



Optimised displaced object identification
and search for long-lived dark matter
candidates in $H \rightarrow aa$ decays with the
ATLAS detector

EMIL HAINES

UNIVERSITY COLLEGE LONDON

Submitted to University College London in fulfilment of the
requirements for the award of the degree of Doctor of Philosophy

April 12, 2024



DECLARATION

I, Emil Haines, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Many extensions to the Standard Model predict new long-lived particles (LLPs) whose existence can be probed by the Large Hadron Collider (LHC). LLPs produce unique detector signatures, and require novel search strategies and analysis techniques to aid in their discovery. This thesis presents a search for light, pseudoscalar LLPs, performed using 140 fb^{-1} of proton-proton collision data with $\sqrt{s} = 13 \text{ TeV}$ recorded at the ATLAS detector between 2015 and 2018. The search targets hadronically decaying LLPs a with masses ranging from 5 to 55 GeV. Benchmark models with LLP pair production from exotic decays of the Higgs boson to LLPs $H \rightarrow aa$ and models featuring long-lived axion-like particles (ALPs) are considered. No excess over the expected background is observed. Upper limits are set on branching ratio of the Higgs boson to pairs of LLPs, the cross-section for ALPs produced in association with a vector boson, and, for the first time, on the branching ratio of the top quark to an ALP and a u/c quark. The limits for the $H \rightarrow aa$ model are the most stringent to date for LLP masses $m_a < 40 \text{ GeV}$ and lifetimes $10^{-1} - 10^{-3} \text{ m}$. Future LLP searches will rely on the efficient reconstruction and identification of displaced objects to maximise discovery potential - this thesis contains further research to that effect. Key performance studies of the large radius tracking algorithm in the ATLAS trigger are presented. Additionally, transformer-based graph neural networks are applied to the task of identifying jets produced in $H \rightarrow aa$ decays, and are found to improve background rejection up $\sim 10\text{--}15\times$, when compared to an existing approach. Further improvements made by incorporating conditioning mechanisms into the network architecture are also studied in depth and detailed.

IMPACT STATEMENT

This thesis contains research in experimental particle physics. The work presented focuses on the search for new physics through the exploration of long-lived particles, using data recorded at the ATLAS experiment at CERN. The discovery or exclusion of a new particle has the potential to greatly enhance our fundamental understanding of the Universe.

The ambitious scientific goals of high energy particle physics drive innovation. The cutting-edge analysis techniques developed for this work, and at CERN for ATLAS and the LHC as a whole, both lay the groundwork for future breakthroughs in particle physics research and have beneficial applications in a wide range of other fields. Technologies developed at CERN have been deployed in domains ranging from medical imaging to distributed computing, and include, among others, the World Wide Web.

The skills and experience acquired by researchers within the field can be transferred to roles within business, industry and the public sector, and used to tackle problems outside the realm of scientific research. In particular, individuals with expertise in machine learning, which features heavily in this work, will have significant societal and economic impact in the coming years, as data driven activities continue to evolve.

Particle physics research is a global pursuit, and serves as a successful blueprint for international research efforts in other fields. The research conducted at CERN is highly visible, encouraging excitement and intrigue in the general public about science and technology, and inspiring the next generations of researchers.

ACKNOWLEDGEMENTS

First, I would like to thank Andreas for his support, advice, and unwavering pragmatism throughout this doctorate. I could not have asked for a better supervisor. I am highly indebted to everyone in the HDMI analysis team, in particular Jackson (Burzynski), Stefanie, Dominique, and Benjamin, whose expertise I could always rely on. I would also like to thank Gabriel for steering me towards the GNN studies, as well as Sam, Jackson (Barr), and Melissa for supporting this work. Both at UCL and CERN, I have had the privilege of meeting countless wonderful people whose companionship has been highly appreciated over the years. Finally, thank you to my Mum, Dad, family, and friends. I owe you all a lifetime of gratitude.

CONTENTS

LIST OF FIGURES [X](#)

LIST OF TABLES [XV](#)

1 INTRODUCTION [1](#)

2 THEORETICAL BACKGROUND [3](#)

2.1 THE STANDARD MODEL OF PARTICLE PHYSICS [3](#)

2.1.1 *Particle content of the Standard Model* [3](#)

2.1.2 *Decay rates, lifetimes and cross sections* [5](#)

2.1.3 *Quantum field theory* [6](#)

2.1.4 *Quantum chromodynamics* [6](#)

2.1.5 *Electroweak theory* [7](#)

2.1.6 *The Higgs mechanism* [10](#)

2.1.7 *The Higgs boson* [13](#)

2.2 LIMITATIONS OF THE STANDARD MODEL [15](#)

2.2.1 *Dark matter* [15](#)

2.2.2 *Hierarchy problem* [16](#)

2.3 PHYSICS BEYOND THE STANDARD MODEL [17](#)

2.3.1 *Long-lived particles* [17](#)

2.3.2 *Exotic Higgs decays* [19](#)

3 THE ATLAS EXPERIMENT [22](#)

3.1 THE LARGE HADRON COLLIDER [22](#)

3.1.1 *Luminosity and pile-up* [24](#)

3.2 THE ATLAS DETECTOR [26](#)

3.2.1	<i>Coordinate system and nomenclature</i>	26
3.2.2	<i>Inner detector</i>	27
3.2.3	<i>Calorimeter</i>	29
3.2.4	<i>Muon spectrometer</i>	31
3.2.5	<i>The ATLAS trigger system</i>	32
3.2.6	<i>Event simulation</i>	33
4	EVENT RECONSTRUCTION	34
4.1	TRACK AND VERTEX RECONSTRUCTION	34
4.1.1	<i>Standard tracking</i>	34
4.1.2	<i>Primary vertices</i>	36
4.2	PHYSICS OBJECT RECONSTRUCTION	36
4.2.1	<i>Jets</i>	36
4.2.2	<i>Electrons</i>	39
4.2.3	<i>Photons</i>	41
4.2.4	<i>Muons</i>	42
4.2.5	<i>Missing transverse momentum</i>	44
4.3	DISPLACED TRACK AND VERTEX RECONSTRUCTION	45
4.3.1	<i>Large radius tracking</i>	45
4.3.2	<i>Displaced vertices</i>	48
5	SEARCH FOR EXOTIC HIGGS DECAYS TO LONG-LIVED PARTICLES	52
5.1	ANALYSIS OVERVIEW	52
5.1.1	<i>Previous searches</i>	52
5.1.2	<i>Analysis strategy</i>	53
5.2	DATA AND SIMULATED SAMPLES	56
5.2.1	<i>Simulated samples</i>	56
5.2.2	<i>Trigger strategy</i>	57
5.3	OBJECT DEFINITIONS	60
5.3.1	<i>Electrons</i>	60
5.3.2	<i>Muons</i>	61

5.3.3	<i>Jets</i>	61
5.3.4	<i>Photons</i>	63
5.3.5	<i>Overlap removal</i>	63
5.4	EVENT SELECTION	64
5.4.1	<i>Preselection</i>	64
5.4.2	<i>Displaced jet selection</i>	65
5.4.3	<i>Displaced vertex selection</i>	68
5.4.4	<i>Control and signal regions</i>	70
5.5	BACKGROUND ESTIMATE	71
5.5.1	<i>Procedure</i>	71
5.5.2	<i>Validation</i>	74
5.6	UNCERTAINTIES	76
5.6.1	<i>Uncertainty on the background prediction</i>	76
5.6.2	<i>Uncertainty on the signal yields</i>	77
5.7	STATISTICAL ANALYSIS	79
5.7.1	<i>The likelihood function</i>	79
5.7.2	<i>Hypothesis tests</i>	81
5.7.3	<i>Limit setting and the CL_s method</i>	82
5.7.4	<i>Fit strategy</i>	82
5.8	RESULTS	83
5.8.1	<i>Unblinded results</i>	83
5.8.2	<i>Exclusion limits</i>	85
5.8.3	<i>Results summary</i>	86
5.9	CONCLUSION	90
6	ADVANCES IN LONG-LIVED PARTICLE IDENTIFICATION	92
6.1	LARGE RADIUS TRACKING PERFORMANCE IN THE ATLAS TRIGGER	93
6.1.1	<i>Data and object reconstruction</i>	93
6.1.2	<i>Trigger large radius tracking efficiency</i>	95
6.1.3	<i>Conclusion</i>	96

6.2	GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION	97
6.2.1	<i>Motivation</i>	97
6.2.2	<i>Neural network theory</i>	98
6.2.3	<i>Datasets</i>	103
6.2.4	<i>Object selections</i>	105
6.2.5	<i>Model specification</i>	106
6.2.6	<i>Training</i>	118
6.2.7	<i>Results</i>	121
6.2.8	<i>Conclusion</i>	138
7	CONCLUSION	144
	BIBLIOGRAPHY	147
A	MONTE CARLO SAMPLES	159
B	LEPTON TRIGGER EFFICIENCIES	168
C	VBF REGION OPTIMISATION	176

LIST OF FIGURES

- 2.1 Particle content of the Standard Model. [4](#)
- 2.2 The production of hadronic jets through hadronisation. [8](#)
- 2.3 Reconstructed invariant mass distribution of the photon pair in candidate $H \rightarrow \gamma\gamma$ decays at the ATLAS experiment [10]. [13](#)
- 2.4 Dominant Higgs boson production modes in proton-proton collisions. [14](#)
- 2.5 Branching ratios of the Higgs boson for masses around 125 GeV [11]. [15](#)
- 2.6 Radiative corrections to the Higgs boson mass via a top quark loop. [17](#)
- 2.7 Particle lifetimes $c\tau$, in metres, as a function of mass, in GeV, for Standard Model particles [16]. [18](#)
- 2.8 A long-lived particle decaying hadronically into a single jet [20] and two jets [21]. [19](#)
- 2.9 Exotics Higgs decay $H \rightarrow ss \rightarrow q\bar{q}q\bar{q}$. [21](#)
- 3.1 An overview of the accelerator complex at CERN in 2022 [33]. [23](#)
- 3.2 The cumulative integrated luminosity versus time delivered to ATLAS between 2015 and 2018 [36]. [24](#)
- 3.3 The luminosity-weighted distribution of pile-up μ for pp collisions in 2015-2018 at 13 TeV and in 2022-2023 at 13.6 TeV [37]. [25](#)
- 3.4 The ATLAS detector and its subdetectors [28]. [26](#)
- 3.5 A cutaway of the barrel region of the inner detector [38]. [28](#)
- 3.6 An r - z cross section of the ATLAS inner detector during Run 2 [38]. [30](#)
- 4.1 Inclusive and technical track reconstruction efficiency for displaced charged particles produced via decays of a pseudoscalar LLP, a , as a function of the production radius of the displaced particles [71]. [47](#)

- 4.2 A comparison of the radial distributions of reconstructed secondary vertices in an LLP sample using the Run 2 and Run 3 LRT configurations [71]. 49
- 4.3 Schematic showing a transverse view of the pixel detector in the barrel region, along with the required, allowed and forbidden pixel layer hits of tracks forming a secondary vertex in between the B-layer and Layer 1 sensors and close to the B-layer [72]. 50
- 5.1 Regions in $BR(H \rightarrow ss)$ versus $c\tau$ excluded at 95% confidence level by the ATLAS analyses listed in the legend [78]. 53
- 5.2 Feynman diagrams depicting the Za , Wa , and $t \rightarrow qa$ ALP production modes. 55
- 5.3 Lepton trigger matching outcome for events in ZH and WH signal MC samples. 58
- 5.4 Normalised distributions of n_{trk} and $m_{\text{DV}}/\Delta R_{\text{max}}$ for vertices in selected exotic Higgs decay signal samples with $H \rightarrow ss \rightarrow 4b$ and data [1]. 70
- 5.5 Normalised distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for events in the *Presel_1lep* region, with the additional requirement that the events contain at least two jets with BDT scores greater than 0.5, in data and simulated exotic Higgs decay signal samples [1]. 71
- 5.6 Distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for the observed data and background prediction along with its uncertainty in the three $n_{\text{DV}} = 1$ validation regions with $0.7 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.9$ [1]. 75
- 5.7 Distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for the observed data and background prediction along with its uncertainty in the two photon validation regions containing events with $n_{\text{DV}} = 1$ and $n_{\text{DV}} > 2$ [1]. 75
- 5.8 Distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for the observed data and background prediction along with its uncertainty after the maximum-likelihood fit to data has been performed [1]. 84

- 5.9 The 95% confidence level exclusion limits on $BR(H \rightarrow ss), \sigma(q\bar{q} \rightarrow Va)$ for $V = Z, W$, and $BR(t \rightarrow aq)$ shown as a function of the mean proper decay length $c\tau$ of the LLP [1]. 87
- 5.10 Regions in LLP mass versus proper decay length excluded at the 95% confidence level for the exotic Higgs decay model. 89
- 5.11 Event display showing a signal-like event in the $2lep$ region. 90
- 5.12 Efficiency of the LRT-based displaced jets triggers plotted against the leading offline jet p_T for a model of exotic decays of the Higgs boson into long-lived pseudoscalars with mass 55 GeV and proper decay length of 100 mm [2]. 91
- 6.1 Radius and mass of offline K_S^0 candidates vertices. 94
- 6.2 Absolute value of the transverse impact parameter $|d_0|$ and p_T of offline full scan tracks associated to K_S^0 candidates vertices. 94
- 6.3 The efficiency of online full scan tracks reconstructed using LRT, with respect to offline tracks associated to K_S^0 vertices reconstructed as a secondary vertex using the offline tracks [2]. 95
- 6.4 The efficiency to reconstruct a displaced secondary K_S^0 vertex from online full scan tracks, as a function of the decay radius of the vertex in the transverse plane r_{xy} [2]. 96
- 6.5 Schematic of a fully connected deep neural network with two inputs, three hidden layers with three neurons each, and a single output. 99
- 6.6 Conditional biasing. 102
- 6.7 Conditional scaling. 102
- 6.8 Conditional affine transformation. 103
- 6.9 Jet p_T for jets labelled as signal from $aa \rightarrow 4u$ and $a \rightarrow 4b$ ($aa \rightarrow 4c$ for $m_a = 5$ GeV) decays, along with background. 107
- 6.10 Displaced p_T fraction for jets labelled as signal from $aa \rightarrow 4u$ and $a \rightarrow 4b$ decays ($aa \rightarrow 4c$ for $m_a = 5$ GeV). 107

- 6.11 Number of tracks associated to jets labelled as signal from $aa \rightarrow 4u$ (left) and $aa \rightarrow 4b$ ($aa \rightarrow 4c$ for $m_a = 5$ GeV) (right) decays, along with background. 108
- 6.12 Distributions of track p_T for tracks associated to jets labelled as signal from $aa \rightarrow 4u$ and $aa \rightarrow 4b$ ($aa \rightarrow 4c$ for $m_a = 5$ GeV) decays, along with background. 108
- 6.13 The total p_T of tracks descending from LLP decays as a fraction of the total p_T of all the tracks associated to light-flavour jets. 111
- 6.14 The inputs to the baseline model. 112
- 6.15 The baseline network architecture. 113
- 6.16 Scaled dot-product attention. 114
- 6.17 Multi-head attention. 115
- 6.18 The inputs to the *Concatenation* model. 116
- 6.19 The *Scaling* model. 118
- 6.20 The *Affine* model. 119
- 6.21 Background jet rejection versus signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7}$, $\mathcal{D}_{\mathcal{M}\setminus 10}$, $\mathcal{D}_{\mathcal{M}\setminus 25}$, and $\mathcal{D}_{\mathcal{M}\setminus 40}$, for jets with p_T between 20 and 300 GeV. 124
- 6.22 Background jet rejection versus signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7,10,25,40}$, for jets with p_T between 20 and 300 GeV. 126
- 6.23 Background jet rejection versus signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}}$, for jets with p_T between 20 and 300 GeV. 129
- 6.24 Model output and background rejection versus signal efficiency for a *Baseline* and *Affine* model. 130
- 6.25 Model output for 7 GeV and 40 GeV signal and background samples for *Baseline* and *Affine* models trained using $\mathcal{D}_{\mathcal{M}}$. 130
- 6.26 Signal LLP jet efficiency and background jet rejection as a function of jet p_T and number of tracks associated to the jet at a fixed signal LLP jet efficiency of 70%. 132

- 6.27 Signal LLP jet efficiency as a function of the jet displaced p_T fraction, fraction of the total sum of transverse momenta of tracks associated to the jet attributed to tracks descending from the decay of an LLP, and the displacement of the parent LLP of the jet, at a fixed signal LLP jet efficiency of 70%. 133
- 6.28 Model output for different flavour signal jets and types of background for an *Affine* model evaluated at $m_a = 7$ GeV and $m_a = 40$ GeV. 135
- 6.29 Background jet rejection versus signal efficiency for light- and heavy-flavour jets in signal samples generated at $m_a = 7$ GeV and $m_a = 40$ GeV for an *Affine* model trained using $\mathcal{D}_{\mathcal{M}}$. 135
- 6.30 Background jet rejection for various SM processes as a function of signal light-flavour and heavy-flavour jet efficiency for an *Affine* model trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated at $m_a = 7$ GeV and $m_a = 40$ GeV. 136
- 6.31 Background jet rejection versus signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7}$, $\mathcal{D}_{\mathcal{M}\setminus 10}$, $\mathcal{D}_{\mathcal{M}\setminus 25}$, and $\mathcal{D}_{\mathcal{M}\setminus 40}$, for jets with p_T greater than 300 GeV. 139
- 6.32 Background jet rejection versus signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}}$, for jets with p_T greater than 300 GeV. 140
- B.1 Pass efficiency for single-electron triggers. 170
- B.2 Pass efficiency for single-muon triggers. 171
- B.3 Pass efficiency for di-electron triggers. 171
- B.4 Pass efficiency for di-muon triggers. 172
- B.5 Match efficiency for single-electron triggers. 173
- B.6 Match efficiency for single-muon triggers. 174
- B.7 Match efficiency for di-electron triggers. 174
- B.8 Match efficiency for di-muon triggers. 175
- C.1 VBF inclusive trigger efficiency as a function of m_{jj} , E_T^0 and E_T^1 for data and MC. 178

LIST OF TABLES

- 2.1 Summary of the electroweak quantum numbers for left- and right-handed chiral fermions. [9](#)
- 2.2 Production cross sections for a Higgs boson with $m_H = 125$ GeV in pp collisions at centre of mass energies $\sqrt{s}$ of 13 TeV and 14 TeV [12]. [14](#)
- 4.1 A description of the five track defining parameters. [35](#)
- 4.2 A comparison of the selection criteria used in standard and large radius tracking. [46](#)
- 4.3 Fiducial selections applied to truth particles entering the tracking efficiency calculation. [48](#)
- 5.1 Single lepton (e or μ) and di-lepton (ee or $\mu\mu$) triggers forming the lepton trigger preselection and the years for which they were active. [59](#)
- 5.2 The VBF inclusive trigger used in the analysis. [59](#)
- 5.3 Photon triggers and the years for which they were active. [60](#)
- 5.4 Summary of electron definition criteria used in the analysis. [61](#)
- 5.5 Summary of muon definition criteria used in the analysis. [61](#)
- 5.6 Summary of jet reconstruction parameters and selections used in the analysis. [62](#)
- 5.7 Summary of photon definition criteria used in the analysis. [63](#)
- 5.8 The overlap removal sequence used in the analysis to disambiguate electrons (e), muons (μ), jets (j), and photons (γ). [64](#)
- 5.9 Summary of selections for each preselection region. [66](#)
- 5.10 Signal and background MC composition of the BDT training set. [66](#)
- 5.11 The full set of selections applied to vertices. [68](#)
- 5.12 Summary of analysis selections for the six signal regions. [71](#)

- 5.13 The per-jet efficiency map binning scheme for the $2lep$ and the VBF search regions. 72
- 5.14 The per-jet efficiency map binning scheme for the $1lep$ search region. 73
- 5.15 The number of bins in $BDT_{j_0} \times BDT_{j_1}$ score used in the maximum-likelihood fit for each signal region. 83
- 5.16 The ranges of values of LLP proper decay length $c\tau_a$ excluded at the 95% confidence level for different values of $BR(H \rightarrow ss \rightarrow 4b/4c)$, for LLP masses in the range 5–55 GeV. 88
- 5.17 The ranges of values of LLP proper decay length $c\tau_s$ excluded at the 95% confidence level for different values of $BR(H \rightarrow ss \rightarrow 4u)$, for LLP masses in the range 5–55 GeV. 89
- 6.1 Summary of the different training sets used in this work. The background jet component of each dataset is split equally between jets from $t\bar{t}$, $Z + \text{jets}$, $W + \text{jets}$, and QCD dijet samples. 105
- 6.2 Summary of the jet and track parameters used as input features to the GNN. 109
- 6.3 Exclusive classes used to categorise tracks for the track origin prediction task. 111
- 6.4 Summary of networks used to perform each classification task. The hidden layers column specifies the number of neurons in each hidden layer. All hidden layers are ReLU activated. 116
- 6.5 Summary of the types of parameterisation used in each model. 119
- 6.6 The background rejection at 70% and 85% signal efficiency working points for models trained using the training sets given in the *Dataset* column and evaluated at the mass point given in the *Mass* column using jets with p_T in between 20 GeV and 300 GeV. 123
- 6.7 The background rejection at 70% and 85% signal efficiency working points for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7,10,25,40}$ and evaluated at the mass point given in the *Mass* column using jets with p_T in between 20 GeV and 300 GeV. 125

- 6.8 The background rejection at 70% and 85% signal efficiency working points for models trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated using jets with p_T in between 20 GeV and 300 GeV. [128](#)
- 6.9 The background rejection at 70% and 85% signal efficiency working points for models trained using the training sets given in the *Dataset* column and evaluated at the mass point given in the *Mass* column using jets with p_T greater than 300 GeV. [141](#)
- 6.10 The background rejection at 70% and 85% signal efficiency working points for models trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated using jets with p_T greater than 300 GeV. [142](#)
- A.1 The dataset identifiers and process names for signal $H \rightarrow aa$ MC samples used in the LLP search. [164](#)
- A.2 The dataset identifiers and process names for background MC samples used in the LLP search. [165](#)
- A.3 The dataset identifiers and process names for signal $H \rightarrow aa$ MC samples used in the graph neural network studies. [167](#)
- B.1 Single lepton (e or μ) and di-lepton (ee or $\mu\mu$) triggers forming the lepton trigger preselection and the years and periods for which they were active. [169](#)
- C.1 Summary of jet pair (j_0, j_1) variable thresholds used in the VBF inclusive trigger. [176](#)
- C.2 Standard Model background samples used to derive the VBF inclusive trigger scale factors. [177](#)

INTRODUCTION

The Standard Model has successfully described particle collisions at experiments such as the Large Hadron Collider. However, the model is not complete and several important questions remain unanswered. A number of theories that aim to solve the limitations of the Standard Model, such as explaining the nature of dark matter, predict the existence of exotic long-lived particles.

This thesis concerns a search for hadronically decaying light long-lived particles, as well as techniques developed to optimise their identification. The thesis is structured as follows:

§2 describes the core theory underpinning the work presented in the rest of the thesis.

§3 describes the CERN accelerator complex and the ATLAS detector.

§4 describes the physics objects relevant to this thesis and the algorithms used to reconstruct them.

§5 presents a search for new, light long-lived particles that decay hadronically within the ATLAS inner detector.

§6 contains two studies related to improving future searches for long-lived particles. The first details the performance of the large-radius tracking algorithm in the trigger. The second investigates using conditioned transformer-based graph neural networks to identify displaced jets from long-lived particle decays.

§7 contains concluding remarks.

This thesis covers work conducted by the author while a member of the ATLAS experiment. The contributions of the author to the work presented in this thesis are as follows:

Search for long-lived particles: The author was an active member of the analysis team responsible for the search for light long-lived particles presented in §5. The author led the implementation of the analysis trigger strategy, including the trigger selection, lepton trigger matching, and lepton trigger scale factors, and performed various trigger efficiency studies. The author was responsible for investigating and optimising the vector boson fusion search channel - the first time such a channel had been included in an ATLAS search of this kind. The author assisted in the development of the background estimation strategy by conducting a number of validation studies. The author also made significant contributions to the analysis software framework, documented work within the internal note, and gave the analysis approval talk to the ATLAS Exotics group. A paper reporting the results of this analysis has recently been published [1].

Large-radius tracking: The author conducted a study of the performance of the large-radius tracking algorithm in the ATLAS trigger for the Inner Detector trigger group (§6.1). This study was included in a recent ATLAS trigger performance paper [2].

Displaced jet tagging: The author led initial investigations into adapting the GN2 algorithm, currently used at ATLAS for flavour tagging, for use in a future search for exotic Higgs decays to long-lived particles. The author contributed to an open source package used in high energy physics for training models such as GN2 [3], by implementing a number of conditioning mechanisms within the framework, allowing users to create parameterised neural networks. The author studied the performance of various new network architectures based on GN2 using these conditioning mechanisms in identifying jets from long-lived particle decays (§6.2), and found the developments could improve the displaced jet tagging performance with respect to current benchmark approaches.

2

THEORETICAL BACKGROUND

This chapter contains a brief overview of the Standard Model in §2.1, followed by a discussion of its limitations in §2.2, and how these motivate theories that go beyond the Standard Model in §2.3. The aim of this chapter is to equip the reader with sufficient knowledge to contextualise and understand the work presented in this thesis. A basic understanding of the Standard Model and quantum field theory is assumed. Those seeking a more complete mathematical description of the Standard Model are referred to Refs. [4, 5, 6].

2.1 THE STANDARD MODEL OF PARTICLE PHYSICS

Particle physics is concerned with the fundamental constituents of the Universe, *particles*, and the interactions between them, *forces*. Our current understanding is underpinned by the the Standard Model (SM) of particle physics.

2.1.1 *Particle content of the Standard Model*

The SM consists of a set of fundamental, or indivisible, particles. These particles interact with each other through the gravitational, electromagnetic, strong and weak forces. Gravity is yet to be integrated into the SM, although its effect at particle physics experiments is negligible when compared to the other three forces. The particles are grouped based on certain intrinsic properties, as shown in Figure 2.1.

Fermions are spin- $\frac{1}{2}$ particles that form the constituents of matter. They are divided

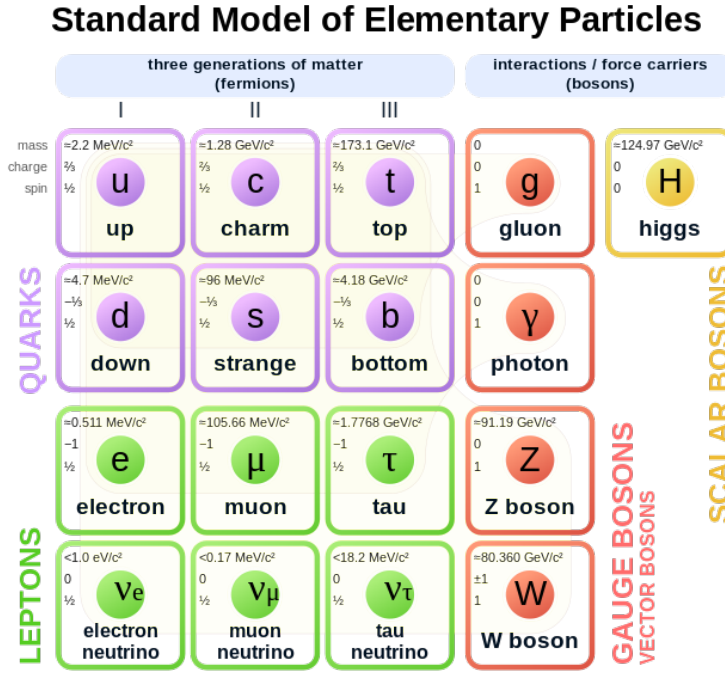


Figure 2.1: Particle content of the Standard Model.

into *quarks* and *leptons*. For each fermion there exists an antiparticle with identical mass but opposite quantum numbers.

There are six flavours of quarks arranged into three generations: the up (u) and down (d), the charm (c) and strange (s), and the top (t) and bottom (b). They can interact via the electromagnetic, strong and weak forces. The first in each pair has electric charge $+\frac{2}{3}e$ and the second $-\frac{1}{3}e$. Quarks also carry colour charge, the quantum number related to the strong interaction.

Leptons are organised into three generations containing a charge- $1e$ particle and a neutral neutrino: the electron (e) and electron neutrino (ν_e), the muon (μ) and muon neutrino (ν_μ), and the tau (τ) and tau neutrino (ν_τ). The charged leptons interact via the electromagnetic and weak forces, while neutrinos only interact via the weak force.

The three forces contained in the SM are mediated by the exchange of spin-1 force-carrying particles, known as *gauge bosons*. The photon (γ) is massless and responsible for electromagnetic interactions between electrically charged particles. The weak bosons, $W^\pm$ and Z^0 , are massive and mediate weak charged- and neutral-current interactions between quarks and leptons. Lastly, the strong interaction is mediated by massless gluons (g); this can be between quarks and gluons or gluons themselves.

The final element of the SM is the Higgs boson (H). The Higgs is a spin-0, or *scalar*, particle, that results from the mechanism through which the weak bosons, quarks and charged leptons acquire mass, as discussed in §2.1.6.

2.1.2 Decay rates, lifetimes and cross sections

Decay rates and cross sections are important experimentally measured phenomena that describe transitions between different particle states. Transitions between states depend on two components: the transition matrix element and the Lorentz-invariant phase space. The former describes the time evolution of all initial and final states in a given interaction. The latter describes the allowed kinematics of the final state.

The decay rate is the probability per unit time of single particle decaying to multiple particles. A given particle may have a number of decay modes j , corresponding to partial decay rates Γ_j . The total decay rate Γ is the sum of all partial decay rates

$$\Gamma = \sum_j \Gamma_j. \quad (2.1.1)$$

Each partial decay rate can be calculated by

$$\Gamma_j = \frac{1}{2m} \int |\mathcal{M}|^2 d\Pi, \quad (2.1.2)$$

where m is the mass of the decaying particle, $\mathcal{M}$ is the matrix element, and $d\Pi$ is the differential Lorentz-invariant phase space available to the decay.

The relative frequency of a particular decay mode is termed the *branching ratio*. The branching ratio for a given decay mode $BR(j)$ is the ratio of the partial decay rate for mode j to the total decay rate

$$BR(j) = \frac{\Gamma_j}{\Gamma}. \quad (2.1.3)$$

The lifetime of a particle τ is determined from the total decay rate,

$$\tau = \frac{1}{\Gamma}, \quad (2.1.4)$$

and is defined as the time taken for a sample of the particle to decay to $1/e$ of its initial

2.1. THE STANDARD MODEL OF PARTICLE PHYSICS

number.

The probability that an interaction between multiple particles will occur is expressed through the interaction *cross section* σ . The cross section can be derived in a similar manner to the decay rate, while additionally accounting for the flux of initial state particles. The cross section has dimensions of area and can be thought of as representing the effective cross sectional area of each particle in an interaction.

In practice, the *differential cross section* with respect to a kinematic variable can provide a deeper understanding of the underlying physics governing an interaction. As such, experimental data is often presented as a differential number of events as a function of some variable for a given integrated luminosity (defined in §3.1.1).

2.1.3 Quantum field theory

The SM is a quantum field theory (QFT). In QFT, particles are described as excitations of a quantum field. The dynamics of a quantum field are derived through a Lagrangian density, $\mathcal{L}$. Within the formalism of the SM, interactions between particles themselves result from the exchange of force-carrying gauge bosons. Vector fields associated to gauge bosons arise from requiring $\mathcal{L}$ to be invariant under local gauge transformations. Interactions are realised as couplings between fields in the SM $\mathcal{L}$. Each gauge symmetry implies a global symmetry with a conserved charge and has a corresponding gauge group. The gauge group of the SM G_{SM} is .

$$G_{SM} = \underbrace{SU(3)_C}_{QCD} \times \underbrace{SU(2)_L \times U(1)_Y}_{EW}, \quad (2.1.5)$$

which is the direct product of the gauge groups for Quantum Chromodynamics and unified electroweak theory [4].

2.1.4 Quantum chromodynamics

Quantum Chromodynamics (QCD) is the QFT of the strong interaction. The symmetry associated to QCD is invariance under $SU(3)_C$ local transformations. QCD is mediated by eight massless gluons corresponding to the eight generators¹ of the $SU(3)_C$ gauge

¹The generators of $SU(3)_C$ T^a are related to the Gell-Mann matrices λ^a by $T^a = \frac{1}{2}\lambda^a$.

symmetry. There are three conserved *colour* charges in QCD, r , g and b . Quarks exist in one of these colour states. Antiquarks carry the opposite colour charge to quarks, $\bar{r}$, $\bar{g}$ and $\bar{b}$. Gluons carry a combination of colour and anticolour, which allows for colour-changing interactions. Gluons are an octet of coloured states.

The QCD coupling constant α_S describes the strength of the strong interaction. The value of α_S depends on the energy scale of the interaction being considered. At high energies (small distances), such as those produced in collider experiments, α_S is small enough such that perturbation theory can be used to compute cross sections; this is known as *asymptotic freedom*. At low energies (large distances), α_S becomes large, QCD becomes non-perturbative and *colour confinement* is observed. This states that objects with non-zero colour charge, quarks and gluons, cannot propagate as free particles and must be bound in colourless states, called *hadrons*. Common types of hadrons include *mesons*, which consist of a quark and an antiquark, and *baryons* (*antibaryons*), which consist of three quarks (three antiquarks).

Colour confinement can be qualitatively understood by considering two free quarks. The two quarks exchange virtual gluons, creating a colour field. As the quarks move further apart, attractive gluon self-interactions cause the field lines to squeeze, increasing the energy stored within the field. Eventually, it becomes energetically favourable to produce a quark antiquark pair from the vacuum, to reduce the energy needed to maintain the separation. This process continues until the quarks and antiquarks have low enough energy to form colourless hadrons. At collider experiments, this *hadronisation* process means quarks and gluons are usually observed as sets of collimated hadrons, known as *hadronic jets*, as depicted in Figure 2.2.

2.1.5 Electroweak theory

The local gauge symmetry associated to the weak interaction $SU(2)_L$ is satisfied by introducing two charged gauge fields and one neutral gauge field. The conserved quantity is the third component of the weak isospin I_W^3 .

The weak charged-current interaction, mediated by the $W^\pm$ bosons, couples only to left-handed particle chiral states (as indicated by the subscript in $SU(2)_L$) and right-

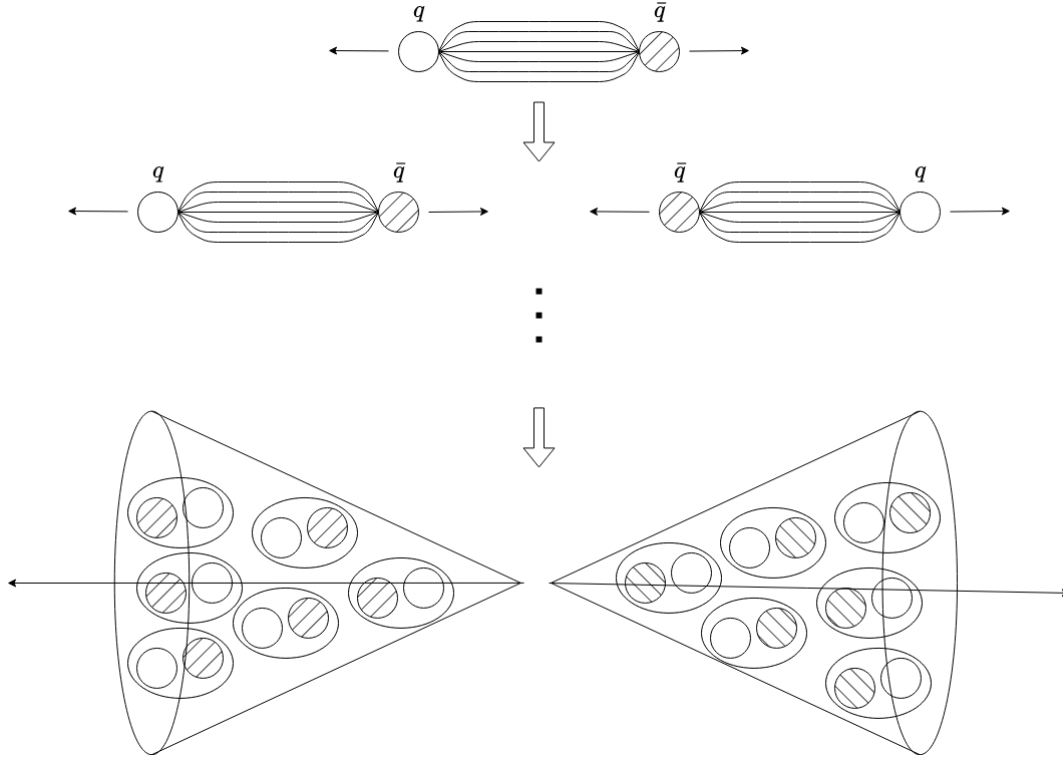


Figure 2.2: The production of hadronic jets through hadronisation. As two quarks produced with opposite momenta in a high energy collision separate, the energy contained in the colour field between the quarks increases, until it becomes energetically favourable to produce a quark antiquark pair. This process continues until the quarks and antiquarks have low enough energy to form hadrons. The result is two collimated hadronic jets.

handed antiparticle chiral states. *Chirality* is a Lorentz invariant property that is equivalent to *helicity*, the projection of the spin onto the direction of momentum, in the zero mass limit.

In the electroweak model of Glashow, Salam and Weinberg (GSW), electromagnetism and the weak force become unified at high energies. The $U(1)$ gauge symmetry of Quantum Electrodynamics (QED), the QFT of electromagnetism, is replaced by a $U(1)_Y$ local gauge symmetry, which has its own gauge field that couples to a new charge called *weak hypercharge* Y . This field mixes with the weak neutral gauge field to give rise to the physical Z^0 and γ fields. The Z^0 boson can couple to both left and right-handed chiral states, but not equally, as its associated current has contributions from both the weak interaction and weak hypercharge fields.

In the GSW model, the underlying gauge symmetry of the electroweak sector is the direct product of the $U(1)_Y$ of weak hypercharge and the $SU(2)_L$ of the weak interaction. The electric charge Q is related to Y and I_W^3 by

Fermion	I_W^3	Y	Q
e_L^-, μ_L^-, τ_L^-	$-\frac{1}{2}$	-1	-1
e_R^-, μ_R^-, τ_R^-	0	-2	-1
$\nu_{e,L}, \nu_{\mu,L}, \nu_{\tau,L}$	$\frac{1}{2}$	-1	0
u_L, c_L, t_L	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$
u_R, c_R, t_R	0	$\frac{4}{3}$	$\frac{2}{3}$
d_L, s_L, b_L	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$
d_R, s_R, b_R	0	$-\frac{2}{3}$	$-\frac{1}{3}$

Table 2.1: Summary of the electroweak quantum numbers for left- and right-handed chiral fermions.

$$Q = \frac{Y}{2} + I_W^3. \quad (2.1.6)$$

Left-handed particle chiral states are grouped into weak isospin doublets of the form

$$\begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L, \quad (2.1.7)$$

where l is e, μ or τ , q_u is an up-type quark u, c or t , and q_d is a down-type quark d, s, b . Right-handed particle chiral states are placed into weak isospin singlets. An overview of the electroweak quantum numbers for fermions is shown in Table 2.1.

Charged-current interactions involving $W^\pm$ bosons allow quarks and leptons to transform into their doublet counterparts, for example, a left-handed electron into an electron neutrino. Additionally, charged-current weak interactions involving quarks do not conserve flavour. This is because the weak eigenstates, that appear in the weak interaction as members of quark isospin doublets, are distinct from the quark mass eigenstates, freely propagating states of definite mass that are created through interactions with the Higgs field (see §2.1.6). The mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The relative strength of a particular quark flavour changing interaction is determined by the relevant element of the CKM matrix.

2.1. THE STANDARD MODEL OF PARTICLE PHYSICS

2.1.6 The Higgs mechanism

While the local gauge principle provides a neat description of interactions within the SM, it seemingly prohibits mass terms for gauge bosons. This leads to an inconsistency with the observation of large masses for the $W^\pm$ and Z^0 bosons. The solution is the *Higgs mechanism*, which can generate masses for the electroweak gauge bosons while preserving the local gauge invariance of the SM.

In the SM, the Higgs mechanism involves the breaking of the *global* $SU(2)_L \times U(1)_Y$ symmetry by the vacuum expectation value (VEV) of a weak isospin doublet of complex scalar fields, called the *Higgs multiplet*,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.1.8)$$

The Lagrangian for ϕ is

$$\mathcal{L}_\phi = (D_\mu \phi)^* (D^\mu \phi) - V(\phi). \quad (2.1.9)$$

Here, local $SU(2)_L \times U(1)_Y$ gauge invariance is achieved through the covariant derivative

$$D_\mu = \partial_\mu + i\frac{g_W}{2}\sigma_j W_\mu^j + i\frac{g'Y}{2}B_\mu, \quad (2.1.10)$$

where W_μ^j , with $j = 1, 2, 3$, are the three $SU(2)_L$ gauge fields, B_μ is the weak hypercharge gauge field, σ_j are the Pauli matrices and g_W and g' are the weak and hypercharge coupling constants.

The potential is given by

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.1.11)$$

When $\mu^2 < 0$ and $\lambda > 0$, $V(\phi)$ has an infinite set of degenerate minima, each with a non-zero VEV v . The choice of the physical vacuum state will spontaneously break the symmetry of $\mathcal{L}_\phi$. The field can then be expanded about the physical vacuum state and written in the *unitary gauge*, in which fields correspond to physical particles, as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (2.1.12)$$

where $H(x)$ is the *Higgs field*.

Expanding Eq. (2.1.9) using Eq. (2.1.12) gives

$$\begin{aligned} \mathcal{L}_\phi = & \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{4}g_W^2 W_\mu^- W^{+\mu}(v + H)^2 \\ & + \frac{1}{8}(g_W^2 + g'^2)Z_\mu Z^\mu(v + H)^2 - \frac{\mu^2}{2}(v + H)^2 - \frac{\lambda}{4}(v + H)^4, \end{aligned} \quad (2.1.13)$$

after transforming into the basis of *physical* boson fields given by

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad (2.1.14)$$

$$Z_\mu = \frac{1}{\sqrt{g_W^2 + g'^2}}(g_W W_\mu^3 - g' B_\mu), \quad (2.1.15)$$

and

$$A_\mu = \frac{1}{\sqrt{g_W^2 + g'^2}}(g_W W_\mu^3 + g' B_\mu). \quad (2.1.16)$$

Through the Higgs mechanism, $\mathcal{L}_\phi$ now yields mass terms for the $W_\mu^\pm$ and Z_μ fields, corresponding to masses for of the $W^\pm$ and Z^0 bosons of

$$m_W = \frac{g_W v}{2} \quad (2.1.17)$$

and

$$m_Z = \frac{v}{2}\sqrt{g_W^2 + g'^2}. \quad (2.1.18)$$

These masses are introduced in a gauge invariant manner. The Higgs field also acquires a mass term, and can be interpreted as a scalar particle with mass

$$m_H = \sqrt{2\lambda}v, \quad (2.1.19)$$

2.1. THE STANDARD MODEL OF PARTICLE PHYSICS

called the *Higgs boson*, described further in §2.1.7. The scale of the masses for both the gauge bosons and the Higgs boson is set by the VEV of the Higgs potential. There is no mass term for the A_μ field, hence the photon remains massless.

Fermion Masses

In addition to generating the masses for the $W^\pm$ and Z^0 bosons, the Higgs mechanism also generates the masses of the fermions through gauge invariant Yukawa terms. For F fermion generations, these are of the form

$$\mathcal{L}_{\text{Yukawa}} = - \sum_{m,n=1}^F [Y_{mn}^u \bar{q}_{mL} \tilde{\phi} u_{nR} + Y_{mn}^d \bar{q}_{mL} \phi d_{nR}] \quad (2.1.20)$$

$$+ Y_{mn}^e \bar{l}_{mL} \phi e_{nR} + Y_{mn}^\nu \bar{l}_{mL} \tilde{\phi} \nu_{nR}] + h.c., \quad (2.1.21)$$

where Y^u, Y^d, Y^e, Y^ν are $F \times F$ matrices containing experimentally determined Yukawa couplings, l_L and q_L are weak isospin doublets, u_R, d_R, e_R and ν_R are weak isospin singlets, ϕ is the Higgs multiplet and $\tilde{\phi}$ is its conjugate. After spontaneous symmetry breaking, $\tilde{\phi}$ in the unitary gauge is given by

$$\tilde{\phi}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix}. \quad (2.1.22)$$

Inserting the unitary gauge expressions for ϕ and $\tilde{\phi}$ in Eqs. (2.1.12) and (2.1.22) into Eq. (2.1.20) yields

$$\mathcal{L}_{\text{Yukawa}} = - \sum_{m,n=1}^F [Y_{mn}^u \bar{u}_{mL} (\frac{v + H}{\sqrt{2}}) u_{nR} + (d, e, \nu) \text{ terms}] + h.c.. \quad (2.1.23)$$

The Lagrangian now contains mass terms for each quark and charged lepton as well as interactions between the fermions and the Higgs boson. Terms involving neutrinos can be omitted as there are no right-handed neutrinos in the SM, meaning they cannot acquire mass through the Yukawa interaction². The mass of each fermion (quark or

²Neutrinos have small, non-zero masses, but are generally considered massless within the SM.

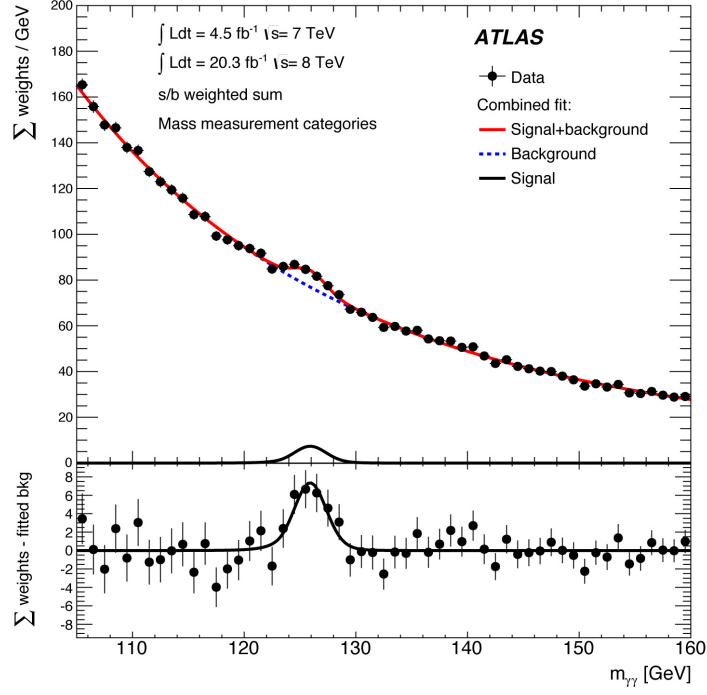


Figure 2.3: Reconstructed invariant mass distribution of the photon pair in candidate $H \rightarrow \gamma\gamma$ decays at the ATLAS experiment [10].

charged-lepton) m_f is proportional to its Yukawa coupling g_f and the VEV of the Higgs, and is given by

$$m_f = \frac{1}{\sqrt{2}} g_f v. \quad (2.1.24)$$

2.1.7 The Higgs boson

In 2012, both the ATLAS and CMS experiments reported evidence of a new particle compatible with the expected properties of the Higgs boson [7, 8]. By fitting the invariant mass of the photon pair (see Figure 2.3) in candidate $H \rightarrow \gamma\gamma$ decays and of the four charged leptons in candidate $H \rightarrow ZZ^* \rightarrow 4l$ decays, the mass of the new particle was found to be approximately 125 GeV. Since the discovery of the Higgs boson, extensive research has been carried out to better understand its properties and further test the predictions of the SM [9].

The SM Higgs boson can be produced in a variety of ways at proton-proton (pp) collider experiments, such as the Large Hadron Collider. The most dominant processes are:

2.1. THE STANDARD MODEL OF PARTICLE PHYSICS

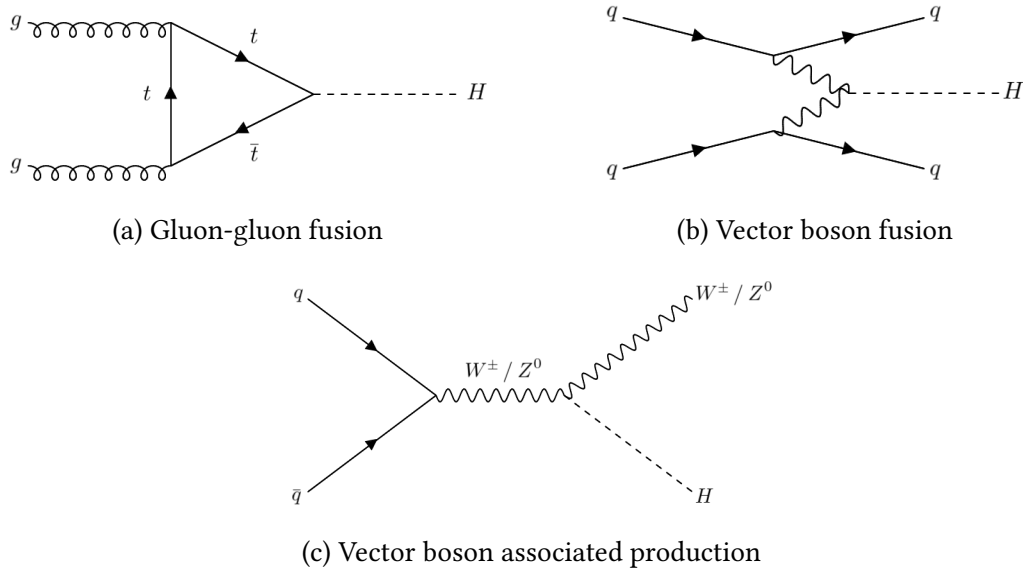


Figure 2.4: Dominant Higgs boson production modes in proton-proton collisions.

$\sqrt{s}$ (TeV)	Production cross section (pb) for $m_H = 125$ GeV			
	ggF	VBF	WH	ZH
13	$48.6^{+5.6\%}_{-7.4\%}$	$3.78^{+2.1\%}_{-2.1\%}$	$1.37^{+2.0\%}_{-2.0\%}$	$0.88^{+4.1\%}_{-3.5\%}$
14	$54.7^{+5.6\%}_{-7.4\%}$	$4.28^{+2.1\%}_{-2.1\%}$	$1.51^{+1.8\%}_{-1.9\%}$	$0.99^{+4.1\%}_{-3.7\%}$

Table 2.2: Production cross sections for a Higgs boson with $m_H = 125$ GeV in pp collisions at centre of mass energies $\sqrt{s}$ of 13 TeV and 14 TeV [12].

- *gluon-gluon fusion* (ggF), depicted in Figure 2.4a.
- *vector boson fusion* (VBF), depicted in Figure 2.4b.
- *vector boson associated production* (VH), depicted Figure in 2.4c.

These reflect the preference of the Higgs to couple to heavy particles. The production cross section for a Higgs boson with $m_H = 125$ GeV in pp collisions at centre of mass energies $\sqrt{s}$ of 13 TeV and 14 TeV for each of these processes is listed in Table 2.2.

The Higgs also has many potential decay modes. Figure 2.5 shows the branching ratios for the dominant decay modes of the Higgs boson as a function of its mass. For $m_H = 125$ GeV the most common decay mode is $H \rightarrow b\bar{b}$ with $BR(H \rightarrow b\bar{b}) = 58.2\%$.

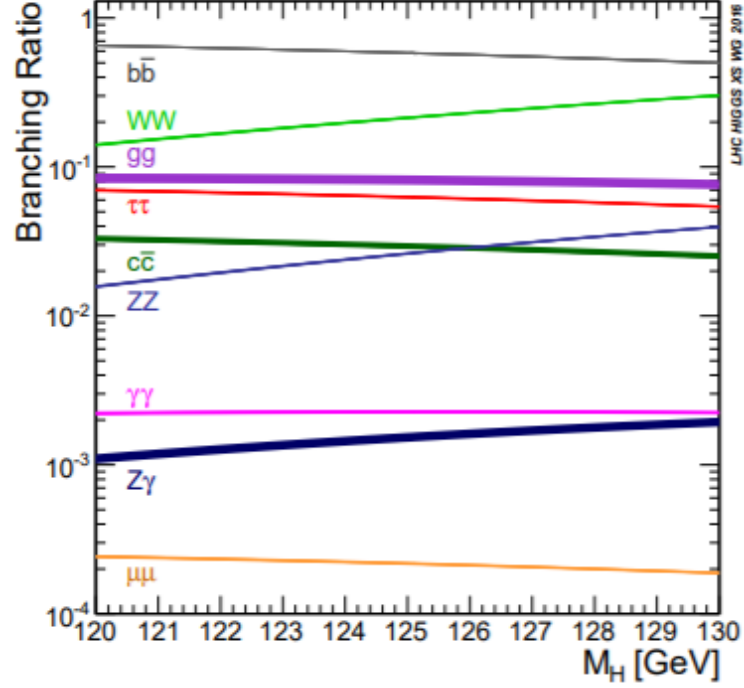


Figure 2.5: Branching ratios of the Higgs boson for masses around 125 GeV [11].

2.2 LIMITATIONS OF THE STANDARD MODEL

The success of the SM in providing accurate descriptions of experimental data is considered a crowning achievement within the field of physics. However, there are a number of significant observed phenomena that the SM fails to explain. These include, but are not limited to, the nature of dark matter, matter-antimatter asymmetry, and a lack of unification with gravity. Additionally, there are several theoretical concerns that, guided by a wanting of naturalness within the SM, suggest an incomplete understanding of fundamental principles, such as the hierarchy problem and the strong CP problem.

2.2.1 Dark matter

The existence of a non-luminous, non-baryonic species of matter is supported by a number of cosmological observations [13], most notably the measurement of the rotation curves of spiral galaxies. Known as *dark matter*, it is thought to comprise as much as $\sim 27\%$ of the energy of the Universe, approximately five times that of visible matter. However, no plausible dark matter candidate exists within the SM, and its nature is completely unknown. Dark matter is one of the most compelling arguments for the need for

2.2. LIMITATIONS OF THE STANDARD MODEL

new physics outside of the SM.

Most experimental searches for dark matter have focused on detecting weakly-interacting massive particles (WIMPs). However, WIMP-like dark matter is becoming increasingly constrained (see e.g. [14, 15]). One alternative to WIMPs is the existence of a dark sector in which dark matter particles are neutral under the SM gauge group, and instead interact with the SM through BSM mediator particles. If the lightest dark sector state is more than half as massive as the mediator, decays of the mediator to this state are forbidden, and the mediator must instead decay back to SM particles. This decay will be highly suppressed due to the small coupling between the SM and the mediator, leading to potentially macroscopic decay lengths (see §2.3.1).

