

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**

November 15, 2022

Declaration

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

Samuel Van Stroud

Abstract

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

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

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

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

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

Impact Statement

impact statement 500 words [link to ucl info](#)

Acknowledgements

Here is an example of how to declare commands for use in a single file that will not be needed elsewhere. Additionally, it serves to illustrate the chapter referencing system.

Brian Moser and Francesco Di Bello provided useful guidance and insights into the studies in Chapter 7.

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

Contents

1	Introduction	11
2	Theoretical Framework	12
2.1	The Standard Model	12
2.1.1	Quantum Electrodynamics	13
2.1.2	Quantum Chromodynamics	15
2.1.3	The Electroweak Sector	16
2.2	The Higgs Mechanism	17
2.2.1	Electroweak Symmetry Breaking	18
2.2.2	Fermionic Yukawa Coupling	21
2.2.3	Higgs Sector Phenomenology	23
3	The Large Hadron Collider and the ATLAS Detector	26
3.1	The Large Hadron Collider	27
3.2	Coordinate System & Collider Definitions	28
3.2.1	ATLAS Coordinate System	29
3.2.2	Track Parameterisation	30
3.2.3	Hadron Collider Definitions	30
3.3	The ATLAS Detector	34
3.3.1	Inner Detector	35
3.3.2	Calorimeters	38
3.3.3	Muon Spectrometer	40
3.3.4	The Trigger	41
3.4	Reconstructed Physics Objects	42
3.4.1	Tracks	42
3.4.2	Vertices	44
3.4.3	Jets	46
3.4.4	Leptons	48

3.4.5	Missing Transverse Momentum	49
4	Tracking and b-tagging	51
4.1	b -hadron Reconstruction	52
4.1.1	Decay Topology	52
4.1.2	Challenges Facing b -hadron Reconstruction	54
4.2	Investigations into High p_T b -hadron Tracking	57
4.2.1	Pseudo-tracking	58
4.2.2	Global χ^2 Fitter Outlier Removal	58
4.3	Conclusion	61
5	Track Classification MVA	63
5.1	Machine Learning Background	63
5.1.1	Neural Networks	65
5.1.2	Training with Gradient Descent	66
5.2	Track Truth Origin Labelling	67
5.3	Fake Track Identification Tool	68
5.3.1	Datasets	69
5.3.2	Model Inputs	71
5.3.3	Results	71
5.4	Conclusion	75
6	Graph Neural Network Flavour Tagger	78
6.1	Motivation	79
6.2	Graph Neural Network Theory	82
6.3	Experimental Setup	82
6.3.1	Datasets	82
6.4	Model Architecture	82
6.4.1	Model Inputs	82
6.4.2	Auxiliary Training Objectives	83
6.4.3	Architecture	85
6.4.4	Training	89
6.5	Results	90
6.5.1	b -tagging Performance	91
6.5.2	c -tagging Performance	96
6.5.3	Jet Display Diagrams	96

6.5.4	Ablations	96
6.5.5	Inclusion of Low-Level Vertexing Algorithms	99
6.5.6	Vertexing Performance	102
6.5.7	Track Classification Performance	105
6.6	Extensions	107
6.6.1	Looser Track Selection	107
6.6.2	Model Optimisations	108
6.6.3	Additional Auxiliary Tasks	108
6.7	Conclusion	108
7	Boosted VHbb Analysis	111
7.1	Analysis Overview	112
7.1.1	Data & Simulated Samples	113
7.1.2	Object Reconstruction	114
7.1.3	Selection Criteria	116
7.1.4	Control Regions	118
7.1.5	Background Composition	118
7.2	Systematic Uncertainties & Background Modelling	120
7.2.1	Sources of Systematic Uncertainties	121
7.2.2	Implementation of Variations	123
7.2.3	Vector Boson + Jets Modelling	124
7.2.4	Diboson Modelling	131
7.3	Statistical Treatment	133
7.3.1	Likelihood Function	134
7.3.2	Background Normalisations	136
7.3.3	Asimov Dataset & Expected Results	137
7.4	Results	137
7.4.1	Post-fit Distributions	138
7.4.2	Observed Signal Strength & Significance	138
7.4.3	Impact of Systematics	142
7.5	Improved b -tagging using GNNs	145
7.6	Conclusion	145
8	Conclusion	147
Bibliography		148

Contents	10
-----------------	-----------

List of figures	161
------------------------	------------

List of tables	172
-----------------------	------------

² Chapter 1

³ Introduction

⁴ This thesis describes various efforts in improving the understanding of the Higgs
⁵ boson and its coupling to b - and c -quarks, primarily through the improvement of
⁶ algorithmic tools, discussed in Chapters ⁵ and ⁶.

⁷ Analysis work on $H \rightarrow b\bar{b}$ decays is also described in Chapter ⁷.

⁸ Chapter 2

⁹ Theoretical Framework

¹⁰ The Standard Model (SM) of particle physics is the theory describing all known
¹¹ elementary particles and their interactions via three of the four fundamental forces.
¹² Developed by merging the successful theories of quantum mechanics and relativity
¹³ in the second half of the 20th century, the SM's position today at the centre of our
¹⁴ understanding of the nature of the Universe is firmly established by an unparalleled
¹⁵ level of agreement between the predictions from the model and experimental results
¹⁶ [2, 3].

¹⁷ The SM has predicted the discovery of the top and bottom quarks [4–6], the W
¹⁸ and Z bosons [7], and the tau neutrino [8]. The last missing piece of the SM to be
¹⁹ discovered was the Higgs boson, first theorised in the 1960s [9–11], and eventually
²⁰ observed at the LHC in 2012 [12, 13]. After its discovery, much ongoing work has
²¹ been carried out performing detailed measurements of its mass and interactions with
²² other particles.

²³ In this chapter, an overview of the SM is given in Section 2.1, and a more detailed
²⁴ discussion of the Higgs sector and Higgs phenomenology is provided in Section 2.2.

²⁵ 2.1 The Standard Model

²⁶ The SM is formulated in the language of Quantum Field Theory (QFT). In this
²⁷ framework, particles are localised excitations of corresponding quantum fields, which
²⁸ are operator-valued distributions across spacetime.

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

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

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

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

Table 2.1: The fermions of the SM [14]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

40

2.1.1 Quantum Electrodynamics

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

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

Table 2.2: The bosons of the SM [14]. The photon, weak bosons and gluons are gauge bosons arising from gauge symmetries, and carry the four fundamental forces of the SM. The recently discovered Higgs boson is the only fundamental scalar particle in the SM.

⁴⁶ Dirac Lagrangian density is

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

⁴⁷ 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

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

⁵⁰ Suppose some fundamental symmetry that requires invariance under a local $U(1)$
⁵¹ gauge transformation

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

⁵² where α varies over every spacetime point x . Under this transformation, the Dirac
⁵³ equation transforms as

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

⁵⁴ For the Dirac equation to remain invariant under the transformation in Eq. (2.4),
⁵⁵ a new field A_μ which transforms as $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$ must be added. The
⁵⁶ transformed interaction term

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

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

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

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

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interactions. Quarks and gluons carry colour charge, which comes in three kinds, called red, green and blue. While the $U(1)$ symmetry group in Section 2.1.1 was Abelian, the QCD Lagrangian is specified by requiring invariance under transformations from the non-Abelian $SU(3)$ group, making it a Yang-Mills theory [15] which requires the addition of self-interacting gauge fields. The infinitesimal $SU(3)$ group generators are given by $T_a = \lambda_a/2$, where λ_a are the eight Gell-Mann matrices. These span the space of infinitesimal group transformations and do not commute with each other, instead satisfying the commutation relation

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

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

⁸³ strength tensors, one for each generator T_a , defined as

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

⁸⁴ where A_μ^a are the gluon fields and g_s is the strong coupling constant. The covariant
⁸⁵ derivative is written as

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

⁸⁶ The full QCD Lagrangian is then given by

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

⁸⁷ Cubic and quartic terms of the gauge fields A_μ^a appear in the Lagrangian, leading to
⁸⁸ the gluon's self interaction.

⁸⁹ The QCD coupling constant g_s varies, or “runs”, with energy. At lower energy
⁹⁰ scales (and corresponding larger distance scales) the interaction is strong. This
⁹¹ leads to quark confinement, whereby an attempt to isolate individual colour-charged
⁹² quarks requires so much energy that additional quark-antiquark are produced. At
⁹³ higher energy scales (and corresponding smaller distance scales), asymptotic freedom
⁹⁴ occurs as the interactions become weaker, allowing perturbative calculations to be
⁹⁵ performed. Hadrons are bound states of quarks. They are invariant under $SU(3)$
⁹⁶ gauge transformations (i.e. are colour-charge neutral, or *colourless*).

⁹⁷ 2.1.3 The Electroweak Sector

⁹⁸ The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam
⁹⁹ (GWS) model of electroweak interaction [16–18]. The Lagrangian is specified by
¹⁰⁰ requiring invariance under the symmetry group $SU(2)_L \otimes U(1)_Y$, as motivated by a
¹⁰¹ large amount of experimental data. Here, $SU(2)_L$ is referred to as weak isospin and
¹⁰² $U(1)_Y$ as weak hypercharge.

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

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

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

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

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

113 The weak force violates parity conservation [19–21], i.e. invariance under parity
114 transformations (mirror reflections). Only left handed fermions participate in the
115 weak interaction. Since there is no other force through which neutrinos interact with
116 other particles, there are no right handed neutrinos in the standard model.

117 2.2 The Higgs Mechanism

118 The Brout-Englert-Higgs mechanism (henceforth just “Higgs mechanism”) is the
119 mechanism through which the fundamental particles of the SM acquire mass [9–11].
120 Experimentally it was known that the weak force had a weak effective strength,
121 which was suggestive of a massive mediating gauge particle. However, directly
122 adding mass to the weak gauge bosons violates the non-Abelian symmetry of the
123 SM. Instead, the gauge bosons can gain mass through the interaction with a scalar
124 Higgs field which results from the spontaneous breakdown of symmetry as discussed
125 in Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as
126 discussed in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the
127 Higgs particle relevant to hadron colliders.

¹²⁸ 2.2.1 Electroweak Symmetry Breaking

¹²⁹ Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It
¹³⁰ is the transition of a physical system from a state of manifest symmetry to a state
¹³¹ of hidden, or *broken*, symmetry. In particular, this applies to physical systems
¹³² where the Lagrangian observes some symmetry, but the lowest energy vacuum states
¹³³ do not exhibit that same symmetry. In other words, the symmetry is broken for
¹³⁴ perturbations around the vacuum state.

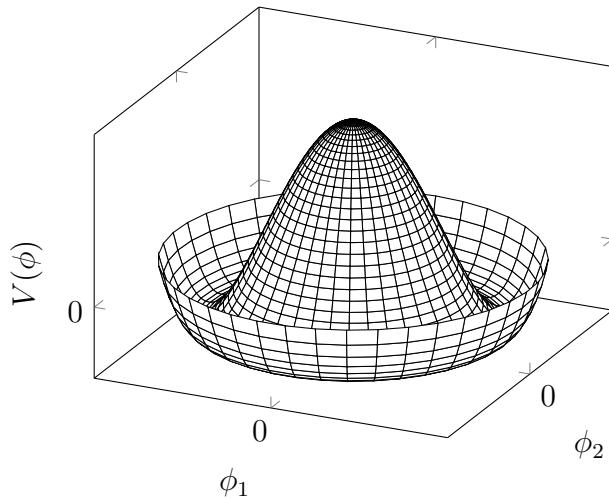


Figure 2.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a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

¹³⁵ Consider gauge fields from the local $SU(2)_L \otimes U(1)_Y$ symmetry group discussed in
¹³⁶ Section 2.1.3 coupled to a complex scalar field $\phi = \phi(x)$. The scalar field ϕ transforms
¹³⁷ as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing
¹³⁸ $\phi^2 \equiv \phi^\dagger \phi$, the Lagrangian is

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

¹³⁹ where the covariant derivative is given by

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

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

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

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

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

¹⁶⁶ in terms of the original fields A_μ^a and B_μ . This gives

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

¹⁶⁷ where θ_W is the weak mixing angle defined by

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

¹⁶⁸ The corresponding masses of the now massive vector bosons can be read off as

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

¹⁶⁹ while the photon remains massless. The Higgs mass is $m_H = v\sqrt{\lambda} = \mu$.

¹⁷⁰ This is the Higgs mechanism. It maintains the renormalisability and unitarity of
¹⁷¹ the SM whilst allowing the weak vector bosons to acquire mass. In summary, an
¹⁷² unphysical complex scalar field ϕ with a nonzero VEV leads to spontaneous symmetry
¹⁷³ breaking. Due to the non-Abelian symmetry breaking, would-be massless Nambu-
¹⁷⁴ Goldstone modes, which arise after expansion around the true vacuum state, are
¹⁷⁵ exactly cancelled out by making a local gauge transformation to the unitary gauge,
¹⁷⁶ and instead are absorbed by the vector bosons, allowing them to acquire mass.

¹⁷⁷ This sector of the SM contains four fundamental parameters that must be determined
¹⁷⁸ from experiment. These can be specified by the Lagrangian parameters g , g' , v
¹⁷⁹ and λ or the physically measurable parameters m_Z , $\sin \theta_W$, m_H and e . In the
¹⁸⁰ local neighbourhood around the true vacuum, the macroscopic symmetry of the
¹⁸¹ system is not realised, and therefore the physical particles do not obey the original
¹⁸² symmetry. However, information about the symmetry is retained through some
¹⁸³ additional constraints on the parameters of the theory. Prior to symmetry breaking,
¹⁸⁴ the potential contained two terms and two constants. After symmetry breaking
¹⁸⁵ there are three terms but still only two constants that relate these terms. This is the
¹⁸⁶ vestige of the original symmetry.

¹⁸⁷ Spontaneous symmetry breaking has modified the original symmetry group of the SM
¹⁸⁸ $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$. Three broken generators from the symmetry
¹⁸⁹ group $SU(2)_L \times U(1)_Y$ have been absorbed into the definition of the physical weak

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

192 2.2.2 Fermionic Yukawa Coupling

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

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

202 Lepton Masses

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

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

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

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

209 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
210 the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore
211 be identified as the electron mass term. An interaction term between the electron
212 field e and the true vacuum Higgs field H is also present. Mass is generated for the
213 other lepton generations in the same way.

²¹⁴ **Quark Masses**

²¹⁵ The down-type quarks acquire their mass analogous to the leptons, with $\psi_L = (u_L, d_L)$
²¹⁶ and $\psi_R = d_R$ for the first quark generation. Mass is generated for the up-type quarks
²¹⁷ using the conjugate field to ϕ which transforms under $SU(2)$ as a doublet with
²¹⁸ $Y = -1/2$. The conjugate field $\tilde{\phi}$ is constructed as

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

²¹⁹ and transforms in the same way as ϕ . This field can be used to write an additional
²²⁰ Yukawa coupling which provides mass for the up-type quarks in a similar way as
²²¹ before.

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

²²² Considering the first generation of up-type quarks with $\psi_L = (u_L, d_L)$ and $\psi_R = u_R$,
²²³ substitution into Eq. (2.24) yields

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

²²⁴ The Yukawa terms mix quarks of different generations of lepton and quark. Physical
²²⁵ particles are detected in their mass eigenstates q , which diagonalise the mass matrix,
²²⁶ but interact via the weak interaction according to their weak eigenstates \tilde{q} , which
²²⁷ are superpositions of the mass eigenstates. This feature of the weak sector leads to
²²⁸ mixing between different generations of quarks and leptons. Quark mixing can be
²²⁹ expressed using the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which specifies the
²³⁰ strength of flavour-changing weak currents. The entries in the matrix are enumerated
²³¹ as

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

²³² where the size of the elements $|V_{pq}|^2$ measures the probability of a transition between
²³³ states p and q .

²³⁴ 2.2.3 Higgs Sector Phenomenology

²³⁵ As previous discussed in this section, the Higgs field plays a key role in the SM,
²³⁶ giving mass to fundamental particles. The strength of the coupling between the
²³⁷ Higgs field and another particle is proportional to that particle's mass. This fact
²³⁸ dictates which production mechanisms and decay modes are dominant at the LHC.
²³⁹ The cross sections for different production mechanisms at a centre of mass energy
²⁴⁰ $\sqrt{s} = 13 \text{ TeV}$ are shown as a function of the Higgs mass m_H in Fig. 2.3. Higgs boson
²⁴¹ production occurs mainly through four modes, shown in Fig. 2.2. The dominant
²⁴² production mode is gluon-gluon fusion ($pp \rightarrow H$), which is predominantly mediated
²⁴³ by a virtual top quark loop. Vector boson fusion ($pp \rightarrow qqH$) is the second most
²⁴⁴ dominant production mechanism, in which a pair of W or Z bosons fuse to produce
²⁴⁵ a Higgs after being radiated by two quarks, which also occur in the final state. Next
²⁴⁶ most common is the associated production of a Higgs boson and a vector boson
²⁴⁷ ($pp \rightarrow VH$), in which a pair of quarks fuse to produce a single W or Z boson which
²⁴⁸ radiates a Higgs. The final of the four leading production modes is top quark fusion,
²⁴⁹ in which two gluons each radiate a quark-antiquark pair, and a quark from each pair
²⁵⁰ fuses to produce a Higgs boson.

²⁵¹ Although gluon-gluon fusion is the dominant production mode, for hadronic decays
²⁵² of the Higgs boson the associated production with a vector boson has the advantage
²⁵³ of leading to a more conspicuous final state due to the likelihood of the vector bosons
²⁵⁴ decaying leptons. Leptons provide a clean signals to detect and trigger on.

²⁵⁵ Since the Higgs boson couples proportional to mass as already mentioned, decays
²⁵⁶ to heavier particles are favoured. The branching ratios of different Higgs boson
²⁵⁷ decay modes are shown as a function of m_H in Fig. 2.4. Approximately 58% of the
²⁵⁸ time the Higgs boson decays to a pair of b -quarks, the dominant decay mode. The
²⁵⁹ next heaviest fermions are the tau lepton and the c -quark, decays to pairs of these
²⁶⁰ particles happen approximately an order of magnitude less often. Decays to pairs
²⁶¹ of vector bosons are via a virtual off shell Higgs boson only. While the $H \rightarrow \gamma\gamma$ and
²⁶² $H \rightarrow ZZ$ branching ratios are small compared with fermionic decay modes (around
²⁶³ 0.2% for $H \rightarrow \gamma\gamma$), these decay channels were instrumental in the initial discovery of
²⁶⁴ the Higgs due to the low level of background processes which mimic the final state.

²⁶⁵ This thesis presents a measurement of the Higgs bosons production rate using events
²⁶⁶ with a Higgs boson produced in association with vector boson and decaying to a pair

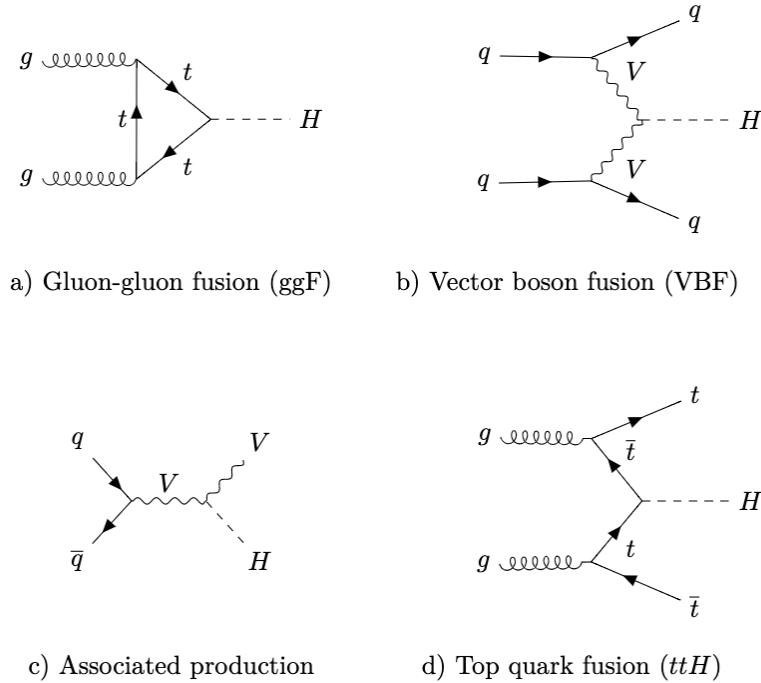


Figure 2.2: Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.

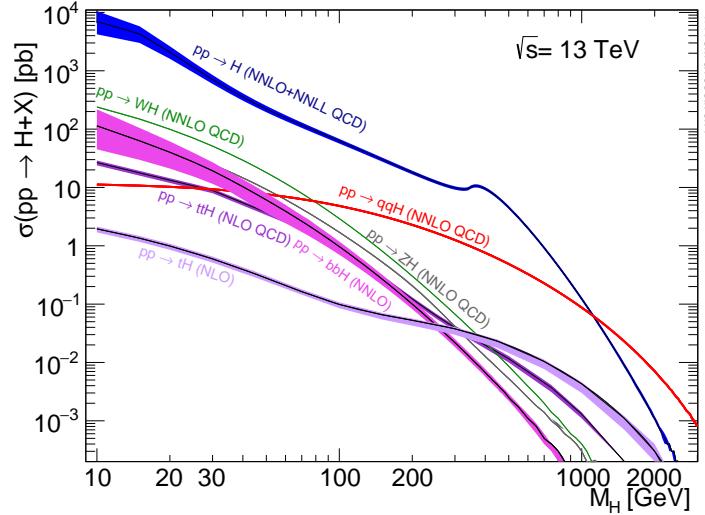


Figure 2.3: Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [24]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fussion, vector boson fusion, associated production with vector bosons, and top quark fusion.

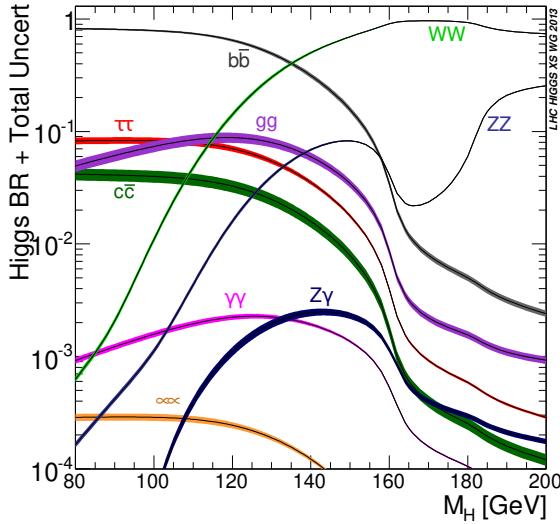


Figure 2.4: Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13 \text{ TeV}$ [24]. Uncertainties are shown in the shaded bands. At $m_H = 125 \text{ GeV}$, the Higgs predominantly decays to a pair of b -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of W bosons.

of b -quarks, i.e. $pp \rightarrow VH(b\bar{b})$. The $H \rightarrow b\bar{b}$ decay mode directly probes the Higgs coupling to the second generation fermions, and more specifically to the bottom quark. This coupling was first observed in 2018 [25, 26]. Ongoing work measuring the coupling strengths, in particular in the high energy regime, is the focus of the analysis presented in this thesis in Chapter 7.

272 Chapter 3

273 The Large Hadron Collider and the
274 ATLAS Detector

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

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

²⁹¹ 3.1 The Large Hadron Collider

²⁹² The LHC is operated in multi-year *runs* during which beams of protons are circulated
²⁹³ and collided. Between runs there are periods of shutdown while the accelerator and
²⁹⁴ detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC
²⁹⁵ collided proton bunches, each containing more than 10^{11} particles, 20 million times
²⁹⁶ per second, providing 7 TeV proton-proton collisions at instantaneous luminosities
²⁹⁷ of up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The centre-of-mass energy was increased to 8 TeV
²⁹⁸ towards the end of Run 1 in 2012. Run 2, which spanned in 2015–2018, further
²⁹⁹ increased the the proton-proton collision energy to 13 TeV. During Run 2 the bunch
³⁰⁰ spacing was reduced, leading to a collisison rate of 40 MHz. Over the course of
³⁰¹ Run 2 a total usable integrated luminosity of 139 fb^{-1} was recorded. 2022 marked the
³⁰² beginning of Run 3 which, with a higher center of mass energy and peak luminosity,
³⁰³ is expected to culminate in the approximate tripling of the dataset size. A summary
³⁰⁴ of key information about each run is listed in Table 3.1.

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

Table 3.1: Overview of the different LHC runs [32,33]. The average number of interactions per bunch-crossing is denoted as $\langle\mu\rangle$ (see Section 3.2.3), and is here averaged over the entire run. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.

³⁰⁵ An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is
³⁰⁶ at the final stage of a chain of accelerators which incrementally step-up the energy
³⁰⁷ of incoming protons. The first accelerator is Linac4, a linear accelerator which
³⁰⁸ accelerates negative hydrogen ions to an energy of 160 MeV. Upon leaving Linac4,
³⁰⁹ the ions are stripped of both electrons and the resulting protons are fed into the
³¹⁰ Proton Synchrotron Booster (PSB), which increases the energy of the protons to
³¹¹ 2 GeV. The protons leaving the PSB are passed to the Proton Synchrotron (PS),
³¹² which increases the energy to 26 GeV, and then from the PS to the Super Proton
³¹³ Synchrotron (SPS) which further increases the energy to 450 GeV. Finally, the proton

314 beams are injected in the LHC where they are accelerated to their final energy of
 315 6.5 TeV (for Run 2).

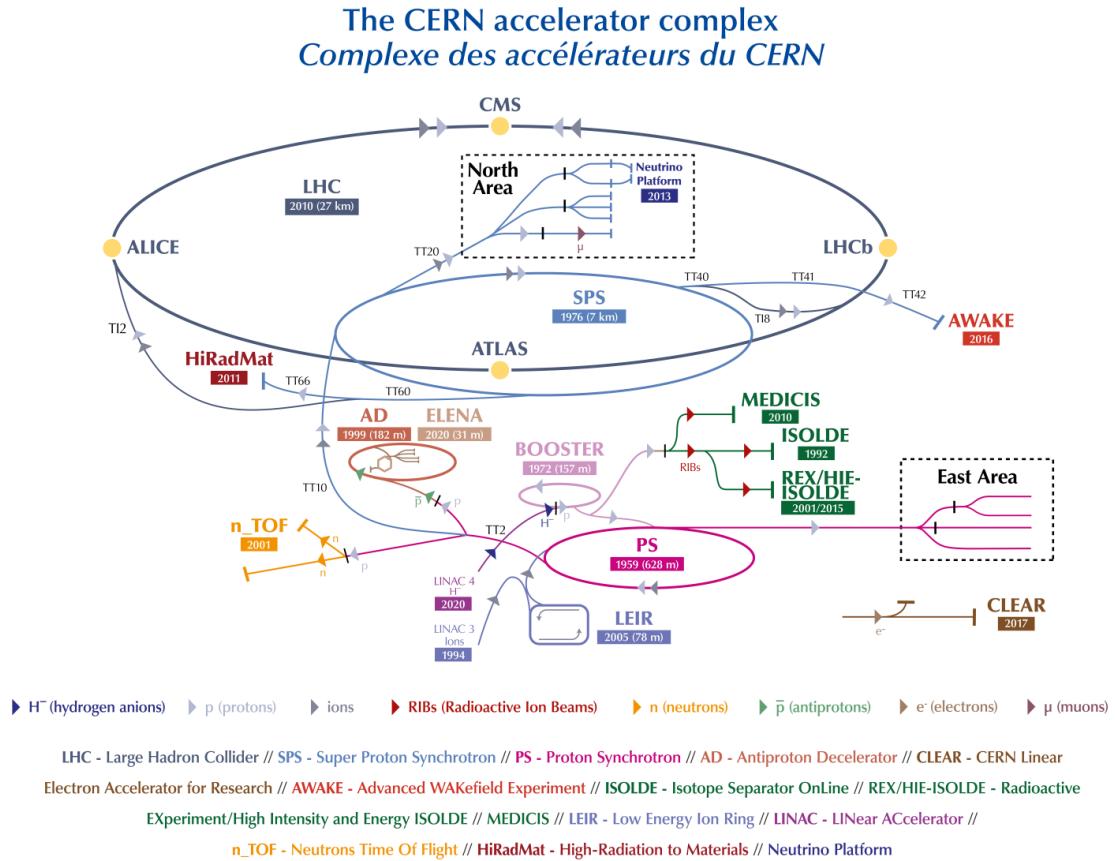


Figure 3.1: An overview of the CERN accelerator complex [34]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

316 3.2 Coordinate System & Collider Definitions

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

321 3.2.1 ATLAS Coordinate System

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

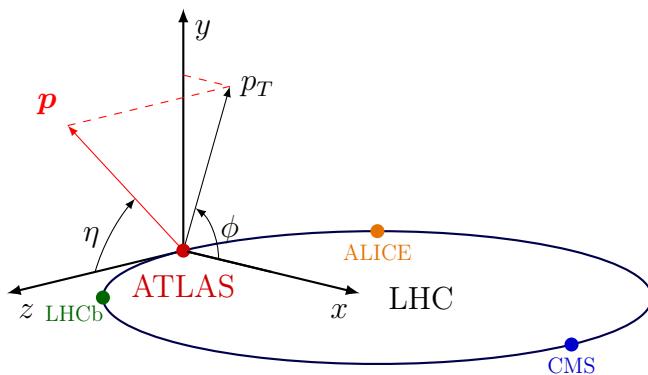


Figure 3.2: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [35].

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

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

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

333 The transverse momentum p_T of an object is the sum in quadrature of the momenta
 334 in the transverse plane

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

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

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

337 3.2.2 Track Parameterisation

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

346 Impact parameter significances are defined as the IP divided by its corresponding
347 uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. When used in flavour tagging
348 (see Chapter 4), track IP significances are lifetime signed according to the track's
349 direction with respect to the jet axis and the primary vertex [37]. The sign IP
350 significances is positive if the track crosses the jet axis in front of the primary vertex
351 and negative if the crossing is behind the primary vertex.

352 3.2.3 Hadron Collider Definitions

353 Cross Section

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

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

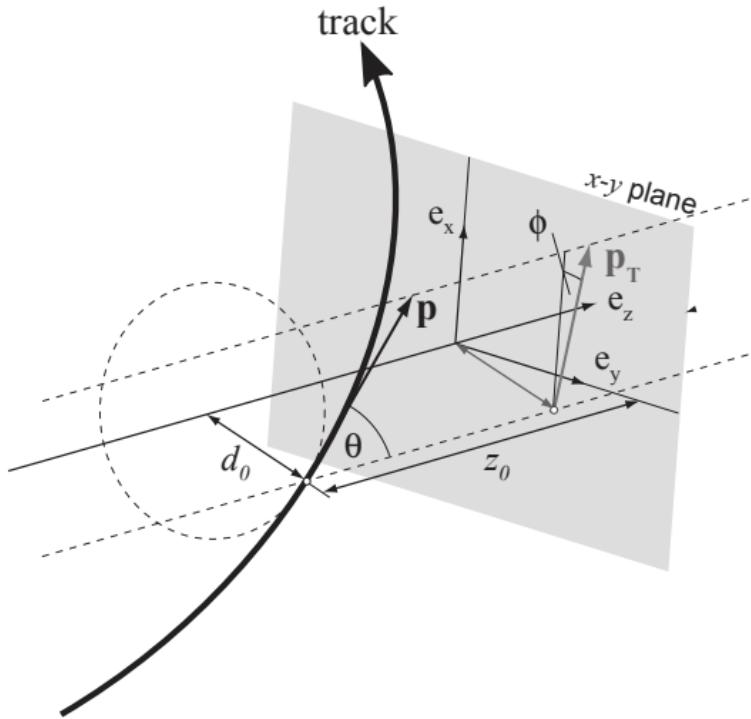


Figure 3.3: The track parameterisation used at the ATLAS detector. Five coordinates ($d_0, z_0, \phi, \theta, q/p$) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum \mathbf{p} and the transverse momentum p_T (defined in Eq. (3.2)). The basis vectors e_x, e_y and e_z are also shown. Reproduced from Ref. [36].

³⁵⁹ section can be factorised as

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

³⁶⁰ The partonic cross section can be calculated at sufficiently high energies such as
³⁶¹ those found at the LHC, while the parton distribution functions (PDFs) have to be
³⁶² extracted from experimental results.

³⁶³ Luminosity

³⁶⁴ The total number of proton-proton collisions N is related to the total pp cross σ
³⁶⁵ section by the integrated luminosity L , as in

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

³⁶⁶ The instantaneous luminosity \mathcal{L} relates the cross section to the number of collisions
³⁶⁷ per unit time. For two colliding bunched proton beams, it is defined as

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

³⁶⁸ where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
³⁶⁹ crossing frequency, and σ_x and σ_y are the rms width of the beam in the horizontal
³⁷⁰ and vertical directions.

³⁷¹ The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In
³⁷² total, 139 fb^{-1} of usable physics data was collected over the three-year run. The
³⁷³ uncertainty on the total integrated luminosity is 1.7% [38].

³⁷⁴ Pile-up

³⁷⁵ At the centre of the ATLAS detector, bunches of more than 10^{11} protons meet at a
³⁷⁶ small crossing angle. Each bunch-crossing is called an *event*. There is generally at
³⁷⁷ most one hard proton-proton scatter per event. Additional interactions are typically
³⁷⁸ relatively soft and are known as *pile-up*. Pile-up from interactions within the same
³⁷⁹ bunch-crossing is known as *in-time* pile-up while residual signatures from previous
³⁸⁰ bunch-crossings is known as *out-of-time* pile-up. The number of pile-up interactions

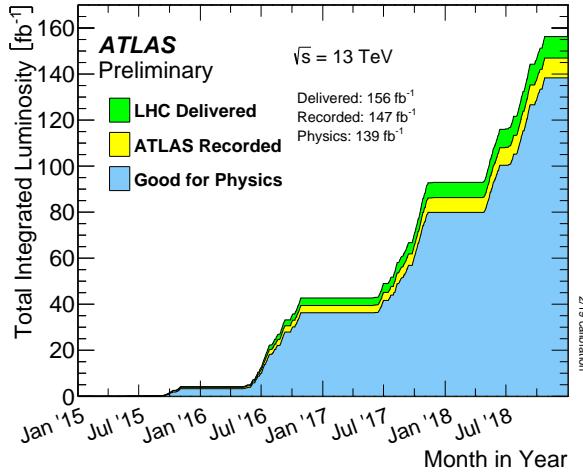


Figure 3.4: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [33]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.

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

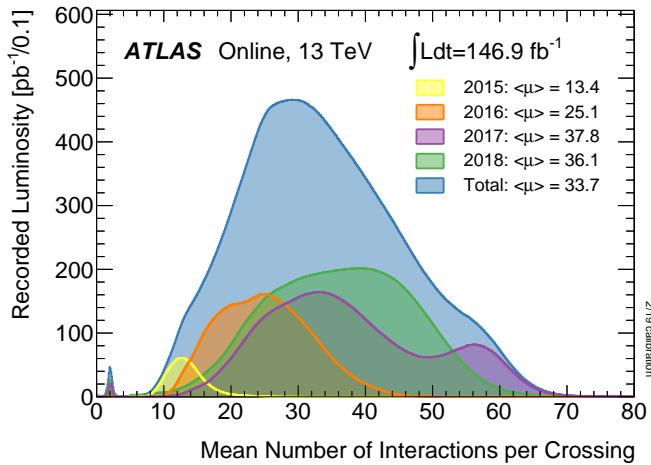


Figure 3.5: Average pile-up profiles measured by ATLAS during Run 2 [33]. Higher levels of pile-up are planned for Run 3.

