



Maurin, Théo (2026) *Piecing together the Standard Model $H \rightarrow WW^* \text{ semi-leptonic decay in ATLAS}$* . PhD thesis.

<https://theses.gla.ac.uk/86192/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University
of Glasgow | Experimental
Particle Physics

Piecing Together the Standard Model $H \rightarrow WW^*$ Semi-Leptonic Decay in ATLAS.

Théo Maurin

of the school of Physics and Astronomy
within the College of Science and Engineering

A dissertation submitted to the University of Glasgow
for the degree of Doctor of Philosophy

Abstract

This thesis presents the design of a novel analysis searching for the previously unexplored Standard Model $H \rightarrow WW^*$ semi-leptonic final state, considering both the ggF and VBF Higgs production modes. The phenomenology study that laid the foundation for the analysis is presented, and the experimental adaptations and improvements are subsequently detailed. The expected 95% CL_s upper limit on the signal strength is found to be $\mu_s = 1.804^{+0.715}_{-0.512}$ for an integrated luminosity of 140 fb^{-1} , which corresponds to the ATLAS data accumulated during Run 2. Projections to the HL-LHC integrated luminosity indicate that a 5σ observation of the channel is achievable, provided the dominant systematic uncertainty arising from the lack of MC statistics for background processes is mitigated. In addition, relevant contributions to flavour tagging within ATLAS and improvements to the Neutrino Weighting algorithm are presented and discussed.

Declaration

The work presented in this thesis is the result of my own work with the Experimental Particle Physics group in the School of Physics and Astronomy at the University of Glasgow, except where explicit reference is made to the work of others. The thesis has not been submitted for another qualification to this or any other university.

This research was performed as part of the ATLAS Collaboration, and thus with the help of expertise, tools and work of thousands of individuals over many years.

Théo Maurin

Acknowledgements

Je tiens à remercier mes parents et mon petit frère, qui m'ont toujours encouragé dans mes aventures, nourrissant ma curiosité tout en m'apportant un soutien physique et moral inébranlable. Dédicace à tous les stous de France, qui ne cessent de me rassasier à chaque fois que je les vois. Ce fut un immense plaisir de vous faire visiter le CERN, Alec, Frankie, Nigolin et Wanda ! J'espère pouvoir te le faire visiter un jour, Raph ; prends bien soin de toi avant que je ne vienne vous gank, Florian et toi, en live au Canada. Merci à Tristan et Paolo de rester des bougs solides, des bastions du Sud. À Oscar, merci pour notre amitié qui traverse les années ; c'était cool de découvrir Glasgow avec toi. Je vous aime tous profondément.

To the people of Glasgow, thank you for making my time as a student in Scotland such an amazing experience. I feel lucky to be part of the Douglas crew, and even luckier that I get to continue through life alongside you all: thank you Jake, James, Lina and Pascale. I could not hope for better PhD brethren: thank you Adam, Aleksandrina, Aurys (see you in the Bayou), elusive Hasret, Kerr, and Dr. Wrishik, mon frère d'une autre mère. I love you all, and I hope that our future paths intertwine (entangle?).

Thank you, Jay, for both your supervision and the many interesting conversations we had, during which you shared your intricate knowledge of experimental particle physics. Thank you, Federica, for your technical guidance. It is a great pleasure to work with you and learn from your expertise. Thank you and good luck to my younger PhD siblings, Ivo and Jay NLO, I always enjoy seeing you both. Thank you Alice for all your advice, I wish you the best. More broadly, I would like to thank the ATLAS collaboration and the particle physics community: the shoulders of giants on which we all stand.

As always, I cannot find the words to express my gratitude to you, Smelia. Thank you for your steadfast kindness and all your support throughout the years. I feel incredibly fortunate to have you. You have certainly been the single greatest contributor to my survival, especially during these last few months of thesis writing! Thank you for drawing the Stubz illustrations for each chapter.

To my Stubz, I did not get to be with you for very long, but no amount of time would have been enough in the end. You blessed a year of my life, and I will remember you forever. Rest in peace, my soft little friend.



Contents

1	Theory	3
1.1	The Standard Model	3
1.1.1	The Higgs Mechanism	6
1.1.2	W boson Higgs coupling and spin	13
1.2	Quantum Information	18
1.2.1	Quantum Entanglement Theory	20
1.2.2	Bell's theorem and Bell operators	22
1.2.3	Measuring Quantum Entanglement in $H \rightarrow WW^*$ at colliders	25
2	The ATLAS Detector	27
2.1	Introduction	27
2.2	The Large Hadron Collider	27
2.3	Coordinate System	31
2.4	Detector Overview	33
2.4.1	Magnets	33
2.4.2	Inner Detector	34
2.4.3	Electromagnetic Calorimeter	37
2.4.4	Hadronic Calorimeter	38
2.4.5	Muon Spectrometer	39
2.5	Trigger	40
2.6	Reconstruction of physics objects	41
2.6.1	Tracks	42
2.6.2	Vertexing	44
2.6.3	Jets	45
2.6.4	Electrons	46
2.6.5	Muons	47
2.6.6	Missing Transverse Energy $\vec{E}_T^{\text{miss}}$	48
2.7	Monte-Carlo generators and detector simulation	48

3	Flavour Tagging	53
3.1	Introduction	53
3.2	The DL1DV01 tagger	53
3.3	The GN2v01 tagger	54
3.4	MC-MC efficiency maps	60
3.5	Flavour sensitive observables in RIVET	66
3.5.1	Introduction to RIVET	66
3.5.2	Comparing generators at truth level	67
4	Neutrino Weighting Improvements	75
4.1	Introduction and definitions	75
4.2	Using MINUIT	77
4.2.1	In dileptonic $t\bar{t}$	78
4.2.2	In semi-leptonic HWW^*	83
4.3	Hessian matrix and the Newton Method	85
4.4	Performance study	87
4.4.1	Dileptonic $t\bar{t}$	88
4.4.2	Semi-leptonic HWW^*	92
4.4.3	Conclusion	98
5	Proposal to Measure $H \rightarrow WW^*$ Entanglement in New Decay Channel	101
5.1	Background Suppression and Event Reconstruction	102
5.1.1	MC and Simulation of Detector Effects	103
5.1.2	Event Selection	104
5.2	Observing Entanglement in Higgs Events	107
5.2.1	Unfolding	109
5.2.2	Expected Precision	111
5.3	Simulating E_T^{miss}	113
5.4	Conclusion	114
6	Search for Semi-Leptonic $H \rightarrow WW^*$	117
6.1	Introduction	117
6.2	Signal MC generation and validation	117
6.3	Background processes	118
6.4	Data and Monte-Carlo samples	120
6.5	Lepton fakes estimation	122
6.5.1	The Asymptotic Matrix Method	124

6.5.2	Run 2 efficiencies	126
6.5.3	Run 3 efficiencies	127
6.6	Preselection	131
6.6.1	JetPairingNet	132
6.6.2	Event Reconstruction	139
6.7	Multivariate Analysis	140
6.7.1	Boosted Decision Trees	141
6.7.2	Model Inputs	144
6.7.3	BDT Training and Performance	150
6.8	Systematic Uncertainties	153
6.8.1	Detector Systematics	155
6.8.2	Theoretical Systematics	156
6.9	Fitting and extracting limits on the cross-section	157
6.9.1	Statistical Model	157
6.9.2	Expected sensitivity	161
6.9.3	Mismodelling	173
6.9.4	Future steps	175
	Bibliography	179
	List of figures	199
	List of tables	207

Introduction

The Standard Model of particle physics provides our current theoretical description of the fundamental constituents of matter and their interactions. Its remarkable success culminated in the discovery of the Higgs boson in 2012, extending the particle content predicted by the theory and confirming the mechanism responsible for electroweak symmetry breaking and the generation of the masses of the elementary particles. Despite this important advancement, many questions remain unanswered, and the quest for a complete understanding of the physical world persists. Since its discovery, the precise measurement of the Higgs boson's properties has become one of the central goals of the LHC physics programme. The properties of the scalar boson are probed through the final state particles resulting from its decay. Experimentally, the $H \rightarrow WW^*$ decay is primarily studied by considering the channel where the W bosons both decay to leptons¹, the *dileptonic decay*. The electrons and muons leave a distinctive detector signature that is relatively easy to distinguish from background processes. The *semi-leptonic decay*, where one W decays to leptons and the other to quarks, is more challenging to separate from the relevant background processes, but possesses a very interesting advantage over its dileptonic counterpart: only one neutrino is produced in the final state. In that configuration, the hadronic W boson can be reconstructed directly and the leptonic W boson can be inferred by using the missing transverse energy and the lepton. In other words, the Higgs can be fully reconstructed, which is crucial to measure the spin correlations of the WW^* system. The reconstructability not only provides a straightforward route to measure observables related to quantum information and the entanglement of the W bosons, but also helps overcome the challenge associated with isolating the signal. This thesis presents the development of the components that together constitute the first baseline analysis targeting the $H \rightarrow WW^*$ semi-leptonic decay, with the goal of setting the first limits on its production.

¹In ATLAS, the term “leptons” is often used to designate charged leptons, excluding the neutrinos that are not detected directly by the detector. This convention is used in this thesis: a distinction is made between leptons and neutrinos.

Chapters 1 and 2 introduce the relevant theory of the Standard Model, and describe the ATLAS detector and the software framework used to produce the data and simulated samples employed throughout the analysis, respectively.

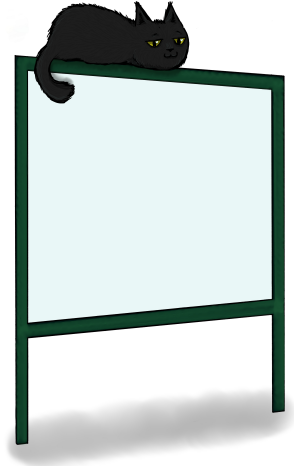
Chapter 3 details the art of flavour tagging, an important component of the analysis, from section 3.1 to section 3.3. The succeeding sections 3.4 and 3.5 present original work undertaken as part of this thesis in the context of the latest flavour-tagging algorithm used in ATLAS. This includes the computation of MC-MC efficiency maps, and the study of flavour-tagging observables in simulation.

Chapter 4 describes the algorithm used to reconstruct the neutrino and provide discrimination power for the signal. Consequent improvements to the optimisation element within the algorithm, focusing both on the scope of the analysis and top quark physics, are presented and discussed. The algorithm predates this thesis, but the improvement, both within the scope of this analysis and in the broader context of top-quark physics, as well as the benchmarking studies, are entirely the result of personal work.

Chapter 5 summarises the published phenomenology study that served as the initial blueprint for the analysis, while also providing a procedure for unfolding a Bell operator observable sensitive to the entanglement of the W bosons.

Finally, the analysis is presented in chapter 6. The development of novel reconstruction methods, a dedicated machine learning framework, the fake lepton background estimation, and the statistical model are detailed. The expected limits, HL-LHC projections, and future steps to optimise the final result are discussed. The mathematical frameworks for the Asymptotic Matrix Method, XGBoost Boosted Decision Trees, and the TRExFitter software are described in detail. Every result presented, some depending on the tools just listed, originates from personal work.

For chapters 3 to 6, unless a reference is given or explicitly stated, the plots and numerical results are the fruit of personal work undertaken as part of this thesis.



1 Theory

1.1 The Standard Model

The Standard Model (SM) of particle physics is a relativistic Quantum Field Theory (QFT) that provides a comprehensive framework for understanding the fundamental forces and their interactions with the particles that constitute the universe¹. Developed over the course of the 20th century, the SM unifies three of the four known fundamental interactions: the electromagnetic force, the weak force, and the strong force.

At the core of the SM are the elementary particles, divided into two primary categories: fermions and bosons. Fermions constitute the building blocks of matter and are described by spinor fields [4], while bosons are described by vector fields that mediate the fundamental forces that connect the fields together with interactions. These two types of particles differ in their intrinsic spin, with fermions having half-integer spin and bosons having integer spin. The connection between spin and quantum statistics is established by the spin-statistics theorem, a fundamental result that follows from Lorentz invariance and a stable vacuum [5–7].

The fermions consist of twelve particles, divided into three generations of quarks and leptons. The quarks combine to form hadrons and their generations are given by the following pairs of up-type and down-type quarks: up and down, charm and strange, top and bottom. The leptons are the electron, the muon and the tau, each with an associated neutrino. Every fermion particle has a corresponding anti-particle with opposite electrical charge (the neutrino is the only neutral fundamental fermion, giving rise to interesting possibilities for the anti-neutrino [8]).

¹Sections 1.1 and 1.1.1 have been written by consulting the following sources [1–3].

The interactions between these particles are governed by fundamental forces that are associated with gauge symmetries of the Lagrangian of the universe. These symmetries are expressible mathematically as the gauge group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, which can be divided into the strong ($SU(3)_C$) and electroweak ($SU(2)_L \otimes U(1)_Y$) forces. The strong interaction is mediated by gluons and the electroweak interaction is mediated by four gauge bosons that get disentangled by the Higgs mechanism described in section 1.1.1. Following Noether's theorem [9], each of these continuous global symmetries comes with a conserved physical quantity: the strong colour charge C , the weak isospin T^3 and the hypercharge Y .

Despite describing and providing predictive power over all these interactions, the SM Lagrangian can be written very densely with contractions and by splitting it in four smaller, physically relevant, components:

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

$$= -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} + i\bar{\psi}\gamma^\mu D_\mu\psi + |D_\mu\Phi|^2 - V(\Phi) + y_{ij}\bar{\psi}_L^i\Phi\psi_R^j + h.c.. \quad (1.2)$$

Each term must correspond to a Lorentz invariant scalar to add up to a relativistic Lagrangian. The $\mathcal{L}_{\text{Gauge}}$ term is composed of the scalar products of rank two or three tensors that characterise the kinetics of gauge fields as described by Yang-Mills theory. The term can be expanded as:

$$\mathcal{L}_{\text{Gauge}} = -\frac{1}{4} \left(B_{\mu\nu}B^{\mu\nu} + W_{\mu\nu}^a W_a^{\mu\nu} + G_{\mu\nu}^a G_a^{\mu\nu} \right), \quad (1.3)$$

where $B_{\mu\nu}$, $W_{\mu\nu}^a$ and $G_{\mu\nu}^a$ are tensors describing the dynamics of the $U(1)_Y$, $SU(2)_L$, and $SU(3)_C$ gauge fields, respectively. The weak and gluon terms have an adjoint index a , in addition to the Lorentz indices μ and ν , which is associated with the generators of their corresponding Lie group [1, 2, 6]. For the Lie group $SU(N)$ the number of generators is given by $N^2 - 1$, adding gauge fields by iterating over a indices thus yields three weak bosons and eight gluons. This additional structure provided by the $SU(N)$ groups makes these gauge fields non-Abelian, unlike the $U(1)$ gauge field, which results in the weak and gluon fields carrying the isospin and strong charge, respectively, and thus, self-interacting. This can be seen more clearly by expanding the tensors:

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon_{bc}^a W_\mu^b W_\nu^c, \quad (1.4)$$

where W_μ^a are the three weak boson fields, g the $SU(2)_L$ coupling constant and ϵ_{bc}^a the structure constant of $SU(2)$. One can obtain the same expression for $G_{\mu\nu}^a$ by swapping W_μ^a for the gluon fields G_μ^a , g for the strong coupling constant and ϵ_{bc}^a for the structure constant of $SU(3)$. The last term characterises the self-interaction of the non-Abelian gauge fields, one can expect cubic and quartic gauge field terms to occur in the sum over Lorentz indices in equation 1.3. For $U(1)_Y$, this additional structure is not present and the tensor simply expands as:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (1.5)$$

The Lagrangian must remain locally gauge invariant. This means that if we apply a transformation $U(x) \in G$ that depends on spacetime x , and where G is one of the gauge groups of our theory, to the fermion fields, the physics must remain unchanged. Additional unphysical “ghost” fields, called Faddeev-Popov fields [10], are required and added to preserve gauge symmetry due to the non-Abelian gauge fields.

The $\mathcal{L}_{\text{Fermions}}$ term contains the fermion spinor fields ψ which obey the Dirac equation. The Dirac equation is obtained by applying the Lagrange-Euler equation to the following Lagrangian:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \quad (1.6)$$

where γ^μ are the Dirac 4×4 matrices. Ignoring the second term for now (see section 1.1.1), it can be seen that the kinematic term contains a differential operator ∂_μ that acts on spacetime. This causes the Lagrangian to violate gauge invariance since, given a gauge transformation $U(x)$, $\partial_\mu(U(x)\psi) \neq U(x)\partial_\mu\psi$. In order to cancel the additional terms that violate gauge symmetry, the differential operator is upgraded to the covariant derivative D_μ by adding one term per gauge symmetry:

$$D_\mu = \partial_\mu - ig_0 Y B_\mu - ig_1 T_a W_\mu^a - ig_2 T'_a G_\mu^a, \quad (1.7)$$

where g_j represents the relevant coupling constants, Y is the hypercharge, T_a the $SU(2)_L$ generators that represent the isospin charge and T'_a the $SU(3)_C$ generators that represent the colour charge. The covariant derivative makes the final expression $i\bar{\psi}\gamma^\mu D_\mu\psi$ gauge invariant provided that the gauge fields transform as follows:

$$T_a W_\mu^a(x) \rightarrow U T_a W_\mu^a(x) U^\dagger - \frac{i}{g_1} (\partial_\mu U) U^\dagger, \quad (1.8)$$

$$B_\mu(x) \rightarrow B_\mu(x) + \frac{1}{g_0} \partial_\mu \theta(x), \quad (1.9)$$

where $U(x)$ is a $SU(2)_L$ transformation, the $U(1)_Y$ transformation is taken as $e^{-iY\theta(x)}$ and the rule for the gluon field G_μ^a is straightforwardly similar to equation 1.8. These new terms that define the covariant derivative convey how the fermionic fields, given they are carrying the relevant charges, interact with the gauge fields upon expanding $\mathcal{L}_{\text{Fermions}}$. The spinor fields ψ can be decomposed into its chiral components, the left- and right-handed Weyl spinors [1, 11]:

$$\psi = \psi_L + \psi_R. \quad (1.10)$$

This is important due to the electroweak force sensitivity to chirality: only ψ_L and $\bar{\psi}_R$ interact with the weak force.

1.1.1 The Higgs Mechanism

In the Lagrangian description of quantum fields, the mass term describes the masses of the field's quantas (particles). Such a term, present in the Dirac Lagrangian shown in equation 1.6, can be expressed after expanding the fermionic field into their chiral component as in equation 1.10:

$$\mathcal{L}_{\text{mass}} = -m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L), \quad (1.11)$$

where m is the Dirac mass, equal to the invariant mass of the fermions that are created or annihilated by the field. This kind of term does not appear directly in the SM Lagrangian, presented in equation 1.1, since it is not invariant under $SU(2)_L \otimes U(1)_Y$ due to ψ_L and ψ_R transforming differently with respect to the gauge group. Experimentally, every fermion is observed to have a mass and it is thus important to add structure to the theory to retrieve a mass term in the Lagrangian. The Higgs mechanism, first postulated in 1964 by Higgs [12], Englert and Brout [13], solves that issue and completes the SM by introducing a new complex scalar field ϕ . The Higgs potential, introduced as $V(\phi)$ in equation 1.1, is given by :

$$V(\phi) = -\mu^2 |\phi|^2 + \lambda |\phi|^4, \quad (1.12)$$

where $\mu, \lambda \in \mathbb{R}$. Initially, one can take ϕ as a simple complex scalar field to illustrate

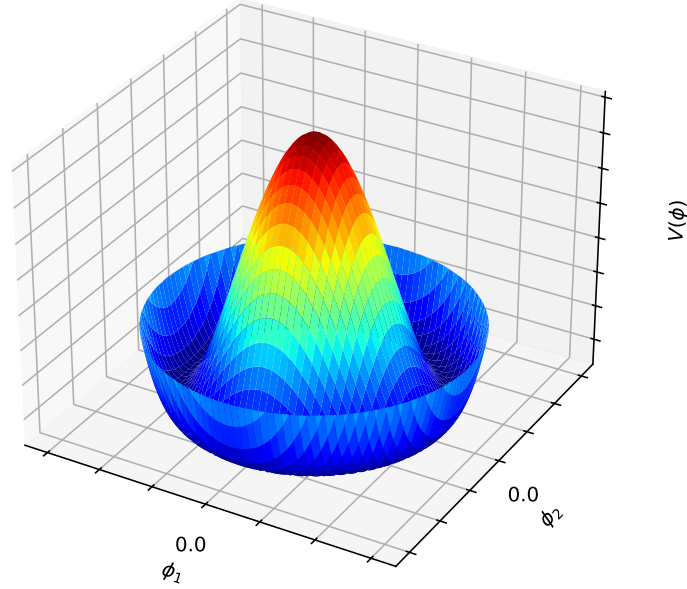


Figure 1.1: Shape of the Higgs potential in the simple case where we define $\phi(x) = \phi_1(x) + i\phi_2(x)$ and $\mu > 0$. The central bump is centred at $\phi = 0$. The values of the potential $V(\phi)$ associated with a colour map, and the fields $\phi_1(x)$ and $\phi_2(x)$, are in arbitrary units.

the Higgs mechanism with a simpler system than the SM, in which case $|\phi|^2 = \phi^\dagger \phi$ and the shape of that potential can be plotted as in Figure 1.1. It becomes clear why the potential is often referred to as the “Sombrero” or “Wine bottle”. The negative sign of the $-\mu^2$ term curves a central maximum (if μ is positive), which results in a continuous range of values that minimises $V(\phi)$. In the complex ϕ_1 - ϕ_2 plane, these values dwell on the circle centred at the origin and of radius $|\phi| = \frac{v}{\sqrt{2}} = \sqrt{\frac{\mu^2}{2\lambda}}$. At $\phi = 0$, the system has a $U(1)$ symmetry from $V(\phi)$: upon defining a continuous real parameter θ , one can scan every direction with $e^{i\theta}$ and observe the potential to be exactly the same. However, the vacuum state of the field lies within the minimum circle, taking the following form: $\langle \phi \rangle = \frac{v}{\sqrt{2}} e^{i\theta}$, thus breaking this $U(1)$ symmetry. This concept, where the Lagrangian holds an intrinsic symmetry that the ground state does not share, is called “Spontaneous

Symmetry Breaking” (SSB). The Lagrangian for the dynamics of ϕ is given by :

$$\mathcal{L} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - V(\phi), \quad (1.13)$$

where we simply added the kinetic term for ϕ . It is useful to expand the field ϕ around a minimum to gain understanding of what the SSB provides:

$$\phi(x) \equiv \frac{1}{\sqrt{2}}(v + \eta(x) + i\chi(x)), \quad (1.14)$$

where we set $\theta = 0$ (one can generalise with θ by mixing the fields $\eta(x)$ and $\chi(x)$ if needed) and the ground states $\langle \eta \rangle$ and $\langle \chi \rangle$ must be zero by definition. By plugging in the expansion into the Lagrangian of equation 1.13, one obtains the following:

$$\mathcal{L} = (\partial_\mu \eta)(\partial^\mu \eta) + (\partial_\mu \chi)(\partial^\mu \chi) - \lambda v^2 \eta^2 - \lambda v \eta(\eta^2 + \chi^2) - \frac{\lambda}{4}(\eta^2 + \chi^2)^2, \quad (1.15)$$

where constant terms have not been written. The fluctuations of $\eta(x)$ are in the direction of the rising sides of the potential (requiring energy to climb uphill) which results in a mass $m_\eta = \sqrt{2\lambda}v$, and can be interpreted as massive bosons (analogous to the Higgs bosons in the SM). The fluctuations of χ are in the direction of the minima circle (travelling on effectively flat potential) which results in a massless boson, referred to as a “Goldstone boson” [14, 15]. So far we discussed the $U(1)$ symmetry as a global symmetry. One can promote the global symmetry to a local symmetry (i.e the phase now depends on spacetime: $\theta \rightarrow \theta(x)$) and couple the scalar field to a gauge field A_μ with coupling strength q by upgrading the differential operator to a covariant derivative $D_\mu = \partial_\mu - iqA_\mu$, exactly as it was discussed with the fermion fields in equation 1.7. The Lagrangian then becomes:

$$\begin{aligned} \mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \mathcal{L}_{\text{Higgs}} \\ &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |D_\mu \phi|^2 - V(\phi), \end{aligned} \quad (1.16)$$

where we also added the kinetic term for the gauge field A_μ . Upon using the same expansion defined in equation 1.14, the following terms appear from the covariant derivative term:

$$|D_\mu \phi|^2 \supset \frac{1}{2}(\partial_\mu \chi)^2 - qvA_\mu \partial^\mu \chi + \frac{1}{2}q^2v^2A_\mu A^\mu. \quad (1.17)$$

One obtains a term that mixes the derivative of the field $\chi(x)$ with A_μ , the gauge field introduced. This makes $\chi(x)$ impossible to interpret as an independent physical field. Instead, one may redefine $B_\mu = A_\mu - \frac{1}{qv}\partial_\mu\chi$ which results in all the terms in expression 1.17 to become simply $\frac{1}{2}q^2v^2B_\mu B^\mu$, a mass term for the field B_μ . This redefinition also leaves the gauge field kinetic term unchanged since one can easily show that the tensors $F_{\mu\nu}$ are the same whether expressed with A_μ or B_μ . The field responsible for the Goldstone boson is thus “absorbed” by the gauge field in our redefinition and the gauge field gains the mass $m_B = qv$.

The Higgs mechanism is a bit more involved in the SM since instead of a $U(1)$ symmetry, the SSB is of the electroweak symmetry $SU(2)_L \otimes U(1)_Y$ into $U(1)_{\text{EM}}$, the electromagnetic symmetry. The Higgs field then interacts with the electroweak interaction as a $SU(2)_L$ doublet and with a hypercharge of $\frac{1}{2}$. Before SSB one can write out the doublet as:

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \quad (1.18)$$

where we define Φ as our SM Higgs field, the fields ϕ^+ and ϕ^0 are the $SU(2)_L$ doublet components of Φ . Electric charge arises from the $U(1)_{\text{EM}}$ symmetry and can be computed from the electroweak charges as $Q = T_3 + Y$. One can compute the electric charge of each field using that definition: $Q(\phi^+) = \frac{1}{2} + \frac{1}{2} = +1$ and $Q(\phi^0) = -\frac{1}{2} + \frac{1}{2} = 0$. After SSB of the electroweak symmetry we can write:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad (1.19)$$

where v is the Vacuum Expectation Value (VEV) of the Higgs, analogous to the parameter with the same name in our $U(1)$ SSB example. Using a similar expansion to our previous example in equation 1.14, we can study the fluctuations of the fields after SSB in the SM:

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} G^+(x) \\ v + h(x) + iG^0(x) \end{pmatrix}, \quad (1.20)$$

where we introduce the fields $G^\pm$ (G^+ and $(G^+)^\dagger$), $h(x)$ and G^0 . Plugging this expression in $\mathcal{L}_{\text{Higgs}}$ results in similar terms to our previous example where $h(x)$ describes massive Higgs boson particles and the Goldstone bosons fields ($G^0, G^\pm$) are massless. It is then

natural for the covariant derivative to couple Φ to the electroweak forces to retrieve gauge invariance:

$$D_\mu = \partial_\mu - ig_1 T^a W_\mu^a - \frac{ig_0}{2} Y B_\mu. \quad (1.21)$$

Plugging the expansion 1.20, considering only the terms associated with $h(x)$, into the covariant derivative term as we did in the simplified example, and using the $SU(2)$ generators identities, one gets the following terms:

$$|D_\mu \Phi|^2 \supset \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{2} \left(\frac{v+h}{2} \right)^2 \left[g_1^2 (W_\mu^1)^2 + g_1^2 (W_\mu^2)^2 + (-g_1 W_\mu^3 + g_0 B_\mu)^2 \right], \quad (1.22)$$

where we expanded the weak fields W_μ^a over their adjoint index. It is clear that the fields W_μ^1 and W_μ^2 are degenerate in mass. One can define $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$, a unique linear combination of the gauge fields that transform as eigenstates of the electromagnetic charge operator and results in the charged-current interactions. This linear combination also makes the ladder operator appear, built out of the $SU(2)$ generators associated with the weak fields: $T^1, T^2, T^3 \rightarrow (T^1 \pm iT^2)T^\pm, T^3$. The ladder operators are responsible for the raising or lowering of the isospin in the charged-current interactions. Then we can diagonalise the rest of the mass matrix (only W_μ^3 and B_μ are left to redefine into mass eigenstates) with the following:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos(\theta_W) & -\sin(\theta_W) \\ \sin(\theta_W) & \cos(\theta_W) \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \quad (1.23)$$

where θ_W is the Weinberg angle which is defined using the coupling constants of the gauge fields : $\sin(\theta_W) = \frac{g_0}{\sqrt{g_0^2 + g_1^2}}$ and $\cos(\theta_W) = \frac{g_1}{\sqrt{g_0^2 + g_1^2}}$. Re-expressing equation 1.22 with the newly defined mass eigenstate fields yields the following terms:

$$|D_\mu \Phi|^2|_{\text{mass}} \supset \frac{1}{4} g_1^2 v^2 W_\mu^+ W^{-\mu} + \frac{1}{8} v^2 (g_1^2 + g_0^2) Z_\mu Z^\mu. \quad (1.24)$$

The mass terms indicate that the $W^\pm$ field has mass $M_W = \frac{1}{2} g_1 v$ and Z_μ has mass $M_Z = \frac{1}{2} \sqrt{g_1^2 + g_0^2} v$. No such term appears for the field A_μ which is then massless. Rewriting the covariant derivative in function of the mass eigenstate fields results in the

following term for A_μ :

$$D_\mu|_{\text{mass}} \supset -i \frac{g_1 g_0}{\sqrt{g_1^2 + g_0^2}} A_\mu (T^3 + Y). \quad (1.25)$$

Here the electromagnetic charge operator Q is recognised, one can identify A_μ as the electromagnetic gauge field and the electromagnetic coupling is then given by:

$$e = \frac{g_1 g_0}{\sqrt{g_1^2 + g_0^2}}. \quad (1.26)$$

After SSB of the electroweak symmetry, one is left with a set of more familiar fields in experimental particle physics: the $W^\pm$ and Z bosons of the weak force as well as the massless photon, with field A_μ , as the boson of the electromagnetic force. The expansion around the vacuum $v \rightarrow v + h(x)$ can also be used in the Higgs potential and the Higgs mass m_h can be extracted from $V(h)$:

$$V(h) = -\mu^2 \frac{(v+h)^2}{2} + \lambda \frac{(v+h)^4}{4} \supset \left(-\frac{\mu^2}{2} + \lambda \frac{3}{2} v^2 \right) h^2. \quad (1.27)$$

One needs to apply the minimisation condition for the vacuum simply given by $\frac{dV}{dh} = 0$, which imply $\mu^2 = \lambda v^2$. Combining this with the relevant term in the expanded $V(h)$ in equation 1.27, one can obtain an expression for the Higgs mass:

$$m_h = 2\lambda v^2, \quad (1.28)$$

where it is reassuring to see that m_h depends on the curvature (steepness) of the potential, the information stored in λ .

The Higgs mechanism is now providing the mass of the weak bosons and the appearance of the photon field but the mass terms of the fermions still needs to be addressed. This is done through the Yukawa Lagrangian term in the SM, shown in equation 1.1. The term, shown as $\mathcal{L}_{\text{Yukawa}}$, is simplified and hides some complexity residing in the difference between the up and down types of quarks. The indices i and j iterate over the three known generations of up quarks, down quarks, and leptons. The Yukawa coupling y_{ij} are free parameters that introduce the observed quark flavour mixing measured in experiments [16] and are encoded in the Cabibbo-Kobayashi-Maskawa matrix [17, 18]. The Higgs field, interacting with the electroweak forces by being a doublet under $SU(2)_L$ and having a hypercharge of $\frac{1}{2}$, makes these terms gauge invariant under $SU(2)_L \otimes U(1)_Y$

by compensating the transformations of fermions. After SSB of the ground state, the Higgs field acquires a VEV as expressed in equation 1.19, and the Yukawa terms become mass terms:

$$y_{ij}\bar{\psi}_L^i\Phi\psi_R^j + h.c \xrightarrow{\text{SSB}} \frac{y_{ij}v}{\sqrt{2}}(\bar{\psi}_L^i\psi_R^j + h.c) = \frac{y_{ij}v}{\sqrt{2}}\bar{\psi}^i\psi^j, \quad (1.29)$$

where the fermionic masses are then retrieved as: $m_{ij} = \frac{y_{ij}v}{\sqrt{2}}$. It is important to note that since there are no right-handed neutrinos or left-handed anti-neutrinos, the Higgs mechanism and Yukawa couplings cannot provide masses for the neutrino particles even though their flavour eigenstates have been observed to oscillate [16, 19, 20], which implies that they have a non-zero mass. The Higgs scalar field, describing the only boson that is not associated with a gauge vector field, is then the keystone that completes the formulation of the SM by providing masses for the fundamental fermion fields (except for neutrinos) and the weak force gauge bosons while preserving the overall $SU(2)_L \otimes U(1)_Y$ gauge symmetry via the Higgs mechanism. A summary of every fundamental field, associated with fundamental particles and anti-particles, is presented in Figure 1.2.

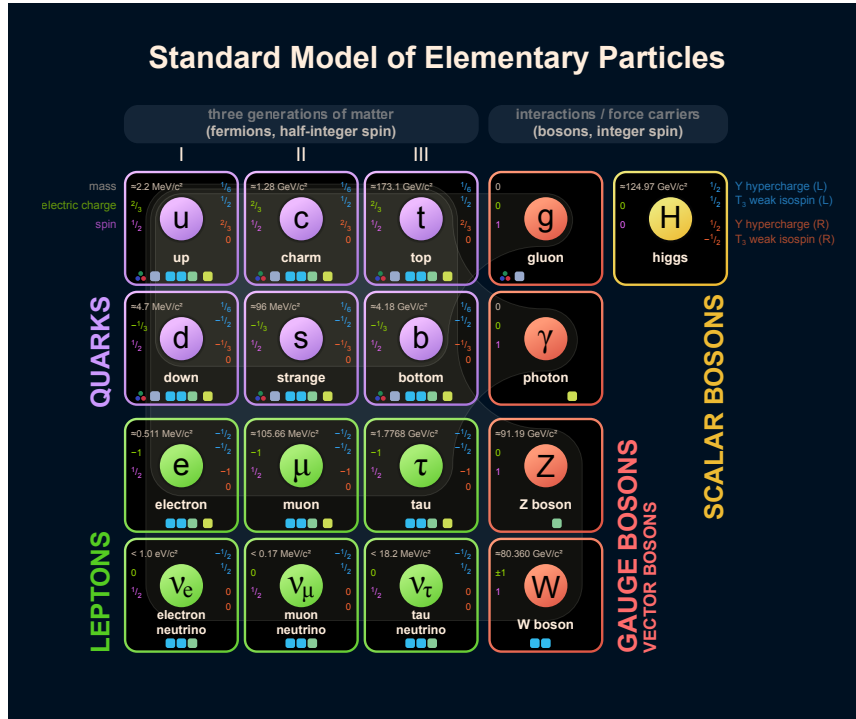


Figure 1.2: The elementary particles of the SM [21]. The coloured squares, displayed below each fermion, represent the gauge boson fields to which it couples. The grey square and coloured circles represent the gluon field. The two blue squares represents the W boson fields, the green square the Z boson field, and the yellow square the photon field.

1.1.2 W boson Higgs coupling and spin

The Higgs field couples to the $W^\pm$ fields via the application of the covariant derivative shown in equation 1.21. The expansion performed in equation 1.22 reveals that the Higgs boson h couples with $W^\pm$ as follows:

$$\mathcal{L}_{hWW} = 2\frac{h}{v} (m_W^2 W_\mu^+ W^{-\mu}) + \left(\frac{h}{v}\right)^2 (m_W^2 W_\mu^+ W^{-\mu}), \quad (1.30)$$

where we used $g_1 = \frac{2m_W}{v}$ to substitute the g_1 parameter for the W mass. Using $v = \frac{2m_W \sin(\theta_W)}{e}$, where e is the electromagnetic coupling defined in equation 1.26, and considering the first term only, the Feynman vertex rule for a Higgs boson interacting with two W bosons is given by the following expression, as discussed in Chapter 29, page 587 of [2]:

$$\begin{array}{c}
 W^+ \\
 \diagup \\
 H \text{ -----} \diagdown \\
 W^-
 \end{array}
 = i \frac{e}{\sin \theta_W} m_W g^{\mu\nu}. \quad (1.31)$$

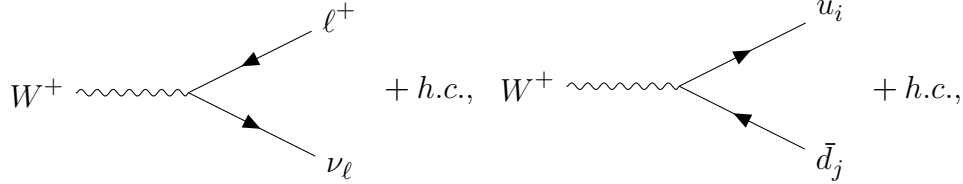
The tensor $g^{\mu\nu}$ is the metric tensor and combines with other tensors resulting from the W decays to form Lorentz scalars that can then be reduced to compute matrix elements, and thus cross sections. The second term in equation 1.30 describes a vertex with two Higgs and two W bosons interacting, which is not relevant for the work presented in this thesis due to its relatively small cross section.

The Lagrangian terms involved in the $W^\pm$ decays to fermions are of the following form:

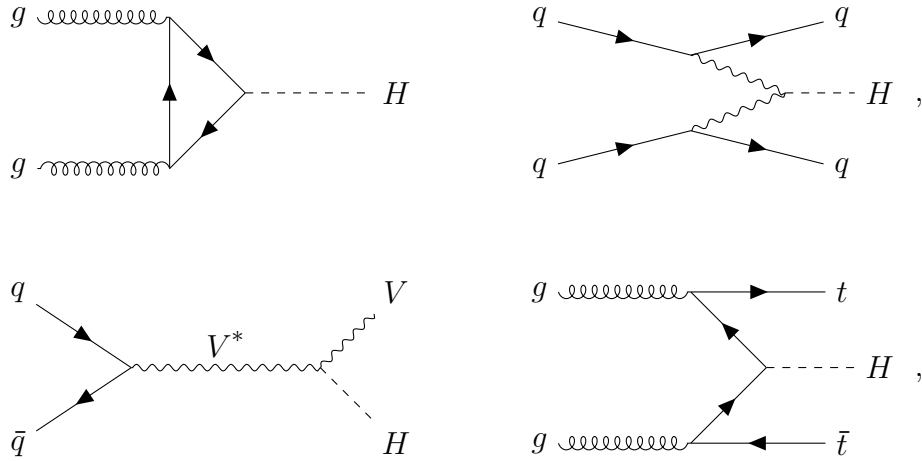
$$\mathcal{L}_{CC} = -\frac{g_1}{\sqrt{2}} \left[\bar{u}_i \gamma^\mu \frac{1-\gamma^5}{2} V^{ij} d_j + \bar{\nu}_\ell \gamma^\mu \frac{1-\gamma^5}{2} \ell \right] W_\mu^+ + h.c., \quad (1.32)$$

where V^{ij} is the CKM matrix, discussed with the Yukawa sector $\mathcal{L}_{\text{Yukawa}}$, that mixes the up quarks u_i and down quarks d_j flavours. The charged leptons l are combined with their neutrinos ν_ℓ . The γ^5 matrix is used to define the projector operators $P_{L,R} = \frac{1 \mp \gamma^5}{2}$ to project the relevant chirality that interacts as $SU(2)_L$ doublets. The Dirac matrices γ^μ are the Lorentz tensors that will get affected by the action of the metric of the hWW vertex described by equation 1.31, and thus connect the W decay diagrams to the Higgs

interaction vertex diagram. The W boson decay can be represented as Feynman diagrams explicitly:



where the *h.c.* (hermitian conjugate) of the first diagram describes $W^- \rightarrow l^- \bar{\nu}_l$ and the *h.c.* of the second diagram describes $W^- \rightarrow \bar{u}_i d_j$. In proton-proton collisions, the two main modes of production for the Higgs are gluon-gluon Fusion (ggF) and Vector Boson Fusion (VBF), followed by the VH and ttH processes. A Feynmann diagram per process is provided as follows:



where V stands for vector boson and can be either $W^\pm$ or Z . In the ggF process (top left), gluons contained in the protons interact via a loop of fermions, mostly the top quark since it has the greatest coupling to the Higgs boson (the heaviest mass) of all fermions. In VBF (top right), quarks contained in the protons scatter weakly by emitting $W^\pm$ or Z bosons that fuse in a vertex (with the Feynman rule given by equation 1.31 when the bosons are W bosons). The other production processes, VH (bottom left) and ttH (bottom right), are provided for completeness, but not considered in detail due to their much lower cross-sections. Figure 1.3 shows the cross section of different Higgs production processes at the LHC against the centre-of-mass energy. The blue line is for ggF and the red line for VBF. It can be seen that the VBF's cross section is roughly about 10 percent of the ggF cross section over the energy range depicted.

The three following processes have been explored as Feynman diagrams: the Higgs production via ggF or VBF, the Higgs decay into two W bosons and the W boson decay to fermions. These three components can be combined together to form the global process $pp \rightarrow H \rightarrow WW^* \rightarrow \ell\nu qq$, which is studied in the analysis presented in chapter 6.

The fundamental parameters described in this section are determined experimentally, either directly or through relations between measured quantities [16]. A selection of relevant parameters and their values is given below:

$$\begin{aligned} m_h &= 125.10 \pm 0.07(\text{stat.}) \pm 0.06(\text{syst.}) \text{ GeV} \\ \alpha^{-1} &= 137.035999084(21) \\ m_W &= 80.3692(133) \text{ GeV} \\ \sin^2(\theta_W) &= 0.22348(10) \\ v &= 246.22 \text{ GeV} \end{aligned}$$

The couplings can be derived from the measured quantities:

$$\begin{aligned} e &= \sqrt{4\pi\alpha} \\ g_0 &= \frac{e}{\cos(\theta_W)} \\ g_1 &= \frac{e}{\sin(\theta_W)} \end{aligned}$$

Considering the measured values for the Higgs mass and the W mass, it is clear that when a Higgs boson decays to two W bosons, as in equation 1.31, at least one of the W has to be off-shell since $m_h < 2m_W (= 160.7384(266) \text{ GeV})$. The branching ratio for the Higgs decaying to two W bosons has been measured to be $\text{BR}(H \rightarrow WW^*) = (25.7 \pm 2.5)\%$ [16]. One can estimate the branching ratio of the W boson using lepton universality (and ignoring the small CKM matrix mixing effects): three lepton-neutrino doublets and two quark doublets (top quarks are too massive to be produced in the decay) times three possible colour charges yields $\text{BR}(W \rightarrow \ell\nu) \approx \frac{3}{3+6} \approx 0.33$ and $\text{BR}(W \rightarrow qq) \approx \frac{6}{3+6} \approx 0.66$. The measured branching ratios for the W boson decays are given by [16]:

$$\begin{aligned} \text{BR}(W \rightarrow \ell\nu) &= (10.86 \pm 0.09)\%, \\ \text{BR}(W \rightarrow qq) &= (67.41 \pm 0.27)\%, \end{aligned}$$

where the branching ratios have been extracted assuming lepton universality, and the branching ratio to leptons is given per individual lepton ($\ell = e, \mu$ or τ). The SM prediction for the total Higgs production cross section at $\sqrt{s} = 13$ TeV is given by $\sigma(pp \rightarrow H) = 55.6 \pm 2.5$ pb [22]. Combining these together yields a prediction for the cross section of the $H \rightarrow WW^*$ decaying semi-leptonically in proton-proton collision: $\sigma(H \rightarrow WW^* \rightarrow \ell\nu qq) = 55.6 \text{ pb} \cdot 0.257 \cdot 2 \cdot ([3 \cdot 0.1086] \cdot 0.6741) \approx 6.28$ pb.

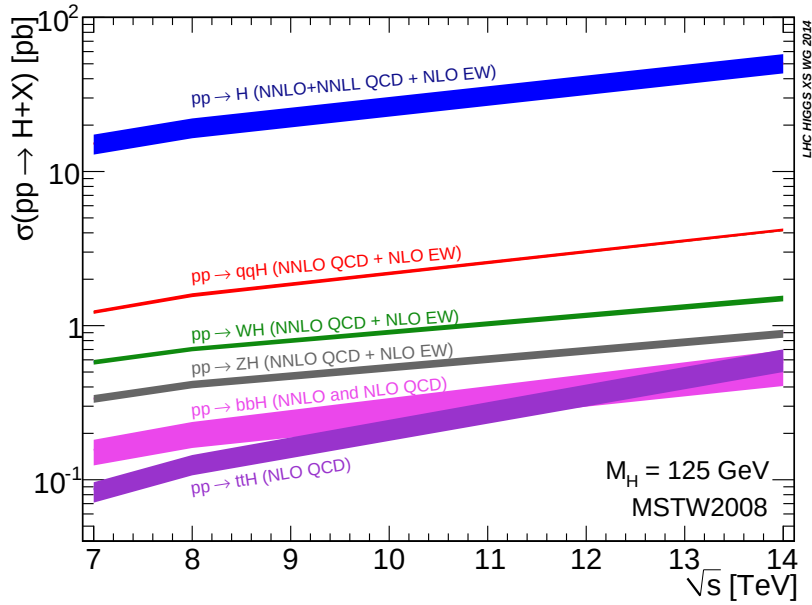


Figure 1.3: Cross section for different Higgs production processes in proton-proton collision against centre-of-mass energy [23].

The spin state of the W boson is accessible from decays occurring via the interaction described in equation 1.32. A solid theoretical framework [24–26] has been studied and improved over the years to measure the spin of the W . The spin state of the weak boson can fully be described as a 3×3 matrix (since the particle is spin-1), the density operator (definition given by equations 1.42 and 1.43) $\rho_{mm'}$ [24]:

$$\rho = \frac{1}{3}\mathbb{I} + \frac{1}{2} \sum_{M=-1}^1 \langle S_M \rangle^* S_M + \sum_{M=-2}^2 \langle T_M \rangle^* T_M, \quad (1.33)$$

where the tensors S_M are in the spherical basis and are made from the spin operators S_i : $S_{\pm 1} = \mp \frac{1}{\sqrt{2}}(S_1 \pm iS_2)$, $S_0 = S_3$.

The tensors T_M are also built from the spin operators S_M :

$$\begin{aligned} T_{\pm 2} &= S_{\pm 1}^2, \\ T_{\pm 1} &= \frac{1}{\sqrt{2}} \{S_{\pm}, S_0\}, \\ T_0 &= \frac{1}{\sqrt{6}} (\{S_{+1}, S_{-1}\} + 2S_0^2), \end{aligned} \quad (1.34)$$

where the brackets $\{, \}$ denote the anti-commutator. When taking the decay to a lepton-neutrino doublet, $W^\pm \rightarrow l^\pm \nu$, the fully differential decay width of the W boson may be related to the density operator as follows:

$$\frac{d\Gamma}{d\cos(\theta^*)d\phi^*} \propto \sum_{m,m'} e^{i(m-m')\phi^*} d_m^1(\theta^*) d_{m'}^1(\theta^*), \quad (1.35)$$

where the angles θ^* and ϕ^* are the polar and azimuthal angles of the charged lepton respectively, and $d_m^j(\theta^*)$ are real functions that depend on the helicity of the daughter particles and whether the W boson is a W^+ or W^- . Equation 1.35 is quite rich, decomposing the $W^\pm \rightarrow l^\pm \nu$ decay into nine observables depending on (θ^*, ϕ^*) that can be used to fully reconstruct the W boson spin state. If one integrates over the azimuthal angle ϕ^* , the exponential term removes the off-diagonal elements and thus three terms are left:

$$\frac{1}{\Gamma} \frac{d\Gamma}{d\cos(\theta^*)} = \frac{3}{8} F_+ (1 + \cos(\theta^*))^2 + \frac{3}{4} F_0 \sin(\theta^*)^2 + \frac{3}{8} F_- (1 - \cos(\theta^*))^2, \quad (1.36)$$

where $F_+ = \rho_{11}$, $F_- = \rho_{-1-1}$, $F_0 = \rho_{00}$ are called the W polarisation fractions, which have been extensively studied in electroweak physics [25, 26] and have been measured a few times in top decays [27–31]. They are related by the normalisation $F_+ + F_- + F_0 = 1$, and each factor is associated with a helicity state for the W boson. Helicity is defined as the projection of the particle spin onto its momentum: $h = \frac{\vec{\sigma} \cdot \vec{p}}{|\vec{p}|}$. There are thus three possible helicity states for vector bosons:

- The spin and momentum are aligned, $h = +1$, and the vector boson is *right-handed*. (Associated with fraction F_+)
- The spin and momentum are anti-aligned, $h = -1$, and the vector boson is *left-handed*. (Associated with fraction F_-)
- The spin and momentum are orthogonal, $h = 0$, and the vector boson is *longitudinal*. (Associated with fraction F_0)

In the SM, the chiral projection in equation 1.32 imposes strong restrictions on the helicity of the W boson in the weak decay of fermions. For example, in the decay $t \rightarrow W^+ b$, the W boson cannot be right-handed ($F_+ = 0$) at tree level. The expression in equation 1.36 can be rewritten as a function of the spin operators S_0 and T_0 introduced earlier:

$$\frac{1}{\Gamma} \frac{d\Gamma}{d\cos(\theta^*)} = \frac{1}{2} + \frac{3}{4} \langle S_0 \rangle \cos(\theta^*) + \frac{3}{4\sqrt{6}} \langle T_0 \rangle (3 \cos(\theta^*)^2 - 1), \quad (1.37)$$

where a two-parameter fit would be enough to characterise the distribution of the decay and measure the two related values $\langle S_0 \rangle$ and $\langle T_0 \rangle$. Equations 1.36 and 1.37 link well-known electroweak decay theory to the concept of quantum tomography [32]. In quantum tomography, each component of ρ is measured, the measurements are thus extended to more distributions that also capture the off-diagonal terms of ρ in where new physics could be hiding. All the expressions shown here are for $W^\pm \rightarrow l^\pm \nu$ at tree-level, this is especially relevant since the charged leptons are known to be good spin analysers [33, 34]: daughter particles whose angular distribution retain spin information about their parent W boson after averaging over helicities. At first glance, it is therefore not trivial to apply these equations and access the W boson's spin when it decays hadronically. Fortunately, it can be shown that the down-type quarks are also excellent spin analysers, as done in [35] with the top quark decay. In that context, the down-type quarks are as good as the charged leptons at tree-level, and very similar at NLO. Provided one can identify and reconstruct the down-type quark from the hadronic W decay, the W spin is available to the same extent that it is for the leptonic decay. This fact enables the $pp \rightarrow H \rightarrow WW^* \rightarrow l\nu qq$ process to be used to measure the spin of the WW^* system, and thus to be sensitive to quantum entanglement of the two W bosons. The basic framework to define quantum entanglement is introduced in section 1.2.1.

1.2 Quantum Information

Quantum Information (QI) has become one of the most important fields of physics in the recent years due to advancement of technology in Quantum Computing and its well-known, regularly advertised, potential. QI measurements typically are performed in experiments designed to reach the lowest possible energies to avoid decoherence (loss of quantum information to the environment). Within the SM framework, it is possible to probe QI at the complete opposite energy regime accessible by humanity using LHC 13 and 13.6 TeV proton-proton collisions that do not suffer from quantum decoherence. Assuming that

Quantum Mechanics (QM) is emergent rather than fundamental, it is conceivable that undiscovered quantum mechanical effects are measurable only at relativistic energies, as scale-dependent new physics that would allow QM to be reformulated or corrected as a physical theory.

Erwin Schrödinger supposedly said: “Entanglement is not one but rather the characteristic trait of Quantum Mechanics.” Quantum entanglement results in correlation between measurements of subsystems described by a single wavefunction in QM such that it cannot be described by a classical theory. The existence of such a feature implies interesting and unintuitive consequences. For instance, the well known Einstein-Podolsky-Rosen (EPR) 1935 paper [36] highlights the non-locality of composite quantum systems: measuring a subsystem entangled with other subsystems causes every correlated quantum state to change instantly regardless of their position in space with respect to the position of the measurement. This feature of QM, famously described by Einstein as a “spooky action at a distance”, led to the proposal that QM is incomplete as a theory and the postulate of local hidden variables that resolve its non-local aspect. The debate about the completeness of QM seemed to be of a philosophical nature: one could come up with infinitely many theories that carry those local hidden variables and which can reproduce the observed experimental results. Then, in 1964, John Bell formulated the Bell Theorem [37], a framework that provides the formulation of a test that can reject the set of all theories containing local hidden variables, and thus the incompleteness of QM. Experiments were famously designed in 1972 (John Clauser and Stuart Freedman [38]) and 1982 (Alain Aspect, Jean Dalibard and Gérard Roger [39]) to measure a reformulation of the Bell inequality, shown in equation 1.53, and found QM to be complete, rejecting the concept of hidden variables and providing a physical answer to the EPR paradox. These results eventually led to John Clauser and Alain Aspect receiving the Nobel prize in 2022. There are still many subtle issues with QM [40–44] and a superseding theory solving or explaining these weaknesses is yet to be formulated. An interesting way to look for a path towards that new theory is to study QM as an axiom of the SM. Furthermore, QI observables can be used to search for beyond the SM physics [45–47].

This section aims at introducing the basics of QI to understand the related literature discussing measurements in proton-proton collision events where a Higgs boson decays into two W bosons, such as the one proposed in the phenomenology study presented in chapter 5.

1.2.1 Quantum Entanglement Theory

In order to discuss quantum entanglement, we must first introduce the concept of a qubit state (two-level quantum mechanical system):

$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (1.38)$$

Such a quantum state $|\phi\rangle$ can be expressed as a superposition of the orthogonal states $|0\rangle$ and $|1\rangle$ forming our qubit basis (α, β being complex numbers called amplitudes, that follow the probability normalisation: $|\alpha|^2 + |\beta|^2 = 1$). Pairs of qubits can then be represented as the tensor product of two qubit states:

$$|\psi\rangle = |\phi_A\rangle \otimes |\phi_B\rangle, \quad (1.39)$$

where the qubits are treated as two subsystems A and B . Working with two qubits, one can define the four Bell states:

$$|\Psi^\pm\rangle = \frac{1}{\sqrt{2}} (|0\rangle \otimes |1\rangle \pm |1\rangle \otimes |0\rangle), \quad (1.40)$$

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}} (|0\rangle \otimes |0\rangle \pm |1\rangle \otimes |1\rangle). \quad (1.41)$$

These states are maximally entangled and hence are maximally correlated: measuring a single qubit lets us know the result for both qubits provided the system is known to have been in a Bell state before the measurement occurs.

Using the familiar formulation of the wavefunction $|\psi\rangle$ can be cumbersome and unintuitive when expressing the entanglement of subsystems. The probability density operator $\hat{\rho}$ provides a more practical framework to define the concept of pure and mixed states, which is itself important to define entanglement. The density matrix for a pure state is defined as the outer product of a pure state wavefunction with itself:

$$\hat{\rho}_{\text{pure}} = |\psi\rangle\langle\psi|, \quad (1.42)$$

whereas a mixed state is given as the sum of multiple pure states weighted by probabilities p_i , with $p_i \in [0, 1]$ and $\sum_i p_i = 1$:

$$\hat{\rho}_{\text{mixed}} = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (1.43)$$

The formulation of mixed state given by the density operator in equation 1.43 is useful to illustrate the confusion that may arise between the concepts of superposition and state mixing. Any pure state $|\psi\rangle$ can always be expressed as a superposition of multiple states $\{ |\phi_i\rangle \}$, as long as the latter set forms an eigenstate basis. Superposition of states is a purely quantum mechanical aspect that naturally occurs from the vector space framework of the theory. Mixed states occur at a higher level: when the quantum system studied can be given by a statistical mixture of multiple different pure states, labelled i , with their corresponding classical probabilities p_i . The density matrix $\hat{\rho}$ can provide the same information contained by the wavefunction:

- Schrödinger time evolution: $i\hbar \frac{d\hat{\rho}}{dt} = [\hat{H}, \hat{\rho}]$, where $[,]$ are the familiar commutator's Poisson brackets and $\hat{H}$ the system's Hamiltonian.
- The expectation value for an operator $\hat{A}$ is given by $\langle \hat{A} \rangle = \text{Tr}(\hat{\rho}\hat{A})$.
- The overlap between the quantum state with another state $|\phi\rangle$ can be obtained as $\langle \phi | \hat{\rho} | \phi \rangle$.
- The quantum state probability normalisation and unitarity is ensured through $\text{Tr}(\hat{\rho}) = 1$ and $\hat{\rho}^\dagger = \hat{\rho}$.

Squaring the density matrix is interesting to differentiate between pure and mixed states; pure states are idempotent $\hat{\rho}_{\text{pure}}^2 = \hat{\rho}$ and thus $\text{Tr}(\hat{\rho}_{\text{pure}}^2) = 1$, whereas for mixed states $\hat{\rho}_{\text{mixed}}^2 \neq \hat{\rho}$ and $\text{Tr}(\hat{\rho}_{\text{mixed}}^2) \leq 1$. One can then define quantum entanglement based on the separability of the quantum state into pure states. A separable (non-entangled) state of two subsystem A and B can be expressed as:

$$\hat{\rho}_{\text{separable}} = \sum_i p_i \hat{\rho}_A^i \otimes \hat{\rho}_B^i. \quad (1.44)$$

Such a state exhibits classical correlations only between subsystems A and B . A more general expression is needed to express a non-separable (entangled) state:

$$\begin{aligned} \hat{\rho}_{\text{non-separable}} &= \sum_{i,j,k,l} \beta_{ij} \beta_{kl}^* \rho_A^{ik} \otimes \rho_B^{jl} \\ &= \sum_{i,j,k,l} \beta_{ij} \beta_{kl}^* |\psi_i\rangle \langle \psi_k|_A \otimes |\phi_j\rangle \langle \phi_l|_B \end{aligned} \quad (1.45)$$

where we expressed explicitly the subsystems A and B using eigenstates $\{ |\psi_i\rangle \}$ and $\{ |\phi_i\rangle \}$ respectively, β_{ij} are wavefunction amplitudes, and the iteration is done over four degrees of freedom instead of one associated with the classical mixing probabilities.

One can then use the definition of partial trace on a separable system:

$$\text{Tr}_B(\hat{\rho}_A \otimes \hat{\rho}_B) = \sum_f \hat{\rho}_A \otimes \langle \phi_f | \hat{\rho}_B | \phi_f \rangle = \hat{\rho}_A, \quad (1.46)$$

and on a non-separable state:

$$\begin{aligned} \text{Tr}_B \left(\sum_{i,j,k,l} \beta_{ij} \beta_{kl}^* |\psi_i\rangle \langle \psi_k|_A \otimes |\phi_j\rangle \langle \phi_l|_B \right) &= \sum_f \sum_{i,j,k,l} \beta_{ij} \beta_{kl}^* |\psi_i\rangle \langle \psi_k|_A \otimes \langle \phi_f | \phi_j \rangle \langle \phi_l | \phi_f \rangle_B \\ &= \sum_{i,j,k} \beta_{ij} \beta_{kj}^* |\psi_i\rangle \langle \psi_k|_A \\ &= \sum_{i,j,k} p_{ijk} \rho_A^{ik}, \end{aligned} \quad (1.47)$$

where the system is reduced with the inner product constraint $j = l (= f)$, and the last line highlights the amplitudes combining to produce probabilities p_{ijk} such that the expression can be related to equation 1.43 explicitly. Observing the effect of the partial trace integrating out the subsystem B , it can be seen that for a separable system one ends up with a pure state, whereas for a non-separable system one ends up with a mixed state. Two systems A and B are then entangled if, and only if, their reduced states are mixed (i.e $\text{Tr}(\hat{\rho}_A^2) \leq 1$ and $\hat{\rho}_B^2 \leq 1$). This result is called the Peres-Horodecki criterion [48, 49], and is crucial in the derivation of the observables used to measure entanglement.

1.2.2 Bell's theorem and Bell operators

A simple bipartite system, as described in 1.2.1, is considered in this subsection. Two measurements, that each act on a different subsystem, are defined as $A(\vec{a})$ and $B(\vec{b})$, where $\vec{a}$ and $\vec{b}$ are two adjustable parameters. Working with this setup, it is possible to measure the correlation function between both measurements, defined as:

$$E(\vec{a}, \vec{b}) = \langle AB \rangle. \quad (1.48)$$

If we are interested in the measurement of spin- $\frac{1}{2}$ particles then we only have two outcomes to consider: spin up, where the measurement yields an angular momentum of $\frac{\hbar}{2}$, and spin down, where the measurement yields an angular momentum of $-\frac{\hbar}{2}$. In that context, our parameters $\vec{a}$ and $\vec{b}$ can be the spatial axis along which spin is measured. The outcomes can be redefined to be $+1$ or -1 , in which case $-1 \leq E(\vec{a}, \vec{b}) \leq 1$.

In order to enforce locality (measurements and systems can only have an effect on their immediate surroundings) and realism (measurement results are predetermined by hidden variable(s) and not purely probabilistic) when formulating Bell's theorem, the correlation function are expressed using the concept of probability function density and the hidden parameter λ :

$$\int d\lambda P(\lambda) [P(A = +1|\lambda) - P(A = -1|\lambda)] \cdot [P(B = +1|\lambda) - P(B = -1|\lambda)] \quad (1.49)$$

where $P(\lambda)$ is the probability density for the hidden variable λ . The notation can be simplified as follows:

$$E(\vec{a}, \vec{b}) = \int d\lambda P(\lambda) A(\vec{a}, \lambda) B(\vec{b}, \lambda). \quad (1.50)$$

The resulting equation describes the measurements with their separate measurement parameter choices, $\vec{a}$ and $\vec{b}$, and their dependence on the hidden variable λ . The core of Bell's theorem is then obtainable through a few algebraic manipulations allowed by the assumptions used so far:

$$\begin{aligned} E(\vec{a}, \vec{b}) - E(\vec{a}, \vec{b}') &= \int d\lambda P(\lambda) [A(\vec{a}, \lambda) B(\vec{b}, \lambda) - A(\vec{a}, \lambda) B(\vec{b}', \lambda)] \\ &= \int d\lambda P(\lambda) [A(\vec{a}, \lambda) B(\vec{b}, \lambda) (1 \pm A(\vec{a}', \lambda) B(\vec{b}', \lambda))] \\ &\quad - \int d\lambda P(\lambda) [A(\vec{a}, \lambda) B(\vec{b}', \lambda) (1 \pm A(\vec{a}', \lambda) B(\vec{b}, \lambda))], \end{aligned} \quad (1.51)$$

where the new free parameters $\vec{a}'$ and $\vec{b}'$ are introduced for the measurements A and B respectively. Keeping the example of spin measurement, these new parameters are then additional spatial axes along which the spin is measured. There are now four possible correlations one can actively measure with this setup: $E(\vec{a}, \vec{b})$, $E(\vec{a}', \vec{b})$, $E(\vec{a}, \vec{b}')$ and $E(\vec{a}', \vec{b}')$. One can show that equation (1.51) holds by expanding the terms composing the two last lines. Since in this case the outcome of both measurement can only be ± 1 , one can write:

$$\left| E(\vec{a}, \vec{b}) - E(\vec{a}, \vec{b}') \right| \leq \int d\lambda P(\lambda) \left[(1 \pm A(\vec{a}', \lambda) B(\vec{b}', \lambda)) \right] - \int d\lambda P(\lambda) \left[(1 \pm A(\vec{a}', \lambda) B(\vec{b}, \lambda)) \right]. \quad (1.52)$$

Identifying the other correlation functions and rearranging yields the following:

$$\left| E(\vec{a}, \vec{b}) - E(\vec{a}, \vec{b}') \right| + \left| E(\vec{a}', \vec{b}') + E(\vec{a}', \vec{b}) \right| \leq 2. \quad (1.53)$$

Equation (1.53) is also known as the Clauser-Horne-Shimony-Holt (CHSH) inequality and can be shown to be violated by QM (but is importantly respected by all hidden variable theories following our assumptions). The maximum value that the Left-Hand Side (LHS) of equation (1.53) can take in QM is $2\sqrt{2}$, known as the “Tsirelson’s bound” [50]. Interestingly, it has been shown that it is possible to write theories, assuming non-locality as an axiom, that predict correlations that exceed this bound up to 4 while still preserving relativistic causality [51]. One might question whether it is possible to create reasonable theoretical models that allow the experimental observation of correlations that violate the $2\sqrt{2}$ boundary, and thus reject QM. The implications of the violation of these boundaries in Bell tests are very profound: a whole class of infinitely many theories containing hidden variables is rejected at once in favour of the fundamental axioms of QM. This is very different from how theories are typically rejected in particle physics, where often the suppression or excess of a physical process is sought after at a specified energy range, corresponding to potential new particles and fields. These additional fields may never be fundamentally rejected for the entire physically accessible energy range, whereas the Bell tests determine elementary properties of the Universe that must be true for all theories. This can be compared with the concept of causality, which provides a similarly powerful constraint.