2.2.2 Hierarchy problem

The hierarchy problem concerns the large discrepancy between the scales of the weak force and gravity (known as the *Planck scale* $M_P \approx 10^{19}$ GeV), and its implication on the formulation of the Higgs mass.

The physical mass of the Higgs boson is $m_H^2 = m_0^2 + \Delta m_H^2$, where m_0^2 is the tree-level mass and

$$\Delta m_H^2 \sim \frac{\lambda^2}{16\pi^2} \Lambda^2 \quad (2.2.1)$$

is the quantum correction due to loop-level diagrams, such as the top quark loop shown in Figure 2.6, where λ is an $\mathcal{O}(1)$ coupling and Λ is the scale up to which the SM is a valid theory. If $\Lambda \sim M_P$, the quadratic dependence in Λ of these corrections means m_0^2 and Δm_H^2 must cancel to within approximately one part in 10^{36} to recover the correct Higgs mass. This is achievable by *fine-tuning* the free parameters of the SM, at the expense of maintaining the overall *naturalness* of the theory.

A more natural solution to the hierarchy problem is possible if $\Lambda \lesssim 1$ TeV, implying new physics at the weak scale $M_{weak} \sim 10$ GeV – TeV. Consequently, theories aimed at solving the hierarchy problem often posit new particles with mass around M_{weak} in order to cancel the quadratic divergences incurred by the SM particles, thereby stabilising the Higgs mass without the need for fine-tuning. Many of these new particles are attractive

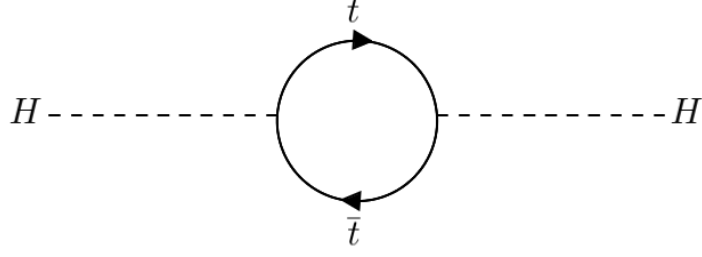


Figure 2.6: Radiative corrections to the Higgs boson mass via a top quark loop.

dark matter candidates. The analysis presented in this thesis considers models that seek to solve the hierarchy problem.

2.3 PHYSICS BEYOND THE STANDARD MODEL

The shortcomings of the SM discussed in §2.2 give a strong indication of new physics existing beyond the Standard Model (BSM). The question is then what this new physics might be. There is a continued effort to develop a wide range of new BSM models. For experimentalists, it is inefficient to rigorously test each model. Instead, developing analyses that are sensitive to features found in number of models, while remaining agnostic to any one in particular, can be more productive. One hallmark of many BSM models is long-lived particles (LLPs).

The analysis presented in §5 is a search for a neutral, scalar LLP produced in an exotic decay of the Higgs boson at the ATLAS detector. The work presented in §6 relates to the development and verification of analysis techniques for future LLP searches. This section contains an overview on LLPs, in §2.3.1, followed by a discussion of exotic Higgs decays as a well motivated potential source of new LLPs, in §2.3.2.

2.3.1 Long-lived particles

Particles within the SM have lifetimes that span many orders of magnitude, as seen in Figure 2.7. LLPs decay over macroscopic distances, significantly greater than the spatial resolution of the detector ($\mathcal{O}(\gtrsim 10 \mu\text{m})$). Examples within the SM include B mesons and τ leptons.

The origin of long lifetimes can be understood by referring back to Eq. (2.1.2) and

2.3. PHYSICS BEYOND THE STANDARD MODEL

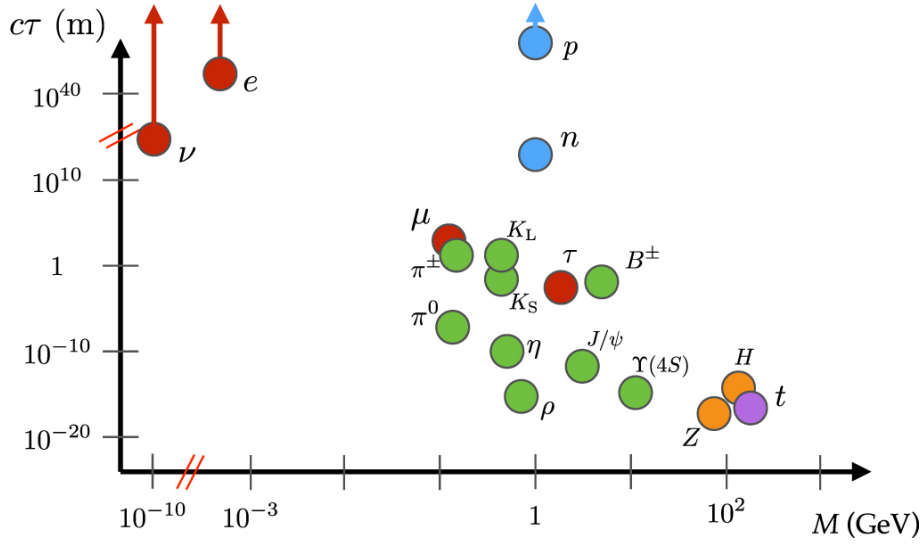


Figure 2.7: Particle lifetimes $c\tau$, in metres, as a function of mass, in GeV, for Standard Model fermions (red), hadrons (blue), mesons (green), bosons (orange) and quarks (purple) [16].

Eq. (2.1.4). For a particle to be long-lived, either the matrix element or the phase space available for its decay must be small. A matrix element can become small if there is an approximate symmetry that prohibits the decay, such as baryon number conservation in the case proton decay, or if it contains a small effective coupling constant. A small coupling can be further attributed to either a small dimensionless coupling, such as a Yukawa coupling or CKM matrix element, or a dimensional scale corresponding to a heavy mediating particle, such as the W boson in weak interactions. Phase space may be suppressed by approximate symmetries that result in near degenerate mass states, such as in the beta-decay of a neutron to a proton and leptons.

Given their ubiquity within the SM, it is unsurprising that many BSM models predict new LLPs. For a comprehensive study of the theoretical motivations for BSM LLPs, readers are referred to Ref. [17]. As the most revealing evidence of new physics would stem from the discovery of a new particle, interest in searching for LLPs has increased in recent years [16, 17, 18, 19]. LLPs are likely to evade searches for particles that decay *promptly* (near the collision point), for which the ATLAS detector and standard analysis software were originally conceived. However, were an LLP to decay within the fiducial volume of a detector, it could produce a unique experimental signature with limited SM background, which could be targeted using dedicated analysis techniques.

For example, the position of the decay of an LLP within the inner tracking detectors

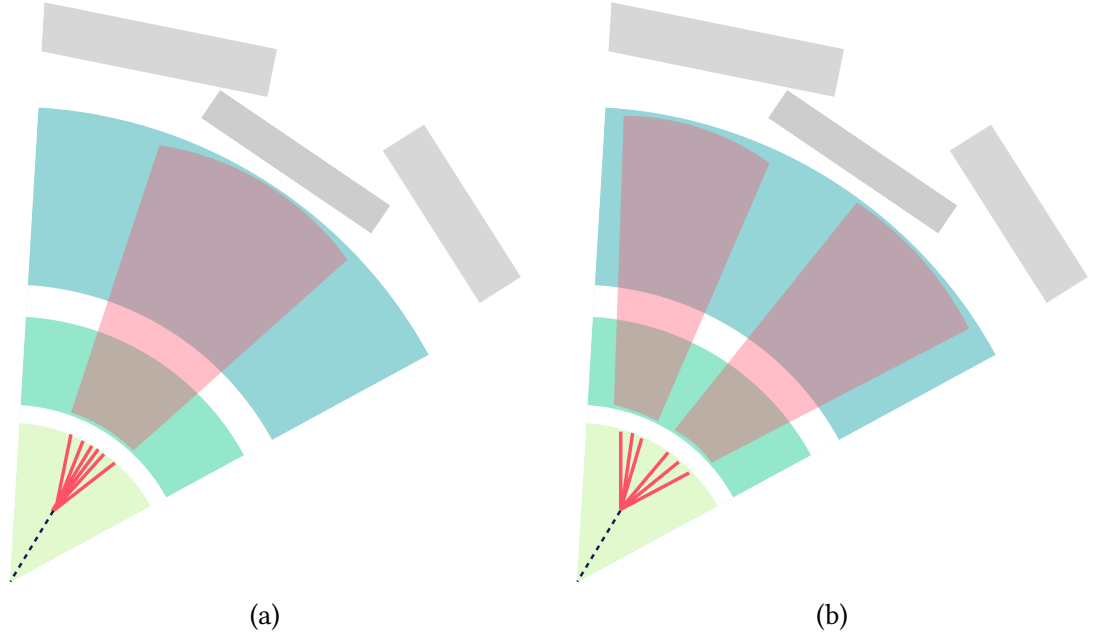


Figure 2.8: A long-lived particle decaying hadronically into a single jet (a) [20] and two jets (b) [21].

(see §3.2.2) of ATLAS could be reconstructed using the trajectories of its charged decay products as a vertex displaced from the collision point (see §4.3.2). Examples of this displaced vertex (DV) signature are shown in Figure 2.8 for a hadronically decaying LLP, where the number of reconstructable jets is dependent on the angular separation of the decay products. Due to the exponential nature of particle decays, inner tracking detector searches for displaced vertices are likely to be sensitive over a wide lifetime range. The search presented in §5 exploits a displaced vertex signature within the inner tracking detectors of the ATLAS experiment.

2.3.2 Exotic Higgs decays

A well established avenue offering discovery potential is the search for new physics produced through exotic decays of the Higgs boson to a hidden sector (known as a Higgs portal model) [22, 23]. Current estimates using ATLAS measurements suggest the upper limit on the branching ratio of the Higgs boson to undetected BSM final states could be as high as 12% [9]. Additionally, the highly suppressed decay width of the SM Higgs, due to small Yukawa couplings, loop factors or multibody phase space limitations, means even small couplings to BSM particles could produce sizeable decay modes and potential signals at detectors.

2.3. PHYSICS BEYOND THE STANDARD MODEL

A simple model leading to exotic Higgs decays is achieved by adding a new, real scalar field S to the SM that couples to the Higgs [22, 23, 24]. The potential for this system, after imposing a Z_2 symmetry $S \rightarrow -S$, is given by

$$V(S, \phi) = \frac{\mu_s^2}{2} S^2 - \frac{\lambda_s^2}{4!} S^4 - \frac{\kappa}{2} S^2 |\phi|^2 + \mu^2 |\phi|^2 - \lambda |\phi|^4, \quad (2.3.1)$$

where ϕ is the Higgs multiplet defined in Eq. (2.1.8) and the couplings μ_s , λ_s and κ represent free parameters in the model. Dependent on the choice of couplings, S may have a non-zero VEV, and the two fields can be written as $S = s + v_s$ and $\phi = (H + v_H)/\sqrt{2}$ after symmetry breaking. The resulting potential contains a term with Hs^2 , that generates the exotic $H \rightarrow ss$ decay with $BR(H \rightarrow ss)$ determined by the strength of its coupling constant, as well as a term containing H^2s^2 that allows s and H to mix into mass eigenstates. This mixing allows s to inherit couplings to the SM fermions and gauge bosons. The mixing angle between s and H determines $\Gamma(s \rightarrow \text{SM})$ and therefore also the lifetime of s . Provided this decay width is sufficiently small, s may be long-lived. Since s couples to the SM in the same manner as the Higgs, for masses of s above ~ 10 GeV, s should predominantly decay to b -quarks.

The model described above provides a simple process through which the Higgs boson decays to two long-lived, neutral scalars s which then proceed to decay to SM particles. This *signal process* is the basis of the search described in §5 and is shown in Figure 2.9. Despite its simplicity, similar mechanisms arise in a number of more complex models describing phenomena such as dark matter, or offering solutions to the hierarchy problem, such as the many realisations of neutral naturalness, including folded supersymmetry [25] and Quirky Little Higgs [26]. Therefore, this model can potentially provide constraints on a wide range of BSM scenarios.

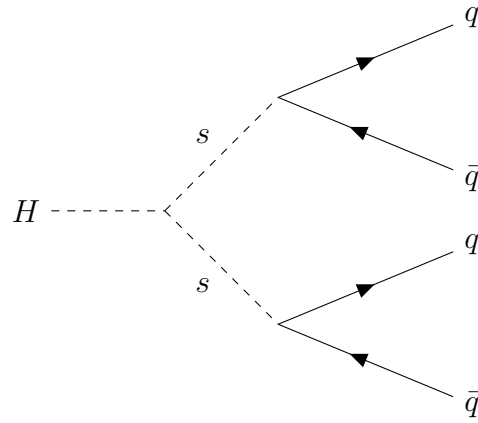


Figure 2.9: Exotics Higgs decay $H \rightarrow ss \rightarrow qq \bar{q} \bar{q}$.

3

THE ATLAS EXPERIMENT

The work in this thesis uses data collected by the ATLAS experiment at the Large Hadron Collider during high energy pp collisions. In this chapter, a description of these two machines is provided, with particular attention given to aspects of their operation and performance that are most relevant to later chapters. The Large Hadron Collider is introduced in §3.1. An overview of the ATLAS detector and its constituents is given in §3.2.

3.1 THE LARGE HADRON COLLIDER

The Large Hadron Collider (LHC) [27] is a multi-purpose particle accelerator and collider located at the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. Occupying a tunnel approximately 100 m underground, the LHC is capable of accelerating both protons and heavy ions along a 27 km ring. Within this ring, two particle beams travel in opposite directions through separate beampipes and collide at one of four interaction points which house the ATLAS [28], CMS [29], LHCb [30] and ALICE [31] detectors. ATLAS, described further in §3.2, and CMS are general purpose detectors, LHCb specialises in B-physics, and ALICE is a dedicated ion experiment.

The LHC has been operational since 2008 and is currently during its third period of data-taking. Run 1 took place between 2010 and 2012 at $\sqrt{s} = 7\text{--}8\text{ TeV}$. Run 2 took place from 2015 to 2018 at $\sqrt{s} = 13\text{ TeV}$. Run 3 commenced in 2022, during which the first pp collisions at $\sqrt{s} = 13.6\text{ TeV}$ were measured. Between runs, the LHC is shut

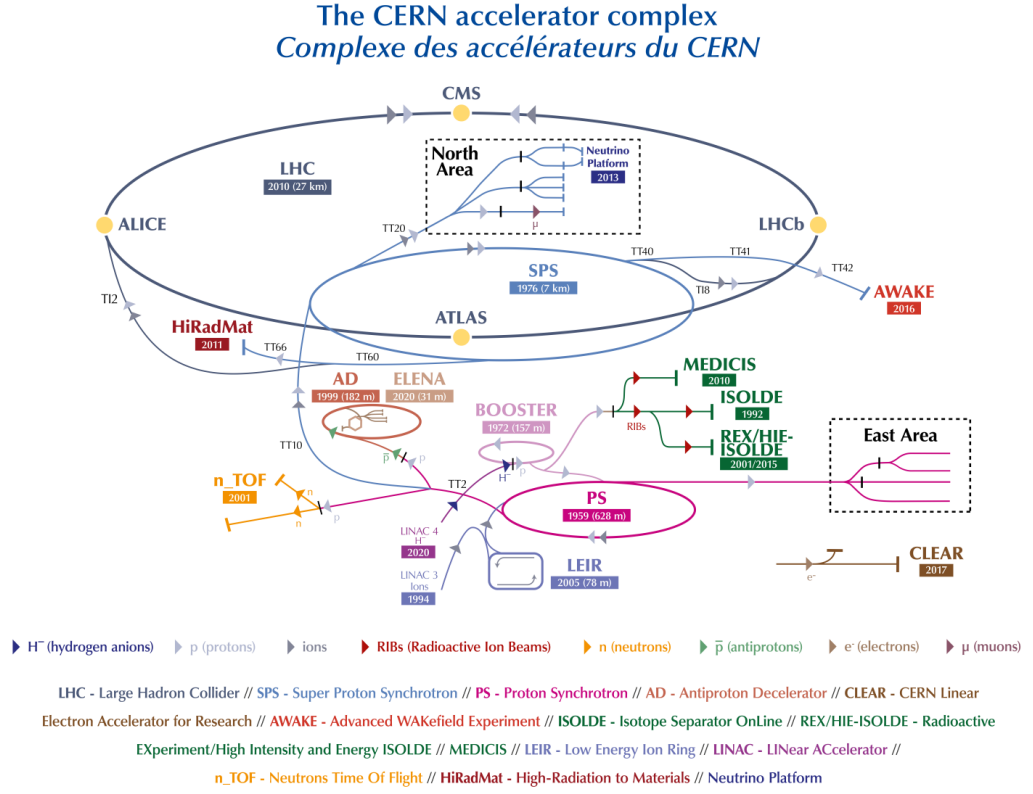


Figure 3.1: An overview of the accelerator complex at CERN in 2022 [33].

down for upgrades.

Figure 3.1 shows a diagram of the CERN accelerator complex. The LHC ring is the last in a chain of accelerators that successively increase the beam to higher energies. During Runs 1 and 2, the first step in this chain was the linear accelerator *Linac2*, which accelerated protons to 50 MeV. For Run 3, *Linac2* was replaced by its successor *Linac4* [32], which accelerates negative hydrogen ions to 160 MeV, after which they are stripped of their electrons, leaving only protons. Next, the protons pass through three circular accelerators, the *Proton Synchrotron Booster*, the *Proton Synchrotron*, and the *Super Proton Synchrotron*, which increase their energy to 1.4, 25 and 450 GeV, respectively, before being split into two beams and injected into the main LHC ring. The ring comprises a series of radio frequency cavities, which accelerate the beams using oscillating electric fields, and superconducting magnets to steer and focus their path. In Run 2, each proton beam had an energy of 6.5 TeV. This was increased to 6.8 TeV for Run 3. Throughout the injection chain, protons in the beam are organised into bunches. Within the LHC ring, bunches are currently spaced at 25 ns intervals and contain approximately 10^{11} protons.

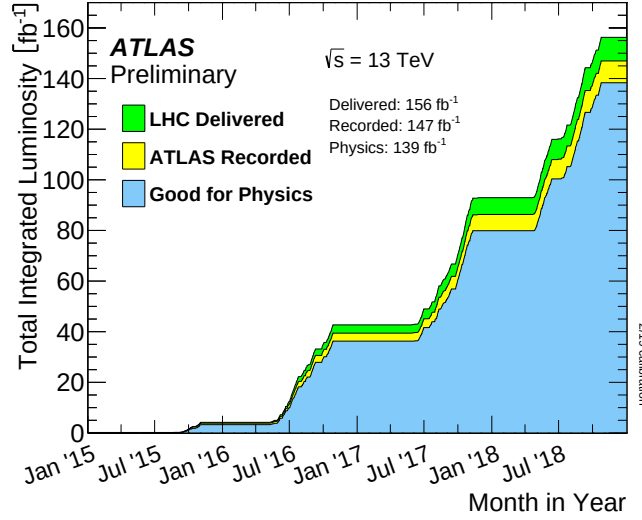


Figure 3.2: The cumulative integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow) and classed as good quality data (blue) between 2015 and 2018 [36].

3.1.1 Luminosity and pile-up

In addition to achieving high $\sqrt{s}$, accelerators are carefully designed to produce collisions at a rate that generates the statistics required for a successful physics program. The instantaneous luminosity characterises the potential collision rate and its relation to the conditions of the beam. The instantaneous luminosity for a single colliding bunch pair $\mathcal{L}_b$ is given by

$$\mathcal{L}_b = \frac{f_r n_1 n_2}{2\pi \Sigma_x \Sigma_y} \quad (3.1.1)$$

where f_r is the revolution frequency, n_1 and n_2 are the number of protons in each bunch and $\Sigma_{x,y}$ are the horizontal and vertical convolved beam sizes [34, 35]. The total instantaneous luminosity $\mathcal{L}$ is the sum of $\mathcal{L}_b$ over all colliding bunch pairs. The design instantaneous luminosity of the LHC is $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, although a maximum of $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ was observed at ATLAS during Run 2.

The integrated luminosity $\mathcal{L}_{int}$ is given by

$$\mathcal{L}_{int} = \int \mathcal{L} dt \quad (3.1.2)$$

and is a measure of the statistics accumulated through collisions over some time frame. Figure 3.2 shows the cumulative integrated luminosity versus time delivered to

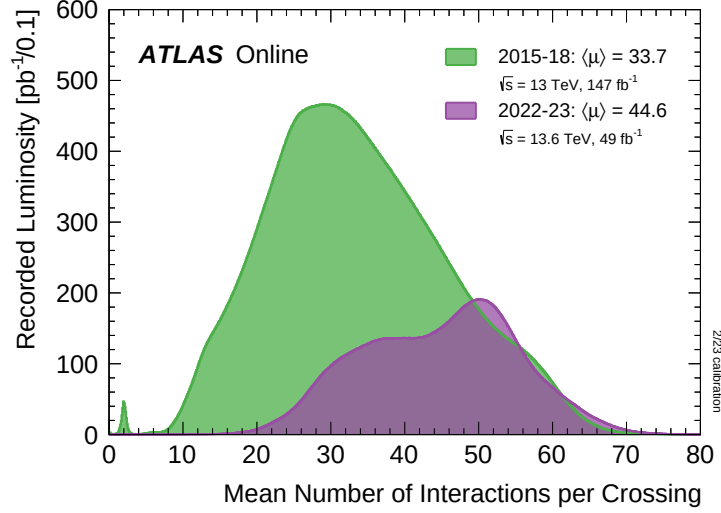


Figure 3.3: The luminosity-weighted distribution of pile-up μ for pp collisions in 2015-2018 at 13 TeV and in 2022-2023 at 13.6 TeV [37].

ATLAS, recorded by ATLAS and classed as good quality data between 2015 and 2018. The full Run 2 dataset comprises 139 fb^{-1} of integrated luminosity and is used in §5.

The integrated luminosity relates the number of times a process occurs at the LHC, N , to the cross section for the process, σ , as follows

$$N = \mathcal{L}_{int} \sigma. \quad (3.1.3)$$

The high luminosity at the LHC results in multiple pp interactions per bunch crossing, of which a single hard scatter interaction is usually of interest. The time-averaged number of pp interactions per bunch crossing μ is known as pile-up. Pile-up can be *in-time*, meaning from an interaction that occurred in the same bunch crossing as the primary interaction of interest, or *out-of-time*, where the interaction occurs in a different bunch crossing. Out-of-time pile-up is possible when the readout time of the detecting element is greater than the bunch spacing. Larger μ leads to several interaction vertices and makes it more difficult for reconstruction algorithms to disambiguate the effects of hard scatter from softer, inelastic interactions. Figure 3.3 shows the luminosity-weighted distribution of μ for pp collisions during Runs 2 and 3. The mean pile-up for Run-2 was $\mu = 33.7$. The mean pile-up for Run-3 during 2022 was $\mu = 42.5$.

3.2. THE ATLAS DETECTOR

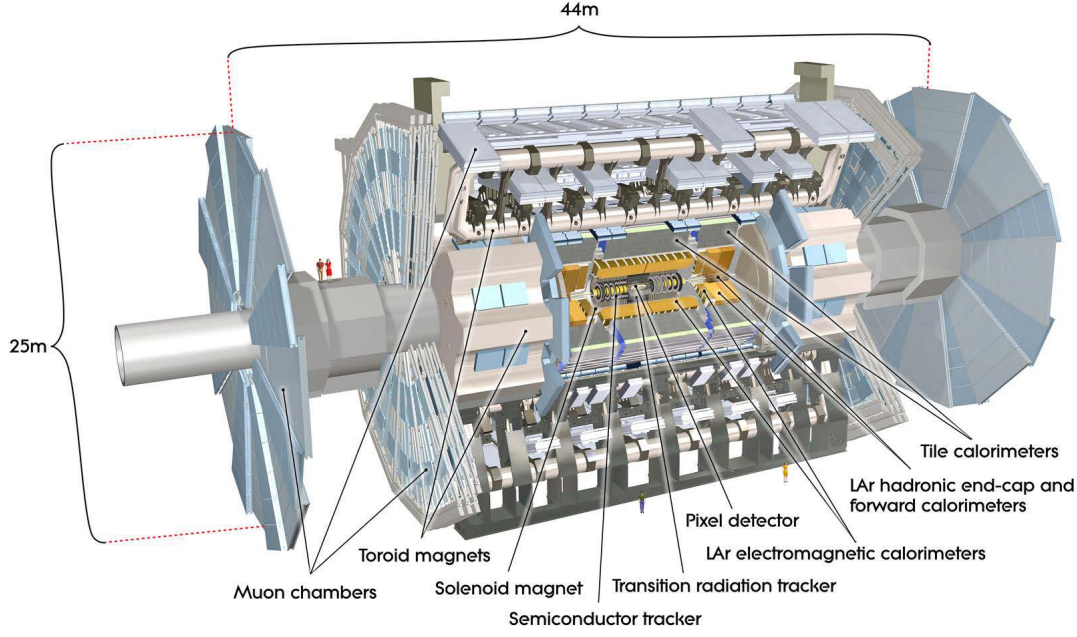


Figure 3.4: The ATLAS detector and its subdetectors [28].

3.2 THE ATLAS DETECTOR

ATLAS (A Toroidal LHC ApparatuS) is a large particle detector at the LHC designed to precisely measure the products of pp collisions and facilitate precision measurements of the SM and searches for BSM physics.

ATLAS is composed of three subdetectors that are arranged coaxially around the beampipe, forming a cylinder consisting of a central barrel and two end-cap regions that together provide close to 4π radians of solid angle coverage. With increasing radii from the beampipe, the subdetectors are: the inner detector, the calorimeter, and the muon spectrometer. Each is designed to identify specific types of particles produced in pp collisions. Figure 3.4 shows a schematic of the ATLAS detector.

3.2.1 Coordinate system and nomenclature

ATLAS uses a right-handed coordinate system with the nominal interaction point at its origin in the centre of the detector. The beam direction defines the z -axis with the positive z direction pointing counter-clockwise around the LHC ring. The x - y plane is transverse to the beam direction. The positive x -axis is directed toward the centre of the LHC ring and the positive y -axis is directed vertically upwards. The azimuthal angle ϕ

is measured around the z -axis where $\phi = 0$ points in the positive x direction. The polar angle θ is the angle from the z -axis where $\theta = 0$ is in the positive z direction. The radius $r = \sqrt{x^2 + y^2}$.

The projection of the momentum $p = (p_x, p_y, p_z)$ in the x - y plane is the transverse momentum, defined as $p_T = |p|\sin\theta$. Similarly, the transverse energy $E_T = |E|\sin\theta$. At collider experiments, $p_T = 0$ prior to a collision, therefore conservation laws can be used to infer the transverse kinematics of particles produced in collisions.

The rapidity, defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \quad (3.2.1)$$

is used to express the angle of a particle with respect to the collision axis. Differences in the rapidity between two particles have the useful property of being Lorentz invariant with respect to boosts along the collision axis. However, it is frequently replaced by the *pseudorapidity*,

$$\eta = -\ln \tan(\theta/2). \quad (3.2.2)$$

Pseudorapidity is easier to calculate, as it only depends on the polar angle θ of the particle. A pseudorapidity $\eta = 0$ is a vector pointing in the transverse plane, while $\eta = \infty$ corresponds to a vector pointing along the z -axis. The terms *central* and *forward* are often used to describe particles or regions in the detector with low- and high- η , respectively. In the massless limit, pseudorapidity is equal to the rapidity.

Particles or systems of particles are usually specified in terms of (p_T, η, ϕ) . A common distance measure between particles is

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (3.2.3)$$

3.2.2 Inner detector

The Inner Detector (ID) is the innermost ATLAS subdetector. The ID provides position measurements of electrically-charged particles. As a charged particle traverses the ID, it leaves energy deposits (*hits*) in the various detecting elements. These hits are used

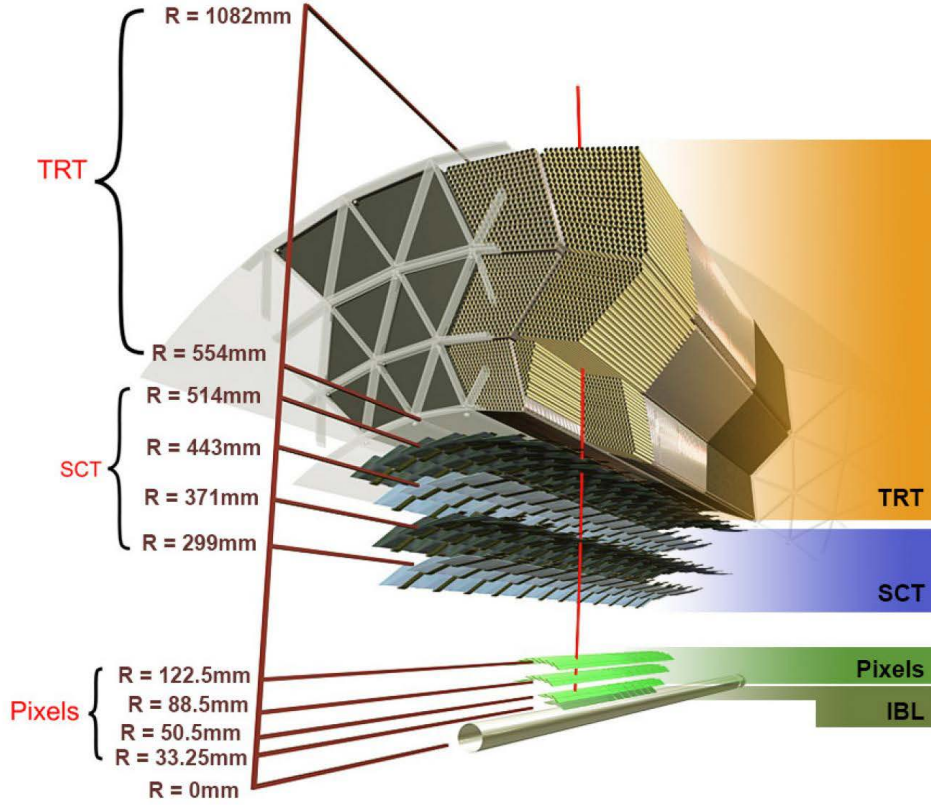


Figure 3.5: A cutaway of the barrel region of the inner detector. Each subsystem is labelled, along with its radial distance from the beamline [38].

to reconstruct the trajectory of the particle (*track*). A 2 T solenoidal magnetic field that immerses the ID bends the path of charged particles as they pass through the ID. The track is used to measure the direction, sign of the charge and momentum of the charged particle, within a pseudorapidity range $|\eta| < 2.5$. The ID comprises three complementary subsystems arranged concentrically around the beampipe: the pixel detector, the SemiConductor Tracker (SCT), and the Transition Radiation Tracker (TRT). A cutaway of the ID showing these subsystems can be seen in Figure 3.5, and a more detailed cross section in Figure 3.6.

The pixel detector consists of four layers of silicon-based pixel modules in the barrel region at $r = 33.5, 50.5, 88.5, 122.5$ mm, and three discs in each end-cap region at $|z| = 495, 580, 650$ mm. The innermost barrel layer is called the Insertable B-Layer (IBL) [39] and was installed prior to Run 2 to improve tracking efficiency and vertex resolution, as well as compensate for radiation damaged incurred by the other pixel layers during Run 1 [40]. In high μ environments, high vertex resolution (at least on the order of B hadron

lifetimes) is needed to distinguish primary vertices (defined in §4.1.2) from secondary vertices, such as those produced through B hadron or exotic LLP decays and material interactions. The IBL consists of pixels with area $50 \times 250 \mu\text{m}^2$, while the remaining layers use pixels with area $50 \times 400 \mu\text{m}^2$. The pixel detector provides precise three-dimensional spatial measurements.

Surrounding the pixel detector is the SCT. The SCT consists of silicon-microstrip pairs forming four layers in the barrel region at $r = 299, 371, 443, 514$ mm, and nine discs in each end-cap region spanning $|z| = 854$ mm to 2720 mm. In each pair, one strip is parallel to the beamline while the other is offset at a stereo angle of 40 mrad. Measurements from each strip are combined to produce three-dimensional position measurements.

The TRT is a gaseous detector consisting of 4 mm diameter straw tubes filled with a gas mixture surrounding a central wire. A particle passing through a straw tube ionises the gas mixture, producing free electrons that drift towards the central wire, generating a signal. In the barrel region, straw tubes extend from $r = 563$ mm to 1066 mm, and in the two end-caps from $|z| = 827$ mm to 2774 mm. Despite the TRT having a lower spatial resolution than the pixel and SCT detectors, it generally produces more position measurements per track (around 36) and over a longer path, further constraining the track curvature and improving momentum resolution.

The TRT also has particle identification capabilities. Spaces between straw tubes are filled with radiator material. As relativistic charged particles traverse materials of varying dielectric constants, they emit photons, known as transition radiation, at intensities dependent on their Lorentz factor. These photons are reabsorbed by the xenon in the straw tubes. The detection of transition radiation generally indicates the traversing of an electron, as opposed to a pion, since an electron with equal momentum to a pion would have a higher Lorentz factor, due to its lower mass.

3.2.3 Calorimeter

The ATLAS calorimetry system surrounds the ID and comprises two subsystems, the electromagnetic calorimeter (ECal) and the Hadronic Calorimeter (HCal). The ECal and

3.2. THE ATLAS DETECTOR

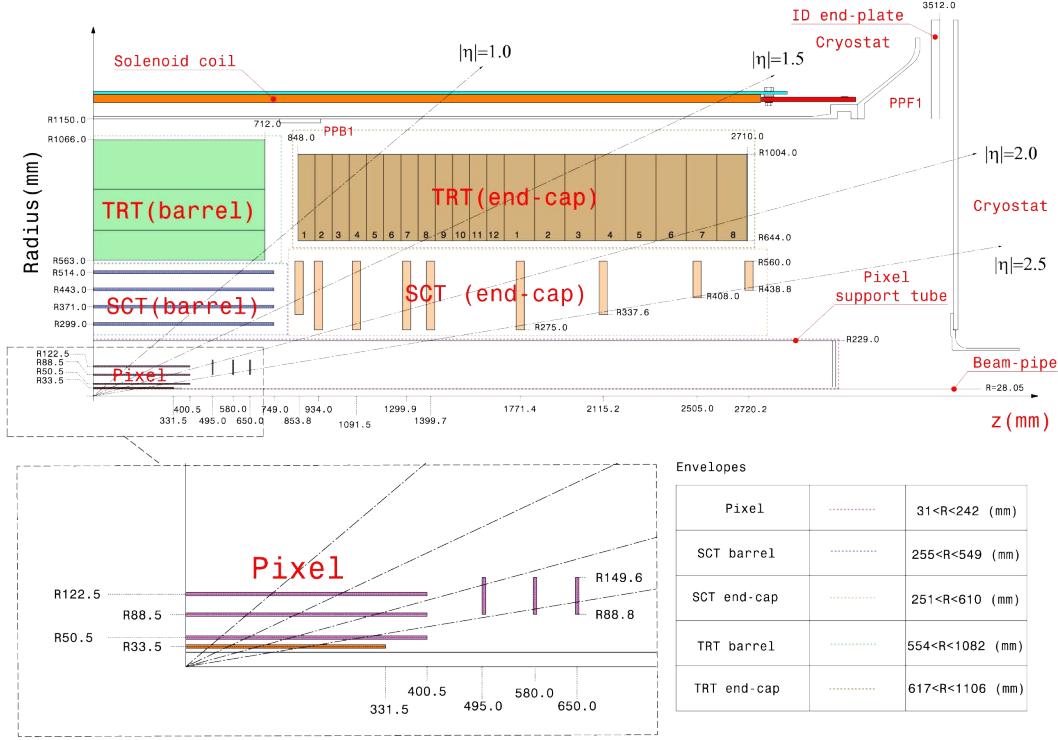


Figure 3.6: An r - z cross section of the ATLAS inner detector during Run 2. The top panel shows the whole inner detector. The bottom left panel shows a more detailed view of the pixel layer, including the Insertable B-Layer in red. The bottom right panel shows the radial envelope of each subsystem [38].

HCal use alternating layers of passive and active material to absorb and measure the energy of particles within the range $|\eta| < 4.9$. The dense passive, or absorbing, material incites the formation of electromagnetic or hadronic showers that generate detectable signals in the active layers. The energy sampled at each layer is combined to give an estimate of the total energy of the incident particle.

The ECal measures the energy of electrons and photons and consists of a barrel region ($|\eta| < 1.475$) and the two end-cap regions ($1.375 < |\eta| < 3.2$) with accordion-like geometry, allowing for complete ϕ symmetry. The ECal uses liquid argon in its active layers and lead in its absorbing layers. The ECal is characterised by the radiation length X_0 of its absorbing material, defined as the mean distance travelled by an electron in the material before it loses $1/e$ of its energy. High Z materials, such as lead, have low X_0 , making them well suited for electromagnetic calorimeters.

The HCal measures the energy of hadronic jets. The barrel region ($\eta < 1.7$), directly outside the ECal, is subdivided into three tile calorimeters that use scintillating tiles as an active medium with steel absorbers. The end-cap regions ($1.5 < \eta < 3.2$) each

form two wheels using liquid argon active material and copper absorbers. The HCal is characterised by the nuclear interaction length λ of its absorbing material, defined as the mean distance travelled by a hadronic particle in the material before it undergoes an inelastic nuclear interaction. Denser materials have lower λ . High-Z materials have a $\lambda > X_0$, therefore hadronic showers will proceed electromagnetic showers.

Calorimetry up to $|\eta| < 4.9$ is supported via additional liquid argon forward calorimeters integrated within the end-cap regions with one copper and two tungsten absorbing layers.

The ATLAS calorimeters illicit a lower response from hadrons than from electrons and photons, making them non-compensating, since a fraction of the energy contained in the hadronic component of a shower does not produce detectable signals [41]. A calibration is performed to correct for these energy losses.

3.2.4 Muon spectrometer

The outermost ATLAS subdetector is the Muon Spectrometer (MS). Due to their comparatively high mass, muons are minimally ionising and produce less Bremstrahlung than electrons [42]. As a result, they generally pass through the calorimeters, leaving only small detectable signals. The MS contains separate tracking and triggering (see §3.2.5) systems for the identification of muons.

The tracking system provides precise track coordinates and comprises monitored drift tubes (MDTs) covering $|\eta| < 2.7$ and cathode strip chambers covering $2.0 < |\eta| < 2.7$. Muon tracks are bent by a large barrel toroidal magnet covering $|\eta| < 1.4$ and smaller magnets in each end-cap covering $1.6 < |\eta| < 2.7$, enabling momentum measurements. Both sets of magnets bend the tracks in the transition region, $1.4 < |\eta| < 1.6$. The MS is designed to measure the p_T of muons with $p_T > 3$ GeV with a resolution 3% for $p_T < 250$ GeV and 10% at p_T of 1 TeV. Triggering and additional coordinate information is provided by resistive plate chambers in the barrel and thin gap chambers in the end-cap within the range $|\eta| < 2.4$.

The MS is the largest of the ATLAS subdetectors, with cylindrical barrel chambers arranged concentrically around the beam axis at $5 < r < 10$ m, and large end-cap wheels

3.2. THE ATLAS DETECTOR

located perpendicular to the z -axis at $7.4 < |z| < 21.5$ m.

3.2.5 The ATLAS trigger system

In the ATLAS detector, bunch crossings occur at a rate of 40 MHz. Significant computing resources would be required to record and store data for each crossing, or *event*, therefore ATLAS employs a trigger system to select only those events with the greatest physics potential. These are events that contain hard-scatter collisions able to produce particles at energy scales required to probe for new physics.

The ATLAS trigger system [43] consists of two levels: the Level 1 (L1) trigger and the High-Level Trigger (HLT).

The hardware-based L1 trigger reduces the event rate from 40 MHz to 100 kHz using reduced granularity information from the calorimeters and MS, within a latency of $2.5\ \mu\text{s}$. Trigger decisions are made based on event-level quantities, multiplicities of objects exceeding thresholds and topological information. The L1 trigger also identifies Regions-of-Interest (RoIs) in the detector around potential high- p_T jet, lepton or photon candidates. Events selected by the L1 trigger are sent to the HLT.

The HLT is software-based and uses dedicated computing farms to run precise, CPU-intensive selection algorithms (largely based on those used for offline reconstruction described in §4). Within RoIs, the HLT uses full detector granularity, including tracking information from the ID, which enables a more detailed reconstruction of the physics objects in an event and therefore more accurate selections than the L1 trigger. The HLT reduces the event rate from 100 kHz to an average of 1.2 kHz, within a latency of 0.2 s. Events selected by the HLT are stored for offline analysis.

Each physics signature has its own dedicated trigger chain. A trigger chain consists of an L1 trigger item which seeds a set of HLT reconstruction and selection algorithms. The full set of trigger chains used for data-taking, along with additional prescale values, form the trigger menu. Prescale values are used to control the rate of acceptance of a given chain and help ensure efficient distribution of L1 and HLT resources within the trigger menu. Applying prescales can limit the sensitivity and increase the complexity of an analysis, therefore triggers are kept unprescaled in BSM searches when possible.

3.2.6 Event simulation

Event simulation is integral to analyses performed at ATLAS. In BSM searches, samples of simulated signal and background processes are used to optimise analysis selections, as well as determine the expected signal efficiency and yield, as is required to perform statistical tests on the real data.

Simulation of events is implemented within the ATLAS simulation framework [44]. Monte Carlo (MC) event generators are used to simulate pp collisions, beginning with the hard scatter process. This is followed by a simulation of the branching of initial and final state partons (quarks and gluons), the so-called *parton shower*, and their subsequent hadronisation, using tools such as PYTHIA [45] and HERWIG [46].

Particles with lifetimes $\mathcal{O}(\gtrsim 1\ \mu\text{s})$ are then propagated through the detector using the GEANT4 framework [47], which records the energy deposited by each particle in the sensitive areas of the detector. These energy deposits, along with those from pile-up simulations, are digitised, such that they emulate the detector readout. This allows the simulated data to be passed through the same trigger and reconstruction algorithms that are used on real data. Records of the history of the particles generated in an event are retained, and is referred to as *truth* information.

Full details of the MC samples used in this thesis, including the specific generators used, is given in §5.2.1 and Appendix A.

EVENT RECONSTRUCTION

After an event has been selected by the online trigger system, offline reconstruction algorithms convert the raw detector-level data recorded by each ATLAS subsystem (described in §3.2) into analysis-level representations of candidate particles, or *physics objects*, produced in collisions. These objects leave different signatures in the various components of the detector. Standard reconstruction algorithms are optimised for reconstructing prompt objects that originate near the interaction point. Dedicated reconstruction algorithms have also been developed to target non-standard objects associated with LLP decays, which are expected to be displaced from the interaction point. The work presented in this thesis uses both standard and non-standard objects, the most relevant of which are described in this chapter.

4.1 TRACK AND VERTEX RECONSTRUCTION

The first step in the reconstruction chain is the reconstruction of tracks in the ID and the pp collision vertices at which they originate.

4.1.1 Standard tracking

Track reconstruction uses sequences of pattern recognition algorithms [48, 49] to identify ID hits produced by the same charged particle. These hits are used to measure the track parameters, defined in Table 4.1, with respect to a reference point. Standard track reconstruction proceeds in two stages: *inside-out* and *outside-in*.

Parameter	Description
d_0	Transverse impact parameter, defined as the distance of closest approach between the track and the reference point in the transverse plane
z_0	Longitudinal impact parameter, defined analogously to d_0
ϕ	Azimuthal angle of the track momentum in the transverse plane
θ	Polar angle of the track momentum in the longitudinal plane
q/p	Ratio of the charge of the track to the magnitude of its momentum

Table 4.1: A description of the five track defining parameters.

Inside-out tracking begins by forming three-dimensional *space-points* from clusters of energy deposits in the pixel and SCT detectors. Track seeds are formed from sets of three space-points located either entirely within the pixel or SCT detectors. A combinatorial Kalman filter [50] is used to build track candidates from seeds by iteratively incorporating additional space-points that are compatible with the seed. An ambiguity solver step aims to remove combinatorial tracks, often referred to as *fakes*, and limit the number of space-points shared between tracks. Each track is given a score that favours high p_T and good fit quality, while penalising missing clusters (*holes*). Clusters shared between multiple tracks are preferentially assigned to tracks with higher scores, then tracks failing to meet a set of quality criteria are rejected. The remaining tracks are extended into the TRT by finding TRT hits compatible with the trajectory of tracks reconstructed within the silicon layers.

After the inside-out sequence is run, the outside-in sequence is performed. Track segments are first identified in the TRT, before being extended into the silicon layers using compatible hits not used in the inside-out stage. This additional pass recovers efficiency lost through ambiguous hits and tracks from secondary interactions (e.g. K_S decays and photon conversions) that have insufficient silicon hits but low enough impact parameters to pass the quality criteria. The final standard track collection comprises tracks resulting from both the inside-out and outside-in strategies.

4.2. PHYSICS OBJECT RECONSTRUCTION

4.1.2 Primary vertices

Following the reconstruction of tracks, primary vertices, points in space where pp collisions have occurred, are located. As discussed in §3.1.1, there may be many primary vertices in a single event. The hard-scatter primary vertex, often referred to as *the* primary vertex, is selected as the vertex with the highest $\sum p_T^2$ of its associated tracks. Other primary vertices, usually from soft-QCD interactions, are called pile-up. Efficiently identifying and assigning tracks to the correct hard-scatter primary vertex is essential in understanding the full kinematic properties of an event.

Primary vertex reconstruction [51] starts by finding vertex seeds using tracks. A track selection is applied to reduce contamination from fakes and tracks from secondary interactions. The vertex position is estimated from the seed and input tracks using an iterative annealing procedure. In each iteration, tracks that are the least compatible with the vertex position are down-weighted, and the vertex position is refit. After the final iteration, the least compatible tracks are removed from the vertex. This process is repeated using tracks yet to be assigned to a vertex. A primary vertex must have at least two tracks.

4.2 PHYSICS OBJECT RECONSTRUCTION

4.2.1 Jets

Quarks and gluons produced in pp collisions cannot exist in isolation due to colour confinement (see §2.1.4). As a result, they undergo hadronisation, forming showers of collimated particles that deposit energy in the calorimeters. Collections of calorimeter energy deposits are reconstructed as *jets*, which represent the hadronic energy content of a given collision. Jets can be produced in the initial pp collision, as well as through hadronic decays of particles including the Z boson, τ leptons and exotic LLPs.

Reconstruction

The first step of jet reconstruction is the formation of three-dimensional, topologically connected clusters of calorimeter cells called *topo-clusters* [52]. Topo-cluster formation

is governed by the cell signal significance $\zeta_{\text{cell}}^{\text{EM}}$, defined as

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}}, \quad (4.2.1)$$

where $E_{\text{cell}}^{\text{EM}}$ is the cell signal and $\sigma_{\text{noise,cell}}^{\text{EM}}$ is the average expected noise of the cell, both of which are calibrated to the electromagnetic (EM) scale. Topo-cluster seeds are formed from cells with high signal significance satisfying $\zeta_{\text{cell}}^{\text{EM}} > 4$. Neighbouring seed cells are merged into a single cluster. Cells neighbouring the seed cells are added to the cluster using the threshold $\zeta_{\text{cell}}^{\text{EM}} > 2$, and this process repeats for the newly added cells until no such neighbours remain. Lastly, cells with $\zeta_{\text{cell}}^{\text{EM}} > 0$ in the perimeter are added. This set of thresholds is referred to as ‘4-2-0’ topo-cluster reconstruction. The resulting topo-cluster is characterised by a core of cells with highly significant signals enveloped by cells with less significant signals.

Jets reconstructed using only calorimeter topo-clusters are called *EMTopo jets*. The angular and energy resolution of reconstructed jets, particular at low p_T , can be improved using the particle flow (PFlow) algorithm [53] to produce *PFlow jets*, that uses tracking information from the ID in addition to calorimeter measurements. However, the PFlow algorithm places strict hit requirements on the tracks entering the jet reconstruction, making it unsuitable for reconstructing jets from LLP decays.

The *anti- k_t* algorithm [54] is used to form jets from topo-cluster four vectors. The anti- k_t algorithm is an infrared and collinear safe sequential recombination algorithm that outputs jets whose shape are not affected by soft radiation. The algorithm defines the distance metrics d_{ij} , between the i^{th} and j^{th} jet constituents (e.g. topo-clusters), and d_{iB} , between the i^{th} constituent and the beamline as:

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2} \quad (4.2.2)$$

$$d_{iB} = k_{ti}^{-2} \quad (4.2.3)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti} is the transverse momentum of the i^{th} constituent, and R is a radius parameter (usually $R = 0.4$).

4.2. PHYSICS OBJECT RECONSTRUCTION

The algorithm proceeds for an input constituent i by computing d_{iB} and d_{ij} for all $j \neq i$. If the smallest distance is a d_{ij} , constituents i and j are recombined, while if it is d_{iB} , constituent i is considered a jet and removed from the set of available constituents. The distances are then recalculated and the procedure repeats until there are no constituents remaining.

Calibration

The energy of reconstructed jets is defined at the EM scale, meaning energy deposits in the calorimeter produced by particles in EM showers are correctly measured. However, this does not account for detector effects such as calorimeter non-compensation (see §3.2.3), energy losses due to inactive material and particles escaping the calorimeter, the effects of pile-up, and inefficiencies in clustering and jet reconstruction. A series of jet energy scale (JES) calibrations are used to restore the jet energy to that of the truth jets [55].

First, an origin correction is applied to the four vectors of the jets such that they originate at the primary vertex.

The expected additional energy due to particles from pile-up interactions is then subtracted from the jet energy using the *pile-up density* and *residual pile-up* corrections, derived from MC simulation.

The absolute JES and η calibration corrects the jet four-vector to the particle-level energy scale. Corrections are derived in MC by comparing kinematic variables in reconstructed and truth jets. The corrections account for varying responses to hadronic showers in different parts of the detector.

After the previous calibrations, residual dependencies of the JES on the flavour of its originating particle and the energy and transverse distribution of its constituents are observed. The global sequential calibration (GSC) procedure reduces these effects. For EMTopo jets, the GSC uses five multiplicative corrections derived from jet observables that remove the dependence of the JES on these observables, while conserving the average JES and improving the jet energy resolution (JER), defined as the standard deviation of a Gaussian fit to the ratio $\frac{p_T^{\text{reco}}}{p_T^{\text{truth}}}$.

The last stages of the JES calibration account for differences in the jet response between MC and data, that arise due to imperfect simulation of the detector material and underlying physics processes.

Precise knowledge of the systematic uncertainties introduced by performing the JES and JER calibrations is important when conducting studies of SM and BSM processes involving particles decaying to jets [56].

Flavour tagging

A vital component of jet studies at ATLAS is flavour tagging. The aim of flavour tagging is to label each jet according to the flavour of the quark that produced it.

For Run 2, the primary algorithm used for flavour tagging is DL1r [57]. DL1r is a neural network based tagger. The inputs to DL1r are the outputs of a number of low-level taggers, that exploit variables such as track multiplicity, hit patterns and impact parameters, and the secondary vertex invariant mass. DL1r has a multidimensional output corresponding to the probabilities of a jet originating from a b -quark, c -quark or light flavour quark. A b -tagging discriminant is derived from these outputs that allows the model to be used at various *working points* (WPs). As an example, a 77% WP would correspond to the discriminant value for which 77% of b -jets have a higher score.

More recently, the GN1 algorithm [58] and its successor GN2 have been developed, which use a graph neural network (GNN) to classify jets. The architecture of the GN2 algorithm forms the basis of the model presented in §6.2 for identifying displaced jets from LLP decays.

4.2.2 Electrons

Electrons¹ are charged particles that leave tracks in the ID and deposit most of their energy in the ECal. The search presented in §5 uses electrons to identify $Z \rightarrow ee$ and $W \rightarrow e\nu$ decays. A detailed overview of electron reconstruction can be found in Refs. [59] and [60].

¹Throughout this thesis, *electrons* is usually used to refer to both electrons and positrons.

4.2. PHYSICS OBJECT RECONSTRUCTION

Reconstruction

An electron is defined as an object consisting of a cluster built from energy deposits in the ECal and a matched track (or tracks).

Electron reconstruction begins with the formation of topo-clusters from deposits in ECal cells², using the ‘4-2-0’ set of thresholds as described in §4.2.1. The topo-clusters are matched to ID tracks. These tracks are refitted to account for energy loss due to bremsstrahlung [61]. An electron can lose energy via bremsstrahlung due to interactions with the material layers in the detector. The radiated photon may convert into an e^+e^- pair. These electrons, positrons and photons are usually highly collimated and are often reconstructed in the same cluster. It is therefore possible for a single energy cluster to be matched to multiple tracks, which are ranked according to the ΔR separation between the track and the cluster, with a preference for tracks with hits in the pixel detector.

Topo-clusters are then grouped into variable-sized clusters called *superclusters*. Seed clusters are identified as having a minimum E_T of 1 GeV and a matched track with at least four silicon hits. Satellite clusters that lie within a $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ window about the barycentre of the seed are identified. Additionally, for electrons, a cluster is also considered a satellite if it lies within a $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ window of the seed, and the two clusters share a highest ranking track. The seed and its associated satellite clusters form a supercluster. A supercluster and its matched tracks constitute an electron candidate.

Identification

A likelihood-based identification is used to select prompt electrons (produced in the decay of a short-lived particle), with $|\eta| < 2.47$, while rejecting background. The background consists of jets that mimic the signature of prompt electrons, electrons from photon conversions and non-prompt electrons from heavy flavour hadron decays. The inputs to the algorithms include measurements from the tracking and calorimeter systems. For a full description of the algorithm, including descriptions of the inputs, see Ref. [59].

²Except in the transition region $1.37 < |\eta| < 1.63$, where energy from the presampler between the HCal and ECal is included.

For each electron candidate, a discriminant is computed. Values of this discriminant are used to define four WPs: *Very Loose*, *Loose*, *Medium* and *Tight*, ordered by increasing background rejection. The efficiencies to identify a prompt electron with $E_T = 40$ GeV are 93%, 88% and 80% for the *Loose*, *Medium* and *Tight* WPs, respectively.

Isolation

Background rejection can be enhanced by imposing isolation requirements on candidate electrons. A characteristic signature of a prompt electron is reduced activity in an area of $\Delta\eta \times \Delta\phi$ surrounding the candidate object. Calorimeter- and track-based isolation variables are derived by summing the transverse energy of topo-clusters within a cone around an electron candidate, or by summing the p_T of tracks within a cone aligned with an electron track. These variables are then used to define electron isolation working points, whose specific implementations are dependent on the analysis being performed.

Calibration

After reconstruction, the electron energy measurement is calibrated [62, 63]. The calibration can be summarised in three steps.