383 3.3 The ATLAS Detector

384 The ATLAS² detector is made up of several specialised sub-detectors which are
385 arranged concentrically around the nominal interaction point at the centre of the
386 detector. The detector is designed to cover nearly the entire solid angle around the
387 collision point. In this section a condensed overview of each sub-detector is given, in
388 order of increasing radial distance from the point of collision. The inner tracking
389 detector is described in Section 3.3.1, the electromagnetic and hadronic calorimeters
390 in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the trigger is
391 described in Section 3.3.4. More complete information on the detector can be found
392 in Ref. [31], while an overview of physics performance is given in [39].

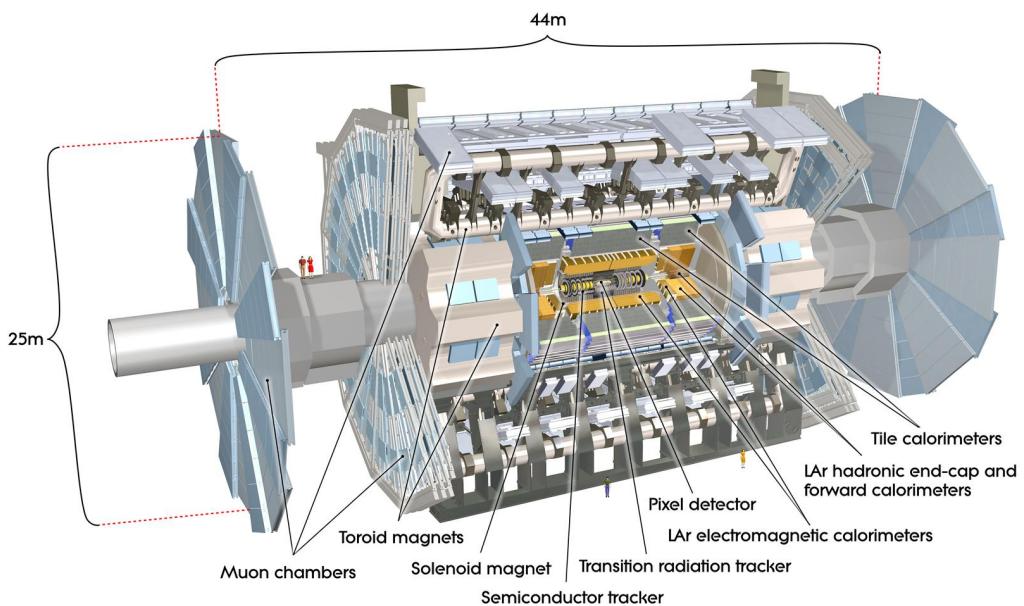


Figure 3.6: A 3D model of the entire ATLAS detector [40]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.

²A Toroidal LHC ApparatuS.

3.3.1 Inner Detector

The inner-detector system (ID) provides high-resolution charged particle trajectory tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field, produced by a superconducting solenoidal magnet, which enables the measurement of particle momentum and charge. After Run 3, the ID will be replaced by the ITk [41, 42].

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

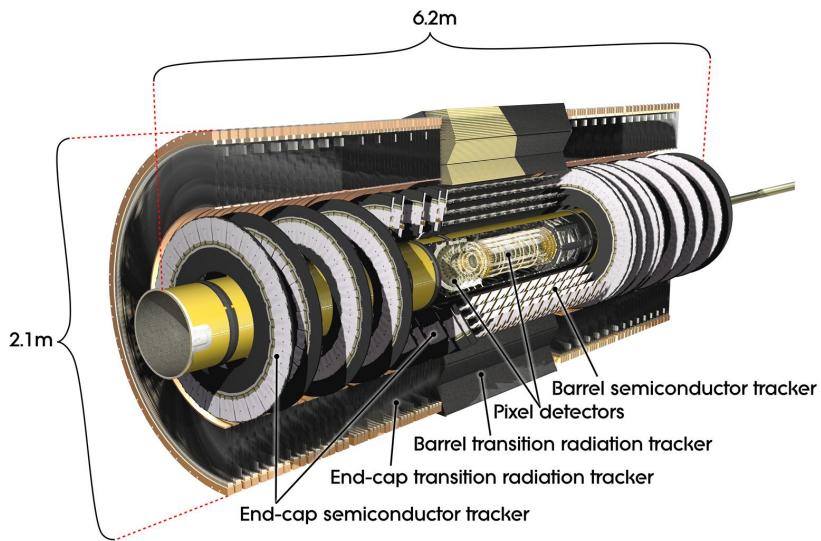


Figure 3.7: A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [43].

The target inverse momentum resolution for the combined ID measurement is parameterised as a function of the track transverse momentum and polar angle [39]. The parameterisation is given by

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

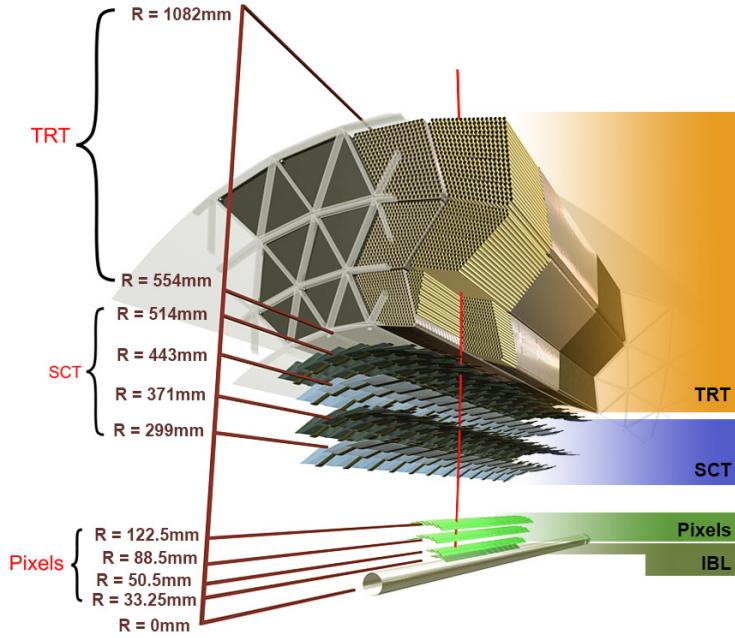


Figure 3.8: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [36].

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

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

413 Pixel Detector

414 The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
 415 from the beamline, and four disks on each side. The innermost barrel layer is
 416 the insertable B-layer (IBL), which was installed before Run 2 [44, 45] and lies
 417 approximately just 33 mm from the beam axis. The second-to-innermost layer is
 418 often referred to as the B-layer. The specification of the pixel detector determines the
 419 impact parameter resolution and the ability to reconstruct primary and secondary

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

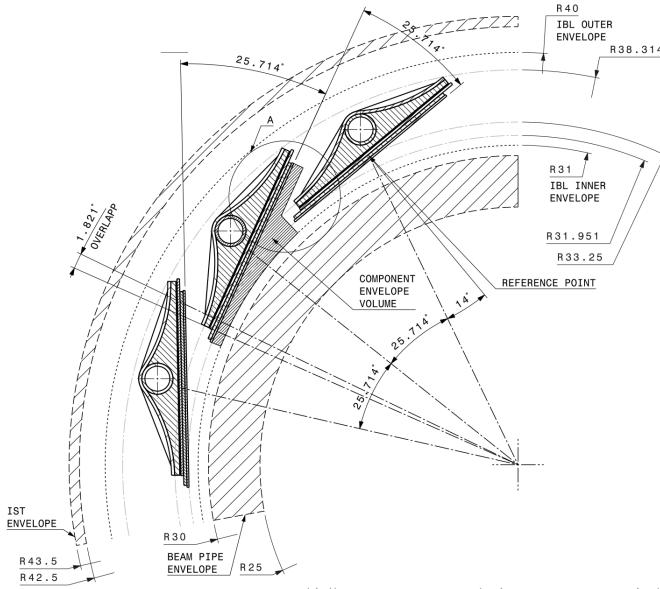


Figure 3.9: A schematic cross-sectional view of the ATLAS IBL [44].

425 Semi-Conductor Tracker (SCT)

426 The SCT is made up of four concentric barrel layers in the central region, and nine
 427 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
 428 with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate
 429 to be measured from a pair of strip measurements. The SCT typically provides four
 430 precision spacepoint measurements (eight strip measurements) per track in the barrel
 431 region. These have intrinsic uncertainties of 17 μm in the transverse direction $R\phi$, and
 432 580 μm in the longitudinal direction z [46]. The measurements provide a contribution
 433 to the measurement of charged particle momentum and impact parameter, along
 434 with vertex position. Charge-particle tracks can be distinguished if separated by
 435 more than $\sim 200 \mu\text{m}$.

436 Transition Radiation Tracker (TRT)

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

450 3.3.2 Calorimeters

451 The calorimeter system measures the energy of incident particles over the range
452 $|\eta| < 4.9$. There are two main sub-systems: the electromagnetic calorimeter (ECal),
453 which focuses on the measurement of electrons and photons, and the hadronic
454 calorimeter (HCal), which measures the energy of hadrons. Upon entering the
455 calorimeter, incident particles will interact with the detector material to produce a
456 shower of secondary particles with reduced energies. The charge deposited in this
457 process is measured to reconstruct the energy of the initial incident particle. The
458 two calorimeter sub-systems must provide strong containment of showering particles
459 to prevent punch-through of EM and non-muon particles to the HCal and muon
460 system respectively.

461 Liquid Argon (LAr) Electromagnetic Calorimeter

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

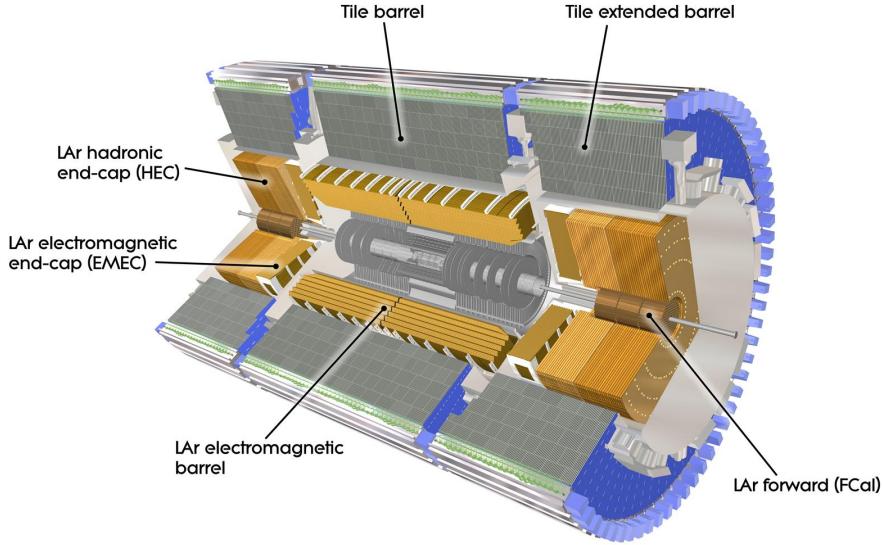


Figure 3.10: The ATLAS calorimeters [47]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic endcaps.

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

470 The energy resolution of the LAr calorimeter is made up of a sampling and a constant
 471 term, which are summed in quadrature to produce the overall energy resolution. The
 472 sampling term contributes approximately $10\%/\sqrt{E}$, while the constant term adds an
 473 additional 0.7%. Photons with moderate transverse energy $E_T \approx 50 \text{ GeV}$ have an
 474 overall energy resolution of at most 1.6% over most of the pseudorapidity range. At
 475 lower $E_T \approx 10 \text{ GeV}$, the resolution is degraded to approximately 5%. The resolution
 476 measurements are obtained from test beam data [39].

477 Hadronic Tile Calorimeter

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

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

484 The jet energy resolution of the HCal is parameterised as a function of the jet
485 transverse energy

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

486 corresponding to a jet energy resolution of 10% at a jet p_T of approximately 100 GeV
487 [48].

488 3.3.3 Muon Spectrometer

489 Due to their higher mass, muons easily pass unimpeded through the ID and calorime-
490 ters and therefore require specialised detectors for their measurement. The Muon
491 Spectrometer (MS) is made up of dedicated tracking and triggering hardware. The
492 precision tracking system uses three layers of monitored drift tubes with a barrel
493 region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$. The inner layers of
494 the end-caps use cathode strip chambers to better cope with the high occupancy
495 in the forward region. Precision tracking resolution is approximately 50 μm . The
496 trigger system is comprised of resistive plate chambers in the barrel region covering
497 $|\eta| < 1.0$ and thin gap chambers in the end-cap regions covering $1 < |\eta| < 2.4$. A set
498 of three superconducting air-core toroidal magnets, each made up of eight coils, is
499 used in each of the barrel and end-caps to deflect the muons as they pass through
500 the MS, allowing their momentum and charge to be measured from the direction
501 and magnitude of curvature. The toroidal magnets generate a field which is largely
502 orthogonal to the muon trajectories which allows for maximum deflection. The
503 transverse momentum resolution has been measured to be approximately 1.7% in the
504 central region for low- p_T muons, increasing to 4% for high- p_T muons in the forward
505 regions [49].

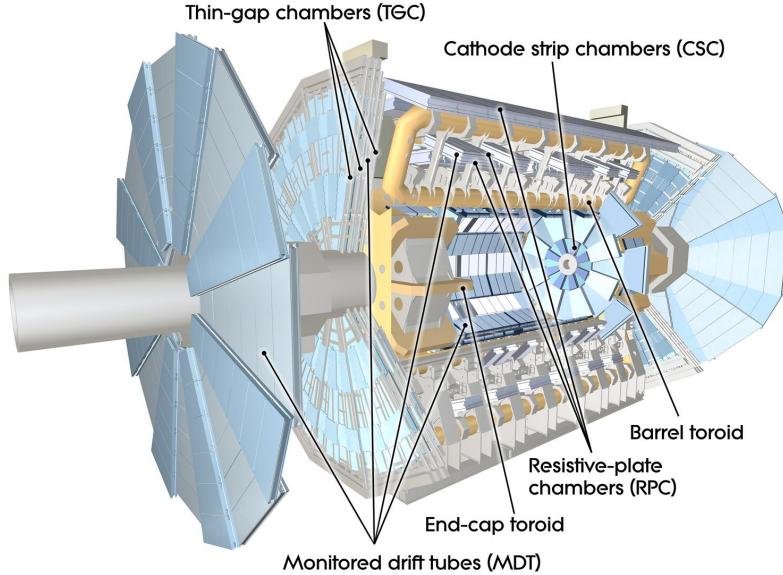


Figure 3.11: The ATLAS muon spectrometer [50].

506 3.3.4 The Trigger

507 The 25 ns bunch spacing used over the course of Run 2 corresponds to a bunch-
 508 crossing or event rate of 40 MHz (see Table 3.1). If the full information for the
 509 detector was written out for each event, this would correspond to the generation
 510 of 60 TB of data each second. This is more than can be feasibly read out from
 511 the hardware, processed and stored, requiring the use of a trigger system which
 512 quickly makes a decision about whether or not an event is potentially interesting
 513 and should be kept for further analysis. The trigger system is comprised of two
 514 levels which search for signs of electrons, muons, taus, photons, and jets, as well as
 515 events with large total or missing transverse energy. The hardware-based Level-1
 516 (L1) trigger uses coarse information from the calorimeters and MS to accept events at
 517 an average rate of 100 kHz approximately 2.5 μ s after the event. After the L1 trigger,
 518 the software-based High Level Trigger (HLT) makes use of 40 000 CPU cores to make
 519 a final selection on surviving events in approximately a few hundred milliseconds.
 520 The final event read-out rate is approximately 1.2 kHz, corresponding to 1.2 GB s^{-1}
 521 of permanent data storage. More information is provided in [51].

522 3.4 Reconstructed Physics Objects

523 Event reconstruction is the process of analysing the output from the detector to
524 determine the type and properties of particles present in an event. The reconstructed
525 event provides information about the underlying physics process that led to these
526 observable final state particles. Events passing the trigger selection (described in
527 Section 3.3.4) undergo offline reconstruction, which makes use of the full information
528 from the detector. Reconstruction and analysis of events relies on the extensive
529 ATLAS software stack, see Ref. [52] for more information.

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

532 3.4.1 Tracks

533 The reconstructed trajectories of charged particles are referred to as *tracks*. Track are
534 reconstructed from the energy depositions (called *hits*) left by the particles as they
535 traverse the the inner detector. Tracks are widely used for a variety of downstream
536 applications, including vertexing and jet tagging, so their accurate reconstruction
537 is a critical task. A comprehensive introduction to ATLAS tracking is available
538 in Ref. [53], while specific optimisations for dense environments are detailed in
539 Refs. [54, 55]. An overview of track reconstruction is given below.

540 Space-point Formation (Clustering)

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

547 Track Finding

548 Space-points are used to build track seeds. These are groups of three hits which
549 are geometrically compatible with being part of a track segment. A combinatorial
550 Kalman filter (KF) is used to build track candidates by extending track seeds. The
551 filter can create multiple track candidates per seed, with bifurcations along the track
552 occurring when more than one compatible space-point exists on a given layer. In
553 this way, the KF creates an excess of *track candidates*, which are only required to
554 satisfy basic quality requirements. Track candidates are allowed to reuse or *share*
555 hits freely (a single hit may be used by multiple track candidates). Typically, the
556 presence of shared hits is a predictor of a bad track due to the high granularity of
557 the ATLAS tracking detectors. At this stage, there can also be a large number of
558 incorrect hits assigned to otherwise good tracks, and additionally large number of
559 *fake* tracks, which are comprised of a majority of wrong hits and do not correspond
560 to the trajectory of any one physical particle (fake tracks are defined as those where
561 the majority of associated hits do not originate from one single truth particle, see
562 Eq. (5.5)). The low quality of tracks at this stage necessitates an ambiguity solving
563 step, in which candidates are cleaned, and the highest quality track are selected.

564 Ambiguity Solving

565 Ambiguity solving was introduced as part of the ATLAS New Tracking effort [53],
566 which was intended to improve track reconstruction performance in dense envi-
567 ronments. In the ambiguity solver, track candidates are processed individually in
568 descending order of a track score. The track score quantifies the likelihood of the
569 track corresponding to the trajectory of a real particle. Scoring uses a number of
570 variables, including the number and positions of hits (preferring hits in more precise
571 regions of the detector), the transverse momentum of the track and the track fit
572 quality. The track fit quality describes the quality of the track as the χ^2 divided
573 by the number degrees of freedom on the track. A preference for high transverse
574 momentum tracks promotes the successful reconstruction of the more physically
575 interesting energetic particles, and suppresses the large number of wrong hits assigned
576 to low momentum tracks. The ambiguity solver also penalises tracks with missing
577 hits on the innermost detector layers.

578 During the processing of a given highest-scoring track candidate, the track is cleaned
579 (whereby problematic hits are removed), and, if the resulting track satisfies the quality
580 selection criteria, a high precision fit of the track parameters using the surviving hits
581 is performed. The high precision fit makes full use of all available information, and
582 uses an updated position and uncertainty estimate for each cluster obtained from
583 a Neural Network (NN) [56]. If the track has reached this stage without rejection
584 by passing various quality regiments, it is re-scored and returned to the list of track
585 candidates. If the same track is then processed again without requiring modification,
586 it is added to the final track collection. Track candidates that fall below a certain
587 quality cut are rejected. This selection does allow for the possibility of a track having
588 small number of shared hits.

list shared
hit cut?

589 Neural Network Cluster Splitting

590 As part of track cleaning, shared hits are classified by a NN to determine if they are
591 compatible with the characteristic features of a merged cluster [54, 56]. A merged
592 cluster is one made up of a combination of energy deposits from more than one
593 particle, which have become merged due to the closeness of the associated particles
594 and the limited resolution of the detector. While in general this event is rare, it
595 is common for clusters to become merged in dense environments, as discussed in
596 Section 4.1. If the cluster is predicted to be merged it is labelled as being freely
597 shareable, or *split*. Hits not compatible with the merged hypothesis can still be
598 shared by a limited number of tracks, but come with a penalty for the track which
599 may hinder its acceptance into the final track collection.

600 3.4.2 Vertices

601 Groups of reconstructed tracks can be examined to determine whether the particles
602 originated from a common spatial point of origin. This occurs when proton-proton
603 collisions take place (primary vertices), when a particle decays or radiates, and also
604 as a result of interaction with the detector material (secondary vertices). Vertex
605 reconstruction is made up of two stages. First, vertex finding takes place, which
606 is the process of grouping tracks into compatible vertices. Second, vertex fitting

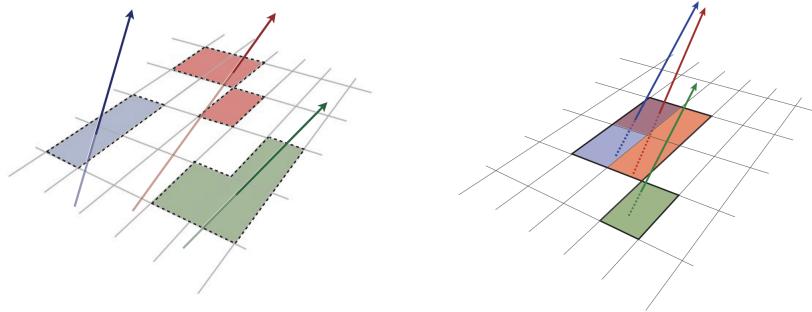


Figure 3.12: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [55].

607 combines information from compatible tracks to reconstruct the physical properties
608 of the vertex, such as mass and position.

609 Primary Vertices

610 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices
611 are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [57].
612 In Run 3, the IVF will be replaced with an adaptive multi-vertex finder (AMVF) [58].
613 The *hard scatter vertex* of an event is chosen as the primary vertex whose associated
614 tracks have the largest sum of transverse momentum squared, $\Sigma(p_T^2)$.

615 Secondary Vertices

616 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient
617 distance from the primary vertex to be resolved from the primary vertex (see
618 Section 4.1.1). Two widely used secondary vertexing tools are used within ATLAS:
619 SV1 and JetFitter [59]. Each attempts to reconstruct secondary vertices inside a jet
620 using the tracks associated to that jet (see Section 3.4.3 for more information about
621 track association). SV1 by design attempts to reconstruct only a single inclusive
622 vertex per jet. This inclusive vertex groups all b -hadron decay products, including
623 tracks from the b -hadron decay itself and tracks from $b \rightarrow c$ decays. The second tool,
624 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary

625 vertices from b -hadron decays are reconstructed separately to tertiary vertices from
626 $b \rightarrow c$ decay chains.

627 3.4.3 Jets

628 Jets are an aggregate reconstructed object corresponding to a collection of collimated
629 stable particles which results from a decay chain of an quark or gluon progenitor. Jets
630 are built by clustering constituent objects (e.g. tracks or calorimeter clusters) using
631 a jet finding algorithm, for example the anti- k_t algorithm [60], which is implemented
632 in FASTJET [61].

633 EMTopo Jets

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

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

637 where E_{cell} is the energy measured in a given calorimeter cell, and $\sigma_{\text{noise, cell}}$ is the
638 expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are
639 formed from a seed cell with a large S_{cell} , and expanded by iteratively adding neigh-
640 bouring cells with a sufficiently large energy significance. Collections of topoclusters
641 are then clustered into a jet using the anti- k_t algorithm with a radius parameter of
642 0.4 (small- R jets) or 1.0 (large- R jets). More information is available in Ref. [62].

643 Particle Flow Jets

644 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [63] using
645 the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate
646 information from both the ID and the calorimeters, improving the energy resolution
647 at high transverse momenta and reducing pile-up contamination. The PFlow jet
648 energy scale is calibrated according to Ref. [64].

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

654 Jet flavour labels are assigned according to the presence of a truth hadron within
 655 $\Delta R(\text{hadron}, \text{jet}) < 0.3$ of the jet axis. If a b -hadron is found the jet is labelled a b -jet.
 656 In the absence of a b -hadron, if a c -hadron is found the jet is called a c -jet. If no b -
 657 or c -hadrons are found, but a τ is found in the jet, it is labelled as a τ -jet, else it is
 658 labelled as a light-jet.

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

660 Large- R Jets

661 Large- R jets have a radius parameter $R = 1.0$ and are built by clustering topological
 662 calorimeter clusters using the anti- k_t algorithm [65]. The large radius parameter
 663 is especially useful for containing the decay products of a boosted Higgs boson, as
 664 discussed in Chapter 7. Due to their large size, large- R jets benefit from a grooming
 665 procedure called trimming which remove soft contaminants inside the jet [66, 67].
 666 Trimming aims to remove jet constituents from pile-up and the underlying event,
 667 which helps to improve the jet mass resolution and its robustness to varying levels
 668 of pile-up. The jet mass is computed using a combination of information from the
 669 calorimeters and ID, and a calibration to data is applied [68].

670 Track-jets

671 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm and
 672 are used in the analysis described in Chapter 7. The radius parameter is allowed
 673 to vary with transverse momentum such that a broader cone (up to $R = 0.4$) is
 674 used for low- p_T track-jets and a narrower cone (down to $R = 0.02$) for high- p_T
 675 track-jets [69, 70]. The narrower cone is better suited to clustering highly collimated
 676 jet constituents at high- p_T . Truth flavour labels for track-jets are derived using the
 677 same $\Delta R(\text{hadron}, \text{jet}) < 0.3$ matching scheme as used for PFlow jets.

678 3.4.4 Leptons

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

682 Electrons

683 Electrons candidates are reconstructed by matching PV-compatible³ inner detector
 684 tracks to topological calorimeter clusters. The track-cluster matching criteria takes
 685 into account the significant energy loss of the electron due to bremsstrahlung. If a
 686 match is found, a refit of the track is performed using the Gaussian Sum Filter (GSF)
 687 [71], which better handles trajectory reconstruction in the presence of bremsstrahlung.
 688 Various identification criteria are then applied to the candidates using a likelihood-
 689 based (LH) method to improve purity. These include requirements on the track
 690 quality and cluster matching, the shape of electromagnetic shower in the ECal,
 691 leakage into the HCal, and the amount of transition radiation detected in the TRT.
 692 Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may
 also be applied. A full description can be obtained from Ref. [72].

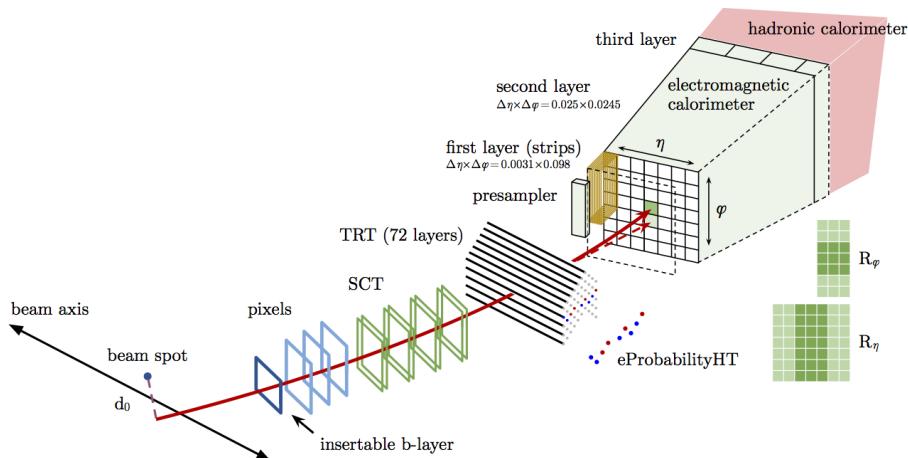


Figure 3.13: A sketch of electron reconstruction using the ATLAS detector [72]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

693

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.

694 Muons

695 Muon reconstruction makes use of the dedicated MS (see Section 3.3.3), the tracks
696 from the ID, and the presence of characteristic signatures in the calorimeters. Muon
697 tracks (i.e. a track reconstructed in the MS) are reconstructed by connecting straight-
698 line track segments, which are identified via a Hough transform, and combined into
699 a approximately parabolic trajectory. Finally, a global χ^2 fit is performed, taking
700 into account possible interactions between the muon and the detector material. A
701 reconstructed muon is called *combined* if it completes successful matching to an
702 ID track. Combined muons undergo a further fit with the combined ID and MS
703 hits, with the energy loss due to the traversal of the calorimeters being taking into
704 account.

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

716 More information on muon reconstruction, identification and isolation can be found
717 in Ref. [73].

718 3.4.5 Missing Transverse Momentum

719 An imbalance in the final state transverse momentum can occur as a result of
720 incomplete measurement of the final state particles. In particular, neutrinos are
721 not measured by the detector and contribute to the missing transverse momentum
722 E_T^{miss} . Incomplete detector acceptance and inaccuracies in the reconstruction of the
723 final state can also contribute to the missing transverse momentum of an event. In
724 order to calculate the missing transverse momentum, the negative vector sum of

- 725 the momentum of all photons, leptons and small- R jets with $p_T > 20\text{ GeV}$ is taken.
726 The momenta of tracks associated to the primary vertex are also taken into account.
727 The magnitude of $\mathbf{E}_T^{\text{miss}}$ is written E_T^{miss} . More information about missing transverse
728 momentum reconstruction is provided in [74].

729 **Chapter 4**

730 **Tracking and b -tagging**

731 Many ATLAS analyses rely on b -tagging, which is the selection of instantiated by
732 b -quarks (b -jets) and rejecting jets created from other quarks (c and light flavours u ,
733 d , s). These b -tagging algorithms (also called *taggers*) work by discriminating for the
734 unique signatures of b -jets, which are outlined in Section 4.1. The various b -tagging
735 algorithms ultimately take as their input information about the reconstructed jet
736 and its associated tracks. Successful b -tagging relies therefore on the efficient and
737 accurate reconstruction of tracks, and especially those tracks corresponding to the
738 products of b -hadron decays.

739 Historically a two tiered approach to b -tagging has been taken, in which so called
740 *low-level* taggers take as inputs information about the jet and associated tracks, and
741 attempt to reconstruct or identify some aspect of a b -jet, such as displaced tracks
742 or secondary vertices. The outputs of several low-level taggers are then fed into a
743 *high-level* tagger, which uses a multivariate approach to discriminate between jet
744 flavours.

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

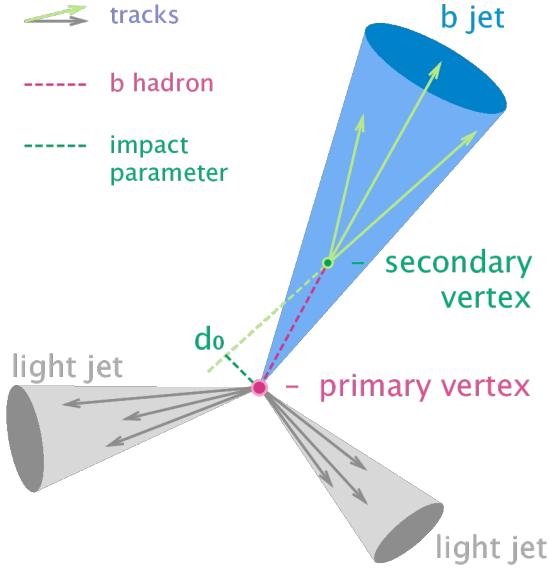


Figure 4.1: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [75]. The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

750 4.1 b -hadron Reconstruction

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

753 4.1.1 Decay Topology

754 b -hadrons are quasi-stable bound states of a bottom quark and one or more lighter
 755 quarks. Collectively, these are the b -mesons (e.g. $B^+ = u\bar{b}$, $B^0 = d\bar{b}$) and baryons
 756 with (e.g. $\Lambda_b^0 = udb$). After a b -quark is produced as the result of some proton-proton
 757 collision, they quickly hadronise. The hadronisation process is hard – around 70-80%
 758 of the b -quark’s momentum is passed to the b -hadron, with the rest being radiated
 759 as prompt hadronisation or fragmentation particles. See Ref. [76] for a more in
 760 depth discussion on hadronisation and the closely related process of fragmentation.
 761 Henceforth the combined hadronisation and fragmentation products will be referred
 762 to collectively as fragmentation.

763 *b*-hadrons are interesting objects of study due to their relatively long proper lifetimes
764 $\tau \approx 1.5$ ps [77]. This lifetime corresponds to a proper decay length $c\tau \approx 450$ μm .
765 Experiment has shown that *b*-hadrons do not couple strongly to light-flavour quarks
766 [78]. The lifetime of *b*-hadrons is therefore approximately determined only by a single
767 CKM matrix element V_{cb} (see Section 2.1.3). In the rest frame of the detector, the
768 typical *b*-hadron travels a distance

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

769 before decaying, where in the high energy limit $\gamma = E_b/m_b$ and $\beta = v/c = 1$. For
770 a 1 TeV *b*-hadron, this gives $d \approx 90$ mm – well beyond the radius of the first pixel
771 layer (the IBL) at a radius of 33 mm (see Fig. 4.2). This significant displacement is
772 characteristic of *b*-jets and makes it possible to reconstruct secondary vertices at the
773 *b*-hadron decay point.

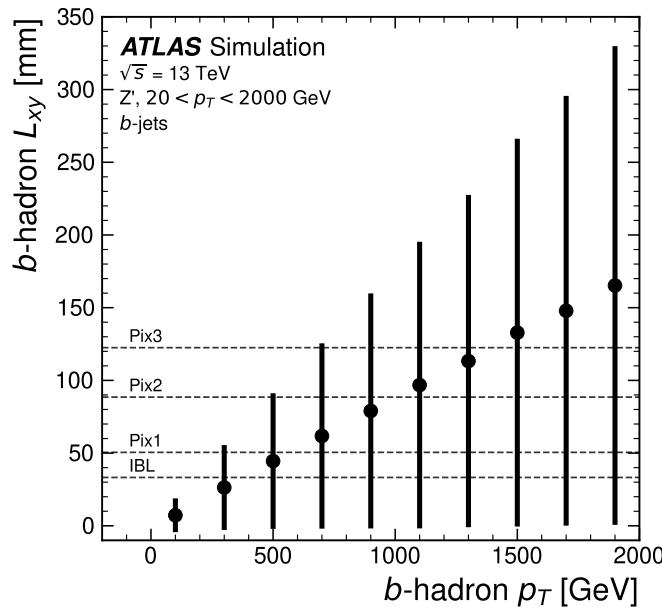


Figure 4.2: The truth *b*-hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed *b*-jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

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

These features stem from the significant displacement of the b -hadron that can occur, in particular for high transverse momentum b -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. 4.1). Additionally, since it is common for the b -hadron to decay to a c -hadron with non-negligible lifetime, tertiary vertices can be found within b -jets resulting from $b \rightarrow c$ decay chains.

These signatures are primarily identified using tracks associated to jets. As such, efficient and accurate track reconstruction is essential for high performance flavour tagging.