In this work, the interest is in measuring the entanglement of two W bosons resulting from the Higgs decay. The W boson has a spin of 1, meaning that the system is composed of qutrits and not qubits, like those introduced in equation 1.38. In other words, the bosons have three outcomes ($|0\rangle, |1\rangle, |2\rangle$, or $|-\rangle, |0\rangle, |+\rangle$, the latter being the helicity eigenstates, more relevant to particle physics) upon spin measurements instead of two. It can be shown that the Collins-Gisin-Linden-Massar-Popesc (CGLMP) inequality is optimal to test the entanglement of such systems [52, 53]. It is important to note that this reformulation of Bell’s inequality has already been measured experimentally [54–56]. The CGLMP inequality observable can be computed by taking the expectation value of the three (xy , yz and zx) Bell operators. Using the Gell-Mann matrices [57] in the standard representation for qutrit systems, the xy Bell operator is given as [58]:

$$\mathcal{B}_{CGLMP}^{xy} = -\frac{2}{\sqrt{3}} (S_x \otimes S_x + S_y \otimes S_y) + \lambda_4 \otimes \lambda_4 + \lambda_5 \otimes \lambda_5 \quad (1.54)$$

where, in this representation, $S_x = \frac{1}{\sqrt{2}}(\lambda_1 + \lambda_6)$, $S_y = \frac{1}{\sqrt{2}}(\lambda_2 + \lambda_7)$ and λ_i are the Gell-Mann matrices, not to be confused with the hidden variable introduced earlier. The CGLMP inequality observable $\mathcal{I}_3$ is then given by:

$$\mathcal{I}_3 = \max(\langle \mathcal{B}_{CGLMP}^{xy} \rangle, \langle \mathcal{B}_{CGLMP}^{yz} \rangle, \langle \mathcal{B}_{CGLMP}^{zx} \rangle), \quad (1.55)$$

such that our inequality takes the simple form of $\mathcal{I}_3 \leq 2$.

1.2.3 Measuring Quantum Entanglement in $H \rightarrow WW^*$ at colliders

The power of Bell's theorem resides in the rejection of locality and realism of the hidden variable theories [37]. Experiments that attempt to perform a Bell test and reject this set of theories must rely on measurements with independent and freely chosen parameters, like A and B described in section 1.2.2. This is not the case in a high-energy collider setting where the momentum of decay products are observed statistically to produce distributions in differential measurements. Furthermore, it has been shown that some local hidden variable models can reproduce cross sections and angular distributions measured assuming quantum theory [59]. From these observations, it is clear that one may not perform a Bell test in unpolarised collider experiments. However, the increased spin correlation between particles occurring due to the non-separability (entanglement) of their quantum state can be measured. ATLAS [60] and CMS [61] both measured the presence of entanglement using spin correlation observables. The Bell operator is such an observable that is sensitive to entanglement and can be used to measure its existence in a spin system.

It can be shown that the expectation value of the Bell operator described by equation 1.54 can be written as the weighted sum of the average of three observables [58]:

$$\begin{aligned} \langle \mathcal{B}_{CGLMP}^{xy} \rangle = & \frac{8}{\sqrt{3}} \langle \xi_x^+ \xi_x^- + \xi_y^+ \xi_y^- \rangle_{av} \\ & + 25 \langle (\xi_x^+)^2 - (\xi_y^+)^2 + (\xi_x^-)^2 - (\xi_y^-)^2 \rangle_{av} \\ & + 100 \langle \xi_x^+ \xi_y^+ \xi_x^- \xi_y^- \rangle_{av} \end{aligned} \quad (1.56)$$

where $\xi_i^\pm$ is the dot product between the direction of the spin analyser daughter particle of the $W^\pm$ with the axis i (where i can be x, y or z) in the rest frame of the W boson. As described in [58], the axes can be defined in the WW^* centre-of-mass frame by using

the direction of one of the beams $\hat{\mathbf{p}}$, and the considered W 's momentum direction $\hat{\mathbf{k}}$ as follows:

$$\hat{\mathbf{k}}, \hat{\mathbf{r}} = \frac{1}{r}(\hat{\mathbf{p}} - y\hat{\mathbf{k}}), \hat{\mathbf{n}} = \frac{1}{r}(\hat{\mathbf{p}} \times \hat{\mathbf{k}}), \quad (1.57)$$

where $y = \hat{\mathbf{p}} \cdot \hat{\mathbf{k}}$ and $r = \sqrt{1 - y^2}$. This right-handed orthonormal basis $\{\hat{\mathbf{n}}, \hat{\mathbf{r}}, \hat{\mathbf{k}}\}$ is defined in the Higgs boson rest frame. A boost into each W boson rest frame (along direction $\hat{\mathbf{k}}$) is then performed and the final basis is obtained: $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\} = \{\hat{\mathbf{n}}, \hat{\mathbf{r}}, \hat{\mathbf{k}}'\}$, where the two first vectors remained unchanged by the boost and $\hat{\mathbf{k}}'$ is maintained to unit length and parallel to $\hat{\mathbf{k}}$. The spin analyser is defined as the daughter particle from which the spin information of its W boson parent can be extracted. As described in section 1.1.2, leptons and down/strange quarks have a high spin analysing power and thus may be used to perform the measurement of the $\xi_i^\pm$ observables.

The $H \rightarrow WW^*$ channel has already been studied (for other purposes than measuring quantum entanglement) using the dileptonic decay [62], the other decay modes (semi-leptonic and fully hadronic), however, suffer from important QCD background and are yet to be explored. The dileptonic decay suggested in [58] is complicated to use for our purpose due to the presence of two neutrinos in the final state. The measurement of the $\xi_i^\pm$'s require boosting into the rest frame of the Higgs, which itself requires the Higgs to be fully reconstructed. Two neutrinos in the final state makes the system under-constrained. Instead, a new approach at measuring the Higgs is proposed: the semi-leptonic decay of $H \rightarrow WW^*$. The phenomenology work, presented in chapter 5, shows that it is possible to suppress the background by imposing certain conditions on the final state, and a simulated Bayesian unfolding of the Bell operator's three observables, shown in equation 1.56, is performed.



2 The ATLAS Detector

2.1 Introduction

The ATLAS (A Toroidal LHC ApparatuS) collaboration involves thousands of scientists and engineers from institutions around the world, making it one of the largest international scientific endeavours in history. The detector is one of the four primary experiments located at a collision point on the 26.7 km Large Hadron Collider (LHC) ring at CERN (Conseil Européen pour la Recherche Nucléaire). The ATLAS detector was built from 2003 to 2008 and is continuously subject to upgrades and repairs over time. The detector is a cylinder with a length of 44m and a diameter of 25m, making it the largest collider detector ever built to date.

2.2 The Large Hadron Collider

The LHC [63] is housed in an underground tunnel, dug between 1984 and 1989 to initially host the Large Electron-Positron collider (LEP), and is the last stage in a complex chain of accelerators. The CERN accelerator infrastructure, and the manner in which the experiments are connected, are shown in Figure 2.1. The work presented here concerns proton-proton collisions, and thus this section focuses on their acceleration.

The proton journey begins bounded to an electron as a hydrogen atom, itself bounded to another hydrogen atom trapped inside a bottle of molecular hydrogen gas at room temperature. The gas is injected into a duoplasmatron [64] that ionises the hydrogen atoms to free the protons and eject them into the first acceleration step towards the LHC ring shown in Figure 2.1: the linear accelerator LINAC2 (replaced by LINAC4 since 2020, which now uses negative hydrogen ions to make the acceleration easier). The latter

brings the proton's energy from 750 keV to 50 MeV [65] (LINAC4 [66] now brings it to 160 MeV) before injecting the beam into the BOOSTER ring, which is a synchrotron that elevates the energy to 1.4 GeV [67] (2 GeV since 2020 [68]). The beam next enters the Proton Synchrotron (PS) [69] and Super Proton Synchrotron (SPS), circular accelerators that push the beam to 25 and 450 GeV, respectively, before finally reaching the LHC where the protons are accelerated to 13 TeV (13.6 TeV since 2020). The LHC is made of superconducting Niobium-Titanium (Nb-Ti) magnets, most of which are dipoles (1232 segments) responsible for bending the trajectory of the particles into the ring using the Lorentz force.

Data acquisition is achieved over multiple years and is split into runs: Run 1 was from 2009 to 2013¹, Run 2 from 2015 to 2018 and, currently, Run 3 started in 2022 and ended in 2026. Between different runs, the LHC and the detectors are upgraded in different ways to withstand higher luminosities and increase the centre-of-mass energy. The work presented here makes use of data collected during Run 2 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and Run 3 with 13.6 TeV.

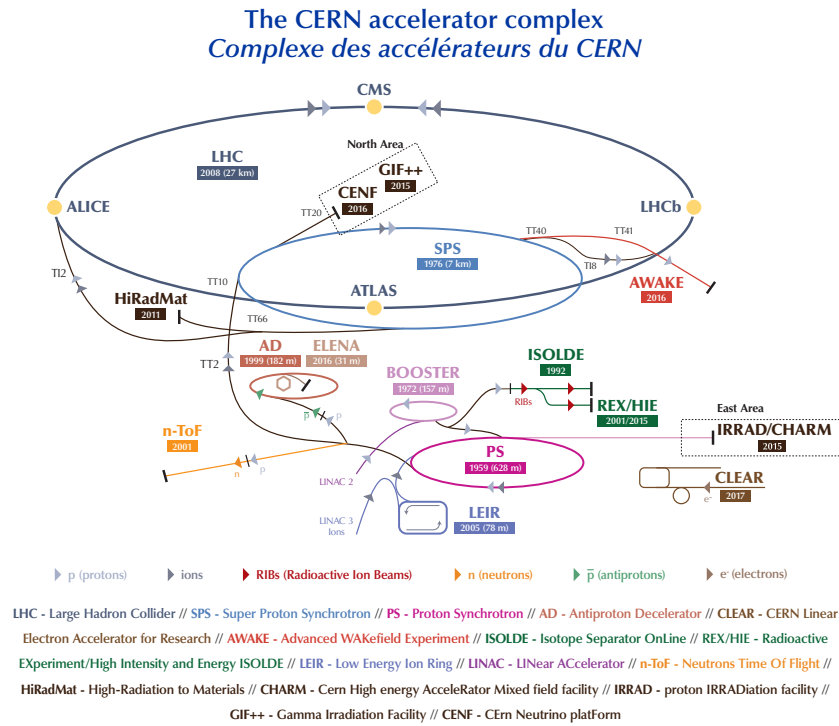


Figure 2.1: Schematic of the accelerators and experiments in CERN [70].

¹Run 1 had collisions ranging between 7 and 8 TeV

The number of collisions N occurring at the LHC is typically computed using the concept of luminosity:

$$N = \int L(t) dt \cdot \sigma, \quad (2.1)$$

where L is the instantaneous luminosity and σ is the cross section of the interaction process of interest. The instantaneous luminosity itself depends on the beam parameters and, assuming that the beams have a Gaussian profile, can be expressed as [63]:

$$L = \frac{N_b^2 n_b f \gamma}{4\pi\epsilon\beta^*} F, \quad (2.2)$$

where N_b is the number of protons in a bunch², n_b the number of bunches in each beam, f the frequency of ring revolution, γ the relativistic gamma factor, ϵ the normalised transverse beam emittance, β^* the beta function at the interaction point and finally F the reduction factor caused by the beams crossing angle at the interaction point. Assuming a proton energy of 6.5 TeV yields $\gamma = \frac{E}{mc^2} \approx 6930$, which allows us to estimate the frequency based on the magnitude of the velocity of the protons v and the circumference of the LHC, $l = 26.7$ km :

$$f \approx \frac{v}{l} = \frac{c\sqrt{1 - \frac{1}{\gamma^2}}}{l} \approx 11.228 \text{ kHz},$$

which is not far from the actual value estimated at 11.245 kHz [71], the difference explained by the travelled length l being slightly overestimated. The beta function is a concept of accelerator physics that describes the transverse size of the beam [72]. The mean number of interactions per bunch crossing $\langle\mu\rangle$, averaged over all pairs of bunches colliding, quantifies the amount of additional soft interactions that accompany the hard scatter, commonly referred to as “pile-up”. It is another interesting related quantity to study the LHC delivery and ATLAS data-taking performances, and is defined as:

$$\langle\mu\rangle = L \frac{\sigma_{\text{inelastic}}}{n_b f}, \quad (2.3)$$

where $\sigma_{\text{inelastic}}$ is the cross section of the inelastic scattering of protons, which has been measured to be roughly 80 mb for $\sqrt{s} = 13$ TeV [73]. The LHC is designed to operate at

²The number of proton per bunch is assumed to be the same in both beams, but in practice it can differ. In that case N_b^2 would be replaced by the multiplication of the number of protons in each of the two beams, N_{b1} and N_{b2} .

a peak instantaneous luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ that, assuming the number of bunches to be around 2800, yields $\langle \mu \rangle \approx 25.4$. Figure 2.2 shows the distribution of μ in ATLAS for different years. The actual measured value for $\langle \mu \rangle$ tends to be higher than our estimation based on the numbers given in the LHC design report of 2004 [71]. The explanation for this observation is complicated and involves multiple factors. In essence, there are fundamental trade-offs between accelerator physics capacities (beam stability, heat load, machine preservation), peak luminosity achieved, and detector constraints (pile-up and radiation damage), which is balanced by changing the operating parameters of the LHC. The 2017 Run [74] is an interesting example to show how dynamic and continuous the tuning of the LHC can be. In figure 2.2, it can be seen that 2017 has a bi-modal pile-up distribution that stands out. During that run, the electron-cloud effect [75] imposed some constraints on the beam configuration. In particular, the electron-clouds generate heat, which must be handled to maintain the magnets in a superconducting state and avoid beam instability. This phenomenon led to the adoption of a new bunch filling scheme: the number of colliding bunches was reduced to mitigate the electron-cloud build-up. In order to keep the luminosity as high as it was, the number of proton per bunch was increased. The cost of decreasing the number of bunches n_b for a given luminosity L is an increase in pile-up, which can be clearly expected from equation 2.3. The change in LHC configuration thus creates the observed bi-modality, which shows how the pile-up depends intricately on the machine parameters and operational choices (such as the bunch filling scheme, for example).

Figure 2.3 displays the total integrated luminosity over Run 2 and part of Run 3 over time. The larger increments for Run 3 corroborate the discussion of the increase in $\langle \mu \rangle$. Both plots show the integrated luminosity delivered by the LHC in green and, of that data, what has been recorded by ATLAS. The recorded luminosity is not the same as the total delivered luminosity because of the data acquisition inefficiency (detector and read-out electronics are busy and cannot keep up), as well as the inefficiency of the “warm start”: when the beam becomes stable the tracking detectors undergo a ramp of the high-voltage and some need to turn on pre-amplifiers before data recording can occur. The Run 2 plot additionally shows in blue the data “good for physics” amongst the recorded events (this plot is not yet available for Run 3). This sub-set of data is flagged based on a certification process taking place after recording and reconstruction, the detector status is checked, in particular it is checked that each subdetector was operational, calibrated, aligned and without readout problems, as well as ensuring that the beam was stable. Run 2 has an overall recording efficiency (fraction of data good for physics) of 95.6%, and Run 3 has between 93.1% and 96.7% efficiency in years 2022-2026 [76].

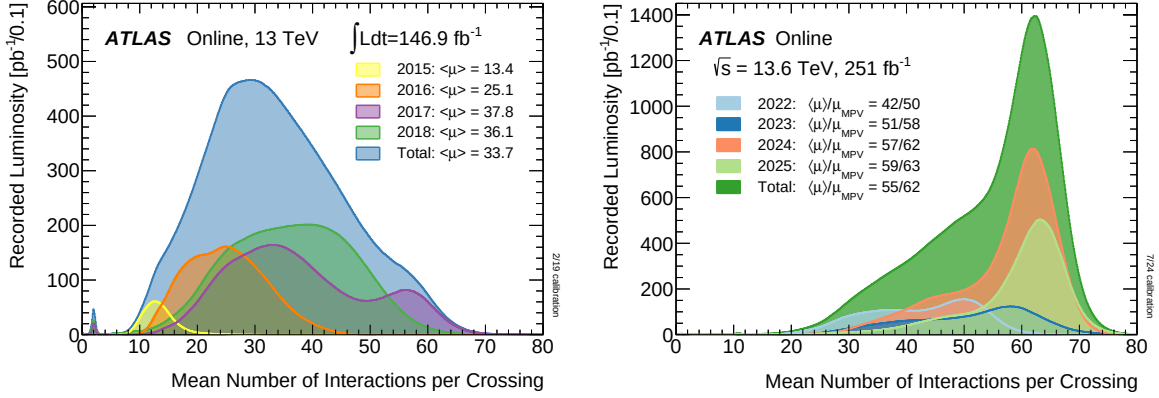


Figure 2.2: Plots showing the luminosity-weighted distribution of the mean pile-up parameter per bunch for physics data, from 2015 to 2018 (Run 2) at 13 TeV (left) and from 2022 to the 17th of September of 2025 (Partial Run 3) at 13.6 TeV (right) [77, 78]. The mean number of interactions per crossing averaged over all bunch pairs and the Most Probably Value (MPV) are shown for each year.

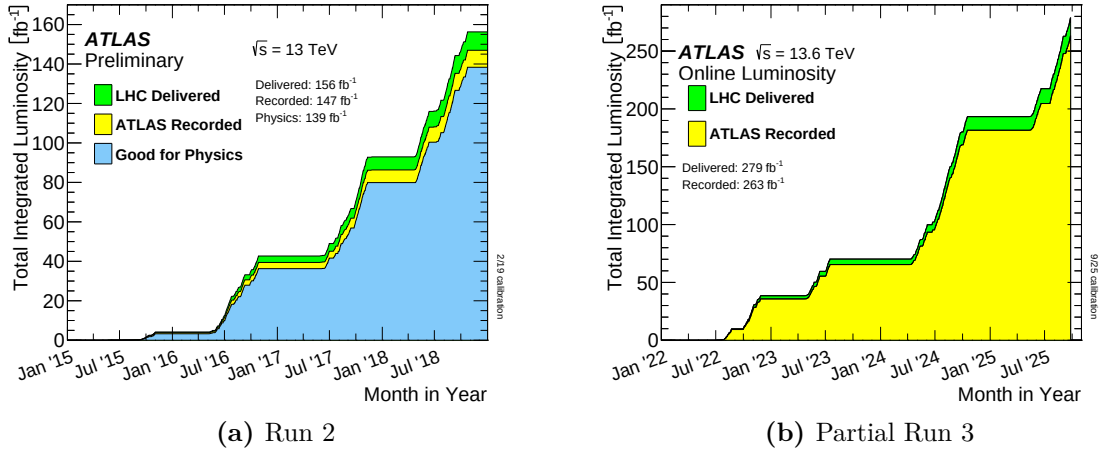


Figure 2.3: Integrated luminosity in ATLAS [77, 78].

2.3 Coordinate System

The ATLAS coordinate system is Cartesian and right-handed and has as its origin the collision point, close to the centre of the detector. The z -axis is aligned with the beam, the x -axis points towards the centre of the LHC ring, and the y -axis is a normal vector to the surface of the Earth pointing away from its centre. It is natural to use cylindrical polar coordinates due to the shape of the experiment: ϕ represents the azimuthal angle in the transverse plane x - y and θ represents the polar angle to the z -axis. The shape of the experiment itself is physically motivated: before collision, the momentum in the transverse direction is 0 for both protons (fully longitudinal) and thus the sum of the

transverse momentum of all particles produced should be 0 as well. This motivates the definition of the missing transverse energy computed from the sum of the transverse momentum of all final particles detected i :

$$\vec{E}_T^{miss} = - \sum_i \vec{p}_T. \quad (2.4)$$

Neutrinos do not interact with the detector and are thus typically very difficult to reconstruct. Their momentum can be estimated from $\vec{E}_T^{miss}$ by using algorithms like the one presented in chapter 4 or machine learning techniques [79]. Rapidity is another very useful physical quantity that carries information about the momentum and energy of particles. In Special Relativity, rapidity provides a linear measure of relative velocity between Lorentz reference frames along a specified axis. In other words, boosting from a reference frame to others along a direction corresponds to simple additions of rapidities, fulfilling a role similar to velocity in Galilean relativity. Rapidity in the z -direction is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (2.5)$$

Pseudo-rapidity, η , is a very good ultra-relativistic approximation of rapidity in the z -direction and is used in detectors like ATLAS because angles are very well measured. It is defined as :

$$\eta = - \ln \left(\tan \left(\frac{\theta}{2} \right) \right). \quad (2.6)$$

For massless particles, y and η are equivalent. A pseudo-rapidity of 0 corresponds to a particle traveling perpendicularly to the beam line, while as η approaches infinity the particle trajectory tends to being parallel to the beamline. High pseudo-rapidity particles are found in the the “forward” region of the detector. It is very useful (for example in the definition of jets as described in section 2.6.3) to define a distance in the $\eta - \phi$ plane between reconstructed objects:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (2.7)$$

Since $\Delta\phi$ is an angle difference between vectors in the transverse plane and $\Delta\eta$ a difference between two pseudorapidities along the z -direction, ΔR is approximately invariant under Lorentz boosts in the beam direction.

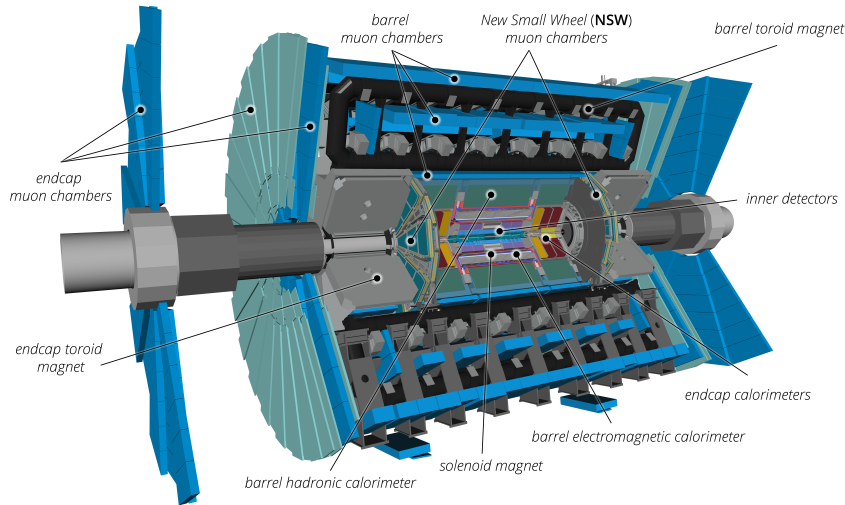


Figure 2.4: Cut-away view of the ATLAS detector in the latest Run 3 configuration with labelled sub-systems [80].

2.4 Detector Overview

The ATLAS detector, displayed in Figure 2.4, is designed to study a wide range of phenomena in particle physics, including the search for new particles, the measurement of particle properties, and the investigation of fundamental forces. To accomplish all of these tasks it must therefore measure, reconstruct, or otherwise infer all particles of the SM. The detector itself consists of several distinct subsystem layers, each optimized for specific tasks. In particular, the detector can be divided into three subdetectors ordered by proximity to the beam: the Inner Detector (ID), the calorimeters, and the Muon Spectrometer (MS). The detection (or absence) of the signatures of particles as they traverse the detector provides valuable information at each layer to identify and reconstruct them. This concept is illustrated in Figure 2.5, where the paths of the particles are cut short by interacting with different parts of the detector depending on their nature. This section provides a description for the magnets, and each subdetector, starting from the pixel detector (the most central component) and moving outward.

2.4.1 Magnets

The ATLAS superconducting magnet system [82] can be divided into two main parts: the solenoid responsible for the magnetic field used in the ID, and the toroidal magnets that generate the field bending muon trajectories in the MS. Figure 2.6 shows an illustration

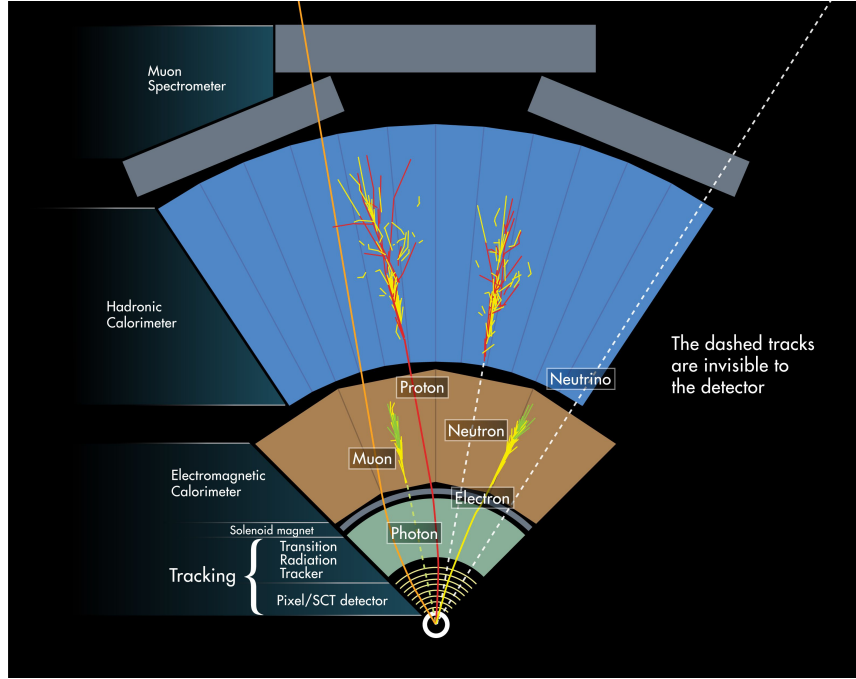


Figure 2.5: Diagram of the path of particles within the transverse plane as they travel through the different subdetectors [81].

of the magnet system, the solenoid is depicted as a central grey cylinder, and the three sets of eight toroidal loops (one set per region) are clearly visible in red. The solenoid is aligned with the beam line and provides an approximately uniform magnetic field of 2 T at its centre, where the ID resides. The toroidal magnetic field strength varies depending on position and detector region, being on average 0.5 T in the barrel and 1 T in the end-caps.

2.4.2 Inner Detector

The ID, shown in Figure 2.7, is the closest subdetector to the collision point and is primarily responsible for the precise measurement of charged particle trajectories using a mostly silicon-based tracking system. It covers the entirety of the ϕ azimuthal range and $|\eta| < 2.5$ for pseudo-rapidity. It consists of three sub-components, represented in Figure 2.8: the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). An approximately uniform magnetic field (magnitude varying by less than 1%) of 2 T permeates the ID, orientated parallel to the beam axis. The magnetic field, generated by the central solenoid, bends the trajectory of charged particles in the transverse plane, which helps to determine their charge and momentum.

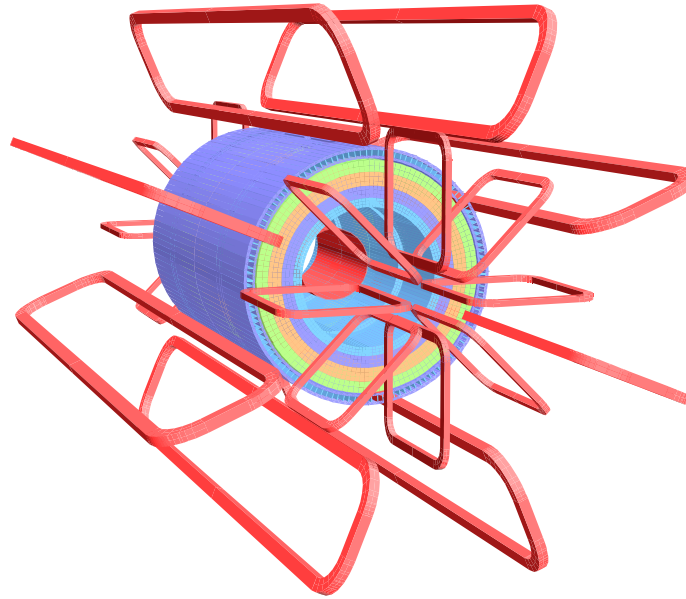


Figure 2.6: Diagram of the ATLAS superconducting magnet system taken from [82].

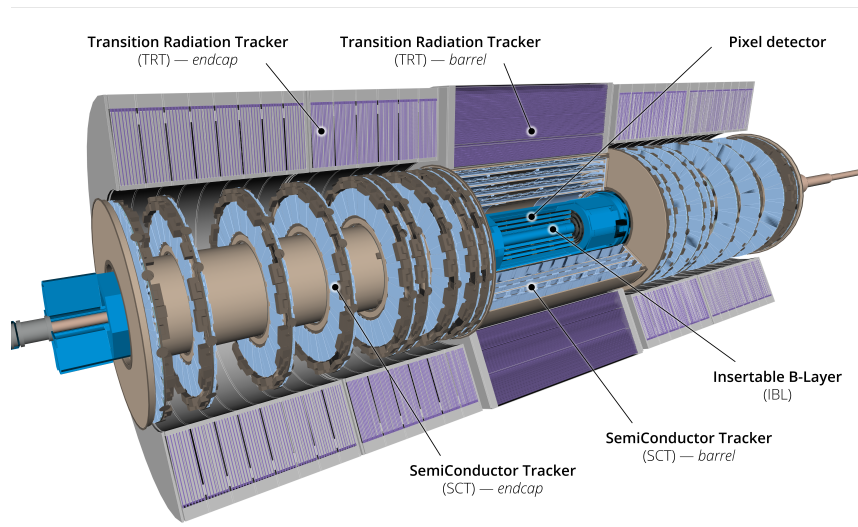


Figure 2.7: The ATLAS ID [80].

Pixel Detector

The pixel detector is the innermost tracking layer in ATLAS, closest to the beam line. This allows the pixels to capture precise and important information about the particle trajectories. The pixels are silicon solid-state detectors that detect the passage of charged particles ionizing the semi-conductor material. Silicon detectors are good at withstanding harsh radiation, which is important since the pixels are so close to the beam line. Each pixel has dimension of $50 \times 400 \mu\text{m}^2$ (50 microns in the ϕ direction and 400 microns

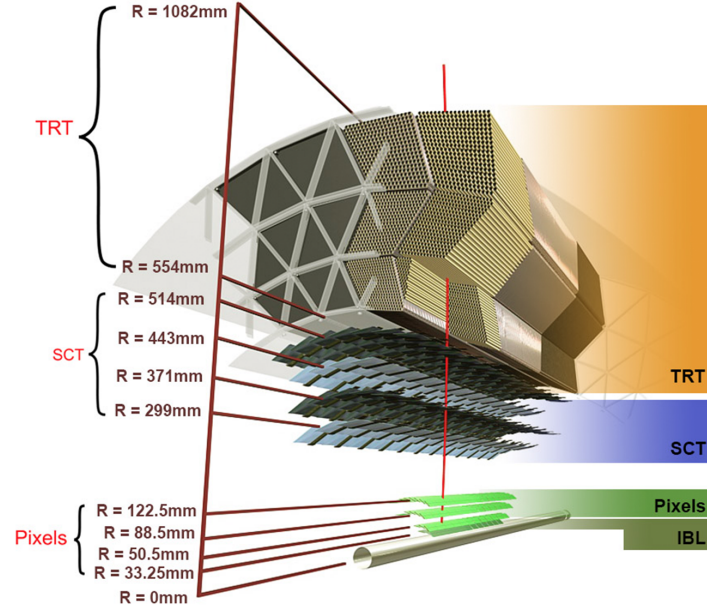


Figure 2.8: The subdetectors composing the ID [83].

in the z direction). The pixel detector is made of four such high-granularity layers with different radial distances to the beam line in the barrel region, and three disks per end-cap, resulting in an overall relatively high spatial tracking resolution. The Insertable B-layer (IBL) is a fourth innermost tracking layer introduced before the start of Run 2. It has an even higher granularity than the other three layers with pixel dimensions of $50\ \mu\text{m}$ in the ϕ direction and $250\ \mu\text{m}$ in the z direction. The IBL improves the resolution of tracks considerably and thus the resolution of vertices (points where multiple tracks cross). This information is very useful for flavour tagging, an important field that benefits many analyses in ATLAS, discussed in more detail in Chapter 3. The pixel has a pseudo-rapidity coverage of $|\eta| < 3$ and includes the entire azimuthal angle range. Charged particles going through the detector typically leave four track hits [84].

Semi-Conductor Tracker

The SCT is the next layer that composes the ID after the pixel detector, going outward from the beam axis. The SCT is composed of 4088 silicon microstrip detectors instead of pixels. Pixels are very small independent rectangles that measure the position of charged particles in two directions (in the ϕ - z plane). Microstrips are larger rectangular sensors with dimensions of $80\ \mu\text{m} \times 12\ \text{cm}$ that measure the position of particles very

accurately in one direction (perpendicularly to the length of the strips, along the widths). Microstrip sensors provide thus less spatial resolution than pixels, but are cheaper to manufacture. The SCT has four layers in the barrel region, as can be seen in Figure 2.8 and 2.7, and nine disks in each end-cap as can be seen in Figure 2.7. Each of these layers are made of modules that are composed of a pair of microstrips sensors glued back-to-back with a stereo angle of 40 mrad between them. This small angle is enough to provide two-dimensional tracking information in the ϕ - z plane. Typically, charged particles leave eight hits in the SCT [84].

Transition Radiation Tracker

Whereas the pixel detector and the SCT are made out of silicon, the TRT is made out of kapton tubes with 4 mm diameter filled with XeO_2CO_2 gas (70% Xe , 27% CO_2 and 3% O_2), known as “straws”. Each tube is coated with aluminium on the inside wall and contains a grounded tungsten wire in its centre. The aluminium coating is kept at around 1.5 kV, acting as a cathode and making the wire an anode. Incident particles ionize the gas, removing electrons from atoms that accelerate and drift towards the anode producing a signal that can be read out. The spatial resolution of the straws is worse than for the silicon pixels and microstrips but the TRT, being radially larger (see Figure 2.8), mitigates this with more hits over greater distances per particle which enhances the overall momentum measurement. The space between the straws is filled with “propylene radiators”. As charged particles travel through this material, they cross boundaries into new mediums and emit “Transition Radiation” (TR), consisting of soft x-rays, as predicted by the Maxwell equations. At high momentum, the electrons produce significantly more radiation than pions (due to their mass difference), and thus this radiation provides very useful information for their discrimination upon its detection [85]. Charged particles travelling through the TRT typically generate more than 30 hits [84].

2.4.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is a sampling calorimeter (as opposed to a homogeneous calorimeter) designed to measure the energy of charged particles and photons. It is composed of alternating layers of lead (high-Z absorber) and liquid argon (active material) sections composed into an accordion shape that covers the full ϕ range and $|\eta| < 3.2$. Incoming electrons and photons interact in the lead via bremsstrahlung

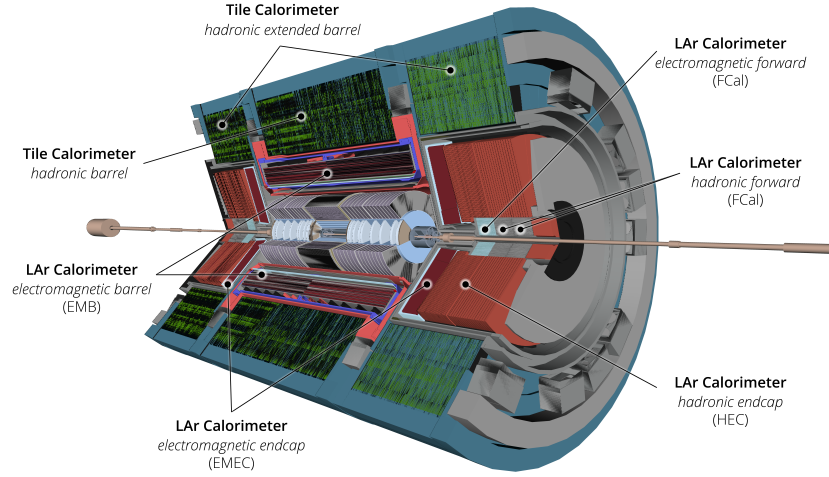


Figure 2.9: The calorimeters in ATLAS [80].

and pair production, which initiates an electromagnetic shower. The charged particles produced in the shower then ionize atoms in the liquid argon thus generating a current upon applying a voltage. This current is then read out and is proportional to the energy of the incoming particle. The ECAL has three barrel layers (EMB) with different granularity in η and two end-caps (EMEC) composed of an inner- and outer-wheel as can be seen in Figure 2.9. The first barrel layer, sometimes referred to as the “strip layer” due to its fine resolution in η ($\Delta\eta \approx 0.003$), provides important information for discriminating between single photons and $\pi^0 \rightarrow \gamma\gamma$ decays, the latter being the source of photons that can be collimated at high energy. When single photons do not convert to e^-e^+ pairs in the ID, the ECAL energy depositions can still be used to estimate the photon’s direction. A pre-sampler liquid argon detector is located between the ID and the ECAL to take into account the energy loss to material by travelling particles.

2.4.4 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) measures the energy of hadrons and is composed of three sampling calorimeter sub-systems: the Tile Calorimeter (TileCal) [86], the Hadronic End-cap Calorimeter (HEC) [87] and the Forward Calorimeter (FCal) [88]. All these sub-systems can be visualised in Figure 2.9. The TileCal itself is a series of smaller calorimeters that use alternating layers of iron plates (absorber) and plastic scintillator tiles (active material), covering $|\eta| < 1.7$. The incoming hadrons produce hadronic showers, through nuclear interactions within the iron plates, which contains

many charged particles (mostly pions, protons and electrons). These charged particles traverse the plastic scintillator and deposit energy by ionising and exciting the molecules of the active material. The excited molecules return to the ground state by emitting a photon isotropically. This light emitted in the scintillator tiles is collected at their edges and redirected to PhotoMultiplier Tubes (PMTs) by wavelength-shifting fibres. The HECs consist of two wheels that use copper plates as the absorbing medium: one thinner for low energy hadronic showers and one thicker for greater shower containment. The signal is read out by kapton-copper electrodes immersed in liquid argon (active medium) and redirected to external readout boards. As its name suggests, the FCal is located in the forward regions that covers up to $|\eta| \approx 4.9$, and therefore is exposed to a large amount of radiation being so close to the beam line. The FCal is composed of three modules per end-cap: an electromagnetic section using copper as absorber and two hadronic sections using tungsten absorbers. Each section is a highly radiation resistant compact array of liquid argon filled tubes with rod electrodes embedded in blocks of the absorber material.

2.4.5 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost layer of the ATLAS detector and is dedicated to the identification and precise measurement of muons, the only detectable particle that makes it this far without being absorbed. The MS plays an important role in the trigger system. It spans the pseudo-rapidity range $|\eta| < 2.7$ and offers nearly complete azimuthal coverage, except for localized gaps required for the services and electronics of other subdetectors and for parts of the ATLAS mechanical support structure. The MS is divided into three main parts: the barrel region ($|\eta| < 1$) and the two end-cap regions ($1 < |\eta| < 2.7$), each having their own large air-core toroid magnet generating a non-uniform magnetic field, of strength ranging from 0.15 to 3.5 T, that bends the muon trajectories. The MS makes use of four different technologies to extract information from the magnetically deflected muons: Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC) that handle precise tracking of muon trajectories, as well as Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) which are used for rapid muon identification, important for triggering (see section 2.5). The MDTs use the same straw-tube technology as the TRT detector in the ID. These aluminium drift tubes are 3 cm in diameter and are filled with a mixture of Argon (93%) and CO_2 (7%). The CSCs, used in the high background rate forward region that covers $2 < |\eta| < 2.7$ in

the end-caps, are multi-wire proportional chambers. They have arrays of several anode wires running through a gas volume and cathode planes segmented into strips that lay perpendicularly to the anodes. This disposition reduces the average drift time, which makes the CSC capable of running at higher incoming rates. When the avalanche effect occurs close to the anodes as a result of incident muons ionising the gas, charge is induced in multiple neighbouring strips. Performing charge interpolation between multiple strips results in a precise measurement of position. The MDTs and CSCs have good tracking performance in the “bending plane” (r - z) but suffer from a slow time response and lack of measurements in the (r - ϕ) plane. Three layers of RPCs provide an answer to both of these issues in the barrel region. The chambers consist of a gas mixture (93% $C_2H_2F_4$, 5% C_4H_{10} and 0.3% SF_6) sandwiched between two parallel bakelite high-resistance plates that operate a very large electric field (≈ 9.5 kV applied). This important electric field causes the ionisation of the gas to cause avalanches that provide a strong and quick signal. The η - (parallel to the beam axis) and ϕ -strips (oriented azimuthally) are connected to the plates to read the signals providing orthogonal coordinate measurements. The TGCs provide similar information in the end-cap regions and are a type of multi-wire proportional chamber optimised for very short drift times and high rate capabilities.

2.5 Trigger

Operating at nominal luminosity during Run 2, 40 million bunches of protons cross and interact at the centre of the ATLAS detector every second (one collision every 25 ns). Despite the large computing capabilities available, it is not possible to store every single event produced. Instead, the Trigger and Data Acquisition (TDAQ) system [89] is used to select events based on how likely they are to be of interest using the signature they leave in the detector. First the L1 trigger, built from hardware to be able to keep up with the high input rate of 40 MHz, filters events and reduces the event rate to ≈ 100 kHz using information from the calorimeters and the dedicated MS trigger chambers (RPCs and TGCs). The η - ϕ phase space is scanned for interesting physics objects (see section 2.6), such as high- p_T leptons or jets for example, and an associated coordinate is produced for each of them. Each coordinate describes the centre of a Region of Interest (RoI), a small area in the η - ϕ plane. The events passing the L1 trigger then go through the High Level Trigger (HLT), which are software-based algorithms (running on a computer farm) that are inherently slower but possess a higher filtering precision. The RoI found by the L1 trigger are used by the HLT algorithms to decide whether or not the event

should be saved. The HLT trigger decision is produced by using the concept of *trigger chains*: sequences of algorithms that are increasingly more computationally intensive and where events may be rejected at any stage to reduce CPU usage. The trigger chains are all gathered into a *trigger menu*, and may be selected and used to identify specific physics objects. A fixed fraction of events passing certain trigger chains may be rejected to further decrease the number of events saved, a method called “prescaling”. The HLT reduces the event rate to the more reasonable frequency of 1000 Hz. Those events passing any trigger chain on the menu, and surviving prescaling, are then saved for analysis. A set of single lepton triggers targetting muons and electrons are used in the analysis presented in chapter 6.

An example of single electron trigger chain is **HLT_e26_lhtight_nod0_ivarloose**, where **e26** refers to the transverse momentum threshold imposed on the electron (26 GeV), **lhtight** the likelihood cut identification requirement, and **ivarloose** the isolation requirement (see section 2.6.4 for more information about lepton identification and isolation). The label **nod0** indicates that there is no requirement on the transverse impact parameter. Some trigger chains might have labels similar to **1EM20VH**, which refers to a requirement on the L1 trigger firing on an electromagnetic trigger object (electron or photon) with a transverse momentum of at least 20 GeV. ‘**V**’ stands for *variable threshold* and indicates that the p_T threshold is η -dependent to compensate for variations in detector response and the material in front of the calorimeter. ‘**H**’ stands for *hadronic veto* and is a cut applied on the energy deposited in the hadronic calorimeter “behind” an EM object: if it is too high the object is likely originating from hadronisation rather than the hard scatter. It is important to note that there is a fundamental lower limit on the energy of accessible leptons due to the trigger p_T thresholds, which is very relevant to the analysis presented in chapter 6 where soft leptons are targeted in the final state.

2.6 Reconstruction of physics objects

This section describes how physics objects, used in analyses to perform measurements and searches to test theoretical models, are extracted from the raw data: a process commonly referred to as “reconstruction”. In particular, the focus is on objects relevant to the work presented in this thesis: tracks, jets, electrons, muons and missing transverse energy.

2.6.1 Tracks

Charged particles leave hits in the multitude of layers that make up the ID detector, described in section 2.4.2, and in the MS, described in section 2.4.5, in the case of muons. These hits are connected to reconstruct the helical trajectories of the charged particles, called *tracks*, bent by the magnetic fields generated by the ID solenoid and the MS toroid. The momentum of the particle associated with the track can be obtained by considering the curvature of the trajectory due to the magnetic field. In the transverse plane (that is perpendicular to the field of strength B) the particle of charge q oscillates in a circle with radius R such that their momentum can be evaluated as $p_T = qRB$. The tracks themselves are parametrised by five variables in a perigee representation as is displayed in Figure 2.10:

$$\vec{x} = (d_0, z_0, \phi, \theta, \frac{q}{p}), \quad (2.8)$$

where d_0 , z_0 are the transverse and longitudinal spatial impact parameters, respectively, ϕ and θ are the angles of the momentum at that spatial point (d_0, z_0) with respect to the beamline as described in section 2.3, and q/p is the charge divided by the momentum. The impact parameters and the angles are all defined with respect to a reference point: the average position of the proton-proton interactions, the “beamspot position”. The perigee representation is thus defined by establishing the point of closest approach to the reference for each track as well as the momentum at that point. These variables are also important in the field of flavour tagging, which is discussed in chapter 3. Tracks are required to have a minimum p_T of 400 MeV and to leave hits in the ID, the MS or both. Muon tracks may be reconstructed as “combined” muons (using the combined hits of both the ID and the MS) or “standalone” muons (using only information from the MS). In the latter case, the muon tracks are bounded as $|\eta| < 2.7$, whereas in the former the trajectories must have $|\eta| < 2.5$ to leave hits in the ID.

The ID tracking is performed by first going “inside out”, i.e from the Pixel to the TRT subdetector, referred to as primary tracking. Then the back-tracking procedure goes “outside in” in the opposite direction from the TRT to the Pixel subdetector. The primary track reconstruction starts by the formation of seeds consisting of three space points in either the Pixel or SCT detector. Space points hold 3D spatial information and are reconstructed from one hit in the Pixel and two in the SCT. Some loose selection criteria are applied to prevent the reconstruction of seeds that will provide low-quality tracks and save on computation time. A fourth confirmation space-point from a different

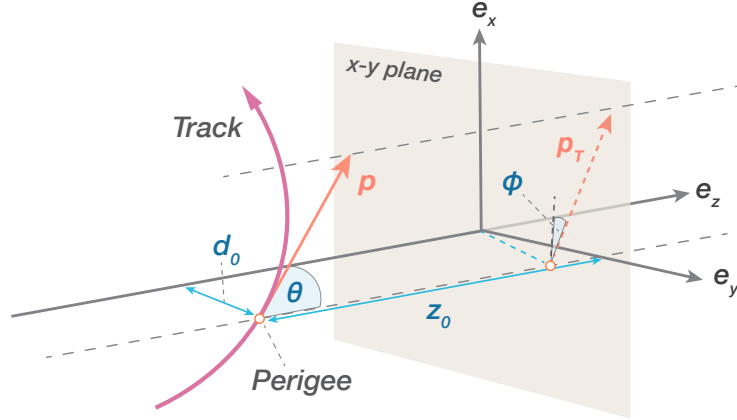


Figure 2.10: The track parameters as a perigee representation. The reference point can be seen as the origin where the axis vectors e_i meet. [90].

layer can be used to replace the inner- or outer-most space-point of a seed to create a new one if its curvature is compatible with the original. A set of detector modules that may contain clusters compatible with the seeds, called “search roads”, are built through the rest of the Pixel and SCT layers based on the seeds layout. These search roads limit the number of clusters to consider and thus reduce the combinatorics when computing the end result. The seeds are then extended along the search roads through a combinatorial Kalman filter [91, 92] that searches for adjacent clusters in both the inward and outward directions. An ambiguity resolution step is then performed to resolve overlaps between track candidates and reject incorrect combinations of clusters that are not related which correspond to fake tracks. The track candidates get scores based on multiple quality criteria: the candidates with a low score sharing a large number of hits with higher-score candidates are rejected since shared hits are often an indicator of fake or low-quality tracks. The remaining tracks are then re-fitted using a global χ^2 method that takes into account the measurements errors and expected uncertainties to obtain the final high-precision track parameter estimate. An extension into the TRT subdetector is attempted by building search roads and applying Kalman filtering in the TRT volume starting from the position estimation of each track candidate and a new fit is performed including these new hits. If the fit gets worse, the fit containing only the hits in the Pixel and SCT is kept. The primary tracking is optimised to reconstruct the tracks of particles produced in the proton hard-scattering. The secondary back-tracking is performed to increase acceptance of particles produced at a greater distance from the beamline such as electrons originating from photon conversions within the detector material. Back-tracking begins in the TRT by considering segments of hits that are compatible with Regions Of

Interest (ROI), that are themselves determined by deposits in the EM calorimeter. Seeds of only two spatial points are constructed in the Pixel and SCT by considering only hits that are close to the TRT. Using the same procedure as in the primary tracking, the seeds are extended into track candidates using road searching, Kalman filtering, ambiguity resolution and a global χ^2 fit within the silicon detectors. The track candidates are then extended back into the TRT again using the hit segments. The segments that are not used in the back-tracking are fitted separately and are typically low quality tracks that are likely to be fakes.

2.6.2 Vertexing

A vertex is a reconstructed point in space where multiple tracks coincide. It is an important concept in experimental particle physics. A vertex is expected at the position of the hard scatter of the protons, there is only one such vertex per event and it is referred to as the Primary Vertex (PV). The reconstruction [93] begins by searching for a seed using the tracks. First, a seed is selected by using the current beam-spot estimation for the x- and y-coordinates, and by looking for the densest cluster of tracks using the half-sample mode method [94] along the beam axis for the z-coordinate. After the initial seed is found, it is used in combination with the tracks to estimate the best vertex position in an iterative fit where a weight is applied to all tracks [95]. The definition of the weight contains a variable called “temperature” that parametrises the acceptance of tracks (high temperature results in higher acceptance), which is used to perform annealing: the temperature is gradually reduced which makes the fit more stable and avoids potential local minima. At each iteration, the less compatible tracks are weighted down and the optimal vertex position is re-evaluated by χ^2 minimisation. After the last temperature step is performed, the tracks found incompatible with the computed vertex are rejected and moved back to the pool of unused tracks. The vertex is then created and the procedure is repeated on the unused tracks until there are none left or until no more vertices can be computed. At the end of the algorithm, a set of vertices is produced from which the PV is identified as having the largest $\sum_{tracks} p_T^2$. All vertices are kept in the recorded data but analyses mainly use objects compatible with the PV, the other vertices are considered to be caused by pile-up.

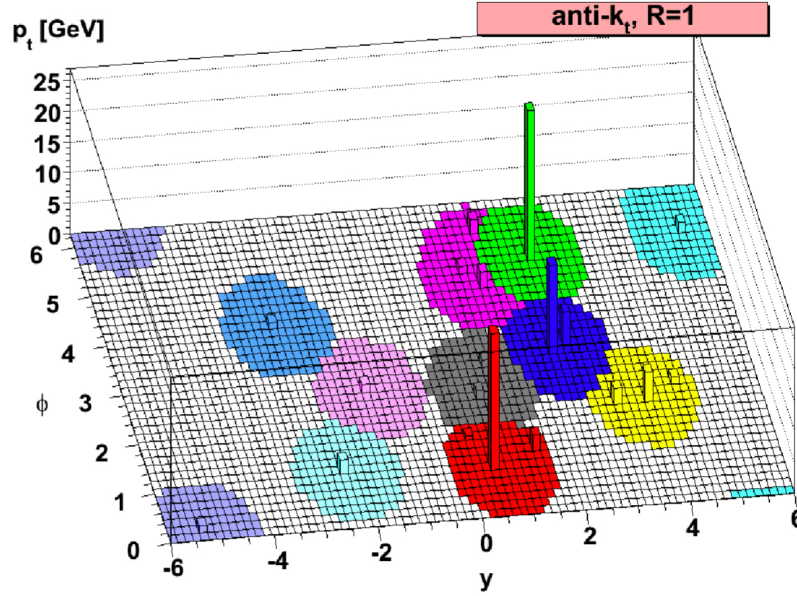


Figure 2.11: The shape of anti- k_t jets with $R = 1$ in the y - ϕ plane, formed from the particles in a simulated event [96]. Hard jets are very circular while softer jets have more complex shapes. The pair of jets around $y = 2$ and $\phi = 5$ provides an interesting example, where the purple jet on the left is much softer than the green jet on the right.

2.6.3 Jets

Quarks and gluons have a colour charge and are thus subject to the QCD interaction. In particular, these particles are subject to colour confinement: coloured particles may not travel freely at our energy scale such that, when they are emitted in the final state of the hard scatter, they undergo a process called *hadronisation* instead. Quark-antiquark pairs are produced very promptly ($\tau_{\text{hadronisation}} \approx \frac{1}{\Lambda_{\text{QCD}}} \approx 10^{-24} \text{ s}$) and form colour-neutral bounded states called hadrons which are collimated around the original particle's direction. Jets are physical objects that are computed by aggregating all particles coming from a hadronised particle in an effort to approximate its original state. One can think of jets as having a cone-like shape originating from the reconstructed PV. Jets are reconstructed by applying a clustering algorithm on tracks and/or calorimeter clusters. The anti- k_t algorithm [96] is mainly used as an infrared- and collinear-safe method, i.e a method that is not sensitive to the addition or removal of soft emissions, which is important when studying QCD objects [97–99]. Figure 2.11 displays the interesting shapes of the anti- k_t jets, in particular how it is affected by other nearby jets based on their relative transverse momentum. The algorithm receives as input a set of four-momenta which may be taken from the tracks or the calorimeter energy deposits, and has as a jet radius parameter R

defined in η - ϕ space. In the work presented here, the jets used are formed with radius $R = 0.4$, and an algorithmic combination of both the tracks and the calorimeter energy deposits is inputted following the particle-flow (PFlow) technique [100]. The algorithm performs an energy subtraction to remove overlaps between momentum and energy measurements in the ID and calorimeters, respectively. Good quality tracks associated with the PV are selected and matched with topo-clusters [101]. The expected energy left in the calorimeter by the charged particle that created the track is then computed based on the matched topo-cluster's position and the momentum of the track. For each match between track and topo-cluster the algorithm evaluates the probability that the particle deposited energy in multiple topo-clusters and uses that information to decide if it is necessary to add more topo-clusters to the system to recover the entirety of the energy. Finally, the topo-clusters attributed to charged particles are removed such that only the neutral particle energy is accounted for in the calorimeter thus avoiding double counting. Left over topo-cluster energy can be removed if it is found consistent with shower fluctuations. Compared to more naive approaches, the PFlow algorithm improves the reconstruction of the soft activity, i.e the energy contribution that does not come from the physics objects and thus is not associated with the hard scatter, but rather with the soft interaction between the protons. This is an important quantity for the computation of E_T^{miss} [102], and it benefits from the pile-up suppression coming from selecting high quality tracks linked to the PV and from the cleaner separation between neutral and charged energy/momentum contributions.

2.6.4 Electrons

The electron reconstruction [103] makes use of topologically connected clusters in the ECAL and the ID tracks. First, a seed-cluster reconstruction occurs within the ECAL: a sliding-window algorithm [104] (sliding across the calorimeter cells) based on transverse energy is applied to create the cluster candidates. If two candidates are very close in space, the cluster with smaller transverse energy is discarded. Each track that matches these clusters geometrically, by extending the trajectories into the ECAL, is considered an electron candidate. The tracks corresponding to electron candidates are then refitted using the Gaussian Sum Filter (GSF) method [105], a generalisation of the Kalman filter that takes into account non-linear effects associated with bremsstrahlung. These refitted tracks are then matched to the ECAL seed-clusters, similarly as before the GSF method is applied but with tighter requirements. If multiple tracks match the same cluster, an

algorithm selects a track based on its η - ϕ distance to the barycenter of the cluster within the second layer of the ECAL and the number of hits in the ID. Finally, the clusters are reconstructed by extending the window used to build the seeds in the ECAL (in η or ϕ) and a calibration, using both data and simulation in multivariate methods [106, 107], is performed to recover the energy of the original electron.

Physics analyses require prompt electrons that originate from the hard scatter, as opposed to fake electrons originating from $\pi^0 \rightarrow \gamma\gamma$ showers, converted photons, heavy flavour decays or mismatched tracks and clusters. Two independent procedures are defined to suppress the backgrounds: electron identification and isolation. Identification tells us whether the object looks like an electron, isolation tells us whether it behaves like a prompt one. The electron identification is based on a likelihood that is built from the probability distributions of some observables related to calorimeter shower shapes, track properties, and track-cluster matching variables, which are extracted from simulations, to determine how likely a given reconstructed electron is to be prompt. Cuts at different thresholds on the likelihood discriminant provide multiple identification working points with different tightness, i.e trade-off between background rejection and efficiency, that the analysis may use. The electron isolation is based on variables reconstructed from the scalar sum of transverse momentum of tracks or calorimeter energy deposits within a cone around the electron direction, excluding the electron itself. Similarly as for the identification, working points are defined by cutting on these observables based on the distribution computed in simulations.

2.6.5 Muons

Tracks in the MS are reconstructed separately [108], starting with the identification of short local straight-line segments from hits in the MS subsections. These segments are found with a pattern recognition method, called the Hough transform [109], and are then connected to form track candidates using parabolic trajectories that are first order approximations for the path taken by charged particles in the MS toroidal magnetic field. A global χ^2 fit, taking into account the effects of potential interactions in the detector as well as the effects of possible misalignments between the different chambers, is performed to obtain the tracks. These fitted tracks are used to remove outliers and add unconsidered hits found along the computed path. Similarly to the creation of ID tracks discussed in section 2.6.1, an ambiguity resolution step occurs to remove tracks sharing hits. The tracks are re-fitted taking into account the energy loss in the electromagnetic

calorimeters and extrapolating back the trajectories to the beam line. The MS tracks are then matched to the ID tracks and another final fit completes the muon reconstruction.

Just like the reconstructed electrons, identification and isolation working points are defined to produce analysis-level selections gauging the quality and probability of promptness for the muons. The muon identification is not based on the definition of a multivariate likelihood like the electrons, but instead on requirements on the track fit properties, the detector coverage, the consistency between measurements in the ID and the MS, and the number of hits in both subdetectors. The muon isolation is similar to the electron: variables of objects within a cone defined in η - ϕ around the lepton are studied and cut on. Muons deposit minimal energy in the calorimeters and are less affected by radiative loss compared to electrons, thus the isolation relies on track-based observables.

2.6.6 Missing Transverse Energy $\vec{E}_T^{\text{miss}}$

Equation 2.4 was defined to illustrate the concept of missing transverse energy and how it is possible to measure it due to the shape of the detector. In practice, $\vec{E}_T^{\text{miss}}$ is computed from both reconstructed and calibrated physics objects (jets, electrons and muons, all described in this section), and additional charged-particle tracks associated with the hard-scatter but none of the calibrated objects [102]. Specifically, one can write:

$$\vec{E}_T^{\text{miss}} = \vec{E}_T^{\text{miss,e}} + \vec{E}_T^{\text{miss},\gamma} + \vec{E}_T^{\text{miss,jets}} + \vec{E}_T^{\text{miss},\mu} + \vec{E}_T^{\text{miss,soft}}, \quad (2.9)$$

where each physics object has an associated term (the τ term was left out since the particle is not reconstructed in the work presented in this thesis), and the *soft* term encapsulates the contributions that are not from the physics objects. The soft term is important but challenging to compute due to the presence of pile-up and the underlying event. Amongst the existence of a few computation methods for the soft term, the Track Soft Term (TST) [102] is used and relies only on the tracks as its name suggests.

2.7 Monte-Carlo generators and detector simulation

The data measured by the detector must be compared to theoretical predictions in order to test and improve our physical models and thus our understanding of the Universe. The cross section for two protons interacting in a collisions to yield the final state X is

given by [99]:

$$\sigma_{pp \rightarrow X} = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b f_a(x_a, \mu_F) f_b(x_b, \mu_F) \int d\sigma_{ab \rightarrow X}(x_a x_b s, \mu_F, \mu_R). \quad (2.10)$$

Several comments can be made about equation 2.10:

- The partons interacting in the collision are denoted a and b . The sum of a, b is over the parton types, i.e quarks, anti-quarks and the gluon.
- $f_i(x_i, \mu_F)$ is the Parton Distribution Function (PDF), it describes the probability that the parton i carries the fraction of proton momentum x_i , and depends on the factorisation scale μ_F .
- The PDFs and the hard scatter cross section of the partons are expressed separately, which is possible due to factorization theorems [110]. The physics is thus split into proton- and parton-scale terms respectively, both sharing the μ_F unphysical parameter that determines how collinear QCD radiation is attributed between them.
- $\sigma_{ab \rightarrow X}$ is the cross-section for the production of the final state X through interaction of the partons a and b . Its differential may be expressed as: $d\sigma_{ab \rightarrow X} = \frac{1}{2x_a x_b s} d\Phi_X |\mathcal{M}_{ab \rightarrow X}|^2(\Phi_X, \mu_F, \mu_R)$ where Φ_X is the final state phase-space, $\mathcal{M}_{ab \rightarrow X}$ the QFT matrix element for the interaction of the partons, μ_R the renormalisation scale and s the proton-proton centre-of-mass energy squared.
- There exists no first principle method to pick values for μ_F and μ_R , they are typically set as $\mu_F = \mu_R = Q^2$ where Q^2 is a hard scale identified from the process of interest. For example, in the production of a resonance with mass M via s-channel, one can set $Q^2 = M^2$.

The proton PDFs are extracted from deep inelastic scattering data gathered by multiple experiments with different methods [111], and are accessible via the LHAPDF interface [112]. Many ATLAS analyses focus on measuring the hard scattering cross section parameters. The gluons and quarks emitted in the final state that carry a large amount of the hard scatter energy do not confine, but instead produce more partons via gluon splitting and radiation. This process is called a parton shower and can be simulated using perturbative QCD. As the partons reach energies of order roughly $O(1 \text{ GeV})$, hadronisation occurs and perturbative calculations are no longer applicable, QCD inspired phenomenological models are used instead. There exist two main classes of such models: the Lund string model [113, 114] and the cluster model [115, 116]. When two protons

collide, it is possible for Multiple Partonic Interactions (MPI) to occur, or for the remnant components of the protons to interact. Both processes are examples of interactions that contribute to the observed particles without originating from the hard-scatter, and any such contribution is referred to as the Underlying Event (UE). Several underlying event models exist [117] for simulation and rely on tuning to experimental data [118]. Measurements of the UE have been performed to test the solidity of certain of these models within ATLAS [119]. The effect of pile-up, described in section 2.2, is difficult to suppress and thus also important to simulate. This is done by overlaying additional *minimum bias* events onto the primary interaction [120], minimum bias referring to events that pass the loosest available trigger. The correct pile-up profile, defined as the distribution of $\langle\mu\rangle$, such as the ones presented in Figure 2.2, is obtained by weighting each event based on the corresponding data-taking period considered.

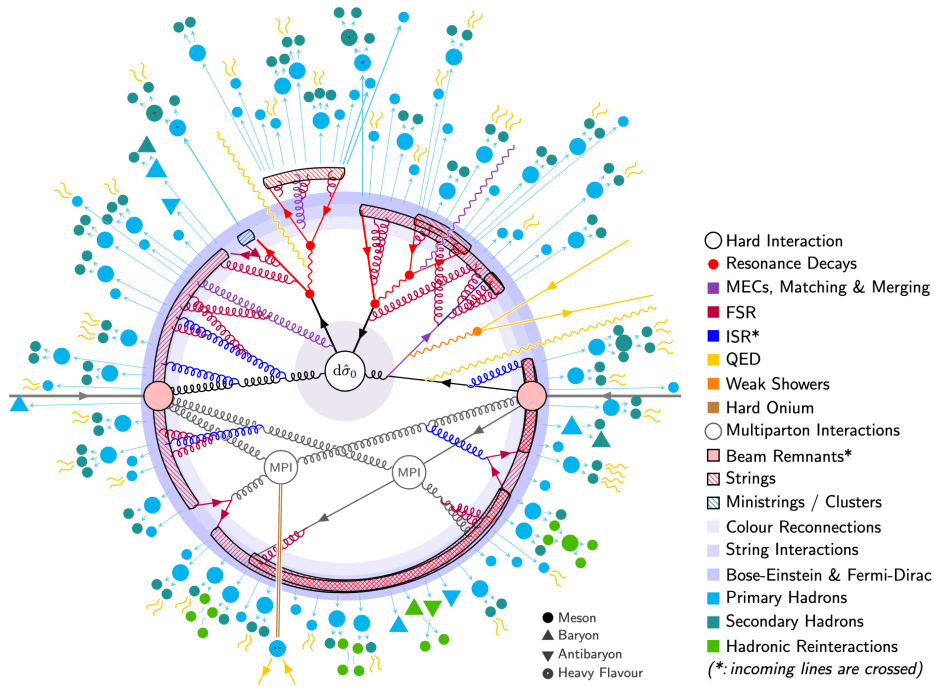


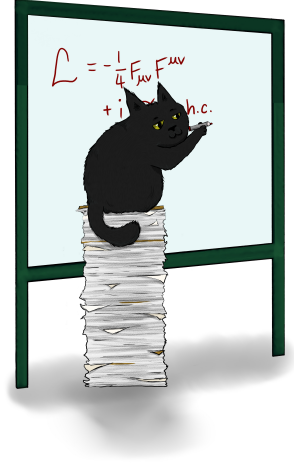
Figure 2.12: Illustration of a simulated $pp \rightarrow t\bar{t}$ MC event with parton showering where the hadronisation is simulated with the Lund string model. Several simplifications are made, not all decays are shown and particle kinematics are treated symbolically. Taken from the PYTHIA8.3 documentation [114].

