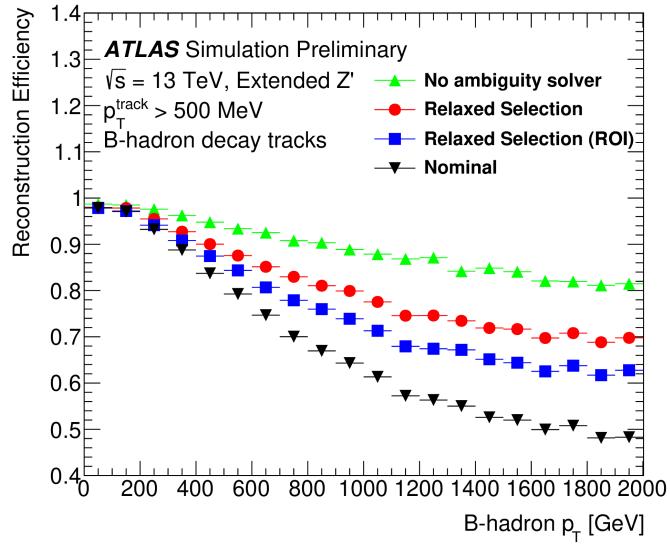


Figure 4.4: b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [88]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T b -hadrons,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

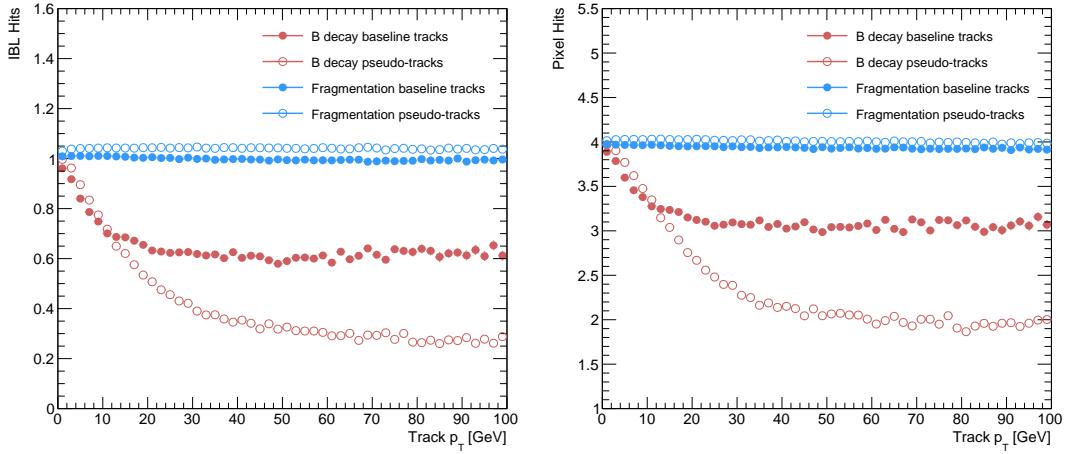


Figure 4.5: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track for tracks in jets in a Z' sample at $\sqrt{s} = 13$ TeV. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in 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.

930 track (see Section 3.2.2). The quality of this measurement is expected to be adversely
931 affected by wrong inner-layer hits on the track. Furthermore, multiple tracks sharing
932 an incorrect hit can lead to the creation of spurious secondary vertices, which can
933 cause further problems for the b -tagging algorithms.

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

939 4.4 Investigations into High p_T b -hadron Tracking

940 In Section 4.4.1 pseudotracks, a key tool for studying the ideal tracking performance
941 of the ATLAS detector, are used to study the shared hit requirements on tracks in
942 the dense cores of high- p_T b -jets. Section 4.4.2 details a study which investigated
943 modifying the global track fitter to improve reconstruction performance in this
944 regime.

945 4.4.1 Shared Hits

946 The ambiguity solver is not run for pseudotracks. However, if the standard track
947 collection is produced alongside the pseudotracks, then cluster splitting neural
948 networks will be run for the standard tracks, and the resulting classification of
949 clusters will be propagated to hits on pseudotracks. This quirk allows one to study
950 the inefficiencies of the cluster splitting process, and relatedly to determine whether
951 shared hit cuts in the ambiguity solver are too loose or too tight. The fraction of hits
952 that are shared for the IBL and the B-layer is shown in Fig. 4.6. The shared hits on
953 pseudotracks represent correctly assigned hits from merged clusters that were not
954 able to be classified as split by the cluster splitting neural networks. As such, these
955 represent the number of shared hits the ambiguity solver should aim to allow given
956 the current performance of the cluster splitting algorithm. For shared hits on the IBL
957 for particles produced before the IBL, the baseline selection appears to be successful
958 in disallowing excessive numbers of shared hits. However, the ambiguity solver fails to

959 limit shared hits for those particles produced after the IBL, reflecting the previously
 960 discussed problem of displaced tracks picking up incorrect hits. Meanwhile, it is clear
 961 that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of
 962 shared hits.

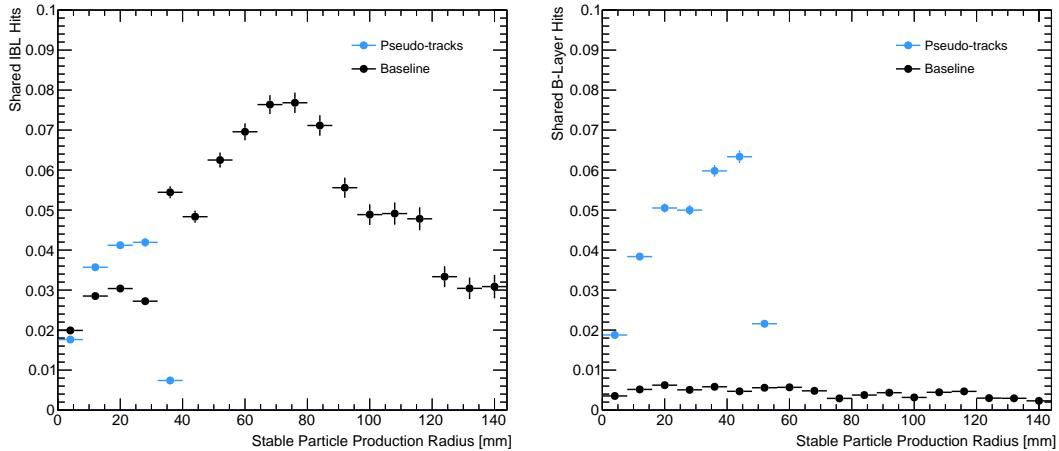


Figure 4.6: The fraction of IBL (left) and B-layer (right) hits which are shared on b -hadron decay tracks as a function of the production radius of the b -hadron decay product for tracks in jets in a Z' sample at $\sqrt{s} = 13$ TeV. Pseudotrack represent the ideal performance given the ATLAS detector, see Section 3.4.1.

963 4.4.2 Global χ^2 Fitter Outlier Removal

964 This section documents ongoing studies into improving hit-to-track assignment by
 965 using the Global χ^2 Fitter (GX2F) to identify and prevent incorrect hits from
 966 being assigned to tracks during the track fit. This is in contrast to a previously
 967 investigated approach [89] which attempted to identify and remove incorrect hits
 968 after the reconstruction of the track. As part of the track fit, an outlier removal
 969 procedure is run, in which suspicious hits are identified and removed.

970 The GX2F code, as a relatively low-level component of track reconstruction, has
 971 not undergone significant modification for several years, and was originally only
 972 optimised in the context of prompt, isolated tracks. During this time, a new tracking
 973 sub-detector, the IBL, was installed. The motivation for looking at the GX2F is that
 974 these changes may require re-optimisation of the GX2F code, and in particular the
 975 outlier removal procedures. Further motivation for this approach comes from the

976 low rate of labelled outliers in baseline tracking, in contrast to the relatively higher
 977 rate of tracks with an incorrect IBL hit.

978 Implementation

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

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

988 This pull is computed for the transverse and longitudinal coordinates of the mea-
 989 surement, and the maximum of the two is selected and checked to see if it exceeds
 990 a certain selection threshold. If it does, the hit will be removed if the track also
 991 exceeds a threshold on the total χ^2/n , where n is the number of degrees of freedom
 992 on the track. The results of varying the outlier selection and χ^2/n thresholds are
 993 described below.

994 Selection Optimisation

995 A systematic variation of the outlier selection and χ^2/n thresholds has been carried
 996 out. Both thresholds were reduced in fixed step sizes of 0.25 for the outlier selection
 997 threshold and 1 for the χ^2/n threshold. The results for the best performing selections
 998 are discussed below. The value of the outlier selection threshold was reduced from 4
 999 down to 1.75, a change which affects the silicon layers (the TRT has separate outlier
 1000 removal logic). Furthermore, a specific cut for the IBL was introduced, and after
 1001 optimisation is set to 1.25. The second threshold on the track χ^2/n was also reduced
 1002 from 7 to 4. Finally, instead of taking the maximum of the pulls in the longitudinal

and transverse directions, a quadrature sum is taken of these two values and used.
 This variation is labelled “Mod GX2F” and was found to improve performance.

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

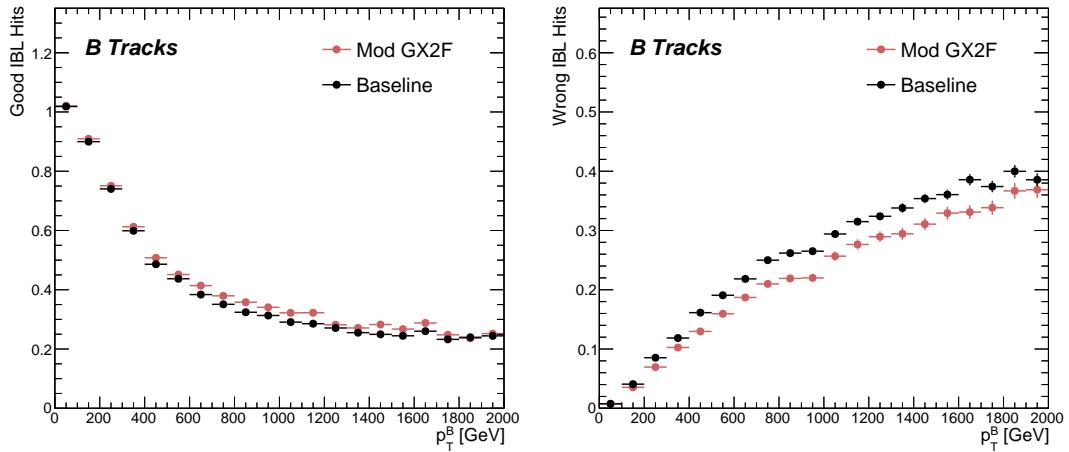


Figure 4.7: The average number of good (left) and wrong (right) IBL hits as a function of b -hadron p_T for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red).

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

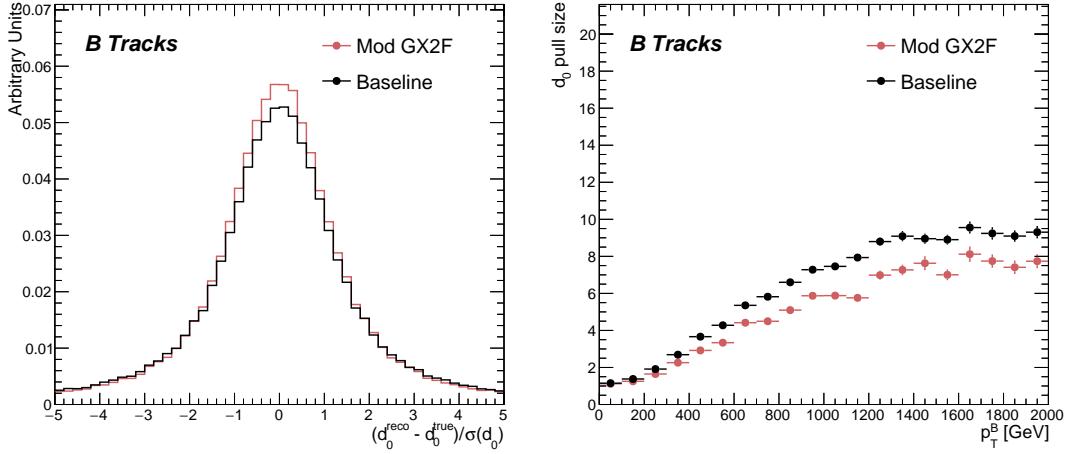


Figure 4.8: (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified GX2F tracks. (right) The absolute value of the d_0 pull as a function of the truth b -hadron transverse momentum.

4.5 Conclusion

In this section, the difficulties facing efficient and accurate track reconstruction, and hence performant b -tagging, have been outlined. The ambiguity solver, which attempts to clean or reject tracks which have an excessive number of shared hits, is shown to be overly aggressive in the removal of b -hadron decay product track candidates. The ambiguity solving process relies on a complicated pre-defined selection which has not been optimised for high transverse momentum b -hadron track reconstruction. These conclusions have motivated further ongoing studies into the improvement of the track reconstruction in dense environments and the high- p_T regime, such as those in Ref. [88].

An optimisation of the outlier removal process in the global χ^2 fitter was carried out. The results of the optimisation show that more aggressive removal of outlier hits can lead to fewer wrong hits being assigned to tracks, and improvements in the pulls of the track parameters.

Future Work

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

1035 in the baseline tracking configuration. It is also necessary to study the impact of
1036 the improved outlier removal on the downstream *b*-tagging algorithms. Thanks to
1037 the all-in-one flavour tagging approach described in Chapter 6, this will in future be
1038 easier to study.

1039 As there are some known data-MC discrepancies, fine tuned optimisation such as the
1040 work presented here presents an opportunity to over-optimise the tracking algorithms
1041 on MC. As such, further studies validating the improved outlier removal procedure
1042 on data are required.

1043 **Chapter 5**

1044 **Track Classification MVA**

1045 This chapter details work on implementing a multivariate algorithm (MVA) to predict
1046 the truth origin of reconstructed tracks. An introduction to formalisms of machine
1047 learning is given in Section 5.1. In Section 5.3, the truth origin label is defined,
1048 and in Section 5.4 these labels are used to train a machine learning model that can
1049 effectively discriminate between good and fake tracks. Several studies motivated this
1050 work by demonstrating that at high p_T , b -tagging performance was degraded by the
1051 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
1052 could be trained to detect fake tracks, these could be removed before their input to
1053 the b -tagging algorithms with the aim of improving performance. In addition, other
1054 groups that are sensitive to the presence of fake tracks would also benefit from this
1055 work.

1056 **5.1 Machine Learning Background**

1057 Over the past few decades, machine learning (ML) techniques have become increas-
1058 ingly popular in High Energy Physics experiments due the increased volumes of
1059 high-dimensional data and improvements in the techniques used (in particular deep
1060 learning). Machine learning is the process by which a computer program uses data
1061 to learn suitable parameters for a predictive model. This is opposed to explicitly
1062 providing instructions on how to perform a task. A subfield known as *supervised*
1063 *learning* is used in this work, and consists of exposing a model to a large number of
1064 labelled examples in order to extract relationships between the input data and their

1065 labels. These relationships are often complex, and explicitly programmed rules can
1066 fail to fully capture the relationships between inputs and outputs.

1067 In the simplest case, a set of m labelled training examples $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$
1068 is collected. Each element (x_i, y_i) consists of a input vector $x_i \in \mathbb{R}^{\text{input}}$, and the
1069 corresponding label y_i . In classification problems, these labels are integer *class*
1070 *labels* $y_i \in \{0, \dots, N - 1\}$, where N is the number of classes, which specify which
1071 of a pre-determined set of categorical classes the training example belongs to. The
1072 rest of the discussion in this chapter is limited to binary classification problems
1073 ($N = 2$). The two classes are often referred to as signal ($y_i = 1$) and background
1074 ($y_i = 0$), which need to be separated. Collecting sufficient and suitable data is one
1075 of the primary challenges of machine learning, as such data is not always readily
1076 available. Fortunately, sophisticated tools to simulate particle collisions have already
1077 been developed by the scientific community [90, 91]. These tools play a key role in
1078 generating a suitably large amount of labelled data which is used to train algorithms.
1079 More detail on the input datasets is given in Section 5.4.1.

1080 After obtaining suitable training data, the next step is to define a model. Given an
1081 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
1082 parameterised functional mapping from input space to output space. Given an input
1083 example x_i and a set of parameters θ , the model outputs a prediction $\hat{y}_i \in (0, 1)$ for
1084 the true label y_i , as in

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

1085 The output \hat{y}_i is in the interval $(0, 1)$ so as to be interpreted as the probability that
1086 the input example x_i belongs to the signal class. The parameters θ of the model are
1087 randomly initialised, and the model is designed to be expressive enough to correctly
1088 map the inputs x_i to the outputs y_i given a reasonable optimisation of the parameters.
1089 To perform this optimisation, the model is then trained, which amounts to showing
1090 the model a series of labelled training examples and modifying the parameters of the
1091 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].

Artificial Neurons

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

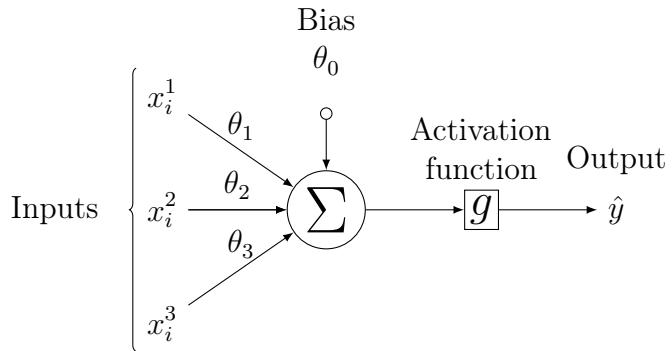


Figure 5.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression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

The output of the dot product and bias term z is fed into an activation function $g(z)$. The activation function has several uses, most notably acting as a source of non-linearity and bounding the output of the neuron. Some common activation functions (sigmod, tanh, ReLU and SiLU) are shown in Fig. 5.2. The choice of activation function can have implications for the performance and convergence of the network, since the gradient of $g(z)$ is used to compute the weight updates during

1110 training. This is also why input data is typically normalised to have zero mean and
 1111 unity variance [96].

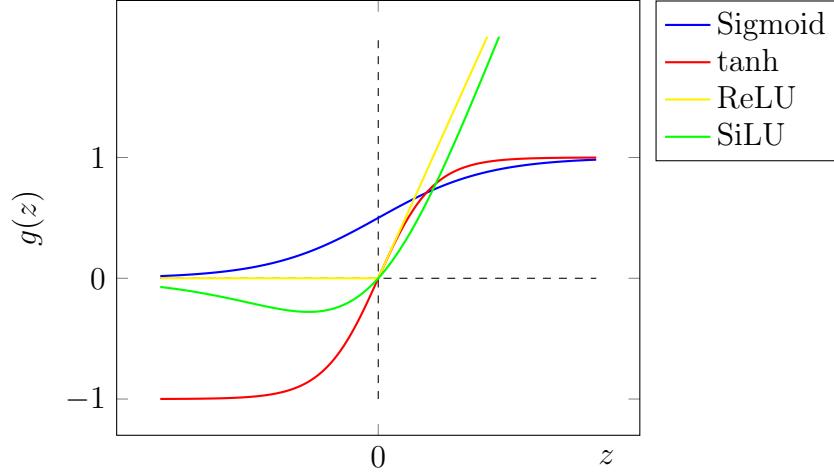


Figure 5.2: The output of several common choices for the activation function $g(z)$ of an artificial neuron. The input z is the output of the dot product between the activation and the weights, plus a bias term.

1112 Networks

1113 Several neurons are linked together in layers to form a neural network. The inputs
 1114 are propagated layer-by-layer through the network until reaching the final output
 1115 layer. The number of layers and neurons per layer are important hyperparameters
 1116 (those parameters which are not optimised as part of the training process) which
 1117 influence the performance of the model. In the case of binary classification, the final
 1118 output layer generally consists of a single neuron with a sigmoid activation

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

1119 where z is the output from the dot product of the inputs and the weights, plus the
 1120 bias term. This value is bounded between zero and one allowing the final output to
 1121 be interpreted as the probability that the input sample belongs to the signal class.
 1122 NNs have the crucial property of being differentiable functions, which facilitates the
 1123 training process described in the next section.

¹¹²⁴ 5.1.2 Training with Gradient Descent

¹¹²⁵ A training algorithm is used to optimise the weights and biases of a NN after
¹¹²⁶ exposure to the training data. The training algorithm works by minimising a loss
¹¹²⁷ function L , which quantifies the error in the model's predictions. NNs are commonly
¹¹²⁸ trained using backpropagation in combination with a variant of the stochastic gradient
¹¹²⁹ descent algorithm to iteratively update the model parameters. In binary classification
¹¹³⁰ problems, the binary cross entropy given in Eq. (5.3) loss if often used.

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

¹¹³¹ Since the model f is differentiable, a correction for each parameter θ_i can be computed
¹¹³² by taking the partial derivative of L with respect to the parameter. Updated
¹¹³³ parameters θ'_i are calculated by updating the original parameter in the direction
¹¹³⁴ which reduces the loss.

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

¹¹³⁵ The hyperparameter α is known as the *learning rate* and dictates the size of the
¹¹³⁶ step taken in the direction of the slope. The errors for each parameter are efficiently
¹¹³⁷ calculated using the backpropagation algorithm [93]. The process of updating weights
¹¹³⁸ is repeated until the weights are judged to have converged, which means the network
¹¹³⁹ is trained. In practice, small batches of the input data are shown to the network at
¹¹⁴⁰ a time. For each batch the average loss is calculated and the network's weights are
¹¹⁴¹ updated. There are many extensions and variations of the gradient descent algorithm.
¹¹⁴² This work uses the Adam optimiser which adds momentum to the weight updates
¹¹⁴³ (dampening oscillations) and an adaptive per-parameter learning rate [97].

¹¹⁴⁴ 5.2 Graph Neural Network Theory

¹¹⁴⁵ Graph neural networks are a more sophisticated neural network model (see Sec-
¹¹⁴⁶ tion 5.1.1) that are designed to operate on graph structured data. A brief introduction
¹¹⁴⁷ to GNNs is provided in this section following the formalism in Ref. [98].

1148 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 $\mathcal{E} =$
1149 $\{e_i\}_{i=1:N^e}$, and a global representation u . Each node represents an individual object,
1150 and edges are directed connections between two nodes, called the *sender* and *receiver*
1151 nodes. The connectivity of the graph therefore encodes information about the
1152 relationships between objects that exist in the graph.

1153 A single graph network layer consists of three separate update functions ϕ^e , ϕ^h and
1154 ϕ^u one for each of the nodes, edges, and global graph representation, and similarly
1155 three aggregation functions $\rho^{e \rightarrow h}$, $\rho^{e \rightarrow u}$ and $\rho^{h \rightarrow u}$. The aggregation functions combine
1156 information across different edges or nodes for input into the update functions,
1157 which produce new representations for the nodes, edges and global objects based on
1158 the information in the previous layer and the aggregated information. The update
1159 functions are typically each implemented as a dense feedforward neural network (as
1160 described in Section 5.1.1). The edges e_i are updated by a edge network ϕ^e as in

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

1161 where h_s and h_r are the sender and receiver nodes respectively. The nodes are
1162 updated with a node network ϕ^h as in

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

1163 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
1164 referred to as the edge aggregation function. The global representation is updated
1165 using the global network ϕ^u as in

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

1166 where \bar{e}' is the aggregation $\rho^{e \rightarrow u}$ over all updated edges e'_i and \bar{h}' is the aggregation
1167 $\rho^{h \rightarrow u}$ over all updated nodes h'_i .

1168 The graph network layer performs a graph convolution, in an analogous way to a
1169 convolutional neural network operating on a grid of pixels. The above description
1170 is general, and not all concrete implementations of GNNs need to implement every
1171 aspect. For example, the global graph representation need not be present, and
1172 it is also possible that no dedicated edge features are present. In such cases the
1173 corresponding update and aggregation functions are not needed. Fig. 5.3 shows two

1174 possible graph network update layers. The layer used in the GN1 model is specified
1175 in more detail in Section 6.3.3.

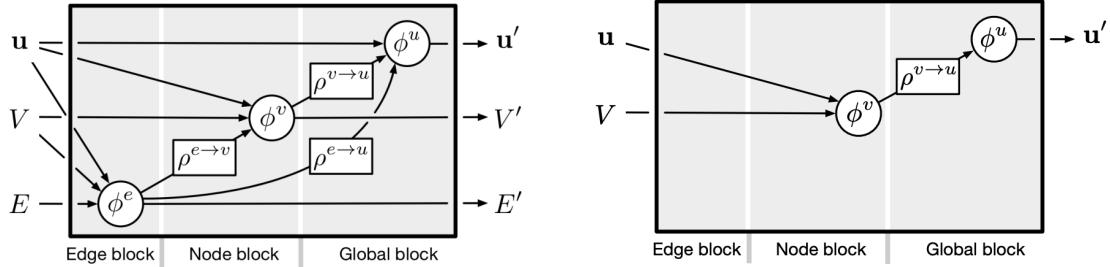


Figure 5.3: The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [99]. Reproduced from Ref. [98].

1176 5.3 Track Truth Origin Labelling

1177 Crucial to supervised learning techniques are the ground truth class labels which the
1178 machine learning model is trained to predict. A set of track truth labels which a
1179 high degree of granularity have been implemented in the ATLAS software stack, and
1180 are listed in Table 5.1. The labelling scheme has been designed to be useful beyond
1181 the classification of good and fake tracks. The origins are determined by analysing
1182 the simulated record to determine the physical process that led to the creation of
1183 the truth (i.e. simulated) particle which is associated with each reconstructed track.
1184 Tracks are associated with truth particles by selecting the particle with the highest
1185 *truth-matching probability* (TMP), defined in Eq. (5.8). For a given truth particle,
1186 the TMP is a weighted sum of the number of hits on a reconstructed track which
1187 are matched to the truth particle N^{match} , divided the total number of hits on the
1188 track N^{total} . The weights are subdetector-dependent and are designed to account for
1189 the varying importance of the different ID subdetectors (based upon their precision)
1190 in the reconstruction of a track.

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

1191 For the fake track classification tool, the track truth origins in Table 5.1 are used
1192 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 but due to the presence of some wrong hits on the track will not accurately reproduce the true trajectory. In such cases the fake track can still be identified as having an origin: it is for example possible to have a fake track which is from the decay of a b -hadron.

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 physical 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 0.75 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 4.3. 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 efficiency to mistag a light-jet decreases by up to 35% at a b -jet efficiency of 35% if such tracks are removed.

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, whereas 0.5 was used in Chapter 6.

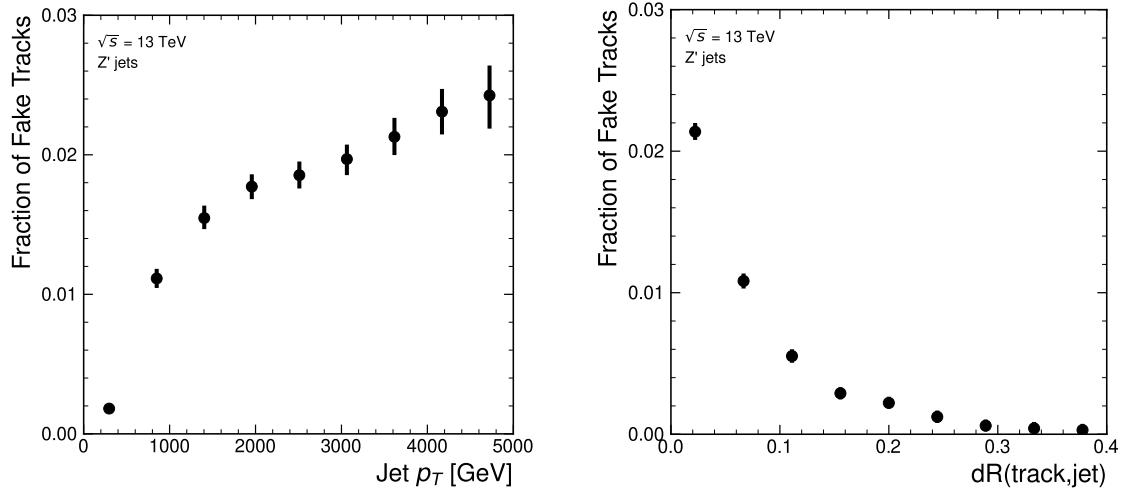


Figure 5.4: Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track}, \text{jet})$ (right) for jets in the Z' sample. The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases.

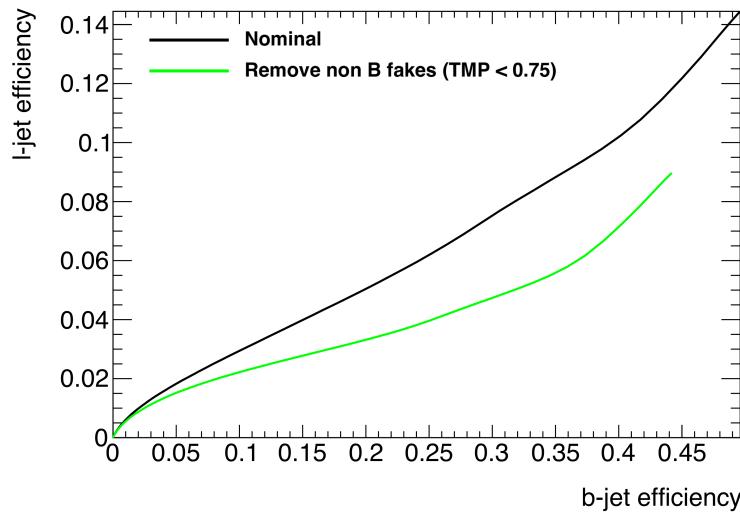


Figure 5.5: The light-jet efficiency of the low level tagger SV1 for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$, as a function of b -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a b -hadron are removed.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

1208 To identify and remove fake tracks, a NN classification tool was trained with all
1209 non-fake tracks as the signal class and fake tracks as the background class. Inputs to
1210 the model are described in Section 5.4.2, while fake track removal performance is
1211 given in Section 5.4.4.