First, uniformity corrections are applied to data such that the response of the longitudinal layers of the ECal are in agreement between data and simulation. Then, corrections to the electron energy that account for the energy lost in the material preceding the calorimeter, in the cells neighbouring the reconstructed cluster, and beyond the calorimeter, are computed using a multivariate regression algorithm and applied to both data and simulation. Lastly, corrections are then made to account for residual differences of the energy scale and resolution between data and simulation. Energy scale corrections are computed using the $Z \rightarrow e^+e^-$ resonance.

4.2.3 *Photons*

Photons, like electrons, deposit most of their energy in the ECal through electromagnetic showers. Photon reconstruction therefore proceeds in a similar fashion to electrons [59, 62, 63, 64]. However, an additional complexity arises due to photon conversions. A photon may convert into an e^+e^- pair before reaching the ECal (a *converted* photon), or

4.2. PHYSICS OBJECT RECONSTRUCTION

not be associated with a conversion at all (an *unconverted* photon). Accurate photon reconstruction relies on identifying both converted and unconverted prompt photons from hadronic collisions, and differentiating them from non-prompt photons produced by hadronic decays in jets and fake candidates associated with hadronic energy deposits in the ECal.

First, tracks that are loosely matched to ECal topo-clusters are input to the conversion vertex reconstruction. A two-track conversion vertex is reconstructed from a pair of collinear tracks. A single-track conversion vertex is built from tracks without innermost pixel detector hits. Tracks used to build conversion vertices must be compatible with an electron traversing the TRT. A cluster that has been matched to a conversion vertex is considered a converted photon, while an unconverted photon is a cluster without a matching track or conversion vertex.

Photon superclusters are formed in a similar manner to electrons. Seed clusters must have a minimum E_T of 1.5 GeV, with no track or conversion vertex matching requirement. For photons, in addition to clusters that lie within a $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ window about the seed, a cluster is added as a satellite if its best matched track belongs to a conversion vertex that has been matched to the seed cluster. A given seed cluster can produce an electron or a photon candidate, in such cases, an ambiguity resolution procedure is performed. The photon energy scale is calibrated in a similar manner to electrons, described in §4.2.2.

Photon identification WPs are defined using a cut-based selection based on calorimetric *shower-shape variables*, designed to give good separation between prompt and non-prompt photons. A full description of these variables can be found in [59]. The two primary reference selections are *Loose* and *Tight*.

4.2.4 Muons

Muons are minimally ionising charged particles, identifiable by a track in the MS or characteristic energy deposits in the calorimeters. The search presented in §5 uses muons to identify $Z \rightarrow \mu\mu$ and $W \rightarrow \mu\nu$ decays. A detailed overview of muon reconstruction can be found in Refs. [65] and [66].

Reconstruction

Muon reconstruction starts with a search for localised hit segments in each muon chamber. Segments from at least two different chambers are combined to form muon track candidates in the MS. A global reconstruction is then performed, using information from the ID, MS and calorimeters. Four muon *types* are defined depending on the subdetectors being used in the reconstruction, one of which, combined (CB) muons, is relevant to this thesis.

CB muons are obtained by performing a global track fit, that accounts for energy loss in the calorimeters, using both ID and MS tracks. Most muons are reconstructed following the *outside-in* regime, where MS tracks are extrapolated inward and matched to ID tracks, although a complimentary *inside-out* approach is also performed.

Identification

Muon candidates are identified using a set of quality requirements designed to suppress background from light hadron decays, while efficiently selecting prompt muons and maintaining precise momentum measurements. The variables used to identify CB muons are:

- the q/p *significance*, defined as the absolute value of the difference between q/p (defined in Table 4.1) measured in the ID and MS divided by the sum in quadrature of the corresponding uncertainties.
- ρ' , defined as the absolute value of the difference in p_T as measured in ID and MS divided by the p_T of the combined track.
- the normalised χ^2 of the combined track fit.

Additional requirements on the number of hits in the ID and MS are used.

Three standard WPs are defined, which are, in order of increasing purity and decreasing efficiency, *Loose*, *Medium* and *Tight*. The *Medium* criteria provides the default selection for most analyses at ATLAS, and is used in the work presented here. It requires CB tracks with ≥ 3 hits in at least two MDT layers, except for $|\eta| < 0.1$, where

4.2. PHYSICS OBJECT RECONSTRUCTION

tracks with hits in at least one MDT layer but no more than one MDT hole³ are allowed. Isolation WPs for muons are defined in a similar fashion to electrons.

Calibration

A calibration of the momentum scale of the muon candidates is then performed [67]. First, corrections are made in data that account for charge-dependent biases induced by detector misalignment in the ID and MS, using the $Z \rightarrow \mu^+\mu^-$ resonance. Then, corrections to the reconstructed muon p_T in simulation are identified such that better agreement between simulation and data is observed. Derived in $\eta - \phi$ regions of the detector homogeneous in terms of technology and performance, these corrections compensate for, among other effects, the mismodelling of magnetic fields in simulation.

4.2.5 *Missing transverse momentum*

Due to conservation of momentum, the vector sum of the transverse momentum in an event should be zero. Experimentally, an imbalance is observed. This missing transverse momentum E_T^{miss} usually serves as a proxy for the p_T of SM particles that escape the detector (such as neutrinos), but can also arise due to detector acceptance effects or reconstruction inefficiencies. Exotic LLPs that do not interact with and decay outside of the detector will also contribute to the E_T^{miss} .

The missing transverse momentum is first calculated in the $x(y)$ plane using [68]

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss},\text{jets}} + E_{x(y)}^{\text{miss},\mu} + E_{x(y)}^{\text{miss},\text{soft}}, \quad (4.2.4)$$

where $E_{x(y)}^{\text{miss},i}$ is the negative vector sum of the momentum of all objects of type i , and $E_{x(y)}^{\text{miss},\text{soft}}$ is a *soft* term, consisting of the negative vector sum of the momentum of ID tracks that are associated with the hard-scatter vertex but not with any hard charged particle or jet.

The total E_T^{miss} is then

³A hole is a missing hit, where one is expected, in a detecting layer between two other hits on the track.

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}. \quad (4.2.5)$$

4.3 DISPLACED TRACK AND VERTEX RECONSTRUCTION

Standard track and vertex reconstruction algorithms, described in §4.1, are optimised for particles that originate near the primary interaction point. They place stringent requirements on the transverse and longitudinal impact parameters of track candidates to ensure particles from primary interactions are efficiently reconstructed with high purity and within manageable computation time. But for BSM searches that aim to detect the decay products of LLPs with decay vertices displaced by more than a few mm from the interaction point, such algorithms are insufficient. Two dedicated algorithms are used in this thesis to reconstruct the decays of LLPs in the ID. First is the *large radius tracking* algorithm, described in §4.3.1, an additional tracking pass that uses loosened impact parameter and hit selections to improve tracking efficiency at larger radii. Then a custom secondary vertexing algorithm is used to identify the decay positions of LLPs, using as inputs both standard and large-radius tracks. This is described in §4.3.2.

4.3.1 Large radius tracking

The large radius tracking (LRT) algorithm [69, 70] is a tertiary tracking pass that follows and is complimentary to the standard inside-out and outside-in tracking sequences. The reconstruction follows the same procedure outlined in §4.1.1 for inside-out tracking but is optimised for LLP signatures. For Run 2, LRT was exclusively optimised for high signal efficiency, leading to a per event computation time twice that of the standard pass, owing to the large number of fake tracks from random combinations of hits being reconstructed. This was manageable as LRT was only applied to an $\mathcal{O}(10\%)$ fraction of recorded events. In preparation for Run 3, where a larger number of pp collisions per bunch crossing is expected, the LRT algorithm was improved, leading to lower processing times and a significantly reduced number of fake tracks. Consequently, LRT reconstruction for LLP signatures was fully integrated into the standard reconstruction chain. The updated reconstruction chain was used to process the the data recorded during Run

Selection Criteria	Standard	Large radius
max. $ d_0 $ [mm]	5	300
max $ z_0 $ [mm]	200	500
max. $ \eta $	2.7	3.0
min. p_T [GeV]	0.5	1
max. silicon holes	2	1
max. double holes	1	0
seeding	Pixel and SCT	SCT only
Common Selecton Criteria		
min. silicon hits	8	8
min. unshared silicon hits	6	5

Table 4.2: A comparison of the selection criteria used in standard and large radius tracking [70]. A hole is a missing hit, where one is expected, in a detecting layer between two other hits on the track. Shared hits are hits used on multiple tracks.

2 that is used in §5.

The key differences and commonalities between standard and LRT reconstruction configurations are listed in Ref. [70], a subset of which are reproduced in Table 4.2. LRT track candidates are seeded using only SCT space points (with a maximum of one seed per middle SP allowed), since displaced tracks are less likely to have three pixel hits, and are created using silicon hits that were not used in the two standard tracking passes, ensuring independence between the standard and LRT track collections. Both the primary and LRT tracking passes use $|d_0|$ to order the relevance of seeds for the track finding stage (the seed with the lowest $|d_0|$ being ranked the highest). In addition, the LRT seed ordering includes a $\Delta\eta = |\eta_{\text{seed}} - \eta'|$ weight factor, where η' is the pseudorapidity of the vector connecting the beam spot with the point of closest approach to the seed. Seeds from LLP decays tend towards low values of $\Delta\eta$, due to the correlated direction between the LLP and its decay products, in contrast to fake tracks. Requirements on the impact parameters and number of hits are relaxed to reflect the extended range of production radii for charged particles produced by LLP decays. The p_T threshold is increased to reduce the number of soft charged pion tracks caused by material interactions. These optimisations significantly improve the execution time by reducing the number of combinatorial permutations to process during the track seeding and finding stages.

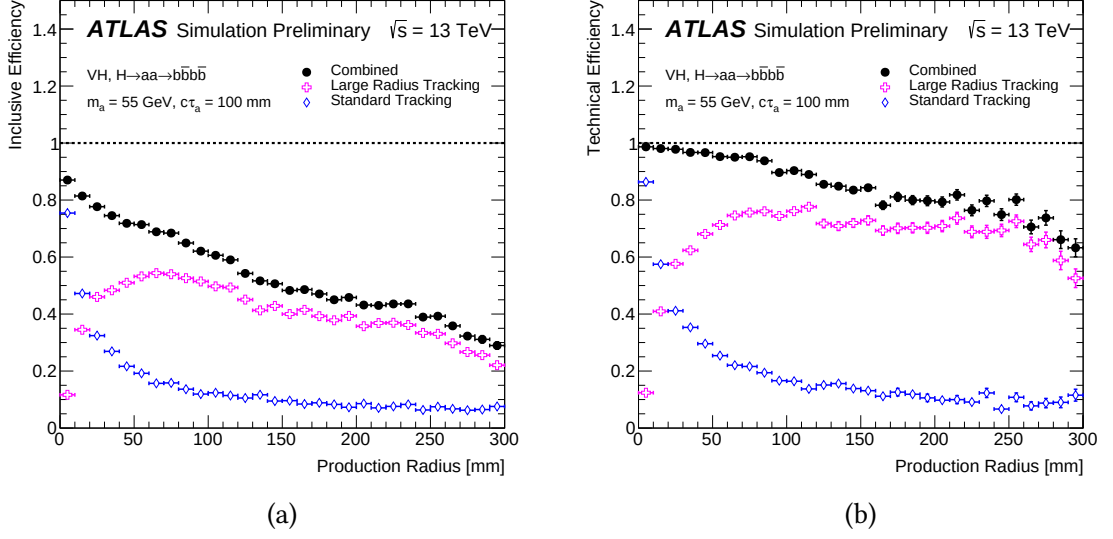


Figure 4.1: Inclusive (a) and technical (b) track reconstruction efficiency for displaced charged particles produced via decays of a pseudoscalar LLP, a , as a function of the production radius of the displaced particles [71].

Performance

To quantify the effect of LRT on displaced particle reconstruction, the efficiency for reconstructing charged particles from BSM LLP decays is determined. This efficiency is defined as the ratio of the number of signal truth particles matched⁴ to reconstructed tracks divided by the number of signal truth particles, where truth particles must satisfy the baseline fiducial requirements outlined in Table 4.3. A simulated data set of a benchmark model featuring a long-lived neutral pseudoscalar boson a , produced in pairs via a Higgs boson decay, that decays to pairs of b -quarks ($H \rightarrow aa \rightarrow b\bar{b}b\bar{b}$), is used. Figure 4.1a shows the inclusive track reconstruction efficiency for charged particles produced in a boson decays as a function of the production radius of the charged particle r_{prod} for standard and LRT tracks, as well for a combination of both collections.

i

At $r_{\text{prod}} < 5$ mm, the standard tracking is highly efficient, but this efficiency drops off sharply as r_{prod} increases. LRT is able to recover some of this efficiency up to higher r_{prod} . The approximately linear drop off in reconstruction efficiency with r_{prod} can be attributed to both algorithmic effects and the fact that charged particles with larger r_{prod} are less likely to produce the minimum number of silicon hits required by the track recon-

⁴The matching scheme used is described in Ref. [70].

Baseline selection
$p_T > 1 \text{ GeV}$
$ \eta < 2.5$
$r_{\text{prod}} < 440 \text{ mm}$
charge = ± 1
Technical efficiency selection
$r_{\text{prod}} < 300 \text{ mm}$
Number of silicon hits ≥ 8

Table 4.3: Fiducial selections applied to truth particles entering the tracking efficiency calculation.

struction. To isolate the algorithmic effect on the efficiency, the *technical reconstruction efficiency* is used. The technical reconstruction efficiency only considers truth particles that could theoretically be reconstructed as tracks within the ID. To achieve this, truth particles used in the efficiency calculation are subject to an additional selection, also shown in Table 4.3, that mimics those used in the track reconstruction. The technical reconstruction efficiency is shown in Figure 4.1b. For the combined track collection, it is observed to be over 60% up to $r_{\text{prod}} = 300 \text{ mm}$.

To access the impact of the improved LRT configuration on searches for displaced vertices, such as the one presented in §5, the rate of vertices reconstructed from fake tracks, which form the dominant background in such searches, is studied. Figure 4.2 shows the transverse displacement L_{xy} of vertices matched⁵ and not matched to truth LLP decays for both the Run 2 and Run 3 LRT configurations. Vertices that are not matched to LLP decays are considered to be fakes. Secondary vertices are reconstructed using the algorithm described in §4.3.2. The improved LRT leads to a small improvement in the number of LLP vertices but substantially fewer fake vertices, owing to the increased purity of the track reconstruction.

4.3.2 Displaced vertices

Once the LRT algorithm, described in §4.3.1, has been used to reconstruct charged particles originating from an LLP decay, the kinematic properties of the LLP can be further

⁵Vertices are considered matched to an LLP decay if the p_T -weighted fraction of tracks in the vertex that are matched to LLP descendants is greater than 0.5.

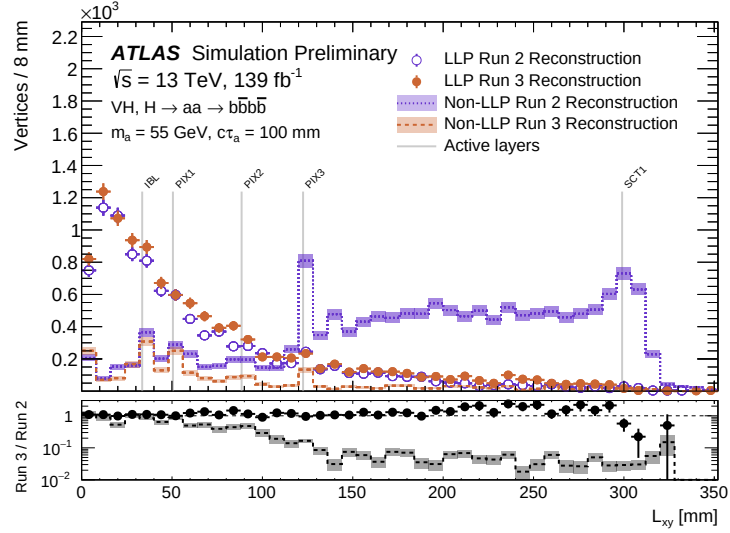


Figure 4.2: A comparison of the radial distributions of reconstructed secondary vertices in an LLP sample using the Run 2 and Run 3 LRT configurations. Circle markers denote vertices that have been matched to truth LLP decays (LLP). Dashed lines denote vertices that have not been matched to truth LLP decays (non-LLP). These are considered fake vertices. Vertical lines show the positions of pixel and SCT layers. The lower panel shows the ratio of the number of vertices reconstructed using the Run 3 configuration to the Run 2 configuration for LLP (circle markers) and non-LLP (dashed lines) vertices [71].

understood by reconstructing a secondary vertex using the resulting tracks. This vertex may be significantly displaced from the primary pp collision and provides an effective signature with which to discriminate signal from background.

The displaced vertices in this thesis are reconstructed using the *Vertex-Secondary-Inclusive* (VSI) algorithm [72], which has a reconstruction efficiency of up to $\sim 60\%$ for the process $H \rightarrow aa \rightarrow b\bar{b}b\bar{b}$. An overview of the algorithm is given here.

First, tracks are preselected for the formation of seed vertices. Tracks from both the standard and LRT track collections are used, so long as they are not associated to a primary vertex. Tracks must have a $p_T > 1$ GeV and pass a set of hit-based requirements designed to reject low-quality tracks.

Vertex finding starts by finding all pairs of preselected tracks with $|d_0| > 2$ mm. A fast calculation of the vertex position is performed using the track parameters. If the track impact parameters are small relative to this initial position, an accurate track extrapolation is performed and the vertex candidate is rejected if the fitting procedure does not produce a valid solution. Tracks in the remaining vertex candidates must fulfil the *hit-pattern consistency*, that requires them to only have hits in tracker layers at radii

4.3. DISPLACED TRACK AND VERTEX RECONSTRUCTION

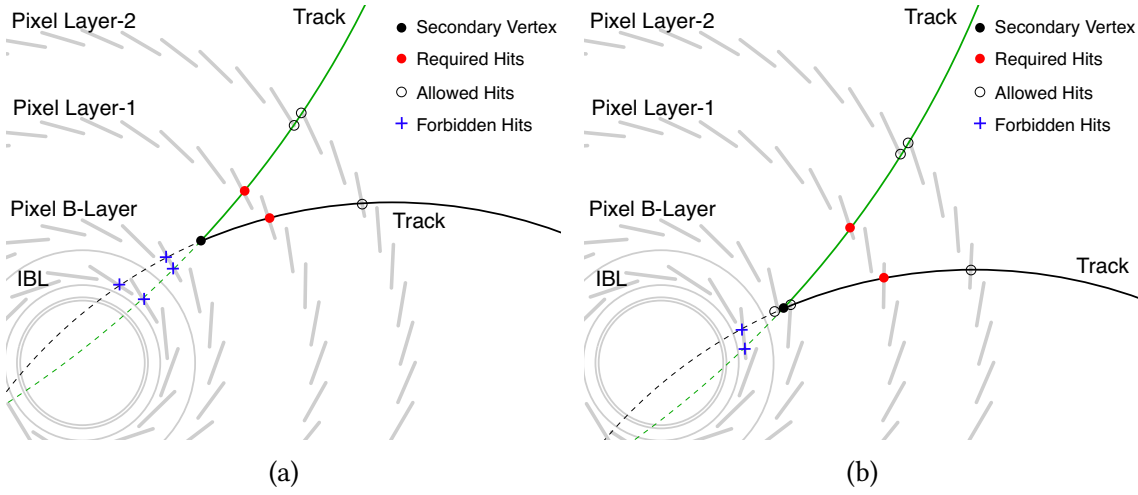


Figure 4.3: Schematic showing a transverse view of the pixel detector in the barrel region, along with the required, allowed and forbidden pixel layer hits of tracks forming a secondary vertex in between the B-layer and Layer 1 sensors (a) and close to the B-layer (b). The tracks must have hits on the closest layer outside the vertex, and must not have hits on any layers within the vertex, unless the vertex is close to detecting layer [72].

beyond the vertex position. This is not enforced when the vertex position is close to a tracker layer, to account for vertex position mis-measurement. An illustration of this is shown in Figure 4.3. After the two-track vertex finding step, multi-track vertices are constructed by finding groups of tracks that are fully compatible with each other.

Multiple secondary vertices may be reconstructed from a single LLP decay, in a process termed *splitting*. This is undesirable as it deteriorates the vertex track multiplicity and invariant mass, both of which are useful discriminants when identifying a BSM LLP signal. Splitting is most likely for vertices with high track multiplicity or if the decay of the LLP contains tertiary vertices, such as for $a \rightarrow b\bar{b} \rightarrow XX$. To mitigate splitting, nearby vertices can be merged. Candidates for merging are chosen in ascending order of track multiplicity, and attempts are made to merge them with vertices in decreasing order of track multiplicity. Two vertices are merged if they pass a series of compatibility tests. Merged vertices are refit using the tracks from both input vertices.

At this point, tracks that originated from the secondary vertex but did not pass the seed track preselection are not included in the vertex fit. Including them can increase the vertex track multiplicity and improve the invariant mass measurement. A selection of unused tracks, including those associated to a primary vertex, are considered for vertex association if they have low impact parameter significances (defined as the ratio of the impact parameter to its uncertainty) with respect to the target vertex. The track is

associated to the vertex if a refit of the vertex including the track passes a loose quality requirement, designed to accept as many tracks as possible for each vertex⁶. Each track can be associated to at most one secondary vertex. Once the track attachment procedure is complete, a final refit is performed and the track parameters with respect to the vertex are recomputed.

⁶Tracks may be subject to analysis specific selections later on.

5

SEARCH FOR EXOTIC HIGGS DECAYS TO LONG-LIVED
PARTICLES

This chapter describes a search for a new, light, long-lived scalar particle s that decays hadronically in the ID of the ATLAS detector, producing displaced jets and displaced hadronic vertices. An overview of the analysis is given in §5.1, including details of related previous searches and the analysis strategy. The data and simulated samples used to conduct the analysis are detailed in §5.2. The object-level and event-level selections used to identify signal-like events are described in §5.3 and §5.4, respectively, followed by an overview of the data-driven procedure used to estimate the residual background once these selections have been applied, in §5.5. The sources of systematic uncertainty on the predicted signal and background yields are described in §5.6. The statistical framework used to interpret the results are outlined in §5.7, before the results themselves are presented in §5.8. Conclusions are drawn in §5.9.

5.1 ANALYSIS OVERVIEW

5.1.1 *Previous searches*

Several searches for Higgs boson decays to LLPs have been performed at ATLAS using data from both Run 1 and Run 2. A summary of the exclusion limits on $BR(H \rightarrow ss)$ set by these searches is shown in Figure 5.1. Stringent limits have been set for LLPs within the proper lifetime range 10^{-1} – 10^2 m [73, 74, 75, 76]. These analyses used dedicated

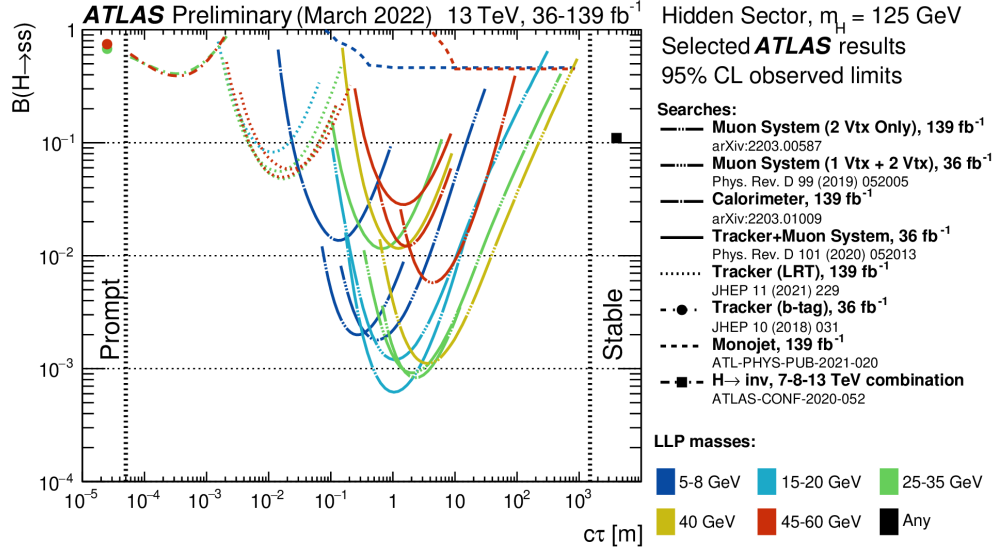


Figure 5.1: Regions in $BR(H \rightarrow ss)$ (where s is an LLP) versus $c\tau$ excluded at 95% confidence level by the ATLAS analyses listed in the legend. The colour of the lines refer to the mass of the long-lived scalar being considered [78].

triggers to select events with LLPs decaying in the calorimeters and MS. The absence of a comparable trigger to identify hadronic decays of LLPs in the ID has hindered sensitivity in the region with LLP lifetimes below 10^{-1} m, where branching ratios $BR(H \rightarrow ss)$ of $\mathcal{O}(1\%)$ are yet to be excluded. Such searches must instead exploit the associated production modes of the Higgs to select events. The analysis presented in this thesis is an extension to a previous analysis [77] (labelled *Tracker (LRT)* in Figure 5.1), that considered only the ZH production mode. This search was hampered by large backgrounds due to the high rate of fake tracks produced by the implementation of LRT available at the time (see §4.3.1). Consequently, only events where both LLPs were reconstructed as displaced vertices (DVs) in the ID, i.e. events with DV multiplicity $n_{DV} \geq 2$, were considered, limiting sensitivity for proper lifetimes below 10^{-3} m and above 10^{-1} m.

5.1.2 Analysis strategy

The search described in this chapter targets hadronic decays of LLPs in the ATLAS ID. The benchmark model is detailed further in §2.3.2 and the main signal process that is studied is shown in Figure 2.9. The objective of the analysis is to extend the $BR(H \rightarrow ss)$ exclusion limit to the $\mathcal{O}(1\%)$ level in the proper lifetime range 10^{-3} – 10^{-1} m.

The analysis considers events with a Higgs boson produced through the ZH , WH

and VBF production modes. Events consistent with a Higgs produced in association with a W or Z boson are identified using prompt leptons from the gauge boson decay, while events containing two forward jets with a large angular separation are characteristic of the VBF production mode. Events are separated into three search regions according to each Higgs production mode using selections designed to effectively remove SM background. An additional region containing photons with low signal contamination is used as a control region.

An LLP that travels a considerable distance before decaying hadronically in the ID will produce two distinct signatures. First, jets resulting from the decay will be displaced from the IP, and have few prompt-tracks and few tracks matched to a primary vertex, when compared to a jet from a promptly decaying particle. Second, the position of the decay itself will be reconstructed as a secondary DV. These features are used to differentiate between signal and SM background.

A per-jet boosted decision tree (BDT) is used to select jets with a topology consistent with an LLP decay. The BDT scores of the two most signal-like jets are used to compute an event level discriminant that separates the search regions into control, validation and signal regions. A data-driven background estimate is derived using the control region of each search region.

The signal regions for each search region are binned according to n_{DV} . Due to improvements in LRT (see §4.3.1) and incorporation of machine learning techniques in the event selection, this iteration of the analysis has significantly reduced background when compared to the previous iteration. As a result, the analysis includes events with only one reconstructed LLP decay within the ID, such that $n_{DV} = 1$, into the signal region, improving sensitivity to a wider range of LLP lifetimes. Each search region is separated into two signal regions, with $n_{DV} = 1$ or $n_{DV} \geq 2$, giving six signal regions in total. The signal regions are kept blinded during the development of the analysis. A maximum-likelihood fit is performed in the signal regions to determine the expected and, as no significant excess of events over the expected background yield is observed, observed exclusion limits.

In addition to the benchmark exotic Higgs decay process, the results are interpreted

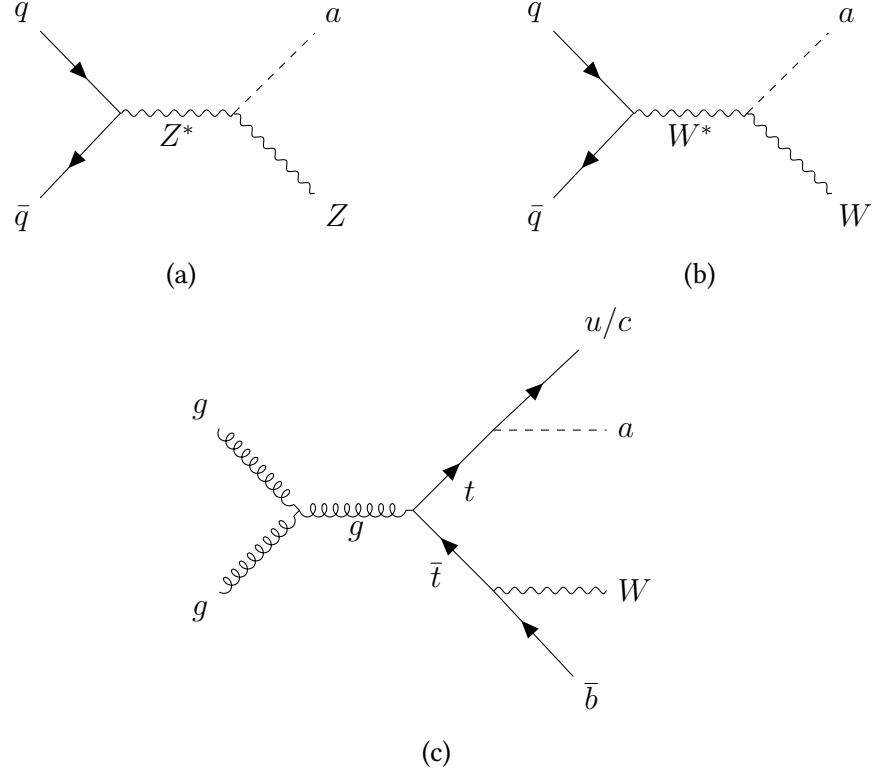


Figure 5.2: Feynman diagrams depicting the Za (a), Wa (b), and $t \rightarrow qa$ ALP (c) production modes.

in terms of models containing Axion-like particles (ALPs). ALPs are new pseudoscalar particles, associated to the breaking of the global $U(1)$ symmetry, that offer a potential solution to the strong CP problem [79], while also being predicted by a number of BSM theories (e.g. [80, 81]). ALPs are generally weakly coupled to the SM, hence they are naturally long-lived. Two benchmark ALP models are considered in this analysis. The first extends the SM with an ALP a that couples to gluons and vector bosons [82], with suppressed couplings to photons. This gives rise to the production of a in association with a vector boson (W/Z) and subsequent decay into gluons, producing a displaced jet topology similar to the exotic Higgs decay. The second allows couplings between a and up-type quarks [83], leading to production of a through an exotic top decay $t \rightarrow ac/au$ in $t\bar{t}$ events. In the exotic top decay, the dominant ALP decay mode is $a \rightarrow c\bar{c}$. Diagrams of these processes are shown in Figure 5.2.

5.2 DATA AND SIMULATED SAMPLES

Both real and simulated pp collision data is used to conduct the analysis. The real data, referred to from now as *data*, was collected by the ATLAS experiment between 2015 and 2018 during Run 2, when the LHC was operating at a centre of mass energy of 13 TeV. The data, collected during stable run conditions when all ATLAS sub-systems were operational, has a total integrated luminosity of 140 fb^{-1} . The Monte-Carlo simulated data samples, referred to as *MC*, are described in §5.2.1. After reconstruction, data and MC is formatted as Analysis Object Data (AOD). The ATLAS derivations framework converts AODs into derived AODs (DAODs), which contain less information than DAODs [84]. The information that is removed depends on the needs of the analysis; this analysis uses the *LLP1* derivation configuration. The DAODs are processed using the custom analysis framework.

Events in both data and MC are required to have at least one primary vertex (defined in 4.1.2) with two tracks, and to pass a trigger selection. As described in §3.2.5, the ATLAS trigger is used to select interesting events for further processing. The trigger strategy employed in the analysis is described in §5.2.2

5.2.1 Simulated samples

The analysis uses both simulated signal and background samples. The full list of datasets used is given in Appendix §A. For further details on event simulation procedure at ATLAS see §3.2.6.

Signal events with a Higgs boson produced in association with a Z or W boson or through VBF are generated with POWHEG-Box v2 [85, 86] using the NNPDF3 parton distribution function set [87]. The decay of the Higgs bosons to pairs of scalars s with masses $m_s = 5, 16, 40, 55 \text{ GeV}$ and the subsequent decays of each scalar to a pair of quarks is performed using PYTHIA 8.212. Each mass point is generated with proper lifetimes $c\tau = 1, 10, 100, 1000 \text{ mm}$. The scalars decay to pairs of quarks with $BR(s \rightarrow q\bar{q}) = 1$, where $q = u, c, b$. The ensuing showers and hadronisation is simulated using PYTHIA8 using the AZNLO tune [88].

Additional signal samples are needed to perform the interpretation in terms of ALP

models. These samples are not used to optimise selections or evaluate systematic uncertainties. Events containing ALPs produced in association with a Z or W boson are generated using MADGRAPH 2.9.9 [89] and PYTHIA 8.307 with the ALP_linear_UFO_WIDTH model. Here, the branching ratio of the ALP to gluons is set to 100%. ALP production through the exotic top decay $t \rightarrow qa$ is simulated in MADGRAPH 2.9.9 and PYTHIA 8.307 using the Charming_ALP_UFO model. Here, the ALP is allowed to decay to both gluons and charm quarks. ALP masses below 40 GeV are not considered by the analysis. While the analysis strategy is to focus on events with two displaced jets, the decay products of lower mass ALPs tend to merge into a single jet.

Background MC is needed to optimise selections and aid in the development of the background estimation procedure. The analysis uses samples of $t\bar{t}$, $V + \text{jets}$, $\gamma + \text{jets}$ and dijet events. Further details on the production of background MC can be found in the analysis paper [1].

Pile-up conditions during Run 2 at the LHC varied for data collected in 2015/16, 2017, and 2018. The MC samples are therefore generated three times according to each of the pile-up profiles. The three MC campaigns are labelled MC20a for 2015/16, MC20d for 2017 and MC20e for 2018.

5.2.2 Trigger strategy

Lepton triggers

A vector boson produced in association with a Higgs can decay leptonically, either through $Z \rightarrow l^+l^-$ or $W \rightarrow l\nu_l$. Therefore, to select events consistent with a ZH or WH topology, unrescaled single- and di-lepton triggers are used. These are designed to identify events containing isolated electrons or muons above a certain p_T threshold. The lepton triggers used, and the years in which they were active, are shown in Table 5.1. If an event has passed at least one of the triggers listed in Table 5.1, trigger matching is performed. Trigger matching is the process by which objects reconstructed offline are matched to those reconstructed in the trigger. It helps to remove events that have passed a lepton trigger, but whose offline reconstructed leptons did not fire the trigger. This reduces the effect by which cross section dependent calculations performed later become

5.2. DATA AND SIMULATED SAMPLES

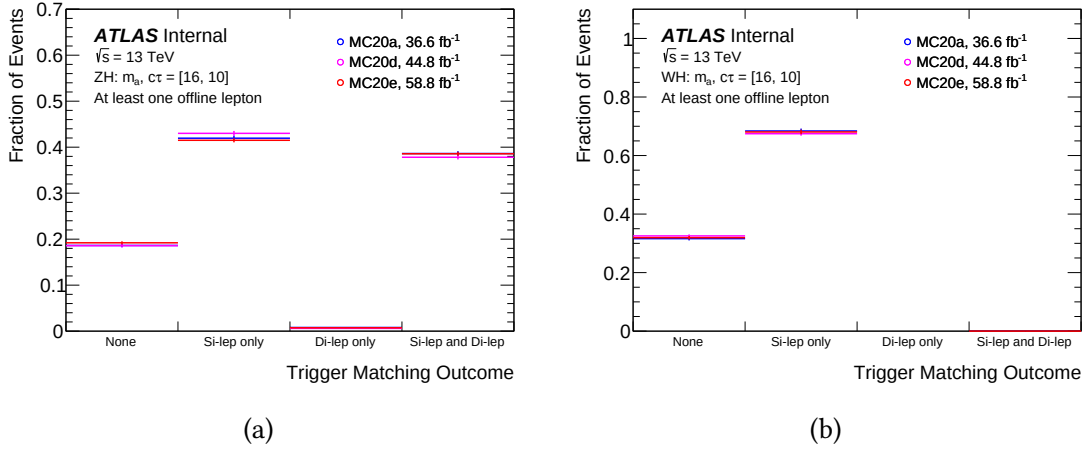


Figure 5.3: Lepton trigger matching outcome for events in ZH (a) and WH (b) signal MC samples. Si-lep and Di-lep refer to single- and di-lepton triggers, respectively.

dependent on physical processes not considered by the analysis. A lepton reconstructed offline is matched to one reconstructed in the trigger if the two lie within a ΔR threshold (0.07 for electrons, 0.1 for muons). For an event to be matched to a di-lepton trigger, both its offline leptons must be matched to distinct trigger leptons. An event passes the lepton trigger preselection if it is matched to at least one lepton trigger that was active during the period¹ in which the event occurred.

Figures 5.3a and 5.3b show a breakdown of lepton trigger matching results for events with at least one offline lepton passing the analysis level criteria in Tables 5.4 or 5.5 (including the looser p_T cut) using ZH ($m_s, c\tau = [16, 10]$) and WH ($m_s, c\tau = [16, 10]$) MC samples. Events are exclusively categorised by whether they matched: none of the triggers, single-lepton triggers only, di-lepton triggers only or both single- and di-lepton triggers. Approximately 81.4%, 81.4% and 80.8% of events were matched to at least one single- or di-lepton trigger and passed the trigger pre-selection for the ZH MC20a, MC20d and MC20e samples, respectively. The inclusion of di-lepton triggers resulted in an overall increase in accepted events of 0.8%, 0.6%, 0.7% for the ZH MC20a, MC20d and MC20e samples, respectively. Approximately 68.4%, 67.5% and 68.1% of events were matched to at least one single- or di-lepton trigger and passed the trigger pre-selection for the WH MC20a, MC20d and MC20e samples, respectively. Additional lepton trigger efficiency studies can be found in Appendix B.

¹A period is a collection of events taken with similar run conditions.

	Trigger	2015	2016	2017	2018
e	HLT_e24_lhmedium_L1EM20VH	✓			
	HLT_e24_lhtight_nod0_ivarloose		✓		
	HLT_e26_lhtight_nod0_ivarloose		✓	✓	✓
	HLT_e60_lhmedium	✓			
	HLT_e60_lhmedium_nod0		✓	✓	✓
	HLT_e120_lhloose	✓			
	HLT_e140_lhloose_nod0		✓	✓	✓
μ	HLT_mu20_iloose_L1MU15	✓			
	HLT_mu26_ivarmedium		✓	✓	✓
	HLT_mu50	✓	✓	✓	✓
ee	HLT_2e12_lhvloose_L12EM10VH	✓			
	HLT_2e15_lhvloose_nod0_L12EM13VH		✓		
	HLT_2e17_lhvloose_nod0		✓		
	HLT_2e17_lhvloose_nod0_L12EM15VHI			✓	✓
	HLT_2e24_lhvloose_nod0			✓	✓
$\mu\mu$	HLT_2mu10	✓			
	HLT_2mu14		✓	✓	✓
	HLT_mu18_mu8noL1	✓			
	HLT_mu22_mu8noL1		✓	✓	✓

Table 5.1: Single lepton (e or μ) and di-lepton (ee or $\mu\mu$) triggers forming the lepton trigger preselection and the years for which they were active.

Trigger	2018
HLT_j70_j50_0eta490_invm1100j70_dphi20_deta40_L1MJJ-500-NFF	✓

Table 5.2: The VBF inclusive trigger used in the analysis. The trigger was only active during 2018.

VBF triggers

To select events consistent with a VBF topology, an unprescaled VBF inclusive trigger is used. The trigger, listed in Table 5.2, uses jet level selections to capture events with two high- η (or *forward*) jets, which are characteristically produced in association with VBF, as shown in Figure 2.4b. The trigger was only active during 2018, therefore the integrated luminosity available to the VBF search region is only 37.5 fb^{-1} . Trigger matching is not required for an event to pass the VBF trigger preselection.

5.3. OBJECT DEFINITIONS

Trigger	2015	2016	2017	2018
HLT_g120_loose	✓			
HLT_g200_etcut	✓			
HLT_g140_loose		✓	✓	✓
HLT_g300_etcut		✓	✓	✓

Table 5.3: Photon triggers and the years for which they were active.

Photon triggers

A photon region is defined to facilitate with the validation of the background estimate. Events in this region must pass at least one of the triggers listed in Table 5.3.

5.3 OBJECT DEFINITIONS

Detailed descriptions of the algorithms used to reconstruct the physics objects relevant to this analysis were given in §4. In this section, the analysis specific criteria used for each object are summarised. The objects as defined here inform the event level selections described in §5.4.

5.3.1 *Electrons*

Electron reconstruction is described in §4.2.2. The electron selection criteria used in this analysis are summarised in Table 5.4. Certain criteria are split into those used in the preselection prior to the overlap removal procedure (see §5.3.5) and those used at the analysis level.

The *Medium* likelihood identification and *Loose_VarRad* isolation working points are used to reduce the background from non-prompt electrons and photon conversions. Electron candidates are required to be within the fiducial region $|\eta| < 2.47$, excluding the calorimeter transition region ($1.37 < |\eta| < 1.52$). Constraints of $|z_0 \sin \theta| < 0.5$ mm and the transverse impact parameter significance $|\frac{d_0}{\sigma(d_0)}| < 5$, where $\sigma(d_0)$ is the uncertainty on the reconstructed d_0 , are placed on the track-to-vertex association.

In the ZH region, exactly two same flavour leptons are required (see §5.4 for the specific selections used to define each region). For electrons, the leading electron must have $p_T > 27$ GeV, while the subleading electron must have $p_T > 10$ GeV. The WH

Feature	Criterion	
	Preselection	Analysis level
Identification	<i>Loose</i>	<i>Medium</i>
Isolation	-	<i>Loose_VarRad</i>
p_T	$> 10 \text{ GeV}$	$> 27(\text{leading}) / > 10 \text{ GeV}$
η range	$ \eta < 2.47$ except $1.37 < \eta < 1.52$	
$ z_0 \sin \theta $	$< 0.5 \text{ mm}$	
$ \frac{d_0}{\sigma(d_0)} $	< 5	

Table 5.4: Summary of electron definition criteria used in the analysis.

Feature	Criterion	
	Preselection	Analysis level
Identification	<i>Loose</i>	<i>Medium</i>
Isolation	-	<i>Loose_VarRad</i>
p_T	$> 10 \text{ GeV}$	$> 27(\text{leading}) / > 10 \text{ GeV}$
η range	$ \eta < 2.5$	
$ z_0 \sin \theta $	$< 0.5 \text{ mm}$	
$ \frac{d_0}{\sigma(d_0)} $	< 3	

Table 5.5: Summary of muon definition criteria used in the analysis.

region requires exactly one lepton with $p_T > 27 \text{ GeV}$.

5.3.2 Muons

Muon reconstruction is described in §4.2.4. The muon selection criteria used in this analysis are summarised in Table 5.5. Muon identification, isolation and p_T criteria are defined analogously to electrons as described in §5.3.1. Muon candidates are required to have $|\eta| < 2.5$. Constraints of $|z_0 \sin \theta| < 0.5 \text{ mm}$ and $|\frac{d_0}{\sigma(d_0)}| < 3$ are placed on the track-to-vertex association.

5.3.3 Jets

Jet reconstruction is described in §4.2.1. This analysis uses *EMTopo* jets. A preselection is applied to jets prior to overlap removal (see §5.3.5), then two analysis level collections of jets are defined, *forward jets*, which are only used in the *VBF* region, and *central jets*, which are used as inputs to the per-jet BDT described in §5.4.2 and matched to DVs. A

5.3. OBJECT DEFINITIONS

Jet reconstruction parameters			
Parameter	Value		
Algorithm	anti- k_t		
R-parameter	0.4		
Input constituent	$EMTopo$		
Selection criteria			
Feature	Preselection	Forward Jets	Central Jets
Jet cleaning	-	$LooseBad$	
p_T		$> 20 \text{ GeV}$	
$ \eta $	< 4.9		< 2.5
(E_T^0, E_T^1)	-	$> (100, 80) \text{ GeV}$	-
(η_0 , η_1)	-	$< (3.2, 4.9)$	-
m_{jj}	-	$> 1200 \text{ GeV}$	-
$\Delta\eta_{jj}$	-	> 4	-
$\Delta\phi_{jj}$	-	< 2	-

Table 5.6: Summary of jet reconstruction parameters and selections used in the analysis.

summary of the jet selections is given in Table 5.6.

Forward jets

A defining signature of a Higgs boson produced via VBF is the presence of two forward jets. A forward jet pair (j_0, j_1) is formed from jets with $p_T > 20 \text{ GeV}$ and $|\eta| < 4.9$. The pair must satisfy $(E_T^0, E_T^1) > (100, 80) \text{ GeV}$, $(|\eta_0|, |\eta_1|) < (3.2, 4.9)$ as well as cuts on the dijet invariant mass of $m_{jj} > 1200 \text{ GeV}$ and the difference in η and ϕ between the two jets of $\Delta\eta_{jj} > 4$ and $\Delta\phi_{jj} < 2$. The selected forward jet pair is the one with the highest E_T^0 . The methodology used to optimise the forward jet selection is described in Appendix C.

Central jets

Central jets are defined as having $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$. In the VBF search region an overlap removal is applied such that jets in selected forward jet pairs do not enter the selected central jet collection.

Feature	Criterion	
	Preselection	Analysis level
Identification	<i>Loose</i>	<i>Tight</i>
Isolation	<i>FixedCutLoose</i>	<i>FixedCutTight</i>
p_T	$> 15 \text{ GeV}$	
η range	$ \eta < 2.47$ except $1.37 < \eta < 1.52$	

Table 5.7: Summary of photon definition criteria used in the analysis.

5.3.4 Photons

Photon reconstruction is described in §4.2.3. The photon selection criteria used in this analysis are summarised in Table 5.7. A preselection is applied prior to overlap removal (see §5.3.5), after which the analysis level selections are used.

Photons are required to have $p_T > 15 \text{ GeV}$ and be within the fiducial region $|\eta| < 2.37$, excluding the calorimeter transition region $1.37 < |\eta| < 1.52$. To be considered in the analysis, they must also pass the *Tight* identification and *FixedCutTight* isolation working points.

5.3.5 Overlap removal

In ATLAS, objects are reconstructed by separate algorithms using the same detector information. This can result in ambiguities in the reconstruction of electrons, muons, photons, and jets, as the same energy deposits may be used to reconstruct different objects. A sequential procedure, known as overlap removal (OR), is performed to resolve these ambiguities and remove any duplication of physics objects.

OR proceeds through a sequence of comparisons between pairs of objects in distinct collections. If the objects overlap geometrically (either through ΔR matching, a shared ID track, or ghost association²), then one of the objects is removed based on a pre-defined priority. The result is a complete set of physics objects containers without any ambiguity. The OR sequence used in this analysis is given in Table 5.8.

²See §6.2.4 for a description of ghost association.

5.4. EVENT SELECTION

Reject	Against	Criteria
μ	e	is calorimeter-tagged and shared ID track
e	μ	shared ID track
γ	e	$\Delta R < 0.4$
γ	μ	$\Delta R < 0.4$
j	e	$\Delta R < 0.2$
e	j	$\Delta R < 0.4$
j	μ	jet n -tracks < 3 and (ghost-associated to μ track or $\Delta R < 0.2$)
μ	j	$\Delta R < 0.4$
γ	j	$\Delta R < 0.4$

Table 5.8: The overlap removal sequence used in the analysis to disambiguate electrons (e), muons (μ), jets (j), and photons (γ).

5.4 EVENT SELECTION

This section describes the event-level selections used to define regions with high sensitivity to the $H \rightarrow ss \rightarrow q\bar{q}q\bar{q}$ signal, as well as signal-free regions that are used to derive and validate the background estimation procedure. A set of preselections targeting certain topologies are defined, from which control and signal regions are formed by imposing requirements based on the displaced jet and DV content of the event. The leptons, jets and photons entering the selections must pass the analysis level requirements listed in Tables 5.4, 5.5, 5.6, and 5.7.

5.4.1 Preselection

Single lepton

The *Presel_1lep* region is defined by the presence of exactly one lepton with $p_T > 27$ GeV, $E_T^{\text{miss}} > 30$ GeV and at least two central jets. Events must pass the lepton trigger preselection. These requirements serve as a baseline to identify events containing a W boson for the WH search channel. The most prominent SM backgrounds passing this preselection are $W + \text{jets}$ and $t\bar{t}$ events.

Di-lepton

The *Presel_2lep* region is defined by the presence of exactly two leptons, with the same flavour and opposite charge, and at least two central jets. The leading lepton must have $p_T > 27$ GeV, and the invariant mass of the di-lepton system m_{ll} must be between 76 and 106 GeV. Events must pass the lepton trigger preselection. These requirements serve as a baseline to identify events containing a Z boson for the ZH search channel. The most prominent SM backgrounds passing this preselection are Z +jets and $t\bar{t}$ events.

VBF

The *Presel_VBF* region is defined by the presence of a forward jet pair passing the selection defined in Table 5.6, zero leptons and at least two central jets. Events must pass the VBF trigger preselection. These requirements serve as a baseline to identify events for the VBF search channel. The most prominent SM backgrounds passing this preselection are QCD multijet events.

Photon

The *Presel_Phot* region is defined by the presence of at least one photon, exactly zero leptons and at least two central jets. The leading photon must have $p_T > 160$ GeV. Events must pass the photon trigger preselection. This region effectively rejects events from the signal models and provides an independent set of data events to validate the background estimation procedure.

A summary of each preselection region is given in Table 5.9.

5.4.2 Displaced jet selection

The LLP being considered in this analysis decays hadronically, therefore the presence of one or more jets in an event originating at a secondary vertex displaced from the IP is a key signature of the signal process. To identify signal-like displaced jets, and differentiate them from prompt jets, a BDT [90] is trained on a custom set of discriminating jet observables.

The BDT is trained using *XGBoost* [91] with approximately 200,000 jets from both

5.4. EVENT SELECTION

	<i>Presel_1lep</i>	<i>Presel_2lep</i>	<i>Presel_VBF</i>	<i>Presel_Phot</i>
<i>n</i> -leptons	1	2 (SFOS)	0	0
Central Jets		≥ 2		
Forward Jets	-		2	-
<i>n</i> -photons		-		≥ 1
E_T^{miss}	$> 30 \text{ GeV}$	-	-	-
m_{ll} range	-	$76 < m_{ll} < 106 \text{ GeV}$	-	-
Trigger	Single- + Di-Lepton		VBF incl.	Photon

Table 5.9: Summary of selections for each preselection region. The abbreviation SFOS stands for same-flavor opposite sign leptons.

Signal		
Samples	m_s (GeV)	$c\tau$ (mm)
ZH	16, 55	10, 100
WH	16, 55	10, 100
Background		
Samples	Channels	Decay Modes
$Z + \text{jets (1/3)}$	$\mu\mu, ee$	light-, b -, c -quarks
$W + \text{jets (1/3)}$	$\mu\nu, e\nu$	
$t\bar{t}$ (1/3)	-	-

Table 5.10: Signal and background MC composition of the BDT training set. The signal samples are produced at each lifetime for each mass, giving four samples in total for both ZH and WH .

signal and background MC samples. For signal, only samples with $a \rightarrow b\bar{b}$ are used in training. Additionally, each signal jet is required to be matched to a truth b -jet to ensure a clean sample of displaced jets. For background jets, the two jets with the highest p_T per event are included in the training set. The composition of the training set is given in Table 5.10.

The inputs to the BDT are as follows:

- the *charged hadron fraction* (CHF), defined as the ratio of the p_T of all prompt tracks, with $d_0 < 0.5 \text{ mm}$, associated to the jet to the p_T of the jet:

$$\text{CHF} = \frac{(\sum_{\text{prompt tracks}} \vec{p})_T}{p_T^{\text{jet}}}. \quad (5.4.1)$$

Displaced jets will be associated to fewer prompt tracks, and therefore have a

lower CHF.

- the *displaced charged hadron fraction* (dCHF), defined as the ratio of the p_T of all displaced tracks, with $d_0 > 0.5$ mm, associated to the jet to the p_T of the jet:

$$\text{dCHF} = \frac{(\sum_{\text{displaced tracks}} \vec{p})_T}{p_T^{\text{jet}}}. \quad (5.4.2)$$

Displaced jets will be associated to more displaced tracks, and therefore have a higher dCHF.

- α_{max} , defined as

$$\alpha_{\text{max}} = \max(\alpha_i), \quad (5.4.3)$$

where α_i is computed for each primary vertex PV_i by taking the ratio of the p_T of all tracks in the jet that are $z_0 \sin \theta < 0.3$ mm from PV_i to the p_T of the jet:

$$\alpha_i = \frac{(\sum_{\text{tracks matched to } PV_i} \vec{p})_T}{(\sum_{\text{tracks in jet}} \vec{p})_T}. \quad (5.4.4)$$

Prompt jets have more tracks matched the primary vertex than displaced jets and therefore a higher α_{max} .

- the *median log impact parameter significance* (IPSIG), defined as the median (log) transverse impact parameter significance of tracks associated to the jet:

$$\text{IPSIG} = \text{median}[\log_{10} \left(\frac{d_0^{\text{track}}}{\sigma(d_0^{\text{track}})} \right)]. \quad (5.4.5)$$

Prompt tracks tend to have $d_0 \approx \sigma(d_0)$ and therefore $\text{IPSIG} \approx 0$. A well reconstructed displaced track will have $\text{IPSIG} \gg 0$, therefore displaced jets should often have $\text{IPSIG} > 0$.

- d_0^{max} , the maximum d_0 of all tracks in the jet. A displaced jet is associated to more displaced tracks and will tend to have a greater d_0^{max} than prompt jets.

A systematic uncertainty accounting for the mismodelling of these input variables between data and MC, which is found to vary between 10 – 20%, is described in §5.6.2.

Feature	Criterion
Material veto	True
$\max \chi^2 / n_{\text{DoF}}$	5
$\max r_{\text{DV}}$	300 mm
$\max z_{\text{DV}} $	300 mm
$\min n_{\text{trk}}$	3
$\min m_{\text{DV}} / \Delta R_{\text{max}}$	4 GeV
$\min H_{\text{T}}$	10 GeV
$\min d_{0,\text{max}}$	3 mm
$\min d_{0,\text{min}}$	0.1 mm
$\max \Delta R_{\text{jet}}$	0.6

Table 5.11: The full set of selections applied to vertices.

The BDT outputs a score between 0 and 1, where more displaced jets are expected to have a score nearer to 1. For the purpose of the analysis, a displaced jet is defined as having a BDT score greater than 0.5.

The analysis uses an event level discriminator, the $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ score, computed by taking the product of the two highest per-jet BDT scores greater than 0.5 in an event. Signal events should have $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ scores close to 1, indicating the presence of at least two highly signal-like displaced jets in the event. Events with fewer than two displaced jets are assigned a $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ score of 0.