The calculations mentioned in this section are quite complicated and involve, in particular, solving multi-dimensional integrals. The Monte-Carlo (MC) numerical method [121] is used to produce simulated events that follow the theory of interest, which can then be used statistically as a prediction. A variety of tools, called MC generators, fully

implement the physics (as stand-alone software or with interfacing of multiple generators) and can be used to generate these events that constitute the theoretical prediction that is compared with the data. An illustration of a MC-simulated proton-proton collision resulting in a $t\bar{t}$ pair is shown in Figure 2.12. Despite the fact that the particle information has been considerably simplified, the schematic representation of the event gives an idea of its great complexity. The relevant energy scales are represented as concentric circles, the energy decreases going outward and the innermost component represents the hard-scatter labelled $d\hat{\sigma}_0$. The black fermion lines connected to the hard scatter represent the top quarks that then decay weakly to three quarks (as can be inferred from the string connections). The Initial State and Final State Radiation (ISR and FSR) is shown in blue and dark red respectively. It can be seen that several gluons are emitted in the initial and final states, and later connect with the quarks originating from the hard-scatter with colour strings. Scale determines the physical effects that have to be modelled, namely Colour Reconnection (CR), the string interactions, the Bose-Einstein enhancements and Fermi-Dirac suppressions. The concept of CR is important, its necessity coming from the UE, which is visible on the illustration as the MPIs, and the interactions of beam remnants. In the Lund string hadronisation model, the gluons and quarks are connected with colour strings which are computed separately for the hard-scatter and the MPIs. The MPIs introduce additional quarks and gluons that may have colour strings that overlap in phase-space with colour strings from the hard-scatter (or other MPIs), in which case there is an ambiguity on the physically optimal configuration. CR is then a physically motivated method that rearranges the colour strings in a dense environment after hard-scatter and MPIs to resolve these ambiguities. Optionally, overlapping string interactions can be modelled with the concept of *rope*: nearby colour strings are combined into higher colour-representation states, leading to an effective increase in string tension. It is only after the event is generated that the Bose-Einstein and Fermi-Dirac statistical effects are considered by applying relatively small corrections to the particles four-momenta. In particular, the production of identical bosons emitted with a small four-momentum difference is enhanced. Beyond that last circle is outside the parton and hadronisation scale, where the hadrons (represented by coloured circles and triangles) travel and decay into other hadrons or fundamental particles.

In ATLAS, the ATHENA software [122, 123] is used to produce the simulated samples through multiple stages. First an MC generator is used to generate the events by calculating $\mathcal{M}_{ab \rightarrow X}$, the parton shower and the hadronisation (using an implementation of one of the two models). The information accessible at this stage is referred to as **parton level** (or sometimes as the truth level) when looking at individual partons

of the hard scatter or **particle level** when looking at hadrons after the shower and hadronization took place. Next the particles travelling through the detector geometry is simulated with Geant4 [123, 124] and interaction with the material is recorded and later digitalised to simulate the detector response. The exact same procedure applied to real data can then be applied to this simulated detector data to reconstruct comparable physics objects. The information present after this procedure is called the reconstruction level, or more simply the **reco level**.



3 Flavour Tagging

3.1 Introduction

Flavour tagging is the art of tagging jets with flavour information, specifically identifying the flavour of bounded quarks in the potential hadrons contained within the jets. This process is crucial to many analyses, including the one presented here in chapter 6. Flavour tagging in ATLAS focuses on discriminating between b , c , l and τ flavours. The b - and c -jets refer to jets containing hadrons with b - and c -quarks respectively, while l stands for “light” jets, originating from u, d and s quarks as well as gluons. The τ flavour refers to jets coming from τ decays. The main physics that is levered to discriminate between jet flavours is the relatively long lifetime of B -mesons ($\tau \approx 1.5 \cdot 10^{-12}$ s, $c\tau \approx 450 \mu\text{m}$), which is shorter but still measurable for D -mesons ($\tau \approx 10^{-12}$ s, $c\tau \approx 150 \mu\text{m}$). The decay of B - or D -mesons is then expected to produce tracks tied to a secondary vertex within the jet. Light-jets will mostly contain pions that either decay very promptly (π^0) or survive long enough to pass through the detector ($\pi^\pm$). Considerable improvements are still being made and future taggers are being developed to also discriminate gluon- and s -jets. This chapter covers work concerning GN2v01 tagger, and one of its predecessors, the DL1dv01 tagger.

3.2 The DL1dv01 tagger

The DL1 taggers [125, 126] were the state-of-the-art for the previous generation of taggers and are based on deep neural networks. The network is composed of a mixture of fully connected layers and maxout layers [127] and has three output classes: p_b , p_c and p_l , each corresponding to the probability of the tagged jet to be of the matching flavour. A

series of low-level algorithms are manually optimised or trained to provide the inputs for the high-level DL1 taggers, as displayed in Figure 3.1. The DL1 inputs consist of: the jet p_T and η , the outputs of the IP2D and IP3D (IP standing for Impact Parameter) algorithms (per-track likelihood ratio based on simulated impact parameter significance distributions), the outputs of the Secondary Vertex Finder algorithm [128] (fits vertices within a jet using tracks) and the outputs of the JetFitter algorithm [129] (reconstructs the decay chain of long-lived heavy-flavour hadrons inside a jet using tracks). DL1R also takes as input the output of a Recurrent Neural Network (RNN) that uses the impact parameter information, called RNNIP. For DL1D, the RNNIP, IP2D and IP3D inputs are replaced with the outputs of a Deep Sets neural network [130] called DIPS, which enforces track permutation invariance, and has the same architecture and inputs as described in [131].

3.3 The GN2v01 tagger

The creatively named GN1 [132] and GN2 [133] taggers mark the introduction of Graph Neural Networks (GNNs) [134, 135] to the field of flavour tagging. The arrival of GNNs comes with an interesting and convenient paradigm shift illustrated as a diagram in Figure 3.1: a single algorithm is trained directly on tracks to learn about the jet’s internal structure and reconstruct vertices within it to finally predict its flavour. The two auxiliary training targets, namely the prediction of the origins of tracks associated with a jet and the attribution of common vertices to pairs of tracks, are introduced to eliminate the need for the low-level algorithms (shown as blue blocks on the diagram for DL1R [136]). The low-level algorithms were manually optimised or trained to encapsulate the physics of the jet structure before running on the high-level neural network. For GNNs, this physics is kept in the training in the form of these auxiliary tasks that produce the same predictions.

The architecture of the GN2 algorithm, as well as a pictogram representing jets and some of their input variables, are shown in Figure 3.2. The tracks are shown as coloured vectors in the b -jet pictogram, and are correspondingly represented by coloured boxes in the architecture diagram. The tracks (coloured boxes) are combined with copies of the jet features (light grey box) and encoded in a latent space as an effective fully connected graph using an initial deep neural network “Track initialiser”. The graph then goes through a GNN transformer using a multi-head attention mechanism that updates

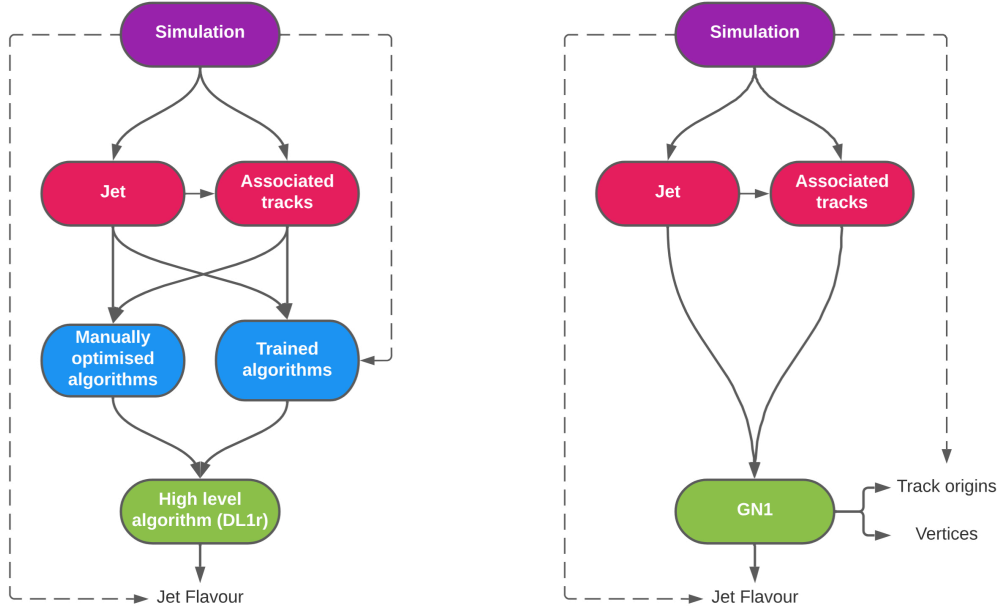


Figure 3.1: Comparison of the DL1R and GN1 flavour tagging schemes [132]. Previously, the jet and associated tracks information was passed into a set of trained or manually optimised algorithms that generates inputs for the final neural network. For GNNs, the intermediary set of algorithms is not needed and is instead replaced by auxiliary training targets, shown as arrows leaving GN1. The dashed arrows represent truth information used in training while the solid arrows indicates reconstructed information.

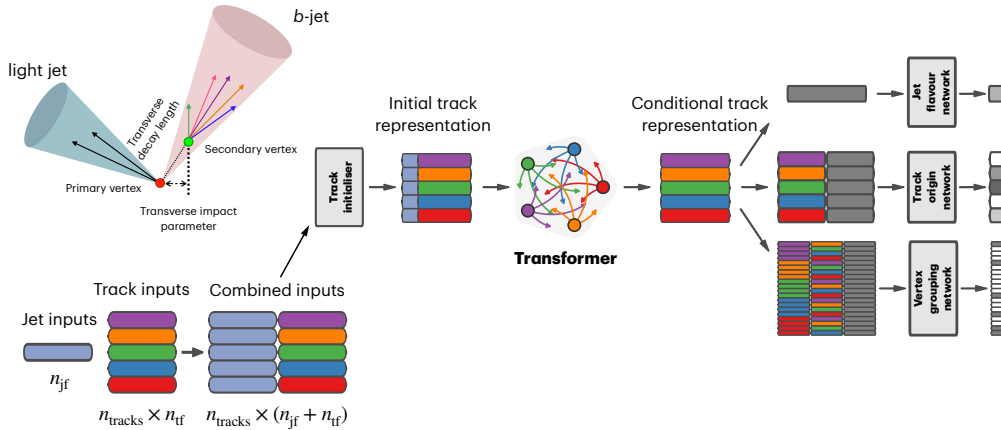


Figure 3.2: Illustration of the GN2 network architecture and the jet structure features used as input [133]. The jet features are duplicated for each track associated with the jet: n_{jf} and n_{tf} correspond to the number of jet and track features respectively. The combined vectors are then fed into an encoding network that outputs an initial latent representation for each track, which is then followed by a transformer encoder returning a global representation of the jet. Finally that jet representation is pooled and its information is fed into task-specific networks: one that predicts the jet flavour and one per auxiliary task.

Table 3.1: Input features to GN1 and GN2 [132].

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 transverse plane
$z_0 \sin \theta$	Closest distance from the track to the PV in the longitudinal 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

the nodes of the graph by aggregating the features of neighbouring nodes. The node features of the output graph are vector representations of the tracks. Each node's feature vector is passed through a dense layer to produce a scalar score, which is interpreted as an attention weight. These attention weights are then used to compute a weighted sum of the original node feature vectors. The resulting vector is called the “global jet representation”, represented by the grey box in Figure 3.2. The output graph information is then reorganised into three different ways: an input per final task neural network running on different parts of the graph.

The jet flavour classification task simply uses the global jet representation and passes it directly through a graph classification Multi-Layer Perceptron (MLP) that outputs the flavour scores: p_b , p_c , p_l and p_τ for the probability of the jet being a b -, c -, light- or τ -jet respectively.

The track classification task is performed with a node classification MLP that takes the node representations separately, each concatenated with the global jet representation. The output is a vector of size 7 containing probabilities assigned to different origins for the track summarised in table 3.2.

The track pairing vertex task is handled with an edge classification MLP: each possible permutation of track representation (node feature) pair is concatenated together with a copy of the global jet representation vector and inputted separately. The output is a scalar that represents the probability that both tracks come from the same point in space and that neither of the two tracks are fake or from pile-up.

Table 3.2: Truth origins which are used to categorise the physics process that led to the production of a track [132].

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 b -hadron decay, which itself is from the decay of a b -hadron
fromC	From the decay of a c -hadron
OtherSecondary	From other secondary interactions and decays

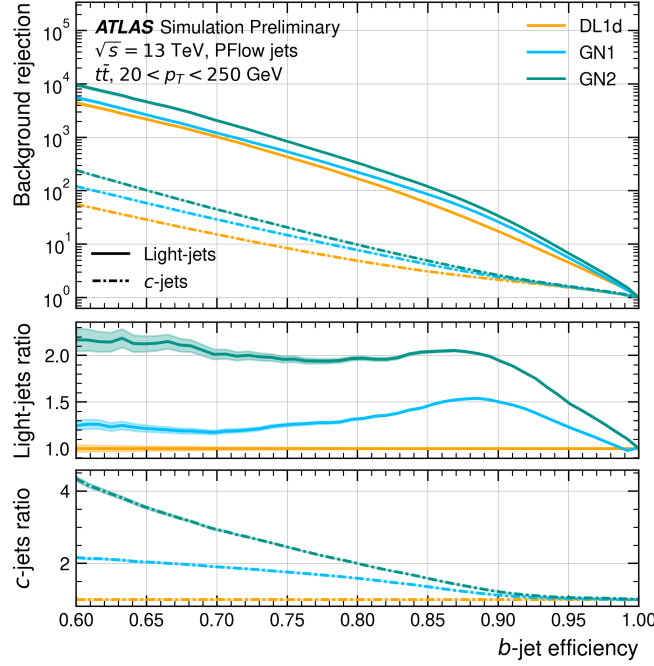


Figure 3.3: ROC curves, evaluated on a $t\bar{t}$ MC sample, displaying the flavour tagging performance of the DL1D, GN1 and GN2 taggers [137].

The comprehensive list of input variables per jet is listed in Table 3.1. The track inputs may be divided into three categories: the five track parameters (from q/p to $z_0 \sin(\theta)$), the uncertainty and significance of these parameters (from $\sigma(q/p)$ to $s(z_0)$), and the hits and holes multiplicities in the ID (from nPixHits to nSCTHoles). As pointed out in the introduction, the flavour discrimination comes mostly from the displacement of vertices in jets. The parameters of the tracks are, of course, sensitive to the presence of secondary vertices which are reconstructed in the dedicated auxiliary task. Specifically, the angular variables $d\phi$ and $d\eta$, and the curvature q/p , are expected to be sensitive to the difference in hadron mass and fragmentation within the jet. The impact parameters of tracks are expected to be correlated with the distance between the primary and secondary vertex, which is directly associated with the lifetime of the hadrons. The method of integrated gradients [138] has been used by the flavour tagging group to investigate the impact of inputs towards the final predictions of the GN2 tagger, the methodology and results are presented in [139]. This analysis found that the output of GN2 is indeed physics-driven and that the identification of heavy-flavour in jets relies strongly on the significance of the impact parameters. It also found that $d\phi$, $d\eta$, $s(\phi)$, $s(\eta)$ and the curvature q/p play an important role in the discrimination of jet flavour, whereas hit-related inputs are of limited importance in that regard. The hit multiplicity in the different subdetectors of the ID is not expected to carry direct information about flavour, these variables are

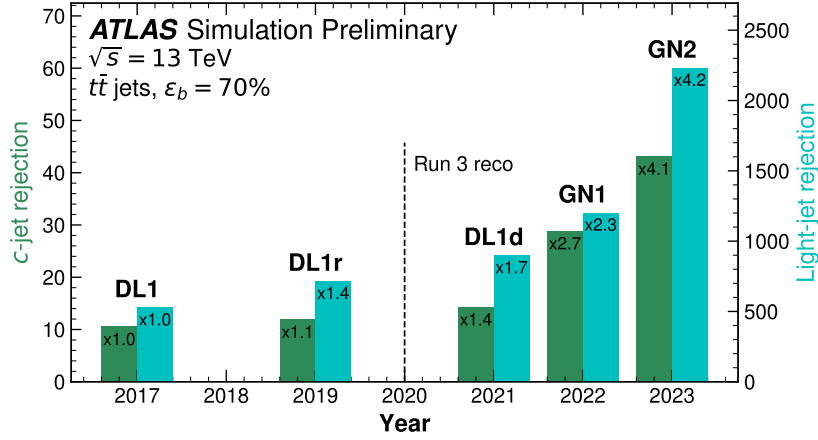


Figure 3.4: b -tagging background rejection at the 70% WP in $t\bar{t}$ for algorithms developed by the ATLAS flavour tagging group since 2017 [137]. The improvements are quantified as multiplicative factors relative to the c - and light-jet rejection rates of the DL1 tagger.

necessary features associated with track quality and measurement precision, helping the network to assess the reliability of each track during inference.

Figure 3.3 highlights the performance of the GNN algorithms and, in particular, of GN2. The Receiver Operating Characteristic (ROC) curves [140] for c -jet and light-jet rejection are shown against b -jet efficiency ranging from 60% to 100%. The GN1 rejection rates are already significantly better than the pre-GNN best algorithm DL1D. In light of these considerations, the performance of GN2 is even more remarkable, showing a considerable improvement over GN1, for instance more than doubling the light-jet and c -jet rejection increase that GN1 achieved over the previous generation of taggers at around 77% b -jet efficiency. The architecture of GN1 is a Graph Attention neTwork (GAT) [141] that started the paradigm shift of the usage of GNNs. Despite the known limitations of GATs [142], the performance results show that the usage of this new method of operating end-to-end on tracks, without intermediary taggers or algorithms, is a clear improvement. GN2 further improves the architecture by switching the GAT to a Transformer [143], the current state-of-the-art in machine learning, which contributes considerably to the further improvements shown here. Figure 3.4 compares the c - and light-jets rejection rates of the algorithms developed in the last few years by the ATLAS flavour tagging group. Taking the DL1 tagger as reference, the plot shows the great amount of progress done in a relatively short time and how GN2 stands out even from the most recent results.

The GN2v01 tagger can be used to describe the concept of working point (WP), noting that the WPs of DL1DV01 (and other taggers) are defined following exactly the same method. All taggers are historically used for b -tagging mainly, GN2 is the first tagger to also introduce collaboration-wide calibrated WPs for c -tagging. The b -tagging discriminant is defined as:

$$D_b = \log \left(\frac{p_b}{p_c f_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_l} \right), \quad (3.1)$$

where f_c and f_τ are parameters found in optimisation studies, called c - and τ -fractions respectively. The c -tagging discriminant is similarly given by:

$$D_c = \log \left(\frac{p_c}{p_b f_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_l} \right). \quad (3.2)$$

The WPs can then be defined by cutting on those discriminants at strategic values corresponding to different jet efficiencies. For b -tagging there used to be 5 calibrated WPs: 65, 70, 77, 85 and 90% b -jet efficiency, each corresponding to a cut on D_b . In the case of c -tagging, WPs definitions use both discriminants by cutting in the $D_b - D_c$ plane. A jet is c -tagged if it is different enough from a b -jet (does not pass the b -tagging 77% WP) and if its D_c value is high enough. The cut values on D_c define the WPs with 10, 30 and 50% c -jet efficiency. In the past, flavour tagging calibration was applied such that every jet in the event had to be tagged with the same WP. The calibration evolved to the so-called Pseudo Continuous Flavour Tagging (PCFT) scheme, which allows for jets to be associated to different WP independently. Figure 3.5 shows the seven 2D bins in the D_b - D_c plane, referred to as the *quantiles*, that are calibrated regions of different tightness that can be used in analyses to simultaneously b - and c - tag jets. One can recover the behaviour of fixed cuts by simply combining multiple quantiles (for example the c -tagging 50% WP is the combination of quantiles 1 to 3). When only using b -tagging, the PCFT is one dimensional, which can be thought of as marginalising over D_c .

3.4 MC-MC efficiency maps

The efficiency of the flavour tagging algorithms used within ATLAS is computed and studied using the “POWHEG + PYTHIA8” Monte-Carlo (MC) generator, also referred to as the *nominal* in this context. This name stands for the interfacing of the two generators POWHEG [144] and PYTHIA8 [114], the latter being used to simulate the parton showers.

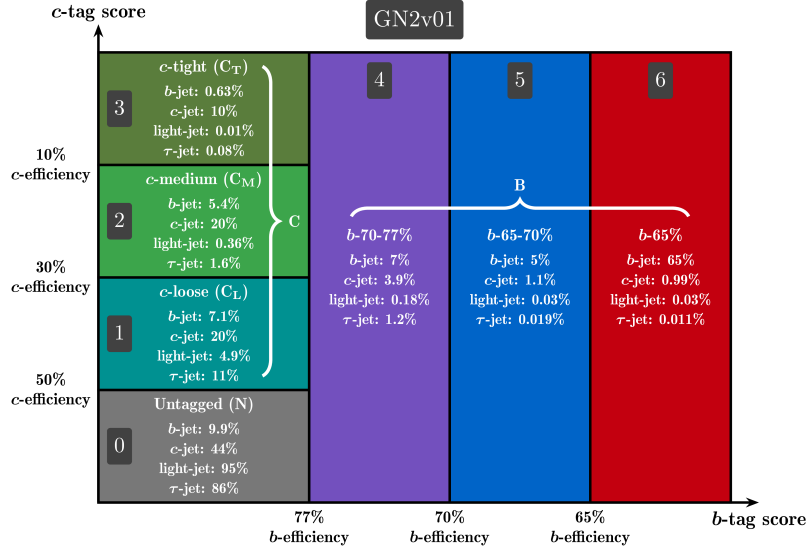


Figure 3.5: The seven PCFT quantiles, labelled 0 to 6. On each region is displayed the percentage of b -, c -, light- and τ - jets that contribute to its population. The percentage shown on both score axes correspond to the associated b - or c -jet efficiency obtained by summing from the tightest to the loosest regions. For example, the 77% b -jet efficiency is obtained by considering the union of quantiles 4, 5 and 6. The illustration is a courtesy of the flavour tagging group documentation.

There exists a few other generators, or generator combinations, that are used to produce samples for analysis within ATLAS and each use different methods to model interactions and parton showering. As briefly described in section 2.7, hadronisation and parton showering are difficult processes to simulate due to their non-perturbative-ness, the models used thus are phenomenological in nature. Because of this, the physical observables used by flavour taggers are sensitive to the choice of model which suggests that one should perform a calibration for each existing generator combination for each algorithm. A calibration is performed by comparing data to MC by hand and takes a significant amount of time and effort. Alternatively, a more practical approach is to use MC-MC efficiency maps: the efficiency of every other (alternative) MC generator is compared to the nominal by taking the ratio of the binned jet p_T - η phase space. This leads to the definition of the following scaling factors (SF):

$$SF = \frac{\epsilon_{MC,nominal}}{\epsilon_{MC,alternative}}, \quad (3.3)$$

where ϵ represents the jet efficiency for a WP and is typically parametrized by the binned phase space of the jet.

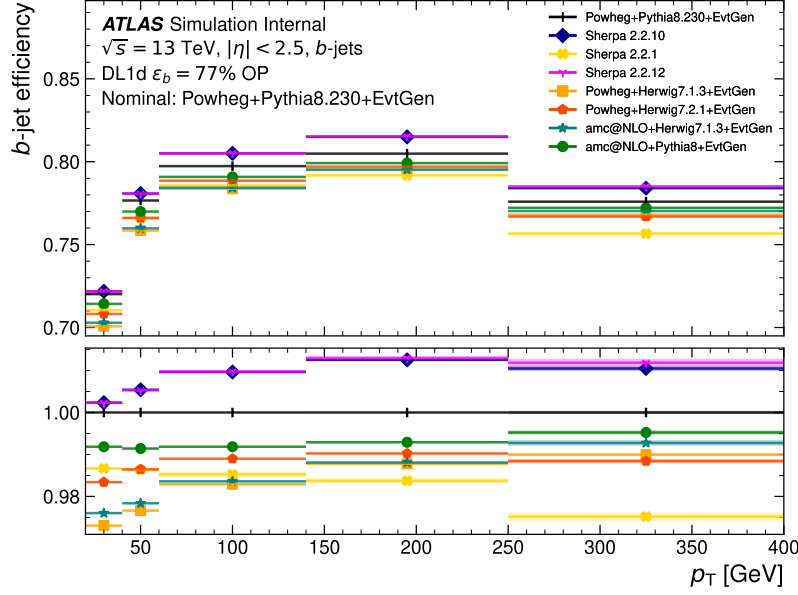


Figure 3.6: The DL1DV01 tagger’s b -jet efficiency at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.

These SFs then should be able to absorb the modelling differences and be used to “correct” an alternative generator sample subjected to a flavour tagging algorithm to make it behave more like the nominal generator. With that setup, only the nominal generator requires a calibration that can then also be applied to the “corrected” samples. One ends up with the following overall SF:

$$SF_{total} = \frac{\epsilon_{data}}{\epsilon_{MC,nominal}} \cdot \frac{\epsilon_{MC,nominal}}{\epsilon_{MC,alternative}}, \quad (3.4)$$

where the first term expresses the calibration performed with data. The tagging efficiency in data ϵ_{data} is measured by performing a likelihood fit between MC and data. An example of how it is done for b -tagging with the DL1 taggers can be found in [145]. The overall flavour tagging event weight is calculated as the product of every jet SF.

Figure 3.6 shows the typical shape of the b -jet efficiency histogram, when integrated over η , for the DL1DV01 tagger. The efficiency curve lies mostly between 0.75 and 0.82 throughout which is expected for an overall b -jet efficiency of $\epsilon_b = 77\%$. The lower plot shows the inverse of the SFs (ratio plot) for that configuration and it can be seen that they are generally close to 1, which is the desired behaviour. Generally, a SF too far from 1 is an abnormality and indicates an issue with the samples used. It can be seen that while the SHERPA2.2.10 and 2.2.12 curves are expectedly very similar, the curve

for SHERPA2.2.1 is quite different with a correction that scales down instead of up. The reasons behind this curious behaviour are studied in more detail in section 3.5.

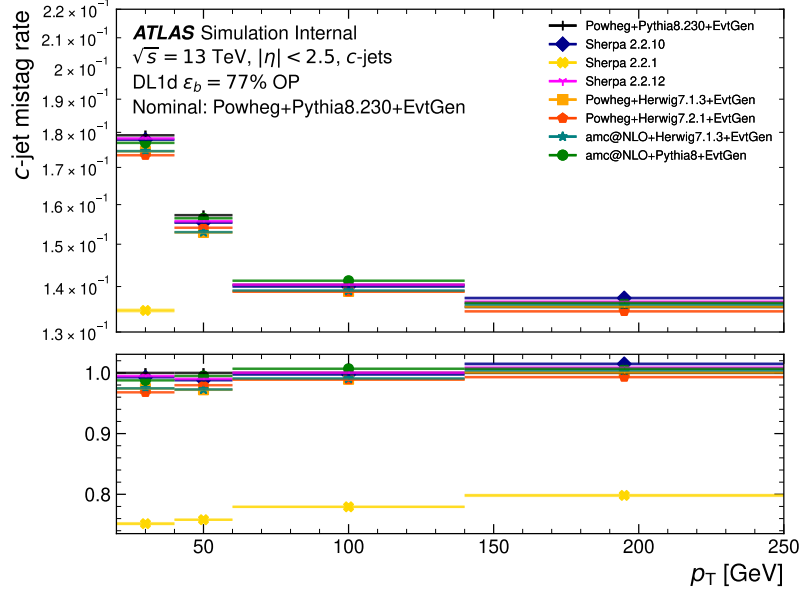


Figure 3.7: The DL1DV01 tagger's c-jet mistag rate at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.

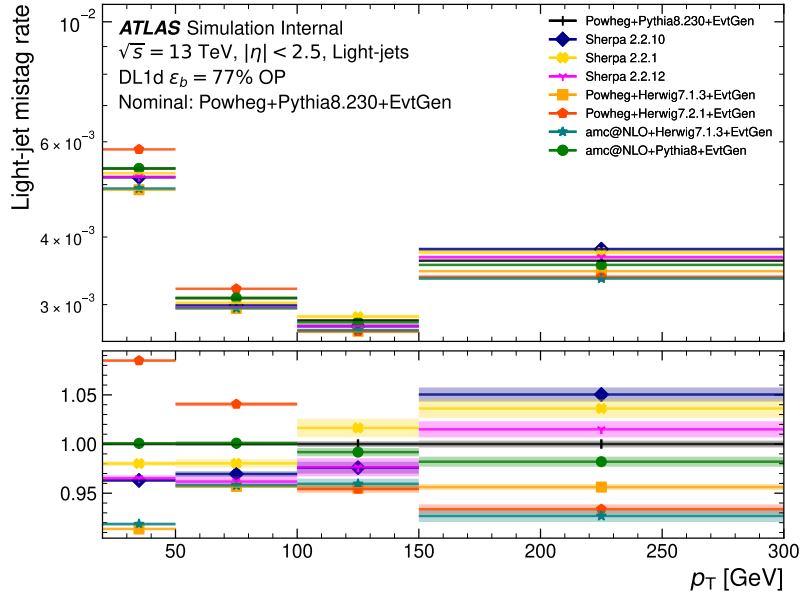


Figure 3.8: The DL1DV01 tagger's l-jet mistag rate at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.

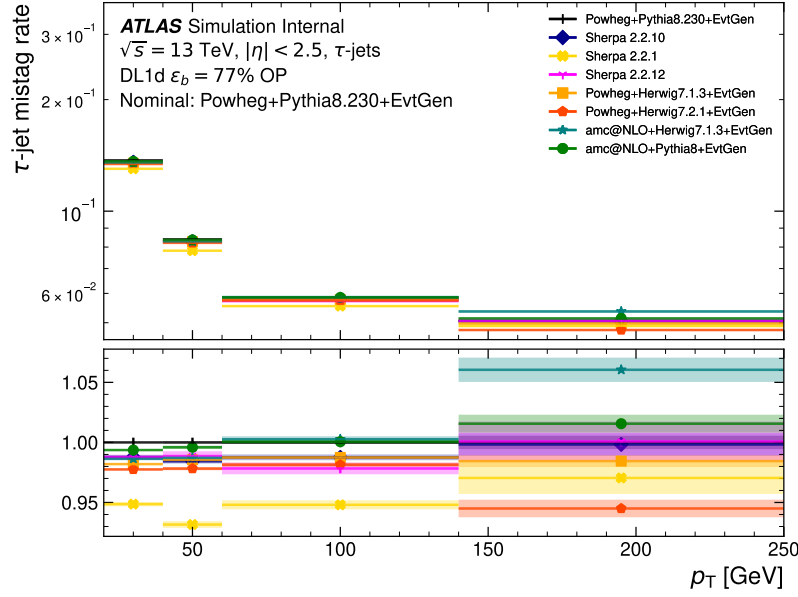


Figure 3.9: The DL1DV01 tagger’s τ -jet mistag rate at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.

Figures 3.7, 3.8 and 3.9, display curves with different shapes corresponding to the c-jet, light-jet and tau-jet mistag rates respectively. The c-jet mistag rate plot stands out due to the abnormally low efficiencies (and thus SFs) for the SHERPA 2.2.1 generator. The shape of the distribution increasing with p_T is also quite unusual as can be seen when comparing with all other generators on the plot. The source of this kind of unexpected deviation can sometimes be tracked down by exploring the truth level of the concerned samples. This is the purpose for which the RIVET routine, described in the following section 3.5, was written.

Figure 3.10 shows the b-tagging efficiencies of the new GN2V01 tagger on the top plot. Compared with the efficiencies for DL1DV01, in Figure 3.6, the position of each generator curve with relation to each other is quite similar: SHERPA 2.2.12 still has a higher efficiency than the nominal, while both POWHEG + HERWIG generators are scaled up by applying the SFs. The shape of the scaling factors as functions of p_T , on the other hand, differs significantly between DL1DV01 and GN2V01, as can be seen by comparison of the ratio plots of both Figures. The ratio of SHERPA 2.2.12 and POWHEG+HERWIG 7.1.3 to the nominal both seem to have opposite behaviour as the jet p_T increases to 250 GeV: increasing for DL1DV01 and decreasing for GN2V01. This amounts to SHERPA 2.2.12 becoming closer to nominal as p_T increases while POWHEG+HERWIG 7.1.3 gets more different for GN2V01 and opposite for DL1DV01. The curve for POWHEG+HERWIG

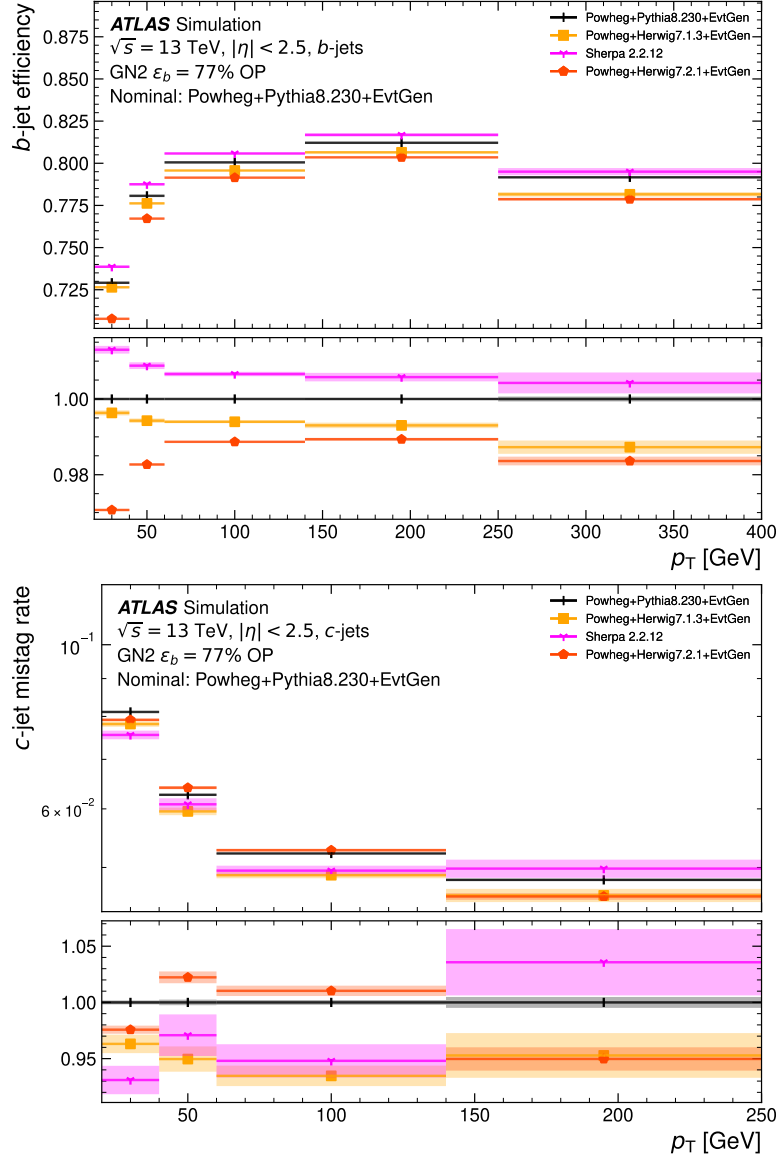


Figure 3.10: The GN2v01 tagger's b -jet efficiency (top) and c -jet mistag rate (bottom) at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.

7.2.1 is very similar for both taggers, except for the low p_T ($\in [0, 60]$ GeV) region, where the alternative generator needs greater correction for GN2v01 than for DL1DV01.

Figure 3.10 bottom panel shows the c -jet mistag rate for GN2v01. Comparing this plot bin by bin with Figure 3.7 (ignoring the abnormal SHERPA 2.2.1 curve), it can be seen that the rates are roughly between 2.25 and 2.8 times higher for DL1DV01, which is in agreement with the performance shown in the ROC curves ratio at 77% b -jet efficiency presented in Figure 3.3. Some of the MC-MC efficiency maps plots presented here have been publicly released and the integral over the ratio of efficiency distributions between different generators was released in a table published in [133].

3.5 Flavour sensitive observables in RIVET

3.5.1 Introduction to RIVET

RIVET stands for “Robust Independent Validation of Experiment and Theory” [146]. It is a powerful tool that reads parton-level and parton shower information produced by MC generators as HepMC3 event files [147], and builds parametrised reconstructed particle-level objects from the accessible final state stable particles. In particular, the stable particles are used to cluster jets, identify dressed leptons and compute event shapes according to experimental definitions, effectively replicating the particle-level stage of an analysis before detector effects. These particle-level objects are the same between different generators, and can then be used to produce particle-level observables that can be compared to experimental data unfolded to the same level. The output of different MC generators can vary significantly depending on their inner workings, making their direct comparison difficult. RIVET provides an abstraction layer that makes meaningful comparison between generators possible and consistent. Every theoretical model that can be simulated with MC generators can thus be easily compared against experimental results in that manner. RIVET is designed to create consistent and reproducible outputs, as well as to capture the essence of analyses as relatively short algorithms. The software also provides tools for detector simulation and event smearing. At this time, more than 1600 analyses are preserved in the form of RIVET routines and can be consulted here [148]. In the work presented in this section 3.5, RIVET is used to probe the particle level of $t\bar{t}$ samples made with different generators, and used to train and calibrate flavour tagging algorithms in ATLAS.

3.5.2 Comparing generators at truth level

A RIVET routine runs on the truth level of samples of interest and computes several jet-level observables that are (or related to) inputs of the flavour tagging neural networks as well as other event-level quantities that can be useful to study differences in modeling. The MC-MC efficiency maps described in Section 3.4 are computed and applied to absorb such differences that occur between generators. Studying truth level objects naturally gives access to a lot more information than simulated objects that are reconstructed with the geometry of the detector, giving insight directly into the behaviour of different Monte Carlo generators and showering algorithms. In particular one can look at the properties of the hadron associated with each jet. The ratio of the hadron p_T to the jet p_T is shown in Figure 3.11, and it can be seen that, as expected, the b -hadron carries most of the energy in the jet, the ratio peaking at around 0.90. Hadrons can have a p_T ratio to their jet greater than 1 because the particle-level jets do not include neutrinos resulting from hadronic decays. The hadrons of the c -jet carry less of the jet's energy, the distribution has a broader peak around 0.6. This is expected due to the mass difference between the quarks, the b -quark being heavier than the c -quark by a factor greater than three (4.18 GeV against 1.27 GeV [16]). The relationship between momentum fraction and the heavy-quark mass is a consequence of QCD fragmentation dynamics. Heavier quarks undergo less gluon radiation, a behaviour known as the *dead-cone effect* [149], which was measured directly by the ALICE collaboration [150]. The b -quarks, being significantly heavier, lose less energy than the c -quarks during the parton shower, resulting in the B -hadrons (containing the valence b -quarks) carrying a larger fraction of the original quark momentum than the D -hadrons (containing the valence c -quarks). This information about the fraction of energy carried by the hadron is interesting in Secondary Vertex (SV) reconstruction [151] where a vertex (point of convergence for multiple tracks) is searched for within a jet. This vertex can be interpreted physically as the decay of the hadron and thus carry valuable flavour tagging information. The SV is illustrated in the pictogram of Figure 3.2.

Figure 3.12 shows the heavy hadron population within jets. It is established heuristically that these distributions do not typically differ greatly between generators. It can be seen that roughly a bit more than half of the hadrons contained in b -jets possess a c quark, mainly D -mesons, which is expected since b -hadrons mostly decay to c -hadrons. The c -hadron population has very similar shape for the b - and c -jets, the majority of hadrons being D^0 s.

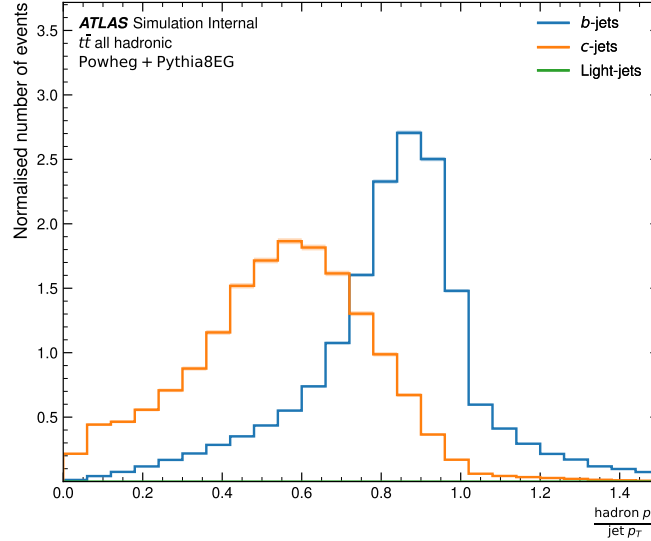


Figure 3.11: The heavy hadron p_T divided by the total p_T of the jet for jets taken in $t\bar{t}$ events decaying hadronically. The light-jets histogram has no entries in any bin since only heavy flavour hadrons are considered. These events are generated with POWHEG+PYTHIA8.

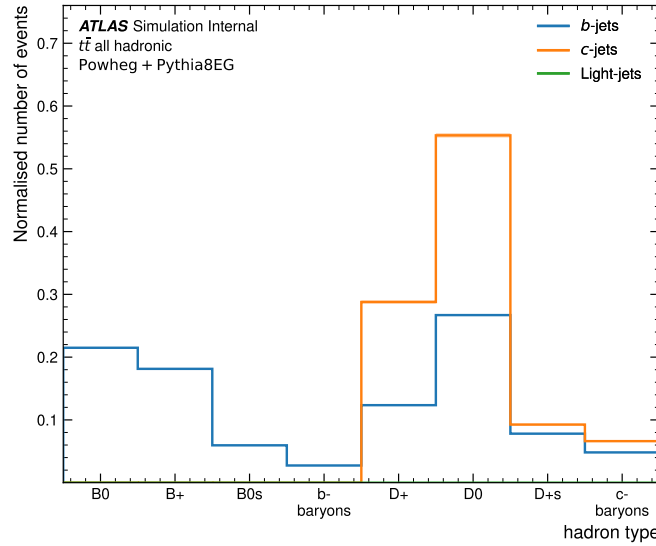


Figure 3.12: The frequency of each heavy hadron type present in jets taken in $t\bar{t}$ events decaying hadronically. The light-jets histogram has no entries in any bin since only heavy flavour hadrons are considered. These events are generated with POWHEG+PYTHIA8

Reco-level observables, such as the impact parameter (IP), represented in Figure 3.13 and in the pictogram of Figure 3.2, can also be emulated and studied. At truth level, there is no beam smearing and the PV is exactly at the space coordinates (0,0,0). Assuming that the particles travel in straight lines (no magnetic field simulation), we can

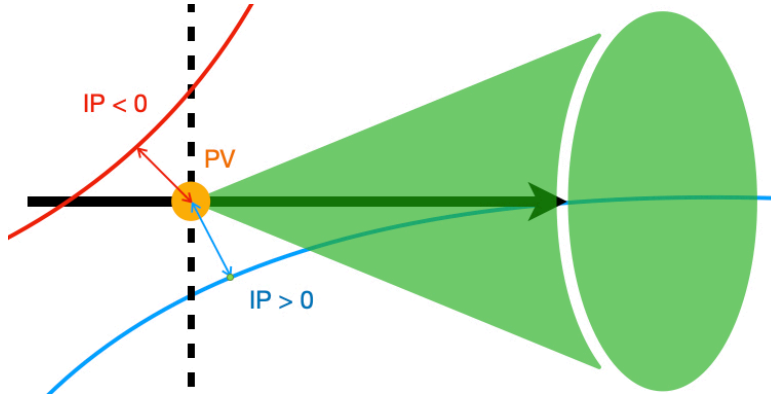


Figure 3.13: Pictogram displaying the concept of signed IP. The jet is represented as a green cone, a black arrow at its center defines the jet axis. The plane perpendicular to the jet axis and passing through the PV (dashed line) defines the sign attributed to the IP. The blue and red curves represent tracks and the double ended arrows points to their respective point of closest approach.

describe their full trajectory using their origin in space, $\vec{q}$, and their momentum direction $\vec{u}$. The vector pointing to the point of closest approach to the PV is then given by:

$$\vec{d} = (-\vec{u} \cdot \vec{q})\vec{u} + \vec{q}. \quad (3.5)$$

Taking the transverse component of $\vec{d}$ retrieves d_0 and its longitudinal component z_0 , both contributing IPs to the inputs of GN2, as shown in Table 3.1, and generally all flavour taggers. The transverse IP of each track can be signed using the geometry of the event, as illustrated in Figure 3.13. The plane perpendicular to the jet axis and passing through the PV determines the sign: points of closest approach that are on the same side of the plane as the jet emission are positive, while IPs on the other side are negative. This sign attribution can be done using information from the full geometry of the event (3D signed) or only the transverse plane (2D signed). At reco-level, in addition to carrying more information useful for flavour discrimination, these signed IPs are used to calibrate the light-jet efficiency ϵ_l [152]. The taggers have such low light-jet mistag rates that it is difficult to get a meaningful ϵ_b to ϵ_l ratios in data to perform a calibration. Figure 3.14 displays the distribution of signed d_0 and z_0 of tracks within jets at reco-level. It can be seen that the distributions are mostly symmetric only for the light-jet while the tracks in b - and c - jets are much more likely to have positive IPs. Furthermore, there is almost no difference in the distribution shape between the three jets flavours when looking only at the negative values for both IPs. It can be assessed that ϵ_l is mostly unaffected by flipping the sign of the IP of all tracks by assuming that the light-jet mistag rate is mostly due to track resolution. This assumption is actually

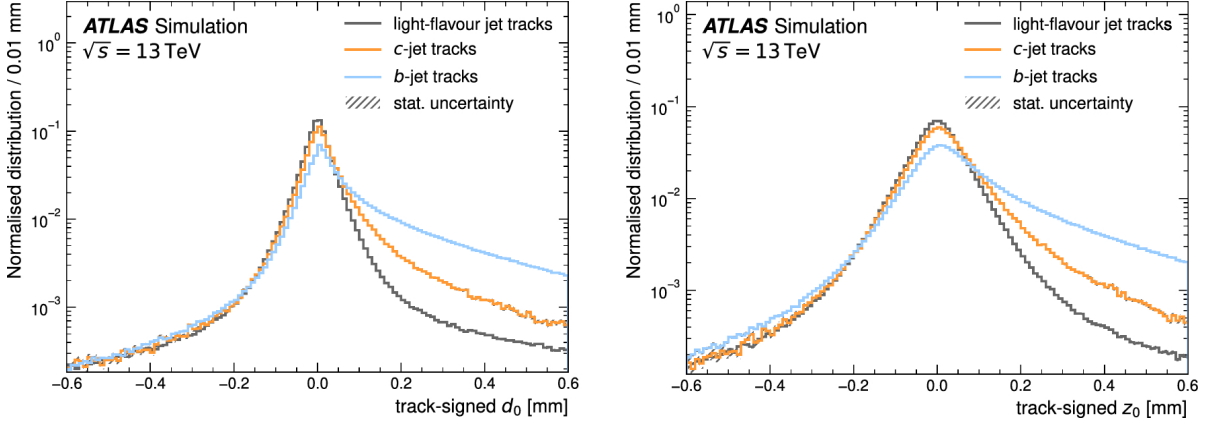


Figure 3.14: Distributions of the signed d_0 and z_0 IPs of tracks matched to b -, c - and light-jets in simulated $t\bar{t}$ events at reco-level. The events are generated with POWHEG+PYTHIA8 [152].

challenged by a tail appearing on the positive side of the signed IP distributions, which is dealt with dedicated extrapolation uncertainties in practice. Separate flavour taggers, called “Flipped taggers”, are then trained on tracks with inverted signed IPs to have lower b - and c -jet efficiencies but roughly the same ϵ_l which can then be compared with the original tagger and calibrated. This is known as the “negative tag methods”.

Figure 3.15 shows d_0 and its signed versions at truth level, computed using the Rivet routine. The unsigned d_0 is histogrammed as $|d_0|$ because the distribution is symmetric about 0. These plots are made using the jet particle constituents with $p_T > 500$ MeV and ignoring photons (a lot of photons are produced for each jet with low IP due to promptly decaying π_0 mesons). It can be seen that the b -jets are significantly more likely to have particles with larger d_0 than c -jets. This is expected since, as described previously, the b -hadrons have a longer lifetime than the c -hadrons and thus should produce SV s further from the PV which in turn causes the resulting tracks to be more displaced and have greater IPs. The comparison of the signed 3D d_0 with the simulated reco-level, shown in Figure 3.14 on the left plot, yields reassuring similarities but also interesting differences. All three distributions are not Gaussian-like around 0 and, in particular, the light-jets do not exhibit the symmetric Gaussian property used for flipped taggers. This is due to the absence of detector resolution effects at truth-level and thus all of the particles contained in the light-jets have an IP exactly equal or very close to 0. The effect of the detector can also be seen at negative IP values, where the histograms of each jet flavour behave very similarly at reco-level when compared to truth-level.

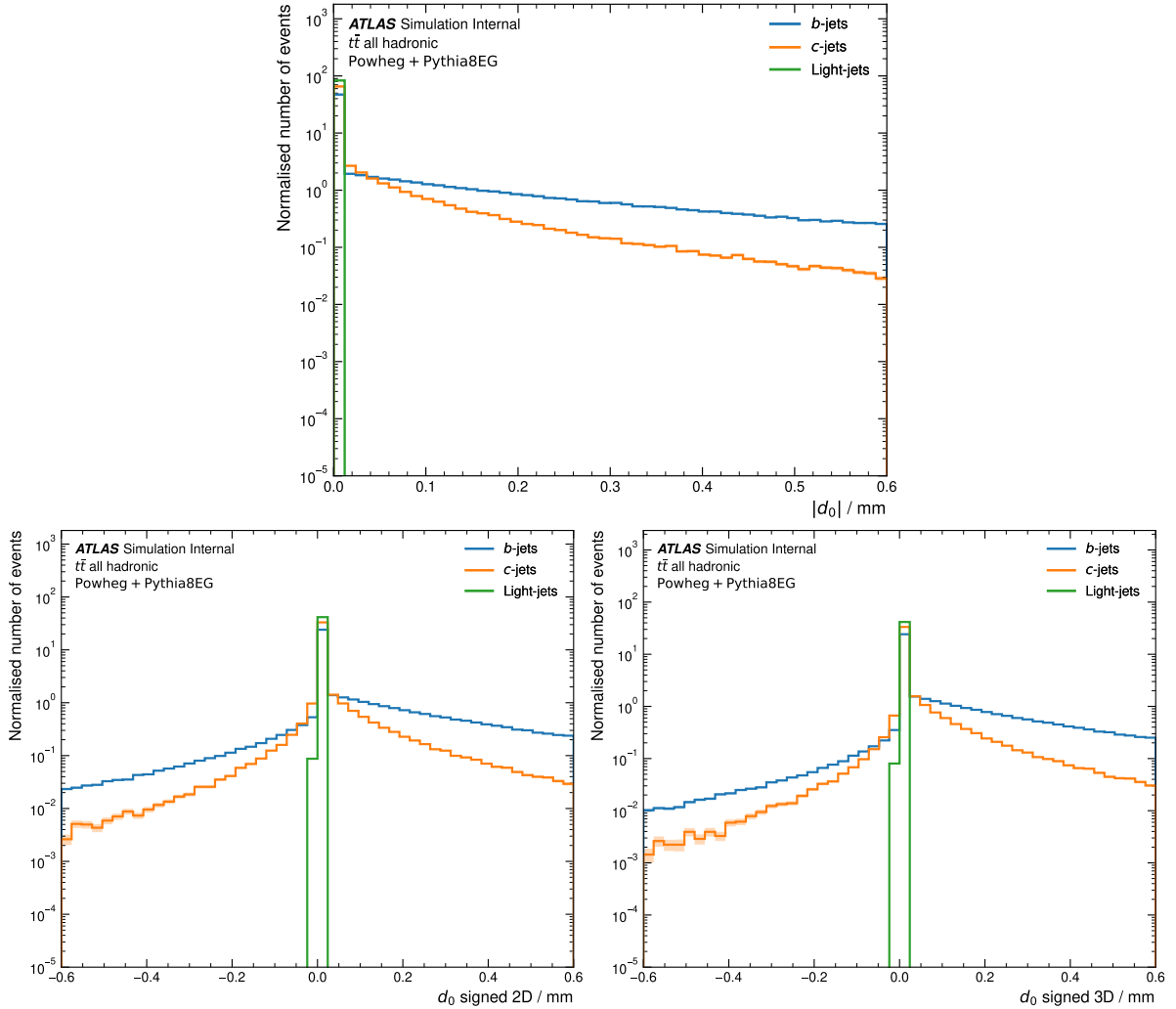


Figure 3.15: Truth-level transverse component of the IP, signed with 2D information (second plot) and 3D information (third plot). The events are $t\bar{t}$ events decaying all-hadronically generated by POWHEG+PYTHIA8.

The Rivet routine can also be used to understand abnormal behaviour in MC-MC maps. For example: section 3.4 mentions the suspiciously low c-jet mistag rates of the SHERPA 2.2.1 sample compared to the other generators in Figure 3.7. The Rivet routine is run over the positive single lepton $t\bar{t}$ decay samples for SHERPA 2.2.1, 2.2.10 and 2.2.12. Ideally, the comparison should also include the nominal generator, but it is not possible here since this exact final state is unique to SHERPA samples. Di-leptonic samples are available for all generators, we do not use them in this example because we are interested in c-jet mistag rates and thus in W bosons decaying hadronically. Figure 3.16 shows the hadron population per type for each generator compared: the first obvious difference is how more than 50% of c-jets contain c-baryons for version 2.2.1

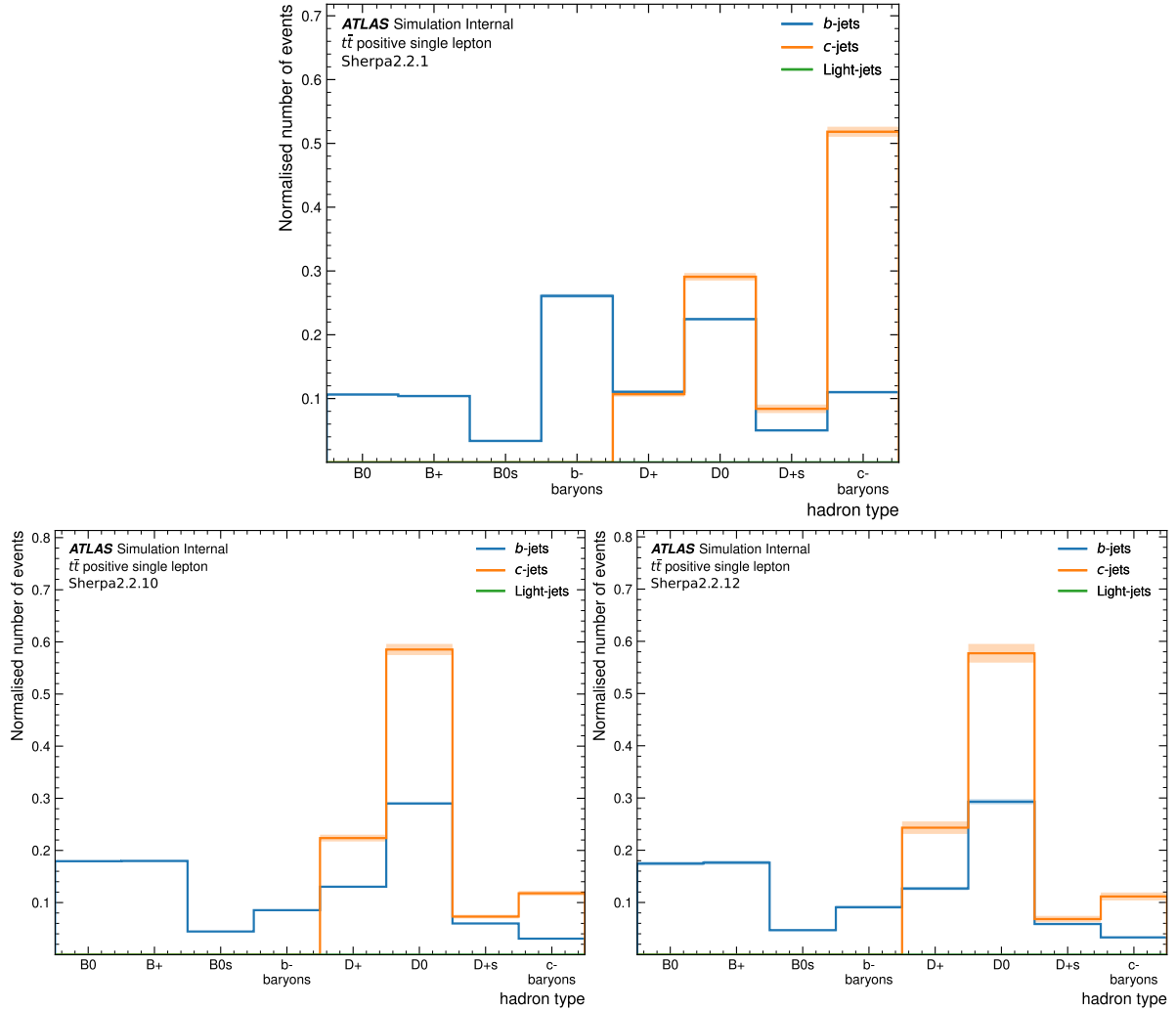
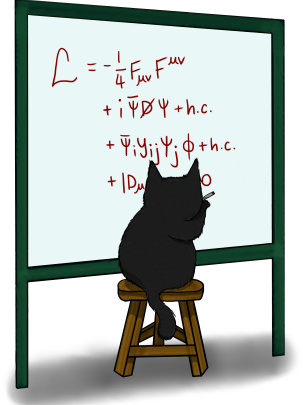


Figure 3.16: The population of each hadron type present in jets taken in $t\bar{t}$ events decaying to a single positive lepton for different SHERPA versions sample. At the top is SHERPA 2.2.1, below are 2.2.10 and 2.2.12.

whereas the other two behave more similarly to the nominal all-hadronic sample, shown in Figure 3.12, with c-jets containing mostly D-mesons. A reasonable hypothesis, explaining the c-jet mistag rate being significantly lower, can be formulated: the c-baryons jets behave physically differently and produce observable distributions that allow for better discrimination from b-jets resulting in lower mistagging. The same can be observed with the number of b-baryon jets being significantly greater than that of B-mesons jets, explaining analogously why the b-tagging efficiencies in Figure 3.6 are much lower than the nominal for that generator unlike the other two SHERPA versions.

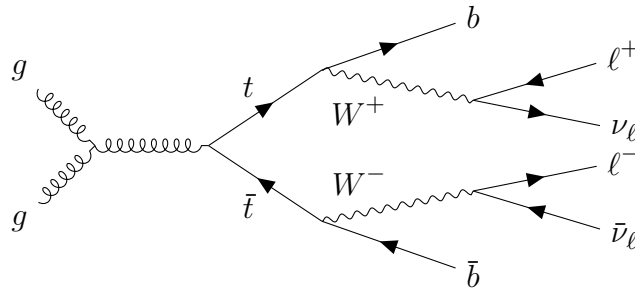
The Rivet routine might have some potential for other applications related to MC-MC efficiency maps. Currently, some generators (or interfacing of generators) do not have an available mapping since it would be difficult and tedious to generate, validate, and maintain $t\bar{t}$ samples for all of them. Individual analyses can compare their samples of interest against each other generator available in the MC-MC efficiency maps and determine which map is closest to their unsupported generators. However, this process can be quite expensive, requires a considerable amount of work from the ATLAS analysers and encourages inconsistencies between analyses. An alternative is proposed here: use the Rivet routine to compute and store strategically chosen and binned histograms for each generator available. Each unsupported sample is then passed through the same Rivet routine and the output histograms compared to the centrally stored histograms for each available generator, computing a “similarity score” using different metrics. This process would be: much cheaper to run, easy for the individual analyser who would run a centralised framework, and consistent across analyses. It is important to clarify that this project remains a proposed idea and requires more work and proofs of concepts that establish the aforementioned metrics required to compute the similarity score.



4 Neutrino Weighting Improvements

4.1 Introduction and definitions

Neutrino Weighting (NW) is a technique that has been used since the D0 experiment, which measured proton-antiproton collision data from 1992 to 2011, to reconstruct the four-momenta of neutrinos in dileptonic $t\bar{t}$ events [153]. The process of $t\bar{t}$ creation via ggF in proton-proton collision, and the leptonic weak decay of both top quarks, is shown as a Feynman diagram below.



The final state is composed of a b -quark, lepton ℓ and corresponding neutrino ν_ℓ per top quark. The sign of the lepton charge depends on whether its parent top quark is a particle or anti-particle, therefore two leptons of opposite charge are expected. The NW method relies on solving the kinematics for the decay chain $t^\pm \rightarrow W^\pm b \rightarrow \ell^\pm \nu b$, individually for both top quarks. A neutrino four-momentum has, of course, four degrees of freedom to resolve. Three physical constraints can be exploited if we assume the particles to be on-shell: the mass of the top quarks, the mass of the W bosons and setting the mass of the neutrinos to zero. This leaves one degree of freedom per neutrino, so two in total for the $t\bar{t}$ event. The resulting systems of two equations (the $m_\nu = 0$ constraint removes

a degree of freedom directly without resulting in an equation) may then be rewritten in terms of the neutrino kinematics: p_x , p_y , η , and rearranged to express p_x and p_y as a function of η . One may then iterate through many solutions $[p_x(\eta), p_y(\eta)]$ by trying different η values numerically. The next step is to define a quantity to optimise in order to find the best solution for $\{[p_x(\eta), p_y(\eta)]_{1,2}\}$, where we labelled the top quark decay as 1, and the anti-top quark decay as 2. Assuming that the missing transverse energy $\vec{E}_T^{\text{miss}}$, measured in the detector, should correspond to the sum of the transverse momentum of the neutrinos, one may write the following comparison:

$$|\Delta_{t\bar{t}}(\eta_{\nu_1}, \eta_{\nu_2})| = |\vec{p}_{T\nu_1}(\eta_{\nu_1}) + \vec{p}_{T\nu_2}(\eta_{\nu_2}) - \vec{E}_T^{\text{miss}}|, \quad (4.1)$$

where $\vec{p}_{T\nu_{1,2}} = (p_{x_{1,2}}, p_{y_{1,2}})$, the transverse momentum of the neutrino from the corresponding top quark. A weight can be defined by wrapping the difference $\Delta_{t\bar{t}}(\eta_{\nu_1}, \eta_{\nu_2})$ into a Gaussian function, mapping the result onto the interval $[0, 1]$: the closer a weight gets to one, the better the solution is at reproducing the missing transverse energy observed in the detector. This weight depends on the kinematics to be optimised, and is referred to as the *NW function* or NW weight in this thesis. The NW weight then compounds two one-dimensional functions: one for each top quark, as shown in equation 4.1. The NW function for di-leptonic $t\bar{t}$ events is then defined as follows:

$$NW(\eta_{\nu_1}, \eta_{\nu_2}) = \exp\left(-\frac{|\Delta_{t\bar{t}}(\eta_{\nu_1}, \eta_{\nu_2})|^2}{2|\vec{E}_T^{\text{miss}}|^2}\right). \quad (4.2)$$

The magnitude of the missing transverse momentum $|\vec{E}_T^{\text{miss}}|$ is used to normalise the Gaussian exponent, which makes the weight distribution well behaved. It can be noted that this is a simple choice that does not have an important impact on the study presented here, and that other quantities can be used for the normalisation instead. For example, one might find a physical observable correlated with the neutrino resolution provided by $\vec{E}_T^{\text{miss}}$, and use it such that the NW weight gets reduced when a low resolution is expected.

In the work presented here, the NW algorithm is adapted to be used in the semi-leptonic decay of $H \rightarrow WW^*$, a physical process presented in detail in section 1.1.2. This is particularly important for the analysis, presented in chapter 6, that makes use of this method. The formulation of NW in the semi-leptonic channel of HWW^* is based on the same technique as $t\bar{t}$, swapping the top quark for the Higgs and the b -jet for the hadronic W (reconstructed experimentally as the sum of two jets) and with only one neutrino

to solve for. The leptonic W will be off-shell roughly half of the time and, in fact, the proposal described in chapter 5 aims at final states where the leptonic W is off-shell. This changes the mass of the boson from a constraint to a parameter that needs to be optimised. Since there is only one neutrino, there are four degrees of freedom to recover in total. Following a similar procedure as for dileptonic $t\bar{t}$, one may assume that the Higgs is on-shell and set the neutrino mass to zero, which leaves two degrees of freedom in total. An equation may be written for the Higgs decay and another one for the leptonic W decay. The system may then be rearranged to express, as before, the components of the transverse momentum of the neutrino: p_x and p_y , as a function of the neutrino pseudo-rapidity η , and the potentially off-shell mass of the leptonic W boson m_{W_ℓ} . It is important to note that, unlike in the dileptonic $t\bar{t}$ case with two individual systems of equations (one per parameter), here the two unknown parameters that constitute the degrees of freedom are coupled in a single system of two equations. Analogous to the $t\bar{t}$ case, one may then define the comparison to $\vec{E}_T^{\text{miss}}$ as:

$$|\Delta_{h \rightarrow WW^*}(\eta_\nu, m_{W_\ell})| = |\vec{p}_{T\nu}(\eta_\nu, m_{W_\ell}) - \vec{E}_T^{\text{miss}}|, \quad (4.3)$$

where there is only one neutrino whose transverse momentum depends on the two parameters to optimise for: η_ν and m_{W_ℓ} . The NW function is then given by:

$$NW(\eta_\nu, m_{W_\ell}) = \exp \left(-\frac{|\Delta_{h \rightarrow WW^*}(\eta_\nu, m_{W_\ell})|^2}{2|\vec{E}_T^{\text{miss}}|^2} \right), \quad (4.4)$$

where $\vec{E}_T^{\text{miss}}$ is used to normalise the exponent similarly to the $t\bar{t}$ case.

