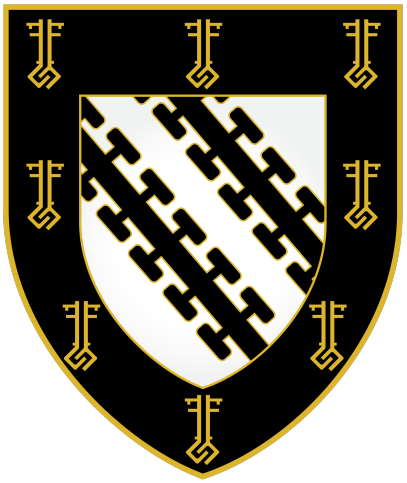
Searching for Dark Matter in the Light of Dark-Higgs Strahlung

A first search for a resonant di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ signature with the ATLAS experiment



Tim Lukas Brückler

Exeter College, Oxford

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Oxford

Michaelmas 2024

CERN-THESIS-2024-310
09/12/2024

Abstract

This thesis summarises the research conducted by the author on the first search for a new particle decaying in a di-Higgs signature with a significant missing momentum component. The work presented utilises the 140 fb^{-1}, $\sqrt{s} = 13$ TeV proton-proton collision dataset collected with the ATLAS detector. The analysis targets new physics in the di-Higgs decay into four-b channel and exploits the unique kinematic features of this resonant decay topology. In particular, the search is optimised for a scenario in which a hypothetical heavy Z' mediator is produced, radiating a dark-Higgs boson before decaying invisibly into dark matter particles. The dark-Higgs boson then decays into two Standard Model Higgs bosons: $pp \to Z'^{*} \to Z'(\to \chi\chi)s(\to hh(\to b\bar{b}b\bar{b}))$. The corresponding model is the Two-Mediator Dark-Higgs Model. Optimised kinematic selections, together with the innovative use of machine learning algorithms, are employed to achieve high sensitivities. Measures are explored to ensure that the machine learning classifier does not distort the reconstructed di-Higgs mass distribution, thus preserving its physical interpretability in the fit. As no significant excesses are observed, limits on the dark-Higgs model are presented in the dark-Higgs and Z' mass plane. Dark-Higgs masses are excluded between 250 and 405 GeV for Z' masses up to 2400 GeV in an benchmark model with a dark matter mass of 200 GeV. These derived limits significantly extend the current best constraints on this model, particularly in the high dark-Higgs and high Z' mass regions.

Additionally, this thesis presents a calibration analysis for a boosted double-b tagger, developed to identify two collimated b-jets originating from the decay of a heavy resonance. Scale factors are derived to correct for any differences in the mis-tagging efficiencies between simulation and data, using simulated dijet events enriched in $g \to b\bar{b}$ splittings.

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my supervisors, without whose invaluable input and guidance, the research presented in this thesis would not have been possible. I have spent countless hours discussing the details of the analyses as well as the broader context with Dr. James Frost - whether in his office, via various video conferencing platforms, or over lunch at Wagamama's. His profound insights into particle physics and the ATLAS experiment have been instrumental in shaping my research and thinking. I am also sincerely grateful to Prof. Todd Huffman, who was always available when I sought advice. His ability to ask the right questions prompted me to reassess and refine my ideas, a process from which I have greatly benefited.

Furthermore, I would like to thank my colleagues and friends in the Oxford ATLAS group for the stimulating input and the wonderful atmosphere they have provided over the years. Special thanks are due to Iza and Tom, with whom I collaborated closely on the two projects detailed in this thesis, and to Holly and Ondrej for their valuable contributions. I would also like to acknowledge Alessandro, Eimear, Maggie, Simon, Ynyr and many others, who not only provided valuable input throughout my DPhil but have also become good friends along the way.

I am equally thankful to the other physics friends I made at Oxford, CERN and various conference visits around the globe, including Ben, Chris, Daniela, Dom, Estelle, Florian, Giovanna, Hank, Jenny, Josh, Lars, Mei and Naeema - to name but a few. My time at Oxford was also enriched by the friendships I formed in the Oxford Austrian Society, with Felix, Veronika, and many others. I am also grateful to the wonderful people I shared accommodation with during my DPhil, whether at College in block R, at Meyrin

in Switzerland, or near Southpark in Oxford. Additionally, I am deeply appreciative of the friends I made at College, as well as through rowing and climbing. A special mention goes to the friends from the ECBC, including Fiona, Manuel, Milko, Min, and many more. I am also thankful to my friends back home, such as Annerose, Mark, Matthias, Priya and Stefan, as well as the many friends I made in London, including Hannah, Huw, Jess, and Julian, who provided a much-needed distraction from the stresses of academic life.

I will forever be grateful for the unwavering support of my family. I am profoundly thankful for the love and support of my grandparents, whose presence has always been a source of strength. Especially the memory of my two grandfathers, Rudi and Toni, continues to inspire me. My parents, Andrea and Christian, have been my steadfast supporters, providing more encouragement than I could have ever hoped for. My brother Alexander, who always enticed me into thrilling adventures in the mountains whenever I visited, Nikita and Klaus, and my parents-in-law, Sunita and Brahma, have all been essential to this journey.

Lastly, to my wife Parul, who has been by my side since the first day of university and who embarked on the adventure of moving to a foreign country with me. Even in the most challenging times, her support has been unfailing, and for that, I am eternally thankful.

Contents

Abstract i

Acknowledgements ii

Contents iv

Preface and Authors Contributions 1

I Theory and Experiment 3

1 The Standard Model and Beyond 4

1.1 The Standard Model of Particle Physics 5

1.1.1 The Fundamental Particles and Interactions 5

1.1.2 The Electromagnetic Interaction 7

1.1.3 The Electroweak Unification 8

1.1.4 The Higgs Mechanism 9

1.1.5 Fermion Masses 12

1.1.6 The Strong Force 13

1.1.7 The Standard Model Lagrangian 15

1.1.8 The Free Parameters of the Standard Model 17

1.2 Shortcomings of the Standard Model 19

1.2.1 Free Parameters 19

1.2.2 Neutrino Masses 19

1.2.3 The Hierarchy Problem 21

1.2.4 Unification of Forces . . . 22
1.2.5 Gravity . . . 23
1.2.6 The Strong CP Problem . . . 24
1.2.7 Matter-Antimatter Asymmetry . . . 25
1.2.8 Dark Matter . . . 25
1.3 Dark Matter Searches: Di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 31
1.3.1 Dark Matter Searches at the LHC . . . 31
1.3.2 The Dark-Higgs Model . . . 34
1.3.3 Other Models . . . 40

2 The ATLAS Experiment at the LHC 46
2.1 The Large Hadron Collider . . . 47
2.1.1 The Accelerator Complex . . . 47
2.1.2 Experiments at the LHC . . . 49
2.1.3 The LHC Run 2 . . . 49
2.2 The ATLAS Detector . . . 52
2.2.1 The Coordinate System . . . 52
2.2.2 Particles in the ATLAS Detector . . . 54
2.2.3 The Inner Detector . . . 56
2.2.4 The Calorimeters . . . 61
2.2.5 The Muon Spectrometer . . . 64
2.2.6 The Trigger System . . . 66
2.3 Event Simulation . . . 69
2.3.1 The Hard-Scatter Interaction . . . 71
2.3.2 Parton Shower . . . 72
2.3.3 Hadronisation . . . 73
2.3.4 The Underlying Event . . . 73
2.3.5 Hadron Decays . . . 74
2.3.6 Simulation of the ATLAS Detector . . . 75
2.4 Reconstruction of Physics Objects . . . 76

2.4.1 Low-Level Objects . . . 76
2.4.2 Jets . . . 79
2.4.3 Muons . . . 88
2.4.4 Missing Transverse Energy . . . 89

II Statistics and Machine Learning 92

3 Statistical Methods 93

3.1 The Likelihood Function . . . 94
3.1.1 Definition . . . 94
3.1.2 The Binned Likelihood Function . . . 95
3.1.3 Systematic Uncertainties . . . 96
3.1.4 Maximum Likelihood Fits . . . 97
3.2 Hypothesis Testing . . . 98
3.2.1 The Test Statistic: Profile Likelihood Ratio . . . 98
3.2.2 Searches for New Physics . . . 100
3.3 The Asimov Dataset . . . 103

4 Machine Learning 104

4.1 Machine Learning in High Energy Physics . . . 105
4.2 Neural Networks . . . 106
4.2.1 The Perceptron . . . 106
4.2.2 Multi-Layer Perceptrons . . . 108
4.2.3 Training Neural Networks: Backpropagation . . . 109
4.2.4 Advanced Neural Network Architectures . . . 110
4.3 Overfitting and Regularisation . . . 112
4.3.1 Bias-Variance Dilemma . . . 113
4.3.2 Curse of Dimensionality . . . 114
4.3.3 Regularisation Methods . . . 114
4.4 Shapley Values . . . 115

III A Search for the dark-Higgs Boson 117

5 Introduction and Analysis Setup 118

5.1 Introduction . . . 119
5.2 Data Samples . . . 119
5.3 Monte Carlo Samples . . . 120
5.3.1 Signal Samples . . . 120
5.3.2 Standard Model Background Samples . . . 121
5.4 Trigger Strategy . . . 124
5.5 Reconstructed Physics Objects . . . 125
5.5.1 Jets . . . 125
5.5.2 Electrons . . . 126
5.5.3 Muons . . . 126
5.5.4 Missing Transverse Energy . . . 126
5.5.5 Overlap Removal . . . 127

6 Event Selection and Higgs Reconstruction 129

6.1 Event Cleaning . . . 130
6.2 Preselection . . . 130
6.2.1 Lepton Vetoes . . . 130
6.2.2 Kinematic Preselection . . . 131
6.3 Optimised Selection . . . 131
6.3.1 Discriminating Kinematic Variables . . . 132
6.3.2 The QCD Multijet Background . . . 135
6.3.3 Kinematic Dependence on Signal Masses . . . 136
6.4 Dark-Higgs Reconstruction . . . 137

7 Machine Learning Analysis 139

7.1 Introduction . . . 140
7.1.1 Signal Kinematics . . . 140
7.1.2 Training Strategy . . . 140

7.2 The Training Dataset . . . 141
7.2.1 The Right Balance . . . 141
7.2.2 Cross-Validation . . . 143
7.2.3 Input Variables . . . 143
7.2.4 Preprocessing of Training Data . . . 147
7.3 Correlation of m_{hh} and Output-Score . . . 147
7.3.1 The Issue: Sculpting . . . 147
7.3.2 First Approach: Adversarial Network . . . 148
7.3.3 Second Approach: Direct Decorrelation . . . 151
7.4 The Training of the Neural Network . . . 153
7.4.1 Hyperparameters . . . 153
7.4.2 Regularisation . . . 155
7.4.3 The Neural Network Output-Scores . . . 155

8 Analysis Regions and Background Estimation 157
8.1 Signal Regions . . . 158
8.1.1 Model-Dependent Exclusion Signal Region . . . 158
8.1.2 Model-Independent Discovery Regions . . . 159
8.2 Background Estimation Strategy . . . 159
8.2.1 $Z+$jets Control Region . . . 160
8.2.2 Control Regions for Top-Quark Processes . . . 161
8.2.3 Validation Region . . . 163
8.3 Systematic Uncertainties . . . 165
8.3.1 Experimental Systematic Uncertainties . . . 165
8.3.2 Theoretical Uncertainties . . . 165

9 Results 168
9.1 The Fit Setup . . . 169
9.2 Validation of the Fit Setup . . . 169
9.2.1 Normalisation Factors: . . . 170

9.2.2 Nuisance Parameters: . . . 172
9.2.3 Postfit Agreement: . . . 173
9.3 The Discovery Fit . . . 173
9.4 Results of the Exclusion Fit . . . 175
9.5 Conclusion . . . 177

IV Calibration of a Boosted Higgs → $b\bar{b}$ Tagger 178

10 The $X \to b\bar{b}$ Algorithm 179
10.1 Introduction . . . 180
10.2 Samples, Objects, and Selection . . . 180
10.3 The Machine Learning Algorithm . . . 181
10.3.1 Training and Test Dataset . . . 182
10.3.2 Input Features . . . 182
10.3.3 The Neural Network . . . 182
10.4 Results . . . 183

11 The $g \to b\bar{b}$ Calibration 185
11.1 Introduction . . . 186
11.2 Data and Simulation . . . 187
11.3 Triggers, Objects, and Selections . . . 187
11.3.1 Triggers . . . 187
11.3.2 Physics Objects and Candidate Selection . . . 188
11.4 Flavour Discriminating Templates . . . 189
11.4.1 Labelling of Flavour Fractions . . . 190
11.4.2 Flavour Discriminating Variables . . . 191
11.5 The Likelihood Fit . . . 194
11.5.1 Scale Factors and Flavour Fraction Correction . . . 195
11.5.2 The Likelihood . . . 196
11.5.3 Analysis Regions . . . 197

11.6 Systematic Uncertainties . . . 199
11.6.1 Experimental Uncertainties . . . 199
11.6.2 Theoretical Uncertainties . . . 200
11.6.3 Smoothing and Pruning . . . 201
11.7 The Prefit Templates . . . 201
11.7.1 $\langle s_{d_0} \rangle$ Templates . . . 201
11.7.2 Secondary Vertex Mass Templates . . . 202

12 Results of the Calibration **207**
12.1 Overview . . . 208
12.2 Scale Factors and Normalisations . . . 208
12.2.1 $\langle s_{d_0} \rangle$ Fit . . . 208
12.2.2 Secondary Vertex Mass Fit . . . 210
12.3 Mismodelling and Nuisance Parameters . . . 211
12.3.1 $\langle s_{d_0} \rangle$ Fit . . . 211
12.3.2 Secondary Vertex Mass Fit . . . 212
12.4 The Postfit Templates . . . 214
12.5 Conclusion . . . 218

Thesis Summary and Outlook **219**

Bibliography **220**

List of Acronyms **237**

Preface and Authors Contributions

This thesis presents both the author's research and a comprehensive overview of the theoretical concepts, the experimental setup, and methodologies that underpin it.

Part I: Chapter 1 introduces the fundamental theoretical concepts upon which the research in the later chapters is based. It begins with an introduction to the Standard Model of Particle Physics, highlighting both its successes and limitations. This chapter also explores beyond the Standard Model signals that could result in a final state with four b-jets and a significant missing momentum component. A particular emphasis is placed on the dark-Higgs model. Given the author's involvement in an analysis targeting Higgsino models [1], reinterpretation efforts involving long-lived particles and other supersymmetric scenarios, as well as preliminary studies of the tri-scalar model, these models—producing similar final states to the dark-Higgs model—are briefly introduced in Section 1.3.3, but are otherwise not further explored in this thesis. Chapter 2 describes the experimental setup, including an overview of the Large Hadron Collider and the ATLAS experiment, along with its various subdetectors. The chapter also covers the simulation of physics events and the reconstruction and calibration of physics objects from the collected data.

Part II: Chapter 3 and Chapter 4 provide an overview of the essential statistical and machine learning methods employed in high-energy physics, with a particular focus on those utilised in this thesis. While these chapters cover topics more broadly than is typical in a thesis, they are included to provide a thorough understanding of the methodologies used. Readers already familiar with these topics, or those primarily interested in the specific research findings, may wish to selectively read these chapters, focusing on the sections referenced in the later parts of the thesis.

Part III: Chapter 5 introduces the main project of this thesis: A search for beyond the Standard Model physics in final states with four b-jets and significant missing momentum,

interpreted in terms of the dark-Higgs model. Together with his supervisor, the author instigated and conceptualised the analysis, developing the analysis strategy from signal model through to a preliminary version of the statistical fit treatment. Once the team expanded to include others, he served as the primary analysis contact, coordinating the analysis and communicating its progress and results within the relevant ATLAS physics group. His contributions included investigating and validating the signal models, and processing data and Monte Carlo simulated samples, including systematic uncertainties. The author developed the kinematic selections (Chapter 6) and background estimation (Chapter 8). The machine learning approach (Chapter 7) was conceptualised entirely by the author, who also handled its implementation, validation and incorporation into the background estimation strategy. A preliminary fit configuration was prepared and tested by the author to demonstrate feasibility, before handing over final fit validation and systematic smoothing to another student. At the time of submitting this thesis, the analysis had been approved by ATLAS to unblind and was undergoing internal review within the collaboration, with publication expected around the author's thesis defence.

Part IV: This part is dedicated to the authors work on the multijet calibration analysis of the ATLAS $X \to b\bar{b}$ tagger. Chapter 10 introduces the tagging algorithm, without any direct involvement of the author. Chapter 11 details the calibration analysis, where the author contributed extensively, particularly in the study of the fit, systematic uncertainties, and flavour-discriminating templates. The author contributed also to earlier efforts concerning this calibration analysis, which are detailed in Ref. [2]. The author worked on this project collaboratively with another student, who has summarised her efforts with a primary focus on the alternative secondary vertex mass fit in Ref. [3]. All the results presented in Chapter 12 were produced by the author.

Publications by the author during the course of this DPhil:

- Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays. [2]
- Search for Higgsino pair production. [1]
- Search for dark-Higgs boson: In preparation.

Part I

Theory and Experiment

Chapter 1

The Standard Model and Beyond

Contents

1.1 The Standard Model of Particle Physics 5

1.2 Shortcomings of the Standard Model 19

1.3 Dark Matter Searches: Di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ 31

This chapter provides the essential theoretical foundation necessary to motivate the research presented in this thesis. It focuses on introducing the Standard Model (SM), including the Higgs mechanism, and offers a general overview of the concept of Dark Matter (DM). Additionally, it explores specific Beyond the Standard Model (BSM) theories that present viable DM candidates. The key sections essential for understanding the research presented in later chapters are Sections 1.1, 1.2.8, and 1.3.2.

1.1 The Standard Model of Particle Physics

The SM of particle physics is a theory describing the fundamental particles and three of the four fundamental interactions in nature, namely the electromagnetic, the weak and the strong force. It is built from multiple theories and ideas which are, together with experimental input, combined into a framework which has been remarkably successful in describing most phenomena in fundamental physics.

The elementary particles are implemented within a Quantum Field Theory (QFT) and the fundamental interactions arise from the principle of local gauge invariance. Special relativity and the Dirac equation [4] provide the tools for describing fermionic dynamics and the Higgs mechanism [5–7] generates particle masses. Additionally, experimental data is required to determine the values of numerous parameters in the theory, such as particle masses, couplings, the vacuum expectation value and mixing angles. While the SM has been confirmed by numerous experiments and is among the most tested theories in physics, it does leave some physical phenomena unexplained.

The following chapters will provide an overview of the most important theoretical concepts relevant to the research presented in this thesis. For a more detailed introduction to the SM, please refer to [8], which serves as a reference for this chapter and provides guidance for the used nomenclature.

1.1.1 The Fundamental Particles and Interactions

Fundamental particles are conceptualised as excited states of quantum fields in spacetime. Their defining characteristics, known as quantum numbers, are a result of underlying symmetries. Specifically, the quantum numbers associated with mass and spin are closely linked to the Poincaré symmetry, which ensures the consistency of physical laws across translations, rotations, and Lorentz boosts. Particles are classified into two categories based on their spin: fermions and bosons. There are twelve spin-1/2 particles, the fermions, in the SM. They are further categorised into three generations with similar properties but increasing mass. An overview of the fermions in the SM and their quantum numbers is

given in Table 1.1.

Type	Generation	Spin	Q	I	Y	Colour
Quarks	1st (u, d)	1/2	+2/3, -1/3	+1/2, -1/2	+1/3, +1/3	Yes
	2nd (c, s)	1/2	+2/3, -1/3	+1/2, -1/2	+1/3, +1/3	Yes
	3rd (t, b)	1/2	+2/3, -1/3	+1/2, -1/2	+1/3, +1/3	Yes
Leptons	1st (e, ν_e)	1/2	-1, 0	-1/2, +1/2	-1, -1	No
	2nd (μ, ν_μ)	1/2	-1, 0	-1/2, +1/2	-1, -1	No
	3rd (τ, ν_τ)	1/2	-1, 0	-1/2, +1/2	-1, -1	No

Table 1.1: Quantum numbers of the fundamental fermions in the SM: The electric charge Q, the weak isospin I and the weak hypercharge Y.

The dynamics of the fundamental particle fields are determined by minimising the action, a functional generated by the Lagrangian of the system. To avoid a dependence of the physical predictions on the chosen gauge, the Lagrangian must be invariant under certain gauge transformations. The transformations between the allowed gauges form a Lie group known as the symmetry group of the theory. The gauge symmetry group of the SM [9] is

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

For each generator of the symmetry group, a gauge field is introduced in the Lagrangian of the theory to ensure its invariance under local gauge transformations. A Lagrangian is said to be locally gauge invariant if it remains unaffected by different symmetry transformations applied at each point in spacetime. The quanta of these gauge fields are the gauge bosons.

The group $\mathrm{SU}(3)_C$ relates to the symmetry responsible for the strong force, dictating the interactions of quarks and gluons. Meanwhile, the combined symmetries of $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ are associated with the unified electroweak force. The generators of these groups introduce additional spin-1 vector boson fields, thereby extending the SM with the photon, three electroweak vector bosons, and eight gluons. See Figure 1.1 for the full particle content of the SM.

Building on Noether's theorem [10] on continuous symmetries it can be shown that

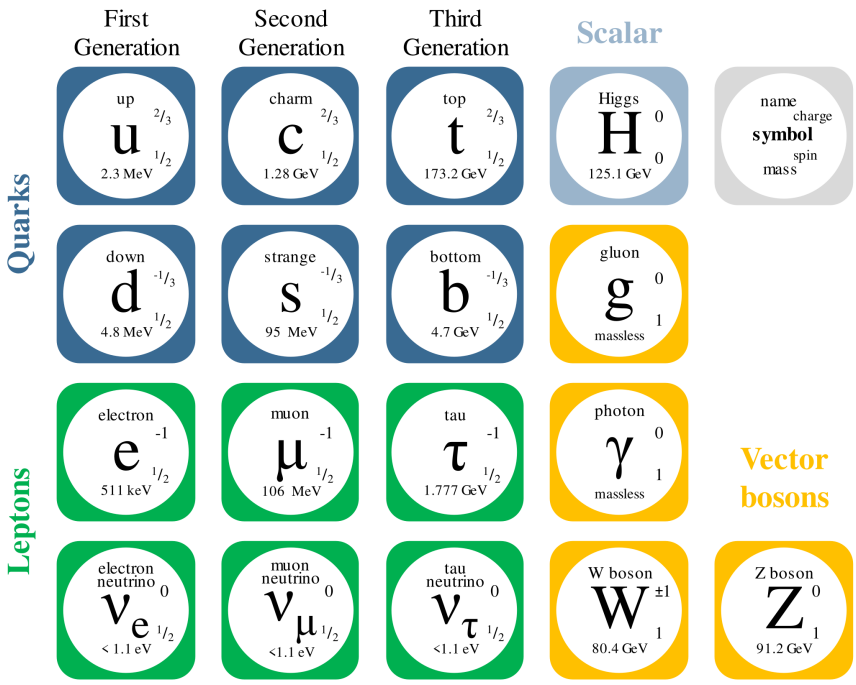


Figure 1.1: The fundamental particles in the SM.

the invariance of the SM Lagrangian under such gauge transformations necessitates the presence of more quantum numbers: electric charge, colour charge, weak isospin and weak hypercharge. Beyond these gauge symmetries, the SM Lagrangian also exhibits accidental symmetries, such as U(1) phase transformations resulting in the conservation of baryon, electron, muon, and tau numbers.

1.1.2 The Electromagnetic Interaction

Phenomena that involve the electromagnetic interaction of electrically charged particles through the exchange of photons are explained by the abelian gauge theory of Quantum Electrodynamics (QED) [11–13]. This theory is based on special relativity and quantum mechanics, and has the gauge symmetry Lie group U(1). The following provides a brief introduction to the concept of local gauge invariance, gauge fields and the covariant derivative using the example of a stand-alone electromagnetic interaction.

The dynamics of fermions are described by the Dirac equation and the corresponding

Lagrangian density

$$0 = (i\gamma^\mu \partial_\mu - m)\psi \tag{1.2}$$

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi \ , \tag{1.3}$$

where the Dirac spinor ψ denotes a spin-1/2 field and the γ^μ the four gamma matrices generating the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}I$. The tensor $g^{\mu\nu}$ represents the metric of the Minkowski spacetime. The Lagrangian density's invariance under local U(1) phase transformations

$$\psi(x) \to \psi'(x) = \psi(x) \exp^{iq\xi(x)} \tag{1.4}$$

is preserved by introducing the covariant derivative

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu \ . \tag{1.5}$$

The newly introduced field A_μ is identified as the photon field and transforms under a local phase transformation as

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \xi(x) \ . \tag{1.6}$$

Consequently, the Lagrangian density for QED is expressed as

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi - q\overline{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{1.7}$$

where the last term denotes the photon field's kinetic term with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ being the electromagnetic field strength tensor. This formalism and the concept of the covariant derivative can be extended to a complete electroweak theory and the strong force.

1.1.3 The Electroweak Unification

The Weinberg-Salam theory [14–16] explains how the electromagnetic and the weak force can be unified into the fundamental electroweak force. Although the two forces present

themselves separately at low energies, they are encompassed in a single theory. At very high energies, such as those present in the very early stages of the universe, they converge into a single force.

The combined electroweak symmetry group is $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. Four gauge fields, namely the weak hypercharge field B_μ and the three weak isospin fields $\mathbf{W}_\mu$ are introduced (similarly to Equation 1.5) to preserve the gauge invariance of the Lagrangian. The next section 1.1.4 will show that these fields are however not the physically observed ones. The resulting covariant derivative is expressed as

$$D_\mu = \partial_\mu + i\frac{g'}{2}B_\mu + i\frac{g''}{2}\mathbf{T}\cdot\mathbf{W}_\mu \ , \tag{1.8}$$

where the $\mathbf{T} = \frac{1}{2}\boldsymbol{\sigma}$ are the three generators of SU(2).

1.1.4 The Higgs Mechanism

While the photon is massless, the $W^\pm$ and Z bosons are notably massive. Similarly, all known fermions, including neutrinos, have non-zero mass. However, introducing a mass term directly into a Lagrangian, as shown in Equation 1.7, would violate this invariance. This raises the question of how particle masses can be explained while maintaining the necessary gauge invariance.

The following discussion will introduce the Higgs mechanism, offering an explanation for the masses of fundamental particles that preserves gauge invariance.

The Higgs potential is given by

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \ . \tag{1.9}$$

In the minimal Higgs model, the weak isospin doublet ϕ is constructed from two complex scalar fields:

$$\phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2\\ \phi_3 + i\phi_4\end{pmatrix} . \tag{1.10}$$

To generate masses for both neutral and charged gauge bosons, one scalar field is assumed to be neutral and the other one charged. For $\mu^2 < 0$, this potential has infinitely many stable degenerate non-zero minima at

$$\langle\phi\rangle = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}} \,, \tag{1.11}$$

where v denotes the vacuum expectation value (VEV) of the Higgs field. Below a certain temperature, the choice of a specific vacuum state breaks the global symmetry of the Lagrangian. To ensure a massless photon, only the lower neutral component of the isospin doublet can acquire a non-zero VEV. Hence, the vacuum state of the field ϕ and its fluctuations around it are described by

$$\langle 0|\,\phi\,|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix} \quad \rightarrow \quad \phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2\\ v + \eta(x) + i\xi(x)\end{pmatrix} \,. \tag{1.12}$$

To simplify calculations, the unitary gauge is applied to minimise the number of scalar degrees of freedom. This results in the three Goldstone bosons being *eaten* by the now massive gauge fields, thereby providing their longitudinal degrees of freedom. For instance, the Goldstone field $\xi(x)$ can be gauged away by applying the gauge transformation

$$\phi(x) \rightarrow \phi(x)\exp\left(\frac{-i\xi(x)}{v}\right) \,. \tag{1.13}$$

Hence, in the unitary gauge, the field ϕ is written as

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v + h(x)\end{pmatrix} \,. \tag{1.14}$$

The total Lagrangian density in this model, including the kinetic term, is expressed as

$$\mathcal{L} = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \,. \tag{1.15}$$

Using the electroweak covariant derivative from Equation 1.8, Equation 1.14 and $\mu^2 =$

$-\lambda v^2$, $\mathcal{L}$ takes the form

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2}\partial_\mu h \partial^\mu h - \lambda v^2 h^2 && (1.16)\\
&- \lambda v h^3 - \frac{\lambda}{4} h^4 && (1.17)\\
&+ \frac{\lambda}{4} v^4 && (1.18)\\
&+ (v+h)^2 \times \mathcal{I}_{gauge} \ . && (1.19)
\end{aligned}$$

The first line represents the Higgs boson with $m_h = \sqrt{2\lambda}v$, the second line its triple and quartic self-interactions and the third line is an irrelevant constant. $\mathcal{I}_{gauge}$ includes the mass and interaction terms of the gauge bosons. In particular, the mass terms take the form

$$\begin{aligned}
v^2 \mathcal{I}_{gauge} &= \frac{g'^2 v^2}{8}\left(W^1_\mu W^{1\mu} + W^2_\mu W^{2\mu}\right) + && (1.20)\\
&+ \frac{v^2}{8}\begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}\begin{pmatrix} g'^2 & -g'g'' \\ -g'g'' & g''^2 \end{pmatrix}\begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} \ . && (1.21)
\end{aligned}$$

The first line allows the determination of the mass of the W bosons as $m_W = \frac{g'v}{2}$ and the physical, charged gauge boson fields emerge as linear combinations of the W^i fields

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2) \ . \tag{1.22}$$

Diagonalising the mass matrix in Equation 1.21 results in one eigenvalue being zero, which is subsequently associated with the massless photon field. The second eigenvalue yields $m_Z = \frac{v}{2}\sqrt{g'^2 + g''^2}$. Writing the ratio of the two couplings as $\tan\theta_W = g'/g''$ enables the physical photon and Z fields to be expressed as a linear combination of the fundamental fields

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B^\mu \\ W^{3\mu} \end{pmatrix} \ . \tag{1.23}$$

This results in an interesting prediction for the ratio of the boson masses

$$\frac{m_W}{m_Z} = \cos\theta_W \ . \tag{1.24}$$

The couplings of the massive gauge bosons to the Higgs can be read off the $vh\mathcal{I}_{gauge}$ terms and are proportional to the boson masses.

1.1.5 Fermion Masses

Adding a fermionic mass term directly to the Lagrangian breaks the gauge symmetry as left- and right-handed components transform differently. However, it can be shown that the combinations $\bar{L}\phi R$ and $\bar{R}\phi L$ are invariant under SU(2) × U(1) transformations, where ϕ is the SU(2) doublet of the Higgs mechanism. In the unitary gauge, the mass term of e.g. the electron is thus expressed as

$$\mathcal{L}_e = -g_e(L^\dagger \phi e_R + e_R^\dagger \phi^\dagger L) \tag{1.25}$$

$$= -\frac{g_e}{\sqrt{2}} v(\overline{e}_L e_R + \overline{e}_R e_L) - \frac{g_e}{\sqrt{2}} h(\overline{e}_L e_R + \overline{e}_R e_L) \tag{1.26}$$

$$= -m_e \overline{e} e - \frac{m_e}{v} \overline{e} e h \ . \tag{1.27}$$

Here, $L = \binom{\nu_e}{e}$ is a left-handed doublet and e_R is a right-handed singlet under SU(2). Introducing the Yukawa coupling as $g_e = \frac{\sqrt{2} m_e}{v}$ gives the mass of the fermion as

$$m_f = \frac{g_f v}{\sqrt{2}} \ . \tag{1.28}$$

For down-type quarks and the other charged leptons the formalism is the same as for electrons. A similar result is obtained for up-type quarks by utilising the conjugate of the ϕ doublet. The coupling of the Higgs field to the fermions is, in accordance with the Higgs gauge boson coupling, proportional to the mass. Interestingly, the heaviest fermion, the top quark, has a mass of 172.52 ± 0.33 GeV [17] and thus $g_f \approx 1$.

1.1.6 The Strong Force

Our understanding of the strong force is based on the theory of Quantum Chromodynamics (QCD) [18, 19].

The strong force binds quarks together to form nucleons. Consequently, it plays a fundamental role in the stability and structure of matter encountered in everyday life. However, QCD becomes particularly important when analysing proton-proton or hadron-hadron collisions at the Large Hadron Collider (LHC). Specifically, the production of hadronic jets at the LHC, a key concept in this thesis, is a direct consequence of the unique properties of QCD.

The theory's mathematical structure is based on the non-Abelian gauge group $\mathrm{SU}(3)_C$. The dynamics of these interactions are encapsulated by the QCD Lagrangian:

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a + \sum_f \bar{\psi}_f (i\gamma^\mu D_\mu - m_f)\psi_f, \tag{1.29}$$

where $G^a_{\mu\nu}$ is the gluon field strength tensor, which encapsulates the dynamics of the gluon fields, and ψ_f represents the quark fields for each flavour f. The covariant derivative of QCD is introduced as

$$D_\mu = \partial_\mu - i g_s G^a_\mu T^a \ , \tag{1.30}$$

and ensures invariance under local SU(3) transformations. Here, G^a_μ denotes the gluon fields, T^a the generators of $SU(3)$, and g_s is the strong coupling constant.

The eight generators, associated with the eight gluons, couple to particles carrying colour charge, namely the quarks and the gluons themselves. The generators are closely related to the Gell-Mann matrices λ^a via $T^a = \lambda^a/2$ and can be represented by 3×3 matrices. While a single electric charge is conserved in QED, in QCD there are three conserved charges, corresponding to the three orthogonal states in SU(3) colour space. They are labelled by the three colours: red, green, and blue. Similar to QED, where anti-particles carry negative charge, anti-quarks carry "anti-colours".

The gluon fields $G^a_\mu(x)$ transform under gauge transformations as

$$G^a_\mu \to G^a_\mu - \partial_\mu \alpha(x)_a - g_S f_{ija} \alpha_i G^j_\mu \tag{1.31}$$

where the α_a are general functions of the space-time coordinates x. SU(3) is non-abelian, meaning its generators do not commute, resulting in the term that includes the structure functions f_{ija}. The latter are defined by the commutation relations of the Gell-Mann matrices. This term gives rise to gluon self-interactions through triple and quartic gluon vertices. Consequently, this results in a rich structure of QCD interactions and *phenomena*.

Colour confinement is a hypothesis stating that no objects carrying non-zero colour charge can propagate freely. So far, no colourless state has been observed experimentally. Quarks are thus confined within colourless hadrons. This phenomenon is a direct result of gluon self-interaction. It can qualitatively be explained by the concept of gluon tubes formed between two quarks as they are pulled apart. The virtual gluons between the two quarks self-interact and form a tube, storing an increasing amount of energy as the distance between the quarks grows. At a certain distance, the energy becomes sufficient to create a new quark-antiquark pair, resulting in the gluon tube breaking.

This process is called hadronisation and continues till the energies are low enough for all quarks to be bound in colourless states. For the LHC, this means that gluons or quarks produced in the interactions are always observed in the form of jets. Most calculations in particle physics are based on perturbation theory. However, the bosonic loop corrections from gluon self-interaction lead to a decreasing α_S with increasing transferred momentum scale Q:

$$\alpha_s(Q^2) \approx \frac{1}{\beta_0 \ln(Q^2/\Lambda^2)} \tag{1.32}$$

Consequently, the strong coupling constant takes values of $\mathcal{O}(1)$ at approximately 1 GeV, and thus pertubation theory cannot be used in the low energy regime. Therefore, to describe hadronisation processes at the LHC, non-perturbative models must be employed. The energies of typical hard scattering processes at the LHC are, however, high enough to allow the use of pertubation theory.

This phenomenon is called *asymptotic freedom* because quarks can be described as almost free particles at very high energies. The latest measurements result in a value for α_S at the scale of the Z boson mass of $\alpha_S(m_Z) = 0.1183 \pm 0.0009$ [20]. In contrast to QED, these values are large enough to necessitate higher order corrections to the matrix element calculations to provide accurate predictions for LHC physics.

1.1.7 The Standard Model Lagrangian

Merging insights from previous sections leads to the formulation of the SM Lagrangian. This essential mathematical framework captures the interactions and dynamics of particles within the SM, efficiently expressed through the use of covariant derivatives and field strength tensors introduced as:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.33}$$

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g'' \epsilon^{ijk} W^j_\mu W^k_\nu, \tag{1.34}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \ . \tag{1.35}$$

Consequently, the SM Lagrangian can be expressed as

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{fermion}} \ , \tag{1.36}$$

where the individual terms are

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W^{\mu\nu i} - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu a} \tag{1.37}$$

$$\mathcal{L}_{\text{Higgs}} = \left(D_\mu \phi\right)^\dagger \left(D^\mu \phi\right) - \mu^2 \phi^\dagger \phi - \lambda \left(\phi^\dagger \phi\right)^2 \tag{1.38}$$

$$\mathcal{L}_{\text{Yukawa}} = -\sum_f^{e,\mu,\tau,d,s,t} g_f \overline{L}_f \phi R_f - \sum_f^{u,c,b} g_f \overline{L}_f \tilde{\phi} R_f + (\text{h.c.}) \tag{1.39}$$

$$\mathcal{L}_{\text{fermion}} = \sum_f^{\text{all}} \overline{L}_f i\gamma^\mu D_\mu L_f + \sum_f^{\text{all}} \overline{R}_f i\gamma^\mu D_\mu R_f + (\text{h.c.}) \ . \tag{1.40}$$

Here, the action of the full SM gauge covariant derivative

$$D_\mu = \partial_\mu - i\frac{g'}{2}Y_W B_\mu - i\frac{g''}{2}\mathbf{W}_\mu \cdot \boldsymbol{\sigma} - i\frac{g_S}{2}\mathbf{G}_\mu \cdot \boldsymbol{\lambda} \tag{1.41}$$

depends on the specific field it acts upon:

$$D_\mu L_{\text{lep}} = D_\mu \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} = \left[\partial_\mu + i\frac{g'}{2}B_\mu - i\frac{g''}{2}\mathbf{W}_\mu \cdot \boldsymbol{\sigma}\right] \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} , \tag{1.42}$$

$$D_\mu L_{\text{quark}} = D_\mu \begin{pmatrix} u_L \\ d_L \end{pmatrix} = \left[\partial_\mu - i\frac{g'}{6}B_\mu - i\frac{g''}{2}\mathbf{W}_\mu \cdot \boldsymbol{\sigma} - i\frac{g_S}{2}\mathbf{G}_\mu \cdot \boldsymbol{\lambda}\right] \begin{pmatrix} u_L \\ d_L \end{pmatrix} , \tag{1.43}$$

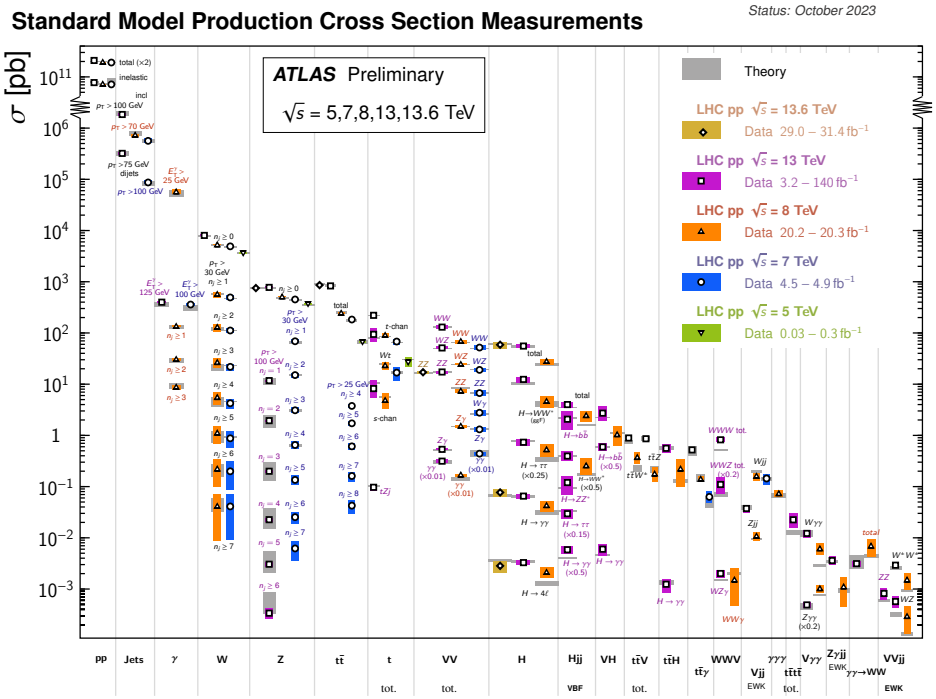
$$D_\mu R_{\text{charged lep}} = D_\mu e_R = \left[\partial_\mu - ig'B_\mu\right] e_R , \tag{1.44}$$

$$D_\mu R_{\text{quark}} = D_\mu u_R = \left[\partial_\mu - i\frac{g'}{3}Y_W B_\mu - i\frac{g_S}{2}\mathbf{G}_\mu \cdot \boldsymbol{\lambda}\right] u_R, \tag{1.45}$$

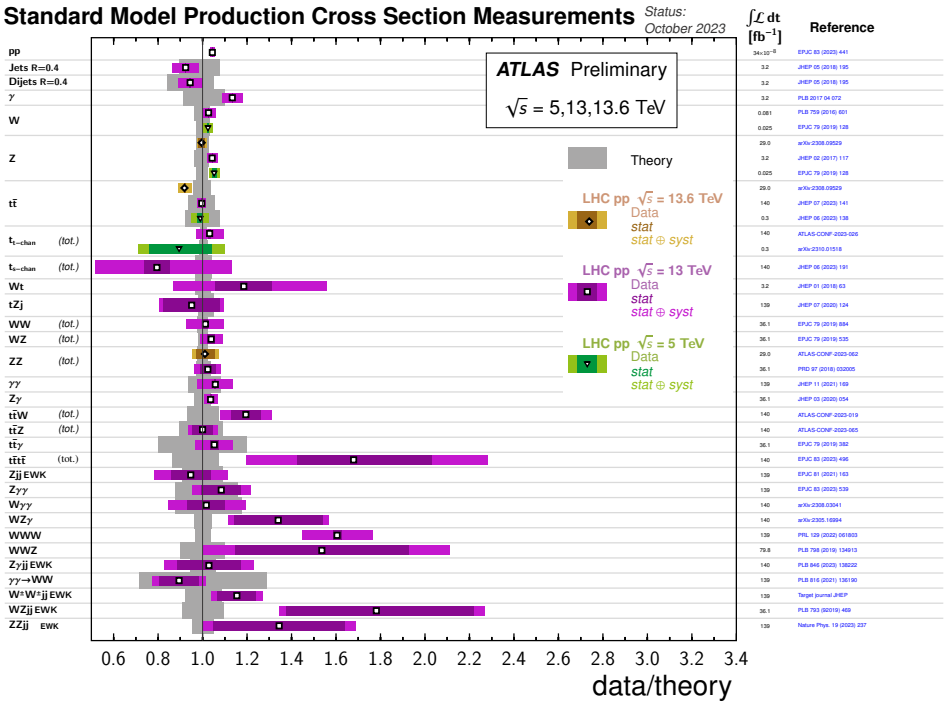
$$D_\mu \phi = \left[\partial_\mu - i\frac{g'}{2}B_\mu - i\frac{g''}{2}\mathbf{W}_\mu \cdot \boldsymbol{\sigma}\right] \phi . \tag{1.46}$$

The term $\mathcal{L}_{\text{gauge}}$ covers the dynamics of the gauge boson fields. The Higgs component of the Lagrangian encompasses both the kinetic and mass terms of the Higgs boson, as well as its interactions with the gauge fields and their mass terms. The Yukawa section provides for the fermions' masses (except for neutrinos) and their interactions with the Higgs field. Finally, $\mathcal{L}_{\text{fermion}}$ captures the dynamic evolution of the fermionic fields.

The physical predictions of this theoretical framework have been confirmed to remarkable precision through countless experiments. The discovery of the Higgs boson [21, 22] at the A Toroidal LHC Apparatus (ATLAS) and Compact Muon Solenoid (CMS) experiments in 2012 served as a crowning achievement, filling the last major gap in the SM's particle zoo. See Figure 1.2 for a summary of cross-sections of SM processes and their current best measurements by the ATLAS collaboration.



(a) SM cross-section measurements.



(b) Comparison to SM predictions.

Figure 1.2: Summary of several ATLAS SM total and fiducial production cross-section measurements (a) and comparisons to the theoretical predictions (b) [23].

1.1.8 The Free Parameters of the Standard Model

To complete the SM, we rely on experimental data to ascertain the values of its numerous free parameters. These include the three coupling constants, the masses or Yukawa

couplings of the twelve fermions (including neutrinos), the Higgs boson's mass, and the VEV. Additionally, experimental determination is required for the eight mixing angles and phases in the Cabibbo-Kobayashi-Maskawa (CKM) and Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrices, which describe the mixing and oscillation between quark and neutrino flavours, respectively, as well as the strong CP phase, which is notably small and often assumed to be zero. For the most accurately measured values of these parameters, see [24].

The existence of these 26 free parameters, alongside the not yet understood observed patterns among their values, is a point of dissatisfaction for many physicists. Coupled with other more critical limitations of the SM, which are discussed in the following chapter, these aspects strongly suggest the existence of physics beyond the Standard Model (SM).

1.2 Shortcomings of the Standard Model

The SM stands as one of the most rigorously tested theories in the realm of physics, validated by countless experiments. Despite its success, several notable deficiencies persist. The following sections provide a brief overview of certain inconsistencies and shortcomings within the SM, with a particular focus on the unresolved issue of identifying a viable DM candidate. While some solutions to these issues will be discussed, a more detailed exploration of specific theories and BSM models relevant to the research presented within this thesis is detailed in Chapter 1.3.

1.2.1 Free Parameters

Section 1.1.8 describes the necessity of external input of certain parameters. The SM's inability to predict these fundamental quantities from first principles are sometimes seen as a hint towards a more fundamental theory.

1.2.2 Neutrino Masses

For a long time, experimental observations were consistent with massless neutrinos. Furthermore, electroweak theory only necessitates the existence of left-handed neutrinos and right-handed anti-neutrinos. To incorporate a Yukawa-like mass term for neutrinos into the SM Lagrangian, it would be necessary to include right-handed neutrinos. Because of these considerations, the SM Lagrangian, developed in the 20th century, did not include a neutrino mass term, and neutrinos were assumed to be massless. However, the discovery of neutrino oscillations, as evidenced in several key experiments [25–27], has conclusively demonstrated that neutrinos have non-zero, albeit very small, masses. We now understand that there are three distinct mass eigenstates, identifiable through measurements of mass differences in neutrino oscillation experiments.

Similar to up-type quarks, neutrino masses could be added to the SM Lagrangian

using the Higgs doublet's conjugate, resulting in a gauge-invariant Dirac mass term

$$L_D = -m_D(\nu_R\nu_L + \nu_L\nu_R) \tag{1.47}$$

suggesting the existence of ν_R. However, right-handed neutrino states ν_R are theorised not to engage in SM interactions, which casts doubt on their existence.

The significantly smaller neutrino masses compared to other fermions hint at an alternative mass-generation mechanism. Given that right-handed neutrinos and left-handed anti-neutrinos behave as SM gauge singlets, additional Lagrangian terms involving these fields can be introduced without violating gauge invariance. Specifically, the Majorana mass term

$$L_M \propto M(\nu_R^c\nu_R + \nu_R\nu_R^c) \ , \tag{1.48}$$

does not disrupt local gauge invariance and can thus be added to the SM Lagrangian. Different to all other fermions in the SM, such Majorana particles are their own anti-particles.

The Seesaw Mechanism

The seesaw mechanism [28, 29], which encompasses both Dirac and Majorana mass terms, provides a framework for explaining the extremely low upper limits on neutrino masses. The resulting neutrino mass matrix from incorporating both these terms can be diagonalised, revealing the physical neutrino states and their masses as its eigenstates and eigenvalues. If the Majorana mass M is significantly larger than the Dirac mass m_D, this mechanism yields one light and one heavy neutrino state, thereby explaining the observed small neutrino masses. The heavy neutrinos would have very high masses, putting them out of reach for current collider experiments. The predicted light neutrinos are predominantly left-handed, and thus interacting weakly, whereas the heavy neutrinos, being right-handed, do not partake in weak interactions.

Although intriguing for explaining neutrino masses, the seesaw mechanism remains hypothetical until experimental evidence, possibly through neutrino-less double beta

decay [30], further substantiates the concept of neutrinos as Majorana particles.

The KATRIN experiment [31] measured the current best upper limit on the neutrino masses to < 0.45 eV [32]. Other large neutrino physics experiments such as DUNE [33] and Hyper-Kamiokande [34] are currently under construction and are expected to solve some of the mysteries around neutrino masses in the coming years.

1.2.3 The Hierarchy Problem

The observation of a 125 GeV Higgs boson by the ATLAS and CMS experiments presents a fundamental question known as the hierarchy problem: Why is the Higgs boson mass markedly lower than the theoretical maximum energy scales of the theory, such as the grand unification scale (10^{16} GeV) or the Planck scale (10^{19} GeV)?

In contrast to fermions, scalar particles such as the Higgs boson are prone to substantial quantum corrections. The corrections to the quadratic Higgs mass due to fermion and boson loops are given by [35] :

$$\text{Fermionic loops: } \Delta m_h^2(f) = -\frac{2g_f^2\Lambda^2}{16\pi^2} \tag{1.49}$$

$$\text{Bosonic loops: } \Delta m_h^2(s) = \frac{\lambda_s^2\Lambda^2}{16\pi^2}\ , \tag{1.50}$$

where g_f is the Yukawa coupling to fermions, λ_s is the scalar quartic coupling, and Λ represents the theory's ultraviolet cut-off. The fermion masses are proportional to the Yukawa couplings and the Higgs boson couples thus most strongly to the most massive particle. Consequently, the most significant corrections to the Higgs mass will originate from the top quark. However, it is expected that new physics exists somewhere between the electroweak scale (~ 100 GeV) and the Planck scale. As a result, the Higgs mass would be very sensitive to corrections from such potential heavy BSM states.

Notably, the different signs of fermionic and bosonic contributions allow for a potential cancellation, thus keeping the Higgs mass within the experimentally observed range. Yet, this cancellation requires fine-tuning of the couplings and must be meticulously

recalibrated at each perturbative order, embodying the essence of the hierarchy problem.

The absence of a protective symmetry in the SM for the Higgs mass implies the electroweak VEV is also unshielded, leading to implications for the entire SM structure, such as the masses of the fundamental particles and subsequently the value of the Fermi constant.

The technical problem of a low Higgs boson mass is thus directly related to one of the fundamental hierarchy problems of physics, namely the large discrepancy between the strengths of the weak force and gravity.

Why is the weak force around 24 orders of magnitude stronger than gravity? Or considered from a different perspective: why are the masses of the fundamental particles so low? Two objects with Planck mass and charge would be equally affected by gravitational and electromagnetic forces. However, while the magnitude of the elementary charge is relatively close to that of the Planck charge, the mass of the top quark is around 17 orders of magnitude below the Planck mass.

Thus, resolving the hierarchy problem motivates BSM theories that introduce new symmetries capable of stabilising the Higgs boson's mass against quantum corrections, see Section 1.3.3.

1.2.4 Unification of Forces

The SM faces significant challenges in unifying the fundamental forces at high energy scales. At the electroweak scale $Q^2 \approx m_Z^2$, the coupling constants of the electromagnetic, the weak and the strong force take on values of approximately 1/128, 1/30 and 1/9. However, the energy dependence of these coupling constants is described by

$$\alpha_i^{-1}(Q^2) = \alpha_i^{-1}(\mu^2) + \beta_i \ln\left(\frac{Q^2}{\mu^2}\right), \tag{1.51}$$

where μ^2 denotes a reference scale and the β_i are coefficients determined by particle loop contributions to gauge boson self-energies.

Notably, the electromagnetic coupling increases with energy, whereas α_S decreases, see Section 1.1.6. The weak force coupling also decreases with increasing energy, albeit

at a slower rate than α_S. This dependence of the coupling strengths on the energy scale hints to a unification of the forces at high energies. However, while this behaviour brings the couplings closer at high energies, the SM does not naturally predict force unification at a single energy scale. The inadequacy in explaining a unified underlying theory of all fundamental interactions is seen as another limitation of the SM.

In theoretical extensions beyond the SM, such as Grand Unified Theories (GUT) [36] and models incorporating Supersymmetry (SUSY), new particles at higher mass scales influence the running of these constants, suggesting a possible convergence at high energy scales.

1.2.5 Gravity

The SM, while successfully describing electromagnetic, weak, and strong interactions, fundamentally fails to integrate gravity. This marks a significant limitation, highlighting the SM's incompleteness as a theory that encompasses all known forces within a single theoretical framework.

Gravity is described by General Relativity [37–39], which explains gravitational forces through the curvature of spacetime, a concept that does not align with the QFT foundations of the SM. At the Planck scale, where gravitational effects become significant, the SM's QFT approach results in non-renormalisable divergences.

Proposed solutions, such as string theory [40] and loop quantum gravity [41] are able to address these issues. They propose a new massless spin-2 particle responsible for mediating the gravitational force, the graviton. However, detecting the graviton is exceptionally challenging due to its incredibly weak interaction with matter. Thus, experimental verification of these theories with current technology is out of reach.

The inability of the SM to describe gravitational interactions hints to it being an effective theory, constrained to domains where gravitational forces are negligible.

1.2.6 The Strong CP Problem

In addition to the hierarchy problem discussed in Section 1.2.3, particle physics faces another fine-tuning challenge known as the strong CP problem. CP symmetry combines charge conjugation symmetry (C), which transforms a particle into its antiparticle, with parity symmetry (P), the inversion of spatial coordinates. Despite CP symmetry being conserved in experimental observations of QCD, there are no underlying theoretical reasons compelling this conservation, thus presenting the strong CP problem.

Within the QCD Lagrangian, see Equation 1.29, one may introduce a term that potentially breaks CP symmetry if the angle θ is non-zero. This term, known as the theta term, can be expressed as:

$$\mathcal{L}_\theta \propto \theta \; G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \; . \tag{1.52}$$

Here, $G^a_{\mu\nu}$ represents the gluon field strength tensor, and $\tilde{G}^{a\mu\nu}$ the dual of the gluon field strength tensor, defined by $\tilde{G}^{a\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} G^a_{\rho\sigma}$, where $\epsilon^{\mu\nu\rho\sigma}$ is the Levi-Civita symbol.

Experimental measurements, particularly of the neutron electric dipole moment (nEDM), have imposed stringent limits on the value of θ [42]. This unexpectedly small upper bound challenges natural expectations in theoretical physics, where θ is anticipated to be around unity without requiring fine-tuning. The question of why θ is that small and potentially zero, encapsulates the core of the strong CP problem.

The most notable theoretical solution proposed to address this issue is the Peccei-Quinn mechanism [43, 44], which introduces an additional global symmetry that dynamically drives θ to zero. Besides providing a solution to the strong CP problem, this approach, similar to the Higgs mechanism, also predicts the existence of a new pseudoscalar particle. This additional particle, the axion, is also a potential candidate for DM.

Looking forward, experimental advancements, such as enhanced measurements of the nEDM or potential axion detection, are expected to provide deeper insights into the strong CP problem.

1.2.7 Matter-Antimatter Asymmetry

It is assumed that the Big Bang produced equal amounts of matter and antimatter. However, current observations reveal a significant imbalance between baryonic matter and anti-baryonic matter. Sakharov outlined several conditions [45] necessary to explain this asymmetry. These conditions include the violation of baryon number and CP symmetry. To date, no baryon number violating interactions have been observed within the SM. Although CP violation does occur in weak interactions [46], the magnitude of this effect is insufficient to account for the asymmetries observed. This suggests the existence of yet unknown baryon number and CP violating mechanisms in BSM physics.

1.2.8 Dark Matter

DM is a hypothetical class of matter that has only been observed indirectly through its gravitational effects. Its existence has been confirmed through a variety of astrophysical observations including gravitational lensing [47], the anomalous rotational velocities of objects in galaxies [48], as well as anisotropies observed in the Cosmic Microwave Background (CMB) [49]. Despite attempts to explain these phenomena through modified Newtonian dynamics (MOND) [50], no existing theory can fully explain them without postulating the existence of a new type of matter that is yet to be discovered.

The Λ Cold Dark Matter Model (ΛCDM model), the standard model of cosmology, describes the evolution of the universe starting with the Big Bang approximately 13.8 billion years ago. In conjunction with the latest measurements of the Planck collaboration [51], the ΛCDM model predicts that DM constitutes about 26.4% of the total energy density in the universe. In contrast, visible baryonic matter accounts for only 4.9%, indicating that DM is more than five times more abundant than visible matter. The remaining 68.5% of the universe's energy content is composed of dark energy.

Despite DM being the dominant form of matter and its importance to the evolution of the universe, we know very little about it. While we can infer its abundance and distribution in galaxies and the universe from its gravitational effects, we don't know what

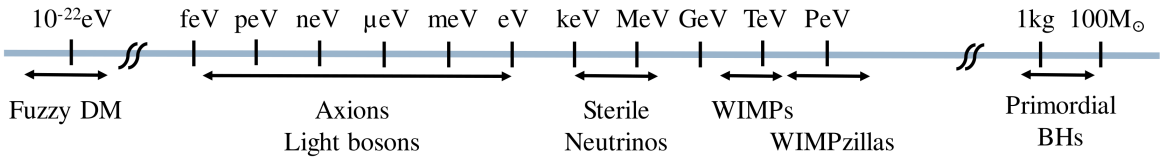


Figure 1.3: Approximate mass ranges of a selection of DM candidates. The potential objects responsible for the DM energy density in our universe span a huge range over many orders of magnitude.

DM is composed of. Consequently, identifying and discovering the building blocks of DM poses one of the most profound challenges in modern physics.

There exist numerous hypotheses about the constituents of DM, ranging from very light particles to extremely massive black holes. Figure 1.3 provides an overview of potential DM candidates and their estimated mass ranges. If DM consists of particles, the SM lacks a viable candidate that satisfies the necessary criteria. This suggests the existence of novel massive, stable and electrically neutral particles that remain to be discovered.

Evidence for Dark Matter

Today, the existence of DM is widely accepted due to compelling evidence from various astrophysical phenomena. This section introduces evidence from galactic rotation curves and the Bullet Cluster. However, these are merely two examples among numerous observations supporting the existence of DM. Other examples [52] are the characteristic peaks in the acoustic power spectrum of the CMB and cosmic expansion measurements from Type Ia supernovae.

Galaxy rotation curves have provided the first source of evidence for the existence of DM. Astrophysical observations reveal that the rotational velocities of stars and gas in galaxies do not conform to the predictions based on their visible matter alone. According to Newtonian dynamics and Kepler's third law, the velocities of objects should decrease with the square root of their distance from the galactic centre, assuming most mass is concentrated there. However, observations show that the rotational velocities remain constant or even increase slightly at greater distances, suggesting the presence of an unseen

mass component distributed also outside the galaxy centre. These measured velocity distributions can be explained by the hypothesis that galaxies are embedded in DM halos, extending far beyond the visible edge of the galaxies.

Gravitational lensing is a method that uses the gravitational fields of massive objects such as galaxy clusters to distort the light from background objects. This technique allows the estimation of the mass profiles of galaxies and clusters. The results indicate that the total masses of galaxies are significantly higher than the estimates obtained from models that only consider visible matter.

A particularly interesting example is the Bullet Cluster, consisting of two colliding galaxy clusters. The three components of the two galaxies, the stars, the gas clouds and the DM behave differently during collision and can be observed using various techniques. The stars emit visible light, while the gas clouds emit X-rays due to the heat generated from interactions during the collision. The DM distribution is studied utilising its gravitational lensing effects.

During the collision, the gas clouds slowed down due to friction from electromagnetic interactions, while the visible stars and the DM passed through without any deceleration. This caused the gas clouds to separate from the DM halos and the stars. See Figure 1.4 for a composition of an X-ray image and a gravitational lensing image of the Bullet Cluster, highlighting the different components and their spatial separations.

Although the gas clouds are expected to contain most of the visible mass, gravitational lensing techniques reveal the gravitational centres of the clusters to be located around the visible stars. This observation indicates the existence of large DM halos dominating the gravitational effects and sets constraints on DM self-interaction.

Dark Matter Detection

Our understanding of the nature of DM has been primarily derived from indirect astrophysical observations. However, the measured relic density can be explained by a process

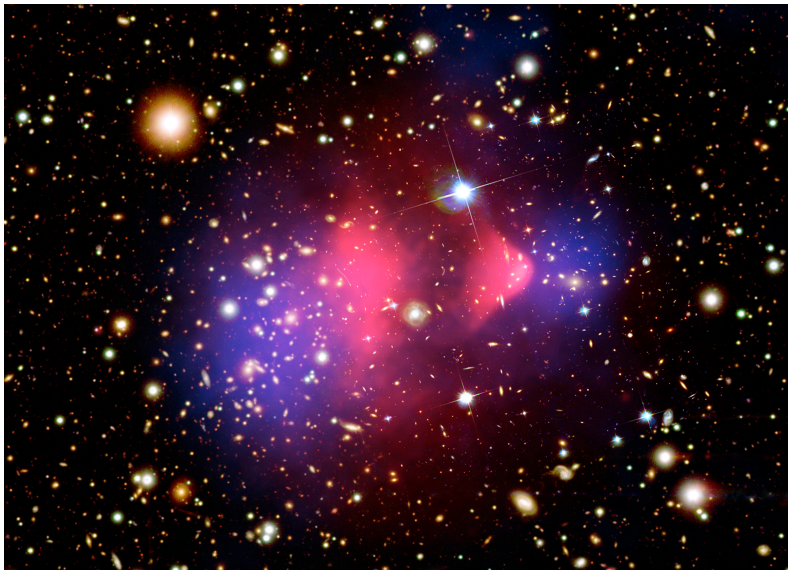


Figure 1.4: The Bullet Cluster. The DM distributions computed from gravitational lensing effects are shown in blue and are superimposed with an X-ray image depicting the gas clouds in red. [53]

known as thermal freeze-out, which suggests interactions between DM and SM particles in the early universe. In this process, DM was in thermal equilibrium with the SM until interactions became rare due to cosmic expansion. The interaction strength between DM and SM determined the timing of freeze-out and, consequently, the relic density. A variety of experimental approaches based on these interactions between DM and the SM are being explored to detect DM and gain further insights into its constituents and properties.

Indirect detection experiments with telescopes or satellites search for signs of DM self-annihilation or decay in areas of high DM density, focusing on analysing the resulting decay products. Direct detection experiments seek to detect DM particles directly through their interactions with atomic nuclei within extensive, shielded underground detectors. Examples of direct detection experiments include the XENON experiment [54], the LZ experiment [55] and the PandaX experiment [56]. In contrast, the research documented in this thesis explores the potential to produce and analyse DM via high-energy particle collisions at the LHC. Despite these efforts, DM has continued to be elusive so far. For a visual representation of the processes involved in various DM detection strategies, refer to Figure 1.5.

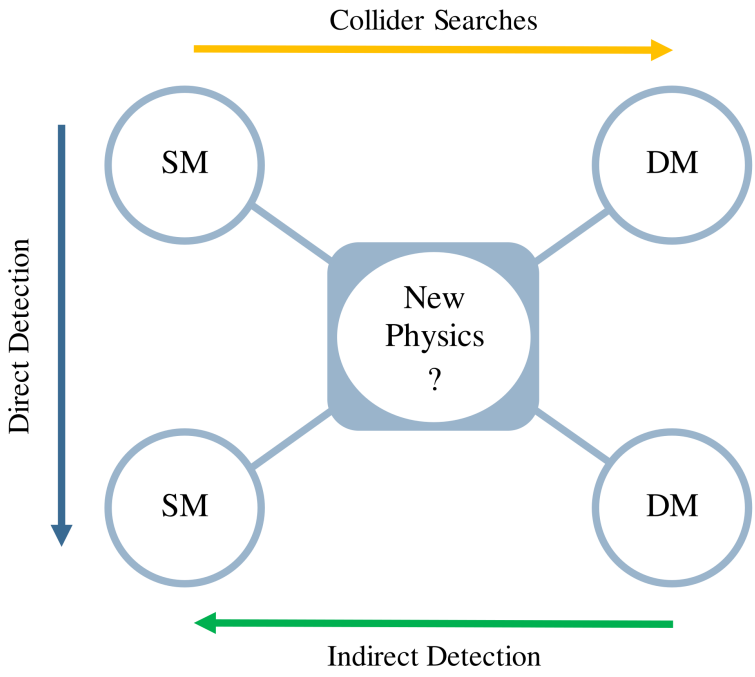


Figure 1.5: This figure illustrates the primary experimental approaches to studying DM. The arrows indicate the direction of the various interactions. Direct detection experiments observe nuclei scattered by interactions with DM particles. Indirect detection experiments detect SM particles resulting from DM annihilation. Collider experiments produce DM particles via high-energy SM particle collisions, potentially revealing new mediators between the SM and DM.

Overview of Dark Matter Candidates

The landscape of DM candidates is extensive, covering a wide range of hypothetical particles and astronomical objects, see Figure 1.3.