4.1.2 Challenges Facing b -hadron Reconstruction

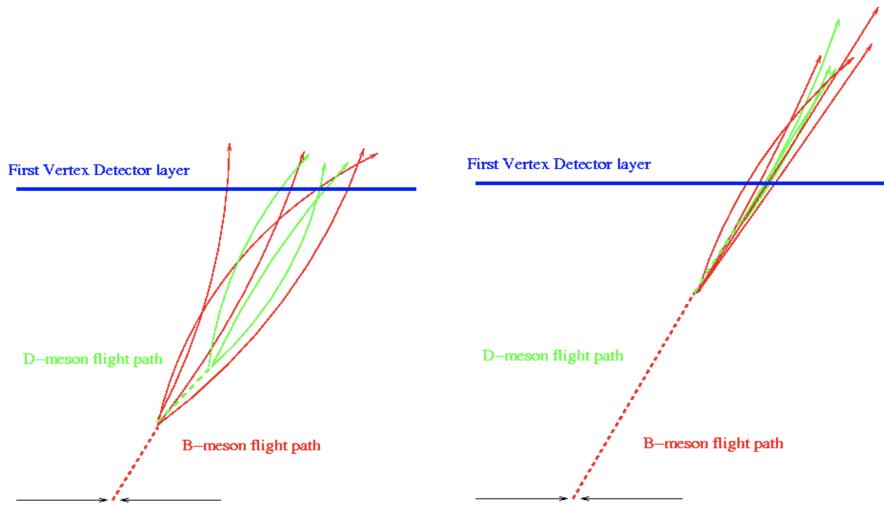


Figure 4.3: At lower p_T (left) the decay length of the b -hadron is reduced, and the resulting decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching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

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

799 At high energies, the increased decay length of b -hadrons (and c -hadrons) means
800 that decay products have less of an opportunity to diverge before reaching the
801 first tracking layers of the detector (shown in Fig. 4.3). If the weak decay takes
802 place close enough to a detector layer, or if the particles are otherwise sufficiently
803 collimated, charge deposits left by nearby particles may not be resolved individually,
804 instead being reconstructed as merged clusters. As discussed in Section 3.4.1, merged
805 clusters are generally rare, and so shared hits generally predict bad tracks and are
806 correspondingly penalised during track reconstruction. However, in the core of high
807 p_T b -jets the density of particles is high enough that the probability of cluster merging
808 increases dramatically. Successful reconstruction of such tracks requires the presence
809 of shared hits, but the presence of these can paradoxically end up impairing the
810 successfully reconstruction of the track. Furthermore, decays may also take place
811 inside the tracking detectors themselves, which at best leads to missing measurements
812 on the most sensitive detector layers, and at worst can lead to wrong inner layer hits
813 being added to displaced tracks, since the reconstruction process penalises tracks
814 without inner layer hits.

can't find
source of
this image,
will have to
reproduce

815 While the multiplicity of the fragmentation products increases with the b -hadron
816 p_T , the multiplicity of the products of the weak decay is unaffected. However, both
817 fragmentation and weak decay products have increased collimation at higher p_T . The
818 combination of effects described above makes reconstructing tracks in the core of
819 high p_T b -jets particularly challenging.

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

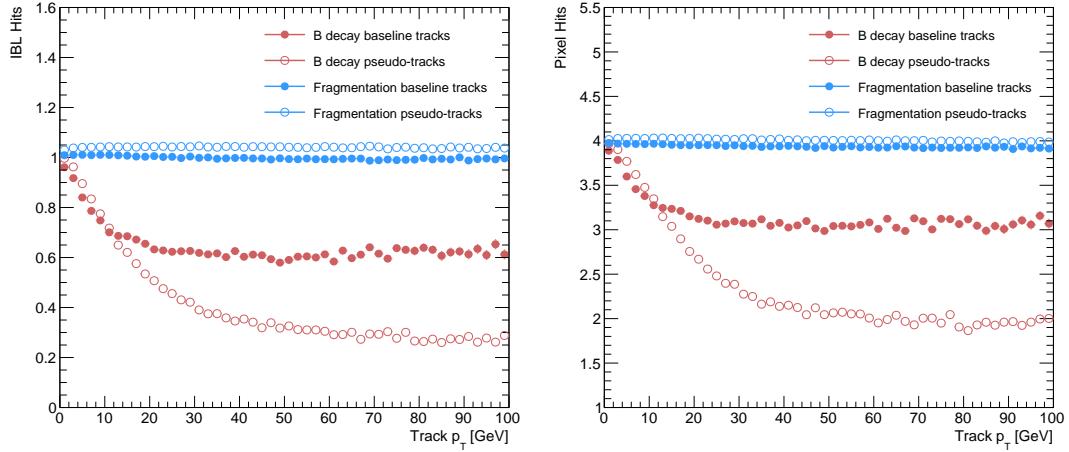


Figure 4.4: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. For each of these, standard tracks and pseudo-tracks are plotted. Hit multiplicities on the pseudo-tracks at high p_T due to the increased flight of the b -hadron. The baseline tracks have more hits than the pseudo-tracks, indicating that they are being incorrectly assigned additional hits.

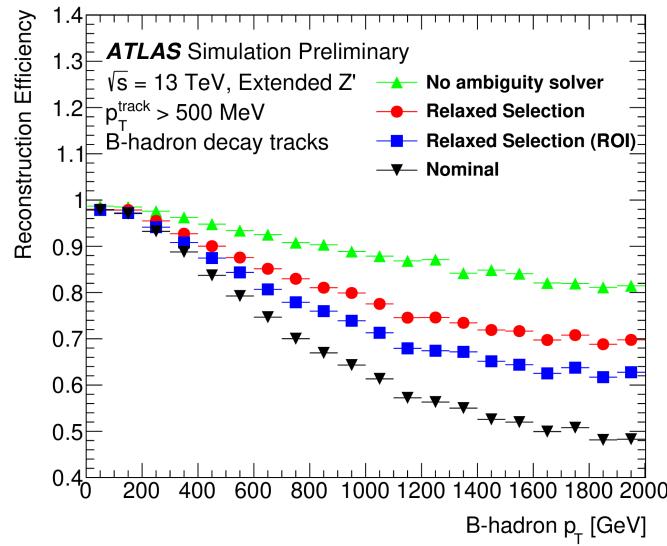


Figure 4.5: b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [79]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T b -hadrons, 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.

825 case of displaced decays. The presence of shared and missing hits reduces a track's
826 score in the ambiguity solver meaning that higher ranking, but potentially worse,
827 track candidates are processed first and take ownership of the hits. This can make
828 it difficult for otherwise reasonable b -hadron decay tracks to meet the ambiguity
829 solver's stringent track quality requirements, leading to their rejection at this stage
830 as shown in Fig. 4.5.

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

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

include plot
from sebs
study show-
ing they are
approx simi-
lar impacts?
or just men-
tion result?

846 4.2 Investigations into High p_T b -hadron Tracking

847 In Section 4.2.1 pseudotracks, a key tool for studying the ideal tracking performance
848 of the ATLAS detector, are used to study the baseline hit-to-track association in
849 the dense cores of high- p_T b -jets. Section 4.2.2 details a study which investigated
850 modifying the global track fitter to improve reconstruction performance in this
851 regime.

852 4.2.1 Pseudo-tracking

853 Pseudo-tracking uses Monte Carlo truth information to group together all the hits left
 854 by each truth particle. Each group of hits which passes basic quality requirements is
 855 directly used in a full resolution track fit. If the track fit is successful, a “pseudo-track”
 856 track is created and stored. If the track fit fails, or the collection of hits does not pass
 857 the basic quality requirements (for example because of a lack of hits) then the particle
 858 is said to be un-reconstructable. In this way, pseudo-tracking performance represents
 859 the ideal reconstruction performance given the ATLAS detector, with perfect hit-
 860 to-track association and and track reconstruction efficiency. The approach was
 861 introduced in [80] as a way to obtain a fast approximation of tracking reconstruction
 862 for simulated data, however the technique has become a useful tool for studying
 863 tracking performance in general [54].

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

880 4.2.2 Global χ^2 Fitter Outlier Removal

882 This section documents ongoing progress into improving hit assignments using the
 883 Global χ^2 Fitter (GX2F) to prevent wrong hits from being assigned to tracks during

discuss
whether to
keep this
section

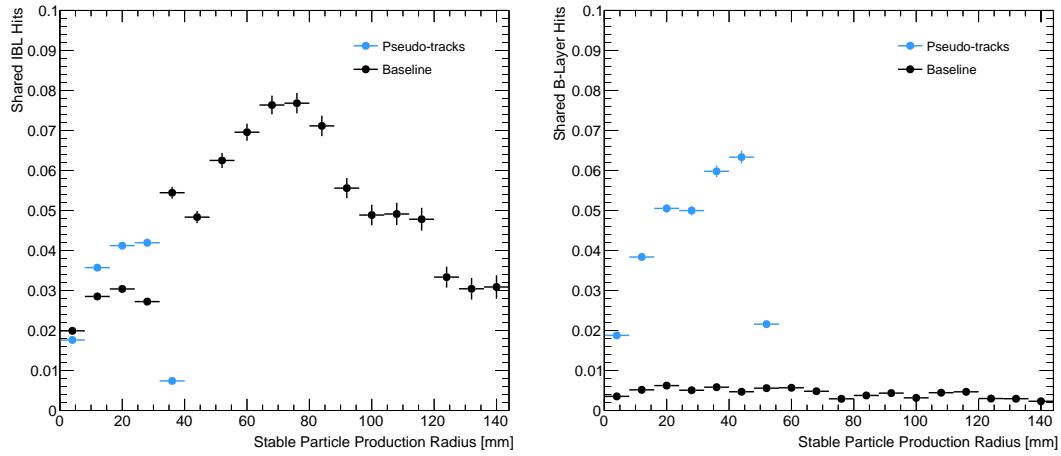


Figure 4.6: The rate of shared hits on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudo-tracks represent ideal performance given the ATLAS detector, and given also the efficiency of the NN cluster splitting algorithms.

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

904 approximately 15% of B hadron decay tracks have a wrong IBL hit (a value which
 905 only increases with the p_T of the B), less than 1% of these tracks have had their IBL
 906 hit labelled and removed as an outlier.

907 Implementation

908 The outlier removal procedure for the pixel detector is described in this section. The
 909 states (also called measurements, or hits) on the track are looped over in order of
 910 increasing radial distance to the beam pipe. For each state, errors $\sigma(m_i)$ on the
 911 measurement of the transverse and longitudinal coordinates are calculated. These
 912 errors are dependent on the sub-detector which recorded the measurement (as some
 913 sub-detectors are more precise than others). Additionally, a residual displacement
 914 $r_i = m_i - x_i$ between the predicted position of the track x_i (inclusive of the current
 915 measurement), and the position of the measurement itself, m_i , is calculated. The
 916 pull p_i on the track state due to the current measurement is calculated according to

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

917 This pull is computed for the transverse and longitudinal coordinates of the mea-
 918 surement, and the maximum of the two is selected and checked to see if it exceeds a
 919 certain threshold. If it does, the hit will be removed, after some additional checks are
 920 made to confirm or deny the presence of the outlier. The threshold is set as a member
 921 variable `m_outlcut`. The results of varying this cut are described in Section 4.2.2.

922 Cut Optimisation

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

931 The results, demonstrating a reduction in wrong hit assignment whilst also improving
 932 slightly the good good hits assigned to tracks, are shown in Fig. 4.7. The improve-
 933 ments are also observed when looking inclusively in all tracks, which removes the
 need for a specific b -jet ROI. An improvement, though modest, of all track parameter

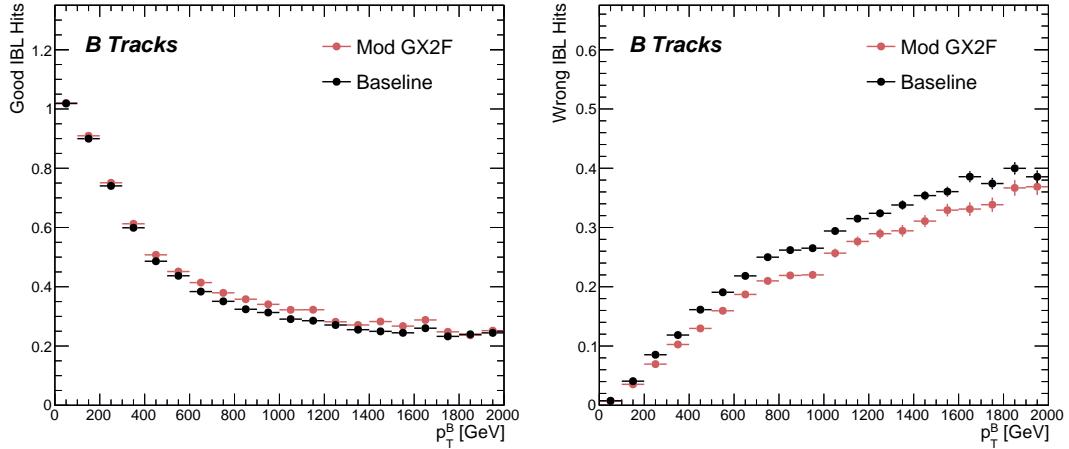


Figure 4.7: Profiles, as a function of parent b -hadron p_T , of good (left) and wrong (right) hit assignment rates on the IBL for tracks using baseline tracking (black), the modified version of the outlier removal procedure (red).

934 resolutions and pulls is observed. Some results are shown in Figs. 4.7 and 4.8, whilst
 935 the remainder of the plots can be found in the talks linked on the task’s Jira page.
 936 The results demonstrate an improvement in hit assignment, unchanged reconstruction
 937 efficiency, and modest improvement in track parameter resolutions and pulls. In
 938 addition, the truth match probability of track is unchanged, suggesting that there is
 939 no increase in fake track rates. The changes are expected to have a negligible impact
 940 on computational resources.

942 4.3 Conclusion

943 In this section, the difficulties facing efficient and accurate track reconstruction,
 944 and hence performant b -tagging, have been outlined. The ambiguity solver, which
 945 attempts to clean or reject tracks which have excessive number shared hits, is shown
 946 to be overly aggressive in the removal of b -hadron decay product track candidates.
 947 The ambiguity solving process relies on many pre-defined cuts which have not been

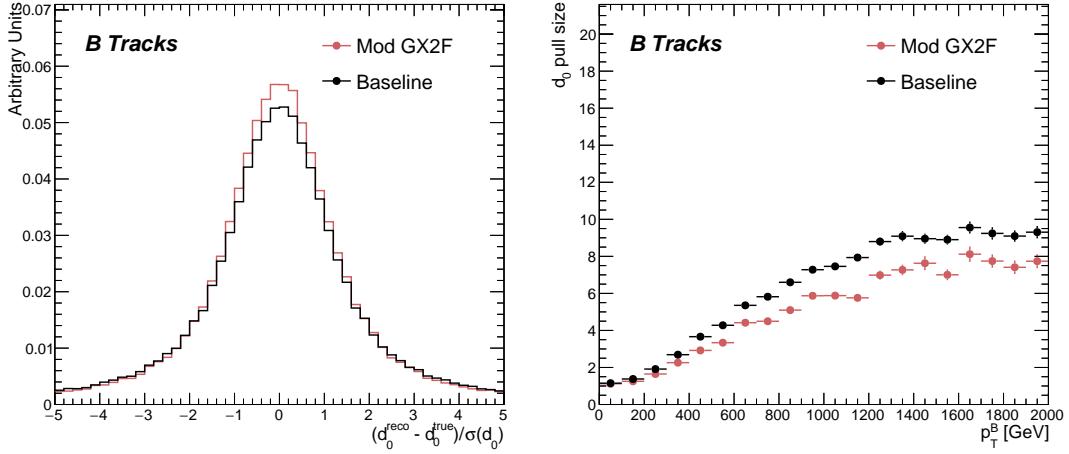


Figure 4.8: (left) B hadron decay track d_0 pulls for baseline and modified GX2F tracks. (right) The magnitude of the decay track d_0 pull as a function of B hadron transverse momentum.

948 optimised for high transverse momentum b -hadron track reconstruction. These
 949 conclusions have motivated further ongoing studies into the improvement of the
 950 track reconstruction in dense environments and the high- p_T regime, such as those in
 951 Ref. [79].

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

961 **Chapter 5**

962 **Track Classification MVA**

963 The chapter details work on implementing a multivariate algorithm (MVA) to predict
964 the truth origin of reconstructed tracks. An introduction to formalisms of machine
965 learning is given in Section 5.1. In Section 5.2, the truth origin label is defined,
966 and in Section 5.3 these labels are used to train a machine learning model that can
967 effectively discriminate between good and fake tracks. Several studies motivated this
968 work by demonstrating that at high p_T , b -tagging performance was degraded by the
969 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
970 could be trained to detect fake tracks, these could be removed before their input to
971 the b -tagging algorithms with the aim of improving performance.

972 **5.1 Machine Learning Background**

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

982 programmed rules can fail to fully capture the relationships between inputs and
 983 outputs.

984 In the simplest case, a set of m labelled training examples $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$
 985 is collected. Each element (x_i, y_i) consists of a input vector $x_i \in \mathbb{R}^{\text{input}}$, and the
 986 corresponding label y_i . In classification problems, these labels are integer *class*
 987 *labels* $y_i \in \{0, \dots, N - 1\}$, where N is the number of classes, which specify which
 988 of a pre-determined set of categorical classes the training example belongs to. The
 989 rest of the discussion in this chapter is limited to binary classification problems
 990 ($N = 2$). The two classes are often referred to as signal ($y_i = 1$) and background
 991 ($y_i = 0$), which need to be separated. Collecting sufficient and suitable data is one
 992 of the primary challenges of machine learning, as such data is not always readily
 993 available. Fortunately, sophisticated tools to simulate particle collisions have already
 994 been developed by the scientific community [81, 82]. These tools play a key role in
 995 generating a suitablly large amount of labelled data which is used to train algorithms.
 996 More detail on the input datasets is given in Section 5.3.1.

997 After obtaining suitable training data, the next step is to define a model. Given an
 998 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
 999 parameterised functional mapping from input space to output space. Given an input
 1000 example x_i and a set of parameters θ , the model outputs a prediction $\hat{y}_i \in (0, 1)$ for
 1001 the true label y_i , as in

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

1002 The output \hat{y}_i is in the interval $(0, 1)$ so as to be interpreted as the probability
 1003 that the input example x_i belongs to the signal class. The parameters θ of the
 1004 model are randomly initialised, however the model is designed to be expressive
 1005 enough to correctly map the inputs x_i to the outputs y_i given the correct choice of
 1006 parameters. The model is then trained, which amounts to showing the model a series
 1007 of labelled training examples and modifying the parameters of the model based on
 1008 its performance.

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 [83] and are easy to train via backpropagation [84].

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 [85, 86]. An artificial 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.

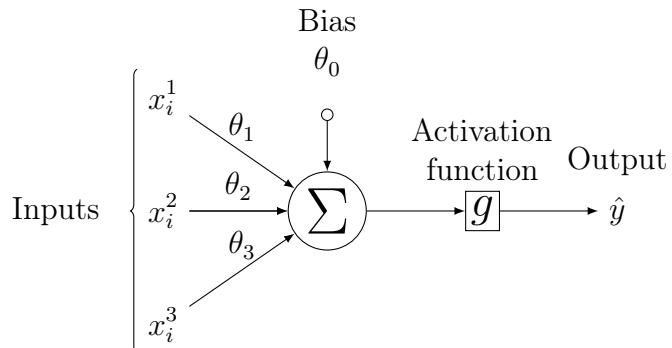


Figure 5.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result is passed to an activation function. Each neuron can be thought of as a logistic regression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

The dot product is fed into an activation function g . 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 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 is used to compute the weight updates during training. This is also why input data is normalised to have zero mean and unity variance [87].

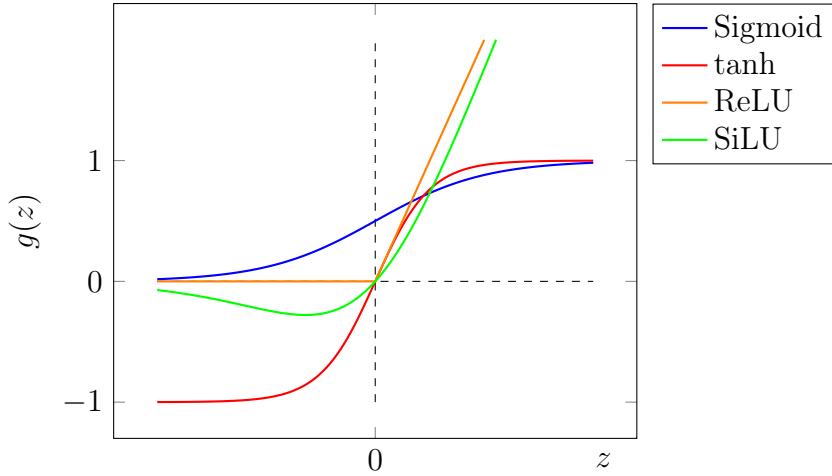


Figure 5.2: Several common choices for the activation function g of an artificial neuron.

1027 Networks

1028 Several neurons are linked together in layers to form a neural network. The inputs
 1029 are propagated layer-by-layer through the network until reaching the final output
 1030 layer. The number of layers and neurons per layer are important hyperparameters
 1031 (those parameters which are not optimised as part of the training process) which
 1032 influence the performance of the model. In the case of binary classification, the final
 1033 output layer consists of a single neuron with a sigmoid activation

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

1034 which is bounded between zero and one allowing the final output to be interpreted
 1035 as the probability that the input sample belongs to the signal class. NNs have the
 1036 crucial property of being differentiable functions, which facilitates training process
 1037 described in the next section.

1038 5.1.2 Training with Gradient Descent

1039 A training algorithm is used to optimise the weights of a NN after exposure to the
 1040 training data. The training algorithm works minimising a loss function L , which
 1041 quantifies the error in the model's predictions for a given input. NNs are commonly
 1042 trained using backpropagation in combination with a variant of stochastic gradient

1043 descent to iteratively update the model parameters. In binary classification problems,
1044 the binary cross entropy given in Eq. (5.3) loss if often used.

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

1045 Since the model f is differentiable, the error for each parameter θ_i can be computed
1046 by taking partial derivative of L with respect to the parameter. Updated parameters
1047 θ'_i are calculated by updating the original parameter in the direction which reduces
1048 the loss.

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

1049 The hyperparameter α is known as the *learning rate* and dictates the size of the
1050 step taken in the direction of the slope. The errors for each parameter are efficiently
1051 calculated using the backpropagation algorithm [84]. The process of updating weights
1052 is repeated until the weights converge the network is trained. In practice, small
1053 batches of the input data are shown to the network at a time. For each batch the
1054 average loss is calculated and the network's weights are updated. There are many
1055 extensions and variations of the gradient descent algorithm. This work uses the Adam
1056 optimiser which adds momentum to the weight updates (dampening oscillations)
1057 and an adaptive per-parameter learning rate [88].

1058 5.2 Track Truth Origin Labelling

1059 Crucial to supervised learning techniques are the ground truth class labels which
1060 the machine learning model is trained to predict. A set of track truth labels which
1061 a high degree of granularity have been implemented in the ATLAS software stack,
1062 and are listed in Table 5.1. The labelling scheme has designed to be useful beyond
1063 the classification of good and fake tracks. The origins are determined by analysing
1064 the detailed simulated truth record for the truth particle associated with each track.
1065 Tracks are associated with truth particles by selecting the truth particle with the
1066 highest *truth-matching probability* (TMP), defined in Eq. (5.5). This is a weighted
1067 sum of the number of hits on a track which are from the same truth particle, versus
1068 the total number of hits on the track. The weights are subdetector-dependent are
1069 designed to account for the varying number of layers (and therefore expected number

1070 of hits) in each of the subdetectors.

$$\text{TMP} = \frac{10N_{\text{Pix}}^{\text{good}} + 5N_{\text{SCT}}^{\text{good}} + N_{\text{TRT}}^{\text{good}}}{10N_{\text{Pix}}^{\text{all}} + 5N_{\text{SCT}}^{\text{all}} + N_{\text{TRT}}^{\text{all}}} \quad (5.5)$$

1071 For the fake track classification tool, the origins in Table 5.1 are used to construct a
1072 binary label by labelling all fake tracks background, and all other tracks as signal.
1073 The fake track classifier is then trained to distinguish between these two categories
1074 of tracks. Fake tracks are also defined using the TMP, with a $\text{TMP} < 0.5$ giving a
1075 track the label of fake. Fake tracks are made up of combinatorial fakes, which are
1076 tracks which do not correspond to the trajectory of any truth particle, and poorly
1077 reconstructed tracks, which may somewhat resemble the trajectory of a truth particle
1078 but may be off due to the presence of some wrong hits on the track.

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

Table 5.1: Truth origins which are used to categorise the physics process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [55]. A truth-matching probability of less than 0.5 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions, K_S^0 and Λ^0 decays, and hadronic interactions.

1079 5.3 Fake Track Identification Tool

1080 The rate of fake tracks increases at high transverse momentum as shown in Fig. 5.3 due
1081 to the difficulties in track reconstruction outlined in Section 4.1.2. The performance
1082 of b -tagging algorithms is reduced as a direct result of the presence of these fake
1083 tracks as shown for SV1 in Fig. 5.4, where the light-efficiency decreases by up to 35%
1084 at a b -efficiency of 35%.

To identify and remove fake tracks, a NN classification tool was trained with good tracks as the signal class and fake tracks as the background class. Due to the imbalance between the two classes (with fake tracks being relatively uncommon), a weight was added to the loss function for the background class to account for this. The NN was made up of two hidden layers with 100 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 has been carried out to ensure good performance with a relatively small number of learnable parameters. The model was trained using 40 million tracks with a further 1 million tracks each used for validation and testing.

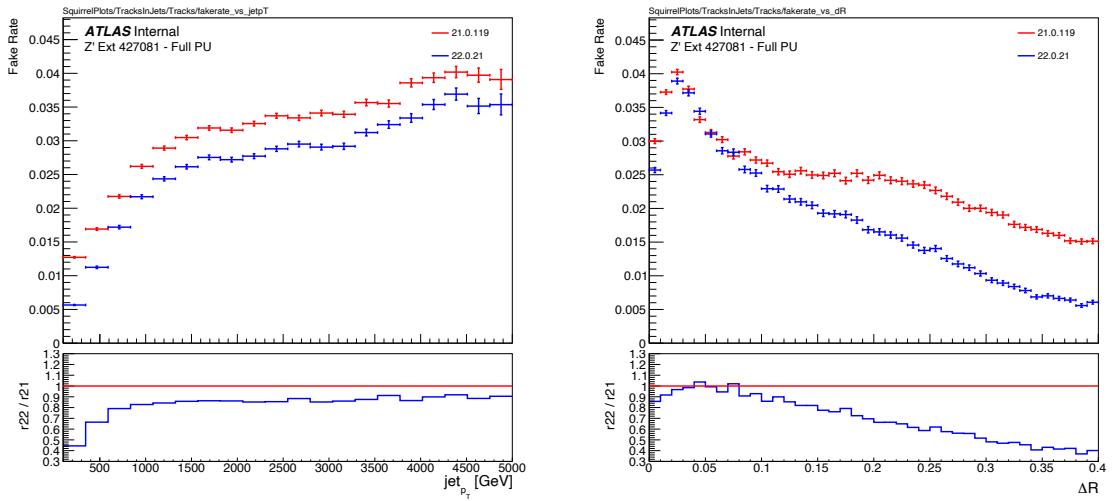


Figure 5.3: Rate of fake tracks as a function of jet transverse momentum (left) and ΔR (track, jet) (right). The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases, and the number of tracks in the jet increases (not shown).

Inputs to the model are described in Section 5.3.2, while fake track removal performance is given in Section 5.3.3.

5.3.1 Datasets

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

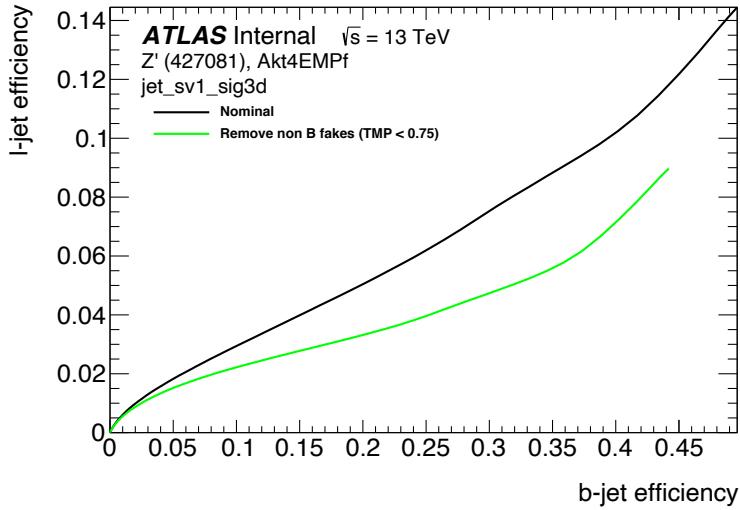


Figure 5.4: The light-jet efficiency of the low level tagger SV1 as a function of b -jet efficiency for the nominal tracking setup (black) and for the case where fake tracks which are not from the decay of a b -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to algorithm performance.

- 1101 up to 5 TeV and decays democratically to an equal numbers of b -, c - and light- jets.
 1102 The generation of the simulated event samples includes the effect of multiple pp
 1103 interactions per bunch crossing with an average pileup of $\langle \mu \rangle = 40$, which includes
 1104 the effect on the detector response due to interactions from bunch crossings before
 1105 or after the one containing the hard interaction.
- 1106 The $t\bar{t}$ events are generated using the POWHEGBOX [89–92] v2 generator at next-
 1107 to-leading order with the NNPDF3.0NLO [93] set of parton distribution functions
 1108 (PDFs). The h_{damp} parameter¹ is set to 1.5 times the mass of the top-quark (m_{top}) [94],
 1109 with $m_{\text{top}} = 172.5$ GeV. The events are interfaced to PYTHIA 8.230 [95] to model the
 1110 parton shower, hadronisation, and underlying event, with parameters set according
 1111 to the A14 tune [96] and using the NNPDF2.3LO set of PDFs [97]. Z' events are
 1112 generated with PYTHIA 8.2.12 with the same tune and PDF set. The decays of b -
 1113 and c -hadrons are performed by EVTGEN v1.6.0 [98]. Particles are passed through
 1114 the ATLAS detector simulation [99] based on GEANT4 [100].

¹The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high- p_T radiation against which the $t\bar{t}$ system recoils.

5.3.2 Model Inputs

The fake track MVA is given two jet variables and 20 tracking related variables for each track fed into the network. The jet transverse momentum and signed pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in Table 5.2. The track parameters and hit pattern are key indicators of whether or not a track is fake. The FracRank variable is the ordered index of the track divided by the total number of tracks in the event. The ambiguity solver processes track candidates iteratively (see Section 3.4.1), and the order in which tracks are accepted is preserved. Tracks which do not require cleaning (i.e. the removal of suspect shared hits) are likely to be accepted earlier on. Hence the FracRank variable gives an indication of how easy it was for the track to be reconstructed.

Track selection follows the loose selection described in Ref. [101] and outlined in Table 5.3, which was found to improve the flavour tagging performance compared to previous tighter selections, whilst ensuring good resolution of tracks and a low fake rate [55]. Inputs are scaled to have a central value of zero and a variance of unity before training and evaluation.

5.3.3 Results

In order to evaluate the of the fake track classification tool, a orthogonal test sample of 1 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 good track (i.e. not fake). Fig. 5.5 shows the performance of the fake track classification MVA. The signal and background classes are well separated in the output of the tool. Also shown is a receiver operating characteristic (ROC) curve, which plots the rate of true positives against the rate of false positives over a scan of cut points on the NN output ranging from zero to one. The area under the curve (AUC) gives a summary of the aggregate classification power of the model. The fake track classification tool achieves an AUC of 0.935 for all tracks, which is indicative of a well-performing model. Considering only tracks from b -hadron decays, this value drops slightly to 0.928.

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

Table 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 [55], while split hits are hits used on multiple tracks which have been identified as merged.

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

Table 5.3: Quality selections applied to tracks, where d_0 is the transverse IP of the track, z_0 is the longitudinal IP with respect to the PV and θ is the track polar angle. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55]. Shared hits on pixel layers are given a weight of 1, while shared hits in the SCT are given a weight of 0.5. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

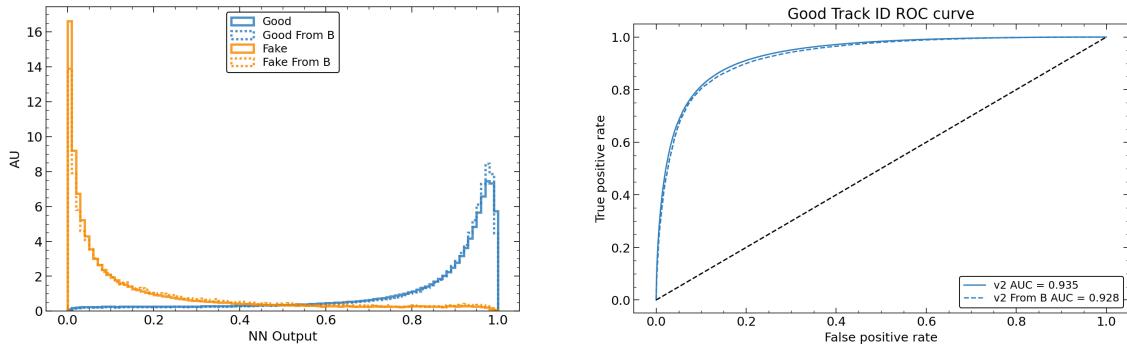


Figure 5.5: (left) Normalised histogram of the model output separated for good and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line).

1144 Good and fake track efficiencies at two different NN output cut points are shown
1145 in Table 5.4. The results demonstrate that the tool is effective in retaining 98.8%
1146 of good tracks, while correctly identifying (and therefore enabling the removal of)
1147 45.6% of fake tracks. Table 5.4 also shows that a significant amount of tracks which
1148 are labelled as both fake and from the decay of a b -hadron are also removed.

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

Table 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. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.

1149 After initial tests and investigation, it was found that fake tracks which were the
1150 result of b -hadron decays actually aided b -tagging performance. The application of a
1151 single tool which removed all fake tracks was therefore not optimal. A second tool
1152 was trained in the same manner of the first, this one was designed to distinguish
1153 between those tracks which were from the decay of a b -hadron and those which were
1154 not. A 2-dimensional cut was then used to only reject those tracks that had a high
1155 probability of being fake, and also a low probability of being a b -hadron decay track.

1156 The light-jet efficiency of SV1 is successfully reduced when using the combined tools
1157 to remove fake tracks that are not from a b -hadron decay, as shown in Fig. 5.6. At
1158 a b -efficiency of 70%, the light-jet mistag rate for jets with $250 < p_T < 400$ GeV is
1159 reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For jets
1160 in $400 < p_T < 1000$ GeV the mistage rate drops from 0.1 to 0.08 for a similar relative
1161 improvement of 20%. The performance of the track classification tools is also shown
1162 for JetFitter in Fig. 5.7. The results are generally consistent as for those from SV1,
1163 with up to a 20% improvement in the light-jet rejection for a given b -efficiency WP
1164 in the two jet p_T bins.

