Search for Heavy ZZ Resonances in the 4ℓ Final State with the ATLAS Detector

and

Identifying Muons in the ATLAS Calorimeter using Machine Learning

Ricardo Wölker

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2021

Abstract

This doctoral thesis explores two lines of work: (i) a search for heavy Higgs-like resonances decaying via two Z bosons to four leptons, and (ii) the development of a novel algorithm for tagging muons in the ATLAS calorimeters.

The search concerns signatures of an extended Higgs sector motivated by extensions to the Standard Model of particle physics that address the large discrepancy between the gravitational and the electroweak energy scales known as the hierarchy problem. A search for heavy ZZ resonances decaying to $\ell^+\ell^-\ell'^+\ell'^-$, where ℓ stands for either an electron or a muon, is conducted in the mass range between 200 and 2000 GeV using 139 fb^{-1} of proton-proton collision data collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with the ATLAS detector during 2015–2018 data-taking at the LHC. The search is performed separately in two analysis categories corresponding to the resonance production mechanisms presumed to be dominant: the fusion of two gluons (ggF) and the fusion of two vector bosons (VBF). A novel method for categorising candidate events according to their production mechanism based on deep learning is developed. It is validated in the context of the search in the $\ell^+\ell^-\ell'^+\ell'^-$ final state and is sensitive to high-mass signals with up to 40% higher discovery significance compared to a previously employed, simpler categorisation scheme. No significant excess in the four-lepton invariant mass distribution is found, and upper limits on the production cross-section times branching fraction of a heavy scalar are set. Using a combination of results with the $\ell^+\ell^-\nu\bar{\nu}$ analysis channel, where ν stands for a neutrino, the observed limits on the production cross-section times branching fraction are between 215 fb at 240 GeV and 2.0 fb at 1900 GeV in the ggF production mode, and between 87 fb at 255 GeV and 1.5 fb at 1800 GeV in the VBF production mode. Upper limits on the production cross-section times branching fraction of a spin-2 Randall-Sundrum graviton are set. Randall-Sundrum gravitons are excluded up to a mass of 1830 GeV at 95% confidence level.

The efficient identification of muons is central to the search in the $\ell^+\ell^-\ell'^+\ell'^-$ decay channel and many other physics analyses. The ATLAS detector has a dedicated instrument for muon identification: the muon spectrometer. There are, however, regions of the detector where the muon spectrometer is only partially instrumented, and therefore cannot be used to identify muons. A novel approach to tagging muons in the ATLAS calorimeter called `CaloMuonScore` is developed based on convolutional neural networks trained on calorimeter energy deposits. The method is validated on simulation and its performance is assessed on data in a tag-and-probe analysis. Compared to the previous approach to calorimeter-based muon identification, `CaloMuonScore` identifies muons with up to 9% higher efficiency while rejecting background coming predominantly from pi mesons up to 20 times more effectively. The method is now commissioned within the central ATLAS software repository for use in future physics analyses.

Acknowledgements

The completion of this thesis would not have been possible without the support of a number of people who have generously offered their guidance, inspiration, expertise, and time, and who have been exceptional collaborators.

First, I would like to thank my supervisor Ian Shipsey for his constant encouragement and support throughout the last four years. Without him this research would not have been possible. His mentorship, infectious enthusiasm, trust, and valuable feedback during this time have been immensely appreciated, in addition to his thoughtful support of my development above and beyond physics.

I am very grateful to Giacomo Artoni for his guidance and time throughout my doctorate. He has greatly shaped my understanding of particle physics and the ATLAS experiment, from the broad and conceptual to the hands-on technicalities of data analysis and ATLAS software. Giacomo has been a consistent source of support and provided ongoing, valuable input throughout this research.

Throughout my time at Oxford and at CERN, I have had the fortune to learn from and work with a number of exceptional physicists. I wish to thank the group of graduate students, post-docs, and professors in the Bortoletto–Shipsey group. Thanks to Luca Ambroz, Giacomo Artoni, Daniela Bortoletto, Maxence Draguet, Maria Giovanna Foti, Maria Mironova, Karolos Potamianos, Elisabeth Schopf, Ian Shipsey, Cecilia Tosciri, Yingjie Wei, Philipp Windischhofer, Siyuan Yan, and Miha Zgubič for many stimulating discussions and invaluable feedback.

I would like to thank Lailin Xu for introducing me to the high-mass $H \to 4\ell$ analysis, and for his guidance and help throughout my time in it. Thanks to Joey Carter and Heling Zhu for collaboration, discussions, and help with many aspects of the high-mass analysis. I wish to thank Nicolas Köhler and Johannes Junggeburth, who have introduced me to efficiency measurements and helped me navigate the ATLAS muon software system. I would like to thank Craig Sawyer for his help and support early on during my qualification task at RAL. I am grateful to many

other ATLAS collaborators, of which there are too many to be listed here. Among them, I would like to thank RD Schaffer, Gaetano Barone, and Andrea Gabrielli for discussions and guidance on the high-mass analysis. I also thank Kyle Cranmer, Aishik Ghosh, and David Rousseau for discussions and advice on the high-mass machine-learning classifier, and Magnar Kopangen Bugge, Davide Cieri, and Marco Vanadia for their help and advice on MCP-related work.

My deep gratitude goes to my family and friends for their support throughout the last four years. To Johannah, whose unwavering support, encouragement, and patience have been essential throughout this journey, go my final thanks and love.

Preface

Research efforts in experimental particle physics are typically conducted in large international teams with many collaborators, and the results reported in this thesis inevitably draw on the work of many individuals. Chapters 2 to 4 provide general introductions to the theory of particle physics, high-energy physics experiments, as well as machine learning, while Chapters 5 and 6 report on my original contributions.

My contributions can be broadly categorised into three areas: a physics analysis, performance work related to muons, as well as a hardware-based study for the ATLAS inner tracker upgrade. Specific contributions are highlighted below.

Physics analysis: Search for heavy $H \to ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ resonances

I was a core member of an analysis team searching for heavy resonances beyond the Standard Model decaying via two Z bosons to the $\ell^+\ell^-\ell'^+\ell'^-$ final state. The search is the subject of Chapter 5. My specific contributions were:

- The development of an event classifier based on deep learning designed to discriminate VBF- and ggF-produced signal events from background arising from Standard Model processes. I designed, optimised, and trained the machine-learning algorithm, integrated it into the analysis workflow, and validated the approach on simulated Higgs signal and Standard Model Monte Carlo samples.

- Validation of the previous event-classification cuts. I ran a study to investigate the optimal working point for the legacy cut-based approach to event categorisation, in which simple cuts were placed on two kinematic event variables. This contribution is not part of this thesis.

- Comparison between data and simulation. I validated the simulation samples used by the analysis group by comparing kinematic variables among different

Monte Carlo simulation campaigns, as well as between simulation and data in a control region where no signal was expected (at masses below 200 GeV). This contribution is not part of this thesis.

- Preliminary signal-modelling studies of radion invariant-mass distributions in preparation for a planned ATLAS paper combining several resonance searches. This contribution is not part of this thesis.

During my time in the $H \to ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ analysis team, the group wrote an internal note of which I was a co-author (Reference [1]), published one conference paper (Reference [2]), and one journal paper (Reference [3]).

Physics performance: Tagging muons in the ATLAS calorimeter

The muon software/performance work presented in Chapter 6 is entirely my own. I developed a tagging algorithm for the ATLAS calorimeter that extrapolates particle tracks from the inner detector into the muon spectrometer, forms three-dimensional representations of energy deposits in the calorimeter, and runs a convolutional neural network on these energy deposit patterns in order to assign a likelihood score corresponding to whether the particle is a muon or not. I developed the algorithm, validated it on data, and commissioned it within the ATLAS software framework, making it available for all physics analyses. The algorithm known as `CaloMuonScore` is now a core component of the main ATLAS software project `Athena Release 22` (released in November 2020), and will be available for use in legacy Run-2 analyses and all analyses from Run 3 onwards.

Hardware: High-statistics irradiations of the ABC130 chip for the ATLAS ITk

I led irradiation damage studies of the read-out chips to be used in the Phase-II upgrade of the ATLAS Inner Tracker (ITk) for the High-Luminosity era of the LHC beginning in 2024. The chips called ABC130 exhibit leakage-current effects that are correlated with the total ionising dose received. These vary greatly between wafer production batches delivered by the silicon foundry, among individual wafers in a batch, and among different chip locations within one wafer. Specifically, I conducted X-ray irradiations of a large number of ABC130 chips in order to quantify the batch-by-batch, wafer-by-wafer, and chip-by-chip variations of the total ionising dose effects. The work was performed at the Rutherford Appleton Laboratory in the UK as part of my ATLAS authorship qualification task from

October 2017 until June 2018. This thesis does not report on the methods and results of these high-statistics irradiations. I presented public results of this work at the *2018 Topical Workshop on Electronics for Particle Physics*, and published the accompanying conference proceedings in Reference [4].

Contents

List of Figures

List of Tables

Chapter 1

Introduction

High-energy scattering experiments are one of the most fundamental ways of probing nature at small scales. Historically, through many generations of scattering experiments combined with decades of theoretical progress, a theory of elementary particles and their interactions through force mediators has been devised, known as the Standard Model (SM) of particle physics. Among the many machines dedicated to the study of the SM through the scattering of elementary particles, the Large Hadron Collider (LHC) at CERN near Geneva, Switzerland, is the one that has achievedc the highest energies and collision rates to date. Designed to measure the properties of the SM and to look for new phenomena, two experiments located at the LHC achieved a significant milestone by observing the Higgs boson, the last missing particle predicted by the SM almost half a century earlier.

The SM predicts all phenomena observed in scattering experiments with exceptional accuracy and precision, yet it presents several theoretical shortcomings and does not account for many phenomena observed outside scattering laboratories. Among these, two of the most notable ones are the question of why the energy scales between the weak force and gravity differ so greatly—the weak force is some 24 orders of magnitude stronger than the gravitational force—and the lack of an explanation for a form of matter known as dark matter, which is at least five times more abundant in our universe than baryonic matter.

After the discovery of the SM Higgs boson in 2012 by the CMS and ATLAS collaborations at the LHC [5, 6], the physics programme of the LHC experiments has focused on (i) measuring the parameters of the SM with greater precision in order to find discrepancies between theoretical prediction and measurement that could be explained by theories beyond the Standard Model (BSM), and (ii) searching directly for signatures of phenomena not explained by the SM such as new particles

or force carriers. Essential to both of these avenues are reliable experiments that can observe and study the collisions of elementary particles at high energies and identify the decay products of the scattering processes reliably.

This thesis addresses both lines of work and is structured as follows: A theoretical introduction to the SM of particle physics is given in Chapter 2, with emphasis on the Higgs mechanism and phenomena beyond the SM that are the subject of this work. Chapter 3 gives an overview of the experimental apparatus used throughout this thesis: the ATLAS detector at the LHC. Both original contributions presented in this thesis use several techniques from the field of machine learning, and Chapter 4 provides an introduction to the mathematical basis underlying the specific methods used in this work.

Chapter 5 presents a search for heavy BSM resonances decaying via two Z bosons to $\ell^+\ell^-\ell'^+\ell'^-$ using 139 fb^{-1} of proton-proton collision data collected with the ATLAS detector during Run 2 of the Large Hadron Collider in 2015–2018. The main target of the search is a heavy spin-0 resonance similar to the SM Higgs boson, and it is assumed that the main production mechanisms of the resonance are the fusion of two gluons as well as the fusion of two W or Z bosons. Unlike in the case of the SM Higgs boson, however, the search does not make assumptions as to a specific BSM scenario that could give rise to the spin-0 resonance, and therefore the ratio of the two production mechanisms is presumed to be unknown. As a result, the search is conducted separately in the two production categories presumed to be dominant. The chapter gives particular focus to the development, optimisation, and performance of a novel event-categorisation algorithm based on deep learning designed to classify potential signal events according to their production mechanism.

The focus of Chapter 6 is muons and their identification in the ATLAS calorimeter. While ATLAS is equipped with a dedicated muon spectrometer, this system has lower acceptance in the central region of the ATLAS detector. Here, muons are frequently identified on the basis of particle tracks in the innermost layer of the ATLAS detector in combination with energy deposits in the surrounding electromagnetic and hadronic calorimeters. Historically, ATLAS has approached muon identification in the calorimeter by examining the energy deposits left by candidate tracks in the following way: If the sum of a track's energy deposits lies above a noise threshold and is low enough to be consistent with a minimum ionising particle, it is likely to be a muon. In this chapter, a new method for identifying muons in the ATLAS electromagnetic and hadronic calorimeters called `CaloMuonScore` is

developed based on convolutional neural networks trained on calorimeter images. The chapter focuses on the development of this new method, the determination of its performance compared to the previous method, and a validation of the approach performed on simulated events as well as recorded collision data.

Finally, Chapter 7 presents concluding remarks.

Chapter 2

Theoretical Motivation

This chapter introduces the Standard Model of particle physics, describing the interaction of matter through three of the four known fundamental forces. The limitations of the SM are highlighted, and a brief overview of efforts to search for physics beyond the SM is given. While the spectrum of searches for BSM physics is vast, emphasis will be given to searches for signatures of an extended Higgs sector.

2.1 The Standard Model of particle physics

The SM of particle physics [7, 8, 9] is a quantum field theory (QFT) that classifies all known elementary particles as well as the interactions between them through three fundamental forces of nature. The three forces accounted for in SM are the electromagnetic force, the weak nuclear force, and the strong nuclear force. The SM does not provide a description of gravity.

The SM of particle physics is an extension of quantum electrodynamics (QED). While QED postulates gauge invariance for the EM fields, the SM extends the concept of gauge invariance to the remaining fields. In the SM, the universe is described by a set of quantum fields, the excitations of which manifest themselves as physical particles.

The dynamics of the SM are fully characterised by a quantity called the *Lagrangian density*, $\mathscr{L}$. The theory of the SM should be independent of the reference frame in which it is considered and consequently $\mathscr{L}$ must be a Lorentz-invariant scalar.

In the Lagrangian formalism of classical field theory, the dynamics of a system are fully described by the action, which is given by

$$S = \int d^4x \, \mathscr{L}[\phi(x), \partial_\mu \phi(x)], \tag{2.1}$$

where $\mathscr{L}[\phi(x), \partial_\mu \phi(x)]$ indicates that the Lagrangian density is a functional of a field and its covariant derivatives. Assuming the principle of stationary action, which states that the path taken by a system is such that its action is stationary, and varying the field according to $\phi \rightarrow \phi + \delta\phi$, it can be shown that $\mathscr{L}$ must obey the following *Euler-Lagrange equation*, which serves as the equation of motion for the field $\phi(x)$:

$$\frac{\partial \mathscr{L}}{\partial \phi(x)} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi(x))} = 0. \tag{2.2}$$

2.1.1 Scattering theory

Transition rates in quantum mechanics are computed using Fermi's golden rule [10]. It relates the transition rates between two states, Γ, to the matrix element, $\mathcal{M}$, of the process, i.e. [11]

$$\Gamma \sim |\mathcal{M}|^2. \tag{2.3}$$

States in QFT are commonly expressed in the Heisenberg picture, in which the evolution of states is contained in the operator. In the case of collider experiments, the momentum eigenstates evolve over the time interval $t \in [-\infty, +\infty]$, and the time-evolution operator is called the S-matrix. It contains full information about how the initial state, $|\Phi_i\rangle$, and the final state, $|\Phi_f\rangle$, evolve in time. Therefore in the S-matrix approach, states evolve according to

$$\langle \Phi_f \mid S \mid \Phi_i \rangle. \tag{2.4}$$

The S-matrix assumes that the states in the distant past at $t = -\infty$ and in the distant future at $t = \infty$ are *free states*, that is, they do not experience any interactions [11]. While the S-matrix contains all phenomena of interest in high-energy physics, it cannot be probed directly. Instead, high-energy experiments prepare known initial states $|\Phi_i\rangle$, and measure final states $|\Phi_f\rangle$, and infer from knowledge of these two states properties about the S-matrix. The matrix element, $\mathcal{M}$, of Equation (2.3) is closely related to the S-matrix, in that it comprises the part of the S-matrix that is physically realisable where initial-state and final-state momenta are equal, and it is therefore sufficient to measure $\mathcal{M}$ experimentally.

The following section shows how known initial states and measurable final states relate to the S-matrix via quantities that can be measured in an experiment: cross-sections and decay rates.

2.1.2 Relating the scattering matrix to observables

The differential cross-section of a process in a collision experiment, measured in units of area, is given by [11]

$$d\sigma = \frac{1}{\tau \times F} dP, \tag{2.5}$$

where dP is the differential scattering probability, τ is the time over which the experiment is conducted, and F is the flux of the incident beam, i.e. the particle number density of the beam times the velocity of the beam. $d\sigma$ and dP are differential in the final-state properties measurable by experiment, such as the scattering angles and energies of the final-state particles. The differential cross-section is related to the number of scattered events measured in a collision experiment per unit time, $d\dot{N}$, according to [11]

$$d\dot{N} = \mathcal{L} d\sigma, \tag{2.6}$$

where $\mathcal{L}$ is the *instantaneous luminosity* of the collider.

In all collision experiments that can be practicably measured, one is constrained to two initial states with four-momenta $p_i = (E_i, \gamma_i m_i \boldsymbol{v_i})$ with $i = 1, 2$, and an arbitrary number of final-state particles with index $j = 1, \ldots, n$. Here, E_i are the initial states' energies, γ_i are their Lorentz factors, m_i are their masses, and $\boldsymbol{v}_i$ are their velocities. Instead of the full S-matrix, it is sufficient to consider only the matrix element $\mathcal{M}$, and what one measures experimentally is the matrix element squared, i.e. $|\mathcal{M}|^2 = |\langle \Phi_f | \mathcal{M} | \Phi_i \rangle|^2$. The differential cross-section is related to the matrix element squared via

$$d\sigma = \frac{1}{4E_1 E_2 |\boldsymbol{v}_1 - \boldsymbol{v}_2|} |\mathcal{M}|^2 d\Pi, \tag{2.7}$$

where $d\Pi$ is the phase space over final-state particles $j = 1, \ldots, n$.

2.1.3 Gauge invariance

Inherent to the internal structure of the SM are *symmetries*. Noether's theorem [12] states that if a Lagrangian has a continuous symmetry, then there exists a conserved current associated with that symmetry under the equations of motion [11]. The SM is a non-Abelian[1] gauge field theory, i.e. a *gauge-invariant* or *gauge-symmetric* field theory in which the transformation of a field to another, called a *gauge transformation*, does not result in a change in any measurable quantity.

The SM is based on the gauge group $\mathrm{SU(3)_{QCD}} \times \mathrm{SU(2)_{weak}} \times \mathrm{U(1)_Y}$, which defines all fundamental interactions of leptons and quarks. The group $\mathrm{SU(2)_{weak}} \times \mathrm{U(1)_Y}$ describes the electroweak (EW) theory, which, at a scale of $v = 246$ GeV, the vacuum expectation value of the Higgs field, breaks down spontaneously to $\mathrm{U(1)_{EM}}$ (Section 2.1.5) [11]. The electroweak theory is also known as Glashow-Salam-Weinberg (GSW) theory, named after its discoverers [7, 8, 13]. The gauge group $\mathrm{SU(3)_{QCD}}$ describes *quantum chromodynamics* (QCD), which is the theory of strong interactions. The group has eight generators, corresponding to the eight colour charges of QCD.

2.1.4 Particle content of the SM

The SM describes two types of particles: those with integer spin, called *bosons*, which obey Bose-Einstein statistics, and those with half-integer spin, called *fermions*,

[1] *Non-Abelian* means that the generators of a gauge group do not commute with each other, i.e. among the group there exists at least one pair of elements a, b such that $a * b \neq b * a$.

obeying Fermi-Dirac statistics. All known matter particles, i.e. quarks and leptons, are fermions, whereas all force carriers are bosons. There exist three copies of each fermion, different from each other only in mass. By ordering the three copies of quarks and charged leptons, along with each lepton's corresponding neutrino, according to their mass one obtains the three *generations* of fermions. There are four spin-1 bosons, known as *gauge bosons*, which are the carriers of the fundamental forces: one of the electromagnetic, two of the weak, and one of the strong force. The fifth boson is a scalar particle known as the *Higgs boson*, and is responsible for a mechanism through which the remaining gauge bosons (except for the photon) and the charged fermions acquire mass (see Section 2.1.5).

Leptons carry electric charge as well as weak hypercharge, and therefore interact through the electromagnetic and weak forces. Neutrinos, on the other hand, do not carry electric charge, and as a result only couple to the weak force through their weak hypercharge. Similarly, quarks interact through the electromagnetic and weak forces through their electric charge and weak hypercharge, and, in addition, interact strongly due to their colour charge. The so-called *up-type* quarks (u, c, t) have electric charges of $2/3$, while the *down-type* quarks (d, s, b) have charges of $-1/3$. The gluon is the mediator of the strong force, and it is massless. As the mediator of electromagnetic interactions, the photon is also massless, while the weak force carriers, W^+, W^-, and Z are massive, as is the Higgs boson. The particles of the SM are summarised in Figure 2.1.

The gravitational force is not part of the SM of particle physics. However, a hypothetical massless spin-2 particle called *graviton* has been postulated to be the mediator of the gravitational force, despite theoretical challenges of finding a consistent way of describing a graviton within QFT.

The total number of free parameters in the SM is 25: six lepton masses, six quark masses, three angles and one phase of the Cabibbo–Kobayashi–Maskawa (CKM) matrix [15] determining the degree of CP violation[2] in the quark sector, three angles and a complex phase in the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [16] determining the degree of neutrino mixing and CP violation in the neutrino sector, three gauge couplings, the Higgs vacuum expectation value, as well as the Higgs mass. In principle, there is a 26^{th} parameter that specifies the degree of CP violation in strong interactions, which has so far been found to be

[2] CP violation is the violation of CP invariance. CP describes the simultaneous transformation of one field through charge conjugation and parity transformation. A field is said to be invariant under CP if a CP transformation leaves all observables unchanged.

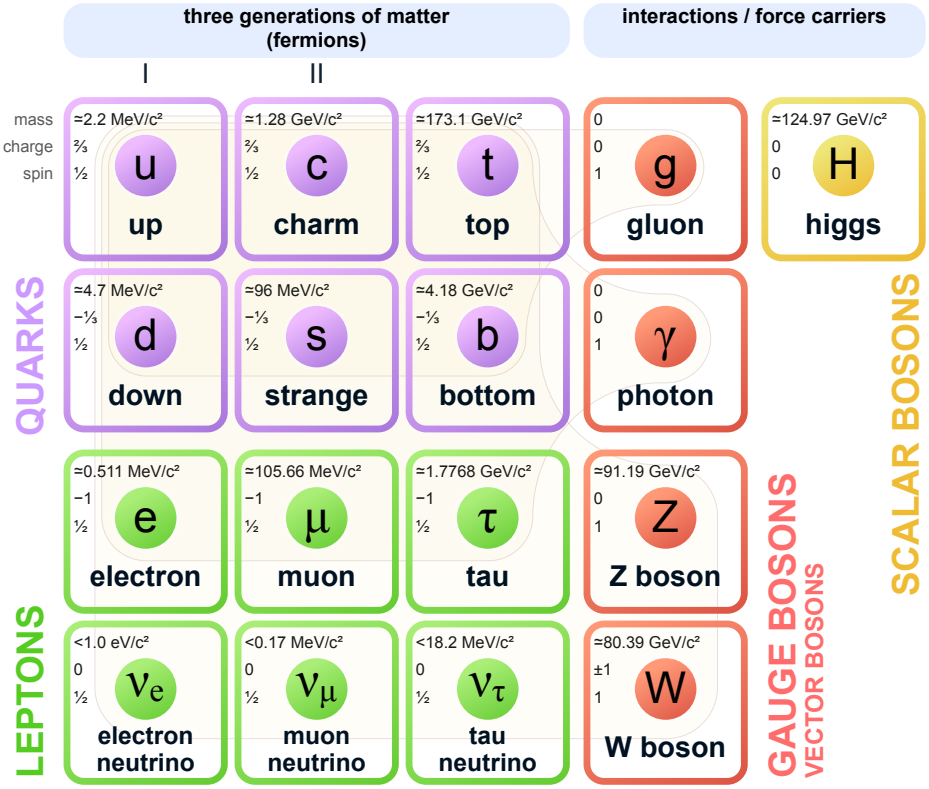


Figure 2.1: The Standard Model of particle physics. Shown are the twelve fermions (quarks and leptons), the four gauge bosons (red), as well as the Higgs boson (yellow). The fermions are ordered column-wise into the three generations of matter. Indicated for a each particle are its mass, its electric charge in units of the elementary charge, and its spin in units of $\hbar$. Figure adapted from Reference [14].

zero experimentally [17]. Determination of these 25 (26) parameters characterises the SM completely. The predictive power of the SM is illustrated in Figure 2.2, which shows the theoretically predicted and experimentally determined production cross-sections of different processes as measured by the ATLAS experiment at the LHC.

2.1.5 The Brout-Englert-Higgs mechanism

In order to generate massive gauge bosons, a mechanism is needed to break down the local gauge symmetry $SU(2)_{weak} \times U(1)_Y$. Known as the *Brout-Englert-Higgs (BEH) mechanism*, it was discovered in 1964, independently by Brout and Englert [19], Higgs [20], as well as Kibble et al. [21]. Unless stated otherwise, the following derivation follows those given in References [11] and [17].

The symmetry breakdown can be generated in the following way: Consider a complex doublet under SU(2), namely

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.8}$$

which is described by the Lagrangian

$$\mathscr{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{2.9}$$

where

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.10}$$

Here, μ is the mass of the complex scalar field, and λ describes the self-interactions of the two complex doublets. D_μ describes the covariant derivative that ensures invariance under a local $SU(2)_{weak} \times U(1)_Y$ gauge transformation, i.e.

$$D_\mu = \partial_\mu + \frac{i}{2} g_W \vec{\sigma} \cdot \vec{W}_\mu + i g' Y B_\mu, \tag{2.11}$$

where g_W is the weak coupling constant, and g' is the coupling strength of the gauge field B_μ to the weak hypercharge, Y. $\vec{W}_\mu$ and B_μ are the $SU(2)_{weak}$ and $U(1)_Y$ gauge bosons. σ are the generators of the $SU(2)_{weak}$ gauge symmetry group, i.e. the Pauli matrices. For the potential V to have a finite minimum, it is required that $\lambda > 0$. Finding a minimum of the Higgs potential (Equation (2.10)) by setting $\frac{\partial V}{\partial(\phi^\dagger \phi)} = 0$ yields two cases depending on the value of μ^2. If $\mu^2 > 0$, the potential has a single

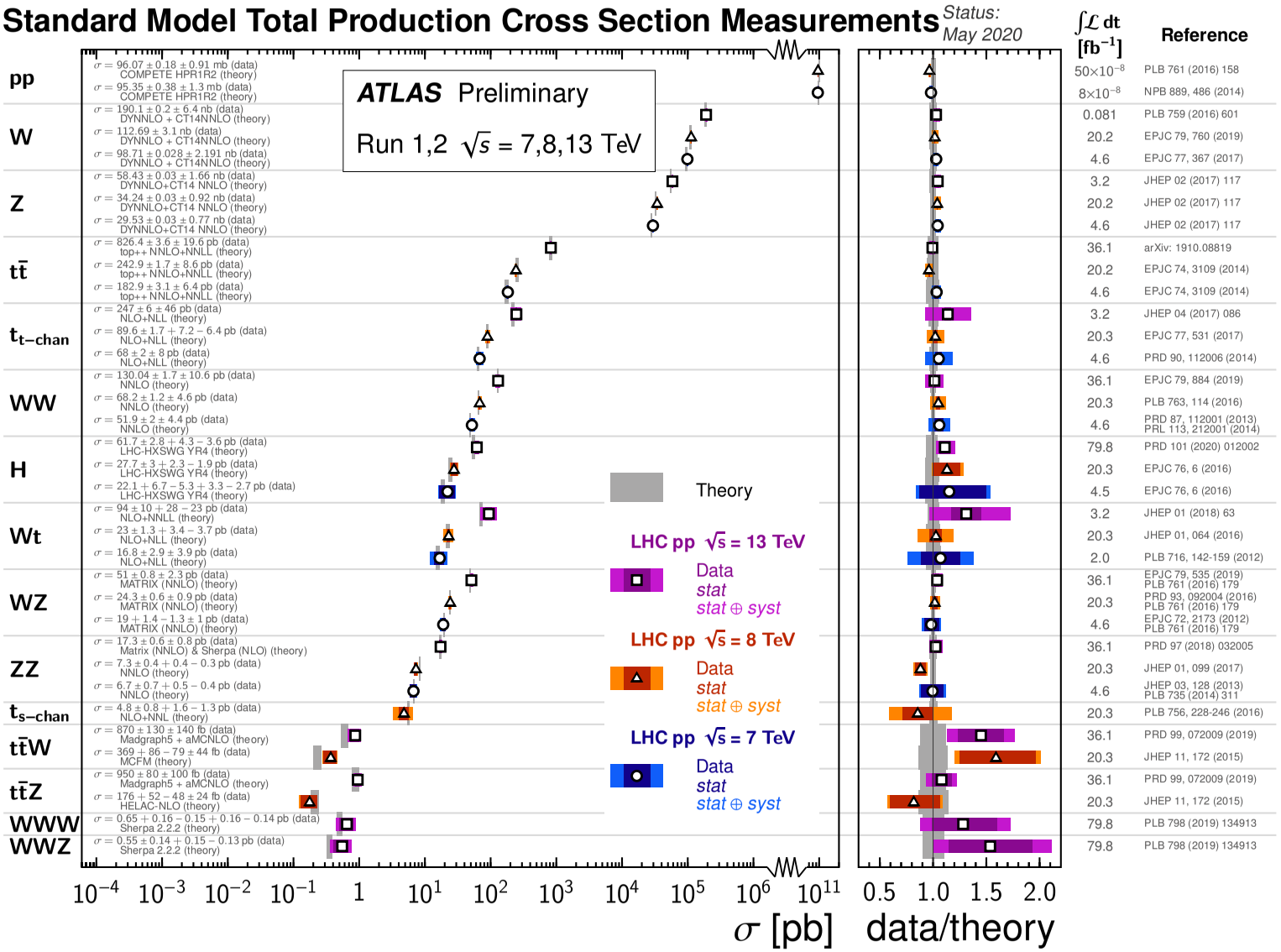


Figure 2.2: The Standard Model total production cross-sections measured by the ATLAS experiment compared to their theoretical expectations. The produced states are indicated on the y-axis. The panel on the right indicates the fractional deviation between measurement and theory, labelled 'data/theory' on the x-axis. The column labelled '$\int \mathcal{L}dt$' indicates the total integrated luminosity, in units of fb^{-1}, each measurement is based on, whereas the rightmost column contains references to the journal publication in which each measurement was reported. Figure reproduced from Reference [18].

minimum at $\phi = 0$. In this case, $V(\phi)$ describes a complex doublet with mass μ and self-interactions proportional to λ. If, on the other hand, $\mu^2 < 0$, V has infinitely many non-zero minima, namely

$$\phi^\dagger\phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}, \tag{2.12}$$

where

$$v = \pm\sqrt{\mu^2/\lambda} \tag{2.13}$$

denotes the Higgs field's *vacuum expectation value*. The shape of the Higgs potential, including its degenerate minima, is illustrated in Figure 2.3. The choice of a vacuum state away from the unstable minimum at $\phi^\dagger\phi = 0$ is known as *electroweak symmetry breaking* (EWSB).

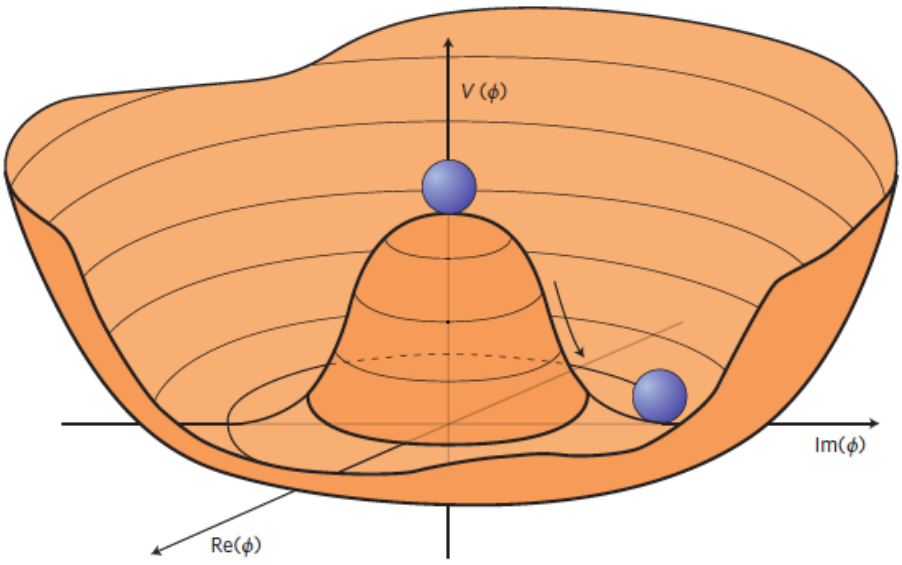


Figure 2.3: The Higgs potential $V(\phi)$ for $\mu^2 < 0$. The degenerate minima are located at $|\phi|^2 = -\mu^2/2\lambda$. The vertical axis shows the Higgs potential, $V(\phi)$, as a function of the real (Re(ϕ)) and imaginary (Im(ϕ)) components of ϕ. Figure reproduced from Reference [22].

Among the infinitely many field configurations of the complex doublet, we choose $\phi^+ = 0$ and $\phi^0 = v$ such that the field configuration that minimises V is given by

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}. \tag{2.14}$$

Now, consider small perturbations around the minimum value of the Higgs potential V, i.e. around its vacuum expectation value $\sim (0, v)$. The two components of the doublet are perturbed according to

$$\phi^+ = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{2.15}$$

and

$$\phi^0 = \frac{1}{\sqrt{2}}(v + \eta + i\phi_4). \tag{2.16}$$

By making an appropriate gauge transformation known as the *unitary gauge transformation*, the Higgs doublet can be rewritten as

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{2.17}$$

where $h(x) \equiv \eta(x)$ is identified as the *Higgs field*. Note that $v + h(x)$ is entirely real.

Returning to $\mathscr{L}_{\text{Higgs}}$ of Equation (2.9) and applying the covariant derivatives, D^μ, to Equation (2.17), it becomes evident how the gauge-boson masses are acquired. We have

$$(D^\mu\phi)^\dagger(D_\mu\phi) = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g_W^2\left(W_\mu^{(1)} + iW_\mu^{(2)}\right)\left(W^{(1)\mu} - iW^{(2)\mu}\right)(v+h)^2 + \frac{1}{8}\left(g_W W_\mu^{(3)} - ig'B_\mu\right)\left(g_W W^{(3)\mu} - ig'B^\mu\right)(v+h)^2, \tag{2.18}$$

where the charged vector bosons $W_\mu^\pm$ are identified as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^{(1)} \mp iW_\mu^{(2)}\right). \tag{2.19}$$

The masses of the gauge bosons appear in the terms that are quadratic in the W and B fields. The terms proportional to $W_\mu^{(1)}W^{(1)\mu}$ and $W_\mu^{(2)}W^{(2)\mu}$ appear with factors $\frac{1}{2}(\frac{1}{2}g_W v)^2$, and therefore the mass of the W bosons is

$$m_{W^\pm} = \frac{1}{2}g_W v. \tag{2.20}$$

The third massive gauge boson is given by

$$Z_\mu = \frac{1}{\sqrt{g_W^2 + g'^2}}\left(g_W W_\mu^{(3)} - g'B_\mu\right), \tag{2.21}$$

with mass

$$m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2}. \tag{2.22}$$

Lastly, the photon is identified as

$$A_\mu = \frac{1}{\sqrt{g_W^2 + g'^2}} \left(g' W_\mu^{(3)} + g_W B_\mu \right) \tag{2.23}$$

with mass

$$m_A = 0. \tag{2.24}$$

The fields corresponding to the physical gauge bosons, including three massive ones, are therefore linear superpositions of the massless bosons that emerged from the imposed $\mathrm{SU(2)_{weak} \times U(1)_Y}$ gauge symmetry. After EWSB, the photon is identified as the massless gauge boson of electromagnetism, preserving gauge invariance under $U(1)_{\mathrm{EM}}$. The ratio of the heavy-gauge-boson masses, $m_W/m_Z = \cos\theta_W$, has been experimentally verified, providing compelling evidence for the BEH mechanism. By using measured values for m_W and g_W, the vacuum expectation value of the Higgs field is predicted to be

$$v = 246 \text{ GeV}, \tag{2.25}$$

which, together with the recently measured values of the Higgs boson mass of $m_H \approx 125$ GeV, allows determination of the parameter λ quantifying the triple and quartic Higgs self-couplings. In addition to the gauge-boson masses, the BEH mechanism is also responsible for generating fermion masses through a Yukawa interaction [8], in which the coupling strength between the Higgs boson and the fermions is proportional to the fermion masses. The details of this mechanism are omitted here for brevity and may be found in References [11] and [17]. .

2.1.6 Discovery of the Higgs boson

A high-energy proton-proton (pp) collider such as the Large Hadron Collider, which operates at a centre-of-mass energy of up to 14 TeV and which is the subject of Chapter 3, is sensitive to all four Higgs production mechanisms: the fusion of two gluons (ggF, sometimes referred to as $pp \to H$), vector boson fusion (VBF, sometimes referred to as $pp \to qqH$), the associated production of a Higgs with a

vector boson ('Higgs-strahlung', or $pp \to WH/ZH$), and the associated production of a Higgs with a pair of top quarks (ttH, or $pp \to t\bar{t}H$) (c.f. Figure 2.4). The SM Higgs boson production cross-sections and branching fractions are illustrated in Figure 2.5. The 125-GeV Higgs boson's branching fractions are illustrated in Table 2.1.

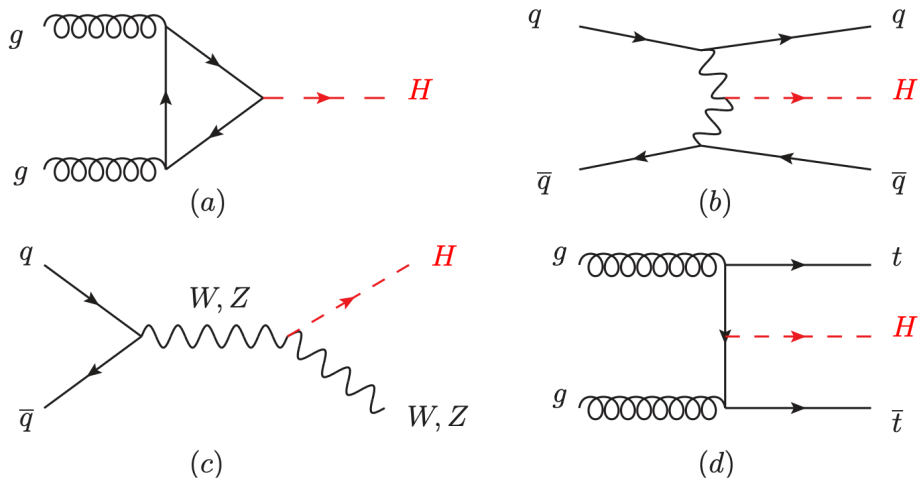
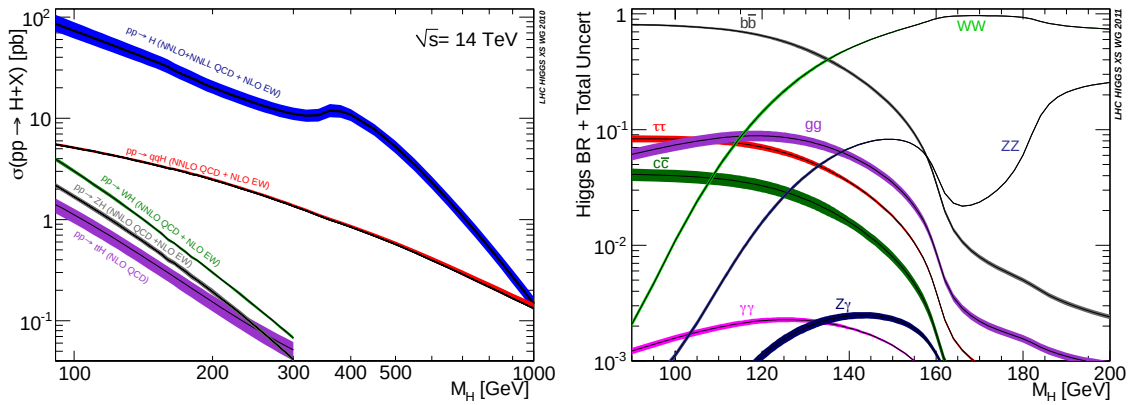


Figure 2.4: Feynman diagrams of the production of a SM Higgs boson through (a) gluon-gluon fusion, (b) vector-boson fusion, (c) associated production with a vector boson, and (d) associated production with a pair of top quarks. Figure reproduced from Reference [23].

Decay channel	Branching fraction	Rel. uncertainty
$H \to \gamma\gamma$	2.27×10^{-3}	$\pm 2.1\%$
$H \to ZZ$	2.62×10^{-2}	$\pm 1.5\%$
$H \to W^+W^-$	2.14×10^{-1}	$\pm 1.5\%$
$H \to \tau^+\tau^-$	6.27×10^{-2}	$\pm 1.6\%$
$H \to b\bar{b}$	5.82×10^{-1}	$^{+1.2\%}_{-1.3\%}$
$H \to c\bar{c}$	2.89×10^{-2}	$^{+5.5\%}_{-2.0\%}$
$H \to Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \to \mu^+\mu^-$	2.18×10^{-4}	$\pm 1.7\%$

Table 2.1: Branching fractions and relative uncertainties of the 125-GeV SM Higgs boson. Table adapted from Reference [26].

Using LHC pp collision data recorded at $\sqrt{s} = 7$ and 8 TeV in 2011 and 2012, the CMS and ATLAS collaborations independently observed a new boson with a mass of $m_H \approx 125$ GeV that is fully compatible with the SM Higgs boson. The primary production process involved in the discoveries was ggF, with the Higgs



(a) Higgs production cross-section at $\sqrt{s} = 14$ TeV as a function of the Higgs boson mass.

(b) Higgs branching fraction as a function of the Higgs boson mass.

Figure 2.5: (a) Standard Model Higgs production cross-section at $\sqrt{s} = 14$ TeV, and (b) branching fractions as a function of Higgs mass. The ggF and VBF production cross-sections are the two uppermost lines in plot (a), labelled '$pp \to H$' and '$pp \to qqH$'. Both quantities are shown as a function of the Higgs boson mass, m_H. The uncertainty bands represent one standard deviation. Figures reproduced from References [24] and [25].

decaying further into $ZZ \to \ell^+\ell^-\ell'^+\ell'^-$, $\gamma\gamma$, and $W^+W^- \to \ell^+\ell^-\nu\bar{\nu}$. While having rather low branching fractions at the observed value of m_H, as shown in Figure 2.5b, the $\ell^+\ell^-\ell'^+\ell'^-$ and $\gamma\gamma$ final states showed the highest sensitivity due to very clean detector signatures and backgrounds from other SM processes that are well understood. In a process called *triggering* (c.f. Section 3.2.6), particle detectors rapidly decide which events among a large number of them that occur nearly simultaneously in a single pp bunch crossing could be of interest and should therefore be recorded. The $\gamma\gamma$ and $ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ final states have the added advantage of being easy to trigger on due to their clear e/γ signatures in the calorimeter, as well as their μ signals in the muon system. Combining all three decay channels, the CMS collaboration observed the Higgs boson with a local significance of 5.0 σ, while ATLAS measured a local significance of 5.9 σ.

2.1.7 Limitations of the Standard Model

A number of phenomena are not explained by the SM, motivating searches for physics beyond the SM. A brief overview of the main phenomena is given, with a dedicated section on searches for an extended Higgs sector presented in Section 2.2.1.

- **Dark matter**. It is known from cosmological observations that there must exist a form of matter that does not interact electromagnetically, and which is five times more abundant than ordinary (baryonic) matter, called *dark matter*. Direct evidence for dark matter comes from studying the rotational velocities of stars around their galactic centres. For a star outside the centre of the galaxy, that is, in the absence of matter, its rotational velocity around the galactic centre should be predicted by classical mechanics in a Newtonian gravitational field. If the distance from the galactic centre of a star is r, its velocity should decrease like $1/\sqrt{r}$, which is not what is observed. Instead, astronomical observations indicate that the mass distribution of the galaxy cannot be concentrated in the galactic centre, but is distributed roughly like $M(r) \sim r$ [17].

- **Dark energy**. The SM of particle physics does not account for dark energy. Experimental surveys of the cosmic microwave background (CMB) have provided strong evidence for the cosmological standard model referred to as the ΛCDM model. In the ΛCDM model, only 5% of the total energy density of the universe is provided by baryonic matter. 23% consist of dark matter, and the majority of it, 72%, consists of *dark energy*. In this model, dark energy is associated with a non-zero cosmological constant in General Relativity, and is therefore thought to be linked to the expansion of the universe [17].

- **Baryon asymmetry**. The universe is evidently made up mostly of matter, yet the SM predicts that matter and antimatter should have been created in equal amounts. In order to generate the apparent relative ubiquity of baryons over antibaryons observed today from the CMB, there must have existed $10^9 + 1$ bayons for every 10^9 antibaryons in the early universe [17]. Matter-antimatter annihilation between them would then yield 10^9 photons and one leftover baryon. The SM provides a mechanism for CP violation, which is one of the Sakharov conditions [27] necessary to generate a baryon asymmetry. Experimental evidence of CP violation in the quark sector—gathered by the BaBar detector [28] at SLAC, the Belle experiment [29] at KEK, and the LHCb Collaboration [30] at CERN—only accounts for a small fraction of the observed matter-antimatter asymmetry.

- **Neutrino masses and oscillations**. In the SM, neutrinos are massless particles, yet neutrino oscillation experiments have shown that neutrinos have mass [31,

32]. One possible explanation is that neutrinos are Majorana particles, that is, they are their own antiparticles. The Majorana nature of neutrinos can be experimentally tested through neutrinoless double beta decay, which has to date not been observed, and puts strong constraints on the neutrino masses of $\mathcal{O}(1)$ eV [33].

Several additional theoretical limitations suggest that the SM cannot be a complete theory. Perhaps the most striking one of these is that the SM does not include the gravitational force, nor does it explain why gravity is 38 orders of magnitude weaker than the strong force. Further, a theoretical tension known as the *hierarchy problem* highlights that the BEH mechanism responsible for mass generation requires that the Higgs scalar field have a mass of $\mathcal{O}(100)$ GeV in order to give the observed values of the massive weak gauge bosons. The mass term of a scalar field, however, receives quadratic radiative corrections in a prescription known as renormalisation. If the SM has a cut-off scale around the grand unification energy (10^{19} GeV), the Higgs field's bare mass would have to receive cancellations of order 10^{38} GeV so as to yield the physically observed Higgs pole mass[3] of $m_H \approx 125$ GeV. This extreme sensitivity of observable quantities, such as the Higgs mass, to tiny variations in the parameters of the theory is called *fine-tuning* [11, 34]. Theories that propose an extended Higgs sector, such as the two-Higgs-doublet model we will encounter in Section 2.2.1, aim to provide theoretical solutions to the hierarchy problem.

2.2 Phenomena beyond the Standard Model

As outlined in the previous section, the SM has a number of limitations, and numerous efforts are underway to find experimental signatures that hint at BSM physics. Among these, this section focuses on searches for an extended Higgs sector and graviton signatures at the LHC.

2.2.1 Searches for an extended Higgs sector

Supersymmetry (SUSY) is a proposed theory that addresses the hierarchy problem by positing the existence of fermionic superpartners for all bosons in the SM, as well as bosonic superpartners of all fermions in the SM. Quantum loop corrections to the

[3]The *pole mass* is the physical mass of a particle, corresponding to the rest mass in special relativity. It is what experiments typically report. It is independent of any renormalisation scheme used to subtract infinities in loop corrections.

bare Higgs mass squared, m_H^2, through a generic fermion loop yield a contribution of order $-\Lambda^2$, where Λ is the ultraviolet cut-off, i.e. the scale up to which the SM is valid [22]. If the SM is valid up to the Planck scale—where quantum-gravitational effects become relevant, i.e. $\Lambda \approx 10^{19}$ GeV—the Higgs mass squared receives corrections $O(10^{38})$ GeV. The loop of a generic scalar field on the other hand yields corrections of order Λ^2. SUSY imposes the existence of a bosonic superpartner for every fermion, and vice versa. Therefore, in a SUSY scenario, the quadratic radiative corrections to the bare Higgs boson mass would cancel between SM particles and their superpartners. As a result, the discrepancy between the electroweak gauge bosons and the grand unification scale is obtained naturally. The corrections to Higgs mass squared due to fermionic and bosonic loops, as imposed by SUSY, are illustrated in Figure 2.6. As yet, no experimental signatures of SUSY have been found, and a number of experiments have placed strong limits on potential SUSY scenarios. Historically, these have been obtained by collider experiments at the SPS [35], LEP [36], and Tevatron [37]. The most stringent limits to date have been set at the LHC [38].

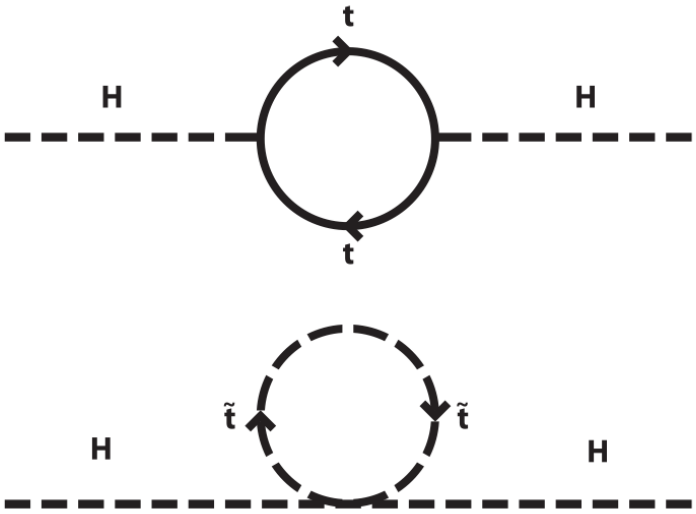


Figure 2.6: Feynman diagrams of one-loop corrections to the Higgs mass squared. The top part of the figure shows the one-loop correction due to a loop of a top quark, yielding a correction of order $-\Lambda^2$. The bottom part shows the one-loop correction due to the bosonic superpartner of the top quark, the *stop* quark, with a contribution of order Λ^2. Figure reproduced from Reference [39].

The minimal SUSY model is called the *Minimal Supersymmetric Standard Model* (MSSM) [40], and it provides the simplest SUSY extensions to the SM. The MSSM

postulates an extended Higgs sector in the form of a specific type of a two-Higgs-doublet model (2HDM) [41]. In the 2HDM, the Higgs sector consists of two doublets, Φ_1 and Φ_2, which, after electroweak symmetry breaking, yield five physical Higgs bosons: two charged Higgs bosons ($H^\pm$), a CP-even Higgs boson (H), a CP-even neutral scalar corresponding to the observed SM Higgs boson (h), and another CP-odd neutral scalar (A). The two doublets can be written as [42]

$$\Phi_i = \begin{pmatrix} \phi_i^+ \\ (v_i + \phi_i^0 + i\varphi_i)/\sqrt{2} \end{pmatrix}, \tag{2.26}$$

where $i = 1, 2$, and the vacuum expectation values must satisfy $\sqrt{v_i^2 + v_2^2} = v = 246$ GeV. The free parameters in the 2HDM are the five physical Higgs masses (m_h, m_H, m_A, and $m_{H^\pm}$), the ratio between the vacuum expectation values of the Φ_i, $\tan\beta = v_1/v_2$, and α, the mixing angle between h and H.

There are several types of 2HDM, classified according to the coupling behaviour of the Φ_i to up-type or down-type quarks. The SUSY-compatible 2HDM[4] is referred to as 'Type II', and in it, Φ_1 couples to down-type quarks and all charged leptons, while Φ_2 couples to up-type quarks. Another commonly considered scenario is the 'Type I' 2HDM, in which Φ_2 couples to all quarks and leptons (in this scenario, Φ_1 does not couple to any charged fermion). 2HDMs can be broadly categorised into flavour-conversing and flavour-violating types. The four flavour-conserving types and their fermion couplings are summarised in Table 2.2.

Model	u	d	ℓ
Type I	Φ_2	Φ_2	Φ_2
Type II	Φ_2	Φ_1	Φ_1
Lepton-specific	Φ_2	Φ_2	Φ_1
Flipped	Φ_2	Φ_1	Φ_2

Table 2.2: Summary of the fermion couplings of the four 2HDMs which lead to flavour conservation. The column labelled 'u' corresponds to couplings to up-type quarks, the one labelled 'd' to couplings to down-type quarks. The column labelled 'ℓ' refers to couplings to charged leptons. Table reproduced from Reference [41].

There are, broadly speaking, two approaches to testing the existence of an extended Higgs sector: through precision measurements of the couplings of the

[4] 2HDMs need not necessarily be considered in a SUSY or MSSM context. SUSY models, however, cannot be constructed with a single complex Higgs doublet, necessitating at least two complex doublets.

SM Higgs boson, and direct searches for any of the other, as-yet-unobserved Higgs bosons [42]. Direct 2HDM searches typically separately consider the mass-degenerate case, in which some of the physical 2HDM Higgs bosons have the same, or very similar, masses, and the hierarchical case. Current 2HDM searches at the LHC usually focus on the decays of $H, H^{\pm}$, and A to $f\bar{f}$, W^+W^-, and ZZ, as well as final states involving the SM Higgs boson. If there exists a hierarchy between the five Higgs boson mass parameters, a richer phenomenology emerges, allowing for exotic decays such as $H \to AZ$, $H \to HZ$, $A \to AZ$, $A \to HZ$, $H \to H^{\pm}W^{\mp}$, and several others [42]. Presented later in this work is a search for heavy resonances beyond the SM, one interpretation of which is as a search for a 2HDM scalar (H) decaying to two Z bosons. Both ATLAS and CMS have conducted searches for a number of decay modes of A/H to SM particles, which are summarised in Table 2.3.

2HDM Higgs decay channel	CMS search	ATLAS search
$A/H \to \mu^+\mu^-$	[43]	[44]
$A/H \to b\bar{b}$	[45]	[46]
$A/H \to \tau^+\tau^-$	[47, 48]	[49]
$A/H \to \gamma\gamma$	[50, 51]	[52, 53]
$A/H \to t\bar{t}$	[54]	-
$H \to W^+W^-$	[55]	[56]
$H \to ZZ$	[57]	[58]

Table 2.3: Summary of CMS and ATLAS results of direct searches for an extended Higgs sector in the 2HDM. Building and improving on the result obtained in Reference [58] will be the subject of Chapter 5. Table adapted from Reference [42].

2.2.2 Graviton searches

One way of incorporating gravity into the SM is through extra spatial dimensions. These can be probed by searching for signatures of gravitational-field excitations in the form of *gravitons*. In Kaluza-Klein (KK) theory of the 1920s [59, 60], the unification of electromagnetic and gravitational forces was provided by adding a fifth dimension to the Newtonian gravitational potential and reconsidering electromagnetism in five dimensions, instead of the ordinary four spacetime dimensions. This idea was extended more recently; in models beyond the SM with large extra dimensions, the SM fields propagate in the four spacetime dimensions, while gravitons can propagate freely across all dimensions. This so-called ADD model [61], named after its originators, Arkani-Hamed, Dimopoulos and Dvali, explains the observed

weakness of gravity compared to the remaining SM forces and proposes that the combined effect of the gravitons in the extra dimensions and their excitations in the five-dimensional KK theory—called KK excitations—become observable at the TeV scale.

The ADD model confines the SM particles and forces to a four-dimensional subspace called the *3-brane* [62]. While it provides a solution to the hierarchy discrepancy between the weak scale and the grand unification scale, it introduces an additional hierarchy between the size of the 3-brane compared to the extra dimensions and the electroweak scale. Overcoming this problem, two models proposed by Randall and Sundrum, RS1 [62] and RS2 [63], suggest the existence of a single *warped* extra dimension that separates the 3-brane associated with the SM fields from the one in which gravity dominates (called the *Planckbrane*). In the RS1 model, the energy scale in the warped extra dimension is large at the end connected to the Planckbrane, and small at the end connected to the 3-brane. The RS2 model is a modified version of RS1, in which there does not exist a 3-brane, and the SM particles and fields exist in the Planckbrane.

Both CMS and ATLAS have searched for RS gravitons, with CMS so far placing the most stringent limits on their production cross-section. Depending on the choice of the natural energy scale in the RS1 model, CMS excluded gravitons with masses below 0.5–4.5 TeV [64]. The most stringent limits set by ATLAS exclude RS1 gravitons in the mass range 1.3–1.8 TeV [65].

Chapter 3

High Energy Physics Experiments

The ability to resolve distinct objects that are close to one another is limited by the de Broglie wavelength, which relates wavelength of a massive particle, λ, to its momentum, p, according to $\lambda \sim 1/p$. Thus, probing the structure of massive particles at small length scales requires high momenta, and the scale that can be probed, in other words, the resolution of the probe, decreases with increasing momenta.

Experimental high energy physics has therefore endeavoured to build machines with higher and higher energies. The most powerful machine ever built for this purpose is the Large Hadron Collider (LHC) at CERN. Its aim is to accelerate bunches of protons to very high energies and collide them head-on at four distinct collision points. Around these collision points are located detectors that aim to measure the fundamental properties of the elementary particles resulting from these collisions. The LHC hosts two general-purpose detectors designed to measure the broadest possible range of phenomena, namely the ATLAS detector [66], which recorded the data which is the subject of this thesis, and the CMS detector [67].

The LHC is briefly introduced in Section 3.1. As this work studies collisions observed in the ATLAS detector, it is described in detail in Section 3.2.

3.1 The Large Hadron Collider

The LHC is a circular proton-proton (pp) collider, consisting of a number of machines aimed at accelerating protons, keeping them on a curved trajectory, and colliding them at dedicated points. The protons circulate in two beams in opposing directions around a 27-km ring located 100 m underground at CERN. Along the ring are located radio frequency (RF) cavities that accelerate protons, as well as superconducting magnets designed to force them onto a circular trajectory, and focus the beams within which they travel. The protons originate from a bottle of hydrogen gas; the electrons are stripped from the H_2 gas, and the resulting H^+ ions are accelerated in a linear accelerator called LINAC 2, generating protons with a momentum of 50 MeV. The protons are directed into the Proton Synchrotron Booster (PSB), in which they are further accelerated to 1.4 GeV, followed by another acceleration stage in the Proton Synchrotron (PS) to a momentum of 26 GeV. Within the PS, they reach a momentum of 450 GeV, after which they are injected into the main LHC ring. Here, they reach their final energy of 6.5 TeV, leading to a centre-of-mass collision energy of $\sqrt{s} = 13$ TeV. The two circulating proton beams each contain a maximum of 2808 populated bunches, with each bunch containing 1.15×10^{11} protons. The spacing between two bunch crossings is 25 ns, giving a collision rate at each of the four interaction points of 40 MHz. The CERN accelerator complex, including the LHC machine, is illustrated schematically in Figure 3.1.

The rate of events produced in a head-on pp collision in the LHC is given by

$$\dot{N} = \mathcal{L}\sigma, \tag{3.1}$$

where $\mathcal{L}$ is the instantaneous luminosity of the LHC, and σ is the cross-section of the events under consideration [69]. Assuming that each particle beam has a normally distributed profile, the instantaneous luminosity is given by

$$\mathcal{L} = \frac{N_b^2 n_b f \gamma}{4\pi\epsilon_n \beta^*}, \tag{3.2}$$

where N_b is the number of protons per bunch, n_b is the number of bunches, f is the revolution frequency of the bunches, γ is the relativistic Lorentz factor, β^* is the beta function[1] at the collision point, and ϵ_n is the normalised transverse

[1]The *beta function* is a function related to the transverse size of the particle beam along the beam direction. For location x along the beam, it is given by $\beta(x) = {}^{\sigma^2(x)}/_{\epsilon_n}$, where $\sigma(x)$ is the width of the Gaussian beam profile at position x.

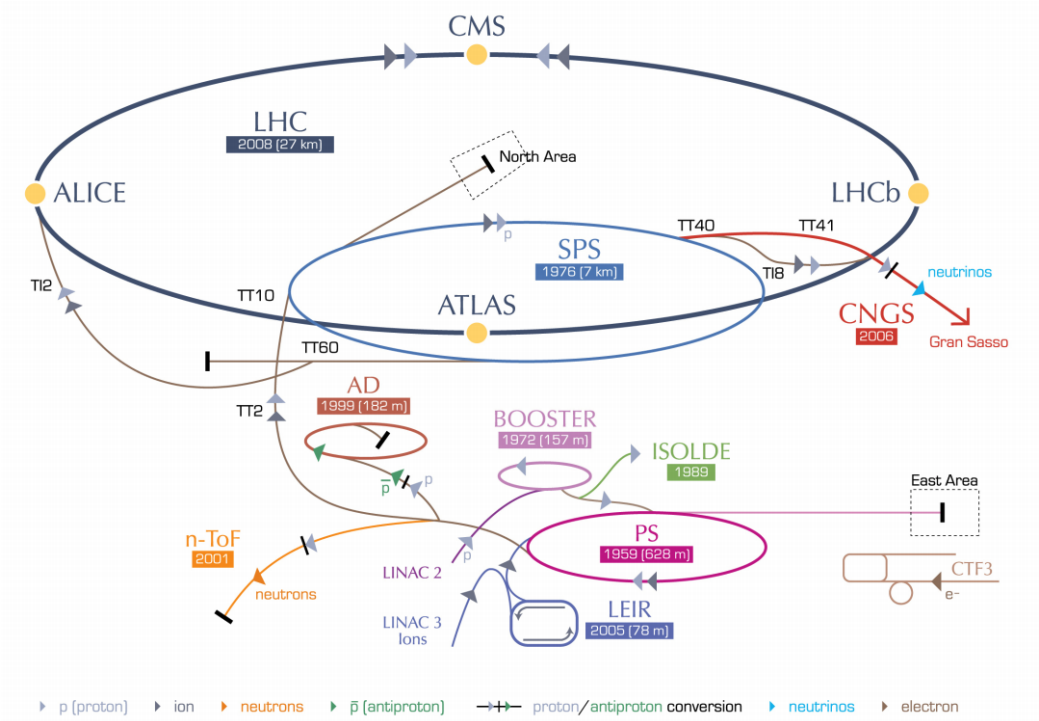


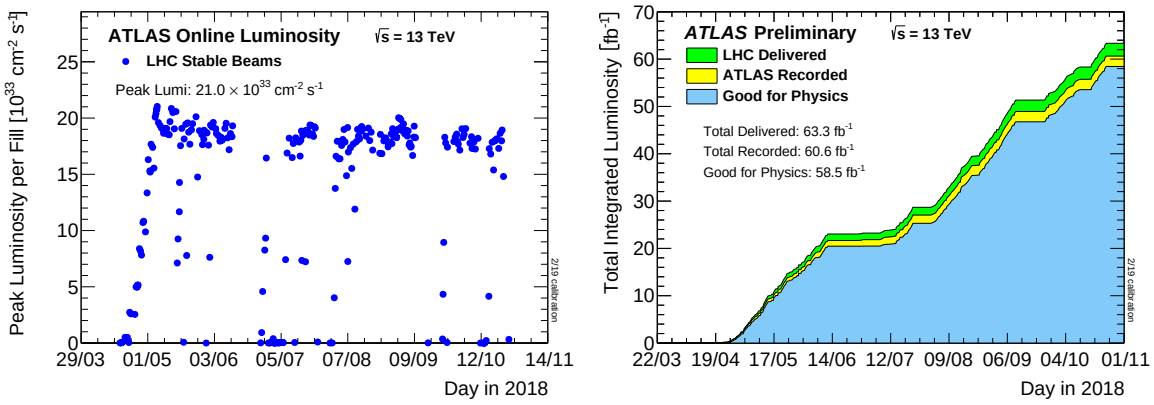
Figure 3.1: The CERN accelerator complex, including the 27-km LHC ring (dark blue). The hydrogen source is located near LINAC 2 (labelled 'p'), and the protons' ensuing sequence of acceleration can be traced through the PSB, the PS, and into the main LHC ring. Figure reproduced from Reference [68].

beam emittance [69]. In general, the instantaneous luminosity is a function of time, i.e. $\mathcal{L} \Rightarrow \mathcal{L}(t)$. In order to determine the total number of events observed for a given process with cross-section σ over a time interval $[t_1, t_2]$, one can integrate Equation (3.1), namely

$$N = \int_{t_1}^{t_2} dt\ \mathcal{L}(t)\sigma = \mathcal{L}_{\text{int}}\sigma, \tag{3.3}$$

where $\mathcal{L}_{\text{int}}$ is the *integrated luminosity*.

The instantaneous luminosity delivered by the LHC varies depending on the experiment the machine caters to. However, in the case of the two general-purpose experiments ATLAS and CMS, the targeted design luminosity for each experiment is $\mathcal{L} = 10^{34}$ cm^{2}s^{-1}. The peak instantaneous and the total integrated luminosity measured during 2018 at the LHC, and delivered to ATLAS, are shown in Figure 3.2.



(a) Peak instantaneous luminosity delivered to ATLAS during 2018.

(b) Total integrated luminosity delivered by the LHC to ATLAS, as well as that recorded by ATLAS in 2018.

Figure 3.2: The LHC's peak instantaneous luminosity (a) as well as the total integrated luminosity delivered to ATLAS (b). Subfigure (b) also shows the total luminosity recorded by ATLAS as well as the fraction of it that is deemed usable in physics analyses. Both figures pertain to the data-taking year 2018. Figures reproduced from Reference [70].

The LHC delivered an integrated luminosity of 30 fb^{-1} during the data-taking period 2011–2012, referred to as *Run 1*. The centre-of-mass energy during Run 1 was $\sqrt{s} = 7 - 8$ TeV, and the bunch spacing was 50 ns. The LHC subsequently underwent a two-year-long shutdown in which upgrades to the machine allowed for an increase in both instantaneous luminosity and collision energy. The period

following the shutdown is referred to as *Run 2*, and in it the bunch spacing was 25 ns, while the centre-of-mass energy was increased to $\sqrt{s} = 13$ TeV. Comprising a total of four data-taking years, the LHC delivered a total integrated luminosity of 150 fb^{-1} in Run 2.

The point at which hard-scatter events occur within a pp bunch crossing is referred to as the *primary vertex* (PV). In any given bunch crossing, a number of other inelastic pp collisions are expected to occur, and one of the tasks of the experimentalists and the detector is to appropriately assign particle tracks and energy deposits to the PV, while reducing contamination from tracks originating from additional inelastic collisions. The additional collisions are referred to as *pile-up*. The pile-up occurring during a single bunch crossing is referred to as *in-time pile-up*, while pile-up from neighbouring bunch crossings is referred to as *out-of-time pile-up*. The latter can occur when events are misattributed to a bunch crossing, for example due to detector read-out rates that are lower than the bunch-crossing frequency. The mean number of interactions per bunch crossing is denoted by μ, which follows a Poisson distribution with mean $\langle\mu\rangle$. Figure 3.3 shows the mean number of interactions per bunch crossing recorded by ATLAS during Run 2 of the LHC.

3.2 The ATLAS experiment

The ATLAS (A Toroidal Lhc ApparatuS) detector is a cylindrical general-purpose detector designed to measure particles produced in high-energy pp collisions, covering nearly the entire solid angle. It consists of several subsystems, each of which is designed to measure different properties of the particles resulting from the pp collisions.

ATLAS uses a right-handed coordinate system with the centre of the detector as the origin [66]. The z-axis corresponds to the LHC's beam line in the counter-clockwise direction, while the plane transverse to the beam line is defined by the x- and y-axes. Positive x is taken to point towards the centre of the LHC ring, while positive y points upwards. The radial distance from the origin in the transverse plane is given by $r = \sqrt{x^2 + y^2}$. ϕ describes the azimuthal angle, with the vector pointing to $\phi = 0$ directed along positive x towards the centre of the LHC ring. The polar angle is measured from the beam axis along positive z, and is defined by $\theta = \arctan(r/z)$. The pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. The distance between two vectors separated in the transverse plane by $\Delta\phi$ and in

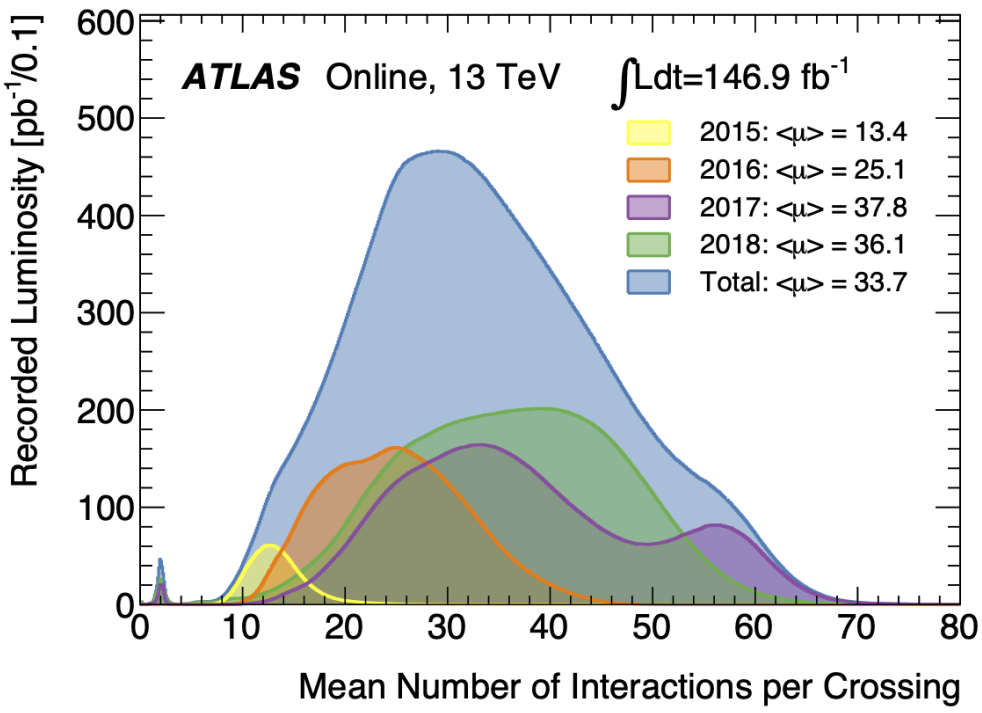


Figure 3.3: Distribution of the mean number of interactions per bunch crossing, μ, for the four data-taking years of Run 2 as recorded by ATLAS. The y-axis shows the total recorded luminosity per $0.1\,\mu$. The legend indicates the mean of the distribution of μ, denoted by $\langle\mu\rangle$, in the four data-taking years as well as in total. Figure reproduced from Reference [70].

pseudorapidity by $\Delta\eta$ is given by $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The projection of a particle with momentum vector $\boldsymbol{p}$ onto the transverse plane is referred to as the *transverse momentum*, $\boldsymbol{p}_{\mathrm{T}}$, with magnitude p_{T}. The analogous projection for the energy gives the *transverse energy*, $\boldsymbol{E}_{\mathrm{T}}$, with magnitude E_{T}.

A schematic overview of the ATLAS detector is shown in Figure 3.4. A tracking system at the centre of the detector is designed to measure the trajectory of charged particles inside a solenoidal magnetic field (Section 3.2.1). Surrounding this is a system of electromagnetic and hadronic calorimeters aimed at stopping and measuring the energy left by particles that enter it (Section 3.2.2), followed by a spectrometer system based on toroidal magnets that is specialised in identifying muons (Section 3.2.4). Coordinating the read-out electronics of all subsystems is the task of an intricate triggering system, which aims to reduce the LHC's bunch-crossing rate of 40 MHz to a manageable value for all sub-detectors and data-processing systems (Section 3.2.6).