Based on their velocities, the candidates are classified primarily into two categories, namely Hot Dark Matter (HDM) and Cold Dark Matter (CDM). As HDM travels at ultra-relativistic speeds, it makes it difficult to explain the universe's smaller-scale structures such as galaxies. For this reason, at least part of the DM content is assumed to be cold.

Weakly interacting massive particles (WIMPs) are the most-studied and most popular CDM candidates. WIMPs interact with the visible sector only via gravity and another force at least as weak as the weak nuclear force. Their popularity as CDM candidates originates from the "WIMP Miracle", which refers to the observation that particles with a mass and interaction strength at the weak scale, typically between 10 GeV

and a few TeV, naturally freeze out from thermal equilibrium in the early universe with an abundance close to the observed relic DM density. The freeze-out occurs when the expansion rate of the universe outpaces the rate at which WIMPs can interact with each other. Subsequently, this leads to a sharp decline in the WIMPs annihilation rate and thus results in a stable relic density. The WIMPs miracle thus elegantly ties the weak scale with the cosmological scale of DM, making WIMPs one of the most popular DM candidates. Furthermore, many popular BSM theories such as SUSY predict particles on the weak mass scale with the required properties, further increasing the appeal of the WIMP hypothesis.

There are various **other DM candidates**, such as the axions. As detailed in Section 1.2.6, they provide both a valid DM candidate as well as a solution for the strong CP problem. Axions are expected to be very light with masses ranging multiple orders of magnitude below the eV scale. Because they are predicted to interact weakly with electromagnetic fields, many axion detection efforts involve converting them into photons in the presence of strong magnetic fields.

On the other end of the mass scale, there are Massive Compact Halo Objects (MACHOs) and primordial black holes. MACHOs, which could be realised as black holes, neutron stars or other heavy, not-luminous objects, could be detected through gravitational lensing. Primordial black holes, formed from density fluctuations in the early universe, are also a popular CDM candidate.

Heavy sterile neutrinos, another DM candidate, could be created through the seesaw mechanism, see Section 1.2.2. To account for the observed DM relic density, sterile neutrinos are expected to have masses on the keV scale.

This section has introduced only a small subset of DM candidates. Other hypothesised DM candidates include ultra-light, fuzzy DM, heavy WIMPzillas [57] and dark photons [58], to name just a few.

1.3 Dark Matter Searches: Di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$

While the previous chapter provided an overview of the shortcomings of the SM, motivating the general necessity for BSM physics, this chapter will introduce specific models that address some of these issues. The first part serves as a brief general introduction to BSM searches at collider experiments.

In accordance with the DM search presented in this thesis, the second part discusses models that predict the production of a pair of SM Higgs bosons accompanied by an imbalance in the measured transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$), attributed to the presence of DM. While di-Higgs production is a rare process in the SM, with the dominant gluon-gluon fusion mode predicted to have a cross-section of approximately 31 fb at $\sqrt{s} = 13$ TeV, current experimental searches by ATLAS [59] and CMS [60] have set upper limits that remain significantly larger than this prediction. These searches primarily target SM di-Higgs final states and do not specifically address events with additional large $E_{\mathrm{T}}^{\mathrm{miss}}$. Signatures with additional $E_{\mathrm{T}}^{\mathrm{miss}}$ can arise in BSM scenarios, such as those involving DM candidates. Given the substantial gap between current experimental limits and the SM prediction, searches for BSM processes with enhanced production rates and distinctive final states, including $E_{\mathrm{T}}^{\mathrm{miss}}$, remain well-motivated. Particular attention is given to the dark-Higgs model, which presents a novel signature and favourable relic density properties. Additionally, other models generating similar signatures will be explored, reflecting the author's research in these areas.

1.3.1 Dark Matter Searches at the LHC

Once produced, hypothetical DM particles would escape detection within the detector, resulting in non-zero $E_{\mathrm{T}}^{\mathrm{miss}}$. These signatures are commonly targeted by X+$E_{\mathrm{T}}^{\mathrm{miss}}$ searches, where the X represents one or more recoiling SM particles. At the LHC, searches for X+$E_{\mathrm{T}}^{\mathrm{miss}}$ events have been conducted using various channels, involving photons, jets, or vector bosons. The main result of this thesis presents a search aimed at identifying BSM events with a final state containing four b-jets recoiling from $E_{\mathrm{T}}^{\mathrm{miss}}$.

Results from BSM searches at the LHC are typically interpreted within the framework of DM theories. These interpretations range from relatively general model-independent approaches to model-specific analyses. While the primary goal is to discover new physics, in the absence of such a discovery, the focus shifts to establishing limits on cross-sections or defining exclusion limits for specific models.

In models with a large number of parameters, it's common practice to simplify the interpretation by limiting the number of free parameters to a small number. This is achieved by fixing other parameters, such as the masses of new particles or the couplings of new interactions, at predetermined values.

Often benchmark values are used, determined from experimental or theoretical arguments. These benchmarks are also crucial for comparing results across different searches. Results from various decay channels or even different experiments, investigating a similar model, can not only be compared to each other, but also statistically combined.

The computed limits are usually illustrated using "Brazilian plots", showing the expected and observed limits along with their uncertainty bands. For instance, Figure 1.10 details exclusion limits for the dark-Higgs model from the ATLAS experiment, parametersied by the masses of two BSM states in the theory.

A large number of theories have been developed to guide and interpret searches at the LHC. These models range from high-level effective theories to detailed complete models, each with their specific advantages and limitations, see Figure 1.6.

Effective Field Theories

EFTs offer a high-level description of DM interactions with the SM, serving as an approximation of more complete theories. Their goal is to describe physical effects observable at LHC energy scales, while decoupling the details of physics at higher scales. Thus EFTs are especially useful when there is a clear separation between the scale of interest and the scale of the underlying interactions. At the scale of interest, the effects of BSM physics are captured through higher-dimensional operators. An early instance of an EFT is Fermi's theory, which describes beta decay at much lower scales than the

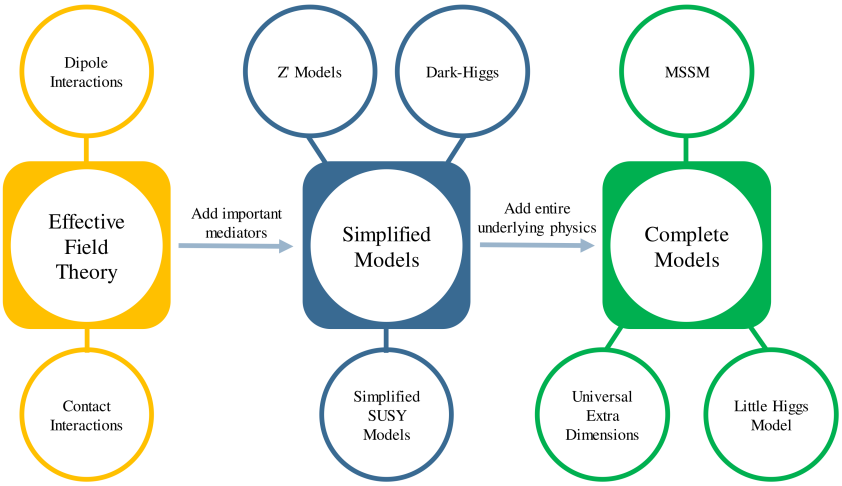


Figure 1.6: DM models used in collider physics are classified into Effective Field Theories (EFT)s, simplified models and complete models. A few examples per category are depicted. Illustration adapted from Ref. [61].

electroweak scale, without any knowledge of the mediating vector bosons. In terms of DM models, EFTs assume that DM is the only accessible state at the LHC. The interactions between DM and SM, mediated by inaccessible particles, are included independently of the exact underlying physics. This approach allows to set constraints on Λ, the suppression scale for higher-order operators.

EFTs are beneficial for their simplicity and model-independent constraints but don't allow conclusions about the underlying physics and lose reliability at energy scales close to new physics phenomena.

Complete Models

On the other end of the spectrum are complete BSM models, providing exhaustive frameworks that incorporate DM candidates while sometimes addressing broader particle physics questions. These models usually introduce many more particles and thus include many additional parameters. However, the rich structure of such models may make it difficult to determine the underlying physics from the limited data available. Examples include the Minimal Supersymmetric Standard Model (MSSM) and theories of Universal Extra Dimensions. Despite their complexity and the difficulty of constraining them experimentally, these models offer precise predictions for DM properties and interactions.

Simplified Models

Simplified models strike a balance between the generality of EFTs and the specificity of complete models, which makes them particularly useful for the interpretation of results of BSM searches at the LHC. They can be employed to generate a broad spectrum of signatures, effectively capturing the possible signatures of more comprehensive models. Additionally to the DM candidate, these models also include the most important mediators between the SM and DM. In particular, this allows to describe the details of the mechanisms of DM production at the LHC. Simplified models usually include a small number of parameters including the masses of the new particles as well as their couplings to each other and to the SM.

1.3.2 The Dark-Higgs Model

One common approach to construct a simplified model is to incorporate an additional dark mediator that communicates between the SM and DM, for example a single s-channel vector mediator, denoted as Z'. However, to ensure agreement with experimental bounds, the couplings of this mediator to SM particles must be very small, making it difficult to avoid DM overproduction. To mitigate this issue, a complex dark-Higgs scalar field S is introduced, which is charged under a new $U(1)'$ gauge group. The dark-Higgs field acquires a VEV, breaking the $U(1)'$ gauge symmetry, thus giving mass to the Z' gauge boson. The resulting physical dark-Higgs boson s, the Z', and an additionally introduced fermionic DM particle χ together form the dark sector in this Two-Mediator Dark Matter Model (2MDM). The following discussion of the dark-Higgs mechanism is guided by Ref. [62].

The total Lagrangian of this model is given by

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{\chi} + \mathcal{L}_{S} + \mathcal{L}'_{\text{gauge}}, \tag{1.53}$$

where $\mathcal{L}_{SM}$ is the SM Lagrangian. Similar to the mechanism generating fermion masses in the SM (see Eq. 1.25), the DM obtains its mass through a Yukawa coupling (y_{χ}) to the

dark-Higgs:

$$\mathcal{L}_{M\chi} = -\frac{1}{2} y_\chi \overline{\chi}(P_L S + P_R S^*)\chi \ . \tag{1.54}$$

Here, S denotes the complex dark-Higgs scalar field, χ the DM Majorana field, and P_L and P_R the projection operators for left and right-handed states, respectively.

$\mathcal{L}_\chi$ also includes terms describing the dynamics of χ as well as its coupling to the Z':

$$\mathcal{L}_\chi = \frac{i}{2}\bar{\chi}\gamma^\mu\partial_\mu\chi - \frac{1}{2} g' q_\chi Z'_\mu \bar{\chi}\gamma^5\gamma^\mu\chi + \mathcal{L}_{M\chi} \ . \tag{1.55}$$

Note that the gauge coupling g' should not be confused with that introduced in Eq. 1.8. It is not a free parameter and is absorbed into the fermion charges.

The two mediators can interact with the SM either through the mixing of the Z' with the neutral gauge bosons or through a direct coupling to SM particles. For instance, models of gauged baryon number generate a coupling of the Z' to SM quarks [63]. The model used for the search detailed in this thesis assumes the latter scenario. The gauge term of the Lagrangian is thus written

$$\mathcal{L}'_{\text{gauge}} = -g' Z'_\mu \sum_{f=\text{quarks}} g_f \bar{f}\gamma^\mu f - \frac{1}{4} F'_{\mu\nu} F'^{\mu\nu} \ , \tag{1.56}$$

where $F'^{\mu\nu} = \partial^\mu Z'^\nu - \partial^\nu Z'^\mu$.

Additionally, the dark-Higgs interacts with the SM through mixing with the SM Higgs. Technically this is implemented by adding the following term to the dark-Higgs potential:

$$\mathcal{L}_{\text{mixing}} = -\lambda_{hs}(H^\dagger H)(S^\dagger S) \ . \tag{1.57}$$

Due to this mixing, the dark-Higgs couples to SM particles proportional to their mass. Depending on the mass of the dark-Higgs, it will preferably decay to SM Higgs, vector bosons or heavy quarks.

The full dark-Higgs part of the Lagrangian, including the potential and kinetic terms

is expressed as

$$\mathcal{L}_S = (D^\mu S)^\dagger (D_\mu S) + \mu_s^2 S^\dagger S - \lambda_s (S^\dagger S)^2 + \mathcal{L}_{\text{mixing}} \ . \tag{1.58}$$

Here, the covariant derivative is defined as $D^\mu = \partial^\mu + i g' q_s Z'^\mu$.

Similar to Section 1.1.4, the unitary gauge can be applied after spontaneous symmetry breaking, allowing the dark-Higgs field to be expanded around its VEV w as

$$S = \frac{1}{\sqrt{2}}(w + s) \ . \tag{1.59}$$

The mass and coupling parameters are defined as $m_\chi = 1/\sqrt{(}2) y_\chi w$, $m_{Z'} = 2 g' q_\chi w$, $g_\chi = g' q_\chi$, $g_f = g' q_f$ and the mixing angle θ introduced by

$$\tan(2\theta) = \frac{\lambda_{hs} v w}{\lambda_h v^2 - \lambda_s w^2} \ . \tag{1.60}$$

The mixing of the Higgs bosons in the visible and dark sectors is handled by the substitution

$$h \to h \cos\theta + s \sin\theta \tag{1.61}$$

$$s \to s \cos\theta - h \sin\theta \ . \tag{1.62}$$

In total, the dark-Higgs model introduces six independent parameters, namely

$$g_\chi, \ g_f, \ m_\chi, \ m_{Z'}, \ m_s \text{ and } \ \theta \ . \tag{1.63}$$

The masses of the Higgs and dark-Higgs, as well as the parameters in the dark-Higgs potential, can be expressed in terms of these parameters. For details, refer to Ref. [62].

Including two mediators in the dark sector also mitigates current experimental constraints. For instance, models with only one additional spin-1 mediator Z' are constrained by mono-jet searches [64], while models with spin-0 mediators are constrained by Higgs to invisible searches and SM Higgs cross-section measurements [65]. However, by introducing the dark-Higgs boson along with appropriate annihilation pathways to SM particles, the

model is more naturally able to satisfy the observed relic density and explain the DM mass, without being highly constrained by existing searches.

Finally, the full dark-Higgs model Lagrangian in terms of the physical dark-Higgs and SM Higgs boson fields after symmetry breaking is detailed by:

$$\begin{aligned}\mathcal{L} &= \frac{i}{2}\bar{\chi}\gamma^{\mu}\partial_{\mu}\chi - \frac{1}{2}g_{\chi}Z'_{\mu}\bar{\chi}\gamma^{5}\gamma^{\mu}\chi - \frac{1}{2}m_{\chi}\bar{\chi}\chi - \frac{m_{\chi}g_{\chi}}{m_{Z'}}\bar{\chi}\chi s + \\ &+ \frac{1}{2}m^{2}_{Z'}Z'_{\mu}Z'^{\mu} + \frac{1}{2}\partial^{\mu}s\partial_{\mu}s + 2g_{\chi}Z'_{\mu}Z'^{\mu}\left(g_{\chi}s^{2} + m_{Z'}s\right) + \\ &+ \frac{\mu_{s}^{2}}{2}\left(s + \frac{m_{Z'}}{2g_{\chi}}\right)^{2} - \frac{\lambda_{s}}{4}\left(s + \frac{m_{Z'}}{2g_{\chi}}\right)^{4} + \frac{\lambda_{hs}}{4}\left(s + \frac{m_{Z'}}{2g_{\chi}}\right)^{2}(h+v)^{2} + \\ &- Z'_{\mu}\sum_{f=\text{quarks}} g_{f}\bar{f}\gamma^{\mu}f - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} - \frac{1}{2}\sin\epsilon F'^{\mu\nu}B_{\mu\nu}\ .\end{aligned} \tag{1.64}$$

This Lagrangian includes mass terms for the Z', the dark-Higgs, and DM, as well as interactions between particles within the dark sector and the coupling of the Z' to the SM quarks.

The broken $U(1)'$ symmetry, from which both mediators originate, allows for their interaction, leading to processes such as dark-Higgs strahlung $Z'^{*} \to Z's$, as shown in Figure 1.7. This process is the main target of the analysis described in Part III.

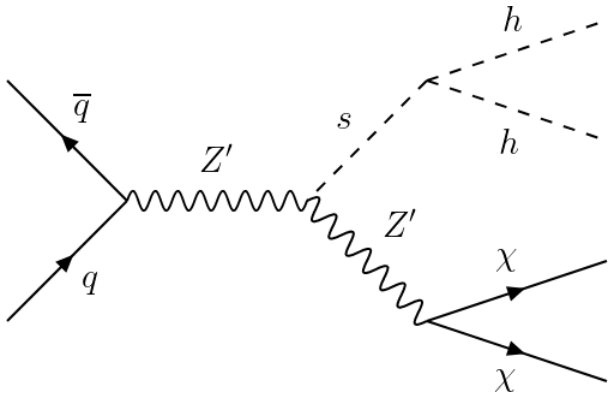


Figure 1.7: Dark-Higgs strahlung in the 2MDM model, which represents the dominant process targeted by the analysis in this thesis. While the dark Higgs s can decay via various channels depending on its mass, this work focuses specifically on the $s \to hh$ channel.

In many events, the dark-Higgs will be relatively soft and thus the two DM particles back-to-back. In this case the $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions cancel and hence such events are

difficult to analyse. High $E_{\mathrm{T}}^{\mathrm{miss}}$ can only be achieved in case of an strongly off-shell Z' or via additional initial state radiation (ISR): $pp \to Z' + j \to \chi\chi + j$. The cross-section for radiating a dark-Higgs boson of several hundred GeV from a low $m_{Z'}$ is strongly suppressed, while for a very massive Z', a boosted dark-Higgs and thus high $E_{\mathrm{T}}^{\mathrm{miss}}$ are likely. Therefore, the cross-section for dark-Higgs radiation declines relatively slowly with increasing $m_{Z'}$, see Figure 1.8. While searches for $pp \to Z' + j \to \chi\chi + j$ benefit from higher cross-sections at low $m_{Z'}$, searches for dark-Higgs radiation are competetitive at high masses. The decay of the dark-Higgs in a strahlung scenario also gives rise to unique resonant signatures, which allow better discrimination agains SM background processes.

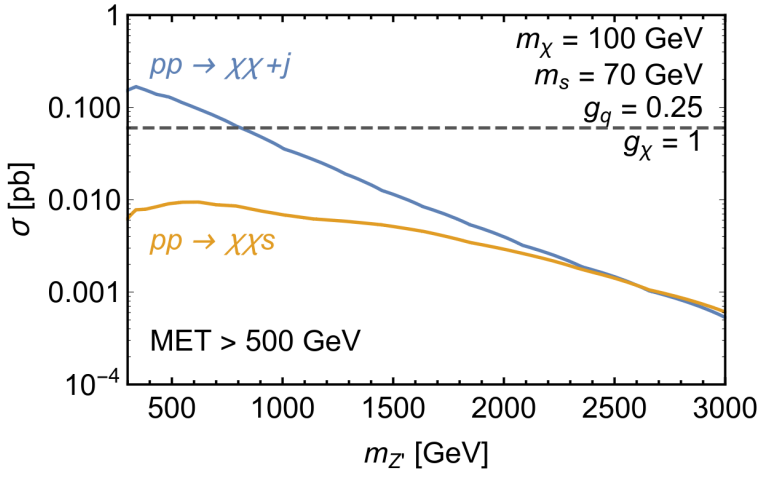


Figure 1.8: Cross-sections for $pp \to \chi\chi + j$ and $pp \to \chi\chi s$ as a function of the Z' mass. Figure taken from Ref. [66].

Although the couplings between SM particles and dark-Higgs bosons are relatively small, the decay to SM particles is dominant if the decay $s \to \chi\chi$ is kinematically not allowed. The branching ratios of the dark-Higgs decay to SM particles depend on m_s. In the mass range targeted by the search presented in this thesis $2m_h \leq m_s \leq 2m_\chi$, the decay modes $s \to VV$ and $s \to hh$ are dominant. In the case of lower dark-Higgs masses ($m_s < 2m_W$), the decay $s \to bb$ is favoured, while for $m_s > 2m_\chi$, decays to SM particles are suppressed. Figure 1.9 provides an overview of the dark-Higgs decay branching ratios.

Due to the dependency of the final states on m_s, dedicated searches for different decay channels and mass-ranges exist for these models. Low m_s are targeted by a search using the $s \to bb$ channel [67], while the fully hadronic final state of $s \to VV$ [68] and

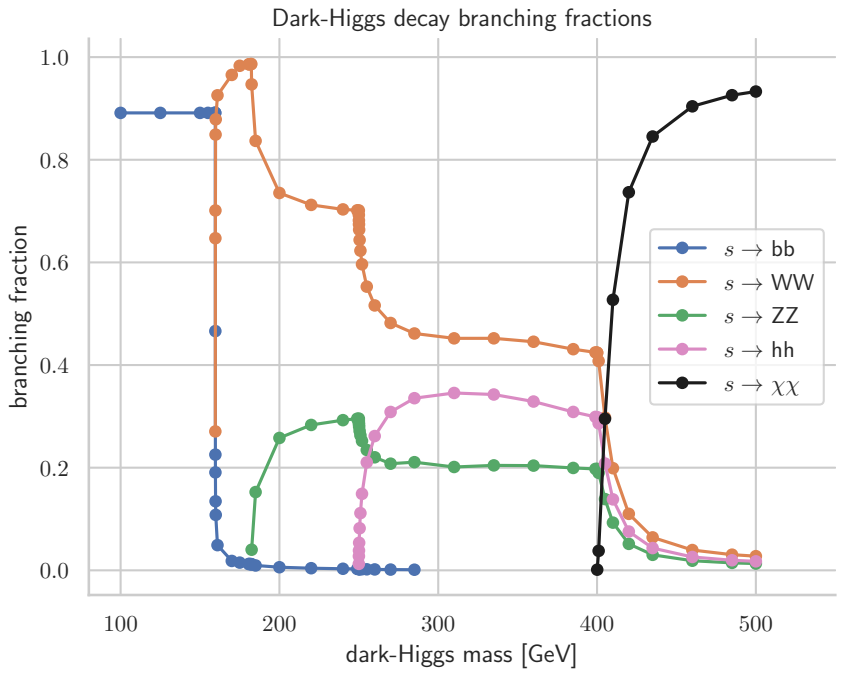


Figure 1.9: Branching ratios of the dark-Higgs bosons decay to b-quarks, W^+W^-, ZZ and HH. The branching ratio was computed using MadGraph v2.9.9 at leading order, for the benchmark case where $m_\chi = 200$ GeV.

semi-leptonic decay of $s \to WW$ [69] are associated with higher m_s. The search presented in this thesis is the first to investigate the $s \to hh \to 4b$ case, targeting similar mass ranges as covered by Ref. [69]. See Figure 1.10 for the current best exclusion limits derived with the $s \to WW$ search and others.

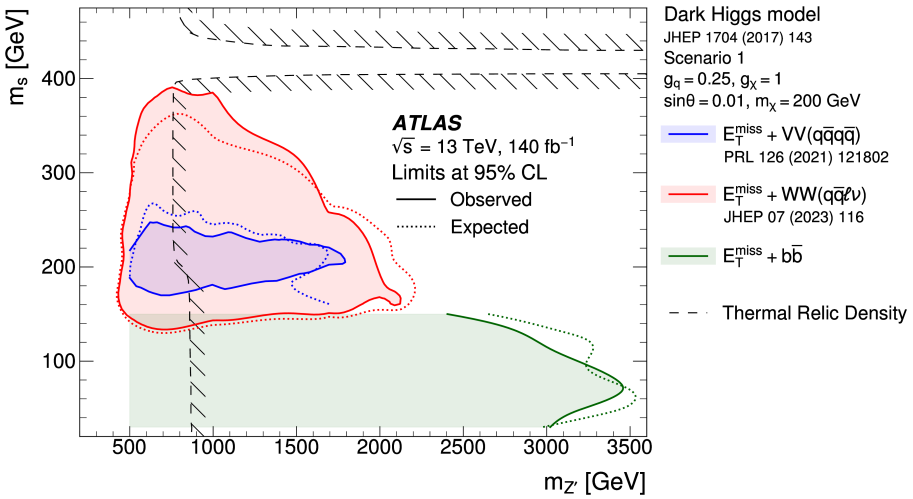


Figure 1.10: The red contours represent the exclusion limits for the dark-Higgs model derived via the semi-leptonic $s \to WW$ channel [69]. The observed exclusion limits from the hadronic $s \to VV$ search are shown in blue [68], while the green contours correspond to the $s \to b\bar{b}$ channel [67]. This summary plot is taken from Ref. [70].

1.3.3 Other Models

The dark-Higgs model is not unique in predicting a di-Higgs + E_T^{miss} signature. However, the specifics of the decay topologies, masses and lifetimes of involved particles, as well as spin effects can result in different kinematic behaviour of the final state particles. It is therefore essential to optimise the analysis strategies towards different models to systematically explore the full range of the di-Higgs + E_T^{miss} channel.

Besides implementing the dark-Higgs search presented in Part III, the author has studied some of the following models in more or less detail.

Supersymmetry

Introduction SUSY theories extend the SM by introducing a symmetry between bosonic and fermionic particles, assigning a superpartner, or "sparticle", to each SM particle that differs by half a unit in spin.

SUSY addresses several of the issues outlined in Section 1.2, including providing a DM candidate, enabling grand unification, explaining the anomalous magnetic moment of the muon, and resolving the hierarchy problem. The latter is solved through the cancellation of the loop corrections to the Higgs boson mass due to the opposite sign of fermionic and bosonic corrections (see Eq. 1.49 and 1.50).

To prevent the rapid decay of the proton in these theories, the conservation of R-parity is required. R-parity is introduced as $R = (-1)^{3(B-L)+2S}$, where B is the baryon number, L is the lepton number, and S denotes the spin of the involved particles and sparticles. SM particles have even R-parity and sparticles have odd R-parity. As a result, sparticles can only be produced in pairs, and the Lightest Supersymmetric Particle (LSP) cannot decay to SM particles and is therefore stable. Electrically neutral LSPs, such as neutralinos or gravitinos, are thus viable candidates for DM.

The MSSM introduces the minimum number of new states to form a complete model, including two Higgs supermultiplets to avoid gauge anomalies. The MSSM adds 120 new parameters and is thus very difficult to constrain in collider searches. Consequently, simplified SUSY models are often employed, assuming that only a small number of new

states are accessible within the energy range of the LHC.

If SUSY were an exact symmetry, the superpartners would share the same electrical charge, colour charge, weak isospin, and mass as their SM partners. However, as no SUSY particles with SM-equivalent masses have been discovered, SUSY must be broken. To maintain the cancellation of the loop corrections to the Higgs mass, SUSY is assumed to be broken softly. Such soft SUSY breaking would affect low energy physics while retaining the cancellation of the loop corrections. It is thus expected that at least a few SUSY states are realised at relatively low energies in the TeV range, and thus potentially within reach of the LHC.

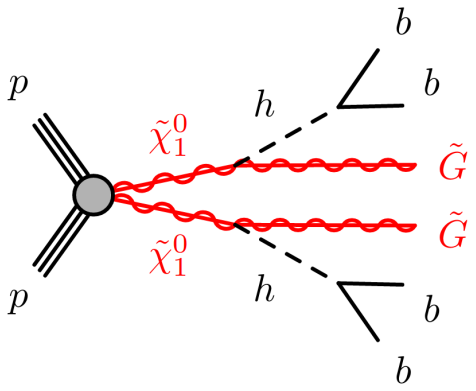
Various theories have been proposed to explain SUSY breaking. In Gauge-Mediated SUSY Breaking (GMSB), SUSY is broken in a hidden sector and communicated to the supermultiplets via massive messenger fields, thereby introducing additional states such as the goldstino and the gravitino. For a more detailed introduction to SUSY, refer to [71].

Di-Higgs + $E_{\mathbf{T}}^{\mathbf{miss}}$ in SUSY In SUSY models, a di-Higgs + E_T^{miss} signature can arise through Goldstino production in models with gauge-mediated SUSY breaking. It can also result from chain decays of superpartners into Higgs bosons and neutralinos in the MSSM and its extensions. Figure 1.11(a) illustrates a typical decay chain in such simplified models. In these theories the LSP, typically the lightest neutralino or the Gravitino, is a promising candidate for DM. Due to R-parity conservation the LSP is assumed to be stable and avoids detection within the detector, and is thus responsible for a large E_T^{miss} component.

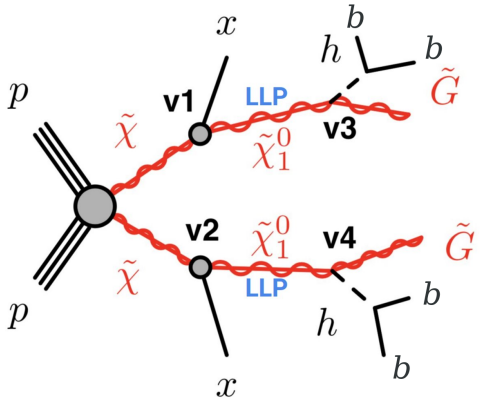
These SUSY models are targeted by numerous searches at ATLAS. The author of this thesis was involved in such an analysis targeting E_T^{miss} and multiple b-jets in the final state [1]. In particular, the author was working on an additional interpretation of the results in terms of long-lived signals, see Figure 1.11(b). Besides the gauge mediated models, also Axino LSP models (Figure 1.11(c)) or models with massive electroweakino LSPs (Figure 1.11(d)) justify a closer look at the di-Higgs + E_T^{miss} signature.

Due to R-parity conservation, sparticles are always created in pairs and thus all these

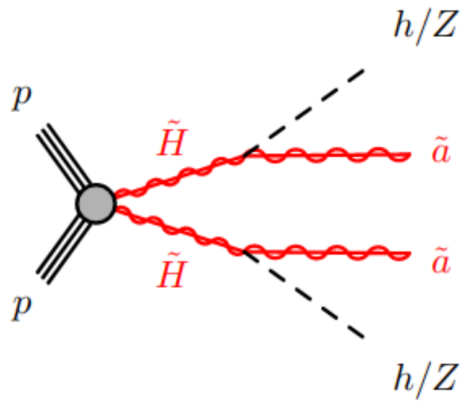
SUSY models have a symmetric decay chain in common. This differentiates them from the dark-Higgs model discussed in the previous section. In the latter, one resonance decays completely visibly, and the other completely invisibly. Due to the very different kinematics of the final state particles in these scenarios, re-interpretations of SUSY searches have limited sensitivity for models such as the dark-Higgs one.



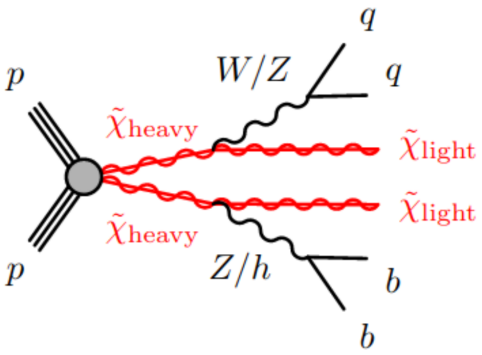
(a) GMSB models with nearly degenerate neutralinos and nearly massless Gravitino LSP (a)



(b) Similar to a), but the lightest neutralino $\tilde{\chi}_1^0$ is assumed to be long-lived with a lifetime between 0.01ns and 10ns. Targeted by the re-interpration of the ATLAS "multi-b" SUSY analysis.



(c) Axino LSP models.



(d) Electroweakino LSP models.

Figure 1.11: Different SUSY models allowing for a di-Higgs + $E_{\text{T}}^{\text{miss}}$ signature.

The tri-Scalar Models

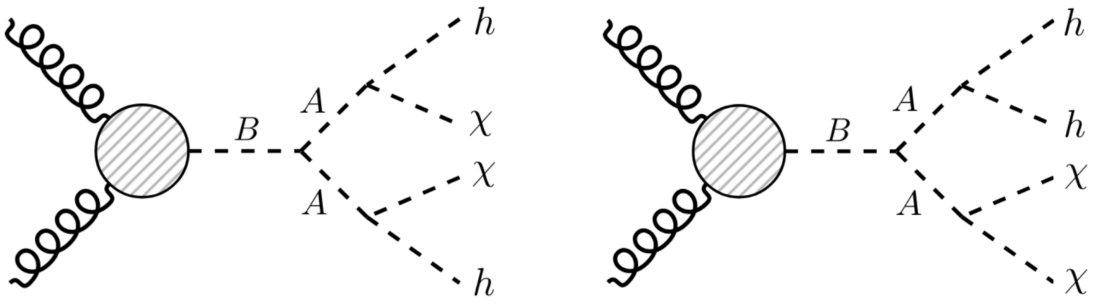
The dark-Higgs model is not unique in its resonant decay topology. In the tri-scalar models [72], both symmetric and resonant decay chains can occur, as illustrated in Fig. 1.12. The motivation for the development of these models lies in their ability to provide a minimal

and simple framework for exploring extended hidden scalar sectors interacting with the Higgs boson, particularly through di-Higgs plus missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) signatures. Three new scalars are postulated, namely the invisible χ, which acts as a candidate for DM, and the two heavy scalars, A and B. The decay chains are given by

$$\text{symmetric: } pp \to B \to AA \to (h\chi)(h\chi) \tag{1.65}$$

$$\text{resonant: } pp \to B \to AA \to (hh)(\chi\chi). \tag{1.66}$$

See Fig. 1.12 for the respective diagrams.



(a) Symmetric decay topology. (b) Resonant decay topology.

Figure 1.12: Diagrams of decays resulting in a di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ signature in the tri-scalar model.

The hypothetical scalar particles A, B, and χ are singlets under the SM gauge group and have a mass hierarchy of $m_B > 2m_A > 2m_\chi$. An additional Z_2 symmetry is introduced, where particles in the hidden sector are labelled as odd and particles in the visible sector are labelled as even. At the LHC, the heaviest scalar B can be produced via gluon-gluon fusion and is thus even under Z_2, decaying with a branching fraction of almost 100% to AA. Couplings for the decays $B \to \chi\chi$, $B \to A\chi$ and $B \to hh$ are assumed to be small. χ belongs to the hidden sector and, as the lightest odd particle, is stable, making it a valid candidate for DM.

A Z_2-odd A can decay via $A \to h\chi$, representing the symmetric event topology. On the other hand, if A is Z_2-even, it can decay via $A \to hh$ and $A \to \chi\chi$, resulting in the resonant topology. Because the analysis described within this thesis targets resonant signatures and the symmetric model has similarities to the MSSM electroweakino production, the

following will focus on the resonant case.

The interaction Lagrangian of this model can be expressed as follows:

$$\begin{aligned}\mathcal{L} &= \frac{C_{Bgg}}{\Lambda} B G^{a}_{\mu\nu} G^{\mu\nu a} + \frac{m_{BAA}}{2} BAA + \\ &+ m_{AHH} A H^{\dagger} H + \frac{m_{A\chi\chi}}{2} A\chi\chi + \lambda_{A\chi HH} A\chi H^{\dagger} H \ . \end{aligned} \tag{1.67}$$

In the symmetric topology, the third and fourth term are zero, and in the resonant topology, the last term is zero, as only χ belongs to the hidden sector. After electroweak symmetry breaking, in the resonant case, the coupling m_{AHH} introduces a mixing of A and the SM Higgs boson. Thus, A also couples to other SM particles, and hence the decay modes $A \to WW$ and $A \to ZZ$ are allowed.

The signal rate for the resonant topology can be estimated by

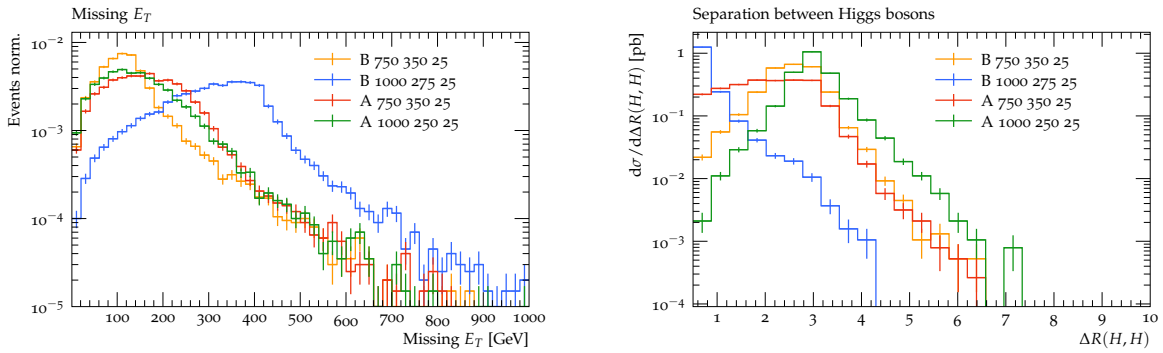
$$\begin{aligned}&\sigma(pp \to (4b + \chi\chi)) = \\ &\quad = \sigma(pp \to B) \times \mathcal{B}(B \to AA) \times 2\mathcal{B}(A \to hh)\mathcal{B}(A \to \chi\chi) \times \mathcal{B}^2(h \to bb) \ , \end{aligned} \tag{1.68}$$

where $\mathcal{B}(B \to AA) \approx 1$. The maximal decay rate into the two-Higgs final state is achieved by $\mathcal{B}(A \to \chi\chi) = 0.5$. $\mathcal{B}(A \to hh)$ is reduced to 0.125, due to the possible decay rates to other gauge bosons.

The cross section for $pp \to B$ production can be estimated by rescaling the SM Higgs cross section. The LHC Higgs cross section group provides the necessary contributions of gluon-gluon fusion to the production cross section [73].

The tri-scalar models demonstrate the distinct kinematics arising from different decay topologies, as shown in Figure 1.13, which presents missing transverse energy distributions from various signal points. The plot illustrates four distinct scenarios, where the kinematics of the final state are determined by the masses of the three particles and the associated mass splittings, namely Δ_{BA}, Δ_{Ah}, and $\Delta_{A\chi}$. The first of these determines the boost of the A states, and consequently splits the possible scenarios into compressed (small Δ_{BA}) and split spectra (large Δ_{BA}). In the resonant case, the E_T^{miss} distribution coincides with the transverse momentum p_T of the A state, and is thus directly determined by Δ_{BA}.

This results in a larger missing energy distribution for the resonant, split scenario (blue curve in Fig. 1.13(a)), where the system has a high boost and the angular separation ΔR of the two Higgs bosons is smaller, see Fig. 1.13(b)). In contrast, in the symmetric case, the E_T^{miss} distribution is limited by cancellations because the two A states are produced back-to-back. As our analysis relies on E_T^{miss} triggers, we will therefore be more sensitive to the resonant case with high mass splittings. Additionally, the ΔR distribution of the two Higgs in the symmetric, compressed scenario is relatively uniform due to the low boost, while it is shifted to higher values in the split distribution due to the back-to-back nature of the process.



(a) Missing transverse momentum.

(b) ΔR of two Higgs bosons.

Figure 1.13: Truth level missing transverse momentum and Higgs boson separation distributions for different signal points. Model A represents the symmetric topology and model B the resonant topology. The legend shows the masses of the three scalars in the format Model_m_B_m_A_m_χ.

Chapter 2

The ATLAS Experiment at the LHC

Contents

2.1 The Large Hadron Collider 47

2.2 The ATLAS Detector 52

2.3 Event Simulation 69

2.4 Reconstruction of Physics Objects 76

The research presented in this thesis utilises data from proton-proton collisions recorded by the ATLAS detector during the LHC's Run 2 between 2015 and 2018. This chapter provides an overview of the LHC, the ATLAS experiment, and the methods used to simulate and reconstruct the physics objects used in the author's research. For more detailed information on the LHC and the ATLAS experiment, please refer to Ref. [74] and Ref. [75], respectively.

2.1 The Large Hadron Collider

The LHC is the world's largest and most powerful particle accelerator and collider, located at the European Organisation for Nuclear Research (CERN) near Geneva, Switzerland. It spans a circumference of approximately 27 kilometres, buried around 100 meters beneath the Franco-Swiss border. It is designed for colliding both protons and heavy ions at unparalleled energies.

2.1.1 The Accelerator Complex

Protons are produced by ionising hydrogen atoms from a simple gas cylinder and are then accelerated through a series of linear and circular interconnected machines, each further increasing the particle energies.

The first stage of the accelerator complex was the Linear Accelerator 2 (LINAC2), which accelerates protons to up to 50 MeV. Higher energies are achieved using synchrotrons, which accelerate particles in closed loops by gradually increasing the magnetic fields used to bend the particle trajectories. The subsequent stages of synchrotrons, including the Proton Synchrotron Booser (PSB), Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), enable the acceleration of protons to energies of 1.4 GeV, 25 GeV and 450 GeV, respectively. Only the final stage of the acceleration complex is formed by the LHC itself, originally designed to accelerate protons to up to 7 TeV, resulting in unprecedented centre-of-mass energies of 14 TeV. Due to difficulties with the training of the superconducting magnets, this target energy has not yet been achieved. For an illustration of the full accelerator complex, see Figure 2.1.

The acceleration is achieved using 16 superconducting radio frequency (RF) cavities, operating at a frequency of 400 MHz. Protons have been successfully accelerated to a record of 13.6 TeV during Run 3, a process that takes approximately half an hour and thus involves more than ten million rotations.

The particle trajectories are bent by 1232 superconducting dipole magnets, each 15 meters long, generating fields of over 8 T. The magnets are cooled with liquid helium

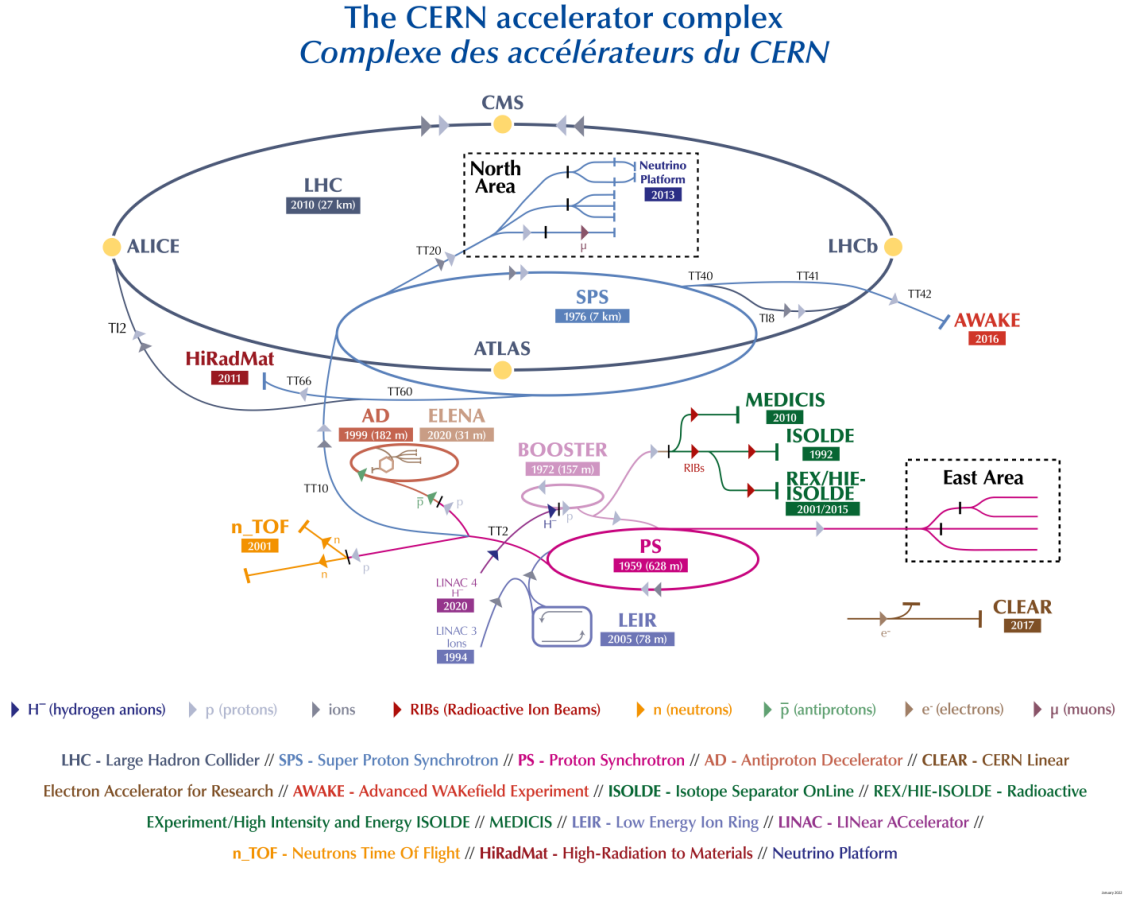


Figure 2.1: The CERN accelerator complex including the LHC, the SPS, the PS, the PSB and the LINAC2. The figure is taken from Ref. [76]

to operate at extremely low temperatures of 1.9 K. Quadrupole magnets are employed to focus the beams during the acceleration phase and to squeeze the beams to a few micrometers in the transverse plane at the four interaction points where the beams are crossed and the collisions occur. Additionally, sextupole, octopole, and decapole magnets are utilised to correct for other small destabilising effects.

The accelerator houses two beams travelling in opposite directions within separate tubes, maintained at an ultra-high vacuum to prevent interactions with gas molecules. Vacuum is also used as thermal insulator in the magnets and the helium distribution line. With a total volume of approximately 15,000 m^3, the combined vacuum system is the largest in the world.

2.1.2 Experiments at the LHC

The LHC has four major interaction points, each housing a distinct experiment, designed to explore different aspects of particle physics. Among these experiments are ATLAS and CMS [77], which are general-purpose detectors, capable of studying a wide range of phenomena. A Large Ion Collider Experiment (ALICE) [78] specialises in studying quark-gluon plasma, a state of matter that existed shortly after the Big Bang. Large Hadron Collider beauty (LHCb) [79] focuses on studying the behaviour of particles containing b- and c-quarks. Besides the four major experiments, smaller experiments such as TOTEM [80], LHCf [81] or FASER [82] focus on specific aspects of particle physics, such as measuring forward particle production.

2.1.3 The LHC Run 2

During Run 2 of the LHC (from 2015 to 2018), proton-proton collisions were conducted at a centre-of-mass energy of 13 TeV, a significant increase from the maximum energy of 8 TeV achieved during Run 1.

The beams consisted of up to 2544 bunches, each containing approximately 10^{11} protons. The individual bunches were spaced by intervals of 25 ns, resulting in a collision frequency of 40 MHz.

Besides the centre-of-mass energy, another critical measure of a particle collider's performance is the rate of physics events it can achieve. The rate of a certain process N is determined by its cross-section σ, which depends on the centre-of-mass energy, and the luminosity achieved at the interaction point:

$$N = L * \sigma = \sigma \int dt\mathcal{L} \ . \tag{2.1}$$

Here L denotes the integrated luminosity, and $\mathcal{L}$ the instantaneous luminosity. Assuming Gaussian beam profiles and head-on collisions, the instantaneous luminosity can be expressed as

$$\mathcal{L} = \frac{f N_1 N_2}{4\pi\sigma_x\sigma_y}, \tag{2.2}$$

where f is the bunch crossing frequency, N_1 and N_2 are the number of particles in each bunch, and σ_x and σ_y are the transverse beam sizes. At the LHC, the beams cross not head-on but in a small crossing angle, which slightly reduces the achieved luminosities.

Multiple dedicated detectors measure the luminosity at the LHC, which are calibrated in dedicated runs through van der Meer (vdM) scans [83]. The maximum recorded instantaneous luminosity during Run 2 was $2.07 \times 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [84]. The integrated luminosity delivered to the ATLAS experiment usable for physics research was measured using the LUCID-2 detector [85] as the primary luminometer, resulting in $(140.1 \pm 1.2)\,\mathrm{fb}^{-1}$ [86]. Figure 2.2 shows the cumulative luminosity delivered to and recorded by the ATLAS experiment between 2015 and 2018.

Due to the high luminosities, multiple hard proton-proton interactions typically occur simultaneously within a single bunch crossing. This phenomenon, known as pile-up, presents significant challenges for both the hardware and read-out systems, as well as for reconstruction software and data analysis, complicating the identification of individual collision events. During Run 2, the average pile-up was 33.7, as illustrated in Figure 2.3.

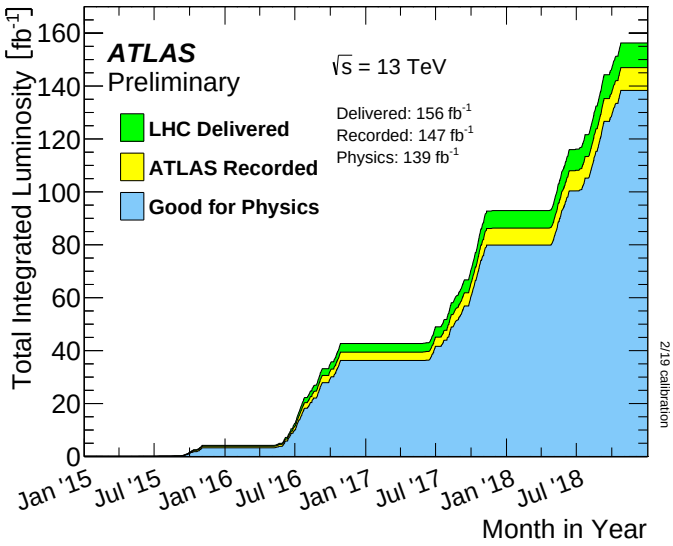


Figure 2.2: Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in Run 2. The delivered luminosity accounts for the luminosity delivered from the start of stable beams till the beam dump or beam studies. The recorded luminosity reflects inefficiencies in the data acquisition. Data quality criteria have to be satisfied to label data as good for physics. The figure and caption are adapted and taken from Ref. [87].

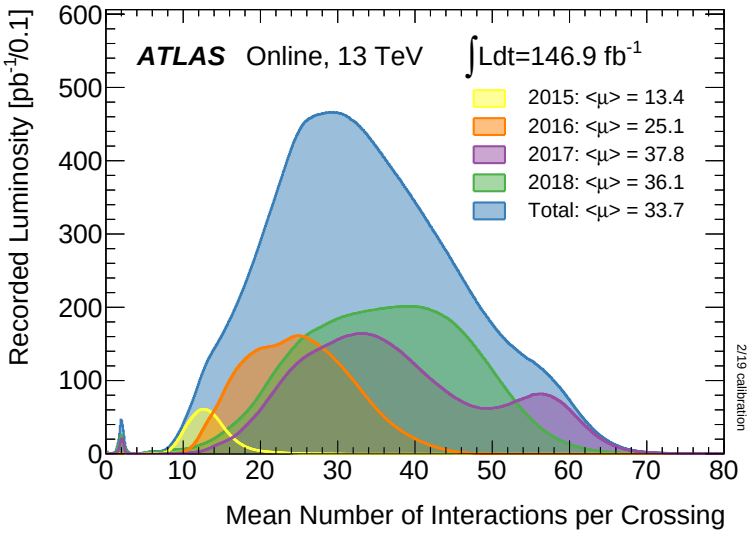


Figure 2.3: Shown is the luminosity-weighted distribution of the mean number of interactions per crossing for the 2015 - 2018 pp collision data at 13 TeV centre-of-mass energy. All data recorded by ATLAS during stable beams is shown, and the integrated luminosity and the mean mu value are given in the figure. The mean number of interactions per crossing corresponds to the mean of the Poisson distribution of the number of interactions per crossing calculated for each bunch. The figure and caption are adapted and taken from Ref. [87].

2.2 The ATLAS Detector

The construction of the ATLAS detector was completed at interaction point 1 of the LHC in 2008. With a diameter of 25 meters and a length of 44 meters, it is one of the largest particle physics detectors ever built.

ATLAS is a general-purpose detector, engineered to conduct a wide range of research activities, including precision measurements of the SM, the search for the Higgs boson, and the exploration of potential new phenomena such as SUSY and DM.

The physics analyses are based on objects reconstructed from measurements of energy, tracks, and momenta of particles produced in the collisions. The cylindrical detector consists of multiple layered subsystems, each tasked with specific functions. See Figure 2.4 for an illustration of the entire detector and its components. The innermost layer, the Inner Detector (ID), is responsible for high-precision tracking of charged particles. Encasing the ID, the calorimeter system measures the energies of particles. The outermost subsystem is the Muon Spectrometer (MS), designed to detect and measure the tracks and momenta of muons. Strong magnetic fields generated by superconducting magnets bend the trajectories of charged particles, allowing for precise measurements of particle momenta.

Combining the measurements from all subsystems enables the reconstruction of physical objects to perform high-precision measurements and searches for new physics.

2.2.1 The Coordinate System

The ATLAS experiment uses a right-handed coordinate system, as depicted in Figure 2.5. The x-axis points towards the centre of the LHC, the y-axis points upwards, and the z-axis in the direction of the beam towards Lake Geneva.

Given the symmetry of the experiment, cylindrical coordinates (r, ϕ, z) are employed. Here, ϕ denotes the azimuthal angle, r represents the radial distance, and z is the distance along the beam axis measured from the interaction point. Additionally, the angle θ with respect to the beam axis z is used to indicate the directions in which objects are moving.

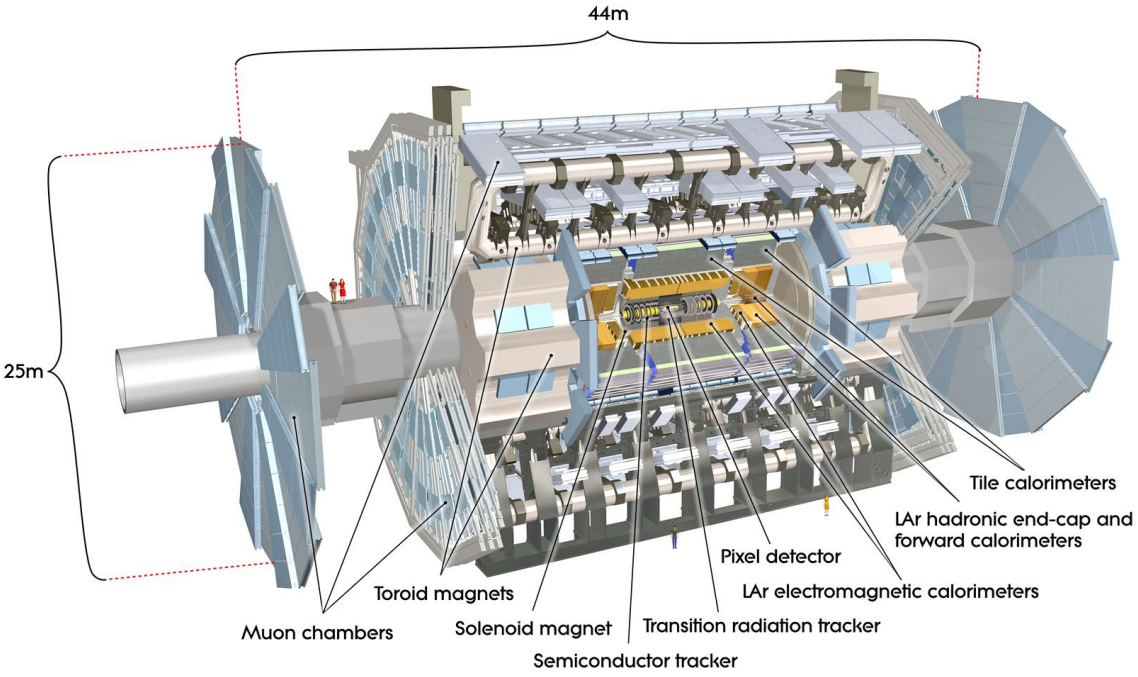


Figure 2.4: Illustration of the whole ATLAS detector and its subsystems during Run 2. The figure is taken from Ref. [88].

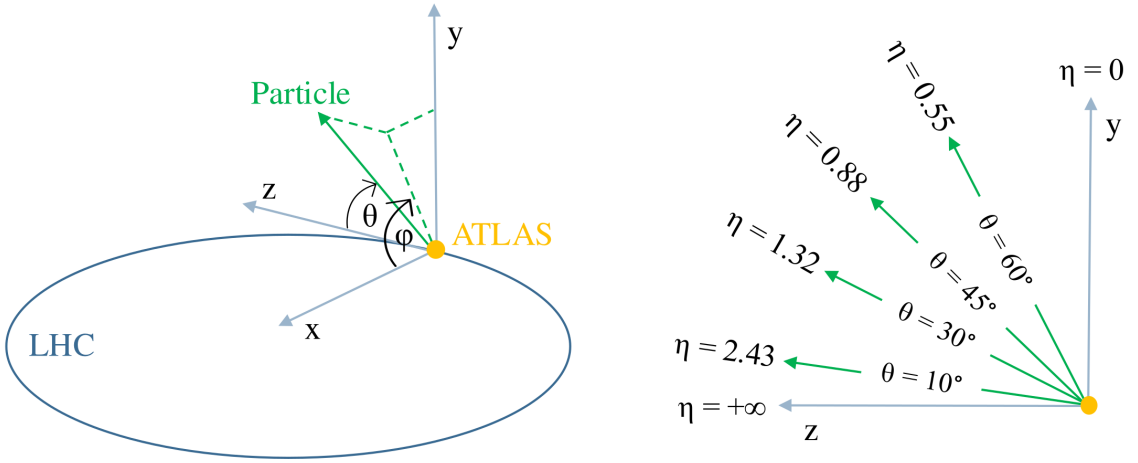


Figure 2.5: Illustration of the ATLAS coordinate system (left) relative to the LHC. Relationship of pseudorapidity η and polar angle θ (right). The momentum of a particle emerging from the point of interaction is depicted in green. The figure is inspired by Ref. [89].

In practice, however, the polar angle θ is replaced by the pseudorapidity η, defined as:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \ . \tag{2.3}$$

In the ultra-relativistic limit, pseudorapidity converges to the definition of rapidity. Rapidity differences are Lorentz invariant under longitudinal boosts, and thus transform additively. This property is crucial because the partons within the colliding protons carry varying and unknown momentum fractions, leading to distinct boosts of the rest-frames in the collisions. Consequently, pseudorapidity differences $\Delta\eta$ remain invariant regardless of the boost of the collision's rest-frame. Figure 2.5 illustrates the relationship between η and θ.

A common measure of the angular separation of particles is

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \ , \tag{2.4}$$

which also remains Lorentz invariant under longitudinal boosts.

Since the momentum fractions of partons along the beam axis are unknown, kinematic variables are frequently defined in the transverse plane (x-y). The key variables include the transverse momentum p_{T} and the transverse energy E_{T} of a particle.

While the ATLAS detector is capable of performing measurements over the full range of ϕ, there is a limit on measurements in the forward direction. Particles can be detected up to a pseudorapidity of $|\eta| < 4.9$. However, different subsystems cover different ranges, as shown in Table 2.1.

2.2.2 Particles in the ATLAS Detector

To classify different particles with the detector, the design ensures that various particles leave distinct signatures in different subsystems.

Charged particles leave traces in the ID, and the curvature of their trajectories allows conclusions about their momenta. Quarks and gluons undergo hadronisation and subsequent decay, resulting in jets containing many particles. These jets leave tracks

System	Sub-Detector	Coverage in $\lvert\eta\rvert$
Inner Detector	Pixel Detector	< 2.5
	Semiconductor Detector	< 2.5
	Transition Radiation Tracker	< 2.0
Calorimeters	Electromagnetic Calorimeter	< 3.2
	Hadronic Calorimeter Endcap	[1.5, 3.2]
	Forward Calorimeter	[3.1, 4.9]
	Hadronic Calorimeter Barrel	< 1.7
Muon Spectrometer	Resistive Plate Chambers	< 1.05
	Thin Gap Chambers	[1.05, 2.4]
	Monitored Drift Tubes	< 2.7
	Cathode Strip Chambers	[2.0, 2.7]

Table 2.1: Summary of subdetectors and their coverage in pseudorapidity.

within the trackers and dissipate their energy within the Electromagnetic Calorimeter (ECal) and Hadronic Calorimeter (HCal). Electrons and photons deposit their energy within the ECal and muons are able to traverse all layers, leaving tracks in the ID and the MS. Short-lived particles like the top quark, as well as $W^{\pm}$, Z, and Higgs bosons, decay before interacting with the detector and are thus reconstructed from their decay products. Particles with slightly longer but finite lifespans decay during their transit through the detector, creating observable decay vertices away from the initial interaction point. An important example are B-hadrons with lifetimes of around 10^{-12} s and decay lengths of multiple millimetres. Kaons composed of strange quarks and charged pions exhibit lifespans around 10^{-8} s, allowing them to traverse the entire tracker system. Neutral pions decay almost instantly (around 10^{-16} s) into photons, depositing energy in the ECal. As primary products of hadronisation, pions are among the most frequently produced particles in high-energy collisions. See Figure 2.6 for an illustration of interactions of different particles with the detector.

Table 2.2 summarises the interactions of the SM particles with the different detector

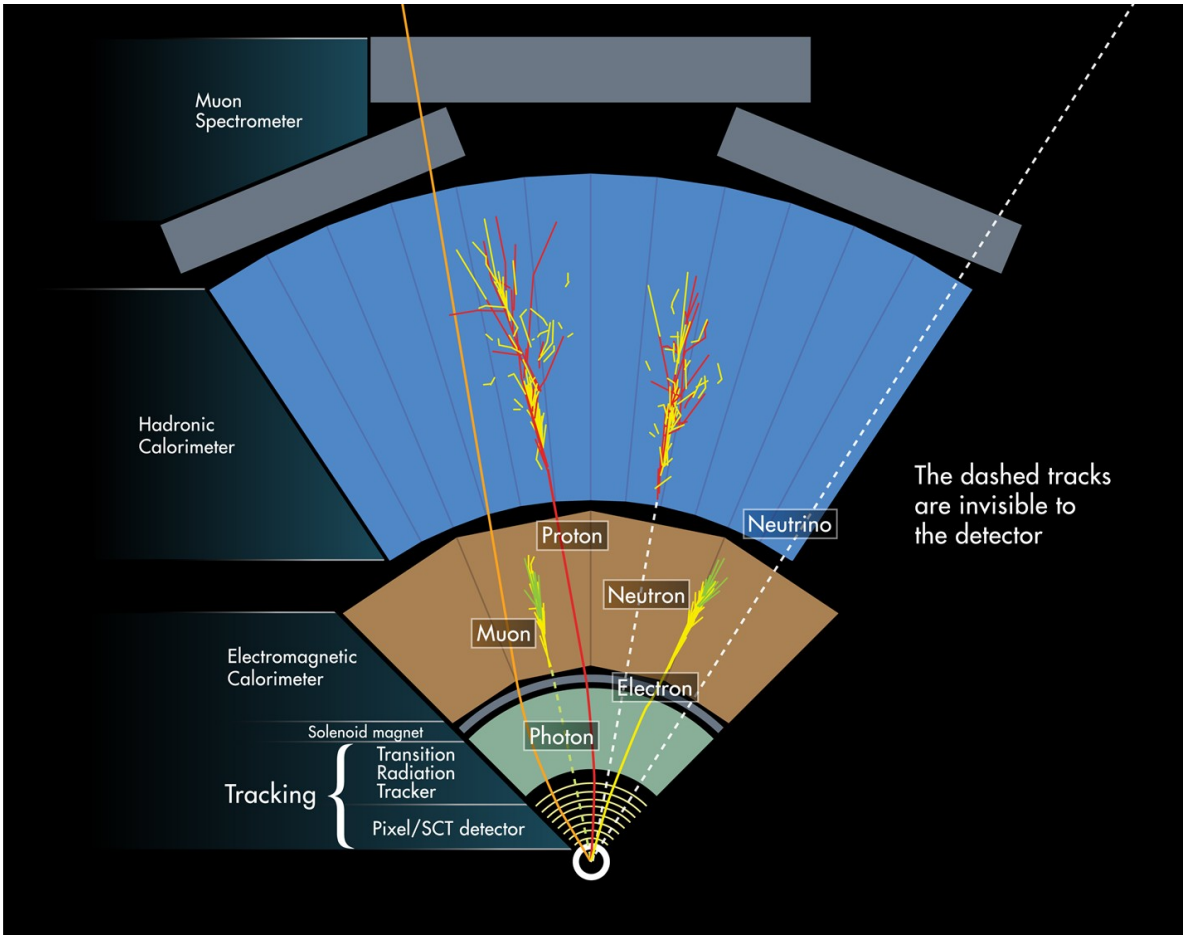


Figure 2.6: A slice through the ATLAS detector depicting the subsystems and its layered structure. Typical interactions of particles with the different subdetectors are shown. The figure is taken from Ref. [90].

subsystems.

The subsequent sections will detail the detection methods and technologies utilised by the ATLAS subdetectors and the specific particle signatures they are designed to capture.

2.2.3 The Inner Detector

The innermost layers of the detector, which serve as particle trackers, form the ID. Located closest to the beampipe and the interaction point, they are subjected to high radiation. These layers are enclosed by a solenoid magnet that provides a 2 T magnetic field parallel to the beam pipe.