5.4.3 Displaced vertex selection

There are a number of background processes that mimic the signature of a displaced secondary vertex originating from the decay of an LLP. These may be irreducible SM backgrounds, as well as instrumental or algorithmic in nature. This background contribution is reduced by applying a selection on the location and kinematics of reconstructed vertices. The full criteria are given in Table 5.11, and explained below.

A common source of displaced hadronic vertices is from particles interacting with nuclei within the detector material. Such vertices tend to be located around areas in the detector containing a high density of material, and can be effectively removed by a material veto. The veto uses three-dimensional maps of detector material constructed from the locations of low mass and low track multiplicity vertices in data and MC. The material veto step removes vertices whose coordinates coincide with detector material.

Cuts on r_{DV} and $|z_{\text{DV}}|$ are dictated by the maps, which extend from $0 < r < 300$ mm. Poorly fitted background vertices are removed by requiring $\chi^2/n_{\text{DoF}} < 5$.

Once the material veto is applied the remaining background vertices will mostly be from SM processes or combinatorial. Requiring the vertex to have at least three tracks effectively removes secondary vertices originating from two-pronged SM process such as K_S^0 decays and photon conversions, as well as fake vertices, without sacrificing too much signal MC. Vertices with $n_{\text{trk}} > 2$ are predominantly from b -hadron decays and *random crossings*. A random crossing occurs when an unrelated track passing in close proximity to a decay is erroneously included in the vertex fit, artificially increasing its track multiplicity. Figure 5.4a shows the distributions of n_{trk} in selected exotic Higgs decay signal samples with $H \rightarrow ss \rightarrow 4b$ and data for events satisfying the *CR1lep* selections (see §5.4.4).

Two discriminating variables designed to exploit the geometry of vertices with random crossings are included in the selection. The first is termed the *reduced mass* of the vertex, and is computed by dividing the mass of the vertex m_{DV} by

$$\Delta R_{\text{max}} = \max_i \left\{ \Delta R(\mathbf{P}_i, \sum_{j \neq i} \mathbf{P}_j) \right\}, \quad (5.4.6)$$

where $\mathbf{P}_i$ is the four-vector of the i -th track belonging to the vertex. Vertices with random crossings tend to have a larger ΔR_{max} than signal vertices, and therefore a lower reduced mass. Vertices are required to have $m_{\text{DV}}/\Delta R_{\text{max}} > 4$ GeV. Figure 5.4b shows the distributions of $m_{\text{DV}}/\Delta R_{\text{max}}$ in selected exotic Higgs decay signal samples with $H \rightarrow ss \rightarrow 4b$ and data for events satisfying the *CR1lep* selections (see §5.4.4).

The second is defined as the scalar sum p_T of tracks with respect to the DV, H_T :

$$H_T = \left(\sum_i p_{i, \text{wrt DV}} \right)_T, \quad (5.4.7)$$

where $p_{i, \text{wrt DV}}$ is the momentum of the i -th track in the vertex with respect to the (x, y, z) coordinates of the vertex. This variable is correlated to the reduced mass for signal, and is a useful discriminant. Vertices are required to have $H_T > 10$ GeV.

Placing cuts on the minimum and maximum $|d_0|$ of tracks associated to the vertex

5.4. EVENT SELECTION

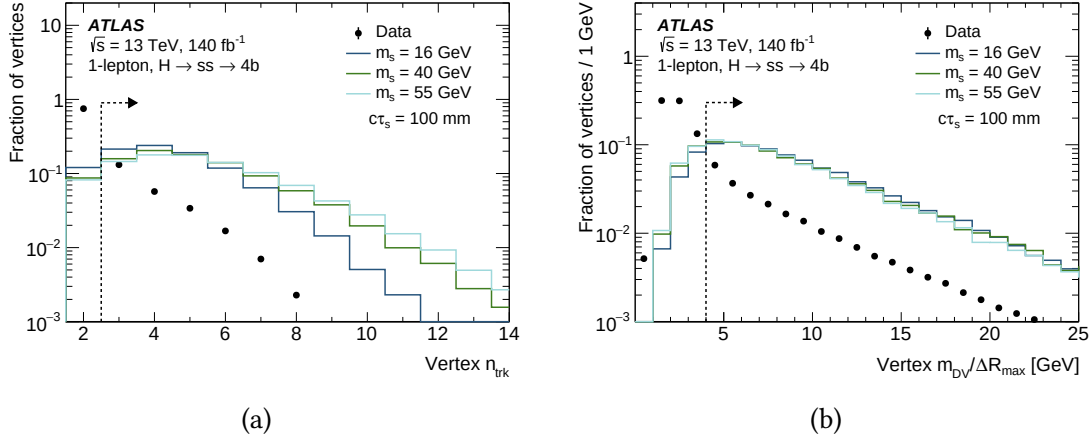


Figure 5.4: Normalised distributions of n_{trk} (a) and $m_{\text{DV}}/\Delta R_{\text{max}}$ (b) for vertices in selected exotic Higgs decay signal samples with $H \rightarrow ss \rightarrow 4b$ and data. Vertices in events satisfying the $CR1lep$ region selection are considered, and are required to pass all selections listed in Table 5.11, except the selection on the observable being shown. The arrow indicates the selection requirement that is placed on the observable [1].

further reduces the acceptance of background vertices from QCD processes, such as b -hadron decays, whose tracks tend to have lower $|d_0|$ than signal.

Finally, all selected vertices are required to be matched to a jet (with BDT score greater than 0.5). The matching proceeds iteratively where in each step the vertex with the fewest jets within $\Delta R < 0.6$ is chosen. This DV is matched to the displaced jet with which it has the lowest ΔR , then both the vertex and jet are removed from further consideration before the process repeats, such that a DV can be matched to at most one jet and vice-versa.

5.4.4 Control and signal regions

Having defined the event-level discriminant $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$, as well as the DV selection, each preselected search region can now be divided into control and signal regions.

Under the conservative assumption of $BR(H \rightarrow ss) = 100\%$, the signal contamination was found to be below 10% for events in each preselected search region with $0.25 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.7$ [92]. Therefore, control regions $CR1lep$, $CR2lep$, and $CRVBF$ can be defined for each search region by requiring $0.25 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.7$. Regions with additional requirements on the DV multiplicity n_{DV} are identifiable through a suffix e.g. $CR1lep1DV$. Unlike the signal regions, data in control regions can be unblinded during the development of the analysis, as there is little risk of exposing any potential

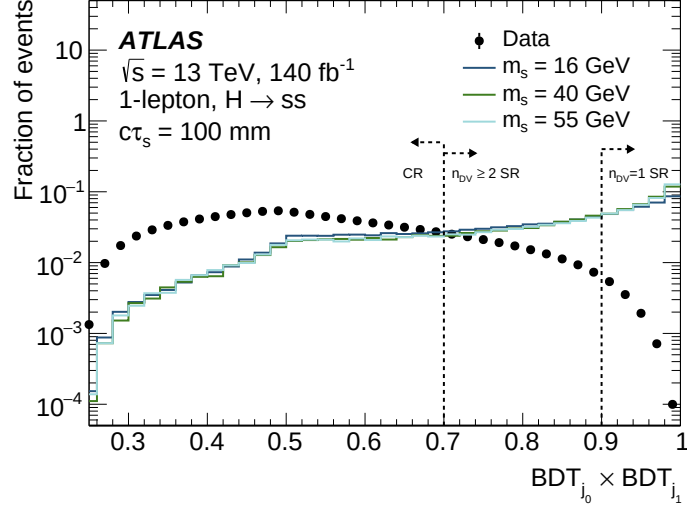


Figure 5.5: Normalised distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for events in the *Presel_1lep* region, with the additional requirement that the events contain at least two jets with BDT scores greater than 0.5, in data (black points) and simulated exotic Higgs decay signal samples (solid lines). Arrows indicate the selections that are made to split events into a control region, and signal regions for events with $n_{\text{DV}} = 1$ and $n_{\text{DV}} \geq 2$ [1].

	SR1lep1DV	SR1lep2DV	SR2lep1DV	SR2lep2DV	SRVBF1DV	SRVBF2DV
Preselection	<i>Presel_1lep</i>		<i>Presel_2lep</i>		<i>Presel_VBF</i>	
$\text{BDT}_{j_0} \times \text{BDT}_{j_1}$	> 0.9	> 0.7	> 0.9	> 0.7	> 0.9	> 0.7
n_{DV}	$= 1$	≥ 2	$= 1$	≥ 2	$= 1$	≥ 2

Table 5.12: Summary of analysis selections for the six signal regions.

signal.

The six signal regions used in this analysis are summarised in Table 5.12. These are based on the production mechanism of the Higgs boson, the $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ score, and n_{DV} . Figure 5.5 demonstrates how the $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ score is used to separate events in the *Presel_1lep* region, with the additional requirement that the events contain at least two jets with BDT scores greater than 0.5, into control and signal regions.

5.5 BACKGROUND ESTIMATE

5.5.1 Procedure

A data-driven method is used to estimate the residual background contribution in each of the signal regions. For an event to pass the full signal selection, it must contain at least two displaced jets, where at least one has been matched to a DV. The probability that any given jet will be matched to a DV is found to be highly correlated with the p_T

Jet observable	Bin edges
p_T	[20, 40, 60, 100, 150, 200, 300, 1000]
DL1r	[−10, 2, 15]
BDT score	[0.5, 0.525, 0.55, 0.575, . . . , 0.925, 0.95, 0.975, 1.0]

Table 5.13: The per-jet efficiency map binning scheme for the $2lep$ and the VBF search regions.

of the jet, its DL1r b -tagging score, as well as its per-jet BDT output. To compute the background estimate, a per-jet vertex efficiency map is derived that parameterises the probability of a jet containing a DV with respect to these three variables. The map is computed for a given region by constructing two three-dimensional histograms binned in jet p_T , DL1r score, and BDT score: one containing all displaced jets matched to a DV passing the selection described in §5.4.3, the other all central jets, entering the region. The efficiency map is defined as the ratio of the former to the latter. The vertex matching efficiencies for jets with BDT score below 0.5 are zero by construction. Per-jet efficiency maps are computed for each search region using events from its corresponding control region, where $0.25 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.7$ (as described in §5.4.4). To produce the background estimate, per-event weights, equivalent to the probability of an event containing a certain number of DVs, are derived for each region, using the map obtained from that region.

The binning schemes for the per-jet efficiency maps, given in Tables 5.13 and 5.14, are chosen to balance the accuracy and precision of the background estimate: too fine of a binning would lead to a large statistical uncertainty in the final estimate, whereas too coarse of a binning would poorly encapsulate intra-bin variations and lead to a less accurate result. The binning in the BDT score axis of the maps varies between the $2lep$ and VBF regions, and the $1lep$ region; it is 0.025 for the former, and 0.01 for the latter. A courser binning is used in the $2lep$ and VBF regions due to them having containing fewer events. The systematic uncertainty due to the choice of binning is detailed in §5.6.

When computing the maps, jets with $p_T > 1000$ GeV or $\text{DL1r} > 15$ are included in the final $p_T/\text{DL1r}$ bin. Similarly, when extracting efficiencies from the map for jets with $p_T > 1000$ GeV or $\text{DL1r} > 15$, the final bins are used.

Maps derived using sufficiently low statistics may contain empty bins if the corre-

Jet observable	Bin edges
p_T	[20, 40, 60, 100, 150, 200, 300, 1000]
DL1r	[−10, 2, 15]
BDT score	[0.5, 0.51, 0.52, 0.53, . . . , 0.97, 0.98, 0.99, 1.0]

 Table 5.14: The per-jet efficiency map binning scheme for the $1lep$ search region.

sponding bins in the numerator are not populated with at least one vertex-matched jet. Such maps will result in an underprediction of the background, due to certain jets in p_T -DL1r-BDT score space being artificially assigned zero probability of being matched to a DV. To rectify this, an interpolation is performed. Empty bins are assigned values based on the arithmetic mean of up to nine surrounding bin values (depending on their availability): the two nearest non-zero bins in both directions of the p_T and BDT score axis, and (as it only contains two bins) the neighbouring bin along the DL1r axis.

The probability that a given jet is matched to a DV $P(1 \text{ DV}|j)_{\text{jet}}$ is extracted from the interpolated maps. The probability that an event contains exactly one DV $P(1 \text{ DV})_{\text{event}}$ can then be computed from a binomial distribution over all central jets in the event:

$$P(1 \text{ DV})_{\text{event}} = \sum_{i=1}^{n_{\text{jet}}} P(1 \text{ DV}|j_i)_{\text{jet}} \times \prod_{k \neq i} (1 - P(1 \text{ DV}|j_k)_{\text{jet}}). \quad (5.5.1)$$

Similarly, the probability that an event contains at least two DVs $P(2 \text{ DV})_{\text{event}}$ is given by:

$$P(2 \text{ DV})_{\text{event}} = 1 - P(1 \text{ DV})_{\text{event}} - P(0 \text{ DV})_{\text{event}}, \quad (5.5.2)$$

where

$$P(0 \text{ DV})_{\text{event}} = \prod_i^{n_{\text{jet}}} (1 - P(1 \text{ DV}|j_i)_{\text{jet}}) \quad (5.5.3)$$

is the probability of an event having no DVs.

Both $P(1 \text{ DV})_{\text{event}}$ and $P(2 \text{ DV})_{\text{event}}$ are applied separately as per-event weights to all events passing the preselection in each region to estimate the expected number of background events in data with $n_{DV} = 1$ and $n_{DV} \geq 2$, respectively.

The three efficiency maps derived in $CR1lep$, $CR2lep$ and $CRVBF$ are used to derive background predictions for each of the six signal regions defined in 5.12. These predic-

tions are computed as a function of the $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ score.

5.5.2 Validation

Before the maps are used to predict the background yield in the signal regions, the method must first be validated using control and validation regions. As the method relies on applying per-event weights to the events in data, predicted distributions for event-level variables can be obtained and compared to the observed distributions. Good agreement in both the normalisation and shape of these distributions implies that the per-jet efficiency maps can effectively model the interdependence between these variables and the expected DV probabilities, and the approach is able to produce robust predictions. In addition to a closure test in which the background yield in the control regions is estimated by applying the maps to the same set of events with which they were derived, two tests are performed to validate the application of the data-driven background estimation from the control regions to larger values of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$.

First, three validation regions are defined using a subset of events in each preselected search region with $n_{\text{DV}} = 1$ and $0.7 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.9$. The maps derived in the control regions are used to predict the background in each corresponding validation region. Figure 5.6 shows the observed data and predicted background in the *1lep*, *2lep* and *VBF* validation regions. The observed data is found to agree with the predicted background within uncertainties.

The second test uses the photon region to validate the application of a map derived in a low $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ region to a higher $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ region, and to events with $n_{\text{DV}} \geq 2$. The *Presel_Phot* region, defined in Table 5.9, is divided into a control region *CRPhot* with $n_{\text{DV}} = 1$ and $0.25 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.7$ and two validation regions: one with $n_{\text{DV}} = 1$ and $\text{BDT}_{j_0} \times \text{BDT}_{j_1} > 0.7$, and another with $n_{\text{DV}} \geq 2$ and $\text{BDT}_{j_0} \times \text{BDT}_{j_1} > 0.25$. An additional map is derived using events in *CRPhot*, then a closure test is performed by predicting the number of events in each validation region. Figure 5.7 shows the predicted and observed distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ in the two photon validation regions. The data distributions are found to be well modelled by the predicted background for events containing both $n_{\text{DV}} = 1$ and $n_{\text{DV}} > 2$.

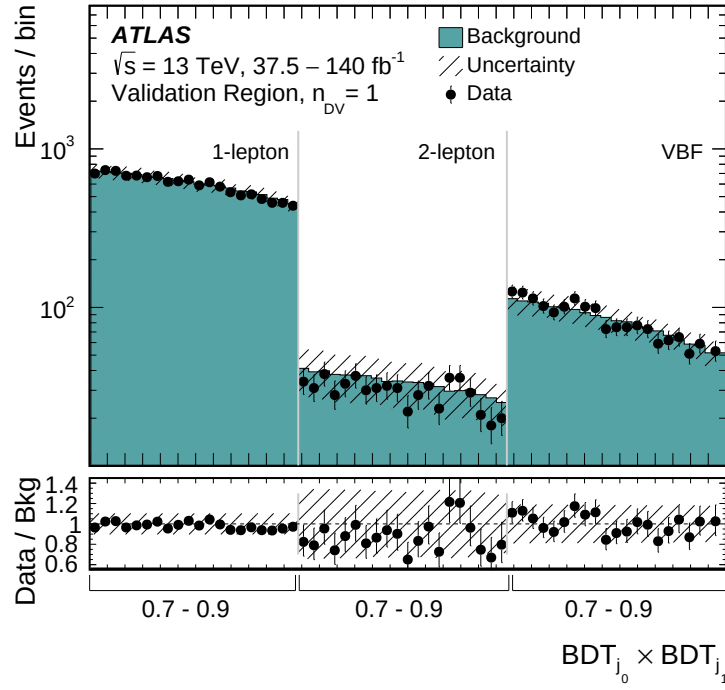


Figure 5.6: Distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for the observed data (black points) and background prediction (teal) along with its uncertainty in the three $n_{\text{DV}} = 1$ validation regions with $0.7 < \text{BDT}_{j_0} \times \text{BDT}_{j_1} < 0.9$ [1].

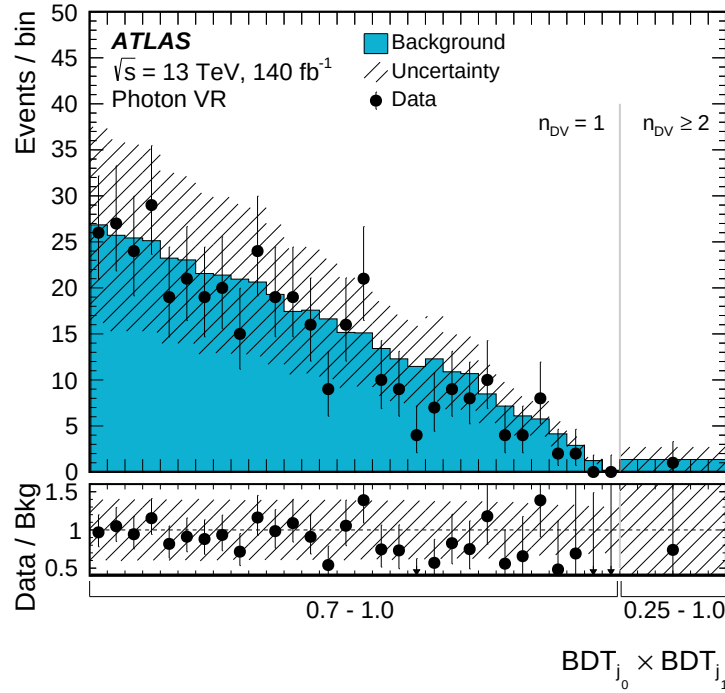


Figure 5.7: Distributions of $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for the observed data (black points) and background prediction (blue) along with its uncertainty in the two photon validation regions containing events with $n_{\text{DV}} = 1$ and $n_{\text{DV}} > 2$ [1].

5.6. UNCERTAINTIES

Both these tests indicate that the background estimation strategy is sufficiently robust. They demonstrate that the efficiency maps are able to accurately predict the background yield in regions containing data that is independent from that used to derive them, and can be applied to events with $n_{\text{DV}} \geq 2$.

5.6 UNCERTAINTIES

Both the final data-driven background prediction, described in §5.5, and predicted signal yield are impacted by various sources of uncertainty. These must be considered in the statistical fit described in §5.7. In this section, the sources of uncertainty on the predicted background and signal yields are described.

5.6.1 Uncertainty on the background prediction

The background prediction is subject to both statistical and systematic uncertainties.

Statistical uncertainty

The control regions used to derive the per-jet probability maps described in §5.5 contain limited statistics. The efficiency of each bin i in a map, given by $\epsilon_i = n_{\text{jet},i}^{\text{DV matched}}/n_{\text{jet},i}$, has an uncertainty given by

$$\sigma(\epsilon_i) = \epsilon_i \sqrt{\frac{1}{n_{\text{jet},i}^{\text{DV matched}}} + \frac{1}{n_{\text{jet},i}}}. \quad (5.6.1)$$

To derive the statistical uncertainty on the background prediction, these per-bin uncertainties are propagated to the final background estimate using a toy method. For each region, the nominal efficiency map is varied bin-by-bin using a Gaussian probability density function with width given by the uncertainty on the bin to create $N = 100$ alternate efficiency maps. These varied maps translate to N approximately Gaussian distributed per-event weights, calculated using Eqs. (5.5.1)-(5.5.3). The standard deviations of the distributions of these weights are used to compute up and down variations on the predicted background yield. Each signal region has its own statistical uncertainty. This is because different maps are used to derive the background estimate in each region, so

the statistical uncertainty between the regions is uncorrelated. The statistical uncertainty varies between 10 – 20%, depending on the region.

Systematic uncertainty

A systematic uncertainty on the background estimate is introduced accounting for the choice of binning used in the efficiency maps (as shown in Tables 5.13 and 5.14). In theory, the finest binning would produce the most accurate prediction, albeit at the cost of a large statistical uncertainty. As described in §5.5.1, a binning in the BDT score axis of 0.025 was chosen for the $2lep$ and VBF regions. However, the predictions given by these maps are found to differ slightly with respect to those given by maps using a binning of 0.01 in the BDT score axis. A systematic uncertainty is introduced to account for these differences.

To compute the uncertainty, background estimates using maps derived in $CR2lep$ and $CRVBF$ with a BDT score binning of 0.01 and 0.025 are computed for the four corresponding signal regions ($SR2lep1DV$, $SR2lep2DV$, $SRVBF1DV$, and $SR2VBF2DV$). The difference between the predictions produced by each map in each signal region is taken as a systematic uncertainty that quantifies the effect of the choice of binning in the probability map. No systematic uncertainty is applied to the $1lep$ region, since the number of events are sufficient to nominally use the map with BDT score binning of 0.01.

5.6.2 Uncertainty on the signal yields

The predicted signal yield, i.e. the number of events in signal MC that pass the final signal region selections, is subject to a number of theoretical and experimental sources of uncertainty. Experimental uncertainties primarily originate from the mismodelling of physics objects and pileup between MC and data. They are evaluated by dedicated ATLAS performance groups and incorporated into the analysis as event re-weightings or prescales of the energy or momentum of an object, before their effect is propagated to the overall signal yield.

5.6. UNCERTAINTIES

Luminosity and pile-up

The uncertainty on the integrated luminosity on all data collected by ATLAS during Run 2 is 1.7% [34]. A pile-up re-weighting is applied to each MC campaign to ensure the MC pile-up profiles match those in data. A systematic uncertainty associated to this re-weighting is evaluated following a standard ATLAS procedure.

Leptons

Efficiency scale factors are used to account for mismodeling of electrons and muons between MC and data [93, 62, 65]. These consist of electron and muon reconstruction, identification, and isolation scale factors, as well as scale factors used to correct for differences in the efficiency of electron and muon triggers between MC and data. Additionally, correction factors that account for uncertainties in the muon p_T scale and resolution and electron energy scale and resolution are used.

Jets

Systematics are included on the JES [55]. These were derived for *EMTopo* jets following recommendations from the ATLAS *JetEtMiss* combined performance group. Uncertainties on the JER are also included, but these have a subdominant effect.

Tracking and vertexing

A systematic uncertainty is evaluated based on the uncertainty in the reconstruction efficiency of both standard and LRT tracks. These uncertainties are propagated to the secondary vertex reconstruction using a track removal procedure, which proceeds as follows:

1. For each track, extract the uncertainty on its reconstruction efficiency.
2. Generate a random number between 0 and 1. If this number exceeds the relative uncertainty on the efficiency for the track, the track is retained, otherwise the track is removed.
3. Rerun the secondary vertex reconstruction using the modified track collection.

4. Compare the signal selection efficiency between the nominal and modified secondary vertex collections. The difference in the final yields is symmetrised and taken as a systematic uncertainty.

BDT discriminant

A systematic uncertainty is used to account for mismodeling of the per-jet BDT input variables between data and MC. First, a per-jet uncertainty is derived by finding the ratio of the distributions in data and MC of the per-jet BDT scores of the two leading jets in events in the *Presel_2lep* region. This uncertainty increases with BDT score, and is approximately 5–10% in the highest BDT score bins. A per-event uncertainty is obtained by applying the uncertainties of the two jets used to compute the $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ score as scale factors to produce a modified $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ distribution. The modified distribution is symmetrised and taken as a systematic uncertainty on the final signal yield in each region.

5.7 STATISTICAL ANALYSIS

In searches for new physics, a statistical analysis is performed to determine which of two hypotheses are most consistent with the observed data. These are the background-only hypothesis and the signal-plus-background hypothesis. Broadly speaking, a claim of discovery is made if the observed data is shown to be incompatible with the background-only hypothesis. If the background-only hypothesis cannot be rejected, limits are set that express the sensitivity of the analysis. In this section, a summary of the statistical framework used to interpret the results of the analysis presented in this thesis is given. The description largely follows the formalism of Ref. [94].

5.7.1 *The likelihood function*

The likelihood function quantifies the likelihood of the observed data, as a function of the hypothesised signal strength $\mu = BR(H \rightarrow ss \rightarrow 4q)$, where $q = c, u, b$. The background-only hypothesis is obtained by setting $\mu = 0$, while $\mu > 0$ points to a signal-plus-background hypothesis. The likelihood function is given by

$$\mathcal{L}(\mathbf{n}, \boldsymbol{\theta}^0 \mid \mu, \boldsymbol{\theta}) = \prod_{r \in \text{SR}} \prod_{i \in \text{bins}} \mathcal{P}_{\text{Pois}}(n_{ri} \mid \mu S_{ri}(\boldsymbol{\theta}) + B_{ri}(\boldsymbol{\theta})) \times \prod_{j \in \text{syst}} \mathcal{P}_{\text{Gaus}}(\theta_j^0 \mid \theta_j, \Delta\theta_j). \quad (5.7.1)$$

The first term results from treating each bin in each signal region as a distinct Poisson counting experiment, and computing the product of the Poisson probabilities for each bin, given by

$$\mathcal{P}_{\text{Pois}}(n_{ri} \mid \mu S_{ri}(\boldsymbol{\theta}) + B_{ri}(\boldsymbol{\theta})) = \frac{(\mu S_{ri}(\boldsymbol{\theta}) + B_{ri}(\boldsymbol{\theta}))^{n_{ri}}}{n_{ri}!} \exp[-(\mu S_{ri}(\boldsymbol{\theta}) + B_{ri}(\boldsymbol{\theta}))], \quad (5.7.2)$$

where $S_{ri}(\boldsymbol{\theta})$ is the expected number of signal events for bin i in signal region r assuming $BR(H \rightarrow ss \rightarrow 4q) = 1$, $B_{ri}(\boldsymbol{\theta})$ is the expected number of background events for bin i in signal region r , and n_{ri} is the observed number of events.

Explicit in Eq. (5.7.1) and Eq. (5.7.2) is the dependence of the predicted signal and background yields on a set of nuisance parameters (NPs) $\boldsymbol{\theta}$. The NPs are distinct from μ , the parameter of interest, as their post-fit values are not of interest, however the presence of NPs decreases the sensitivity of the analysis and their impact on the expected yield must be accounted for. Each NP corresponds to a different uncertainty, and is subject to a Gaussian constraint in the likelihood fit, given by

$$\mathcal{P}_{\text{Gaus}}(\theta_j^0 \mid \theta_j, \Delta\theta_j) = \frac{1}{\sqrt{2\pi\Delta\theta_j^0}} \exp\left[\frac{-(\theta_j^0 - \theta_j)^2}{2(\Delta\theta_j^0)^2}\right], \quad (5.7.3)$$

where θ_j^0 is the nominal value of the j NP derived through an auxiliary measurement, θ_j is the fitted value, and $\Delta\theta_j^0$ is the prior uncertainty on the NP.

A simultaneous maximum-likelihood fit is performed on all bins in the $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ distribution for all six signals regions in Table 5.12 to compute the maximum likelihood estimates (MLEs) $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$. In the fit, μ is treated as a free parameter that is common to all signal regions. The fitted NP values $\hat{\boldsymbol{\theta}}$ represent deviations from the nominal values that globally provide the best fit to the observed data.

5.7.2 Hypothesis tests

The likelihood is also used to construct a one-sided hypothesis test. For limit setting the test statistic q_μ is used to differentiate the hypothesis with signal rate μ from an alternate hypothesis with signal rate $\mu' < \mu$:

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \quad (5.7.4)$$

where $\lambda(\mu)$ is the profile likelihood ratio, given by

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}. \quad (5.7.5)$$

Here, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the global MLEs, and $\hat{\boldsymbol{\theta}}_\mu$ are the conditional MLEs (CMLEs), or *profiled value*, of $\boldsymbol{\theta}$ that maximise the likelihood for the specified value of μ . The *RooFit* package is used to evaluate the test statistic [95].

The test statistic represents the outcome of an experiment under a specific hypothesis. For the observed data it will have some value q_μ^{obs} , but across many experiments it would have its own distribution $f(q_\mu \mid \mu, \boldsymbol{\theta})$. This distribution can be used to compute the p -value for an observation under the hypothesis μ

$$p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \mid \mu, \hat{\boldsymbol{\theta}}_\mu^{\text{obs}}) dq_\mu, \quad (5.7.6)$$

defined as the probability of observing an equal or more *extreme* result under the assumed hypothesis. The p -value may be transformed into a related quantity called the *significance* Z , given by

$$Z = \Phi^{-1}(1 - p), \quad (5.7.7)$$

where Φ^{-1} is the inverse of the cumulative standard Gaussian distribution. The test statistic is used to define a critical region for which the significance of a result under a given hypothesis exceeds a threshold, and the hypothesis can be rejected. In particle physics, a significance of $Z = 5$ (corresponding to a p -value or test size of 2.9×10^{-7}) of

the observed data assuming the background-only hypothesis is the standard discovery convention.

5.7.3 Limit setting and the CL_S method

If the background-only hypothesis cannot be rejected, the sensitivity of the analysis to various signal-plus-background hypotheses can still be tested. For a test of size α , the values of μ that are not rejected defines a *confidence interval* with a *confidence level* $CL = 1 - \alpha$. The highest value of μ that is not rejected at a certain CL is called the *upper limit* μ_{up} .

In this analysis, the CL upper limits are computed using the CL_S method [96]. The standard 95% CL_S upper limit μ_{up} is found by solving $p'_{\mu_{\text{up}}} = 5\%$, where

$$p'_\mu = \frac{p_\mu}{1 - p_b}, \quad (5.7.8)$$

where p_μ is defined in Eq. (5.7.9), and p_b is the p -value derived using the same test statistic under the background-only hypothesis

$$p_b = 1 - \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \mid \mu = 0, \hat{\theta}_0^{\text{obs}}) dq_\mu. \quad (5.7.9)$$

A value of $\mu > \mu_{\text{up}}$ is regarded as excluded at the 95% CL.

Computing p'_μ requires evaluating the two complex integrals in Eqs. (5.7.9) and (5.7.8). These are approximated using asymptotic formulae [97], and validated against toy MC. In the latter, an ensemble of pseudo-experiments are generated by randomly sampling the model observables according to Eq. (5.7.1). The test statistic q_μ^{obs} is computed for each pseudo-experiment, and the resulting distribution is numerically integrated.

5.7.4 Fit strategy

This analysis uses the *TRexFitter* framework [98, 99] to perform the statistical analysis.

A simultaneous maximum-likelihood fit is performed on all bins in the six signal regions listed in Table 5.12. The number of bins in $\text{BDT}_{j_0} \times \text{BDT}_{j_1}$ for each signal region is given in Table 5.15. The binning for each $2DV$ region is as fine as possible while

$SR1lep1DV, SR2lep1DV, SRVBF1DV$	$SR2lep2DV$	$SR1lep2DV$	$SRVBF2DV$	
No. of Bins	10	1	6	3

Table 5.15: The number of bins in $BDT_{j_0} \times BDT_{j_1}$ score used in the maximum-likelihood fit for each signal region.

ensuring that each bin has at least one predicted background event. The uncertainties on the predicted signal and background yields described in §5.6 are included in the fit as NPs. In the $1lep$ region, where no systematic uncertainty is applied to the background prediction, a linear shape uncertainty is included as an additional degree of freedom in the fit to allow the value of the fitter statistical uncertainty to vary from bin-to-bin.

Expected upper limits are used to quantify the sensitivity of the analysis. The expected upper limit is the upper limit that is expected to be obtained if the background-only hypothesis is true. Expected upper limits are evaluated through an *Asimov fit*, in which the observed data is replaced with the expected background plus the expected signal with μ set to 0.1.

5.8 RESULTS

Having demonstrated the robustness of the background estimation strategy in §5.5, and outlined the fit strategy in §5.7.4, the analysis may now proceed with *unblinding*, or revealing the data in the signal regions, and interpreting the results, following the theory described in §5.7.

5.8.1 Unblinded results

Figure 5.8 shows distributions of $BDT_{j_0} \times BDT_{j_1}$ for the observed data and the background prediction after the maximum-likelihood fit to the data has been performed, for each of the six signal regions listed in Table 5.12. The figure also contains the expected distributions of several signal models as a reference. Overall, no significant deviation from the expected SM background has been observed.

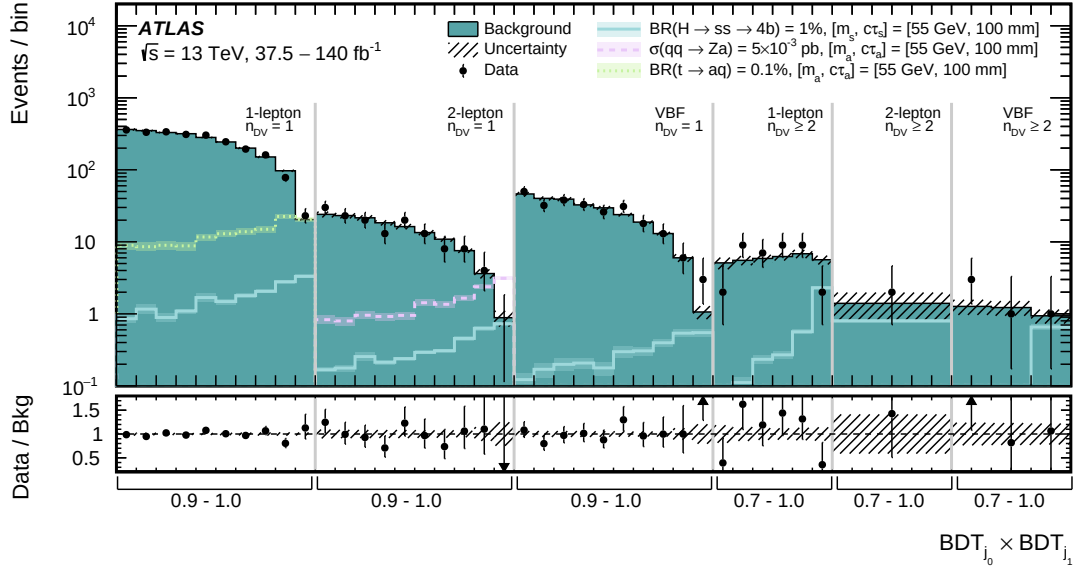


Figure 5.8: Distributions of $BDT_{j_0} \times BDT_{j_1}$ for the observed data (black points) and background prediction (teal) along with its uncertainty after the maximum-likelihood fit to data has been performed. The expected signal yield for the exotic Higgs decay model with $m_s = 55$ GeV and $c\tau_s = 100$ mm is shown in blue, scaled to $BR(H \rightarrow ss \rightarrow 4b) = 1\%$. The expected signal yield for ALP production, with $m_s = 55$ GeV and $c\tau_s = 100$ mm, in association with a Z boson, scaled $\sigma(q\bar{q} \rightarrow Za) = 5 \times 10^{-3}$ pb, and for $t \rightarrow aq$, scaled to $BR(t \rightarrow aq) = 0.1\%$, are shown in green and pink, respectively. The observed data in the $1lep$ and $2lep$ regions correspond to an integrated luminosity of 140 fb^{-1} , while the VBF region corresponds to an integrated luminosity of 37.5 fb^{-1} [1].

5.8.2 Exclusion limits

As no excess of signal-like events have been observed, limits may be derived on the branching ratio of Higgs boson decays to pairs of LLPs, $BR(H \rightarrow ss)$, following the formalism of §5.7. Interpretations in terms of ALP-like models are considered by additionally setting limits on the cross section $\sigma(q\bar{q} \rightarrow Va)$ and $BR(t \rightarrow aq)$.

First, in order to present exclusion limits as a function of LLP proper decay length $c\tau_a$, a lifetime reweighting procedure is applied to derive the expected signal yields for LLPs with lifetimes other than those generated for this analysis. In the remainder of this section, the lifetime reweighting procedure is detailed, then the exclusion limits are then presented along with a comparison of existing constraints on $BR(H \rightarrow ss)$ set by previous searches at the LHC.

Lifetime reweighting

As it is resource-intensive to generate MC simulated samples, it is preferable to generate signal samples for a small number of mean LLP lifetimes, then use a *lifetime reweighting* procedure to extrapolate the signal efficiency to different mean lifetimes. To reweight from the generated lifetime τ_1 to the target lifetime τ_2 , a weight for each LLP i in an event is computed using [100]

$$w_i = \frac{\tau_1}{\tau_2} e^{-(\frac{1}{\tau_2} - \frac{1}{\tau_1})t_i}, \quad (5.8.1)$$

where t_i is the proper lifetime of LLP i . A weight is then applied to each event in the sample, equal to the product of the weights for each LLP in the event. This adjusts the generated distributions to approximate those expected for a signal sample generated at the target lifetime, and the expected signal yields at this lifetime can be computed.

In this analysis, signal samples are generated using four different lifetimes. To reweight to the target lifetime τ_2 , from generated samples with lifetimes τ_1 and τ_3 , where $\tau_1 < \tau_2 < \tau_3$, a critical lifetime t_c , given by

$$t_c = \frac{2 \ln(\frac{\tau_3}{\tau_1})}{\frac{1}{\tau_1} - \frac{1}{\tau_3}}, \quad (5.8.2)$$

5.8. RESULTS

is used to determine which reference sample to use in Eq. (5.8.1). When the sum of the proper decay time t for the two LLPs in the event is less than t_c , events from the τ_1 sample are used, and when this sum is greater than t_c , events from the τ_3 sample are used.

Limits

In both the exotic Higgs decay and ALP signal models, the proper lifetimes of the new scalar or ALP is treated as a free parameter, therefore the limits are presented as a function of the mean LLP proper decay length $c\tau_a$. Limits are computed by performing the CL_s method, as described in §5.7, across the full range of masses and proper lifetimes considered by the analysis. The expected and observed exclusion limits for each signal model are shown in Figure 5.9. The Figure contains exclusion limits at the 95% confidence level for $BR(H \rightarrow ss)$, $\sigma(q\bar{q} \rightarrow Va)$ for $V = Z, W$, and $BR(t \rightarrow aq)$.

5.8.3 Results summary

Having now computed the 95% confidence level upper limits on the parameters of interest for each signal model, the remainder of this section will be dedicated to discussing and interpreting the results, as well as contextualising their significance through comparisons to previous searches.

Tables 5.16 and 5.17 shows the ranges of values of $c\tau_a$ excluded at the 95% confidence level for different values of $BR(H \rightarrow ss \rightarrow 4b/4c)$ and $BR(H \rightarrow ss \rightarrow 4u)$, respectively, for LLP masses in the range 5-55 GeV.

For $BR(H \rightarrow ss \rightarrow 4b)$, branching ratios greater than 1% are excluded for $m_s = 55$ GeV and $m_s = 40$ GeV between $5.7 < c\tau_a < 67.9$ mm and $9.1 < c\tau_a < 33.4$ mm. The exclusion limits are stronger for light-quark final states, due to better displaced vertex reconstruction efficiency. For $BR(H \rightarrow ss \rightarrow 4u)$, branching ratios greater than 1% are excluded for $m_s = 55$ GeV and $m_s = 40$ GeV between $4.4 < c\tau_a < 107.4$ mm and $4.5 < c\tau_a < 80.9$ mm. This is the first search to exclude LLP lifetimes of $\mathcal{O}(10^{-3}-10^{-1}$ mm) for branching ratios $BR(H \rightarrow ss)$ at the 1% level for $m_s > 40$ GeV, filling the gap in exclusion left by previous searches that targeted lifetimes outside this range, as shown in Figure 5.1.

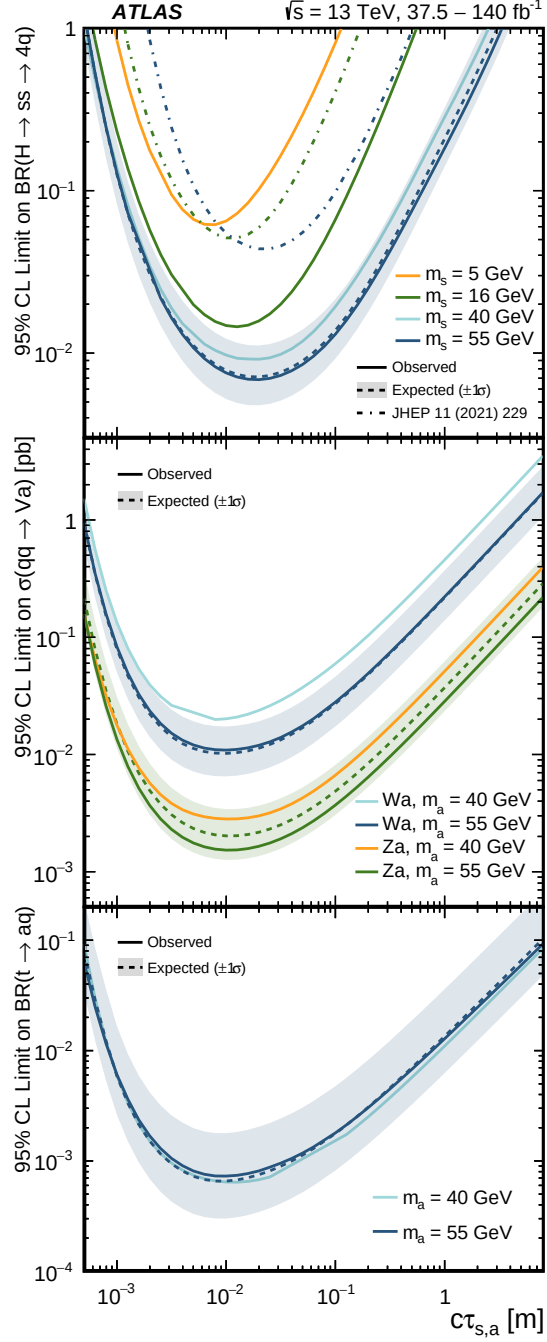


Figure 5.9: The 95% confidence level exclusion limits on $BR(H \rightarrow ss)$ (upper), where s denotes the LLP, $\sigma(q\bar{q} \rightarrow Va)$ for $V = Z, W$ (middle), and $BR(t \rightarrow aq)$ (lower) shown as a function of the mean proper decay length $c\tau$ of the LLP. In each plot, the observed limits are shown with a solid line and the expected limits and $\pm 1\sigma$ uncertainty bands are shown with dashed lines and shaded bands, respectively. In the upper plot, the limits shown are for the branching ratio $H \rightarrow ss \rightarrow 4b$ for $m_s = 16, 40, 55$ GeV, and $H \rightarrow ss \rightarrow 4c$ for $m_s = 5$ GeV. In the upper plot, the observed limits for the previous ATLAS search [77] for exotic Higgs decays are shown with the dotted lines [1].

$BR(H \rightarrow ss \rightarrow 4b/4c)$	Excluded range of $c\tau_s$ (mm)			
	$m_s = 55 \text{ GeV}$	$m_s = 40 \text{ GeV}$	$m_s = 16 \text{ GeV}$	$m_s = 5 \text{ GeV}$
1%	5.7-67.9	9.1-33.4	-	-
5%	1.8-361.2	1.6-254.2	2.2-80.3	-
10%	1.1-626.8	1.2-447.5	1.5-133.3	3.1-19.5
20%	0.8-1070.0	0.9-761.9	1.1-210.2	1.9-37.0

Table 5.16: The ranges of values of LLP proper decay length $c\tau_a$ excluded at the 95% confidence level for different values of $BR(H \rightarrow ss \rightarrow 4b/4c)$, for LLP masses in the range 5-55 GeV.

For $m_s < 16 \text{ GeV}$ and $c\tau_a < 100 \text{ mm}$, this search is the first at the LHC to probe branching ratios of $H \rightarrow ss$ beyond the limit of 12% imposed on Higgs boson decays to BSM states by combining measurements made at ATLAS [9]. As an example, for $BR(H \rightarrow ss \rightarrow 4u)$, branching ratios greater than 5% are excluded for $m_s = 16 \text{ GeV}$ and $m_s = 5 \text{ GeV}$ between $1.8 < c\tau_a < 104.7 \text{ mm}$ and $5.6 < c\tau_a < 8.5 \text{ mm}$.

Figure 5.10 shows the regions in the LLP mass versus proper decay length plane excluded at the 95% confidence level for the exotic Higgs decay model, for this and a number of previous ATLAS searches. The constraints set on the exotic Higgs decay model are a significant improvement over those set by the previous analysis [77], which used the same dataset. The increase in sensitivity is driven by the updated LRT configuration, the addition of the $1lep$ and VBF signal regions, the per-jet BDT to discriminate between signal and background jets, and the inclusion of events with n_{DV} in the signal regions. Notably, the VBF region contributed a similarly level of sensitivity as the $1lep$ and $2lep$ regions, despite having a small fraction of their integrated luminosity. This is the first search for hadronically-decaying LLPs at the LHC to include the VBF topology.

This analysis also sets the first limits on long-lived ALP models with suppressed couplings to photons, cross sections greater than 0.1 pb for $qq \rightarrow Va$ are excluded for $1.2 < c\tau_a < 192 \text{ mm}$ and $m_a > 40 \text{ GeV}$, and long-lived ALPs produced in exotic top decays are probed for the first time. Branching ratios $BR(t \rightarrow aq) > 0.1\%$ are excluded between $4 < c\tau_a < 36 \text{ mm}$ for $m_a > 40 \text{ GeV}$.

An event display showing a signal-like event in the $2lep$ region containing two displaced vertices is shown in 5.11.

$BR(H \rightarrow ss \rightarrow 4u)$	Excluded range of $c\tau_s$ (mm)			
	$m_s = 55$ GeV	$m_s = 40$ GeV	$m_s = 16$ GeV	$m_s = 5$ GeV
1%	4.4-107.4	4.5-80.9	7.5-19.8	-
5%	1.6-398.6	1.4-329.2	1.8-104.7	5.6-8.5
10%	1.2-651.9	1.1-547.2	1.3-168.4	2.4-23.3
20%	0.9-1076.0	0.8-928.6	1.0-263.4	1.5-42.2

Table 5.17: The ranges of values of LLP proper decay length $c\tau_s$ excluded at the 95% confidence level for different values of $BR(H \rightarrow ss \rightarrow 4u)$, for LLP masses in the range 5-55 GeV.

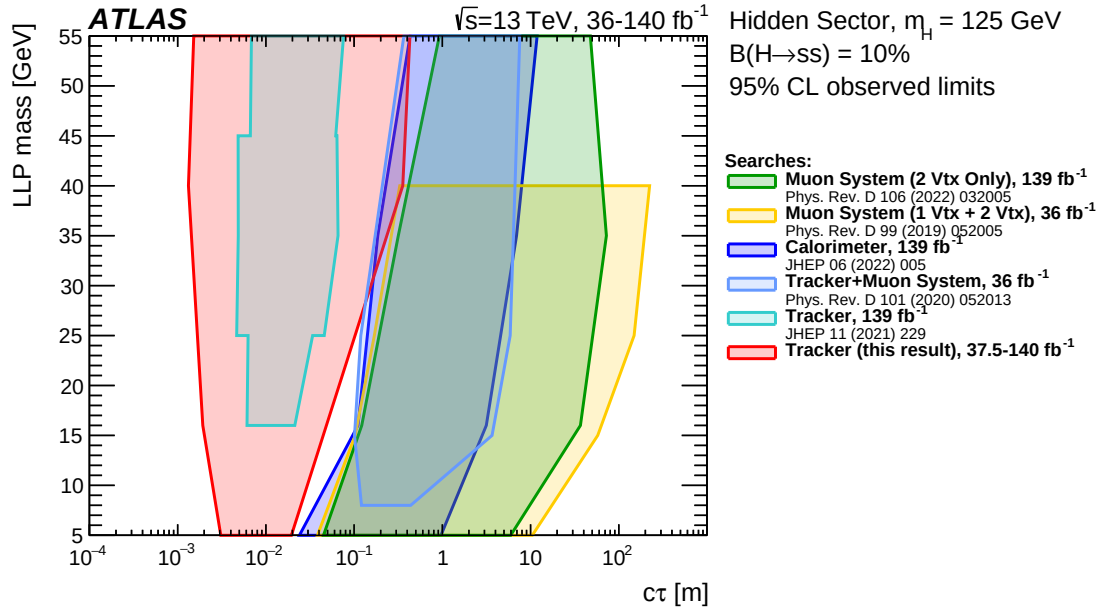


Figure 5.10: Regions in LLP mass versus proper decay length excluded at the 95% confidence level for the exotic Higgs decay model. The coloured lines represent the regions excluded by the different ATLAS searches listed in the legend. The figures shows the limits obtained assuming $BR(H \rightarrow ss) = 10\%$. This search (red) excludes a large region of phase space previously unexplored by ATLAS searches.

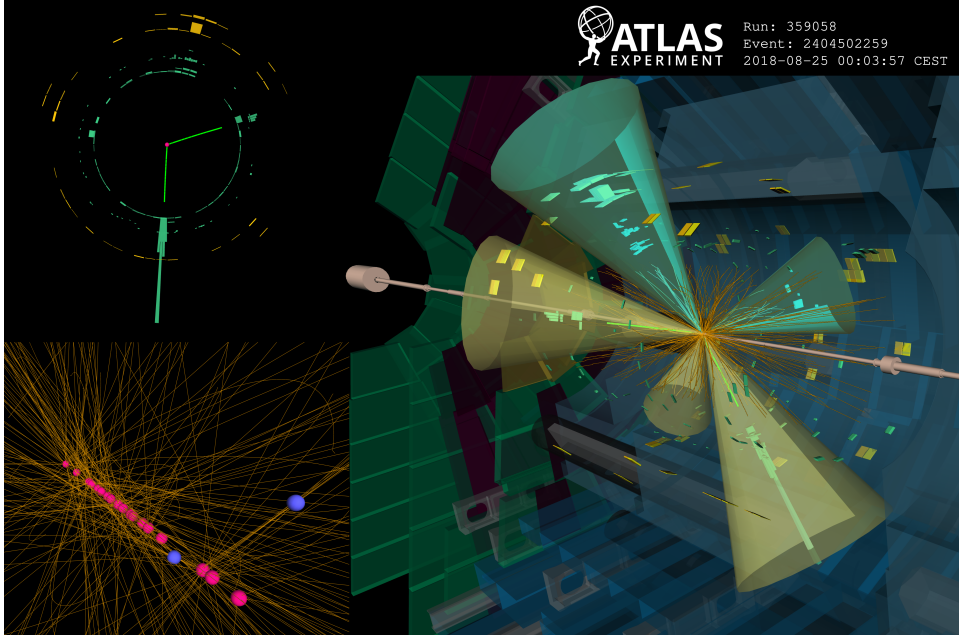


Figure 5.11: Event display showing a signal-like event in the $2lep$ region. The upper left portion shows the lepton tracks and calorimeter deposits from the Z decay. The bottom left portion shows the primary vertices (pink) and two tagged displaced vertices (purple). The right portion shows the jets in the event. Those tagged as displaced jets are coloured cyan [1].

5.9 CONCLUSION

A search for LLPs using 140 fb^{-1} of pp collision data at $\sqrt{s} = 13 \text{ TeV}$ recorded at the ATLAS experiment at the LHC between 2015 and 2018 is presented. The search targets LLPs with masses between 5 and 55 GeV that decay hadronically in the ID. The primary benchmark model considered by the analysis features LLP pair production through an exotic decay of the Higgs boson. Additionally, models that feature long-lived ALPs are included in the interpretation of the results. No excess of events over the background prediction is observed. Upper limits are set on $BR(H \rightarrow ss)$, $\sigma(q\bar{q} \rightarrow Va)$ for $V = Z, W$, and $BR(t \rightarrow aq)$. The limits reported in §5.8 for the exotic Higgs models are the most stringent to date for LLP masses $m_s < 40 \text{ GeV}$ and lifetimes 10^{-1} – 10^{-3} m . The constraints placed on long-lived ALP models are the first of their kind.

A further iteration of the search is planned using data recorded during Run 3 at ATLAS at $\sqrt{s} = 13.6 \text{ TeV}$. There are several potential ways to improve on the limits set by this analysis.

LRT has been introduced into the trigger reconstruction chain for the first time in

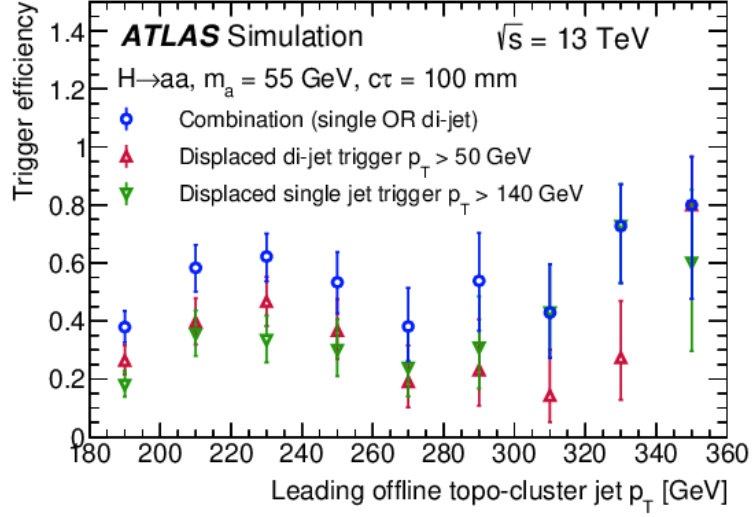


Figure 5.12: Efficiency of the LRT-based displaced jets triggers plotted against the leading offline jet p_T for a model of exotic decays of the Higgs boson into long-lived pseudoscalars with mass 55 GeV and proper decay length of 100 mm. Only statistical uncertainties are shown [2].