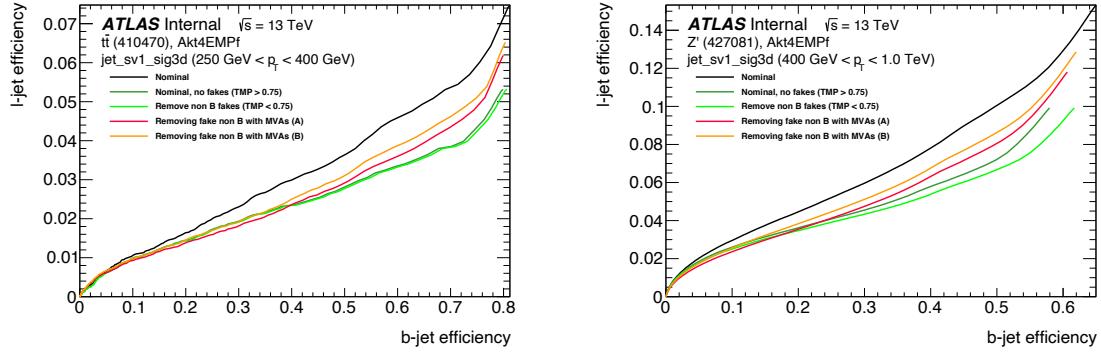


Figure 5.6: The effect of applying the fake track identification algorithm alongside the b -hadron decay track identification on the jet tagging performance of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) is compared to two working points, labelled “A” (red) and “B” (orange) of fake track removal, 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.

5.4 Conclusion

Fake tracks, which are prevalent in the core of high p_T jets, are shown to 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 inputted to downstream b -tagging algorithms. It was found that, since many b -hadron decay tracks are poorly reconstructed and thus marked as fake, it was 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.

While removing tracks prior to their input to the low level tagging algorithms is shown here 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, potentially still making use of some of the information. This is not straightforward with manually optimised taggers such as SV1 and JetFitter, but is possible with approaches such as those described in Chapter 6.

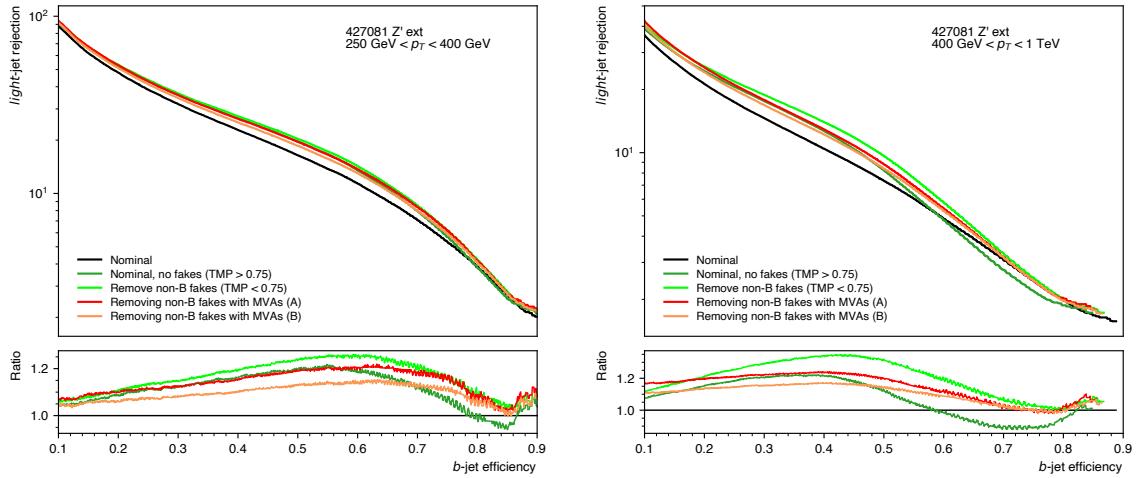


Figure 5.7: The effect of applying the fake track identification algorithm alongside the b -hadron decay track identification on the jet tagging performance of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) is compared to two working points of fake track removal (orange and red). Removal of fake tracks based on truth information is shown by the green curves.

1182 Tools which identify the origin of a given track have other potential uses. One
1183 application is to isolate a relatively pure sample of fake tracks which can be used
1184 to estimate the fake track rate in data, which would be useful for producing the
1185 recommendations for tracking systematic uncertainties. Another application would
1186 be to use the b -hadron track identification tool to improve the track-to-jet association.
1187 Both applications are currently under investigation.

1188 The approach here works on a track-by-track basis, but a more sophisticated approach
1189 would consider the correlations between the tracks inside a jet, as shown in Chapter 6.

1190 Also left for future work is to simultaneously train a single tool which discriminates
1191 between all the truth origins listed in Table 5.1. Such a tool would be useful as a
1192 general purpose multiclass classifier.

₁₁₉₃ **Chapter 6**

₁₁₉₄ **Graph Neural Network Flavour
Tagger**

₁₁₉₆ Some of the work in this chapter has previously been published in Ref. [102]. The
₁₁₉₇ author of this thesis was on the editorial team. Figures, tables and text from the
₁₁₉₈ published note are reproduced here.

₁₁₉₉ As discussed in Chapter 4, flavour tagging is the identification of jets originating from
₁₂₀₀ b - and c -quarks. Flavour tagging is a critical component of the physics programme of
₁₂₀₁ the ATLAS experiment. It is crucial importance for the study of the Standard Model
₁₂₀₂ (SM) Higgs boson and the top quark, which decay preferentially to b -quarks [103, 104],
₁₂₀₃ and additionally for several Beyond Standard Model (BSM) resonances that readily
₁₂₀₄ decay to heavy flavour quarks [105].

₁₂₀₅ Current flavour tagging algorithms rely on the outputs of a number of low-level
₁₂₀₆ algorithms, which reconstruct various properties of jets using charged particle tracks,
₁₂₀₇ that are then combined using machine learning techniques. This chapter introduces
₁₂₀₈ GN1, a novel ML-based flavour tagging algorithm based on graph neural networks
₁₂₀₉ (GNNs).

₁₂₁₀ As opposed to the previous two-tiered approach described in Chapter 4, which
₁₂₁₁ requires the use of both low- and high-level algorithms, GN1 uses information from
₁₂₁₂ a variable number of charged particle tracks within a jet to predict the jet flavour
₁₂₁₃ without requiring intermediate low-level algorithms. In addition to predicting the
₁₂₁₄ flavour of the jet, the model predicts which physical processes produced the various
₁₂₁₅ jet tracks, and groups the tracks in the jet into vertices. These auxiliary training

1216 objectives provide valuable additional information about the contents of the jet and
1217 enhance the performance of the primary flavour prediction task.

1218 GN1 outperforms the existing ATLAS flavour tagging algorithms as shown in Sec-
1219 tions 6.5.1 and 6.5.2. For a b -jet efficiency of 70%, the light (c)-jet rejection is
1220 improved by a factor of ~ 1.8 (~ 2.1) for jets coming from $t\bar{t}$ decays with transverse
1221 momentum $20 < p_T < 250$ GeV. For jets coming from Z' decays with transverse
1222 momentum $250 < p_T < 5000$ GeV, the light (c)-jet rejection improves by a factor ~ 6
1223 (~ 2.8) for a comparative 30% b -jet efficiency.

1224 In Section 6.1, an overview of the approach used for GN1 is provided. An introduction
1225 to the theory of GNNs is provided in Section 6.2. Details of the experimental setup
1226 are provided in Section 6.3, while the architecture of GN1 is specified in Section 6.4.3.
1227 In Section 6.4.4, the training procedure is described, and in Section 6.5 the results
1228 are shown.

1229 6.1 Motivation

1230 GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. The use
1231 of GNNs offers a natural way to classify jets with variable numbers of unordered
1232 associated tracks (see Section 6.2), while allowing for the inclusion of auxiliary
1233 training objectives [106, 107].

1234 The current ATLAS flavour tagger, DL1r [108], is a deep neural network which takes
1235 the outputs of a number of independently optimised “low-level” algorithms [59] as
1236 its inputs. Each of these low-level algorithms reconstructs a distinct feature of the
1237 experimental signature of heavy flavour jets using the tracks associated to the jet,
1238 and outputs some variables which are discriminating in the jet flavour.

1239 The low-level algorithms are a combination of manually optimised reconstruction
1240 algorithms, for example the SV1 and JetFitter algorithms that reconstruct displaced
1241 decay vertices, and trained taggers such as RNNIP and DIPS that use the IP and
1242 hit information from a variable number of tracks to identify the flavour of the
1243 jet [59, 101, 109, 110]. In contrast GN1 consists of only a single neural network, which
1244 takes the tracks as inputs along with some kinematic information about the jet. As a
1245 result, it does not depend on the outputs of any other flavour tagging algorithm. A

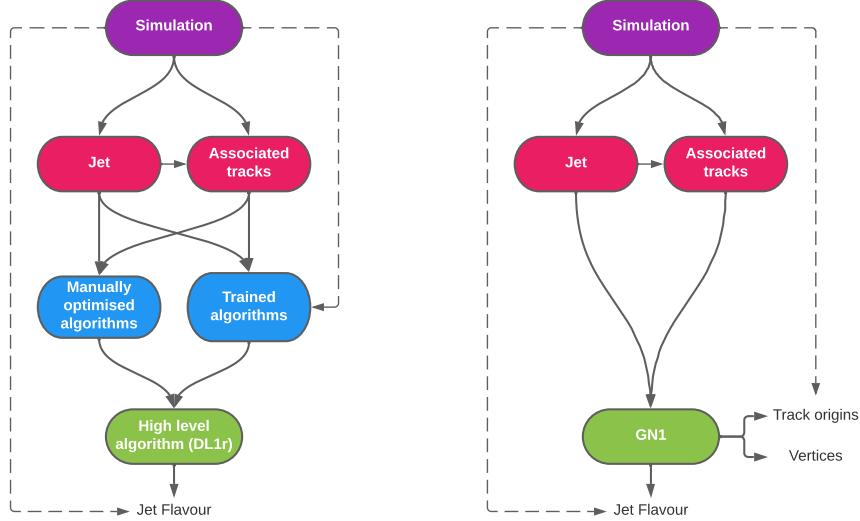


Figure 6.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right). The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional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

1246 simple training of the model fully optimises its parameters, representing a significant
1247 simplification with respect to the optimisation procedure for DL1r. This is particularly
1248 important when optimising the tagger for new regions of phase space (e.g. c -tagging or
1249 high- p_T b -tagging), or when the detector or charged particle reconstruction algorithms
1250 are updated.

1251 GN1 is trained to learn about the internal structure of the jet through the use of two
1252 auxiliary training objectives: the prediction of the underlying physics process from
1253 which each track originated, and the grouping of tracks originating from a common
1254 spatial position (i.e. a common vertex). These auxiliary objectives are meant to
1255 guide the neural network towards a more complete understanding of the underlying
1256 physics inside the jet, thereby removing the need for the low-level algorithms, which
1257 previously contained information about the underlying physics in their design. The
1258 training targets for the primary and auxiliary objectives are extracted from truth
1259 information, i.e. information that is only available in simulation, as opposed to
1260 reconstructed quantities available in both collision data and simulation.

1261 In this chapter, the following advantages of the GN1 approach will be demonstrated:

- 1262 1. GN1 boasts improved performance with respect to the current ATLAS flavour
1263 tagging algorithms, with significantly larger background rejection rates for a
1264 given signal efficiency. Alternatively the rejection rates can be kept fixed for a
1265 substantial increase in signal efficiency, in particular at high- p_T .
- 1266 2. The same network architecture can be easily optimised for a wider variety of
1267 use cases (e.g. c -jet tagging and high- p_T jet tagging) since there are no low-level
1268 algorithms to retune.
- 1269 3. There are fewer algorithms to maintain.
- 1270 4. Alongside the network's prediction of the jet flavour, the auxiliary vertex and
1271 track origin predictions provide more information on why a jet was (mis)tagged
1272 or not. This information can also have uses in other applications, for instance
1273 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. [55].

1274 6.2 Graph Neural Network Theory

1275 6.3 Experimental Setup

1276 6.3.1 Datasets

1277 Datasets used to train the GN1 tagger are the same as described in Section 5.3.1.

1278 Jets are also required not to overlap with a generator-level electron or muon from
1279 W boson decays. All jets are required to have a pseudorapidity $|\eta| < 2.5$ and
1280 $p_T > 20 \text{ GeV}$. Additionally, a standard selection using the Jet Vertex Tagger (JVT)
1281 algorithm at the tight working point is applied to jets with $p_T < 60 \text{ GeV}$ and $|\eta| < 2.4$
1282 in order to suppress pile-up contamination [111].

1283 As previously, truth labelled b -, c - and light- jets are kinematically re-sampled in
1284 p_T and η to ensure identical distributions in these variables. The resulting dataset
1285 contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40% of which are Z' jets. While
1286 DL1r uses 70% $t\bar{t}$ jets and 30% Z' jets, the change in sample composition did not
1287 affect the final performance of GN1. To evaluate the performance of the model, 500k
1288 jets from both the $t\bar{t}$ and Z' samples, which are statistically independent from the
1289 training sample, are used. Track- and jet-level inputs are scaled to have a central
1290 value of zero and a variance of unity before training and evaluation.

1291 6.4 Model Architecture

1292 6.4.1 Model Inputs

1293 GN1 is given two jet variables and 21 tracking related variables for each track
1294 fed into the network. The jet transverse momentum and signed pseudorapidity
1295 constitute the jet-level inputs, with the track-level inputs listed in Table 6.1. If a jet
1296 has more than 40 associated tracks, the first 40 tracks with the largest transverse

1297 IP significance² $s(d_0)$ are selected as inputs. Full track parameter information
1298 and associated uncertainties, along with detailed hit information, carry valuable
1299 information about the jet flavour. In the dense cores of high- p_T jets, tracks are highly
1300 collimated and separation between tracks can be of the same order as the active
1301 sensor dimensions, resulting in merged clusters and tracks which share hits [55]. Due
1302 to the relatively long lifetimes of b -hadrons and c -hadrons, which can traverse several
1303 layers of the ID before decaying and have highly collimated decay products, the
1304 presence of shared or missing hits is a critical signature of heavy flavour jets. Full
1305 track parameter information and associated uncertainties, along with detailed hit
1306 information, carry valuable information about the jet flavour.

1307 Dependence on the absolute value of the azimuthal jet angle ϕ is explicitly removed
1308 by providing only the azimuthal angle of tracks relative to the jet axis. The track
1309 pseudorapidity is also provided relative to the jet axis.

1310 Since heavy flavour hadrons can decay semileptonically, the presence of a recon-
1311 structed lepton in the jet carries discriminating information about the jet flavour.
1312 In addition to the baseline GN1 model, the GN1 Lep variant includes an additional
1313 track-level input, leptonID, which indicates if the track was used in the reconstruction
1314 of an electron, a muon or neither. The muons are required to be combined [112],
1315 and the electrons are required to pass the *VeryLoose* likelihood-based identification
1316 working point [113].

1317 Track selection follows the loose selection used for the fake track classification MVA
1318 in Chapter 5 and described in Table 5.3. This selection was found to improve the
1319 flavour tagging performance compared to previous tighter selections, whilst ensuring
1320 good resolution of tracks and a low fake rate [55]. Section 6.6.1 demonstrates that
1321 further relaxation of the track selection requirements may be warranted.

1322 6.4.2 Auxiliary Training Objectives

1323 In addition to the jet flavour classification, two auxiliary training objectives are
1324 defined. Each auxiliary training objective comes with a training target which, similar

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 [37].

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

Table 6.1: Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55], while split hits are hits used on multiple tracks which have been identified as merged. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1 Lep model.

1325 to the jet flavour label, are truth labels derived from the simulation. The presence
1326 of the auxiliary training objectives improves the jet classification performance as
1327 demonstrated in Section 6.5.4.

1328 The first auxiliary objective is the prediction of the origin of each track within the
1329 jet. Each track is labelled with one of the exclusive categories defined in Table 5.1
1330 after analysing the particle interaction that led to its formation. Since the presence
1331 of different track origins is strongly related to the flavour of the jet, training GN1 to
1332 recognise the origin of the tracks may provide an additional handle on the classification
1333 of the jet flavour. This task may also aid the jet flavour prediction by acting as a
1334 form of supervised attention [114] - in detecting tracks from heavy flavour decays
1335 the model may learn to pay more attention to these tracks.

1336 Displaced decays of b - and c -hadrons lead to secondary and tertiary vertices inside
1337 the jet. Displaced secondary vertices can also occur in light-jets as a result of material
1338 interactions and long-lived particle decays (e.g. K_S^0 and Λ^0). The second auxiliary
1339 objective is the prediction of track-pair vertex compatibility. For each pair of tracks
1340 in the jet, GN1 predicts a binary label, which is given a value 1 if the two tracks
1341 in the pair originated from the same point in space, and 0 otherwise. To derive the
1342 corresponding truth labels for training, truth production vertices within 0.1 mm are
1343 merged. Track-pairs where one or both of the tracks in the pair have an origin label
1344 of either Pileup or Fake are given a label of 0. Using the pairwise predictions from
1345 the model, collections of commonly compatible tracks can be grouped into vertices.
1346 The addition of this auxiliary training objective removes the need for inputs from a
1347 dedicated secondary vertexing algorithm.

1348 Both auxiliary training objectives can be considered as “stepping stones” on the way
1349 to classifying the flavour of the jet. By requiring the model to predict the truth
1350 origin of each track and the vertex compatibility of each track-pair, the model is
1351 guided to learn representations of the jet which are connected to the underlying
1352 physics and therefore relevant for classifying the jet flavour.

1353 6.4.3 Architecture

1354 As discussed above, the GN1 model combines a graph neural network architec-
1355 ture [115] with auxiliary training objectives in order to determine the jet flavour.

1356 Coarse optimisation of the network architecture hyperparameters, for example number
 1357 of layers and number of neurons per layer, has been carried out to maximise the
 1358 tagging efficiency.

1359 The model architecture is based on a previous implementation of a graph neural
 1360 network jet tagger [107]. As compared to the previous approach, GN1 uses a
 1361 only a single graph neural network and makes use of a more sophisticated graph
 1362 neural network layer [116], described below. These changes yield improved tagging
 1363 performance and a significant reduction in training time with respect to the previous
 1364 approach.

1365 The model takes jet- and track-level information as inputs, as detailed in Section 6.4.1.
 1366 The jet inputs are concatenated with each track’s inputs, as shown in Fig. 6.2. The
 1367 combined jet-track vectors are then fed into a per-track initialisation network with
 1368 three hidden layers, each containing 64 neurons, and an output layer with a size of
 1369 64, as shown in Fig. 6.3. The track initialisation network is similar to a Deep Sets
 1370 model [117], but does not include a reduction operation (mean or summation) over
 1371 the output track representations.

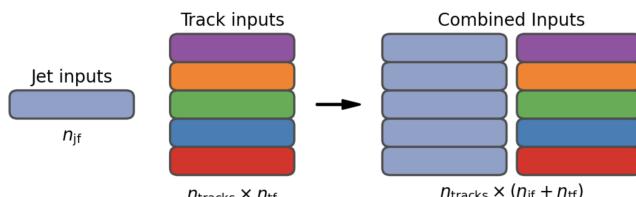


Figure 6.2: The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of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

1372 A fully connected graph is built from the outputs of the track initialisation network,
 1373 such that each node in the graph neighbours every other node. Each node h_i in
 1374 the graph corresponds to a single track in the jet, and is characterised by a feature
 1375 vector, or representation. The per-track output representations from the initialisation
 1376 networks are used to populate the initial feature vectors of each node in the graph.
 1377 In each layer of the graph network, output node representations h'_i are computed by
 1378 aggregating the features of h_i and neighbouring nodes \mathcal{N}_i as described in Ref. [116].
 1379 First, the feature vectors of each node are fed into a fully connected layer \mathbf{W} , to

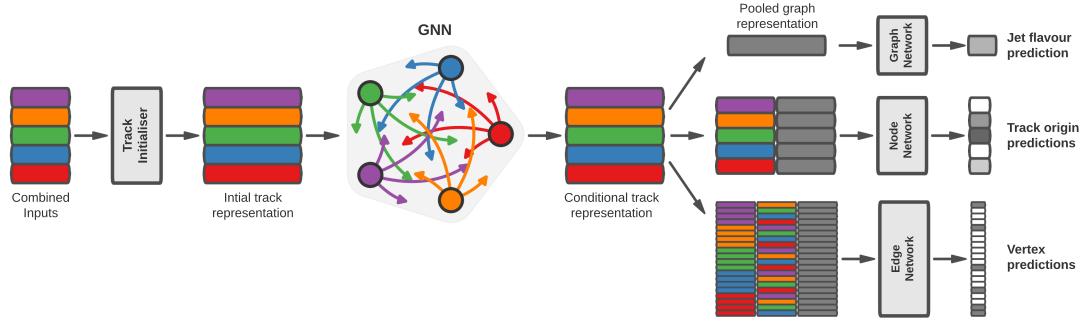


Figure 6.3: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting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

1380 produce an updated representation of each node $\mathbf{W}h_i$. These updated feature vectors
 1381 are used to compute edge scores $e(h_i, h_j)$ for each node pair,

$$e(h_i, h_j) = \mathbf{a}^\perp \theta [\mathbf{W}h_i \oplus \mathbf{W}h_j], \quad (6.1)$$

1382 where \oplus denotes vector concatenation, θ is a non-linear activation function, and \mathbf{a} is
 1383 a learned vector. These edge scores are then used to calculate attention weights a_{ij}
 1384 for each pair of nodes using the softmax function over the edge scores

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

1385 Finally, the updated node representation h'_i is computed by taking the weighted sum
 1386 over each updated node representation $\mathbf{W}h_i$, with weights a_{ij}

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

1387 The above set of operations constitute a single graph network layer. Three such layers
 1388 are stacked to construct the graph network, representing a balance between achieving
 1389 optimal performance and preventing overtraining. The final output node feature
 1390 vectors from the network are representations of each track that are conditional on the
 1391 other tracks in the jet. The output representation for each track is combined using a
 1392 weighted sum to construct a global representation of the jet, where the attention
 1393 weights for the sum are learned during training. Three separate fully connected
 1394 feedforward neural networks are then used to independently perform the different
 1395 classification objectives of GN1. Each of the objectives makes use of the global
 1396 representation of the jet. A summary of the different classification networks used for
 1397 the various training objectives is shown in Table 6.2.

Table 6.2: A summary of GN1’s different classification networks used for the different training objectives. The hidden layers column contains a list specifying the number of neurons in each layer.

Network	Hidden layers	Output size
Node classification network	128, 64, 32	7
Edge classification network	128, 64, 32	1
Graph classification network	128, 64, 32, 16	3

1398 A node classification network, which takes as inputs the features from a single output
 1399 node from the graph network and the global jet representation, predicts the track
 1400 truth origin, as defined in Table 5.1. This network has three hidden layers containing
 1401 128, 64 and 32 neurons respectively, and an output size of seven, corresponding to
 1402 the seven different truth origins.

1403 An edge classification network, which takes as inputs the concatenated representations
 1404 from each pair of tracks and the global jet representation, is used to predict whether
 1405 the tracks in the track-pair belong to a common vertex. The edge network has three
 1406 hidden layers containing 128, 64 and 32 neurons respectively, and a single output,
 1407 which is used to perform binary classification of the track-pair compatibility. These
 1408 predictions are used for the auxiliary training objectives discussed in Section 6.4.2.

1409 A graph classification network takes only the global jet representation as an input,
 1410 and predicts the jet flavour. The graph classification network is comprised of four

1411 fully connected hidden layers with 128, 64, 32 and 16 neurons respectively, and has
1412 three outputs corresponding to the b -, c - and light- jet classes.

1413 **6.4.4 Training**

1414 The full GN1 training procedure minimises the total loss function L_{total} , defined in
1415 Eq. (6.4). This loss is composed of three terms: L_{jet} , the categorical cross entropy
1416 loss over the different jet flavours; L_{vertex} , the binary track-pair compatibility cross
1417 entropy loss averaged over all track-pairs; and L_{track} , the categorical cross entropy loss
1418 for the track origin prediction. L_{vertex} is computed by averaging over all track-pairs
1419 in the batch, and L_{track} is computed by averaging over all tracks in the batch.

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

1420 The different losses converge to different values during training, reflective of differences
1421 in the relative difficulty of the various objectives. As such, L_{vertex} and L_{track} are
1422 weighted by $\alpha = 1.5$ and $\beta = 0.5$ respectively to ensure they converge to similar values,
1423 giving them an equal weighting towards L_{total} . The values of α and β also ensure
1424 that L_{jet} converges to a larger value than L_{vertex} and L_{track} , reflecting the primary
1425 importance of the jet classification objective. In practice, the final performance of
1426 the model was not sensitive to modest variations in the loss weights α and β , or to
1427 pre-training using L_{total} and fine tuning on the jet classification task only. As there
1428 was a significant variation in the relative frequency of tracks of different origins, the
1429 contribution of each origin class to L_{track} was weighted by the inverse of the frequency
1430 of their occurrence. In L_{vertex} , the relative class weight in the loss for track-pairs
1431 where both tracks are from either a b - or c -hadron is increased by a factor of two as
1432 compared with other track-pairs.

1433 The track classification and vertexing objectives are supplementary to the jet clas-
1434 sification objective and trainings can be performed with either the node or edge
1435 networks, or both, removed, as discussed in Section 6.5.4. In these cases, the cor-
1436 responding losses L_{vertex} and L_{track} are removed from the calculation of L_{total} . The

1437 resulting trainings demonstrate how useful the different auxiliary training objectives
1438 are for the primary jet classification objective.

1439 GN1 trainings are run for 100 epochs on 4 NVIDIA V100 GPUs, taking around 25
1440 mins to complete each epoch over the training sample of 30 million jets described
1441 in Section 6.3.1. The Adam optimiser [118] with an initial learning rate of 1e−3,
1442 and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically the
1443 validation loss, calculated on 500k jets, stabilised after around 60 epochs. The epoch
1444 that minimized the validation loss was used for evaluation. GN1 has been integrated
1445 into the ATLAS software [52] using ONNX [119], and jet flavour predictions for the
1446 test sample are computed using the ATLAS software stack.

1447 6.5 Results

1448 The performance of the GN1 tagger is evaluated for both b -tagging and c -tagging use
1449 cases, and for both jets with $20 < p_T < 250$ GeV from the $t\bar{t}$ sample and jets with
1450 $250 < p_T < 5000$ GeV from the Z' sample. Performance is compared to the DL1r
1451 tagger [108], which has been retrained on 75 million jets from the same samples as
1452 GN1. The input RNNIP tagger [110] to DL1r has not been retrained.

1453 The taggers predict the probability that a jet belongs to the b -, c - and light- classes.
1454 To use the model for b -tagging, these probabilities are combined into a single score
1455 D_b , defined as

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

1456 where f_c is a free parameter that determines the relative weight of p_c to p_l in the
1457 score D_b , controlling the trade-off between c - and light-jet rejection performance.
1458 This parameter is set to a value of $f_c = 0.018$ for the DL1r model, obtained through
1459 an optimisation procedure designed to maximise the c - and light-jet rejection of
1460 DL1r [108]. For the GN1 models a value of $f_c = 0.05$ is used, based on a similar
1461 optimisation procedure. The choice of f_c is arbitrary, with the different optimised
1462 values reflecting the relative c - versus light-jet rejection performance of the various

1463 taggers. A fixed-cut working point (WP) defines the corresponding selection applied
1464 to the tagging discriminant D_b in order to achieve a given inclusive efficiency on the
1465 $t\bar{t}$ sample.

1466 The technical implementation of GN1 results in any jet with no associated tracks or
1467 exactly one associated track to be classified as a light-jet. The impact of this on the
1468 tagging performance of GN1 was found to be negligible, with 0.12% of b -jets in the
1469 $t\bar{t}$ sample and 0.02% of b -jets in the Z' sample affected. Of those, 89% of the b -jets
1470 in the $t\bar{t}$ sample and 98% of the b -jets in the Z' sample are classified as light-jets by
1471 DL1r at the 70% $t\bar{t}$ WP.

1472 A comparison of the b -tagging discriminant D_b between DL1r and GN1 is given in
1473 Fig. 6.4. The shapes of the distributions are broadly similar for b -, c - and light-
1474 jets, however, the GN1 model shifts the b -jet distribution to higher values of D_b in
1475 the regions with the best discrimination. The GN1 c -jet distribution is also shifted
1476 to lower values of D_b when compared with DL1r, enhancing the separation and
1477 indicating that GN1 will improve c -jet rejection when compared with DL1r.

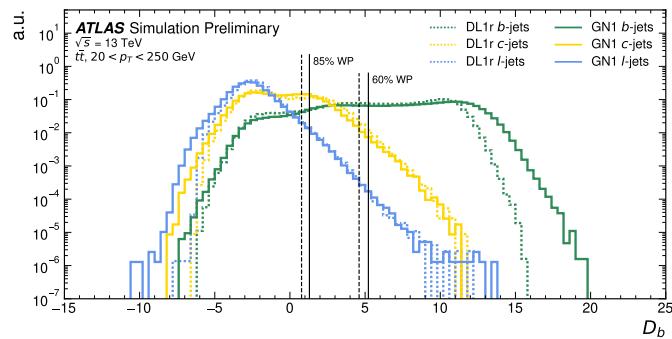


Figure 6.4: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs used by analyses. A value of $f_c = 0.018$ is used in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

1478 6.5.1 b -tagging Performance

1479 The performance of a b -tagging algorithm is quantified by its power to reject c - and
1480 light-jets for a given b -jet tagging efficiency, or WP. In order to compare the b -tagging
1481 performance of the different taggers for the b -jet tagging efficiencies in the range

1482 typically used by analyses, the corresponding c - and light-jet rejection rates are
1483 displayed in Figs. 6.5 and 6.6 for jets in the $t\bar{t}$ and Z' samples respectively. Four
1484 standard WPs with b -jet tagging efficiencies of 60%, 70%, 77% and 85% are used
1485 by physics analyses depending on their specific signal and background requirements.
1486 These WPs are defined using jets in the $t\bar{t}$ sample only. The b -jet tagging efficiencies
1487 for jets in the Z' sample are lower than the corresponding WPs calculated in the $t\bar{t}$
1488 sample, due to the much higher jet p_T range in the Z' sample. For instance the WP
1489 defined to provide a 70% b -jet tagging efficiency on the $t\bar{t}$ sample results in a b -jet
1490 tagging efficiency of $\sim 30\%$ on the Z' sample. To account for this, the range of b -jet
1491 tagging efficiencies displayed in Fig. 6.6 is chosen to span the lower values achieved
1492 in the Z' sample.

1493 For jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$, GN1 demonstrates considerably
1494 better c - and light-jet rejection compared with DL1r across the full range of b -jet
1495 tagging efficiencies probed. The relative improvement depends on the b -jet tagging
1496 efficiency, with the largest improvements found at lower values. At a b -jet tagging
1497 efficiency of 70%, the c -rejection improves by a factor of ~ 2.1 and the light-jet
1498 rejection improves by a factor of ~ 1.8 with respect to DL1r. For high- p_T jets in the
1499 Z' sample with $250 < p_T < 5000 \text{ GeV}$, GN1 also brings considerable performance
1500 improvements with respect to DL1r across the range of b -jet tagging efficiencies
1501 studied. Again, the largest relative improvement in performance comes at lower
1502 b -jet tagging efficiencies. At a b -jet tagging efficiency of 30%, GN1 improves the
1503 c -rejection by a factor of ~ 2.8 and the light-jet rejection by a factor of ~ 6 . An
1504 increasing statistical uncertainty due to the high rejection of background affects the
1505 comparison at lower b -jet tagging efficiencies. It is estimated that for a b -jet tagging
1506 efficiency of 70% in the $t\bar{t}$ sample, $\sim 5\%$ ($\sim 30\%$) of the relative improvement in the
1507 c -jet (light-jet) rejection comes from loosening the track selection and for a b -jet
1508 tagging efficiency of 30% in the Z' the corresponding number is $\sim 10\%$ for both c -jets
1509 and light-jets. Given the sophisticated exploitation of low-level information, further
1510 studies are needed to confirm if the performance gain is also observed in experimental
1511 data.

1512 The GN1 Lep variant shows improved performance with respect to the baseline GN1
1513 model, demonstrating the additional jet flavour discrimination power provided by the
1514 leptonID track input. For jets in the $t\bar{t}$ sample, the relative c -rejection improvement
1515 with respect to DL1r at the 70% b -jet WP increases from a factor of ~ 2.1 for GN1 to

1516 a factor of ~ 2.8 for GN1 Lep. The improvement in light-jet rejection also increases
1517 from a factor of ~ 1.8 to ~ 2.5 at this WP. For jets in the Z' sample, the relative
1518 c -rejection (light-jet rejection) improvement with respect to DL1r increases from a
1519 factor of ~ 2.8 to ~ 3 (~ 6 to ~ 7.5) at a b -jet tagging efficiency of 30%. As shown in
1520 Fig. 6.7, the greatest improvement of GN1 Lep over GN1 is seen at low p_T .

1521 The performance of the taggers is strongly dependent on the jet p_T . Charged particle
1522 reconstruction is particularly challenging within high- p_T jets [55]. The multiplicity of
1523 fragmentation particles increases as a function of p_T , while the number of particles
1524 from heavy flavour decays stays constant. Collimation of particles inside the jet
1525 increases and approaches the granularity of the tracking detectors, making it difficult
1526 to resolve the trajectories of different particles. Furthermore, at high p_T , heavy
1527 flavour hadrons will travel further into the detector before decaying. For hadrons
1528 which traverse one or more layers of the ID before decaying, the corresponding decay
1529 tracks may pick up incorrect hits, left by the hadron itself or fragmentation particles,
1530 in the inner layers of the detector, reducing the accuracy of the reconstructed track
1531 parameters. These factors contribute to a reduced reconstruction efficiency for heavy
1532 flavour tracks, and a general degradation in quality of tracks inside the core of a jet,
1533 which in turn reduces the jet classification performance.

1534 In order to study how the b -jet tagging efficiency of the taggers varies as a function
1535 of jet p_T , the b -jet tagging efficiency as a function of p_T for a fixed light-jet rejection
1536 of 100 in each bin is shown in Fig. 6.7. For jets in the $t\bar{t}$ sample, at a fixed light-jet
1537 rejection of 100, GN1 improves the b -jet tagging efficiency by approximately 4%
1538 across all jet p_T bins. GN1 Lep shows improved performance with respect to GN1,
1539 in particular at lower p_T , with the relative increase in the b -jet tagging efficiency
1540 going from 4% to 8%. For jets in the Z' sample, GN1 has a higher b -jet tagging
1541 efficiency than DL1r across the p_T range, with the largest relative improvement in
1542 performance, approximately a factor of 2, found at jet $p_T > 2$ TeV. GN1 outperforms
1543 DL1r across the entire jet p_T spectrum studied. The performance was also evaluated
1544 as a function of the average number of pileup interactions in an event, and was found
1545 to have no significant dependence on this quantity.