The main detection principle relies on sampling the trail of ionised atoms and liberated electrons that a charged particle leaves behind when traversing detector material. The

Particle	**Signature**	**Sub-Detectors**
u, d, s	jet	ID, ECal, HCal
c, b	jet + displaced vertex	ID, ECal, HCal
t	$t \to bW$: invariant m_t	ID, ECal, HCal, (MS)
ν_e, ν_μ, ν_τ	E_T^{miss}	all (inferred)
e	track	ID, ECal
μ	tracks	ID, MS
τ (hadronic)	jets + displaced vertex	ID, ECal, HCal, (MS)
g	jet	ID, ECal, HCal
γ	no track	ECal
$W^{\pm}$, Z, h	invariant m_X	depends on decay

Table 2.2: SM particles and their typical signatures within the detector.

two main tracking technologies used are drift tubes and semiconductors, primarily silicon pixels and strips. In drift tubes, a tube is kept at high voltage relative to a wire at its centre. The tubes are filled with gas that is ionised by passing particles, and the liberated electrons are collected. The semiconductor technology operates on a similar principle. Particles ionise atoms in doped silicon wafers. The created holes drift towards the negative pole and are collected at p-n junctions.

The ID consists of three subdetectors: the Pixel detector (including the Insertible B-Layer (IBL)), the Semiconductor Detector (SCT), and the Transition Radiation Tracker (TRT). The first two subdetectors use silicon sensors, while the TRT employs drift tubes and enhances its detection capabilities using transition radiation. The silicon detectors are designed for high-precision tracking, whereas the TRT, located further from the interaction point, provides a large number of measurement points, improving track reconstruction. The interactions of charged particles with the different layers are called hits. These hits are used for track reconstruction of the charged particles, allowing to infer a particle's momentum and charge.

Each subdetector comprises two systems. The cylindrical layers around the beam axis form the barrel part of the detector. The endcaps, positioned on both sides of the detector, consist of disks perpendicular to the beam pipe. The subdetectors, their η coverage, active

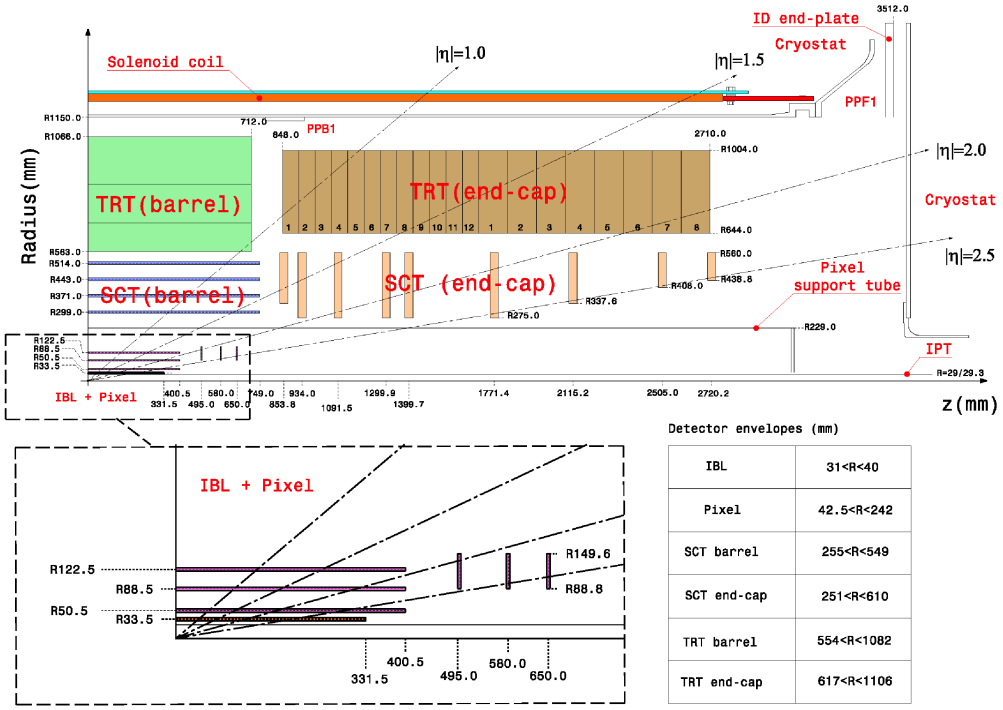


Figure 2.7: The layout of the ATLAS ID detector. The figure and caption are taken and adapted from Ref. [91]

material, and resolution are summarised in Table 2.3. See Figure 2.7 for the layout of the ID including all subsystems and layers. The ID is designed to achieve a momentum resolution of

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = 0.05\% p_\mathrm{T} \oplus 1\% \ . \tag{2.5}$$

The first term is proportional to the p_T, as it becomes increasingly difficult to determine the curvature of very high p_T particles, which appear almost as straight lines despite the magnetic fields in the tracker. The following sections will briefly introduce the three subdetectors.

The Pixel Detector

The Pixel detector consists of four layers of silicon pixels in the barrel and three layers in the endcaps. The innermost layer is the IBL [93], which was installed during the long shutdown before Run 2 to improve the impact parameter resolution [91]. This enhancement allows better Secondary Vertex (SV) reconstruction and thus more effective flavour tagging. The other layers are the B-Layer, Layer-1, and Layer-2. The layers

Sub-Detector	**Coverage in** $\lvert\eta\rvert$	**Active Material**	**Resolution in** $r-\phi$ [µm]	**Resolution in** z [µm]
IBL	< 3.0	Silicon	10	60
Pixel	< 2.5	Silicon	10	115
SCT	< 2.5	Silicon	17	580
TRT	< 2.0	Ar and Xe Gas	130	

Table 2.3: Summary of the ID subdetectors, their coverage in pseudorapidity, active material and spatial resolution. The values for the detector resolutions are taken from Ref. [92]

are arranged such that a particle traversing the detector passes through at least three layers. As depicted in Figure 2.8, the four layers are placed at radii of $r = 33.25$ mm, $r = 50.5$ mm, $r = 88.5$ mm, and $r = 122.5$ mm from the beam axis. Consequently, the Pixel detector must tolerate very high radiation doses. The layers of the two endcaps are located about half a meter from the interaction point in the z direction. The pixels of the IBL have dimensions of 50 µm×250 µm, while the sensors in the other layers have dimensions of 50 µm×400 µm [92]. Overall, the Pixel detector consists of over 92 million pixels arranged in 2024 modules.

The Semiconductor Detector

The SCT forms the middle part of the ID. To cover much larger areas than the Pixel detector, it uses long strips instead of small pixels. In the barrel region, the SCT consists of four concentric layers enclosing the Pixel detector. The first layer is located at 299 mm and the fourth layer at 514 mm from the beam. The endcaps are formed of nine disks each, positioned between $z = 853.8$ mm and $z = 2720.2$ mm. Both the barrel and the endcaps consist of around 2000 modules, each built of four rectangular silicon-strip sensors. Two almost square layers, each approximately 64 mm×64 mm, are positioned on top of the baseboard and two below it. The strips are around 12 cm long. In the endcaps, the sizes of the sensors vary slightly depending on the layer. In total, the silicon sensors of the SCT span around 60 m^2 and are composed of approximately six million silicon

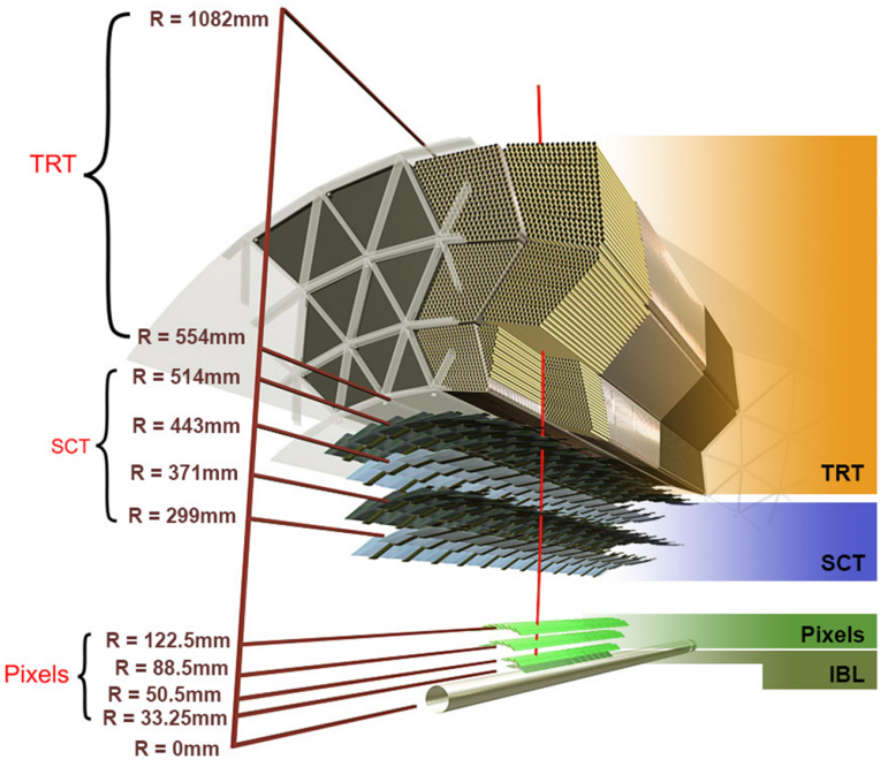


Figure 2.8: Illustration of the ATLAS ID in the barrel region. The structure of the ID made of highly granular silicon pixels, silicon strips and straw tubes. The figure and caption are taken and adapted from Ref. [94].

micro-strips. This configuration results in a resolution of 17 µm in r-ϕ and 580 µm in the z direction [95].

The Transition Radiation Tracker

The outermost layers of the ID are constructed from around 300,000 drift tubes, referred to as straws. The TRT utilises two detection mechanisms. The straw tubes serve as gaseous ionisation detectors, providing a pattern of hits used for track reconstruction. Additionally, materials with different refraction indices cause relativistic charged particles to emit transition radiation. The amount of transition radiation emitted depends on the γ factor of a particle, allowing the distinction between lighter particles, such as electrons, and heavier particles, such as pions.

The thin-walled tubes, with a diameter of 4 mm, are filled with Argon or Xenon-based gas mixtures and contain a 30 µm thick gold-plated tungsten wire in their center. The 50,000 straws in the barrel, each 144 cm long, are arranged in 73 co-axial layers, with the

outermost layer located at $r = 1082\,\text{mm}$. In the endcaps, 250,000 straws, each 39 cm long, form 160 planes where the tubes are arranged perpendicular to the beam axis.

The TRT contributes on average 30 hits with an intrinsic resolution of approximately 120 µm for charged particle tracks with $|\eta| < 2$ and $p_{\text{T}} > 0.5$ GeV [96].

2.2.4 The Calorimeters

Calorimeters in ATLAS are designed to absorb the majority of particles created in the collisions. With the exception of muons, neutrinos, and in rare cases punch-through jets, SM particles deposit all their energy and thus stop within the calorimeters. ATLAS uses sampling calorimeters, built from alternating layers of high-density absorber materials and active material. The incoming high-energy particles interact with the absorber, initiating rapidly evolving showers that break down their energy into a multitude of lower-energy particles. The energy of the particles in the shower is measured within the active layers once their energies are low enough to be stopped.

The ATLAS calorimeter system consists of two subsystems: the ECal and the HCal. The fine granularity of the first is designed to precisely determine the energies of photons and electrons, while the coarser granularity of the latter is sufficient to measure the energies of hadrons. Figure 2.9 illustrates the layout of the calorimeter system, and Figure 2.10 shows a sketch of an electromagnetic shower developing within a calorimeter. The following sections provide a brief introduction to the ECal and HCal.

The Electromagnetic Calorimeter

The ECal utilises LAr as the active sampling material and is therefore also referred to as the LAr calorimeter. LAr was chosen due to its radiation hardness and intrinsic linear response to the amount of energy deposited. However, maintaining LAr requires cooling the calorimeter to −184 °C. The absorber plates are made of lead. The active layers, absorber layers, and electrodes are arranged in an accordion-like shape. The barrel covers $|\eta| < 1.475$, and the endcaps cover $1.375 < |\eta| < 3.2$. The granularity of the calorimeters decreases with r. Incoming particles ionise the LAr in the active layers, and the energies

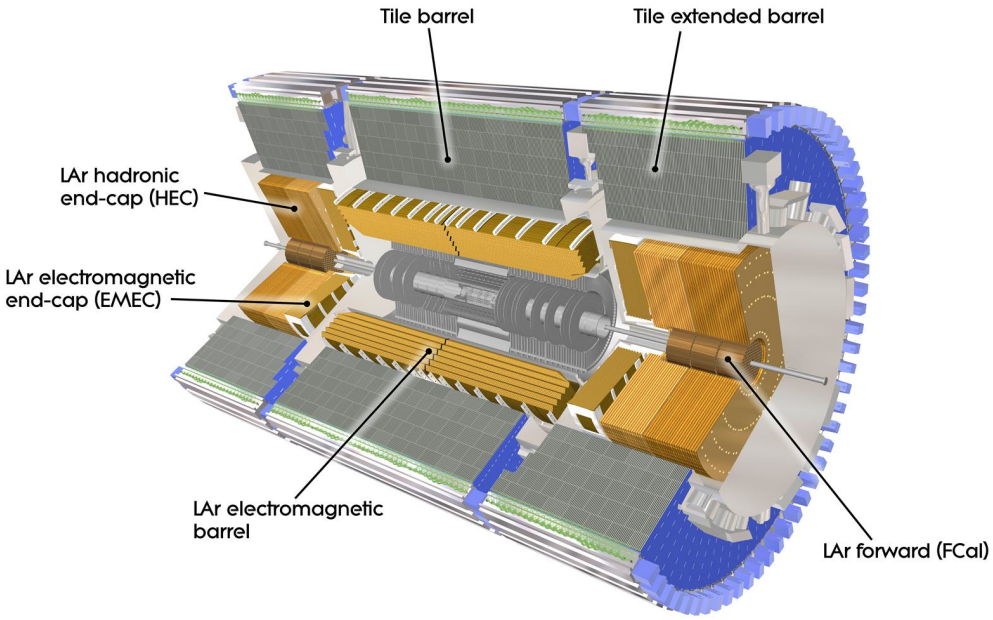


Figure 2.9: Illustration of the ATLAS calorimeter system, including the ECal with Liquid Argon (LAr) as active material and the HCal with plastic scintillator tiles. Both the barrel and endcaps are depicted. The figure is taken from Ref. [97].

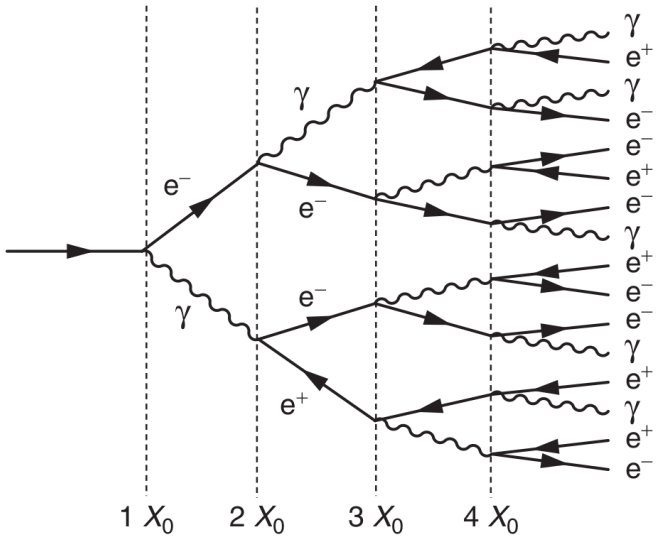


Figure 2.10: The development of an electromagnetic shower, where the number of particles roughly doubles after each radiation length. The figure and caption are taken from Ref. [8].

of the electrons and photons are sampled and measured. Combining signals from all layers allows the reconstruction of particle energies. The energy resolution of the LAr calorimeters is

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \ . \tag{2.6}$$

The first term is of a stochastic nature, while the second term includes effects such as energy deposition outside the active layers of the calorimeter, LAr impurities, leakages, and temperature effects [98].

The ECal has a depth of around 22 radiation lengths, approximately half a metre, in the barrel, and 24 radiation lengths in the endcaps. A radiation length X_0 is the depth at which an electron's energy is reduced by $1/e$. The ECal is large enough to contain full electromagnetic showers. Hadrons also start showering within the ECal and are stopped in the HCal.

The Hadronic Calorimeter

The HCal barrel, covering $0.8 < |\eta| < 1.7$, is also a sampling calorimeter, composed of sandwiched layers of steel (as the absorber) and plastic scintillator tiles (as active material). In the scintillator, the energies of the particles are absorbed and emitted as photons, which are then converted to electric signals through photomultipliers. It is specifically designed to measure the energies of hadrons, which are not stopped in the LAr calorimeters.

Similar to the ECal, the endcaps use LAr as the active material and copper as the absorber. They cover $1.5 < |\eta| < 3.2$. Additionally, the forward hadron calorimeters, composed of LAr and tungsten, are located further up the beam line and cover $3.1 < |\eta| < 4.9$.

In contrast to electrons and photons, which are the primary targets of the ECal, hadrons also interact strongly with the nuclei of the calorimeter material. However, the nuclear interaction lengths are significantly larger than the radiation lengths. Thus, the central HCal is built from 64 wedges, each 2.6 m long, which equals about ten nuclear interaction lengths. Overall, the ATLAS HCal consists of about 420,000 scintillator tiles and weighs almost 2900 tonnes.

The HCal is designed to achieve energy resolutions of:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \quad \text{in the barrel and endcap} \tag{2.7}$$

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \quad \text{in the forward calorimeter .} \tag{2.8}$$

It stops all SM particles created at the interaction point, except for muons, neutrinos, and very high energetic particles in jets, which can punch through into the muon spectrometer.

2.2.5 The Muon Spectrometer

The MS [99] forms the outermost part of the ATLAS detector. Due to their much higher mass, muons interact only weakly with the detector and consequently traverse it without losing much of their energy. Compared to electrons, which lose a significant amount of energy to Bremsstrahlung, and hadrons, which interact with the material through the strong nuclear force, the interaction of relativistic muons with the detector material can be well modelled by the Bethe-Bloch equation [100]. Consequently, muons with energies of multiple GeV travel many metres in high-density materials. With a lifetime of more than 2 µs, they are also stable on the timescales relevant to detection. Thus, muons usually traverse the full detector.

The MS is designed to determine the presence of muons in an event, measuring the muon p_{T}, and serving as a trigger. It contains multiple different technologies, such as Monitored Drift Tubes (MDTs), Cathode Strip Chambers (CSCs), Resistive Plate Chambers (RPCs), and Thin Gap Chambers (TGCs). Figure 2.11 shows the layout of the muon system and the placement of the different detector types. To bend the muon trajectories and allow for good momentum resolution, the MS is filled with a magnetic field, which is generated by one large 830 tonne air-core toroid in the barrel and two smaller ones in the endcaps. The toroids generate magnetic fields up to 3.5 T with field lines following the ϕ coordinate and thus approximately perpendicular to the incoming

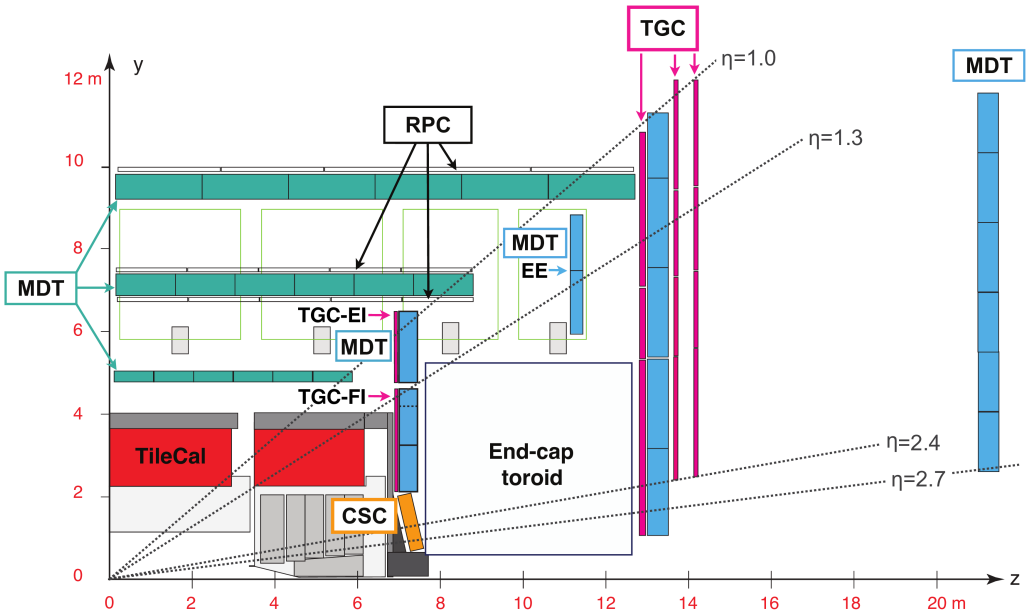


Figure 2.11: A schematic picture showing a quarter-section of the muon system in a plane containing the beam axis, with MDTs and cathode strip CSCs for momentum determination and RPCs and TGCs for triggering. The Forward Inner and Endcap Inner TGCs are marked as TGC-FI and TGC-EI, respectively. The Extended Endcap MDTs chambers are referred to as EE. The figure is taken from Ref. [101].

muon trajectories. The MS is designed to achieve a muon momentum resolution of

$$\frac{\sigma_{p_T}}{p_T} = 10\% \tag{2.9}$$

at $p_T = 1$ TeV. The following sections will briefly introduce the different detector technologies employed within the MS.

Monitored Drift Tubes

Similar to the straws in the TRT, the MDTs consist of 3 cm wide aluminium tubes with 400 µm wall thickness, filled with a gas mixture based on Argon and CO_2. The tubes have different lengths ranging between 85 cm and 630 cm. The gas atoms are ionised by the muons, and the electrons are collected at a tungsten-rhenium wire in the centre of the tube. More than 350,000 drift tubes are installed in multiple layers in both the barrel and the endcaps, see Figure 2.11. The MDTs are the main detectors for muon precision tracking, and single drift tubes are designed to achieve a resolution of 80 µm.

Cathode Strip Chambers

In the forward region, for high $|\eta|$ between 2.0 to 2.7, CSCs are employed. These detectors offer higher radiation hardness and finer granularity. CSCs are multiwire proportional chambers filled with Argon and CO_2 gas. The principle is very similar to that of drift tubes. However, in this case, larger chambers with multiple wires are used. The ionised atoms and liberated electrons cause avalanches of charged particles, which are collected by the wires closest to the trajectory of the ionising particle. The sampled charge is proportional to the ionisation caused by the muon. Resolutions of 60 µm are achieved.

Resistive Plate Chambers

RPCs are used to trigger on muons and are installed in the barrel. In addition to their function as triggers, they also provide additional measurements to the MDTs and CSCs in the direction parallel to the magnetic field. In RPCs, the avalanche of charged particles resulting from the ionisation of the gas in the chambers is collected by two parallel plates, serving as the electrodes, and not by dedicated wires, allowing for lower read-out times. Time resolution of $\mathcal{O}$(1ns) allows triggering on the bunch-crossing level. The downside is a relatively low spatial resolution of $\mathcal{O}$(1cm).

Thin Gap Chambers

TGCs serve a similar purpose as the RPCs. They are also primarily used for triggering on muons but are located in the endcaps. They consist of parallel 30 µm wires within a gas mixture with a high percentage of CO_2 and provide low enough time resolutions to distinguish between bunch crossings.

2.2.6 The Trigger System

As collisions occur at a frequency of 40 MHz, the amount of data created is enormous. Combined with a pile-up of around 30, the high bunch-crossing frequency results in $\mathcal{O}(10^9)$ collisions per second. Since a raw data event requires more than a MB of storage space,

this results in many terabytes of data per second, which is too much to be processed and stored. However, the majority of collisions involve only small momentum transfers. Since such events are not the main target of the ATLAS physics program, a trigger system is used to decide which events are to be stored and which are discarded, thereby reducing the amount of required storage by multiple orders of magnitude.

The triggers are designed to accept events with interesting signatures, such as high $E_{\mathrm{T}}^{\mathrm{miss}}$ or a certain number of b-tagged jets or leptons. During Run 2, a two-stage system, consisting of a hardware-based Level-1 Trigger (L1) and a software-based High-Level Trigger (HLT), was employed. The requirements for an interesting signature are combined in a trigger chain, consisting of a L1 and a HLT item. Many chains together form a trigger menu, which is specified for each run. Furthermore, pre-scales can be applied to store only a fraction of events of a certain signature that pass the trigger criteria to further reduce the overall acceptance rate. The following sections will briefly discuss the two stages of the trigger. For further details, refer to Ref. [102], which serves as a general reference for this section.

The Level-1 Trigger

The L1 trigger is a hardware-based system utilising custom electronics to process information from the calorimeters and MS at reduced granularity [103].

The L1 calorimeter trigger (L1Calo) [104] digitises and calibrates the analogue signals from the calorimeters, then routes them to both the Cluster Processor (CP) and the Jet/Energy-sum Processor (JEP). The CP identifies candidates for electrons, photons, and tau-leptons above set thresholds, while the JEP focuses on identifying jet candidates and calculating global sums of total and missing transverse energy. The signals from the LAr calorimeter span multiple bunch crossings, thus measurements can be affected by neighbouring collisions, known as out-of-time pile-up.

The L1 muon trigger (L1Muon) [101] utilises hit patterns from the RPCs and TGCs and compares them to a pattern caused by a muon with infinite momentum.

The L1 topological trigger (L1Topo) [105] applies requirements to kinematic and

geometric variables, such as invariant masses or angular distances, computed by combining objects received from L1Muon and L1Calo.

The Central Trigger Processor (CTP) combines signals from the L1Calo, L1Muon, L1Topo, and several other subdetectors to form the L1 decision, effectively managing the overall trigger strategy. The CTP also enforces dead times to prevent read-out overlaps and buffer overflows, ensuring the detector operates within its read-out latency constraints.

The L1 trigger reduces the event rate from 40 MHz down to 100 kHz by leveraging the topological data as well as measurements like total calorimeter energy and object multiplicities. Additionally, the L1 identifies regions of interest (RoIs) in η and ϕ.

Once an event is accepted, it is read by the front-end (FE) electronics and sent to the ReadOut Drivers (RODs), where it is processed and buffered before being sent, together with the RoIs, to the next stage, the HLT.

The High-Level Trigger

The HLT operates on a software basis, utilising a farm of approximately 40,000 dedicated processing units. It first employs fast algorithms for initial data reduction, followed by more intensive algorithms for final event selection. These algorithms operate within designated RoIs, extracting features and validating trigger criteria within milliseconds. In cases requiring comprehensive analysis, the HLT requests data from the entire detector to reconstruct objects and quantities such as $E_{\mathrm{T}}^{\mathrm{miss}}$. The HLT accepts events at an average rate of 1.2 kHz. Accepted events are subsequently sent to permanent storage for detailed offline analysis.

2.3 Event Simulation

The majority of physics analyses at ATLAS rely on simulated data. For instance, in searches for BSM physics, events of signal models, such as those detailed in Section 1.3, are simulated. Regions with a high concentration of signal events are isolated and investigated. Observed data are compared to expectations from SM processes, and potential excesses and discrepancies are studied and statistically analysed (see Chapter 3). While SM backgrounds can be estimated by data-driven methods, many analyses rely on simulations to determine the yields and kinematic behaviour of SM background processes.

Hadron-hadron collisions allow us to achieve very high centre-of-mass energies, but the composite nature of hadrons necessitates complex multi-particle interactions in very high-dimensional phase spaces. Some particles created in the collisions traverse the detector, others initiate entire decay chains, and some of their decay products hadronise and shower, subsequently interacting with the detector material in different ways. The ATLAS detector provides data consisting of hits in the tracker and the MS, as well as energy deposits in the calorimeters originating from all these particles, as detailed in Section 2.2. Monte-Carlo (MC) simulations are utilised to simulate all these different processes, starting with the hard-scatter interaction, the parton shower (PS), the hadronisation, as well as multiple parton interactions (MPI) and the interactions of particles with the detector. Figure 2.12 shows a sketch of a $pp \to tt$ event generated through Pythia [106].

As it is impractical to generate MC samples for every possible process in a proton-proton collision, only processes of interest and significant backgrounds that could produce similar signatures in the detector are simulated. Markov-Chain Monte Carlo techniques are used to sample representative event ensembles from the probability distributions of the processes.

The following sections provide a short introduction to the different techniques used for the various stages of the simulation process. For detailed information on simulations employed in ATLAS and LHC experiments in general, refer to Refs. [107] and [108].

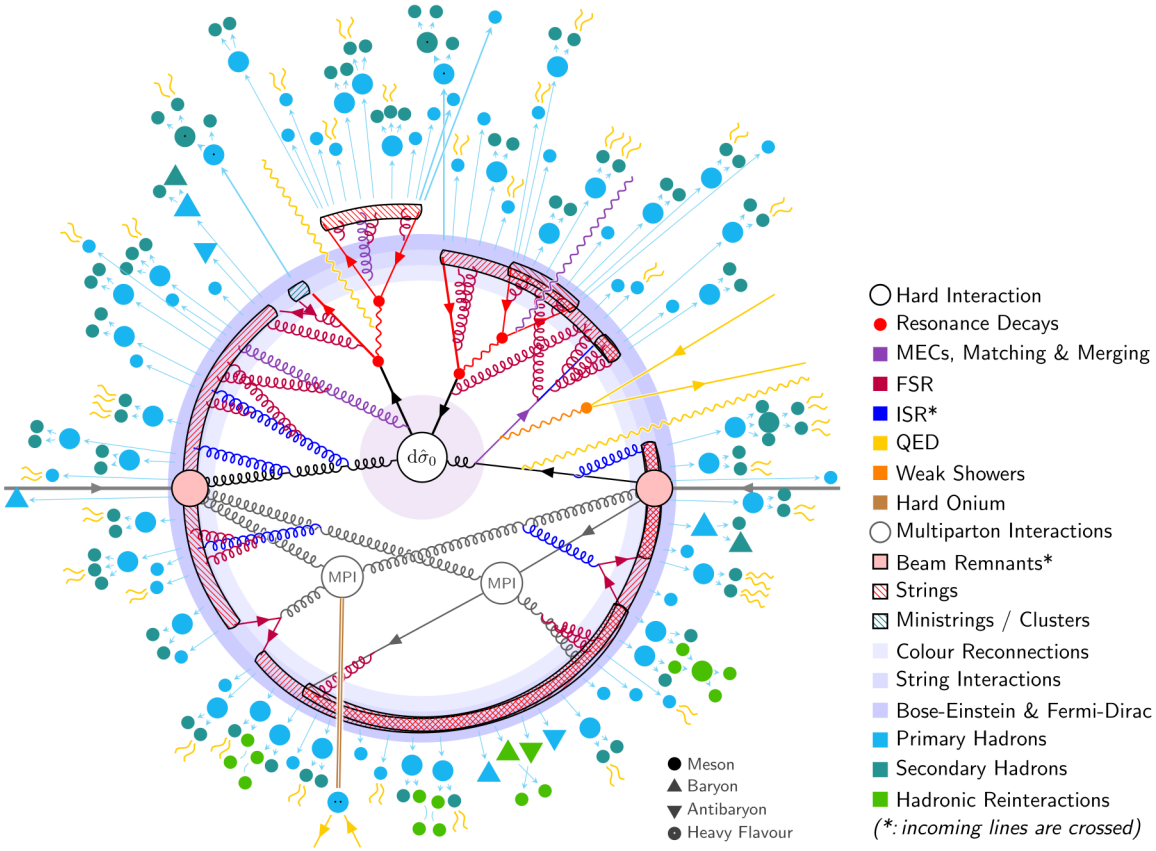


Figure 2.12: Schematic of the structure of a $pp \to tt$ event, as modelled by PYTHIA. To keep the layout relatively clean, a few minor simplifications have been made: 1) shower branchings and final state hadrons are slightly less numerous than in real PYTHIA events, 2) recoil effects are not depicted accurately, 3) weak decays of light-flavour hadrons are not included (thus, e.g. a K_S^0 meson would be depicted as stable in this figure), and 4) incoming momenta are depicted as crossed ($p \to -p$). The latter means that the beam remnants and the pre- and post-branching incoming lines for ISR branchings should be interpreted with "reversed" momentum, directed outwards towards the periphery of the figure; this avoids beam remnants and outgoing ISR emissions having to criss-cross the central part of the diagram. The figure and caption are taken from Ref. [106].

2.3.1 The Hard-Scatter Interaction

The hard-scatter interaction in a proton-proton collision refers to the parton-parton interaction with the highest momentum transfer. The cross-section for this interaction, involving partons a and b, can be computed using

$$\sigma = \sum_{a,b} \int_0^1 \int_0^1 dx_a \, dx_b \, f_1(x_a, \mu_F) \, f_2(x_b, \mu_F) \, d\hat{\sigma}(ab \to X)(\mu_F, \mu_R) \ , \tag{2.10}$$

where f_i represents the parton distribution function (PDF) that describe the momentum distributions of the partons within the proton. The x_i denote the momentum fractions of the interacting partons.

The factorisation scale μ_F delineates the boundary between non-perturbative soft effects and perturbative hard physics. While hard interactions are included in the matrix element, non-perturbative effects are incorporated into the PDFs. This separation is referred to as the factorisation theorem. The PDFs are extracted from experimental data, particularly deep inelastic scattering experiments. The dependence on the unphysical μ_F can be reduced by including higher-order terms in the matrix element calculations. However, these higher-order calculations are computationally expensive and are often truncated at leading order (LO), next-to leading order (NLO), or next-to-next-to leading order (NNLO).

$d\hat{\sigma}(ab \to X)$ represents the parton-level cross-section for producing the state X via an interaction of partons a and b. It is computed from the squared matrix element $|\mathcal{M}_{ab\to X}|^2$, averaged over the possible initial spin and colour states. The matrix element is determined by summing over the Feynman diagrams for the relevant processes. The renormalisation scale μ_R is introduced to handle ultraviolet divergences in the matrix element calculations, incorporating them into the coupling constant.

The large phase space at the LHC leads to the creation of $\mathcal{O}(100) - \mathcal{O}(1000)$ particles with momenta spanning several orders of magnitude [109]. Consequently, the phase space integral is performed using MC techniques, implemented in dedicated event generators such as MadGraph5_aMC@NLO [110] and Powheg Box [111–113]. These generators

compute the matrix elements of hard-scatter processes and sample the phase space to generate representative events.

2.3.2 Parton Shower

Equation 2.10 specifies the total inclusive cross-section for the production of the final state X, without detailing the subsequent processes. This implies that X can be accompanied by any number of additional particles and can have any kinematic configuration.

To derive exclusive cross-sections, one would need to constrain the phase space for X or specify the number of accompanying particles. However, event generators initially generate an inclusive final state containing only X. Additional particles are then incorporated through four-momentum-conserving Markov processes, resulting in a high-multiplicity final state while maintaining the inclusive nature of the initial production [109].

Parton showering, the phenomenon of high-momentum incoming and outgoing partons emitting gluons and quarks, which can further emit additional particles, resulting in cascades of partons, is modelled through such a Markov process. Emissions occurring before the hard-scatter are termed ISR, while those after are known as final state radiation (FSR).

Parton showers transition the momentum scale from the perturbative regime of the hard-scatter down to the non-perturbative regime below around 1 GeV. This evolution is described by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations, which describe PDFs at varying momentum scales [114–116].

However, several approximations are used, such as the emission of soft and collinear gluons. Additionally, cross-section calculations of the parton shower are typically truncated at LO or NLO.

In summary, parton shower generators simulate the emission of cascades of partons from the hard-scatter down to a scale where non-perturbative effects dominate and hadronisation becomes important.

2.3.3 Hadronisation

Hadronisation, or fragmentation, is the process by which outgoing partons transition into colourless states called hadrons. This transformation occurs after partons lose sufficient energy through parton showering, and thus at non-perturbative scales. Consequently, the simulation of hadronisation processes relies on phenomenological models. Two widely employed models are the Lund string model [117, 118], used in Pythia, and the cluster model [119], used in Herwig [120, 121].

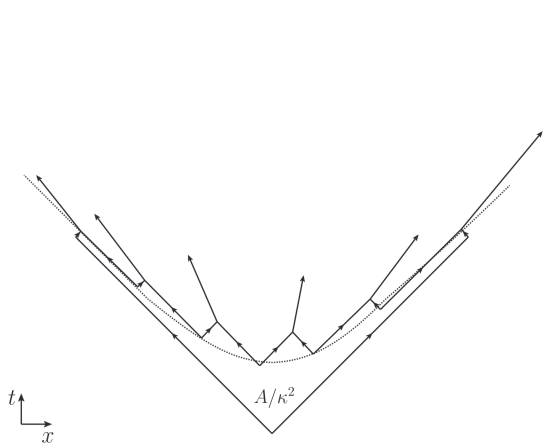
The Lund string model describes hadronisation by representing the strong QCD field between partons as a string under tension. As quarks move apart, the energy in the string increases until it becomes energetically favourable for new quark-antiquark pairs to form from the vacuum, breaking the string. This process continues until only colourless stable hadrons remain. Figure 2.13(a) illustrates hadronisation within the Lund string model.

The cluster hadronisation model leverages the pre-confinement property of QCD. At energy scales much lower than the hard subprocess scale, partons in a shower group into colourless clusters with an invariant mass distribution independent of the subprocess' nature and scale. These clusters, formed at the hadronisation scale, act as proto-hadrons that decay into the final state hadrons [122]. See Figure 2.13(b) for an illustration of the cluster model.

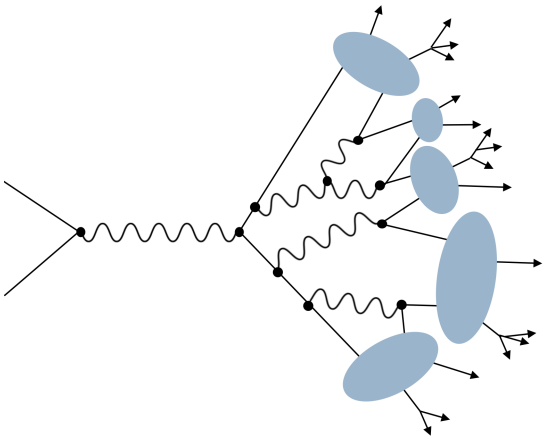
2.3.4 The Underlying Event

In events containing a hard-scatter, extra hadron production is observed that cannot be ascribed to showering from the coloured partons associated to the hard-scatter process and its subprocesses. This extra activity, called the underlying event, is greater than in minimum-bias events, where no hard subprocess is present. It arises from interactions between partons in the colliding hadrons that are not involved in the hard subprocess.

The dominant hard subprocess in proton-proton collisions is gluon-gluon scattering ($gg \to gg$). As the LO cross-section of this process diverges at low momentum transfer, the probability for MPI is relatively high.



(a) A quark-antiquark string breaking into hadrons.. The original pair is moving outwards along light-like trajectories. New qq pairs are produced around a hyperbola in (x, t), and combine into hadrons. The figure and caption are taken from Ref. [106].



(b) In the cluster model, partons form colourless clusters, which can be identified as proto-hadrons (depicted in blue). The figure is inspired by Ref. [122].

Figure 2.13: Illustration of two widely used phenomenological hadronisation models: The Lund string model (a) and the cluster model (b).

To model the underlying event, event generators account for the impact parameter, the transverse distance between colliding hadrons. Smaller impact parameters lead to central collisions with higher overlap, more MPIs, and increased underlying event activity, while larger impact parameters result in peripheral collisions with fewer interactions. [122]

2.3.5 Hadron Decays

Hadronisation results in many unstable particles. Thus, the final stage of event generation involves their decay into final state particles that are stable on the timescales of detection. Generators rely on experimental data and particle properties provided by the Particle Data Group (PDG) [24]. However, information on some particles is insufficient, making it necessary to rely on models. Theoretical models are particularly needed for many excited mesons and hadrons containing b and c quarks. [108] Modern generators also consider multi-particle decay correlations and spin correlations.

2.3.6 Simulation of the ATLAS Detector

Finally, the interaction of the final state particles with the ATLAS detector is simulated. GEANT4 [123] is used to model particle interactions in media, using a full, detailed detector model. Particle interactions with active layers of the detectors are converted to electrical signals and subsequently digitised, resulting in raw data of the same format as data from measured physical collisions. However, in contrast to physics data, for the simulated data the underlying true physics process is known and detailed information on its interaction with the detector and its reconstruction can be assesed, which is critical to the understanding of calibrations, efficiencies and performance. The next step is to reconstruct physical objects from this raw data, consisting of hits and energy deposits.

2.4 Reconstruction of Physics Objects

Signals from interactions of physical particles with the detector are stored in the form of raw physics data. This data includes signals from various subdetectors, such as hits in the ID and the MS, as well as energy deposits in the calorimeters. This section describes how to reconstruct the original physical states from these signals.

First, low-level objects such as tracks, decay vertices, and energy clusters are reconstructed. These low-level objects are then combined into high-level physics objects and quantities, which can be connected to actual physical states. The reconstructed objects are then used in physics analyses, allowing ATLAS physicists, for example, to perform selections or fits on their multiplicities and kinematic distributions. Reconstructed objects require calibrations to correct their properties to match the truth values on particle level. These calibrations come with uncertainties that must be accounted for in statistical analyses based on the reconstructed objects.

This section focuses on the reconstruction of jets, particularly b-tagged jets, as they are central to the physics analyses detailed in this thesis. Additionally, $E_{\mathrm{T}}^{\mathrm{miss}}$ plays a significant role, being reconstructed from many different physical objects. Although not the primary objects of interest, muons are used in both analyses detailed in this thesis. Therefore, the relevant reconstruction methods, together with those of low-level objects and jets, are discussed. Electrons, however, do not play a significant role except for being vetoed. For detailed information on electron reconstruction, refer to [124].

2.4.1 Low-Level Objects

This section details the reconstruction of low-level objects, which form the building blocks of high-level reconstructed physics object candidates such as particles or jets. Different particles are constructed from different combinations of these low-level objects.

Tracks

Track reconstruction is based on hits in the ID and the MS. The shape of tracks in the magnetic field of the detector allows the inference of the charge sign and momentum of charged particles.

The reconstruction process begins with clusters of hits in the silicon Pixel and SCT trackers, formed by grouping adjacent pixels or strips with signals above a threshold. These clusters are used to determine three-dimensional space points, or hits. Tracks are initialised by combining three hits and are extended with additional points using Kalman filters. The high number of hits originating from pile-up interactions makes correct track identification a challenging task. Ambiguities arising from multiple tracks being associated with the same hits are resolved by employing a neural network, prioritising tracks with high quality, meaning high p_{T}, few holes, a large number of associated hits, and low χ^2. Tracks must meet pre-defined quality criteria to be considered. A high-resolution χ^2 fit is then applied, and the track trajectory is extrapolated to the TRT. For further details on track reconstruction during Run 2, refer to Ref. [125].

Muon tracks are reconstructed by combining straight-line segments from different stations into initial track candidates. A χ^2 fit refines the muon track, considering detector interactions and chamber misalignments, updating included hits, and removing low-quality tracks. Ambiguities are resolved by discarding tracks that share multiple hits with higher-quality ones. Subsequently, a final fit, which also considers energy losses in the calorimeters, is performed. See Ref. [126] for details on muon reconstruction during Run 2.

Vertices

The Primary Vertex (PV) reconstruction is based on ID tracks with $p_{\mathrm{T}} > 500$ MeV. Initially, a vertex seed is defined based on the beam spot. Subsequently, an iterative process, consisting of vertex fitting and weighting associated tracks based on their compatibility, is utilised [127]. Lastly, incompatible tracks are removed and used for the reconstruction of additional vertices. The hard-scatter PV is identified as the vertex with the largest sum of squared transverse momentum ($\sum p_T^2$) of associated tracks.

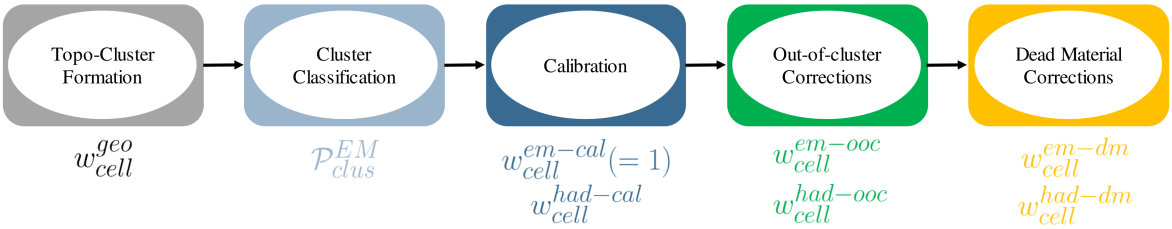


Figure 2.14: Overview of the local hadronic cell weighting (LCW) calibration scheme for topo-clusters. The figure is adapted from Ref. [128].

Energy Clusters

The fundamental low-level objects used for calorimeter-based measurements are topo-clusters. Their reconstruction and calibration are detailed in Ref. [128]. Topo-clusters represent three-dimensional topologically connected energy deposits resulting from showers of incoming particles in the active calorimeter volume. Depending on the incoming particle type, topo-clusters can contain the entire or a fractional response to a single particle, or the merged response of several particles.

Topo-clusters are reconstructed through a topological growing-volume clustering algorithm. This algorithm begins by identifying seed cells with high signal significance, defined as the ratio of the signal to the expected noise. Surrounding cells are then added if they meet certain significance criteria. The process is iterative, expanding the cluster by incorporating neighbouring cells until no neighbours with signals above the threshold are found. The resulting proto-clusters can be very large and are split if they include multiple spatially separated signal maxima. The kinematics of clusters are then determined by recombining cell energies and directions.

The local hadronic cell weighting (LCW) calibration scheme aims to provide a calibrated calorimeter signal for physics object reconstruction in ATLAS, irrespective of the object type. This is crucial for accurately reconstructing hadronic jets and for the precise determination of $E_{\mathrm{T}}^{\mathrm{miss}}$. Calibration of topo-clusters involves multiple steps to correct for various detector effects such as non-compensating calorimeter responses, signal losses due to clustering strategies, and energy losses in inactive material. Figure 2.14 provides a schematic overview of the topo-cluster calibration. The calibration is performed locally, at the level of individual cells. Separate corrections are derived for electromagnetic and

hadronic clusters. All calibrations and corrections are derived from MC simulations of charged and neutral pions, for hadronic and electromagnetic clusters respectively, across various energies and directions. First, clusters are assigned a probability $\mathcal{P}_{clus}^{EM}$ of being of electromagnetic nature, a process called cluster classification. To correct for the non-compensating nature of the signal, differences in the cell response between electrons and hadrons are evaluated and corresponding weights are applied. These corrections are only applied to the hadronic part of the energy deposit and thus scaled by $\mathcal{P}_{clus}^{EM}$ for each cell. The cell weights applied are thus computed by

$$w_{\text{cell}}^{\text{cal}} = P_{\text{clus}}^{\text{EM}} \cdot w_{\text{cell}}^{\text{em-cal}} + (1 - P_{\text{clus}}^{\text{EM}}) \cdot w_{\text{cell}}^{\text{had-cal}}. \tag{2.11}$$

Additionally, out-of-cluster corrections are applied in the form of weights $w_{\text{cell}}^{\text{em-ooc}}$ to incorporate energy deposits in cells discarded by the clustering algorithm due to low signal significances. Finally, dead material corrections are applied to account for energy deposits outside the active material.

2.4.2 Jets

Due to the phenomena of parton showering and subsequent hadronisation, quarks and gluons produced at the LHC are only observed indirectly in the form of jets. Jets are collimated, cone-shaped sprays of hadrons and other particles, resulting in regions of high hit density in the ID and large energy deposits in the calorimeters. The reconstruction of jets involves multiple steps. First, potential jet constituents must be correctly grouped into a jet, referred to as clustering. Subsequently, the jets must be calibrated to ensure that the reconstructed properties match the ones at particle level. The goal is to assign properties, such as momentum and direction, of the initial parton to the jet.

Reconstruction

While jets in Run 1 were reconstructed using information from single subdetectors, Run 2 introduced Particle Flow Objects (PFOs), which combine information from the ID and

the calorimeters, resulting in improved energy resolution [129]. Jet constituents, such as calorimeter clusters, tracks, and PFOs, are used as input to jet clustering algorithms. The most commonly used clustering algorithm in ATLAS is the anti-k_t algorithm [130]. It allows the creation of different types of jets, built from various objects and of different sizes, such as small-R jets, large-R jets, or jets of variable size. After a jet is formed, additional objects can be associated with it through methods based on their proximity to the jet or through ghost-association [131–133]. The following sections introduce these concepts in more detail.

Particle Flow Algorithm To reconstruct PFOs, high-quality ID tracks passing certain criteria are selected. Each of these tracks is matched to a single topo-cluster and the expected energy deposit of the particle associated with the track is computed. If energy deposits in other topo-clusters are likely to come from the same particle, these topo-clusters are added to the track/topo-cluster system. This strategy allows for the recovery of the full shower energy. Next, the expected energy of the particle associated with the track is subtracted on a calorimeter cell level in each of the topo-clusters in the system. If the remaining energy in the system is consistent with expected fluctuations, the entire topo-cluster system is removed from the algorithm.

This procedure is applied to all selected tracks, starting with the highest p_{T} one. Priority is given to tracks with only a single associated topo-cluster. After looping over all selected tracks, the resulting track/topo-cluster systems are the PFOs used as building blocks for jet clustering. Details on the algorithm are provided in Ref. [129].

The use of PFOs in jet reconstruction leads to significant improvements in energy resolution. It combines the advantages of the high momentum resolution of the ID for low-p_{T} objects and the high performance of the calorimeters for high-p_{T} particles. Additionally, using the calorimeter signals allows for incorporating contributions from electrically neutral particles.

The Anti-k_t Algorithm The anti-k_t algorithm is an infrared (IR) and collinear safe jet clustering algorithm. An algorithm is IR safe if the addition of soft radiation doesn't impact

the resulting reconstructed jet. Collinear safety refers to the algorithm's insensitivity to the collinear splitting of partons. The anti-k_t algorithm fulfils both these requirements while also resulting in well-defined jet boundaries. Contrary to other popular clustering algorithms such as the k_t algorithm [134] or the Cambridge/Aachen algorithm [135], the anti-k_t algorithm tends to produce more regular, circular shaped jets. This makes the resulting jets easier to work with and allows for better comparison of results. The anti-k_t algorithm clusters the hardest particles first, making the resulting jets less sensitive to soft particles, thus more stable and less affected by underlying events or pile-up. All kinds of objects can be used as inputs to the algorithm, as long as they have a p_{T} and a detector position vector associated with them. PFOs, but also calorimeter clusters or tracks can be used as inputs to the clustering algorithm. Furthermore, even entire jets can be re-clustered to form larger jets.

The clustering algorithm is sequential, clustering the hardest particles (or other objects) first. The first step is to compute the distances d_{ij} between all pairs of particles i and j, defined as

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{R_{ij}^2}{R^2} \ . \tag{2.12}$$

Here, R_{ij} are the Lorentz-invariant angular separations between the two objects, and R is a parameter defining the size of the jet. Additionally, the quantity $d_{iB} = p_{Ti}^{2p}$ is computed, where i is the index of the harder of the two particles. The sequential algorithm starts with the two inputs resulting in the smallest d_{ij}. The two input particles i and j are clustered into a single pseudo-jet with a single index and a combined four-vector if d_{ij} is smaller than d_{iB}. Otherwise, the particle or pseudo-jet i is recognised as a jet, removed from the set of inputs, and the algorithm starts again with the remaining inputs. The procedure is repeated until no entities remain.

The difference between the sequential algorithms is the value of the parameter p. While Cambridge/Aachen uses $p = 0$ and the k_t algorithm uses $p = 1$, the anti-k_t algorithm uses $p = -1$. In contrast to the k_t algorithm, this choice of the parameter leads to the hardest components being clustered first.

Types of Jets ATLAS analyses make use of different types of jets. These are defined based on different values of the radius parameter R in the clustering algorithm, as well as the types of inputs fed to the algorithm. One of the most widely used type of jets are fixed-radius jets with a parameter $R = 0.4$. The calibration described in Part IV utilises $R = 0.4$ jets, which are reconstructed from topo-clusters. Additionally, large-R jets with $R = 1.0$ can be used to reconstruct the collimated system of decay products from boosted heavy particle decays. These jets are also based on topo-clusters. Variable radius (VR) jets can also be defined, where the size of the jet depends on its boost. The DM search presented in Part III uses fixed-radius small-R jets based on PFOs. Further details are discussed in the relevant chapters.

Associating Objects with Jets Physics objects can be associated with a jet based on two different strategies. Geometric association simply associates objects within a predefined angular distance ΔR with the jet.

Ghost association relies on the IR safety of the jet clustering algorithm. A ghost of an object is defined as an infinitely soft particle with the same direction of movement (ϕ and η) as the object itself. The ghost is added to the inputs to the clustering algorithm. If it is clustered with the jet, the object is ghost-associated with it. Thus, ghost association can more accurately determine whether an object is part of a jet, even if the jet has an irregular shape, unlike the geometric association based on a fixed cone size. The IR safety ensures that the other jet constituents remain unaffected.

Calibration of Jets

The calibration of jets involves two main components: the jet energy scale (JES) calibration and the jet energy resolution (JER) corrections. The JES calibration aims to correct the four-momenta of reconstructed jets to match the true particle-level energies in simulations. This procedure involves multiple corrections, as depicted in Figure 2.15. On the other hand, the JER smears the simulated jets to ensure their resolution matches that of jets in the observed data. This section introduces the different stages of jet calibration based on Ref. [136].

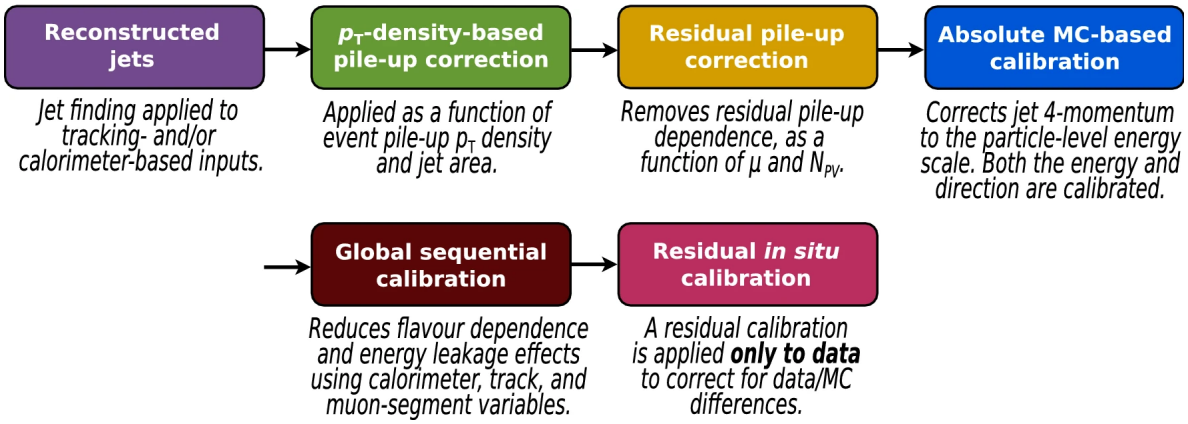


Figure 2.15: Flowchart of the stages of the JES calibration. Each stage is applied to the four-momentum of the jet. The figure and caption are taken and adapted from Ref. [136].

Jet Energy Scale Calibration The first two steps in the JES calibration apply pile-up corrections. First, the average pile-up p_{T} density ρ in the jet direction is estimated, and the jet p_{T} is corrected based on this estimate and the jet area A. The correction is given by $p_T^{\mathrm{corri}} = p_T^{\mathrm{reco}} - \rho \times A$. Due to systematic uncertainties in the derivation of ρ, a residual dependency on pile-up remains after this correction. Therefore, an additional residual correction, derived from truth information in MC simulations, is applied based on the number of primary vertices and the mean number of interactions per bunch crossing μ.

The next stage involves the absolute JES and η calibration, which corrects the reconstructed jet four-momentum for losses in passive material, out-of-cone effects, and biases in the jet η reconstruction [136]. These corrections are derived from $R = 0.4$ anti-k_t jets in PYTHIA MC dijet events by calculating the jet energy response $\mathcal{R}$, the mean of the $\frac{E^{reco}}{E^{true}}$ distribution. The corrections are applied in bins of p_{T} and η of the truth jets.

After the absolute calibration, the response still depends on the flavour and energy distribution of the jet constituents. The composition of a jet, and thus its shower shape, varies depending on the initiating parton. Quark-induced jets tend to contain fewer hadrons with high p_{T}, while gluon-induced jets typically exhibit a larger number of soft particles. Thus, gluon jets result in wider showering, while jets from quarks penetrate deeper into the calorimeter. A series of multiplicative corrections to different observables is applied in bins of p_{T} and η. These corrections are derived by the response $\frac{p_T^{reco}}{p_T^{true}}$ of PYTHIA MC dijet simulations. The six observables include fractions of energy deposited in different

calorimeter layers and information on the composition, such as track multiplicities and the fraction of charged particles.

The final stage of the JES calibration is the in situ calibration, applied only to data to account for differences in jet response between data and simulation. These deviations arise from imperfect physics and detector simulations. The jet response is computed by balancing the p_{T} of a jet against that of a well-calibrated reference object, applying the following correction factor to data:

$$c = \frac{\mathcal{R}_{\mathrm{data}}}{\mathcal{R}_{\mathrm{MC}}} = \frac{\langle \frac{p_T^{probe}}{p_T^{ref}} \rangle_{\mathrm{data}}}{\langle \frac{p_T^{probe}}{p_T^{ref}} \rangle_{\mathrm{MC}}} \ . \tag{2.13}$$

The in situ calibration consists of three independent stages, applied sequentially. The first stage, the η-intercalibration, updates the energy scale of forward jets to match that of central jets using the p_{T} balance in dijet events. The second stage balances the hadronic recoil against the p_{T} of a calibrated Z boson or photon. The third and final stage uses the multi-jet balance analysis to balance a single high-p_{T} jet against a system of well-calibrated low-p_{T} jets.

Jet Energy Resolution The relative JER on the p_{T} of a jet is a functional of the form

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C \ , \tag{2.14}$$

with a noise term N, a stochastic term S, and a constant term C. N includes contributions from electronic noise and pile-up, independent of the energy deposited, and is dominant in the low-p_{T} region. For objects with p_{T} below several hundred GeV, the stochastic term is relevant. C incorporates fluctuations from energy deposits in passive material and non-uniformities in the response, and it is dominant in the high-p_{T} region. The total resolution is extracted from measurements of the asymmetry in p_{T} of dijet events, which are abundant and theoretically well understood.

Additionally, the JER estimate can be refined by directly estimating the noise term using data collected with random unbiased triggers. In the random cone method, energy

deposits in two $\Delta R = 0.4$ sized areas at random ϕ and opposite η are summed, and their difference serves as a measure of random fluctuations in the deposited energy.

To match the resolution between data and simulation, a smearing procedure is applied. In p_{T} regions where the data resolution is larger, the simulation is smeared until the average resolution matches. In regions where no smearing is applied to avoid altering the observed data, a systematic uncertainty is defined as $\sigma = \sigma_{\mathrm{data}} - \sigma_{\mathrm{MC}}$ [136].

Systematic Uncertainties The complete JES calibration introduces 125 systematic uncertainties. These include effects from pile-up, dependencies on flavour compositions, and uncertainties from MC generators. However, the majority of uncertainties originate from the three analyses performed for the in situ calibration. These include the effects of analysis selection cuts, reconstruction uncertainties in other objects used, MC mis-modelling, and statistical limitations.

Additionally, the JER corrections introduce uncertainties, which are propagated through analyses as Gaussian smearing of the simulation.

Given the correlations among many uncertainties, most analyses can utilise reduced sets of systematic uncertainties. These are derived through eigenvalue decomposition of the covariance matrix. Depending on the analysis, the JES uncertainties can be reduced to around 30 [136]. The final set of systematic uncertainties is propagated through analyses in the form of nuisance parameters.

Flavour Tagging

ATLAS uses dedicated algorithms to identify jets initiated by b-quarks and c-quarks. This section focuses on the tagging of b-jets, which are the primary physics objects in the analyses described within this thesis. These taggers exploit several characteristic features of b-jets, which distinguish them from light jets (initiated by up, down and strange quarks).

These distinguishing features include the relatively long lifetime, high mass, and decay multiplicity of b-hadrons compared to other hadrons. b-hadrons have a lifetime of about 1.5 ps, resulting in significant flight lengths, typically several mm. This leads to displaced

decay vertices.

The field of b-tagging algorithms has advanced rapidly over the last few years, with many modern algorithms based on the latest advancements in Graph Neural Network (GNN). However, during Run 2, the baseline b-tagging algorithm employed by ATLAS was the DL1r tagger, based on a feed-forward Deep Neural Network (DNN). It classifies jets into three categories: b-jets, c-jets, and light jets, assigning three output values p_b, p_c, and p_l, which sum to unity and can be interpreted as probabilities. These probabilities are combined to form one output value, the DL1r score:

$$D_{DL1r} = \log\Big(\frac{p_b}{f_c p_c - (1 - f_c) p_l}\Big) \ , \tag{2.15}$$

where f_c is the expected fraction of c-jets. Different working points (WPs) balance tagging efficiency and background rejection. Jets with a DL1r score above the threshold of a certain WP are identified as b-jets. The WPs are defined based on the algorithm's performance on jets in $t\bar{t}$ samples [137]. For a comprehensive overview of flavour tagging algorithms used in Run 2, refer to Ref. [138]

The DL1r tagger is trained on simulated data, with labels derived from truth hadrons geometrically matched to jets in their vicinity. In addition to properties of the jet itself, such as p_{T} and η, it uses quantities computed by low-level taggers. These low-level taggers identify specific characteristics distinguishing heavy flavour jets, particularly b-jets, from others.

The IP2D and IP3D are impact parameter-based algorithms [139]. IP2D uses the signed transverse impact parameter significance s_{d_0} of tracks to construct discriminating variables, while IP3D also utilises the track longitudinal impact parameters z_0 and their correlation. These algorithms work on a track-by-track basis, and their results are combined to produce three discriminants for each algorithm, which are then used as inputs to the DL1r tagger.

The Single Secondary Vertex Finder (SSVF) algorithm reconstructs a single displaced vertex within a jet [140]. It starts by building two-track vertices from tracks associated with the jet. Two-track vertices associated with long-lived particles such as K_s, photon conversions, or hadronic interactions with detector material are removed [137]. The vertex

fitter then combines all remaining tracks into a single vertex, iteratively removing the track with the highest χ^2 until a χ^2 below a predefined threshold is found. If the vertex has an invariant mass below 6 GeV, it is accepted. The SV1 algorithm uses the SVs provided by the SSVF algorithm to construct discriminating variables. Eight SV-based variables, including the number of associated tracks and the SV mass, are used as inputs to the DL1r tagger.

The JetFitter algorithm [141] aims to reconstruct the full b-hadron decay chain, exploiting the topological structure of $b \to c \to s$ decays inside the jet. Kalman Filters [142] determine the line on which the primary vertex and the two secondary vertices from the b-hadron and c-hadron are found. Eight flavour discriminating variables, such as the invariant masses of the SVs and the number of tracks associated with them, are constructed and used as inputs to the DL1r tagger.

The last of the low-level taggers feeding information to the DL1r tagger is the RNNIP algorithm [143], based on Recurrent Neural Networks (RNNs) (see Section 4.2.4). While this network uses similar information to the IP2D and IP3D taggers, it is applied to all tracks associated with the jet that fulfil certain criteria. This allows it to utilise correlations between different track impact parameters, resulting in increased performance. The RNNIP output score was added to the DL1 tagger at a later stage, resulting in the baseline DL1r tagger used in most Run 2 legacy analyses.

Figure 2.16 shows the background rejection perfromance of the low-level taggers and the DL1r tagger.

To account for potential differences in the performance of b-tagging algorithms on data and simulation, dedicated calibrations derive scale factors (SFs), which are applied to MC. The SFs are given by the ratio of the efficiencies $\frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$. These calibrations are performed separately for light [144], c [145], and b-jets [137]. The mis-tagging calibration of light jets is based on $Z+$jets events, whereas the c-jet and b-jet calibrations target semi-leptonic $t\bar{t}$ decays.