4.2 Using MINUIT

The NW algorithm has been used to reconstruct the neutrinos and access interesting spin observables in multiple $t\bar{t}$ analyses [153, 154]. In its current formulation, the NW algorithm relies on an optimisation performed for each event by a simple grid search with between four and six parameters. This process is quite CPU-intensive and time consuming, in particular when running on large MC samples with several systematic variations. Due to the large computational power and run time required for NW, analyses now use other methods (namely the *Sonnenschein* [155] and *Ellipse* [156] methods), which find neutrino reconstruction solutions for a strictly smaller set of events. From now on, the fraction of events for which the algorithm finds a reconstruction solution will be

referred to as the *reconstruction rate*. There is space for investigation of alternative, more sophisticated optimisation methods to use with the NW algorithm instead, and bring back this technique as an accessible tool to collider analyses. In particular, looking at Figure 4.1, it can be seen that the NW function, to be optimised, tends to be relatively well behaved and resembles a smooth manifold. This indicates that *gradient-based* methods can be used to perform the optimisation. Gradient-based methods rely on the differentiability (smoothness) of a function to find an extremum. Specifically, the existence of a first-order derivative, and possibly higher orders, is leveraged to travel on the hyper-surface that constitutes the function to be optimised, and reach the optimum sought after in that manner.

In this work, the MIGRAD minimiser of the MINUIT software [157] is used. MIGRAD estimates the gradient and the Hessian matrix numerically at every point to find the minimum. The reason why MINUIT was chosen is because this study is a proof of concept and the package is easily accessible through the CERN ROOT library [158], the latter being omnipresent within the ATLAS software. It is important to note that other gradient-based methods could improve the efficiency even further and the results presented here are thus a baseline, showing potentially pessimistic performance. A custom optimiser, making use of an analytically computed Hessian matrix for the NW function and the Newton method [159], is built in section 4.3 in an effort to maximise efficiency by not relying on the numerical computation of the Hessian matrix (and gradient) at every step, and to compare results with the MINUIT implementation.

Crucially, every gradient-based method needs a starting point on the hyper-surface being optimised. Figure 4.1 shows that the NW function is not defined at every point, some NW weights are complex-valued and displayed as being equal to -1 , which makes finding an initial starting point challenging. Two solutions to this complication are provided for both the dileptonic $t\bar{t}$ and semi-leptonic HWW^* : a first simple method that uses a cheap line-search, and a second method that uses an analytical solution for the initial point.

4.2.1 In dileptonic $t\bar{t}$

NW function with smearing

Due to detector effects, the energy of the final reconstructed objects are subject to smearing, and thus the condition on the top quarks and W bosons to be considered

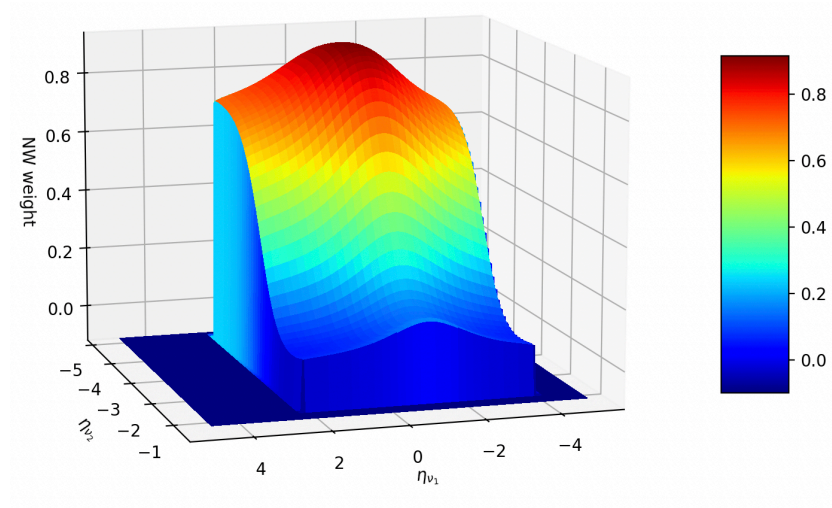
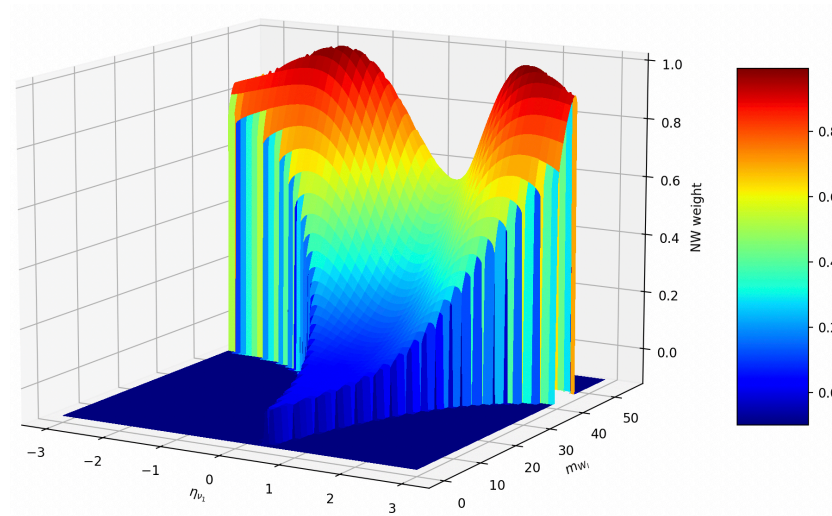
(a) Dileptonic $t\bar{t}$ event.(b) Semi-leptonic HWW^* event.

Figure 4.1: The NW function for a single dileptonic $t\bar{t}$ event (a), parametrised by the two neutrino pseudo-rapidities, and for a single semi-leptonic HWW^* event (b), parametrised by the neutrino pseudo-rapidity and the mass of the leptonic W . All the unphysical complex-valued solutions are set to -1 , resulting in the dark-blue flat surface “below” the shape of interest. The shape of the NW function can change drastically between different events in both cases.

exactly on-shell is too strict, resulting in events failing the reconstruction. The smearing of top and W masses themselves is then important to improve the reconstruction rate. In current analyses, this is done using Gaussian sampling around the expected mass peak of both masses respectively (for each top decay, so four masses to smear in total). In terms of complexity, this is equivalent to using a grid search in six dimensions which is quite expensive. Using MINUIT, the smearing of the top and W masses can be effectively performed by extending the parameter space instead: $\{\eta_{\nu_1}, \eta_{\nu_2}\} \rightarrow \{\eta_{\nu_1}, \eta_{\nu_2}, m_t, m_{\bar{t}}, m_{W^+}, m_{W^-}\}$ and imposing physical constraints on m_t , $m_{\bar{t}}$ and $m_{W^\pm}$. In other words, by enforcing $m_t, m_{\bar{t}} \in [172.0 - \delta_t, 172.0 + \delta_t]$ GeV and $m_{W^\pm} \in [80 - \delta_W, 80 + \delta_W]$ GeV, where δ_t and δ_W are adjustable parameters and whose determination can be informed by simple studies on the detector effects. As observed in the studies displayed in section 4.4, it is mainly because of this extension to six dimensions that the potential performance of gradient-based optimisers stands out as being much more efficient due to their excellent scaling (compared to grid searches). The results presented in section 4.4 include top mass smearing (but not for the W mass).

Finding a starting point

As briefly discussed in section 4.2, the NW algorithm solutions for p_x and p_y of either neutrino can be complex-valued (unphysical) and finding an initial point in the valid region of the NW function domain is thus not trivial. A real-valued initial point $(\eta_{\nu_1}, \eta_{\nu_2})$ can be found on the surface by making use of the fact that the functions $\vec{p}_{T\nu_1,2}(\eta_{\nu_1,2})$ are one-dimensional, and perform a cheap line search (i.e a 1D grid search) for both function simultaneously. Alternatively, an analytical solution (the derivation of which is given further below) is proposed to compute the component of the neutrino momentum p_x , p_y and p_z corresponding to the starting point in a valid part of the domain (from which the initial η_ν can easily be computed). The analytical solution is given by:

$$p_z = \frac{-b}{2a}, \quad p_{y1,2} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \quad p_{x1,2} = \frac{-p_y q_y - p_z q_z - c_3}{q_x} \quad (4.5)$$

with :

$$a = \left(\frac{2E_k^2 q_y q_z}{q x^2} - 2\alpha_y \alpha_z \right)^2 - 4A \left(E_k^2 \left(1 + \frac{q z^2}{q x^2} - \alpha_z^2 \right) \right), \quad (4.6)$$

$$b = -4A \left(\frac{2c_3 E_k^2 q_z}{q_x^2} - 2c_4 \alpha_z \right) + \frac{4c_3 E_k^2 q_y \left(\frac{2E_k^2 q_y q_z}{q_x^2} - 2\alpha_y \alpha_z \right)}{q_x^2} - 4c_4 \alpha_y \left(\frac{2E_k^2 q_y q_z}{q_x^2} - 2\alpha_y \alpha_z \right), \quad (4.7)$$

$$A = \left(1 + \frac{q_y^2}{q_x^2} \right) E_k^2 - \alpha_y^2, \quad (4.8)$$

$$B = p_z \left(\frac{2q_y q_z}{q_x^2} E_k^2 - 2\alpha_y \alpha_z \right) + 2\frac{q_y}{q_x} c_3 E_k^2 - 2\alpha_y c_4, \quad (4.9)$$

$$C = p_z^2 \left(\left(1 + \frac{q_z^2}{q_x^2} \right) E_k^2 - \alpha_z^2 \right) + p_z \left(2\frac{q_z}{q_x} c_3 E_k^2 - 2\alpha_z c_4 \right) + \frac{c_3^2 E_k^2}{q_x^2} - c_4^2, \quad (4.10)$$

$$k^\mu = l^\mu + b^\mu, \vec{q} = \frac{\vec{p}_k}{E_k} - \frac{\vec{p}_l}{E_l} \quad (4.11)$$

$$\alpha_i = \frac{q_i}{q_x} k_x + k_i, \quad (4.12)$$

$$c_1 = \frac{m_t^2 - m_k^2}{2}, \quad c_2 = \frac{m_W^2}{2}, \quad c_3 = \frac{c_1}{E_k} - \frac{c_2}{E_l}, \quad c_4 = c_3 \frac{k_x}{q_x} + c_1. \quad (4.13)$$

These computations yield two solutions per neutrino, so four points to test in total. The results presented in section 4.4 use the average of the solutions (i.e the centre of the rectangle formed by the four points in the η_{ν_1} - η_{ν_2} plane). This analytical solution and the one presented in section 4.2.2 for HWW^* are particularly useful when running the NW algorithm on background processes with lower reconstruction rates: if there are no solutions, events are evaluated and skipped in $O(1)$ complexity, against the current $O(N^{2+m})$ where N is the *grid size* (i.e number of points sampled per dimension optimised over) and m the number of additional physical parameters considered for smearing. Concerning the ideal case, which includes top and W mass smearing, one can use the given analytical solution to find a starting point by either assuming the particles to be exactly on-shell or with a sparse 4D grid search (which can be stopped as soon as a point is found). The latter is recommended to have optimal reconstruction efficiency and

is the method used in the results presented in section 4.4 (only for top mass smearing, so a 2D sparse grid search). In the derivation of the analytical method, presented in the following part, an improvement is suggested to remove the need for this exploration of smearing dimensions.

Derivation of the starting point analytical solution

The details of the neutrino kinematic equations used in the NW framework can be found in [153]. The main difficulty in manipulating the expressions algebraically and finding an analytical solution from the equations in that framework comes from the usage of the transcendental hyperbolic functions cosh and sinh, that appear from the decision to use the neutrino η as a variable. The first step is then to rewrite the equations in terms of p_x , p_y and p_z instead. Starting from the on-shell mass equations using the four-momenta of particles, one may write:

$$\begin{aligned} t^2 &= (k + \nu)^2 = m_t^2 \\ W^2 &= (\ell + \nu)^2 = m_W^2, \end{aligned} \tag{4.14}$$

where $k = \ell + b$, t , b , ℓ , W and ν are the four-momenta of the relevant particles. One can linearise in terms of the variables of interest by getting rid of the energy term for the neutrino E_ν , by combining both equations:

$$E_\nu = \frac{1}{E_k}(c_1 + \vec{p}_k \cdot \vec{p}_\nu) = \frac{1}{E_\ell}(c_2 + \vec{p}_\ell \cdot \vec{p}_\nu), \tag{4.15}$$

where the definition for the constants c_i are given in equation 4.13. After this step, one can expand the three-momenta of the neutrino and rearrange to obtain a linear expression of the form $p_y = a_z p_x + b_z$, where both a_z and b_z depend on p_z . Substituting that expression to eliminate p_x in the top quark mass equation, one obtains the quadratic form whose solution for p_y is shown in equation 4.5, where B and C depend on p_z . The square root term is the source of complex-valued solutions that are unphysical and discarded. One can thus establish that if the maximum value of the function $f(p_z) = B(p_z)^2 - 4A \cdot C(p_z)$ is negative then no valid solution exists for the event. Expressing $f(p_z)$ explicitly and collecting the terms in powers of p_z yields another quadratic expression of the form $f(p_z) = ap_z^2 + bp_z + c$, where the constants a and b are given in equations 4.6 and 4.7 respectively. It is important to note that here the assumption is made that the quadratic has a well-defined maximum (i.e $a < 0$) for events with an existing physical solution.

The maximum of f then occurs at the value for p_z given by the simple expression in equation 4.5. There is room to improve the method further at this step by extending the maximisation to the other two parameters of interest m_t and m_W used in smearing, i.e one would have to solve the linear system of equation $\vec{\nabla} f(p_z, m_t, m_W) = 0$. Finally, the solution for p_x and p_y can trivially be recovered by using the value for p_z , and together be used to compute η_ν for each neutrino. The quadratic nature of the solution for p_y means that two solutions per neutrino exists, so four solutions are obtained in total.

4.2.2 In semi-leptonic HWW^*

NW function with smearing

In this process, only the Higgs mass m_h may be smeared by adding a dimension to the grid searches or extending the parameter space in gradient-based methods with a constraint $m_h \in [125.5 - \delta_h, 125.5 + \delta_h]$ GeV. For this case, the gradient-based methods are expected to be considerably faster than grid searches even without smearing. This is due to the coupling of the parameters η_ν and m_{W_ℓ} that forces the grid search to be two dimensional at minimum.

Finding a starting point

In the $t\bar{t}$ case, described in section 4.2.1, it was presented that the dependency is “separable” into two one-dimensional functions, resulting in the real solution to dwell in rectangular regions over the η_{ν_1} - η_{ν_2} phase-space, as can be seen in the subplot (a) of Figure 4.2. This separability is not present in equation 4.4 and the shape of the region holding real solutions is thus a bit more interesting, as can be seen from the subplot (b) of Figure 4.2. Fortunately, the function has a very useful property: the region of real solutions concentrates around the lepton pseudo-rapidity η_l as $m_{W_\ell} \rightarrow 0$, which can be seen in Figure 4.2. This seems to suggest that for reconstructable events (i.e events with real-valued solutions), there should always be real solutions along the line $\eta_\nu = \eta_l$. One may then find an initial point (η_ν, m_{W_ℓ}) to start the MINUIT optimisation by performing a cheap line search along that line. Alternatively, as for the $t\bar{t}$ case, an analytical solution is proposed (the derivation being presented further below):

$$\eta_\nu = \eta_l, \quad m_{W_\ell} = -\frac{B_W}{2A_W}, \quad (4.16)$$

with :

$$A_W = k_T^2 - E_k^2 \cosh(\eta_l)^2 - k_z^2 \sinh(\eta_l)^2 + E_k k_z \sinh(2\eta_l), \quad (4.17)$$

$$B_W = (m_b^2 - m_h^2 + 2b^\mu l_\mu)(\alpha - \beta \cosh(2\eta_l) + \gamma \sinh(2\eta_l)), \quad (4.18)$$

$$\alpha = -(E_b E_l) - E_l^2 + 2(\vec{b}_T \cdot \vec{l}_T + l_T^2) + b_z l_z + l_z^2, \quad (4.19)$$

$$\beta = (E_b E_l + E_l^2 + l_z(b_z + l_z)), \quad (4.20)$$

$$\gamma = (E_b l_z + E_l(b_z + 2l_z)), \quad (4.21)$$

where b^μ is the hadronic W boson four-momentum and $k^\mu = b^\mu + l^\mu$. It is important to note that the $\eta_\nu = \eta_l$ line search and the analytical solution are very useful to reject non-reconstructable events efficiently without performing a full grid search. In the $HW W^*$ case, roughly 58% of events are reconstructable without mass smearing, as opposed to between 70 and 90% for $t\bar{t}$, depending on smearing. The reconstruction rate falls even lower for backgrounds (around 12 and 9% for W +jets and Z +jets respectively) which tend to require many more simulated MC events, making this feature very valuable to save overall computation time.

Derivation of the starting point analytical solution

Finding an analytical solution for the $H \rightarrow WW^*$ case is a bit more tricky than in $t\bar{t}$ because of how the parameters are coupled in a system of equations. One could follow the same strategy as in $t\bar{t}$ and start from the solutions given in equation 4.5, set $f(p_z, m_{W_\ell}) = B(p_z, m_{W_\ell})^2 - 4A \cdot C(p_z, m_{W_\ell})$ and solve the linear system of equation given by $\vec{\nabla} f(p_z, m_{W_\ell}) = 0$, noting that this calculation would also be useful for the smearing of the W mass in the $t\bar{t}$ case. Alternatively, a simpler solution is possible by noticing that for every reconstructable event, there is always a large range of real solutions along the line $\eta_\nu = \eta_l$ as shown in Figure 4.2. Setting $\eta_\nu = \eta_l$ makes the problem one dimensional and $f(m_{W_\ell})|_{\eta_\nu=\eta_l}$ may be rearranged as a quadratic form by collecting powers of m_{W_ℓ} : $f = A_W m_{W_\ell}^2 + B_W m_{W_\ell} + C_W$, where A_W and B_W are

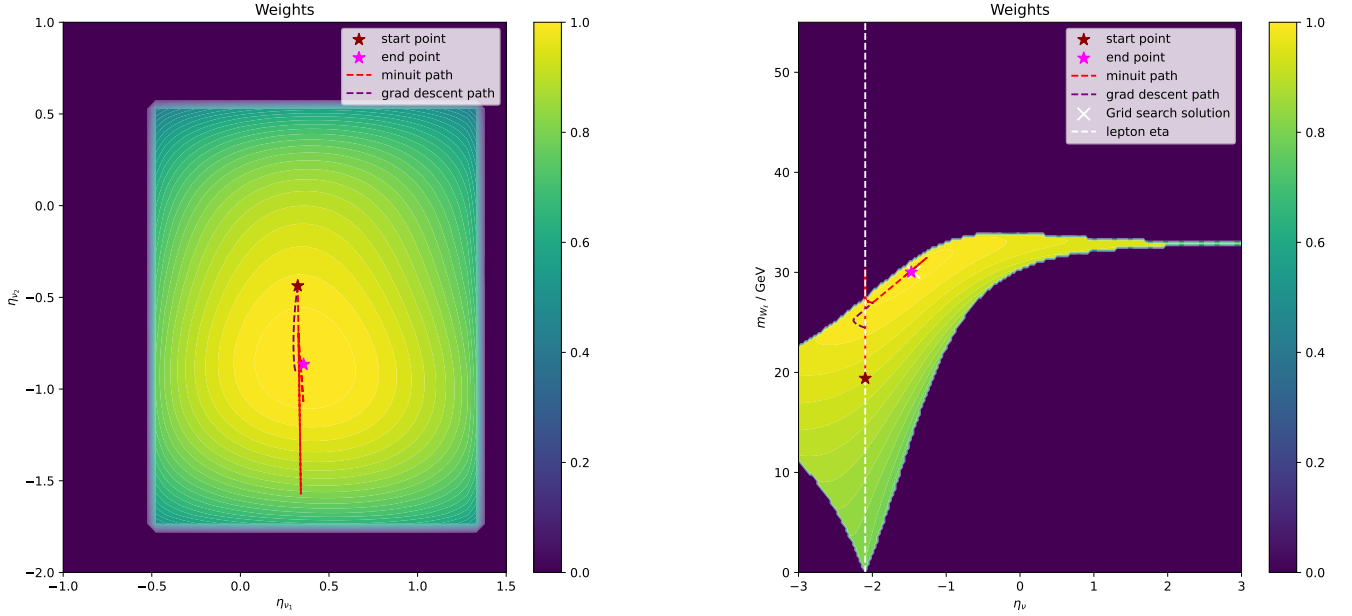
(a) Dileptonic $t\bar{t}$ event.(b) Semi-leptonic HWW^* event.

Figure 4.2: The NW function for a dileptonic $t\bar{t}$ event (a), parametrised by the two neutrino pseudo-rapidities, and for a semi-leptonic HWW^* event (b), parametrised by the neutrino pseudo-rapidity and the mass of the leptonic W . The MIGRAD starting and end point are given as well as the path taken by the algorithm. A naive gradient descent is also shown in purple to compare with the MINUIT path.

constants defined in equations 4.17 and 4.18 respectively. Finally, the solution for an initial point is then more simply given by equation 4.16.

4.3 Hessian matrix and the Newton Method

The NW algorithm relies on the computation of several terms that originate from on-shell mass constraints given by the system of equation 4.14 for $t\bar{t}$. These terms can be partially differentiated twice with respect to η_ν and m_{W_ℓ} , such that the Hessian matrix is computed analytically. This has been done for the semi-leptonic $H \rightarrow WW^*$ topology, and is used to produce a custom optimiser. Several optimisation algorithms can make use of this information to find the NW maximum. One such algorithm is *Newton's method*, which

takes the following simple formulation :

$$\vec{x}_{n+1} = \vec{x}_n - H^{-1} \nabla NW(\eta_\nu, m_{W_\ell}), \quad (4.22)$$

where $\vec{x}_n = (\eta_\nu, m_{W_\ell})_n$ is the parameter vector at the n th step and H the Hessian. Since the NW function may have saddle points where the derivative goes to zero, the Hessian matrix can become non-definite negative (i.e its eigenvalues are not all negative) and the algorithm gets stuck. This issue can be addressed by perturbing H whenever one of its eigenvalues gets close to 0 or becomes positive:

$$H \rightarrow H + \alpha \cdot |\max(\{\lambda_i\})| \cdot \mathbb{I}, \quad (4.23)$$

where α is a real number greater than 1, $\{\lambda_i\}$ the set of eigenvalues of H and $\mathbb{I}$ the identity matrix. This makes the Newton step behave more like gradient descent and enables the optimisation to escape saddle points with a little “push”. The calculation of H^{-1} and the eigenvalues of H is quite efficient in our case where the NW function has 2 parameters (4(3) if including Top(Higgs) mass smearing).

Just like MIGRAD and other algorithms that navigate the NW surface using its derivative, a step may “overshoot” and land outside the region of real solutions. For example, in Figure 4.2 (b), the optimum is very close to the edge of the domain of valid solutions, and the algorithm may find a next step with $m_{W_\ell} > 40$ GeV, which would yield an unphysical momentum. When this happens, a possible solution is to pick a new point between the current and next coordinates calculated as follows:

$$\vec{x}_{n+2} = \beta(\vec{x}_{n+1} - \vec{x}_n), \quad (4.24)$$

where $\vec{x}_i$ is the vector of parameters at step i , $n + 1$ is the step outside the region of real solutions and $\beta \in [0, 1)$. Setting $\beta = 0.5$ then corresponds to picking the point halfway between both steps. This can be repeated until a real solution is computed and the optimisation may continue. In order to stop the algorithm, a tolerance on the step size ϵ_{tol} can be defined such that if, for a step i , $|\vec{x}_{i+1} - \vec{x}_i| \leq \epsilon_{\text{tol}}$ then the algorithm terminates and $\vec{x}_{i+1}$ is the final result for the parameters. All the ingredients used to build the custom algorithm, which will now be referred to as $hWW\text{NEWTON}$, have now been laid out. The method relies on three simple parameters: α , β and ϵ_{tol} . $hWW\text{NEWTON}$ is also evaluated and compared to the other methods in section 4.4, better run time performances are expected over MIGRAD, which relies on computing the Hessian numerically by evaluating the kinematics multiple times per steps. The

calculation of the Hessian for the semi-leptonic $H \rightarrow WW^*$ topology can be used to build the $t\bar{t}$ NEWTON algorithm.

4.4 Performance study

A benchmarking software was developed to compare different methods and measure the run time in a realistic setting. The software is designed to be user-friendly and for repeated use: it reads a simple YAML configuration file where each method is parametrisable. The code is written with a concern for modularity: other neutrino reconstruction methods can be appended to the framework easily. In particular, the Sonnenschein [155] and the Ellipse [156] analytical methods, which rely on algebraic and geometric concepts respectively, were added to the software to include them in future benchmarking. These methods have a smaller reconstruction rate than the NW algorithm and are only used in $t\bar{t}$ analysis to reduce run time. They will not be studied here since this work is about the improvement of the NW method. All the results here are obtained by running the algorithms on reco-level MC samples that resulted from the full ATLAS simulation chain, described in more detail in section 2.7. Some simple metrics have been defined to compare the performance of the different algorithms: the *mean duration*, the *mean best weight* found and the reconstruction fraction introduced earlier (abbreviated as “reco fraction”). The mean duration and mean best weight are given by the average time taken by the algorithm to run and average best NW weight found respectively, for all events where a solution exists. Dileptonic $t\bar{t}$, ggF and VBF semi-leptonic $H \rightarrow WW^*$ reco-level MC events are produced with the ATLAS tool chain. The importance of the missing transverse energy in the NW algorithm makes the realistic simulation of its detector resolution equally crucial. For $t\bar{t}$ events, the lepton and b -jets are paired (one pair for each top) by minimising the ΔR between the two physical objects. In practice, a more advanced way to pair leptons and b -jets is by running the NW algorithm on both configurations and picking the solution with the highest weight. This is of little importance in our case since we seek to optimise run time. For semi-leptonic HWW^* events, the hadronic W consists of the sum of the four-momenta of two jets objects which are selected by the JetPairingNet neural network, presented in more details as part of the analysis of chapter 6 in section 6.6.1.

4.4.1 Dileptonic $t\bar{t}$

The algorithm performing a grid search on $t\bar{t}$ events is called $t\bar{t}$ GridSearch, and the grid in the η_1 - η_2 plane (the pseudo-rapidity of both neutrinos) is made of 20×20 linearly spaced points with range $\eta_1, \eta_2 \in [-3, 3]$. The $t\bar{t}$ GridSearch_{MS} algorithm also performs a grid search but with the addition of two mass smearing dimensions, one per top quark, with five points between 162.5 and 182.5 GeV. Only five points are used to establish a more favourable scenario for the grid search algorithms, since grid search algorithms scale very badly with additional dimensions. The results presented here can then be treated as a lower limit on the performance gained by switching to gradient-based algorithms. The algorithm using MIGRAD to perform the optimisation instead of a grid search is referred to as $t\bar{t}$ MINUIT, and $t\bar{t}$ MINUIT_{MS} refers to the same algorithms but with two additional parameters to solve for: one for each top quark mass. The top quark masses are both constrained between 162.5 and 182.5 GeV (effectively setting $\delta_t = 10$ GeV), capping the amount of mass smearing allowed for the solutions considered. The gradient-based methods use the analytical solution, presented in section 4.2.1, to find a starting point on the hyper-surface in the domain of real solutions. All algorithms compared here can either differ in their method of optimisation or on whether they use top mass smearing (or both), but all share the same core mechanism: solving the NW kinematics. It is thus important to check that the solutions outputted by the different algorithms are consistent with respect to one another. This consistency is expected for a reasonably high grid density, and reasonably low tolerance for the gradient-based algorithms. The following distance is defined to study the consistency of the algorithms with one another:

$$s_w = \sqrt{\left(\frac{\delta\eta_{\nu_1}}{6}\right)^2 + \left(\frac{\delta\eta_{\nu_2}}{6}\right)^2}, \quad (4.25)$$

where $\delta\eta_{\nu_i}$ represents the distance between the solutions of two algorithms along the parameter's axis. The distances are weighted by a factor of six, since it is equal to the length for the grid used by the grid search algorithms. Intuitively, most neutrino solutions are expected to lie within a “maximum” distance of $s_{w_{\max}} = \sqrt{2}$, which can then be seen as an upper bound for difference between algorithm solutions.

Figure 4.3 shows the distribution of s_w , which reveals interesting details about the algorithm comparisons. It is clear that the means (red bars) are consistently greater than the medians (blue bars), being inflated by data points lying in the long tails of the distributions. Despite the high mean distance between $t\bar{t}$ GridSearch and $t\bar{t}$ MINUIT,

distribution 1 shows that most values lie below the line $s_w = 0.1$, indicating that the algorithms are actually quite similar in their outputs. Distribution 0, comparing grid search with and without top mass smearing, displays a unique undulating shape, which can be explained by understanding the method used to obtain the “violins”. The plots are computed with Gaussian kernel density estimators [160] and Scott’s rule [161], assuming the inputted data to follow a smooth distribution. When comparing two grid search algorithms, the distributions are far from smooth when s_w is close to 0, since the data points accumulate discretely at grid points, separated by distances s_w that are multiples of $\frac{1}{20}$ along the optimised parameter axes and multiples of $\frac{\sqrt{2}}{20}$ diagonally, for a grid sampling of 20 points. The appearing multimodality is then artificial and the peaks occur at values that suggest that for most events the algorithm disagrees by one grid step only. Distributions 2 and 3 show the greatest difference, which is expected since they compare algorithms with different optimisation mechanisms as well as opposite mass smearing (top mass smearing against none). Distribution 4 has a similar tail to 2 and 3, producing a similarly high mean and median. However, the distribution shape flattening more towards low s_w values indicates that the algorithms are more similar than in distributions 2 and 3. This is reassuring since, while they use different optimisation techniques, they both make use of mass smearing. This suggests that the source of differences between algorithms is mainly due to the optimisation technique used rather than its potential usage of mass smearing. Finally, distribution 5 shows the greatest agreement with a peak very close to zero and a short tail becoming invisible around the line $s_w = 0.3$. This corroborates the idea that the main source of difference between algorithms comes from the choice of optimiser.

Now that it is clear that the outputs of all $t\bar{t}$ NW algorithms are consistent with each other, their performances can be assessed. Table 4.1 shows the different metrics for each algorithm executed on $t\bar{t}$ events. The row displaying the mean durations (run times) shows that the grid search-based optimisation is faster than the gradient-based methods without mass smearing. This is expected as the grid searching can be made very efficient when performed only on the η_1 - η_2 plane by first looking for solutions for both top quarks simultaneously, making use of the fact that the $\vec{p}_{T\nu_1}(\eta_{\nu_1})$ and $\vec{p}_{T\nu_2}(\eta_{\nu_2})$ are independent of each other in equation 4.2. The NW function is then only evaluated for all possible combinations of $\vec{p}_{T\nu_1}$ and $\vec{p}_{T\nu_2}$ found. When adding the top quark mass dimensions, the kinematics equations get coupled and it is no longer possible to solve them simultaneously which results in a considerable increase in run time. The mean run times for the optimisations including the mass smearing dimensions are significantly greater as expected. $t\bar{t}\text{GridSearch}_{\text{MS}}$ has a very high average run time of roughly 1900 μs

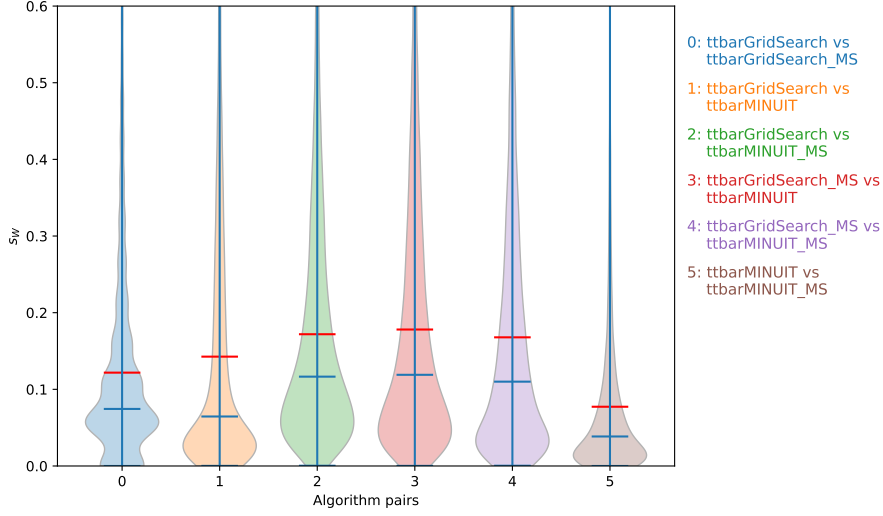


Figure 4.3: Violin plots displaying the distribution for the distance s_W between the solution of different algorithm pairs. The red line shows the mean of the distribution while the blue line highlights its median

whereas $t\bar{t}$ MINUIT_{MS} is faster by a factor of more than two on average with a mean of about $780 \mu\text{s}$. The strength of the good scaling of time complexity of gradient-based optimisation with respect to the number of parameters can be seen in these results: adding the mass smearing multiplies by a factor of almost twenty the average run time for grid search-based algorithms while the gradient-based algorithms increase only by a factor of roughly two. In practice, the mass smearing might also happen for the W bosons which would add two additional dimensions and thus make the usage of gradient-based methods even more beneficial than the more naive alternative.

$t\bar{t}$ -	GridSearch	GridSearch _{MS}	MINUIT	MINUIT _{MS}
mean duration (μs)	114.044	1914.761	436.29	782.757
mean best weight	0.813	0.909	0.837	0.899
reco fraction	0.873	0.907	0.877	0.907

Table 4.1: Metrics computed using roughly 890,000 $t\bar{t}$ events in total. The mean duration and best weight are computed by taking into account successfully reconstructed events only.

Figure 4.4 shows the distribution of run times for $t\bar{t}$ events where a NW solution was found for the algorithms. The shapes of the curves resemble one-sided crystal ball functions. One might expect the grid searching algorithms to have Gaussian distributions

that would be caused by a constant number of iterations. However, because the trick of solving for the kinematics simultaneously in the η_1 - η_2 plane is used, the number of steps is not constant between events. The plot illustrates the great scaling of gradient-based methods. Adding the top quark mass smearing with a grid search optimisation is represented by going from the blue curve peaking around 100 μs to the flat orange curve with run times extending up to 3000 μs , whereas doing so when optimising with MINUIT results in going from the green curve to the red curve.

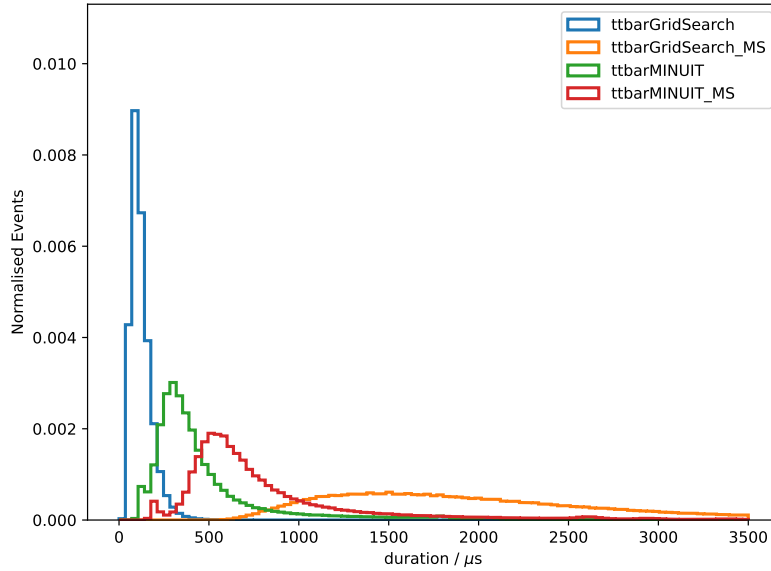


Figure 4.4: Distribution of algorithm run times for $t\bar{t}$ events with a NW solution.

The mean best weights are higher for the algorithms with mass smearing. This is expected since mass smearing simply extends the phase-space explored for the best possible weight. The average best weight are not significantly different between $t\bar{t}$ GridSearch and $t\bar{t}$ MINUIT, as well as between $t\bar{t}$ GridSearch_{MS} and $t\bar{t}$ MINUIT_{MS}, which is also expected from the analysis of consistency between algorithms performed with the violin plots in Figure 4.3.

The percentage of reconstructed events is about 88%, which increases by 3% when including mass smearing. The algorithm using MINUIT without mass smearing has a slightly larger reconstruction rate than the grid search-based counterpart, by 0.4%. These events have one side of their rectangular real solution domains thinner than the grid spacing such that no grid points find a real solution. The mass smearing increases

the probability of finding these rare edge-cases as can be seen from the fact that the reconstruction rates are identical for both mass smearing algorithms. This observation highlights the robustness of the analytical method for finding initial solutions even when the area of the domain of real solutions is very small. Figure 4.5 shows the distribution of one of the two top quark masses found by the algorithms. The events accumulate mostly on the edges of the constrained range with a noticeable overall preference towards lower mass values. This indicates that the solutions with the highest weights are located outside the constrained range. Loosening the constraint on the top mass parameters can be considered but needs to be carefully studied, it is possible that the resolution of the neutrino solutions gets negatively impacts if the masses are smeared too far away from the nominal expected value of 172.0 GeV. The situation can become more complicated with the addition of the smearing of the W boson masses.

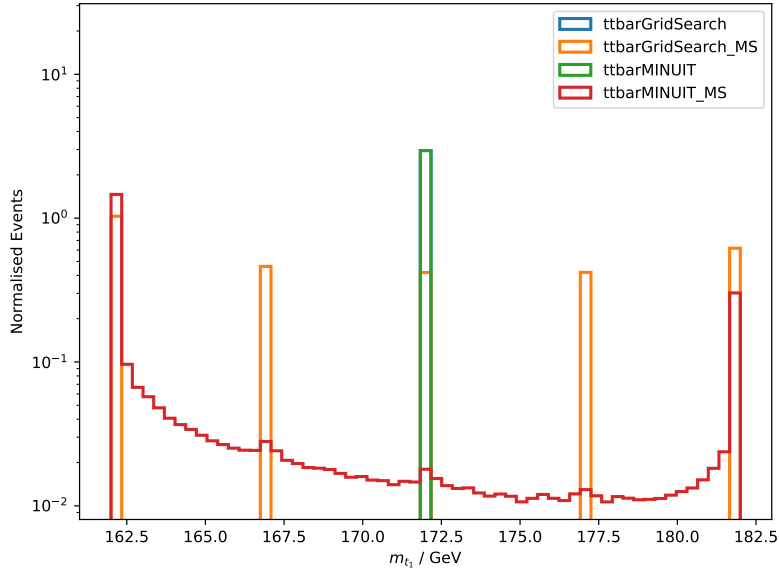


Figure 4.5: Distribution of the best top quark mass found. The algorithms without mass smearing all have $m_{t_1} = 172.0$ GeV exactly, which is why the blue histogram is not visible, hidden by the green distribution.

4.4.2 Semi-leptonic HWW^*

The algorithm that performs a grid search on $H \rightarrow WW^*$ events is called hWW GridSearch, and the grid is defined in the η - m_{W_ℓ} plane following a 20×20 set of linearly spaced points

with ranges: $\eta \in [-3, 3]$ and $m_{W_\ell} \in [0, 50]$ GeV. The $hWW\text{GridSearch}_{\text{MS}}$ algorithm is also a grid search but including a Higgs mass smearing dimension, with five points between 124 and 126 GeV. Just like for $t\bar{t}$, only five mass points are used to keep the run time realistic. The algorithm that makes use of MIGRAD to perform the optimisation is referred to as $hWW\text{MINUIT}$. $hWW\text{MINUIT}_{\text{MS}}$ has the Higgs mass as an additional parameter, which is constrained between 124 and 126 GeV to limit the mass smearing. $hWW\text{Newton}$ is the custom optimiser that uses the Newton method with the elements described in section 4.3. For the results presented in this section, $\alpha = 2.0$, $\beta = 0.5$ and $\epsilon_{\text{tol}} = 5.0 \cdot 10^{-10}$. No mass smearing version of the Newton algorithm is presented here, it would require the computation of an additional set of second order derivatives with respect to the Higgs mass to obtain the analytical Hessian. The purpose of the $hWW\text{Newton}$ algorithm is to demonstrate that better run time performance is easily achievable with alternative gradient-based methods. Every method uses the analytical solution, presented in section 4.2.2, to find a starting point on the hyper-surface of real solutions (except for the grid search based algorithms that have no use for it). Exactly like for $t\bar{t}$, it is important to show that the algorithms outputs are consistent with each other. A weighted dimension-less distance s_w is defined, analogously to equation 4.25 for $t\bar{t}$, as follows:

$$s_w = \sqrt{\left(\frac{\delta\eta_\nu}{6}\right)^2 + \left(\frac{\delta m_{W_\ell}}{50 \text{ GeV}}\right)^2}. \quad (4.26)$$

Similarly to the parametric distance used for $t\bar{t}$, the factors are chosen such that the longest possible distance for grid search algorithms (the diagonal of the grid) would be given more intuitively by $s_w = \sqrt{2}$. Figure 4.6 and 4.7 display the violin plots for all the possible pairing of studied algorithms, each corresponding to a non-diagonal entry of the distance matrices, for ggF and VBF respectively. The distributions are almost undistinguishable between both Higgs production process. Just as in the $t\bar{t}$ case, the means of the distributions are relatively high for most algorithm comparisons due to long tails, which explains the inflated values seen in the distance matrices. Distribution 0 shows the difference between the grid search algorithm with and without mass smearing. Just as in the $t\bar{t}$ case, an artificial multimodality can be observed at small values of s_w , but with less undulations in the shape, approaching a cleaner bimodal distribution. The more complex shape in $t\bar{t}$ is explained by the fact that there are two mass dimensions instead of one here. Distributions 1 to 6 have the longest tails and widest spread, which is expected as they illustrate the comparison of algorithms with very different optimisation methods: grid search-based against gradient-based. This effect was also observed in

the previous section where it was established that the main source of difference is the optimisation method choice and not whether it incorporates mass smearing. Distribution 1 and 5 show that if the algorithm pair agree on the usage of mass smearing then the distance s_w is smaller. This is a small effect, which corroborates what has been found with $t\bar{t}$, but is reassuring as it is expected. It is clear with distributions 3 and 6 that hWW Newton seems to be the most different optimisation method to the simple grid search, agreeing with the distance matrices, even if it is by a marginal amount overall. Finally, distributions 7, 8 and 9 show that the gradient-based methods are quite similar with each other. Distribution 7 is interesting, the peak of the distribution is clearly offset from 0 even though the only source of difference is the mass smearing. In the $t\bar{t}$ case the peak was much closer to 0. Distribution 8 shows very good agreement between the MINUIT implementation and the custom Newton optimiser, peaking at 0 and having a short tail. Distribution 9 shows the impact of adding the mass dimension: the spread is increased and the peak is clearly offset from 0.

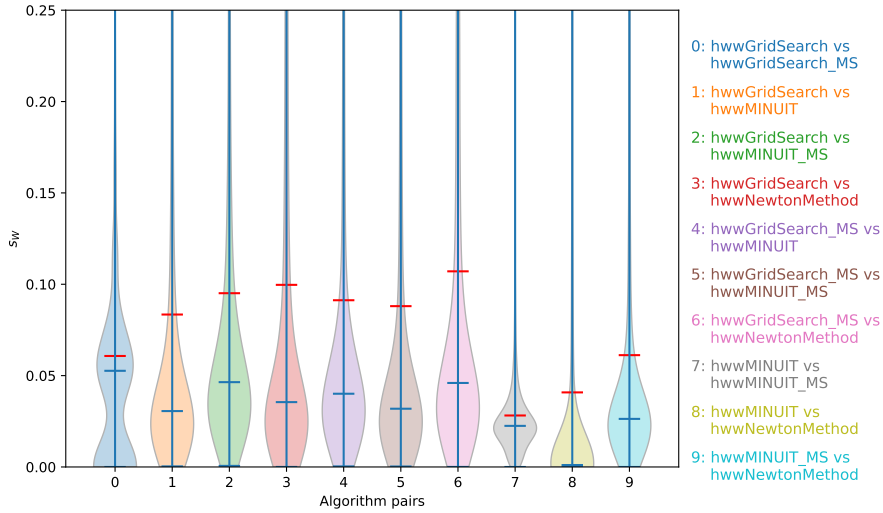


Figure 4.6: Violin plots displaying the distribution for the distance s_w between the solution of different algorithm pairs running on ggF events. The red line shows the mean of the distribution while the blue line highlights its median.

Having assessed the differences between algorithms in more detail, and confirmed that they agree well overall, their individual performances can now be considered. Table 4.2 shows the different metrics for each algorithms, on ggF and VBF events separately. The mean duration (run time) for the algorithms shows the expected better performance of gradient-based methods over grid searches. Adding a mass dimension to the grid search algorithm is quite expensive, multiplying by five the run time required per event

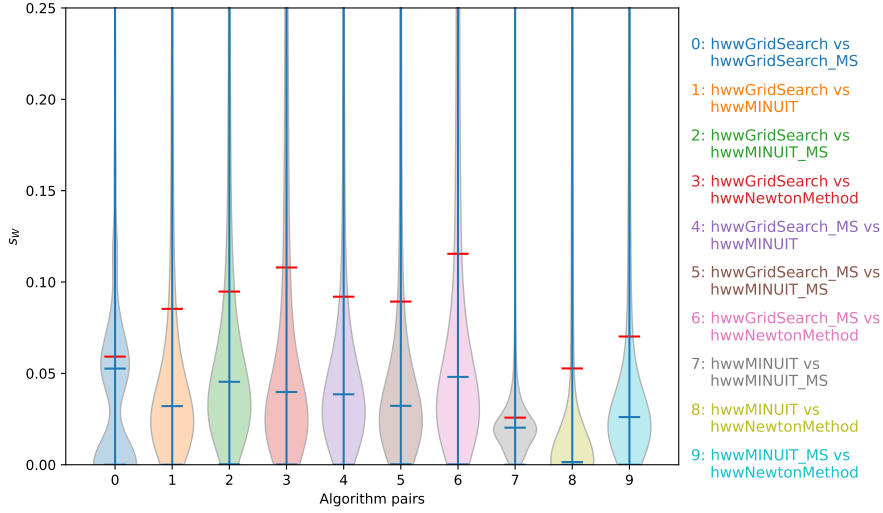


Figure 4.7: Violin plots displaying the distribution for the distance s_W between the solution of different algorithm pair running on VBF events. The red line shows the mean of the distribution while the blue line highlights its median

hWW -	GridSearch	GridSearch _{MS}	MINUIT	MINUIT _{MS}	Newton
ggF					
mean duration (μs)	333	1607	233	394	55
mean best weight	0.776	0.789	0.892	0.895	0.855
reco fraction	0.575	0.582	0.575	0.582	0.575
VBF					
mean duration (μs)	193	941	139	240	29
mean best weight	0.774	0.787	0.904	0.906	0.843
reco fraction	0.445	0.45	0.445	0.45	0.445

Table 4.2: Metrics computed using roughly 200 thousands events in total, for both ggF and VBF. The mean duration and best weight are computed by taking into account successfully reconstructed events only.

on average, for both ggF and VBF. For the algorithms using MINUIT, adding the mass parameter less than doubles the average run time, displaying the excellent scaling of the gradient-based methods. $hWWMINUIT_{MS}$ can be faster than the grid search algorithm on average without mass smearing, which is an impressive result. The Newton algorithm is faster than the others by an order of magnitude, which shows the power of the analytical computations used in the creation of this custom method. Figure 4.8 shows the run time distribution for the ggF events successfully reconstructed by the

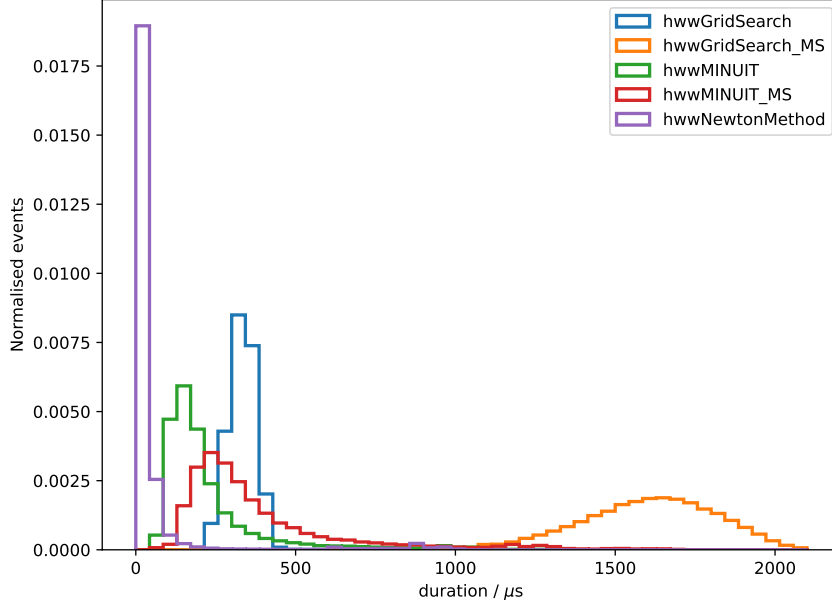


Figure 4.8: Distribution of algorithm run times for ggF events with a NW solution.

algorithms. The grid search algorithm run times are distributed as Gaussians around fixed values, which is exactly what is expected from algorithms with a fixed number of operations. The Gaussian aspect comes from stochastic fluctuation from the system, a noise arising from the CPU being subject to other instructions during the runs. The algorithms making use of MINUIT instead behave like one-sided crystal ball functions that peak before the hWW GridSearch algorithm, highlighting their overall superior run time performance. Adding mass smearing to the gradient-based algorithms as a parameter seems very efficient since it changes the distribution very slightly compared to the grid search algorithms. This nicely illustrates again the very good scaling of gradient-based algorithms. hWW Newton clearly outperforms all methods in run time by peaking very close to zero. The mean best weight is interesting because it is clearly lower for the grid search algorithms than for the gradient-based algorithms, for both ggF and VBF. This behaviour, not trivially expected since the algorithms should be consistent, as demonstrated above with the distance s_w , can easily be understood by looking at the distribution of the leptonic W mass outputted by the algorithms. This distribution is shown for ggF in Figure 4.9, where it can clearly be seen that the two grid search algorithms behave differently: the gradient-based algorithms find maximum values with $m_{W_\ell} > 50$ GeV, which was the upper limit we defined for the grid used by

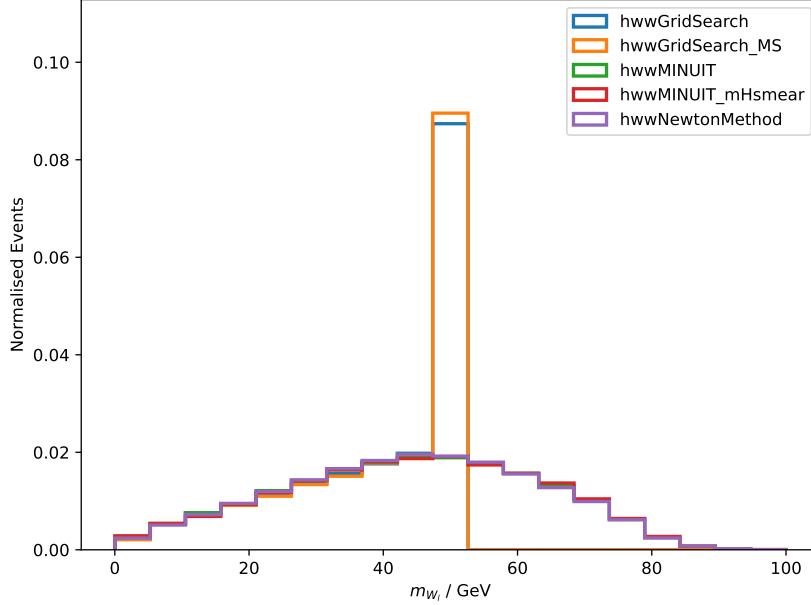


Figure 4.9: Distribution of the output leptonic W mass for ggF events.

grid searches. This illustrates a key practical advantage of the gradient-based algorithms: no grid definition is required, which makes them more natural and robust. This is particularly useful in multiple dimensions when including mass smearing, one does not have to manually study what limits to use by pre-running the algorithms at least once per dataset. Setting the upper limit of the grids to 90 GeV and re-running the algorithms yield the results presented in table 4.3. It is clear that extending the grid range makes the

hWW -	GridSearch	GridSearch _{MS}	MINUIT	MINUIT _{MS}	Newton
ggF	0.882	0.895	0.892	0.895	0.855
VBF	0.895	0.908	0.904	0.906	0.843

Table 4.3: The best NW weight averaged over successfully reconstructed events for the ggF and VBF samples after extending the grid upper limit for the m_{W_ℓ} parameter from 50 GeV to 90 GeV.

average of the best NW weight more consistent across every algorithms. The algorithms with mass smearing will find higher weights than their non-smearred counterparts since they optimise along an additional axis. Some fluctuation is expected for gradient-based algorithms because of how much the shape of the NW function can change between events, it may include multiple peaks to choose from or saddle points that are difficult to

escape. Overall, after updating the grid to more reasonable values for m_{W_ℓ} , the algorithms find very similar best weights on average, as expected from our earlier consistency study.

The reconstruction rate is different between ggF and VBF, even though the algorithms seem to behave very similarly for both production processes. The VBF events reconstruction rate is, interestingly, roughly 13% lower, with and without mass smearing for all algorithms. This interesting difference between the production mode could be explained by the fact that the JetPairingNet neural network used to find the jet pair corresponding to the hadronic W has only been trained on ggF samples. As expected, the inclusion of mass smearing increases the reconstruction rate, by 0.7% for ggF and 0.2% for VBF. This increase is relatively small and can be understood by studying the distribution of the best Higgs mass found by the algorithms, shown in Figure 4.10. For algorithms with mass smearing, the events clearly accumulate on the edges of the defined range, similarly to when the grid size is too small in m_{W_ℓ} . The constraint on the Higgs mass, as a parameter or as a grid dimension, is then clearly too tight for optimal performance and needs to be loosened to better accommodate for the detector smearing effects, which in turn will improve the reconstruction rate. Table 4.4 shows the mean best weights and the fraction of reconstructed events when increasing the Higgs mass smearing to 10 GeV. The increase in reconstruction rate gained by including mass smearing is now around 6% for ggF and 4% for VBF. The increase in average best weight is also significantly greater. If the NW algorithm is used for separating signal and background in an analysis, such as the one presented in chapter 6, then it might be important to study the effect of mass smearing on background processes as well. Tables 4.2 and 4.4 show that the fraction of reconstructed events is exactly the same for algorithms that use the same mass smearing policy but different optimising methods, indicating that the analytical solution, used by gradient-based methods only and described in section 4.2.2, works robustly to establish whether an event is reconstructable.

4.4.3 Conclusion

The NW algorithm was presented in its original form, used to reconstruct $t\bar{t}$ events, and in an adapted version that is used to reconstruct the semi-leptonic decay of Higgs to W bosons. The NW method requires an optimisation over two or more parameters, which is typically performed using a grid search due to the non-trivial domain of real solutions produced by the NW function. For both the $t\bar{t}$ and $H \rightarrow WW^*$ cases, two

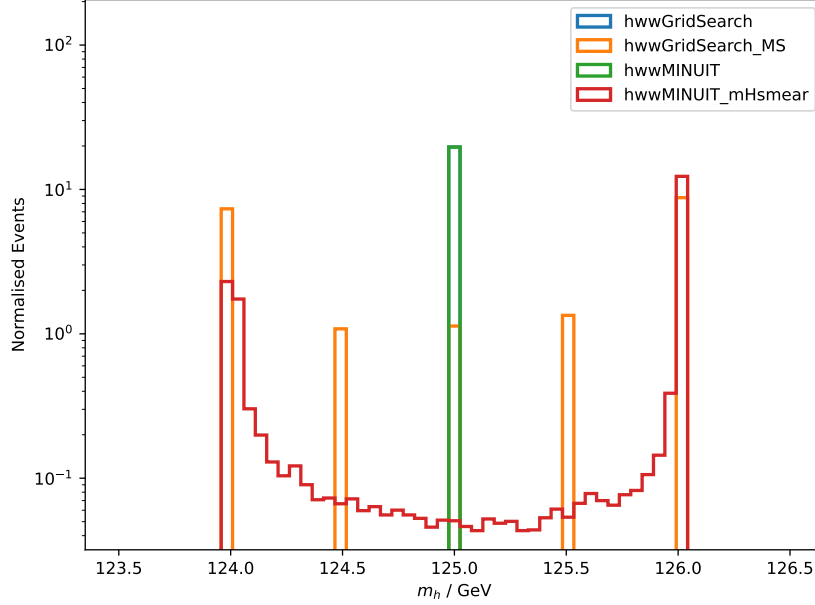


Figure 4.10: Distribution of the best Higgs mass found for ggF events. Algorithms without mass smearing have every event with $m_h = 125.0$ GeV, which is why the blue histogram is not visible, hidden behind the green distribution.

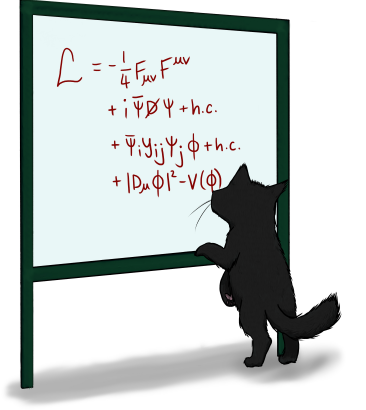
hWW -	GridSearch	GridSearch _{MS}	MINUIT	MINUIT _{MS}	Newton
ggF					
mean best weight	0.776	0.853	0.892	0.912	0.855
reco fraction	0.575	0.634	0.575	0.634	0.575
VBF					
mean best weight	0.774	0.85	0.904	0.92	0.843
reco fraction	0.445	0.485	0.445	0.485	0.445

Table 4.4: The best NW weight averaged over successfully reconstructed events, and the fraction of reconstructed events for the ggF and VBF samples after increasing the Higgs mass smearing from 1 GeV to 10 GeV.

techniques were presented to find an initial point within the domain of real solutions such that the optimisation may alternatively be carried by gradient-based algorithm: a first strategy using an inexpensive line search and a second strategy using an analytical solution derived from the NW kinematic equations. The latter approach was tested (for both cases) and found to be robust during the benchmarking results presented in section 4.4. The NW function was optimised using both a classic grid search algorithm and the

MIGRAD optimiser of the MINUIT software, with and without mass smearing. For the $H \rightarrow WW^*$ NW, an additional custom and rudimentary gradient-based optimiser was built following the Newton Method by deriving analytically the Hessian of the NW function. This was done to highlight the very low achievable run times, noting that MINUIT is chosen for practical reasons and is not ideal when running event-by-event. The custom optimiser demonstrates the higher performance that can be achieved by using better suited methods. In a first time, the consistency between algorithms is verified, then the run time, average best weights and reconstruction rates are assessed as part of the benchmarking. For $t\bar{t}$ events, the usage of gradient-based methods shines when including the mass smearing, which is important to improve reconstruction rate and might play a role in neutrino resolution (further study being required to assert the latter). One can note that a custom method is possible and would considerably improve run time performance, as shown with the Newton optimiser used in the Higgs case. For $H \rightarrow WW^*$ events, the gradient-based algorithms are strictly superior concerning run time regardless of the inclusion of mass smearing. The top quark and Higgs mass smearing distributions show that the best NW weight is often far from their nominal values (more than 10 GeV), with top quarks masses preferentially decreasing and Higgs masses mildly more likely to increase than decrease. The question of how tight the constraints on mass parameters should be is non-trivial and relates to the jet resolution provided by the detector, and thus should be studied and evaluated specifically for the analysis of interest.

The quality of the NW solution should be studied by comparing it to the MC truth-level information. The E_T^{miss} resolution is the main limiting factor on the achievable neutrino resolution, but mass smearing and the usage of alternative algorithms to NW might have a noticeable effect as well. Finally, a strategy is proposed here to squeeze more background discrimination power out of the NW algorithm. First collect the NW best weight, the entries of the Hessian matrix containing curvature information (estimated numerically by a software like MINUIT or computed analytically), and define a real distance δ_c from the origin to the global NW peak after extending the considered domain to complex solutions by keeping the real part of the computed neutrino transverse momentum only. Then combine all these quantities into a vector space, compute their covariance matrix and diagonalise it following the well-known Principle Component Analysis (PCA) [162] procedure. The diagonalisation yields new (linearly) uncorrelated quantities that are linear combinations of the old ones. These new variables can then be compared to the NW weight by computing their ROC curves for signal and background processes.



5 Proposal to Measure $H \rightarrow WW^*$ Entanglement in New Decay Channel

The ATLAS and CMS collaborations have both observed the decay of Higgs bosons to W bosons via both the ggF and VBF production modes [163, 164]. This decay has only been studied by using the dileptonic decay of the W bosons, the two charged leptons leaving a clean signature in the detectors, whereas the hadronic decays of the W produce a final state that resembles significant SM background processes. However, the fully leptonic final state has two main disadvantages: the relatively low leptonic branching ratio (roughly 30% using lepton universality arguments, as done in section 1.1.2) and the presence of two neutrinos. The former limits the event yield and results in large statistical uncertainties, while the latter makes the reconstruction of the Higgs boson kinematics incredibly challenging to perform due to the lack of constraint on the system. Full kinematic reconstruction would enable access to the Higgs and W bosons rest frames, which is crucial to measure spin information and observe interesting phenomena such as the maximum entanglement of the WW^* qutrit system induced by the scalar nature of the Higgs [58]. The semi-leptonic decay strikes an interesting balance, providing both a fully reconstructable final state and a signature that can be isolated from background, while benefitting from a larger branching ratio than the traditional channel. This chapter presents the phenomenological study published in [165], the proposal of a method to isolate the semi-leptonic decay of $H \rightarrow WW^*$ using the NW algorithm (see section 4) and c-tagging. Additionally, the NW neutrino reconstruction capabilities are used to investigate access to spin information: the unfolding of the observables composing the Bell Operator derived in [58] (and presented in section 1.2.3) is performed.

5.1 Background Suppression and Event Reconstruction

Targetting a final state with one charged lepton and a minimum of two jets yields a phase-space dominated by the following backgrounds involving at least one W boson leptonic decay, sorted by expected contribution: W +jets, semi-leptonic $t\bar{t}$, semi-leptonic diboson and single top ($t+W$). In the analysis presented in chapter 6, it was discovered that the Z +jets background has a very important contribution from events where one of the charged leptons from the Z boson decay is not reconstructed.

In $H \rightarrow WW^*$ decay, one of the W bosons has to be off-shell due to the mass of the Higgs boson being smaller than twice the on-shell W mass (~ 125 GeV vs ~ 160 GeV). One could make the assumption that the leptonically decaying W is on-shell as an additional constraint, and then use the missing transverse energy to reconstruct that W separately. The hadronic W would then have to be reconstructed by combining jets with some criterion or algorithm that could make use of their combined invariant mass, expected to be lower than 80 GeV. This approach is similar to a technique employed in the reconstruction of semi-leptonic $t\bar{t}$ decay [166] and would have a decent efficiency on the signal but also on the W +jets process, which is the main background. The semi-leptonic decay mode has not been considered yet mainly because of the W +jets process, which acts as an overwhelming background with a cross-section four orders of magnitude greater than the signal (4.3 pb [16, 23] against ~ 42000 pb [167]). It is therefore crucial to suppress the W +jets background as much as possible. Instead of relying on an on-shell leptonic W , the proposed strategy is to sacrifice the direct usage of $\vec{E}_T^{\text{miss}}$ and require the hadronic W to be on-shell (unlike in the W +jets process). The missing transverse energy is then instead used in the NW algorithm (described in more detail in chapter 4) to reconstruct the neutrino and thus the complete system. The weight output by the NW algorithm can then also be used to further suppress the background processes. Requiring the hadronic W to be on-shell has the added benefit of providing a simple criterion to pair jet objects for its reconstruction: the minimisation of $|m_{jj} - m_W|$, where $m_W = 80.6$ GeV. Making use of the high branching ratio for the semi-leptonic decay and the large amount of statistics provided by the LHC, one may restrict the targeted final state to the hadronic W decaying to a charm and strange quark in an effort to suppress backgrounds with different signatures. For instance, the W +jets background events are more likely to contain light quarks and gluons in the final state. The handling of selection on quark flavours is experimentally done with jet flavour tagging, a technique

described in detail in chapter 3. The pre-selection should thus filter out events that do not have exactly one c -jet and at least one light-jet. Additionally, the requirement for zero b -jet can be enforced to suppress backgrounds further, in particular, the very prominent $t\bar{t}$ process that contains two b -quarks in its final state.