Run 3, allowing for the possibility to trigger directly on displaced objects, and circumventing the need to use lepton triggers to select events. Two new LRT-based displaced jet triggers have been introduced into the trigger menu that target hadronic LLP decays by selecting events with a jet from initial-state radiation and at least one or two displaced jets [2]. Figure 5.12 shows the expected performance of these triggers based on MC simulation (using Run 2 conditions) of the exotic Higgs model for an LLP with mass 55 GeV and proper decay length of 100 mm. Combined, the triggers are up to 80% efficient for the highest p_T jets. The triggers offer the potential to include the ggF production mode as an additional search channel in the analysis, as well as increase the signal selection efficiency of the existing channels, which would be expected to improve the overall sensitivity of future searches.

The dedicated displaced object triggers included in the trigger menu for Run 3 so far use ROI-based LRT. However, the potential benefit of triggers that run LRT over the full ID volume should be studied as a means to increase signal selection efficiency. Further increases in signal selection efficiency could be made possible by using more sophisticated techniques to discriminate between signal-like displaced jets and SM background at the analysis level. These two avenues form the basis of the studies presented in §6.

6

ADVANCES IN LONG-LIVED PARTICLE IDENTIFICATION

In §5.8, exclusion limits on $BR(H \rightarrow aa)$ set by the second iteration of a search for new, hadronically decaying LLPs are shown to vastly improve those set by the previous iteration, despite the two analyses using the same set of Run 2 data. This increase in sensitivity stems from developments in the ATLAS reconstruction chain, and the introduction of sophisticated analysis techniques that leverage machine learning to more effectively identify LLPs and reject background. As the LHC continues its third run of data taking, the next generation of ATLAS LLP searches will strive to improve even further on those limits set using Run 2 data. This necessitates further advances in reconstruction software and analysis techniques, such as the recent integration of LRT into the trigger system, and the shift towards GNNs as the preferred algorithm for jet flavour tagging at ATLAS.

This chapter includes two studies that explore the impact of these advances on future ATLAS LLP searches. In §6.1, results are presented that demonstrate the performance of LRT in the trigger using tracks associated to K_S^0 vertices, while §6.2 contains an investigation into the use of a GNN-based tagger to identify jets from hadronic decays of LLPs.

6.1 LARGE RADIUS TRACKING PERFORMANCE IN THE ATLAS TRIGGER

Many LLP signatures rely on precise tracking information to be detected, such as displaced vertices. For Run 2, only the standard tracking pass was run in the trigger, which, while effective up to $|d_0| \lesssim 5$ mm, is less viable for most LLP searches. Searches instead used calorimeter/MS information, or prompt leptons, as noted in §5.1.1, to trigger events. However, due to the reduced processing time and fake rate of the updated LRT algorithm (see §4.3.1), LRT has been added to the trigger in Run 3, making the option to trigger directly on displaced objects in the ID possible for the first time. The LRT configuration is based on that used offline.

LRT-based triggers for electrons, muons and jets have already been developed and validated [2]. These only run LRT in lepton ROIs. However, full scan¹ LRT in the trigger may be useful for signatures with no obvious ROI, as well as improve jet reconstruction. Full scan trigger LRT is therefore important for the development of displaced jet and vertex triggers, that may be included in the trigger menu at a later stage. Here, a study of the LRT performance in the trigger is performed using full scan tracks associated to offline K_S^0 vertices.

The lifetime of K_S^0 particles is compared with other SM particles in Figure 2.7. Due to their long lifetimes, K_S^0 decays are well suited for assessing the efficiency for reconstructing displaced tracks using LRT in the trigger. The most common K_S^0 decay mode is to two pions, $K_S^0 \rightarrow \pi^+ + \pi^-$, producing the signature of two tracks with opposite charge forming a secondary vertex.

6.1.1 Data and object reconstruction

This study uses data recorded during a single run in 2022 at $\sqrt{s} = 13.6$ TeV that was reprocessed to run the full scan trigger tracking configuration of LRT for every event. A collection of offline K_S^0 vertices is produced by selecting pairs of standard and LRT offline tracks with $p_T > 1$ GeV and opposite charge that form vertex candidates with mass within 25 MeV of 497 MeV [101]. The radius and mass of the K_S^0 vertex candidates

¹‘Full scan’ refers to tracking that is being run over the entire ID volume.

6.1. LARGE RADIUS TRACKING PERFORMANCE IN THE ATLAS TRIGGER

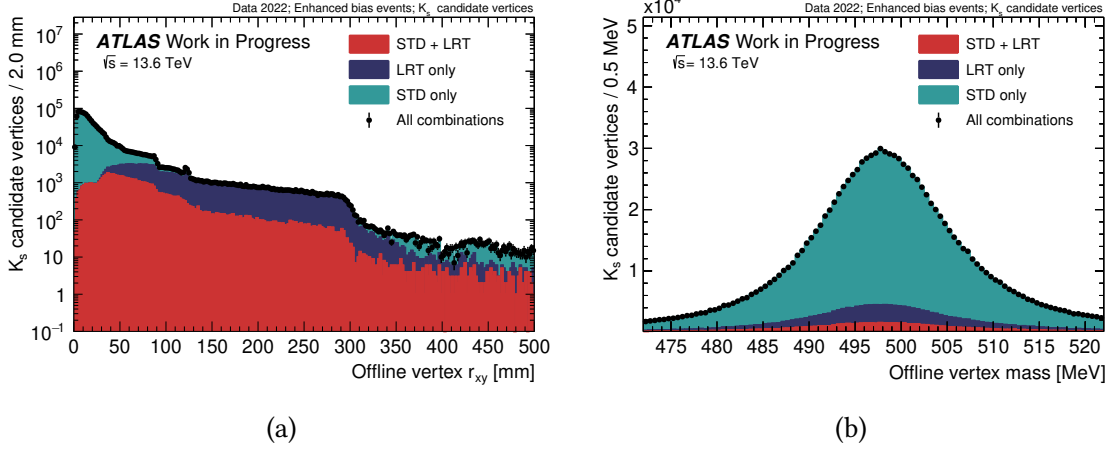


Figure 6.1: Radius (a) and mass (b) of offline K_S^0 candidates vertices. The stacked plot shows the track content of the vertex: two standard (STD) (teal), two LRT (blue), or a combination of both STD and LRT offline tracks (red). The black markers show the overall distribution.

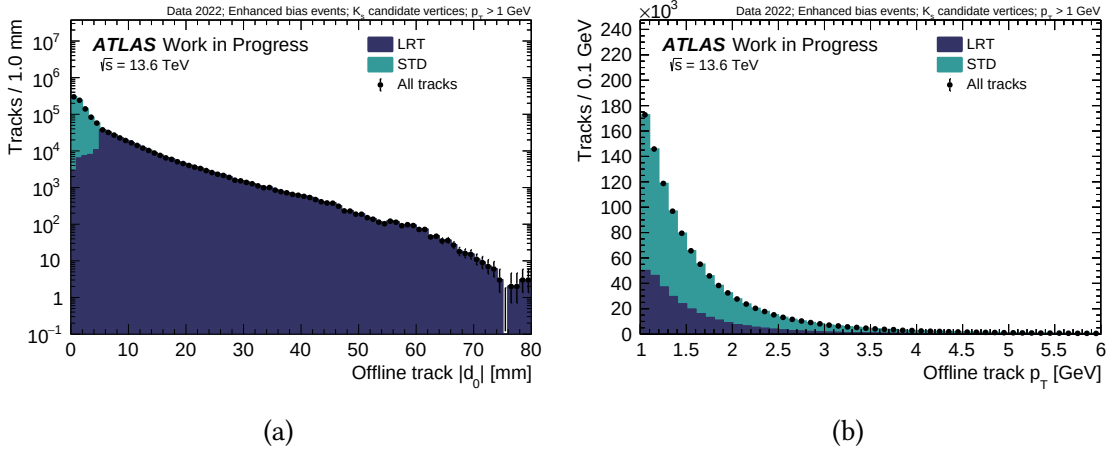


Figure 6.2: Absolute value of the transverse impact parameter $|d_0|$ (a) and p_T (b) of offline full scan tracks associated to K_S^0 candidates vertices. The stacked plot is divided into tracks reconstructed using the standard (STD) (teal) or LRT (blue) tracking passes. The black markers show the overall distribution.

are shown in Figure 6.1. At radii $\gtrsim 100$ mm, excluding > 300 mm where the vertices are likely combinatorial fakes, the majority of vertices are formed purely of LRT tracks. The absolute value of the transverse impact parameter $|d_0|$ with respect to the beamline and the p_T of offline full scan tracks associated to the K_S^0 candidate vertices are shown in Figure 6.2. Offline tracks from K_S^0 decays tend to have low p_T ; there are very few tracks with $p_T > 5$ GeV and $|d_0| > 80$ mm.

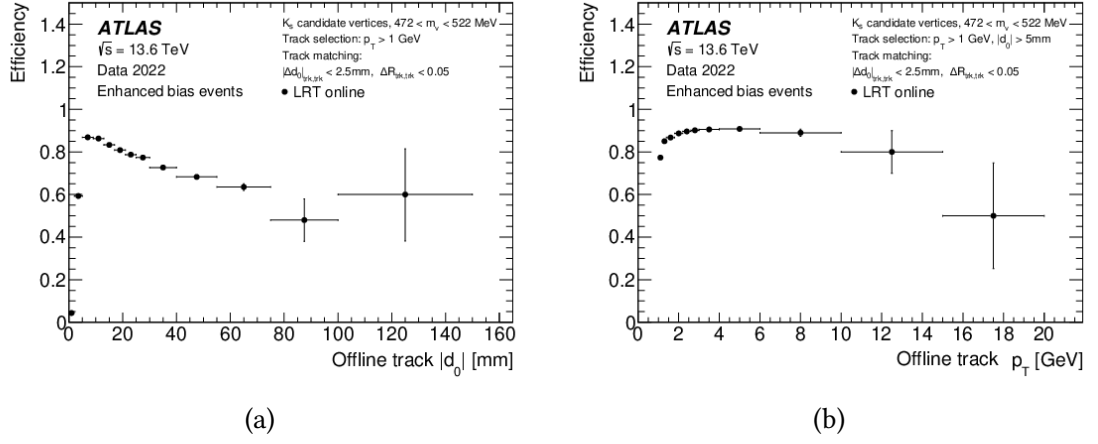


Figure 6.3: The efficiency of online full scan tracks reconstructed using LRT, with respect to offline tracks associated to K_S^0 vertices reconstructed as a secondary vertex using the offline tracks. The efficiency is shown as a function of the absolute value of the transverse impact parameter $|d_0|$ with respect to the beamline (a) and p_T (b) of the corresponding offline track [2].

6.1.2 Trigger large radius tracking efficiency

The efficiency to reconstruct tracks online using LRT is computed with respect to offline tracks associated to K_S^0 candidate vertices. First, offline tracks, standard or LRT, are matched to online standard full scan tracks if they fall within $\Delta R < 0.05$ and $\Delta d_0 < 2.5$ mm of each other. The remaining tracks are used in the denominator of the online LRT efficiency calculation. Those that are matched to online LRT full scan tracks, using the same criteria, form the numerator in the efficiency calculation. The trigger LRT efficiency can be formulated as

$$\epsilon_{\text{LRT, trigger}} = \frac{\text{offline} \wedge \text{not matched to online STD} \wedge \text{matched to online LRT}}{\text{offline} \wedge \text{not matched to online STD}}, \quad (6.1.1)$$

where STD and LRT refer to standard and LRT full scan tracks, respectively.

Figure 6.3 shows the trigger LRT efficiency as a function of offline track $|d_0|$ and p_T . When computing the efficiency as a function of p_T , there is an additional cut of $|d_0| > 5$ mm on all tracks, in order to remove tracks with low displacement, which are expected to be reconstructed by the standard tracking. For $|d_0|$ greater than 5 mm, where the standard tracking pass is less effective, up to 80 mm, beyond which the number of tracks significantly drops off, trigger LRT achieves efficiencies of $\sim 60\text{--}85\%$.

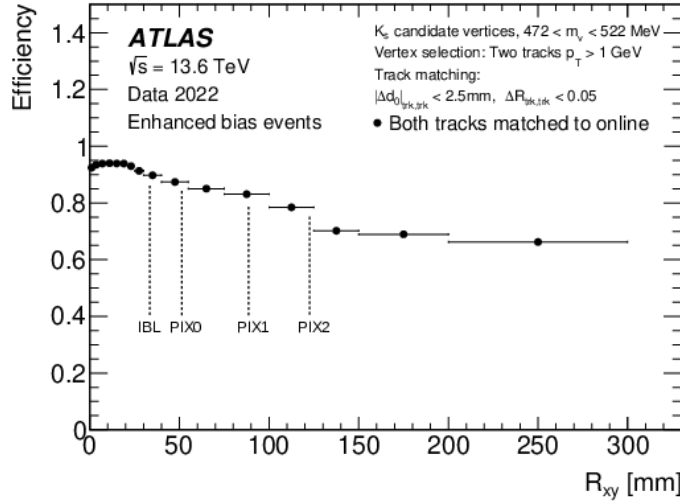


Figure 6.4: The efficiency to reconstruct a displaced secondary K_S^0 vertex from online full scan tracks, as a function of the decay radius of the vertex in the transverse plane r_{xy} . Dotted black lines indicate the mean radial positions of the insertable B-layer and other barrel layers in the pixel detector [2].

Figure 6.4 shows the efficiency to match both offline tracks used to reconstruct the vertex to online tracks, standard or LRT, versus the reconstructed radius r_{xy} of the K_S^0 vertex. Between r_{xy} of 100 mm and 300 mm, where vertices reconstructed purely from offline LRT tracks are dominant, secondary vertex reconstruction efficiency in the trigger is found to be over 60% and up to $\sim 80\%$. Particles produced in decays at larger radii leave fewer potential detector hits for track reconstruction, hence drops in efficiency are observed at radial positions corresponding to barrel layers in the pixel detector.

6.1.3 Conclusion

During Run 3, LRT has been integrated into the ATLAS trigger for the first time. The performance of full scan LRT in the trigger has been reported for early Run 3 data collected in 2022, by computing its efficiency using tracks associated to offline K_S^0 vertices as a reference. The LRT recovers tracking efficiency out to large $|d_0|$ and secondary vertexing efficiency out to large r_{xy} . This strong performance motivates the development and addition to the trigger menu of triggers for LLP signatures that rely on full scan tracking information. Such triggers would be valuable to many future LLP searches, including the next iteration of the analysis presented in §5.

6.2 GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

This section investigates algorithms based on GNNs for tagging jets from LLP decays. In §6.2.1, an overview of the strategy used is given. An introduction to the key concepts behind deep neural networks, GNNs and *parameterised neural networks* is given in §6.2.2. The datasets used are summarised in §6.2.3, followed by object definitions and selections in §6.2.4. A description of the models tested is given in §6.2.5. Results are presented in §6.2.7, before conclusions are drawn in §6.2.8.

6.2.1 Motivation

Since the introduction of GN1 [58], flavour tagging algorithms based on GNNs have become widely used at ATLAS. GN1 boasts greater performance than previous flavour tagging algorithms, such as DL1r, while being highly adaptable and easier to maintain. GN1 takes a variable number of tracks and jet information as inputs. The training incorporates two auxiliary tasks to guide the learning process, as well as enable predictions of track origin and spatial track-pair compatibility, in addition to the jet flavour. Training targets are extracted from MC truth information. The success of GN1 led to further optimisations and the development of GN2 [102], a transformer-based GNN. Here, the GN2 algorithm has been adapted to identify jets from hadronic decays of LLPs.

Instead of predicting the flavour of a jet, the task is now a binary signal-background classification, where the signal is a jet produced in the decay of a new LLP with an unknown mass, and the background are jets from SM processes. In the absence of a theoretical prediction on the exact mass of a new particle, analyses will typically generate MC samples for a range of mass hypotheses $\mathcal{M}$, and conduct a search at each fixed mass value. For each mass $m \in \mathcal{M}$, the classification problem will be related, but unique. Conventional approaches using machine learning might include the following:

- Train a single model using a single mass point and apply it to all masses. This approach, while the simplest, degrades performance at masses not used in training.
- Train a single model using an unlabelled mixture of all $m \in \mathcal{M}$ and apply it to

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

all masses. This model will exhibit more even performance across the mass range compared to the previous approach, but the performance at each mass will be suboptimal, due to smearing of the signal. This is the approach that was taken when developing the per-jet BDT used in §5.

- Train an independent set of individual models for each $m \in \mathcal{M}$. Each model will give the best performance at the mass with which they were trained, however, the performance will be discontinuous between the individual models. Additionally, this approach requires the analyst to train, optimise and maintain multiple models, increasing the management complexity, with each model using only a subset of the available data.

Parameterised neural networks (PNNs) [103] incorporate additional parameters, such as the generated signal mass, into the training of the model, conditioning its output to change with the value of these parameters. This strategy is claimed to solve many of the issues highlighted in the approaches outlined above. A single PNN can replace a set of individual classifiers, benefiting from shared weights and representations learned using the full parameter-dependent dataset. Also, since the additional parameters, e.g. the mass, represent a *smoothly varying learning task*, the model should be able to smoothly interpolate between different parameter values. In effect, the performance of a PNN should be close to or better than isolated networks for each $m \in \mathcal{M}$, while also generalising to mass values not used during training. In the following, three methods of parameterisation are implemented within GN2 and their performance is compared to baseline models without parameterisation.

6.2.2 Neural network theory

Deep neural networks

A deep neural network [104] consists of multiple layers of neurons. Information is passed into the input layer $\mathbf{x}$, and feeds forward through intermediate, or *hidden*, layers $\mathbf{h}$, via differentiable operations, that define a mapping to the output layer $\mathbf{y}$. A network in which each neuron is connected to every neuron in the subsequent layer is called a fully-connected, or *dense*, network, as shown in Figure 6.5.

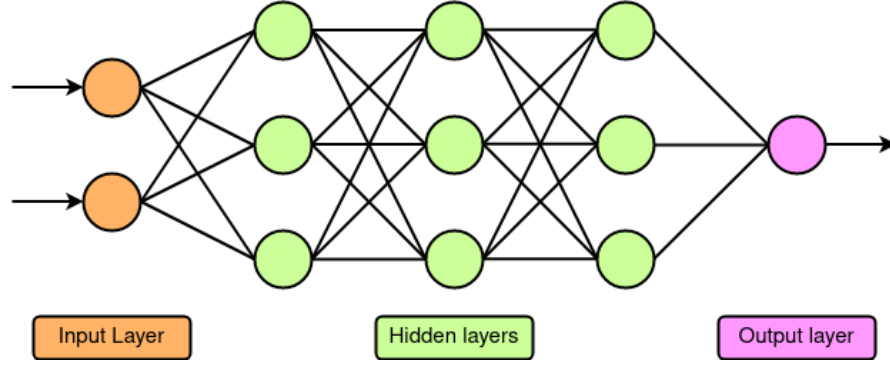


Figure 6.5: Schematic of a fully connected deep neural network with two inputs, three hidden layers with three neurons each, and a single output. Information flows from left to right. Black lines denote connections between neurons.

Each layer is parameterised by a set of weights ω and biases $\mathbf{b}$. The first hidden layer is given by

$$\mathbf{h}^{(1)} = g^{(1)}(\omega^{(1)\top} \mathbf{x} + \mathbf{b}^{(1)}) \quad (6.2.1)$$

where $g(\mathbf{z})$ is a non-linear activation function. Each subsequent layer is a function of the layer that preceded it. The output $\mathbf{y}$ of a model with n hidden layers is given by

$$\mathbf{y} = g^{(y)}(\omega^{(y)\top} \mathbf{h}^{(n)} + \mathbf{b}^{(y)}). \quad (6.2.2)$$

A training algorithm is used to optimise ω and $\mathbf{b}$. In supervised learning, each input $\mathbf{x}$ has a label $\mathbf{y}^*$, that specifies the intended output $\mathbf{y}$. For classification, $\mathbf{y}^*$ would denote the true class of a particular sample. During training, parameters are updated to minimise a loss function $\mathcal{L}(\mathbf{y}, \mathbf{y}^*; \theta)$, which quantifies the separation between $\mathbf{y}$ and $\mathbf{y}^*$, using gradient descent:

$$\theta^{n+1} = \theta^n - \alpha \frac{\partial \mathcal{L}(\mathbf{y}, \mathbf{y}^*; \theta^n)}{\partial \theta^n}, \quad (6.2.3)$$

where θ^n represents all trainable parameters at iteration n , and α is the *learning rate*. Each update of θ is computed using the average loss of a subset of the training data, called a *batch*. A complete pass through all the training data is called an *epoch*. Parameters are repeatedly updated until they are found to converge. Partial derivatives of the loss with respect to each parameter are efficiently calculated using the backpropagation algorithm

[105]. Large deep neural networks tend to use optimisations of the gradient descent method to increase training speed. This work uses the AdamW optimiser [106].

Model *hyperparameters* control the learning process, or the topology and size of the network, but are not optimised through gradient descent. They include the learning rate, batch size, number of layers, and number of neurons per layer.

Non-linear activation functions used in this work include: the Rectified Linear Unit (ReLU) [107], where

$$\text{ReLU}(x) = \max(0, x), \quad (6.2.4)$$

the sigmoid function,

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (6.2.5)$$

and the softmax function, where

$$\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^m e^{x_j}}, \quad (6.2.6)$$

where m is the size of vector $\mathbf{x}$.

Graph neural networks

GNNs are a more advanced type of neural network used to learn representations of graph structured data [108, 109, 110]. An overview of their applications within particle physics can be found in Ref. [111].

A typical graph consists of a set of nodes $\mathcal{V} = \{1, \dots, n\}$ and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, directed connections between two nodes, where $(j, i) \in \mathcal{E}$ is an edge from *sender* node j to *receiver* node i . The intuition underpinning graphs is that nodes represent objects, and edges represent their relationship. Nodes and edges may each be represented by real vectors, or *representations*. The representations of node i and edge (j, i) are denoted by $\mathbf{h}_i$ and $\mathbf{e}_{ji}$.

Each layer of a GNN updates every node representation by aggregating the representations of its connections. A layer takes as inputs the set of node representa-

tions $\{\mathbf{h}_i \mid i \in \mathcal{V}\}$ and edges $\mathcal{E}$, and outputs an updated set of node representations $\{\mathbf{h}'_i \mid i \in \mathcal{V}\}$. The outputs are the product of a parametric function applied to each node given its neighbours $\mathcal{N}_i = \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}\}$:

$$\mathbf{h}'_i = f_\theta(\mathbf{h}_i, \text{AGGREGATE}(\mathbf{h}_i, \{\mathbf{h}_j, \epsilon_{ji} \mid j \in \mathcal{N}_i\})). \quad (6.2.7)$$

The AGGREGATE function ensures the updated representations of each node are conditioned on the representations of its neighbours. The output of the final layer can be fed to downstream algorithms to aid in tasks such as classification. The distinguishing feature of a particular GNN is the choice of f_θ and AGGREGATE.

Parameterised neural networks

PNNs are neural networks that use auxiliary inputs to condition the output of the network. A typical neural network can be denoted as $f(\mathbf{x}; \boldsymbol{\theta})$, a function of its input $\mathbf{x}$ and a set of trainable parameters $\boldsymbol{\theta}$. In contrast, a PNN is also parameterised in terms of its conditioning representation, and can be denoted, for example, by $f(\mathbf{x}, m; \boldsymbol{\theta})$, when the output is conditioned on the mass m of a hypothetical particle.

The conditioning mechanism can take a number of forms. The canonical approach, as described by Baldi et al. in Ref. [103], and used in many ATLAS searches, see e.g. Refs. [112, 113] for recent examples, is to include m as an additional input to the model, by concatenating $\mathbf{x}$ with m before feeding the inputs into the first layer. This approach is parameter-efficient, only requiring additional weights equal to the number of neurons in the first hidden layer, but makes the assumption that the input is where the conditioning information is most valuable.

Alternatively, *feature-wise transformations*, such as *conditional biasing* and *conditional scaling*, as described by Dumoulin et al. in Ref. [114] and applied within high energy physics by Anzalone et al. in Ref. [115], use the conditioning representation, e.g. m , to bias or scale the features of a layer. This can be achieved using a separate dense network to map m to a bias or scaling vector $\mathbf{b}$ or $\mathbf{s}$, to which an element-wise addition or multiplication with the input features $\mathbf{x}$ is applied, outputting $\mathbf{x} \oplus \mathbf{b}$ or $\mathbf{x} \odot \mathbf{s}$, as shown in Figures 6.6 and 6.7. The dimension of the vectors $\mathbf{b}$ and $\mathbf{s}$ must match the dimension

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

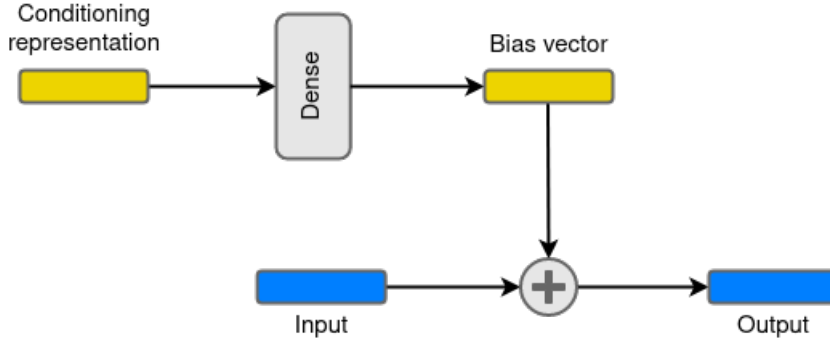


Figure 6.6: Conditional biasing: the conditioning representation is mapped to a bias vector that is added element-wise to the input features of a layer. The parameters, e.g. the mass of a new particle, form the conditioning representation.

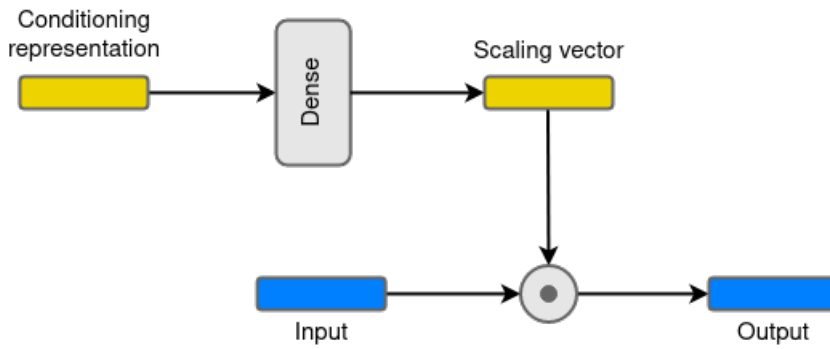


Figure 6.7: Conditional scaling: the conditioning representation is mapped to a scaling vector that is multiplied element-wise by the input features of a layer. The parameters, e.g. the mass of a new particle, form the conditioning representation.

of x . Interestingly, conditional biasing can be thought of as an alternate implementation of the concatenation-based approach described above [114].

A special type of conditional scaling is *sigmoidal gating*, in which the features are scaled by a value between 0 and 1, by first applying the sigmoid function (defined in Eq. (6.2.5)) element wise to the scaling vector. Intuitively, this allows the conditioning information to select the most important features to pass through the model, while zeroing out others.

Conditional biasing and conditional scaling can be combined into a *conditional affine transformation*,

$$\text{Affine}(x) = x \odot s(m) \oplus b(m), \quad (6.2.8)$$

where the dependence on s and b on the conditioning representation m is made explicit. This mechanism is shown in Figure 6.8.

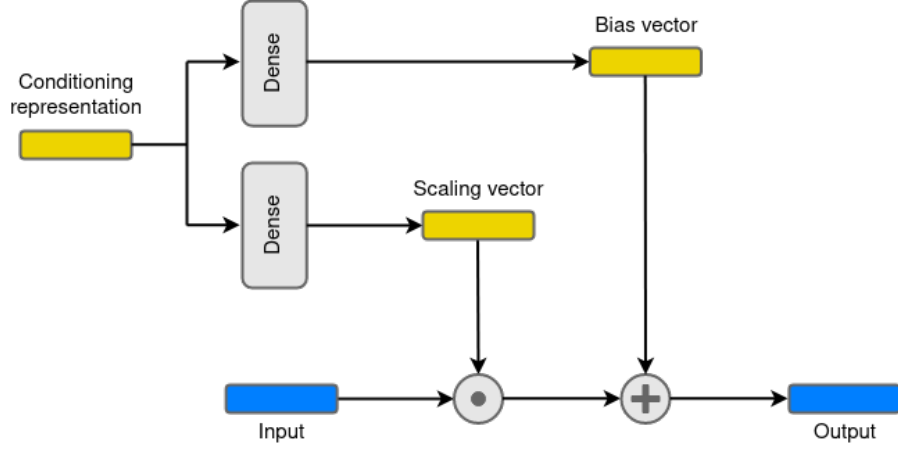


Figure 6.8: Conditional affine transformation: the conditioning representation is mapped to both a bias and scaling vector, the inputs are first multiplied element-wise by scaling vector, then the result is added element wise to the bias vector. The parameters, e.g. the mass of a new particle, form the conditioning representation.

Feature-wise transformations can be applied to the features of any layer in the model, allowing the analyser to find the combination of transformations that best suit the task, without needing to make assumptions about where the conditioning information is most valuable. Furthermore, the option to include both additive and multiplicative transformations removes assumptions on how best to leverage the additional parameters. However, feature-wise transformations can be more memory intensive and computationally expensive than the canonical concatenation-based approach, as they require storing and optimising the parameters of a number of additional dense networks.

6.2.3 Datasets

The datasets used to train the GNN tagger include a subset of those described §5.2.1 in addition to samples generated at new mass points. The samples used are noted in Appendix §A. Only simulated samples from the MC20e campaign are used. Signal samples with masses $\mathcal{M} = \{5, 7, 10, 16, 25, 35, 40, 55\}$ GeV are used. Background jets are taken from $t\bar{t}$, $V + \text{jets}$, and QCD dijet samples. Statistically independent sets of training, validation and test data are formed from the signal and background samples.

Training sets contain an equal number of signal and background jets. The signal component is comprised of an approximately equally represented mixture of mass points. Each mass point consists of a non-uniform mixture of jets from samples generated with varying Higgs production modes (ZH , WH , VBF, ggF), LLP proper lifetime (10 or 100 mm),

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

and LLP decay mode ($a \rightarrow uu, bb, cc$). The background component is comprised of an equal mixture of jets from $t\bar{t}$, $Z + \text{jets}$, $W + \text{jets}$ and QCD dijet samples. The $Z + \text{jets}$ component contains an equal mixture of $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events. The $W + \text{jets}$ component contains an equal mixture of $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ events. Both $V + \text{jets}$ components contain uniform mixtures of b -, c -, and light-jets. Jets are kinematically re-sampled in p_T (in the range $20 < p_T < 300$ GeV) and η (in the range $-2.5 < \eta < 2.5$) from each background sample to ensure similar distributions in these variables between signal and background. The resampling proceeds by estimating the number of background jets required in each (p_T, η) bin to produce a background sample of a certain size with a (p_T, η) distribution that approximates the signal. In each (p_T, η) bin, jets are sampled from the background until the required number is reached, or until no more background jets remain. In the latter case, the rest of the jets are chosen at random from this bin with replacement, until the required number is reached. As a result of the resampling, jets may be repeated in the background component of each training set, however the effect is negligible here.

The training set $\mathcal{D}_{\mathcal{M}}$ uses all $m_a \in \mathcal{M}$. Interpolation performance to unseen masses is accessed by training models using training sets that exclude certain masses $m_a \in \mathcal{M}$, then testing the models using data generated at the excluded mass points. Five such training sets are used, $\mathcal{D}_{\mathcal{M} \setminus 7}$, $\mathcal{D}_{\mathcal{M} \setminus 10}$, $\mathcal{D}_{\mathcal{M} \setminus 25}$, $\mathcal{D}_{\mathcal{M} \setminus 40}$, which exclude the 7, 10, 25 and 40 GeV samples, respectively, and $\mathcal{D}_{\mathcal{M} \setminus 7, 10, 25, 40}$, which exclude samples from all four of these masses. Lastly, four training sets contain only a single mass point. These are $\mathcal{D}_7$, $\mathcal{D}_{10}$, $\mathcal{D}_{25}$, and $\mathcal{D}_{40}$, which contain only 7, 10, 25 and 40 GeV samples, respectively. A summary of the composition of the training sets used is given in Table 6.1. Each training set has a corresponding validation set with an equivalent signal and background composition, to evaluate the performance of the model during training. To test the performance of each trained model and produce performance plots, test sets for each $m_a \in \mathcal{M}$ containing $\sim 200,000$ jets, and for each background containing 3 million jets, are used. In these test sets, the jets lie in the same p_T range as the training sets.

Dataset	Mass points (GeV)	Jets per mass point	Signal jets	Background jets	Total jets
$\mathcal{D}_{\mathcal{M}}$	5, 7, 10, 16, 25, 35, 40, 55		14.4 m	14.4 m	28.8 m
$\mathcal{D}_{\mathcal{M}\setminus 7}$	5, 10, 16, 25, 35, 40, 55				
$\mathcal{D}_{\mathcal{M}\setminus 10}$	5, 7, 16, 25, 35, 40, 55	~ 1.8 m	12.6 m	12.6 m	25.2 m
$\mathcal{D}_{\mathcal{M}\setminus 25}$	5, 7, 10, 16, 35, 40, 55				
$\mathcal{D}_{\mathcal{M}\setminus 40}$	5, 7, 10, 16, 25, 35, 55				
$\mathcal{D}_{\mathcal{M}\setminus 7,10,25,40}$	5, 16, 35, 55		7.2 m	7.2 m	14.4 m
$\mathcal{D}_7$	7				
$\mathcal{D}_{10}$	10	1.8 m	1.8 m	1.8 m	3.6 m
$\mathcal{D}_{25}$	25				
$\mathcal{D}_{40}$	40				

Table 6.1: Summary of the different training sets used in this work. The background jet component of each dataset is split equally between jets from $t\bar{t}$, Z + jets, W + jets, and QCD dijet samples.

6.2.4 Object selections

This work uses jets from the *EMTopo* jet collection. Details on jet reconstruction can be found in §4.2.1. Jets are required to have $p_T > 20$ GeV, $|\eta| < 2.5$ and be associated to at least one track. Background jets are sampled from background MC and given a label of 0. Signal jets are sampled from signal MC and selected by cutting on the *displaced p_T fraction*, defined per-jet as the ratio of the sum of the p_T of truth particles descending from the decay of the truth LLP that are matched ($\Delta R < 0.4$) to the jet, to the reconstructed p_T of the jet. Signal jets require a displaced p_T fraction greater than 0.35. This reduces contamination in the signal sample from pile-up jets that happen to be in the vicinity of LLP descendent particles. Signal jets are given a label of 1.

Tracks from both standard and LRT collections are associated to jets via ghost-association, where tracks are given negligible momentum and clustered within jets after their formation [116]. These tracks pass the selections given in Table 4.2, with some additional constraints. All tracks are required to have $|\eta| < 2.5$, at most one hole in the pixel layers, and at most two holes and one shared hit across all silicon layers. LRT tracks are required to have $|z_0| < 300$ mm.

Jet kinematics

In each Figure in this section, jet variables are plotted for signal jets, passing the selection stated above, in samples with $m_a = 5, 25, 55$ GeV and $c\tau_a = 10, 100$ mm, formed from

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

either $aa \rightarrow 4u$ or $aa \rightarrow 4b$ (or $aa \rightarrow 4c$ for $m_a = 5$ GeV) decays, along with jets from the combined SM background. These masses are chosen as they span the whole mass range.

Figure 6.9 shows jet p_T . Jets from decays of lighter LLPs have a higher p_T , since less energy is required to produce the LLP mass in the exotic Higgs decay. Light-flavour jets have a higher p_T for similar reasons. Signal jets tend to have a higher p_T than those produced in background SM processes.

Figure 6.10 shows the jet displaced p_T fraction. The distributions are centred close to, but slightly less than 1, owing to inefficiencies in the reconstruction of jet p_T . The distribution is sharper for jets produced in decays of lighter LLPs. These jets tend to have a higher p_T and are therefore reconstructed with higher precision. Jets produced at longer lifetimes tend to have a lower displaced p_T fraction, as jet p_T reconstruction becomes less efficient at larger radii.

Figure 6.11 shows the number of tracks associated to the jets. Jets produced from decays of heavier LLPs have more associated tracks than those from decays of lighter LLPs, but these tracks will tend to have a lower p_T , as shown in Figure 6.12, which contains the p_T distribution of all tracks associated to jets. A higher mass object is able to decay to a greater number of particles while conserving energy, producing more tracks, but the initial p_T of the particle will be lower, hence so will the p_T of the resulting tracks. Jets produced at longer lifetimes have fewer tracks as track reconstruction efficiency decreases at larger radii; this also explains why background jets tend to have more associated tracks than signal jets.

6.2.5 Model specification

The models used in this work are adapted from GN2, a transformer-based flavour-tagging algorithm. Readers are encouraged to seek Refs. [58] and [117] for a deeper look at the motivation behind the core architecture of the model.

Model inputs

Each model uses as inputs two jet kinematic variables (p_T and η), and a set of up to 40 tracks associated to the jet. If more than 40 tracks are associated to the jet, the 40 tracks

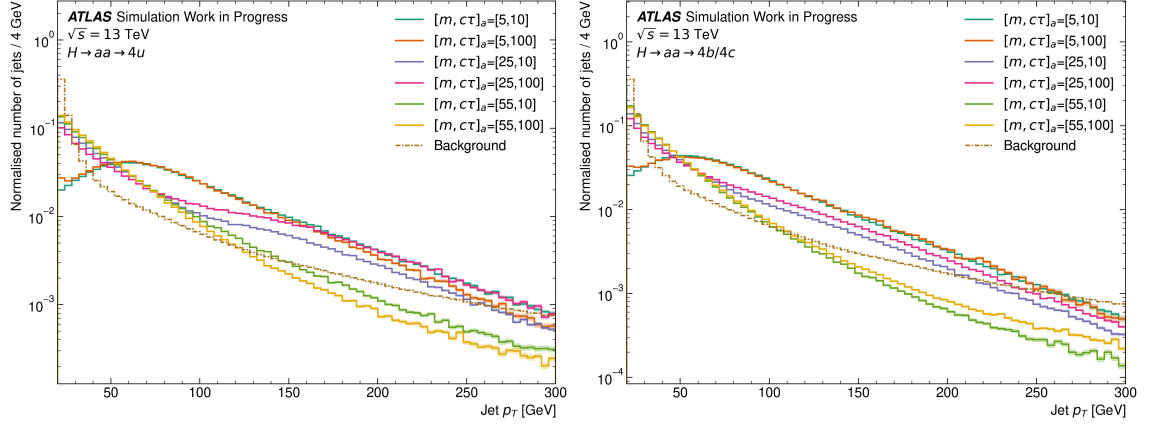


Figure 6.9: Jet p_T for jets labelled as signal from $aa \rightarrow 4u$ (left) and $a \rightarrow 4b$ ($aa \rightarrow 4c$ for $m_a = 5$ GeV) (right) decays. The dotted lines show the background jet p_T distribution, formed from a combination of $t\bar{t}$, Z +jets, W +jets and dijet samples. The distributions are normalised to unit area.

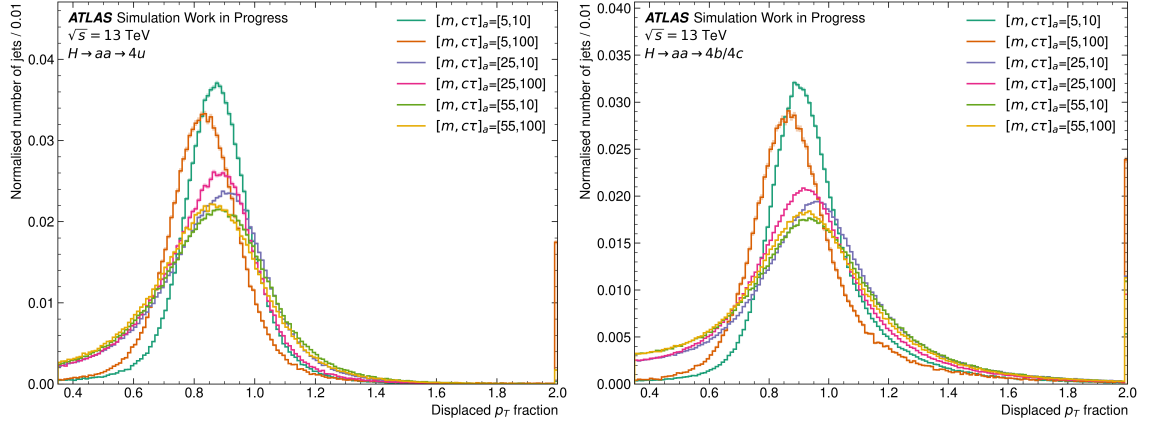


Figure 6.10: Displaced p_T fraction (defined as the ratio of the sum of the p_T of truth particles descending from the decay of the truth LLP that are matched ($\Delta R < 0.4$) to the jet, to the reconstructed p_T of the jet) for jets labelled as signal from $aa \rightarrow 4u$ (left) and $a \rightarrow 4b$ decays ($aa \rightarrow 4c$ for $m_a = 5$ GeV) (right). Signal jets are required to have displaced p_T fraction greater than 0.35. The final bins contain the overflow. The distributions are normalised to unit area.

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

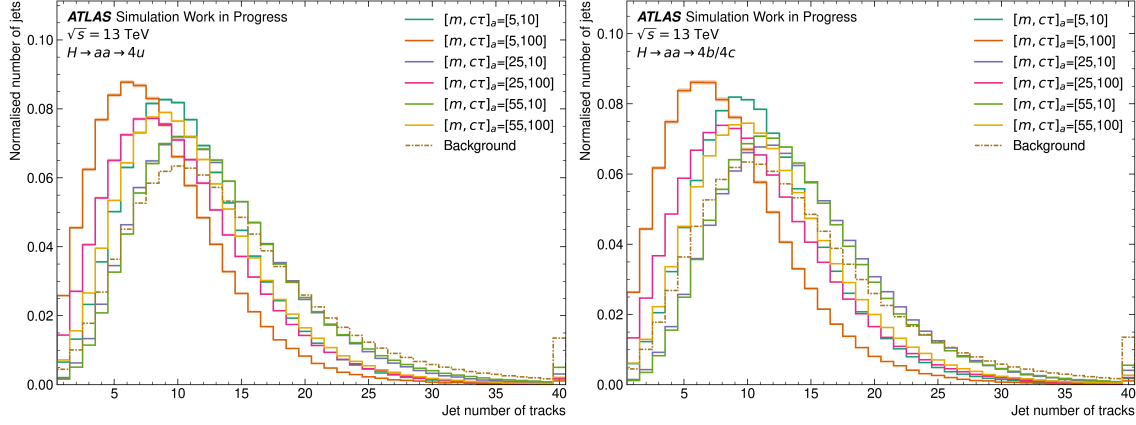


Figure 6.11: Number of tracks associated to jets labelled as signal from $aa \rightarrow 4u$ (left) and $a \rightarrow 4b$ ($aa \rightarrow 4c$ for $m_a = 5$ GeV) (right) decays. The dotted lines show the background jet p_T distribution, formed from a combination of $t\bar{t}$, Z +jets, W +jets and dijet samples. The final bins contain the overflow. The distributions are normalised to unit area.

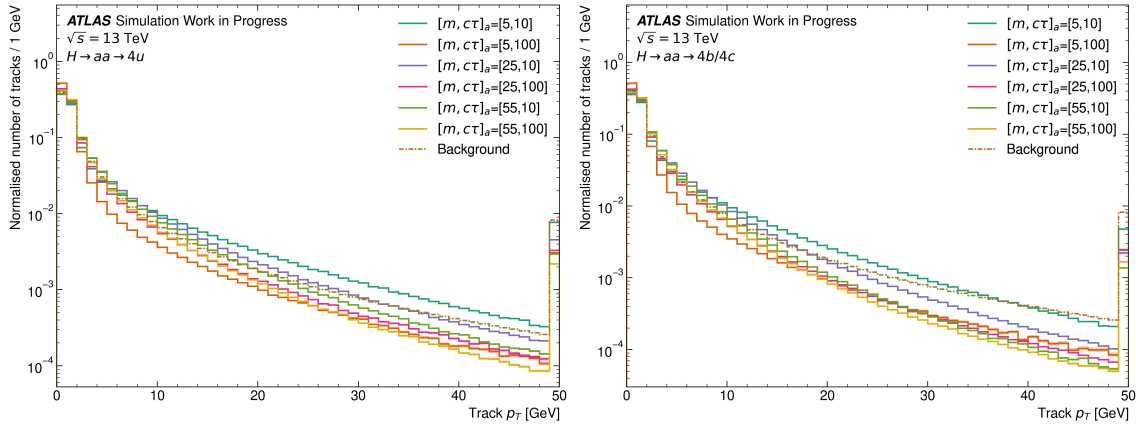


Figure 6.12: Distributions of track p_T for tracks associated to jets labelled as signal from $aa \rightarrow 4u$ (left) and $aa \rightarrow 4b$ ($aa \rightarrow 4c$ for $m_a = 5$ GeV) (right) decays. The dotted lines show track p_T distribution for background jets, formed from a combination of $t\bar{t}$, Z +jets, W +jets and dijet samples. The final bins contain the overflow. The distributions are normalised to unit area.

Jet input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track input	Description
d_0	Distance of closest approach between the track and primary vertex in the transverse plane
$z_0 \sin(\theta)$	Distance of closest approach between the track and primary vertex in the longitudinal plane, multiplied by the sin of the polar angle of the track relative to the jet axis
$d\phi$	Azimuthal angle of the track momentum relative to the jet axis
$d\eta$	Pseudorapidity of the track momentum relative to the jet axis
q/p	Ratio of the charge of the track to the magnitude of its momentum
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(q/p)$	Uncertainty on q/p
$s(d_0)$	Signed d_0 significance
$s(z_0)$	Signed z_0 significance
nIBLHits	Number of innermost pixel layer (IBL) hits
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nPixShared	Number of shared pixel hits
nSCTShared	Number of shared SCT hits

Table 6.2: Summary of the jet and track parameters used as input features to the GNN.

with the highest transverse impact parameter significance, defined as $s(d_0) = \frac{d_0}{\sigma(d_0)}$, are used. Each track is specified by the 15 variables listed in Table 6.2, which include some of it defining parameters and their uncertainties, as well as information on the hit content of the track. The azimuthal angle and pseudorapidity of the tracks is computed relative to the jet axis, to remove the dependence of the model on the azimuthal angle of the jet. Each of the jet and tracks variables are normalised to have a mean of 0 and a variance of 1 before being passed through the model.

Auxiliary tasks

Two auxiliary objectives are incorporated into the model: one to predict the physical process that produced each track (track origin), and one to predict the track-pair vertex compatibility. Their inclusion has been shown to improve the performance of the primary jet classification task [117].

For the track origin prediction, each track is classed exclusively using the categories listed in Table 6.3. One-hot encoding is used to produce labels for each track during training. In simulation, tracks are classified using a *truth matching probability* [49], computed by dividing a weighted sum of the hits on a reconstructed track that are matched to a given truth particle, by a weighted sum of all hits on the track. Tracks with a truth matching probability less than 0.5 are labelled as fakes. The categories in Table 6.3 are adapted from the default set used at ATLAS to suit the needs of this project. Tracks originating from the decay of a b -hadron, c -hadron, or a c -hadron that was itself from a b -hadron decay, are grouped into a single class. An additional class is also used containing all tracks matched to truth particles that descend from the decay of an LLP, regardless of how they might otherwise have been classed. The presence of such a track in a jet would be a strong indicator that the jet is itself from the decay of an LLP, highlighting the potential benefit of including the track origin prediction in the training. Figure 6.13 shows the total p_T of tracks descending from LLP decays as a fraction of the total p_T of all the tracks associated to the jet, $\sum p_T^{\text{track, LLP}} / \sum p_T^{\text{track}}$, for light-flavour jets labelled as signal in samples with $m_a = 5, 25, 55$ GeV and $c\tau_a = 10, 100$ mm. This fraction is generally larger for jets produced at shorter lifetimes, as track reconstruction efficiency decreases away from the beamspot. If the decay of the LLP occurs outside the ID, then no tracks will be reconstructed at all.

The vertexing task predicts whether two tracks belong to the same vertex. Each track pair in a jet is given a binary label, where the label has a value of 1 if the truth particles associated to the tracks originated at the same spatial point, and a value of 0 otherwise. To derive the labels, truth vertices within 1 mm of each other are merged. Any pair of tracks that includes a track without an associated truth particle, e.g. fakes, are given a label of 0. The presence of displaced secondary and tertiary vertices is indicative of,

Track truth origin	Description
Pile-up	From a pp -collision other than the primary interaction
Fake	Created combinatorially using hits from multiple particles
Primary	Originating from a primary, as oppose to secondary, decay
FromHF	From the decay of a b - or c -hadron
FromTau	From the decay of a τ
FromLLP	Descending from the decay of an LLP
OtherSecondary	From other secondary interactions and decays

Table 6.3: Exclusive classes used to categorise tracks for the track origin prediction task. LLP refers to a hypothetical BSM long-lived particle. “FromLLP” tracks are classed as such, regardless of how they might otherwise be classed. Examples of “Other Secondary” tracks include those from photon conversions, K_S^0 decays, and hadronic interactions.

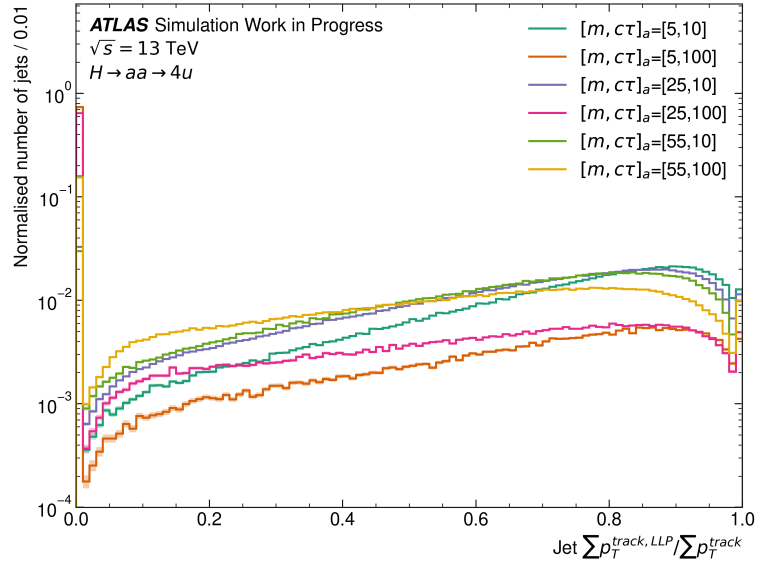


Figure 6.13: The total p_T of tracks descending from LLP decays as a fraction of the total p_T of all the tracks associated to light-flavour jets. The distribution is normalised to unit area.

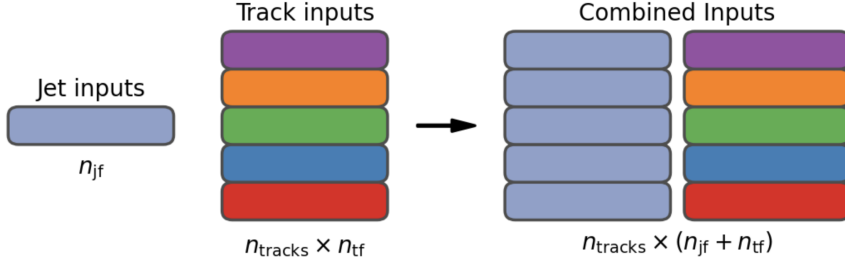


Figure 6.14: The inputs to the baseline model are $n_{\text{jf}} = 2$ jet features and $n_{\text{tf}} = 15$ track features. The jet features are copied and concatenated with the features of each track, giving combined jet-track feature vectors of size 17 [58].

although not unique to, a jet produced by a hadronically decaying LLP.

The choice of auxiliary tasks is strongly motivated by the intuition behind graphical models. A jet containing constituent tracks related to each other through shared production vertices can be seen as analogous to a graph formed of nodes connected to each other via edges. The two auxiliary tasks are therefore designed to guide the GNN to learn representations of the jet that reflect its underlying structure, facilitating the overall classification of the jet.

Baseline architecture

The baseline GN2 architecture takes as inputs information about the jet and a number of its associated tracks, d_{trk} . The jet variables are concatenated with the variables for each track, as shown in Figure 6.14, before being fed into a per-track initialisation network with a single ReLU activated hidden layer of size 256, and an output of size $d_{\text{embed}} = 256$, as shown in Figure 6.15.

The per-track outputs of the initialisation network form the initial representations $\{\mathbf{h}_i \mid i \in \mathcal{V}\}$ of each node in a fully connected graph, one in which every node in $\mathcal{V}$ is connected to every other node (including itself). Nodes represent tracks in the jet. Each layer of the graph network, or *encoder*, consists of two sub-layers: a multi-head attention mechanism with $n_{\text{heads}} = 8$, as described in Ref. [118], and a dense network (this architecture is known as a *transformer*), and produces a new set of node representations $\{\mathbf{h}'_i \mid i \in \mathcal{V}\}$, by aggregating each $\mathbf{h}_i$ with its neighbours $\mathcal{N}_i$. The encoder consists of four such layers. The following operations define a single layer.