1212 **5.4.1 Datasets**

1213 To train and evaluate the model, simulated SM $t\bar{t}$ and BSM Z' events initiated by
1214 proton-proton collisions at a center of mass energy $\sqrt{s} = 13 \text{ TeV}$ were used. The Z'
1215 sample is constructed in such a manner that it has a relatively flat jet p_{T} spectrum
1216 up to 5 TeV and decays democratically to equal numbers of b -, c - and light-jets.
1217 The generation of the simulated event samples includes the effect of multiple pp
1218 interactions per bunch crossing with an average pile-up of $\langle \mu \rangle = 40$, which includes
1219 the effect on the detector response due to interactions from bunch crossings before
1220 or after the one containing the hard interaction.

1221 The $t\bar{t}$ events are generated using the POWHEGBox v2 generator [100–103] at
1222 next-to-leading order with the NNPDF3.0NLO [104] set of parton distribution
1223 functions (PDFs). The h_{damp} parameter² is set to 1.5 times the mass of the top-quark
1224 (m_{top}) [105], with $m_{\text{top}} = 172.5 \text{ GeV}$. The events are interfaced to PYTHIA 8.230 [106]
1225 to model the parton shower, hadronisation, and underlying event, with parameters
1226 set according to the A14 tune [107] and using the NNPDF2.3LO set of PDFs [108].
1227 Z' events are generated with PYTHIA 8.2.12 with the same tune and PDF set. The
1228 decays of b - and c -hadrons are performed by EVTGEN v1.6.0 [109]. Particles are
1229 passed through the ATLAS detector simulation [110] based on GEANT4 [111].

1230 Jets are required to have a pseudorapidity $|\eta| < 2.5$ and $p_{\text{T}} > 20 \text{ GeV}$. Additionally,
1231 a standard selection using the Jet Vertex Tagger (JVT) algorithm (see Section 3.4.3)
1232 at the tight working point is applied to jets with $p_{\text{T}} < 60 \text{ GeV}$ and $|\eta| < 2.4$ in order
1233 to suppress pile-up contamination [65].

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.

1234 5.4.2 Model Inputs

1235 The fake track MVA is given two jet variables and 20 tracking related variables
1236 for each track fed into the network. The jet transverse momentum and signed
1237 pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in
1238 Table 5.2.

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.

1239 The track parameters and hit pattern are key indicators of whether or not a track
1240 is fake. The FracRank variable is the ordered index of the tracks that pass the
1241 ambiguity solver's selection divided by the total number of successfully reconstructed
1242 tracks in the event. The ambiguity solver processes track candidates iteratively

in order of an internal score (see Section 3.4.1), and the order in which tracks are accepted is preserved. Since tracks with shared hits have lower scores, tracks which do not require the removal of shared hits are likely to be processed and accepted earlier on, whereas tracks with shared hits will be processed later and potentially have their shared hits removed. Hence the FracRank variable gives an indication of the track quality and how likely it is that hits would have been removed (tracks processed later on are more likely to have hits removed).

Track selection follows the loose selection described in Ref. [86] and outlined in Table 3.2, which was found to improve the performance compared to previous tighter selections, whilst ensuring good resolution of the track’s parameters and a low fake rate [57]. Inputs are scaled to have a central value of zero and a variance of unity before training and evaluation.

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 balance their relative weights. 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 – 54,000. The model was trained using 40 million tracks with a further 4 million tracks each used for validation and testing. A full list of the model hyperparameters is given in Table 5.3.

5.4.4 Results

In order to evaluate the fake track classification tool, a orthogonal test sample of 4 million tracks in jets in the combined $t\bar{t}$ and Z' samples was 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 5.3: 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.

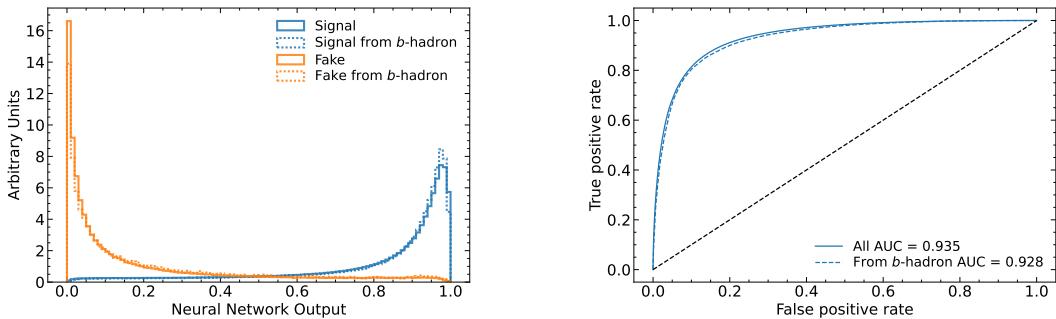


Figure 5.6: (left) Normalised histograms of the fake track classification model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line). The plots show tracks in the combined $t\bar{t}$ and Z' testing sample. 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 5.4. The results demonstrate that the tool is effective in retaining 98.8% of signal tracks, while correctly identifying (and therefore enabling the removal of)

1281 45.6% of fake tracks. Table 5.4 also shows that a significant amount of tracks which
1282 are labelled as both fake and from the decay of a b -hadron are also removed. This can
1283 happen because fake tracks with $\text{TMP} < 0.75$ are still matched to a truth particle,
1284 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 5.4: Good and fake track selection efficiencies for the combined $t\bar{t}$ and Z' samples.
Two working points are defined, cutting on the NN output at 0.06 and 0.12.

1285 5.5 b -hadron Track Identification

1286 After initial tests and investigation, it was found that fake tracks which were the result of b -hadron decays actually aided b -tagging performance, as demonstrated in Fig. 5.8. The application of a single tool which removed all fake tracks was therefore not optimal. A second tool was therefore trained in the same manner as the first, this one was designed to distinguish between those tracks which were from the decay of a b -hadron (FromB and FromBC in Table 5.1) and those which were not (all other truth origins). Fake tracks which were from the decay of a b -hadron were included in the signal class. The b -hadron decay track MVA was trained using the same setup as described above, with the same tracks, input variables, and training procedure. The performance of the model to separate b -hadron decay tracks from other tracks is shown in Fig. 5.7. Using a selection WP of 0.1, the model can retain 98.5% of b -hadron tracks and reject 46.2% of tracks not from the decay of a b -hadron. In Section 5.6, this model is used in conjunction with the fake track identification MVA to identify and remove fake tracks which are not from the decay of a b -hadron.

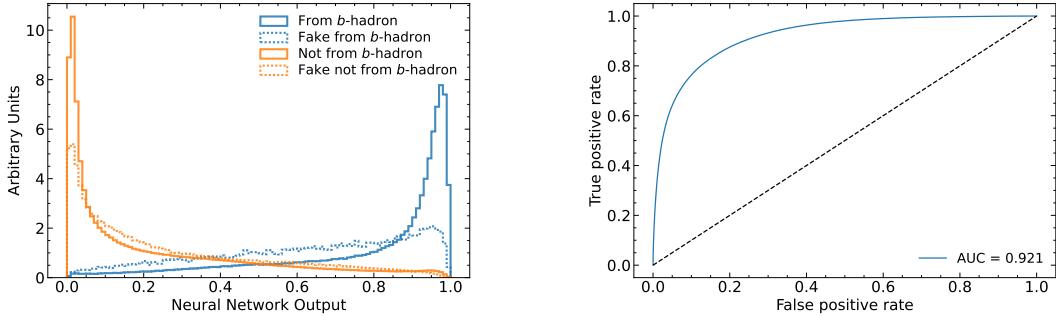


Figure 5.7: (left) Normalised histogram of the b -hadron track identification model output separated for tracks from the decay of a b -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line). The plots show tracks in the combined $t\bar{t}$ and Z' testing sample.

1300 5.6 Combined Approach

1301 A 2-dimensional cut was then used to only reject those tracks that had a high
 1302 probability of being fake, and also a low probability of being a b -hadron decay track.
 1303 The results of the combined approach are provided in Table 5.5, which shows that
 1304 for the working point “A”, 98.6% of b -hadron decay tracks (both good and fake) are
 1305 retained, while 50.7% of fake tracks which are not from b -hadron decays are rejected.

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

Table 5.5: 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).

1306 The light-jet efficiency of SV1 is successfully reduced when using the combined tools
 1307 to remove fake tracks that are not from a b -hadron decay, as shown in Fig. 5.8. At a
 1308 b -jet efficiency of 70%, the light-jet mistag rate for jets with $250 < p_T < 400$ GeV
 1309 is reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For
 1310 jets with $400 < p_T < 1000$ GeV the mistage rate drops from 0.1 to 0.08 for a similar

relative improvement of 20%. The performance of the fake track removal approach was also tested for the other low level vertexing algorithm – JetFitter. A similar level of improvement in the light-jet mistag rate was observed with a reduction of up to a 20% reduction for both low- and high- p_T jets in the Z' sample achieved. 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 by an amount which is comparable to the improvement that would be observed if the tracks were selected at truth level1.

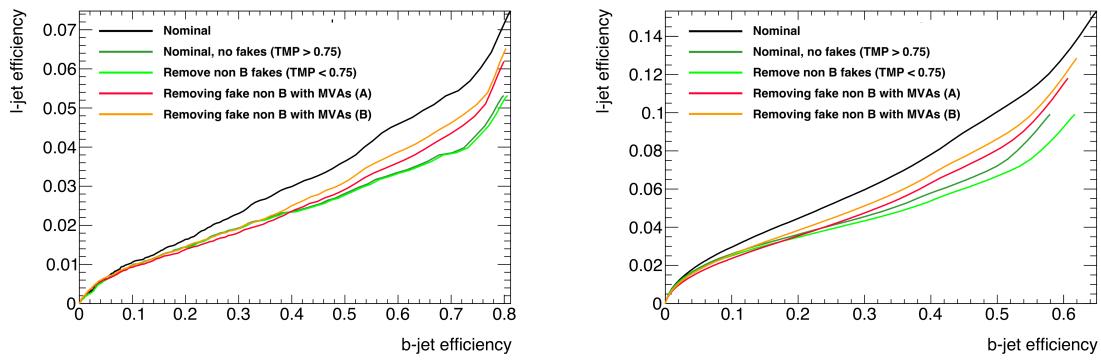


Figure 5.8: The effect of applying the fake track identification algorithm together with the b -hadron decay track identification on the jet tagging performance of SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). 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. An advantage of the approach is that the continuous output of the model allows for the tuning of good and fake track identification efficiencies. 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 that achieved when using the truth information to remove the fake tracks not from the decay of a b -hadron.

Future Work

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

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

The approach here works on a track-by-track basis, but a more sophisticated approach would consider the correlations between the tracks inside a jet. Also left for future work is to simultaneously train a single tool which discriminates between all the truth origins listed in Table 5.1. Such a tool would be useful as a general purpose multiclass classifier. An algorithm which takes both these aspects into consideration is discussed in Chapter 6.

₁₃₅₂ **Chapter 6**

₁₃₅₃ **Graph Neural Network Flavour
Tagger**

₁₃₅₅ As discussed in Chapter 4, flavour tagging is the identification of jets instantiated
₁₃₅₆ from b - and c -hadrons. Flavour tagging is a critical component of the physics
₁₃₅₇ programme of the ATLAS experiment. It is of crucial importance for the study of the
₁₃₅₈ Standard Model (SM) Higgs boson and the top quark, which decay preferentially to
₁₃₅₉ b -quarks [112, 113], and additionally for several Beyond the Standard Model (BSM)
₁₃₆₀ resonances that readily decay to heavy flavour quarks [114].

₁₃₆₁ This chapter introduces GN1, a novel ML-based flavour tagging algorithm based
₁₃₆₂ on graph neural networks (GNNs). In Section 6.1, an overview of the approach
₁₃₆₃ used for GN1 is provided. An introduction to the theory of GNNs is provided in
₁₃₆₄ Section 5.2. Details of the experimental setup are provided in Section 6.2, while
₁₃₆₅ the architecture of GN1 is specified in Section 6.3.3. In Section 6.3.4, the training
₁₃₆₆ procedure is described, and in Section 6.4 the results are shown.

₁₃₆₇ **6.1 Motivation**

₁₃₆₈ GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. As opposed
₁₃₆₉ to the existing approach to flavour tagging described in Chapter 4, which relies
₁₃₇₀ on a two tiered approach requiring the use of both low- and high-level algorithms,
₁₃₇₁ GN1 takes as inputs information directly from an unordered variable number of
₁₃₇₂ 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, 115].

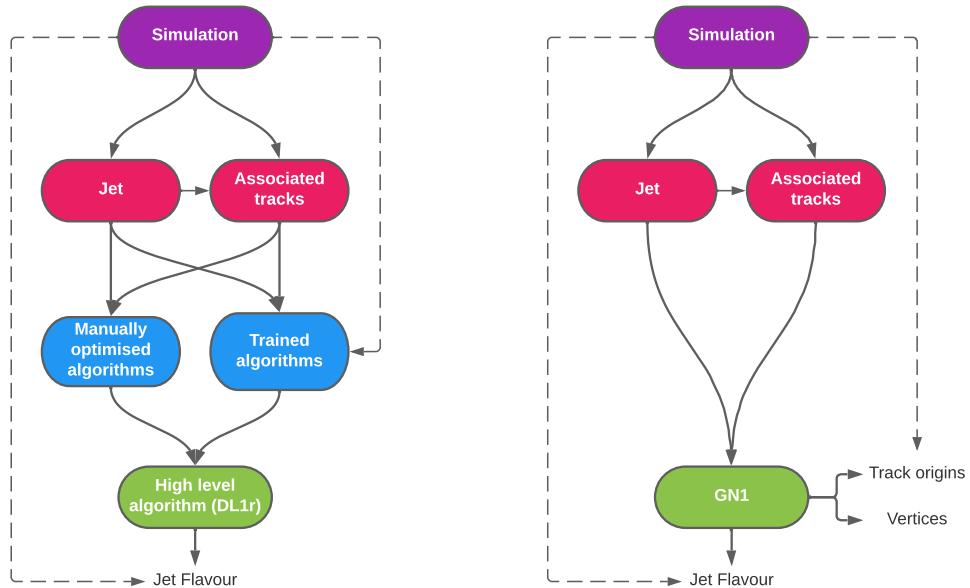


Figure 6.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right) [3]. The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms,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

1387 DL1r. This is particularly important when optimising the tagger for new regions of
1388 phase space (e.g. c -tagging or high- p_T b -tagging), or when the detector is upgraded
1389 or the charged particle reconstruction or selection algorithms are re-optimised.

1390 GN1 is trained to learn about the internal structure of the jet through the use of two
1391 auxiliary training objectives: the prediction of the underlying physics process from
1392 which each track originated, and the grouping of tracks originating from a common
1393 spatial position (i.e. a common vertex). These auxiliary objectives are meant to
1394 guide the neural network towards a more complete understanding of the underlying
1395 physics inside the jet, thereby removing the need for the low-level algorithms, which
1396 previously contained information about the underlying physics in their design. The
1397 training targets for the primary and auxiliary objectives are extracted from truth
1398 information, i.e. information that is only available in simulation, as opposed to
1399 reconstructed quantities available in both collision data and simulation.

1400 In this chapter, the following advantages of the GN1 approach will be demonstrated:

- 1401 1. GN1 boasts improved performance with respect to the current ATLAS flavour
1402 tagging algorithms, with significantly larger background rejection rates for a
1403 given signal efficiency. Alternatively the rejection rates can be kept fixed for a
1404 substantial increase in signal efficiency, in particular at high- p_T .
- 1405 2. The same network architecture can be easily optimised for a wider variety of
1406 use cases (e.g. c -jet tagging and high- p_T jet tagging) since there are no low-level
1407 algorithms to retune.
- 1408 3. There are fewer algorithms to maintain.
- 1409 4. Alongside the network's prediction of the jet flavour, the auxiliary vertex and
1410 track origin predictions provide more information on why a jet was (mis)tagged
1411 or not. This information can also have uses in other applications, for instance
1412 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].

1413 6.2 Experimental Setup

1414 6.2.1 Datasets

1415 The datasets used to train the GN1 tagger are the same as described in Section 5.4.1.
1416 The training dataset contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40%
1417 of which are Z' jets. In order to evaluate the performance of the model during, a
1418 statistically independent set of 500k testing jets from both the $t\bar{t}$ and Z' samples are
1419 used. For the final testing of the model and the creation of the performance plots,
1420 a further 1 million independent testing jets from each of the $t\bar{t}$ and Z' samples are
1421 used. Before being fed into the model, the track- and jet-level inputs are normalised
1422 to have a mean of zero and a variance of unity. The jet flavour labels are assigned
1423 as described in Section 3.4.3. Truth labelled b -, c - and light-jets are kinematically
1424 re-sampled in p_T and η to ensure identical distributions in these variables.

1425 6.3 Model Architecture

1426 6.3.1 Model Inputs

1427 As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40
1428 tracks which have been associated to the jet. Each track consists of 21 variables. The
1429 kinematic jet variables are the jet transverse momentum and signed pseudorapidity.
1430 The input variables which are provided for each track are listed in Table 6.1. For
1431 each track, variables containing the track parameters and uncertainties, and detailed
1432 information on the hit content are provided as inputs to the model.

1433 In cores of high- p_T jets, track density is high due to the increased multiplicity and
1434 collimation of tracks (see Chapter 4). As a result, the separation between tracks
1435 can be of the same order as the active sensor dimensions, resulting in an increase
1436 in merged clusters and tracks which share hits [57]. Due to the relatively long
1437 lifetimes of b -hadrons and c -hadrons, which can traverse several layers of the ID
1438 before decaying and have highly collimated decay products, the presence of shared
1439 or missing hits is a critical signature of heavy flavour jets.

1440 Dependence of the model on the absolute value of the azimuthal jet angle ϕ is
1441 explicitly removed by providing only the azimuthal angle of tracks relative to the jet
1442 axis. The track pseudorapidity is also provided relative to the jet axis.

1443 Since heavy flavour hadrons can decay semileptonically approximately 20% of the time,
1444 the presence of a reconstructed lepton in the jet carries discriminating information
1445 about the jet flavour. To exploit this, a variant of GN1 called GN1Lep is trained in
1446 addition to the baseline model. The GN1Lep variant is identical to the baseline model,
1447 except for the inclusion an additional track-level input, leptonID, which indicates
1448 if the track was used in the reconstruction of an electron, a muon or neither. The
1449 variable is signed by the charge of the reconstructed lepton. The leptons used in the
1450 definition of the leptonID variable are required to satisfy basic quality requirements.
1451 The muons are required to be combined [116], and the electrons are required to pass
1452 the *VeryLoose* likelihood-based identification working point [117].

1453 The selections applied to the tracks is the same as that used for the fake track
1454 classification MVA described in Chapter 5. The full set of track selections is listed
1455 in Table 3.2. This selection was found to improve the flavour tagging performance
1456 compared to previous tighter selections, whilst ensuring good resolution of tracks
1457 and a low fake rate [57]. However, Section 6.4.8 demonstrates that further relaxation
1458 of the track selection requirements may be warranted.

1459 If more than 40 tracks are associated to a given jet, only the first 40 tracks with the
1460 largest transverse IP significance² $s(d_0)$ are fed into the model as inputs.

1461 6.3.2 Auxiliary Training Objectives

1462 In addition to the jet flavour classification, two auxiliary training objectives are
1463 defined. The first auxiliary objective is the prediction of the physical process that
1464 gave rise to each track within the jet (i.e. the track origin), while the second is the
1465 prediction of track-pair vertex compatibility. Each auxiliary training objective comes
1466 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.

1467 from the simulation. The presence of the auxiliary training objectives improves the
1468 jet classification performance as demonstrated in Section 6.4.3.

1469 For the track origin prediction objective, each track is labelled with one of the
1470 exclusive categories defined in Table 5.1 of Section 5.3 after analysing the particle
1471 interaction (or lack thereof) which led to its formation. Since the presence of different
1472 track origins is strongly related to the flavour of the jet, training GN1 to recognise
1473 the origin of the tracks provides an additional handle on the classification of the
1474 jet flavour. This task may also aid the jet flavour prediction by acting as a form of
1475 supervised attention [118] - in detecting tracks from heavy flavour decays the model
1476 may learn to pay more attention to these tracks.

1477 The vertexing auxiliary objective makes use of the fact that displaced decays of b -
1478 and c -hadrons lead to secondary and tertiary vertices inside the jet, as described in
1479 Section 4.1.1. The presence of displaced secondary vertices is not a completely clean
1480 signal of a heavy flavour jet, as displaced secondary vertices can also occur in light-jets
1481 as a result of material interactions, conversions, and long-lived particle decays (e.g.
1482 K_S^0 and Λ^0). For the auxiliary object, GN1 predicts a binary label for each pair of
1483 tracks in the jet. The label has a value of 1 if the truth particles associated with the
1484 two tracks in the pair originated from the same spatial point, and 0 otherwise. To
1485 derive the corresponding truth labels for training, truth production vertices within 0.1
1486 mm are merged. Track-pairs where one or both of the tracks in the pair have an origin
1487 label of either Pile-up or Fake are given a label of 0. Using the pairwise predictions
1488 from the model, groups of tracks that have common compatibility can be formed,
1489 resulting in the finding of vertices. Two existing low-level tagging algorithms, SV1
1490 and JetFitter (introduced in Section 3.4.2), are currently used to find and reconstruct
1491 vertices inside jets and are used as inputs to the existing jet flavour tagger DL1r.
1492 The addition of this auxiliary training objective removes the need for inputs from a
1493 dedicated secondary vertexing algorithm.

1494 Both of the auxiliary training objectives described here can be considered as “stepping
1495 stones” on the way to classifying the flavour of the jet. By requiring the model to
1496 predict the truth origin of each track and the vertex compatibility of each track-pair,
1497 the model is guided to learn representations of the jet which are connected to the
1498 underlying physics and therefore relevant for classifying the jet flavour.

1499 6.3.3 Architecture

1500 As discussed in the previous sections, GN1 is a graph neural network which makes
1501 use of auxiliary training objectives in order to determine the jet flavour. A coarse
1502 optimisation of the network architecture hyperparameters (for example number of
1503 layers and number of neurons per layer) has been carried out in order to maximise
1504 the flavour tagging performance, but it is likely that further dedicated optimisation
1505 studies could lead to further performance improvements.

1506 The model architecture builds on a previous implementation of a GNN-based jet
1507 tagger [115]. The previous approach was comprised of two separate graph neural
1508 networks with the auxiliary tasks being performed at an intermediate stage after the
1509 first and before the second. This two stage approach was found to be unnecessary and
1510 as such GN1 simplifies the architecture into a single graph neural network with the
1511 auxiliary tasks being performed at the end, alongside the primary jet classification
1512 task. GN1 makes use of a more sophisticated graph neural network layer [119],
1513 which is described in more detail below. The changes significantly improved tagging
1514 performance and also led to a significant reduction in training time.

1515 As inputs, the model takes information about the jet and a number of associated
1516 tracks, as detailed in Section 6.3.1. The jet variables are concatenated with the
1517 variables for each track as shown in Fig. 6.2. The combined jet-track input vectors
1518 are then fed into a per-track initialisation network with three hidden layers, each
1519 containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.3. The
1520 track initialisation network is similar to a deep sets model [99], but does not include
1521 a reduction operation (mean or summation) over the output track representations.
1522 The initialisation network allows for initial per-track input processing without the
1523 associated parameter count cost of the graph convolutional layers described below.

1524 The outputs of the track initialisation network are used to populate the nodes of a
1525 fully connected graph, such that each node in the graph neighbours every other node.
1526 Each node h_i in the graph corresponds to a single track in the jet, and is characterised
1527 by a feature vector, also called a representation. The per-track output representations
1528 from the initialisation networks are used as the initial feature vectors of each node
1529 in the graph. In each layer of the graph network, output node representations h'_i
1530 are computed by aggregating the features of h_i and neighbouring nodes \mathcal{N}_i using
1531 a multi-head attention mechanism ($n = 2$) as described in Ref. [119, 120]. First,

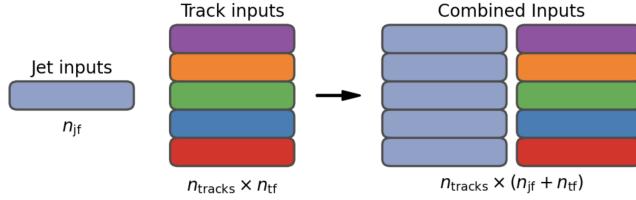


Figure 6.2: The inputs to GN1 are the two jet features ($n_{jf} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{tf} = 21$) [3]. The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1.

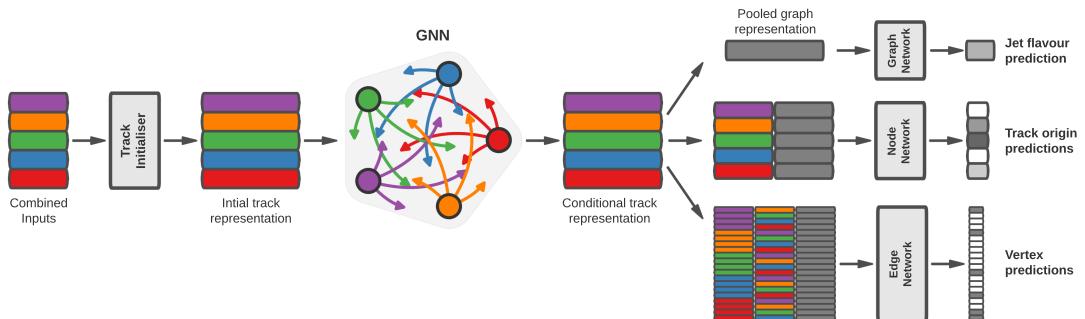


Figure 6.3: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting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

1532 the feature vectors of receiver and sender nodes are fed into two fully connected
 1533 linear layers \mathbf{W}_r and \mathbf{W}_s , to produce an updated representation for each sender and
 1534 receiver node $\mathbf{W}_r h_i$ and $\mathbf{W}_s h_j$. These updated feature vectors are used to compute
 1535 edge scores $e(h_i, h_j)$ for each node pair, as in

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

1536 where, θ is a non-linear activation function, and \mathbf{a} is a learned vector. These edge
 1537 scores are then used to calculate attention weights a_{ij} for each pair of nodes using
 1538 the softmax function over the edge scores

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

1539 Finally, the updated representations for the receiver nodes h'_i are computed by taking
 1540 the weighted sum over each updated node representation $\mathbf{W}_r h_i$, with weights a_{ij}

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

1541 The set of operations described above constitute a single graph network layer. Three
 1542 such layers are stacked to construct the graph network, representing a balance
 1543 between achieving good performance in a reasonable time and avoiding overtraining
 1544 due to inflation of the parameter count of the model. The final output from the graph
 1545 neural network is a set of per-node (i.e. per-track) feature vectors that are conditional
 1546 representations of each track given the other tracks in the jet. In order to perform
 1547 the jet flavour prediction, a flattened global representation of the jet is needed. To
 1548 produce this, the output track representations are combined using a weighted sum,
 1549 where the weights are learned during training and therefore act as a form of attention
 1550 over the different tracks. The flattened outputs from the sum are then fed into a
 1551 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.

6.3.4 Training

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

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

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

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

1603 useful indication of the benefit of including the auxiliary tasks to the primary jet
1604 classification objective.

1605 GN1 was trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes
1606 approximately 25 mins to complete over the training sample of 30 million jets
1607 described in Section 6.2.1. The Adam optimiser [121] with an initial learning rate of
1608 $1e-3$, and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically
1609 the validation loss, calculated on 500k jets, became stable after around 60 epochs.
1610 The epoch that minimized the validation loss was used for evaluation. GN1 has
1611 been integrated into the ATLAS software [54] using ONNX [122]. The test sample
1612 jet flavour predictions scores are computed using the ATLAS software stack as a
1613 verification of this process.

1614 6.4 Results

1615 The GN1 tagger is evaluated both as a b -tagging and c -tagging algorithm in Sec-
1616 tion 6.4.1 and Section 6.4.2 respectively. Evaluation is performed separately on
1617 both jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV and jets in the Z' sample with
1618 $250 < p_T < 5000$ GeV. The performance of the model is compared to the DL1r
1619 tagger [83, 123], which has been retrained on 75 million jets from the same samples
1620 as GN1. The input RNNIP tagger [85] to DL1r has not been retrained. As discussed,
1621 each tagger predicts the probability that a jet belongs to the b -, c - and light-classes.
1622 To use the model for b -tagging, these probabilities are combined into a single score
1623 D_b , which is defined as

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

1624 where f_c is a free parameter that determines the relative weight of p_c to p_l in the
1625 score D_b , controlling the trade-off between c - and light-jet rejection performance.
1626 The choice of f_c is arbitrary, and is optimised based upon the desired light- vs c -jet
1627 rejection performance. This parameter is set to a value of $f_c = 0.018$ for the DL1r
1628 model, obtained through an optimisation procedure described in Ref. [83]. Based on
1629 a similar optimisation procedure, a value of $f_c = 0.05$ is used for the GN1 models.