5.1.1 MC and Simulation of Detector Effects

The hard scatter of signal and background events, with only one lepton in the final state, are simulated using the POWHEG [144] MC generator and the parton showers with PYTHIA [114]. The diboson processes are generated with SHERPA [168] instead of PYTHIA. Only the dominant ggF production process is simulated for the signal, the VBF process is not studied in this work due to its relatively small cross-section in comparison. The RIVET software (described in a bit more detail in section 3.5), is used to read the HEPMC event files and reconstruct particle-level physics objects. Electrons and muons are associated with a W boson decay (including charged leptons originating from indirect decay of τ particles). These leptons are *dressed* with photon radiation found within a cone $\Delta R < 0.1$: the photons are removed from the event record and summed into the lepton four-momentum to take into account the QED final-state radiation. Jets are reconstructed using the anti- k_t algorithm with the classic distance parameter $R = 0.4$, and are flavour tagged with the potential possession of b - and c -hadrons through ghost matching. $\vec{E}_T^{\text{miss}}$ is defined simply using neutrinos originating from W boson decays (again including the neutrinos from τ decays). This results in neglecting the experimental effect of the missing transverse energy resolution, which is discussed in more detail in section 5.3. The effect of the detector response is simulated by applying publicly available acceptances and efficiencies from the ATLAS collaboration to the physics objects. The electrons [169] and muons [170] trigger efficiencies are approximated with a p_T threshold of 20 GeV. The lepton identification and reconstruction efficiencies are taken to be 90% and 96.1%, following the ATLAS definition for medium lepton quality [171, 172]. No efficiencies are applied to jets and effects arising due to limited resolution are ignored. The jet energy smearing was simulated using the fractional Jet Energy Scale (JES) uncertainties shown in [173]. The fractional uncertainties were used to smear the energy of the jets following a Gaussian distribution. Specifically, the jet energy E was transformed as $E \rightarrow E \times \mathcal{N}(1, \sigma_j(p_T))$, where σ_j is the JES uncertainty that depends on the p_T of the jet. The jet energy smearing was found to have a negligible effect on the results presented in sections 5.2.1 and 5.2.2.

Flavour tagging is simulated in two steps: first, the true flavour of the particle-level jet is determined using ghost-matching, then jet tagging efficiencies are applied to emulate experimental conditions. Ghost-matching is a technique for determining whether a jet contained an unstable B - or D -hadron by setting their energy to zero and letting their momentum be clustered into the stable particle jets with the anti- k_t algorithm. Any jet that clustered a B -hadron is a true b -jet, any jet that clustered a D -hadron and no B -hadron is a c -jet. The hadrons considered in the ghost-matching are required to have a p_T above 5 GeV. Table 5.1 shows the ATLAS-like [136, 145] efficiencies and mis-tagging rates applied to the jets given its true flavour. The values reflect that the true light-jets are rejected efficiently with flavour tagging in real experiments like ATLAS, but that b -jets and c -jets are more difficult to discriminate from one-another. The values in that table are taken from older performance papers, which are less performant than the results shown for the latest state-of-the-art flavour tagger in chapter 3. This is because the GN2 tagger did not yet exist at the time of the publication of this work.

Table 5.1: c - and b -tagging efficiencies of the DL1R tagger used in the ATLAS experiment [136, 145]. The numbers refer to the probability for the tagging algorithms to assign the c or b flavour to jets containing a B -hadron (b -jet), a c -hadron and no B -hadrons (c -jet) or no heavy-flavour hadrons (l -jet). These numbers are directly used in the analysis presented in this paper to emulate realistic c -tagging performances.

true flavour	c -tagging efficiency	b -tagging efficiency
b -jet	0.14	0.77
c -jet	0.4	0.2
l -jet	0.016	0.008

5.1.2 Event Selection

The considerations highlighted in section 5.1 are now compiled to produce the following event selections:

- **Exactly 1 lepton with $p_T > 20$ GeV**
- **Exactly 0 b -tagged jets** to suppress $t\bar{t}$ events.
- **2 or more jets, exactly one of which must be c -tagged**, since we expect only one charm quark from W hadronic decays.

- **At least 1 (c -jet, l -jet) pair with $|m_{cl} - 80.6| < 10$ GeV** to suppress events with off-shell hadronic W bosons.
- **A reconstructed leptonic W boson from NW with $w > 0.7$** to suppress all backgrounds and have events with good reconstruction quality.
- **A maximum of 2 light jets** to suppress backgrounds like $t\bar{t}$ and $t+W$ where jet misidentification may supply additional light-jets.
- **The invariant mass of the lepton and the c -tagged jet must satisfy $m(\ell c) < 80$ GeV¹** to suppress $t\bar{t}$ events where the lepton and misidentified c -jet come from the same top quark.

For events with multiple combinations of light-jets and c -jets, the pair whose combined system has an invariant mass closest to $m_W = 80.6$ GeV is chosen as the reconstructed hadronic W . Two charm tagging setups are used throughout this work: an idealised case with perfect identification efficiencies and zero mis-tag rates, as well as a realistic case with the efficiencies and mis-tag rates shown in Table 5.1 (with true c -jet efficiency $\epsilon_c = 40\%$). The former setup is provided as an indication of how significant potential flavour tagging improvements could impact the result, whereas the latter is given as a more realistic scenario. Every result presented uses the realistic case by default, unless specified otherwise. Since the hadronic W is an important input of the NW algorithm, it is expected that the c -tagging has some impact on the NW reconstruction performance. The NW algorithm is found to have a reconstruction rate of 78% on the signal events in the ideal tagging against 72% for the realistic tagging case.

The expected purity, computed as the fraction of the selected signal events (S) compared to the total yield including the background (B) events (i.e $\frac{S}{S+B}$), is shown as a function of the weight output by the NW algorithm in Figure 5.1.

Applying the one lepton and b -veto requirements alone, the purity is of the order of 10^{-4} . The c -tagging selection helps in isolating the signal further by increasing this value by a factor of ten, but it is clear that most of the isolation of the signal must be performed by cutting on the NW weight. Requiring the NW weight to be greater than 0.7 has a 45% efficiency on the signal against a 0.5% efficiency on the backgrounds. Two additional criteria have been added to suppress the $t\bar{t}$ process to a greater extent. Amongst the $t\bar{t}$ events that pass the selection, the expected two b -jets must be mis-tagged as either light-

¹ $m(\ell c)$ is the invariant mass of the lepton + c -jet system, not to be confused with the invariant mass of the c - and light-jet system m_{cl} .

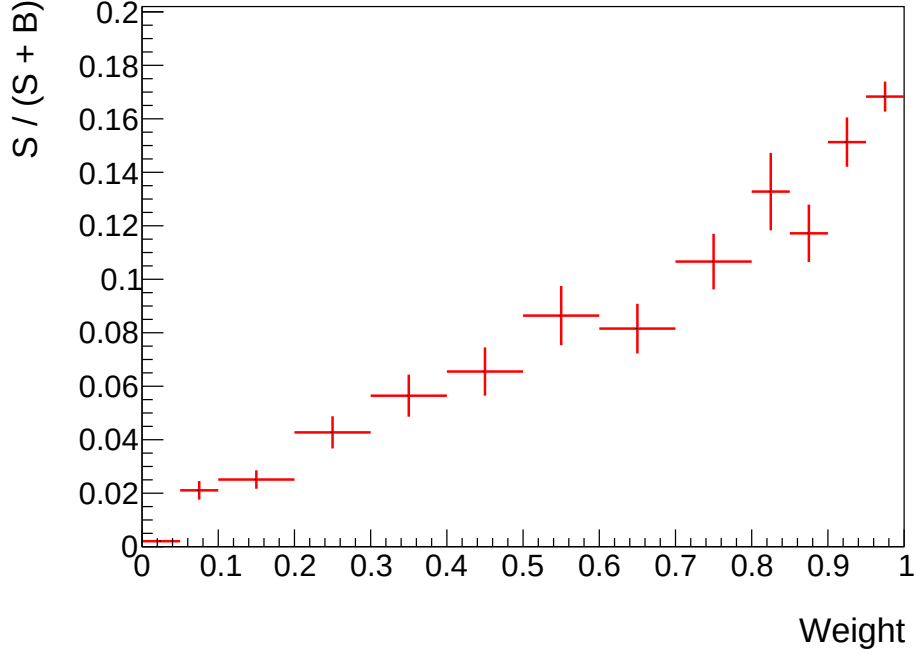
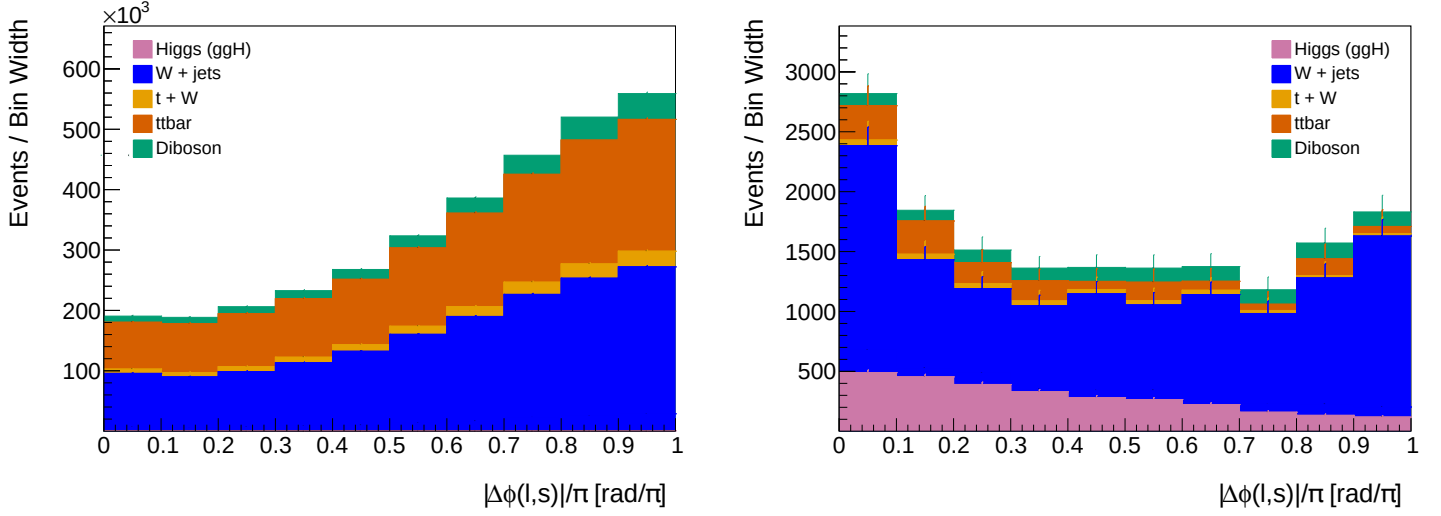


Figure 5.1: Signal purity as a function of the requirement applied on the NW weight to be greater than some value. Only events passing the pre-selection and the c -tagging selection are considered.

or c -jets. Limiting the number of light-jets in the event aims at discarding events in the first category, whereas the cut on $m(\ell c)$ targets events in the second category. The latter criterion can be understood better by realising that in the case a b -jet is mis-tagged as a c -jet, both the lepton and the c -jet will originate from the same top quark, and thus the invariant mass $m(\ell c)$ has a distribution unique to $t\bar{t}$ events. For example, the distribution is bounded by the top mass. The expected composition of the data after application of only the pre-selection and c -tagging requirements, and after the entire selection is shown in Figure 5.2. Despite the combination of the requirements on the reconstructed hadronic W mass and the veto on b -jets, the W +jets and $t\bar{t}$ processes dominate significantly over the signal, which is completely invisible in the stacked histogram of panel 5.2 (a). After applying the NW weight cut, the background processes are heavily reduced, and the signal becomes well visible, as shown in 5.2 (b).

The expected yield of the signal and background processes for 300 fb^{-1} of data at $\sqrt{s} = 13 \text{ TeV}$, the total luminosity expected by combining Run2 and Run3, is presented in Table 5.2. Run3 is operated at the slightly higher energy of 13.6 TeV, which has a negligible effect on the cross-sections of the physical processes involved in this study.


 (a) With pre-selection and c -tagging requirements.

(b) With application of the entire selection.

Figure 5.2: Distribution of $|\Delta\phi_{ls}|$, the absolute value of the difference in azimuthal angle between the light-jet identified as originating from the s -quark and the lepton. The angle difference is divided by π to bound the distribution between 0 and 1.

With realistic charm-tagging conditions, roughly 2800 signal events are expected with a purity of 18%. This signal purity is remarkable considering that the cross-section of the background processes are a few orders of magnitude higher than for the signal. Assuming perfect charm-tagging, these numbers improve to around 6000 signal events and a purity of 27%. It is clear that such a performance is unrealistic, but it can be taken as an upper bound on the results provided future improvements to the experimental techniques employed in this work (for example on flavour tagging, lepton efficiencies or jet resolution). Including the effect of jet smearing degrades the purity to 13% for the realistic charm-tagging scenario, which is more realistic. Advanced analysis methods, such as multivariate techniques that make use of machine learning, can be used to improve signal isolation from the background by including additional observables that behave differently between the signal and the backgrounds of importance.

5.2 Observing Entanglement in Higgs Events

Section 1.2.2 and 1.2.3 describe the concept of Bell Inequality and how it can be used to produce an observable that measures quantum entanglement. In particular, the strategy proposed by Barr [58] is followed, with small modifications to adapt the method

Process	idealised			$\epsilon_c = 40\%$		
$W + \text{jets}$	13131	$\pm$	785	10444	$\pm$	664
WW	2298	$\pm$	31	1137	$\pm$	22
$t\bar{t}$	601	$\pm$	76	1453	$\pm$	119
tW	217	$\pm$	8	350	$\pm$	11
Higgs	5967	$\pm$	76	2843	$\pm$	56
$S/(S+B)$	0.27			0.18		

Table 5.2: The expected yields for an integrated luminosity of 300 fb^{-1} in a selection with idealised charm tagging and Run2-like charm tagging. The purity ($S/(S+B)$) is also shown. The uncertainties are statistical only.

from the dileptonic channel. As briefly discussed in section 1.1.2, the down-type quark preserves the spin information of their parent W boson upon its decay very similarly to the lepton (both have maximum spin analysing power at tree level). Consequently, the technique described in [58] can also be applied to the signal studied in this work if we can identify the light-jet associated with the down-type quark and use it as a proxy for its four momentum. The expectation values for the Bell operators $\langle \mathcal{B}_{\text{CGLMP}}^{xy} \rangle$, $\langle \mathcal{B}_{\text{CGLMP}}^{yz} \rangle$, and $\langle \mathcal{B}_{\text{CGLMP}}^{zx} \rangle$, are composed of the average of three observables, explicitly written in equation 1.56. One can define the following expressions:

$$\begin{aligned}
 \mathcal{O}'^1_{xy} &= \frac{8}{\sqrt{3}} \langle O^1_{xy} \rangle \\
 O^1_{xy} &= \xi_x^+ \xi_x^- + \xi_y^+ \xi_y^- \\
 \mathcal{O}'^2_{xy} &= 25 \langle O^2_{xy} \rangle \\
 O^2_{xy} &= ((\xi_x^+)^2 - (\xi_y^+)^2)((\xi_x^-)^2 - (\xi_y^-)^2) \\
 \mathcal{O}'^3_{xy} &= 100 \langle O^3_{xy} \rangle \\
 O^3_{xy} &= \xi_x^+ \xi_y^+ \xi_x^- \xi_y^-,
 \end{aligned}$$

where the ξ_i have the same meaning as in equation 1.56, and the observables for the yz and zx axes can be straightforwardly obtained by swapping the relevant ξ_i . The sign of each W is determined by using the sign of the electric charge of the lepton. Each observable O^k_{ij} is constructed event-by-event and used to fill histograms to study their distributions separately. The means are computed for each distribution and used to construct the

corresponding operators $\mathcal{O}_{ij}^k$, which are necessary to compute the entanglement witness $\mathcal{I}_3$.

5.2.1 Unfolding

Unfolding is a technique used in collider measurements to correct for detector effects and, in case of parton-level measurements (as opposed to particle-level measurements), non-perturbative effects such as hadronisation, in an effort to recover the truth-level distributions. One can see unfolding as a method that aims to invert the mapping from parton-level to reco-level:

$$\begin{array}{ccc} \text{parton-level} & \xrightarrow{\text{hadronisation} + \text{detector effects}} & \text{reco-level} \\ \text{reco-level} & \xrightarrow{\text{unfolding}} & \text{parton-level.} \end{array}$$

The detector response is encoded in a migration (or response) matrix M_{ij} , which describes the probability for an event generated in a given truth bin i to be reconstructed in a given reco-level bin j . This setup can be described mathematically as follows:

$$r_j = \sum_i M_{ij} t_i, \quad (5.1)$$

where r_j and t_i are components of vectors representing the content of bins for the reco-level and truth-level distributions, respectively. The distributions are constrained to the *fiducial region* at truth-level, the central region where the detector is considered reliable. The unfolding is the effective inversion of the migration matrix, which is typically done with more sophisticated method than direct linear algebra matrix inversion. This is because of the nature of the migration matrix: events from one truth bin can migrate over many reco-level bins, which causes the columns to be strongly correlated and make the matrix close to singular. Additionally, the finite size of the simulated samples used to estimate the migration matrix leads to the presence of statistical fluctuations whose effects are amplified by the matrix inversion. The alternative methods make use of *regularisation* techniques that generally stabilise the inversion by suppressing solutions that are too noisy and promoting smoothness. Iterative Bayesian Unfolding [174] is used here to remove the smearing effects imposed by the simulated detector efficiencies and jet mis-tagging. This method is based on the Bayes theorem and re-expresses the problem

as follows:

$$t_i^{(n+1)} = \sum_j r_j P^{(n)}(i | j), \quad (5.2)$$

with

$$P^n(i | j) = \frac{M_{ij} \pi_i^n}{\sum_k M_{kj} \pi_k^n}, \quad (5.3)$$

where π_i^n describes the Bayesian *prior* on the truth distribution in iteration n . At every new iteration $n + 1$, the prior is replaced with the *posterior* probability from the previous iteration ($\pi_i^n := t_i^n$). In our case, the prior is computed from MC simulation assuming the SM. The regularisation comes from the prior that imposes a smooth physical shape from MC simulation: a low amount of iteration would result in a strong regularisation because the result remains close to the prior. Increasing the number of iterations reduces the bias but results in higher variance for the resulting distribution.

In the study, a procedure described in several ATLAS top quark measurements [154, 166] is followed and implemented with the ROOUNFOLD package [175]. The nine observables are unfolded using two Bayesian iterations and extrapolated to the full semi-leptonic phase space, at which point the expectation value of the Bell operators $\langle \mathcal{B}_{\text{CGLMP}}^{xy} \rangle$, $\langle \mathcal{B}_{\text{CGLMP}}^{yz} \rangle$, and $\langle \mathcal{B}_{\text{CGLMP}}^{zx} \rangle$, are computed. It is found that with the choice of axes and selection presented in section 5.1.2, $\langle \mathcal{B}_{\text{CGLMP}}^{zx} \rangle$ always has the largest value and thus corresponds to the overall value of $\mathcal{I}_3$. A side note on the choice of axes: it is possible to combine two $SU(3)$ matrices (both generated with 8 parameters, adding up to 16 in total) to rotate the Bell Operator in an effort to maximise the observed $\mathcal{I}_3$. Figure 5.3 shows the normalised migration matrices for the observable used to build this operator. The binning is defined based on a simple optimisation, which followed the constraint that 70% of events must lie on the diagonal bins, to obtain roughly diagonal matrices and thus a stable unfolding. The simulated detector resolution effects can be seen in the smearing of the distribution towards zero, the central value.

Figure 5.4 shows the unfolded distributions of all three observables used to compute the averages composing $\langle \mathcal{B}_{\text{CGLMP}}^{zx} \rangle$. Although the shape of the observable O_{zx}^1 indicates that it will have the furthest mean from zero, its contribution to the final value of $\langle \mathcal{B}_{\text{CGLMP}}^{zx} \rangle$ (and thus $\mathcal{I}_3$) will be limited by the relatively small coefficient $\frac{8}{\sqrt{3}}$ in the definition of $\mathcal{O}_{zx}^1$. Conversely, the distribution of the observable O_{zx}^3 shows a mean that will be relatively

small, but its overall contribution will be considerably enhanced by the coefficient of 100 in the definition of $\mathcal{O}_{zx}^3$.

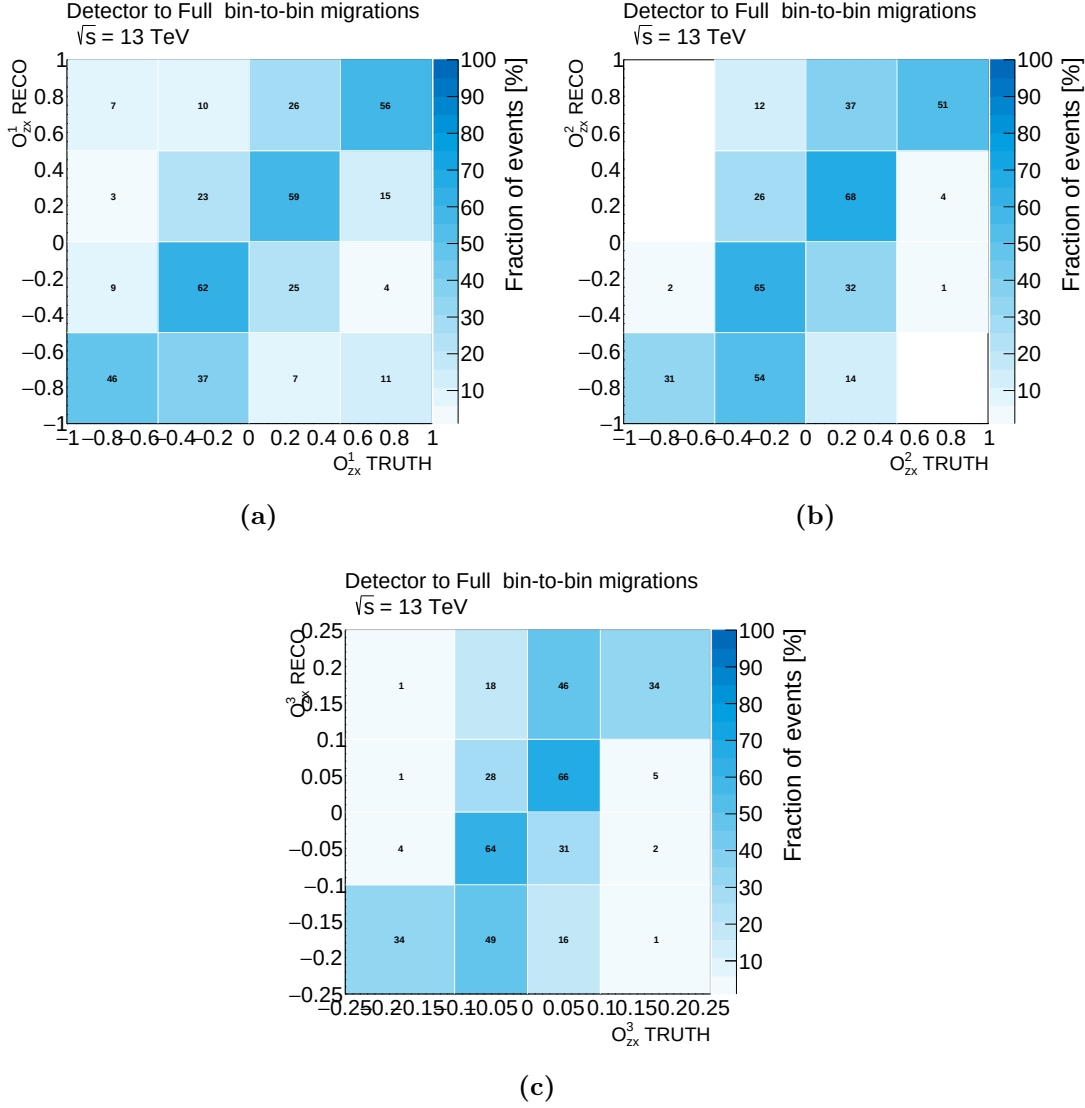


Figure 5.3: Migration matrices for the observables O_{zx}^1 (a), O_{zx}^2 (b) and O_{zx}^3 (c).

5.2.2 Expected Precision

The expected measurement and sensitivity for $\mathcal{I}_3$ is computed, using the event selection, reconstruction simulation and unfolded distributions described throughout this chapter. The results are shown for three integrated luminosities: 139 fb^{-1} (data collected in Run2), 300 fb^{-1} (expected data accumulated by Run2 and Run3 combined) and 3000 fb^{-1}

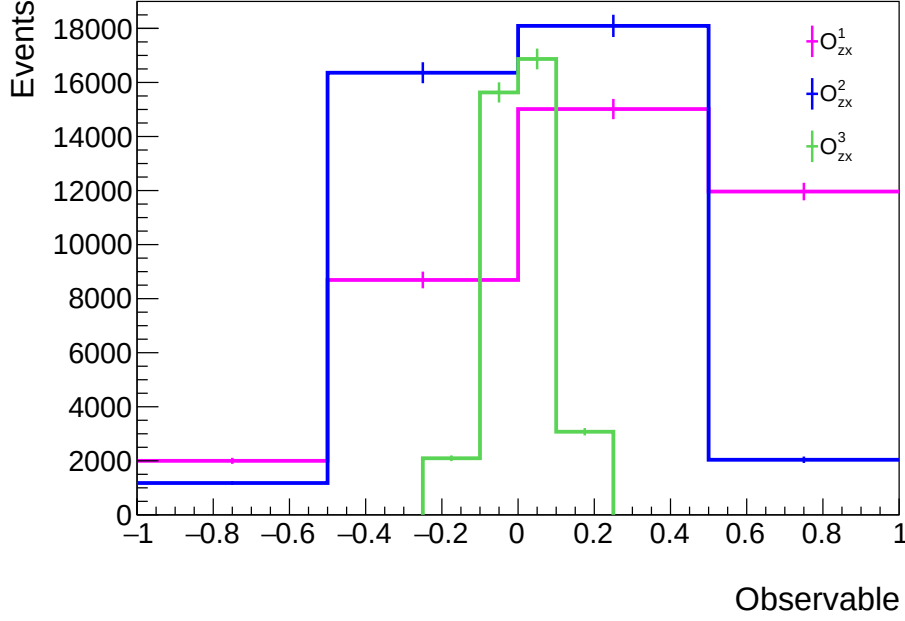


Figure 5.4: Unfolded histograms for the three observables composing the expectation value of the zx Bell Operator. The uncertainties on each bin are statistical.

(expected integrated luminosity for the High-Luminosity LHC program). For all cases, the centre of mass energy is taken to be 13 TeV even though Run3 is operated at a slightly higher energy (13.6 TeV) and High-Luminosity LHC (HL-LHC) may reach up to 14 TeV. The increase in energy has a small effect on the results presented here since the cross-sections of the processes involved barely change (for example, this can be seen for the signal production in Figure 1.3). Additionally, it is possible to combine multiple results obtained at different energies to improve the significance. The uncertainty considered is only statistical, and its evaluation takes into account the unfolding procedure. The uncertainty is estimated by using pseudo-experiments: the input histograms have their bins fluctuated with respect to a Poisson distribution and the standard deviation of the unfolded pseudo-experiments in each bin is taken as its uncertainty. Finally, this uncertainty is propagated to the mean of the histograms and $\mathcal{I}_3$. The expected values and significances of the measurement for each luminosity scenario are presented in Table 5.3. The central value is 2.45, a number greater than two, indicating the presence of quantum entanglement, i.e strong spin correlations that cannot be explained classically. The expected significance of these values, simply taken as $\frac{|2-2.45|}{\sigma}$ where σ is the statistical uncertainty, is shown for the realistic and idealised c-tagging cases. At 300 fb^{-1} the expected significance is below 3σ for the realistic case and close to 4σ in the idealised

Luminosity [fb^{-1}]	$\langle \mathcal{B}_{\text{CGLMP}}^{zx} \rangle$	(idealised)	Significance [σ](idealised)
139	2.45	± 0.25 (0.18)	1.8 (2.5)
300	2.45	± 0.17 (0.12)	2.65 (3.75)
3000	2.45	± 0.05 (0.04)	9.0 (11.25)

Table 5.3: The expected sensitivity for different values of integrated luminosity for a realistic and idealised case. The uncertainties are statistical only.

case, suggesting that evidence is attainable by the LHC towards the end of Run3. When jet energy smearing is simulated, the significances are diluted to 1.6, 2.4 and 7.5σ for 139 fb^{-1} , 300 fb^{-1} , and 3000 fb^{-1} respectively.

5.3 Simulating E_T^{miss}

The following section is not part of the publication [165], it presents a short study performed after gaining experience with ATLAS experimental simulation tools. As seen in section 5.1, in particular in subsection 5.1.2, the suppression of the background relies dominantly on the discriminating power of the NW algorithm weight. As briefly mentioned previously and in the publication [165], the missing transverse energy is taken to be directly the transverse momentum of the parton-level neutrino daughter of the leptonic W . This assumption is, of course, unrealistic since $\vec{E}_T^{\text{miss}}$ is affected by other undetected emissions occurring in the parton showers, and suffers from a poor resolution in ATLAS and CMS. The NW algorithm is very sensitive to $\vec{E}_T^{\text{miss}}$ and will therefore perform considerably worse with ATLAS data and full simulation. This is the main reason why the reconstruction rates are observed to be lower in chapter 4, and why the performance and strategy in measuring the $H \rightarrow WW^* \rightarrow \ell\nu jj$ process differs somewhat from expectations in the analysis presented in chapter 6. The impact of the NW algorithm can be studied phenomenologically, but it is crucial to simulate $\vec{E}_T^{\text{miss}}$ realistically. Figure 5.5 shows the resolution on the neutrino momentum provided by $\vec{E}_T^{\text{miss}}$ along the x-axis. A fit is performed to highlight the expected Gaussian shape of the resolution, and to extract its main properties: a mean of zero and a width of roughly $\sim 20 \text{ GeV}$. With this information, one can propose a simple, but effective, method to emulate the transverse missing energy resolution effects: sample a Gaussian distribution, centred at 0 with a width of 20 GeV, twice and add the obtained numbers to the neutrino momentum components along the x- and y-axis. Figure 5.6 shows the same physical

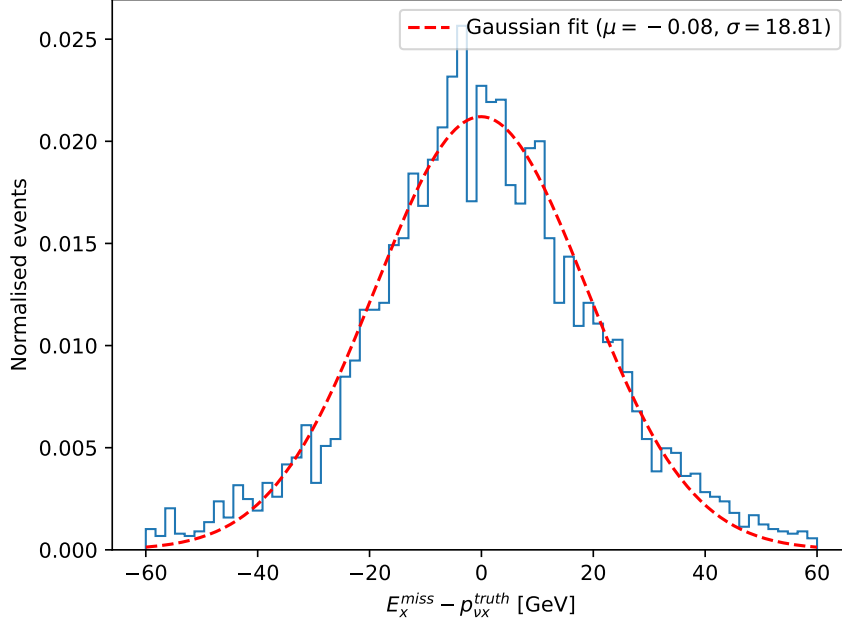
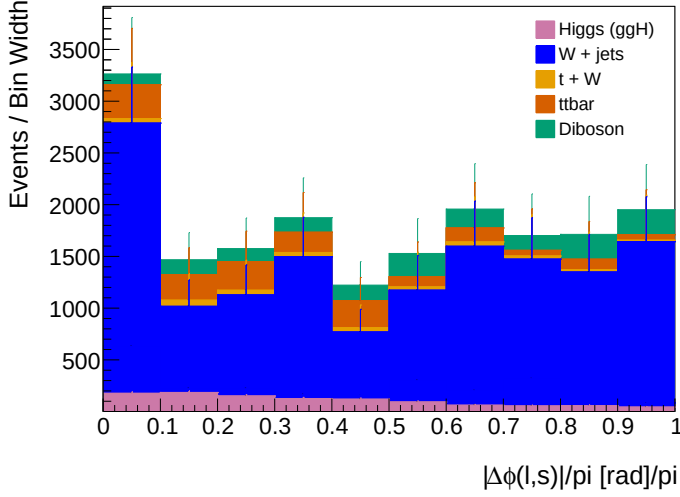


Figure 5.5: The difference between the ATLAS reco-level $\vec{E}_T^{\text{miss}}$ and the parton-level neutrino momentum along the x-axis for signal events.

distribution shown in section 5.1.2 with the addition of the simulation of the missing transverse energy resolution with Gaussian sampling. The signal is still visible, but much harder to see than in the previous result shown. The yield table for the realistic c-tagging scenario is also shown in the Figure. Compared to Table 5.2, one can see that every background has increased slightly in amount and that the signal decreased significantly by a factor of about 2.6, from roughly 2600 events to approximately 1000. The purity of the new yield is reduced by a factor of three, which is considerably worse, as expected. It is also important to remember that the effects of jet smearing are not included in this section, even though it was mentioned in section 5.1.2 that this effect also deteriorates purity. Despite the degradation of the isolation, it can be noted that the NW weight is still the most powerful observable reviewed and remains efficient at selecting the signal while discarding background considering the fundamental differences in cross-sections.

5.4 Conclusion

Several ingredients have been presented and combined into a proposed selection to isolate and measure the semi-leptonic $H \rightarrow WW^*$ process at the LHC. In particular,



Process	$\epsilon_c = 40\%$		
$W + \text{jets}$	12875	$\pm$	2149
WW	1693	$\pm$	137
$t\bar{t}$	1672	$\pm$	132
tW	365	$\pm$	22
Higgs	1058	$\pm$	32
$S/(S+B)$	0.06		

Figure 5.6: Expected yields when simulating $\vec{E}_T^{\text{miss}}$ resolution for 300 fb^{-1} .

the adaptation of the NW method to the Higgs decay to W bosons is found to be very powerful to suppress backgrounds. The method proposed is relatively simple, relying on inefficient cuts that have been chosen with the goal to maximise purity. The usage of more sophisticated multivariate techniques can improve the signal isolation considerably. Additionally, it is shown that the NW algorithm and c-tagging can be incorporated into a method to fully reconstruct the final state and thus access the WW^* system spin information. A measurement of the CGLMP Bell Operator is prototyped using the unfolding of nine observables and shown to be sensitive to entanglement. Because the proposed method fully reconstructs the final state, it can be used in the measurement of any other entanglement witness, which could facilitate its observation in $H \rightarrow WW^*$ events with the available data. It was also shown that the c-tagging efficiency is a limiting factor for both the isolation and reconstruction, which suggests that improvements to flavour tagging algorithms should enhance the background discrimination and the reconstruction resolution. A considerable amount of improvements occurred in that domain since the publication of this work, and is presented in chapter 3 with the development of the GN2 tagger, which did not exist at the time. In light of the unexpected experimental challenges discovered throughout the analysis presented in chapter 6, the simulation of the $\vec{E}_T^{\text{miss}}$ resolution was added to the detector simulation framework described in section 5.1.1. The latter is shown to have a strong degrading effect on purity, decreasing it from 18% to 6%. The inclusion of the Z+jets process was not considered since the final state has two leptons, but the analysis has shown

116 Proposal to Measure $H \rightarrow WW^*$ Entanglement in New Decay Channel

that this background can occur in our region of interest when one of the leptons is not reconstructed, and it is not negligible due to its relatively large cross-section.



6 Search for Semi-Leptonic $H \rightarrow WW^*$

6.1 Introduction

The analysis presented in this chapter builds on the techniques developed in the phenomenology study, discussed in chapter 5, to search for the semi-leptonic decay of $H \rightarrow WW^*$. The GN2 flavour tagger, described in detail in chapter 3, is used to label b - and c -jets, which is useful to both the reconstruction and the definition of regions for the statistical model. The NW algorithm, of which the core mechanism is introduced and improved in chapter 4, is used to reconstruct the neutrino as proposed in the phenomenology study. The presence of interesting and complicated mismodelling with respect to the transverse missing prevents fitting the data. The main result presented is thus the expected CL_s limits on the signal process, using the analysis pieced together throughout this thesis.

6.2 Signal MC generation and validation

The two signal processes of interest in this analysis, the ggF and VBF production of the Higgs decaying semi-leptonically via W bosons, were to be produced and verified to be used for the measurement. This section briefly describes the validation of the ggF sample only, noting that the validation of the VBF sample was done similarly using the same procedure. The configuration for a pre-existing validated sample was modified by removing a 20 GeV p_T filter applied on the lepton at parton-level. The removal of the filter is important because the targeted events tend to have soft leptons (from an off-shell

leptonic W boson). The generated ggF sample will be referred to as the “New MC” in this section, in opposition to the “Old MC” that will refer to the previous sample with a filter. The dileptonic sample was already generated and accessible for comparison. The signal was generated using the POWHEG-BOXV2 MC generator, interfaced with PYTHIA8 for the hadronic showering. A RIVET routine is written to inspect the particle-level of tens of thousand generated events and validate the settings. After these were deemed acceptable, millions of events are generated and put through the derivation chain, briefly described in section 2.7, to obtain central samples that contain truth-, particle- and reco-level MC information.

The comparison is performed to ensure that the only differences come from the lepton p_T filter, which was removed in our configuration, by applying a 20 GeV p_T cut. Figure 6.1 (a) shows the particle-level invariant mass for the combination of the two jets with the greatest p_T , and (b) the invariant mass of the combination of the lepton and neutrino momenta. Panel (a) shows that simply taking the two jets with largest p_T is not going to yield the off-shell and on-shell W mass spectrum. The observed distribution is expected, the off-shell hadronic W mass peak is difficult to access due to gluon emissions that produce additional high energy jets. Panel (b) displays both the off- and on-shell leptonic W peaks at around 35 GeV and 80 GeV respectively. This distribution is well known in Higgs to W bosons physics. Both plots show clearly that the new sample has shape differences but that a 20 GeV p_T cut on the lepton makes the distribution almost identical to the validated $l\nu q\bar{q}$ sample, which is the expected behaviour.

6.3 Background processes

Several physical processes leave a signature in the detector that passes the requirements of the preselection described in section 6.6. The regions used in the analysis are dominated by backgrounds, and it is thus crucial to understand their modelling with as much precision as possible. The main background is the W +jets process, which dominates the signal region by being responsible for roughly 90% of the events. The Z +jets process is the second most contributing background, responsible for roughly 9% of the events. Examples of Feynman diagrams for the production of a vector boson and some jets are

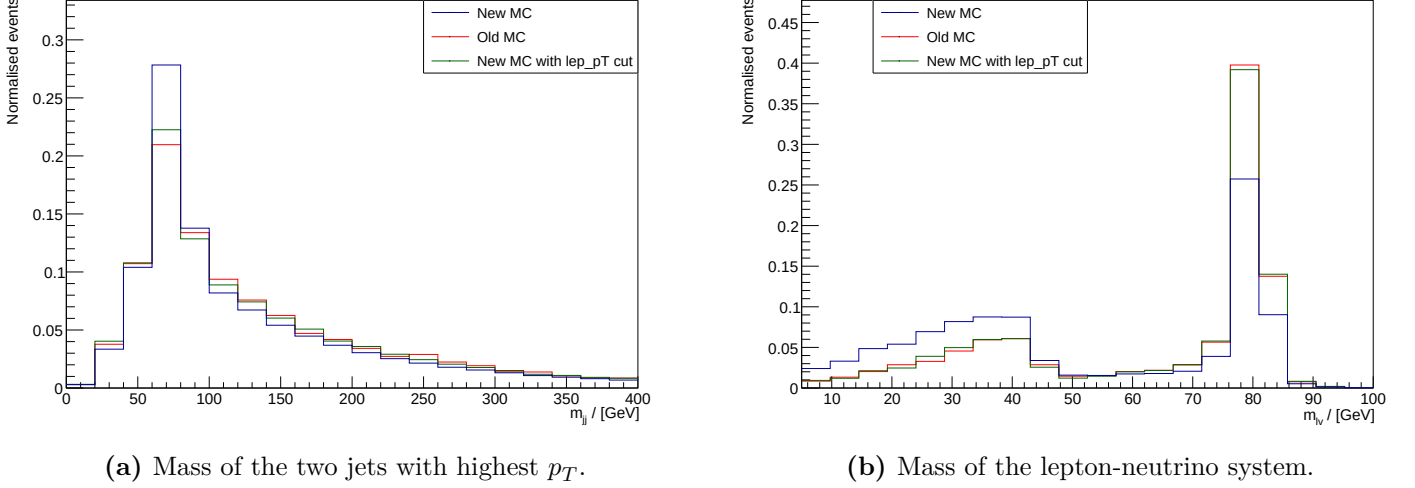
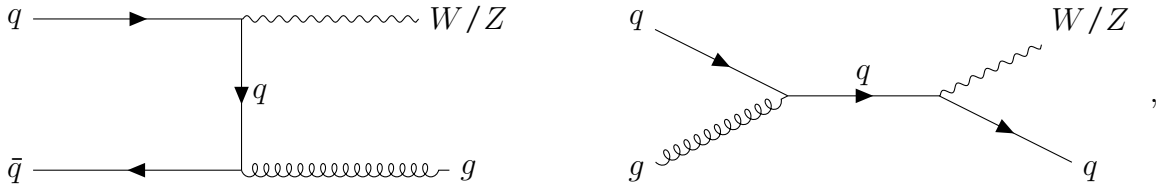


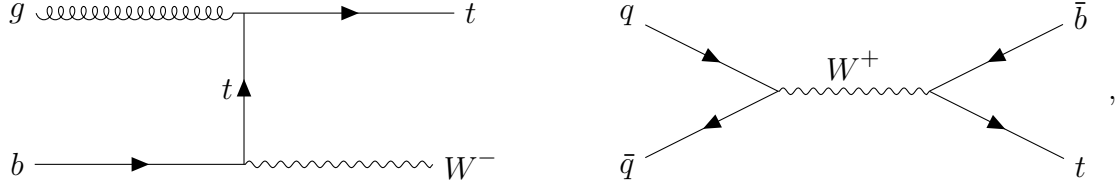
Figure 6.1: Invariant masses computed from the particle level objects obtained with the RIVET routine. The Old MC compared here is the $\nu q q$ sample with a 20 GeV lepton p_T filter. The new MC sample is shown with and without a 20 GeV lepton p_T cut for comparison.

given below:

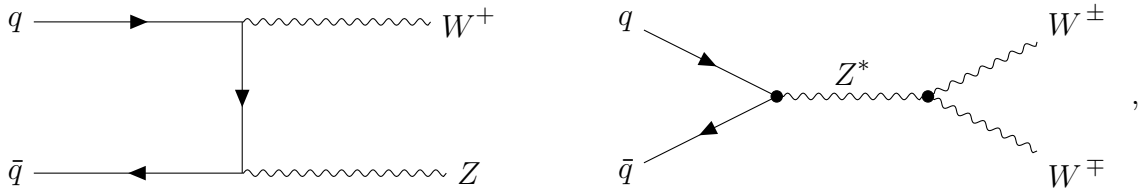


where the gluon or quark in the final state will result in one or multiple jets, and additional gluons radiated may result in more jets. The W +jets process has the W boson decaying leptonically, producing a detectable lepton and a neutrino. In the Z +jets events, the Z boson decays leptonically to two leptons, one of which is discarded due to the detector and trigger efficiencies. The absence of a neutrino means that the Z +jets process will mostly contribute at low missing transverse energy. This mimics our signal well due to the neutrino being soft when the leptonic W boson is off-shell. Despite the very high prompt lepton efficiencies (see sections 6.5.2 and 6.5.3), the Z +jets process has a relatively large cross section that results in a correspondingly important contribution. The $t\bar{t}$ and single top processes may also result in the same final state as the targeted signal. A Feynman diagram for the $t\bar{t}$ process is shown at the beginning of section 4.1,

and a couple of diagrams for the single top process are displayed below:



where the first diagram typically is referred to as tW , and the second one the single top s-channel. In this chapter, the single top process will mostly be referred to as tW despite also containing the s- and t-channel processes. This is done because both processes have relatively small contributions and behave similarly. In the tW process, either the W boson decays leptonically and the top quark into hadronically or the inverse. For the s- and t-channel cases, the top quark decays leptonically. Concerning the $t\bar{t}$ process, the overall decay must be semi-leptonic. Both $t\bar{t}$ and single top have at least one b -jet in the final state, making flavour tagging very useful to suppress these processes and define the signal and control regions. The last background modelled with MC simulation is the diboson process, which can be produced via the following Feynman diagrams:



where the left diagram may result in two W bosons, and the right vertex of the right diagram may be rotated and result in a $W + Z$ boson pair that decay to the final state. When a $W + Z$ boson pair occurs, the main contributing decay combination is $W \rightarrow \ell\nu$ and $Z \rightarrow qq$. For the two W bosons case, the main decay contributing is semi-leptonic. Finally, the non-prompt electrons and muons constitute an important background for this analysis, a dedicated data-driven estimation is constructed and described in section 6.5.

6.4 Data and Monte-Carlo samples

This analysis uses Run 2 data at a collision energy of $\sqrt{s} = 13$ TeV, collected by the ATLAS detector during the 2015-2018 period. This data corresponds to an integrated luminosity of roughly 140 fb^{-1} . The Run 3 data used to perform MC modelling studies

and estimate the fake background is at a collision energy of $\sqrt{s} = 13.6$ TeV, which was collected during the 2022-2024 period, and corresponding to an integrated luminosity of about 159 fb^{-1} . Run 3 is not included in the statistical model due to on-going work on the fake lepton background estimation and the understanding of the effects of systematics.

Every background presented in section 6.3, except for fake leptons estimated with a data-driven method described in section 6.5, is modelled by using simulated MC events centrally produced within ATLAS as separate samples per process. The generated MC events are processed through the full simulation of the ATLAS detector based on the Geant4 software [123, 124], which models the interaction of particles with the detector material and the detector response. Simulated events are reconstructed using the same software chain as used for the collision data. Additional multijet events are simulated with the PYTHIA 8 [114] generator and overlaid onto the hard-scatter event to simulate the pileup, a concept described in more details in section 2.2. The pileup distribution obtained is reweighted to reproduce the distribution observed in data.

The W +jets, Z +jets and diboson backgrounds are simulated using the Sherpa generator [168]. The matrix elements are calculated for several parton multiplicities and matched to the parton shower using the MEPS@NLO prescription [176], providing next-to-leading order (NLO) accuracy for low parton multiplicities (0, 1, 2 additional partons) and leading-order (LO) accuracy for higher multiplicities. This approach provides an accurate description of additional QCD radiation, which is particularly important for processes involving multiple jets in the final state.

The signal processes (ggF and VBF Higgs production), $t\bar{t}$ and single top events are generated using POWHEG-BOX [144], which computes hard scattering matrix element with NLO accuracy, and are interfaced to the PYTHIA 8 for parton showering. Additional hard jet emissions beyond the first prompt emissions are leading logarithmic (LL) accurate and modelled by the parton shower.

The $t\bar{t}$ and single top samples are interfaced to the A14 tune [118], which is optimised for LHC Run 2 measurements of QCD radiation and the underlying event in hadronic final states. The signal samples are showered using the AZNLO tune [177], which is derived from fits to electroweak and Drell-Yan data to describe the underlying event activity in electroweak processes. The tunes parametrise non-perturbative effects such as multiple parton interactions, colour reconnection and hadronisation, all briefly described in section 2.7, and are derived from fits to LHC and LEP data.

In the simulation of the $t\bar{t}$ sample, the POWHEG damping parameter is set to $h_{\text{damp}} = 1.5 m_t$. That parameter ensures a smooth transition between the matrix element description of hard emissions and the parton shower approximation, and is set to that value to regulate hard additional emission to improve the modelling of extra jet activity. The top quark decays are handled using MADSPIN [178], which preserves spin correlations between the production and decay. This is important for the modelling of angular variables that are sensitive to the spin correlations of the top quarks.

The ggF Higgs signal events are generated using the Multi-scale Improved NLO (MiNLO) method [179], where the renormalisation and factorisation scales are chosen event-by-event according to the kinematics of the emitted jets to obtain a more accurate description of hard jet emission. The events are subsequently reweighted such that inclusive Higgs boson observables reproduce the fixed-order NNLO QCD prediction [180]. This results in an NNLOPS sample, which combines NNLO accuracy for inclusive observables, and NLO accuracy for Higgs+jets configurations.

6.5 Lepton fakes estimation

The leptons of interest, muons and electrons generated by the hard-scattering process, are called *prompt* or *real*, as opposed to *non-prompt* or *fake* leptons that come from other sources occurring post-collision or from the underlying event. Fakes cannot be identified directly in data and are thus another background to model with the help of MC. Fakes are very important to consider in this analysis: the strategy of selecting events with off-shell leptonic W has the disadvantage of providing a working phase space with relatively soft leptons and low $\vec{E}_T^{\text{miss}}$, which is the region where most fakes are located. This is physically expected when considering the sources for fake leptons. Fake electrons mainly come from misidentified jets and photon pair production in the detector. Both non-prompt muons and electrons are produced from heavy-flavour hadron decays inside b - and c -jets. In-flight decays of pions and kaons can also contribute slightly to the production of fake muons, but heavy-flavour decays remain the dominant source. The processes all produce secondary leptons that tend to have significantly smaller p_T than leptons produced in electroweak decay of W or Z bosons due to the considerably larger masses of these bosons. Requiring exactly one lepton (electron or muon) per event naturally selects physical processes that include leptons from W boson decay, producing

a prompt neutrino that contributes to E_T^{miss} , which is not the case for events with fake leptons.

There are a few approaches available to estimate the fake lepton rates in regions of interest. The method chosen for this analysis is called the *Asymptotic Matrix Method* (AMM) [181]. Like most other methods, the estimation relies on the comparison of two lepton working points, not to be mistaken with flavour tagging WPs introduced in section 3.1. The electron and muon working points are defined based on two properties of the reconstructed physical objects: their *identification* and *isolation*. These concepts are described in more detail in section 2.6.4 for electrons, and section 2.6.5 for muons. The working points are named *tight* and *loose*, referring to their relative lepton acceptance. The identification and isolation combinations used for each working point, and for each data-taking period, are summarised in Table 6.1. These combinations are constrained by the availability of trigger efficiency corrections computed centrally within ATLAS. In Run 2, the electron identification is the same for both working points, namely **TightLH**, which relies on a multivariate likelihood discriminant. The electron isolations used are **TightVarRad** and **LooseVarRad** for the tight and loose working point respectively. The term **VarRad** refers to Variable Radius, which is an isolation method using a cone with a radius (in the $\phi - \eta$ plane) that scales with the lepton p_T to take into account boosting effects. The muon identification quality is **medium** for both working points, while the muon isolation working points are **TightVarRad** and **LooseVarRad**, similarly to electrons. For Run 3, the only change is for the electron identification for the loose working point. Instead, the **LooseBLayerLH** identification is used, which includes additional information from the innermost pixel B-layer. The analysis uses tight leptons to define the regions used in the likelihood fit and study the physics of the channel, the loose leptons are only kept to estimate the fake lepton background contribution.

Lepton	Working point	Data-taking period	
		Run 2	Run 3
Electron	Tight	TightLH + TightVarRad	TightLH + TightVarRad
Electron	Loose	TightLH + LooseVarRad	LooseBLayerLH + LooseVarRad
Muon	Tight	Medium + TightVarRad	Medium + TightVarRad
Muon	Loose	Medium + LooseVarRad	Medium + LooseVarRad

Table 6.1: Summary of the electron and muon working points used in Run 2 and Run 3. The information is presented as “identification + isolation”.

6.5.1 The Asymptotic Matrix Method

If $N_{\text{real/fake}}^{\text{tight/loose}}$ represents the number of events with real (fake) leptons that pass the tight (loose) working point, then the real and fake efficiencies are defined as follows:

$$\epsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}}, \quad (6.1)$$

$$\epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}. \quad (6.2)$$

Following the matrix method, the number of tight and loose events for a given bin respects the following system of equations [181]:

$$\begin{bmatrix} N_{\text{tight}} \\ N_{\text{loose}} \end{bmatrix} = \begin{bmatrix} \epsilon_{\text{real}} & \epsilon_{\text{fake}} \\ 1 - \epsilon_{\text{real}} & 1 - \epsilon_{\text{fake}} \end{bmatrix} \begin{bmatrix} N_{\text{real}} \\ N_{\text{fake}} \end{bmatrix}, \quad (6.3)$$

and thus the number of events with fake leptons passing the tight working point is then given by inverting the matrix, taking the bottom row and multiplying by ϵ_{fake} :

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} \left(\epsilon_{\text{real}} N_{\text{loose}} + (\epsilon_{\text{real}} - 1) N_{\text{tight}} \right). \quad (6.4)$$

The result in equation 6.4 can then be used to compute the number of fakes per bin per histogram. A more convenient approach is to turn equation 6.4 into a weight to be applied per event instead:

$$w = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (\epsilon_{\text{real}} - f), \quad (6.5)$$

where f is a flag equal to 1 when the lepton is tight and 0 otherwise. These weights are applied to the data for the loose lepton working point, in what is called a *data-driven* approach. The reweighted data is then taken as the estimation for the fake lepton background, and histograms can be built for the distribution of any observable. In this scheme, the efficiencies ϵ_{real} and ϵ_{fake} are binned in the $(p_T - \eta)$ 2D phase space. The fake efficiencies ϵ_{fake} are estimated using data, in a *fake enriched* region obtained by requiring $|\vec{E}_T^{\text{miss}}| < 30$ GeV and $m_{W_T} < 30$ GeV, where m_{W_T} is the expected transverse mass of the W boson. The latter is defined as follows:

$$m_{W_T} = \sqrt{2p_T^\ell |\vec{E}_T^{\text{miss}}| (1 - \cos(\Delta\phi))}, \quad (6.6)$$

where p_T^ℓ is the transverse momentum of the lepton and $\Delta\phi$ the angle difference between the lepton momentum and $\vec{E}_T^{\text{miss}}$ in the transverse plane. The real efficiencies ϵ_{real} are estimated using MC in an orthogonal region defined by requiring $|\vec{E}_T^{\text{miss}}| \geq 30$ GeV, and the truth level information is used to filter out fake leptons. The MC events with a prompt lepton are subtracted from the data to get ϵ_{fake} as follows:

$$\epsilon_{\text{fake}} = \frac{(N_{\text{data}} - N_{\text{MC}}^{\text{real}})|_{\text{tight}}}{(N_{\text{data}} - N_{\text{MC}}^{\text{real}})|_{\text{loose}}}, \quad (6.7)$$

where parentheses are used to indicate the lepton working point. Using the same notation, ϵ_{real} , evaluated only with MC, is given by:

$$\epsilon_{\text{real}} = \frac{(N_{\text{MC}}^{\text{real}})|_{\text{tight}}}{(N_{\text{MC}}^{\text{real}})|_{\text{loose}}} \quad (6.8)$$

This approach has two main drawbacks that require attention: the estimation of fakes is affected by the statistics of the binned efficiencies ϵ_{real} and ϵ_{fake} , and looking at equation 6.5, it is possible for the weights to be negative. The statistical uncertainty on the estimated number of fakes in a given region N_{fake} is given by:

$$\sigma_{N_{\text{fake}}} = \sqrt{\sum_i^N w_i^2}, \quad (6.9)$$

where i iterates over all N events in the relevant region. It is clear that the method becomes valid when efficiencies are computed with a relatively large number of events. The

combination of this well-known statistical asymptotic limit and the matrix formulation of equation 6.3 is what gives the AMM its name.

As described in more detail in section 6.6, the analysis makes use of b -jet vetoing in signal regions and prescription in some control regions. In addition, the regions are split by the number of jets present in the event (2, 3, or 4 jets). All these selection criteria affect the rate of fake leptons and so a set of efficiencies need to be computed for each one of them.

6.5.2 Run 2 efficiencies

Figure 6.2 shows the fake efficiencies for both electrons and muons separately in the 3 jets region, binned in p_T and η , as defined in equation 6.7. The leptons with a transverse momentum greater than 1000 GeV are accumulated in the relevant highest p_T bins. The bins are smaller at low p_T to get a better resolution where the fakes are expected to be dominant, and because it is also where most of our signal yield is expected. Those efficiencies are used to produce event weights for loose data, following equation 6.5, which yields the lepton fakes contribution as a background process. The electron fake efficiency increases with p_T and η , indicating that fake leptons with higher momentum and pseudo-rapidity are more likely to pass the tight working point, which is expected. However, the relatively low statistical uncertainties indicate that many electrons are found with a high p_T and $|\eta|$, which is unexpected and suggests that the fake-enriched region is contaminated with an important number of real leptons. This has an important impact on the final fake lepton estimation. A narrow bin has been defined in the range $|\eta| \in [1.37, 1.52]$ for electrons to reflect the gap between the barrel and end-cap region in the geometry of the detector. The electrons with a pseudo-rapidity in that range are vetoed in our preselection, as described in section 6.6, but some still fall into that bin due to small resolution effects at its edges. The uncertainty on the efficiencies is relatively small overall, except for the bin that was just described, where, as expected, very few electrons are present. Very few fake muons pass the tight working point with $p_T > 50$ GeV, which results in mostly empty bins in that region. Figure 6.3 shows the real electron and muon efficiencies. The range spanned for the bin contents clearly indicates that the real lepton efficiencies are higher than for the fake leptons, which is expected since the detector has been designed to target real leptons. The shape of the real lepton efficiencies seem to be dominated by p_T , with almost no change over $|\eta|$. The statistical uncertainty is very low in all bins except for the small gap bin for

electrons, and a bin with barely any muons displaying an uncertainty close to 1. Figure 6.4 shows the one-dimensional distribution of efficiencies obtained after marginalising over $|\eta|$ for all regions, including the control regions without the b -veto. The regions requiring a b -veto generally have considerably greater fake efficiencies than regions with a b -filter (events with at least 1 b -jet). The fake efficiency quantifies the probability that a non-prompt lepton passing a loose selection also passes the tight selection, as can be seen from equation 6.7. The b -veto removes an important source of fake leptons originating from inside the b -jets. Those leptons perform poorly with respect to the isolation criteria due to the large surrounding activity originating from b -hadrons chain decays. Thus, such leptons are efficiently removed by switching from the loose to the tight working point, which results in lower values for ϵ_{fake} . In events without b -jets, the leptons come mostly from jets with lighter flavours and photon conversions for electrons. Such leptons do not have as many energetic particles surrounding them within a ΔR cone thus making them more likely to pass the isolation criteria, which results in higher values for ϵ_{fake} .

6.5.3 Run 3 efficiencies

Figure 6.5 shows the distribution of the efficiencies marginalised over $|\eta|$ for Run 3. The most important difference between Run 2 and Run 3 is for the electrons. The electron fake efficiencies for regions with a b -veto lie between 0.4 and 0.6, while the regions with a b -filter have values ranging from 0.2 to 0.4. In the Run 2 efficiencies shown in Figure 6.4, the corresponding efficiencies lie in the intervals $[0.6, 0.8]$ and $[0.2, 0.6]$, respectively. There is an overall decrease in efficiency compared to Run 2, which is also visible for ϵ_{real} . The real efficiency increases from 0.9 to 1.0 in Run 2 against 0.8 to 0.96 in Run 3.

This behaviour is explained by the usage of a looser loose working point for the electrons, as can be seen from Table 6.1. Keeping the same tight working point and loosening the looser working point will naturally decrease the efficiencies given their definition in equations 6.7 and 6.8. A greater difference between the working points improves their performance in estimating the fake lepton background contribution. Intuitively, increasing the separation between the tightness of the two working points considered enhances the discrimination power of the set of events that are loose and not tight, which becomes increasingly enriched in fake leptons. Conversely, tightening the tighter working point leads the tight set of events to be more dominated by genuinely prompt leptons. Mathematically, this behaviour can be understood from equation 6.5, where the AMM method depends on the difference between real and fake efficiencies in

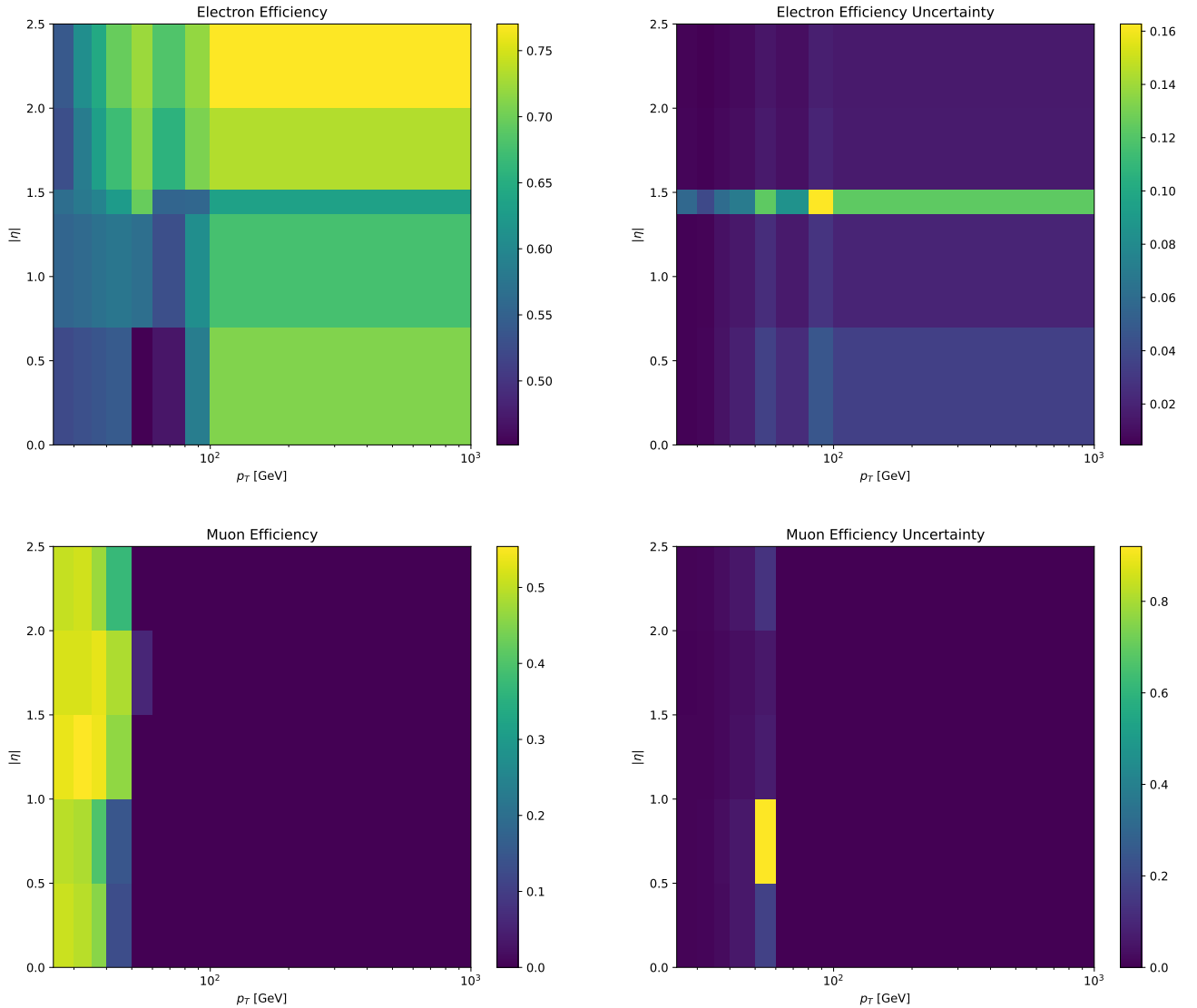


Figure 6.2: Fake charged lepton efficiencies and their statistical uncertainties in Run 2, binned in p_T and η , for the 3 jets region with b -veto.

the denominator. This term originates from the determinant of the matrix, which is used during the inversion. As the difference $\epsilon_{\text{real}} - \epsilon_{\text{fake}}$ increases, the matrix becomes less degenerate, improving the stability of the inversion. Another effect of loosening the lepton working point is to increase the acceptance of the electrons and thus the available statistics. The difference between regions with different number of jets seems to be smaller compared to Run 2, especially for the bins with higher p_T . A clear drop occurs at 60 GeV for the fake efficiencies, an energy coinciding with one of the electron triggers. This bias introduced by the trigger is less important, but still visible in Run 2.

