Searches for dark matter in di-Higgs final states with the ATLAS experiment and measurements of the Higgs potential with multi-Higgs at the FCC





Thomas Benjamin Dingley

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2026

In loving memory of Leonard Dingley and Marilyn Stephens.

Acknowledgements

First, to my supervisors, without whom the research presented in this thesis would not have been possible. Thank you both for being so consistently engaged in our weekly meetings; I have learnt a tremendous amount from your questions alone. To Dr James Frost, for your expansive knowledge, persistence to understand and willingness to explain. I have never met, and may never meet, a person with such attention to detail. It is a great shame our fated badminton clash never came to fruition. I wish you and the family all the very best! To Prof. Todd Huffman, thank you for always capturing the wider context, the vibrant atmosphere you encourage, your many stories, your wacky (and not-so-wacky) ideas, and for the many shared pints. To Holly, thank you for your HHH, $p_{\mathrm{T}}^{\mathrm{miss}}$ and ATLAS know-how, your career advice and for bringing the comically large coffee mug, so symbolic of a physicist's needs, to so many of our weekly meetings. And finally, a very special thank you to Tim, with whom I collaborated closely. Thank you for your patience when I knew very little, and your seemingly limitless enthusiasm.

I've had the pleasure of being a part of not one but two international collaborations, both of which are home to an enormous array of brilliant scientists. Firstly, to Francesco Armando Di Bello, Bing, Dan, Valentina and Angela, for the many insightful FTAG discussions and your calm leadership. Thank you to Brian Moser for the many fun chats about physics and life. Our walks in Genoa and Cinque Terre will hold a fond place in my DPhil memory. My thanks extend to the FCC-hh physics group, especially Birgit for dealing with my constant questions, and to Michele for your leadership and support.

Thank you to all of my colleagues and friends in the Denys Wilkinson Building. The building may appear an abomination from the outside, and wildly impractical from the inside, but it is home to a number of people that I will miss dearly. To Ynyr, Eimear and Zhiying, for welcoming us with open arms to Office 656; to Kim for your unwavering support and kindness, without you this department would surely steer astray; to Yuval, for being the most delightful officemate in Oxford and at CERN; to Alessandro, although likely delaying my thesis by several months - thank you for the endless yaps, whenever I see a horsebox I will think of you; to Simon, for our lengthy coffee walks around Uni Parks and for the many morning pancakes in Madison; to Baris, for our extended rivalry; to Bill, for our shared love of sweet treats; to Malek, for our peaceful (and not-so-peaceful) bike rides, and for introducing me to birdwatching; to Ida, for our many lovely conversations,

introducing me to the art of the cinnamon bun, and for attempting to teach me Swedish: jag hoppas vi ses igen — nästa gång i Sverige! My thanks to the JET collaboration: to Yangfan, a kindred statistical spirit, for all the chats and for the informative Mandarin lessons - the next chocolate milk is on me; to Maoze, for your much-needed calm amongst the collaboration's chaos; and to George, the student who knew infinitely more than the master. My thanks to the Oxford Exotics group, particular Maggie for the warm welcome and to Ruben — good luck with the rest of your PhD! Unfortunately, there is not enough space to list everyone, but also my thanks to Vid, Anjelo, Ben H, Ben G, Lars, Gulliver, Jairus, Tom Y and Florian. And finally, even though you are quite unlikely to see this, thank you to Samir. Our brief interactions were often a much-needed source of joy.

To Charlie, for our numerous (and inconsistent) gym sessions, for our cycling adventures and for introducing me to Brill. Radek put it best, you truly are a ray of sunshine. To my friends and teammates in Oxford Badminton: Sugi, Adrian, Enson, Siam, Rash and to Hui Min, for keeping me company. To Luke, Aimee and Emily for the countless trips to Wagamama and our hiking adventures. To Sohan, Matt, Sam, Elliot and Adam, for the many reunions in the outbuilding and beyond.

Thank you to Manish and Cedric for introducing me to, and welcoming me into, the Sunday sessions and the Geneva Badminton Club; they were the highlight of my year in Switzerland. Thank you, Dharam, for your endless positivity, your infectious enjoyment of badminton, and, most importantly, for all the lifts! Thank you to Bu for organising the sessions, Alexei for your wisdom, AJ for your laughter, Sargar for your warm spirit, and Rene: un pour la rue? Thank you to my doubles partner, Ding – may our undefeated Swiss record continue for all of time. Thank you to Max, our many long conversations in 5 Rue Gilbert were a constant source of fun.

To Colin, for making me feel so very welcome in a time when I needed it the most. To Jakub for all the marvellous meals, to Dinesh, Leandro, Rebecca, Francesca, Chantal, Paul, Angie, and, of course, to Molly and Fluff. Who would've thought that becoming a regular at Café Nero would lead to such treasured souls.

Many of the previously mentioned outings, adventures and exciting encounters have been accompanied by my glorious Polish brother, Radek. The past few years have been remarkably special in many ways, and at times remarkably challenging, and I have had the privilege of experiencing all of it with you. Your appreciation for those who came before, your approach to people, physics and life have taught me a great deal. I will miss you dearly, and I look forward to seeing, now from a slightly more distant standpoint, what's in store for you next. Thank you, Boss[1].

Lastly, to my family. There really is too much to say, so I will keep it brief. To Jess, for becoming like a sister to me over the past few years. To Harry, for being my best friend. To Mum, for your constant love. And to Dad, for believing in me.

[1]P.S. I believe the last table tennis match was winner takes all, so I will graciously take it all.

Abstract

This thesis presents three interconnected strands of research that investigate the Higgs boson as both a central component of the Standard Model and a powerful probe of possible extensions to it.

The main analyses presented involve two Higgs bosons decaying into $b\bar{b}$ pairs, and are therefore strongly reliant on excellent flavour-tagging performance. The first research component contributes to this endeavour by calibrating two advanced machine learning algorithms, `DL1d` and `GN2`, for highly energetic b-jets, thereby enabling the substantially improved algorithms to be deployed across the physics analyses of the ATLAS collaboration.

The second analysis probes, for the first time at colliders, the resonant production of two SM Higgs bosons in association with a large amount of undetected momentum ($HH + p_{\mathrm{T}}^{\mathrm{miss}}$). The analysis targets the dark Higgs boson model and, using machine learning techniques tailored to the resonant topology, achieves strong sensitivity to new physics. Control regions are defined to target the dominant top and $Z+\mathrm{jets}$ background processes, and the statistical interpretation is performed using a signal region binned in m_{HH}. The signal region is consistent with the SM expectation, and limits are set on the dark Higgs model, which extend ATLAS' limits to high dark Higgs boson and Z' mediator masses.

The final component presented is a feasibility study for a high-energy future circular hadron collider (FCC-hh) to probe the experimentally challenging quartic Higgs self-coupling. The $HHH \to 4b2\tau$ final state is used, where the accurate mass reconstruction of two τ-leptons effectively reduces the $t\bar{t} + b\bar{b}$ background. A brief investigation of the complementary information provided by di-Higgs in the $b\bar{b}\gamma\gamma$ final state is presented. The statistical combination is performed to assess the experiment's ultimate sensitivity. The study finds limits of $\mathcal{O}(10)$ on the quartic Higgs self-coupling and constitutes the first study within the FCC collaboration and at the $\sqrt{s} = 84$ TeV benchmark.

Preface and Author Contributions

Introductory Material

Chapter 1 begins with an introduction to the Standard Model of particle physics, which underpins all components of the research presented later. Since the two analyses presented utilise interactions with the Higgs boson in various forms, particular attention is paid to the Higgs mechanism and to the electroweak transition. In chapter 2, the underlying structure of the SM is connected to the outcomes of the ultra-relativistic proton-proton collisions at the Large Hadron Collider (LHC). Particular focus is given to the dominant production modes of the Higgs boson and to the challenge of producing three SM Higgs bosons at the LHC, motivating the possible use of higher-energy Future Circular hadron Collider (FCC-hh) facilities to probe this process. Chapter 3 then delves into the known shortcomings of the SM and their connection with the Higgs sector. The connections to the Higgs sector motivate measurements of the Higgs potential, notably the quartic Higgs self-coupling, and searches for physics beyond the SM (BSM) in Higgs final states. Specifically, the resonant production of two SM Higgs bosons in association with missing transverse momentum. Here, we introduce the dark Higgs model, which is the benchmark used to probe the $HH+p_{\mathrm{T}}^{\mathrm{miss}}$ final state with the ATLAS experiment. Chapter 4 then introduces the LHC, describing the accelerator complex that feeds the four main interaction points along the beamline. The $HH+p_{\mathrm{T}}^{\mathrm{miss}}$ search uses the full amount of data collected during Run 2 of the LHC by the ATLAS detector. The ATLAS detector is described, along with the reconstruction of objects from the signals deposited following collisions. Chapter 5 introduces the statistical techniques which are used extensively in the three main research components: machine learning and statistical inference.

Flavour Tagging

Two flavour tagging algorithms, `DL1d` and `GN2`, are introduced in chapter 6. Particular attention is paid to the performance of these algorithms for highly energetic jets, for which the author calibrated each algorithm in this high-p_{T} phase space, a regime relevant to searches for BSM physics. This work contributed to the publication of Ref [1].

Hunting in the Dark with $HH + p_{\mathbf{T}}^{\mathbf{miss}}$

Chapter 7 then documents a search for a dark Higgs boson. The chapter summarises the full analysis, including signal and background kinematics, event reconstruction and preselection, machine learning, background estimation, statistical inference, subsequent results and an outlook towards the future. A previous student summarised a large portion of the analysis in his thesis [2], and was involved in the analysis setup, preselection, machine learning, and background estimation approach. The author of this thesis implemented and verified a full statistical fitting framework for the analysis, deriving, implementing and validating all theoretical and experimental systematics. Thus, he was responsible for validating and understanding all of the many iterations of analysis results, expectations for the impact of CRs and uncertainties, from the first simple fits to the final published results. During this time, he served as analysis contact, reporting on its status in the relevant ATLAS subgroup and on the numerous collaboration approval processes leading to publication [3]. All plots and results, unless explicitly stated otherwise, were produced by the author. Any plot included with ATLAS collaboration identification is within the paper main body or auxiliary material, all of which were made by the author.

Quartic Higgs self-coupling at the FCC

Chapter 8 documents a feasibility study to probe the quartic Higgs self-coupling using triple-Higgs ($4b2\tau$) events, and a small initial study, proposed and performed entirely by the author, which builds on previous feasibility studies, into the possibility of probing the quartic Higgs self-coupling by combining with di-Higgs measurements in the $b\bar{b}\gamma\gamma$ final state. The author performed all stages of the analysis, including: technical implementation of signal sample generation within the FCC software suite, definition of background samples, object identification optimisation of leptons and of truth association, event preselection, fully implemented and calibrated advanced di-τ mass reconstruction tool `MMC`, $HH \to 4b$ pairing, machine learning implementation, statistical analysis and presentation of results. The analysis is the first of its kind within the FCC software suite and at the updated $\sqrt{s} = 84$ TeV benchmark. A full statistical combination of the two analyses indicates the ultimate sensitivity of the proposed experiment to the quartic Higgs self-coupling. This work was included within the European Strategy Physics Briefing Book [4], and was documented within an FCC-hh note [5].

Conclusions

The thesis concludes in chapter 9 with a brief overview of the three research outputs and an outlook to the future.

Contents

1 The Standard Model **1**

1.1 Particles and Fields . . . 1
1.2 Gauge Invariance . . . 3
1.3 Quantum Electrodynamics . . . 4
1.4 Quantum Chromodynamics . . . 4
1.5 Weak Interactions . . . 6
1.6 Electroweak Unification . . . 6
1.7 Electroweak Symmetry Breaking . . . 7
1.7.1 The Higgs Mechanism applied to Electroweak Theory . . . 8
1.7.2 Fermion Masses . . . 10
1.7.3 Experimental Confirmation . . . 11
1.7.4 Evolution of the Electroweak Vacuum . . . 11

2 The Standard Model at Colliders **14**

2.1 Event Simulation . . . 14
2.1.1 Hard-scatter Event Generation . . . 14
2.1.2 Parton Shower . . . 16
2.1.3 Matching and Merging . . . 17
2.1.4 Hadronisation . . . 17
2.1.5 Underlying Event . . . 19
2.2 Detector Simulation . . . 19
2.3 Higgs Boson Production . . . 20

3 Beyond the Standard Model **23**

3.1 Limitations of the SM . . . 23
3.1.1 Dark Matter . . . 24
3.1.2 Theoretical Shortcomings . . . 27
3.1.3 Electroweak Baryogenesis . . . 27
3.1.4 Connections to the Higgs sector . . . 30
3.2 Quartic Higgs self-coupling . . . 30
3.3 Searching for Dark Matter . . . 33
3.3.1 Direct Detection . . . 33
3.3.2 Indirect Detection . . . 33
3.3.3 Collider Searches . . . 34
3.3.4 Di-Higgs + $p_{\mathrm{T}}^{\mathrm{miss}}$ Phenomenology . . . 35

4 The ATLAS Experiment **40**
4.1 The Large Hadron Collider . . . 40
4.1.1 Acceleration and Magnet Technology . . . 41
4.1.2 Interaction Regions . . . 42
4.1.3 Coordinate System . . . 43
4.1.4 Luminosity and Pile-up . . . 44
4.2 The ATLAS Detector . . . 45
4.2.1 Inner Detector . . . 47
4.2.2 Calorimeters . . . 50
4.2.3 Muon Spectrometers . . . 54
4.3 Objects in the ATLAS experiment . . . 55
4.3.1 Tracks . . . 55
4.3.2 Vertex Reconstruction . . . 56
4.3.3 Calorimeter Clusters . . . 56
4.3.4 Jets . . . 57
4.3.5 Jet Calibration . . . 59
4.3.6 Pile-up Mitigation . . . 61
4.3.7 *b*-jets . . . 61
4.3.8 Muons . . . 63
4.3.9 Missing Transverse Momentum . . . 63
4.3.10 ATLAS Trigger . . . 64

5 Machine Learning and Statistical Inference **68**
5.1 Classification . . . 69
5.1.1 Building the Model . . . 70
5.2 Machine Learning at Colliders . . . 70
5.2.1 The Perceptron . . . 71
5.2.2 The Multi-Layer Perceptron . . . 72
5.2.3 Graph Neural Networks and Transformers . . . 72
5.2.4 Model Outputs and Discriminants . . . 73
5.2.5 Training and Validation . . . 74
5.2.6 Network Interpretation . . . 75
5.3 Statistical Inference . . . 76
5.3.1 Building the Likelihood . . . 76
5.3.2 Profile-Likelihood Fit . . . 78
5.3.3 Quantifying Uncertainty . . . 79
5.3.4 Hypothesis Testing . . . 80

6 Flavour Tagging **84**
6.1 Flavour Tagging at ATLAS . . . 84
6.1.1 DL1 Series . . . 85
6.1.2 GN series . . . 87
6.1.3 Algorithm Training . . . 89
6.1.4 Deployment in Analysis . . . 89
6.1.5 Algorithm performance at high-p_{T} . . . 92
6.2 Algorithm Calibration . . . 94
6.2.1 b-tagging Efficiency . . . 94
6.2.2 Charm-mistag Efficiency . . . 95
6.2.3 Light-mistag Efficiency . . . 95
6.2.4 Extrapolation to High-p_{T} . . . 96
6.2.5 Usage in Analysis . . . 105
6.2.6 Conclusions . . . 106

7 Hunting in the Dark with HH+$p_{\mathbf{T}}^{\mathbf{miss}}$ **108**
7.1 Signal Models and Kinematics . . . 109
7.2 Background Processes and Event Generation . . . 112
7.3 Dataset and Triggers . . . 113
7.4 Analysis Objects . . . 113
7.5 Preselection . . . 116
7.5.1 Higgs Reconstruction . . . 119
7.6 Machine Learning . . . 120
7.6.1 Model Architecture . . . 120
7.6.2 Input Features . . . 121
7.6.3 Decorrelation . . . 124
7.7 Background Estimation . . . 126
7.7.1 Control Regions . . . 127
7.7.2 Signal Region . . . 129
7.8 Systematic uncertainties . . . 132
7.8.1 Experimental Uncertainties . . . 132
7.8.2 Theoretical Uncertainties . . . 132
7.8.3 Systematic Smoothing: single top . . . 133
7.9 Statistical Model . . . 134
7.10 Fit Validation . . . 135
7.10.1 Background-only Fit . . . 136
7.10.2 Exclusion Fit . . . 140
7.11 Unblinded Results . . . 144
7.11.1 Model-independent Fit and Results . . . 144
7.11.2 Model-dependent Results . . . 145
7.12 Conclusions . . . 149

8 Quartic Higgs self-coupling at the FCC **151**
8.1 Measurements of the Higgs Potential . . . 152
8.1.1 Limitations of the LHC . . . 153
8.2 The FCC Project . . . 155
8.2.1 FCC-ee . . . 155
8.2.2 FCC-hh . . . 156
8.3 Triple Higgs at the FCC-hh . . . 158
8.3.1 Signal Generation . . . 159
8.3.2 SM Triple-Higgs cross section . . . 162
8.4 The $4b2\tau$ final state and backgrounds . . . 164
8.4.1 Production Kinematics of Triple Higgs at the FCC . . . 165
8.4.2 Object Identification . . . 169
8.5 Event Reconstruction and Preselection . . . 172
8.6 Di-τ mass reconstruction with `MMC` . . . 175
8.6.1 The Missing Mass Calculator algorithm . . . 175
8.6.2 MMC Weighting . . . 178
8.6.3 Implementation and Results . . . 178
8.6.4 $m_{\tau\tau}^{\mathrm{MMC}}$ performance . . . 182
8.7 $HH \to 4b$ Pairing . . . 185
8.8 Event Preselection . . . 189
8.9 Preselection and Machine Learning . . . 191
8.9.1 Training Performance . . . 194
8.9.2 Robustness to (κ_3, κ_4) variations . . . 196
8.10 Signal region and Statistical Interpretation . . . 198
8.10.1 Statistical Analysis . . . 200
8.10.2 Systematic Uncertainties . . . 201
8.10.3 Fit Behaviour . . . 202
8.11 Results . . . 204
8.12 Di-Higgs at the FCC-hh . . . 207
8.13 Combination . . . 213
8.14 A view to the future . . . 218
8.15 Conclusions . . . 218

9 Thesis Summary and Outlook **220**

References **222**

1
The Standard Model

Contents

1.1 Particles and Fields 1
1.2 Gauge Invariance 3
1.3 Quantum Electrodynamics 4
1.4 Quantum Chromodynamics 4
1.5 Weak Interactions 6
1.6 Electroweak Unification 6
1.7 Electroweak Symmetry Breaking 7
1.7.1 The Higgs Mechanism applied to Electroweak Theory . 8
1.7.2 Fermion Masses 10
1.7.3 Experimental Confirmation 11
1.7.4 Evolution of the Electroweak Vacuum 11

1.1 Particles and Fields

The Standard Model (SM) of particle physics is a theoretical framework describing the known fundamental particles and their interactions. Formulated in the latter half of the 20th century, it describes the electromagnetic, weak and strong forces through the language of Quantum Field Theory (QFT). Its underlying mathematical structure is governed by the local gauge group:

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \tag{1.1}$$

where $\mathrm{SU}(3)_C$ represents Quantum Chromodynamics (QCD) governing the strong interaction, and $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ represents the unified Electroweak (EW) sector.

A desirable property for a physical theory is consistency across space, time, and reference frame. To satisfy this, particles are defined as states that transform under irreducible unitary representations of the Poincaré group — encompassing

translations, rotations, and Lorentz boosts. The particles of the SM, separated into the quarks, leptons and gauge bosons, are shown in Table 1.1, alongside their spin, charge and colour representation under SU(3) [6].

Particle	**Symbol**	**Spin**	**Charge (Q)**	**Color Rep.**
		Quarks		
Up	u	1/2	$+2/3$	3
Down	d	1/2	$-1/3$	3
Charm	c	1/2	$+2/3$	3
Strange	s	1/2	$-1/3$	3
Top	t	1/2	$+2/3$	3
Bottom	b	1/2	$-1/3$	3
		Leptons		
Electron	e	1/2	-1	1
e-neutrino	ν_e	1/2	0	1
Muon	μ	1/2	-1	1
μ-neutrino	ν_μ	1/2	0	1
Tau	τ	1/2	-1	1
τ-neutrino	ν_τ	1/2	0	1
		Gauge Bosons		
Photon	γ	1	0	1
Gluon	g	1	0	8
W boson	$W^\pm$	1	± 1	1
Z boson	Z	1	0	1
		Scalar Boson		
Higgs boson	H	0	0	1

Table 1.1: The fundamental particle content of the Standard Model, organised by spin, charge, colour representation and generation. Colour representations denote how particles transform under $\mathrm{SU}(3)_C$: singlets, triplets or octets.

The dynamics of the SM fields are governed by the principle of least action, $\delta S = 0,\ S = \int \mathcal{L}_{\mathrm{SM}}\, d^4x$, and this chapter will briefly introduce the ingredients of the SM Lagrangian, $\mathcal{L}_{\mathrm{SM}}$; together with a more detailed account of spontaneous

symmetry breaking in the EW sector, giving rise to the Higgs boson of the SM.

1.2 Gauge Invariance

The allowed terms of the SM Lagrangian are constrained by requiring invariance under the local gauge group introduced above. For illustration, consider a fermion field, ψ, charged under $U(1)_Y$. Its free kinetic term is

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi. \tag{1.2}$$

This is invariant under a global phase transformation,

$$\psi \mapsto \psi' = e^{ig'\frac{Y}{2}\alpha}\psi, \tag{1.3}$$

where α is constant.

However, under a local transformation,

$$\psi(x) \mapsto \psi' = e^{ig'\frac{Y}{2}\alpha(x)}\psi, \tag{1.4}$$

the derivative produces an additional term proportional to $\partial_\mu\alpha(x)$, so the free kinetic term is no longer invariant. Local gauge invariance is restored by introducing the gauge field B_μ, and replacing ∂_μ by the covariant derivative,

$$\partial_\mu \mapsto D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu. \tag{1.5}$$

Provided the gauge field transforms as

$$B_\mu \mapsto B'_\mu = B_\mu - \partial_\mu\alpha(x), \tag{1.6}$$

the covariant derivative transforms in the same way as the field itself,

$$D_\mu\psi \mapsto e^{ig'\frac{Y}{2}\alpha(x)}D_\mu\psi, \tag{1.7}$$

and therefore

$$i\bar{\psi}\gamma^\mu D_\mu\psi \tag{1.8}$$

is invariant under local $U(1)_Y$ gauge transformations. By considering a free fermionic Lagrangian and imposing local gauge invariance, we require a new field B_μ which interacts with the fermion field ψ [7].

Gauge invariance is of fundamental importance, not only in its ability to explain existing experimental results, but also due to its predictive power — as will be seen later for neutral currents and the electroweak symmetry breaking mechanism. Contextualising experimental results within a framework that satisfies gauge invariance led to the development of the SM Lagrangian, beginning with Quantum Electrodynamics (QED).

1.3 Quantum Electrodynamics

The approach introduced above, when applied to the Dirac equation, gives rise to the massless photon and QED, with Lagrangian density,

$$\mathcal{L}_{\text{QED}} = \sum_i \bar{\psi}_i \left(i\gamma^\mu D_\mu - m_i\right) \psi_i - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.9}$$

where i runs over all charged quarks and leptons in the SM of mass m_i, the covariant derivative is given by $D_\mu = \partial_\mu + ieQA_\mu$, where the photon field, A_μ, is introduced to restore gauge invariance, and its field strength tensor is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The QED lagrangian describes the behaviour and interactions of electrically charged objects in the SM, and gives rise to the Maxwell equations of electrodynamics.

1.4 Quantum Chromodynamics

At hadron colliders, the proton initial states, comprised of coloured constituent quarks and gluons, collide and predominantly interact by the strong interaction. In much the same way that QED describes the behaviour of charged particles within the SM, the strong interaction, or, quantum chromodynamics (QCD) describes the dynamics of coloured particles. Developed as a non-Abelian gauge theory [8], QCD is comprised of three colours within an SU(3) gauge group, and the eight generators of SU(3) correspond to massless coloured spin-1 bosons known as gluons. There are six flavours of electrically charged coloured fields, known as quarks (q).

The Lagrangian is derived analogously to QED, with the principle of gauge invariance applied to the free Lagrangian:

$$\mathcal{L} = \bar{q}(i\gamma^\mu \partial_\mu - m)q \tag{1.10}$$

with local gauge invariance in SU(3) requiring,

$$q(x) \mapsto q'(x) = e^{i\theta^a(x)T^a} q(x) \tag{1.11}$$

where $a = 1, ..., 8$ label the generators $T^a = \lambda^a/2$, with the Gell-Mann matrices λ^a that span the Lie algebra of SU(3) and $\theta^a(x)$ is the local gauge transformation.

Together with the covariant derivative,

$$D_\mu = \partial_\mu + ig_s T^a G^a_\mu \tag{1.12}$$

gives the Lagrangian density,

$$\mathcal{L}_{\text{QCD}} = \sum_f \bar{q}_f \left(i\gamma^\mu D_\mu - m_f\right) q_f - \frac{1}{4} G^{a\mu\nu} G^a_{\mu\nu} \tag{1.13}$$

where f runs over the six flavours of quark and $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu$.

The last term in this expression encodes gluon self-interactions, and has important consequences for the strength of the strong coupling constant ($\alpha_s = \frac{g_s^2}{4\pi}$) at different energy scales. In contrast to QED, where the charge is screened at large distances, an anti-screening mechanism occurs in QCD which enhances the coupling strength at large distances [9]. Accordingly, quarks produced at high energy (small distances) following a hadron-hadron collision behave as asymptotically free states, with $\alpha_s \ll 1$. However, at low energy (large distances) QCD becomes strongly coupled, and quarks are confined to form colour-neutral bound states, known as hadrons. The behaviour of α_s as a function of scale is known as the running of the strong coupling constant, and the general dependence can be seen with the leading-order solution to the renormalisation group equation for α_s,

$$\alpha_s(Q^2) \simeq \frac{4\pi}{\beta_0 \ln(Q^2/\Lambda^2_{\text{QCD}})} \tag{1.14}$$

where Q is the scale of the interaction, $\beta_0 = 11 - \frac{2}{3} n_f$ with n_f set to the number of flavours included in the calculation and Λ_{QCD} is the scale for which QCD becomes strongly coupled with $\lim_{Q \to \Lambda_{\text{QCD}}} (\alpha_s) \to \infty$. The behaviour of QCD between these two scales is crucial to hadron-collider physics, as will be described later.

1.5 Weak Interactions

The weak nuclear force is the only interaction within the SM that can change particle flavour, i.e. an up-quark to down-quark transition via an intermediate charged vector boson, the W boson ($u \to dW^+$). In the SM, flavour labels the different species of quarks and leptons. Corresponding fermions in different generations have otherwise identical transformation properties under the SM gauge group, but differ in their mass, generated by Yukawa couplings as will be introduced in Section 1.7.2. Governed by the $\mathrm{SU}(2)_L$ gauge group, it is comprised of charged and neutral-current interactions mediated by the W^+, W^- and the Z bosons. Originally formulated as a vector low-energy effective field theory by Fermi to describe β-decay [10], the weak interaction was guided by experimental results throughout the 20th century. Notably, the experiment by C.S. Wu [11] showing that parity is maximally violated in weak decays guided its development into a $V - A$ theory [12]. The weak interaction is *chiral*, meaning that it couples left-handed fermions to right-handed anti-fermions.

1.6 Electroweak Unification

To accommodate bosons into the weak interaction, Glashow postulated that electromagnetism and the weak interaction are low-energy manifestations of the same gauge group: $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ [13]. This gauge group then contains four gauge bosons $(W^1_\mu, W^2_\mu, W^3_\mu, B_\mu)$, with couplings g and g', and electric charge now written as $Q = T_3 + \frac{1}{2}Y$, where T_3 is the weak isospin and Y the weak hypercharge.

The electroweak Lagrangian [14] is written as,

$$\mathcal{L} = \sum_\psi \bar{\psi} i\gamma^\mu D_\mu \psi - \frac{1}{4} W^{i\mu\nu} W^i_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{1.15}$$

where $i = 1, 2, 3$ for the three $\mathrm{SU}(2)_L$ gauge bosons, with field strength tensors,

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon^{ijk} W^j_\mu W^k_\nu, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \tag{1.16}$$

Interactions between the gauge bosons and fermionic matter are incorporated via the covariant derivative,

$$D_\mu = \partial_\mu + i\frac{g}{2}\sigma^i W_\mu^i + i\frac{g'Y}{2}B_\mu \tag{1.17}$$

where σ are the Pauli matrices which operate on left-handed SU(2) doublets, and for right-handed singlets the second term is zero. Given the non-Abelian nature of SU(2), this formalism contains self-interactions of the W^i-bosons.

The experimentally observed short-range nature of the weak interaction motivated massive vector gauge bosons for its description, yet this Lagrangian does not possess a mass term. However, inserting a mass term of the form $m_W^2 W^{i\mu} W_\mu^i$ is not allowed by gauge invariance. A more sophisticated method is therefore required to generate massive vector bosons.

1.7 Electroweak Symmetry Breaking

The process by which a system transitions from a symmetric state down to one of a set of degenerate ground states, selecting a preferred direction and breaking the initial symmetry, is known as *spontaneous symmetry breaking*. Spontaneous symmetry breaking is relatively commonplace, notably explaining the dynamics of ferromagnetic materials, where below the Curie temperature a preferred direction is chosen which breaks the rotational symmetry of the system. For continuous global symmetries, as with the rotational symmetry in ferromagnetism, spontaneous symmetry breaking leads to long-range correlations associated with massless bosons, known as *Goldstone* bosons [15]. Promoting this to a local gauge symmetry, the corresponding gauge field is introduced through the covariant derivative, and the Goldstone boson is absorbed into the gauge field. The gauge boson acquires a longitudinal polarisation and a mass term [16], while preserving the underlying gauge invariance of the theory. This section will explore this phenomenon, known as the *Higgs mechanism* [17–19], and its role within the SM.

1.7.1 The Higgs Mechanism applied to Electroweak Theory

To realise electroweak symmetry breaking, the SM introduces a complex scalar doublet, $\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$, with weak hypercharge $Y = 1$. The two components have $T_3 = +1/2$ and $T_3 = -1/2$ respectively, giving $Q(\phi^+) = 1$ and $Q(\phi^0) = 0$. The lower component can therefore acquire a vacuum expectation value without breaking electromagnetism, leaving the photon massless. We can then write the Lagrangian for the Higgs sector as,

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{1.18}$$

where the most general gauge-invariant renormalisable scalar potential may be written as,

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \qquad \mu^2, \lambda > 0. \tag{1.19}$$

This potential has minima at $\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda}$, with a vacuum expectation value, $v^2 = \frac{\mu^2}{\lambda}$. Since the potential depends only on $\Phi^\dagger \Phi$, all field configurations satisfying $\Phi^\dagger \Phi = \mu^2/(2\lambda)$ are degenerate. However, once a particular minimum is chosen the $\text{SU}(2)_L \times \text{U}(1)_Y$ group is spontaneously broken to the electromagnetic subgroup $\text{U}(1)_{\text{EM}}$. To illustrate this, we can take the real component of the complex Higgs doublet at the vacuum state, starting with an exponential expression for the Higgs doublet,

$$\Phi(x) = e^{\frac{i\sigma^i \pi^i(x)}{v}} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.20}$$

where $\pi^i(x)$ are the Goldstone modes and $h(x)$ is the Higgs field. We are free to apply a gauge transformation to rotate out the Goldstone modes,

$$\Phi(x) \mapsto \Phi'(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.21}$$

this is known as the *unitary gauge* and has been chosen to make the physical W and Z boson mass terms explicit in the final Lagrangian. Under the above gauge transformation, the Goldstone fields are removed from the scalar doublet and are *eaten* to form the longitudinal polarisations of the physical W and Z bosons.

Inserting this form into the Higgs potential of Equation 1.19 yields,

$$V(h) = -\frac{\mu^4}{4\lambda} + \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4. \tag{1.22}$$

The excitation of this field about the minima is the Higgs boson, with an explicit mass given by $m_h^2 = 2\lambda v^2$. The third and fourth terms of this potential identify the trilinear and quartic Higgs self-interaction terms. For the remainder of this thesis, the cubic and quartic terms are identified by λ_{HHH} and λ_{HHHH} respectively. The SM predicts $\lambda_{HHH} = \lambda v$ and $\lambda_{HHHH} = \frac{\lambda}{4}$ [20].

Expanding about the minimum within the kinetic term of the Lagrangian,

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \frac{v^2}{8} \left[g^2 \left((W_\mu^1)^2 + (W_\mu^2)^2 \right) + \left(g W_\mu^3 - g' B_\mu \right)^2 \right] \tag{1.23}$$

and defining $W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right)$ we can identify the first and second term in brackets with the W boson mass term with $m_W = \frac{gv}{2}$. The third term, $\left(g W_\mu^3 - g' B_\mu \right)^2$, may be written as,

$$\mathcal{L} = \frac{v^2}{8} \begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix} \begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix} \begin{pmatrix} W^{3,\mu} \\ B^\mu \end{pmatrix} \tag{1.24}$$

and since the matrix is not diagonal, B_μ and W_μ^3 are not mass eigenstates.

This matrix has two eigenvalues: 0 and $g^2 + g'^2$, corresponding to eigenvectors $\begin{pmatrix} g' \\ g \end{pmatrix}$, $\begin{pmatrix} g \\ -g' \end{pmatrix}$ associated to the photon and the Z boson respectively. Defining the weak mixing angle as $\tan\theta_W = \frac{g'}{g}$, the physical bosons are expressed as a mixture of W_μ^3 and B_μ fields:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{1.25}$$

This predicts that the massive neutral Z boson should be heavier than its charged W counterpart:

$$m_W = m_Z \cos\theta_W. \tag{1.26}$$

The longitudinal modes of the W and Z bosons are a prediction of electroweak symmetry breaking and, without a Higgs boson, the amplitude for longitudinal vector boson scattering becomes non-unitary at high energies. This unitarity

violation shows either that the Higgs boson must exist or that new physics should regulate this scattering at the TeV scale. This energy scale can be expressed in terms of the SM Higgs mass,

$$m_H \lesssim 1 \text{ TeV} \tag{1.27}$$

which is known as the Lee-Quigg-Thacker [21] bound and served as a strong motivation for constructing the Large Hadron Collider (LHC): either the Higgs boson is discovered or new physics will manifest at the TeV scale that would regulate this scattering[1].

1.7.2 Fermion Masses

As introduced earlier for the electroweak Lagrangian, left-handed fermions transform as doublets whereas right-handed fermions transform as singlets under SU(2), with

$$\begin{aligned} L &= \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}, \qquad e_R,\ \mu_R,\ \tau_R \\ Q &= \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \qquad u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R. \end{aligned} \tag{1.28}$$

Unlike in pure QED, where fermion mass terms are gauge invariant and appear in the Dirac equation, the electroweak sector forbids a direct mass term like $\bar{e}_L e_R$ since fermions are left-handed doublets but right-handed singlets under $\mathrm{SU}(2)_L \times U(1)_Y$. Fermion masses are instead generated through Yukawa interactions with the Higgs doublet [14]. Schematically, these take the form

$$\mathcal{L}_{\text{Yukawa}} = -\bar{Q}_L Y_d \Phi d_R - \bar{Q}_L Y_u \tilde{\Phi} u_R - \bar{L} Y_e \Phi e_R + \text{h.c.}, \tag{1.29}$$

where Y_u, Y_d, and Y_e are 3×3 matrices in flavour space and $\tilde{\Phi} = i\sigma_2 \Phi^*$. After electroweak symmetry breaking, these interactions generate fermion mass terms and Higgs-fermion interactions,

$$M_f = \frac{v}{\sqrt{2}} Y_f. \tag{1.30}$$

[1] Or both, as will be explored in the context of the dark Higgs model in Section 3.3.4.

Following their diagonalisation, the Yukawa interaction for a fermion mass eigenstate f becomes

$$\mathcal{L}_{\text{Yukawa}} = -m_f \bar{f} f - \frac{m_f}{v} h \bar{f} f, \tag{1.31}$$

with $m_f = \frac{y_f v}{\sqrt{2}}$, showing that the Higgs coupling to fermions is proportional to the fermion mass.

In the SM, the Yukawa couplings are matrices acting on flavour eigenstates, and the physical mass eigenstates are obtained by diagonalising these matrices through unitary rotations of the fermion fields. In the quark sector the rotations that diagonalise the up- and down-type Yukawa matrices differ, and this mismatch appears as the Cabibbo-Kobayashi-Maskawa (CKM) matrix [22]. The irreducible remaining complex phase present in the 3×3 CKM matrix gives rise to CP violation in the SM. Fermion masses and Yukawa couplings are not a prediction of the SM, and are determined by experimental measurements.

1.7.3 Experimental Confirmation

The experimental observation of neutral currents by the Gargamelle bubble chamber at CERN [23, 24] was the first hint that the electroweak gauge theory of Glashow, Salam and Weinberg is indeed realised in nature. The picture was sharpened by the discovery of the W and Z bosons by the UA1 [25, 26] and UA2 [27, 28] experiments with masses that agreed with estimates based on electroweak observables. The electroweak picture, and the SM, was completed with the discovery of the Higgs boson by the ATLAS [29] and CMS [30] experiments at the LHC in 2012, crowning a remarkable set of achievements in both theoretical and experimental physics.

1.7.4 Evolution of the Electroweak Vacuum

Since the discovery of the Higgs boson, attention has turned to the nature of the boson itself and of the spontaneous symmetry breaking mechanism predicted by the SM. Early in the evolution of the Universe, the electroweak vacuum evolved from a symmetric to a broken phase, giving rise to a non-zero vacuum expectation value

and hence to the masses of the weak gauge bosons and fermions. Experimentally, the precise nature of this evolution is unknown, but the SM makes a clear prediction, and the shape of the Higgs potential is central in this argument.

The Higgs potential introduced in Equation 1.22 is the area of the Higgs potential probed at hadron colliders, where the electroweak symmetry is broken and the Higgs field is at the minimum of the potential. A common parametrisation away from the minimum [31], at finite temperature T is,

$$V_{\text{eff}}(h,T) \sim D(T^2 - T_0^2)h^2 - ETh^3 + \frac{\lambda_T}{4}h^4 \tag{1.32}$$

where in the SM $D = \frac{1}{8v^2}\left(2m_W^2 + m_Z^2 + 2m_t^2 + m_H^2\right)$, $E = \frac{1}{4\pi v^3}\left(2m_W^3 + m_Z^3\right)$ and λ_T is the temperature dependent Higgs self-interaction. For high temperatures, the potential has a minimum at $h = 0$, known as the *symmetric* phase, and for temperatures $T < T_0$ the minimum is no longer at 0 but at non-zero values, known as the *broken* phase.

The behaviour of the potential as a function of temperature, particularly the transition dynamics between the symmetric and broken phase, is of considerable phenomenological importance. The transition is first order if, at some temperature T_c, the potential has degenerate minima, $V(0,T_c) = V(v,T_c)$, separated by a potential barrier. Lattice calculations show that the mass of the Higgs boson must be less than 72.4 ± 1.7 GeV [32] for a first-order phase transition. This is not realised with a measured Higgs boson mass of $m_H = 125.11 \pm 0.11$ GeV [33], meaning that the transition predicted by the SM is a smooth crossover and will have consequences later for models of electroweak baryogenesis. A smooth crossover is a gradual transition between two regimes, rather than a thermodynamic phase transition, and for the electroweak crossover the minimum of the Higgs potential evolves continuously from zero to a minimum at non-zero values of the Higgs field.

Figure 1.1 shows two Higgs potentials at temperatures just below the critical temperature, T_c, for a mass well below the 72.4 ± 1.7 GeV bound predicted by lattice calculations for a first-order phase transition in the early Universe, contrasted against the experimentally measured Higgs boson mass of 125.11 GeV, predicting a smooth crossover.

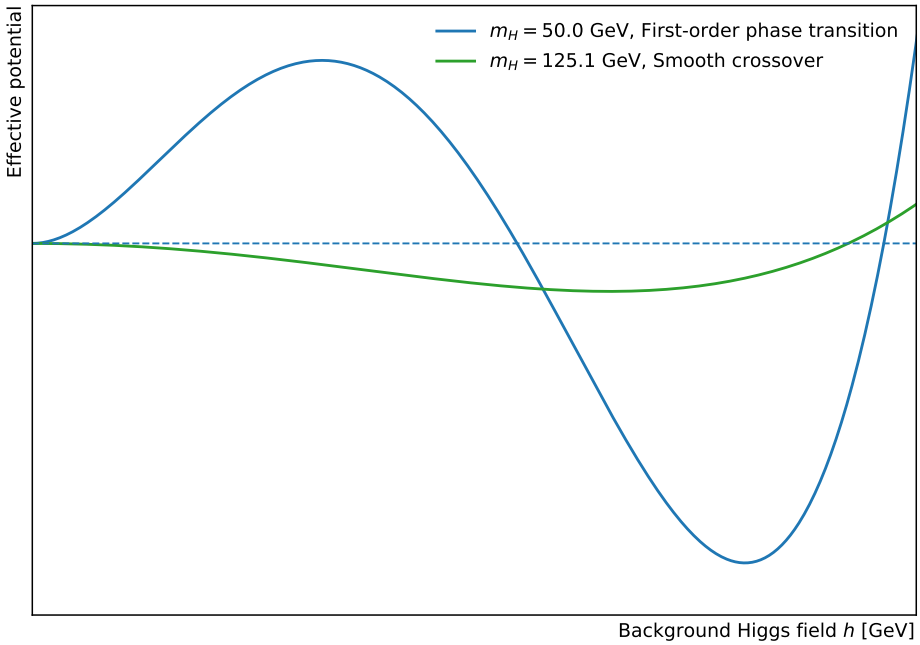


Figure 1.1: Schematic showing the structure of the Higgs potential at temperatures below the critical temperature T_c for the SM with a Higgs mass of 50 GeV (blue) clearly showing a minimum at non-zero Higgs field with a potential barrier separating the symmetric and broken phase indicating a first-order phase transition and 125.1 GeV (green) showing a smooth crossover.

2

The Standard Model at Colliders

Contents

2.1 Event Simulation 14
2.1.1 Hard-scatter Event Generation 14
2.1.2 Parton Shower 16
2.1.3 Matching and Merging 17
2.1.4 Hadronisation 17
2.1.5 Underlying Event 19
2.2 Detector Simulation 19
2.3 Higgs Boson Production 20

2.1 Event Simulation

Connecting the theoretical structure of the SM to the data collected following proton-proton collisions at the LHC is an intensely challenging, high-dimensional problem. The simulation must span orders of magnitude in energy, from the high-energy parton-parton interactions down to the low-energy non-perturbative dynamics responsible for hadronisation within jets and for the interactions of stable particles with the detector material. General-purpose event generators, like Pythia [34] or Herwig [35, 36], seek to model the entire event structure and are based on Monte Carlo (MC) simulation, with the outputs passed to Geant4 [37], which models detector interactions. This chapter will summarise the principles behind the event generators used throughout the thesis.

2.1.1 Hard-scatter Event Generation

The factorisation theorem of QCD [38] enables the cross-section for a given process to be decomposed into two components: the non-perturbative composite structure

of the proton, and the interactions between partons that produce the final states of interest. The hard subprocess is described by a partonic cross-section $\mathrm{d}\hat{\sigma}_{ij\to X}$, evaluated at the parton centre-of-mass energy $\sqrt{\hat{s}}$, where $\hat{s} = x_1 x_2 s$, x_1 and x_2 are the longitudinal momentum fractions of the two interacting partons and $\sqrt{s}$ is the centre-of-mass energy of the collision. The density of parton species i, with momentum fraction x of the incoming protons' momenta, is parametrised by parton distribution functions (PDFs). PDFs are derived from a variety of datasets sensitive to proton structure, such as deep-inelastic scattering (DIS) measurements, fixed target and, more recently, measurements at the LHC.

The full cross-section may be written as,

$$\mathrm{d}\sigma(pp \to X) = \sum_{i,j} \int \mathrm{d}x_1 \mathrm{d}x_2 f_i\left(x_1, \mu_F\right) f_j\left(x_2, \mu_F\right) \mathrm{d}\hat{\sigma}_{ij\to X}\left(x_1 x_2 s; \mu_R, \mu_F\right) \quad (2.1)$$

where i, j are the incoming parton species; μ_F is the factorisation scale of an event which separates the short-distance hard scatter and long-distance PDFs; μ_R is the renormalisation scale introduced to set the scale of the interaction, such as the strength of the strong coupling constant, $\alpha_s(\mu_R)$, when truncating the perturbative expansion; $f_i\left(x, \mu_F\right)$ is the density of parton i with momentum fraction x inside the proton at scale μ_F and $\mathrm{d}\hat{\sigma}_{ij\to X}\left(x_1 x_2 s; \mu_R, \mu_F\right)$ is the cross-section for partons ij scattering to form the final state X. The arbitrary scales μ_R and μ_F cancel in an all-orders prediction, but dependence remains after truncating the perturbative expansion. Varying these scales, together with PDF and α_s variations, provides an estimate of the theoretical uncertainty on the cross-section [39].

An exact calculation of the amplitude requires a sum over all possible Feynman diagrams that contribute to the process under consideration. Pragmatically, these calculations are often truncated under the assumption that, within perturbation theory, more complex diagrams contribute an increasingly small fraction of the total amplitude. This assumption is perfectly valid for coupling factors much less than 1 (e.g. $\alpha_{\mathrm{EM}} \sim \frac{1}{137} \ll 1$). However, with the running of the strong coupling constant, for sufficiently low energies ($\mathcal{O}(1)$ GeV), α_s approaches and exceeds 1 — known as the non-perturbative limit of QCD.

The leading-order (LO) contribution is the lowest non-vanishing order in the relevant expansion. Higher-order corrections (starting with next-to-leading order (NLO)) include virtual-loop and real-emission contributions with additional final-state partons. Event generators such as MadGraph5_aMC@NLO [40] and Powheg [41–43] use importance-sampling techniques to populate phase space efficiently by flattening the integrand, preferentially generating events in regions where the matrix element and phase space give large contributions to the differential cross-section. For each generated event, the matrix-element generator produces states specified by particle four-momenta and an associated event weight, which are then passed to a parton shower.

2.1.2 Parton Shower

The final state produced by the matrix-element generator is not a realistic description of a collider event. Short-lived states are decayed according to their branching ratios (resonances such as H, W, Z and top quarks will decay almost instantaneously with respect to measurements within a detector volume), with modern generators preserving relevant spin correlations where required to model the decay products' angular distributions accurately.

Then, analogously to accelerated charges emitting photons in QED, energetic coloured particles emit gluons. Unlike QED, gluons can self-interact and emit further gluons that can split into $q\bar{q}$ pairs — producing a cascade of quarks and gluons known as a *parton shower* [44]. The shower is evolved down to a cut-off scale $Q \sim \mathcal{O}(1)$ GeV, where non-perturbative effects begin to dominate. In the soft and collinear limit of QCD, branchings in the shower are modelled using Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting functions, such as $P_{q\to qg}(z)$, which describe the probability for a parton to split as a function of the momentum-sharing variable z. Analogous splitting functions are used for QED radiation, such as photon emission from charged particles. Starting from a large-scale Q_0, the shower is evolved downward in an ordering variable, with the probability of no resolvable emission between scales Q_0 and Q_1 encoded by

a no-emission probability known as the Sudakov form factor. The next emission scale is then sampled from the corresponding probability distribution, given by the product of the Sudakov form factor and the local branching rate. The choice of ordering variable also dictates shower evolution, and modern parton showers commonly use either p_{T}-ordering (Pythia) or angular ordering (Herwig).

At hadron colliders, radiation from the incoming partons is also important. The hard-scatter process samples from PDFs with momentum fractions x_1 and x_2, giving a squared partonic centre-of-mass energy $\hat{s} = x_1 x_2 s$. Initial-state radiation (ISR) is then generated using backwards evolution, in which the shower reconstructs a sequence of earlier branchings consistent with the final hard-process.

2.1.3 Matching and Merging

Matrix-element generators provide an accurate description of hard, well-separated partons, while parton showers model the additional radiation produced through softer or more collinear emissions. A consistent event simulation must combine these descriptions without double-counting emissions that could be produced by either the matrix element or the shower. Matching schemes, such as those used by Powheg and MadGraph5_aMC@NLO, combine NLO matrix elements with parton showers while preserving the inclusive NLO cross-section [42, 45]. Merging schemes combine multiple samples at different fixed orders (e.g. Zj and Zjj), separated by a merging scale, so that hard emissions are described by matrix elements and softer emissions by the shower.

2.1.4 Hadronisation

Once partons reach scales close to Λ_{QCD} ($\sim$ 200 MeV), QCD becomes strongly coupled and coloured particles neutralise to form colour-neutral hadrons in a process known as *hadronisation*. Since hadronisation is non-perturbative, only phenomenological models exist and are tuned to match the observed data. The two most common approaches are Lund-string hadronisation in Pythia [46, 47] and cluster hadronisation in Herwig [36].

The Lund-string model introduces connections between coloured partons in the final state, with a string of constant energy density per unit length. As the partons separate, the string increases in tension until it is energetically favourable to form $q\bar{q}$ pairs. This evolution continues if the produced $q\bar{q}$ system has a high invariant mass; otherwise, these form candidate hadrons. The above mechanism naturally leads to a number of mesons in the final state, and to produce baryons, the Lund string model also includes the possibility that a string may break into diquark, di-anti-quark pairs ($qq - \bar{q}\bar{q}$).

The probability of producing a $q\bar{q}$ pair during string breaking is suppressed for heavier quarks. The Lund string model incorporates this with probability ratios proportional to,

$$\frac{P(q\bar{q})}{P(u\bar{u})} \propto \exp\left(-\frac{\pi\left(m_q^2 - m_u^2\right)}{\kappa}\right) \tag{2.2}$$

where κ is the string tension. Non-perturbative production of heavy flavour quark pairs is therefore very strongly suppressed. In Pythia this is set to zero for b and c-splittings, meaning the production of heavy flavour inside jets is purely perturbative, in the hard process or parton shower, which subsequently hadronise into heavy-flavour hadrons [48].

In the Lund string model, the longitudinal momentum sharing in each string break is described by the fragmentation function,

$$f(z) = \frac{(1-z)^a}{z}\exp\left(-\frac{bm_T^2}{z}\right) \tag{2.3}$$

where m_T is the transverse mass of the hadron and a and b are tunable parameters of the model. Since the heavy-quark mass is large compared with the non-perturbative scale, the resulting heavy hadron usually carries a large fraction of the original heavy-quark momentum, corresponding to a harder fragmentation function — a key observable in heavy-flavour identification as will be discussed later.

Cluster hadronisation utilises the pre-confinement property of QCD [49], which is the observation that quarks and gluons produced by a parton shower evolution, between the hard-scatter scale and the non-perturbative regime, naturally form

colour-singlet states with masses independent of the initial scale of the event. Remaining gluons are forced to decay into $q\bar{q}$ pairs, and the resultant clusters are decayed to form final-state hadrons.

2.1.5 Underlying Event

Additional interactions occur beyond that of the hard-scatter, known as the *underlying event*. It receives contributions from multiple parton interactions (MPI), beam remnants, colour reconnection, hadronisation, and radiation not directly associated with the primary hard process. In the scenario that the two colliding protons have a small transverse impact parameter, the probability that multiple partons interact grows, and Pythia uses this relationship to model MPI. Moreover, once the hard subprocess occurs, the extraction of partons with momentum fractions x_1 and x_2 leaves coloured beam remnants carrying the remaining longitudinal momentum. Since a coloured object is involved in the hard-scatter, the beam remnants are connected to the colour-strings used within Pythia for Lund-string hadronisation [48].

2.2 Detector Simulation

Once stable particles on detector timescales have been formed, the expected response of the detector to such a set of states must be simulated. The complete detector geometry of an experiment is implemented in Geant4, which models ionisation, bremsstrahlung, pair production, multiple scattering and hadronic interactions of particles traversing the volume. The output of Geant4 is an event record of hits in the tracker and energy deposits within the calorimeter, which are fed to electronic readout simulation and digitisation, and then enter the same reconstruction algorithms deployed for data-taking [50]. This involves, for ATLAS, all 100 million readout channels, together with a precise understanding of the material budget, composition, and position.

Figure 2.1 is a summary of SM processes measured by ATLAS and the corresponding theoretical prediction. The level of agreement across such a wide range of

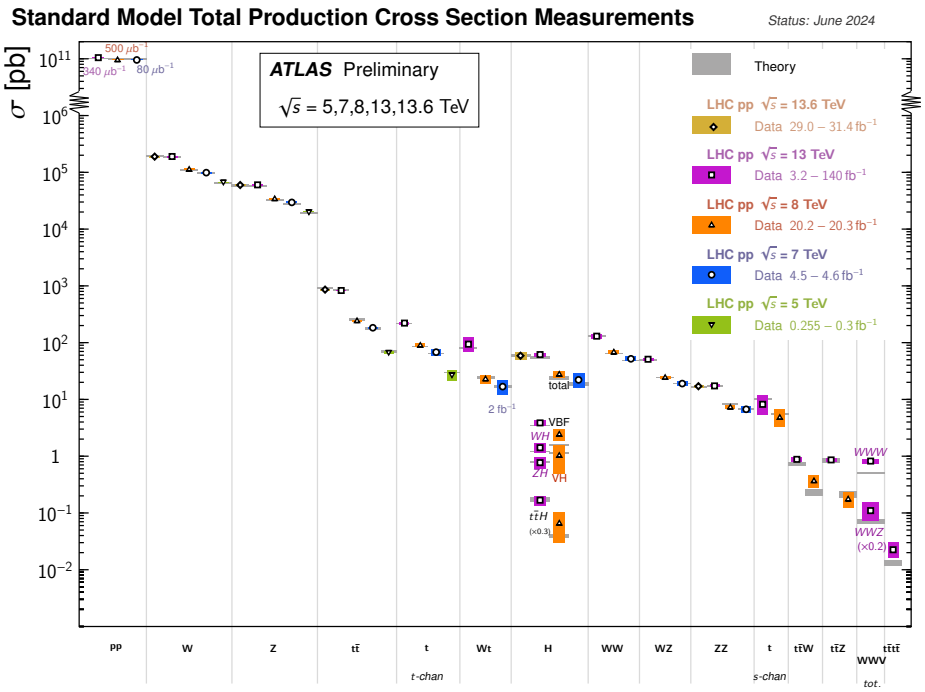


Figure 2.1: Theory predictions and measured cross-sections for a wide range of SM processes by the ATLAS experiment across five centre-of-mass energies. Figure from Ref [51].

processes and energy scales is a striking validation of perturbative SM predictions, MC modelling, detector simulation and experimental operation. One of the two analyses presented in this thesis will study the production of an exceptionally rare process in the SM: the production of three SM Higgs bosons. A brief account of the dominant production modes is given here, along with a motivation to probe this final state at future high-energy proton-proton colliders.

2.3 Higgs Boson Production

The dominant production mechanism for SM Higgs bosons at colliders is via gluon-gluon fusion (ggF), whereby two gluons fuse to form a heavy-quark loop which couples to the Higgs. Other processes include the production in association with a vector boson (VH), via vector-boson fusion (VBF) or in association with a top quark or top-quark pair ($tH/t\bar{t}H$) [52]. ggF is shown with the leading diagram contributing to single, double and triple-Higgs production in Figure 2.2,

and since the Higgs boson couples proportionally to mass, the production is strongly dominated by top-quark loops.

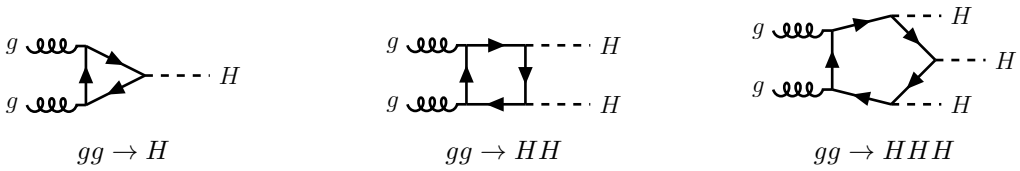


Figure 2.2: Representative leading-order gluon-gluon fusion diagrams for single-, double- and triple-Higgs production via a quark loop.

Figure 2.3 shows the leading-order cross-section as predicted by MadGraph for single, double and triple-Higgs production via `ggF` as a function of the centre-of-mass energy ($\sqrt{s}$) from 14 TeV to 120 TeV, with a ratio to 14 TeV shown in the lower panel. The minimum fraction of the proton's momenta required by the gluons can be shown to scale as $x_{\min} \sim \frac{3m_H}{\sqrt{s}}$ for triple-Higgs. Therefore, for a fixed event scale, the gluon density within the proton to produce such a state sharply rises with $\sqrt{s}$, as shown by the steep increase in cross section for all three processes.

Figure 2.4 shows the gluon density as a function of Bjorken-x, where the increase in gluon density for the higher-scale triple-Higgs process is shown relative to single-Higgs, with a ratio between them shown in the lower panel. Given the larger scale for triple-Higgs ($m^2_{HHH} \gg m^2_{HH} \gg m^2_H$), the gluon enhancement is sharper. Therefore, the production of triple-Higgs events scales the steepest with $\sqrt{s}$ — strongly motivating future high-energy collider facilities to measure such a process. Specifics of triple-Higgs production will be explored later, briefly with some discussion of the κ-framework in Section 3.2 and within the FCC-hh analysis later in Section 8.

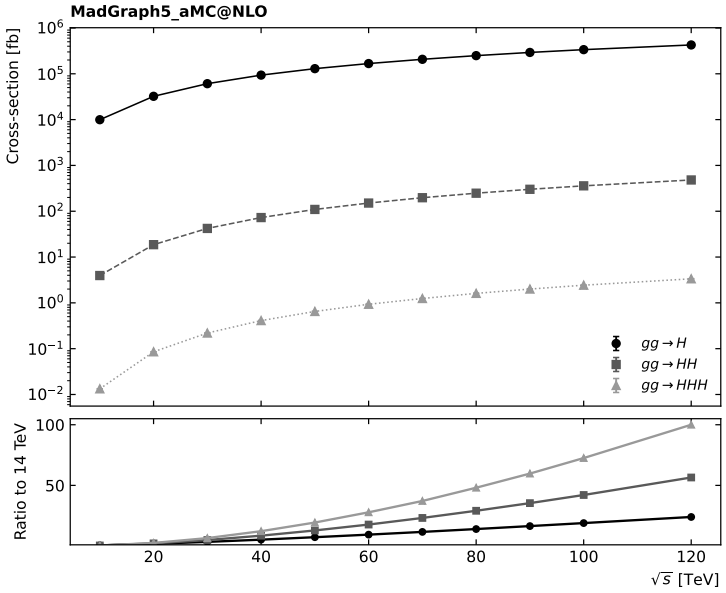


Figure 2.3: Leading-order cross-section for single, double and triple-Higgs production via gluon-gluon fusion as a function of centre-of-mass energy.

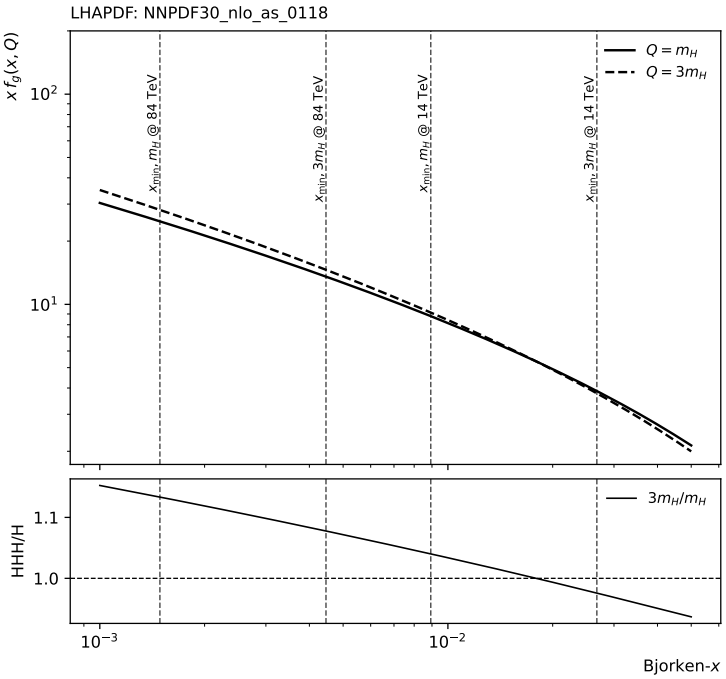


Figure 2.4: Gluon densities, using the `NNPDF30_nlo_as_0118` PDF set [53], for single-Higgs and triple-Higgs production, with vertical lines indicating the minimum Bjorken-x required to produce the respective process at a given energy scale.

3

Beyond the Standard Model

Contents

3.1 Limitations of the SM 23
3.1.1 Dark Matter 24
3.1.2 Theoretical Shortcomings 27
3.1.3 Electroweak Baryogenesis 27
3.1.4 Connections to the Higgs sector 30
3.2 Quartic Higgs self-coupling 30
3.3 Searching for Dark Matter 33
3.3.1 Direct Detection 33
3.3.2 Indirect Detection 33
3.3.3 Collider Searches 34
3.3.4 Di-Higgs + $p_{\mathrm{T}}^{\mathrm{miss}}$ Phenomenology 35

3.1 Limitations of the SM

The SM is not perfect, and a number of experimental results and theoretical arguments are made to show this. Firstly, the SM describes three of the four known fundamental forces — notably omitting the gravitational interaction. Unifying quantum mechanics and gravity is one of the big open questions in physics, and the SM does not accommodate it. Recent measurements of neutrino oscillations show that neutrinos have mass [54], but the lack of right-handed neutrino prevents a mass term through Yukawa-like interactions. Astrophysical observations indicate the existence of a weakly interacting type of matter, *dark matter*, but the SM does not possess such a particle [6, 55]. There is an excess of matter over antimatter, and the CP violation in the SM is insufficient to account for this imbalance. Moreover, there are several aesthetic theoretical concerns, including the fine-tuning of the Higgs mass, the large number of free parameters in the SM, and the strong CP

problem. The above makes it abundantly clear that the SM is not a complete theory of physics. This chapter will expand on a number of these shortcomings and also explore some proposals for physics beyond the SM (BSM).

3.1.1 Dark Matter

Numerous astrophysical measurements support the claim that dark matter exists in the Universe. A brief account of this evidence will be given, starting with the observed anomalous rotation curves of galaxies as a function of distance from the galactic centre.

Rotation Curves

For a massive body in a gravitationally bound orbit, its velocity is given by

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{3.1}$$

where r is the radius of the orbit, G is the gravitational constant and $M(r)$ is the mass contained within a radius r from the galactic centre.

This relation predicts a $v \propto \frac{1}{\sqrt{r}}$ relationship for objects away from the visible galactic centre. However, when observing the rotation of stars on the outer edges of galaxies, a roughly constant relationship emerged [56–59]. A constant relationship implies that the increase in matter content is proportional to the radius ($M(r) \propto r$), leaving $v \propto$ const. — indicating the presence of another form of matter, which is not visible, located on the periphery of the galaxy.

Bullet Cluster

A particularly compelling piece of evidence is that of the bullet cluster [60], an observation of two galaxy clusters colliding. Galaxy clusters are comprised of three components: the galaxies themselves, the intracluster gas, and the hypothetical dark matter. The idea of this measurement is that when two galaxy clusters collide, the behaviour of the three components differs — enabling an independent probe of each.

Given the large spatial separation of galaxies within each cluster, the probability of a collision between two galaxies is remarkably small, so that this system is

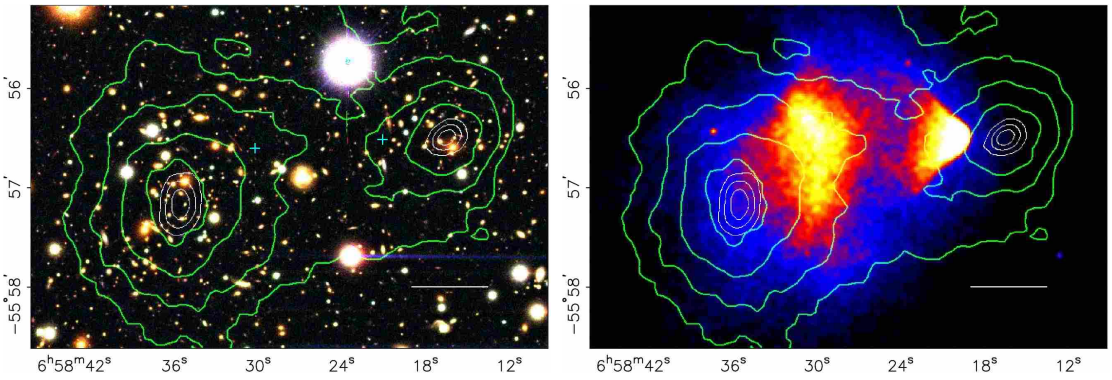


Figure 3.1: (a) Magellan images of the galaxies within each cluster and (b) a heatmap of the density of intracluster gas after the collision, decoupled from the centre of the galaxy. The green contours are the inferred amount of mass via weak gravitational lensing, peaking around the centres of the visible galaxies. Figure taken from Ref [60].

effectively collisionless. The gas contained within each interacts, slowing the passage of the gas relative to the galaxies and can completely separate from the smaller galaxy — known as ram pressure stripping. And finally, the hypothesised dark matter is weakly interacting and will pass through unperturbed.

Measurements are made after the collision, and Figure 3.1(a) shows the distribution of galaxies within each using images by the Magellan telescope. Figure 3.1(b) shows a heatmap of the intracluster gas density after the collision, clearly showing the gas decoupled from the visible galaxies. Inferred from weak gravitational lensing, the green contours indicate the density of matter in the system and peak around the centres of the visible galaxies. This is contradictory to measurements of galaxy clusters before a collision, in which the majority of the visible mass is contained within the gas itself. This indicates the existence of a weakly interacting form of matter surrounding the two clusters of galaxies that passed through effectively interactionless and maintained its form around the galaxies as they drifted through.

Cosmic Microwave Background

Finally, by studying the dynamics of the early Universe, the density of dark matter can be estimated. Moments after the Big Bang, the Universe was a hot, dense plasma where photons had an incredibly short mean free path and mostly interacted

by Thomson scattering off free electrons ($e\gamma \to e\gamma$). As the Universe expands and the temperature decreases, a recombination temperature is reached at which electrons and protons combined to form Hydrogen. This formation of neutral states reduced the number of free electrons, and the previously scattering photons are free to propagate throughout space. The temperature at which photons last scattered is roughly 3000 K, but as the Universe continues to expand, this temperature steadily decreased to a temperature of 2.7 K today and is known as the *cosmic microwave background* (CMB). Fluctuations in the CMB temperature arose via the influence of cold dark matter, which produced stable gravitational potential wells that altered the amplitude of acoustic peaks in the CMB power spectrum. Measurements of this power spectrum enabled Planck [61] to measure both the dark matter and baryonic density of the Universe as,

$$\Omega_c h^2 = 0.1200 \pm 0.0012 \tag{3.2}$$

$$\Omega_b h^2 = 0.02237 \pm 0.00015 \tag{3.3}$$

where Ω is the fractional density of dark matter (c) and baryons (b) compared to the current critical density of the Universe, and h is the reduced Hubble constant.

WIMP Miracle

Numerous hypotheses have been put forth to provide candidates of dark matter; one compelling case is that of the weakly interacting massive particle (WIMP), which is motivated by thermal freeze-out. Suppose a particle candidate for dark matter, χ, exists and has some mechanism to interact with SM particles. At high temperatures, well above m_χ, the creation (SM SM $\to \chi\bar{\chi}$) and annihilation ($\chi\bar{\chi} \to$ SM SM) channels are in thermal equilibrium. As the Universe expands and cools below $T \sim m_\chi$, the production of dark matter becomes Boltzmann suppressed, while the dark matter already present continues to annihilate, but increasingly infrequently given the expansion of the Universe — eventually reaching a stable abundance of dark matter, known as the *relic density*. Then, if the mass and coupling of the dark matter candidate are around the electroweak scale, the

predicted relic density closely matches the measurements made by Planck [55]. The reproduction of the relic density by a WIMP around the electroweak scale is known as the *WIMP miracle* [62]. Masses around the EW-scale are, for the first time, broadly accessible at the LHC, making the prospect of searching for dark matter at colliders a particularly exciting one.

3.1.2 Theoretical Shortcomings

The SM has several shortcomings, so it is often regarded as an effective field theory valid below some higher energy scale. If we connect the SM to reasonable higher scales, such as the scale at which the gauge interactions become of equal strength (the Grand Unification Scale [63]) or the Planck scale, and calculate radiative corrections to the Higgs mass, we see that the mass receives large corrections. This is unique to the scalar sector of the SM: chiral symmetry ensures that corrections to fermion masses remain proportional to the fermion mass itself, whilst gauge symmetry protects gauge-boson masses. There is no such symmetry protecting the Higgs boson mass, which is therefore sensitive to higher scales through loop corrections. The physical mass can be related to such scales and corrections by $m_H^2 = m_{H,\text{bare}}^2 + \delta m_H^2$. By considering the corrections connected to the GUT scale $\delta m_H^2 \propto \Lambda_{\text{GUT}}^2$, with $\Lambda_{\text{GUT}} \sim 10^{15}$ GeV, we see that for a Higgs boson to have a physical mass at the experimentally measured 125.11 ± 0.11 GeV, there has to be an extraordinary cancellation between the bare mass and the corrections [64]. This fine-tuning is considered unnatural and is the essence of the hierarchy problem.

3.1.3 Electroweak Baryogenesis

There is an excess of matter over anti-matter in the Universe, which can be quantified by the difference between number densities of baryons (matter) and anti-baryons (anti-matter), $\Delta = n_B - n_{\bar{B}}$, normalised by the photon density of the CMB, which scales in approximately the same way with the expansion of the Universe [55],

$$\eta_B = \frac{n_B - n_{\bar{B}}}{n_\gamma} \sim 6 \times 10^{-10}, \tag{3.4}$$

where n_X ($X \in \gamma, B, \bar{B}$) is the number density of the particular particle species, the number of baryons is inferred from measurements of the baryon density (e.g. the CMB, consistent with Big Bang Nucleosynthesis), the number of photons is related to the thermal average temperature of the CMB ($\sim$2.7 K) and the number of anti-baryons is assumed to be negligibly small.

Assuming an initially baryon-symmetric Universe which respects CPT invariance, three conditions, postulated by Andrei Sakharov [65], are necessary to dynamically generate a baryon asymmetry:

1. Baryon number violation,

2. C and CP-violation,

3. Departure from thermal equilibrium.

The first postulate is clearly required to generate an asymmetry, and the remaining can then be understood by considering a simple baryon-violating process: $X \to qq$. Charge conjugation violation is required since the process $\bar{X} \to \bar{q}\bar{q}$ would occur at the same rate. Moreover, if the combination of CP was conserved, the following equality: $\Gamma(X \to q_L q_L) = \Gamma(\bar{X} \to \bar{q}_R \bar{q}_R)$ holds and the asymmetry is again washed out. Finally, a departure from thermal equilibrium is required to ensure: $\Gamma\left(X \to qq\right) > \Gamma\left(qq \to X\right)$

The SM does contain the mechanisms sufficient to produce an asymmetry. Since only left-handed fermions transform as a doublet under $\mathrm{SU}(2)_L$, the anomalous currents generated by quantum effects are not cancelled out by right-handed components. This enables transitions between distinct EW vacuum configurations. The saddle point on top of the energy barrier between such configurations is known as a *sphaleron*. At sufficiently high temperatures, efficient transitions between EW vacua can occur — a process that conserves $B - L$ while violating $B + L$. Moreover, the chiral nature of weak interactions violates charge conjugation, and the complex phase present in the CKM matrix gives rise to CP violation in the quark sector, satisfying the second criterion. However, given that the SM predicts a smooth

crossover between the symmetric and broken phases of the electroweak vacuum, sphaleron processes occur in thermal equilibrium, and any baryon asymmetry is washed out to zero. Furthermore, even in the scenario that sphaleron processes produce a sufficient baryon asymmetry, CP violation in the SM is insufficient to replicate the observed asymmetry [66].

If, however, a strong first-order phase transition broke the electroweak symmetry, a mechanism known as *electroweak baryogenesis* [31] could explain the observed baryon abundance. As shown earlier in Figure 1.1, if the phase transition is first-order (blue), then for temperatures around and below the critical temperature, regions of space will tunnel from the symmetric phase into a broken phase via a process known as *bubble nucleation*.

For temperatures below the critical temperature, the bubbles are the energetically preferred configuration and proceed to expand, with a characteristic transition layer between the two regions known as the *bubble wall*. Inside the bubbles, the phase is broken, and sphaleron processes are exponentially suppressed. Within the bubble wall, the vacuum expectation value is position dependent. Together with additional BSM CP-violation, particles and anti-particles interact differently, generating a chiral asymmetry. Outside of the bubbles, since $v = 0$, the barrier between vacuum states is sufficiently small and enables efficient sphaleron processes to occur. Therefore, as the bubble expands and absorbs the regions of sphaleron-enhanced symmetric phase, the CP violation originating from the bubble wall is converted into a baryon asymmetry by non-suppressed sphaleron processes in the symmetric phase and can explain the observed matter-antimatter asymmetry. Interestingly, the interactions between bubbles as they expand and collide can create gravitational waves, which future space-based interferometers may be able to detect [67, 68] — enabling an independent and complementary measurement of the nature of the electroweak phase transition.

If electroweak baryogenesis is to be realised in Nature, new physics must alter the shape of the Higgs potential to generate a first-order phase transition and provide additional sources of CP violation. Various BSM theories account for this,

including regions of 2HDM [69], composite Higgs models [70, 71], extended scalar sectors coupled to the SM [72] and more generally within SMEFT frameworks [73].