1630 A fixed-cut working point (WP) defines the corresponding selection applied to the
 1631 tagging discriminant D_b in order to achieve a given efficiency on the inclusive $t\bar{t}$
 1632 sample.

1633 A comparison of the b -tagging discriminant D_b between DL1r and GN1 is shown in
 1634 Fig. 6.4. The shapes of the D_b distributions are generally similar for b -, c - and light-
 1635 jets between both models, however, GN1 shifts the b -jet distribution to higher values
 1636 of D_b in the regions with the greatest discrimination. The GN1 c -jet distribution is
 1637 also shifted to lower values of D_b when compared with DL1r, enhancing the separation
 1638 and indicating that GN1 is improving c -jet rejection when compared with DL1r.

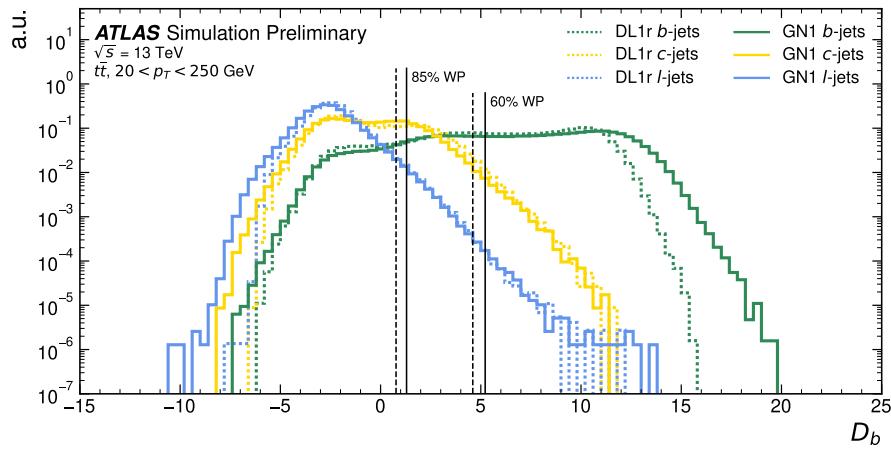


Figure 6.4: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample [3]. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of $f_c = 0.018$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

1639 6.4.1 b -tagging Performance

1640 The performance of b -tagging algorithms is quantified by their ability to reject c - and
 1641 light-jets for a given b -jet selection efficiency WP. In order to compare the b -tagging
 1642 performance of the different taggers for the b -jet tagging efficiencies in the range
 1643 typically used by analyses, the corresponding c - and light-jet rejection rates are
 1644 displayed in Figs. 6.5 and 6.6 for jets in the $t\bar{t}$ and Z' samples respectively. Four
 1645 standard WPs are defined with b -jet tagging efficiencies of 60%, 70%, 77% and 85%

1646 respectively. These WPs are commonly used by physics analyses depending on their
1647 specific signal and background requirements. The WPs are defined based on jets
1648 in the $t\bar{t}$ sample only. Due to the much higher jet p_T range in the Z' sample, and
1649 the increased difficulty in tagging jets at high- p_T (see Chapter 4), the b -jet tagging
1650 efficiencies for jets in the Z' sample are lower than the corresponding WPs calculated
1651 in the $t\bar{t}$ sample. For instance the WP cut value computed to provide a 70% b -jet
1652 tagging efficiency on the $t\bar{t}$ sample results in a b -jet tagging efficiency of just \sim 30%
1653 on the Z' sample. In order to account for this, the range of b -jet tagging efficiencies
1654 displayed for plots showing the performance for jets in the Z' sample (for example
1655 Fig. 6.6) is chosen to span the lower efficiencies achieved in the Z' sample at high- p_T .

1656 For jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV, GN1 demonstrates considerably
1657 better c - and light-jet rejection when compared with DL1r across the full range of
1658 b -jet tagging efficiencies studied. The relative improvement is strongly dependent
1659 on the b -jet tagging efficiency under study. The largest improvements are found at
1660 lower b -jet tagging efficiencies. At a b -jet tagging efficiency of 70%, the c -jet rejection
1661 improves by a factor of \sim 2.1 while the light-jet rejection improves by a factor of \sim 1.8
1662 with respect to DL1r. For high- p_T jets in the Z' sample with $250 < p_T < 5000$ GeV,
1663 GN1 also brings a significant performance improvement with respect to DL1r across
1664 the range of b -jet tagging efficiencies studied. Again, the largest relative improvement
1665 in performance comes at the lower b -jet tagging efficiencies. At a b -jet efficiency of
1666 30%, GN1 improves the c -jet rejection with respect to DL1r by a factor of \sim 2.8 and
1667 the light-jet rejection by a factor of \sim 6. The performance comparison at lower b -jet
1668 tagging efficiencies is made more difficult due to the increased statistical uncertainties
1669 which result from the high rejection of background. Given that GN1 exploits the
1670 low-level detector information in a more complete and sophisticated way than DL1r,
1671 further studies are needed to confirm that the performance gain observed in these
1672 simulated samples is also observed in experimental data.

1673 The GN1Lep variant of GN1 demonstrates further improved performance with respect
1674 to the baseline model. This demonstrates the additional jet flavour discrimination
1675 power provided by the leptonID track input. For jets in the $t\bar{t}$ sample, the relative c -
1676 jet rejection improvement with respect to GN1 at the 70% b -jet WP is approximately
1677 25%. The improvement in light-jet rejection also increases by 40% at the same WP.
1678 For jets in the Z' sample, the relative c -jet rejection (light-jet rejection) performance

1679 with respect to GN1 improves by approximately 10% (25%) at a b -jet tagging
1680 efficiency of 30%.

1681 In general, the performance of all the taggers is strongly dependent on the jet p_T .
1682 This is due to the increased multiplicity and collimation of tracks, and the displaced
1683 decays that result from within the heavy flavour jets (see Chapter 4). Together,
1684 they contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a
1685 general degradation in quality of tracks inside the core of a jet, which in turn reduces
1686 the jet tagging performance.

1687 In order to study how the tagging performance changes as a function of the jet p_T ,
1688 the b -jet tagging efficiency as a function of p_T for a fixed light-jet rejection of 100 in
1689 each bin is shown in Fig. 6.7. For jets in the $t\bar{t}$ sample, at a fixed light-jet rejection
1690 of 100, GN1 improves the b -jet tagging efficiency by approximately 4% across all the
1691 jet p_T bins. Meanwhile, GN1Lep again demonstrates improved performance with
1692 respect to GN1, in particular at lower p_T . The relative increase in the b -jet tagging
1693 efficiency increases from 4% to 8% with respect to DL1r. For jets in the Z' sample,
1694 GN1 again outperforms DL1r across the entire jet p_T range studied. The largest
1695 relative improvement in performance is found at the highest transverse momenta
1696 of jet $p_T > 2 \text{ TeV}$, and corresponds to an approximate factor of 2 improvement in
1697 efficiency with respect to DL1r.

1698 The performance of the model was also evaluated as a function of the average
1699 number of pile-up interactions in the event. No significant dependence of the tagging
1700 performance was observed.

1701 6.4.2 c -tagging Performance

1702 As discussed previously, GN1 does not rely on any inputs from manually optimised
1703 low-level tagging algorithms. Since these algorithms were originally designed and
1704 tuned with the aim of b -tagging, and not c -tagging, the low level tagging algorithms
1705 may perform suboptimally for c -tagging purposes. The tagging of c -jets therefore
1706 presents a compelling use case for GN1. As each of the the models is trained with
1707 three output classes, using it as a c -tagging algorithm is trivially analogous to the
1708 approach used for b -tagging. The model output probabilities are combined into a
1709 single score D_c , which is defined similarly to Eq. (6.5) as

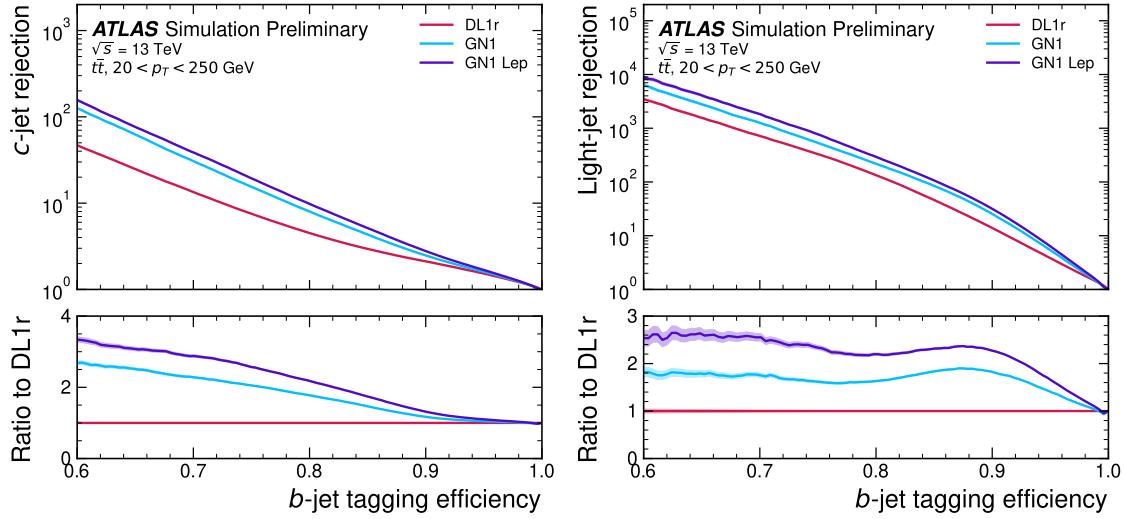


Figure 6.5: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ [3]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$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

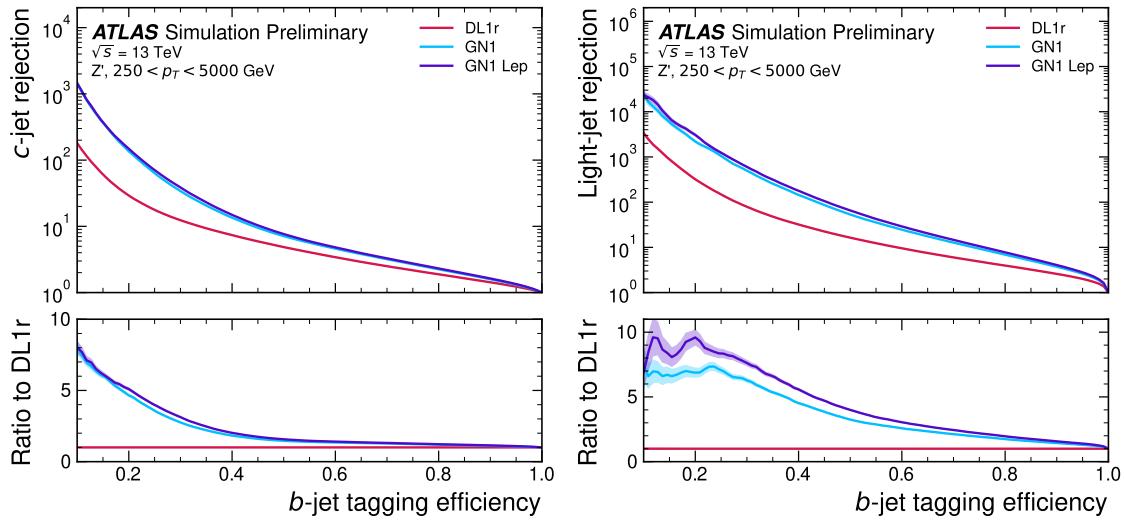


Figure 6.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$ [3]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used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

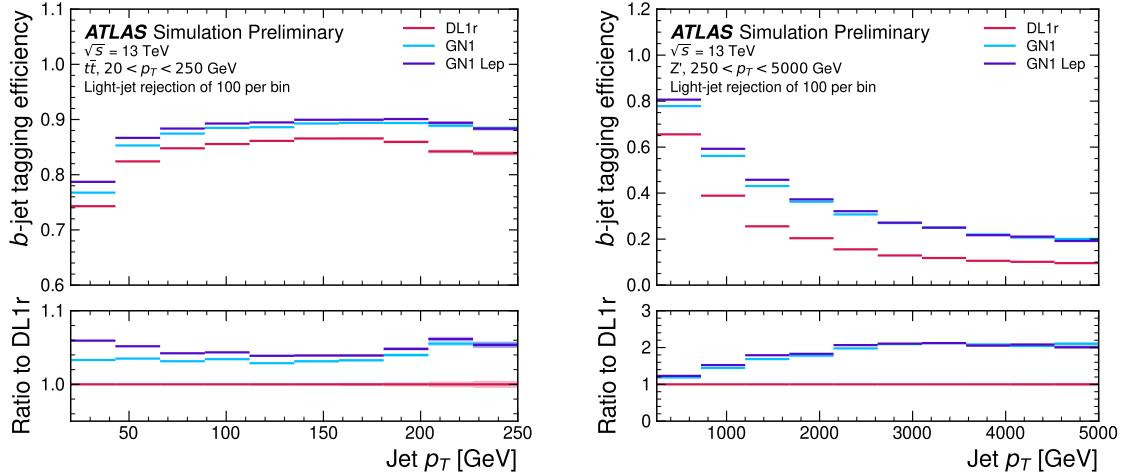


Figure 6.7: The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin [3]. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and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

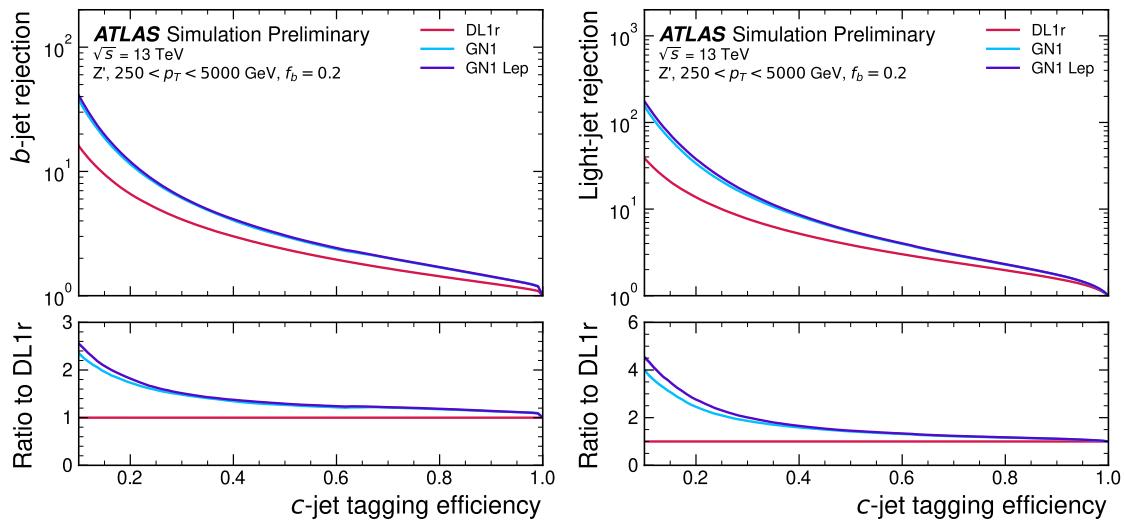
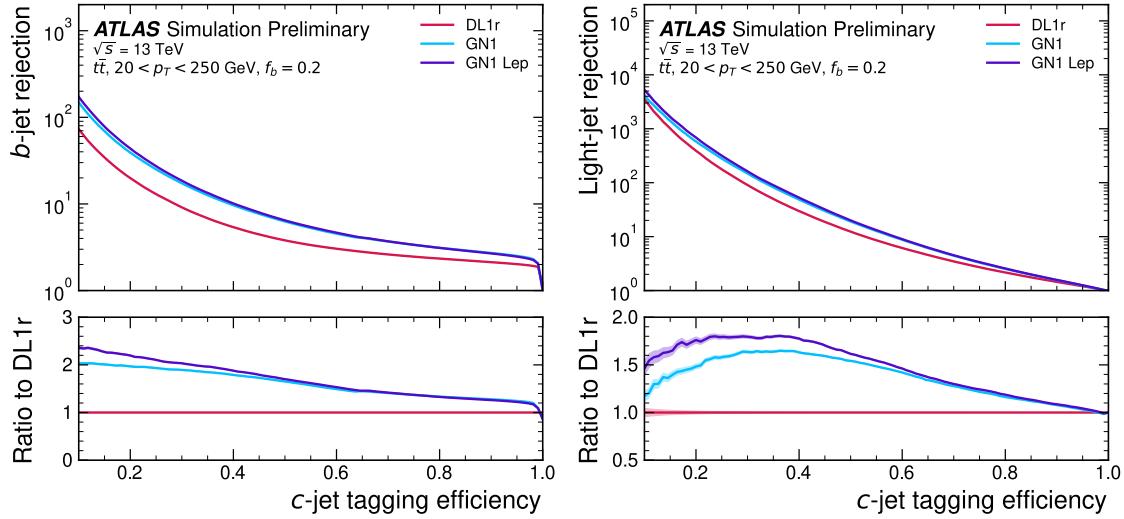
1724 case, with the both the b -jet rejection (light-jet rejection) improving with respect to
1725 GN1 by approximately 10% (20%) at the 25% c -jet WP. Fig. 6.9 shows the c -tagging
1726 performance on the jets in the Z' sample with $250 < p_T < 5000$ GeV. Both GN1 and
1727 GN1Lep perform similarly, improving the b -jet rejection by 60% and the light-jet
1728 rejection by a factor of 2 at the 25% c -jet WP.

1729 6.4.3 Ablations

1730 Ablation studies (the removal of certain components of a given model in order to
1731 study the impact of that component) are carried are carried out to determine the
1732 importance of the auxiliary training objectives of GN1 to the overall performance.
1733 The “GN1 No Aux” variant retains the primary jet classification objective, but
1734 removes both track classification and vertexing auxiliary objectives (see Section 6.3.2)
1735 and correspondingly only minimises the jet classification loss. The “GN1 TC” variant
1736 includes track classification objective but not the vertexing objective. Finally, the
1737 “GN1 Vert” includes the vertexing objective, but not the track classification objective.

1738 For jets in both the $t\bar{t}$ and Z' samples, a general trend is observed that the models
1739 trained without one or both of the auxiliary objectives results in significantly reduced
1740 c - and light-jet rejection when compared with the baseline GN1 model. This result
1741 is shown clearly in Figs. 6.10 and 6.11. For jets in the $t\bar{t}$ sample, the performance of
1742 GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert perform similarly to
1743 each other. For jets in the Z' sample meanwhile, the GN1 No Aux model already
1744 shows a clear improvement in c - and light-jet rejection when compared with DL1r at
1745 lower b -jet tagging efficiencies. Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1
1746 Vert perform similarly, and bring large gains in background rejection when compared
1747 with GN1 No Aux, but the combination of both auxiliary objectives yields the best
1748 performance.

1749 It is notable that the GN1 No Aux model matches or exceeds the performance of
1750 DL1r without the need for inputs from the low-level algorithms. This indicates that
1751 the performance improvements enabled by the improved neural network architecture
1752 used in GN1 appear to be able to compensate for the removal of the low-level
1753 algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform
1754 DL1r, demonstrating that both contribute to the overall high performance of the
1755 baseline model. The overall best performing model is the full version of GN1 trained



1756 with both auxiliary objective, demonstrating that the two auxiliary objectives are
1757 complementary.

1758 6.4.4 Inclusion of Low-Level Vertexing Algorithms

1759 As already mentioned, GN1 does not include any inputs from the low-level tagging
1760 algorithms, including the vertexing algorithms SV1 and JetFitter [61]. Since these
1761 algorithms are known to play a key role in contributing to the performance of DL1r, it
1762 was necessary to study whether their inclusion in GN1 resulted in further performance
1763 improvements. In a dedicated training of GN1 the SV1 and JetFitter tagger outputs
1764 were added to the GN1 jet classification network as an input, similar to how they
1765 are used in DL1r. These outputs include information on the reconstructed vertices,
1766 including the number of vertices, and the properties of the highest ranking recon-
1767 structed vertex (in the case of JetFitter). In addition, the index of the reconstructed
1768 SV1 or JetFitter vertices were included as two track-level inputs to GN1. These
1769 indices were also used to construct an input feature for the edge classification
1770 network used to identify vertices, which was given a value of one if the track-pair
1771 were from a common reconstructed SV1 or JetFitter vertex, and zero otherwise. The
1772 jet classification performance of this GN1 model was not significantly different to the
1773 baseline model, and in some cases the performance was slightly reduced. GN1 does
1774 not benefit from the inclusion of information from SV1 and JetFitter, indicating
1775 that the model is able to reconstruct the relevant information provided by these
1776 low-level algorithms. The study also demonstrates that the model can function as a
1777 highly performant standalone tagger that does not require (beyond retraining) any
1778 manual optimisation to achieve good performance in a wide range of phase spaces.
1779 A dedicated look at the vertexing performance of GN1 with some comparisons to
1780 SV1 and JetFitter is found in Section 6.4.6

determine ranking

1781 6.4.5 Jet Display Diagrams

1782 The auxiliary training objectives of GN1 allow for improved model interpretability,
1783 which is especially important for a monolithic approach as the low level taggers,
1784 which provide useful physical insight, are no longer present. Figs. 6.12 and 6.13
1785 provide example comparisons of the true origin and vertexing information compared

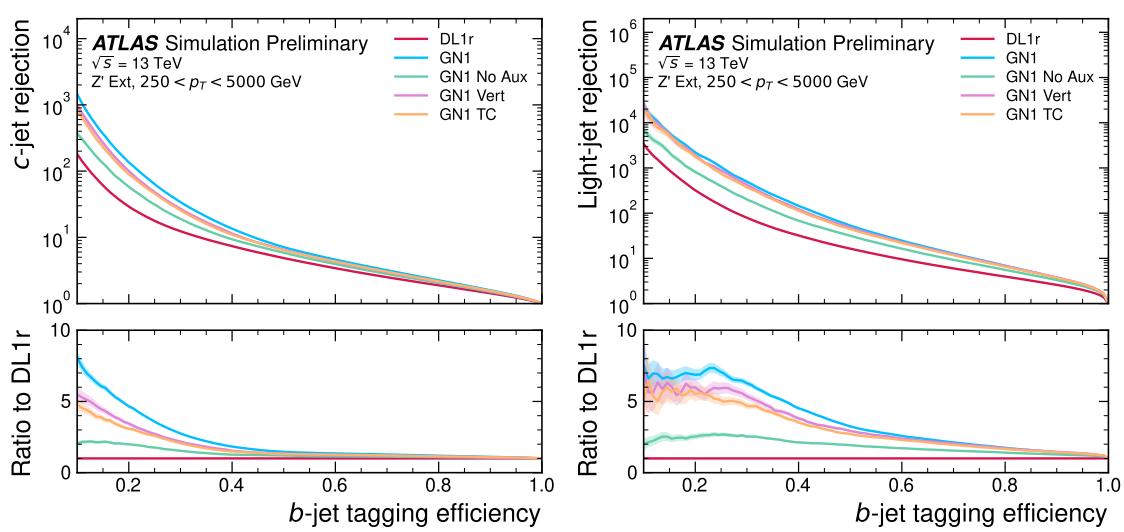
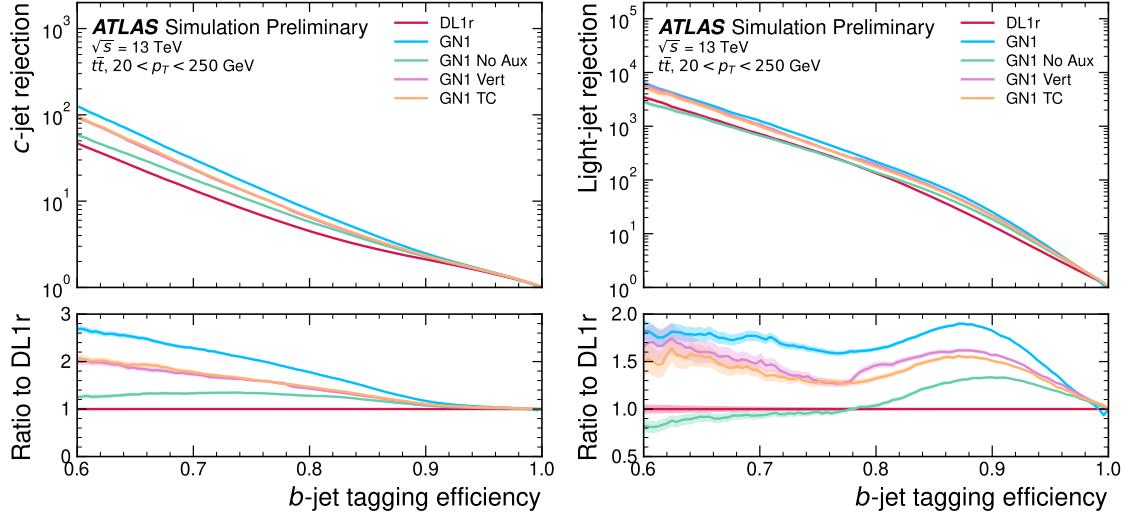


Figure 6.11: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary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

1786 with the predicted values from GN1, SV1 and JetFitter. Such comparisons can be
 1787 used to provide an indication that GN1 reconstructs the correct representation of the
 1788 jet structure, and may also help to identify limitations of the model. In the figures,
 1789 the tracks in the jet are indexed twice on each of the x - and y -axes, and tracks are
 1790 grouped into vertices along with other tracks as indicated by common markings in
 1791 the relevant rows and columns.

1792 In Fig. 6.12, GN1 correctly groups the three primary tracks as having come from the
 1793 primary vertex. The b -hadron and $b \rightarrow c$ -hadron decay vertices are also correctly
 1794 predicted, and the origin of the tracks in each is correct. There is a single OtherSec-
 1795 ondary track which GN1 incorrectly predicts as having come pile-up. Meanwhile SV1
 1796 (by design) merges the two heavy flavour decay vertices, but incorrectly includes a
 1797 track from the primary vertex. JetFitter reconstructs two vertices, one which is a
 1798 combination of two tracks from different truth vertices and two other single track
 1799 vertices in each of the heavy flavour vertices. GN1 also predicts the flavour of the jet
 1800 with a high degree of certainty.

1801 Similarly, Fig. 6.12 shows that GN1 is able to relatively accurately predict the origin
 1802 and vertex information of tracks inside a jet. The pile-up tracks and primary vertex
 1803 tracks are correctly identified, and the heavy flavour decay tracks are also correctly
 1804 identified with the exception of one of the b -hadron decay tracks. Again, SV1 merges
 1805 the two heavy flavour decay vertices along with a track from pile-up, while JetFitter
 1806 shows signs of being underconstrained by reconstructing two single track vertices,
 1807 one with a pile-up track and one with a track from a $b \rightarrow c$ -hadron decay.