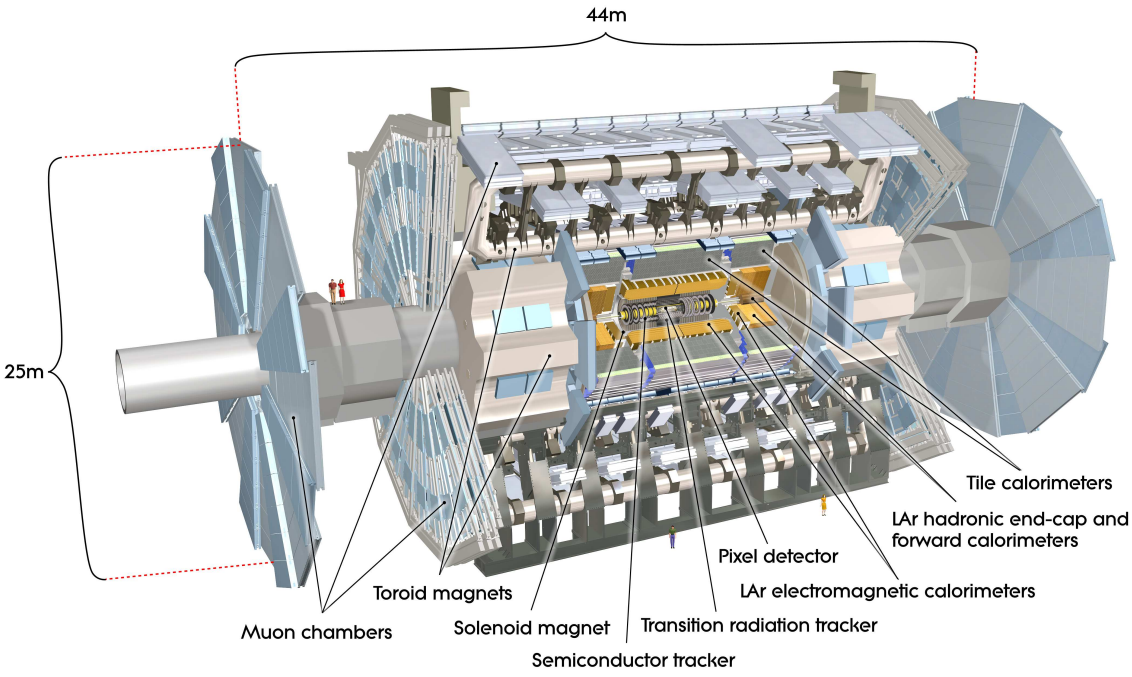


Figure 3.4: Cut-away view of the ATLAS detector. The different detector subsystems are labelled. The dimensions of ATLAS are 44 m in length and 25 m in diameter, with a total weight of 7,000 tonnes. Figure reproduced from Reference [66].

3.2.1 Inner detector

The ATLAS inner detector (ID) is designed to measure the trajectories and momenta of charged particles, and provide accurate vertex information on up to 1000 particles that are expected to emerge from each bunch crossing [66]. The ID achieves high momentum and position resolution through separate subsystems: the pixel detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT). The layout of the ID is shown schematically in Figure 3.5.

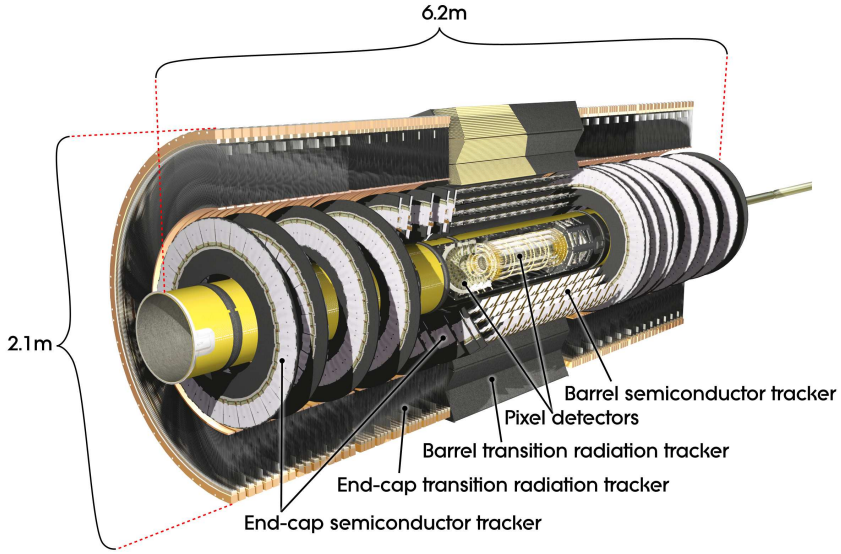


Figure 3.5: Cut-away view of the ATLAS inner detector. The ID consists of three sub-detectors (pixel detector, SCT and TRT), which are labelled. The separate barrel and end-cap regions are shown for each of the ID's sub-detectors. Figure reproduced from Reference [66].

The ID is located inside a 2-T magnetic field generated by a solenoid, designed to force charged particles onto curved trajectories in order to measure their momenta. The central region of the ID is referred to as the *barrel region* and covers a range in pseudorapidity of $|\eta| < 2.5$, while the region beyond the barrel is referred to as the *forward region* ($|\eta| \geq 2.5$). The ID's subsystems are as follows:

- **Pixel detector**. The innermost layer of the ID is the pixel detector, and it is the highest-granularity instrument of the ID. The sensors are arranged in four concentric cylinders, the innermost one of which is called the *insertable B-layer*, which was added during the extended shutdown after Run 1. The size of each pixel is 50×400 μm^2, and the overall spatial resolution of particles that traverse it is 14×115 μm^2. Particles produced within the barrel region traverse four

pixel layers on average, while particles in the forward region traverse three *end-cap* pixel layers. The pixel detector has a total of 80 million channels and a surface area of 1.7 m^2.

- **SCT**. Surrounding the pixel detector is the SCT, which comprises four layers of silicon microstrip detectors, both in the barrel and in the forward region, each with an area of 6.36×6.40 cm^2, and 768 strips with a pitch of 80 µm. The total detector area of the SCT is 61 m^2 with 6.2 million readout channels and a spatial resolution of 16×580 μm^2 [71].

- **TRT**. The TRT surrounds the SCT, and it is designed to detect transition radiation, which occurs when a relativistic charged particle traverses the boundary between two materials. The TRT consists of 550,000 straw tubes, each of which has a diameter of 4 mm and length of at most 1.5 m. The tubes are filled with xenon gas, which is ionised by transition radiation. Due to its greater distance from the interaction point, it is sufficiently instrumented with a lower spatial resolution 140 µm while comprising the majority of the ID's overall volume (12 m^3) [66].

Surrounding the ID is the central solenoid generating a 2-T magnetic field with a length of 5.3 m and a diameter of 2.5 m.

3.2.2 Calorimetry

The calorimeters are designed to contain electromagnetic and hadronic showers, and in doing so to measure their energy. The key parameter of a calorimeter designed to stop photons and electrons is its radiation length, X_0, which is defined as the mean distance over which a relativistic electron loses all but $1/e$ of its energy by bremsstrahlung [72]. For stopping hadronic showers, the key parameter is the interaction length, λ_n, defined as the mean free path of a particle between two inelastic interactions [73]. The ATLAS calorimetry system consists of two separate calorimeters: the electromagnetic (EM) calorimeter (ECAL) with a radiation length of $X_0 > 22$ and the hadronic calorimeter (HCAL), which has an interaction length, λ_n, of at least 10 [66].

The fractional calorimeter resolution, $\sigma(E)/E$, is generally a function of particle energy, E, and is commonly parametrised as [74]

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E/\text{GeV}}} \oplus \frac{b}{E/\text{GeV}} \oplus c, \tag{3.4}$$

where a is a stochastic coefficient accounting for statistical fluctuations in the development of particle showers, b is an electronic-noise coefficient, and c is a constant term accounting for inhomogeneities in the detector due to construction and miscalibration.

A schematic overview of the ATLAS calorimeter system is given in Figure 3.6. The energy losses experienced by particles traversing matter are described in detail in Appendix A.1; additional information on the absorption of hadrons in the HCAL is provided in Appendix A.2.

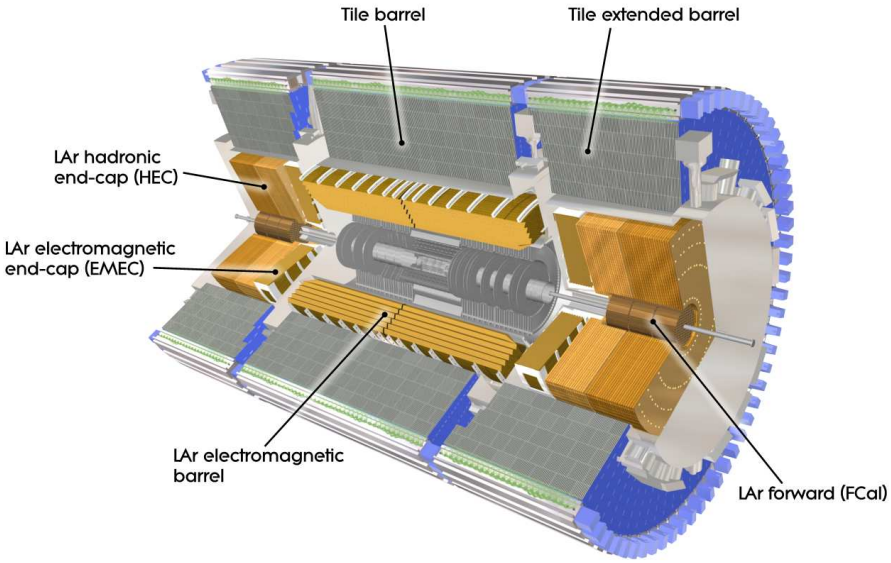


Figure 3.6: Cut-away view of the ATLAS calorimeters. The different components of the EM and hadronic calorimeters are shown, with their respective barrel, extended barrel, and end-cap regions indicated separately. Figure reproduced from Reference [66].

3.2.2.1 Electromagnetic calorimeter

The EM calorimeter comprises two parts: a barrel region ($|\eta| < 1.475$) and end-caps ($1.375 < |\eta| < 3.2$). The calorimeter material consists of interleaved layers of lead as the absorbing material and liquid argon (LAr) as the active material. The barrel region comprises three samplings with varying spatial granularities. The innermost sampling in the barrel region has a granularity in $\Delta\eta \times \Delta\phi$ of 0.003×0.1, followed by 0.025×0.025 in the subsequent sampling, and $0.005 \times 0.0.025$ in the outermost one. In the central region defined by $|\eta| < 1.8$, the EM calorimeter has an additional LAr presampler providing improved precision in that region with a granularity

0.025×0.1. The segmentation in the end-cap region is split up more finely in η, and varies between 0.025×0.1 and 0.05×0.025. The total number of channels in the EM calorimeter is 236,160. A detailed description can be found in Reference [75]. A schematic of an ECAL module in $\eta - \phi$-space is shown in Figure 3.7. The position of all ECAL modules within ATLAS is presented in Figure 3.8, which shows the electronic noise for all ECAL and HCAL modules. The four layers in the central barrel region of ATLAS, which will revisited in Chapter 6, are labelled 'PS', 'EM1', 'EM2', and 'EM3'.

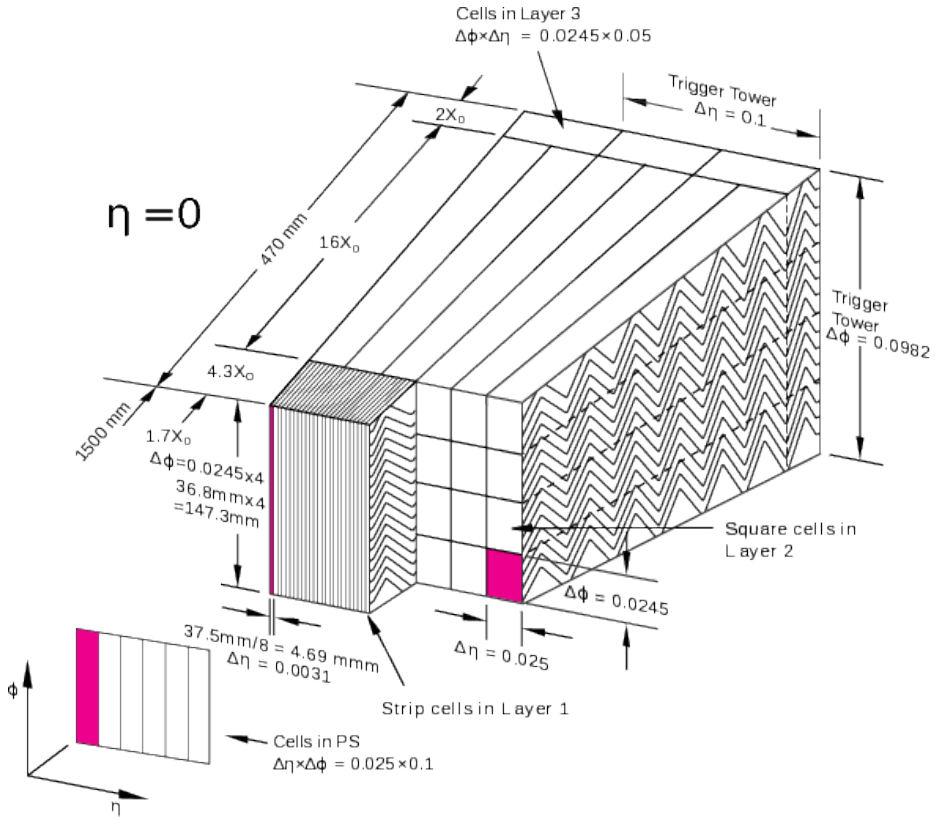


Figure 3.7: Schematic of an ECAL barrel module, located at $\eta = 0$. All four barrel layers are depicted, i.e. one presampler (PS) and three subsequent calorimeter layers. The granularity in η and ϕ is shown for each layer. Figure reproduced from Reference [76].

The resolution of the ECAL as a function of energy has been measured in test-beam campaigns [66]. The result of a performance study of a barrel ECAL module based on an electron test beam is presented in Figure 3.9, which shows the fractional energy resolution as a function of electron energy. Electronic noise has been subtracted from the data, and a fit to Equation (3.4) yields $a = 10.1\%$ and $c = 0.17\%$. The subtracted electronic noise corresponds to a noise coefficient of

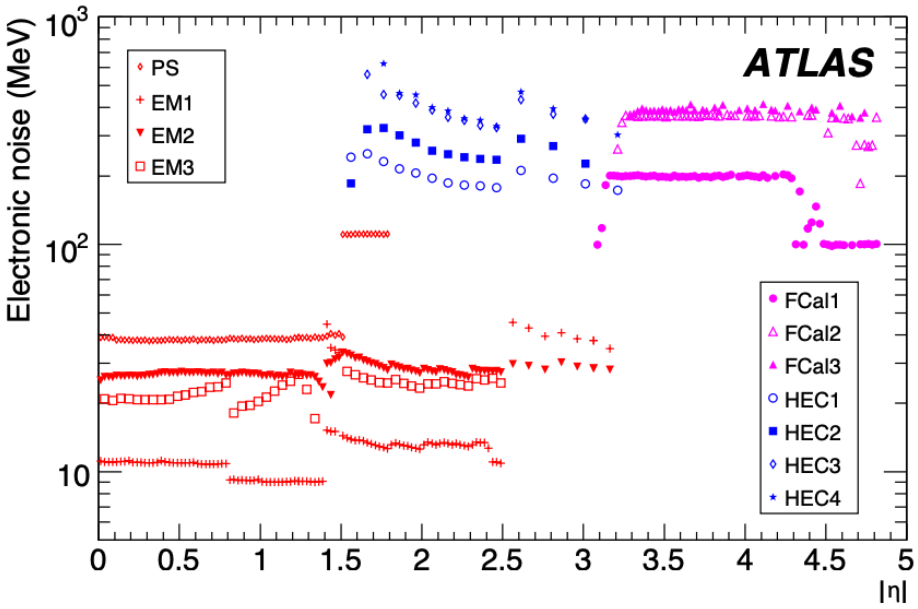


Figure 3.8: Measured electronic noise in all sections of the ECAL and HCAL at zero luminosity. The elements of the barrel region of the electromagnetic calorimeter are plotted in red. Figure reproduced from Reference [77].

$b = 0.2$. An electron with an energy of 100 GeV, such as one that would be common in the $H \to ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ analysis that is the subject of Chapter 5, consequently has an energy resolution in the ECAL of about 1.4 GeV (approximately 1.4%).

3.2.2.2 Hadronic calorimeter

The HCAL is a tile calorimeter comprising layers of steel as the absorbing material and scintillating tiles as the active material [66]. The scintillating tiles are read out by photomultiplier tubes (PMTs). The barrel region of the tile calorimeter has a pseudorapidity coverage of $|\eta| < 1.0$, with a separate extended barrel region covering $0.8 < |\eta| < 1.7$. The HCAL barrel region consists of three samplings with a granularity in $\Delta\eta \times \Delta\phi$ of 0.1×0.1 for the inner two samplings, and 0.2×0.1 for the outer one. The same granularities apply to the extended barrel region, consisting of three samplings on either side of the detector. The end-caps of the HCAL are LAr calorimeters that have a coverage of $1.5 < |\eta| < 3.2$. Consisting of three samplings on either side, the HCAL has a granularity of 0.1×0.1 in the region $1.5 < |\eta| < 2.5$, and 0.2×0.2 beyond that. A third forward calorimeter within the HCAL covers the range $3.1 < |\eta| < 4.9$, consisting of three samplings on either side with a granularity of 0.2×0.2 in each sampling. The total number of readout channels in the HCAL is 17,024. Figure 3.10 shows the structure of an individual HCAL barrel module. The

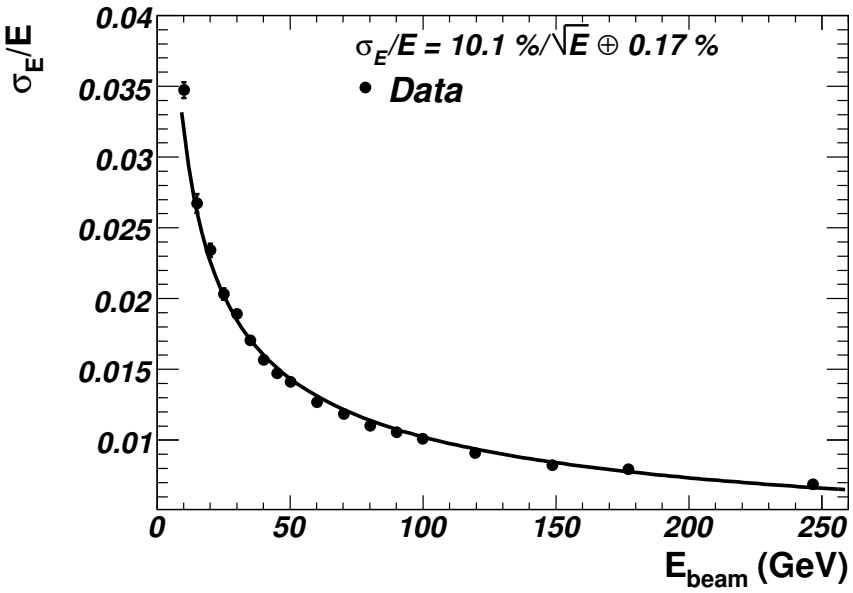


Figure 3.9: Fractional energy resolution as a function of the incident beam energy, E_{beam}, measured in an electron test-beam campaign, for an ECAL module in the barrel region at $|\eta| = 0.687$. Electronic noise was subtracted from the data before plotting the results. The curve represents the results of a fit to the data using Equation (3.4) with $b = 0$. Figure reproduced from Reference [66].

overall segmentation and location of all HCAL modules within the ATLAS detector is depicted in Figure 3.11.

The HCAL energy resolution has been studied as a function of the incident beam energy in pion test beams [66]. The result of one such campaign is shown in Figure 3.12, which shows the fractional energy resolution of combined ECAL and HCAL calorimetry in the barrel region as a function of the inverse square root of the incident beam energy. The data allow for a fit to Equation (3.4), yielding $a = 52\%$, $b = 1.59$, and $c = 3\%$. A pion with an energy of 40 GeV, such as one that would be expected in the background studies presented in Chapter 6, has a resolution of 6.1 GeV, corresponding to a fractional resolution of approximately 15%.

3.2.3 Electron reconstruction

Electrons are reconstructed combining information from the electromagnetic calorimeter and the ID. Electrons are reconstructed using a topological cell-clustering approach based on variable-size, or *dynamical*, calorimeter clusters [3, 78, 79]. The dynamical approach allows for an improved measurement of the electron energy in situations where the electron radiates bremsstrahlung. Electron candidates are

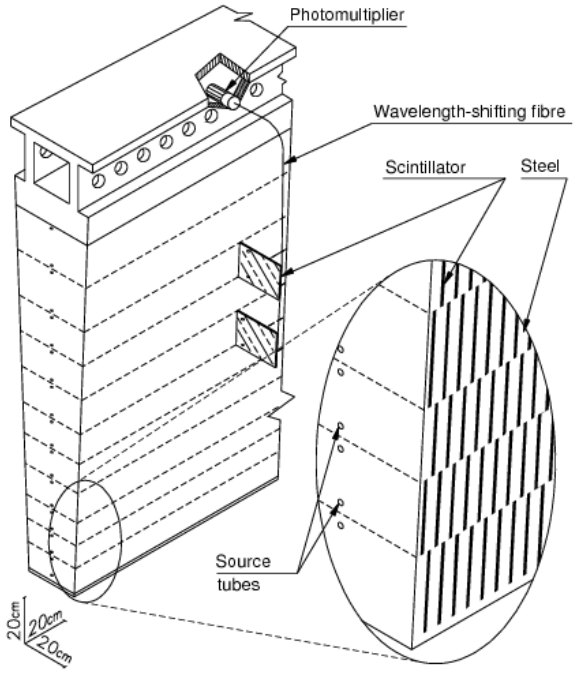


Figure 3.10: Mechanical structure of an HCAL module, showing the layered arrangement of steel and scintillating elements, as well as the light collection through PMTs. Figure reproduced from Reference [66].

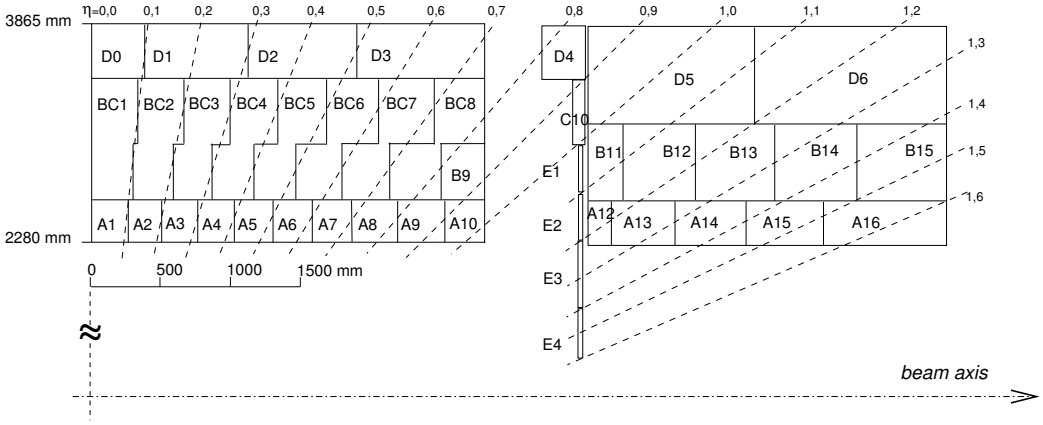


Figure 3.11: Segmentation in depth and η of the HCAL modules in the central (left) and extended (right) barrels. The HCAL is symmetric about the interaction point at the origin. The y-axis corresponds to the radial distance from the beam pipe. Figure reproduced from Reference [66].

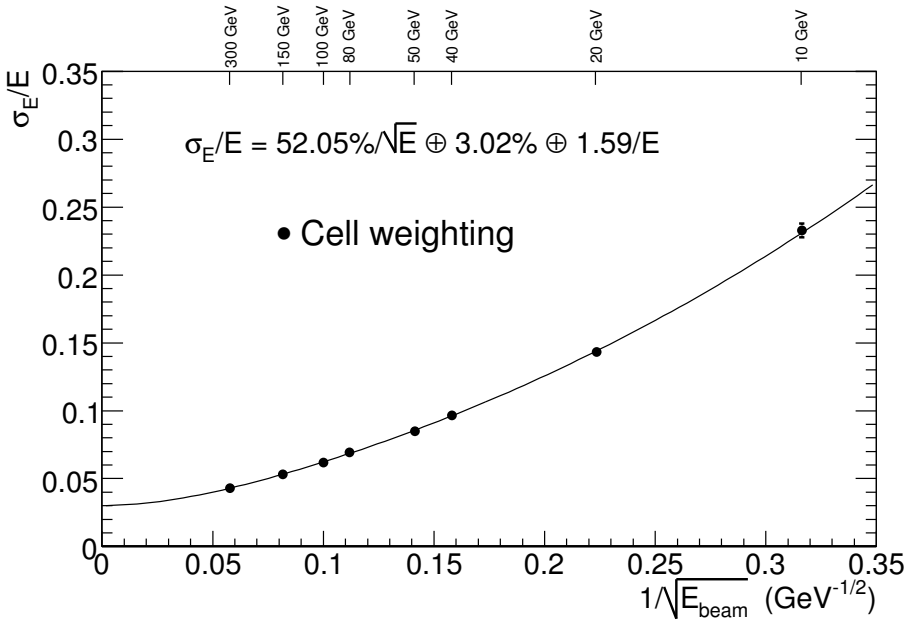


Figure 3.12: Fractional energy resolution obtained for pions as a function of the inverse square root of the beam energy, E_{beam}, for combined ECAL and HCAL calorimetry at an angle of incidence equivalent to $|\eta| = 0.35$. Figure reproduced from Reference [66].

constructed by forming clusters of energy deposits in the EM calorimeter that are associated with ID tracks. The matching of ID tracks to calorimeter clusters is performed after fitting the tracks with a Gaussian sum filter (GSF) [80]. The GSF aims to correct deviations from the original track after the electrons radiate bremsstrahlung photons when traversing significant amounts of material in the ID. The energy of the calorimeter cluster associated with the ID track, as well as the direction of the ID track at the interaction point are used to compute the electron transverse momentum. Background—mainly arising from pions which mimic the electron signature in the EM calorimeter—is suppressed by relying on the longitudinal and transverse shapes of the calorimeter showers, the matching between the tracks and the calorimeter clusters, and various properties of the ID track; comprising 13 variables in total, they are combined into a multivariate likelihood discriminant [79]. The values of the likelihood discriminant are segmented into three bins which define the *electron working points* [81]: 'Loose', 'Medium', and 'Tight'. They correspond to identifying a *prompt* electron, that is, an electron originating from the primary interaction vertex, with $p_{\text{T}} = 40$ GeV with an efficiency of 93%, 88%, and 80%, respectively. The corresponding fake rates for the 'Loose', 'Medium', and 'Tight' working points are 0.9%, 0.5%, and, 0.3%, respectively [82]. All reconstructed electrons are required to

satisfy (i) $p_{\mathrm{T}} > 4.5$ GeV in order to lie above the EM calorimeter's electronic-noise threshold, and (ii) $\eta < 2.47$ as this is where the 'precision region' of the ID with a fully instrumented pixel detector and SCT ends.

3.2.4 Muon spectrometry

The ATLAS muon spectrometer (MS) is designed to measure the trajectories of muons inside a toroidal magnetic field. The magnetic field is provided by a large barrel toroid for $|\eta| < 1.4$ surrounding the HCAL, whereas for $1.6 < |\eta| < 2.7$ the magnetic field is provided by smaller end-cap toroids inserted into the barrel toroidal magnet [66]. The intermediate region is referred to as the *transition region*, and here magnetic deflection is provided by both the barrel and the end-cap toroidal fields. The MS consists of several subsystems: the monitored drift tubes (MDTs) provide the main track coordinates over most of the η range, while cathode strip chambers are used in the forward region. The muon triggering system is provided by resistive plate chambers (RPC) in the barrel, as well as thin gap chambers (TGC) in the end-cap region. A schematic overview of the ATLAS MS is given presented in Figure 3.13. The total number of readout channels in the MS is 1.08 million.

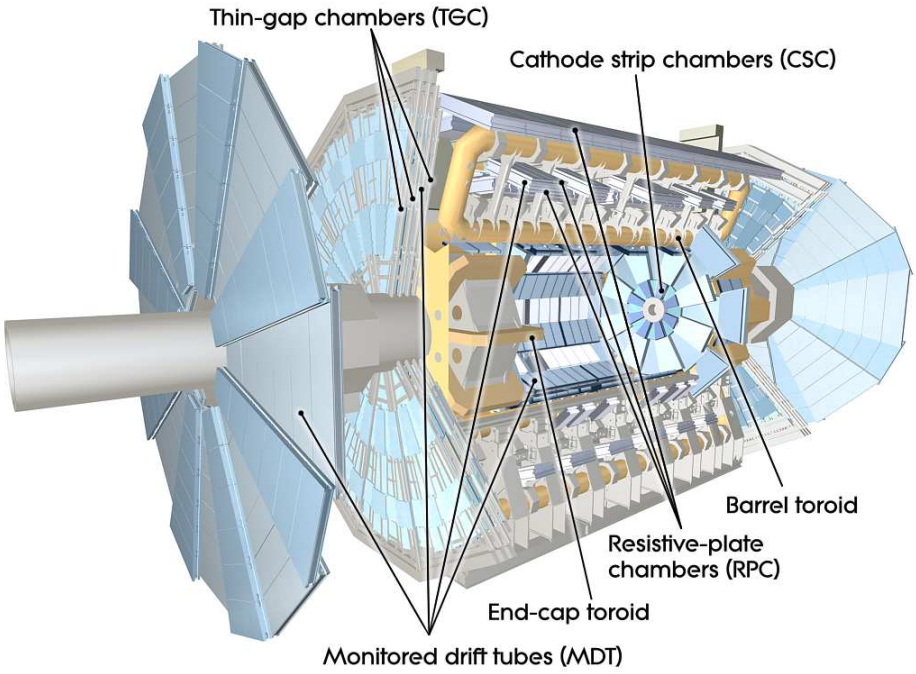


Figure 3.13: Cut-away view of the ATLAS muon spectrometer. Indicated are the different toroidal magnet components as well as the four subsystems of the MS. Figure reproduced from Reference [66].

The geometrical acceptance of the MS generally varies due to the segmentation of the individual chambers as well as constrains imposed by the other detector components. The MS has a region of nearly zero acceptance for $|\eta| < 0.03$ due to a 30-cm gap required for cryogenic pipes and other service lines to the electromagnetic calorimeter. This loss of muons for $|\eta| < 0.03$ translates to an overall loss in muon efficiency over the central barrel region of the detector ($|\eta| < 1$) of roughly 3%, the partial recovery of which is the subject of Chapter 6. There are, in addition, other regions of reduced acceptance due to mechanical structures to support the barrel magnet—the *transition region* of the MS between barrel and end-caps near $|\eta| \approx 1$—as well as 'cracks' between end-cap MDT chambers that do not perfectly overlap. Figure 3.14 shows the MS acceptance as a function of pseudorapidity. More details on the MS and its subsystems can be found in Reference [83].

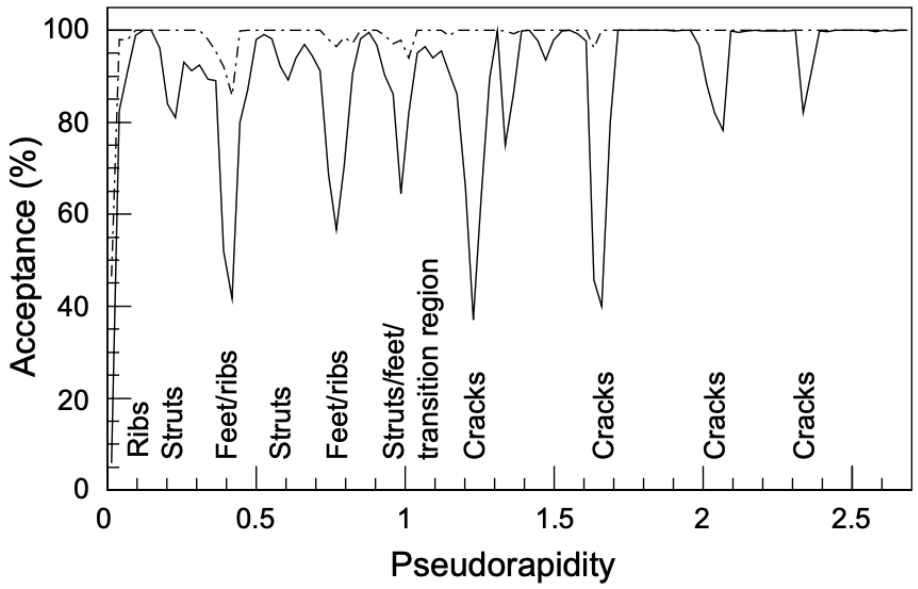


Figure 3.14: Muon spectrometer acceptance as a function of pseudorapidity. The acceptances are shown separately for a requirement of hits occurring in at least three stations (solid line) and a hit occurring in at least one station (dashed line). Figure reproduced from Reference [83].

3.2.5 Muon reconstruction

The ATLAS detector performs muon reconstruction independently in the ID and the MS in order to eventually combine the reconstruction information gathered from different parts of the detector and create muon objects [84]. Every reconstructed muon receives a tag associated with its *type* that indicates which parts of the detector

were involved in reconstructing the muon. ATLAS distinguishes between four different muon types [84] that are reconstructed by algorithms using information available in the ID, the MS, as well as the calorimeters. The four muon types are as follows:

- *Combined muons* have a successful track reconstruction performed independently in the ID and the MS. The two tracks are combined in a global fit. The fitting procedure starts from the outside (the MS), and extrapolates the track towards the ID and tries to match it with a track in the ID.

- *Segment-tagged muons* are muons that have a track in the ID that is extrapolated to the MS, and has at least one track segment in one of two parts of the MS: the monitored drift tube chambers or the cathode strip chambers. This muon type applies when only one layer of the MS is crossed. This could either be because of a low muon p_{T}, or because the muon traverses a region of the MS with low acceptance.

- A muon type called *extrapolated muon* occurs when the muon is reconstructed only in the MS and cannot be matched with an ID track. A requirement is placed on the track in the MS that ensures it is compatible with having originated from the primary interaction vertex. In order to qualify as an extrapolated muon, the track is required to traverse two or more layers of the MS in general and at least three layers in the forward region of the detector. The primary use of extrapolated muons is to extend the muon-reconstruction acceptance into the forward region of the detector that is not covered by the ID ($2.5 < |\eta| < 2.7$). The efficiency of correctly assigning MS tracks to interaction vertices in this pseudorapidity region is $\sim 70\%$ at a muon p_{T} of 10 GeV, increasing to $\sim 90\%$ for $p_{\mathrm{T}} > 30$ GeV [85].

- *Calorimeter-tagged* muons have a track in the ID that can also be matched with an energy deposit in the calorimeter that is found to be consistent with a minimum ionising particle[2]. While calorimeter-tagged muons recover muon-reconstruction acceptance in the regions of the ATLAS detector where the MS reduced acceptance, they have the lowest purity among all muons types. The selection criteria for this muon type are optimised for the low-eta region

[2]A minimum ionising particle, or MIP, is a particle whose energy loss is close to the minimum energy loss. The minimum energy loss occurs when the minimum ionisation energy threshold is reached, that is, when the energy loss is just sufficient to remove the most loosely bound electrons from the atoms in the material traversed.

of $|\eta| < 0.1$ (where the MS has a gap to allow access for cryogenic and other service lines to the ECAL) and a momentum range of $25 \leq p_{\mathrm{T}} \leq 100$ GeV.[3] Only ATLAS physics analyses that admit muons with a 'loose' quality requirement use the calorimeter-tagged muon type.

In cases where a muon can be associated with more than one of the types mentioned above, the muon to be kept is chosen in the following way: If two muons have the same track in the ID, which is possible in the case of combined, segment-tagged, and calorimeter-tagged muons, the combined type is preferred, followed by the segment-tagged type. If a muon of any of these three types overlaps with an extrapolated muon, the muon type with the better-quality fit and larger number of hits in the MS is selected [84].

3.2.6 Triggering and data acquisition

The LHC's bunch-crossing rate of 40 MHz is beyond what any of the ATLAS subsystems can process in real time, and a subset of events that contains potentially interesting phenomena needs to be selected in a process called *triggering*. ATLAS handles triggering using two successive triggers; a hardware-based level-1 (L1) trigger and a software-based high-level trigger (HLT), performing step-wise reductions of the event rate [66]. In the first stage, a subset of the detectors is used to decide whether to keep or discard a given event within 2.5 µs. The subsequent levels reduce the final event rate written to disk to about 1 kHz, and produce events with a size of roughly 1.3 MB [66].

The triggering system works as follows: The L1 trigger looks for signature of particles with high transverse momentum and indications of large missing transverse energy (hinting at the presence of high-p_{T} neutrinos). High-p_{T} muons are identified in the MS's RPCs and TGCs, whereas high-p_{T} electrons, photons, jets, and τ decays are identified based on calorimeter energy deposits that are read out at reduced granularity [86]. A central trigger processor combines the information collected in the calorimeter and MS triggering systems, and passes events that fulfil a set of pre-defined criteria in the so-called *trigger menu* on to the next level. Each interesting feature identified by the L1 trigger is associated with its coordinates in η and ϕ, which can be exploited by higher-level triggers. The coordinates of potentially interesting features are called regions of interest (ROIs) [66].

[3] This description of calorimeter-tagged muons pertains to their definition prior to the work outlined in this thesis. Due to improvements made to the reconstruction method of calorimeter-tagged muons presented in this work in Chapter 6, the definition of this muon type changes.

The HLT uses the ROIs and performs a full read-out of all detector subsystems within the ROIs. The HLT is a software-based trigger that runs in a computing farm consisting of 40,000 CPUs [87], and it runs a series of reconstruction algorithms that are very close to the ones run in *offline* physics analyses. The HLT processing time for a given event is $\mathcal{O}(1)$ s. The trigger thresholds of the HLT are designed to reduce the event rate further to 1 kHz.

The L1-trigger configurations relevant to Chapters 5 and 6 operate at fairly high transverse-momentum thresholds of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. Those that are used to seed the single-muon HLTs used in both chapters are 70% efficient in the barrel region, and have efficiency of 90% in the end-caps, where it is higher as the end-caps do not suffer from the reduced geometrical coverage affecting the barrel (c.f. Section 3.2.4) [88]. The L1 electron triggers that are used to seed the single-electron triggers in Chapter 5 are $90-95\%$ efficient depending on p_{T} and pile-up conditions [89].

3.2.7 Software and simulation

Monte Carlo (MC) simulations are central to the ATLAS experiment, as they provide a way to study the detector response for a wide range of physics processes [90]. They are required for physics analyses in order to simulate SM and BSM processes and obtain an understanding of their expected behaviour, for the development and validation of reconstruction and identification algorithms, and for the evaluation of the performance and degradation of detector components. The full software infrastructure in ATLAS, from MC event generation to reconstruction, is provided by the `Athena` [91] framework. The ATLAS simulation software chain is commonly divided into three components, which are described in more detail below.

3.2.7.1 Event generation

Event generation refers to the production of particles which can be passed on to the detector simulation. Event generators are provisioned within `Athena`, however, they are usually written and maintained by external authors. Each generated event contains all particles that originate from a vertex at the geometric origin, Here, a number modifications to account for specific beam profiles can be made. Typically, particles with a decay length of less than 10 mm (e.g. W or Z bosons) are considered unstable by the generator, and their decays are immediately simulated. Those with a decay length of 10 mm or more (e.g. electrons, muons, and b quarks) are passed on to the detector simulation, and any subsequent decays are handled therein. The

most common event generators are *general-purpose generators* such as `Pythia` [92], `Powheg` [93], `Herwig` [94], and `Sherpa` [95]. They produce events starting from a proton-proton, proton-nucleus, or nucleus-nucleus initial state. They calculate the matrix element of a hard-scattering interaction to lowest order in QCD, to which simulated QED and QCD radiation is added in a so-called *shower approximation*. Short-lived particles are decayed, before, finally, the particles are passed on to the detector simulation.

3.2.7.2 Detector simulation

The simulation of the interaction of the generated events with the ATLAS detector is based on the `Geant4` [96] toolkit. `Geant4` simulates the passage of particles through matter and allows for the implementation of specific geometries—such as that of the ATLAS detector—in great detail, including any misalignments. In the case of ATLAS, the `Geant4` detector geometry—referred to as the `GeoModel`—consists of more than 4.8 million elements. During simulation, `Geant4` records a list of energy deposits in sensitive detector regions, each of which is associated with an exact location in the detector. This list is written to an output file—the *hit file*—and is passed on to the next step in the simulation chain for further processing.

3.2.7.3 Digitisation

The digitisation processes hits from the detector simulation. The overlay of pile-up, the effect of detector noise, and the response of the hardware-based L1 trigger are simulated in this stage. The main goal of the digitisation step is to create *digits*, which serve as input to the simulated detector readout electronics. At a basic level, a *digit* is produced if the voltage or current of a detector readout channel reaches a pre-configured threshold value within a particular time window. The signal of the detector readout electronics is fully emulated. The resulting detector signals are saved in an output format known as *raw data object* (RDO). RDO files are passed on to the reconstruction software, in which objects suitable for physics analyses, such as electrons and muons with full kinematic information, are constructed from lower-level information. At this stage, the reconstruction usually does not distinguish between simulation and collision data.

3.2.8 Computing

Providing the simulation infrastructure to ATLAS requires immense computing resources. Small-scale experimentation and MC production is often conducted on institutional computer clusters, however, large-scale MC production is typically performed on the Worldwie LHC Computing Grid (WLCG) [97]. Shared among many LHC experiments, it consists of over 900,000 CPU cores distributed across more than 170 computing facilities in 42 countries. The data rates processed by the WLCG regularly exceed 80 GB/s.

Chapter 4

Introduction to Machine Learning and its Applications to High Energy Physics

Traditionally, the way particle physics data have been analysed is through human expertise, that is, through careful engineering of algorithms that extract useful information from the data. This approach is becoming more and more impractical as the size of the available data grows and limits the interpretation and understanding of the data to a preconceived set of event variables. Machine learning offers a different approach by enabling the discovery of interesting patterns in data without the need for human intervention [98].

The two main avenues of the physics programme at the LHC aim to search for new physics beyond the Standard Model as well as to make precision measurements of Standard Model processes. Both of these avenues benefit from higher luminosities, increasing the probability of the production of rare events while at the same time increasing the amount of pile-up to be expected in every bunch crossing. Consequently, the expected amount of data to emerge from the LHC experiments in the coming decades is unprecedented. Running at increased luminosity in Run 3 and beyond, the experimental particle physics domain of data analysis will face compounded challenges of interpreting vast amounts of data in increasingly complex environments. The amount of information that can potentially be extracted from larger, more complex datasets would be significantly limited using traditional methods without the aid of machine learning algorithms.

While the high-energy physics community has used neural networks for many decades, recent advances in computing power, coupled with new methods for

cheaply performing backpropagation on GPUs[1], have led to the widespread application of deep neural architectures. The primary forms of machine learning that have been used at the LHC are deep neural networks and boosted decision trees. In last decade or so, newer methods such as variational autoencoders and generative adversarial networks have become more widely adopted. These techniques and others have been applied to a wide range of problems in particle physics, a few of which are highlighted here:

1. *Event classification.* One of the most common problems in which machine learning is applied is in the classification of events into one out of two or more classes. A typical challenge that machine learning tries to solve is identifying a rare signal event among background events that occur orders of magnitude more frequently. Here, the algorithm is typically trained on a simulated set of signal and background events in the form of a *supervised learning* problem. Here, class 1 are the signal events, and class 2 are the background events. Machine-learning techniques for event classification have become widely adopted in particle physics. Two notable examples are the D0 and CDF Collaborations, which used BDTs and neural networks to classify events as signal- or background-like in the observation of the electroweak production of a single top quark [100, 101]. More recently, the CMS Collaboration used machine learning in the observation of the Higgs boson in 2012 [5], in which a boosted decision tree was adopted to discriminate Higgs-like signals in the diphoton channel from Standard Model background. Another related application is one that we will encounter later on in this work in Chapters 5 and 6: machine learning algorithms are adept at identifying the production mechanism of, in this case, a heavy Higgs-like scalar (a task that before the introduction of machine learning to this problem had been achieved through a simultaneous cut on two high-level event variables), as well as by discriminating muons in the calorimeter from other particles.

2. *Event generation.* Full physics event simulations typically simulate the passage of particles through matter using a software package called `Geant4` [96]. The most computationally expensive part of the simulation chain is the showering

[1]The work presented in Chapter 6 of this thesis benefited from the availability of GPUs, allowing for the exploration of more complex models as well as larger search spaces in the selection of optimal training parameters, which would have been infeasible on CPUs. The author is grateful for and would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility [99] in carrying out this work.

process, and in an ATLAS-like calorimeter, this step can take several minutes per event. *Generative adversarial networks* (GANs) are a promising way to generate physics events more cheaply than conventional methods. In a project called CaloGAN [102], GANs are trained to provide fast simulations of realistic particle showers in the calorimeter. GANs comprise two components: The first, called the *generator*, tries to generate realistic samples from initially random noise in order to fool the second part of the GAN, the so-called the *discriminator*. The discriminator tries to distinguish fully simulated events (from `Geant4`) from those synthetically produced by the generator. It calculates a score corresponding to the probability that the seen example is real or fake, i.e. generated. The generator, in turn, uses this score as feedback on its ability to generate convincing events in order to further improve it. Initial studies have shown that such an approach could deliver a factor in speed-up of 10^5 compared to `Geant4`-simulated showers [102]. Reducing the overall computing load through the speed-up of event generation would free up significant resources, as currently it is estimated that more than 50% of computing resources at the LHC are spent on Monte Carlo event generation [103].

3. *Particle physics theory.* Machine learning has broad areas of application in particle physics theory, from determining physical parameters by constraining effective field theories, to model-independent searches, to fitting parton distribution functions. One proof-of-concept approach to performing model-independent searches using machine learning considers the difference between two datasets: the simulated Standard Model events, which constitute the background, as well as recorded collision data, including a new physics signature and Standard Model events, as an unsupervised learning problem, without hypothesising a specific model in which to interpret the two datasets [104]. The approach compares the two datasets in a high-dimensional space using a nearest-neighbour search. It is non-parametric in the sense that it does not make any assumptions as to the underlying probability distributions that generated the datasets. It retains model independence by not requiring a physical model to determine whether the two datasets differ in some way. In a case study that looked for new physics in two samples, one containing Standard Model events, and the other containing Standard Model events and a small fraction of injected monojet signals, the model-independent search strongly disfavoured the background-only, i.e. Standard Model, hypothesis [104].

4. *Anomaly detection.* The quality of data-taking during an LHC run is usually monitored by physicists who study the output histograms of many diagnostic variables and compare them to reference versions in order to spot potential deviations, or anomalies. Many such anomalies, however, manifest themselves not in a large deviation in one diagnostic variable alone, but in a large number of small deviations in many variables, making them very hard to detect if they have not occurred before. A machine learning method called anomaly detection seems promising in its potential to improve this situation. It uses a class of neural networks called *autoencoders*, which have a twofold structure: The first part of the network, called the encoder, produces a representation of the inputs in a greatly reduced number of dimensions (this is called the latent space[2]), which is then restored by the second component, the decoder, which tries to recreate as much of the inputs as possible from the latent-space representation. Anomalous data appear much more rarely than regular data, such that when the autoencoder detects an anomalous example, its output differs greatly from its input, as the latent-space representation of the input does not constitute a meaningful compression of the inputs. In particle physics, autoencoders can be especially useful when coupled with a call to action upon seeing an anomalous example, such as alerting a physicist or shutting down and restarting the part of the detector that likely is responsible for the anomaly through malfunctioning [98, 105].

Machine learning powers many parts of society with examples including medicine, image recognition and processing, natural-language understanding, processing and generation, recommendation systems, fraud detection, autonomous driving, and virtual assistants. Particle physics is not exempt from this trend. The application of machine learning to particle physics has enabled advances in the development of new algorithms that extract interesting features from data, an improved understanding of physical theory, and a reduction in the required computational resources to process LHC-scale datasets. This chapter provides an introduction to the theoretical basics of machine learning, and outlines several fundamental principles of some of the machine learning algorithms used presently in particle physics. Special attention will be given to deep neural networks in general, as well as convolutional neural networks in particular, as these two methods will turn out to be of importance

[2] *The latent space* is a high-dimensional vector space spanned by *latent variables*. Latent variables are variables that are not directly observable, but that are sufficient to describe the data, as inferred by an algorithm.

in this work through their application in the two following chapters: a search for high-mass 4ℓ resonances (Chapter 5), and the identification of muons in the ATLAS calorimeters (Chapter 6). Finally, practical aspects relevant to machine learning as applied in this work related to training and optimising machine learning algorithms are discussed.

4.1 Machine learning basics

Machine learning is defined as a set of methods used to detect patterns in data and further use these detected patterns to make predictions about unseen data [106].

Approaches to machine learning can be broadly classified into *supervised* and *unsupervised* machine learning. The main difference between the two approaches is that in the case of supervised learning, prior knowledge as to what the true labels of the output data are is present, whereas in the unsupervised case this knowledge is not available. This chapter focuses on the supervised-learning case, as this approach is employed in Chapters 5 and 6.

4.1.1 Supervised learning

Given a set of N labelled input-output pairs, called the *training set*,

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}, \tag{4.1}$$

the goal of supervised machine learning is to learn a mapping g from the *inputs* $\boldsymbol{x}$ to the *output* y

$$g : \boldsymbol{x} \to y. \tag{4.2}$$

Each member of the training set x_i is a D-dimensional vector (typically of numbers). Each dimension of the input space corresponds to a *feature* or *variable*, which in the case of particle physics is frequently associated with an event-level observable such as the p_{T} or η of a particle. The elements of $\boldsymbol{x}$ may in principle be objects with a more complex structure such as an image (represented as a matrix containing pixel values), a sentence (represented using a word embedding), or a graph (represented as a set of vertices connected by edges).

The outputs y_i can have various forms, but in the two most common types of machine learning problems, they are either *categorical* variables or real scalars. When y_i is categorical, it can take one of any values of a finite set of *classes* or *categories*, namely

$$y \in \{1, \ldots, C\}, \tag{4.3}$$

where C is some positive integer. A typical example of such classes in particle physics comes from the task of *binary classification* with two class labels for *signal*

(class label 1) and background (class label 0). When the outputs y_i can take on values corresponding to one of more than two categories, the problem is called *multiclass* or *multi-label classification*[3]. One formal definition for the problem of supervised learning is that there exists some function f such that

$$y = f(x). \tag{4.4}$$

Moreover, the goal of the learning algorithm is to estimate the function f given a training set $\mathcal{D}$ [106]. The approximation to f obtained using machine learning is called a *classifier* or a *model* and is denoted by $\hat{f}$. $\hat{f}$ can be used to make predictions, i.e.

$$\hat{y} = \hat{f}(\boldsymbol{x}). \tag{4.5}$$

When the elements of y are real scalars, the problem is called *regression*, in which case the task is to predict a numerical value. The learning task is the same as that of classification, except that y can take on continuous numerical values, i.e.

$$y_i \in \mathbb{R}. \tag{4.6}$$

An example from particle physics is the reconstruction of a particle's energy based on the calorimeter energy deposits.

4.1.2 Generalisation and training

The challenge for a machine learning model $\hat{f}$ is to perform well on unseen data, that is, when making predictions (called *inference*) on inputs that do not belong to the training set and have therefore not been seen before. This ability is called *generalisation* [107]. When evaluating the performance of a model under consideration, it is usually evaluated on a *test set*, that is, a set of labelled input-output pairs that do not belong to $\mathcal{D}$.

[3]One example of multi-label classification in particle physics is the problem of discriminating two different types of signal events, e.g. a new particle produced through one of two considered production mechanisms, a and b, from a single type of background, e.g. Standard Model events. In this case, one would assign the class label 0 to the background, class label 1 to the signal produced through mechanism a, and class label 2 to the signal produced through mechanism b.

Obtaining a model $\hat{f}$ is achieved through *training*, which denotes the process of using a training set in order to reduce some measure of error between the predicted outputs $\hat{y}$ and the true outputs y, called the *training error* [107]. The crucial difference between traditional optimisation and machine learning is that in machine learning, there is an additional goal to reduce the *test error*, that is, the expected error on an unseen input.

A machine learning algorithm samples the training set and minimises the *training error* using an optimisation algorithm, which modifies its internal parameters. When the resulting model is applied to the test set, it is expected that the test error is equal to or greater than the training error. A model is said to suffer from *underfitting* if it does not achieve a sufficiently low training or test error, whereas *overfitting* denotes the condition where the training error is much lower than the test error. There is no general prescription for diagnosing over- or underfitting. Machine-learning practitioners frequently study the evolution of the training and errors as a function of training progress. Typically, if the test error flattens off while the training error continues to decrease, the model has entered the overfitting regime. Conversely, if neither the training nor the test error have stopped decreasing, the model is in the underfitting regime.

4.1.3 Gradient-based optimisation

In order to be able to evaluate a machine learning model, both on seen and unseen inputs, a measure of performance is needed. Typically, it quantifies the difference between the predicted output and the true output. A fundamental measure of performance is the *mean squared error* of the training set.

Machine learning algorithms typically involve optimisation, which is defined as minimising some *cost function*, $h(\boldsymbol{\theta})$, by changing the *parameters*, $\boldsymbol{\theta}$. This function is also called an *objective function*, *loss function*, or *error function*. Minimising $h(\boldsymbol{\theta})$ is typically achieved through *gradient-based optimisation*, which relies on the gradient $\boldsymbol{\nabla} h(\boldsymbol{\theta})$ in order to find $\boldsymbol{\theta}$ such that h is minimal.

Suppose that the cost function at time step t is given by $h(\boldsymbol{\theta}_t)$. Given the first-order Taylor series of h about $\boldsymbol{\theta}_t$,

$$h(\boldsymbol{\theta}_t + \Delta\boldsymbol{\theta}) \approx h(\boldsymbol{\theta}_t) + \boldsymbol{\nabla} h|_{\boldsymbol{\theta}_t} \, \Delta\boldsymbol{\theta}^T, \tag{4.7}$$

we know that for small changes to the parameters given by $\Delta\boldsymbol{\theta}$, h can be reduced by moving along the negative gradient [108]. This method is called *gradient descent* [109], and it chooses $\Delta\boldsymbol{\theta}$ such that it is parallel to $-\boldsymbol{\nabla} h|_{\boldsymbol{\theta}_t}$. At the subsequent time step $t+1$, the parameters are then given by

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha \boldsymbol{\nabla} h|_{\boldsymbol{\theta}_t}, \tag{4.8}$$

where α is the *learning rate*, which is a parameter that needs to be chosen. Since the first-order Taylor approximation in Equation (4.7) only holds for small $\Delta\boldsymbol{\theta}$, α in general needs to be small, and choosing a learning rate that is too large can lead to a situation where a local minimum in h is missed. On the other hand, smaller learning rates lead to longer training times; finding the right value for α is often a delicate matter in practice.

4.2 Boosted decision trees

A class of machine-learning classifiers commonly used in particle physics is called *boosted decision trees* (BDTs)[4]. BDTs consist of an ensemble of decision trees, all of which are generally weak classifiers, and which are combined into a single robust classifier through a process called *boosting*.

4.2.1 Decision tree classifiers

A decision tree works as follows: Suppose a dataset consisting of N features contains two classes, a and b. At the beginning of the training process, a decision tree finds the *root node*, which is a decision boundary[5] on the single most discriminating feature. A training example can either pass or fail this cut condition, creating two *leaves* in the tree. The training examples that pass this requirement are split further according to the second-most discriminating feature, and similarly for those that fail the condition of the root node. A decision tree iteratively chooses the features that most effectively split the input data into the two classes through this method. The values at which to split the data for a given feature are chosen so as to minimise the loss function. Figure 4.1 illustrates the basic structure of a decision tree. The

[4] In the machine learning literature, they are more commonly referred to as *gradient-boosting trees*.

[5] *Decision boundary* refers to the hyperplane that separates the underlying vector space into two sets. In this section's example, the decision boundary separates the N-dimensional feature space into two sets corresponding to classes a and b.

splitting continues until a stopping condition is met; this happens when a tree reaches a maximum number of splittings, i.e. a maximum *depth*[6], or when further splitting does not lead to a notable decrease of the loss function.

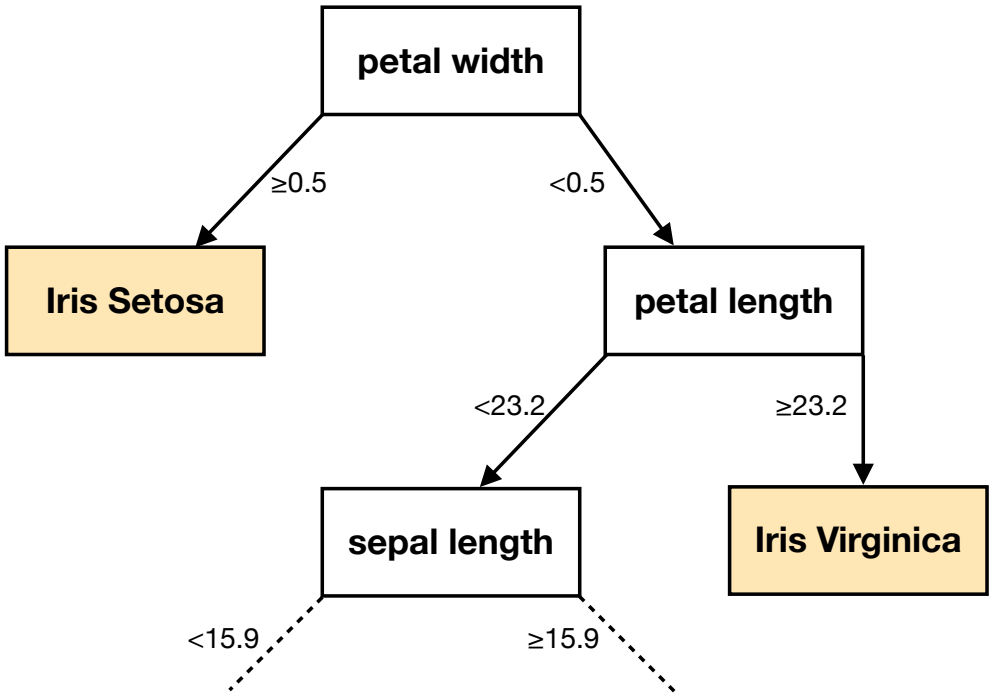


Figure 4.1: Example structure of a decision tree. The root node 'petal width' is the most discriminating feature in this dataset. If a sample has a 'petal width' value of at least 0.5, it is classified as an 'Iris Setosa'. If examples fail this cut, they are discriminated further according to the remaining features, of which 'petal length' and 'sepal length' are shown. N.B.: The variable names and classes are inspired by the Iris flower classification dataset [110].

4.2.2 Boosting

Decision tree classifiers such as the one described above are prone to overfitting, that is, they do not show robust performance when presented with unseen samples. In order to overcome this limitation, decision tree classifiers are frequently combined into an *ensemble* of decision trees, creating a strong classifier out of many weak ones. This ensemble is referred to as a BDT.

BDTs are created by training many individual decision tree classifiers sequentially. Let T be the decision tree classifier trained at iteration t. Each training example

[6]The depth of a decision tree classifier is a parameter set by the human practitioner, often motivated by a constraint on computational resources. It may be the case that the loss function has not stopped decreasing when the depth condition is reached.

receives a weight that increases when it has been misclassified by a given tree, and decreases when correctly classified. In this way, misclassified examples are assigned greater importance in subsequent trees, increasing the probability of being classified correctly. Furthermore, each tree T is assigned a weight, w_t, that is proportional to the overall *accuracy*[7] achieved by T on the training set. The output of the decision tree ensemble is therefore given by

$$\hat{y}(x) = \sum_t w_t h_t(x), \tag{4.9}$$

where $h_t(x)$ is the output of tree T. The scheme follows the AdaBoost algorithm [111], although many other variations of gradient-boosted decision trees exist (LightGBM [112], XGBoost [113], and random forests [114] to name but a few). In particle physics, a popular implementation of a boosted decision tree is provided by the `TMVA` package [115].

4.3 Deep learning

While traditional machine learning systems are confined to processing data that humans structure according to a set of engineered variables, modern methods have moved away from this constraint by employing *representation learning*. Machine learning methods are said to use representation learning if they can be provided with unstructured raw data and discover representations of that data relevant to detecting interesting patterns in an automated way [116]. *Deep learning* methods employ representation learning in that they form internal representations of the data in a way that is not determined by a human. Typically, machine learning algorithms based on deep learning form multiple representations of the input data by combining the representation at one level with a nonlinear transformation to obtain a representation at a subsequent level. The representation of the input data becomes more abstract with each layer, and the aim here is to represent arbitrarily complex functions through the combination of a sufficient number of abstract representations of the input data. In classification problems, each level of abstraction in representation-learning algorithms emphasises aspects of the input data that are more relevant to the discrimination problem while discounting parts

[7] Accuracy is a measure of how well the decision tree identifies examples correctly. In the case of binary classification, it is given by $\text{Acc} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$, where TP denotes number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives.

less relevant [116]. By combining more processing layers and increasing the number of nodes in each layer, deep learning models can represent functions of increasing complexity [107]. Crucially in the case of deep learning, the abstract representations are not human-engineered but are obtained by the algorithm itself.

4.3.1 Deep learning architectures

More specifically, deep learning is defined as a branch of machine learning concerned with models based on *artificial neural networks*, which aim to mimic aspects of the structure found in the human brain. Deep learning has proven to be very successful at supervised-learning tasks with a large number of training examples, benefiting from an increase in computing power that allows networks that are composed of many processing layers to be trained efficiently.

4.3.1.1 Deep feedforward neural networks

The most common class of deep neural networks are *deep feedforward neural networks* (DNNs)[8], which are also known as *multi-layer perceptrons* (MLPs). DNNs are artificial neural networks comprised of many layers that aim to approximate a function f given a set of inputs $\boldsymbol{x}$. During the training process, deep neural networks change their internal parameters, $\boldsymbol{\theta}$, until they reach a satisfactory approximation

$$\hat{y} = \hat{f}(\boldsymbol{x}; \boldsymbol{\theta}). \tag{4.10}$$

The term *feedforward* comes from the observation that information in these networks, starting from the inputs $\boldsymbol{x}$, flows forward and traverses each layer of the network until reaching the output y. The internal parameters, $\boldsymbol{\theta}$, are *weights*, $\boldsymbol{W}$, and *biases*, $\boldsymbol{c}$, such that the output of the network at layer n given inputs $\boldsymbol{x}$ is [107]

$$f_{\mathrm{n}}(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{c}) = g(\boldsymbol{W}^T \boldsymbol{x} + \boldsymbol{c}). \tag{4.11}$$

Feedforward neural networks are based on the principle that the output of the first layer, f_{input}, can serve as input to a subsequent layer after being multiplied by a nonlinear *activation function*[9], g. This process happens iteratively until the final layer

[8] Frequently, the terms *deep neural network* and *deep feedforward neural network* are used synonymously, and the same convention is used throughout this work.

[9] The activation function of a layer, consisting of one or more *neurons*, in a deep feedforward neural network defines the output of the layer given an input. In a sense, it determines how "active" the neuron is.

is reached.

The structure of deep neural networks is analogous to a directed acyclic graph, which describes how the functions are chained [107]. For instance, in a deep neural network with two *hidden layers*[10], the output of the first layer f_{input} becomes the input to the first hidden layer f_1, whose output is the input to the second hidden layer f_2. Finally, the output of the second hidden layer becomes the input of the output layer f_{output}, meaning given inputs $\boldsymbol{x}$, this deep neural network with two hidden layers performs the operation

$$f(\boldsymbol{x}) = f_{\text{output}}(f_2(f_1(f_{\text{input}}(\boldsymbol{x})))). \tag{4.12}$$

Such a network is illustrated in Figure 4.2.

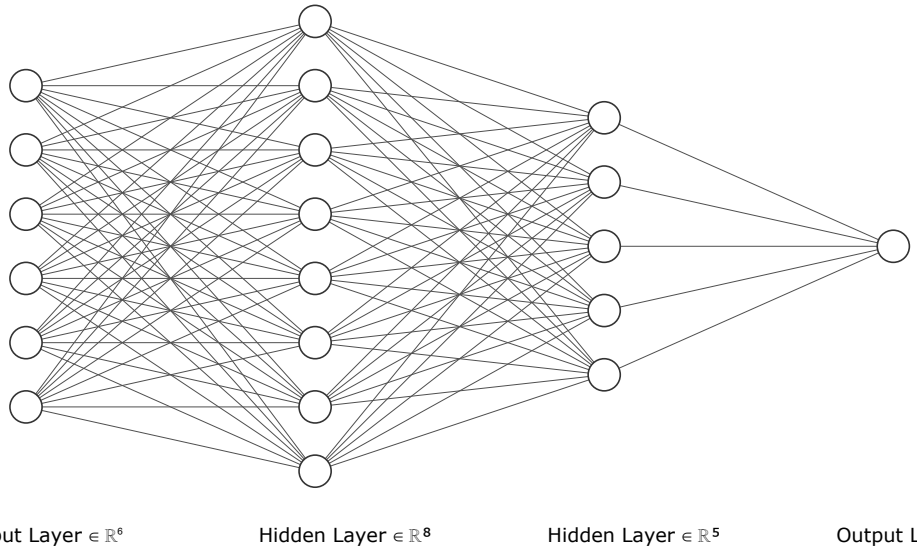


Figure 4.2: Example of a deep neural networks with two hidden layers. The input layer receives a six-dimensional input vector, which is fed forward through two hidden layers to reach the output layer. Such a network could correspond to a binary classification task, where the output takes values between 0 and 1. Illustration created using the `NN-SVG` software package [117].

4.3.1.2 Nonlinearities in deep neural networks

For neural networks to be able to approximate real-world problems, which are usually nonlinear, the activation functions must be nonlinear. In practice, one

[10]Layers are considered *hidden* if they are located anywhere between the input and the output layer of a deep neural network.

often distinguishes between the activation functions at the hidden layers and the activation function at the final, or output, layer.

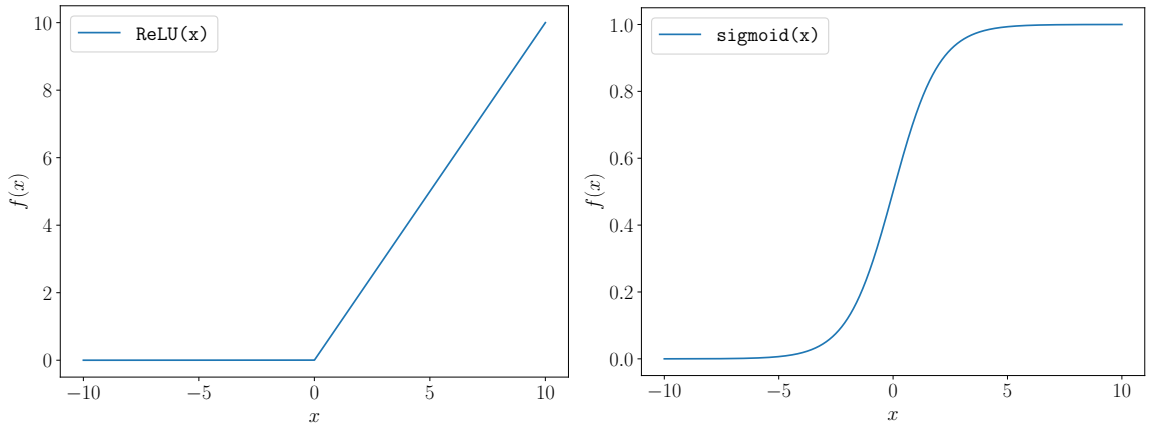
One of the most common activation functions used for the hidden layers in DNNs is the rectified linear unit, or *ReLU*. It is defined as the positive part of its argument, namely

$$\mathrm{ReLU}(x) = \max(0, x). \tag{4.13}$$

In most cases, the type of problem one is trying to solve determines the activation function at the final layer. In the case of binary classification, for instance, the choice of output-layer activation function must meet the constraint that the output be mapped to a binary class label. Usually, this is achieved by choosing an activation at the final layer that can take values in the interval $[0, 1]$, rounding down values < 0.5 to the class label `0`, and rounding up values ≥ 0.5 to the class label `1`. The activation function used for the output layers throughout this work is the *sigmoid* function, which is defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{4.14}$$

The two activation functions are illustrated in Figure 4.3.



(a) ReLU activation function.

(b) Sigmoid activation function.

Figure 4.3: ReLU (a) and sigmoid (b) activation functions, as functions of their input, x. It is clear that the range of the ReLU function is $[0, \infty]$, while the range of the sigmoid function is $[0, 1]$. This makes the latter a suitable choice for the output-layer activation in binary-classification tasks.

4.3.1.3 Backpropagation

Central to finding a local optimum in the loss function of a DNN is the ability to calculate the gradient in a given network layer in order to perform gradient-based optimisation of the kind encountered in Section 4.1.3. This is done using an algorithm called *backpropagation*, which is based on applying the chain rule to the weights in an artificial neural network successively from the last layer back to the first one (hence the prefix 'back'). The backpropagation algorithm was first discovered in 1974 [118], but it took at least a decade until its applicability to neural networks was fully appreciated [119, 120, 121]. Computing the gradients as a function of the weights allows one to use gradient-based optimisation in order to adjust, or *tune*, the weights towards lower values of the overall cost function each time the error from a *batch* of training examples propagates back through the network.

What follows is a brief introduction to the backpropagation algorithm for DNNs, which is an implementation of the general method of gradient-based optimisation introduced in Section 4.1.3. Consider the n^{th} layer of a feedforward neural network with a total of N layers with activation function f_n, inputs x_n, and weights w_n. The output of layer n is given by [122]

$$y_n = f_n(x_n), \tag{4.15}$$

and the inputs to the subsequent layer, x_{n+1}, are just

$$x_{n+1} = w_n y_n = w_n f_n(x_n). \tag{4.16}$$

Suppose $\mathcal{L}$ denotes the loss function of the network. We define the gradient of the loss function with respect to the input vector at layer n as

$$\delta_n = \frac{\partial \mathcal{L}}{\partial x_n}. \tag{4.17}$$

Similarly, the gradient at the next layer $n+1$ is given by

$$\delta_{n+1} = \frac{\partial \mathcal{L}}{\partial x_{n+1}}. \tag{4.18}$$

Rewriting Equation (4.17) using the chain rule, we have

$$\begin{aligned}\delta_n &= \frac{\partial \mathcal{L}}{\partial x_n} \\ &= \frac{\partial \mathcal{L}}{\partial x_{n+1}} \frac{\partial x_{n+1}}{\partial x_n} \\ &= \delta_{n+1} \frac{\partial x_{n+1}}{\partial x_n}.\end{aligned} \tag{4.19}$$

Using Equation (4.16), this can be written as

$$\begin{aligned}\delta_n &= \delta_{n+1} \frac{\partial (w_n y_n)}{\partial x_n} \\ &= \delta_{n+1} \frac{\partial (w_n y_n)}{\partial y_n} \frac{\partial y_n}{\partial x_n} \\ &= \delta_{n+1} w_n \frac{\partial y_n}{\partial x_n}.\end{aligned} \tag{4.20}$$

Using the definition of y_n from Equation (4.15), and defining $\frac{\partial f_n(x_n)}{\partial x_n} = f_n'(x_n)$, this becomes

$$\begin{aligned}\delta_n &= \delta_{n+1} w_n \frac{\partial f_n(x_n)}{\partial x_n} \\ &= \delta_{n+1} w_n f_n'(x_n).\end{aligned} \tag{4.21}$$

Lastly, we find the partial derivative of the loss function with respect to w_n

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_n} &= \frac{\partial \mathcal{L}}{\partial x_{n+1}} \frac{\partial x_{n+1}}{\partial w_n} \\ &= \delta_{n+1} \frac{\partial x_{n+1}}{\partial w_n} \\ &= \delta_{n+1} \frac{\partial (w_n f_n(x_n))}{\partial w_n} \\ &= \delta_{n+1} y_n,\end{aligned} \tag{4.22}$$

where we have made use of Equations (4.16) and (4.18).