3.1.4 Connections to the Higgs sector

Many indications for physics beyond the SM have connections to the Higgs sector: mass generation via the Higgs mechanism could allow gravitationally interacting dark matter to acquire a mass, the intense fine-tuning required for the Higgs mass indicates that perhaps the Higgs is not fundamental and the insufficient baryon asymmetry with the smooth cross-over predicted by the SM Higgs potential points to extended scalar sectors. The above, amongst a host of others, necessitates the measurement and characterisation of the Higgs boson in many regions of phase space. The remainder of this chapter will focus on the phenomenology of the two analyses presented in this thesis which seek to contribute to this endeavour: searches for a Higgs portal model positing a dark Higgs boson in the di-Higgs + $p_{\mathrm{T}}^{\mathrm{miss}}$ final state, and a feasibility study for measurements of the quartic Higgs self-coupling at a proposed $\sqrt{s} = 84$ TeV hadron-hadron collider in di- and tri-Higgs final states.

3.2 Quartic Higgs self-coupling

Deviations away from the SM in the Higgs sector are often parametrised within the *kappa* framework [74]. This introduces multiplicative (κ) factors for all Higgs couplings, both to fermions and gauge bosons, and also the Higgs trilinear and quartic self-coupling. The SM is taken with all κ-factors fixed at 1. This framework is easily interpretable for comparing results across collaborations, and this thesis will use the Higgs self-coupling parameters κ_3 and κ_4, which multiply Hermitian operators and are thus strictly real parameters. Constraining the trilinear and quartic Higgs self-couplings requires measurements sensitive to deviations in κ_3 and κ_4.

For the trilinear self-coupling, this is achieved either through loop-induced single-Higgs measurements or direct probes via the production of two SM Higgs bosons (di-Higgs). Isolating the quartic Higgs self-coupling requires either loop-induced constraints from di-Higgs production or tree-level constraints from the production

of three SM Higgs bosons (tri-Higgs), which are poorly constrained by the HL-LHC [75]. The loop-induced LO diagrams contributing to triple-Higgs production are shown in Figure 3.2, and the decomposition of the amplitude as a function of the triple-Higgs invariant mass is shown in Figure 3.3. The pentagon diagram (a) is dominant across the kinematic range, and destructive interference between the pentagon and box (c) significantly reduces the amplitude. The interplay of the constructive and destructively interfering terms as a function of the coupling modifiers κ_3 and κ_4 directly influences the production kinematics and cross section. Kinematics are correlated with the scale, and events near threshold ($Q \sim 3m_H$) generally contain softer Higgs bosons. An analysis targeting such signatures should therefore ensure that event selections transfer efficiently across the (κ_3, κ_4)-plane.

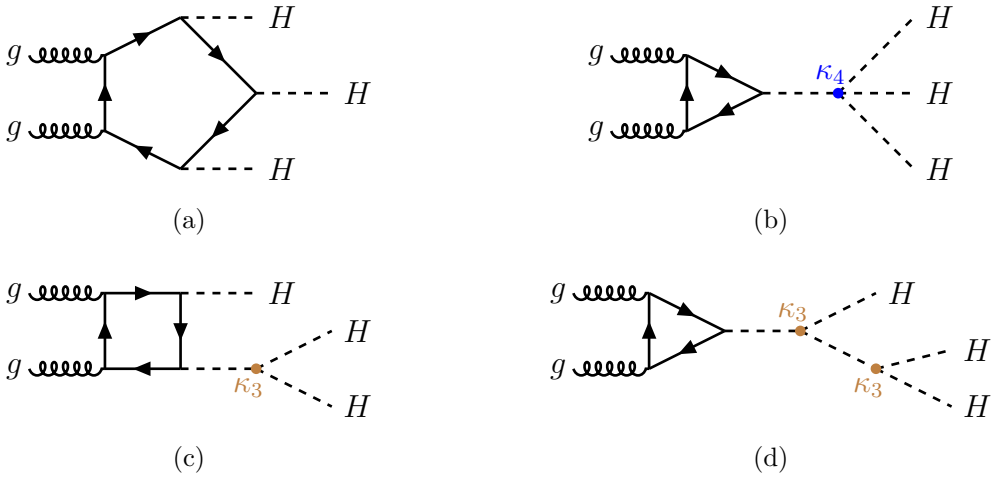


Figure 3.2: Feynman diagrams for leading-order $gg \to HHH$ production via a quark loop. The dominant production mechanism is shown in (a), whilst sub-leading diagrams (b), (c) and (d) are sensitive to κ_4, κ_3 and κ_3^2 respectively, as indicated by labels.

Two examples of NLO EW di-Higgs Feynman diagrams are shown in Figure 3.4. Since these are EW loop-induced, the dependence on the cross section and kinematics with κ_4 is less significant compared with tri-Higgs. Still, it may provide complementary information to κ_4 given that di-Higgs is produced with a greatly enhanced rate (roughly 300 times larger at an 84 TeV collider). Upon completion of the high-luminosity LHC runs, assuming SM-like production, ATLAS and CMS are expected to constrain $\kappa_3 \in [0.7, 1.3]$ using single- and double-Higgs production [77].

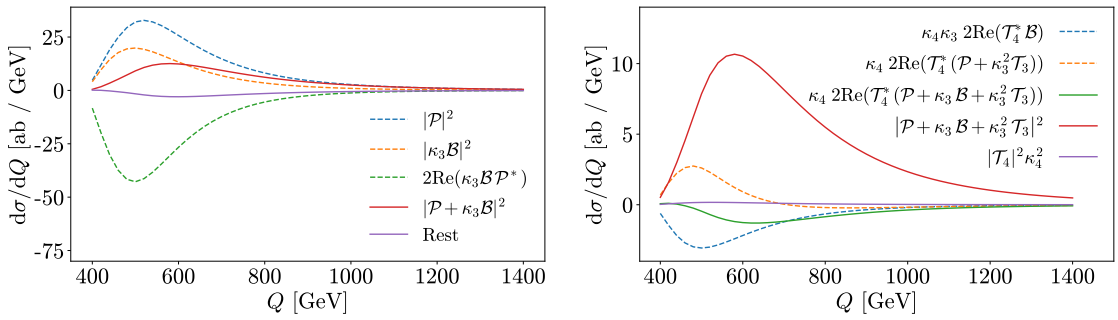


Figure 3.3: Contributions to the triple-Higgs amplitude as a function of the triple-Higgs invariant mass Q, separated into the terms independent and sensitive to κ_3 on the left and those sensitive to κ_3 and κ_4 on the right. $\mathcal{P}$, $\mathcal{T}_4$, $\mathcal{B}$, $\mathcal{T}_3$ refer to the amplitudes corresponding to the Feynman diagrams of Figure 3.2 a, b, c and d respectively. Figure from Ref [76].

ATLAS are expected to constrain $\kappa_4 \in [-84, 96]$ at 95% confidence level (CL) in the $6b$-channel [75]. Notably, appreciable limits on κ_4 beyond those already excluded by perturbative unitarity constraints from $HH \to HH$ scattering [78] are not expected in the near future.

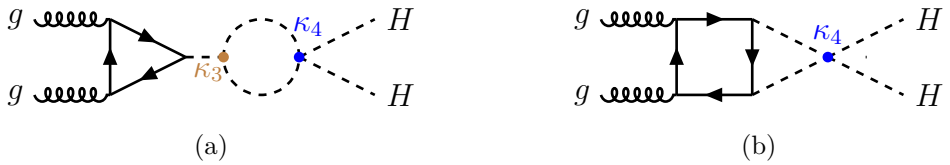


Figure 3.4: Feynman diagrams for 2-loop $gg \to HH$ production via a quark loop and a loop-correction to the Higgs self-energy (a) and additional self-interactions present for the box configuration, with vertices sensitive to κ_3 and κ_4 illustrated.

Since the trilinear and quartic Higgs self-couplings are fixed and fully correlated in the SM, a natural question is why we need to measure this parameter if the trilinear term will be known to a good precision at the end of the HL-LHC? Given that we know the SM to be incomplete, a first argument is indeed that this correlation is a prediction of a model we know to be incomplete, and before each parameter is measured this assumption carries little weight. Secondly, models of BSM physics consistent with a strong first-order phase transition decouple these parameters, and, to fully characterise such a model, we require precision on both κ_3 and κ_4, with the global behaviour of the potential dominantly driven by κ_4. For example, SMEFT above dimension six decouples the two parameters, and certain

regions of this space are compatible with electroweak baryogenesis [79]. Finally, if the measurements of κ_3 at the end of the HL-LHC are consistent with the SM, the measurement of the quartic Higgs self-coupling will become increasingly crucial to understand possible hints of new physics.

3.3 Searching for Dark Matter

A number of avenues are explored across astrophysics and particle physics to search for candidates of dark matter. Searches are defined broadly within three categories: direct detection, indirect detection and collider-based.

3.3.1 Direct Detection

Direct detection searches seek to probe individual dark matter scatterings, typically with a large volume of noble-liquid scintillator. Experiments like XENONnT [80], LZ [81, 82], PandaX-4T [83] and DarkSide-50 [84] are dual-phase time projection chamber (TPC) experiments comprised of liquid xenon/argon. Thus far, no evidence[1] for dark matter has been found at such experiments, and exclusion limits are set on spin-dependent and spin-independent DM interactions. A recent spin-independent summary plot is shown in Figure 3.5, with the liquid-argon-based DarkSide experiment leading at low mass due to the favourable recoil kinematics of a lighter noble gas inducing a sufficiently large signal, before xenon-based experiments dominate the high-mass region where the cross section is significantly enhanced with the square of the atomic mass, leading to more stringent constraints.

3.3.2 Indirect Detection

Indirect detection experiments search for evidence of dark matter annihilation or decay into SM states [86]. Experiments like Fermi-LAT [87, 88], H.E.S.S. [89],

[1] A recent notable, but isolated, event was published by LZ [85] amounting to a 2.6σ tension with the background-only hypothesis.

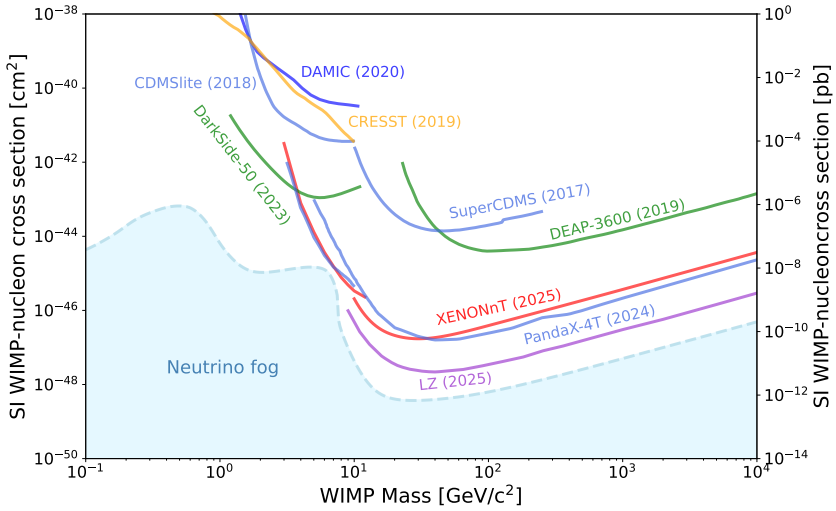


Figure 3.5: Summary plot of the upper limits on the spin-independent WIMP-nucleon cross section as a function of the DM candidate mass. The neutrino fog here refers to the region of parameter space for which neutrino-induced recoil kinematics closely match those of genuine dark matter. This difficult to distinguish background significantly diminishes the sensitivity of the experiments in this region of parameter space. Figure from Ref [6].

CTA [90] and VERITAS [91] search for anomalous production of γ-rays in astrophysical objects, while neutrino-based experiments like IceCube use high-energy neutrino events to constrain DM arising from annihilation in the Sun [92].

3.3.3 Collider Searches

Finally, collider searches seek to find DM produced as a result of a high-energy collision. The DM candidate is taken to be very weakly interacting and thus escapes detection. The presence of an invisible particle is inferred through the imbalance of transverse momentum in an event, defined as the negative vector sum of transverse momenta of all visible objects ($p_{\mathrm{T}}^{\mathrm{miss}} = -\sum \vec{p}_T$). To trigger and measure such a particle, the dark matter candidate must therefore be produced in association with SM particles, otherwise the $p_{\mathrm{T}}^{\mathrm{miss}}$ of the event will vanish. Thus, these final states lead to a signature of missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) in association with visible activity in the event - known as $X+p_{\mathrm{T}}^{\mathrm{miss}}$ searches.

ATLAS and CMS have searched for DM using both simplified and more complete models. Simplified models do not attempt to describe a fully realised theory of nature, but rather attempt to construct extensions of the SM that may be used to generically

search for BSM physics [93]. More complete models, such as the composite Higgs model or the Minimal Supersymmetric SM (MSSM) [94], seek to address a number of the SM's open questions and predict a wide array of signatures to be searched for.

A prominent example of a simplified model is the massive vector spin-1 Z'-extension, which couples to the SM and DM. The s-channel production is a typical avenue for exploration, with the Z' then decaying into SM fermions ($pp \to Z' \to f\bar{f}$) leaving a resonant peak on an otherwise smoothly falling $m_{f\bar{f}}$ spectrum, or to DM with ISR (e.g. $q\bar{q} \to Z'j \to \chi\bar{\chi}j$). Since the experimental bounds are fairly stringent, the coupling between the Z' and the SM is required to be very small — suppressing the dark-matter annihilation into the SM, resulting in these models often predicting an overproduction of dark matter relative to the observations of Planck [95].

To simplify the search strategy, for simplified and complete models, a number of model parameters are set to values not excluded by previous measurements but still generate sufficiently large cross sections at the LHC, and leave one or two parameters to perform the search across (e.g. $m_{Z'}, g_q$).

3.3.4 Di-Higgs + $p_{\mathrm{T}}^{\mathrm{miss}}$ Phenomenology

The SM predicts non-resonant Higgs-boson pair production but contains no state that can be produced on shell and decay into two Higgs bosons; thus, any evidence for such a decay chain may indicate the existence of a broader scalar sector. Moreover, the production cross section of two Higgs bosons (resonant and non-resonant) in association with an imbalance of transverse momentum is negligibly small (e.g. $pp \to HHZ(\to \nu\nu)$ with $\sigma(pp \to Z(\nu\nu)HH) \sim 0.1$fb [52]) in the SM. However, a wide range of BSM phenomenology accounting for DM predicts the HH+$p_{\mathrm{T}}^{\mathrm{miss}}$ signature with a sizeable cross section. Therefore, a search with an excess of events in this region of phase space is strongly sensitive to new physics. Analyses within ATLAS [96] and CMS [97] have targeted the symmetric decay topology in the context of a gauge-mediated SUSY-breaking model, whereby two neutralinos are produced and each decays into a SM Higgs boson and a gravitino. The almost-massless

gravitino is taken to be the lightest supersymmetric state and provides the source of $p_{\mathrm{T}}^{\mathrm{miss}}$. The two Higgs bosons decay into two b-quarks, forming a $4b + p_{\mathrm{T}}^{\mathrm{miss}}$ final state.

The remainder of the chapter will focus on the dark Higgs model, which predicts the resonant production of $HH + p_{\mathrm{T}}^{\mathrm{miss}}$ and guides the development of the analysis presented later in Section 7 — which is the first search for resonant $HH + p_{\mathrm{T}}^{\mathrm{miss}}$.

The Dark Higgs Model

A dark Higgs mechanism [95] is introduced, where a complex scalar field S, charged under an additional $\mathrm{U}(1)^{'}$ gauge group, acquires a vacuum expectation value, giving rise to three additional massive states: a Majorana DM candidate χ, a vector mediator Z' and a dark Higgs boson s.

Starting with the free kinetic term for a DM candidate,

$$\mathcal{L} = \frac{1}{2} i \bar{\chi} \gamma^{\mu} \partial_{\mu} \chi, \tag{3.5}$$

and for a Majorana DM candidate the field satisfies $\chi = \chi^{c}$ with left and right-handed components oppositely charged under $\mathrm{U}(1)^{'}$. To ensure local gauge invariance, the covariant derivative is defined as,

$$D_{\mu} = \partial_{\mu} + i g_{D} q_{\chi} \gamma^{5} Z'_{\mu} \tag{3.6}$$

with the transformation,

$$Z' \mapsto (Z'_{\mu})' = Z'_{\mu} - \frac{1}{g_{D}} \partial_{\mu} \alpha(x), \tag{3.7}$$

leaving axial interaction terms between the Z' and the DM candidate, χ, of the form $-\frac{g_{\chi}}{2} \bar{\chi} \gamma^{\mu} \gamma^{5} \chi Z'_{\mu}$, defining $g_{\chi} = g_{D} q_{\chi}$. Purely axial couplings are significantly less constrained by direct detection experiments, since the dominant vector contributions to the spin-independent bounds are suppressed. A direct coupling is introduced between the Z', which is leptophobic, and SM quarks. Models typically require this, since searches for leptonically decaying Z' resonances at the LHC have proved unsuccessful, leading to stringent bounds on Z' mediators coupled to leptons [98].

Moreover, the number of annihilation pathways for dark matter is increased compared with simpler Z' models. Pathways such as $\chi\chi \to ss$ open up, and therefore

the dark Higgs model more naturally reproduces the observed dark matter relic density. It also provides more flexibility with respect to the benchmark, since the relic properties of the dark Higgs model are controlled by not only g_q, but also g_χ [95].

The Higgs mechanism, as described in Section 1.7.1, is used to provide massive mediators and massive DM, by introducing a complex singlet S, together with the SM Higgs doublet, to form the scalar potential,

$$V(H,S) = -\mu_H^2 H^\dagger H + \lambda_H (H^\dagger H)^2 - \mu_S^2 S^\dagger S + \lambda_S (S^\dagger S)^2 + \lambda_{HS}(H^\dagger H)(S^\dagger S). \quad (3.8)$$

where λ_H and λ_S are the Higgs and dark Higgs boson self-interactions, and λ_{HS} encodes the interaction between Higgs and dark Higgs bosons. Choosing the unitary gauge and expanding about the minima of the scalar potential leads to quadratic terms of the form,

$$\mathcal{L} = -\frac{1}{2}\begin{pmatrix} h' & s' \end{pmatrix} \begin{pmatrix} 2\lambda_H v^2 & \lambda_{HS} v w \\ \lambda_{HS} v w & 2\lambda_S w^2 \end{pmatrix} \begin{pmatrix} h' \\ s' \end{pmatrix} \quad (3.9)$$

diagonalising this matrix, we find the physical Higgs and dark Higgs bosons as mixtures of h' and s',

$$\begin{pmatrix} h' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} h \\ s \end{pmatrix}, \qquad \tan 2\theta = \frac{\lambda_{HS} v w}{\lambda_S w^2 - \lambda_H v^2} \quad (3.10)$$

where h' and s' are the interaction eigenstates and the physical mass eigenstates h and s correspond to the Higgs and dark Higgs bosons, respectively. Given the mixing, the dark Higgs can decay into SM states with equal branching fractions to a SM Higgs boson of different mass and for masses $m_s < 2m_\chi$, the decay into SM states is dominant.

The benchmark scenario adopted for Run 2 at the LHC fixes the dark matter mass to $m_\chi = 200$ GeV, and its coupling to the Z' set to $g_\chi = 1$ [99]. The coupling between SM quarks and the Z' is fixed at $g_q = 0.25$, and the mixing angle between the Higgs boson and the dark Higgs boson is set to $\sin\theta = 0.01$. Owing to the small value of $\sin\theta$, the phenomenology of the SM Higgs boson remains largely unaltered. The model introduces interactions of the form $2g_\chi m_{Z'} Z'^\mu Z'_\mu s$, which arise from the kinetic term in the Lagrangian after symmetry breaking, and relate

to the dark-Higgs strahlung process ($q\bar{q} \to Z'^{*} \to sZ'$). For high Z' masses, the cross-section at high-$p_{\mathrm{T}}^{\mathrm{miss}}$ for dark-Higgs strahlung, as shown in Figure 3.6, begins to become competitive with conventional monojet approaches, since large $p_{\mathrm{T}}^{\mathrm{miss}}$ can be generated by the recoil kinematics of the $Z's$ system.

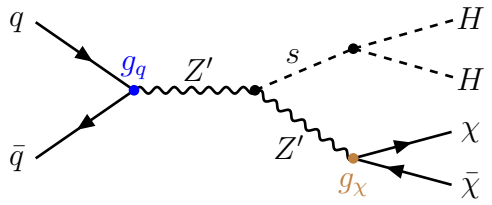


Figure 3.6: Feynman diagram for dark-Higgs strahlung in the dark Higgs model, coupling modifiers are shown for the Z' to quarks, g_q, and the coupling in the dark-sector, g_χ.

A plot of branching fractions as a function of dark Higgs mass is shown in Figure 3.7. For low values of the dark Higgs mass ($m_s < 2m_W$), its decay into two b-quarks is dominant and is searched for in the $b\bar{b}$+$p_{\mathrm{T}}^{\mathrm{miss}}$ final state by ATLAS [100] and more recently by CMS [101]. In the intermediate mass window ($2m_W < m_s < 2m_H$), the decay into vector bosons is favoured, and is searched for in the hadronic and semi-leptonic final states by ATLAS [99, 102] and CMS [103]. For dark Higgs masses $2m_h < m_s < 2m_\chi$, the di-Higgs final state becomes competitive with that of vector bosons and above $2m_t$ the decay into top-quark pairs opens up, but constitutes a small contribution to the branching ratio. Once the invisible decay into DM opens up via $s \to \chi\chi$ for $m_s > 2m_\chi$, this final state is favoured and the branching fraction quickly approaches unity.

A summary of the exclusion contours, as will be formally defined in Section 5.3, for the previous ATLAS searches is shown in Figure 3.8. The HH final state targets the upper-right quadrant of unexplored parameter space and is the final state targeted by the analysis presented in this thesis. Given the high-$p_{\mathrm{T}}^{\mathrm{miss}}$ final state and distinct kinematics of this process, the $HH \to 4b$ process is chosen to maximise signal statistics, resulting in a $4b + p_{\mathrm{T}}^{\mathrm{miss}}$ final state — which is searched for using the full Run 2 dataset collected by the ATLAS experiment at the LHC.

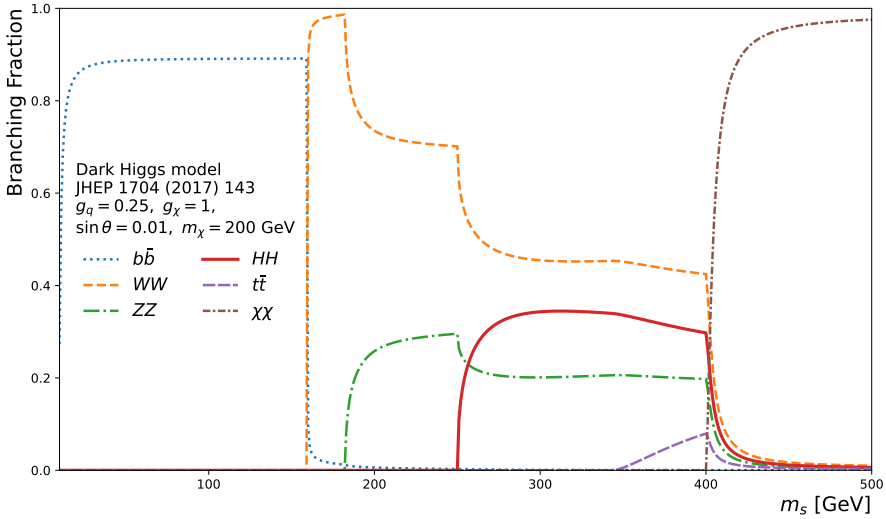


Figure 3.7: Branching fractions of the dark Higgs boson for a range of masses in the Run 2 benchmark. The $b\bar{b}$ final state dominates at low-mass $m_s < 2m_W$, with the decay into vector bosons in the window $2m_W < m_s < 2m_H$, after which the di-Higgs final state opens up, and the branching fractions are comparable between the VV and HH channels, favouring the WW final state broadly over the mass window $2m_H < m_s < 2m_\chi$. Once the channel $s \to \chi\chi$ opens up, the branching fraction rapidly approaches unity, diminishing the sensitivity for the model.

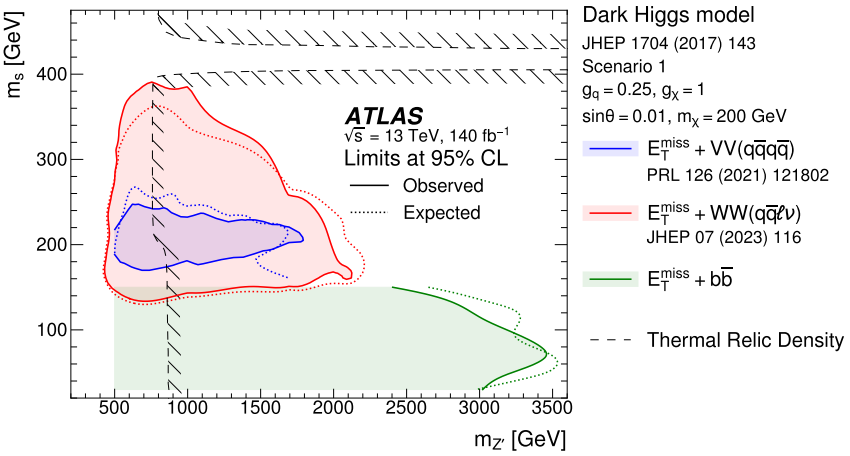


Figure 3.8: Summary of exclusion contours, as will be formally introduced later in Section 5.3, for all previous ATLAS dark Higgs analyses. The dark Higgs analysis presented in Section 7 targets the upper-right quadrant, which is not excluded by previous measurements. Figure from Ref [100].

4

The ATLAS Experiment

Contents

4.1 The Large Hadron Collider 40
4.1.1 Acceleration and Magnet Technology 41
4.1.2 Interaction Regions 42
4.1.3 Coordinate System 43
4.1.4 Luminosity and Pile-up 44
4.2 The ATLAS Detector 45
4.2.1 Inner Detector 47
4.2.2 Calorimeters 50
4.2.3 Muon Spectrometers 54
4.3 Objects in the ATLAS experiment 55
4.3.1 Tracks 55
4.3.2 Vertex Reconstruction 56
4.3.3 Calorimeter Clusters 56
4.3.4 Jets 57
4.3.5 Jet Calibration 59
4.3.6 Pile-up Mitigation 61
4.3.7 *b*-jets 61
4.3.8 Muons 63
4.3.9 Missing Transverse Momentum 63
4.3.10 ATLAS Trigger 64

4.1 The Large Hadron Collider

The European Organisation for Nuclear Research (CERN), located in Meyrin, Geneva, on the Swiss-French border, has been at the forefront of experimental particle physics since its formation in 1954. Numerous accelerator complexes have operated throughout its history, feeding high-energy particles into experimental caverns for analysis. The following chapter will outline the accelerator complex required for Run 2 of the LHC [104] to collide protons and describe the ATLAS experiment [105] that records the resultant proton-proton collisions.

4.1.1 Acceleration and Magnet Technology

Protons are prepared and accelerated before injection into the 27 km LHC ring. First, hydrogen is ionised, and the resultant protons are accelerated by a linear particle accelerator (LINAC2 for Run 2), with the most recent iteration, LINAC 4 [106] installed in 2020 for Run 3 and the High-Luminosity LHC (HL-LHC), which instead accelerates negative hydrogen ions. The protons are accelerated by a series of radiofrequency (RF) cavities, and controlled into a tight beam using quadrupole magnets. At injection, the protons enter the Proton Synchrotron (PS) Booster, and are accelerated in four rings of 25 m from 50 MeV to 1.4 GeV. The PS Booster directly feeds the PS, which accelerates protons up to 26 GeV into the Super Proton Synchrotron (SPS), which brings the beams to 450 GeV. Finally, the proton bunches enter the LHC via two beryllium vacuum beampipes travelling clockwise and anti-clockwise. They are ramped up to 6.5 TeV in Run 2 using a series of 8 RF cavities per beam and kept within a circular trajectory with 1232 superconducting dipole magnets (each 8 T, operated at 1.9 K).

The maximum achievable centre-of-mass energy is dictated by both the ring radius and the bending magnetic field. A higher energy collider requires a larger ring or a higher dipole field, or both. The LHC was originally designed to collide protons at 14 TeV. However, operation at the design energy was delayed due to issues with the superconducting magnet interconnections, as well as the need for magnet consolidation, commissioning, and training at the required operating current [107]. Run 1 operated at $\sqrt{s} = 7$ and 8 TeV, corresponding to dipole fields of 4.2 and 4.8 T respectively. The energy was ramped up to 13 TeV (7.7 T) in Run 2, 13.6 TeV (8.1 T) in Run 3. Following the HL-LHC upgrades, the LHC is expected to run at 14 TeV with 8.33 T dipole magnets. High-field superconducting magnets have historically been a central technical limitation for circular hadron colliders, and remain one of the defining challenges for future colliders. A feasibility study for measuring triple-Higgs production at the Future Circular Collider (FCC), specifically the hadron–hadron stage (FCC-hh), is presented later in Section 8. The target centre-of-mass energy is $\sqrt{s} = 84$ TeV within a 91 km ring. Even under the idealised

assumption that the full circumference is available for bending, the required dipole field would be approximately 10 T. In a more realistic scenario, where only part of the circumference is occupied by dipole magnets, the required field is closer to 14 T [108]. Such fields may be achievable using superconductors such as Nb_3Sn, or potentially through future high-temperature-superconductor technology. Nb_3Sn was studied during the development of the LHC, for which NbTi was ultimately chosen, and is now being deployed in the HL-LHC upgrade for the new inner-triplet quadrupoles. However, thousands of these magnets will be required for operation, and large-scale manufacturing, together with the stability, longevity, and readiness of high-current magnet wires, poses a significant challenge.

4.1.2 Interaction Regions

Inner triplet quadrupole magnets are used to cross two proton bunches (a bunch-crossing) every 25 ns at four interaction points along the LHC beamline. These four interaction points house the LHC experiments: Large Hadron Collider beauty [109] (LHCb), A Large Ion Collider Experiment [110] (ALICE), Compact Muon Solenoid (CMS) [111] and finally A Toroidal LHC Apparatus (ATLAS). LHCb targets the forward region, where the production of b-hadrons is enhanced, with a pseudo-rapidity coverage of $2 < \eta < 5$. LHCb exploits highly efficient particle identification to perform high-precision measurements of heavy-flavour physics, CP violation and the discovery of a vast array of new hadrons. The LHC is also capable of accelerating and colliding heavy ions (e.g. iron, lead, oxygen), and the ALICE experiment primarily focuses on measurements of these collisions — where highly dense environments enable the study of the quark-gluon plasma (QGP). ATLAS and CMS are general-purpose detectors designed for a wide range of physics, like the discovery of the Higgs boson and, with almost full solid angle coverage, searches for a large number of BSM states and phenomena. Outside of the four main LHC beamline experiments, downstream of ATLAS are: the LHC forward (LHCf) experiment [112], designed to constrain cosmic-ray modelling using forward decay products; the ForwArd Search ExpeRiment (FASER) [113], designed to investigate

the existence of possible long-lived particles produced at the ATLAS interaction point; and the Scattering and Neutrino Detector (SND) [114], to measure the flux of neutrinos from the ATLAS interaction point, alongside feebly interacting particles.

4.1.3 Coordinate System

The coordinate system [105] used by the ATLAS experiment is right-handed, with the origin at the nominal interaction point. The direction of the beam defines the $\hat{z}$-direction; the direction pointing into the centre of the ring is taken as $\hat{x}$, and $\hat{y}$ as vertically upwards. The azimuthal angle, ϕ, is measured around the beam axis, and the polar angle, θ, is measured from the positive $\hat{z}$-axis.

The rapidity of a particle is defined as $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$, and in the ultra-relativistic regime this approaches the massless limit, pseudo-rapidity:

$$\eta = -\ln\tan\frac{\theta}{2}. \tag{4.1}$$

The difference in rapidities between two particles is invariant under longitudinal boosts. This invariance is essential for hadron colliders, since the momentum balance along $\hat{z}$ varies for each collision depending on each parton's momenta. Thus, distances in collider events are written as,

$$\Delta R^2 = \Delta\phi^2 + \Delta\eta^2 \tag{4.2}$$

and strictly only longitudinally boost-invariant using Δy; however, the use of η is an excellent approximation for highly relativistic reconstructed particles.

The four-vector of a particle with known mass can be fully expressed with a measurement of p_{T}, y and ϕ:

$$p^\mu = (m_{\mathrm{T}}\cosh y, p_{\mathrm{T}}\cos\phi, p_{\mathrm{T}}\sin\phi, m_{\mathrm{T}}\sinh y) \tag{4.3}$$

where m_{T} is the transverse mass and for massless particles: $m_{\mathrm{T}} \mapsto p_{\mathrm{T}}$, $y \mapsto \eta$.

4.1.4 Luminosity and Pile-up

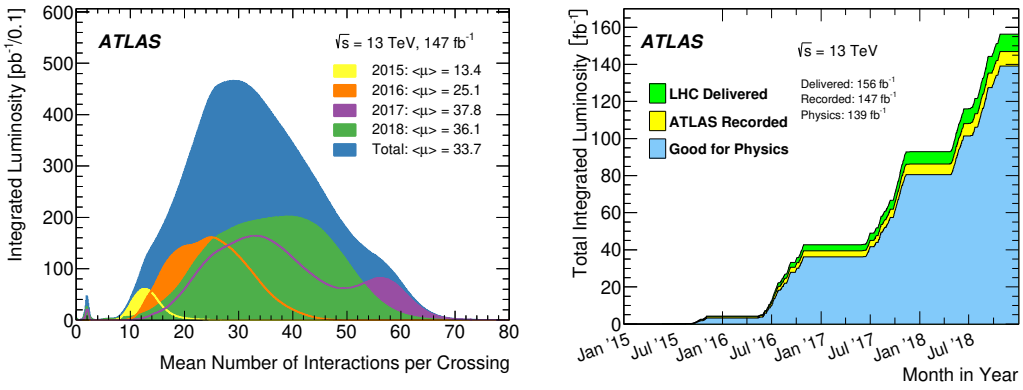
The instantaneous luminosity of two colliding beams is given by,

$$\mathcal{L} = \frac{N_1 N_2 f_{\text{rev}} n_b}{4\pi\sigma_x\sigma_y} \tag{4.4}$$

where N_1, N_2 are the number of protons in each colliding bunch, f_{rev} is the revolution frequency of the beams, n_b is the number of bunches, and the transverse size of the beam is parametrised by Gaussian widths σ_x, σ_y.

For all experiments, the tightness of each colliding bunch at the interaction region is an important consideration. Tighter bunches result in increased instantaneous luminosity, thereby enabling more data to be collected. This is not without cost, since the number of additional proton-proton interactions per bunch-crossing increases, known as pile-up. Pile-up can create busy environments for reconstruction, such as overlapping hits in the tracker or multiple particles depositing energy within the same calorimeter cell. Each experiment must therefore optimise the pile-up profile between the amount of data collected, the radiation dose the detector will be exposed to and the achievable quality of event reconstruction.

With the smaller radiation dose that a low pile-up environment provides, the initial layer of the tracking volume may be placed closer to the beamline — enabling precise vertexing to be performed. LHCb targets highly precise hadron reconstruction and identification, and thus benefits greatly from a clean environment with minimal pile-up (levelled at $\mu = 1.1$ during Run 2 [115]). However, the ATLAS and CMS experiments, with physics aims such as the discovery of the Higgs boson over a wide range of possible masses or searches for BSM states, require large datasets and therefore operate at much higher luminosity and therefore pile-up. During typical Run 2 proton-proton operation, each bunch contained approximately 1.1×10^{11} protons, with around 2500 bunch pairs colliding at a revolution frequency of 11.245 kHz. At the ATLAS and CMS interaction points, the beams were focused to transverse Gaussian widths of approximately 10 μm, producing levelled instantaneous luminosities of order $\mathcal{O}(10^{34})\text{cm}^{-2}\text{s}^{-1}$. The pile-up profile of ATLAS during Run 2 is shown in Figure 4.1(a), where the gradual



(a) **(b)**

Figure 4.1: The mean number of interactions per bunch-crossing (pile-up) of the ATLAS experiment during Run 2 of operation (a) alongside the total integrated luminosity as a function of time across Run 2 (b) with the proportion of delivered, recorded and good-quality data shown in green, yellow and blue. Figures taken from Ref [116].

ramping up of instantaneous luminosity (with higher pile-up) enables larger datasets to be collected, as indicated by the increasing slope of total integrated luminosity collected across Run 2 in Figure 4.1(b).

4.2 The ATLAS Detector

A general-purpose detector must be capable of measuring a wide range of SM processes and of searching for a variety of BSM phenomena — requiring good momentum resolution for both high and low momentum particles and highly efficient lepton, b and τ-identification across a wide kinematic range.

To identify and measure the momentum of charged particles, a central tracking detector is immersed within a solenoidal magnetic field. Charged particles bend in this magnetic field and ionise sensors in the tracker, producing a series of deposits along a helical trajectory. Neutral particles do not ionise in the silicon tracker and thus no track can be resolved[1]. However, their measurement is crucial for all measurements involving photons and jets. Approximately one third of jet energy is carried by neutral constituents, so a significant portion of jet momentum is left

[1]They *can* deposit energy through the strong interaction, or conversions of photons, within the tracking material — but the design of thin silicon pixels means that this is a relatively sub-leading effect.

completely unresolved within a tracking volume alone [117]. Moreover, even for charged particles, the momentum measurement relies on the track curvature in a magnetic field and becomes imprecise for high-momentum particles, since the curvature is so slight. Both issues are addressed through the use of calorimeters. Here, highly dense passive layers (lead, for example) initiate a shower of charged particles to be absorbed and measured by an active layer (for ATLAS, a liquid argon or plastic scintillating tile layer).

The distinct differences in energy deposition between electrons/photons and hadrons motivate the use of two separate calorimeter systems. A compact and granular electromagnetic calorimeter, primarily tasked with measuring the shower evolution of electrons and photons, and a deeper hadronic calorimeter used to measure hadronic particles within jets, whose constituents undergo both electromagnetic (for charged constituents) and strong interactions with the calorimeter material.

Finally, since the emission of photons from a charged particle whilst in the vicinity of an atomic nucleus, *bremsstrahlung*, is suppressed with the square of the parent particle mass, muons deposit comparatively little energy within the calorimeters. Muons also have lifetimes around 2 μs and will not often decay within the inner tracker and calorimeter volume. Muon spectrometers (MS) then surround the detector. Here, a particle that registers hits in the inner tracker, deposits only a small fraction of its energy in the calorimeters, and is later measured in the MS is highly likely to be a muon.

The remainder of this chapter will detail each of the elements of the ATLAS detector shown schematically in Figure 4.2, with particular focus on the detector elements that are instrumental in reconstructing the physical objects used throughout the thesis: tracks, jets, b-jets and missing transverse momentum.

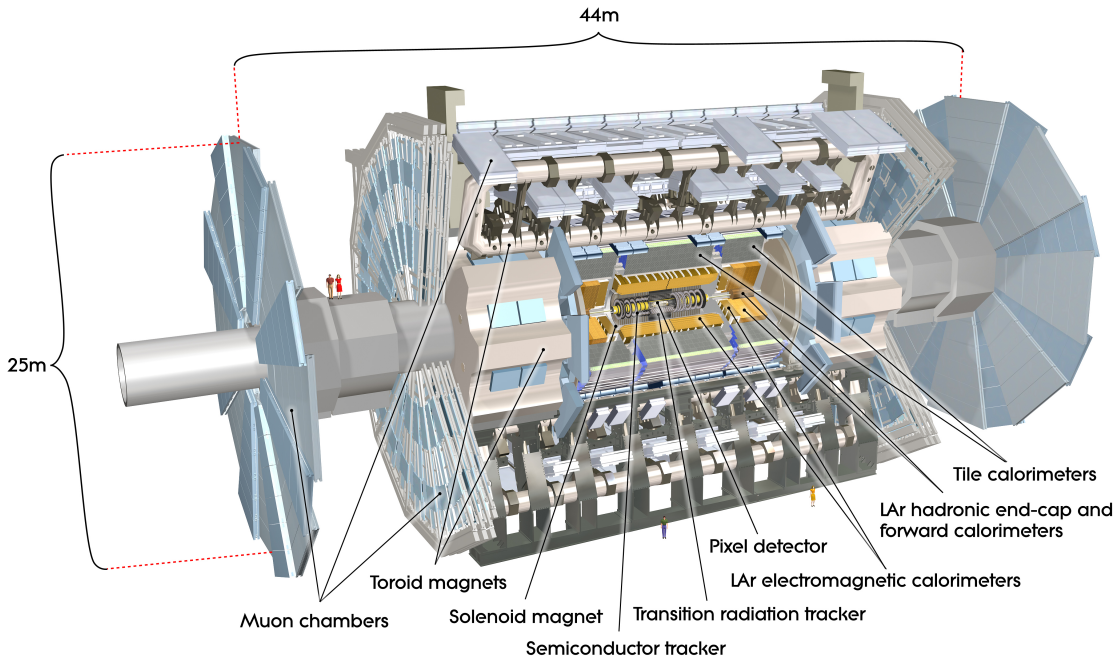


Figure 4.2: An overview of the ATLAS detector. Starting from closest to the beamline, the labelled components are the pixel detector, semiconductor tracker (SCT), transition radiation tracker (TRT), liquid argon (LAr) electromagnetic calorimeter (ECAL), tile hadronic calorimeters (HCAL), the muon chambers and in the forward regions the LAr HCAL end-cap and forward calorimeters (FCAL). Figure taken from Ref [105].

4.2.1 Inner Detector

The tracking volume of the ATLAS detector, the inner detector, measures ionisation deposits left by charged particles as they bend in a solenoidal magnetic field. ATLAS utilises silicon semiconducting sensors, in which charged particles entering a sensor will liberate electrons (approximately 3.6 eV is required) from the valence to the conduction bands. The resultant electron-hole pairs are made to drift using an applied voltage across the sensor, and if the number collected exceeds a set threshold[2], a *hit* is recorded. Hits across multiple sensor layers can be used to form helical-track trajectories through the magnetic field. The inner detector (ID) is placed to cover the central region of the interaction point ($|\eta| < 2.5$), striking a compromise among physics coverage, radiation tolerance, and pragmatic cost considerations.

[2]The exact threshold is tuned per sensor, and varies over time with radiation damage, but is typically thousands of electron-hole pairs.

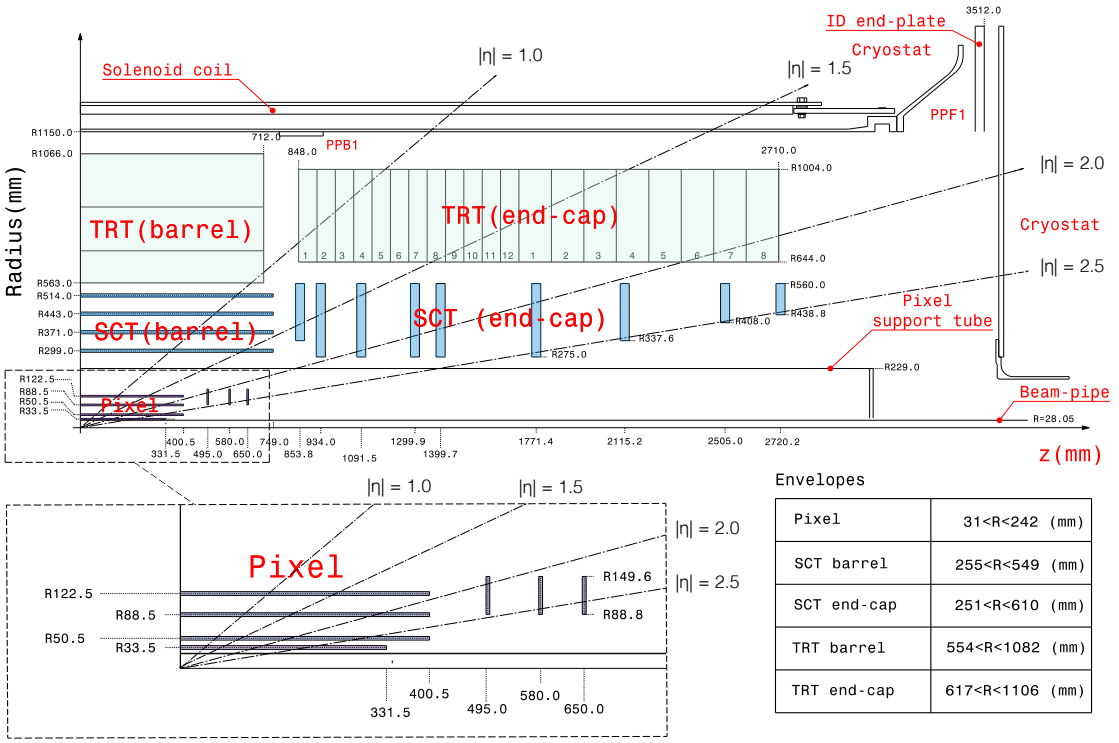


Figure 4.3: An overview of the ATLAS ID, with the Pixel Detector, SCT and TRT shown with the radius from the beamline, displacement along the $\hat{z}$ direction and example pseudo-rapidities overlaid. Figure taken from Ref [119].

The ID is embedded in a 2 T axial magnetic field to measure charged-particle trajectories between radii of 33.5 and 1066 mm from the beampipe. It is comprised of three main sub-systems: the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). In total, it is comprised of over 90 million readout channels[3]. The pseudo-rapidity coverage of the ID is shown in Figure 4.3, with the corresponding barrel and end-cap regions shown.

Pixel Detector

The Pixel Detector, consisting of 1744 silicon pixel modules, is the innermost system and covers the radial distance between 31 mm and 242 mm. The placement of the innermost layer is important for vertex reconstruction, and a compromise between radiation damage, occupancy and vertexing performance is made. In preparation for the increase in luminosity following Run 1, ATLAS installed the Insertable Barrel

[3]Remarkably, the inner tracker (ITk) upgrade[118], replacing the ID to cope with the high pile-up environment at the HL-LHC, is comprised of over 5 billion readout channels!

Layer (IBL) at a radial distance of 33 mm [120, 121]. The IBL comprises 14 staves, each consisting of 20 sensor modules. Modules house the nearly 12 million total pixel sensors of the IBL, and, with tracking information in the transverse plane being most relevant, the shorter dimension (50 μm) is oriented in the $R\phi$ direction, with the longer along $\hat{z}$ (250 μm). The remaining three layers, the B-layer, Layer-1 and Layer-2, are made of pixel modules. A pixel module consists of one silicon sensor, with 46080 pixels, and 16 front-end chips that are bump-bonded to the sensor, each with 2880 readout channels — enabling a full module readout. Silicon pixel modules are loaded onto three concentric barrel layers at 50.5 mm, 88.5 mm and 122.5 mm and two endcaps with three layers.

Semiconductor Tracker (SCT)

The SCT consists of 4088 modules of silicon strips arranged in four concentric barrels and two endcaps, each with nine discs. A second identical sensor is glued back-to-back with the first at a stereo angle of 40 mrad. This enables a 2D point to be defined as the intersection between the two strips with registered hits. The barrel consists of 2112 modules mounted on barrels at radii: 299 mm, 371 mm, 443 mm and 514 mm respectively. Each endcap consists of 988 modules mounted on 9 disks. The layout is such that any particle will traverse four layers of the SCT, as shown in Figure 4.3, and, together with a back-to-back layout, provide four spatial points for reconstruction per track.

Transition Radiation Tracker (TRT)

The TRT is the outermost layer of the ID between $563 < r < 1066$ mm for the barrel and $644 < r < 1004$ mm in the end-caps. Structurally, it consists of 298304 (4 mm diameter) gas-filled drift (straw) tubes interleaved with radiation-absorbing material, polypropylene or polyethylene. The radiation material induces transition radiation, since the tubes and radiator have different dielectric constants. The yield of photons depends on $\gamma = \frac{E}{mc^2}$ and thus enables particle ID to be performed for high-energy charged particles, notably between electrons and charged hadrons. To efficiently

capture photons from transition radiation, a gas mixture consisting predominantly of a heavy noble gas (70% Xe, 27% CO_2 and 3% O_2) is used for the straw tubes[4].

Inner Detector Performance

The inner detector subsystems, together with their pseudo-rapidity coverage, element sizes and intrinsic resolutions are provided in Table 4.1. The introduction of the IBL, 17 mm closer to the interaction point than the B-layer, was vital for strong b-jet identification performance during Run 2[5]. The overall momentum resolution of the ATLAS inner detector was designed[6] to reach [105],

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = 0.05\% p_{\mathrm{T}}[\mathrm{GeV}] \oplus 1\% \tag{4.5}$$

where the first term reflects the challenge of resolving curvature for high-momentum particles and the 1% can be attributed to multiple-Coulomb scattering with the inner detector material.

Table 4.1: Summary of the main characteristics of the ATLAS Inner Detector subsystem.

Subsystem	Pseudorapidity coverage	Element size	Intrinsic resolution [μm]
IBL	$\lvert\eta\rvert < 3.0$	(50×250) μm	10×60
Pixel	$\lvert\eta\rvert < 2.5$	(50×400) μm	10×115
SCT	$\lvert\eta\rvert < 2.5$	80 μm	17
TRT	$\lvert\eta\rvert < 2.0$	4 mm	130

4.2.2 Calorimeters

As introduced briefly earlier, the distinctive properties of electrons/photons and hadrons motivate two separate calorimeter systems for optimal measurement performance of each. An electromagnetic calorimeter (ECAL) is placed directly following the inner detector, and seeks to identify and measure the energy of electrons and photons. Hadronic calorimeters (HCAL) follow the ECAL, and seek

[4]Due to gas leaks, much of the xenon was replaced with argon during Run 2 and 3, degrading performance.

[5]For example, a close to four times enhanced light-jet rejection was achieved for a fixed 77% b-jet tagging efficiency [122].

[6]This resolution function is particle dependent and this parametrisation was derived using charged pions.

to measure all hadronic activity in the event, which will deposit a fraction of energy within the ECAL, but are often measured predominantly within the HCAL. Both the ECAL and HCAL follow a sampling calorimeter design, where alternating layers of dense passive layers initiate a spray of charged particles, which are measured as ionisation within an active layer. An overview of the ATLAS calorimeter system is shown in Figure 4.4, where the central region is covered by the electromagnetic and hadronic (tile) barrel calorimeters, the forward region with ECAL / HCAL end-caps and a liquid-argon (LAr) forward calorimeter.

Electromagnetic Calorimeters

Electrons can dissipate energy in a medium through several mechanisms, with the dominant process depending on the electron energy. For electron energies above $\sim 10 - 20$ MeV, the dominant mechanism are photons emitted as bremsstrahlung. Photons then pair-produce electrons and positrons, and the process continues iteratively to form an electromagnetic shower. For prompt photons, this begins with pair-production and continues as above. Once the shower reaches energies less than 10 MeV, the dominant process is via ionisation, and this proceeds until the energy is fully depleted within the calorimeter. An electromagnetic calorimeter (ECAL) is designed so that, for a range of interesting electron and photon energies ($\sim 10^0 - 10^2$ GeV), the shower is fully contained within the ECAL material. In contrast, hadronic showers penetrate further into the hadronic calorimeter (HCAL).

The passive material should have a high atomic number to induce a sufficiently large shower with a short radiation length, optimised such that numerous passive/active layers may be placed to improve the measured energy resolution. The active material should then have the appropriate properties to accurately measure the deposited signal. A noble gas achieves this well, and an active layer of liquid argon (LAr) with a passive layer of lead is chosen for the ECAL. LAr is beneficial for its linear energy response (meaning that a particle at 1 GeV and 100 GeV should deposit an approximately fixed fraction of its energy, which is important for universal calibrations as will be discussed later) and for its radiation hardness. This

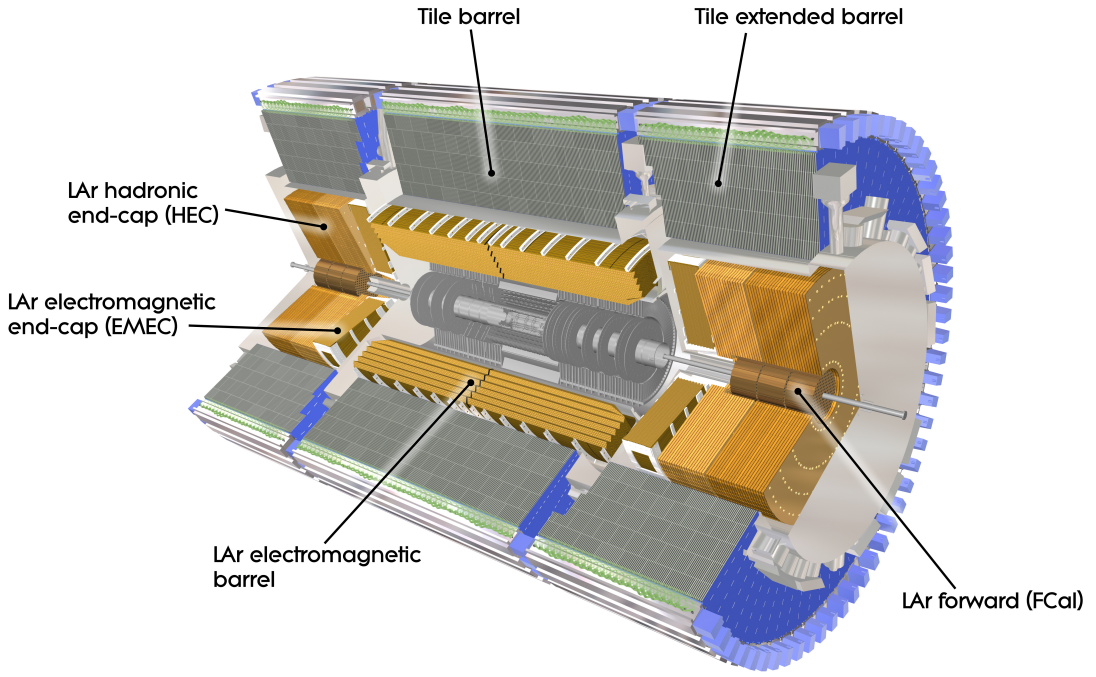


Figure 4.4: An overview of the ATLAS Calorimeter subsystems. Figure taken from Ref [124].

property will also be beneficial for the forward hadronic calorimeter (FCAL). Rather than having sequential calorimeter segments, ATLAS arranges the ECAL with an accordion geometry to ensure the response of the calorimeter is uniform in ϕ. Other philosophies in which the calorimeters are uniformly radially segmented enable more granular calorimetry for potential imaging calorimeters in future electron-positron detectors [123]. The electromagnetic calorimeter is divided into a barrel section (EMB), covering $|\eta| < 1.475$, and two end-cap sections (EMEC), covering $1.375 < |\eta| < 3.2$. Both the barrel and end-cap calorimeters are longitudinally segmented into three layers, the first with fine granularity to improve separation between neutral pions producing a $\gamma\gamma$ system and singular isolated photons. To cover the scenarios in which a particle interacts with material before the ECAL, a presampler is placed in the $|\eta| < 1.8$ region. The EM barrel is contained within a barrel cryostat, which houses the EMB and the 2 T solenoid.

Hadronic Calorimeters

Following the ECAL, the hadronic calorimeter (HCAL) seeks to fully absorb and measure the remaining hadronic particles following a collision. The Tile Calorimeter is a sampling calorimeter with a passive steel layer followed by active plastic scintillating tiles, and covers radially between 2.28 m and 4.23 m. Three-layers are mounted onto a central barrel, and onto two extended barrels, each with 64 azimuthal modules of width $\Delta\phi = 2\pi/64$. Charged particles, initiated by interactions within the steel layer, excite molecules within the plastic scintillating tile, and the de-excitation produces photon signals, which are transported by wavelength-shifting fibres to photo-multiplier tubes, where the received signal is amplified. Ttwo wheels of copper followed by LAr cover the forward region ($1.5 < |\eta| < 3.2$). Forward calorimeters cover $3.2 < |\eta| < 4.9$, with tungsten passive layers followed again by LAr.

Calorimeter Performance

The energy performance of a calorimeter can be parametrised as a function of the leading effects,

$$\frac{\sigma_E}{E} = C \oplus \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \tag{4.6}$$

where C, S and N are the constant, stochastic and noise terms respectively. As can be gathered by the energy dependence; the noise term (N), originating from electronic noise in the sensors, dictates the energy resolution at low energy; the stochastic term (S), in the intermediate range of energy, mainly represents fluctuations in shower development and in the fraction of energy deposited in the active layers; and the constant term (C) dominates at high energies, accounting for variations of the calorimeter response across its volume. The design specifications for the ATLAS detector for each term are provided in Table 4.2.

Table 4.2: Design energy-resolution parameters of the ATLAS calorimeters. The lack of a noise term specified does not mean the hadronic calorimeters are noiseless, but rather that, for the energies of jets reaching the HCAL, the noise term contributes negligibly [105].

Calorimeter region	Constant term	Stochastic term	Noise term
Electromagnetic	0.7%	$10\%\sqrt{\text{GeV}}$	~ 0.2 GeV
Hadronic ($\|\eta\| < 3.2$)	3%	$50\%\sqrt{\text{GeV}}$	–
Forward ($3.2 < \|\eta\| < 4.9$)	10%	$100\%\sqrt{\text{GeV}}$	–

4.2.3 Muon Spectrometers

The muon spectrometer (MS) is the outermost layer of the ATLAS detector, designed to identify and measure the momentum of muons. It comprises high-precision tracking chambers covering $|\eta| < 2.7$, together with dedicated trigger chambers, arranged within a toroidal magnetic field providing a bending integral of approximately 2.5–6 T, m. There are three regions of magnetic field configuration: a large eight-coil barrel magnet surrounding the HCAL covers $|\eta| < 1.0$, and two end-cap toroidal magnets cover $1.4 < |\eta| < 2.7$. The transition region between $1.0 < |\eta| < 1.4$ is covered by a combination of the two fields.

Three stations of muon drift tubes (MDT) chambers are used in the central region, and the innermost MDT is replaced by a cathode-strip chamber (CSC) in the forward $2 < |\eta| < 2.7$ region. Within the MDTs, muons will ionise the gas atoms, and the liberated electrons drift to, and are collected by, a tungsten-rhenium wire placed in the centre of the tube, analogously to the design philosophy of the TRT. CSCs are comprised of 32 four-layer chambers, and muons will ionise the gas mixture, and the liberated electrons will again drift to the closest wire in the chamber, producing an avalanche, for which the sampled charge is proportional to the ionisation induced by the muon. Spatial resolutions of 80(60) μm are achieved in the MDTs (CSCs).

The drift time within the MDTs and CSCs is typically around 700 ns, significantly larger than the bunch-crossing spacing of 25 ns. Therefore, dedicated trigger systems are deployed for the muon spectrometer, where resistive plate chambers (RPCs) and thin-gap chambers (TGCs) in the barrel and end-cap are used, respectively.

4.3 Objects in the ATLAS experiment

The ATLAS detector in Run 2 had $\mathcal{O}(10^8)$ readout channels (dominated by the pixel detector), and this enormous space of information is often impractical for use in physics analysis as an ensemble. Frequently, the information within a particular detector system is compressed to form an *object*. This section will outline the objects relevant to the analyses presented in this thesis, including: tracks, calorimeter clusters, jets, b-jets and muons. The principle of b-tagging will be briefly summarised here, with a more detailed account within ATLAS explored later in Section 6. The reconstruction of τ leptons will be briefly discussed later in Section 8. Further information regarding electron and photon reconstruction in Run 2 can be found within Ref [125].

4.3.1 Tracks

A charged particle will leave spatially correlated deposits across the various ID subsystems. Raw measurements from pixels and strips in the ID are clustered according to the amount of charge deposited in neighbouring elements, and form three-dimensional space points known as *hits*. Seeds, consisting of three hits, are built from the Pixel Detector and SCT layers to begin forming a helical track trajectory. This seed is then extended to include the remaining layers of the ID using a combinatorial Kalman filter, which is flexible and can branch to form multiple track candidates.

To reduce the contribution of tracks arising combinatorially from an unphysical set of seeds (multiple tracks associated with the same hit, for example), an ambiguity-solving algorithm is applied whereby the tracks are ranked according to several metrics, and lowly-ranked tracks that overlap with too many clusters of a higher-ranked track are removed. To aid the ambiguity-solving task for highly collimated sprays of particles, such as those produced by a high-momentum b-jet, a neural network is deployed to identify clusters consistent with two charged particles. Here,

the measured charge, position of pixels within a cluster and the incident angle of the particle are used to identify merged clusters.

Tracks are parametrised using a perigee-based formalism, defined as the point where the projection of the fitted track trajectory onto the transverse plane passes closest to the beamline. A track is defined with a set of 5 parameters: (transverse impact parameter d_0, longitudinal impact parameter z_0, $\phi, \theta, \frac{q}{p}$). The final candidates enter a track-fitting procedure, minimising the difference between predicted and observed hit positions along the fitted trajectory to extract the final perigee coordinates [126].

4.3.2 Vertex Reconstruction

In the presence of multiple pile-up interactions along the beamline, accurately determining the vertex position of the hard-scatter interaction is vital for downstream reconstruction. An iterative vertex finding algorithm associates tracks above 500 MeV to a possible vertex position, and a vertex fitting procedure is performed to extract a position vector for each vertex, $\boldsymbol{r}_{\mathrm{PV}}$. The hard-scatter vertex is then defined as the primary vertex with the largest sum of square track-p_{T} [127].

4.3.3 Calorimeter Clusters

Particles will often deposit energy across multiple calorimeter cells and layers. To begin reconstructing such deposits, calorimeter cells are combined to form composite objects, called *proto*-clusters. Proto-clusters are formed from seed cells with a signal significance above four times the expected cell noise, $\mathcal{S} = \frac{E_{\mathrm{cell}}}{\sigma_{\mathrm{cell}}}, |\mathcal{S}| > 4$. The significance can be negative because the ATLAS calorimeter readout uses a $\mathtt{CR\text{-}RC}^2$ shaping filter, meaning out-of-time pile-up occurring more than ∼100 ns before the triggered event can therefore produce a negative reconstructed cell signal [128]. Including negative cells allows cancellation of positive and negative noise fluctuations. These seed cells then follow an iterative algorithm, where neighbouring deposits with significance above 2 are sequentially added to the initial

cluster until a perimeter is set of clusters with $0 < |\mathcal{S}| < 2$. The growth of topo-clusters is fully three-dimensional, and cells in the next layer of the calorimeter are also considered for combination.

The preceding cluster growth does not account for structure, such as local maxima, which could indicate the presence of multiple particles within a single proto-cluster. Information from the high-granularity ECAL is used for this purpose, and local maxima are defined as $E_{\text{cell}} > 500$ MeV, with all four neighbours being lower energy. The resulting split clusters are known as *topo*-clusters.

Hadronic showers can lose energy through mechanisms not measured by the ECAL or HCAL (e.g. neutrino production or atomic capture of low-energy hadrons). The response of the ATLAS calorimeters therefore depends on the mechanism that produced the energy deposit: whether it was electromagnetic or hadronic. This difference is corrected for using a local cell weighting (LCW) calibration, which is also used to correct out-of-cluster signal losses and dead material effects. First, the probability of a deposit originating from an electromagnetic interaction is defined, where the depth and density of the shower throughout the cell are used to differentiate between electromagnetic and hadronic. The depth and average cell density of the topo-cluster, binned in $(E^{\text{EM}}_{\text{clus}}, \eta_{\text{clus}})$, are the most differentiating variables and a probability of an EM deposit is formed, $\mathcal{P}^{\text{EM}}_{\text{clus}}$. To provide a smooth calibration, this probability is used in a weighting procedure [129]:

$$w_{\text{cell}} = \mathcal{P}^{\text{EM}}_{\text{clus}} \cdot w^{\text{EM-calib}}_{\text{cell}} + (1 - \mathcal{P}^{\text{EM}}_{\text{clus}}) \cdot w^{\text{Had-calib}}_{\text{cell}} \tag{4.7}$$

where $w_{\text{cell}} = \frac{E^{\text{deposit}}_{\text{cell}}}{E^{\text{EM}}_{\text{cell}}}$, and is defined as 1 for $w^{\text{EM-calib}}_{\text{cell}}$ by construction. Clusters measure the deposits of single, and occasionally multiple, particles and can be combined into objects known as *jets* — which are used to reconstruct the underlying hard-scatter interaction.

4.3.4 Jets

Quarks and gluons, due to confinement, cannot exist as free coloured particles and instead undergo a parton shower, as introduced in Section 2.1.2. Once

sufficiently low-energy, the shower of partons hadronises to form colour-neutral hadrons, forming a jet.

The clustering of stable final state particles to form jets has historically been performed using a number of algorithms. Theoretically, jet definitions should be invariant to the emission of infinitely soft or collinear gluons, known as infrared and collinear (IRC) safety. Experimentally, a jet algorithm should be computationally efficient, and the boundaries of such jets should be invariant under soft emission. The latter ensures that consistent jet calibration can be performed. These conditions are satisfied by the anti-k_t jet clustering algorithm [130].

Two distances are used within the algorithm; the distance between two *pseudo*-jets within the detector and the distance between a pseudo-jet and the beam:

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T},i}^2}, \frac{1}{p_{\mathrm{T},j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = p_{\mathrm{T},i}^{-2} \tag{4.8}$$

where $p_{\mathrm{T},i}$ is the transverse momentum of particle i, the parameter R defines the radius of the jet, and for almost all studies within this thesis, a jet radius of 0.4 is chosen. For each pseudo-jet (i), all pairwise distances (d_{ij}) are formed, together with the distance to the beam, d_{iB}. For the smallest d_{ij}, which is also less than d_{iB}, ij are clustered together and re-enter the algorithm with updated kinematics. Once all pairwise distances for a given pseudo-jet are greater than d_{iB}, this object is considered a final jet. The inverse square distance measure in anti-k_t ensures the hardest objects are clustered first and defines a boundary, with radius R, for which softer particles cluster around. This algorithm continues until the full event is clustered.

The input objects that enter the anti-k_t jet clustering algorithm within ATLAS are particle-flow (PFlow) objects [117]. PFlow reconstruction seeks to exploit the strong low-p_{T} track momentum and angular resolution of the ID combined with the strong high-p_{T} cluster-based resolution of the calorimeters, to maximise overall jet reconstruction performance.

The PFlow algorithm begins with tracks, as defined in Section 4.3.1, and attempts to match each track to a single topo-cluster. Using the measured track

momentum and the position of the reconstructed topo-cluster, the expected energy deposited within the calorimeter is computed. Often a particle, particularly at high-momentum, will deposit energy within multiple topo-clusters, and therefore the probability that the first track–topo-cluster system fully contains the particle's deposited energy is computed, and if below a threshold, more topo-clusters are added to this track–topo-cluster system.

With the set of topo-clusters associated with the track selected, a cell-by-cell subtraction is performed to iteratively remove energy deposits within each topo-cluster consistent with the track momentum to avoid double counting. Any remaining cluster energy is retained as a neutral-particle contribution. Outside the ID coverage ($|\eta| < 2.5$), only calorimetric information is used. The inputs to the jet reconstruction are the remaining positive-energy topo-clusters above the noise threshold after track energy subtraction, and the selected tracks from the primary vertex. They are fed into the anti-k_t jet clustering algorithm to define PFlow jets — denoted `Antikt4EMPFlow` for anti-k_t jets of radius 0.4 clustered with particle flow objects calibrated at the electromagnetic scale.

4.3.5 Jet Calibration

The goal of jet reconstruction is to closely approximate the kinematics of the quark or gluon that produced it. There are several effects, some of which have already been discussed, that cause the reconstructed energy and direction to be distorted. Jet calibration in ATLAS first corrects the energy scale to the expectation within simulation, known as the jet energy scale (JES) calibration. Second, the difference between the reconstructed jet energy resolution as predicted in MC and that measured in data is corrected for, a correction known as the jet energy resolution (JER) calibration.

The calibration follows a staged procedure shown in Figure 4.5. First, assuming pile-up is uniform in $\eta - \phi$, an average pile-up p_{T} subtraction is performed: $p'_{\mathrm{T}} = p_{\mathrm{T}} - \rho A$, where ρ is the estimated energy density, and A is the jet area. After this initial subtraction, any residual pile-up dependencies are corrected using a linear

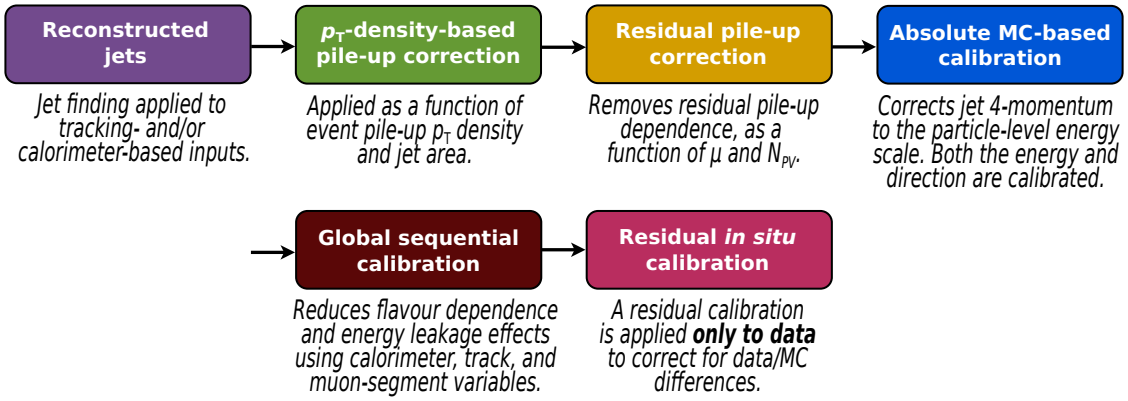


Figure 4.5: An overview of the jet energy scale (JES) calibration within ATLAS. Figure taken from Ref [131].

correction depending on the number of reconstructed primary vertices and the mean number of pile-up interactions, μ. Jets are then corrected on an $(|\eta|, E_{\text{true}})$ plane, at each bin calculating $R = \frac{E_{\text{jet}}^{\text{reco}}}{E_{\text{jet}}^{\text{truth}}}$ using dijet MC, with the inverse applied to reconstructed jets to correct back to truth momenta. A global sequential calibration step seeks to calibrate the jet response as a function of kinematics sensitive to the origin of the jet (i.e. quark or gluon initiated) using a sequential calibration of charged particle fraction, fraction of energy deposited within the first tile calorimeter layer, the depth of shower in the ECAL, number of tracks, the momentum-weighted width of the jet using only tracks and the number of muon tracks associated to the jet. Finally, an in-situ JES calibration corrects residual differences between the jet response as simulated in MC and observed in data. To perform this, processes containing a well-calibrated object (such as $Z \to \mu\mu$) can be used to probe the expected behaviour of the jet. Firstly, the response as a function of η can be probed using the transverse balance of di-jet events, with the energy response of a forward jet to that of the central well-reconstructed jet. Secondly, Z+jet and γ+jet events allow a clean $Z \to \mu\mu$ or photon system to be balanced with the jets, ideal for lower-p_{T} jets. Finally, to calibrate jets at high momentum, multi-jet events are used where a single high-p_{T} jet is balanced with a system of a few well-calibrated softer jets.

The jet energy resolution (JER) quantifies the precision with which the ATLAS detector reconstructs jet momentum [131]. Accurate determination of the resolution

is crucial for all areas of the ATLAS physics programme, but relevantly for the ATLAS analysis presented in Section 7, the JER is the limiting factor determining the missing transverse momentum resolution.

JER has the same functional form as the calorimeter energy resolution, incorporating noise, stochastic and constant terms. In ATLAS, the JER is calibrated using in-situ momentum-balance techniques in dijet events. A smearing factor is then applied to MC simulations to match the resolution in data; if in a certain p_{T} window the MC already has a larger resolution, a systematic uncertainty is applied to account for the difference.

4.3.6 Pile-up Mitigation

Suppressing jets that do not originate from the primary vertex is a crucial objective in ATLAS analyses. There are two forms: jets from a single pile-up interaction, QCD pile-up, and the reconstruction of hadronic jets from random hits and deposits made by multiple particles within the detector, known as *stochastic* pile-up. Two quantities are defined to extract the fraction of momenta contained within a jet from tracks originating from the primary vertex. First, a fully track-based quantity, `corrJVF`, is the fraction of track p_{T} originating from the hard-scatter vertex relative to the total — large for genuine hard-scatter jets and small for pile-up. Then, the ratio of the scalar sum of track transverse momentum associated with the hard-scatter vertex to that of fully calibrated PFlow jet p_{T}, $R_{p_{\mathrm{T}}}$, is defined. A heatmap of these two observables is defined, with each bin containing the fraction of hard-scatter jets relative to the total in the event using MC simulation. A probability is then derived using a k-nearest neighbours-based algorithm on a 2D corrJVF$-R_{p_{\mathrm{T}}}$ plane to form the final jet vertex discriminant score, $\texttt{JVT}(\texttt{corrJVF}, R_{p_{\mathrm{T}}})$ [132], and is applied to jets below 60 GeV in the dark Higgs analysis.

4.3.7 b-jets

Both analyses presented in this thesis involve four jets initiated by b-quarks, known as b-jets, and must be separated from processes involving light-quark (u, d, s) jets.

Light-quark and gluon-induced jet production is overwhelmingly dominant at the LHC, as can be seen in Figure 2.1, and the separation between these processes and those of interest to analyses (e.g. $H \to b\bar{b}$) starts with flavour-tagging. Should the initiating quark of the jet, the *flavour*, be identified (or *tagged*), events not containing b-jets can be rejected. Identifying flavour, *flavour tagging*, is therefore of great importance to separate any hadronic process at the LHC into respective final states.