Many BSM searches target heavy resonances or boosted Higgs bosons that can decay to b-quarks. In these boosted topologies, b-tagging is particularly challenging, necessitating

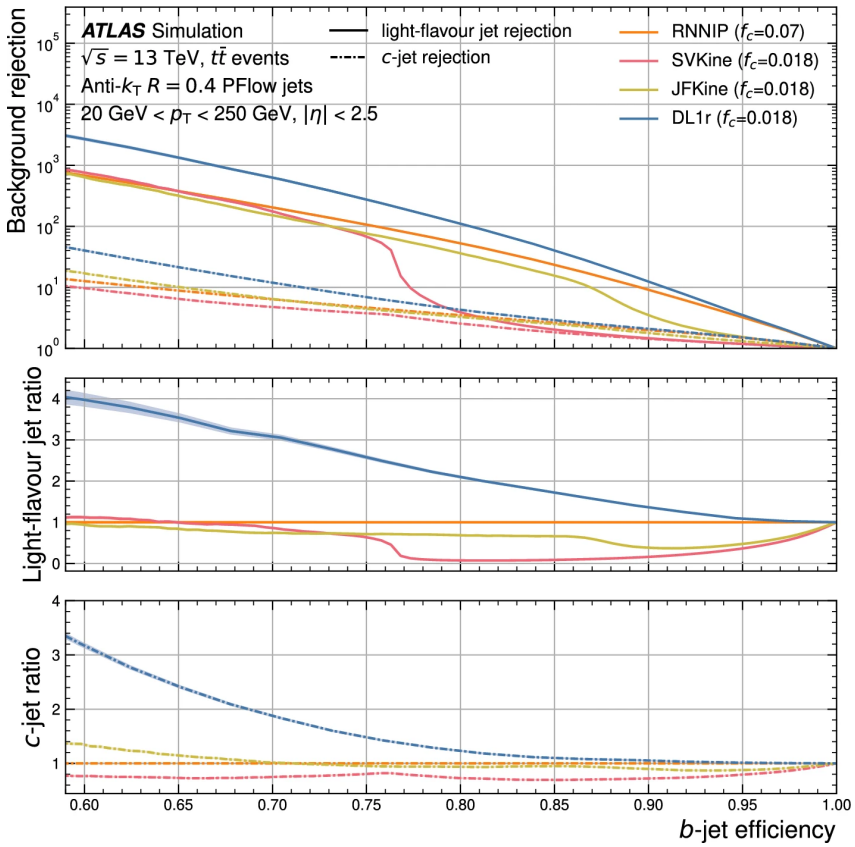


Figure 2.16: The light-flavour jet and c-jet rejection factors as a function of ϵ_b for DL1r and the low-level b-taggers RNNIP, SVKine, and JFKine. The latter two taggers rely only on information provided by the SV1 and JetFitter algorithms. SV1 and JetFitter have secondary-vertex-finding efficiencies of approximately 80% and 90%, respectively; this causes the rapid growth in light-flavour jet rejection around these values for SVKine and JFKine. The lower two panels show the ratio of the light-flavoured and the c-flavoured jet rejection of all other algorithms to the ones of RNNIP. The figure and caption are adapted from Ref. [138].

the development and calibration of a dedicated tagger, described in detail in Part IV.

2.4.3 Muons

Muon reconstruction primarily relies on signals from the ID and MS. Additionally, calorimeter information is used to associate ID tracks with muons independently of the MS and to account for cases of large energy losses in the calorimeters [126].

Muons can be reconstructed either from MS stand-alone tracks or based on complete detector information. The first method is detailed in Section 2.4.1. Global muon reconstruction, the second method, utilises signals from the ID, MS, and calorimeters. Five separate strategies are used to reconstruct five types of muons: combined (CB),

inside-out combined (IO), muon-spectrometer extrapolated (ME), segment-tagged (ST), and calorimeter-tagged (CT) muons.

CB muons are identified by a combined fit of MS and ID hits, taking into account the energy loss in the calorimeters. The fits may be repeated with an updated list of MS hits. IO muons use ID tracks extrapolated to the MS, where at least three compatible hits are required. This method does not require independently reconstructed MS tracks, enhancing efficiency in certain detector areas or for low-p_{T} muons that do not traverse the entire MS. If no ID track is compatible with a reconstructed MS track, the track parameters are used to define an ME muon. Similarly, ST muons are identified by matching ID tracks to incomplete MS tracks, consisting of hits in only one MS segment. CT muons are identified by extrapolating ID tracks to compatible energy deposits in the calorimeters. These energy deposits can be used to tag the ID tracks as muons, and the ID track parameters are used.

After reconstruction, high-quality muon candidates are selected based on criteria such as the number of hits in various subdetectors and goodness-of-fit properties. Specific requirements for each muon type, known as selection WPs, are tailored to the needs of different physics analyses. The WPs balance the efficiency of prompt-muon identification, momentum resolution, and background rejection. The background stems from fake and non-prompt muons. The latter are further categorised into muons from light hadron decays, which result in lower-quality tracks, and those from heavy-flavour hadron decays, which produce better-quality tracks. Prompt muons are associated with the primary vertex and usually satisfy tight isolation criteria. For further details on muon reconstruction and identification, refer to Ref. [126].

2.4.4 Missing Transverse Energy

$E_{\mathrm{T}}^{\mathrm{miss}}$ is a measure of the momentum imbalance in the transverse plane. Accurate reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$ is crucial for many BSM searches, as many hypothetical particles, such as DM, are expected to traverse the detector without interacting with it. Therefore, $E_{\mathrm{T}}^{\mathrm{miss}}$ can be identified as an estimate for the p_{T} of a DM particle produced in the hard-

scatter. In the SM, the only genuine source of $E_{\mathrm{T}}^{\mathrm{miss}}$ is neutrinos. Its reconstruction is challenging because it involves all detector components, various types of physics objects and is very sensitive to pile-up. The following introduction to $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is based on Ref. [146].

$E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is based on two components: the hard-event and the soft-event. The hard-event includes all reconstructed and calibrated objects such as jets, electrons, muons, photons, and tau leptons. The soft-event consists of tracks associated with the hard-scatter but not linked to any physics object.

$E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed as the negative vector sum of the p_{T} of all hard objects, passing certain selection criteria, and the soft term:

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = -\underbrace{\sum_{\text{electrons}} \mathbf{p}_{\mathrm{T}}^{\mathrm{e}} - \sum_{\text{photons}} \mathbf{p}_{\mathrm{T}}^{\gamma} - \sum_{\text{taus}} \mathbf{p}_{\mathrm{T}}^{\tau} - \sum_{\text{muons}} \mathbf{p}_{\mathrm{T}}^{\mu} - \sum_{\text{jets}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\text{hard term}} - \underbrace{\sum_{\text{unused tracks}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}}}_{\text{soft term}} . \tag{2.16}$$

The notation $E_{\mathrm{T}}^{\mathrm{miss}}$ is used to refer to both the vector quantity $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = (E_{\mathrm{x}}^{\mathrm{miss}},\ E_{\mathrm{y}}^{\mathrm{miss}})$ and the scalar quantity $E_{\mathrm{T}}^{\mathrm{miss}} = |E_{\mathrm{T}}^{\mathrm{miss}}| = \sqrt{(E_{\mathrm{x}}^{\mathrm{miss}})^2 + (E_{\mathrm{y}}^{\mathrm{miss}})^2}$. As not all objects might be reconstructed, and the p_{T} of hard objects is reconstructed with limited resolution, an observation bias towards non-zero values is introduced, even in events with no genuine $E_{\mathrm{T}}^{\mathrm{miss}}$.

Ambiguities can arise because different types of physics objects are reconstructed separately, potentially leading to double-counting of signals. For example, an energy deposit in the calorimeter could be counted twice, once for a jet and once for an electron. To prevent this issue in $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, objects are incorporated sequentially, and objects reconstructed from signals that have already entered are rejected. First, electrons are included, followed by photons, tau leptons, and jets. Muons usually have little overlap, as they are primarily reconstructed from tracks in the MS and ID. For details on more refined overlap removal procedures, refer to Ref. [146].

Systematic uncertainties are provided for both the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution. These uncertainties depend on both the magnitude of the soft-term and the composition of the hard-term, which varies with the selections applied in a given analysis. The uncertainties

for the hard-term stem from the contributing calibrated objects and are assumed to be uncorrelated. The systematic uncertainties associated with the soft term are derived from data-MC comparisons of the parallel and orthogonal projections of the soft term onto the hard term. These uncertainties are derived from $Z \to \mu\mu$ events without additional jets, where the soft term is expected to point in the opposite direction of the hard term. The average ratio of the longitudinal projection of the soft term onto the hard term in data and simulation is used to determine the scale uncertainty. Fluctuations in the transverse projection are used to determine the resolution uncertainty. Figure 2.17 shows the individual components used to derive these uncertainties. Differences in response between

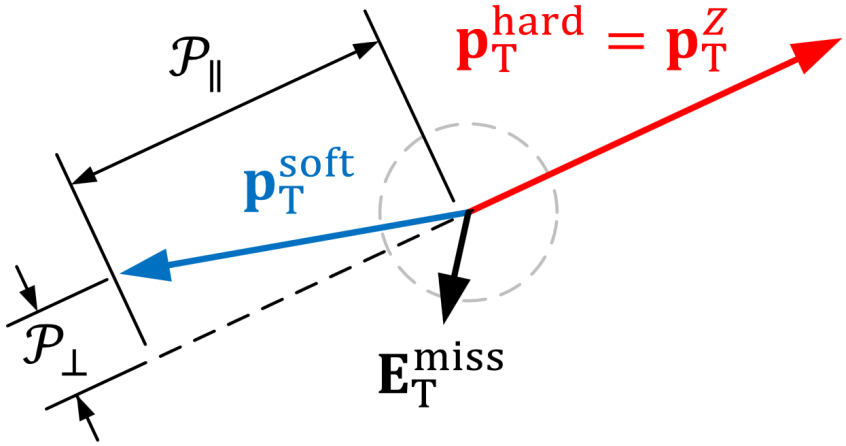


Figure 2.17: Schematic view of the parallel and perpendicular projections of the soft and hard term in $Z \to \mu\mu$ events without genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ for a final state without additional jets. The figure and caption are taken from Ref. [146].

data and simulation are corrected by applying a scaling correction to the longitudinal component in simulation. Offsets in resolution are addressed by applying Gaussian smearing to the longitudinal and transverse components of the soft-term.

Part II

Statistics and Machine Learning

Chapter 3

Statistical Methods

Contents

3.1 The Likelihood Function 94

3.2 Hypothesis Testing 98

3.3 The Asimov Dataset 103

This chapter introduces the statistical methods used to analyse and interpret the results of analyses in High Energy Physics (HEP). It follows the description of these methods in the Refs. [147–150].

Introduction

Typically, HEP analyses first define empirical event selections using multiple observables, such as transverse momenta, invariant masses, angular relationships or flavour information, such as whether a jet is of a certain flavour (e.g., b-tagged). Subsequently, statistical tests, such as hypothesis testing and parameter estimation, are performed on this dataset of reduced size and dimensionality. This chapter explores essential statistical methods, including likelihood fits, hypothesis testing, and the computation of exclusion limits. These tools provide the necessary framework for analysing and interpreting complex collision data.

3.1 The Likelihood Function

The likelihood is a fundamental concept in particle physics analyses. Likelihood fits enable the precise determination of parameters describing the data, such as process normalisations, nuisance parameters, or signal strength. These fits are also used to extract scale factors in calibrations, as detailed in Part IV. Moreover, likelihood ratios serve as test statistics in hypothesis testing to claim discoveries or determine upper limits on parameters of BSM models.

3.1.1 Definition

The observed outcome of a HEP experiment is represented by a set of random variables $\mathbf{X} = (X_1, X_2, X_3, \ldots, X_n)$. For instance, X_i might represent the number of observed events in different bins of a kinematic distribution in an analysis region. The randomness arises from sources in both the measurement itself, such as signal noise, detector resolution, or efficiencies, and from the underlying theoretical model. The measurement and theory are characterised by a set of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \ldots, \theta_m)$ with unknown values.

Suppose the observed values are $X_1 = x_1$, $X_2 = x_2$, $\ldots$, $X_n = x_n$, with $\mathbf{x} =$

$(x_1, x_2, x_3, \ldots, x_n)$. The likelihood function is defined as

$$L(\boldsymbol{\theta}; \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) \ . \tag{3.1}$$

Here, $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$ denotes the joint probability distribution functions (ProbDFs) of the observed data, given the parameter values $\boldsymbol{\theta}$. The likelihood can be intuitively interpreted as a measure of plausibility of $\mathbf{x}$ given a specific set of parameter values $\boldsymbol{\theta}$, represented as the function $\boldsymbol{\theta} \to f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$. It is important to note that this is not the same as the probability of $\boldsymbol{\theta}$ being the true values given an observation $\mathbf{x}$. Although the notation resembles that of conditional probabilities, it is crucial to recognise that $\boldsymbol{\theta}$ are not random variables but unknown quantities with fixed true values.

3.1.2 The Binned Likelihood Function

In many analyses, such as those detailed in Part III and Part IV, the data are represented as binned histograms. In event-counting analyses, the bin contents follow Poisson distributions and are independent of each other. Thus, the likelihood L is defined as a product of the Poisson ProbDFs $\mathcal{P}$ of the event counts in the bins of each analysis region or category. The histogram is typically a superposition of multiple templates, which are histograms of event yields from individual subprocesses. To account for systematic uncertainties, additional ProbDFs $\mathcal{C}_{\text{syst}}$ are included. In a BSM search with multiple single-binned Control Regions (CRs) and a multi-binned Signal Region (SR), the combined likelihood [151] can be written as:

$$\begin{aligned} L(\mu_{sig}, \boldsymbol{\mu_p}, \boldsymbol{\theta}; \boldsymbol{n}) &= \mathcal{P}_{SR} \times \mathcal{P}_{CRs} \times \mathcal{C}_{syst} && (3.2) \\ &= \prod_{i \in \text{SR bins}} \mathcal{P}(n_i|\lambda_i(\mu_{sig}, \boldsymbol{\mu_p}, \boldsymbol{\theta})) \times && (3.3) \\ &\prod_{i \in \text{CRs}} \mathcal{P}(n_i|\lambda_i(\mu_{sig}, \boldsymbol{\mu_p}, \boldsymbol{\theta})) \times \mathcal{C}_{syst}(\boldsymbol{\theta}) && (3.4) \end{aligned}$$

where the n_i represent the observed data, typically the count of events in bin or region i. The expected number of events λ_i in the ith bin is a function of the unconstrained

parameters, such as the signal strength μ_{sig} and the normalisations of the background processes $\boldsymbol{\mu_p}$, as well as the constrained nuisance parameters $\boldsymbol{\theta}$ associated with systematic uncertainties. These parameters include the nominal MC yields for each background process and the signal being considered.

3.1.3 Systematic Uncertainties

Measurements are affected by various systematic uncertainties, such as the JES uncertainties discussed in Section 2.4.2. To incorporate these uncertainties into the likelihood function, they must be modelled with one or more parameters. Additionally, the distribution of these parameters must be defined, and a response model describing the impact of the parameters on the measurement must be established.

For experimental effects, such as calibrations or detector-related uncertainties, the central limit theorem often applies, allowing the parameter distributions to be assumed Gaussian. For instance, a simple 10% JES uncertainty can be modelled by a parameter whose exact value is unknown but follows a Gaussian distribution with a width of 10%. The response is for example defined by increasing the energy of all jets by 1% if the calibration changes by 1%. The corresponding likelihood for such a measurement, with n observed events, s expected signal events, and b expected background events, where the background estimation is affected by a $\sigma = 10\%$ JES uncertainty, is:

$$L(s, \theta; n, \tilde{\theta}) = \mathcal{P}(n|s + \tilde{b}(\theta, \tilde{\theta}))\mathcal{G}(\theta|\tilde{\theta}, \sigma = 0.1) \tag{3.5}$$

Here, $\tilde{b}(\theta, \tilde{\theta})$ is the response function and $\tilde{\theta}$ is the nominal value of the unknown parameter θ. For example, if $\tilde{b}(\theta, \tilde{\theta}) = \frac{\theta}{\tilde{\theta}}$, a 1% change in JES would increase the background prediction by 1%. The Gaussian constraint term $\mathcal{G}(\theta|\tilde{\theta}, \sigma)$ contains information about the parameter θ from a subsidiary measurement, the calibration analysis.

In practice, this expression is often simplified by renormalising the subsidiary measurement to $\mathcal{G}(\theta|0, 1)$ by scaling the parameter such that values of $\theta = \pm 1$ correspond to the nominal uncertainty, e.g., $\pm 10\%$ in the example above. The response function can

then be expressed as $\tilde{b}(\theta)$, taking the nominal value for $\theta = 0$. The resulting likelihood is referred to as the profile likelihood:

$$L(s, \theta; n, 0) = \mathcal{P}(n|s + \tilde{b}(\theta))\mathcal{G}(\theta|0, 1) \tag{3.6}$$

MC statistical uncertainties are treated similarly but with a Poisson distribution. The treatment of theoretical uncertainties is more subtle, as the distribution of underlying parameters is not always clear. Two-point systematics, resulting for example from comparisons of samples with different shower algorithms such as HERWIG and PYTHIA, can be modelled either with a Gaussian centred at one of the predictions or with a box distribution with Gaussian wings.

Assuming Gaussian distributions for the underlying systematic parameters, the full constraint term can be factorised and modelled by the product:

$$\mathcal{C}_{syst} = \prod_{i \in \text{syst.}} \mathcal{G}(\theta_i|0, 1) \prod_{j \in \text{bins/cats.}} \mathcal{P}_j \tag{3.7}$$

The subsidiary measurements are treated as uncorrelated, but the resulting estimators for the parameter values can show correlations, which can be encoded within a joint response function $b(\boldsymbol{\theta})$.

3.1.4 Maximum Likelihood Fits

The aim of statistical inference is to draw conclusions from data subject to random variations. These methods are particularly useful for obtaining information about the values of unknown parameters, such as their central value and the associated uncertainty. For instance, the calibration analysis in Part IV uses maximum likelihood estimators to determine an efficiency correction scale factor.

The best-fit values for the parameters are ultimately obtained through a simultaneous maximum likelihood fit. However, often the negative log-likelihood is minimised instead, with the minimisation performed using specialised software, such as MINUIT [152]. The parameter of interest can be, for example, the signal strength of a certain BSM model or

an efficiency correction SF for a b-tagging calibration.

3.2 Hypothesis Testing

Hypothesis testing refers to statistical procedures that determine the compatibility of data with different models. To assess the compatibility of data with a specific hypothesis or model, a test statistic is computed. The latter is a functional of the data that summarises the information contained within it.

3.2.1 The Test Statistic: Profile Likelihood Ratio

Given two hypotheses H_b (e.g. only SM physics) and H_{s+b} (e.g. additional Beyond the Standard Model (BSM) processes), the likelihood ratio

$$\lambda(\boldsymbol{n}) = \frac{L(H_{s+b}; \boldsymbol{n})}{L(H_b; \boldsymbol{n})} \tag{3.8}$$

can be used as a test statistic. The Neyman-Pearson lemma shows that this ratio is the uniformly most powerful test statistic in this case [153]. Intuitively, λ takes on large values if the data are compatible with H_{s+b} and smaller values if it is compatible with H_b.

In practice, however,

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}_\mu; \boldsymbol{n})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}}; \boldsymbol{n})} \tag{3.9}$$

is used as test statistic for the computation of upper limits or discovery tests in LHC searches. The concept of constraining the parameters $\boldsymbol{\theta}$ through data by using the best-fit values is called profiling, and thus the test statistic defined above is called the Profile Likelihood Ratio (PLR).

Here, $\hat{\hat{\boldsymbol{\theta}}}_\mu$ is the best-fit value, or maximum likelihood estimator, for a given fixed signal strength μ. The parameters $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the overall best-fit values obtained by fitting both μ and $\boldsymbol{\theta}$ simultaneously. The signal strength μ serves as the Parameter of Interest (PoI).

Although the PLR is represented by a single value for a certain observation or

measurement, it would take different values in the case of repeated experiments. Thus, the PLR has a distribution that depends on the PoI and the other parameters. The ProbDF of the logarithmic test statistic $q_\mu = -2\log(\lambda(\mu))$ is denoted as $f(q_\mu|\mu, \boldsymbol{\theta})$.

Asymptotic Limit

The reason $\lambda(\mu)$ is chosen as a test statistic lies in its convenient asymptotic properties, where the asymptotic limit refers to infinite statistics. However, this limit is approached relatively quickly and can serve as a good approximation for reasonably large statistics of $\gtrsim 10$ events per histogram bin. Wilks showed [154] that $f(q_\mu|\mu, \boldsymbol{\theta})$ is asymptotically distributed as a χ^2 distribution with m degrees of freedom, where m is the number of PoIs. It is important to note that this asymptotic distribution $f(q_\mu|\mu)$ is independent of the nuisance parameters $\boldsymbol{\theta}$. This independence allows for the direct calculation of the p-value (discussed below) from $\lambda(\mu)$ of a single observation.

Toys

In cases of very low event counts, the asymptotic approximation breaks down, and pseudo data, called toys, must be generated. The analysis detailed in Part III, for example, required additional checks with toys.

To compute the p-value, $f(q_\mu|\mu, \boldsymbol{\theta})$ must be approximated by generating a large number of toys using MC methods. First, the ProbDFs for the hypotheses are constructed, and a toy histogram is drawn. Subsequently, the PLR is computed. This procedure is repeated many times to determine $f(q_\mu|\mu, \boldsymbol{\theta})$.

It is common to sample several thousand toys to generate a meaningful distribution that does not suffer from fluctuations. This process is computationally expensive, so in practice, the asymptotic distribution is used whenever possible. In cases of relatively low event statistics, the results of using toys and the asymptotic limit can be compared for a small subset of potential signals to determine if the use of the asymptotic approximation is justified.

3.2.2 Searches for New Physics

If an analysis provides convincing evidence for new physics, it can claim a discovery. Otherwise, the results are still valuable as they can be used to compute upper limits on the magnitude of a potential BSM signal and draw conclusions on what signals are excluded. These upper limits can be provided for the cross-section or specific parameters describing the model. Additionally, this allows for the exclusion of certain regions in the parameter space.

The p-value

To measure the consistency of the observed outcome of an experiment with a hypothesis of a specified signal strength μ, a p-value is computed. For instance, to claim a discovery, one must demonstrate that the observed outcome is inconsistent with the background-only hypothesis $\mu = 0$ (see Section 3.2.2). The test statistic, introduced in the previous section, ranks hypotheses based to their consistency with the observation. The p-value quantifies the probability of observing an outcome at least as extreme as the observed one under a given hypothesis (e.g. the background-only hypothesis), mapping this measure of consistency to a uniform distribution between 0 and 1. It takes large values for observations highly consistent with the tested hypothesis and small values close to zero otherwise.

The p-value for measuring the consistency of the data with a signal strength hypothesis μ is given by:

$$p_{\mu,\boldsymbol{\theta}} = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu,\boldsymbol{\theta})\, dq_\mu \ . \tag{3.10}$$

A hypothesis can be excluded if its p-value is lower than a specified threshold α, called the significance level.

However, this definition of the p-value also depends on $\boldsymbol{\theta}$. A hypothesis could only be rejected if the p-value is smaller than α for all possible $\boldsymbol{\theta}$. In the asymptotic limit, this issue is resolved, as Wilks' theorem shows that the distribution of the test statistic is independent of $\boldsymbol{\theta}$. For realistic scenarios with limited statistics, the supremum of the p-value $p_{\mu,\boldsymbol{\theta}} = \sup_{\boldsymbol{\theta}}$ is well approximated by $\hat{\hat{\boldsymbol{\theta}}}_\mu$. Thus, a definition of the p-value

independent of $\boldsymbol{\theta}$ is given by:

$$p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu, \text{obs}))\, dq_\mu \ . \tag{3.11}$$

Alternatively to the p-value, the results of a discovery test are often quantified through its significance. The p-value can be converted to a significance via $Z = \Phi^{-1}(1-p)$, where Φ^{-1} is the inverse of the cumulative standard Gaussian distribution.

Discovery Test

In the case of a discovery test for potential new physics, the test statistic is defined as:

$$q_0 = -2\log(\lambda(\mu = 0)) \tag{3.12}$$

if $\hat{\mu} \geq 0$, and otherwise $q_0 = 0$, as only positive signal strengths are considered in a BSM hypothesis. The p-value is thus computed by:

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0|0, \hat{\hat{\boldsymbol{\theta}}}(0, \text{obs}))\, dq_0 \ . \tag{3.13}$$

In the asymptotic limit, the significance of a discovery test can be directly approximated by $Z_0 \approx \sqrt{q_0}$ [148]. Typically, a discovery is claimed when $Z \geq 5$, referred to as 5σ. This corresponds to a threshold $\alpha = 2.87 \times 10^{-7}$, which can be interpreted as the probability of such a result or a more extreme one caused by a fluctuation, assuming a background-only hypothesis.

It should be noted that the above significance is the local significance. For example, if one scans over an interval of a certain parameter, such as the mass of a particle in the theory, one must consider that the background could fluctuate upwards at any mass value. This is known as the look-elsewhere effect. Methods to approximate the global significance, which is usually lower than the local significance, are detailed in Ref. [155].

Upper Limits: The CLs Method

For setting limits or excluding a specific value of μ, the PLR

$$q_\mu = -2\log(\lambda(\mu)) \tag{3.14}$$

is utilised for $\hat{\mu} \leq \mu$. If, however, $\hat{\mu} > \mu$, the test statistic is set to $q_\mu = 0$ to ensure that no value of μ is excluded due to an upward fluctuation in the data.

Solving Equation 3.10 for $p = \alpha = 0.05$ allows the extraction of one-sided 95% confidence intervals (upper limits) for the signal strength. This frequentist procedure to construct the confidence interval is technically performed by inverting the Neyman belt [156]. However, downward fluctuations of the observed data below the background prediction can lead to the exclusion of very small signal strengths. This is undesirable as one typically lacks sensitivity to small signal contributions. This method excludes hypotheses that it has no sensitivity to with a probability close to α (5% in the case of a 95% confidence level).

Consider, for example, two experiments looking for the same signal. The first expects one background event, and the second expects five background events. In the case of zero observed events (a downward fluctuation), the second experiment will exclude lower signal strength values than the first. This is counter-intuitive, as the second experiment, despite its worse background rejection, results in lower limits.

To address this issue, the purely frequentist approach is adapted to the modified frequentist CL_s method. For this purpose, the p-value for the background-only hypothesis p_b is introduced as:

$$p_b = \int_{-\infty}^{q_\mu^{\text{obs}}} f(q_\mu|0, \hat{\hat{\boldsymbol{\theta}}}(0, \text{obs}))\, dq_\mu \ . \tag{3.15}$$

Often, p_b is replaced by $CL_b = 1 - p_b$, and p_μ is referred to as CL_{s+b}. Given these redefinitions, the CL_s quantity is defined by:

$$CL_s(\mu) = \frac{CL_{s+b}(\mu)}{CL_b(\mu)} = \frac{p_\mu}{1 - p_b} \ . \tag{3.16}$$

The $1-\alpha$ CL_s upper limit on the signal strength is subsequently computed by:

$$CL_s \leq \alpha \ . \tag{3.17}$$

In the case of an exclusion fit, the threshold p-value is typically set to $\alpha = 0.05$, corresponding to $Z = 1.64$. The resulting signal strength values are thus said to be excluded at a confidence level of 95%.

In the case of a downward fluctuation where both the numerator and denominator take small values, the fraction, and thus the derived p-value, is protected from being very small. Thus, the CL_s method avoids the exclusion of signals to which the measurement is not sensitive. The downside of this method is that one must be careful when interpreting the resulting upper limit. It does not provide the typical 95% coverage in a frequentist way. In fact, it is conservative, meaning it over-covers, and the limit is less stringent than those computed with a purely frequentist approach.

3.3 The Asimov Dataset

The Asimov dataset is an artificially generated dataset with the set of model parameters such that the best-fit estimators of all those parameters resulting from this dataset equals the true values from which the data are generated. In the case of a binned dataset, as the ones discussed within this thesis, the Asimov dataset can be constructed by setting the yields in all bins and analysis regions to their expected values, with the nuisance parameters taking their nominal values. Fits on Asimov data are often used for testing and validation. They allow a detailed investigation of the fit setup and the treatment of nuisance parameters and constraints. Asimov data fits are also used to compute approximate limits of the expected sensitivity of the analysis, while avoiding the computationally expensive generation of many pseudo-experiments. For technical details on the Asimov dataset, see Ref. [157].

Chapter 4

Machine Learning

Contents

4.1 Machine Learning in High Energy Physics 105

4.2 Neural Networks 106

4.3 Overfitting and Regularisation 112

4.4 Shapley Values 115

This chapter provides an introduction to modern Machine Learning (ML) algorithms. Supervised learning models, with a focus on classification tasks, are discussed. Basic ML models like perceptrons and Mulit-Layer Perceptrons (MLPs), along with more advanced algorithms are described. The general reference for Neural Networks (NNs), their training, and corresponding regularisation methods is Ref. [158].

The essential section for understanding the setup and training of the neural network used in the dark-Higgs analysis is Section 4.2. The other sections offer valuable context for the interested reader.

4.1 Machine Learning in High Energy Physics

ML has become an indispensable tool in HEP for analysing the vast datasets generated by the LHC experiments. The tagger described in Part IV utilises deep NNs, the SUSY analysis mentioned in Section 1.3.3 employed Boosted Decision Trees (BDTs), and the DM search detailed in Part III also uses NNs in various forms. Moreover, advanced techniques like GNNs form the base of modern Run 3 flavour tagging algorithms [159], and even unsupervised learning has been applied in first ATLAS physics analyses [160].

In supervised learning, models are trained on labelled data, where each input is associated with a known output. Since observed data lacks labels, one relies on simulated data and the corresponding truth information for the training. Supervised learning has numerous applications, including regression and classification tasks. In physics analysis, supervised learning is mostly used for the latter, such as distinguishing between SM background processes and BSM signals, or the tagging and identification of physics objects. Mathematically, this involves mapping an input (or feature) vector $\mathbf{X}$ (e.g. kinematic event features) to an output $\boldsymbol{Y}$ through a function $f(\mathbf{X}|\boldsymbol{\theta})$. The model is parameterised through the parameters $\boldsymbol{\theta}$ which are determined by the data. In a simple background/signal discrimination, one could label the signal by 1 and the background by 0.

The use of simulated data in training and the evaluation of the resulting algorithm on observed data presents challenges and necessitates dedicated data/MC calibrations, see Part IV. Additionally, generalisability, meaning that the algorithm performs well on data not included in the training set, is crucial. In particular, multiple measures can be taken to avoid overfitting to the training set.

Unsupervised learning, on the other hand, deals with unlabelled data, aiming to uncover hidden patterns or structures without predefined labels. While unsupervised learning is valuable for anomaly detection or clustering, it is, up to now, only sparingly used in LHC physics analyses.

The focus of this chapter will thus be on supervised learning, in particular the training of algorithms for classification tasks.

4.2 Neural Networks

Loosely inspired by biological neural systems, NNs are designed to learn from data and improve through experience. They are built from artificial neurons connected by edges that carry trainable weights. This section delves into the fundamental concepts of NNs, starting from the simplest unit, the perceptron, and extending to MLPs and DNNs.

4.2.1 The Perceptron

The perceptron is the simplest form of a NN and consists of a single unit. It is built from an input layer, a set of weights, a bias term, and an activation function. Fig. 4.1 illustrates the structure of a perceptron.

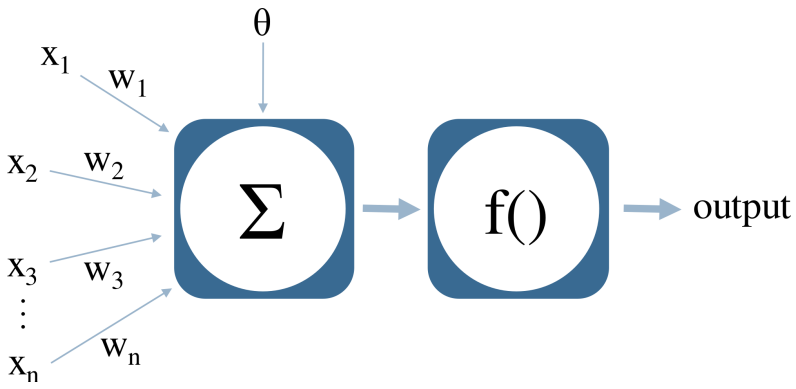


Figure 4.1: Model of a perceptron with the input features x_i, the corresponding weights w_i, and the bias θ. The output of the perceptron is given by applying the activation function f on the weighted sum of the inputs.

A perceptron receives a set of inputs x_i from a dataset with n features. Each input x_i is associated with a weight w_i, representing the strength of the connection. The activation potential (the weighted sum) of the perceptron is computed as:

$$ap = \boldsymbol{w}^T \boldsymbol{x} + \theta, \tag{4.1}$$

where $\boldsymbol{x} = \left(x_1, x_2, \ldots, x_n\right)$ is the input feature vector, $\boldsymbol{w} = \left(w_1, w_2, \ldots, w_n\right)$ is the weight vector, and θ is the bias term.

The output o of the perceptron is obtained by applying an activation function f to

the activation potential: $o(\boldsymbol{x}) = f(ap)$. Common activation functions include the step function, logistic sigmoid function, hyperbolic tangent, Rectified Linear Unit (ReLU), and Exponential Linear Unit (ELU). The ReLU is particularly popular for training DNNs with many layers, as it helps avoiding the vanishing gradient problem and leads to faster convergence.

Training a Perceptron

To train a perceptron, a loss function $L(\boldsymbol{w})$ must be defined. The objective of training is to minimise this loss function, thereby improving the perceptron's predictive accuracy. A widely used loss function is the mean squared error (MSE):

$$L(\boldsymbol{w}) = \frac{1}{2}\sum_{l=1}^{N}(y_l - o(\boldsymbol{x}_l, \boldsymbol{w}))^2, \tag{4.2}$$

where y_l are the target values of the N events in the training set. For binary classification tasks, the most common choice for the loss function is the binary cross-entropy (BCE), see Equation 7.2.

The training process involves adjusting the weights to minimise the loss function using optimisation algorithms, with stochastic gradient descent being a popular choice. In stochastic gradient descent, the weights are updated iteratively in the direction of the negative gradient of the loss function:

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \gamma \nabla_{\boldsymbol{w}} L(\boldsymbol{w}(t)), \tag{4.3}$$

where t is the iteration step, γ is the learning rate, and $\nabla_{\boldsymbol{w}} L(\boldsymbol{w}(t))$ is the gradient of the loss function with respect to the weight vector.

While the gradient and the weight update can be computed for each individual training event (stochastic gradient descent) or only once for the entire training set (batch gradient descent), a common approach in practice is mini-batch gradient descent. In mini-batch gradient descent, the training set is divided into batches. The gradients are accumulated, and the weights are updated after a full batch has been processed. This procedure is

advantageous for both computational efficiency and stability of the training.

The learning rate and the batch size are hyperparameters that can significantly impact the convergence and training time of the algorithm. The optimal values for such hyperparameters are often determined by testing multiple different choices.

4.2.2 Multi-Layer Perceptrons

A single perceptron is limited to solving linearly separable problems. To overcome this limitation, multiple artificial neurons are combined to form a MLP, also known as a fully-connected feedforward NN. A MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of neurons, and the output of each neuron in one layer serves as the input to the neurons in the subsequent layer. The architecture of a MLP with one hidden layer is illustrated in Fig. 4.2. The addition of hidden layers introduces non-linearity, enabling the MLP to model complex functions and solve more sophisticated problems.

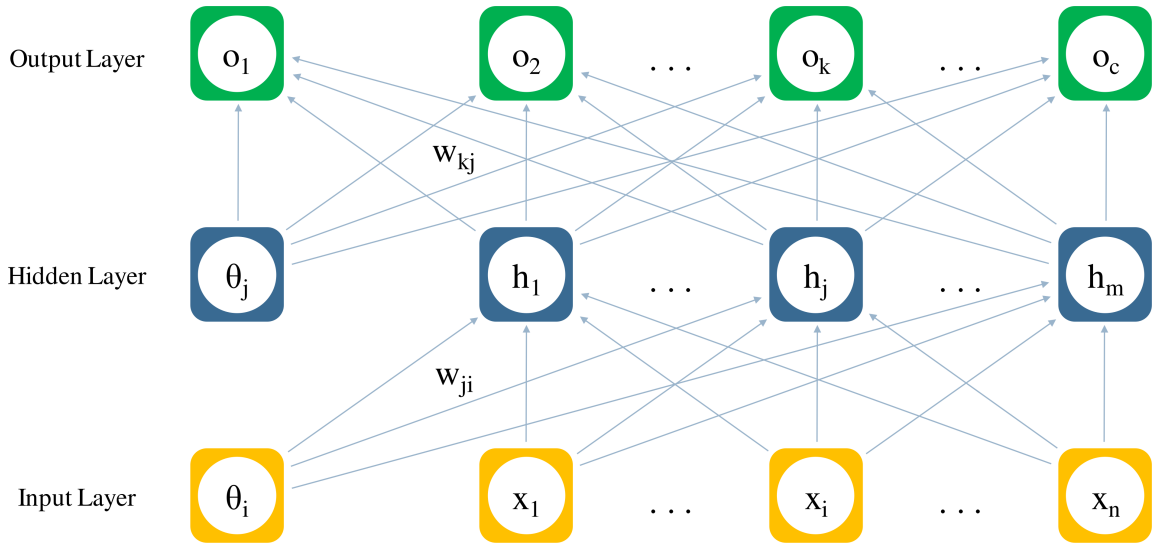


Figure 4.2: Illustration of a MLP with an input layer with n features, one hidden layer with m nodes and c output nodes. Such a network could be used to classify events into c different categories, such as physics processes or flavour compositions.

The output of unit k of a MLP with one hidden layer is computed as follows:

$$o_k = f\left(\sum_j^m w_{kj} f\left(\sum_i^n w_{ji} x_i + \theta_j\right) + \theta_k\right), \tag{4.4}$$

where w_{ji} are the weights between the input and hidden layer, w_{kj} are the weights between the hidden and output layer, θ_j and θ_k are the bias terms, and f is the activation function.

While theoretically, a MLP with one hidden layer can implement arbitrary decision boundaries, it is observed that a larger number of hidden layers increases the network's performance on complex problems. This approach, known as deep learning, involves NNs with many hidden layers and has achieved significant success in fields such as image recognition, natural language processing, and HEP.

4.2.3 Training Neural Networks: Backpropagation

The question arises of how to update the weights of the neurons in the hidden layers, as one typically only has target values for the output nodes. This is called the credit assignment problem and is solved by the backpropagation algorithm, which efficiently computes the gradients of the loss function with respect to each weight by applying the chain rule of calculus.

The backpropagation algorithm operates in two phases: the forward pass and the backward pass. During the forward pass, the input data is propagated through the network to generate predictions. The loss function, a measure of the discrepancy between the predicted outputs and the true labels, is then evaluated.

During the backward pass, the gradients of the loss function with respect to the weights are computed by backpropagating the errors from the output layer to the input layer. For a single training event with features $\boldsymbol{x}_l$ and loss E_l, the weights are updated similarly to Equation 4.3:

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \Delta\boldsymbol{w} = \boldsymbol{w}(t) - \gamma \nabla_{\boldsymbol{w}} E_l(\boldsymbol{w}(t)) \tag{4.5}$$

First, the gradients are computed straightforwardly with respect to the output weights

w_{kj}:

$$\frac{\partial E_l}{\partial w_{kj}} = \underbrace{\frac{\partial E_l}{\partial o_k}\frac{\partial o_k}{\partial ap_k}}_{\delta_k}\frac{\partial ap_k}{\partial w_{kj}} = \delta_k h_j \ , \tag{4.6}$$

where the sensitivity δ_k takes, in the case of MSE loss, the form $\delta_k = (y_k - o_k)f'(ap_k)$. h_j denotes the output of the hidden layer node with index j. The gradients with respect to the hidden layer weights are subsequently calculated as:

$$\frac{\partial E_l}{\partial w_{ji}} = \sum_k^c \underbrace{\frac{\partial E_l}{\partial o_k}\frac{\partial o_k}{\partial ap_k}}_{\delta_k}\underbrace{\frac{\partial ap_k}{\partial h_j}}_{w_{kj}}\underbrace{\frac{\partial h_j}{\partial ap_j}\frac{\partial ap_j}{\partial w_{ji}}}_{f'(ap_j)x_i} = \tag{4.7}$$

$$= \underbrace{f'(ap_j)\sum_k^c w_{kj}\delta_k}_{\delta_j} x_i = \delta_j x_i \tag{4.8}$$

The weights of the output layer and hidden layer are thus updated by

$$\Delta w_{kj} = \gamma \delta_k h_j \ , \tag{4.9}$$

$$\Delta w_{ji} = \gamma \delta_j x_i \ . \tag{4.10}$$

The forward pass includes the computation of the outputs of the individual layers (x_i, y_j, o_k), and the backward pass involves the computation of the sensitivities δ. This procedure can be applied to NNs of arbitrary depth, and the required sensitivities can be very efficiently computed by the weighted sum of sensitivities of the previous layer.

4.2.4 Advanced Neural Network Architectures

While MLPs and backpropagation provide a strong foundation for NN learning, advanced architectures and techniques have been developed to address specific challenges and improve performance across various tasks. The following sections provide a brief overview of the most commonly used architectures and algorithms.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) were originally designed for processing grid-like data structures, such as images or videos. In ATLAS, CNNs could be used to identify patterns of energy deposition in the calorimeters, potentially improving electron identification [161]. CNNs employ convolutional layers with shared weights that apply filters to local regions of the input, capturing spatial patterns. This approach significantly reduces the number of parameters compared to fully-connected feedforward NNs, enabling faster convergence and reducing the risk of overfitting.

Graph Neural Networks

GNNs extend the concept of CNNs to capture not only spatial patterns but also other types of connections or structures within the input data. In CNNs, spatially close inputs (e.g. pixels in an image) are connected, whereas in GNNs, arbitrary connections between different objects in a graph are introduced, with edges defining the connections. The resulting graph representation serves as the input to the GNN, and the nodes are updated iteratively using information from connected nodes.

These algorithms are highly effective for tasks such as flavour tagging, where the track constituents of jets are used as nodes in the graph. Relationships between individual tracks in a jet are exploited to assign a flavour label [159, 162]. This is advantageous compared to RNNs (see next paragraph), as it does not require a predefined ordering of the constituents. Thus, GNNs form the basis of the flavour tagging algorithms in Run 3.

Recurrent Neural Networks

RNNs are designed for sequential data, where the output at each time step depends on previous time steps. RNNs maintain a hidden state that captures information from previous inputs, making them suitable for tasks such as time series analysis and natural language processing. In HEP, RNNs are useful for processing inputs of variable length, such as the varying number of constituents in a jet or the different numbers of certain particles in an event. These objects are usually ordered, for example, by their p_{T} and

then fed as a sequence into a RNN.

RNNs are used in the RNNIP algorithm described in Section 2.4.2. Additionally, the use of Long short-term memory networks (LSTMs) [163], a special kind of RNN, was tested in the signal/background discrimination network used in the analysis described in Part III, due to the varying number of jets per event.

Adversarial Networks

Adversarial networks consist of two networks competing against each other. In industry, they are mostly used in the form of Generative Adversarial Networks (GANs), consisting of a generator and a discriminator. The generator produces synthetic data, while the discriminator attempts to distinguish between real and synthetic data. The two networks are trained simultaneously in a competitive way, where the generator aims to fool the discriminator, and the discriminator aims to correctly identify real and synthetic data. Outside of physics, such algorithms are popular for generating images and videos. In ATLAS, GANs are also being tested for the fast simulation of calorimeter showers for photons, electrons, and pions [164].

Within this thesis, adversarial networks are used slightly differently. A second network is added to a signal/background classifier to constrain the main classifier by requiring decorrelation with an auxiliary variable. This ensures that the background distribution remains a good variable for discovery and exclusion tests without being distortet by the classifier. For detailed information on the adversarial setup, see Part III.

4.3 Overfitting and Regularisation

A NN can be seen as a parameterised approximation $\hat{f}_{\boldsymbol{w}}(\boldsymbol{x})$ of an unknown target function $f(\boldsymbol{x})$. The goal of training is to adjust the parameters $\boldsymbol{w}$ to minimise the expected deviation between predictions and true function values. This deviation, known as overall

risk, is defined as:

$$\mathbb{E}\left[\left(\hat{f}_{\boldsymbol{w}}(\mathbf{x}) - f(\mathbf{x})\right)^2\right] = \int \left(\hat{f}_{\boldsymbol{w}}(\mathbf{x}) - f(\mathbf{x})\right)^2 p(\mathbf{x}) d\mathbf{x}, \tag{4.11}$$

where $p(\boldsymbol{x})$ is the probability density function of $\boldsymbol{x}$. Noise and incomplete knowledge often complicate the task of modelling the unknown function.

4.3.1 Bias-Variance Dilemma

A model that is too simple will lead to underfitting, resulting in large training errors. These models are said to exhibit a large bias. Conversely, overly complex models may perform well on the training set but fail to generalise to unseen data, leading to overfitting. Such models minimise the empirical risk (e.g., $\sum_i^N (\hat{f}_{\boldsymbol{w}}(\mathbf{x}) - y_i)^2$) instead of the overall risk. Overfitting models are typically highly dependent on noise or the spacing of sparse data, thus exhibiting large variance and leading to poor approximations of the target function.

The expected overall risk can be decomposed into bias squared and variance:

$$\mathbb{E}\left[\left(\hat{f}_{\boldsymbol{w}}(\mathbf{x}) - f(\mathbf{x})\right)^2\right] = \underbrace{\left(\mathbb{E}\left[\hat{f}_{\boldsymbol{w}}(\mathbf{x}) - f(\mathbf{x})\right]\right)^2}_{\text{bias}^2} + \underbrace{\mathbb{E}\left[\left(\hat{f}_{\boldsymbol{w}}(\mathbf{x}) - \mathbb{E}\left[\hat{f}_{\boldsymbol{w}}(\mathbf{x})\right]\right)^2\right]}_{\text{variance}} . \tag{4.12}$$

To minimise overall risk, one must balance bias and variance. This is challenging because reducing bias with a more complex model increases variance and vice versa. Hence, this feature is also known as the bias-variance trade-off and can be interpreted as a conservation law of errors.

The choice of model complexity (e.g., number of layers, nodes per layer, and parameters in a MLP) must be optimised based on the problem complexity and available data. Ideally, a model with infinite parameters and data would solve these issues, but data is typically limited. Controlling model complexity is known as regularisation and is discussed in 4.3.3.

4.3.2 Curse of Dimensionality

Increasing the number of input features enlarges the feature space, causing training data to become sparse and leading to high variance and overfitting. This "curse of dimensionality" means that exponentially larger datasets are required when increasing the input feature vector dimension. In practice, however, features are often correlated and occupy only a small subspace, somewhat mitigating the issue.

4.3.3 Regularisation Methods

To mitigate the risk of overfitting in NNs, several techniques can be employed, often involving the application of regularisation methods that impose constraints on the model. As discussed in Section 4.3.2, overfitting can be combated by dimensionality reduction. This can be achieved by using Principle Component Analysis (PCA) to eliminate redundant and irrelevant features, thereby simplifying the model.

Additionally, regularisation techniques such as L2 and L1 regularisation are commonly used to reduce model complexity by adding penalty terms for large weights to the loss function, thereby reducing the variance. L2 regularisation, also known as weight decay, adds a penalty proportional to the sum of the squares of the weights:

$$L_\lambda(\boldsymbol{w}, \boldsymbol{x}) = L(\boldsymbol{w}, \boldsymbol{x}) + \underbrace{\lambda \sum_i w_i^2}_{L_2} , \tag{4.13}$$

which helps prevent large weights and thus applies a smoothing constraint on the model.

L1 regularisation, on the other hand, adds a penalty proportional to the sum of the absolute values of the weights:

$$L_1 = \lambda \sum_i |w_i|, \tag{4.14}$$

which encourages sparsity in the model by driving some weights to zero, effectively performing feature selection. The hyperparameter λ determines the magnitude of the regularisation constraint and must be chosen depending on the task.

Selecting the appropriate values for the hyperparameters and model complexity, such as the number of layers or nodes per layer, is often done through cross-validation. In cross-validation, the training data is divided into disjoint subsets, and the model's performance on the training subset (minimised empirical risk) is compared to its performance on a validation set, which provides an estimate of the overall risk.

Network pruning is another technique that involves removing unnecessary neurons or connections, thereby simplifying the model without significantly impacting performance. Dropout is a regularisation method used in DNNs, where randomly selected neurons are ignored during training, reducing the chance of the model becoming overly reliant on specific neurons and improving generalisation. Early stopping is another effective strategy, where the training process is halted once the model's performance on a validation set begins to deteriorate, thus preventing overfitting to the training data.

4.4 Shapley Values

Shapley values, originating from cooperative game theory [165], provide a measure of the contribution of individual features to the prediction of a machine learning model. This is particularly useful as it allows the determination of how valuable a certain input variable is, for example, in the task of discriminating a BSM signal from SM background. For instance, the analysis detailed in Part III uses Shapley values to reduce the set of potential kinematic variables to those with a significant impact on the classification task, thereby avoiding issues such as the one discussed in Section 4.3.2.

Shapley values can be computed for various ML algorithms without requiring any knowledge of the model's internal structure or parameters. The Shapley value for a feature i in a model f, trained on N features, is defined as the average marginal contribution of that feature across all possible subsets S of the features. Mathematically, the Shapley value ϕ_i for feature i is given by:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f(S \cup \{i\}) - f(S) \right] \tag{4.15}$$

where $f(S)$ is the model prediction using the feature subset S [166].

When computing the output value $f(S)$ of a machine learning model with a subset of features S, it is necessary to handle the absence of certain features. One common approach is to marginalise over the missing features by integrating out their values. Since it is not feasible to retrain the model with all subsets of input features, this can be approximated by replacing the missing features with their expected values based on the data distribution.

Computing exact Shapley values can be computationally expensive, particularly for models with a large number of features, as it requires evaluating the model on all possible subsets of features. However, several approximation methods and software packages, such as SHAP [167, 168], have been developed to make this computation feasible. Examples of Shapley value distributions, generated with the SHAP package, are shown in Figure 7.1.

Part III

A Search for the dark-Higgs Boson

Chapter 5

Introduction and Analysis Setup

Contents

5.1 Introduction 119

5.2 Data Samples 119

5.3 Monte Carlo Samples 120

5.4 Trigger Strategy 124

5.5 Reconstructed Physics Objects 125

This chapter introduces a search for a dark-Higgs boson and outlines the analysis setup, including the samples used, reconstructed objects, and triggers.

5.1 Introduction

The following chapters present a search aimed at identifying BSM events with a final state containing four b-jets recoiling from $E_{\mathrm{T}}^{\mathrm{miss}}$. Specifically, the BSM production of two SM Higgs bosons, decaying into four b-quarks, in association with $E_{\mathrm{T}}^{\mathrm{miss}}$ resulting from DM production. Previous analyses, such as the one detailed in Ref. [1], have targeted this final state within the context of SUSY models. Despite having the same particles in the final state, the sensitivity of such reinterpretations is highly dependent on the kinematics of the final state, which are influenced by factors such as the mass of the invisible particles, decay topologies, spin effects, and particle lifetimes. Modern analyses often rely on highly optimised machine learning algorithms, utilising kinematic specificities of different processes to distinguish between them. Thus changes in kinematics can significantly affect performance and sensitivity for different signal models. To expand the scope of ATLAS searches in the di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ final state, the following chapters introduce the first search optimised for the resonant decay topology. An illustrative diagram of such a generic resonant signature (within the tri-scalar model, see Section 1.3.3) is shown in Figure 1.12(b). This search, however, targets the dark-Higgs model, detailed in Section 1.3.2, in the $s \to hh$ channel. The primary processes targeted in this search are depicted in Figure 5.1. Here, $E_{\mathrm{T}}^{\mathrm{miss}}$ arises from a second BSM state decaying invisibly, often in the opposite direction to the visible final state particles. This kinematic behaviour distinguishes these signals from many SM processes and other BSM signals with di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ signatures, such as the symmetric decays typical of SUSY theories.

5.2 Data Samples

The data used in this analysis were recorded during pp collisions at a centre-of-mass energy of 13 TeV with a 25 ns bunch-crossing interval, collected during LHC Run 2 from 2015 to 2018. After excluding runs where detector subsystems were not fully operational and ensuring high data quality [169], the dataset corresponds to an integrated luminosity of

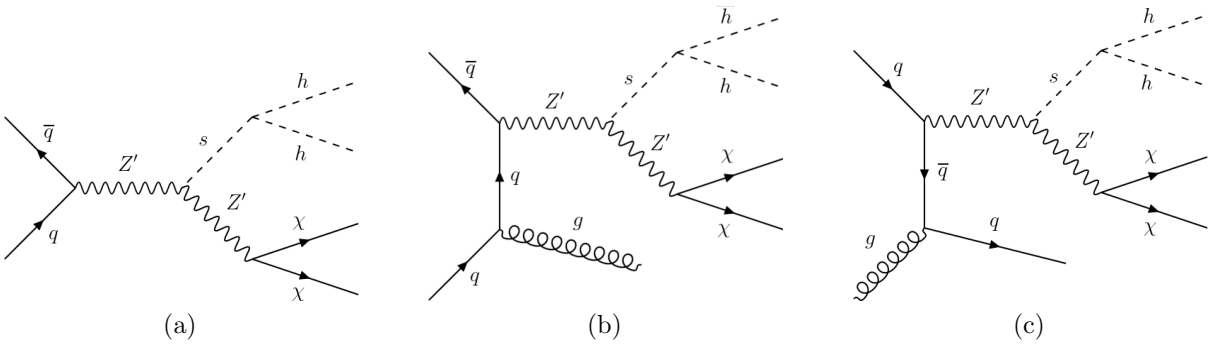


Figure 5.1: Diagrams of the three main processes contributing to the signal. Diagram (a) illustrates the dominant lowest-order dark-Higgs strahlung process, and (b) and (c) include additional initial state radiation.

$140\,\text{fb}^{-1}$.

5.3 Monte Carlo Samples

Predictions for signal and SM background processes in this analysis are based on simulated MC samples. Detailed information on the sample generation setup, including the utilised generators, PDFs, tuned parameters sets, and cross-section normalisations for all relevant processes, is summarised in Table 5.1.

All nominal samples are processed through a full ATLAS detector simulation [107] implemented in GEANT4 [123]. In-time and out-of-time pileup interactions from the same or neighbouring bunch-crossings are simulated by overlaying additional pp collisions generated by PYTHIA v8.186 [170], using the NNPDF2.3LO PDF set [171] and A3 tune [172].

5.3.1 Signal Samples

In the dark-Higgs signal samples, matrix elements (ME) are generated with up to two additional partons using MADGRAPH v2.9.9 [110] at LO with the NNPDF 2.3 PDF sets [173] and the A14 tune [174]. PYTHIA v8.307 [106] is subsequently used to model PS, hadronisation, and underlying event (UE).

Signals are generated for a wide range of dark-Higgs masses and Z' masses. The

dark-Higgs masses range from 251 GeV to 435 GeV, with 25 GeV increments. To ensure good coverage of the critical mass ranges, additional points were generated at the lower end of the mass spectrum at $m_s \gtrsim 2m_h$ and at $m_s \sim 2m_\chi$. The Z' mass ranges from 420 GeV to 3.3 TeV, with narrower spacing at the lower end and 400 GeV increments at the upper end. An overview of the generated signal grid and the relevant signal cross-sections is shown in Figure 5.2.

The values of the remaining model parameters (see Equation 1.63) are chosen to align with benchmarks used in other analyses. The DM mass m_χ is set to a benchmark value of 200 GeV, selected to strike a balance between lower values, which allow access to lighter Z' masses favoured by relic density constraints, and higher values, which avoid opening the $s \to \chi\bar{\chi}$ channel too early. Specifically, a smaller m_χ would cause the $s \to \chi\bar{\chi}$ decay to become kinematically accessible at $m_s \geq 2m_\chi$, diminishing sensitivity to competing channels such as $s \to VV$ and $s \to hh$, which only open at higher m_s.

This benchmark choice has been consistently adopted in all ATLAS and CMS searches targeting dark-Higgs strahlung. However, in theory, this analysis in the $s \to hh$ channel could extend its sensitivity to higher dark-Higgs masses, above 400 GeV, by selecting a DM mass greater than 200 GeV.

The coupling constant in the dark sector, g_χ, is set to unity, the coupling between the SM quarks and the Z' is $g_f = 0.25$, and the mixing angle between the SM Higgs and dark-Higgs is set to 0.01. For a detailed introduction to the dark-Higgs model and its parameters, see Section 1.3.2.

5.3.2 Standard Model Background Samples

Overview of Contributing Background Processes

After the preselection of significant $E_{\mathrm{T}}^{\mathrm{miss}}$ and high b-jet multiplicities (see Section 6.2), the dominant remaining background consists of semi-leptonically decaying $t\bar{t}$ events. In these events, one top quark decays fully hadronically, while the other decays into a b-quark and a W, which subsequently decays either leptonically or to a hadronically decaying tau lepton. If the lepton is not reconstructed and at least one additional jet is identified as a b-jet, this

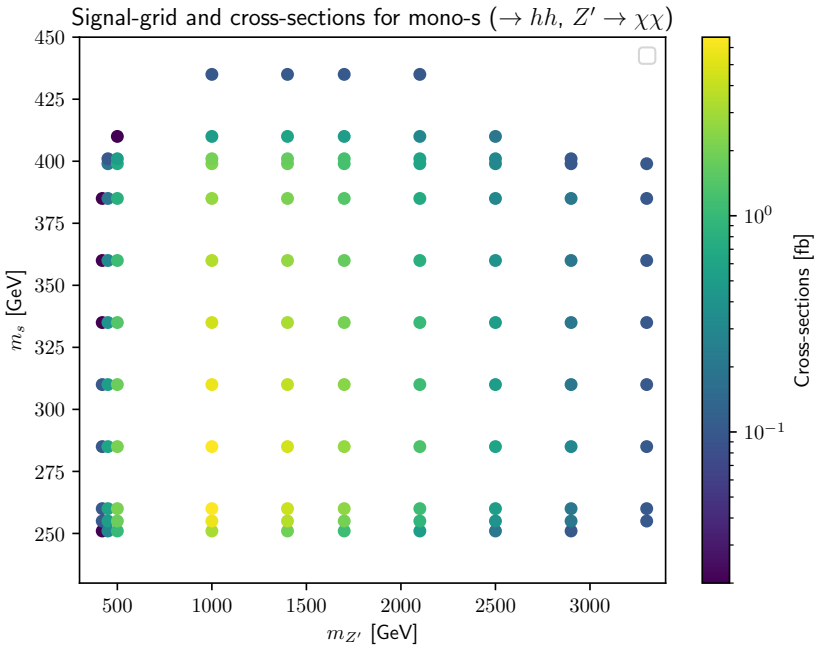


Figure 5.2: The dark-Higgs signal grid used in the analysis and the corresponding cross-sections parameterised in $m_{Z'}$ and m_s.

background can mimic the final states of the targeted signal models. However, applying cuts on transverse-mass (m_{T}) based variables in the dark-Higgs search, as described in Chapter 6.3, can significantly reduce this background.

The $t\bar{t}$ processes are further categorised based on the flavour content of the truth-jets in the MC events. An event is labelled as $t\bar{t} + \geq b$ if three or more truth-jets are associated with b-quarks. Events with two b-jets and at least one c-jet are labelled as $t\bar{t} + \geq c$, while all remaining events are categorised as $t\bar{t} + \geq light$. These $t\bar{t}$ subprocesses are treated separately, as they may populate different analysis regions to varying extents, as discussed in Chapter 8.

Other background processes involving top quarks are grouped into the categories of single-top and top + electroweak (topEW) processes. These categories include processes such as $t +$ jets, $t + W$, and $t\bar{t} + W$, $t\bar{t} + Z$, $t\bar{t} + h$, and events with four top quarks. Additionally, $Z +$ jets and $W +$ jets with neutrinos from leptonic vector boson decays also play a significant role in this analysis. Requirements on the significance of $E_{\mathrm{T}}^{\mathrm{miss}}$ and angular separation between it and energetic jets effectively remove the multijet background (see Section 6.3.2). Finally, events with two vector bosons contribute a small but important

background, while events with three vector bosons and VH processes are found to be negligible.

Background Sample Generation

The sample for the estimation of the $t\bar{t}$ background is generated using the POWHEG BOX generator [111–113] at NLO with the NNPDF3.0 PDF set [175] and interfaced to PYTHIA v8.230 for PS, hadronisation, and UE using the NNPDF2.3 PDF set. Heavy-flavour hadron decays are modelled using the EVTGEN v1.6.0 package [176]. The h_{damp} parameter in POWHEG, which controls the p_{T} of the first additional emission beyond the Born level and regulates the p_{T} of the recoil emission against the $t\bar{t}$ system, is set to 1.5 times the mass of the top quark. All events with at least one semi-leptonically decaying top quark are included. Fully hadronic $t\bar{t}$ events do not contain sufficient $E_{\text{T}}^{\text{miss}}$ to contribute significantly to the background. To ensure sufficient MC statistics in the signal-sensitive regions, H_{T}- and $E_{\text{T}}^{\text{miss}}$-sliced $t\bar{t}$ samples [177] are used in addition to the inclusive sample. Orthogonality between the combined samples is maintained by the following logic: Events from the inclusive sample are selected if $H_{\text{T}} < 600$ GeV and $E_{\text{T}}^{\text{miss}} < 200$ GeV, H_{T} sliced samples are used for $E_{\text{T}}^{\text{miss}} < 200$ GeV, and $E_{\text{T}}^{\text{miss}}$ sliced samples cover the remaining phase space.

The single-top sample is generated using POWHEG BOX v2 interfaced to PYTHIA v8.230. For the $Z+$jets background, matrix elements are computed using the Comix [178] and OpenLoops [179] generators and merged with the SHERPA v2.2.1 parton shower [180] using the ME+PS@NLO prescription [181]. For the Wt-channel of the single-top background, H_{T}- and $E_{\text{T}}^{\text{miss}}$-sliced samples are used in addition to the inclusive ones, which are combined following the same logic as for the $t\bar{t}$ samples.

Additional backgrounds come from the production of $t\bar{t}$ pairs in association with $W/Z/h$ and additional jets. Processes such as $t\bar{t}W$, $t\bar{t}Z$, four top-quark, and $t\bar{t}h$ are grouped into a single category, referred to as topEW. The $t\bar{t}h$ sample is generated using POWHEG BOX v2 interfaced to PYTHIA v8.230, while the $t\bar{t}+W/Z$ and $t\bar{t}\,t\bar{t}$ samples use MADGRAPH5_AMC@NLO interfaced with PYTHIA v8.210 and v8.186 [182], respectively.

Other smaller backgrounds include $W+$jets and $WW/WZ/ZZ$ events. The latter are collectively referred to as diboson events. For diboson and $W+$jets events, matrix elements and showering are computed similarly to the $Z+$jets samples. The matrix elements for all these background samples use the NNPDF3.0 PDF set, except the $t\bar{t}t\bar{t}$ sample, which uses NNPDF2.3. All Pythia8 samples use the A14 [174] tune and the NNPDF2.3 PDF set for modelling the UE.

Process	**Generator + fragmentation/hadronisation**	**Tune**	**PDF set**	**Cross-section Order**
dark-Higgs signals	MadGraph v2.9.9 + Pythia v8.307	A14	NNPDF2.3	LO
$t\bar{t}$	Powheg Box v2 + Pythia v8.230	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NNLO+NNLL [183]
Single top	Powheg Box v2 + Pythia v8.230	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NLO [184] (t/s-channel) NLO+NNLL [185] (Wt)
$t\bar{t}W/t\bar{t}Z$	MadGraph5_aMC@NLO v2.3.3 + Pythia-8.210	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NLO [110]
4-tops	MadGraph5_aMC@NLO v2.2.2 + Pythia v8.186	A14	NNPDF2.3	NLO [110]
$t\bar{t}h$	Powheg Box v2 + Pythia v8.230	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NLO [186]
Dibosons WW, WZ, ZZ	Sherpa v2.2.1	Default	NNPDF3.0	NLO [181, 187]
W/Z**+jets**	Sherpa v2.2.1	Default	NNPDF3.0	NNLO [188]
VH	Pythia v8.186	A14	NNPDF2.3	LO [186]
Triboson	Sherpa v2.2.14	Default	NNPDF3.0	NLO [187]

Table 5.1: Details of the sample generation setup for the various processes relevant to this analysis. The UE tunes, PDF sets, and the pQCD highest-order accuracy used for the normalisation of the samples are listed. Note that in the following chapters, the $t\bar{t}W$, $t\bar{t}Z$, four top-quark, and $t\bar{t}h$ processes are combined and treated as a single $t\bar{t}+$ electroweak (topEW) category.

5.4 Trigger Strategy

Since the final state targeted in this search includes a large amount of $E_{\mathrm{T}}^{\mathrm{miss}}$, the analysis uses the lowest threshold unprescaled $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger in each year of data taking for the initial event selection. The utilised $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger thresholds are summarised in Table 5.2.

It should be noted that while the $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds of the triggers vary by year, all triggers are fully efficient at $E_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV, in line with the offline $E_{\mathrm{T}}^{\mathrm{miss}}$ cut applied in the preselection of the analysis.

Year	Online selections	Offline selections
2015	$E_{\mathrm{T}}^{\mathrm{miss}}(\text{inv. } \mu) > 70$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
2016	$E_{\mathrm{T}}^{\mathrm{miss}}(\text{inv. } \mu) > 90$ GeV	
2017	$E_{\mathrm{T}}^{\mathrm{miss}}(\text{inv. } \mu) > 100$ GeV	
2018	$E_{\mathrm{T}}^{\mathrm{miss}}(\text{inv. } \mu) > 110$ GeV	

Table 5.2: Summary of the $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds for the various $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers used during LHC Run 2 data-taking. Online selections refer to trigger conditions applied during data collection, while offline selections refer to those applied to the fully reconstructed objects used in the analysis. As muons are invisible to the trigger, their momentum is included in the reconstructed online $E_{\mathrm{T}}^{\mathrm{miss}}$ value.