First, defining $\mathbf{H} \in \mathbb{R}^{d_{\text{trk}} \times d_{\text{embed}}}$ as the matrix containing all node (track) representa-

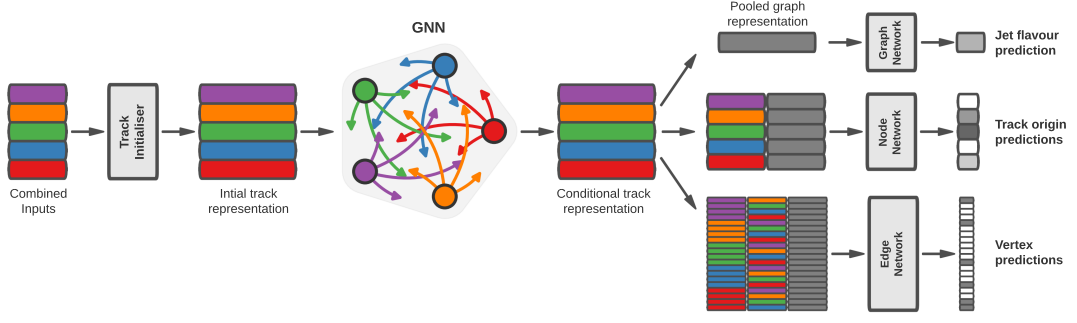


Figure 6.15: The baseline network architecture. The inputs are fed into a per-track initialisation network that outputs the initial track representations. These representations populate the nodes of a fully-connected graph network. The resulting conditional track representations are used as inputs to classifiers that predict the jet flavour, track origin, and track-pair vertex compatibility [58].

tions, three linear projections, parameterised by the trainable weight matrices $\mathbf{W}_R$, $\mathbf{W}_S$, $\mathbf{W}_A \in \mathbb{R}^{d_{\text{embed}} \times d_{\text{embed}}}$, are applied to $\mathbf{H}$, followed by a reshaping to add the head dimension, to give the matrices $\mathbf{R}, \mathbf{S}, \mathbf{A} \in \mathbb{R}^{d_{\text{trk}} \times d_{\text{embed}}/n_{\text{heads}} \times n_{\text{heads}}}$. Each head h in the multi-head mechanism then evaluates the scaled dot-product attention, as shown in Figure 6.16, over every pair of sender and receiver nodes in the projections, outputting the matrix

$$\text{head}_h(\mathbf{R}, \mathbf{S}, \mathbf{A}) = \text{softmax} \left(\frac{\mathbf{R}^h \mathbf{S}^{h\top}}{\sqrt{d_{\text{embed}}}} \right) \mathbf{A}^h, \quad (6.2.9)$$

where $\mathbf{R}^h, \mathbf{S}^h, \mathbf{A}^h \in \mathbb{R}^{d_{\text{trk}} \times d_{\text{embed}}/n_{\text{heads}}}$ are the h -th slices of $\mathbf{R}, \mathbf{S}, \mathbf{A}$, respectively, through their third dimension. The input to the softmax function can be interpreted as a matrix of edge scores. The edge scores indicate the importance of the projected features of one node on another node. Following the softmax normalisation, the edge scores for each track will lie in the interval $[0, 1]$ and will sum to 1, meaning they can be interpreted as probabilities. The outputs of each head $_h$, which have dimension $d_{\text{trk}} \times d_{\text{embed}}/n_{\text{heads}}$, are concatenated along their second dimension and fed into a dense layer parameterised by matrix $\mathbf{O} \in \mathbb{R}^{d_{\text{embed}} \times d_{\text{embed}}}$ to give the multi-head output,

$$\text{Multi-head}(\mathbf{R}, \mathbf{S}, \mathbf{A}) = \text{Concat}(\text{head}_1(\mathbf{R}, \mathbf{S}, \mathbf{A}), \dots, \text{head}_{n_{\text{heads}}}(\mathbf{R}, \mathbf{S}, \mathbf{A}))\mathbf{O}. \quad (6.2.10)$$

The multi-head attention mechanism is shown in Figure 6.17. The output of the

Scaled dot-product attention

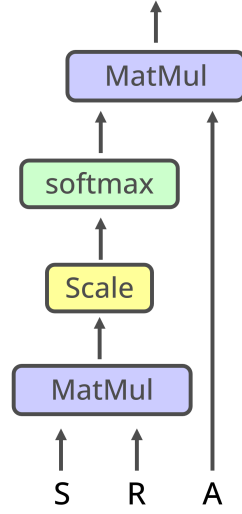


Figure 6.16: Scaled dot-product attention: the input sender S and receiver R matrices are multiplied, scaled, then activated using the softmax function. The result is multiplied by an additional matrix of weights A .

multi-head attention mechanism can be seen as an implementation of the AGGREGATE function, using the rubric discussed in §6.2.2.

The matrix of updated track representations $\mathbf{H}'$ outputted by the layer are computed using a residual connection [119] between each sub-layer, such that

$$\mathbf{H}' = \mathbf{L} + \text{Dense}(\mathbf{L})$$

where $\mathbf{L} = \text{LayerNorm}(\mathbf{H} + \text{Multi-head}(\mathbf{H}))$, (6.2.11)

where LayerNorm is a layer normalisation [120]. Additional layer normalisations are applied to the node representations before they are passed into an encoder layer, and to the output of the final encoder layer. Layer normalisation computes the mean and variance used for normalisation across all the features in a layer i.e. across all 256 features of a single encoded track representation in this instance. It is known to stabilise training and improve generalisation.

The output of the encoder network is a set of per-track representations, or *embeddings*, that are conditional on the representations of all tracks in the jet. To perform the jet classification, the track representations are pooled using a weighted sum, where the

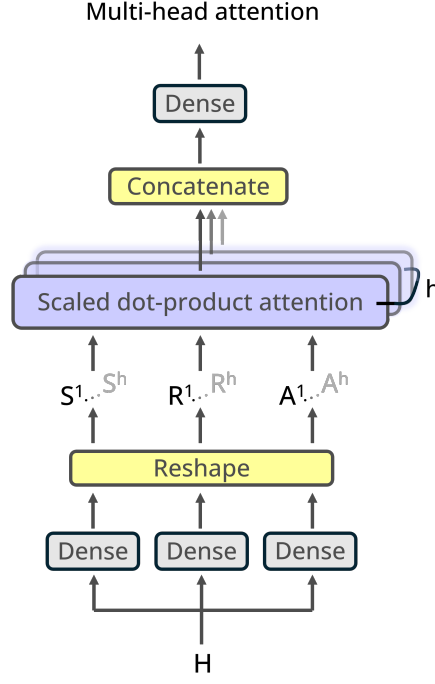


Figure 6.17: Multi-head attention: The track representations are fed into three linear projections whose outputs are reshaped to add an additional dimension for the h heads. Each head evaluates the scaled dot-product attention using its corresponding projections. The results are then concatenated, to recover the initial input shape, and fed into a final dense layer.

weights are learned during training, to produce a flattened global jet representation of size $d_{\text{global}} = 128$. This is fed into the *graph classification network*, which has two outputs, corresponding to background and signal jets, to which a softmax activation is applied, converting the outputs into probabilities. The output corresponding to signal is used as a discriminant throughout this section, and ranges between 0 and 1, with values close to 1 representing a more signal-like jet.

The *node classification network* predicts the track truth origin, as defined in Table 6.3. The network takes as input the conditional representation of a single track as outputted by the graph network, concatenated with the global jet representation. The network has seven outputs, corresponding to the different track classes.

The *edge classification network* predicts the vertex compatibility of each track-pair. The network takes as input the conditional representations of two tracks, concatenated with the global jet representation. The network has a single output, which is used to perform the binary classification of the track-pair compatibility.

Each classification network has three ReLU activated hidden layers containing 128,

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

Network	Prediction	Hidden layers	Output size
Graph classification network	Jet origin		2
Node classification network	Track origin	128, 64, 32	7
Edge classification network	Track-pair compatibility		1

Table 6.4: Summary of networks used to perform each classification task. The hidden layers column specifies the number of neurons in each hidden layer. All hidden layers are ReLU activated.

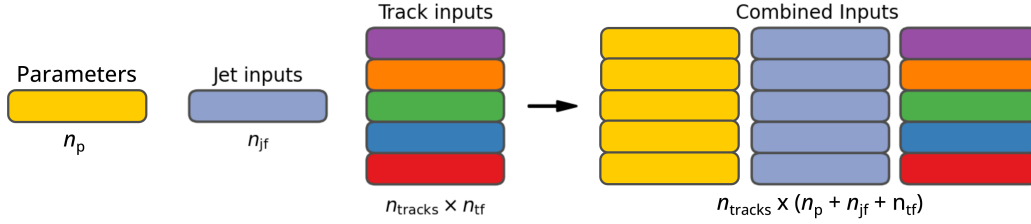


Figure 6.18: The inputs to the *Concatenation* model, that uses concatenation-based parameterisation, are $n_p = 1$ parameters, $n_{jf} = 2$ jet features, and $n_{\text{tf}} = 15$ track features. The jet features and parameters are copied and concatenated with the features of each track, giving combined parameter-jet-track feature vectors of size 18. Adapted from Ref. [58].

64 and 32 neurons, respectively. A softmax function is applied to each of their outputs, to obtain probabilities for each class. A summary of the networks is given in Table 6.4.

This architecture is referred to in the following as the *Baseline* model.

Parameterisation

Parameterisation of GN2 is achieved by implementing the conditioning mechanisms described in §6.2.2. The models are conditioned on the value of a single parameter, the mass m_a of a hypothetical, hadronically decaying, long-lived particle, as motivated in §2.3.2.

A model utilising concatenation-based parameterisation uses the baseline GN2 architecture, but with m_a concatenated to the input variables for each track. The resulting combined inputs, shown in Figure 6.18, consist of the track features concatenated with copies of both the jet features and mass parameter, giving 18 input features per-track in total. This model is referred to in the following as the *Concatenation* model.

A second parameterised model uses conditional scaling, as shown in Figure 6.19. First, the output of a dense network is used to scale each jet-track feature vector, before

they are passed into the initialisation network. The network maps m_a to a scaling vector of size 17, to match the total number of jet and track variables, via a single ReLU activated hidden layer with 4 neurons. Conditional scaling is then applied to the input features of each layer of the transformer-based encoder using four dense networks that map m_a to a scaling vector of size 256 (to match d_{embed}), via a single ReLU activated hidden layer with 128 neurons. The output of a final dense network is used to scale each track representation outputted by the encoder. This network maps m_a to a scaling vector of size 128, to match the number of features in the conditional track representations, via a single ReLU activated hidden layer with 64 neurons. When scaling the encoder and global representations, sigmoidal gating is used, to ensure each feature is scaled by a value between 0 and 1. This model is referred to in the following as the *Scaling* model.

A final parameterised model uses conditional affine transformations, as shown in Figure 6.20. First, the output of two dense networks is used to scale then bias each jet-track feature vector, before they are passed into the initialisation network. The networks map m_a to a transforming vector of size 17, to match the total number of jet and track variables, via a single ReLU activated hidden layer with 4 neurons. Conditional affine transformations are then applied to the input features of each layer of the transformer-based encoder using eight dense networks that map m_a to a scaling or biasing vector of size 256 (to match d_{embed}), via a single ReLU activated hidden layer with 128 neurons. A layer normalisation is applied to the features after each affine transformation. Sigmoidal gating is then applied to each track representation outputted by the encoder, using the output of a final dense network. This network maps m_a to a scaling vector of size 128, to match the number of features in the conditional track representations, via a single ReLU activated hidden layer with 64 neurons. This model is referred to in the following as the *Affine* model.

A summary of the four models tested, the type of parameterisation used in each, and their total trainable parameters, is given in Table 6.5. The models become increasingly complex, with the *Affine* model, featuring both conditional scaling and biasing across multiple layers, containing the most trainable parameters. However, there is no guarantee that increased complexity will result in better performance. There is no consensus

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

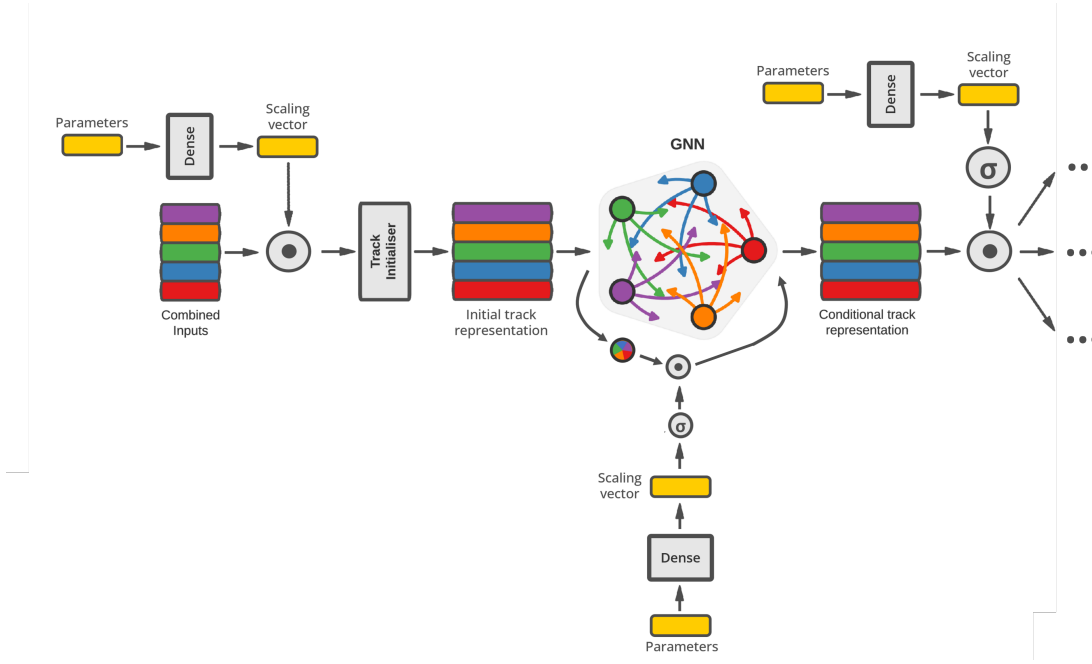


Figure 6.19: The *Scaling* model uses the baseline architecture in addition to conditional scaling applied to the combined jet-track input features, the node representations inputted to each encoder layer, and the conditional track representations outputted by the encoder. Sigmoidal gating is used when scaling the encoder layers and track embeddings. Parameters includes the long-lived particle mass m_a . The task layers are omitted for brevity. Adapted from Ref. [58].

as to which type of parameterisation is best, and the observed performance will likely be task dependent. Additionally, the specific architectures of the models using feature-wise transformations (*Scaling* and *Affine*) are the result of experimentation. The models presented are those found to give the most promising performance, given the type of transformation being incorporated, during these experiments. However, they may not be *the* most performative, and could be improved upon with further experimentation. The intent of this work is to demonstrate the advantages and disadvantages of different conditioning mechanisms for the task of jet classification, and to serve as a guide for those who wish to use them in a similar fashion in the future.

6.2.6 Training

The training procedure minimises the total loss function L_{total} , given by

$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}}, \quad (6.2.12)$$

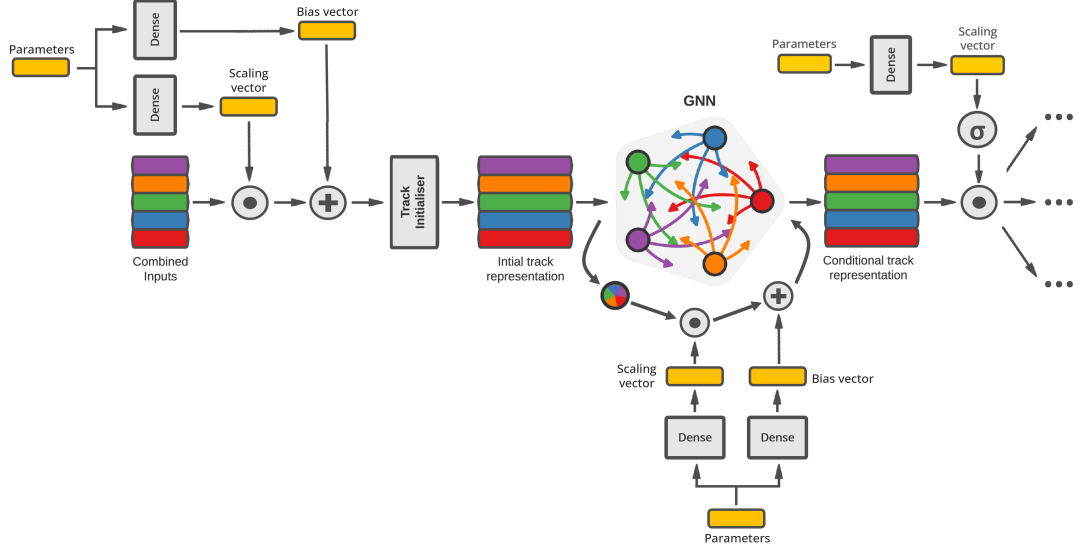


Figure 6.20: The *Affine* model uses the baseline architecture, in addition to affine transformations applied to the combined jet-track input features, and the node representations inputted to each encoder layer. Sigmoidal gating is then applied to the conditional track representations outputted by the encoder. Parameters includes the long-lived particle mass m_a . The task layers are omitted for brevity. Adapted from Ref. [58].

Model	Parameterisation										Parameters
	Inputs			Encoder			Embeddings				
	Type	h	o	Type	h	o	Type	h	o		
<i>Baseline</i>	-			-			-			2,342,379	
<i>Concatenation</i>	Concatenate m_a to inputs	-		-			-			2,342,635	
<i>Scaling</i>	Conditional scaling	4	17	Conditional scaling with sigmoidal gating	128	256	Conditional scaling with sigmoidal gating	64	128	2,484,040	
<i>Affine</i>	Affine transformation	4	17	Affine transformation and layer normalisation	128	256	Conditional scaling with sigmoidal gating	64	128	2,619,301	

Table 6.5: Summary of the types of parameterisation used in each model. The layer in which the conditioning information is used is indicated. This is either: ‘Inputs’, meaning applied to the track features prior to the initialisation network, ‘Encoder’, meaning applied to the features of each layer in the encoder, or ‘Embeddings’, meaning applied to the conditional track representations outputted by the encoder prior to pooling. When feature-wise transformations are used, h and o indicate the number of nodes in the hidden and output layers on the dense networks producing the transforming vectors. The final column contains the number of trainable parameters in the model.

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

where L_{jet} is the categorical cross entropy over the two jet classes, L_{vertex} is the binary track-pair compatibility loss, and L_{track} is the categorical cross entropy for the track origin prediction. L_{jet} is computed by averaging the loss over all jets in a batch. L_{vertex} is computed by averaging the loss of all intra-jet track-pairs in a batch. L_{track} is a weighted average over all tracks in a batch, where tracks are weighted by the inverse of their frequency within the training set.

Each loss in Eq. (6.2.12) converges to a different value, reflecting the relative difficulty of each task. To account for this, weights α and β are applied to L_{vertex} and L_{track} when computing L_{total} . These weights effectively act as a form of regularisation, and can prevent individual tasks from overfitting by reducing their impact in the overall loss. Values of $\alpha = 0.15$ (or 0.5 for models trained using $\mathcal{D}_7$, $\mathcal{D}_{10}$, $\mathcal{D}_{25}$, and $\mathcal{D}_{40}$) and $\beta = 0.015$ were chosen using a coarse optimisation. These values produce minimal overfitting for each task and a strong overall jet classification performance.

The mass of a hypothetical particle has no meaning for background jets. Therefore, during the training of a parameterised network, background jets are randomly assigned masses according to their distribution in the signal component of the training set being used. For example, when using $\mathcal{D}_{\mathcal{M}}$, background jets are assigned one of $m_a \in \mathcal{M}$ with equal probability. Assigning the masses during training means the mass value of a background jet may vary between epochs. This ensures there is no correlation between the mass parameter and the properties of the background jets. To evaluate a model at a particular mass, the mass parameter of all jets in the test sample is set to the desired value.

Each model is trained for 40 epochs² with a batch size of 600 on a single NVIDIA V100 Tensor Core GPU using the Salt framework [3]. The AdamW optimiser [106] is used, with an adaptive learning rate following the 1cycle policy [121]. The 1cycle policy anneals the learning rate from an initial value to some maximum value, then from that maximum value to some minimum value, across the full training cycle. In this work, the initial value is 1×10^{-7} , the maximum value is 5×10^{-4} , and the minimum value is 1×10^{-5} . The learning rate is set to increase for the initial 1% of the total training cycle. The results presented in §6.2.7 are obtained using the epoch that minimised the

²This is reduced to 30 for models trained using $\mathcal{D}_7$, $\mathcal{D}_{10}$, $\mathcal{D}_{25}$, and $\mathcal{D}_{40}$.

validation loss of the jet classification task.

6.2.7 Results

The primary purpose of the models described in §6.2.5 is to differentiate between displaced jets produced through hadronic decays of new, LLPs and jets produced in SM background processes. As such, the results presented in this section focus on the main jet classification task. Each model is assessed on its ability to correctly classify unseen jets associated to LLPs with masses that were included in the training data, as well as jets associated to LLPs with masses that were not, the latter quantifying the interpolation power of the model. The jet classification performance of the models is compared by computing the background jet rejection (defined as the inverse of the background jet efficiency) with respect to the signal jet efficiency. The performance of the tracking and vertexing auxiliary tasks are not included in this thesis, and could warrant dedicated studies in the future.

Interpolation to unseen masses

The interpolation power of a PNN is its capacity to generalise to novel LLP mass points that lie between values that were used in training. Smooth interpolation between different masses, and strong performance in cases where samples for certain mass hypotheses are missing, are some of the key possible benefits of PNNs, as discussed in §6.2.1.

Two tests are performed to assess each class of model on its ability to interpolate solutions to unseen masses. In the first, models are trained using jets generated at all signal mass points in $\mathcal{M}$ except for one (giving seven total mass points), then evaluated using jets generated at the missing mass point, forcing the models to interpolate their output to this single mass value. Models are trained using $\mathcal{D}_{\mathcal{M}\setminus 7}$, $\mathcal{D}_{\mathcal{M}\setminus 10}$, $\mathcal{D}_{\mathcal{M}\setminus 25}$, $\mathcal{D}_{\mathcal{M}\setminus 40}$, then evaluated on test sets containing only jets from $m_a = 7, 10, 25$ and 40 GeV signal samples, respectively. This is followed by a more challenging test in which each class of model is trained using $\mathcal{D}_{\mathcal{M}\setminus 7,10,25,40}$, then evaluated separately using jets from the excluded $7, 10, 25$ and 40 GeV signal samples. By reducing the number of mass points in the training set (to a total of four in this case), more stress is placed on the interpolation. The interpolation power of each model is therefore tested in the more realistic scenario of

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

having a limited number of available mass points, and using a single model to interpolate solutions to multiple mass points.

In both tests, the interpolation performance of the parameterised models is compared to *Baseline* models trained using the same training set. This isolates the effect of the parameterisation on the interpolation performance, by keeping all other aspects of the model and training procedure constant. The models are also compared to *Baseline* models trained using only jets from the interpolated mass point, referred to here as an *isolated* network. This tests how the performance of a model benefiting from shared weights, increased sample size (as training as a single mass point naturally reduces the sample size, as shown in Table 6.1), and any parameterisation, compares to an isolated network, even when forced to interpolate its solution to the desired mass.

Figure 6.21 shows the background rejection as a function of signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7}$, $\mathcal{D}_{\mathcal{M}\setminus 10}$, $\mathcal{D}_{\mathcal{M}\setminus 25}$, $\mathcal{D}_{\mathcal{M}\setminus 40}$, along with isolated networks trained using $\mathcal{D}_7$, $\mathcal{D}_{10}$, $\mathcal{D}_{25}$, $\mathcal{D}_{40}$, and evaluated using test sets containing only jets from $m_a = 7, 10, 25$ and 40 GeV signal samples, respectively. Jets in the test samples are required to have p_T within the range used in the resampling ($20 < p_T < 300$ GeV). Table 6.6 shows the background rejection of each model (not including the isolated networks) at signal efficiency working points (WPs) of 70% and 85%. When interpolating to lower masses, the parameterised models demonstrate a significant increase in the background rejection across the entire signal efficiency range studied when compared to the *Baseline* unparameterised networks trained using the same training sets. At the 70% WP, parameterisation increases the background rejection by up to 88% and 72% when interpolating to 7 GeV and 10 GeV, respectively. At the 85% WP, increases of up to 68% and 77% are observed for the same mass points. When interpolating to higher masses, the parameterised networks show better performance than the *Baseline* models at lower signal efficiencies, with increases in background rejection of up to 9.8% and 18% when interpolating to 25 GeV and 40 GeV, respectively, at the 70% WP. However, a loss in performance is observed at higher signal efficiencies. The *Baseline* models outperform each parameterised model at the 85% WP, where parameterisation results in losses in background rejection of up to 7.8% and 17% when interpolating to 25 GeV and 40 GeV,

Dataset	Mass (GeV)	<i>Baseline</i>	<i>Concatenation</i>	<i>Scaling</i>	<i>Affine</i>	Max. $\pm\%$ diff. from <i>Baseline</i>
70% WP background rejection						
$\mathcal{D}_{\mathcal{M}\setminus 7}$	7	277 ± 1	520 ± 3	522 ± 3	499 ± 3	+88%
$\mathcal{D}_{\mathcal{M}\setminus 10}$	10	558 ± 4	892 ± 8	960 ± 9	850 ± 7	+72%
$\mathcal{D}_{\mathcal{M}\setminus 25}$	25	1110 ± 10	1160 ± 10	1150 ± 10	1220 ± 10	+9.9%
$\mathcal{D}_{\mathcal{M}\setminus 40}$	40	2010 ± 30	2120 ± 30	2370 ± 30	2260 ± 30	+18%
85% WP background rejection						
$\mathcal{D}_{\mathcal{M}\setminus 7}$	7	73.8 ± 0.2	130.3 ± 0.4	129.1 ± 0.4	127.8 ± 0.4	+77%
$\mathcal{D}_{\mathcal{M}\setminus 10}$	10	121.8 ± 0.4	188.6 ± 0.7	204.2 ± 0.8	186.4 ± 0.7	+68%
$\mathcal{D}_{\mathcal{M}\setminus 25}$	25	112.3 ± 0.3	104.5 ± 0.3	103.5 ± 0.3	107.5 ± 0.3	-7.8%
$\mathcal{D}_{\mathcal{M}\setminus 40}$	40	110.1 ± 0.3	91.6 ± 0.3	96.4 ± 0.3	95.9 ± 0.3	-17%

Table 6.6: The background rejection at 70% and 85% signal efficiency working points for models trained using the training sets given in the *Dataset* column and evaluated at the mass point given in the *Mass* column, using jets with p_T in between 20 GeV and 300 GeV. The best performing model at a given working point is highlighted in bold. The final column shows the maximum percentage difference between the background rejection of the parameterised models (*Concatenation*, *Scaling* and *Affine*) and the *Baseline* model. The binomial uncertainties are shown.

respectively. The performance of the three parameterised models is comparable, with no particular architecture clearly outperforming any other.

Figure 6.21 also highlights the benefit of an increased sample size and the sharing of weights resulting from training set comprised of multiple mass points. Each parameterised model (and in most cases *Baseline* model) trained using $\mathcal{D}_{\mathcal{M}\setminus m}$ for $m = 7, 10, 25, 40$ GeV considerably outperforms the corresponding isolated network, despite the parameterised models having to interpolate their solution to the mass m . The effect is more pronounced for higher LLP masses.

Figure 6.22 shows the background rejection as a function of signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7,10,25,55}$, along with isolated networks trained using $\mathcal{D}_7$, $\mathcal{D}_{10}$, $\mathcal{D}_{25}$, $\mathcal{D}_{40}$, and evaluated using separate test sets containing only jets from $m_a = 7, 10, 25$ or 40 GeV signal samples. Table 6.7 shows the background rejection of each model (not including the isolated networks) at signal efficiency WPs of 70% and 85%. The background rejection for all models is generally ~ 10 -20% lower than those in Table 6.6 for the same LLP mass. This is expected as having fewer mass points in the training set, along with fewer jets to train on overall, places more strain on the interpolation. The parameterised models still generally outperform the *Baseline* model, once again more

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

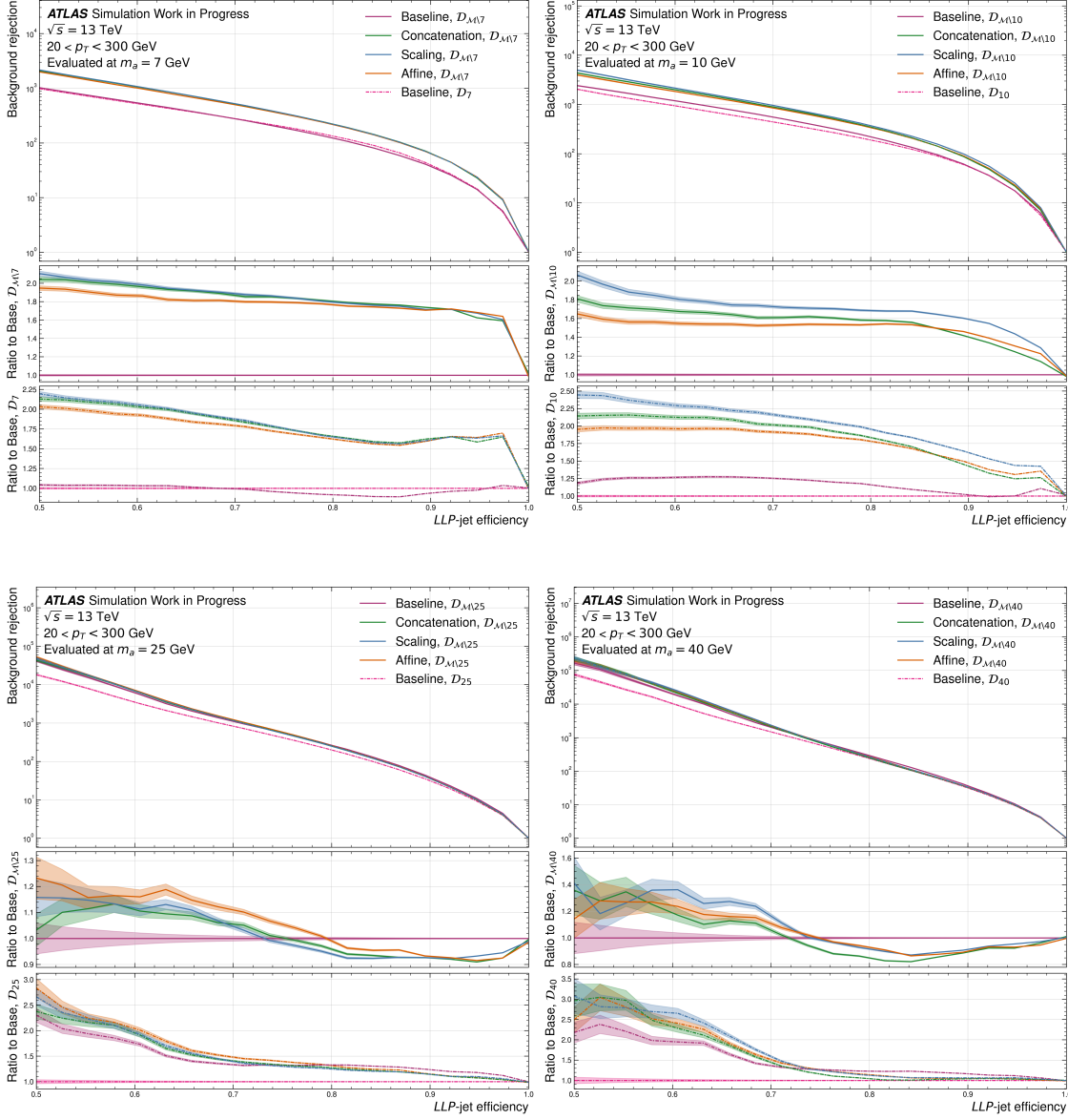


Figure 6.21: Background jet rejection versus signal jet efficiency for models trained using $\mathcal{D}_{M\backslash 7}$ (top left), $\mathcal{D}_{M\backslash 10}$ (top right), $\mathcal{D}_{M\backslash 25}$ (bottom left), and $\mathcal{D}_{M\backslash 40}$ (bottom right). Each model is evaluated using a test set comprised of jets with p_T between 20 and 300 GeV generated at the excluded mass point. The dashed lines correspond to isolated *Baseline* models trained using jets generated at one of the excluded mass points with $\mathcal{D}_m$ for $m = 7, 10, 25, 40$ GeV. The top ratio panel is computed with respect to the *Baseline* model trained using one of $\mathcal{D}_{M\backslash m}$ for $m = 7, 10, 25, 40$ GeV, while the bottom ratio panel is computed with respect to corresponding isolated model. Binomial error bands are denoted by the shaded regions.

Mass (GeV)	<i>Baseline</i>	<i>Concatenation</i>	<i>Scaling</i>	<i>Affine</i>	Max. $\pm\%$ diff. from <i>Baseline</i>
70% WP background rejection					
7	275 ± 1	471 ± 3	444 ± 3	437 ± 3	+71%
10	456 ± 3	703 ± 5	657 ± 5	699 ± 5	+54%
25	1120 ± 10	1027 ± 9	1070 ± 10	1190 ± 10	+6.3%
40	1910 ± 20	2100 ± 30	2030 ± 30	2150 ± 30	+13%
85% WP background rejection					
7	72.2 ± 0.2	121.5 ± 0.4	117.3 ± 0.4	117.3 ± 0.4	+68%
10	102.3 ± 0.3	150.2 ± 0.5	148.1 ± 0.5	154.7 ± 0.6	+51%
25	105.0 ± 0.3	99.1 ± 0.3	100.2 ± 0.3	107.7 ± 0.3	+2.6%
40	104.3 ± 0.3	93.3 ± 0.3	94.7 ± 0.3	97.0 ± 0.3	-11%

Table 6.7: The background rejection at 70% and 85% signal efficiency working points for models trained using $\mathcal{D}_{\mathcal{M} \setminus \{7,10,25,40\}}$ and evaluated at the mass point given in the *Mass* column using jets with p_T in between 20 GeV and 300 GeV. The best performing model at a given working point is highlighted in bold. The final column shows the maximum percentage difference between the background rejection of the parameterised models (*Concatenation*, *Scaling* and *Affine*) and the *Baseline* model. Positive differences are prioritised even if there is a greater negative difference. The binomial uncertainties are shown.

significantly at lower masses, with the *Affine* model being most consistent.

Classifying jets from known mass points

To test the ability of each model to classify jets associated to LLPs with masses present in the training sample, each class of model is trained using the full dataset $\mathcal{D}_{\mathcal{M}}$, and evaluated separately using test sets generated at each $m_a \in \mathcal{M}$. Jets in the test samples are required to have p_T within the range used in the resampling ($20 < p_T < 300$ GeV). In this instance, neither the *Baseline* or parameterised networks need to interpolate their output. Figure 6.23 shows the background rejection as a function of signal efficiency for each model at each mass point. Table 6.6 shows the background rejection of each model for each mass point at signal efficiency WPs of 70% and 85%.

The BDT described in §5.4 and used in the $H \rightarrow aa$ search had a background efficiency of approximately 12% (averaged over samples with $m_a = 16$ and 55 GeV) at an 85% WP, corresponding to a background rejection of ~ 8 . Therefore, by replacing the BDT with the core GN2 architecture alone one would expect to improve background rejection by a factor of $\sim 10\text{--}15\times$, with further improvements possible through addi-

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

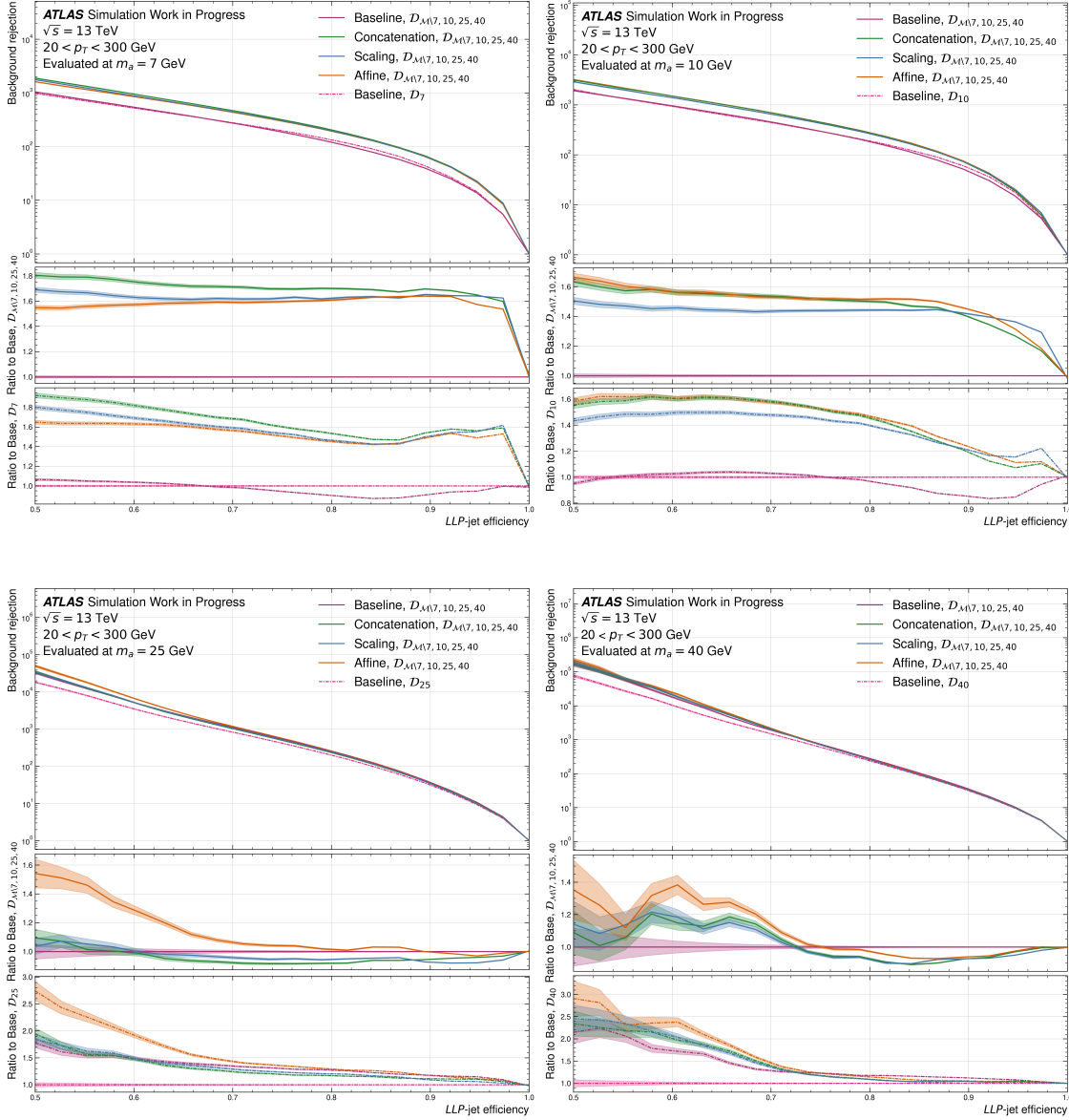


Figure 6.22: Background jet rejection versus signal jet efficiency for models trained using $\mathcal{D}_{M\backslash 7,10,25,40}$ and evaluated using test sets comprised of jets with p_T greater than 300 GeV generated at 7 GeV (top left), 10 GeV (top right), 25 GeV (bottom left), and 40 GeV (bottom right). The dashed lines correspond to isolated *Baseline* models trained using jets generated at one of the excluded mass points with $\mathcal{D}_m$ for $m = 7, 10, 25, 40$ GeV. The top ratio panel is computed with respect to the *Baseline* model trained using one of $\mathcal{D}_{M\backslash 7,10,25,40}$ while the bottom ratio panel is computed with respect to corresponding isolated model. Binomial error bands are denoted by the shaded regions.

tional parameterisation. For lower LLP masses, the parameterised networks produce a considerable increase in background rejection across the entire signal efficiency range studied when compared to the *Baseline* model. For m_a between 5 GeV and 16 GeV, increases in background rejection up to 32-90% and 21-100% are observed at the 70% and 85% WPs, respectively. The additional contextual information provided by the mass parameter can therefore be effectively exploited through parameterisation, and increase the discriminatory power of a model at seen mass points as well as interpolated mass points. For $m_a \geq 25$ GeV, while the parameterised networks have a higher background rejection at lower signal efficiencies, the *Baseline* model outperforms them at higher signal efficiencies. The performance of the parameterised models is comparable for higher masses, while the *Concatenation* model performs the best most often for lower masses.

Figure 6.24 shows the output of *Baseline* and *Affine* models trained using $\mathcal{D}_{\mathcal{M}}$ when evaluated using an inclusive background test sample and signal samples for each $m_a \in \mathcal{M}$. The figure also contains the background rejection as a function of signal efficiency for each model evaluated at each mass point. Figure 6.24 highlights the dependency of the output of the parameterised *Affine* model on the choice of mass parameter. For the same background test sample, the *Affine* model output varies considerably when evaluated at different masses in $\mathcal{M}$. The same is true of the other parameterised models. At higher masses, the distribution is smeared, resulting in fewer background jets with output scores close to both 0 (i.e. highly background-like) and 1 (i.e. highly signal-like), reflecting a lesser degree of confidence in the model in its ability to correctly classify background jets. This results in a loss in discriminatory power for higher signal efficiencies and an increase at lower signal efficiencies, as observed in Figure 6.23. This is emphasized in Figure 6.25, which directly compares the *Baseline* and *Affine* model outputs for background and signal samples when evaluated at $m_a = 7$ and 40 GeV.

For a search such as the one presented in §5, which uses a data-driven background estimate, having a background distribution that is dependent on the signal hypothesis means computing a separate background estimate using the data for each mass point. Depending on how many mass points are being used, and how demanding the background estimation procedure is, a parameterised approach may produce additional com-

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

Mass (GeV)	<i>Baseline</i>	<i>Concatenation</i>	<i>Scaling</i>	<i>Affine</i>	Max. $\pm\%$ diff. from <i>Baseline</i>
70% WP background rejection					
5	329 ± 2	625 ± 5	601 ± 4	625 ± 5	+90%
7	329 ± 2	585 ± 4	592 ± 4	574 ± 4	+80%
10	614 ± 4	991 ± 9	849 ± 7	943 ± 8	+61%
16	694 ± 5	917 ± 8	911 ± 8	836 ± 7	+32%
25	1300 ± 10	1240 ± 10	1190 ± 10	1280 ± 10	-8.5%
35	2000 ± 30	2120 ± 30	2120 ± 30	2110 ± 30	+6.0%
40	2200 ± 30	2410 ± 30	2450 ± 30	2460 ± 40	+12%
55	6300 ± 100	8200 ± 200	8200 ± 200	8100 ± 200	+30%
85% WP background rejection					
5	77.9 ± 0.2	156.1 ± 0.6	145.3 ± 0.3	153.0 ± 0.5	+100%
7	87.0 ± 0.2	155.7 ± 0.6	153.2 ± 0.5	150.0 ± 0.5	+79%
10	136.7 ± 0.5	206.7 ± 0.9	178.9 ± 0.7	200.9 ± 0.8	+51%
16	129.0 ± 0.4	156.6 ± 0.6	156.3 ± 0.6	147.3 ± 0.5	+21%
25	119.3 ± 0.4	108.6 ± 0.3	109.1 ± 0.3	108.2 ± 0.3	-9.3%
35	120.0 ± 0.4	110.4 ± 0.3	110.2 ± 0.3	107.0 ± 0.3	-11%
40	115.4 ± 0.4	103.0 ± 0.3	103.1 ± 0.3	100.6 ± 0.3	-13%
55	118.0 ± 0.4	115.1 ± 0.4	114.9 ± 0.4	113.3 ± 0.3	-4.0%

Table 6.8: The background rejection at 70% and 85% signal efficiency working points for models trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated at the mass point given in the *Mass* column, using jets with p_T in between 20 GeV and 300 GeV. The best performing model at a given working point is highlighted in bold. The final column shows the maximum percentage difference between the background rejection of the parameterised models (*Concatenation*, *Scaling* and *Affine*) and the *Baseline* model. The binomial uncertainties are shown.

plexity that outweighs the potential gains in displaced jet identification performance. This would need to be considered by the analysers. The parameterised approach is more suited for an MC-driven background estimate.

Figure 6.24 also shows that for each class of model the background rejection is significantly less for jets associated to lighter LLPs at low signal efficiencies, when compared to jets associated to heavier LLPs. The mass of the LLP plays an important role in the classification task. Jets associated to lighter LLPs are kinematically more comparable to background jets from heavy-flavour decays, making their differentiation harder (as explored further below), resulting in proportionately fewer jets being given very high scores close to 1. At higher signal efficiencies the background rejection is greater for jets associated to lighter LLPs.

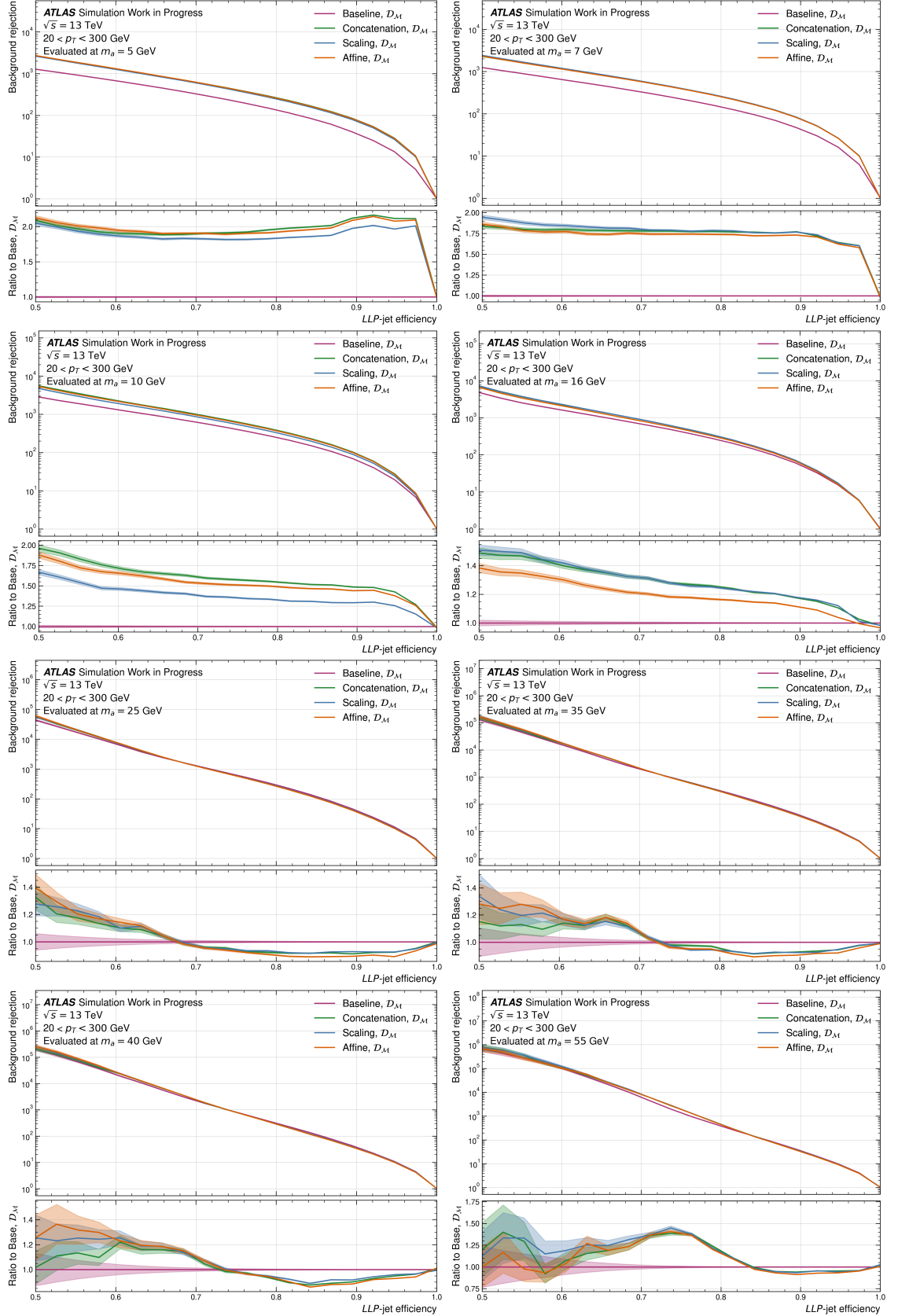


Figure 6.23: Background jet rejection versus signal jet efficiency for models trained using $\mathcal{D}_M$ and evaluated separately using jets with p_T between 20 and 300 GeV from samples generated with (moving left to right and top to bottom) m_a equal to 7, 10, 16, 25, 35, 40, and 55 GeV. The ratio panel is computed with respect to the *Baseline* model. Binomial error bands are denoted by the shaded regions.

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

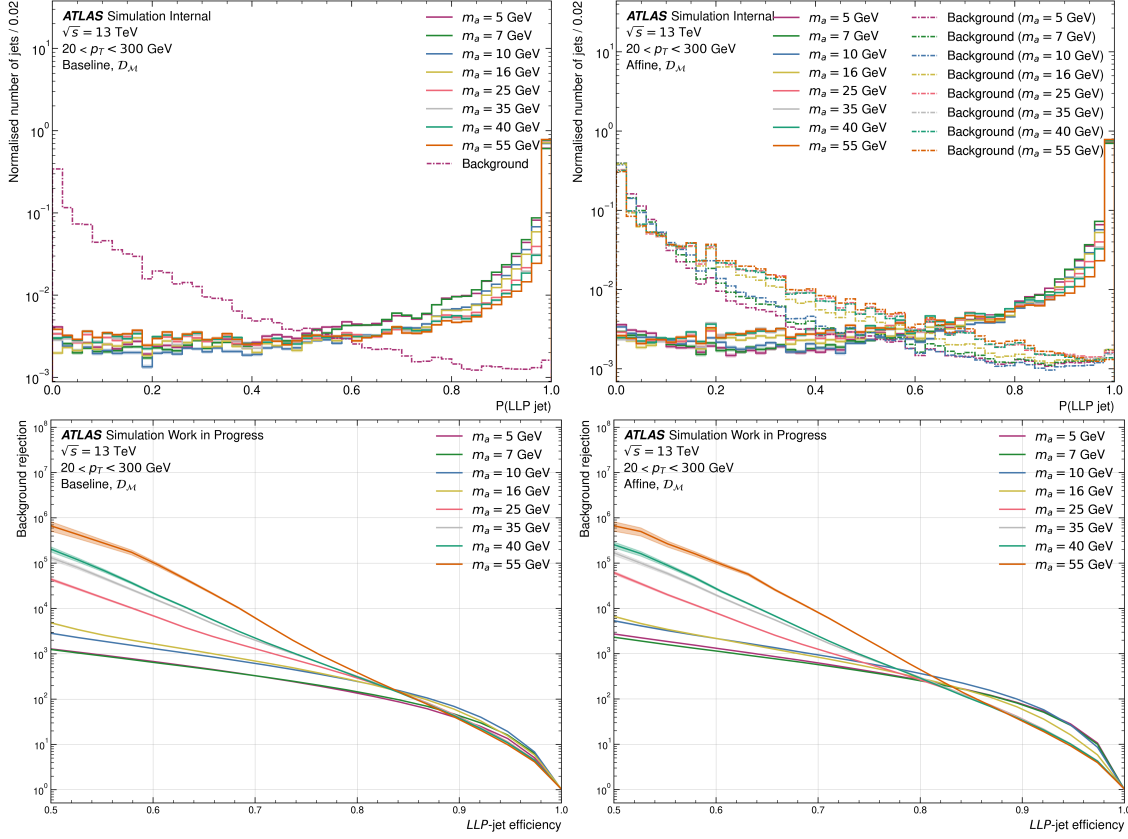


Figure 6.24: Model output (labelled as $P(\text{LLP jet})$) (top row) and background rejection (bottom row) for a Baseline model (left) and Affine model (right). The model output is shown for the combined background and each mass point. The background rejection is computed separately for each mass point. Note that background distribution for the *Affine* model is dependent on the choice of mass parameter.

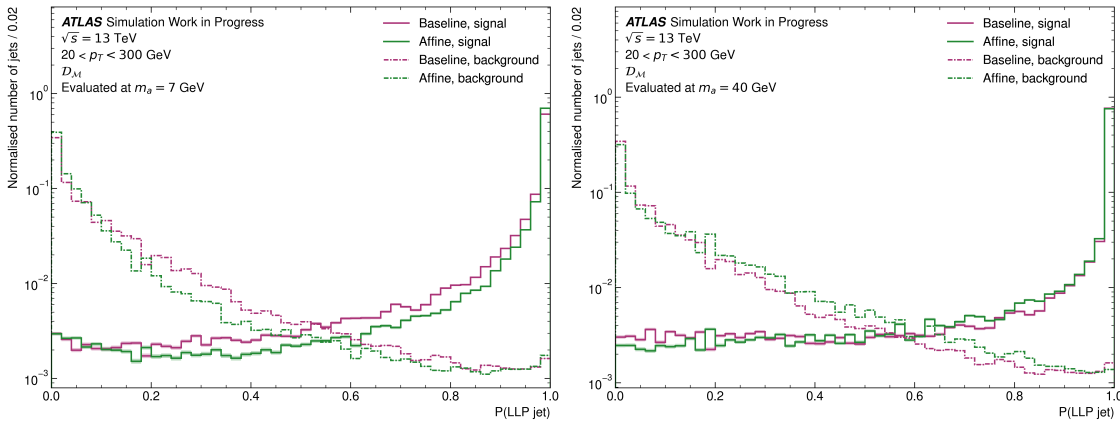


Figure 6.25: Model output (labelled as $P(\text{LLP jet})$) for 7 GeV (left) and 40 GeV (right) signal (solid lines) and background (dotted lines) samples for *Baseline* (purple) and *Affine* (green) models trained using $\mathcal{D}_{\mathcal{M}}$. Note that background distribution for the *Baseline* model is the same in each plot, while for the *Affine* model it is dependent on the choice of mass parameter.

Jet observable dependence

Studying the dependence of the performance of the model on some key jet observables can provide further insight into the classification problem and the functionality of the model itself.

Figure 6.26 shows the signal jet efficiency as a function of jet p_T for a fixed overall signal jet efficiency of 70% for the *Affine* model (although similar distributions are observed for each architecture for this and every other observable explored in this section) evaluated at $m_a = 5, 25$ and 55 GeV. The performance of the model is strongly dependent on jet p_T . The efficiency generally increases with jet p_T as higher p_T jets will in general have an increased track multiplicity and therefore more geometric information for the model to exploit. The dependence on jet p_T is greater for jets associated to lighter LLPs. Figure 6.26 also shows the signal jet efficiency as a function of the number of tracks associated to the jet.

Figure 6.27 shows the signal jet efficiency as a function of the jet displaced p_T fraction, $\sum p_T^{\text{track, LLP}} / \sum p_T^{\text{track}}$, and the displacement of the parent LLP of the jet. The latter is defined as the distance between the production and decay vertices of the LLP as recorded in the truth information.

The signal efficiency increases as the displaced p_T fraction of the jet increases. The efficiency peaks close to 1, where the quality of the reconstruction of particles within the jet is best, before steadily decreasing.

The signal efficiency is higher for jets with greater $\sum p_T^{\text{track, LLP}} / \sum p_T^{\text{track}}$. This could be due to the auxiliary track origin classification task guiding the model to identify LLP descendent tracks, and associating them to signal-like jets.