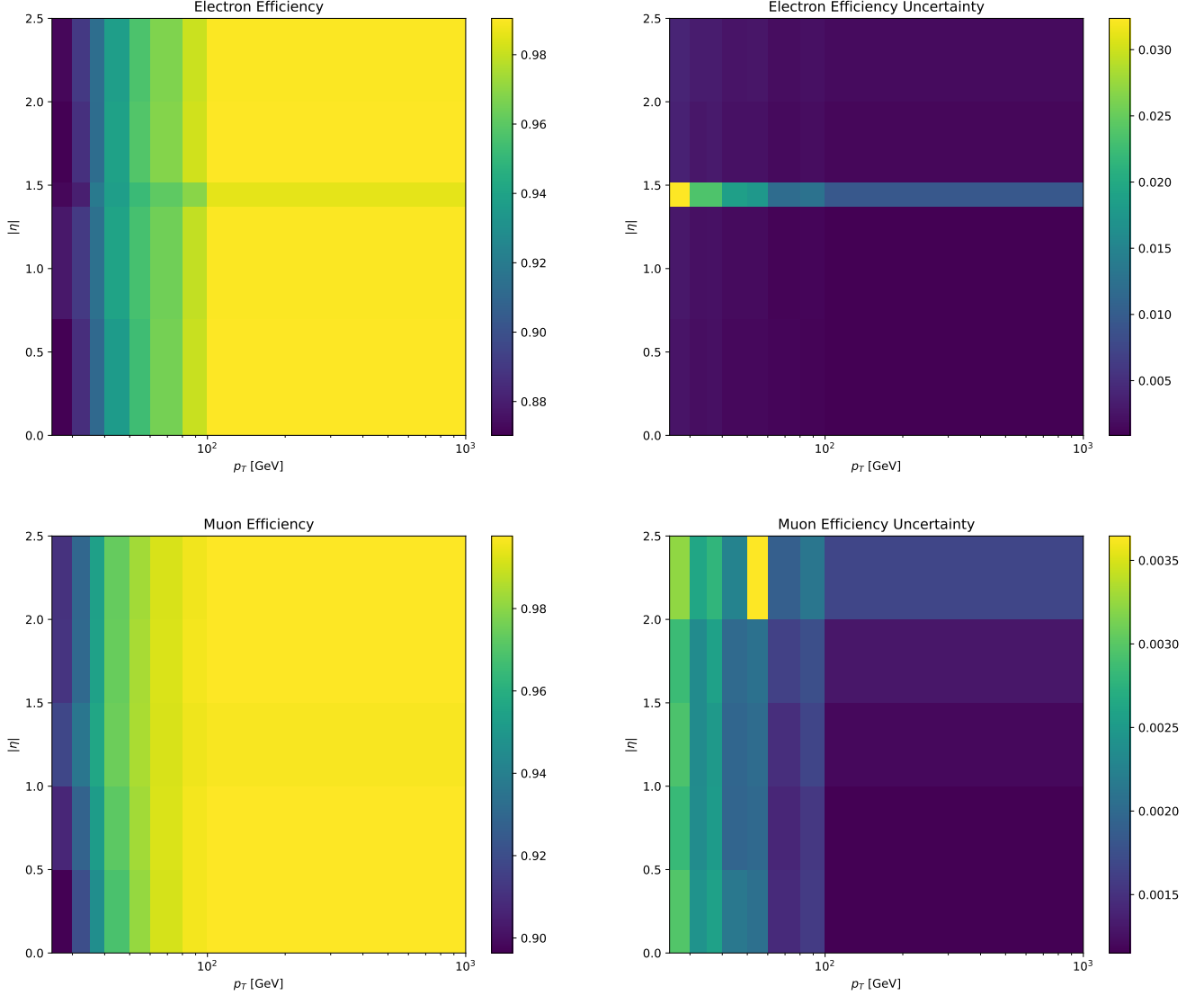
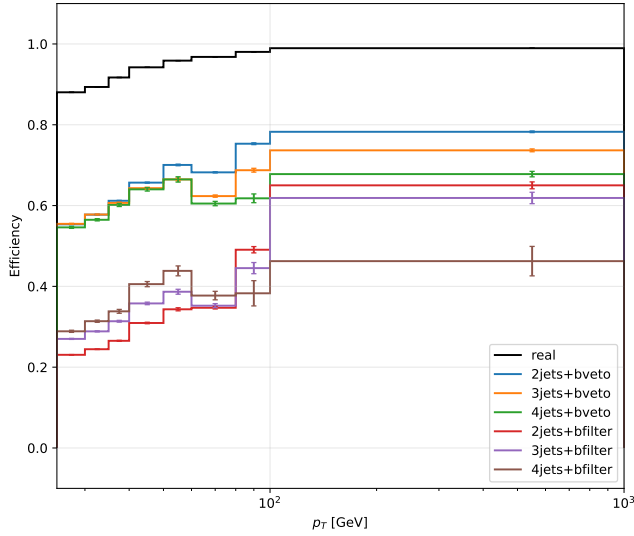
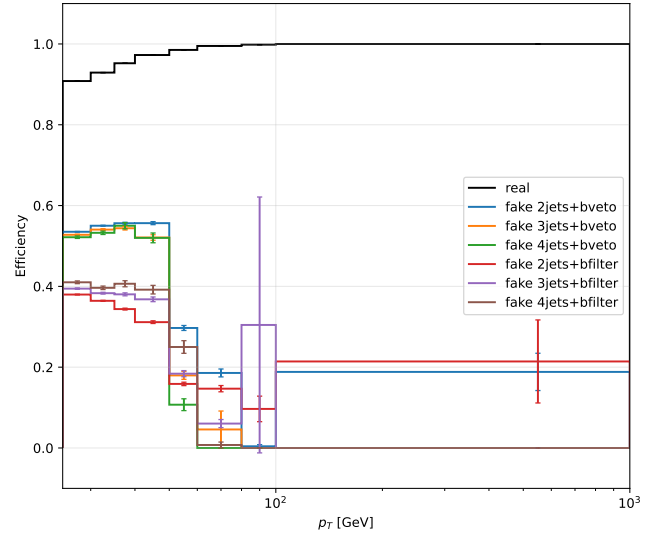


Figure 6.3: Real charged lepton efficiencies and their statistical uncertainties in Run 2, binned in p_T and η .

The muon efficiencies are much more similar between Run 2 and Run 3, which is expected since the tightness of the muon working points remains the same. The largest difference is at high p_T , where the fake efficiencies for the regions with a b -veto have much more statistics and a relatively high value. This makes apparent the trigger bias, visible as a sudden gap in the fake efficiency curves at 50 GeV, which coincides with the trigger energy threshold. The fake efficiencies at low p_T are also rising to higher values, up to $\epsilon_{\text{fake}} \approx 0.7$, whereas in Run 2 the curves remain below the 0.6 line. Following the same reasoning as for the electrons, because fake efficiencies are greater while the real

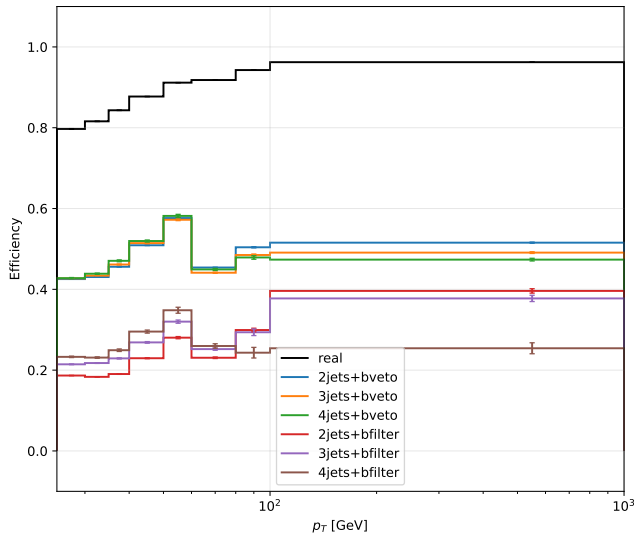


(a) Electrons

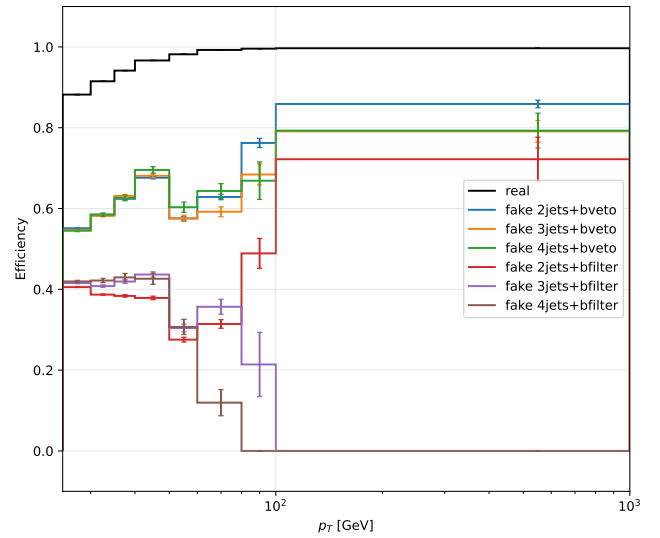


(b) Muons

Figure 6.4: Charged lepton efficiencies, marginalised over η for electrons and muons, for Run 2.



(a) Electrons



(b) Muons

Figure 6.5: Charged lepton efficiencies, marginalised over η for electrons and muons, for Run 3.

efficiencies remain very similar compared to Run 2, the estimation of the fake leptons is expected to be slightly less performant in Run 3. The overall decrease in efficiency can be explained by the lepton isolation cuts that have been added to the triggers during Run 3,

which suppress the contribution of fake leptons in every analysis. A greater acceptance of fake leptons in the tighter working point is also expected as a result.

6.6 Preselection

The events to be used for the measurement are first selected based on a few requirements on the physics objects, called a *preselection*. The event reconstruction, where physical observables are computed and stored, a process described in section 6.6.2, is performed for all preselected events. Finally, these events are split into multiple *regions* by applying more selections that use the observables computed with the event reconstruction step.

The preselection can be detailed by iterating through the physical objects of interest. Every event is required to have exactly one electron and no muons, or exactly one muon and no electrons. The muons are required to have $p_T > 25$ GeV and $|\eta| < 2.50$. Electrons are required to have $p_T > 25$ GeV and $|\eta| < 2.47$, and additionally to be outside the detector transition region between the barrel and the end-cap, delimited by $1.37 < |\eta| < 1.52$. Two lepton working points of interest are defined taking into account identification and isolation criteria, their definitions varying slightly between Run 2 and Run 3. The consideration of two lepton workings points is necessary to estimate the background contribution from fake leptons, a procedure that is described in more detail in section 6.5.

Events selected need to have between two and four jets of radius $R = 0.4$ (small- R), with $p_T > 20$ GeV and $|\eta| < 2.5$ as kinematic requirements. The flavour tagging information is not used in the preselection, but the c -tagging PCFT quantiles (defined at the end of section 3.3) are instrumental to the event reconstruction. Both b - and c -tagging quantiles are saved to be used in later selections and in the multivariate analysis.

Since jets and leptons are physical objects built from information measured within a finite radius R in the η - ϕ phase-space, it is important to avoid double counting by using *overlap removal*. This procedure will then depend on the isolation of the leptons and thus needs to be performed once per lepton working point (see section 6.5). This results in two sets of jets per event: a loose lepton might not pass the tight working point and might be reconstructed as a jet instead, depending on the overlap removal. In this scenario, the event will only pass the preselection at the loose working point since at the tight working point it is lepton-less. The set of jets associated with the loose leptons

should be the same for events with tight leptons, and is thus used to count the number of jets per event regardless of the lepton working point. Overlap removal is also important in the computation of transverse missing energy, whose definition involves the sum of the transverse momentum of all physical objects. The lepton working points can change the physical objects considered, and thus will affect the components of $\vec{E}_T^{\text{miss}}$. For that reason, two missing transverse energies are obtained per event as well, one per lepton working point.

6.6.1 JetPairingNet

The hadronic W is reconstructed by combining two jets in the events. When an event has a number of jets N_{jet} , there are $\frac{N_{\text{jet}}^2 - N_{\text{jet}}}{2}$ possible pairs of jets to select from. A reconstructable event is defined as an event in which both parton-level daughters of the hadronic W can be ΔR -matched to one of the reco-level jets. It is possible for the two parton-level daughters to be collimated and thus matched to the same jet, occurring for about 20% of the wrongly-reconstructed reconstructable events. A wrongly-reconstructed event is defined as an event where at least one of the jets selected to reconstruct the hadronic W cannot be matched to one of its parton-level daughters. Inspection of the parton-level information revealed that roughly 37% of events passing the pre-selection are reconstructable. Non-reconstructable events mainly occur when a final state quark has too low p_T to produce an observable jet and additional gluon radiation makes the event pass the jet multiplicity requirements. Initially, the algorithm used to select the jet pair of interest was designed based on the findings presented in the study of chapter 5:

- Use the 50% c -tagging WP and reject events that do not contain a least one c -jet (c -jet filtering) and one light-jet.
- Compute the invariant mass of the sum of four-momentum of every possible c -jet and light-jet pair.
- Make the assumption that events of interest have dominantly on-shell hadronic W s. The invariant mass closest to 80 GeV then corresponds to the best pair.

This algorithm was found to be too simple: most jet pairs have an invariant mass close to 80 GeV, making this requirement a weak discriminating criterion. Only 57% of reconstructable events (increasing to 63% when NW reconstruction is required to find a solution) have their pair successfully selected. Instead, a more sophisticated, neural network-based approach was chosen to select the jet pairs. Figure 6.6 shows the different

components of the neural network, creatively named “JetPairingNet”. The black boxes (figuratively and literally) represent sub-networks, the trapezoidal shapes are Multi-Layer Perceptrons (MLPs) and the thin rectangles are a single dense layer outputting a single scalar quantity. First, the red boxes j_{ij} , each representing a jet pair (j_i, j_j) , are assembled by concatenating the sum and absolute subtraction of the jet vectors j_i and j_j : $j_{ij} = [j_i + j_j, |j_i - j_j|]$. This is done to make the jet pair representation invariant under the swapping of the two jets $i \leftrightarrow j$. Each jet vector contains five entries: the jet’s p_T , η , ϕ , energy and flavour tagging quantile, thus making the jet pair representation vector j_{ij} of size 10. It can be noted that, while ϕ is not a useful quantity to learn from directly, the addition and subtraction of the azimuthal angles of the two jets stored in the pair representation contain useful information. The jet pairs are then each individually input into an MLP with a single hidden- and output-layer, both of size 16. The augmented new jet pair representations are labelled z_{ij} and represented by orange boxes. In parallel, the event features vector is processed. It contains the lepton’s three-momentum and energy as well as both missing transverse energy components, thus making six entries in total. The event features vector is input into two separate MLPs with the same size as the MLP processing the jet pairs, each outputting a vector of same dimension as the jet pair representations z_{ij} . These vectors, labelled γ and β , are used to apply a linear transformation to the components of each z_{ij} element, updating it into a new representation h_{ij} : $h_{ij} = \gamma z_{ij} + \beta$. This mechanism is referred to as Feature-wise Linear Modulation (FiLM) [182]. The FiLM method is used here to inject information from the other objects of interest within the event that are independent from the jets, while keeping the jet pairs the central focus architecturally. That last jet pair representation elements h_{ij} are then passed through a dense layer that outputs a scalar per element. Those scalars are subject to a softmax normalisation to obtain probabilities which are taken as the final “pair scores”, used to choose the best jet pair. Finally, the jet pair representations h_{ij} are summed, weighted by their score s_{ij} into an event-level representation that is passed through a final dense layer that outputs a scalar. The sigmoid function is used to convert the scalar into a probability, which is taken as the so-called “no-pair score”, used to establish whether an event possesses a valid jet pair. The loss function has two terms, one to steer the training of the pair scores and the other for the no-pair score. The loss function can be written as follows:

$$L = \alpha L_{CC}(y_{\text{pairs}}, \hat{y}_{\text{pairs}})(1 - \hat{y}_{\text{no-pair}}) + L_{BC}(y_{\text{no-pair}}, \hat{y}_{\text{no-pair}}), \quad (6.10)$$

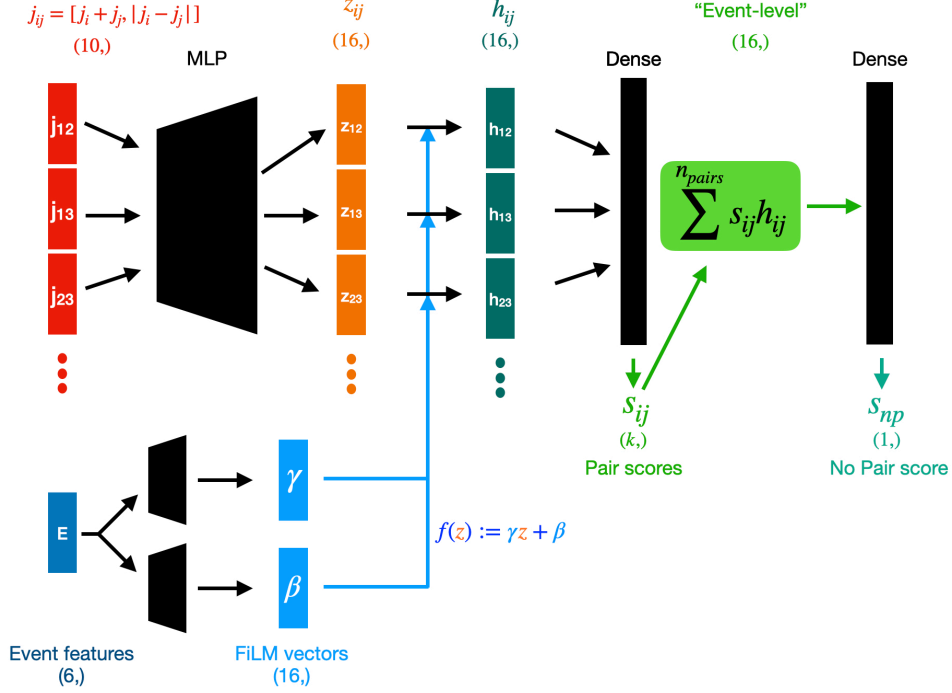


Figure 6.6: Illustration of the JetPairingNet neural network.

where L_{CC} is the Categorical Cross-entropy (CC) function, L_{BC} the Binary Cross-entropy (BC) function, y_{pairs} and $y_{no-pair}$ are the outputs of JetPairingNet as described previously, $\hat{y}_i$ denote the corresponding true label used for training, and α is a real number parameter that may be used for balancing of the terms (set to one in the work presented here). The label $\hat{y}_{no-pair}$ will be equal to one when there are no valid pairs in the event, thus removing the effect of the loss that affects the pair score. This ensures that the parameters of the neural network responsible for computing the pair scores do not learn from events without pairs. It is interesting to note that JetPairingNet may be used on other objects than jets, and can easily be converted to a situation where more than one pair is sought after by swapping the CC function to a BC function in the loss.

Training

The training is performed on ggF signal events organised in two folds of a hundred thousand events each (20% of which are used for validation). The events are split following the parity (even vs odd) of their event number. The training dataset itself is divided into batches of two hundred events, the loss and accuracy is averaged over every batch per epoch. The labels are produced for the training dataset by using ΔR matching

of the reco-level jets with the parton-level daughters of the hadronic W . Specifically, a “valid” pair of jets is found in an event when two jets have $\Delta R < 0.4$ with the hadronic W daughter particles. Figure 6.7 shows the evolution of the loss against training epochs. An interesting bump can be seen around 30 epochs for both the pair score and no-pair score losses. The no-pair loss seems to flatten out relatively earlier at around epoch 175, whereas the pair score loss is still decreasing visibly until epoch 250. The dashed training curves and the validation solid curves mostly overlap throughout the training, a clear indication that the overtraining is very limited.

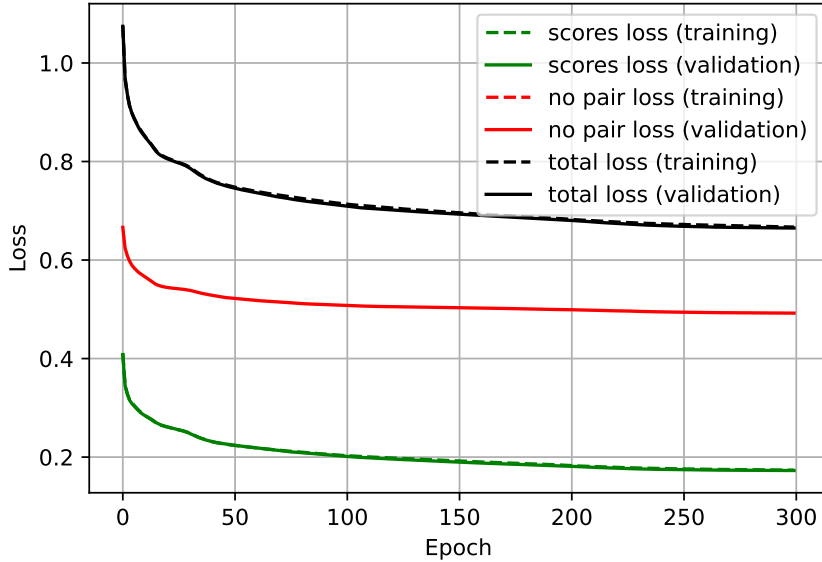


Figure 6.7: Loss against epochs during the training of the JetPairingNet on the even fold.

The improvement of accuracy is shown in Figure 6.8 for the same network during the same training session. Similarly to the loss function curves, the pair and no-pair accuracy curves can be cut into three sections: an initial fast increase from epochs 0 to roughly epoch 20, the interesting bump from epoch 20 to epoch 40, and beyond a slower steady progression. The first section shows that the network quickly moves away from random guessing, with significantly higher accuracy, around 15% greater than what random guesses provide. It is also in that first section that the performance of the network goes beyond the previous simpler invariant mass algorithm. The bump section might show a potential to be overcome to finally reach the last section that resembles a slow fine tuning of the parameters. At the end of the training, an accuracy of around 85% is reached, representing an overall improvement of 22% over the simpler algorithm. Just as

shown by the loss curves, the no-pair score accuracy stops progressing around epoch 175, stopping around 76%.

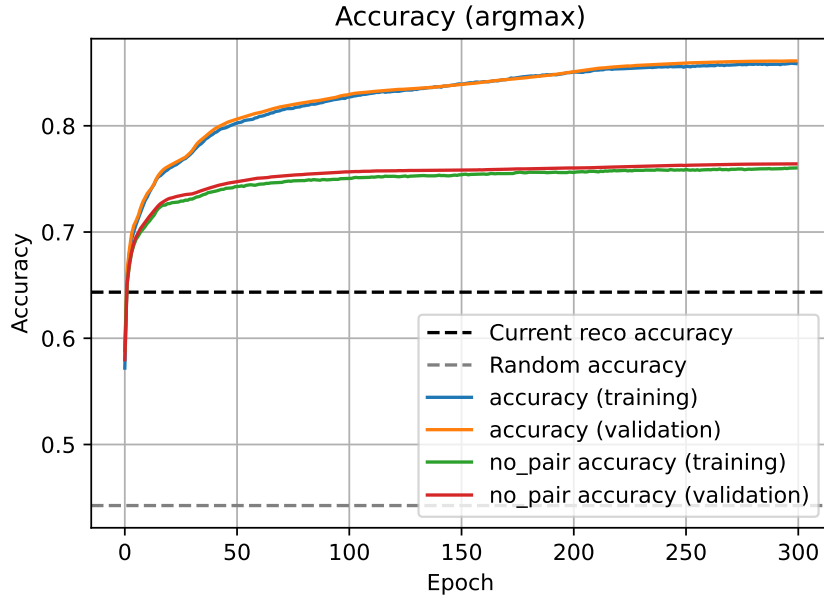


Figure 6.8: Accuracy against epoch during the training of the JetPairingNet on the even fold. The black dashed horizontal line shows the accuracy of the naive algorithm relying on c-tagging and the invariant mass of combination of jets. The gray dashed horizontal line displays the accuracy of guessing at random.

The ROC curve for the no-pair score is shown in Figure 6.9. The Area Under the Curve (AUC) metric is around 0.84, which is an acceptable performance, but indicates that simply cutting on the no-pair score might not be very efficient. Looking at the curve itself, cutting at 80% true positive rate leaves about 27% of events without a true pair, which is a large fraction considering only 37% of events have a valid pair. The no-pair score has thus more potential to define regions with very tight cuts or as an input to other machine learning models used in background discrimination.

Performance

The overall improvement can be seen indirectly in the NW reconstruction: switching the simple invariant mass algorithm for JetPairingNet reduces the number of ggF events without a NW solution by approximately 42%. A more solid and concrete method for asserting the performance is to look at the parton- and reco-level matching. Figure 6.10 shows the distribution of the parton-level hadronic W mass m_{W_h} , and displays the

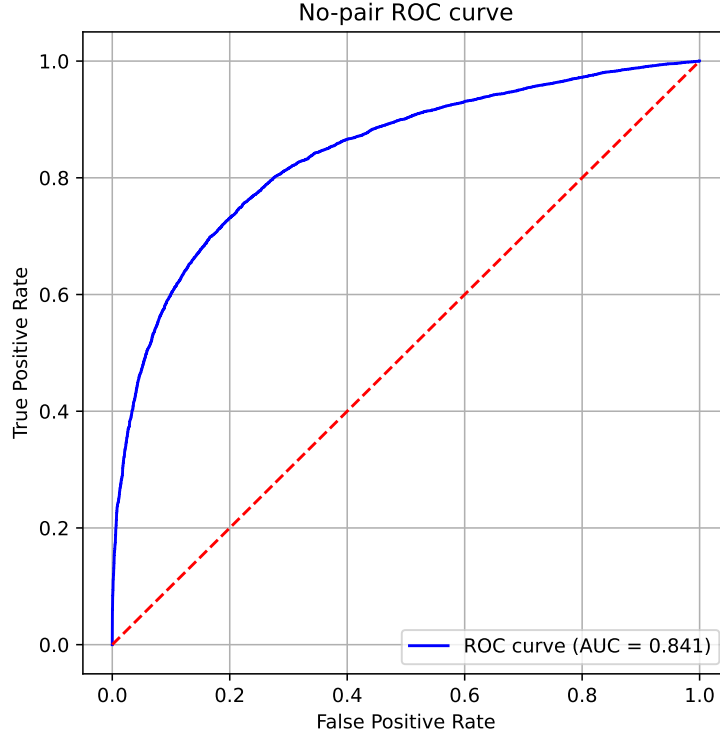


Figure 6.9: ROC curve for the no-pair score of the JetPairingNet model trained on the even fold.

fraction of events where the jet pair matches the parton-level W daughters. JetPairingNet increases the reconstructed event rate from 56.70% to 77.35%, which is a significant improvement. The plots also highlight another key change of using JetPairingNet: more events with off-shell hadronic W s are reconstructed. The previous algorithm made the assumption that the hadronic W is on-shell, working with the hypothesis that a post-selection cut on the hadronic W mass would be enough to discard most events with off-shell W s. This turns out to be incorrect and off-shell W events pass the selection with a high probability that the jet pair is misidentified. JetPairingNet does not make that assumption and considers events with an on-shell or off-shell W the same, which results in this performance improvement. Another key upgrade is the removal of a requirement for c -tagging: the simple invariant mass algorithm requires at least one c -jet and one light-jet, leading to the requirement that there is at least one c -jet in the event (c -jet filtering). This comes at the cost of discarding a considerable number of events that do not have a c -jet (and the few events with only c -jets and no light-jets). JetPairingNet does not require a c -jet, but still uses c -tagging information by taking the PCFT quantile of each jet as

an input, a flavour tagging concept described in more detail in section 3.3. Figure 6.11 shows the fraction of reconstructed events with the m_{W_h} distribution on events without c -jet filtering. The algorithm requiring a c -jet only reconstruct about 16% of events, a poor performance which is expected by design. JetPairingNet, despite being trained on c -jet filtered events only, learned information about the event that is more relevant than the c -tagging and on-shell mass condition alone, and performs unexpectedly well with an approximate 73% reconstruction rate. This ability to reconstruct the hadronic W for events without c -jets extends the available phase-space considerably.

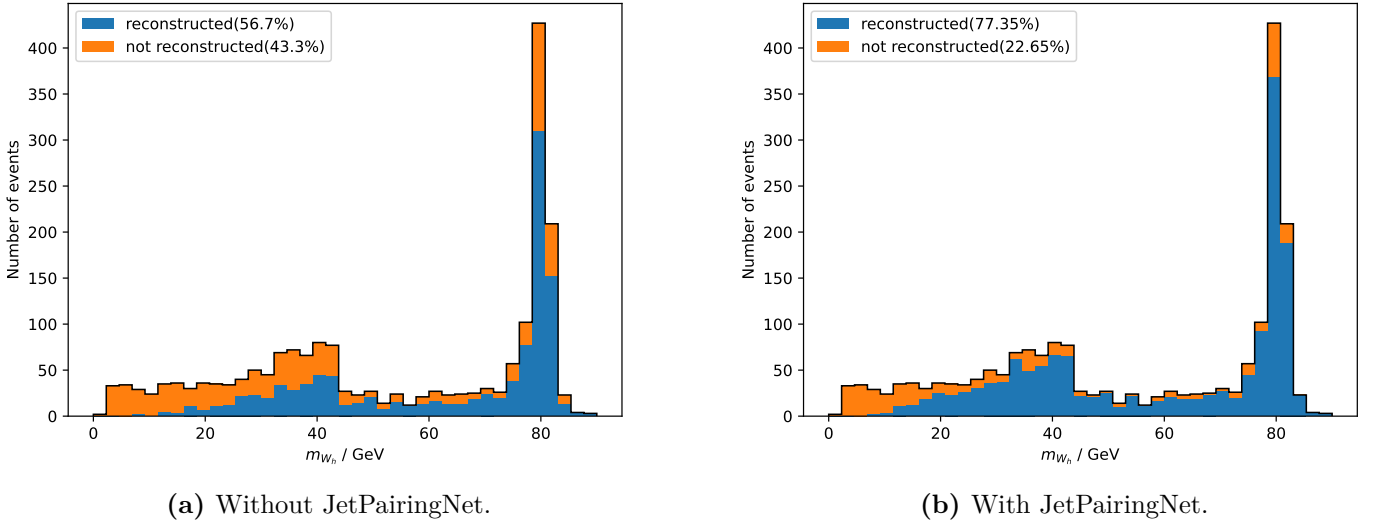


Figure 6.10: Distribution of the parton-level mass of the hadronic W , for reconstructable events only. The stacked histogram separates events that have both the parton-level W daughters ΔR match reco-level jets (reconstructed events) and events that do not (not reconstructed events).

The improvements that bring JetPairingNet are very encouraging, and it is important to note that there is still room for several optimisations. Firstly, no hyperparameter optimisation has been performed: the size of the MLPs, the dimensionality of the intermediary jet pair representations, the α coefficient in the loss and the learning rates are all obvious parameters that would need to be fine-tuned to maximise performance. The training only involved ggF produced signal for Run 2, but might benefit from also including the VBF production mode (it was noted in section 4.4.2 that the ggF events have a higher NW reconstruction rate than the VBF samples) and Run 3 events. The no-pair score could be upgraded to a background rejecting variable by adding background into the training dataset, although this must not come at the detriment of the jet pairing

performance, the core task of JetPairingNet. Studies on class balancing and feature importance would be very instructive and could steer the potential future improvements.

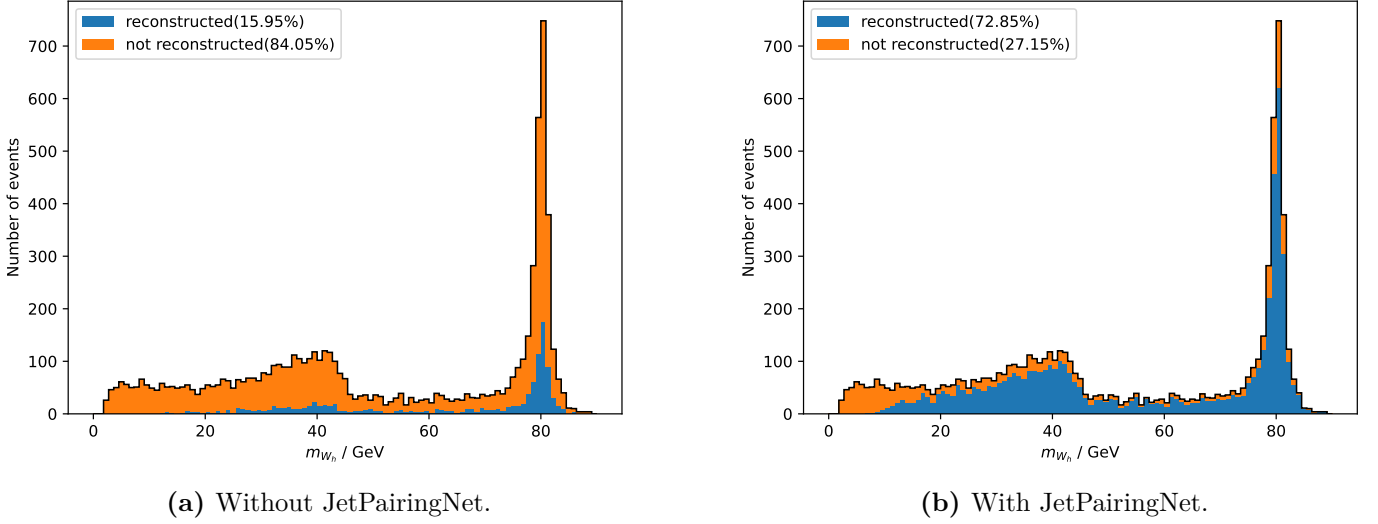


Figure 6.11: Distribution of the parton-level mass of the hadronic W , for reconstructable events only. The same information as Figure 6.10 is presented but for events that do not have c -jet filtering.

6.6.2 Event Reconstruction

An algorithm is applied on every preselected event to perform the reconstruction of important observables that are used to isolate the signal from the backgrounds. The four momentum of the lepton, the four momentum of every jet in the event, the jet PCFT quantiles, and the missing transverse energy are gathered. All these quantities are input into JetPairingNet, described in section 6.6.1, which outputs a score per possible jet pair, as well as a no-pair score, associated with the belief that there is no valid pair in the event. Two models of JetPairingNet, trained on event numbers with different parity, are used to avoid bias from the training. The jet pair with the largest score is taken as coming from the hadronic W and the no-pair score is saved for studying and potential later use in multivariate analysis. The jets, labelled $jet1$ and $jet2$, have their four momentum combined to produce the reconstructed hadronic W . The newly produced hadronic W , the lepton and missing transverse energy are then used by the NW algorithm (detailed in chapter 4) to reconstruct the four momentum of the neutrino

and the leptonic W . The Higgs four momentum is then obtained by combining the two W four momenta and the NW weight is saved for its background discrimination power. When an event does not have a NW solution, the Higgs four momentum is defined as the sum of the transverse missing energy and the hadronic W . Every four momentum is saved to be studied, produce additional observables post reconstruction, and serve as input to the multivariate analysis.

6.7 Multivariate Analysis

A small piece of software, daringly named **MLTrainer**, was developed to streamline the creation and training of simple Machine Learning (ML) based models. The software is fed a yaml configuration file that specifies the number of classes, paths to input ROOT files, plotting options and region definitions. For each region definition, a set of ML algorithms can be defined with different parameters (learning rate, tree depth, neural network architecture, regularisation terms, loss functions and more) and a list of variables to take as input. The ML models available to build and train are: Boosted Decision Trees (BDT) made with the XGBoost library [183], MLPs and Variational AutoEncoders (VAE) made with both the Tensorflow [184] and DeepMind Sonnet [185] libraries. The encoder part of the trained VAEs can be detached and plugged onto MLPs such that they can benefit from their smaller and (hopefully) disentangled output space. A dataset, containing inputs, labels and weights, is created and shuffled per region. The dataset is split into training and validation datasets either by directly providing the desired fraction of events dedicated to validation, or by specifying a number of folds. If a number of folds is specified, a k -fold cross-validation procedure is followed based on the MC event number, where a separate ML model is trained per fold. The weights are transformed based on the sum of weights of each class to balance the training. The prepared datasets can then be cached locally to make subsequent runs much faster by directly skipping to the training stage. For the MLPs and VAEs, the dataset bounds are computed and added to the models as constant parameters that transform linearly the input data values to be in the range $[-1, 1]$. Several plots describing the dataset, training evolution and model performance are output. The models are saved as Open Neural Network eXchange (ONNX) files, an open source format [186] that makes the export of the trained ML models considerably easier. The MultiVariate Analysis (MVA) studies have shown that the performances of BDTs and MLPs are very similar. The over-fitting was observed to be generally easier to control for the MLPs, via the usage of *dropout*

layers [187]. The robustness of the BDTs response against the change in input variables makes their usage more natural to set a baseline, and their simplicity avoids ambiguities during hyperparameter optimisation. For these reasons, the BDT was the model chosen to perform the separation between signal and background in this analysis, leaving the potential for improvement with more sophisticated methods to future analyses.

6.7.1 Boosted Decision Trees

BDTs can be described as an *ensemble learning* method, in which multiple binary decision trees are combined sequentially to progressively improve the predictive performance. Each tree is trained to correct the errors of the previous ones, accumulating into a strong classifier composed of several weak learners. The BDTs built from the XGBoost library use the gradient boosting method. The trees are added iteratively to minimise a differentiable objective function, which can be written as the sum of two terms:

$$\mathcal{L} = \sum_i \ell(y_i, F(\vec{x}_i)) + \sum_m^M \Omega(f_m), \quad (6.11)$$

where ℓ is a loss function that takes as input the true labels y_i and the output of the BDT model $F(\vec{x}_i)$. The second term is for *regularisation*: a function Ω acts on each tree f_k and contributes to limit overtraining during the iterative optimisation. The overall BDT model is given by the sum of the trees:

$$F(\vec{x}_i) = \sum_m^M f_m(\vec{x}_i), \quad (6.12)$$

where $\vec{x}_i$ are the input vectors (in feature space) fed to the M trees. The trees f_m are built with a *splitting* procedure, and analytically computed weights w_j are attributed to each output leaf j . The function Ω , that composes the regularisation term in equation 6.11, describes the complexity of the trees f , and is mathematically defined as follows:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_j^T w_j^2 + \alpha \sum_j |w_j|, \quad (6.13)$$

where γ , λ and α are regularisation hyperparameters, and T is the number of leaves in the tree f . The first term penalises trees with too many leaves, the second term limits the weights which reduces the presence of local large corrections, and the last term promotes sparsity in tree leaves by suppressing weak weights. The latter term with

coefficient α is ignored in the rest of this section to keep the derivations simple without affecting the overall methodology. The gradient boosting method can be thought of as a gradient descent performed in function space: the architecture of the overall tree ensemble becomes more complicated with the addition of each tree, rather than simply updating parameters of the function. The impact of adding a tree at a step m is scaled by the learning rate η such that $F_m = F_{m-1} + \eta f_m$, analogous to the step size in gradient descent, which controls the magnitude of each update to the model. The gradient and hessian of the loss function can be defined for each data point:

$$\begin{aligned} g_i &= \frac{\partial \ell}{\partial F(\vec{x}_i)}, \\ h_i &= \frac{\partial^2 \ell}{\partial F(\vec{x}_i)^2}. \end{aligned} \tag{6.14}$$

The loss function may then be simplified for the tree addition step by using a second-order Taylor expansion with respect to the full BDT function $F(\vec{x}_i)$. This approximation transforms the optimisation problem into a quadratic form in f_m , enabling the analytical determination of the leaf weights and a closed-form expression to be used in the leaf splitting procedure. The expansion can be expressed in terms of g_i and h_i for the m -th step:

$$\ell(y_i, F_{m-1}(\vec{x}_i) + f_m(\vec{x}_i)) \approx \ell(y_i, F_{m-1}(\vec{x}_i)) + g_i f_m(\vec{x}_i) + \frac{1}{2} h_i f_m(\vec{x}_i)^2. \tag{6.15}$$

Defining I_j as the set of data point indices i assigned to the j -th leaf, and adding the regularisation term detailed in equation 6.13 (setting $\alpha = 0$ for simplification as mentioned earlier), one obtains the following approximation for the objective function:

$$\begin{aligned} \mathcal{L} &\approx \sum_j^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T, \\ &= \sum_j^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T, \end{aligned} \tag{6.16}$$

where we defined G_j and H_j using I_j , and replaced trees with their effective contribution which is the weight of their leaves w_j . The weights are independent with respect to each other and the quadratic form they obey is easy to manipulate. Taking the derivative of

this expression with respect to w_j and setting it to zero yields the optimal weight w_j^* :

$$w_j^* = -\frac{G_j}{H_j + \lambda}, \quad (6.17)$$

and plugging this expression back into equation 6.16 yields the optimal objective function value:

$$\mathcal{L}^* = -\frac{1}{2} \sum_j^T \frac{G_j^2}{H_j + \lambda} + \gamma T. \quad (6.18)$$

The analytical expression for w_j^* is used to compute the leaf weights and $\mathcal{L}^*$ measures the optimality of a given tree structure given the data. In particular, $\mathcal{L}^*$ provides a score that is used in a greedy optimisation to obtain the best possible tree. The splitting procedure used to establish the tree structure relies on the gain $\Delta_{LR}\mathcal{L}^*$, which is computed and optimised one level at a time. Specifically, $\Delta_{LR}\mathcal{L}^*$ describes how much the objective function score is improved by splitting a leaf into two: left (L) and right (R). The gain can be written mathematically as :

$$\Delta_{LR}\mathcal{L}^* = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma, \quad (6.19)$$

where the third term represents the score before the splitting occurs. Using this framework, the splitting procedure is then carried out for each feature by sorting the data and scanning for an optimal split at each node until the maximum tree-depth is reached, or until no more gain is achieved. Finally, after the learned tree ensemble is built, the inference is performed by summing the contribution of all trees as described by the function $F(\vec{x}_i)$ in equation 6.12. In our case, the software is designed to potentially discriminate between more than two classes, which is not directly compatible with binary trees. This is handled by producing an ensemble of trees per class instead, with the final scores passed through the softmax function to obtain the final class probabilities as follows:

$$p_k = \frac{e^{z_k}}{\sum_i^K e^{z_j}}, \quad (6.20)$$

where z_k represents the logit output for the tree $F_k(\vec{x}_i)$ associated with the class k , and the denominator is normalising by summing over the outputs of all K classes.

6.7.2 Model Inputs

Several observables are used to train the BDTs, and the importance of their contribution is assessed using the gain, defined in equation 6.19, which is calculated during training. Specifically, the mean gain per split is the metric taken as the importance of a given observable, and can be intuitively understood as evaluating how much separation power is obtained from splitting on that observable on average. Figure 6.12 shows a bar plot displaying the gain-based importance for each input of a BDT trained in a fold of the 3-jets b -veto region (events requiring exactly three jets and zero b -jets). The ordering of the importance is very consistent and almost identical between folds for most inputs. The ordering of the inputs with the lowest importance values varies considerably, suggesting a high statistical fluctuation arising from the limited usage of these variables in the trees.

Figure 6.12 also shows the correlation matrix of the inputs, which provides complementary information to evaluate the importance of features. Some observables may be correlated and thus contain shared information, which results in redundancy in the overall training dataset. As a result, some observables might seem to have powerful separability based on their distributions, but will have a limited effect because of their correlation. It is also important to note that the correlation matrix is computed via Pearson coefficients, which measure linear correlations. Thus, there could be more complicated correlations that would not be directly visible from the correlation matrix. From the correlation matrix, it is clear that several inputs are correlated to different degree. For example, the features whose names start with “boosted” can be seen to be generally correlated between themselves and other observables. It is therefore not surprising to see that these features tend to have a low importance ranking. High correlations and inconsistent importance ranking between folds are generally not desirable, but, as can be seen from section 6.7.3, there is no visible overfitting during the training. The regularisation hyperparameters constrain the overtraining well, this can be seen in the performance, which is identical between different folds for both the training and validation datasets.

As can be seen in Figure 6.12, there are 29 different inputs, which is quite substantial. The inputs are grouped and described into categories to simplify the discussion. The observable distributions will be shown as “shape histograms”, meaning that the histogram for each process is normalised to focus on the shape differences. The first category regroups the four inputs computed using only jet kinematics, shown in Figure 6.13. The p_T distribution of j_1 and j_2 are very similar between ggF and W +jets, where both are strictly falling and tend to have lower values than the other processes. The VBF-produced

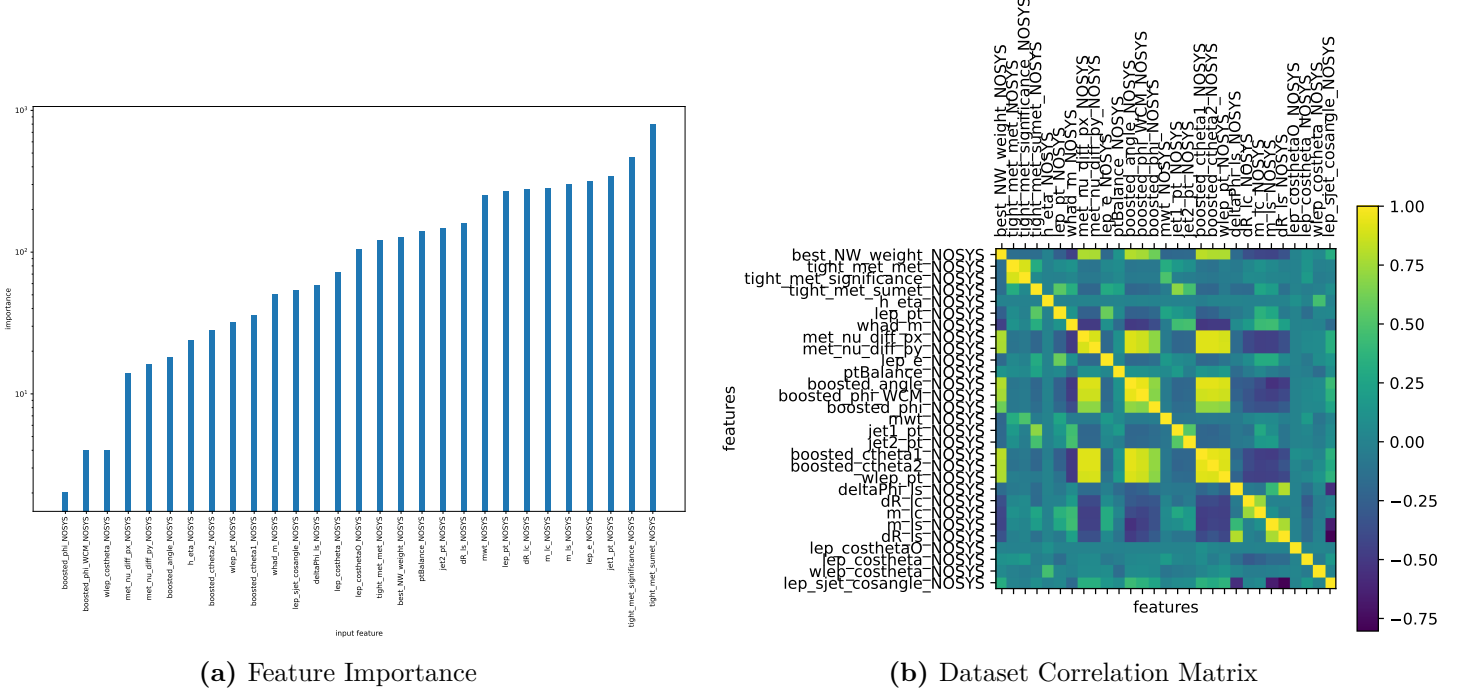


Figure 6.12: Feature importance (a) computed using the average gain from splitting with a given observable, and correlation matrix (b) computed from the signal training dataset (ggF + VBF).

signal has a broader distribution that rises up to roughly 40 GeV before exponentially falling, and thus looks quite different from the W +jets background. However, this distribution is very similar to the top quark processes. It is not surprising to see W +jets producing the softest jets since they originate from QCD radiation and not from W boson decays. The p_T balance observable is defined as:

$$B_{p_T} = \frac{|\vec{p}_{Tj_1} + \vec{p}_{Tj_2}|}{|\vec{p}_{Tj_1}| + |\vec{p}_{Tj_2}|}, \quad (6.21)$$

where the numerator may be identified as the transverse momentum of the reconstructed hadronic W . Intuitively, p_T balance measures the angle between the momentum of the two jets in the transverse plane. When both jets originate from the same resonant particle decay, the jets are expected to be back-to-back in the rest-frame of the resonant particle (the W boson in this case), and thus to have some angular separation in the lab frame after boosting. Therefore, low p_T balance values indicate a resonance, and high values indicate collinearity atypical of resonant decays. In the plot, it can be seen that the ggF-produced signal events differ slightly from all other physical processes by being more likely to have jet systems with a low p_T balance. The invariant mass of the jet

system, $m_{j_1 j_2}$, is quite interesting. The V +jets processes are very similar and peak at low mass, around 50 GeV. For the ggF signal, the distribution has a higher and broader peak, between 50 and 80 GeV, and falls faster than for V +jets at higher mass. The broadness of the peak is caused by the hadronic W being off-shell for half of the signal events. The diboson, top quark and VBF processes show the typical W boson peak at 80 GeV.

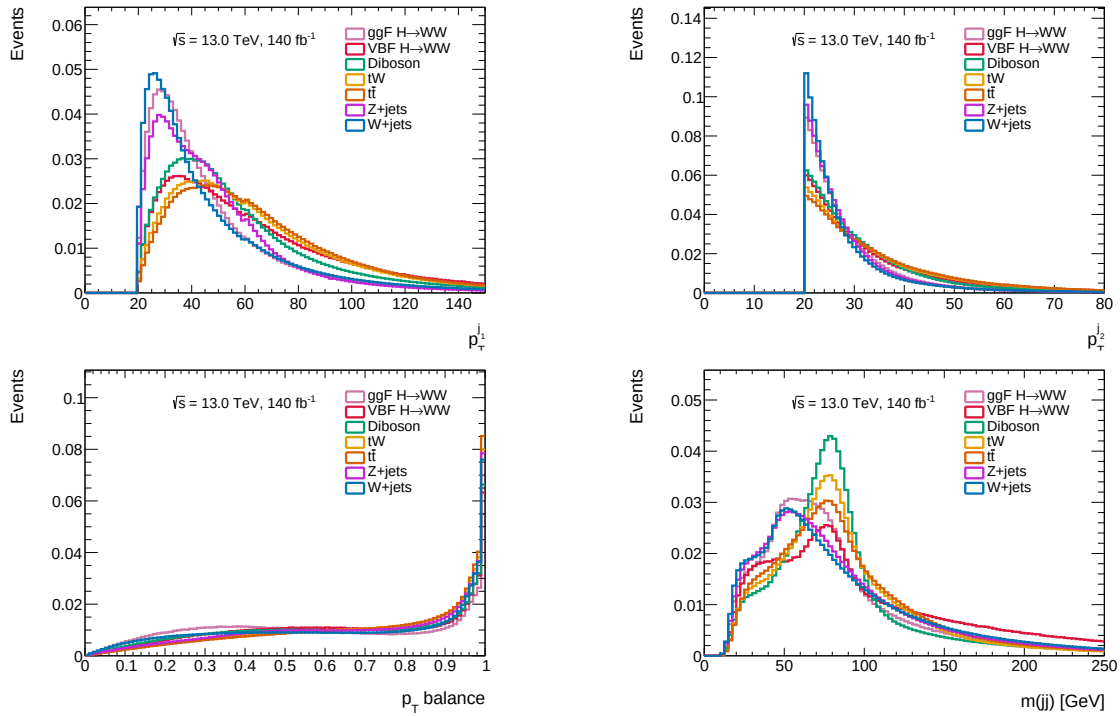


Figure 6.13: Shape plots of the BDT inputs computed using the kinematics of j_1 and j_2 only.

The next group of features is composed of observables produced by combining or comparing the lepton and one of the jets, j_1 or j_2 . Figure 6.14 shows the shape histograms of the six observables used in the training. The distribution of the angular distances between the lepton and each jet in the η - ϕ plane, $\Delta R_{\ell j_1}$ and $\Delta R_{\ell j_2}$, are displayed in the two plots at the top of the Figure. For both observables, the ggF and VBF produced signals peak at very low values, close to the minimum of 0.4, and at a value around 3.1. Every background process behaves differently to the signal, and instead have a rising distribution from 0.4 to 3.1, after which every process decreases exponentially. The VBF Higgs events are significantly more distributed at that first peak than the ggF events, making the observable more powerful at isolating the former than the latter.

The histograms start at 0.4 due to the overlap removal procedure that occurs between the $R = 0.4$ jets and the lepton physics objects. Lower angular separation between the targeted jets and the lepton is expected for the signal processes, the Higgs has a non-zero momentum in the lab frame, which orients the W bosons and their decays towards the same direction. In the background processes, there is no resonant decay that connects the lepton and the jets, and the physical objects tend to be back-to-back from the recoil of the hard scatter. The lepton and jets being emitted in the opposite direction explains the peak at around 3.1. The next two plots, in the middle row, show the distribution of the invariant mass of the lepton with each jet, $m_{\ell j_1}$ and $m_{\ell j_2}$. The signal processes both peak at low mass values, at around 50 GeV for $m_{\ell j_1}$ and 60 GeV for $m_{\ell j_2}$. The backgrounds have generally broader distributions (except for Z +jets, which exhibits a sharp maximum) that peak at higher mass values, at around 100 GeV for $m_{\ell j_1}$ and 70 GeV for $m_{\ell j_2}$. The Z +jets background peak stands out from the other physical processes in the distribution of $m_{\ell j_1}$, due to its narrowness, and a high mass centring at around 110 GeV. This peak is present only for the electron channel, which strongly suggests that it is caused by detector effects. The Z +jets is expected to display extreme effects that differ between electron and muons due to the preselection requirements, specifically, one of the two leptons resulting from the Z decay must be discarded or not reconstructed by the detector. The electron and muon physics objects reconstructions make use of different algorithms and parts of the detector, and will thus behave very differently from one another. A hypothesis that explains this high mass peak could be the overlapping of the “missing” electron with a jet. The (ℓ, j_1) system would then reconstruct the Z boson with a mass $m_Z \sim 91 \text{ GeV}$, offset by the additional energy of the jet. This interesting feature needs to be studied further at particle- and truth-level. The two plots on the bottom row show the distribution of angles defined between the lepton and the j_2 jet. The plot on the left shows the distribution of the cosine of the angle between the momentum directions of the lepton and the jet. It can be defined as follows:

$$\cos(\Theta_{\ell j_2}) = \frac{\vec{p}_\ell \cdot \vec{p}_{j_2}}{|\vec{p}_\ell| |\vec{p}_{j_2}|}. \quad (6.22)$$

The ggF and VBF signal events are both significantly more likely to have the lepton and jet aligned towards the same direction, whereas the distribution is flatter for the background processes, with smaller peaks of equal heights at -1 and 1. The plot on the right shows the absolute angular difference between the lepton and j_2 in ϕ , $\Delta\phi_{\ell j_2}$. The distributions behave very similarly to $\Delta R_{\ell j_2}$, which is of course expected from the

observable direct relation: $\Delta R_{\ell j_2} = \sqrt{\Delta\phi_{\ell j_2}^2 + \Delta\eta_{\ell j_2}^2}$. The distribution shows a strong discrimination power for the signal processes isolation from the backgrounds, as was noted in the phenomenology study described in Chapter 5.

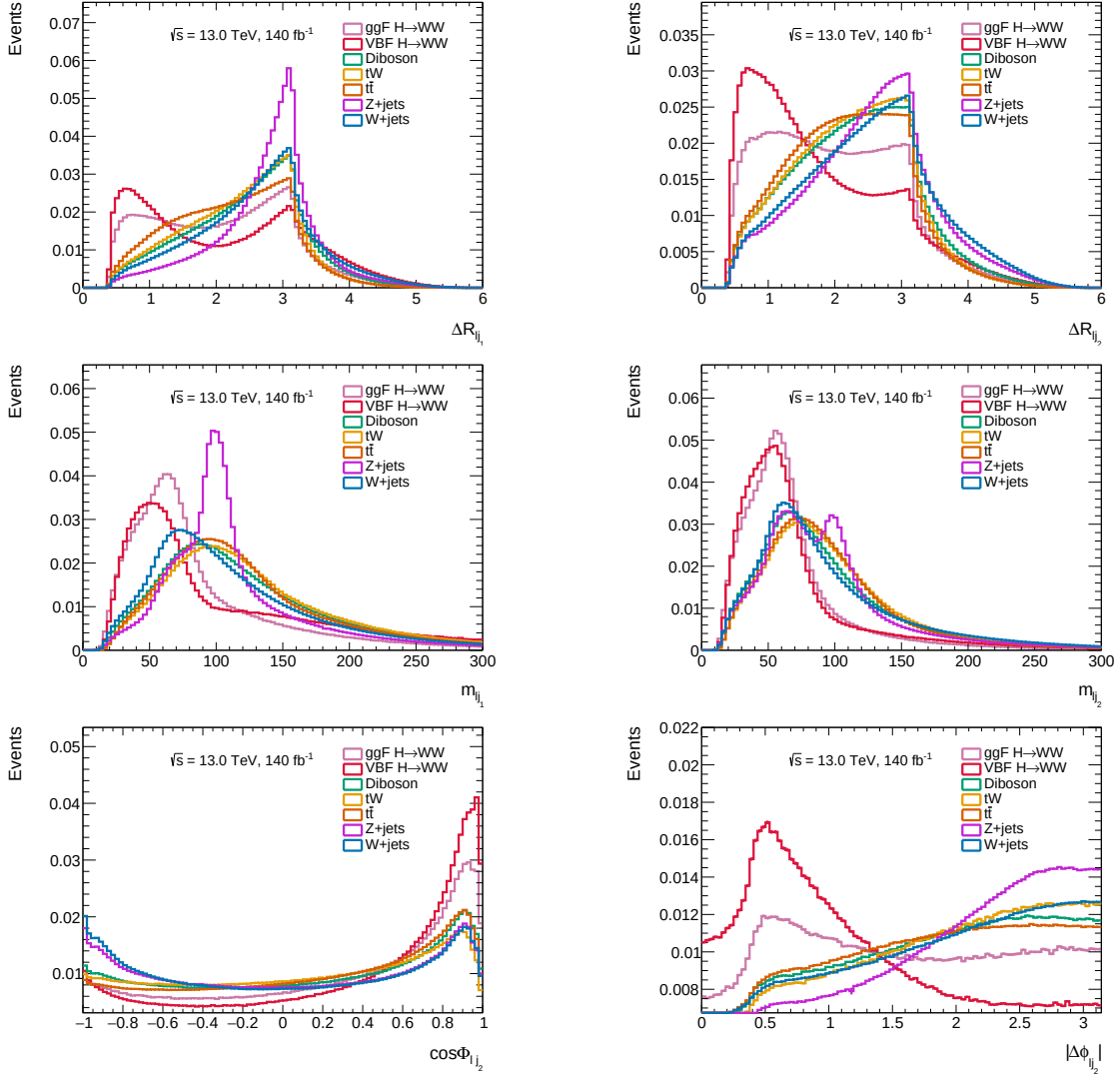


Figure 6.14: Shape plots of the BDT inputs computed using the kinematics of j_1 or j_2 , combined with the lepton.

The last category of plots discussed concerns observables relying on the missing transverse energy, which tend to be the best performing features. Figure 6.15 shows the distributions of the six relevant observables. In the top row, the left plot shows the magnitude of the missing transverse energy, and the right plot shows its significance. The magnitude of $\vec{E}_T^{\text{miss}}$ is very low for Z +jets, which is expected since this process does

not have neutrinos produced in the final state. Both signal processes are very similar to the W +jets and diboson processes, being slightly lower on average. The top quark processes have broader distributions and at higher values, which is expected from the top decays producing more energetic W bosons. The significance of the transverse energy is defined as follows:

$$S_{\vec{E}_T^{\text{miss}}} = \left(\vec{E}_T^{\text{miss}} \right)^T V_m^{-1} \left(\vec{E}_T^{\text{miss}} \right), \quad (6.23)$$

where V_m is the covariance matrix associated with the resolution of every physical object included in the computation of the missing transverse energy. This distance measures how far the missing transverse energy lies from zero in units of the expected $\vec{E}_T^{\text{miss}}$ resolution. Intuitively, this can be interpreted as testing how likely it is for the observed missing transverse energy to be caused by detector resolution and statistical fluctuation (low significance), or by genuine missing physical momentum (high significance). Based on this definition, it is expected to see the Z +jets process remain quite low. Interestingly, both signal processes peak at a lower significance than the diboson and W +jets process. This is one of the effects of the neutrino being softer due to the leptonic W boson being off-shell in the Higgs decay. In the second row, the plot on the left shows the distribution of the sum of the magnitude of all transverse momenta used in the computation of the missing transverse energy. This observable is correlated with QCD activity and multiplicity. The V +jets processes look very similar with respect to that observable, peaking at low values, around 150 GeV. The ggF-produced signal is similar to diboson, peaking at slightly higher values around 175 GeV, and both have broader distributions with larger tails. The VBF-produced signal is similar to the top quark processes, with even broader distributions at significantly higher values, around 250 GeV. This is expected because this observable is affected by the shape of the event: the VBF process has two additional hard jets, originating from the final state quarks that scattered weakly to produce the Higgs. Similarly, the top quark process will have additional hard b -jets from the top quark decays. The right plot on the middle row shows the distribution of the transverse mass of the leptonic W , which is defined in equation 6.6. This observable is correlated with the alignment of the lepton and the missing transverse energy. The Z +jets process looks very different from other processes, with a distribution simply decaying from 0 (the decay rate changes around 50 GeV). This is expected since it is the only process without real W boson. The other processes all share the characteristic peak of the W boson mass around 80 GeV, with different broadness and tail sizes. The bottom left plot shows the distribution of the NW weight for events that can be reconstructed

with the NW algorithm. This requirement is very selective and represents a small portion of phase space, as described in more detail in chapter 4. It can be seen that the shape of the NW distribution itself has discrimination power, with the signal processes tending to be close to 1.0, whereas the background processes are more flatly distributed. The NW distribution is a bit saturated, with many events being exactly 1.0 and most events above 0.0. This can be optimised by changing the coefficient used in the NW weight definition in equation 4.4. The events at the 0.0 bin represent the rare cases where the NW optimisation cannot find an optimal weight because the NW surface is very flat. Finally, the last plot on the bottom right shows the distribution of the angle between the lepton and the jet j_2 in the Higgs rest frame. The distribution is shown only for events where the NW reconstruction finds a solution, since otherwise the Higgs itself is not fully reconstructed. It is clear that the distributions are quite flat for most processes, except for the Z +jets events, which peak at 1.0, and are fully concentrated above 0. The signal is more likely to be at -1.0, which is expected since the jet and lepton should be mostly back-to-back in the Higgs rest frame. Every other background process is quite flat with respect to this angle, showing they are not sensitive to the Higgs rest frames, which is expected. The other angles related to the boost in the Higgs or W boson rest frames are not shown since they have a similarly low discriminating power and do not contribute much to the models.

6.7.3 BDT Training and Performance

The k -fold method is used to produce multiple models such that they do not infer a score based on events they were trained on, avoiding potential bias. Five folds are used per region which results in 80% of the MC events being used for training and 20% for validation. In total, 15 BDTs are trained (5 folds $\times$ 3 jet-multiplicity regions) per scenario (full and simplified models).

A simple BDT hyperparameter optimiser, based on the Optuna framework [188], was added to the `MLTrainer` software and used to optimise the BDT training. The optimisation is performed over three folds for cross validation, and the metrics studied are averaged over all folds to find the optimum. The objective function to maximise is defined as the ROC curve AUC minus the overtraining, where the latter is defined by taking the loss difference in the last boosting round between the training and validation datasets. This is done to maximise the separation power of the BDT while keeping the over-fitting reasonably low. Initially, the regularization parameters are kept to 0 (except

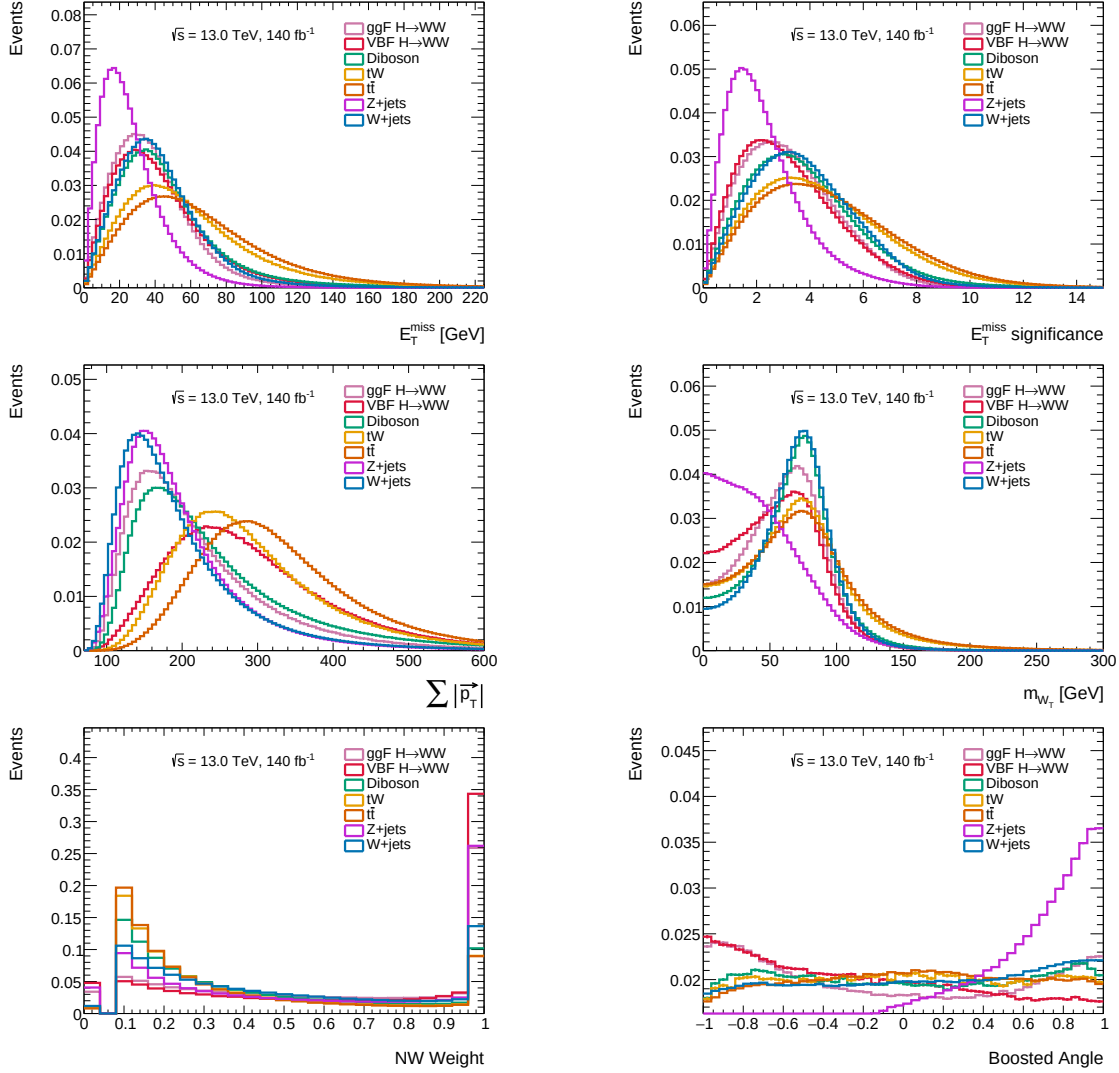


Figure 6.15: Shape plots of the BDT inputs computed using the missing transverse energy, or related physical quantities.

for the default setting $\lambda := 1.0$) and the lowest tree depth and number of boosting rounds M are found for the best ROC curve AUC (Area Under the Curve) achievable. These two parameters are the main drivers of the discrimination performance, but also have a consequent effect on the overtraining, which is why they are determined first independently of the other hyperparameters. Then in a second step, these two parameters are kept constant while η , λ , α , γ are free to float, mainly concerned with minimising the overtraining while maintaining the achieved discrimination power.