5.5 Reconstructed Physics Objects

The primary objects used in this analysis are small-R b-jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. Except for one CR used for background estimation (see Chapter 8), electrons and muons are vetoed throughout the analysis. Three distinct definitions are established for the various physical objects, each employed at different stages of the analysis. `Loose` electrons and muons are used for vetoes as recommended by the ATLAS SUSY combination team, simplifying potential future statistical combinations of analyses with orthogonal selections. `Baseline` objects are used to build $E_{\mathrm{T}}^{\mathrm{miss}}$ and serve as inputs to the overlap removal (OR). Ultimately, the kinematic properties and multiplicities of `Signal` objects are utilised for preselections, analysis region definitions, and as inputs to the NN. In addition to the detailed descriptions below, the definitions of the various objects used in the analysis are summarised in Table 5.3.

5.5.1 Jets

Small-R ($R = 0.4$) jets are reconstructed using the anti-k_t algorithm with PFOs as inputs (see Section 2.4.2). `Baseline` jets must have $p_{\mathrm{T}} > 20$ GeV, while `Signal` jets must have

$p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.8$. To suppress pile-up jets, `Signal` jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ must have a jet vertex tagger (JVT) score > 0.2 (the `Medium` JVT working point) [189]. JVT is a k-nearest neighbour based algorithm [190], used to suppress pile-up by testing the closeness of events based on different metrics. The 77% b-tagging efficiency WP of the DL1r algorithm (see Section 2.4.2) is used to b-tag `Baseline` jets within the acceptance of the ID ($|\eta| < 2.5$).

5.5.2 Electrons

`Loose` electrons must pass the centrally defined `Loose` likelihood identification criteria with at least one B-layer hit [124, 191], and must satisfy $p_{\mathrm{T}} > 8$ GeV and $|\eta| < 2.47$. `Baseline` electrons are the subset of `Loose` electrons with $p_{\mathrm{T}} > 20$ GeV and $z_0 \sin(\theta) < 0.5$ mm. `Signal` electrons are the subset of `Baseline` electrons passing OR, centrally defined `Loose` isolation and medium likelihood identification criteria [124, 191], with $s_{d_0} < 5$.

5.5.3 Muons

`Loose` muons must be of centrally defined `Medium` quality [126, 192] and have $p_{\mathrm{T}} > 8$ GeV and $|\eta| < 2.7$. `Baseline` muons are the subset of `Loose` muons with $p_{\mathrm{T}} > 20$ GeV and $z_0 \sin(\theta) < 0.5$ mm. `Signal` muons are the subset of `Baseline` muons passing OR, tight isolation requirements and have $|\eta| < 2.5$ and $s_{d_0} < 3$ [126, 192]. The `Baseline` muons are also used to correct the four-momentum of semileptonic decays in b-jets. If a muon is found within $\Delta R < 0.4$ of a jet, the muon's four-momentum is added to that of the jet. This correction uses muons that fail OR, as no such muons would survive OR.

5.5.4 Missing Transverse Energy

The $E_{\mathrm{T}}^{\mathrm{miss}}$ of each event is defined as the magnitude of the negative vector sum p_{T} of all selected and calibrated `Baseline` physics objects in the event, with an extra term added to account for soft energy not associated with any selected objects (see Section 2.4.4). To minimise pile-up effects, the analysis uses the centrally defined `Loose` WP [193]. Additionally, the object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, defined in Ref. [194], plays an important

role in the analysis. The latter is a measure of the amount of genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event. It is calculated as the $E_{\mathrm{T}}^{\mathrm{miss}}$ divided by its resolution built from the resolutions of the individual objects from which $E_{\mathrm{T}}^{\mathrm{miss}}$ is constructed.

5.5.5 Overlap Removal

OR is applied to avoid double-counting low-level objects in the reconstruction of small-R jets, electrons, and muons. The OR algorithm is applied to the `Baseline` objects defined above. The sequential algorithm is as follows: Firstly, electrons sharing ID tracks with muons are removed. Secondly, jets within $\Delta R < 0.2$ of electrons are discarded. Subsequently, electrons within the proximity of the remaining jets are removed if they satisfy $\Delta R(j, e) < \min(0.4, 0.04 + \frac{10\mathrm{GeV}}{p_T^e})$. Finally, jets within $\Delta R < 0.2$ of muons and jets with a ghost-associated muon track are discarded, and muons within the proximity of the remaining jets are removed if they satisfy $\Delta R(j, \mu) < \min(0.4, 0.04 + \frac{10\mathrm{GeV}}{p_T^\mu})$.

Physics Object	Requirement	`Loose`	`Baseline`	`Signal`
Electrons	subset of	-	`Loose`	`Baseline`
	p_{T} [GeV]	> 8	> 20	> 20
	$\lvert\eta\rvert$	< 2.47	< 2.47	< 2.47
	s_{d_0}	-	-	< 5
	$z_0 \sin(\theta)$ [mm]	-	< 0.5	< 0.5
	ID	`Loose`	`Loose`	`Loose`
	Isolation	-	-	`Loose`
	passOR	-	-	true
Muons	subset of	-	`Loose`	`Baseline`
	p_{T} [GeV]	> 8	> 20	> 20
	$\lvert\eta\rvert$	< 2.7	< 2.7	< 2.5
	s_{d_0}	-	-	< 3
	$z_0 \sin(\theta)$ [mm]	-	< 0.5	< 0.5
	ID	`Medium`	`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	true
	b-tag WP DL1r	-	77%	77%
	b-tag $\lvert\eta\rvert$	-	< 2.5	< 2.5

Table 5.3: Summary of the selections used for the `Loose`, `Baseline`, and `Signal` object definitions for electrons, muons, and jets. The table is adapted from Ref. [1].

Chapter 6

Event Selection and Higgs Reconstruction

Contents

6.1 Event Cleaning 130

6.2 Preselection 130

6.3 Optimised Selection 131

6.4 Dark-Higgs Reconstruction 137

This chapter details the event selection and Higgs reconstruction algorithm used in the dark-Higgs search.

6.1 Event Cleaning

To ensure high data quality, centrally defined recommendations are followed to discard corrupted data affected by issues in the ID and the calorimeters [169].

Additionally, fake jets can arise from non-collision backgrounds or cosmic muons depositing energy in the calorimeters [195]. The centrally defined `Loose` *bad-jet* WP [196], which is based on the quality of the fit on the calorimeter pulse shape, the fraction of jet energy associated with specific calorimeter cells, and the charged momentum fraction of the jet, is applied.

Fake muons can be reconstructed from high hit-multiplicities in the muon spectrometer due to very energetic punch-through jets, or from poorly measured ID tracks wrongly matched to muon spectrometer segments. Such fake muons are a potential source of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ and are thus vetoed by centrally defined muon momentum resolution criteria.

Moreover, muon tracks with large longitudinal and transverse impact parameters $|z_0| > 1$ mm and $|d_0| > 0.2$ mm are vetoed to reject muons originating from cosmic rays.

6.2 Preselection

For historical reasons, and to utilise studies conducted for the analysis detailed in Ref. [1], the event selection is divided into two sets of cuts: the common preselection cuts, which this analysis shares with others targeting the same final state (see, e.g., Ref. [1]), and the optimised selection, which introduces cuts tailored to the resonant decay topology of the dark-Higgs signals.

6.2.1 Lepton Vetoes

Events with more than one `Loose` lepton (muon or electron) are discarded. A single `Loose` lepton is allowed to retain events with semileptonic b-hadron decays. This cut aligns with the recommendations of the ATLAS SUSY combination group, facilitating future statistical combinations of analyses with orthogonal lepton number selections.

All subsequent cuts discussed in this and the following chapters refer to objects of `Signal` quality unless otherwise specified.

6.2.2 Kinematic Preselection

To target the $4b + E_{\mathrm{T}}^{\mathrm{miss}}$ final state, events are required to have zero leptons and at least four jets. Since no additional b-jets are expected in the signal, but they are in the background through top pair processes, a maximum jet multiplicity of seven is imposed. Of the selected jets, at least three must be b-tagged by the DL1r algorithm at the 77% WP. The significant $E_{\mathrm{T}}^{\mathrm{miss}}$ recoiling against the b-jets in the signal topology allows for raising the required $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold to > 200 GeV, compared to similar analyses that use > 150 GeV. This ensures that the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers are fully efficient.

Furthermore, the variable $\Delta\phi_{\mathrm{min}}^{4j}$ is defined as the minimum separation in ϕ between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and any of the four leading jets (in p_{T}):

$$\Delta\phi_{\mathrm{min}}^{4j} = \min_{i}(\Delta\phi(\mathrm{jet}_i, E_{\mathrm{T}}^{\mathrm{miss}})) \ . \tag{6.1}$$

Spurious $E_{\mathrm{T}}^{\mathrm{miss}}$ resulting from jet mismeasurements is per definition closely aligned with the jets. Therefore, a cut $\Delta\phi_{\mathrm{min}}^{4j} > 0.4$ is introduced to reduce the contribution of QCD multijet background events. Table 6.2 summarises the applied selections on the `Signal` objects.

6.3 Optimised Selection

The distinct decay topology associated with dark-Higgs radiation allows for a tighter selection than the one detailed in the previous section. The three main processes targeted by this search, illustrated in Figure 5.1, offer additional kinematic information that can be leveraged to enhance the sensitivity of the analysis.

For instance, the resonant topology enables the reconstruction of the masses of the BSM states using the invariant masses of the visible final state objects. The back-to-back nature of the b-jets relative to the $E_{\mathrm{T}}^{\mathrm{miss}}$ is particularly useful, as it helps to distinguish

Variable	Definition
$\Delta R_{\text{min}}(b,b)$	$\min_{ij}(\Delta R(\text{b-jet}_i, \text{b-jet}_j))$
$\Delta R_{\text{min}}(b,b)/E_{\text{T}}^{\text{miss}}$	$\frac{\Delta R_{\text{min}}(b,b)}{E_{\text{T}}^{\text{miss}}}$
m_{T}^{4b}	$\sqrt{(p_{\text{T}}^{4b}+E_{\text{T}}^{\text{miss}})^2-(p_{\text{x}}^{4b}+E_{\text{x}}^{\text{miss}})^2-(p_{\text{y}}^{4b}+E_{\text{y}}^{\text{miss}})^2}$
$m_{\text{T},min}^{b}$	$\min_{1\leq i\leq 3}^{\text{b-jets}} \sqrt{(p_{\text{T}}^{i}+E_{\text{T}}^{\text{miss}})^2-(p_{\text{x}}^{i}+E_{\text{x}}^{\text{miss}})^2-(p_{\text{y}}^{i}+E_{\text{y}}^{\text{miss}})^2}$

Table 6.1: Key variables highlighting the different kinematic properties of the dark-Higgs signal and the SM backgrounds. The transverse momentum p_{T}^{4b} is the vectorial sum of the p_{T} of all jets selected in the Higgs reconstruction (see Section 6.4). In the case of the signal, it is an estimate of the p_{T} of the dark-Higgs boson.

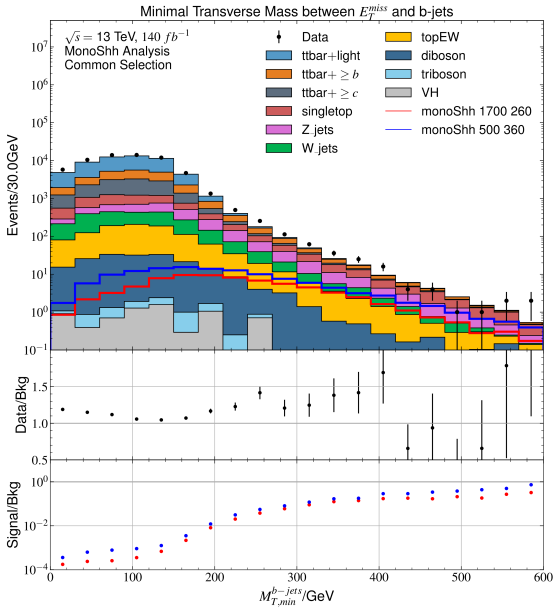
the signal from the primary SM backgrounds. In the predominant $t\bar{t}$ background, for example, one b-jet and a neutrino, responsible for the $E_{\text{T}}^{\text{miss}}$, originate from the same top decay and are thus in general spatially closer to each other.

6.3.1 Discriminating Kinematic Variables

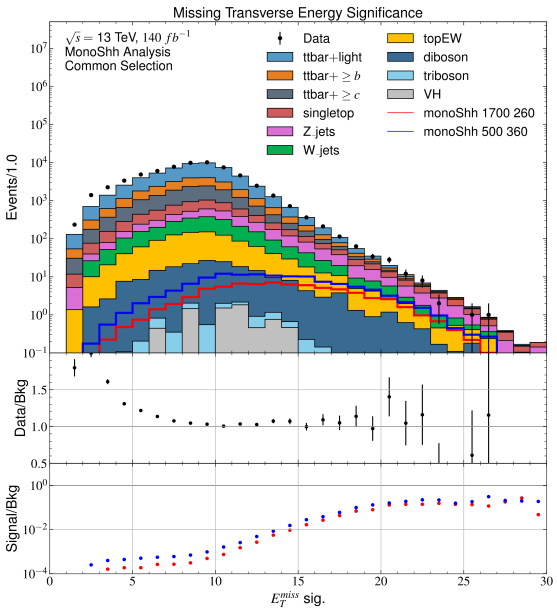
To effectively reduce the backgrounds while maintaining high signal efficiency, the discriminative power of specific kinematic variables are leveraged. This allows the definition of cuts that significantly reduce general background yields, including minimising the contribution of QCD multijet processes to negligible levels.

Four of the most effective variables employed in this analysis are: $m_{\text{T},min}^{b}$, $S(E_{\text{T}}^{\text{miss}})$, $\Delta R_{\text{min}}(b,b)/E_{\text{T}}^{\text{miss}}$, and m_{T}^{4b}. These variables are defined in Table 6.1, and their distributions for various SM background processes, as well as two benchmark signal points, are shown in Figure 6.1. Note that no dedicated multijet estimate has been included in these plots, which is evident in the data/MC mismatch observed in certain regions, such as for low $S(E_{\text{T}}^{\text{miss}})$ values in Figure 6.1(b).

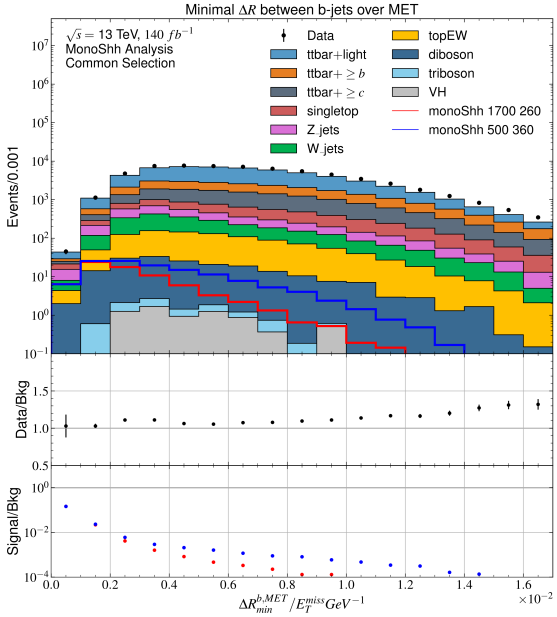
Based on these variables, selections with high signal acceptance and significant background rejection can be established. In theory, a background rejection of more than two orders of magnitude can be achieved while maintaining signal efficiencies above 50%. However, detailed studies indicate that applying only relatively loose cuts and relying on an multivariate analysis (MVA) algorithm, such as the NN detailed in Chapter 7, for



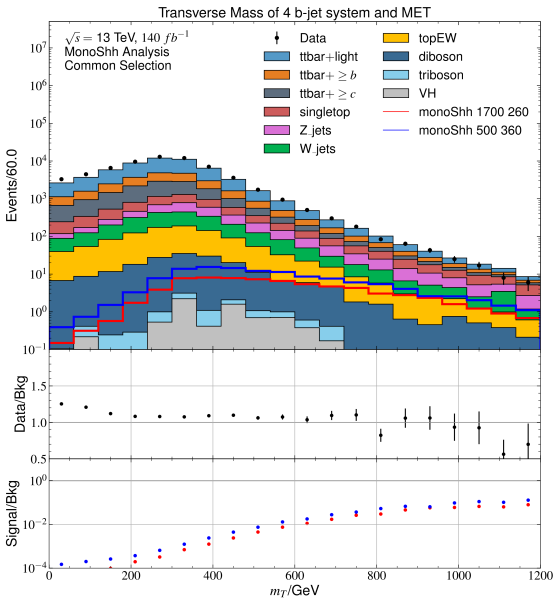
(a) $m^{b}_{\mathrm{T},min}$ distributions.



(b) $S(E^{\mathrm{miss}}_{\mathrm{T}})$ distributions.



(c) $\Delta R_{\mathrm{min}}(b,b)/E^{\mathrm{miss}}_{\mathrm{T}}$ distributions.



(d) m^{4b}_{T} distributions.

Figure 6.1: Distributions of some highly discriminating kinematic variables after applying the preselection shared with other $4b + E^{\mathrm{miss}}_{\mathrm{T}}$ analyses. Note that no dedicated multijet estimate has been included in these plots.

Variable	Cut value
Preselection	
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	> 150
n_{leptons}	0
n_{jets}	$(4, 7)$
$n_{b\text{-jets}}$	≥ 3
$\Delta\phi_{\mathrm{min}}^{4j}$	> 0.4
Optimised selection	
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	> 200
$m_{\mathrm{T},min}^{b}$ [GeV]	> 80
$S(E_{\mathrm{T}}^{\mathrm{miss}})$	> 7

Table 6.2: Overview of the selections applied in this analysis. All mentioned objects are of `Signal` quality (see Table 5.3).

the majority of background rejection yields the best performance (see Section 7.2). The selections consistently applied throughout the analysis are summarised in Table 6.2 and are referred to as the optimised selection for the dark-Higgs analysis.

This selection reduces the overall background yield, especially the dominant $t\bar{t}$ processes, by more than a factor of five. Compared to the common preselection, the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut was raised to > 200 GeV. This ensures that the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers described in Section 5.4 operate above their plateau threshold, eliminating the need for custom scale-factors. Additionally, the raised $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold reduces the dominant $t\bar{t}$ background significantly.

Implementing a cut on $m_{\mathrm{T},min}^{b}$ which is shown in Figure 6.1(a), effectively further reduces the contamination with $t\bar{t}$, as the transverse mass has an upper bound of $m_t \approx 175$ GeV for $t \to bW(\to l\nu)$ processes. Stricter constraints on this variable were avoided due to the limited number of background events available, which are crucial for training an optimised machine learning algorithm. For more details on the training set, see Section 7.2.

The final cut in the optimised selection is applied to $S(E_{\mathrm{T}}^{\mathrm{miss}})$ and is discussed in Section 6.3.2

As depicted in Figure 6.1(d), m_{T}^{4b} exhibits high values for the dark-Higgs signals since the four-b-jet system is back-to-back with the $E_{\mathrm{T}}^{\mathrm{miss}}$. For background processes with

semi-leptonically decaying top quarks, these values are generally lower since the neutrino originates from the same top quark as one of the b-jets. Similarly, $\Delta R_{\text{min}}(b,b)/E_{\text{T}}^{\text{miss}}$ (see Figure 6.1(c)) is a highly discriminating variable because dark-Higgs signals tend to exhibit high $E_{\text{T}}^{\text{miss}}$ values and closely spaced b-jets.

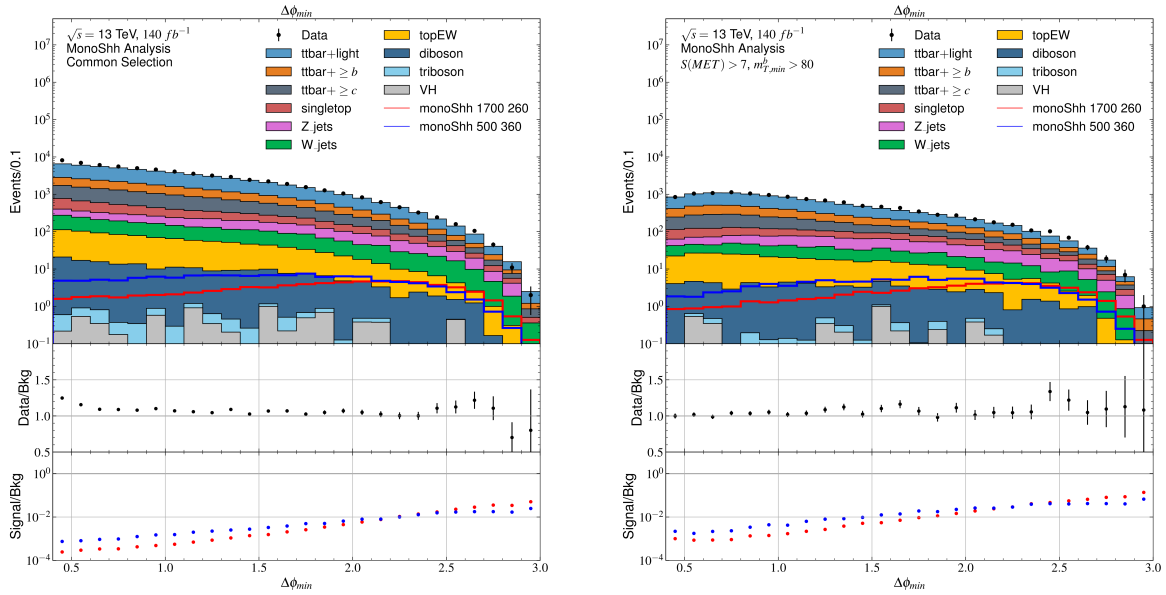
Although m_{T}^{4b} and $\Delta R_{\text{min}}(b,b)/E_{\text{T}}^{\text{miss}}$ offer strong signal-to-background discrimination, the final analysis strategy chose not to include these variables in the optimised selection. Instead, a NN was employed to leverage these variables and their correlations with others for optimal separation.

6.3.2 The QCD Multijet Background

In addition to the preselection cut on $\Delta\phi_{\text{min}}^{4j}$, both the raised $E_{\text{T}}^{\text{miss}}$ and the $m_{\text{T},min}^{b}$ cuts further reduce the remaining QCD background, particularly in cases where the $E_{\text{T}}^{\text{miss}}$ results from mis-measurements and aligns with one of the jets. Furthermore, $S(E_{\text{T}}^{\text{miss}})$ is specifically designed to distinguish events where the reconstructed $E_{\text{T}}^{\text{miss}}$ originates from invisible particles from those where it is consistent with contributions from particle mis-measurement, resolutions, and inefficiencies [197].

The mismodelling at low values in Figure 6.1(b) captures the multijet events not included in these plots. Thus, applying a cut on this variable particularly reduces the remaining QCD multijet background and allows us to rely entirely on background processes with genuine $S(E_{\text{T}}^{\text{miss}})$, modelled by MC simulation and CRs, to estimate the backgrounds. A comparison with the background estimation of the analysis detailed in Ref. [1] further confirms that the multijet background is exclusively located in regions removed by the tightened cuts on $E_{\text{T}}^{\text{miss}}$, $S(E_{\text{T}}^{\text{miss}})$ and $m_{\text{T},min}^{b}$.

Figure 6.2 compares the $\Delta\phi_{\text{min}}^{4j}$ distributions before and after applying the optimised selection. The QCD background is primarily concentrated at the lower end of the $\Delta\phi_{\text{min}}^{4j}$ spectrum, which was also verified by the analysis detailed in Ref. [1]. After applying the optimised selection criteria, the agreement between observed data and MC improves significantly at low values, demonstrating the negligibility of the QCD background in the analysis selection.



(a) Preselection shared with other analyses.

(b) Full selection.

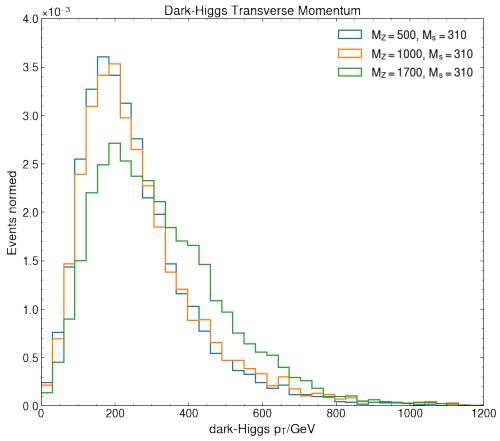
Figure 6.2: Distributions of $\Delta\phi^{4j}_{\text{min}}$ with common selection (a) and optimised selection (b) applied. The mismodelling at low values in (a) is a result of the unestimated multijet estimation. The application of the optimised selection (b) significantly removes this mismodelling by eliminating the multijet contribution.

6.3.3 Kinematic Dependence on Signal Masses

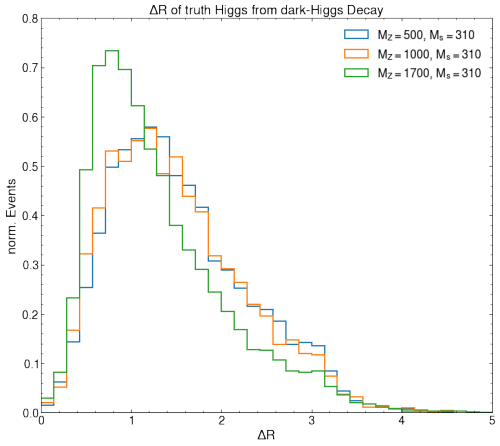
Although different signal mass points show relatively similar behaviour in most kinematic variables, discrimination improves slightly for higher $m_{Z'}$. The trend in $m_{Z'}$ can be attributed to the angular separation of the dark-Higgs decay products and the increased $E_{\text{T}}^{\text{miss}}$. A higher $m_{Z'}$ mass leads to a harder dark-Higgs, as shown in Figure 6.3, resulting in more collimated decay products.

Furthermore, higher mass Z' are typically produced more centrally, resulting in increased $E_{\text{T}}^{\text{miss}}$ and $S(E_{\text{T}}^{\text{miss}})$. The larger $E_{\text{T}}^{\text{miss}}$ and smaller ΔR of the decay products are exploited in the $\Delta R_{\text{min}}(b,b)/E_{\text{T}}^{\text{miss}}$ variable, providing better discrimination power for high mass signals. Additionally, due to the more collimated four-b-jet signature in the high mass scenario, the back-to-back nature of the process is more pronounced, enhancing the effectiveness of m_{T}-based variables.

Overall, the kinematics of the dark-Higgs strahlung process are relatively robust against variations in the masses of the dark-Higgs boson and Z'.



(a) p_{T} of dark-Higgs boson.



(b) ΔR of truth SM Higgs bosons originating from the dark-Higgs decay vertex.

Figure 6.3: Kinematic truth distributions of simulated dark-Higgs radiation events with a dark-Higgs mass of 310 GeV. Higher Z' masses result in more boosted Higgs bosons (a), which are therefore more collimated (b), enhancing the back-to-back nature that discriminates the signals from many SM processes.

6.4 Dark-Higgs Reconstruction

The fully visible final states of the dark-Higgs decay enable the reconstruction of its mass. The dark-Higgs mass, or di-Higgs mass, is reconstructed as $\sqrt{(p_{h_1} + p_{h_2})^2}$, where p_{h_i} represents the four-momenta of the reconstructed Higgs candidates. The reconstruction process involves two steps.

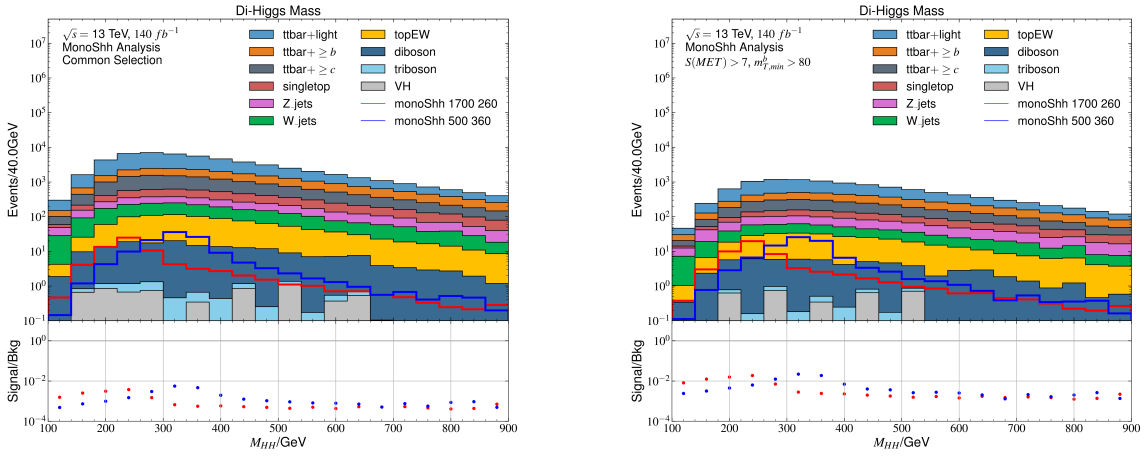
In the first step, jets are selected based on the number of b-jets and their mass. If $n_{b\text{-jets}} = 4$, all b-jets are selected. If there are more than four b-tagged jets, the four leading jets in p_{T} are used. For cases where $n_{b\text{-jets}} = 3$ and the mass of one of the jets exceeds 100 GeV, it is assumed to be a merged jet and used as a Higgs candidate. However, if none of the three jet masses exceeds this threshold, the untagged jet that minimises the mass of the heavier Higgs candidate is selected in addition to the tagged ones. In future efforts, a double-b tagger, such as the one introduced in Part IV, could be employed to enhance sensitivity to merged topologies.

In the second step of the reconstruction process, the jets are matched to form the Higgs candidates. In the case of a heavy merged jet, the other two b-tagged jets are matched

to form the second Higgs. For all other cases, the four selected jets are matched by minimising $\Delta R_{\max}(b,b)$, the maximum angular separation between the two jets matched to each Higgs:

$$\min \Delta R_{\max}(b,b) = \min(\max(\Delta R(h_1), \Delta R(h_2))) \ . \tag{6.2}$$

To not strongly sculpt the background distributions, no cuts are applied on the two Higgs mass distributions. The resolution of the reconstructed masses of the individual Higgs bosons is relatively poor and thus not suitable for a tight mass-window cut. Instead, the useful sensitivity in these variables is leveraged by the NN introduced in Chapter 7. However, the di-Higgs mass distribution is used as the key variable in this analysis, as it is utilised in the final data fits. Figure 6.4 shows the distributions of the reconstructed di-Higgs mass with and without the optimised selection applied. The signal is observed as a localised bump on much flatter SM background distributions, which all exhibit relatively similar broad shape. The signal peak is clearly visible for both shown benchmark signals with both the preselection and the optimised selection applied. As expected, the signal with the higher (lower) truth dark-Higgs mass peaks at higher (lower) m_{hh} values.



(a) Preselection. (b) Optimised Selection.

Figure 6.4: Reconstructed di-Higgs mass (m_{hh}) distributions of all relevant background processes and two representative signals with $m_{Z'} = 1700$ GeV and $m_s = 260$ GeV (red) and $m_{Z'} = 500$ GeV and $m_s = 360$ GeV (blue), are shown with preselection (a) and optimised selection (b) applied. The ratios of signal to total background event yields per bin are displayed on the lower axes.

Chapter 7

Machine Learning Analysis

Contents

7.1 Introduction 140

7.2 The Training Dataset 141

7.3 Correlation of m_{hh} and Output-Score 147

7.4 The Training of the Neural Network 153

This chapter discusses the use of MVA algorithms, particularly NNs, for further background rejection in the dark-Higgs analysis. The initial sections cover the training setup, followed by a discussion on the challenges of overcoming the sculpting of the background distribution of the fit variable m_{hh}. Finally, the results of the NN training are presented.

7.1 Introduction

While cuts are already applied to the most discriminating kinematic variables, valuable information is also encoded in many other variables and their correlations. To utilise this additional information, a machine learning classifier is trained and employed in addition to the cuts discussed in Chapter 6.

7.1.1 Signal Kinematics

In contrast to, for example, the Higgsino models (see Section 1.3.3), the kinematic behaviour across various mass points within the dark-Higgs models is relatively similar. To some extent this different behaviour can be attributed to the different decay modes.

In the Higgsino model's symmetrical decay topology, the two LSPs originate from separate Higgsino decays and have the potential to cancel each other out, resulting in low $E_{\mathrm{T}}^{\mathrm{miss}}$ values. This cancellation effect is particularly likely at lower Higgsino masses, as the decay products are more boosted and receive a weaker perpendicular kick relative to the direction of motion of the Higgsinos. Consequently, the two particles responsible for the $E_{\mathrm{T}}^{\mathrm{miss}}$ signature are likely to be back-to-back, leading to a $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution shifted to lower values.

Conversely, in the dark-Higgs scenario, such cancellation is not possible. The heavy Z' decays entirely into invisible particles, and thus the $E_{\mathrm{T}}^{\mathrm{miss}}$ is exclusively defined by the p_{T} of the Z', which is only marginally influenced by the masses involved.

Most of the important kinematic variables used in this analysis depend in some way on $E_{\mathrm{T}}^{\mathrm{miss}}$ and/or the angular separation of the b-jets, either among themselves or with the $E_{\mathrm{T}}^{\mathrm{miss}}$. These distributions are only weakly affected by the mass of the dark-Higgs and the Z'.

7.1.2 Training Strategy

Consequently, a more general approach than that used in the SUSY analysis is chosen, where a MVA is trained using simulated signal events from all mass points collectively.

The training set is selected by incorporating an equivalent quantity of signal events from each available mass point. This approach eliminates the need to train a separate classifier for each signal model, avoiding complications in interpreting intermediate hypotheses. Tests demonstrated no significant reduction in the performance to distinguish signal from background. Moreover, this strategy offers substantial advantages in terms of generalisability of results and streamlines the technical aspects of the analysis, particularly the fitting process. Most importantly, it removes the need to define unique SRs, CRs, and validation regions (ValRs) for each signal hypothesis individually.

7.2 The Training Dataset

When preparing the training set for the ML algorithms, two primary and somewhat competing factors must be considered. Firstly, the relative sizes of the input MC samples for each signal and background process. It is essential to provide the algorithm with a balanced representation of signals and backgrounds to ensure the network focuses equally on all areas of phase space. Secondly, the tightness of the preselection cuts applied to the samples before training. Providing a more signal-like selection at the point of training provides the algorithm with a more challenging yet more useful task to discriminate between signals and the more signal-like remaining background events.

7.2.1 The Right Balance

In preparing the training set, ensuring a balanced representation of signal and background events is crucial. Given the analysis is statistically constrained by the limited number of available background events, downsampling the signal datasets becomes necessary. Additionally, it is important to select an appropriate number of events from each background sample for training, as some background processes are overrepresented in our MC simulated data, particularly evident in the top+electroweak (topEW) samples. There are different strategies for downsampling individual background samples.

The first strategy involves selecting an equal number of events from each background

category, thereby enabling the NN to weight all background processes equally, irrespective of their actual impact on the analysis. The second strategy is to create a combined background training set that mirrors the composition of events passing the pre-selection and optimised selection criteria. This approach allows for focused training on more critical processes, thus including, for example, more $t\bar{t}$ and $Z+$jets events than diboson events. A third, more nuanced approach is to assign variable weights to different processes during training, giving higher weights to those with fewer events.

Given that the size of the background samples, excluding those from top+EW processes, reasonably approximates the relative significance of each background process in the analysis, the second strategy is chosen. This choice was substantiated by various checks confirming the NN's performance remains robust despite the unequal event representation from each background sample. Studies demonstrated the NN's proficiency in assigning low scores to all background processes, even those underrepresented in the training set.

Turning to the discussion of sample preselections: As the unweighted yield of signal events exceeds that of background events, the optimised selection is deliberately chosen to be relatively loose. As detailed in Section 6.3, it is possible to substantially reduce the background yield while preserving the majority of signal events. However, relaxing the selection criteria for the NN training expands our background dataset, which consequently also permits the inclusion of more signal events. A tighter selection, on the other hand, would result in training the network on background events that are more kinematically similar to the signal events, thereby challenging the network to discern subtle differences between the two. Generally, larger datasets can support more complex models without succumbing to overfitting (see Section 4.3) and can enhance the discriminative capacity of the network. However, caution is warranted since many background events introduced by loosening the selection may be kinematically distant from the signal, thus offering limited additional value.

To balance these competing ideas and determine the most advantageous strategy, multiple training iterations were conducted. Ultimately, a moderate cut of $m^{b}_{\mathrm{T},min} > 80$ GeV proved beneficial compared to using the full dataset, albeit modestly so. Table 7.1

shows the unweighted event yields per process used in the training set.

Process	Number of Selected Events
signal	430387
$t\bar{t}$	338703
$Z+$jets	17127
$W+$jets	4280
single-top	12597
top+EW	56491
di-boson	962
tri-boson	217
VH	10

Table 7.1: Number of selected events in the training set.

After defining the kinematic selection and the desired number of background and signal events for training, the events to be used for training are randomly selected from the given datasets. Finally, the background and signal events are shuffled randomly to ensure a smooth training procedure.

7.2.2 Cross-Validation

The training dataset is split into k equally sized sets to utilise k-fold cross-validation. This technique allows all the available data to be used in the analysis without ever applying the network to events on which it was trained. For a given 'fold', a separate algorithm is applied to $1/k$ of the events in the final analysis and trained on the remaining $(k-1)/k$ events. Overall, k algorithms were trained, which studies confirmed perform very similarly. For this analysis, $k = 5$ was a reasonable choice to maximise training statistics while ensuring confidence in the validation.

7.2.3 Input Variables

For training the NN, a broad spectrum of kinematic variables was considered, as listed in Table 7.2. This approach allows the network to access as much information about the

data as possible. Initially, variables with clear discrimination potential were identified by comparing signal and background distributions. Additionally, extra variables were retained, if they could provide additional discriminative power through correlations when combined with other variables. Studies of correlations between potential input variables were conducted. In cases of significant differences in correlation between signal and background samples, additional discrimination power can be expected. One illustrative case is the correlation between $S(E_{\mathrm{T}}^{\mathrm{miss}})$ and H_{T}, which are strongly correlated for signal events but show almost no correlation for background events. This allows the NN to enhance its discriminative power by using both variables as inputs, even though, for example, H_{T} does not exhibit significant discriminative power on its own.

Furthermore, the impact of each variable on the output-score was assessed using Shapley values (see Section 4.4). Figure 7.1 shows the Shapley value distribution of 100 randomly selected events in the training set. Variables with low impact were excluded from our model, while those with high impact were retained. As expected and described in Section 6.3.1, $m_{\mathrm{T},min}^{b}$ is the highest-ranked variable, followed by $S(E_{\mathrm{T}}^{\mathrm{miss}})$ and variables that rely on the angular separation of the b-jets, such as $\Delta\eta_{\mathrm{max}}(b,b)$.

Additionally, the Python package *featuretools* [198] was used to generate new variables by applying the four basic arithmetic operations—multiplication, division, addition, and subtraction—to the existing variables. These artificially created variables were then evaluated based on their Shapley values, and those showing promise were incorporated into the training. Given the moderate size of the training samples, these pre-combined variables can speed up the training process and allow the neural network to learn finer details of the datasets. The improvement in discrimination gained from the additional variables depends on the size of the training set and the training set selection. Detailed studies showed a significant improvement for tight selections, whereas no improvement was observed for loose selections. For the relatively loose selection chosen in the final analysis strategy, a modest performance increase was observed.

Variable Name	**Description**
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$m_{\mathrm{T},min}^{b}$	Minimal transverse mass of $E_{\mathrm{T}}^{\mathrm{miss}}$ and one b-jet
$S(E_{\mathrm{T}}^{\mathrm{miss}})$	Object based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance
p_{T}^{4b}	p_{T} of the di-Higgs system
m_{T}^{4b}	Transverse mass of di-Higgs system and $E_{\mathrm{T}}^{\mathrm{miss}}$
n_{jets}	Number of jets
$n_{b\text{-jets}}$	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
X_{Wt}	Variable for top-rejection
$\Delta\eta_{\mathrm{max}}(b, b)$	Maximal $\Delta\eta$ between b-jets
Combined Variables	**Description**
$E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{H_{\mathrm{T}}}$	Estimate of $S(E_{\mathrm{T}}^{\mathrm{miss}})$
$E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{4b}$	$E_{\mathrm{T}}^{\mathrm{miss}}$ divided by p_{T}^{4b}
$N_{\mathrm{jet}}/n_{b\text{-jets}}$	Ratio of jets and b-jet counts
$n_{\mathrm{jets}} - n_{b\text{-jets}}$	Number of untagged jets
$p_{\mathrm{T}}^{b1}/n_{\mathrm{jets}}$	Ratio of p_{T} of leading jet and number of jets
$S(E_{\mathrm{T}}^{\mathrm{miss}})/n_{b\text{-jets}}$	Ratio of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance and number of b-jets
$S(E_{\mathrm{T}}^{\mathrm{miss}}) \times p_{\mathrm{T}}^{4b}$	$E_{\mathrm{T}}^{\mathrm{miss}}$ 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 $E_{\mathrm{T}}^{\mathrm{miss}}$
$\Delta\eta_{\mathrm{RMS}}(b, b) \times \Delta\eta_{\mathrm{max}}(b, b)$	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

Table 7.2: The full set of NN input variables. The second part of the table lists combinations of variables which lead to improved signal/background discrimination.

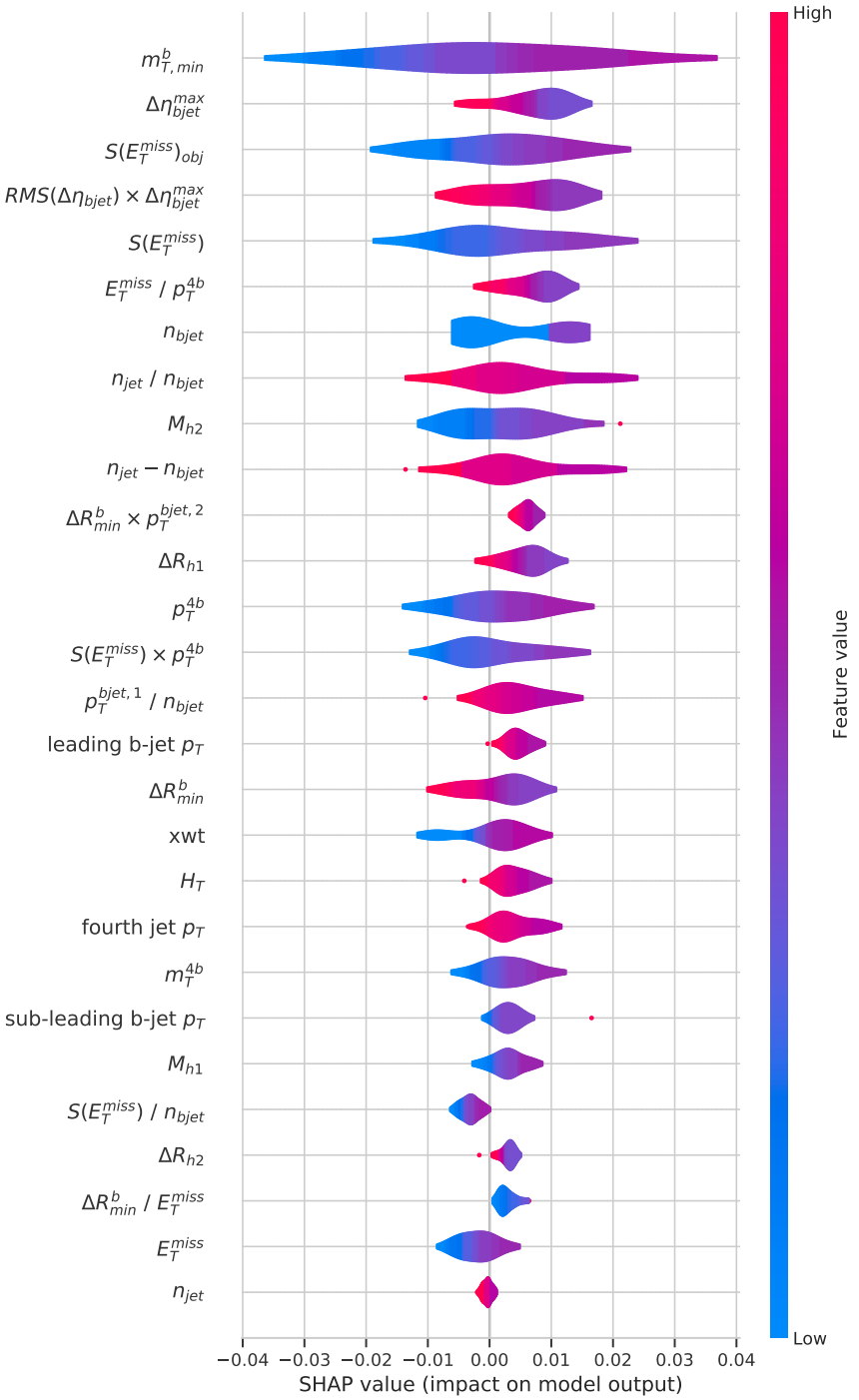


Figure 7.1: Shapley values of individual neural network input variables. Variables higher up in the plot have a higher impact on the output-score, while those further down have a lower impact.

7.2.4 Preprocessing of Training Data

Standard preprocessing steps were applied to the training data. This involves normalising the variables by subtracting the sample mean and dividing by the sample standard deviation. Additionally, a 'Box-Cox' transformation was applied after mean-scaling to adjust the input distributions towards more Gaussian-like profiles, further fine-tuning the NN's performance. The Box-Cox transformation is defined as:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \text{for } \lambda \neq 0, \\ \log(y) & \text{for } \lambda = 0. \end{cases} \tag{7.1}$$

The optimal value of λ is computed via the python package SciPy and its statistics sub-module [199]. Studies demonstrated a modest performance increase resulting from the application of the Box-Cox transformation.

7.3 Correlation of m_{hh} and Output-Score

The analysis strategy of this search entails fitting the di-Higgs mass distribution after all cuts are placed on kinematic variables (see Table 6.2) and the MVA output-score.

7.3.1 The Issue: Sculpting

An optimised MVA may be capable of indirectly extracting information regarding the di-Higgs mass through its correlation with other input variables, even if the di-Higgs mass m_{hh} is not explicitly used as an input variable. The signal hypotheses cover a di-Higgs mass range that is relatively narrow, extending from 250 GeV to just above 400 GeV. By contrast, the background processes show a considerably wider distribution in m_{hh}. This disparity poses a risk of sculpting the m_{hh} distribution in our background samples when a cut is applied to the MVA output-score. Thus, the idea of fitting a physically interpretable distribution would be compromised.

Two different strategies were tested to overcome this issue, which also led to the

decision to use a NN instead of a BDT, to provide more flexibility. The first approach, involving the training of an adversarial network, is elegant but very challenging in its practical implementation. The second approach includes the addition of a dedicated decorrelation term to the loss function minimised during the training.

7.3.2 First Approach: Adversarial Network

The first option is to train an adversarial network composed of three distinct components. The first component, termed the *Feature Extractor*, is designed to transform the input features (the kinematic variables) into a representation that maximises information about whether an event is categorised as signal or background. Simultaneously, it aims to minimise information about the reconstructed m_{hh}. The second component, the *Classifier*, is tasked with identifying events as either signal or background. The third component, the *Protector*, aims to decouple the *Classifier's* output from the m_{hh} distribution. In essence, it "protects" m_{hh} from being learned by the *Classifier*, which it, counterintuitively, achieves by attempting to predict m_{hh} itself.

Each component of the adversarial network is structured as a feedforward network, comprising multiple fully-connected layers. The output of the *Feature Extractor* is passed to both the *Classifier* and the *Protector*. See Figure 7.2 for a sketch of the structure of this network.

Training of an Adversarial Network

The main idea to train this network lies in inverting the gradients during the backward pass (see Section 4.2.3) between the *Protector* and the *Feature Extractor*. This is accomplished through a customly implemented gradient-reversal layer in the Python-based Pytorch library. Consequently, a competitive dynamic emerges between the *Protector* and the *Feature Extractor*. The latter aims to transmit minimal information about m_{hh} to the *Protector*, which in turn is trained to predict m_{hh}. Meanwhile, the *Classifier* is trained with standard training procedures.

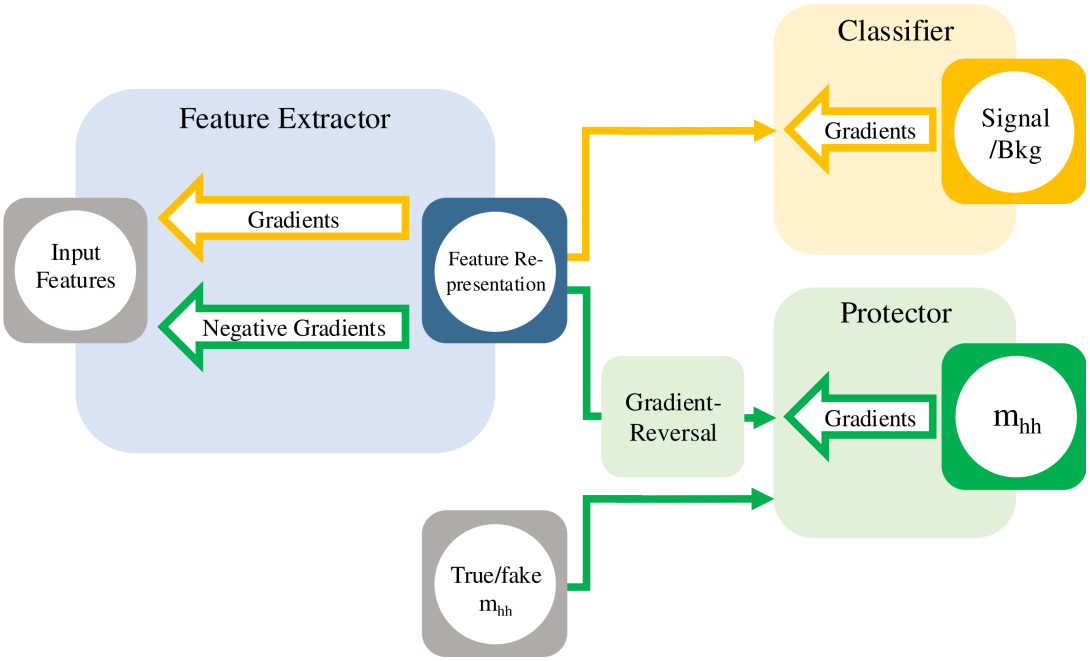


Figure 7.2: Sketch of the architecture of the adversarial network tested to overcome the issue of sculpting the background di-Higgs mass distribution. The three large areas in light colours represent the feedforward NNs.

Challenges

This complex training methodology introduces several challenges. Achieving an equilibrium between the *Feature Extractor* and the *Protector* is difficult, as there's a risk of one overpowering the other. Furthermore, the gradient updates from the *Classifier* can complicate this balancing act.

Another challenge is managing the simultaneous minimisation of two distinct loss functions. For the task of signal/background classification, we utilise the conventional BCE loss, defined as:

$$L_{\text{BCE}}(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \ . \tag{7.2}$$

Here, N represents the number of events (e.g. in a mini-batch), y_i the true label for the i-th event (0 for background and 1 for signal), and $\hat{y}_i$ is the predicted probability for the i-th event to be a signal event. The BCE loss is a common choice for a classification model whose output is a probability value between 0 and 1.

For the prediction of a continuous variable such as m_{hh}, a common choice for the loss

function is the MSE, defined as:

$$L_{\text{MSE}}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{7.3}$$

A challenge in the simultaneous use of the MSE and BCE loss functions is the typically larger values of MSE. This disparity can lead to the gradients from the *Protector* disproportionately influencing the *Feature Extractor's* training. To address this, experiments with scaling the loss functions to achieve balanced magnitudes were conducted. However, it proved difficult to completely remove this effect, as the value of the loss evolves during training.

Thus, alternative methods were tested. One such approach involved presenting the *Protector* with an m_{hh} either the correct one, or one of a different event in the mini-batch. The protector then has to classify the value of m_{hh} as either true or false, thus allowing the prediction of a continuous variable to reduce to a classification problem and subsequently allowing a BCE loss to be used. It is ensured that the fake m_{hh} value has a minimal difference to the true value to account for resolution issues.

Multiple other strategies were tested to increase the balance of the adversarial training, as achieving good convergence was difficult. These strategies included training the *Classifier* for a few epochs with a relatively low learning rate first before adding the *Protector*, as it was found that starting the optimisation of both losses simultaneously did not lead to convergence. Furthermore, a fine-tuning step was added at the very end of the training, where again only the *Classifier* was trained with a very low learning rate, to achieve slightly better discrimination results without altering the decorrelation of the m_{hh} distribution and the NN output-score.

Results

Ultimately, this approach yielded modest results, achieving good classification but only mediocre decorrelation. Time constraints necessitated the exploration of other strategies, as optimising the numerous parameters proved too time-intensive. Nonetheless, this elegant approach remains interesting for future exploration, potentially yielding more

insights with further investigation.

7.3.3 Second Approach: Direct Decorrelation

The second approach aims to directly decorrelate the m_{hh} and NN output-score distributions by adding an additional term to the loss function. This method is easier to implement, allowing a simple feed-forward NN architecture with multiple layers to be used.

To overcome the issue of sculpting the di-Higgs mass, the correlation coefficient can be added to the loss function. Thus, a linear combination of a BCE loss and a correlation coefficient is utilised. Two different expressions were tested for the correlation term: the Pearson correlation coefficient and the distance correlation.

Pearson Correlation

The Pearson correlation coefficient measures the linear correlation between two variables X and Y. It has a value between +1 and -1, where 1 indicates total positive linear correlation, 0 indicates no linear correlation, and -1 indicates total negative linear correlation. The coefficient is defined as:

$$r(X,Y) = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \,, \tag{7.4}$$

where $\overline{X}$ and $\overline{Y}$ are the means of the X and Y variables, respectively, and N is the number of events. In this analysis, the variables X and Y are given by the NN output-score and m_{hh}.

Distance Correlation

In contrast to the Pearson correlation, the distance correlation can detect both linear and non-linear associations. The distance correlation between two random vectors X and Y is defined as:

$$\mathcal{R}(X,Y) = \frac{\mathcal{V}(X,Y)}{\sqrt{\mathcal{V}(X,X)\mathcal{V}(Y,Y)}} \tag{7.5}$$

where $\mathcal{V}(X,Y)$ is the distance covariance between X and Y, and $\mathcal{V}(X,X)$ and $\mathcal{V}(Y,Y)$ are the distance variances of X and Y, respectively.

Given two random vectors X and Y with respective elements X_j and Y_k, the distance covariance is defined as:

$$\mathcal{V}^2(X,Y) = \frac{1}{n^2}\sum_{j,k} a_{jk}b_{jk} \tag{7.6}$$

where a_{jk} and b_{jk} are the elements of the double-centred distance matrices for X and Y, respectively. These matrices are computed as follows:

$$a_{jk} = |X_j - X_k| - \overline{|X_j - X|}_k - \overline{|X_k - X|}_j + \overline{|X_j - X_k|}, \tag{7.7}$$

$$b_{jk} = |Y_j - Y_k| - \overline{|Y_j - Y|}_k - \overline{|Y_k - Y|}_j + \overline{|Y_j - Y_k|}. \tag{7.8}$$

Here, $|X_j - X_k|$ is the distance between the j-th and k-th elements of X, and $\overline{|X_j - X|}_k$ is the average distance of the j-th element of X from all elements of X, averaged over k.

Combined Loss Function

Both additions to the loss function - the Pearson correlation and the distance correlation - have been tested. The differences between the results with the two approaches are negligible. The distance correlation was chosen as it provides additional protection from non-linear correlation and was used in similar scenarios before, see Ref. [200].

The BCE and distance correlation are combined linearly as:

$$L_{\text{total}} = \alpha L_{BCE}(Y,\hat{Y}) + (1-\alpha)\mathcal{R}(\hat{Y}, m_{hh}), \tag{7.9}$$

where α is a scaling parameter ranging from 0 to 1. Various experiments to determine the optimal value of α were conducted, considering both static and dynamically adjusted values during the training process. The aim was to maintain a balance between the two loss functions throughout the training. Ultimately, a static value of $\alpha = 0.7$ was chosen, which proved particularly beneficial in the initial epochs, which are crucial for the overall effectiveness of the training process.

Practical Implementation

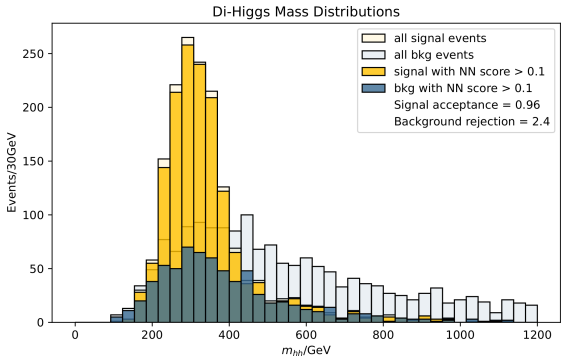
The correlation is evaluated mini-batch wise and separately for signal and background. While the ideal approach would be to decorrelate each background process individually, this method necessitates considerably larger batch sizes, which would negatively impact training and classification performance. As no large discrepancies in the m_{hh} distributions of the SM background processes are observed, treating the backgrounds as a single sample for decorrelation purposes is justifiable. Tests to decorrelate at least the large backgrounds separately were conducted but did not yield significant improvements.

In Figure 7.3, a side-by-side comparison of the training outcomes with and without incorporating the correlation term in the loss function is shown. The addition of the correlation term results in a substantial reduction in the sculpting of the background distribution, allowing subsequent use of m_{hh} as a fit variable. Notably, the decorrelation does not lead to any apparent decrease in overall performance. However, it does influence the output-score distribution, leading to a flatter curve for the SM background processes. At very low output-scores (see Figure 7.3(b)), the background rejection is significantly lower, as the model is restricted from assigning extremely low scores to events in the tails of the m_{hh} distribution. This effect is balanced by better performance at higher output-scores. In the critical region for defining a signal region (output-score > 0.9), the inclusion of the correlation term results in enhanced background rejection, albeit with a slight reduction in signal acceptance (see Figure 7.3(d)). Overall, the performance remains comparable to the original model, affirming that the additional decorrelation does not compromise the classifier's effectiveness.

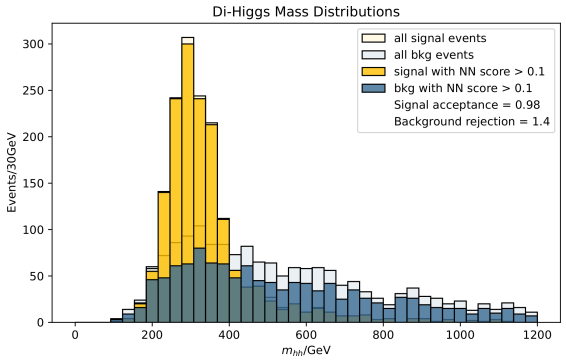
7.4 The Training of the Neural Network

7.4.1 Hyperparameters

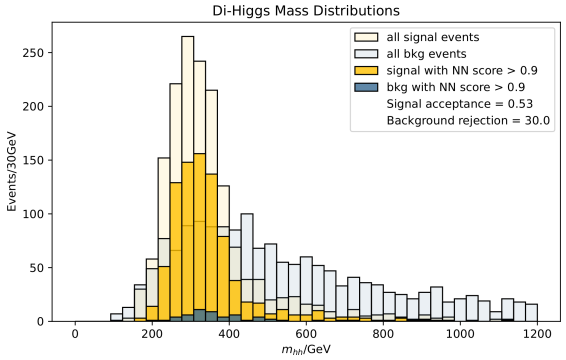
The size of the mini-batches plays a crucial role in ensuring good training performance. Ideally, a single batch encompassing the entire dataset can be used to achieve optimal



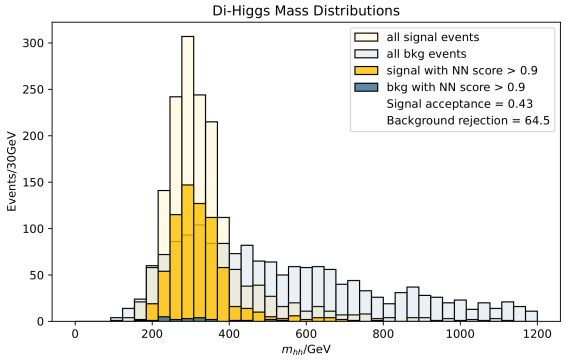
(a) Classifier only: NN-score > 0.1.



(b) With decorrelation applied: NN-score > 0.1.



(c) Classifier only: NN-score > 0.9.



(d) With decorellation applied: NN-score > 0.9.

Figure 7.3: Unweighted m_{hh} distributions for a training with (right) and without (left) decorrelation. The histograms show the distributions after applying NN output-score cuts of > 0.1 (top) and > 0.9 (bottom), alongside the original distributions with only the training selection applied. The depicted signal includes events from multiple models, with m_{hh} values ranging from 260 GeV to 385 GeV.

decorrelation. However, a compromise to use smaller batch sizes is made for better classification performance. The learning rate is a significant factor that interacts with batch size. Extensive testing with various combinations of batch sizes and learning rates was conducted. The most effective results were obtained using a batch size of 4096 events coupled with an initial learning rate of 5×10^{-6}.

A learning rate scheduler halving the learning rate every five epochs was implemented. The training extends over a span of 25 epochs, and the epoch that exhibits the lowest total loss is selected for the final analysis.

The network architecture includes eight fully-connected hidden layers, featuring a progressively decreasing number of nodes. The initial hidden layer contains 1024 nodes, and this number is halved in each subsequent layer. To increase the training stability, a

batch normalisation layer is incorporated after each hidden layer. The ReLU activation function is employed consistently across the network, with the exception of the output layer, which utilises a sigmoid function.

7.4.2 Regularisation

To minimise the risk of overfitting, several generalisation measures (see Section 4.3.3) have been implemented. Firstly, the number of layers and nodes per layer were optimised to be as compact as possible, without compromising classification performance. Secondly, dropout layers with $p = 0.2$ are incorporated after the non-linearity in each hidden layer, preventing over-reliance on any single node. Finally, to constrain the magnitude of individual weights, an $L2$-norm weight decay is integrated into the loss function. For optimisation purposes, the stochastic Adam algorithm [201] is utilised.

Moreover, as detailed in Section 7.2.2, five-fold cross-validation is applied in the training process. Studies were performed to ensure similar behaviour over all five trainings and that the output-score distributions were concordant. When evaluating the model on physics data, each of these five trained parameter sets is applied to a separate fifth of the dataset. This method ensures that any minor variations between the five training sessions are effectively averaged out, enhancing the robustness of our evaluation.

7.4.3 The Neural Network Output-Scores

Trainings were conducted using two approaches: a combined training on the entire dataset and separate trainings for $n_{b\text{-jets}} = 3$ and $n_{b\text{-jets}} \geq 4$ categories. The results show that the combined training yields superior performance, likely due to the increased dataset statistics. The output-scores obtained from this method are shown in Figure 7.4, and subsequently used for defining signal, validation, and control regions.

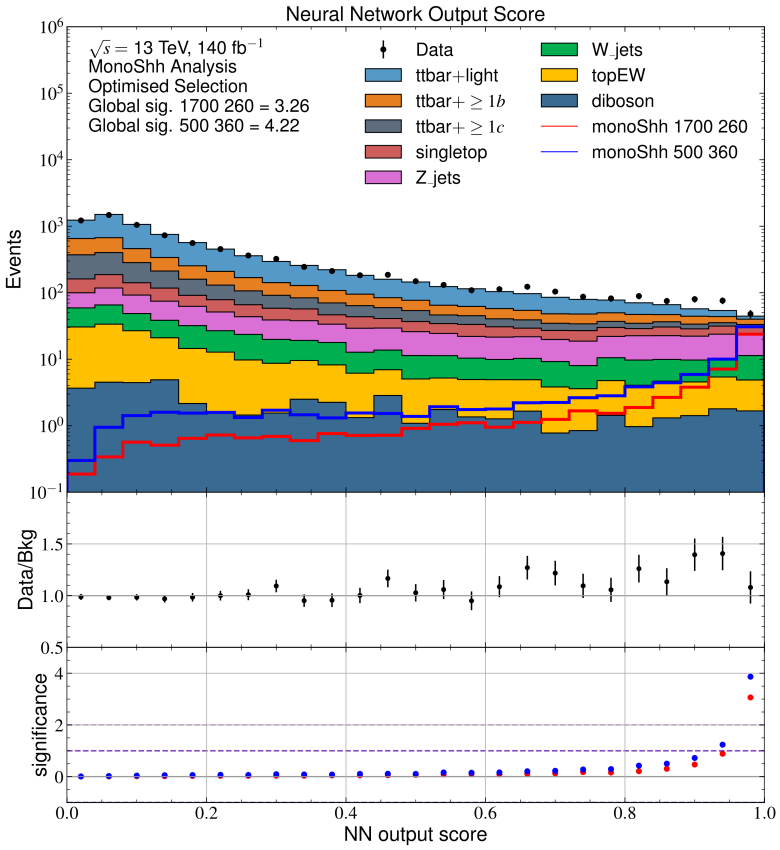


Figure 7.4: Neural network output score for various SM background processes and two signal models.