Lastly, the signal efficiency decreases at short displacements of the parent LLP of the jet. Signal jets produced at shorter displacements are harder to distinguish from prompt jets produced in SM background processes, as the topological features of the jets will be more alike. For example, tracks associated to jets produced in similar regions of the detector will have more comparable hit content than those in jets produced further apart, since the tracks would have to pass through the same active layers. The loss in efficiency is greater when the mass of the parent LLP is similar to the mass of heavy quarks. At the

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

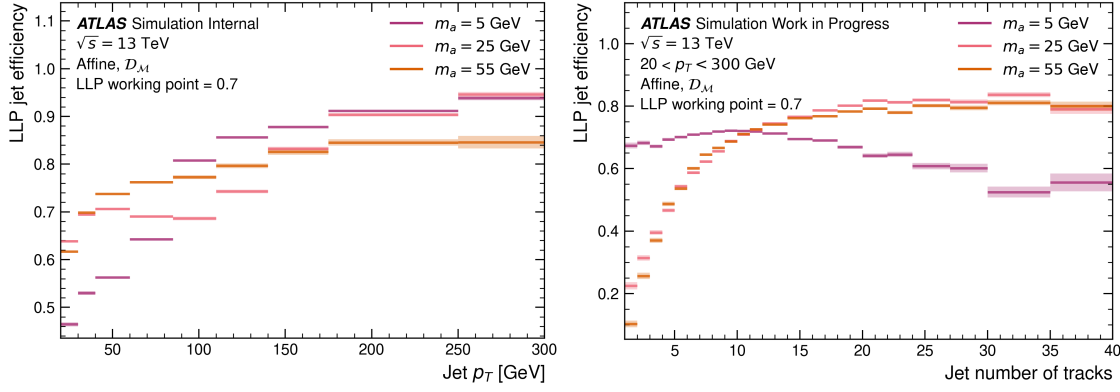


Figure 6.26: Signal LLP jet efficiency as a function of jet p_T (left) and number of tracks associated to the jet (right) at a fixed signal LLP jet efficiency of 70%, for an *Affine* model trained using $\mathcal{D}_M$ and evaluated at $m_a = 5, 25$ and 55 GeV.

70% WP, the signal jet efficiency for the $m_a = 5$ GeV sample is on the order of a couple of percent when the LLP displacement is below 10 mm, which is significantly lower than the $\sim 54\%$ efficiency observed for the $m_a = 55$ GeV sample. As previously discussed, jets produced in decays of lighter LLPs will be closer kinematically to heavy-flavour background jets, making their differentiation harder. The signal efficiency plateaus at LLP displacements of approximately 20, 40 and 100 mm for LLP masses of 55, 25 and 5 GeV, respectively. At larger displacements, the quality of the reconstruction of the tracks in the jet will degrade as there are fewer active layers to record hits. As a result, the signal efficiency decreases.

Flavour dependence

The training sets listed in Table 6.1 each contain a mixture of signal jets formed of light- and heavy-quarks, in addition to different flavour background jets produced in a range of SM processes. The properties of a jet are strongly tied to its flavour, and therefore so is the performance of the model, as shown in Figure 6.28, which shows variation in the output of an *Affine* model trained using $\mathcal{D}_M$ and evaluated at 7 GeV and 40 GeV for each flavour of signal and type of background (where $Z + \text{jets}$ and $W + \text{jets}$ have been grouped into a single $V + \text{jets}$ class and divided by jet flavour). Distinguishing each flavour of signal from each class of background is itself a unique classification task.

Figure 6.29 shows the background efficiency as a function of signal light-flavour and heavy-flavour jet efficiency in signal samples generated at $m_a = 7$ GeV and $m_a =$

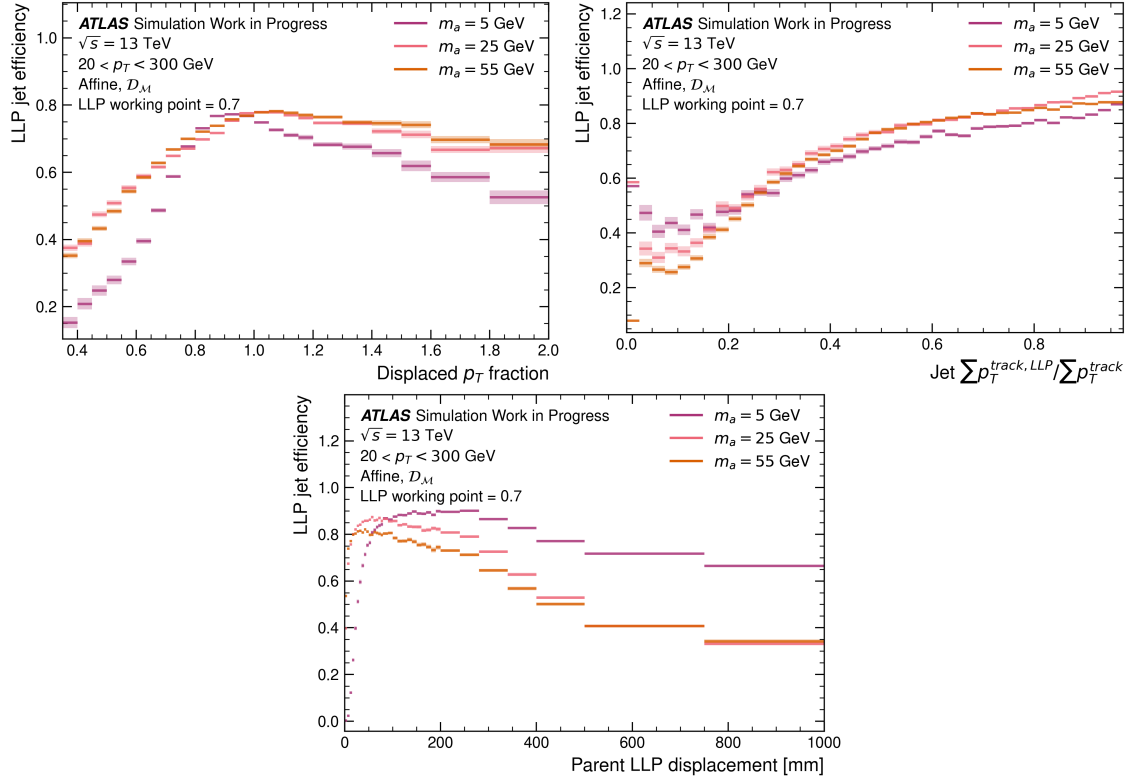


Figure 6.27: Signal LLP jet efficiency as a function of the jet displaced p_T fraction (top left), fraction of the total sum of transverse momenta of tracks associated to the jet attributed to tracks descending from the decay of an LLP (top right), and the displacement of the parent LLP of the jet (bottom middle), at a fixed signal LLP jet efficiency of 70%, for an *Affine* model trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated at $m_a = 5, 25$ and 55 GeV.

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

40 GeV for an *Affine* model trained using $\mathcal{D}_{\mathcal{M}}$. The heavy-flavour jets are comprised of c -quarks at 7 GeV and b -quarks at 40 GeV. In general, the background rejection is lower when identifying heavy-flavour signal jets. The effect of flavour on the model performance could be further disambiguated by training separate models using only light- or heavy-flavour signal jets, and individually studying their performance.

Figure 6.30 shows the background rejection for each SM background process computed separately as a function of signal light-flavour and heavy-flavour jet efficiency for an *Affine* model evaluated at $m_a = 7$ GeV and $m_a = 40$ GeV. For $m_a = 7$ GeV, the lowest background rejection is observed for $t\bar{t}$ events, followed by $V + b$ -jet and dijet events (depending on the signal efficiency). The best rejection is observed for $V + c$ -jet and $V + \text{light-jet}$ events. For $m_a = 40$ GeV, the lowest background rejection is observed for $V + b$ -jet events, followed by $t\bar{t}$ then dijet events. Once again the best rejection is observed for $V + c$ -jet and $V + \text{light-jet}$ events. The ordering is the same for both light-flavour and heavy-flavour signal jets, with the background rejection for heavy-flavour signal jets being generally lower as noted previously.

Heavy-flavour jets, such as those taken from $V + b$ -jet and $t\bar{t}$ samples, contain secondary vertices from displaced decays of heavy-flavour quarks, making their topology more signal-like. Furthermore, the presence of additional displaced decays makes track reconstruction less efficient, leading to a degradation in the quality of tracks inside the jet, resulting in lower rejection of heavy-flavour background jets and worse identification of heavy-flavour signal jets. One way of improving this may be to introduce additional weights for background jets during training, such that those containing heavy-flavour decays contribute more to the jet classification loss, therefore encouraging the model to classify these jets with greater accuracy.

Classifying high- p_T jets

Until now, the performance of each model has been evaluated using jets with p_T within the range chosen to perform the kinematic resampling ($20 < p_T < 300$ GeV) to the signal p_T distribution, as described in §6.2.3. While this p_T range captures ~ 98 –99% of the signal jets in MC (depending on the LLP mass), a sufficiently large dataset, such

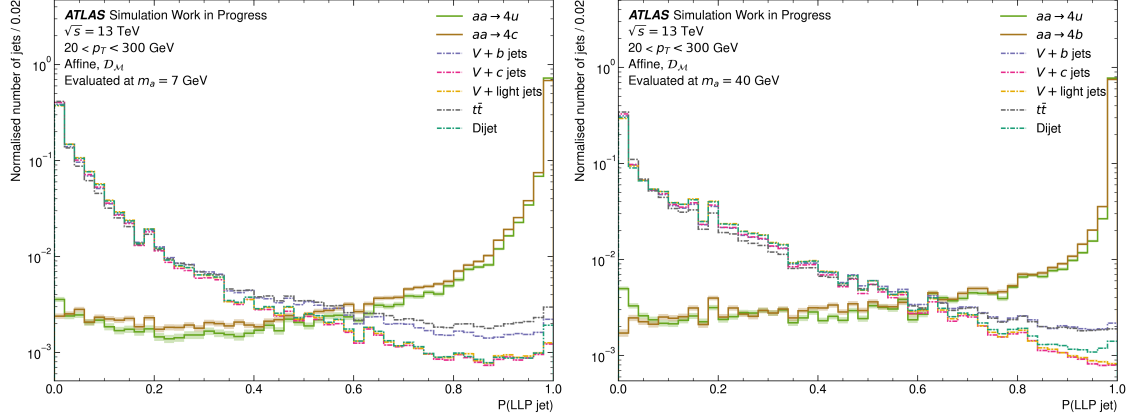


Figure 6.28: Model output (labelled as $P(\text{LLP jet})$) for different flavour signal jets and types of background for an *Affine* model evaluated at $m_a = 7 \text{ GeV}$ (left) and $m_a = 40 \text{ GeV}$ (right). The heavy jets are comprised of c -quarks and b -quarks for the 7 GeV and 40 GeV samples, respectively. Jets from $Z + \text{jets}$ and $W + \text{jets}$ samples have been grouped into a single $V + \text{jets}$ class, and divided by the flavour of the associated jet.

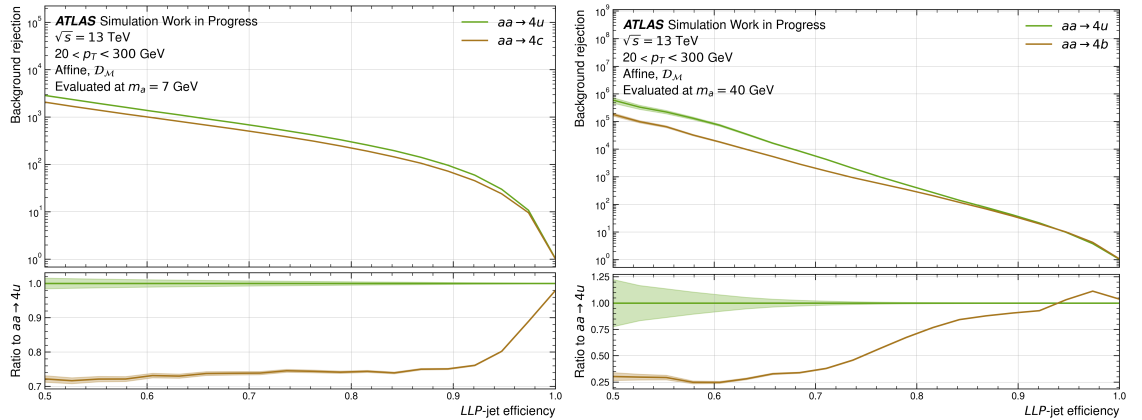


Figure 6.29: Background jet rejection versus signal efficiency for light- (green) and heavy-flavour (brown) jets in signal samples generated at $m_a = 7 \text{ GeV}$ (left) and $m_a = 40 \text{ GeV}$ (right) for an *Affine* model trained using $\mathcal{D}_{\mathcal{M}}$. The heavy jets are comprised of c -quarks and b -quarks for the 7 GeV and 40 GeV samples, respectively. The ratio is computed with respect to the background rejection for the light jets. Binomial error bands are denoted by the shaded regions.

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

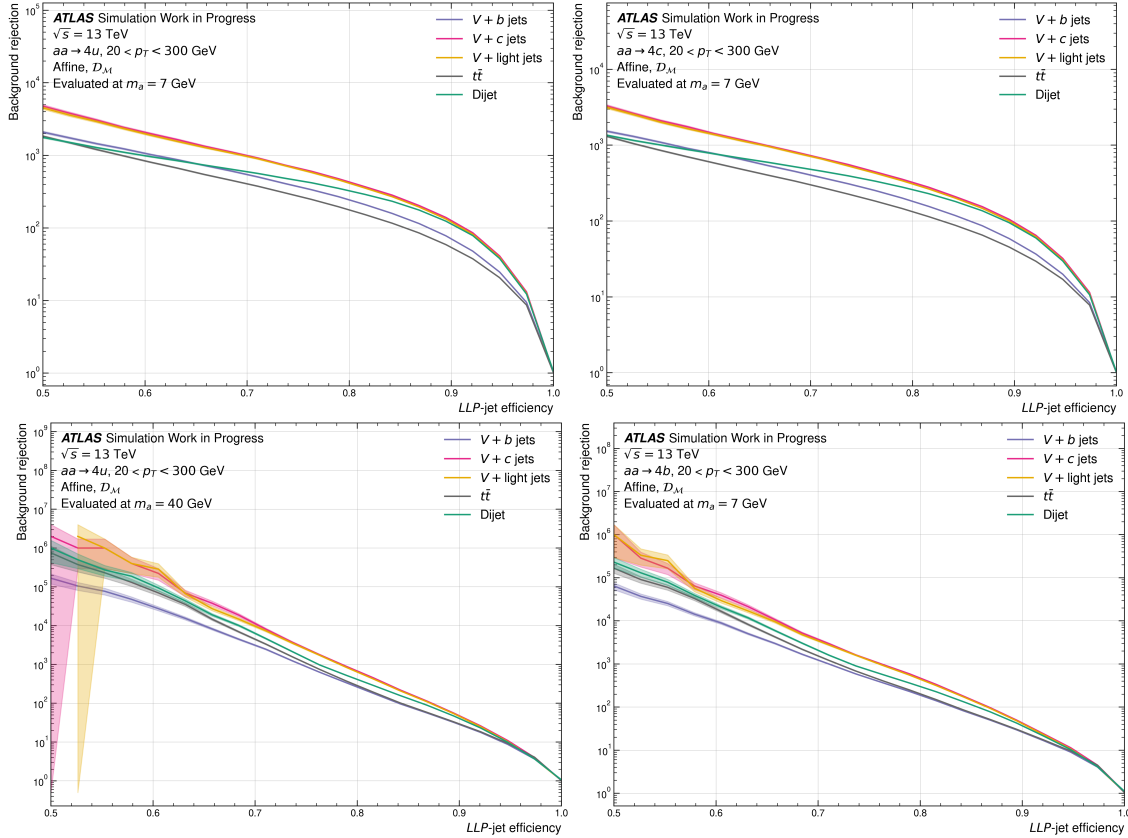


Figure 6.30: Background jet rejection for various SM processes as a function of signal light-flavour (left column) and heavy-flavour (right column) jet efficiency for an *Affine* model trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated at $m_a = 7$ GeV (top row) and $m_a = 40$ GeV (bottom row). The heavy jets are comprised of c -quarks and b -quarks for the 7 GeV and 40 GeV samples, respectively. Jets from Z +jets and W +jets samples have been grouped into a single V +jets class, and divided by the decay mode of the jets. Binomial error bands are denoted by the shaded regions. Truncated distributions result from there being no background jets available to compute the rejection at the given signal efficiency.

as that recorded at ATLAS during an entire LHC run, will contain a large number of high- p_T jets that fall outside of this range, which, if efficiently misclassified as signal, would harm the sensitivity of an analysis with low expected signal yields. It is therefore important to test how well the performance of a model generalises to jets with p_T that fall outside the range chosen to perform the resampling.

Figure 6.31 shows the background rejection as a function of signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7}$, $\mathcal{D}_{\mathcal{M}\setminus 10}$, $\mathcal{D}_{\mathcal{M}\setminus 25}$, $\mathcal{D}_{\mathcal{M}\setminus 40}$, along with isolated networks trained using $\mathcal{D}_7$, $\mathcal{D}_{10}$, $\mathcal{D}_{25}$, $\mathcal{D}_{40}$, and evaluated separately using test sets containing only jets from $m_a = 7, 10, 25$ and 40 GeV signal samples, respectively, where jets in the test samples, both signal and background, are required to have p_T greater than 300 GeV. Here, each model (except the isolated networks) is interpolating its output. Table 6.9 shows the background rejection of each model (not including the isolated networks) at signal efficiency WPs of 70% and 85% .

Figure 6.32 shows the background rejection as a function of signal efficiency for models trained using $\mathcal{D}_{\mathcal{M}}$ and evaluated separately using test sets generated at each $m_a \in \mathcal{M}$, where jets in the test samples are required to have p_T greater than 300 GeV. Table 6.10 shows the background rejection of each model trained using $\mathcal{D}_{\mathcal{M}}$ for each mass point at signal efficiency WPs of 70% and 85% .

Given the increase in signal efficiency and background rejection for higher p_T jets within the range $20 < p_T < 300$ GeV, one would expect the background rejection to continue to improve for jets with p_T greater than 300 GeV. This is generally observed.

When both interpolating to a new mass point and classifying jets from seen mass points, the *Affine* model is consistently better at generalising to jets with $p_T > 300$ GeV, when compared to the other models. Unlike the *Concatenation* and *Scaling* models, the *Affine* model outperforms the *Baseline* model across the majority of the signal efficiency range studied for all mass points. The relative improvement in background rejection tends to increase for larger mass points and at lower signal efficiencies, although low statistics at higher masses leads to large uncertainties in these results. The *Concatenation* model is the least effective at generalising to higher- p_T jets, with background rejections that tend to fall below those presented in Table 6.6 and Table 6.8.

The varied performance of the models in the high- p_T range suggests that the re-sampling procedure has introduced some undesirable dependence on the jet p_T in the output of the models. The p_T range used captures most of the signal, so while one could extend the range to higher p_T values, the lower signal stats in this region will mean few additional high- p_T background jets enter into the training. Furthermore, any sufficiently large dataset will always have a considerable number of jets that fall into the tails of the p_T distribution. Were the studies to be repeated, an alternate method of re-sampling could be considered. The current approach resamples the background such that its p_T distribution matches the signal. The signal is comprised of data from multiple mass points, each with their own unique p_T distributions. Therefore the resampled background distribution is not representative of the distribution of any one signal mass point, rather an average distribution computed over samples with different LLP mass. Resampling both signal and background to a uniform distribution may help to further decorrelate the output from jet p_T . Alternatively, removing resampling entirely could be considered. Preliminary studies using models trained with samples without kinematic resampling demonstrate more consistent improvement in background rejection due to parameterisation for each mass across the entire signal efficiency range, and similar levels of performance for each model in the high- p_T range. Nonetheless, the ability of the *Affine* architecture to effectively generalise to high p_T jets when resampling is used is a potentially interesting feature and could warrant further study. Generalising to areas of phase space outside of that seen in training could be useful if there is limited training data available, or if the available training data poorly represents the full kinematics of the problem.

6.2.8 Conclusion

In this section, the GN2 jet tagger, currently used at ATLAS for flavour tagging, has been adapted to identify displaced jets from $H \rightarrow aa$ decays, where a is a new, scalar LLP. The baseline GN2 improves background rejection with respect to the BDT used in §5 by $\sim 10\text{--}15\times$. Incorporating GN2 into a future iteration of a $H \rightarrow aa$ search will therefore significantly improve the sensitivity of the analysis.

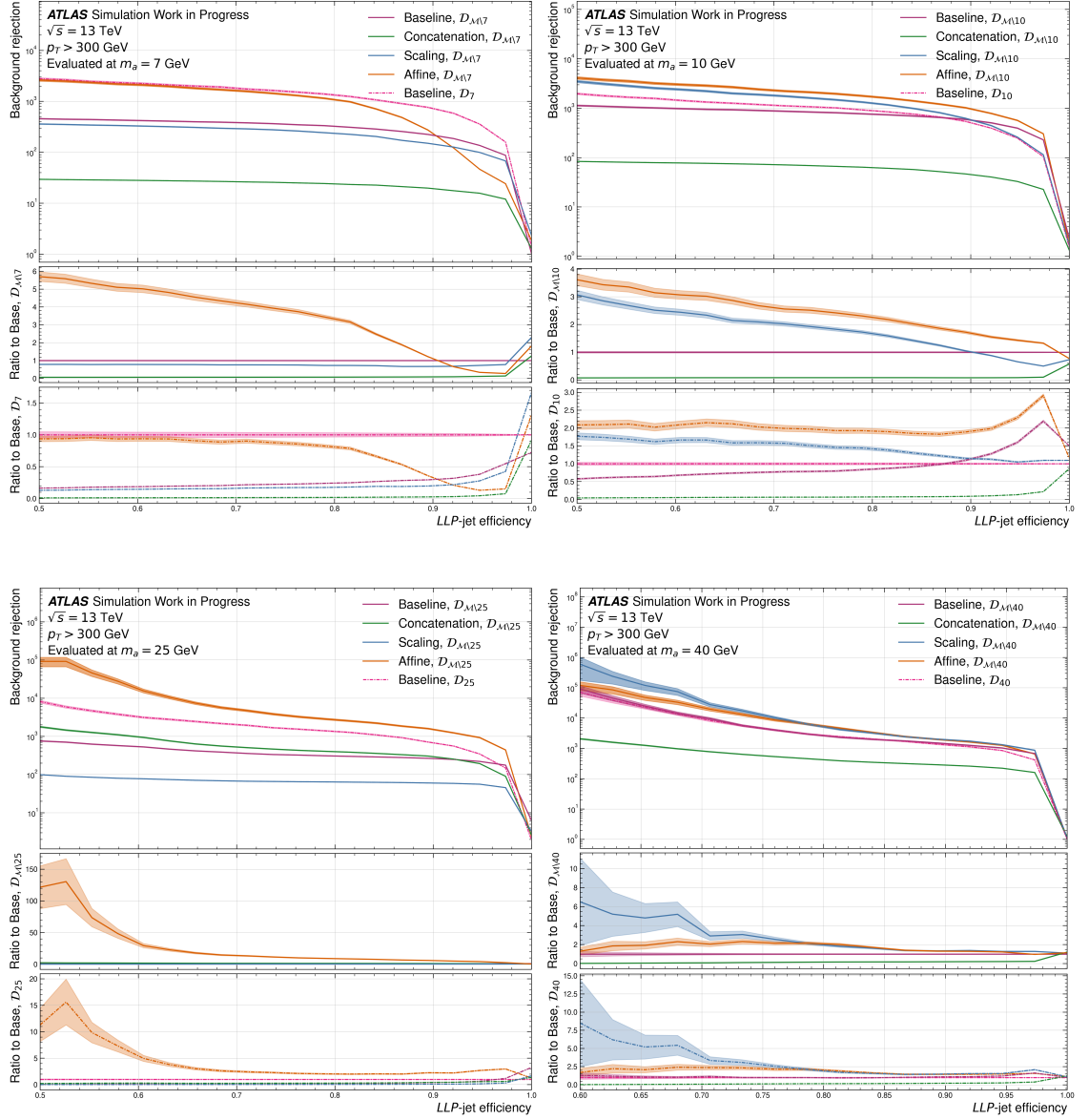


Figure 6.31: Background jet rejection versus signal jet efficiency for models trained using $\mathcal{D}_{\mathcal{M}\setminus 7}$ (top left), $\mathcal{D}_{\mathcal{M}\setminus 10}$ (top right), $\mathcal{D}_{\mathcal{M}\setminus 25}$ (bottom left), and $\mathcal{D}_{\mathcal{M}\setminus 40}$ (bottom right). Each model is evaluated using a test set comprised of jets with p_T greater than 300 GeV generated at the excluded mass point. The dashed lines correspond to isolated *Baseline* models trained using jets generated at one of the excluded mass points with $\mathcal{D}_m$ for $m = 7, 10, 25, 40$ GeV. For the latter, the signal efficiency range has been truncated at 0.6. The top ratio panel is computed with respect to the *Baseline* model trained using one of $\mathcal{D}_{\mathcal{M}\setminus m}$ for $m = 7, 10, 25, 40$ GeV, while the bottom ratio panel is computed with respect to corresponding isolated model. Binomial error bands are denoted by the shaded regions.

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

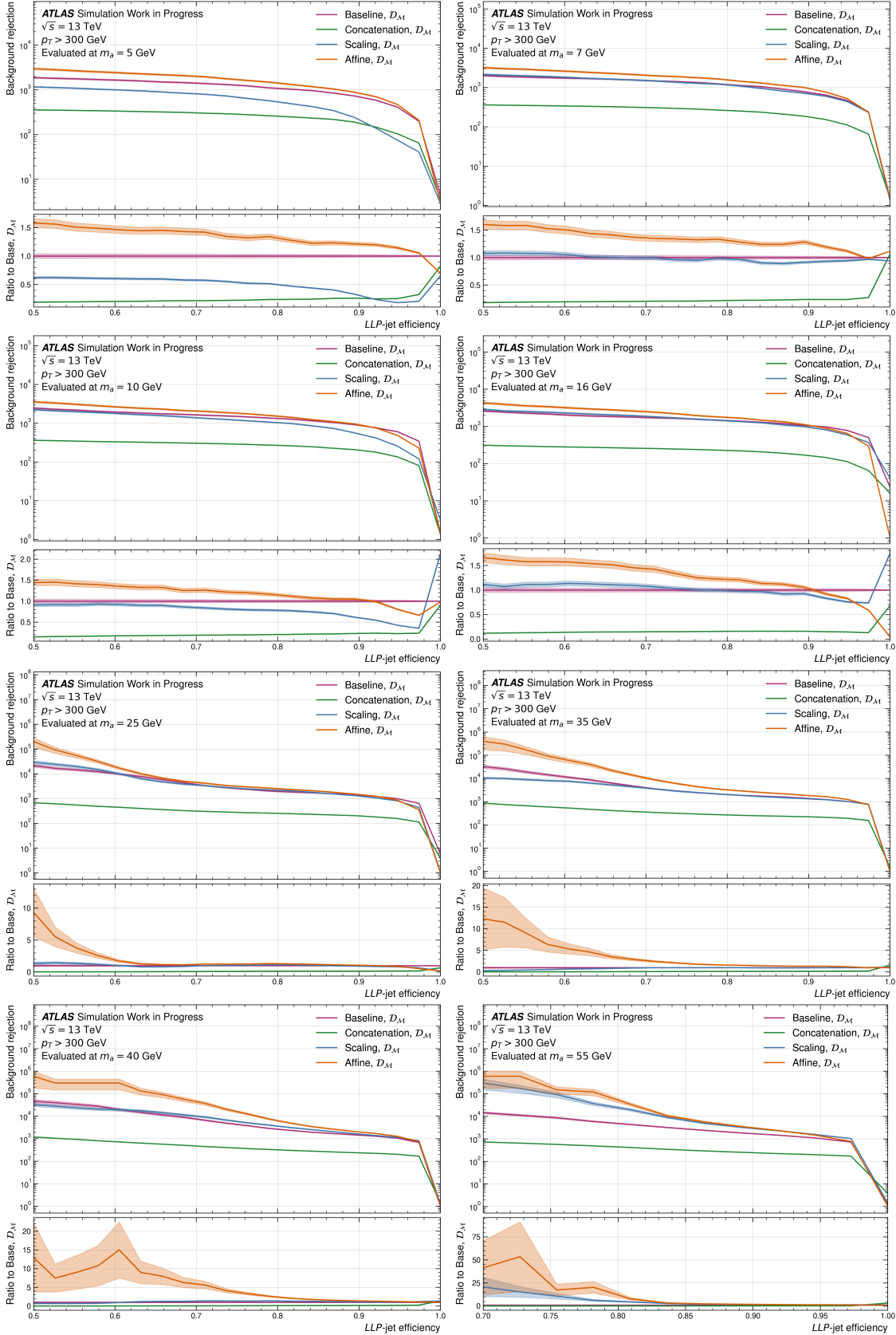


Figure 6.32: Background jet rejection versus signal jet efficiency for models trained using $\mathcal{D}_M$ and evaluated separately using jets with p_T greater than 300 GeV from samples generated with (moving left to right and top to bottom) m_a equal to 7, 10, 16, 25, 35, 40, and 55 GeV. For the latter the signal efficiency range has been truncated at 0.7. The ratio is computed with respect to the *Baseline* model. Binomial error bands are denoted by the shaded regions.

Dataset	Mass (GeV)	Baseline	Concatenation	Scaling	Affine	Max. $\pm\%$ diff. from Baseline
70% WP background rejection						
$\mathcal{D}_{\mathcal{M}\setminus 7}$	7	384 ± 7	26.4 ± 0.1	291 ± 5	1620 ± 60	+320%
$\mathcal{D}_{\mathcal{M}\setminus 10}$	10	890 ± 20	72.1 ± 0.6	1830 ± 70	2300 ± 100	+160%
$\mathcal{D}_{\mathcal{M}\setminus 25}$	25	371 ± 7	5310 ± 10	67.5 ± 0.5	5100 ± 300	+1270%
$\mathcal{D}_{\mathcal{M}\setminus 40}$	40	10000 ± 1000	820 ± 20	32000 ± 5000	23000 ± 3000	+220%
85% WP background rejection						
$\mathcal{D}_{\mathcal{M}\setminus 7}$	7	282 ± 4	22.1 ± 0.1	196 ± 2	630 ± 10	120%
$\mathcal{D}_{\mathcal{M}\setminus 10}$	10	680 ± 20	56 ± 0.4	940 ± 30	1350 ± 50	+99%
$\mathcal{D}_{\mathcal{M}\setminus 25}$	25	286 ± 4	352 ± 6	62.3 ± 0.4	2070 ± 90	+620%
$\mathcal{D}_{\mathcal{M}\setminus 40}$	40	1900 ± 80	336 ± 6	2900 ± 100	3000 ± 100	+58%

Table 6.9: The background rejection at 70% and 85% signal efficiency working points for models trained using the training sets given in the *Dataset* column and evaluated at the mass point given in the *Mass* column, using jets with p_T greater than 300 GeV. The best performing model at a given working point is highlighted in bold. The final column shows the maximum percentage difference between the background rejection of the parameterised models (*Concatenation*, *Scaling* and *Affine*) and the *Baseline* model. Positive differences are prioritised even if there is a greater negative difference. The binomial uncertainties are shown.

Each LLP mass m_a represents a unique classification task. Three architectures incorporating mechanisms to condition the output of GN2 on m_a have been explored. These include a standard input concatenation approach, as well as feature-wise transformations such as conditional scaling and affine transformations, the latter for the first time at ATLAS. The parameterised networks are found to generally improve performance when interpolating to unseen masses, as well as when classifying jets from LLPs with masses seen during training, when compared to the baseline architecture. Background rejection increases up to 72-88% and 67-76% at the 70% and 85% WPs, respectively, when interpolating to $m_a = 7-10$ GeV. Background rejection increases up to 32-90% and 21-100% at the 70% and 85% WPs, respectively, for mass points ranging from $m_a = 5-16$ GeV, after training with the full parameter dependent training set. At higher masses, improvements are generally observed for lower signal efficiencies only, with small losses in performance at higher signal efficiencies. This may be an artefact of the kinematic resampling procedure.

Each parameterised model is found to give comparative performance. The simplicity of the *Concatenation* model, along with the fact its performance within the resampled p_T range is similar to (and sometimes stronger than) the more complex models featuring

6.2. GRAPH NEURAL NETWORKS FOR LONG-LIVED PARTICLE IDENTIFICATION

Mass (GeV)	<i>Baseline</i>	<i>Concatenation</i>	<i>Scaling</i>	<i>Affine</i>	Max. $\pm\%$ diff. from <i>Baseline</i>
70% WP background rejection					
5	1410 ± 50	308 ± 5	830 ± 20	2020 ± 80	+43%
7	1520 ± 50	310 ± 5	1510 ± 50	2080 ± 90	+37%
10	1650 ± 60	304 ± 5	1390 ± 50	2060 ± 90	+25%
16	1740 ± 70	258 ± 4	1880 ± 70	2500 ± 100	+44%
25	3700 ± 200	306 ± 5	3500 ± 200	4500 ± 300	+22%
35	6300 ± 500	413 ± 8	5000 ± 300	24000 ± 3000	+280%
40	6300 ± 500	423 ± 8	9000 ± 800	35000 ± 6000	+450%
85% WP background rejection					
5	930 ± 30	231 ± 3	404 ± 7	1150 ± 40	+27%
7	1020 ± 30	230 ± 3	900 ± 20	1250 ± 40	+23%
10	1120 ± 30	240 ± 3	820 ± 20	1200 ± 40	+7.1%
16	1270 ± 40	204 ± 3	1220 ± 40	1450 ± 50	+14%
25	1660 ± 60	226 ± 3	1660 ± 60	2060 ± 90	+24%
35	1790 ± 70	249 ± 4	1630 ± 60	2500 ± 100	+40%
40	1690 ± 60	259 ± 4	2200 ± 100	2800 ± 100	+66%

Table 6.10: The background rejection at 70% and 85% signal efficiency working points for models trained using $\mathcal{D}_M$ and evaluated at the mass point given in the *Mass* column, using jets with p_T greater than 300 GeV. The best performing model at a given working point is highlighted in bold. The final column shows the maximum percentage difference between the background rejection of the parameterised models (*Concatenation*, *Scaling* and *Affine*) and the *Baseline* model. Positive differences are prioritised even if there is a greater negative difference. Rejections for 55 GeV are omitted as the statistics are too low to form a reliable comparison. The binomial uncertainties are shown.

feature-wise transformations, may make it the preferred architecture for use in analyses. However, the *Concatenation* model is least effective at generalising to jets with p_T outside the resampled range, with the *Affine* model consistently demonstrating the strongest performance in this regard. Should alternative resampling procedures be explored in the future, it will be worthwhile to continue to study the properties of each conditioning mechanism proposed here.

CONCLUSION

The Standard Model (SM) is an incomplete theory that leaves a number of questions unanswered. The discovery of a new particle would represent a major breakthrough in our understanding of physics beyond the SM (BSM). Many BSM models contain new long-lived particles (LLPs). LLPs have macroscopic decay lengths, meaning they may elude standard searches for promptly decaying particles. However their unique detector signatures make them attractive targets in dedicated searches for new physics.

This thesis presented a search for light, pseudoscalar LLPs using 140 fb^{-1} of pp collision data with $\sqrt{s} = 13 \text{ TeV}$ recorded at the ATLAS detector between 2015 and 2018. The search targeted LLPs with mass between 5 and 55 GeV that decay hadronically within the ATLAS inner detector (ID), producing the distinct signature of displaced jets and displaced vertices. The main benchmark model considered by the analysis was that of an exotic decay of the Higgs boson to pairs of LLPs, $H \rightarrow aa$. Additionally, models featuring axion-like particles (ALPs) were considered. Although no significant excess over the expected background was observed, stringent upper limits were set on branching ratio of the Higgs boson to pairs of LLPs, the cross-section for ALPs produced in association with a vector boson, and, for the first time, on the branching ratio of the top quark to an ALP and a u/c quark. The limits presented for the $H \rightarrow aa$ model are the most stringent to date for LLP masses $m_a < 40 \text{ GeV}$ and lifetimes $10^{-1} - 10^{-3} \text{ m}$.

The high-luminosity phase of the LHC is expected to begin in 2029 with an instantaneous luminosity of $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and average pile-up of 200. To cope with the increased number of interactions, major upgrades are planned for the ATLAS detectors.

This includes a new, all-silicon Inner Tracker (ITk) to replace the current ID. The ITk will feature higher granularity, improved radiation hardness, faster readout electronics, and a longer trigger latency. The ITk tracking performance is expected to be comparable or better than during Run 2, despite the increase in pile-up [122]. The ITk will also extend tracking coverage to $\eta = 4$, increasing the acceptance of processes like VBF. These improvements are expected to increase the overall acceptance of searches for displaced vertices in the tracker.

While the advent of the high-luminosity LHC promises even larger statistics with which to hunt for signs of new physics, improving the sensitivity, and therefore discovery potential, of future LLP searches will also rely on the development and optimisation of novel analysis techniques that best exploit this data. This thesis focused on two areas of development: the performance of large radius tracking (LRT) in the ATLAS trigger, and transformer-based graph neural networks (GNNs) for identifying displaced jets.

The LRT algorithm was recently integrated within the ATLAS trigger system for the first time. In this thesis, the performance of the LRT algorithm in the ATLAS trigger was studied using K_S^0 vertex candidates. The strong performance motivated the inclusion of LRT-based triggers that use the entire ID tracking volume in the ATLAS trigger menu for Run 3. Such triggers would enhance the selection of events in searches that rely on signatures such as displaced vertices.

The GN2 jet tagger algorithm was adapted to identify jets from $H \rightarrow aa$ decays. The algorithm was found to improve background rejection by up to $\sim 10\text{--}15\times$, when compared to the boosted decision tree that was used in the search herein. The baseline architecture was also extended to produce three new parameterised models that incorporated different conditioning mechanisms, including feature-wise transformations. The conditioned networks were tested on their ability to classify jets associated to LLPs with masses that were used in training, as well as interpolate to new masses. They were found to generally improve background rejection in both cases across a range of possible LLP masses. Issues arising due to the kinematic resampling procedure used were highlighted and discussed. This work demonstrated the substantial gains that could be achieved by incorporating GNN-based taggers in future LLP searches, and lays the groundwork for

a number of further studies featuring conditioning mechanisms.

BIBLIOGRAPHY

- [1] ATLAS Collaboration. “Search for light long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV using displaced vertices in the ATLAS inner detector”. 2024. arXiv: 2403.15332 [hep-ex].
- [2] ATLAS Collaboration. “The ATLAS Trigger System for LHC Run 3 and Trigger performance in 2022”. 2024. arXiv: 2401.06630 [hep-ex].
- [3] J. Barr et al. “Salt: Multimodal Multitask Machine Learning for High Energy Physics”. (2024). <https://cds.cern.ch/record/2886492>.
- [4] M. D. Schwartz. “Quantum Field Theory and the Standard Model”. Cambridge University Press, 2014. ISBN: 978-1-107-03473-0.
- [5] M. Thomson. “Modern Particle Physics”. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525367.
- [6] P. Langacker. “The Standard Model and Beyond”. CRC Press, 2009. ISBN: 9781420079074.
- [7] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. *Phys. Lett. B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.020>.
- [8] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. *Phys. Lett. B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [9] ATLAS Collaboration. “A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery”. *Nature* 607.7917 (2022), pp. 52–59. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex].

- [10] ATLAS Collaboration. “Measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data. Measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data”. *Phys. Rev. D* 90 (2014), p. 052004. DOI: 10.1103/PhysRevD.90.052004.
- [11] CERN. “CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector”. (2017). DOI: 10.23731/CYRM-2017-002.
- [12] Particle Data Group. “Review of Particle Physics”. *PTEP* 2022.8 (2022), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097.
- [13] M. Roos. “Dark Matter: The evidence from astronomy, astrophysics and cosmology”. 2010. arXiv: 1001.0316 [astro-ph.CO].
- [14] D. S. Akerib et al. “Results from a Search for Dark Matter in the Complete LUX Exposure”. *Phys. Rev. Lett.* 118 (2 2017), p. 021303. DOI: 10.1103/PhysRevLett.118.021303.
- [15] C. Pérez de los Heros. “Status, Challenges and Directions in Indirect Dark Matter Searches”. *Symmetry* 12.10 (2020), p. 1648. ISSN: 2073-8994. DOI: 10.3390/sym12101648.
- [16] J. Alimena et al. “Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider”. *J. Phys. G Nucl. Part. Phys.* 47.9 (Sept. 2020), p. 090501. DOI: 10.1088/1361-6471/ab4574.
- [17] D. Curtin et al. “Long-lived particles at the energy frontier: the MATHUSLA physics case”. *Rep. Prog. Phys.* 82.11 (2019), p. 116201. DOI: 10.1088/1361-6633/ab28d6.
- [18] L. Lee et al. “Collider searches for long-lived particles beyond the Standard Model”. *Prog. Part. Nucl. Phys.* 106 (2019), pp. 210–255. ISSN: 0146-6410. DOI: <https://doi.org/10.1016/j.pnpnp.2019.02.006>.
- [19] S. Knapen and S. Lowette. “A guide to hunting long-lived particles at the LHC”. 2022. arXiv: 2212.03883 [hep-ph].

- [20] H. Russell. “Collimated hadronic ID decay”. Licensed under CC BY 4.0. URL: https://hrussell.web.cern.ch/hrussell/images/graphics/LLP_IDHadronic_1jet.png.
- [21] H. Russell. “Non-collimated hadronic ID decay”. Licensed under CC BY 4.0. URL: https://hrussell.web.cern.ch/hrussell/images/graphics/LLP_IDHadronic_2jets.png.
- [22] D. Curtin et al. “Exotic decays of the 125 GeV Higgs boson”. *Phys. Rev. D* 90.7 (2014). DOI: 10.1103/physrevd.90.075004.
- [23] M. Cepeda et al. “Exotic Higgs Decays”. *Annu. Rev. Nucl. Sci.* 72.1 (2022), pp. 119–149. DOI: 10.1146/annurev-nucl-102319-024147.
- [24] Donal O’Connell, Michael J. Ramsey-Musolf, and Mark B. Wise. “Minimal extension of the standard model scalar sector”. *Phys. Rev. D* 75.3 (2007). DOI: 10.1103/physrevd.75.037701.
- [25] G. Burdman et al. “Folded supersymmetry and the LEP paradox”. *JHEP* 2007.02 (Feb. 2007), p. 009. DOI: 10.1088/1126-6708/2007/02/009.
- [26] H. Cai, H. Cheng, and J. Terning. “A quirky little Higgs model”. *JHEP* 2009.05 (May 2009), p. 045. DOI: 10.1088/1126-6708/2009/05/045.
- [27] L. Evans and P. Bryant. “LHC Machine”. *JINST* 3.08 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.
- [28] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. *JINST* 3.08 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.
- [29] CMS Collaboration. “The CMS experiment at the CERN LHC”. *JINST* 3.08 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.
- [30] LHCb Collaboration. “The LHCb Detector at the LHC”. *JINST* 3.08 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.
- [31] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. *JINST* 3.08 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

- [32] M. Vretenar et al. “Linac4 design report”. Vol. 6. CERN Yellow Reports: Monographs. 2020. doi: 10.23731/CYRM-2020-006.
- [33] E. Lopienska. “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”. (2022). <https://cds.cern.ch/record/2800984>.
- [34] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. Tech. rep. <https://cds.cern.ch/record/2677054>. CERN, 2019.
- [35] W. Herr and B. Muratori. “Concept of luminosity”. (2006). doi: 10.5170/CERN-2006-002.361.
- [36] ATLAS Collaboration. “Public ATLAS Luminosity Results for Run-2 of the LHC”. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [37] ATLAS Collaboration. “Public ATLAS Luminosity Results for Run-3 of the LHC”. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>.
- [38] ATLAS Collaboration. “Study of the material of the ATLAS inner detector for Run 2 of the LHC”. *JINST* 12.12 (2017), P12009–P12009. doi: 10.1088/1748-0221/12/12/p12009.
- [39] B. Abbott et al. “Production and integration of the ATLAS Insertable B-Layer”. *JINST* 13.05 (2018), T05008. doi: 10.1088/1748-0221/13/05/T05008.
- [40] T. Heim. “Status and performance of the ATLAS Pixel Detector after 3 years of operation”. *Nucl. Instrum. Methods Phys. Res. A* 765 (2014), pp. 227–231. issn: 0168-9002. doi: <https://doi.org/10.1016/j.nima.2014.04.058>.
- [41] S. Lee. “On the limits of the hadronic energy resolution of calorimeters”. *JPCS* 1162.1 (2019), p. 012043. doi: 10.1088/1742-6596/1162/1/012043.
- [42] H. Kolanoski and N. Wermes. “Interactions of particles with matter”. *Particle Detectors: Fundamentals and Applications*. Oxford University Press, 2020. doi: 10.1093/oso/9780198858362.003.0003.
- [43] ATLAS Collaboration. “Operation of the ATLAS trigger system in Run 2”. *JINST* 15.10 (2020), P10004–P10004. doi: 10.1088/1748-0221/15/10/p10004.

- [44] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. *EPJ C* 70.3 (2010), pp. 823–874. DOI: 10.1140/epjc/s10052-010-1429-9.
- [45] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. *Computer Physics Communications* 191 (2015), pp. 159–177. DOI: <https://doi.org/10.1016/j.cpc.2015.01.024>.
- [46] J. Bellm et al. “Herwig 7.2 release note”. *EPJ C* 80.5 (2020). DOI: 10.1140/epjc/s10052-020-8011-x.
- [47] S. Agostinelli et al. “Geant4—a simulation toolkit”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [48] T. Cornelissen et al. “The new ATLAS track reconstruction (NEWT)”. *JPCS* 119.3 (2008), p. 032014. DOI: 10.1088/1742-6596/119/3/032014.
- [49] ATLAS Collaboration. “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2”. *EPJ C* 77.10 (2017), p. 673. DOI: 10.1140/epjc/s10052-017-5225-7.
- [50] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. DOI: [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [51] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. *EPJ C* 77.5 (2017). <https://cds.cern.ch/record/2235651>, p. 332.
- [52] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeter and its performance in LHC Run 1”. *EPJ C* 77 (2017). <https://cds.cern.ch/record/2138166>, p. 490.
- [53] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. *EPJ C* 77.7 (2017). <https://cds.cern.ch/record/2257597>, p. 466.

- [54] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-kt jet clustering algorithm”. *JHEP* 2008.04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063.
- [55] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *EPJ C* 8 (2021). <https://cds.cern.ch/record/2722869>, p. 689.
- [56] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Phys. Rev. D* 96.7 (2017). <https://cds.cern.ch/record/2257300>, p. 072002.
- [57] ATLAS Collaboration. “ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset”. (2022). arXiv: 2211.16345 [physics.data-an].
- [58] ATLAS Collaboration. “Graph Neural Network Jet Flavour Tagging with the ATLAS Detector”. ATL-PHYS-PUB-2022-027 (2022). <https://cds.cern.ch/record/2811135>.
- [59] 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”. *EPJ C* 79.8 (2019). DOI: 10.1140/epjc/s10052-019-7140-6.
- [60] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data”. *JINST* 14.12 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006.
- [61] ATLAS Collaboration. “Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung”. ATLAS-CONF-2012-047 (2012). <https://cds.cern.ch/record/1449796>.
- [62] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data”. *JINST* 14.03 (2019), P03017–P03017. DOI: 10.1088/1748-0221/14/03/p03017.
- [63] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV”. ATL-PHYS-PUB-2016-015 (2016). <https://cds.cern.ch/record/2203514>.

- [64] ATLAS Collaboration. “Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016”. *EPJ C* 79.3 (2019). doi: 10.1140/epjc/s10052-019-6650-6.
- [65] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV”. *EPJ C* 81 (2021), p. 578. arXiv: 2012.00578 [hep-ex].
- [66] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV”. *EPJ C* 76.5 (2016). doi: 10.1140/epjc/s10052-016-4120-y.
- [67] ATLAS Collaboration. “Studies of the muon momentum calibration and performance of the ATLAS detector with pp collisions at $\sqrt{s} = 13$ TeV”. (2022). arXiv: 2212.07338 [hep-ex].
- [68] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV”. *EPJ C* 78.11 (2018). doi: 10.1140/epjc/s10052-018-6288-9.
- [69] ATLAS Collaboration. “Performance of the reconstruction of large impact parameter tracks in the ATLAS inner detector”. ATL-PHYS-PUB-2017-014 (2017). <https://cds.cern.ch/record/2275635>.
- [70] ATLAS Collaboration. “Performance of the reconstruction of large impact parameter tracks in the inner detector of ATLAS”. (2023). arXiv: 2304.12867 [hep-ex].
- [71] ATLAS Collaboration. “Expected Large-Radius Tracking Performance in Run 3”. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2021-003/>.
- [72] ATLAS Collaboration. “Performance of vertex reconstruction algorithms for detection of new long-lived particle decays within the ATLAS inner detector”. ATL-PHYS-PUB-2019-013 (2019). <https://cds.cern.ch/record/2669425>.

- [73] ATLAS Collaboration. “Search for long-lived, weakly interacting particles that decay to displaced hadronic jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. *Phys. Rev. D* 92.1 (2015). ISSN: 1550-2368. DOI: 10.1103/physrevd.92.012010.
- [74] ATLAS Collaboration. “Search for pair-produced long-lived neutral particles decaying to jets in the ATLAS hadronic calorimeter in pp collisions at $\sqrt{s} = 8$ TeV”. *Phys. Lett. B* 743 (2015), pp. 15–34. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2015.02.015.
- [75] ATLAS Collaboration. “Search for long-lived particles produced in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS muon spectrometer”. *Phys. Rev. D* 99 (5 2019), p. 052005. DOI: 10.1103/PhysRevD.99.052005.
- [76] ATLAS Collaboration. “Search for long-lived neutral particles in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS calorimeter”. *EPJ C* 79.6 (2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6962-6.
- [77] ATLAS Collaboration. “Search for exotic decays of the Higgs boson into long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV using displaced vertices in the ATLAS inner detector”. *JHEP* 11 (2021), p. 229. DOI: 10.1007/JHEP11(2021)229. arXiv: 2107.06092 [hep-ex].
- [78] ATLAS Collaboration. “Long-lived particle summary plots for Hidden Sector and Dark Photon models”. ATL-PHYS-PUB-2022-007 (2022). <https://cds.cern.ch/record/2803320>.
- [79] S. Weinberg. “A New Light Boson?”. *Phys. Rev. Lett.* 40 (4 Jan. 1978), pp. 223–226. DOI: 10.1103/PhysRevLett.40.223.
- [80] G. Cacciapaglia et al. “Light scalars in composite Higgs models”. *Front. in Phys.* 7 (2019), p. 22. DOI: 10.3389/fphy.2019.00022. arXiv: 1902.06890 [hep-ph].
- [81] H. Cheng, L. Li, and E. Salvioni. “A theory of dark pions”. *JHEP* 2022.1 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep01(2022)122.

- [82] I. Brivio et al. “ALPs effective field theory and collider signatures”. *EPJ C* 77.8 (2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5111-3.
- [83] A. Carmona et al. “The ALPs from the top: searching for long lived axion-like particles from exotic top decays”. *JHEP* 2022.7 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep07(2022)122.
- [84] J. Catmore et al. “A New Petabyte-scale Data Derivation Framework for ATLAS”. Tech. rep. 7. Geneva: CERN, 2015. DOI: 10.1088/1742-6596/664/7/072007.
- [85] P. Nason and C. Oleari. “NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG”. *JHEP* 2010.2 (2010). ISSN: 1029-8479. DOI: 10.1007/jhep02(2010)037.
- [86] S. Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. *JHEP* 2010.6 (2010). ISSN: 1029-8479. DOI: 10.1007/jhep06(2010)043.
- [87] R. D. Ball et al. “Parton distributions for the LHC Run II”. *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040.
- [88] ATLAS Collaboration. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. *JHEP* 2014.9 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)145.
- [89] 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* 2014.7 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079.
- [90] Y. Coadou. “Boosted Decision Trees”. *Artificial Intelligence for High Energy Physics*. World Scientific, 2022, pp. 9–58. DOI: 10.1142/9789811234033_0002.
- [91] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785.

- [92] B. P. Brau et al. “Search for light long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV using displaced vertices in the ATLAS inner detector”. (2021). <https://cds.cern.ch/record/2783923>.
- [93] ATLAS Collaboration. “Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment”. *JHEP* 2024.5 (2024). ISSN: 1029-8479. DOI: 10.1007/jhep05(2024)162.
- [94] K. Cranmer. “Practical Statistics for the LHC”. 2015. arXiv: 1503.07622 [physics.data-an].
- [95] W. Verkerke and D. Kirkby. “The RooFit toolkit for data modeling”. 2003. arXiv: physics/0306116 [physics.data-an].
- [96] A L Read. “Presentation of search results: the CLs technique”. *J. Phys. G Nucl. Part. Phys.* 28.10 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.
- [97] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. *EPJ C* 71.2 (2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0.
- [98] “TRexFitter framework documentation”. (2024). <https://trexfitter-docs.web.cern.ch/trexfitter-docs/>.
- [99] “TRexFitter framework GitLab project”. (2024). <https://gitlab.cern.ch/TRExStats/TRExFitter>.
- [100] ATLAS Collaboration. “Search for long-lived particles in final states with displaced dimuon vertices in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Phys. Rev. D* 99.1 (2019), p. 012001. DOI: 10.1103/PhysRevD.99.012001. arXiv: 1808.03057.
- [101] ATLAS Collaboration. “ K_s^0 and Λ production in pp interactions at $\sqrt{s}=0.9$ and 7 TeV measured with the ATLAS detector at the LHC”. *Phys. Rev. D* 85 (1 2012), p. 012001. DOI: 10.1103/PhysRevD.85.012001.
- [102] ATLAS Collaboration. “Public plots for MC/MC and Data/MC comparisons of D11d and GN1, and simulation performance of GN2”. URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/>.

- [103] P. Baldi et al. “Parameterized neural networks for high-energy physics”. *EPJ C* 76.5 (Apr. 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4099-4.
- [104] I. J. Goodfellow, Y. Bengio, and A. Courville. “Deep Learning”. <http://www.deeplearningbook.org>. Cambridge, MA, USA: MIT Press, 2016.
- [105] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. *Nature* 323 (1986). <https://api.semanticscholar.org/CorpusID:205001834>, pp. 533–536.
- [106] I. Loshchilov and F. Hutter. “Decoupled Weight Decay Regularization”. 2019. arXiv: 1711.05101 [cs.LG].
- [107] A. F. Agarap. “Deep Learning using Rectified Linear Units (ReLU)”. 2019. arXiv: 1803.08375 [cs.NE].
- [108] M. Gori, G. Monfardini, and F. Scarselli. “A new model for learning in graph domains”. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*. Vol. 2. 2005, 729–734 vol. 2. DOI: 10.1109/IJCNN.2005.1555942.
- [109] F. Scarselli et al. “The Graph Neural Network Model”. *IEEE Trans. Neural Netw* 20.1 (2009), pp. 61–80. DOI: 10.1109/TNN.2008.2005605.
- [110] P. Battaglia et al. “Relational inductive biases, deep learning, and graph networks”. (2018). arXiv: 1806.01261 [cs.LG].
- [111] J. Shlomi, P. Battaglia, and J. Vlimant. “Graph neural networks in particle physics”. *MLST* 2.2 (Dec. 2020), p. 021001. DOI: 10.1088/2632-2153/abfb9a.
- [112] ATLAS Collaboration. “Search for a new scalar resonance in flavour-changing neutral-current top-quark decays $t \rightarrow qX$ ($q = u, c$), with $X \rightarrow b\bar{b}$, in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *JHEP* 2023.7 (2023). ISSN: 1029-8479. DOI: 10.1007/jhep07(2023)199.
- [113] ATLAS Collaboration. “Search for a new heavy scalar particle decaying into a Higgs boson and a new scalar singlet in final states with one or two light leptons and a pair of τ -leptons with the ATLAS detector”. *JHEP* 2023.10 (2023). ISSN: 1029-8479. DOI: 10.1007/jhep10(2023)009.