1808 6.4.6 Vertexing Performance

1809 From the track-pair vertex prediction described in Section 6.3.2, tracks can be
 1810 partitioned into compatible groups representing vertices (see [115]). As such, GN1
 1811 can perform vertex “finding”, but not vertex “fitting”, i.e. the reconstruction of
 1812 a vertex’s properties, which currently still requires the use of a dedicated vertex
 1813 fitter. In order to study the performance of the different vertexing tools, the truth
 1814 vertex label of the tracks, discussed in Section 6.3.2, are used. To estimate the
 1815 efficiency with which GN1 manages to find vertices inclusively, vertices containing
 1816 tracks identified as coming from a b -hadron are merged together and compared to the
 1817 inclusive truth decay vertices that result from a b -hadron decay (where if there are

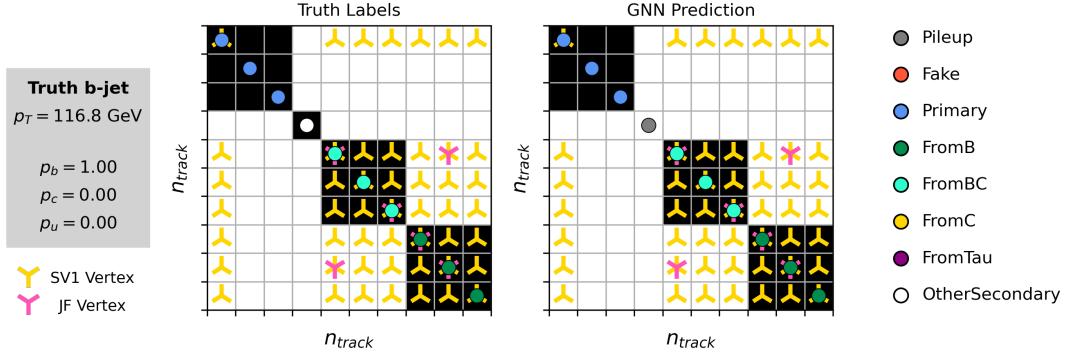


Figure 6.12: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed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

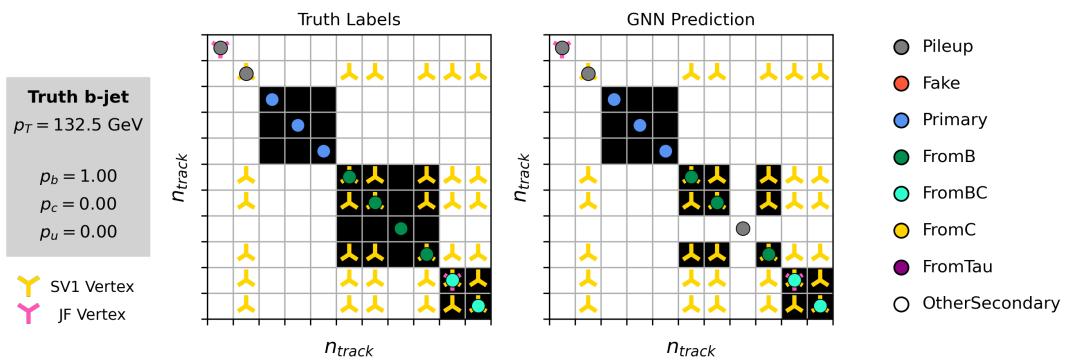


Figure 6.13: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed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

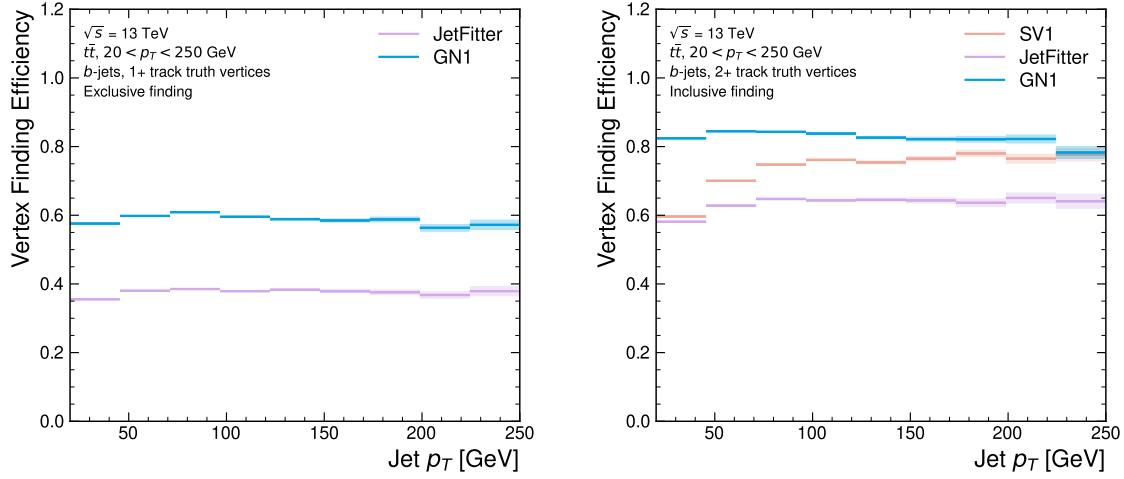


Figure 6.14: Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices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

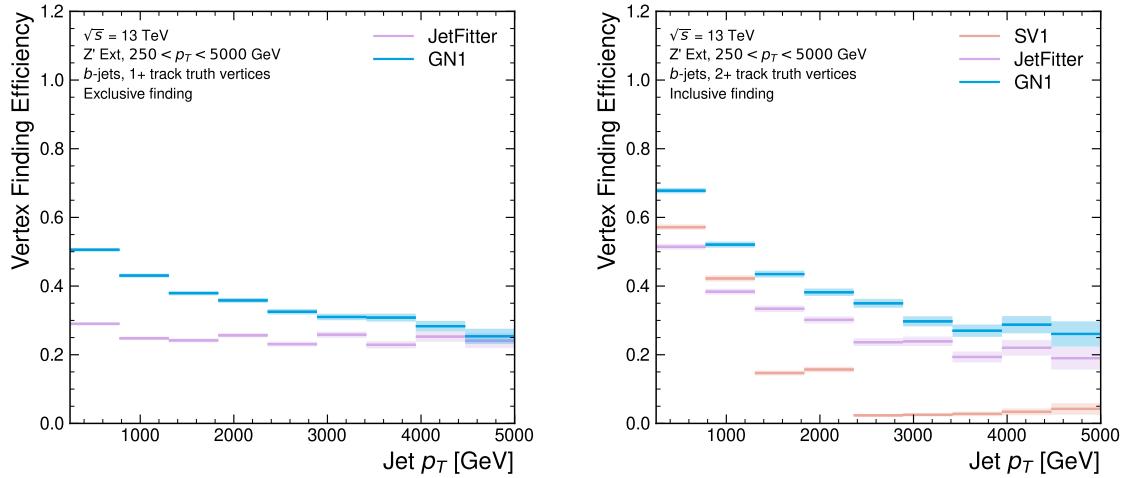


Figure 6.15: Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in the Z' sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain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

1883 While Figs. 6.14 and 6.15 indicate that GN1 is able to successfully find displaced heavy
 1884 flavour vertices in b -jets, it is also important to consider the vertexing performance
 1885 inside light-jets. light-jets may also contain displaced vertices due to long lived
 1886 secondary particles and material interactions. These tracks have a truth origin of
 1887 OtherSecondary in the truth labelling scheme enumerated in see Table 5.1. The
 1888 efficiency to reconstruct vertices comprised of OtherSecondary tracks can be computed
 1889 in an analogous way to the heavy flavour vertexing efficiency, which is described above.
 1890 Figs. 6.16 and 6.17 show the efficiency to reconstruct displaced OtherSecondary
 1891 vertices in light-jets as a function of p_T for jets in the $t\bar{t}$ sample and jets in the Z'
 1892 sample respectively. The figures demonstrate that GN1 is able to more effectively find
 1893 such vertices in light-jets as compared with SV1 and JetFitter. Since the properties
 1894 of the displaced vertices in light-jets are likely to be significantly different to heavy
 1895 flavour vertices found in heavy flavour jets, the improved reconstruction of such
 1896 vertices may help to differentiate between different flavour of jet.

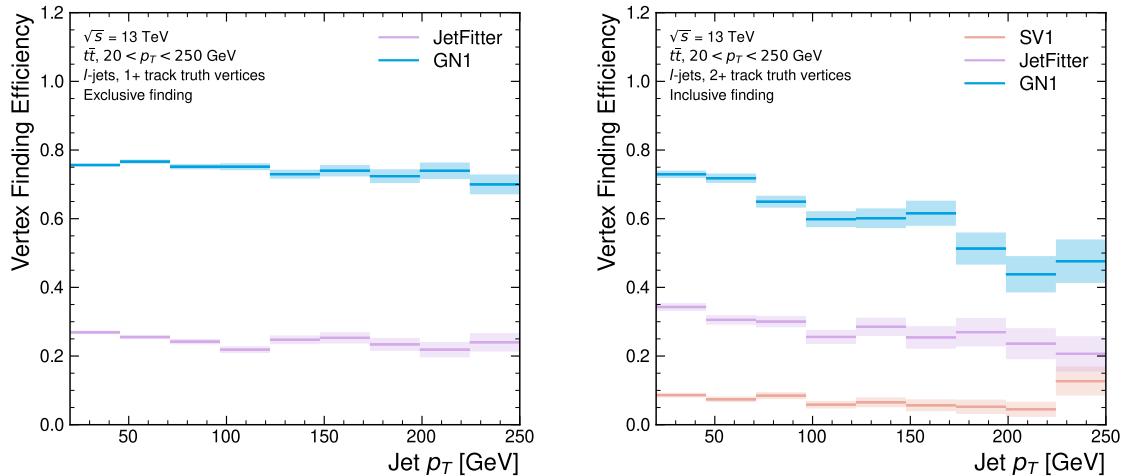


Figure 6.16: Vertex finding efficiency as a function of jet p_T for light-jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain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

1897 Collectively, the results in this section demonstrate that GN1 is able to accurately
 1898 group tracks by their spatial origin in both b -jets and light-jets. The purity of
 1899 the found vertices was also investigated and was found to be comparable or better
 1900 than that of SV1 and JetFitter. Using a vertex fitting algorithm to compare the

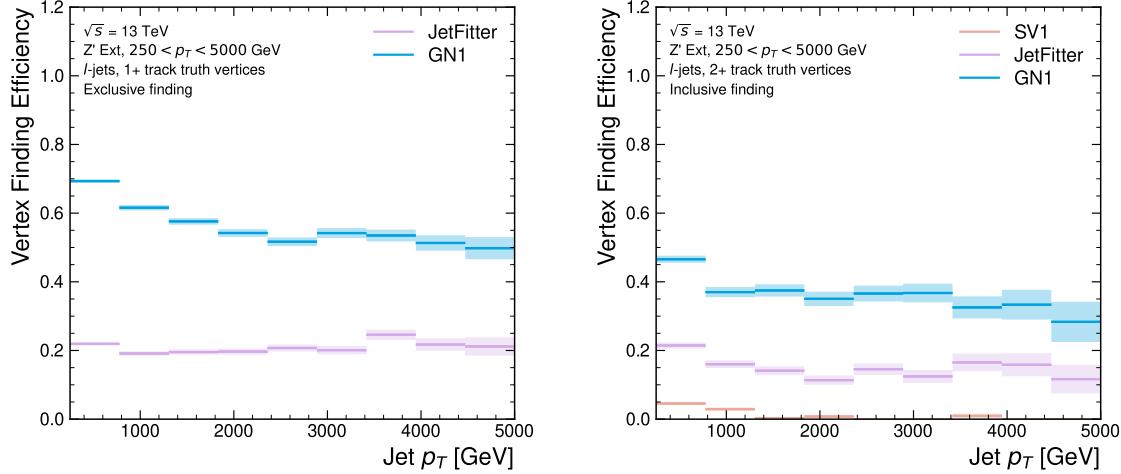


Figure 6.17: Vertex finding efficiency as a function of jet p_T for light-jets in the Z' sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least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

1901 reconstructed vertex quantities with those from SV1 and JetFitter is left for future
 1902 work.

1903 6.4.7 Track Classification Performance

1904 One of the two auxiliary training objectives used by GN1 is to predict the truth origin
 1905 of each track associated to the jet, as discussed in Section 6.3.2. Since the equivalent
 1906 information is not provided by any of the existing flavour tagging tools, a benchmark
 1907 model used to predict the truth origin of each track is trained based on a standard
 1908 multi-class feed-forward classification network. The benchmark model is trained
 1909 on the same tracks used for the baseline GN1 training. The model uses precisely
 1910 the same concatenated track-and-jet inputs as used by GN1 (see Section 6.3.1),
 1911 but processes only a single track at a time, meaning it cannot take into account
 1912 the correlations between tracks when determining the track origin. The model is
 1913 made up of five densely connected linear layers with 200 neurons in each layer. The
 1914 performance of the model was found to be unsensitive to changes in the network
 1915 structure.

1916 To measure the track classification performance, the area under the curve (AUC)
 1917 of the receiver operating characteristic (ROC) curve is computed for each origin
 1918 class, using a one-versus-all classification approach. The AUCs for the different
 1919 truth origins are averaged using both an unweighted and a weighted mean. The
 1920 unweighted mean treats the performance of each class equally, while the weighted
 1921 mean uses as a weight the relative abundance of tracks of each class. Table 6.3
 1922 demonstrates clearly that GN1 outperforms the MLP both at $20 < p_T < 250 \text{ GeV}$
 1923 for jets in the $t\bar{t}$ sample and at $250 < p_T < 5000 \text{ GeV}$ for jets in the Z' sample. For
 1924 example, GN1 can reject 65% of fake tracks in jets in the $t\bar{t}$ sample, while retaining
 1925 more than 99% of good tracks (i.e. those tracks which are not fake). The GN1 model
 1926 has two advantages over the MLP which can explain the performance improvement.
 1927 Firstly, the graph neural network architecture enables the sharing of information
 1928 between tracks as discussed in Section 6.3.3. This is likely to be beneficial since the
 1929 origins of different tracks within a jet are correlated. Secondly, the jet classification
 1930 and vertexing objectives may be complementary to the track classification objective,
 1931 and so the track classification performance is improved by the combined training of
 1932 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.

1933 Fig. 6.18 shows the track origin classification ROC curves for the different track
 1934 origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve visual readability
 1935 of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and
 1936 FromC) have been combined (weighted by their relative abundance), as have the
 1937 Primary and OtherSecondary origins. In jets in both the $t\bar{t}$ and Z' samples, the AUC

1938 of all the different origin groups exceeds 0.9, representing strong overall classification
 1939 performance. In both samples fake tracks are the easiest to classify, followed by
 1940 pile-up tracks. The FromC tracks which are c -hadron decay products, are the hardest
 1941 to classify, possibly due to their similarity to both fragmentation tracks and b -hadron
 1942 decay tracks, depending on the c -hadron species in question.

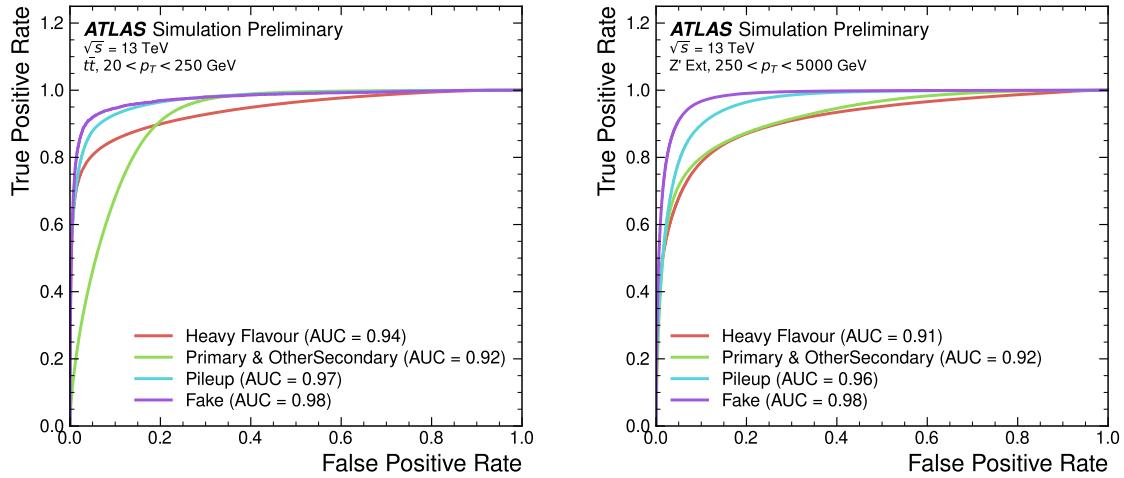


Figure 6.18: ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) [3]. The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and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

1943 6.4.8 Looser Track Selection

1944 The track selections used to produce the main results are listed in Table 3.2. This
 1945 selection includes a cut on the number of shared silicon modules used to reconstruct
 1946 the track $N_{\text{shared}}^{\text{Si}}$. This value is calculated as

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

1947 where $N_{\text{shared}}^{\text{Pix}}$ is the number of shared pixel hits and $N_{\text{shared}}^{\text{SCT}}$ is the number of shared
 1948 SCT modules on a track. The nominal cut used elsewhere in this thesis is $N_{\text{shared}}^{\text{Si}} < 2$.
 1949 As the rate of shared hits is significantly higher for b -hadron decay tracks than for

1950 other tracks, especially at high- p_T , this cut rejects a significant proportion of these
1951 tracks.

1952 Figs. 6.19 and 6.20 show the result of training the GN1 tagger with the full relaxation
1953 of this cut, i.e. allowing tracks with any number of shared hits. The shared hit
1954 requirements applied by the ambiguity solver as part of track reconstruction (see
1955 Section 3.4.1) are still applied. In addition, the maximum allowed value of d_0 is
1956 increased from 3.5 mm to 5.0 mm. The results show that optimisation of the input
1957 track selection can lead to significant improvements in performance over the default
1958 selection. For the jets in the $t\bar{t}$ sample shown in Fig. 6.19, the effect of loosening
1959 the track selection is limited. This is expected due to the lower prevalence of shared
1960 hits at highly displaced tracks at lower transverse momenta. However for jets in the
1961 Z' sample as shown in Fig. 6.20, the light-jet rejection improves with respect to the
1962 baseline GN1 model by 30%, while the light-jet rejection improves by 70% at the
1963 50% b -jet WP.

1964 Although the results demonstrate a significant performance improvement at high- p_T ,
1965 it is also possible that additional studies on further loosening the selection could yield
1966 further improved results. For example the selections on the number of number of
1967 holes and the longitudinal impact parameter could be further relaxed. The maximum
1968 number of tracks provided as input to the model could also be increased from the
1969 default value of 40. In order to change the default tracking setup, studies investigating
1970 the modelling uncertainties of the additional tracks need to be carried out.

1971 6.5 Other Implementations of GN1

1972 The implementation of GN1 described in this chapter has been re-used in several
1973 other contexts, demonstrating its flexibility to easily provide good jet flavour tagging
1974 performance with minimal overhead. The model has been implemented as a b -jet
1975 tagger in the High Level Trigger (HLT) (see Section 3.3.4). The inputs to the
1976 model are the running on precision tracks and jet level quantities reconstructed after
1977 primary vertexing. Fig. 6.21 shows the performance of GN1 versus a comparable
1978 DL1d model [123], and two versions of DIPS [86], with EMTopo and PFlow jets (see
1979 Section 3.4.3) based on a low-precision region-of-interest based tracking pass. The

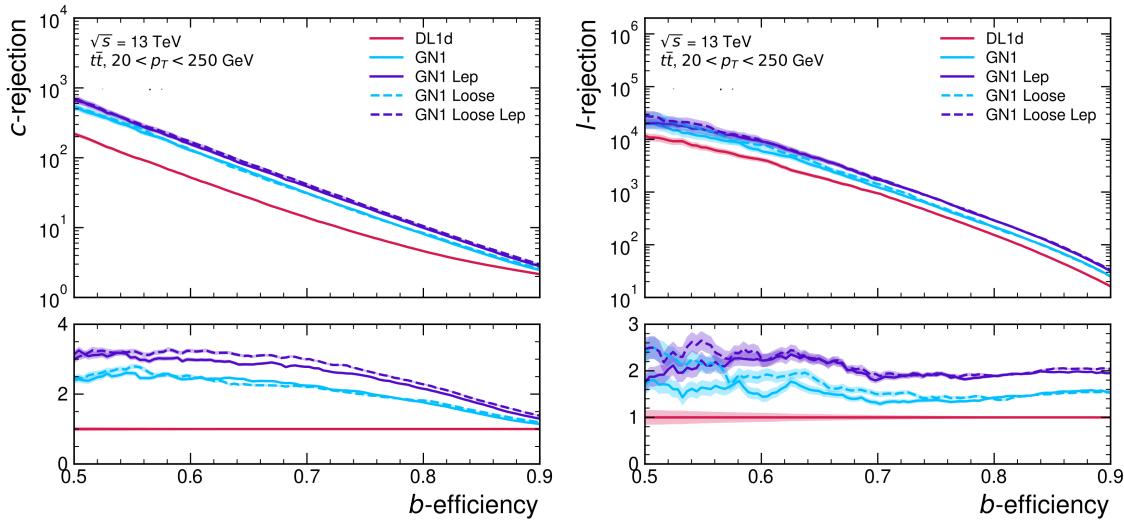


Figure 6.19: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [123] is shown in the bottom panels. A value 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.

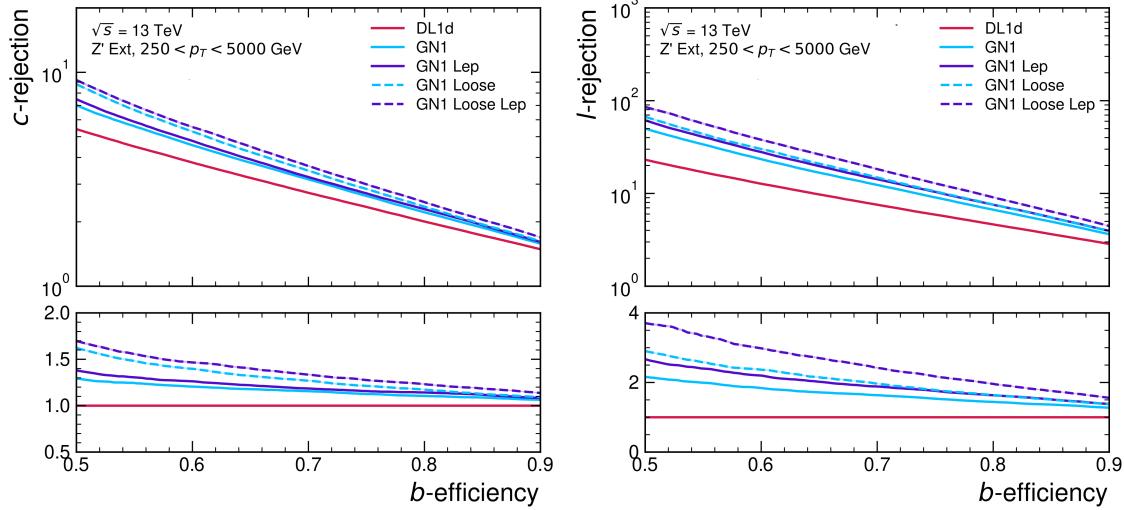


Figure 6.20: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [123] is shown in the bottom panels. A value of $f_c = 0.018$ 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.

1980 trigger implementation of GN1 improves upon the light-jet rejection of DL1d by 50%
1981 at the 60% b -jet WP for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV.

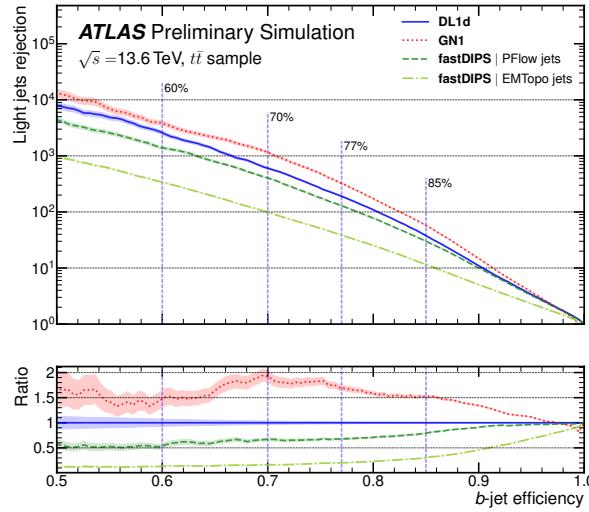


Figure 6.21: The light-jet rejection as a function of the b -jet efficiency jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 13.6$ TeV [124]. The ratio to the performance of the DL1d algorithm [123] is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b -tagging.

1982 The model also demonstrates strong performance for the High Luminosity LHC
1983 (HL-LHC), as documented in Ref. [125]. Figs. 6.22 and 6.23 are reproduced from
1984 Ref. [125]. The results show that GN1 outperforms other existing flavour tagging
1985 algorithms when trained on an entirely different detector geometry, the ITk (see
1986 Section 3.3.1). When compared with DL1d [123], GN1 improves on the c -jet rejection
1987 (light-jet rejection) by a factor of ~ 2 (~ 2.5) for jets in the $t\bar{t}$ sample at the 60%
1988 b -jet WP. Significant improvements in rejections are also observed for jets in the Z'
1989 sample.

1990 6.6 Conclusion

1991 In this chapter a novel jet tagger, GN1, is presented. The model has a graph neural
1992 network architecture and is trained with auxiliary training objectives, which are

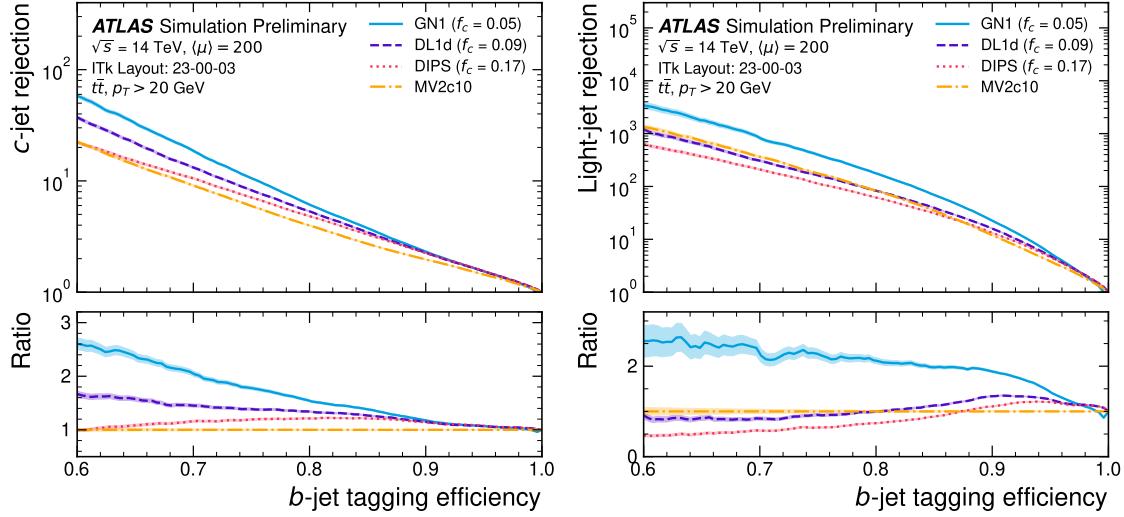


Figure 6.22: The c -jet rejection (left) and light-jet rejection (right) as a function of the b -jet efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ for events with a centre of mass energy $\sqrt{s} = 14 \text{ TeV}$ [125]. The ratio to the performance of the DL1d algorithm 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.

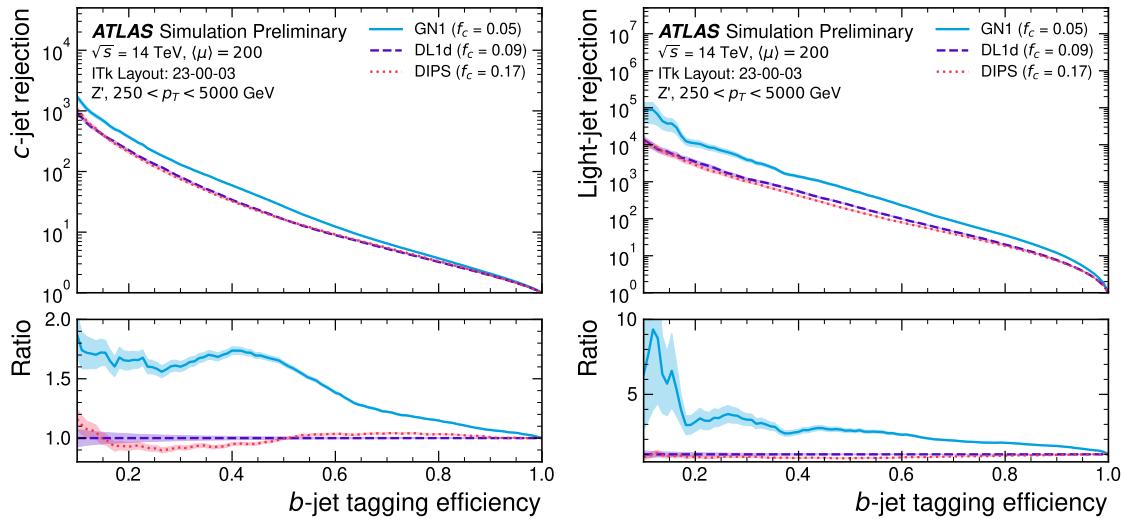


Figure 6.23: The c -jet rejection (left) and light-jet rejection (right) as a function of the b -jet efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$ for events with a centre of mass energy $\sqrt{s} = 14 \text{ TeV}$ [125]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions.

1993 shown to improve the performance of the basic model. GN1 significantly improves
1994 flavour tagging performance with respect to DL1r, the current default ATLAS flavour
1995 tagging algorithm, when compared in simulated collisions. GN1 improves c - and
1996 light-jet rejection for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV by factors of
1997 ~ 2.1 and ~ 1.8 respectively at a b -jet tagging efficiency of 70% when compared with
1998 DL1r. For jets in the Z' sample with $250 < p_T < 5000$ GeV, GN1 improves the
1999 c -jet rejection by a factor of ~ 2.8 and light-jet rejection by a factor of ~ 6 for a
2000 comparative b -jet efficiency of 30%.

2001 Previous multivariate flavour tagging algorithms relied on inputs from low-level
2002 tagging algorithms, whereas GN1 needs no such inputs, making it more flexible.
2003 It can be easily fully optimised via a retraining for specific flavour tagging use
2004 cases, as demonstrated with c -tagging and high- p_T b -tagging, without the need for
2005 time-consuming retuning of the low-level tagging algorithms. The model is also
2006 simpler to maintain and study due to the reduction of constituent components.

2007 GN1 demonstrates improved track classification performance when compared with
2008 a simple per-track MLP and an efficiency of $\sim 80\%$ for inclusive vertex finding in
2009 b -jets. The model is also able to perform vertex finding, and preliminary studies
2010 suggest it outperforms previous manually optimised approaches. The auxiliary track
2011 classification and vertex finding objectives are shown to significantly contribute to
2012 the performance in the jet classification objective, and, along with the more advanced
2013 graph neural network architecture, are directly responsible for the improvement over
2014 DL1r.