Chapter 8

Analysis Regions and Background Estimation

Contents

8.1 Signal Regions 158

8.2 Background Estimation Strategy 159

8.3 Systematic Uncertainties 165

This chapter details the analysis regions utilised in the dark-Higgs search. The CRs and ValR used for background estimation, as well as model-dependent and model-independent SRs are introduced. The last section introduces the systematic uncertainties affecting the estimation of background and signal events.

8.1 Signal Regions

8.1.1 Model-Dependent Exclusion Signal Region

For the purpose of conducting exclusion tests (see Section 3.2.2), a SR with a high concentration of signal events and low background contamination is defined. The analysis strategy is based on fitting the reconstructed di-Higgs mass distribution. This approach has the advantage of preserving a single physically meaningful distribution to ease interpretability of results, e.g. if evidence of an excess is observed, in comparison to, for example, fitting many ML discriminant distributions directly. The SR selection is defined to maximise the overall significance, obtained by adding the bin-wise significances in quadrature, estimated as outlined in Ref. [202]. In contrast to the CRs, see Section 8.2, the SR is binned in m_{hh}. Studies indicate that a NN output-score cut of ≥ 0.95, in addition to the selection outlined in Table 6.2, yields the highest overall significance. Due to the kinematic behaviour discussed in Section 7.1.1, the optimal cut value is relatively independent of the specific signal mass point, and thus only a single SR is defined.

While large event numbers per bin can be beneficial for maximising the significance, it is essential to retain the resolution of the di-Higgs mass peak. Therefore, the choice of bin width is crucial for this analysis. In the central bins, a bin width of around 50 GeV in m_{hh} is used to capture the achieved mass peak resolution, as shown in Figure 8.2(a). In the tails, where the expected event counts are much lower, wider bins are used. A range of 100-900 GeV is considered for the m_{hh} distribution. The signal and background yields outside this range are negligible. The prefit yields of the various processes contributing to the SR are listed in Table 8.4 and are further discussed in Section 8.2.

Further subcategorisation of the SR based on the number of b-jets ($n_{b\text{-jets}} = 3$ and $n_{b\text{-jets}} \geq 4$) was also explored. After applying only the kinematic selections, the sensitivity of the analysis appears to be almost exclusively contained in the higher b-jet category. However, this advantage dissipates once the NN output-score cut is applied. The NN is particularly effective in discriminating between background and signal in scenarios with

lower b-jet multiplicities. Consequently, both categories achieve similar sensitivities at similar NN output-score cut values. Therefore, dividing the SR into b-jet categories did not enhance the analysis's sensitivity, and ultimately a single inclusive SR is utilised.

8.1.2 Model-Independent Discovery Regions

For the purpose of discovery tests (see Section 3.2.2) on unspecified BSM physics within the SR phase space, a set of single-bin regions in m_{hh} are defined. Seven such discovery regions (DRs) are introduced to comprehensively cover the mass range, while avoiding overly low statistics. Table 8.1 lists the respective m_{hh} intervals for the DRs.

Region	DRa	DRb	DRc	DRd	DRe	DRf	DRg
m_{hh} **in GeV**	$[100, 250)$	$[200, 300)$	$[250, 350)$	$[300, 400)$	$[350, 450)$	$[400, 900)$	$[500, 900)$

Table 8.1: Model-independent DR m_{hh} intervals.

8.2 Background Estimation Strategy

This analysis relies on fitting MC simulated signals and SM background estimates to data. Therefore, an accurate estimation of the SM backgrounds in the SR is crucial. To achieve this, dedicated CRs, rich in specific background sources, are implemented. By ensuring that the events in the CRs are kinematically similar to those in the SR, corrections derived from the data in the CRs can be applied to refine the MC-based background estimates in the SR. Overall, the shape of the background estimation is constrained by the uncertainties on the MC simulation (see Section 8.3), but the normalisations are derived from the data.

As multijet processes are notoriously difficult to model with MC simulations, it is common practice to resort to data-driven methods in this case. In this analysis, however, the cuts applied on $\Delta\phi^{4j}_{\text{min}}$ and the object-based $E_{\text{T}}^{\text{miss}}$ significance have been highly effective in suppressing this background, as discussed in Section 6.3.2. Therefore, no dedicated multijet background estimation is required.

The other SM processes can be simulated using MC and only require corrections derived from auxiliary measurements in regions kinematically similar to the SR. As detailed in

Table 8.4, after applying the SR selection criteria, the two predominant background processes are $t\bar{t}$ and $Z+$jets. In the case of $Z+$jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ is produced when the Z boson decays to neutrinos, with the $E_{\mathrm{T}}^{\mathrm{miss}}$ recoiling against one or more hadronic jets. Multiple b-jets can be generated through gluon splitting, additional radiation, or mis-tagged light-flavour jets. For selected $t\bar{t}$ events, $E_{\mathrm{T}}^{\mathrm{miss}}$ primarily arises from the leptonic decays of W bosons, with two b-quarks produced from the top decays. Additional b-jets can come from heavy-flavour radiation or mis-tagged light jets, contributing to a high b-jet multiplicity. To enhance the precision of the background estimates for these processes, the analysis incorporates freely floating normalisation factors within the fitting procedure for $Z+$jets, $t\bar{t}+>1b$, and $t\bar{t}+$light processes, as detailed in Chapter 9.

While the SR comprises multiple bins in m_{hh}, the CRs consist of a single bin each. The following sections detail the selections for the regions utilised for background estimation and its validation.

8.2.1 $Z+$jets Control Region

To define a selection for a dedicated $Z+$jets CR without altering the kinematics of the process compared to the SR, while significantly suppressing all other processes in the phase space, $Z \to \nu\nu$ events are replaced by $Z \to \mu\mu$ events. This is achieved by lifting the lepton veto in the preselection and instead requiring two muons of opposite sign. To maintain similar kinematics, the SR selection is mimicked by adding the p_{T} of the muons to the $E_{\mathrm{T}}^{\mathrm{miss}}$ to simulate the neutrinos in the $Z \to \nu\nu$ decays. Events in this CR must have >200 GeV of this muon-corrected missing transverse energy $E_T^{miss-\mu}$, while the genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is adjusted to <75 GeV. Furthermore, the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is estimated as $E_T^{miss-\mu}/\sqrt{H_{\mathrm{T}}}$ and, in agreement with the optimised selection, required to be >7. Similarly, $m_{\mathrm{T},min}^{b}$ is redefined using $E_T^{miss-\mu}$, and the standard selection cut >80 GeV is applied. The rest of the kinematic selection requirements remain unchanged. Additionally, a m_Z-window cut is applied to further minimise contributions from other processes. With this set of cuts, a purity of approximately 80% for $Z+$jets events is achieved in this CR with moderate statistics. As shown in Figure 8.2(b), the di-Higgs

mass distribution of the background estimation shows decent shape agreement. However, given the slight underestimation of the observed overall data yields, the floating $Z+$jets normalisation nuisance parameter is expected to exhibit an upward pull of approximately 1.2. See Table 8.2 for a summary of the selection, and Table 8.4 for the process-specific composition of the event yield in the $Z+$jets CR.

Z+jets CR	
Variable	**Cut Value**
Preselection	Pass (except n_{leptons} veto)
n_μ	2
$\lvert m_{\mu\mu} - m_Z \rvert$ [GeV]	< 15
$E_{\text{T}}^{\text{miss}}$ [GeV]	< 75
$E_{\text{T}}^{\text{miss}-\mu}$ [GeV]	> 200
$E_{\text{T}}^{\text{miss}-\mu}/\sqrt{H_{\text{T}}}$	> 7
$m_{\text{T,ZCR}}^{b,min}$ [GeV]	> 80

Table 8.2: Cut selection defining the $Z+$jets CR. The last three cuts mirror the selection in the SR.

8.2.2 Control Regions for Top-Quark Processes

To select regions with a high purity of $t\bar{t}$ processes, regions complementary to the SR selection are used. The main sub-processes are $t\bar{t}+$ light jets and $t\bar{t}+b$-jets, which peak at relatively low and high $m_{\text{T},min}^{b}$, respectively. For this reason, a cut at $m_{\text{T},min}^{b} = 200$ GeV is used to define two separate CRs for these processes. While the low-$m_{\text{T},min}^{b}$ region shows high purity in $t\bar{t}+$ light jets, the high-$m_{\text{T},min}^{b}$ region includes, besides $t\bar{t}+b$-jets, significant contributions from various other processes, mostly single-top and $Z+$jets. Whereas single-top processes dominate in the $n_{b\text{-jets}} = 3$ case, $t\bar{t}+b$-jets lead to higher b-jet multiplicities. Consequently, the regions are further divided into separate CRs, namely for $n_{b\text{-jets}} = 3$ and $n_{b\text{-jets}} \geq 4$. This strategy results in three CRs: the $3b$ low-$m_{\text{T},min}^{b}$ CR enriched in $t\bar{t}+$ light jets, the $4b$ CR targeting $t\bar{t}+b$-jets, and a $3b$ high-$m_{\text{T},min}^{b}$ CR with roughly equal contributions from single-top, $t\bar{t}+b$-jets, and $Z+$jets.

The NN output-score is used to make the CRs orthogonal to the SR and the ValR. A specific range of the NN output-score is selected for defining the CRs, with the upper boundary aligning with the ValR. The lower boundary is chosen to balance the available statistics and the exclusion of background events that exhibit markedly different kinematic characteristics from those in the SR. Figure 8.1 illustrates the definitions of the three CRs, the ValR, and SR in terms of the key variables $m^{b}_{\mathrm{T},min}$, $n_{b\text{-jets}}$, and NN output-score.

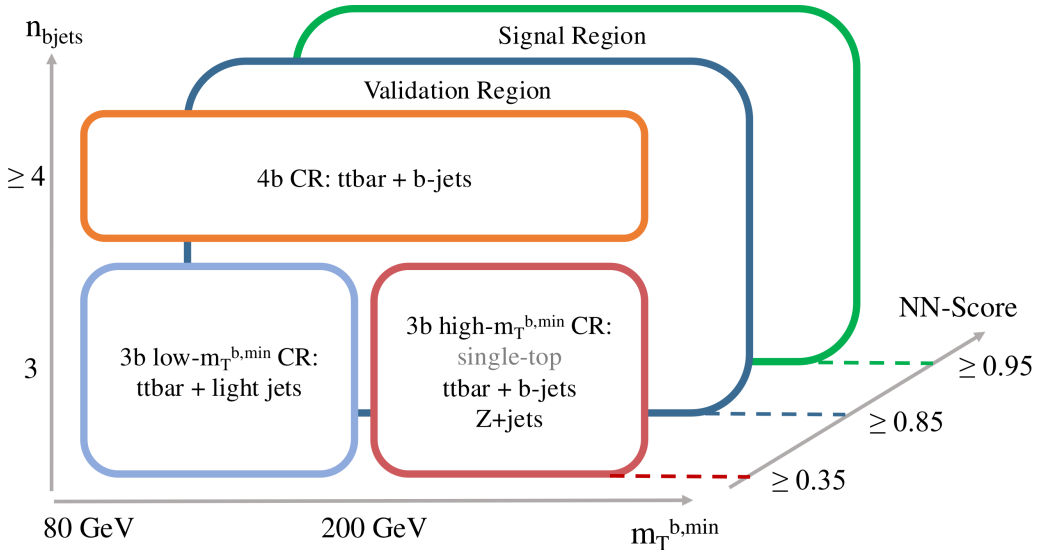


Figure 8.1: Schematic overview of the definitions of the SR, the ValR, and the CRs (with the exception of the Z+jets CR) in terms of b-jet multiplicity, $m^{b}_{\mathrm{T},min}$, and NN output-score.

In the $3b$ low-$m^{b}_{\mathrm{T},min}$CR, a purity of approximately 57% for $t\bar{t}$ + light jets is achieved, with the majority of the remaining background yield coming from other top subprocesses. In the $3b$ high-$m^{b}_{\mathrm{T},min}$CR, the contributions from $Z+$jets, $t\bar{t}$ + b-jets, and single-top processes amount to approximately 25%, 16%, and 20%, respectively. The purity of $t\bar{t}$ + b-jets in the $4b$ CR is around 45%. See Table 8.3 for a summary of the selections. Furthermore, see Figures 8.2(c), 8.2(d), and 8.2(e) for the prefit di-Higgs mass distributions in the $t\bar{t}$ and single-top CRs and Table 8.4 for the prefit yields. Given the slight underestimation of the background yield in the $4b$ CR, an increased normalisation of the $t\bar{t}$ + b-jets processes is anticipated. This, along with the expected upward pull of the $Z+$jets normalisation, aligns with the observed mismodelling in the $3b$ high-$m^{b}_{\mathrm{T},min}$CR. Conversely, no significant normalisation changes are expected for the $t\bar{t}$ + light jets processes, due to the excellent prefit modelling observed in the $3b$ low-$m^{b}_{\mathrm{T},min}$CR.

Crontrol Regions:	**3b low-$m^b_{\mathrm{T},min}$**	**3b high-$m^b_{\mathrm{T},min}$**	**4b**
Variable	**Cut Value**		
Preselection	Pass		
NN_{score}	$(0.35, 0.85)$		
$n_{b\text{-jets}}$	$== 3$	$== 3$	$== 4$
$m^b_{\mathrm{T},min}$ [GeV]	$(80, 200)$	> 200	> 80

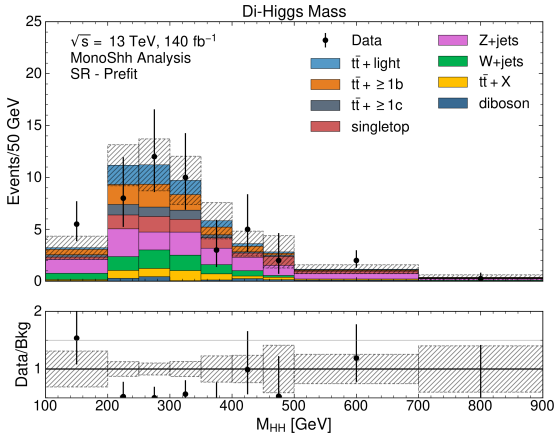
Table 8.3: Cut selections defining the single-top and $t\bar{t}$ CRs.

8.2.3 Validation Region

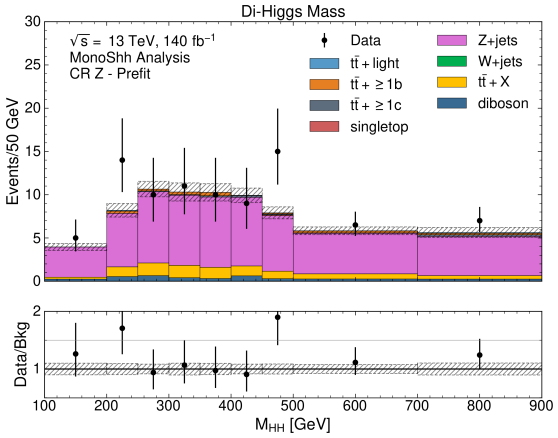
To test the accuracy of the background estimation near the SR, a dedicated ValR is established. It populates the region $0.85 < \mathrm{NN_{score}} < 0.95$, situated between the CRs and the SR. See Figure 8.2(f) for the prefit di-Higgs mass distribution in the ValR, and Table 8.4 for the prefit contributions of the individual processes to the ValR. In line with the anticipated increased postfit normalisations of the $t\bar{t}$ + b-jets and Z + jets processes, an underestimation of the data yields are observed in the ValR before the fit.

Process	**Control Regions**				**Validation Region**	**Signal Region**
	Z+jets	**3b low-$m^b_{\mathrm{T},min}$**	**3b high-$m^b_{\mathrm{T},min}$**	**4b**		
signal	0.0	8.6	4.4	1.5	10.8	25.6
data	130	1162	262	216	200	60
total background	111.4	1137.5	209.2	191.7	146.7	57.0
$t\bar{t}$ + light	0.6	651.3	19.0	31.2	35.6	6.8
$t\bar{t}$ + $\geq$ 1 b-jet	2.8	128.6	34.4	86.1	24.8	8.6
$t\bar{t}$ + $\geq$ 1 c-jet	1.6	116.7	13.3	27.6	12.3	4.0
Z + jets	88.6	97.5	52.4	13.4	31.7	15.9
W+ jets	0.0	45.2	24.3	6.0	11.8	7.7
single-top	0.1	68.5	41.4	13.3	18.0	7.8
top+EW	11.7	19.3	16.5	11.8	8.1	4.3
di-boson	5.4	9.3	6.7	2.2	3.7	1.9
tri-boson	0.3	0.4	0.4	0.1	0.1	0.2
VH	0.4	0.6	0.7	0.0	0.5	0.0

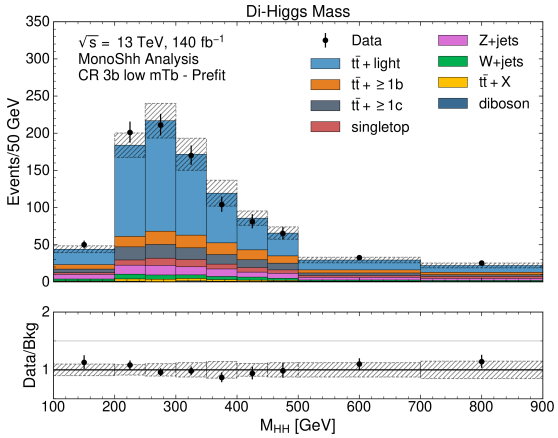
Table 8.4: Expected event yields per process in all analysis regions. The reference signal point used in this table has the masses: $M_{Z'} = 1700$ and $M_s = 260$ GeV.



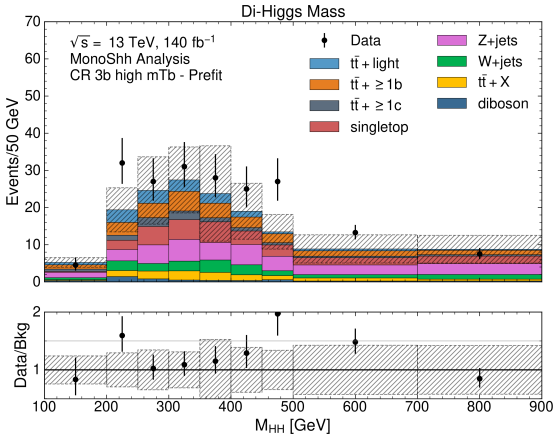
(a) Signal region.



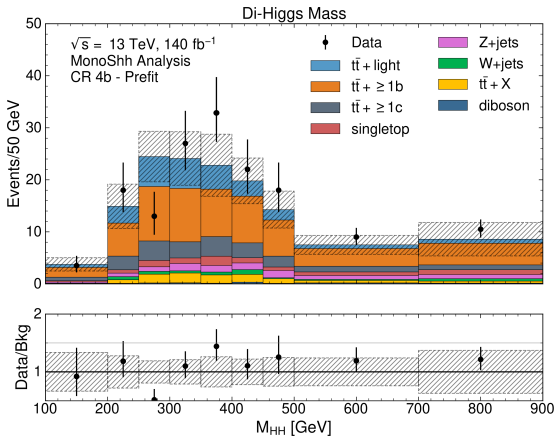
(b) $Z+$jets CR.



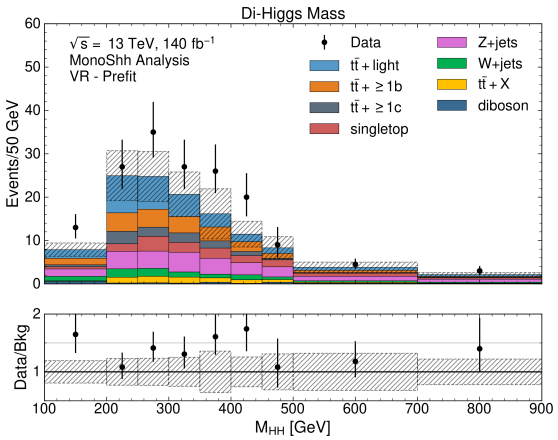
(c) $3b$ low-$m^b_{\mathrm{T},min}$CR.



(d) $3b$ high-$m^b_{\mathrm{T},min}$CR.



(e) $4b$ CR.



(f) Validation region.

Figure 8.2: The prefit di-Higgs mass distributions in the SR (a), the $Z+$jets (b), $3b$ low-$m^b_{\mathrm{T},min}$(c), $3b$ high-$m^b_{\mathrm{T},min}$ (d), $4b$ (e) CRs, and the ValR (f). While all these histograms display multiple bins in m_{hh}, the final CRs utilised in the fit comprise only a single bin.

8.3 Systematic Uncertainties

The estimation of background and signal yields in the analysis regions is subject to both statistical and systematic uncertainties. The latter are categorised into experimental uncertainties, which account for calibration and detector effects, and theoretical uncertainties, associated with the modelling of simulated processes. Theoretical uncertainties on the main backgrounds, such as $t\bar{t}$, $Z+$jets, and the subleading single-top processes, outweigh the relatively small experimental uncertainties. The systematic uncertainties employed in this analysis follow standard ATLAS recommendations.

8.3.1 Experimental Systematic Uncertainties

Systematic uncertainties associated with the JES and resolution [136] introduce over 20 nuisance parameters. Uncertainties related to the anti-k_t $R = 0.8$ reclustered jets are derived from the propagation of the small-R jet systematic uncertainties. Additionally, b-tagging uncertainties [137], and uncertainties linked to the soft $E_{\mathrm{T}}^{\mathrm{miss}}$ terms [146], alongside those associated with pileup conditions [189, 203] and luminosity are incorporated. Variations in the objects used for $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction are propagated through to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. Uncertainties related to the lepton reconstruction efficiency are found to be negligible and are consequently dropped from the fit. Experimental uncertainties are correlated across the analysis regions and processes. Compared to the dominant theoretical uncertainties discussed in the subsequent section, the impact of the relatively small experimental uncertainties is minimal.

8.3.2 Theoretical Uncertainties

$t\bar{t}$-related uncertainties

Systematic uncertainties associated with models used in the MC generator, the parton shower models, and the modelling of ISR and FSR are applied to the $t\bar{t}$ and single-top backgrounds. For the $t\bar{t}$ processes, generator uncertainties are evaluated by comparing

the nominal sample generated with POWHEG +PYTHIA against an alternative sample generated with MADGRAPH5_AMC@NLO +PYTHIA. Similarly, uncertainties related to the parton shower model are derived by comparing the nominal sample with one generated using POWHEG +HERWIG 7. Finally, uncertainties associated with ISR and FSR are estimated following centrally defined recommendations, which involve varying the renormalisation and factorisation scales, as well as the PYTHIA tuning parameters. These uncertainties, with sizes of around 10%, are the ones with the largest impact on the results due to $t\bar{t}$ being the dominant background process. To account for the cross-section uncertainty in this less-explored phase space, an additional 50% cross-section uncertainty is assigned to the sub-dominant $t\bar{t}+ \geq 1c$ process, implemented as a single nuisance parameter with a Gaussian prior. For the $t\bar{t}+ \geq 1b$ and $t\bar{t}$ + light processes, both the normalisations and their uncertainties are directly determined within the fit.

Single-top

Theoretical uncertainties for the single-top background are assessed using the same methodology as for the $t\bar{t}$ background, incorporating uncertainties related to the generator, parton shower, and radiation modelling. Additionally, for the tW process, the uncertainty associated with interference with $t\bar{t}$ production is evaluated by comparing the nominal sample to an alternative sample generated using the diagram subtraction scheme [204, 205]. While generator and interference systematics rank among the larger uncertainties in this analysis, their overall impact is limited due to the subdominant contribution of single-top processes in the signal region. Similar to the treatment of $t\bar{t}+ \geq 1c$ events, a 50% cross-section uncertainty is assigned to single-top events due to the limited knowledge of this process in the relevant phase space.

Statistical limitations in the alternative samples used to determine the systematic uncertainties of the single-top background can introduce noise into the fit procedure, potentially leading to spurious fitting behaviour. To ensure statistically robust systematic uncertainties, the cut on the NN output-score is loosened, providing sufficient statistics for comparing the nominal and varied samples. Validation studies were conducted to support

this approach: the ratio between the total nominal and varied single-top yields remains consistent across different output score-intervals, and the uncertainty shapes in m_{hh} are similarly consistent.

W/Z+jets

Uncertainties related to the renormalisation and factorisation scales are applied to the $W+$jets and $Z+$jets backgrounds. As variations in the renormalisation and factorisation scales are correlated, a 7-point scale variation is defined using pairwise variations of the two scales: $\{\mu_r, \mu_f\} \times \{\{0.5, 0.5\}, \{1, 0.5\}, \{0.5, 1\}, \{1, 1\}, \{2, 1\}, \{1, 2\}, \{2, 2\}\}$. Following centrally defined recommendations, two nuisance parameters are employed, corresponding to the largest (up and down) uncertainties within the envelope of these seven-point variations. Additional matching uncertainties associated with the CKKW merging scheme have been found to be negligible. For the $W+$jets process, a cross-section uncertainty of 50% is included and implemented as a single nuisance parameter with a Gaussian prior.

Uncertainties on Other Backgrounds

A cross-section uncertainty of 20% is applied to the minor topEW and diboson backgrounds to account for modelling and other theoretical uncertainties associated with these processes. Given the exceptionally low yields of the VH and triboson processes, they are excluded from the fitting process, and consequently, no further systematic uncertainties are derived for them.

Uncertainties on the Signal

The signal model is also subject to theoretical uncertainties. To calculate the theoretical systematic uncertainties for the signal Monte Carlo samples, event weights related to the renormalisation and factorisation scales are derived following the same procedure as for the $Z+$jets and $W+$jets processes.

Chapter 9

Results

Contents

9.1 The Fit Setup 169

9.2 Validation of the Fit Setup 169

9.3 The Discovery Fit 173

9.4 Results of the Exclusion Fit 175

9.5 Conclusion 177

This chapter presents preliminary results of this analysis. It begins with an overview of the fit setup and its validation, followed by a discussion of the discovery fit. The chapter concludes with the presentation of the derived exclusion limits for the dark-Higgs model.

9.1 The Fit Setup

The fits of the simulated data to the observed data distributions in the analysis regions are performed using `HistFitter` [206]. To conduct hypothesis tests, it is essential to obtain an accurate estimation of the background yields in the SR and DRs.

This is accomplished by determining dedicated normalisation factors for $Z+$jets, $t\bar{t}$ + light jets, and $t\bar{t}$ + $\geq 1b$-jets processes through a simultaneous fit to data in the background-rich CRs (defined in Section 8.2) along with the SR or DR. Given the high statistics in the CRs, these normalisation factors are primarily constrained there and subsequently propagated to the SR, the ValR, and the DRs.

Furthermore, the yields in the bins of the analysis regions are affected by systematic uncertainties. Alternative histograms that account for these uncertainties are constructed from the $\pm 1\sigma$ variations of the uncertainties detailed in Section 8.3. To define continuous response functions, used within the profile likelihood method, a sixth-order polynomial is utilised to interpolate between the 1σ up and down variations, while linear functions are applied for extrapolation up to $\pm 5\sigma$.

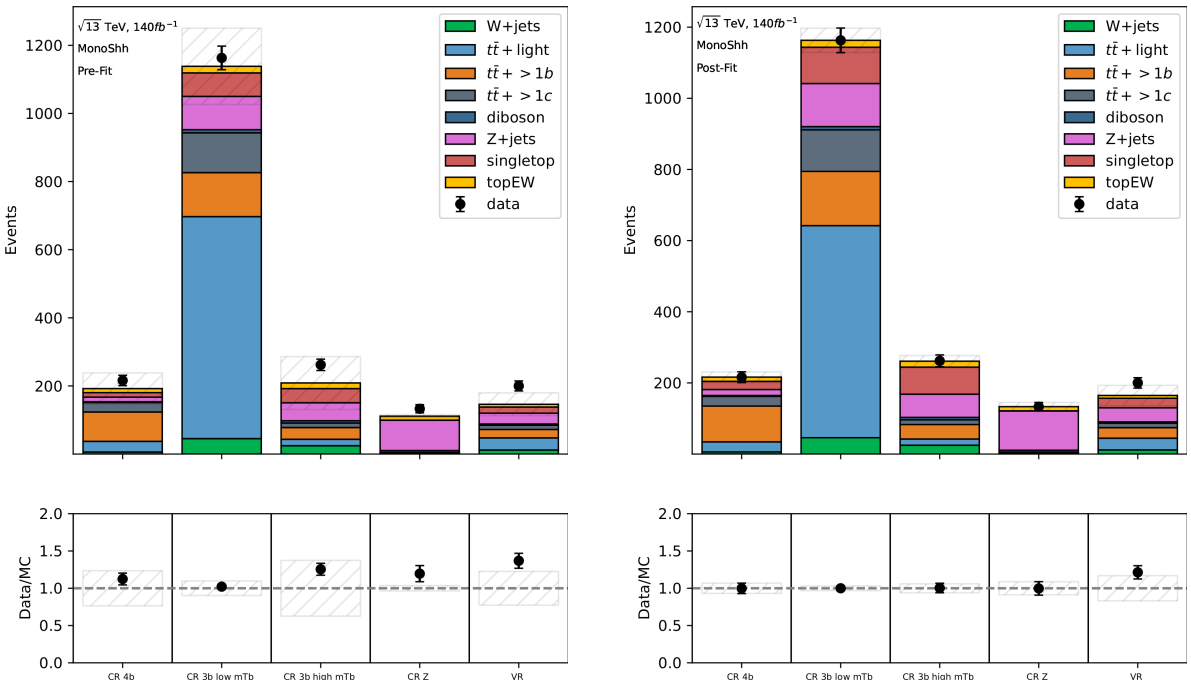
To extract results, two different profile-likelihood fits are performed: discovery fits and exclusion fits. Additional fits are conducted for validation purposes.

9.2 Validation of the Fit Setup

Background-only fits and fits to Asimov data (see Section 3.3) are conducted for validation purposes. These fits are used to study constraints, the behaviour of nuisance parameters, and the background estimation strategy before unblinding the signal sensitive SRs. In the background-only fit, the fit is applied only to the CRs, neglecting the SRs, and assuming no signal contamination in the CRs. The dedicated ValR, introduced in Section 8.2.3, allows for the validation of the semi-data-driven background estimation strategy. In this strategy, normalisation factors for $t\bar{t}+ \geq 1b$, $t\bar{t}$ + light jets, and $Z+$jets are constrained using CR data. For this purpose, the fit is applied only to the CRs, but the ValR is not

actively included in the fit model.

Prefit yields in each CR and the ValR are shown in Figure 9.1(a). As detailed in Section 8.2, the simulation underestimates the data yields in both the CRs and the VR. The expected background yields are adjusted through pulls on the normalisation factors of the main background processes, as well as on cross-section and other uncertainties, resulting in closure and excellent postfit agreement between data and MC yields in the CRs, as shown in Figure 9.1(b).



(a) Prefit yields.

(b) Postfit yields.

Figure 9.1: The prefit (a) and postfit (background-only fit) (b) yields in all CRs and the ValR.

9.2.1 Normalisation Factors:

The normalisation factors obtained in the fit are consistent with the expectations outlined in Section 8.2. The $Z+$jets normalisation is adjusted by a factor of $\mu_{Z+\text{jets}} = 1.24 \pm 0.14$ to account for the underestimation in the corresponding CR. Similarly, the $t\bar{t}$ + b-jets yields are increased by a factor of $\mu_{t\bar{t}bb} = 1.18 \pm 0.59$, addressing the underestimation of data in the 4b CR. The good prefit modelling in the 3b low-$m^{b}_{\text{T},min}$CR is maintained by a $t\bar{t}$ + light-jets normalisation of $\mu_{t\bar{t}\text{light}} = 0.91 \pm 0.14$, compensating for the upward pulls in

the other normalisations. The pulls and uncertainties on these normalisation factors for the three main backgrounds, along with the profiled nuisance parameters (discussed next), are summarised in Figure 9.2. The relatively small uncertainties, particularly on $\mu_{Z+\text{jets}}$ and $\mu_{t\bar{t}\text{light}}$, indicate that the analysis regions are well designed, increasing confidence in the accuracy of the background estimation also in the signal-sensitive regions.

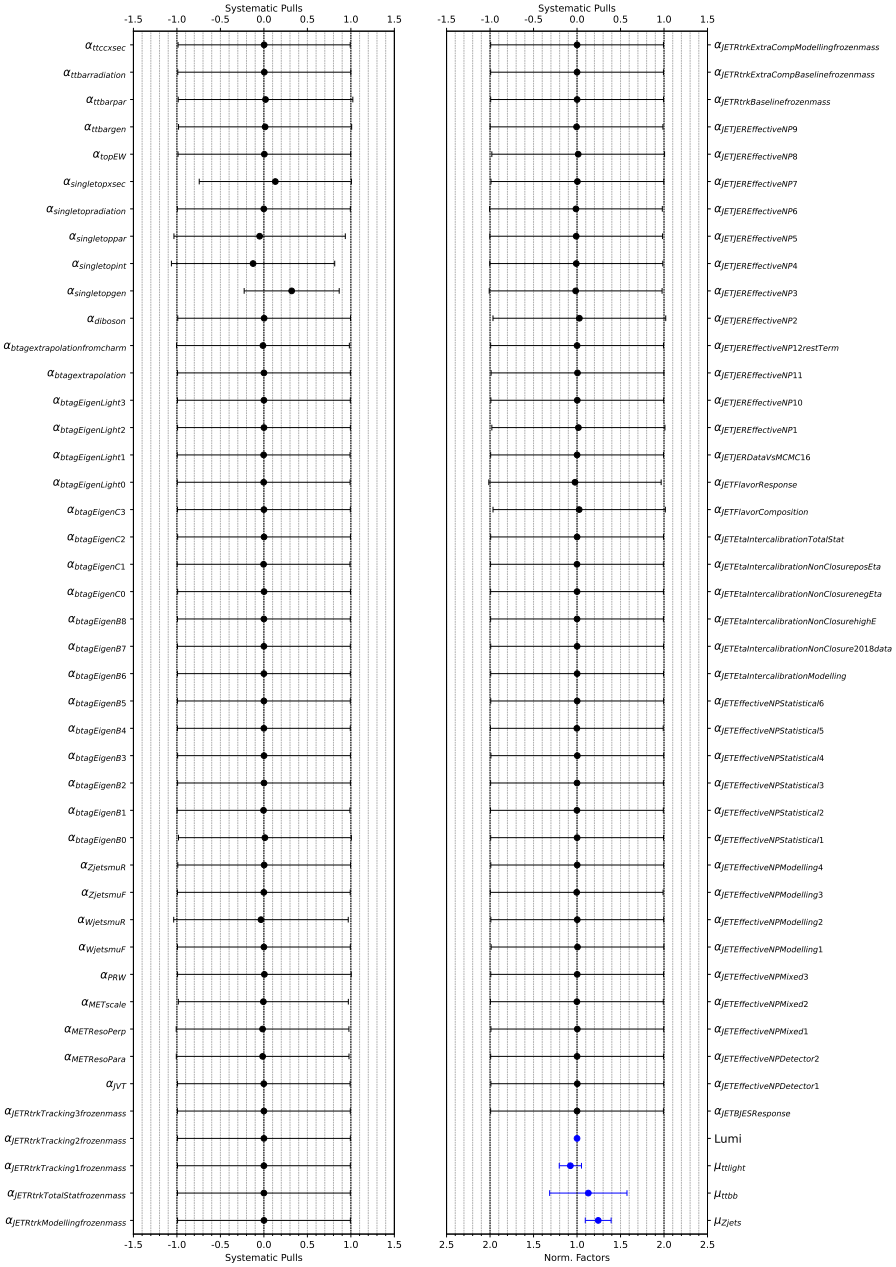


Figure 9.2: The nuisance parameters (α) and floating normalisation factors (μ) in the background-only fit. The resulting normalisation factors for the $t\bar{t}$ + b-jets, $t\bar{t}$ + light jets, and Z + jets procsses match the expectations detailed in Section 8.2. The only significant pulls are observed on the nuisance parameters associated with the theory uncertainties on the single-top background samples.

9.2.2 Nuisance Parameters:

As illustrated in Figure 9.2, the vast majority of the included nuisance parameters are either unconstrained or constrained only to a negligible extent in the fit. The only parameters that are significantly affected are those associated with the single-top background, which will be discussed below. Due to difficulties in isolating the single-top background in a dedicated CR, no freely floating normalisation for this process was included in the fit. Instead, an additional cross-section uncertainty of 50% was incorporated. However, the $3b$ high-$m^{b}_{\mathrm{T},min}$CR, which has some sensitivity to single-top, favours increased background yields, in agreement with but exceeding the adjusted normalisations of the three main background processes. As a result, the cross-section uncertainty is pulled up by approximately 20%, leading to an increased normalisation of the yields by around 10%.

Furthermore, the single-top background prediction is subject to large theoretical prefit uncertainties, which exhibit small partial pulls that further increase the single-top yields and additionally improve the postfit modelling of the m_{hh} distribution in the $3b$ high-$m^{b}_{\mathrm{T},min}$CR. The single-top generator and interference systematics are the only ones showing non-negligible postfit constraints. The profiling of the theoretical uncertainties for the single-top background were somewhat expected, as they were also observed in the Asimov fit, and the analysis detailed in Ref. [1], which operates in a similar phase space, exhibited comparable behaviour. While it should be noted that the uncertainties on the single-top processes are among the highest ranked based on their impact on the signal strength, as shown in Figure 9.3, their impact is still very low with approximately 5% or less. This low impact is consistent with expectations given the small single-top contribution to the SRs. Similarly, while the $t\bar{t}$ and $Z+$jets backgrounds have larger relative contributions to the signal region yields, their associated theoretical uncertainties still result in very small impact on the signal strength. This demonstrates that no single nuisance parameter significantly drives the result.

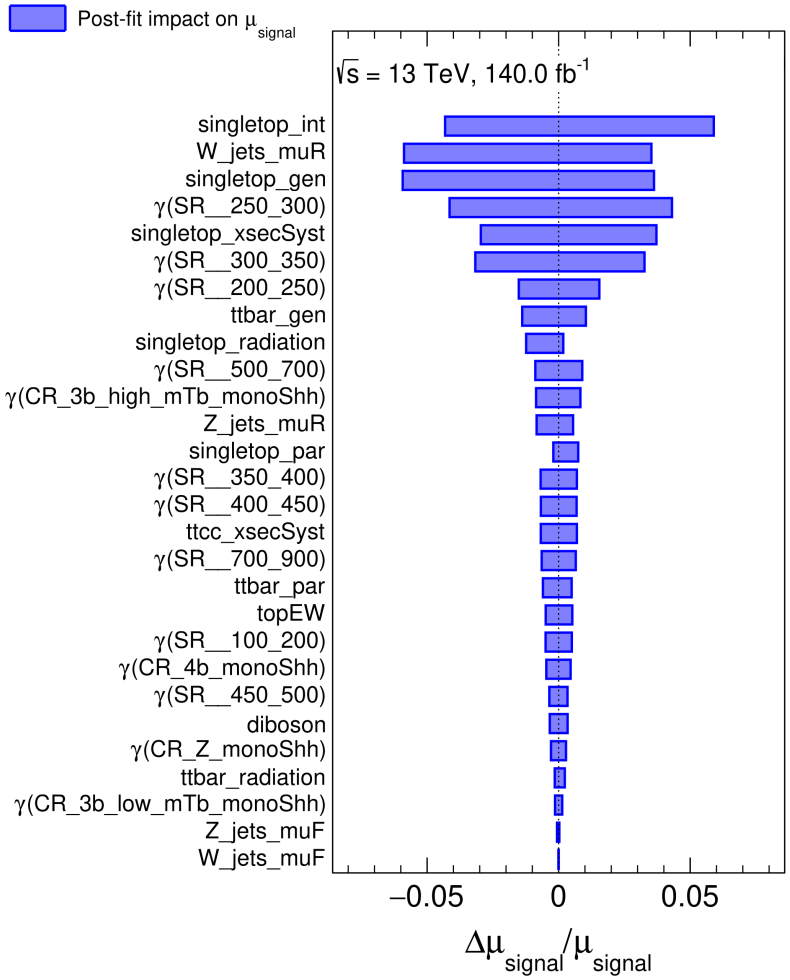


Figure 9.3: Systematic uncertainties and statistical uncertainties in SR bins ranked based on their impact on the signal strength μ_{signal} in a fit on Asimov data.

9.2.3 Postfit Agreement:

The CRs show excellent agreement of expected and observed data yields. Furthermore, the background-only fit results in good agreement between simulated and observed data within $\pm 1\sigma$ uncertainties in the ValR, as depicted in Figure 9.1, thereby validating the background estimation strategy.

9.3 The Discovery Fit

For the discovery of unspecified signals with a resonant di-Higgs + $E_{\text{T}}^{\text{miss}}$ signature, model-independent limits are set based on the yields in the seven DRs (DRa-DRg). For the

definitions of the DRs, see Section 8.1.2. In the discovery fit, both the CRs and the ValR are included, while potential signal contamination in the CRs is neglected, making the extracted limits conservative. To avoid additional assumptions about the signal sample, the DRs is treated as single-binned, and a dummy signal sample of one event is used in each bin. A hypothesis test, as outlined in Section 3.2.2, was performed to exclude the SM-only hypothesis based on the observed number of events in the DRs. As shown in Figure 9.4, no significant excess is observed in any of the model-independent DRs, and subsequently exclusion fits are conducted to establish model-dependent limits for the dark-Higgs model specified in Section 1.3.2.

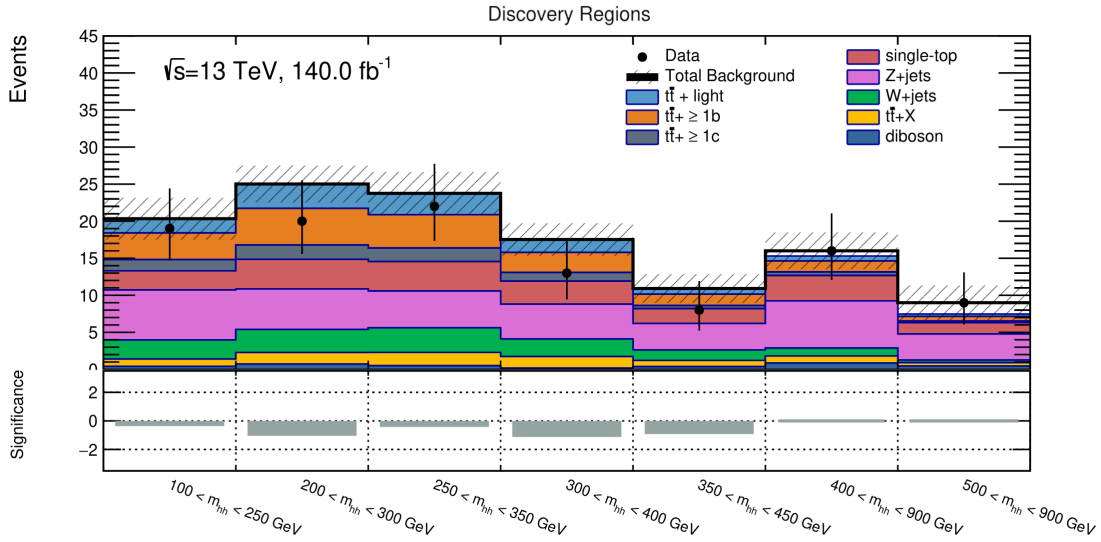


Figure 9.4: The post-fit event yields in all DRs. No significant excess is observed. The significances shown in the lower panel are estimated using the formula from Ref. [202].

Figure 9.5 shows the unblinded m_{hh} distribution with the SR selection applied. Good agreement between the observed data and the postfit background yields, as obtained in a background-only fit, are observed. The agreement between data and the expected background is within the $\pm 1\sigma$ uncertainties across all bins, with a slight underfluctuation observed in some of the central signal-sensitive bins. The data show relatively high statistical uncertainties, indicating that the analysis is statistically limited. This suggests that the sensitivity will significantly improve with the larger dataset available from Run 3.

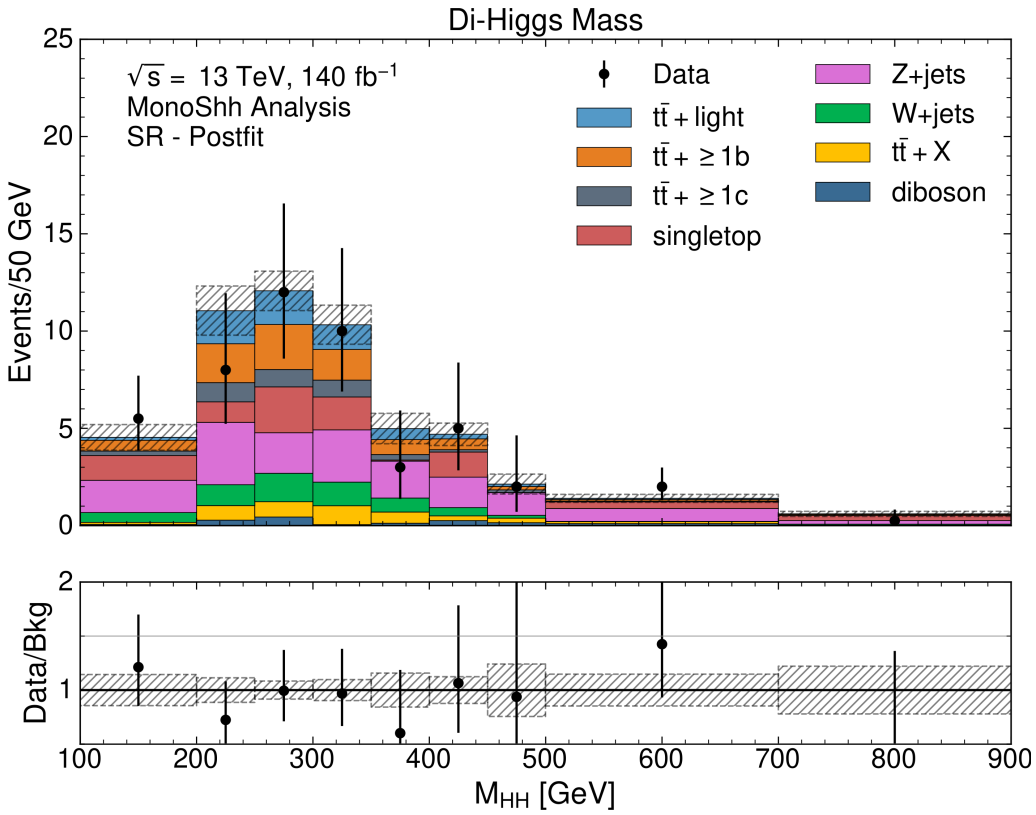


Figure 9.5: The unblinded binned SR displaying the postfit background yields obtained in a background-only fit.

9.4 Results of the Exclusion Fit

As no excess is observed in the DRs with the discovery fit, the strategy shifts towards setting model-dependent limits. The exclusion fits include both the CRs and the binned SRs, accounting for signal events in all regions. The results of the background estimation in this fit may differ slightly from the background-only fit, but any such differences are expected to be minimal due to the design of the CRs, ensuring high statistics and minimising signal contamination. An additional parameter, μ_{sig}, which scales the nominal signal prediction in the SR, is included. Despite the relatively low event yields in the bins of the SR, the distribution of the test statistic, $\tilde{q}_\mu$ (see Section 3.2.2), is approximated in the asymptotic regime of large statistics. This approach was verified by tests with 2000 toy experiments, which showed good agreement with the CLs values obtained in the asymptotic limit.

The resulting limits and the excluded phase space, along with the expected limits

obtained from a fit to Asimov data, are shown in Figure 9.6. These limits fully cover the kinematically accessible regions of phase space considered in the analysis. Despite the relatively low cross-sections at very low $m_{Z'}$, the exclusion limits extend down to $m_{Z'} \lesssim 500$ GeV. At the upper end of the mass spectrum, the observed limits reach up to $m_{Z'} \approx 2400$ GeV for high m_s and $m_{Z'} \approx 2200$ GeV for lower m_s. The dark-Higgs model is excluded for m_s values ranging from the lower kinematic limit of $m_s = 250$ GeV up to $m_s = 2m_\chi = 400$ GeV, where the invisible dark-Higgs decay channel becomes dominant. A slight underfluctuation in the data within some signal-sensitive bins in the SR leads to a marginally larger excluded region in the $m_{Z'} - m_s$ plane than the median expectation, but within the 1σ range of expected exclusions.

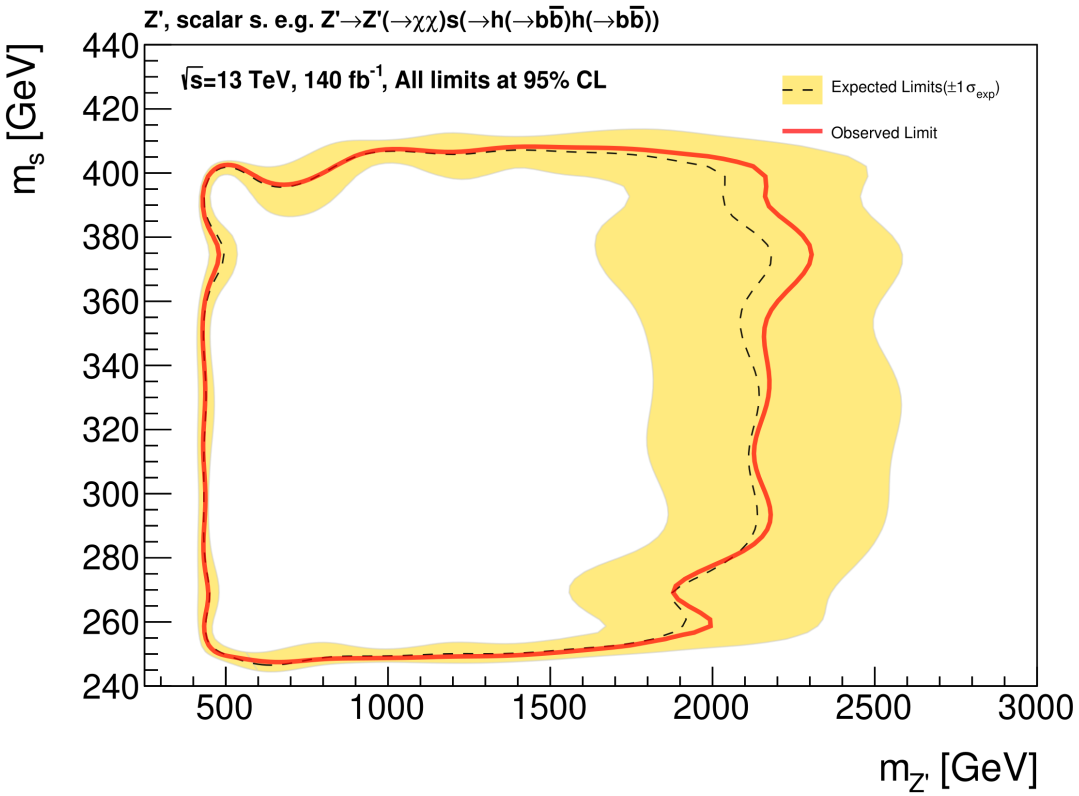


Figure 9.6: Observed and expected exclusion limits with $\pm 1\sigma$ uncertainty bands for the dark-Higgs model in terms of the dark-Higgs and Z' masses. The phase space inside the closed contours is excluded.

9.5 Conclusion

This thesis presents the first search for a resonant di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ signature, focusing on the four-b decay channel and optimised for the dark-Higgs strahlung process in a 2MDM model. The search was conducted using the 140 fb^{-1}, $\sqrt{s} = 13$ TeV proton-proton collision dataset, collected by the ATLAS experiment. No significant excess beyond the SM prediction was observed. The presented analysis excluded dark-Higgs bosons with masses between 250 and 405 GeV, for Z' masses up to 2400 GeV, significantly extending the current limits from previous searches targeting different dark-Higgs decay channels. As detailed in Figure 9.7, the best limits in the high dark-Higgs boson mass region, previously set by searches for $s \to WW$ decays in both ATLAS [207] and CMS [208], were improved by several hundred GeV in the high m_s and $m_{Z'}$ region. Compared to a direct reinterpretation of the analysis detailed in Ref. [1], in which the author was also involved, the optimised analysis presented in this thesis improved the exclusion limits by several hundred GeV in every direction in the mass plane. The increased sensitivity was achieved through the use of advanced machine learning techniques and optimised kinematic selections, exploiting the unique resonant decay structure of the targeted signals.

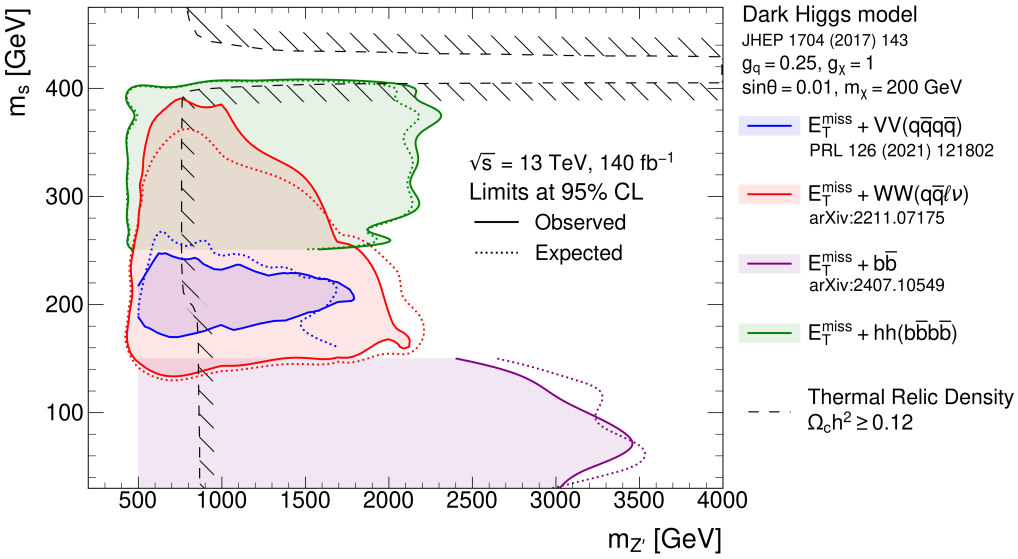


Figure 9.7: The current best exclusion limits for the dark-Higgs model are presented in the dark-Higgs and Z' mass plane, based on various ATLAS searches targeting different dark-Higgs decay channels. The exclusion limits resulting from the search in the $s \to hh$ channel are shown in green.

Part IV

Calibration of a

Boosted Higgs $\rightarrow b\bar{b}$ Tagger

Chapter 10

The $X \to b\bar{b}$ Algorithm

Contents

10.1 Introduction 180

10.2 Samples, Objects, and Selection 180

10.3 The Machine Learning Algorithm 181

10.4 Results 183

This chapter presents the boosted $X \to b\bar{b}$ tagger, used for identifying heavy resonances decaying into $b\bar{b}$ pairs. The primary reference for this chapter is Ref. [209].

10.1 Introduction

The high energies and luminosities achieved during the LHC Run 2 make ATLAS sensitive to various new physics scenarios, many of which predict new particles manifesting as heavy resonances decaying into SM particles such as the Higgs boson. Furthermore, these BSM theories often involve resonances with strong couplings to the b-quark. Therefore, algorithms for detecting and identifying $b\bar{b}$-pairs are crucial in searches for such hypothetical resonances.

Due to significant expected mass splittings between BSM resonances and final state SM particles, the latter are often boosted, meaning they possess high momentum. Consequently, the decay products, such as the $b\bar{b}$-pair, become strongly collimated.

As detailed in Sections 2.3.2 and 2.3.3, b-quarks and other partons undergo showering, hadronisation, and subsequent decay, leading to jet formation. When a $b\bar{b}$-pair results from the decay of a boosted Higgs or another heavy resonance X, the two resulting jets can overlap and form a single large-radius jet. This overlap complicates the reconstruction and identification of individual b-quarks, necessitating an understanding of the substructure of large jets alongside standard b-hadron identification techniques.

This chapter introduces the boosted $X \to b\bar{b}$ tagger [209], which identifies heavy resonances decaying to $b\bar{b}$-pairs by using information from the substructure of large jets. Identifying these jets as containing two b-quarks provides better performance than seeking two independent tags within them. The algorithm is based on a DNN that uses information from the three leading-p_T subjets to classify a jet as being top-like, Higgs-like, or light. To assess and account for any differing performances of this tagger in MC events and observed data, calibration is necessary, as discussed in Chapter 11.

10.2 Samples, Objects, and Selection

The dataset used for training and evaluating the tagger is derived from simulated Higgs $\to b\bar{b}$ events. Specifically, two Higgs bosons, originating from Randall-Sundrum

(RS) gravitons [210], are required to decay into pairs of b quarks. In addition to the signal events, two types of simulated background samples are generated: multijet events and top quark jets from $Z' \to t\bar{t}$ decays. To ensure the tagger performs well across a wide range of large-R jet p_{T}, both signal and top-jet background samples are simulated for various mass scenarios. This includes decays of RS gravitons with masses between 300 GeV and 6 TeV, and decays of Z' with masses ranging from 750 GeV to 5 TeV.

The primary objects used in this study are topo-cluster large-R jets, clustered using the anti-k_t algorithm. The resultant jets are then trimmed [211] by removing subjets of radius $R = 0.2$ that contribute less than 5% to the total p_{T} of the large-R jet, which helps in reducing pile-up contamination. Large-R jets are selected based on the criteria $250 < p_{\mathrm{T}} < 3000$ GeV and $\eta < 2.0$.

In addition, b-jets are reconstructed from ID tracks and clustered using the anti-k_t algorithm with a variable size parameter R. These jets are referred to as VR track jets. The size parameter is defined by $R = \frac{30}{p_{\mathrm{T}}}$, with the p_{T} given in GeV. To avoid extreme values, minimum and maximum sizes of $R = 0.02$ and $R = 0.4$ are implemented. Allowing the jet size to vary accounts for the fact that higher p_{T} jets tend to be narrower due to their boosted nature. The resulting VR track jets are ghost-associated (see Section 2.4.2) with the large-R jets if they meet specific requirements, as outlined in Ref. [209].

Lastly, the VR track jets are assigned a flavour label. If a b-hadron is found within $\Delta R < 0.3$, the jet is labelled as a b-jet. If no b-hadron is present, the jet is labelled as a c-jet if a c-hadron is found within $\Delta R < 0.3$. If neither b- nor c-hadrons are geometrically associated with the jet, it is labelled as a light jet.

10.3 The Machine Learning Algorithm

The motivation for developing the boosted $X \to b\bar{b}$ tagger was to utilise information from both the large-R jet and the VR track jets associated with it to improve the ability to identify heavy particle decays to $b\bar{b}$.

10.3.1 Training and Test Dataset

The algorithm is trained on events with a large-R jet that has a ghost-associated truth Higgs boson and tested on events selected as described in Section 10.2. Additionally, a large-R jet mass window cut between 76 GeV and 146 GeV is applied. This selection ensures that almost all large-R jets ($> 97.5\%$) have at least two sub-jets. The majority of the signal jets contain two b-jets, most of the top background jets contain one b-jet, and the multijet background predominantly contains no heavy-flavour jets.

To prevent the algorithm from classifying an event based solely on the large-R jet p_{T}, the three datasets representing the three different processes are downsampled to have the same large-R jet p_{T} distribution.

10.3.2 Input Features

The input features include kinematic variables of the large-R jet, specifically the p_{T} and η. Additionally, the three DL1r (see Section 2.4.2) output scores, which represent the probabilities of the jet being of light, charm, or bottom flavour, are provided for up to three associated VR track jets. This results in a total of up to 11 input variables. In cases where there are fewer than three sub-jets, the values for the other features are populated by the corresponding mean values.

10.3.3 The Neural Network

The algorithm is based on a feedforward neural network with six fully-connected layers, each containing 250 nodes and using a ReLU activation function. Batch normalisation is applied to stabilise the training, which is performed through the mini-batch gradient descent. The final layer employs a softmax function with three output nodes, resulting in three output values that give the probabilities for the event being classified as a Higgs, a multijet, or a top process. These three output probabilities are combined into a single D_{Xbb} score:

$$D_{Xbb} = \ln\left(\frac{p_{\mathrm{Higgs}}}{f_{\mathrm{top}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{top}}) \cdot p_{\mathrm{multijet}}}\right) , \tag{10.1}$$

where f_{top} denotes the fraction of top events, which is set to 0.25 in this thesis.

See Figure 10.1 for a detailed illustration of how the input features propagate through the various taggers to the final D_{Xbb} score.

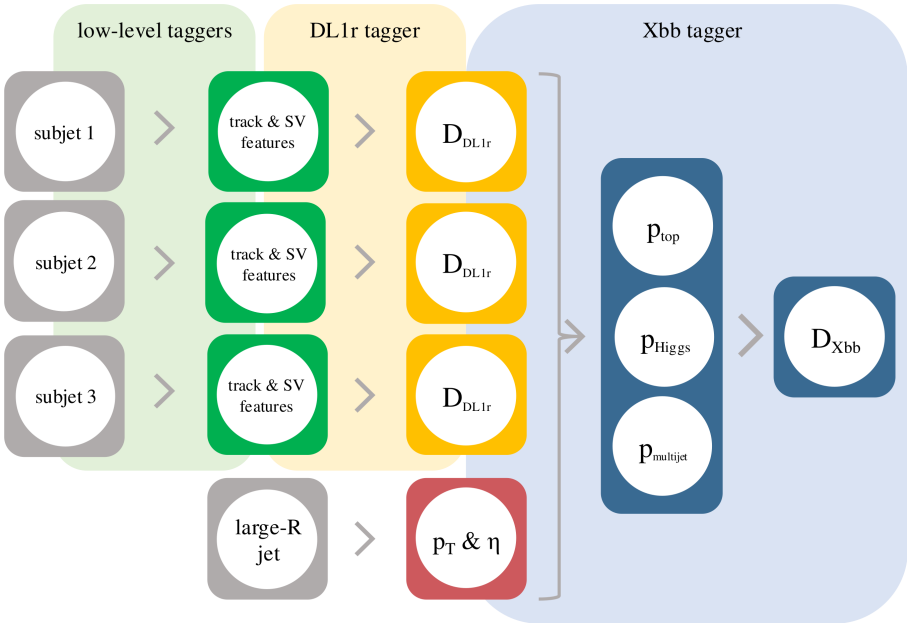


Figure 10.1: Schematic illustrating the propagation of input features through the low-level taggers, DL1r tagger, and finally the Xbb tagger, culminating in the D_{Xbb} score of the $X \to b\bar{b}$ tagger.

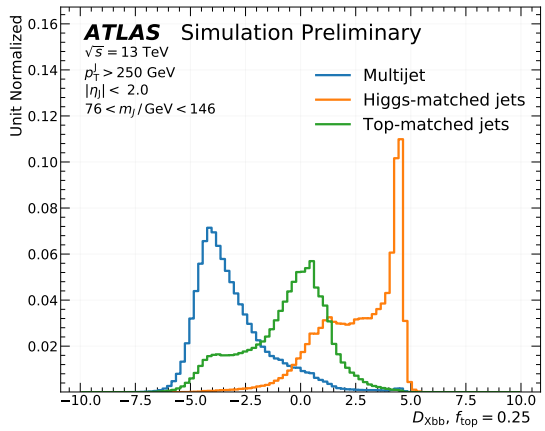
10.4 Results

The results are evaluated in terms of the Higgs tagging efficiency and the background rejection with respect to the two considered background processes. A large-R jet is tagged if its D_{Xbb} score exceeds a certain threshold. Three WPs (Tight, Medium, and Loose) are defined based on signal tagging efficiencies of 50%, 60%, and 70%, corresponding to D_{Xbb} values of 3.13, 2.55, and 1.92, respectively.

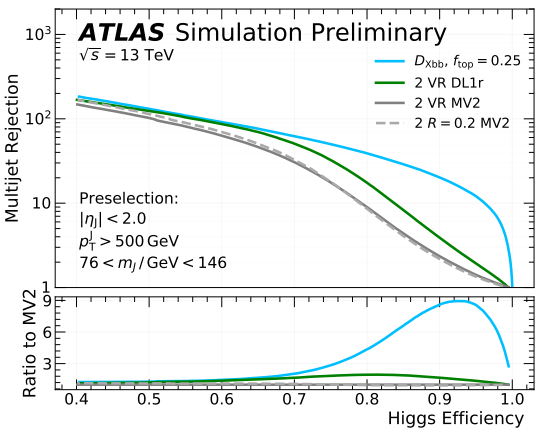
Compared to two individual DL1r tags, the $X \to b\bar{b}$ tagger achieves equal or better background rejection across the full range of tagging efficiencies for both background processes. Notably, the performance is significantly improved for high-p_{T} jets, where the collimation effect is stronger and the entire Higgs decay topology is contained within the large-R jet.

Figure 10.2(a) shows the D_{Xbb} score distribution for the three samples, and Figure 10.2(b) details the achieved multijet background rejection as a function of the tagging efficiency.

As this is a central feature in Part III, it should be noted that none of the input variables strongly correlate with the jet mass. Therefore, no significant sculpting of the jet mass is observed after applying cuts on the D_{Xbb} score.