The loss function used is the *multiclass softmax cross-entropy* (also called multinomial negative log-likelihood) which is defined as:

$$\ell = - \sum_k^K y_k \log(p_k) = -\log(p_{\text{true}}), \quad (6.24)$$

where we simplified the sum over K classes to a single term due to the true labels y_k being one-hot encoded, the probabilities p_k are the softmax normalised logits for each class, and p_{true} correspond to the probability attributed to the true class for the events being discriminated. Figure 6.16 shows the loss against the boosting round for a fold per region, for both the full and simple models. The difference between the validation and training datasets is very small thus indicating that the overtraining is negligible. The curve flattens towards the last boosting rounds, which suggests that the BDTs will not learn much more information by increasing the number of trees in each ensemble. Figure 6.17 shows the ROC curves for the same BDTs whose learning curves are shown in Figure 6.16. Two curves are displayed, one per class discriminated. For our case, with only two classes, the curves will look very similar and be perfectly symmetrical since the BDT output distribution for both scores is identical but flipped. For every curve, the AUC and Matthew's Correlation Coefficient (MCC) [189] are shown. The latter is an interesting metric derived from the confusion matrix that accounts simultaneously for true and false positives and negatives (1 indicates perfect classification and 0 is equivalent to random guessing). The curves and metrics indicate that the full models have a significantly better performance than the simple models, which is expected given the restricted inputs and region cuts applied for the latter BDTs. It is also clear that as the number of jets allowed in the region increases, the discrimination power of the BDT improves. It would be interesting to check if that behaviour continues with five jets events, even though such a region would have less statistics it is possible that the potential enhanced separability makes a difference. Figure 6.18 shows the distribution for the signal probability score attributed by one of the BDTs trained on the 3 jets region, for both the full and simple scenarios. The training and validation curves overlap well, which indicates that there is a limited effect from overtraining on the prediction. This is in good agreement with the overlapping of the learning curves. As expected from the ROC curves, the full model distributions are more separated than for the simple model. The distributions for BDTs trained on other regions share the same characteristics.

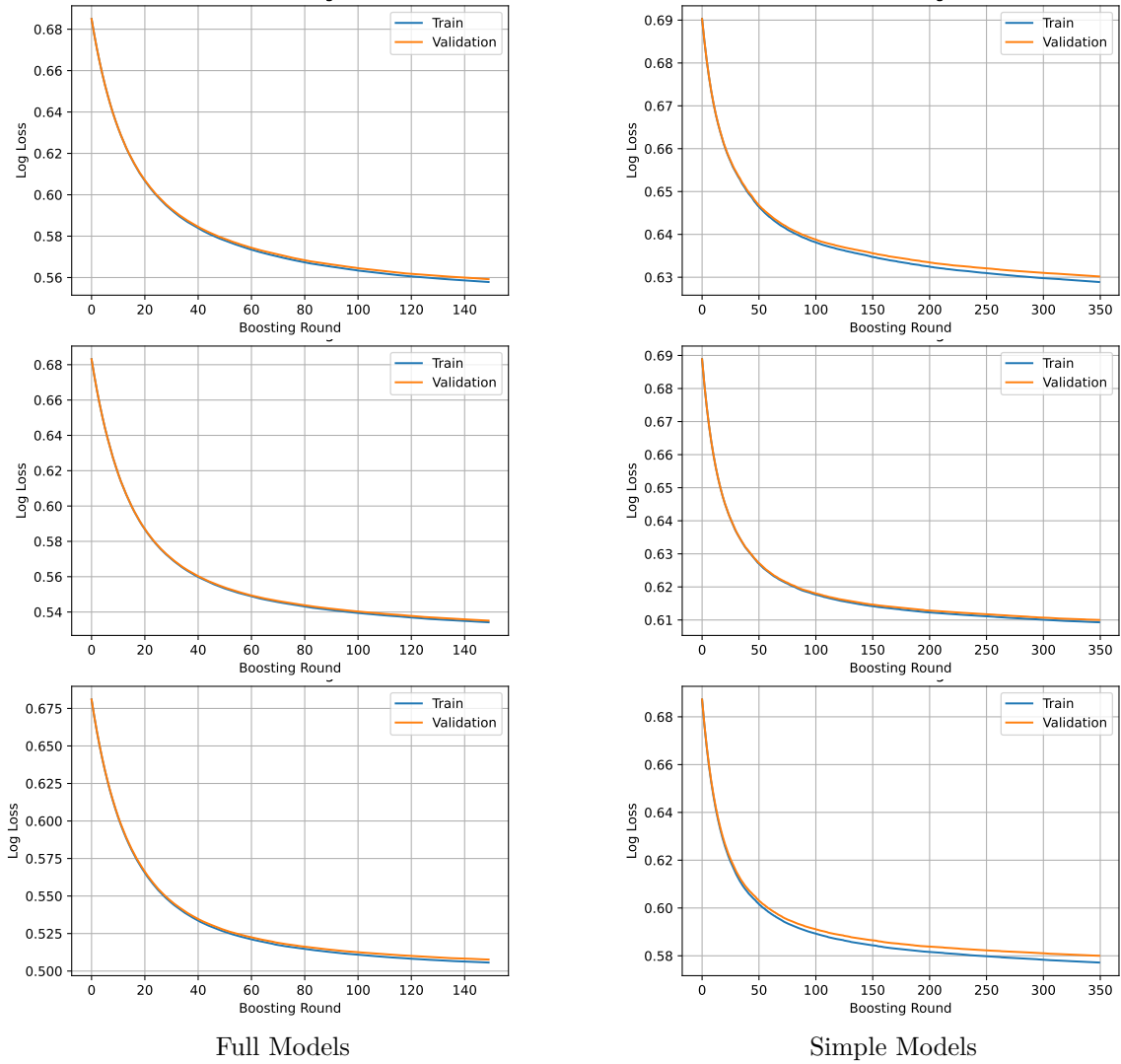


Figure 6.16: Learning curves corresponding to a single training fold in each region. The rows, from top to bottom, show the BDTs trained for the 2-, 3-, and 4-jets regions, respectively.

6.8 Systematic Uncertainties

There are several sources of systematic uncertainty that arise from the experimental methods used and the finite precision of the prediction of theoretical models. This encourages the classification of the systematics into two main categories: the detector uncertainties, which are associated with the reconstruction and calibration of physics objects, and the theoretical uncertainties associated with the modelling of the physical processes. Systematic uncertainties are evaluated by propagating dedicated *variations* through the full analysis chain. Depending on the uncertainty considered, these vari-

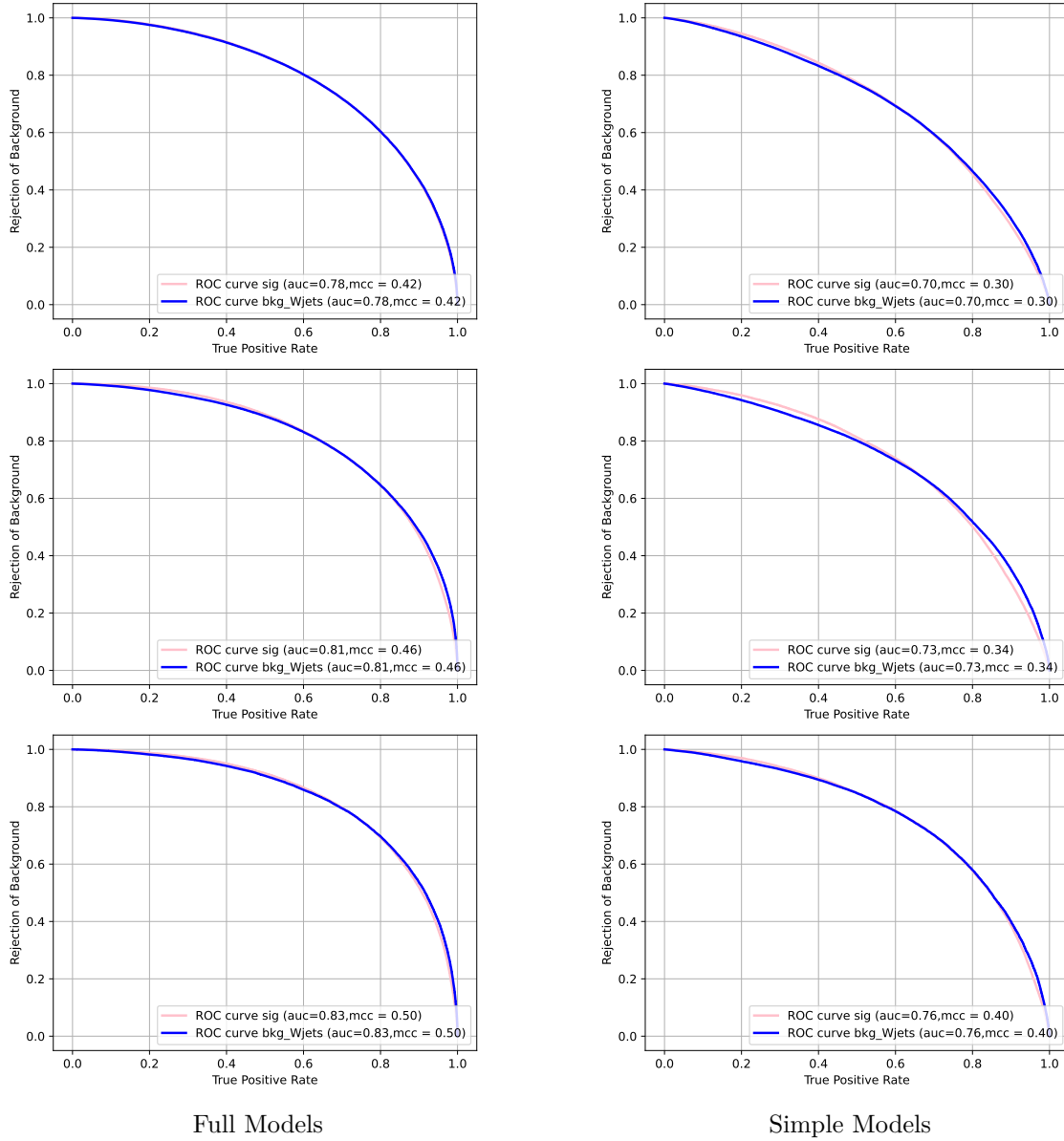


Figure 6.17: ROC curves corresponding to a single training fold in each region. The rows, from top to bottom, show the BDTs trained for the 2-, 3-, and 4-jets regions, respectively.

ations may correspond to modified reconstructed objects, alternative event weights or independent simulated samples.

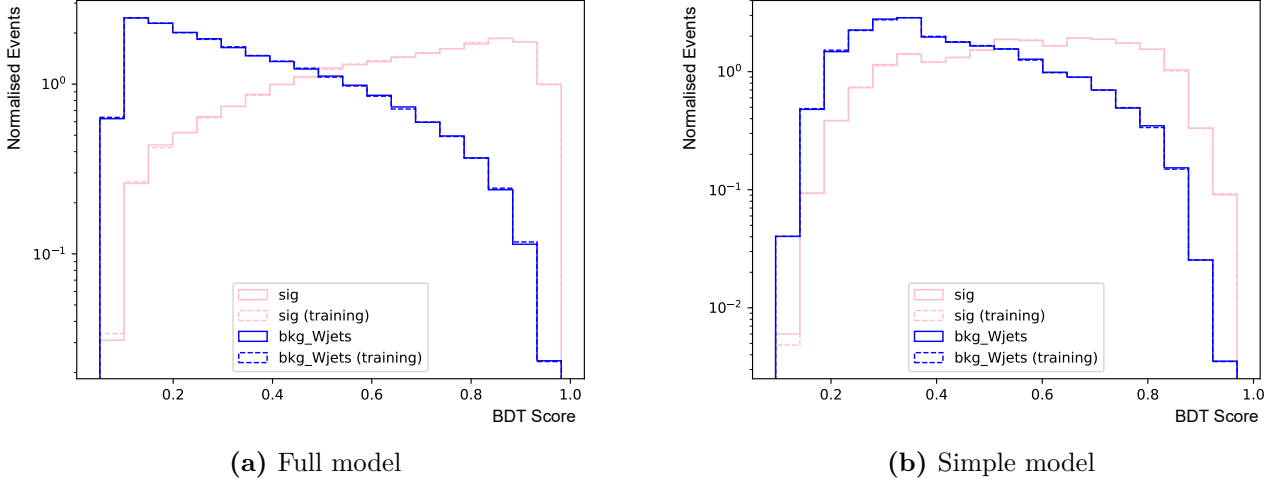


Figure 6.18: Output score distribution for the 3 jets region.

6.8.1 Detector Systematics

Jets

The Jet Energy Scale (JES) calibration [190] is a sequence of corrections, applied to both MC simulation and data, performed to recover the expected jet energy at particle level. The uncertainty associated with the calibrations is parametrised by 35 independent variations that modify the energy of the jets. The Jet Energy Resolution (JER) needs to be understood precisely for measurements involving SM jet production and also affects the missing transverse energy. The JER uncertainty is encapsulated into 15 independent variations, which modify the jet energy smearing in simulation. The flavour tagging uncertainty associated with the calibration of the GN2v01 tagging efficiencies is contained within 133 independent variations that take into account the cross-flavour correlations (i.e correlations between b - and c -tagging). These uncertainties are propagated through variations of the tagger scaling factor correction applied per event.

Electrons and Muons

The electrons and muons both have systematic variations that represent the uncertainty associated with energy and momentum scales and resolutions, which act on their four-momenta directly. Additional uncertainties arise from the reconstruction, identification, isolation, and trigger efficiencies. These are evaluated through variations of the corre-

sponding scale factors applied to simulated events in order to correct the efficiencies to those measured in data.

Missing transverse energy

The missing transverse energy is composed of hard terms and the soft term, as shown in equation 2.9. The systematic uncertainty associated with the hard terms does not have dedicated variations. Instead, their uncertainty arises from the propagation of the uncertainty of the reconstructed physics objects used in their computation. The soft term uncertainty is decomposed into scale and resolution components, where the resolution uncertainty is evaluated separately along directions parallel and perpendicular to the hard recoil axis. There are overall four variations: up and down for the scale, parallel and perpendicular resolution.

6.8.2 Theoretical Systematics

Parton Distribution Function Uncertainty

The PDFs, described in section 2.7, model the momentum distribution of partons within the proton prior to the collision. The uncertainty arising from the imperfect knowledge of the PDFs is evaluated by varying their parameters within their own uncertainty. Dedicated error sets are provided by the PDF fitting groups and used to reweight the simulated events, resulting in multiple variations. A few different prescriptions to handle those error sets, in order to produce the final variations that estimate the systematic uncertainty, exist based on how these sets are produced [112, 191]. The PDF systematics are only considered for the signal (ggF and VBF), and the W +jets process, which is the main background. The former processes are following the *replica* prescription, meaning that a final pair of up and down PDF variations is obtained by combining all replicas into the per-bin offset from the nominal using the root mean squared as follows:

$$\Delta X = \sqrt{\frac{1}{N_{\text{rep}} - 1} \sum_i (X_i - \langle X \rangle)^2}, \quad (6.25)$$

where X is a binned observable, i denotes the indexing of the replicas and N_{rep} is the total number of replicas. The variance ΔX is then added and subtracted to obtain the bin up and down variations respectively. The W +jets process follows the *symmetric hessian*

prescription, where the 100 provided variations can be used directly as independent symmetric variations in the statistical model.

Scale Uncertainty

The scale uncertainty comes from choosing the renormalisation and factorisation scales, μ_R and μ_F , both unphysical and described in more detail in section 2.7. Physical observables should be, in principle, independent of these scales, but, at finite perturbative order, a residual dependence remains. The systematic uncertainty is evaluated by computing additional MC weights following a 7-point prescription. The scales (μ_R, μ_F) are multiplied by the following combination of factors: $(1, 1)$, $(2, 1)$, $(1, 2)$, $(2, 2)$, $(0.5, 1)$, $(1, 0.5)$, $(0.5, 0.5)$, where $(1, 1)$ corresponds to the nominal. The uncertainty is then derived as the envelope of the resulting variations. The scale systematic is computed and included in the statistical model for every physical process.

Parton Shower Uncertainty

The simulated events generated with the Sherpa MC generator include variations of the renormalisation and factorisation scales that also propagate to the parton shower evolution and the matching between the matching between the matrix element and the shower. These variations thus probe both matrix-element and parton-shower modelling uncertainties. The processes involved include the W +jets and Z +jets backgrounds, which are the two main contributing processes in the regions of interest. The nominal $t\bar{t}$ process has the parton shower simulated with PYTHIA8, the associated uncertainty is estimated by using an alternative sample with the same generator for the matrix element (POWHEG-BOX) and showered with HERWIG7 instead.

6.9 Fitting and extracting limits on the cross-section

6.9.1 Statistical Model

The cross-sections are extracted using a simultaneous profile likelihood fit over the Signal Regions (SRs) and Control Regions (CRs). The likelihood function is given by:

$$\mathcal{L}(\vec{n}|\mu_s, \mu_b^j, \vec{s}, \vec{b}^j, \boldsymbol{\theta}) = \prod_i \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i} \times \prod_k f_k(\theta_k), \quad (6.26)$$

where $\vec{n}$ is the vector of events observed in the SRs and CRs, $\vec{s}$ and the set $\{\vec{b}^j\}$, are vectors of the predicted yield for signal and background processes respectively, $\boldsymbol{\theta}$ is the vector of systematic uncertainties, μ_s is the Normalisation Factor (NF) for the signal and μ_b^j a set of NFs for the backgrounds. The vectors $\vec{n}$, $\vec{s}$ and $\{\vec{b}^j\}$ share the same vector space and are indexed with $i \in \{\text{Bins} \otimes \text{Regions}\}$ (written n_i , s_i and $\{b_i^j\}$ respectively). The vector of systematics $\boldsymbol{\theta}$ lives in its own vector space and is indexed with $k \in \{\text{Systematics}\}$. The signal NF μ_s is the Parameter Of Interest (POI) such that $\mu_s = 0$ represents the background only hypothesis and $\mu_s = 1$ the expected signal hypothesis from the SM. There is one NF per background process μ_b^j such that $j \in \{\text{Backgrounds}\}$. Every parameter that is not the POI can be referred to as a Nuisance Parameter (NP). The physics modelling dwells in the Poisson means λ_i , expressed as follows:

$$\lambda_i(\mu_s, \mu_b^j, \vec{s}, \vec{b}, \boldsymbol{\theta}) = \mu_s \cdot s_i(\boldsymbol{\theta}) + \sum_j \mu_b^j \cdot b_i^j(\boldsymbol{\theta}), \quad (6.27)$$

where the effect of systematic uncertainties are included by adjusting the bin contents s_i and b_i^j . The global fit is performed on data to find the NFs and the systematic NPs $\boldsymbol{\theta}$ simultaneously by minimising the negative log likelihood $-2 \ln L$, where L is defined in equation 6.26. The NFs μ_s and μ_b^j are free floating in the fit, whereas the adjustments to the NPs $\boldsymbol{\theta}$ are constrained by the second set of terms in the likelihood, $\prod_k f_k(\theta_k)$. These penalising terms encode auxiliary measurements constraining the systematics and enforcing the level of confidence attributed to them. These terms are labelled as functions f_k because they take a different form depending on the nature of each systematic uncertainty. Most of the systematics are constrained by a Gaussian probability density which can be written as:

$$f(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\theta^2}{2\sigma^2}\right), \quad (6.28)$$

where the width σ characterises the degree of confidence in θ . A single nuisance parameter, called a *gamma*, is defined per bin to constrain the uncertainty originating from the MC statistics by scaling multiplicatively their yield content. The gamma NPs are constrained by Poisson terms given as follows:

$$f(\theta) = \frac{(\theta\tau)^\tau}{\tau!} e^{-\theta\tau}, \quad (6.29)$$

where τ is the effective MC yield, defined as the ratio of the nominal MC yield and the MC statistical uncertainty squared. For weighted events with weights w_n :

$$\tau = \frac{\left(\sum_n^{N_i} w_n\right)^2}{\sum_n^{N_i} w_n^2}.$$

The potential observation of the signal process, and the computation of exclusion limits with respect to its expected cross-section, rely on the formulation of a hypothesis test. The first step towards a hypothesis testing framework is to construct the *test statistic*. Let the profile likelihood ratio be defined as :

$$\tilde{\lambda}(\mu_s) = \begin{cases} \frac{L(\mu_s, \hat{\mu}_b(\mu_s), \hat{\boldsymbol{\theta}}(\mu_s))}{L(\hat{\mu}_s, \hat{\mu}_b, \hat{\boldsymbol{\theta}})}, & \hat{\mu}_s > 0 \\ \frac{L(\mu_s, \hat{\mu}_b(\mu_s), \hat{\boldsymbol{\theta}}(\mu_s))}{L(0, \hat{\mu}_b(0), \hat{\boldsymbol{\theta}}(0))}, & \hat{\mu}_s \leq 0 \end{cases} \quad (6.30)$$

where the notation for the likelihood L was simplified by ignoring bin contents and discarding the background index j . The quantities $\hat{\mu}_s$, $\hat{\mu}_b$ and $\hat{\boldsymbol{\theta}}$ correspond to the best-fit parameters where $\hat{\mu}_s$ is allowed to float (unconditional maximum-likelihood estimate), while $\hat{\mu}_b$ and $\hat{\boldsymbol{\theta}}$ denote the best-fit nuisance parameters with fixed μ_s (profiled conditional likelihood). The piece-wise definition of $\tilde{\lambda}$ accounts for the physical boundary of the NF: $\hat{\mu}_s$ cannot be negative since it would lead to a negative event yield. The test statistic for μ_s is then defined as:

$$q_{\mu_s} = \begin{cases} -2 \ln \tilde{\lambda}(\mu_s), & \hat{\mu}_s \leq \mu_s \\ 0, & \hat{\mu}_s > \mu_s \end{cases} \quad (6.31)$$

where q_{μ_s} is set to 0 when the data is more signal-like than the tested hypothesis (i.e $\hat{\mu}_s > \mu_s$) to ensure that such a fluctuation does not artificially strengthen the exclusion when estimating limits. The p -value, representing the probability of obtaining an equal or greater incompatibility with the tested hypothesis than the one observed in data, can be defined using the probability density function of q_{μ_s} :

$$p_{\mu_s} = \int_{q_{\mu_s}^{\text{obs}}}^{\infty} f(q_{\mu_s} | \mu_s) dq_{\mu_s}, \quad (6.32)$$

where the probability density function depends on the tested μ_s value. Estimating $f(q_{\mu_s} | \mu_s)$ can be done by evaluating several pseudo-experiments generated by fluctuating

the histograms input to the model, which is quite computationally expensive. For that reason, Wald's approximation [192] is used instead:

$$q_{\mu_s} = -2 \ln \tilde{\lambda}(\mu_s) \approx \frac{(\mu_s - \hat{\mu}_s)^2}{\sigma^2}, \quad (6.33)$$

where σ is the asymptotic standard deviation associated with the signal strength $\hat{\mu}_s$, and is extracted from the curvature of the profile likelihood evaluated on the so-called *Asimov dataset*, in which the observed event counts are replaced by the MC prediction. This approximation is only applicable in the asymptotic limit where sufficient statistics are available, and the distribution of $\hat{\mu}_s$ is very close to $\mathcal{N}(\mu_s, \sigma^2)$. The corresponding p -value under the signal-plus-background and background-only hypotheses can then be evaluated analytically using the Gaussian cumulative distribution function Φ :

$$p_{s+b} = 1 - \Phi \left(\sqrt{q_{\mu_s}^{\text{obs}}} \right), \quad (6.34)$$

$$p_b = \Phi \left(\sqrt{q_{\mu_s}^{\text{obs}}} - \frac{\mu_s}{\sigma} \right). \quad (6.35)$$

The frequentist confidence level CL_s can then be evaluated for a given signal strength μ_s :

$$CL_s = \frac{p_{s+b}}{1 - p_b}. \quad (6.36)$$

The 95% confidence limit then corresponds to the value of μ_s satisfying $CL_s = 0.05$.

The **TRExFitter** [193] (Top Related Experiment Fitter) software is used to perform the binned profile likelihood fit and extract limits. The likelihood is optimised via a multi-dimensional minimisation of the negative log likelihood, using the **MIGRAD** algorithm embedded in the **MINUIT** framework. The uncertainties are extracted by using the **MINOS** procedure on the signal NF, and the **HESSE** method on the other parameters of the fit. The latter determines symmetric uncertainties from the inverse of the Hessian matrix evaluated at the likelihood minimum, making the assumption that it is locally parabolic. The former is more robust and computes asymmetric confidence intervals by scanning the profile likelihood instead.

Each systematic uncertainty is represented by a pair of histograms (“up” and “down” variations) derived from modified detector calibrations, reconstruction conditions or theoretical modelling choices within ATLAS. Figure 6.19 shows an example of such a

pair of histograms for two jet systematics. The plotted SRs are introduced in section 6.9.2. These variations encode the expected effect of the uncertainty on the event yields and distributions. In the statistical model, each systematic uncertainty is associated with a nuisance parameter θ , constrained by Gaussian terms described by equation 6.28, which continuously deforms the nominal template (histogram) towards the up or down variation. That deformation, referred to as *template morphing*, consists of a per-bin polynomial interpolation between the variations such that $\theta = -1$ and $\theta = 1$ correspond to the down- and up- variation respectively. Outside the interpolation range $|\theta| > 1$, exponential interpolation is used to preserve positive-definite yields. Smoothing and symmetrisation procedures are performed on the up and down systematic template histograms before constructing the template morphing. The former suppresses the effect of noise from low MC statistics, and the latter reduces potential bias that arises from statistical or non-physical asymmetries between variations. Both transformations are important to improve the stability and robustness of the template morphing. On the left plot of Figure 6.19, the effects of the smoothing procedure can be appreciated from the difference between the dashed and solid line histograms. There are several sources of systematic uncertainty, and letting them all contribute to the statistical model would make the fit unstable and lead to poorly constrained nuisance parameters. In particular, it is expected that some systematics have a negligible effect on the fitted regions and the POI. A pruning procedure is performed to exclude systematics based on thresholds set on normalisation (testing the overall yield change) and shape (testing the Kolmogorov-Smirnov distance [194, 195] to the nominal template), thus reducing model complexity and improving fit stability while having a negligible impact on the final result.

6.9.2 Expected sensitivity

Regions and separability

The expected sensitivity to the POI is estimated by performing a fit on the Asimov dataset. Despite the fact that the analysis workflow integrates Run 3 data and simulation, more work is required to study the systematics of this Asimov dataset before it can be included in the fit. The six SRs used for the statistical model are shown in Figure 6.20. The binned variables are built by taking the logit transform of the signal probability score, obtained by running the BDTs described in section 6.7.1. The variable can be

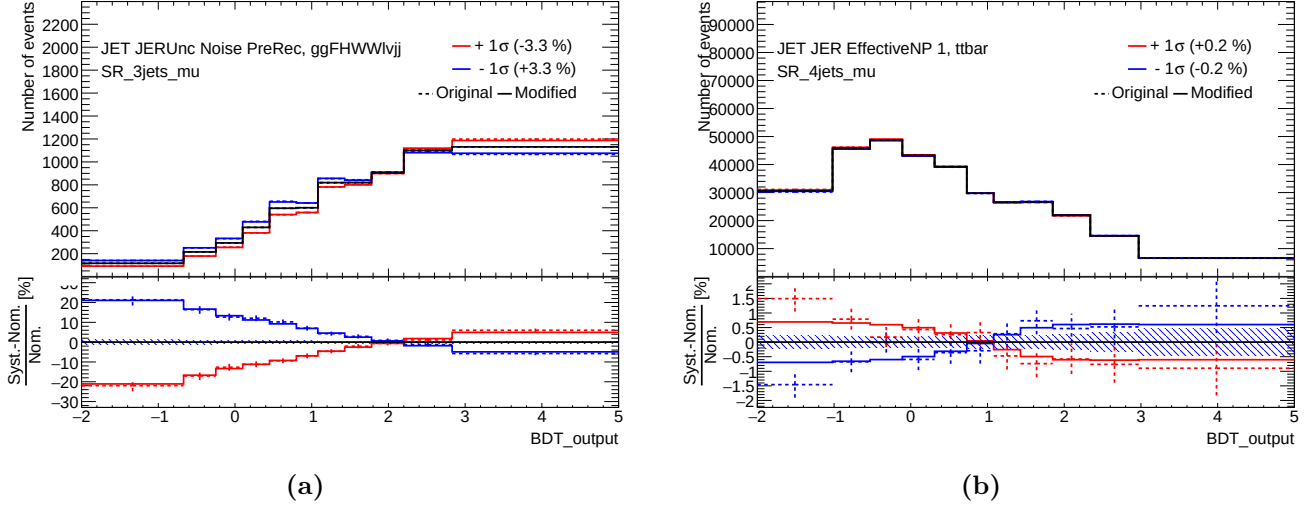


Figure 6.19: Histograms representing the variations from the nominal associated with a systematic. The black histogram is the nominal, the red and blue histograms represent the up and down variations respectively. The dashed lines represent the variation template before the symmetrisation and smoothing procedures.

mathematically expressed as follows:

$$\text{logit}(p_s) = \ln \left(\frac{p_s}{1 - p_s} \right), \quad (6.37)$$

where p_s is the probability associated with the event being a signal event by the relevant BDT. The logit is defined in the interval $(-\infty, +\infty)$, and expands the probability score into a distribution with signal-rich and background-rich tails. This improves the use of shape information in the likelihood fit. The regions are defined by applying the b -veto and then splitting into jet requirements: $N_{\text{jets}} = 2, 3, 4$. The regions are further split into electron and muon channels for two main reasons. Firstly, the electron and muon systematic uncertainties are independent such that the associated nuisance parameters will only affect half of the SRs, which helps in constraining them during fitting. Secondly, as discussed in more detail in section 6.5, the fake lepton background behaves very differently for electrons and muons, and thus splitting into lepton channels helps in constraining the associated NF. From the plots it can be seen that an unusually large amount of statistics is present in the regions, with bin contents of order $\sim 10^6$ to 10^7 events, suggesting a sensitivity to fluctuations of the order of 0.01%. As a consequence, the regions are very sensitive to systematics and will tend to over-constrain those with large template differences, even with the Asimov dataset. The threshold used in the systematic pruning is set to 0.01%, for both shape and normalisation, to ensure that

no relevant uncertainty is discarded. This results in several nuisance parameters in the statistical model, which is thus very complex and requires the most numerically stable fitting strategy available in **TrexFitter**. The histogram binning for the SRs is computed using an algorithm that maximises the signal sensitivity: the histograms start from a 100 bins, with linearly spaced edges, which are progressively merged until the target number of bins is reached.

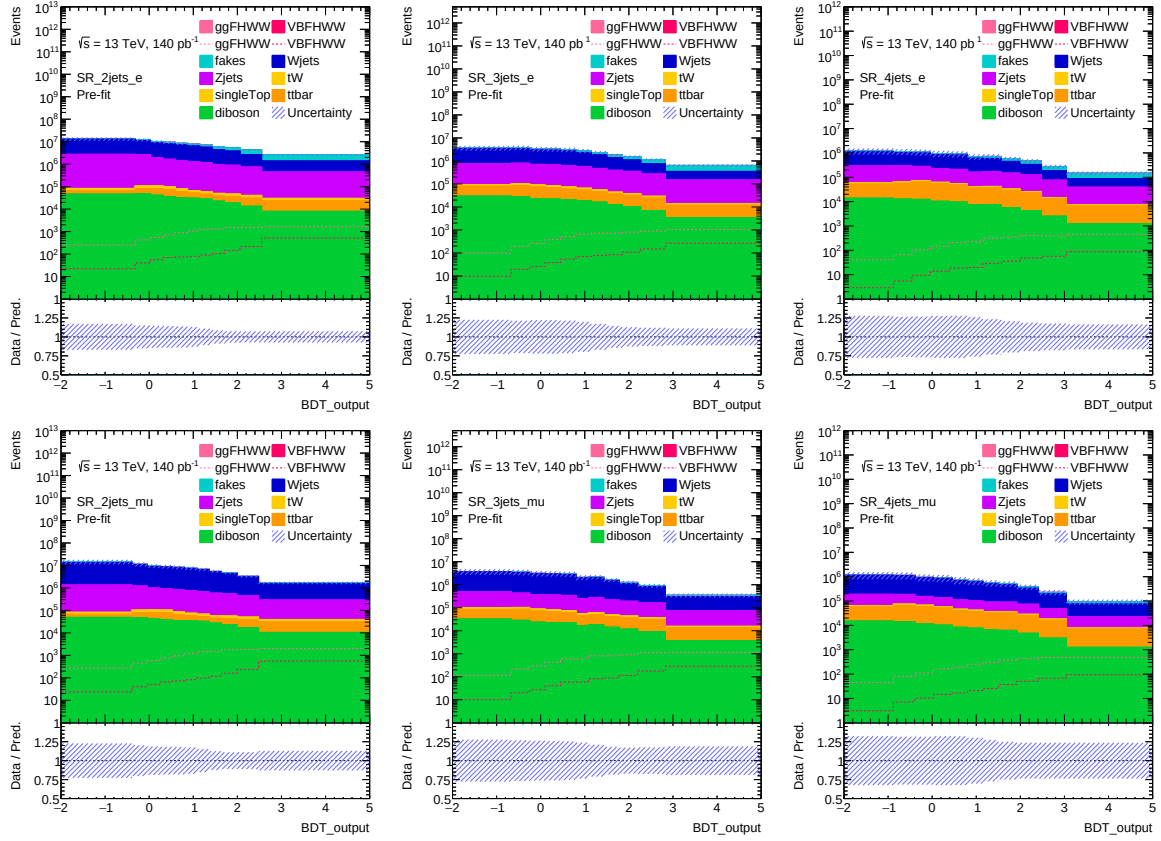


Figure 6.20: The SRs used in the fit. The regions are obtained by applying a b -veto. Each column corresponds to the number of jets allowed in the region (2, 3 and 4), and the electron and muon channels are shown on the top and bottom row respectively.

Figure 6.21 shows the plots of the three CRs, which target the top quark processes by requiring at least one b -jet. As the requirement on the number of jets increases to 4, the concentration in $t\bar{t}$ events increases as expected. The binned variable is the invariant mass of the lepton- j_2 system, $m_{\ell j_2}$. In the future, a dedicated BDT score could be used for these regions to further separate the $t\bar{t}$ process from the other backgrounds. The CRs follow the same binning optimisation as for the SRs.

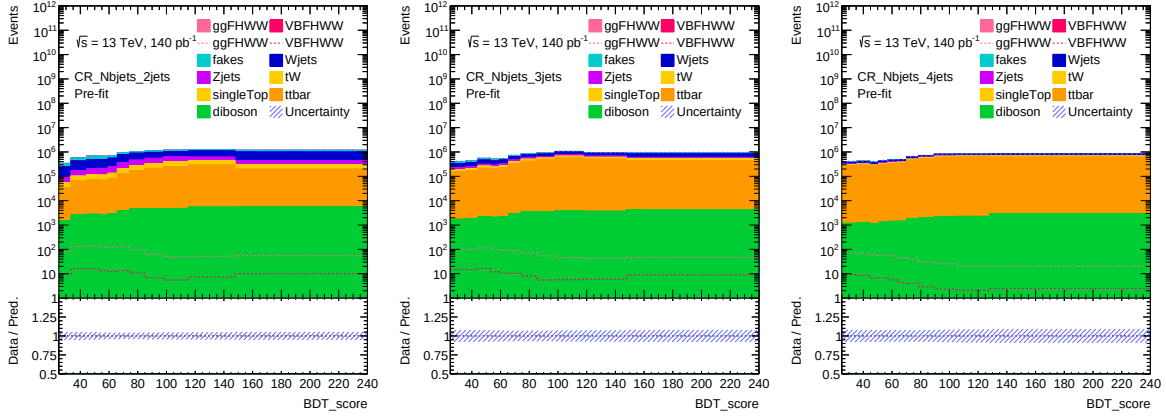


Figure 6.21: The CRs used in the fit. The regions require at least 1 b -jet in the event, and are split per number of jets (2,3 and 4).

The separation of the ggF signal process with respect to the additively combined backgrounds histogram is shown for each SR in Figure 6.22. The separation percentage is evaluated as follows:

$$\text{Separation} = 100 \cdot \frac{1}{2} \sum_i \frac{(s_i - b_i)^2}{s_i + b_i}, \quad (6.38)$$

where the index i iterates over the bins of the normalised signal histogram s_i , and the normalised background histogram b_i . This metric has the convenient property of being bounded between 0 and 100, where it is equal to 0 for identical shapes, and equal to 100 for perfect separation. As expected from the performance of the BDTs, pertinently evaluated through ROC curves in section 6.7.3, the separation increases as the requirement for the number of jets increases. The separation is noticeably better in the muon regions than in the electron regions.

Asimov fit

A NF is attributed to almost every single physical process. A single NF parametrises the normalisation of all the top quark processes (single top + $t\bar{t}$) due to their shape being very similar. Similarly, the POI is a single NF for the ggF and VBF signal processes because the statistical model currently does not hold enough information to separate the production mode effectively. Concerning the fake lepton processes, two NFs are

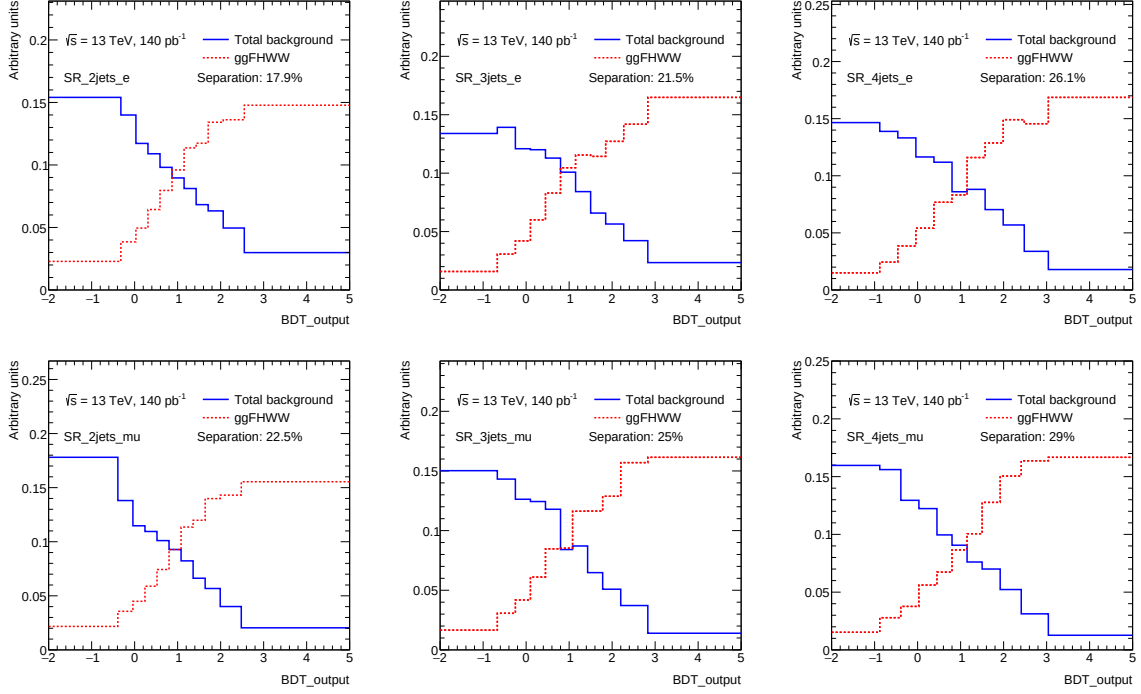


Figure 6.22: The normalised histograms of the ggF signal and the combined backgrounds for each SR.

attributed: one for the SRs and one for the CRs. This is motivated by the fact that the fake efficiencies look very different in b -veto and b -filter regions, as seen in section 6.5.

Figure 6.23 shows the post-fit value of the systematic NP values and their uncertainty. There are a relatively large number of NPs entering the fit: roughly 230 are shown in the pulls, which does not contain the additional NF and gamma parameters. Every NP is found to be 0, which is the exact behaviour expected from an Asimov fit. However, their uncertainties do defy expectations by not consistently spanning a standard deviation. In particular, the jet systematics and a few theoretical NPs, shown at the top of the Figure, are very constrained. This over-constraining that occurs with the Asimov dataset is caused by the unusually large amount of statistics in the regions used by the statistical model. In order to further investigate the strong constraints on the NPs, a simpler fit including only three jet systematics is performed. These jet uncertainties are decorrelated per physical process, thus producing eight NPs per systematic. Figure 6.24 shows the post-fit value and uncertainty of the NPs associated with the decorrelated systematics. It can be seen that the constraining is directly correlated with the expected yield of the physical processes. The NPs associated with the W +jets process are the most constrained,

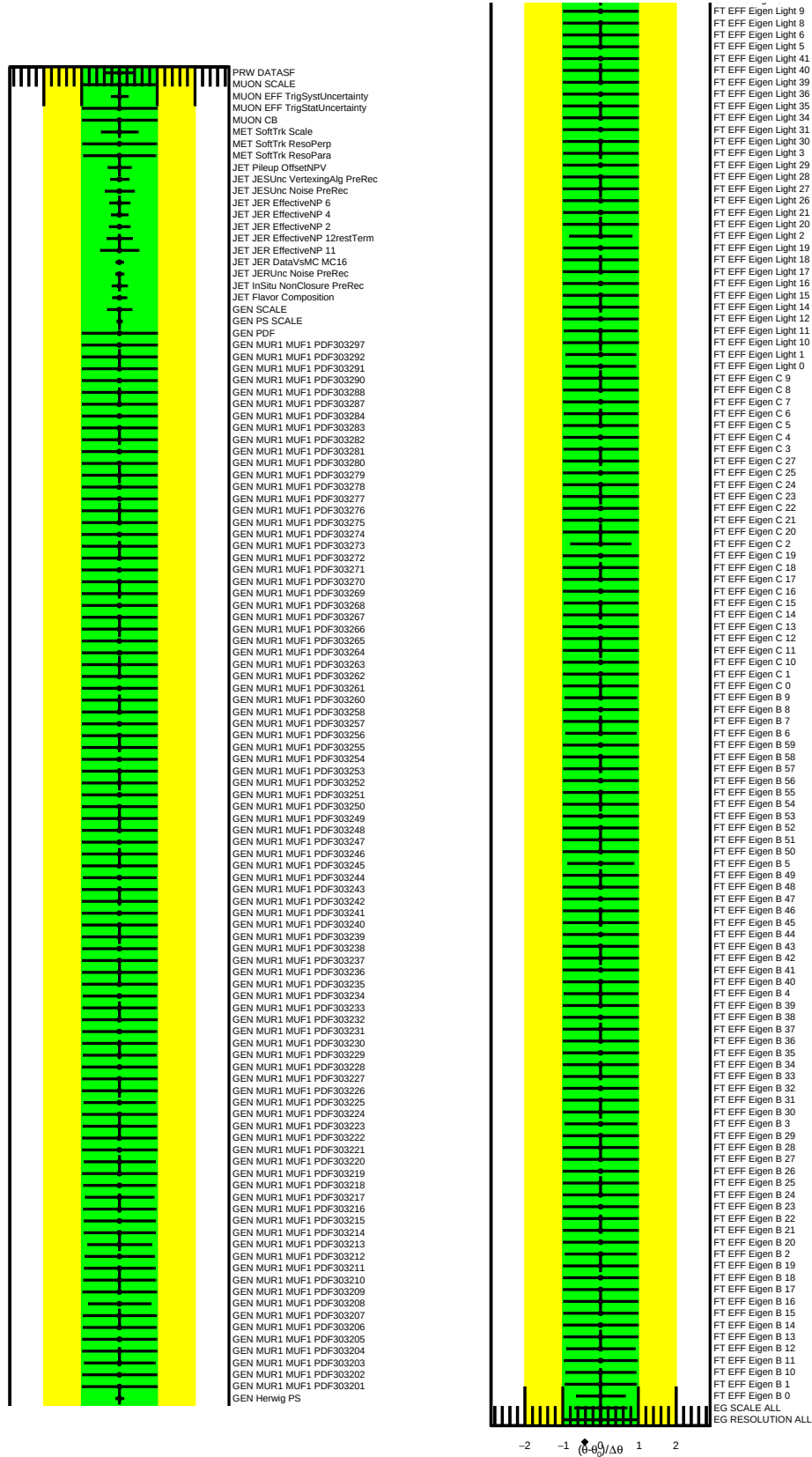


Figure 6.23: Post-fit pulls of the systematic NPs.

followed by Z +jets and $t\bar{t}$. The NPs attached to the signal and diboson processes have the expected uncertainty and thus do not show any signs of over-constraining, which corroborates the high statistics explanation. Discussions about this behaviour are still on-going, and it is expected that some decorrelation strategies may be used to relax this effect if it becomes unphysical.

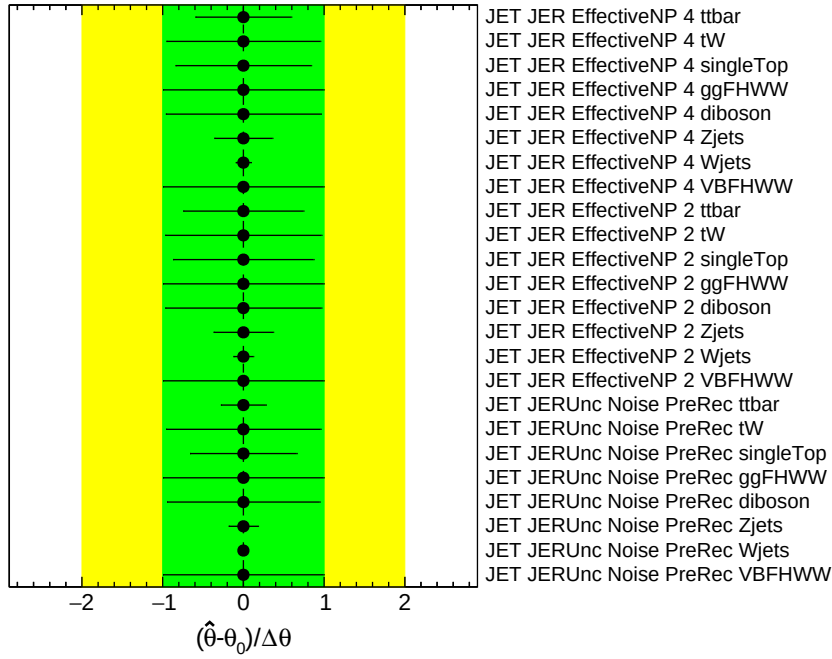


Figure 6.24: Post-fit pulls of a fit with only three jet systematics. The systematics are decorrelated per physical process.

Figure 6.25 shows the log-likelihood scan over the POI. The minimum is located at 0 because the background-only hypothesis is tested when performing the search for the signal. A quadratic polynomial is fitted to the scan points to highlight any potential deviation that would indicate that the Gaussian approximation does not hold. The quadratic is expressed with two parameters as follows:

$$f(\mu_s) = \frac{(\mu_s - \mu)^2}{2\sigma^2}, \quad (6.39)$$

where the parameters are μ (shown as “mean” in the legend) and σ . Two small deviations stand out: first a small bump appears close to 0, and second, a slight asymmetry when comparing the points at the opposite end of the scan, at $\mu_s \pm 1.0$. Both deviations are not concerning, and are to be expected based on the large amount of NPs floating in the fit. Overall, the likelihood scan is not symptomatic and behaves well. The usage of MINOS

on the POI avoids reliance on the local curvature, and provides a better estimation of the error by also taking into account any asymmetry.

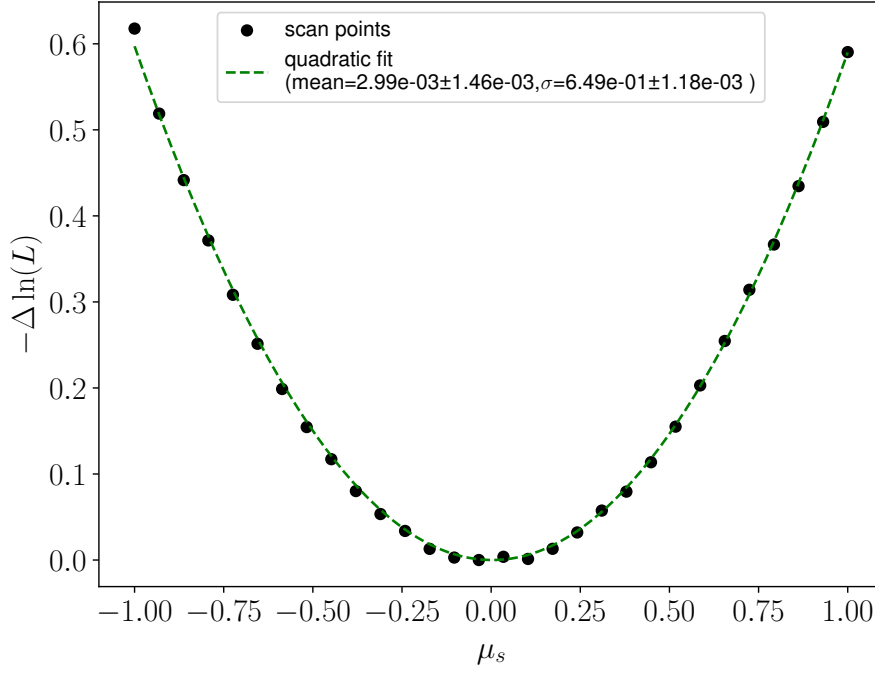


Figure 6.25: Scan of the negative log-likelihood over the POI.

Uncertainty decomposition

The uncertainty is decomposed using the post-fit covariance matrix V_{ij} , which is defined as the inverse of the Hessian matrix at the minimum of the likelihood L :

$$V_{ij} = \left[-\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \Big|_{\vec{\theta}=\hat{\vec{\theta}}} \right]^{-1}. \quad (6.40)$$

The Hessian and covariance are estimated by MINUIT during the minimisation procedure. The uncertainty on the POI generated by a group of NPs g is approximated by summing over the relevant contributions among the off-diagonal terms of the covariance matrix as follows:

$$\sigma_{\mu_s, g} = \sqrt{\sum_{i \in g} \frac{V_{\mu_s i}^2}{V_{\mu_s \mu_s}}}, \quad (6.41)$$

where μ_s is used to indicate the index of the POI. The contribution of groups of NPs to the POI uncertainty is provided in Table 6.2. It is clear that the dominating source of systematic uncertainty comes from the lack of MC statistics in the regions of interest. This source of uncertainty is relatively easy to handle compared to the others, and can be reduced by simply generating more MC events. The total systematic uncertainty drops to 0.413 when ignoring the MC statistics, which is a remarkable and physically accessible improvement. The second main source of systematic uncertainty is from theory, which mostly originate from the W +jets PDF variations, accounting for 95% of the theoretical uncertainty (a variance of 0.316^2 out of a total of 0.324^2), when decomposing further. The JER and JES uncertainties are important contributions, which is not surprising since jets are used intensively in the reconstruction, and in establishing the regions for the MVA and the statistical model. Most NPs are from flavour tagging systematic variations. It is expected for the b - and c -eigenvariations to have a large contribution since the analysis relies heavily on b -tagging to define the regions, and c -tagging in the reconstruction of the hadronic W boson. The systematic uncertainty associated with the missing transverse energy, which is also used extensively in the reconstruction of the Higgs boson with the NW algorithm, as well as an important input to the BDT models under multiple forms, is non negligible. The muon and electron systematic uncertainties are less important, which might be a result of the constraining occurring from splitting the regions into lepton channels. Finally, the pileup reweighting SFs are a source of systematic uncertainty, and their associated NP contributes non-negligibly to the total POI uncertainty. The statistical uncertainty on the POI is given by: $\sigma_{\text{stat}} = 0.459$, which is smaller than the total systematic uncertainty if the lack of MC statistics is taken into account.

Figure 6.26 shows the correlation matrix using the same set of systematic groups as used for the uncertainty decomposition shown in Table 6.2. The correlations of grouped systematics are obtained by defining the asymmetric “grouping” matrix A , which has as many rows as there are groups, and as many columns as there are parameters (i.e the dimension of the covariance matrix). The entries A_{gi} are 1 if the corresponding NP i belongs to the group g , and 0 otherwise. The new covariance matrix is then obtained by transforming the original covariance matrix: $V_{\text{group}} = AVA^T$. This is equivalent to summing over all the pairwise covariances for two groups of NPs. The Pearson

NP Group	Uncertainty on POI
MC Statistics	0.677
Theoretical	0.324
JER	0.127
JES	0.114
Ftag b	0.114
Ftag c	0.097
Ftag light	0.040
Missing Transverse Energy	0.079
Muon	0.056
Electron	0.039
Pileup Reweighting Data SF	0.040
Total	0.793

Table 6.2: Uncertainty on the signal strength for different groups of systematics. The total is computed by adding each contribution in quadrature.

correlations are then obtained using the following well-known definition:

$$\rho_{ij} = \frac{(V_{\text{group}})_{ij}}{\sqrt{(V_{\text{group}})_{ii}(V_{\text{group}})_{jj}}}. \quad (6.42)$$

Some interesting comments can be made about the observed correlations. The electron systematics group seems to have stronger correlations than the muon systematics group, and is correlated with the fakes NF in the SRs (denoted mu_fakes), and mildly anti-correlated with the fakes in the CRs (named mu_fakes_Nb). The b -tagging systematics are strongly correlated with the top quark processes NF, which is expected since they tend to decrease the yield in the b -filtered CRs. In comparison, the c -tagging systematics group does not have correlations with any other group or individual parameter. The effects of c -tagging are expected to be more subtle since it is not used directly in the definition of regions, but instead at the reconstruction stage, to identify the jets reconstructing the hadronic W . The JES and JER groups are anti-correlated, which suggests that they tend to transform the nominal shape in the same way. Interestingly, the JES and JER group share the same correlation magnitude with the pileup reweighting scale factors systematics (named PRW_DATASF), but opposite sign. The gammas show no correlation, which is expected given their statistical nature. Concerning NFs, the JES group is anti-correlated with the diboson process, whereas the JER group is anti-correlated with the Z +jets

process. The W +jets NF shows a similar anti-correlation with fakes in both the SRs and CRs. The Z +jets NF, on the other hand, is anti-correlated with the fakes in the SRs, and correlated with the fakes in the CRs. The strongest correlations are between the diboson and SR fakes NFs, with a magnitude of 0.38. The dibosons are somewhat anti-correlated to the fakes in the CRs. The POI (named mu_signal in the figure) does not have any significant correlation except for an anti-correlation with the fakes in the SR.

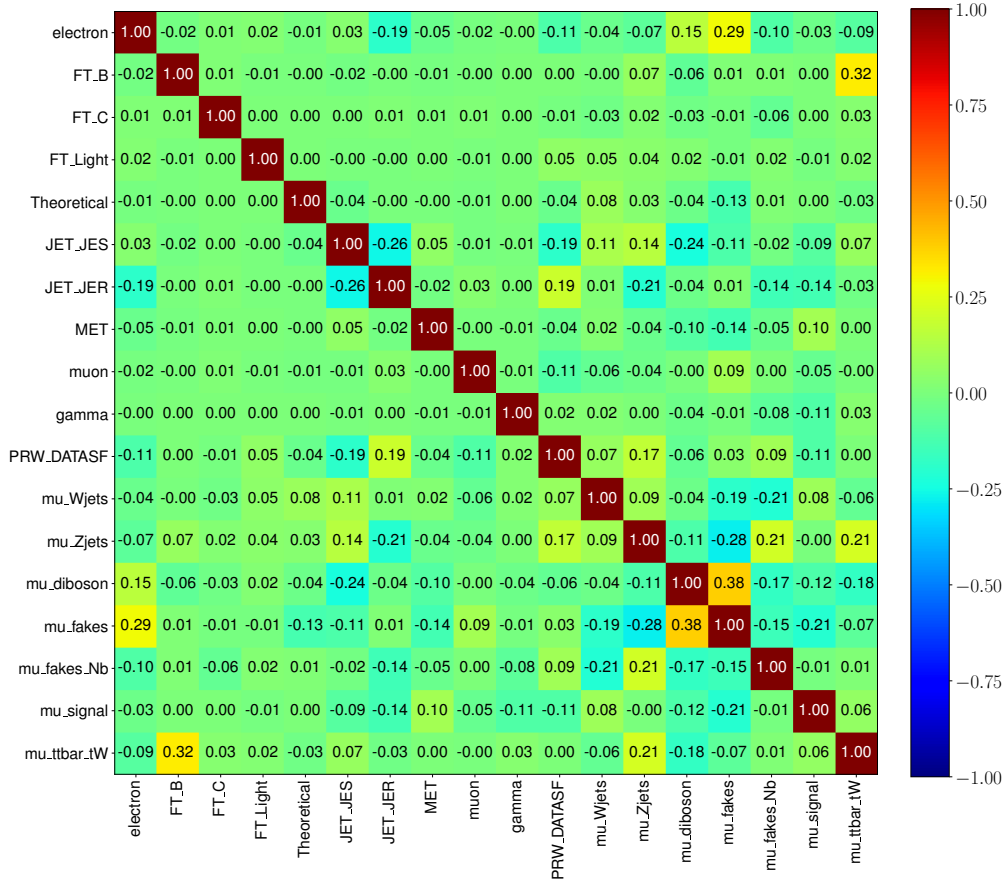


Figure 6.26: Correlation matrix for groups of systematic uncertainties and the NFs of the physical processes.

Limits

The expected CLs upper limit exclusion on the POI is given by $\mu_s = 1.804^{+0.715}_{-0.512}$ for Run 2 data only (140 fb^{-1}). An illustration of the limits is provided in Figure 6.27, where the $\pm 2\sigma$ bands are also indicated.

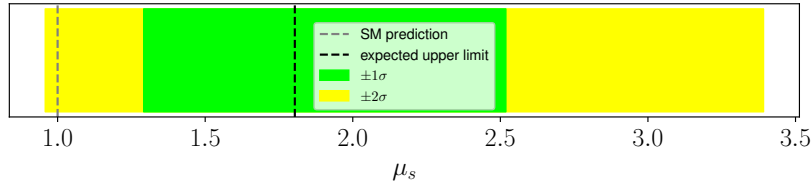


Figure 6.27: Illustration of the CLs limits obtained on the signal strength.

Projection of the expected precision with the HL-LHC

One can estimate the improvement in significance as the available integrated luminosity increases up to 3000 fb^{-1} , which should be reached by the High-Luminosity LHC (HL-LHC) program. The projections are produced by assuming that the grouped systematic uncertainties, detailed in Table 6.2, are Gaussian and independent from each other, which is not a bad approximation given the statistics and the low off-diagonal terms in the correlation matrix shown in Figure 6.26. The statistical uncertainty and every systematic uncertainty, except for the MC Statistics and Theoretical groups, are varied with luminosity with the following relation:

$$\sigma(L) = \sigma_0 \sqrt{\frac{L_0}{L}}, \quad (6.43)$$

where L denotes the luminosity, $L_0 = 140 \text{ fb}^{-1}$ and σ_0 is the uncertainty at the luminosity L_0 , corresponding to the Asimov data used for the fit. Some systematics may be scaled down by a constant factor to reflect potential improvements that come from a better understanding of the detector, relevant calibrations, or theory. Using this method, three scenarios are explored, in which only the MC statistics and theoretical systematics are affected. The theoretical uncertainty is assumed to be scaled down by a factor of two, and the three scenarios are then defined by the MC statistics being reduced by factors of 2, 5 and 1000, respectively. The scenarios are targeting the two dominating systematics, which additionally are not reduced with the luminosity increase since they are not being directly constrained within the statistical model. The theoretical uncertainty is scaled

down by a factor of two, which relies on the assumption that the current uncertainty is artificially amplified due to the large number of associated NPs entering the fit as a result of the unusually large amount of available statistics. For example, different W +jets PDF prescriptions may be used to model this theoretical effect without requiring several NPs. It is also possible that some theoretical progress is made with hadronic final states during HL-LHC. The MC statistical uncertainty currently determines the achievable sensitivity, and it is an effect that can be managed by generating more MC events with filters to target the regions of interest. It is for these two reasons that the scenarios focus on that systematic. Figure 6.28 shows the expected significance against the integrated luminosity for each scenario in plot (a). Plot (b) shows the CLs limits for scenario 3, computed by injecting the uncertainty into Wald's approximation shown in equation 6.33. From the difference between scenario 1 and scenario 2, and the fact that Scenario 1 flattens quickly around 2σ , it is clear that the uncertainty associated with the MC statistics is driving the overall sensitivity, as expected from the relative contributions. In Scenario 3, representing the ideal situation where the effect of MC statistics is negligible, the current model reaches the coveted 5σ sensitivity required for observation.

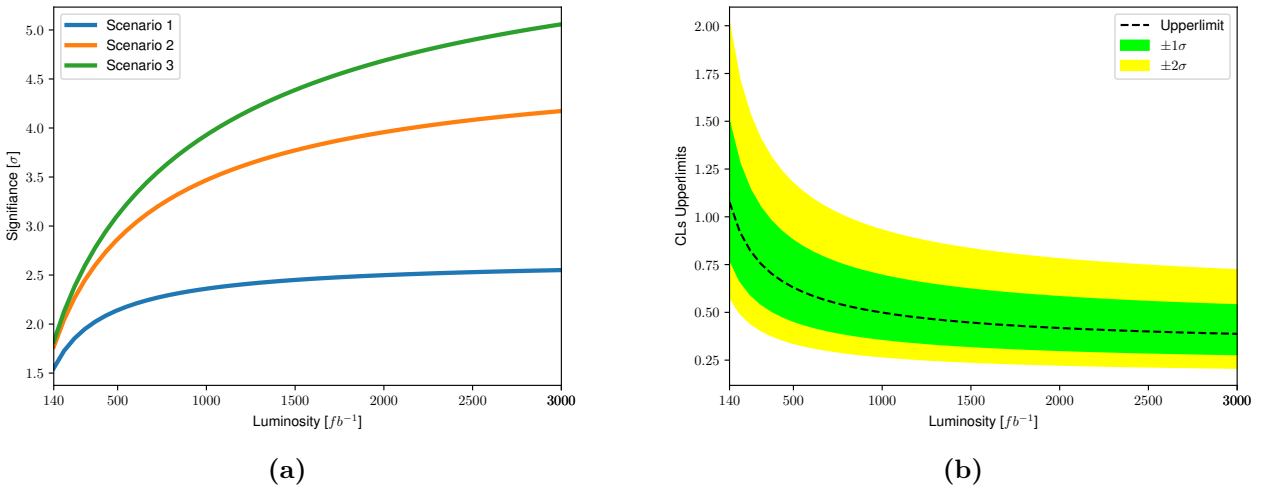


Figure 6.28: Projection for the significance under three distinct scenarios concerning the systematic uncertainties (a), and the CLs limits in the third scenario (b).