Keeping the overall aim in mind of reducing the loss function through an iterative process, Equation (4.22) provides a recipe for updating the weights in the network. In each training iteration, the change in weights at layer n, Δw_n, should be such as

to reduce $\mathcal{L}$:

$$\begin{aligned}\Delta w_n &= -\alpha \frac{\partial \mathcal{L}}{\partial w_n} \\ &= -\alpha \delta_{n+1} y_n,\end{aligned} \tag{4.23}$$

where α is the learning rate from Equation (4.8).

4.3.1.4 Loss functions

The backpropagation algorithm only works for loss functions, $\mathcal{L}$, that are differentiable. There are a great number of loss functions to choose from, and here only a brief overview is given of one type of loss function that has found application in this work.

A common choice of loss function for binary classification tasks is *binary cross-entropy*, which is defined as

$$\mathcal{L}_{\text{BCE}} = -\sum_{c=0}^{1} y^c_{\text{true}} \log y^c_{\text{pred}}. \tag{4.24}$$

Here, the sum over c runs over the two possible class labels (0 and 1). y^c_{true} denotes the true label of the example of class c (this is either 0 or 1). y^c_{pred} is the output of the classifier for the training example being considered; it can take on continuous values in the interval $[0, 1]$.

4.3.1.5 Regularisation

Ensuring that an algorithm does not only perform well on the training set, but also on new inputs, is a central problem in machine learning. The set of techniques concerned with achieving sufficiently low test errors is collectively referred to as *regularisation* [107]. One of the most common regularisation methods for DNNs is *dropout* [123]. It refers to randomly removing, or *dropping out*, neurons—along with their connections—during training of the network with a certain probability that is usually set as a training parameter by the practitioner. Applying dropout to a network effectively leads to sampling from a large number of thinned networks, which has the effect of preventing neurons from co-adapting[11] too much. At test time, one uses the full network, i.e. without dropped neurons, which is equivalent

[11] *Co-adaptation* refers to a configuration of multiple neurons which produce very similar outputs for some given input data.

to averaging the predictions of all thinned networks that were sampled during training [123]. Overall, dropout leads to a significant reduction in overfitting, and is used in both applications of deep learning presented in this thesis in Chapters 5 and 6.

4.3.2 Convolutional neural networks

Convolutional neural networks (CNNs) are a class of deep neural networks designed to process information that can be represented in a grid-like way. Most frequently, CNNs find application in the area of computer vision, which is concerned with processing digital images or videos. In the case of (moving) images, the CNN inputs are two-dimensional pixel grids [107]. Convolutional neural networks perform *convolutions* on their inputs instead of ordinary multiplications as in the case of feedforward neural networks.

The convolution operation involves convolving a grid-like input (such as an image) with a *kernel matrix* (frequently referred to as *kernel* or *filter*). During training, the weights of the convolutional kernels are iteratively adjusted until training finishes. As with all internal neural network parameters, the weights are initially assigned random values. The kernels are sensitive to local spatial features in the input images and preserve translational invariance. This means that translating spatial features in the input images by a small amount does not change the internal representation of the image. This is especially useful if one tries to capture whether certain features are present in the input images or not, as opposed to where features are located with respect to one another.

4.3.2.1 Convolution operation

Suppose we have an input image represented as a 2D matrix of size (m, n). The convolution operation requires a kernel, which is a two-dimensional matrix whose size (m', n') is smaller than that of the input matrix. In the convolution operation, subsections of the input image (called the *image patch*) are multiplied with the convolutional kernel. Each such multiplication produces one real number, which forms one element of the output matrix (called the *feature map*).

Figure 4.4 illustrates the convolution operation on a 3×3 input image. Here, a submatrix of size 2×2, the image patch, is convolved with the convolutional kernel of size 2×2. The resulting output matrix, or feature map, is of size 2×2. The multiplication operation denoted by $*$ is the sum of all matrix elements of

the element-wise multiplication of the image patch (red shaded area) with the convolutional kernel. The element-wise product of matrices A and B is defined as

$$(A \circ B)_{ij} = A_{ij} B_{ij}. \tag{4.25}$$

Taking the sum of the resulting matrix (also referred to as the *grand sum*) results in a scalar. The illustrated multiplication is repeated for all submatrices of the input image, with the result of each multiplication forming one element of the feature map. The process as a whole is referred to as the convolution operation, and it differs strongly from the usual mathematical meaning of the term.

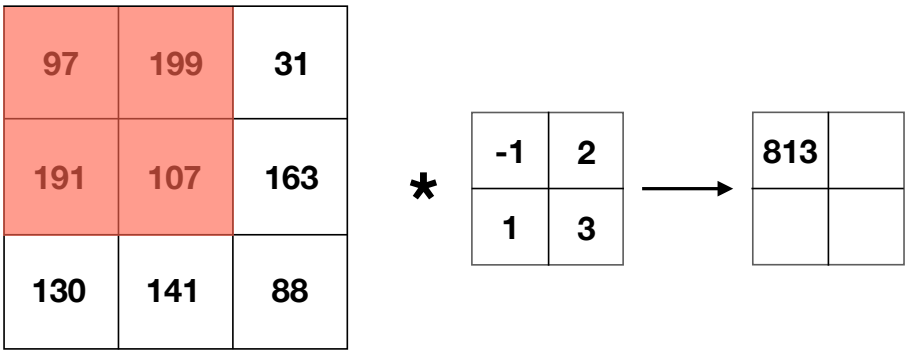


Figure 4.4: Example of a convolution operation on a 3×3 input matrix using a 2×2 convolutional kernel. The red shaded 2×2 image patch is multiplied element-wise with the convolutional kernel of size 2×2. The symbol $*$ denotes the summation of all matrix elements of the result of the element-wise multiplication. The resulting 2×2 matrix is called the *feature map*.

The visualisation in Figure 4.4 pertains only to a single colour channel. In practice, one usually encounters at least three channels (such as for RGB images), in which case the convolution operation is performed separately in each of the colour channels, and the resulting three output matrices are added element-wise to form a single feature map. Furthermore, one typically chooses more than one convolutional kernel in order to create sensitivity to multiple spatial features in the input image. The number of kernels is a tunable hyperparameter of the network, and there is no general guideline as to the correct number of kernels to choose.

At this stage of the flow of an input image through a convolutional neural network, we have only applied linear activations in the sense that the element-wise multiplication is of linear nature. In order to retain nonlinearity in the overall network, the feature maps are run through a nonlinear function, such as the ReLU activation function encountered in Section 4.3.1.2.

4.3.2.2 Pooling

In order to reduce the size of intermediate, or hidden, representations of the input image, an operation called *pooling* is commonly applied. There exist various forms of pooling, but its usual purpose is to reduce the size of the image by replacing parts of the image with summary statistics. Pooling can be interpreted as adding the assumption that the function learnt in a given layer must be translationally invariant [107]. The two most common forms of pooling are *max pooling*, in which the summary statistic is based on the maximum element within a submatrix, as well as *average pooling*, in which the summary statistic is the average value of a submatrix.

The max-pooling operation is illustrated in Figure 4.5, where the pooling filter chooses maximum elements within submatrices of size 2×2. The pooling filter is moved across the input image until the entire matrix has been traversed. The average-pooling operation works in a similar fashion (the average value of a given submatrix is computed instead of the maximum value). Max pooling can help make a convolutional neural network translationally invariant by choosing only maximum values in a given neighbourhood as opposed to the exact location of features [107].

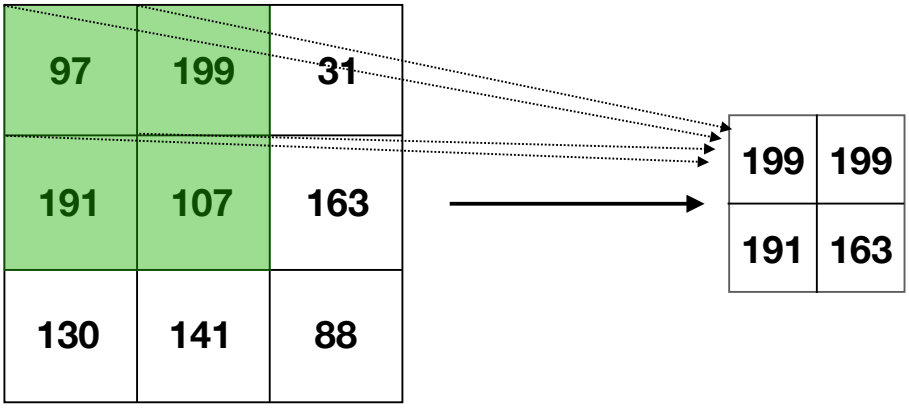


Figure 4.5: Diagram showing the max-pooling operation on an example 3×3 input matrix. The submatrix (green shaded area) contains a maximum value of 199, which becomes the entry at position $(0, 0)$ in the output matrix.

4.3.3 Recurrent neural networks

Traditional neural networks, both feedforward networks and convolutional neural networks alike, cannot incorporate any notion of *sequence*. That is, information in different parts of sequential input data is not available when the network is applied to other parts of the input. *Recurrent neural networks* (RNNs) overcome this

shortcoming by learning representations of sequential inputs [124]. At the core of an RNN lies a *cell*, which iteratively handles a specific section of the input to compute a new output. The next section of the input is handled in the subsequent iteration, and a new output is computed based on the previous output state and the new inputs. This process is repeated until the entire input sequence is exhausted. In this way, an RNN creates a fixed-size representation for an input sequence of arbitrary length. This fixed-length representation is then usually passed into a feedforward neural network [124]. The fundamental operating principle of an RNN cell is illustrated in Figure 4.6. In the case of particle physics, RNNs find application where the input data follow sequential patterns that can be exploited, such as in b-jet tagging, where the presence of one track with a large impact parameter means a second track with a large impact parameter is more likely to occur [124].

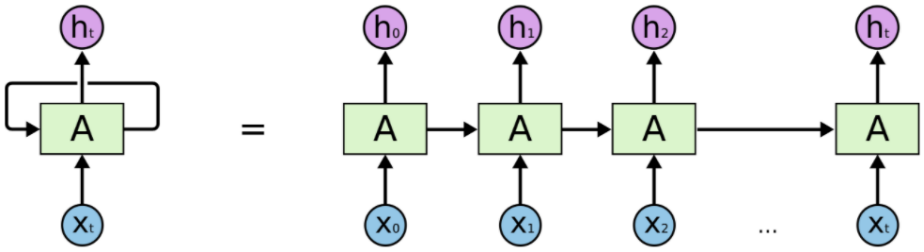


Figure 4.6: Diagram of an unrolled recurrent neural network. The RNN cell on the left handles an input sequence at timestep t, labelled x_t. The cell applies an activation A to x_t by handling all subsequences $(x_0, x_1, \ldots, x_t)$ until the input sequence is exhausted. Figure reproduced from Reference [125].

4.4 Hyperparameters and validation

Machine-learning models have settings that can be adjusted and that are separate from the internal parameters. The settings are called *hyperparameters* and are adjusted either by hand or in an automated process called *hyperparameter optimisation*. Hyperparameters are those classifier parameters that should not be learnt automatically, as doing so would likely result in overfitting to the training set or, in rarer cases, because the hyperparameter is too difficult to learn through training. Most of the hyperparameters chosen not to be learnt are responsible for the overall complexity of the model [107].

A data-driven approach to finding an appropriate set of hyperparameters is to reserve a specific subset of the training data for *validation*. Typically, 20% of the training set are reserved for this purpose. The validation set is not observed by

the training algorithm but is used to evaluate the generalisation capacity during or after training and to tune the classifier's hyperparameters. Note that this is different from the test set, which is used to make statements about the classifier's capacity to generalise once the training is finalised. Under no circumstances is the test set used to make decisions about a classifier's internal or hyperparameters [107].

4.4.1 k-fold cross-validation

In cases where one has access to only a relatively small amount of data, the observed generalisation capacity on the test set can suffer from great statistical uncertainty. One solution to this problem is called *k-fold cross-validation*. It involves splitting the entire dataset (training and test examples) into K *folds* of equal size. Let k denote the index of a fold such that $k \in \{1, \ldots, K\}$. The training process is repeated K times, and each time the k^{th} partition of the data is reserved for testing, while the remaining data are used for training. In this way, each training example is used once for testing and $K - 1$ times for training. Let m be the size of each fold. The cross-validation estimate of the prediction error of a classifier $\hat{f}$ is given by [126, 127]

$$\text{CV}(\hat{f}) = \frac{1}{K} \sum_{k=1}^{K} E_k(\hat{f}). \tag{4.26}$$

Here, $E_k(\hat{f})$ is the prediction error of the k^{th} classifier, tested on the k^{th} fold, namely [126]

$$E_k(\hat{f}) = \frac{1}{m} \sum_{i \in k^{\text{th}} \text{ fold}} (y^i_{\text{true}} - y^i_{\text{pred}})^2, \tag{4.27}$$

where the sum runs over all examples of the k^{th} fold, y^i_{true} denotes the true label of example i, and y^i_{pred} is the output score of the classifier on example i. $E_k(\hat{f})$ estimates the model's expected error on an unseen sample and can therefore be interpreted as a measure of the model's *bias*. It is common to repeat the k-fold cross-validation multiple times in order to obtain an estimate of the model's *variance*, which is the standard deviation squared of many $E_k(\hat{f})$ estimates.

Chapter 5

Search for Heavy ZZ Resonances in the $\ell^+\ell^-\ell'^+\ell'^-$ Final State

With the discovery of a new particle in 2012 consistent with the Higgs boson predicted by the Standard Model, the CMS and ATLAS collaborations provided experimental evidence for and contributed to the understanding of the mechanism of electroweak symmetry breaking. Many studies published since [128, 129, 130, 131] indicate that the particle is consistent with the SM Higgs boson. A number of models [132, 133, 134] attempt to solve the problems of hierarchy and naturalness, in which the SM Higgs boson is part of an extended Higgs sector.

This chapter describes a search for heavy resonances decaying to two Z bosons, subsequently decaying to four charged leptons, referred to as the $\ell^+\ell^-\ell'^+\ell'^-$ final state, where $\ell = e, \mu$. The results of this search are combined with those of a separately conducted analysis, which looks for heavy resonances decaying via ZZ to two charged leptons and two neutrinos ($\ell^+\ell^-\nu\bar{\nu}$, where $\ell = e, \mu$ and $\nu = \nu_e, \nu_\mu, \nu_\tau$). The two neutrinos cannot be reconstructed directly, and their presence must be inferred from missing transverse energy. The focus of this chapter is on the $\ell^+\ell^-\ell'^+\ell'^-$ decay channel; however, overall results are presented using a combination of both final states. The search uses proton-proton collisions recorded at a centre-of-mass energy of $\sqrt{s} = 13$ TeV collected with the ATLAS detector during Run 2 of the LHC (2015–2018), with a total integrated luminosity of $\mathcal{L}_{\text{int}} = 139$ fb^{-1}.

This chapter focuses on the development and implementation of a multivariate classification method aimed at improving the sensitivity in each of the two analysis channels probing different production modes of scalar resonances. It is structured as follows: a brief overview of the analysis is given in Section 5.1, followed by a description of the recorded dataset and simulation samples used in Section 5.2. The

reconstruction of physics objects and events is outlined in Section 5.3, followed by details on the selection of events in Section 5.4. The background estimation and modelling of signal and background are outlined in Sections 5.5 and 5.6. A review of a cut-based approach to the categorisation of events is given in Section 5.7, followed by a description and performance analysis of the novel multivariate event classifiers in Section 5.8. Sections 5.9 and 5.10 provide the detailed statistical interpretation of the search and its results. Finally, concluding remarks are made in Section 5.11.

5.1 Analysis overview

The search in the $\ell^+\ell^-\ell'^+\ell'^-$ channel aims to look for excesses in the distribution of the $\ell^+\ell^-\ell'^+\ell'^-$ invariant mass, $m_{4\ell}$, in the range $200 < m_{4\ell} < 2000$ GeV. In the $\ell^+\ell^-\nu\bar{\nu}$ channel, the final state cannot be fully reconstructed, and the transverse mass, m_T, is used as a discriminant. It is defined as

$$m_T = \sqrt{\left[\sqrt{m_Z^2 + (p_{\mathrm{T}}^{\ell\ell})^2} + \sqrt{m_Z^2 + (E_T^{\mathrm{miss}})^2}\right]^2 - |\vec{p}_{\mathrm{T}}^{\,\ell\ell} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|^2}, \tag{5.1}$$

where m_Z is the mass of the Z boson, $\vec{p}_{\mathrm{T}}^{\,\ell\ell}$ is the transverse momentum of the charged-lepton pair with magnitude $p_{\mathrm{T}}^{\ell\ell}$, and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is the missing transverse energy with magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$.

The search considers the following resonance hypotheses:

- A heavy Higgs-like scalar with mass m_H in the *narrow-width approximation* (NWA), that is, a heavy Higgs boson whose natural width is the same as that of the SM Higgs boson ($\Gamma_H = 4.07$ MeV [135]).
- A heavy Higgs-like scalar with mass m_H in the *large-width assumption* (LWA), that is, a heavy Higgs boson with a non-negligible natural width. Different natural widths are considered: 1, 5, 10, and 15% of the Higgs boson mass, m_H.
- A spin-2 resonance in the Randall-Sundrum (RS) model [62, 136], which gives rise to a Kaluza-Klein (KK) excitation with mass $m_{G_{kk}}$ (Section 2.2.2). The Randall-Sundrum (RS) model aims to solve the hierarchy problem by introducing extra dimensions in which all SM fields can propagate. Their ability to propagate causes the SM fields to create a tower of KK excitations, including excitations of the gravitational field that appear as TeV-scale spin-2

gravitons [137]. The key parameter in the RS model is $k/\bar{M}_{\text{Planck}}$, where k defines the curvature scale of the extra dimension, and $\bar{M}_{\text{Planck}} = M_{\text{Planck}}/\sqrt{8\pi}$ is the reduced Planck mass. It is assumed to have a value of 1 in this work. Typically, gravitons in the RS model have non-negligible couplings to all SM fields, which means they do not propagate into the extra dimensions. Here, a simplified model is considered, in which the SM fields are allowed to propagate into the extra dimensions (the *bulk*). In this scenario, the couplings of the graviton to light fermions are suppressed, leading to significantly reduced production rates from vector-boson fusion and lower branching fractions to leptons and photons. This leads to a dominance of the ggF production mode. $k/\bar{M}_{\text{Planck}}$ is the only free parameter in this simplified 'bulk RS' model, and its value of 1 is typical.

In the absence of an excess of events in either the $m_{4\ell}$ or the m_T spectrum, upper limits on the production cross-section times branching fraction of a heavy resonance, $\sigma \times B(H \to ZZ)$, are set. The scalar search assumes that the heavy Higgs (hereafter referred to as H) is predominantly produced through the vector-boson fusion (VBF) process as well as the gluon-gluon fusion (ggF) process, but that, unlike in the case of the SM Higgs boson, the ratio between these two processes is unknown in the absence of a specific model. Consequently, limits on the production cross-section times branching fraction are set separately in the two event categories, i.e. $\sigma_{\text{VBF}} \times B(H \to ZZ)$ and $\sigma_{\text{ggF}} \times B(H \to ZZ)$. The upper limits on the production cross-section times branching fraction of a narrow resonance are translated into exclusion contours within the context of a 2HDM encountered in Section 2.2.1. Limits on the production cross-section times branching fraction of a spin-2 graviton are used to constrain the RS model.

The previous iteration of this analysis [58], which combined results in the $\ell^+\ell^-\ell'^+\ell'^-$ final with those in the $\ell^+\ell^-\nu\bar{\nu}$ final state, found two excesses in 36 fb^{-1} of data collected in 2015–16: at 240 GeV and 700 GeV, each with a local significance of 3.6 σ. The excess at 240 GeV was seen predominantly in the $4e$ channel, while that at 700 GeV was observed in all 4ℓ channels. The 700 GeV excess seen in the $4e$ final state is excluded at 95% confidence level by the $\ell^+\ell^-\nu\bar{\nu}$ channel due to its higher sensitivity in this mass range. The analysis was repeated in 2018 using 80 fb^{-1} of data collected in 2015–17 [138] in order to obtain an early indication as to the persistence of the two excesses. Both excesses disappeared nearly completely when only considering the data collected in 2017, bringing down the local significance of the excess at 240 GeV to 2.5 σ, and that of the excess at 700 GeV to 3.0 σ when

considering the 2015–17 dataset. The excesses come exclusively from the 2015–16 component of the dataset.

The search presented here aims to further confirm or reject the previously observed excesses by including data collected in 2018, using a total integrated luminosity of 139 fb^{-1}. Apart from switching to new ATLAS recommendations on the version of the analysis software and the jet-reconstruction algorithm [86], this search implements improvements leading to increased sensitivity in the $\ell^+\ell^-\ell'^+\ell'^-$ final state by switching from a cut-based definition of the ggF and VBF event categories to a multivariate one.

5.2 Data and simulation

5.2.1 Data

The data used in this search were collected by the ATLAS detector between 2015 and 2018 at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV with a bunch spacing of 25 ns. The recorded data are subject to quality requirements, and events are rejected if parts of the detector are not fully operational. The average data-taking efficiency across the whole of Run 2 of the LHC is 95.76% [139]. The total amount of usable data for this analysis amounts to (139.0 ± 2.4) fb^{-1}. Data were reconstructed using `Athena Release 21.2` [91].

5.2.2 Simulation

Simulated events are used widely in this analysis. They form the basis of the modelling of background events expected from the SM, the calculation of acceptances in all final states and event categories, the optimisation of the VBF and ggF event categories (including the training of a multivariate algorithm), as well as the estimation of systematic uncertainties. The samples used were produced in three official ATLAS Monte Carlo (MC) simulation campaigns: 'mc16a', which mimics data-taking conditions in 2015–16, 'mc16d', which accounts for the 2017 data-taking period, and 'mc16e', which covers data-taking during 2018. The major differences between these three simulation campaigns are the pile-up conditions of the data-taking periods they are designed to mimic.

All events were processed by the ATLAS simulation framework, which operates in the context of `Geant4` [96]. Collisions that are not part of the primary interaction vertex are referred to as *pile-up*, and all events are reweighted in such a way as to

match the observed distribution of the mean number of interactions per bunch crossing (a process referred to as *pile-up reweighting*).

What follows is a description of the generation of high-mass signal and SM background events using the ATLAS simulation framework [90]. After the generation of events, data and simulation are treated on an equal footing, that is, they are processed in the same way by using the same trigger settings and reconstruction algorithms.

5.2.2.1 Simulation of signal

Signal samples in the NWA were produced using the `POWHEG-BOX v2` [140] event generator with next-to-leading order (NLO) accuracy in QCD, which was interfaced with `PYTHIA8` [141] for the simulation of parton showering and hadronisation, as well as decaying the Higgs boson into the two final states under consideration. `EvtGen v1.2.0` [142] was used to simulate b and c hadron decays, while the hard-scattering process used the leading-order (LO) `CT10` PDF set [143]. NWA signals were generated separately in the ggF and VBF production mode at discrete mass points between 200 and 2000 GeV, in steps of 100 GeV up to a mass of 1000 GeV, and in steps of 200 GeV thereafter. In addition, the 2400 and 3000 GeV mass points were generated in order to validate the extrapolation of the signal-shape parametrisation to higher masses. The NWA signals were mainly generated using the full detector simulation. Two of the samples were generated using the fast detector simulation (300 and 700 GeV in the VBF production mode in the 'mc16e' campaign) [90], in which the response of the electromagnetic and hadronic calorimeters uses a parametrisation, while the response of the ID and the MS is fully simulated.

LWA signal samples in the ggF production mode were generated at NLO accuracy in QCD using `MADGRAPH5_MC@NLO v2.3.2` [144], interfaced with `PYTHIA8` for parton showering, hadronisation, and Higgs boson decays into the two final states. As in the case of the NWA samples, b and c hadron decays were generated using `EvtGen v1.2.0`. LWA signal samples were generated with a natural width of $\Gamma_H = 0.15\, m_H$ between 400 and 2000 GeV, with a step size of 100 GeV up to 1000 GeV, and a step size of 200 GeV thereafter. In addition, LWA signals were generated with natural widths of $\Gamma_H = 0.05\, m_H$ and $\Gamma_H = 0.1\, m_H$ at $m_H = 900$ GeV in order to validate the signal parametrisation of the shape of the $m_{4\ell}$ distribution in the $\ell^+\ell^-\ell'^+\ell'^-$ channel.

Spin-2 gravitons within the RS framework were generated using `MADGRAPH5` interfaced with `PYTHIA8`. The same method was used to generate the b and c hadron

decays as for the NWA and LWA samples. The graviton samples were generated with $k/\bar{M}_{\text{Planck}} = 1$.

The mass points of the NWA and LWA samples, the widths of the LWA samples, and the $k/\bar{M}_{\text{Planck}}$ parameter of the graviton samples were chosen for compatibility with the previous iteration of the analysis using 36 fb^{-1} of data.

5.2.2.2 Simulation of background

The $\ell^+\ell^-\ell'^+\ell'^-$ analysis considers SM background from diboson (ZZ), triboson (VVV), Z + jets, $t\bar{t}$, and ttV processes. In the generation of all backgrounds, `EvtGen v1.2.0` is used to simulate b and c hadronisation.

The SM $q\bar{q} \to ZZ$ background was generated using `SHERPA v2.2.2` [95], as well as the `NNPDF3.0` PDF set at NLO [145] for the hard-scattering process. The accuracy in QCD was NLO for events with up to one jet, and LO for events with two or three jets. Parton showers were simulated using `SHERPA`'s built-in showering algorithm [146], and were merged using `MePs` [147] at NLO.

The SM $gg \to ZZ$ process, including the production of a SM Higgs boson and the interference between the two processes, were modelled using `SHERPA v2.2.2` in combination with `OPENLOOPS` [148]. The calculation of matrix elements for zero and one jet was performed at LO, and combined with `SHERPA` parton showers. The electroweak $q\bar{q} \to ZZ$ process, $q\bar{q} \to ZZ$ (EW), containing two additional jets produced via vector-boson scattering was generated using `SHERPA v2.2.2`. The PDF set used was `NNPDF3.0` at NLO. Events containing a single Z boson with associated jets (Z + jets) were generated using `SHERPA v2.2.2`. The calculation of matrix elements for up to two partons was performed at NLO, and up to four partons at LO, using the `COMIX` [149] and `OPENLOOPS` matrix-element generators. These were merged with `SHERPA` parton showers using `MePs` at NLO. The SM backgrounds arising from VVV processes were generated using `SHERPA v2.2.1`, while `MADGRAPH5_MC@NLO` was used for the $t\bar{t}$ + Z background with four prompt leptons. The $t\bar{t}$ background was generated using `POWHEG-BOX v2` interfaced with `PYTHIA v6.428` [92], using the `PERUGIA 2012` set [150] of tuned parameters for parton showering and hadronisation. The generator was further interfaced with `PHOTOS` [151] for the application of radiative corrections from QED and with `TAUOLA` [152, 153] for the simulation of τ decays.

5.3 Object and event reconstruction

The reconstruction of electrons (Section 3.2.3) follows a new approach based on dynamical topological cell clustering in order to improve the electron energy measurement. Details on this approach are available in Reference [154]. The use of this approach is new with respect to the previous 36 fb^{-1} analysis [58]. The resulting electrons have a p_{T} of at least 4.5 GeV and lie within $|\eta| < 2.47$. The working point for the electron selection is 'loose', yielding an efficiency of more than 90% for electrons with a p_{T} of at least 30 GeV.

The reconstruction of muons is based on the ID, the calorimeters, and the MS. All four muon types used in ATLAS are used in the $\ell^+\ell^-\ell'^+\ell'^-$ final state. They correspond to the combined, calorimeter-tagged, segment-tagged, and extrapolated muon types introduced in Section 3.2.5. Segment-tagged muons are restricted to the central barrel region ($|\eta < 0.1|$). Calorimeter-tagged muons are accepted in the same η region if the ID track satisfies $p_{\mathrm{T}}^{\mathrm{track}} > 15$ GeV. Extrapolated muons are included in the forward region defined by $2.5 < |\eta| < 2.7$ if they have hits in three traversed MS stations. As is the case for electrons, the muon identification in this round of the analysis implements several improvements. The improvements are related to the extrapolation of ID tracks to the MS. The overall muon identification working point is 'loose', which has an efficiency of 98.5% in the $\ell^+\ell^-\ell'^+\ell'^-$ final state.

Another improvement with respect to the 36 fb^{-1} analysis [58] comes from a new jet-reconstruction approach based on the particle-flow algorithm [86]. The particle flow objects are used as input to the anti-k_t algorithm [155] implemented in the `FASTJET` package [156]. Once jets are reconstructed, several corrections are applied in order to calibrate the energy scale, as described in Reference [157]. All jets in the $\ell^+\ell^-\ell'^+\ell'^-$ final state must satisfy $p_{\mathrm{T}} > 30$ GeV as well as $|\eta| < 4.5$. The contamination from pile-up jets is reduced by placing additional requirements on the tracks and vertices for jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ [158].

Four-lepton events are required to have at least one vertex with two associated ID tracks, each satisfying $p_{\mathrm{T}}^{\mathrm{track}} > 500$ MeV. The primary vertex is chosen to be the one with the largest $\sum(p_{\mathrm{T}}^{\mathrm{track}})^2$. Potential overlap between leptons and jets is resolved in the following way: If electrons and muons share the same ID track, the muon is preferred if it is a combined or an extrapolated muon; in all other cases, the electron is preferred. If jets overlap with electrons, the jets are removed within a cone of $\Delta R = 0.2$ around the electron. For muons, the cone within which overlapping jets are removed has a size of $\Delta R = 0.1$.

5.4 Event selection

The $\ell^+\ell^-\ell'^+\ell'^-$ final state has very good sensitivity to heavy resonances due to its excellent mass resolution and small experimental and theoretical uncertainties. Four-lepton events are selected and classified into one of three channels according to their flavour, i.e. $4e$, 4μ, or $2e2\mu$. The selection is based on single-lepton, di-lepton and tri-lepton triggers. The di- and tri-lepton triggers include triggers for $e\mu$ events. Single-electron triggers need to satisfy a 'medium' or 'tight' likelihood requirement, while di- and tri-electron triggers must satisfy a 'loose' or 'medium' requirement. The minimum trigger thresholds vary in p_{T} due to an increase in peak luminosity over the four data-taking years. In the case of single muons, they vary between 20 and 26 GeV, and in the case of electrons, minimum trigger thresholds vary between 24 and 26 GeV. Overall, the trigger efficiency for events passing the selection requirements is 98%.

In each of the flavour channels, charged-lepton quadruplets are formed by finding two pairs of same-flavour, opposite-sign leptons. Each selected electron must satisfy $p_{\mathrm{T}} > 7$ GeV and be measured within a pseudorapidity range of $|\eta| < 2.47$. Each selected muon must satisfy $p_{\mathrm{T}} > 5$ GeV and $|\eta| < 2.7$. There are additional requirements on the leptons' transverse momenta: The lepton with the highest p_{T} must satisfy $p_{\mathrm{T}} > 20$ GeV, the one with the second-highest p_{T} must satisfy $p_{\mathrm{T}} > 15$ GeV, while the one with the third-highest p_{T} must satisfy $p_{\mathrm{T}} > 10$ GeV. If two or four muons are present in the quadruplet, at most one of them is allowed to be calorimeter-tagged, segment-tagged, or an extrapolated muon in the range $2.5 < |\eta| < 2.7$ (see Section 3.2.5 for more details).

In events with ambiguity as to which charged leptons constitute the quadruplet coming from the ZZ decay, the quadruplet is kept with the highest dilepton , m_{12}, closest to the Z boson mass, and the second-highest dilepton invariant mass, m_{34}, second-closest to the Z boson mass. The first charged-lepton pair is the *leading pair*, while the second one is referred to as the *subleading pair*. Constraints are placed on the proximity of the dilepton pairs' invariant masses to the Z boson mass: The leading pair with mass m_{12} needs to satisfy $50 < m_{12} < 106$ GeV, while the subleading pair with mass m_{34} has to satisfy $50 < m_{34} < 115$ GeV. All charged leptons in the quadruplet need to be separated by $\Delta R > 0.1$. In order to suppress background from J/Ψ decays, $4e$ and 4μ events that have a charged-lepton pair with $m_{\ell\ell} < 5$ GeV are discarded. If, at this stage, events contain multiple quadruplets from the different flavour channels, the quadruplet in the channel with the highest expected signal

yield is kept, in the order $4\mu, 2e2\mu, 4e$. The fraction of events for which this occurs is expected to be negligible. The inclusion of the quadruplet-ambiguity resolution is a remnant of several related SM $H \to 4\ell$ analyses such as [159], where events with more than four leptons can occur due to SM $t\bar{t}H$ and VH production.

Background from Z + jets and $t\bar{t}$ events is suppressed by placing requirements on the impact parameter as well as on track- and calorimeter-based lepton-isolation variables. Background from heavy-flavour hadrons is reduced by placing a requirement on the transverse impact parameter significance, d_0/σ_{d_0}, where d_0 is the impact parameter, and σ_{d_0} is its uncertainty. It is required to be less than 3 for muons, and less than 5 for electrons. The lepton isolation requirements are as follows: a track-isolation discriminant is computed based on all tracks with $p_{\mathrm{T}}^{\mathrm{track}} > 0.5$ GeV found close to the electron or muon within $\Delta R < 0.3$. The tracks must either originate from the primary vertex, or otherwise have a longitudinal impact parameter, z_0, that satisfies $z_0 \sin\theta < 3$ mm. If the lepton has a p_{T} greater than 33 GeV, the cone size reduces to 0.2, falling linearly from 0.3 to 0.2 between a p_{T} of 33 and 50 GeV, and remaining at a constant value of 0.2 thereafter. A calorimeter-based isolation variable is computed based on topological calorimeter clusters with positive energy (to suppress electronic noise) that are not associated with the track of a muon or an electron, within $\Delta R < 0.2$. The requirement is such that the sum of the track isolation and 40% of the calorimeter isolation must amount to less than 16% of the lepton p_{T}. This requirement is referred to as 'FixedCutPFlowLoose'. Further measures are taken to reduce the contamination from Z + jets and $t\bar{t}$ background: The algorithm responsible for fitting the quadruplet of four charged leptons under the requirement that they originate from a common vertex needs to fulfil a requirement on the goodness-of-fit of $\chi^2/N_{\mathrm{dof}} < 6$ in the case of the 4μ channel, and $\chi^2/N_{\mathrm{dof}} < 9$ in the case of all other flavour channels. Here, N_{dof} stands for the number of degrees of freedom in the quadruplet fit.

Photons from final-state radiation (FSR) are included in the reconstruction of the Z bosons, due to their being identifiable in the electromagnetic calorimeter as well as being well-modelled by simulation. The incorporation of FSR photons works by searching for a single FSR photon per event, which is consistent with having been radiated from one of the charged leptons in the leading pair. The four-momentum of this photon is then added to that lepton pair. Events with more than one FSR photon are vetoed. Following the FSR correction, the four-momenta of both dilepton pairs are recomputed using a kinematic fit that imposes a Z-mass constraint following the method described in [160].

An overview of the event selection in the $\ell^+\ell^-\ell'^+\ell'^-$ final state is provided in Table 5.1.

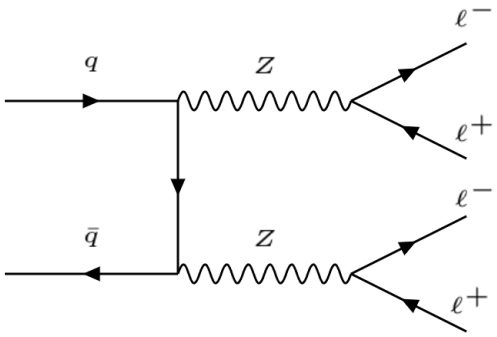
Physics Objects	
Electrons	
Loose Likelihood quality electrons with hit in innermost layer, $\eta > 7$ GeV and $\|\eta\| < 2.47$	
Interaction point constraint: $\|z_0 \cdot \sin\theta\| < 0.5$ mm (if ID track is available)	
Muons	
Loose identification with $p_T > 5$ GeV and $\|\eta\| < 2.7$	
Calo-tagged muons with $p_T > 15$ GeV and $\|\eta\| < 0.1$, segment-tagged muons with $\|\eta\| < 0.1$	
Stand-alone and silicon-associated forward restricted to the $2.5 < \|\eta\| < 2.7$ region	
Combined, stand-alone (with ID hits if available) and segment-tagged muons with $p_T > 5$ GeV	
Interaction point constraint: $\|d_0\| < 1$ mm and $\|z_0 \cdot \sin\theta\| < 0.5$ mm (if ID track is available)	
Jets	
anti-k_T jets with *bad-loose* identification, $p_T > 30$ GeV and $\|\eta\| < 4.5$	
Overlap removal	
Jets within $\Delta R < 0.2$ of an electron or $\Delta R < 0.1$ of a muon are removed	
Vertex	
At least one collision vertex with two or more associated tracks	
Primary vertex	
Vertex with the largest p_T^2 sum	
Event Selection	
Quadruplet Selection	- Require at least one quadruplet of leptons consisting of two pairs of same-flavour opposite-charge leptons fulfilling the following requirements: - p_T thresholds for three leading leptons in the quadruplet: 20, 15 and 10 GeV - Maximum one calo-tagged or stand-alone muon or silicon-associated forward per quadruplet - Leading di-lepton mass requirement: $50 < m_{12} < 106$ GeV - Sub-leading di-lepton mass requirement: $m_{\text{threshold}} < m_{34} < 115$ GeV - $\Delta R(\ell, \ell') > 0.10$ for all leptons in the quadruplet - Remove quadruplet if alternative same-flavour opposite-charge di-lepton gives $m_{\ell\ell} < 5$ GeV - Keep all quadruplets passing the above selection
Isolation	- Contribution from the other leptons of the quadruplet is subtracted - FixedCutPFlowLoose WP for all leptons
Impact Parameter Significance	- Apply impact parameter significance cut to all leptons of the quadruplet - For electrons: $d_0/\sigma_{d_0} < 5$ - For muons: $d_0/\sigma_{d_0} < 3$
Best Quadruplet	- If more than one quadruplet has been selected, choose the quadruplet with highest Higgs decay ME according to channel: 4μ, $2e2\mu$, $2\mu2e$ and $4e$
Vertex Selection	- Require a common vertex for the leptons: - χ^2/ndof < 5 for 4μ and < 9 for others decay channels

Table 5.1: Summary of the object and event selection requirements. The masses of the leading and subleading leptons pairs are denoted by m_{12} and m_{34}, respectively. Table reproduced from Reference [1].

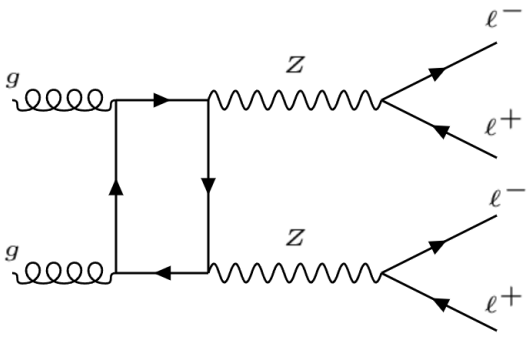
5.5 Background estimation

The main background in the $\ell^+\ell^-\ell'^+\ell'^-$ final state, accounting for 97% of expected events, comes from non-resonant ZZ production. The ZZ continuum background

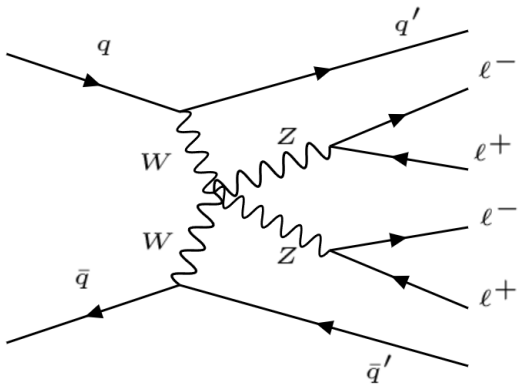
is irreducible[1], and, in the ggF category, 89% of it arises from quark-antiquark annihilation (referred to as $q\bar{q} \to ZZ$), 10% from gluon-initiated production (referred to as $gg \to ZZ$), and 1% from electroweak (EW) vector-boson scattering produced in association with two jets (denoted by $q\bar{q} \to ZZ$ (EW)). The balance shifts slightly in the VBF category, where the overall contribution of the $q\bar{q} \to ZZ$ (EW) backgrounds to all backgrounds is expected to be 16-20% depending on the categorisation method used (Section 5.8). Feynman diagrams of the three leading SM background processes are depicted in Figure 5.1.



(a) Feynman diagram of the process $q\bar{q} \to ZZ \to 4\ell$.



(b) Feynman diagram of the process $gg \to ZZ \to 4\ell$.



(c) Feynman diagram of the process $q\bar{q} \to ZZ$ (EW) with four charged leptons and two associated jets in the final state.

Figure 5.1: Feynman diagrams of the leading SM backgrounds with four charged leptons in the final state: (a) $q\bar{q} \to ZZ \to 4\ell$, (b) $gg \to ZZ \to 4\ell$, as well as (c) the electroweak production of two Z bosons via vector-boson scattering decaying to four charged leptons in association with two jets (referred to as $q\bar{q} \to ZZ$ (EW)).

[1]A background is *irreducible* if its final-state particles are indistinguishable from those expected from the decay of the signal. Conversely, a *reducible* background is one whose final-state particles are in principle distinguishable from the final-state particles of the signal decay.

The ZZ background is modelled using MC simulation combined with a partly data-driven approach: The shape in $m_{4\ell}$ of the ZZ background is modelled from MC simulation, while the overall background normalisation is taken from data. This helps reduce the total uncertainty in the SM ZZ yield by constraining the theoretical uncertainties associated with the modelling of this background. The shape-modelling procedure using MC simulation is outlined in Section 5.6.2. Another source of background comes from Z + jets and $t\bar{t}$ processes. Contributing at the per-cent level, they are reducible and are estimated using data-driven methods where possible, following those outlined in Reference [161]. Minor contributions are expected from triboson (VVV) and $t\bar{t}+V$ events (here, V stands either for a $W^{\pm}$ or a Z), which are estimated using MC simulation.

5.6 Signal and background modelling

The discriminating variable in the $\ell^+\ell^-\ell'^+\ell'^-$ is the four-lepton invariant mass, $m_{4\ell}$. In order to apply the statistical interpretation to arbitrary resonance masses, m_H, the $m_{4\ell}$ distributions of signal and background are parametrised as a function of the signal mass hypothesis. The shape of the $m_{4\ell}$ distribution is estimated from MC simulation, both for the expected SM background and for the high-mass signal.

5.6.1 Signal modelling

The signal modelling follows different approaches depending on the type of resonance. The signal modelling methods for the narrow- and large-width scalar hypotheses and the graviton signal hypothesis are briefly described below.

5.6.1.1 NWA scalar

High-mass signals in the NWA are modelled by fitting the $m_{4\ell}$ distribution obtained from MC simulation to an empirical function. As the natural width of the resonance is negligible compared to the detector resolution, the $m_{4\ell}$ shape is modelled using the sum of a Crystal Ball [162, 163] and a Gaussian function to model the detector resolution. The function is given by

$$P_s(m_{4\ell}) = f_C \cdot C(m_{4\ell}; \mu, \sigma_C, \alpha_C, n_C) + (1 - f_C) \cdot \mathcal{N}(m_{4\ell}; \mu, \sigma_{\mathcal{N}}), \tag{5.2}$$

where C is the Crystal Ball function centred at a peak, which is denoted by μ with standard deviation σ_C. The parameters α_C and n_C determine the shape and position

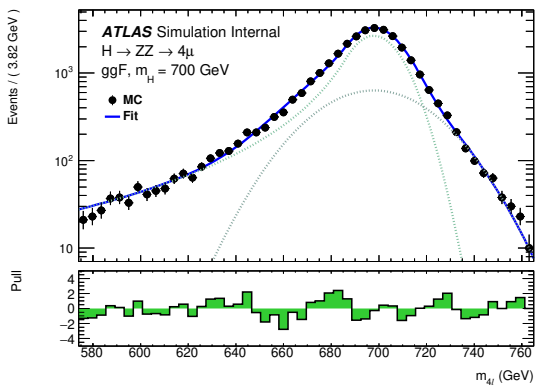
of the tail of the Crystal Ball function. $\mathcal{N}$ is the normal distribution with the same mean μ and standard deviation $\sigma_{\mathcal{N}}$. The parameter f_C is required to ensure the relative normalisation of the probability density functions.

The fitting procedure is performed separately in the ggF and VBF categories, yielding a set of fit parameters at every available mass point. A polynomial is fitted to these parameters, allowing for an interpolation of fit parameters along continuous values of $m_{4\ell}$ for the entire search range ($200 < m_{4\ell} < 2000$ GeV). The degree of the polynomial is established by beginning with a polynomial of degree three, and decreasing the degree until the coefficient of the highest-order term is greater than its standard error. This is done in order to avoid overfitting. The polynomial fitting procedure is conducted separately for all six fitted parameters in Equation (5.2), i.e. $f_C, \mu, \sigma_C, \alpha_C, n_C$, and $\sigma_{\mathcal{N}}$. As a result, the final polynomial degrees are generally different for each fitted parameter, and include polynomials of first, second, and third degree. The bias in signal yield associated with this fitting procedure is 0.8% in the 4μ channel, 0.9% in the $4e$ channel, and 2% in the $2e2\mu$ channel, and is treated as a systematic uncertainty. The bias on the extracted signal yield was assessed by comparing the signal yield extracted from the fit with that from a simple count. It is given by $\frac{N_{\text{Rec}}-N_{\text{Fitted}}}{N_{\text{Fitted}}}$, where N_{Rec} is the number of reconstructed events and N_{Fitted} is the number of events obtained by integrating the fitted spectrum over the full $m_{4\ell}$ range.

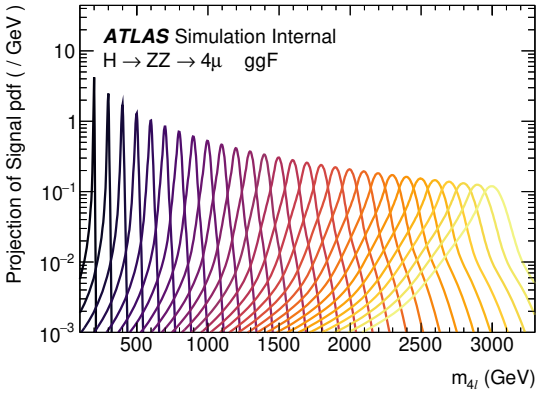
Figure 5.2 shows an example fit of a ggF signal sample at a mass point of $m_H = 700$ GeV to the sum of a Crystal Ball and a Gaussian function, as well as a series of projections of $m_{4\ell}$ signal shapes obtained from the parametrisation procedure described above. The signal $m_{4\ell}$ spectra are very similar in the ggF and VBF regions and are for that reason only illustrated for the ggF category. In the figures of this thesis, panels labelled 'pull' show the standardised residual between an observation and a fit. For an observation x_i with fitted value $\bar{x}_i$ and fit uncertainty σ_i, it is given by $(x_i - \bar{x}_i)/\sigma_i$.

5.6.1.2 LWA scalar and graviton hypotheses

The shape of the $m_{4\ell}$ distribution in the case of LWA scalar particles and spin-2 gravitons are modelled as a convolution of a truth-level distribution with a detector-resolution function. It is given for LWA scalar particles in Appendix B.1, and for gravitons in Appendix B.2. The detector-resolution function is taken to be the same as the one obtained for the NWA signal parametrisation (Section 5.6.1), since



(a) ggF sample fitted to a Crystal Ball + Gaussian function.



(b) Projection of signal $m_{4\ell}$ shapes between 200 and 3000 GeV.

Figure 5.2: (a) $m_{4\ell}$ spectrum of a ggF sample in the 4μ channel at a mass point of 700 GeV (black) fitted to a Crystal Ball + Gaussian function (blue). The bottom panel shows the standardised residuals between the simulated distribution and its fit (pull). (b) The projection of $m_{4\ell}$ shapes based on interpolating the fit parameters obtained from fits at all available mass points, such as as the one shown in (a). Figures reproduced from Reference [1].

due to the negligible width of the resonance the $m_{4\ell}$ shape in the NWA is entirely dominated by the detector resolution.

The modelling of the LWA scalar hypothesis must take into account interference effects between the heavy resonance and SM processes. The SM background process of $gg \to ZZ \to \ell^+\ell^-\ell'^+\ell'^-$, the ggF-produced SM Higgs boson decay to $ZZ \to \ell^+\ell^-\ell'^+\ell'^-$, and the ggF-produced heavy resonance in the LWA decaying to $ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ all share the same initial state of gg and the same final state of $\ell^+\ell^-\ell'^+\ell'^-$, and therefore interfere with one another. The interference contribution becomes significant when the resonance width is non-negligible, affecting signal yields by up to 10% [164]. It becomes more significant as the natural width of the scalar increases. In the case of the interference between the heavy scalar and the SM $gg \to ZZ$ process, dedicated MC simulation samples are generated. The overall interference contribution is estimated by subtracting the $m_{4\ell}$ distributions of the signal-only and background-only samples from the dedicated interference samples at truth level. The resulting $m_{4\ell}$ distribution is convolved with the detector-resolution function obtained previously. In the case of a heavy scalar interfering with the SM Higgs boson, it is assumed that the only difference between the two types of Higgs bosons in the production cross-section must originate from the propagator. The interference contribution is obtained by reweighting the truth-level

$m_{4\ell}$ shape of the generated LWA signal samples with a function based on the different propagators associated with the SM Higgs boson and the heavy Higgs boson, the details of which are given in Appendices B.1 to B.3.

In the case of the graviton hypothesis, the natural width is assumed to be relatively small, and it therefore does not take into account interference modelling.

5.6.2 Background modelling

The SM ZZ continuum background from the processes $q\bar{q} \to ZZ$, $gg \to ZZ$, and $q\bar{q} \to ZZ$ (EW) is modelled by fitting an empirical function to the MC-simulated $m_{4\ell}$ distributions, while the overall normalisation is a free parameter in the statistical fit (Section 5.9). The background modelling function is given by

$$f_{m_{4\ell}} = C_0 \times H(m_0 - m_{4\ell}) \times f_1(m_{4\ell}) + H(m_{4\ell} - m_0) \times f_2(m_{4\ell}), \tag{5.3}$$

where

$$f_1(m_{4\ell}) = \left(\frac{m_{4\ell} - a_4}{a_3}\right)^{a_1-1} \left(1 + \frac{m_{4\ell} - a_4}{a_3}\right)^{-a_1-a_2}, \tag{5.4}$$

$$f_2(m_{4\ell}) = \exp\left[b_1 \left(\frac{m_{4\ell} - b_5}{b_4}\right)^{b_2-1} \left(1 + \frac{m_{4\ell} - b_5}{b_4}\right)^{-b_2-b_1}\right], \tag{5.5}$$

$$C_0 = \frac{f_2(m_0)}{f_1(m_0)}, \tag{5.6}$$

and a_i and b_i are real-valued fit parameters.

Incorporating the overall normalisation as a free parameter in the fit helps constrain systematic uncertainties in the ZZ yield associated with theoretical uncertainties on the SM ZZ production rate. This approach has been validated in Reference [1], and reduces systematic uncertainties on the main backgrounds by up to 30%.

Most other sources of background are modelled using MC simulation, both in shape and normalisation. Exceptions to this are light-flavour jets, where the $m_{4\ell}$ distribution is extracted from a dedicated data control region; this method is described in Reference [1].

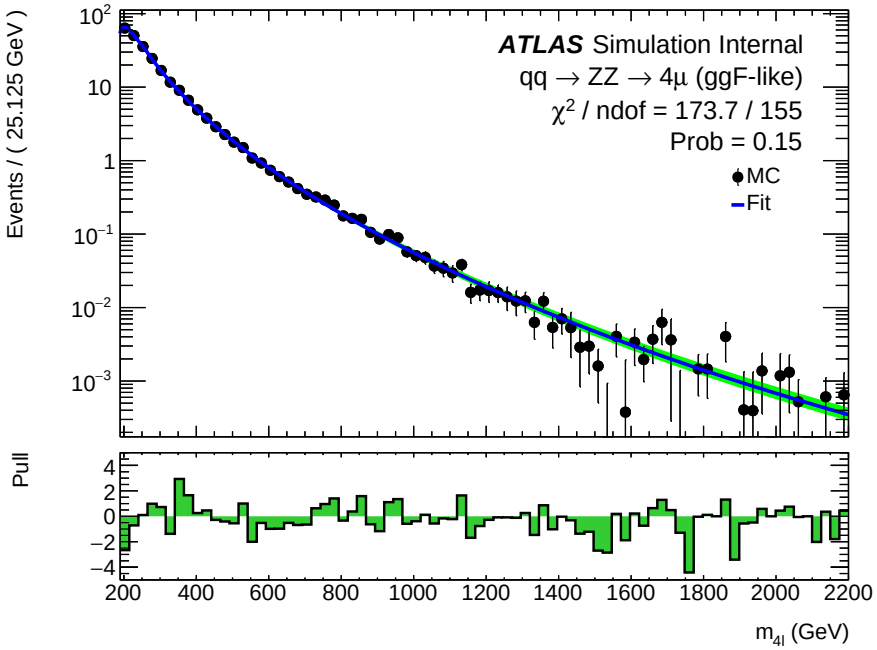


Figure 5.3: $m_{4\ell}$ distribution of the MC-simulated $q\bar{q} \to ZZ$ background in the ggF category and the 4μ final state (black), fitted to an empirical function given in Equation (5.3) (blue). The bottom panel shows the standardised residuals between the simulated distribution and its fit (pull). The green band in the upper panel corresponds to a statistical uncertainty in the fit of $\pm 1\,\sigma$. Figure reproduced from Reference [1].

5.7 Cut-based event categorisation

The previous iteration of the analysis [58] using 36 fb^{-1} of data employed the following strategy for categorising events as ggF-like and VBF-like: if an event contains two or more jets ($N_{\text{jets}} \geq 2$) that are well separated such that $\Delta\eta_{jj} > 3.3$, and their invariant mass, m_{jj}, is at least 400 GeV, it is categorised as VBF-like[2]. Otherwise, an event is classified as ggF-like. This includes all events with fewer than two jets ($N_{\text{jets}} < 2$). Due to the two simple cuts on two high-level event variables, this event-categorisation method is referred as the *cut-based categorisation*. Figure 5.4 shows the Feynman diagrams for the two production mechanisms under consideration.

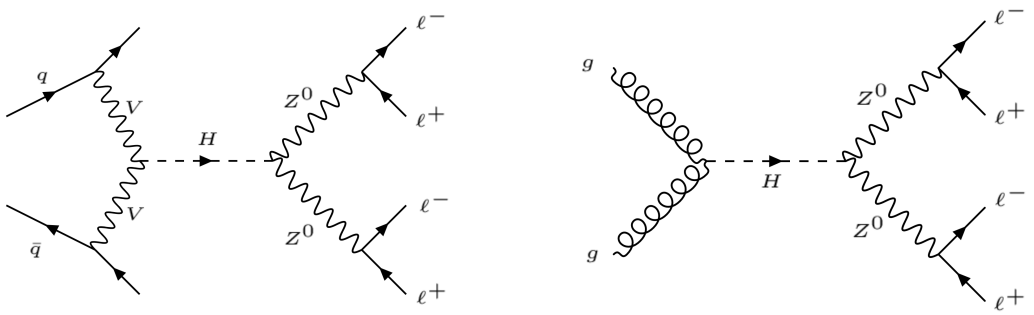


Figure 5.4: Leading-order Feynman diagrams for the two Higgs production mechanisms under consideration. (a) VBF production of a Higgs, decaying to two Z bosons, each decaying to two charged leptons. (b) ggF production of a Higgs boson decaying to two Z bosons, each decaying to two charged leptons.

One advantage of the cut-based categorisation method is its model independence, in that the two dijet variables it uses are assumed to be uncorrelated with the internal structure of the resonance that is produced in the ggF or VBF process (such as whether it is a scalar or a vector particle). This is because the VBF resonance production occurs after the two incoming quarks each radiate a vector boson, which subsequently are identified as jets. For reasons of model independence, the cut-based categorisation continues to be used in this iteration of the analysis in the search for large-width scalars and in the search for spin-2 gravitons. Furthermore, the cut-based categorisation scheme is offered as an alternative strategy in the search for narrow resonances. The latter interpretation is not presented in this chapter; it is available in the appendix to Reference [3].

[2]The validity of these cuts was investigated as a preliminary study to the work presented in this chapter. All possible simultaneous cuts on $\Delta\eta_{\text{jj}}$ and m_{jj} were assessed with respect to their impact on the discovery significance of a potential ggF or VBF signal at all available mass points. The study confirmed that the previous cuts are very close to optimal.

5.8 Event categorisation using deep neural networks

While the cut-based categorisation method outlined in Section 5.7 is well-motivated by the kinematics of the VBF production mechanism, which in the majority of cases produces a well-separated, high-mass jet pair, it is ultimately limiting, as it does not take into account other event variables that may contain information that is correlated with the production mechanism.

In order to improve the classification of events into ggF-like and VBF-like, two multivariate classifiers are developed. Both classifiers are based on RNNs and MLPs[3]. The first classifier is trained on simulated VBF events as signal and SM events as background, and is intended to improve the separation between VBF events and SM background events. The second classifier is trained on ggF events as signal, and, again, SM events as background. The classifiers are applied in sequence: A given event under consideration is considered VBF-like if it scores above a certain threshold on the VBF classifier's output score. Failing that, if it scores above another threshold on the ggF classifier's output score, it is considered ggF-like. Failing both cuts, the event is considered background-like.

5.8.1 Classifier architecture

The two classifiers are similar in architecture. The VBF classifier comprises two RNNs and an MLP. The first RNN takes as input the p_{T}-ordered pairs of p_{T} and η of the four charged leptons. The second RNN takes as input the p_{T}-ordered pairs of p_{T} and η of the jets in the event. The MLP receives a set of high-level event variables as input.

The ggF classifier comprises a single RNN, which, just like in the case of the VBF classifier, takes the p_{T}-ordered pairs of p_{T} and η of the four charged leptons as input. In addition, it consists of an MLP, which as its input receives the remaining high-level event variables, as well as the p_{T} and η of up to one jet. In both cases, the layers are concatenated and fed into several additional deep layers, the last layer of which produces an output score. The network architectures are summarised in Figures 5.5 and 5.6. The models were implemented using the `TensorFlow` library [165].

[3]It is worth mentioning that because several neural-network layers are involved, the classifiers are frequently referred to as DNNs throughout this chapter.

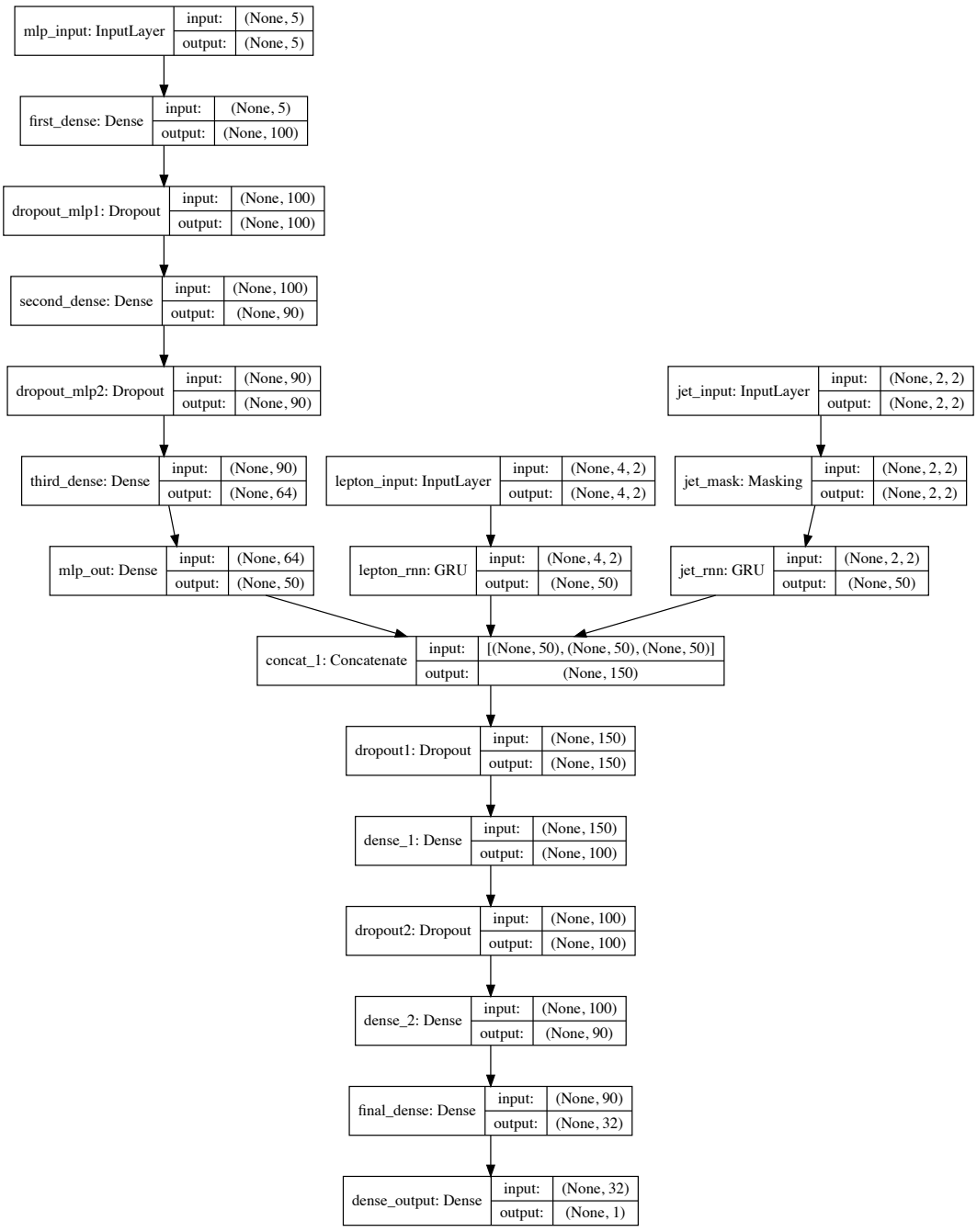


Figure 5.5: Diagram of the VBF classifier architecture. The upper part of the diagram shows the three different network components: the MLP layer on the left with an input dimension of 1 (labelled `mlp_input`), the lepton RNN in the middle with an input dimension of 4×2 (labelled `lepton_input`), and the jet RNN on the right with an input dimension of 2×2 (labelled `jet_input`). The lower part of the diagram shows how the three input layers are combined in a *concatenation*, run through several dense and dropout layers, in order to eventually end on a single output neuron. The 'None' labels in the first dimension of each layer are a placeholder for the batch size, which is a variable training parameter and of no importance to the classifier architecture.

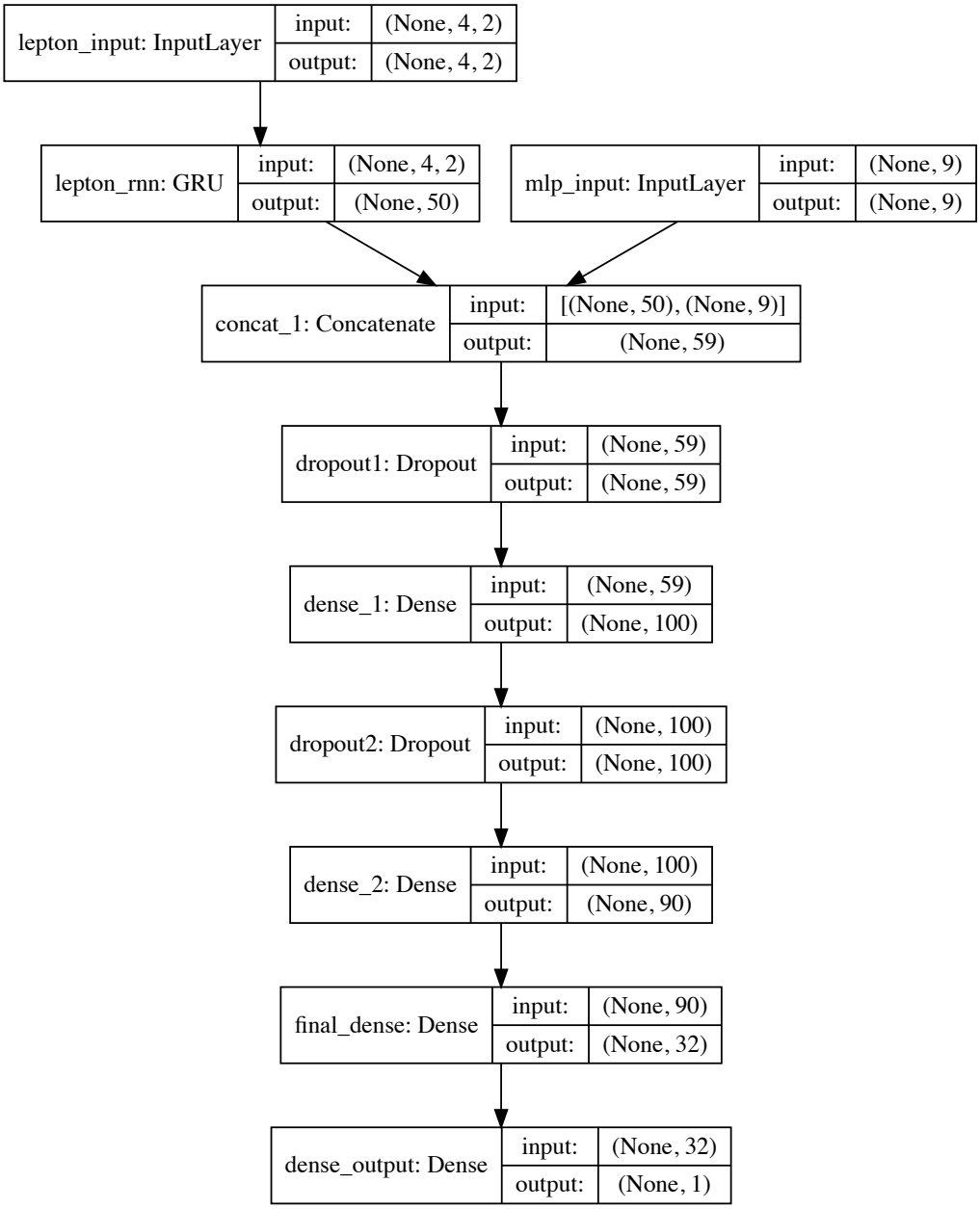


Figure 5.6: Diagram of the ggF classifier architecture. The top of the diagram shows the two different input layers: the lepton RNN on the left with a dimension of 4×2 (labelled `lepton_input`) and the MLP on the right with a dimension of 9 (labelled `mlp_input`). The two input layers are concatenated and run through several dense and dropout layers, eventually ending on a single output neuron. The 'None' labels in the first dimension of each layer are a placeholder for the batch size, which is a variable training parameter and of no importance to the classifier architecture.

5.8.2 Input features

5.8.2.1 Feature-selection method

The two classifiers' input features are primarily selected in order to exploit the dijet kinematics in the VBF category (and the absence thereof in the ggF category). Both classifiers were trained with a large set of high-level event variables, which was iteratively reduced in size in a process called *pruning*.

The selection of features follows slightly different methods for the two classifiers due to the availability of a more mature software library used for feature selection at the time the ggF classifier was developed (several months after the VBF classifier).

In the case of the ggF classifier, the *SHAP valuse* [166] were used as a proxy for the importance of a given variable. Based on Shapley values, developed initially in the context of game theory [167], SHAP values aim to quantify the contribution that each input feature makes to the output score of a classifier. Details on the computation of SHAP values are available in Reference [168]; here, only a brief, intuitive explanation is provided. Given N input features, models are trained on a sample from the power set formed by $N-1$ features, and the difference in model output is compared with the original model trained on all N features. For a given feature k, the SHAP value is computed from the difference in output score between the model trained on all features, $\hat{f}_N$, and the model trained on all features *except for* feature k, $\hat{f}_{N\setminus k}$. The SHAP values were implemented using the `shap` software package [166].

The feature pruning for the VBF classifier is based on a related measure of variable importance, called the *permutation importance* [169]. The calculation of the permutation importance is based on a single classifier trained on all N features. A test set is evaluated on the trained classifier by replacing one feature at a time with random noise, and comparing its output on the test set with that of the classifier trained on the complete set of features. The difference between the output of the noisy test set and the full test set is taken as a proxy for the importance of a given feature. The feature ranking according to permutation importance was implemented using the `eli5` package [170].

5.8.2.2 Choice of input features

The classifiers were initially trained with a large set of N input features, ranked according to their mean SHAP value. The lowest-ranking feature was removed, and the classifier was trained on the remaining $N-1$ input features, and so forth. The

pruning procedure was halted once the lowest-ranking feature had a SHAP value above a threshold value of 0.001. The cut-off value of 0.001 was chosen as it roughly corresponds to the uncertainty in SHAP value observed for a given feature when repeating performing repeated training runs with the same overall set of features. In other words, the set of input variables was chosen such that the average impact on the model output of each feature was at least 0.1%.

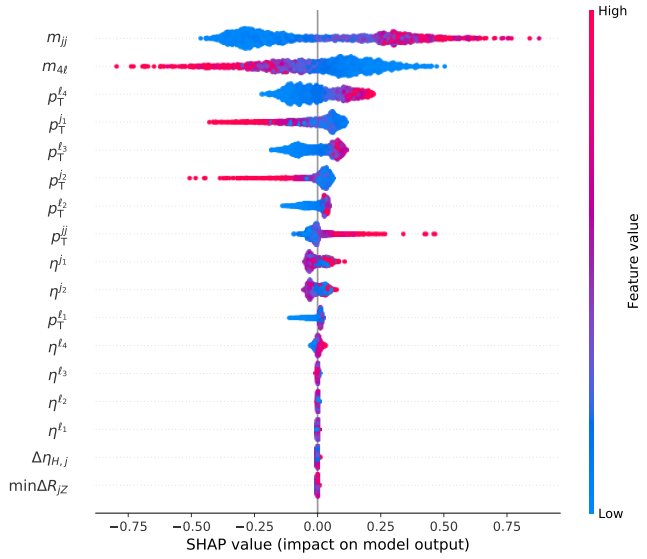
The final choice of input features to the two classifiers is summarised in Tables 5.2 and 5.3. The SHAP value rankings for the two classifiers are presented in Figure 5.7. It is noteworthy that the dijet invariant mass, m_{jj}, is the highest-ranking feature for the VBF classifier, indicating that the classifier chose this variable—the one most sensitive to the dijet topology along with the two jets' pseudorapidities—as the most important one. This is similar to the cut-based categorisation, in which m_{jj} is one of the two dijet variables used.

variable	description
$m_{4\ell}$	4ℓ invariant mass
m_{jj}	dijet invariant mass
p_{T}^{jj}	dijet transverse momentum
$\Delta\eta_{H,j}$	difference in pseudorapidities between the 4ℓ system and the leading jet
$\min \Delta R_{jZ}$	minimum angular separation between one of the two $\ell\ell$ pairs and a jet
p_{T}^{j}	transverse momenta of the two leading jets
η^{j}	pseudorapidities of the two leading jets
p_{T}^{ℓ}	transverse momenta of the four charged leptons
η^{ℓ}	pseudorapidities of the four charged leptons

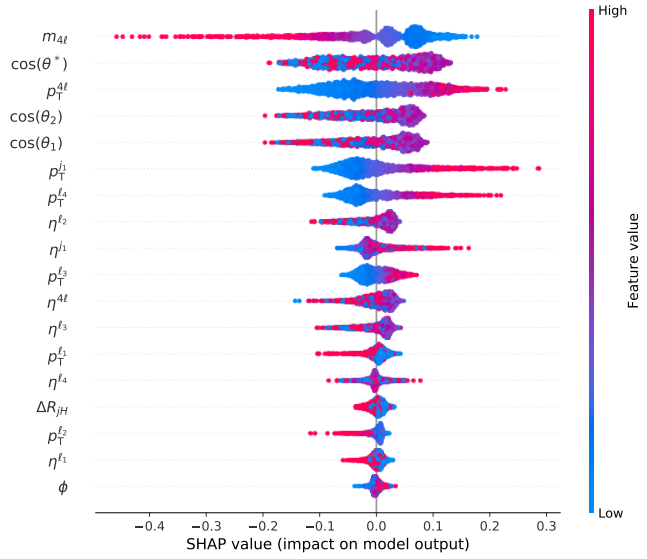
Table 5.2: Input features to the VBF classifier.