(a) The discriminant distribution of the $X \to b\bar{b}$ algorithm with a top fraction of $f = 0.25$. Distributions are normalised to unity.



(b) Multijet rejection as a function of the $H \to b\bar{b}$ tagging efficiency, for large-R jets with $p_T > 500$ GeV. Performance of the D_{Xbb} algorithm is compared to DL1r and to two variants of MV2, one evaluated on VR jets, the other on fixed-radius $R = 0.2$ jets.

Figure 10.2: Distribution of D_{Xbb} (a) and performance of the tagger in terms of Higgs tagging efficiency and multijet rejection (b). The figures and captions are taken and adapted from Ref. [209].

Chapter 11

The $g \to b\bar{b}$ Calibration

Contents

11.1 Introduction 186

11.2 Data and Simulation 187

11.3 Triggers, Objects, and Selections 187

11.4 Flavour Discriminating Templates 189

11.5 The Likelihood Fit 194

11.6 Systematic Uncertainties 199

11.7 The Prefit Templates 201

This chapter details the setup and methods used in the multijet mis-tag efficiency calibration of the boosted $X \to b\bar{b}$ tagger, using events enriched in $g \to b\bar{b}$ splittings. The calibration aims to correct for potential performance discrepancies of the tagger between simulation and observed data.

11.1 Introduction

As discussed in Chapter 10, the tagger is trained using simulated events. However, due to the complexity of the physics in high-energy proton-proton collisions, MC simulations cannot perfectly replicate reality. To correct for the performance discrepancies of the tagger between MC and observed data, SFs are defined as:

$$\mathrm{SF} = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}} \ , \tag{11.1}$$

where ϵ_{data} and ϵ_{MC} are the double b-tagging efficiencies in observed data and MC, respectively.

This chapter outlines the calibration of the mis-tag efficiency in multi-jet events, which are the predominant background in numerous physics analyses targeting fully hadronic final states. In particular, multi-jet events enriched with jets containing b-hadrons present a significant background in searches for heavy resonances or SM Higgs bosons decaying into b-quark pairs. Multi-jet events with $g \to b\bar{b}$ splittings are especially important, as they frequently produce large-R jet topologies with two associated b-flavoured sub-jets and potentially additional light sub-jets.

A full calibration requires also the signal and single-fake topologies to be verified. The corrections for the signal tagging efficiency are derived from $Z(\to b\bar{b})$+jets and $Z(\to b\bar{b})\gamma$ events. This calibration is not included in this thesis and is detailed in Ref. [2]. Also not covered in this chapter is the second mis-tag calibration for large-R jets with one sub-jet containing a b-hadron and additional light sub-jets. Details on this calibration, which utilises semi-leptonically decaying $t\bar{t}$ events, are also available in Ref. [2].

Efficiency correction SFs are required for all supported WPs of the tagger, but this chapter focuses on the results obtained with the Tight WP, corresponding to a Higgs tagging efficiency of 50%.

11.2 Data and Simulation

The studies presented within this and the next chapter utilise the full Run 2 dataset recorded by ATLAS from 2015 to 2018 at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 140 fb^{-1}.

The di-jet MC samples were generated with Pythia 8.235 [106] using the A14 PS and hadronisation tune [212, 213] and the NNPDF2.3LO PDF set [171]. The interaction effects of the particles with the detector were simulated using Geant4 [123].

In addition to an inclusive MC sample, a muon-filtered sample requiring at least one muon with $p_{\mathrm{T}} > 3$ GeV and $|\eta| < 2.8$ was used. This requirement enriches the sample with semi-muonically decaying hadrons, thereby increasing the fraction of events containing b-quarks. The inclusive and muon-filtered samples are combined to enhance the available statistics. To ensure orthogonality and avoid double counting of events, all events that include a truth muon matching the filter of the muon-filtered samples are removed from the inclusive sample.

Additionally, another multi-jet sample was generated using Herwig 7.13 [120, 121] with the MMHT2014LO PDF set [214] and dedicated tunes [215]. This sample is used to evaluate uncertainties associated with alternative modelling choices.

11.3 Triggers, Objects, and Selections

To match the setup used for developing the tagger, the event selections and jet objects used in the calibration are very similar to those defined in Section 10.2.

11.3.1 Triggers

To trigger on suitable candidates for the calibration analysis, an additional $R = 0.4$ (small-R) jet is required in the selected events. A logical OR on the triggers used in each year of data taking is applied to select the events. Table 11.1 lists the employed jet triggers and the respective jet p_{T} thresholds that the offline calorimeter small-R jets must meet to

ensure full trigger efficiency. These jets are referred to as trigger jets in this chapter.

Triggers	**Trigger jet p_T threshold (GeV)**			
	2015	2016	2017	2018
HLT_j175	205.5	205.5	-	-
HLT_j260	297.5	297.5	282.5	282.5
HLT_j360	416.5	416.5	395.5	388.5
HLT_j380	-	436.0	-	-
HLT_j420	-	-	463.5	454.5

Table 11.1: Offline trigger jet p_T thresholds in GeV for the employed HLT triggers, depending on the year of data taking.

An additional challenge arises from the fact that most triggers (except HLT_j420) are pre-scaled. To reduce the trigger acceptance rate, only a fraction of the events passing the triggers are saved. For instance, a pre-scale of five means that only every fifth event is stored for further processing and analysis. Consequently, the observed data yields are scaled up, which leads to additional challenges when applying statistical uncertainties to the data yields.

11.3.2 Physics Objects and Candidate Selection

All events must have a primary vertex with at least two 500 MeV tracks associated with it. A trigger jet (as described in Section 11.3.1) and at least one large-R jet ($R = 1.0$) with $250 < p_{\mathrm{T}} < 1000$ GeV, $|\eta| < 2.0$, and a mass greater than 50 GeV are required. The trigger and large-R candidate jet must be separated by $\Delta R > 1.5$. In the case of multiple large-R candidate jets, the one leading in p_{T} is used.

To match the selection criteria applied for training the $X \to b\bar{b}$ tagger, the large-R jet must have at least two VR track jets ghost-associated with it. Each VR track jet must satisfy $p_{\mathrm{T}} > 7$ GeV and have at least two tracks associated with it. These tracks must have $p_{\mathrm{T}} > 0.5$ GeV, $|d_0| < 5.0$ mm, and $|z_0 \sin(\theta)| < 3.0$ mm, which reduces pile-up effects while preserving events with large impact parameters, expected from b-hadron decays.

Additionally, at least one of the ghost-associated track jets must have a muon of medium quality [216] associated with it. For a muon to be matched to the VR track jet, it must be found within its ΔR. This ensures the sample is enriched in b-jets, as high-p_{T} muons are frequently produced in semi-leptonic b-hadron decays. These non-isolated muons must satisfy the following requirements: $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.4$, $|d_0| < 2$ mm, and $|z_0 \sin(\theta)| < 2.0$ mm.

Events with overlapping subjets are vetoed. Figure 11.1 illustrates a typical event topology selected for this calibration analysis.

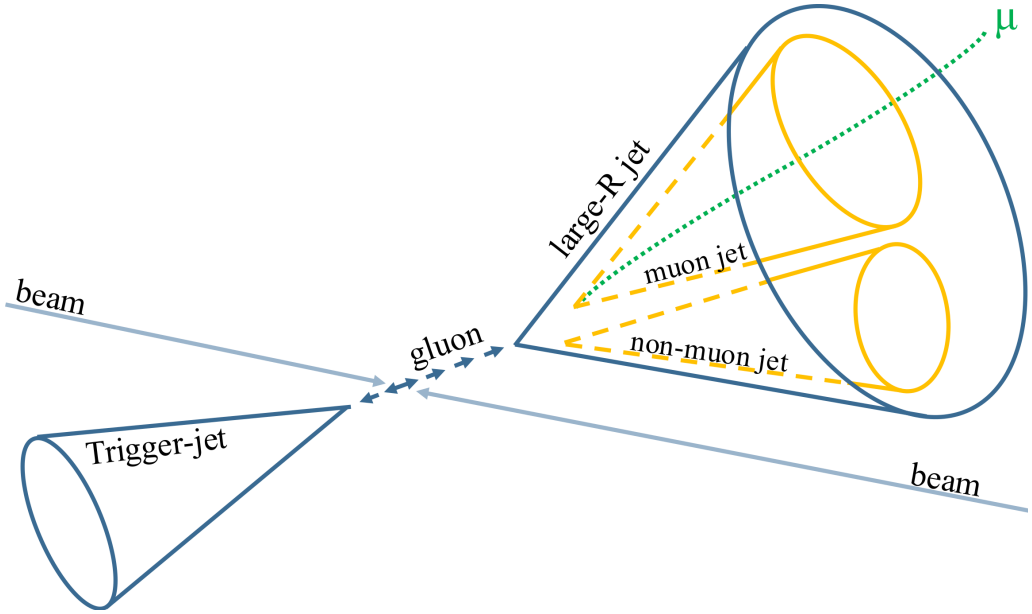


Figure 11.1: Illustration of the event topology selected for the $X \to b\bar{b}$ calibration, including the back-to-back trigger jet and large-R candidate jet with at least two VR subjets, one of which has a muon associated with it.

11.4 Flavour Discriminating Templates

Due to the complexities of QCD, multi-jet events, particularly their flavour composition, are difficult to model. Therefore, to accurately determine the tagging efficiencies, the flavour composition of the MC simulation must be corrected to match the observed data by reweighting the individual flavour components. This correction is achieved by fitting MC to observed data in flavour-sensitive distributions, referred to as templates. However, before correcting the flavour fractions in MC, candidates must be labelled based on their

flavour content.

11.4.1 Labelling of Flavour Fractions

To categorise candidate jets, the flavours of the three leading ghost-associated VR track jets, which are also referred to as subjets, are utilised. As detailed in Section 10.2, VR track jets are categorised into b-jets, c-jets, and light jets based on the geometric association of heavy-flavour hadrons in their vicinity. VR track jets originating from b- or c-hadron decays are labelled B and C, respectively, while subjets of light flavour are labelled L. If there are only two ghost-associated track jets, x is used for the *third jet* to indicate its absence.

The leading subjet (in p_{T}) with an associated muon is referred to as the *muon jet*, while the leading one among the remaining subjets is referred to as the *non-muon jet*. The flavour labels of the three subjets are then combined to form a label for the candidate jet, with the first letter representing the *muon jet*, the second the *non-muon jet*, and the third the optional *third jet*. This results in 36 three-letter categories.

However, it is not feasible to individually correct the contributions of such a large number of flavour categories. Firstly, many of these flavour categories contribute only a small fraction to the overall event yield. Secondly, some of the three-letter categories are difficult to distinguish from one another, as they exhibit very similar template shapes, making it impossible for the fit to correct them independently. To simplify the task of correcting the flavour contributions, the number of categories is reduced by merging similar ones into five two-letter categories: BB, BL, CC, CL, and LL. The contents of these five categories, referred to as the flavour fractions, are listed in Table 11.2.

There are different measures of similarity which can be used to merge categories, and therefore no unique choice for the labelling scheme exists. While various schemes were implemented and tested, the labelling scheme used is based on the similarity of the $X \to b\bar{b}$ score. This scheme ensures that all candidates with two b-flavoured subjets are included in the BB category. An alternative labelling scheme would merge categories based on similar template shapes, allowing better discrimination. With the current scheme, 16% of

Flavour Fractions	Three-Letter Flavour Categories
BB	BBB, BBC, BBL, BBx, BCB, BLB, CBB, LBB
BL	BCC, BCL, BCx, CBC, CBL, CBx, CCB, BLC, BLL, BLx, LBC, LBL, LBx, CLB, LCB, LLB
CC	CCC, CCL, CCx, CLC, LCC
CL	CLL, CLx, LCL, LCx, LLC
LL	LLL, LLx

Table 11.2: The five flavour fractions and their contents. The first, second, and third letters represent the flavour of the muon jet, non-muon jet, and third jet, respectively. An x indicates the absence of a third jet.

the simulated events are categorised as BB, 22% as BL, 11% as CC, 23% as CL, and 28% as LL.

11.4.2 Flavour Discriminating Variables

To correct the normalisation of the flavour fractions, the predicted events are fitted to data using templates with shapes sensitive to the flavour fractions. Two of the numerous studied variables used to construct these templates are discussed in this section, namely the signed impact parameter significance s_{d_0} and the reconstructed mass of the secondary vertices.

Signed Impact Parameter Significance s_{d_0}

The transverse impact parameter d_0 is defined as the distance of closest approach of a track to the PV in the transverse plane. It is crucial for b-tagging because b-hadrons have relatively long lifetimes, creating displaced SVs. Hence, the tracks of decay products of b-hadrons tend to have larger impact parameters.

Additionally, a sign s can be assigned to the impact parameter to enhance its value as a discriminator for tracks originating from decays of hadrons of different flavours. The sign is positive if the interpolated track intersects the jet axis in front of the primary vertex, and negative otherwise. As depicted in Figure 11.2, this is equivalent to assigning $s = +1$ if the angle between the jet axis and the transverse impact parameter vector is less than 90°, and otherwise $s = -1$. The resulting quantity is called the signed transverse

impact parameter significance:

$$s_{d_0} = s\frac{d_0}{\sigma_{d_0}} \ , \tag{11.2}$$

where σ_{d_0} is the associated uncertainty. Tracks associated with the decay of b-hadrons or hadrons of other heavy flavours tend to have large and positive s_{d_0} values. Negative values, however, can appear due to finite impact parameter resolution. Tracks in light jets, on the other hand, typically result in low s_{d_0} values and exhibit relatively equal probability for negative and positive signs, resulting in symmetric distributions centred around 0.

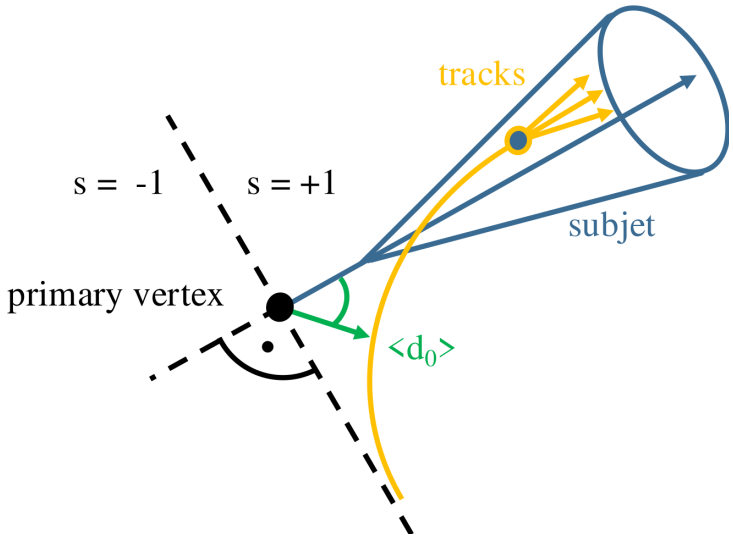


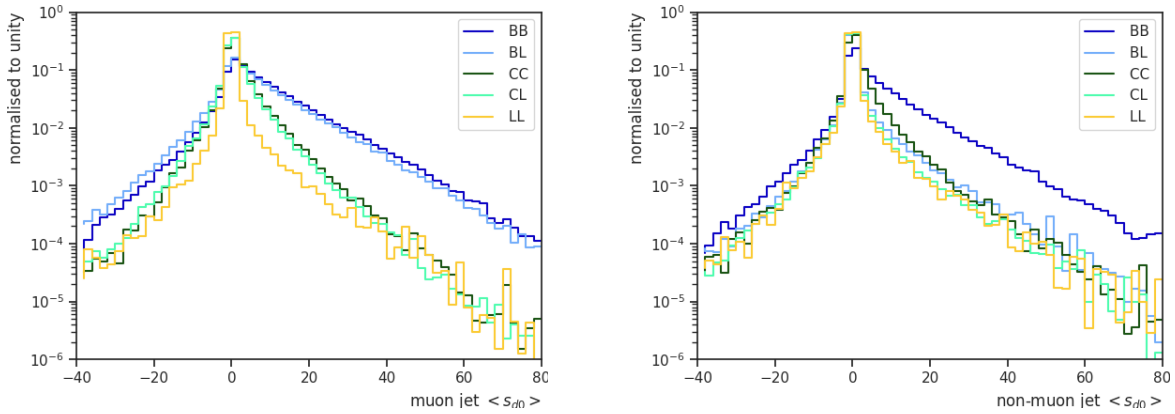
Figure 11.2: Schematic illustration of a subjet, and the mean impact parameter $< d_0 >$ resulting from the three leading tracks associated with it. The sign s used to evaluate $< s_{d0} >$ is determined by the angle between the impact parameter vector and the subjet axis.

To derive robust flavour-discriminating quantities and reduce the impact of single-track pile-up contamination and from individual tracks from relatively long-lived light hadrons (e.g. K_s), the average of the s_{d_0} values of the up to three leading tracks per subjet is computed. The resultant mean transverse impact parameter significance is referred to in this thesis as $\langle s_{d_0} \rangle$.

The normalised $\langle s_{d_0} \rangle$ distributions for the *muon jet* and the *non-muon jet* are shown in Figures 11.3(a) and 11.3(b), respectively. Both distributions clearly show different shapes for light, c-, and b-flavoured tracks. As expected, the latter show relatively flat distributions, with significant contributions in the tails, while the light jets exhibit peaky distributions with exponentially decreasing tails centred around 0. Also, clearly visible are

the enhanced tails, especially for positive $\langle s_{d_0} \rangle$ values for the BB component. However, the *muon jet* templates show very similar distributions for BB and BL as well as for CC and CL. This is due to the fact that both only consider jets with the same flavour. Thus, the *muon jet* templates alone do not allow discrimination between all the five flavour categories implemented in Section 11.4.1. The *non-muon jet* template, however, breaks this degeneracy between BB and BL, and between CC and CL, as depicted in Figure 11.3. Thus a simultaneous fit to both distributions should be able to determine corrections of all flavour fractions independently.

While the flavour templates of the *muon jet* and *non-muon jet* together provide good flavour discrimination, a closer look at the template shapes also discloses the flaws of the labelling scheme. For example, the shape of the *non-muon jet* BL template clearly has a different shape than the CL and LL ones, which ideally should be identical. This results from the C and B contributions to the *non-muon jet* category of the BL category, which make the tails decrease less than a pure light template. Furthermore, for the BB category, the *non-muon jet* template exhibits steeper tails than its *muon jet* template, due to the non-negligible contribution from the BLB sub-category. In general, the *non-muon jet* templates are less pure than the *muon jet* ones. However, the templates for both subjets together provide clear discrimination between all five flavour fractions and thus allow correction of the individual contributions.



(a) Muon jet $\langle s_{d_0} \rangle$ templates.

(b) Non-muon jet $\langle s_{d_0} \rangle$ templates.

Figure 11.3: The $\langle s_{d_0} \rangle$ templates for the five flavour fractions for the *muon jet* and the *non-muon jet*.

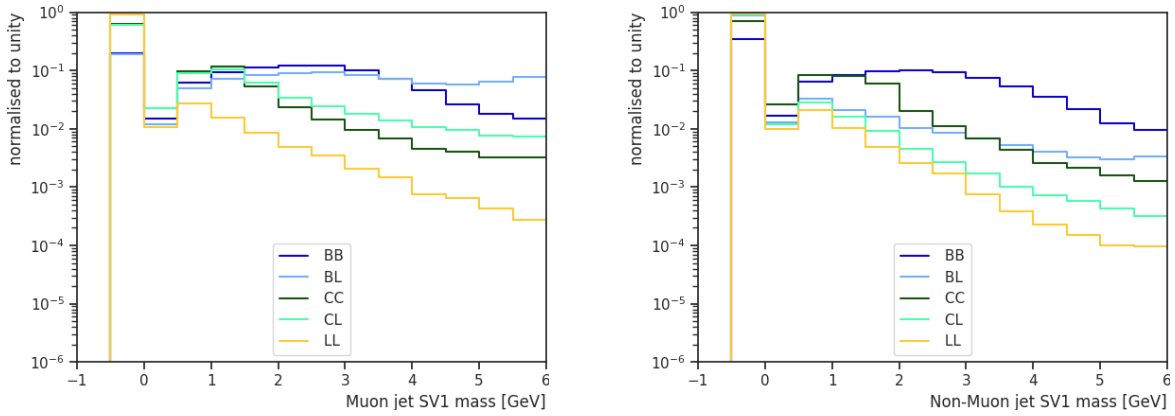
Secondary Vertex Mass

Another variable with discriminatory power between the five flavour fractions is the mass of the SV associated with a subjet. For more details on the reconstruction of decay vertices, see Section 2.4.1. Hadrons including b-quarks are generally heavier than those with c-hadrons. Subsequently, the reconstructed mass of the SV associated with the decay of the hadron shows the same ordering. Moreover, jets originating from the decay of very short-lived light hadrons typically don't allow the reconstruction of SVs. As SVs originating from long-lived light hadrons are, together with those from material interactions and photon conversions, removed in the reconstruction step, light subjets are either associated with a fake vertex or, in the majority of cases, do not have an associated vertex and are thus assigned a negative value for their SV mass.

Figures 11.4(a) and 11.4(b) show the SV mass templates for the *muon jet* and *non-muon jet*, respectively. Again we see a clear separation of the B, C, and L contributions to the templates of both subjets. While we also see some discrepancies in the templates of the same flavours, such as in the CC and CL *muon jet* templates, which can be understood by the impurities resulting from the labelling scheme, there is one particular feature that stands out. The *muon jet*, and to a lesser extent also the *non-muon jet* BL templates, show a second peak at high SV masses close to 6 GeV. This behaviour is explained by the inclusion of very strongly collimated $g \to b\bar{b}$ events, where the b-jets cannot be resolved and are reconstructed as a single VR track jet containing two b-hadrons. Subsequently, these jets exhibit high SV masses and distort the template shape.

11.5 The Likelihood Fit

To extract the efficiency correction factors and simultaneously adjust the flavour fractions, the templates described in the previous section are fitted to the observed data.



(a) Muon jet $\langle s_{d_0} \rangle$ templates.

(b) Non-muon jet $\langle s_{d_0} \rangle$ templates.

Figure 11.4: The SV-mass templates for the five flavour fractions for the *muon jet* and the *non-muon jet*.

11.5.1 Scale Factors and Flavour Fraction Correction

The primary PoI in the fit is the efficiency correction SF. As detailed in Equation 11.1, it is introduced as the ratio of double b-tagging efficiencies in observed data and MC simulation, where the efficiencies are estimated by the number of events in the BB flavour category:

$$\mathrm{SF} = \frac{\epsilon^{\mathrm{BB}}_{\mathrm{data}}}{\epsilon^{\mathrm{BB}}_{\mathrm{MC}}} \ . \tag{11.3}$$

The efficiencies are given by:

$$\epsilon^{\mathrm{BB}}_{\mathrm{data}} = \frac{N^{\mathrm{BB,pass}}_{\mathrm{data}}}{N^{\mathrm{BB,incl}}_{\mathrm{data}}} \quad \text{and} \quad \epsilon^{\mathrm{BB}}_{\mathrm{MC}} = \frac{N^{\mathrm{BB,pass}}_{\mathrm{MC}}}{N^{\mathrm{BB,incl}}_{\mathrm{MC}}} \ . \tag{11.4}$$

Here, $N^{\mathrm{BB,pass}}$ represents the number of BB events passing the tagging criteria, and $N^{\mathrm{BB,incl}}$ denotes the total number of BB events. However, the number of BB events in MC must be corrected.

To adjust the flavour composition of the MC simulation, a normalisation factor NF^{flav} is introduced for each of the five flavour fractions: flav $\in$ {BB, BL, CC, CL, LL}. Each NF^{flav} has a nominal value of 1 and serves as a multiplicative correction to the normalisation of the background sample across all regions.

Consequently, the corrected MC simulated event yields in Equation 11.4 are given by:

$$N_{\text{MC}}^{\text{BB,pass}} = N_{\text{prefit}}^{\text{BB,pass}} \times NF^{\text{BB}} \text{ and} \tag{11.5}$$

$$N_{\text{MC}}^{\text{BB,incl}} = N_{\text{prefit}}^{\text{BB,incl}} \times NF^{\text{BB}} \ , \tag{11.6}$$

where $N_{\text{prefit}}^{\text{flav,reg}}$ represents the nominal estimated event yield for a given flavour fraction and region (e.g., *pass* or *fail*, as discussed in Section 11.5.3) before any corrections.

Assuming a valid correction of the flavour fractions, the BB components in the observed data can be approximated by:

$$N_{\text{data}}^{\text{BB,incl}} = N_{\text{MC}}^{\text{BB,incl}} \text{ and} \tag{11.7}$$

$$N_{\text{data}}^{\text{BB,pass}} = N_{\text{data}}^{\text{pass}} - N_{\text{MC}}^{\text{non-BB,pass}} \ , \tag{11.8}$$

where $N_{\text{data}}^{\text{pass}}$ is the total yield of data that meet the tagging criteria, and $N_{\text{MC}}^{\text{non-BB,pass}}$ is the small (less than 10% of total yield) MC background contribution from non-BB fractions that pass the tagging criteria.

While the SF is applied to the BB component of the tagged events (*pass* region), its impact on those that fail the tagging (*fail* region) is represented by an anti-SF, which is applied to the BB component of the untagged events and is defined as:

$$\text{anti-}SF = \frac{1 - \epsilon_{\text{data}}}{1 - \epsilon_{\text{MC}}} = \frac{1 - \epsilon_{\text{MC}} \times SF}{1 - \epsilon_{\text{MC}}} \ . \tag{11.9}$$

11.5.2 The Likelihood

The likelihood function to be maximised is constructed as described in Section 3.1. It is defined in terms of the PoI (the efficiency correction SF), the flavour fraction normalisations **NF**, the nuisance parameters $\boldsymbol{\theta}$ (discussed in Section 11.6), and the observed data yields in all bins of all analysis regions $\mathbf{N}_{\text{data}}$:

$$L(SF, \boldsymbol{NF}, \boldsymbol{\theta}; \boldsymbol{N}_{\text{data}}) = \prod_{r \in \text{regions}} \prod_{b \in \text{bins}} \mathcal{P}(N_{\text{data}}^{r,b} | \lambda^{r,b}(SF, \boldsymbol{NF}, \boldsymbol{\theta})) \times \mathcal{C}_{syst}(\boldsymbol{\theta}) \ . \tag{11.10}$$

Here, $\lambda^{r,b}$ is the estimated expected event yield in region r and bin b:

$$\lambda^{r,b} = \sum_{\text{flav}} \alpha^{\text{flav},r} \times NF^{\text{flav}} \times N^{\text{flav},r,b}_{\text{MC,prefit}} \ , \tag{11.11}$$

where $\alpha^{\text{flav},r} = 1$ for all flavour fractions except for BB, where it is given by:

$$\alpha^{\text{flav},r} = \begin{cases} SF & \text{for flav} = \text{BB and } r = \text{pass} \\ \text{anti-SF} & \text{for flav} = \text{BB and } r = \text{fail} \\ 1 & \text{else} \end{cases} \tag{11.12}$$

11.5.3 Analysis Regions

To derive the efficiency correction SFs, it is necessary to split the available data into events tagged by the $X \to b\bar{b}$ algorithm and those that are not. This chapter presents the results obtained with the 50% WP of the $X \to b\bar{b}$ tagger, using the threshold $D_{Xbb} > 3.13$ to divide the selected events into the *pass* and *fail* regions. Furthermore, both the *muon jet* and the *non-muon jet* templates must be fitted to the data distributions to enable discrimination between all five flavour fractions and correct their normalisation (see Section 11.4.2). This results in four preliminary analysis regions: *muon jet pass*, *non-muon jet pass*, *muon jet fail*, and *non-muon jet fail*.

Muon Categories

Additionally, studies were conducted to further split the regions based on muon multiplicity. Introducing $n_\mu = 1$ and $n_\mu = 2$ categories doubles the number of regions to eight.

The $n_\mu = 2$ regions provide doubly-enhanced heavy flavour distributions, thus increasing the sensitivity. Moreover, this split allows for a slight reduction in the strong observed anti-correlation between the SF and NF^{BB}. Both of these factors affect the BB component. While NF^{BB} serves as an overall normalisation, the SF is only applied in the *pass* regions but is linked to the *fail* regions through its correlation with the anti-SF. However, due to the very good tagging efficiencies, the contribution of the BB flavour

category in the *fail* region is very low ($< 10\%$). Consequently, the effect of the *fail* regions on both NF^{BB} and the SF is limited, and their values are predominantly determined in the *pass* regions, with their effects merely propagated to the *fail* regions. The effect of both quantities is the same in the *pass* regions, resulting in a very strong anti-correlation.

The newly introduced $n_\mu = 2$ region exhibits a higher fraction of BB events. While the inclusive $n_\mu = 1$ region has a BB fraction of around 12%, the inclusive $n_\mu = 2$ region has a BB contribution of more than 35%. Consequently, about a quarter of the events failing the b-tagging belong to the BB flavour category. This results in a stronger impact of the *fail* regions on the SF and NF^{BB} and subsequently a reduced anti-correlation, thus increasing the confidence in the resultant efficiency correction SF.

Table 11.3 summarises the contributions of the different flavour fractions to the analysis regions.

	Pass		**Fail**		**Inclusive**
	1 Muon	**2 Muon**	**1 Muon**	**2 Muon**	
BB	96	97	9	36	16
BL	4	3	24	19	23
CC	< 1	< 1	12	27	11
CL	< 1	< 1	25	10	23
LL	< 1	< 1	30	8	28

Table 11.3: Flavour fractions in percentages in the various regions used in the fit.

Transverse Momentum of Large-R Jets

The flavour fractions listed in Table 11.3 are only part of the picture, as a dependency on the p_{T} of the candidate jets is observed. For increasing large-R jet p_{T}, the fraction of light jets in the simulation increases, while that of heavy-flavour jets decreases. For instance, for jets with p_{T} between 600 GeV and 750 GeV, the prefit LL fraction is 39% and the BB fraction is 12%, compared to overall fractions of 28% and 16%, respectively, as detailed in Table 11.3. It cannot be excluded that the necessary corrections of the flavour fractions

are also correlated with the large-R jet p_{T}.

To ensure that potential dependencies of the efficiency differences between data and simulation are captured, the fit is performed independently in multiple large-R jet p_{T} bins. Due to higher available statistics, narrower bins are used in the low-p_{T} region and increasingly wider bins for high-p_{T} candidate jets. The utilised binning is: 250 GeV, 280 GeV, 310 GeV, 340 GeV, 380 GeV, 420 GeV, 460 GeV, 500 GeV, 550 GeV, 600 GeV, 750 GeV, and 1000 GeV.

11.6 Systematic Uncertainties

This section discusses the systematic uncertainties affecting the fit templates and results. As detailed in Section 3.1.3, these systematic uncertainties are parameterised in terms of nuisance parameters and incorporated in the profile likelihood fit. Apart from statistical uncertainties associated with the limited number of observed and MC simulated events, the uncertainties can be grouped into two categories: experimental uncertainties associated with the calibration of physics objects and detector effects, and theoretical uncertainties from models used in the simulations.

11.6.1 Experimental Uncertainties

As detailed in Section 2.4, reconstructed physics objects, such as muons and the various types of jets utilised in this calibration analysis, have systematic uncertainties associated with their calibration and detector effects. The size of these uncertainties was determined by dedicated auxiliary measurements, and the resulting $\pm 1\sigma$ values are propagated to the physics objects. The varied objects are propagated through the analysis by applying the same selections and forming alternative flavour templates, which are subsequently used to define the uncertainties.

Uncertainties associated with the JES and resolution are implemented for both large-R jets [217, 218] and the subjets [219]. Furthermore, reconstruction and momentum uncertainties, as well as uncertainties on the efficiencies of identification and isolation of

muons, are implemented [126]. The relative impact of the majority of these uncertainties on $\langle s_{d_0} \rangle$ is significantly below 1% and thus they don't survive the applied pruning (see below). Moreover, a set of systematic tracking efficiency uncertainties are included and extended by dedicated uncertainties accounting for the fake rates and efficiencies in dense environments [125]. The latter are, in most regions, the dominant experimental uncertainties, with relative values up to 5%.

In addition, systematic uncertainties are introduced to account for additional processes that can cause large s_{d_0} values and SVs, such as photon conversion, long-lived light particles, and hadronic interactions with detector material. The size of the relative uncertainties associated with these processes is, in most regions, on the percent level.

11.6.2 Theoretical Uncertainties

Many of the dominant uncertainties in this analysis are related to the modelling of the multijet events in MC.

A comparison of the flavour template shapes resulting from Pythia and Herwig generators is used to estimate the uncertainties related to the modelling of PS and hadronisation. These uncertainties are decorrelated between the flavour fractions. These generator uncertainties are by far the largest in this calibration analysis, with relative shape differences of up to 20%.

Furthermore, uncertainties on the renormalisation scale are included by separately varying the ISR and FSR scales up and down by factors of two. Similar to the PS and hadronisation uncertainties, only the shape difference is considered for the variation, as normalisations are corrected through the flavour fraction normalisation factors. After smoothing (see below), the shapes of these uncertainties differ by up to 5% compared to the nominal ones.

Finally, the fractions of specific b-hadrons are subject to uncertainties, affecting the flavour templates due to different lifetimes and decay modes. The variations are derived by reweighting the simulated events so that the overall composition of the truth b-hadrons in the sample matches the one derived by the Heavy Flavour Averaging Group in Ref. [220].

These b-hadron fraction uncertainties are applied only to the BB and BL templates. While the relative size of the uncertainty depends on the region, flavour template, and bin, they typically range between 1% and 2%.

11.6.3 Smoothing and Pruning

Smoothing is applied to the ratio of the nominal and systematic templates to reduce statistical noise and enhance the stability of the fit. The MAXVARIATION smoothing algorithm from the TRExFitter framework [221, 222], based on HistFactory [149] and RooFit [223], is used for smoothing. This algorithm first rebins the templates until the relative MC statistical uncertainty is below a specified threshold and then applies the *353QH twice* algorithm detailed in Ref. [224].

Additionally, to stabilise the fit, pruning is applied to remove small uncertainties of less than 0.5%, either in their shape or normalisation component. Pruning is applied separately for each flavour template and region, allowing individual systematics to be removed in only a subset of the regions or flavour templates. Consequently, it is common for the heavy flavour template uncertainties to be pruned in the fail region, and vice versa for the light flavour fractions in the pass regions. While most tracking uncertainties and uncertainties related to light long-lived particles remain in at least some regions, a large fraction of other experimental uncertainties are pruned due to their negligible size. As expected, the dominant theoretical uncertainties, however, remain untouched by the pruning.

11.7 The Prefit Templates

11.7.1 $\langle s_{d_0} \rangle$ Templates

Figures 11.5 and 11.6 show the $\langle s_{d_0} \rangle$ prefit templates for the 600-750 GeV large-R jet p_{T} range with all systematic and statistical uncertainties applied. As detailed in Section 11.5.3, the $n_\mu = 2$ templates exhibit relatively low statistics compared to the $n_\mu = 1$ templates, primarily serving to break the anti-correlation between the SF and the BB normalisation.

While many bins exhibit decent data/MC agreement within uncertainties, two main types of mismodelling are observed. In the pass region, the MC underestimates the data for low $\langle s_{d_0} \rangle$ values and overestimates it for high values, especially prominent in the *non-muon jet* templates (see, for example Figure 11.6(a)).

In the fail regions, the MC underestimates the data in the tails, leading to a U-shaped data/MC ratio for these templates. This effect is observed for both the *muon jet* and *non-muon jet* templates and is more pronounced for the high-p_{T} candidate jets. See, for example, Figures 11.5(b) and 11.6(b).

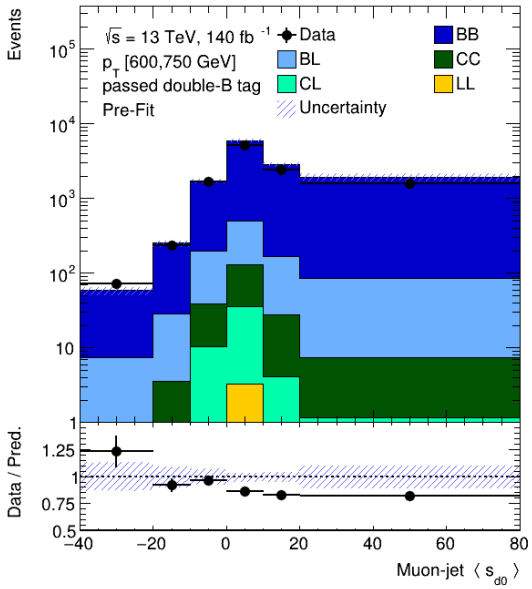
The observed prefit mismodelling makes it challenging to ensure an accurate correction of the flavour fractions and subsequently derive valid SFs.

Studies showed that a significant portion of the $\langle s_{d_0} \rangle$ disagreement between data and simulation can be attributed to a strong mismodelling of σ_{d_0}. Therefore, as an additional check to validate the robustness of the derived normalisations and SFs despite the mismodelling, fits were also performed using the SV-mass templates.

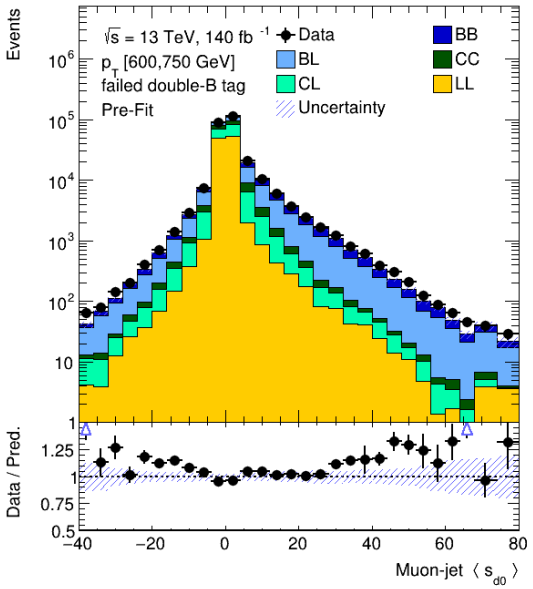
11.7.2 Secondary Vertex Mass Templates

Figures 11.7 and 11.8 show the *muon jet* and *non-muon jet* SV-mass templates for one large-R jet p_{T} bin, respectively.

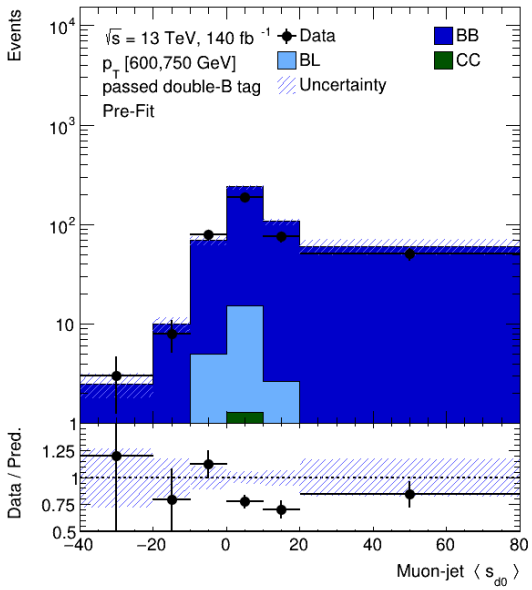
With the exception of the *non-muon jet fail* template, the SV-mass templates generally show relatively good modelling. The former exhibits a clear underestimation of the data in the tail. This effect worsens with decreasing p_{T} and is likely due to a mismodelling of the rate of collimated $g \to b\bar{b}$ radiation in a single jet.



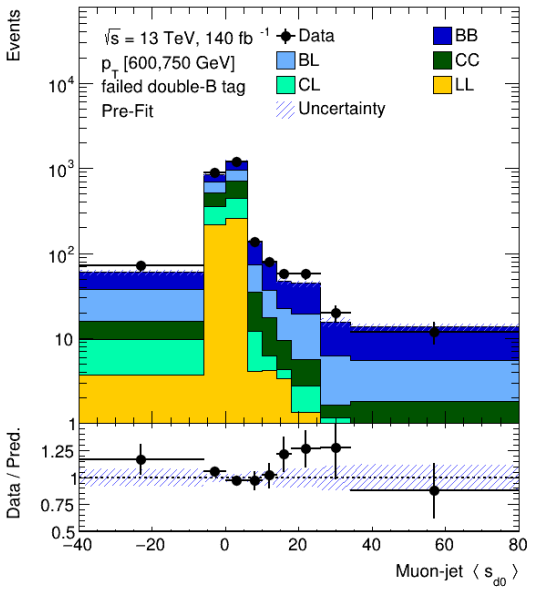
(a) *pass* and $n_\mu = 1$ template.



(b) *fail* and $n_\mu = 1$ template.

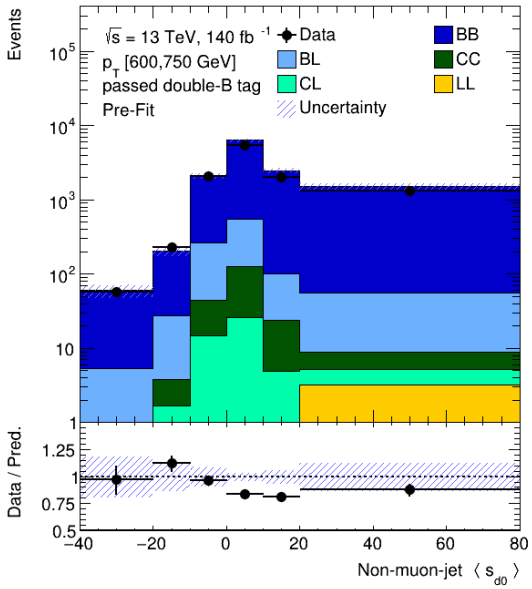


(c) *pass* and $n_\mu = 2$ template.

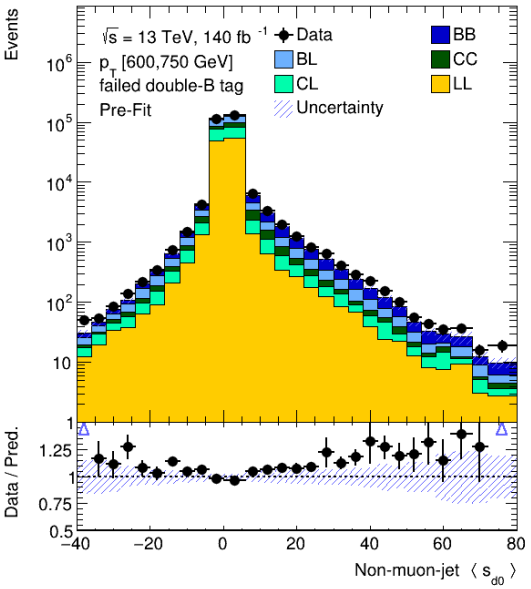


(d) *fail* and $n_\mu = 2$ template.

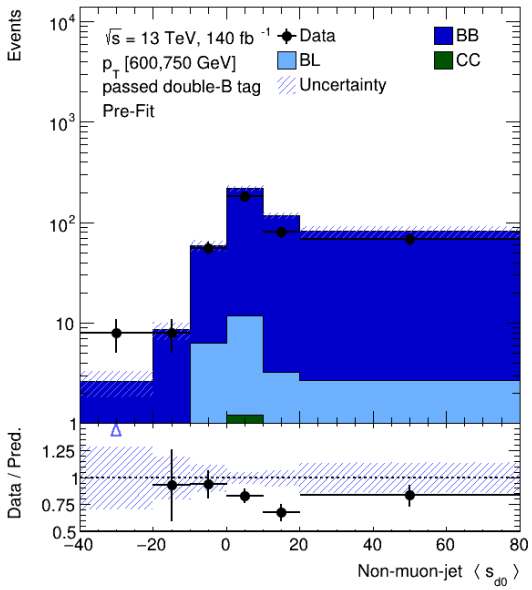
Figure 11.5: The *muon jet* $\langle s_{d_0} \rangle$ prefit templates for the separate muon categories in the 600-750 GeV large-R jet p_{T} bins.



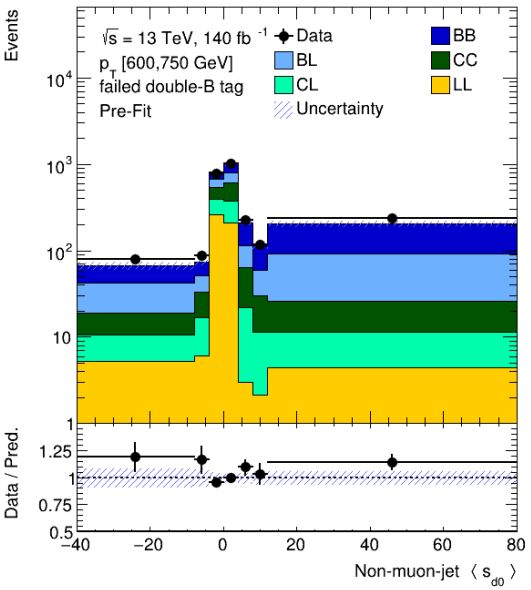
(a) *pass* and $n_\mu = 1$ template.



(b) *fail* and $n_\mu = 1$ template.

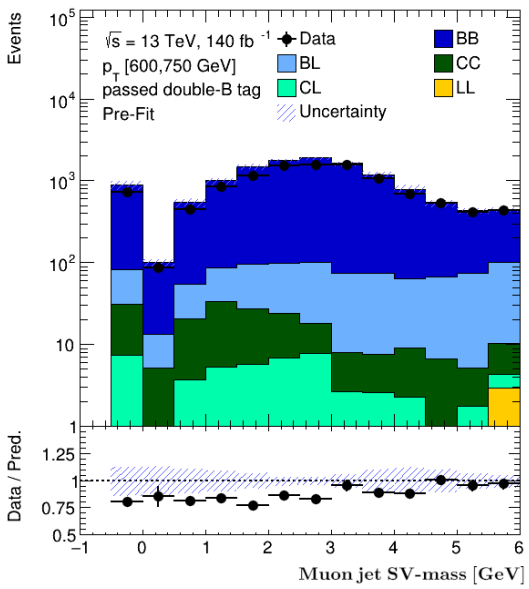


(c) *pass* and $n_\mu = 2$ template.

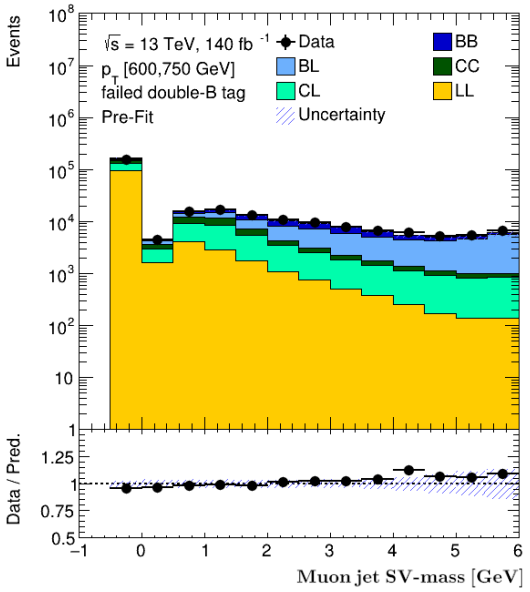


(d) *fail* and $n_\mu = 2$ template.

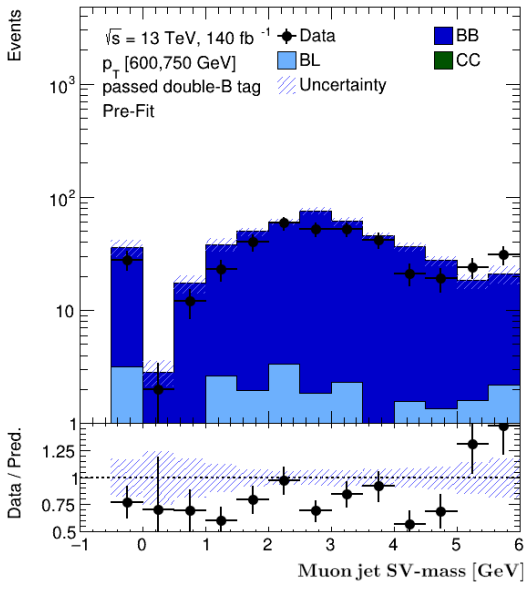
Figure 11.6: The *non-muon jet* $\langle s_{d_0} \rangle$ prefit templates for the separate muon categories in the 600-750 GeV large-R jet p_{T} bins.



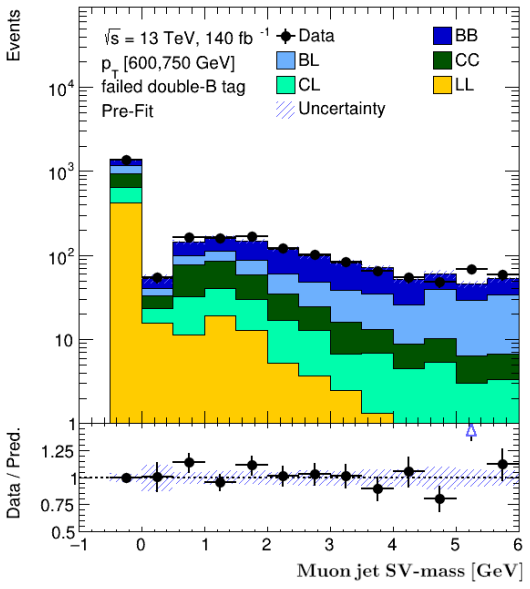
(a) *Pass* and $n_\mu = 1$ template.



(b) *Fail* and $n_\mu = 1$ template.

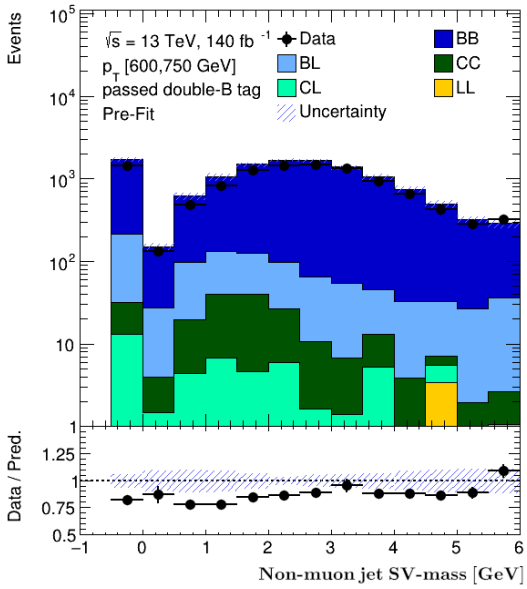


(c) *Pass* and $n_\mu = 2$ template.

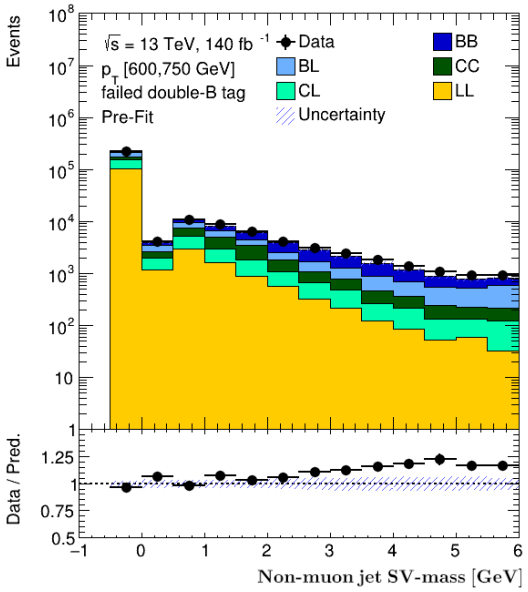


(d) *Fail* and $n_\mu = 2$ template.

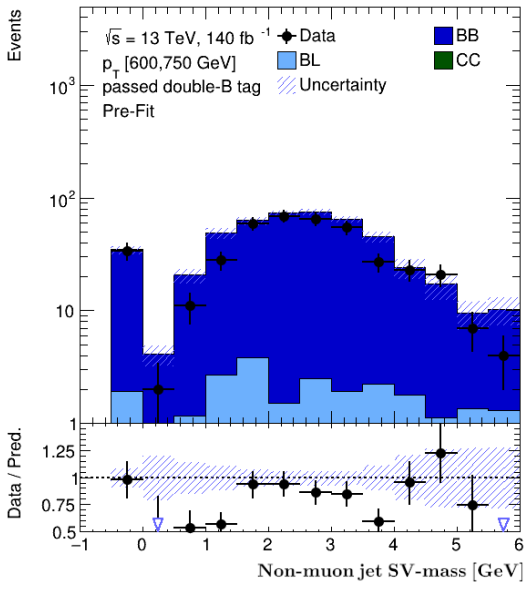
Figure 11.7: The *muon jet* SV-mass prefit templates for the separate muon categories in the 600-750 GeV large-R jet p_{T} bins.



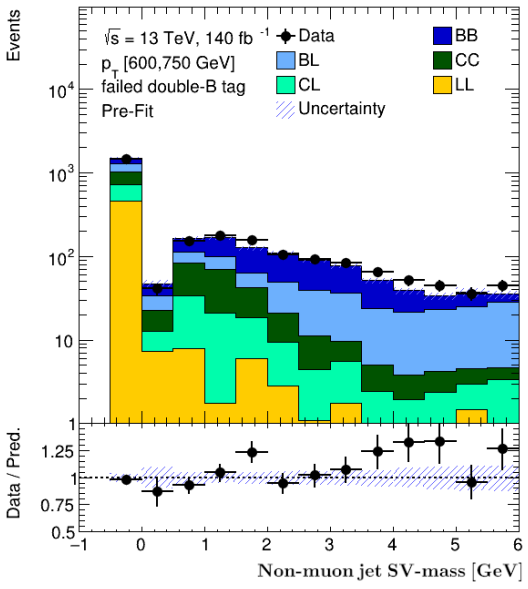
(a) *Pass* and $n_\mu = 1$ template.



(b) *Fail* and $n_\mu = 1$ template.



(c) *Pass* and $n_\mu = 2$ template.



(d) *Fail* and $n_\mu = 2$ template.

Figure 11.8: The *non-muon jet* SV-mass prefit templates for the separate muon categories in the 600-750 GeV large-R jet p_{T} bins.

Chapter 12

Results of the Calibration

Contents

12.1 Overview 208

12.2 Scale Factors and Normalisations 208

12.3 Mismodelling and Nuisance Parameters 211

12.4 The Postfit Templates 214

12.5 Conclusion 218

This chapter presents the preliminary results of the $g \to b\bar{b}$ calibration analysis for the ATLAS boosted $X \to b\bar{b}$ tagger.

12.1 Overview

A simultaneous fit was conducted across all eight regions and in each of the large-R jet p_{T} bins independently. The profile likelihood fit was performed using the TRExFitter framework, with the likelihood maximisation achieved through MINUIT.

This section highlights the results obtained from fitting the $\langle s_{d_0} \rangle$ templates, with a detailed focus on the 600-750 GeV large-R jet p_{T} bin. Additionally, the results from fitting the SV-mass templates are presented for validation and compared to those of the $\langle s_{d_0} \rangle$ fit, but are not discussed in great detail.

As detailed in Section 11.5, fitting the MC templates to data allows for the simultaneous determination of flavour fractions and the efficiency correction factor. It was found that the fit stability improved by fixing the normalisations of the CC and CL flavour fractions to unity. These two flavour fractions are small in both the *pass* and *fail* regions and the fit is struggling to resolve them independently. This results in a strong anti-correlation, which can result in unphysical pulls in opposite directions. The results for the SFs and the normalisations for the dominant flavour fractions BB, BL, and LL are presented in the following.

12.2 Scale Factors and Normalisations

Figure 12.1 presents the derived SFs and flavour fraction normalisations for both the $\langle s_{d_0} \rangle$ and SV-mass template fits. In most of the large-R jet p_{T} bins, the values from the two different template fits agree within their 1σ uncertainties, thereby increasing confidence in the results.

12.2.1 $\langle s_{d_0} \rangle$ Fit

The resultant SFs range between 0.75 and 0.95, with a trend towards lower values at higher large-R jet p_{T}. SFs below unity are expected due to the general overprediction of the data yields in simulation within the *pass* regions. Values below 1.0 are not surprising

since the tagger was trained on simulated data and is thus unlikely to perform better on recorded data.

SFs below unity lead to anti-SFs above unity, which increase the BB fraction in the *fail* regions. However, this increase is insufficient to correct the mismodelling in the *fail* region tails, causing an additional upward pull on the BB normalisation nuisance parameters. This, in turn, results in lower SFs than might be expected when considering only the *pass* regions.

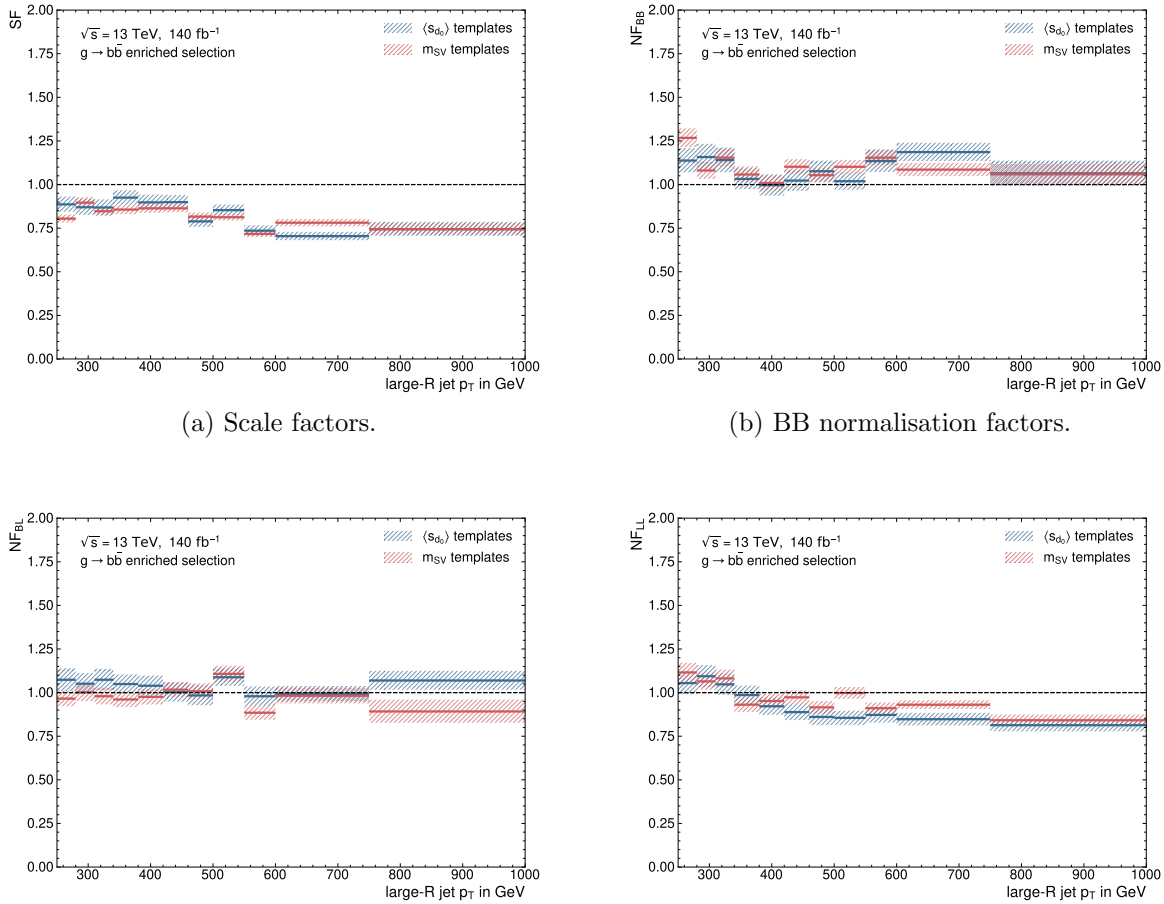
While the overall BB fraction is relatively small in the *fail* region, it is the dominant component in the tails and can thus compensate for much of the mismodelling.

The relatively low SFs, combined with the slightly increased BB normalisations, result in a nominal reduction of the event yield in the *pass* region, which clearly improves the data/MC agreement. However, the shape mismodelling cannot be corrected by a pure normalisation adjustment, leading to the pull of certain nuisance parameters, as discussed in Section 12.3.

While the $n_\mu = 1$ templates are statistically dominant, the underprediction of the event yields is also present in the $n_\mu = 2$ *non-muon jet fail* templates, where the BB fraction is significantly higher. Thus, the increased BB fraction due to the anti-SFs above unity already results in relatively good data/MC agreement, requiring only a slightly increased BB normalisation. This reduces the anti-correlation between the SF and NF_{BB} by around 10%, bringing the postfit values for both quantities slightly closer to unity compared to a fit without the muon multiplicity split.

The BL flavour fraction requires only minor normalisation corrections but shows significant pulls on its shape uncertainties, as discussed in Section 12.3.

The LL component exhibits a slight p_{T} dependence, with normalisations above unity in the low-p_{T} cases and below unity at high p_{T}. The latter is driven by a slight overestimation of the event yield in the centre of the $\langle s_{d_0} \rangle$ templates, where LL is the dominant fraction. The p_{T} dependence results from different mismodelling in the *fail* regions. While the aforementioned U-shaped data/MC ratio is more pronounced in the high-p_{T} bins, the ratio is flatter in the low-p_{T} ones.



(a) Scale factors.

(b) BB normalisation factors.

(c) BL normalisation factors.

(d) LL normalisation factors.

Figure 12.1: Resultant mis-tagging efficiency correction SFs and flavour fraction normalisation factors in all large-R jet p_{T} bins for the 50% WP of the $X \to b\bar{b}$ tagger.

12.2.2 Secondary Vertex Mass Fit

Similar to the fit with the $\langle s_{d_0} \rangle$ templates, the data/MC agreement of the SV-mass templates in the statistically dominant *fail* regions can be improved by increasing the BB normalisations to boost the predicted yields in the tails, and by slightly reducing LL to enhance the modelling of the events where no SV was reconstructed.

While the BL normalisations agree within uncertainties, they exhibit a slightly different trend at low large-R jet p_{T} compared to the $\langle s_{d_0} \rangle$ case. This discrepancy is likely related to the challenges in modelling the fraction of reconstructed jets containing two truth b-jets from $g \to b\bar{b}$ splittings. However, the derived SFs seem to be relatively robust against this effect.

12.3 Mismodelling and Nuisance Parameters

12.3.1 $\langle s_{d_0} \rangle$ Fit

While the normalisations of the individual flavour fractions can correct some mismodelling, primarily in the *fail* regions, they cannot address the shape mismodelling in the $\langle s_{d_0} \rangle$ *pass* regions. The mismodelling of the prefit templates leads to strong pulls on multiple nuisance parameters, complicating the interpretation of the results and challenging the confidence in the derived SFs.

The following is an attempt to explain the pulls in terms of the two main sources of mismodelling observed (see Section 11.7.1). However, it should be noted that the interplay of a large number of nuisance parameters and normalisation factors in a relatively complex fit with eight regions is a challenging task.

Figure 12.2 shows the systematic variations for some of the dominant systematics capturing the mismodelling in the various templates. Additionally, Figure 12.3 ranks the most important systematic uncertainties based on the impact of a postfit $\pm 1\sigma$ pull on the SF.

While most systematics remain unprofiled, there is a pull of the generator-specific nuisance parameters towards the HERWIG shape for the BB and CC templates, and a negative pull for the BL and LL templates. In particular, due to the very pure *pass* regions with decent data statistics, the BB generator systematic is also strongly constrained.

The aforementioned characteristic mismodelling in the *pass* regions is significantly improved by multiple systematic uncertainties, as detailed below. The largest correction stems from the PS and hadronisation uncertainty on the BB template, which, as shown in Figure 12.2(a), captures the shape disagreement between data and simulation very well. Despite this, the corresponding nuisance parameter shows only a pull of approximately $+0.5\sigma$, explained by the shape of the corresponding templates in the *fail* regions, where it would lead to reduced BB fractions in the tails, the opposite of what is needed. This, in turn, leads to the pull of other nuisance parameters to compensate for this effect.

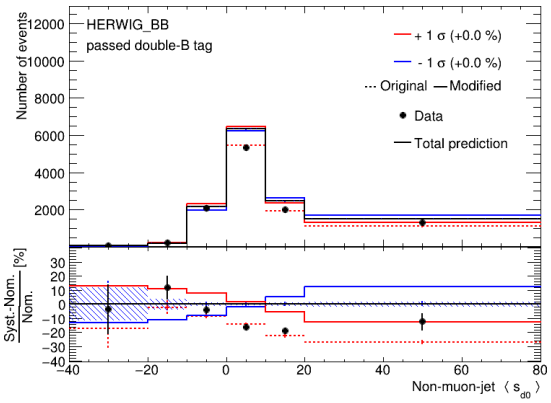
The negative pull of the LL generator uncertainty slightly broadens the peak of the *non-muon jet fail* templates, thus improving the post-fit modelling. The pull on the CC generator uncertainty serves a similar effect, addressing the mismodelling at medium to high absolute values of $\langle s_{d_0} \rangle$, where neither LL nor BB are dominant. Additionally, the CC generator nuisance parameter plays a crucial role in the $n_\mu = 2$ *fail* regions, where it is the sub-dominant component after BB. The BL PS and hadronisation nuisance parameter also plays a role in the $n_\mu = 2$ templates as well as in the tails of the $n_\mu = 1$ *muon-jet fail* template (see Figure 12.2(b)), where it is the dominant uncertainty with a prefit value of around 20%. It is pulled to $\approx -0.5\sigma$, thus significantly increasing the BL component in the tails. A larger negative pull of the BL generator nuisance parameter is prevented as it would worsen the mismodelling observed in the *pass* regions.