Light-flavour hadrons predominantly decay through the strong interaction whenever kinematically allowed. Ground-state hadrons for which no such strong decay exists instead decay electromagnetically or weakly, leading to much longer lifetimes. With the strong interaction preserving flavour, the lightest b-hadrons can only decay weakly and, since $m_t \gg m_b$, only via CKM-suppressed transitions (e.g. $b \to cW^-$, $b \to uW^-$). The combination of these two factors leaves the b-hadron with a relatively long[7] lifetime of $\mathcal{O}(10^{-12})$s. Accounting for the Lorentz boost, b-hadrons produced at 50 GeV therefore travel around $\langle L \rangle = \beta\gamma c\tau \sim 3$ mm. This displacement is substantially larger than the primary-vertex position resolutions, which reach better than $\sim 20, \mu$m in the transverse plane and $\sim 30, \mu$m along the beam axis [127]. Therefore, the longitudinal and transverse impact parameters, and their corresponding significances, are often highly discriminating high-level features in this classification task. Moreover, the weak decays of b-hadrons most often proceed via a charm quark ($|V_{cb}|^2 \gg |V_{ub}|^2$), and the tertiary decay of the charm hadron can provide useful discriminating information. Shower-shape features and charged-track multiplicity also provide valuable information for distinguishing between categories. Modern flavour tagging at ATLAS and CMS utilises machine learning to perform this task. LHCb take a different approach, whereby the highly efficient particle identification enables the full decay chains to be reconstructed into specific hadron species — rather than a jet-wide classification task. A detailed account of flavour tagging at ATLAS is given later in Section 6, when describing the author's contribution to the algorithm's calibration at high transverse momentum.

[7] Here, a long lifetime is relative to hadrons that decay strongly or electromagnetically. The much larger mass and greater number of accessible decay channels of a b-hadron give it a substantially larger decay width, and hence a much shorter lifetime, than a muon for example.

4.3.8 Muons

Muons are identified by matching MS tracks to ID tracks and performing a combined χ^2-based track fit. The muons must also pass a momentum-charge compatibility measure between the inner detector and muon spectrometers, and the `medium` working point as defined in Ref [133] is used in the dark Higgs analysis presented later.

4.3.9 Missing Transverse Momentum

The longitudinal momentum fraction of the probed partons within each colliding proton varies, leaving a momentum imbalance along the $\hat{z}$-direction. At such high energies, the small transverse imbalance can be considered negligible. The reconstructed imbalance of momentum perpendicular to the beam following a collision is therefore sensitive to the production of particles that do not appreciably interact with the detector material, neutrinos in the SM or a hypothetical dark matter candidate, and is known as the *missing transverse momentum* ($p_{\mathrm{T}}^{\mathrm{miss}}$).

Calculating this quantity requires a precise understanding of every detector subsystem and the subsequent object reconstruction. First, the event is categorised into two components. Reconstructed and calibrated particles (electrons, photons, τ-leptons, muons and jets) form the *hard*-component. Charged tracks associated to the hard scatter vertex, and not to any other object, form the *soft*-component[8]. The missing transverse momentum for a given event is then defined as,

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \underbrace{\underbrace{-\sum_{e}\vec{p}_{\mathrm{T}}^{\,e}}_{\text{electrons}} \underbrace{-\sum_{\gamma}\vec{p}_{\mathrm{T}}^{\,\gamma}}_{\text{photons}} \underbrace{-\sum_{\tau_{\mathrm{had}}}\vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{had}}}}_{\text{hadronic taus}} \underbrace{-\sum_{\mu}\vec{p}_{\mathrm{T}}^{\,\mu}}_{\text{muons}} \underbrace{-\sum_{\text{jets}}\vec{p}_{\mathrm{T}}^{\,\text{jet}}}_{\text{jets}}}_{\text{hard term}} - \vec{p}_{\mathrm{T}}^{\,\text{soft}} \tag{4.9}$$

There is a possibility that objects entering this calculation are reconstructed from common calorimeter deposits, and care is taken to avoid double counting contributions to $p_{\mathrm{T}}^{\mathrm{miss}}$. Objects therefore enter the calculation sequentially, within an ambiguity-resolving algorithm, starting with electrons, photons, τ_{had} and finally jets [134]. The order is generally set starting with the most unambiguously defined

[8] Alternative definitions that include calorimeter soft-terms have been used in the past but do not scale well with pile-up.

objects first; for example, electrons form similar ECAL deposits as photons but also possess an associated track in the ID and are thus more cleanly defined. If a photon or τ_{had}-candidate shares calorimeter deposits with a higher-priority object, they are not considered in the calculation. Muons deposit a negligible amount of energy within the calorimeters and thus are not considered for ambiguity resolution. However, if a jet is matched to a muon and is consistent with the muon track momenta, it is excluded from the $p_{\text{T}}^{\text{miss}}$ computation. Tracks that are not matched to physics objects in the event and are associated to the primary vertex are then used to construct the $p_{\text{T}}^{\text{miss}}$ soft-term. Various working points are used for $p_{\text{T}}^{\text{miss}}$-reconstruction, where the jet requirements are made tighter from the `loose` to `tenacious` working points.

Another useful quantity to define is the $p_{\text{T}}^{\text{miss}}$-resolution ($\sigma(p_{\text{T}}^{\text{miss}})$). Per-object p_{T}-resolutions are taken from dedicated object-specific calibrations and are then projected onto $p_{\text{T}}^{\text{miss}}$ parallel and perpendicular directions. Then, the significance of the observed $p_{\text{T}}^{\text{miss}}$, $S_{\text{obj}}(p_{\text{T}}^{\text{miss}})$, can be defined using a likelihood ratio between the null hypothesis of zero true $p_{\text{T}}^{\text{miss}}$ sources and the alternative of genuine-$p_{\text{T}}^{\text{miss}}$ [135].

Uncertainties in the reconstruction of $p_{\text{T}}^{\text{miss}}$ arise from both the hard and soft-terms. For the hard-term, uncertainties from the dedicated object calibrations are propagated through to the final estimation of $p_{\text{T}}^{\text{miss}}$ at the analysis level. Three dedicated uncertainties are derived for the soft term in the basis parallel and perpendicular to $p_{\text{T}}^{\text{hard}}$. Here, the scale of the parallel component is calibrated[9], together with the parallel and perpendicular width.

4.3.10 ATLAS Trigger

ATLAS cannot read out every proton-proton bunch-crossing delivered to the experiment. A decision is therefore made in real-time whether an event is considered sufficiently interesting for physics analysis. With the overwhelming majority of scatterings observed at the LHC mediated by the strong interaction, for which

[9]There is no direction preferred with respect to the hard-component, meaning that positive and negative contributions cancel out. The mean of the perpendicular projection is therefore zero ($\langle p_{\text{T}}^{\text{soft}} \perp p_{\text{T}}^{\text{hard}} \rangle = 0$) and does not require a calibration.

forward low-p_{T} jets are dominant, reconstructing and identifying potentially interesting signatures that differ from this is the goal of the ATLAS *trigger* system. The trigger system [136, 137] uses numerous detector signatures to make this decision, and only $\mathcal{O}(10^{-5})$ events are retained relative to the total. In Run 2, the ATLAS trigger was a two-level system, comprised of a hardware-based Level-1 (L1) trigger and a software-based high-level trigger (HLT). The general principle is that the 40 MHz bunch-crossing rate is reduced sequentially by each. L1 reduces the rate to 100 kHz, and finally the events are read out by the HLT at a rate of 1 kHz.

One such interesting object is the $p_{\mathrm{T}}^{\mathrm{miss}}$ of an event, and the $p_{\mathrm{T}}^{\mathrm{miss}}$-trigger attempts to reconstruct this to accept events with a large amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ [138]. $p_{\mathrm{T}}^{\mathrm{miss}}$-triggers are used in the analysis presented in Section 7. Relevant for the $p_{\mathrm{T}}^{\mathrm{miss}}$-trigger is the L1 calorimeter (L1Calo) trigger system, here signals left in the calorimeter cells are summed to form trigger towers. If such a tower does not exceed a pre-defined noise threshold, it is set to zero. The remaining towers are summed to form an L1 $p_{\mathrm{T}}^{\mathrm{miss}}$-estimate. Since this estimate does not use calibrated objects, the threshold is deliberately low at 50 GeV and remained stable throughout Run 2 operations. Events passing the L1 $p_{\mathrm{T}}^{\mathrm{miss}}$-trigger then proceed onto the HLT trigger algorithms, where the entire set of calorimeter information is used to perform more sophisticated reconstruction to reduce the trigger rate further. There are three different algorithms deployed during Run 2 of the ATLAS experiment. `cell` is similar to the L1 estimate and is the vector sum of all calorimeter deposits, now using the full calorimeter readout. `mht` exploits the observation that a large portion of momenta within events are contained within hadronic jets and approximates $p_{\mathrm{T}}^{\mathrm{miss}}$ by a vector sum of HLT-jet p_{T}. The `HLT` jets follow a similar calibration procedure as discussed above for offline jet reconstruction. Finally, `pufit` arranges LCW topological clusters on an $\eta - \phi$ map. Areas with large energy deposits are assumed to originate from the hard-scatter, and the remaining low-energy regions are used for a pile-up adjusting fit, where the fit is constrained assuming that summed x and summed y components individually cancel. Using this fit, a directionally aware area-subtraction scheme is applied to the high-E_{T} regions, which are then used to form the `pufit` $p_{\mathrm{T}}^{\mathrm{miss}}$-estimate.

Year	Trigger name	HLT algorithm	L1 [GeV]	HLT [GeV]
2015	`HLT_xe70_mht_L1XE50`	`mht`	50	70
2016	`HLT_xe90_mht_L1XE50`	`mht`	50	90
2016	`HLT_xe110_mht_L1XE50`	`mht`	50	110
2017	`HLT_xe90_pufit_L1XE50`	`pufit, cell`	50	90, 50
2017	`HLT_xe100_pufit_L1XE50`	`pufit, cell`	50	100, 50
2017	`HLT_xe110_pufit_L1XE50(55)`	`pufit, cell`	50 (55)	110, 50
2018	`HLT_xe110_pufit_xe65_L1XE50`	`pufit, cell`	50	110, 65
2018	`HLT_xe110_pufit_xe70_L1XE50`	`pufit, cell`	50	110, 70

Table 4.3: The evolution of the primary $p_{\mathrm{T}}^{\mathrm{miss}}$ physics triggers through the years of LHC Run 2 from 2015 to 2018. For each year, the table shows the algorithms used and the L1 and HLT thresholds applied. Where two HLT thresholds are given, the first corresponds to the `pufit` algorithm and the second to the `cell` algorithm. Table from Ref [138]. Taking one name to demonstrate the entire set, the early 2018 data-taking trigger `HLT_xe110_pufit_xe70_L1XE50` requires $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV (where XE refers to $p_{\mathrm{T}}^{\mathrm{miss}}$) to pass L1, and the HLT requires $p_{\mathrm{T}}^{\mathrm{miss}} > 110$ GeV for `pufit` and $p_{\mathrm{T}}^{\mathrm{miss}} > 70$ GeV for `cell`.

A summary of all algorithms used by ATLAS during Run 2 is given in Table 4.3, together with the respective $p_{\mathrm{T}}^{\mathrm{miss}}$-thresholds in L1 and HLT.

Only calorimetric information[10] is used in the determination of trigger-based $p_{\mathrm{T}}^{\mathrm{miss}}$, and muons are therefore considered invisible[11]. Taking the muons to be invisible enables the trigger efficiencies to be accurately defined in a well-modelled process, and are shown as a function of the offline reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ for the `cell`, `pufit` and the combination of `cell + pufit` algorithms in Figure 4.6 — where the trigger turn-on region is clear and is approximately fully efficient above 200 GeV. The performance, quantified via the efficiency of an event passing the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger, differs in simulation and data close to threshold. Therefore, analyses with large $p_{\mathrm{T}}^{\mathrm{miss}}$ present in the final state typically apply tighter cuts around 200 GeV — well above this threshold region such that the efficiency in data and MC is 100% and no dedicated corrections are required between simulation and data.

[10] Inner detector tracking information is too computationally expensive to be used within 100 ms.

[11] This enables analyses to define dedicated muon control regions to estimate backgrounds, a more detailed discussion of which follows later in Section 7.7.1.

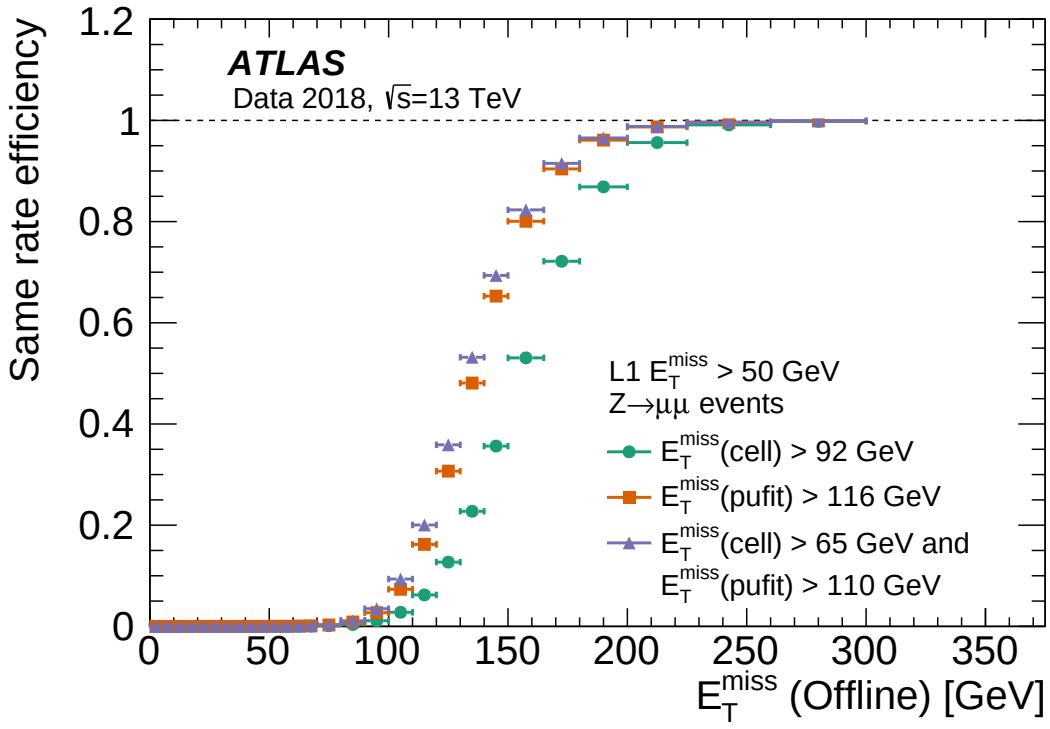


Figure 4.6: Same rate efficiency curves for the `HLT cell`, `pufit` and `cell + pufit` algorithms as a function of the offline reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ in $Z \to \mu\mu$ events that pass the first L1 trigger threshold of 50 GeV, where the muons are considered invisible in $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. Figure from Ref [139].

5

Machine Learning and Statistical Inference

Contents

5.1 Classification 69
5.1.1 Building the Model 70
5.2 Machine Learning at Colliders 70
5.2.1 The Perceptron 71
5.2.2 The Multi-Layer Perceptron 72
5.2.3 Graph Neural Networks and Transformers 72
5.2.4 Model Outputs and Discriminants 73
5.2.5 Training and Validation 74
5.2.6 Network Interpretation 75
5.3 Statistical Inference 76
5.3.1 Building the Likelihood 76
5.3.2 Profile-Likelihood Fit 78
5.3.3 Quantifying Uncertainty 79
5.3.4 Hypothesis Testing 80

Searches and measurements at the LHC seek to measure particular parameters of interest. For a search, this could be quantifying the presence, or lack thereof, of a BSM process, and for a measurement this could be, for example, determining the strength of the Higgs trilinear self-coupling. Achieving this often benefits from the use of machine learning techniques to provide enhanced signal–background separation, together with sophisticated statistical fit procedures to appropriately extract the parameter of interest. This chapter will introduce machine learning and statistical inference, both of which are used extensively in the remaining chapters of this thesis.

5.1 Classification

At the dataset level, the goal of many collider analyses is to separate two hypotheses, typically a null hypothesis H_0 (background), and an alternative hypothesis H_1 (signal + background). The information available to perform this task are the signals stored by the detector, recorded as a high-dimensional pattern of energy deposits, hits and timing information. As discussed in Section 4.3, these signals are compressed to form common objects, such as jets, leptons, and event-wide objects such as $p_{\mathrm{T}}^{\mathrm{miss}}$. To aid this separation task, analyses form higher-level kinematic observables, herein denoted $\mathbf{x}$, such as invariant masses, angular separations, or event-shape features. Using simulation, density functions for such observables can be defined for the signal ($p(\mathbf{x} \mid s)$) and the background ($p(\mathbf{x} \mid b)$), which in an analysis setting may be approximated with binned histograms. Following from Bayes' theorem, the probability that an event is signal given an input feature set is,

$$p(s \mid \mathbf{x}) = \frac{\pi_s p(\mathbf{x} \mid s)}{\pi_s p(\mathbf{x} \mid s) + \pi_b p(\mathbf{x} \mid b)} \tag{5.1}$$

where π_s and π_b are the fraction of signal and background events in the training set. These fractions need not match the genuine composition of the analysis, but often are chosen such that the training set is comprised equally of signal and background to aid classification performance. A simple redefinition finds, with $\pi_s = \pi_b$,

$$\frac{p(s \mid \mathbf{x})}{1 - p(s \mid \mathbf{x})} = \frac{p(\mathbf{x} \mid s)}{p(\mathbf{x} \mid b)} \tag{5.2}$$

where the right-hand side can be identified as the *likelihood ratio*. Statistically, the likelihood ratio is the most optimal discriminant between two hypotheses, known as the Neyman-Pearson lemma [140]. A classifier trained to learn the posterior signal probability, $p(s \mid \mathbf{x})$, can therefore provide a low-dimensional summary statistic which approximates the likelihood ratio.

5.1.1 Building the Model

Developing such a model to map between a high-dimensional input feature set and the underlying probability distribution relies on MC simulation, containing truth labels for signal ($y = 1$) and background ($y = 0$). Here, the Bernoulli distribution captures the probability per event and defines the likelihood function,

$$L(\theta \mid \{\mathbf{x}_i, y_i\}_{i=1}^{N}) = \prod_{i}^{N} \hat{p}_\theta(\mathbf{x}_i)^{y_i} \left(1 - \hat{p}_\theta(\mathbf{x}_i)\right)^{1-y_i} \tag{5.3}$$

where θ is the set of model parameters considered, N is the number of events, and $\hat{p}_\theta(\mathbf{x}_i)$ is the probability that the event is signal given an input feature vector $\mathbf{x}_i$.

Taking the average negative log of the Bernoulli distribution over the set of events gives the binary cross-entropy [141],

$$\mathcal{L}_{\text{BCE}}(\theta \mid \{\mathbf{x}_i, y_i\}_{i=1}^{N}) = -\frac{1}{N} \sum_{i}^{N} \left[y_i \ln \hat{p}_\theta\left(\mathbf{x}_i\right) + (1 - y_i) \ln\left(1 - \hat{p}_\theta\left(\mathbf{x}_i\right)\right) \right]. \tag{5.4}$$

Minimising the binary cross-entropy over the model parameters θ can provide an optimal discriminant for a signal–background separation problem given an input feature set, $\mathbf{x}$. In practice, the performance of the classifier depends on the model architecture, training procedure, MC modelling, event weighting and finite sample size. The final analysis sensitivity also depends on the statistical model, the control of systematic uncertainties and the treatment of the classifier output in the final fit, all of which will be discussed later.

Simplistic approaches involve defining a set of cuts on input features which form rigid boundaries. This often discards subtler correlations between input features that form non-trivial boundaries. A model that is sufficiently flexible to capture both linear and non-linear correlations will be superior; machine learning has proved exceptionally powerful for finding this optimal function.

5.2 Machine Learning at Colliders

Machine learning has been relatively commonplace in particle physics for the last 20 years, starting with jet classification at the Large Electron Positron (LEP)

experiments [142], and moving to its almost universal use today — from experimental operation to future experiment development, object identification, and event-wide classification. Beyond the inevitable gains that more collected data bring, machine learning has been instrumental in increasing measurement sensitivity over the past few decades.

5.2.1 The Perceptron

A particular class of machine learning algorithm is the neural network, inspired by the biological model of neurons in the brain [143]. The perceptron, introduced by Rosenblatt in 1958 [144], feeds an input feature vector $\mathbf{x}$ into neurons, each with a set of tunable weights ($\mathbf{W}$) of the same length as the input vector together with a bias (b). Each neuron then passes the sum $z = \sum_i w_i x_i + b$ into an activation function, like a sigmoid,

$$y(z) = \frac{1}{1 + e^{-z}} \tag{5.5}$$

which maps z to an output on [0,1].

Although y is non-linear in z, the boundary obtained by making a cut on y remains linear. For example, the requirement $y > 0.5$ is equivalent to $z > 0$, corresponding to

$$\sum_i w_i x_i + b > 0. \tag{5.6}$$

The boundary itself is therefore defined by,

$$\sum_i w_i x_i + b = 0, \tag{5.7}$$

and does not capture the aforementioned non-trivial correlations between input features.

5.2.2 The Multi-Layer Perceptron

The multi-layer perceptron (MLP) seeks to remedy this linearity, and, indicative of its name, combines multiple perceptron-like layers. To illustrate the idea, consider a simple MLP consisting of two hidden layers with 128 and 64 neurons respectively, with an input feature vector of length N. The raw inputs are fed into the first layer with 128 neurons, each with N trainable weights. Now, the mth neuron forms $z_m = \sum_{p=1}^{N} W_{mp} x_p + b_m$, and after activation ($y_m = \phi(z_m)$) is fed into each of the following hidden layer's 64 neurons, each consisting of 128 trainable weights. The activations from the 64 neurons are then passed to an output layer, which combines them to form the final discriminant,

$$y = \phi\left[\sum_{k=1}^{64} W_k^{(3)} \phi\left(\sum_{m=1}^{128} W_{km}^{(2)} \phi\left(\sum_{p=1}^{N} W_{mp}^{(1)} x_p + b_m^{(1)}\right) + b_k^{(2)}\right) + b^{(3)}\right] \tag{5.8}$$

where $W^i(b^i)$ are the set of weights (biases) for the ith layer. If we apply the example shown earlier for the perceptron, the decision boundary is now non-linear in the input feature space and capable of forming non-trivial boundaries.

5.2.3 Graph Neural Networks and Transformers

The fixed-neuron structure of an MLP forbids variable-length input feature vectors and forces a consistent ordering. Thus, it relies on sufficiently informative high-level features to discriminate between signal and background. However, the construction of objects and high-level features naturally compresses information, and potentially discards informative correlations among lower-level detector signals. A graph neural network (GNN) instead represents an event by a graph structure, where nodes in the graph (a track, for example) contain a feature vector of properties (p_{T}, η, track-fit score, charge, etc). Connections between nodes then encode inter-nodal information, such as distance, invariant mass or angular separation. This construction thus naturally forms a number of the composite features that an MLP would exploit, whilst preserving the low-level feature correlations that would otherwise be discarded. First, the input nodes enter a trainable embedding stage —

mapping the raw input to a higher-dimensional space that better represents the physical process the network seeks to learn.

The Transformer

The goal of a transformer block is to take the embedded objects, assign inter-nodal connection weights and finally use a feed-forward MLP to exploit the richer representation through highly non-linear, optimisable hidden layers. A dynamic trainable method of assigning weight to relevant inter-nodal connections is the *attention*-mechanism [145], and the transformer in Section 6.1.2 uses self-attention (where all objects are of the same type, tracks in this instance).

A particularly informative example is in heavy-flavour tagging, where a long-lived heavy-flavour hadron will form a displaced vertex with multiple tracks. For a light-flavour jet, displaced vertices can arise from interactions with the detector material, but often are produced with low multiplicity and kinematically differ from that of a heavy hadron decay vertex, whose secondary vertex mass will be large. The attention mechanism helps distinguish the two scenarios: for the heavy-flavour jet the displaced tracks will have high cross-compatibility scores and be upweighted. In contrast, the isolated track in a light-flavour jet will not have a similarly informative track in its vicinity or, if it does, the node-node connection will not form a secondary vertex with large mass.

After the attention mechanism, the embedded objects are passed into an MLP, which non-linearly transforms the object representation, and the output serves as the updated embedding. Global representations of the high-level object under consideration are then built from the embedded objects, typically with a similar attention mechanism (e.g. attention pooling).

5.2.4 Model Outputs and Discriminants

The architectures introduced above map a high-dimensional input feature space to a learned representation, which is passed to an output layer for classification. For binary classification, a single output, known as a *logit*, is mapped to a probability

using a sigmoid function. For classification into multiple categories, the output layer provides one logit per class, such as flavours of jet with z_b, z_c, z_l or z_τ and a softmax function maps the logits to estimate class probabilities [141],

$$p_i = \frac{\exp z_i}{\sum_j \exp z_j} \tag{5.9}$$

which can be used in a likelihood-ratio discriminant.

5.2.5 Training and Validation

The MLP and transformer-based GNN architectures are clearly highly adaptable networks, consisting of an enormously large space of unspecified tunable parameters[1]. Optimising the network involves selecting the values of each parameter. First, a full run through of the network is performed with a random initialisation of parameter values, known as the *forward pass*. The performance of the model is computed using a loss function. Taking the MLP as an example, the network should attempt to minimise the loss function,

$$\hat{\mathbf{W}}, \hat{\mathbf{b}} = \arg\min_{\mathbf{W},\mathbf{b}} \mathcal{L}_{\text{BCE}} \tag{5.10}$$

where the BCE loss has been taken since it is used throughout the remaining machine learning architectures of the thesis.

Since the network output is a nested set of functions, the gradients of the loss with respect to the model parameters can be computed using the chain rule. For an MLP with L layers, to compute the gradient for the weights of layer i (W^i), this may be written schematically as,

$$\frac{\partial \mathcal{L}_{\text{BCE}}}{\partial W^{(i)}} = \frac{\partial \mathcal{L}_{\text{BCE}}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z^{(L)}} \cdots \frac{\partial h^{(i)}}{\partial z^{(i)}} \frac{\partial z^{(i)}}{\partial W^{(i)}}, \tag{5.11}$$

which are efficiently computed using *backpropagation* [141] and are used by optimisers to update the weights,

$$W \to W' = W - \alpha \frac{\partial \mathcal{L}_{\text{BCE}}}{\partial W}, \tag{5.12}$$

[1]For illustration, the simplest MLP we introduced, consisting of only two layers (assuming a reasonable set of 10 features), contains close to 10^4 parameters.

where α is known as the *learning rate* of the network, with larger values enabling more responsive network training, but too large a value can lead to training instabilities. In practice, more sophisticated optimisers such as `Adam` [146] are often used, which aid training stabilisation by computing moving averages of gradients from previous timesteps. In mini-batch training, each update is computed using a subset of the training sample, and one *epoch* corresponds to a full pass over the training dataset.

To avoid the network over-optimising to statistical fluctuations in the training sample, a behaviour known as *overtraining*, a subset of the dataset is withheld from training and used as a *validation* set. Ideally, the network should learn physical features that distinguish signal from background, rather than features specific to the finite training sample, and should therefore perform similarly on both the training and validation datasets. During training, the loss is minimised using only the training sample, while the validation loss is evaluated separately to monitor the network's generalisability. Training is typically halted once the validation loss stops improving, or begins to degrade while the training loss continues to decrease, indicating the onset of overtraining.

5.2.6 Network Interpretation

Modern networks in particle physics are highly non-linear, involving multiple high-dimensional transformations which distort the physically interpretable input feature space. Decoding precisely what the network has learnt can be challenging, but a number of techniques exist to begin understanding the most relevant features and feature correlations. One such method is that of Shapley values, named after game theorist Lloyd Shapley, which treat features as players in a cooperative game [147]. The model response is computed assuming a feature is masked, and for a single feature the Shapley score is,

$$\phi_f = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(M - |S| - 1)!}{M!} \left[f\left(S \cup x_i\right) - f\left(S\right) \right] \tag{5.13}$$

where $|S|$ is the number of features in subset S, M is the total number of features and $f(X)$ is the model output considering X features.

All possible subsets are considered because the contribution of a feature can depend strongly on which other features are already included. For example, a feature may provide little additional information when a highly correlated feature is present, but may be highly informative when that feature is absent. Positive Shapley scores for a given feature value indicate a signal-like characteristic. A common ML summary plot is the Shapley beeswarm, which shows the distribution of Shapley values for each input feature of the model ordered by feature importance. The Shapley approach relies on a fixed input feature set per event, and therefore does not scale well with more dynamic models with variable input lengths, like GNNs. Here, Shapley values can be approximated [148], or methods such as integrated gradients can be used to interpret important features [149].

5.3 Statistical Inference

Once an analysis is built to separate the hypotheses considered, an inference procedure attempts to estimate the value of a given model parameter of interest (POI). There are a number of methods to achieve this, and this chapter will summarise the techniques used in both analyses presented in this thesis: the profile-likelihood fit.

5.3.1 Building the Likelihood

Collider data analyses are highly complex, incorporating perturbative SM predictions, MC modelling, detector simulation and experimental operation. There is a degree of uncertainty attributed to each component which should be accounted for, such that any hint of new physics beyond the SM, or measurement of a SM parameter, can be appropriately considered in the context of these effects. This accounting is achieved by including parameters within the likelihood function that are allowed to vary, subject to constraints, for each source of uncertainty, known as *nuisance* parameters. To quantify the degree of agreement between a set of

observed data, D, and a given hypothesis with parameter μ, together with a set of nuisance parameters $\boldsymbol{\theta}$, the likelihood function is used,

$$L(D \mid \boldsymbol{\theta}, \mu) = \prod_i f(d_i \mid \boldsymbol{\theta}, \mu) \tag{5.14}$$

where $f(d_i \mid \boldsymbol{\theta}, \mu)$ is the probability density function for a given observation d_i. In practice, the analyses considered in this thesis use binned likelihoods, where the observed event counts are compared with the predicted yields across a set of analysis regions and discriminant bins. This likelihood can then be written as a product of Poisson and Gaussian probability density functions [150, 151],

$$L(D \mid \boldsymbol{\theta}, \mu) = \prod_{r \in R} \text{Pois}\left(N_r \mid \nu\left(\boldsymbol{\theta}, \mu\right)\right) \prod_{s \in \boldsymbol{\theta}} \text{Gaus}\left(\theta_s \mid 0, 1\right) \tag{5.15}$$

where R is the set of regions considered by the analysis, $\nu\left(\boldsymbol{\theta}, \mu\right)$ is the predicted yield and N_r is the observed data in that region. The Gaussian constraint term is chosen to encode the natural smearing of the likelihood function in the presence of multiple systematic uncertainties. The increased flexibility introduced by a nuisance parameter is encoded in a response function that describes how the nominally predicted yield varies as the nuisance parameter is adjusted over the range associated with the corresponding uncertainty.

Analysis regions consist of control, validation and signal regions. Control regions are designed to constrain one or more important background processes while remaining kinematically close to the signal region and largely free from signal contamination. They therefore provide data-driven information on the normalisation and, in some cases, the shape of the relevant background predictions in a phase space close to that used for extracting behaviour of the signal.

Validation regions are then used to test the reliability of this background estimate. They are chosen to be kinematically as similar to the signal region as possible, with minimal signal contamination, while remaining orthogonal to both the control and signal regions, and are not used to constrain the fit. Agreement between data and prediction in these regions provides confidence that the fitted

background normalisations and nuisance-parameter constraints extrapolate reliably towards the signal-like phase space.

Finally, the signal region is defined to maximise sensitivity to the signal hypothesis under study. Typically, to avoid biasing the strategy, analyses seeking to measure a parameter of the SM, or search for BSM physics, use only the MC predictions within the signal region to design and optimise the analysis. Data is inspected in the signal region only after understanding the behaviour of the background estimate and fitting behaviour. Then, final hypothesis tests are performed, using the background model constrained by the control regions and validated in the validation regions.

5.3.2 Profile-Likelihood Fit

The global best-fit value for the POI and NPs together is known as the maximum likelihood estimate,

$$\left(\hat{\mu}, \hat{\boldsymbol{\theta}}\right) = \arg\max_{\mu,\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mu \mid D). \tag{5.16}$$

This maximisation is practically implemented using the `MINUIT` [152] package in `ROOT` [153].

In the MLE formalism, the POI and NPs are considered together in a global fit, $\left(\hat{\mu}, \hat{\boldsymbol{\theta}}\right)$. However, the NPs are not the target of a search or measurement, and a statement about a physical parameter μ should not depend arbitrarily on the values of $\hat{\boldsymbol{\theta}}$. Instead, the NPs can be re-fitted to best describe the observed data given a particular choice of μ in a procedure known as *profiling*,

$$\hat{\hat{\boldsymbol{\theta}}}\left(\mu\right) = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mu \mid D) \tag{5.17}$$

where $\hat{\hat{\boldsymbol{\theta}}}$ is the conditional best-fit set of NPs. Profiling therefore constructs a likelihood depending only on μ, while still propagating the impact of systematic uncertainties through the fit.

Taking the ratio between the profiled likelihood and the maximum likelihood estimate is the profile-likelihood ratio (PLR) [154]:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{5.18}$$

where the value of the PLR quantifies how preferred this specific configuration is with respect to the best fit. The curvature of the PLR encodes the sensitivity of the analysis to μ. If $\lambda(\mu)$ varies only slowly, then a wide range of values provide similarly good descriptions of the data, indicating limited sensitivity. Conversely, a sharply peaked profile likelihood indicates that the data strongly constrain the POI.

In a search, the uncertainty on μ determines how strongly the data can exclude or favour the background-only hypothesis. In a measurement, it determines the precision with which the parameter is extracted and whether it is compatible with the SM prediction. Quantifying the statistical and systematic uncertainty on the extracted POI is therefore a vital component of an analysis.

5.3.3 Quantifying Uncertainty

The *score* is a measure of how sensitive the log-likelihood function is to changes in the model parameters, and at the maximum likelihood estimate (MLE), the score is zero. Once such a maximum is found, uncertainties on model parameters are computed using the Hessian, defined as the second derivative of the log-likelihood function:

$$\mathcal{H}_{ij}(\boldsymbol{\alpha}) = \frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\alpha_i} \partial \alpha_j} \tag{5.19}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\theta}, \mu)$.

The Hessian matrix, representing curvature, indicates how rapidly the log-likelihood falls away from this maximum, thus directly relating to the uncertainty in the parameter estimates. Sharper peaks in the likelihood correspond to smaller uncertainties. Moreover, if the Hessian is negative definite everywhere, the log-likelihood function is concave and the derived maximum is global.

The negative expectation value of the Hessian for a given parameter is the Fisher information matrix [155], defined as,

$$\mathcal{I}(\boldsymbol{\alpha}) = -\mathbb{E}\left[\mathcal{H}(\boldsymbol{\alpha})\right]. \tag{5.20}$$

$\mathcal{I}$ is directly proportional to the expected amount of information one obtains about a particular parameter α_i given the observed data and statistical model. Here, utilising the Cramér-Rao bound [156], the variance of the MLE $\hat{\alpha}_i$ is bounded by the inverse of the Fisher information matrix:

$$\mathrm{var}(\hat{\alpha}_i) \geq \left[I^{-1}(\alpha)\right]_{ii}. \tag{5.21}$$

The *constraint* for parameter α_i is then given by:

$$C(\hat{\alpha}_i) = \frac{\sigma_{\text{pre-fit}} - \sigma_{\text{post-fit}}}{\sigma_{\text{pre-fit}}} = 1 - \sigma_{\text{post-fit}} \tag{5.22}$$

where $\sigma_{\text{pre-fit}}$ is taken to be 1 as per convention. Under certain conditions and in the asymptotic limit, the covariance of the MLE approaches the inverse Fisher information. In this limit, the curvature of the likelihood provides an estimate of the achievable parameter uncertainty provided by the Cramér-Rao bound.

5.3.4 Hypothesis Testing

Given the profile likelihood, three related inference questions can be addressed. First, what confidence interval may be set on a given POI estimate? Second, are the data incompatible with the background-only hypothesis, corresponding to evidence for a signal? Third, if no significant excess is observed, which signal hypotheses can be excluded?

To begin, it is useful to define the test statistic,

$$t_\mu = -2\ln\lambda(\mu) \tag{5.23}$$

where smaller values of t_μ indicate increased compatibility between the hypothesised POI μ and the data, with the best-fit $\mu = \hat{\mu}$ at $t_\mu = 0$. In the asymptotic limit, the test statistic approaches a χ^2-distribution and enables t_μ to be converted into p-values.

For SM measurements, the two-sided profile-likelihood test statistic is typically used. The best-fit value lies at the minimum of $t_\mu = -2\ln\lambda(\mu)$, and confidence intervals are obtained by finding the values of μ for which the test statistic increases by a fixed threshold. For a single POI, the one-standard-deviation interval is defined by $\Delta t_\mu = 1$.

However, if we want to check for excesses above the background-only hypothesis, a negative μ loses its meaning and therefore a one-sided test statistic is built,

$$t_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} \geq 0, \\ 0, & \hat{\mu} < 0. \end{cases} \tag{5.24}$$

where $\mu = 0$ corresponds to the background-only hypothesis.

The discovery p-value is thus,

$$p_d = \int_{t_0^{\text{obs}}}^{\infty} f\left(t_0|H_0\right) dt_0 \tag{5.25}$$

where t_μ^{obs} is observed value of the test statistic, which can be converted to a significance,

$$Z = \Phi^{-1}(1 - p_d) \tag{5.26}$$

where Φ^{-1} is the inverse cumulative distribution function for a Gaussian distribution[2] [157, 158].

In the event that no deviation from the SM expectation is observed, a search then sets limits on existing BSM signatures. First, defining the test statistic now as a function of μ,

$$t_\mu = \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > \mu. \end{cases} \tag{5.27}$$

where the test statistic is set to zero for scenarios in which the data prefers a signal strength greater than the nominal signal prediction, avoiding incorrectly excluding excessively signal-like characteristics.

The p-value is defined similarly to before,

$$p_{s+b} = \int_{t_\mu^{\text{obs}}}^{\infty} f\left(t_\mu|H_1\right) dt_\mu \tag{5.28}$$

[2]A 5σ discovery therefore relates to a 1 in 3.5 million chance that the background-hypothesis could generate data at least as extreme as the observed data.

and is the probability of observing data at least as incompatible with the signal hypothesis H_1, and a naive exclusion bound can be set requiring $p_{s+b} < 0.05$. However, this approach can lead to misleading exclusions when a downward fluctuation in the data is observed. Here, the data will appear highly incompatible with the signal model that predicts an excess of events. This is problematic if the analysis also has limited sensitivity to the signal model, with the PLR slowly varying with μ, meaning that the distributions for the test statistic under H_0 and H_1 are similar whilst giving a small value of p_{s+b}, potentially giving a misleading exclusion result.

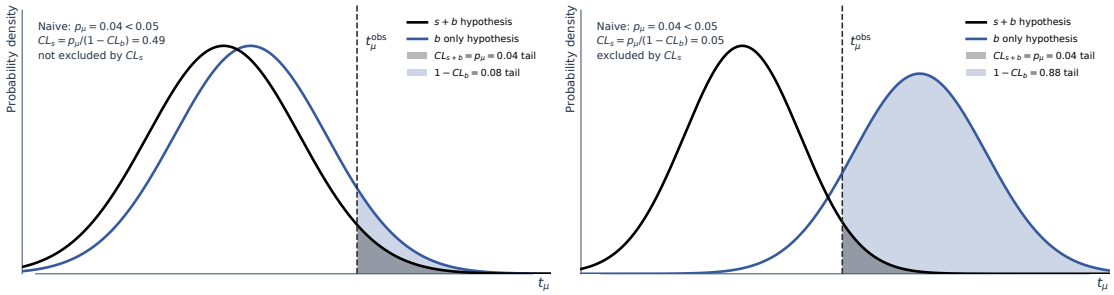
The CL_s method accounts for this with,

$$\mathrm{CL}_s = \frac{p_{s+b}}{1 - p_b} \tag{5.29}$$

where the denominator suppresses exclusions in regions where the observed value of the test statistic is also likely under the background-only hypothesis, as occurs when the analysis has little sensitivity to the tested signal model. A signal hypothesis is conventionally excluded at the 95% confidence level when $\mathrm{CL}_s < 0.05$. Figure 5.1 shows the CL_s method for two illustrative scenarios, on the left for an analysis with poor sensitivity to μ, which spuriously excludes a signal model with a naive CL_{s+b} approach but does not with the CL_s-method. In the high-sensitivity case, shown on the right, the two hypotheses are well separated; data that are incompatible with the signal-plus-background model are compatible with the background-only hypothesis, and both the naive and CL_s criteria exclude the signal model.

As introduced in Section 3.3.3, analyses searching for new physics often choose a benchmark which fixes a set of parameters to pre-defined values and scans over two remaining parameters, like $(m_s, m_{Z'})$ for the dark Higgs model. Exclusion results of this analysis will then be shown on the 2D-plane defined by $(m_s, m_{Z'})$, and a contour drawn through signal models satisfying $\mathrm{CL}_s < 0.05$, with the interior defining the excluded region.

Furthermore, a model-independent upper limit on the visible cross-section, $\sigma_{\mathrm{vis}} = A\varepsilon\sigma_{\mathrm{model}}$, can be defined. Here, σ_{model} is the predicted cross-section of a



(a) **(b)**

Figure 5.1: Two example configurations to illustrate the CL_s method. An example analysis with poor sensitivity to the signal model considered is shown on the left, where an observed deficit pushes p_μ to below 0.05 and excludes the signal. However, the distribution of the test statistic in the b-only scenario closely matches that of the signal, and the CL_s method nullifies the aggressive exclusion by normalising to CL_b. On the right we see an example where the analysis is very sensitive to variations in μ, and both the naive and CL_s approaches correctly exclude the signal model.

hypothetical BSM model, and $A\varepsilon$ is the acceptance and efficiency for which the BSM model passes all selections defining the signal region. Such a limit extends the utility of the analysis beyond the specific benchmark model targeted. Provided that the selections are externally reproducible[3], a new BSM can be excluded, or remain viable, without undertaking a full re-analysis [159, 160].

Deriving this quantity utilises the CL_s prescription described above. To minimise assumptions about a possible signal's kinematic dependence, the signal region is a single bin. The assumed signal yield is increased iteratively until CL_s reaches 0.05, defining the upper limit at the 95% confidence level on the number of signal events S_{obs}^{95}, and converted to the visible cross-section by normalising to the total integrated luminosity used by the analysis, $\sigma_{\mathrm{vis}}^{95} = S_{\mathrm{obs}}^{95}/\mathcal{L}$.

[3]If machine learning is used, this requires the full set of network weights, input definitions and preprocessing procedures to be made publicly available.

"If they can't identify the flavour, they shouldn't be cooking the dish."

— Gordon Ramsay

6 Flavour Tagging

Contents

6.1 Flavour Tagging at ATLAS 84
6.1.1 DL1 Series 85
6.1.2 GN series 87
6.1.3 Algorithm Training 89
6.1.4 Deployment in Analysis 89
6.1.5 Algorithm performance at high-p_{T} 92
6.2 Algorithm Calibration 94
6.2.1 b-tagging Efficiency 94
6.2.2 Charm-mistag Efficiency 95
6.2.3 Light-mistag Efficiency 95
6.2.4 Extrapolation to High-p_{T} 96
6.2.5 Usage in Analysis 105
6.2.6 Conclusions 106

This chapter details two approaches to heavy-flavour jet identification in the ATLAS collaboration, both utilising machine learning: the deep-learning (DL) series of algorithms `DL1r` and `DL1d`, and the latest attention-based graph-neural-network (GNN) tagger `GN2`, together with the subsequent calibration of each — notably the calibration's extension to high jet transverse momentum, which the author performed. The `DL1r` algorithm, briefly introduced here, is used to tag the flavour of jets in the $HH + p_{\mathrm{T}}^{\mathrm{miss}}$ analysis of Section 7.

6.1 Flavour Tagging at ATLAS

The algorithms developed in the ATLAS experiment during Run 2 relied on a two-staged approach. Key features, as described in Section 4.3.7, including the presence of at least one displaced vertex, the large secondary vertex mass and high charged-track multiplicity, are reconstructed first with a set of low-level

algorithms, with the outputs and correlations between these low-level algorithms exploited using high-level taggers. A key progression in Run 3 is the use of more sophisticated GNN-style architectures, starting with a Graph Attention Network `GN1` and moving to a transformer-based architecture `GN2` — which is the most advanced algorithm currently deployed within ATLAS, and boasts significant performance gains relative to its `DL1` predecessors.

6.1.1 `DL1` Series

The deep-learning (DL) series [161] of flavour-tagging algorithms deployed a set of low-level feature taggers to exploit the distinct kinematics of heavy-flavour jets, a secondary vertex tagger `SV1`, a multi-vertex fitter `JetFitter` and, for `DL1r`, a recurrent neural network-based impact parameter reconstruction algorithm `RNNIP`. `DL1d` improved on this architecture first through the inclusion of c-specific feature outputs of `JetFitter`, and also a deep-sets-based impact parameter reconstruction network `DIPS`, replacing `RNNIP`.

SV1

The Secondary-Vertex-Fitter (`SV1`) [162] algorithm attempts to reconstruct a single displaced vertex consistent with a heavy-flavour hadronic decay. For computational efficiency, the algorithm first reconstructs two-track vertices and then attempts to combine spatially close two-track vertices into an n-track vertex. This reduces the computational complexity from $\mathcal{O}(N^n)$ to $\mathcal{O}(N^2)$. Various techniques are employed to remove fake two-track vertices, but it remains very challenging to efficiently reconstruct vertices in the highly collimated jet core. The inevitable reduction in vertex-finding efficiency for high-p_{T} jets leads to a deterioration in flavour-tagging performance, which is discussed in detail later. Four features are then used to construct a likelihood ratio between b and non-b hypotheses: the invariant mass of particles assigned to the reconstructed secondary vertex, the ratio between the sum of track energies assigned to the vertex to the total track

energy within the jet, the number of clean two-track vertices and finally the spatial separation between SV − PV and jet directions.

JetFitter

With the majority of b-hadrons decaying via a charmed hadron, the ability to reconstruct a common flight path between the primary, secondary and tertiary vertices will provide discriminating power. `JetFitter` seeks to achieve this by using a modified Kalman filter, with the constraint that the two displaced vertices lie on a line connected to the hard-scatter vertex, and outputs a number of features to be used as inputs to high-level taggers [163].

Impact-parameter Networks

`DIPS` exploits track-level impact parameter, kinematic and hit information using a permutation-invariant deep sets architecture [164]. Tracks within jets have no inherent ordering, and unlike `RNNIP`, which orders on impact parameter significances, `DIPS` takes as input an unordered set of tracks. Track features include the longitudinal and transverse impact parameter significances, together with hit information in the ID, distance between tracks and the jet axis, and the fraction of jet p_{T} carried by the track. Each track then enters a neural network to form a track embedding, and is pooled to form jet-level probabilities — much the same as the mechanisms used within the transformer block introduced in Section 5.2.3.

The impact-parameter sign is assigned relative to the jet axis, which is used as an approximation to the heavy-hadron flight direction. Tracks are assigned a positive-signed impact parameter when their point of closest approach to the primary vertex lies in the downstream direction defined by the jet.

Therefore, the tracks originating from a b-hadron, displaced along the jet direction, will more often have a positive-signed IP, whereas light-jets from the PV will form a symmetrically signed-IP distribution. `DIPS` and previous IP-based low-level taggers exploit this property to enhance separation. This property is also used to *flip* a flavour-tagging algorithm, where the sign of lifetime-sensitive features is flipped, and the output of the network recomputed; this greatly reduces the

number of heavy-flavour jets whilst maintaining a similar composition of light-jets, useful for calibrations as will be discussed later.

High-level Taggers

For genuine and well-reconstructed b-jets, the outputs of each low-level tagger will be highly correlated. For non-b cases, individual low-level taggers could predict highly b-like features whilst the remaining scores are low. The degree of separability for each tagger varies as a function of jet kinematics, and to fully exploit the set of low-level tagger outputs, an 8-layer MLP is trained to predict probabilities p_b, p_c and p_l for bottom, charm and light jets respectively.

6.1.2 GN series

Unlike DL1d, GN2 [1] is an end-to-end architecture with no low-level feature taggers, taking as input a feature set of track and jet kinematics and outputting per-class probabilities directly.

For each jet, up to 40 associated tracks are used, with zero-padding applied to jets with fewer selected tracks. Track-level inputs, together with the jet p_{T} and η, are listed within Table 6.1. Tracks are fed into an initialiser, consisting of a single hidden layer and an output layer of size 256. Then, the tracks enter a transformer block, stylistically the same as described in Section 5.2.3. The transformer layers allow the representation of each track to be updated with information from all other tracks in the jet through self-attention. This provides a natural mechanism for learning correlations among displaced tracks, and, to aid with the task of classification, two auxiliary tasks are defined to predict the origin of a particular track and to determine which tracks form common vertices. These auxiliary tasks enter the total loss function,

$$\mathcal{L}_{\texttt{GN2}} = \mathcal{L}_{\text{jet}} + \alpha\mathcal{L}_{\text{vertex}} + \beta\mathcal{L}_{\text{track}} \tag{6.1}$$

with $\alpha = 1.5$ and $\beta = 0.5$,

The auxiliary vertexing objective can be viewed as a directly trainable analogue of the information sought by SV1 and JetFitter. Rather than requiring a discrete

Jet Input	**Description**
p_{T}	Jet transverse momentum
η	Jet pseudorapidity
Track Input	**Description**
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of the track, relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Transverse impact parameter with respect to the PV
$z_0 \sin\theta$	Longitudinal impact parameter, projected by $\sin\theta$, to the PV
$\sigma(q/p)$	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
$s(z_0 \sin\theta)$	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

Table 6.1: Jet-level and track-level input variables used by the flavour-tagging algorithm. The inclusion of shared and split-hit information is particularly relevant in dense jet cores, whereby charged particles may produce merged pixel clusters. Table taken from Ref [1].

reconstructed vertex, `GN2` learns a continuous representation of whether tracks are mutually compatible with a common displaced vertex. This is particularly valuable in dense jets, where explicit vertex reconstruction is inefficient and where the low-level tagger optimisation environment differs substantially.

The track-origin auxiliary task seeks to determine the origin of each track, distinguishing prompt tracks, pile-up or fake tracks, tracks from b and c-hadron decays, tracks from the $b \to c$ cascade, tracks from τ-lepton decays, and tracks from other secondary interactions (material interactions in the ID, for example).

Studies in which GN2 was retrained without the vertex-compatibility and track-origin objectives showed reductions of up to 30% in charm- and light-jet rejection. Removing only one of the two auxiliary objectives recovers most of this loss, indicating a high degree of correlation between the tasks [1]. The auxiliary objectives therefore act as a mechanism to guide the latent track representation to encode displaced-track and vertex-like information.

6.1.3 Algorithm Training

Algorithm training requires a truth-labelling prescription that assigns each reconstructed jet to a flavour category in simulation. The work presented uses ΔR-associations, and should a b-hadron above $p_{\mathrm{T}} > 5$ GeV lie within $\Delta R < 0.3$ then the jet is considered a truth b-jet, and if no b-hadron is found then the labelling is hierarchical and jets matched to a c-hadron are labelled as c-jets, followed by hadronically decaying τ-leptons, with the remaining jets labelled as light flavour. Recent advances in IRC-safe jet labelling [165] may improve the purity of these selections by neutralising $g \to b\bar{b}$ splittings that are mistakenly labelled as b or bb-jets.

The training dataset consists of samples containing each object class, including b, c, light-jets and, for GN2, also τ-jets. At low-p_{T} ($20 < p_{\mathrm{T}} < 250$ GeV) $t\bar{t}$ events are used and to populate the high-p_{T} region ($250 < p_{\mathrm{T}} < 6000$ GeV) a Z' sample is used, which preferentially generates events at higher-$\hat{s}$ according to the scaling $w(\hat{s}) = \exp(-9 + 0.002\sqrt{\hat{s}})$. The Z' decay branching fractions are chosen as 30% for b, c and light-quark jets, 5% for τ-leptons and 5% for electrons.

6.1.4 Deployment in Analysis

The output class probabilities for each algorithm are used to form a discriminant,

$$D_b = \ln\left(\frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_l}\right) \tag{6.2}$$

where f_c, f_τ control the relative weighting of each category and are tuned for the desired balance between the rejection of charm and light-flavour jets. No dedicated τ-class existed for DL1d and therefore f_τ is set to 0. The values of $f_{c(\tau)}$

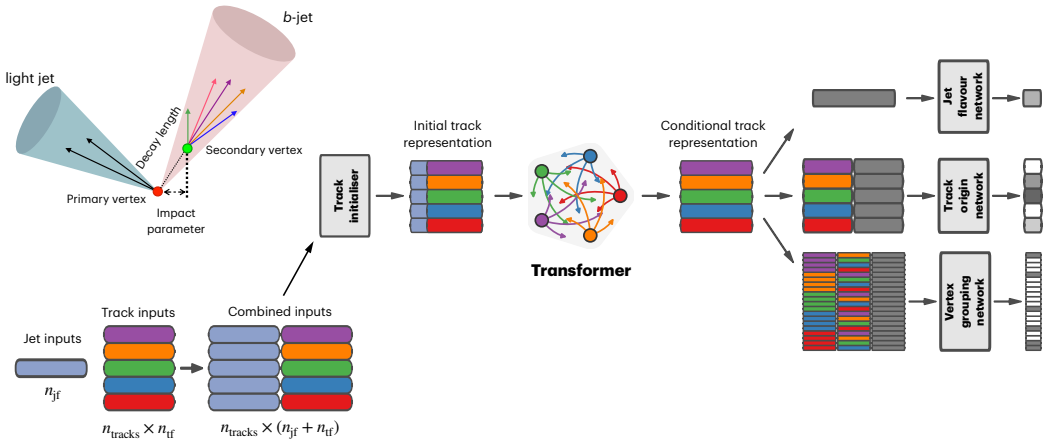


Figure 6.1: Schematic of the GN2 flavour tagging algorithm. Figure taken from Ref [1].

are 0.2 (0.05) for GN2 and 0.018 (0) for DL1d, where GN2 targeted significantly enhanced charm rejection.

For each jet, this forms a discriminant score that may be used to define working points. Working points (WP) are defined in bins of the discriminant according to a target efficiency in $t\bar{t}$ MC, such that, for example, a fixed-cut on D_b corresponds to an average of a 77% efficient selection of b-jets, known as the 77% fixed-cut WP. For a large portion of Run 2, this was the baseline approach, and more recent approaches involve placing two-sided cuts on the discriminant, between 70% and 77% working points, for example, to define a pseudo-continuous b-tagging scheme (PCBT). This approach provides greater flexibility for analyses sensitive to flavour composition to define a number of control regions. Moreover, by segmenting the calibration into more granular bins, systematic uncertainties can be considered locally to a particular PCBT bin, rather than increasingly inclusive for fixed-cut WPs.

The distribution of discriminant scores for the two leading truth-matched b-, c- and light-jets in a Z' sample is shown in Figure 6.2 with vertical lines indicating the fixed-cut WPs. The separation of the flavour templates demonstrates that increasingly tight cuts on D_b preferentially retain true b-jets while suppressing the charm- and light-flavour counterparts.

The most recent update to this scheme for ATLAS has been 2D discriminant cuts, where the classification task for b-tagging can easily be rephrased as a charm-

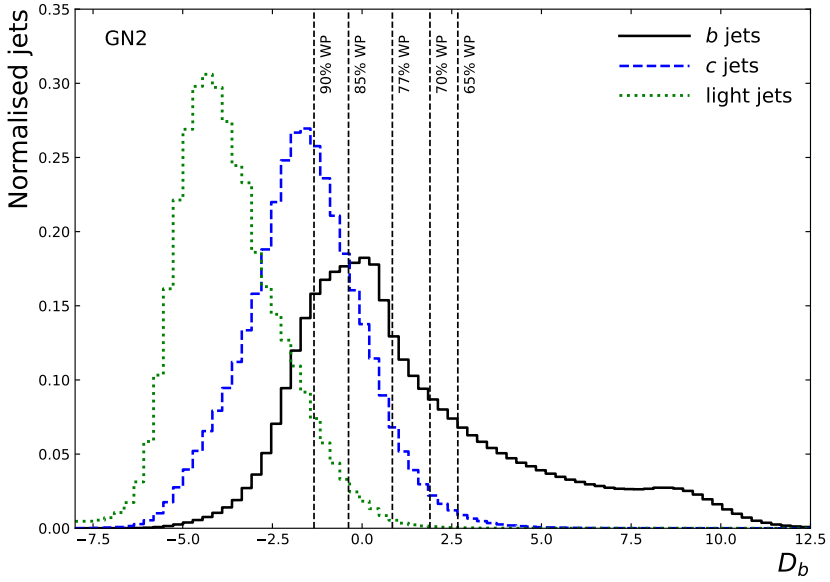


Figure 6.2: The distribution of discriminant scores for truth b (black), c (blue) and light-jets (green) for the leading two jets in a Z' sample shown together with all fixed-cut WP cuts on the discriminant score. This sample includes very high-p_{T} jets, far beyond the kinematic range considered to define the working point. Thus, the labelled thresholds should not be interpreted as corresponding to their nominal average efficiencies in this sample.

tagging objective with the discriminant rewritten as,

$$D_c = \ln\left(\frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_l}\right) \tag{6.3}$$

where $f_{b(\tau)}$ are set to 0.3 (0.01) for `GN2` and 0.1 (0) for `DL1d`.

This enables a 2D classification to be performed, which is highly beneficial for analyses targeting $c\bar{c}$ and $b\bar{b}$ final states, like $VH \to b\bar{b}/c\bar{c}$ to measure the Higgs bottom and charm Yukawa coupling [166]. Here, b and c-enhanced regions can be defined according to Figure 6.3 — known as pseudo-continuous flavour tagging (PCFT). The author migrated the MC-based uncertainty frameworks from fixed-cut to PCBT and, finally, to PCFT, performing the calibration at high transverse momentum for each iteration. Future deployment within the collaboration may utilise the technique of optimal transport to continuously, rather than pseudo-continuously, calibrate the algorithm's discriminant [167, 168] — providing analyses maximal flexibility with respect to their flavour-tagging selections.

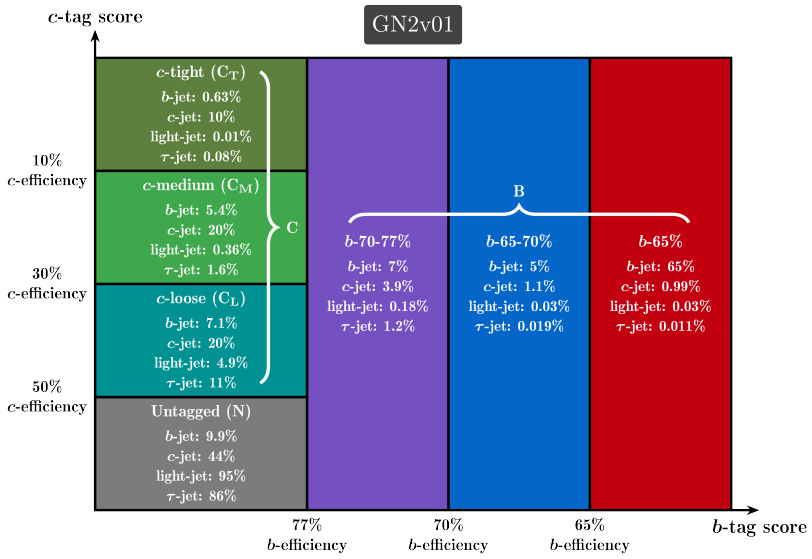


Figure 6.3: Schematic of the 2D pseudo-continuous flavour tagging (PCFT) working points. Here, a combination of cuts on D_c and D_b enables regions of relative charm and bottom-purity simultaneously. Figure taken from Ref [1].

6.1.5 Algorithm performance at high-p_{T}

ATLAS has measured, and searched for, a wide variety of SM and BSM processes, the majority of which show strong agreement with SM predictions. Increasingly, therefore, the collaboration is searching in regions of phase space where the SM is less well constrained. One such region is at high transverse momentum, where a high-mass BSM state could decay into two high-momentum, well-separated jets that benefit from flavour tagging — requiring efficient jet flavour tagging at high-p_{T}.

Flavour tagging at high-p_{T} presents several challenges. Firstly, since the opening angle of the decay products scales with $\theta \propto \frac{1}{\gamma} \sim \frac{m_B}{p_B}$, high-momentum b-hadrons will produce tracks contained within a tight cone. For example, the decay products of a 1 TeV b-hadron will be contained within 0.005 rad, and at the radius of the IBL this corresponds to a transverse scale of $\Delta x \sim r_{\mathrm{IBL}}\theta \sim 150$ μm. Comparing this to the segmentation of the IBL pixel layers of 50×250 μm in the transverse and longitudinal directions, this structure can, in principle, be resolved. Still, often multiple particles will traverse the same IBL pixel — relying on the pixel clustering algorithm to resolve the underlying structure. The track reconstruction challenge is compounded by the tiny crossing angle of each track, making the task of vertex

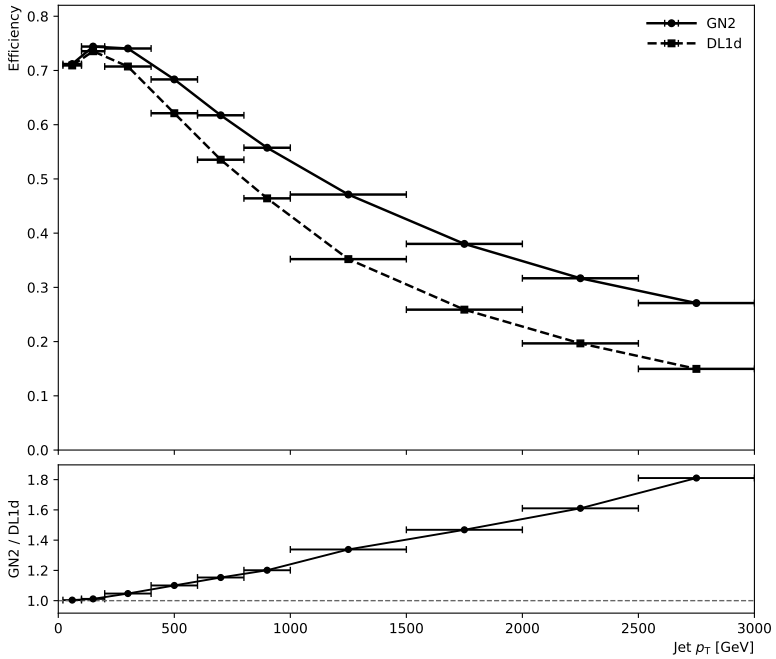


Figure 6.4: Plot of the b-tagging efficiency for the two leading truth b-jets with the 77% fixed-cut WP as a function of the jet p_{T} in Z' events for the `GN2` (solid) and `DL1d` (dashed) algorithms, with the ratio panel showing the increase in b-tagging efficiency for identifying high-momentum b-jets in `GN2` relative to `DL1d`.

finding exceptionally difficult. Finally, the mean path length of a b-hadron at 1 TeV is around 10 cm, meaning that b-hadrons will often decay after the IBL. The inclusion of the IBL was central for improvements in b-tagging between Run 1 and 2. Losing hit information, together with the aforementioned effects, inevitably leads to a degradation in b-tagging efficiency with jet momentum.

Figure 6.4 shows the b-tagging efficiency as a function of the jet p_{T} for the `GN2` (solid) and `DL1d` (dashed) algorithms each at the 77% fixed-cut working point, with the ratio panel showing the increase in b-tagging efficiency with `GN2` for high-momentum b-jets. `GN2` does not rely on lower-level taggers to reconstruct characteristic features directly, but rather relies directly on correlations between tracks, which in a challenging high-p_{T} environment may not reconstruct a clean high-level feature but still be indicative of a heavy hadron decay. This is supported by the improved secondary vertex reconstruction efficiency within GN2 compared with that of `SV1`.

6.2 Algorithm Calibration

MC simulation is used for the development, training and optimisation of DL1d and GN2. However, the observables exploited by flavour-tagging algorithms depend sensitively on the modelling of track reconstruction, impact-parameter resolutions, detector material and heavy-quark fragmentation. The tagging efficiency rates predicted by simulation therefore cannot be assumed to reproduce those observed in collision data. A validation of modern taggers such as GN2 is therefore that their improved performance in MC simulation is reflected in collision data after calibration, rather than arising from simulation-specific artefacts.

Dedicated calibration analyses are used to derive data-to-simulation scale factors for the b-jet tagging efficiency and for the c- and light-flavour mistag efficiencies. These scale factors correct the simulated efficiency to match the data, while their uncertainties are propagated as nuisance parameters in the statistical model of ATLAS physics analyses.

6.2.1 b-tagging Efficiency

To measure the b-tagging efficiency within data, a relatively pure sample of b-jets is found by selecting $t\bar{t}$ events, which, with $|V_{tb}| \sim 1$, decay almost exclusively into $W^+W^-b\bar{b}$. The dileptonic $t\bar{t}$ channel improves the purity of the $b\bar{b}$ selection whilst reducing the otherwise large multi-jet background. However, the presence of backgrounds pollutes the calibration, leaving the data a mixture of different flavour compositions. Therefore, the flavour composition is fit simultaneously with the efficiency in a combined likelihood fit [169]. An example of the extracted data-to-simulation scale factors is shown in Figure 6.5 for the $70 < \varepsilon_b < 77\%$ PCBT tag-weight bin, where the agreement between MC and collision data is strong, indicating that the GN2 architecture is learning features of heavy-flavour jets that are well modelled by simulation. A full breakdown of the algorithm performance of GN2 and DL1d is shown in Figure 6.6, for both the baseline MC and with the data-to-simulation scale factors applied. This shows that the significant performance

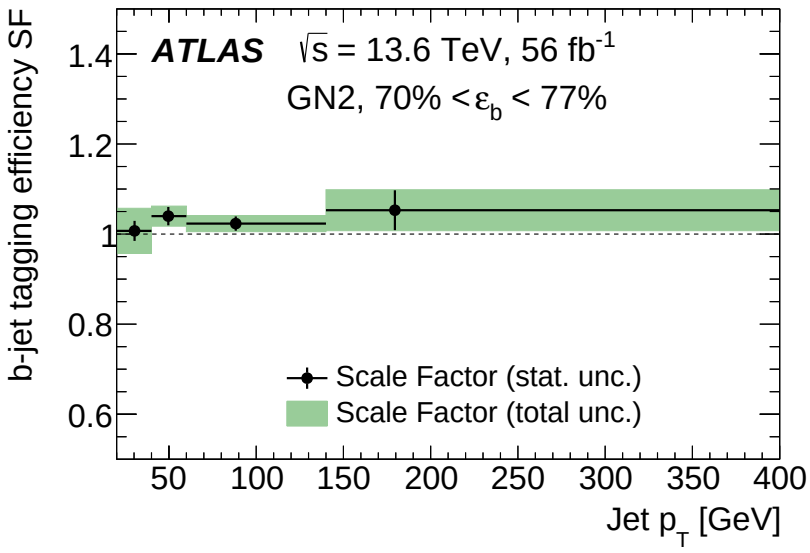


Figure 6.5: Plot of the Run 3 b-tagging efficiency data-to-simulation b-jet SFs as a function of the jet p_{T} as measured in $t\bar{t}$ events for the GN2 algorithm in the pseudo-continuous bin between the 77% and 70% fixed-cut working points. The total uncertainty is shown in the green band, and the statistical uncertainty as an error bar, which becomes significant in the final bin of the calibration. The data are shown at the geometric bin centre. Figure from Ref [170].

gains achieved by GN2 persist beyond the correction of any mismodelling. However, the data-to-simulation corrections are larger for GN2 than for DL1d, indicating that the lower-level features exploited by GN2 exhibit poorer modelling within MC.

6.2.2 Charm-mistag Efficiency

To calibrate the charm-jet mistag rate, a sample of semi-leptonic $t\bar{t}$ decays is used. Here, one side decays via $W \to l\nu$ and the other via $W \to jj$, with the latter's charm-enhanced flavour composition dependent on the size of the CKM elements: $p(W \to cj) \sim 0.5$. With both genuine b and c-jets in the final state, the b and c scale factors are derived in a simultaneous likelihood fit to D_b [171].

6.2.3 Light-mistag Efficiency

For the light-jet mistag rate, since radiative production of heavy flavour is suppressed, Z+jets events enable a light-enhanced sample to be defined. To further enhance the light-flavour composition, a flipped tagger is used. This tagger reverses the

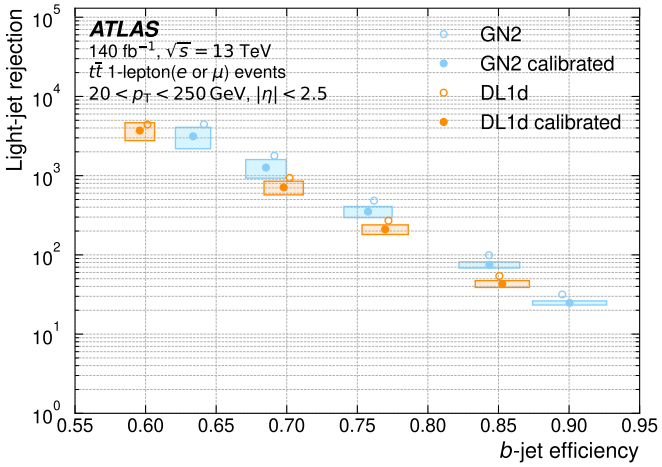


Figure 6.6: Plot of the Run 2 light-jet rejection as a function of the b-tagging efficiency within the $20 < p_{\mathrm{T}} < 250$ GeV bin as measured in $t\bar{t}$ events for the `GN2` (blue) and `DL1d` (green) algorithms, with the baseline (hollow) and calibrated (filled) performance. The performance gains made by `GN2` exist beyond the data-to-simulation scale-factor corrections. Figure from [1].

sign of lifetime-sensitive quantities before evaluating the discriminant, suppressing genuine heavy-flavour jets while leaving the approximately symmetric light-flavour component largely unchanged. A likelihood fit is then performed on the secondary-vertex mass to extract the light-jet mistag efficiency scale factors.

As illustrated with the discriminant plot of Figure 6.2, the rejection of light-flavour jets by the `GN2` algorithm is extremely strong. Consequently, the tightest working point of the calibration analysis, even after using a flipped tagger, is dominated by genuine heavy-flavoured jets and the remaining light-fraction is strongly statistically limited. In contrast to `DL1d`, the calibration of the tightest bin for `GN2` therefore relied on an adjusted-MC method [172], for which the author was involved in initial feasibility studies and in the development of the software frameworks used by the analysis.

6.2.4 Extrapolation to High-p_{T}

The mainstream calibrations rely on a sufficiently well populated region of phase space to perform the fit. In the final kinematic bin of the b-jet calibration, as shown in Figure 6.5, the statistical uncertainty begins to dominate the total uncertainty band, with insufficient statistical precision to provide a reliable representation for

jet p_{T} above this. The b, c and light calibrations suffer from low statistics above a jet p_{T} of 400 GeV, 300 GeV and 250 GeV respectively. Dedicated simulation-based extrapolation uncertainties are therefore assigned to cover the modelling of flavour-tagging performance outside of the directly calibrated region. By observing the approximate p_{T} independence of the scale factors up to the high p_{T} threshold, the extrapolation procedure takes the central value of the final kinematic bin of the mainstream calibration as constant up to 3 TeV. Support for this approach comes from alternative calibration methods that use multi-jet events, and enhance the b-flavour composition by exploiting the kinematics of b-hadron decays into muons within the jets, providing estimates of the scale factors up to 1.2 TeV, where the flat shape of the scale factors is observed well beyond the mainstream calibration [173].

The general methodology for an MC-based uncertainty involves varying a specific element of the simulation and recomputing the tagger discriminant. The relative difference of the tagging efficiency between the nominal, ε_i, and varied sample, $\varepsilon_{i,s}$, is taken as an uncertainty,

$$\delta\varepsilon_{i,s}(p_{\mathrm{T}}) = \frac{\varepsilon_i(p_{\mathrm{T}}) - \varepsilon_{i,s}(p_{\mathrm{T}})}{\varepsilon_i(p_{\mathrm{T}})}. \tag{6.4}$$

Variations to parameters that are important for jet flavour discrimination, such as the choice of parton shower generator or the track impact parameter resolution, are amongst the parameters considered. Since all uncertainties considered are also applied to the mainstream calibration analyses, the value of the systematic in the final kinematic bin ($p_{\mathrm{T,ref}}$) is subtracted to form an independent extrapolation uncertainty.

The total extrapolation uncertainty is quantified as the quadrature sum of all contributions,

$$\sigma^2_{\mathrm{extrap}}\left(p_{\mathrm{T}}; p_{T;\mathrm{ref}}\right) = \sum_{s\in S}\left[\delta\varepsilon^{\mathrm{MC}}_{i,s}\left(p_{\mathrm{T}}\right) - \delta\varepsilon^{\mathrm{MC}}_{i,s}\left(p_{\mathrm{T,ref}}\right)\right]^2. \tag{6.5}$$

Where S is the full set of systematics and $p_{\mathrm{T,ref}}$ is the transverse momentum at which the extrapolation begins, 400 GeV for b-jets, 300 GeV for charm-jets and 250 GeV for light-jets [174].

For DL1d, each uncertainty component was required to be monotonically increasing to avoid artificial reductions in uncertainty from statistical fluctuations. However, when moving to the PCBT and PCFT schemes, each jet is assigned to a particular bin of the discriminant. Therefore, the unitarity of the tagger ($\sum_{j\in\text{PCBT bins}} \Delta\varepsilon_j = 0$) is a relevant quantity to be considered. The monotonicity requirement in each bin can therefore artificially create or remove jets across the full set of bins, and this requirement is relaxed for the PCBT and PCFT GN2 calibrations.

The full set of systematic uncertainties considered for the calibration is given in Table 6.2. A few of these systematics are particularly relevant for flavour-tagging, and the largest uncertainty components are briefly summarised, starting with the parton shower modelling uncertainty.

Parton Shower

The evolution of a final-state b-quark down to the stable hadrons that are measured inside a jet depends on both the perturbative emissions inside the parton shower and the non-perturbative hadronisation. The various choices made within a parton shower, including the ordering variable, the treatment of heavy-quark masses, and the hadronisation model, directly affect flavour-tagging observables, notably the number of charged tracks, the angular displacement of decay products, the secondary-vertex masses and the rate of particular hadron-species production.

Defining 1σ variations for parton showers is challenging, but in principle the variation should encapsulate the different possible choices. Typically, the systematic is chosen as the difference between Pythia and Herwig, which probes substantially different modelling assumptions, including p_{T}-ordered versus angular-ordered shower evolution and Lund-string versus cluster hadronisation. Successive iterations are tuned to increasingly precise data and their predictions for flavour-sensitive observables improve; this systematic uncertainty therefore decreases as the phenomenological models become more precise.