It is worth noting that in the feature ranking based on permutation importance, all features shown in Table 5.2 scored sufficiently highly, however, Figure 5.7a showing the SHAP values indicates that the five lowest-ranking variables could have been safely removed without affecting the overall model output significantly. The inclusion of these features in the final classifier is not expected to be problematic, as partially redundant features generally do not negatively impact model accuracy.

The distributions of input features for four signal samples and the combined SM background are shown in Figures 5.8 to 5.11. In the figures of this thesis, distributions that are normalised to unit area are denoted by a.u. (arbitrary units) on the ordinate.



(a) SHAP value rankings for the VBF classifier.



(b) SHAP value rankings for the ggF classifier.

Figure 5.7: Ranked SHAP values for all input features of the VBF (a) and ggF (b) classifiers. The ranking is according to the magnitude of the SHAP value across 5000 tested events. The y-axes show the input features. The z-axes correspond to the value of a given feature, ranging from its minimum value (blue) to its maximum (magenta).

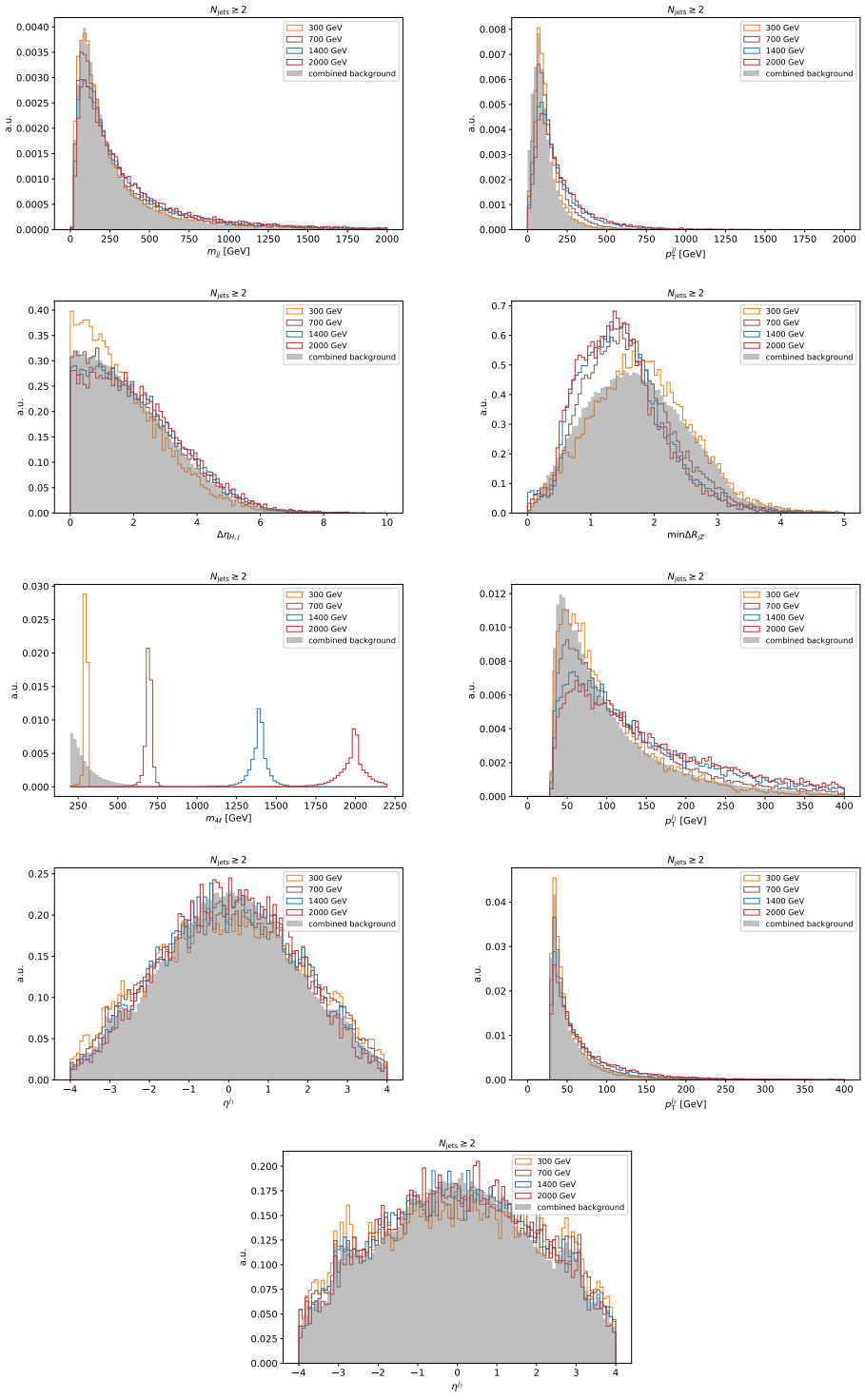


Figure 5.8: The MLP and jet-RNN input features to the VBF classifier for four different signal mass points (300, 700, 1400, 2000 GeV) (coloured) and the combined background (grey). Only events satisfying the training selection cut of $N_{\text{jets}} \geq 2$ are shown. The distributions, from left to right, top to bottom, are m_{jj}, p_{T}^{jj}, $\Delta\eta_{H,j}$, $\min\Delta R_{jZ}$, $m_{4\ell}$, $p_{\text{T}}^{j_1}$, η^{j_1}, $p_{\text{T}}^{j_2}$, and η^{j_2}. All distributions are normalised to unit area to facilitate comparison (a.u.).

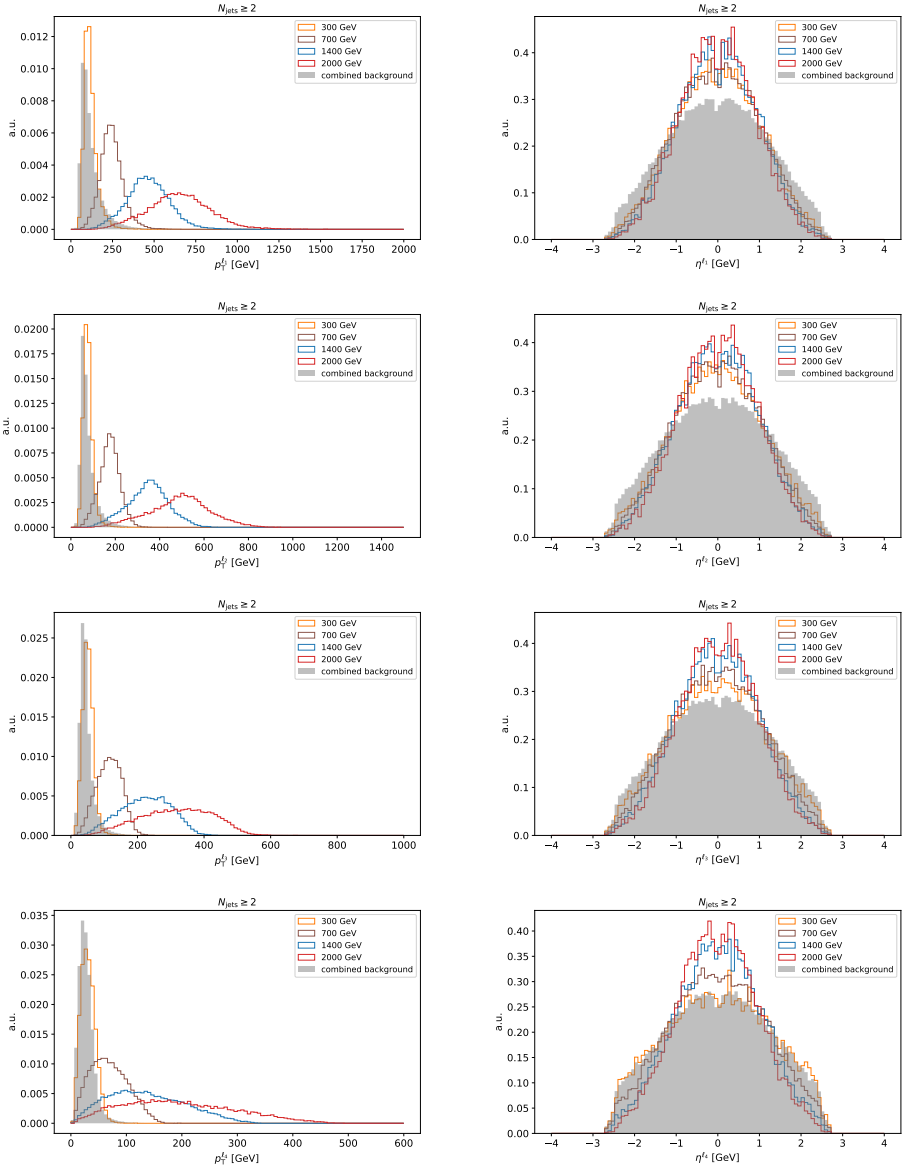


Figure 5.9: Input variables used in the VBF classifier that are passed to the lepton RNN for four different signal mass points (300, 700, 1400, 2000 GeV) (coloured) and the combined background (grey). Only events satisfying the training selection cut of $N_{\text{jets}} \geq 2$ are shown. The distributions, from left to right, top to bottom, are $p_{\text{T}}^{\ell_1}$, η^{ℓ_1}, $p_{\text{T}}^{\ell_2}$, η^{ℓ_2}, $p_{\text{T}}^{\ell_3}$, η^{ℓ_3}, $p_{\text{T}}^{\ell_4}$, and η^{ℓ_4}. All distributions are normalised to unit area to facilitate comparison (a.u.).

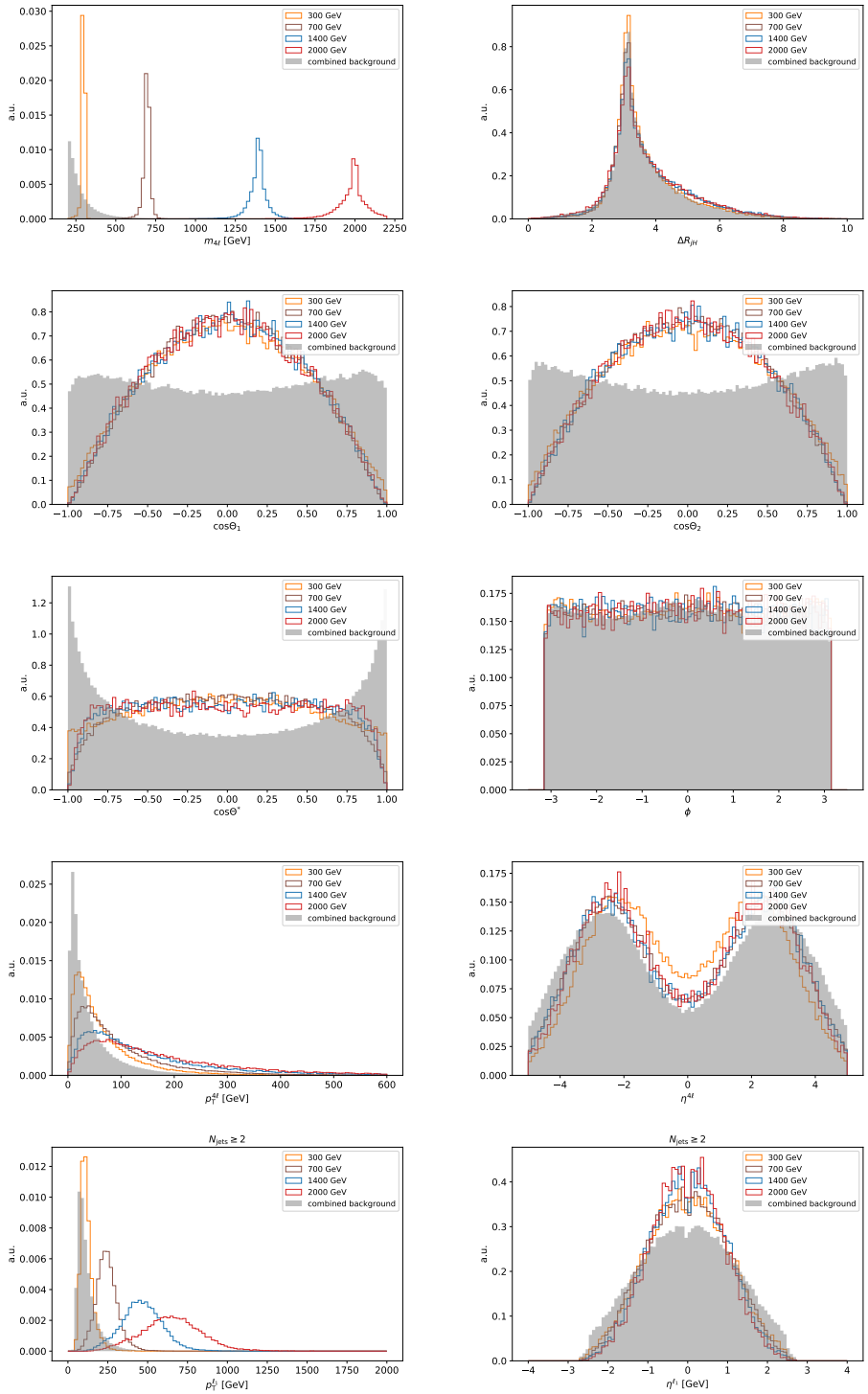


Figure 5.10: MLP input features to the ggF classifier for four different signal mass points (300, 700, 1400, 2000 GeV) (coloured) and the combined background (grey). Events with any jet multiplicity are shown, as this classifier is used in both categories ($N_{\text{jets}} \geq 2$ and $N_{\text{jets}} < 2$). The distributions, from left to right, top to bottom, are $m_{4\ell}$, ΔR_{jH}, $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, ϕ, $p_{\text{T}}^{4\ell}$, $\eta^{4\ell}$, $p_{\text{T}}^{\ell_1}$, and η^{ℓ_1}. All distributions are normalised to unit area to facilitate comparison (a.u.).

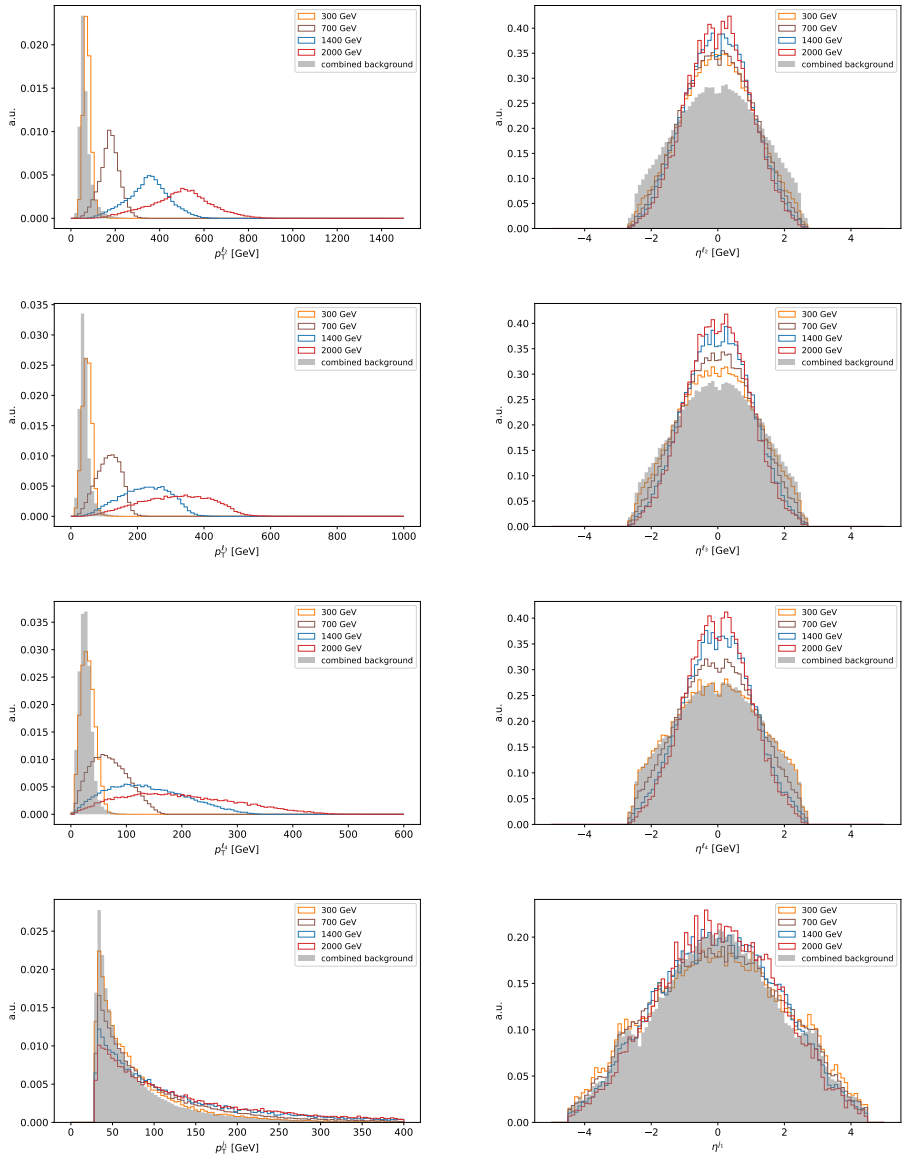


Figure 5.11: Input features to the ggF classifier used in the lepton RNN for four different signal mass points (300, 700, 1400, 2000 GeV) (coloured) and the combined background (grey). Events with any jet multiplicity are shown, as this classifier is used in both categories ($N_{\text{jets}} \geq 2$ and $N_{\text{jets}} < 2$). The distributions, from left to right, top to bottom, are $p_{\text{T}}^{\ell_1}$, η^{ℓ_1}, $p_{\text{T}}^{\ell_2}$, η^{ℓ_2}, $p_{\text{T}}^{\ell_3}$, η^{ℓ_3}, $p_{\text{T}}^{\ell_4}$, and η^{ℓ_4}. All distributions are normalised to unit area to facilitate comparison (a.u.).

variable	description
$m_{4\ell}$	4ℓ invariant mass
$\cos\theta_1$	decay angle of the leading Z
$\cos\theta_2$	decay angle of the sub-leading Z
$\cos\theta^*$	production angle of the ZZ system
ΔR_{jH}	angular separation between the 4ℓ system and the leading jet
ϕ	azimuthal angle of the ZZ system
$p_{\mathrm{T}}^{4\ell}$	transverse momentum of the $\ell^+\ell^-\ell'^+\ell'^-$ system
$\eta^{4\ell}$	pseudorapidity of the $\ell^+\ell^-\ell'^+\ell'^-$ system
p_{T}^{j}	transverse momentum of up to one jet
η^{j}	pseudorapidity of up to one jet
p_{T}^{ℓ}	transverse momenta of the four charged leptons
η^{ℓ}	pseudorapidities of the four charged leptons

Table 5.3: Input features to the ggF classifier.

5.8.3 Training statistics

Both networks are trained on VBF and ggF signal samples generated at the following mass points: $m_H = 200$, 300, 400, 500, 600, 700, 800, 900, 1000, 1200 and 1400 GeV. The background samples used for training comprise the processes $q\bar{q} \to ZZ$, $gg \to ZZ$ and $q\bar{q} \to ZZ$ (EW). All events considered for training are required to pass the event-selection cuts outlined in Section 5.4. In addition, events are required to lie within the mass range $200 < m_{4\ell} < 1300$ GeV, as the available background statistics above 1300 GeV are far smaller than those available for signal, and meaningful training in that mass range proved infeasible. As will be demonstrated in Section 5.8.11, the networks show enough generalisation capacity to extrapolate to masses the classifiers were not provided during training well beyond the cut-off value of $m_{4\ell} = 1300$ GeV applied during training.

The VBF network, aimed at exploiting the kinematics of events with a jj signature, is trained on events that contain two or more jets. The ggF network, on the other hand, is trained on events with at most one jet. It can, however, be very effectively applied to events with more than one jet, in which case the leading (i.e. highest-p_{T}) jet's η and p_{T} are fed into the MLP part of the network as input to fill the variables η^j and p_{T}, while all other jets are ignored. The training of the ggF network is restricted to events with at most one jet so as to have a statistically independent background component during training (given the $N_{\mathrm{jets}} > 2$ requirement on the VBF network). To assess the soundness of this approach, two versions of the ggF

network were trained (one with at most one jet in a given event, and another one with any jet multiplicity), and no significant difference in network output, when applied to events with any jet multiplicity, was observed. For events with zero jets, η^j and p_{T}^j are assigned a value of 0. Note that this happens both during training and during evaluation of the network. The overall number of events used for training are summarised in Tables 5.4 to 5.7.

m_H	Raw events per MC campaign			Total
	mc16a	mc16d	mc16e	
200 GeV	2213	2176	2210	6599
300 GeV	1306	1623	24699	27628
400 GeV	7020	7310	7181	21511
500 GeV	6406	7576	7531	21513
600 GeV	7335	7799	7508	22642
700 GeV	1529	1718	28486	31733
800 GeV	7322	7756	7592	22670
900 GeV	7473	7801	7584	22858
1000 GeV	7381	7681	7639	22701
1200 GeV	7157	7404	7362	21923
1400 GeV	435	575	481	1491
Total	55577	59419	108273	223269

Table 5.4: Number of raw VBF signal events used during training of the VBF DNN passing the basic and training selection cuts for the three simulation campaigns and in total. The mc16a campaign is designed to mimic data-taking conditions in 2015–2016, mc16d those in 2017, and mc16e those in 2018.

Process	Raw events per MC campaign			Total
	mc16a	mc16d	mc16e	
$q\bar{q} \to ZZ$	192234	342893	319883	855010
$q\bar{q} \to ZZ$ (EW)	6812	8612	6777	22201
$gg \to ZZ$	13170	18732	18732	50634
Total	212216	370237	345392	927845

Table 5.5: Number of raw background events used during training of the VBF DNN passing the basic and training selection cuts for the three simulation campaigns and in total.

m_H	Raw events per MC campaign			Total
	mc16a	mc16d	mc16e	
200 GeV	2944	2879	2765	8588
300 GeV	17203	19447	27795	64445
400 GeV	9365	8808	8791	26964
500 GeV	9213	8861	8820	26894
600 GeV	9253	8728	8683	26664
700 GeV	18338	22453	32601	73392
800 GeV	9097	8417	8617	26131
900 GeV	8772	8299	8289	25360
1000 GeV	8668	8073	7957	24698
1200 GeV	8245	7584	7694	23523
1400 GeV	446	498	502	1446
Total	101544	104047	122514	328105

Table 5.6: Number of raw ggF signal events used during training of the ggF DNN passing the basic and training selection cuts for the three simulation campaigns and in total.

Process	Raw events per MC campaign			Total
	mc16a	mc16d	mc16e	
$q\bar{q} \to ZZ$	791848	988576	1143735	2924159
$q\bar{q} \to ZZ$ (EW)	1330	1443	1178	3951
$gg \to ZZ$	59845	70216	70216	200277
Total	853023	1060235	1215129	3128387

Table 5.7: Number of raw background events used during training of the ggF DNN passing the basic and training selection cuts for the three simulation campaigns and in total.

5.8.4 Reweighting

The signal has been generated at discrete mass points; therefore, merely combining all signal samples would lead to a classifier that would preferentially assign signal-like scores to events with values of $m_{4\ell}$ close to any of the mass points it was provided during training. In order to overcome this limitation in the absence of a continuous mass spectrum of signal samples, events are reweighted before training such that the sum of all $m_{4\ell}$ distributions of the available signal samples matches the $m_{4\ell}$ spectrum of the smoothly falling ZZ continuum background. This ensures equal importance is assigned to reweighted signal and background events across the entire mass spectrum. N.B.: the events are only reweighting during training of the classifier. When the trained classifiers are applied to test sets or to data, event weights are not modified.

The reweighting procedure is as follows: a background event with index i receives a sample weight during training given by

$$w_i^{\text{bkg}} = w_i, \tag{5.7}$$

where w_i, the event weight, is a number associated with every MC-generated event that reflects the cross-section of the underlying process and the desired luminosity of the generated sample. For N_0 raw generated events of a process with production cross-section σ, and a recorded integrated luminosity $\mathcal{L}_{\text{int}}$, it is defined for event i as

$$w_i = \frac{\sigma \mathcal{L}_{\text{int}}}{N_0}. \tag{5.8}$$

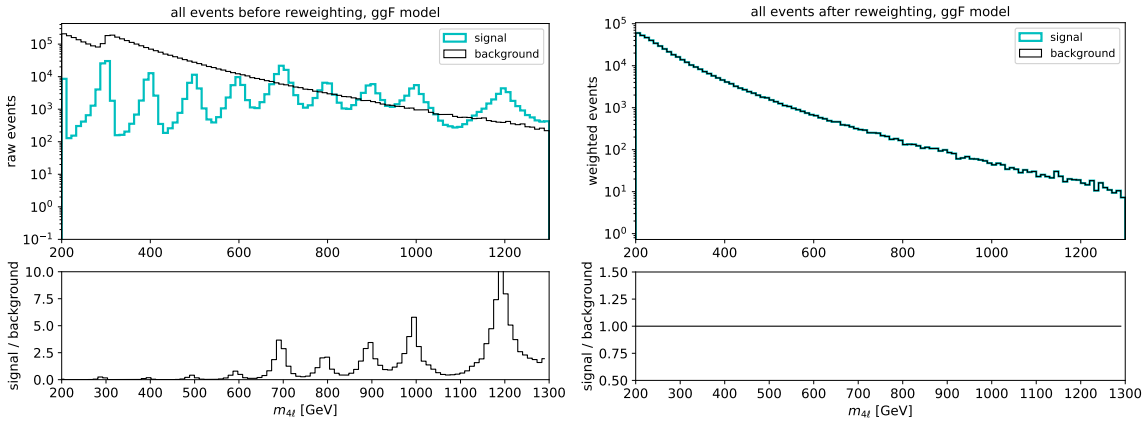
A signal event with index j, event weight w_j, and 4ℓ invariant mass m_j receives the following sample weight during training:

$$w_j^{\text{sig}} = w_j \frac{\sum_l w_l}{\sum_k w_k}, \tag{5.9}$$

where the index k (l) runs over signal (background) events with a 4ℓ invariant mass, $m^{k(l)}$, in units of GeV, that satisfies

$$\lfloor m_j \rfloor \leq m^{k(l)} < \lceil m_j \rceil. \tag{5.10}$$

Here, $\lfloor\rfloor$ and $\lceil\rceil$ are the floor and ceiling functions[4], respectively. The $m_{4\ell}$ distributions before and after performing the reweighting procedure are shown in Figure 5.12 for the ggF classifier. The distribution on the right-hand side of the plot shows that the reweighting procedure was applied correctly; the $m_{4\ell}$ distribution of the combined ggF signal samples perfectly overlaps with the background distribution (their ratio is unity everywhere). The reweighting of VBF samples in preparation for training of the VBF classifier follows the same procedure.



(a) Distribution of raw (unweighted) signal training events for ggF signal (blue) and background (black) before the application of training weights.

(b) Distribution of weighted signal training events for ggF signal (blue) and background (black) after application of training weights.

Figure 5.12: Unweighted (a) and weighted (b) training events used to train the ggF classifier. The bottom panels show the ratio between the number of signal events to the number of background events (signal / background).

The reweighting is performed inclusively, that is, for all types of final state, in the case of the VBF classifier. In the case of the ggF classifier, it is performed separately for each of the three possible final states: $4e$, 4μ and $2e2\mu$. This is due to the relatively large fraction of events that are expected to be categorised as ggF-like, compared to VBF-like and background-like.

5.8.5 Hyperparameter optimisation

Hyperparameter searches were performed using the `Hyperopt` package [171] in order to find the optimal set of training parameters. As a result of the hyperparameter

[4]The *floor function* is a function that takes as input a real number x and gives as output the greatest integer less than or equal to x, denoted $\lfloor x \rfloor$. Similarly, the *ceiling function* gives for x the smallest integer greater than or equal to x, denoted $\lceil x \rceil$.

optimisation, the networks are trained over 20 epochs[5] with a batch size of 512 (VBF classifier) or 256 (ggF classifier). The optimisation algorithm is `Adagrad` [172], with binary cross-entropy (Equation (4.24)) as the objective function. Dropout (c.f. Section 4.3.1.5) between network layers is applied with a probability of 0.05. The hidden-layer activation function is $\tanh(x)$, while the final-layer activation function is the sigmoid function (Equation (4.14)).

5.8.6 Choice of classification cut

The optimal classification cut is chosen for good overall sensitivity, in terms of signal acceptance and background rejection, across the entire mass range. The measure of significance, Z, quantifying this is given by

$$Z = \frac{n - b}{\sqrt{b}}, \tag{5.11}$$

where n is the number of observed events for b expected events [173].

Figures 5.13 and 5.14 show the improvement in Z over the cut-based categorisation as a function of VBF and ggF signal mass in the VBF and ggF categories for different cuts on the classifiers scores. For the VBF category, a cut on the VBF classifier score of 0.8 for events with two or more jets produces the best overall performance improvement. For the ggF category, a combination of cuts (a VBF classifier score of less than 0.8 and a ggF classifier score of greater than 0.5 for events with two or more jets, and simply a ggF classifier score greater than 0.5 for events with fewer than two jets), produces the best overall improvement over the cut-based categorisation. The category definitions are summarised diagrammatically in Figure 5.15.

5.8.7 Classifier output

Figure 5.16 shows the classifier output for the SM background samples as well as the 700-GeV VBF and ggF signal samples. Figure 5.17 shows the VBF and ggF classifier response for different VBF and ggF signal samples, respectively. Figure 5.18 compares the VBF and ggF classifier output between the different ATLAS MC simulation campaigns, using the 700-GeV mass point as an example. The plot

[5] During training of a machine-learning classifier, an *epoch* corresponds to an entire pass of the training algorithm over the full training set.

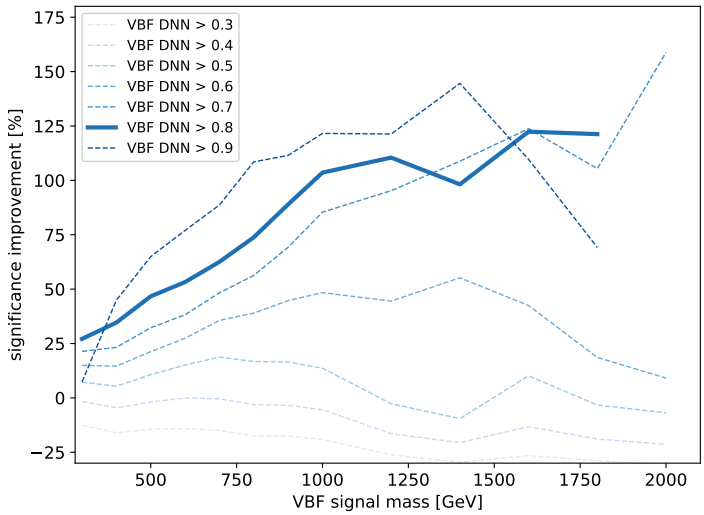


Figure 5.13: Significance improvement of the multivariate over the cut-based definition of the VBF category as a function of VBF signal mass for VBF signal samples between 300 and 2000 GeV for seven different cuts on the VBF classifier. The significance improvement on the ordinate is expressed in percent. The optimal cut of 0.8 is represented by a solid line, while other alternative cuts are plotted using dashed lines. Entries at 2000 GeV for cuts of 0.8 and 0.9 are not included due to a lack of background events passing this selection.

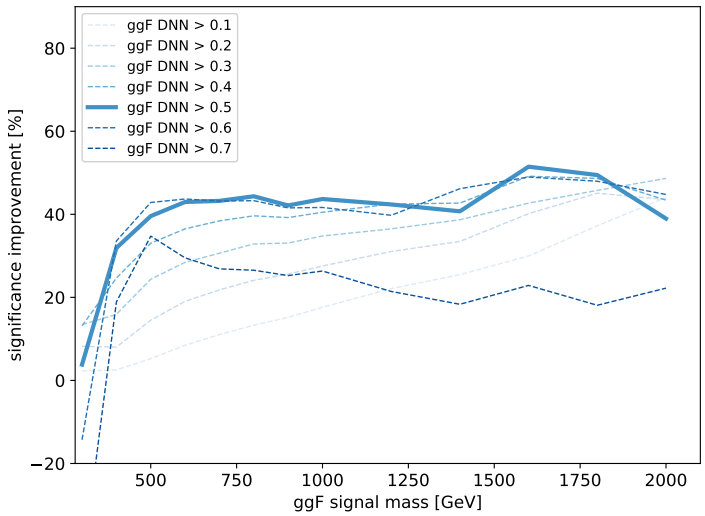


Figure 5.14: Significance improvement of the multivariate over the cut-based definition of the ggF category as a function of ggF signal mass for ggF signal samples between 300 and 2000 GeV for seven different cuts on the ggF classifier. The significance improvement on the ordinate is expressed in percent. The optimal cut of 0.5 is represented by a solid line, while other alternative cuts are plotted using dashed lines.

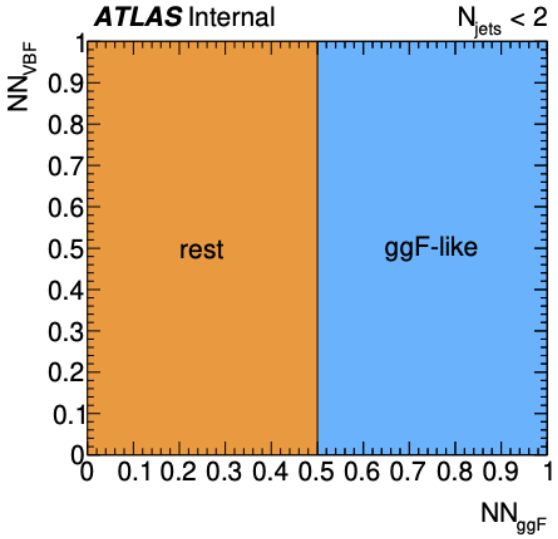
indicates no significant difference in classifier output between the campaigns within statistical uncertainties.

5.8.8 Agreement between data and simulation

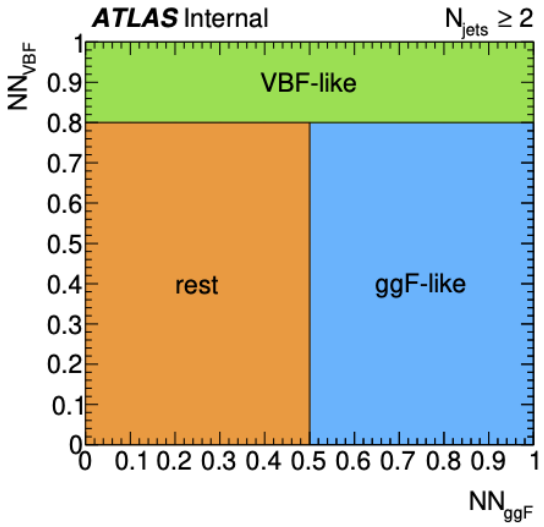
It remains to be shown that the classifier output on data is well-modelled by simulation. Figure 5.19 presents a comparison between the VBF and ggF classifier outputs on data and MC on the entire Run-2 dataset, indicating that the classifier scores are generally well-modelled. The indication of some degree of mismodelling at low values of the ggF classifier (Figure 5.19b) is expected to have a negligible impact on the results of this analysis, as events are classified as ggF-like if their output score is greater than 0.5.

5.8.9 $m_{4\ell}$ signal distributions

Figure 5.20 show the $m_{4\ell}$ distributions of the VBF and ggF samples inclusively and in the VBF and ggF categories, indicating that the VBF and ggF selection cuts do not bias the $m_{4\ell}$ distributions significantly. The distributions before and after the

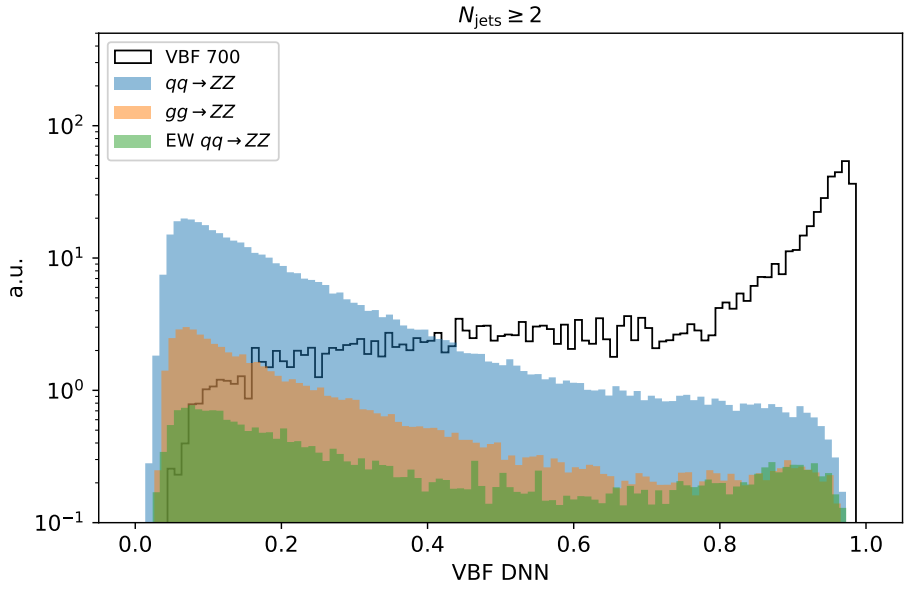


(a) Category definitions for events with fewer than two jets.

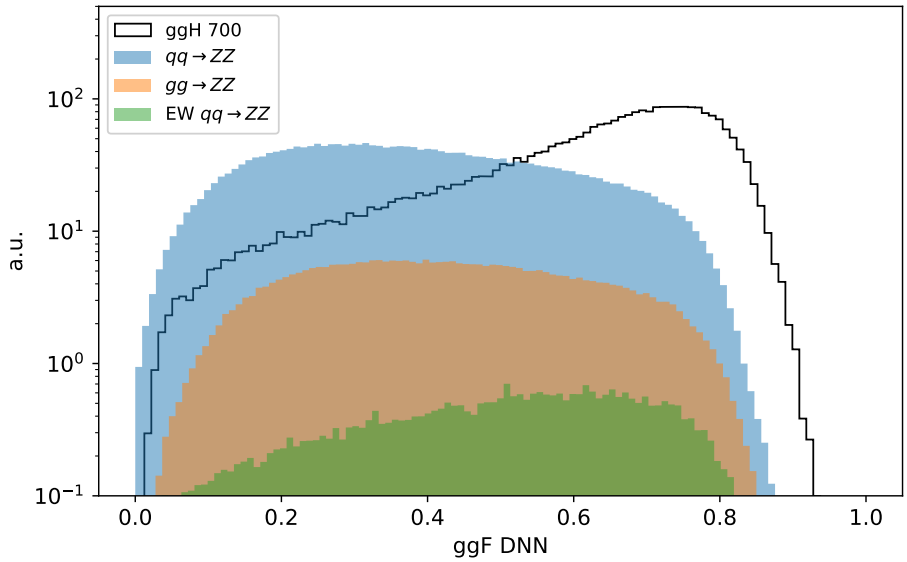


(b) Category definitions for events with two or more jets.

Figure 5.15: VBF, ggF and background category definitions for events passing the basic $\ell^+\ell^-\ell'^+\ell'^-$ selection cuts. The category definitions are shown separately for events with fewer than two jets (a), and for events with two or more jets (b). Here, $\mathrm{NN_{VBF(ggF)}}$ is the VBF (ggF) classifier's output score. The *rest* category is referred to as *background-like* elsewhere in this work. Figures reproduced from Reference [1].

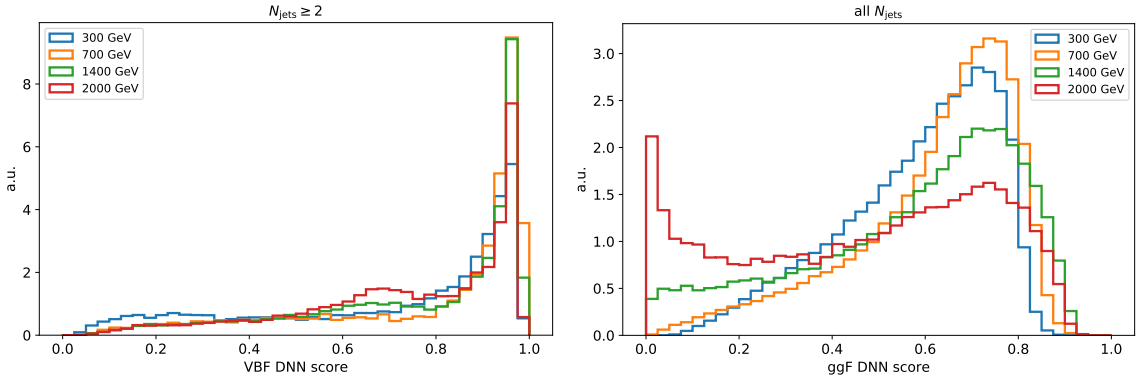


(a) The VBF classifier response for the background samples (filled) and the 700-GeV VBF signal sample (black), for events with two or more jets.



(b) The ggF classifier response for the background samples (filled) and the 700-GeV ggF signal sample (black), for all events regardless of jet multiplicity.

Figure 5.16: VBF (a) and ggF (b) classifier outputs for the three background samples (coloured), as well as one example VBF (a) and ggF (b) signal sample. The distributions are normalised to unit area to facilitate comparison (a.u.).



(a) The VBF classifier response for four VBF signal mass points (300, 700, 1400 and 2000 GeV) for events with two or more jets.

(b) The ggF classifier response for four ggF signal mass points (300, 700, 1400 and 2000 GeV), for events with any jet multiplicity.

Figure 5.17: VBF (a) and ggF (b) classifier output at four different signal mass points: 300, 700, 1400 and 2000 GeV. The distributions are normalised to unit area to facilitate comparison (a.u.).

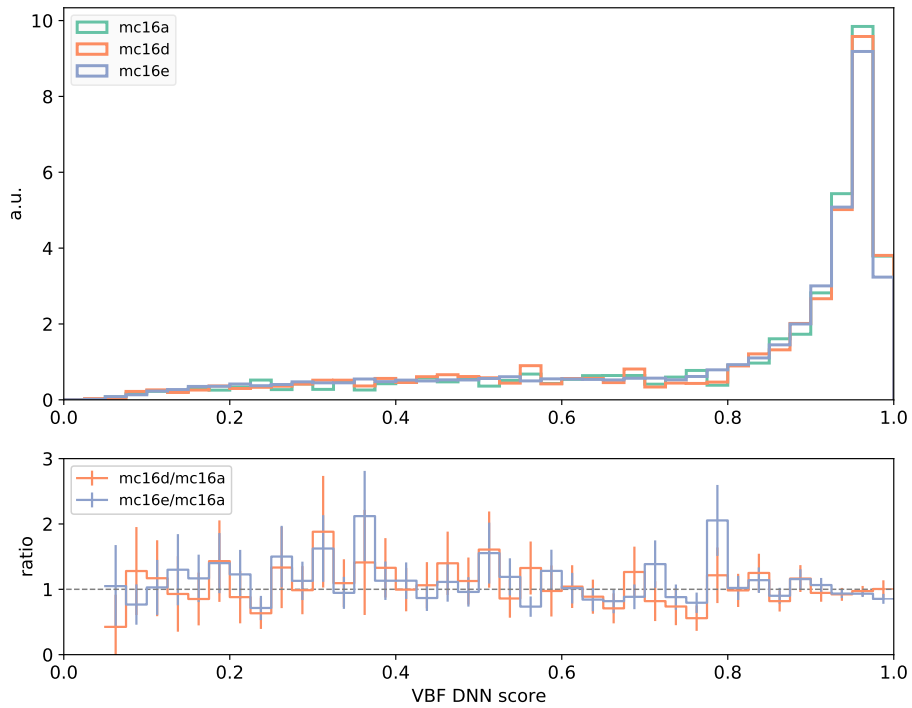
classification cuts overlap nearly perfectly. Any residual bias in the signal shape is taken into account by the signal modelling outlined in Section 5.6.1.

5.8.10 $m_{4\ell}$ background distribution in the VBF and ggF categories

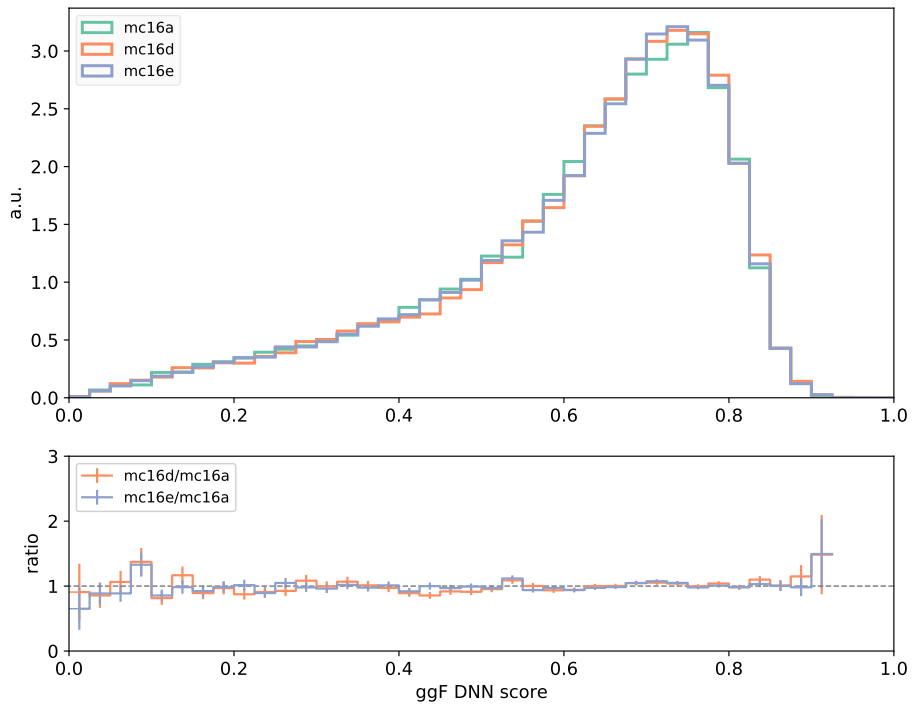
Figure 5.21 compares the shapes of the $m_{4\ell}$ distributions in the inclusive category with those in the VBF and ggF categories, respectively. The overall changes in shape of the $m_{4\ell}$ distributions in the two categories are accounted for in the parametrised modelling of the ZZ continuum background due to the fact that the fits are performed separately in all event categories. Importantly, the classification cuts do not introduce any bumps into the ZZ $m_{4\ell}$ distributions, which could be misinterpreted as signal. The noticeable change in $m_{4\ell}$ shape at low values in the ggF case (Figure 5.21b) indicate that more SM ZZ events migrate into the ggF categories in that region. This is likely caused by the fact that the low-mass region is much richer in SM ZZ background, which makes it a more challenging region for discriminating signal from background.

5.8.11 Classifier bias

While the reweighting procedure of the training samples ensures equal importance is assigned to signal and background events across the entire mass range, the raw event statistics are limited to thirteen mass points for each production mode.

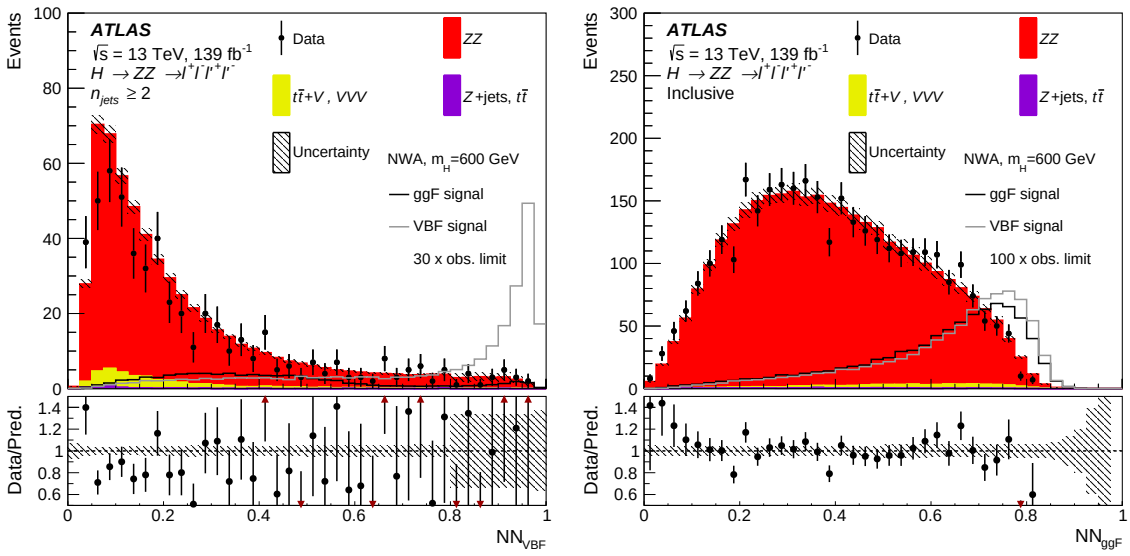


(a) The VBF classifier output for 700-GeV VBF MC samples for MC campaigns mc16a, mc16d and mc16e.



(b) The ggF classifier output for 700-GeV ggF MC samples for MC campaigns mc16a, mc16d and mc16e.

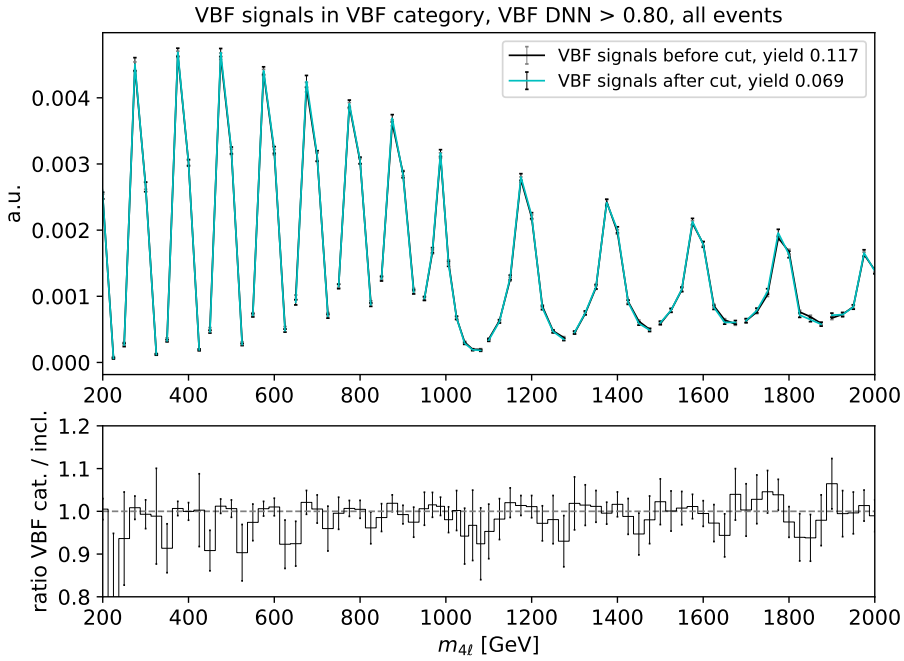
Figure 5.18: Comparison between the three MC campaigns for the VBF DNN score (a) and the ggF DNN score (b). The distributions are normalised to unit area to facilitate comparison (a.u.).



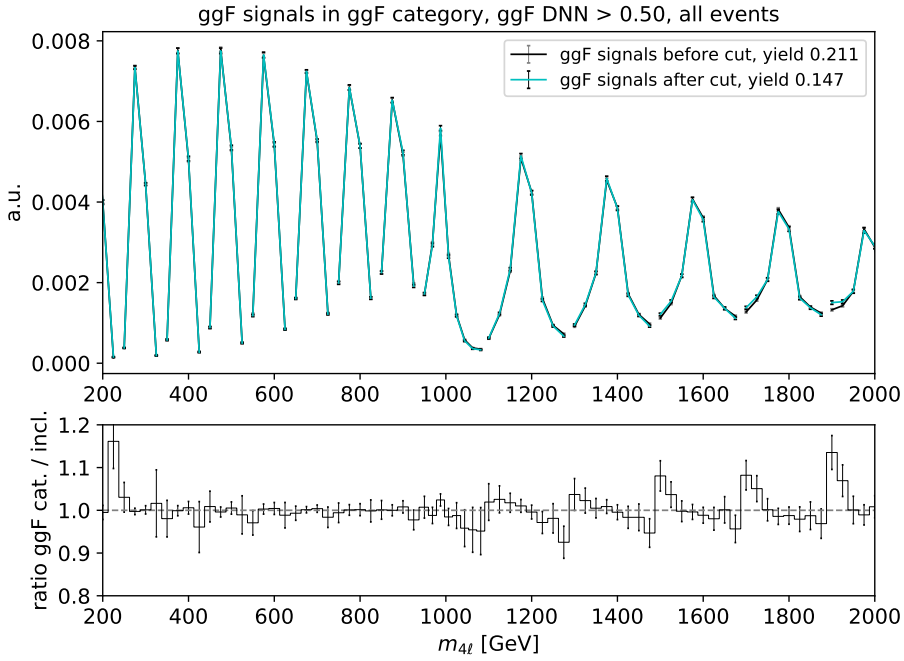
(a) VBF classifier output comparison between data and simulation.

(b) ggF classifier output comparison between data and simulation.

Figure 5.19: Comparison between data (black) and different MC samples (coloured) of the two classifier scores. The comparison is shown for the VBF classifier for events with two or more jets (a), and for the ggF classifier for events with any jet multiplicity (b). The output score of a VBF-produced signal with mass $m_H = 600$ GeV is shown in grey, that of a ggF-produced Higgs boson with mass $m_H = 600$ GeV is shown in black. The bottom panels show the ratio of the number of observed events to the number of predicted events. The grey bands show the magnitudes of the $\pm 1\,\sigma$ statistical uncertainties of the data/prediction estimates. Figures reproduced from Reference [3].

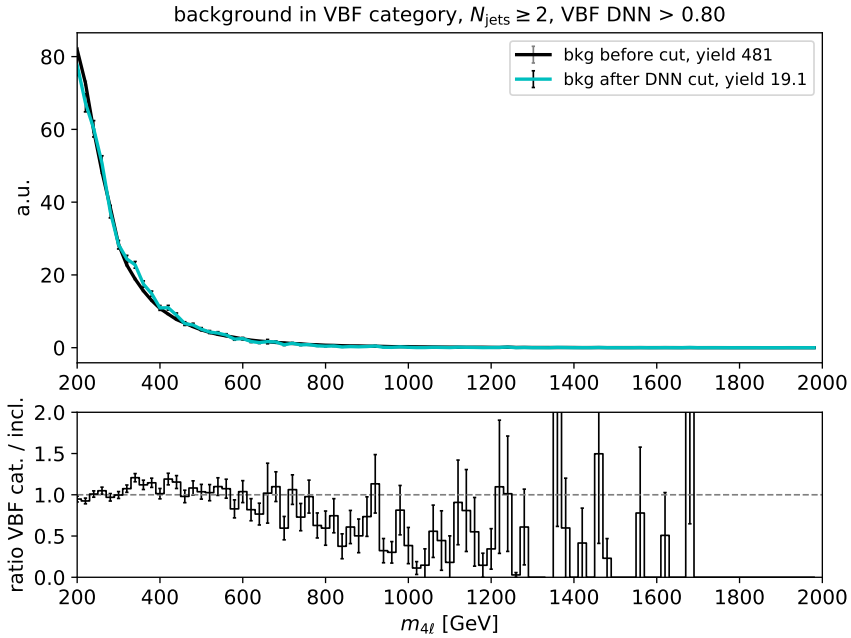


(a) Distribution in $m_{4\ell}$ of 13 VBF samples before any classification cut (black), and in the VBF category (blue).

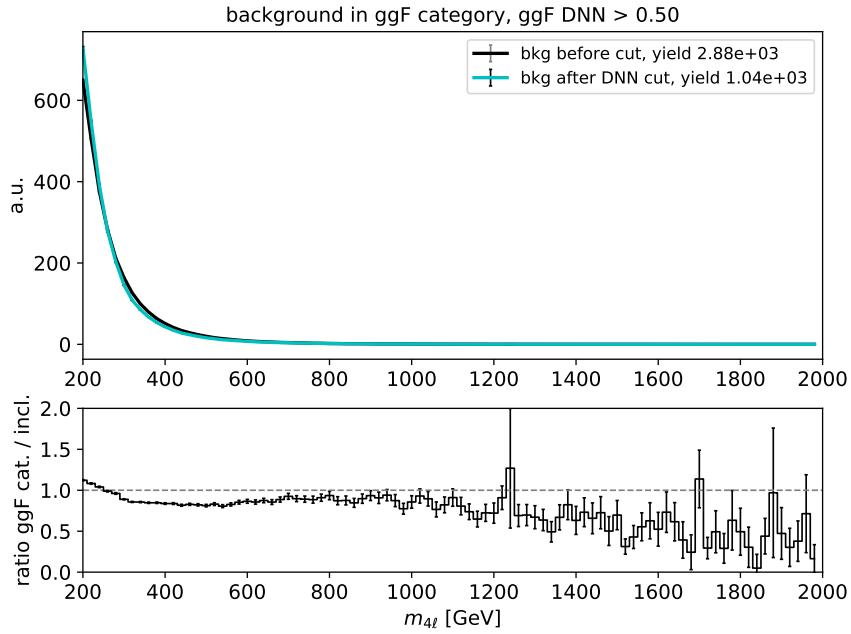


(b) Distribution in $m_{4\ell}$ of 13 ggF samples before any classification cut (black), and in the ggF category (blue).

Figure 5.20: Simulated $m_{4\ell}$ VBF (a) and ggF (b) signal distributions before and after the VBF and ggF category cuts. In order to facilitate comparison, the distributions for each mass point after applying the classification cut (blue) are scaled to the same area as the distributions before any classification cut (black). This is denoted by 'a.u.' on the ordinates.



(a) Distribution in $m_{4\ell}$ of the combined SM background before any classification cut (black), and in the VBF category (blue).



(b) Distribution in $m_{4\ell}$ of the combined SM background before any classification cut (black), and in the ggF category (blue).

Figure 5.21: (a) Distribution in $m_{4\ell}$ of the simulated, combined SM background before any classification cut (black), and in the VBF category (blue). The distributions are normalised to the same area. (b) Distribution in $m_{4\ell}$ of the combined SM background before any classification cut (black), and in the ggF category (blue). In order to facilitate comparison, the distributions after applying the classification cut (blue) are scaled to the same area as the distributions before any classification cut (black). This is denoted by 'a.u.' on the ordinates. The bottom panels show the ratio of the number of events after applying the ggF- or VBF-category cuts to the number of events before applying any classification cut (ratio ggF (VBF) cat./incl.).

Therefore, it is important to determine whether the classifier is biased towards values of $m_{4\ell}$ that it has been provided as a mass point during training. In order to assess this bias, the classifier is evaluated on a specially generated ggF signal mass point at 450 GeV, which it has not seen during training. A mass of 450 GeV was chosen because it is located exactly between two mass points used for training (400 and 500 GeV), and mid-way between the two masses where the previous publication [58] observed excesses (240 and 700 GeV). Furthermore, 450 GeV corresponds to a region in $m_{4\ell}$ that is still relatively high in SM background. In Figure 5.22, the efficiencies in four event categories (ggF $4e$, ggF 4μ, ggF $2\mu 2e$ and background) are interpolated using fifth-degree polynomial fits. In these fits, the 450 GeV mass point has been excluded. This allows for a comparison of the predicted efficiency at 450 GeV with the measured efficiency at that mass point. In order to quantify the deviation between the efficiencies at 450 GeV predicted by the interpolation curves and the measured ones, it is helpful to compare these with the standard deviation of the residuals (excluding 450 GeV). Table 5.8 compares the standard deviation of residuals, σ_r, with the residuals at the 450 GeV mass point, r_{450}.

category	σ_r	r_{450}
VBF	0.00073	0.00028
ggF $4e$	0.0023	0.0015
ggF 4μ	0.0020	0.0019
ggF $2e2\mu$	0.0067	0.0097
background	0.0095	0.0099

Table 5.8: Standard deviation of residuals and residuals at 450 GeV in the five event categories.

The residuals at 450 GeV all lie within, or very close to, one standard deviation of the remaining residuals, indicating that the classifier is not significantly biased towards values of $m_{4\ell}$ it has been provided as mass points during training.

5.8.12 Overall improvement

The overall improvement obtained using the multivariate categorisation over the cut-based categorisation is assessed by comparing expected upper limits on the production cross-section times branching fraction calculated using each of the two methods (Figure 5.23). The multivariate method provides an overall improvement on the VBF production cross-section, σ_{VBF}, of up to 25%. In terms of the ggF

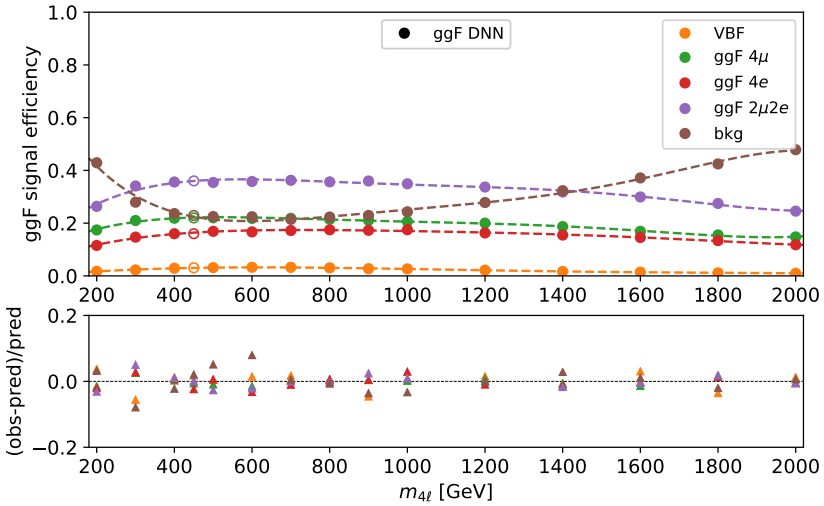


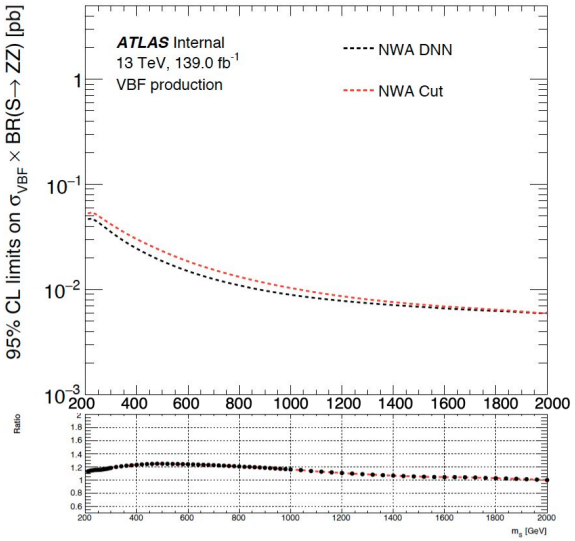
Figure 5.22: Efficiencies of four event categories (ggF ($4e$), ggF (4μ), ggF ($2\mu 2e$), and background) as a function of ggF signal mass. The dashed lines show fifth-degree polynomial fits. The open circles at $m_{4\ell} = 450$ GeV indicate the measured efficiencies on a ggF signal sample with $m_H = 450$ GeV, which was not available during training. The bottom panel shows the residuals between the efficiency measurements and their polynomial fits, normalised to the value of the fits ($^{(\text{obs-pred})}/_{\text{pred}}$).

production cross-section, σ_{ggF}, we observe an improvement of up to 40%. The lack of improvement in the high-mass region can be attributed to the fact that that regime is nearly background-free, leaving little room for improvement in sensitivity.

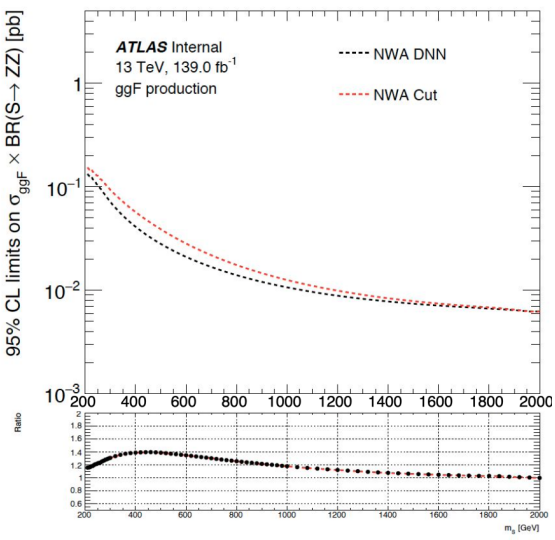
5.9 Statistical interpretation and systematic uncertainties

Limits on the production cross-section times branching fraction are set using a test statistic based on the profile likelihood ratio, $q(\alpha^i, \boldsymbol{\theta})$, where α_i are the parameters of interest, i.e. the production cross-section times branching fraction of a heavy resonance in the ggF and VBF production modes, and $\boldsymbol{\theta}$ are the nuisance parameters. The profile likelihood ratio is the probability of correctly rejecting the null hypothesis assuming that the opposite, i.e. the signal hypothesis, is true [174]. For a given signal or background hypothesis, $\mathcal{H}$, it is defined as

$$q(\alpha^i, \boldsymbol{\theta}) = -2 \ln \frac{\mathcal{L}(\alpha^i_{\mathcal{H}}, \hat{\boldsymbol{\theta}}_{\mathcal{H}})}{\mathcal{L}(\hat{\alpha}^i, \hat{\boldsymbol{\theta}})}, \tag{5.12}$$



(a) Upper limits at 95% CL on the cross-section times branching fraction as a function of the heavy Higgs mass m_H in the VBF production mode. The limits are compared between the cut-based analysis (red) and the DNN-based analysis (black).



(b) Upper limits at 95% CL on the cross-section times branching fraction as a function of the heavy Higgs mass m_H in the ggF production mode. The limits are compared between the cut-based analysis (red) and the DNN-based analysis (black).

Figure 5.23: Upper limits at 95% CL on the production cross-section times branching fraction of a heavy H as a function of m_H, compared between the cut-based and the DNN-based categorisation. The limits are shown separately for the VBF category (a), and the ggF category (b). The bottom panels show the ratio of the cross-section limits obtained using the cut-based categorisation (red) to the limits obtained using the DNN-based categorisation (black). Figures reproduced from Reference [1].

where $\alpha^i_{\mathcal{H}}$ are the parameters of interest under the tested hypothesis (with $\mathcal{H} = \mathcal{H}_0$ in the case of the background-only hypothesis), $\hat{\boldsymbol{\theta}}_{\mathcal{H}}$ is the best fit of the nuisance parameters under the tested hypothesis, $\hat{\alpha}^i$ is the best fit of the parameters of interest under the signal hypothesis, and $\hat{\boldsymbol{\theta}}$ is the best fit of the nuisance parameters under the signal hypothesis.

The likelihood function $\mathcal{L}$ is given by the product of a Poisson distribution, which represents the probability of observing n events under the null or signal hypothesis, and a weighted sum of signal and background probability distribution functions. In the case of the NWA signal hypothesis, and for an observation of n events, it is given by

$$\mathcal{L}(x_1, \ldots, x_n | \sigma_{\text{ggF}}, \sigma_{\text{VBF}}) = \text{Poisson}(n | S_{\text{ggF}} + S_{\text{VBF}} + B) \times \prod_{i=1}^{n} \frac{S_{\text{ggF}} f_{\text{ggF}}(x_i) + S_{\text{VBF}} f_{\text{VBF}}(x_i) + B f_{\text{B}}(x_i)}{S_{\text{ggF}} + S_{\text{VBF}} + B}, \quad (5.13)$$

where S_{ggF} is the number of ggF-produced signal events, S_{VBF} is the number of VBF-produced signal events, and B is the number of expected SM events. The functions f_x are determined from the signal and background modelling procedures described in Section 5.6. The probability density functions $f_{\text{ggF}}(x_i)$ and $f_{\text{VBF}}(x_i)$ give the ggF and VBF signal probability of observing the event x_i, whereas $f_B(x_i)$ is the value of the overall background probability distribution function for event x_i. The nuisance parameters $\boldsymbol{\theta}$ are estimates of the systematic uncertainties, and are constrained by normal distributions.

Upper limits on the production cross-section times branching fraction at a resonance mass value under test, m_H, are obtained, separately in the ggF and VBF categories, using the CL_s method [175, 176, 177]. It is an iterative process that involves defining two p-values for all hypotheses, $\mathcal{H}$, under test, each of them corresponding to a specific production cross-section times branching fraction of a heavy Higgs. The first, referred to as $p_{\mathcal{H}}$, corresponds to the probability that the value of q under the tested hypothesis is equal to or greater than the observed one, q_{obs}. The second one, referred to as p_B, is the probability that the value of q under the background-only hypothesis is equal to or greater than q_{obs}. One then defines

$$CL_S = \frac{p_{\mathcal{H}}}{1 - p_B}. \quad (5.14)$$

Finding the exclusion limit involves adjusting $\mathcal{H}$, and, with it, the implied cross-section times branching fraction of a heavy Higgs, until the desired value of CL_S

is reached. To quote a 95% CL_S limit, as is done in this analysis, $p_{\mathcal{H}}$ is adjusted until a target value of $CL_S = 0.05$ is reached. The exclusion limits as a function of m_H are obtained by repeating this procedure for all possible values of m_H within $200 < m_H < 2000\,\text{GeV}$. In each step, one fixes m_H in $\mathcal{L}$ (c.f. Equation (5.13), where m_H is contained implicitly in the probability density functions, f_x, and finds the nuisance parameters $\boldsymbol{\theta}$ that minimise $\mathcal{L}$.

5.9.1 Systematic uncertainties

The impact of systematic uncertainties is discussed in detail in References [1, 3, 178], and only a brief overview is given here. The degree to which a systematic uncertainty impacts the results depends on the final state, the production mode, as well as the resonance mass. In the $\ell^+\ell^-\ell'^+\ell'^-$ final state at lower masses, uncertainties associated with the parametrisation of the ZZ continuum background (Section 5.6.2) are dominant in the ggF production mode, while at higher masses, uncertainties from parton showering become the most relevant ones. In the $\ell^+\ell^-\ell'^+\ell'^-$ final state in the VBF production mode, the dominant systematic uncertainty is the jet flavour composition, while at higher masses the most important one comes from higher-order QCD corrections to the Higgs production cross-section. Table 5.9 shows the four leading systematic uncertainties at four hypothesised values of m_H for the ggF and VBF production modes. Above a mass of $m_H = 600\,\text{GeV}$, the systematic uncertainties in the $\ell^+\ell^-\nu\bar{\nu}$ final state are dominant; hence the systematic uncertainties in the $\ell^+\ell^-\ell'^+\ell'^-$ final state are not shown. The leading systematic uncertainty in the $\ell^+\ell^-\ell'^+\ell'^-$ final state in this region has an impact on the final yield of 3% or less.