6.9.3 Mismodelling

The distribution of some observables show interesting disagreement with data. Figure 6.29 shows the missing transverse energy, including the fake lepton background contribution

whose estimation is described in section 6.5. At low values of $|\vec{E}_T^{\text{miss}}|$, the excess in data can be attributed to the fake leptons which are clearly imperfectly modelled, in particular for the muons. At around 50 GeV for electrons and 30 GeV for muons, the MC is greater than the data without the fake background, until around 140 GeV above which the MC and data agree nicely. The difference in mismodelling between electrons and muons can be explained by the fake lepton background whose contribution is much more important for electrons.

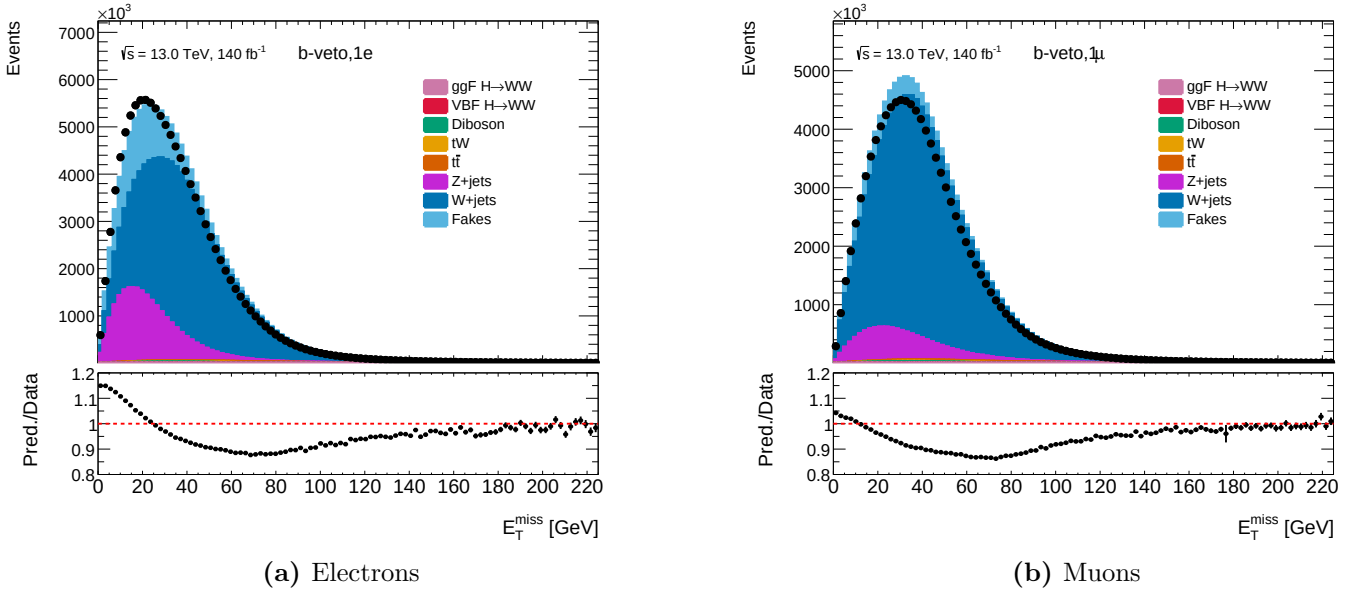


Figure 6.29: Distribution of the missing transverse energy for electrons (a) and muons (b).

The excess of MC occurs without addition of the fake lepton estimation, and thus indicates the presence of theoretical mismodelling, which could have effects that extend to the low E_T^{miss} region. This is highly problematic since the estimation of the fake lepton contribution relies on the MC, as can be seen from equation 6.7. In addition, it is likely that the fake-enriched region used to compute the fake efficiencies is contaminated by an important amount of real leptons given the large overestimation of fake leptons at relatively high E_T^{miss} and p_T , which is particularly visible for the muon channel.

These complicated considerations remain open questions and indicate that additional steps are needed before fitting to the observed data. Physics modelling and fake lepton experts should be contacted to plan accordingly.

6.9.4 Future steps

The most important task to perform next is the study of the mismodelling presented in section 6.9.3. It is likely that new regions will have to be defined, based on observables that are sensitive to the missing transverse energy, to derive reweighting factors to correct the MC.

The estimation of fake leptons currently follows a standard data-driven methodology used in many analyses. However, the fake background contribution is over-estimated, which can be seen when comparing the prediction to the observed data. It is still unclear whether this is caused by the MC background mismodelling, a prompt lepton contamination in the region used to measure the fake efficiency, or both.

The presence of the Z +jets background needs to be studied in more details. In particular, the way this process passes the preselection by having a muon or electron not reconstructed as a physics object may occur for different reasons. Identifying the mechanisms that cause the leptons to not be reconstructed by the detector might inform methods to further constrain the Z +jets events.

The inclusion of JetPairingNet into the reconstruction algorithm has drastically increased the available statistics by removing the c -tagging requirement, and the fitting strategy has not been updated to optimise the usage of these additional events. The signal regions, defined with the distributions of BDT output scores, should be split into control regions. A simple and natural way to do so is simply to cut on the scores themselves, keeping the high score region as a signal region, and attributing the low score events to a new control region.

Those new control regions should then be further split into finer control regions that target specific background processes. A control region for the Z +jets background can be defined by cutting on the missing transverse energy variables, and potentially the peak appearing in the $m_{\ell j_{1,2}}$ distributions. A diboson control region is more difficult to setup using simple cuts, and will probably require the usage of ML methods that consider several observables instead. The multi-class feature of **MLTrainer** should be used to discriminate the diboson process against the main other backgrounds: W +jets, Z +jets and $t\bar{t}$.

JetPairingNet has many hyperparameters, which have not been optimised. Additionally, the neural network was trained using only the ggF signal and it is thus expected to have a worse performance for the VBF-produced Higgs events. The no-pair score

output has been shown to be highly correlated with the observables obtained in the reconstruction, in particular the NW related variables. The no-pair head, which is currently of limited use, could be replaced by a background discriminating head instead, and JetPairingNet would then be trained to produce a score that separates the signal from background events. If this change does not have a negative impact on the jet pairing performance, this new setup would greatly simplify the workflow by removing the need for separate MVA studies dedicated to signal-background discrimination.

The 4-jets control region with b -filtering is dominated by $t\bar{t}$, which has been shown, via MVA studies, to be easier to separate from the signal compared to the W +jets events. This can also be seen from the BDT input variables plots, shown in section 6.7.2. Despite the fact that very few signal events are expected in that region, the usage of ML methods to potentially extract more significance should be studied.

The work presented here focuses on ggF, the main signal production mode. While the VBF production mode has a cross-section roughly ten times smaller, it also results in events that are considerably different to the main backgrounds compared to ggF-produced events. Dedicated studies should be performed to find optimal variables to exploit this separability. The multi-class feature of `MLTrainer` could then be further exploited to train models that take into account the VBF process as an independent class, and produce a score that can be used to define new regions used in the statistical model.

The input dataset to the ML models should be simplified by taking into account the correlations between features. The simplification can occur by simply removing variables with redundant information, or by using a more sophisticated method such as principal component analysis or auto-encoding.

Fitting the statistical model can be CPU intensive and numerically unstable due to the large number of floating NPs. The number of systematics can be reduced by using a different set of PDF variations for the W +jets process. The current systematics rely on the symmetric Hessian prescription, but as briefly described in section 6.8.2, other prescriptions exist which result in fewer NPs overall. A similar approach can be attempted for the flavour tagging NPs, which are even more numerous.

Conclusion

The semi-leptonic decay of $H \rightarrow WW^*$ is a physical process with strong potential to further probe the Higgs sector. First, a methodology to reconstruct the full final state was established in a phenomenology study, proposing a specific procedure to access the hadronic W and neutrino. In the same study, the power of this reconstructability was leveraged to develop an unfolding method for observables related to spin correlations and Quantum Information tests. Using this blueprint, an ATLAS analysis was built to search for the physical process, setting the baseline and foundational method to look for the signal in this region. Leaving the phenomenological world to engage with the technical nature of experimental physics revealed interesting challenges to resolve.

The analysis makes important use of flavour tagging and the NW algorithm, two techniques that have been studied and improved in the work presented here. The MC-MC efficiency maps were derived for the latest state-of-the-art flavour tagger GN2, used in the analysis for both b - and c -tagging. The run time and efficiency of the NW algorithm were considerably improved with the derivation of analytical methods.

The various components of the analysis are iteratively upgraded using more sophisticated approaches to improve sensitivity. In particular, the hadronic W is reconstructed with an advanced machine learning technique that uses c -tagging, but no longer requires a c -jet in the event, while removing the requirement on the W boson to be on-shell. The number of jets per event was found to affect the sensitivity, which encouraged the splitting of SRs based on jet multiplicity, and b -tagging is used to define CRs that target the top quark processes. Five BDTs are trained per SR in folds to avoid bias, and the scores are used for the profile-likelihood fit. The statistical model is complicated due to the important number of NPs that have to be considered because of the large amount of statistics in the SRs. The presence of mismodelling with respect to missing transverse energy in the SR makes fitting to data unfeasible. More studies are required to understand, and potentially correct, the shapes of the distributions.

The Asimov dataset is used to compute the expected CLs upper limit on the ggF+VBF combined signal strength, and is found to be $\mu_s = 1.804^{+0.715}_{-0.512}$ for 140 fb^{-1} . This result is very preliminary, and many improvements can still be brought to the analysis. In particular, the region definitions in the statistical model combined with the usage of multi-class ML models should be able to reach higher sensitivity. Most of the uncertainty is caused by the lack of MC statistics, which is a manageable systematic source. Despite the current setup being non-optimal, simple projections to the HL-LHC luminosity (that are made conservative by assuming no major improvement in experimental systematics) reveal that handling that uncertainty can lead to a 5σ observation of the channel.

An interesting lesson can be learned from the differences between the phenomenology study and the experimental performance. The study focused on the signal purity to improve entanglement interpretation, instead of looking pragmatically into the sensitivity to the channel. In essence, we presumptuously asked what the channel could do for us before assessing its existence.

Bibliography

- [1] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, 1995), 1995, DOI: <https://doi.org/10.1201/9780429503559>.
- [2] M. D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, 2014), 2014, DOI: <https://doi.org/10.1017/9781139540940>.
- [3] J. Illana, *Field theory & the EW standard model*, Lecture at the CERN School of High-Energy Physics, 2025, Slides available at: <https://indico.cern.ch/event/1531763/contributions/6625074/>.
- [4] R. Penrose and W. Rindler, *Spinors and Space-Time, Vol. 1* (Cambridge University Press, 1984), 1984, DOI: <https://doi.org/10.1017/CB09780511564048>.
- [5] W. Pauli, Phys. Rev. **58**, 716 (1940), *The Connection Between Spin and Statistics*, DOI: <https://doi.org/10.1103/PhysRev.58.716>.
- [6] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations* (Cambridge University Press, 2005), 2005, DOI: <https://doi.org/10.1017/CB09781139644167>.
- [7] R. F. Streater and A. S. Wightman, *PCT, spin and statistics, and all that* (Princeton University Press, 1989), 1989, DOI: <https://doi.org/10.1515/9781400884230>.
- [8] K.-i. Hikasa, PTEP **2024**, 08C101 (2024), 2405.19774, *CPT, Majorana Fermions, and Particle Physics Beyond the Standard Model*, DOI: <https://doi.org/10.1093/ptep/ptae081>.
- [9] E. Noether, Gott. Nachr. **1918**, 235 (1918), physics/0503066, *Invariant Variation Problems*, DOI: <https://doi.org/10.1080/00411457108231446>.
- [10] L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967), *Feynman Diagrams for the Yang-Mills Field*, DOI: [https://doi.org/10.1016/0370-2693\(67\)90067-6](https://doi.org/10.1016/0370-2693(67)90067-6).

- [11] J. L. Diaz-Cruz, B. L. Lopez, O. Meza-Aldama, and J. R. Perez, Rev. Mex. Fis. E **61**, 104 (2015), 1511.07477, *Weyl spinors and the helicity formalism*, DOI: <https://doi.org/10.48550/arXiv.1511.07477>.
- [12] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964), *Broken Symmetries and the Masses of Gauge Bosons*, DOI: <https://doi.org/10.1103/PhysRevLett.13.508>.
- [13] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964), *Broken Symmetry and the Mass of Gauge Vector Mesons*, DOI: <https://doi.org/10.1103/PhysRevLett.13.321>.
- [14] J. Goldstone, Nuovo Cim. **19**, 154 (1961), *Field Theories with "Superconductor" Solutions*, DOI: <https://doi.org/10.1007/BF02812722>.
- [15] J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962), *Broken Symmetries*, DOI: <https://doi.org/10.1103/PhysRev.127.965>.
- [16] Particle Data Group, S. Navas *et al.*, Phys. Rev. D **110**, 030001 (2024), *Review of particle physics*, DOI: <https://doi.org/10.1103/PhysRevD.110.030001>.
- [17] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963), *Unitary Symmetry and Leptonic Decays*, DOI: <https://doi.org/10.1103/PhysRevLett.10.531>.
- [18] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973), *CP Violation in the Renormalizable Theory of Weak Interaction*, DOI: <https://doi.org/10.1143/PTP.49.652>.
- [19] Super-Kamiokande, Y. Ashie *et al.*, Phys. Rev. Lett. **93**, 101801 (2004), hep-ex/0404034, *Evidence for an oscillatory signature in atmospheric neutrino oscillation*, DOI: <https://doi.org/10.1103/PhysRevLett.93.101801>.
- [20] KamLAND, K. Eguchi *et al.*, Phys. Rev. Lett. **90**, 021802 (2003), hep-ex/0212021, *First results from KamLAND: Evidence for reactor anti-neutrino disappearance*, DOI: <https://doi.org/10.1103/PhysRevLett.90.021802>.
- [21] S. W. "Cush", *Standard Model of Elementary Particles (dark background)*, https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_dark.svg, 2017, Public domain (CC0), Wikimedia Commons.

- [22] The ATLAS Collaboration, G. Aad *et al.*, JHEP **05**, 028 (2023), 2207.08615, *Measurement of the total and differential Higgs boson production cross-sections at $\sqrt{s} = 13$ TeV with the ATLAS detector by combining the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ decay channels*, DOI: [https://doi.org/10.1007/JHEP05\(2023\)028](https://doi.org/10.1007/JHEP05(2023)028).
- [23] LHC Higgs Cross Section Working Group, *LHC Higgs Cross Section Working Group: Cross Sections and Figures*, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGCrossSectionsFigures>, CERN TWiki, accessed 22 January 2026.
- [24] J. A. Aguilar-Saavedra and J. Bernabeu, Phys. Rev. D **93**, 011301 (2016), 1508.04592, *Breaking down the entire W boson spin observables from its decay*, DOI: <https://doi.org/10.1103/PhysRevD.93.011301>.
- [25] Z. Bern *et al.*, Phys. Rev. D **84**, 034008 (2011), 1103.5445, *Left-Handed W Bosons at the LHC*, DOI: <https://doi.org/10.1103/PhysRevD.84.034008>.
- [26] W. J. Stirling and E. Vryonidou, JHEP **07**, 124 (2012), 1204.6427, *Electroweak gauge boson polarisation at the LHC*, DOI: [https://doi.org/10.1007/JHEP07\(2012\)124](https://doi.org/10.1007/JHEP07(2012)124).
- [27] The D0 Collaboration, V. M. Abazov *et al.*, Phys. Rev. D **83**, 032009 (2011), 1011.6549, *Measurement of the W Boson Helicity in Top Quark Decays using 5.4 fb^{-1} of $p\bar{p}$ Collision Data*, DOI: <https://doi.org/10.1103/PhysRevD.83.032009>.
- [28] The CDF Collaboration, T. Aaltonen *et al.*, Phys. Rev. D **87**, 031104 (2013), 1211.4523, *Measurement of W-Boson Polarization in Top-quark Decay using the Full CDF Run II Data Set*, DOI: <https://doi.org/10.1103/PhysRevD.87.031104>.
- [29] The ATLAS Collaboration, G. Aad *et al.*, JHEP **06**, 088 (2012), 1205.2484, *Measurement of the W boson polarization in top quark decays with the ATLAS detector*, DOI: [https://doi.org/10.1007/JHEP06\(2012\)088](https://doi.org/10.1007/JHEP06(2012)088).
- [30] The CMS Collaboration, S. Chatrchyan *et al.*, JHEP **10**, 167 (2013), 1308.3879, *Measurement of the W-Boson Helicity in Top-Quark decays from $t\bar{t}$ Production in Lepton + Jets Events in pp Collisions at $\sqrt{s} = 7$ TeV*, DOI: [https://doi.org/10.1007/JHEP10\(2013\)167](https://doi.org/10.1007/JHEP10(2013)167).

- [31] The CMS Collaboration, V. Khachatryan *et al.*, JHEP **01**, 053 (2015), 1410.1154, *Measurement of the W boson helicity in events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 8$ TeV*, DOI: [https://doi.org/10.1007/JHEP01\(2015\)053](https://doi.org/10.1007/JHEP01(2015)053).
- [32] R. Ashby-Pickering, A. J. Barr, and A. Wierchucka, JHEP **05**, 020 (2023), 2209.13990, *Quantum state tomography, entanglement detection and Bell violation prospects in weak decays of massive particles*, DOI: [https://doi.org/10.1007/JHEP05\(2023\)020](https://doi.org/10.1007/JHEP05(2023)020).
- [33] The ATLAS and CMS Collaborations, F. Hubaut, *Top physics at LHC with $t\bar{t}$ events*, in *41st Rencontres de Moriond: QCD and Hadronic Interactions*, pp. 169–172, 2006, hep-ex/0605029.
- [34] A. Czarnecki, M. Jezabek, and J. H. Kuhn, Nucl. Phys. B **351**, 70 (1991), *Lepton Spectra From Decays of Polarized Top Quarks*, DOI: [https://doi.org/10.1016/0550-3213\(91\)90082-9](https://doi.org/10.1016/0550-3213(91)90082-9).
- [35] A. Brandenburg, Z. G. Si, and P. Uwer, Phys. Lett. B **539**, 235 (2002), hep-ph/0205023, *QCD corrected spin analyzing power of jets in decays of polarized top quarks*, DOI: [https://doi.org/10.1016/S0370-2693\(02\)02098-1](https://doi.org/10.1016/S0370-2693(02)02098-1).
- [36] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935), *Can Quantum-Mechanical Description of physical reality be considered complete?*, DOI: <https://doi.org/10.1103/PhysRev.47.777>.
- [37] J. S. Bell, Physics Physique Fizika **1**, 195 (1964), *On the Einstein-Podolsky-Rosen paradox*, DOI: <https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195>.
- [38] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972), *Experimental Test of Local Hidden-Variable Theories*, DOI: <https://doi.org/10.1103/PhysRevLett.28.938>.
- [39] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982), *Experimental Test of Bell's Inequalities Using Time-Varying Analyzers*, DOI: <https://doi.org/10.1103/PhysRevLett.49.1804>.
- [40] A. A. Tomaz, R. S. Mattos, and M. Barbatti, (2025), 2502.19278, *The Quantum Measurement Problem: A Review of Recent Trends*, DOI: <https://doi.org/10.1080/14786435.2025.2601922>.

- [41] J. Bell, *Physics World* **3**, 33 (1990), *Against ‘measurement’*, DOI: <https://doi.org/10.1088/2058-7058/3/8/26>.
- [42] M. Schlosshauer, *Rev. Mod. Phys.* **76**, 1267 (2004), quant-ph/0312059, *Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics*, DOI: <https://doi.org/10.1103/RevModPhys.76.1267>.
- [43] S. Carlip, *Rept. Prog. Phys.* **64**, 885 (2001), gr-qc/0108040, *Quantum gravity: A Progress report*, DOI: <https://doi.org/10.1088/0034-4885/64/8/301>.
- [44] R. Penrose, *Gen. Rel. Grav.* **28**, 581 (1996), *On gravity’s role in quantum state reduction*, DOI: <https://doi.org/10.1007/BF02105068>.
- [45] M. Fabbrichesi, R. Floreanini, and E. Gabrielli, *Eur. Phys. J. C* **83**, 162 (2023), 2208.11723, *Constraining new physics in entangled two-qubit systems: top-quark, tau-lepton and photon pairs*, DOI: <https://doi.org/10.1140/epjc/s10052-023-11307-2>.
- [46] R. Aoude, E. Madge, F. Maltoni, and L. Mantani, *JHEP* **12**, 017 (2023), 2307.09675, *Probing new physics through entanglement in diboson production*, DOI: [https://doi.org/10.1007/JHEP12\(2023\)017](https://doi.org/10.1007/JHEP12(2023)017).
- [47] A. J. Barr, M. Fabbrichesi, R. Floreanini, E. Gabrielli, and L. Marzola, *Prog. Part. Nucl. Phys.* **139**, 104134 (2024), 2402.07972, *Quantum entanglement and Bell inequality violation at colliders*, DOI: <https://doi.org/10.1016/j.pnpnp.2024.104134>.
- [48] A. Peres, *Phys. Rev. Lett.* **77**, 1413 (1996), quant-ph/9604005, *Separability criterion for density matrices*, DOI: <https://doi.org/10.1103/PhysRevLett.77.1413>.
- [49] M. Horodecki, P. Horodecki, and R. Horodecki, *Phys. Lett. A* **223**, 1 (1996), quant-ph/9605038, *On the necessary and sufficient conditions for separability of mixed quantum states*, DOI: [https://doi.org/10.1016/S0375-9601\(96\)00706-2](https://doi.org/10.1016/S0375-9601(96)00706-2).
- [50] B. S. Cirelson, *Lett. Math. Phys.* **4**, 93 (1980), *Quantum Generalizations of Bell’s Inequality*, DOI: <https://doi.org/10.1007/BF00417500>.
- [51] S. Popescu and D. Rohrlich, *Found. Phys.* **24**, 379 (1994), *Quantum nonlocality as an axiom*, DOI: <https://doi.org/10.1007/BF02058098>.

- [52] D. Kaszlikowski, L. C. Kwek, J.-L. Chen, M. Żukowski, and C. H. Oh, Phys. Rev. A **65**, 032118 (2002), *Clauser-Horne inequality for three-state systems*, DOI: <https://doi.org/10.1103/PhysRevA.65.032118>.
- [53] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Phys. Rev. Lett. **88**, 040404 (2002), *Bell Inequalities for Arbitrarily High-Dimensional Systems*, DOI: <https://doi.org/10.1103/PhysRevLett.88.040404>.
- [54] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson, Nature Physics **7**, 677–680 (2011), *Experimental high-dimensional two-photon entanglement and violations of generalized Bell inequalities*, DOI: <https://doi.org/10.1038/nphys1996>.
- [55] T. Ikuta and H. Takesue, Physical Review A **93** (2016), *Enhanced violation of the Collins-Gisin-Linden-Massar-Popescu inequality with optimized time-bin-entangled ququarts*, DOI: <https://doi.org/10.1103/physreva.93.022307>.
- [56] D. Zhang, X. Qiu, and L. Chen, Physical Review A **110** (2024), *Experimental test of the Collins-Gisin-Linden-Massar-Popescu inequality for multisetting and multidimensional orbital angular momentum systems*, DOI: <https://doi.org/10.1103/physreva.110.012202>.
- [57] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962), *Symmetries of Baryons and Mesons*, DOI: <https://doi.org/10.1103/PhysRev.125.1067>.
- [58] A. J. Barr, Physics Letters B **825**, 136866 (2022), *Testing Bell inequalities in Higgs boson decays*, DOI: <https://doi.org/10.1016/j.physletb.2021.136866>.
- [59] S. Li, W. Shen, and J. M. Yang, Eur. Phys. J. C **84**, 1195 (2024), 2401.01162, *Can Bell inequalities be tested via scattering cross-section at colliders ?*, DOI: <https://doi.org/10.1140/epjc/s10052-024-13584-x>.
- [60] The ATLAS Collaboration, G. Aad *et al.*, Nature **633**, 542 (2024), 2311.07288, *Observation of quantum entanglement with top quarks at the ATLAS detector*, DOI: <https://doi.org/10.1038/s41586-024-07824-z>.
- [61] The CMS Collaboration, A. Hayrapetyan *et al.*, Rept. Prog. Phys. **87**, 117801 (2024), 2406.03976, *Observation of quantum entanglement in top quark pair production in proton–proton collisions at $\sqrt{s} = 13$ TeV*, DOI: <https://doi.org/10.1088/1361-6633/ad7e4d>.

- [62] E. L. Berger *et al.*, Physical Review D **82** (2010), *Higgs boson search sensitivity in the $H \rightarrow WW$ Dilepton decay mode at $\sqrt{s} = 7$ and 10 TeV*, DOI: <https://doi.org/10.1103/physrevd.82.053003>.
- [63] L. Evans and P. Bryant, Journal of Instrumentation **3**, S08001 (2008), *LHC Machine*, DOI: <https://doi.org/10.1088/1748-0221/3/08/S08001>.
- [64] M. A. Hone, Report CERN-PS-LR-79-37, 1979, *The duoplasmatron ion source for the new CERN Linac preinjector*, <https://cds.cern.ch/record/2640736>.
- [65] E. Boltezar *et al.*, IEEE Trans. Nucl. Sci. **26**, 3674 (1979), *Performance of the new CERN 50 MeV Linac*, DOI: <https://doi.org/10.1109/TNS.1979.4330576>.
- [66] J. Vollaie *et al.*, *Linac4 design report* Vol. 6/2020 (CERN, Geneva, 2020), 2020, DOI: <https://doi.org/10.23731/CYRM-2020-006>.
- [67] S. Dewan, R. Holmes, K. Reiniger, F. Völker, and J. Zeng, Report CERN-PS-PO-Note-2000-052, 2000, *High Power Switch-mode Power Converters for the 1.4 GeV PS-Booster Beam Transfer Lines at CERN*, <https://cds.cern.ch/record/2830275>.
- [68] K. Hanke *et al.*, Report CERN-ACC-2013-0024, 2013, *A possible scheme to deliver 2 GeV Beams from the CERN PS Booster to the ISOLDE Facility*, <https://cds.cern.ch/record/1566843>.
- [69] J.-P. Burnet *et al.*, *Fifty years of the CERN Proton Synchrotron* (CERN, Geneva, 2011), 2011, DOI: <https://doi.org/10.5170/CERN-2011-004>.
- [70] E. Mobs, (2018), *The CERN accelerator complex - August 2018*, General Photo.
- [71] O. S. Brüning *et al.*, *LHC Design Report* (CERN, Geneva, 2004), 2004, DOI: <https://doi.org/10.5170/CERN-2004-003-V-1>.
- [72] B. J. Holzer, (2013), *Beam optics and lattice design for particle accelerators*, DOI: <https://doi.org/10.5170/CERN-2013-001.171>.
- [73] The ATLAS Collaboration, M. Aaboud *et al.*, Phys. Rev. Lett. **117**, 182002 (2016), 1606.02625, *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, DOI: <https://doi.org/10.1103/PhysRevLett.117.182002>.

- [74] *Proceedings, 8th Evian Workshop on LHC beam operation: Evian Les Bains, France, December 12-14, 2017*, No. CERN-ACC-2019-012, Geneva, 2019, CERN.
- [75] F. Zimmermann, CERN Yellow Rep. Conf. Proc. **7**, 1 (2020), *Electron cloud effects in accelerators*, DOI: <https://doi.org/10.23732/CYRCP-2020-007.1>.
- [76] The ATLAS Collaboration, *ATLAS Public Data Quality Results*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/DataQualityResults>, Accessed: 2026-03-17.
- [77] The ATLAS Collaboration, *ATLAS Luminosity Public Results, Run 2*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [78] The ATLAS Collaboration, *ATLAS Luminosity Public Results, Run 3*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>.
- [79] M. Leigh *et al.*, SciPost Physics **14** (2023), *ν -flows: Conditional neutrino regression*, DOI: <https://doi.org/10.21468/scipostphys.14.6.159>.
- [80] The ATLAS Collaboration, G. Aad *et al.*, JINST **19**, P05063 (2024), 2305.16623, *The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3*, DOI: <https://doi.org/10.1088/1748-0221/19/05/P05063>.
- [81] J. Pequeno and P. Schaffner, *How ATLAS detects particles: diagram of particle paths in the detector*, <https://cds.cern.ch/record/1505342>, 2013.
- [82] The ATLAS Collaboration, G. Aad *et al.*, JINST **3**, S08003 (2008), *The ATLAS Experiment at the CERN Large Hadron Collider*, DOI: <https://doi.org/10.1088/1748-0221/3/08/S08003>.
- [83] J. Pequeno, *Computer generated image of the ATLAS inner detector*, <https://cds.cern.ch/record/1095926>, 2008.
- [84] T. A. Collaboration, (2015), *Charged-particle distributions in $\sqrt{s}=13$ TeV $p p$ interactions measured with the ATLAS detector at the LHC*.

- [85] A. Andronic and J. Wessels, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **666**, 130 (2012), *Transition radiation detectors*, DOI: <https://doi.org/https://doi.org/10.1016/j.nima.2011.09.041>, Advanced Instrumentation.
- [86] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **84**, 1313 (2024), 2401.16034, *Operation and performance of the ATLAS tile calorimeter in LHC Run 2*, DOI: <https://doi.org/10.1140/epjc/s10052-024-13151-4>.
- [87] D. M. Gingrich *et al.*, JINST **2**, P05005 (2007), *Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter*, DOI: <https://doi.org/10.1088/1748-0221/2/05/P05005>.
- [88] A. Artamonov *et al.*, Journal of Instrumentation **3**, P02010 (2008), *The ATLAS Forward Calorimeter*, DOI: <https://doi.org/10.1088/1748-0221/3/02/P02010>.
- [89] T. A. T. Collaboration, Journal of Instrumentation **11**, P06008 (2016), *The ATLAS Data Acquisition and High Level Trigger system*, DOI: <https://doi.org/10.1088/1748-0221/11/06/P06008>.
- [90] L. Wickremasinghe, *Illustration of track related parameters used in the ATLAS experiment.*, 2025.
- [91] R. Fruhwirth, Nucl. Instrum. Meth. A **262**, 444 (1987), *Application of Kalman filtering to track and vertex fitting*, DOI: [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [92] T. Cornelissen *et al.*, J. Phys. Conf. Ser. **119**, 032014 (2008), *The new ATLAS track reconstruction (NEWT)*, DOI: <https://doi.org/10.1088/1742-6596/119/3/032014>.
- [93] The European Physical Journal C **77** (2017), *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, DOI: <https://doi.org/10.1140/epjc/s10052-017-4887-5>.
- [94] T. Robertson and J. D. Cryer, Journal of the American Statistical Association **69**, 1012 (1974), *An Iterative Procedure for Estimating the Mode*, DOI: <https://doi.org/10.2307/2286181>.
- [95] W. Waltenberger, Report CERN-CMS-NOTE-2008-033, 2008, *Adaptive vertex reconstruction*, <https://cds.cern.ch/record/1166320>.

- [96] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), 0802.1189, *The anti- k_t jet clustering algorithm*, DOI: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [97] G. P. Salam, Eur. Phys. J. C **67**, 637 (2010), 0906.1833, *Towards Jetography*, DOI: <https://doi.org/10.1140/epjc/s10052-010-1314-6>.
- [98] T. D. Lee and M. Nauenberg, Phys. Rev. **133**, B1549 (1964), *Degenerate Systems and Mass Singularities*, DOI: <https://doi.org/10.1103/PhysRev.133.B1549>.
- [99] A. Buckley *et al.*, Phys. Rept. **504**, 145 (2011), 1101.2599, *General-purpose event generators for LHC physics*, DOI: <https://doi.org/10.1016/j.physrep.2011.03.005>.
- [100] The ATLAS Collaboration, M. Aaboud *et al.*, Eur. Phys. J. C **77**, 466 (2017), 1703.10485, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, DOI: <https://doi.org/10.1140/epjc/s10052-017-5031-2>.
- [101] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **84**, 455 (2024), 2310.16497, *Improving topological cluster reconstruction using calorimeter cell timing in ATLAS*, DOI: <https://doi.org/10.1140/epjc/s10052-024-12657-1>.
- [102] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **77**, 241 (2017), 1609.09324, *Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions in the ATLAS detector*, DOI: <https://doi.org/10.1140/epjc/s10052-017-4780-2>.
- [103] The ATLAS Collaboration, M. Aaboud *et al.*, Eur. Phys. J. C **79**, 639 (2019), 1902.04655, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV*, DOI: <https://doi.org/10.1140/epjc/s10052-019-7140-6>.
- [104] W. Lampl *et al.*, Report ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003, 2008, *Calorimeter Clustering Algorithms*, <https://cds.cern.ch/record/1099735>.
- [105] R. Frühwirth, Computer Physics Communications **154**, 131 (2003), *A Gaussian-mixture approximation of the Bethe-Heitler model of electron energy loss by bremsstrahlung*, DOI: [https://doi.org/https://doi.org/10.1016/S0010-4655\(03\)00292-3](https://doi.org/https://doi.org/10.1016/S0010-4655(03)00292-3).

- [106] The ATLAS Collaboration, M. Aaboud *et al.*, JINST **14**, P03017 (2019), 1812.03848, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data*, DOI: <https://doi.org/10.1088/1748-0221/14/03/P03017>.
- [107] The ATLAS Collaboration, G. Aad *et al.*, JINST **19**, P02009 (2024), 2309.05471, *Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data*, DOI: <https://doi.org/10.1088/1748-0221/19/02/P02009>.
- [108] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **81**, 578 (2021), 2012.00578, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, DOI: <https://doi.org/10.1140/epjc/s10052-021-09233-2>.
- [109] J. Illingworth and J. Kittler, Comput. Vision Graphics Image Proc. **44**, 87 (1988), *A survey of the hough transform*, DOI: [https://doi.org/10.1016/s0734-189x\(88\)80033-1](https://doi.org/10.1016/s0734-189x(88)80033-1).
- [110] J. C. Collins, D. E. Soper, and G. F. Sterman, Adv. Ser. Direct. High Energy Phys. **5**, 1 (1989), hep-ph/0409313, *Factorization of Hard Processes in QCD*, DOI: https://doi.org/10.1142/9789814503266_0001.
- [111] C. Lorcé, A. Metz, B. Pasquini, and P. Schweitzer, (2025), 2507.12664, *Parton Distribution Functions and their Generalizations*, DOI: <https://doi.org/10.48550/arXiv.2507.12664>.
- [112] M. R. Whalley, D. Bourilkov, and R. C. Group, p. 575 (2005), hep-ph/0508110, *The Les Houches accord PDFs (LHAPDF) and LHAGLUE*, DOI: <https://doi.org/10.48550/arXiv.hep-ph/0508110>.
- [113] B. Andersson, **7** (1998), *The Lund Model*, DOI: <https://doi.org/10.1017/9781009401296>.
- [114] C. Bierlich *et al.*, (2022), 2203.11601, *A comprehensive guide to the physics and usage of PYTHIA 8.3*, DOI: <https://doi.org/10.48550/arXiv.2203.11601>.
- [115] B. R. Webber, Nucl. Phys. B **238**, 492 (1984), *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, DOI: [https://doi.org/10.1016/0550-3213\(84\)90333-X](https://doi.org/10.1016/0550-3213(84)90333-X).

- [116] J.-C. Winter, F. Krauss, and G. Soff, Eur. Phys. J. C **36**, 381 (2004), hep-ph/0311085, *A Modified cluster hadronization model*, DOI: <https://doi.org/10.1140/epjc/s2004-01960-8>.
- [117] C. M. Buttar *et al.*, (2005), *The underlying event*, DOI: <https://doi.org/10.5170/CERN-2005-014.192>.
- [118] The ATLAS Collaboration, Report ATL-PHYS-PUB-2014-021, 2014, *ATLAS Pythia 8 tunes to 7 TeV data*, <https://cds.cern.ch/record/1966419>.
- [119] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **74**, 2965 (2014), 1406.0392, *Measurement of the underlying event in jet events from 7 TeV proton-proton collisions with the ATLAS detector*, DOI: <https://doi.org/10.1140/epjc/s10052-014-2965-5>.
- [120] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **70**, 823 (2010), 1005.4568, *The ATLAS Simulation Infrastructure*, DOI: <https://doi.org/10.1140/epjc/s10052-010-1429-9>.
- [121] N. Metropolis and S. Ulam, Journal of the American Statistical Association **44**, 335 (1949), *The Monte Carlo Method*, DOI: <https://doi.org/10.2307/2280232>.
- [122] The ATLAS Collaboration, (2019), *Athena*, DOI: <https://doi.org/10.5281/zenodo.2641997>.
- [123] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **85**, 234 (2025), 2404.06335, *Software and computing for Run 3 of the ATLAS experiment at the LHC*, DOI: <https://doi.org/10.1140/epjc/s10052-024-13701-w>.
- [124] J. Allison *et al.*, Nucl. Instrum. Meth. A **835**, 186 (2016), *Recent developments in Geant4*, DOI: <https://doi.org/10.1016/j.nima.2016.06.125>.
- [125] The ATLAS Collaboration, Report ATL-PHYS-PUB-2017-013, 2017, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, <https://cds.cern.ch/record/2273281>.
- [126] The ATLAS Collaboration, Report ATL-PHYS-PUB-2022-047, 2022, *Neural Network Jet Flavour Tagging with the Upgraded ATLAS Inner Tracker Detector at the High-Luminosity LHC*, <https://cds.cern.ch/record/2839913>.

- [127] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, (2013), 1302.4389, *Maxout Networks*, DOI: <https://doi.org/10.48550/arXiv.1302.4389>.
- [128] The ATLAS Collaboration, Report ATL-PHYS-PUB-2017-011, 2017, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, <https://cds.cern.ch/record/2270366>.
- [129] The ATLAS Collaboration, Report ATL-PHYS-PUB-2018-025, 2018, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, <https://cds.cern.ch/record/2645405/>.
- [130] M. Zaheer *et al.*, (2018), 1703.06114, *Deep Sets*, DOI: <https://doi.org/10.48550/arXiv.1703.06114>.
- [131] The ATLAS Collaboration, Report ATL-PHYS-PUB-2020-014, 2020, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, <https://cds.cern.ch/record/2718948>.
- [132] The ATLAS Collaboration, Report ATL-PHYS-PUB-2022-027, 2022, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, <https://cds.cern.ch/record/2811135>.
- [133] The ATLAS Collaboration, G. Aad *et al.*, (2025), 2505.19689, *Transforming jet flavour tagging at ATLAS*, DOI: <https://doi.org/10.1038/s41467-025-65059-6>.
- [134] P. W. Battaglia *et al.*, (2018), 1806.01261, *Relational inductive biases, deep learning, and graph networks*, DOI: <https://doi.org/10.48550/arXiv.1806.01261>.
- [135] J. Shlomi, P. Battaglia, and J.-R. Vlimant, Machine Learning: Science and Technology **2**, 021001 (2020), *Graph neural networks in particle physics*, DOI: <https://doi.org/10.1088/2632-2153/abfb9a>.
- [136] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **83**, 681 (2023), 2211.16345, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, DOI: <https://doi.org/10.1140/epjc/s10052-023-11699-1>.
- [137] The ATLAS Collaboration, (2023), *Jet Flavour Tagging with GN1 and DL1d. Generator dependence, Run 2 and Run 3 agreement studies.*, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/>.

- [138] M. Sundararajan, A. Taly, and Q. Yan, (2017), 1703.01365, *Axiomatic Attribution for Deep Networks*, DOI: <https://doi.org/10.48550/arXiv.1703.01365>.
- [139] The ATLAS Collaboration, Report ATL-PHYS-PUB-2025-029, 2025, *Explaining the ATLAS GN2 Flavour Tagging Algorithm with Integrated Gradients*, <https://cds.cern.ch/record/2936634>.
- [140] T. Gneiting and P. Vogel, (2018), 1809.04808, *Receiver Operating Characteristic (ROC) Curves*, DOI: <https://doi.org/10.48550/arXiv.1809.04808>.
- [141] P. Veličković *et al.*, (2018), 1710.10903, *Graph Attention Networks*, DOI: <https://doi.org/10.48550/arXiv.1710.10903>.
- [142] S. Brody, U. Alon, and E. Yahav, (2022), 2105.14491, *How Attentive are Graph Attention Networks?*, DOI: <https://doi.org/10.48550/arXiv.2105.14491>.
- [143] A. Vaswani *et al.*, (2023), 1706.03762, *Attention Is All You Need*, DOI: <https://doi.org/10.48550/arXiv.1706.03762>.
- [144] S. Frixione, P. Nason, and C. Oleari, Journal of High Energy Physics **2007**, 070–070 (2007), *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, DOI: <https://doi.org/10.1088/1126-6708/2007/11/070>.
- [145] The ATLAS Collaboration, G. Aad *et al.*, The European Physical Journal C **79** (2019), *ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, DOI: <https://doi.org/10.1140/epjc/s10052-019-7450-8>.
- [146] A. Buckley *et al.*, Computer Physics Communications **184**, 2803 (2013), *Rivet user manual*, DOI: <https://doi.org/https://doi.org/10.1016/j.cpc.2013.05.021>.
- [147] A. Buckley *et al.*, Comput. Phys. Commun. **260**, 107310 (2021), 1912.08005, *The HepMC3 event record library for Monte Carlo event generators*, DOI: <https://doi.org/10.1016/j.cpc.2020.107310>.
- [148] Rivet Analysis Coverage, <https://rivet.hepforge.org/rivet-coverage>.
- [149] Y. L. Dokshitzer, V. A. Khoze, and S. I. Troian, J. Phys. G **17**, 1602 (1991), *On specific QCD properties of heavy quark fragmentation ('dead cone')*, DOI: <https://doi.org/10.1088/0954-3899/17/10/023>.

- [150] The ALICE Collaboration, S. Acharya *et al.*, Nature **605**, 440 (2022), 2106.05713, *Direct observation of the dead-cone effect in quantum chromodynamics*, DOI: <https://doi.org/10.1038/s41586-022-04572-w>, [Erratum: Nature 607, E22 (2022)].
- [151] The ATLAS Collaboration, G. Aad *et al.*, JINST **11**, P04008 (2016), 1512.01094, *Performance of b-Jet Identification in the ATLAS Experiment*, DOI: <https://doi.org/10.1088/1748-0221/11/04/P04008>.
- [152] The ATLAS Collaboration, G. Aad *et al.*, The European Physical Journal C **83** (2023), *Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton–proton collision data at $\sqrt{s} = 13\text{ TeV}$* , DOI: <https://doi.org/10.1140/epjc/s10052-023-11736-z>.
- [153] J. M. Meyer, *Measurement of the Top Quark Mass using Dilepton Events and a Neutrino Weighting Algorithm with the D0 Experiment at the Tevatron (Run II)*, PhD thesis, Bonn U., 2007.
- [154] The ATLAS Collaboration, M. Aaboud *et al.*, Eur. Phys. J. C **80**, 754 (2020), 1903.07570, *Measurements of top-quark pair spin correlations in the $e\mu$ channel at $\sqrt{s} = 13\text{ TeV}$ using pp collisions in the ATLAS detector*, DOI: <https://doi.org/10.1140/epjc/s10052-020-8181-6>.
- [155] L. Sonnenschein, Phys. Rev. D **72**, 095020 (2005), hep-ph/0510100, *Algebraic approach to solve $t\bar{t}$ dilepton equations*, DOI: <https://doi.org/10.1103/PhysRevD.72.095020>.
- [156] B. A. Betchart, R. Demina, and A. Harel, Nucl. Instrum. Meth. A **736**, 169 (2014), 1305.1878, *Analytic solutions for neutrino momenta in decay of top quarks*, DOI: <https://doi.org/10.1016/j.nima.2013.10.039>.
- [157] F. James and M. Winkler, (2004), *MINUIT User’s Guide*, <https://inspirehep.net/files/c92c2ba4dac7c0a665cce687fb19b29c>.
- [158] I. Antcheva *et al.*, Comput. Phys. Commun. **180**, 2499 (2009), 1508.07749, *ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization*, DOI: <https://doi.org/10.1016/j.cpc.2009.08.005>.
- [159] A. Galántai, Journal of Computational and Applied Mathematics **124**, 25 (2000), *The theory of Newton’s method*, DOI: [https://doi.org/https://doi.org/10.1016/S0377-0427\(00\)00435-0](https://doi.org/https://doi.org/10.1016/S0377-0427(00)00435-0), Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

- [160] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986), 1986, DOI: <https://doi.org/10.1201/9781315140919>.
- [161] D. W. Scott, *Biometrika* **66**, 605 (1979), *On Optimal and Data-Based Histograms*, DOI: <https://doi.org/10.1093/biomet/66.3.605>.
- [162] H. Hotelling, *Journal of Educational Psychology* **24**, 417 (1933), *Analysis of a Complex of Statistical Variables into Principal Components*, DOI: <https://doi.org/10.1037/h0071325>.
- [163] The ATLAS Collaboration, *Phys. Lett. B* **789**, 508 (2019), 1808.09054, *Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \rightarrow WW^* \rightarrow e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, DOI: <https://doi.org/10.1016/j.physletb.2018.11.064>.
- [164] The CMS Collaboration, *Phys. Lett. B* **791**, 96 (2019), 1806.05246, *Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV*, DOI: <https://doi.org/10.1016/j.physletb.2018.12.073>.
- [165] F. Fabbri, J. Howarth, and T. Maurin, *Eur. Phys. J. C* **84**, 20 (2024), 2307.13783, *Isolating semi-leptonic $H \rightarrow WW^*$ decays for Bell inequality tests*, DOI: <https://doi.org/10.1140/epjc/s10052-023-12371-4>.
- [166] The ATLAS Collaboration, *Phys. Lett. B* **810**, 135797 (2020), 2006.13076, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, DOI: <https://doi.org/10.1016/j.physletb.2020.135797>.
- [167] J. Alwall *et al.*, *JHEP* **07**, 079 (2014), 1405.0301, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, DOI: [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [168] The Sherpa Collaboration, *SciPost Phys.* **7**, 034 (2019), 1905.09127, *Event Generation with Sherpa 2.2*, DOI: <https://doi.org/10.21468/SciPostPhys.7.3.034>.
- [169] The ATLAS Collaboration, *Eur. Phys. J. C* **80**, 47 (2020), 1909.00761, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, DOI: <https://doi.org/10.1140/epjc/s10052-019-7500-2>.

- [170] The ATLAS Collaboration, JINST **15**, P09015 (2020), 2004.13447, *Performance of the ATLAS muon triggers in Run 2*, DOI: <https://doi.org/10.1088/1748-0221/15/09/p09015>.
- [171] The ATLAS Collaboration, JINST **14**, P12006 (2019), 1908.00005, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, DOI: <https://doi.org/10.1088/1748-0221/14/12/P12006>.
- [172] The ATLAS Collaboration, Eur. Phys. J. C **76**, 292 (2016), 1603.05598, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, DOI: <https://doi.org/10.1140/epjc/s10052-016-4120-y>.
- [173] The ATLAS Collaboration, G. Aad *et al.*, (2023), 2303.17312, *New techniques for jet calibration with the ATLAS detector*, DOI: <https://doi.org/10.1140/epjc/s10052-023-11837-9>.
- [174] G. D’Agostini, (2010), 1010.0632, *Improved iterative Bayesian unfolding*, DOI: <https://doi.org/10.48550/arXiv.1010.0632>.
- [175] L. Brenner *et al.*, Int. J. Mod. Phys. A **35**, 2050145 (2020), 1910.14654, *Comparison of unfolding methods using RooFitUnfold*, DOI: <https://doi.org/10.1142/S0217751X20501456>.
- [176] T. Gehrmann, S. Hoche, F. Krauss, M. Schonherr, and F. Siegert, JHEP **01**, 144 (2013), 1207.5031, *NLO QCD matrix elements + parton showers in $e^+e^- \rightarrow$ hadrons*, DOI: [https://doi.org/10.1007/JHEP01\(2013\)144](https://doi.org/10.1007/JHEP01(2013)144).
- [177] The ATLAS Collaboration, Report ATL-PHYS-PUB-2013-017, 2013, *Example ATLAS tunes of Pythia8, Pythia6 and Powheg to an observable sensitive to Z boson transverse momentum*, <https://cds.cern.ch/record/1629317>.
- [178] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, JHEP **03**, 015 (2013), 1212.3460, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, DOI: [https://doi.org/10.1007/JHEP03\(2013\)015](https://doi.org/10.1007/JHEP03(2013)015).
- [179] K. Hamilton, P. Nason, and G. Zanderighi, JHEP **10**, 155 (2012), 1206.3572, *MINLO: Multi-Scale Improved NLO*, DOI: [https://doi.org/10.1007/JHEP10\(2012\)155](https://doi.org/10.1007/JHEP10(2012)155).

- [180] K. Hamilton, P. Nason, C. Oleari, and G. Zanderighi, *JHEP* **05**, 082 (2013), 1212.4504, *Merging $H/W/Z + 0$ and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, DOI: [https://doi.org/10.1007/JHEP05\(2013\)082](https://doi.org/10.1007/JHEP05(2013)082).
- [181] The ATLAS Collaboration, G. Aad *et al.*, *JINST* **18**, T11004 (2023), 2211.16178, *Tools for estimating fake/non-prompt lepton backgrounds with the ATLAS detector at the LHC*, DOI: <https://doi.org/10.1088/1748-0221/18/11/T11004>.
- [182] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, (2017), 1709.07871, *FiLM: Visual Reasoning with a General Conditioning Layer*, DOI: <https://doi.org/10.48550/arXiv.1709.07871>.
- [183] T. Chen and C. Guestrin, *Xgboost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 785–794, ACM, 2016, DOI: <https://doi.org/10.1145/2939672.2939785>.
- [184] M. Abadi *et al.*, (2015), *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from <https://www.tensorflow.org/>.
- [185] DeepMind, *Sonnet*, <https://github.com/google-deeppmind/sonnet>, 2019, TensorFlow-based neural network library.
- [186] ONNX Community, *Open neural network exchange (onnx)*, <https://onnx.ai/>, 2019.
- [187] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Journal of Machine Learning Research* **15**, 1929 (2014), *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, DOI: <https://doi.org/10.9734/ajrcos/2025/v18i2569>.
- [188] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, *Optuna: A next-generation hyperparameter optimization framework*, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, DOI: <https://doi.org/10.48550/arXiv.1907.10902>.
- [189] B. W. Matthews, *Biochimica et Biophysica Acta (BBA)-Protein Structure* **405**, 442 (1975), *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*, DOI: [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9).

- [190] The ATLAS Collaboration, G. Aad *et al.*, Eur. Phys. J. C **81**, 689 (2021), 2007.02645, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, DOI: <https://doi.org/10.1140/epjc/s10052-021-09402-3>.
- [191] J. Butterworth *et al.*, J. Phys. G **43**, 023001 (2016), 1510.03865, *PDF4LHC recommendations for LHC Run II*, DOI: <https://doi.org/10.1088/0954-3899/43/2/023001>.
- [192] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C **71**, 1554 (2011), 1007.1727, *Asymptotic formulae for likelihood-based tests of new physics*, DOI: <https://doi.org/10.1140/epjc/s10052-011-1554-0>, [Erratum: Eur.Phys.J.C **73**, 2501 (2013)].
- [193] The ATLAS Collaboration, *TRExFitter*, DOI: <https://doi.org/10.5281/zenodo.14845712>.
- [194] A. N. Kolmogorov, G. Ist. Ital. Attuari (1933), *Sulla determinazione empirica di una legge di distribuzione*.
- [195] N. V. Smirnov, Matematicheskii Sbornik (1939), *On the estimation of the discrepancy between empirical curves of distribution*.

List of figures

1.1	Shape of the Higgs potential in the simple case where we define $\phi(x) = \phi_1(x) + i\phi_2(x)$ and $\mu > 0$. The central bump is centred at $\phi = 0$. The values of the potential $V(\phi)$ associated with a colour map, and the fields $\phi_1(x)$ and $\phi_2(x)$, are in arbitrary units.	7
1.2	The elementary particles of the SM [21]. The coloured squares, displayed below each fermion, represent the gauge boson fields to which it couples. The grey square and coloured circles represent the gluon field. The two blue squares represents the W boson fields, the green square the Z boson field, and the yellow square the photon field.	12
1.3	Cross section for different Higgs production processes in proton-proton collision against centre-of-mass energy [23].	16
2.1	Schematic of the accelerators and experiments in CERN [70].	28
2.2	Plots showing the luminosity-weighted distribution of the mean pile-up parameter per bunch for physics data, from 2015 to 2018 (Run 2) at 13 TeV (left) and from 2022 to the 17th of September of 2025 (Partial Run 3) at 13.6 TeV (right) [77, 78]. The mean number of interactions per crossing averaged over all bunch pairs and the Most Probably Value (MPV) are shown for each year.	31
2.3	Integrated luminosity in ATLAS [77, 78].	31
2.4	Cut-away view of the ATLAS detector in the latest Run 3 configuration with labelled sub-systems [80].	33
2.5	Diagram of the path of particles within the transverse plane as they travel through the different subdetectors [81].	34

2.6	Diagram of the ATLAS superconducting magnet system taken from [82].	35
2.7	The ATLAS ID [80].	35
2.8	The subdetectors composing the ID [83].	36
2.9	The calorimeters in ATLAS [80].	38
2.10	The track parameters as a perigee representation. The reference point can be seen as the origin where the axis vectors e_i meet. [90].	43
2.11	The shape of anti- k_t jets with $R = 1$ in the y - ϕ plane, formed from the particles in a simulated event [96]. Hard jets are very circular while softer jets have more complex shapes. The pair of jets around $y = 2$ and $\phi = 5$ provides an interesting example, where the purple jet on the left is much softer than the green jet on the right.	45
2.12	Illustration of a simulated $pp \rightarrow t\bar{t}$ MC event with parton showering where the hadronisation is simulated with the Lund string model. Several simplifications are made, not all decays are shown and particle kinematics are treated symbolically. Taken from the PYTHIA8.3 documentation [114].	50
3.1	Comparison of the DL1R and GN1 flavour tagging schemes [132]. Previously, the jet and associated tracks information was passed into a set of trained or manually optimised algorithms that generates inputs for the final neural network. For GNNs, the intermediary set of algorithms is not needed and is instead replaced by auxiliary training targets, shown as arrows leaving GN1. The dashed arrows represent truth information used in training while the solid arrows indicates reconstructed information. . .	55
3.2	Illustration of the GN2 network architecture and the jet structure features used as input [133]. The jet features are duplicated for each track associated with the jet: n_{jf} and n_{tf} correspond to the number of jet and track features respectively. The combined vectors are then fed into an encoding network that outputs an initial latent representation for each track, which is then followed by a transformer encoder returning a global representation of the jet. Finally that jet representation is pooled and its information is fed into task-specific networks: one that predicts the jet flavour and one per auxiliary task.	55

3.3	ROC curves, evaluated on a $t\bar{t}$ MC sample, displaying the flavour tagging performance of the DL1D, GN1 and GN2 taggers [137].	58
3.4	b -tagging background rejection at the 70% WP in $t\bar{t}$ for algorithms developed by the ATLAS flavour tagging group since 2017 [137]. The improvements are quantified as multiplicative factors relative to the c - and light-jet rejection rates of the DL1 tagger.	59
3.5	The seven PCFT quantiles, labelled 0 to 6. On each region is displayed the percentage of b -, c -, light- and τ - jets that contribute to its population. The percentage shown on both score axes correspond to the associated b - or c -jet efficiency obtained by summing from the tightest to the loosest regions. For example, the 77% b -jet efficiency is obtained by considering the union of quantiles 4, 5 and 6. The illustration is a courtesy of the flavour tagging group documentation.	61
3.6	The DL1DV01 tagger's b -jet efficiency at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.	62
3.7	The DL1DV01 tagger's c -jet mistag rate at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.	63
3.8	The DL1DV01 tagger's l -jet mistag rate at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.	63
3.9	The DL1DV01 tagger's τ -jet mistag rate at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.	64
3.10	The GN2V01 tagger's b -jet efficiency (top) and c -jet mistag rate (bottom) at $\epsilon_b = 77\%$ for different generators against p_T with $\eta \in [-2.5, 2.5]$. The black bins represent the nominal. The ratio plot effectively shows the SFs.	65
3.11	The heavy hadron p_T divided by the total p_T of the jet for jets taken in $t\bar{t}$ events decaying hadronically. The light-jets histogram has no entries in any bin since only heavy flavour hadrons are considered. These events are generated with POWHEG+PYTHIA8.	68

3.12	The frequency of each heavy hadron type present in jets taken in $t\bar{t}$ events decaying hadronically. The light-jets histogram has no entries in any bin since only heavy flavour hadrons are considered. These events are generated with POWHEG+PYTHIA8	68
3.13	Pictogram displaying the concept of signed IP. The jet is represented as a green cone, a black arrow at its center defines the jet axis. The plane perpendicular to the jet axis and passing through the PV (dashed line) defines the sign attributed to the IP. The blue and red curves represent tracks and the double ended arrows points to their respective point of closest approach.	69
3.14	Distributions of the signed d_0 and z_0 IPs of tracks matched to b -, c - and light-jets in simulated $t\bar{t}$ events at reco-level. The events are generated with POWHEG+PYTHIA8 [152].	70
3.15	Truth-level transverse component of the IP, signed with 2D information (second plot) and 3D information (third plot). The events are $t\bar{t}$ events decaying all-hadronically generated by POWHEG+PYTHIA8.	71
3.16	The population of each hadron type present in jets taken in $t\bar{t}$ events decaying to a single positive lepton for different SHERPA versions sample. At the top is SHERPA 2.2.1, below are 2.2.10 and 2.2.12.	72
4.1	The NW function for a single dileptonic $t\bar{t}$ event (a), parametrised by the two neutrino pseudo-rapidities, and for a single semi-leptonic HWW^* event (b), parametrised by the neutrino pseudo-rapidity and the mass of the leptonic W . All the unphysical complex-valued solutions are set to -1 , resulting in the dark-blue flat surface “below” the shape of interest. The shape of the NW function can change drastically between different events in both cases.	79
4.2	The NW function for a dileptonic $t\bar{t}$ event (a), parametrised by the two neutrino pseudo-rapidities, and for a semi-leptonic HWW^* event (b), parametrised by the neutrino pseudo-rapidity and the mass of the leptonic W . The MIGRAD starting and end point are given as well as the path taken by the algorithm. A naive gradient descent is also shown in purple to compare with the MINUIT path.	85

4.3	Violin plots displaying the distribution for the distance s_W between the solution of different algorithm pairs. The red line shows the mean of the distribution while the blue line highlights its median	90
4.4	Distribution of algorithm run times for $t\bar{t}$ events with a NW solution. . .	91
4.5	Distribution of the best top quark mass found. The algorithms without mass smearing all have $m_{t_1} = 172.0$ GeV exactly, which is why the blue histogram is not visible, hidden by the green distribution.	92
4.6	Violin plots displaying the distribution for the distance s_W between the solution of different algorithm pairs running on ggF events. The red line shows the mean of the distribution while the blue line highlights its median.	94
4.7	Violin plots displaying the distribution for the distance s_W between the solution of different algorithm pair running on VBF events. The red line shows the mean of the distribution while the blue line highlights its median	95
4.8	Distribution of algorithm run times for ggF events with a NW solution. .	96
4.9	Distribution of the output leptonic W mass for ggF events.	97
4.10	Distribution of the best Higgs mass found for ggF events. Algorithms without mass smearing have every event with $m_h = 125.0$ GeV, which is why the blue histogram is not visible, hidden behind the green distribution.	99
5.1	Signal purity as a function of the requirement applied on the NW weight to be greater than some value. Only events passing the pre-selection and the c -tagging selection are considered.	106
5.2	Distribution of $ \Delta\phi_{ls} $, the absolute value of the difference in azimuthal angle between the light-jet identified as originating from the s -quark and the lepton. The angle difference is divided by π to bound the distribution between 0 and 1.	107
5.3	Migration matrices for the observables O_{zx}^1 (a), O_{zx}^2 (b) and O_{zx}^3 (c). . .	111
5.4	Unfolded histograms for the three observables composing the expectation value of the zx Bell Operator. The uncertainties on each bin are statistical.	112
5.5	The difference between the ATLAS reco-level $\vec{E}_T^{\text{miss}}$ and the parton-level neutrino momentum along the x-axis for signal events.	114

5.6	Expected yields when simulating $\vec{E}_T^{\text{miss}}$ resolution for 300 fb^{-1}	115
6.1	Invariant masses computed from the particle level objects obtained with the RIVET routine. The Old MC compared here is the $l\nu qq$ sample with a 20 GeV lepton p_T filter. The new MC sample is shown with and without a 20 GeV lepton p_T cut for comparison.	119
6.2	Fake charged lepton efficiencies and their statistical uncertainties in Run 2, binned in p_T and η , for the 3 jets region with b -veto.	128
6.3	Real charged lepton efficiencies and their statistical uncertainties in Run 2, binned in p_T and η	129
6.4	Charged lepton efficiencies, marginalised over η for electrons and muons, for Run 2.	130
6.5	Charged lepton efficiencies, marginalised over η for electrons and muons, for Run 3.	130
6.6	Illustration of the JetPairingNet neural network.	134
6.7	Loss against epochs during the training of the JetPairingNet on the even fold.	135
6.8	Accuracy against epoch during the training of the JetPairingNet on the even fold. The black dashed horizontal line shows the accuracy of the naive algorithm relying on c -tagging and the invariant mass of combination of jets. The gray dashed horizontal line displays the accuracy of guessing at random.	136
6.9	ROC curve for the no-pair score of the JetPairingNet model trained on the even fold.	137
6.10	Distribution of the parton-level mass of the hadronic W , for reconstructable events only. The stacked histogram separates events that have both the parton-level W daughters ΔR match reco-level jets (reconstructed events) and events that do not (not reconstructed events).	138
6.11	Distribution of the parton-level mass of the hadronic W , for reconstructable events only. The same information as Figure 6.10 is presented but for events that do not have c -jet filtering.	139

6.12	Feature importance (a) computed using the average gain from splitting with a given observable, and correlation matrix (b) computed from the signal training dataset (ggF + VBF).	145
6.13	Shape plots of the BDT inputs computed using the kinematics of j_1 and j_2 only.	146
6.14	Shape plots of the BDT inputs computed using the kinematics of j_1 or j_2 , combined with the lepton.	148
6.15	Shape plots of the BDT inputs computed using the missing transverse energy, or related physical quantities.	151
6.16	Learning curves corresponding to a single training fold in each region. The rows, from top to bottom, show the BDTs trained for the 2-, 3-, and 4-jets regions, respectively.	153
6.17	ROC curves corresponding to a single training fold in each region. The rows, from top to bottom, show the BDTs trained for the 2-, 3-, and 4-jets regions, respectively.	154
6.18	Output score distribution for the 3 jets region.	155
6.19	Histograms representing the variations from the nominal associated with a systematic. The black histogram is the nominal, the red and blue histograms represent the up and down variations respectively. The dashed lines represent the variation template before the symmetrisation and smoothing procedures.	162
6.20	The SRs used in the fit. The regions are obtained by applying a b -veto. Each column corresponds to the number of jets allowed in the region (2, 3 and 4), and the electron and muon channels are shown on the top and bottom row respectively.	163
6.21	The CRs used in the fit. The regions require at least 1 b -jet in the event, and are split per number of jets (2,3 and 4).	164
6.22	The normalised histograms of the ggF signal and the combined backgrounds for each SR.	165
6.23	Post-fit pulls of the systematic NPs.	166

6.24	Post-fit pulls of a fit with only three jet systematics. The systematics are decorrelated per physical process.	167
6.25	Scan of the negative log-likelihood over the POI.	168
6.26	Correlation matrix for groups of systematic uncertainties and the NFs of the physical processes.	171
6.27	Illustration of the CLs limits obtained on the signal strength.	172
6.28	Projection for the significance under three distinct scenarios concerning the systematic uncertainties (a), and the CLs limits in the third scenario (b).	173
6.29	Distribution of the missing transverse energy for electrons (a) and muons (b).	174

List of tables

3.1	Input features to GN1 and GN2 [132].	56
3.2	Truth origins which are used to categorise the physics process that led to the production of a track [132].	57
4.1	Metrics computed using roughly 890,000 $t\bar{t}$ events in total. The mean duration and best weight are computed by taking into account successfully reconstructed events only.	90
4.2	Metrics computed using roughly 200 thousands events in total, for both ggF and VBF. The mean duration and best weight are computed by taking into account successfully reconstructed events only.	95
4.3	The best NW weight averaged over successfully reconstructed events for the ggF and VBF samples after extending the grid upper limit for the m_{W_ℓ} parameter from 50 GeV to 90 GeV.	97
4.4	The best NW weight averaged over successfully reconstructed events, and the fraction of reconstructed events for the ggF and VBF samples after increasing the Higgs mass smearing from 1 GeV to 10 GeV.	99
5.1	c - and b -tagging efficiencies of the DL1R tagger used in the ATLAS experiment [136, 145]. The numbers refer to the probability for the tagging algorithms to assign the c or b flavour to jets containing a B -hadron (b -jet), a c -hadron and no B -hadrons (c -jet) or no heavy-flavour hadrons (l -jet). These numbers are directly used in the analysis presented in this paper to emulate realistic c -tagging performances.	104

5.2	The expected yields for an integrated luminosity of 300 fb^{-1} in a selection with idealised charm tagging and Run2-like charm tagging. The purity ($S/(S+B)$) is also shown. The uncertainties are statistical only.	108
5.3	The expected sensitivity for different values of integrated luminosity for a realistic and idealised case. The uncertainties are statistical only.	113
6.1	Summary of the electron and muon working points used in Run 2 and Run 3. The information is presented as “identification + isolation”. . . .	124
6.2	Uncertainty on the signal strength for different groups of systematics. The total is computed by adding each contribution in quadrature.	170