Final state radiation

The amount of final-state radiation (FSR) produced in the parton shower depends on the value of the strong coupling, $\alpha_s(\mu_R)$, evaluated at the shower renormalisation scale μ_R. Lowering μ_R increases α_s, typically enhancing radiation, impacting the formation of the jet. A systematic is defined by taking the envelope of renormalisation scale variations,

$$\delta\varepsilon_{\text{FSR}} = \max_{A} \left| \frac{\varepsilon(A\mu_R) - \varepsilon(\mu_R)}{\varepsilon(\mu_R)} \right|. \tag{6.6}$$

where $A \in [0.625, 0.75, 0.875, 1.25, 1.5, 1.75, 2.0]$.

Tracking in Dense Environments

The accurate reconstruction of charged tracks, and therefore a variety of flavour-sensitive observables, is central for algorithm performance at high-p_{T}. Derived in the core of high-p_{T} jets, the efficiency systematic accounts for the scenario in which two tracks traverse spatially close within the detector. The fraction of lost tracks, F_{Lost}, is derived by constructing templates of $\frac{dE}{dx}$ in pixel clusters. Those consistent with two particles will have a distinct template that can be separated from that of the single-particle cluster. Therefore, if a cluster has only one reconstructed track but a $\frac{dE}{dx}$ shape consistent with two, then the second track is considered lost. The uncertainty attributed to this effect is F_{Lost}, and for each track there is a probability equal to F_{Lost} for this track to be removed from reconstruction. Following this reduction in track reconstruction efficiency, the flavour-tagging discriminant is recomputed, and the difference in efficiency taken as an uncertainty.

Category	**Variation**
Tracking	**Description**
d_0 IP resolution	Transverse IP resolution
z_0 IP resolution	Longitudinal IP resolution
d_0 dead modules	Impact of dead-material in ID on d_0
z_0 dead modules	Impact of dead-material in ID on z_0
q/p sagitta bias	Track-curvature sagitta-bias
z_0 bias	Longitudinal IP bias
d_0 bias	Transverse IP bias
Fake rate	Rate of fake tracks, extracted from non-linearity in the rate of primary and secondary tracks
Passive material IBL	Passive material in the Insertable B-Layer
Passive material PP0	Passive material in Patch Panel 0
Passive material ID	Passive material in the inner detector
GEANT4 physics model	Detector-simulation physics model
Track efficiency in dense environments	F_{Lost} variation to account for tracks missed in reconstruction inside dense jet core
Jet	**Description**
JER effective NP 1–7	Effective components of the JER uncertainty
JES grouped NP 1–3	Grouped components of the JES uncertainty
JER noise term	Contribution to the JER uncertainty from detector noise and pile-up fluctuations
η-intercalibration non-closure	Residual uncertainty from imperfect closure of the JES calibration between central and forward jets
In situ non-closure	Residual JES uncertainty from imperfect closure of the in situ calibration procedure in simulation
Jet Vertex Tagger	Uncertainty associated with the jet–vertex association used to reject pile-up jets
Modelling	**Description**
Parton shower	Choice of parton shower model between PYTHIA and HERWIG
Final-state radiation	Envelope of renormalisation scale variations, changing the rate of emissions in showering

Table 6.2: Systematic variations considered for the high-p_{T} extrapolation.

Results of the Extrapolation

Figure 6.7 shows the Run 2 DL1d extrapolation uncertainties for b-tagging efficiency and for charm and light mistag efficiencies in the 77% fixed-cut WP. The leading systematic for bottom and charm jets is the choice of parton shower, which grows steadily with p_{T} into the extrapolation region. Differences in track and jet reconstruction are subleading, with only minor contributions from FSR variations. For light jets, the contributions are more comparable, and variations in tracking resolution constitute the largest uncertainty. The author prepared the sample migration from Herwig 7.1 to Herwig 7.2, which, with the inclusion of improved heavy flavour fragmentation modelling, is more consistent with data and with Pythia flavour-tagging predictions, reducing this systematic component for future calibrations. Extrapolation uncertainties were also derived for DL1d in Run 3, but a bug was later found in the MC campaign corresponding to the training of the NN that attempts to resolve merged pixel clusters; this bug impacted track reconstruction for all physics simulations within ATLAS, but particularly at high-p_{T} given the dense collimated jet core and consequently the results of the extrapolation were not used. This was resolved in the subsequent MC23 release, which is the MC campaign used for the Run 3 GN2 results presented here.

A summary plot of the extrapolation uncertainties for Run 2 and 3 of the GN2 algorithm in the 77%–70% PCFT tag-weight bin is shown in Figure 6.8, where the parton shower and renormalisation scale variations are leading for charm mistagging and the newly introduced tracking efficiency in dense environments variation leading for b-tagging and light mistagging. Figure 6.9 shows the nominal b-tagging efficiency of the GN2 algorithm together with the F_{Lost} tracking variation (a) which is dominant for b and light jet uncertainties, where the reduction in tracking efficiency greatly reduces the tagging efficiency of genuine b-jets and in (b) for light-jets being mistagged as b-jets, where the reduction in track efficiency increases the likelihood for a fake heavy-flavour jet.

The author found improperly implemented bias maps, which account for detector misalignment that is missed by χ^2-based detector alignment methods [175], and

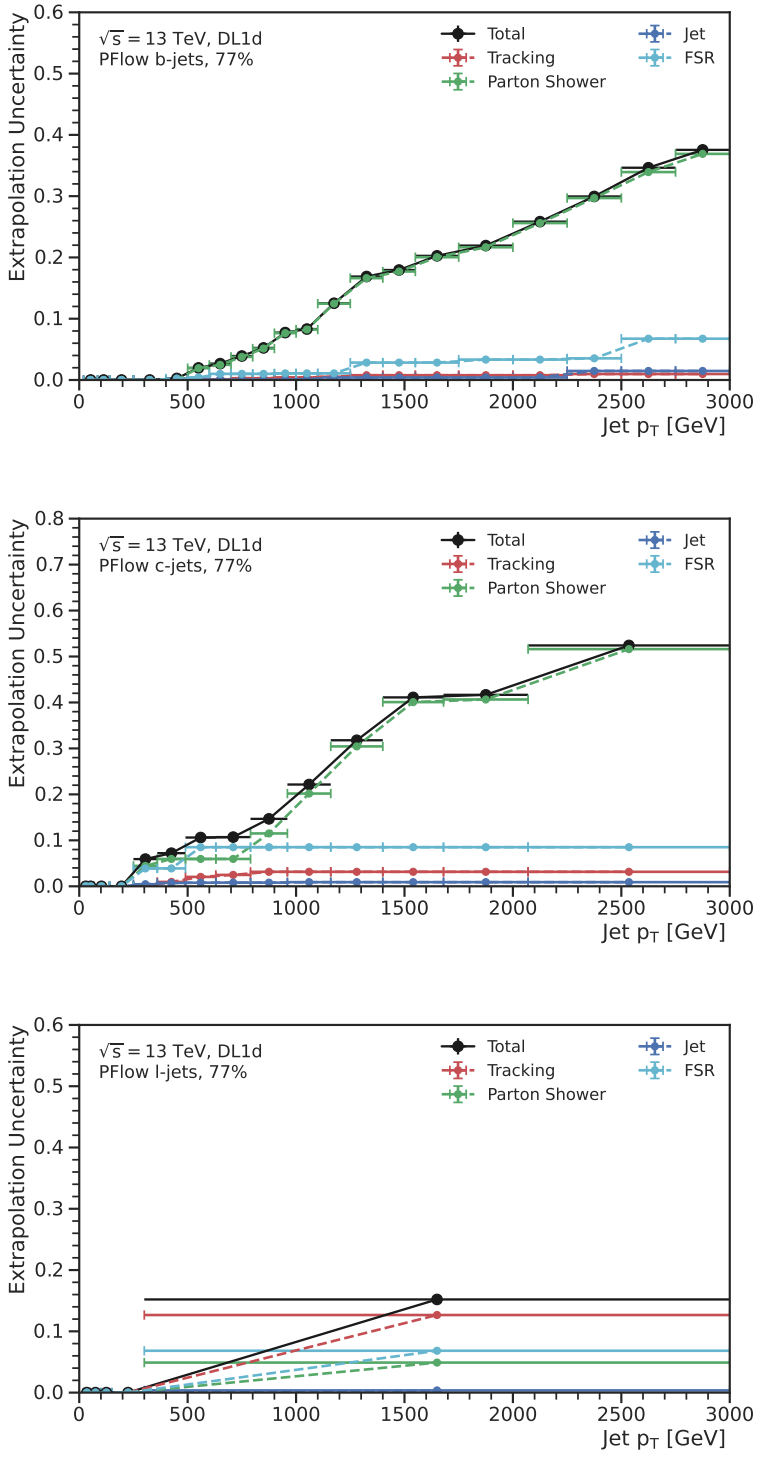


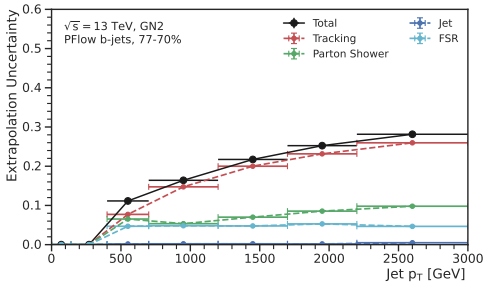
Figure 6.7: Extrapolation uncertainties for truth b, c and light-jets for the `DL1d` flavour-tagging algorithm in the 77% fixed-cut working points. The monotonically increasing uncertainties are shown for Run 2 simulation. Differences between Pythia and Herwig are dominant for very high-p_{T} jets.

the results presented use an earlier release of maps with a conservative inflation of $\sigma \mapsto \sigma^{'} = 1.5\sigma$. For both genuine b-jet tagging and light-jet mistagging, resolution systematics that smear the distribution of track impact parameters induce non-negligible differences at low-p_{T}.For b-jets, due to the shorter flight distance of a lower-momentum b-hadron, track are reconstructed with smaller impact-parameter significance and thus more sensitive to resolution variations, and for light jets the resolution smearing can make tracks appear more displaced, mimicking the characteristics of a genuine b-jet.

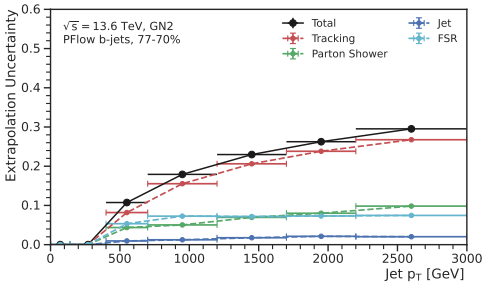
Although hadronic interactions with detector material are included in the detector simulation, the material nuisance parameters do not provide a dedicated variation of the rate at which such interactions produce secondary-vertex-like signatures in light-flavour jets. Initial studies were performed by preferentially up-weighting jets in the sample containing tracks originating from material interactions in the inner detector, thereby mimicking an increase in material budget. Efficiency differences of order $\mathcal{O}(10)\%$ were observed — which, together with similar studies at low-p_{T} for the light-jet calibration, motivated dedicated samples with varied material budgets to be requested for more rigorous systematics to be defined accounting for hadronic interactions within the ID in the future.

Jet scale and resolution systematics shift the overall kinematics but only mildly affect the internal structure of the jet that is sensitive to heavy flavour; therefore, jet variations constitute a small fraction of the overall uncertainty. Moreover, with the relatively minimal detector and reconstruction changes relevant to b-tagging between Run 2 and Run 3, the extrapolation uncertainty remains stable across the campaigns. A marginal increase in jet uncertainties is observed due to a conservative non-closure systematic between Run 2 and 3 calibrations, which is applied only in Run 3.

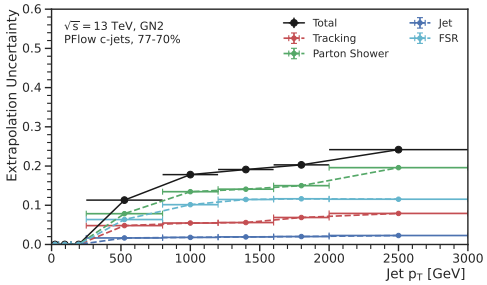
The calibration for `DL1d` was finalised in October 2023, whereas the latest samples used for the results presented in this thesis for `GN2` were produced in October 2024. Therefore, direct comparisons between the two taggers are difficult, because even within the same MC campaign, the applied variations differ substantially. Between these two periods, the parton shower systematic variation was updated to the



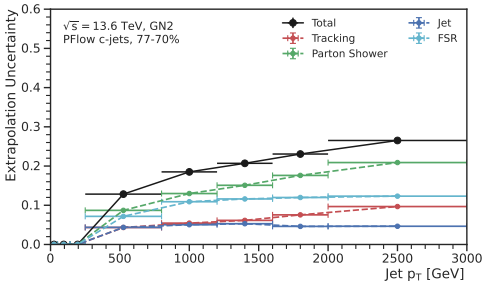
(a) Bottom jets, Run 2 simulation.



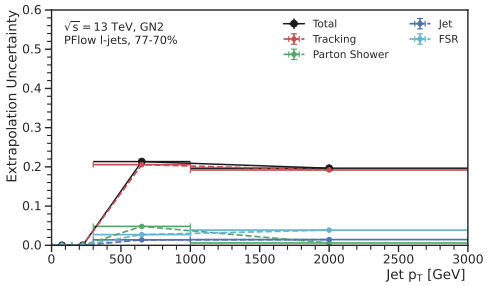
(b) Bottom jets, Run 3 simulation.



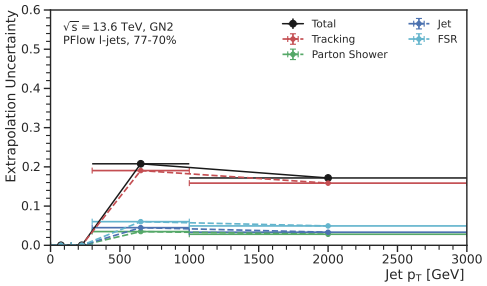
(c) Charm jets, Run 2 simulation.



(d) Charm jets, Run 3 simulation.



(e) Light jets, Run 2 simulation.



(f) Light jets, Run 3 simulation.

Figure 6.8: Extrapolation uncertainties for the GN2 flavour-tagging algorithm in the pseudo-continuous bin between the 77% and 70% fixed-cut working points are shown. The b-tagging extrapolation uncertainties are shown for Run 2 simulation in (a) and Run 3 simulation in (b). Track-reconstruction efficiency in the dense core of highly boosted jets is the dominant component in both runs. The charm-mistag extrapolation uncertainties are shown for Run 2 simulation in (c) and Run 3 simulation in (d). Modelling uncertainties of parton shower and variations tuning the renormalisation scale $\alpha_s(\mu_R)$. The light-mistag extrapolation uncertainties are shown for Run 2 simulation in (e) and Run 3 simulation in (f). Track-reconstruction efficiency in the dense core is again the leading systematic uncertainty.

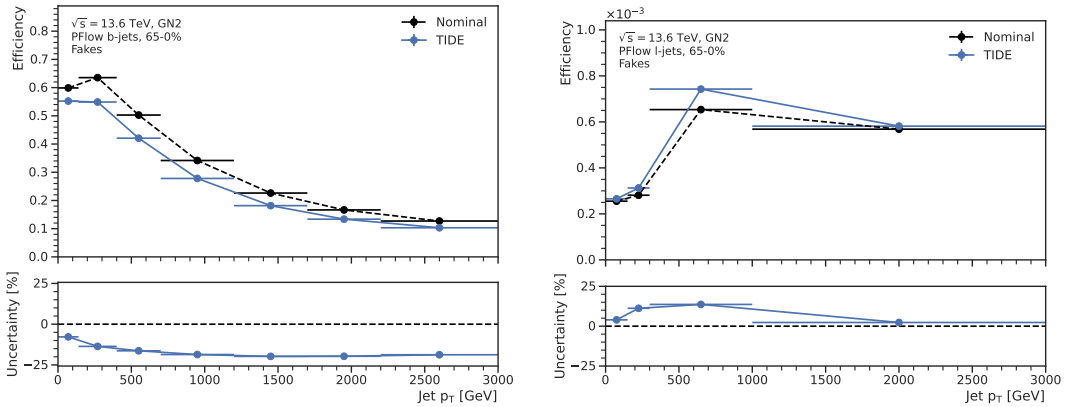


Figure 6.9: Plot of the b-tagging efficiency the two leading truth b-jets as a function of the jet p_{T} in Z' events for the GN2 algorithm for the nominal sample, and one with the F_{Lost} tracking variation (a) for genuine b-jets and (b) for light-flavoured jets in the tightest b-tagging bin of 65%–0%. The percentage difference between the two is shown in the lower panel, and can be converted into an extrapolation uncertainty by subtracting the uncertainty for the 400 GeV bin.

latest models, substantially reducing the observed systematic uncertainty, together with major updates to the tracking nuisance parameters — chiefly the inclusion of the very large tracking-in-dense-environments systematic uncertainty which dominates the extrapolation. Nonetheless, qualitative statements can be made using earlier versions of GN2 compared with DL1d. Here, the GN2 architecture, despite the increased model complexity, is less dependent on parton shower variations — perhaps in part due to its use of lower-level input feature correlations which form more physically robust representations of heavy-flavour jets as opposed to more rigid higher-level representations and low-level tagger assumptions within DL1d. Further investigation could utilise the Lund-jet plane [176] to more precisely understand where the differences between the two generators lie and what exactly the network is learning, and would constitute an interesting follow-up study.

6.2.5 Usage in Analysis

Once the scale factors have been derived in dedicated calibration analyses and the full set of uncertainties provided to extrapolate the calibration to high transverse momentum, they must be deployed within analyses across ATLAS. The scale factors

are applied for each jet with weight,

$$SF_f(p_{\mathrm{T}}) = \frac{\varepsilon_f^{\mathrm{data}}(p_{\mathrm{T}})}{\varepsilon_f^{\mathrm{MC}}(p_{\mathrm{T}})} \tag{6.7}$$

with untagged jets receiving weight,

$$w_{\mathrm{untagged}}(p_{\mathrm{T}}) = \frac{1 - SF_f(p_{\mathrm{T}})\varepsilon_f^{\mathrm{MC}}}{1 - \varepsilon_f^{\mathrm{MC}}} \tag{6.8}$$

where f denotes the truth flavour assigned to the jet. A product over individual jet scale factors is performed to obtain the final b-tagging weight for the event.

The calibration analyses described above are performed using a set of baseline generators and parton shower models, leaving $\varepsilon_f^{\mathrm{MC}}$ dependent on the exact MC used. To account for this, MC-MC scale factors are derived and used,

$$SF_f^{\mathrm{alt}}(p_{\mathrm{T}}) = \frac{\varepsilon_f^{\mathrm{data}}(p_{\mathrm{T}})}{\varepsilon_f^{\mathrm{MC}}(p_{\mathrm{T}})} \frac{\varepsilon_f^{\mathrm{MC}}(p_{\mathrm{T}})}{\varepsilon_f^{\text{alt-MC}}(p_{\mathrm{T}})} = \frac{\varepsilon_f^{\mathrm{data}}(p_{\mathrm{T}})}{\varepsilon_f^{\text{alt-MC}}(p_{\mathrm{T}})} \tag{6.9}$$

to transport the central value of the mainstream calibration to the MC used by the analysis.

The set of uncertainties considered by each calibration analysis is large ($\mathcal{O}(100)$). Often, these variations have little impact on the calibration, and an ATLAS analysis using such a scheme would be insensitive to their variation. It is therefore chosen to provide a smaller number of uncertainty eigenvariations. The smallest eigenvalues are often negligibly small, and the 9 leading eigenvariations from the b calibration and the 4 leading from c and light calibrations are provided as nuisance parameters encoding the 1σ variation to the scale factor at the analysis level. A number of nuisance parameters in the calibration fits are common amongst analyses, like JES variations, and the original covariance matrices can be recombined by an analysis using an eigenvector recomposition tool [177] — particularly useful for cross-collaboration combinations for which the eigenvectors have no direct physical meaning.

6.2.6 Conclusions

The extrapolation presented extends the validity of the calibration derived at low p_{T} to much higher p_{T} — enabling flavour tagging to be utilised by analyses

operating outside the usual phase space, such as searches for additional heavy Higgs bosons [178]. The author performed the calibration for the `DL1d` and `GN2` architectures, across multiple iterations of flavour-tagging working points, for use in Run 2 and Run 3 data collected by the ATLAS experiment. With the significant performance gains provided by `GN2`, the work presented is used across the entire ATLAS physics programme — particularly relevant for analyses containing multiple heavy-flavour jets, but also for analyses explicitly vetoing them. Moreover, numerous software improvements were made by the author, for the extrapolation itself but also for the wider flavour-tagging community — making future calibration procedures faster and more reliable.

Modelling uncertainties, including FSR and parton shower variations, are large, but increasingly refined tuning and modelling reduce this systematic over time. However, increasingly expressive algorithms may learn structures inherent to these simulations, and this systematic will remain a crucial validation stage. The highly dense tracking environment within the cores of high-p_{T} b-jets makes the identification of secondary vertices extremely challenging, and the fraction of tracks lost in reconstruction constitutes the largest uncertainty in the extrapolation procedure.

Machine learning is a rapidly evolving landscape, and the significant step made by `GN2` will be made yet again with `GN3` [179], for which both the calibration and extrapolation are necessary ingredients for these advancements to be utilised for increasingly sensitive searches and measurements within ATLAS.

"Have you tried looking under the shed?"

— Steve and Harry Dingley

7

Hunting in the Dark with $HH+p_{\mathrm{T}}^{\mathrm{miss}}$

Contents

7.1 Signal Models and Kinematics . . . 109
7.2 Background Processes and Event Generation . . . 112
7.3 Dataset and Triggers . . . 113
7.4 Analysis Objects . . . 113
7.5 Preselection . . . 116
7.5.1 Higgs Reconstruction . . . 119
7.6 Machine Learning . . . 120
7.6.1 Model Architecture . . . 120
7.6.2 Input Features . . . 121
7.6.3 Decorrelation . . . 124
7.7 Background Estimation . . . 126
7.7.1 Control Regions . . . 127
7.7.2 Signal Region . . . 129
7.8 Systematic uncertainties . . . 132
7.8.1 Experimental Uncertainties . . . 132
7.8.2 Theoretical Uncertainties . . . 132
7.8.3 Systematic Smoothing: single top . . . 133
7.9 Statistical Model . . . 134
7.10 Fit Validation . . . 135
7.10.1 Background-only Fit . . . 136
7.10.2 Exclusion Fit . . . 140
7.11 Unblinded Results . . . 144
7.11.1 Model-independent Fit and Results . . . 144
7.11.2 Model-dependent Results . . . 145
7.12 Conclusions . . . 149

This chapter will detail a search for a dark Higgs boson in the $HH(\to 4b)+p_{\mathrm{T}}^{\mathrm{miss}}$ final state, performed using the full 140 fb^{-1} of proton-proton data collected during Run 2 by the ATLAS experiment. A loose preselection is applied to select the high-$p_{\mathrm{T}}^{\mathrm{miss}}$ and $4b$ final state, with additional requirements to suppress $t\bar{t}$ and $Z+$jets backgrounds. The distinctive recoil kinematics of dark-Higgs strahlung are exploited with an NN classifier. The dark Higgs mass is reconstructed using

the fully visible final state, $m_s \sim m_{HH}$, and is therefore a key kinematic signature for the analysis. However, without sufficient care, the NN architecture may indirectly learn to exploit m_{HH} in the separation task. This means that any cut on the classifier score will distort this kinematic feature, rejecting background events that do not lie in the simulated signal window, known as *sculpting*. To avoid this, a modified loss function is used to ensure statistical independence between the NN score and m_{HH} — preserving this key signature to be exploited in the statistical interpretation. A complete account of the analysis is given, but since the author contributed predominantly to background validation, systematic derivation, statistical interpretation and subsequent results, greater attention is paid to these elements.

7.1 Signal Models and Kinematics

The dark Higgs model, as detailed in Section 3.3.4, introduces three additional massive states to the SM: a Majorana dark matter candidate (χ), a massive vector mediator (Z') and a dark Higgs boson (s). The dark Higgs strahlung process is targeted due to its distinct recoil kinematics, producing a large $p_{\mathrm{T}}^{\mathrm{miss}}$ final state back-to-back with the $4b$ system and is shown in Figure 7.1.

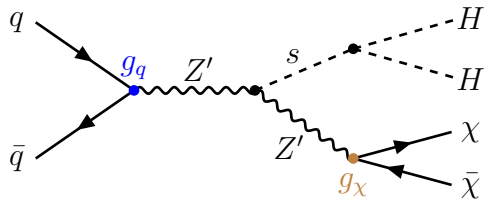


Figure 7.1: Feynman diagram for dark-Higgs strahlung in the dark Higgs model, coupling modifiers are shown for the Z' to quarks, g_q, and the coupling in the dark-sector, g_χ.

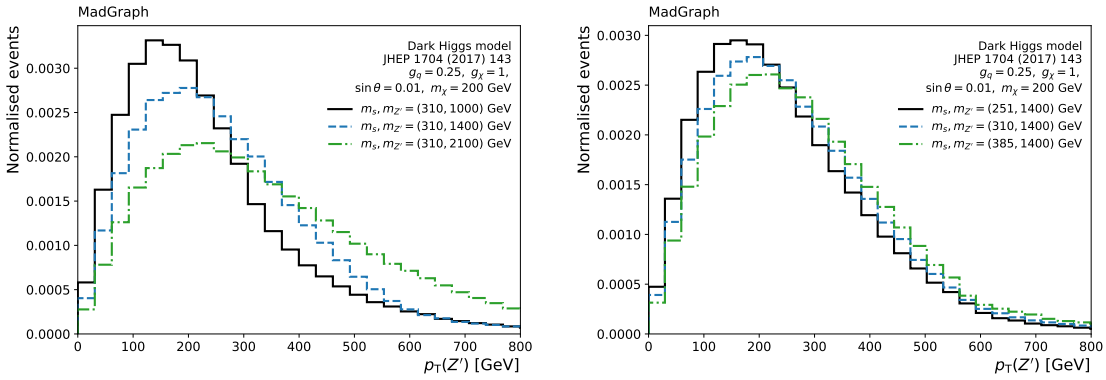
There are six parameters in the model, and the benchmark targeted in the analysis fixes $g_q = 0.25$, $g_\chi = 1$, $\sin\theta = 0.01$ and the dark matter mass to $m_\chi = 200$ GeV. Since the only two free remaining parameters from the benchmark are $m_{Z'}$ and m_s, one must choose a sensible range for both. The range of dark Higgs masses is extracted from the branching fraction plot of Figure 3.7, where the

HH final state opens up from the kinematic threshold of $2m_H$, with appreciable branching fraction until the $s \to \chi\chi$ channel opens up and becomes dominant for masses above $2m_\chi$. Therefore, for this benchmark, the mass of the dark Higgs boson is chosen in the range $251 < m_s < 435$ GeV.

If the dark Higgs boson is produced at rest in the transverse plane, the Z' decay products ($\chi\bar{\chi}$) will be back-to-back and cancel out the key kinematic signature of $p_{\mathrm{T}}^{\mathrm{miss}}$. The mass range of the Z' is therefore motivated by requiring sufficiently large $p_{\mathrm{T}}^{\mathrm{miss}}$, above trigger thresholds, with a large enough cross section such that a search may be sensitive with Run 2 at the LHC. For $2m_\chi < m_{Z'} < m_s + 2m_\chi$, the Z' is first produced off-shell, proceeding via $pp \to Z'^* \to Z's$, with $Z' \to \chi\chi$ and $s \to HH$. For fixed m_s and m_χ, increasing $m_{Z'}$ increases the available phase space and allows increasingly hard recoil between the dark Higgs and the invisible system. The behaviour for fixed $m_s(m_{Z'})$ and varied $m_{Z'}(m_s)$ is shown in Figure 7.2a(b), where larger candidate masses raise the scale of the event, increasing the amount of transverse recoil. The larger amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ generated by larger mediator masses competes with the sharply falling cross section with both $m_{Z'}$ in the propagator and PDF suppression for sufficiently high-mass Z' candidates (approaching $x \sim 0.25$ for masses above 3 TeV). There is also a lower bound set by the requirement that $m_{Z'} > 2m_\chi$ for the Z' to decay into DM. The final chosen mass window of $420 < m_{Z'} < 3300$ GeV enables a hard recoil of the dark Higgs, producing a high-$p_{\mathrm{T}}^{\mathrm{miss}}$ final state, whilst maintaining sufficient cross section to perform the search.

The range of dark Higgs and mediator masses targeted remain unconstrained by previous searches, as shown earlier in Figure 3.8, allowing previously unexplored regions of parameter space to be probed. More broadly, the analysis probes the signature of resonant di-Higgs production in association $p_{\mathrm{T}}^{\mathrm{miss}}$ for the first time.

The cross sections for the selected signal points are shown in Figure 7.3, and a balance between signal cross section, shape of the m_{HH} background distribution and favourable kinematics will ultimately dictate where the analysis is most sensitive.



(a) **(b)**

Figure 7.2: Truth transverse momentum of the Z' in the dark Higgs signal model for the Run 2 benchmark with three mass points: $(m_s, m_{Z'}) = (310, 1000), (310, 1400), (310, 2100)$ GeV as generated at leading-order in MadGraph 5. The transverse momentum of the Z' is a direct handle for the truth $p_{\mathrm{T}}^{\mathrm{miss}}$ present in the event, and the amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ increases with $m_{Z'}$ and m_s.

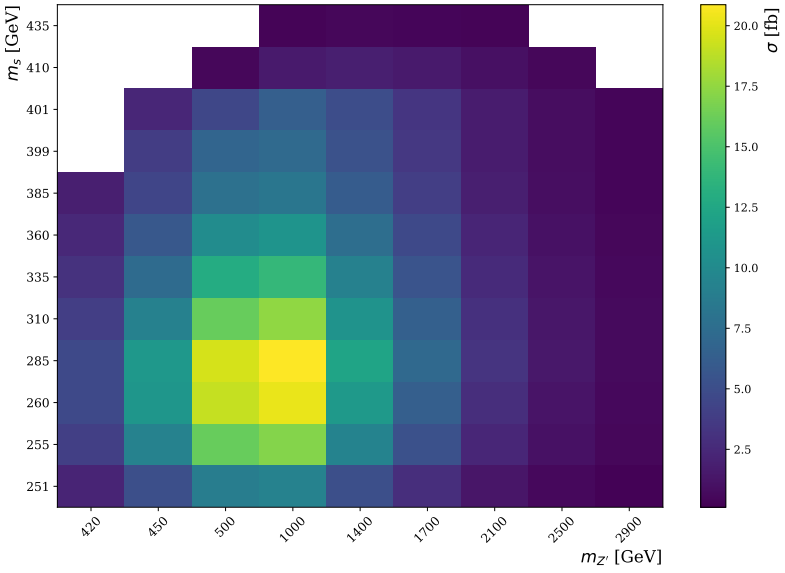


Figure 7.3: Cross sections as predicted by MadGraph across the full range of signal masses for the dark Higgs model in the Run 2 benchmark.

7.2 Background Processes and Event Generation

Given the $4b + p_{\mathrm{T}}^{\mathrm{miss}}$ final state, dominant SM backgrounds include $t\bar{t}$ produced in various forms. Three classes of $t\bar{t}$ production are defined for the analysis: the production of $t\bar{t}$ in association with additional truth b-jets ($t\bar{t} + \geq 1b$), and two categories involving mistagged charm jets ($t\bar{t} + \geq 1c$) and mistagged light jets ($t\bar{t} + \text{light}$). The source of $p_{\mathrm{T}}^{\mathrm{miss}}$ in these processes may arise from mismeasured jet activity or via the leptonic decay of a W-boson, with the lepton subsequently being missed in the reconstruction, alongside the weakly interacting neutrino. Another dominant background is that of $Z(\rightarrow \nu\nu) + \text{jets}$, which is kinematically very similar to the signal, with a pair of neutrinos forming the contribution of $p_{\mathrm{T}}^{\mathrm{miss}}$, back-to-back with jet activity. Other, sub-leading, background processes include single top (Wt), $W+\text{jets}$, top quarks in association with electroweak bosons (ttW, ttZ, ttH), and diboson (WW, WZ, ZZ) production. Triboson and VH background samples were also tested but contributed negligibly to the final background yield. Since the signal final state has a large $p_{\mathrm{T}}^{\mathrm{miss}}$, which is well-separated from the $4b$-system, the multi-jet QCD contribution is negligible — more detail will be provided later in Section 7.

The MC simulated events are processed through a full simulation of the ATLAS detector [50] based on GEANT4. The effect of pile-up in the same and neighbouring bunch crossings was modelled by overlaying the simulated hard-scattering event with inelastic proton–proton events generated with PYTHIA8.186 [180] using the NNPDF2.3lo set of parton distribution functions (PDF) [181] and the A3 set of tuned parameters [182]. All samples used are listed in Table 7.1, along with the matrix element generator, parton shower, relevant parton distribution function (PDF) sets, and the cross section order. For PYTHIA (SHERPA) samples the `A14` [183] (SHERPA default [184]) tunes are used.

Table 7.1: Simulated event samples with the corresponding matrix-element and parton-shower generators, cross section order used to normalise the event yield, and matrix-element PDF set.

Process	Generator (ME)	Parton shower	Normalisation	PDF (ME)
dark Higgs	MadGraph5_aMC@NLO[2.9.9] [40]	Pythia 8.307 [185]	LO	NNPDF2.3LO [181]
$Z+$jets	Sherpa 2.2.1 [186, 187]	Sherpa 2.2.1	NNLO [188]	NNPDF3.0NNLO [53]
$W+$jets	Sherpa 2.2.1	Sherpa 2.2.1	NNLO	NNPDF3.0NNLO
diboson	Sherpa 2.2.1	Sherpa 2.2.1	NLO	NNPDF3.0NNLO
$t\bar{t}$	Powheg Box v2 [41–43, 189]	Pythia 8.230	NNLO+NNLL [190]	NNPDF3.0NLO
single top	Powheg Box v2	Pythia 8.230	NLO [191, 192]	NNPDF3.0NLO
ttW/ttZ	MadGraph5_aMC@NLO[2.3.3]	Pythia 8.210	NLO	NNPDF3.0NLO
four-top	MadGraph5_aMC@NLO[2.2.2]	Pythia 8.186	NLO	NNPDF2.3LO
ttH	Powheg Box v2	Pythia 8.230	NLO	NNPDF3.0NLO

7.3 Dataset and Triggers

The analysis uses the full Run 2 dataset of the ATLAS experiment at the LHC, amounting to 140 fb^{-1} of data across the 2015-2018 data-taking periods [193]. The requirements made on the data follow collaboration-wide recommendations, which aim to remove periods of data-taking with poor detector quality [194]. Given the high $p_{\mathrm{T}}^{\mathrm{miss}}$ final state, $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers are used. As shown in Figure 7.2, many of the signals have long $p_{\mathrm{T}}^{\mathrm{miss}}$-tails, whereas many of the SM backgrounds typically peak at lower $p_{\mathrm{T}}^{\mathrm{miss}}$ values. Lowering the $p_{\mathrm{T}}^{\mathrm{miss}}$ selection below the 200 GeV trigger plateau region (see Figure 4.6) does not benefit the analysis. Therefore, an offline selection at 200 GeV is chosen to significantly simplify the analysis, whilst negligibly impacting sensitivity.

7.4 Analysis Objects

Multiple objects and object definitions are established for use at various stages of the analysis. `Baseline` objects are used to build $p_{\mathrm{T}}^{\mathrm{miss}}$ and serve as inputs to the overlap removal (OR). `Signal` objects are used to define the event preselection and to build high-level kinematic observables for input to the ML architecture. The object definitions are summarised in Table 7.2, with a brief account of each given below.

`Baseline` jets are calibrated small-radius ($R = 0.4$) jets, as defined in Section 4.3.4, and are required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.8$ to be considered in the analysis. `Signal` jets are built from the subset of `Baseline` jets above

Table 7.2: Summary of object definitions.

Object	Requirement	`Baseline`	`Signal`
electrons	subset of	–	`Baseline`
	p_{T} [GeV]	> 8	> 8
	$\lvert\eta\rvert$	< 2.47	< 2.47
	$S(d_0)$	–	< 5
	$z_0 \sin(\theta)$	< 0.5 mm	< 0.5 mm
	ID	`Loose`	`Loose`
	Isolation	–	`Loose`
	passOR	–	true
muons	subset of	–	`Baseline`
	p_{T} [GeV]	> 8	> 8
	$\lvert\eta\rvert$	< 2.7	< 2.5
	$S(d_0)$	–	< 3
	$z_0 \sin(\theta)$	< 0.5 mm	< 0.5 mm
	ID	`Medium`	`Medium`
	Isolation	–	`Tight`
	passOR	–	True
jets	subset of	–	`Baseline`
	p_{T} [GeV]	> 20	> 25
	$\lvert\eta\rvert$	< 2.8	< 2.8
	JVT	`Medium`	`Medium`
	passOR	–	True
	`DL1r` WP	77%	77%
	b-tag $\lvert\eta\rvert$	< 2.5	< 2.5

25 GeV that pass overlap removal, which is defined below. To suppress pile-up, jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ are required to have JVT > 0.2 (the `Medium` JVT working point as defined in Ref [132]). The `DL1r` algorithm is used to b-tag `Baseline` jets with the 77% efficiency working point, corresponding to a probability of 0.52% (17.7%) for misidentifying light (charm) jets - as measured in a dileptonic $t\bar{t}$ sample with $20 < p_{\mathrm{T}} < 400$ GeV. Although not available in time to be considered by this analysis, the `GN2` algorithm would be a clear improvement[1] on `DL1r` for the analysis presented, with a probability of 0.2% (6%)

[1]For example, in the $3b(4b)$ categories `GN2` would approximately enhance the rejection of $t\bar{t} + \geq 1c$ and $t\bar{t} +$ light assuming one (two) mistagged jets, by a factor of 3(9) — reaching close to an order of magnitude in increased rejection within the most sensitive analysis category. Other background processes, such as $Z +$ jets, will also be reduced by similar orders.

for misidentifying light (charm)-flavour jets, enhancing the rejection of mistagged backgrounds, such as $t\bar{t} + \geq 1c$ and $t\bar{t}$ + light.

With the exception of muons in the Z + jets control region, events containing electrons and muons are vetoed. `Baseline` electrons (muons) are required to have $|\eta| < 2.47$ ($|\eta| < 2.7$) and $p_{\mathrm{T}} > 8$ GeV, and to satisfy the `Loose` likelihood-based identification working point defined in Ref [195] (`Medium` identification working point as defined in Ref [196]). The significance of the transverse impact parameter must be less than 5 (3) for electrons (muons). `Signal` electrons are required to satisfy the `Loose` isolation working point of Ref [195] to reduce contamination from semileptonic hadron decays and jets misidentified as electrons. `Signal` muons are required to satisfy the `Tight` isolation working point of Ref [196] to reduce contamination from semileptonic heavy flavour and hadron decays.

b-hadrons decay into a final state involving a muon around 10% of the time. The muon will not deposit a large fraction of energy within the calorimeters, and since PFlow object reconstruction does not include information from the MS, the jet p_{T} will be underestimated. Therefore, if a `Baseline` muon fails the isolation requirement and lies within $\Delta R < 0.4$ of a `Baseline` jet, the muon four-momentum is added to that of the jet.

Events in this final state are highly active, and multiple physics objects may overlap within the detector. This can lead to double-counting of physical states, potentially biasing event selections or the construction of kinematic observables. Overlap removal is applied sequentially to `Baseline` objects in the following order:

1. **Electron–muon overlap:** If a muon shares an ID track with an electron, the electron is removed.

2. **Electron–jet overlap:** Jets satisfying $\Delta R(j, e) < 0.2$ are removed. For the surviving jets, electrons satisfying $\Delta R(j, e) < \min\left(0.4,\ 0.04 + \frac{10\ \mathrm{GeV}}{p_{\mathrm{T}}^{e}}\right)$ are removed.

3. **Muon–jet overlap:** If $\Delta R(j, \mu) < 0.2$ or the muon ID track is ghost-associated to the jet, the jet is removed provided it has fewer than three

associated tracks with $p_{\mathrm{T}} > 500$ MeV. For the surviving jets, muons satisfying $\Delta R(j,\mu) < \min\left(0.4,\ 0.04 + \frac{10\ \mathrm{GeV}}{p_{\mathrm{T}}^{\mu}}\right)$ are removed.

Events with significant activity not consistent with that of the bunch-crossing are rejected as non-collision background [197]. There are various forms of non-collision background, one example of which is the beam halo. This refers to high-energy muons produced by beam halo interactions with material upstream of the collision, which record signals in the detector and can be a source of anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$. A set of cuts is made which has a high rejection of non-collision backgrounds, while preserving an efficiency close to 100% for `Signal` objects [198]. Cosmic muons are identified and removed after the muon-jet overlap removal as muons with large longitudinal and transverse impact parameters $|z_0| > 1$ mm and $|d_0| > 0.2$ mm, respectively.

7.5 Preselection

An event preselection is defined targeting the $4b$+$p_{\mathrm{T}}^{\mathrm{miss}}$ signature of the signal. Given the range of dark Higgs and Z' masses, and the transverse momentum of the dark Higgs (a close proxy of which is shown in Figure 7.2), the final state b-jets are expected to be well separated and resolved within small-radius ($R = 0.4$) jets. With the signal process containing at most four b-jets, and $t\bar{t}$ backgrounds dominating higher jet multiplicities, a lower bound of 4 and an upper bound of 7 `Signal` jets is placed. Since the signal process will produce no genuine leptons, a `Signal` lepton veto is applied to reduce background.

For the signal, the final state b-jets will be back-to-back in the transverse plane with $p_{\mathrm{T}}^{\mathrm{miss}}$ from the $Z' \to \chi\bar{\chi}$ decay. This kinematic handle enables various $p_{\mathrm{T}}^{\mathrm{miss}}$-based features to be utilised to improve signal-background separation. First, the minimum transverse mass is defined:

$$m_{\mathrm{T,min}}^{b} = \min_{b-\mathrm{jets}} \left\{ \sqrt{2 p_{\mathrm{T}}^{\mathrm{miss}} p_{\mathrm{T}}^{b,i} \left(1 - \cos\left(\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, b^i)\right)\right)} \right\} \tag{7.1}$$

where p_{T}^{b} is the transverse momentum of the ith b-jet. For a system where a parent particle, X, decays into a b-jet and an invisible particle (or an object missed in

reconstruction) — $m^b_{\mathrm{T,min}}$ is bounded from above by the mass of the parent. An example of a background satisfying such a criterion is the top process, in which $m^b_{\mathrm{T,min}}$ from $t \to bl^+\nu$ is bounded by m_t. The large separation between the b-jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ results in higher values of $m^b_{\mathrm{T,min}}$ for the signal process.

Naturally, given the multi-jet final state, we must consider contributions from the multi-jet QCD background. However, since the dark Higgs signal is characterised by large $p_{\mathrm{T}}^{\mathrm{miss}}$ and $m^b_{\mathrm{T,min}}$ values, the QCD background is expected to be small. Furthermore, the $p_{\mathrm{T}}^{\mathrm{miss}}$ contributions from QCD often arise from incorrectly associated or mismeasured jets and thus tend to align with the jet activity in the event. Therefore, a cut on the $p_{\mathrm{T}}^{\mathrm{miss}}$-significance [199] ($S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})> 7$) is used alongside a cut on the minimum azimuthal angle between any of the jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ in the final state,

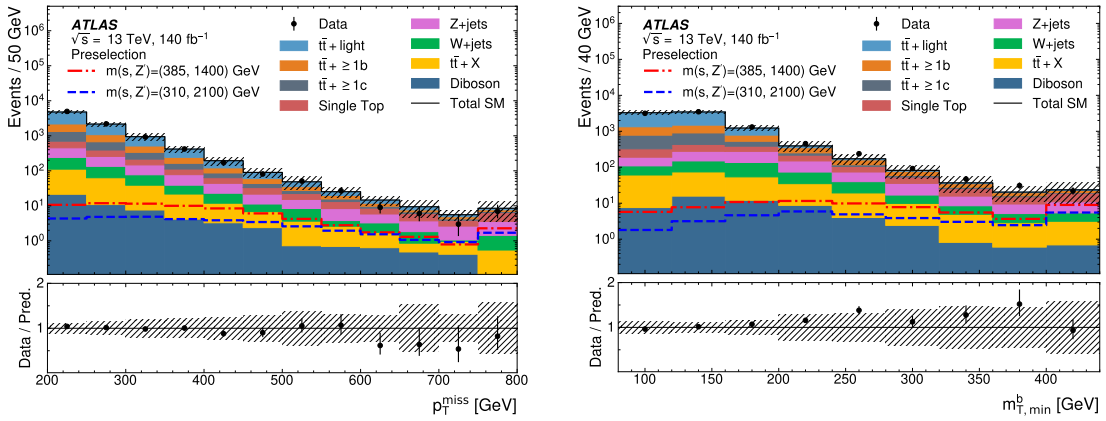
$$\Delta\phi^{4j}_{\mathrm{min}} = \min_{i=1,\ldots,4} \Delta\phi\Big(j_i, \mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}\Big) > 0.4 \tag{7.2}$$

where j_i is the ith leading jet in the final state. $\Delta\phi^{4j}_{\mathrm{min}}$ peaks at low values for QCD, with $p_{\mathrm{T}}^{\mathrm{miss}}$ aligning with the direction of jet activity in the event, and at large values for the well-separated $4b$-$p_{\mathrm{T}}^{\mathrm{miss}}$ system of the signal. Finally, the transverse mass of the $4b$–$p_{\mathrm{T}}^{\mathrm{miss}}$ system, which is well-separated for signal and peaks at high values, must exceed $m_{\mathrm{T}}^{4b} > 300$ GeV. A combination of the selections summarised in Table 7.3 renders the QCD background negligible whilst significantly reducing the otherwise dominant $t\bar{t}$ background.

Table 7.3: Preselection requirements.

Variable	Requirement
n_{leptons}	0
n_{jets}	$[4, 7]$
$n_{b-\mathrm{jets}}$	≥ 3
$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	> 200
$\Delta\phi^{4j}_{\mathrm{min}}$	> 0.4
$m^b_{\mathrm{T,min}}$ [GeV]	> 80
$S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})$	> 7
m_{T}^{4b} [GeV]	> 300

The $p_{\mathrm{T}}^{\mathrm{miss}}$ and $m^b_{\mathrm{T,min}}$ distributions after the event preselection are shown in Figure 7.4. The signal exhibits a characteristic high-$p_{\mathrm{T}}^{\mathrm{miss}}$ tail; however, no additional



(a) **(b)**

Figure 7.4: $p_{\mathrm{T}}^{\mathrm{miss}}$ (a) and $m_{\mathrm{T,min}}^{b}$ (b) distributions after the application of the event preselection as defined in Table 7.3, with all background processes shown together with the observed data. Two signals $m(s, Z') = (385, 1400)$ GeV and $m(s, Z') = (310, 2100)$ GeV are overlaid. The panel below shows the ratio of the observed data to the SM prediction, with all points agreeing with the SM prediction within uncertainties.

requirement on $p_{\mathrm{T}}^{\mathrm{miss}}$ provides an unambiguous improvement in signal–background separation, as is the case for a number of reconstructed kinematic features — pointing towards the use of machine learning to aid in this task. In the $m_{\mathrm{T,min}}^{b}$ distribution, the $t\bar{t}+$light background is concentrated at lower values, while the signal, single-top background and $t\bar{t}+\geq 1b$ component extend to larger $m_{\mathrm{T,min}}^{b}$. A tighter requirement could therefore suppress a substantial fraction of the $t\bar{t}+$light background. However, as discussed later, the associated reduction in the number of events available for training degrades the multivariate classifier's performance. The analysis therefore ultimately benefits from retaining a comparatively loose requirement on $m_{\mathrm{T,min}}^{b}$. The remaining backgrounds considered throughout the analysis are: $Z+$jets, $t\bar{t}+b\bar{b}$, $t\bar{t}+\geq 1c$, $t\bar{t}+$light, single top, $W+$jets, $t\bar{t}+X$ and diboson processes.

An informative display of an event passing the preselection (and eventually the SR) requirements is shown in Figure 7.5. Here, the key kinematic signature of the analysis can be seen: a four b-jet system with p_{T}^{4b} of 715 GeV back-to-back, and relatively balanced, with a significant $p_{\mathrm{T}}^{\mathrm{miss}}$ of 825 GeV. One of the tagged jets contains a semi-leptonically decaying b-hadron whose muon failed isolation.

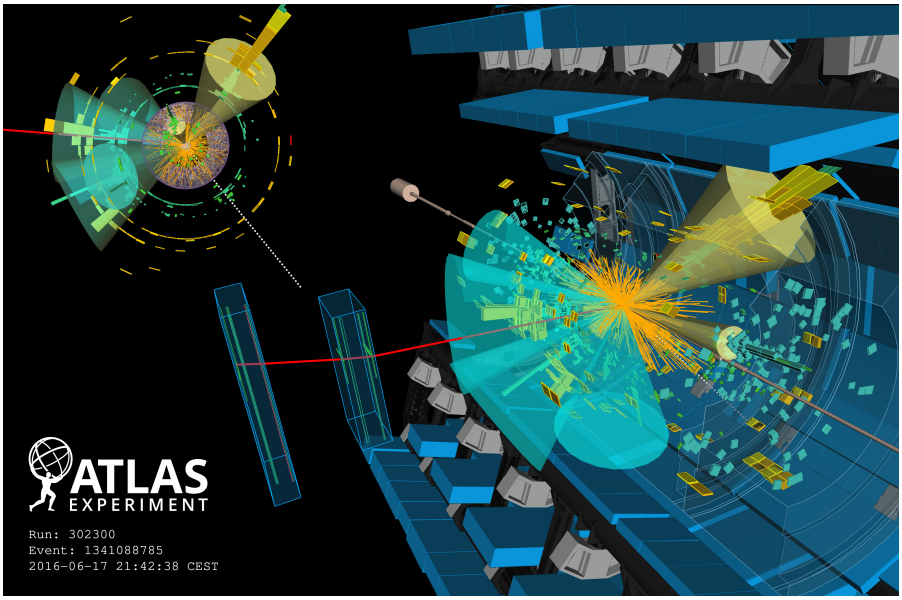


Figure 7.5: Display of a candidate event, projected onto the $x-y$ plane (upper left) and as a three-dimensional view of the ATLAS detector (right). The event contains six reconstructed jets, four of which satisfy the 77% `DL1r` b-tagging working point (shown in cyan), with two further jets (shown in yellow). A significant $p_{\mathrm{T}}^{\mathrm{miss}}$ of 825 GeV is present, indicated by the white dashed line, and is nearly back-to-back with a four b-tagged jet system whose combined transverse momentum (p_{T}^{4b}) is 715 GeV. A single muon candidate, originating from a decay within a b-tagged jet, is shown in red, with hits in the relevant muon spectrometer chambers illustrated in green.

7.5.1 Higgs Reconstruction

The reconstructed di-Higgs invariant mass is a direct proxy of the dark Higgs boson mass, and therefore a powerful discriminant. The analysis distinguishes between regions with three and four b-jets, and in the $4b$ case, jet selection is simple, with $m_{HH} = m_{4b}$. The situation is more sophisticated in the $3b$ case. If the mass of one of these b-jets is greater than 100 GeV, then it is assumed to be a 'merged' jet formed from a high-momentum Higgs decaying into two collimated products, which, with the other two b-tagged jets, forms m_{HH}[2]. If the mass threshold is not met by any of the b-tagged jets, then the fourth jet is chosen from the remaining untagged jets. This jet is chosen to minimise the mass of the heaviest possible Higgs boson candidate constructed from any two of the set containing the three

[2]Further optimisation of this channel may utilise modern large-radius jet reconstruction and tagging algorithms. Here the jet can be tagged as a $H \to b\bar{b}$ candidate with the `GN2X` tagger [200], which utilises a similar network architecture to `GN2`.

b-tagged jets (α, β, γ) and the considered light jets (j):

$$\min_{j}\left[\max_{\alpha,\beta,\gamma}\left(m_{\alpha\beta}, m_{\gamma j}\right)\right].$$

Individual Higgs boson kinematics are informative and used as inputs to the NN architecture, and Higgs candidates are formed as follows. When there is a merged jet present, it is chosen as one Higgs boson candidate, whilst the other two jets form the second Higgs boson candidate. Otherwise, the four jets $(j = 1, 2, 3, 4)$ are paired to minimise the maximal separation between the Higgs boson candidates' decay products (jets i, k, l, m). This is expressed as:

$$\min(\Delta R_{\max}(b, b)) = \min_{i,k,l,m\in j}(\max(\Delta R(i, k), \Delta R(l, m))),$$

and the two Higgs boson candidates are then numbered in descending mass.

7.6 Machine Learning

The preselection effectively removes events that are clearly incompatible with the signal kinematics. However, even after preselection, the remaining background remains substantial, as shown in Figure 7.4. Neural networks, as introduced in Section 5.2, exploit correlations among multiple input features to form high-dimensional, non-linear latent spaces that enhance signal–background separation. This is particularly beneficial when multiple features, in isolation, appear to have insignificant discriminating power but show strong differences in correlation between the signal and background processes. Moreover, the network deployed typically benefits from increased training statistics, which are preserved by avoiding overly tight selections. For this reason, a neural network is trained on events that pass the preselection outlined in Table 7.3.

7.6.1 Model Architecture

A feed-forward MLP NN architecture is used. The MLP has 8 fully connected hidden layers, with 1024 nodes in the first layer and the number of nodes halving at each subsequent layer. Training is performed in batches, with a batch size of

4096. Rectified linear unit (ReLU) activation [201, 202], batch normalisation [203], an Adam optimiser with weight decay, and dropout regularisation [204] (with a dropout rate of 0.2) are used between layers. An initial learning rate of 5×10^{-6} is used, with a learning rate scheduler that halves the learning rate every five epochs. The network is trained for 25 epochs; the model with the lowest validation loss is used in the analysis.

For the signal training set, since the kinematics do not vary significantly across the plane (see Figure 7.4), simulated MC events for all signal points are combined[3]. For the background training set, all MC simulated events passing the preselection are used, as they provided a good match to the expected weighted background composition. To avoid overtraining and benefit from the full set of available events, orthogonal training and testing/validation datasets are used through a k-fold splitting procedure with $k = 5$ [205]. k-folding randomly splits the dataset into five slices, and trains five NNs. For each network, four subsets are used for training, and the remaining subset is reserved for validation and inference. Each event is therefore evaluated only by a network for that was not used during training, yielding out-of-fold predictions for the complete dataset while retaining the full statistical power of the dataset.

7.6.2 Input Features

The kinematic input features used in the NN are listed and defined in Table 7.4. These were selected from a wider set as the kinematic variables that provided the best signal-background discrimination and improved the NN performance, either alone or in combination with other variables. The impact of each variable is quantified using Shapley values, as defined in Section 5.2.6. This is complemented by an ablation study, in which the network is retrained after removing individual variables. Features associated with large Shapley values and a degradation in performance when omitted are retained. The complex combined features listed in the lower section of the table are generated using the Python `featuretools` [206] package, which pairs variables

[3]This choice also significantly simplifies the statistical interpretation later, since a single signal region definition is used for all signal models.

from the existing list using basic arithmetic operations. Such variables that improved performance are retained for training. Whilst these variables are correlated with the existing input features, they can help the NN to converge more quickly and to resolve more subtle differences between the signal and background distributions.

Table 7.4: Neural network input variables and their descriptions. The second part of the table lists combinations of variables that proved powerful.

Variable Name	Description
$p_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$m_{\mathrm{T,min}}^{b}$	Minimal transverse mass of $p_{\mathrm{T}}^{\mathrm{miss}}$ and one b-jet
$S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})$	Object based $p_{\mathrm{T}}^{\mathrm{miss}}$ significance
p_{T}^{4b}	p_{T} of the di-Higgs system
m_{T}^{4b}	Transverse mass of di-Higgs system and $p_{\mathrm{T}}^{\mathrm{miss}}$
n_{jets}	Number of jets
$n_{b-\mathrm{jets}}$	Number of b-jets
H_{T}	Scalar sum of p_{T} of signal jets
p_{T}^{b1}	Transverse momentum of leading $b-$jet
p_{T}^{b2}	Transverse momentum of sub-leading $b-$jet
p_{T}^{j4}	Transverse momentum of fourth-leading jet
m_{h1}	Mass of the leading Higgs candidate
m_{h2}	Mass of the sub-leading Higgs candidate
$\Delta R(b_{h1}^{1}, b_{h1}^{2})$	ΔR of the two jets associated with the leading Higgs
$\Delta R(b_{h2}^{1}, b_{h2}^{2})$	ΔR of the two jets associated with the sub-leading Higgs
$\Delta R_{\mathrm{min}}(b,b)$	Minimal ΔR of b-jets
`xwt`	Variable for top-rejection
$\Delta\eta_{\mathrm{max}}(b,b)$	Maximal $\Delta\eta$ between b-jets
Combined Variables	**Description**
$p_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{H_{\mathrm{T}}}$	Estimate of $S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})$
$p_{\mathrm{T}}^{\mathrm{miss}}\ /p_{T}^{4b}$	$p_{\mathrm{T}}^{\mathrm{miss}}$ divided by p_{T}^{4b}
$n_{\mathrm{jets}}/\ n_{b-\mathrm{jets}}$	Ratio of jets and b-jet counts
$n_{\mathrm{jets}} - n_{b-\mathrm{jets}}$	Number of untagged jets
$p_{\mathrm{T}}^{b1}/n_{\mathrm{jets}}$	Ratio of p_{T} of leading jet and number of jets
$S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})/n_{b-\mathrm{jets}}$	Ratio of $p_{\mathrm{T}}^{\mathrm{miss}}$ significance and number of b-jets
$S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}}) \times p_{\mathrm{T}}^{4b}$	$p_{\mathrm{T}}^{\mathrm{miss}}$ significance times p_{T} of di-Higgs system
$\Delta R_{\mathrm{min}}(b,b)/E_{\mathrm{T}}^{\mathrm{miss}}$	Minimal ΔR of b-jets divided by $p_{\mathrm{T}}^{\mathrm{miss}}$
$\Delta\eta_{\mathrm{RMS}}(b,b) \times \Delta\eta_{\mathrm{max}}(b,b)$	RMS of $\Delta\eta$ of b-jet pairs times maximal $\Delta\eta$ between b-jets
$\Delta R_{\mathrm{min}}(b,b) \times p_{\mathrm{T}}^{b2}$	Minimal ΔR of b-jets times p_{T} of sub-leading b-jet

Figure 7.6 shows the Shapley beeswarm plot for the classifier used in the analysis. Here, $p_{\mathrm{T}}^{\mathrm{miss}}$-based kinematic features, $m_{\mathrm{T,min}}^{b}$ and $S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})$ are highly ranked. The

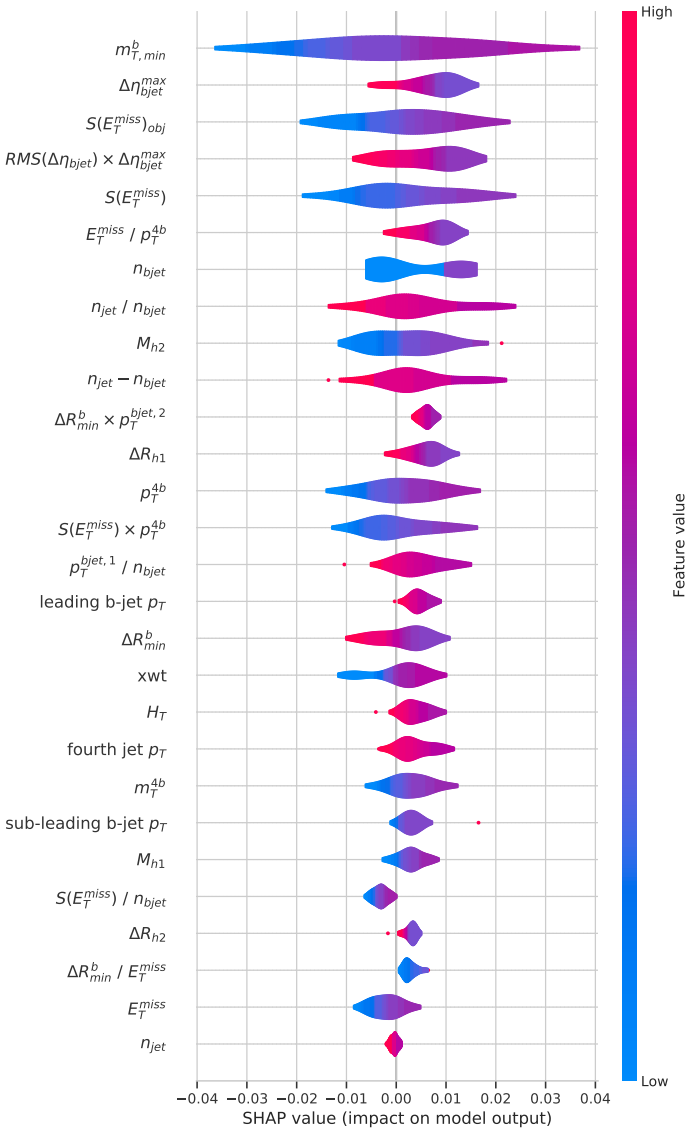


Figure 7.6: Shapley beeswarm for the classifier used in the analysis, with $p_{\mathrm{T}}^{\mathrm{miss}}$ based features and b-jet kinematics being highly ranked. Figure from Ref [2].

ratio $p_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{4b}$ is also assigned high importance. This variable probes the transverse-momentum balance between the invisibly decaying Z' system and the visible dark Higgs candidate, for which the signal distribution is expected to peak near unity. In contrast, the two systems are less correlated for $t\bar{t}$ backgrounds and do not peak near unity. Moreover, since the decay of a relatively high-p_{T} HH system produces b-jets spatially close within the detector and with $t\bar{t}$ events generally producing more separated b-jets, kinematics relating to average b-jet separations, like $\Delta\eta_{\mathrm{max}}(b,b)$ and $\Delta\eta_{\mathrm{RMS}}(b,b) \times \Delta\eta_{\mathrm{max}}(b,b)$, are effective discriminants between signal and background.

7.6.3 Decorrelation

As discussed, the physical interpretability of the di-Higgs mass as a proxy for the dark Higgs mass is an essential kinematic feature of the analysis. Any localised excess above the SM in m_{HH} would be clear evidence of a new particle with a well-defined mass. Given that the neural network output score is a non-linear combination of input features, despite not directly feeding m_{HH} as an input, the classifier can indirectly learn information correlated with m_{HH}. Thus, a cut on the output score will sculpt the background distribution, both reducing the physical interpretability of the signal region whilst diminishing sensitivity to m_{HH} shape within the fit.

To prevent this, a term penalising correlations between the neural network output score and m_{HH} is included in the loss function alongside the binary cross-entropy,

$$\mathcal{L} = \alpha\mathcal{L}_{\text{BCE}}(\hat{y}) + (1-\alpha)\mathcal{L}_{\text{decorr.}}(\hat{y}, m_{HH}) \tag{7.3}$$

where $\mathcal{L}_{\text{BCE}}$ is the cross-entropy loss, $\mathcal{L}_{\text{decorr.}}$ is a term penalising correlations between the NN output score, $\hat{y}$, and the value of m_{HH} for the event, and α controls the relative weight of the decorrelation.

The choice of decorrelation term involves defining a measure for correlation between the two observables, $\hat{y}$ and m_{HH}, and this analysis uses distance correlation [207, 208], with a weight of $\alpha = 0.7$. Distance correlation has the highly desirable feature that it is zero if and only if the two distributions are independent, capturing not only linear, but also non-linear correlations that a Pearson correlation measure would miss. The metric is defined as

$$d^2(\hat{y}, m_{HH}) = \frac{\text{dCov}^2(\hat{y}, m_{HH})}{\sqrt{\text{dCov}^2(\hat{y}, \hat{y})\text{dCov}^2(m_{HH}, m_{HH})}} \tag{7.4}$$

where $\text{dCov}(\hat{y}, m_{HH})$ is the distance covariance between $\hat{y}$ and m_{HH}, and is computed from double-centred pairwise distance matrices. For events i and j in a mini-batch containing n events,

$$a^{ij} = |\hat{y}^i - \hat{y}^j|, \qquad b^{ij} = |m^i_{HH} - m^j_{HH}|. \tag{7.5}$$

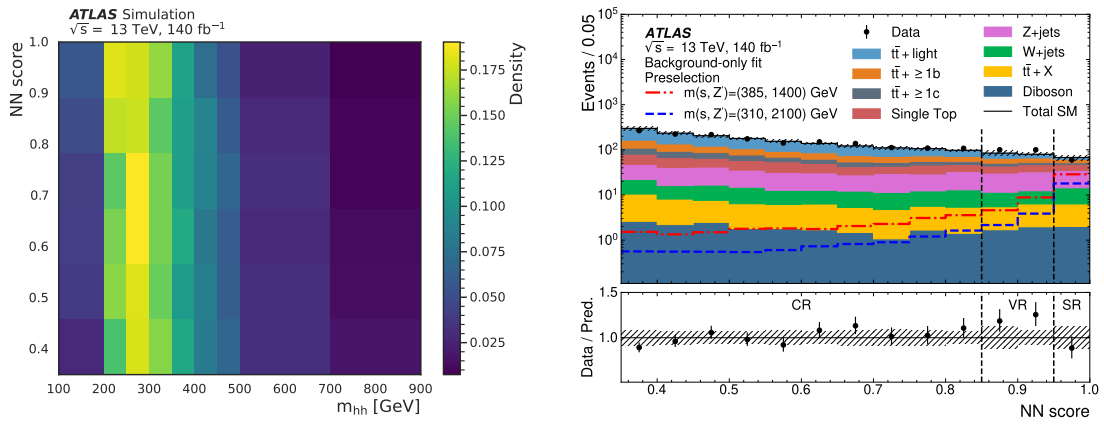
The corresponding double-centred matrices are

$$A^{ij} = a^{ij} - \frac{1}{n}\sum_{k=1}^{n} a^{ik} - \frac{1}{n}\sum_{k=1}^{n} a^{kj} + \frac{1}{n^2}\sum_{k,l=1}^{n} a^{kl}, \tag{7.6}$$

and analogously for B^{ij}. The distance covariance is then

$$\mathrm{dCov}_n^2(\hat{y}, m_{HH}) = \frac{1}{n^2}\sum_{i,j=1}^{n} A^{ij}B^{ij}. \tag{7.7}$$

The network then updates the weights according to the total loss. The successful application of this decorrelation technique is shown in Figure 7.7(a). The normalised SM background m_{HH} distribution is shown as a function of the NN output score, and the shape, shown by the heatmap, remains stable across the range of NN scores considered in the analysis. Despite the more sophisticated loss function, the classification performance remains strong for high NN scores, and is showcased for two benchmark signals in Figure 7.7(b). The retention of separation power may be explained by decorrelation, which forces the network to rely on non-trivial kinematic correlations for classification, rather than overly relying on an arbitrary reconstruction of an m_{HH} value within the signal window. The plot is shown after performing the background-only fit, as will be discussed in detail later, where strong agreement is observed across the score distribution, validating that the features learnt from MC simulation transfer effectively to data.



(a) (b)

Figure 7.7: (a) Two-dimensional histogram of m_{HH} and the NN score distributions for the total pre-fit SM background prediction, after the preselection requirements. This illustrates the minimal correlation between the variables resulting from the choice of NN loss function. (b) Neural network output score across the top control regions, the `VR` and `SR` after the background-only fit, as will be detailed later. The hashed bands show the total systematic and MC statistical uncertainty. Benchmark signal distributions are overlaid as dashed lines. The lower panel shows the ratio of data and predicted background yields.

7.7 Background Estimation

To reliably test for physics beyond the SM expectation, the modelling of MC-simulated backgrounds in the signal region is validated. A set of control regions targeting specific background processes, alongside a validation region, is defined to achieve this. Any mismodelling observed in the control regions will be accounted for using dedicated normalisation factors and a combination of nuisance parameter pulls relating to changes in modelling of the background processes in the phase space probed by the analysis. The success of this approach is checked in the validation region and, should sufficient agreement be observed, the analysis can have confidence in the background estimate when moving to the signal region. Control, validation and signal regions are defined by the event preselection of Table 7.3 with additional cuts of $[0.35, 0.85)$, $[0.85, 0.95)$ and $[0.95, 1.0]$ on the NN score, respectively. A lower bound of 0.35 is placed on the NN score to ensure that the derived normalisation factors and nuisance parameter pulls reflect a phase space similar to the signal region, whilst retaining sufficient statistics to avoid instabilities due to fluctuations.

The contamination of signal in the validation region is below 10% for all mass points not already excluded by the previous WW analysis [99]. Four control regions are defined, three of which lie within the $[0.35, 0.85)$ NN score range, targeting the top backgrounds, and a separate region targeting $Z+$jets.

7.7.1 Control Regions

The three categories of $t\bar{t}$ background ($t\bar{t}+\geq 1b$, $t\bar{t}+\geq 1c$, $t\bar{t}+$light) have distinct kinematic properties that enable the analysis to check the modelling of each separately. For sensitivity to $t\bar{t}+\geq 1b$, a cut of $n_{b-\text{jets}}=4$ is placed, defining `CR4b` — with the additional genuine b-jets in the $t\bar{t}+\geq 1b$ process dominating the region (45%). Within the $n_{b-\text{jets}}=3$ region, the fake-tag categories ($t\bar{t}+$light and $t\bar{t}+\geq 1c$) tend to peak at lower values of $m^b_{\text{T,min}}$. By contrast, $t\bar{t}+\geq 1b$ events, in which the additional b-jets do not originate from the $t\bar{t}$ system and are consequently less likely to align with $p_{\text{T}}^{\text{miss}}$, together with single top events, populate larger values of $m^b_{\text{T,min}}$. Thus, a cut of $m^b_{\text{T,min}}<200$ GeV is placed to define `CR3b-low`, populated by $t\bar{t}+$light (57%), and $m^b_{\text{T,min}}>200$ GeV to define `CR3b-high` — equally populated by single top, $Z+$jets and $t\bar{t}+\geq 1b$ (20-25%). The full selection criteria defining the $t\bar{t}$ related control regions are shown in Table 7.5, with a schematic in Figure 7.8 illustrating the set of top control regions, the validation region (`VR`) and the signal region (`SR`).

Table 7.5: Selections for all single-top and $t\bar{t}$ control regions, and the validation region.

Variable	Region requirement			
	`CR3b-low`	`CR3b-high`	`CR4b`	`VR`
Preselection	pass			pass
NN output score	$[0.35, 0.85)$			$[0.85, 0.95)$
$n_{b-\text{jets}}$	$=3$	$=3$	$=4$	–
$m^b_{\text{T,min}}$ [GeV]	$[80, 200)$	≥ 200	≥ 80	–

$Z+$jets within the signal region is dominated by $Z(\to \nu\nu)+$jets. However, since muons are considered invisible to the $p_{\text{T}}^{\text{miss}}$-trigger, $Z(\to \mu\mu)+$jets can be used as a proxy to define a control region. Two `Signal` muons are required in the

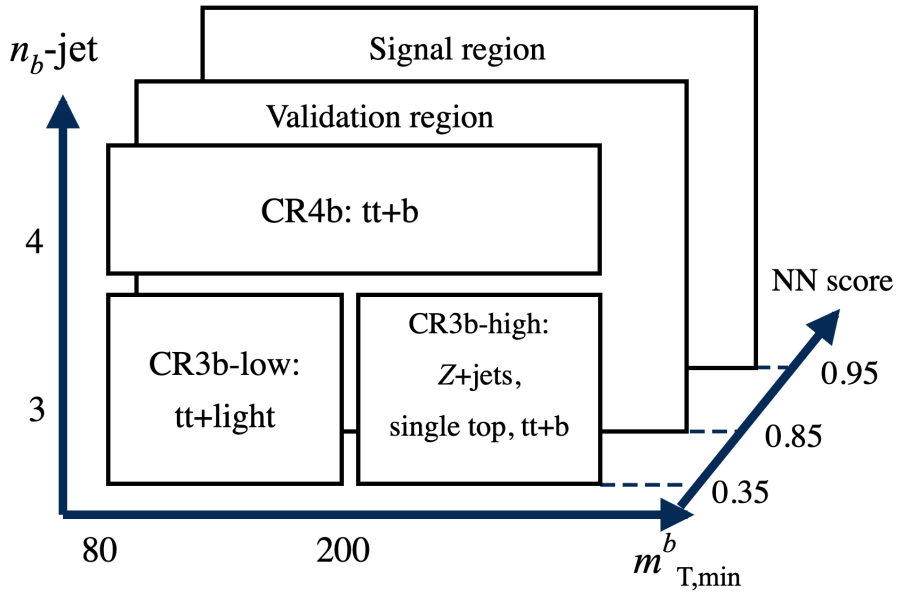


Figure 7.8: Schematic definition of the signal, validation and top control regions in $n_{b-\text{jets}}$, $m^b_{\text{T,min}}$, and NN output score as defined in Table 7.3, with the process targeted in each. `CRZ` follows the definition given in Table 7.6.

final state and assumed to be invisible to reconstruct $p_{\text{T}}^{\text{miss}}$ kinematics similar to $Z(\to \nu\nu)+\text{jets}$. To replicate the event preselection, an approximation for $S_{\text{obj}}(p_{\text{T}}^{\text{miss}})$ is defined. Mismeasurements of reconstructed objects can fake $p_{\text{T}}^{\text{miss}}$ and arise from random, uncorrelated contributions across multiple objects. These contributions sum in quadrature, meaning that fake $p_{\text{T}}^{\text{miss}}$ will scale approximately with $\sqrt{H_{\text{T}}}$. The $p_{\text{T}}^{\text{miss}}$-significance in this region is therefore defined as $p_{\text{T}}^{\text{miss}}/\sqrt{H_{\text{T}}}$. To further enhance purity, a window selection around the Z-mass of $|m_Z - m_{\mu\mu}| < 15$ GeV is applied. To preserve sufficient statistics within the fit, and because the constructed kinematics do not exactly reconstruct the input feature set, a selection is not applied to the NN output score. The full set of selections defining `CRZ` is shown in Table 7.6. There is strong shape agreement of the $Z+\text{jets}$ m_{HH} distributions between `CRZ`, `VR` and `SR`, giving confidence in this proxy approach.