2015 Further improvements in the b - and c -tagging performance are likely possible with
2016 a more thorough optimisation of the model architecture, and the integration of
2017 additional information from other parts of the ATLAS detector. The addition of
2018 other auxiliary training objectives, such as the truth b -hadron decay radius and
2019 transverse momentum, may also yield additional performance gains on top of the
2020 gains achieved by loosening the input track selection (demonstrated in Section 6.4.8).

2021 Additional future work includes the verification of the performance of GN1 on collision
2022 data, and the full calibration of the model so it can be used by analyses. The flexible
2023 nature of the model means it can also be readily applied to other related problems
2024 outside of standard b - and c -tagging applications, as demonstrated in Section 6.5.
2025 Additional applications for the architecture include $X \rightarrow bb$ and $X \rightarrow cc$ tagging.

- 2026 The model could also be repurposed as a pile-up jet tagger, or general primary and
2027 secondary vertexing tool.

2028 Chapter 7

2029 Boosted VHbb Analysis

2030 The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [14, 15],
2031 is predicted by the standard model to decay primarily to a pair of b -quarks, with
2032 a branching fraction of 0.582 ± 0.007 for $m_H = 125$ GeV [26]. Observation of this
2033 decay mode was reported by ATLAS [112] and CMS [28] in 2018, establishing the
2034 first direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks
2035 (see Section 2.2.2). The $H \rightarrow b\bar{b}$ process is also important for constraining the total
2036 decay width of the Higgs [126].

2037 Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion
2038 as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet
2039 background and so overall sensitivity to the Higgs is low. The QCD multijet
2040 background refers to events containing one or more strongly produced jets which
2041 are not the decay product of heavy resonances, for example $g \rightarrow q\bar{q}$. The gluon-
2042 gluon fusion channel contains to leading order only jets in the final state, and
2043 therefore it is extremely difficult to distinguish signal events from the overwhelming
2044 multijet background. The $H \rightarrow b\bar{b}$ observation therefore searched for Higgs bosons
2045 produced in association with a vector boson V (where V can be a W or Z boson)
2046 which subsequently decays leptonically. The leptonic final states allow for leptonic
2047 triggering whilst at the same time significantly reducing the multijet background.

2048 Two full Run 2 dataset analyses were carried out as a follow-up to the $H \rightarrow$
2049 $b\bar{b}$ observation [112]. Similar to the observation, both measured the associated
2050 production of a Higgs with a vector boson, with the Higgs boson decaying to a pair
2051 of b -quarks. The first analysis [127] was focussed on the resolved phase-space, where
2052 the Higgs-jet candidate is reconstructed as two distinct jets with radius parameter

2053 $R = 0.4$. The second analysis [128] was focussed on the boosted phase-space, where
2054 the Higgs-jet candidate has a sufficiently large transverse momenta such that it can
2055 be reconstructed as a single jet with a radius parameter of $R = 1.0$. This chapter
2056 will focus on the latter analysis. The analysis is outlined in Section 7.1. Modelling
2057 studies performed are detailed in Section 7.2, and the results of the analysis are
2058 presented in Section 7.4. This analysis has been published in Ref. [128]. Figures and
2059 tables from Ref. [128] are reproduced here.

2060 7.1 Analysis Overview

2061 The boosted VH , $H \rightarrow b\bar{b}$ analysis is focused on the high transverse momentum
2062 regime, which has the benefit of being more sensitive to physics beyond the Standard
2063 Model [129], but the disadvantage of being more challenging due to the increased
2064 difficulty in the accurate reconstruction of high transverse momentum physics objects
2065 (discussed in Chapter 4). In order to focus on the high- p_T regime, the reconstructed
2066 vector boson p_T^V is required to be $p_T^V > 250$ GeV (see Section 7.1.2). Events are
2067 also split into two p_T^V bins with the first bin covering $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and
2068 the second covering $p_T^V > 400 \text{ GeV}$, which allows the analysis to benefit from the
2069 improved signal-to-background in the high- p_T regime.

2070 The previous ATLAS analysis in Ref. [112] was primarily sensitive to vector bosons
2071 with a more modest p_T^V boost in the region of 100–300 GeV. In this regime, the
2072 Higgs candidate was reconstructed using a pair of jets with radius parameter of
2073 $R = 0.4$, called small- R jets. However in the high- p_T regime, the decay products
2074 of the Higgs boson become increasingly collimated and the small- R jets may
2075 not be individually resolved. In order to enhance the reconstruction of the Higgs
2076 boson candidate, this analysis uses a large- R jet with radius parameter $R = 1.0$ to
2077 reconstruct the Higgs boson candidate (see Section 3.4.3). The Higgs candidate is
2078 required to have exactly two ghost-assciated (see Section 3.4.3) and b -tagged variable-
2079 radius track-jets. The candidate large- R jet is reconstructed using jet substructure
2080 techniques, in particular it is trimmed by removing soft and wide-angle components,
2081 which helps to remove particles from the underlying event and pile-up collisions [130].
2082 Refer to Section 3.4.3 for more details on jet reconstruction.

2083 On top of the binning in p_T^V , selected events are further categorised into the 0-, 1-
 2084 and 2-lepton channels depending on the number of charged leptons (electrons and
 2085 muons) present in the reconstructed final state (also referred to as the 0L, 1L, and 2L
 2086 channels respectively). The 0-lepton channel targets the $ZH \rightarrow \nu\nu b\bar{b}$ process, the
 2087 1-lepton channel targets $WH \rightarrow \ell\nu b\bar{b}$, and the 2-lepton channel targets $ZH \rightarrow \ell\ell b\bar{b}$,
 2088 where ℓ is an electron or muon and ν is a neutrino. Each channel has a dedicated
 2089 set of selections which are listed in more detail in Section 7.1.3. Events in the 0- and
 2090 1-lepton channels are further split depending on the number of additional small- R
 2091 jets not matched to the Higgs-jet candidate. The high-purity signal region (HP SR)
 2092 has zero such jets, while the low-purity signal region (LP SR) has one or more, and
 2093 therefore absorbs a larger number of background $t\bar{t}$ events. Maintaining a high purity
 2094 signal region is important for the extraction of the signal yield. The 0- and 1-lepton
 2095 channels also make use of a dedicated $t\bar{t}$ control region for jets with one or more
 2096 additional b -tagged small- R jets, described described in Section 7.1.4. A complete
 2097 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 . 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 [128].

2098 The signal $VH, H \rightarrow b\bar{b}$ yields is extracted from a profile likelihood fit to the large- R
 2099 jet mass over several signal and control analysis regions, which are described in
 2100 Sections 7.1.3 and 7.1.4. The diboson background $VZ, Z \rightarrow b\bar{b}$ yield is simultaneously
 2101 extracted from the fit, and provides a cross check on the signal extraction. The fit

2102 model (described henceforth only as ‘‘the fit’’) is described in more detail in described
2103 in Section 7.3.

2104 7.1.1 Data & Simulated Samples

2105 Data from centre-of-mass energy $\sqrt{s} = 13$ TeV proton-proton collisions at the LHC
2106 recorded over the course of Run 2 (between 2015 and 2018) were used for the analysis.
2107 The resulting dataset corresponds to a total integrated luminosity of 139 fb^{-1} (see
2108 Fig. 3.4).

2109 An overview of the MC simulated samples used in the analysis is given in Table 7.2.
2110 These samples are used to model the signal and background processes relevant to the
2111 analysis, with the exception of the multijet background which is modelled using a
2112 data-driven technique. Data and simulated events are reconstructed using the same
2113 algorithms, and a reweighting is applied to the simulated events in order to match
2114 the pile-up distribution observed in the data.

2115 7.1.2 Object Reconstruction

2116 The presence of neutrinos in the $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \ell\ell b\bar{b}$ signatures can be
2117 inferred from a momentum imbalance in the transverse plane Section 3.4.5. The
2118 vector boson transverse momentum p_T^V is reconstructed as the missing transverse
2119 energy E_T^{miss} in the 0-lepton channel, as the magnitude of the summed $\mathbf{E}_T^{\text{miss}}$ and
2120 charged-lepton momentum in the 1-lepton channel, and as the transverse momentum
2121 of the 2-lepton system in the 2-lepton channel (see Section 3.4.5).

2122 Electrons and muons are reconstructed as outlined in Section 3.4.4, and following the
2123 approach described in Ref. [112]. Leptons are required to satisfy the selections listed
2124 in Table 7.3. *Baseline* electrons are required to pass the likelihood-based method
2125 described in Section 3.4.4, and *Signal* electron additionally are required to satisfy
2126 a tighter likelihood identification selection. *Baseline* muons are required to pass
2127 the ‘loose’ identification described in Ref. [77], while *signal* muons are required to
2128 pass the ‘medium’ identification working point. All signal leptons are required to
2129 additionally satisfy a $p_T > 27$ GeV selection criteria, except for muons in the 1-lepton
2130 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 [131] + GoSAM [133] + MiNLO [134, 135]	NNPDF3.0NLO (*) [104]	Pythia 8.212 [106]	AZNLO [132]	NNLO(QCD) + NLO(EW) [136–142]
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	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 [143–147]
Top quark ($m_t = 172.5$ GeV)					
$t\bar{t}$	PowHEG-Box v2 [131, 148]	NNPDF3.0NLO	Pythia 8.230	A14 [107]	NNLO+NNLL [149]
s-channel	PowHEG-Box v2 [131, 150]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [151]
t-channel	PowHEG-Box v2 [131, 150]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [152]
Wt	PowHEG-Box v2 [131, 153]	NNPDF3.0NLO	Pythia 8.230	A14	Approximate NNLO [154]
Vector boson + jets					
$W \rightarrow \ell\nu$	SHERPA 2.2.1 [155–158]	NNPDF3.0NNLO	SHERPA 2.2.1 [159, 160]	Default	NNLO [161]
$Z/\gamma^* \rightarrow \ell\ell$	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 [128]. 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 [162] 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 [163, 164].

2131 categorise the event into the 0-, 1- or 2-lepton channels. The 1- and 2-lepton channels
2132 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.

2133 The analysis makes use of large- R and variable-radius small- R track-jets, which
2134 are described in Section 3.4.3. The large- R jets are used to reconstruct the Higgs
2135 boson candidate, while the small- R jets are used for b -tagging and for selection of
2136 the analysis region. The track-jets matched to the Higgs candidate are b -tagged
2137 using the MV2c10 b -tagging algorithm (see Chapter 4). The efficiency of the tagging
2138 algorithm is calibrated to events in data [165–167]. The jet tagging strategy relies
2139 on extensive studies into track-jet b -tagging in boosted topologies [168, 169].

2140 7.1.3 Selection Criteria

2141 An extensive list of selection cuts are applied to each event in order to reject
2142 background events whilst retaining as many signal events as possible. A full list of
2143 selection cuts applied to the different analysis regions is given in Table 7.4, while
2144 some key selections are listed below.

2145 All channels require events with at least one large- R jet with $p_T > 250 \text{ GeV}$
2146 and $|\eta| < 2.0$. The vector boson transverse momentum is also required to satisfy
2147 $p_T^V > 250 \text{ GeV}$. The Higgs candidate is chosen as the highest p_T large- R jet satisfying
2148 these requirements. As mentioned, the candidate large- R jet is required to have
2149 two ghost-assciated and b -tagged variable-radius track-jets. These track-jets are
2150 required to have at least two constituent tracks with $p_T > 500 \text{ MeV}$ and $|\eta| < 2.5$.
2151 The track-jets themselves must satisfy $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$.

2152 In the 0-lepton channel, trigger selections are applied using an E_T^{miss} trigger with a
2153 luminosity-dependent threshold between 70–110 GeV. In the 1-lepton electron sub-
2154 channel a combination of single electron triggers is used with minimum p_T thresholds
2155 between 24–26 GeV. In the muon sub-channel the same E_T^{miss} trigger as the 0-lepton
2156 channel is used. Since muons are not used for the E_T^{miss} trigger calculations, this
2157 is in effect a p_T requirement on the muon-neutrino system, which in the analysis
2158 phase space is more efficient than a single-muon trigger. The 2-lepton channel uses
2159 the same triggering strategy as the 1-lepton channel. In all channels, the trigger
2160 selections applied are fully efficient for events selected using the full requirements in
2161 Table 7.4.

2162 The combined selections in Table 7.4 result in a signal efficiency ranging from 6–16%
2163 for the WH and ZH processes depending on the channel and p_T^V bin.

2164 7.1.4 Control Regions

2165 The $t\bar{t}$ process presents a major background in the 0- and 1-lepton channels. In these
2166 events, the Higgs candidate is often reconstructed from a correctly tagged b -jet from
2167 the top decay $t \rightarrow Wb$, and an incorrectly tagged c - or light-jet from the subsequent
2168 decay of the W , as shown in Fig. 7.1.

2169 The top quark predominately decays to a W and a b -quark. Hence, the second top
2170 quark from the $t\bar{t}$ pair is also likely to result in a second tagged b -tagged track-jet
2171 outside of the large- R Higgs candidate. To ensure sufficient $t\bar{t}$ rejection, 0- and
2172 1-lepton channel signal regions are defined using a veto on events with b -tagged
2173 track-jets outside the Higgs-jet candidate. These events are used to construct a
2174 control region (CR) which is enriched in $t\bar{t}$ events. The CR is used to constrain the
2175 normalisation of the $t\bar{t}$ background in the fit.

2176 7.1.5 Background Composition

2177 After the selections described in Section 7.1.3 the number of background events
2178 mimicking the VH , $H \rightarrow b\bar{b}$ signal is greatly reduced. However, the number of
2179 background events still greatly outnumbers that of signal events. The background
2180 processes are channel dependent. In the 0-lepton channel the dominant sources of

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

Table 7.4: Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels and sub-channels [128]. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region. The $\min[\Delta\phi(\mathbf{E}_{\text{T}}^{\text{miss}}, \text{small-}R \text{ jets})]$ selection is used to remove jets when the missing transverse momentum $\mathbf{E}_{\text{T}}^{\text{miss}}$ is pointing in the direction of the Higgs candidate, and the $\Delta\phi(\mathbf{E}_{\text{T}}^{\text{miss}}, \mathbf{E}_{\text{T, trk}}^{\text{miss}})$ is used to reject events where the calorimeter missing transverse momentum $\mathbf{E}_{\text{T}}^{\text{miss}}$ is not pointing in the direction of the track-based missing transverse momentum $\mathbf{E}_{\text{T, trk}}^{\text{miss}}$. The $\Delta y(V, H_{\text{cand}})$ quantifies the rapidity difference between the reconstructed vector boson and Higgs candidate.

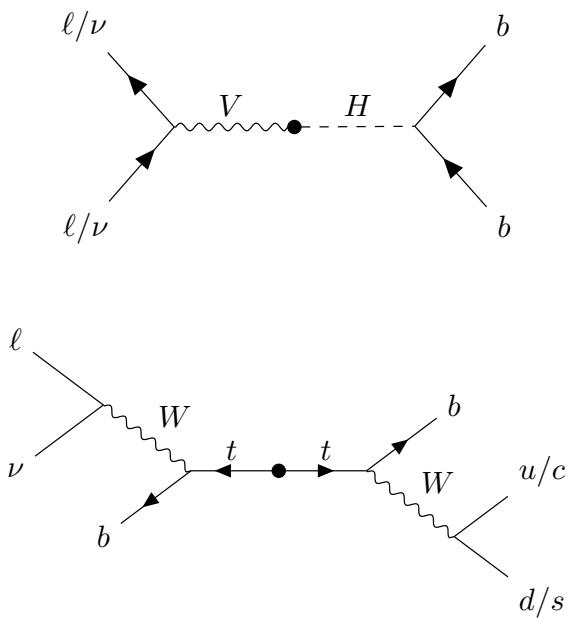


Figure 7.1: Diagrams of the signal process (top) and $t\bar{t}$ background (bottom). Object to the right of centre are reconstructed within the large- R jet. For the $t\bar{t}$ background, the large- R jet contains a mis-tagged c -jet or (less often) a mis-tagged light-jet. The contribution in the 0-lepton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

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

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

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

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

7.2 Systematic Uncertainties & Background Modelling

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

2211 an increase in the number of observed events with respect to the background-only
2212 hypothesis is looked for. The excess is often relatively small against the total number
2213 of background events, and hence accurate modelling of the expected number of
2214 background and signal events is crucial for successfully performing the analysis.
2215 Particular care is paid to the uncertainties on the modelling predictions as discussed
2216 in this section.

2217 Modelling uncertainties are described in detail in the following sections. *Nominal*
2218 samples are used as a reference to which different variations can be compared.
2219 The nominal samples are chosen as the best possible representation of the underlying
2220 physical process. *Alternative* samples are used to understand inaccuracies that may
2221 be present in the nominal samples. Some aspect of the nominal model is varied, and
2222 the discrepancy with respect to the nominal model is quantified. The discrepancy is
2223 used to estimate a systematic uncertainty associated with the model parameter which
2224 was varied. The alternative samples are sometimes obtained via internal weight
2225 variations or parameterisation methods, rather than by re-running the simulation.
2226 This is discussed in more detail in Section 7.2.1.

2227 Modelling studies involving c - and light-jets is hampered by the low number of events
2228 available after the analysis selection is applied, due to the high rejection rates of the
2229 b -tagging algorithm MV2c10. For modelling studies, truth tagging (TT) is therefore
2230 employed to ensure sufficient numbers of jets are available to calculate uncertainties.
2231 TT works by computing a 2-dimensional efficiency map using the jet p_T and η . The
2232 two leading track-jets associated to the large- R jet are weighted based on their p_T
2233 and η using the pre-calculated efficiency map, rather than being required to explicitly
2234 pass the b -tagging requirement.

2235 7.2.1 Implementation of Variations

2236 Modelling variations are implemented in different ways, depending on the associ-
2237 ated uncertainty. Table 7.5 lists the different sources of uncertainty described in
2238 Section 7.2.2 and for each lists the implementation. As production of high-stastic
2239 MC samples is computationally expensive, a technique in state of the art simulation
2240 packages is to store some sources of variation as internal weights, which can be
2241 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.

2242 sample then effectively contains information about an ensemble of different samples,
2243 corresponding to different model parameters, which are accessible via reweightings.
2244 When filling histograms for the variations, bins are incremented by the internal
2245 weight of the event associated with the variation in question.

2246 While the inclusion of internal weight variation in MC event generators has decreased
2247 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently
2248 some sources of systematic uncertainty that are unable to be stored as internal weight
2249 variations due to technical limitations. Two examples are the choice of resummation
2250 and merging scales. A method to parameterise the systematic variation using
2251 one sample, and to then apply this parameterisation to another sample, has been
2252 developed by ATLAS [170]. This method was used to derive resummation and
2253 merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous (lower
2254 statistic) SHERPA 2.1 alternative sample. The resulting uncertainties were studied
2255 and found to be negligible in comparison with systematics from other sources.

2256 7.2.2 Sources of Systematic Modelling Uncertainties

2257 This section briefly describes the different sources of uncertainty in the analysis, and
2258 how each is implemented. For each source of uncertainty, acceptance and shape
2259 uncertainties are derived. Acceptance uncertainties account for the uncertainty in
2260 the overall number of events in each channel, and for the migration of events between

2261 different analysis regions. Meanwhile, shape uncertainties account for differences in
2262 the shapes, but not overall normalisations, of the large- R jet mass.

2263 QCD Scales

2264 The $V + \text{jets}$ matrix element calculations contains infrared and ultraviolet divergences.
2265 These are handled by introducing arbitrary parameters corresponding to the renor-
2266 malisation scale (μ_R) and factorisation scale (μ_F). Physical observables are not
2267 dependent on these parameters when using the infinite perturbation series expansion,
2268 however at some fixed order in QCD a limited dependence is present. To assess the
2269 impact of this, both μ_R and μ_F are independently varied from their nominal values
2270 by factors of 0.5 and 2 to account for higher order corrections to the calculation of
2271 the matrix element used to simulate the process.

2272 PDF Sets

2273 Parton distribution functions (PDFs) specify the probability of finding a parton with
2274 a given momentum inside a hadron (in this case, colliding protons). PDFs have
2275 to be derived from data and are a significant source of uncertainty in analyses of
2276 hadronic collision data. There are three sources of PDF uncertainties: the statistical
2277 and systematic errors on the underlying data used to derive the PDFs, the theory
2278 which is used to describe them (which is based on some fixed order perturbative
2279 QCD expansion), and finally the procedure which is used to extract the PDFs from
2280 the data. PDF-related uncertainties were derived following Ref. [162]. This involves
2281 considering 100 PDF replicas which, when combined, form a central value and
2282 associated uncertainty, and also in parallel direct changes to the central values of
2283 PDFs using the MMHT2014 [171] and CT14NLO [172] PDF sets.

2284 Event Generator

2285 The choice of parton shower (PS) and underlying event (UE) generators can affect
2286 the analysis outcome. Changing these models modifies several aspects of the event
2287 generation at the same time, such as the accuracy of matrix element predictions and

2288 different approaches to parton showering. This change tends to lead to the largest
2289 discrepancy with respect to the nominal samples.

2290 **Resummation and Merging Scales**

2291 Resummation is a technique used in QCD to help cope with calculations involving
2292 disparate energy scales, and involves the introduction of an associated resummation
2293 scale, the choice of which introduces some systematic uncertainty into the model.
2294 Parton showering models are accurate when simulating low- p_T radiation, however
2295 inaccuracies start to arrive when simulating hard emissions. To combat this, par-
2296 ton showering models utilise more precise matrix element calculations above some
2297 momentum threshold. The choice of threshold, or *merging scale* introduces some
2298 uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale
2299 variations are available for a subset of the SHERPA samples. The number of available
2300 events is significantly lower than the number of events in the nominal sample, and no
2301 statistically significant discrepancy with respect to the nominal samples is observed.
2302 The corresponding uncertainties and therefore neglected.

2303 **7.2.3 Sources of Experimental Uncertainties**

2304 The main experimental uncertainties in the analysis are due to the following sources:

- 2305 • The small- R jet energy scale and resolution, which are informed by in situ
2306 calibration studies and the dependence of the jet energy on the level of pile-
2307 up [68].
- 2308 • The large- R jet energy and mass scales and resolutions. The scales are calibrated
2309 as described in Ref. [72], and an uncertainty of 2% and 20% is applied for the
2310 jet energy and mass resolutions, respectively.
- 2311 • b -tagging uncertainties, which are computed separately for b -, c - and light-flavour
2312 jets as described in the calibration studies in Refs. [165–167]. An additional
2313 extrapolation uncertainty is added to account for jets with transverse momenta
2314 above that which is accessible in the calibration analyses.

- 2315 • Uncertainties associated with the lepton energy and momentum scales, and
2316 reconstruction and identification efficiencies.
- 2317 • Uncertainty on the pile-up models which are used in the simulated samples,
2318 described in Ref. [173].
- 2319 • Uncertainties associated with the reconstruction of the missing transverse energy
2320 E_T^{miss} , which have various sources as described in Ref. [78].

2321 The impact of these uncertainties on the analysis can be found in Table 7.16.

2322 7.2.4 Vector Boson + Jets Modelling

2323 After event selection, the $V+jets$ background is a dominant background in all three
2324 analysis channels as described in Section 7.1.5. The $V+jets$ samples are split into
2325 categories depending on the truth flavour of the track-jets which are ghost-associated
2326 to the large- R jet Higgs candidate. The categories are $V+bb$, $V+bc$, $V+bl$, $V+cc$,
2327 $V+cl$, $V+ll$, and $V+hf$ refers collectively to the categories containing at least one b -
2328 or c -jet. $V+bb$ is dominant generally accounting for 80% of the events, while $V+hf$
2329 accounts for around 90% of the events. The full flavour composition breakdown for
2330 each channel and analaysis region are given in Tables 7.6, 7.8 and 7.9.

2331 In order to access uncertainties associated with the use of MC generators, variations
2332 of the simulation are produced using alternative generators or variation of nominal
2333 generator parameters as described in Section 7.2.1. As described in Section 7.1.1,
2334 the nominal MC event generator used for $V+jets$ events is SHERPA 2.2.1, while
2335 MADGRAPH5_AMC@NLO+PYTHIA8 (which uses a different parton showering
2336 model) is used as an alternative generator.

2337 Modelling systematics can have several impacts, including affecting the overall
2338 normalisation for different processes, the relative acceptances between different
2339 analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR,
2340 and between p_T^V bins), and the shapes of the m_J distributions. Since the fit model
2341 fits only the large- R jet mass m_J to data, all shape uncertainties are estimated with
2342 respect to this observable. Several sources of uncertainty, summarised in Section 7.2.2,
2343 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 [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).

Sample	Mp_T^V HP SR	Hp_T^V HP SR	Mp_T^V LP SR	Hp_T^V LP SR	Mp_T^V CR	Hp_T^V CR
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 [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).

Channel	Mp_T^V HP SR	Hp_T^V HP SR	Mp_T^V LP SR	Hp_T^V LP SR	Mp_T^V CR	Hp_T^V CR
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 [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).

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

2344 Acceptance Uncertainties

2345 Several different types of acceptance uncertainties have been calculated and im-
 2346 plemented as nuisance parameters in the fit. These account for the uncertainty
 2347 in the overall number of events in each channel, and for the migration of events
 2348 between different analysis regions. The acceptance uncertainties relevant to the
 2349 $V+jets$ processes are summarised below.

- 2350 • **Overall normalisation:** only relevant where normalisation cannot be left
 2351 unconstrained (or “floating”, i.e. determined as part of the fit). The $V+hf$
 2352 component is left floating in the fit, with independent normalisations used for
 2353 $W+hf$ and $Z+hf$. The normalisations are mainly determined by the 1-lepton
 2354 (for $W+hf$) and 2-lepton (for $Z+hf$) regions respectively and then extrapolated
 2355 to the 0-lepton channel. The negligible $V+jets$ backgrounds were constrained to
 2356 their cross-sections in the fit. were constrained to their cross-sections

- **SR-to-CR relative acceptance:** the uncertainty on the relative number of $V+jets$ events in the signal and control regions.
- **HP-to-LP relative acceptance:** the uncertainty on the relative number of $V+jets$ events in the HP and LP SRs.
- **Medium-to-high p_T^V relative acceptance:** the uncertainty on the relative number of $V+jets$ events in the medium and high p_T^V bins.
- **Flavour relative acceptance:** for each flavour $V+xx$, where $xx \in \{bc, bl, cc\}$ the ratio of $V+xx/V+bb$ events is calculated. This corresponds to the uncertainty on the heavy flavour composition of the $V+hf$ background.
- **Channel relative acceptance:** the uncertainty on the relative number of $V+jets$ events between the channels.

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

V+jets Acceptance Uncertainties				
Boson	W		Z	
Channel	0L	1L	0L	2L
Vbb Norm.	30%	-	-	-
SR-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%			
VI Norm.	30%			

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

²³⁷³ **Shape Uncertainties**

²³⁷⁴ In order to derive shape uncertainties for a given background or signal process,
²³⁷⁵ normalised distributions of the reconstructed large- R Higgs candidate jet mass m_J
²³⁷⁶ are compared for the nominal sample and variations. For each variation, the ratio of
²³⁷⁷ the variation to nominal is calculated, the up and down variations are symmetrised,
²³⁷⁸ and an analytic function is used to parameterise the ratio. If different analysis regions
²³⁷⁹ or channels show the same pattern of variation, a common uncertainty is assigned.

²³⁸⁰ An example of a significant source of uncertainty, arising from choice of factorisation
²³⁸¹ scale μ_R is shown in Fig. 7.2. The HP SRs in the medium and high p_T^V bins are
²³⁸² shown for the 0-lepton channel for the $W+\text{hf}$ and $Z+\text{hf}$ jets. The 0- and 1-lepton
²³⁸³ channels for the $W+\text{hf}$ contribution and the 0- and 2-lepton channels for the $Z+\text{jets}$
²³⁸⁴ contribution were found to have compatible shapes in m_J across channels, and so
²³⁸⁵ were jointly measured. An exponential function $e^{p_0+p_1x} + p_2$ has been fitted to the
²³⁸⁶ ratio of the normalised distributions. The magnitude of the variation is p_T^V dependent,
²³⁸⁷ and so separate uncertainties are implemented in the fit for each p_T^V region.

²³⁸⁸ The shape uncertainties for μ_R were derived on the SRs but are also applied to the
²³⁸⁹ CRs, as the low statistics in the CRs make it difficult to derive dedicated shape
²³⁹⁰ uncertainties. All the shape uncertainties are fully correlated across regions.

²³⁹¹ A comparison of the m_J shapes between SHERPA and MADGRAPH is shown in
²³⁹² Fig. 7.3. The plots are split by process and channel, but merged in SR purity and p_T^V
²³⁹³ bins reflecting similarities between the m_J shapes and variations across these regions.
²³⁹⁴ Due to the low statistics available for the alternate MADGRAPH sample, and the
²³⁹⁵ lack of statistically significant variation between the samples, no additional shape
²³⁹⁶ uncertainty was added to the fit in this case.

²³⁹⁷ The impacts of variations in the factorisation scale μ_F and the choice of PDF set on
²³⁹⁸ m_J shape were also found to be negligible in comparison with μ_R and are hence no
²³⁹⁹ additional uncertainty was added to the fit.