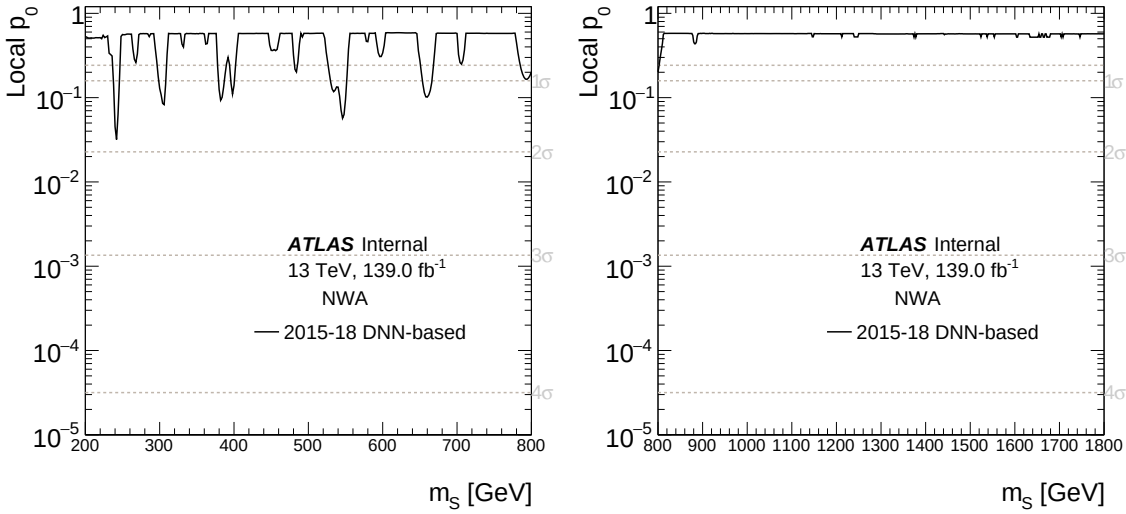
5.10 Results

In order to enhance the sensitivity of the overall results, the $\ell^+\ell^-\ell'^+\ell'^-$ analysis is combined with the $\ell^+\ell^-\nu\bar{\nu}$ channel. Full details on the $\ell^+\ell^-\nu\bar{\nu}$ analysis can be found in References [3] and [178]. No significant excess in the $\ell^+\ell^-\ell'^+\ell'^-$ or $\ell^+\ell^-\nu\bar{\nu}$ was found, and consequently, limits are placed on the production cross-section times branching fraction of the different signal hypotheses. The total number of observed events in the $\ell^+\ell^-\ell'^+\ell'^-$ final state is 3275, while 2794 events are observed in the $\ell^+\ell^-\nu\bar{\nu}$ final state. The deviation with the highest significance found in the $\ell^+\ell^-\ell'^+\ell'^-$ channel occurred at 240 GeV with an observed local significance of $1.9\,\sigma$, rejecting the excess observed previously in the $36\,\text{fb}^{-1}$ publication. The values of

ggF production		VBF production	
Systematic source	Impact [%]	Systematic source	Impact [%]
$m_H = 300\,\text{GeV}$			
ZZ parameterisation ($\ell^+\ell^-\ell'^+\ell'^-$)	4.5	Jet flavor composition	3.0
Z + jets modelling ($\ell^+\ell^-\nu\bar{\nu}$)	2.3	$q\bar{q} \to ZZ$ QCD scale (VBF-enriched category, $\ell^+\ell^-\ell'^+\ell'^-$)	2.8
Parton showering of ggF ($\ell^+\ell^-\ell'^+\ell'^-$)	2.2	ZZ parameterisation ($\ell^+\ell^-\ell'^+\ell'^-$)	2.3
$e\mu$ statistical uncertainty $\ell^+\ell^-\nu\bar{\nu}$	2.0	Jet energy scale (*in-su* calibration)	1.8
Data stat. uncertainty	53	Data stat. uncertainty	58
Total uncertainty	55	Total uncertainty	60
$m_H = 600\,\text{GeV}$			
Electroweak corrections for $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	4.9	QCD scale of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	7.6
QCD scale of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	2.5	Jet energy resolution	5.4
Z + jets modelling ($\ell^+\ell^-\nu\bar{\nu}$)	2.5	Parton showering ($\ell^+\ell^-\nu\bar{\nu}$)	3.3
PDF of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\ell'^+\ell'^-$)	2.2	Electroweak corrections for $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	3.0
Data stat. uncertainty	54	Data stat. uncertainty	61
Total uncertainty	57	Total uncertainty	63
$m_H = 1000\,\text{GeV}$			
Electroweak corrections for $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	9.3	Parton showering ($\ell^+\ell^-\nu\bar{\nu}$)	6.8
Parton showering ($\ell^+\ell^-\nu\bar{\nu}$)	5.2	Electroweak corrections for $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	4.7
QCD scale of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	4.8	QCD scale of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	2.4
Z + jets modelling ($\ell^+\ell^-\nu\bar{\nu}$)	2.4	Jet flavor composition	2.4
Data stat. uncertainty	57	Data stat. uncertainty	58
Total uncertainty	59	Total uncertainty	59
$m_H = 1500\,\text{GeV}$			
Parton showering ($\ell^+\ell^-\nu\bar{\nu}$)	9.6	Parton showering ($\ell^+\ell^-\nu\bar{\nu}$)	9.0
Electroweak corrections for $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	6.8	Electroweak corrections for $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	4.6
PDF of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	5.4	PDF of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	3.4
QCD scale of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	4.6	QCD scale of $q\bar{q} \to ZZ$ ($\ell^+\ell^-\nu\bar{\nu}$)	2.8
Data stat. uncertainty	57	Data stat. uncertainty	55
Total uncertainty	59	Total uncertainty	57

Table 5.9: Impact of the leading systematic uncertainties, the data statistical uncertainties, and the total uncertainties on the predicted signal event yield with the cross-section times branching fraction being set to the expected upper limit, expressed as a percentage of the signal yield for the ggF (left) and VBF (right) production modes at m_H = 300, 600, 1000, and 1500 GeV. Table reproduced from Reference [3].

the local p_0—corresponding to a rejection of the background-only hypothesis—as a function of signal mass in the NWA is presented in Figure 5.24. The previous excess at 700 GeV all but disappears. The second-largest observed excess occurs at 540 GeV with a local significance of 1.5 σ.



(a) Local p_0 in the NWA in the mass range 200–800 GeV.

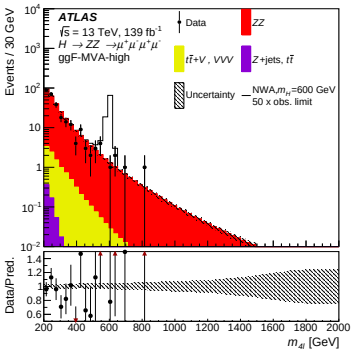
(b) Local p_0 in the NWA in the mass range 800–1800 GeV.

Figure 5.24: Local p_0 as a function of scalar-resonance mass, m_S, in the NWA. All five event categories are combined. Plot (a) shows the local p_0 in the mass range 200–800 GeV, while plot (b) shows the local p_0 in the mass range 800–1800 GeV. Figures reproduced from Reference [1].

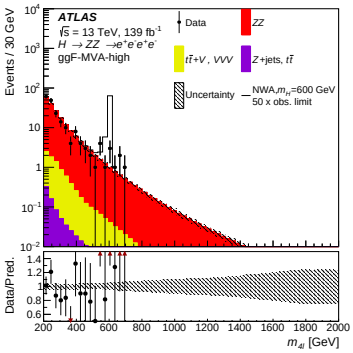
The $m_{4\ell}$ spectra in each of the five event categories are shown in Figure 5.25.

5.10.1 Scalar resonance in the NWA

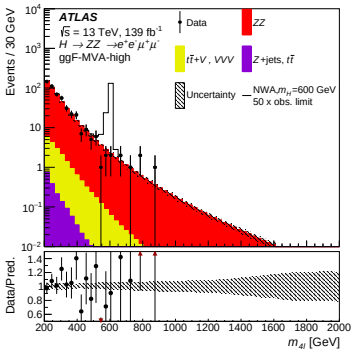
Upper limits on the production cross-section times branching fraction, $\sigma \times B(H \to ZZ)$, for a heavy resonance are computed using the CL_S method described in Section 5.9. The limits are calculated separately for the two presumably dominant production mechanisms, ggF and VBF. When setting limits on one, the other parameter is allowed to float freely in the fit. The expected and observed limits are shown in Figure 5.26, combining results from both the $\ell^+\ell^-\ell'^+\ell'^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states. In the ggF production mode, the observed limits on $\sigma_{\text{ggF}} \times B(H \to ZZ)$ range from 215 fb at 240 GeV to around 2.0 fb at 1900 GeV. In the VBF production mode, the observed limits on $\sigma_{\text{VBF}} \times B(H \to ZZ)$ range from 87 fb at 255 GeV to 1.5 fb at 1800 GeV.



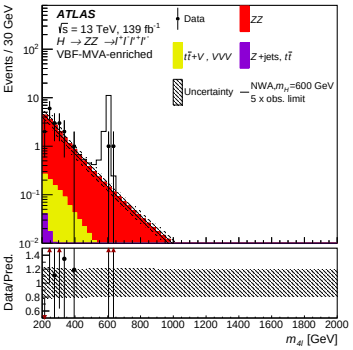
(a) Distribution of $m_{4\ell}$ in the ggF (4μ) category.



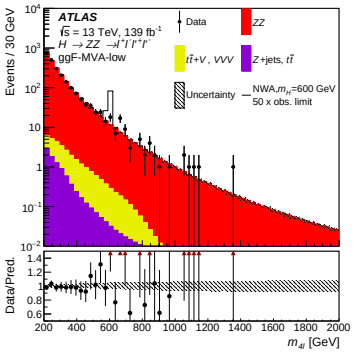
(b) Distribution of $m_{4\ell}$ in the ggF ($4e$) category.



(c) Distribution of $m_{4\ell}$ in the ggF ($2e2\mu$) category.

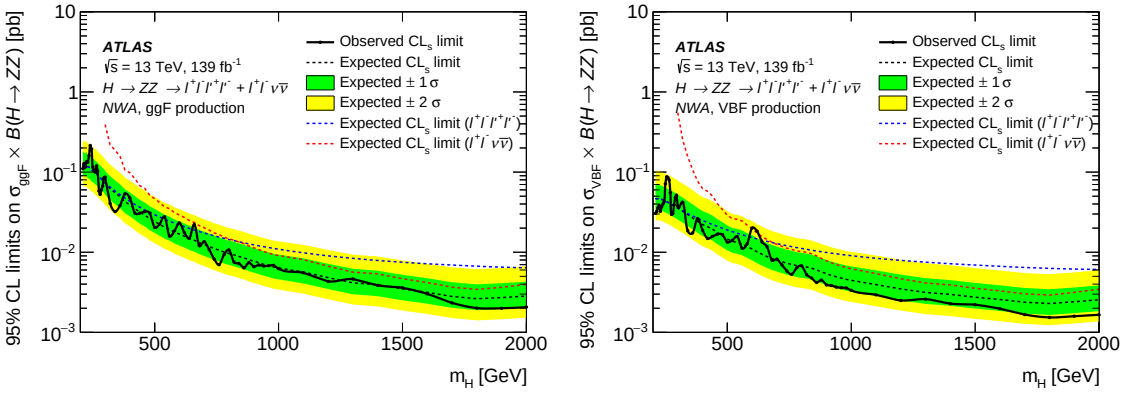


(d) Distribution of $m_{4\ell}$ in the VBF category.



(e) Distribution of $m_{4\ell}$ in the background category.

Figure 5.25: Distribution of $m_{4\ell}$ in the five event categories: (a) ggF (4μ), (b) ggF (4μ), (c) ggF (4μ), (d) VBF, and (e) background (labelled 'ggF-MVA-low'). The background contributions are determined from a combined likelihood fit to the data under the background-only hypothesis. For illustrative purposes, the plots include a simulated heavy-Higgs signal with a mass of $m_H = 600\,\text{GeV}$, which has been scaled to a cross-section corresponding to 50 times the observed limit in the ggF production mode (a, b, c, and e). In the VBF production mode, the signal cross-section has been scaled to five times the observed limit (d). The lower panels show the ratio of the data to the simulated SM prediction. The red arrows in the lower panels indicate data points that are outside the displayed range. Figure reproduced from [3].



(a) Limits on $\sigma_{\text{ggF}} \times B(H \to ZZ)$ in the NWA.

(b) Limits on $\sigma_{\text{VBF}} \times B(H \to ZZ)$ in the NWA.

Figure 5.26: Upper limits at 95% CL on the production cross-section times branching fraction of a heavy scalar resonance with mass m_H in the NWA. The upper limits are shown separately for the ggF production mode (a) and the VBF production mode (b). The observed limits are shown as black solid lines, while the expected limits are shown as black dashed lines. The blue dashed lines show the expected limits in the $\ell^+\ell^-\ell'^+\ell'^-$ final state, while the red dashed lines show the expected limits in the $\ell^+\ell^-\nu\bar{\nu}$ final state. The green bands represent a deviation from the expected limit of $\pm 1\,\sigma$, while the yellow bands show a deviation of $\pm 2\,\sigma$. Figures reproduced from Reference [3].

5.10.2 Scalar resonance interpreted in the 2HDM

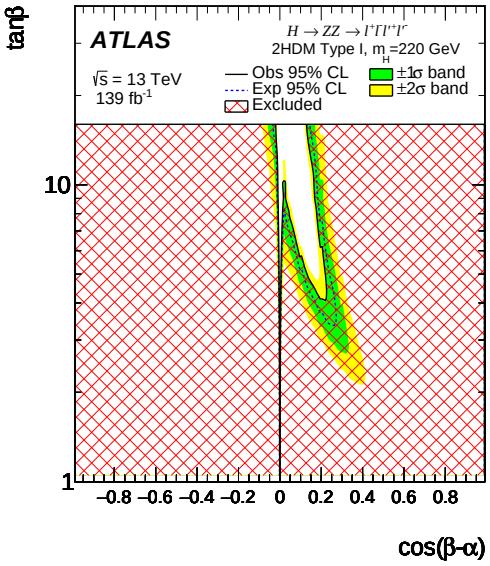
The upper limits on the production cross-section times branching fraction of a heavy scalar in the NWA are translated into exclusion contours in the context of a CP-conserving 2HDM introduced in Section 2.2.1. While direct couplings to leptons are possible in other types of 2HDMs, in the context of the $H \to ZZ$ process, only Type I, in which Φ_2 couples to all quarks and leptons, and Type II, in which Φ_1 couples to down-type quarks and leptons and Φ_2 couples to up-type quarks, are relevant. The coupling of the heaviest CP-even Higgs boson to W and Z bosons is proportional to $\cos(\beta - \alpha)$, and in the limit of $\cos(\beta - \alpha) \to 0$, the light CP-even Higgs boson would be indistinguishable from the SM Higgs boson. Furthermore, it is assumed that the mass hierarchy between the Higgs bosons is small enough so that the heavy CP-even Higgs boson does not decay to other Higgs bosons.

The exclusion limits in the parameter space spanned by $\cos(\beta - \alpha)$ and $\tan\beta$ as well as the space $\tan\beta$ vs. m_H are presented in Figure 5.27 for the Type-I and Type-II scenarios. In both scenarios, the couplings of the heavier CP-even Higgs to vector bosons are proportional to $\cos(\beta - \alpha)$. In the projection of $\cos(\beta - \alpha)$ vs. $\tan\beta$, m_H is fixed to a value of 220 GeV. The choice of m_H is such that the NWA is valid over most of the parameter space (i.e. β and α), and the experimental sensitivity of this analysis is maximal. In the complementary projection of $\tan\beta$ vs. m_H, the value of $\cos(\beta - \alpha)$ is fixed to -0.1, which is chosen so that (i) it is non-zero, and therefore compatible with a non-zero coupling to Z bosons, and (ii) the properties of the light Higgs boson in the 2HDM are still compatible with the recently measured couplings of the SM Higgs boson and the resulting excluded regions of 2HDM parameter space [179]. These results significantly improve on those obtained in the previous iteration of the analysis using 36 fb^{-1} of data [58]. For example, the region excluded in the Type-II scenario in the space spanned by $\tan\beta$ and m_H is 60% larger for $200 < m_H < 400$ GeV than before.

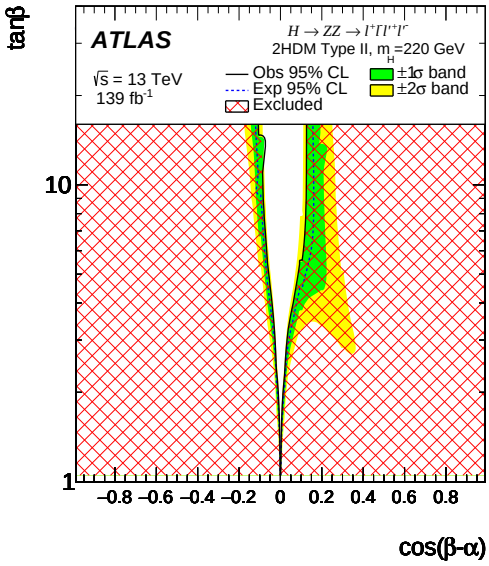
5.10.3 Scalar resonance in the LWA

Upper limits on the production cross-section times branching fraction are set for a scalar resonance in the ggF production mode in the LWA, as shown in Figure 5.26. The upper limits are computed separately for four natural widths[6]: 1%, 5%, 10%, and 15% of m_H. The observed limits are very similar across the four studies, ranging from

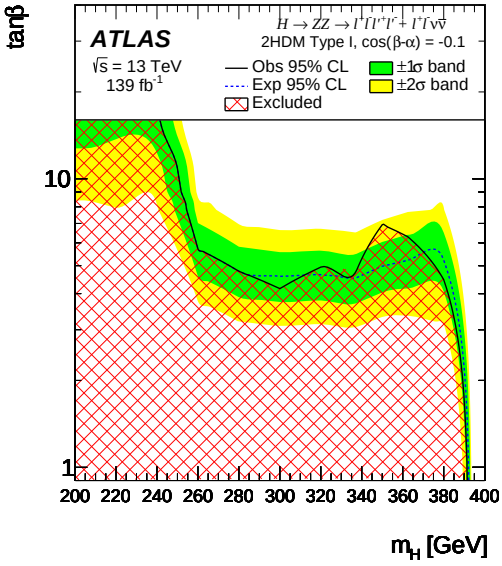
[6]These four widths are a standard configuration used in ATLAS searches for BSM scalars, allowing for an easier combination of results from different analyses.



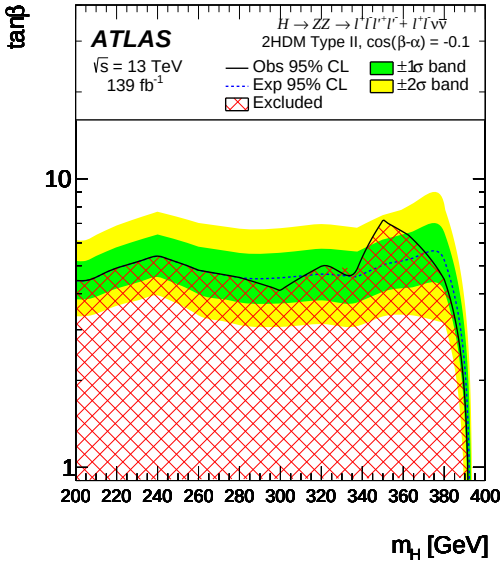
(a) Excluded region in the parameter space $\tan\beta$ vs. $\cos(\beta-\alpha)$ with m_H = 220 GeV in the Type-I 2HDM.



(b) Excluded region in the parameter space $\tan\beta$ vs. $\cos(\beta-\alpha)$ with m_H = 220 GeV in the Type-II 2HDM.



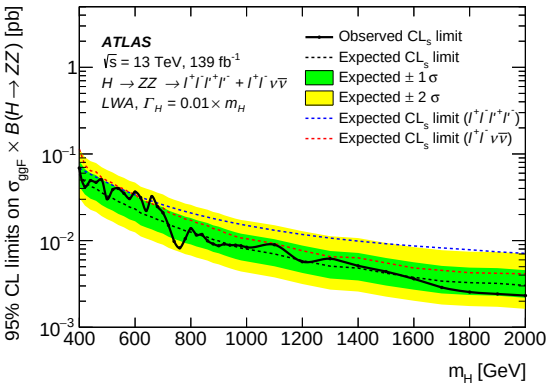
(c) Excluded region in the parameter space $\tan\beta$ vs. m_H with $\cos(\beta-\alpha)$ = −0.1 in the Type-I 2HDM.



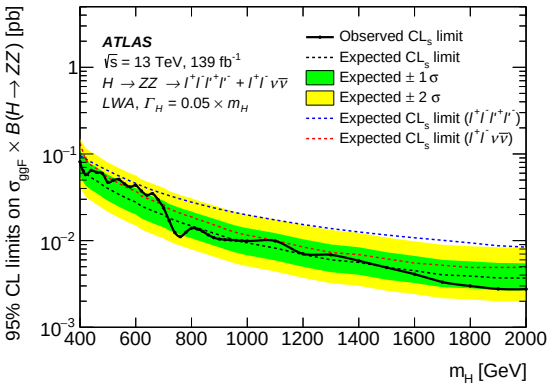
(d) Excluded region in the parameter space $\tan\beta$ vs. m_H with $\cos(\beta-\alpha)$ = −0.1 in the Type-II 2HDM.

Figure 5.27: The exclusion contours in the 2HDM Type-I (a) and Type-II (b) models for m_H = 220 GeV shown as a function of the parameters $\cos(\beta-\alpha)$ and $\tan\beta$. The Type-I and Type-II models as a function of the parameters $\tan\beta$ and m_H with $\cos(\beta-\alpha)$ is shown in panels (c) and (d), respectively. The green and yellow bands represent deviations from the expected limits of ±1 σ and ±2 σ, respectively. The hatched areas show the observed excluded regions. Figures reproduced from Reference [3].

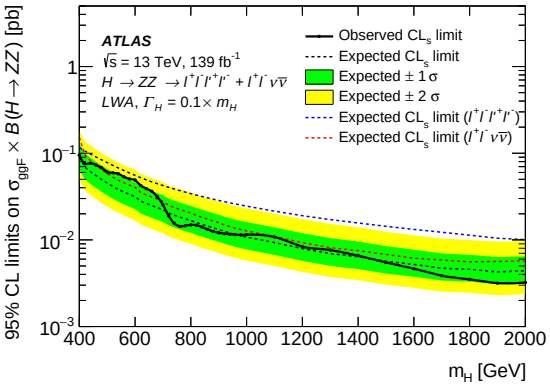
~ 100 fb at 400 GeV to ~ 2 fb at 2000 GeV. The cut-based categorisation (Section 5.7) is used to define the ggF-enriched category.



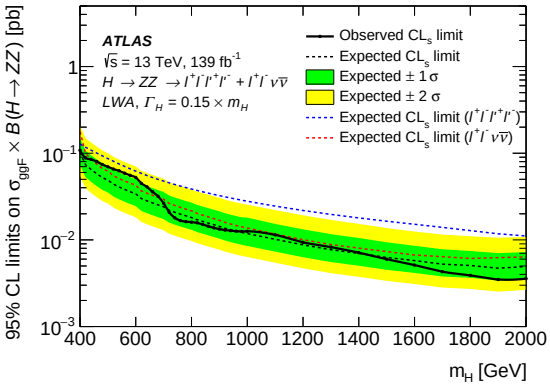
(a) Limits on $\sigma_{\text{ggF}} \times B(H \to ZZ)$ in the LWA with $\Gamma_H = 0.01\ m_H$.



(b) Limits on $\sigma_{\text{ggF}} \times B(H \to ZZ)$ in the LWA with $\Gamma_H = 0.05\ m_H$.



(c) Limits on $\sigma_{\text{ggF}} \times B(H \to ZZ)$ in the LWA with $\Gamma_H = 0.1\ m_H$.



(d) Limits on $\sigma_{\text{ggF}} \times B(H \to ZZ)$ in the LWA with $\Gamma_H = 0.15\ m_H$.

Figure 5.28: Upper limits at 95% CL on the production cross-section times branching fraction of a heavy scalar resonance with mass m_H in the large-width assumption (LWA). The upper limits are shown in the ggF production mode for different natural widths of the resonance: 1% (a), 5% (b), 10% (c), and 15% (d). The observed limits are shown as solid black lines, while the expected limits are shown as dashed black lines. The dashed blue lines show the expected limits in the $\ell^+\ell^-\ell'^+\ell'^-$ final state, while the dashed red lines show the expected limits in the $\ell^+\ell^-\nu\bar{\nu}$ final state. The green bands represent a deviation from the expected limit of $\pm 1\,\sigma$, while the yellow bands correspond to a deviation of $\pm 2\,\sigma$. Figures reproduced from Reference [3].

5.10.4 Spin-2 graviton

The results are interpreted additionally as a search for a KK excitation in the context of the bulk RS model with $k/\bar{M}_{\text{Planck}} = 1$. The expected and observed limits, as well

as the predicted cross-section, are shown in Figure 5.29. The cut-based categorisation (Section 5.7) is used to define the ggF-enriched category. The results exclude a spin-2 graviton up to a mass of 1830 GeV at 95% CL.

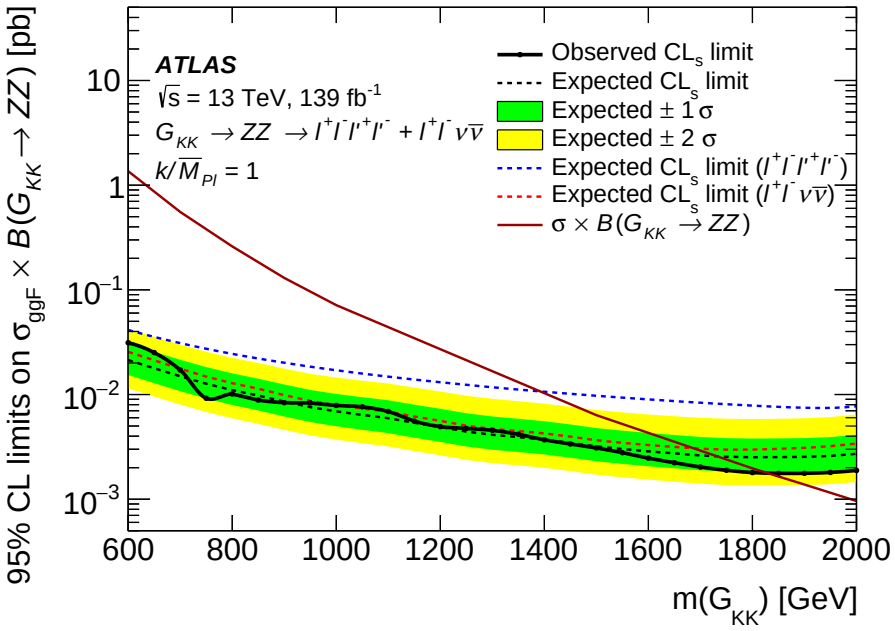


Figure 5.29: Upper limits at 95% CL on the production cross-section times branching fraction of a graviton with mass $m_{G_{kk}}$ in the ggF production mode with $k/\bar{M}_{\text{Planck}} = 1$. The observed limits are shown as black solid lines, while the expected limits are shown as black dashed line. The blue dashed lines show the expected limits in the $\ell^+\ell^-\ell'^+\ell'^-$ final state, while the red dashed lines show the expected limits in the $\ell^+\ell^-\nu\bar{\nu}$ final state. The green bands represent a deviation from the expected limit of $\pm 1\,\sigma$, while the yellow bands correspond to a deviation of $\pm 2\,\sigma$. The red solid line shows the predicted cross-section times branching fraction within the context of the bulk RS model. Figure reproduced from Reference [3].

5.11 Conclusion

A search for heavy ZZ resonances decaying to $\ell^+\ell^-\ell'^+\ell'^-$ has been presented, focusing specifically on a multivariate classification method to categorise events as either ggF-like, VBF-like, or background-like. Two classifiers were described, each based on a combination of RNNs and MLPs.

To facilitate a training of the classifiers in a way that is unbiased towards the four-lepton reconstructed mass, the signal samples used for training were reweighted so as to follow the smoothly falling ZZ continuum background. Studies were

presented that demonstrate the classifier's ability to (i) perform well at intermediate masses that were not available during training, and, (ii) perform well in the mass range that lies beyond the cut-off value of $m_{4\ell} = 1300\,\text{GeV}$ at which events were excluded from the training. It was further shown that cutting on the classifiers does not bias the background $m_{4\ell}$ distribution. After selecting optimal cut values on the classifiers using a discovery-significance estimation, the sensitivity in terms of upper limits on the production cross-section times branching fraction of VBF and ggF resonances was presented, showing improved upper limits on the production cross-section of up to 25% (VBF) and 40% (ggF) compared with the previously used cut-based categorisation.

The search results in the $\ell^+\ell^-\ell'^+\ell'^-$ final state were combined with the $\ell^+\ell^-\nu\bar{\nu}$ final state, and limits on the production cross-section times branching fraction were set for heavy Higgs-like scalars in the NWA and in the LWA. The NWA limits were used to expand the excluded parameter space in the context of a Type-I and Type-II 2HDM. Furthermore, limits on the production cross-section times branching fraction of a spin-2 graviton within the context of the bulk RS model were set, excluding KK excitations with a mass of $m_{G_{\text{KK}}} < 1830\,\text{GeV}$ at 95% CL.

Chapter 6

Calorimeter-Tagging Muons Using Machine Learning

Identifying muons efficiently and measuring their momenta accurately is central to many ATLAS analyses. Several recent physics results would not have been possible without a highly performant ATLAS muon system. Some examples of physics analyses in which muons played an important role are the discovery of the Higgs boson in 2012 [180], where efficient muon identification was important in the $\ell^+\ell^-\ell'^+\ell'^-$ decay channel, as well as BSM searches for high-mass dilepton resonances [181], or the high-mass $\ell^+\ell^-\ell'^+\ell'^-$ search presented in Chapter 5.

ATLAS distinguishes between different muon types that depend on which parts of the detector are involved in their reconstruction. This work focuses on a specific type called *calorimeter-tagged* muons, i.e. muons with a track in the inner detector, but without a matching track in the muon spectrometer (see Section 3.2.5). In this chapter, a novel method for calorimeter-tagging muons in the ATLAS detector called `CaloMuonScore` is developed. This method aims to distinguish muons that have a track in the inner detector from other particles based on the energy-loss patterns they create in the ATLAS electromagnetic and hadronic calorimeters. Prior to the introduction of this method, ATLAS used a cut-based approach in the identification of such muons; `CaloMuonScore` improves on this approach by taking into account a greater amount of low-level detector information.

Muons that originate from the primary interaction vertex are referred to as *prompt* muons; those originating from the decay of light hadrons (such as $\pi^\pm$ and $K^\pm$) are referred to as *non-prompt* muons. While these light-hadron decays produce true muons, they can in principle be distinguished from prompt muons based on the presence of larger energy deposits in the calorimeters. `CaloMuonScore` is

a binary classifier, and was, consequently, developed with prompt muons as the *signal* component and non-prompt muons as the *background* component. Among the light hadrons, pions are the most abundant, and serve, therefore, as the source of non-prompt muons during development and training of the algorithm. In addition, kaons are explored as another source of background. It will be shown that while the `CaloMuonScore` algorithm was trained on pions as the background component, it performs equally well when discriminating prompt muons from kaon background.

This chapter first discusses the existing cut-based and a newly introduced feature-based approach to calorimeter-tagging muons. This is followed by the description of a novel method for calorimeter-tagging muons (`CaloMuonScore`) based on a machine-learning approach using convolutional neural networks. The chapter concludes with a tag-and-probe analysis, in which the performance of the novel method is validated on collision data.

6.1 Muon efficiency measurement

In the case of MC simulation, the efficiency with which a given algorithm identifies muons is easily determined since the associated truth-level particle is generally known. Here, the reconstruction efficiency is simply given by

$$\epsilon_{\mu}^{\mathrm{MC}} = \frac{N_{\mu}^{\mathrm{reco}}}{N_{\mu}^{\mathrm{truth}}}, \tag{6.1}$$

where N_{μ}^{reco} is the number of muons reconstructed by the reconstruction algorithm under study, and N_{μ}^{truth} is the number of simulated muons within the acceptance region of the algorithm.

In the case of data, the true particle type is unknown and has to be inferred from the overall event topology. One way of measuring the muon reconstruction efficiency on recorded collision data is the so-called *tag-and-probe method* [84], and it provides an important, complementary measurement to the efficiency derived based on simulation only.

6.1.1 The tag-and-probe method

The tag-and-probe method is based on selecting an almost pure sample of $Z \to \mu\mu$ events; this is achieved by selecting candidate events with one well-reconstructed muon, the *tag*, and another muon candidate with opposite charge, the *probe*. The two muons are required to have an invariant mass compatible with the Z boson mass, meaning that the reconstructed pair very likely originate from a $Z \to \mu\mu$ decay. No specific muon selection criteria are applied to the probe, making it an unbiased tool for studying the muon reconstruction efficiency. Applying the tag-and-probe method to MC simulation instead of data can be used to extract an *efficiency scale factor*. Quantifying the deviation of simulation from data, it is defined as the ratio of the muon reconstruction efficiency measured on data to the one measured on simulation, and is used in physics analyses to correct simulation-derived event yields [84].

6.1.1.1 The tag-and-probe method with $Z \to \mu\mu$ events

The method requires the invariant mass of the di-muon pair to be close to the Z-boson mass ($81 < m_{\mu\mu} < 101$ GeV). The tag muon is required to have triggered the

event readout. Further, it needs to fulfil the *loose*[1] isolation and *medium* identification criteria, which comprise the default muon selection in ATLAS. Medium muons are either combined or extrapolated muons (c.f. Section 3.2.5) [84]. Aimed at measuring the calorimeter-tagging efficiency, this study uses probes that are identified as muons in the MS (so-called *MS probes*). In the low-η region ($|\eta| < 0.1$), where the MS is not fully instrumented, the study uses probes that are based on tracks in the inner detector (so-called *ID probes*).

A further requirement on the tag is a p_{T} of at least 24 GeV (this is the single-muon-trigger threshold), and a transverse impact parameter significance of at most three. In addition, the longitudinal impact parameter must satisfy $|z_0| < 10$ mm, and the tag muon must have triggered the event readout. The requirements on the probe muon are $p_{\mathrm{T}} > 10$ GeV and a loose isolation criterion.

6.1.1.2 Measuring the efficiency of calorimeter-tagged muons using the tag-and-probe method

Assume that the efficiency of a muon-reconstruction algorithm, X, is to be measured. X is a method involving ID tracks as well as calorimeter information. The efficiency of this reconstruction method is then given by

$$\epsilon(X) = \frac{N_{\mathrm{X|MS}}}{N_{\mathrm{MS}}}, \tag{6.2}$$

where $N_{\mathrm{X|MS}}$ is the number of muons passing the selection criterion defined by X given that they are reconstructed as muons in the MS, and N_{MS} is the number of muons reconstructed in the MS.

The scale factor relevant to physics analyses that use calorimeter-tagged muons using method X is given by

$$\mathrm{SF}_X = \frac{\epsilon(X)_{\mathrm{data}}}{\epsilon(X)_{\mathrm{MC}}}, \tag{6.3}$$

[1]The *loose* isolation requirement places simultaneous cuts on two discriminating variables such that they achieve 99% muon efficiency that is constant in η and p_{T}, as measured on data in a $Z \to \mu\mu$ tag-and-probe analysis [84]. The two discriminating variables are $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}}^{\mu}$ and $E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}}^{\mu}$. Here, p_{T}^{μ} is the muon's transverse momentum, and $p_{\mathrm{T}}^{\mathrm{varcone30}}$ is a track-based isolation variable, defined as the sum of the magnitudes of the transverse momenta of tracks with $p_{\mathrm{T}} > 1$ GeV within a cone of size $\Delta R = \min(10\ \mathrm{GeV}/p_{\mathrm{T}}^{\mu}, 0.3)$. $E_{\mathrm{T}}^{\mathrm{topocone20}}$ denotes the sum of the transverse energy of all calorimeter topological clusters within a cone of size $\Delta R = 0.2$ surrounding the muon, excluding the energy deposit from the muon itself as well as pile-up contributions.

where $\epsilon(X)_{\text{data}}$ denotes the calculation of the reconstruction efficiency given in Equation 6.2 applied to data, and $\epsilon(X)_{\text{MC}}$ is the same efficiency measured on simulation.

A small fraction of tag-and-probe pairs originates from events other than $Z \to \mu\mu$ decays (about 0.1%). This small contribution to the $Z \to \mu\mu$ background is not considered in this work. It has to be estimated precisely and subtracted for high-precision measurements of the muon-reconstruction efficiency using dedicated simulation samples. Details of this approach can be found in Reference [84].

6.2 Calorimeter-tagging muons using CaloMuonTag

The algorithm for identifying calorimeter-tagged muons prior to this work is known as CaloMuonTag. The workings of this algorithm are described extensively in [182], and only a brief overview is given here. CaloMuonTag works on the basis of tracks in the inner detector, which satisfy the track-preselection criteria outlined in Section 6.2.1. The inner detector tracks are then extrapolated to the electromagnetic and hadronic calorimeters, and the energy deposits in the calorimeter cells traversed are recorded.

6.2.1 Track preselection

Particle tracks in the inner detector need to fulfil isolation criteria and are selected to ensure that they originate from the primary interaction vertex. The preselection of tracks is described in detail in Reference [182], and here only a brief description is given. Tracks are selected if their transverse momentum, $p_{\text{T}}^{\text{track}}$, satisfies $p_{\text{T}}^{\text{track}} > 2$ GeV. The requirement that the track originate from the primary interaction vertex is such that tracks need to have at least two hits in the pixel detector and six hits in the SCT subdetector. In addition, tracks are required to satisfy $d_0 < 1.2$ mm and $d_0/\sigma_{d_0} < 7$ in order to eliminate tracks that originate from FSR.

Two additional requirements are aimed at ensuring that the candidate track is sufficiently isolated: The first such requirement on the tracks is $\log_{10}(p_{\text{T}}^{\text{iso}}/p_{\text{T}}^{\text{track}}) < 0.7$, where $p_{\text{T}}^{\text{iso}}$ is the sum of the p_{T} of the ID tracks within $\Delta R < 0.45$ of the candidate track. Finally, a requirement is placed on the isolation of the track when extrapolated to the calorimeters: The energy of all calorimeter cells within a cone of $\Delta R < 0.45$ around the candidate track, $E_{\text{T}}^{\text{iso}}$, is required to satisfy $\log_{10}(E_{\text{T}}^{\text{iso}}/p_{\text{T}}^{\text{track}}) < 0.4$ [182].

6.2.2 Muon identification

A muon candidate receives a calorimeter tag if the energy deposits are consistent with a minimum ionising particle. Selection cuts are applied to these calorimeter deposits in order to filter out electronic noise. `CaloMuonTag` determines the energy threshold, E_{th}, above which muons are calorimeter-tagged as a function of the scattering angle, according to

$$E_{\text{th}} = \frac{50\ \text{MeV}}{\sin^2\theta}, \tag{6.4}$$

where θ is the track scattering angle, i.e. the polar angle measured between the beam line along positive z and the muon ID track. The functional form of Equation (6.4) stems from an empirical approximation to the distribution of a muon's deposited energy in a calorimeter layer, which increases with its path length. 50 MeV is chosen to ensure a constant tagging efficiency for a large range in θ [182].

Table 6.1 summarises the muon efficiency, ϵ_{sig}, as well as the fake rate, ϵ_{bkg}, estimated using a sample of truth-matched muons and truth-matched pions in the range $|\eta| < 0.1$ (the range for which the method is optimised).

ϵ_{sig}	ϵ_{bkg}
87.3%	1.5%

Table 6.1: Muon efficiency (ϵ_{sig}) and fake rate (ϵ_{bkg}) of the `CaloMuonTag` algorithm.

6.3 Training data generation

The data used for both the evaluation of the existing `CaloMuonTag` approach and the training of the machine-learning algorithms presented in this chapter originate from two samples of MC-generated $Z \to \mu\mu$ and $t\bar{t}$ events. $Z \to \mu\mu$ events were chosen as a source of prompt muons from Z decays, while the $t\bar{t}$ sample was chosen to provide non-prompt muons. The $t\bar{t}$ sample is a common choice for studies on muon reconstruction, identification, and isolation in ATLAS [84, 85], as it contains muons of several origins, such as prompt muons from $W^{\pm}$ decays, non-prompt muons from light-hadron decays, and non-isolated, non-prompt muons from b-jets. While pions originating from t decays are kinematically different from muons originating from Z decays, as will be seen in Section 6.3.3, this is not expected to affect the results of

this study as the performance of the algorithms developed here—and usually that of any muon-reconstruction algorithm—is generally considered as a function of p_{T} and η.

Some of the algorithms that will be considered aim to identify muons and distinguish them from other particles, and therefore need to interpret the detector readout at a point that occurs prior to the ordinary physics object reconstruction. More precisely, the algorithms are designed to work at the level of *track particles*, which are objects that are associated with a track in the inner detector. They are a set of hits that the ATLAS track-fitting procedure [183] has identified as consistent with a charged particle. Track particles associated with a true muon are extracted from the $Z \to \mu\mu$ sample and used as the muon, or signal, component for training, whereas track particles associated with a true pion are extracted from the $t\bar{t}$ sample and are used as the pion, or background, component. The MC samples are processed in *RAW Data Object* (RDO) format, which means at this stage the sample contains information as it is read out from the various detector components. Samples in RDO format do not yet contain reconstructed physics objects, and in general, the information is at the level of 'amount of energy collected in each calorimeter cell' and 'trajectory of a track in the inner detector'. This level of derivation is ideal for the problem at hand, in which the interest lies in information at the level of calorimeter cells. Table 6.2 provides more details on the two selected samples.

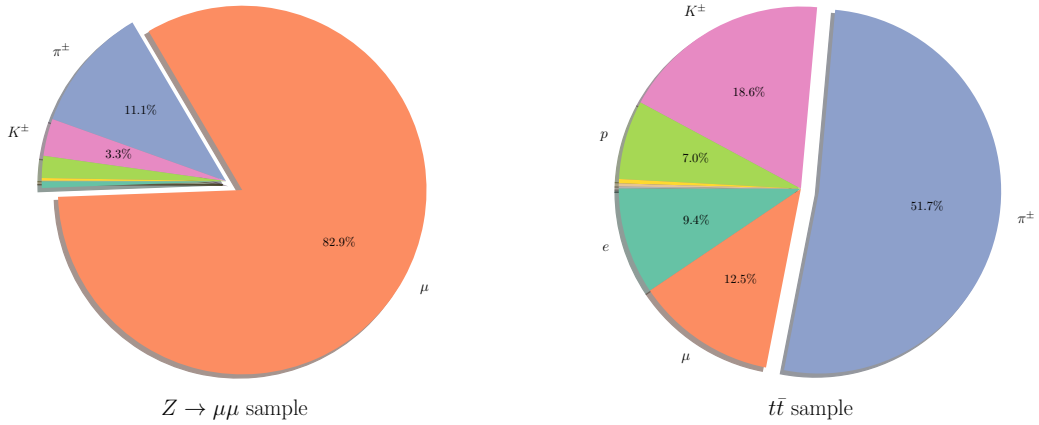
label	process	sample ID	generator	N_{events}
signal	$Z \to \mu\mu$	361107	`Powheg + Pythia`	100k
background	$t\bar{t}$	410470	`Powheg + Pythia`	1M

Table 6.2: Monte Carlo samples used for the generation of training data.

6.3.1 Truth matching

Track particles that are considered to be used for training and evaluation data are associated with truth-level particles in order to confirm that the track does indeed

originate from a truth-level muon or pion[2][3]. The remaining particles, such as true muons from t decays in the $t\bar{t}$ sample, are not used; they could, however, in principle be used to enhance the available training statistics in future iterations of this work. The distributions of truth-level particles among all track particles in the $Z \to \mu\mu$ and $t\bar{t}$ samples are summarised in Figure 6.1.



(a) Truth-level particles matched to ID tracks in the $Z \to \mu\mu$ sample.

(b) Truth-level particles matched to ID tracks in the $t\bar{t}$ sample.

Figure 6.1: Pie charts showing the distribution of truth-level particles matched to ID tracks. The (a) $Z \to \mu\mu$ and (b) $t\bar{t}$ samples were used for the generation of training data. Only muons from the $Z \to \mu\mu$ sample were used for the signal component of the training sample. Conversely, only pions from the $t\bar{t}$ sample were used for the background component of the training sample.

6.3.2 Calorimeter-layer naming convention

The MC samples in RDO format contain energy-deposit information at the level of individual cells. Each of these cells is mapped to a particular part of the ATLAS calorimeter system. This part is called *calorimeter sampling*. Calorimeter-tagged muons are generally used at very low η values, and it is useful to restrict the track particles and energy deposits to the barrel region of the detector, i.e. $|\eta| < 1$. Out of the 24 calorimeter samplings, only seven cover the η region of interest, and the

[2]Traditional ΔR-based truth-matching is not required; the association between reconstruction-level and truth-level particles is available at truth level in the ATLAS MC samples in RDO format.

[3]N.B.: Due to a bug in the central ATLAS reconstruction software, `Athena`, flagged in May 2020, a significant fraction of track particles (around 49% in the case of $Z \to \mu\mu$ and 27% in the case of $t\bar{t}$) could not be correctly associated with truth particles. The affected tracks were not considered for training. The bug affected track particles randomly based on their MC event number, and, as a result, is not expected to have a systematic effect on this study.

calorimeter samplings used in the generation of training data are chosen on the basis of their η coverage. The presampler and the barrel regions of the electromagnetic and the hadronic tile calorimeters all cover the η region of interest [75]. Table 6.3 summarises the seven calorimeter samplings used, as well as their η coverage and layer indices. The layer indices are a helpful reminder of where the calorimeter samplings are located in relation to the beam pipe; the higher the layer index, the further away from the beam pipe. The radiation length of layer 1 (presampler) is $2X_0$. Layers 2–4 (EM barrel samplings 1, 2, and 3) have radiation lengths of 4.3, 16, and $2X_0$, respectively. The interactions lengths of layers 5–7 (hadronic calorimeter samplings 1, 2, and 3) in terms of λ_n, are 1.5, 4.1, and 1.9, respectively [184, 185].

layer index	calorimeter sampling	η coverage
1	presampler	$\|\eta\| < 1.52$
2	EM barrel, sampling 1	$\|\eta\| < 1.475$
3	EM barrel, sampling 2	$\|\eta\| < 1.475$
4	EM barrel, sampling 3	$\|\eta\| < 1.475$
5	hadronic tile barrel, sampling 1	$\|\eta\| < 1.0$
6	hadronic tile barrel, sampling 2	$\|\eta\| < 1.0$
7	hadronic tile barrel, sampling 3	$\|\eta\| < 1.0$

Table 6.3: Calorimeter-layer indexing system used in this work. Each calorimeter sampling is assigned an index, starting at 1. The layer indices increase with increasing distance from the beam pipe, r.

6.3.3 Kinematics of the training samples

In order not to bias the training samples towards specific phase-space regions, all true muons and pions in the $|\eta| < 1$ range are considered in the training and evaluation datasets. This is done so as to allow the application of any developed algorithm to data, for muons and pions over a wide kinematic region to potentially allow for an expanded use of calorimeter-tagged muons in broader regions in η, as calorimeter-tagged muons so far have only been used in the $|\eta| < 0.1$ region.

Figure 6.2 compares the p_{T} distribution between muons and pions. While the pion p_{T} distribution is peaked at low p_{T} (roughly 8 GeV), the muon p_{T} distribution peaks at higher values of around 40 GeV. Figure 6.2 compares the distribution in η between muons and pions from the two samples.

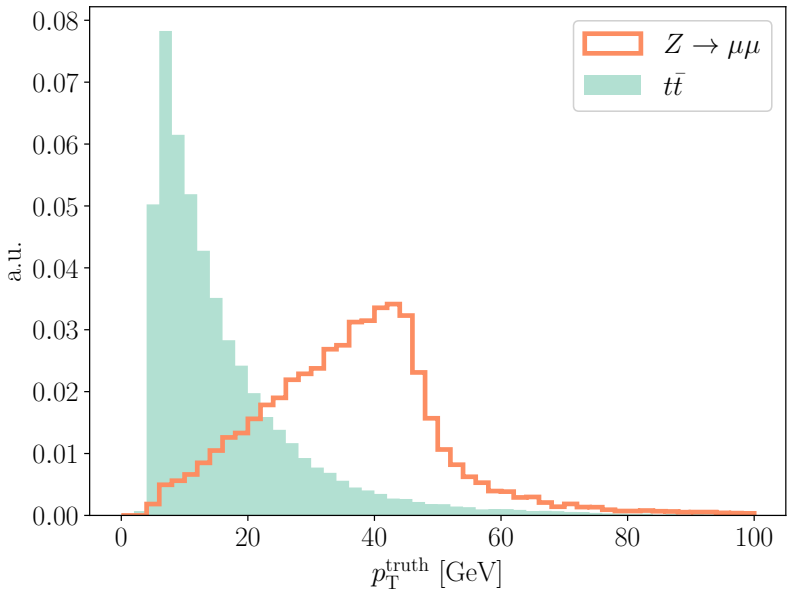


Figure 6.2: Distribution of the transverse momentum, $p_{\mathrm{T}}^{\mathrm{truth}}$, for muons from the $Z \to \mu\mu$ sample (orange) and pions from the $t\bar{t}$ sample (green). The distributions are normalised to unit area to facilitate comparison (a.u.).

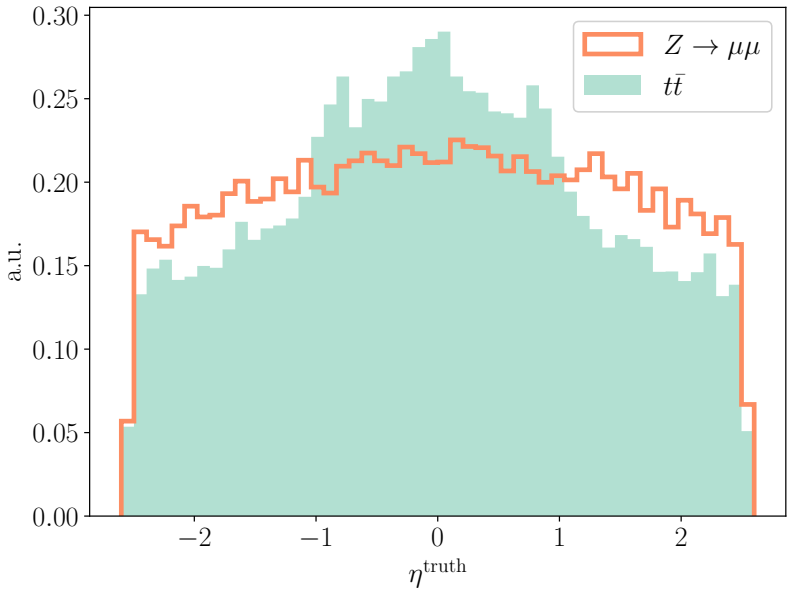


Figure 6.3: Distribution of the pseudorapidity, η^{truth}, for muons from the $Z \to \mu\mu$ sample (orange) and pions from the $t\bar{t}$ sample (green). The distributions are normalised to unit area to facilitate comparison (a.u.).

6.4 Evaluation of calorimeter-tagging methods

A practical way of evaluating and comparing the performance of different binary classifiers is by visualising them in terms of their *receiver operating characteristic* (ROC) curve. ROC curves indicate the diagnostic ability of a binary classifier as the discrimination threshold is varied. ROC curves are created by plotting the true positive rate, which in the problem at hand corresponds to the signal efficiency, ϵ_{sig}, against the false positive rate, which is referred to as the *fake rate* or, equivalently, the *background efficiency*, ϵ_{bkg}, in this work. A convenient metric that quantifies a classifier's overall performance as a single number, which can be derived from a ROC curve, is the *area under the ROC curve* (AUC). The AUC can take values in the interval $[0, 1]$, and is equal to the probability that a classifier will rank a randomly chosen positive event (a signal event) higher than a randomly chosen negative one (a background event).

6.5 Calorimeter-tagging using a feature-based classifier

As a first step towards improving the calorimeter-tagging of muons using the `CaloMuonTag` approach outlined in Section 6.2, a method is investigated that is based on a set of variables that can be extracted from particle tracks as they traverse the calorimeter. This method is referred to as a *feature-based* approach, as it relies on a set of human-engineered features as opposed to features generated by an algorithm, or an algorithm working on low-level data such as energy deposits as will be seen later on. It shall serve, along with the `CaloMuonTag` approach mentioned above, as a baseline for the deep neural networks developed later on in this chapter.

The feature-based approach uses candidate tracks that satisfy the preselection criteria outlined in Section 6.2.1. Particle tracks that pass the preselection are extended into the calorimeter and are associated with all calorimeter cells within a cone of $\Delta R < 0.2$.

6.5.1 Variables used in the feature-based model

The feature-based model investigated is a gradient-boosting tree (Section 4.2), implemented using the `XGBoost` library [113]. The feature-based gradient-boosting classifier is referred to as `XGB`. It uses a set of engineered features based on physics variables commonly used in calorimetry to characterise particle-shower shapes

and energy deposits [186]. In total, the classifier uses 39 features, which are summarised in Table 6.4. The second moments aim to capture the shower widths in x, y, and z. The higher moments are included to capture even more shower-shape information. The authors of Reference [186] do not provide explicit reasons for including the higher moments, however, it is likely that they attempt to capture as much shower-shape information as feasible.

variable	EM calorimeter	hadronic calorimeter	$N_{\text{variables}}$
total energy deposited	✓	✓	2
total number of hits	✓	✓	2
ratio of energy in first layer to energy in second layer	✓	✓	2
ratio of energy in first layer to the entire calorimeter	✓	✓	2
2^{nd}–6^{th} moments in x, y, and z of energy deposits	✓	✓	30
ratio of EM calorimeter energy to hadronic calorimeter energy	×	×	1

Table 6.4: List of variables used in the feature-based approach. All variables except the last one ('ratio of EM calorimeter energy to hadronic calorimeter energy') apply to both the EM and the hadronic calorimeters. The total number of variables is 39.

6.5.2 Optimisation of XGB training parameters

Training parameters for the feature-based classifier were optimised using a grid search. Table 6.5 summarises the parameters over which the grid search was performed, as well as the optimal values found as a result of the search. The objective function was configured to a logistic-regression function for binary classification[4]. For a given observed label y_{pred} and a true label y_{true}, the logistic-regression objective function is given by

$$\mathcal{L} = -\left[y_{\text{true}} \log\left(y_{\text{pred}}\right) + \left(y_{\text{true}} - 1\right) \log\left(1 - y_{\text{pred}}\right)\right]. \tag{6.5}$$

[4]This is the recommended objective function for binary classification within the `XGBoost` library; other objective functions were not explored.

Values of $\mathcal{L}$ above 0.5 are rounded to one, values up to 0.5 are rounded to 0, turning the logistic regression into a binary classification objective function.

parameter	search space	optimum
learning rate	$[0, 1]$	0.3
max. tree depth	$[1, 100]$	8
boosting rounds	$[1, 100]$	10
Γ	$[0, 100]$	0

Table 6.5: Training parameters for the `XGB` classifier. The parameters were optimised using a grid search. Γ denotes the minimum loss reduction required to create a new partition in the tree [113].

6.5.3 Training statistics

As the `XGB` classifier aims to be compared directly to `CaloMuonTag`, which is optimised in the range $|\eta| < 0.1$, `XGB` is trained on the same range in η. The events available for training and testing in this η region are summarised in Table 6.6.

η range	N_{total}	N_{train}	N_{test}
$\lvert\eta\rvert < 0.1$	30,700	24,560	6,140

Table 6.6: Train and test dataset statistics available for the `XGB` classifier. The total dataset of size 30,700 is divided into a training and a test set according to an 80/20 split.

6.5.4 `XGB` training

Figure 6.4 shows the output score of `XGB` on the training and test sets, indicating that the classifier converged well during training, and that no over- or underfitting occurred.

6.5.5 Performance

Figure 6.5 shows the performance of the feature-based model in terms of its ROC curve and compares it to the point in ROC space (ϵ_{sig} vs. ϵ_{bkg}) obtained using the `CaloMuonTag` method. It is not surprising that the feature-based model outperforms the `CaloMuonTag` method, having access to 39 variables instead of just one in the

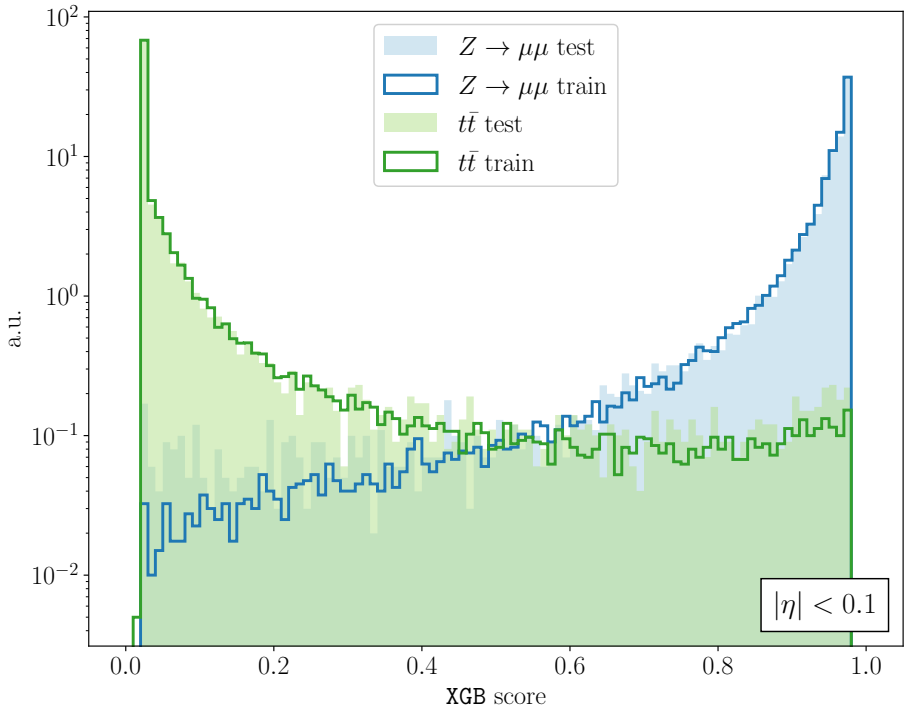


Figure 6.4: Distribution of the XGB output score on the training set compared with the test set in the range $|\eta| < 0.1$ for muons from the $Z \to \mu\mu$ sample (blue), as well as for pions from the $t\bar{t}$ sample (green). The outputs on the test set are shown as filled histograms, whereas the outputs on the training set are unfilled. The near-perfect overlap of the distributions associated with the two particle types indicates that no under- or overfitting occurred during training. The distributions are normalised to unit area to facilitate comparison (a.u.).

case of `CaloMuonTag` (the total calorimeter energy deposit). The gain in performance from using a relatively large number of engineered features is very small, on the other hand, indicating that the `CaloMuonTag` method is highly optimised. The two approaches introduced thus far, `XGB` and `CaloMuonTag`, comprise the set of performance benchmarks against which the more complex convolutional models presented in later sections are evaluated.

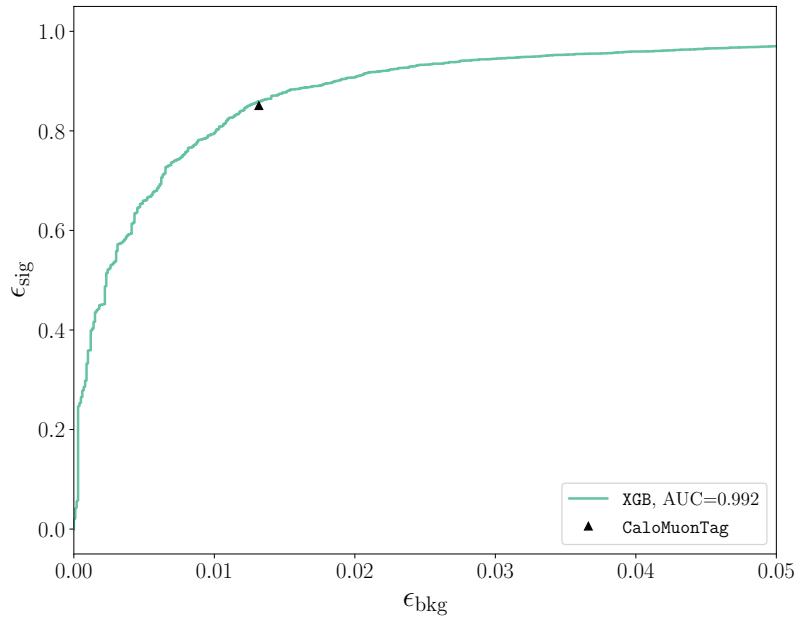


Figure 6.5: Receiver operating characteristic (ROC) curve for the feature-based classifier (labelled `XGB`). The signal efficiency is shown on the y-axis, while the background efficiency is shown on the x-axis. 'AUC' refers to the area under the ROC curve. The `CaloMuonTag` working point is shown for comparison. The evaluation was performed in the range $|\eta| < 0.1$.

6.6 Calorimeter-tagging using convolutional neural networks

The ML-based approach to calorimeter-tagging muons focuses on the central barrel region of the ATLAS detector, as this is where calorimeter-tagged muons are used most frequently. In practice, this means that the selection of events that are considered for training, as well as the translation of calorimeter information to a format a machine-learning algorithm can process, will focus on events that satisfy the following criteria:

- Events need to have a track in the ATLAS inner detector, that is, they need to have associated with them an *ID track*.
- The track particle's pseudorapidity, η_{track}, satisfies $|\eta_{\text{track}}| < 1$.
- Track particles need to be truth-matched, that is, associated with a truth-level MC particle (see Section 6.3.1).
- The truth-matched particle is either a muon (in the case of the signal sample), or to a pion (in the case of the background sample).

In this work, the track particle's η range was expanded to cover $|\eta| < 1$ in order to allow for a potential use of calorimeter-tagged muons in a broader pseudorapidity range compared with the previous optimisation in the very narrow region of $|\eta| < 0.1$.

6.6.1 Event selection

As in the case of the feature-based approach, inner detector tracks need to fulfil the preselection criteria detailed in Section 6.2.1. Again, these particle tracks are extended into the calorimeters, and all calorimeter cells within a cone of $\Delta R < 0.2$ are considered for a given event.

6.6.2 Conversion into a 2D computer vision problem

A matrix representation of calorimeter energy deposits is developed so that machine-learning techniques from the domain of computer vision, such as convolutional neural networks, can be applied. The scheme devised in this work extends the RGB, i.e. three-channel, digitisation of images to seven channels. The advantage is that reliable approaches from computer vision can be employed, as they can be readily adapted to handle a higher number of 'colour' channels[5]. As this work is concerned with the central barrel region of the ATLAS detector, which is covered by seven distinct calorimeter layers, the convolutional neural networks are constructed such that they can process seven 'colour' channels.

[5]The 'colour' analogy to describe the energy deposits read out by the individual calorimeter layers is convenient for several reasons: The input data processed by the convolutional neural networks described in this chapter can be thought of as 2D matrices with a multiplet of seven numbers for each entry. Further, in the computer vision literature, the term *channels* refers to the depth of the input image matrices or the number of distinct colours that are processed. While technically not describing colours in the ATLAS calorimeter case, the *channel* terminology is adopted throughout this work.

Traditional computer vision approaches deal with two-dimensional images using the RGB colour model[6], where three *channels* represent the brightness values of the individual colours. When dealing with energy deposits in several layers of the ATLAS calorimeter, the situation is very similar: In a given layer, the smallest unit of the calorimeter that can be read out can be thought of as a pixel that stores a number corresponding to an energy deposit in the cell. This energy deposit can be considered to correspond to a colour brightness, just like in the case of RGB images. Stacking several such layers on top of one other is analogous to the combination of three colour channels in the RGB colour model. Figures 6.6 and 6.7 illustrate this idea. As a result, a given track particle that traverses the ATLAS calorimeter is converted into an 'image' with seven colour channels (more on this in Section 6.6.3).

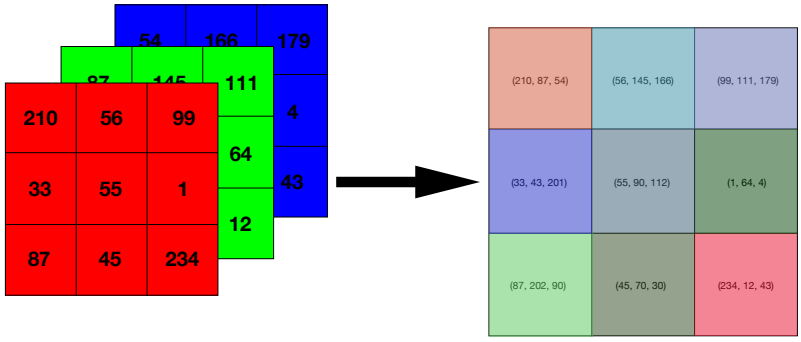


Figure 6.6: A three-dimensional RGB matrix of size $3 \times 3 \times 3$ corresponding to a 3×3 px image. The input image on the left contains three layers, where each layer is a two-dimensional matrix containing a brightness value for either the *red*, *green* or *blue* colour channel. The two-dimensional matrix on the right is an equivalent representation, i.e. a two-dimensional matrix in which each entry is a triplet of numbers. For illustrative purposes, cell colours on the right correspond to the RGB values of the brightness triplets contained within them.

6.6.3 Data preprocessing

In order to be able to use the computer-vision technique outlined above (Section 6.6.2), one needs to choose the granularity of the resulting seven-channel 2D images (this corresponds to the *resolution* of the 2D image). The approach to finding the optimal resolution taken in this work was to train and evaluate the machine-learning algorithms on images with increasing resolution and to stop at the resolution at

[6]Digital images that are stored in the RGB colour model can be thought of as a matrix of pixels, where each pixel is a triplet of numbers containing brightness values for the colours *red*, *green*, and *blue*.

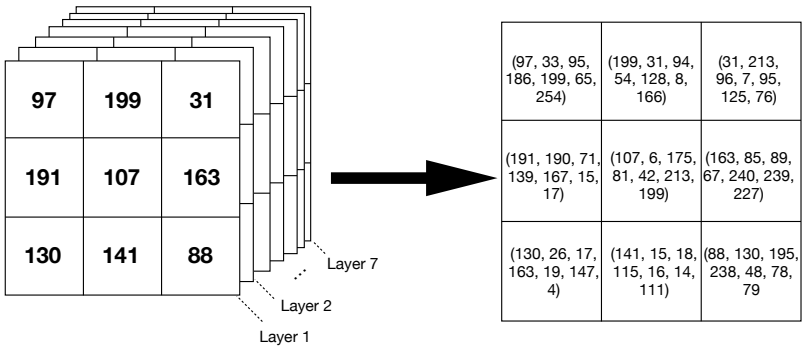


Figure 6.7: A three-dimensional matrix of size $3 \times 3 \times 7$ corresponding to a 3×3 px image consisting of seven layers. The entries in the matrix on the left are brightness values corresponding to the energy deposited in each pixel. The resulting 3×3 matrix on the right contains a septuple of layer brightnesses for every pixel.

which no further increase in performance was observed. The resulting optimal resolution was found to be 30×30 px. Using this resolution, each pixel corresponds to an area in $\eta \times \phi$ of $1/60 \times 1/60 = 0.017 \times 0.017$. It is worth noting that while both the optimal image size in terms of pixels and the optimal pixel size in terms of η and ϕ were found to be square, this was not a requirement; rectangular configurations were investigated, yet generally performed slightly worse than square configurations. The presumed reason is that for the low pseudorapidities explored ($\eta < 1$), $\eta \sim -\theta$, resulting in (lateral) shower sizes in $\eta - \phi$-space that correspond almost exactly to distances in $\eta - \theta$-space.

The calorimeter hit patterns are converted into images of size 30×30 px. The conversion takes place for the generation of training data and in the preprocessing of events during the application of the classifier alike; there is one difference in treatment between the two cases, and it is highlighted below. The conversion of hit patterns into images works as follows:

1. Select all track particles that satisfy $|\eta_{\text{track}}| < 1$.

   (a) For training data only: Keep only the subset of track particles that are truth-matched to a muon or a pion.

2. Select all energy deposits within a cone of $\Delta R < 0.2$ around the ID track for all layers (1–7).

3. Split the energy deposits into seven layers corresponding to the central-barrel calorimeter layers (Table 6.3).

4. Shift in η: Shift the resulting hit pattern in each layer in η by η_{track}, such that

$$\eta' \rightarrow \eta - \eta_{\text{track}}. \tag{6.6}$$

   The largest possible shifts occur for $\eta_{\text{track}} = \pm 1$.

5. Shift in ϕ: Shift the resulting hit pattern in each layer in ϕ by ϕ_{track}, such that

$$\phi' \rightarrow \phi - \phi_{\text{track}}. \tag{6.7}$$

6. Segment the resulting seven hit maps into 30 bins in η', as well as 30 bins in ϕ'.

The result of this preprocessing are 3D matrices with dimensions of $\eta \times \phi \times$ channel, which are easily visualised as 2D images. The conversion from 3D calorimeter hit patterns into 2D images is illustrated by way of example in Figures 6.8 to 6.10, which look at a randomly chosen muon from the $Z \rightarrow \mu\mu$ sample in calorimeter layer 3.

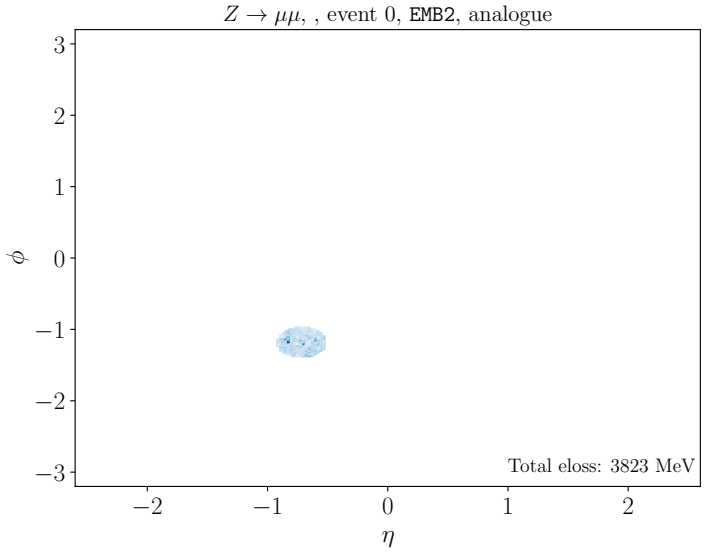


Figure 6.8: 2D heat map of ϕ vs. η showing the energy deposits left by a muon in calorimeter layer 3 (`EMB2`, which stands for EM barrel 2). The opacity of each blob is proportional to the energy deposited. The total energy deposited by this muon in layer 3 is 3823 MeV.

At this stage, it is instructive to examine the digitised hit patterns (30×30 px images) for both muons and pions in order to highlight a few noteworthy differences. An important observation is that pions tend to lose more of their energy in the first few calorimeter layers traversed compared to muons. This is because muons lose

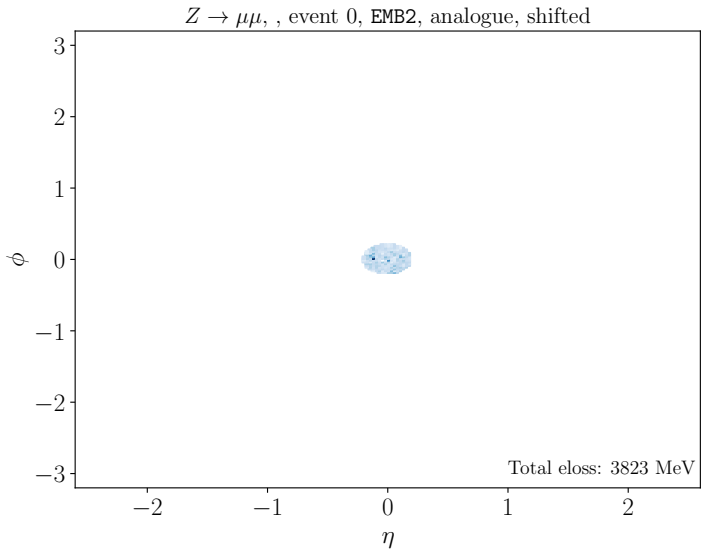


Figure 6.9: 2D heat map of ϕ vs. η corresponding to the same muon shown in Figure 6.8, with η and ϕ shifted by η^{track} and ϕ^{track}, respectively.