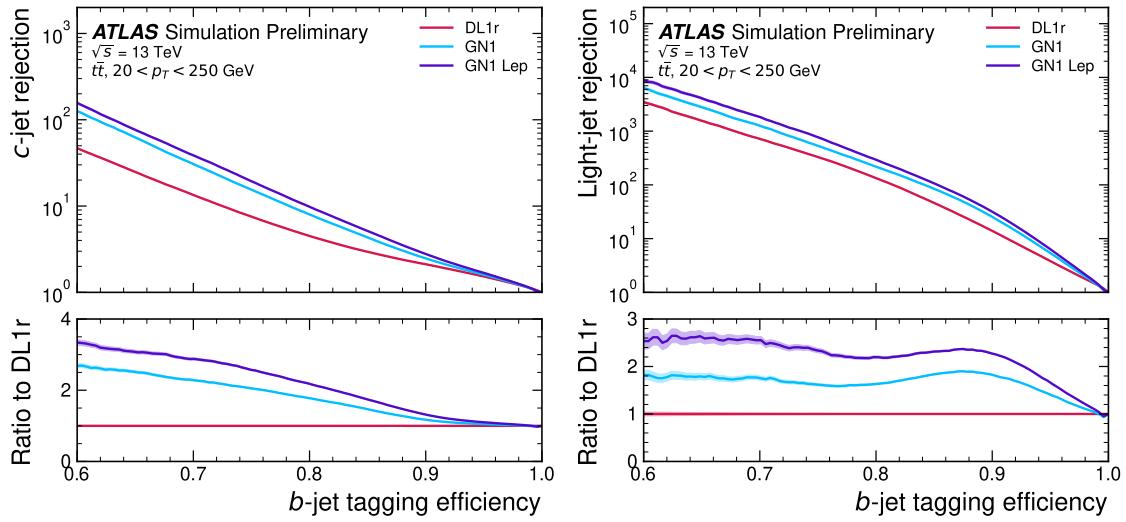


Figure 6.5: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 75\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions of phase space.

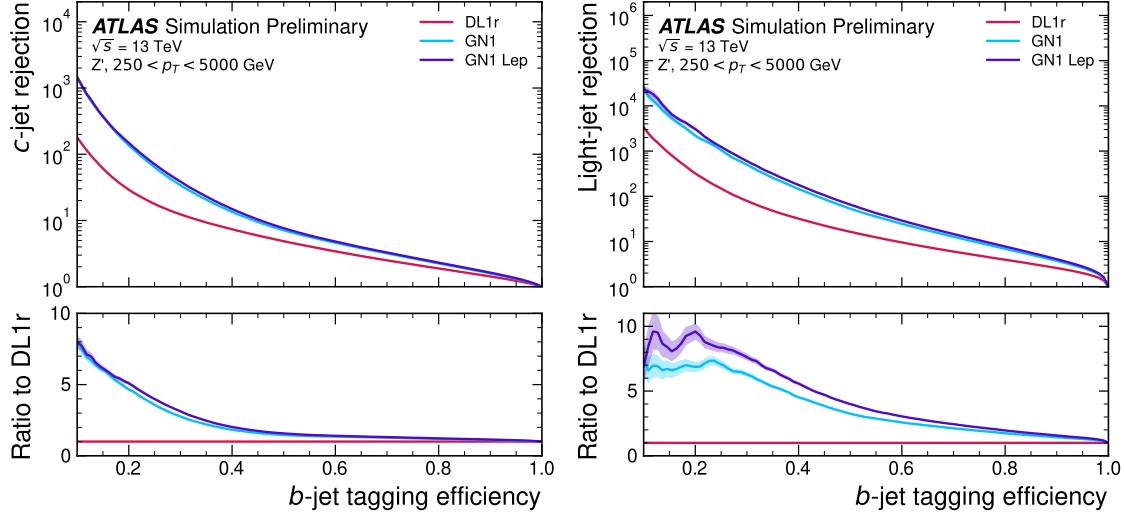


Figure 6.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 20\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions of phase space.

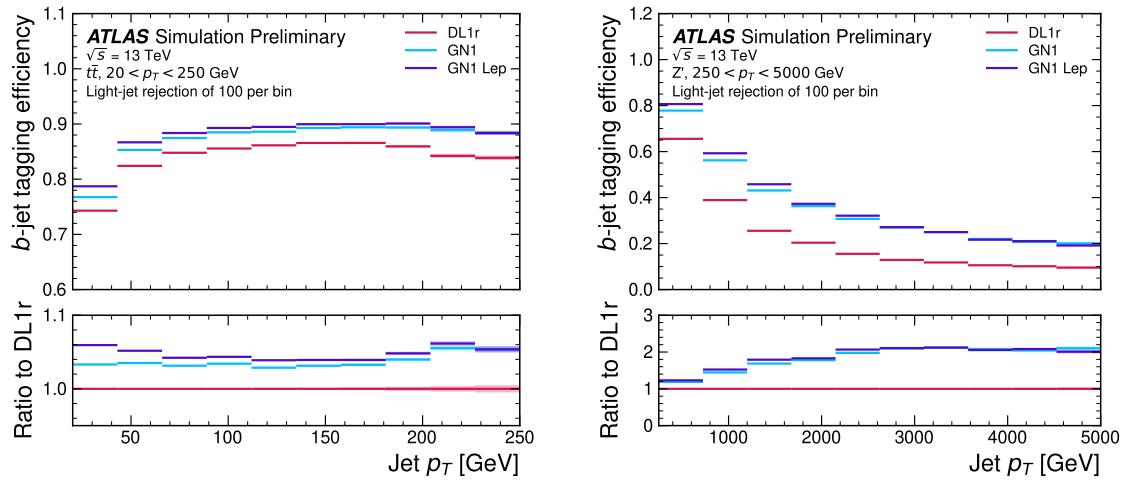


Figure 6.7: The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1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

6.5.2 c -tagging Performance

Since GN1 does not rely on any manually optimised low-level tagging algorithms, which may not have been optimised for c -tagging, tagging c -jets presents a compelling use case for GN1. To use the model for c -tagging, the output probabilities are combined into a single score D_c , defined similarly to Eq. (6.5) as

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

A value of $f_b = 0.2$ is used for all models. Similar to Section 6.5.1, performance of the different taggers is compared by scanning through a range of c -jet tagging efficiencies and plotting the corresponding b - and light-jet rejection rates. As in Section 6.5.1, WPs are defined using jets in the $t\bar{t}$ sample. Standard c -jet tagging efficiency WPs are significantly lower in comparison with the b -tagging WPs in order to maintain reasonable b - and light-jet rejection rates. This is reflected in the range of c -jet tagging efficiencies used in Figs. 6.8 and 6.9. In Fig. 6.8, which displays the c -tagging performance of the models on the jets in the $t\bar{t}$ sample, GN1 performs significantly better than DL1r. The b - and light-jet rejection improve most at lower c -jet tagging efficiencies, with both background rejections increasing by a factor of 2 with respect to DL1r at a c -jet tagging efficiency of 25%. GN1 Lep outperforms GN1, with the b -rejection (light-jet rejection) relative improvement increasing from a factor of 2 to 2.1 (2 to 2.3) at the 25% c -jet WP. Fig. 6.9 shows the c -tagging performance on the jets in the Z' sample. Both GN1 and GN1 Lep perform similarly, improving the b -rejection by 60% and the light-jet rejection by a factor of 2 at the 25% c -jet WP.

6.5.3 Jet Display Diagrams

write

6.5.4 Ablations

Several ablations, the removal of components in the model to study their impact, are carried out to determine the importance of the auxiliary training objectives of

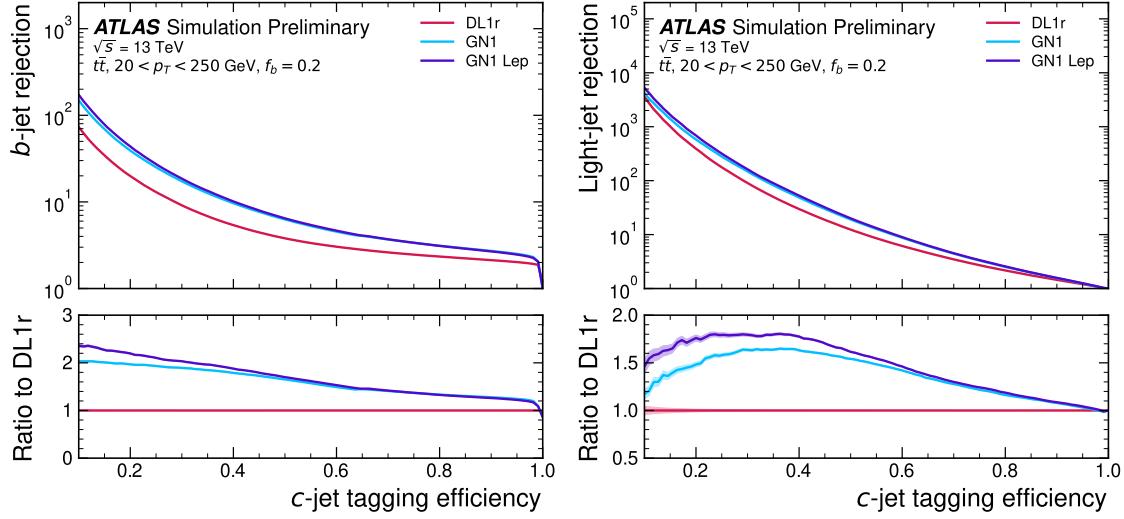


Figure 6.8: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. At c -jet tagging efficiencies than $\sim 25\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

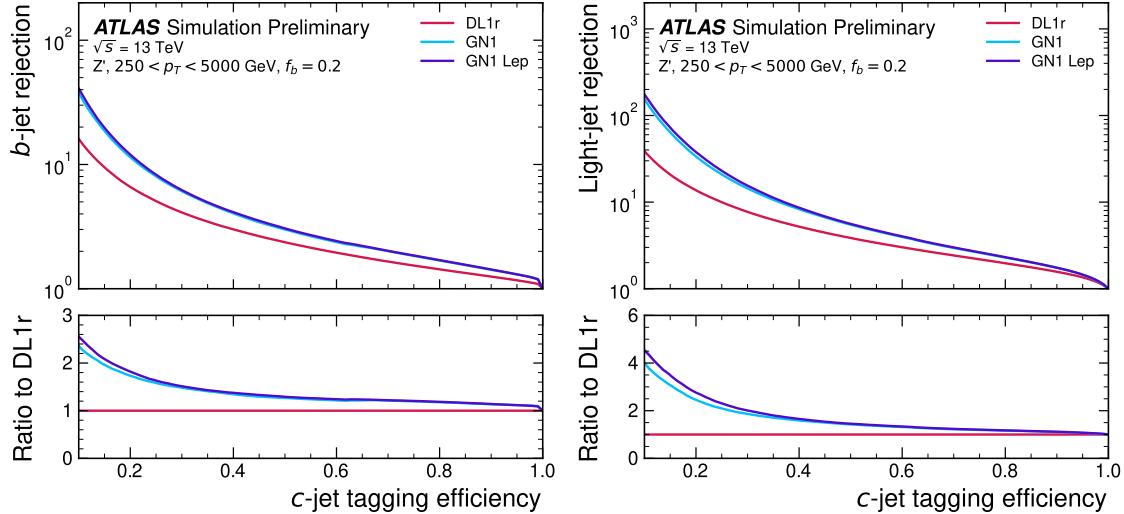


Figure 6.9: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000 \text{ GeV}$. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The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

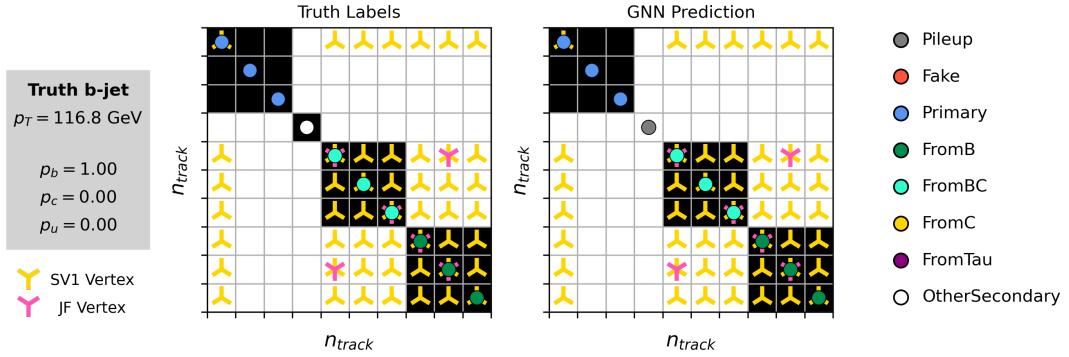


Figure 6.10: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked.

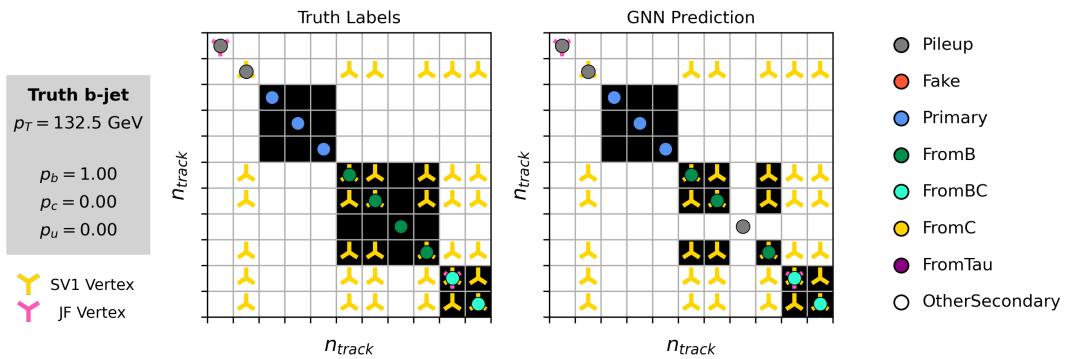


Figure 6.11: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked.

1572 GN1 to the overall performance. The “GN1 No Aux” variant retains the primary jet
 1573 classification objective, but removes both track classification and vertexing auxiliary
 1574 objectives (see Section 6.4.2) and as such only minimises the jet classification loss.
 1575 The “GN1 TC” variant includes track classification but not vertexing, while “GN1
 1576 Vert” includes vertexing, but not track classification.

1577 For jets in both the $t\bar{t}$ and Z' samples, the models without one or both of the auxiliary
 1578 objectives display significantly reduced c - and light-jet rejection when compared with
 1579 the baseline GN1 model, as shown in Figs. 6.12 and 6.13. For jets in the $t\bar{t}$ sample,
 1580 the performance of GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert
 1581 perform similarly to each other. For jets in the Z' sample, the GN1 No Aux model
 1582 shows a clear improvement in c - and light-jet rejection when compared with DL1r at
 1583 lower b -jet tagging efficiencies. Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1
 1584 Vert perform similarly, and bring large gains in background rejection when compared
 1585 with GN1 No Aux, but the combination of both auxiliary objectives yields the best
 1586 performance.

1587 It is notable that the GN1 No Aux model matches or exceeds the performance of
 1588 DL1r without the need for inputs from the low-level algorithms. This indicates that
 1589 the performance improvements enabled by GN1 appear to be able to compensate for
 1590 the removal of the low-level algorithm inputs. The GN1 TC and GN1 Vert variants
 1591 each similarly outperform DL1r, demonstrating that both contribute to the overall
 1592 high performance of the baseline model.

1593 6.5.5 Inclusion of Low-Level Vertexing Algorithms

1594 GN1 does not include inputs from low-level tagging algorithms, including the vertexing
 1595 tools SV1 and JetFitter [59]. Since these algorithms are known to improve the
 1596 performance of DL1r, it was feasible that their inclusion in GN1 may further improve
 1597 on the performance of the GN1 models. In a dedicated training of GN1 the SV1 and
 1598 JetFitter tagger outputs were added to the GN1 jet classification network as an input,
 1599 similar to their use in DL1r. These outputs include information on the reconstructed
 1600 vertices, including the number of vertices, the vertex mass, displacement, and other
 1601 properties. In addition, the index of the reconstructed SV1 or JetFitter vertices were
 1602 included as two track-level inputs to GN1. The jet classification performance of this
 1603 GN1 model was not significantly different to the baseline model, and in some cases

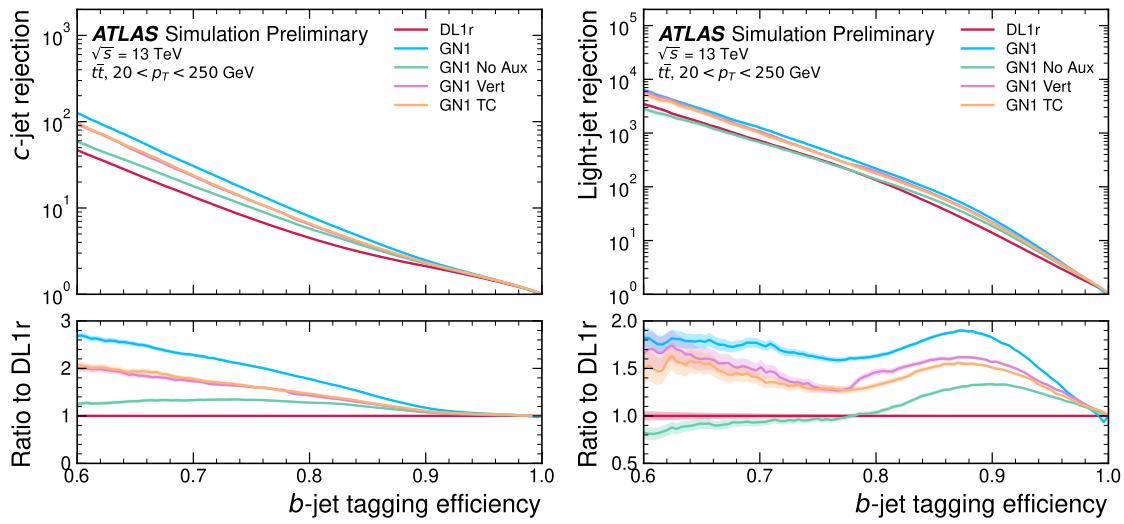


Figure 6.12: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 65\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.

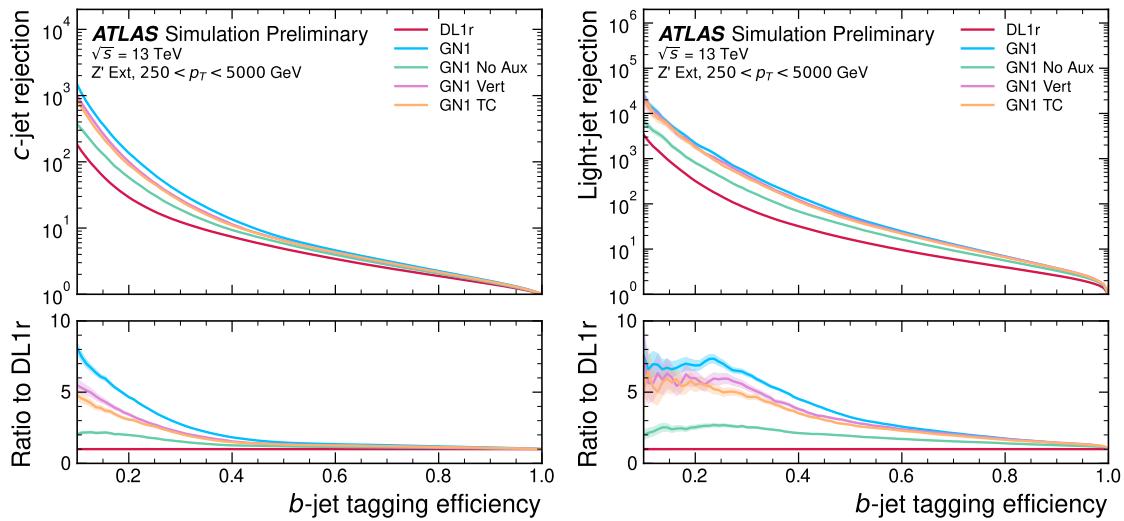


Figure 6.13: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 25\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.

1604 the performance was slightly reduced. A dedicated look at the vertexing performance
1605 of GN1 with some comparisons to SV1 and JetFitter is found in Section 6.5.6

1606 **6.5.6 Vertexing Performance**

1607 From the track-pair vertex prediction described in Section 6.4.2, tracks can be
1608 partitioned into compatible groups representing vertices (see [107]). As such, GN1
1609 is able to be used to perform vertex “finding”, but not vertex “fitting”, i.e. the
1610 reconstruction of a vertex’s properties, which currently still requires the use of a
1611 dedicated vertex fitter. In order to study the performance of the different vertexing
1612 tools inside b -jets, the truth vertex label of the tracks, discussed in Section 6.4.2, are
1613 used. To estimate the efficiency with which GN1 manages to find vertices inclusively,
1614 vertices from GN1 containing tracks identified as coming from a b -hadron are merged
1615 together and compared to the inclusive truth decay vertices that result from a
1616 b -hadron decay (where if there are multiple distinct truth vertices from a b -hadron
1617 decay they are also merged together). Vertices are compared with the target truth
1618 vertex and the number of correctly and incorrectly assigned tracks is computed.
1619 Since secondary vertex information is only recovered for reconstructed tracks, an
1620 efficiency of 100% here denotes that all possible secondary vertices are recovered
1621 given the limited track reconstruction efficiency. A vertex is considered matched if
1622 it contains at least 65% of the tracks in the corresponding truth vertex, and has a
1623 purity of at least 50%. GN1 manages to achieve an inclusive reconstruction efficiency
1624 in b -jets of $\sim 80\%$, demonstrating that it effectively manages to identify the displaced
1625 vertices from b -hadron decays.

1626 **More detail**

1627 In order to study the performance of the different vertexing tools inside b -jets, the
1628 truth vertex label of the tracks, discussed in Section 6.4.2, is used. The reconstructed
1629 vertices from GN1, SV1 and JetFitter are compared to the target truth vertices in
1630 order to calculate the efficiencies of the different vertexing tools. Since secondary
1631 vertex information is only recovered for reconstructed tracks, an efficiency of 100%
1632 here denotes that all possible secondary vertices are recovered given the limited track
1633 reconstruction efficiency.

1634 There are several caveats to a comparison of the vertexing tools which are a result
1635 of the different approaches they take to vertexing. SV1 and JetFitter are designed
1636 to only find secondary vertices in the jet, whereas GN1 is also trained to determine
1637 which tracks in the jet belong to the primary vertex (the vertex of the hard scatter
1638 pp interaction). To account for this the GN1 vertex with the largest number of
1639 predicted primary tracks is excluded from the vertex finding efficiency calculation.
1640 While JetFitter and GN1 aim to resolve each displaced vertex inside the jet, such
1641 that secondary vertices from b -hadron decays are found separately to tertiary vertices
1642 from $b \rightarrow c$ decay chains, SV1 by design attempts to find a single inclusive vertex
1643 per jet. This inclusive vertex groups inclusive b -hadron decays. These are tracks
1644 from the b -hadron decay itself (FromB) and tracks from $b \rightarrow c$ decays (FromBC).
1645 In order to fairly compare the performance of the different tools, both the exclusive
1646 and inclusive vertex finding efficiency is studied. For the exclusive vertex finding
1647 case JetFitter and GN1 can be directly compared, while a comparison with SV1 is
1648 not possible due to aforementioned design constraints. The inclusive vertex finding
1649 performance of all three tools can be compared using the procedure outlined below.

1650 The starting point for the secondary vertex finding efficiency in both the exclusive
1651 and inclusive cases is to select truth secondary vertices as those containing only
1652 inclusive b -hadron decays to be considered as initial targets. For exclusive vertex
1653 finding, these truth secondary vertices can be used directly as the denominator for the
1654 efficiency calculation. Meanwhile for the inclusive efficiency all such truth secondary
1655 vertices in the jet are merged into a single inclusive target vertex. Correspondingly,
1656 for the inclusive vertex finding case, the vertices found by JetFitter are merged into
1657 a single vertex, and the vertices found by GN1 with at least one predicted inclusive
1658 b -hadron decay track are also merged similarly. SV1 does not require any vertex
1659 merging.

1660 Next, in both cases for each truth secondary vertex, vertices in the jet found by the
1661 different vertexing tools are compared with the target truth vertex. The number
1662 of correctly and incorrectly assigned tracks is computed. In order to call a vertex
1663 efficient, it is required to contain at least 65% of the tracks in the corresponding
1664 truth vertex, and to have a purity of at least 50%. Single track vertices are required
1665 to have a purity of 100%. Additionally, for GN1 only, at least one track in the vertex
1666 is required to have a predicted heavy flavour origin.

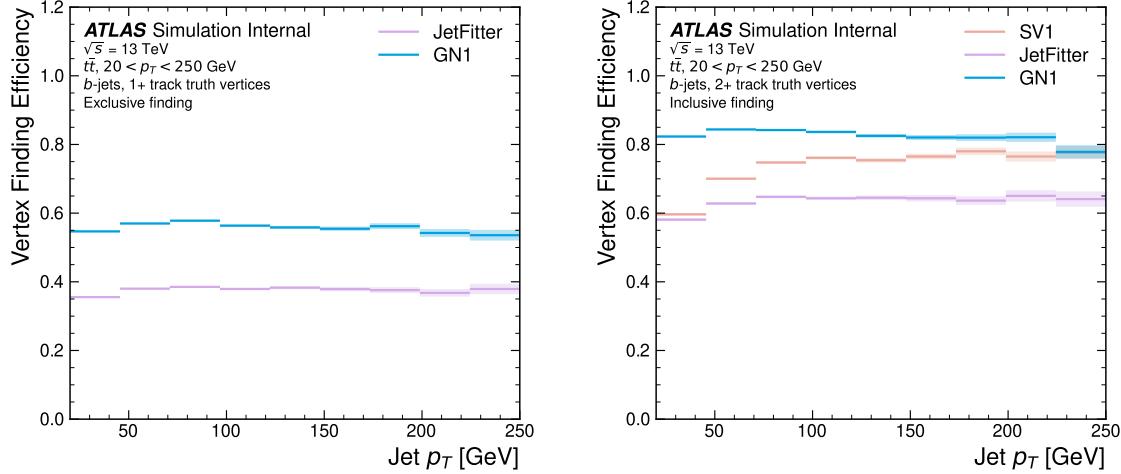


Figure 6.14: Vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50%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

1667 Vertex finding efficiencies for b -jets in the $t\bar{t}$ sample are displayed as a function of p_T
 1668 separately for the inclusive and exclusive approaches in Fig. 6.14. For b -jets in the $t\bar{t}$
 1669 sample with $20 < p_T < 250$ GeV, the exclusive vertex finding efficiency of JetFitter
 1670 and GN1 is relatively flat as a function of p_T . Of the truth secondary vertices in this
 1671 p_T region, JetFitter efficiently finds approximately 40% and GN1 finds approximately
 1672 55%. When finding vertices inclusively the vertex finding efficiency is generally higher.
 1673 An increased dependence on p_T is also visible for JetFitter and SV1. As the jet p_T
 1674 increases from 20 GeV to 100 GeV, the efficiency of JetFitter increases from 55% to
 1675 65%. In the same range, the efficiency of SV1 increases from 55% to 75%. GN1
 1676 displays less dependence on p_T than JetFitter and SV1, efficiently finding upwards
 1677 of 80% of vertices in b -jets in this p_T region. For b -jets with $p_T > 100$ GeV, JetFitter
 1678 finds approximately 65% of vertices, SV1 finds approximately 75% of vertices, and
 1679 GN1 finds approximately 80% of vertices.
 1680 For b -jets in the Z' sample, the vertex finding efficiency drops steeply with increasing
 1681 p_T up until $p_T = 3$ TeV. GN1 outperforms SV1 and JetFitter across the p_T spectrum.
 1682 In the first bin, the efficiency of GN1 is 75%, while the efficiencies of SV1 and JetFitter
 1683 are around 60%. The efficiency of SV1 drops rapidly to almost zero above 3 TeV,
 1684 while JetFitter and GN1 retain approximately 30% efficiency. Fig. 6.15 compares

the exclusive vertex finding efficiencies of JetFitter and GN1 for multitrack vertices. JetFitter finds 45-50% of vertices in b -jets in the $t\bar{t}$ sample, while GN1 finds 60-65%. For b -jets in the Z' sample, JetFitter finds 35% of vertices in the first bin, dropping to 20% of vertices above 2 TeV. GN1 finds 55% of vertices in the first bin, dropping to 30% above 2 TeV.

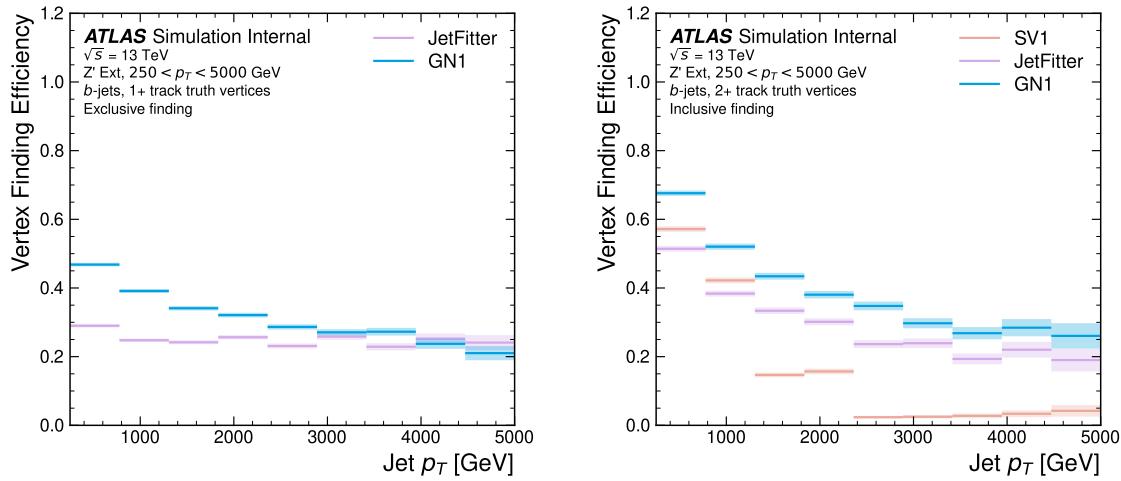


Figure 6.15: Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly.

6.5.7 Track Classification Performance

As discussed in Section 6.4.2, one of the auxiliary training objectives for GN1 is to predict the truth origin of each track in the jet. Since the equivalent information is not provided by any of the existing flavour tagging tools, as a benchmark a multi-class classification multilayer perceptron (MLP) is trained on the same tracks used for the baseline GN1 training. The model uses the same concatenated track-and-jet inputs as GN1 (see Section 6.4.1), but processes only a single track at a time. The model is comprised of five densely connected layers with 200 neurons per layer, though the performance was not found to be strongly sensitive to changes in the network structure. To measure the track classification performance, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve is computed for each origin class using a one versus all classification approach. The AUCs for the

1702 different truth origin classes are averaged using both an unweighted and a weighted
 1703 approach. The unweighted mean treats the performance of each class equally, while
 1704 the weighted mean uses the fraction of tracks from each origin as a weight. As seen
 1705 in Table 6.3, GN1 outperforms the MLP, both at $20 < p_T < 250 \text{ GeV}$ for jets in
 1706 the $t\bar{t}$ sample, and at $250 < p_T < 5000 \text{ GeV}$ for jets in the Z' sample. For tracks in
 1707 jets in the $t\bar{t}$ sample, GN1 can reject 65% of fake tracks while retaining more than
 1708 99% of good tracks. The GN1 model has two advantages over the MLP which can
 1709 explain the performance improvement. Firstly, the mixing of information between
 1710 tracks, enabled by the fully connected graph network architecture as discussed in
 1711 Section 6.4.3, is likely to be beneficial since the origins of different tracks within
 1712 a jet are to some extent correlated. Secondly, the jet classification and vertexing
 1713 objectives can be considered auxiliary to the track classification task, and may bring
 1714 improved track classification performance with respect to the standalone MLP.

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

1715 Fig. 6.16 shows the track origin classification ROC curves for the different track
 1716 origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve legibility of the
 1717 figure, the heavy flavour truth origins have been combined weighted by their relative
 1718 abundance, as have the Primary and OtherSecondary labels. In jets in both the $t\bar{t}$
 1719 and Z' samples, the AUC of the different (grouped) origins is above 0.9, representing
 1720 good classification performance. Fake tracks, followed by pileup tracks, are the
 1721 easiest to classify in both samples.

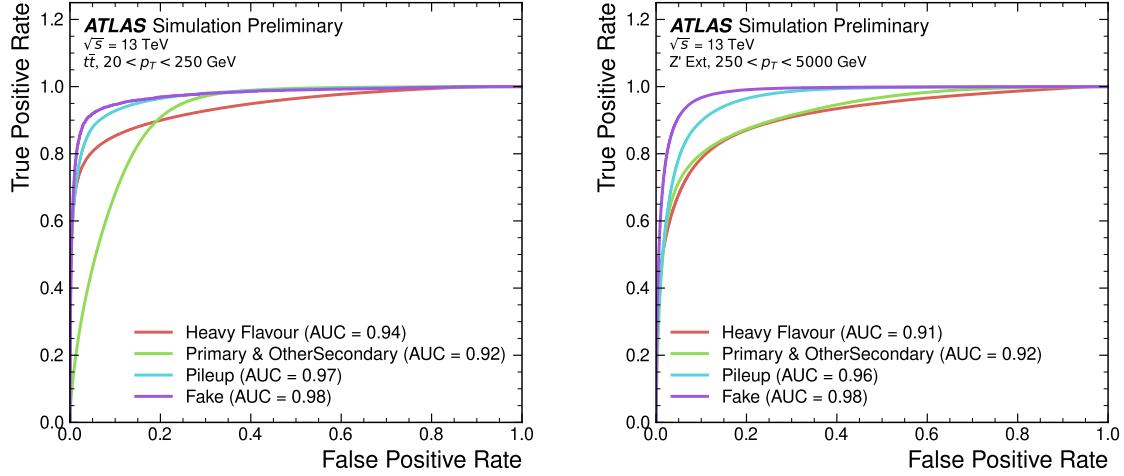


Figure 6.16: ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the 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.

1722 6.6 Extensions

1723 6.6.1 Looser Track Selection

1724 The track selections used to produce the main results are listed in Table 5.3. This
 1725 selection includes a cut on the number of shared silicon modules $N_{\text{shared}}^{\text{Si}}$. This value
 1726 is calculated as

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

1727 where $N_{\text{shared}}^{\text{Pix}}$ is the number of shared pixel hits and $N_{\text{shared}}^{\text{SCT}}$ is the number of shared
 1728 SCT modules on a track. The nominal cut used elsewhere in this thesis is $N_{\text{shared}}^{\text{Si}} < 2$.
 1729 As the rate of shared hits is significantly higher for b -hadron decay tracks than for
 1730 other tracks, this cut rejects a significant proportion of these tracks. Figs. 6.17
 1731 and 6.18 show the result of training the GN1 tagger with the full relaxation of this
 1732 cut, i.e. allowing tracks with any number of shared hits. The shared hit requirements
 1733 applied by the ambiguity solver as part of track reconstruction (see Section 3.4.1)

1734 are still applied. In addition, the maximum allowed value of d_0 is increased from
1735 3.5 mm to 5.5 mm.