²⁴⁰⁰ **7.2.5 Diboson Modelling**

²⁴⁰¹ The procedure to derive the uncertainties for the diboson background generally
²⁴⁰² follows that of $V+\text{jets}$.

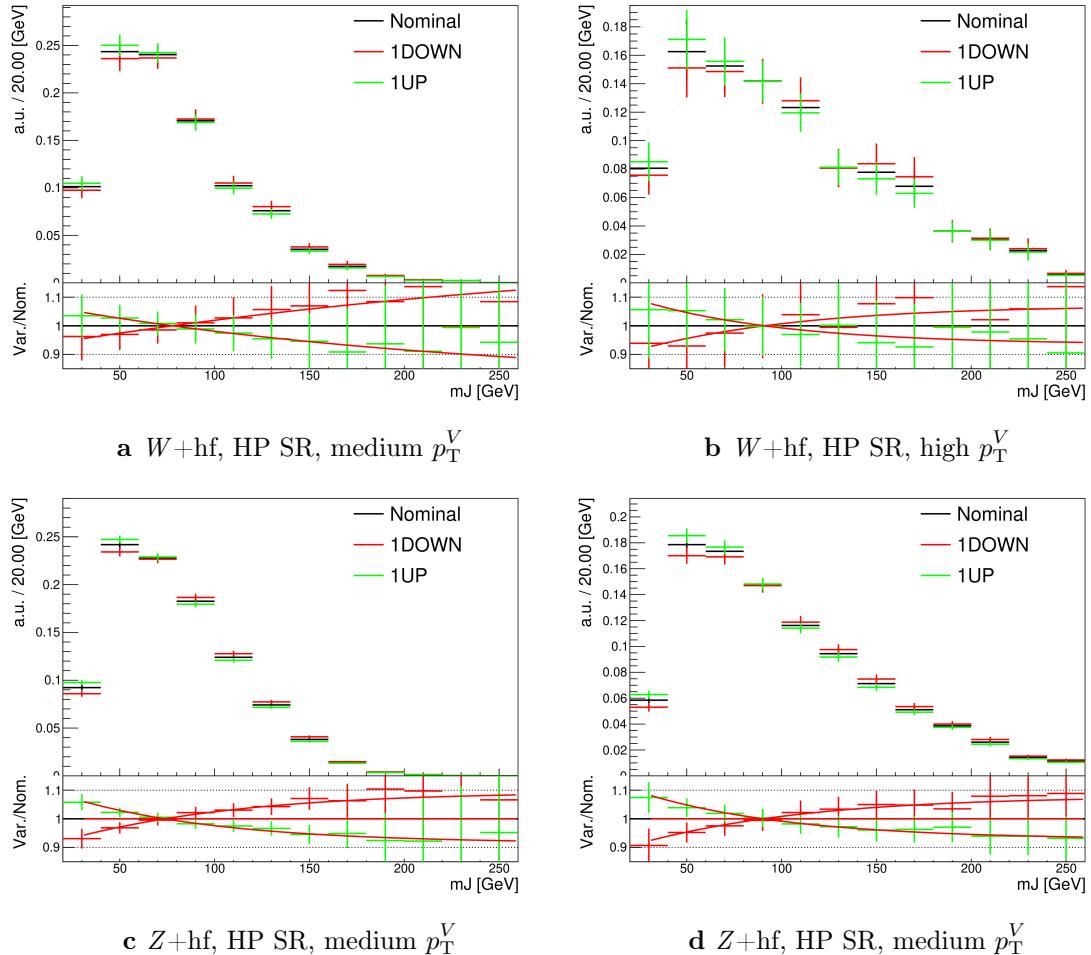


Figure 7.2: Leading large- R jet mass for the Z and $W + \text{hf}$ processes in the HP SR of the 0-lepton channel [174]. The renormalisation scale μ_r has been varied by a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to the ratio between the nominal and alternate samples.

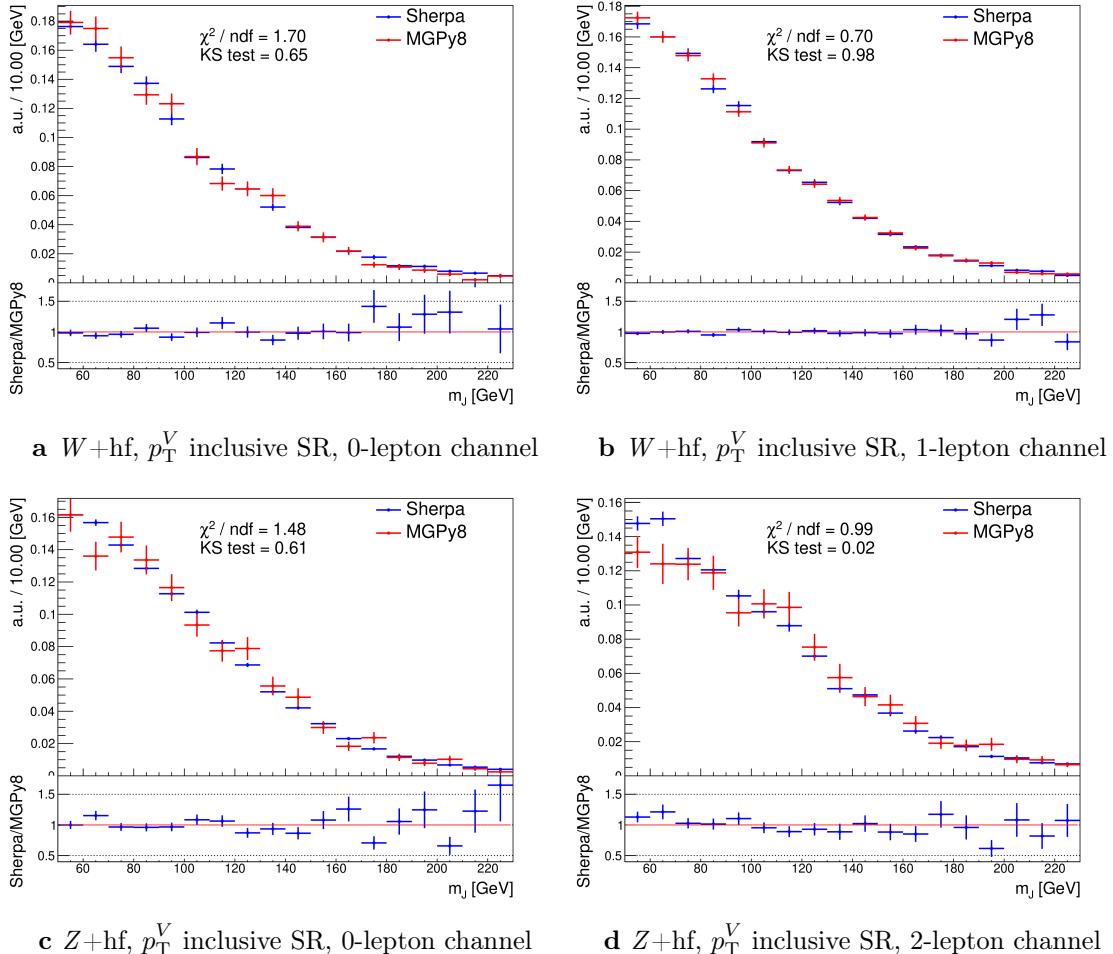


Figure 7.3: Leading large- R jet m_J inclusive in p_T^V for the $V + \text{hf}$ process modelled using both the SHERPA (blue) and MADGRAPH (red) samples [174]. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.

2403 The nominal diboson samples are generated using SHERPA 2.2.1 (except for $gg \rightarrow VV$
2404 which uses SHERPA 2.2.2) with the NNPDF3.0NNLO tune. Alternative samples were
2405 generated using POWHEG interfaced with PYTHIA8, using the AZNLO shower tune
2406 with the CTEQ6L1 PDFs [175]. Unlike SHERPA, POWHEG models the off-shell Z
2407 contribution at NLO.

2408 Acceptance and shape uncertainties are derived in an analogous fashion to $V+jets$
2409 as described below.

2410 Acceptance Uncertainties

2411 Diboson acceptance uncertainties are summarised in Table 7.11. Variations from μ_R ,
2412 μ_F , PDF choice and an alternative generator are considered and are combined via
2413 a sum in quadrature as described in Section 7.2.4. The largest modification to the
2414 nominal acceptance results from the POWHEG+PYTHIA8 alternate sample. Since
2415 the diboson contribution to the $t\bar{t}$ control region is negligible, no SR-to-CR relative
2416 acceptance uncertainty is necessary.

2417 For the WZ contribution, uncertainties are derived using the 1-lepton channel
2418 and applied to all three channels. The 1-lepton channel was used as it has the
2419 largest amount of available statistics. As far as was possible given the limited
2420 statistics available in the other channels, compatibility was checked between the
2421 derived uncertainties and the other channels. An additional 8% channel migration
2422 uncertainty is applied on the WZ 0-lepton channel. For the ZZ contribution, the
2423 normalisation uncertainty is calculated using the 2-lepton channel and applied to all
2424 three channels. The 0- and 1-lepton channels were found to have a similar HP-to-LP
2425 relative acceptance uncertainty of 18%. The 1-lepton medium-to-high p_T^V relative
2426 acceptance is based off the value obtained from the 2-lepton channel, since the
2427 1-lepton channel had an insufficient number of events to estimate the uncertainty
2428 directly. 30% and 18% channel migration uncertainties are applied in the 0- and
2429 1-lepton channels respectively.

2430 Since the contribution from WW is negligible, dedicated studies are not performed,
2431 but a 25% normalisation uncertainty is applied in all the three channels which is
2432 based on the modelling studies performed for the previous analysis [112].

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

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

2433 Shape Uncertainties

2434 Diboson shape uncertainties are derived in a similar fashion to $V+jets$. Only the
 2435 uncertainties associated with systematic variation of μ_R and the alternate event
 2436 generator have a non-negligible impact on the m_J shape. Variation of μ_R produces
 2437 consistent m_J shape changes across all regions and channels, and hence only a single
 2438 associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted
 2439 to the symmetrised ratio.

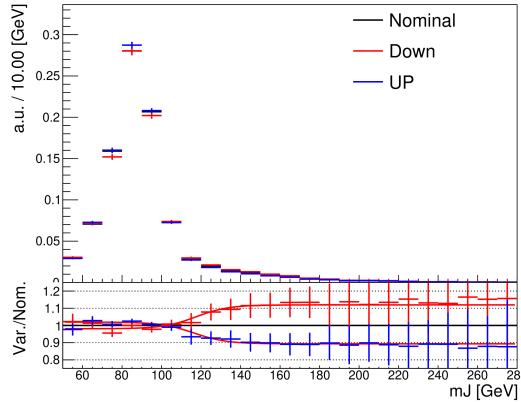


Figure 7.4: Leading large- R jet mass distribution for the WZ and ZZ processes, inclusive across all signal regions and lepton channels [174]. The renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.

2440 The comparison between the nominal SHERPA and alternative POWHEG+PYTHIA8
 2441 samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both the WZ and

2442 ZZ processes. For these channels, the shape of m_J varies in opposite directions
2443 in the LP and HP signal regions. Shapes are similar between p_T^V bins, the 0- and
2444 1-lepton channels and for WZ and ZZ . In order to reduce the effects of statistical
2445 fluctuations on the fit, these regions are merged before deriving the shape uncertainty.
2446 A third order polynomial is fitted to the ratio, and this function transitions to a
2447 constant piecewise function in the high mass region to accurately represent the shape.
2448 Dependence on the event generator was found to be negligible within statistical
2449 uncertainty in the 2-lepton channel, and so no uncertainty was applied. All diboson
2450 shape uncertainties are fully correlated in the fit.

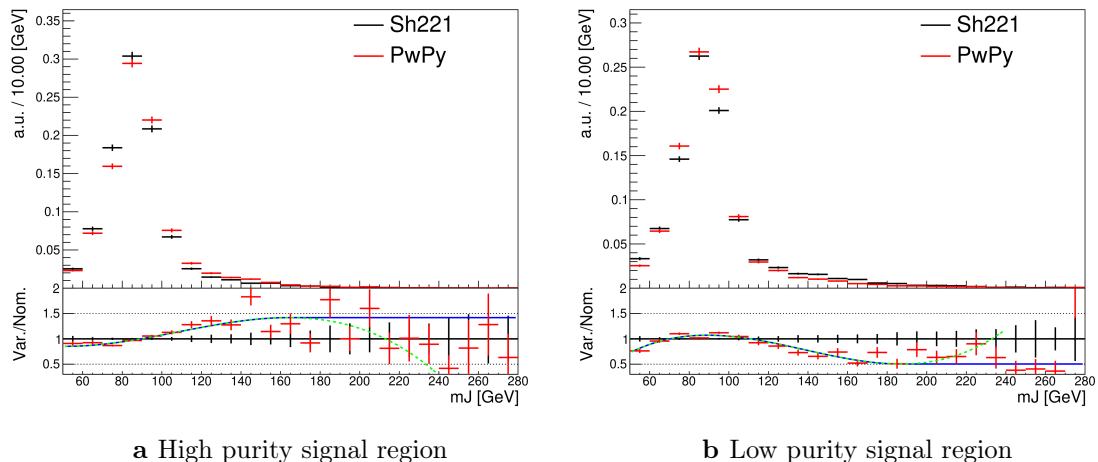


Figure 7.5: The comparison of the shapes of the large- R jet mass m_J between SHERPA (black) and POWHEG+PYTHIA8 (red) samples from WZ and ZZ process in high and low purity signal regions, integrated over p_T^V regions and the 0- and 1-lepton channels [174]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after a protection is added in the high mass region.

2451 7.2.6 $t\bar{t}$ and single-top Modelling

2452 The main features of the systematic uncertainties on the remaining two modelled
2453 backgrounds, $t\bar{t}$ and single-top, are described below.

2454 The modelling of the $t\bar{t}$ background uses a POWHEG+PYTHIA8 nominal sample.
2455 Two alternate samples were considered: POWHEG+HERWIG7 (providing an alternate
2456 parton shower model) and MADGRAPH5+PYTHIA8.2 (providing an alternate hard
2457 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. [176] 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, 177, 178]. 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 [179, 180]. 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.

2488 7.3 Statistical Treatment

2489 A binned global maximum-profile-likelihood fit of the m_J distribution is performed
 2490 to extract information on the signal, combining all the analysis regions defined
 2491 in Table 7.1. The signal strength $\mu = \sigma/\sigma_{\text{SM}}$ is defined as the ratio between the
 2492 observed and predicted cross-sections, where $\mu = 0$ corresponds to the background-
 2493 only hypothesis and $\mu = 1$ corresponds to the SM prediction. This is a parameter of
 2494 interest (POI) which acts to scale the total number of signal events, and is determined
 2495 during the fit procedure.

2496 The present analysis makes use of two POIs. The first, μ_{VH}^{bb} , is the signal strength
 2497 for the VH , $H \rightarrow b\bar{b}$ process, the primary process under investigation. The diboson
 2498 production strength μ_{VZ}^{bb} for the VZ , $Z \rightarrow b\bar{b}$ process is measured simultaneously and
 2499 provides a validation of the analysis apparatus used for the primary $H \rightarrow b\bar{b}$ measure-
 2500 ment. Alongside the two POIs, the predictive model depends on several parameters
 2501 which are not the primary target of measurement, and represent the systematic
 2502 uncertainties discussed previously. These parameters are called nuisance parameters
 2503 (NPs), collectively referred to as θ . Freely floating background normalisations are
 2504 implemented as NPs and are also extracted during the fitting processes.

2505 7.3.1 Likelihood Function

2506 The statistical setup treats each bin as a Poisson counting experiment and is based on
 2507 the ROOSTATS framework [181]. The combined likelihood over N bins is constructed
 2508 as the product of Poisson probabilities in each bin. Considering the simplified case
 2509 of a single signal strength parameter μ , and neglecting sources of systematic or
 2510 statistical uncertainty, this is given by

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

2511 where s_i (b_i) is the expected number of signal (background) events in bin i , and n_i is
 2512 the number of observed data events in bin i .

2513 **Treatment of Uncertainties**

2514 Systematic uncertainties can modify the predicted signal and background yields s_i
2515 and b_i . Each source of systematic uncertainty is taken into account by adding an
2516 additional NP θ_j to the likelihood in the form of a Gaussian cost function. The
2517 combined effect of the NPs is then

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

2518 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
2519 value, and σ_j is the corresponding associated prior uncertainty on the NP. As the
2520 fitted value $\hat{\theta}_j$ deviates from its nominal value, a cost is introduced. The presence of
2521 NPs modifies the likelihood as

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

2522 The predicted signal and background yields are also modified by the presence of the
2523 NPs with

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

2524 For NPs which are left freely floating in the fit, no corresponding Gaussian constraints
2525 are added to the likelihood.

2526 The pull of a NP is defined as the difference between the fitted value $\hat{\theta}_j$ and the
2527 nominal value θ_j , divided by the uncertainty on the NP σ_j . To obtain the uncertainty
2528 on the pull of a NP, the following procedure is used. The Hessian matrix \mathbf{H} is
2529 calculated as

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

2530 Taking the inverse of the Hessian matrix \mathbf{H}^{-1} yields the covariance matrix, from
 2531 which the post-fit uncertainties on the different NPs can be extracted. If the post-fit
 2532 uncertainty is smaller than the nominal uncertainty, additional information about
 2533 the NP has been extracted by the fit, and NP is said to be *constrained*.

2534 The statistical uncertainty on the simulated events is implemented using a dedicated
 2535 NP for each bin which can scale the background yield in that bin. Statistical NPs
 2536 are also implemented using a Gaussian constraint.

2537 Smoothing and Pruning

2538 To simplify the fit to reduce and improve its robustness, systematic uncertainties
 2539 are smoothed and pruned. Smoothing accounts for the large statistical uncertainty
 2540 present in some samples that can lead to unphysical fluctuations in the shape
 2541 systematics. The smoothing procedure relies on the assumption that the impact of
 2542 systematics should be approximately monotonic and correlated between neighbouring
 2543 bins.

2544 In addition to smoothing, pruning is the process of removing from the fit those
 2545 systematics which only have a very small effect. This improves the stability of the
 2546 fit by reducing the number of degrees of freedom. Acceptance uncertainties are
 2547 pruned in a given region if they have a variation of less than 0.5%, or if the up and
 2548 down variations have the same sign in that region. Shape uncertainties are pruned
 2549 in a given region if the deviation in each bin is less than 0.5% in that region. In
 2550 addition, acceptance and shape uncertainties are neglected in a given region for any
 2551 background which makes up less than 2% of the total background in a given region.

2552 Fit Procedure and Statistical Tests

2553 The best-fit value of μ , denoted $\hat{\mu}$, is obtained via an unconditional maximisation of
 2554 the likelihood. The likelihood is also used to construct a statistical test which can
 2555 confirm or reject the background-only hypothesis. The test statistic q_μ is constructed
 2556 from the profile likelihood ratio,

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

2557 where $\hat{\mu}$ and $\hat{\theta}$ are chosen to maximise the likelihood \mathcal{L} , and the profile value $\hat{\theta}_\mu$ is
 2558 obtained from a conditional maximisation of the likelihood for a specific choice of
 2559 $\mu = 0$ corresponding to the background-only hypothesis.

2560 The test statistic is used to construct a p -value which is used to probe the background-
 2561 only hypothesis. The p -value is typically reported in terms of the significance Z ,
 2562 defined as the number of standard deviations for a Gaussian Normal distribution
 2563 which will produce a one-sided tail integral equal to the p -value, as in

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

2564 Typically a value of $Z = 3$ constitutes *evidence* of a processes, while $Z = 5$ is required
 2565 for a *discovery*, or observation. Alongside the p -value, the best-fit value of the signal
 2566 strength $\hat{\mu}$ and its corresponding uncertainty are quoted, and compared to their
 2567 expected values.

2568 7.3.2 Background Normalisations

2569 The backgrounds which can be constrained by the fit are left freely floating and
 2570 the corresponding normalisation factors are extracted. Normalisation factors (NF),
 2571 represent the value by which the predicted normalisations are scaled, and are im-
 2572 plemented for the dominant backgrounds ($t\bar{t}$, $Z+hf$, $W+hf$). The NFs are also
 2573 subdivided into different regions of phase-space for $t\bar{t}$, given it is possible to obtain
 2574 a strong constraint in the individual channels. This also removes the need for an
 2575 extrapolation uncertainty.

2576 The normalisations and shapes of all other backgrounds, with the exception of the
2577 multijet background which is estimated using a data driven technique, are initialised
2578 using the nominal samples and the state-of-the art process normalisations, as outlined
2579 in Table 7.2.

2580 **7.3.3 Asimov Dataset & Expected Results**

2581 The Asimov dataset is constructed by replacing the data with the sum of the signal
2582 and background predictions $n_i = s_i + b_i$. A fit to this dataset using the nominal
2583 values of the NPs from the simulation will recover the input values and is useful for
2584 studying the expected result, in addition to constraints on and correlations between
2585 the NPs.

2586 Alternatively, a conditional fit to the Asimov dataset can be performed using values
2587 of the background NPs which are determined from an unconditional fit to data. The
2588 signal NPs and POIs are fixed at their nominal values from the SM simulation. The
2589 result of this fit can be used to calculate expected (median) significances given a
2590 more realistic background model, which can be compared to their observed values,
2591 as is done in Section 7.4.3.

2592 **7.4 Results**

2593 In the present analysis, the two signal strength parameters μ_{VH}^{bb} and μ_{VZ}^{bb} are extracted
2594 from a simultaneous maximisation of the likelihood described in Section 7.3. The
2595 results of the analysis are summarised in this section. The corresponding post-fit
2596 background normalisations are listed in Table 7.12. Post-fit m_J distributions are
2597 shown in Section 7.4.1. The observed signal strengths are given in Section 7.4.3,
2598 along with observed and expected significances. Finally in Section 7.4.4 the impact
2599 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 [128]. The errors represent the combined statistical and systematic uncertainties.

2600 7.4.1 Post-fit Distributions

2601 In addition to the observed significance and signal strength, it is also necessary to
 2602 study the post-fit m_J distributions to compare level of the agreement between the
 2603 simulation (using the best-fit values of the signal strength $\hat{\mu}$ and the NP $\hat{\theta}$) and the
 2604 data. The best-fit values $\hat{\mu}$ and $\hat{\theta}$ are obtained from an unconditional fit to data over
 2605 all analysis regions. Post-fit m_J distributions are given for the signal regions in the
 2606 0-, 1- and 2-lepton channels in Fig. 7.6. The LP and HP regions are merged for the
 2607 0- and 1-lepton channels for the sake of simplicity. In general there is a good level of
 2608 agreement between the simulation and data, indicating the fit model is performing as
 2609 expected. Fig. 7.7 shows the post-fit plots for the $t\bar{t}$ control regions. Again, a good
 2610 level of agreement is observed given the statistical uncertainties on the distributions.

2611 7.4.2 Post-fit Yields

2612 The post-fit yields resulting from the unconditional fit to data are shown in Table 7.13,
 2613 Table 7.14, and Table 7.15 for the 0-, 1- and 2-lepton channels, respectively.

2614 7.4.3 Observed Signal Strength & Significance

2615 The measured signal strength is computed as the ratio between the measured signal
 2616 yield to the prediction from the SM. The combined result for all three lepton channels
 2617 and all analysis regions is given for μ_{VH}^{bb} in Eq. (7.8), and for μ_{VZ}^{bb} in Eq. (7.9). Both

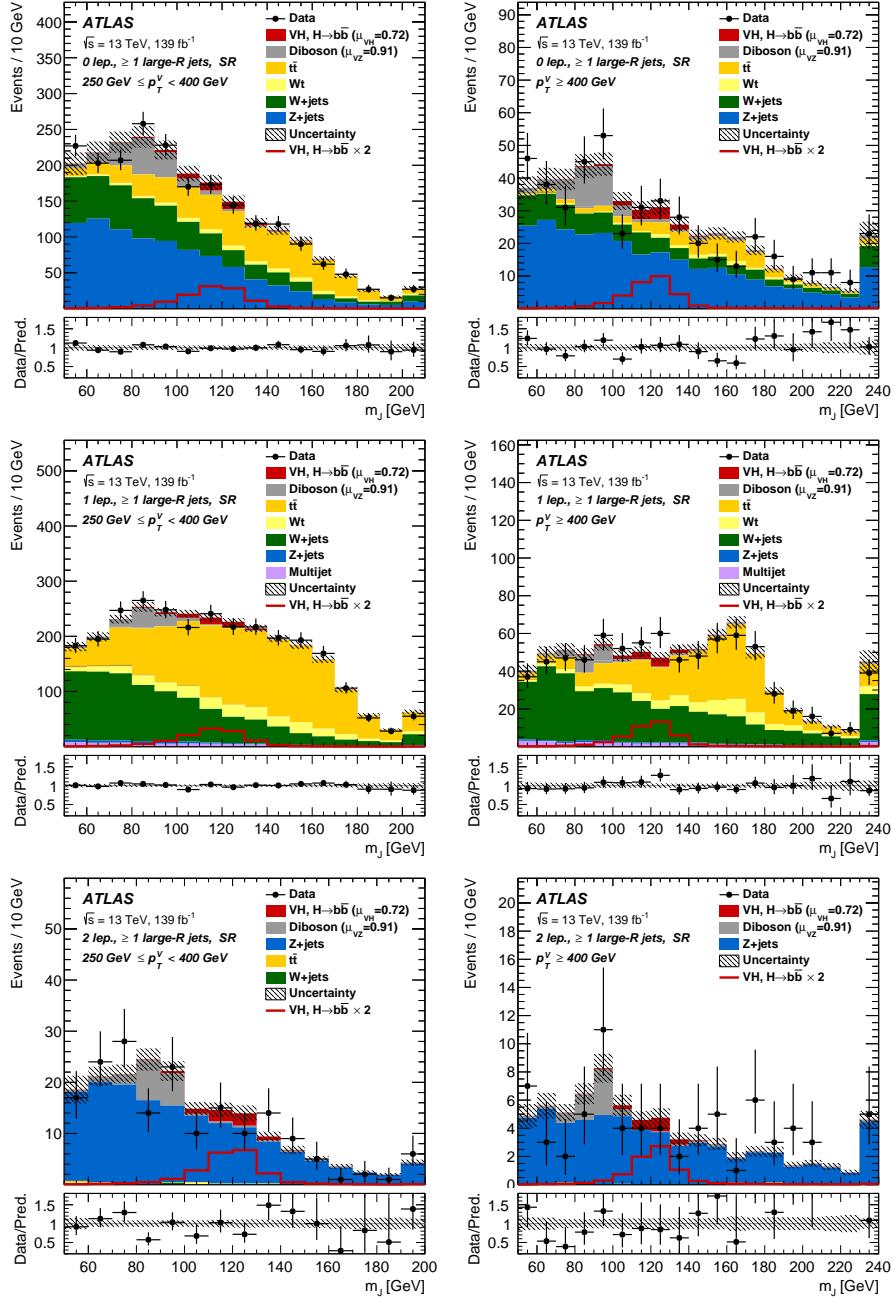


Figure 7.6: The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions 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 [128].

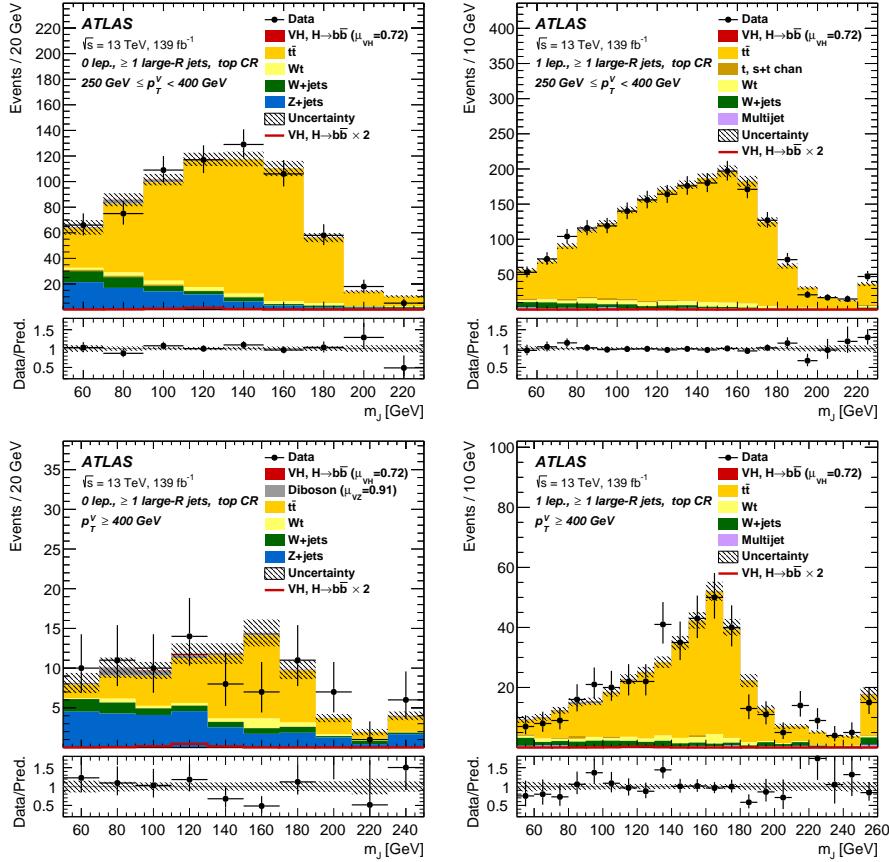


Figure 7.7: The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. The background contributions after 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 [128].