The defined control regions enable the analysis to float three normalisation factors for $t\bar{t}+\geq 1b$, $t\bar{t}+\text{light}$ and $Z+\text{jets}$ processes, which are constrained by their relative purity in `CR4b`, `CR3b-low` and `CRZ`.

Table 7.6: Selection for the $Z+$jets control region. The muons are considered visible in the offline reconstruction, hence the comparatively low $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement.

Variable	`CRZ` requirement
n_μ	$= 2$
n_{jets}	$[4, 7]$
$n_{b-\mathrm{jets}}$	≥ 3
$\Delta\phi_{\mathrm{min}}^{4j}$	> 0.4
$p_{\mathrm{T}}^{\mathrm{miss}-\mu}$ [GeV]	> 200
$p_{\mathrm{T}}^{\mathrm{miss}-\mu}/\sqrt{H_{\mathrm{T}}}$ [$\sqrt{\mathrm{GeV}}$]	> 7
$m_{\mathrm{T,min}}^{b-\mu}$ [GeV]	> 80
$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	< 75
$\lvert m_{\mu\mu} - m_Z \rvert$ [GeV]	< 15

7.7.2 Signal Region

The decorrelated NN score enables the analysis to exploit the resonant signal peaks in a statistical fit to the m_{HH} distribution. An approximate signal m_{HH} resolution of 50 GeV defines the bin-width covering the range of dark Higgs signal masses ($200 < m_{HH} < 500$), with wider bins at the low-statistics tails. Two displays of events passing the `SR` requirements are shown, first earlier in Figure 7.5, and in Figure 7.9, where four b-tagged jets are shown back-to-back with the $p_{\mathrm{T}}^{\mathrm{miss}}$ vector.

The acceptance $\times$ efficiency of all signal models considered by the analysis after the `SR` requirements are shown in Figure 7.10, where the selection of at least three b-jets, alongside the tight cut on $p_{\mathrm{T}}^{\mathrm{miss}}$, are the two most significant reductions in signal yield. The behaviour with increasing m_s is explained by the increased separation of the Higgs bosons from a higher-mass dark Higgs, producing a cleaner resolved $4b$ system and a small increase in acceptance due to increased $p_{\mathrm{T}}^{\mathrm{miss}}$, which also increases the $4b$ pairing efficiency. The strong correlation of $m_{Z'}$ with $p_{\mathrm{T}}^{\mathrm{miss}}$ (see Figure 7.2) explains the trend of acceptance $\times$ efficiency towards higher-Z' mediator masses.

The pre-fit background m_{HH} distributions for all control regions are shown in Figure 7.11, where good agreement is observed in `CR3b-low` and underpredictions are seen for `CR4b`, `CR3b-high` and `CRZ`. Positive values of normalisations and nuisance parameter pulls for $t\bar{t} + \geq 1b$, single top and $Z+$jets are expected in the fit to correct the underprediction, which is similarly observed in the `VR`.

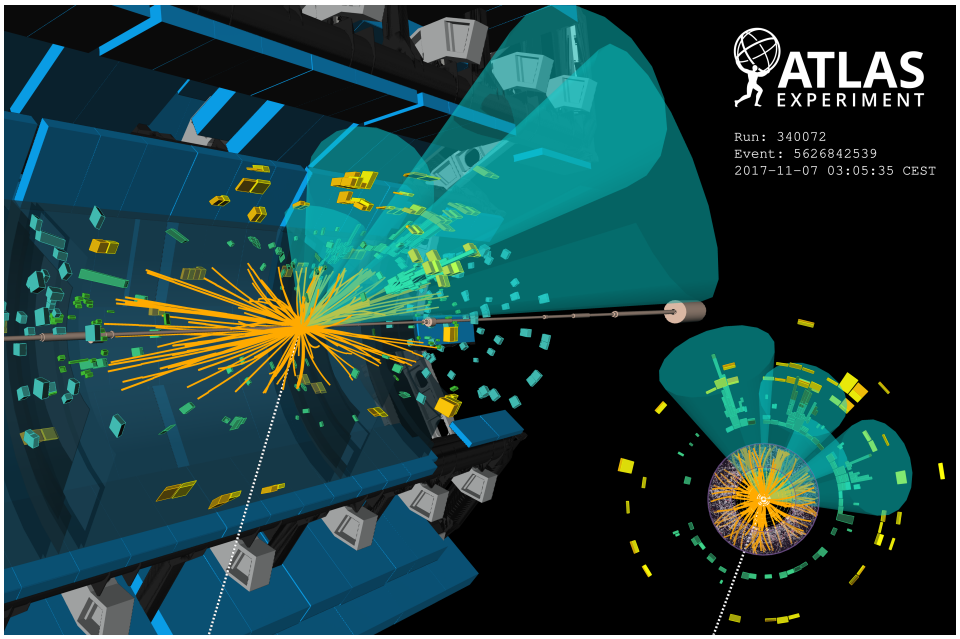


Figure 7.9: Display of a candidate event, projected onto the $x-y$ plane (lower right) and as a three-dimensional view of the ATLAS detector (left). The event contains four reconstructed jets, all of which satisfy the 77% `DL1r` b-tagging working point (shown in cyan). A missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) of 242 GeV is present, indicated by the white dashed line, and is nearly back-to-back with a four b-tagged jet system whose combined transverse momentum (p_{T}^{4b}) is 241 GeV.

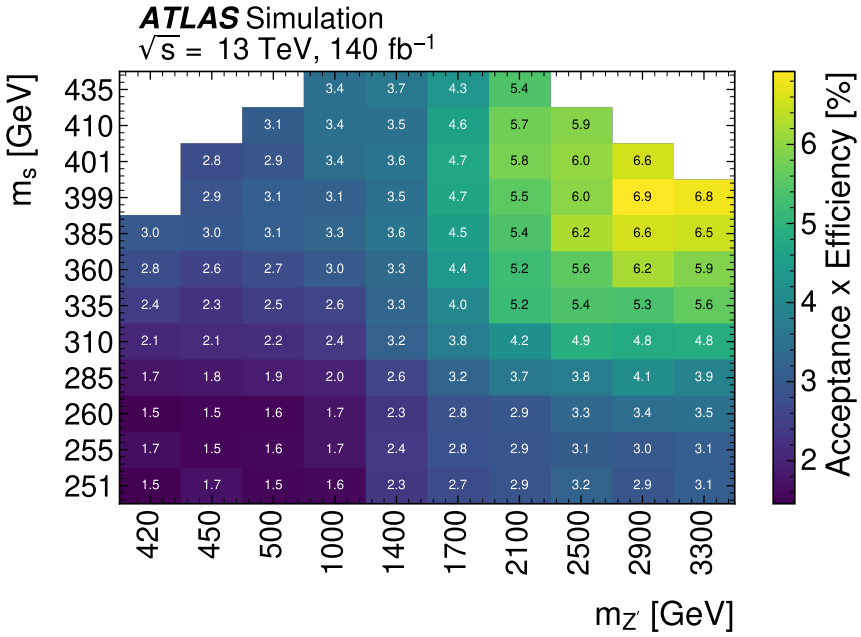
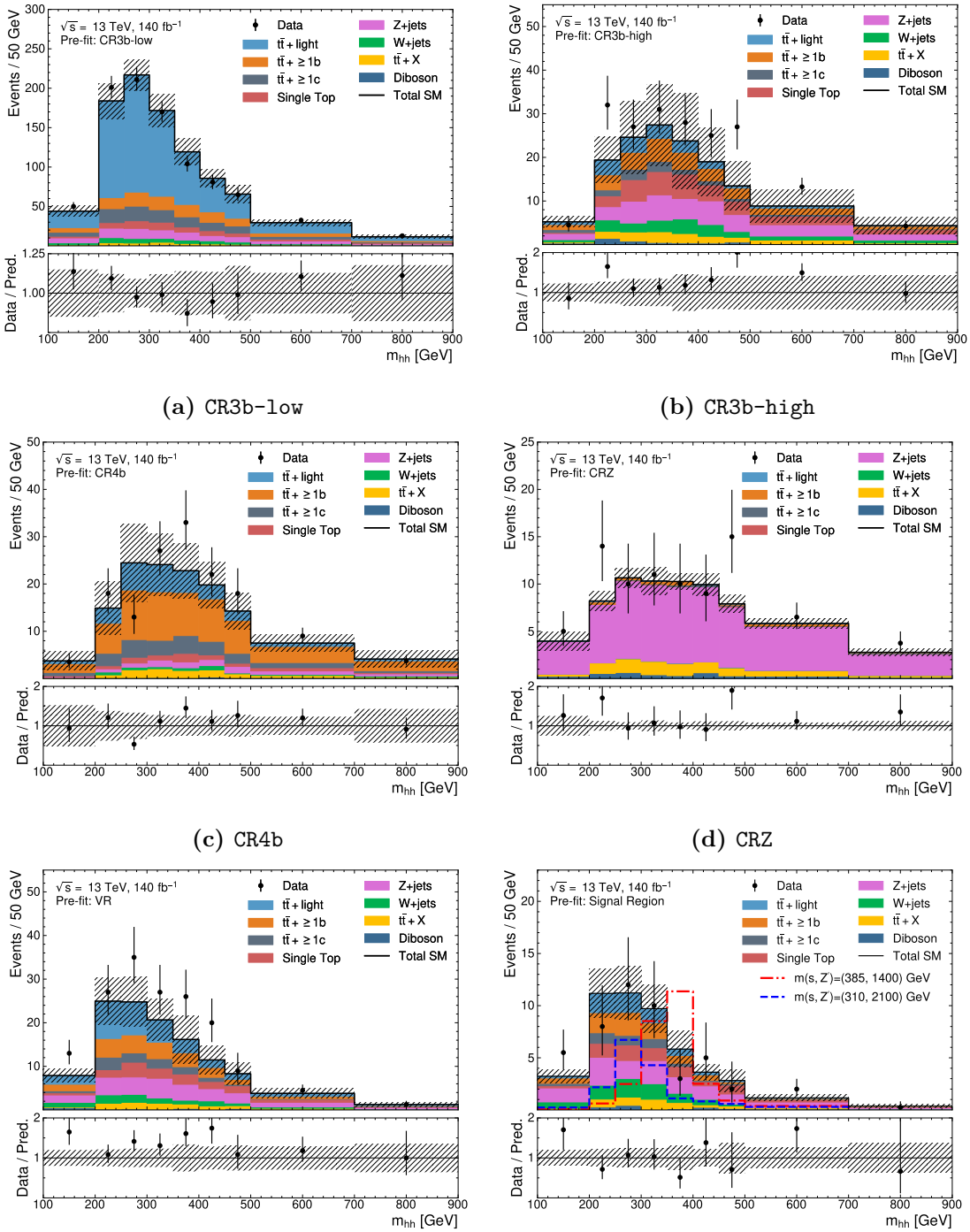


Figure 7.10: The overall dark Higgs signal acceptance × efficiency as a function of the Z' and dark Higgs boson masses. White areas indicate phase space not considered.



(a) CR3b-low (b) CR3b-high

(c) CR4b (d) CRZ

(e) VR (f) SR

Figure 7.11: The observed data and expected background m_{HH} distribution in CR3b-low (a), CR3b-high (b), CR4b (c), CRZ (d), VR (e) and SR (f). The hashed bands show the total systematic and MC statistical uncertainty. The lower panel shows the ratio of data and predicted background yields.

7.8 Systematic uncertainties

There are a number of sources of systematic uncertainty considered in the analysis, and can be broadly defined in two categories: experimental and theoretical.

7.8.1 Experimental Uncertainties

Experimental uncertainties arise from the reconstruction of physical objects in the detector, such as the flavour-tagging efficiency scale factor applied to each of the three or four b-jets in the final state. Often, as briefly mentioned in Section 6.2.5, these uncertainties are delivered to analyses as $\pm 1\sigma$ uncertainty eigenvariations of a large space of kinematic bins considered in the calibration. For flavour tagging, this amounts to 17 nuisance parameters for b-tagging efficiency and charm/light mistag efficiency scale factors. Variations on the jet energy scale and resolution total 31 and 13 nuisance parameters, respectively. The uncertainty in the `JVT` efficiency is estimated following the technique in Ref [132]. The uncertainties associated with lepton reconstruction efficiency are negligible and are not included in the fit.

Three nuisance parameters are included for the determination of the $p_{\mathrm{T}}^{\mathrm{miss}}$ soft term, one for the variation of the scale, and two for the resolution. The variations of the objects described above are also propagated to the missing transverse momentum calculation. The uncertainty derived from the mean number of interactions per bunch-crossing is applied with $\pm 1\sigma$ variations on the pileup weight. The impact of luminosity is included as a normalisation factor and allowed to float within an uncertainty of 0.83%.

7.8.2 Theoretical Uncertainties

A variety of modelling uncertainties related to choices made in the MC simulation pipeline are included for the SM backgrounds and the signal. The analysis targets an extreme region of phase space, where predictions rely on substantial extrapolation beyond the kinematic regimes most strongly constrained by data and generator tuning. Differences in the treatment of the matrix element, parton shower, hadronisation and additional QCD radiation can therefore lead to sizeable variations

in the predicted event yields and kinematic distributions, and are accounted for through dedicated nuisance parameters.

For the $t\bar{t}$ and single top processes, a dedicated matrix element generator uncertainty is defined between the nominal (Powheg +Pythia 8) and varied (MadGraph5_aMC@NLO+Pythia 8) samples, and the variation accounting for the choice of parton shower is taken as the difference between the nominal (Powheg +Pythia 8) and varied (Powheg +Herwig 7) samples. Following previous work in a similar phase space [96], cross-section uncertainties in this obscure phase space of 50%(20% [209]) are attributed to $t\bar{t} + \geq 1c$ and single top ($t\bar{t} + X$ and diboson) backgrounds. No such cross-section uncertainty is applied to $t\bar{t} + \geq 1b$ and $t\bar{t}+$light since their normalisation is constrained by data. For $t\bar{t}$, single top and the signal, the uncertainty due to initial-state radiation is estimated with variations to the `Var3c` Pythia parameter for the `A14` tune, and the h_{damp} parameter controlling the kinematics of the leading emission in Powheg. There is an ambiguity in the definition of single top and $t\bar{t}$ at NLO, with both contributing to the same $W^+W^-b\bar{b}$ final state. A modelling uncertainty is defined as the difference between the two accepted approaches for resolving the ambiguity, between the nominal (diagram removal [210]) and varied (diagram subtraction [211]), which is applied to the single top background. Uncertainties related to missing higher-order corrections are assessed by variations of the choice of renormalisation and factorisation scales for the $W+$jets and $Z+$jets backgrounds and the dark Higgs signal models, evaluated as the envelope of a set of seven variations [212].

7.8.3 Systematic Smoothing: single top

Single top is a sub-leading background process in the signal region, and the calculation of two-point systematics is therefore statistically limited. Statistically limited systematics can introduce noise into the fit procedure, potentially generating spurious fitting behaviour. To increase the available statistics, the selection on NN score is loosened to provide sufficient statistics in the nominal and varied samples. After loosening the selection to cover both the validation and signal

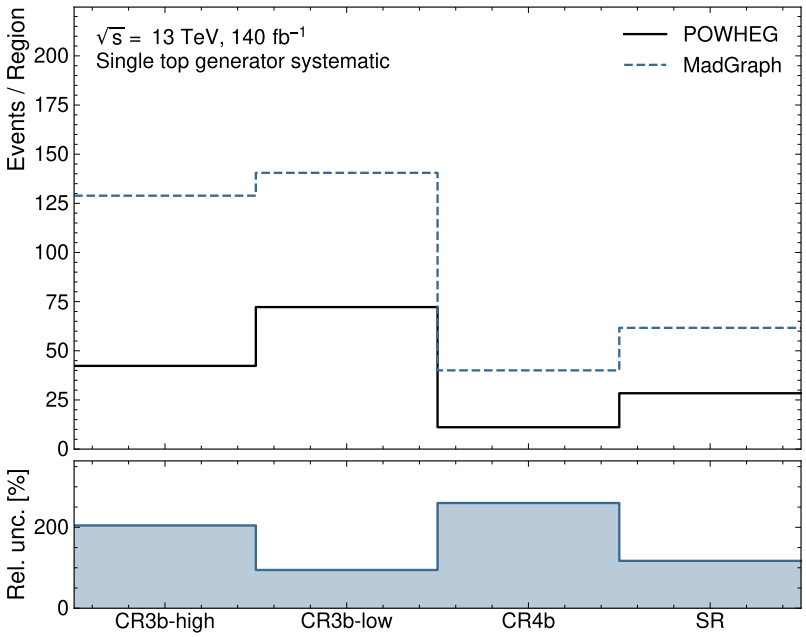


Figure 7.12: Shown are the predicted yields with the nominal (POWHEG) and alternative (MADGRAPH5_aMC@NLO) generator approaches for the single top process in `CR3b-high`, `CR3b-low`, `CR4b` and the `SR`, with the score cut relaxed to include the `VR`, as introduced in the main body. The lower panel shows the relative percentage uncertainty.

regions, the ratio and shape between the total varied and nominal samples are consistent — showing that the overall impact of the systematic is retained between selections. Gaussian kernel smoothing for both the nominal and varied histograms is implemented. Here, a desirable property is the interpretation of the bandwidth, and consequently a 50 GeV width is chosen[4]. The fit utilises the generator systematic uncertainty for single top to improve the modelling of this background process across the analysis regions, and the relative size is shown for the relevant control and signal regions in Figure 7.12.

7.9 Statistical Model

A statistical model is built following the general principle introduced in Section 5.3 and implemented within the `HistFitter` [151] software package. The binned

[4]The size of the systematic uncertainty is generally robust to both the chosen smoothing algorithm and chosen bandwidth.

likelihood function for the analysis is written as,

$$L(\mu, \boldsymbol{\theta}) = \prod_{i \in \mathrm{CR}} \mathrm{Pois}\Big(n_i \mid \lambda_i(\mu, \boldsymbol{\theta})\Big) \prod_{j \in \mathrm{SR}} \mathrm{Pois}\Big(n_j \mid \lambda_j(\mu, \boldsymbol{\theta})\Big) \prod_k G\Big(\theta_k \mid \hat{\theta}_k, \sigma_k\Big), \quad (7.8)$$

where $n_{i(j)}$ are observed yields in control (signal) regions and $\boldsymbol{\theta}$ represents nuisance parameters that encode systematic uncertainties. A Gaussian prior $G(\theta_k \mid \hat{\theta}_k, \sigma_k)$ constrains each systematic to its nominal value $\hat{\theta}_k$ with uncertainty σ_k, defined with zero mean and unity variance, with the impact of the systematic on yield absorbed into the Poisson terms with $\lambda(\mu, \boldsymbol{\theta})$.

The predicted yields in each bin $\lambda_i(\mu, \boldsymbol{\theta})$ and $\lambda_j(\mu, \boldsymbol{\theta})$ are modelled as:

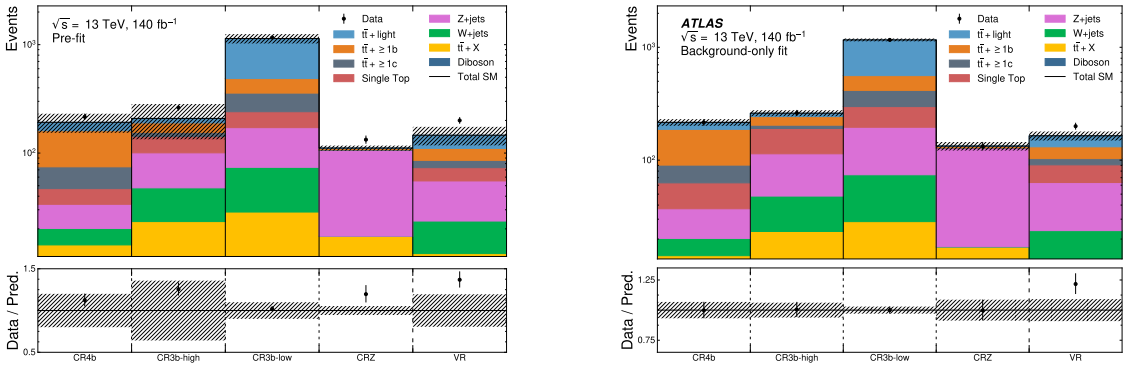
$$\lambda_i(\mu, \boldsymbol{\theta}) = \sum_{\mathrm{b}} \mu_{\mathrm{b}} \, \lambda_{\mathrm{b}}^{\mathrm{MC}}\Big(\boldsymbol{\theta}\Big) + \mu_{\mathrm{sig}} \lambda_{\mathrm{sig}}^{\mathrm{MC}}\Big(\boldsymbol{\theta}\Big), \quad (7.9)$$

where μ_{b} are background normalisation factors (if allowed to float), and $\lambda^{\mathrm{MC}}(\boldsymbol{\theta})$ includes all shape and normalisation variations for the relevant nuisance parameters. In general, systematics are calculated as $\pm 1\sigma$ variations on the nominal sample, or by taking the difference between the nominal sample and an alternative modelling approach. Two histograms, one for each of the $\pm 1\sigma$ variations, for this particular systematic are defined for both the CR and SR. However, to ensure the fit has sufficient flexibility to vary nuisance parameters to accommodate the data, the impact of the systematic, the *response*, must be defined as a continuous function of σ. There are multiple methods of interpolating between $\pm 1\sigma$; for this analysis a 6th order polynomial[5] is used, which is the default within `HistFitter` and is extrapolated linearly beyond this.

7.10 Fit Validation

A profile-likelihood fit is performed across all analysis regions, considering the signal normalisation across all regions together with the modelling of SM background processes via nuisance parameters and normalisation factors. The systematic

[5]The coefficients of the polynomial are fixed by requiring that the derived response function passes through the three input templates, the first derivatives at the $\pm 1\sigma$ boundaries match the expected difference between the nominal and $\pm 1\sigma$ and a vanishing second derivative at the boundaries, leaving no unconstrained degrees of freedom.



(a) Pre-fit **(b)** Post-fit

Figure 7.13: Background-only pre-fit (a) and post-fit (b) distributions in all control and validation regions.

implementation cannot be assumed correct a priori, and a number of fit-validation stages are performed to ensure the behaviour is correct before performing hypothesis tests in the signal region and following the collaboration-wide blinding guidelines, as outlined in Section 5.3.

7.10.1 Background-only Fit

First, to verify the background prediction, control-region fits to data are performed, in which no signal model is assumed, and only background-process related nuisance parameters are used, with the modelling checked in the `VR`.

From the pre-fit control region distributions in Figure 7.13(a), there is an underprediction of MC in `CR4b` and `CR3b-high` — predominantly composed of $t\bar{t}$ and single top processes — and in the `CRZ`. Thus, it is expected that normalisations and systematics pertaining to these processes will be utilised and profiled to correct this underprediction. Pulls of systematics and normalisation factors that result in an increase of MC yields post-fit are observed, as showcased in Figure 7.14, with strong post-fit control region agreement shown in Figure 7.13(b).

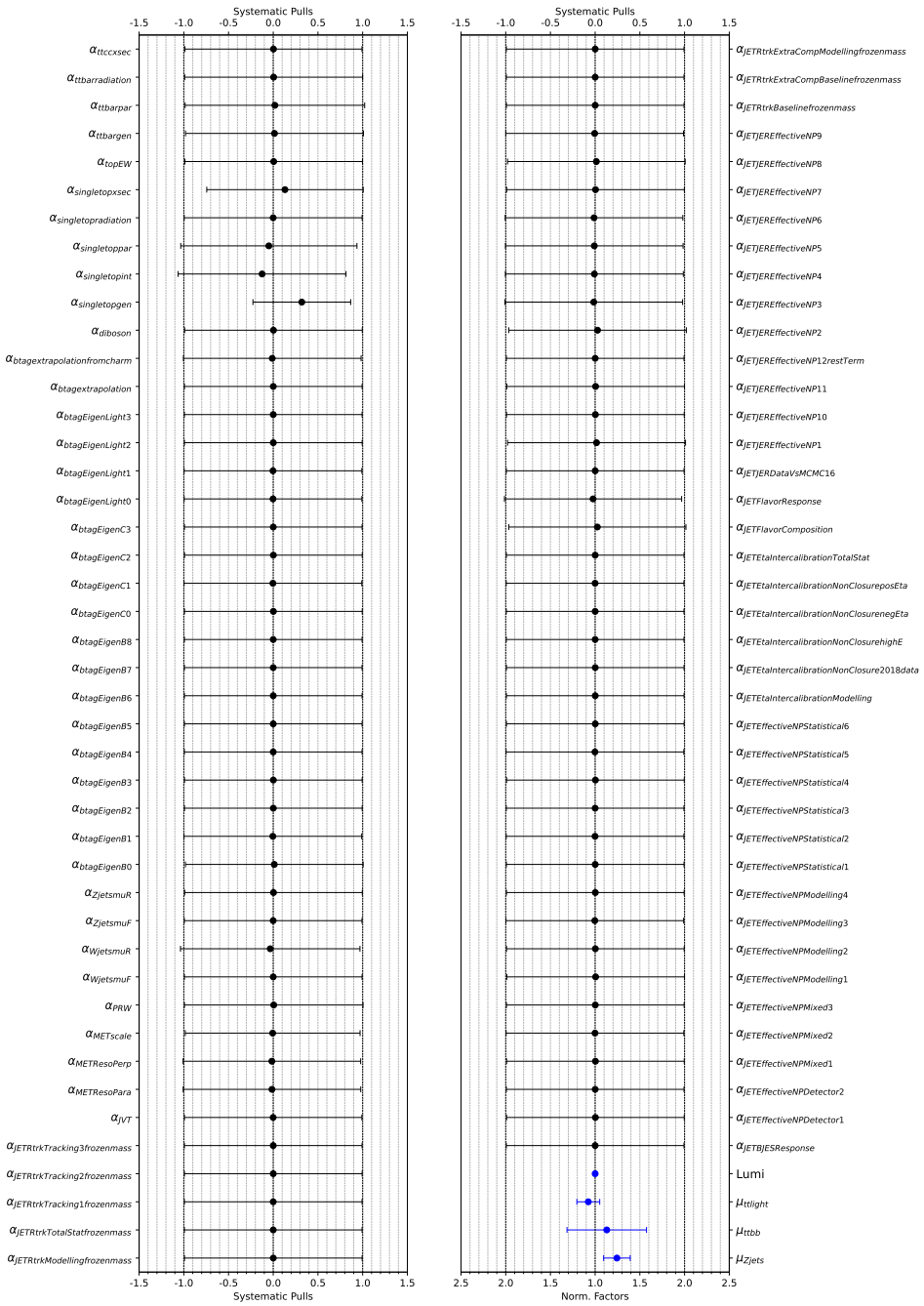


Figure 7.14: Best-fit values of the nuisance parameters in the background-only fit.

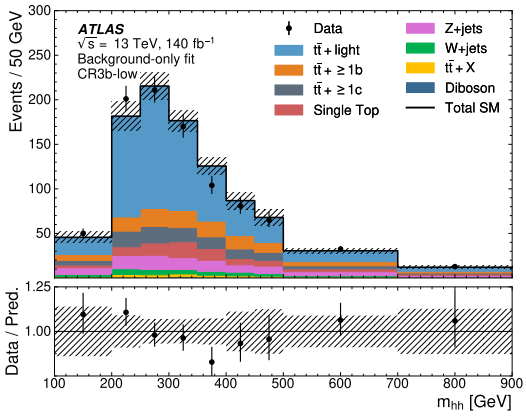
The purity of the Z+jets region leads to a well-constrained normalisation factor ($\mu_{Z+\text{jets}} = 1.24 \pm 0.15$), consistent with the data/MC differences in `CRZ`. Pre-fit agreement in the `CR3b-low` region is good, with the compatibility of the $t\bar{t}$ + light normalisation ($\mu_{t\bar{t}+\text{light}} = 0.92 \pm 0.13$) with unity reflecting this. An underprediction is observed in `CR4b` which is corrected for with the $t\bar{t}$ + $\geq 1b$

normalisation ($\mu_{t\bar{t}\ +\geq 1b} = 1.13 \pm 0.44$). The most critical systematics are the theoretical uncertainties on $Z+$jets and $t\bar{t}$, given that these are the two primary processes in the SR. However, the pulls and constraints on these systematics are negligible, since the impact of these processes on the normalisation is corrected by the normalisation factors. CR3b-high is underestimated by MC, supporting the increase in rate for $Z+$jets and $t\bar{t}+\geq 1b$ derived in CRZ and CR4b, and the residual differences are consistent with a marginal increase in rate for single top production using a combination of alternative modelling choices for the process. Given the large pre-fit size of the single top generator uncertainty in the CRs, as shown in Figure 7.12, some constraint of this parameter is expected[6]. The CR3b-high is most sensitive, where the data favour the increased yield and the modelling of an admixture with the alternative generator.

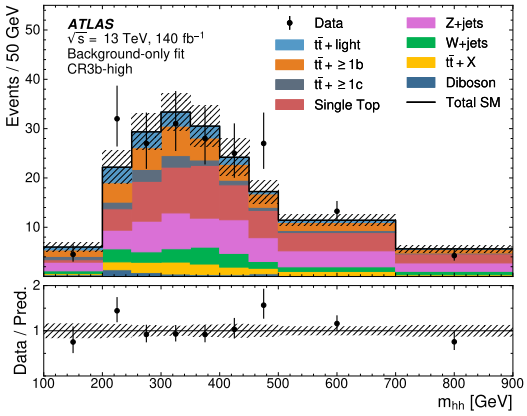
The post-fit background predictions, along with the observed data, in all analysis regions, binned in m_{HH}, are shown in Figure 7.15, where, although not guaranteed beforehand, the normalisation factors and nuisance parameter pulls derived in the single-bin control regions translate into improvements in shape. An improved shape is observed particularly for CR3b-high, in which the pull towards the alternative generator single top sample tends towards better agreement in m_{HH}[7]. A small upwards fluctuation ($< 1.7\sigma$) is observed in the VR. The agreement between data and MC is stable across the NN score in the VR, with slight fluctuations within the [0.85,0.875] and [0.925,0.95] bins, and no clear mismodelling trends between data and MC in m_{HH} — providing confidence in the background estimation approach.

[6]The size of the constraint was verified analytically with the Fisher information formalism of Section 5.3.3 using the full analysis likelihood function of Equation 7.8, also verifying that CR3b-high is where the sensitivity to single top modelling resides, with 59% of the total information for the process across all analysis regions.

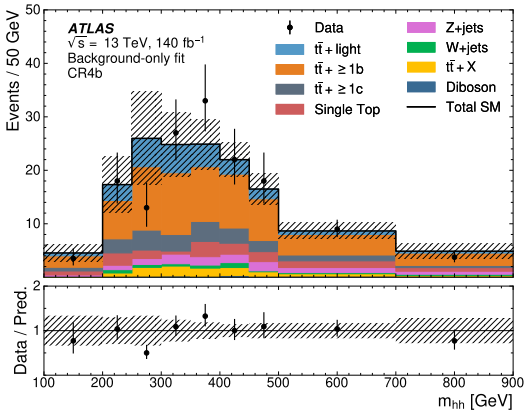
[7]Additional fit configurations were tested using a binned CR3b-high, where a slightly larger pull is observed on the single top generator systematic, but ultimately the change in agreement within the VR was negligible.



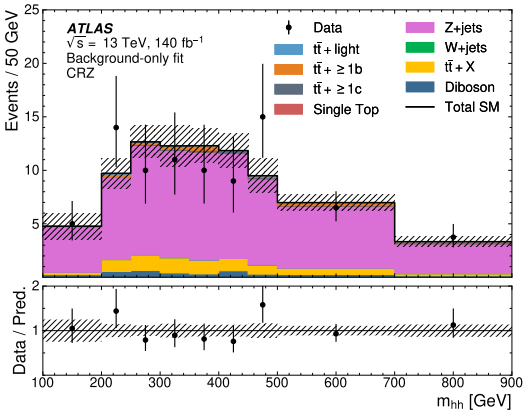
(a) CR3b-low



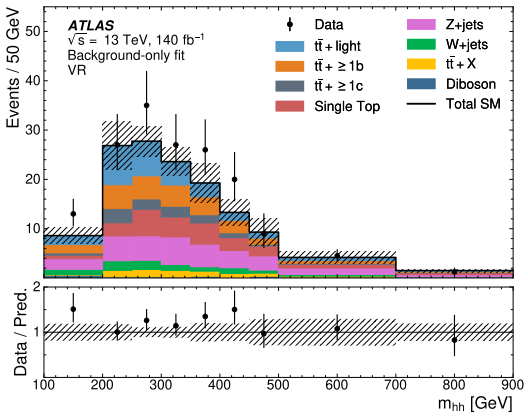
(b) CR3b-high



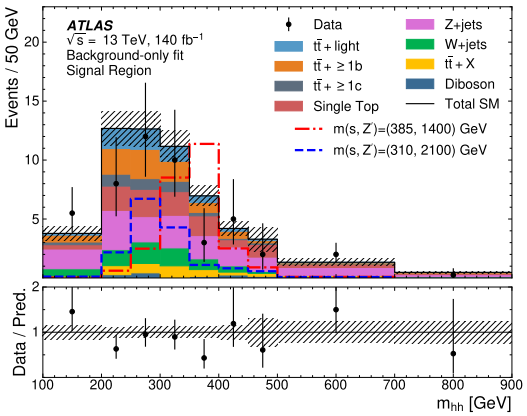
(c) CR4b



(d) CRZ



(e) VR



(f) SR

Figure 7.15: The observed data and expected background m_{HH} distribution in CR3b-low (a), CR3b-high (b), CR4b (c), CRZ (d), VR (e) and SR (f) after the background-only fit. The hashed bands show the total systematic and MC statistical uncertainty. The lower panel shows the ratio of data and predicted background yields.

Table 7.7: Fitted background yields and observed data after the background-only fit in each control and validation region.

Region	CR3b-low	CR3b-high	CR4b	CRZ	VR
Data	1163	262	216	133	200
Fitted background	1160 ± 34	261 ± 16	216 ± 15	133 ± 12	164 ± 15
$t\bar{t}+$ light	602 ± 77	17.7 ± 3.9	28.9 ± 6.9	0.56 ± 0.28	33.1 ± 6.9
$t\bar{t}+\geq 1b$	146 ± 64	39 ± 16	97 ± 31	3.2 ± 1.5	28 ± 13
$t\bar{t}+\geq 1c$	117 ± 65	13.4 ± 7.9	28 ± 17	1.6 ± 1.2	12.3 ± 7.3
single top	102 ± 37	76 ± 30	26 ± 12	$0.49^{+0.88}_{-0.49}$	28 ± 10
$Z+$jets	121 ± 16	66 ± 11	16.8 ± 3.2	110 ± 12	39.6 ± 6.5
$W+$jets	46 ± 22	25 ± 13	6.1 ± 3.4	–	11.9 ± 5.3
$t\bar{t}+X$	19.3 ± 4.5	16.6 ± 3.8	11.8 ± 2.8	11.7 ± 2.7	8.1 ± 2.1
diboson	9.2 ± 2.7	6.8 ± 1.7	2.4 ± 0.7	5.4 ± 1.4	3.7 ± 1.5

7.10.2 Exclusion Fit

Once the background estimate is validated and the behaviour of the background-only fit understood, the analysis moves to CR+SR fits. Here, to avoid biasing fit validation with respect to the observed data, the fit is performed using the Asimov dataset [157] — in which the observed dataset is replaced with the expected background prediction pre-fit. Such a fit provides an expectation of the analysis' sensitivity to the signal model, assuming the MC estimate equals the data. An Asimov fit also provides an understanding of the analysis' sensitivity to specific nuisance parameters — quantified via constraints. The Asimov exclusion fit parameters are shown in Figure 7.16 for the $m_{Z'} = 1700$ GeV and $m_s = 310$ GeV signal model, where the constraints on background-specific nuisance parameters are similar to those observed in the background-only fit, as expected given that the background processes are predominantly constrained in high-statistics control regions.

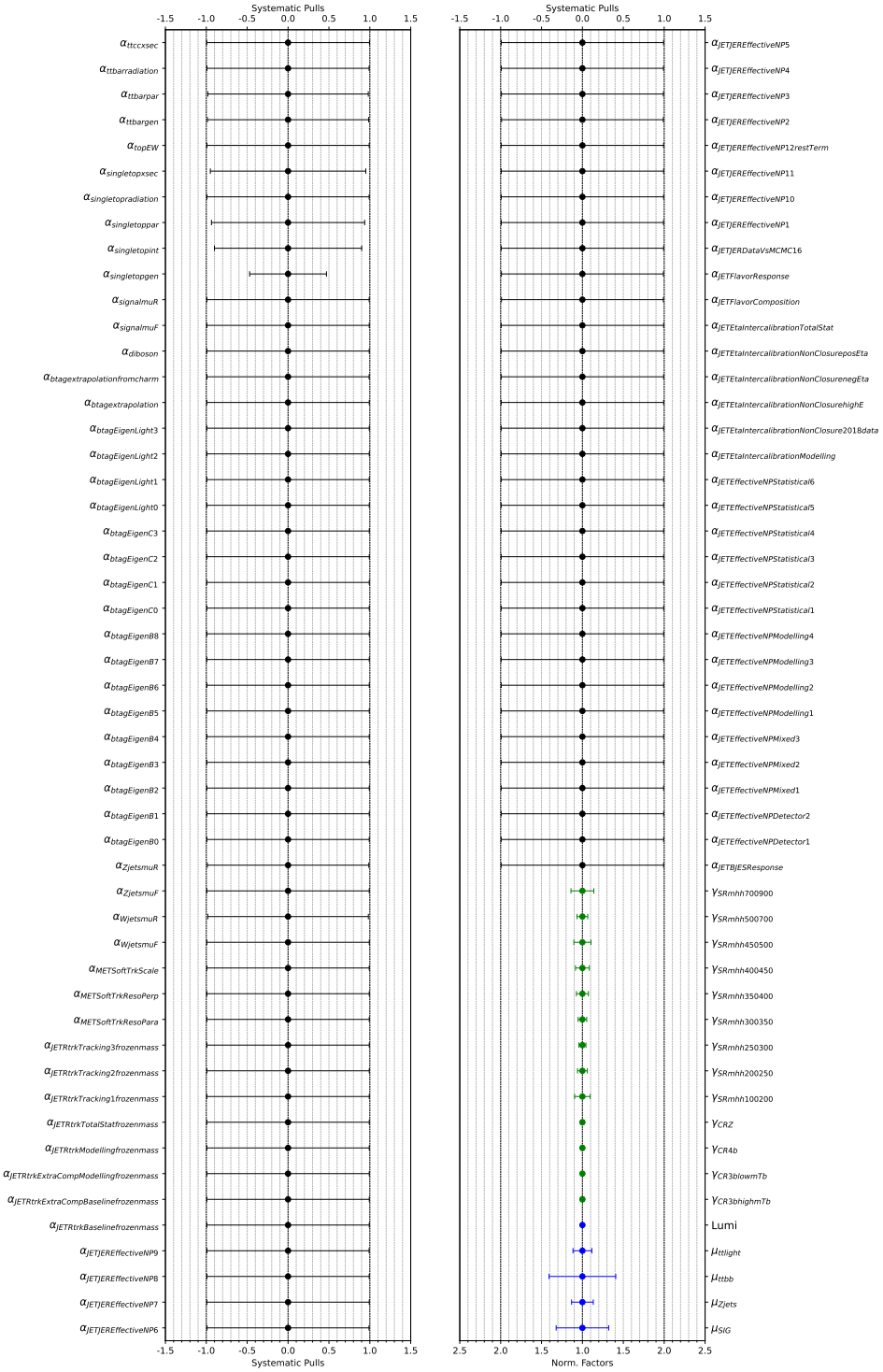


Figure 7.16: Fit parameters for a combined control and signal region fit to an Asimov dataset with the signal model $(m_s, m_{Z'}) = (310, 1700)$ GeV.

The ranking of nuisance parameters for $\pm 1\sigma$ deviations from an Asimov exclusion-fit is shown in Figure 7.17. All systematic uncertainties result in $< 5\%$ impacts on the best-fit signal normalisation, emphasising the flexibility of the fit through utilising dedicated control regions for the dominant background processes. More specifically, nuisance parameters for dominant backgrounds ($t\bar{t} + \geq 1b$, $t\bar{t} +$ light and $Z +$ jets) that carry an associated normalisation factor have relatively small impacts on the signal normalisation, with the largely mass-independent systematic uncertainties being easily corrected for with the dedicated normalisation factors. The nuisance parameter variations for sub-leading background processes are not compensated by floating normalisations. Consequently, variations that affect the signal and control regions differently, together with other nuisance parameters adjusted to accommodate deviations from the pre-fit expectation, can correlate with the fitted signal strength and thereby influence the inferred signal normalisation. Furthermore, a well-populated control-region contribution does not constrain the $W+$ jets background, its correlation with SR-sensitive parameters and, consequently, its impact on the fitted signal normalisation, is increased.

It is also instructive to investigate the response of the fit with respect to variations smaller than $\pm 1\sigma$. Rankings for deviations in increments of 0.1 from nominal to $+0.5\sigma$ are explored. The impacts on the signal normalisation up to 0.5σ for single top nuisance parameters are shown in Figure 7.18. The largest systematic size (single top generator) corresponds to the largest impact on the signal normalisation, as expected. Linear behaviour is observed for the leading nuisance parameters up to and beyond the largest pull observed (0.3) in the analysis.

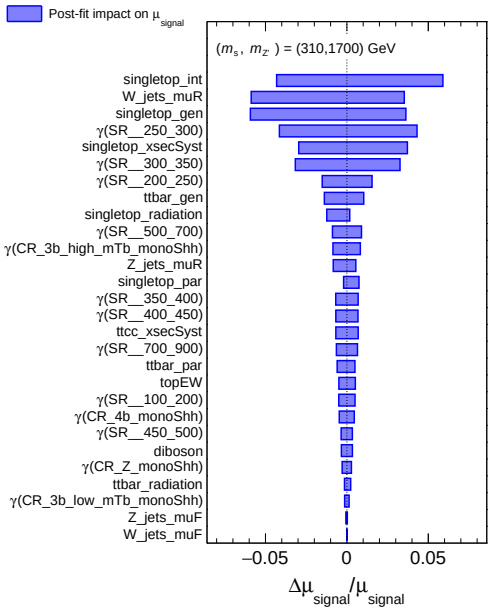


Figure 7.17: Fit parameter ranking, calculated by setting the parameter in question constant at the $\pm\sigma$ values and re-running the fit, for an Asimov exclusion-fit with the $(m_s, m_{Z'}) = (310, 1700)$ GeV signal point at $\mu = 1$. $\Delta\mu$ corresponds to the post-fit impact on the signal normalisation.

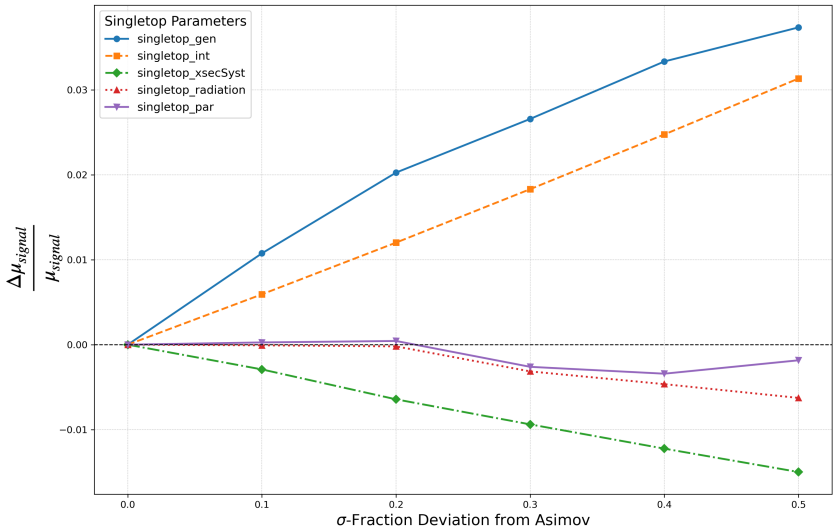


Figure 7.18: Impacts on the signal normalisation, $\frac{\Delta\mu_{signal}}{\mu_{signal}}$, shown for a range $[0, 0.5]$ in increments of 0.1σ for single top theory nuisance parameters for the $(m_s, m_{Z'}) = (310, 1700)$ GeV Asimov exclusion fit.

7.11 Unblinded Results

With the background estimate validated in a kinematically similar phase space to the signal region, and the behaviour of the fit understood with the signal region and specific signal models included, the analysis can reliably test for physics beyond the SM expectation using the observed data in the signal region. An unblinded signal region after the background-only fit is shown in Figure 7.19, with two illustrative signal models overlaid — highlighting the sensitivity the analysis achieves to such resonant signals through m_{HH}. No excess is observed above the SM expectation, with the data agreeing well with the background expectation.

7.11.1 Model-independent Fit and Results

Before injecting specific signal models into the unblinded fit, hypothesis tests are performed to assess the compatibility of the data with the background-only hypothesis. To achieve this, a set of single-bin discovery regions is defined, each spanning a wide range of m_{HH} and chosen to ensure a meaningful number of expected events. Given the optimised event selection and the use of advanced NN architectures targeting the dark Higgs process, these results are not completely model-independent but are as close as a modern particle physics analysis may construct. The fit is performed simultaneously across all control regions and the chosen discovery region, allowing the background nuisance parameters to adjust within their uncertainties. Agreement within 1σ is observed across the complete set of discovery regions, as shown in Figure 7.20.

Following the formalism introduced earlier in Section 5.3.4, the significance is evaluated using asymptotic formulae with the test statistic (t_0) and p_d-value. In the absence of any excess, an upper limit on the number of events above the SM expectation at the 95% confidence level is computed. These are converted into upper limits on the visible cross section for each discovery region and summarised in Table 7.8. Minor deficits are observed in `DRa-DRe`; consequently, the p_d-value is set to 0.5. The excesses of `DRf` and `DRg` are mild, reflected in the p_d-values of

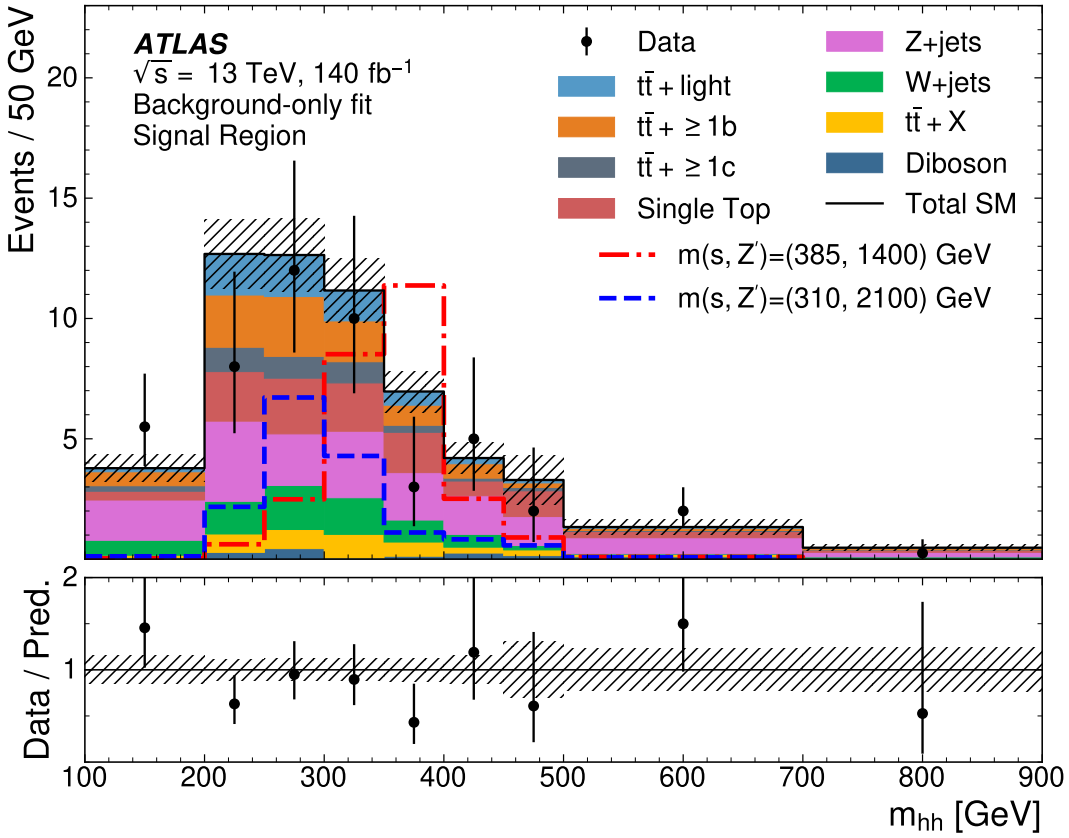


Figure 7.19: Post-fit m_{HH} distribution in the signal region after the unblinded background-only fit.

0.40 and 0.33, respectively. The analysis sets limits on signal models with visible cross sections ranging from 0.042 to 0.079 fb.

7.11.2 Model-dependent Results

Dark Higgs signals are now injected, and fits are performed separately for each signal mass point (m_s, $m_{Z'}$) across all control and signal regions. The best-fit value of μ_{sig} here reflects the presence, or lack thereof, of a particular signal given the observed data. Exclusion contours, showcasing the median expectation from an Asimov dataset and the observed exclusion contours from the unblinded data, are shown in Figure 7.21. The observed limit of the analysis agrees closely with that of an Asimov expectation, with the slightly more stringent limits explained via the slight deficits seen in Figure 7.19. Due to the cross sections decreasing steadily as a function

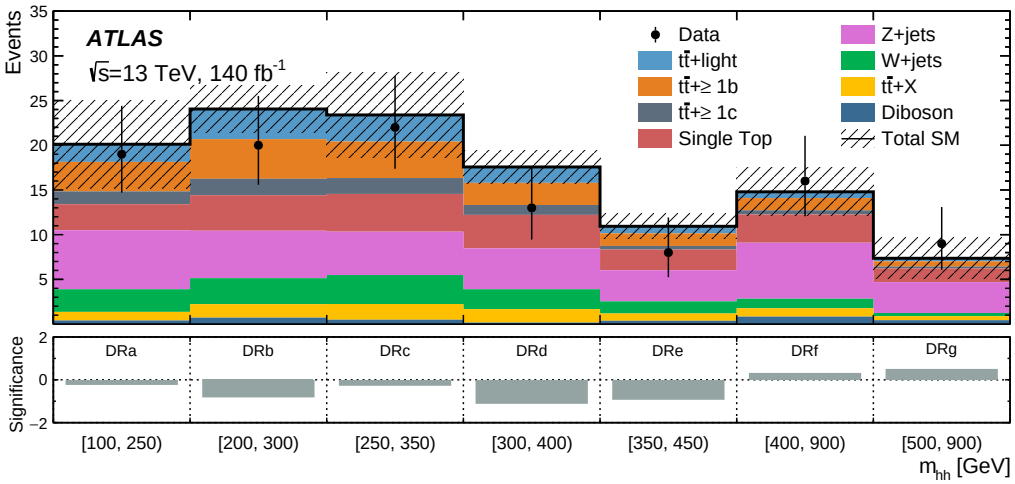


Figure 7.20: Observed data and fitted background yields in each discovery region (DR). The hashed band shows the total systematic and MC statistical uncertainty. In the lower panel, the significance (measured using Equation 1 of Ref [213]) of fluctuations between the data and the SM expectation is shown.

Table 7.8: Upper limits on the expected (observed) number of BSM events, S^{95}_{exp} (S^{95}_{obs}), at 95% confidence level, the $\pm 1\sigma$ variations on the expected upper limit, and the observed limit on the effective visible BSM cross section, $\langle \epsilon\sigma^{95}_{\text{obs}} \rangle$, for each of `DRa`–`DRg`. The discovery p_d-value is also shown and is capped at 0.5 when the data yield is smaller than the SM prediction.

Region	m_{HH}	Data	Fitted background	S^{95}_{exp}	S^{95}_{obs}	$\langle \epsilon\sigma^{95}_{\text{obs}} \rangle$ [fb]	p_d
`DRa`	$[100, 250)$	19	20.1 ± 4.9	$11.1^{+5.1}_{-3.3}$	10	0.072	0.50
`DRb`	$[200, 300)$	20	24.1 ± 2.7	$12.1^{+5.5}_{-3.6}$	8.8	0.063	0.50
`DRc`	$[250, 350)$	22	23.4 ± 4.7	$11.9^{+5.5}_{-3.6}$	11	0.076	0.50
`DRd`	$[300, 400)$	13	17.6 ± 1.9	$10.2^{+4.9}_{-3.1}$	6.9	0.049	0.50
`DRe`	$[350, 450)$	8	10.9 ± 1.4	$8.2^{+4.1}_{-2.5}$	5.9	0.042	0.50
`DRf`	$[400, 900)$	16	14.8 ± 2.7	$10.0^{+4.7}_{-3.0}$	11	0.079	0.40
`DRg`	$[500, 900)$	9	7.4 ± 2.3	$7.2^{+3.6}_{-2.3}$	8.7	0.062	0.33

of $m_{Z'}$ (see Figure 7.3), $m_{Z'} \approx 2$ TeV is the edge of the analysis' sensitivity. The lower edge in $m_{Z'}$ coincides with the opening of $Z' \to \chi\chi$ with sufficient branching fraction. The relevant decay thresholds similarly determine the accessible range in m_s. The decay $s \to HH$ becomes kinematically allowed for $m_s > 2m_H$, with phase space suppression close to threshold. For the benchmark value $m_\chi = 200$ GeV, the competing decay $s \to \chi\chi$ becomes accessible for $m_s > 2m_\chi = 400$ GeV. The opening of this channel reduces the branching fraction to HH and therefore weakens

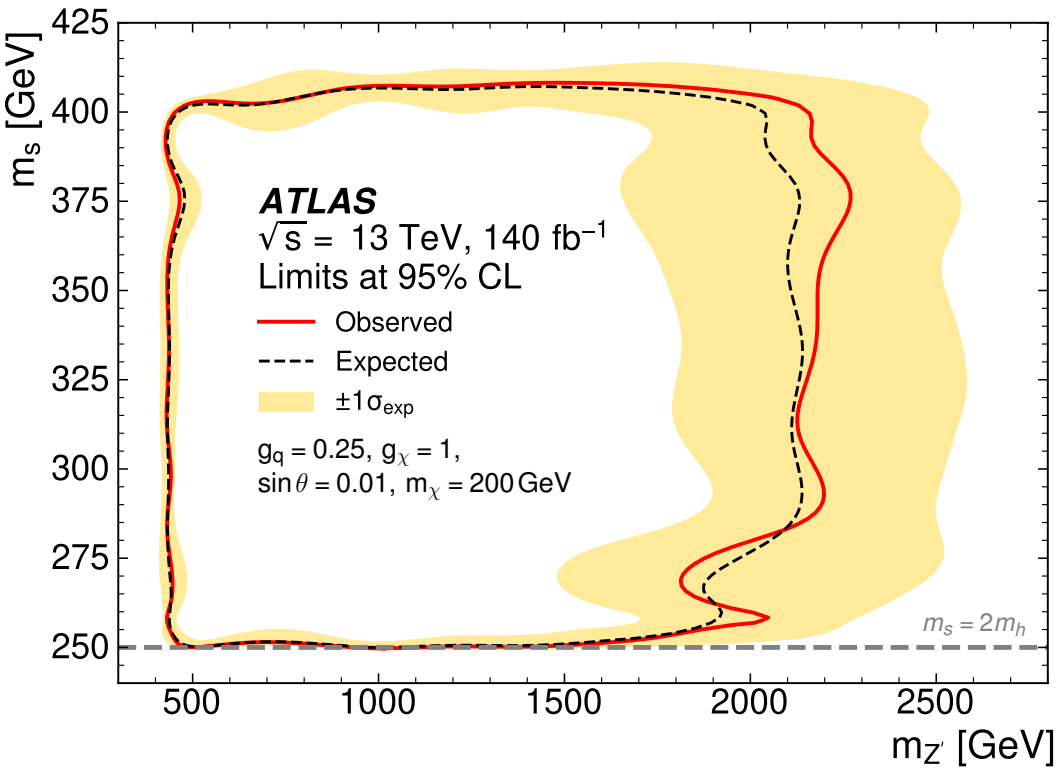


Figure 7.21: Observed and expected exclusion limits in the $(m_{Z'}, m_s)$ plane at 95% CL, with the excluded part inside the closed contour. The $\pm 1\sigma$ expected exclusion intervals are shown as the filled bands. The horizontal dashed line at $m_s = 2m_H$ indicates the limit where the $s \to HH$ decay mode becomes kinematically accessible.

the sensitivity in the $HH + p_{\mathrm{T}}^{\mathrm{miss}}$ final state. The exclusions therefore cover most of the kinematically accessible region of the benchmark mass plane in which both the signal production rate and the branching fraction to HH remain sufficiently large.

Table 7.9 shows the contribution of each systematic category to the total uncertainty of the signal normalisation for the $m_{Z'}$= 1400 GeV, m_s = 385 GeV signal model. The limited data statistics within the signal region dominate, with 95% of the total uncertainty. Other, sub-leading, uncertainties include the modelling of top-related backgrounds and the limited availability of MC statistics.

Figure 7.22 is a contextualised limit plot, where the $HH+p_{\mathrm{T}}^{\mathrm{miss}}$ analysis is shown together with the previous ATLAS searches for the same model. Here it can be seen that the $HH+p_{\mathrm{T}}^{\mathrm{miss}}$ final state probes high m_s and high $m_{Z'}$, significantly extending ATLAS' reach within the dark Higgs model. Furthermore, the results extend reach

Table 7.9: Summary of the relative impact of each category of uncertainty on the SR yield in a signal-plus-background fit for a signal model with $m_{Z'} = 1400$ GeV and $m_s = 385$ GeV.

Source of uncertainty	Percentage of total uncertainty
Signal modelling	< 1
b-jet identification	5
Jet calibrations	9
Other experimental	2
$t\bar{t}$ modelling and normalisation	11
$Z+$jets modelling and normalisation	9
Single-top modelling and cross section	18
Other background modelling and cross section	13
MC statistical uncertainty	17
Overall systematic uncertainty	33
Data statistical uncertainty	95
Total uncertainty	100

beyond the searches performed by CMS [101, 103] in the $s \to b\bar{b}$ and WW final states. The dashed line indicates the precise region of $(m_s, m_{Z'})$-space that exactly replicates the observed dark matter relic density [55], with the region following the diagonal lines favouring a relic density above $\Omega_c h^2 = 0.12$. The resonant annihilation channel via $\chi\chi \to s \to$SM becomes kinematically viable at $m_s = 2m_\chi$ — leading to large DM depletion below the observed relic density. The smearing to higher m_s arises from the distribution of possible DM velocities, allowing this channel to be kinematically viable frequently enough up to $m_s \sim 430$ GeV, above which it becomes sub-dominant. The small deviation of the relic density line at $m_s = 200$ GeV is explained by the opening of the $\chi\chi \to ss$ channel. This additional annihilation channel reduces the relic abundance and prevents dark matter overproduction even when annihilation through the Z' mediator is inefficient.

Despite the lower branching fraction for $s \to HH$ than for $s \to WW$, the reconstruction efficiency of a higher-mass boson decaying into a fully resolved and visible final state, together with the $H \to b\bar{b}$ branching ratio and novel machine learning techniques, enables the HH channel to reach higher-s and Z' masses.

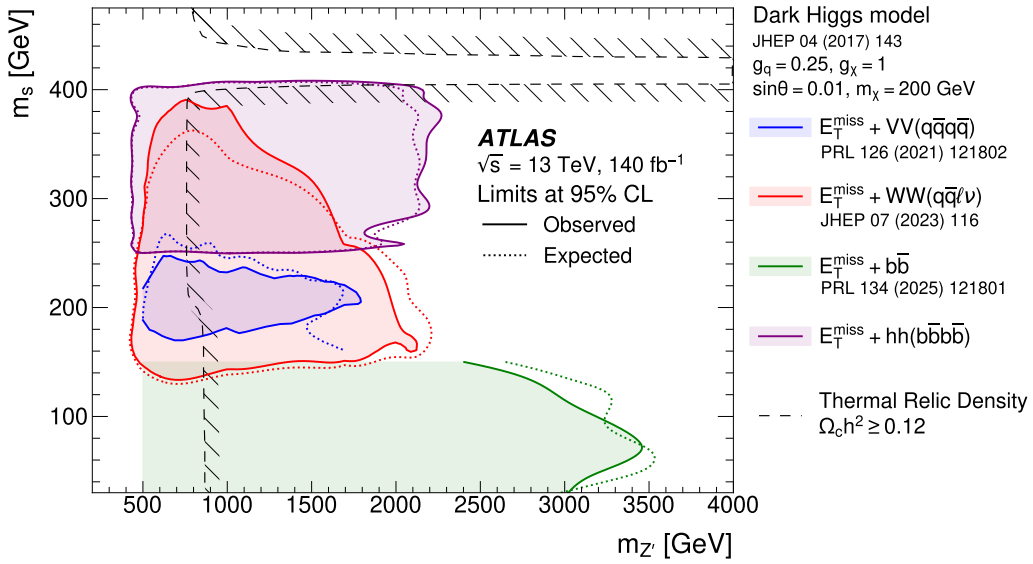


Figure 7.22: Contextualised expected and observed $HH \to 4b+p_{\mathrm{T}}^{\mathrm{miss}}$ limits with the previous ATLAS dark Higgs searches. The parameter points where the relic density, as calculated with `MadDM` [214], matches the value observed by Planck are shown by the short-dashed line.

7.12 Conclusions

A search for a dark Higgs boson in the $HH + p_{\mathrm{T}}^{\mathrm{miss}}$ final state is presented using the full 140 fb^{-1} of proton-proton data collected during Run 2 by the ATLAS experiment. The analysis employs a neural-network classifier tailored to the resonant HH system, with a dedicated mass-decorrelation strategy to distinguish SM backgrounds from the dark Higgs signal. Control regions are defined to constrain and improve the modelling of the top and $Z+$jets background processes, and the background predictions are compared with the observed data in a `SR` binned in m_{HH}. No excess is observed above the SM expectation, and limits are therefore set that extend ATLAS' exclusion reach across the full dark Higgs boson mass range between $250 < m_s < 405$ GeV and up to 2.3 TeV in $m_{Z'}$.

This is the first search in the resonant $HH + p_{\mathrm{T}}^{\mathrm{miss}}$ final state at colliders. The resonant signature is predicted in a broad range of theories positing dark matter through scalar-portal interactions with the Higgs boson, making the result relevant both to the dark Higgs benchmark considered here and to possible reinterpretations

in related models. Moreover, although this analysis provides strong first insights, its statistical limitations point toward further exploration into the $HH+p_{\mathrm{T}}^{\mathrm{miss}}$ final state with the more than two-times-larger dataset collected at $\sqrt{s} = 13.6$ TeV in Run 3. Further refinements of the analysis, including the latest GN2 b-tagging algorithm and considering both resolved and boosted jet reconstruction will extend the sensitivity of searches within the same final state.

"If the laws of physics no longer apply in the future,
God help you."

— GLaDOS, Portal 2

8

Quartic Higgs self-coupling at the FCC

Contents

8.1 Measurements of the Higgs Potential 152
8.1.1 Limitations of the LHC 153
8.2 The FCC Project 155
8.2.1 FCC-ee 155
8.2.2 FCC-hh 156
8.3 Triple Higgs at the FCC-hh 158
8.3.1 Signal Generation 159
8.3.2 SM Triple-Higgs cross section 162
8.4 The $4b2\tau$ final state and backgrounds 164
8.4.1 Production Kinematics of Triple Higgs at the FCC . . . 165
8.4.2 Object Identification 169
8.5 Event Reconstruction and Preselection 172
8.6 Di-τ mass reconstruction with `MMC` 175
8.6.1 The Missing Mass Calculator algorithm 175
8.6.2 MMC Weighting 178
8.6.3 Implementation and Results 178
8.6.4 $m_{\tau\tau}^{\mathrm{MMC}}$ performance 182
8.7 $HH \to 4b$ Pairing 185
8.8 Event Preselection 189
8.9 Preselection and Machine Learning 191
8.9.1 Training Performance 194
8.9.2 Robustness to (κ_3, κ_4) variations 196
8.10 Signal region and Statistical Interpretation 198
8.10.1 Statistical Analysis 200
8.10.2 Systematic Uncertainties 201
8.10.3 Fit Behaviour 202
8.11 Results 204
8.12 Di-Higgs at the FCC-hh 207
8.13 Combination 213
8.14 A view to the future 218
8.15 Conclusions 218

This chapter details a feasibility study for measuring the quartic Higgs self-coupling at the proposed proton-proton stage of the Future Circular Collider

(FCC-hh) at $\sqrt{s} = 84$ TeV. There are two elements of this chapter, the first and most detailed will be an analysis probing the production of three Higgs bosons (HHH). The triple-Higgs final state with two τ-leptons and four b-jets ($4b2\tau$) is targeted. A complete implementation and calibration of an advanced di-τ mass reconstruction tool `MMC` in the FCC software suite is given. The remaining $4b$ system is paired to form kinematic features for an NN trained for both the fully hadronic ($\tau_{\text{had}}\tau_{\text{had}}$) and semi-leptonic ($\tau_{\text{lep}}\tau_{\text{had}}$) analysis channels. Following this is a short initial study into the complementary power that the production of two SM Higgs bosons provides to κ_4 via indirect constraints in the $b\bar{b}\gamma\gamma$ final state. Both analyses use neural networks to enhance signal–background separation, with profile-likelihood fits performed to extract results simultaneously on (κ_3, κ_4).

8.1 Measurements of the Higgs Potential

Measurements probing the electroweak symmetry breaking mechanism (EWSB) of the SM via the shape of the Higgs potential are a crucial component of the LHC physics programme. The Higgs potential within the κ-framework may be parametrised as

$$V(h) = \frac{1}{2}m_H^2 h^2 + \kappa_3 \lambda v h^3 + \frac{\kappa_4}{4}\lambda h^4 \tag{8.1}$$

where m_H is the mass of the SM Higgs boson, v is the vacuum expectation value, $\lambda = \frac{m_h^2}{2v^2}$ and κ_3 and κ_4 are the ratios of the tri-linear and quartic Higgs self-couplings to their SM predictions ($\kappa_i = \lambda_i/\lambda_i^{SM}$), such that $\kappa_3 = \kappa_4 = 1$ is the SM. To constrain κ_3 directly, the production of two Higgs bosons (HH) is needed. The measurement of the inclusive and differential triple-Higgs (HHH) cross section provides direct tree-level constraints on κ_4. The four leading diagrams are reproduced from earlier discussions in Section 3.2 in Figure 8.1.

Di-Higgs production receives loop-level contributions that depend on the quartic Higgs self-coupling; examples are shown in Figure 8.2. These contributions modify the HH cross section and shape information in m_{HH}, which in principle allow both κ_3 and κ_4 to be constrained simultaneously. Given that these corrections

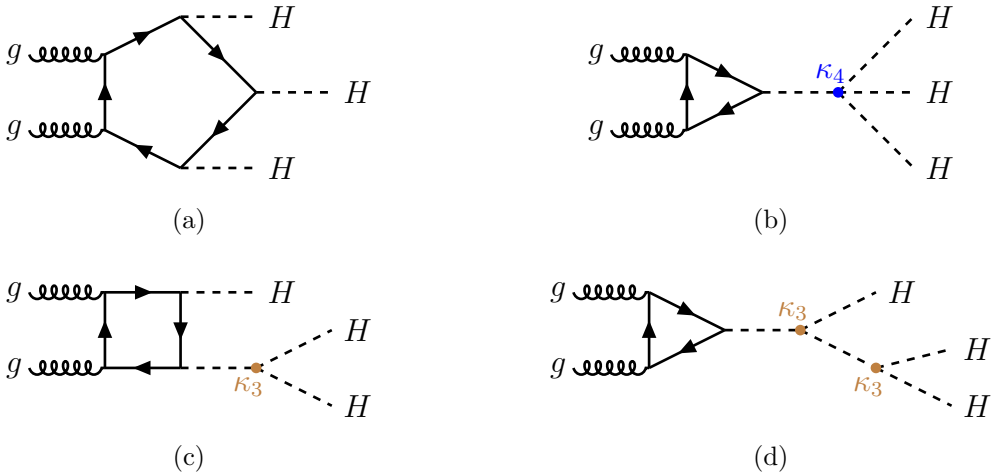


Figure 8.1: Feynman diagrams for leading-order $gg \to HHH$ production via a quark loop. The dominant production mechanism is shown in (a), whilst sub-leading diagrams (b), (c) and (d) are sensitive to κ_4, κ_3 and κ_3^2 respectively, as indicated by labels. Figure reproduced from Section 3.2 for ease of reference.

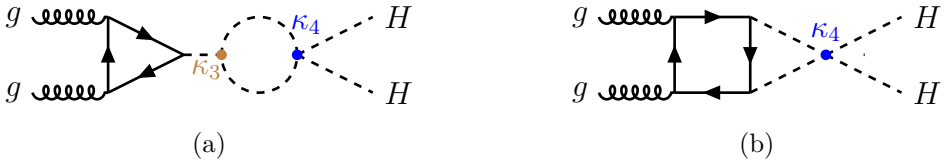


Figure 8.2: Feynman diagrams for 2-loop $gg \to HH$ production via a quark loop and a loop-correction to the Higgs self-energy (a) and additional self-interactions present for the box configuration, with vertices sensitive to κ_3 and κ_4 illustrated. Figure reproduced from Section 3.2 for ease of reference.

are loop-induced, the modifications are small in comparison with those of κ_3 and direct sensitivity to κ_4 via triple-Higgs production. Nonetheless, the complementary information is used in a combination of HH and HHH-analyses and will be addressed later.

8.1.1 Limitations of the LHC

The LHC has vastly outperformed expectations in measurements of the Higgs boson. Relevantly, the first measurements of the trilinear Higgs self-coupling have been performed using Run 1, 2 and 3 of the LHC. The current experimental picture shows good agreement for the trilinear Higgs self-coupling, with the Run 2 ATLAS and CMS combination shown in Figure 8.3. Looking ahead to the end of the HL-LHC, if significant deviations are not observed, κ_3 will be constrained to within 30% of

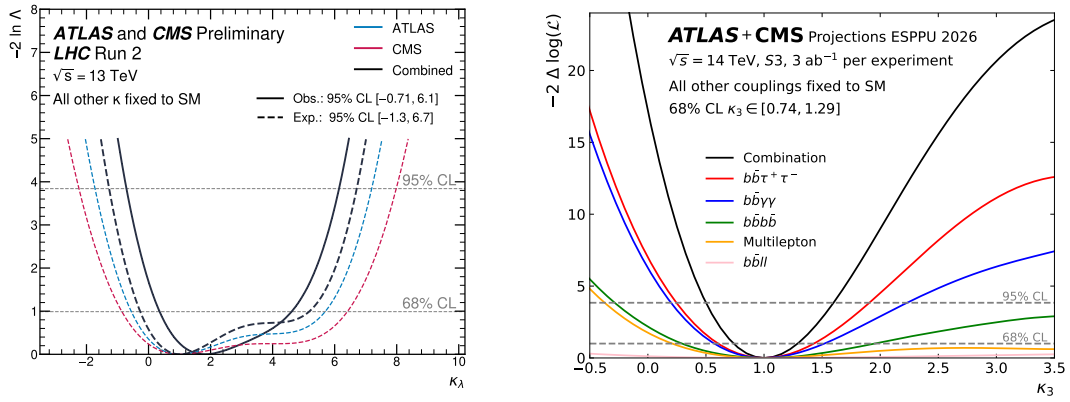


Figure 8.3: Negative log-likelihood scans over κ_λ (equivalent to κ_3) for the Run 2 di-Higgs ATLAS and CMS combination (left), together with the HL-LHC projections (right). Figures taken from [77, 218]

the SM value. Despite this success, however, it notably leaves the quartic Higgs self-coupling mostly unconstrained. Both the ATLAS and CMS experiments have performed direct searches for triple-Higgs production and included κ_4-interpretations. The first experimental result came from ATLAS in the $6b$ final state and, assuming a SM value for κ_3, set limits at $\kappa_4 \in [-230, 240]$ [215]. CMS have also studied the $6b$ final state, utilising both boosted and resolved event reconstruction strategies, together with the contributions of the di-Higgs background reweighted with κ_3, to set a more stringent limit of $\kappa_4 \in [-190, 190]$ [216]. Finally, CMS have also performed a search in the $4b2\gamma$ final state, which, given the small branching fraction and statistically limited signal, set weaker limits at $\kappa_4 \in [-533, 541]$ [217]. These limits, albeit impressive first investigations of this coupling space, are orders of magnitude away from the SM prediction and leave large areas of parameter space that are viable for first-order phase transitions unexplored. Upon completion of the HL-LHC runs, ATLAS and CMS are expected to constrain $\kappa_4 \in [-84, 96]$ for $\kappa_3 = 1$ at 95% confidence level (CL) in the $6b$-channel [75]. Only modest improvements of this limit are expected through improved reconstruction techniques and combination with additional channels. Notably, it is not foreseen that these limits will appreciably probe the space not excluded by perturbative unitarity constraints [78].

The limiting factor in all LHC triple-Higgs measurements is the exceptionally small cross section. At NNLO in QCD, the cross section for HHH production at 14 TeV is $\sigma_{HHH}^{\text{NNLO}} = 0.103\text{fb}$ [76]. Assuming a total integrated luminosity of 3ab^{-1} of pp collision data at the HL-LHC, this corresponds to just 309 expected events. When factoring in branching ratios, object efficiencies, detector acceptance, and analysis selections, only a few candidate events remain for a given final state. Escaping this limitation requires a machine with a significantly larger centre-of-mass energy. A simple argument for this follows from the brief introduction in Section 2.3. The increase in $\sqrt{s}$ greatly enhances the cross section given the exponentially rising gluon PDF at low-x, which is more pronounced for processes at higher scales, like the production of three SM Higgs bosons. The ratio of NNLO triple-Higgs cross sections at 84 TeV and 14 TeV is approximately 36, meaning for each unit of luminosity collected by an 84 TeV collider, 36 times more HHH events are expected compared to the HL-LHC, enabling the study of the quartic Higgs self-coupling. A proposed experiment to achieve this is the Future Circular Collider, the next flagship collider at CERN.