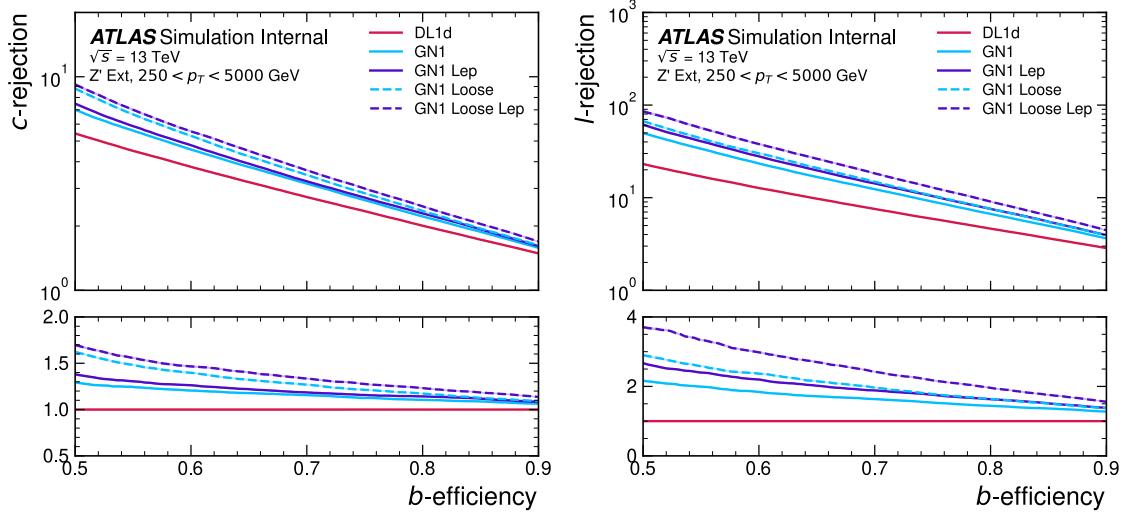


Figure 6.17: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value of $f_c = 0.018$ 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.

1736 6.6.2 Model Optimisations

1737

write

1738 6.6.3 Additional Auxiliary Tasks

1739

write

1740 6.7 Conclusion

1741 A novel jet tagger, GN1, with a graph neural network architecture and trained
1742 with auxiliary training targets, is presented and now fully implemented in the
1743 ATLAS software. GN1 is shown to improve flavour tagging performance with respect
1744 to DL1r, the current default ATLAS flavour tagging algorithm, when compared
1745 in simulated collisions. GN1 improves c - and light-jet rejection for jets in the $t\bar{t}$
1746 sample with $20 < p_T < 250$ GeV by factors of ~ 2.1 and ~ 1.8 respectively at a b -jet

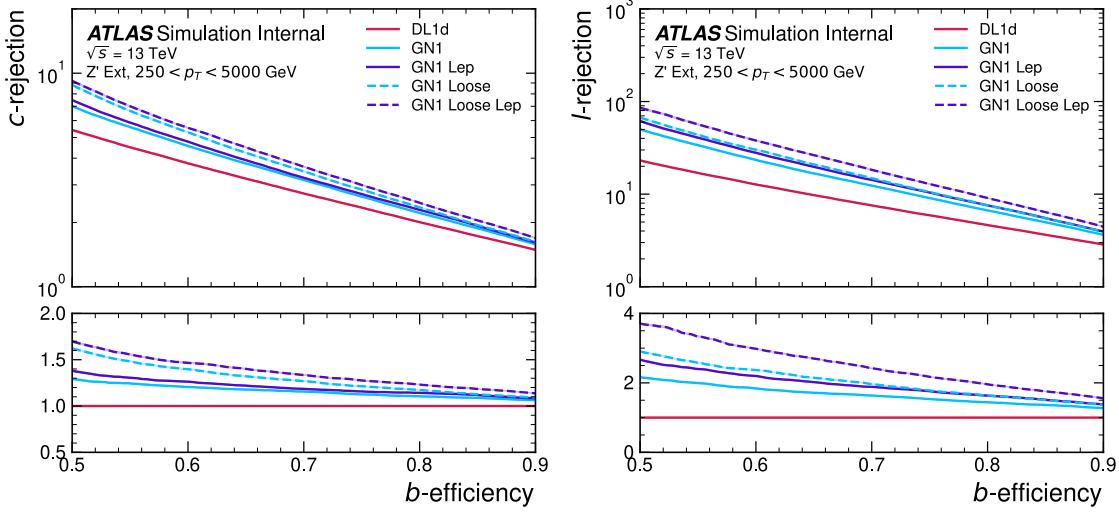


Figure 6.18: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions.

tagging efficiency of 70% when compared with DL1r. For jets in the Z' sample with $250 < p_T < 5000$ GeV, GN1 improves the c -rejection by a factor of ~ 2.8 and light-jet rejection by a factor of ~ 6 for a comparative b -jet efficiency of 30%. Previous multivariate flavour tagging algorithms relied on inputs from low-level tagging algorithms, whereas GN1 needs no such inputs, making it more flexible. It can be easily fully optimised via a retraining for specific flavour tagging use cases, as demonstrated with c -tagging and high- p_T b -tagging, without the need for time-consuming retuning of the low-level tagging algorithms. The model is also simpler to maintain and study due to the reduction of constituent components. GN1 demonstrates improved track classification performance when compared with a simple per-track MLP and an efficiency of $\sim 80\%$ for inclusive vertex finding in b -jets. The auxiliary track classification and vertex finding objectives are shown to significantly contribute to the performance in the jet classification objective, and are directly responsible for the improvement over DL1r. Further studies need to be undertaken to verify the performance of GN1 on collision data.

future work

- 1764 • more aux tasks, regression
- 1765 • Future work will include adding modalities from other subdetectors and com-
- 1766 bining these in an efficient way.

₁₇₆₇

Chapter 7

₁₇₆₈

Boosted VHbb Analysis

₁₇₆₉ The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [12, 13],
₁₇₇₀ is predicted by the standard model to decay primarily to a pair of b -quarks, with a
₁₇₇₁ branching factor of 0.582 ± 0.007 for $m_H = 125\text{ GeV}$ [24]. Observation of this decay
₁₇₇₂ mode was reported by ATLAS [103] and CMS [26] in 2018, establishing the first
₁₇₇₃ direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks
₁₇₇₄ (see Section 2.2.2). The $H \rightarrow b\bar{b}$ process is also important for constraining the total
₁₇₇₅ decay width of the Higgs [120].

₁₇₇₆ Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion
₁₇₇₇ as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet
₁₇₇₈ background and so overall sensitivity to the Higgs is low. The QCD multijet
₁₇₇₉ background refers to events containing one or more strongly produced jets which are
₁₇₈₀ not the decay product of heavy resonances, for example $g \rightarrow q\bar{q}$. The gluon-gluon
₁₇₈₁ fusion channel contains to leading order only jets in the final state, and therefore
₁₇₈₂ it is extremely difficult to distinguish signal events from the overwhelming multijet
₁₇₈₃ background. The $H \rightarrow b\bar{b}$ observation therefore searched for Higgs bosons produced
₁₇₈₄ in association with a vector boson V (where V can be a W or Z boson) which
₁₇₈₅ subsequently decays leptonically. The leptonic final states from the decay of the
₁₇₈₆ vector boson allow for leptonic triggering whilst at the same time significantly
₁₇₈₇ reducing the multijet background.

₁₇₈₈ A closely related analysis [121] has more recently measured the associated production
₁₇₈₉ of a Higgs boson decaying to b -quarks in events where the vector and Higgs bosons are
₁₇₉₀ highly boosted. The analysis is outlined in Section 7.1. Modelling studies performed
₁₇₉₁ by the author are detailed in Section 7.2, and the results of the analysis are presented

1792 in Section 7.4. The author contributed to various signal and background modelling
1793 studies, fit studies, and to the diboson unblinding effort. This analysis has been
1794 published in Ref. [121]. Figures and tables from Ref. [121] are reproduced here.

1795 7.1 Analysis Overview

1796 The boosted VH , $H \rightarrow b\bar{b}$ analysis is focused on the high transverse momentum
1797 regime, which has the benefit of being more sensitive to physics beyond the Standard
1798 Model [122], but the disadvantage of being more challenging due to the increased dif-
1799 ficulty in the accurate reconstructed of highly energy events (discussed in Chapter 4).
1800 In order to focus on the high- p_T regime, the reconstructed vector boson is required to
1801 have $p_T^V > 250$ GeV (see Section 7.1.2). Events are also split into two p_T^V bins with the
1802 first bin covering $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and the second covering $p_T^V > 400 \text{ GeV}$,
1803 which allows the analysis to account for the improved signal-to-background in the
1804 high- p_T regime.

1805 The previous ATLAS analysis in Ref. [103] was primarily sensitive to vector bosons
1806 with a more modest p_T^V boost in the region of 100–300 GeV. In this regime, the Higgs
1807 candidate was reconstructed using a pair of jets with radius parameter of $R = 0.4$,
1808 called small- R jets. However in the high- p_T regime, the decay products of the Higgs
1809 boson become increasingly collimated and the small- R jets may overlap. In order to
1810 avoid the associated problems and to aid in the reconstruction of the Higgs boson
1811 candidate, the present analysis uses instead a large- R jet with radius parameter
1812 $R = 1.0$ to reconstruct the Higgs boson candidate in all channels (see Section 3.4.3).
1813 The Higgs candidate is required to have exactly two ghost-assciated and b -tagged
1814 variable-radius track-jets. The candidate large- R jet is reconstructed using jet
1815 substructure techniques, for example it is trimmed by removing soft and wide-angle
1816 components, which helps to remove particles from the underlying event and pileup
1817 collisions [123]. Refer to Section 3.4.3 for more details on jet reconstruction.

1818 On top of the binning in p_T^V , selected events are further categorised into 0-, 1- and
1819 2-lepton channels depending on the number of selected charged leptons (electrons
1820 and muons) are present in the reconstructed final state (also referred to as 0L, 1L,
1821 and 2L respectively). The 0-lepton channel targets the $ZH \rightarrow \nu\nu b\bar{b}$ process, the
1822 1-lepton channel targets $WH \rightarrow \ell\nu b\bar{b}$, and the 2-lepton channel targets $ZH \rightarrow \ell\ell b\bar{b}$,

where ℓ is an electron or muon and ν is a neutrino. Each channel has a dedicated set of selections which are listed in more detail in Section 7.1.3. Events in the 0- and 1-lepton channels are further split depending on the number of additional small- R jets not matched to the Higgs-jet candidate. The high-purity signal region (HP SR) has zero such jets, while the low-purity signal region (LP SR) has one or more. The 0- and 1-lepton channels also make use of a dedicated $t\bar{t}$ control region, described in Section 7.1.4. A complete overview of the different analysis regions is given in Table 7.1.

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

Table 7.1: Summary of the definition of the analysis regions. Signal enriched regions are marked with the label SR. There are regions with relatively large signal purity (HP SR) and with low purity (LP SR). Background enriched regions are marked with the label CR. The shorthand ‘‘add’’ stands for additional small- R jets, i.e. number of small- R jets not matched to the Higgs-jet candidate.

7.1.1 Data & Simulated Samples

The analysis uses pp collision data recorded between 2015 and 2018 by the ATLAS detector [31] during Run 2 at the LHC. This dataset corresponds to an integrated luminosity of 139 fb^{-1} .

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

An overview of the MC simulated samples used in the analysis is given in Table 7.2. These samples are used to model the signal and background processes relevant to the

1840 analysis, with the exception of the multijet background which is modelled using a
1841 data-driven technique. Data and simulated events are reconstructed using the same
1842 algorithms, and a reweighting is applied to the simulated events in order to match
1843 the pile-up distribution observed in the data.

1844 **7.1.2 Object Reconstruction**

1845 The presence of neutrinos in the $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \ell\ell b\bar{b}$ signatures can be
1846 inferred from a momentum imbalance in the transverse plane Section 3.4.5. The
1847 vector boson transverse momentum p_T^V is reconstructed as the missing transverse
1848 energy E_T^{miss} in the 0-lepton channel, as the magnitude of the summed $\mathbf{E}_T^{\text{miss}}$ and
1849 charged-lepton momentum in the 1-lepton channel, and as the transverse momentum
1850 of the 2-lepton system in the 2-lepton channel (see Section 3.4.5).

1851 Leptons are used for the channel classification and to select relevant events as outlined
1852 in Section 7.1.3. Electrons and muons are reconstructed as outlined in Section 3.4.4.
1853 Electron identification follows the approach outlined in Ref. [103]. In addition to the
1854 likelihood-based method described in Section 3.4.4, *baseline* electrons are required to
1855 satisfy $p_T > 7 \text{ GeV}$, $|\eta| < 2.47$, $s(d_0) < 5$, and $|z_0 \sin(\theta)| < 0.5 \text{ mm}$. *Signal* electron
1856 additionally are required to satisfy a tighter likelihood identification selection. Muons
1857 are required to satisfy $p_T > 7 \text{ GeV}$, $|\eta| < 2.7$, $s(d_0) < 3$, and $|z_0 \sin(\theta)| < 0.5 \text{ mm}$.
1858 *Baseline* muons are required to pass the ‘loose’ identification described in Ref. [73],
1859 while *signal* muons are required to pass the ‘medium’ identification working point.
1860 All signal leptons are required to additionally satisfy a $p_T > 27 \text{ GeV}$ selection criteria,
1861 except for muons in the 1-lepton channel where a cut of 25 GeV is used. The number
1862 of baseline leptons is used to categorise the event into the 0-, 1- or 2-lepton channels.
1863 The 1- and 2-lepton channels additionally require one signal lepton to be present.

1864 The track-jets matched to the Higgs candidate are b -tagged using the MV2c10 b -
1865 tagging algorithm [59, 108, 158]. MV2c10 is a machine learning algorithm using a
1866 Boosted Decision Tree (BDT) which is tuned to achieve an average b -jet efficiency of
1867 70% on simulated $t\bar{t}$ events. At this efficiency working point, rejection factors for
1868 c -jets and light-jets are approximately 9 and 304 respectively. The MV2 algorithm
1869 takes inputs from the outputs of a number of low-level algorithms (IPxD, SV1 and
1870 JetFitter). The outputs of the low-level algorithms are provided as inputs to the
1871 boosted decision tree. The efficiency of the tagging algorithm is calibrated to events

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%)					
$qg \rightarrow W H \rightarrow \ell\nu b\bar{b}$	Powheg-Box v2 [124] + GoSAM [126] + MiNLO [127, 128]	NNPDF3.0NLO (*) [93]	PyTHIA 8.212 [95]	AZNLO	NNLO(QCD) + NLO(EW) [129–135]
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	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 [136–140]
Top quark ($m_t = 172.5$ GeV)					
$t\bar{t}$	Powheg-Box v2 [124, 141]	NNPDF3.0NLO	PyTHIA 8.230	A14 [96]	NNLO+NNLL [142]
s-channel	Powheg-Box v2 [124, 143]	NNPDF3.0NLO	PyTHIA 8.230	A14	NLO [144]
t-channel	Powheg-Box v2 [124, 143]	NNPDF3.0NLO	PyTHIA 8.230	A14	NLO [145]
Wt	Powheg-Box v2 [124, 146]	NNPDF3.0NLO	PyTHIA 8.230	A14	Approximate NNLO [147]
Vector boson + jets					
$W \rightarrow \ell\nu$	SHERPA 2.2.1 [148–151]	NNPDF3.0NNLO	SHERPA 2.2.1 [152, 153]	Default	NNLO [154]
$Z/\gamma^* \rightarrow \ell\ell$	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. 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 [155] 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 [156, 157].

1872 in data [159–161]. The jet tagging strategy relies on extensive studies into track-jet
1873 b -tagging in boosted topologies [162, 163].

1874 The jet flavour labelling scheme is described in Section 3.4.3.

1875 7.1.3 Selection Criteria

1876 An extensive list of selection cuts are applied to each event in order to reject
1877 background events whilst retaining as many signal events as possible. A full list of
1878 selection cuts applied to the different analysis regions is given in Table 7.3, while
1879 some key selections are listed below.

1880 All channels are require events with at least one large- R jet with $p_T > 250 \text{ GeV}$
1881 and $|\eta| < 2.0$. The vector boson transverse momentum is also required to satisfy
1882 $p_T^V > 250 \text{ GeV}$. The Higgs candidate is chosen as the highest p_T large- R jet satisfying
1883 these requirements. As mentioned, the candidate large- R jet is required to have
1884 two ghost-assciated and b -tagged variable-radius track-jets. These track-jets are
1885 required to have at least two constituent tracks with $p_T > 500 \text{ MeV}$ and $|\eta| < 2.5$.
1886 The track-jets themselves must satisfy $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$.

1887 In the 0-lepton channel, trigger selections are applied using an E_T^{miss} trigger with a
1888 luminosity-dependent threshold between 70–110 GeV. In the 1-lepton electron sub-
1889 channel a combination of single electron triggers is used with minimum p_T thresholds
1890 between 24–26 GeV. In the muon sub-channel the same E_T^{miss} trigger as the 0-lepton
1891 channel is used. Since muons are not used for the E_T^{miss} trigger calculations, this
1892 is in effect a p_T requirement on the muon-neutrino system, which in the analysis
1893 phase space is more efficient than a single-muon trigger. The 2-lepton channel uses
1894 the same triggering strategy as the 1-lepton channel. In all channels, the trigger
1895 selections applied are fully efficient for events selected using the full requirements in
1896 Table 7.3.

1897 The combined selections in Table 7.3 result in a signal efficiency ranging from 6–16%
1898 for the WH and ZH processes depending on the channel and p_T^V bin.

Selection	0 lepton channel	1 lepton channel	2 leptons channel
Trigger	E_T^{miss}	e sub-channel Single electron	μ sub-channel E_T^{miss}
Leptons	0 <i>baseline</i> leptons	$p_T > 27 \text{ GeV}$ 1 <i>signal</i> lepton no second <i>baseline</i> lepton	2 <i>baseline</i> leptons among which $p_T > 25 \text{ GeV}$ both leptons of the same flavour - opposite sign muons
E_T^{miss}	$> 250 \text{ GeV}$	$> 50 \text{ GeV}$	-
p_T^V			$p_T^V > 250 \text{ GeV}$
Large- R jets		at least one large- R jet, $p_T > 250 \text{ GeV}, \eta < 2.0$	
Track-jets		at least two track-jets, $p_T > 10 \text{ GeV}, \eta < 2.5$, matched to the leading large- R jet	
b -tagged jets		leading two track-jets matched to the leading large- R must be b -tagged (MV2c10, 70%)	
m_1			$> 50 \text{ GeV}$
$\min[\Delta\phi(E_T^{\text{miss}}, \text{small-}R \text{ jets})]$	$> 30^\circ$		-
$\Delta\phi(E_T^{\text{miss}}, H_{\text{cand}})$	$> 120^\circ$		-
$\Delta\phi(E_T^{\text{miss}}, E_{T, \text{trk}}^{\text{miss}})$	$< 90^\circ$		-
$\Delta y(V, H_{\text{cand}})$	-		$ \Delta y(V, H_{\text{cand}}) < 1.4$
$m_{\ell\ell}$	-		$66 \text{ GeV} < m_{\ell\ell} < 116 \text{ GeV}$
Lepton p_T imbalance	-		$(p_T^{\ell_1} - p_T^{\ell_2})/p_T^Z < 0.8$

Table 7.3: Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels and sub-channels. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region.

¹⁸⁹⁹ 7.1.4 Control Regions

¹⁹⁰⁰ The $t\bar{t}$ process presents a major background in the 0- and 1-lepton channels. In these
¹⁹⁰¹ events, the Higgs candidate is often reconstructed from a correctly tagged b -jet from
¹⁹⁰² the top decay $t \rightarrow Wb$, and an incorrectly tagged c - or light-jet from the subsequent
¹⁹⁰³ decay of the W , as shown in Fig. 7.1.

not sure where the 0L ttbar ETmiss comes from

¹⁹⁰⁴ The only known decay mode of the top quark is via the weak force to a W and
¹⁹⁰⁵ a down-type quark. (it is the only quark heavy enough to decay into an on-shell
¹⁹⁰⁶ W). Overwhelmingly (96% of the time) the down-type quark is a b -quark Hence, the
¹⁹⁰⁷ second top quark is also likely to result in a second tagged b -tagged track-jet outside
¹⁹⁰⁸ of the large- R Higgs candidate. To ensure sufficient $t\bar{t}$ rejections, 0- and 1-lepton
¹⁹⁰⁹ channel signal regions are defined using a veto on events with b -tagged track-jets
¹⁹¹⁰ outside the Higgs-jet candidate. These events are used to construct a control region
¹⁹¹¹ (CR) which is enriched in $t\bar{t}$ events. The CR is used to constrain the normalisation
¹⁹¹² of the $t\bar{t}$ background in the fit.

¹⁹¹³ 7.1.5 Background Composition

¹⁹¹⁴ After the selections described in Section 7.1.3 the number of background events
¹⁹¹⁵ mimicking the VH , $H \rightarrow b\bar{b}$ signal is greatly reduced. However, the number of
¹⁹¹⁶ background events still greatly outnumbers that of signal events. The background
¹⁹¹⁷ processes are channel dependent. In the 0-lepton channel the dominant sources of
¹⁹¹⁸ backgrounds are $Z+jets$ ($Z \rightarrow \nu\nu$) and $t\bar{t}$, with $W+jets$ and diboson events being
¹⁹¹⁹ subdominant. In the event of $W \rightarrow \tau\nu$, and subsequent hadronic decay of the τ or
¹⁹²⁰ lack of successful reconstruction/selection of the leptonic decay products, $W+jets$
¹⁹²¹ can also contribute to the 0-lepton channel. $t\bar{t}$ and $W+jets$ (with a leptonic decay
¹⁹²² of the W as in $W \rightarrow \ell\nu$) are dominant in the 1-lepton channel, while single-top is
¹⁹²³ subdominant. In the 2-lepton channel, $Z+jets$ ($Z \rightarrow \ell\ell$) is again dominant followed
¹⁹²⁴ by $Z Z$ diboson events.

¹⁹²⁵ The diboson background VV consists primarily of WZ and ZZ events in which the
¹⁹²⁶ Z decays to a pair of b -quarks. This process very closely matches the signal, with
¹⁹²⁷ a resonant peak occurring at $m_Z = 91$ GeV and so is considered as an irreducible
¹⁹²⁸ background ($V+b$ -jets is also irreducible).

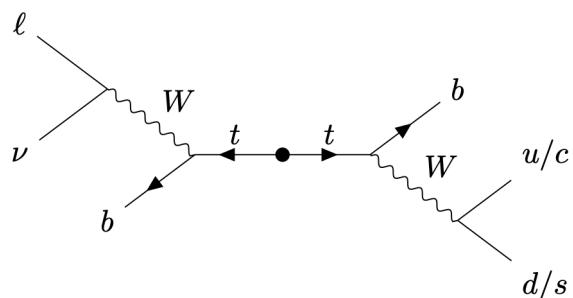
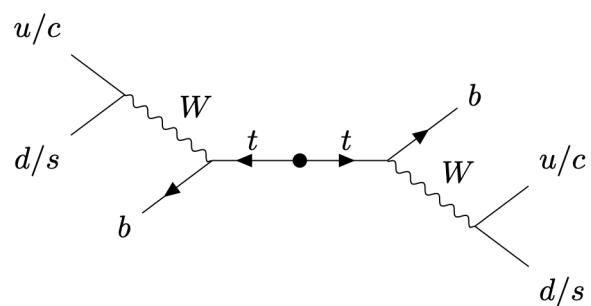
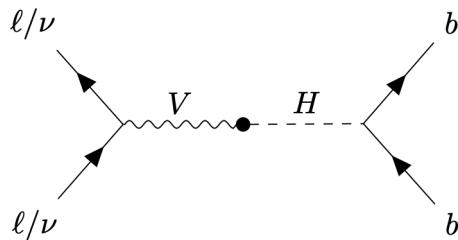


Figure 7.1: Diagrams of the signal process (top) and the 0-lepton and 1-lepton $t\bar{t}$ backgrounds (middle, bottom). Objects to the right of centre are reconstructed within the large- R jet. For the backgrounds, the large- R jet contains a mis-tagged c - or light-jet.

¹⁹²⁹ The $t\bar{t}V$, $t\bar{t}H$ and multijet backgrounds are negligible in the analysis phase space
¹⁹³⁰ after the selections have been applied, with the exception of the 1-lepton electron
¹⁹³¹ sub-channel, in which multijet background is not ignored. The multijet background
¹⁹³² is made up of jets with semileptonic heavy-flavour-hadron decays (e.g. $b \rightarrow c\ell\nu$) and
¹⁹³³ jets which are mis-tagged by the flavour tagging algorithm MV2c10.

¹⁹³⁴ The contributions from the different backgrounds are modelled using Monte Carlo
¹⁹³⁵ event generators and the impacts on the analysis are studied in Section 7.2. The
¹⁹³⁶ multijet background is not modelled but instead estimated using a data-driven
¹⁹³⁷ technique.

¹⁹³⁸ 7.2 Systematic Uncertainties & Background ¹⁹³⁹ Modelling

¹⁹⁴⁰ Systemic uncertainties are extensively employed to give the fit model described in
¹⁹⁴¹ Section 7.3 enough flexibility to account for inaccuracies in the various inputs to the fit.
¹⁹⁴² Two main types of systematic uncertainty are considered: experimental and modelling.
¹⁹⁴³ Experimental uncertainties arise due to the limited due to limited detector precision,
¹⁹⁴⁴ imperfect reconstruction algorithms (in particular the b -tagging algorithms), and
¹⁹⁴⁵ due to the imperfect measurement of pile-up and integrated luminosity. Modelling
¹⁹⁴⁶ is the simulation processes relevant to the analysis using Monte-Carlo (MC) event
¹⁹⁴⁷ generators, and is used to predict the outcome of the analysis. Modelling uncertainties
¹⁹⁴⁸ arise due to the imperfections in the simulation of signal and background events, for
¹⁹⁴⁹ example differences between event generators, or use of different model parameters
¹⁹⁵⁰ when producing simulated events. In order to observe a certain process, for example
¹⁹⁵¹ VH , $H \rightarrow b\bar{b}$, an increase in the number of observed events with respect to the
¹⁹⁵² background-only hypothesis is looked for. The excess is often relatively small against
¹⁹⁵³ the total number of background events, and hence accurate modelling of the expected
¹⁹⁵⁴ number of background and signal events is crucial for successfully performing the
¹⁹⁵⁵ analysis. Particular care is paid to the uncertainties on the modelling predictions as
¹⁹⁵⁶ discussed in this section.

¹⁹⁵⁷ Modelling uncertainties are described in detail in the following sections. Modelling
¹⁹⁵⁸ uncertainties: *Nominal* samples are used as a reference to which different vari-

ations can be compared. The nominal samples are chosen as the best possible representation of the underlying physical process. *Alternative* samples are used to understand inaccuracies that may be present in the nominal samples. Some aspect of the nominal model is varied, and the discrepancy with respect to the nominal model is quantified. The discrepancy is used to systematic uncertainty associated with the model parameter which was changed.

Modelling studies involving c - and light-jets is hampered by the low available statistics of jets pass the analysis selections, due to the high rejection rates of the b -tagging algorithm MV2c10. For modelling studies therefore, truth tagging (TT) is employed to ensure sufficient numbers of jets are available to calculate uncertainties. TT works by computing a 2-dimensional efficiency map using the jet p_T and jet η . The two leading track-jets associated to the large- R jet automatically passes the b -tagging requirement, and are weighted based on their p_T and η using the pre-calculated efficiency map.

7.2.1 Sources of Systematic Uncertainties

This section briefly describes the different sources of uncertainty in the predictive model used in the analysis, and how each source of uncertainty is implemented within the analysis framework. Considered sources of systematic uncertainty are listed in Table 7.4. For each source of uncertainty, acceptance and shape uncertainties are derived.

QCD Scales

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

1988 PDF Sets

1989 Parton distribution functions (PDFs) specify the probability of finding a parton
1990 with a given momentum inside a hadron (in this case, inside colliding protons).
1991 PDFs have to be derived from data and are a significant source of uncertainty in
1992 analyses of hadronic collision data. There are three sources of PDF uncertainties:
1993 the statistical and systematic errors on the underlying data used to derive the PDFs,
1994 the theory which is used to describe them (which is based on some fixed order
1995 perturbative QCD expansion), and finally the procedure which is used to extract the
1996 PDFs from the data. PDF-related uncertainties were derived following Ref. [155].
1997 This involves considering 100 PDF replicas which, when combined, form a central
1998 value and associated uncertainty, and also in parallel direct changes to the central
1999 values of PDFs using the MMHT2014 [164] and CT14NLO [165] PDF sets.

2000 Event Generator

2001 The choice of parton shower (PS) and underlying event (UE) generators can affect
2002 the analysis outcome. Changing these models modifies several aspects of the event
2003 generation at the same time, such as the accuracy of matrix element predictions and
2004 different approaches to parton showering. This change tends to lead to the largest
2005 discrepancy with respect to the nominal samples.

2006 Resummation and Merging Scales

2007 Resummation is a technique used in QCD to help cope with calculations involving
2008 disparate energy scales, and involves the introduction of an associated resummation
2009 scale, the choice of which introduces some systematic uncertainty into the model.
2010 Parton showering models are accurate when simulating low- p_T radiation, however
2011 inaccuracies start to arrive when simulating hard emissions. To combat this, par-
2012 ton showering models utilise more precise matrix element calculations above some
2013 momentum threshold. The choice of threshold, or *merging scale* introduces some
2014 uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale
2015 variations are available for a subset of the SHERPA samples. The number of available
2016 events is significantly lower than the number of events in the nominal sample, and no

- 2017 statistically significant discrepancy with respect to the nominal samples is observed.
2018 The corresponding uncertainties and therefore neglected.

2019 7.2.2 Implementation of Variations

- 2020 Modelling variations are implemented in different ways, depending on the associated
2021 uncertainty. Table 7.4 lists the different sources of uncertainty described in Section 7.2.1
and for each lists the implementation. As production of high-stastic MC samples

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 7.4: 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.

- 2022
- 2023 is computationally expensive, a technique in state of the art simulation packages
2024 is to store some sources of variation as internal weights, which can be generated
2025 alongside the nominal samples, saving computation time. The nominal sample then
2026 effectively contains information about an ensemble of different samples, corresponding
2027 to different model parameters, which are accessible via reweightings. When filling
2028 histograms for the variations, bins are incremented by the internal weight of the
2029 event associated with the variation in question.

- 2030 While the inclusion of internal weight variation in MC event generators has decreased
2031 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently
2032 some sources of systematic uncertainty that are unable to be stored as internal weight
2033 variations due to technical limitations. Two examples are the choice of resummation
2034 and merging scales. A method to parameterise the systematic variation using
2035 one sample, and to then apply this parameterisation to another sample, has been

2036 developed by ATLAS [166]. This method was used to derive resummation and
2037 merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous (lower
2038 statistic) SHERPA 2.1 alternative sample. The resulting uncertainties were studied
2039 and found to be negligible in comparison with systematics from other sources.

2040 **7.2.3 Vector Boson + Jets Modelling**

2041 After event selection, the $V+jets$ background is a dominant background in all three
2042 analysis channels as described in Section 7.1.5. The $V+jets$ samples are split into
2043 categories depending on the truth flavour of the track-jets which are ghost-associated
2044 to the large- R jet Higgs candidate. The categories are $V+bb$, $V+bc$, $V+bl$, $V+cc$,
2045 $V+cl$, $V+ll$, and $V+hf$ refers collectively to the categories containing at least one
2046 b - or c -jet. $V+bb$ is dominant generally accounting for 80% of the jets, while $V+hf$
2047 accounts for around 90% of jets. The full flavour composition breakdown for each
2048 channel and analysis region are given in Tables 7.5, 7.7 and 7.8.

2049 In order to access uncertainties associated with the use of MC generators, variations
2050 of the data are produced using alternative generators or variation of nominal
2051 generator parameters as described in Section 7.2.2. As described in Section 7.1.1,
2052 the nominal MC event generator used for $V+jets$ events is SHERPA 2.2.1, while
2053 MADGRAPH5_AMC@NLO+PYTHIA8 (which uses a different parton showering
2054 model) is used as an alternative generator.

2055 Modelling systematics can have several impacts, including affecting the overall
2056 normalisation for different processes, and the relative acceptances between different
2057 analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR,
2058 and between p_T^V bins), and the shapes of the m_J distributions. Since the fit model
2059 fits only the large- R jet mass m_J to data, all shape uncertainties are estimated with
2060 respect to this observable. Several sources of uncertainty, summarised in Section 7.2.1,
2061 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 7.5: 0-lepton $W+jets$ nominal sample flavour composition and total event yield.

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 7.6: 1-lepton $W+jets$ nominal sample flavour composition and total event yield.

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 7.7: 0-lepton $Z+jets$ nominal sample flavour composition and total event yield.

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 7.8: 2-lepton $Z+jets$ nominal sample flavour composition and total event yield.

2062 **Acceptance Uncertainties**

2063 Several different types of acceptance uncertainties have been calculated and imple-
2064 mented as nuisance parameters in the fit. These account for uncertainty in the overall
2065 number of events in each channel, and for the migration of events between different
2066 analysis regions. The acceptance uncertainties relevant to the $V+jets$ processes are
2067 summarised below.

- 2068 • **Overall normalisation:** only relevant where normalisation cannot be left
2069 floating (determined as part of the fit). The $V+hf$ component is left floating in
2070 the fit. For other components, independent normalisations used for $W+hf$ and
2071 $Z+hf$. The contributions are mainly determined by the 1-lepton (for $W+hf$)
2072 and 2-lepton (for $Z+hf$) SRs respectively and then extrapolated to 0-lepton
2073 channel.
- 2074 • **SR-to-CR relative acceptance:** the uncertainty on the normalisation of the
2075 signal region due to events migrating between the signal and control regions.
- 2076 • **HP-to-LP relative acceptance:** the uncertainty on the normalisation of the
2077 high-purity (HP) signal region due to events migrating between the high- and
2078 low-purity signal regions.
- 2079 • **Medium-to-high p_T^V relative acceptance:** describes any shape effect in p_T^V
2080 distribution, given that the analysis only uses two p_T^V bins (medium and high).
- 2081 • **Flavour relative acceptance:** for each flavour $V+xx$, where $xx \in \{bc, bl, cc\}$
2082 the ratio of $V+xx/V+bb$ events is calculated. This corresponds to the uncer-
2083 tainty of Vbb events due to the miss-tagging of other flavours Vxx .
- 2084 • **Channel relative acceptance:** corresponding to the uncertainty in the nor-
2085 malisation of $V+jets$ events events due to the migration of events between
2086 channels.

2087 The uncertainties arising from the different sources described in Section 7.2.1 are
2088 summed in quadrature to give a total uncertainty on each region. A summary of the
2089 different acceptance uncertainties that were derived in this way and subsequently
2090 applied in the fit are given in Table 7.9. An effort has been made, wherever possible,
2091 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/CR	90% [†]	40% [†]	40%	-
HP/LP	18%		18%	-
High/Medium p_T^V	30%	10%*	10%	
Channel Extrap.	20%	-	16%	-
Vbc/Vbb	30%			
Vbl/Vbb	30%			
Vcc/Vbb	20%			
Vcl Norm.	30%			
VL Norm.	30%			

Table 7.9: V+jets acceptance uncertainties. W+jets SR and CR uncertainties marked with a superscript † are correlated. The 1L W+jets H/M uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions. The 0L W+jets Wbb Norm uncertainty is only applied when a floating normalisation for Wbb cannot be obtained from the 1L channel. A 30% uncertainty for $Z \rightarrow b\bar{b}$ norm is applied in the 1L channel when a floating normalisation for $Z \rightarrow b\bar{b}$ cannot be obtained from the 0L or 2L channels.