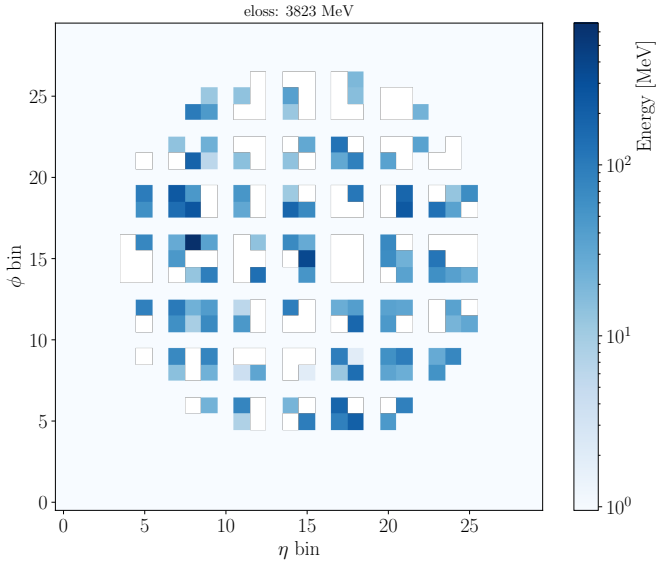
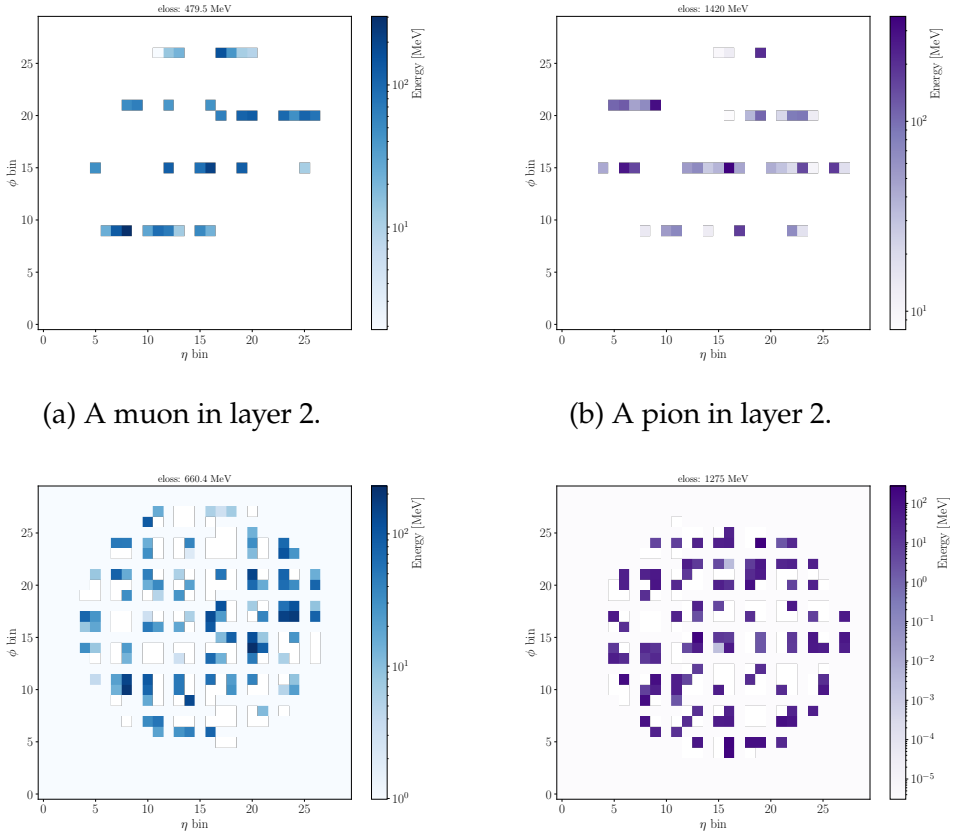


Figure 6.10: 2D $\phi - \eta$ grid of 30×30 px showing the discretised energy-loss pattern for the same muon seen in Figures 6.8 and 6.9. This pixel grid corresponds to the area in Figure 6.9 defined by $-0.25 < \eta < 0.25$ and $-0.25 < \phi < 0.25$.

relatively little energy through minimum ionisation, whereas pions lose most of their energy through nuclear interaction with the detector material. The difference is illustrated in Figure 6.11, which shows the images produced by two randomly selected muon and pion events in calorimeter layers 2 and 3.



(a) A muon in layer 2.

(b) A pion in layer 2.

(c) A muon in layer 3.

(d) A pion in layer 3.

Figure 6.11: 2D $\phi - \eta$ grids of 30×30 px showing the discretised energy-loss patterns for a randomly chosen muon event (a and c), and a randomly chosen pion event (b and d) in calorimeter layers 2 and 3.

A convenient quantity used to compare the density of hits, or 'lit pixels', between calorimeter images is *sparsity*, ξ, defined as the fraction of pixels with a non-zero brightness. The average images of 20,000 muons and 20,000 pions are compared for several calorimeter layers in Figure 6.12. Figures 6.12a and 6.12b show clearly that the energy-loss patterns vary greatly in the inner layers of the calorimeter between muons and pions. In the middle layers of the calorimeter, such as layer 4, as shown in Figures 6.12c and 6.12d, the difference is more minor in terms of sparsity, but large in terms of deposited energy. In the outermost calorimeter layer considered

here, the difference between the two particle types is very small (Figures 6.12e and 6.12f), both in terms of sparsity and energy loss. The average energy-loss values and sparsities for muons and pions in the seven calorimeter layers are summarised in Table 6.7. It is evident that the scale of the particles' total energy deposits carries important information—the mean energy losses of pions are significantly higher than of muons in almost all layers—and, as a result, the calorimeter images were chosen to not be normalised to the particles' total energy.

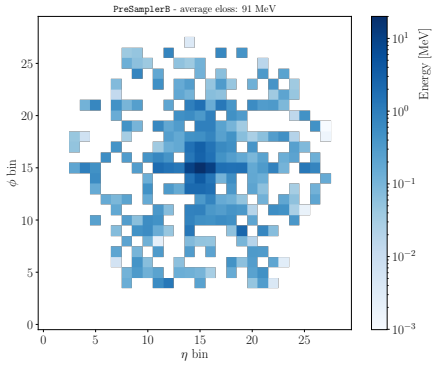
layer	$\langle \xi^\mu \rangle$ [%]	$\langle \xi^\pi \rangle$ [%]	$\langle E^\mu \rangle$ [MeV]	$\langle E^\pi \rangle$ [MeV]	ΔE [MeV]
1	1.46	1.82	91	1452	1361
2	1.64	2.65	502	6980	6478
3	3.54	6.68	1408	17580	16172
4	0.36	0.75	77	803	726
5	0.34	0.76	996	8463	6790
6	0.24	0.57	1673	6874	5201
7	0.11	0.07	622	524	−98

Table 6.7: Sparsities, average energy losses, and differences in energy loss for muons and pions in seven calorimeter layers. $\langle \xi^\mu \rangle$ denotes the sparsity of muon images, and $\langle \xi^\pi \rangle$ that of pion images. $\langle E^\mu \rangle$ is the average energy loss of muons, $\langle E^\pi \rangle$ that of pions. ΔE is the difference in average energy losses between muons and pions, and is defined as $\Delta E = \langle E^\pi \rangle - \langle E^\mu \rangle$.

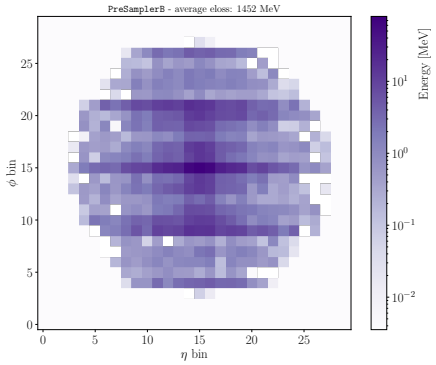
6.6.4 Classifier architecture

Four convolutional architectures are developed and compared with each other. The first two networks, 'CNN 1' and 'CNN 2' are designed from scratch, with their architectures inspired vaguely by the *AlexNet* architecture [187]. AlexNet was one of the most successful implementations of a convolutional neural network and occupied top ranks on the popular image-classification benchmark *ImageNet* [188] for several years. The remaining two networks' architectures are adapted from state-of-the-art image-classification networks introduced after *AlexNet*: *Xception* [189] and *ResNet* [190]. The four network architectures investigated here can be summarised as follows:

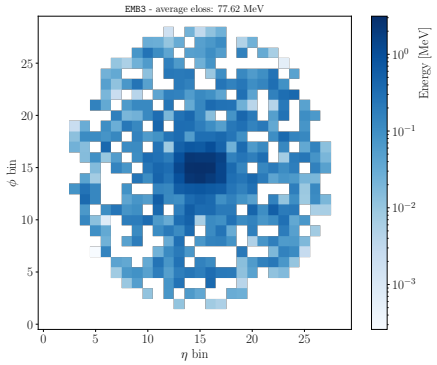
- **CNN 1:** This straightforward convolutional architecture consists of a series of three combinations of convolutional layers directly followed by max-pooling layers. The output of the third max-pooling layer is fed into a wide dense



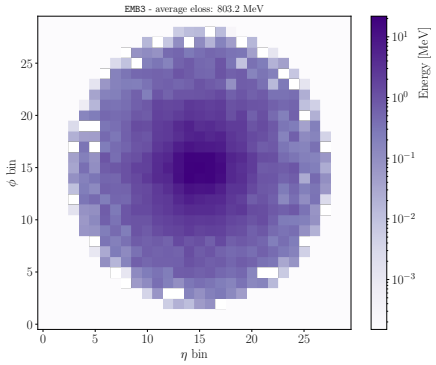
(a) Average of 20,000 muon images in layer 1.



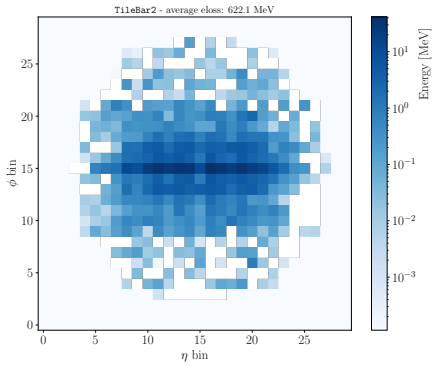
(b) Average of 20,000 pion images in layer 1.



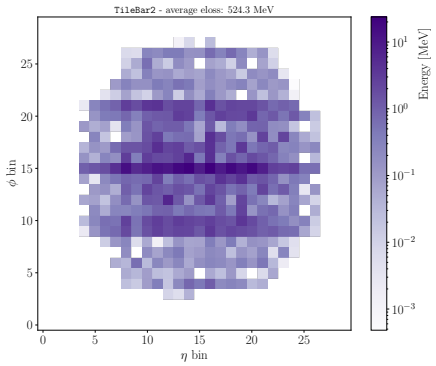
(c) Average of 20,000 muon images in layer 4.



(d) Average of 20,000 muon images in layer 4.



(e) Average of 20,000 muon images in layer 7.



(f) Average of 20,000 muon images in layer 7.

Figure 6.12: Average images produced by 20,000 muons and 20,000 pions in different calorimeter layers. The muon images are shown on the left (a, c, and e), whereas the pion images are shown on the right (b, d, and f). The plots are 2D $\phi - \eta$ grids of 30×30 px showing the binned energy losses.

layer (it has a width of 2,304 neurons). The number of combinations of convolutional and max-pooling layers and the width of the pre-final dense layer were optimised using a grid search. The architecture of this network is summarised in Figure 6.13.

- **CNN 2:** The architecture of 'CNN 2' makes use of a modified version of an ordinary convolutional layer, called a *depthwise separable convolution*. Depthwise separable convolutions were first introduced in Reference [191], and have been recently used in the *Inception* [192, 193] and *MobileNet* [194] architectures; they achieve a significant reduction in computational steps required to train a CNN by restricting the convolutional kernels to one colour channel each. The network consists of three units of two combined depthwise separable convolutions, followed by a global average pooling layer[7], and eventually a dense layer. The number of depthwise separable convolutional units and the width of the final dense layer of 32 neurons were optimised using a grid search. The full network architecture is summarised in Figure 6.14.

- **Xception:** *Xception* is a convolutional neural network that is very deep with 36 convolutional layers. It outperformed previous models on the *ImageNet* image-classification benchmark test when it was introduced in 2017. The original *Xception* architecture was developed for RGB image-classification problems and is summarised in Reference [189]. In order for a classifier of this kind to be used on the calorimeter images at hand in this work, it was adapted slightly to handle the input matrices containing seven colour channels.

- **ResNet:** The *ResNet* model architecture achieved state-of-the-art performance on the *ImageNet* classification task. As in the case of *Xception*, *ResNet* was developed for computer vision problems in RGB colour space. *ResNet* stands for *residual network*, and its groundbreaking feature is the introduction of *skip connections*, which allow it to skip layers in the network. This makes it possible for extremely deep networks to be trained with finite computational resources. The default *ResNet* architecture outlined in Reference [190] was adapted to handle input images with seven colour channels.

[7]Global average pooling layers are very similar to max-pooling layers in that they perform a dimensionality reduction. The degree of reduction in dimensionality, however, is much greater than in the case of max pooling. For a matrix with dimensions $w \times h \times d$, global average pooling reduces the dimension to $1 \times 1 \times d$, where the matrix layers projected onto $w - h$ space are replaced by a single number (their averages) [195].

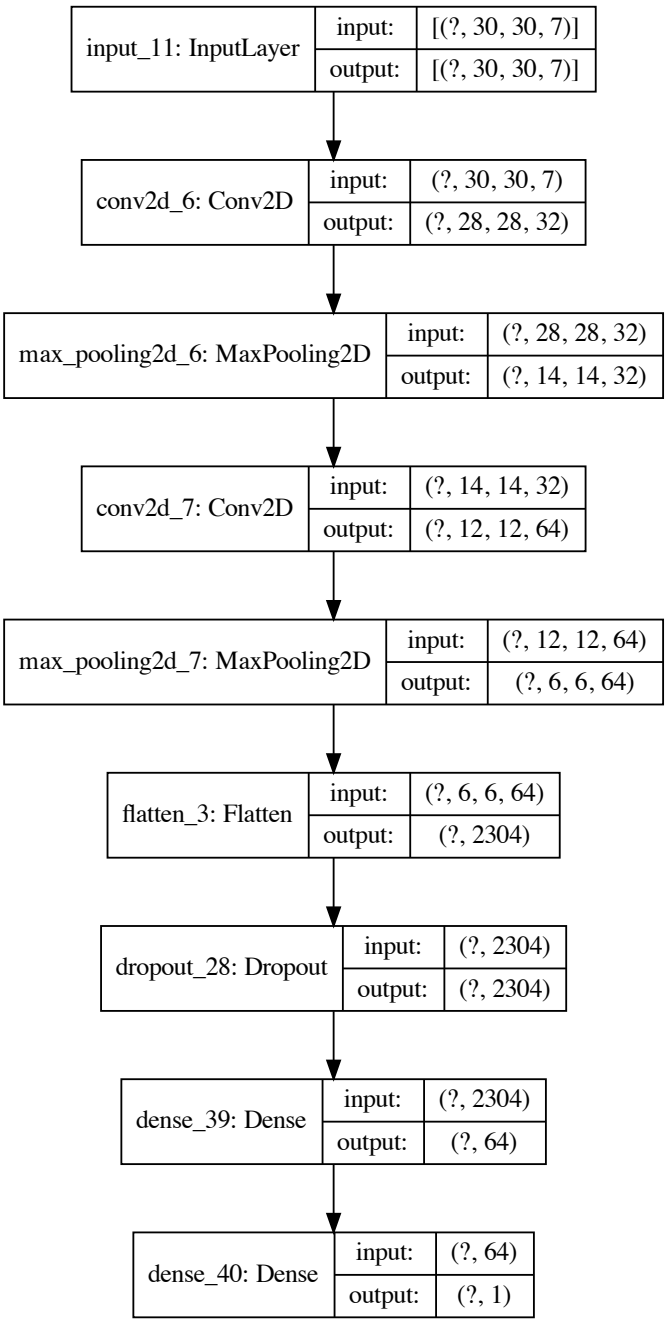


Figure 6.13: Diagram summarising the architecture of 'CNN 1'. The input images of size $30 \times 30 \times 7$ are fed into two successive combinations of convolutional and max-pooling layers. The output of the last max-pooling layer is flattened into a one-dimensional array of size 2304, followed by a layer that applies a dropout. The last hidden layer is a dense layer that is 64 neurons wide. The final layer is just a single neuron with sigmoid as its activation function (this is needed for binary classification). The question marks ('?') in the first dimension of each layer indicate the batch size, which is a variable training parameter and of no importance to the classifier architecture.

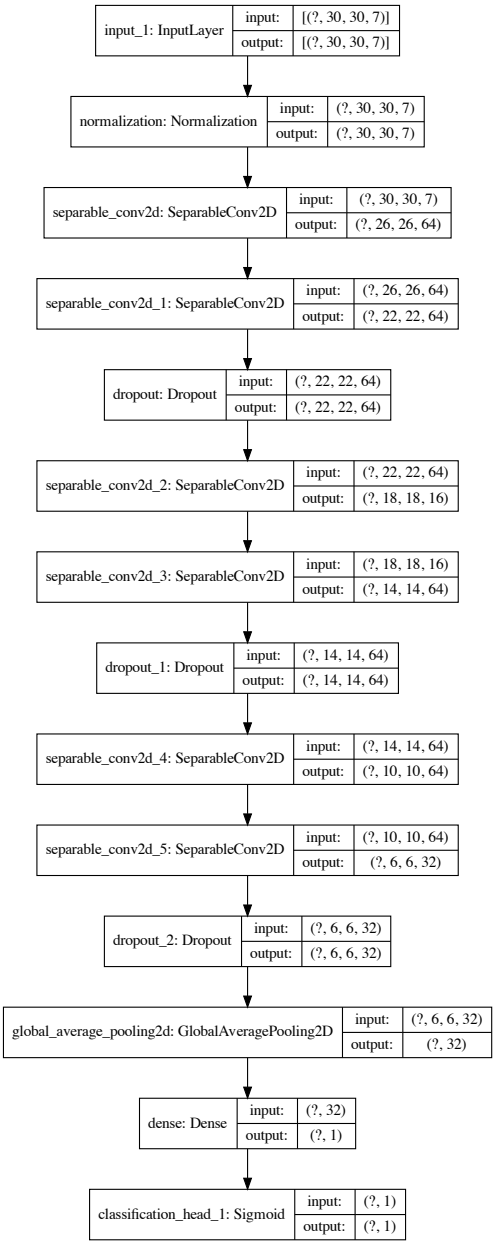


Figure 6.14: Diagram summarising the architecture of 'CNN 2'. The input images of size $30 \times 30 \times 7$ are first fed into a batch-normalisation layer. In this layer, the images are normalised using the mean and standard deviation of the current training batch. The images are then fed into two depthwise separable convolutional layers, followed by a dropout layer. This combination of layers is repeated twice more. The last two hidden layers are a global average pooling layer and a dense layer 32 neurons wide. The final layer is a dense layer (one neuron wide) with sigmoid as its activation function. The question marks ('?') in the first dimension of each layer indicate the batch size, which is a variable training parameter and of no importance to the classifier architecture.

The architectures were implemented using the `TensorFlow` library [165]. Where mentioned, parameters related to the network architecture – such as how many convolutional layers to use or how many neurons a dense layer should contain – were optimised using grid searches. Training hyperparameters, on the other hand, were optimised using a random search approach[8], implemented using the `Hyperopt` package [171]. The final set of hyperparameters obtained from this optimisation and the total number of trained parameters in each of the classifier architectures are summarised in Table 6.8.

architecture	optimiser	learning rate	dropout	epochs	batch size	parameters
CNN 1	Adam	0.01	0.5	22	256	168,129
CNN 2	Adam	0.001	0.25	30	32	25,136
Xception	Adam	0.001	—	20	128	1,057,296
ResNet	Adam	0.001	—	50	128	15,617,693

Table 6.8: Training hyperparameters and number of trainable parameters for the four convolutional neural network architectures investigated. *Dropout* refers to a regularisation technique that is used to reduce overfitting (c.f. Section 4.3.1.5).

All available muon and pion events are split into three classes. Out of all available events, 20% are held for testing (they are not used during training). The remaining 80% are further split into a training set comprising 64% of the entire dataset and a validation set comprising 16% of the entire dataset. The datasets are balanced so that there is an equal number of muons and pions in each class (training, validation, and test). The relative proportions of the three datasets are visualised in Figure 6.15.

Figure 6.15: Proportion of the entire dataset reserved for training, validation, and testing.

The training statistics available for the three datasets in the η region of $|\eta| < 1$ are summarised in Table 6.9.

[8]A grid search samples every point in the defined hyperparameter search space at some predefined granularity. A random search, on the other hand, randomly samples a subset of points in hyperparameter space. The choice of random search over grid search was made so as to keep the computational resources required for training manageable.

η range	N_{total}	N_{train}	N_{val}	N_{test}
$\lvert\eta\rvert < 1$	129,800	83,072	20,768	25,960

Table 6.9: Training, validation, and test statistics for events that satisfy $|\eta| < 1$.

6.7 The CaloMuonScore classifier

Out of the four convolutional architectures investigated, 'CNN 2' performs significantly better than all others. More details on the performance of the different classifiers will be given in this section, however, in order to facilitate the comparison between the convolutional methods and previous ones, the classifier obtained from 'CNN 2' will from here on be referred to as CaloMuonScore.

The performance of CaloMuonScore is studied separately in two η regions: (1) the wide η range for which training data are generated ($|\eta| < 1$), and, therefore, over which calorimeter-tagging muons using the CNN-based approach developed here is feasible, and (2) the narrow η range for which CaloMuonTag is optimised ($|\eta| < 0.1$).

6.7.1 Performance in the range $|\eta| < 1$

Trained on the wider η range ($|\eta| < 1$), the four convolutional neural networks are compared with each other, and the best performing one is selected. Figure 6.16 shows this comparison. It reveals that 'CNN 2' (i.e. CaloMuonScore) outperforms all other models by a significant margin.

CaloMuonScore achieves excellent performance on the test set in the wide η range, which is illustrated in Figure 6.17, indicating that the convolutional architecture devised learns the binary-classification task very well. It is at this point instructive to compare the classifier output on the test set with the output on the training set, i.e. the data the classifier observed during training. Figure 6.18 highlights this comparison, from which it is evident that no under- or overfitting has occurred.

6.7.2 Definition of CaloMuonScore working points

CaloMuonTag is not defined on a sliding scale (it is a binary label), and therefore constitutes a single *working point*. In order to facilitate the comparison between CaloMuonTag and CaloMuonScore, it is convenient to define four working points for CaloMuonScore as follows:

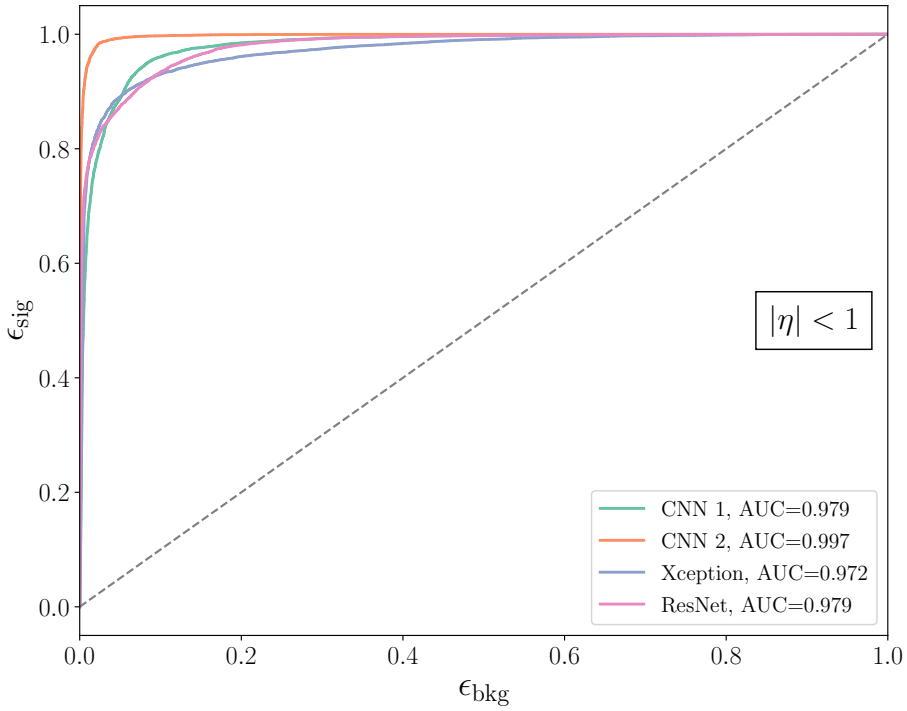


Figure 6.16: Receiver operating characteristic (ROC) curves for the four CNNs investigated. The signal efficiency is shown on the y-axis, while the background efficiency is shown on the x-axis. 'AUC' refers to the area under the ROC curve. Training and evaluation of these networks were performed over the full η range ($|\eta| < 1$). The grey dashed line represents the theoretical ROC curve one would obtain by making random guesses.

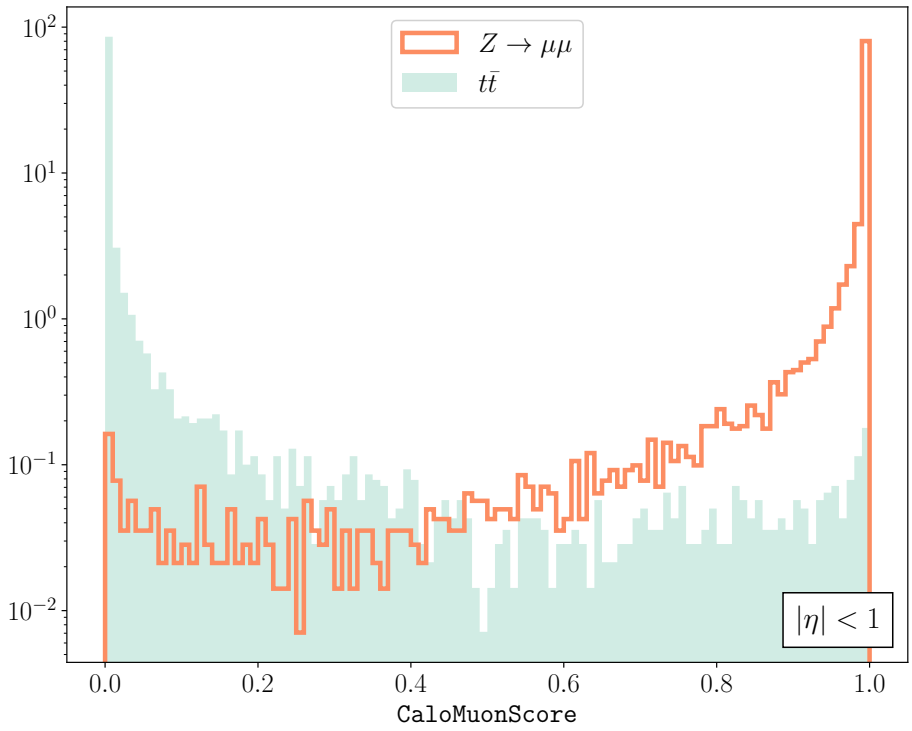


Figure 6.17: Distribution of `CaloMuonScore` on the test set in the range $|\eta| < 1$ for muons from the $Z \to \mu\mu$ sample (orange), as well as for pions from the $t\bar{t}$ sample (green). The y-axis is in arbitrary units (a.u.), and the two distributions are normalised to unit area to facilitate comparison.

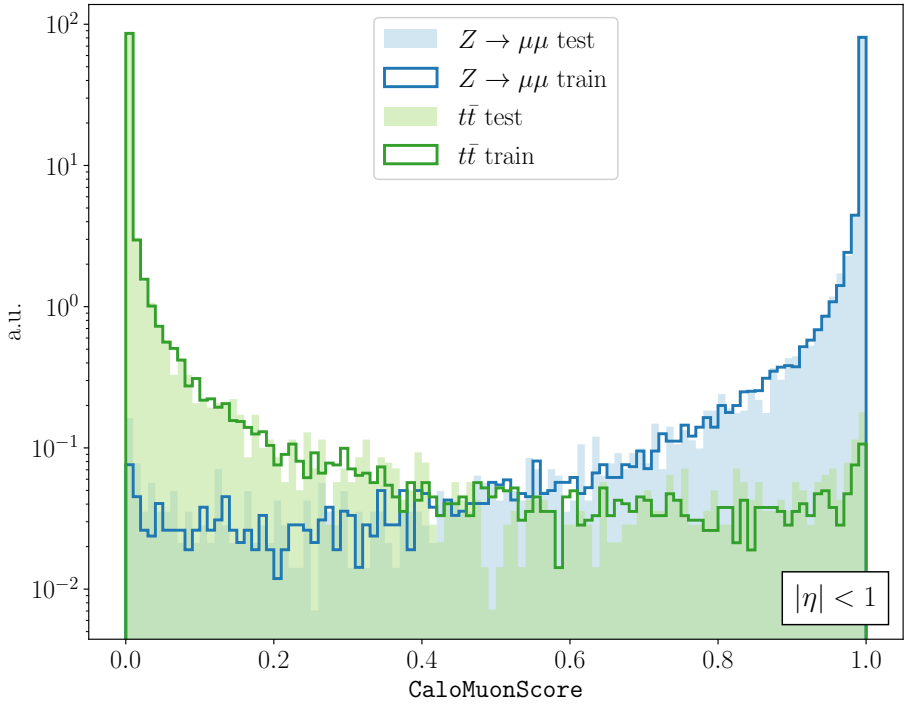


Figure 6.18: Distribution of `CaloMuonScore` on the training set compared with the test set in the range $|\eta| < 1$ for muons from the $Z \to \mu\mu$ sample (blue), as well as for pions from the $t\bar{t}$ sample (green). The outputs on the test set are shown as filled histograms, whereas the outputs on the training set are unfilled. The near-perfect overlap of the distributions associated with the two particle types indicates that no under- or overfitting has occurred during training. The distributions are normalised to unit area to facilitate comparison (a.u.).

- WP1: The cut on CaloMuonScore that achieves the same muon efficiency as CaloMuonTag for $|\eta| < 0.1$ and $p_T \in [30, 40]$ GeV. This efficiency is 90.1%.
- WP2: The cut on CaloMuonScore that achieves a muon efficiency of 97.0% for $|\eta| < 0.1$ and $p_T \in [30, 40]$ GeV.
- WP3: The cut on CaloMuonScore that achieves the same background efficiency as CaloMuonTag for $|\eta| < 0.1$. These are 10.0% for $p_T \in [5, 15]$ GeV, 8.2% for $p_T \in [15, 20]$ GeV, and an average of 6.8% thereafter.
- WP4: The cut on CaloMuonScore that achieves a flat fake rate of 6.8% for $|\eta| < 0.1$ and all p_T.

The background efficiencies are based on non-prompt muons (i.e. muons that do not come from the primary $Z \to \mu\mu$ interaction vertex). The working points can be broadly categorised into those that are constant in p_T, expressed as a simple cut on the value of CaloMuonScore, and those that vary in p_T, expressed as a cut on CaloMuonScore that varies as a function of p_T.

The ϵ_{sig} targets for WP1 and WP2 are optimised for the $p_T \in [30, 40]$ GeV range as this where the majority of true muons from $Z \to \mu\mu$ decays is expected (c.f. Figure 6.2). Targeting a desired muon efficiency in a fairly narrow p_T window, WP1 and WP2 are expressed as constant cuts on CaloMuonScore. Referring to Figure 6.20, and recording the values for which the two conditions are achieved, we find that WP1 corresponds to a cut value on CaloMuonScore of 0.92, whereas WP2 corresponds to a cut value of 0.59.

The background efficiency achieved by CaloMuonTag rises steeply with p_T, leading to very high fake rates (60% or more) in kinematic regions that could be of interest to physics analyses involving, for instance, prompt muons from Z, $W^{\pm}$, or H decays. This behaviour can be suppressed using CaloMuonScore owing to its continuous output by using p_T-dependent cuts. WP3 and WP4 are two such p_T-dependent working points, and are expressed as step functions in muon p_T. The background efficiency achieved by CaloMuonTag varies especially strongly up to a p_T of 20 GeV, and is relatively constant thereafter. As a result, WP3 and WP4 are expressed as step functions in p_T, consisting of cuts defined by third-order polynomials up to a p_T of 20 GeV, and constant cuts thereafter. The precise definitions of the functions and the polynomial parameters are as follows:

$$f_{\text{WP3}}(p_{\text{T}}) = \begin{cases} \text{Poly3}(p_{\text{T}})^{\text{WP3}} & p_{\text{T}} \leq 20\ \text{GeV} \\ 0.76 & p_{\text{T}} > 20\ \text{GeV} \end{cases}$$

and

$$f_{\text{WP4}}(x) = \begin{cases} \text{Poly3}(p_{\text{T}})^{\text{WP4}} & p_{\text{T}} \leq 20\ \text{GeV} \\ 0.76 & p_{\text{T}} > 20\ \text{GeV}. \end{cases}$$

The coefficients of the third-order polynomials $\text{Poly3}(p_{\text{T}})^{\text{WP3}}$ and $\text{Poly3}(p_{\text{T}})^{\text{WP4}}$ are summarised in Table 6.10. The fitting procedure is illustrated in Figure 6.19: The red dots indicate the cuts on `CaloMuonScore` required to achieve the desired target fake rates in four p_{T} bins ($p_{\text{T}} \in [5, 10]$, $p_{\text{T}} \in [10, 15]$, $p_{\text{T}} \in [15, 20]$ and $p_{\text{T}} \in [20, \infty]$ GeV), corresponding to `WP3`. The blue line is a third-order polynomial fit, with the constraint that the function be monotonically decreasing so as to avoid cut values that increase at low p_{T}.

name	a	b	c	d
$\text{Poly3}(p_{\text{T}})^{\text{WP3}}$	$-1.98 \cdot 10^{-4}$	$-6.04 \cdot 10^{-3}$	$-6.13 \cdot 10^{-2}$	1.16
$\text{Poly3}(p_{\text{T}})^{\text{WP4}}$	$-1.80 \cdot 10^{-4}$	$5.02 \cdot 10^{-3}$	$-4.62 \cdot 10^{-2}$	1.12

Table 6.10: Polynomial fit parameters for the p_{T}-dependent cuts defining `WP3` and `WP4` describing third-order polynomials of the form $f(x) = ax^3 + bx^2 + cx + d$.

6.7.3 Performance in the range $|\eta| < 0.1$

The performance of `CaloMuonScore` is benchmarked against `CaloMuonTag` and the feature-based classifier (`XGB`), as shown in Figure 6.20. It is apparent that `CaloMuonScore` performs significantly better than both `CaloMuonTag` and `XGB`. As a result, the following performance comparisons will focus on `CaloMuonTag` and `CaloMuonScore`.

6.7.4 ϵ_{sig} and ϵ_{bkg} in the two η regions

Having defined four `CaloMuonScore` working points, this section compares the specific signal and background efficiencies as functions of η and p_{T}, separately in the $|\eta| < 0.1$ and $|\eta| < 1$ regions.

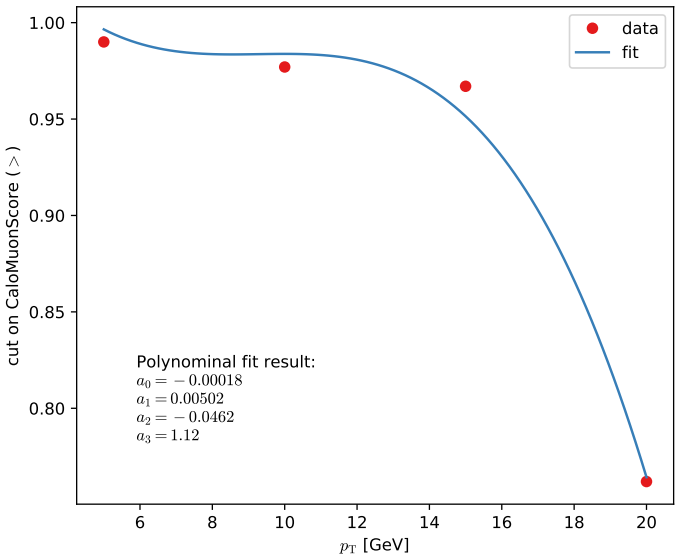


Figure 6.19: Third-order polynomial fit (blue line) to cuts on CaloMuonScore in four p_T bins (red dots). The blue line defines WP3 up to a p_T of 20 GeV. For $p_T > 20$ GeV, WP3 is defined as a constant cut on CaloMuonScore. The y-axis is the minimum value of CaloMuonScore, whereas the x-axis is the muon transverse momentum.

6.7.4.1 ϵ_{sig} and ϵ_{bkg} in the region $|\eta| < 0.1$

Regarding specific efficiency values of the different methods, it is helpful to compare the signal efficiency at a given background efficiency. As CaloMuonTag does not allow for a sliding cut (it is simply a binary tag), the background efficiency obtained by CaloMuonTag for $p_T \in [30, 40]$ GeV is chosen for comparison, namely $\epsilon_{bkg} = 48.18\%$. Table 6.11 compares the signal and background efficiencies obtained by CaloMuonTag and the four CaloMuonScore working points. The comparison shows that CaloMuonScore, operating at WP1, can identify true muons three percentage points more effectively than CaloMuonTag, while suppressing background more efficiently by a factor of 19.9. At WP2, CaloMuonScore would identify muons 7 percentage points more efficiently than CaloMuonTag, while also reducing the fake rate by a factor of 7.6.

Figures 6.21 to 6.24 show the performance of CaloMuonTag and the four CaloMuonScore working points in terms of ϵ_{sig} and ϵ_{bkg} vs. p_T and η in the low-η region ($|\eta| < 0.1$).

6.7.4.2 ϵ_{sig} and ϵ_{bkg} in the region $|\eta| < 1$

Figure 6.25 shows the performance of CaloMuonTag and CaloMuonScore in the $|\eta| < 1$ region in terms of ϵ_{sig} and ϵ_{bkg} vs. p_T and η, where CaloMuonTag operates

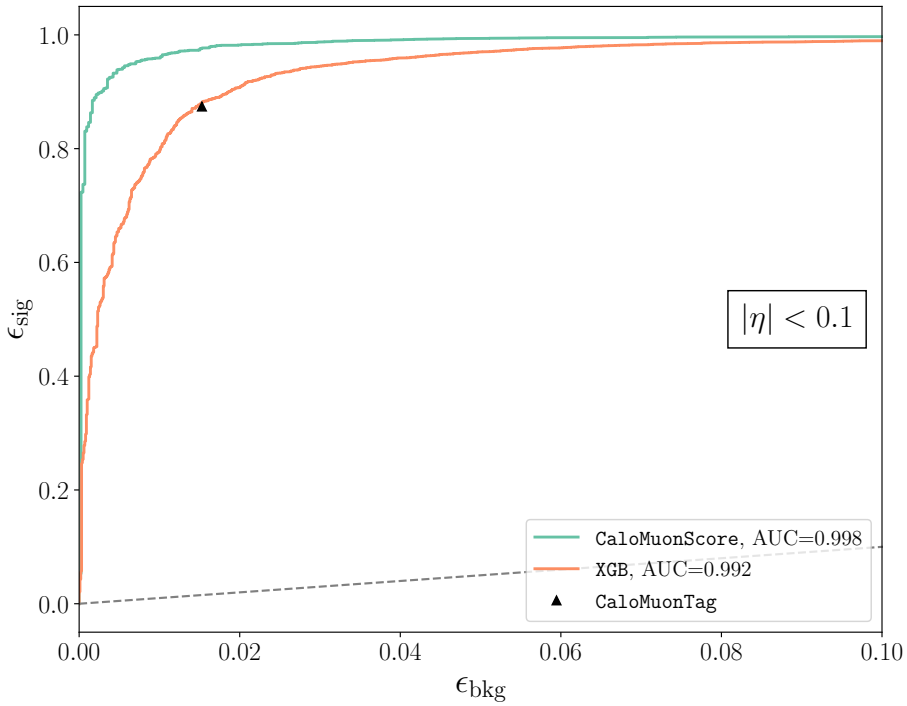
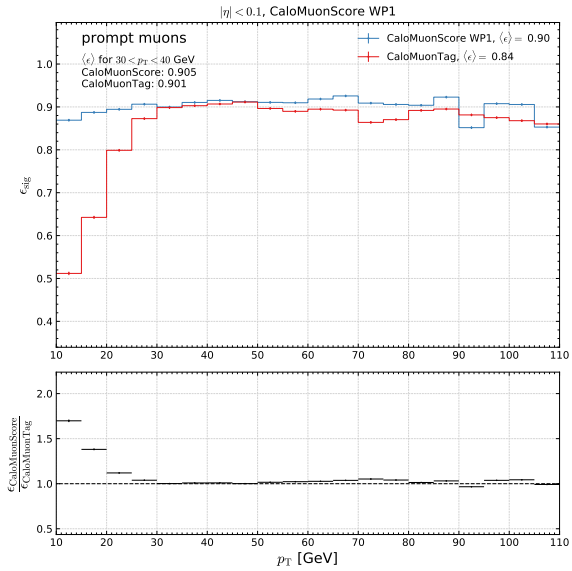


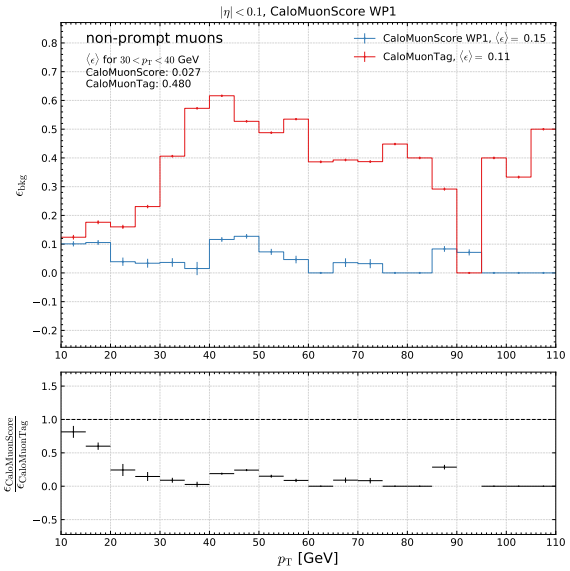
Figure 6.20: Receiver operating characteristic (ROC) curves for `CaloMuonScore` (green) and `XGB` (orange) on the held-out test set in the range $|\eta| < 0.1$. The signal efficiency is shown on the y-axis, while the background efficiency is shown on the x-axis. 'AUC' refers to the area under the ROC curve. The `CaloMuonTag` working point is shown for comparison. The grey dashed line represents the theoretical ROC curve one would obtain by making random guesses.

classifier	ϵ_{sig}	ϵ_{sig} improvement	ϵ_{bkg}	ϵ_{bkg} improvement
`CaloMuonTag`	88.06%	—	48.18%	—
`CaloMuonScore WP1`	90.99%	2.94 p.p.	2.42%	×19.9
`CaloMuonScore WP2`	95.23%	7.17 p.p.	6.36%	×7.6
`CaloMuonScore WP3`	95.14%	7.09 p.p.	3.94%	×12.2
`CaloMuonScore WP4`	95.14%	7.09 p.p.	3.94%	×12.2

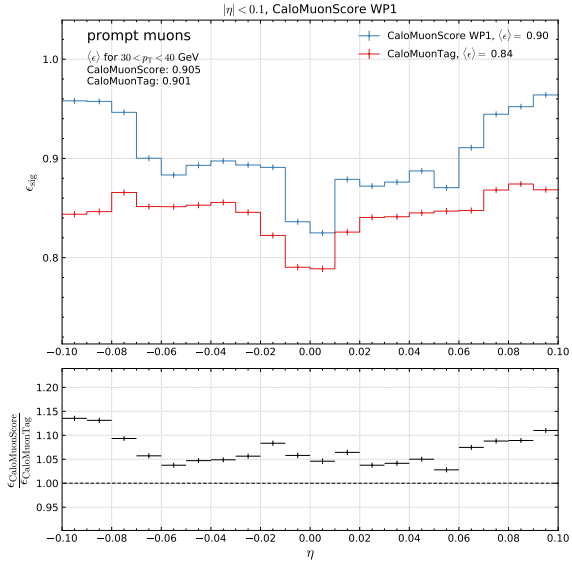
Table 6.11: ϵ_{sig} and ϵ_{bkg} for `CaloMuonTag` and four `CaloMuonScore` working points in $|\eta| < 0.1$ and $p_{\text{T}} \in [30, 40]$ GeV. The columns labelled 'improvement' indicate the performance improvement of `CaloMuonScore` over `CaloMuonTag` in terms of ϵ_{sig} and ϵ_{bkg}, respectively. The column labelled 'ϵ_{sig} improvement' lists the percentage-point improvement, whereas the one labelled 'ϵ_{bkg} improvement' indicates the improvement factor.



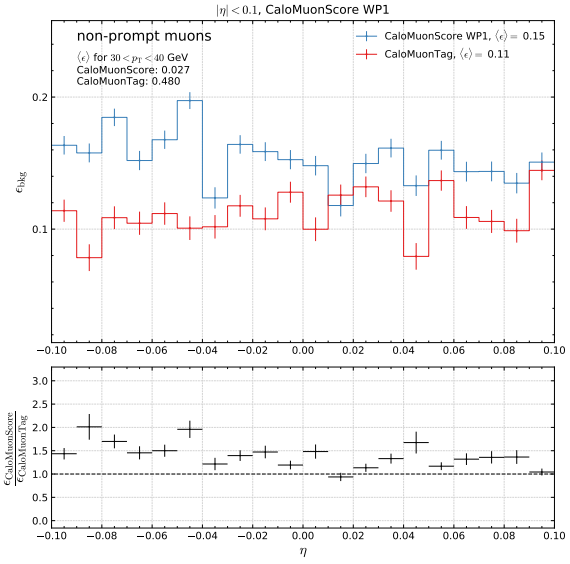
(a) ϵ_{sig} vs. p_{T} for CaloMuonTag and CaloMuonScore operating at WP1.



(b) ϵ_{bkg} vs. p_{T} for CaloMuonTag and CaloMuonScore operating at WP1.

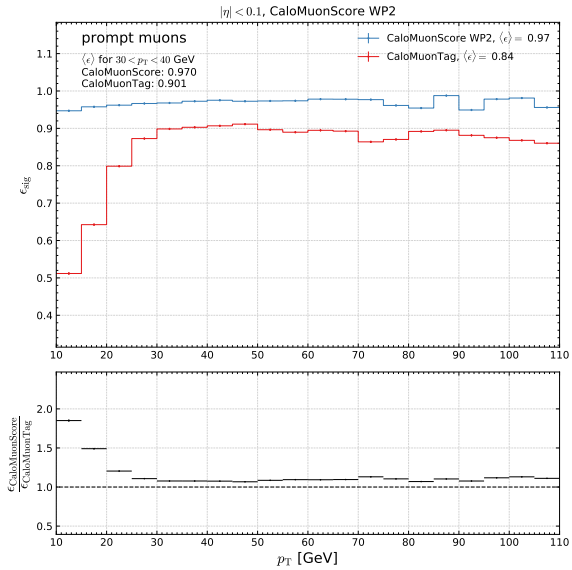


(c) ϵ_{sig} vs. η for CaloMuonTag and CaloMuonScore operating at WP1.

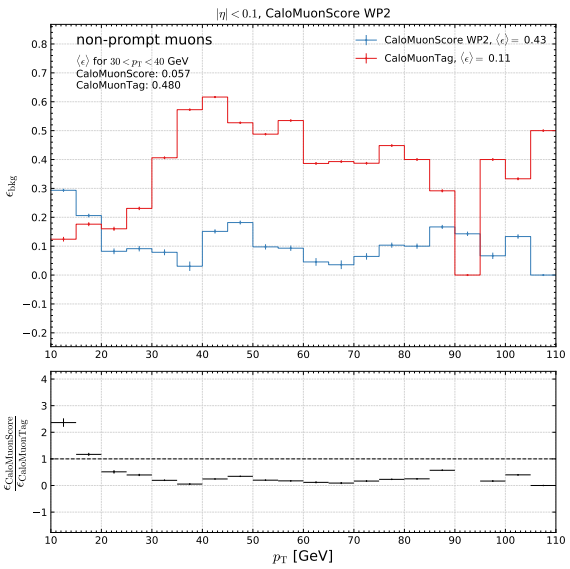


(d) ϵ_{bkg} vs. η for CaloMuonTag and CaloMuonScore operating at WP1.

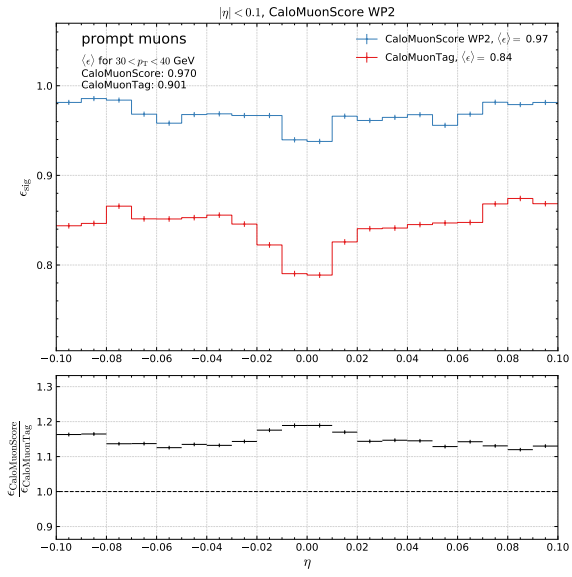
Figure 6.21: Signal and background efficiencies for CaloMuonTag (red) and CaloMuonScore at WP1 (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 0.1$. The bottom panel in each plot shows the ratio of the CaloMuonScore efficiency to the CaloMuonTag efficiency. The label 'prompt muons' used in these plots and below refers to true muons originating from a $Z \to \mu\mu$ decay. The label 'non-prompt muons' refers to particles that are either true muons not originating from the primary $Z \to \mu\mu$ interaction vertex, or hadrons that are misidentified as muons.



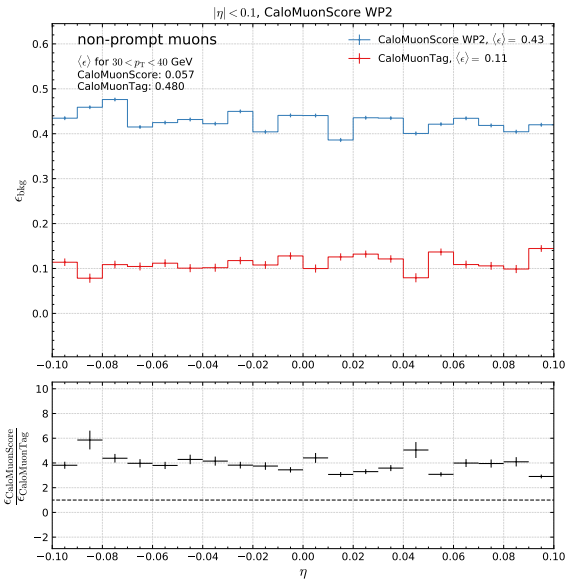
(a) ϵ_{sig} vs. p_{T} for CaloMuonTag and CaloMuonScore operating at WP2.



(b) ϵ_{bkg} vs. p_{T} for CaloMuonTag and CaloMuonScore operating at WP2.

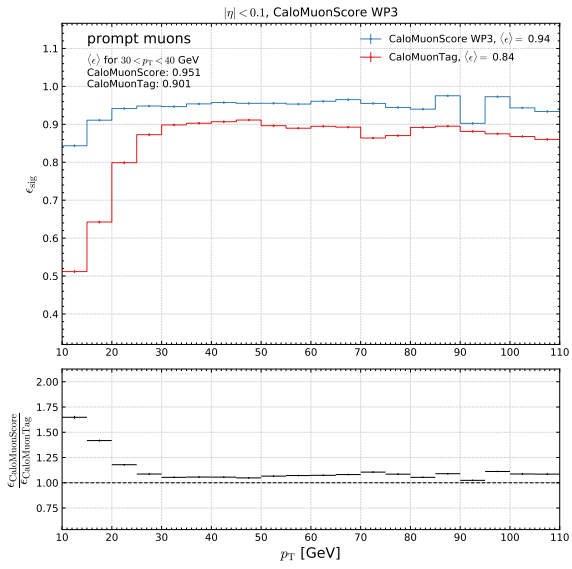


(c) ϵ_{sig} vs. η for CaloMuonTag and CaloMuonScore operating at WP2.

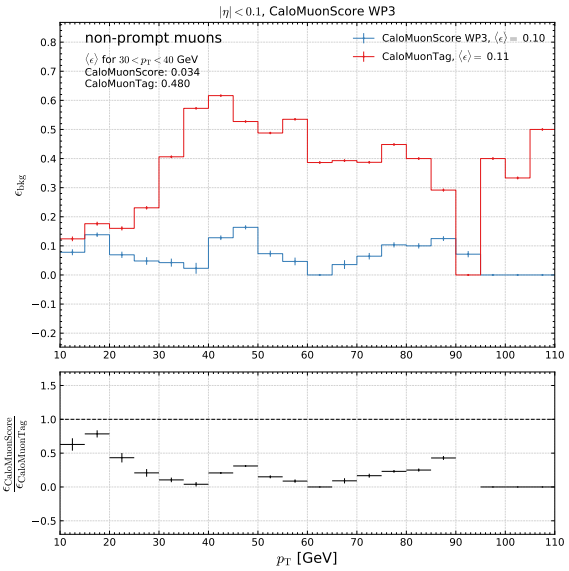


(d) ϵ_{bkg} vs. η for CaloMuonTag and CaloMuonScore operating at WP2.

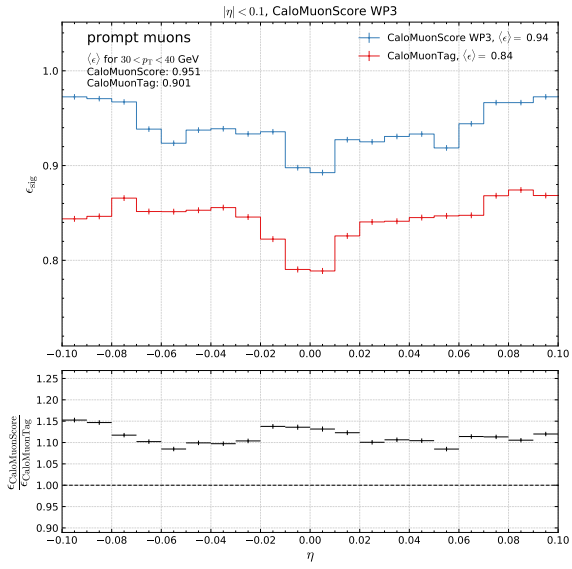
Figure 6.22: Signal and background efficiencies for CaloMuonTag (red) and CaloMuonScore at WP2 (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 0.1$. The bottom panel in each plot shows the ratio of the CaloMuonScore efficiency to the CaloMuonTag efficiency.



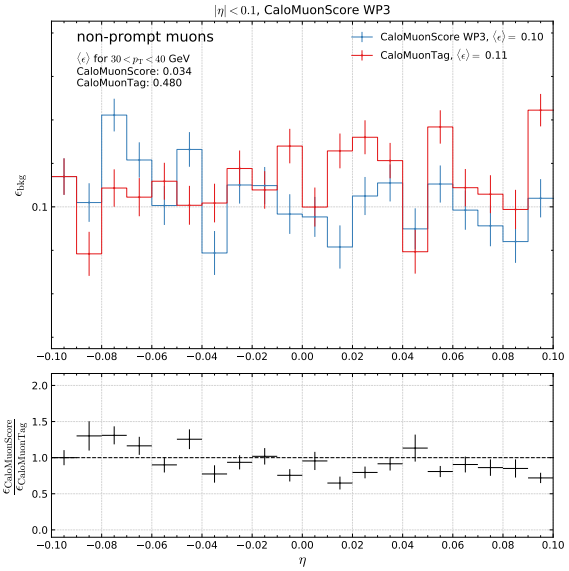
(a) ϵ_{sig} vs. p_T for CaloMuonTag and CaloMuonScore operating at WP3.



(b) ϵ_{bkg} vs. p_T for CaloMuonTag and CaloMuonScore operating at WP3.

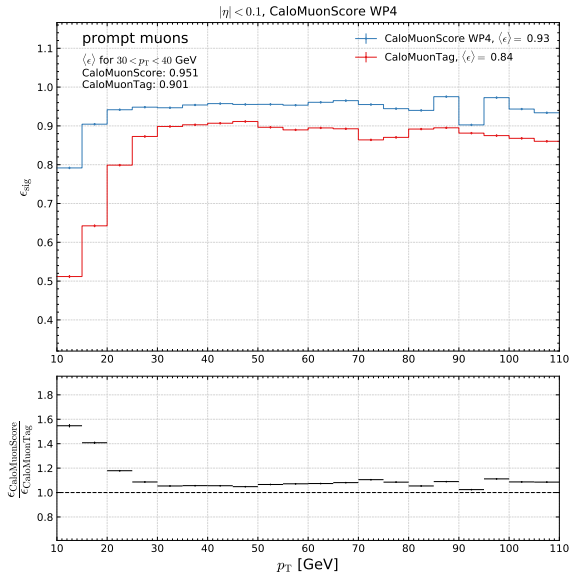


(c) ϵ_{sig} vs. η for CaloMuonTag and CaloMuonScore operating at WP3.

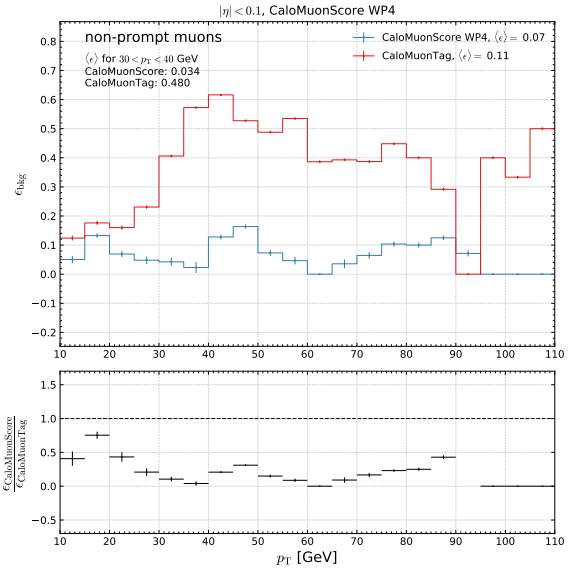


(d) ϵ_{bkg} vs. η for CaloMuonTag and CaloMuonScore operating at WP3.

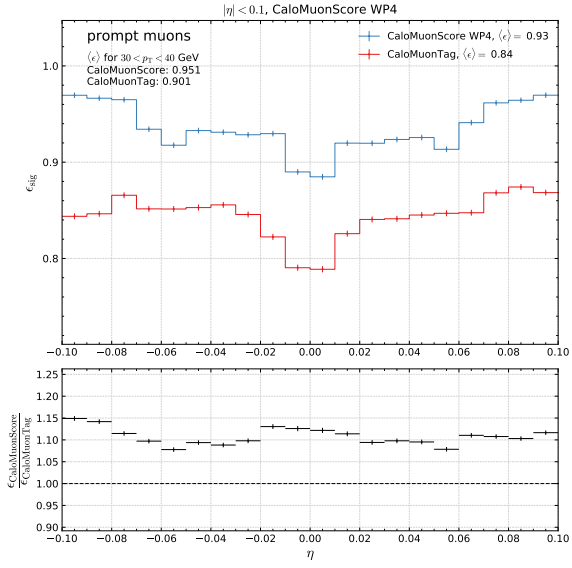
Figure 6.23: Signal and background efficiencies for CaloMuonTag (red) and CaloMuonScore at WP3 (blue) as a function of p_T (a, b) and η (c, d) in the region $|\eta| < 0.1$. The bottom panel in each plot shows the ratio of the CaloMuonScore efficiency to the CaloMuonTag efficiency.



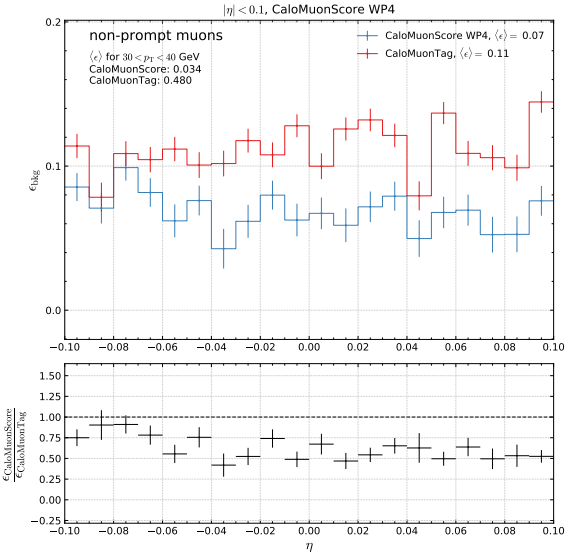
(a) ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.



(b) ϵ_{bkg} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.



(c) ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.



(d) ϵ_{bkg} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

Figure 6.24: Signal and background efficiencies for `CaloMuonTag` (red) and `CaloMuonScore` at `WP4` (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 0.1$. The bottom panel in each plot shows the ratio of the `CaloMuonScore` efficiency to the `CaloMuonTag` efficiency.

at WP1. The distribution showing the signal efficiency vs. η (Figure 6.25c) reveals that ϵ_{sig} obtained by CaloMuonScore is significantly higher than CaloMuonTag, and, furthermore, that the overall signal efficiency of CaloMuonScore is much flatter across the entire η range compared to that of CaloMuonTag. A significant result to emerge from this comparison is that the performance of CaloMuonScore in terms of signal efficiency is much less p_{T}-dependent than that of CaloMuonTag (see Figure 6.25a).

A similar overview is shown for WP2 in Figure 6.26. The same observations as for WP1 regarding trends in signal efficiency hold here. Compared with WP1, this working point, by design, has a very similar ϵ_{sig} to CaloMuonTag while achieving a much better rejection of fakes than CaloMuonTag. It is worth noting that the background rejection for both working points is much more p_{T}-dependent than that of CaloMuonTag.

The strong p_{T} dependence of ϵ_{bkg} for WP1 and WP2 for $p_{\text{T}} < 20$ GeV is greatly reduced in WP3 and WP4 (see Figures 6.27 and 6.28).

All four working points deliver excellent performance up to $|\eta| < 1$ with significantly higher signal efficiencies than those achieved by CaloMuonTag.

6.7.5 Bias and variance of CaloMuonScore

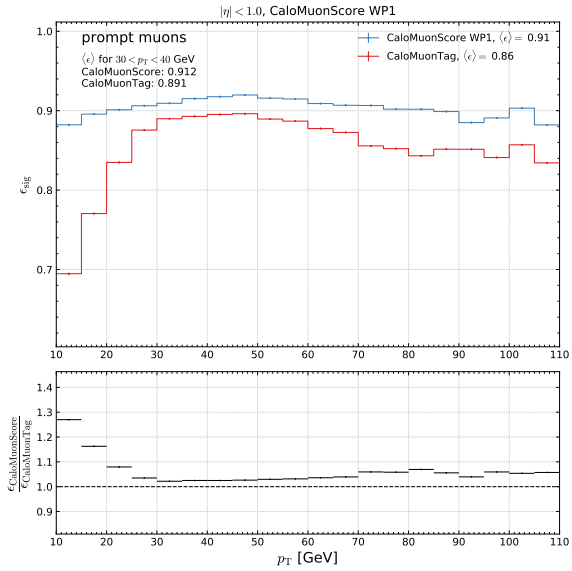
In order to assess the bias and variance of CaloMuonScore, a k-fold cross-validation (Section 4.4.1) was performed with $k = 5$. The k-fold cross-validation was repeated ten times. Using Equations (4.26) and (4.27), the following estimate of the prediction error, PE, was obtained:

$$\text{PE} = (1.87 \pm 0.14)\%.$$

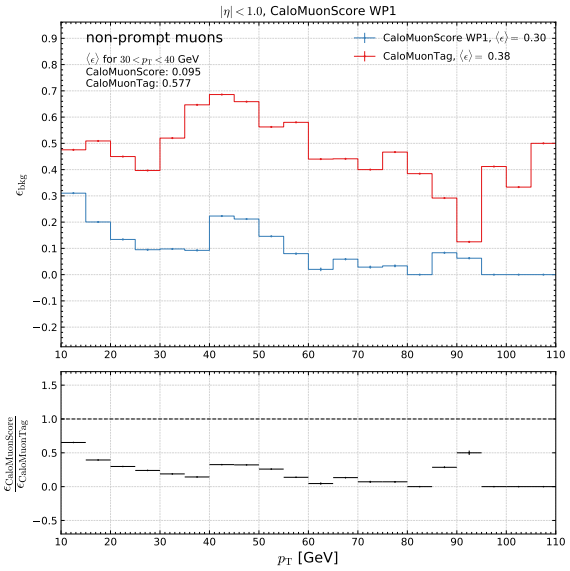
This result indicates that the model has low bias and very low variance. The prediction errors across all ten runs, as well as individual ROC curves for one k-fold cross-validation run are shown in Figure 6.29.

6.7.6 Effect of the training data size

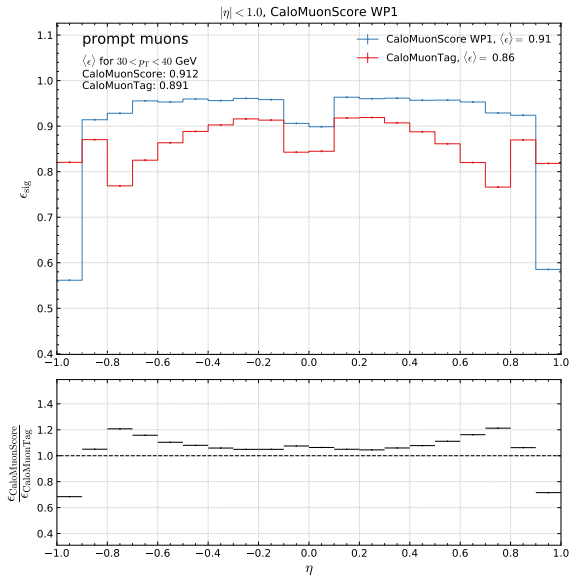
In order to assess whether CaloMuonScore used a sufficient number of training examples, the model performance was evaluated as a function of size of the training set. Twenty-two experiments were conducted, in which different fractions of the total number of events (Table 6.9) were used, varying from 1% to 100%. The chosen



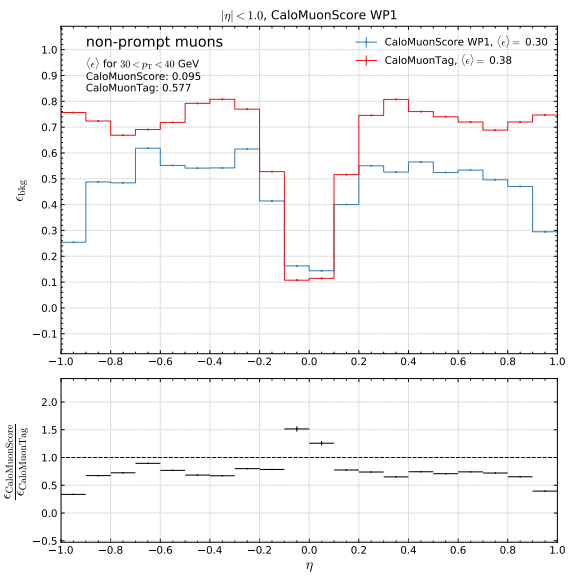
(a) ϵ_{sig} vs. p_{T} for CaloMuonTag and CaloMuonScore operating at WP1.



(b) ϵ_{bkg} vs. p_{T} for CaloMuonTag and CaloMuonScore operating at WP1.

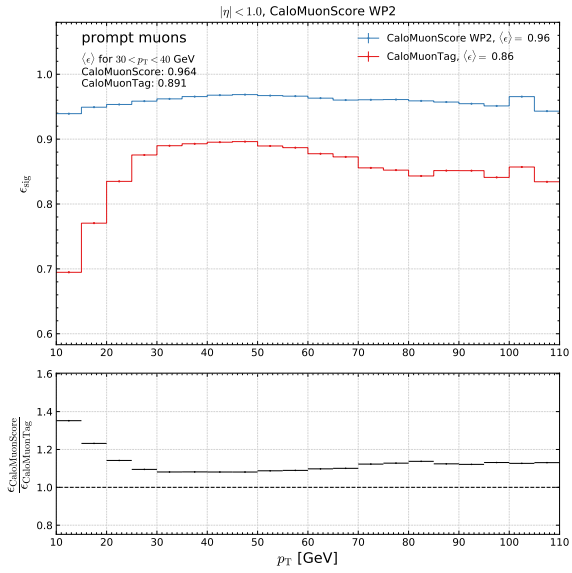


(c) ϵ_{sig} vs. η for CaloMuonTag and CaloMuonScore operating at WP1.

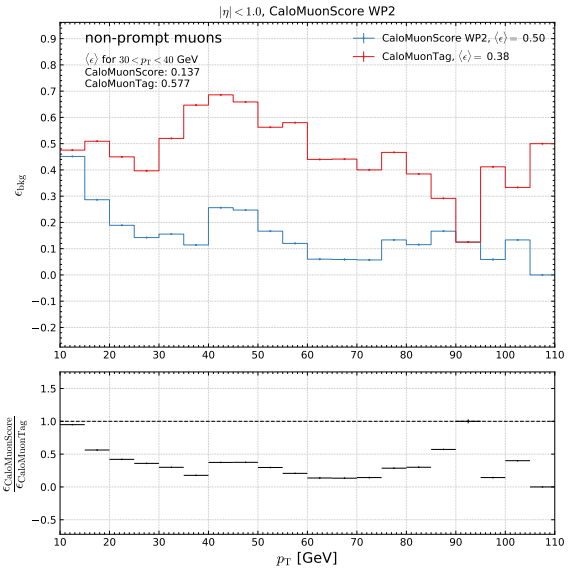


(d) ϵ_{bkg} vs. η for CaloMuonTag and CaloMuonScore operating at WP1.

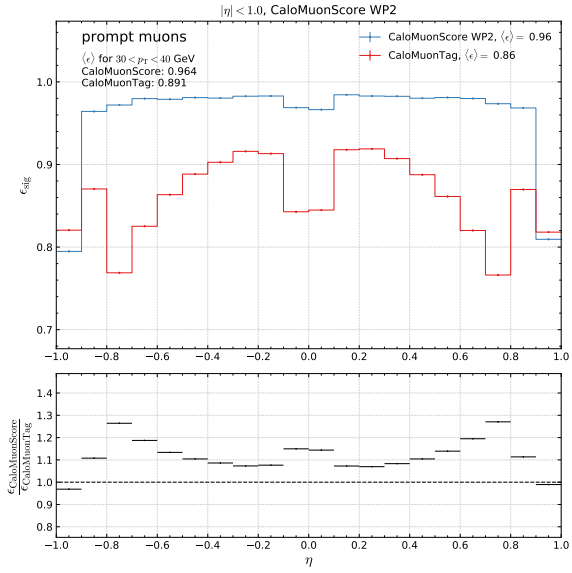
Figure 6.25: Signal and background efficiencies for CaloMuonTag (red) and CaloMuonScore at WP1 (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 1$. The bottom panel in each plot shows the ratio of the CaloMuonScore efficiency to the CaloMuonTag efficiency.



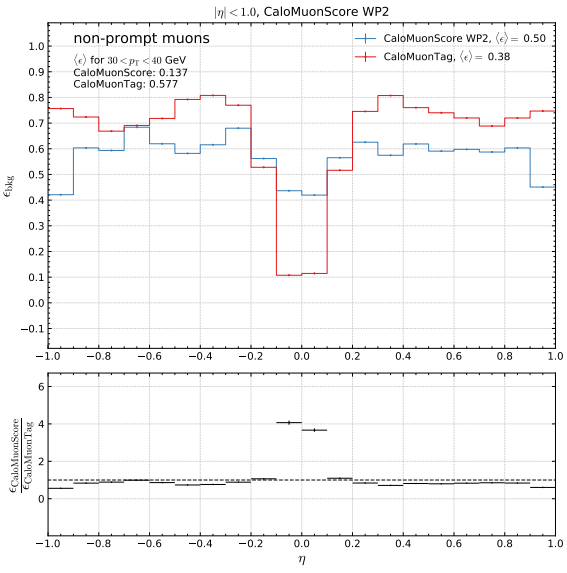
(a) ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.



(b) ϵ_{bkg} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.