8.2 The FCC Project

The Future Circular Collider (FCC) project is a proposed collider based at CERN. It is proposed in two stages, first as an electron-positron collider (FCC-ee) and, once magnet technology meets the demands of a high-energy proton-proton collider, a proton-proton run (FCC-hh) will operate.

8.2.1 FCC-ee

Electron-positron collisions offer clean environments for precision measurements of SM processes. The following is a brief and non-exhaustive summary of the physics potential of the FCC-ee; for a comprehensive summary, see the FCC CDR Volume 1 in Ref. [219]. The broad programme requires a variety of centre-of-mass energies, with the Z-pole, WW-threshold, ZH and finally at the $t\bar{t}$-threshold. Firstly, the FCC-ee will operate around the Z-peak (collecting 2×10^5 times the data collected by

the LEP experiment) — enabling orders of magnitude improvement on electroweak precision observables: m_Z, Γ_Z, forward-backward asymmetries and finally the Z lineshape. Operation around the WW and $t\bar{t}$-threshold will enable precision measurements of W-boson and top masses and widths. Precision measurements are sensitive to high-mass particles running in loops, enabling the FCC-ee to also be a discovery-based experiment. For the Higgs, the FCC-ee will be the first opportunity to realistically measure its coupling to second- and possibly first-generation fermions. The recoil method also enables the most precise measurement of the Higgs mass, and combined with measurements of $H \to ZZ^*$ extract the Higgs width, Γ_H, to within 1%. To enable the study of Higgs couplings to electrons, the accelerator requires excellent precision of the beam-energy spread — monochromatisation is a proposed method to achieve this [220]. This will also be the first model-independent measurement of the Higgs couplings, in which the recoil-mass method with ZH enables the rate of Higgs production to be measured irrespective of the Higgs final state. Loop-corrections to single-Higgs production will enable constraints to be set on the trilinear Higgs self-coupling [4], complementary to the HL-LHC.

8.2.2 FCC-hh

Following 14 years of e^+e^- data-taking, the accelerator complex would be re-purposed for the FCC-hh. The goal of such an experiment is to precisely measure the Higgs potential and to perform direct searches for new physics at significantly higher mass scale than at the LHC.

Similar in layout to the ATLAS detector as described in Section 4.2, the FCC-hh detector is designed to achieve excellent momentum resolution over an enormous range of energies, from low-p_{T} particles produced by possible extensions of the SM [221] to particles upwards of 40 TeV. A longitudinal cross section of the detector is shown in Figure 8.4.

The inner detector is placed within a 4 T solenoidal magnetic field, extending the tracking coverage to $|\eta| < 4$ — the same as that achieved by the ITk for the HL-LHC. It is an all-silicon tracker, with the innermost pixel layer placed at 25 mm

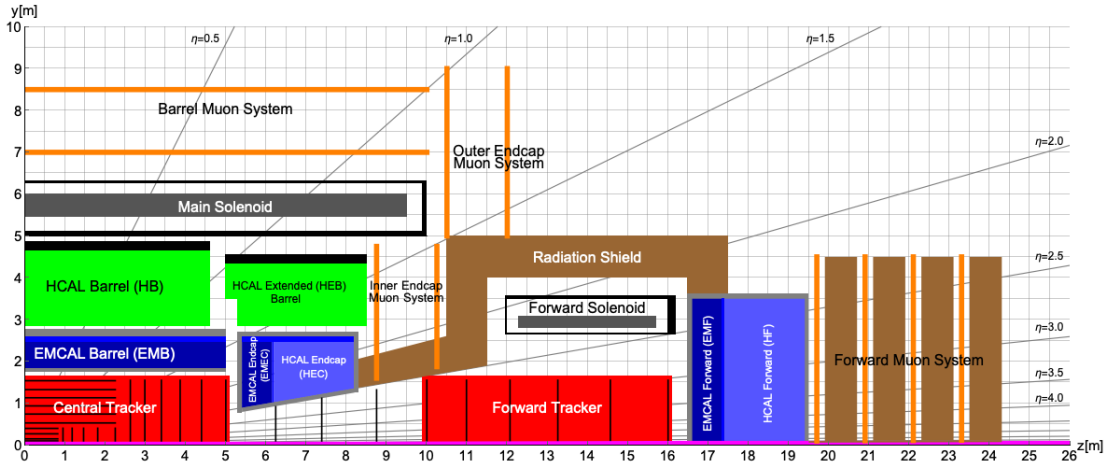


Figure 8.4: Longitudinal cross section of the FCC-hh reference detector. Figure from Ref [222].

from the beamline (compared to 33 mm for the IBL, or 34 mm for the ITk) with (25×50) μm granularity. The tracker extends to approximately 1.5 m from the interaction point (compared to 1 m for the ITk), providing an extended lever arm to measure the tracks of charged particles within a two times larger magnetic field.

The ECAL (HCAL) surrounds this, with a barrel ECAL (HCAL) covering the tracker volume, following the same lead (steel) passive/LAr (scintillating-tile) active-layer sampling design as ATLAS. It is, however, more finely segmented, with eight (ten) layers of $\Delta\eta \times \Delta\phi = 0.01(0.025) \times 0.009(0.025)$ cell dimension. The ECAL (HCAL) is extended into the forward region with endcaps covering $4 < |\eta| < 6$ with six layers of $\Delta\eta \times \Delta\phi = 0.025(0.025) \times 0.025(0.025)$ cell dimension. The finer segmentation of calorimeters enables the reconstruction of boosted objects, such as a $H \to b\bar{b}$ decay, at much higher momentum.

The calorimeters target resolution functions,

$$\frac{\sigma_E}{E} = \frac{10(50)\%}{\sqrt{E/\text{GeV}}} \oplus 0.15(1)\%. \tag{8.2}$$

where the resolutions for the ECAL (HCAL) are given. The smaller constant term compared with the ATLAS calorimeters, 0.7(3)%, enables precision measurements to much higher energies.

Finally, a muon spectrometer encloses the detector and covers the region $|\eta| < 6$ with a momentum resolution of 10% for muons up to 4 TeV.

The detector design, as outlined above, is simulated using the full geometry in GEANT4 [4]. However, since many of the design choices remain flexible, feasibility studies instead utilise an approximate implementation of the detector response. DELPHES is a parametrised detector simulation framework where, instead of a full-simulation like GEANT4, it takes stable truth particles from an event-generator and applies approximate resolution maps to construct objects and applies efficiency maps for particle identification. This is significantly faster than GEANT4 based simulation, while still capturing the main experimental effects needed for feasibility and sensitivity studies.

Average responses in momenta and angular spread are parametrised as a function of p_{T} and η for each object to form a DELPHES [223] configuration. To define the *truth* component, stable truth-particles are used as inputs to the anti-k_t jet clustering algorithm with $R = 0.4$. Truth-jets are then matched to reconstructed jets if they are within $\Delta R < 0.3$. Current state-of-the-art object identification performance from ATLAS and CMS is taken as the baseline for the FCC-hh, with improvements expected prior to operation. The corresponding identification efficiencies are parametrised as a function of p_{T} for $|\eta| < 4$.

With the large centre-of-mass energy and instantaneous luminosity, the pileup profile is significant, reaching $\mathcal{O}(600)$ additional interactions along the beamline. Despite this, pile-up is not considered in the analysis, since studies into the FCC-hh pile-up mitigation performance are ongoing and techniques are foreseen to significantly improve with the experience gained during the HL-LHC — operating nominally around an average pile-up of $\mathcal{O}(200)$. Including the harsh pile-up profile without such advanced techniques deployed therefore would be overly conservative.

8.3 Triple Higgs at the FCC-hh

With the nominally proposed centre-of-mass of $\sqrt{s} = 84$ TeV, collecting a total integrated luminosity of $30\,\mathrm{ab}^{-1}$, high-precision single and double-Higgs measure-

ments can be performed, whilst opening up the possibility to attempt measurements of triple-Higgs production. The following chapter outlines a feasibility study for triple-Higgs production at the FCC-hh, enabling limits to be set on the quartic Higgs self-coupling via κ_4. All aspects were performed by the author, including: simulated samples, background estimation, event reconstruction, event selection, machine learning, statistical interpretation, results and conclusions for the study.

8.3.1 Signal Generation

Standard Model HHH samples are generated at $\sqrt{s} = 84$ TeV using the MadGraph5_aMC@NLO [224] event generator and decayed using MadSpin [225]. Parton distribution functions are taken from the `MMHT2015qed_nlo` PDF set [226] and interfaced with Pythia 8 [227] for parton showering and hadronisation. The samples are generated at the SM point, $(\kappa_3, \kappa_4) = (1, 1)$, and then reweighted to describe arbitrary values of κ_3 and κ_4. This is achieved using a *moment morphing* technique, which exploits the known form of the matrix element squared as a polynomial in κ_3 and κ_4,

$$\mathcal{M}(\kappa_3, \kappa_4) = \mathcal{M}_a + \kappa_3 \mathcal{M}_c + \kappa_3^2 \mathcal{M}_d + \kappa_4 \mathcal{M}_b \tag{8.3}$$

where the $\mathcal{M}_i$ are functions determined by the underlying Feynman diagrams shown in Figure 8.1. The squared matrix element can be written in terms of nine independent monomials:

$$|\mathcal{M}(\kappa_3, \kappa_4)|^2 = m_1 + m_2\kappa_3 + m_3\kappa_3^2 + m_4\kappa_3^3 + m_5\kappa_3^4 + m_6\kappa_4 + m_7\kappa_4^2 + m_8\kappa_3\kappa_4 + m_9\kappa_3^2\kappa_4, \tag{8.4}$$

where m_i are to be determined per-event.

Benchmark points

To determine the moments, nine benchmark points in the (κ_3, κ_4) plane are chosen, and the matrix element squared is sampled for each, such that they span the space

of monomials. The benchmark set used in this analysis is

$$(\kappa_3, \kappa_4)_j = \begin{cases} (0,0), & j = 1, \\ (1,-1), & j = 2, \\ (-1,1), & j = 3, \\ (1,0), & j = 4, \\ (0,1), & j = 5, \\ (-1,0), & j = 6, \\ (0,-1), & j = 7, \\ (0.5,0), & j = 8, \\ (-0.5,0), & j = 9. \end{cases} \tag{8.5}$$

For each generated event, the matrix element reweighting, performed within MadGraph5_aMC@NLO [228], provides a set of weights

$$\mathbf{w}_{\text{event}} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_9 \end{pmatrix}, \qquad w_j \equiv w(\kappa_{3,j}, \kappa_{4,j}), \tag{8.6}$$

corresponding to these nine benchmark points. This definition yields a linear system:

$$\mathbf{w}_{\text{event}} \;=\; F\,\mathbf{m}, \tag{8.7}$$

where F has entries

$$F_{ji} \;=\; f_i(\kappa_{3,j}, \kappa_{4,j}) \tag{8.8}$$

Equivalently, the j-th row of F is simply the vector of basis functions evaluated at the j-th benchmark point,

$$F_j = \Big(1,\, \kappa_{3,j},\, \kappa_{3,j}^2,\, \kappa_{3,j}^3,\, \kappa_{3,j}^4,\, \kappa_{4,j},\, \kappa_{4,j}^2,\, \kappa_{3,j}\kappa_{4,j},\, \kappa_{3,j}^2\kappa_{4,j}\Big). \tag{8.9}$$

The moments for a given event are obtained by inverting Eq. (8.7):

$$\mathbf{m} \;=\; F^{-1}\,\mathbf{w}_{\text{event}}. \tag{8.10}$$

Weights at arbitrary (κ_3, κ_4)

With the moments $\mathbf{m}$ known for a given event, the weight at any desired (κ_3, κ_4)-point is obtained by evaluating the basis functions at that point and forming

$$w(\kappa_3, \kappa_4) \;=\; \sum_{i=1}^{9} m_i \, f_i(\kappa_3, \kappa_4) \;=\; \mathbf{f}(\kappa_3, \kappa_4)^{\mathrm{T}} \, \mathbf{m}. \tag{8.11}$$

This reweighting scheme allows a single underlying event sample to be reinterpreted at arbitrary values of the Higgs self-couplings. The resulting variation of the total cross section across the (κ_3, κ_4) plane is shown in Fig. 8.5, where each point is normalised to the SM prediction. There is a region of smaller cross section relative to the SM for small positive deviations from the SM in (κ_3, κ_4), mostly driven by the destructive interference between the pentagon and box diagrams (see Figure 3.3). Outside of this region, strong enhancement of the cross section is observed and generally dictates the ultimate sensitivity of the analysis. The relative weighting of the matrix element also changes the kinematics, with the destructive region generally corresponding to strongly boosted Higgs bosons — making event reconstruction challenging in this phase space.

Implementation in the FCC software framework

The author implemented this approach, and the nine benchmark weights w_j produced at generator level using the built-in matrix-element reweighting of MadGraph5_aMC@NLO are stored in LHE files as separate weight entries. These weights are propagated through Pythia 8 and the `Key4SimDelphes` simulation into the `EDM4hep` event record as per-event scalar weight collections within the `Key4hep` software suite [229]. Within the analysis, the fixed matrix F and its inverse are constructed once from the chosen benchmark points (Eq. 8.5), and the moments $\mathbf{m}$ for each event are computed according to Eq. (8.10). Arbitrary (κ_3, κ_4) configurations are then obtained on the fly by evaluating Eq. (8.11) inside the `FCCAnalyses` framework, ensuring a consistent treatment of the kinematic dependence of the cross section throughout the analysis.

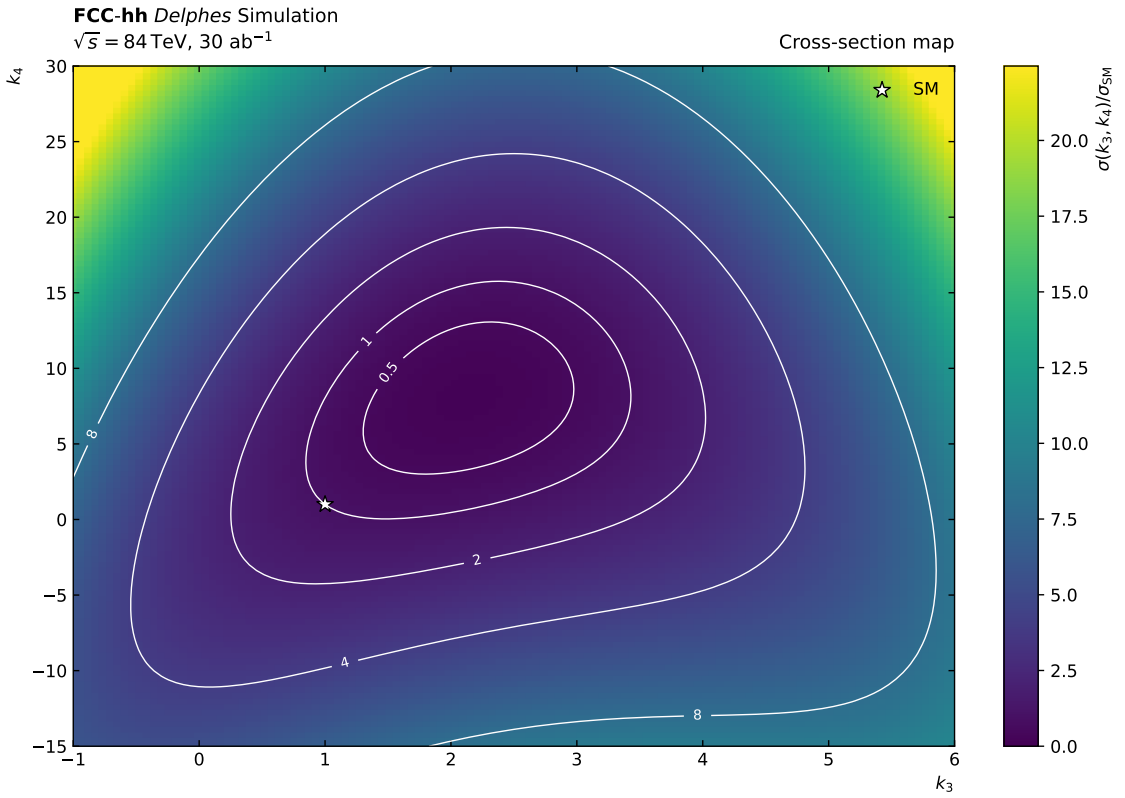


Figure 8.5: Ratio of cross section at each point in (κ_3, κ_4) to the SM prediction.

8.3.2 SM Triple-Higgs cross section

The amplitude for triple-Higgs production at leading order (one-loop) is known, first numerically [230] and more recently analytically [231]. Due to the multi-scale loops present in the box and pentagon diagrams, as shown in Figure 8.1 (a,c), computing the exact corrections at NLO for triple Higgs is extremely challenging. An approximation made to make the calculation manageable is the heavy-top-limit (HTL), in which the top mass is taken to infinity ($m_t \to \infty$), collapsing the loops to effective ggH vertices. Here, the QCD corrections are qualitatively the same as the known calculation of QCD corrections to single-Higgs production. Further corrections are applied to approximate the effects of a finite top mass. The validity of the HTL depends on the scale of the event: $Q^2 = m^2_{hhh}$. This approximation is powerful for low-scale processes ($Q^2 < m_t^2$), but begins to suffer at high scales — where a large portion of triple-Higgs events lie, as shown by Figure 8.6. This approximation is pulled more into question with a high-energy

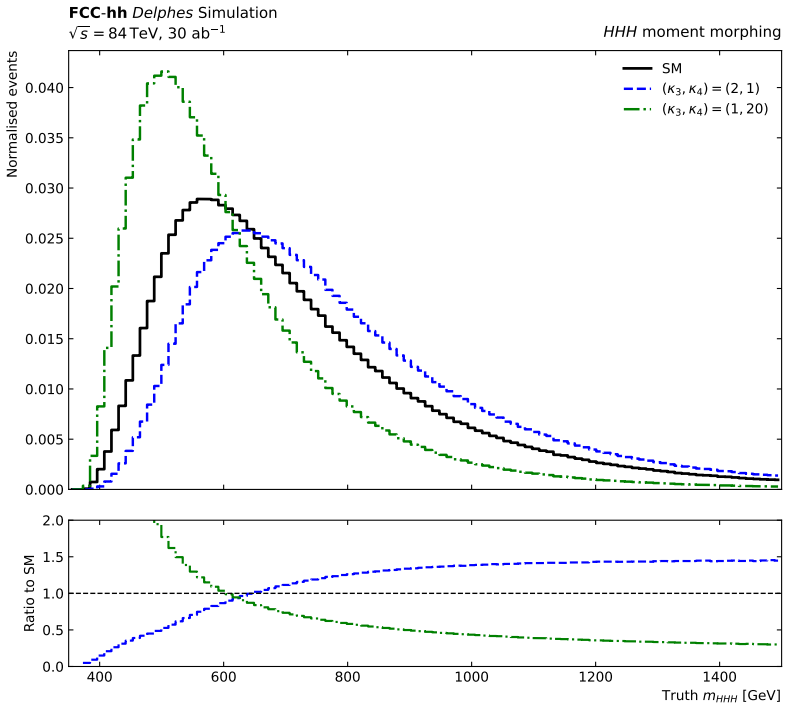


Figure 8.6: Distributions of the truth HHH invariant mass, m_{HHH}^{truth}, for the SM prediction and two example (κ_3, κ_4) variations. No selection requirements are applied. Histograms are normalised to unit area, and the lower panel shows the ratio of the varied (κ_3, κ_4) models to the SM distribution.

collider, like the FCC, given that the events occur with an enhanced rate at high-Q^2, as shown in Figure 8.7.

Nevertheless, this NNLO QCD calculation including approximate finite-top-mass effects has been performed for 14, 27 and 100 TeV centre-of-mass energies [76]. To extrapolate to 84 TeV, the 100 TeV cross section is multiplied by the ratio of cross sections at leading order using MadGraph5_aMC@NLO, yielding a cross section of $\sigma_{HHH}^{84\text{TeV}} = 3.56$ fb.

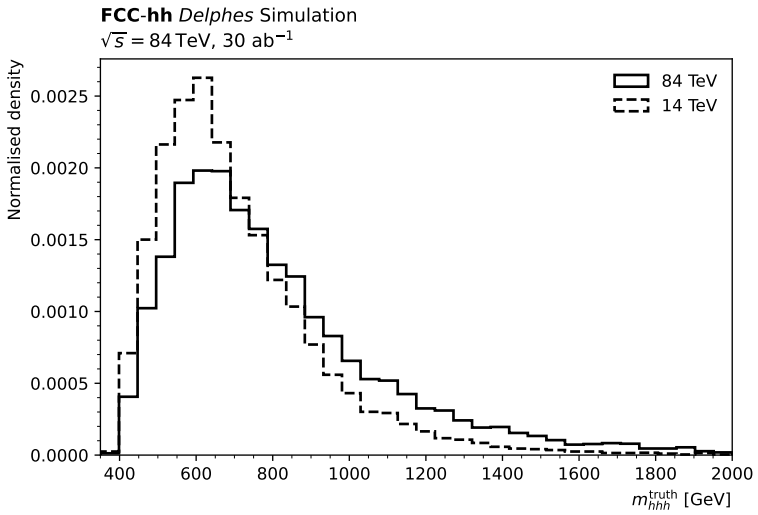


Figure 8.7: Normalised distributions of the triple-Higgs invariant mass at truth-level for the SM at $\sqrt{s} = 84$ TeV for an FCC-hh and $\sqrt{s} = 14$ TeV for the HL-LHC.

8.4 The $4b2\tau$ final state and backgrounds

With the large datasets collected by the FCC-hh, final states other than the high-branching-fraction $6b$ can be considered. A number of final states are listed in Table 8.1, with the branching fraction expressions shown and the expected number of events at the FCC-hh. This study chooses to target the $4b2\tau$ final state, where, despite the smaller branching fraction, the selection of two τ-leptons enhances rejection of the background — enabling a purer signal selection to be defined.

Naturally, for this final state, $t\bar{t}$ production, with the W-boson decaying via $\tau\nu$, in association with $b\bar{b}$ is the dominant background. Other final states involving a $t\bar{t}$ pair, such as $t\bar{t}Z$, $t\bar{t}H$ and $t\bar{t}t\bar{t}$ are also considered together with multi-boson ZZZ production. Finally, a di-Higgs background is considered and is simulated at NLO in QCD and includes the two-loop amplitudes relevant for the trilinear and quartic Higgs self-coupling variations using POWHEG, including variations on (κ_3, κ_4) [79, 232–234]. Rejection of the non-resonant background is therefore crucial for the performance of the analysis, where the efficient reconstruction of three SM Higgs bosons should effectively reduce $t\bar{t} + b\bar{b}$. Moreover, the presence of 2(3) neutrinos in the final state for $\tau_{\text{had}}\tau_{\text{had}}(\tau_{\text{lep}}\tau_{\text{had}})$ complicates the reconstruction of

the di-τ invariant mass; reconstructing this will again prove powerful at reducing non-resonant backgrounds. Given the non-perturbative nature of QCD, accurately simulating QCD multi-jet backgrounds is extremely challenging. The selection of two τ-leptons will effectively eliminate a large portion of this background component, and is expected to be sub-leading in this final state. Conservative k-factors are applied to the dominant background $t\bar{t} + b\bar{b}$, and additional QCD backgrounds will not impact the result significantly. Nonetheless, studies of this final state with more complete estimation of this background component, alongside the inclusion of pileup, constitute interesting further investigation. The aforementioned backgrounds are simulated at leading order in MadGraph 5 [235]. As with the signal process, parton distribution functions are taken from the `MMHT2015qed_nlo` PDF set [226] and interfaced with Pythia 8 [227] for parton showering and hadronisation.

Table 8.1: Expected number of HHH events in various Higgs boson decay channels at $\sqrt{s} = 84$ TeV and $\mathcal{L} = 30$ ab^{-1}, assuming SM Higgs boson branching ratios (BRs) from Ref. [76].

Final state	BR expression	Expected events (30 ab^{-1})
$6b$	$(0.58)^3$	20,819
$4b2\tau$	$3 \times (0.58)^2 \times 0.063$	6,791
$4\tau 2b$	$3 \times (0.063)^2 \times 0.58$	738
$4b2\gamma$	$3 \times (0.58)^2 \times 0.0023$	247
$2\tau 2\gamma 2b$	$6 \times 0.58 \times 0.063 \times 0.0023$	54
6τ	$(0.063)^3$	27
$4b2W$	$3 \times (0.58)^2 \times 0.215$	23,200
$2b4W$	$3 \times 0.58 \times (0.215)^2$	8,571
$6W$	$(0.215)^3$	1,055

8.4.1 Production Kinematics of Triple Higgs at the FCC

With regard to the detector layout, one important consideration is the tracking volume's pseudorapidity coverage. Assuming sharply falling gluon PDFs, a reasonable approximation of the maximum Bjorken-x is $x_{\max} \sim 0.5$. The maximal pseudorapidity of the system, as a function of centre-of-mass energy, is then given by $y_{\max} = \ln\left(\frac{\sqrt{s}}{2m_{HHH}}\right)$. Therefore, Higgs bosons are produced more often at higher pseudo-rapidities for a high-energy collider. The distribution of the maximum

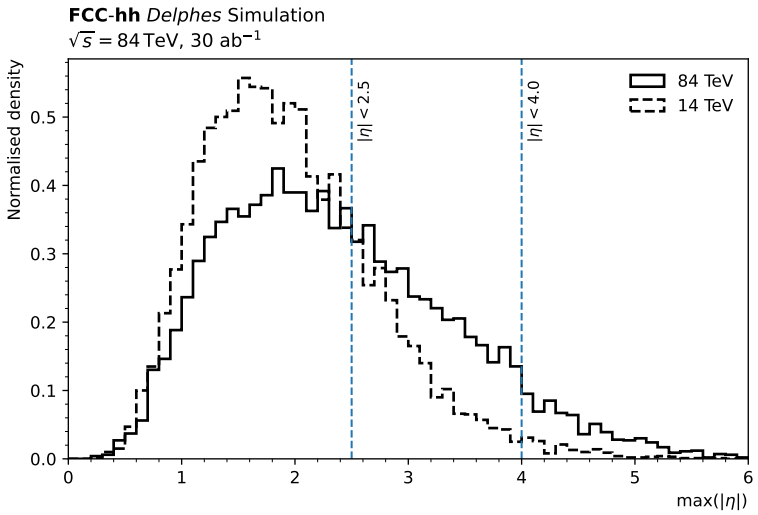


Figure 8.8: Maximum pseudorapidity of decay products for Higgs daughters in HHH events at the HL-LHC (dashed) overlaid with those produced at the FCC-hh (solid).

pseudorapidity of any Higgs daughter in a given event is shown in Figure 8.8. At the FCC-hh, 7% of events contain a parton too forward to be reconstructed by a tracker with an acceptance of $|\eta| < 4$, compared with the more central HL-LHC events, which lose only 2% — highlighting the requirement for strong tracking performance in the forward region.

With the larger phase space, more HHH events are produced at high momentum. Highly boosted particles produce decay products that are more collimated along the direction of motion. Reconstructing the decay products typically relies on small-radius jets ($R = 0.4$); however, for sufficiently large boosts, the partons can be spatially too close within the detector to be resolved by two small-R jets. By investigating the truth separation of Higgs boson decay products, quantified by the distribution of $\min_{H_i}\left(\Delta R(f\bar{f})\right)$ shown in Figure 8.9, these boosted configurations are more common at the FCC-hh than at the HL-LHC. Approximately 32% of events contain a Higgs boson with decay products within $\Delta R < 0.8$.

This is only shown for SM HHH production, but can be more important in regions of parameter space away from the SM. For example, the transverse momenta of the produced Higgs bosons depend strongly on the targeted point in (κ_3, κ_4). A close proxy for the degree of overall event activity is the scalar sum of Higgs

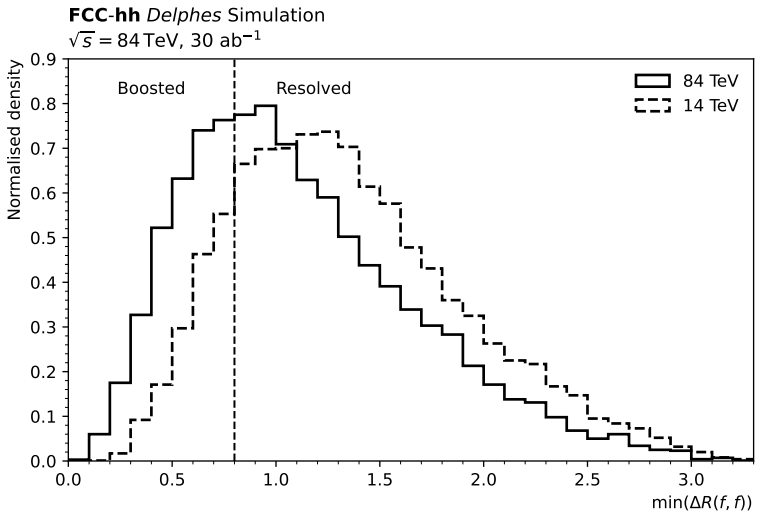
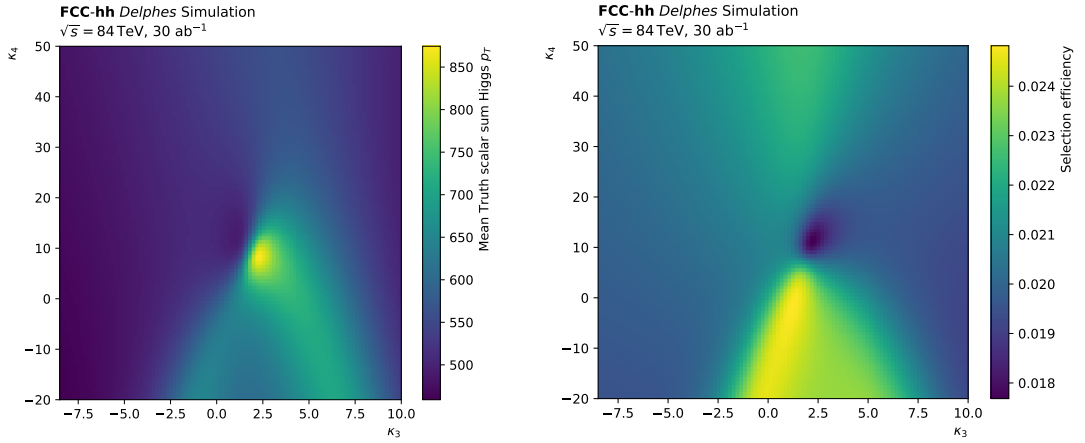


Figure 8.9: Minimum separation of decay products for Higgs daughters in HHH events at the HL-LHC (dashed) overlaid with those produced at the FCC-hh (solid).

transverse-momenta, $H_T\,(h) = \sum_i p_{\mathrm{T}}(H_i)$. To avoid biases from the reconstruction strategy (small-R vs. large-R jets), the truth-level Higgs transverse momenta $p_{\mathrm{T}}(H_i)$ are used when defining $H_T\,(h)$. A 2D histogram of the $H_T\,(h)$ across the (κ_3, κ_4)-plane is shown in Figure 8.10(a). Comparing the reconstruction efficiency for the resolved case, shown in Figure 8.10(b), with the average Higgs momenta, the region of largest boost coincides with the least efficient event reconstruction. Moreover, this region aligns closely with the region of largest destructive interference, as shown in the cross-section plot of Figure 8.5. The potential benefit of large-radius jet reconstruction is clear, given the large number of boosted events at the SM, which is only compounded in regions of (κ_3, κ_4) space where the destructive interference is larger compared with the SM — producing highly boosted final states.

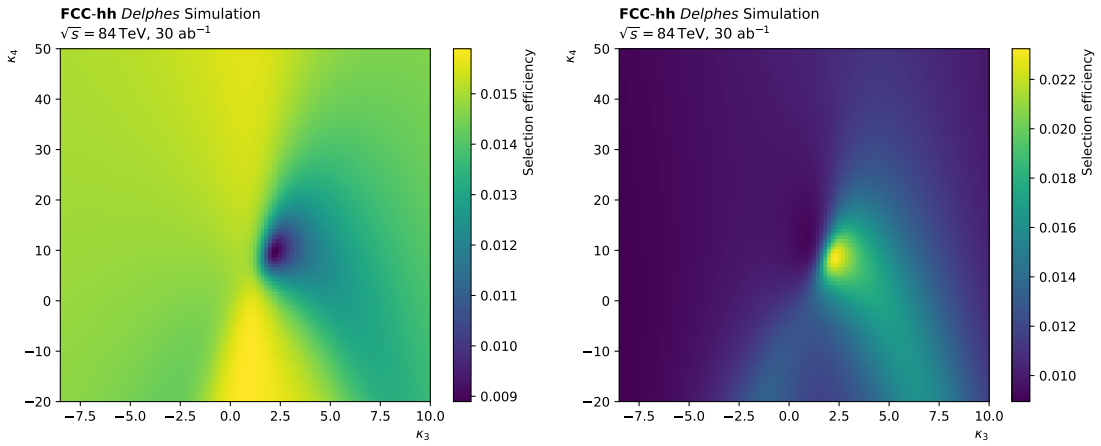
Within Delphes, $R = 0.8$ anti-k_t jets are reconstructed using particle flow objects. The substructure within these jets is resolved using $R = 0.2$ anti-k_t track-jets. Despite pile-up not being considered, track-jets are generally more robust to it and are used at current colliders for this purpose. Partons are associated with track-jets with a radius of 0.2 and passed through identification maps. A large radius jet is considered double-b tagged if two b-tagged sub-jets are contained within $R = 0.8$ of the large-radius jet axis. The bb-tag efficiency for a boosted



(a) **(b)**

Figure 8.10: Scalar sum of truth Higgs boson p_{T} ($H_T\,(h)$) of MadGraph5_aMC@NLO HHH events at leading order per bin in (κ_3, κ_4) (a) and the reconstruction efficiency of the resolved channel across the (κ_3, κ_4)-plane (b).

$H \to b\bar{b}$ is therefore $\varepsilon_b^2 \varepsilon_{\mathrm{match}}^2$, where $\varepsilon_{\mathrm{match}}$ is the matching efficiency of a parton to a track-jet. Small-radius jets overlapping with bb-tagged large-radius jets are removed such that orthogonal resolved and boosted regions can be defined. The benefit in overall event reconstruction efficiency can be seen in Figure 8.11, where a sizeable increase in reconstruction efficiency is observed across the parameter space, alongside the loss in reconstruction efficiency at the highest-$H_T\,(h)$ region recovered when considering large-radius jets. From this initial study, the benefit of large radius jet reconstruction is clear, and the analysis could consider both *boosted* and *resolved* channels to optimise performance. However, due to time constraints, this is left for future work, and the baseline analysis outlined in this thesis uses only resolved reconstruction.



(a) **(b)**

Figure 8.11: Selection efficiency in the resolved (a) and boosted (b) channel across the (κ_3, κ_4) plane.

8.4.2 Object Identification

In this section we study the truth-level kinematics and reconstruction performance of the $HHH \to 4b2\tau$ signal. These studies guide the choice of jet-parton matching parameters, flavour-tagging strategy, and lepton isolation working points used in the analysis.

Truth Matching

Distributions of the truth b-hadron, τ_{had} and τ_{lep} transverse momenta are shown in Figure 8.12. The tagging efficiency as a function of reconstructed jet-p_{T} for hadronic tau lepton decays and for b-jets is shown for the HHH-signal in Figure 8.13. These plots follow the parametrised implementation within the DELPHES configuration, which is guided by the current state-of-the-art within ATLAS and CMS — with the efficiency here assumed more robust with p_{T} when compared with ATLAS due to the finer granularity (25×50) μm, and closer placement, of the innermost pixel layers.

DELPHES relies on truth-association to assign flavour to reconstructed jets. A truth parton matched to a jet within a default radius of $\Delta R = 0.5$ is passed through the efficiency maps for a parton of that flavour. If two truth stable partons enter the same jet, the efficiency map for each flavour is applied. Thus, for a reconstructed jet containing both a b and τ it has probability $\varepsilon_b \varepsilon_\tau$ of being

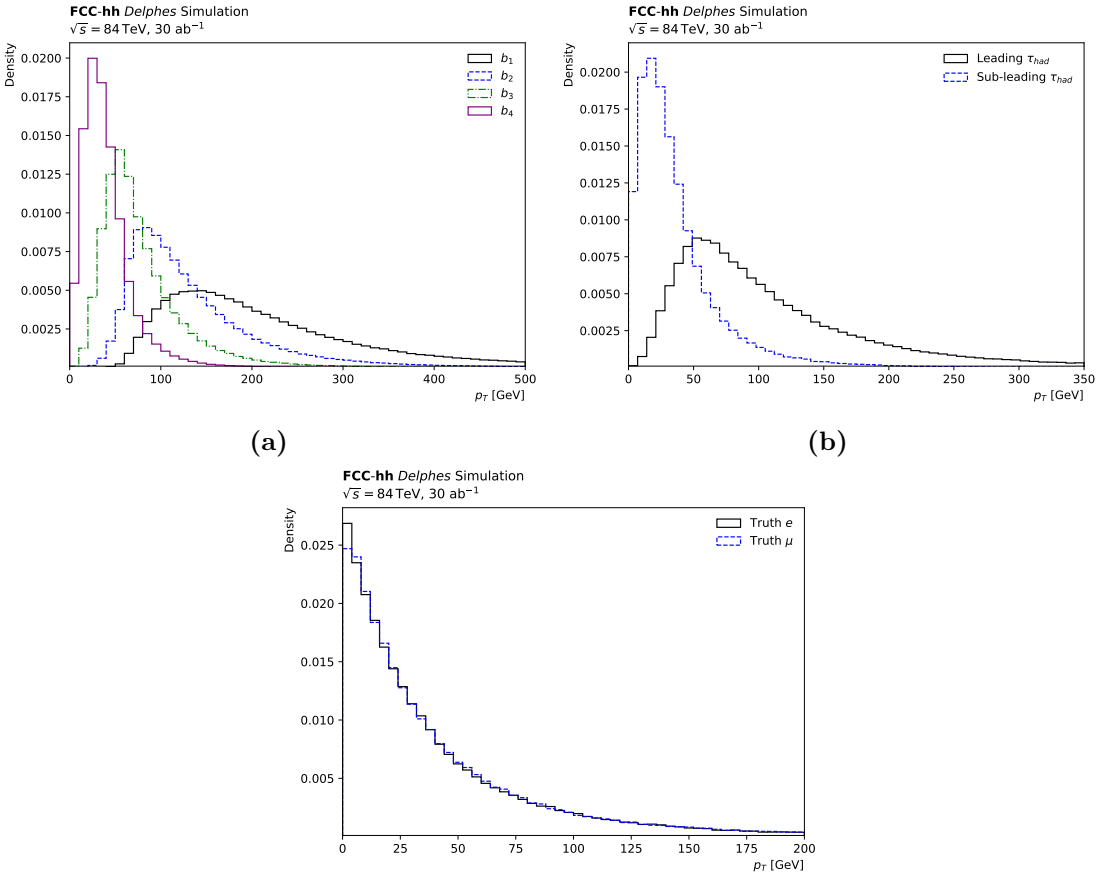


Figure 8.12: Truth transverse momenta distribution of the b-quarks (a), (τ_{had}) (b) and (τ_{lep}) (c) of the HHH signal.

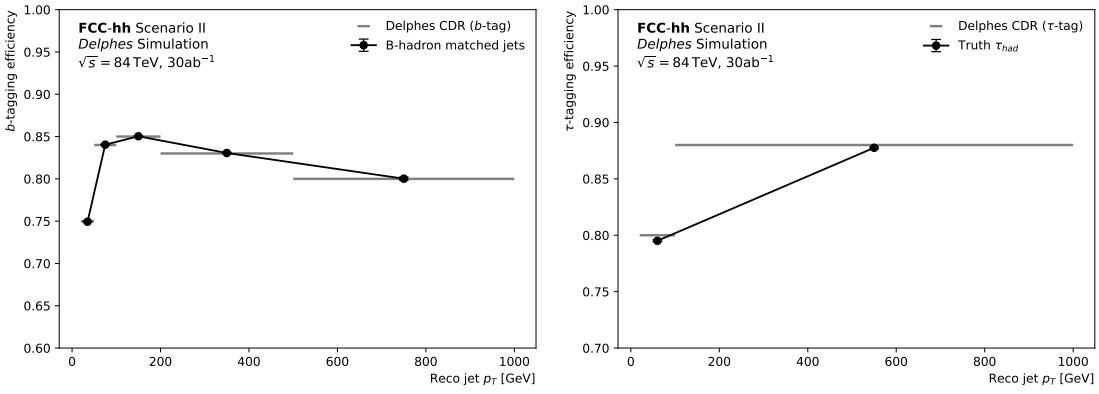


Figure 8.13: Identification efficiency of reconstructed jets with a truth Higgs b-quark (a) and hadronic tau (τ_{had}) (b). The binning is chosen to replicate the granularity of the Delphes efficiency map.

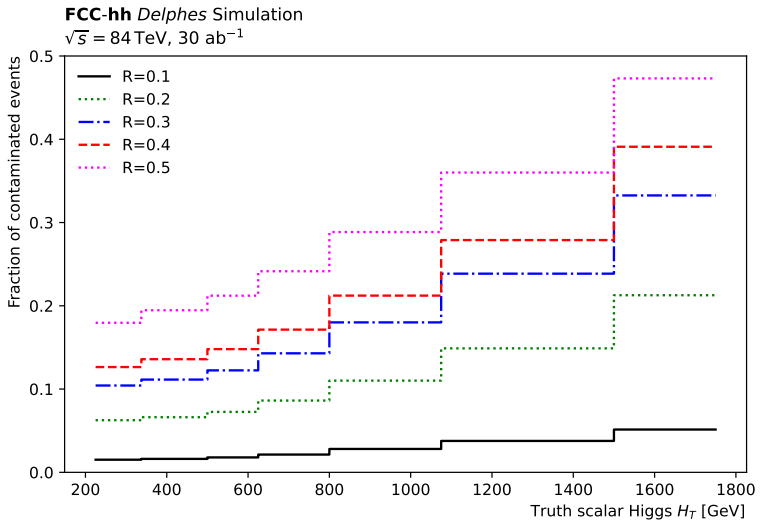


Figure 8.14: Percentage of events containing a jet with both a signal b and τ within $\Delta R \in [0.1, 0.2, 0.3, 0.4, 0.5]$.

simultaneously tagged as a b and τ_{had} candidate. This is investigated because contamination can be problematic for the HHH-signal events, which are highly active and contain 6 partons originating from the hard scatter. With active events, multiple truth partons are more likely to appear spatially close to one another within the detector. Physically, one can imagine production mechanisms akin to diagrams (c) or (d) from Figure 8.1 with a single-Higgs recoiling against the di-Higgs system — producing a collimated HH pair. If this system decays via $HH \to bb\tau\tau$, there is some probability of one reconstructed jet containing two truth-signal partons (b, τ). Again, H_T (h) is a useful proxy for overall event activity, and the percentage contamination as a function of H_T (h) is shown in Figure 8.14. Decay products from a highly boosted Higgs boson will often fall within the same reconstructed jet in the detector. Therefore, the contamination rises from 20% near threshold to around 50% at large boost. To target high matching efficiency whilst minimising potential signal contamination, a matching radius of 0.3 is chosen. Since τ_{had} candidates are more easily faked by b-jets than b-jets are faked by τ_{had}, jets with both tags are treated as τ_{had} candidates and excluded from the b-jet selection.

Lepton Identification

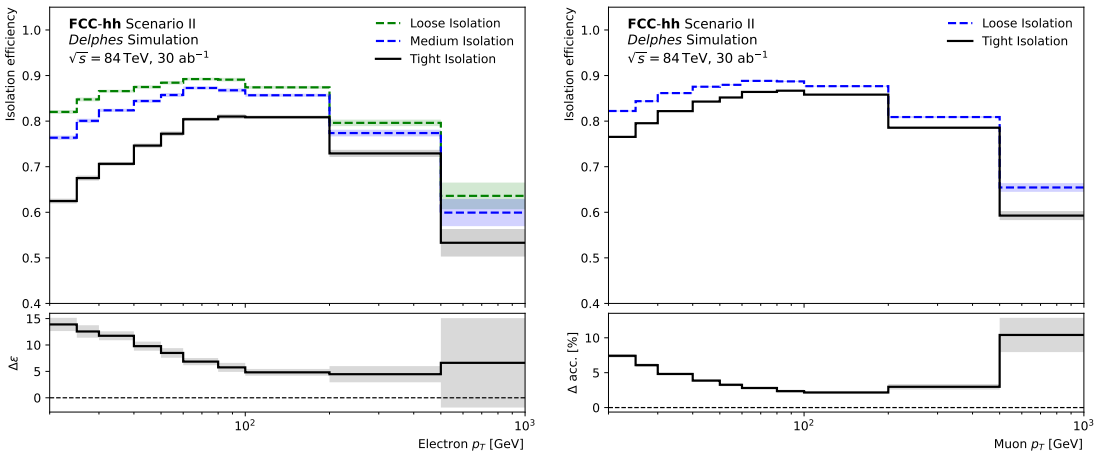
Electron and muon candidates, defined by DELPHES, are required to have a $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 4$. Identification and reconstruction efficiencies are parametrised as functions of p_{T} and η. To remove the contribution of both non-prompt leptons and leptons originating from a b-hadron decay, the following isolation criterion is applied. For each given electron (muon) candidate, all particle flow objects within $\Delta R < 0.3$ are used to define:

$$I(l) = \frac{\sum_{i[\Delta R<0.3]} p_{\mathrm{T}}^{i}}{p_{\mathrm{T}}(l)} \tag{8.12}$$

The original isolation implementation of the FCC-hh experiment for electrons ($I < 0.1$) was deemed too tight, and rejects a large portion of prompt leptons. To perform this re-optimisation, samples of prompt and non-prompt leptons are selected. For prompt leptons, the signal leptons originating from leptonic τ-decays are considered. Leptons originating from semi-leptonic b-hadrons are used for the non-prompt sample. The goal of isolation is to maintain high acceptance of prompt leptons whilst rejecting large portions of non-prompt background. Three working points are defined: Tight, Medium, and Loose, corresponding to cuts of $I(l) < 0.1, 0.2, 0.3$, respectively. The isolation efficiency of each for signal electrons and muons is shown in Figure 8.15, where the low-p_{T} behaviour is targeted for optimisation. The efficiency in the high-p_{T} region begins to decrease due to signal contamination with a τ_{had} jet. The Medium electron working point is chosen, while the default Tight isolation is deemed sufficiently efficient and is retained for muons.

8.5 Event Reconstruction and Preselection

To target the desired $4b2\tau$ final state, a loose preselection is first defined to select the six signal objects: four b-jets and two τ-candidates. Four jets are required to be b-tagged at the medium working point, with loose cuts placed on the transverse momenta of $p_{\mathrm{T}}(b_i)$ greater than $[40, 35, 30, 25]$ for each b-jet, respectively. τ-leptons, with a mass of 1.776 GeV, can decay both leptonically and hadronically, with



(a) **(b)**

Figure 8.15: Isolation efficiency as a function of reconstructed lepton p_{T} for each working point: Tight (black), Medium (Blue) and Loose (Green) for electrons, (a) and Tight (black), Loose (Blue) for muons, (b). The acceptance on the lower panel shows the increase in acceptance for the Medium working point compared with Tight.

branching fractions 35.2% and 64.8% respectively. Therefore, the two dominant channels are when both τs decay hadronically ($\tau_{\mathrm{had}}\tau_{\mathrm{had}}$) and when one decays hadronically with the other leptonically ($\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$). These two channels constitute approximately 88% of the channel statistics.

To target the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ final state, two oppositely charged (OS) τ_{had} candidates are required at the medium working point. Hadronic τ-decays produce 1, 3 or 5 charged tracks — referred to as *prongs*. The branching fraction for 5-prong τ decays is negligible compared to 1- and 3-prong decays and is not considered. Tracks are associated to τ-candidates if a track with transverse momentum above 1 GeV is within $\Delta R < 0.2$ of the τ vertex. To target the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ final state, one hadronically decaying τ-lepton is required alongside an isolated lepton of opposite charge.

The full $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ loose preselection is outlined in Table 8.2. A cutflow showing the application of these cuts on all signal and background processes is shown for $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ in Table 8.3 and Table 8.4 respectively. The identification and reconstruction above p_{T} thresholds for six objects in the final state greatly reduces the event yield, particularly those for the third and fourth b-jet.

Table 8.2: Summary of the loose event Preselection for the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels

Variable	$\tau_{\textbf{lep}}\tau_{\textbf{had}}$	$\tau_{\textbf{had}}\tau_{\textbf{had}}$
Preselection		
$n(b)$	==4	
$p_{\text{T}}(b_1)$ [GeV]	> 40	
$p_{\text{T}}(b_2)$ [GeV]	> 35	
$p_{\text{T}}(b_3)$ [GeV]	> 30	
$p_{\text{T}}(b_4)$ [GeV]	> 25	
$n(\tau_{\text{had}})$	==1	==2
$n(l = e/\mu)$	==1	==0
τ charges	OS(τ_{had}, l)	OS($\tau_{\text{had}}, \tau_{\text{had}}$)
$p_{\text{T}}(\tau_{\text{had1}})$ [GeV]	> 25	
$p_{\text{T}}(\tau_{\text{had2}})$ [GeV]	-	> 25
$p_{\text{T}}(l)$ [GeV]	> 20	-

	n_b^M	n_τ^M	OS($\tau_h\tau_h$)	Lepton Veto
HHH	1.05×10^3	212	189	180
$t\bar{t}b\bar{b}$	1.26×10^7	1.58×10^6	1.38×10^6	1.31×10^6
$t\bar{t}t\bar{t}$	2.40×10^7	8.20×10^5	4.30×10^5	2.84×10^5
$t\bar{t}H$	9.18×10^5	1.18×10^5	1.05×10^5	9.92×10^4
$t\bar{t}Z$	3.72×10^5	4.56×10^4	4.04×10^4	3.83×10^4
HH	2.21×10^5	5.83×10^3	4.68×10^3	4.43×10^3
ZZZ	6.67×10^3	168	139	135

Table 8.3: Cutflow of the triple-Higgs signal and all background processes in the $\tau_{\text{had}}\tau_{\text{had}}$ channel through the Loose Preselection outlined in Table 8.2.

	n_b^M	n_τ^M	$\tau_e\tau_h$ + $\tau_\mu\tau_h$		Loose Preselection
HHH	1.05×10^3	516	55.9	65.3	121
$t\bar{t}b\bar{b}$	1.19×10^8	2.70×10^7	5.83×10^6	6.57×10^6	1.24×10^7
$t\bar{t}t\bar{t}$	2.40×10^7	5.78×10^6	5.21×10^5	5.80×10^5	1.10×10^6
$t\bar{t}H$	9.18×10^5	4.01×10^5	3.14×10^4	3.68×10^4	6.82×10^4
$t\bar{t}Z$	3.72×10^5	1.61×10^5	1.24×10^4	1.47×10^4	2.71×10^4
HH	2.21×10^5	2.86×10^4	1.33×10^3	1.57×10^3	2.89×10^3
ZZZ	6.67×10^3	1.14×10^3	37.0	43.0	79.9

Table 8.4: Cutflow of the triple-Higgs signal and all background processes in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel through the Loose Preselection outlined in Table 8.2.

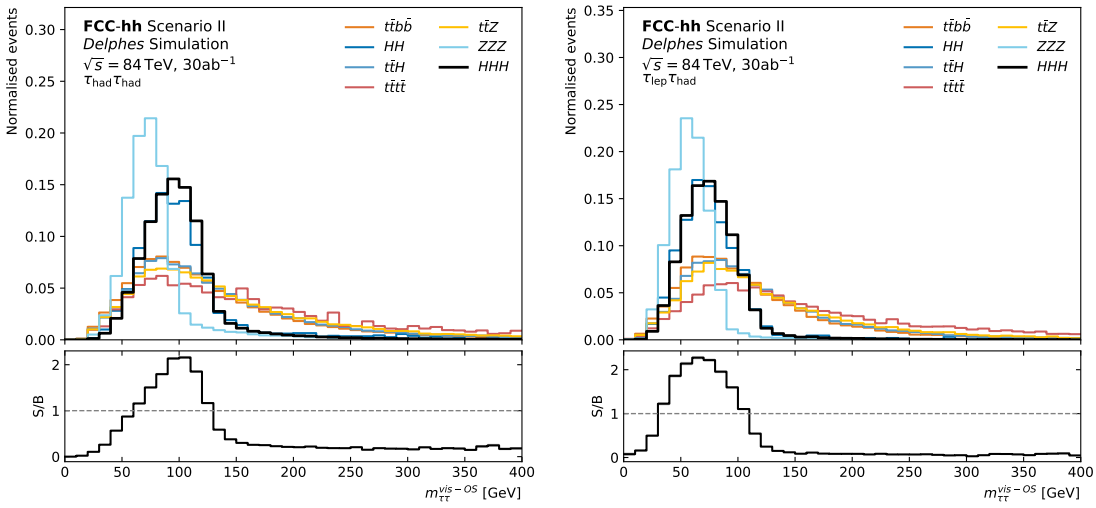
8.6 Di-τ mass reconstruction with MMC

With non-resonant production of τ-leptons via $t\bar{t}$ being a dominant source of background, the accurate reconstruction of the $\tau\tau$ invariant mass ($m_{\tau\tau}$) is a crucial objective. This is challenging because each τ-lepton decay contains at least one neutrino, such that the $\tau_{\text{had}}\tau_{\text{had}}$ channel contains two neutrinos and the $\tau_{\text{lep}}\tau_{\text{had}}$ channel contains three. With the neutrinos escaping detection, the full τ four-momenta cannot be reconstructed, making the corresponding $m_{\tau\tau}$ hard to construct. The simplest attempt at reconstructing the mass of the parent particle is by taking the invariant mass of the visible decay products of each tau, $m_{\tau\tau}^{\text{vis}}$. As shown in Figure 8.16, this is a poor estimate for the signal Higgs mass with a large width. Moreover, both the signal and background peak in a similar region of visible mass — making this feature a poor discriminant.

To improve the accuracy of the $m_{\tau\tau}$ reconstruction, the analysis utilises the Missing Mass Calculator method [236]. Relevant for the overall performance of the algorithm is the purity of the $\tau\tau$ selection. This is assessed as the percentage of events that pass $\tau_{\text{had}}\tau_{\text{had}}$ preselection that are at truth level $\tau_{\text{had}}\tau_{\text{had}}$. For a performant algorithm, one should expect these configurations to yield a valid $m_{\tau\tau}^{\text{MMC}}$ solution. For events that pass the $\tau_{\text{had}}\tau_{\text{had}}$ preselection, the purity is 87%. Of the remaining events, 12.5% are $\tau_{\text{lep}}\tau_{\text{had}}$ events at truth level that contain low-p_{T} leptons that fail isolation. With the poorer isolation performance for low-p_{T} electrons, they are more abundant in this region — constituting 8%. Another jet in the event is then mis-tagged as a τ-jet and passes $\tau_{\text{had}}\tau_{\text{had}}$ preselection. Only 0.5% of these events are $\tau_{\text{lep}}\tau_{\text{lep}}$ at truth level - containing two leptons failing isolation and two mis-tagged τ-jets. A brief description of the algorithm and its calibration for the FCC-hh detector is given below.

8.6.1 The Missing Mass Calculator algorithm

With τ-decays including at least one neutrino (one for hadronic, two for leptonic), there is an unavoidable ambiguity in the source of genuine $p_{\text{T}}^{\text{miss}}$ in the event. To



(a) **(b)**

Figure 8.16: Normalised histogram of visible mass, $m_{\tau\tau}^{\text{vis}}$, for all analysis samples in the $\tau_{\text{had}}\tau_{\text{had}}$ (a) and $\tau_{\text{lep}}\tau_{\text{had}}$ (b) channels.

fully reconstruct the invariant mass, the missing 3-momenta of each τ must be known - constituting 6 unknown components. For leptonic decays, with more than one neutrino, an additional unknown of m_{miss} is included. With the assumption that all $p_{\text{T}}^{\text{miss}}$ in the event arises from the neutrinos, one has 4 equations for the: τ invariant masses ($m_{\tau_{1/2}}$) and $p_{\text{x}}^{\text{miss}}/p_{\text{y}}^{\text{miss}}$. By scanning over missing-momentum vector azimuthal angles, (ϕ_1, ϕ_2), the system is fully specified.

For each point in parameter space, the measured $p_{\text{T}}^{\text{miss}}$ in the x-y plane can be equated with the two missing momentum components:

$$\begin{pmatrix} p_x^{\text{miss}} \\ p_y^{\text{miss}} \end{pmatrix} = \begin{pmatrix} \cos\phi_{\nu_1} & \cos\phi_{\nu_2} \\ \sin\phi_{\nu_1} & \sin\phi_{\nu_2} \end{pmatrix} \begin{pmatrix} p_{\text{T}}(\tau_1^{\text{miss}}) \\ p_{\text{T}}(\tau_2^{\text{miss}}) \end{pmatrix} \tag{8.13}$$

Inverting to find each missing component,

$$\begin{pmatrix} p_{\text{T}}(\tau_1^{\text{miss}}) \\ p_{\text{T}}(\tau_2^{\text{miss}}) \end{pmatrix} = \frac{1}{\sin(\phi_{\nu_2} - \phi_{\nu_1})} \begin{pmatrix} \sin\phi_{\nu_2} & -\cos\phi_{\nu_2} \\ -\sin\phi_{\nu_1} & \cos\phi_{\nu_1} \end{pmatrix} \begin{pmatrix} p_x^{\text{miss}} \\ p_y^{\text{miss}} \end{pmatrix} \tag{8.14}$$

This form assumes that all $p_{\text{T}}^{\text{miss}}$ arises from the neutrinos. However, the two τ-leptons can be back-to-back in the transverse plane — in which the $p_{\text{T}}^{\text{miss}}$ is dominated by detector effects such as jet mismeasurement. Thus, an additional scan over the x and y components of the $p_{\text{T}}^{\text{miss}}$ is performed according to the per-event $\sigma(p_{\text{T}}^{\text{miss}})$, with this system of equations solved at each scanned point.

Assuming on-shell τ-leptons, using the four-vectors of the visible and missing components of the decay:

$$p_\tau^\mu = p_{\text{visible}}^\mu + p_{\text{missing}}^\mu \tag{8.15}$$

Squaring one finds:

$$m_\tau^2 = m_{\text{vis}}^2 + m_{\text{miss}}^2 + 2E_{\text{vis}}E_{\text{miss}} - 2\vec{p}_{\text{vis}} \cdot \vec{p}_{\text{miss}} \tag{8.16}$$

Explicitly expanding out $\vec{p}_{\text{vis}}$ and $\vec{p}_{\text{miss}}$ in terms of $p_{\text{x}}, p_{\text{y}}, p_{\text{z}}$, a quadratic is found in p_{z}:

$$(p_{\text{z}}^{\text{miss}})^2 \left[(p_{\text{z}}^{\text{vis}})^2 - E_{\text{vis}}^2\right] + p_{\text{z}}^{\text{miss}} \left[2Ap_{\text{z}}^{\text{vis}}\right] + \left[A^2 - (E_{\text{vis}}p_{\text{T}}^{\text{miss}})^2 - (E_{\text{vis}}m_{\text{miss}})^2\right] = 0 \tag{8.17}$$

where $A = \frac{1}{2}\left(m_\tau^2 - m_{\text{vis}}^2 - m_{\text{miss}}^2\right) + \vec{p}_{\text{T}}^{\text{miss}} \cdot \vec{p}_{\text{T}}^{\text{vis}}$.

For each point in $(p_{\text{x}}^{\text{miss}}, p_{\text{y}}^{\text{miss}}, \phi_1, \phi_2)$-space, τ four-vectors are formed for any real p_{z} solution and the value of $m_{\tau\tau}$ computed. A per-event histogram is built from all real solutions to the p_z quadratic. The mode of this histogram corresponds to the value of $m_{\tau\tau}^{\text{MMC}}$ for the event. However, not all real solutions are equally likely, and knowledge of the expected τ-decay kinematics can help guide the algorithm towards the kinematically most likely configurations of the visible and missing components, improving upon the unweighted $m_{\tau\tau}^{\text{MMC}}$ estimate. The reweighting takes into account the angular spread of the hypothesised neutrino to the measured visible component, the fraction of energy carried away by the neutrino, the deviation in $p_{\text{x}}^{\text{miss}}$ and $p_{\text{y}}^{\text{miss}}$ and, in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, also the missing mass attributed to the leptonic tau decay. The log-likelihood for each tau is thus,

$$-\log L = -\log\left[\mathcal{P}\left(\theta_{3D}^1(p_\tau^1)\right)\mathcal{P}\left(\theta_{3D}^2(p_\tau^2)\right)\mathcal{P}\left(\Delta p_x^{\text{miss}}\right)\mathcal{P}\left(\Delta p_y^{\text{miss}}\right)\mathcal{P}\left(x_{\text{miss}}^1\right)\mathcal{P}\left(x_{\text{miss}}^2\right)\right] \tag{8.18}$$

where θ_{3D} is the opening angle of the missing and visible components, $\Delta p_{x(y)}^{\text{miss}}$ is the difference between the scanned and measured $p_{\text{x}}^{\text{miss}}(p_{\text{y}}^{\text{miss}})$, and x_{miss} is the fraction of energy carried away by the invisible component in each τ.

8.6.2 MMC Weighting

Accurately deriving each of the input probability density functions for the algorithm enables the most likely solutions to be preferentially up-weighted to improve the performance of `MMC`. To achieve this, a sample of $Z \to \tau\tau$ events is generated using MadGraph 5, partons are showered using Pythia 8 and the Scenario-II FCC-hh detector smearing is applied using Delphes.

Since the collimation of decay products is strongly dependent on the momenta of the parent, the θ distribution is simultaneously fit across all bins in p_τ by a combination of Gaussian and Moyal distributions to define a continuous PDF. The parameters of the Gaussian-Moyal mixture are parametrised as smooth functions of $\ln|p_\tau|$, motivated by the visible–neutrino opening angle shrinking approximately as a power of $|p_\tau|$, $\theta \sim \mathcal{O}(1/\gamma) \propto 1/p_\tau$, so that power-law behaviour in p_τ becomes linear in $\ln p_\tau$. A dependence on $\ln p_\tau$ therefore provides a smooth and flexible description over the wide $|p_\tau|$ range covered by the analysis, while avoiding over-fitting in individual p_τ bins. These fits are performed separately for the leptonic, 1- and 3-prong hadronic decays. Example fits in the same bin of p_τ are shown in Figure 8.17, where the mixture model describes the observed distribution well over the distributions for each of the τ decay modes considered.

8.6.3 Implementation and Results

A grid scan is implemented, where solutions are built for each set of parameters scanned and weighted towards the most-likely mass configurations. The window considered when scanning over azimuthal angles is restricted to ± 0.4 around the visible τ direction. Moreover, the measured mass of the visible τ-candidate often overshoots the true underlying visible mass. Therefore, a hypothesised mass of the ρ-meson (0.8 GeV) and the a-meson (1.2 GeV) are given to 1-prong and 3-prong hadronic τ decays, respectively. The four-momentum of the visible tau is then recomputed according to this mass hypothesis. To enhance the performance of the algorithm, an accurate description of the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is highly beneficial, such that the scan is performed over a suitable window around the measured value.

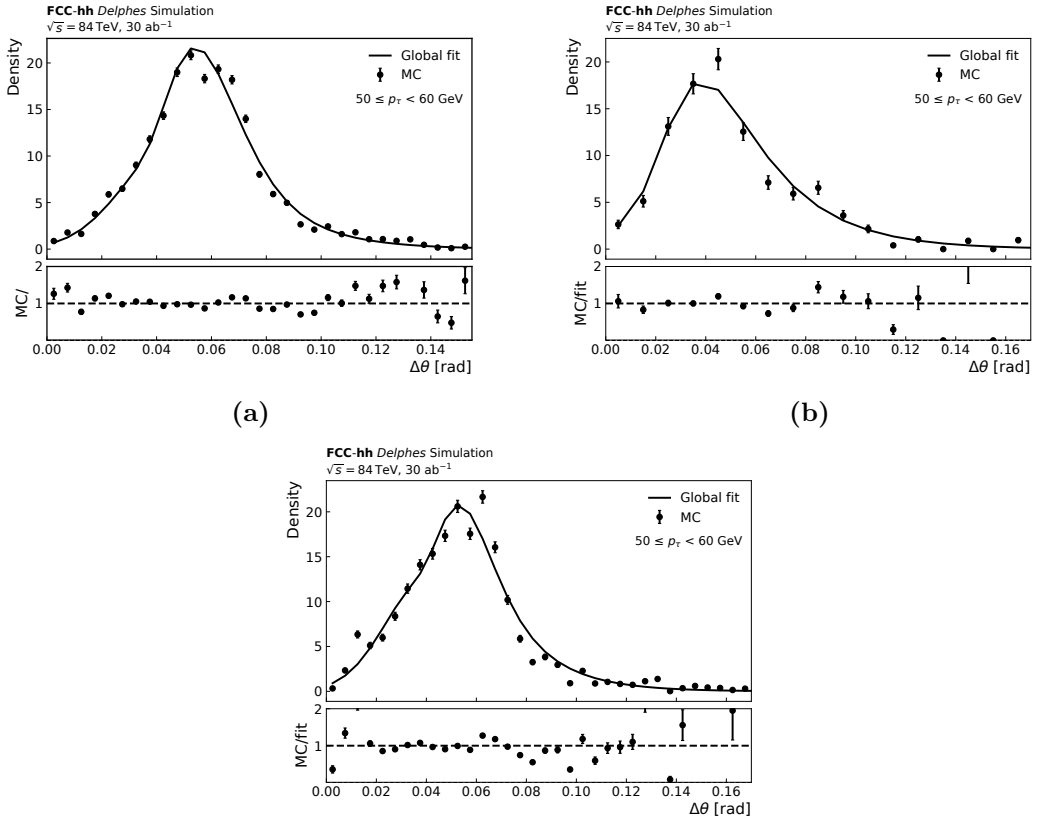


Figure 8.17: Example fit configurations for 1-prong (a), 3-prong (b) and leptonic (c) τ-decays.

A parametrisation of $\sigma(p_{\mathrm{T}}^{\mathrm{miss}})$ is performed to accurately achieve this for the high jet-multiplicity final states considered in this analysis. Taking inspiration from more comprehensive calibrations at current collider experiments, the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is parametrised as the quadratic sum of two contributions: a stochastic term and a per-object term,

$$\sigma\left(p_{\mathrm{T}}^{\mathrm{miss}}\right)^2 = a^2 + b^2 n \tag{8.19}$$

where a is the stochastic term, b is the per-jet resolution and n is the number of jets in the event.

Generator-level $p_{\mathrm{T}}^{\mathrm{miss}}$ is stored, and the true $\sigma(p_{\mathrm{T}}^{\mathrm{miss}})$ is defined as the width of the difference between the measured and true $p_{\mathrm{T}}^{\mathrm{miss}}$: $\left(p_{\mathrm{T}}^{\mathrm{miss,\,meas}} - p_{\mathrm{T}}^{\mathrm{miss,\,true}}\right)$. An example of this distribution is shown in Figure 8.18 in the 5-jet bin, centred around zero with a width of 10.96 ± 0.13 GeV.

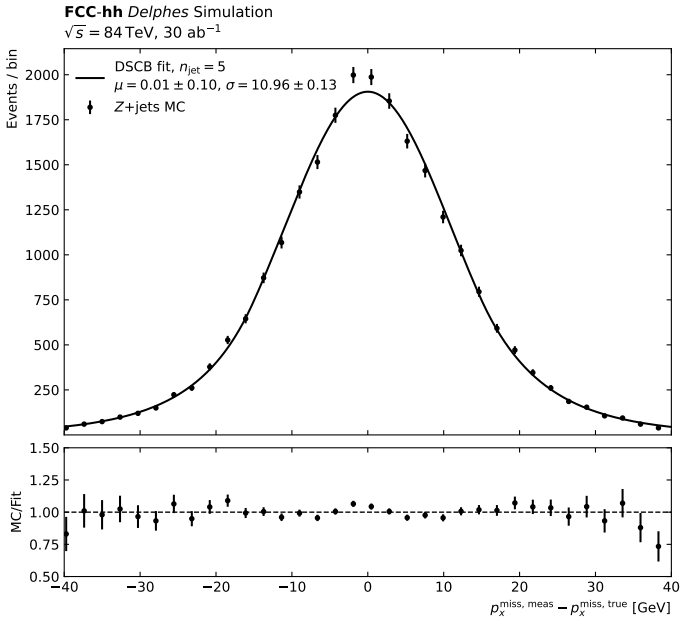
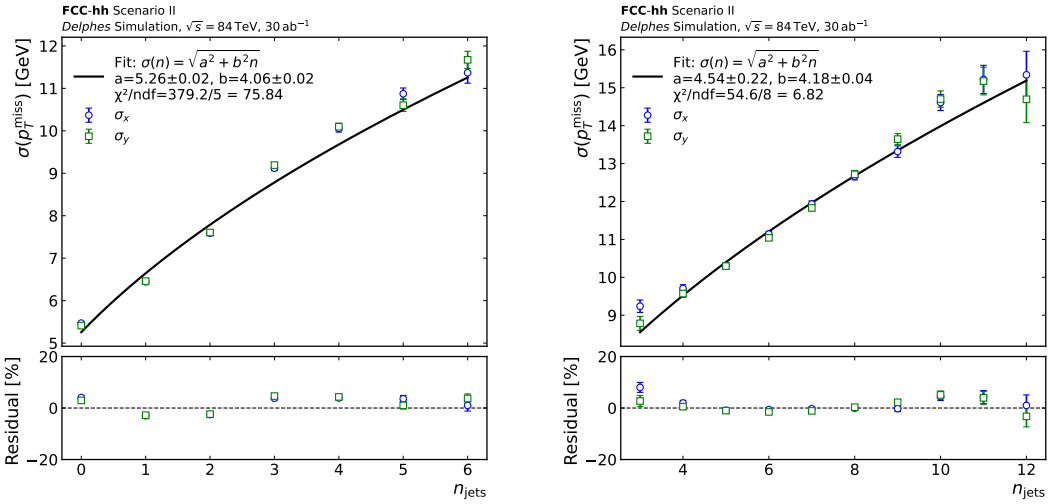


Figure 8.18: Fit to $(p_x^{\text{miss, meas}} - p_x^{\text{miss, true}})$ events by a double-sided crystal ball function for the n_{jets}=5 bin.

The calibration as a function of n_{jets} for Z+jets events is shown in Figure 8.19(a) with an independent fit to HHH events in Figure 8.19(b). Extrapolation tests were performed using both fits and agreement is found across the range of jet multiplicities considered by the analysis, with the final implementation taken from the higher-statistics control sample of Z+jets.

This calibration is performant and captures the average event-level $\sigma(p_{\text{T}}^{\text{miss}})$. However, despite this approach capturing the $\sigma(p_{\text{T}}^{\text{miss}})$ on average, it does not factor in per-event asymmetries between the x and y directions in the transverse plane. For example, events with jets that predominantly lie along x should have $\sigma(p_{\text{x}}^{\text{miss}}) > \sigma(p_{\text{y}}^{\text{miss}})$. The following technique attempts to resolve this and is inspired by the per-object $p_{\text{T}}^{\text{miss}}$ calibrations used within the ATLAS and CMS experiments. Each object considered has a resolution map, according to the DELPHES parametrisation. These are explicitly derived for truth-reconstructed object matches and binned as a function of p_{T} and η. The jet angular spread receives contributions from the finite detector angular resolution, including calorimeter granularity, as well as from the jet clustering procedure and the resolution of the reconstructed constituents.



(a) **(b)**

Figure 8.19: The fitted $\sigma(p_{\mathrm{T}}^{\mathrm{miss}})$ as a function of the number of jets for Z+jets (a) and HHH (b)

With the high-jet-multiplicity final state and jets providing the dominant source of mismeasured $p_{\mathrm{T}}^{\mathrm{miss}}$, only jet mismeasurement is considered. For all jets above 25 GeV and within the central tracking volume ($|\eta| < 4$), the response function is taken from the derived map. This is then projected onto the basis parallel and perpendicular to the measured $p_{\mathrm{T}}^{\mathrm{miss}}$:

$$\sigma(p_{\parallel}^{\mathrm{miss}})^2 = \sum_i \left(\sigma_i \cos\Delta\phi\right)^2, \quad \sigma(p_{\perp}^{\mathrm{miss}})^2 = \sum_i \left(\sigma_i \sin\Delta\phi\right)^2$$

where σ_i is the average p_{T} resolution for jet i and $\Delta\phi$ is the difference in azimuthal angle between the jet and $p_{\mathrm{T}}^{\mathrm{miss}}$.

This method captures per-event asymmetries in the transverse plane, which, when projected back onto the $x-y$ plane, show non-zero differences between directions, compared with the previous implementation, which resulted in identical x, y resolution in each event. Moreover, this method enables the study of $S_{\mathrm{obj}}(p_{\mathrm{T}}^{\mathrm{miss}})$, similar to more sophisticated methods deployed by the ATLAS experiment [135] to enhance rejection of fake-$p_{\mathrm{T}}^{\mathrm{miss}}$ backgrounds.