2092 **Shape Uncertainties**

2093 In order to derive shape uncertainties (which as the name suggests affect shapes but
2094 not overall normalisations of distributions), the following procedure is carried out.
2095 Normalised distributions of the reconstructed large- R Higgs candidate jet mass m_J
2096 are compared for the nominal sample and variations. For each variation, the ratio of
2097 the variation to nominal is calculated, the up and down variations are symmetrised,
2098 and an analytic function is fit to the symmetrised ratio. If different analysis regions
2099 or channels show the same pattern of variation, a common uncertainty is assigned.
2100 An example of a significant source of uncertainty, arising from choice of factorisation
2101 scale μ_R is shown in Fig. 7.2. HP SRs split into medium and high p_T^V bins are shown
2102 for the 0-lepton channel for $W+\text{hf}$ and $Z+\text{hf}$ jets. The 0- and 1-lepton channels for
2103 the $W+\text{hf}$ contribution and the 0- and 2-lepton channels for the $Z+\text{jets}$ contribution
2104 are merged, since the shapes in m_J are consistent across channels. An exponential
2105 function $e^{p_0 + p_1 x} + p_2$ has been fitted to the ratio of the normalised distributions.
2106 The magnitude of the variation does show p_T^V dependence, and so two separate
2107 uncertainties are added in the fit, and applied individually in each p_T^V region.

2108 The shape uncertainties for μ_R were derived on the SRs but are also applied to the
2109 CRs, as the low statistics in the CRs make it difficult to derive dedicated shape
2110 uncertainties. All the shape uncertainties are fully correlated accross regions.

2111 A comparison of the m_J shapes between SHERPA and MADGRAPH is shown in
2112 Fig. 7.3. The plots are split by process and channel, but merged in SR purity and p_T^V
2113 bins reflecting similarities between the m_J shapes across these regions. Due to the low
2114 statistics available for the alternate MADGRAPH sample, and the lack of statistically
2115 significant variation between the samples, no associated shape uncertainty is added
2116 to the fit in this case.

2117 The impacts of variations in the factorisation scale μ_F and the choice of PDF set on
2118 m_J shape were also found to be negligible in comparison with μ_R and are hence not
2119 associated uncertainty was added to the fit.

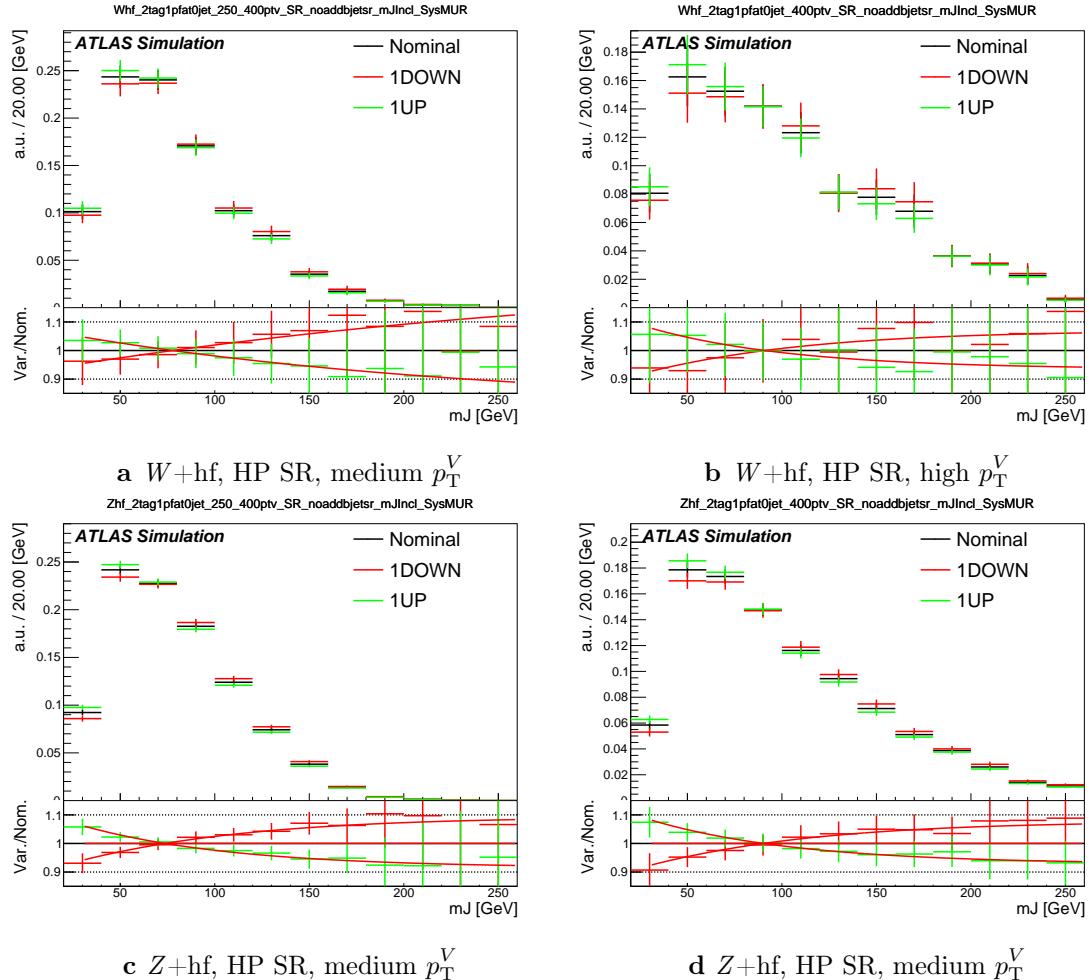


Figure 7.2: Normalised leading large- R jet mass distribution from Z and $W + hf$ processes in the HP SR of the 0-lepton channel. The renormalisation scale μ_r has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential function is fitted to the ratio between the nominal and variation samples.

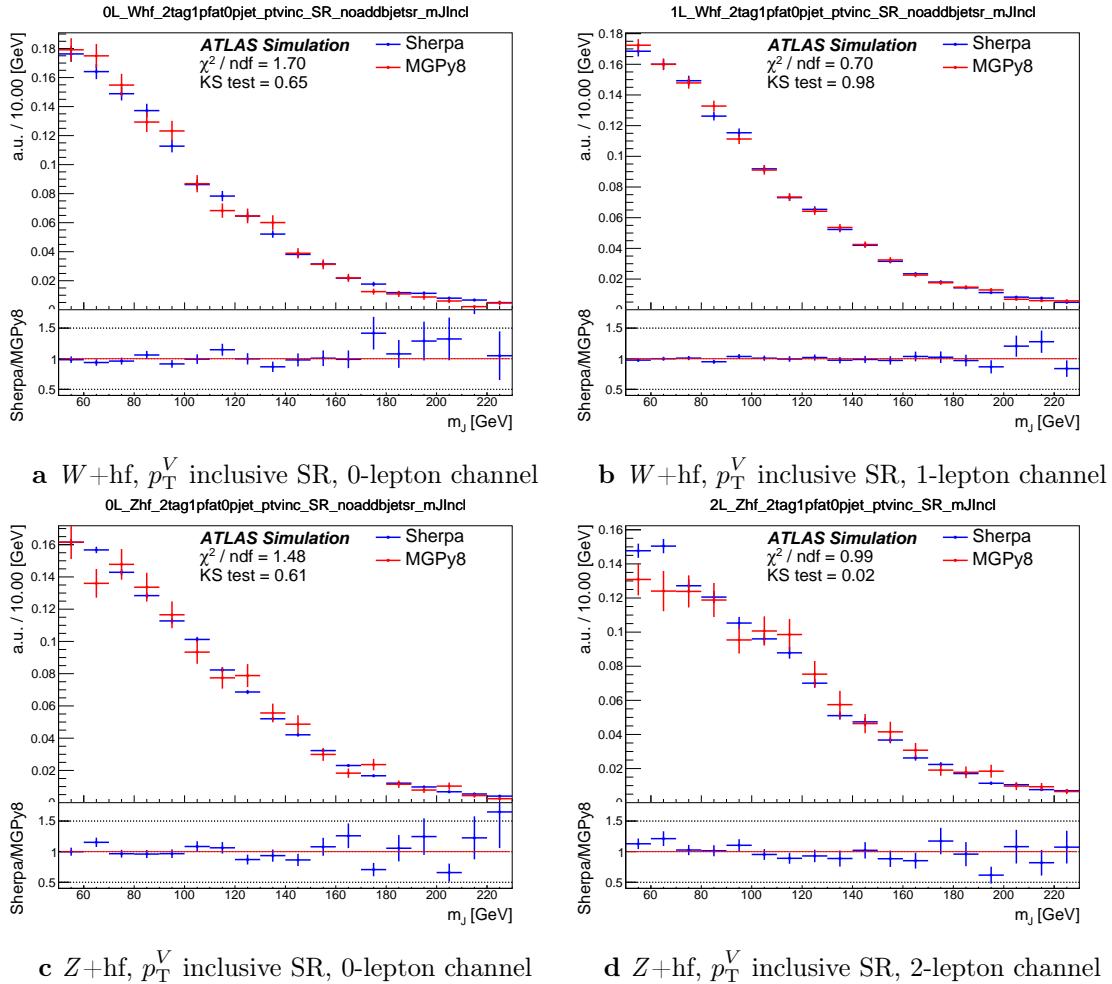


Figure 7.3: The comparison on m_J shapes between SHERPA and MADGRAPH samples from $V + \text{hf}$ process in p_T^V inclusive signal regions. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.

2120 7.2.4 Diboson Modelling

2121 The uncertainties for the diboson background generally follows that of $V+jets$. How-
2122 ever an alternative sample was generated using POWHEG interfaced with PYTHIA8,
2123 using the AZNLO shower tune with the CTEQ6L1 PDFs [167]. Unlike SHERPA,
2124 POWHEG models the off-shell Z contribution at NLO.

2125 Acceptance and shape uncertainties are derived in an analogous fashion to $V+jets$.

2126 Acceptance Uncertainties

2127 Diboson acceptance uncertainties are summarised in Table 7.10. Variations from μ_R ,
2128 μ_F , PDF choice and alternative generator are considered and are combined combined
2129 through a sum in quadrature as described in Section 7.2.3. The largest modification
2130 to the nominal acceptance results from the POWHEG+PYTHIA8 alternate sample,
2131 which modifies several parts of the generative model at the same time. Since the
2132 diboson contribution to the $t\bar{t}$ control region is small, no SR-to-CR relative acceptance
2133 uncertainty is required.

2134 For the WZ contribution, uncertainties are derived using the 1-lepton channel and
2135 applied to all three channels. An additional 8% channel migration uncertainty
2136 is applied on the 0-lepton channel. For the ZZ contribution, the normalisation
2137 uncertainty is calculated using the 2-lepton channel and applied to all three channels.
2138 The 0- and 1-lepton channels have a similar HP-to-LP relative acceptance uncertainty
2139 of 18%. The 1-lepton medium-to-high p_T^V relative acceptance is based off the value
2140 obtained from the 2-lepton channel. 30% and 18% channel migration uncertainties
2141 are applied in the 0- and 1-lepton channels respectively.

2142 Since the contribution from WW is small, dedicated studies are not performed, but
2143 a 25% normalisation uncertainty is applied in all the three channels which is based
2144 on the modelling studies performed for the previous analysis [103].

2145 Shape Uncertainties

2146 Diboson shape uncertainties are derived in a similar fashion to $V+jets$. Only the
2147 uncertainties associated with systematic variation of μ_R and the event generator

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 7.10: Diboson acceptance uncertainties. All uncertainties except channel extrapolation uncertainties are fully correlated between ZZ and WZ processes and channels.

have a non-negligible impact on the m_J shape. Variation of μ_R produces consistent m_J shape impacts across all regions and channels, and hence only a single associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted to the symmetrised ratio.

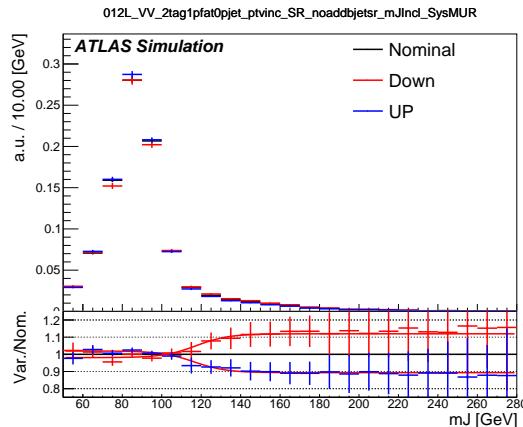


Figure 7.4: Normalised leading large- R jet mass distribution from WZ and ZZ process, merged among all the signal regions and lepton channels. The renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.

The comparison between the nominal SHERPA and alternative POWHEG+PYTHIA samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both WZ and ZZ processes. For these channels, the shape of m_J varies in opposite directions in the LP and HP signal regions. Shapes are similar between p_T^V bins, the 0- and 1-lepton channels and for WZ and ZZ . A third order polynomial is fitted to the ratio, and this function transitions to a constant piecewise function in the high

mass region to accurately represent the shape taking into account large statistical uncertainties. Dependence on event generator was found to be negligible within statistical uncertainty in the 2-lepton channel. All diboson shape uncertainties are fully correlated in the fit.

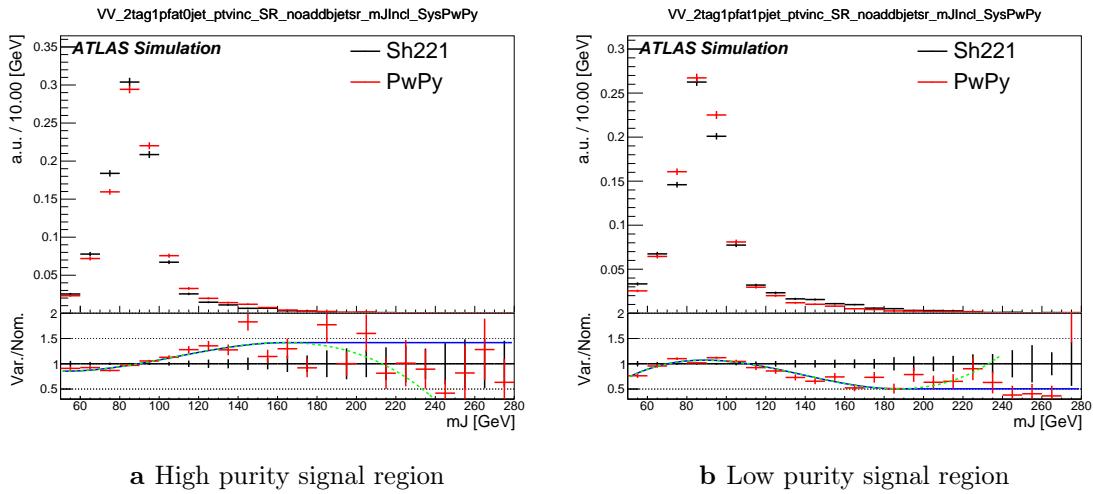


Figure 7.5: The comparison on m_J shapes between SHERPA and POWHEG+PYTHIA 8 samples from WZ and ZZ process in high and low purity signal regions. p_T^V regions and 0- and 1-lepton channels are merged. The dashed green line shows the fitted third order polynomial function and the blue lines show the function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

7.3 Statistical Treatment

Selected events are used to perform a statistical test of the background-only hypothesis, i.e. a model which does not include the VH , $H \rightarrow b\bar{b}$ process. The test involves a binned global maximum-profile-likelihood fit from the model to the data using the m_J distribution, and combines all the analysis regions defined in Table 7.1. The test is based on the profile likelihood ratio test statistic. The signal strength $\mu = \sigma/\sigma_{SM}$ is defined as the ratio between the observed and predicted cross-sections, where $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ corresponds to the SM prediction. It is a parameter of interest (POI) which acts to scale the total number of signal events.

The present analysis makes use of two POIs. The first, μ_{VH}^{bb} , is the signal strength for the VH , $H \rightarrow b\bar{b}$ process, the primary process under investigation. The diboson

2174 production strength μ_{VZ}^{bb} for the VZ , $Z \rightarrow b\bar{b}$ process is measured simultaneously
2175 and provides a validation of the analysis apparatus used for the primary $H \rightarrow b\bar{b}$
2176 measurement. Alongside the two POIs, the predictive model depends on several
2177 uninteresting parameters which are not the primary target of measurement. These
2178 parameters are called nuisance parameters (NPs), collectively referred to as θ . Freely
2179 floating background normalisations are implemented as NPs and are also extracted
2180 during the fitting processes.

2181 7.3.1 Likelihood Function

2182 The statistical setup treats each bin as a Poisson counting experiment and is based on
2183 the ROOSTATS framework [168]. The combined likelihood over N bins is constructed
2184 as the product of Poisson probabilities in each bin. Considering the simplified case
2185 of a single signal strength parameter μ , and neglecting sources of systematic or
2186 statistical uncertainty, this is given by

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

2187 where s_i (b_i) is the expected number of signal (background) events in bin i , and n_i is
2188 the number of observed data events in bin i .

2189 Treatment of Uncertainties

2190 Systematic uncertainties can modify the predicted signal and background yields s_i
2191 and b_i . Each source of systematic uncertainty is taken into account by adding an
2192 additional NP θ_j to the likelihood in the form of a Gaussian cost function. The
2193 combined effect of the NPs is then

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

2194 where N_θ is the number of NPs, $\bar{\theta}_j$ is the nominal value of the j th NP, θ_j is the fitted
2195 value, and σ_j is the corresponding associated prior uncertainty on θ_j . As the fitted
2196 value of the θ_j deviates from its nominal value, a cost is introduced. The presence

2197 of NPs modifies the likelihood as

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

2198 The predicted signal and background yields are also modified by the presence of the
2199 NPs with

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

2200 For NPs which are left freely floating in the fit, no corresponding Gaussian constraint
2201 is added to the likelihood.

2202 Statistical uncertainty is also present, and implemented using a dedicated NP for
2203 each bin which can scale the background yield in that bin. Statistical NPs are also
2204 implemented using a Gaussian constraint.

2205 Smoothing and Pruning

2206 Systematic uncertainties are smoothed and pruned in the fit. Smoothing accounts
2207 for the large statistical uncertainty present in some bins that are subject to large
2208 fluctuations. The smoothing procedure relies on the assumption that the impact of
2209 systematics should be approximately monotonic and correlated between neighbouring
2210 bins.

2211 In addition to smoothing, pruning is the process of removing from the fit those
2212 systematics which only have a very small effect. This improves the stability of the
2213 fit by reducing the number of degrees of freedom. Acceptance uncertainties are
2214 pruned in a given region if they have a variation of less than 0.5%, or if the up and
2215 down variations have the same sign in that region. Shape uncertainties are pruned
2216 in a given region if the deviation in each bin is less than 0.5% in that region. In
2217 addition, acceptance and shape uncertainties are neglected in a given region for any
2218 background which makes up less than 2% of the total background in a given region.

2219 **Fit Procedure and Statistical Tests**

2220 The best-fit value of μ , denoted $\hat{\mu}$, is obtained via an unconditional maximisation
2221 of the likelihood. The likelihood is also used to construct a statistical test which can
2222 confirm or reject the background-only hypothesis. The test statistic q_μ is constructed
2223 from the profile likelihood ratio, as in

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

2224 where $\hat{\mu}$ and $\hat{\theta}$ are chosen to maximise the likelihood \mathcal{L} , and the profile value $\hat{\theta}_\mu$ is
2225 obtained from a conditional maximisation fo the likelihood for a specific choice of
2226 $\mu = 0$ corresponding to the background-only hypothesis.

2227 The test statistic is used to construct a p -value which is used to confirm or accept
2228 the background-only hypothesis. The p -value is typically reported in terms of the
2229 significance Z , defined as the number of standard deviations for a Gaussian Normal
2230 distribution which will produce a one-sided tail integral equal to the p -value, as in

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

2231 Typically a value of $Z = 3$ constitutes *evidence* of a processes, while $Z = 5$ is required
2232 for a *discovery*. Alongside the p -value, the best-fit value of the signal strength $\hat{\mu}$ and
2233 its corresponding uncertainty are typically quoted, and compared to their expected
2234 values (see Section 7.3.3).

2235 **7.3.2 Background Normalisations**

2236 The normalisation of the largest backgrounds are left floating and are determined in
2237 the fit. The corresponding postfit background normalisations are listed in Table 7.11.
2238 A single normalisation factor is used for $W+\text{hf}$ and $Z+\text{hf}$, which constitue more
2239 than 90% of the total $V+\text{jets}$ background, since the use of independent factors in
2240 different channels were found to be compatible.

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 7.11: 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. The errors represent the combined statistical and systematic uncertainties.

2241 The normalisations and shapes of all other backgrounds, with the exception of the
2242 multijet background which is estimated using a data driven technique, are initialised
2243 using the simulated samples.

2244 7.3.3 Asimov Dataset & Expected Results

2245 The Asimov dataset is constructed by replacing the data with the sum of the signal
2246 and background predictions $n_i = s_i + b_i$. A fit to this dataset using the nominal
2247 values of the NPs from the simulation will recover the input values and is useful for
2248 studying constraints on and correlations between the NPs.

2249 Alternatively, a conditional fit to the Asimov dataset can be performed using values
2250 of the background NPs which are determined from an unconditional fit to data. The
2251 signal NPs and POIs are fixed at their nominal values from the SM simulation. The
2252 result of this fit can be used to calculate expected (median) significances, which can
2253 be compared to their observed values.

2254 7.4 Results

2255 In the present analysis, the two signal strength parameters μ_{VH}^{bb} and μ_{VZ}^{bb} are extracted
2256 from a simultaneous maximisation of the likelihood described in Section 7.3. The
2257 results of the analysis are summarised in this section. Post-fit m_J distributions are
2258 shown in Section 7.4.1. The observed signal strengths are given in Section 7.4.2,
2259 along with observed and expected significances. Finally in Section 7.4.3 the impact
2260 of systematic uncertainties on the results is examined.

2261 7.4.1 Post-fit Distributions

2262 In addition to the observed significance and signal strength, it is also useful to study
 2263 the post-fit m_J distributions to compare the simulation and data using the best-fit
 2264 values $\hat{\mu}$ and $\hat{\theta}$. Post-fit m_J distributions are given for the signal regions in the 0-,
 2265 1- and 2-lepton channels in Fig. 7.6. The LP and HP regions are merged for the
 2266 0- and 1-lepton channels. The plots show large falling backgrounds, predominantly
 2267 made up of W -jets and Z -jets events, and a signal distribution corresponding to
 2268 the Standard Model Higgs boson peaking around $m_H = 125$ GeV. In general there is
 2269 a good level of agreement between the simulation and data, indicating the fit model
 2270 is performing as expected. Fig. 7.7 shows the post-fit plots for the $t\bar{t}$ control regions.
 2271 Again, a good level of agreement is observed given the statistical uncertainties on
 2272 the distributions.

2273 7.4.2 Observed Signal Strength & Significance

2274 The measured signal strength is computed as the ratio between the measured signal
 2275 yield to the prediction from the SM. The combined result for all three lepton channels
 2276 and all analysis regions is given for μ_{VH}^{bb} in Eq. (7.7), and for μ_{VZ}^{bb} is given in Eq. (7.8).
 2277 Both results include a full breakdown of the systematic and statistical uncertainties.

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

2278

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

2279 The results for μ_{VH}^{bb} and μ_{VZ}^{bb} agree with the expectation from the SM within their
 2280 combined uncertainty. The μ_{VH}^{bb} measurement is dominated by statistical uncertainty,
 2281 while the μ_{VZ}^{bb} measurement is dominated by systematic sources of uncertainty. These
 2282 measured signal strength for μ_{VZ}^{bb} corresponds to an observed significance of 2.1
 2283 standard deviations, with an expected (median) significance given the SM prediction
 2284 of 2.7 standard deviations obtained using the method described in Section 7.3.3. The
 2285 diboson observed (expected) signal strength significance is 5.4 (5.7). These results
 2286 are summarised in Fig. 7.8, which shows the background-subtracted m_J distribution.
 2287 A clear signal excess is visible around the Higgs mass of $m_H = 125$ GeV.

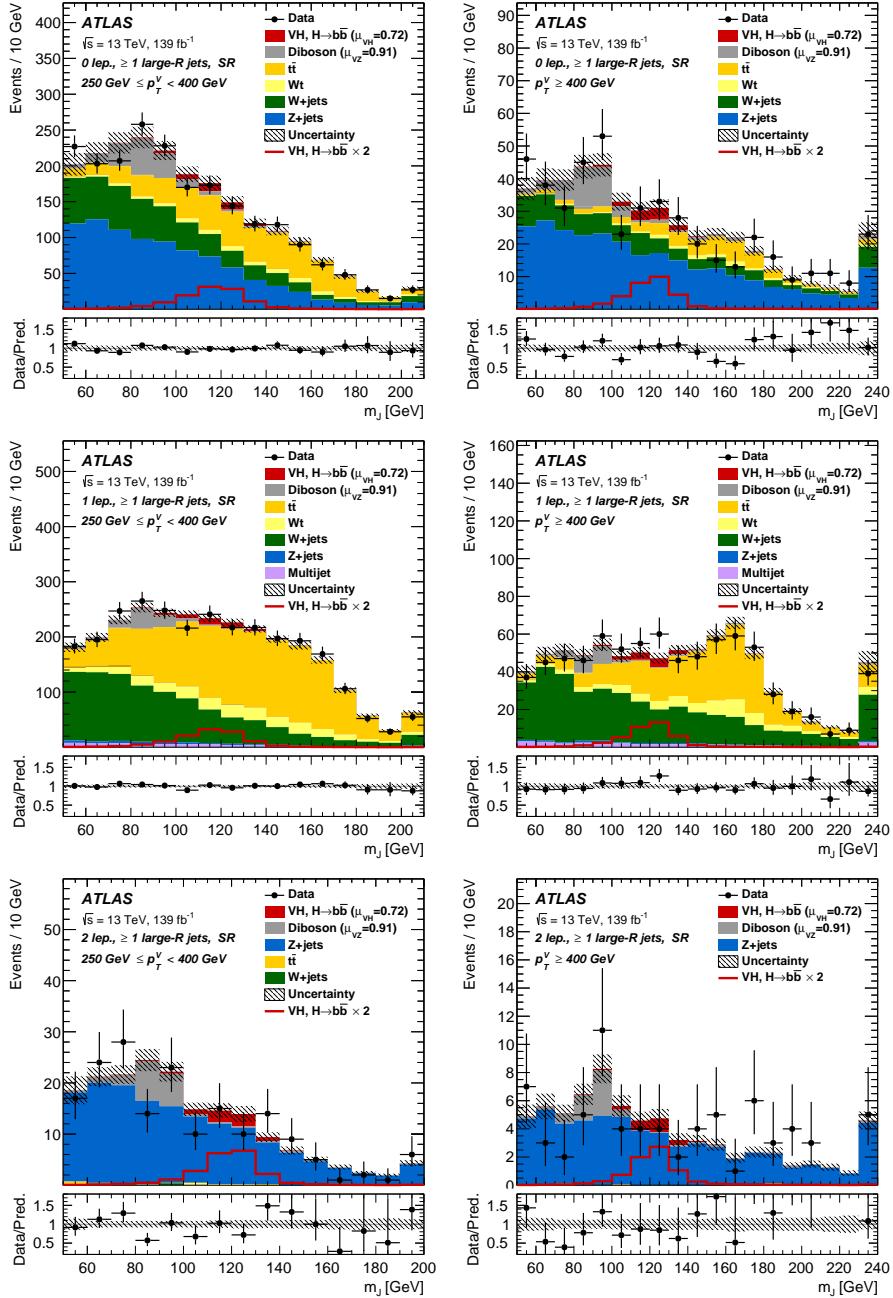


Figure 7.6: The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions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

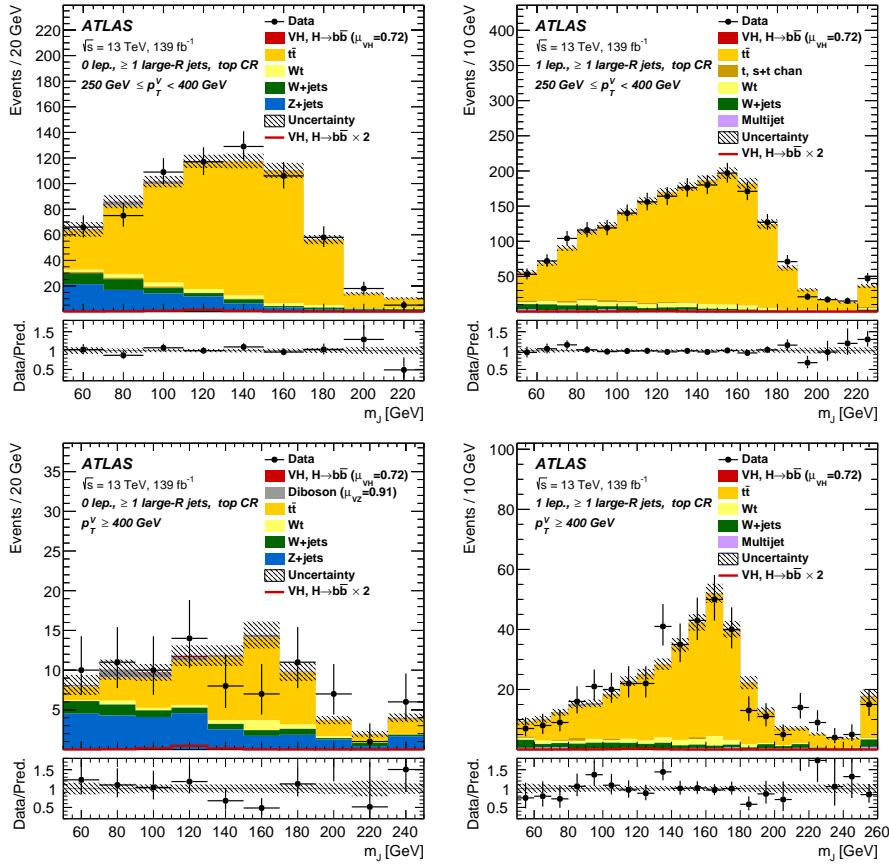


Figure 7.7: The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. The background contributions after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

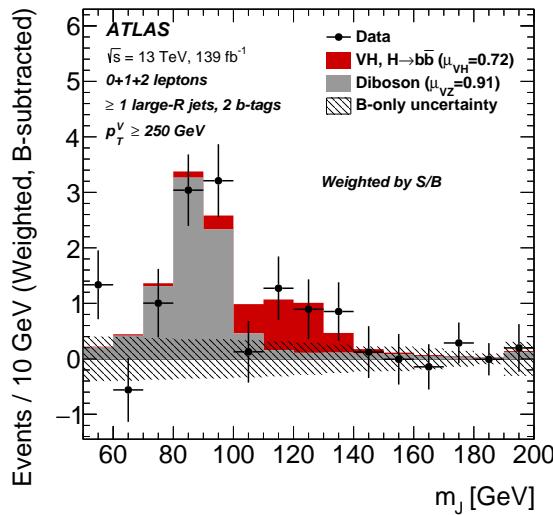


Figure 7.8: m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected 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.

2288 **Compatability Studies**

2289 Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting μ_{VH}^{bb}
2290 into three separate POIs, one for each channel. A simultaneous fit to the channel
2291 specific signal strengths can then be performned, which allows a comparison of the
2292 contributions from each channel. Fig. 7.9 compares the best-fit signal strengths.
2293 The 0- and 1-lepton channels show a signal strength which is consistent with the
2294 SM prediction, while the 2-lepton channel shows a small deviation within the 1σ
2295 uncertainty. Overall, good compatibility is observed via a χ^2 test with a corresponding
2296 p -value of 49%.

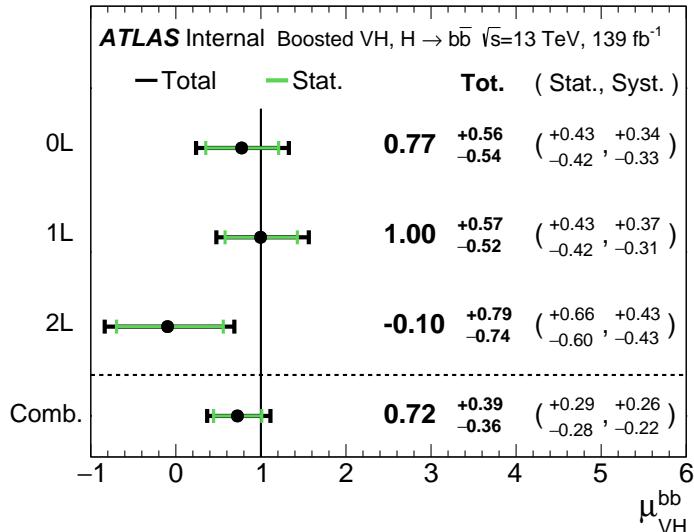


Figure 7.9: Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels splitted) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a χ^2 difference test and results in a p -value of 49%.

2297 **7.4.3 Impact of Systematics**

2298 The impact of systematic uncertainties on the final fitted value $\hat{\mu}_{VH}^{bb}$ is studied looking
2299 at the NP rankings, and the uncertainty breakdown.

2300 Fig. 7.10 shows the NP ranking, which is used to visualise which out of the many
2301 NPs involved in the fit have the largest impact on the sensitivity to the fitted POI.
2302 To obtain the ranking, a likelihood scan is performed for each NP θ_j . First, an

unconditional fit is used to determine $\hat{\theta}_j$. From this best-fit point, the NP is varied in steps and the likelihood is recomputed until the $\pm 1\sigma_{\hat{\theta}_j}$ values are reached. For each corresponding value of θ_j , the change in the best-fit value of the POI, $\Delta\hat{\mu}_{VH}^{bb}$ is calculated and used to rank the NPs. Also shown in Fig. 7.10 are the pulls and constraints for the highest ranked NPs.

The experimental uncertainty on the signal large- R jet mass resolution (JMR) has the largest impact of any NP. It is a significant contributor to the overall uncertainty on μ_{VH}^{bb} in Eq. (7.7). JMR and jet energy scale (JES) uncertainties also have impacts for the $V+jets$ background and for the diboson background. The freely-floating $Z+hf$ normalisation is the second highest ranked NP, and is heavily constrained by the fit. The VZ POI μ_{VZ}^{bb} is also a significant NP when considering the primary μ_{VH}^{bb} measurement.