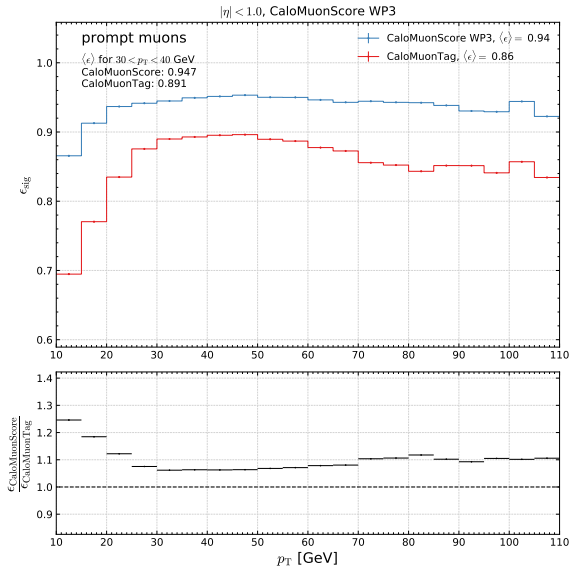


(c) ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.

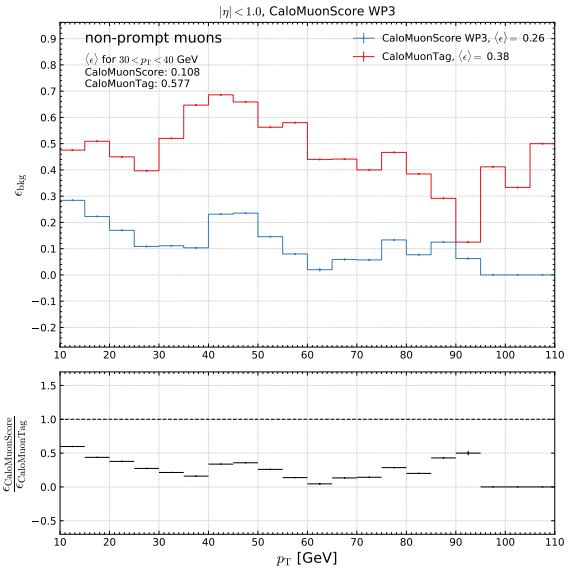


(d) ϵ_{bkg} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.

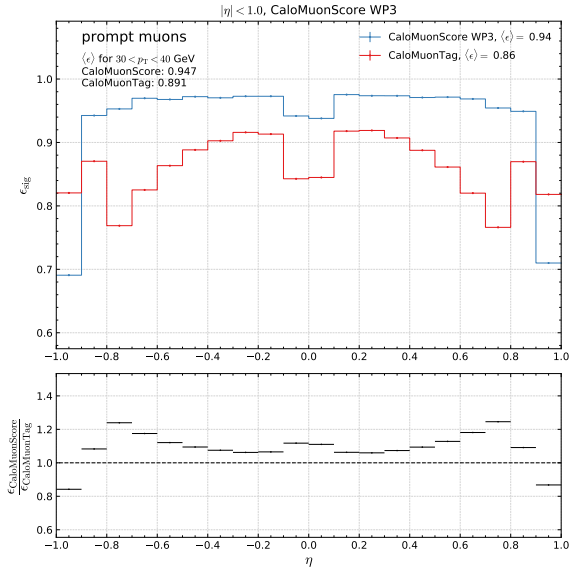
Figure 6.26: Signal and background efficiencies for `CaloMuonTag` (red) and `CaloMuonScore` at `WP2` (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 1$. The bottom panel in each plot shows the ratio of the `CaloMuonScore` efficiency to the `CaloMuonTag` efficiency.



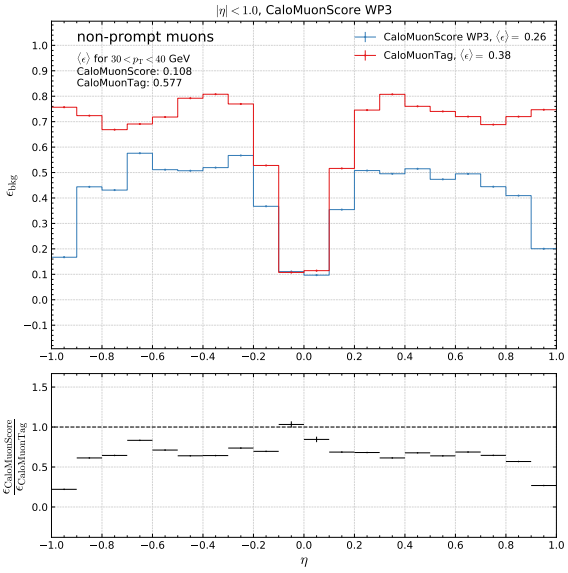
(a) ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.



(b) ϵ_{bkg} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.

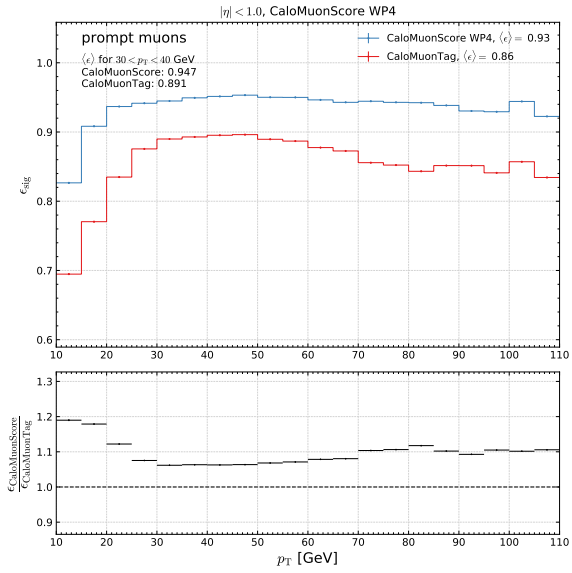


(c) ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.

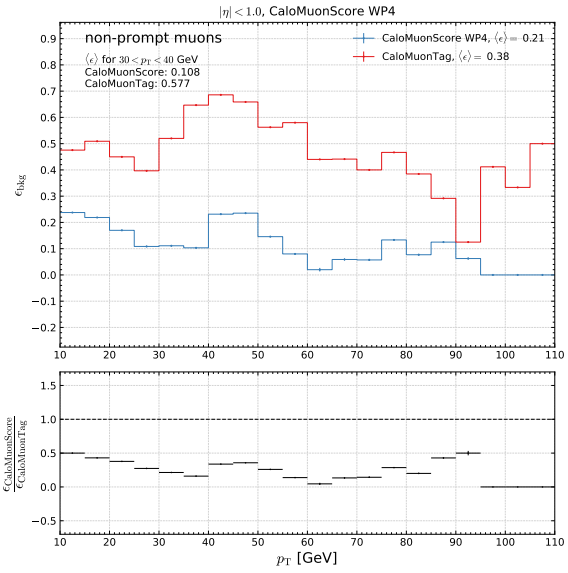


(d) ϵ_{bkg} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.

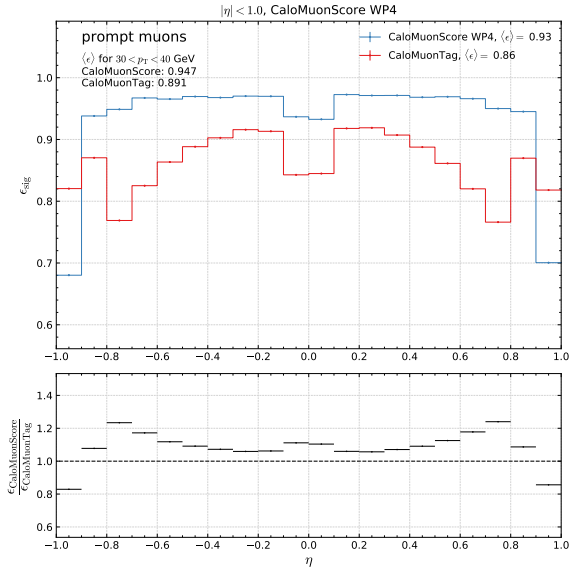
Figure 6.27: Signal and background efficiencies for `CaloMuonTag` (red) and `CaloMuonScore` at `WP3` (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 1$. The bottom panel in each plot shows the ratio of the `CaloMuonScore` efficiency to the `CaloMuonTag` efficiency.



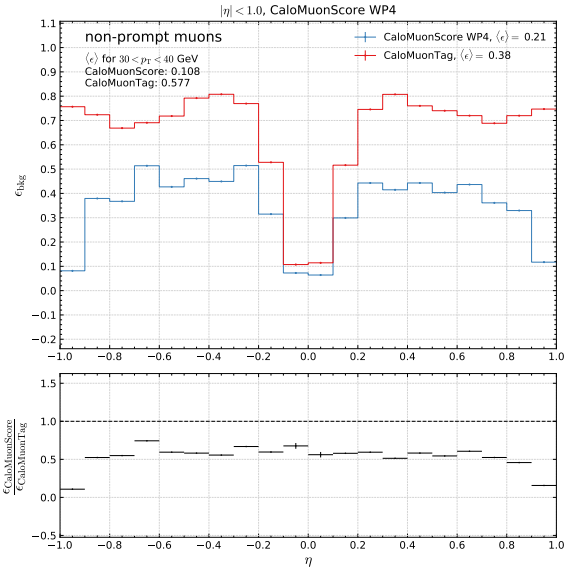
(a) ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.



(b) ϵ_{bkg} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

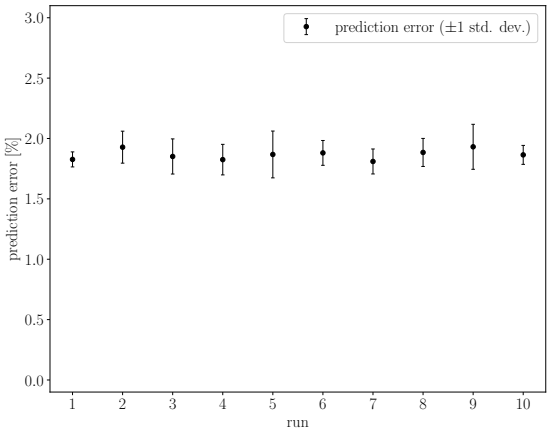


(c) ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

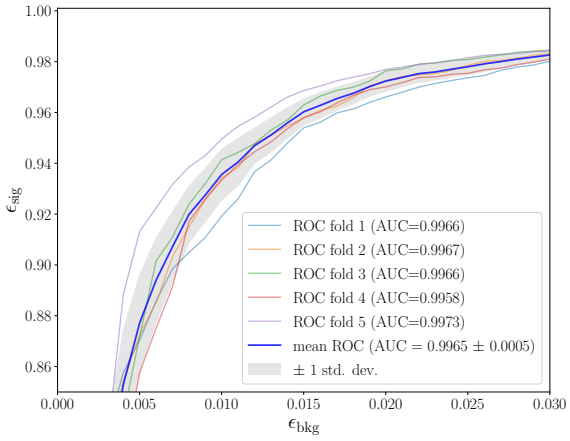


(d) ϵ_{bkg} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

Figure 6.28: Signal and background efficiencies for `CaloMuonTag` (red) and `CaloMuonScore` at `WP4` (blue) as a function of p_{T} (a, b) and η (c, d) in the region $|\eta| < 1$. The bottom panel in each plot shows the ratio of the `CaloMuonScore` efficiency to the `CaloMuonTag` efficiency.



(a) Mean prediction error over 10 runs.



(b) Individual and mean Receiver operating characteristic (ROC) curves for run 5.

Figure 6.29: (a) Mean prediction error over ten k-fold cross-validation runs using $k = 5$. The error bars indicate one standard deviation from the mean across the five folds of a given run. (b) Receiver operating characteristic (ROC) curves for all folds in a randomly chosen run (run number 5). The signal efficiency is shown on the y-axis, while the background efficiency is shown on the x-axis. 'AUC' refers to the area under the ROC curve. The mean ROC curve is shown in blue with an error band covering one standard deviation from the mean (grey).

fractions at the low end are 1% and 3%, followed by regularly spaced 5% intervals from 5% to 100%. The experiment at each sampling point was repeated ten times with a different set of events, and each time 20% of events were reserved for testing. In each repetition, the sets were randomly chosen from all available events; this may lead to some events not appearing at all in the training set in any of the repetitions (especially for sampling points at the lower end), or to events appearing in more than one repetitions (especially for sampling points at the higher end). Figure 6.30 summarises the results, showing the prediction error on the test set as a function of the fraction of training data available for training. It is clear from the plot that the model performance saturates at around 50% of all training events, establishing confidence that the available training statistics are more than sufficient to learn the task. Therefore, it is reasonable to suspect that generating additional training data would not lead to a significant improvement in model performance.

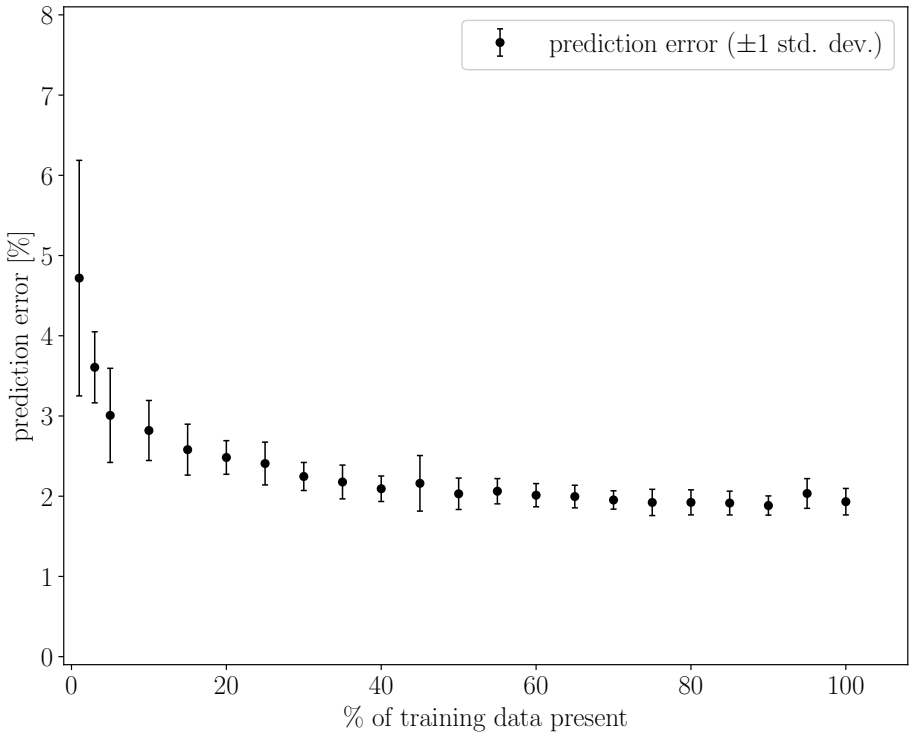
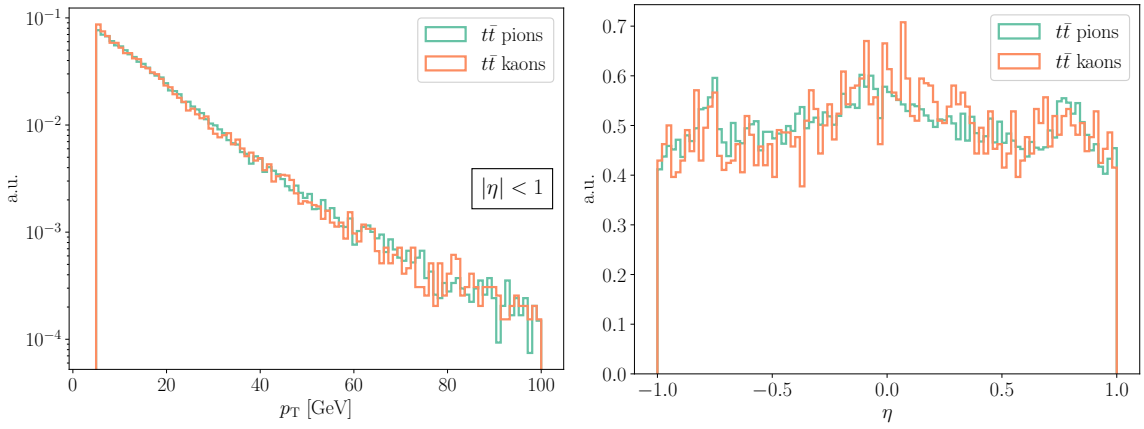


Figure 6.30: Prediction error as a function of the fraction of all training events that were used during training. The error bars indicate the standard deviation on the prediction error, measured over ten runs.

6.7.7 Generalisation to kaons

Having been trained only on pions as the background component, it is important to verify the performance of `CaloMuonScore` on other background sources. Referring to Figure 6.1b, it is apparent that the second most frequently occurring particles in the $t\bar{t}$ sample are kaons. In this section, the performance of `CaloMuonScore` is evaluated on kaons from the $t\bar{t}$ sample as background, and, as before, muons from the $Z \to \mu\mu$ sample as signal.

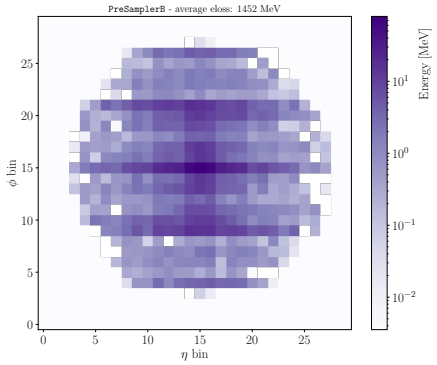
As a first step, it is instructive to investigate whether or not there are any qualitative differences between the pion and kaon components of the $t\bar{t}$ sample. Figure 6.31 compares the p_{T} and η distributions between pions and kaons coming from the $t\bar{t}$ sample. Both plots indicate that there are no notable kinematic differences between the two particle types. Furthermore, there is no appreciable difference in the average calorimeter images the two background types produce. This is illustrated in Figure 6.32, which shows the average of 20,000 pion and kaon images in three different calorimeter layers, indicating that the two particle types' calorimetric signatures are comparable. The calorimetric energy-deposit patterns are further examined in Figure 6.33, which compares the average energy deposited by 20,000 pions with that deposited by 20,000 kaons along η and ϕ in the same three calorimeter layers. The plots indicate a large degree of similarity in calorimetric signature between pions and kaons.



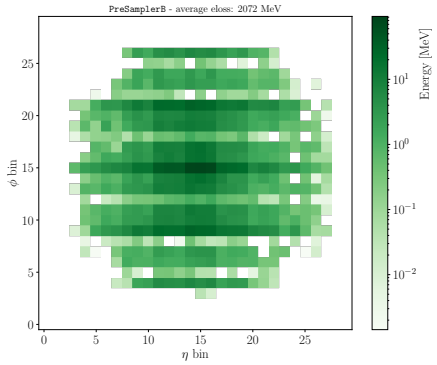
(a) p_{T} distribution for pions (green) and kaons (orange).

(b) η distribution for pions (green) and kaons (orange).

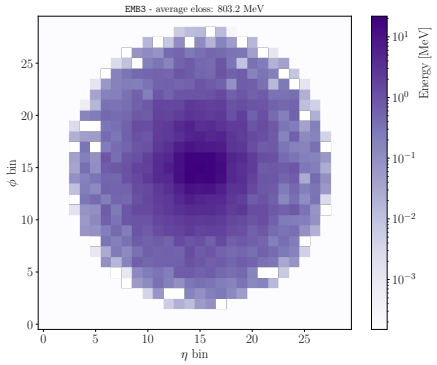
Figure 6.31: Distributions of muon p_{T} (a) and η (b) for pions (green) and kaons (orange), both originating from the $t\bar{t}$ sample. The distributions are normalised to unit area to facilitate comparison (a.u.).



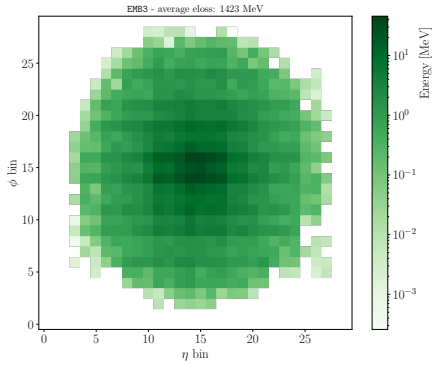
(a) Average of pion images in layer 1.



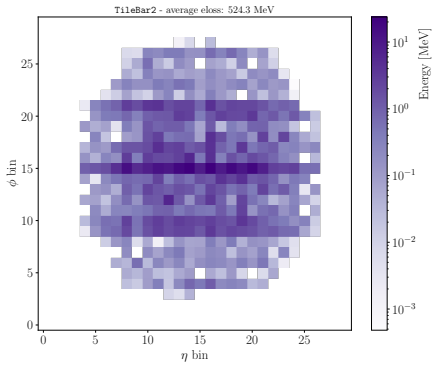
(b) Average of kaon images in layer 1.



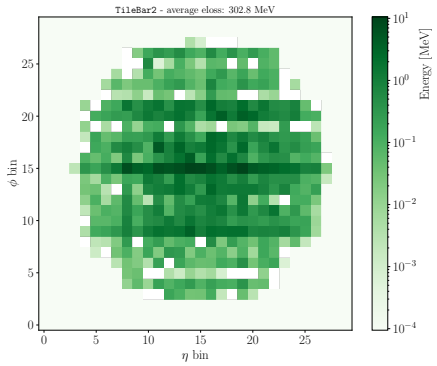
(c) Average of pion images in layer 4.



(d) Average of kaon images in layer 4.

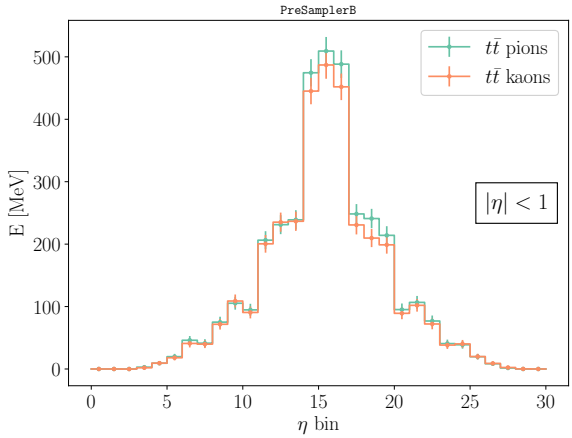


(e) Average of pion images in layer 7.

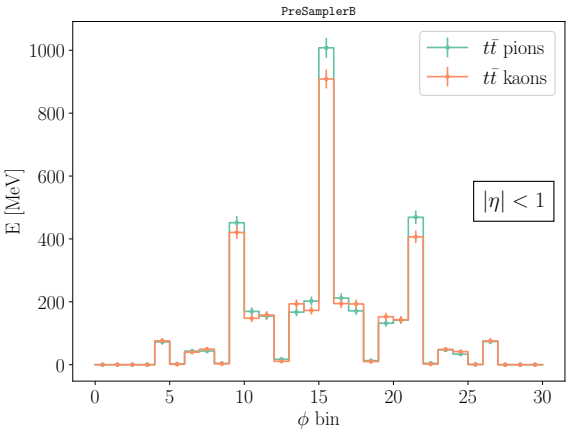


(f) Average of kaon images in layer 7.

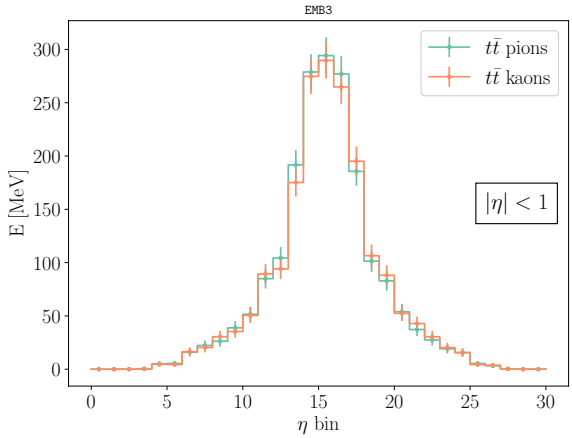
Figure 6.32: Average images produced by 20,000 pions and 20,000 kaons in different calorimeter layers. The pion images are shown on the left (a, c, and e), whereas the kaon images are shown in the right (b, d, and f). The plots are 2D $\phi - \eta$ grids of 30×30 px showing the binned energy losses.



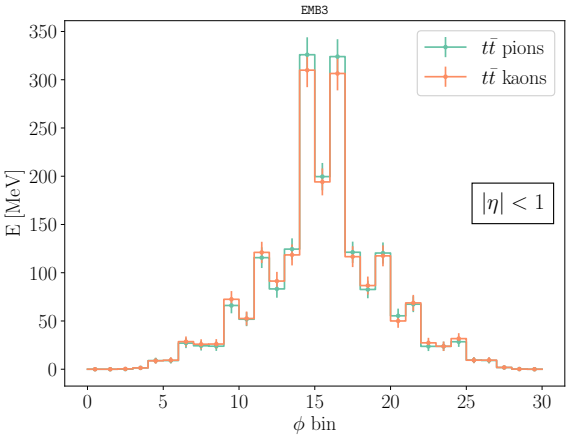
(a) Energy deposit in η in layer 1.



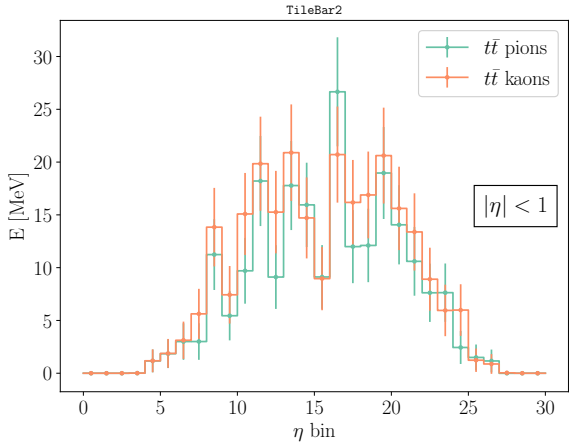
(b) Energy deposit in ϕ in layer 1.



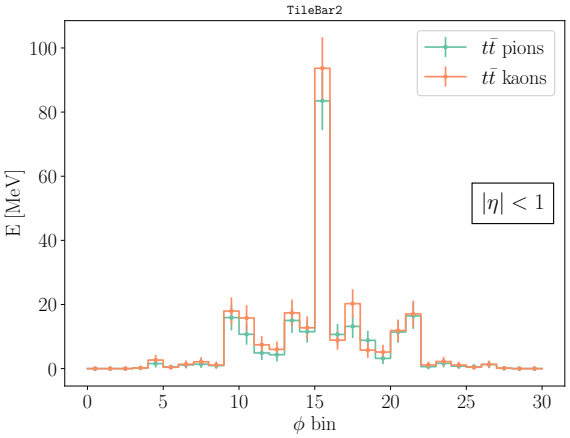
(c) Energy deposit in η in layer 4.



(d) Energy deposit in ϕ in layer 4.



(e) Energy deposit in η in layer 7.



(f) Energy deposit in ϕ in layer 7.

Figure 6.33: Mean energy-deposit signatures of 20,000 pions and 20,000 muons from the $t\bar{t}$ sample in three calorimeter layers. The comparison is shown separately in bins of η (a, c, e) and ϕ (b, d, f). The error bars shown are statistical only.

The Figure 6.34 compares the output histogram of CaloMuonScore between pions (green) and kaons (orange). The distributions are virtually indistinguishable between the two particle types, demonstrating that CaloMuonScore is equally as effective in discriminating prompt from non-prompt muons originating from pion decays as it is in discriminating prompt from non-prompt muons of kaonic origin.

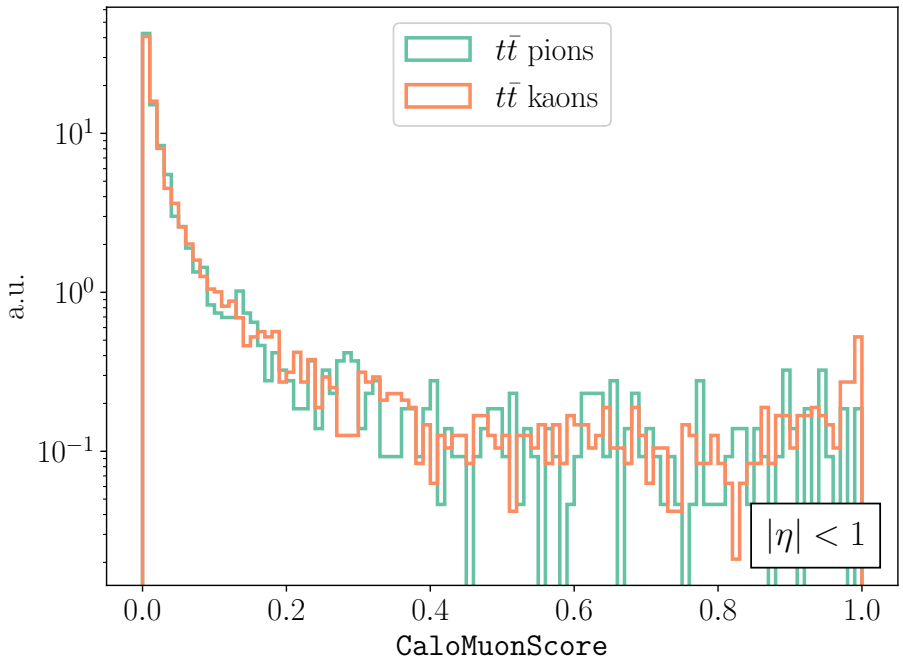
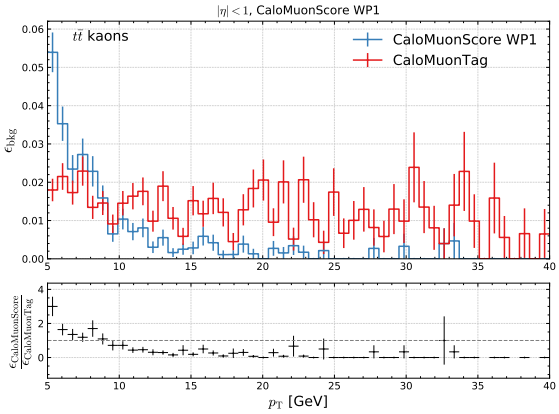


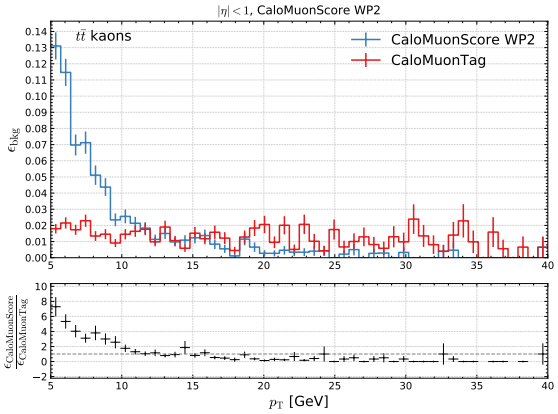
Figure 6.34: Distribution of CaloMuonScore in $|\eta| < 1$ for pions (green) and kaons (orange) from the $t\bar{t}$ sample. The distributions are normalised to unit area to facilitate comparison (a.u.).

Furthermore, it remains to be seen how the probabilities of faking a muon compare between pions and kaons. Table 6.12 shows the background efficiency of kaons being misidentified as muons, separately for CaloMuonTag and CaloMuonScore at all four working points. CaloMuonScore achieves a significant improvement in fake suppression compared to CaloMuonTag at all working points.

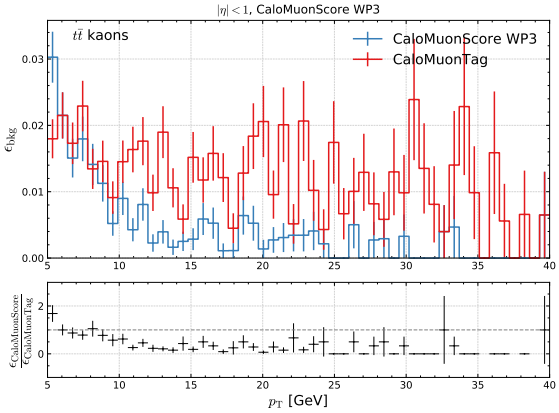
The dependence of the kaon-background efficiency on p_T and η is presented in Figures 6.35 and 6.36 for CaloMuonTag and CaloMuonScore at all four working points. As is the case for pions, the background efficiency has a stronger dependence on p_T than CaloMuonTag; it shows, however, little dependence on η up to $\eta = \pm 0.7$ and an increase in background efficiency beyond that.



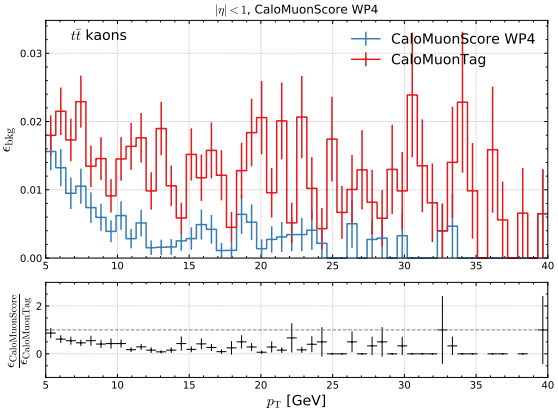
(a) Kaon ϵ_{bkg} as a function p_T for `CaloMuonTag` and `CaloMuonScore` at `WP1` in $|\eta| < 1$.



(b) Kaon ϵ_{bkg} as a function p_T for `CaloMuonTag` and `CaloMuonScore` at `WP2` in $|\eta| < 1$.

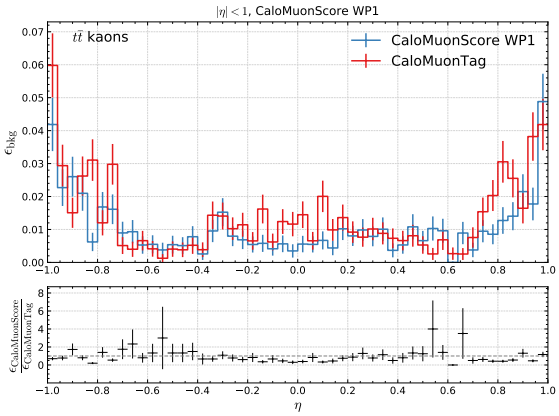


(c) Kaon ϵ_{bkg} as a function p_T for `CaloMuonTag` and `CaloMuonScore` at `WP3` in $|\eta| < 1$.

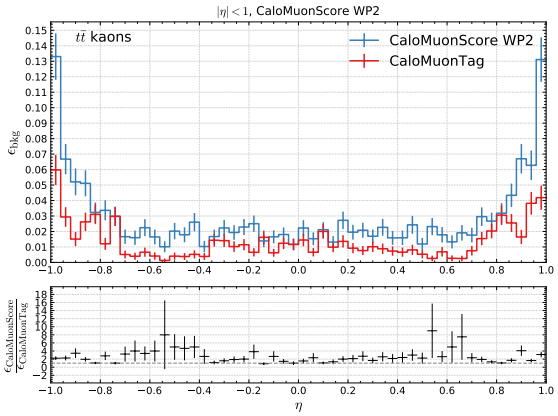


(d) Kaon ϵ_{bkg} as a function p_T for `CaloMuonTag` and `CaloMuonScore` at `WP4` in $|\eta| < 1$.

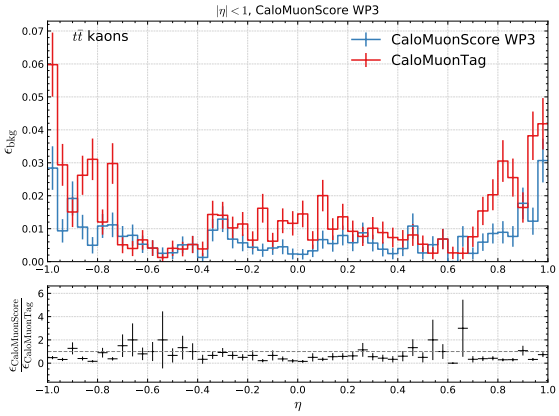
Figure 6.35: Background efficiencies for kaons as a function of kaon η for $|\eta| < 1$ for `CaloMuonTag` (red) and `CaloMuonScore` operating at four different working points (blue).



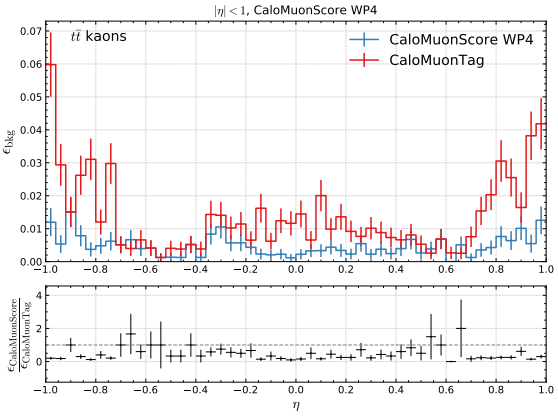
(a) Kaon ϵ_{bkg} as a function η for `CaloMuonTag` and `CaloMuonScore` at `WP1` in $|\eta| < 1$.



(b) Kaon ϵ_{bkg} as a function η for `CaloMuonTag` and `CaloMuonScore` at `WP2` in $|\eta| < 1$.



(c) Kaon ϵ_{bkg} as a function η for `CaloMuonTag` and `CaloMuonScore` at `WP3` in $|\eta| < 1$.



(d) Kaon ϵ_{bkg} as a function η for `CaloMuonTag` and `CaloMuonScore` at `WP4` in $|\eta| < 1$.

Figure 6.36: Background efficiencies of kaons as a function of kaon p_T for `CaloMuonTag` (red) and `CaloMuonScore` operating at four different working points (blue) in $|\eta| < 1$. The bottom panel in each plot shows the ratio of the `CaloMuonScore` efficiency to the `CaloMuonTag` efficiency.

classifier	ϵ_{bkg}	ϵ_{bkg} improvement
`CaloMuonTag`	1.04%	—
`CaloMuonScore WP1`	0.07%	×14.9
`CaloMuonScore WP2`	0.16%	×6.5
`CaloMuonScore WP3`	0.14%	×7.4
`CaloMuonScore WP4`	0.14%	×7.4

Table 6.12: Kaon ϵ_{bkg} for `CaloMuonTag` and four `CaloMuonScore` working points in $|\eta| < 1$ and $p_{\text{T}} \in [30, 40]$ GeV. The column labelled 'ϵ_{bkg} improvement' indicates the performance improvement factor of `CaloMuonScore` over `CaloMuonTag` in terms of ϵ_{bkg}.

6.8 Tag-and-probe results

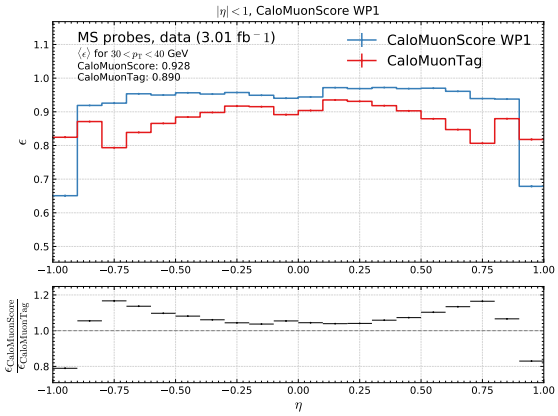
The performance results shown in the previous sections are entirely based on simulated events. For a new classification algorithm to be used in physics analyses, its performance needs to be carefully evaluated on data. In this chapter, the results of a tag-and-probe analysis, as described in Section 6.1.1, are presented.

The tag-and-probe results are based on 3.01 fb^{-1} of data recorded by ATLAS in 2017. 20 million $Z \to \mu\mu$ events simulated using `Powheg + Pythia` are used as the MC component. The results are presented separately in two detector regions: $|\eta| < 1$, using MS probes, and $|\eta| < 0.1$, using ID probes.

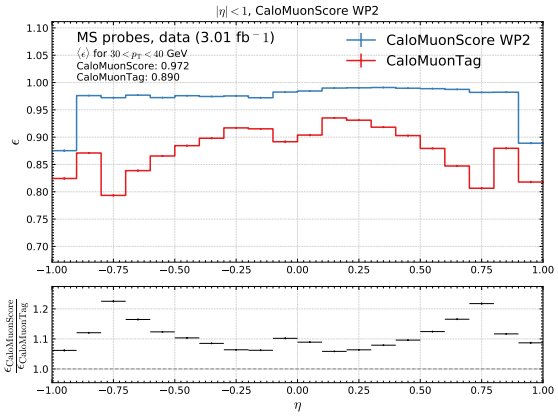
6.8.1 Tag-and-probe results in $|\eta| < 1$

6.8.1.1 Overall performance

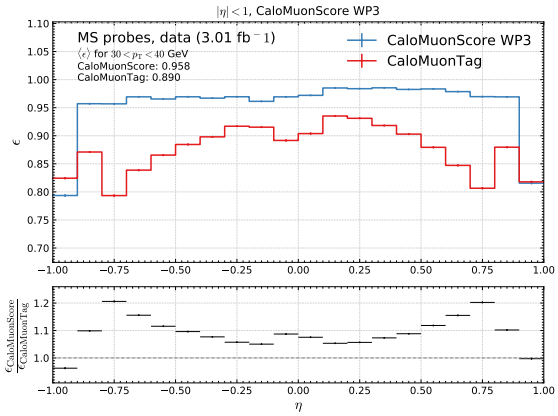
The signal efficiency of `CaloMuonScore` at all four working points is compared with that of `CaloMuonTag`, separately as a function of η in Figure 6.37 and as a function of p_{T} in Figure 6.38. The efficiencies are measured on 3.01 fb^{-1} of data and provide the most direct in-situ measurement of the performance of the new method. N.B.: The tag-and-probe analysis only allows for a measurement of the signal efficiency; in order to estimate the background efficiency, one has to rely on the MC-only numbers presented in Section 6.7.3. The observed signal efficiencies generally agree very well with the numbers measured on simulation in the same section.



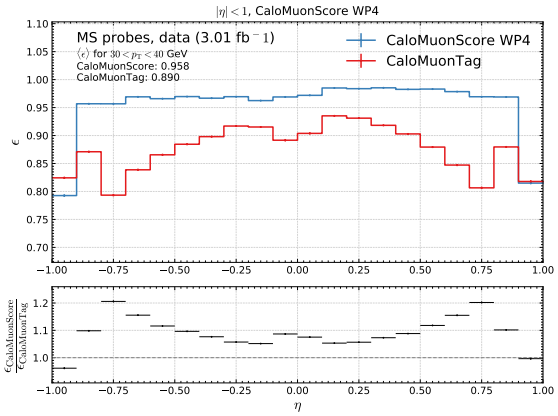
(a) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP1`.



(b) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.

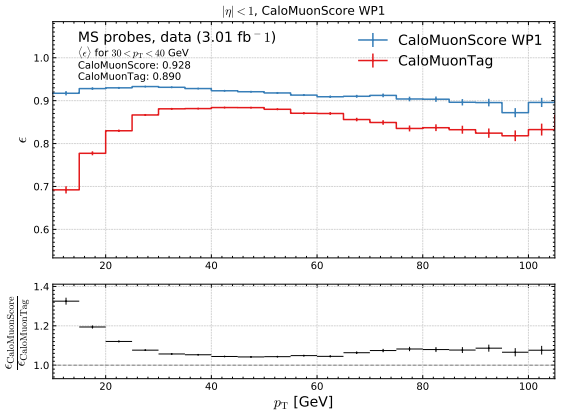


(c) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.

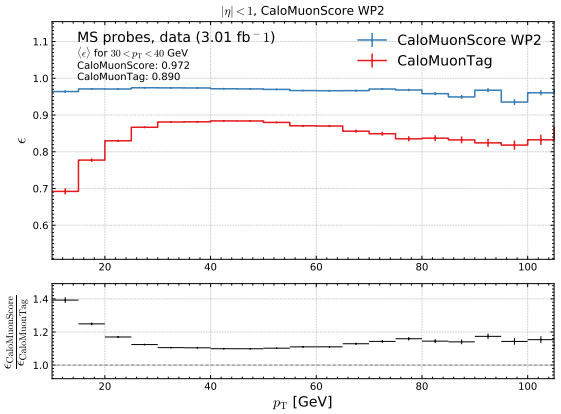


(d) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

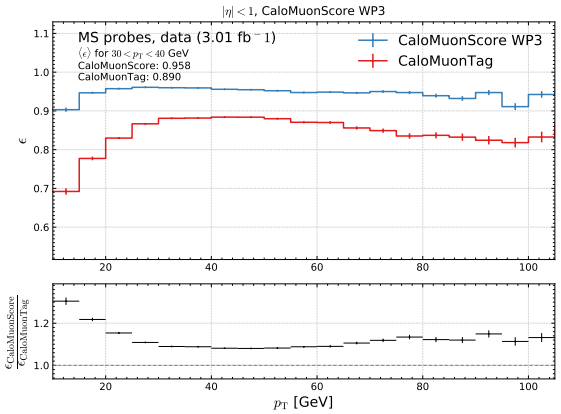
Figure 6.37: Tag-and-probe signal efficiencies vs. η for `CaloMuonTag` vs. `CaloMuonScore` at four working points using MS probes. The efficiencies are measured on 3.01 fb^{-1} of data. The efficiencies achieved by the different `CaloMuonScore` working points are shown in blue, while those achieved by `CaloMuonTag` are shown in red. The panels show the ratio of the `CaloMuonScore` efficiencies to the `CaloMuonTag` efficiencies. The error bars shown are statistical only.



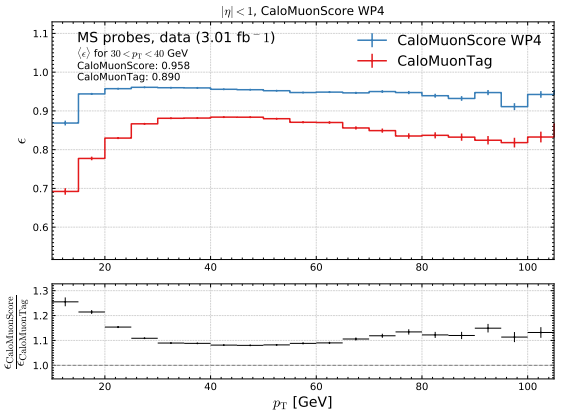
(a) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonTag` and `CaloMuonScore` operating at `WP1`.



(b) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.



(c) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.



(d) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

Figure 6.38: Tag-and-probe signal efficiencies vs. p_T in $|\eta| < 1$ for `CaloMuonTag` vs. `CaloMuonScore` at four working points using MS probes. The efficiencies are measured on 3.01 fb^{-1} of data. The efficiencies achieved by the different `CaloMuonScore` working points are shown in blue, while those achieved by `CaloMuonTag` are shown in red. The panels show the ratio of the `CaloMuonScore` efficiencies to the `CaloMuonTag` efficiencies. The error bars shown are statistical only.

6.8.1.2 Reduced `CaloMuonScore` efficiencies in three detector regions

From the efficiencies as a function of η (Figure 6.37), two notable regions of slightly reduced `CaloMuonScore` efficiencies emerge. First, the working points are significantly less efficient in the outermost bins near $\eta = \pm 1$. For a particle occurring just inside the acceptance range of `CaloMuonScore`, i.e. at $\eta = \pm 1$, only roughly half of its energy is deposited within the calorimeter cells accessible to `CaloMuonScore`. This leads to significantly lowered classification confidences, manifesting in a `CaloMuonScore` distribution that is shifted towards lower values.

Second, it is apparent that the `CaloMuonScore` efficiencies are slightly higher for $\eta \geq 0$ than $\eta < 0$. The reason for this is a dead tile calorimeter module located in the region spanned by $-1 \leq \eta \leq 0$ and $-0.2 \leq \phi \leq -0.1$. The module, named 'LBC63', experienced an severe electronic defect during data-taking in 2017, and was subsequently switched off. In a process referred to as *masking*, those Monte Carlo samples produced so as to match data-taking conditions in 2017 also have this part of the calorimeter switched off (i.e. the energy readout in the cells within LBC63 is set to zero). The tag-and-probe analysis at hand is based on data recorded in 2017, and consequently uses a Monte Carlo production that replicates the 2017 detector conditions. As a result, the calculation of `CaloMuonScore` for particles penetrating LBC63 is severely affected. The module went back online in 2018, and any subsequently recorded data, as well as simulation matching the 2018 data-taking period, are not affected.

Both regions of lowered `CaloMuonScore` are illustrated in Figure 6.39, which shows the mean value `CaloMuonScore` in small bins of ϕ and η. The two vertical bands on the far left and the far right mark the end of the acceptance region. In these bands, the mean value of `CaloMuonScore` is around 0.8. The horizontal strip near $\phi = -0.1$ and $\eta \leq 0$ clearly shows the region corresponding to LBC63. Here, the mean value of `CaloMuonScore` is around 0.3. Elsewhere in $\phi - \eta$ space, the mean value of `CaloMuonScore` is 0.97. The empty bins at very central η values, the so-called crack region, indicate that no events have been recorded; this is due to a partial lack of MS instrumentation[9].

The shift in the `CaloMuonScore` distribution towards lower values for η close to ± 1 is detailed in Figure 6.40a, which shows two marginal slices of `CaloMuonScore` in η. The red histogram, showing the marginal distribution of Figure 6.39 for

[9]The tag-and-probe analysis in the $|\eta| < 1$ region uses MS probes, and is therefore limited to measuring efficiencies in regions where the MS is instrumented. The analysis in the $|\eta| < 0.1$ region uses ID probes and does not exhibit these gaps in acceptance.

$0.98 < |\eta| < 1.00$, is shifted significantly to the left of the blue histogram, showing a marginal slice for $0.00 < |\eta| < 0.02$. Similarly, Figure 6.40b shows the distribution of `CaloMuonScore` in two $\phi - \eta$ regions. The red histogram shows the region in ϕ and η corresponding to LBC63, while the blue histogram shows the same slice in ϕ but on the opposite side in η. The histograms illustrate the severe effect of the masked module on the output of `CaloMuonScore`.

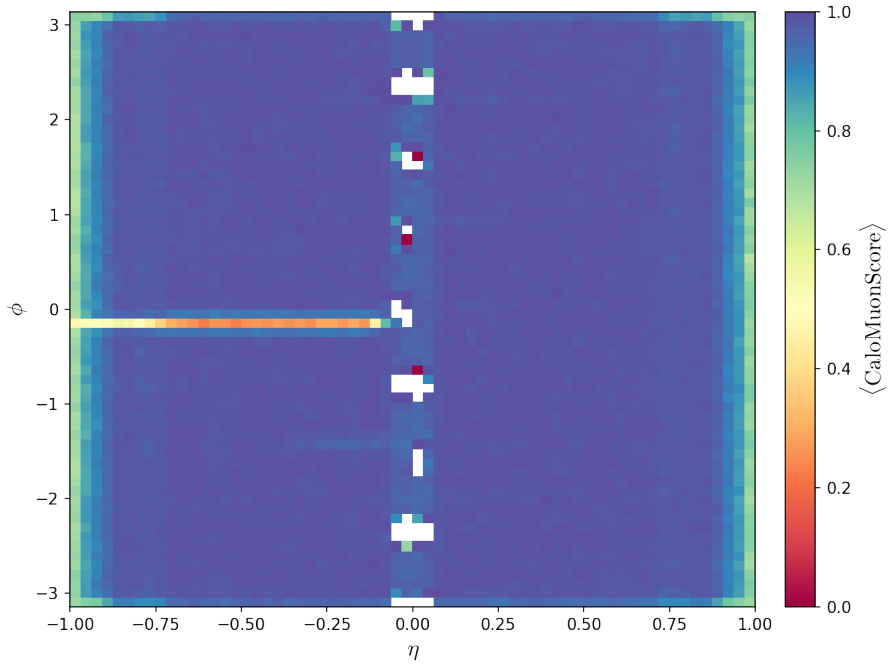
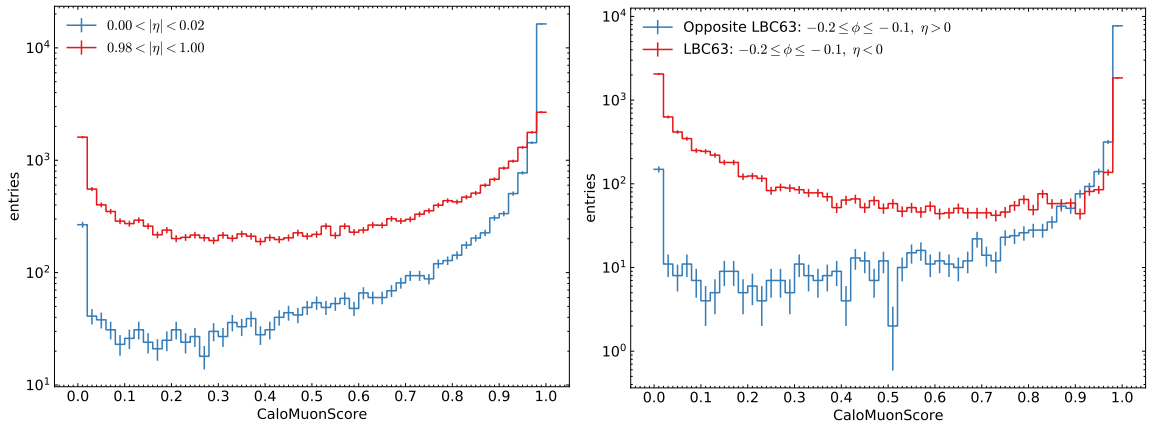


Figure 6.39: Mean value of `CaloMuonScore` in bins of ϕ and η. The size of each bin in $\phi \times \eta$ is $\pi/32 \times 1/32$. The plot is based on 7 million tag-and-probe muons from simulated $Z \to \mu\mu$ decays and uses MS probes.

6.8.1.3 Scale factors

The scale factors for the four `CaloMuonScore` working points, and `CaloMuonTag` for comparison, are derived separately in η and p_{T}. Figure 6.41 shows the muon efficiency as a function of η in the region $|\eta| < 1$ for the four `CaloMuonScore` working points as well as `CaloMuonTag`. The scale factors are around 0.3% or less in most bins, indicating a small and expected level of mismodelling. The level of mismodelling in the case of `CaloMuonScore` is of the same magnitude as observed for `CaloMuonTag`. The scale factors of the `CaloMuonScore` workings points tend to differ from unity in the left- and rightmost bins in η; a higher degree of mismodelling is expected



(a) `CaloMuonScore` in two slices of $|\eta|$.

(b) `CaloMuonScore` in the region of LBC63, and the on the opposite η side.

Figure 6.40: Distribution of `CaloMuonScore` in different $\eta - \phi$ regions for true muons from $Z \to \mu\mu$ decays. Plot (a) explores the regions of low efficiency near $\eta = \pm 1$, while plot (b) illustrates the drop in the region corresponding to LBC63. The error bars shown are statistical only.

here due to gaps between the barrel and extended-barrel regions of the hadronic calorimeter which house ID cabling and electronics [185].

The scale factors are additionally derived as a function of p_{T}; this is shown in Figure 6.42. The scale factors here, again, are generally of order 0.3% or less.

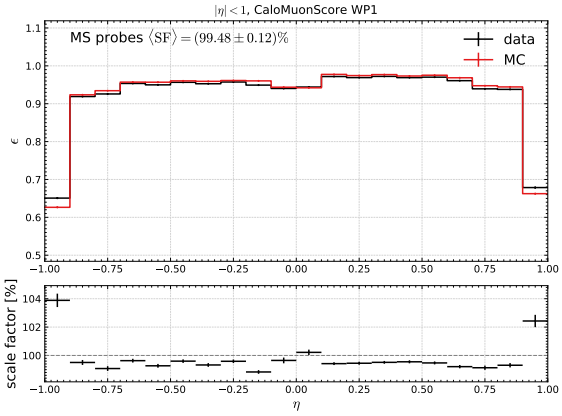
6.8.2 Tag-and-probe results in $|\eta| < 0.1$

In this section, the tag-and-probe results are presented in the region $|\eta| < 0.1$.

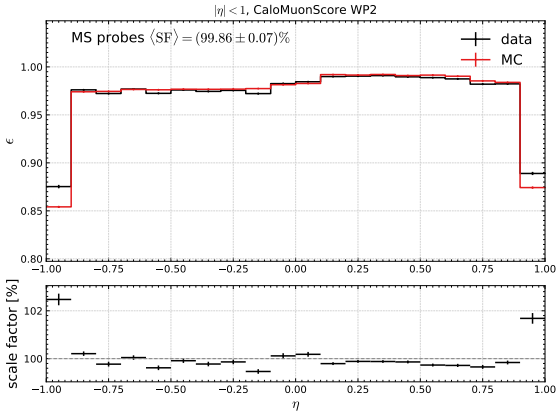
6.8.2.1 Overall performance

Figures 6.43 and 6.44 show the signal efficiency of the four `CaloMuonScore` working points compared with `CaloMuonTag` as a function of η and p_{T}, respectively. As before, the efficiencies are measured on 3.01 fb^{-1} of data.

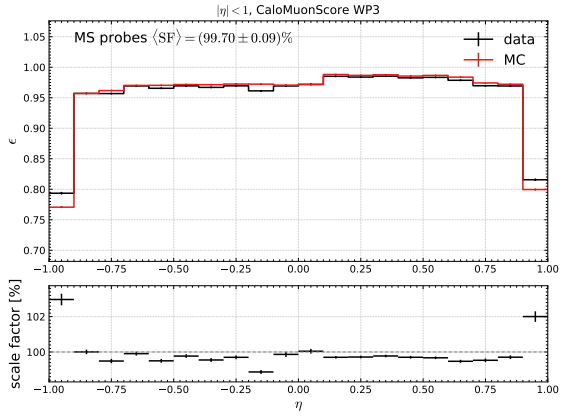
The efficiencies as a function of η (c.f. Figure 6.43) highlight that `WP1` does not offer a significant improvement over `CaloMuonTag` up to $|\eta| = \pm 0.07$. Beyond that, `WP1` achieves an improvement of around 5%. Working points 2, 3, and 4 offer a less η-dependent performance improvement; all three of them show a minimum efficiency improvement of 5% in the central-η region, and an increase to 10% beyond that. The performance gain as a function of p_{T} (c.f. Figure 6.44) is largest up to $p_{\mathrm{T}} = 30$ GeV for all four working points and reaches a roughly constant 5–10% beyond that.



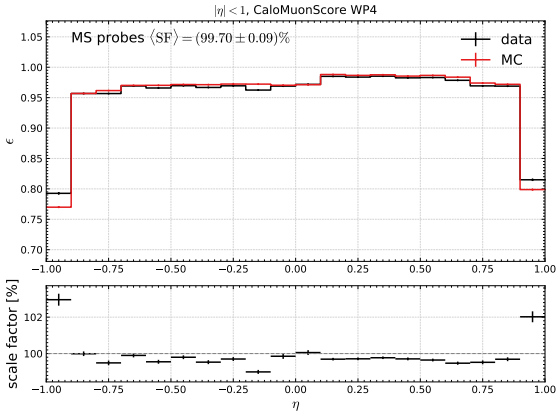
(a) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP1`.



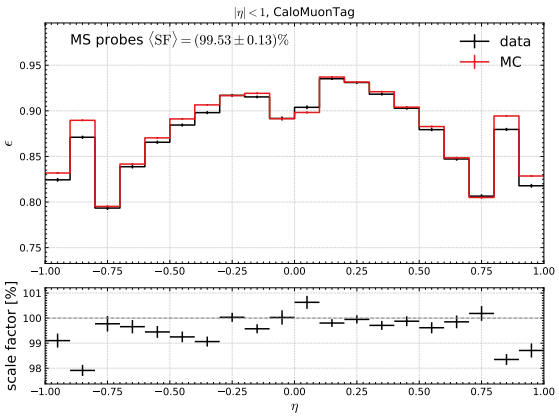
(b) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP2`.



(c) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP3`.

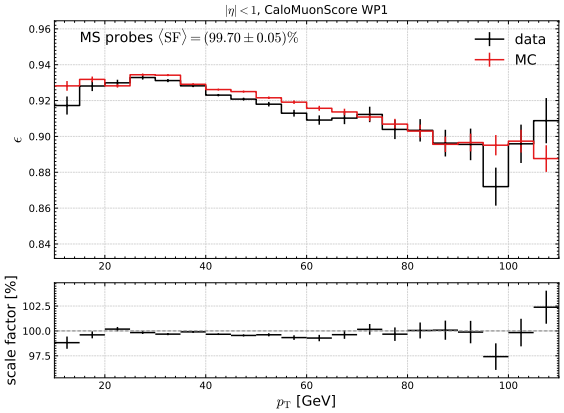


(d) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP4`.

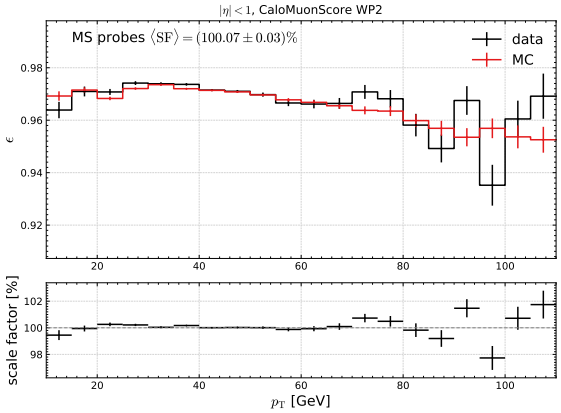


(e) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag`.

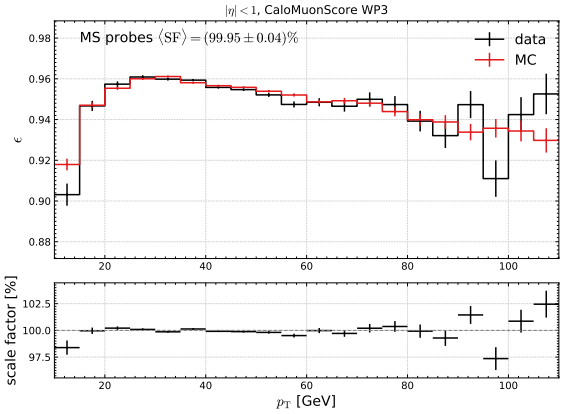
Figure 6.41: Tag-and-probe signal efficiencies vs. η for `CaloMuonScore` at four working points (a–d), as well as `CaloMuonTag` (e). The measurements are based on MS probes. The efficiencies measured on data are shown in black, those measured on simulation are shown in red. The panels show the ratio of the efficiencies measured on data to those measured on simulation (i.e. the scale factors). The error bars shown are statistical only. ⟨SF⟩ denotes the mean scale factor weighted by bin content and its standard error.



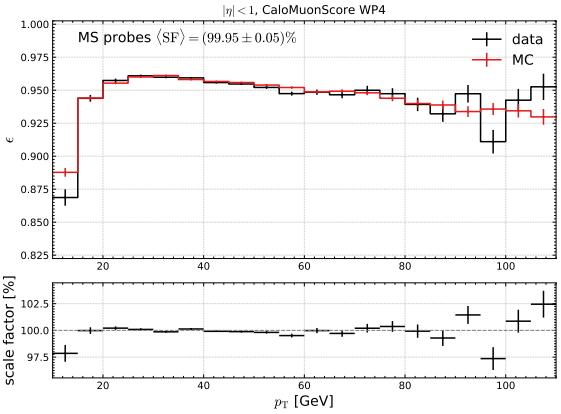
(a) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP1`.



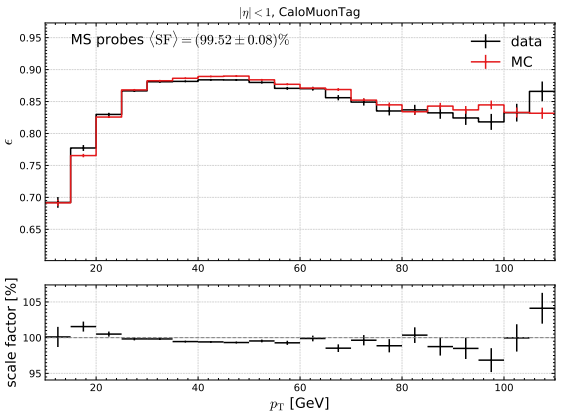
(b) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP2`.



(c) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP3`.

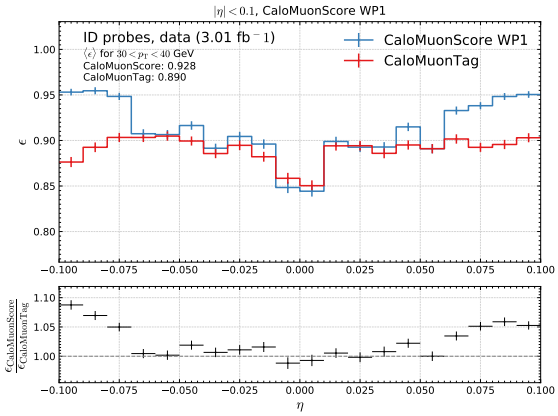


(d) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP4`.

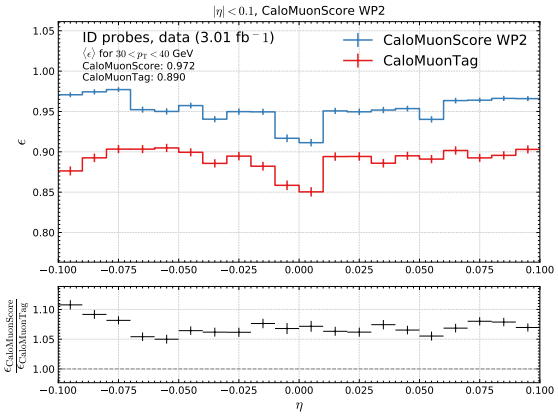


(e) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonTag`.

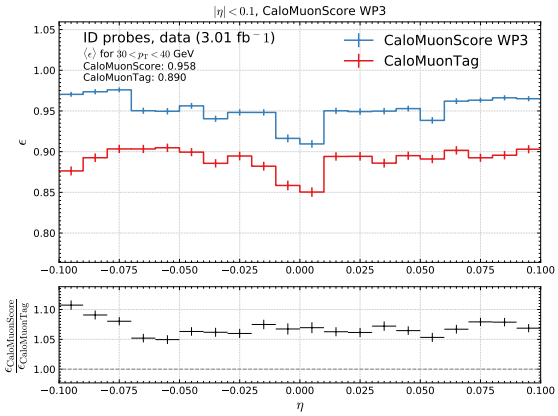
Figure 6.42: Tag-and-probe signal efficiencies vs. p_T in $|\eta| < 1$ for `CaloMuonScore` at four working points (a–d), as well as `CaloMuonTag` (e). The measurements are based on MS probes. The efficiencies measured on data are shown in black, those measured on simulation are shown in red. The panels show the ratio of the efficiencies measured on data to those measured on simulation (i.e. the scale factors). The error bars shown are statistical only.



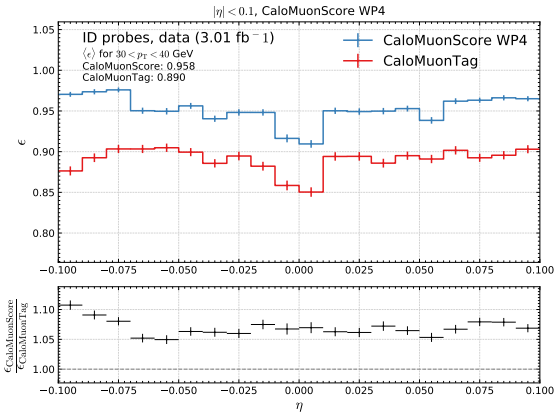
(a) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP1`.



(b) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.

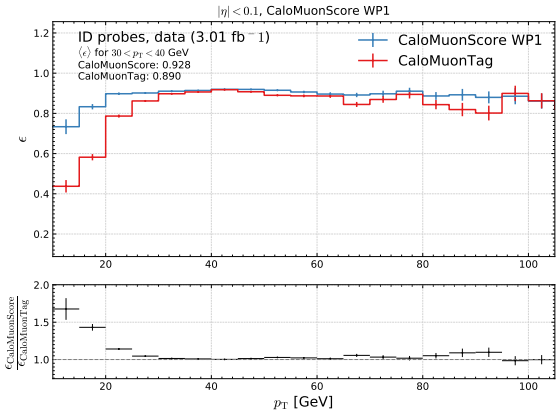


(c) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.

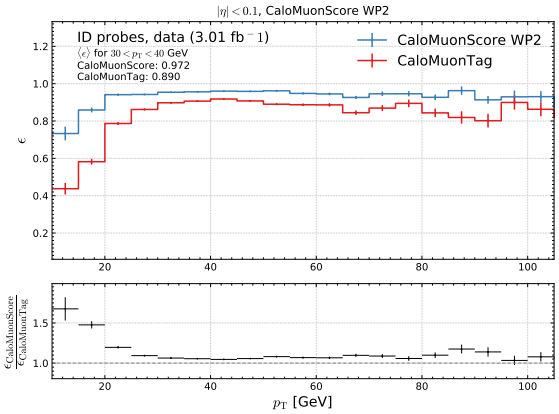


(d) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

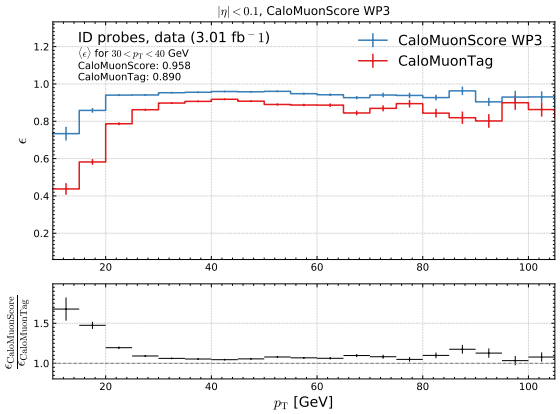
Figure 6.43: Tag-and-probe signal efficiencies vs. η for `CaloMuonTag` vs. `CaloMuonScore` at four working points using ID probes. The efficiencies are measured on 3.01 fb^{-1} of data. The efficiencies achieved by the different `CaloMuonScore` working points are shown in blue, while those achieved by `CaloMuonTag` are shown in red. The bottom panels show the ratio of the `CaloMuonScore` efficiencies to the `CaloMuonTag` efficiencies. The error bars shown are statistical only.



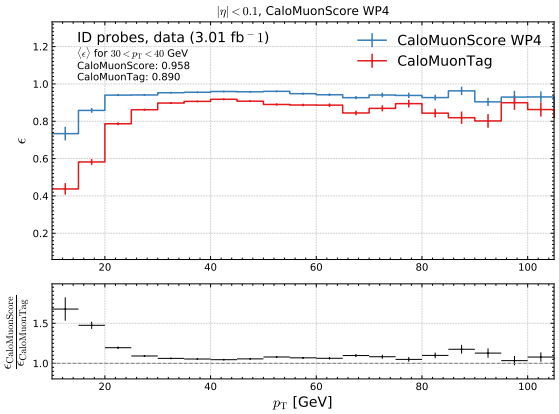
(a) Tag-and-probe ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP1`.



(b) Tag-and-probe ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP2`.



(c) Tag-and-probe ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP3`.



(d) Tag-and-probe ϵ_{sig} vs. p_{T} for `CaloMuonTag` and `CaloMuonScore` operating at `WP4`.

Figure 6.44: Tag-and-probe signal efficiencies vs. p_{T} in $|\eta| < 0.1$ for `CaloMuonTag` vs. `CaloMuonScore` at four working points using ID probes. The efficiencies are measured on 3.01 fb^{-1} of data. The efficiencies achieved by the different `CaloMuonScore` working points are shown in blue, while those achieved by `CaloMuonTag` are shown in red. The panels show the ratio of the `CaloMuonScore` efficiencies to the `CaloMuonTag` efficiencies. The error bars shown are statistical only.

6.8.2.2 Scale factors

The scale factors in the region $|\eta| < 0.1$ are shown separately as a function of η in Figure 6.45 and p_{T} in Figure 6.46. The scale factors do not show a significant η or p_{T} dependence. Moreover, they do not differ significantly from unity anywhere, indicating that the degree of calorimeter mismodelling in the low-η region is negligible for the purposes of `CaloMuonScore`.