8.6.4 $m_{\tau\tau}^{\text{MMC}}$ performance

Using the more dynamic $\sigma(p_{\text{T}}^{\text{miss}})$ calibration as described above, a histogram is built with each event entering with weight given by Equation 8.18, with the mode of this histogram being the $m_{\tau\tau}^{\text{MMC}}$ estimate. Distributions for all analysis samples for $m_{\tau\tau}^{\text{MMC}}$ are shown in Figure 8.20. The improvement relative to $m_{\tau\tau}^{\text{vis}}$ is clear, with $m_{\tau\tau}^{\text{MMC}}$ peaking around m_H for the signal and HH background, and around m_Z for the ZZZ background. There is no assumption placed about the shared kinematics of the two τ leptons, and therefore the non-resonant $\tau\tau$ backgrounds remain flat in $m_{\tau\tau}^{\text{MMC}}$. The performance of the algorithm is similar between the two analysis channels, with the missing-mass weighting applied in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and fewer jets in the final state that could distort the $p_{\text{T}}^{\text{miss}}$ estimate, leading to a marginally sharper peak compared with $\tau_{\text{had}}\tau_{\text{had}}$. The impact of reweighting is shown for the $\tau_{\text{lep}}\tau_{\text{had}}$ channel in Figure 8.21, where the resolution of the resonant backgrounds is improved around m_H and m_Z for (HHH, HH) and ZZZ respectively. The weighting retains the flat shape of the non-resonant backgrounds, with a small preference towards more massive $m_{\tau\tau}^{\text{MMC}}$ values within the overflow bin.

High rejection of non-resonant backgrounds is achieved, with 28% of $t\bar{t} + b\bar{b}$ events yielding no valid solution, compared with 3% of triple-Higgs events. Even for the scenario in which $t\bar{t} + b\bar{b}$ is produced with the W decaying via $\tau\nu$, the additional, often higher momentum neutrino from the W further distorts the $p_{\text{T}}^{\text{miss}}$ assumption and yields worse solution rates. A number of configurations that yield no solution could be recovered via wider scans in $\sigma(p_{\text{T}}^{\text{miss}})$; however, since this impacts only background processes, it proves a useful rejection tool whilst keeping the computational time low for genuine $X \to \tau\tau$ final states. Moreover, the shape of $m_{\tau\tau}$ remains flat - enabling loose mass preselections of $50 < m_{\tau\tau}^{\text{MMC}} < 300$ to be defined, exploiting the narrower peaks in $m_{\tau\tau}^{\text{MMC}}$. This selection improves the signal-to-background purity by 26%, whilst remaining a discriminating feature to be used for machine learning to further enhance signal–background separation.

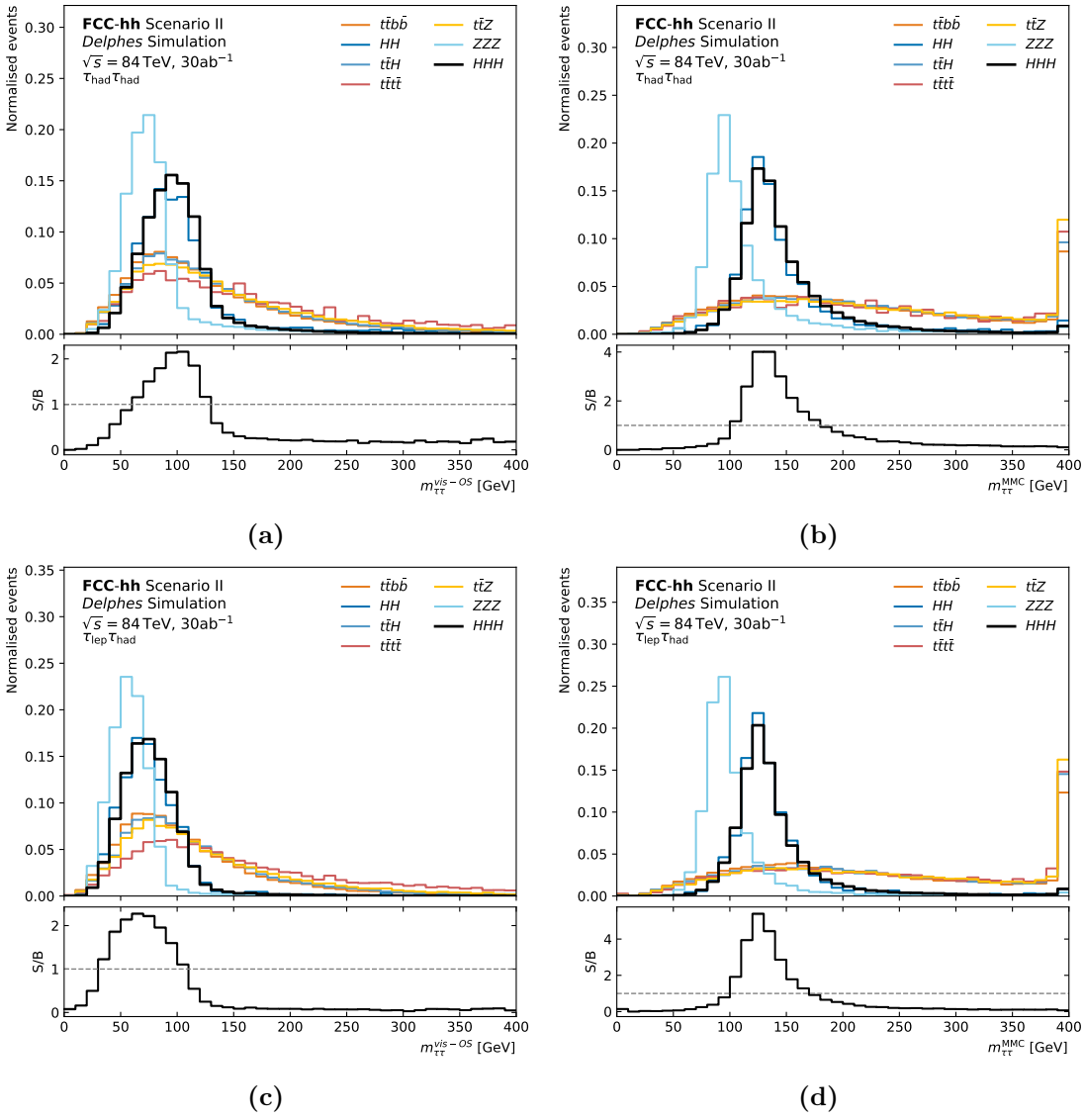
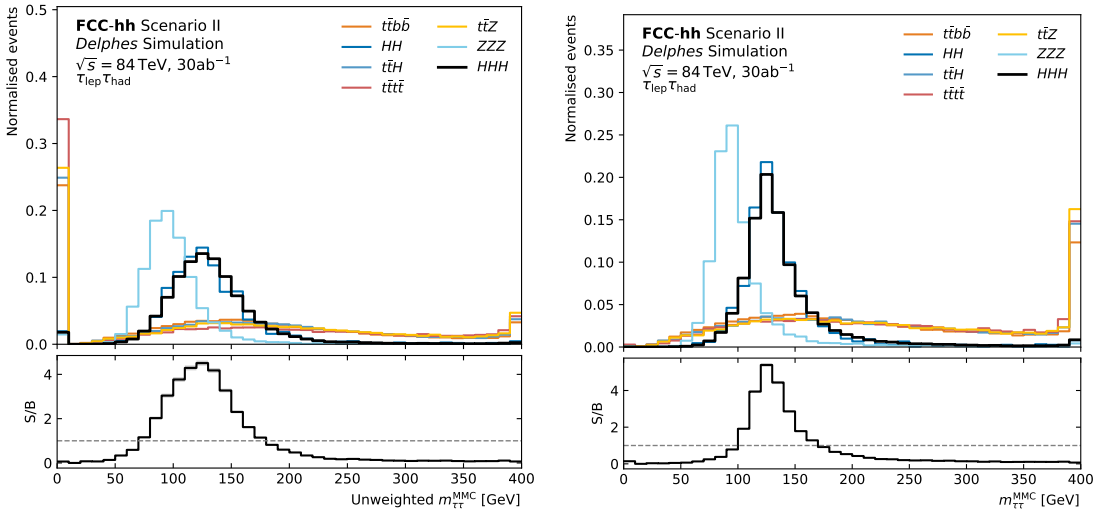


Figure 8.20: Normalised histogram of visible mass, $m_{\tau\tau}^{\text{vis}}$, and Missing Mass Calculator mass, $m_{\tau\tau}^{\text{MMC}}$, for all analysis samples in the $\tau_{\text{had}}\tau_{\text{had}}$ (a, b) and $\tau_{\text{lep}}\tau_{\text{had}}$ (c, d) channels. The ratio between the HHH-signal and weighted background is shown below. An overflow bin is set at 400 GeV.



(a) **(b)**

Figure 8.21: Comparison of unweighted to weighted Missing Mass Calculator di-tau mass ($m_{\tau\tau}^{\mathrm{MMC}}$) estimates in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel. The ratio between the HHH-signal and weighted background is shown below. An overflow bin is set at 400 GeV.

8.7 $HH \rightarrow 4b$ Pairing

The reconstruction of the two resonant $H \rightarrow b\bar{b}$ decays in the signal is an effective discriminant against the non-resonant backgrounds. There are three possible $b\bar{b}$ pairing combinations from the four final-state b-jets. Defining a metric to pair these b-jets is therefore beneficial to provide approximate Higgs candidate kinematics. Both the angular and momentum response of the detector dictate the achievable resolution on the invariant mass of a decaying particle,

$$m_{ij}^2 \sim 2p_{\mathrm{T}}^i p_{\mathrm{T}}^j \left(\cosh \Delta\eta - \cos \Delta\phi\right) \tag{8.20}$$

and the task of pairing the $4b$-jets can be phrased as a likelihood maximisation problem. The likelihood function can be comprised of a number of expected kinematic features of a Higgs decay, often the invariant mass of the daughters. To provide optimal target distributions, the leading and sub-leading $H \rightarrow b\bar{b}$ mass distributions, where the b-jets are matched to truth b-quarks, are shown in Figure 8.22. The Higgs peak is not cleanly reconstructed, even in the scenario of correctly matched b-jets, since there is an inevitable loss in energy per jet.

Assuming the correctly-paired reconstructed invariant mass distribution follows a Gaussian, the likelihood may be written as:

$$\mathcal{L} = \frac{1}{\sqrt{2\pi}\sigma_{H_1}} \exp\left(\frac{-\left(m_{ij} - m_{H_1}\right)^2}{2\sigma_{H_1}^2}\right) \frac{1}{\sqrt{2\pi}\sigma_{H_2}} \exp\left(\frac{-\left(m_{kl} - m_{H_2}\right)^2}{2\sigma_{H_2}^2}\right)$$

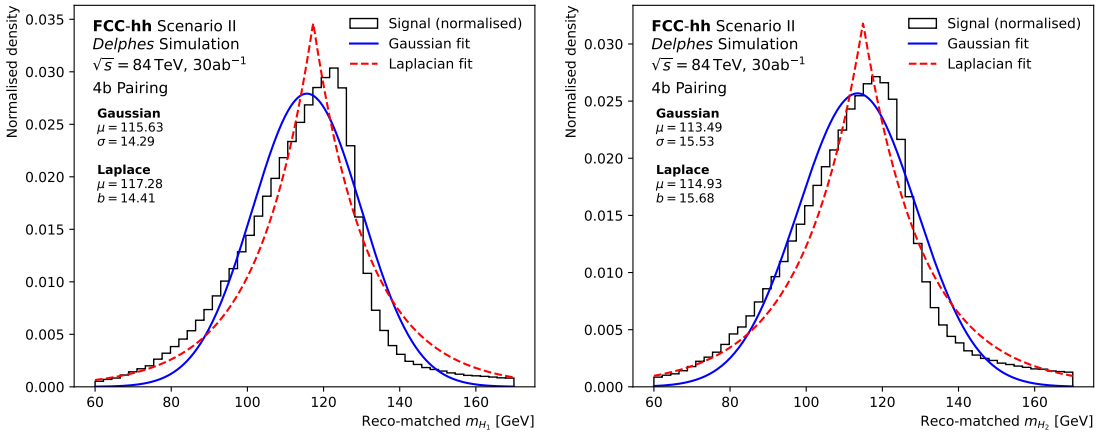
where σ is the width of the invariant mass distribution, m_{ij} is the invariant mass of the pair ij. Pairs are then chosen that maximise the likelihood function, which for the case of a Gaussian corresponds to minimising the following metric:

$$\chi^2_{ij,kl} = \left(m_{ij} - m_{H_1}\right)^2 + \left(m_{kl} - m_{H_2}\right)^2$$

where m_{H_1} and m_{H_2} may be taken as the mode of the truth-matched leading and sub-leading Higgs m_{bb} distributions and the widths taken to be equal.

Different choices of likelihood function lead to different χ^2-minimisations, such as a Laplace distribution giving rise to:

$$\chi^2_{ij,kl} = \left|m_{ij} - m_{H_1}\right| + \left|m_{kl} - m_{H_2}\right|.$$



(a) **(b)**

Figure 8.22: The leading (a) and sub-leading (b) Higgs masses with Gaussian and Laplace fits overlaid, corresponding to the likelihood function of `square-mass` and `abs-mass` pairing respectively.

Akin to the dark Higgs analysis, angular pairing methods, seeking to exploit the proximity of b-jet daughters from the Higgs decay, can be used, where the pairs are chosen according to

$$\min(\Delta R_{\max}(b,b)) = \min_{i,k,l,m}(\max(\Delta R(i,k), \Delta R(l,m))),$$

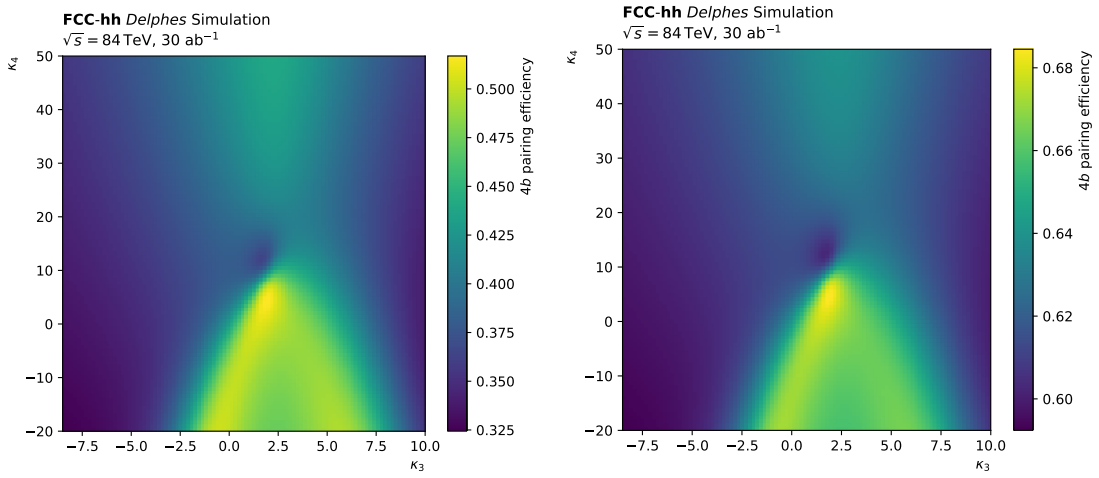
where i, k, l, m are the four selected b-jets in the final state.

As a baseline, the random-chance correct pairing assignment is $\varepsilon_{\text{Rand}} = \frac{1}{3}$. Comparisons against this baseline are shown in Table 8.5, where the two mass-based algorithms are most performant, as demonstrated by the sharper peak in the mass distributions shown in Figure 8.25 (a), (b). Around the peak, these sculpt into either Laplace-like (triangular peak) or Gaussian-like (smooth peak) features for `abs-mass` pairing and `square-mass` pairing, respectively. Away from the peak, the low-mass shoulder can be understood via the phase-space effects preferentially creating lower-mass pairs, shown by the random baseline distribution for the HHH signal and $t\bar{t} + b\bar{b}$.

Moreover, kinematic variations across the (κ_3, κ_4) parameter space result in efficiency differences for each pairing strategy. The pairing efficiency across the plane for the angular and the absolute-mass algorithm are shown in Figure 8.23. Two representative points on the (κ_3, κ_4) plane are chosen to showcase the impact

Algorithm	Pairing efficiency
Random	0.33
Angular	0.48
Abs-mass	0.66
Square-mass	0.66

Table 8.5: Pairing efficiency for $HH \to 4b$ events for the three pairing strategies. The random baseline corresponds to the combinatorial chance of selecting the correct pairing out of the three possible dijet combinations for four b-jets.

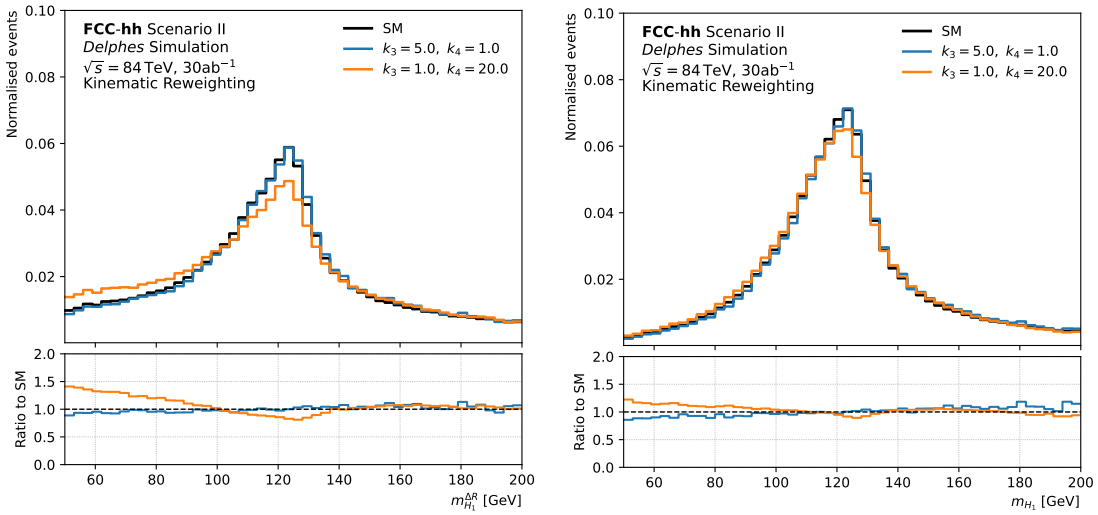


(a) **(b)**

Figure 8.23: Pairing efficiencies across the (κ_3, κ_4) plane for angular (a) and absolute-mass (b) pairing strategies.

on the resolution in Figure 8.24. Angular pairing showcases wider variations across the plane, since the algorithm is optimally performant for moderately high-p_{T} Higgs decays, whereby the decay products are spatially closer in the detector. These regions also align with larger contributions from the triangle configuration (d) in Figure 8.1, whereby a di-Higgs system recoils against a single-Higgs — leaving a potential $4b$ system more often within the same hemisphere of the detector.

Given the non-resonant nature of b-jet pairs produced in the dominant $t\bar{t} + b\bar{b}$ background process, avoiding sculpting of the background mass distributions towards the Higgs mass is greatly beneficial. Therefore, evaluating the most performant pairing strategy must also consider the background. The Higgs candidate masses for the $t\bar{t} + b\bar{b}$ process are shown for each pairing strategy in Figure 8.25 (c),(d). Angular-based pairing places no assumptions on the specific mass of a decaying



(a) **(b)**

Figure 8.24: Kinematically reweighted leading Higgs candidate mass for angular (a) and absolute-mass (b) pairing strategies at the SM, $(\kappa_3, \kappa_4) = (5, 1)$ and $(\kappa_3, \kappa_4) = (1, 20)$.

resonance and shows significantly reduced sculpting when compared with mass-based pairing strategies. Notably, it shifts the invariant mass away from the Higgs peak for $t\bar{t} + b\bar{b}$ since angular pairing more often picks spatially closer b-jets from each of the two bb-pairs: $t\bar{t} \to b\bar{b}W^+W^-$ and $g \to b\bar{b}$. Random assignments often overshoot the mass, with well-separated candidates mixed between the two bb-systems, as can be understood via Equation 8.20. Angular pairing leads to the best performance, with the moderate signal efficiency allowing loose mass preselections to be defined whilst minimising sculpting of the dominant $t\bar{t} + b\bar{b}$ background.

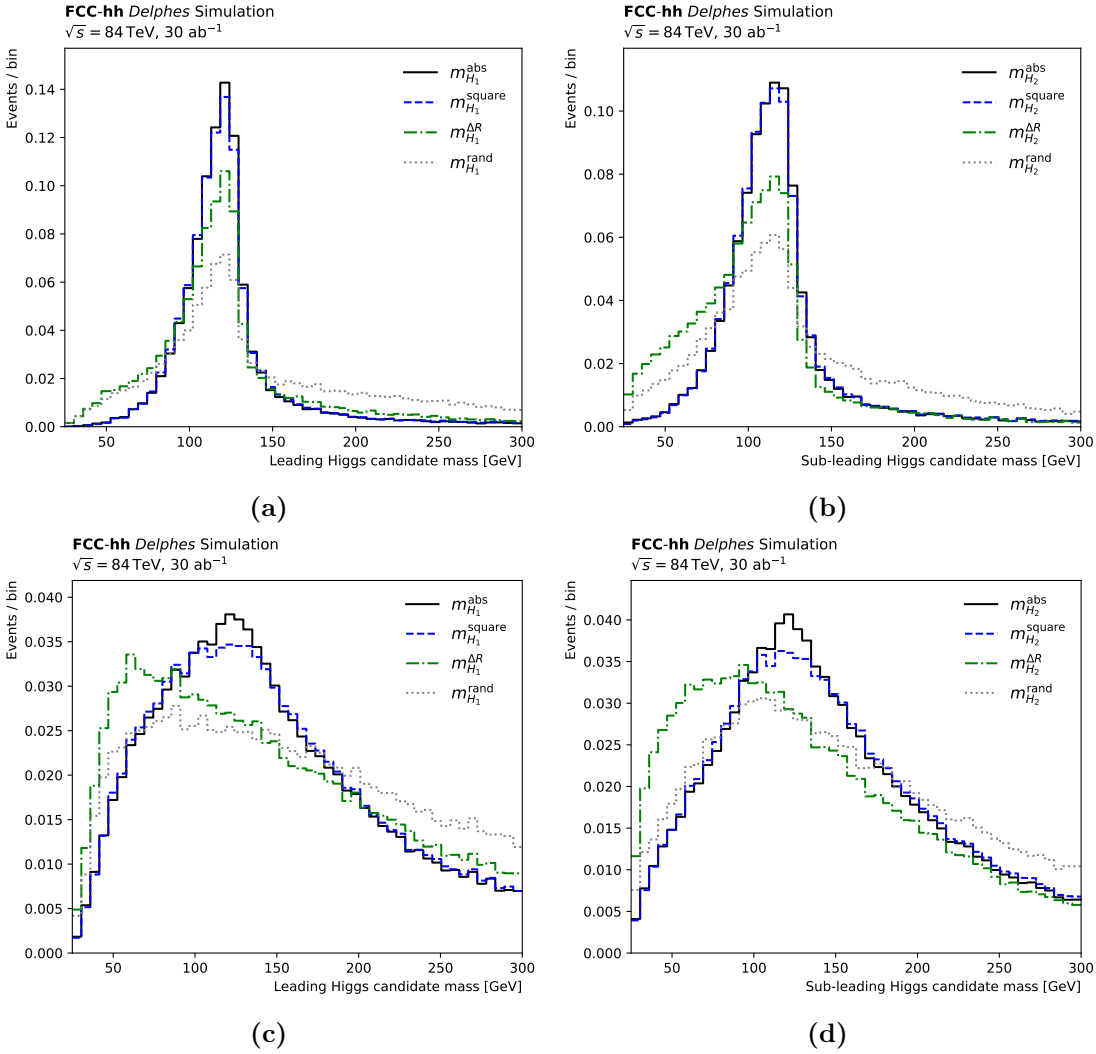


Figure 8.25: Leading and sub-leading Higgs masses in the HHH signal (a), (b) and $t\bar{t} + b\bar{b}$-background (c), (d).

8.8 Event Preselection

A tight preselection can be defined, building on the Loose preselection of Tables 8.3, 8.4, targeting the reconstruction of three resonant Higgs boson decays in $m_{\tau\tau}^{\text{MMC}}$, m_{H_1} and m_{H_2}. The tight preselection is summarised in Table 8.6, where, to retain the small number of signal events, relatively conservative mass-window cuts are placed on the three masses.

Yield tables with all signal and background processes after the tight preselection in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels are shown in Tables 8.7 8.8. Notably, the rejection of $t\bar{t}+b\bar{b}$ as a result of the MMC algorithm is powerful and rejects a significant

Table 8.6: Summary of the tight preselection for the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels

Tight Preselection		
Variable	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
Loose Preselection	Pass	
m_{H_1}	$40 < m_{H_1} < 175$	
m_{H_2}	$20 < m_{H_2} < 160$	
MMC	Pass	
$m_{\tau\tau}^{\text{MMC}}$	$50 < m_{\tau\tau}^{\text{MMC}} < 300$	

portion of the $t\bar{t} + b\bar{b}$ background while retaining close to 95% of the signal.

	Loose Preselection	m_{H_1}	m_{H_2}	Pass MMC	$m_{\tau\tau}^{\text{MMC}}$
HHH	180	145	132	128	124
$t\bar{t}b\bar{b}$	1.31×10^6	6.90×10^5	4.41×10^5	3.04×10^5	2.31×10^5
$t\bar{t}t\bar{t}$	2.84×10^5	1.11×10^5	6.46×10^4	4.18×10^4	2.98×10^4
$t\bar{t}H$	9.92×10^4	6.63×10^4	4.80×10^4	3.30×10^4	2.42×10^4
$t\bar{t}Z$	3.83×10^4	2.42×10^4	1.60×10^4	1.08×10^4	7600
HH	4430	2740	2080	1910	1830
ZZZ	135	118	112	108	105
$\frac{S}{\sqrt{B}}$	0.14	0.15	0.17	0.20	0.23

Table 8.7: Cutflow of the triple-Higgs signal and all background processes in the $\tau_{\text{had}}\tau_{\text{had}}$ channel through the tight preselection outlined in Table 8.6.

	Loose Preselection	m_{H_1}	m_{H_2}	Pass MMC	$m_{\tau\tau}^{\text{MMC}}$
HHH	121	99.1	90.5	88.8	86.6
$t\bar{t}b\bar{b}$	1.24×10^7	6.69×10^6	4.30×10^6	2.86×10^6	2.05×10^6
$t\bar{t}t\bar{t}$	1.10×10^6	4.34×10^5	2.52×10^5	1.67×10^5	1.09×10^5
$t\bar{t}H$	6.82×10^4	4.55×10^4	3.26×10^4	2.45×10^4	1.66×10^4
$t\bar{t}Z$	2.71×10^4	1.72×10^4	1.13×10^4	8.28×10^3	5.39×10^3
HH	2.89×10^3	1.77×10^3	1.33×10^3	1.30×10^3	1.27×10^3
ZZZ	79.9	71.5	68.5	67.5	66.5
$\frac{S}{\sqrt{B}}$	0.03	0.04	0.04	0.05	0.06

Table 8.8: Cutflow of the triple-Higgs signal and all background processes in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel through the tight preselection outlined in Table 8.6.

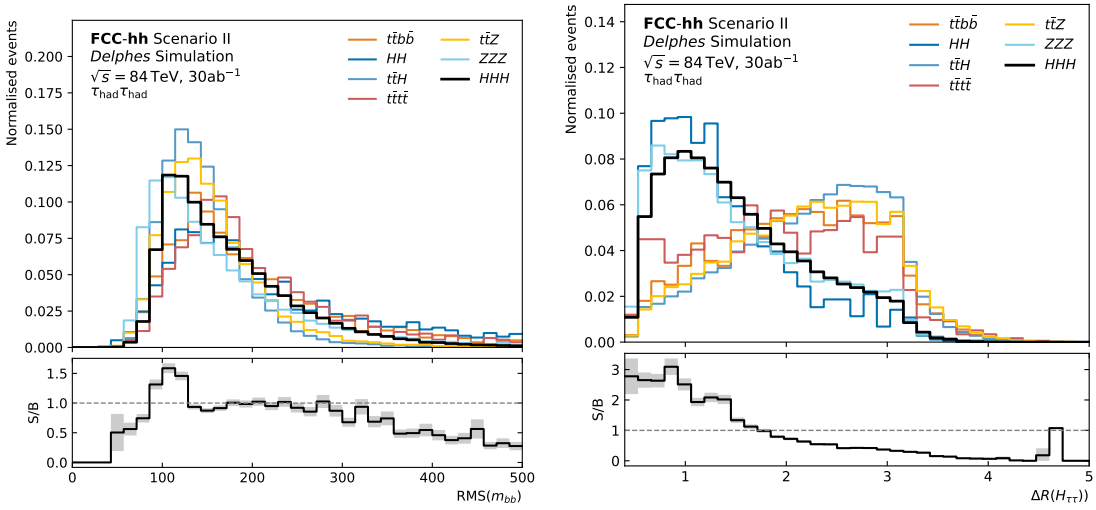
8.9 Preselection and Machine Learning

Although the preselection is effective at eliminating a large portion of the SM background, the HHH signature is of low cross section and benefits from machine-learning classifiers to enhance sensitivity. For each of the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels, a NN classifier ($\text{NN}_{\text{lephad}}$, $\text{NN}_{\text{hadhad}}$) is trained to discriminate between the HHH signal and SM backgrounds, and a cut is placed on the resulting NN output score to define the signal region in that channel.

Input features to the NNs are chosen from variables related to the kinematics of the b-jets, τ-leptons, reconstructed Higgs boson candidates, $p_{\text{T}}^{\text{miss}}$ and event-wide variables. The full set of input features is listed in Table 8.9, with two examples of particularly discriminating variables, $\text{RMS}(m_{bb})$ and $\Delta R(\tau, \tau)$ shown in Figure 8.26. The reconstruction of two on-shell $H \to b\bar{b}$ systems leads to the signal peaking around the Higgs mass with a wider spread for the background, which all have fewer than two such systems. Moreover, the two τ-leptons will be spatially close to one another in the signal, whereas in $t\bar{t} + b\bar{b}$ events they originate from separate W-boson decays within the two top-quark systems.

Feature	Description
$\Delta R(\tau_1, \tau_2)$	Angular distance between the two reconstructed τ candidates.
RMS(ΔR_{jj})	Root-mean-square of ΔR values over selected jet pairs in the event.
RMS(m_{jj})	Root-mean-square of dijet invariant masses m_{jj} over selected jet pairs.
RMS($\Delta\eta_{jj}$)	Root-mean-square of pseudorapidity separations $\Delta\eta$ over selected jet pairs.
$m_T^{\min}(b, \vec{p}_T^{\,\text{miss}})$	Minimum transverse mass between $\vec{p}_T^{\,\text{miss}}$ and the selected b-jets
$S_{\text{obj}}(p_{\text{T}}^{\text{miss}})$	Estimated missing transverse momentum significance, MET/σ_{MET}.
m_{H_1}	Invariant mass of the leading-p_{T} Higgs ($b\bar{b}$) candidate.
m_{H_2}	Invariant mass of the sub-leading p_{T} Higgs ($b\bar{b}$) candidate.
$m_{\tau\tau}^{\text{MMC}}$	Reconstructed MMC mass.
$m_{\tau\tau}^{\text{vis, OS}}$	Visible invariant mass of the opposite-sign τ pair.
$p_{\text{T}}^{\text{miss}}$	Missing transverse momentum magnitude, $\lvert\vec{p}_T^{\,\text{miss}}\rvert$.
$p_T(b_1)$	Transverse momentum of the leading b-tagged jet.
$p_T(b_2)$	Transverse momentum of the subleading b-tagged jet.
$p_T(b_3)$	Transverse momentum of the third b-tagged jet.
$p_T(b_4)$	Transverse momentum of the fourth b-tagged jet.
$p_T(\tau_1)$	Transverse momentum of the leading τ candidate.
$p_T(\tau_2)$	Transverse momentum of the subleading τ candidate.
$N_{\text{jets}}^{\text{sel}}$	Number of selected jets passing the analysis jet requirements.
Aplanarity	A measure of the extent to which the event momentum lies out of a plane, 0 for planar events and reaching 1/2 for a fully isotropic topology.
Sphericity	A measure of how isotropic the visible momentum is distributed in the event, ranging from 0 for a pencil-like topology to 1 for a fully spherical event.
Thrust	Quantifies how strongly the event is aligned along a single axis, taking values near 1 for pencil-like events and smaller values for more isotropic momentum configurations.

Table 8.9: Kinematic features used in the machine learning training.



(a) **(b)**

Figure 8.26: Example distributions of kinematic variables important to each NN for the HHH signal and each SM background: (a) the root-mean-square of m_{bb} between all possible pairs of b-jets in the $\tau_{\text{had}}\tau_{\text{had}}$ channel and (b) the ΔR between the τ_{had} candidates in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The tight preselection is applied. Histograms are normalised to unit area.

8.9.1 Training Performance

The NNs are implemented using PyTorch [237]. The networks are trained on the SM HHH signal and SM background events that pass the tight preselection. To avoid overtraining and benefit from the full set of available simulated samples, orthogonal training and testing/validation datasets are used. These are defined through a k-fold splitting procedure with $k = 2$ [205]. The `Optuna` framework [238] is used to optimise the hyperparameters in each network, with the mean AUC between the two folds in the validation sets used to quantify performance within the search space defined in Table 8.10[1].

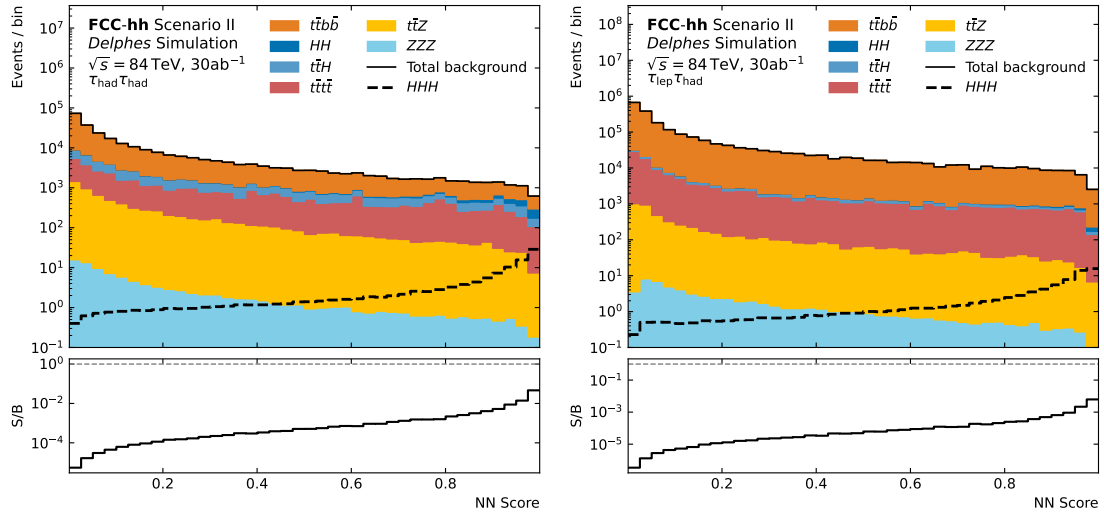
Quantity	Search space
Hidden layer sizes	$(256, 128, 64)$, $(256, 128)$, $(128, 128, 64)$, $(128, 64)$, $(128, 128)$, $(64, 64, 32)$
Weight decay	$x \in [-8, -2]$, with $\lambda = 10^x$
Learning rate	$x \in [-4, -1.5]$, with $\eta = 10^x$
Batch size	256, 512, 1024, 2048
Activation	`ReLu`, `Leaky ReLu`, `ELu`, `tanh`, `GeLu`
Early-stopping patience	$[12, 30]$
Adam β_1	$[0.85, 0.95]$
Adam β_2	$[0.995, 0.9999]$
Adam ϵ	$x \in [-9, -7]$, with $\epsilon = 10^x$
Dropout	$[0, 0.2]$

Table 8.10: Hyperparameter search space and optimisation objective used for neural network training.

For both NNs, an architecture with three hidden layers is chosen, with 128 hidden nodes in the first layer, 128 in the second and 64 in the third. Gaussian error linear unit (GeLU) activation is used for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, and rectified linear unit (ReLU) activation is used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. A standard Binary Cross-Entropy loss function is used. For $\text{NN}_{\text{hadhad}}$ ($\text{NN}_{\text{lephad}}$), a batch size of 2048 (1024) is used, a learning rate of 0.000762 (0.00108) and an early-stopping patience of 15 (21) epochs.

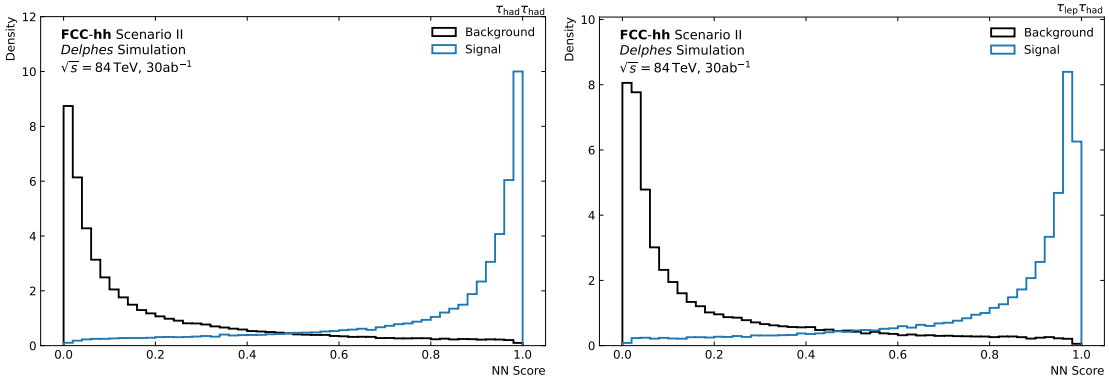
Both networks efficiently separate signal and background, as shown by the score distribution in Figure 8.27, with an AUC of 0.928 (0.931) for the $\tau_{\text{had}}\tau_{\text{had}}(\tau_{\text{lep}}\tau_{\text{had}})$ channel. A normalised plot of signal against weighted background is shown in Figure 8.28. The importance-ordered Shapley beeswarm plot is shown in Figure 8.29.

[1]The networks are not strongly dependent on the choice of hyperparameters.



(a) **(b)**

Figure 8.27: Distribution of the (a) $\mathrm{NN}_{\mathrm{hadhad}}$ and (b) $\mathrm{NN}_{\mathrm{lephad}}$ scores for the signal and background events. The lower panel shows the signal-to-background ratio in each bin.



(a) **(b)**

Figure 8.28: Normalised histograms of NN output score for signal and background processes in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ (a) and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ (b) channels.

In particular, there is significant discriminating power in the reconstructed Higgs boson candidate masses and the spatial separation between the $\tau\tau$-pair, as the SM backgrounds lack the triply-resonant structure of the HHH signal. `MMC` provides strong discriminatory power within the tight preselection, retaining large portions of signal while removing clearly incompatible background $\tau\tau$ pairs. It is similarly helpful when utilised within the network, exploiting correlations with other event features, particularly $\Delta R(\tau,\tau)$, to enhance separation.

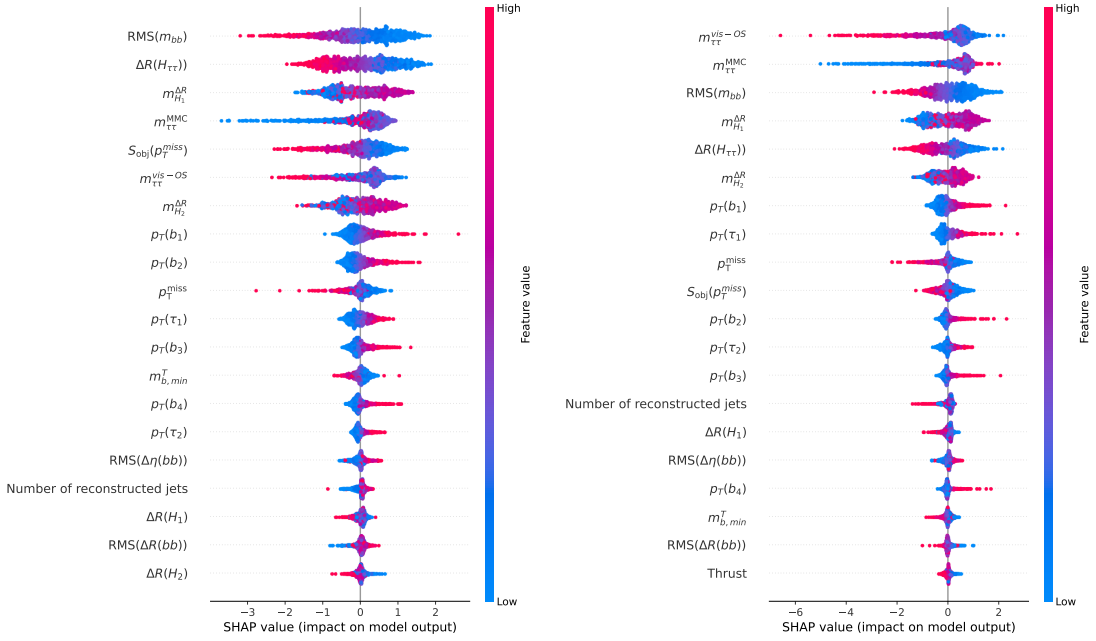
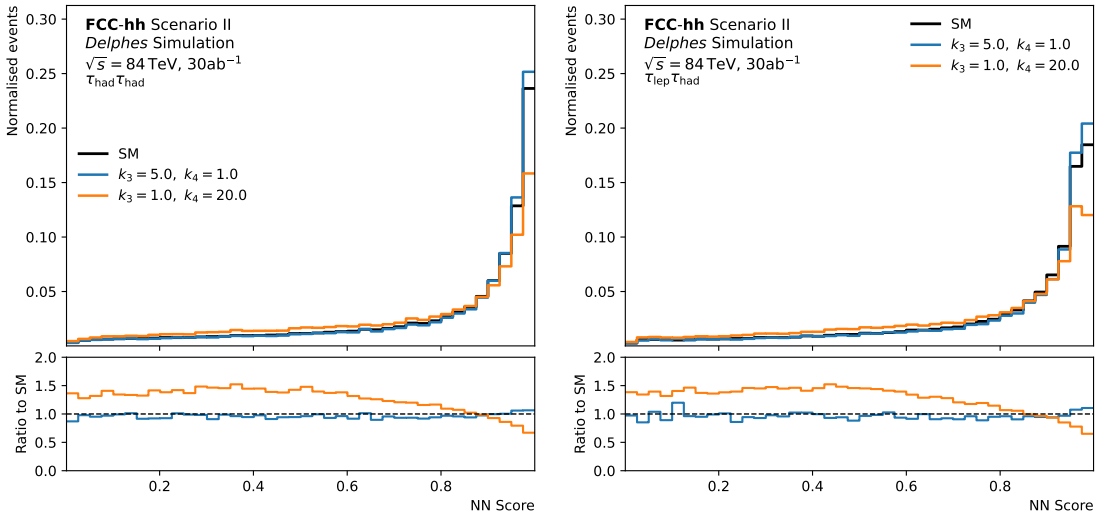


Figure 8.29: Shapley beeswarm plots for $\tau_{\text{had}}\tau_{\text{had}}$ (a) and $\tau_{\text{lep}}\tau_{\text{had}}$ (b), each ordered from most to least important from top to bottom. Higher Shapley indicate more signal-like characteristics, and lower values more background-like.

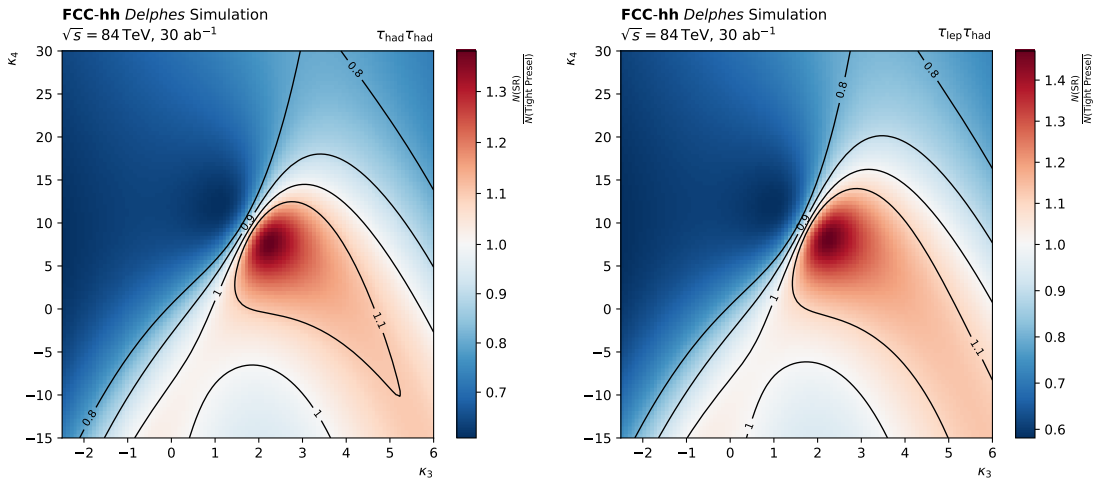
8.9.2 Robustness to (κ_3, κ_4) variations

Variations in κ_3 and κ_4 manifest in kinematically disparate regions of parameter space for the signal process; therefore, care is taken when choosing the classifier's input feature set to avoid performance loss away from the SM signal used for training. Discriminating features like m_{HHH} and the transverse momentum of reconstructed Higgs boson candidates were excluded for this reason. Resonant features targeting specific Higgs boson kinematics are invariant across the (κ_3, κ_4) plane, and thus prove beneficial in rejecting background whilst remaining robust across the targeted parameter space. Variations on the NN score for two (κ_3, κ_4) points are shown in Figure 8.30. A complete representation of this across the full (κ_3, κ_4) plane is shown in Figure 8.31, where the networks do not vary significantly across the plane — but ultimately point to perhaps a more sophisticated approach being beneficial, whereby the networks are parametrised across the plane or use a mixed (κ_3, κ_4)-signal training set. It is seen that more highly boosted Higgs bosons (see the regions of Figure 8.10(b)) result in stronger-NN performance, in alignment



(a) **(b)**

Figure 8.30: Normalised histogram of signal NN scores for the SM point and (κ_3, κ_4)=$(5, 1), (1, 20)$ in the $\tau_{\text{had}}\tau_{\text{had}}$ (a) and $\tau_{\text{lep}}\tau_{\text{had}}$ (b) channels.



(a) **(b)**

Figure 8.31: Relative loss of the NN across the (κ_3, κ_4) plane, defined as the ratio: $R = \frac{N_{\text{SR}}(\kappa_3,\kappa_4)}{N_{\text{SR}}(\text{SM})} \frac{N_{\text{Preselection}}(SM)}{N_{\text{Preselection}}(\kappa_3,\kappa_4)}$

with the improved ΔR pairing-algorithm performance for the $b\bar{b}$ Higgs candidates and the more collimated $\tau\tau$ pairs — each important to the network. Given the general kinematic similarities between the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels, the two networks exhibit similar behaviour across the grid.

8.10 Signal region and Statistical Interpretation

Analyses using ML-based classifiers generally have two approaches to performing a fit. The first is to place a tight selection on the NN output score and perform a yield-only statistical analysis — commonly referred to as a cut-and-count approach. The second is placing a moderately tight selection on the NN output score, and performing a yield and shape-fit in the signal region binned in a given observable, as was performed for the dark Higgs search in m_{HH}. The second approach is powerful if the analysis maintains a kinematic feature that discriminates between signal and background post-NN selection. The triple-Higgs invariant mass, m_{HHH}, as shown previously, is a strong discriminant across the (κ_3, κ_4) plane.

To optimise the chosen signal region, an estimate of the expected sensitivity is obtained by taking the likelihood ratio test statistic of two Poisson distributions, yielding

$$(-2\Delta \ln L)_{\text{binned}} = 2\left[\sum_i \mu_i - n_i + n_i \ln\left(\frac{n_i}{\mu_i}\right)\right] \tag{8.21}$$

$$(-2\Delta \ln L)_{\text{single}} = 2\left[\mu_{\text{tot}} - n_{\text{tot}} + n_{\text{tot}} \ln\left(\frac{n_{\text{tot}}}{\mu_{\text{tot}}}\right)\right] \tag{8.22}$$

where L is the product of Poisson distributions over all bins i, μ_i is the expected total event yield, n_i is the total Asimov data in bin i.

The binned m_{HHH} distribution is informative for separating a particular (κ_3, κ_4)-signal point and the SM background, as shown by the shape variations as a function of the parameters in Figure 8.6. Utilising the shape information available in m_{HHH} also lifts the degeneracy in the cross-section contour (Figure 8.32) — enabling a clear global minimum of the SM to be achieved in the Asimov-approximation as shown in Figure 8.33. The sensitivity dominance of the $\tau_{\text{had}}\tau_{\text{had}}$ channel is clear, as can be expected with the significantly larger $t\bar{t} + b\bar{b}$ background in Table 8.4. Further optimisation of the $\tau_{\text{lep}}\tau_{\text{had}}$ channel using more sophisticated lepton kinematic features in the NN training to distinguish between $W \to \nu e$ and $W \to \nu(\tau \to \nu e)$ in $t\bar{t}$ decays will decrease this discrepancy, but, due to time constraints, this is left

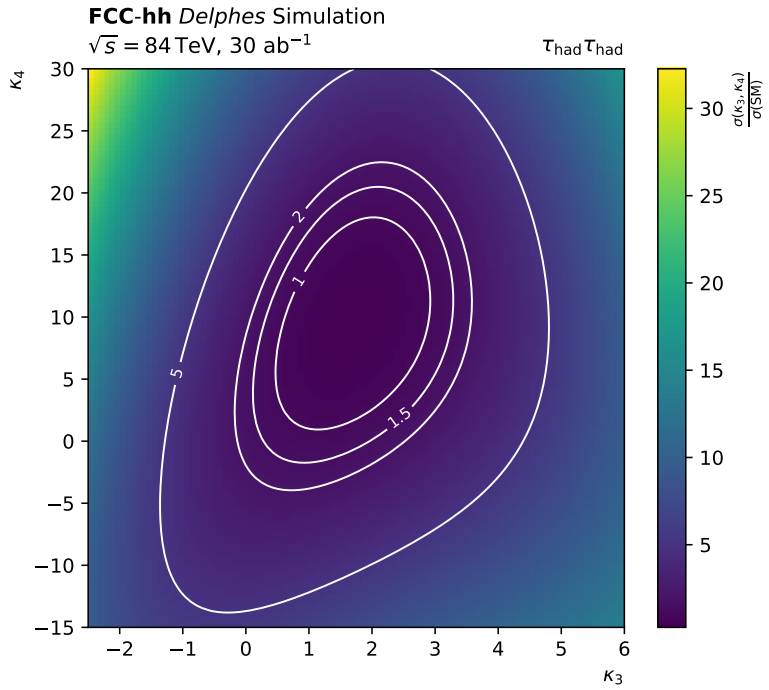


Figure 8.32: Heatmap of the ratio between the cross section at a given (κ_3, κ_4) point and the SM cross section.

for future work. The signal region is defined as a 4-bin m_{HHH} region with an NN score cut of 0.975. Yields in the SR-$\tau_{\text{had}}\tau_{\text{had}}$ and SR-$\tau_{\text{lep}}\tau_{\text{had}}$ are shown in Table 8.11. Here, given the very signal-like kinematics of the di-Higgs background, its rate becomes relevant within the SR despite its small cross section. Nonetheless, the $t\bar{t}$+$b\bar{b}$ background dominates in each region and ultimately dictates the analysis' sensitivity.

	SR-$\tau_{\text{had}}\tau_{\text{had}}$	SR-$\tau_{\text{lep}}\tau_{\text{had}}$
HHH	28.7	15.0
$t\bar{t}b\bar{b}$	324	2080
HH	121.35	52.6
$t\bar{t}t\bar{t}$	101	117
$t\bar{t}H$	66.2	27.3
$t\bar{t}Z$	7.19	6.15
ZZZ	0.18	0.08
$\frac{S}{\sqrt{B}}$	1.15	0.32

Table 8.11: Event yields and summary metrics for the triple-Higgs signal and background processes in the SR-$\tau_{\text{had}}\tau_{\text{had}}$ and SR-$\tau_{\text{lep}}\tau_{\text{had}}$ signal regions.

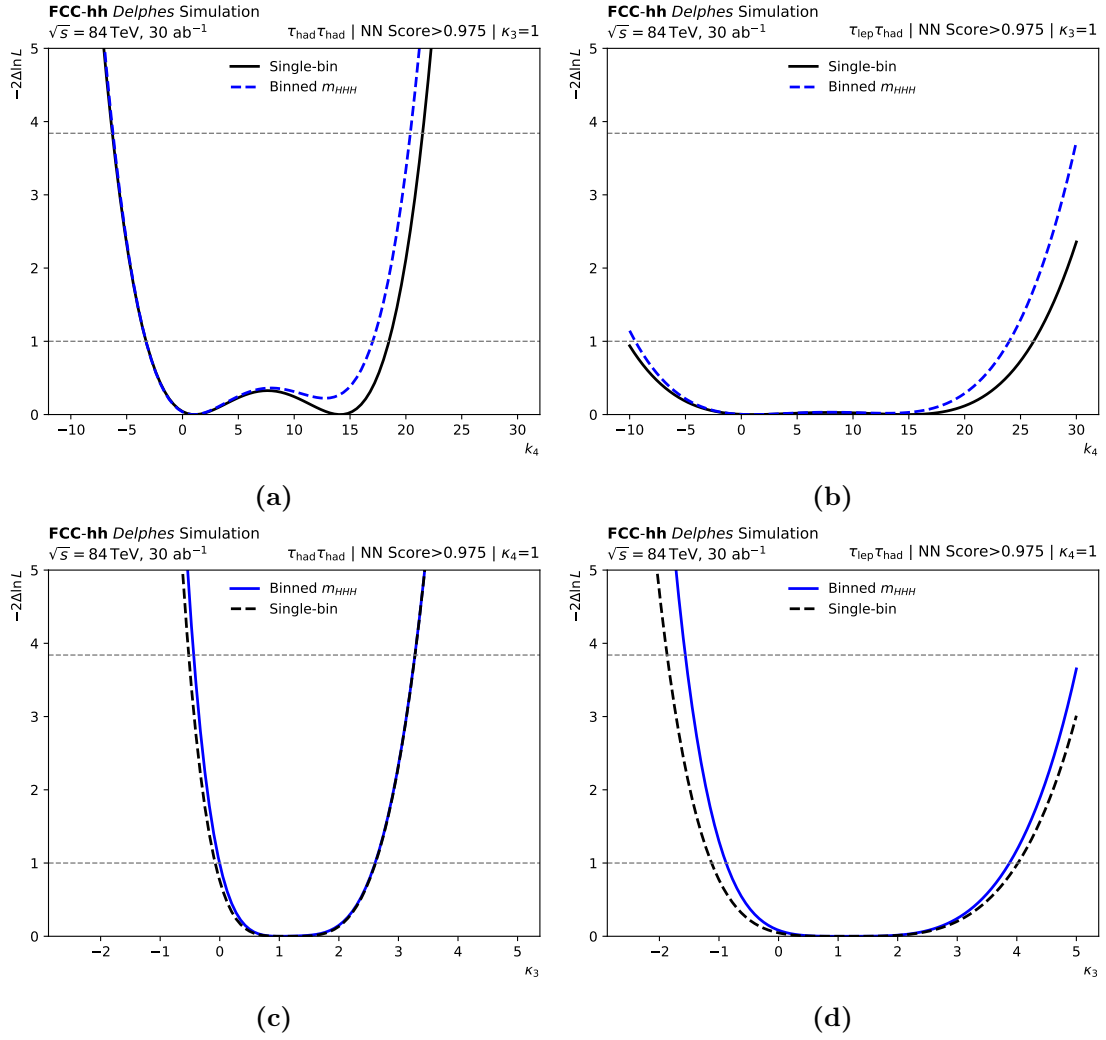


Figure 8.33: $-2\Delta \ln L$ distributions as a function of κ_4 (a,b) and κ_3 (c,d) for the single-bin and 4-bin m_{HHH} distribution for the $\tau_{\text{had}}\tau_{\text{had}}$ (a,c) and $\tau_{\text{lep}}\tau_{\text{had}}$ (b,d) channels.

8.10.1 Statistical Analysis

The profile-likelihood method, as described in Section 5.3.4 and used in Section 7.9, is used to perform the statistical interpretation. The likelihood function is comprised of a product of Poisson terms for each bin of m_{HHH} and Gaussian terms encoding the constraints of systematic nuisance parameters. For the purposes of a feasibility study, an Asimov dataset is assumed — in which the data matches the pre-fit expectation. For all fits, the Asimov dataset is defined as the SM hypothesis ($\mu_{HHH} = 1$ for upper limits, and $(\kappa_3, \kappa_4) = (1, 1)$ otherwise).

A combined likelihood is constructed for the two independent channels. The

common fitted parameter of interest, μ, is the signal strength of a given (κ_3, κ_4) prediction, defined as the ratio of the observed to the predicted value of the signal cross section times branching ratio for that parameter point. For upper limits on the SM HHH cross section, the 95% confidence level is obtained using the frequentist CL_s prescription with the profile likelihood ratio test statistic. Confidence regions in the (κ_3, κ_4) plane and one-dimensional intervals on each coupling are constructed from the profile likelihood using the asymptotic approximation of Ref. [157], taking fixed thresholds in $-2\Delta \ln \mathcal{L}$ corresponding to 68% and 95% CL.

Four fits are performed using the `TRexFitter` framework [239]. Firstly, an upper-limit scan is performed for the SM HHH cross section, with the limits corresponding to the 95% CL. Secondly, independent variations to (κ_3, κ_4) are considered across the plane using a 2D likelihood scan. Given that LHC measurements currently favour κ_3 values close to the SM prediction [240], a third fit is performed in which κ_3 is fixed to its SM value while κ_4 is varied, effectively probing the quartic coupling in isolation. Finally, a fit scanning κ_3 values for $\kappa_4 = 1$ is performed, to evaluate the sensitivity to κ_3 alone. Table 8.12 summarises the various fits considered by the analysis.

Fit	Parameter treatment	Asimov dataset
SM HHH upper limit	μ floated, corresponding to the SM signal hypothesis	$\mu_{HHH} = 1$
2D coupling scan	κ_3 and κ_4 varied simultaneously	SM point: $(\kappa_3, \kappa_4) = (1, 1)$
1D κ_4 scan	$\kappa_3 = 1$ fixed, κ_4 varied	SM point: $(\kappa_3, \kappa_4) = (1, 1)$
1D κ_3 scan	$\kappa_4 = 1$ fixed, κ_3 varied	SM point: $(\kappa_3, \kappa_4) = (1, 1)$

Table 8.12: Summary of the four profile-likelihood fits performed in the analysis, and the corresponding Asimov datasets used.

8.10.2 Systematic Uncertainties

Systematic uncertainties relating to the luminosity measurement, object identification and background normalisations are considered. In line with other FCC-hh

feasibility studies, two scenarios are outlined in Table 8.13.

Theoretical uncertainties are not considered in this analysis. With triple-Higgs production in its infancy with respect to experimental measurements, the theoretical calculations required for high-precision measurements do not yet exist — and the feasibility study therefore does not include theoretical uncertainties. Progress has been made in recent years to improve the status of HHH calculations, with a fully analytical leading-order calculation [231] paving the way for more sophisticated NLO predictions to be made — foreseeing significant improvement on the current 20% uncertainty on the cross section by the time the FCC-hh is operational.

Table 8.13: Systematic uncertainty scenarios.

Source	Uncertainty Scenario I	Uncertainty Scenario II
b-jet identification	1%	2%
τ identification	1%	2%
Background norm	1%	1%
Luminosity	1%	1%

8.10.3 Fit Behaviour

Nuisance parameter pulls, constraints and impact are shown post-fit for the κ_4 fit in Figure 8.34. With the Asimov setup, all pulls are zero. Due to the statistically limited signal region and the lack of dedicated control regions, small constraints are imposed on the object identification and luminosity uncertainties. Since these uncertainties are flat with respect to the overall yield, the constraint is identical for each of the three parameters. Small constraints are imposed on the dominant $t\bar{t}+b\bar{b}$ process normalisation uncertainty and negligible constraints on the remaining set. The ranking of nuisance parameters also follows the relative composition of the SR. The full correlation matrix is shown in Figure 8.35, where all correlations are less than 25% and degenerate correlations are observed for nuisance parameters of the same uncertainty magnitude.

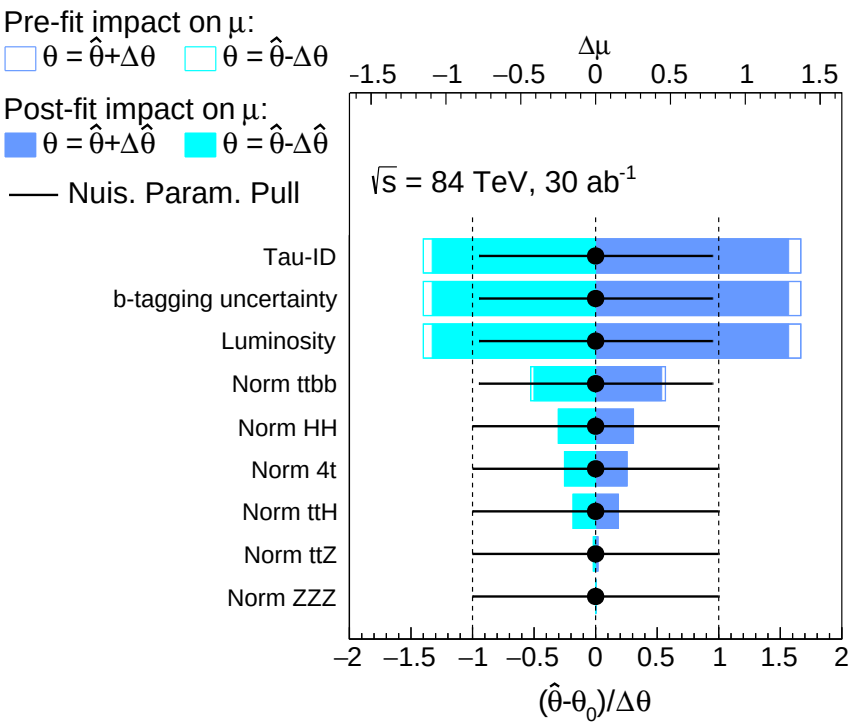


Figure 8.34: Summary of pre-fit and post-fit pulls and constraints for the κ_4-fit. For the Asimov fit, all pulls are fixed to 0.

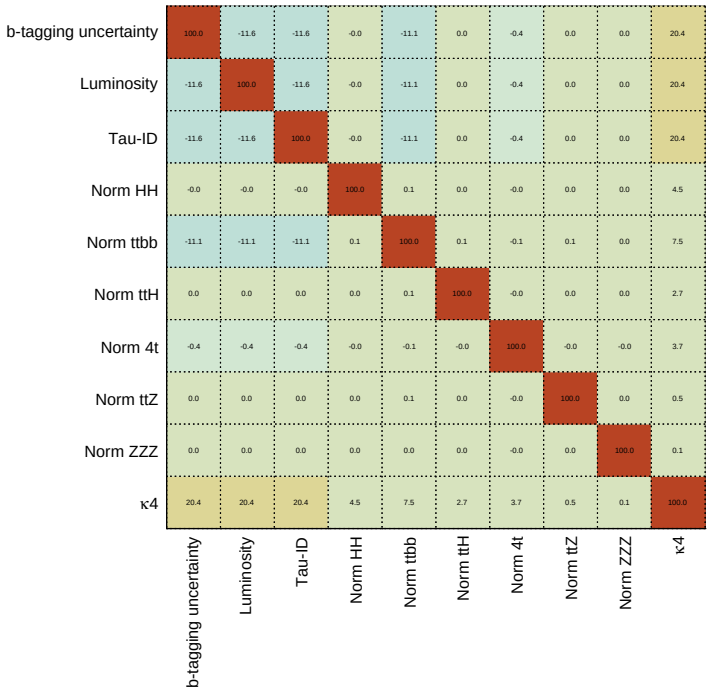


Figure 8.35: Summary of correlations between fit parameters for the κ_4-fit.

8.11 Results

Unless stated otherwise, all quoted results use the Uncertainty Scenario I systematic set-up. The expected 95% CL upper limit on the SM cross section is 6.23 fb, corresponding to $1.75 \times \sigma^{SM}_{HHH}$ at $\sqrt{s} = 84$ TeV. For comparison, the ATLAS Run 2 result [241] placed a limit of 59 fb, $750 \times \sigma^{SM}_{HHH}$ at $\sqrt{s} = 13$ TeV. A discovery test may be performed, and yields a significance of 1.17σ, matching closely to the approximate significance of $S/\sqrt{B}$ given in Table 8.11. The expected simultaneous limits on the Higgs self-coupling modifiers κ_3 and κ_4 are shown in Figure 8.36, with the respective negative log-likelihood values in a heatmap. The shape of the contour is driven predominantly by the cross section as a function of the couplings across the grid.

Independent likelihood scans are performed for each coupling modifier, scanning over one κ-value with the other fixed at the SM prediction of unity. The likelihood scan results are shown in Figure 8.37, with 68% and 95% CL levels indicated. The two systematic uncertainty Scenarios, and the 'No Syst' scenario are shown. As expected given the small signal cross section, the result is strongly statistically limited. Assuming $\kappa_3 = 1$, κ_4 is restricted to be with the range $[-3.3, 17]$ ($[-6.4, 21]$) at 68% (95%) CL. This situation would be indicated if SM HH is observed at HL-LHC or FCC-ee. Instead assuming $\kappa_4 = 1$, κ_3 is restricted to be with the range $[0.0, 2.7]$ ($[-0.5, 3.4]$) at 68% (95%) CL.

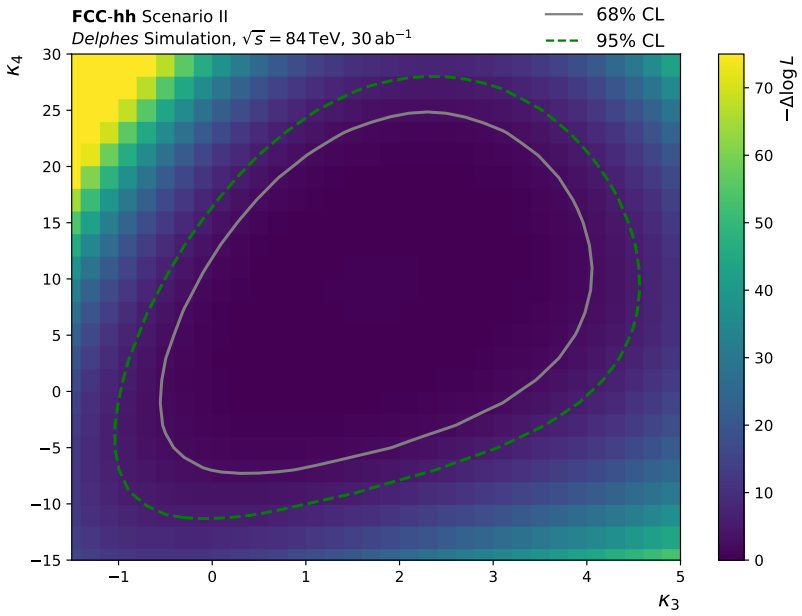
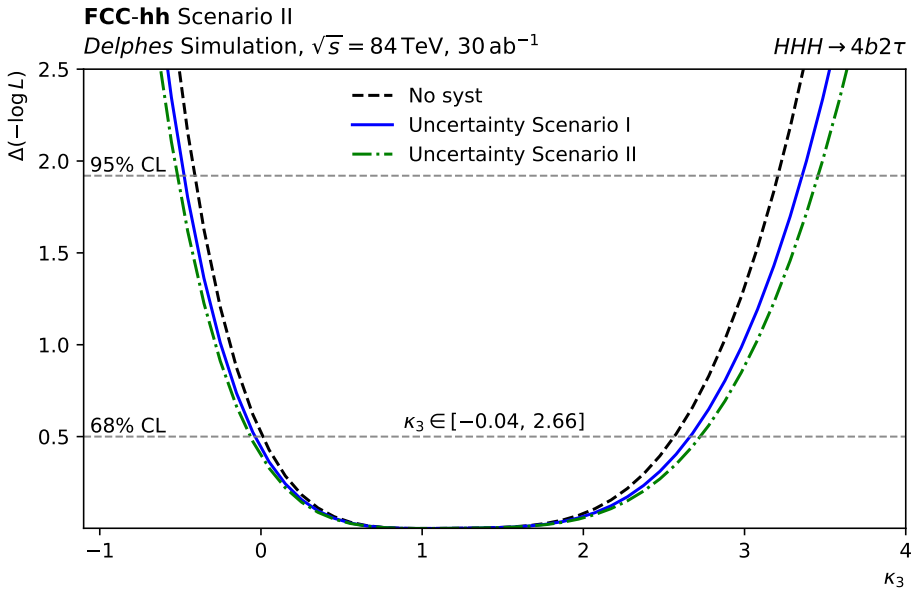
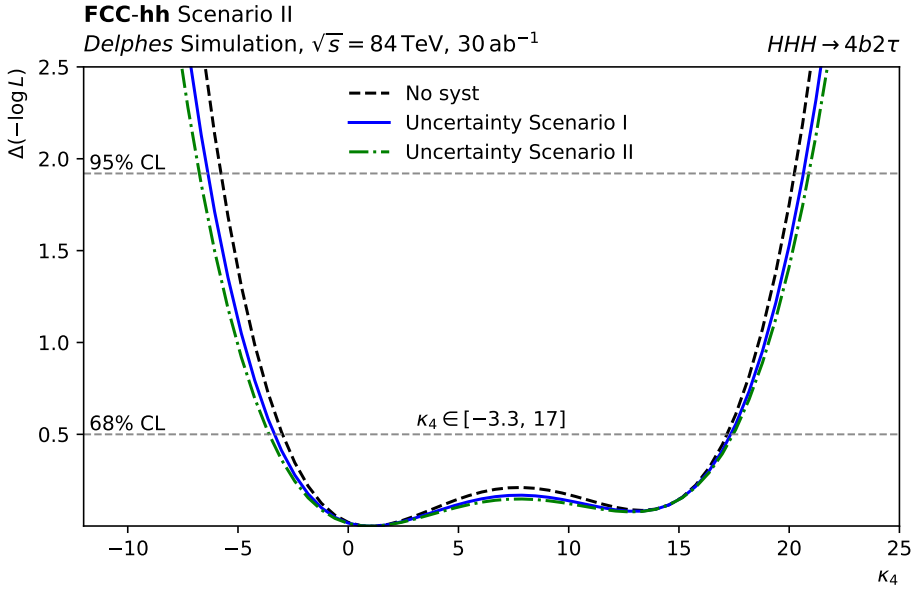


Figure 8.36: Expected 95% and 68% CL constraints on κ_3 and κ_4 from the combined fit of the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels, in dashed and solid lines respectively. The negative log-likelihood values are shown in the coloured heat-map.



(a)



(b)

Figure 8.37: One-dimensional likelihood scans over the Higgs self-coupling modifiers (a) κ_3 and (b) κ_4 from the combined fit of the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. The likelihood scan of each parameter keeps the other fixed at one. The expected likelihood contour is shown, with the 68% and 95% CL horizontal lines overlaid.

8.12 Di-Higgs at the FCC-hh

As discussed previously, there is complementary information regarding (κ_3, κ_4) to be gained from di-Higgs measurements — with the loop-level constraints arising from diagrams such as in Figure 8.2. Di-Higgs production is simulated at NLO in QCD and includes the relevant two-loop EW amplitudes for variations in (κ_3, κ_4) within the POWHEG framework [79, 232–234]. Differently from the implementation discussed in Section 8.3.1, this implementation does not use per-event reweighting, and so a uniform 20×20 grid of samples with different (κ_3, κ_4) values is simulated to estimate the cross-section dependence across the parameter space. A 9-parameter polynomial, with a similar monomial structure as Equation 8.4, is fit across the plane to interpolate between simulated points on the grid and enable a continuous 2D space to be defined.

It is seen later that sensitivity to κ_4 from di-Higgs alone lies outside this region. To validate that this methodology is generalisable to regions outside the directly simulated parameter space, extrapolation tests are performed. Here, the nominal range of (κ_3, κ_4) fit values was restricted to $([-3, 3], [-15, 25])$ and tested out to the full range of $([-5, 5], [-30, 50])$. The results of the extrapolation test are shown in Figure 8.38, where the extrapolated fit behaves well with all regions within 2% of the generated cross section — validating that the fit may be applied to regions outside the explicitly simulated region. This is generally true as the 9-parameter monomial space is sufficiently spanned by the input dataset, similar to the choice of basis points in the triple-Higgs analysis, regardless of the magnitude of the parameters. The fit is then performed across the full range, and the fitted polynomial is,

$$\begin{aligned}\sigma_{HH}^{84\,\mathrm{TeV}}(\kappa_3, \kappa_4) \approx 1.8 &- 1.1\,\kappa_3 + 3.5 \times 10^{-5}\,\kappa_4 + 0.23\,\kappa_3^2 \\ &+ 2.8 \times 10^{-3}\,\kappa_3\kappa_4 + 5.4 \times 10^{-5}\,\kappa_4^2 - 1.1 \times 10^{-3}\,\kappa_3^2\kappa_4 \\ &- 3.4 \times 10^{-5}\,\kappa_3\kappa_4^2 + 7.9 \times 10^{-6}\,\kappa_3^2\kappa_4^2\,.\end{aligned} \tag{8.23}$$

This fit is applied across the targeted parameter space, which is shown in Figure 8.39.