The NP ranking highlights individual NPs which have a large impact on the POI measurement sensitivity. Complementary information is provided at a higher level by considering the overall impact of different groups of systematics. The groups are constructed from NPs which have similar physical origin. The impact on each group is calculated by running a fit with all the NPs in the given group fixed to their nominal values. The uncertainty on the POI extracted from this fit is subtracted in quadrature from the the uncertainty on the POI from the nominal fit, and the resulting values are provided as the impact for each group. The full breakdown for the observed impact of uncertainties on the μ_{VH}^{bb} signal strength is provided in Table 7.12. The total systematic impact is the difference in quadrature between the nominal uncertainty on μ_{VH}^{bb} and the combined statistical impact. The “data stat only” group fixes all NPs at their nominal value, while the total statistical impact fixes all NPs except floating normalisations. The floating normalisations group fixes only the NPs associated with normalisation which are left floating in the fit. The uncertainty on μ_{VH}^{bb} is dominated by combined statistcal effects (0.28), although the combined impact of systematics (0.24) is of a comparable size. The signal largest group is the data stat uncertainty (0.25), demonstrating that the analysis would benefit from an increased integrated luminosity or improved efficiency to select signal relevant events (recall from Section 7.1.3 the signal efficiency is in the range of 10%). Of the experimental systematic sources of uncertainty, the dominant impact is the experimental uncertainties associated with the reconstruction of large- R jets (0.13). Other experimental sources of uncertainty are small in comparison. Modelling

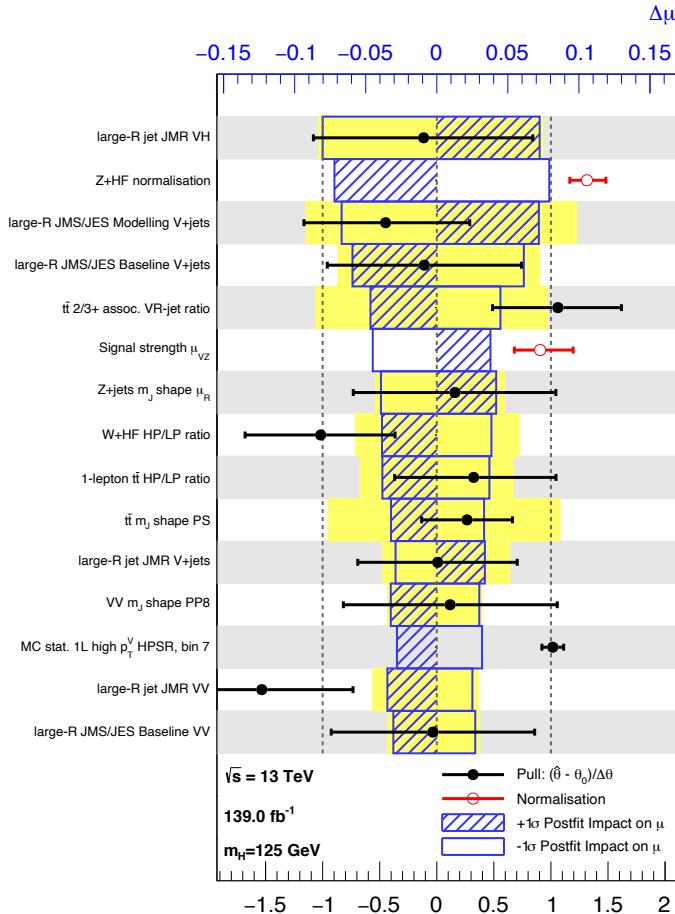


Figure 7.10: Impact of systematic uncertainties on the fitted VH signal-strength parameter $\hat{\mu}_{VH}^{\text{bb}}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value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

2337 uncertainties also have a large contribution to the overall systematic uncertainty. The
2338 biggest contribution to the overall uncertainty is the combined statistical uncertainty
2339 on the simulated samples (0.09), which contain only a finite number of events. Out of
2340 the backgrounds, the $W + \text{jets}$ and $Z + \text{jets}$ have the highest (0.06) and second-highest
2341 (0.05) impact respectively.

2342 7.5 Improved b -tagging using GNNs

2343 7.6 Conclusion

2344 Work has been carried out as part of the boosted VHbb analysis group to understand,
2345 and implement in the global profile likelihood fit, systematic uncertainties on $V + \text{jets}$
2346 samples. This background modelling work is an essential part of the success of
2347 the analysis. So far the fit has proved stable with the inclusion of the $V + \text{jets}$
2348 uncertainties, and detailed studies are now underway to determine the causes behind
2349 any observed pulls of the added NPs. Additional work is ongoing to help with the
2350 derivation of uncertainties on diboson samples, another important background. The
2351 analysis is already advanced, and is now progressing into its final stages. Publication
2352 is expected in the new year.

2353 This analysis would benefit greatly from the improved high p_T b -tagging enabled by
2354 GN1.

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
b-tagging	b-jets	+0.028 / -0.004
	c-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 + jets	+0.049 / -0.047	0.048
↔ W + jets	+0.059 / -0.056	0.058
↔ t̄t	+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 7.12: Breakdown of the observed 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.

2355 Chapter 8

2356 Conclusion

2357 flavour tagging at high pT is important because

2358 tracking suffers from several problems which might be improved by

2359 algorithmic gains in flavour tagging can improve things

2360 ultimately analyses will benefit greatly

2361 future work:

2362 • improve tracking

2363 • use more info for flavour tagging

2364 • improve jet labelling

2365 • boosted xbb tagging

2366 • lead to dihiggs

₂₃₆₇

Bibliography

- ₂₃₆₈ [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- ₂₃₆₉
- ₂₃₇₀ [2] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* **588** (2020) pp. 61–65.
- ₂₃₇₁
- ₂₃₇₂
- ₂₃₇₃ [3] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al., *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606** (2022) pp. 479–483.
- ₂₃₇₄
- ₂₃₇₅
- ₂₃₇₆ [4] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- ₂₃₇₇
- ₂₃₇₈ [5] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) pp. 2632–2637 [[hep-ex/9503003](#)].
- ₂₃₇₉
- ₂₃₈₀ [6] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) pp. 252–255.
- ₂₃₈₁
- ₂₃₈₂ [7] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122** (1983) pp. 103–116.
- ₂₃₈₃
- ₂₃₈₄
- ₂₃₈₅ [8] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) pp. 218–224 [[hep-ex/0012035](#)].
- ₂₃₈₆
- ₂₃₈₇ [9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- ₂₃₈₈
- ₂₃₈₉ [10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) pp. 508–509.
- ₂₃₉₀
- ₂₃₉₁ [11] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*

- 2392 *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 2393 [12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
2394 *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)
2395 p. 1 [[1207.7214](#)].
- 2396 [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*
2397 *experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 2398 [14] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)
2399 p. 083C01.
- 2400 [15] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*
2401 *invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 2402 [16] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)
2403 pp. 579–588.
- 2404 [17] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 2405 [18] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*
2406 *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**
2407 (1968) pp. 367–377.
- 2408 [19] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
2409 *Phys. Rev.* **104** (1956) pp. 254–258.
- 2410 [20] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental*
2411 *test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 2412 [21] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*
2413 *conservation of parity and charge conjugation in meson decays: the magnetic moment*
2414 *of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 2415 [22] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
2416 *Phys. Rev.* **117** (1960) pp. 648–663.
- 2417 [23] J. Goldstone, *Field theories with «superconductor» solutions*, *Il Nuovo Cimento*
2418 (1955–1965) **19** (1961) pp. 154–164.
- 2419 [24] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*
2420 *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](#).
- 2421 [25] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*

- 2422 *ATLAS detector*, ATLAS-CONF-2018-036 (2018).
- 2423 [26] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys. Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](https://arxiv.org/abs/1808.08242)].
- 2425 [27] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2426 [28] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) pp. S08002–S08002.
- 2427
- 2428 [29] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) p. S08004.
- 2429
- 2430 [30] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of Instrumentation* **3** (2008) pp. S08005–S08005.
- 2431
- 2432 [31] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) p. S08003.
- 2433
- 2434 [32] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2435
- 2436
- 2437 [33] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2438
- 2439
- 2440 [34] CERN, “The accelerator complex.” (2012).
- 2441
- 2442 [35] G. C. Strong, *On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case*, [2002.01427](https://arxiv.org/abs/2002.01427).
- 2443
- 2444 [36] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022). <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoveryview/>.
- 2445
- 2446 [37] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* **11** (2016) p. P04008 [[1512.01094](https://arxiv.org/abs/1512.01094)].
- 2447
- 2448 [38] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021 (2019).
- 2449
- 2450 [39] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*, .
- 2451

- 2452 [40] for the ATLAS Collaboration collaboration, *Status of the ATLAS Experiment*, Tech.
2453 Rep., Geneva (2010). 10.3204/DESY-PROC-2010-04/1.
- 2454 [41] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design
2455 Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 2456 [42] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design
2457 Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 2458 [43] ATLAS Collaboration, *The inner detector*, (2022).
2459 <https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2460 [44] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*,
2461 ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 2462 [45] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
2463 *JINST* **13** (2018) p. T05008 [[1803.00844](#)].
- 2464 [46] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor
2465 tracker*, *JINST* **9** (2014) p. P08009 [[1404.7473](#)].
- 2466 [47] ATLAS Collaboration, *Calorimeter*, (2022).
2467 <https://atlas.cern/Discover/Detector/Calorimeter>.
- 2468 [48] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, .
- 2469 [49] G. Artoni, *Muon momentum scale and resolution in pp collisions at $\sqrt{s} = 8$ TeV in
2470 ATLAS*, Tech. Rep., Geneva (2016). 10.1016/j.nuclphysbps.2015.09.453.
- 2471 [50] ATLAS Collaboration, *Muon spectrometer*, (2022).
2472 <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2473 [51] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur.
2474 Phys. J. C* **77** (2017) p. 317 [[1611.09661](#)].
- 2475 [52] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, (2021).
- 2476 [53] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts,
2477 Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2478 [54] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense
2479 Environments*, ATL-PHYS-PUB-2015-006 (2015).
- 2480 [55] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in
2481 dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].

- 2482 [56] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon*
2483 *pixel detector*, *JINST* **9** (2014) p. P09009 [[1406.7690](#)].
- 2484 [57] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment*
2485 *in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332
2486 [[1611.10235](#)].
- 2487 [58] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for*
2488 *LHC Run 3*, ATL-PHYS-PUB-2019-015 (2019).
- 2489 [59] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency*
2490 *measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79**
2491 (2019) p. 970 [[1907.05120](#)].
- 2492 [60] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP*
2493 **04** (2008) p. 063 [[0802.1189](#)].
- 2494 [61] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur. Phys. J. C* **72** (2012)
2495 p. 1896 [[1111.6097](#)].
- 2496 [62] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*
2497 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 2498 [63] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
2499 *the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 2500 [64] ATLAS Collaboration, *Jet energy scale measurements and their systematic*
2501 *uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
2502 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- 2503 [65] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*
2504 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)
2505 p. 242001 [[0802.2470](#)].
- 2506 [66] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*
2507 *decaying into b -quarks with a vector boson at high transverse momentum in pp*
2508 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2509 p. 136204 [[2008.02508](#)].
- 2510 [67] ATLAS collaboration, *Performance of jet substructure techniques for large- R jets in*
2511 *proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
2512 p. 076 [[1306.4945](#)].

- 2513 [68] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*
2514 *13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019)
2515 p. 135 [[1807.09477](#)].
- 2516 [69] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) p. 059
2517 [[0903.0392](#)].
- 2518 [70] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
2519 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010
2520 (2017).
- 2521 [71] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*
2522 *Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).
- 2523 [72] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
2524 *using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024 (2016).
- 2525 [73] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*
2526 *proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) p. 292
2527 [[1603.05598](#)].
- 2528 [74] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
2529 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys.*
2530 *J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 2531 [75] Nazar Bartosik, *Diagram showing the common principle of identification of jets*
2532 *initiated by b-hadron decays*, (2022).
2533 https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png.
- 2534 [76] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)
2535 pp. 577–606 [[hep-ph/9912292](#)].
- 2536 [77] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**
2537 (2018) p. 030001.
- 2538 [78] A. Chen, M. Goldberg, N. Horwitz, A. Jawahery, P. Lipari, G. C. Moneti et al.,
2539 *Limit on the $b \rightarrow u$ coupling from semileptonic b decay*, *Phys. Rev. Lett.* **52** (1984)
2540 pp. 1084–1088.
- 2541 [79] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).
2542 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLLOTS/IDTR-2022-03/>.
- 2543 [80] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*

- 2544 *Experiment*, .
- 2545 [81] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*
2546 *Process Interface for Event Generators*, ArXiv High Energy Physics - Phenomenology
2547 *e-prints* (2001) .
- 2548 [82] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*
2549 *standard format for Les Houches Event Files*, *Computer Physics Communications*
2550 **176** (2007) pp. 300–304.
- 2551 [83] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*
2552 *universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 2553 [84] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*
2554 *back-propagating errors*, *nature* **323** (1986) pp. 533–536.
- 2555 [85] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*
2556 *activity*, *The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 2557 [86] J. Hopfield, *Neural networks and physical systems with emergent collective*
2558 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*
2559 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 2560 [87] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*
2561 *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 2562 [88] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv
2563 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2564 [89] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,
2565 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 2566 [90] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*
2567 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007)
2568 p. 126–126.
- 2569 [91] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*
2570 *shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007)
2571 p. 070–070.
- 2572 [92] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*
2573 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*
2574 *Physics* **2010** (2010) .

- 2575 [93] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)
2576 p. 040 [[1410.8849](#)].
- 2577 [94] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
2578 ATL-PHYS-PUB-2016-020 (2016).
- 2579 [95] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An
2580 introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159
2581 [[1410.3012](#)].
- 2582 [96] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
2583 ATL-PHYS-PUB-2014-021 (2014).
- 2584 [97] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)
2585 p. 244 [[1207.1303](#)].
- 2586 [98] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*
2587 **462** (2001) p. 152.
- 2588 [99] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**
2589 (2010) p. 823 [[1005.4568](#)].
- 2590 [100] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.
2591 Instrum. Meth. A* **506** (2003) p. 250.
- 2592 [101] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter
2593 Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- 2594 [102] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS
2595 Detector*, ATL-PHYS-PUB-2022-027 (2022).
- 2596 [103] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the
2597 ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 2598 [104] ATLAS Collaboration, *Observation of Higgs boson production in association with a
2599 top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173
2600 [[1806.00425](#)].
- 2601 [105] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs
2602 using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **03**
2603 (2020) p. 145 [[1910.08447](#)].
- 2604 [106] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
2605 *Secondary vertex finding in jets with neural networks*, *The European Physical Journal*

- 2606 *C* **81** (2021) .
- 2607 [107] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
2608 *Set2Graph: Learning Graphs From Sets*, arXiv e-prints (2020) p. arXiv:2002.08772
2609 [[2002.08772](#)].
- 2610 [108] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
2611 *b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).
- 2612 [109] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*
2613 *ATLAS detector*, ATL-PHYS-PUB-2017-011 (2017).
- 2614 [110] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
2615 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003 (2017).
- 2616 [111] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*
2617 *detector*, ATLAS-CONF-2014-018 (2014).
- 2618 [112] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13$ TeV data*,
2619 ATL-PHYS-PUB-2015-037 (2015).
- 2620 [113] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*
2621 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
2622 $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 2623 [114] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
2624 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, arXiv e-prints (2020)
2625 p. arXiv:2007.08294 [[2007.08294](#)].
- 2626 [115] J. Shlomi, P. Battaglia and J.-R. Vlimant, *Graph neural networks in particle physics,*
2627 *Machine Learning: Science and Technology* **2** (2021) p. 021001.
- 2628 [116] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,
2629 arXiv e-prints (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 2630 [117] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
2631 *Deep Sets*, arXiv e-prints (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 2632 [118] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv
2633 e-prints (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2634 [119] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
2635 <https://github.com/onnx/onnx>.

- 2636 [120] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs*
2637 *Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- 2638 [121] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson*
2639 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
2640 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2641 p. 136204 [[2008.02508](#)].
- 2642 [122] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated*
2643 *Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039
2644 [[1512.02572](#)].
- 2645 [123] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in*
2646 *proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
2647 p. 076 [[1306.4945](#)].
- 2648 [124] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO*
2649 *calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010)
2650 p. 043 [[1002.2581](#)].
- 2651 [125] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum*
2652 *distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09**
2653 (2014) p. 145 [[1406.3660](#)].
- 2654 [126] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al.,
2655 *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) p. 1889
2656 [[1111.2034](#)].
- 2657 [127] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*,
2658 *JHEP* **10** (2012) p. 155 [[1206.3572](#)].
- 2659 [128] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at*
2660 *NLO with the POWHEG BOX interfaced to GoSam and their merging within*
2661 *MinLO*, *JHEP* **10** (2013) p. 083 [[1306.2542](#)].
- 2662 [129] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to*
2663 *associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003)
2664 p. 073003 [[hep-ph/0306234](#)].
- 2665 [130] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the*
2666 *Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) pp. 149–156
2667 [[hep-ph/0307206](#)].

- 2668 [131] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production*
2669 *at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.*
2670 **107** (2011) p. 152003 [[1107.1164](#)].
- 2671 [132] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects*
2672 *in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- 2673 [133] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated*
2674 *WH production and decay at the LHC*, *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- 2675 [134] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron*
2676 *colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015)
2677 pp. 51–55 [[1407.4747](#)].
- 2678 [135] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson*
2679 *at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].
- 2680 [136] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
2681 *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078
2682 [[1211.5015](#)].
- 2683 [137] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via*
2684 *gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- 2685 [138] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for*
2686 *gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- 2687 [139] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron*
2688 *Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- 2689 [140] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron*
2690 *colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- 2691 [141] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte*
2692 *Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].
- 2693 [142] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair*
2694 *cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) p. 2930
2695 [[1112.5675](#)].
- 2696 [143] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with*
2697 *shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) p. 111
2698 [[0907.4076](#)].

- 2699 [144] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top*
2700 *quark production*, *Phys. Rev. D* **81** (2010) p. 054028 [[1001.5034](#)].
- 2701 [145] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for*
2702 *t-channel single top quark production*, *Phys. Rev. D* **83** (2011) p. 091503 [[1103.2792](#)].
- 2703 [146] E. Re, *Single-top Wt-channel production matched with parton showers using the*
2704 *POWHEG method*, *Eur. Phys. J. C* **71** (2011) p. 1547 [[1009.2450](#)].
- 2705 [147] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated*
2706 *production with a W^- or H^-* , *Phys. Rev. D* **82** (2010) p. 054018 [[1005.4451](#)].
- 2707 [148] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al.,
2708 *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) p. 007 [[0811.4622](#)].
- 2709 [149] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) p. 034
2710 [[1905.09127](#)].
- 2711 [150] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
2712 *Phys. Rev. Lett.* **108** (2012) p. 111601 [[1111.5206](#)].
- 2713 [151] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008)
2714 p. 039 [[0808.3674](#)].
- 2715 [152] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour*
2716 *dipole factorisation*, *JHEP* **03** (2008) p. 038 [[0709.1027](#)].
- 2717 [153] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton*
2718 *showers. The NLO case*, *JHEP* **04** (2013) p. 027 [[1207.5030](#)].
- 2719 [154] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson*
2720 *Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.*
2721 *Rev. Lett.* **103** (2009) p. 082001 [[0903.2120](#)].
- 2722 [155] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**
2723 (2016) p. 023001 [[1510.03865](#)].
- 2724 [156] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to*
2725 *Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* **03**
2726 (2012) p. 075 [[1112.5142](#)].
- 2727 [157] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo*
2728 *program for Higgs production in vector-boson fusion and Higgs strahlung at hadron*
2729 *colliders*, *Comput. Phys. Commun.* **195** (2015) pp. 161–171 [[1412.5390](#)].

- 2730 [158] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*
2731 *Run-2*, ATL-PHYS-PUB-2015-022 (2015).
- 2732 [159] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS*
2733 *detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 2734 [160] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using*
2735 *ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2018-006
2736 (2018).
- 2737 [161] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using*
2738 *a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001 (2018).
- 2739 [162] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*
2740 *ATLAS Detector*, ATL-PHYS-PUB-2014-013 (2014).
- 2741 [163] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark*
2742 *pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836
2743 [[1906.11005](#)].
- 2744 [164] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*
2745 *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) p. 204
2746 [[1412.3989](#)].
- 2747 [165] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*
2748 *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*
2749 *D93* (2016) p. 033006 [[1506.07443](#)].
- 2750 [166] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a*
2751 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 2752 [167] J. Pumplin et al., *New generation of parton distributions with uncertainties from*
2753 *global qcd analysis*, *JHEP* **07** (2002) p. 012 [[0201195](#)].
- 2754 [168] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*
2755 *roostats project*, arXiv preprint [arXiv:1009.1003](#) (2010) .

2756 List of figures

<small>2757</small>	2.1	The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a corresponding 5-dimensional potential.	18
<small>2758</small>			
<small>2759</small>	2.2	Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.	24
<small>2760</small>			
<small>2761</small>	2.3	Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [24]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fusion, vector boson fusion, associated production with vector bosons, and top quark fusion.	24
<small>2762</small>			
<small>2763</small>	2.4	Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [24]. 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.	25
<small>2764</small>			
<small>2765</small>	3.1	An overview of the CERN accelerator complex [34]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.	28
<small>2766</small>			
<small>2767</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. [35].	29
<small>2768</small>			
<small>2769</small>			
<small>2770</small>			
<small>2771</small>			
<small>2772</small>			
<small>2773</small>			
<small>2774</small>			
<small>2775</small>			
<small>2776</small>			
<small>2777</small>			
<small>2778</small>			
<small>2779</small>			
<small>2780</small>			

2781	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. [36].	31
2787	3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [33]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.	33
2791	3.5 Average pile-up profiles measured by ATLAS during Run 2 [33]. Higher levels of pile-up are planned for Run 3.	33
2793	3.6 A 3D model of the entire ATLAS detector [40]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.	34
2797	3.7 A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [43].	35
2799	3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [36].	36
2801	3.9 A schematic cross-sectional view of the ATLAS IBL [44].	37
2802	3.10 The ATLAS calorimeters [47]. 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.	39
2805	3.11 The ATLAS muon spectrometer [50].	41
2806	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 [55].	45
2810	3.13 A sketch of electron reconstruction using the ATLAS detector [72]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.	48

2813	4.1	Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [75]. 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.	52
2814			
2815			
2816			
2817			
2818	4.2	The truth b -hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed b -jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.	53
2819			
2820			
2821	4.3	At lower p_T (left) the decay length of the b -hadron is reduced, and the resulting decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching detector elements. As a result, the ID may be unable to resolve charged depositions from different particles, instead reconstructing merged clusters.	54
2822			
2823			
2824			
2825			
2826			
2827	4.4	Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. For each of these, standard tracks and pseudo-tracks are plotted. Hit multiplicities on the pseudo-tracks at high p_T due to the increased flight of the b -hadron. The baseline tracks have more hits than the pseudo-tracks, indicating that they are being incorrectly assigned additional hits.	56
2828			
2829			
2830			
2831			
2832			
2833			
2834	4.5	b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [79]. 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.	56
2835			
2836			
2837			
2838			
2839			
2840			
2841	4.6	The rate of shared hits on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudo-tracks represent ideal performance given the ATLAS detector, and given also the efficiency of the NN cluster splitting algorithms.	59
2842			
2843			
2844			

2845	4.7 Profiles, as a function of parent b -hadron p_T , of good (left) and wrong (right) hit assignment rates on the IBL for tracks using baseline tracking (black), the modified version of the outlier removal procedure (red).	61
2848	4.8 (left) B hadron decay track d_0 pulls for baseline and modified GX2F tracks. (right) The magnitude of the decay track d_0 pull as a function of B hadron transverse momentum.	62
2851	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 is passed to an activation function. Each neuron can be thought of as a logistic regression model.	65
2855	5.2 Several common choices for the activation function g of an artificial neuron.	66
2856	5.3 Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases, and the number of tracks in the jet increases (not shown).	69
2860	5.4 The light-jet efficiency of the low level tagger SV1 as a function of b -jet efficiency for the nominal tracking setup (black) and for the case where fake tracks which are not from the decay of a b -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to algorithm performance.	70
2865	5.5 (left) Normalised histogram of the model output separated for good and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line).	73
2869	5.6 The effect of applying the fake track identification algorithm alongside the b -hadron decay track identification on the jet tagging performance of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) is compared to two working points, labelled “A” (red) and “B” (orange) of fake track removal, 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.	75

2877	5.7 The effect of applying the fake track identification algorithm alongside 2878 the b -hadron decay track identification on the jet tagging performance 2879 of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with 2880 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) 2881 is compared to two working points of fake track removal (orange and red). 2882 Removal of fake tracks based on truth information is shown by the green 2883 curves.	76
2884	6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right). 2885 The existing approach utilises low-level algorithms (shown in blue), the 2886 outputs of which are fed into a high-level algorithm (DL1r). Instead 2887 of being used to guide the design of the manually optimised algorithms, 2888 additional truth information from the simulation is now being used as 2889 auxiliary training targets for GN1. The solid lines represent reconstructed 2890 information, whereas the dashed lines represent truth information.	80
2891	6.2 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of 2892 n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The 2893 jet features are copied for each of the tracks, and the combined jet-track 2894 vectors of length 23 form the inputs of GN1.	86
2895	6.3 The network architecture of GN1. Inputs are fed into a per-track initialisation 2896 network, which outputs an initial latent representation of each track. 2897 These representations are then used to populate the node features of a 2898 fully connected graph network. After the graph network, the resulting node 2899 representations are used to predict the jet flavour, the track origins, and 2900 the track-pair vertex compatibility.	87
2901	6.4 Comparison between the DL1r and GN1 b -tagging discriminant D_b for 2902 jets in the $t\bar{t}$ sample. The 85% WP and the 60% WP are marked by the 2903 solid (dashed) lines for GN1 (DL1r), representing respectively the loosest 2904 and tightest WPs used by analyses. A value of $f_c = 0.018$ is used in the 2905 calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. The distributions 2906 of the different jet flavours have been normalised to unity area.	91

2907	6.5 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2908 tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV. The 2909 ratio with respect to the performance of the DL1r algorithm is shown in the 2910 bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for 2911 DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands 2912 are denoted by the shaded regions. At b -jet tagging efficiencies less than 2913 $\sim 75\%$, the light-jet rejection becomes so large that the effect of the low 2914 number of jets is visible. The lower x -axis range is chosen to display the 2915 b -jet tagging efficiencies usually probed in these regions of phase space.	94
2916	6.6 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2917 tagging efficiency for jets in the Z' sample with $250 < p_T < 5000$ GeV. The 2918 ratio with respect to the performance of the DL1r algorithm is shown in the 2919 bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for 2920 DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands 2921 are denoted by the shaded regions. At b -jet tagging efficiencies less than 2922 $\sim 20\%$, the light-jet rejection becomes so large that the effect of the low 2923 number of jets is visible. The lower x -axis range is chosen to display the 2924 b -jet tagging efficiencies usually probed in these regions of phase space.	95
2925	6.7 The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the 2926 Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 2927 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of D_b for 2928 DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands 2929 are denoted by the shaded regions.	95
2930	6.8 The b -jet (left) and light-jet (right) rejections as a function of the c -jet 2931 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the 2932 performance of the DL1r algorithm is shown in the bottom panels. Binomial 2933 error bands are denoted by the shaded regions. At c -jet tagging efficiencies 2934 less than $\sim 25\%$, the light-jet rejection becomes so large that the effect of the 2935 low number of jets is visible. The lower x -axis range is chosen to display 2936 the c -jet tagging efficiencies usually probed in these regions of phase space.	97

2937	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. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The lower x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.	97
2943	6.10 A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked.	98
2946	6.11 A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked.	98
2949	6.12 The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 65\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.	100
2960	6.13 The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 25\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.	101

2971	6.14 Vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample 2972 using the exclusive (left) and inclusive (right) vertex finding approaches. 2973 Efficient vertex finding requires the recall of at least 65% of the tracks in 2974 the truth vertex, and allows no more than 50% of the tracks to be included 2975 incorrectly. Binomial error bands are denoted by the shaded regions.	104
2976	6.15 Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in 2977 the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . 2978 Efficient vertex finding requires the recall of at least 65% of the tracks in 2979 the truth vertex, and allows no more than 50% of the tracks to be included 2980 incorrectly.	105
2981	6.16 ROC curves for the different groups of truth origin labels defined in Table 5.1 2982 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The 2983 FromB, FromBC and FromC labels have been combined, weighted by their 2984 relative abundance, into the Heavy Flavour category, and the Primary and 2985 OtherSecondary labels have similarly been combined into a single category. 2986 The mean weighted area under the ROC curves (AUC) is similar for both 2987 samples.	107
2988	6.17 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2989 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track 2990 selection trainings of GN1. The ratio to the performance of the DL1d 2991 algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in 2992 the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial 2993 error bands are denoted by the shaded regions.	108
2994	6.18 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2995 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track 2996 selection trainings of GN1. The ratio to the performance of the DL1d 2997 algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in 2998 the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial 2999 error bands are denoted by the shaded regions.	109
3000	7.1 Diagrams of the signal process (top) and the 0-lepton and 1-lepton $t\bar{t}$ back- 3001 grounds (middle, bottom). Objects to the right of centre are reconstructed 3002 within the large- R jet. For the backgrounds, the large- R jet contains a 3003 mis-tagged c - or light-jet.	119

3004	7.2 Normalised leading large- R jet mass distribution from Z and $W+hf$ processes in the HP SR of the 0-lepton channel. The renormalisation scale μ_r has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential function is fitted to the ratio between the nominal and variation samples.	129
3005		
3006		
3007		
3008	7.3 The comparison on m_J shapes between SHERPA and MADGRAPH samples from $V+hf$ process in p_T^V inclusive signal regions. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.	130
3009		
3010		
3011	7.4 Normalised leading large- R jet mass distribution from WZ and ZZ process, merged among all the signal regions and lepton channels. 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.	132
3012		
3013		
3014		
3015	7.5 The comparison on m_J shapes between SHERPA and POWHEG+PYTHIA 8 samples from WZ and ZZ process in high and low purity signal regions. p_T^V regions and 0- and 1-lepton channels are merged. 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.	133
3016		
3017		
3018		
3019		
3020	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.	139
3021		
3022		
3023		
3024		
3025		
3026		
3027		
3028		
3029		

3030	7.7 The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton 3031 channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) 3032 the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. 3033 The background contributions after the likelihood fit are shown as filled 3034 histograms. The Higgs boson signal ($m_h = 125 \text{ GeV}$) is shown as a filled 3035 histogram on top of the fitted backgrounds normalised to the signal yield 3036 extracted from data ($\mu_{VH}^{bb} = 0.72$), and unstacked as an unfilled histogram, 3037 scaled by the SM prediction times a factor of 2. The size of the combined 3038 statistical and systematic uncertainty for the sum of the fitted signal and 3039 background is indicated by the hatched band. The highest bin in the 3040 distributions contains the overflow.	140
3041	7.8 m_J distribution in data after subtraction of all backgrounds except for the 3042 WZ and ZZ diboson processes. The contributions from all lepton channels 3043 and signal regions are summed and weighted by their respective values of 3044 the ratio of fitted Higgs boson signal and background yields. The expected 3045 contribution of the associated WH and ZH production of a SM Higgs 3046 boson with $m_H = 125 \text{ GeV}$ is shown scaled by the measured combined 3047 signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to 3048 its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and 3049 systematic uncertainty is indicated by the hatched band. This error band is 3050 computed from a full signal-plus-background fit including all the systematic 3051 uncertainties defined in Section 7.2, except for the VH/VZ experimental 3052 and theory uncertainties.	141
3053	7.9 Signal strength compatibility test between the (3+1)-POI fit (with the three 3054 lepton channels splitted) and the default (1+1)-POI fit. The compatibility 3055 of the three channels is evaluated via a χ^2 difference test and results in a 3056 p-value of 49%.	142

- 3057 7.10 Impact of systematic uncertainties on the fitted VH signal-strength pa-
3058 rameter μ_{VH}^{bb} sorted in decreasing order. The boxes show the variations
3059 of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual
3060 nuisance parameter to its post-fit value modified upwards or downwards by
3061 its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of
3062 up- and down-variations can be distinguished via the dashed and plane box
3063 fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying
3064 each nuisance parameter by ± 1 . The filled circles show the deviation of the
3065 fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value
3066 θ_0 expressed in standard deviations with respect to their nominal uncer-
3067 tainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties
3068 on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show
3069 the fitted values and uncertainties of the normalization parameters that are
3070 freely floating in the fit. Pre-fit, these parameters have a value of one. . 144

3071 List of tables

<small>3072</small>	2.1 The fermions of the SM [14]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.	13
<small>3075</small>	2.2 The bosons of the SM [14]. The photon, weak bosons and gluons are gauge bosons arising from gauge symmetries, and carry the four fundamental forces of the SM. The recently discovered Higgs boson is the only fundamental scalar particle in the SM.	14
<small>3079</small>	3.1 Overview of the different LHC runs [32, 33]. The average number of interactions per bunch-crossing is denoted as $\langle \mu \rangle$ (see Section 3.2.3), and is here averaged over the entire run. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.	27
<small>3083</small>	5.1 Truth origins which are used to categorise the physics process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [55]. A truth-matching probability of less than 0.5 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions, K_S^0 and Λ^0 decays, and hadronic interactions.	68
<small>3090</small>	5.2 Input features to the fake track classification NN. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55], while split hits are hits used on multiple tracks which have been identified as merged.	72

3096 5.3	Quality selections applied to tracks, where d_0 is the transverse IP of the track, z_0 is the longitudinal IP with respect to the PV and θ is the track polar angle. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55]. Shared hits on pixel layers are given a weight of 1, while shared hits in the SCT are given a weight of 0.5. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.	73
3103 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. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.	74
3107 6.1	Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55], while split hits are hits used on multiple tracks which have been identified as merged. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1 Lep model.	84
3114 6.2	A summary of GN1's different classification networks used for the different training objectives. The hidden layers column contains a list specifying the number of neurons in each layer.	88
3117 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 6.4.3, outperforms the MLP model for both $t\bar{t}$ and Z' jets.	106
3124 7.1	Summary of the definition of the analysis regions. Signal enriched regions are marked with the label SR. There are regions with relatively large signal purity (HP SR) and with low purity (LP SR). Background enriched regions are marked with the label CR. The shorthand “add” stands for additional small- R jets, i.e. number of small- R jets not matched to the Higgs-jet candidate.	113

3130	7.2 Signal and background processes with the corresponding generators used for the nominal samples. 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 [155] using the internal algorithm in POWHEG-Box v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the $pp \rightarrow ZH$ process already includes the $gg \rightarrow ZH$ contribution. The $qq \rightarrow ZH$ process is normalised using the cross-section for the $pp \rightarrow ZH$ process, after subtracting the $gg \rightarrow ZH$ contribution. An additional scale factor is applied to the $qq \rightarrow VH$ processes as a function of the transverse momentum of the vector boson, to account for electroweak (EW) corrections at NLO. This makes use of the VH differential cross-section computed with HAWK [156, 157].	115
3142	7.3 Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels and sub-channels. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region.	117
3146	7.4 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.	123
3149	7.5 0-lepton $W+jets$ nominal sample flavour composition and total event yield. . .	125
3150	7.6 1-lepton $W+jets$ nominal sample flavour composition and total event yield. . .	125
3151	7.7 0-lepton $Z+jets$ nominal sample flavour composition and total event yield. . .	125
3152	7.8 2-lepton $Z+jets$ nominal sample flavour composition and total event yield. . .	125
3153	7.9 $V+jets$ acceptance uncertainties. $W+jets$ SR and CR uncertainties marked with a superscript † are correlated. The 1L $W+jets$ H/M uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions. The 0L $W+jets$ Wbb Norm uncertainty is only applied when a floating normalisation for Wbb cannot be obtained from the 1L channel. A 30% uncertainty for $Z \rightarrow b\bar{b}$ norm is applied in the 1L channel when a floating normalisation for $Z \rightarrow b\bar{b}$ cannot be obtained from the 0L or 2L channels.	127

3161	7.10 Diboson acceptance uncertainties. All uncertainties except channel extrapolation uncertainties are fully correlated between ZZ and WZ processes and channels.	132
3164	7.11 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. The errors represent the combined statistical and systematic uncertainties.	137
3167	7.12 Breakdown of the observed 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.	146