In addition to the generator uncertainties, the FSR scale variation and the b-hadron composition nuisance parameters are pulled by around $+1\sigma$. The pull in the first is explained by its shape in *muon jet pass* regions, which captures the observed data/MC ratio very closely, as detailed in Figure 12.2(c). The latter, while being quite small, allows for improved postfit modelling both in the *pass* and *fail* regions.

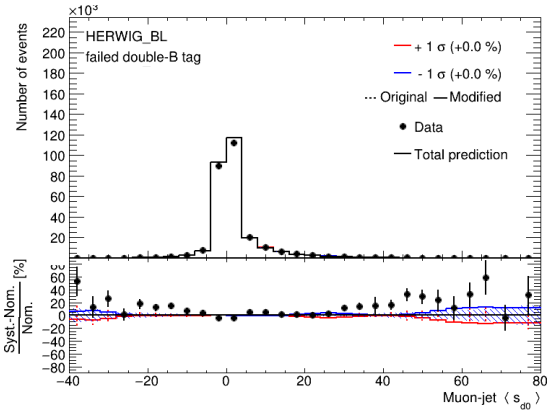
Furthermore, a nuisance parameter associated with the uncertainties on fake rates of tracking in dense environments is pulled to -1σ. It exhibits a very similar (although inverted) shape to the FSR scale variation nuisance parameter, improving the modelling in the *pass* regions (as shown in Figure 12.2(d)) while also increasing the yields in the *fail* tails. All other tracking, jet, and muon uncertainties are smaller, have less impact on the SF, and most of them are either not profiled at all or show very small pulls. It is worth noting that the minimal impact of the assessed tracking uncertainties is particularly surprising, given the substantial observed mismodelling of $\langle s_{d_0} \rangle$.

12.3.2 Secondary Vertex Mass Fit

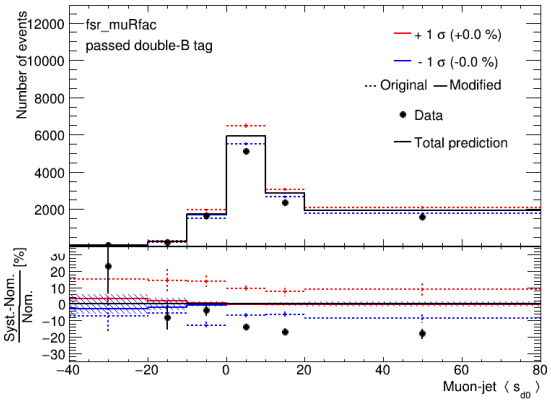
The SV-mass template fit exhibits similar pulls of the PS and hadronisation nuisance parameters. However, as expected, the BL nuisance parameter has the highest impact due to its impact on the modelling at high SV-masses. Additionally, the FSR scale variation



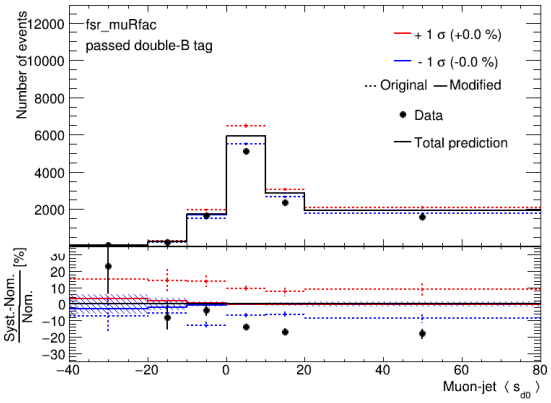
(a) BB PS and hadronisation uncertainty (HERWIG_BB) on $n_\mu = 1$ *non-muon jet pass* combined template.



(b) BL PS and hadronisation uncertainty (HERWIG_BL) on $n_\mu = 1$ *muon jet fail* combined template.



(c) FSR scare variation uncertainty (fsr_muRfac) on $n_\mu = 1$ *muon jet pass* combined template.



(d) Tracking fake rate in dense environments uncertainty (TRK_FAK_RATE_LOOSE_TIDE) on $n_\mu = 1$ *muon jet pass* combined template.

Figure 12.2: The alternative $\langle s_{d_0} \rangle$ templates of some of the dominant uncertainties. The original (dashed) and smoothed (solid) templates corresponding to the $\pm 1\sigma$ variations of the systematic uncertainties are shown for the 600-750 GeV large-R jet p_{T} bin.

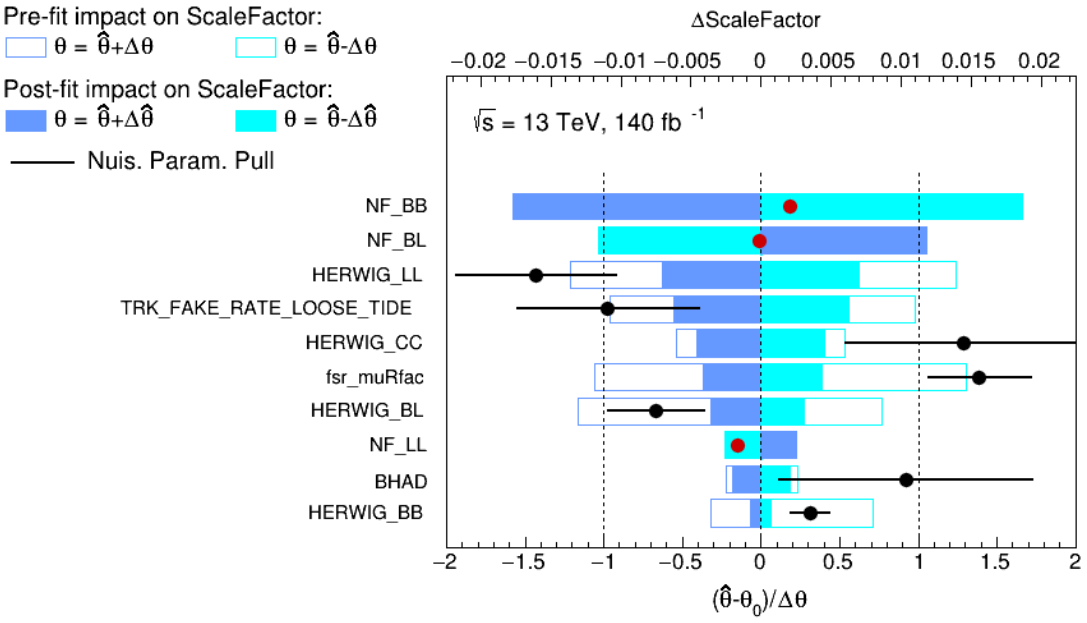


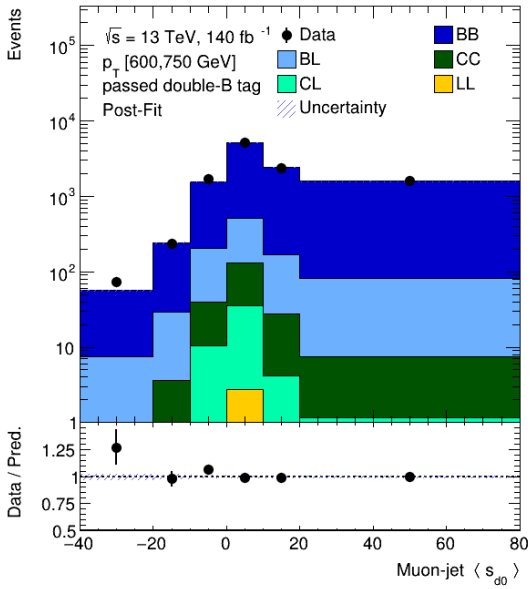
Figure 12.3: Systematic ranking of the nuisance parameters in the $\langle s_{d_0} \rangle$ template fit in the 600-750 GeV bin. The systematic uncertainties are ranked based on the impact of a postfit $\pm 1\sigma$ pull on the SF. The nine highest ranked uncertainties as well as the PS and hadronisation uncertainty for the BB template (HERWIG_BB) are shown.

uncertainty and the b-hadron sample composition uncertainty play significant roles, similar to the $\langle s_{d_0} \rangle$ fit.

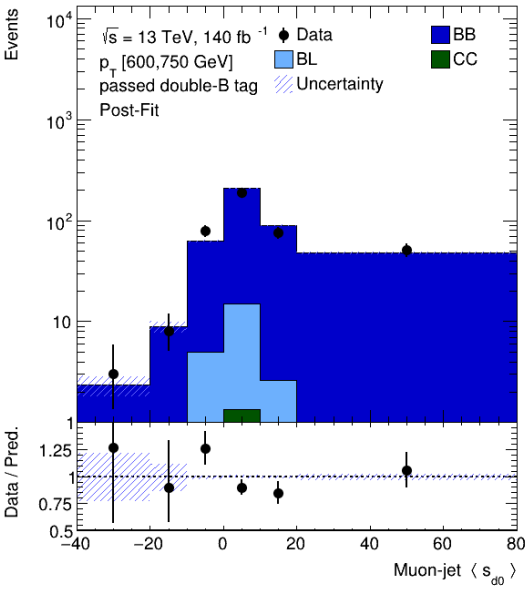
12.4 The Postfit Templates

The interplay between the SF, flavour fraction normalisation corrections, and the profiled nuisance parameters results in decent postfit agreement for most regions. Better agreement is observed for the $n_\mu = 1$ regions compared to the $n_\mu = 2$ regions, which are subject to larger statistical fluctuations. Figures 12.4 and 12.5 show the postfit templates in the 600-750 GeV bin for the *muon jet* and *non-muon jet*, respectively.

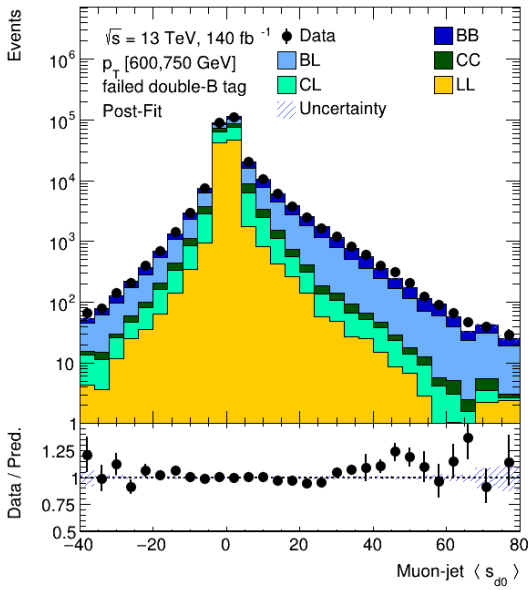
The $n_\mu = 1$ postfit templates for the 600-750 GeV bin of the SV-mass fit are shown in Figure 12.6. The excellent postfit agreement between observed data and simulation strongly motivates further study of this variable for future calibration analyses. See Ref. [3] for a detailed study of the SV-mass templates.



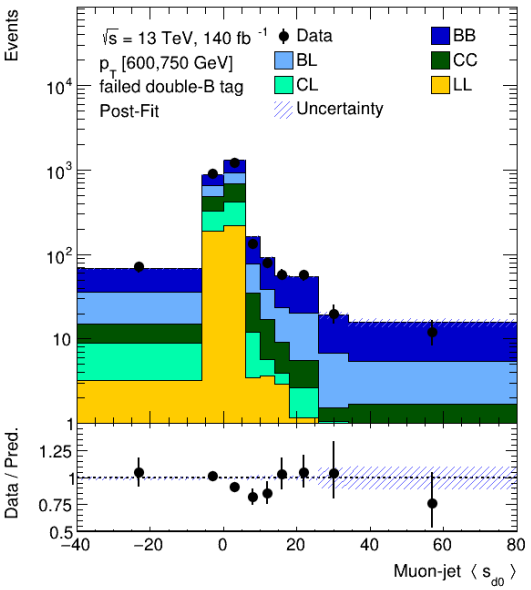
(a) *Muon jet pass* $n_\mu = 1$ template.



(b) *Muon jet pass* $n_\mu = 2$ template.

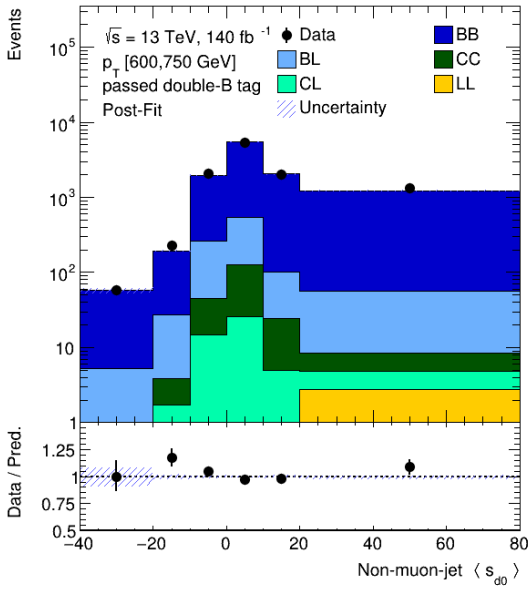


(c) *Muon jet fail* $n_\mu = 1$ template.

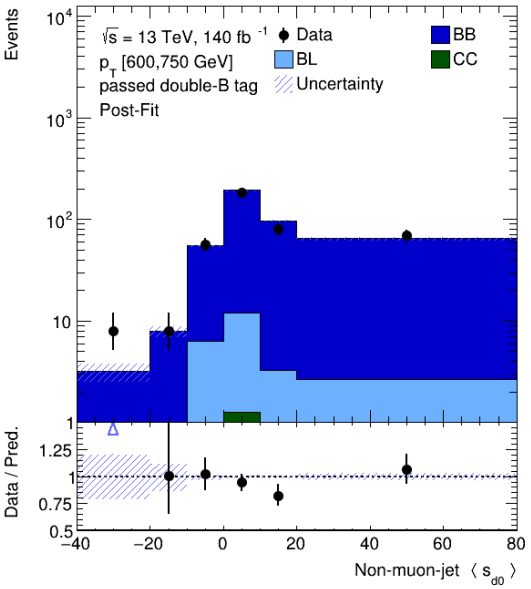


(d) *Muon jet fail* $n_\mu = 2$ template.

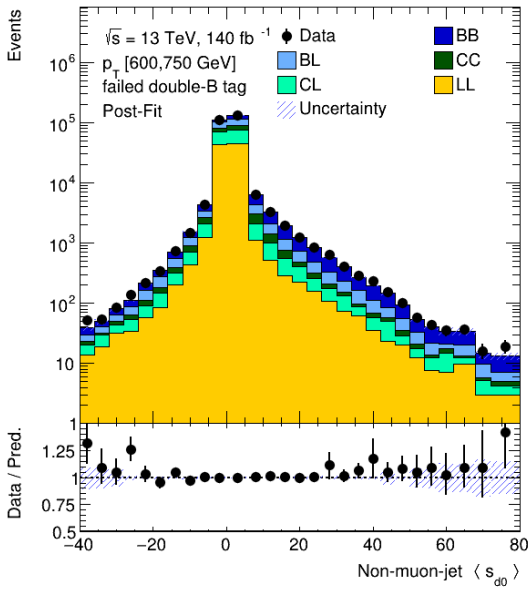
Figure 12.4: The *muon jet* $\langle s_{d_0} \rangle$ postfit templates for the separate muon categories and the 600-750 GeV large-R jet p_{T} bin.



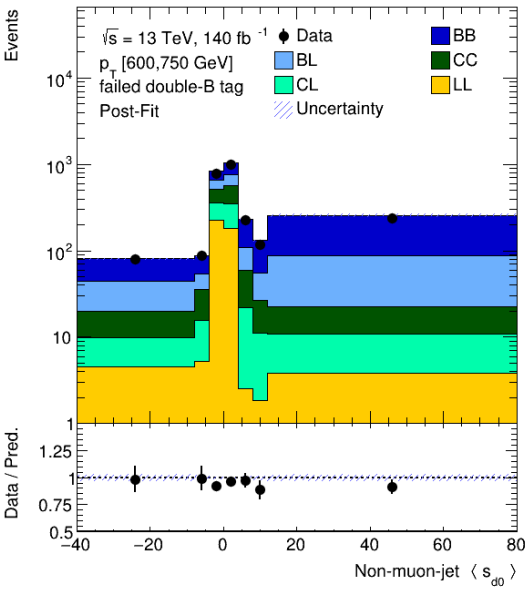
(a) *Non-muon jet pass* $n_\mu = 1$ template.



(b) *Non-muon jet pass* $n_\mu = 2$ template.

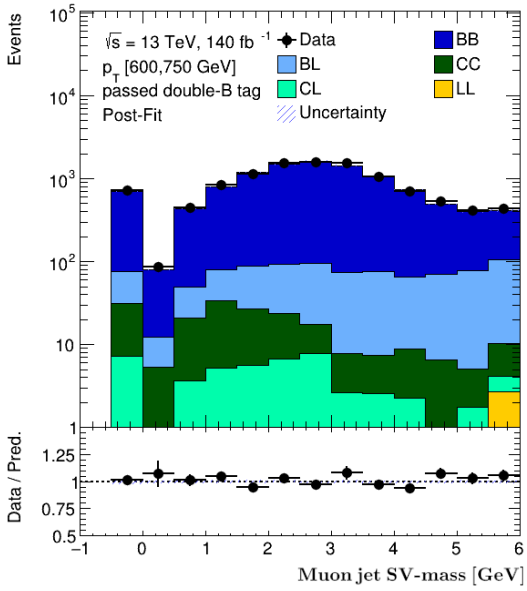


(c) *Non-muon jet fail* $n_\mu = 1$ template.

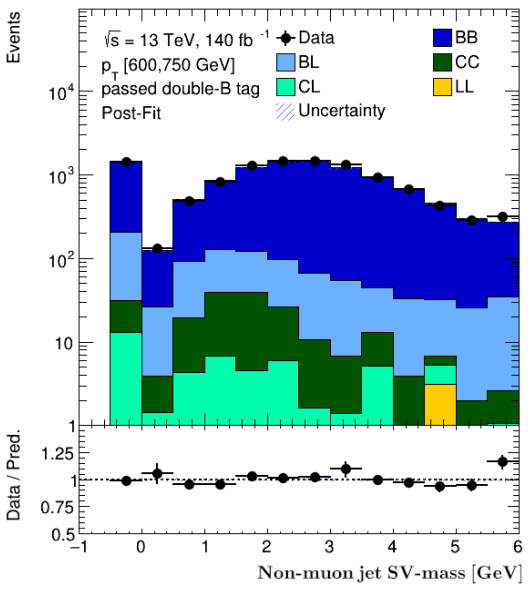


(d) *Non-muon jet fail* $n_\mu = 2$ template.

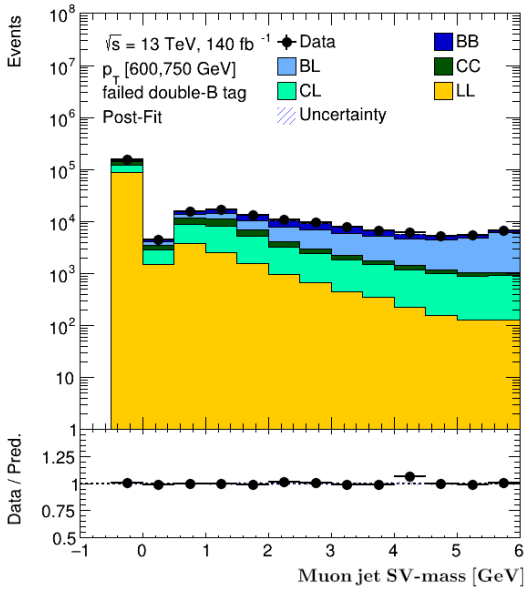
Figure 12.5: The *non-muon jet* $\langle s_{d_0} \rangle$ postfit templates for the separate muon categories and the 600-750 GeV large-R jet p_{T} bin.



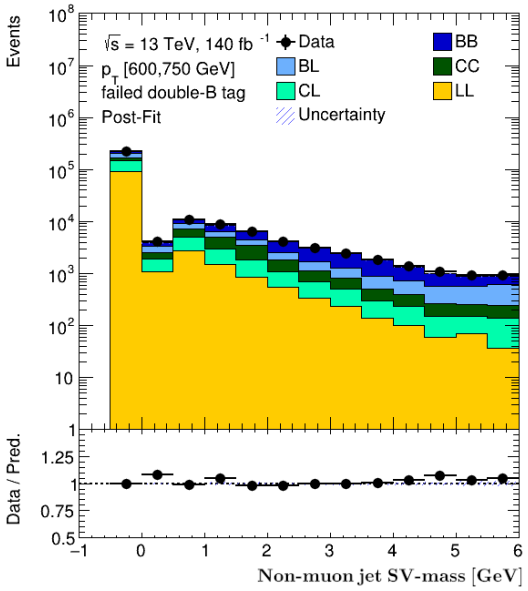
(a) *Muon jet pass* $n_\mu = 1$ template.



(b) *Non-muon jet pass* $n_\mu = 2$ template.



(c) *Muon jet fail* $n_\mu = 1$ template.



(d) *Non-muon jet fail* $n_\mu = 2$ template.

Figure 12.6: The $n_\mu = 1$ SV-mass postfit templates in the 600-750 GeV large-R jet p_{T} bin.

12.5 Conclusion

This chapter presented the multijet mistag efficiency correction SFs for the ATLAS boosted $X \to b\bar{b}$ tagger, obtained through the $g \to b\bar{b}$ calibration. Flavour-discriminating templates were used to correct the flavour composition of the multijet events, and SFs were derived over a wide range of large-R jet p_{T} up to 1 TeV. The resulting SFs and flavour composition corrections are shown in Figure 12.1. Despite significant challenges posed by modelling issues of the crucial flavour-discriminating template variables, the agreement of the results across two different setups enhances confidence in the developed calibration methods. However, these mismodelling issues resulted in a lengthy and complex calibration analysis, coming too late for Run 2 legacy analyses.

Nevertheless, the ideas and concepts developed for the $g \to b\bar{b}$ calibration will provide valuable insights, serving as a blueprint for future calibration efforts. Additionally, the results indicate the necessity for dedicated tracking uncertainties to cover the observed mismodelling in these boosted topologies, as well as improvements in the modelling of multijet events, which are expected to accelerate calibration efforts for future taggers.

The recent rapid advancements in machine learning based tagging algorithms, such as those using GNNs, will require calibrations for their use in Run 3. With more data and higher centre-of-mass energies of up to 14 TeV, high-performing calibrated tagging algorithms are essential for both SM measurements, such as those involving high-p_{T} Higgs bosons, and BSM searches predicting di-Higgs production, new scalars, and heavy resonances.

Thesis Summary and Outlook

This thesis begins with an introduction to the SM, outlining its successes and limitations, followed by an exploration of the dark-Higgs model and other scenarios predicting a di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ final state. Part I also provides an overview of the ATLAS experiment, introducing the detector, event simulation, and physics object reconstruction. A broad overview of statistical and ML methods used in HEP is discussed in Part II. Part III presents the author's primary thesis project, a search for DM optimised for the dark-Higgs model. The thesis concludes with Part IV, detailing a calibration analysis for an algorithm designed to tag large-R jets enclosing two b-flavoured subjets originating from the decay of a Higgs boson or heavy resonances in BSM theories.

Despite constituting approximately 85% of the matter in the universe, the nature of DM remains one of the great mysteries in modern physics. Among the various theoretical frameworks proposed to explain DM, dark-Higgs models are particularly compelling as they naturally satisfy the observed relic density and provide a mass-generation mechanism for DM, while avoiding current experimental constraints. While this analysis is optimised for the dark-Higgs strahlung process, the resulting resonant di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ signature represents a broader phenomenological feature that has not yet been examined at the LHC. This search is not only the first dedicated effort to target this generic signature, but it also significantly extends the exclusion limits on dark-Higgs models. As shown in Figure 9.6, the limits extend up to 2.4 TeV in $m_{Z'}$ and 405 GeV in m_s. Specifically, for $m_s > 250$ GeV, the observed limits have been extended by several hundred GeV towards higher $m_{Z'}$, with greater improvements at higher m_s. This strong sensitivity was achieved through the innovative use of ML algorithms and a sophisticated background estimation strategy. Physical interpretability was ensured by fitting the di-Higgs mass and decorrelating it from the classifier discriminant, providing a novel blueprint for leveraging the power of ML in binary classification while preserving interpretability and avoiding reliance on uninterpretable discriminant distributions. However, the presented search

marks only the beginning of investigations into this decay topology, with higher sensitivities anticipated in future analyses that will have access to more data and advanced ML b-tagging algorithms. Additionally, the analysis would benefit from improved modelling of top-related background processes in the targeted phase space, and could be extended to explore different parameter choices for the dark-Higgs model, particularly higher DM masses. In general, the di-Higgs + $E_{\mathrm{T}}^{\mathrm{miss}}$ channel is an important signature for DM searches, as the Higgs boson naturally provides a portal for interactions with DM, allowing the study of extended Higgs and scalar sectors. Looking ahead to Run-3 and beyond, increased luminosity and refined analysis tools will further enhance the sensitivity of these searches. The High-Luminosity LHC, which will deliver ten times the integrated luminosity of Run-3, combined with significantly upgraded detectors, is expected to provide even greater discovery sensitivity on DM models. Additionally, potential future collider experiments, such as the Future Circular Collider, will enable precision measurements of the Higgs boson and direct searches for low-mass scalar particles. These collider-based searches will also benefit from complementarity with direct and indirect DM detection experiments, jointly constraining the parameter space of DM models and providing a more comprehensive understanding of DM interactions.

The presented multijet mistag calibration analysis for the ATLAS $X \to b\bar{b}$ tagger derived efficiency correction scale factors across a wide range of large-R jet p_{T}, as shown in Figure 12.1(a). The scale factors were extracted from a template fit of flavour-sensitive distributions, such as the mean transverse impact parameter significance, which allows for the correction of the flavour composition in the utilised dijet samples. Although mismodelling issues of key variables presented significant challenges, the stability of the derived efficiency correction factors enhances confidence in the results and methods developed. However, improved modelling of the impact parameter significance and the development of dedicated systematic uncertainties in this boosted high-p_{T} phase space would greatly reduce these challenges and benefit the calibration analysis. Consequently, this calibration serves as a blueprint for future calibrations of even more advanced b-tagging algorithms, which will be applied to data collected in Run 3.

Bibliography

[1] G. Aad et al., *Search for pair production of higgsinos in events with two Higgs bosons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions at the ATLAS experiment*, Phys. Rev. D **109** (11 2024) 112011, URL: https://link.aps.org/doi/10.1103/PhysRevD.109.112011 (cit. on pp. 1, 2, 41, 119, 128, 130, 135, 172, 177).

[2] *Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in pp collisions at $\sqrt{s} = 13$ TeV 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-2021-035: CERN, 2021, URL: http://cds.cern.ch/record/2777811 (cit. on pp. 2, 186).

[3] I. Veliscek, *The search for pair production of Higgs Bosons at the ATLAS experiment*, University of Oxford, 2023, URL: http://cds.cern.ch/record/2905691 (cit. on pp. 2, 214).

[4] P. A. M. Dirac, *The quantum theory of the electron*, Proc. Roy. Soc. Lond. A **117** (1928) 610 (cit. on p. 5).

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, ed. by J. C. Taylor (cit. on p. 5).

[6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, ed. by J. C. Taylor (cit. on p. 5).

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585, ed. by J. C. Taylor (cit. on p. 5).

[8] M. Thomson, *Modern particle physics*, New York: Cambridge University Press, 2013, ISBN: 978-1-107-03426-6 (cit. on pp. 5, 62).

[9] C.-N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191, ed. by J.-P. Hsu and D. Fine (cit. on p. 6).

[10] E. Noether, *Invariante Variationsprobleme*, ger, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235, URL: http://eudml.org/doc/59024 (cit. on p. 6).

[11] S. Tomonaga,
On a relativistically invariant formulation of the quantum theory of wave fields,
Prog. Theor. Phys. **1** (1946) 27 (cit. on p. 7).

[12] J. S. Schwinger, *Quantum electrodynamics. I A covariant formulation*,
Phys. Rev. **74** (1948) 1439, ed. by K. A. Milton (cit. on p. 7).

[13] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*,
Phys. Rev. **76** (6 1949) 769,
URL: https://link.aps.org/doi/10.1103/PhysRev.76.769 (cit. on p. 7).

[14] S. L. Glashow, *Partial Symmetries of Weak Interactions*,
Nucl. Phys. **22** (1961) 579 (cit. on p. 8).

[15] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264 (cit. on p. 8).

[16] A. Salam, *Weak and Electromagnetic Interactions*,
Conf. Proc. C **680519** (1968) 367 (cit. on p. 8).

[17] ATLAS and CMS Collaborations,
Combination of measurements of the top quark mass from data collected by the ATLAS and CMS experiments at $\sqrt{s} = 7$ and 8 TeV, (2023),
arXiv: 2402.08713 [hep-ex] (cit. on p. 12).

[18] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*,
Phys. Rev. Lett. **30** (1973) 1346, ed. by J. C. Taylor (cit. on p. 13).

[19] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*,
Phys. Rev. Lett. **30** (1973) 1343, ed. by J. C. Taylor (cit. on p. 13).

[20] ATLAS Collaboration, *A precise determination of the strong-coupling constant from the recoil of Z bosons with the ATLAS experiment at $\sqrt{s} = 8$ TeV*, (2023),
arXiv: 2309.12986 [hep-ex] (cit. on p. 15).

[21] G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
arXiv: 1207.7214 [hep-ex] (cit. on p. 16).

[22] S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30,
arXiv: 1207.7235 [hep-ex] (cit. on p. 16).

[23] *Standard Model Summary Plots October 2023*, tech. rep.,
All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-039: CERN, 2023, URL: http://cds.cern.ch/record/2882448
(cit. on p. 17).

[24] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01
(cit. on pp. 18, 74).

[25] S. T. Petcov, *The Nature of Massive Neutrinos*,
Adv. High Energy Phys. **2013** (2013) 852987, arXiv: 1303.5819 [hep-ph]
(cit. on p. 19).

[26] Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*,
Phys. Rev. Lett. **81** (1998) 1562, arXiv: hep-ex/9807003 (cit. on p. 19).

[27] Q. R. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, arXiv: nucl-ex/0204008 (cit. on p. 19).

[28] M. Gell-Mann, P. Ramond, and R. Slansky, *Complex Spinors and Unified Theories*, Conf. Proc. C **790927** (1979) 315, arXiv: 1306.4669 [hep-th] (cit. on p. 20).

[29] R. N. Mohapatra and G. Senjanovic, *Neutrino Mass and Spontaneous Parity Nonconservation*, Phys. Rev. Lett. **44** (1980) 912 (cit. on p. 20).

[30] S. Abe et al., *Search for the Majorana Nature of Neutrinos in the Inverted Mass Ordering Region with KamLAND-Zen*, Phys. Rev. Lett. **130** (2023) 051801, arXiv: 2203.02139 [hep-ex] (cit. on p. 21).

[31] C. Fengler, *Overview of KATRIN Results on the Neutrino Mass and New Physics Searches*, PoS **DISCRETE2022** (2024) 011 (cit. on p. 21).

[32] M. Aker et al., *Direct neutrino-mass measurement based on 259 days of KATRIN data*, 2024, arXiv: 2406.13516 [nucl-ex], URL: https://arxiv.org/abs/2406.13516 (cit. on p. 21).

[33] B. Abi et al., *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE*, JINST **15** (2020) T08008, arXiv: 2002.02967 [physics.ins-det] (cit. on p. 21).

[34] K. Abe et al., *Hyper-Kamiokande Design Report*, (2018), arXiv: 1805.04163 [physics.ins-det] (cit. on p. 21).

[35] G. Bhattacharyya, *Hierarchy problem and BSM physics*, Pramana **89** (2017) 53 (cit. on p. 21).

[36] H. Georgi and S. L. Glashow, *Unity of All Elementary Particle Forces*, Phys. Rev. Lett. **32** (1974) 438 (cit. on p. 23).

[37] A. Einstein, *Die Feldgleichungen der Gravitation*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (1915) 844 (cit. on p. 23).

[38] A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Annalen der Physik **354** (1916) 769 (cit. on p. 23).

[39] A. Zee, *Einstein Gravity in a Nutshell*, New Jersey: Princeton University Press, 2013, ISBN: 978-0-691-14558-7 (cit. on p. 23).

[40] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory: 25th Anniversary Edition*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2012 (cit. on p. 23).

[41] A. Ashtekar and E. Bianchi, *A short review of loop quantum gravity*, Reports on Progress in Physics **84** (2021) 042001, URL: https://dx.doi.org/10.1088/1361-6633/abed91 (cit. on p. 23).

[42] C. Abel et al.,
Measurement of the Permanent Electric Dipole Moment of the Neutron,
Phys. Rev. Lett. **124** (2020) 081803, arXiv: 2001.11966 [hep-ex] (cit. on p. 24).

[43] R. D. Peccei and H. R. Quinn,
Constraints Imposed by CP Conservation in the Presence of Instantons,
Phys. Rev. D **16** (1977) 1791 (cit. on p. 24).

[44] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*,
Phys. Rev. Lett. **38** (1977) 1440 (cit. on p. 24).

[45] A. D. Sakharov,
Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe,
Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32 (cit. on p. 25).

[46] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay,
Evidence for the 2π Decay of the K_2^0 Meson, Phys. Rev. Lett. **13** (1964) 138
(cit. on p. 25).

[47] S. Vegetti et al., *Strong gravitational lensing as a probe of dark matter*, 2023,
arXiv: 2306.11781 [astro-ph.CO] (cit. on p. 25).

[48] 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/,
Astrophys. J. **238** (1980) 471 (cit. on p. 25).

[49] A. Challinor, *CMB anisotropy science: a review*,
Proceedings of the International Astronomical Union **8** (2012) 42, ISSN: 1743-9221,
URL: http://dx.doi.org/10.1017/S1743921312016663 (cit. on p. 25).

[50] M. Milgrom, *A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis*, Astrophys. J. **270** (1983) 365 (cit. on p. 25).

[51] Planck Collaboration, *Planck 2018 results - VI. Cosmological parameters*,
Astronomy & Astrophysics **641** (2020) A6,
URL: https://doi.org/10.1051/0004-6361/201833910 (cit. on p. 25).

[52] A. Arbey and F. Mahmoudi, *Dark matter and the early Universe: A review*,
Progress in Particle and Nuclear Physics **119** (2021) 103865, ISSN: 0146-6410,
URL: http://dx.doi.org/10.1016/j.ppnp.2021.103865 (cit. on p. 26).

[53] Markevitch, M. and others, *Composite image of the Bullet Cluster*,
https://example.com/bullet_cluster_image,
X-ray: NASA/CXC/CfA/M.Markevitch et al.; Optical: NASA/STScI;
Magellan/U.Arizona/D.Clowe et al.; Lensing Map: NASA/STScI; ESO WFI;
Magellan/U.Arizona/D.Clowe et al., 2006 (cit. on p. 28).

[54] E. Aprile et al.,
First Dark Matter Search with Nuclear Recoils from the XENONnT Experiment,
Phys. Rev. Lett. **131** (2023) 041003, arXiv: 2303.14729 [hep-ex] (cit. on p. 28).

[55] J. Aalbers et al.,
First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment,
Phys. Rev. Lett. **131** (2023) 041002, arXiv: 2207.03764 [hep-ex] (cit. on p. 28).

[56] X. Cao et al., *PandaX: A Liquid Xenon Dark Matter Experiment at CJPL*, Sci. China Phys. Mech. Astron. **57** (2014) 1476, arXiv: 1405.2882 [physics.ins-det] (cit. on p. 28).

[57] E. W. Kolb, D. J. H. Chung, and A. Riotto, *WIMPZILLAS!*, 1998, arXiv: hep-ph/9810361 [hep-ph], URL: https://arxiv.org/abs/hep-ph/9810361 (cit. on p. 30).

[58] M. Fabbrichesi, E. Gabrielli, and G. Lanfranchi, *The Physics of the Dark Photon: A Primer*, Springer International Publishing, 2021, ISBN: 9783030625191, URL: http://dx.doi.org/10.1007/978-3-030-62519-1 (cit. on p. 30).

[59] G. Aad et al., *Combination of Searches for Higgs Boson Pair Production in pp Collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS Detector*, Phys. Rev. Lett. **133** (10 2024) 101801, URL: https://link.aps.org/doi/10.1103/PhysRevLett.133.101801 (cit. on p. 31).

[60] *Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at sqrt(s) = 13 TeV*, tech. rep., CERN, 2024, URL: https://cds.cern.ch/record/2917252 (cit. on p. 31).

[61] J. Abdallah et al., *Simplified Models for Dark Matter Searches at the LHC*, Phys. Dark Univ. **9-10** (2015) 8, arXiv: 1506.03116 [hep-ph] (cit. on p. 33).

[62] M. Duerr, F. Kahlhoefer, K. Schmidt-Hoberg, T. Schwetz, and S. Vogl, *How to save the WIMP: global analysis of a dark matter model with two s-channel mediators*, Journal of High Energy Physics **2016** (2016), URL: https://doi.org/10.1007/jhep09(2016)042 (cit. on pp. 34, 36).

[63] A. Pais, *Remark on baryon conservation*, Phys. Rev. D **8** (1973) 1844 (cit. on p. 35).

[64] G. Aad et al., *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*, Phys. Rev. D **103** (2021) 112006, arXiv: 2102.10874 [hep-ex] (cit. on p. 36).

[65] G. Aad et al., *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022) 52, [Erratum: Nature 612, E24 (2022)], arXiv: 2207.00092 [hep-ex] (cit. on p. 36).

[66] M. Duerr et al., *Hunting the dark Higgs*, Journal of High Energy Physics **2017** (2017), URL: https://doi.org/10.1007%5C%2Fjhep04%5C%282017%5C%29143 (cit. on p. 38).

[67] *Search for dark matter produced in association with a dark Higgs boson in the* $b\bar{b}$ *final state using* $\sqrt{s} = 13$ *TeV collisions recorded 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-2024-004: CERN, 2024, URL: https://cds.cern.ch/record/2893855 (cit. on pp. 38, 39).

[68] G. Aad, B. Abbott, and D. C. Abbott, *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*, Phys. Rev. Lett. **126** (2021) 121802, 32 pages in total, author list starting page 16, 4 figures, 1 table, submitted to Physical Review Letters. All figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/ EXOT-2018-40/, arXiv: 2010.06548, URL: https://cds.cern.ch/record/2741385 (cit. on pp. 38, 39).

[69] A. 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*, 2022, arXiv: 2211.07175 [hep-ex] (cit. on p. 39).

[70] *Dark matter summary plots for s-channel, 2HDM+a, Higgs portal and Dark Higgs models*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-010: CERN, 2024, URL: http://cds.cern.ch/record/2901527 (cit. on p. 39).

[71] S. P. MARTIN, "A SUPERSYMMETRY PRIMER", *Advanced Series on Directions in High Energy Physics*, WORLD SCIENTIFIC, 1998 1, URL: http://dx.doi.org/10.1142/9789812839657_0001 (cit. on p. 41).

[72] M. Blanke, S. Kast, J. M. Thompson, S. Westhoff, and J. Zurita, *Spotting hidden sectors with Higgs binoculars*, JHEP **04** (2019) 160, arXiv: 1901.07558 [hep-ph] (cit. on p. 42).

[73] CERN, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, en, 2017, URL: https://e-publishing.cern.ch/index.php/CYRM/issue/view/32 (cit. on p. 44).

[74] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08001 (cit. on p. 46).

[75] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 46).

[76] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, (2022), General Photo, URL: https://cds.cern.ch/record/2800984 (cit. on p. 48).

[77] S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 49).

[78] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 49).

[79] A. A. Alves Jr. et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 49).

[80] G. Anelli et al., *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007 (cit. on p. 49).

[81] O. Adriani et al., *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) S08006 (cit. on p. 49).

[82] A. Ariga et al., *Technical Proposal for FASER: ForwArd Search ExpeRiment at the LHC*, (2018), arXiv: 1812.09139 [physics.ins-det] (cit. on p. 49).

[83] S. van der Meer, *Calibration of the Effective Beam Height in the ISR*, (1968) (cit. on p. 50).

[84] R. Steerenberg et al., *Operation and performance of the CERN Large Hadron Collider during proton Run 2*, (2019) MOPMP031, URL: https://cds.cern.ch/record/2696126 (cit. on p. 50).

[85] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 50).

[86] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **83** (2023) 982, arXiv: 2212.09379 [hep-ex] (cit. on p. 50).

[87] A. Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022), URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (cit. on p. 51).

[88] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 53).

[89] G. Strong, *On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case*, Machine Learning: Science and Technology **1** (2020) (cit. on p. 53).

[90] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector", 2013, URL: https://cds.cern.ch/record/1505342 (cit. on p. 56).

[91] B. Abbott et al., *Production and Integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 58).

[92] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run 2*, Eur. Phys. J. C **80** (2020) 1194, arXiv: 2007.07624 [hep-ex] (cit. on p. 59).

[93] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, tech. rep., 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 58).

[94] A. Collaboration, "Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment", General Photo, 2020, URL: https://cds.cern.ch/record/2723878 (cit. on p. 60).

[95] A. Abdesselam and T. Akimoto, *The Barrel Modules of the ATLAS SemiConductor Tracker*, tech. rep., CERN, 2006, URL: https://cds.cern.ch/record/974073 (cit. on p. 60).

[96] A. Vogel, *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*, tech. rep., CERN, 2013, URL: https://cds.cern.ch/record/1537991 (cit. on p. 61).

[97] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", 2008, URL: https://cds.cern.ch/record/1095927 (cit. on p. 62).

[98] A. Airapetian et al., *ATLAS: Detector and physics performance technical design report. Volume 1*, (1999) (cit. on p. 63).

[99] *ATLAS muon spectrometer: Technical design report*, (1997) (cit. on p. 64).

[100] H. Bethe, *Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie.*, Annalen Phys. **5** (1930) 325 (cit. on p. 64).

[101] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv: 2004.13447 [physics.ins-det] (cit. on pp. 65, 67).

[102] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, JINST **15** (2020) P10004, arXiv: 2007.12539 [hep-ex] (cit. on p. 67).

[103] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 67).

[104] R. Achenbach et al., *The ATLAS Level-1 Calorimeter Trigger*, Journal of Instrumentation **3** (2008) P03001, URL: https://dx.doi.org/10.1088/1748-0221/3/03/P03001 (cit. on p. 67).

[105] R. Simoniello, *The ATLAS Level-1 Topological Processor: from design to routine usage in Run-2*, tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2649959 (cit. on p. 67).

[106] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, 2022, arXiv: 2203.11601 [hep-ph] (cit. on pp. 69, 70, 74, 120, 187).

[107] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on pp. 69, 120).

[108] A. Buckley et al., *General-purpose event generators for LHC physics*, Physics Reports **504** (2011) 145, ISSN: 0370-1573, URL: https://www.sciencedirect.com/science/article/pii/S0370157311000846 (cit. on pp. 69, 74).

[109] S. Höche, *Introduction to parton-shower event generators*, 2015, arXiv: 1411.4085 [hep-ph] (cit. on pp. 71, 72).

[110] 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*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on pp. 71, 120, 124).

[111] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on pp. 71, 123).

[112] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on pp. 71, 123).

[113] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on pp. 71, 123).

[114] V. N. Gribov and L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438 (cit. on p. 72).

[115] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126** (1977) 298 (cit. on p. 72).

[116] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics*, Sov. Phys. JETP **46** (1977) 641 (cit. on p. 72).

[117] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation and string dynamics*, Phys. Rept. **97** (1983) 31 (cit. on p. 73).

[118] T. Sjöstrand, *Jet fragmentation of multiparton configurations in a string framework*, Nucl. Phys. B **248** (1984) 469 (cit. on p. 73).

[119] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. B **238** (1984) 492 (cit. on p. 73).

[120] M. Bähr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph] (cit. on pp. 73, 187).

[121] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on pp. 73, 187).

[122] B. Webber, *Parton shower Monte Carlo event generators*, Scholarpedia **6** (2011) 10662 (cit. on pp. 73, 74).

[123] S. Agostinelli et al., GEANT4 – *a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on pp. 75, 120, 187).

[124] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [physics.ins-det] (cit. on pp. 76, 126).

[125] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on pp. 77, 200).

[126] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex] (cit. on pp. 77, 88, 89, 126, 200).

[127] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, arXiv: 1611.10235 [hep-ex] (cit. on p. 77).

[128] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 78).

[129] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 80).

[130] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 80).

[131] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **09** (2013) 076, arXiv: 1306.4945 [hep-ex] (cit. on p. 80).

[132] M. Cacciari, G. P. Salam, and G. Soyez, *The catchment area of jets*, Journal of High Energy Physics **2008** (2008) 005, URL: https://dx.doi.org/10.1088/1126-6708/2008/04/005 (cit. on p. 80).

[133] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Physics Letters B **659** (2008) 119, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269307011094 (cit. on p. 80).

[134] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Physical Review D **48** (1993) 3160, ISSN: 0556-2821, URL: http://dx.doi.org/10.1103/PhysRevD.48.3160 (cit. on p. 81).

[135] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*, 1999, arXiv: hep-ph/9907280 [hep-ph] (cit. on p. 81).

[136] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 689, arXiv: 2007.02645 [hep-ex] (cit. on pp. 82, 83, 85, 165).

[137] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on pp. 86, 87, 165).

[138] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Eur. Phys. J. C **83** (2023) 681, arXiv: 2211.16345 [physics.data-an] (cit. on pp. 86, 88).

[139] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-013: CERN, 2017, URL: https://cds.cern.ch/record/2273281 (cit. on p. 86).

[140] *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: CERN, 2017, URL: https://cds.cern.ch/record/2270366 (cit. on p. 86).

[141] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-PUB-2018-025, 2018, URL: https://cds.cern.ch/record/2645405 (cit. on p. 87).

[142] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444 (cit. on p. 87).

[143] *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003: CERN, 2017, URL: https://cds.cern.ch/record/2255226 (cit. on p. 87).

[144] ATLAS Collaboration, *Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **83** (2023) 728, arXiv: 2301.06319 [hep-ex] (cit. on p. 87).

[145] ATLAS Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, Eur. Phys. J. C **82** (2022) 95, arXiv: 2109.10627 [hep-ex] (cit. on p. 87).

[146] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex] (cit. on pp. 90, 91, 165).

[147] L. Lista, "Practical Statistics for Particle Physicists", *2016 European School of High-Energy Physics*, 2017 213, arXiv: 1609.04150 [physics.data-an] (cit. on p. 93).

[148] L. Lista, *Statistical Methods for Data Analysis in Particle Physics*, vol. 909, Springer, 2016, ISBN: 978-3-319-20175-7, 978-3-319-20176-4 (cit. on pp. 93, 101).

[149] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012, URL: https://cds.cern.ch/record/1456844 (cit. on pp. 93, 201).

[150] K. Cranmer, *Practical Statistics for the LHC*, 2015, arXiv: 1503.07622 [physics.data-an] (cit. on p. 93).

[151] M. Baak et al., *HistFitter software framework for statistical data analysis*, The European Physical Journal C **75** (2015), URL: https://doi.org/10.1140%2Fepjc%2Fs10052-015-3327-7 (cit. on p. 95).

[152] F. James, *MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1*, (1994) (cit. on p. 97).

[153] J. Neyman and E. S. Pearson,
On the Problem of the Most Efficient Tests of Statistical Hypotheses,
Phil. Trans. Roy. Soc. Lond. A **231** (1933) 289 (cit. on p. 98).

[154] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 60 (cit. on p. 99).

[155] E. Gross and O. Vitells,
Trial factors for the look elsewhere effect in high energy physics,
Eur. Phys. J. C **70** (2010) 525, arXiv: 1005.1891 [physics.data-an]
(cit. on p. 101).

[156] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Phil. Trans. Roy. Soc. Lond. A **236** (1937) 333
(cit. on p. 102).

[157] G. Cowan, K. Cranmer, E. Gross, and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics,
Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)],
arXiv: 1007.1727 [physics.data-an] (cit. on p. 103).

[158] C. M. Bishop, *Pattern Recognition and Machine Learning*,
New York, NY, USA: Springer, 2006, ISBN: 978-0-387-31073-2 (cit. on p. 104).

[159] *Graph Neural Network Jet Flavour Tagging 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-2022-027: CERN, 2022, URL: http://cds.cern.ch/record/2811135
(cit. on pp. 105, 111).

[160] ATLAS Collaboration, *Search for New Phenomena in Two-Body Invariant Mass Distributions Using Unsupervised Machine Learning for Anomaly Detection at* $\sqrt{s} = 13$ *TeV with the ATLAS Detector*, Phys. Rev. Lett. **132** (2023) 081801,
arXiv: 2307.01612 [hep-ex] (cit. on p. 105).

[161] *Electron Identification with a Convolutional Neural Network in the ATLAS Experiment*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-001: CERN, 2023, URL: https://cds.cern.ch/record/2850666
(cit. on p. 111).

[162] ATLAS Collaboration, *Jet Flavour Tagging With GN1 and DL1d. Generator dependence, Run 2 and Run 3 data agreement studies.*,
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/,
[Online; accessed June 11, 2024], 2023 (cit. on p. 111).

[163] S. Hochreiter and J. Schmidhuber, *Long short-term memory*,
Neural computation **9** (1997) 1735 (cit. on p. 112).

[164] *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2020-006: CERN, 2020, URL: http://cds.cern.ch/record/2746032
(cit. on p. 112).

[165] L. S. Shapley, "A Value for n-Person Games",
Contributions to the Theory of Games, Vol. 2, Princeton University Press, 1953 307 (cit. on p. 115).

[166] S. M. Lundberg and S.-I. Lee,
"A Unified Approach to Interpreting Model Predictions",
Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017,
URL: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf (cit. on p. 116).

[167] S. M. Lundberg and S.-I. Lee,
SHAP: A Unified Approach to Interpreting Model Predictions,
https://github.com/slundberg/shap, 2017 (cit. on p. 116).

[168] S. M. Lundberg et al.,
From local explanations to global understanding with explainable AI for trees,
Nature Machine Intelligence **2** (2020) 56 (cit. on p. 116).

[169] ATLAS Collaboration,
ATLAS data quality operations and performance for 2015–2018 data-taking,
JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det]
(cit. on pp. 119, 130).

[170] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*,
Computer Physics Communications **178** (2008) 852, ISSN: 0010-4655,
URL: http://dx.doi.org/10.1016/j.cpc.2008.01.036 (cit. on p. 120).

[171] R. D. Ball et al., *Parton distributions with LHC data*,
Nuclear Physics B **867** (2013) 244, ISSN: 0550-3213,
URL: http://dx.doi.org/10.1016/j.nuclphysb.2012.10.003
(cit. on pp. 120, 187).

[172] *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*, tech. rep.,
All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-017: CERN, 2016, URL: https://cds.cern.ch/record/2206965
(cit. on p. 120).

[173] R. D. Ball et al., *Parton distributions with LHC data*,
Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on p. 120).

[174] ATLAS Collaboration, *ATLAS Run-1 Pythia8 tunes*,
ATLAS-PHYS-PUB-2014-021 (2014),
URL: http://cdsweb.cern.ch/record/1966419 (cit. on pp. 120, 124).

[175] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040,
arXiv: 1410.8849 [hep-ph] (cit. on p. 123).

[176] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 123).

[177] *Merging of nominal ttbar and Wt samples*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MergingHTMETSamplesttWt (cit. on p. 123).

[178] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*,
JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 123).

[179] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 123).

[180] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on p. 123).

[181] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on pp. 123, 124).

[182] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 [hep-ph] (cit. on p. 123).

[183] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 124).

[184] P. Kant et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, Comput. Phys. Commun. **191** (2015) 74, arXiv: 1406.4403 [hep-ph] (cit. on p. 124).

[185] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028, arXiv: 1001.5034 [hep-ph] (cit. on p. 124).

[186] S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, (2013), URL: https://cds.cern.ch/record/1559921 (cit. on p. 124).

[187] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 124).

[188] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, Phys. Rev. Lett. **103** (2009) 082001, arXiv: 0903.2120 [hep-ph] (cit. on p. 124).

[189] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on pp. 126, 165).

[190] *Tagging and suppression of pileup jets 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-2014-018: CERN, 2014, URL: https://cds.cern.ch/record/1700870 (cit. on p. 126).

[191] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on p. 126).

[192] ATLAS Collaboration, *Studies of the muon momentum calibration and performance of the ATLAS detector with pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **83** (2023) 686, arXiv: 2212.07338 [hep-ex] (cit. on p. 126).

[193] *E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-023: CERN, 2018, URL: https://cds.cern.ch/record/2625233 (cit. on p. 126).

[194] 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*, (2024), arXiv: 2402.05858 [hep-ex] (cit. on p. 126).

[195] ATLAS Collaboration, *Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton–proton run*, JINST **8** (2013) P07004, arXiv: 1303.0223 [hep-ex] (cit. on p. 130).

[196] *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: CERN, 2015, URL: https://cds.cern.ch/record/2037702 (cit. on p. 130).

[197] *Object-based missing transverse momentum significance in the ATLAS detector*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-038: CERN, 2018, URL: https://cds.cern.ch/record/2630948 (cit. on p. 135).

[198] J. M. Kanter and K. Veeramachaneni, "Deep feature synthesis: Towards automating data science endeavors", *2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, Paris, France, October 19-21, 2015*, IEEE, 2015 1 (cit. on p. 144).

[199] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17** (2020) 261 (cit. on p. 147).

[200] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*, Physical Review Letters **125** (2020), ISSN: 1079-7114, URL: http://dx.doi.org/10.1103/PhysRevLett.125.122001 (cit. on p. 152).

[201] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2017, arXiv: 1412.6980 [cs.LG] (cit. on p. 155).

[202] *Formulae for Estimating Significance*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-025: CERN, 2020, URL: https://cds.cern.ch/record/2736148 (cit. on pp. 158, 174).

[203] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: 1606.02625 [hep-ex] (cit. on p. 165).

[204] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168 (cit. on p. 166).

[205] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber, *Single-top hadron production in association with a W boson*, Journal of High Energy Physics **2008** (2008) 029, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/2008/07/029 (cit. on p. 166).

[206] M. Baak et al., *HistFitter software framework for statistical data analysis*, Eur. Phys. J. **C75** (2015) 153, arXiv: 1410.1280 [hep-ex] (cit. on p. 169).

[207] 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*, JHEP **07** (2023) 116, arXiv: 2211.07175 [hep-ex] (cit. on p. 177).

[208] CMS Collaboration, *Search for dark matter particles in W^+W^- events with transverse momentum imbalance in proton–proton collisions at $\sqrt{s} = 13$ TeV*, (2023), arXiv: 2310.12229 [hep-ex] (cit. on p. 177).

[209] *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-019: CERN, 2020, URL: https://cds.cern.ch/record/2724739 (cit. on pp. 179–181, 184).

[210] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, Phys. Rev. Lett. **83** (1999) 3370, arXiv: hep-ph/9905221 (cit. on p. 181).

[211] D. Krohn, J. Thaler, and L.-T. Wang, *Jet trimming*, Journal of High Energy Physics **2010** (2010), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP02(2010)084 (cit. on p. 181).

[212] *ATLAS Pythia 8 tunes to 7 TeV data*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-021: CERN, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 187).

[213] *Summary of ATLAS Pythia 8 tunes*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2012-003: CERN, 2012, URL: https://cds.cern.ch/record/1474107 (cit. on p. 187).

[214] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, The European Physical Journal C **75** (2015), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-015-3397-6 (cit. on p. 187).

[215] D. Reichelt, P. Richardson, and A. Siodmok,
Improving the simulation of quark and gluon jets with Herwig 7,
The European Physical Journal C **77** (2017), ISSN: 1434-6052,
URL: http://dx.doi.org/10.1140/epjc/s10052-017-5374-8 (cit. on p. 187).

[216] M. K. Bugge et al., *Muon Selection and Identification Working Points*, tech. rep.,
CERN, 2019, URL: https://cds.cern.ch/record/2665711 (cit. on p. 189).

[217] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*,
Eur. Phys. J. C **79** (2019) 135, arXiv: 1807.09477 [hep-ex] (cit. on p. 199).

[218] *Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with 80 fb^{-1} of $\sqrt{s} = 13$ TeV pp data*, tech. rep.,
All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-022: CERN, 2020, URL: https://cds.cern.ch/record/2724442
(cit. on p. 199).

[219] *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, tech. rep.,
All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-015: CERN, 2015, URL: https://cds.cern.ch/record/2037613
(cit. on p. 199).

[220] Y. Amhis et al.,
Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021,
Phys. Rev. D **107** (5 2023) 052008,
URL: https://link.aps.org/doi/10.1103/PhysRevD.107.052008
(cit. on p. 200).

[221] M. Pinamonti, L. Valery, A. Held, T. Dado, and M. Aly, *TRExFitter*,
GitLab Repository, accessed: 2024-07-19, 2024,
URL: https://gitlab.cern.ch/TRExStats/TRExFitter (cit. on p. 201).

[222] M. Pinamonti, L. Valery, A. Held, T. Dado, and M. Aly,
TRExFitter Documentation,
https://trexfitter-docs.web.cern.ch/trexfitter-docs/,
Accessed: 2024-07-19, 2024 (cit. on p. 201).

[223] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003,
arXiv: physics/0306116 [physics.data-an] (cit. on p. 201).

[224] *Proceedings of the 1974 CERN School of Computing: Godøysund, Norway 11 - 24 Aug 1974. 3rd CERN School of Computing*, CERN, Geneva: CERN, 1974,
URL: https://cds.cern.ch/record/186223 (cit. on p. 201).

Acronyms

ΛCDM model Λ Cold Dark Matter Model. 25

2MDM Two-Mediator Dark Matter Model. 34, 37, 177

ALICE A Large Ion Collider Experiment. 49

ATLAS A Toroidal LHC Apparatus. 16, 17, 21, 31, 41, 42, 46, 49–54, 56, 58, 60–64, 67, 69, 75, 76, 78, 80, 82, 85, 86, 111, 112, 119–121, 125, 130, 165, 177, 180, 187, 207, 218–220

BCE binary cross-entropy. 107, 149–152

BDT Boosted Decision Tree. 105, 148

BSM Beyond the Standard Model. 98

BSM Beyond the Standard Model. 4, 19, 21, 22, 25, 30–34, 69, 87, 89, 94, 95, 97, 100, 101, 105, 115, 119, 131, 159, 180, 218, 219

CB combined. 88, 89

CDM Cold Dark Matter. 29, 30

CERN European Organisation for Nuclear Research. 47, 48

CKM Cabibbo-Kobayashi-Maskawa. 18

CMB Cosmic Microwave Background. 25, 26

CMS Compact Muon Solenoid. 16, 21, 31, 49, 121, 177

CNN Convolutional Neural Network. 111

CP Cluster Processor. 67

CR Control Region. 95, 125, 135, 141, 157–164, 169, 170, 172–175

CSCs Cathode Strip Chambers. 64–66

CT calorimeter-tagged. 89

CTP Central Trigger Processor. 68

DGLAP Dokshitzer–Gribov–Lipatov–Altarelli–Parisi. 72

DM Dark Matter. 4, 19, 24–35, 37, 40, 41, 43, 52, 82, 89, 105, 119, 121, 219, 220

DNN Deep Neural Network. 86, 106, 107, 115, 180

DR discovery region. 159, 169, 173–175

ECal Electromagnetic Calorimeter. 55, 57, 61–63

EFT Effective Field Theories. 32–34

ELU Exponential Linear Unit. 107

FE front-end. 68

FSR final state radiation. 72, 165, 166, 200, 212, 213

GAN Generative Adversarial Network. 112

GMSB Gauge-Mediated SUSY Breaking. 41, 42

GNN Graph Neural Network. 86, 105, 111, 218

GUT Grand Unified Theories. 23

HCal Hadronic Calorimeter. 55, 57, 61–64

HDM Hot Dark Matter. 29

HEP High Energy Physics. 93, 94, 105, 109, 111, 219

HLT High-Level Trigger. 67, 68, 188

IBL Insertible B-Layer. 57–59

ID Inner Detector. 52, 54–60, 76, 77, 79, 80, 88–90, 126, 127, 130, 181

IO inside-out combined. 89

IR infrared. 80, 82

ISR initial state radiation. 38, 70, 72, 165, 166, 200

JEP Jet/Energy-sum Processor. 67

JER jet energy resolution. 82, 84, 85

JES jet energy scale. 82–85, 96, 165, 199

JVT jet vertex tagger. 126

L1 Level-1 Trigger. 67, 68

L1Calo L1 calorimeter trigger. 67, 68

L1Muon L1 muon trigger. 67, 68

L1Topo L1 topological trigger. 67, 68

LAr Liquid Argon. 61–63, 67

LCW local hadronic cell weighting. 78

LHC Large Hadron Collider. 13–15, 28, 31–34, 41, 43, 44, 46–50, 52, 53, 69, 71, 79, 98, 105, 119, 125, 219, 220

LHCb Large Hadron Collider beauty. 49

LINAC2 Linear Accelerator 2. 47, 48

LO leading order. 71–73, 120

LSP Lightest Supersymmetric Particle. 40–42

MACHOs Massive Compact Halo Objects. 30

MC Monte-Carlo. 69, 71, 79, 83, 85, 87, 91, 96, 97, 99, 105, 120, 122, 123, 132, 135, 141, 159, 165, 170, 180, 186, 187, 189, 195, 196, 199–202, 208–210, 212

MDTs Monitored Drift Tubes. 64–66

ME muon-spectrometer extrapolated. 89

ML Machine Learning. 104, 105, 115, 141, 158, 219, 220

MLP Mulit-Layer Perceptron. 104, 106, 108–110, 113

MOND modified Newtonian dynamics. 25

MPI multiple parton interactions. 69, 73, 74

MS Muon Spectrometer. 52, 55, 57, 64, 65, 67, 69, 76, 77, 88–90

MSE mean squared error. 107, 110, 150

MSSM Minimal Supersymmetric Standard Model. 33, 40, 41, 43

MVA multivariate analysis. 132, 139, 140, 147

nEDM neutron electric dipole moment. 24

NLO next-to leading order. 71, 72, 123

NN Neural Network. 104–106, 108–112, 114, 125, 132, 135, 138, 139, 142–145, 147–149, 151, 154, 158, 159, 162, 166

NNLO next-to-next-to leading order. 71

OR overlap removal. 125–127

PCA Principle Component Analysis. 114

PDF parton distribution function. 71, 72, 120, 123, 124, 187

PDG Particle Data Group. 74

PFOs Particle Flow Objects. 79–82, 125

PLR Profile Likelihood Ratio. 98, 99, 102

PMNS Pontecorvo-Maki-Nakagawa-Sakata. 18

PoI Parameter of Interest. 98, 99, 195, 196

ProbDF probability distribution function. 95, 99

PS parton shower. 69

PS Proton Synchrotron. 47, 48, 120, 123, 200, 211–214

PSB Proton Synchrotron Booser. 47, 48

PV Primary Vertex. 77, 191

QCD Quantum Chromodynamics. 13, 14, 24, 73, 131, 132, 135, 189

QED Quantum Electrodynamics. 7, 8, 13, 15

QFT Quantum Field Theory. 5, 23

ReLU Rectified Linear Unit. 107, 155, 182

RF radio frequency. 47

RNN Recurrent Neural Network. 87, 111, 112

RODs ReadOut Drivers. 68

RoIs regions of interest. 68

RPCs Resistive Plate Chambers. 64–67

RS Randall-Sundrum. 180, 181

SCT Semiconductor Detector. 57, 59, 77

SF scale factor. 87, 98, 186, 195–198, 201, 202, 208–212, 214, 218

SM Standard Model. 18, 34, 61, 98

SM Standard Model. 4–7, 15–20, 22, 23, 25, 26, 28, 29, 31–38, 40, 41, 43, 44, 52, 55, 57, 64, 69, 90, 105, 115, 119–121, 132, 137, 138, 153, 156, 159, 174, 177, 180, 186, 218, 219

SPS Super Proton Synchrotron. 47, 48

SR Signal Region. 95, 141, 157–164, 169, 172–176

SSVF Single Secondary Vertex Finder. 86, 87

ST segment-tagged. 89

SUSY Supersymmetry. 23, 30, 40–42, 52, 105, 119, 125, 130, 140

SV Secondary Vertex. 58, 87, 191, 194, 195, 200, 202, 205, 206, 208, 210, 212, 214, 217

TGCs Thin Gap Chambers. 64–67

TRT Transition Radiation Tracker. 57, 59–61, 65, 77

UE underlying event. 120, 123, 124

ValR validation region. 141, 157, 162–164, 169, 170, 173, 174

vdM van der Meer. 50

VEV vacuum expectation value. 10, 18, 22, 34, 36

VR variable radius. 82, 170, 181, 182, 184, 188–190, 194

WIMPs Weakly interacting massive particles. 29, 30

WP working point. 86, 89, 126, 130, 131, 183, 186, 197, 210