Processes	$250 \text{ GeV} < p_T^V \leq 400 \text{ GeV}$			$p_T^V > 400 \text{ GeV}$		
	HP	LP	CR	HP	LP	CR
Signal	21.93±11.17	18.99±9.76	1.05±0.54	5.69±2.88	5.85±3.01	0.33±0.17
W+t	14.70±5.37	45.55±19.44	17.18±8.09	2.03±0.98	8.93±6.33	3.76±2.49
other t+X	0.79±0.03	3.18±0.66	4.51±1.28	-	0.66±0.03	0.11±0.00
$t\bar{t}$	75.19±13.60	423.85±36.12	539.21±31.39	7.54±1.77	38.20±6.75	44.07±7.43
VZ	77.01±17.09	87.70±19.36	6.16±1.56	17.30±4.10	28.77±6.55	2.79±0.72
WW	-	2.15±0.05	0.24±0.01	0.33±0.02	1.80±0.06	-
Whf	100.78±20.01	331.31±59.54	29.97±21.85	20.19±6.24	59.82±17.91	6.61±5.09
Wcl	5.13±2.31	8.44±3.24	0.46±0.01	0.99±0.69	2.77±1.14	0.19±0.07
Wl	5.61±3.93	4.61±2.45	0.16±0.00	1.41±2.06	2.67±1.67	0.57±0.36
Zhf	318.76±35.27	548.71±61.84	76.97±21.47	86.79±10.63	184.99±21.43	25.76±7.43
Zcl	3.97±1.63	6.74±2.68	0.83±0.02	-	6.36±2.73	0.93±0.41
Zl	1.34±0.67	3.61±2.14	0.42±0.01	1.05±0.63	3.68±2.47	0.29±0.16
Data	623	1493	683	146	330	85
Background	603±25	1466±36	676±25	138±9	339±15	85±7

Table 7.13: Post-fit yields in the 0-lepton channel. Combined statistical and systematic uncertainties are shown [174].

Processes	$250 \text{ GeV} < p_T^V \leq 400 \text{ GeV}$			$p_T^V > 400 \text{ GeV}$		
	HP	LP	CR	HP	LP	CR
Signal	24.23±12.34	18.02±9.29	0.88±0.45	7.84±3.96	7.50±3.87	0.39±0.20
W+t	64.35±21.12	159.95±75.14	73.44±29.96	16.40±7.31	53.28±41.74	21.16±15.36
other t+X	1.92±0.48	16.33±0.31	21.89±6.18	0.13±0.01	1.70±0.06	3.95±1.40
$t\bar{t}$	234.76±30.21	1189.51±75.91	1758.08±57.99	50.87±7.34	226.85±23.98	340.61±25.32
VZ	35.94±8.87	56.30±13.98	4.93±1.38	8.63±2.30	20.02±5.29	2.61±0.84
WW	-	6.48±1.63	-	-	4.35±1.32	0.93±0.03
Whf	265.13±27.68	617.81±63.56	59.91±21.90	91.42±11.51	238.81±29.53	26.55±9.84
Wcl	7.33±2.95	13.81±5.65	2.10±0.04	6.23±2.49	10.17±4.09	0.63±0.02
Wl	2.99±1.47	5.66±3.39	0.65±0.01	2.21±1.35	7.67±4.98	0.31±0.01
Zhf	10.16±1.24	24.61±2.46	3.45±0.41	2.12±0.30	6.56±0.79	0.98±0.12
Zcl	0.02±0.00	0.75±0.02	-	-	0.33±0.01	0.02±0.00
Zl	-	0.49±0.01	0.03±0.00	0.30±0.19	0.23±0.01	0.02±0.00
ggWW	-	0.35±0.01	0.27±0.01	0.15±0.02	0.33±0.01	-
MultiJet	17.04±8.87	44.29±22.82	21.78±11.22	7.81±4.50	21.85±12.73	7.86±4.01
Data	668	2161	1946	185	597	410
Background	640±26	2136±44	1947±43	186±11	592±21	406±18

Table 7.14: Post-fit yields in the 1-lepton channel. Combined statistical and systematic uncertainties are shown [174].

Processes	$250 \text{ GeV} < p_T^V \leq 400 \text{ GeV}$	$p_T^V > 400 \text{ GeV}$
	SR	SR
Signal	7.62 ± 3.88	2.79 ± 1.41
W+t	1.28 ± 0.39	-
t̄t	1.64 ± 0.35	0.45 ± 0.10
VZ	19.90 ± 4.86	7.49 ± 2.05
Whf	0.41 ± 0.07	0.07 ± 0.01
Zhf	150.94 ± 12.72	57.15 ± 5.81
Zcl	2.20 ± 0.91	1.80 ± 0.76
Zl	0.94 ± 0.67	1.01 ± 0.67
Data	179	73
Background	177 ± 12	68 ± 6

Table 7.15: Post-fit yields in the 2-lepton channel. Combined statistical and systematic uncertainties are shown [174].

2618 results include a full breakdown of the systematic and statistical uncertainties.

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

2619

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

2620 The results for μ_{VH}^{bb} and μ_{VZ}^{bb} are consistent with the expectation from the SM. The
2621 μ_{VH}^{bb} measurement is dominated by statistical uncertainty, while the μ_{VZ}^{bb} measurement
2622 is dominated by systematic sources of uncertainty. The measured signal strength
2623 for μ_{VZ}^{bb} corresponds to an observed significance of 2.1 standard deviations, with an
2624 expected (median) significance given the SM prediction of 2.7 standard deviations.
2625 The diboson observed (expected) signal strength significance is 5.4 (5.7). These
2626 results are summarised in Fig. 7.8, which shows the background-subtracted m_J
2627 distribution. A clear signal excess is visible around the Higgs mass of $m_H = 125 \text{ GeV}$.

2628 Compatability Studies

2629 Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting μ_{VH}^{bb}
2630 into three separate POIs, one for each channel. A simultaneous fit to the channel
2631 specific signal strengths can then be performned, which allows a comparison of the
2632 contributions from each channel. Fig. 7.9 compares the best-fit signal strengths.
2633 The 0- and 1-lepton channels show a signal strength which is consistent with the

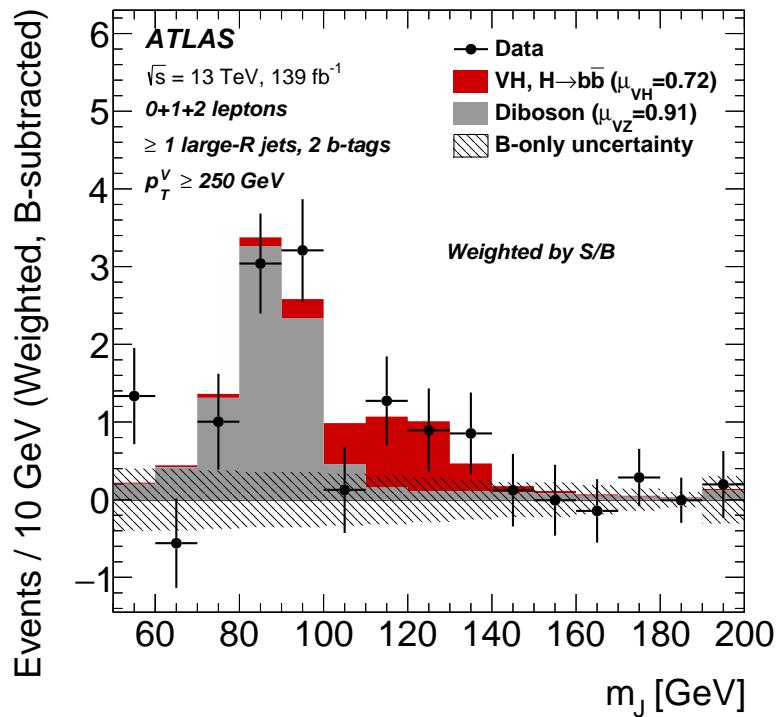


Figure 7.8: m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated WH and ZH production of a SM Higgs boson with $m_H = 125 \text{ GeV}$ is shown scaled by the measured combined signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.2, except for the VH/VZ experimental and theory uncertainties [128].

2634 SM prediction, while the 2-lepton channel shows a small deviation within the 1σ
2635 uncertainty. Overall, good compatibility is observed via a χ^2 test with a corresponding
2636 p -value of 49%.

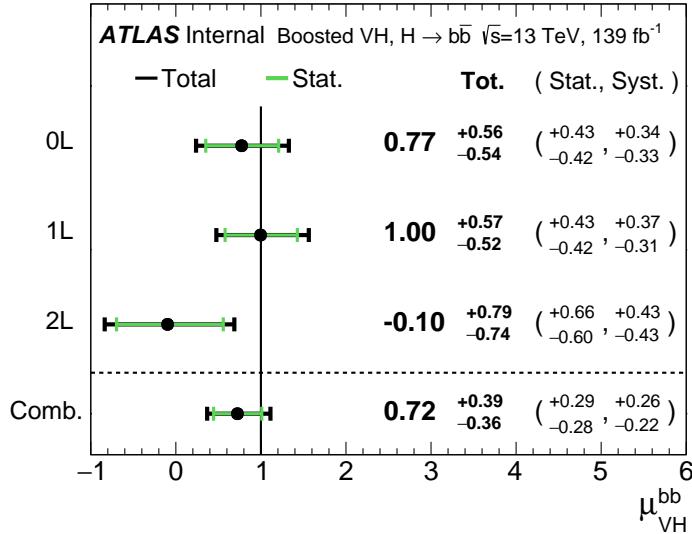


Figure 7.9: Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels splitted) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a χ^2 difference test and results in a p -value of 49% [128].

2637 7.4.4 Impact of Systematics

2638 The impact of systematic uncertainties on the final fitted value $\hat{\mu}^{bb}$ can be studied
2639 using the NP rankings, and the uncertainty breakdown.

2640 Fig. 7.10 shows the NP ranking, which is used to visualise which out of the many
2641 NPs involved in the fit have the largest impact on the sensitivity to the fitted
2642 POI. To obtain the ranking, a likelihood scan is performed for each NP θ_j . First,
2643 an unconditional fit is used to determine $\hat{\theta}_j$. The NP is then fixed to its post-fit
2644 value varied by $\pm 1\sigma$, the fit is repeated and the best-fit value of the POI, $\Delta\hat{\mu}_{VH}^{bb}$, is
2645 calculated, and used to rank the NPs. Also shown in Fig. 7.10 are the pulls and
2646 constraints for the highest ranked NPs.

2647 The experimental uncertainty on the signal large- R jet mass resolution (JMR) has
2648 the largest impact. JMR and jet energy scale (JES) uncertainties also have impacts

2649 for the $V+jets$ background and for the diboson background. The freely-floating
2650 $Z+hf$ normalisation is the second highest ranked NP, and is heavily constrained by
2651 the fit. The VZ POI μ_{VZ}^{bb} is also a significant NP when considering the primary μ_{VH}^{bb}
2652 measurement.

2653 The NP ranking highlights individual NPs which have a large impact on the POI
2654 measurement sensitivity. Complementary information is provided at a higher level
2655 by considering the overall impact of different groups of systematics. The groups
2656 are constructed from NPs which have similar physical origin. The impact of each
2657 group is calculated by running a fit with all the NPs in the given group fixed to their
2658 nominal values. The uncertainty on the POI extracted from this fit is subtracted
2659 in quadrature from the the uncertainty on the POI from the nominal fit, and the
2660 resulting values are provided as the impact for each group. The full breakdown
2661 for the observed impact of uncertainties on the μ_{VH}^{bb} signal strength is provided in
2662 Table 7.16. The total systematic impact is the difference in quadrature between
2663 the nominal uncertainty on μ_{VH}^{bb} and the combined statistical impact. The “data
2664 stat only” group fixes all NPs at their nominal value, while the total statistical
2665 impact fixes all NPs except floating normalisations. The floating normalisations
2666 group fixes only the NPs associated with normalisation which are left floating in
2667 the fit. The uncertainty on μ_{VH}^{bb} is dominated by combined statistical effects (0.28),
2668 although the combined impact of systematics (0.24) is of a comparable size. The
2669 signal largest group is the data stat uncertainty (0.25), demonstrating that the
2670 analysis would benefit from an increased integrated luminosity or improved efficiency
2671 to select signal events (recall from Section 7.1.3 the signal efficiency is in the range of
2672 10%). Of the experimental systematic sources of uncertainty, the dominant impact
2673 is the experimental uncertainties associated with the reconstruction of large- R jets
2674 (0.13). Other experimental sources of uncertainty are small in comparison. Modelling
2675 uncertainties also have a large contribution to the overall systematic uncertainty. The
2676 biggest contribution to the overall uncertainty is the combined statistical uncertainty
2677 on the simulated samples (0.09). Out of the backgrounds, the $W+jets$ and $Z+jets$
2678 have the highest (0.06) and second-highest (0.05) impact respectively.

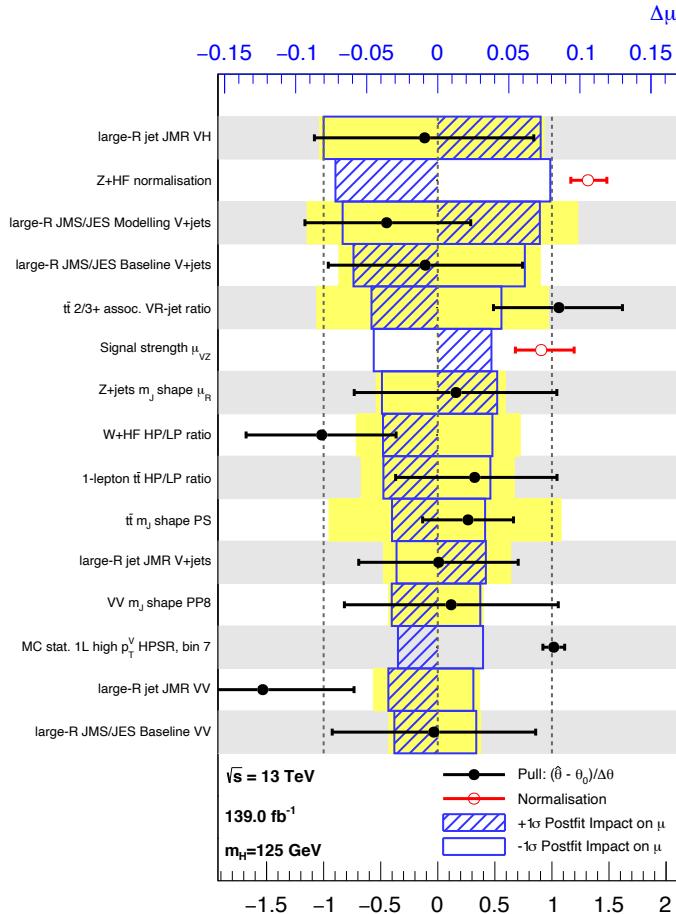


Figure 7.10: Impact of systematic uncertainties on the fitted VH signal-strength parameter $\hat{\mu}_{VH}^{\text{bb}}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value 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 [174].

Source of uncertainty	Signed impact	Avg. impact
Total	+0.388 / -0.356	0.372
Statistical	+0.286 / -0.280	0.283
↔ Data stat only	+0.251 / -0.245	0.248
↔ Floating normalisations	+0.096 / -0.092	0.094
Systematic	+0.261 / -0.219	0.240
<hr/>		
Experimental uncertainties		
small-R jets	+0.041 / -0.034	0.038
large-R jets	+0.161 / -0.105	0.133
E_T^{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 7.16: 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 [174].

2679 7.4.5 STXS Interpretation

2680 The Simplified Template Cross Sections (STXS) framework provides a common
 2681 categorisation of candidate Higgs boson events according to certain truth-level
 2682 properties of the production mode under study [26, 182]. The STXS framework is
 2683 designed to be independent of the decay mode of the Higgs boson, and is therefore
 2684 well suited to the combination of measurements between different decay channels
 2685 and experiments.

2686 The STXS cross sections are independently measured for the ZH and WH production
 2687 modes following the approach described in [177]. For each production mode, two
 2688 bins in the truth vector boson transverse momentum $p_T^{V,t}$ are considered, $250 \text{ GeV} <$
 2689 $p_T^{V,t} < 400 \text{ GeV}$ and $p_T^{V,t} \geq 400 \text{ GeV}$, leading to four independent analysis regions.
 2690 Events from the simulated signal samples are categorised into the regions and used
 2691 to estimate the expected cross section times branching ratio $\sigma \times B$ in each region,
 2692 where

$$B = B(H \rightarrow b\bar{b}) \times B(V \rightarrow \text{leptons}), \quad (7.10)$$

2693 A simultaneous fit of the four cross section times branching ratios is performed.
 2694 The uncertainties described in Section 7.2 are reused for the STXS fit, with the
 2695 exception of the theoretical uncertainties on the signal cross section and branching
 2696 ratios. The result from the fit is shown in Section 7.4.5 and compared with the
 2697 expected prediction from the SM. The expected and observed results agree within
 2698 the given uncertainties.

2699 7.5 Conclusion

2700 The analysis of the associated production of vector bosons with boosted Higgs bosons
 2701 decaying to a pair of b -quarks using large- R jets is presented. The Higgs candidate is
 2702 reconstructed as a large- R jet in order to improve sensitivity in the boosted regime
 2703 in which the Higgs decay products are significantly collimated. The analysis is
 2704 performed using 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13 \text{ TeV}$ collected
 2705 throughout the duration of Run 2 of the LHC.

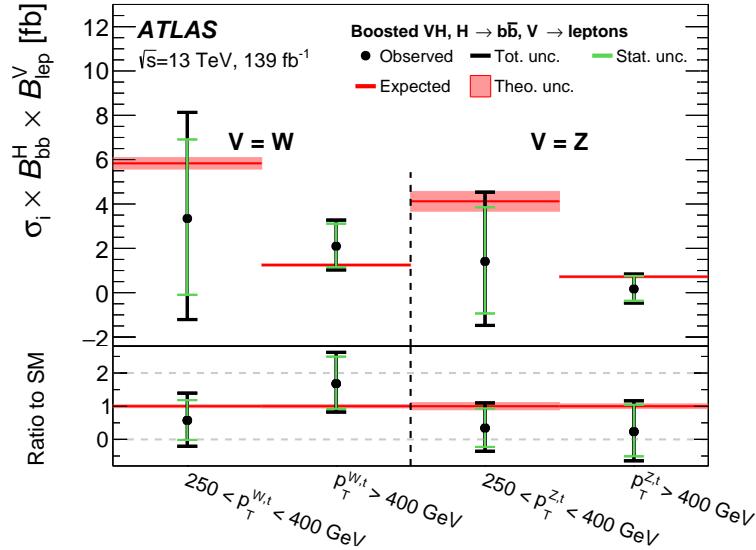


Figure 7.11: Measured VH simplified template cross sections times the $H \rightarrow b\bar{b}$ and $V \rightarrow \text{leptons}$ branching fractions in the medium and high $p_T^{V,t}$ bins [128].

2706 In comparison with the null hypothesis, the Standard Model (SM) VH , $H \rightarrow b\bar{b}$
 2707 process is found to have an observed significance of 2.1 standard deviations, whereas
 2708 the corresponding expected significance is 2.7 standard deviations. The VH , $H \rightarrow b\bar{b}$
 2709 process is measured simultaneously with the diboson VZ , $Z \rightarrow b\bar{b}$ process, which
 2710 provide a cross-check for the main analysis. The observed (expected) significance for
 2711 the diboson process is 5.4 (5.7).

2712 The statistical and systematic sources of uncertainty contribute a similar amount
 2713 to the overall uncertainty on the result. This analysis would therefore likely benefit
 2714 greatly from the improved b -tagging efficiency at high- p_T enabled by GN1 as discussed
 2715 in Chapter 6, due to the associated reduction in statistical uncertainty provided by
 2716 the increased number of events used in the analysis.

2717 The large- R jet mass resolution is found to be the dominant source of systematic
 2718 uncertainty on the μ_{VH}^{bb} measurement. An improved method of reconstructing the
 2719 large- R jet mass, for example by using a machine learning based regression approach,
 2720 possibly as an additional auxiliary task to GN1 (see Chapter 6), would therefore be
 2721 expected to reduce the systematic uncertainty on the μ_{VH}^{bb} measurement.

2722 **Chapter 8**

2723 **Conclusion**

2724 **8.1 Summary**

2725 The current understanding of particle physics contains many unanswered questions,
2726 and improving our understanding of the Standard Model is a promising way to
2727 attempt to answer some of them. One of the key objects which may enhance this
2728 understanding is the Higgs Boson, which was first observed only a decade ago and
2729 remains under intense scrutiny at the LHC. Given it's propensity to decay to heavy
2730 flavour b -quarks, reconstructing and identifying b -jets is of crucial importance to
2731 improving our understanding in this area. As discussed in Chapter 4, this task
2732 becomes increasingly difficult at high transverse momenta.

2733 One of the effects that hampered tracking and b -tagging performance at high- p_T was
2734 identified to be the increased rate of fake tracks. To address this issue, a tools was
2735 developed which was able to successfully identify fake tracks within jets 45% of the
2736 time, with a minimal loss of signal tracks of 1.2%. Removal of such tracks was found
2737 to improve the light-jet mistagging rate of the SV1 and JetFitter algorithms by up
2738 to 20% at high transverse momentum.

2739 A novel approach to b -tagging, GN1 was also developed using a Graph Neural Network
2740 (GNN) architecture. The model is encouraged to learn the topology of the jet through
2741 vertexing and track classification auxiliary tasks. As a single end-to-end trained
2742 model, GN1 simplifies the complexity of the flavour tagging pipeline and is able to
2743 achieve superior performance to the current state-of-the-art algorithms, which rely
2744 on a two-tiered approach. Compared with DL1r, GN1 improves the light-jet rejection

2745 by a factor of ~ 1.8 for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ at the 70%
2746 b -jet WP and by a factor of ~ 6 for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$
2747 for a corresponding b -jet efficiency of 30%. GN1 also demonstrates a significant
2748 improvement in the discrimination between b - and c -jets.

2749 The work in this thesis demonstrates that even with suboptimal track reconstruction
2750 in this regime, it is possible to make algorithmic advancements to the flavour tagging
2751 pipeline to improve the identification of b -jets. This work has impacts for any analysis
2752 which relies on the identification of b -jets, including those which are sensitive to the
2753 Higgs Boson.

2754 Analysis of VH , $H \rightarrow b\bar{b}$ events was also carried out with 139 fb^{-1} of Run 2
2755 ATLAS at $\sqrt{s} = 13 \text{ TeV}$. Various background modelling uncertainties were derived
2756 and investigations into the fit model were carried out. The analysis observed a
2757 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
2758 observed (expected) significance of 2.1σ (2.7σ). The result was validated using a
2759 simultaneous fit to the VZ , $Z \rightarrow b\bar{b}$ process, which acts as a cross check to validate
2760 the primary analysis. The results of the analysis are the most precise measurements
2761 available in the high- p_T for the VH , $H \rightarrow b\bar{b}$ process. The high- p_T region is of
2762 particular interest as it is a region of phase space with good sensitivity to new physics.

2763 8.2 Future Work

2764 Additional algorithmic improvements are likely to yield further improved flavour
2765 tagging performance. Aside from these, large improvements to the flavour tagging
2766 performance will likely be possible if improvements are made to the b -hadron decay
2767 track reconstruction efficiency and accuracy.

2768 At the moment only the tracks from the Inner Detector and kinematic information
2769 about the jet are provided as inputs to the tagging algorithms. In Chapter 6 it
2770 was shown that the addition of a simple track-level variable corresponding to the
2771 ID of the reconstruction lepton to the model improved the performance. However
2772 there is still untapped potential in the form of additional information from the full
2773 parameters of the reconstructed leptons (making full use of the Calorimeters and
2774 Muon Spectrometer), the calorimeter clusters, and even the low level hits. Providing

2775 such additional inputs to the model is likely to complement the information provided
2776 by the tracks and further aid in the improvement of performance.

2777 On the output side, additional auxiliary training objectives may yield improved
2778 performance and also help to add to the explainability of the model. Regression of
2779 jet-level quantities such as the transverse momentum and mass, in addition to the
2780 truth b -hadron decay length are promising regression targets.

2781 The GN1 architecture can also be readily optimised for new use cases and topologies,
2782 as demonstrated by the studies described in Section 6.5. For example, a model
2783 with only hit-level information provided as inputs could be used for a fast trigger
2784 preselection on jets without the need to run the computationally expensive tracking
2785 algorithms. The model could also be repurposed for primary vertexing, or a pile-up
2786 jet tagger. Finally, training a version of GN1 to identify large- R jets would benefit
2787 those analysis investigating such boosted topologies. Ultimately the improved jet
2788 tagging performance enabled by the new algorithm will have a large impact across a
2789 broad spectrum of the ATLAS physics programme.

₂₇₉₀

Bibliography

- ₂₇₉₁ [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- ₂₇₉₂
- ₂₇₉₃ [2] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
₂₇₉₄ *Secondary vertex finding in jets with neural networks*, *The European Physical Journal C* **81** (2021) .
₂₇₉₅
- ₂₇₉₆ [3] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, **ATL-PHYS-PUB-2022-027** (2022).
- ₂₇₉₇
- ₂₇₉₈ [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)
₂₇₉₉ pp. 61–65.
₂₈₀₀
- ₂₈₀₁ [5] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al.,
₂₈₀₂ *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606**
₂₈₀₃ (2022) pp. 479–483.
- ₂₈₀₄ [6] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
₂₈₀₅
- ₂₈₀₆ [7] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995)
₂₈₀₇ pp. 2632–2637 [[hep-ex/9503003](#)].
- ₂₈₀₈ [8] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV*
₂₈₀₉ *Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) pp. 252–255.
- ₂₈₁₀ [9] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy*
₂₈₁₁ *Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122**
₂₈₁₂ (1983) pp. 103–116.
- ₂₈₁₃ [10] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504**
₂₈₁₄ (2001) pp. 218–224 [[hep-ex/0012035](#)].

- 2815 [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2816 *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- 2817 [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.*
2818 **13** (1964) pp. 508–509.
- 2819 [13] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*
2820 *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 2821 [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
2822 *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)
2823 p. 1 [[1207.7214](#)].
- 2824 [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*
2825 *experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 2826 [16] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)
2827 p. 083C01.
- 2828 [17] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*
2829 *invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 2830 [18] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)
2831 pp. 579–588.
- 2832 [19] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 2833 [20] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*
2834 *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**
2835 (1968) pp. 367–377.
- 2836 [21] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
2837 *Phys. Rev.* **104** (1956) pp. 254–258.
- 2838 [22] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hopper and R. P. Hudson, *Experimental*
2839 *test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 2840 [23] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*
2841 *conservation of parity and charge conjugation in meson decays: the magnetic moment*
2842 *of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 2843 [24] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
2844 *Phys. Rev.* **117** (1960) pp. 648–663.

- 2845 [25] J. Goldstone, *Field theories with « superconductor » solutions*, *Il Nuovo Cimento*
2846 (*1955-1965*) **19** (1961) pp. 154–164.
- 2847 [26] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*
2848 *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](https://arxiv.org/abs/1610.07922).
- 2849 [27] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2850 *ATLAS detector*, [ATLAS-CONF-2018-036](https://cds.cern.ch/record/2600000) (2018).
- 2851 [28] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys.*
2852 *Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](https://arxiv.org/abs/1808.08242)].
- 2853 [29] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2854 [30] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of*
2855 *Instrumentation* **3** (2008) pp. S08002–S08002.
- 2856 [31] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008)
2857 p. S08004.
- 2858 [32] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of*
2859 *Instrumentation* **3** (2008) pp. S08005–S08005.
- 2860 [33] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
2861 *JINST* **3** (2008) p. S08003.
- 2862 [34] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data*
2863 *taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2865 [35] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*,
2866 (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2868 [36] CERN, “The accelerator complex.” (2012).
- 2869 [37] G. C. Strong, *On the impact of selected modern deep-learning techniques to the*
2870 *performance and celerity of classification models in an experimental high-energy*
2871 *physics use case*, [2002.01427](https://arxiv.org/abs/2002.01427).
- 2872 [38] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022).
2873 <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoverview/>.
- 2874 [39] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*,

- 2875 *JINST* **11** (2016) p. P04008 [[1512.01094](#)].
- 2876 [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).
- 2878 [41] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*, .
- 2880 [42] for the ATLAS Collaboration collaboration, *Status of the ATLAS Experiment*, Tech. Rep., Geneva (2010). 10.3204/DESY-PROC-2010-04/1.
- 2882 [43] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 2884 [44] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 2886 [45] ATLAS Collaboration, *The inner detector*, (2022).
<https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2888 [46] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 2890 [47] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) p. T05008 [[1803.00844](#)].
- 2892 [48] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, *JINST* **9** (2014) p. P08009 [[1404.7473](#)].
- 2894 [49] ATLAS Collaboration, *Calorimeter*, (2022).
<https://atlas.cern/Discover/Detector/Calorimeter>.
- 2896 [50] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, .
- 2897 [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.
- 2899 [52] ATLAS Collaboration, *Muon spectrometer*, (2022).
<https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2901 [53] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) p. 317 [[1611.09661](#)].
- 2903 [54] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, **ATL-SOFT-PUB-2021-001** (2021).