The di-Higgs analysis targeting the trilinear Higgs self-coupling in the $b\bar{b}\gamma\gamma$ channel is used to showcase the amount of complementary information. A full

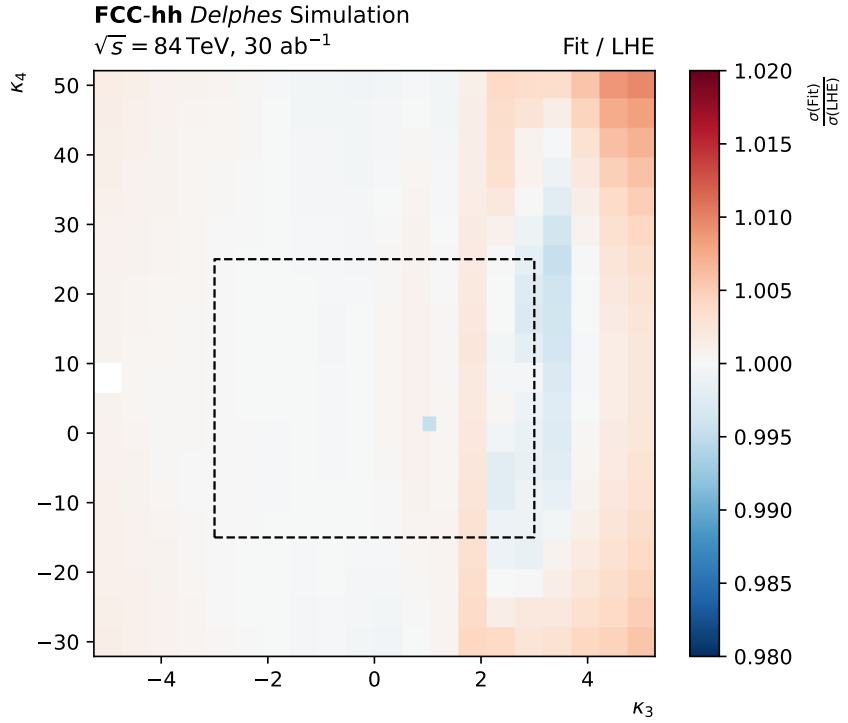


Figure 8.38: Ratio of the fitted cross section to that derived by POWHEG, showing the fit-region within the black dotted rectangle — and the ratio across the entire range of generated events.

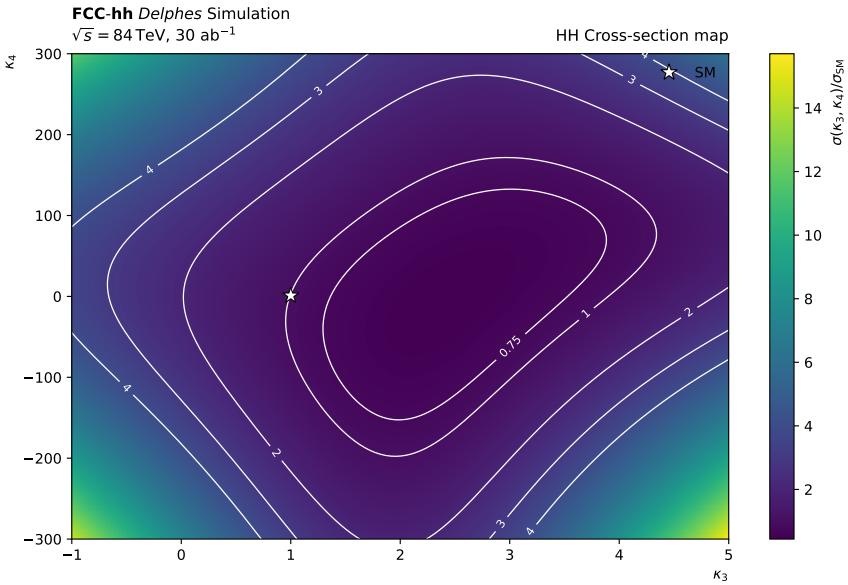
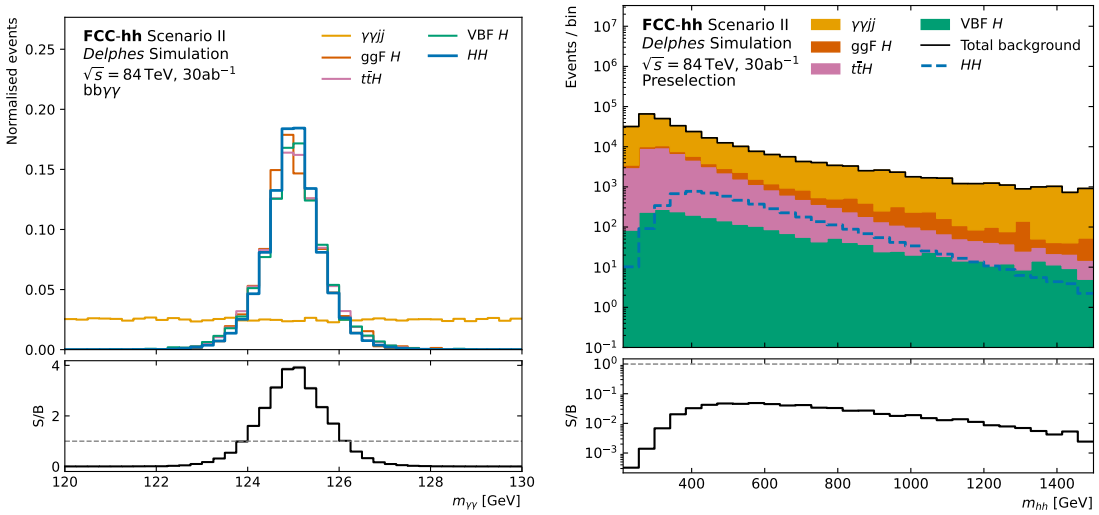


Figure 8.39: Ratio of di-Higgs cross section at (κ_3, κ_4) to the SM-point across the coupling space, including extrapolation to the full range of couplings considered by the analysis.



(a) **(b)**

Figure 8.40: Normalised di-photon invariant mass (a), and the weighted di-Higgs invariant mass (b) for all signal and background processes after event preselection.

account of this analysis is given in Ref [242], with an approximate implementation outlined and used here to illustrate sensitivity. Non-resonant production of QCD in association with photons is the dominant background — where single-Higgs (VH, VBF, ggH) constitute the remaining irreducible background. The excellent reconstruction of the di-photon invariant mass, $m_{\gamma\gamma}$, is a key observable in the analysis and is shown in Figure 8.40(a).

A tight selection window around this mass of $122 < m_{\gamma\gamma} < 128$ is used, with a loose preselection on the $b\bar{b}$-system of $80 < m_{b\bar{b}} < 160$. The di-Higgs invariant mass distribution for all signal and background processes is shown in Figure 8.40(b), where the dominant non-resonant background of $\gamma\gamma jj$ peaks at low-m_{HH} values, with a smoothly falling distribution towards the tail. Variations on κ_3 strongly distort this distribution, but the loop-induced κ_4 variations are much less pronounced, as shown for the truth m_{HH} distribution in Figure 8.41 for SM production (green) and $\kappa_4 = 60$ (blue).

The same neural network architectures as introduced for the triple-Higgs analysis are used, exploiting Collins-Soper rest-frame kinematics [243] in the bb, $\gamma\gamma$, and $bb\gamma\gamma$ frames. A cut on the output score is optimised, and limits are placed on the

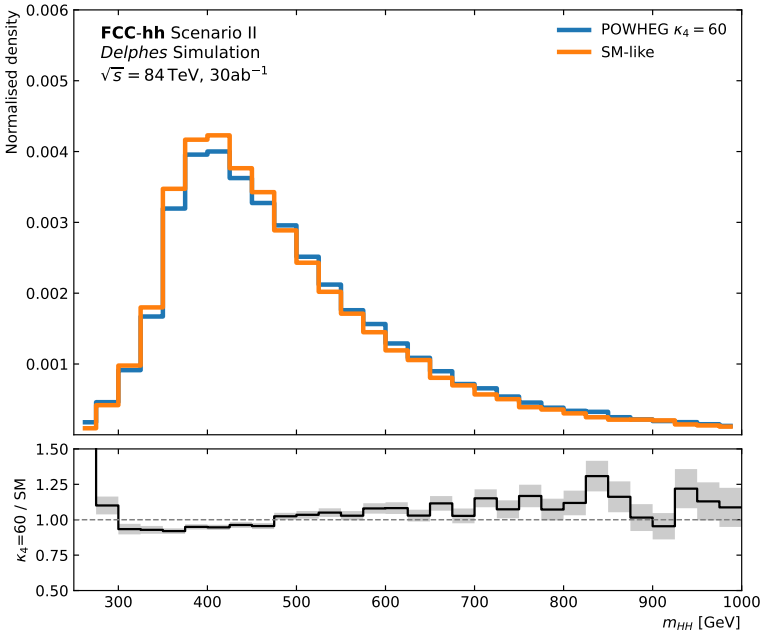


Figure 8.41: Distributions of the truth HH invariant mass, m_{HH}^{truth}, for the SM prediction and the $\kappa_4 = 60$ prediction. No selection requirements are applied. Histograms are normalised to unit area, and the lower panel shows the ratio of the varied (κ_3, κ_4) models to the SM distribution.

yield in the SR. The same set of systematic uncertainties as in the triple-Higgs analysis are used, replacing the tau-ID uncertainties with photon-ID uncertainties.

The cross-section parametrisation of Equation 8.23 is encoded in `TRexFitter`, and results are extracted via profile-likelihood fits to the Asimov dataset. A 2D heatmap of $-\Delta \ln L$ across (κ_3, κ_4) is shown, with 68% and 95% confidence intervals, in Figure 8.42. The limit follows closely with the cross-section contour as a function of the two couplings, showing strong sensitivity in κ_3 but relatively limited sensitivity with κ_4 across the plane. By inspecting the cross-section contours in Figure 8.39, the limits could be improved by binning the final discriminant in m_{HH} (see Figure 8.41), and performing a per-(κ_3, κ_4) reweighting, to resolve the observed degeneracies with shape information. The 1D likelihood scan results are shown in Figure 8.43, with 68% and 95% CL levels indicated. The two systematic uncertainty Scenarios, and the 'No Syst' scenario are shown. The double minimum structure, as discussed

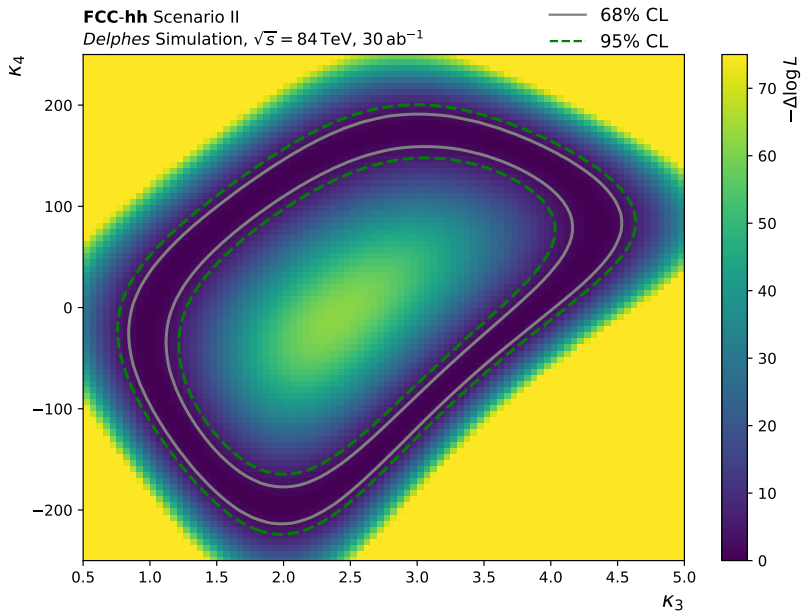
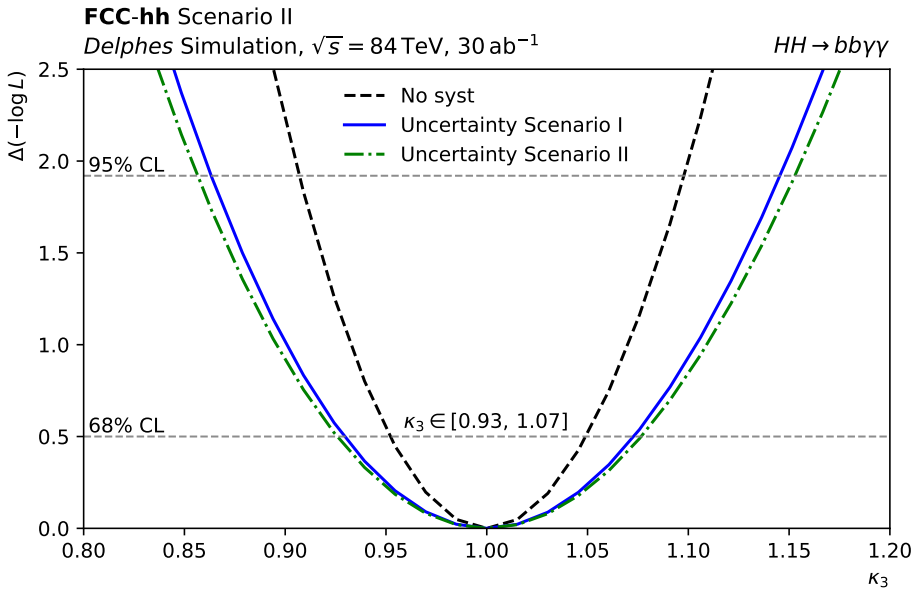
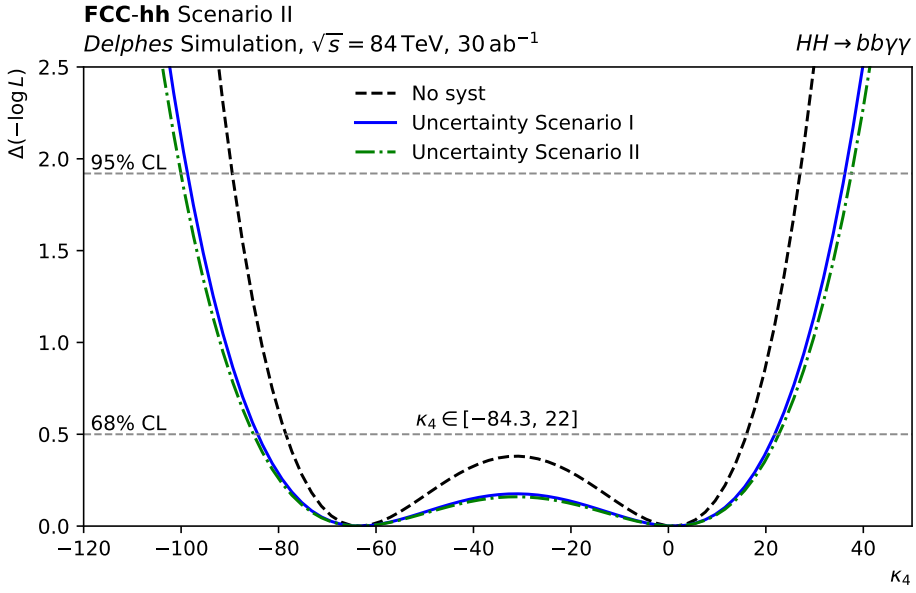


Figure 8.42: Two-dimensional heatmap of $-\Delta \ln L$ values for Uncertainty Scenario I across the (κ_3, κ_4) coupling space for the di-Higgs analysis, with the 68% and 95% CLs shown.

earlier in the triple-Higgs analysis, is observed and occurs at negative κ_4 due to the interference structure of loop-induced HH altering the production cross section in Figure 8.39. Assuming $\kappa_3 = 1$, κ_4 is restricted to be with the range $[-84, 22]$ ($[-99, 36]$) at 68% (95%) CL. Instead assuming $\kappa_4 = 1$, κ_3 is restricted to be with the range $[0.93, 1.07]$ ($[0.86, 1.2]$) at 68% (95%) CL. The results for κ_4 are competitive with the sensitivity to triple-Higgs production for positive values of the coupling, strongly motivating a combination for the most stringent limit to be placed on κ_4.



(a)



(b)

Figure 8.43: One-dimensional likelihood scans over the Higgs self-coupling modifiers (a) κ_3 and (b) κ_4. The likelihood scan of each parameter keeps the other fixed at 1. The expected likelihood contour is shown, with the 68% and 95% CL horizontal lines overlaid.

8.13 Combination

The two analyses' selection criteria are orthogonal and therefore compatible for a combination. `TRexFitter` is used to perform the combination. Unlike the original triple-Higgs implementation, the HH background is reweighted according to the considered (κ_3, κ_4) parameters. This will individually improve the triple-Higgs analysis' sensitivity to κ_3 since the HH background rapidly varies with κ_3. Table 8.14 summarises all systematic uncertainties used in each analysis. b-tagging and luminosity uncertainties are common to both analyses and therefore correlated; all remaining uncertainties are uncorrelated.

Table 8.14: Systematic uncertainty scenarios assumed for the $HHH \to 4b2\tau$ and $HH \to b\bar{b}\gamma\gamma$ analyses.

Analysis	Source	Scenario I	Scenario II
$HHH \to 4b2\tau$	b-jet identification	1%	2%
	τ identification	1%	2%
	Background norm.	1%	1%
	Luminosity	1%	1%
$HH \to b\bar{b}\gamma\gamma$	b-jet identification	1%	2%
	γ identification	1%	2%
	Background norm.	1%	1%
	Luminosity	1%	1%

Nuisance parameter pulls, constraints and impacts are shown post-fit for the combined κ_4 fit in Figure 8.44. With the Asimov setup, all pulls are zero. The high-statistics di-Higgs analysis yields constraints of order 10% on the b-tagging, luminosity and photon-ID nuisance parameters. Parameters correlated across the two analyses show stronger constraints, whereas the τ-ID has fewer correlations and is therefore highly ranked despite HHH being statistically limited. The full correlation matrix is shown in Figure 8.45, where nuisance parameters that are degenerate in size for the HH-analysis show strong cross-correlations, and others are all small and below 20%.

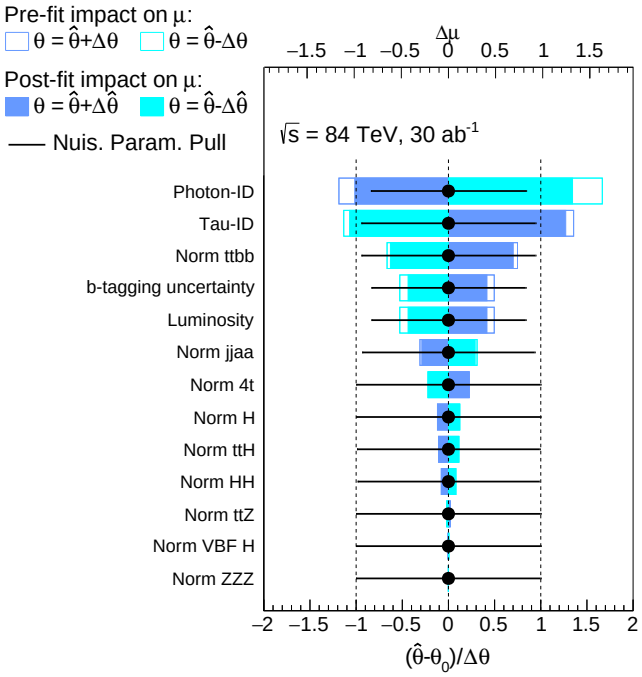


Figure 8.44: Summary of pre- and post-combined fit pulls and constraints for the κ_4-scan. For the Asimov fit, all pulls are fixed to 0.

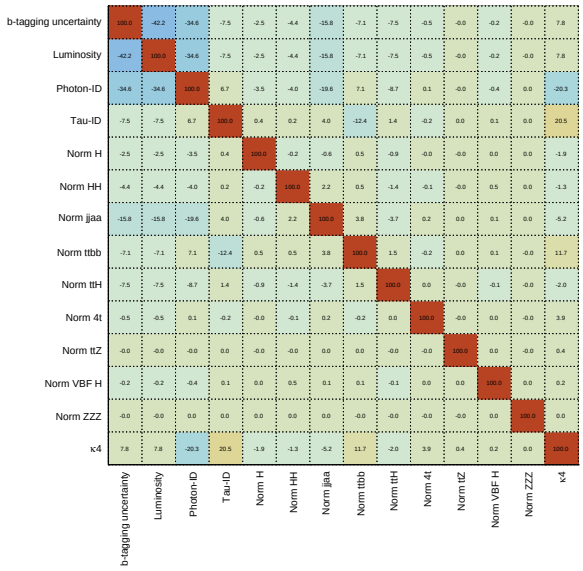


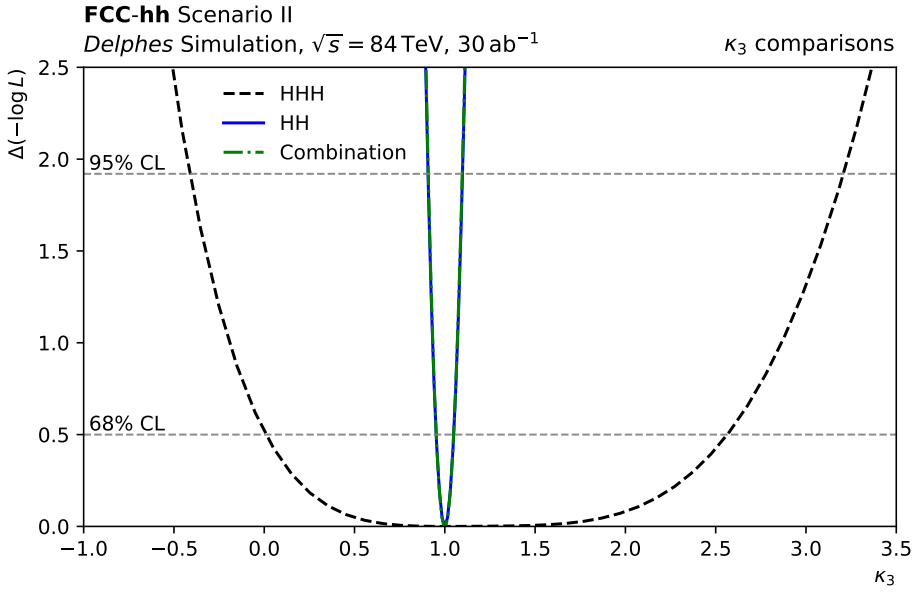
Figure 8.45: Summary of correlations between combined fit parameters for the κ_4-scan.

Comparisons between the two analyses for all three coupling limit scenarios κ_3, κ_4 and (κ_3, κ_4) are shown in Figures 8.46(a,b) and Figure 8.47. Here it is seen that the combination dramatically reduces the allowed space in κ_3 in the (κ_3, κ_4)-limit, whilst reducing the space of allowed positive-κ_4 configurations by including the loop-induced HH κ_4-dependence. Improved constraints on κ_3 and κ_4 are possible with the inclusion of multiple HH final states, such as $bb\tau\tau$ or $4b$. Table 8.15 summarises the 1D-scan configurations across the HHH, HH and combined analyses.

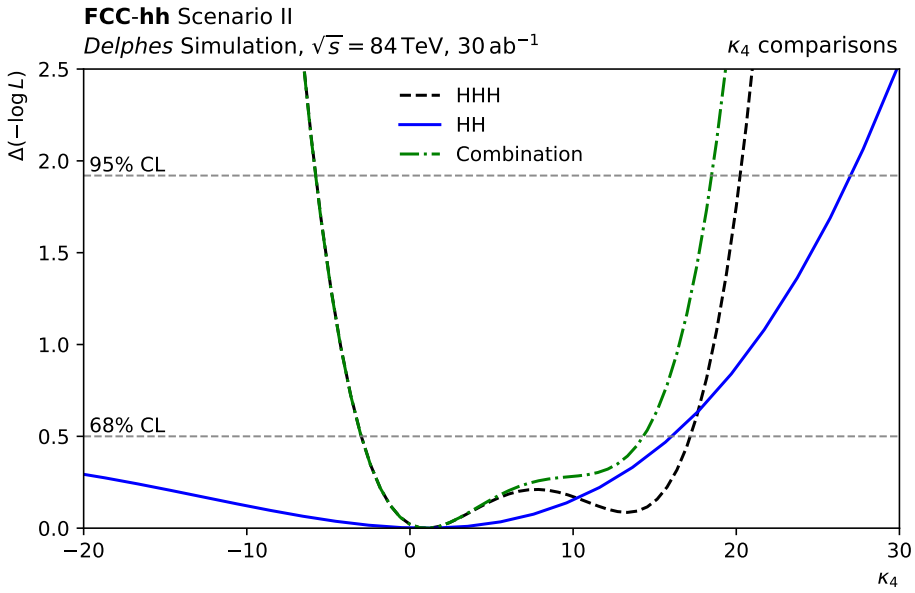
Table 8.15: Summary of expected 68% confidence intervals from 1D scans of κ_3 and κ_4 for the HHH, HH, and combined measurements, under the statistical-only and Uncertainty Scenario I assumptions.

Measurement	κ_3 1D scan		κ_4 1D scan	
Type	Stat-only	Scenario I	Stat-only	Scenario I
HHH	[0.01, 2.56]	[−0.04, 2.7]	[−3.0, 17]	[−3.3, 17]
HH	[0.95, 1.05]	[0.93, 1.07]	[−78, 16]	[−84, 22]
Combination	[0.95, 1.05]	[0.93, 1.07]	[−3.0, 14]	[−3.2, 16]

The expected simultaneous limits on the Higgs self-coupling modifiers κ_3 and κ_4 are shown in Figure 8.48. Systematic uncertainty scenarios I and II are overlaid. The contour is systematically (statistically) limited in $\kappa_3(\kappa_4)$-direction in line with the limitations of the HH (HHH) analysis.



(a)



(b)

Figure 8.46: Comparisons of expected 68% CL constraints on κ_3 (a) and κ_4 (b) from the HHH, HH and combined fits.

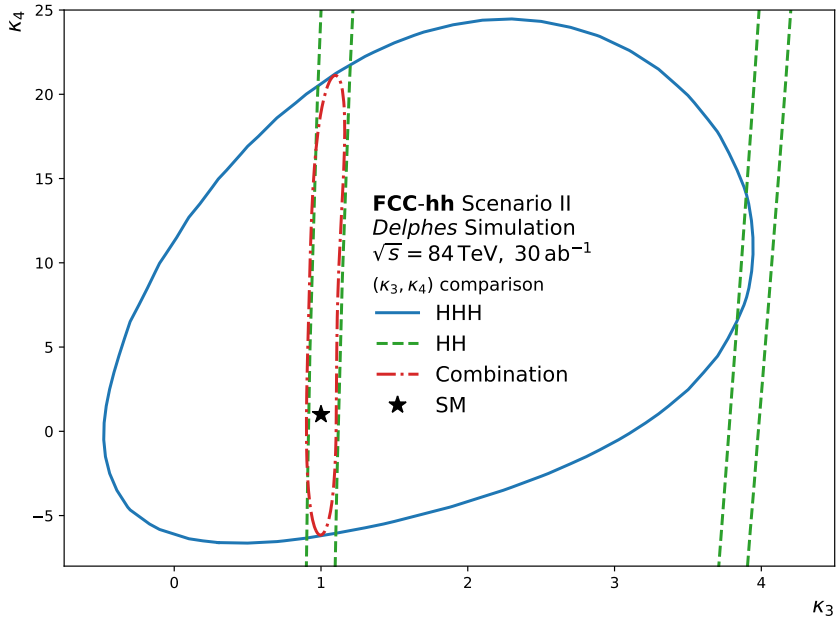


Figure 8.47: Comparisons of expected 68% CL constraints on (κ_3, κ_4) from the HHH, HH and combined fits.

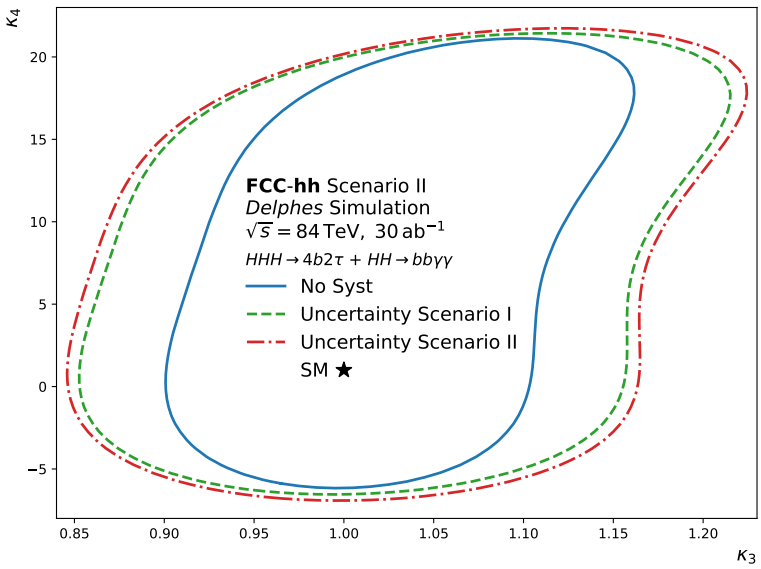


Figure 8.48: Expected 68% CL constraints on (κ_3, κ_4) from the combined fit of $HHH \to 4b2\tau$ and $HH \to b\bar{b}\gamma\gamma$.

8.14 A view to the future

The limits set by this analysis on κ_4 reflect the challenges of probing triple-Higgs production, even at a significantly higher centre-of-mass energy. Nonetheless, several extensions of the analysis, alongside adjustments to detector design, could improve sensitivity to κ_4.

Firstly, the small radius jet reconstruction begins to lose efficiency when the decay products of a boosted $H \to b\bar{b}$ decay overlap. The preliminary large-radius jet study in Section 8.4.1 demonstrates that part of this acceptance can be recovered and improve sensitivity. Furthermore, optimisation of the machine learning approach to deploy a dedicated coupling-dependent model, could tighten constraints in the kinematically challenging destructive interference region — a region coinciding with significant Higgs boson boost. Further gains would be expected from dedicated optimisation of the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and the combination with additional HHH ($6b$, $4b2\gamma$, $4b2W$ and $4\tau 2b$, for example) and HH final states.

For a fixed centre-of-mass energy and integrated luminosity, the most relevant experimental improvements are those that improve signal acceptance and mass reconstruction. Reconstruction of hadronic jets down to transverse momenta below 25 GeV would significantly improve the selection efficiency of the fourth b-jet, while a tighter segmentation of the hadronic calorimeters could reduce the stochastic jet energy resolution, improving the reconstruction of jets — resulting in a sharper peak on the m_{bb} and $m_{\tau\tau}$ distributions, providing efficient discriminants against the $t\bar{t} + b\bar{b}$ background.

8.15 Conclusions

The quartic Higgs self-coupling is central to our understanding of the shape of the Higgs potential, and models that account for electroweak baryogenesis require deviations of the Higgs self-couplings to generate a first-order electroweak phase transition. However, its measurement is made exceptionally challenging by the tiny signal cross section for triple Higgs production predicted by the SM, which is only

improved by higher-energy colliders. A feasibility study targeting the quartic Higgs self-coupling at the proposed $\sqrt{s} = 84$ TeV FCC-hh experiment is presented.

This is the first study of triple-Higgs production within the FCC collaboration and at the updated $\sqrt{s} = 84$ TeV benchmark. With the large 30 ab^{-1} dataset collected at the FCC-hh, the $HHH \to 4b2\tau$ final state is probed, where the selection of two τ-leptons, alongside the use of modern di-τ mass reconstruction techniques, effectively reduces the dominant $t\bar{t} + b\bar{b}$ background process. Neural networks are trained exploiting the triply-resonant signal within the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels, and profile-likelihood fits are performed to set expected constraints of $\kappa_4 \in [-3.3, 17]$ at the 68% CL. This corresponds to over a factor of 5 improvement relative to the HL-LHC projections. Indirect sensitivity using di-Higgs indicates strong potential benefit in combination with triple Higgs to probe κ_4, improving the limit when combined in a global fit to $\kappa_4 \in [-3.2, 16]$.

For now, probing SM-like values of the quartic Higgs self-coupling, even at an $\sqrt{s} = 84$ TeV FCC-hh experiment, will remain exceptionally challenging. Nonetheless, developments have been numerous in the past 10 years, especially regarding measurements of the Higgs trilinear self-coupling, so there must be a degree of optimism when looking ahead. Moreover, there are other avenues to explore the electroweak symmetry-breaking mechanism of the SM, such as space-based interferometers that could measure gravitational waves originating from a first-order phase transition, or indeed the experimentally ambitious muon collider [244].

"And away..."

Bob Mortimer and Paul Whitehouse

9 Thesis Summary and Outlook

The discovery of the Higgs boson in 2012 completed the Standard Model of particle physics, which proves remarkable in its ability to predict a vast array of processes across a wide variety of scales. However, the SM is incomplete and, notably, cannot explain astrophysical observations of dark matter or provide a mechanism to generate the observed matter-antimatter asymmetry of the Universe. Many limitations of the SM are tied to the Higgs sector; mass generation via the Higgs mechanism could allow gravitationally interacting dark matter to acquire a mass, and the insufficient baryon asymmetry predicted by the SM Higgs potential points to extended scalar sectors.

First, since the two analyses presented rely on excellent flavour-tagging performance, calibrations of the `DL1d` and `GN2` algorithms are presented for high-p_{T} b-jets. Tracks within these jets are tightly collimated into a dense core, making charged-track reconstruction and thus the identification of the displaced b-hadron decay vertex exceptionally challenging. The uncertainty associated with the reconstruction efficiency of charged tracks within such cores constitutes the largest systematic uncertainty on the tagging rate. The calibrations presented enable both algorithms to be used in analyses of Run 2 and Run 3 data collected by the ATLAS experiment.

Second, a search for a dark Higgs boson in the $HH + p_{\mathrm{T}}^{\mathrm{miss}}$ final state is presented using the full Run 2 dataset collected by the ATLAS experiment. The dark Higgs model is compelling as an extension of the SM as it provides a mass-generation mechanism for the dark matter candidate, χ, and opens up additional channels for DM annihilation, more naturally reproducing the observed relic density. The analysis employs a neural-network classifier tailored to the resonant HH system, with a dedicated mass-decorrelation strategy to distinguish SM backgrounds from the dark Higgs signal. Control regions are defined to constrain and improve the modelling of the top and $Z+$jets background processes, and the background predictions are

compared with the observed data in a SR binned in m_{HH}. No excess is observed above the SM expectation, and limits are therefore set that extend ATLAS' exclusion reach across the full dark Higgs boson mass range considered and up to 2.3 TeV in $m_{Z'}$.

Finally, the third research element of this thesis turns to the Higgs self-interactions associated with electroweak symmetry breaking in the SM. A feasibility study targeting the quartic Higgs self-coupling at the $\sqrt{s} = 84$ TeV FCC-hh is performed in the $HHH \to 4b2\tau$ final state, where the reconstruction of two τ-leptons, and of their invariant mass with MMC, helps reduce background contributions. The study finds expected sensitivities significantly improved with respect to the HL-LHC, with $\kappa_4 \in [-3.3, 17]$. An extension of the study included the indirect sensitivity to κ_4 with EW loop corrections to di-Higgs production, which tightens the expected limits for positive κ_4 values in the range $\kappa_4 \in [-3.2, 16]$.

ATLAS is currently preparing for the HL-LHC, with its phase-II upgrade fully replacing the inner detector with the all-silicon ITk [118, 245], installing the High-Granularity Timing Detector [246], introducing an updated trigger and data-acquisition system [247], and upgrading the calorimeter [248, 249] and muon readout electronics [250]. The insertion of the ITk will enable roughly stable track reconstruction efficiencies between Run 3 and 4 despite the large increase in pile-up, and enable tracking, and therefore the tagging of heavy-flavour jets, up to $|\eta| < 4$. The innermost barrel layer will use pixels with dimensions of 25×100 μm^2, whose finer granularity should aid track reconstruction within the dense cores of high-p_{T} jets, potentially improving the efficiency of flavour-tagging algorithms, and reducing the large uncertainty in the extrapolation region. The significant leap in performance made by GN2 is already being utilised across the ATLAS experiment and the substantial increase in statistical power that the over-ten-times-larger HL-LHC dataset provides, together with continued advances in machine learning, statistical inference and analysis design, will considerably extend the sensitivity of measurements of and searches beyond the SM. And, given the progress achieved during Runs 1–3 of the LHC, and the further advances anticipated from the HL-LHC and FCC-ee, we can remain optimistic and excited by the prospects of the FCC-hh.

References

[1] ATLAS Collaboration. "Transforming jet flavour tagging at ATLAS". In: *Nature Commun.* 17.1 (2026), p. 541. arXiv: 2505.19689 [hep-ex].

[2] Tim Lukas Bruckler. "Searching for Dark Matter in the Light of Dark-Higgs Strahlung: A first search for a resonant di-Higgs + E_T^{miss} signature with the ATLAS experiment". Presented 09 Dec 2024. U. Oxford, 2024. URL: https://cds.cern.ch/record/2921787.

[3] ATLAS Collaboration. "A search for dark matter produced in association with a dark Higgs boson decaying into a Higgs boson pair in $3b$ or $4b$ final states using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 09 (2025), p. 067. arXiv: 2507.02175 [hep-ex].

[4] Jorge de Blas et al. "Physics Briefing Book: Input for the 2026 update of the European Strategy for Particle Physics". In: (Nov. 2025). arXiv: 2511.03883 [hep-ex].

[5] Thomas Benjamin Dingley, Holly Pacey, and James Frost. *Feasibility studies for triple Higgs boson production in the $4b2\tau$ final state at the FCC-hh*. Sept. 2025. URL: https://doi.org/10.17181/5ywv6-es391.

[6] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001.

[7] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Addison-Wesley, 1995.

[8] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[9] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Nonabelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (1973), pp. 1343–1346.

[10] E. Fermi. "An attempt of a theory of beta radiation. 1." In: *Z. Phys.* 88 (1934), pp. 161–177.

[11] C. S. Wu et al. "Experimental Test of Parity Conservation in β Decay". In: *Phys. Rev.* 105 (1957), pp. 1413–1414.

[12] R. P. Feynman and M. Gell-Mann. "Theory of the Fermi Interaction". In: *Phys. Rev.* 109 (1 Jan. 1958), pp. 193–198. URL: https://link.aps.org/doi/10.1103/PhysRev.109.193.

[13] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588.

[14] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266.

[15] J Goldstone. "Field theories with "superconductor" solutions". In: *Nuovo Cimento* 19 (1961), pp. 154–164. URL: https://cds.cern.ch/record/343400.

[16] Philip W. Anderson. "Plasmons, Gauge Invariance, and Mass". In: *Physical Review* 130.1 (1963), pp. 439–442.

[17] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physical Review Letters* 13 (1964), pp. 321–323.

[18] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13 (1964), pp. 508–509.

[19] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Physical Review Letters* 13 (1964), pp. 585–587.

[20] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". In: (July 2013). Ed. by S Heinemeyer et al. arXiv: 1307.1347 [hep-ph].

[21] Benjamin W. Lee, C. Quigg, and H. B. Thacker. "Weak interactions at very high energies: The role of the Higgs-boson mass". In: *Phys. Rev. D* 16 (5 Sept. 1977), pp. 1519–1531. URL: https://link.aps.org/doi/10.1103/PhysRevD.16.1519.

[22] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657.

[23] F. J. Hasert et al. "Search for Elastic ν_μ Electron Scattering". In: *Phys. Lett. B* 46 (1973), pp. 121–124.

[24] F. J. Hasert et al. "Observation of Neutrino Like Interactions without Muon or Electron in the Gargamelle Neutrino Experiment". In: *Nucl. Phys. B* 73 (1974), pp. 1–22.

[25] Arnison G. et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540GeV$". In: *Phys. Lett. B* 122 (1983), pp. 103–116. URL: https://cds.cern.ch/record/142059.

[26] Banner M. et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\overline{p}$p collider". In: *Phys. Lett. B* 122 (1983), pp. 476–485. URL: https://cds.cern.ch/record/142759.

[27] G. Arnison et al. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider". In: *Phys. Lett. B* 126 (1983), pp. 398–410.

[28] P. Bagnaia et al. "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 129 (1983), pp. 130–140.

[29] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. arXiv: 1207.7214 [hep-ex].

[30] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. arXiv: 1207.7235 [hep-ex].

[31] Jorinde van de Vis, Jordy de Vries, and Marieke Postma. "Bubble Trouble: a Review on Electroweak Baryogenesis". In: *Prog. Part. Nucl. Phys.* 150 (2026), p. 104244. arXiv: 2508.09989 [hep-ph].

[32] F. Csikor, Z. Fodor, and J. Heitger. "Endpoint of the hot electroweak phase transition". In: *Phys. Rev. Lett.* 82 (1999), pp. 21–24. arXiv: hep-ph/9809291.

[33] ATLAS Collaboration. "Combined Measurement of the Higgs Boson Mass from the H→$\gamma\gamma$ and H→ZZ*→4ℓ Decay Channels with the ATLAS Detector Using s=7, 8, and 13 TeV pp Collision Data". In: *Phys. Rev. Lett.* 131.25 (2023), p. 251802. arXiv: 2308.04775 [hep-ex].

[34] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual". In: *JHEP* 05 (2006), p. 026. arXiv: hep-ph/0603175.

[35] G. Corcella et al. "HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)". In: *JHEP* 01 (2001), p. 010. arXiv: hep-ph/0011363.

[36] M. Bähr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. arXiv: 0803.0883 [hep-ph].

[37] S. Agostinelli et al. "GEANT4 - A Simulation Toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303.

[38] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization for Short Distance Hadron-Hadron Scattering". In: *Nucl. Phys. B* 261 (1985), pp. 104–142.

[39] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. arXiv: 1510.03865 [hep-ph].

[40] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph].

[41] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. arXiv: hep-ph/0409146.

[42] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. arXiv: 0709.2092 [hep-ph].

[43] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. arXiv: 1002.2581 [hep-ph].

[44] Andy Buckley et al. "General-purpose event generators for LHC physics". In: *Phys. Rept.* 504 (2011), pp. 145–233. arXiv: 1101.2599 [hep-ph].

[45] Stefano Frixione and Bryan R. Webber. "Matching NLO QCD computations and parton shower simulations". In: *JHEP* 06 (2002), p. 029. arXiv: hep-ph/0204244.

[46] Bo Andersson et al. "Parton fragmentation and string dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145.

[47] Torbjorn Sjöstrand. "Jet fragmentation of multiparton configurations in a string framework". In: *Nucl. Phys. B* 248 (1984), pp. 469–502.

[48] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: *SciPost Phys. Codeb.* (2022), p. 8. arXiv: 2203.11601 [hep-ph].

[49] Daniele Amati and Gabriele Veneziano. "Preconfinement as a property of perturbative QCD". In: *Phys. Lett. B* 83 (1979), pp. 87–92. URL: https://cds.cern.ch/record/133141.

[50] ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *Eur. Phys. J. C* 70 (2010), pp. 823–874.

[51] ATLAS Collaboration. *Standard Model Summary Plots June 2024*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-011. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2903866.

[52] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: (2017). arXiv: 1610.07922 [hep-ph].

[53] NNPDF Collaboration, Richard D. Ball, et al. "Parton distributions for the LHC run II". In: *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph].

[54] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. arXiv: hep-ex/9807003.

[55] Planck Collaboration. "Planck 2018 results. VI. Cosmological parameters". In: *Astron. Astrophys.* 641 (2020), A6. arXiv: 1807.06209 [astro-ph.CO]. Errantum: in: *Astron. Astrophys.* 652 (2021), p. C4.

[56] V. C. Rubin, N. Thonnard, and W. K. Ford Jr. "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/". In: *Astrophys. J.* 238 (1980), p. 471.

[57] V. C. Rubin et al. "Rotation velocities of 16 SA galaxies and a comparison of Sa, Sb, and SC rotation properties". In: *Astrophys. J.* 289 (1985), p. 81.

[58] Edvige Corbelli and Paolo Salucci. "The Extended Rotation Curve and the Dark Matter Halo of M33". In: *Mon. Not. Roy. Astron. Soc.* 311 (2000), pp. 441–447. arXiv: astro-ph/9909252.

[59] K. G. Begeman, A. H. Broeils, and R. H. Sanders. "Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics". In: *Mon. Not. Roy. Astron. Soc.* 249 (1991), p. 523.

[60] Douglas Clowe et al. "A direct empirical proof of the existence of dark matter". In: *Astrophys. J. Lett.* 648 (2006), pp. L109–L113. arXiv: astro-ph/0608407.

[61] Planck Collaboration. "Planck 2018 results. I. Overview and the cosmological legacy of Planck". In: *Astron. Astrophys.* 641 (2020), A1. arXiv: 1807.06205 [astro-ph.CO].

[62] Gary Steigman and Michael S. Turner. "Cosmological Constraints on the Properties of Weakly Interacting Massive Particles". In: *Nucl. Phys. B* 253 (1985), pp. 375–386.

[63] Howard Georgi, Helen R. Quinn, and Steven Weinberg. "Hierarchy of Interactions in Unified Gauge Theories". In: *Phys. Rev. Lett.* 33 (1974), pp. 451–454.

[64] Leonard Susskind. "Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory". In: *Phys. Rev. D* 20 (1979), pp. 2619–2625.

[65] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35.

[66] Patrick Huet and Eric Sather. "Electroweak baryogenesis and standard model CP violation". In: *Phys. Rev. D* 51 (1995), pp. 379–394. arXiv: hep-ph/9404302.

[67] Alexandre Alves et al. "Collider and Gravitational Wave Complementarity in Exploring the Singlet Extension of the Standard Model". In: *JHEP* 04 (2019), p. 052. arXiv: 1812.09333 [hep-ph].

[68] Riccardo Apreda et al. "Gravitational waves from electroweak phase transitions". In: *Nucl. Phys. B* 631 (2002), pp. 342–368. arXiv: gr-qc/0107033.

[69] Wei Su, Anthony G. Williams, and Mengchao Zhang. "Strong first order electroweak phase transition in 2HDM confronting future Z & Higgs factories". In: *JHEP* 04 (2021), p. 219. arXiv: 2011.04540 [hep-ph].

[70] Kohei Fujikura et al. "Cosmological phase transitions in composite Higgs models". In: *JHEP* 09 (2023), p. 053. arXiv: 2306.01305 [hep-ph].

[71] Ligong Bian, Yongcheng Wu, and Ke-Pan Xie. "Electroweak phase transition with composite Higgs models: calculability, gravitational waves and collider searches". In: *JHEP* 12 (2019), p. 028. arXiv: 1909.02014 [hep-ph].

[72] Stefano Profumo, Michael J. Ramsey-Musolf, and Gabe Shaughnessy. "Singlet Higgs phenomenology and the electroweak phase transition". In: *JHEP* 08 (2007), p. 010. arXiv: 0705.2425 [hep-ph].

[73] Eliel Camargo-Molina, Rikard Enberg, and Johan Löfgren. "A catalog of first-order electroweak phase transitions in the Standard Model Effective Field Theory". In: *JHEP* 08 (2025), p. 113. arXiv: 2410.23210 [hep-ph].

[74] LHC Higgs Cross Section Working Group et al. *LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle*. 2012. arXiv: 1209.0040 [hep-ph]. URL: https://arxiv.org/abs/1209.0040.

[75] ATLAS Collaboration. *HL-LHC prospects for the measurement of triple Higgs production in the 6b final state at the ATLAS experiment*. ATL-PHYS-PUB-2025-003. 2025. URL: https://cds.cern.ch/record/2924772.

[76] Daniel de Florian, Ignacio Fabre, and Javier Mazzitelli. "Triple Higgs production at hadron colliders at NNLO in QCD". In: *JHEP* 03 (2020), p. 155. arXiv: 1912.02760 [hep-ph].

[77] ATLAS and CMS Collaborations. "Highlights of the HL-LHC physics projections by ATLAS and CMS". In: (Apr. 2025). arXiv: 2504.00672 [hep-ex].

[78] Panagiotis Stylianou and Georg Weiglein. "Constraints on the trilinear and quartic Higgs couplings from triple Higgs production at the LHC and beyond". In: *Eur. Phys. J. C* 84.4 (2024), p. 366. arXiv: 2312.04646 [hep-ph].

[79] Wojciech Bizoń, Ulrich Haisch, and Luca Rottoli. "Constraints on the quartic Higgs self-coupling from double-Higgs production at future hadron colliders". In: *JHEP* 10 (2019), p. 267. arXiv: 1810.04665 [hep-ph].

[80] E. Aprile et al. "Search for New Physics in Electronic Recoil Data from XENONnT". In: *Phys. Rev. Lett.* 129.16 (2022), p. 161805. arXiv: 2207.11330 [hep-ex].

[81] J. Aalbers et al. "First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment". In: (July 2022). arXiv: 2207.03764 [hep-ex].

[82] J. Aalbers et al. "Dark Matter Search Results from 4.2 Tonne-Years of Exposure of the LUX-ZEPLIN (LZ) Experiment". In: *Phys. Rev. Lett.* 135.1 (2025), p. 011802. arXiv: 2410.17036 [hep-ex].

[83] Zihao Bo et al. “Dark Matter Search Results from 1.54 Tonne · Year Exposure of PandaX-4T”. In: *Phys. Rev. Lett.* 134.1 (2025), p. 011805. arXiv: 2408.00664 [hep-ex].

[84] P. Agnes et al. “DarkSide-50 532-day Dark Matter Search with Low-Radioactivity Argon”. In: *Phys. Rev. D* 98.10 (2018), p. 102006. arXiv: 1802.07198 [astro-ph.CO].

[85] D. S. Akerib et al. “Search for dark matter particle interactions in an extended nuclear recoil energy window with the LUX-ZEPLIN (LZ) experiment”. In: (Sept. 2026). arXiv: 2609.02823 [hep-ex].

[86] Carlos Pérez de los Heros. “Status, Challenges and Directions in Indirect Dark Matter Searches”. In: *Symmetry* 12.10 (2020), p. 1648. arXiv: 2008.11561 [astro-ph.HE].

[87] M. Ackermann et al. “Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data”. In: *Phys. Rev. Lett.* 115.23 (2015), p. 231301. arXiv: 1503.02641 [astro-ph.HE].

[88] M. Ackermann et al. “The Fermi Galactic Center GeV Excess and Implications for Dark Matter”. In: *Astrophys. J.* 840.1 (2017), p. 43. arXiv: 1704.03910 [astro-ph.HE].

[89] H. Abdalla et al. “Search for Dark Matter Annihilation Signals in the H.E.S.S. Inner Galaxy Survey”. In: *Phys. Rev. Lett.* 129.11 (2022), p. 111101. arXiv: 2207.10471 [astro-ph.HE].

[90] A. Acharyya et al. “Sensitivity of the Cherenkov Telescope Array to a dark matter signal from the Galactic centre”. In: *JCAP* 01 (2021), p. 057. arXiv: 2007.16129 [astro-ph.HE].

[91] A. Acharyya et al. “Search for Ultraheavy Dark Matter from Observations of Dwarf Spheroidal Galaxies with VERITAS”. In: *Astrophys. J.* 945.2 (2023), p. 101. arXiv: 2302.08784 [astro-ph.HE].

[92] M. G. Aartsen et al. “Search for neutrinos from decaying dark matter with IceCube”. In: *Eur. Phys. J. C* 78.10 (2018), p. 831. arXiv: 1804.03848 [astro-ph.HE].

[93] Daniel Abercrombie et al. “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”. In: *Phys. Dark Univ.* 27 (2020), p. 100371. arXiv: 1507.00966 [hep-ex].

[94] A. Djouadi et al. “The Minimal supersymmetric standard model: Group summary report”. In: *GDR (Groupement De Recherche) - Supersymetrie*. Dec. 1998. arXiv: hep-ph/9901246.

[95] Michael Duerr et al. “Hunting the dark Higgs”. In: *JHEP* 04 (2017), p. 143. arXiv: 1701.08780 [hep-ph].

[96] ATLAS Collaboration. “Search for pair production of higgsinos in events with two Higgs bosons and missing transverse momentum in s=13 TeV pp collisions at the ATLAS experiment”. In: *Phys. Rev. D* 109.11 (2024), p. 112011. arXiv: 2401.14922 [hep-ex].

[97] Armen Tumasyan et al. "Search for higgsinos decaying to two Higgs bosons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 05 (2022), p. 014. arXiv: 2201.04206 [hep-ex].

[98] Andreas Ekstedt et al. "Constraining minimal anomaly free U(1) extensions of the Standard Model". In: *JHEP* 11 (2016), p. 071. arXiv: 1605.04855 [hep-ph].

[99] ATLAS Collaboration. "Search for dark matter produced in association with a dark Higgs boson decaying into W^+W^- in the one-lepton final state at $\sqrt{s}$=13 TeV using 139 fb^{-1} of pp collisions recorded with the ATLAS detector". In: *JHEP* 07 (2023), p. 116. arXiv: 2211.07175 [hep-ex].

[100] ATLAS Collaboration. "Search for Dark Matter Produced in Association with a Dark Higgs Boson in the $b\bar{b}$ Final State Using pp Collisions at s=13 TeV with the ATLAS Detector". In: *Phys. Rev. Lett.* 134.12 (2025), p. 121801. arXiv: 2407.10549 [hep-ex].

[101] Aram Hayrapetyan et al. "Search for dark matter produced in association with a dark Higgs boson decaying into a bottom quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Apr. 2026). arXiv: 2604.20999 [hep-ex].

[102] ATLAS Collaboration. "Search for Dark Matter Produced in Association with a Dark Higgs Boson Decaying into $W^{\pm}W^{\mp}$ or ZZ in Fully Hadronic Final States from $\sqrt{s} = 13$ TeV pp Collisions Recorded with the ATLAS Detector". In: *Phys. Rev. Lett.* 126.12 (2021), p. 121802. arXiv: 2010.06548 [hep-ex].

[103] Aram Hayrapetyan et al. "Search for dark matter particles in W^+W^- events with transverse momentum imbalance in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 03 (2024), p. 134. arXiv: 2310.12229 [hep-ex].

[104] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001.

[105] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003.

[106] L Arnaudon et al. *Linac4 Technical Design Report*. Tech. rep. revised version submitted on 2006-12-14 09:00:40. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/1004186.

[107] M. Bajko et al. *Report of the Task Force on the Incident of 19 September 2008 at the LHC*. Tech. rep. LHC-PROJECT-Report-1168. Geneva: CERN, 2009. URL: https://cds.cern.ch/record/1168025.

[108] M Benedikt and F Zimmermann. "Future Circular Collider Integrated Programme". In: *EPJ Research Infrastructures* 10.1 (2026), p. 11. URL: https://cds.cern.ch/record/2962941.

[109] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005.

[110] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002.

[111] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004.

[112] Alessio Tiberio. "The LHCf experiment at the Large Hadron Collider: status and prospects". In: *PoS* ICRC2023 (2023), p. 444.

[113] Akitaka Ariga et al. “Technical Proposal for FASER: ForwArd Search ExpeRiment at the LHC”. In: (Dec. 2018). arXiv: 1812.09139 [physics.ins-det].

[114] G. Acampora et al. “SND@LHC: the scattering and neutrino detector at the LHC”. In: *JINST* 19.05 (2024), P05067. arXiv: 2210.02784 [hep-ex].

[115] Roel Aaij et al. “Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC”. In: *JINST* 14.04 (2019), P04013. arXiv: 1812.10790 [hep-ex].

[116] ATLAS Collaboration. “ATLAS data quality operations and performance for 2015–2018 data-taking”. In: *JINST* 15.04 (2020), P04003. arXiv: 1911.04632 [physics.ins-det].

[117] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77.7 (2017), p. 466. arXiv: 1703.10485 [hep-ex].

[118] ATLAS Collaboration. “Technical Design Report for the ATLAS Inner Tracker Pixel Detector”. In: (2017).

[119] ATLAS Collaboration. “The ATLAS inner detector trigger performance in pp collisions at 13 TeV during LHC Run 2”. In: *Eur. Phys. J. C* 82.3 (2022), p. 206. arXiv: 2107.02485 [hep-ex].

[120] M Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. 2010. URL: https://cds.cern.ch/record/1291633.

[121] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report Addendum*. Tech. rep. Addendum to CERN-LHCC-2010-013, ATLAS-TDR-019. 2012. URL: https://cds.cern.ch/record/1451888.

[122] ATLAS Collaboration. *Expected performance of the ATLAS b-tagging algorithms in Run-2*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-022. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2037697.

[123] Mogens Dam. “Detector requirements, design, and technologies for the FCC-ee Higgs, electroweak, and top factory”. In: *Nucl. Instrum. Meth. A* 1080 (2025), p. 170648. arXiv: 2505.06781 [hep-ex].

[124] Joao Pequenao. “Computer Generated image of the ATLAS calorimeter”. 2008. URL: https://cds.cern.ch/record/1095927.

[125] ATLAS Collaboration. “Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment”. In: *JHEP* 05 (2024), p. 162. arXiv: 2308.13362 [hep-ex].

[126] ATLAS Collaboration. “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2”. In: *Eur. Phys. J. C* 77.10 (2017), p. 673. arXiv: 1704.07983 [hep-ex].

[127] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *Eur. Phys. J. C* 77.5 (2017), p. 332. arXiv: 1611.10235 [physics.ins-det].

[128] N. J. Buchanan et al. “Design and implementation of the Front End Board for the readout of the ATLAS liquid argon calorimeters”. In: *JINST* 3 (2008), P03004.

[129] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. arXiv: 1603.02934 [hep-ex].

[130] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. arXiv: 0802.1189 [hep-ph].

[131] ATLAS Collaboration. "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 81.8 (2021), p. 689. arXiv: 2007.02645 [hep-ex].

[132] ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *Eur. Phys. J. C* 76.11 (2016), p. 581. arXiv: 1510.03823 [hep-ex].

[133] ATLAS Collaboration. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 81.7 (2021), p. 578. arXiv: 2012.00578 [hep-ex].

[134] ATLAS Collaboration. "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 78 (2018), p. 903. arXiv: 1802.08168 [hep-ex].

[135] ATLAS Collaboration. "The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions". In: *Eur. Phys. J. C* 85.6 (2025), p. 606. arXiv: 2402.05858 [hep-ex].

[136] ATLAS Collaboration. "ATLAS high-level trigger, data acquisition and controls: Technical design report". In: (July 2003).

[137] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15.10 (2020), P10004. arXiv: 2007.12539 [physics.ins-det].

[138] ATLAS Collaboration. "Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking". In: *JHEP* 08 (2020), p. 080. arXiv: 2005.09554 [hep-ex].

[139] ATLAS Collaboration. "Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking". In: *JHEP* 08 (2020), p. 080. arXiv: 2005.09554 [hep-ex].

[140] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933), pp. 289–337.

[141] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: https://www.deeplearningbook.org/.

[142] B. Brandl et al. "Multivariate analysis methods to tag b quark events at LEP / SLC". In: *Nucl. Instrum. Meth. A* 324 (1993), pp. 307–316.

[143] Warren S. McCulloch and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity". In: *The Bulletin of Mathematical Biophysics* 5.4 (1943), pp. 115–133.

[144] Frank Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”. In: *Psychological Review* 65.6 (1958), pp. 386–408.

[145] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[146] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[147] Benedek Rozemberczki et al. *The Shapley Value in Machine Learning*. 2022. arXiv: 2202.05594 [cs.LG].

[148] Selahattin Akkas and Ariful Azad. “GNNShap: Scalable and Accurate GNN Explanation using Shapley Values”. In: *Proceedings of the ACM Web Conference 2024*. WWW ‘24. ACM, May 2024, pp. 827–838. URL: http://dx.doi.org/10.1145/3589334.3645599.

[149] ATLAS Collaboration. “Explaining the ATLAS GN2 Flavour Tagging Algorithm with Integrated Gradients”. In: (2025).

[150] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. CERN, 2012. URL: https://cds.cern.ch/record/1456844.

[151] M. Baak et al. “HistFitter software framework for statistical data analysis”. In: *Eur. Phys. J. C* 75 (2015), p. 153. arXiv: 1410.1280 [hep-ex].

[152] F. James and M. Roos. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367.

[153] Rene Brun and Fons Rademakers. “ROOT — An object oriented data analysis framework”. In: *Nucl. Instrum. Meth. A* 389.1-2 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86.

[154] Samuel S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62.

[155] R. A. Fisher. “On the Mathematical Foundations of Theoretical Statistics”. In: *Philosophical Transactions of the Royal Society of London. Series A* 222 (1922), pp. 309–368.

[156] C. R. Rao. “Information and the Accuracy Attainable in the Estimation of Statistical Parameters”. In: *Bulletin of the Calcutta Mathematical Society* 37 (1945), pp. 81–91.

[157] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. arXiv: 1007.1727 [physics.data-an]. Erratum: in: *Eur. Phys. J. C* 73 (2013), p. 2501.

[158] Alexander L. Read. “Presentation of search results: the CL_S technique”. In: *J. Phys. G* 28 (2002), p. 2693.

[159] Eamonn Maguire, Lukas Heinrich, and Graeme Watt. “HEPData: a repository for high energy physics data”. In: *J. Phys. Conf. Ser.* 898.10 (2017). Ed. by Richard Mount and Craig Tull, p. 102006. arXiv: 1704.05473 [hep-ex].

[160] "SimpleAnalysis: Truth-level Analysis Framework". In: (2022).

[161] ATLAS Collaboration. "ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset". In: *Eur. Phys. J. C* 83.7 (2023), p. 681. arXiv: 2211.16345 [physics.data-an].

[162] ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-011. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2270366.

[163] ATLAS Collaboration. *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-025. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2645405.

[164] ATLAS Collaboration. *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-014. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2718948.

[165] Fabrizio Caola et al. "Flavored jets with exact anti-kt kinematics and tests of infrared and collinear safety". In: *Phys. Rev. D* 108.9 (2023), p. 094010. arXiv: 2306.07314 [hep-ph].

[166] ATLAS Collaboration. "Measurements of WH and ZH production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp collisions with the ATLAS detector". In: *JHEP* 04 (2025), p. 075. arXiv: 2410.19611 [hep-ex].

[167] Chris Pollard and Philipp Windischhofer. "Transport away your problems: Calibrating stochastic simulations with optimal transport". In: *Nucl. Instrum. Meth. A* 1027 (2022), p. 166119. arXiv: 2107.08648 [physics.data-an].

[168] ATLAS Collaboration. "A continuous calibration of the ATLAS flavour-tagging classifiers via optimal transportation maps". In: *Eur. Phys. J. C* 85.11 (2025), p. 1272. arXiv: 2505.13063 [hep-ex].

[169] ATLAS Collaboration. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.11 (2019), p. 970. arXiv: 1907.05120 [hep-ex].

[170] ATLAS Collaboration. "Measurement of the b-jet identification efficiency in dileptonic $t\bar{t}$ events using proton–proton collision data at $\sqrt{s} = 13.6$ TeV collected with the ATLAS detector". In: (July 2026). arXiv: 2607.05322 [hep-ex].

[171] ATLAS Collaboration. "Simultaneous efficiency measurements of b- and c-jets in $t\bar{t}$ events from $\sqrt{s} = 13.6$ TeV pp collision data collected with the ATLAS detector". In: (July 2026). arXiv: 2607.06199 [hep-ex].

[172] ATLAS Collaboration. *Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2314418.

[173] ATLAS Collaboration. *Measuring the b-jet identification efficiency for high p_T jets using multijet events in proton−proton collisions at $\sqrt{s} = 13$ TeV recorded with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-010. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2804062.

[174] ATLAS Collaboration. "Simulation-based extrapolation of b-tagging calibrations towards high transverse momenta in the ATLAS experiment". In: (2021). All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-003. URL: https://cds.cern.ch/record/2753444.

[175] ATLAS Collaboration. "Alignment of the ATLAS Inner Detector in Run-2". In: *Eur. Phys. J. C* 80.12 (2020), p. 1194. arXiv: 2007.07624 [hep-ex].

[176] Frédéric A. Dreyer, Gavin P. Salam, and Grégory Soyez. "The Lund Jet Plane". In: *JHEP* 12 (2018), p. 064. arXiv: 1807.04758 [hep-ph].

[177] Giacinto Piacquadio et al. *Eigenvector recomposition: a new method to correlate flavor-tagging systematic uncertainties across analyses*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-024. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2791430.

[178] Georges Aad et al. "Search for heavy Higgs bosons with flavour-violating couplings in multi-lepton plus b-jets final states in pp collisions at 13 TeV with the ATLAS detector". In: *JHEP* 12 (2023), p. 081. arXiv: 2307.14759 [hep-ex].

[179] ATLAS Collaboration. *GN3: Multi-task, Multi-modal Transformers for Jet Flavour Tagging in ATLAS*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2026-001. Geneva: CERN, 2026. URL: https://cds.cern.ch/record/2953652.

[180] T. Sjöstrand, S. Mrenna, and P. Skands. "A brief introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. arXiv: 0710.3820 [hep-ph].

[181] NNPDF Collaboration, Richard D. Ball, et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867 (2013), p. 244. arXiv: 1207.1303 [hep-ph].

[182] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2206965.

[183] ATLAS Collaboration. *ATLAS Run 1 Pythia8 tunes*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1966419.

[184] ATLAS Collaboration. *Monte Carlo generator comparisons to ATLAS measurements of vector boson production with jets*. Tech. rep. Geneva: CERN, 2016.

[185] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. arXiv: 1410.3012 [hep-ph].

[186] Enrico Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. arXiv: 1905.09127 [hep-ph].

[187] Tanju Gleisberg and Stefan Höche. "Comix, a new matrix element generator". In: *JHEP* 12 (2008), p. 039. arXiv: 0808.3674 [hep-ph].

[188] Ryan Gavin et al. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". In: *Comput. Phys. Commun.* 182 (2011), p. 2388. arXiv: 1011.3540 [hep-ph].

[189] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". In: *Eur. Phys. J. C* 71 (2011), p. 1547. arXiv: 1009.2450 [hep-ph].

[190] Michal Czakon and Alexander Mitov. "Top++: A program for the calculation of the top-pair cross-section at hadron colliders". In: *Comput. Phys. Commun.* 185 (2014), p. 2930. arXiv: 1112.5675 [hep-ph].

[191] M. Aliev et al. "HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR". In: *Comput. Phys. Commun.* 182 (2011), pp. 1034–1046. arXiv: 1007.1327 [hep-ph].

[192] P. Kant et al. "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions". In: *Comput. Phys. Commun.* 191 (2015), pp. 74–89. arXiv: 1406.4403 [hep-ph].

[193] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: *Eur. Phys. J. C* 83 (2023), p. 982. arXiv: 2212.09379 [hep-ex].

[194] ATLAS Collaboration. "ATLAS data quality operations and performance for 2015–2018 data-taking". In: *JINST* 15.04 (2020), P04003. arXiv: 1911.04632 [physics.ins-det].

[195] ATLAS Collaboration. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data". In: *JINST* 14 (2019), P12006. arXiv: 1908.00005 [hep-ex].

[196] ATLAS Collaboration. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 81 (2021), p. 578. arXiv: 2012.00578 [hep-ex].

[197] ATLAS Collaboration. "Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton-proton run". In: *JINST* 8 (2013), P07004. arXiv: 1303.0223 [hep-ex].

[198] ATLAS Collaboration. *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-029. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2037702.

[199] ATLAS Collaboration. "The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions". In: *Eur. Phys. J. C* 85.6 (2025), p. 606. arXiv: 2402.05858 [hep-ex].

[200] ATLAS Collaboration. *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-021. Geneva: CERN, 2023. URL: https://cds.cern.ch/record/2866601.

[201] Kunihiko Fukushima. "Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements". In: *IEEE Transactions on Systems Science and Cybernetics* 5.4 (1969), pp. 322–333.

[202] Vinod Nair and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, 2010, pp. 807–814.

[203] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

[204] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *JMLR* 15.56 (2014), pp. 1929–1958. URL: https://jmlr.org/papers/v15/srivastava14a.html.

[205] M. Stone. "Cross-Validatory Choice and Assessment of Statistical Predictions". In: *J.Roy.Statist.Soc.B* 36.2 (1974), pp. 111–147.

[206] James Max Kanter and Kalyan Veeramachaneni. "Deep feature synthesis: Towards automating data science endeavors". In: *2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, Paris, France, October 19-21, 2015*. IEEE. 2015, pp. 1–10.

[207] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. "Measuring and testing dependence by correlation of distances". In: *The Annals of Statistics* 35.6 (2007), pp. 2769–2794. URL: https://doi.org/10.1214/009053607000000505.

[208] Gregor Kasieczka and David Shih. "DisCo Fever: Robust Networks Through Distance Correlation". In: *Phys. Rev. Lett.* 125.12 (2020), p. 122001. arXiv: 2001.05310 [hep-ph].

[209] ATLAS Collaboration. *Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} = 13$ TeV ATLAS analyses*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-005. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2120826.

[210] Stefano Frixione et al. "Single-top hadroproduction in association with a W boson". In: *JHEP* 07 (2008), p. 029. arXiv: 0805.3067 [hep-ph].

[211] ATLAS Collaboration. "Studies on top-quark Monte Carlo modelling for Top2016". In: (Oct. 2016).

[212] Enrico Bothmann, Marek Schönherr, and Steffen Schumann. "Reweighting QCD matrix-element and parton-shower calculations". In: *Eur. Phys. J. C* 76.11 (2016), p. 590. arXiv: 1606.08753 [hep-ph].

[213] ATLAS Collaboration. *Formulae for estimating significance in searches for new phenomena*. Tech. rep. Geneva: CERN, 2020.

[214] Federico Ambrogi et al. "MadDM v.3.0: a Comprehensive Tool for Dark Matter Studies". In: *Phys. Dark Univ.* 24 (2019), p. 100249. arXiv: 1804.00044 [hep-ph].

[215] ATLAS Collaboration. "Search for triple Higgs boson production in the 6b final state using pp collisions at s=13 TeV with the ATLAS detector". In: *Phys. Rev. D* 111.3 (2025), p. 032006. arXiv: 2411.02040 [hep-ex].

[216] *Search for nonresonant triple Higgs boson production in the six b-quark final state in proton-proton collisions at 13 TeV*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2945361.

[217] "Search for triple Higgs production in Run2 data of CMS using 4b2gamma final state." In: (2025).

[218] ATLAS and CMS Collaborations. *Combination of ATLAS and CMS searches for Higgs boson pair production at 13 TeV*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2947521.

[219] A. Abada et al. "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1". In: *Eur. Phys. J. C* 79.6 (2019), p. 474.

[220] David d'Enterria, Andres Poldaru, and George Wojcik. "Measuring the electron Yukawa coupling via resonant s-channel Higgs production at FCC-ee". In: *Eur. Phys. J. Plus* 137.2 (2022), p. 201. arXiv: 2107.02686 [hep-ex].

[221] ATLAS Collaboration. "Searches for direct slepton production in the compressed-mass corridor in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". In: *JHEP* 08 (2025), p. 053. arXiv: 2503.17186 [hep-ex].

[222] M. Aleksa et al. "Conceptual design of an experiment at the FCC-hh, a future 100 TeV hadron collider". In: 2/2022 (Nov. 2022). Ed. by Michelangelo L. Mangano and Werner Riegler.

[223] Michele Selvaggi and Birgit Stapf. *A Delphes parameterisation of the FCC-hh detector*. 2025.

[224] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph].

[225] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. arXiv: 1212.3460 [hep-ph].

[226] L. A. Harland-Lang et al. "Ad Lucem: QED Parton Distribution Functions in the MMHT Framework". In: *Eur. Phys. J. C* 79.10 (2019), p. 811. arXiv: 1907.02750 [hep-ph].

[227] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. arXiv: 0710.3820 [hep-ph].

[228] Olivier Mattelaer. "On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy". In: *Eur. Phys. J. C* 76.12 (2016), p. 674. arXiv: 1607.00763 [hep-ph].

[229] Gerardo Ganis, Clément Helsens, and Valentin Völkl. "Key4hep, a framework for future HEP experiments and its use in FCC". In: *Eur. Phys. J. Plus* 137.1 (2022), p. 149. arXiv: 2111.09874 [hep-ex].

[230] T. Binoth et al. "Multi-Higgs boson production in the standard model and beyond". In: *Physical Review D* 74.11 (Dec. 2006). URL: http://dx.doi.org/10.1103/PhysRevD.74.113008.

[231] John M. Campbell, Giuseppe De Laurentis, and R. Keith Ellis. *An analytic result for the* $0 \to ggHHH$ *amplitude*. 2025. arXiv: 2507.19313 [hep-ph]. URL: https://arxiv.org/abs/2507.19313.

[232] Wojciech Bizoń et al. "Addendum to: Constraints on the quartic Higgs self-coupling from double-Higgs production at future hadron colliders [JHEP 10 (2019) 267]". In: *JHEP* 02 (2024), p. 170. arXiv: 2402.03463 [hep-ph].

[233] G. Heinrich et al. "NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers". In: *JHEP* 08 (2017), p. 088. arXiv: 1703.09252 [hep-ph].

[234] G. Heinrich et al. "Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects". In: *JHEP* 06 (2019), p. 066. arXiv: 1903.08137 [hep-ph].

[235] Johan Alwall et al. "MadGraph 5: going beyond". In: *JHEP* 06 (2011), p. 128. arXiv: 1106.0522 [hep-ph].

[236] A. Elagin et al. "A New Mass Reconstruction Technique for Resonances Decaying to di-tau". In: *Nucl. Instrum. Meth. A* 654 (2011), pp. 481–489. arXiv: 1012.4686 [hep-ex].

[237] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: 1912.01703 [cs.LG]. URL: https://arxiv.org/abs/1912.01703.

[238] Takuya Akiba et al. *Optuna: A Next-generation Hyperparameter Optimization Framework*. 2019. arXiv: 1907.10902 [cs.LG]. URL: https://arxiv.org/abs/1907.10902.

[239] Mohamed Aly et al. *TRExFitter (1.2.0)*. 2025.

[240] ATLAS Collaboration. "Combination of Searches for Higgs Boson Pair Production in pp Collisions at s=13 TeV with the ATLAS Detector". In: *Phys. Rev. Lett.* 133.10 (2024), p. 101801. arXiv: 2406.09971 [hep-ex].

[241] ATLAS Collaboration. "Search for triple Higgs boson production in the 6b final state using pp collisions at s=13 TeV with the ATLAS detector". In: *Phys. Rev. D* 111.3 (2025), p. 032006. arXiv: 2411.02040 [hep-ex].

[242] Angela Taliercio et al. *Higgs self-coupling determination in the bbaa final state at the FCC-hh*. Mar. 2025. URL: https://doi.org/10.17181/5mqfv-xnd34.

[243] John C. Collins and Davison E. Soper. "Angular Distribution of Dileptons in High-Energy Hadron Collisions". In: *Phys. Rev. D* 16 (1977), p. 2219.

[244] Mauro Chiesa et al. "Measuring the quartic Higgs self-coupling at a multi-TeV muon collider". In: *JHEP* 09 (2020), p. 098. arXiv: 2003.13628 [hep-ph].

[245] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Strip Detector*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257755.

[246] ATLAS Collaboration. *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2719855.

[247] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285584.

[248] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285582.

[249] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285583.

[250] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285580.