- [114] V. Dumoulin et al. “Feature-wise transformations”. *Distill* (2018). DOI: 10.23915/distill.00011.
- [115] L. Anzalone, T. Diotallevi, and D. Bonacorsi. “Improving parametric neural networks for high-energy physics (and beyond)”. *MLST* 3.3 (2022), p. 035017. ISSN: 2632-2153. DOI: 10.1088/2632-2153/ac917c.
- [116] M. Cacciari, G. P. Salam, and G. Soyez. “The catchment area of jets”. *JHEP* 2008.04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005.
- [117] S. Van Stroud. “Graph Neural Network Flavour Tagging and Boosted Higgs Measurements at the LHC”. <https://cds.cern.ch/record/2869719>. 2023.
- [118] A. Vaswani et al. “Attention Is All You Need”. 2023. arXiv: 1706.03762 [cs.CL].
- [119] K. He et al. “Deep Residual Learning for Image Recognition”. 2015. arXiv: 1512.03385 [cs.CV].
- [120] J. L. Ba, J. R. Kiros, and G. E. Hinton. “Layer Normalization”. 2016. arXiv: 1607.06450 [stat.ML].
- [121] L. N. Smith and N. Topin. “Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates”. (2018). arXiv: 1708.07120 [cs.LG].
- [122] ATLAS Collaboration. “Expected tracking and related performance with the updated ATLAS Inner Tracker layout at the High-Luminosity LHC”. ATL-PHYS-PUB-2021-024 (2021). <https://cds.cern.ch/record/2776651>.



MONTE CARLO SAMPLES

The exotic Higgs decay and SM background MC samples used in the LLP search presented in §5 and the GNN studies presented in §6.2 are detailed here. Sample are generated for each MC campaign: MC20a, MC20d, and MC20e.

Table A.1 lists the $H \rightarrow aa$ signal MC samples used in the LLP search. Samples from the MC20e campaign with means LLP lifetimes of 10 and 100 mm are used in the GNN studies. Table A.2 lists the SM background MC samples used in the LLP search.

Table A.3 lists the additional $H \rightarrow aa$ signal MC samples that are used in the GNN studies. Only samples from the MC20e campaign are used.

DSID	Process Name
601245	PhPy8_WpH_H125_a5a5_4u_ctau10
601246	PhPy8_WpH_H125_a16a16_4u_ctau10
601247	PhPy8_WpH_H125_a16a16_4b_ctau10
601248	PhPy8_WpH_H125_a16a16_4b_ctau100
601249	PhPy8_WpH_H125_a55a55_4u_ctau10
601250	PhPy8_WpH_H125_a55a55_4b_ctau10
601252	PhPy8_WmH_H125_a5a5_4u_ctau10
601253	PhPy8_WmH_H125_a16a16_4u_ctau10
601254	PhPy8_WmH_H125_a16a16_4b_ctau10
601255	PhPy8_WmH_H125_a16a16_4b_ctau100
601256	PhPy8_WmH_H125_a55a55_4u_ctau10
601257	PhPy8_WmH_H125_a55a55_4b_ctau10
601258	PhPy8_WmH_H125_a55a55_4b_ctau100
602106	PhPy8_WpH_H125_a5a5_4u_ctau1

602107 PhPy8_WpH_H125_a5a5_4u_ctau100
602108 PhPy8_WpH_H125_a5a5_4u_ctau1000
602109 PhPy8_WpH_H125_a5a5_4c_ctau1
602110 PhPy8_WpH_H125_a5a5_4c_ctau10
602111 PhPy8_WpH_H125_a5a5_4c_ctau100
602112 PhPy8_WpH_H125_a5a5_4c_ctau1000
602113 PhPy8_WpH_H125_a16a16_4u_ctau1
602114 PhPy8_WpH_H125_a16a16_4u_ctau100
602115 PhPy8_WpH_H125_a16a16_4u_ctau1000
602116 PhPy8_WpH_H125_a16a16_4b_ctau1
602117 PhPy8_WpH_H125_a16a16_4b_ctau1000
602118 PhPy8_WpH_H125_a40a40_4u_ctau1
602119 PhPy8_WpH_H125_a40a40_4u_ctau10
602120 PhPy8_WpH_H125_a40a40_4u_ctau100
602121 PhPy8_WpH_H125_a40a40_4u_ctau1000
602122 PhPy8_WpH_H125_a40a40_4b_ctau1
602123 PhPy8_WpH_H125_a40a40_4b_ctau10
602124 PhPy8_WpH_H125_a40a40_4b_ctau100
602125 PhPy8_WpH_H125_a40a40_4b_ctau1000
602126 PhPy8_WpH_H125_a55a55_4u_ctau1
602127 PhPy8_WpH_H125_a55a55_4u_ctau100
602128 PhPy8_WpH_H125_a55a55_4u_ctau1000
602129 PhPy8_WpH_H125_a55a55_4b_ctau1
602130 PhPy8_WpH_H125_a55a55_4b_ctau1000
602131 PhPy8_WmH_H125_a5a5_4u_ctau1
602132 PhPy8_WmH_H125_a5a5_4u_ctau100
602133 PhPy8_WmH_H125_a5a5_4u_ctau1000
602134 PhPy8_WmH_H125_a5a5_4c_ctau1
602135 PhPy8_WmH_H125_a5a5_4c_ctau10
602136 PhPy8_WmH_H125_a5a5_4c_ctau100
602137 PhPy8_WmH_H125_a5a5_4c_ctau1000
602138 PhPy8_WmH_H125_a16a16_4u_ctau1
602139 PhPy8_WmH_H125_a16a16_4u_ctau100
602140 PhPy8_WmH_H125_a16a16_4u_ctau1000
602141 PhPy8_WmH_H125_a16a16_4b_ctau1
602142 PhPy8_WmH_H125_a16a16_4b_ctau1000
602143 PhPy8_WmH_H125_a40a40_4u_ctau1

602144 PhPy8_WmH_H125_a40a40_4u_ctau10
602145 PhPy8_WmH_H125_a40a40_4u_ctau100
602146 PhPy8_WmH_H125_a40a40_4u_ctau1000
602147 PhPy8_WmH_H125_a40a40_4b_ctau1
602148 PhPy8_WmH_H125_a40a40_4b_ctau10
602149 PhPy8_WmH_H125_a40a40_4b_ctau100
602150 PhPy8_WmH_H125_a40a40_4b_ctau1000
602151 PhPy8_WmH_H125_a55a55_4u_ctau1
602152 PhPy8_WmH_H125_a55a55_4u_ctau100
602153 PhPy8_WmH_H125_a55a55_4u_ctau1000
602154 PhPy8_WmH_H125_a55a55_4b_ctau1
602155 PhPy8_WmH_H125_a55a55_4b_ctau1000
601259 PhPy8_ZllH_H125_a5a5_4u_ctau10
601260 PhPy8_ZllH_H125_a16a16_4u_ctau10
601261 PhPy8_ZllH_H125_a16a16_4b_ctau10
601262 PhPy8_ZllH_H125_a16a16_4b_ctau100
601263 PhPy8_ZllH_H125_a55a55_4u_ctau10
601264 PhPy8_ZllH_H125_a55a55_4b_ctau10
601265 PhPy8_ZllH_H125_a55a55_4b_ctau100
601266 PhPy8_ggZllH_H125_a5a5_4u_ctau10
601267 PhPy8_ggZllH_H125_a16a16_4u_ctau10
601268 PhPy8_ggZllH_H125_a16a16_4b_ctau10
601269 PhPy8_ggZllH_H125_a16a16_4b_ctau100
601269 PhPy8_ggZllH_H125_a16a16_4b_ctau100
601270 PhPy8_ggZllH_H125_a55a55_4u_ctau10
601271 PhPy8_ggZllH_H125_a55a55_4b_ctau10
601272 PhPy8_ggZllH_H125_a55a55_4b_ctau100
602156 PhPy8_ZllH_H125_a5a5_4u_ctau1
602157 PhPy8_ZllH_H125_a5a5_4u_ctau100
602158 PhPy8_ZllH_H125_a5a5_4u_ctau1000
602159 PhPy8_ZllH_H125_a5a5_4c_ctau1
602160 PhPy8_ZllH_H125_a5a5_4c_ctau10
602161 PhPy8_ZllH_H125_a5a5_4c_ctau100
602162 PhPy8_ZllH_H125_a5a5_4c_ctau1000
602163 PhPy8_ZllH_H125_a16a16_4u_ctau1
602164 PhPy8_ZllH_H125_a16a16_4u_ctau100
602165 PhPy8_ZllH_H125_a16a16_4u_ctau1000

602166 PhPy8_ZllH_H125_a16a16_4b_ctau1
 602167 PhPy8_ZllH_H125_a16a16_4b_ctau1000
 602168 PhPy8_ZllH_H125_a40a40_4u_ctau1
 602169 PhPy8_ZllH_H125_a40a40_4u_ctau10
 602170 PhPy8_ZllH_H125_a40a40_4u_ctau100
 602171 PhPy8_ZllH_H125_a40a40_4u_ctau1000
 602172 PhPy8_ZllH_H125_a40a40_4b_ctau1
 602173 PhPy8_ZllH_H125_a40a40_4b_ctau10
 602174 PhPy8_ZllH_H125_a40a40_4b_ctau100
 602175 PhPy8_ZllH_H125_a40a40_4b_ctau1000
 602176 PhPy8_ZllH_H125_a55a55_4u_ctau1
 602177 PhPy8_ZllH_H125_a55a55_4u_ctau100
 602178 PhPy8_ZllH_H125_a55a55_4u_ctau1000
 602179 PhPy8_ZllH_H125_a55a55_4b_ctau1
 602180 PhPy8_ZllH_H125_a55a55_4b_ctau1000
 602181 PhPy8_ggZllH_H125_a5a5_4u_ctau1
 602182 PhPy8_ggZllH_H125_a5a5_4u_ctau100
 602183 PhPy8_ggZllH_H125_a5a5_4u_ctau1000
 602184 PhPy8_ggZllH_H125_a5a5_4c_ctau1
 602185 PhPy8_ggZllH_H125_a5a5_4c_ctau10
 602186 PhPy8_ggZllH_H125_a5a5_4c_ctau100
 602187 PhPy8_ggZllH_H125_a5a5_4c_ctau1000
 602188 PhPy8_ggZllH_H125_a16a16_4u_ctau1
 602189 PhPy8_ggZllH_H125_a16a16_4u_ctau100
 602190 PhPy8_ggZllH_H125_a16a16_4u_ctau1000
 602191 PhPy8_ggZllH_H125_a16a16_4b_ctau1
 602192 PhPy8_ggZllH_H125_a16a16_4b_ctau1000
 602193 PhPy8_ggZllH_H125_a40a40_4u_ctau1
 602194 PhPy8_ggZllH_H125_a40a40_4u_ctau10
 602195 PhPy8_ggZllH_H125_a40a40_4u_ctau100
 602196 PhPy8_ggZllH_H125_a40a40_4u_ctau1000
 602197 PhPy8_ggZllH_H125_a40a40_4b_ctau1
 602198 PhPy8_ggZllH_H125_a40a40_4b_ctau10
 602199 PhPy8_ggZllH_H125_a40a40_4b_ctau100
 602200 PhPy8_ggZllH_H125_a40a40_4b_ctau1000
 602201 PhPy8_ggZllH_H125_a55a55_4u_ctau1
 602202 PhPy8_ggZllH_H125_a55a55_4u_ctau100

602203	PhPy8_ggZllH_H125_a55a55_4u_ctau1000
602204	PhPy8_ggZllH_H125_a55a55_4b_ctau1
602205	PhPy8_ggZllH_H125_a55a55_4b_ctau1000
600914	PhPy8_VBF_H125_a16a16_4b_ctau10
600915	PhPy8_VBF_H125_a16a16_4b_ctau100
600916	PhPy8_VBF_H125_a16a16_4b_ctau1000
600918	PhPy8_VBF_H125_a55a55_4b_ctau10
600919	PhPy8_VBF_H125_a55a55_4b_ctau100
600920	PhPy8_VBF_H125_a55a55_4b_ctau1000
602206	PhPy8_VBF_H125_a5a5_4u_ctau1
602207	PhPy8_VBF_H125_a5a5_4u_ctau10
602208	PhPy8_VBF_H125_a5a5_4u_ctau100
602209	PhPy8_VBF_H125_a5a5_4u_ctau1000
602210	PhPy8_VBF_H125_a5a5_4c_ctau1
602211	PhPy8_VBF_H125_a5a5_4c_ctau10
602212	PhPy8_VBF_H125_a5a5_4c_ctau100
602213	PhPy8_VBF_H125_a5a5_4c_ctau1000
602214	PhPy8_VBF_H125_a16a16_4u_ctau1
602215	PhPy8_VBF_H125_a16a16_4u_ctau10
602216	PhPy8_VBF_H125_a16a16_4u_ctau100
602217	PhPy8_VBF_H125_a16a16_4u_ctau1000
602218	PhPy8_VBF_H125_a16a16_4b_ctau1
602219	PhPy8_VBF_H125_a40a40_4u_ctau1
602220	PhPy8_VBF_H125_a40a40_4u_ctau10
602221	PhPy8_VBF_H125_a40a40_4u_ctau100
602222	PhPy8_VBF_H125_a40a40_4u_ctau1000
602223	PhPy8_VBF_H125_a40a40_4b_ctau1
602224	PhPy8_VBF_H125_a40a40_4b_ctau10
602225	PhPy8_VBF_H125_a40a40_4b_ctau100
602226	PhPy8_VBF_H125_a40a40_4b_ctau1000
602227	PhPy8_VBF_H125_a55a55_4u_ctau1
602228	PhPy8_VBF_H125_a55a55_4u_ctau10
602229	PhPy8_VBF_H125_a55a55_4u_ctau100
602230	PhPy8_VBF_H125_a55a55_4u_ctau1000
602231	PhPy8_VBF_H125_a55a55_4b_ctau1

Table A.1: The dataset identifiers (DSIDs) and process names for signal MC samples used in the LLP search. The process name details the Higgs production mode, the mass of the Higgs boson, the mass of the two LLPs, the decay mode of the LLPs, and the mean LLP lifetime. Samples are generated for each MC campaign: MC20a, MC20d, and MC20e. Samples with mean lifetimes of 10 or 100 mm (MC20e only) are also used in the graph neural network studies in §6.2.

DSID	Process Name	GNN
364701	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1WithSW	*
364702	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2WithSW	*
364703	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3WithSW	*
364704	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4WithSW	*
364705	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5WithSW	*
364706	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6WithSW	
364707	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7WithSW	
364708	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8WithSW	
364709	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9WithSW	
364710	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10WithSW	
364711	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11WithSW	
364712	Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12WithSW	
410470	PhPy8EG_A14_ttbar_hdamp258p75_nonallhad	*
410471	PhPy8EG_A14_ttbar_hdamp258p75_allhad	
700320	Sh_2211_Zee_maxHTpTV2_BFilter	*
700321	Sh_2211_Zee_maxHTpTV2_CFilterBVeto	*
700322	Sh_2211_Zee_maxHTpTV2_CVetoBVeto	*
700323	Sh_2211_Zmumu_maxHTpTV2_BFilter	*
700324	Sh_2211_Zmumu_maxHTpTV2_CFilterBVeto	*
700325	Sh_2211_Zmumu_maxHTpTV2_CVetoBVeto	*
700338	Sh_2211_Wenu_maxHTpTV2_BFilter	*
700339	Sh_2211_Wenu_maxHTpTV2_CFilterBVeto	*
700340	Sh_2211_Wenu_maxHTpTV2_CVetoBVeto	*
700341	Sh_2211_Wmumu_maxHTpTV2_BFilter	*
700342	Sh_2211_Wmumu_maxHTpTV2_CFilterBVeto	*
700343	Sh_2211_Wmumu_maxHTpTV2_CVetoBVeto	*
800662	Py8_gammajet_direct_DP70_140_ordered	
800663	Py8_gammajet_direct_DP140_280_ordered	
800664	Py8_gammajet_direct_DP280_500_ordered	

800665 Py8_gammajet_direct_DP500_800_ordered
800666 Py8_gammajet_direct_DP800_1000_ordered

Table A.2: The dataset identifiers (DSIDs) and process names for background MC samples used in the LLP search. Samples are generated for each MC campaigns: MC20a, MC20d, and MC20e. Samples used in the graph neural network studies in §6.2 are marked with a * (MC20e only).

DSID	Process Name
603015	PhPy8_WpH_H125_a10a10_4u_ctau10
603016	PhPy8_WpH_H125_a10a10_4u_ctau100
603019	PhPy8_WpH_H125_a10a10_4b_ctau10
603020	PhPy8_WpH_H125_a10a10_4b_ctau100
603023	PhPy8_WpH_H125_a25a25_4u_ctau10
603024	PhPy8_WpH_H125_a25a25_4u_ctau100
603027	PhPy8_WpH_H125_a25a25_4b_ctau10
603028	PhPy8_WpH_H125_a25a25_4b_ctau100
603031	PhPy8_WpH_H125_a7a7_4u_ctau10
603032	PhPy8_WpH_H125_a7a7_4u_ctau100
603035	PhPy8_WpH_H125_a7a7_4c_ctau10
603036	PhPy8_WpH_H125_a7a7_4c_ctau100
603039	PhPy8_WpH_H125_a35a35_4u_ctau10
603040	PhPy8_WpH_H125_a35a35_4u_ctau100
603043	PhPy8_WpH_H125_a35a35_4b_ctau10
603044	PhPy8_WpH_H125_a35a35_4b_ctau100
603047	PhPy8_WmH_H125_a10a10_4u_ctau10
603048	PhPy8_WmH_H125_a10a10_4u_ctau100
603051	PhPy8_WmH_H125_a10a10_4b_ctau10
603052	PhPy8_WmH_H125_a10a10_4b_ctau100
603055	PhPy8_WmH_H125_a25a25_4u_ctau10
603056	PhPy8_WmH_H125_a25a25_4u_ctau100
603059	PhPy8_WmH_H125_a25a25_4b_ctau10
603060	PhPy8_WmH_H125_a25a25_4b_ctau100
603063	PhPy8_WmH_H125_a7a7_4u_ctau10
603064	PhPy8_WmH_H125_a7a7_4u_ctau100
603067	PhPy8_WmH_H125_a7a7_4c_ctau10
603068	PhPy8_WmH_H125_a7a7_4c_ctau100
603071	PhPy8_WmH_H125_a35a35_4u_ctau10

603072 PhPy8_WmH_H125_a35a35_4u_ctau100
603075 PhPy8_WmH_H125_a35a35_4b_ctau10
603076 PhPy8_WmH_H125_a35a35_4b_ctau100
603079 PhPy8_ZllH_H125_a10a10_4u_ctau10
603080 PhPy8_ZllH_H125_a10a10_4u_ctau100
603083 PhPy8_ZllH_H125_a10a10_4b_ctau10
603084 PhPy8_ZllH_H125_a10a10_4b_ctau100
603087 PhPy8_ZllH_H125_a25a25_4u_ctau10
603088 PhPy8_ZllH_H125_a25a25_4u_ctau100
603091 PhPy8_ZllH_H125_a25a25_4b_ctau10
603092 PhPy8_ZllH_H125_a25a25_4b_ctau100
603095 PhPy8_ZllH_H125_a7a7_4u_ctau10
603096 PhPy8_ZllH_H125_a7a7_4u_ctau100
603099 PhPy8_ZllH_H125_a7a7_4c_ctau10
603100 PhPy8_ZllH_H125_a7a7_4c_ctau100
603103 PhPy8_ZllH_H125_a35a35_4u_ctau10
603104 PhPy8_ZllH_H125_a35a35_4u_ctau100
603107 PhPy8_ZllH_H125_a35a35_4b_ctau10
603108 PhPy8_ZllH_H125_a35a35_4b_ctau100
603111 PhPy8_ggZllH_H125_a10a10_4u_ctau10
603112 PhPy8_ggZllH_H125_a10a10_4u_ctau100
603115 PhPy8_ggZllH_H125_a10a10_4b_ctau10
603116 PhPy8_ggZllH_H125_a10a10_4b_ctau100
603119 PhPy8_ggZllH_H125_a25a25_4u_ctau10
603120 PhPy8_ggZllH_H125_a25a25_4u_ctau100
603123 PhPy8_ggZllH_H125_a25a25_4b_ctau10
603124 PhPy8_ggZllH_H125_a25a25_4b_ctau100
603127 PhPy8_ggZllH_H125_a7a7_4u_ctau10
603128 PhPy8_ggZllH_H125_a7a7_4u_ctau100
603131 PhPy8_ggZllH_H125_a7a7_4c_ctau10
603132 PhPy8_ggZllH_H125_a7a7_4c_ctau100
603135 PhPy8_ggZllH_H125_a35a35_4u_ctau10
603136 PhPy8_ggZllH_H125_a35a35_4u_ctau100
603139 PhPy8_ggZllH_H125_a35a35_4b_ctau10
603140 PhPy8_ggZllH_H125_a35a35_4b_ctau100
603143 PhPy8_VBF_H125_a10a10_4u_ctau10
603144 PhPy8_VBF_H125_a10a10_4u_ctau100

603147	PhPy8_VBF_H125_a10a10_4b_ctau10
603148	PhPy8_VBF_H125_a10a10_4b_ctau100
603151	PhPy8_VBF_H125_a25a25_4u_ctau10
603152	PhPy8_VBF_H125_a25a25_4u_ctau100
603155	PhPy8_VBF_H125_a25a25_4b_ctau10
603156	PhPy8_VBF_H125_a25a25_4b_ctau100
603159	PhPy8_VBF_H125_a7a7_4u_ctau10
603160	PhPy8_VBF_H125_a7a7_4u_ctau100
603163	PhPy8_VBF_H125_a7a7_4c_ctau10
603164	PhPy8_VBF_H125_a7a7_4c_ctau100
603167	PhPy8_VBF_H125_a35a35_4u_ctau10
603168	PhPy8_VBF_H125_a35a35_4u_ctau100
603171	PhPy8_VBF_H125_a35a35_4b_ctau10
603172	PhPy8_VBF_H125_a35a35_4b_ctau100

Table A.3: The dataset identifiers (DSIDs) and process names for signal MC samples used in the graph neural network studies, in addition to those noted in Table A.1. The process name details the Higgs production mode, the mass of the Higgs boson, the mass of the two LLPs, the decay mode of the LLPs, and the mean LLP lifetime. Only samples from the MC20e campaign are used.

B

LEPTON TRIGGER EFFICIENCIES

One way of measuring the performance of a trigger is by computing its efficiency. In this appendix, studies detailing the pass and matching efficiencies of lepton triggers for MC signal samples are presented. All efficiencies in this appendix are computed using a Bayesian approach, assuming a uniform beta function prior. The uncertainties are also Bayesian and are computed with a 68.3% confidence level. At their core, efficiency calculations rely on finding the number of events that pass a cut out of some total number of events. These two sets are often referred to as the *numerator* and *denominator*, respectively, within a frequentist context. Therefore, although not being treated as such in the Bayesian approach, this terminology is adopted to clearly specify which sets of events are being used in each efficiency calculation.

As described in §5.2.2, single- and di-lepton triggers are used in the analysis to target the ZH and WH search regions. The lepton triggers used and the year and periods for which they were active are shown in Table B.1.

	Trigger	2015	2016	2017	2018
e	HLT_e24_lhmedium_L1EM20VH	D3 - J			
	HLT_e24_lhtight_nod0_ivarloose		A3 - D3		
	HLT_e26_lhtight_nod0_ivarloose		D4 - L	B - K	B - Q
	HLT_e60_lhmedium	D3 - J			
	HLT_e60_lhmedium_nod0		A3 - L	B - K	B - Q
	HLT_e120_lhloose	D3 - J			
	HLT_e140_lhloose_nod0		A3 - L	B - K	B - Q
μ	HLT_mu20_iloose_L1MU15	D3 - J			
	HLT_mu26_ivarmedium		A3 - L	B - K	B - Q
	HLT_mu50	D3 - J	A3 - L	B - K	B - Q
ee	HLT_2e12_lhvloose_L12EM10VH	D3 - J			
	HLT_2e15_lhvloose_nod0_L12EM13VH		A3 - D3 ^a		
	HLT_2e17_lhvloose_nod0		D4 - L ^b		
	HLT_2e17_lhvloose_nod0_L12EM15VHI			B - K ^c	B - Q
	HLT_2e24_lhvloose_nod0			B - K	B - Q
$\mu\mu$	HLT_2mu10	D3 - J			
	HLT_2mu14		A3 - L	B - K	B - Q
	HLT_mu18_mu8noL1	D3 - J			
	HLT_mu22_mu8noL1		A3 - L	B - K	B - Q

^a Except run 300540 due to prescaling

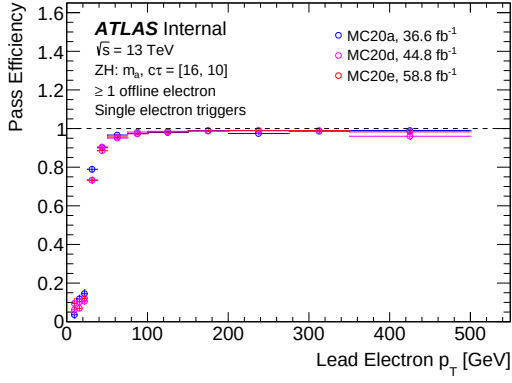
^b Also run 300540

^c Except runs 326834 - 328393 due to prescaling

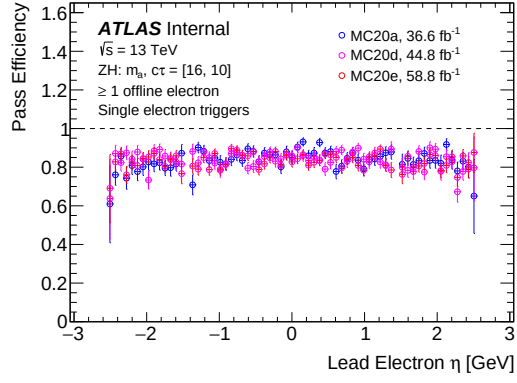
Table B.1: Single lepton (e or μ) and di-lepton (ee or $\mu\mu$) triggers forming the lepton trigger preselection and the years and periods for which they were active.

For an event to pass the trigger pre-selection, it must first pass a trigger that was active during the period in which the event occurred. The pass efficiency for a set of triggers is computed using events that pass at least one of the triggers in the set as its numerator and all events as its denominator. Events are required to have a least one lepton¹, where the lepton flavour is dependent on the type of trigger being considered. Figures B.1a and B.1b (B.1c and B.1d) show the pass efficiency for single-electron triggers as a function of lead electron p_T and η , respectively, for $m_a, c\tau = [16, 10]$ MC ZH (WH) signal samples. Figures B.2a and B.2b (B.2c and B.2d) show the pass efficiency for single-muon triggers as a function of lead muon p_T and η , respectively, for the ZH (WH)

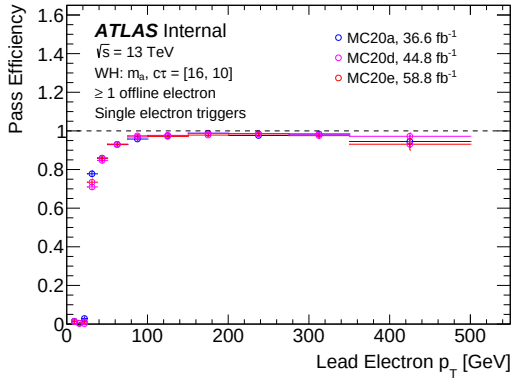
¹The electrons and muons used in these efficiency studies must pass the analysis level criteria listed in Table 5.4 or 5.5, including the looser p_T cut.



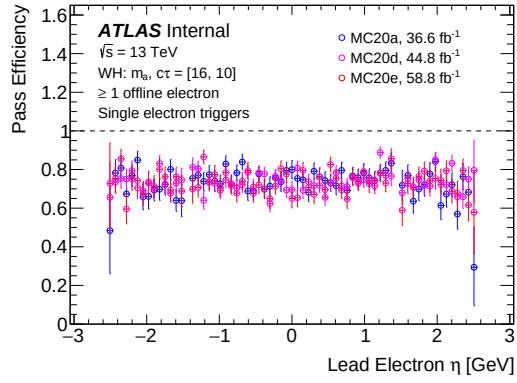
(a)



(b)



(c)

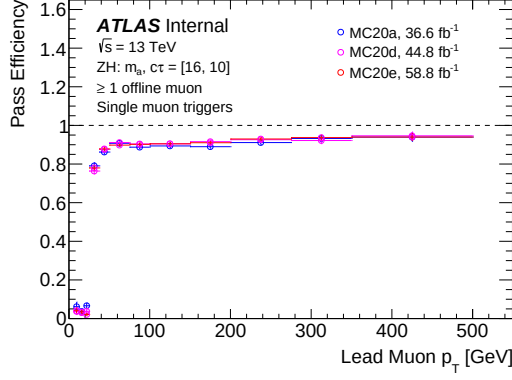


(d)

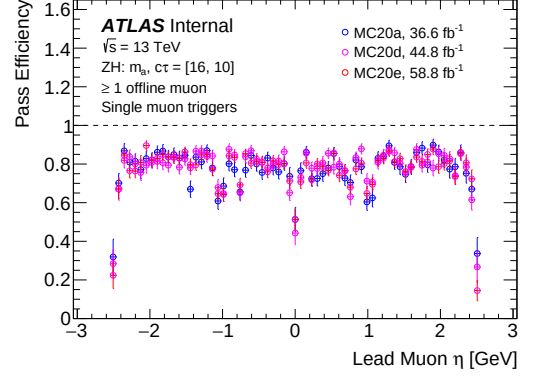
Figure B.1: Top row: Pass efficiency for single-electron triggers as a function of lead electron p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. Bottom row: Pass efficiency for single-electron triggers as a function of lead electron p_T (c) and η (d) for MC20a, MC20d and MC20e WH samples. The single-electron triggers used are listed in Table B.1.

signal samples. High efficiency is observed for each beyond the p_T threshold of the loosest triggers. Figures B.3a and B.3b show the pass efficiency for di-electron triggers as a function of lead electron p_T and η , respectively, for the ZH signal samples. Figures B.4a and B.4b show the pass efficiency for di-muon triggers as a function of lead muon p_T and η , respectively, for the ZH signal samples. The efficiency for each plateaus at around 60 GeV. The WH samples are omitted since signal events from a Higgs produced in association with a W boson are only expected to have a single lepton.

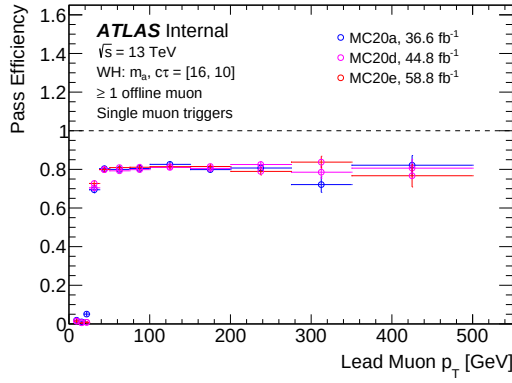
Events that are matched to at least one single- or di-lepton trigger that was active during the period in which the event occurred pass the trigger preselection. The matching efficiency for single-lepton triggers is computed using all events matched to at least one single-lepton trigger as the numerator and all events that passed a single-lepton



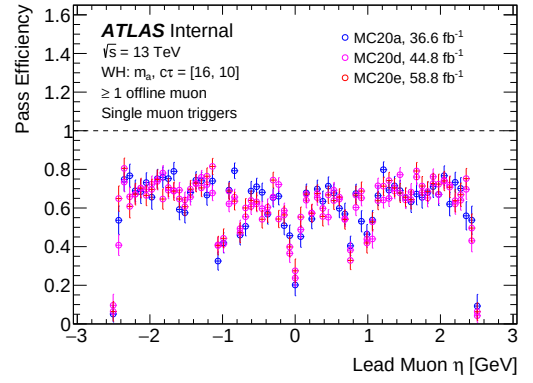
(a)



(b)

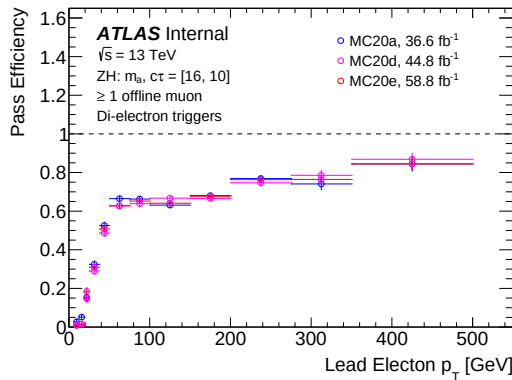


(c)

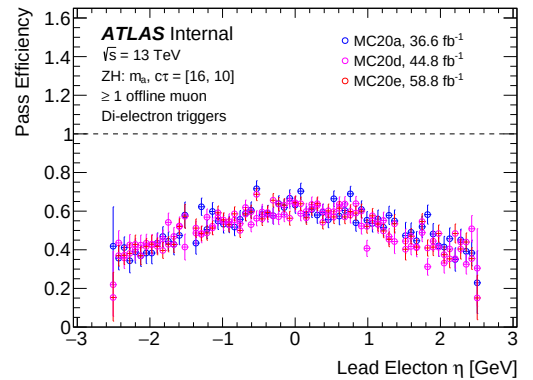


(d)

Figure B.2: Top row: Pass efficiency for single-muon triggers as a function of lead muon p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. Bottom row: Pass efficiency for single-muon triggers as a function of lead muon p_T (c) and η (d) for MC20a, MC20d and MC20e WH samples. The single-muon triggers used are listed in Table B.1.



(a)



(b)

Figure B.3: Pass efficiency for di-electron triggers as a function of lead electron p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. The di-electron triggers used are listed in Table B.1.

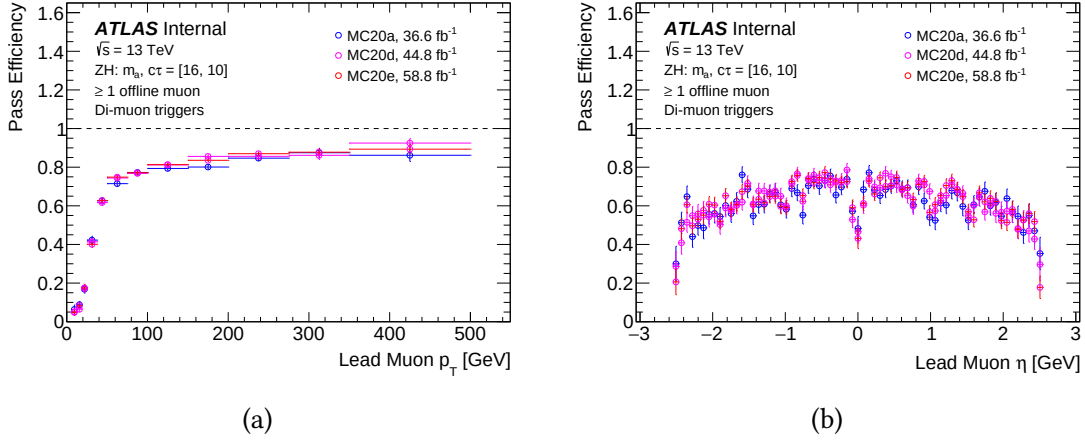


Figure B.4: Pass efficiency for di-muon triggers as a function of lead muon p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. The di-muon triggers used are listed in Table B.1.

trigger with either one offline lepton or two offline leptons as the denominator. Figures B.5a and B.5b (B.5c and B.5d) show the match efficiency for single-electron triggers as a function of lead electron p_T and η , respectively, for the ZH (WH) signal samples. Figures B.6a and B.6b (B.6c and B.6d) show the match efficiency for single-electron triggers as a function of lead electron p_T and η , respectively, for the ZH (WH) signal samples. The matching efficiency for di-lepton triggers is computed using all events matched to at least one di-lepton trigger as the numerator and all events that passed a di-lepton trigger with either two offline electrons or two offline muons as the denominator. Figures B.7a and B.7b show the match efficiency for di-electron triggers as a function of lead electron p_T and η , respectively, for the ZH signal samples. Figures B.8a and B.8b show the match efficiency for di-muon triggers as a function of lead muon p_T and η , respectively, for the ZH signal samples. High matching efficiency is observed for single-lepton and di-electron triggers. Di-muon triggers have a lower matching efficiency. Loss in efficiency at $|\eta| < 0.1$ for muon triggers is due to a gap in the muon system coverage. Both pass and matching efficiencies are comparable between MC campaigns.

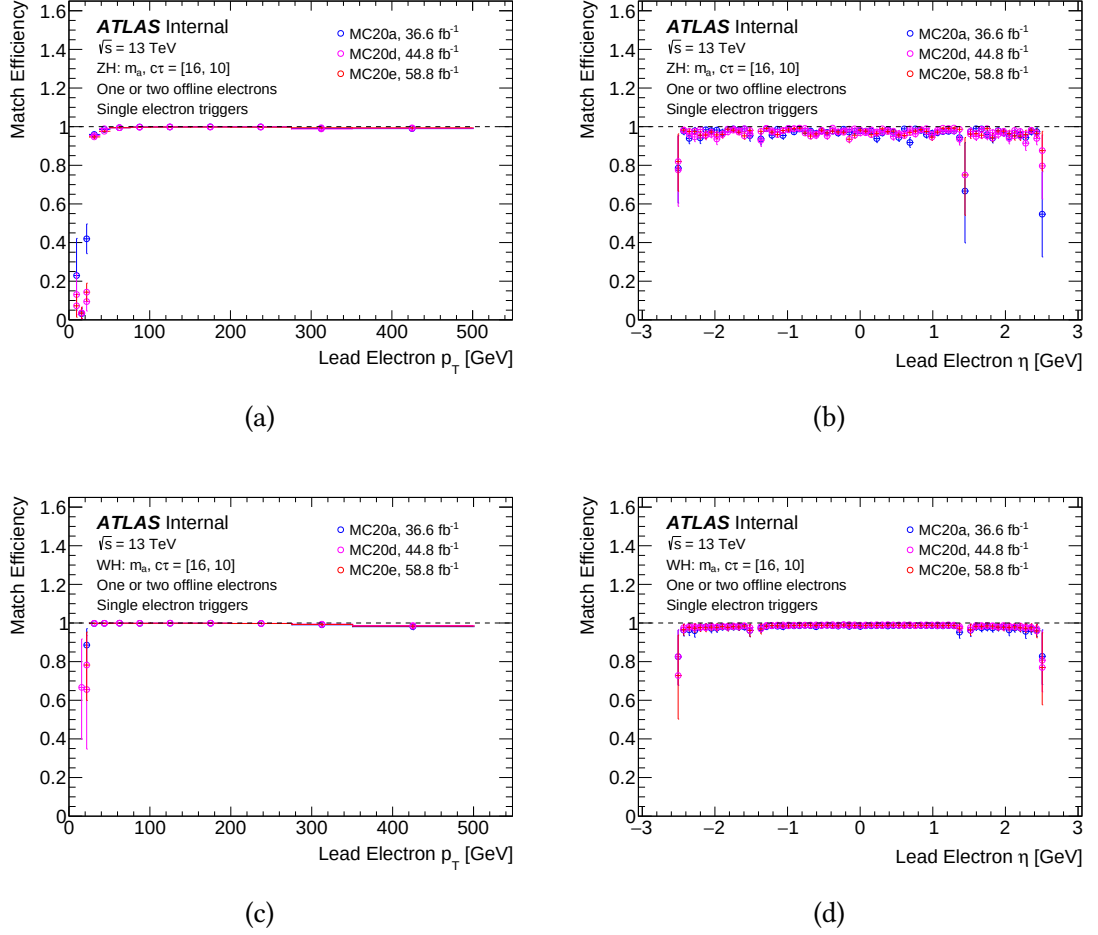
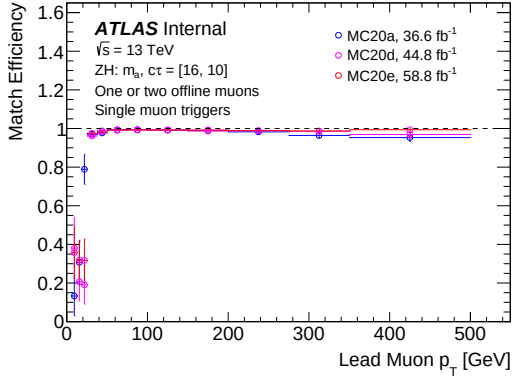
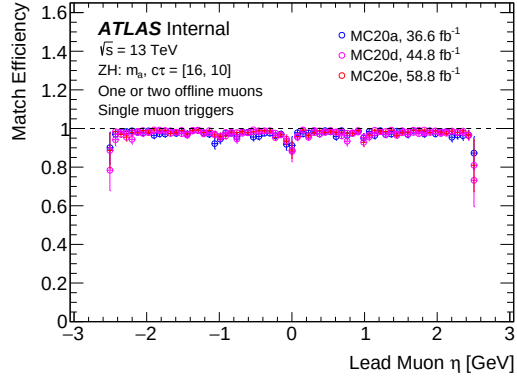


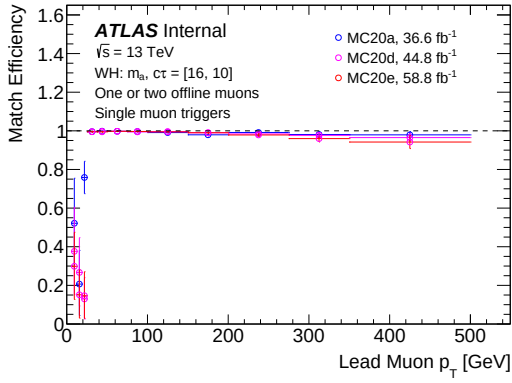
Figure B.5: Top row: Match efficiency for single-electron triggers as a function of lead electron p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. Bottom row: Match efficiency for single-muon electron as a function of lead muon p_T (c) and η (d) for MC20a, MC20d and MC20e WH samples. The single-electron triggers used are listed in Table B.1. Matching is performed for and its efficiency computed with respect to events passing single-electron triggers with one or two offline electrons.



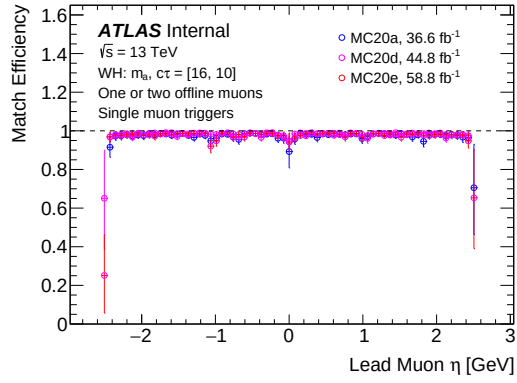
(a)



(b)

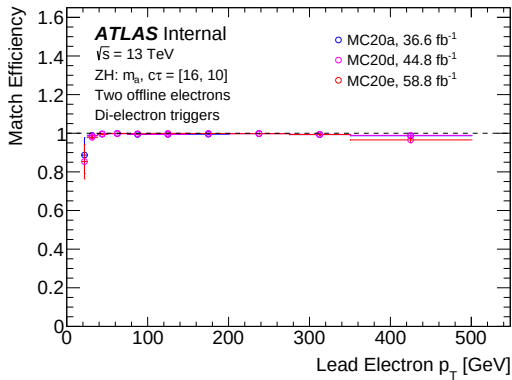


(c)

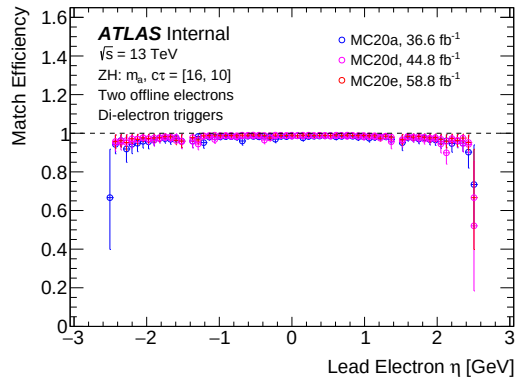


(d)

Figure B.6: Top row: Match efficiency for single-muon triggers as a function of lead muon p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. Bottom row: Match efficiency for single-muon electron as a function of lead muon p_T (c) and η (d) for MC20a, MC20d and MC20e WH samples. The single-muon triggers used are listed in Table B.1. Matching is performed for and its efficiency computed with respect to events passing single-muon triggers with one or two offline muons.



(a)



(b)

Figure B.7: Match efficiency for di-electron triggers as a function of lead electron p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. The di-electron triggers used are listed in Table B.1. Matching is performed for and its efficiency computed with respect to events passing d-electron triggers with two offline electrons.

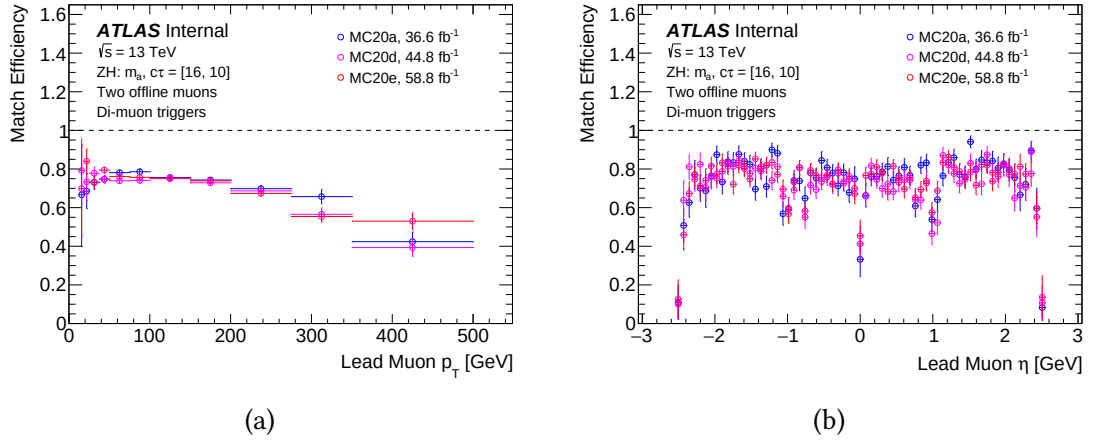


Figure B.8: Match efficiency for di-muon triggers as a function of lead muon p_T (a) and η (b) for MC20a, MC20d and MC20e ZH signal samples. The di-muon triggers used are listed in Table B.1. Matching is performed for and its efficiency computed with respect to eventss passing di-muon triggers with two offline muons.



VBF REGION OPTIMISATION

The VBF inclusive trigger listed in Table 5.2 selects events containing forward jet pairs (j_0, j_1) produced in association with VBF by placing thresholds on the invariant mass m_{jj} , transverse energies (E_T^0, E_T^1) and pseudo-rapidities (η_0, η_1) of the jets, as well as the difference in η and ϕ between the two jets, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$. These are listed in Table C.1.

In the analysis, the trigger is applied to both MC and data to select events for the VBF search region. The signal and background yields in MC are used to compute expected limits on $BR(H \rightarrow aa)$, while the data is used to determine the observed limits. Trigger efficiencies in MC and data may not be identical, due to poor modelling of certain parts of the detector in simulation, for example. Not accounting for this mismatch will result in a systematic difference in yield between MC and data, which would affect the accuracy of the expected limits. Scale factors (SFs) are computed to find the difference in performance of the trigger between MC and data, and are used to inform the forward jet selections detailed in §5.3.3.

Trigger thresholds
$(E_T^0, E_T^1) > (70, 50) \text{ GeV}$
$(\eta_0 , \eta_1) < (3.2, 4.9)$
$m_{jj} > 1100 \text{ GeV}$
$\Delta\eta_{jj} > 4$
$\Delta\phi_{jj} < 2$

Table C.1: Summary of jet pair (j_0, j_1) variable thresholds used in the VBF inclusive trigger.

Background	Sample
$t\bar{t}$	Py8EG_A14_ttbar_hdamp258p75_nonallhad
$V + \text{jets}$	Sh_2211_Zmumu_maxHTpTV2_BFilter
	Sh_2211_Zmumu_maxHTpTV2_CFilterBVeto
	Sh_2211_Zmumu_maxHTpTV2_CVetoBVeto
	Sh_2211_Wmunu_maxHTpTV2_BFilter
	Sh_2211_Wmunu_maxHTpTV2_CFilterBVeto
	Sh_2211_Wmunu_maxHTpTV2_CVetoBVeto

Table C.2: Standard Model background samples used to derive the VBF inclusive trigger scale factors. All samples are from the MC20e MC campaign, which corresponds to the data-taking year 2018 when the VBF trigger was active.

The SF is defined as the ratio between the efficiency to pass the trigger in data ϵ_{data} and the efficiency to pass the trigger in MC ϵ_{MC} . A Tag-and-Probe method is used to compute the SF. Single-muon triggers are used to tag events. The efficiency to pass the VBF inclusive trigger is defined as the ratio of all events that pass single-muon and VBF inclusive triggers to all events passing a single-muon trigger. The efficiency is computed in MC using the SM background samples listed in Table C.2, as well as in data collected during 2018.

Events must first pass cuts of $|\eta_1| < 3.2$, $|\eta_2| < 4.9$, $\Delta\eta_{jj} > 4$ and $\Delta\phi_{jj} < 2$. These mirror those used in the VBF trigger. Cuts are then chosen for m_{jj} , E_T^0 and E_T^1 such that the VBF trigger is fully efficient in both data and simulation and the SF becomes approximately equal to 1, meaning that no scale factor needs to be applied to the MC. This minimises differences in event yield between data and MC.

Nominal selections of $m_{jj} > 1200$ GeV, $E_T^0 > 100$ GeV and $E_T^1 > 80$ GeV were chosen after studying the behaviour of the SF using various event selections. Figures C.1a, C.1b and C.1c show the VBF trigger efficiency and SF (labelled data/MC) as a function of m_{jj} , E_T^0 and E_T^1 , respectively, computed using events with at least one offline muon¹ and containing a forward jet pair satisfying the $|\eta_0|$, $|\eta_1|$, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ cuts listed above as well as the nominal m_{jj} , E_T^0 and E_T^1 cuts, except the one applied to the variable being

¹The muon is required to pass the analysis level criteria given in Table 5.5 (including the looser p_T cut).

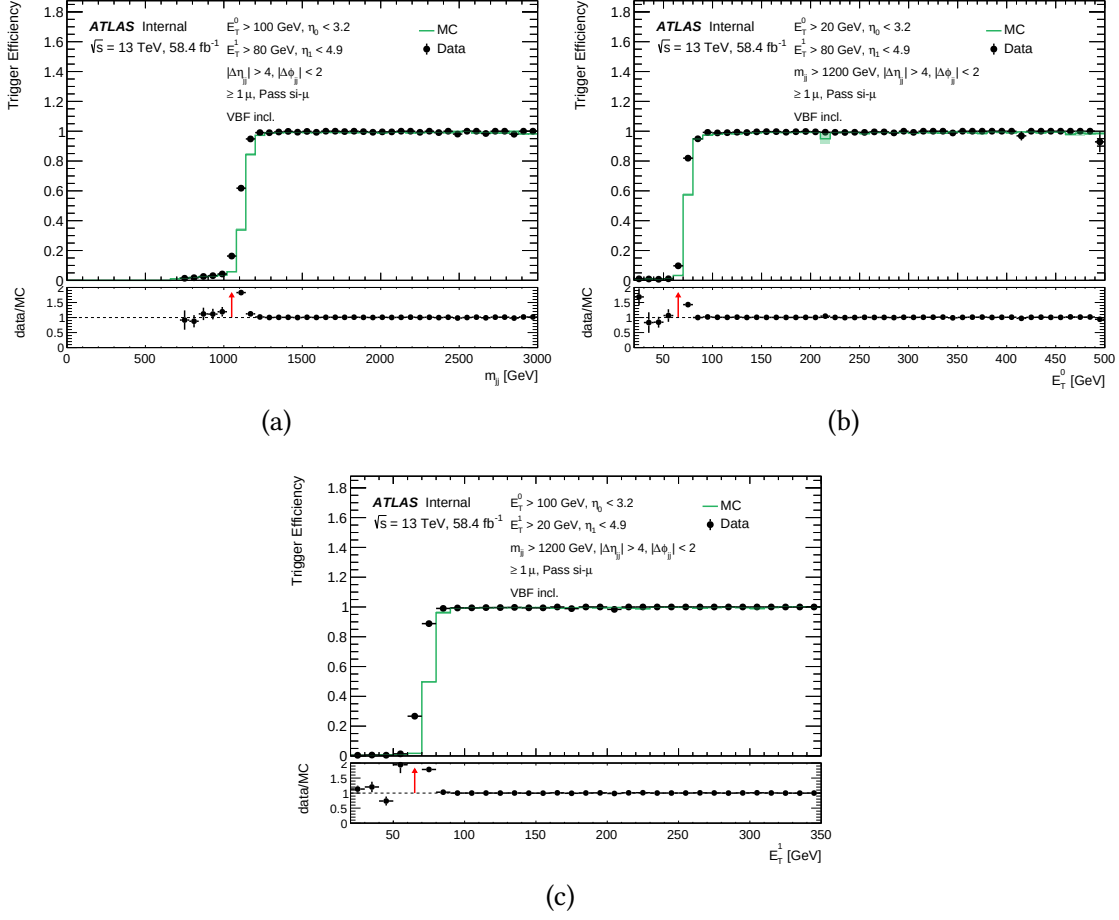


Figure C.1: VBF inclusive trigger efficiency as a function of m_{jj} (a), E_T^0 (b), and E_T^1 (c) for data and MC. Events are required to pass the selection indicated on the Figure. The scale factors, computed by taking the ratio of the data and MC efficiencies, are shown in the ratio plots.

plotted, for which a looser cut is used instead (either $E_T^0 > 20$ GeV, $E_T^1 > 20$ GeV, or no cut for m_{jj}). By observing the behaviour of the SF around the thresholds for each jet pair variable, one can verify that the nominal cuts meet the criteria set out above is met.

Taking Figure C.1a as an example, for values of m_{jj} between 1000 GeV and 1200 GeV an increase of the SF resulting from differences in the efficiency curves between data and MC near the trigger threshold of 1100 GeV is evident. Beyond the nominal cut of 1200 GeV, the trigger nears full efficiency in both data and MC and the SF is approximately equal to 1. Similar conclusions can be made regarding the nominal cuts for E_T^0 and E_T^1 using Figures C.1b and C.1c.