- 2905 [55] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts,
2906 Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2907 [56] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense
2908 Environments*, **ATL-PHYS-PUB-2015-006** (2015).
- 2909 [57] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in
2910 dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].
- 2911 [58] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon
2912 pixel detector*, *JINST* **9** (2014) p. P09009 [[1406.7690](#)].
- 2913 [59] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS
2914 Experiment*, .
- 2915 [60] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment
2916 in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332
2917 [[1611.10235](#)].
- 2918 [61] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency
2919 measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79**
2920 (2019) p. 970 [[1907.05120](#)].
- 2921 [62] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP*
2922 **04** (2008) p. 063 [[0802.1189](#)].
- 2923 [63] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur.Phys.J. C* **72** (2012)
2924 p. 1896 [[1111.6097](#)].
- 2925 [64] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04**
2926 (2008) p. 005 [[0802.1188](#)].
- 2927 [65] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS
2928 detector*, **ATLAS-CONF-2014-018** (2014).
- 2929 [66] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its
2930 performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 2931 [67] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with
2932 the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 2933 [68] ATLAS Collaboration, *Jet energy scale measurements and their systematic
2934 uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
2935 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].

- 2936 [69] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*
2937 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)
2938 p. 242001 [[0802.2470](#)].
- 2939 [70] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*
2940 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
2941 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2942 p. 136204 [[2008.02508](#)].
- 2943 [71] ATLAS collaboration, *Performance of jet substructure techniques for large-R jets in*
2944 *proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
2945 p. 076 [[1306.4945](#)].
- 2946 [72] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*
2947 *13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019)
2948 p. 135 [[1807.09477](#)].
- 2949 [73] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) p. 059
2950 [[0903.0392](#)].
- 2951 [74] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
2952 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010
2953 (2017).
- 2954 [75] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*
2955 *Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).
- 2956 [76] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
2957 *using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024 (2016).
- 2958 [77] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*
2959 *proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) p. 292
2960 [[1603.05598](#)].
- 2961 [78] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
2962 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys.*
2963 *J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 2964 [79] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)
2965 pp. 577–606 [[hep-ph/9912292](#)].
- 2966 [80] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**
2967 (2018) p. 030001.

- 2968 [81] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom*
2969 *and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets*,
2970 **ATL-PHYS-PUB-2014-008** (2014).
- 2971 [82] Nazar Bartosik, *Diagram showing the common principle of identification of jets*
2972 *initiated by b-hadron decays*, (2022).
2973 https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png.
- 2974 [83] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
2975 *b-tagging algorithms for the 2017-18 LHC run*, **ATL-PHYS-PUB-2017-013** (2017).
- 2976 [84] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*
2977 *ATLAS detector*, **ATL-PHYS-PUB-2017-011** (2017).
- 2978 [85] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
2979 *Neural Networks at the ATLAS Experiment*, **ATL-PHYS-PUB-2017-003** (2017).
- 2980 [86] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*
2981 *Flavour Tagging in ATLAS*, **ATL-PHYS-PUB-2020-014** (2020).
- 2982 [87] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*
2983 *Run-2*, **ATL-PHYS-PUB-2015-022** (2015).
- 2984 [88] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).
2985 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2022-03/>.
- 2986 [89] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for*
2987 *New Physics in the Fully Hadronic Final State*, Ph.D. thesis, Geneva U., 2021.
2988 [10.13097/archive-ouverte/unige:158558](https://doi.org/10.13097/archive-ouverte/unige:158558).
- 2989 [90] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*
2990 *Process Interface for Event Generators*, *ArXiv High Energy Physics - Phenomenology*
2991 *e-prints* (2001) .
- 2992 [91] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*
2993 *standard format for Les Houches Event Files*, *Computer Physics Communications*
2994 **176** (2007) pp. 300–304.
- 2995 [92] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*
2996 *universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 2997 [93] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*
2998 *back-propagating errors*, *nature* **323** (1986) pp. 533–536.

- 2999 [94] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*
3000 *activity, The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 3001 [95] J. Hopfield, *Neural networks and physical systems with emergent collective*
3002 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*
3003 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 3004 [96] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*
3005 *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 3006 [97] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*
3007 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 3008 [98] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,
3009 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,
3010 *arXiv e-prints* (2018) p. arXiv:1806.01261 [[1806.01261](#)].
- 3011 [99] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
3012 *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 3013 [100] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,
3014 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 3015 [101] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*
3016 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007)
3017 p. 126–126.
- 3018 [102] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*
3019 *shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007)
3020 p. 070–070.
- 3021 [103] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*
3022 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*
3023 *Physics* **2010** (2010) .
- 3024 [104] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)
3025 p. 040 [[1410.8849](#)].
- 3026 [105] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
3027 *ATL-PHYS-PUB-2016-020* (2016).
- 3028 [106] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*
3029 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159

- 3030 [1410.3012].
- 3031 [107] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
3032 [ATL-PHYS-PUB-2014-021](#) (2014).
- 3033 [108] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)
3034 p. 244 [[1207.1303](#)].
- 3035 [109] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*
3036 **462** (2001) p. 152.
- 3037 [110] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**
3038 (2010) p. 823 [[1005.4568](#)].
- 3039 [111] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*
3040 *Instrum. Meth. A* **506** (2003) p. 250.
- 3041 [112] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
3042 *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 3043 [113] ATLAS Collaboration, *Observation of Higgs boson production in association with a*
3044 *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173
3045 [[1806.00425](#)].
- 3046 [114] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*
3047 *using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **03**
3048 (2020) p. 145 [[1910.08447](#)].
- 3049 [115] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
3050 *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772
3051 [[2002.08772](#)].
- 3052 [116] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13\text{ TeV}$ data*,
3053 [ATL-PHYS-PUB-2015-037](#) (2015).
- 3054 [117] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*
3055 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
3056 *$\sqrt{s} = 13\text{ TeV}$* , *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 3057 [118] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
3058 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, *arXiv e-prints* (2020)
3059 p. arXiv:2007.08294 [[2007.08294](#)].
- 3060 [119] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,

- 3061 *arXiv e-prints* (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 3062 [120] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al.,
3063 *Attention Is All You Need*, *arXiv e-prints* (2017) p. arXiv:1706.03762 [[1706.03762](#)].
- 3064 [121] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*
3065 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 3066 [122] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
3067 <https://github.com/onnx/onnx>.
- 3068 [123] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp*
3069 *collision dataset*, [2211.16345](#).
- 3070 [124] ATLAS Collaboration, *Performance of Run 3 HLT b-tagging with fast tracking,*
3071 *Public b-Jet Trigger Plots for Collision Data*, (2022).
3072 <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/>
3073 *BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi*.
- 3074 [125] ATLAS Collaboration, *Neural Network Jet Flavour Tagging with the Upgraded*
3075 *ATLAS Inner Tracker Detector at the High-Luminosity LHC*,
3076 [ATL-PHYS-PUB-2022-047](#) (2022).
- 3077 [126] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs*
3078 *Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- 3079 [127] ATLAS Collaboration, *Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$*
3080 *decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C*
3081 **81** (2021) p. 178 [[2007.02873](#)].
- 3082 [128] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson*
3083 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
3084 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
3085 p. 136204 [[2008.02508](#)].
- 3086 [129] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated*
3087 *Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039
3088 [[1512.02572](#)].
- 3089 [130] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in*
3090 *proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
3091 p. 076 [[1306.4945](#)].

- 3092 [131] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO*
3093 *calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010)
3094 p. 043 [[1002.2581](#)].
- 3095 [132] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum*
3096 *distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09**
3097 (2014) p. 145 [[1406.3660](#)].
- 3098 [133] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al.,
3099 *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) p. 1889
3100 [[1111.2034](#)].
- 3101 [134] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*,
3102 *JHEP* **10** (2012) p. 155 [[1206.3572](#)].
- 3103 [135] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at*
3104 *NLO with the POWHEG BOX interfaced to GoSam and their merging within*
3105 *MinLO*, *JHEP* **10** (2013) p. 083 [[1306.2542](#)].
- 3106 [136] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to the*
3107 *associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003)
3108 p. 073003 [[hep-ph/0306234](#)].
- 3109 [137] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the*
3110 *Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) pp. 149–156
3111 [[hep-ph/0307206](#)].
- 3112 [138] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production*
3113 *at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.*
3114 **107** (2011) p. 152003 [[1107.1164](#)].
- 3115 [139] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects*
3116 *in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- 3117 [140] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated*
3118 *WH production and decay at the LHC*, *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- 3119 [141] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron*
3120 *colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015)
3121 pp. 51–55 [[1407.4747](#)].
- 3122 [142] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson*
3123 *at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].

- 3124 [143] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
 3125 *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078
 3126 [[1211.5015](#)].
- 3127 [144] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via*
 3128 *gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- 3129 [145] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for*
 3130 *gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- 3131 [146] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron*
 3132 *Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- 3133 [147] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron*
 3134 *colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- 3135 [148] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte*
 3136 *Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].
- 3137 [149] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair*
 3138 *cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) p. 2930
 3139 [[1112.5675](#)].
- 3140 [150] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with*
 3141 *shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) p. 111
 3142 [[0907.4076](#)].
- 3143 [151] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top*
 3144 *quark production*, *Phys. Rev. D* **81** (2010) p. 054028 [[1001.5034](#)].
- 3145 [152] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for*
 3146 *t-channel single top quark production*, *Phys. Rev. D* **83** (2011) p. 091503 [[1103.2792](#)].
- 3147 [153] E. Re, *Single-top Wt-channel production matched with parton showers using the*
 3148 *POWHEG method*, *Eur. Phys. J. C* **71** (2011) p. 1547 [[1009.2450](#)].
- 3149 [154] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated*
 3150 *production with a W^- or H^-* , *Phys. Rev. D* **82** (2010) p. 054018 [[1005.4451](#)].
- 3151 [155] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al.,
 3152 *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) p. 007 [[0811.4622](#)].
- 3153 [156] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) p. 034
 3154 [[1905.09127](#)].

- 3155 [157] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
3156 *Phys. Rev. Lett.* **108** (2012) p. 111601 [[1111.5206](#)].
- 3157 [158] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008)
3158 p. 039 [[0808.3674](#)].
- 3159 [159] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour*
3160 *dipole factorisation*, *JHEP* **03** (2008) p. 038 [[0709.1027](#)].
- 3161 [160] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton*
3162 *showers. The NLO case*, *JHEP* **04** (2013) p. 027 [[1207.5030](#)].
- 3163 [161] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson*
3164 *Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.*
3165 *Rev. Lett.* **103** (2009) p. 082001 [[0903.2120](#)].
- 3166 [162] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**
3167 (2016) p. 023001 [[1510.03865](#)].
- 3168 [163] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to*
3169 *Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* **03**
3170 (2012) p. 075 [[1112.5142](#)].
- 3171 [164] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo*
3172 *program for Higgs production in vector-boson fusion and Higgs strahlung at hadron*
3173 *colliders*, *Comput. Phys. Commun.* **195** (2015) pp. 161–171 [[1412.5390](#)].
- 3174 [165] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS*
3175 *detector using t̄t events at √s = 13 TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 3176 [166] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using*
3177 *ATLAS proton–proton collision data at √s = 13 TeV*, *ATLAS-CONF-2018-006*
3178 (2018).
- 3179 [167] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in t̄t events using*
3180 *a likelihood approach with the ATLAS detector*, *ATLAS-CONF-2018-001* (2018).
- 3181 [168] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*
3182 *ATLAS Detector*, *ATL-PHYS-PUB-2014-013* (2014).
- 3183 [169] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark*
3184 *pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836
3185 [[1906.11005](#)].

- 3186 [170] J. K. Anders and M. D'Onofrio, *V+Jets theoretical uncertainties estimation via a*
3187 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 3188 [171] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*
3189 *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J.* **C75** (2015) p. 204
3190 [[1412.3989](#)].
- 3191 [172] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*
3192 *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*
3193 **D93** (2016) p. 033006 [[1506.07443](#)].
- 3194 [173] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at*
3195 $\sqrt{s} = 13\text{ TeV}$ *with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016)
3196 p. 182002 [[1606.02625](#)].
- 3197 [174] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in*
3198 *association with a vector boson and decaying to a pair of b-quarks using large-R jets*,
3199 *ATL-COM-PHYS-2019-1125* (2019).
- 3200 [175] J. Pumplin et al., *New generation of parton distributions with uncertainties from*
3201 *global qcd analysis*, *JHEP* **07** (2002) p. 012 [[0201195](#)].
- 3202 [176] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top*
3203 *hadroproduction in association with a W boson*, *JHEP* **07** (2008) p. 029 [[0805.3067](#)].
- 3204 [177] ATLAS Collaboration, *Measurement of VH, $H \rightarrow b\bar{b}$ production as a function of the*
3205 *vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*,
3206 *JHEP* **05** (2019) p. 141 [[1903.04618](#)].
- 3207 [178] ATLAS Collaboration, *Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector*,
3208 *JHEP* **12** (2017) p. 024 [[1708.03299](#)].
- 3209 [179] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino
3210 and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential*
3211 *Distributions*, *CERN-2012-002* (2012) [[1201.3084](#)].
- 3212 [180] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3.*
3213 *Higgs Properties*, *CERN-2013-004* (2013) [[1307.1347](#)].
- 3214 [181] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*
3215 *roostats project*, *arXiv preprint arXiv:1009.1003* (2010) .
- 3216 [182] J. R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model*

- ³²¹⁷ *Working Group Report*, in *9th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2015) Les Houches, France, June 1-19, 2015*, 2016, [1605.04692](#).

3219 List of figures

<small>3220</small>	2.1	The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin.	<small>3221</small> 21
<small>3223</small>	2.2	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.	<small>3224</small> 26
<small>3228</small>	2.3	Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.	<small>3229</small> 26
<small>3230</small>	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.	<small>3231</small> 27
<small>3234</small>	3.1	An overview of the CERN accelerator complex [36]. The LHC is fed by a series of accelerators starting with Linac2 (or Linac4 from 2020). Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.	<small>3235</small> 30
<small>3238</small>	3.2	The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [37].	<small>3239</small> 31
<small>3240</small>			<small>3241</small>

3242	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
3248	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
3252	3.5 Average pile-up profiles measured by ATLAS during Run 2 [35].	35
3253	3.6 A 3D model of the entire ATLAS detector [42]. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.	36
3256	3.7 A 3D model of the ATLAS ID showing the pixel, SCT and TRT subdetectors [45].	37
3258	3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [38].	38
3260	3.9 A schematic cross-sectional view of the ATLAS IBL [46].	39
3261	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
3264	3.11 The ATLAS muon spectrometer [52].	43
3265	3.12 Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are recon- structed as a single merged cluster [57].	47
3269	3.13 A sketch of electron reconstruction using the ATLAS detector [76]. Electron reconstruction makes use of the entire ID and the calorimeters.	52

3271	4.1 The truth b -hadron decay radius L_{xy} as a function of truth transverse 3272 momentum p_T for reconstructed b -jets in Z' events. Error bars show the 3273 standard deviation. The pixel layers are shown in dashed horizontal lines.	56
3274	4.2 Diagram of a typical b -jet (blue) which has been produced in an event 3275 alongside two light jets (grey) [82]. The b -hadron has travelled a significant 3276 distance (pink dashed line) from the primary interaction point (pink dot) 3277 before its decay. The large transverse impact parameter d_0 is a characteristic 3278 property of the trajectories of b -hadron decay products.	56
3279	4.3 At lower p_T (left) the decay length of the b -hadron is on average reduced, and 3280 the decay tracks are less collimated. At higher p_T (right) the b -hadron decay 3281 length increases and the resulting decay tracks are more collimated and have 3282 less distance over which to diverge before reaching detector elements. As a 3283 result, the ID may be unable to resolve charged depositions from different 3284 particles, resulting merged clusters.	58
3285	4.4 b -hadron decay track reconstruction efficiency as a function of truth b -hadron 3286 p_T [88]. Nominal track reconstruction is shown in black, while the track 3287 reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver 3288 efficiency) is shown in green. For high- p_T b -hadrons, the ambiguity solver 3289 is overly aggressive in its removal of b -hadron decay tracks. Suggestions for 3290 the improvement of the track reconstruction efficiency in this regime by the 3291 loosening of cuts in the ambiguity solver are shown in blue and red. . . .	60
3292	4.5 Hit multiplicities on the IBL (left) and the pixel layers (right) as a function 3293 of the p_T of the reconstructed track for tracks in jets in a Z' sample at 3294 $\sqrt{s} = 13$ TeV. Tracks from the weak decay of the b -hadron are shown in red, 3295 while fragmentation tracks (which are prompt) are in blue. Baseline tracks 3296 are those produced in the standard reconstruction described in Section 3.4.1, 3297 while pseudotrack represent the ideal performance of the ATLAS detector 3298 and are described in Section 3.4.1.	60
3299	4.6 The fraction of IBL (left) and B-layer (right) hits which are shared on 3300 b -hadron decay tracks as a function of the production radius of the b -hadron 3301 decay product for tracks in jets in a Z' sample at $\sqrt{s} = 13$ TeV. Pseudotracks 3302 represent the ideal performance given the ATLAS detector, see Section 3.4.1.	62

3303	4.7	The average number of 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).	64
3304			
3305			
3306	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 the truth b -hadron transverse momentum.	65
3307			
3308			
3309	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.	69
3310			
3311			
3312			
3313	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.	70
3314			
3315			
3316	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].	73
3317			
3318			
3319	5.4	Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track}, \text{jet})$ (right) for jets in the Z' sample. The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases.	75
3320			
3321			
3322			
3323	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.	75
3324			
3325			
3326			
3327			
3328			

3329	5.6 (left) Normalised histograms of the fake track classification model output 3330 separated for signal and fake tracks, and further separated by those tracks 3331 which are from the decay of a b -hadron. (right) The ROC curve for all 3332 tracks (solid line) and tracks from the decay of a b -hadron (dashed line). 3333 The plots show tracks in the combined $t\bar{t}$ and Z' testing sample. The model 3334 is able to successfully discriminate between signal and fake tracks, and 3335 shows a very similar performance when looking specifically at tracks from 3336 the decay of a b -hadron.	79
3337	5.7 (left) Normalised histogram of the b -hadron track identification model 3338 output separated for tracks from the decay of a b -hadron and tracks from 3339 other sources. The two groups are further separated into those tracks which 3340 are fake. (right) The ROC curve for all tracks (solid line). The plots show 3341 tracks in the combined $t\bar{t}$ and Z' testing sample.	81
3342	5.8 The effect of applying the fake track identification algorithm together with 3343 the b -hadron decay track identification on the jet tagging performance of 3344 SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and 3345 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) 3346 is compared to two working points of fake track removal, labelled “A” (red) 3347 and “B” (orange), which represent two different 2D working points of the 3348 track classification tools. Removal of fake tracks based on truth information 3349 is shown by the green curves.	82
3350	6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right) [3]. 3351 The existing approach utilises low-level algorithms (shown in blue), the 3352 outputs of which are fed into a high-level algorithm (DL1r). Instead 3353 of being used to guide the design of the manually optimised algorithms, 3354 additional truth information from the simulation is now being used as 3355 auxiliary training targets for GN1. The solid lines represent reconstructed 3356 information, whereas the dashed lines represent truth information.	85
3357	6.2 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of n_{tracks} , 3358 where each track is described by 21 track features ($n_{\text{tf}} = 21$) [3]. The jet 3359 features are copied for each of the tracks, and the combined jet-track vectors 3360 of length 23 form the inputs of GN1.	92

3361	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.	92
3362		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.	
3363			
3364			
3365			
3366			
3367	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.	97
3368			
3369			
3370			
3371			
3372			
3373	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.	100
3374			
3375			
3376			
3377			
3378			
3379			
3380	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.	100
3381			
3382			
3383			
3384			
3385			
3386			
3387	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.	101
3388			
3389			
3390			
3391			
3392			

3393	6.8 The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV [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.	103
3394		
3395		
3396		
3397		
3398		
3399	6.9 The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV [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.	103
3400		
3401		
3402		
3403		
3404		
3405	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.	105
3406		
3407		
3408		
3409		
3410		
3411		
3412	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.	105
3413		
3414		
3415		
3416		
3417		
3418		
3419	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.	107
3420		
3421		
3422		
3423		
3424		
3425		
3426		
3427		

3428 6.13 A schematic showing the truth track origin and vertex information (left)
3429 compared with the predictions from GN1 (right). The diagrams show the
3430 truth (left) and predicted (right) structure of a b -jet. The shaded black boxes
3431 show the groupings of tracks into different vertices. Vertices reconstructed
3432 by SV1 and JetFitter are also marked. Class probabilities p_b , p_c and p_u are
3433 rounded to three significant figures. GN1 improves over SV1 and JetFitter
3434 by successfully determining the origins and vertex compatibility of all but
3435 one tracks in the jet. 107

3436 6.14 Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in
3437 the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding
3438 approaches. The exclusive vertexing approach includes single track vertices
3439 while the inclusive approach requires vertices to contain at least two tracks.
3440 Efficient vertex finding requires the recall of at least 65% of the tracks in
3441 the truth vertex, and allows no more than 50% of the tracks to be included
3442 incorrectly. Binomial error bands are denoted by the shaded regions. 110

3443 6.15 Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in
3444 the Z' sample using the exclusive (left) and inclusive (right) vertex finding
3445 approaches. The exclusive vertexing approach includes single track vertices
3446 while the inclusive approach requires vertices to contain at least two tracks.
3447 Efficient vertex finding requires the recall of at least 65% of the tracks in
3448 the truth vertex, and allows no more than 50% of the tracks to be included
3449 incorrectly. Binomial error bands are denoted by the shaded regions. 110

3450 6.16 Vertex finding efficiency as a function of jet p_T for light-jets in the $t\bar{t}$ sample
3451 using the exclusive (left) and inclusive (right) vertex finding approaches.
3452 The exclusive vertexing approach includes single track vertices while the
3453 inclusive approach requires vertices to contain at least two tracks. Efficient
3454 vertex finding requires the recall of at least 65% of the tracks in the truth
3455 vertex, and allows no more than 50% of the tracks to be included incorrectly.
3456 Binomial error bands are denoted by the shaded regions. 111

3457	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.	112
3464	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.	114
3471	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 [123] 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.	116
3477	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 [123] 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.	116
3483	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 [124]. The ratio to the performance of the DL1d algorithm [123] 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.	117

3489	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 [125]. 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.	118
3490		
3491		
3492		
3493		
3494		
3495	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 [125]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions.	118
3496		
3497		
3498		
3499		
3500	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.	129
3501		
3502		
3503		
3504		
3505		
3506	7.2 Leading large- R jet mass for the Z and $W + \text{hf}$ processes in the HP SR of the 0-lepton channel [174]. 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.	140
3507		
3508		
3509		
3510	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 [174]. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.	141
3511		
3512		
3513	7.4 Leading large- R jet mass distribution for the WZ and ZZ processes, inclusive across all signal regions and lepton channels [174]. The renormalisation scale 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.	143
3514		
3515		
3516		
3517	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 [174]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after a protection is added in the high mass region.	144
3518		
3519		
3520		
3521		
3522		

3523	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 [128].	152
3533	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 [128].	153
3544	7.8 m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated WH and ZH production of a SM Higgs boson with $m_H = 125 \text{ GeV}$ is shown scaled by the measured combined signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.2, except for the VH/VZ experimental and theory uncertainties [128].	156

3556	7.9 Signal strength compatibility test between the (3+1)-POI fit (with the three 3557 lepton channels splitted) and the default (1+1)-POI fit. The compatibility 3558 of the three channels is evaluated via a χ^2 difference test and results in a 3559 p-value of 49% [128].	157
3560	7.10 Impact of systematic uncertainties on the fitted VH signal-strength pa- 3561 rameter μ_{VH}^{bb} sorted in decreasing order. The boxes show the variations 3562 of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual 3563 nuisance parameter to its post-fit value modified upwards or downwards by 3564 its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of 3565 up- and down-variations can be distinguished via the dashed and plane box 3566 fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying 3567 each nuisance parameter by ± 1 . The filled circles show the deviation of the 3568 fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value 3569 θ_0 expressed in standard deviations with respect to their nominal uncer- 3570 tainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties 3571 on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show 3572 the fitted values and uncertainties of the normalization parameters that are 3573 freely floating in the fit. Pre-fit, these parameters have a value of one [174].	159
3574	7.11 Measured VH simplified template cross sections times the $H \rightarrow b\bar{b}$ and 3575 $V \rightarrow$ leptons branching fractions in the medium and high $p_T^{V,t}$ bins [128].	162

3576 List of tables

<small>3577</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.	15
<small>3578</small>		
<small>3579</small>		
<small>3580</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.	16
<small>3581</small>		
<small>3582</small>		
<small>3583</small>		
<small>3584</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. The luminosity is the peak instantaneous luminosity. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.	29
<small>3585</small>		
<small>3586</small>		
<small>3587</small>		
<small>3588</small>		
<small>3589</small>	3.2 Quality selections applied to tracks, where d_0 is the transverse IP of the track, z_0 is the longitudinal IP with respect to the PV and θ is the track polar angle (see Section 3.2.2 for the IP definitions). Silicon hits are hits on the pixel and SCT layers. 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.	46
<small>3590</small>		
<small>3591</small>		
<small>3592</small>		
<small>3593</small>		
<small>3594</small>		
<small>3595</small>		

3596 5.1	Truth origins which are used to categorise the physical 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 0.75 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.	74
3603 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.	77
3609 5.3	Hyperparameter for the track classification model	79
3610 5.4	Good and fake track selection efficiencies for the combined $t\bar{t}$ and Z' samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12.	80
3612 5.5	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).	81
3616 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.	89
3624 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.	94

3627 6.3 The area under the ROC curves (AUC) for the track classification from
3628 GN1, compared to a standard multilayer perceptron (MLP) trained on a
3629 per-track basis. The unweighted mean AUC over the origin classes and
3630 weighted mean AUC (using as a weight the fraction of tracks from the
3631 given origin) is provided. GN1, which uses an architecture that allows track
3632 origins to be classified in a conditional manner as discussed in Section 6.3.3,
3633 outperforms the MLP model for both $t\bar{t}$ and Z' jets. 113

3634 7.1 Summary of the definitions of the different analysis regions . Signal enriched
3635 regions are marked with the label SR. There are regions with relatively
3636 large signal purity (HP SR) and with low purity (LP SR). Background
3637 enriched regions are marked with the label CR. The shorthand “add” stands
3638 for additional small- R jets, i.e. number of small- R jets not matched to the
3639 Higgs-jet candidate. The medium and high p_T^V regions are referred to as
3640 Mp_T^V and Hp_T^V , respectively [128]. 123

3641 7.2 Signal and background processes with the corresponding generators used for the
3642 nominal samples [128]. If not specified, the order of the cross-section calculation
3643 refers to the expansion in the strong coupling constant (α_s). (★) The events
3644 were generated using the first PDF in the NNPDF3.0NLO set and subsequently
3645 reweighted to the PDF4LHC15NLO set [162] using the internal algorithm in
3646 PowHEG-Box v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for
3647 the $pp \rightarrow ZH$ process already includes the $gg \rightarrow ZH$ contribution. The $qq \rightarrow ZH$
3648 process is normalised using the cross-section for the $pp \rightarrow ZH$ process, after
3649 subtracting the $gg \rightarrow ZH$ contribution. An additional scale factor is applied to
3650 the $qq \rightarrow VH$ processes as a function of the transverse momentum of the vector
3651 boson, to account for electroweak (EW) corrections at NLO. This makes use of
3652 the VH differential cross-section computed with HAWK [163, 164]. 125

3653 7.3 Selections applied to baseline and signal electrons and muons. 126

3654	7.4 Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels and sub-channels [128]. 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 $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T, \text{trk}})$ 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.	128
3664	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.	132
3667	7.6 0-lepton $W + \text{jets}$ nominal sample flavour composition and total event yield [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).	136
3670	7.7 1-lepton $W + \text{jets}$ nominal sample flavour composition and total event yield [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).	136
3673	7.8 0-lepton $Z + \text{jets}$ nominal sample flavour composition and total event yield [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).	137
3676	7.9 2-lepton $Z + \text{jets}$ nominal sample flavour composition and total event yield [174]. Mp_T^V refers to the medium p_T^V region, and Hp_T^V refers to the high p_T^V region (see Table 7.1).	137
3679	7.10 $V + \text{jets}$ acceptance uncertainties [174]. $W + \text{jets}$ SR and CR uncertainties marked with a superscript † are correlated. The 1L $W + \text{jets}$ H/M uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions.	138
3683	7.11 Diboson acceptance uncertainties [174]. All uncertainties except channel extrapolation uncertainties are fully correlated between ZZ and WZ processes and channels.	143

3686	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 [128]. The errors represent the combined statistical and systematic uncertainties.	151
3689	7.13 Post-fit yields in the 0-lepton channel. Combined statistical and systematic uncertainties are shown [174].	154
3691	7.14 Post-fit yields in the 1-lepton channel. Combined statistical and systematic uncertainties are shown [174].	154
3693	7.15 Post-fit yields in the 2-lepton channel. Combined statistical and systematic uncertainties are shown [174].	155
3695	7.16 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 [174].	160