6.9 Performance summary

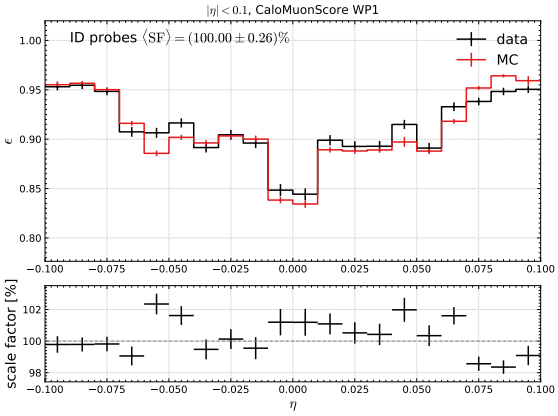
The preceding sections have presented the performance of `CaloMuonScore` in two different detector regions, and, moreover, have relied both on MC-only measurements and tag-and-probe results based on data. It is at this point instructive to summarise the overall performance of `CaloMuonScore` in different kinematic and detector regions. This summary is presented in Table 6.13. Here, the ϵ_{sig} values are based on tag-and-probe measurements, while the ϵ_{bkg} values are based on simulation only.

The table highlights that in terms of signal efficiency, `CaloMuonScore WP2` offers the best performance across all p_{T} and η regions. Its improvement over `CaloMuonTag` varies between 7 and 9 percentage points. All `CaloMuonScore` working points offer higher signal efficiencies than `CaloMuonTag`, and the selection of a working point for physics analyses should likely focus on the degree of background rejection required in a given analysis.

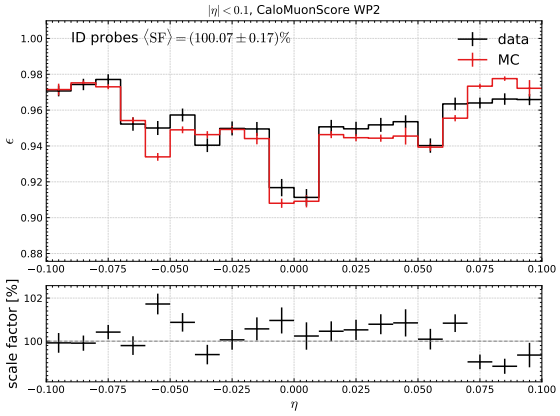
`CaloMuonScore WP2` delivers the best performance in terms of background efficiency. Its improvement over `CaloMuonTag` reaches a factor of 19.9 in the low-η region and for $p_{\mathrm{T}} \in [30, 40]$ GeV.

`CaloMuonScore WP3` offers the best overall performance, with improved signal and background efficiencies compared to `CaloMuonTag` in both detector regions. Its signal efficiency is 7–9 percentage points higher than `CaloMuonTag`, and it suppresses background 1–12 times more effectively than `CaloMuonTag`.

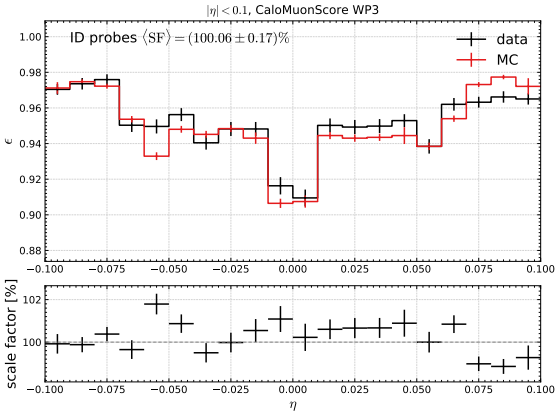
In the region where `CaloMuonScore` and `CaloMuonScore` should be compared most directly, i.e. $|\eta| < 0.1$ and $p_{\mathrm{T}} \in [30, 40]$ GeV, `CaloMuonScore WP3` and `WP4` deliver outstanding performance. Here, their signal efficiency is 7 percentage points higher than `CaloMuonTag`'s, and they suppress background more effectively by a factor of 12.2 (c.f. Table 6.11).



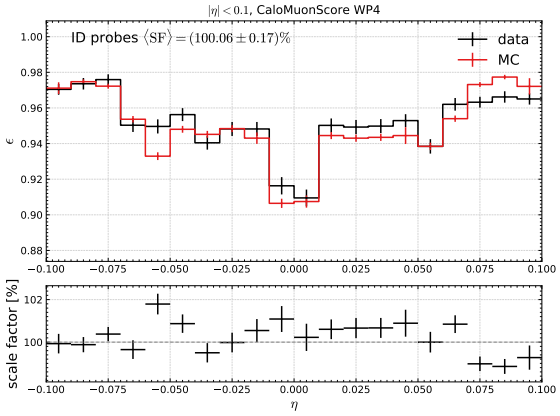
(a) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP1`.



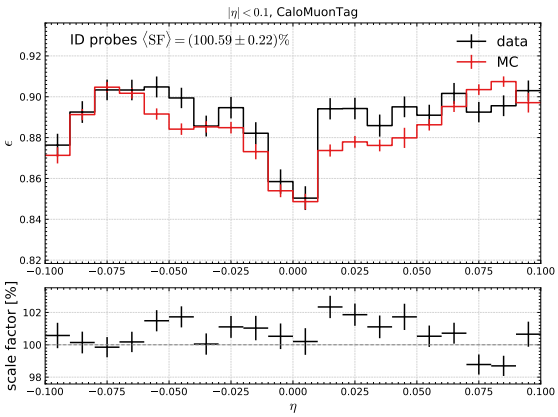
(b) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP2`.



(c) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP3`.

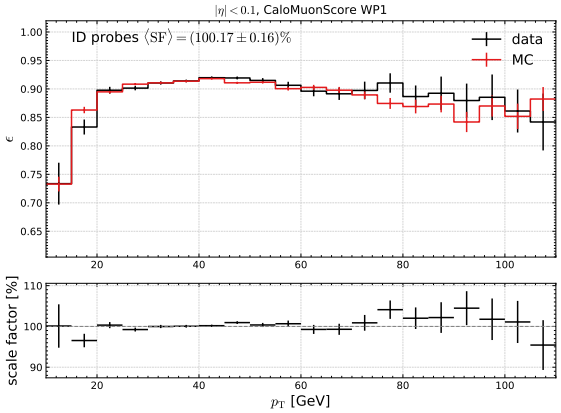


(d) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonScore` operating at `WP4`.

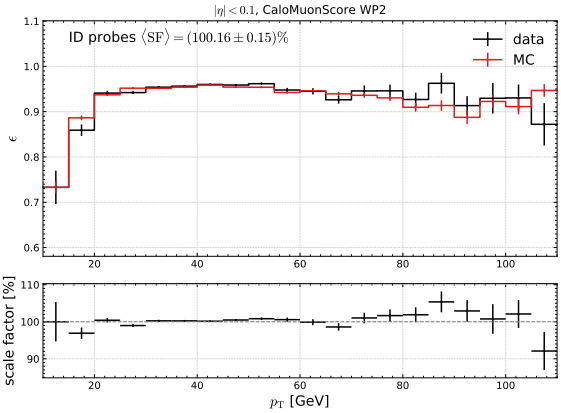


(e) Tag-and-probe ϵ_{sig} vs. η for `CaloMuonTag`.

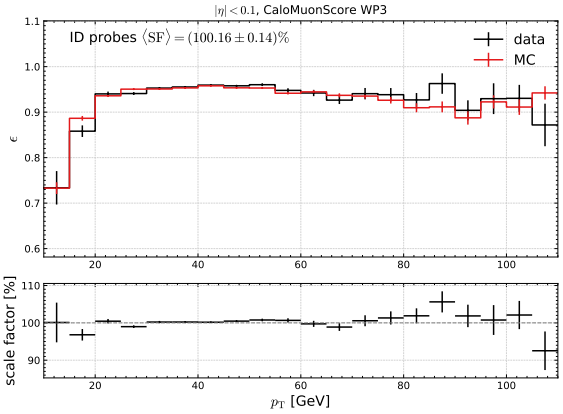
Figure 6.45: Tag-and-probe signal efficiencies vs. η for `CaloMuonScore` at four working points (a–d), as well as `CaloMuonTag` (e). The measurements are based on ID probes. The efficiencies measured on data are shown in black, those measured on simulation are shown in red. The panels show the ratio of the efficiencies measured on data to those measured on simulation (i.e. the scale factors). The error bars shown are statistical only.



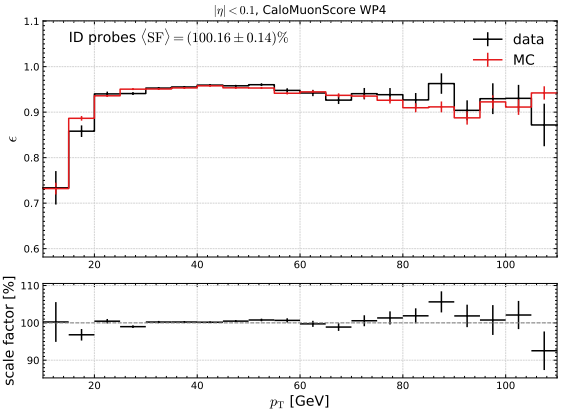
(a) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP1`.



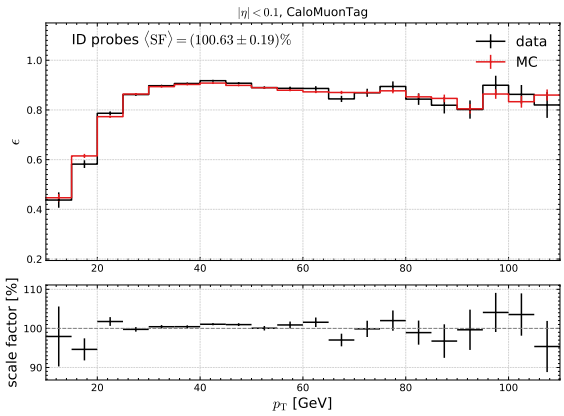
(b) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP2`.



(c) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP3`.



(d) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonScore` operating at `WP4`.



(e) Tag-and-probe ϵ_{sig} vs. p_T for `CaloMuonTag`.

Figure 6.46: Tag-and-probe signal efficiencies vs. p_T in $|\eta| < 0.1$ for `CaloMuonScore` at four working points (a–d), as well as `CaloMuonTag` (e). The measurements are based on ID probes. The efficiencies measured on data are shown in black, those measured on simulation are shown in red. The panels show the ratio of the efficiencies measured on data to those measured on simulation (i.e. the scale factors). The error bars shown are statistical only.

	$\|\eta\| < 0.1$				$\|\eta\| < 1$			
	ID probes				MS probes			
	$p_{\mathrm{T}}^{\mathrm{incl}}$		$p_{\mathrm{T}}^{[30,40]}$		$p_{\mathrm{T}}^{\mathrm{incl}}$		$p_{\mathrm{T}}^{[30,40]}$	
	$\epsilon_{\mathrm{sig}}^{\mathrm{TP}}$	$\epsilon_{\mathrm{bkg}}^{\mathrm{MC}}$	$\epsilon_{\mathrm{sig}}^{\mathrm{TP}}$	$\epsilon_{\mathrm{bkg}}^{\mathrm{MC}}$	$\epsilon_{\mathrm{sig}}^{\mathrm{TP}}$	$\epsilon_{\mathrm{bkg}}^{\mathrm{MC}}$	$\epsilon_{\mathrm{sig}}^{\mathrm{TP}}$	$\epsilon_{\mathrm{bkg}}^{\mathrm{MC}}$
`CaloMuonTag`	85.86	10.70	88.06	48.18	88.05	37.00	88.97	57.56
`CaloMuonScore WP1`	90.03	14.91	90.99	**2.42**	92.21	29.83	92.82	**8.89**
`CaloMuonScore WP2`	**94.15**	43.66	**95.23**	6.36	**97.02**	50.32	**97.25**	13.78
`CaloMuonScore WP3`	94.07	10.19	95.14	3.94	95.41	25.06	95.82	10.67
`CaloMuonScore WP4`	94.07	**6.69**	95.14	3.94	95.38	**20.16**	95.82	10.67

Table 6.13: Performance summary of `CaloMuonTag` and all four `CaloMuonScore` working points. $p_{\mathrm{T}}^{\mathrm{incl}}$ denotes all muons and fakes independent of p_{T}, while $p_{\mathrm{T}}^{[30,40]}$ denotes a selection of particles with a transverse momentum between 30 and 40 GeV. The columns labelled '$\epsilon_{\mathrm{sig}}^{\mathrm{TP}}$' indicate efficiency measurements that are based on a tag-and-probe analysis, while those labelled '$\epsilon_{\mathrm{bkg}}^{\mathrm{MC}}$' are based on simulation. For each column, the best-performing algorithm is highlighted in bold. All efficiencies are given as per-cent values.

6.10 Conclusions and future work

In this chapter, a novel method for identifying muons in the ATLAS calorimeters was presented. This approach aims to improve the quality of calorimeter-tagged muons, that is, muons that have a track in the inner detector but no matching track in the muon spectrometer. This type of calorimeter-tagged muon is important in the low-η barrel region of the ATLAS detector, where the muon acceptance is generally lower as here the muon spectrometer is only partially instrumented.

The method for performing calorimeter-tagging used by ATLAS since 2009 is called `CaloMuonTag`, and it discriminates muons from other particles by an energy-deposit signature in the calorimeters that is consistent with a minimum-ionising particle. In this work, it has been demonstrated that performing muon identification using calorimeter energy deposits at the individual cell level is a viable approach that outperforms the `CaloMuonTag` method. While the new method was developed on simulation, its performance was measured both on simulation and data with a high level of agreement between the two.

Several lines of work could be undertaken to extend, and improve upon, what has been outlined here:

1. Different convolutional neural network architectures. The approach discussed

in this work is based on multi-channel two-dimensional convolutional neural networks. Another class of neural-network layers worth investigating are 3D convolutional layers, in which the convolutional kernel operates in three instead of two dimensions. While computationally much more costly, especially at training time, it could be a suitable approach that can potentially better exploit the spatial evolution of a particle shower along its depth.

2. Introducing a variable convolutional-layer granularity that depends on the calorimeter layer. The calorimeter samplings (indices 1 through 7) used have different intrinsic granularities in η and ϕ. Future investigations into different convolutional architectures should consider combining several layers of single-channel convolutional layers on images of size $m_i \times n_i$, where $i \in \mathbb{Z} : i \in [1, 7]$ is an index associated with each calorimeter sampling. One hypothesis is that such an architecture would be better able to exploit the high granularity of the inner calorimeter samplings, while using images of very coarse granularity in outer calorimeter samplings that are intrinsically less fine-grained.

3. Extending the calorimeter-tagging region for muons beyond $|\eta| = 1$. The methods developed here can likely be applied beyond $|\eta| = 1$ in order to partially recover muons in other regions of low MS acceptance (c.f. Section 3.2.4). This is especially relevant for the transition region between the barrel and end-cap sections of the MS near $|\eta| \approx 1$, as well as the cracks in the end-caps at $|\eta| \approx 1.2, 1.7, 2.1$, and 2.3. The regions near $|\eta| \approx 1.2$ and 1.7 would likely benefit less from this approach due to reduced calorimeter acceptance in the transition region between the ECAL and HCAL barrel and end-cap regions. In order to investigate an adaptation of `CaloMuonScore` extending to higher pseudorapidities, the classifier would have to be partially redesigned. Specifically, it would need to take into account a different selection of calorimeter samplings, since only central-barrel calorimeter samplings are considered in the configuration developed in this chapter.

4. Suppression of electronic noise in the calorimeters. All calorimeter samplings exhibit electronic noise, which manifests itself in noisy, and in some cases negative, energy-deposit readouts. Depending on the particular calorimeter sampling, the noise is on the order of 100–1000 GeV100 to 1000 MeV per readout channel [75]. Future studies should investigate whether 'cleaning' the calorimeter images before training the classifiers can push the performance

even further. As calorimeter images from both muons and pions in this study exhibit the same amount of electronic noise, it is assumed that the classifiers learn to ignore the noise and implicitly subtract background noise from calorimeter images. It is therefore unlikely that such an approach would provide a significant performance improvement.

5. Training dedicated classifiers in different p_{T} and η regions. Training of dedicated convolutional neural networks in different η and p_{T} regions was investigated, and it was found that the performance was worse compared to training on the full η and p_{T} range. However, it is likely that with access to much more training data (at least an order of magnitude more), training separately on different phase-space regions could provide marginal improvements. The two $Z \to \mu\mu$ and $t\bar{t}$ samples mentioned were the only suitable samples available in the required RDO data format when `CaloMuonScore` was under development. The first step towards determining if dedicated classifiers for regions of p_{T} and η would be valuable is to repeat the study on the effect of the available training statistics in Section 6.7.6, and to determine a trend in the classifier performance as a function of the size of the training set. If the performance curve flattens out, an increase in training set size is unlikely to further increase the performance. If, on the other hand, the trend looks as though the performance would keep increasing, future work would benefit from a greater number of MC samples.

The `CaloMuonScore` method outlined in this chapter has been implemented into the central ATLAS code base. It offers for the first time a complementary approach to calorimeter-tagging for all physics analyses that include muons in addition to `CaloMuonTag`, which has been in use since 2009. Based on a study performed on MC events containing truth-matched muons from $Z \to \mu\mu$ events, as well as truth-matched pions and kaons from $t\bar{t}$ events, several working points at which `CaloMuonScore` can be operated have been derived, each of them targeting a different muon efficiency or fake rate.

Depending on the working point at which one chooses to operate `CaloMuonScore`, an improvement in terms of muon-identification efficiency of 7–9 percentage points above `CaloMuonTag` was measured. Further, the suppression of fake muons is improved by up to a factor of 20. `CaloMuonScore` efficiently identifies muons over a wide range in p_{T}, and further exhibits high muon efficiency beyond the narrow range of $|\eta| < 0.1$ up to $|\eta| < 1$. These results, therefore, suggest that `CaloMuonScore`

can be used to recover muons with (i) higher efficiency, (ii) at a greatly reduced fake rate, and (iii) in wider kinematic regions than `CaloMuonTag`.

Chapter 7

Conclusion

Several extensions to the Standard Model predict the existence of additional heavy Higgs-like resonances that may be observable at hadron colliders at the TeV scale. Operating at the highest luminosities ever achieved in a hadron collider, the general-purpose particle detectors ATLAS and CMS at the CERN LHC can test the predictions of the Standard Model with high precision and look for phenomena beyond the Standard Model. Essential to both precision tests and searches for new physics is the reliable identification of muons. Comprising both a search for BSM physics and a step towards improved muon identification, the results of this thesis are briefly summarised below.

A search for heavy resonances decaying via two Z bosons to four leptons using the ATLAS detector was presented in Chapter 5. The search results using 139 fb^{-1} of data collected during 2015–2018 improve significantly on those previously published. This is achieved by a multivariate approach based on deep learning in both the ggF and VBF analysis channels. This approach is shown to provide significantly lower upper limits on the production cross-section of a heavy scalar compared to a cut-based approach. No significant excess in the $m_{4\ell}$ spectrum is found, and stronger limits are placed on the production cross-section of a heavy Higgs-like scalar up to 2 TeV. Using a combination of results with the closely related $\ell^+\ell^-\nu\bar{\nu}$ analysis, the observed limits on the production cross-section times branching fraction are between 215 fb at 240 GeV and 2.0 fb at 1900 GeV in the ggF production mode, and between 87 fb at 255 GeV and 1.5 fb at 1800 GeV in the VBF production mode. The results are interpreted in the context of a Type-I and a Type-II 2HDM scenario as well as in the context of a search for a spin-2 graviton in the RS1 model. In the latter case, gravitons are excluded up to a mass of 1830 GeV at 95% confidence level. Future iterations of this analysis using more data acquired by the ATLAS experiment—an

additional 300 fb^{-1} are to be expected by the end of Run 3 in 2024, and a total of 3000 fb^{-1} by the end of Run 6 in 2037—as well as novel analysis techniques will probe the current $\ell^+\ell^-\ell'^+\ell'^-$ search range with greater sensitivity and at the same time push the reach in $m_{4\ell}$ to higher values of 3 TeV and beyond.

A novel calorimeter-based muon-identification method called `CaloMuonScore` was introduced in Chapter 6. Validated on data and simulation, the method was implemented in the central ATLAS code base and commissioned for use in future physics analyses. Based on convolutional neural networks trained on energy-deposit patterns in the ATLAS electromagnetic and hadronic calorimeters, so-called calorimeter images, the method can effectively discriminate muons from fakes. Four working points offering different signal and background efficiencies were derived; these were benchmarked against the previous algorithm called `CaloMuonTag`, and it was demonstrated that `CaloMuonScore`'s ability to identify muons and its capacity to reject fake muons coming from pions and kaons are both significantly greater than for `CaloMuonTag`. For particles with a transverse momentum between 30 and 40 GeV occurring in the ATLAS central barrel region ($|\eta| < 0.1$), `CaloMuonScore WP3` has an average muon efficiency of 95.2% and a fake rate of 3.9%, compared with a muon efficiency of 88.1% and a fake rate of 48.2% for `CaloMuonTag`. The results indicate that `CaloMuonScore` is a suitable alternative to `CaloMuonTag` that provides better performance, and is fit for use in physics analyses that rely on calorimeter-tagged muons. Accordingly, `CaloMuonScore` is now part of the central ATLAS code release.

Appendices

Appendix A

Particle interactions with matter

The ATLAS calorimeters measure particle energies in two main ways: (i) by measuring the ionisation and radiative energy losses of charged particles, and (ii) by absorbing hadrons through interactions with the detector material. The electromagnetic calorimeter is designed to cover the former case, while the hadronic calorimeter is responsible for the latter.

Charged particles traversing the detector material, such as muons or charged pions, lose energy through three main processes: ionisation, bremsstrahlung, and e^+e^- pair production. For energies typically encountered in the ATLAS detector, ionisation is by far the most significant form of energy loss, however, radiative losses due to bremsstrahlung and pair production become important at higher energies. The *muon critical energy*, E_c^μ, is defined as the muon energy at which energy losses from radiation become as important as energy losses from ionisation. In the main material of the ATLAS electromagnetic calorimeter, argon, it has a value of roughly 500 GeV [26]. Energy losses from e^+e^- pair production become significant at even higher energies. The muon energies encountered in this thesis are typically of order 100 GeV or less; as a result, the following section will focus solely on energy losses due to ionisation.

A.1 Energy losses due to ionisation

The mean rate of energy loss, sometimes referred to as the *mass stopping power*, due to ionisation experienced by relativistic charged particles per unit length of material

traversed is given by the Bethe formula [196],

$$\langle \frac{dE}{dx} \rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{A.1}$$

where $K = 0.31\,\text{MeV}\,\text{mol}^{-1}\,\text{cm}^2$, z is the charge number of the incident particle, Z is the atomic number of the absorber material, A is the atomic mass of the absorber material, β is the incident particle velocity in units of the speed of light, c, m_e is the electron mass, and $\gamma = 1/\sqrt{1-\beta^2}$ is the Lorentz factor. The mass stopping power[1] of $\text{MeV}\,\text{g}^{-1}\,\text{cm}^2$ is expressed in units.

A density-effect correction to the ionisation energy loss is given by $\delta(\beta\gamma)/2$. It accounts for the fact that at high particle energies, the traversed material becomes polarised, which limits electric-field interactions of the charged particle with the medium. This leads to a reduction in energy loss, becoming more significant at higher energies. At very high energies, it is given by

$$\frac{\delta(\beta\gamma)}{2} = \ln(\hbar\omega_p / I) + \ln \beta\gamma - \frac{1}{2}, \tag{A.2}$$

where $\hbar\omega_p$ is the plasma energy. It is equal to $\sqrt{\rho\langle Z/A\rangle}$, where ρ is the density of the absorber material in units of $\text{g}\,\text{cm}^{-3}$. I is the mean excitation energy, which is determined experimentally, and varies as a function of the atomic number of the absorber material. For the ATLAS electromagnetic calorimeter—which chiefly contains argon with $Z = 18$—it has a value of around 0.55 eV. $W_{\max}$ is the maximum possible energy transfer to an electron in a single collision. For an incident particle with mass M, it is given by

$$W_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e / M + (m_e/M)^2}. \tag{A.3}$$

In the limit where $2\gamma m_e \gg M$, one has

$$W_{\max} = Mc^2 \beta^2 \gamma. \tag{A.4}$$

The mass stopping power for positive muons in copper is illustrated in Figure A.1. Equation (A.1) describes the central region of the plot, labelled 'Bethe', for a muon momentum between 10 and 90 GeV. Beyond a muon momentum of 90 GeV, radiative

[1]The mass stopping power is obtained by dividing the stopping power by the density of the material. The stopping power has dimensions of EL^{-1}, where E is energy and L is length. Dividing by the dimensions of material density, i.e. ML^{-3}, where M is mass, shows that the mass stopping power has dimensions of EL^2M^{-1}.

losses become more significant. While the plot applies to muons in copper ($Z = 29$), the behaviour in the ATLAS calorimeter looks very similar, as $\langle \frac{dE}{dx} \rangle$ decreases only slowly with Z. The minimum of the plot in the 'Bethe' region corresponds to *minimum ionisation*, which occurs when the most loosely bound electrons are ionised. The mass stopping power at minimum ionisation for muons is around 1.65 MeV cm^2 g^{-1} [26], and occurs at a muon momentum of around 0.1–10 GeV. Liquid argon, the main material in the ATLAS EM calorimeter, has a density of around 1.40 g cm^{-3} [197], yielding a minimum ionising stopping power for muons in liquid argon of 2.31 MeV cm^{-1}. In the ATLAS hadronic calorimeter, which is made up chiefly of steel as the absorbing material with a density of around 7.86 g cm^{-3} [198], the energy loss of a minimum ionising particle is around 12.97 MeV cm^{-1}. The mass stopping power, and resulting minimum-ionising stopping power, for pions is almost the same, as Equation (A.1), taking into account Equation (A.4), depends only logarithmically on the incident particle's mass.

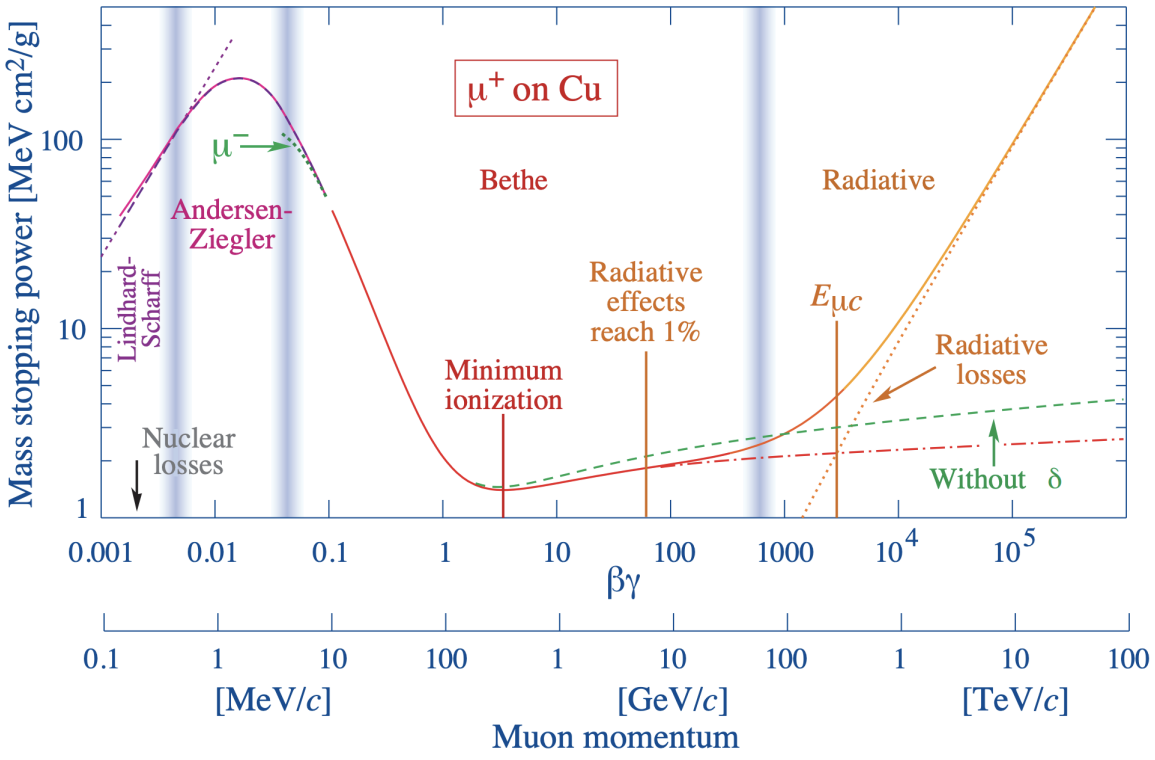


Figure A.1: Mass stopping power, $\langle dE/dx \rangle$, for antimuons in copper as a function of muon momentum. Solid curves indicate the total stopping power. Figure reproduced from Reference [26].

A.2 Hadronic interactions

The main difference in calorimeter signature between muons and pions comes from the propensity of pions to create hadronic showers. Hadronic showers are produced by hadrons interacting with the detector material. The result is that leptons such as muons do not produce any activity in the hadronic calorimeter from hadronic interactions, however, they may leave smaller electromagnetic signatures.

A hadronic shower typically has two components: (i) an electromagnetic component coming from e^+e^- pair production or neutral hadrons decaying into two photons which launch an electromagnetic cascade, and (ii) a hadronic component corresponding to the energy deposited through nuclear collisions, intranuclear cascades (referred to as *spallation*), and subsequent secondary interactions. A shower in the hadronic calorimeter typically contains a 15% electromagnetic component. Electromagnetic showers tend to be much more contained than hadronic ones: for iron—the main element of the absorbing material in the ATLAS hadronic calorimeter—the length scale governing electromagnetic showers is $X_0 \approx 1.76\,\text{cm}$, whereas the hadronic-shower length scale is $\lambda_n \approx 16.8\,\text{cm}$. The ATLAS hadronic calorimeter has a depth of around 11 interaction lengths in the barrel region; the activity in a hadronic shower decreases exponentially with depth. A 300 GeV pion, for example, loses 95% of its energy within $8\,\lambda_n$, and 99% within $11\,\lambda_n$.

Appendix B

Search for Heavy ZZ Resonances in the $\ell^+\ell^-\ell'^+\ell'^-$ Final State

This appendix provides additional details on the signal and interference modelling. It has largely been reproduced from Reference [1].

B.1 Signal modelling in the large-width assumption

The shape of the $m_{4\ell}$ distribution under the large-width assumption (LWA) is described as a convolution of a truth distribution with a detector-resolution function. The truth-level distribution is given below, whereas the detector-resolution function is modelled with the same function used in the NWA case (Equation (5.2)); its truth width is negligible and the $m_{4\ell}$ distribution is therefore dominated by detector effects.

The cross-section of the BSM Higgs signal in the LWA decaying to ZZ is given by [1]

$$\sigma_{pp\to H\to ZZ}(m_{4\ell}) = 2 \cdot m_{4\ell} \cdot \mathcal{L}_{gg} \cdot \frac{1}{|s - s_H|^2} \cdot \Gamma_{gg\to H}(s) \cdot \Gamma_{H\to ZZ}(s), \tag{B.1}$$

where $\mathcal{L}_{gg}$ is the gluon-gluon luminosity [199] and $\frac{1}{s-s_H}$ is the Higgs propagator given by [200, 201]

$$\frac{1}{s - s_H} = \frac{1 + i \cdot \overline{\Gamma}_H/\overline{m}_H}{s - \overline{m}_H^2 + i \cdot s \cdot \overline{\Gamma}_H/\overline{m}_H}, \tag{B.2}$$

with

$$\overline{m}_H = \sqrt{\Gamma_H^2 + m_H^2} \tag{B.3}$$

and

$$\overline{\Gamma}_H = \overline{m}_H \cdot \frac{\Gamma_H}{m_H}, \tag{B.4}$$

where s is the square of the centre-of-mass energy, m_H is the BSM Higgs mass and Γ_H is the total BSM Higgs width in the LWA.

$\Gamma_{gg\to H}(s)$ is the $gg \to H$ partial width, given by

$$\Gamma_{gg\to H}(s) = C_1 \cdot s^{\frac{3}{2}} \cdot |A_t(\tau)|^2 \tag{B.5}$$

with

$$\begin{aligned} A_t(\tau) &= 2\frac{\tau + (\tau-1)f(\tau)}{\tau^2} \\ \tau &= \frac{s}{4m_t^2} \\ f(\tau) &= \begin{cases} \arcsin^2(\sqrt{\tau}), & \tau \leqslant 1 \\ -\frac{1}{4}\left[\log\frac{1+\sqrt{1-\tau^{-1}}}{1-\sqrt{1-\tau^{-1}}} - i\pi\right]^2, & \tau > 1, \end{cases} \end{aligned} \tag{B.6}$$

where C_1 is a constant and m_τ is the mass of the top quark.

$\Gamma_{H\to ZZ}(s)$ is the $H \to ZZ$ partial width, given by

$$\Gamma_{H\to ZZ}(s) = C_2 \cdot s^{\frac{3}{2}} \cdot \left[1 - \frac{4m_Z^2}{s} + \frac{3}{4}\left(\frac{4m_Z^2}{s}\right)^2\right] \cdot \left[1 - \frac{4m_Z^2}{s}\right]^{\frac{1}{2}}, \tag{B.7}$$

where C_2 is a constant and m_Z is the mass of the Z boson.

B.2 Graviton signal modelling

The shape of the graviton $m_{4\ell}$ distribution is described by the convolution of a truth-level distribution with a detector-resolution function. The truth-level distribution is given by the product of a relativistic Breit-Wigner factor, a parton-luminosity factor, $\mathcal{L}_{gg}$, and a factor that corresponds to the matrix element of the production process as given in Reference [202].

The graviton cross-section of a ggF-produced graviton decaying to ZZ is given by

$$\sigma_{pp\to G^*\to ZZ}(m_{4\ell}) = C_3 \cdot \mathcal{L}_{gg} \cdot s^2 \cdot \frac{s(1+s)(1+2s+2s^2)}{(s^2 - m_{G^*}^2\Gamma^2)}, \tag{B.8}$$

where C_3 is a constant, s is the square of the centre-of-mass energy, m_G^* is the mass of the graviton and Γ is the total graviton decay width.

As in the case of the LWA, the detector-resolution function is taken to be the same as in the NWA (Equation (5.2)).

B.3 Interference modelling

Three processes share the same gg initial and ZZ final states, and will therefore interfere with each other [1]:

1. the SM process $gg \to ZZ$ with amplitude A_B,
2. the SM Higgs boson with amplitude A_h, and
3. the BSM heavy Higgs with amplitude A_H.

The full parton cross-section can be written as

$$\begin{aligned}\sigma_{gg\to(X)\to ZZ}(s) &= \frac{1}{2s}\int d\Omega\,|A_h(s,\Omega)+A_H(s,\Omega)+A_B(s,\Omega)|^2 \\ &= \frac{1}{2s}\int d\Omega\left(|A_h(s,\Omega)|^2+|A_H(s,\Omega)|^2+|A_B(s,\Omega)|^2\right) \\ &+ \frac{1}{s}\int d\Omega\big(Re\left[A_h(s,\Omega)\cdot A_B^*(s,\Omega)\right] \\ &+ Re\left[A_H(s,\Omega)\cdot A_B^*(s,\Omega)\right]+Re\left[A_H(s,\Omega)\cdot A_h^*(s,\Omega)\right]\big).\end{aligned} \tag{B.9}$$

The term proportional to $|A_h|^2$ corresponds to the on-shell contribution of the SM Higgs, which is negligible in the high-mass region. The term proportional to $|A_H|^2$ corresponds to the heavy Higgs with a cross-section given by Equation (B.1). The terms proportional to $|A_B|^2$ and $A_h \cdot A_B^*$ correspond to the $gg \to ZZ$ continuum background and the interference between the SM Higgs and the $gg \to ZZ$ continuum background, respectively. Both are accounted for through MC simulation. The terms proportional to $A_H \cdot A_B^*$ and $A_H * A_h^*$ describe the interference between the heavy Higgs and the continuum background, and between the heavy Higgs and the BSM Higgs, respectively. These two terms become significant when the width of the heavy Higgs is large. The relevant interference terms are discussed below.

B.3.1 Interference of the heavy Higgs with the SM $gg \to ZZ$ background

The Higgs amplitude can be written as the product of a propagator, a production amplitude, A_H^P, and a decay amplitude, A_H^D:

$$A_H = A_H^P \cdot \frac{1}{s - s_H} \cdot A_H^D. \tag{B.10}$$

The parton cross-section of the interference can be expressed as

$$\sigma_{gg}(s) = \frac{1}{s}\mathrm{Re}\left[\frac{1}{s - s_H}\int d\Omega \cdot A_H^P(s,\Omega)\cdot A_H^D(s,\Omega)\cdot A_B^*(s,\Omega)\right]. \tag{B.11}$$

The integral in this equation is an unknown complex function of s. It is assumed that this function is smooth, and that it can be replaced with a complex polynomial of arbitrary order, namely

$$\int d\Omega\cdot A_H^P(s,\Omega)\cdot A_H^D(s,\Omega)\cdot A_B^*(s,\Omega) \approx (a_0+a_1\cdot\sqrt{s}+\dots)+i\cdot(b_0+b_1\cdot\sqrt{s}+\dots), \tag{B.12}$$

where a_i and b_i are real-valued polynomial coefficients. They can be obtained by fitting to the truth-level $m_{4\ell}$ distribution after the reconstruction-level event selections in order to take into account signal-acceptance and detector effects. As neither the Higgs mass nor the width enter into this function, the coefficients have the same values for all tested signal hypotheses.

The analytical function of the propagator is chosen to be consistent with the definition used in the `MCFM` generator [203] that is used to produce the interference samples. This propagator is shown in Equation (B.2). The parton cross-section can be transformed to a hadron cross-section as a function of $m_{4\ell}$ as described in Appendix B.1. The final equation describing the interference of the heavy Higgs and $gg \to ZZ$ is given by

$$\sigma_{pp}(m_{4\ell}) = \mathcal{L}_{gg}\cdot\frac{1}{m_{4\ell}}\cdot\mathrm{Re}\left[\frac{1}{s - s_H}\cdot((a_0 + a_1\cdot m_{4\ell}+\dots)+i\cdot(b_0+b_1\cdot m_{4\ell}+\dots))\right]. \tag{B.13}$$

B.3.2 Interference of the heavy Higgs with the SM Higgs

Similar to Equation B.10, the amplitude of the SM Higgs can be written as

$$A_h = A_h^P \cdot \frac{1}{s - s_h} \cdot A_h^D, \tag{B.14}$$

where A_h^P is the production amplitude, and A_h^D is the decay amplitude. The parton cross-section for this process can then be rewritten

$$\sigma_{gg}(s) = \frac{1}{s} \int d\Omega \cdot \mathrm{Re}\left[A_H^P(s,\Omega) \cdot \frac{1}{s - s_H} \cdot A_H^D(s,\Omega) \cdot A_h^{P*}(s,\Omega) \cdot \frac{1}{(s - s_h)^*} \cdot A_h^{D*}(s,\Omega) \right] \tag{B.15}$$

It is assumed that the heavy Higgs possesses similar properties to the one in the SM. Specifically, the production and decay amplitudes are assumed to be the same for the SM and heavy Higgs, since they depend neither on the mass nor on the width of the boson, i.e. $A_h^D = A_H^D; A_h^P = A_H^P$. The cross-section then simplifies to

$$\sigma_{gg}(s) = \frac{1}{s} \int d\Omega \cdot \mathrm{Re}\left[\frac{1}{s - s_H} \cdot \frac{1}{(s - s_h)^*} \right] \cdot \left| A_{gg\to H}(s,\Omega) \right|^2 \left| A_{H\to ZZ}(s,\Omega) \right|^2 \tag{B.16}$$

Using the definition of the partial width of a heavy Higgs decay to a state F, namely

$$\Gamma_{H\to F}(s) = \frac{1}{2\sqrt{s}} \int d\Omega \left| A_{H\to F}(s,\Omega) \right|^2, \tag{B.17}$$

the cross-section further simplifies to

$$\sigma_{gg}(s) = 4 \cdot \mathrm{Re}\left[\frac{1}{s - s_H} \cdot \frac{1}{(s - s_h)^*} \right] \cdot \Gamma_{gg\to H}(s) \cdot \Gamma_{H\to ZZ}(s), \tag{B.18}$$

where the partial widths are described in Equations (B.5) and (B.7), and the propagators are given in Equation (B.2). The parton cross-section can be transformed to a hadron cross-section as a function of $m_{4\ell}$ as described in Appendix B.1. The final equation describing the interference of a heavy Higgs and the SM Higgs is as follows:

$$\sigma_{pp}(m_{4\ell}) = 4 \cdot m_{4\ell} \cdot \mathcal{L}_{gg} \cdot \mathrm{Re}\left[\frac{1}{s - s_H} \cdot \frac{1}{(s - s_h)^*} \right] \cdot \Gamma_{gg\to H}(m_{4\ell}) \cdot \Gamma_{H\to ZZ}(m_{4\ell}). \tag{B.19}$$

This equation is very similar to the differential cross-section of the signal-only

production, with the only difference in the propagator part. However, the propagator does not affect the kinematic properties of the final state, so this interference process has the same acceptance as the one in the signal-only process. As a result, the interference process can be reproduced by reweighting the signal events with weights given by

$$w(m_{4\ell}) = \frac{2 \cdot \mathrm{Re}\left[\frac{1}{s-s_H} \cdot \frac{1}{(s-s_h)^*}\right]}{\frac{1}{|s-s_H|^2}}. \tag{B.20}$$

The reweighting procedure was validated by comparing a reweighted signal distribution for a certain mass and width to its nominal MC distribution for the same signal hypothesis, the details of which can be found in Reference [1].

A set of pseudo-MC samples for the interference process was produced using the reweighting procedure. These samples provide a precise description of the interference including the reconstruction effects, and they were further used for the validation of the interference model.

Modelling of the interference was achieved in the same way as for the large-width signal discussed in Appendix B.1. The truth-level cross-section is described by the computation of the leading-order Feynman diagram given in Equation (B.18) which is further corrected for acceptance effects.

B.3.3 Complete signal model

In the previous sections of this appendix, the pure LWA signal and its interferences with the SM processes are discussed, while this section focuses on the integration of the pure LWA signal with the interference models. Both contributions are added and will be hereafter referred to as the "complete signal model". When searching for signal or setting exclusion limits, the interference contribution is included as part of the signal. The relative importance of the interference within the complete signal model for the reconstructed $m_{4\ell}$ shape is discussed here.

B.3.3.1 Relative importance of interference terms

The relative normalisation of the interference with respect to the pure signal depends on the couplings of the heavy Higgs boson. Here, when referring to the BSM Higgs boson's couplings, we assume that the heavy Higgs has SM-like couplings that are scaled by a factor, κ. In this analysis, upper limits are set on the production cross-section times branching fraction while both the mass and the width of the signal are fixed. The signal cross-section is a function of couplings, mass, and

width only, where the latter two are fixed. Therefore, the cross-section limit is directly related to the heavy Higgs couplings, i.e. $\sigma_H \propto \kappa^2$. The normalisation of both interference contributions will have only a linear dependence on the heavy Higgs couplings, i.e. $\sigma_{\text{interf}} \propto \kappa$. This means that the relative normalisation of the interference with respect to the pure signal contribution is dependent on the analysis sensitivity. This has the consequence that the interference contribution grows with the analysis sensitivity, a fact that is taken into account when setting the cross-section limit in the LWA case.

Bibliography

[1] C. Anastopoulos, G. Artoni, G. Barone, P. Bellos, M. Cano Bret, J. W. Carter, D. Fassouliotis, A. Gabrielli, X. Ju, T. Lagouri, W. A. Leight, R. F. Naranjo Garcia, R. Nicolaidou, P. Podberezko, T. D. Powell, A. Schaffer, W. Su, S. E. Von Buddenbrock, R. Wölker, H. Yang, H. Zhu, L. Xu, R. Y. Gonzalez Andana, M. Haacke, R. I. Hoppe Elsholz, S. A. Olivares Pino, M. E. Thabede, O. Mtintsilana, and D. Bortoletto, "Searches for heavy Higgs bosons in the four-lepton decay channel with the ATLAS detector using 140 fb^{-1} of $\sqrt{s} = 13$ TeV data," Tech. Rep. ATL-COM-PHYS-2018-1697, CERN, Geneva, Dec 2018.

[2] "Search for heavy resonances decaying into a pair of Z bosons in the $\ell^+\ell^-\ell'^+\ell'^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using 139 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," Tech. Rep. ATLAS-CONF-2020-032, CERN, Geneva, Aug 2020.

[3] ATLAS Collaboration, "Search for heavy resonances decaying into a pair of Z bosons in the $\ell^+\ell^-\ell'^+\ell'^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using 139 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 81, no. 4, p. 332, 2021.

[4] R. Wölker and C. Sawyer, "Investigations into the effect of gamma irradiation on the leakage current of 130-nm readout chips for the ATLAS ITk strip detector," in *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP2018)*, vol. 343, p. 121, 2019.

[5] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B*, vol. 716, p. 30, 2012.

[6] 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*, vol. 716, p. 1, 2012.

[7] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear physics*, vol. 22, no. 4, pp. 579–588, 1961.

[8] S. Weinberg, "A model of leptons," *Physical review letters*, vol. 19, no. 21, p. 1264, 1967.

[9] Nobelstiftelsen, *Elementary Particle Theory: Relativistic Groups and Analyticity. Proccedings of the Eighth Nobel Symposium Held May 19-25, 1968 at Aspenäsgarden, Lerum, in the County of Älvsborg, Sweden*. Nobel symposium, Almqvist & Wiksell, 1968.

[10] E. Fermi, *Nuclear physics: a course given by Enrico Fermi at the University of Chicago*. University of Chicago Press, 1950.

[11] M. D. Schwartz, *Quantum field theory and the standard model*. Cambridge University Press, 2014.

[12] E. Noether, "Invariant variation problems," *Transport Theory and Statistical Physics*, vol. 1, no. 3, pp. 186–207, 1971.

[13] A. Salam, "Weak and electromagnetic interactions," in *Selected Papers Of Abdus Salam: (With Commentary)*, pp. 244–254, World Scientific, 1994.

[14] W. Commons, "Standard Model of Elementary Particles," 2019.

[15] M. Kobayashi and T. Maskawa, "Cp-violation in the renormalizable theory of weak interaction," *Progress of theoretical physics*, vol. 49, no. 2, pp. 652–657, 1973.

[16] C. Giganti, S. Lavignac, and M. Zito, "Neutrino oscillations: the rise of the pmns paradigm," *Progress in Particle and Nuclear Physics*, vol. 98, pp. 1–54, 2018.

[17] M. Thomson, *Modern particle physics*. Cambridge University Press, 2013.

[18] "Standard Model Summary Plots Spring 2020," Tech. Rep. ATL-PHYS-PUB-2020-010, CERN, Geneva, May 2020.

[19] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Physical Review Letters*, vol. 13, no. 9, p. 321, 1964.

[20] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Physical Review Letters*, vol. 13, no. 16, p. 508, 1964.

[21] G. S. Guralnik, C. R. Hagen, and T. W. Kibble, "Global conservation laws and massless particles," *Physical Review Letters*, vol. 13, no. 20, p. 585, 1964.

[22] J. Ellis, "Higgs Physics," pp. 117–168. 52 p, Dec 2013. 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report.

[23] P. D. Group *et al.*, "Review of particle physics," *Chinese Physics C*, vol. 40, no. 10, p. 100001, 2016.

[24] S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, J. Baglio, P. Bolzoni, R. Boughezal, O. Brein, C. Collins-Tooth, S. Dawson, *et al.*, "Handbook of LHC Higgs cross sections: 1. Inclusive observables," *arXiv preprint arXiv:1101.0593*, 2011.

[25] S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, S. Alekhin, J. Alwall, E. Bagnaschi, A. Banfi, J. Blumlein, S. Bolognesi, *et al.*, "Handbook of LHC Higgs cross sections: 2. Differential distributions," *arXiv preprint arXiv:1201.3084*, 2012.

[26] J. Beringer, J. Arguin, R. Barnett, K. Copic, O. Dahl, D. Groom, C. Lin, J. Lys, H. Murayama, C. Wohl, *et al.*, "Review of particle physics," *Physical Review D-Particles, Fields, Gravitation and Cosmology*, vol. 86, no. 1, p. 010001, 2012.

[27] A. D. Sakharov, "Violation of CP-invariance, C-asymmetry, and baryon asymmetry of the Universe," in *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16*, pp. 84–87, World Scientific, 1998.

[28] B. Aubert, A. Bazan, A. Boucham, D. Boutigny, I. De Bonis, J. Favier, J.-M. Gaillard, A. Jeremie, Y. Karyotakis, T. Le Flour, *et al.*, "The babar detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 479, no. 1, pp. 1–116, 2002.

[29] B. Collaboration *et al.*, "The belle detector," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 479, no. 1, pp. 117–232, 2002.

[30] A. A. Alves Jr, L. Andrade Filho, A. Barbosa, I. Bediaga, G. Cernicchiaro, G. Guerrer, H. Lima Jr, A. Machado, J. Magnin, F. Marujo, *et al.*, "The LHCb detector at the LHC," *Journal of instrumentation*, vol. 3, no. 08, p. S08005, 2008.

[31] R. Davis Jr, D. S. Harmer, and K. C. Hoffman, "Search for neutrinos from the sun," *Physical Review Letters*, vol. 20, no. 21, p. 1205, 1968.

[32] F. Capozzi, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, "Neutrino masses and mixings: Status of known and unknown 3ν parameters," *Nuclear Physics B*, vol. 908, pp. 218–234, 2016.

[33] P. Guzowski, L. Barnes, J. Evans, G. Karagiorgi, N. McCabe, and S. Söldner-Rembold, "Combined limit on the neutrino mass from neutrinoless double-β decay and constraints on sterile Majorana neutrinos," *Physical Review D*, vol. 92, no. 1, p. 012002, 2015.

[34] M. Peskin, *An introduction to quantum field theory*. CRC press, 2018.

[35] H. Baer, D. Karatas, and X. Tata, "On the squark and gluino mass limits from the cern pp collider," *Physics Letters B*, vol. 183, no. 2, pp. 220–226, 1987.

[36] A. Lipniacka, "Understanding susy limits from lep," *arXiv preprint hep-ph/0210356*, 2002.

[37] M. Jaffré, "Susy searches at the tevatron," in *EPJ Web of Conferences*, vol. 28, p. 09006, EDP Sciences, 2012.

[38] A. Canepa, "Searches for supersymmetry at the large hadron collider," *Reviews in Physics*, vol. 4, p. 100033, 2019.

[39] Wikimedia Commons, "Higgs quadratic mass renormalization corrections due to stop squarks in a supersymmetric extension to the standard model," 2008. https://en.wikipedia.org/wiki/File:Hqmc-vector.svg.

[40] H. E. Haber and G. L. Kane, "The search for supersymmetry: probing physics beyond the standard model," *Physics Reports*, vol. 117, no. 2-4, pp. 75–263, 1985.

[41] G. C. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, "Theory and phenomenology of two-Higgs-doublet models," *Physics reports*, vol. 516, no. 1-2, pp. 1–102, 2012.

[42] F. Kling, S. Su, and W. Su, "2HDM Neutral Scalars under the LHC," *arXiv preprint arXiv:2004.04172*, 2020.

[43] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. E. Del Valle, M. Flechl, *et al.*, "Search for MSSM Higgs bosons decaying to $\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Physics Letters B*, vol. 798, p. 134992, 2019.

[44] ATLAS Collaboration, "Search for scalar resonances decaying into μ+ μ- in events with and without b-tagged jets produced in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2019, no. 7, p. 117, 2019.

[45] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, *et al.*, "Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks," *Journal of High Energy Physics*, vol. 2015, no. 11, p. 71, 2015.

[46] ATLAS Collaboration, "Search for heavy neutral Higgs bosons produced in association with b-quarks and decaying into b-quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Physical Review D*, vol. 102, no. 3, p. 032004, 2020.

[47] CMS Collaboration, "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV," *Physics Letters B*, vol. 713, no. 2, pp. 68–90, 2012.

[48] CMS Collaboration, "Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions," *Journal of High Energy Physics*, vol. 2014, no. 10, p. 160, 2014.

[49] A. collaboration *et al.*, "Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using pp Collisions at $\sqrt{s}$ = 13 TeV," *Physical Review Letters*, vol. 125, no. 5, 2020.

[50] CMS Collaboration, "Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s}$ = 13 TeV," *Physical Review D*, vol. 98, no. 9, p. 092001, 2018.

[51] CMS Collaboration, "Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s}$ = 8 and 13 TeV," *Physics letters B*, vol. 793, pp. 320–347, 2019.

[52] ATLAS Collaboration, "Search for Scalar Diphoton Resonances in the Mass Range 65–600 GeV with the ATLAS Detector in pp Collision Data at $\sqrt{s = 8}$ TeV," *Physical review letters*, vol. 113, no. 17, p. 171801, 2014.

[53] ATLAS Collaboration, "Search for new phenomena in high-mass diphoton final states using 37 fb^{-1} of proton–proton collisions collected at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Physics Letters B*, vol. 775, pp. 105–125, 2017.

[54] C. collaboration *et al.*, "Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s}$ = 13 TeV," *arXiv preprint arXiv:1908.01115*, 2019.

[55] CMS Collaboration, "Search for a heavy Higgs boson decaying to a pair of W bosons in proton-proton collisions at s $\sqrt{s}$ = 13 TeV," *Journal of High Energy Physics*, vol. 2020, no. 3, pp. 1–54, 2020.

[56] ATLAS Collaboration, "Search for heavy resonances decaying into WW in the $e\nu\mu\nu$ final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *The European Physical Journal C*, vol. 78, no. 1, pp. 1–34, 2018.

[57] CMS Collaboration, "Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at $\sqrt{s}$ = 13 TeV," *Journal of High Energy Physics*, vol. 2018, no. 6, p. 127, 2018.

[58] ATLAS Collaboration, "Search for heavy ZZ resonances in the $\ell^+\ell^-\ell^+\ell^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 78, p. 293, 2018.

[59] T. Kaluza, "Zum unitätsproblem der physik," *Sitzungsber. Preuss. Akad. Wiss. Berlin.(Math. Phys.)*, vol. 1921, no. 966972, pp. 45–46, 1921.

[60] O. Klein, "Quantentheorie und fünfdimensionale Relativitätstheorie," *Zeitschrift für Physik*, vol. 37, no. 12, pp. 895–906, 1926.

[61] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, "The hierarchy problem and new dimensions at a millimeter," *Physics Letters B*, vol. 429, no. 3-4, pp. 263–272, 1998.

[62] L. Randall and R. Sundrum, "Large mass hierarchy from a small extra dimension," *Physical review letters*, vol. 83, no. 17, p. 3370, 1999.

[63] L. Randall and R. Sundrum, "An alternative to compactification," *Physical Review Letters*, vol. 83, no. 23, p. 4690, 1999.

[64] CMS Collaboration, "Search for resonant production of high-mass photon pairs in proton-proton collisions at $\sqrt{s}$ = 8 and 13 TeV," *Physical review letters*, vol. 117, no. 5, p. 051802, 2016.

[65] ATLAS Collaboration, "Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2019, no. 9, p. 91, 2019.

[66] ATLAS Collaboration, "The ATLAS experiment at the CERN large hadron collider," *Jinst*, vol. 3, p. S08003, 2008.

[67] CMS Collaboration, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[68] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN." Dec 2008.

[69] L. Evans and P. Bryant, "LHC machine," *Journal of instrumentation*, vol. 3, no. 08, p. S08001, 2008.

[70] ATLAS Collaboration, "Luminosity Public Results Run 2."

[71] *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS, Geneva: CERN, 1997.

[72] M. Gupta, "Calculation of radiation length in materials," tech. rep., 2010.

[73] P. Krivkova and R. Leitner, "Measurement of the interaction length of pions and protons in the TileCal Module 0," Tech. Rep. ATL-TILECAL-99-007, CERN, Geneva, Mar 1999.

[74] M. Aharrouche, J. Colas, L. Di Ciaccio, M. El Kacimi, O. Gaumer, M. Gouanère, D. Goujdami, R. Lafaye, S. Laplace, C. Le Maner, *et al.*, "Energy linearity and resolution of the atlas electromagnetic barrel calorimeter in an electron test-beam," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 568, no. 2, pp. 601–623, 2006.

[75] ATLAS Collaboration, *ATLAS calorimeter performance: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1996.

[76] ATLAS Collaboration, "Measurement of the photon identification efficiencies with the atlas detector using lhc run-1 data," *The European Physical Journal C*, vol. 76, no. 12, pp. 1–42, 2016.

[77] H. Abreu, M. Aharrouche, M. Aleksa, L. Aperio-Bella, J. Archambault, S. Arfaoui, O. Arnaez, E. Auge, M. Aurousseau, S. Bahinipati, *et al.*, "Performance of the electronic readout of the atlas liquid argon calorimeters," *Journal of Instrumentation*, vol. 5, no. 09, p. P09003, 2010.

[78] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," *The European Physical Journal C*, vol. 77, no. 7, pp. 1–73, 2017.

[79] ATLAS Collaboration, "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data," *Journal of Instrumentation*, vol. 14, Dec 2019.

[80] ATLAS Collaboration, "Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung," tech. rep., CERN, Geneva, May 2012.

[81] ATLAS Collaboration, "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s}$ = 13 TeV," *Eur. Phys. J. C*, vol. 79, p. 639. 40 p, Feb 2019. 63 pages in total, author list starting page 47, 16 figures, 4 tables, final version published in EPJC. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2017-01.

[82] ATLAS Collaboration, "Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data," *The European Physical Journal C*, vol. 77, no. 3, pp. 1–45, 2017.

[83] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1997.

[84] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV," *The European Physical Journal C*, vol. 76, no. 5, p. 292, 2016.

[85] ATLAS Collaboration, "Muon reconstruction and identification efficiency in atlas using the full run 2 pp collision data set at $sqrts$ = 13 tev," *The European Physical Journal C*, vol. 81, no. 7, pp. 1–44, 2021.

[86] ATLAS Collaboration, "Jet reconstruction and performance using particle flow with the ATLAS Detector," *The European Physical Journal C*, vol. 77, no. 7, p. 466, 2017.

[87] M. zur Nedden, "The Run-2 ATLAS Trigger System: Design, Performance and Plan," Tech. Rep. ATL-DAQ-PROC-2016-039, CERN, Geneva, Dec 2016.

[88] ATLAS Collaboration, "Performance of the ATLAS muon triggers in Run 2," *JINST*, vol. 15, p. P09015. 60 p, Apr 2020. 60 pages in total, author list starting page 44, 34 figures, 5 tables, submitted to JINST. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2018-01.

[89] ATLAS Collaboration, "Performance of electron and photon triggers in ATLAS during LHC Run 2," *Eur. Phys. J. C*, vol. 80, p. 47. 56 p, Sep 2019. 56 pages in total, author list starting page 40, 26 figures, 10 tables, published in EPJC. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2018-05.

[90] ATLAS Collaboration, "The ATLAS simulation infrastructure," *The European Physical Journal C*, vol. 70, no. 3, pp. 823–874, 2010.

[91] ATLAS Collaboration, "ATHENA: The ATLAS software framework," 2021.

[92] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *Journal of High Energy Physics*, vol. 2006, no. 05, p. 026, 2006.

[93] S. Frixione, P. Nason, and C. Oleari, "Matching nlo qcd computations with parton shower simulations: the powheg method," *Journal of High Energy Physics*, vol. 2007, no. 11, p. 070, 2007.

[94] G. Marchesini, B. R. Webber, G. Abbiendi, I. Knowles, M. H. Seymour, and L. Stanco, "Herwig 5.1-a monte carlo event generator for simulating hadron emission reactions with interfering gluons," *Computer Physics Communications*, vol. 67, no. 3, pp. 465–508, 1992.

[95] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *Journal of High Energy Physics*, vol. 2009, no. 02, p. 007, 2009.

[96] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. . Barrand, *et al.*, "GEANT4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[97] J. Shiers, "The worldwide lhc computing grid (worldwide lcg)," *Computer physics communications*, vol. 177, no. 1-2, pp. 219–223, 2007.

[98] D. Bourilkov, "Machine and deep learning applications in particle physics," *International Journal of Modern Physics A*, vol. 34, no. 35, p. 1930019, 2019.

[99] A. Richards, "University of Oxford Advanced Research Computing," *Zenodo Document DOI 10*, 2015.

[100] D0 Collaboration, "Observation of single top-quark production," *Physical Review Letters*, vol. 103, no. 9, p. 092001, 2009.

[101] CDF Collaboration, "Observation of electroweak single top-quark production," *Phys. Rev. Lett.*, vol. 103, p. 092002, Aug 2009.

[102] M. Paganini, L. de Oliveira, and B. Nachman, "CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks," *Physical Review D*, vol. 97, no. 1, p. 014021, 2018.

[103] J. Albrecht, A. A. Alves, G. Amadio, G. Andronico, N. Anh-Ky, L. Aphecetche, J. Apostolakis, M. Asai, L. Atzori, M. Babik, *et al.*, "A Roadmap for HEP Software and Computing R&D for the 2020s," *Computing and software for big science*, vol. 3, no. 1, p. 7, 2019.

[104] A. De Simone and T. Jacques, "Guiding new physics searches with unsupervised learning," *The European Physical Journal C*, vol. 79, no. 4, pp. 1–15, 2019.

[105] K. Albertsson *et al.*, "Machine Learning in High Energy Physics Community White Paper," *J. Phys. Conf. Ser.*, vol. 1085, no. 2, p. 022008, 2018.

[106] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[107] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[108] A. Cauchy, "Méthode générale pour la résolution des systemes d'équations simultanées," *Comp. Rend. Sci. Paris*, vol. 25, no. 1847, pp. 536–538, 1847.

[109] H. Robbins and S. Monro, "A stochastic approximation method," *The annals of mathematical statistics*, pp. 400–407, 1951.

[110] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[111] R. E. Schapire, "A brief introduction to boosting," in *Ijcai*, vol. 99, pp. 1401–1406, 1999.

[112] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, pp. 3146–3154, Curran Associates, Inc., 2017.

[113] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM.

[114] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1, pp. 278–282, IEEE, 1995.

[115] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, *et al.*, "TMVA-toolkit for multivariate data analysis," *arXiv preprint physics/0703039*, 2007.

[116] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[117] A. LeNail, "Nn-svg: Publication-ready neural network architecture schematics," *Journal of Open Source Software*, vol. 4, no. 33, p. 747, 2019.

[118] P. Werbos, "New tools for prediction and analysis in the behavioral sciences," *Ph. D. dissertation, Harvard University*, 1974.

[119] D. Parker, *Learning-logic: Casting the Cortex of the Human Brain in Silicon*. Technical report: Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science, 1985.

[120] Y. LeCun, "Une procedure d'apprentissage pour reseau a seuil asymetrique. proceedings of Cognitiva 85," 1985.

[121] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[122] J. Makin, "Lecture notes on a University of California, Berkeley course on Deep Neural Network CS182," February 2006.

[123] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[124] "Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment," Tech. Rep. ATL-PHYS-PUB-2017-003, CERN, Geneva, Mar 2017.

[125] C. Olah, "Understanding lstm networks," 2015.

[126] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[127] Y. Bengio and Y. Grandvalet, *Bias in Estimating the Variance of K-Fold Cross-Validation*, pp. 75–95. Boston, MA: Springer US, 2005.

[128] ATLAS Collaboration, "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$ = 7 and 8 TeV in the ATLAS experiment," *Eur. Phys. J. C*, vol. 76, p. 6, 2016.

[129] ATLAS Collaboration, "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector," *Eur. Phys. J. C*, vol. 75, p. 476, 2015.

[130] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV," *Eur. Phys. J. C*, vol. 75, p. 212, 2015.

[131] CMS Collaboration, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at $\sqrt{s}$ = 7 and 8 TeV," *Phys. Rev. D*, vol. 92, p. 012004, 2015.

[132] H. E. Haber and G. L. Kane, "The Search for Supersymmetry: Probing Physics Beyond the Standard Model," *Phys. Rept.*, vol. 117, pp. 75–263, 1985.

[133] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, "The Hierarchy problem and new dimensions at a millimeter," *Phys. Lett. B*, vol. 429, pp. 263–272, 1998.

[134] G. F. Giudice, "Naturally Speaking: The Naturalness Criterion and Physics at the LHC," 2008.

[135] C. Amsler, M. Doser, P. Bloch, A. Ceccucci, G. Giudice, A. Höcker, M. Mangano, A. Masoni, S. Spanier, N. Törnqvist, *et al.*, "Review of particle physics," *Physics Letters B*, vol. 667, no. 1-5, pp. 1–6, 2008.

[136] H. Davoudiasl, J. Hewett, and T. Rizzo, "Bulk gauge fields in the Randall–Sundrum model," *Physics Letters B*, vol. 473, no. 1-2, pp. 43–49, 2000.

[137] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, "Warped gravitons at the CERN LHC and beyond," *Physical Review D*, vol. 76, no. 3, p. 036006, 2007.

[138] M. Cano Bret, A. Schaffer, A. Gabrielli, D. Denysiuk, R. Nicolaidou, W. Su, R. F. Naranjo Garcia, W. A. Leight, J. W. Carter, X. Ju, G. Artoni, C. Anastopoulos, G. Barone, M. Goblirsch-Kolb, P. Podberezko, S. E. Von Buddenbrock, T. Lagouri, L. Xu, and S. Hassani, "Search for heavy ZZ resonances in the $\ell^{|}\ell^{-}\ell^{+}\ell^{-}$final state using proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," tech. rep., CERN, Geneva, Jul 2018.

[139] A. Collaboration, "ATLAS data quality operations and performance for 2015-2018 data-taking," *Journal of Instrumentation*, vol. 15, 2020.

[140] P. Nason and G. Zanderighi, "$W^{+}W^{-}$, WZ and ZZ production in the POWHEG-BOX-V2," *The European Physical Journal C*, vol. 74, no. 1, p. 2702, 2014.

[141] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1," *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, 2008.

[142] D. J. Lange, "The EvtGen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1-2, pp. 152–155, 2001.

[143] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan, "New parton distributions for collider physics," *Physical Review D*, vol. 82, no. 7, p. 074024, 2010.

[144] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *Journal of High Energy Physics*, vol. 2014, no. 7, p. 79, 2014.

[145] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, *et al.*, "Parton distributions for the LHC Run II," *Journal of High Energy Physics*, vol. 2015, no. 4, p. 40, 2015.

[146] S. Schumann and F. Krauss, "A parton shower algorithm based on Catani-Seymour dipole factorisation," *Journal of High Energy Physics*, vol. 2008, no. 03, p. 038, 2008.

[147] S. Hoeche, F. Krauss, M. Schönherr, and F. Siegert, "QCD matrix elements+ parton showers. The NLO case," *Journal of High Energy Physics*, vol. 2013, no. 4, p. 27, 2013.

[148] F. Cascioli, P. Maierhöfer, and S. Pozzorini, "Scattering amplitudes with open loops," *Physical review letters*, vol. 108, no. 11, p. 111601, 2012.

[149] T. Gleisberg and S. Höche, "Comix, a new matrix element generator," *Journal of High Energy Physics*, vol. 2008, no. 12, p. 039, 2008.

[150] P. Z. Skands, "Tuning Monte Carlo generators: the perugia tunes," *Physical Review D*, vol. 82, no. 7, p. 074018, 2010.

[151] P. Golonka and Z. Was, "PHOTOS Monte Carlo: a precision tool for QED correctionsin Z and W decays," *The European Physical Journal C-Particles and Fields*, vol. 45, no. 1, pp. 97–107, 2006.

[152] S. Jadach, Z. Was, R. Decker, and J. Kühn, "The τ decay library TAUOLA, version 2.4," *Computer Physics Communications*, vol. 76, no. 3, pp. 361–380, 1993.

[153] P. Golonka, B. Kersevan, T. Pierzchała, E. Richter-Was, Z. Was, and M. Worek, "The tauola-photos-F environment for the TAUOLA and PHOTOS packages, release II," *Computer Physics Communications*, vol. 174, no. 10, pp. 818–835, 2006.

[154] A. Collaboration *et al.*, "Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach," tech. rep., ATL-PHYS-PUB-2017-022, 2017.

[155] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

[156] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual," *The European Physical Journal C*, vol. 72, no. 3, p. 1896, 2012.

[157] ATLAS Collaboration, "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at s= 13 TeV with the ATLAS detector," *Physical Review D*, vol. 96, no. 7, p. 072002, 2017.

[158] ATLAS Collaboration, "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector," *The European Physical Journal C*, vol. 76, no. 11, p. 581, 2016.

[159] ATLAS Collaboration, "Measurement of inclusive and differential cross sections in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. Measurement of inclusive and differential cross sections in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *JHEP*, vol. 10, p. 132. 49 p, Aug 2017. 49 pages in total, author list starting page 33, 11 figures, 4 tables, published in JHEP. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2016-25/.

[160] M. Agustoni, H. P. Beck, A. Cervelli, A. Ereditato, V. Gallo, S. Haug, T. Kruker, L. Marti, F. Meloni, B. Schneider, *et al.*, "Measurement of the higgs boson mass from the $h \rightarrow \gamma\gamma$ and $h \rightarrow zz4\ell$ collisions at center-of-mass energies of 7 and 8 tev with the atlas detector," *Physical review. D-particles, fields, gravitation, and cosmology*, vol. 90, no. 5, p. 052004, 2014.

[161] ATLAS Collaboration, "Measurements of Higgs boson production and couplings in the four-lepton channel in p p collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector," *Physical Review D*, vol. 91, no. 1, p. 012006, 2015.

[162] M. Oreglia, E. Bloom, F. Bulos, R. Chestnut, J. Gaiser, G. Godfrey, C. Kiesling, W. Lockman, D. Scharre, R. Partridge, *et al.*, "Study of the reaction $\psi \rightarrow \gamma\ \gamma$ J ψ," *Physical Review D*, vol. 25, no. 9, p. 2259, 1982.

[163] J. E. Gaiser, "Charmonium Spectroscopy from Radiative Decays of the J/ψ and ψ," 1984.

[164] N. Kauer and C. O'Brien, "Heavy Higgs signal–background interference in $gg \rightarrow VV$ in the Standard Model plus real singlet," *The European Physical Journal C*, vol. 75, no. 8, p. 374, 2015.

[165] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving,

M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems." Software available from tensorflow.org.

[166] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 4765–4774, Curran Associates, Inc., 2017.

[167] L. S. Shapley, "A value for n-person games," *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 307–317, 1953.

[168] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, pp. 4765–4774, 2017.

[169] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[170] TeamHG-Memex, "eli5," 2021.

[171] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, "Hyperopt: a Python library for model selection and hyperparameter optimization," *Computational Science & Discovery*, vol. 8, no. 1, p. 014008, 2015.

[172] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of machine learning research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[173] W. Buttinger and M. Lefebvre, "Formulae for Estimating Significance," Tech. Rep. ATL-COM-GEN-2018-026, CERN, Geneva, Oct 2018.

[174] S. Algeri, J. Aalbers, K. D. Morå, and J. Conrad, "Searching for new phenomena with profile likelihood ratio tests," *Nature Reviews Physics*, pp. 1–8, 2020.

[175] A. L. Read, "Presentation of search results: the CL_S technique," *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, p. 2693, 2002.

[176] B. Mistlberger and F. Dulat, "Limit setting procedures and theoretical uncertainties in higgs boson searches," *arXiv preprint arXiv:1204.3851*, 2012.

[177] ATLAS and CMSCollaborations, LHC Higgs Combination Group, "Procedure for the lhc higgs boson search combination in summer 2011," tech. rep., Technical Report ATL-PHYS-PUB 2011-11, CMS NOTE 2011/005, 2011.

[178] C. Geng, Y. Wu, B. S. Stapf, G. Barone, A. Gabrielli, A. Schaffer, M. J. Veen, A. E. Mcdougall, P. Kluit, H. L. Snoek, I. Van Vulpen, D. Krasnopevtsev, L. Han, F.-y. Tsai, S. Heim, J. A. Sabater Iglesias, Y. C. Yap, B. Heinemann, J. Li, B. Zhou, H. Yang, K. D. McLean, C. R. Anelli, K. Hamano, M. Lefebvre, A. Basalaev, I. Naryshkin, H. Zhu, M. Cano Bret, W. A. Leight, and J. Gao, "Search for heavy resonances in the $H' \to ZZ \to \ell\ell\nu\nu$ channel with proton-proton collisions at $\sqrt{s}$ = 13 TeV for 2015-2018 data with the ATLAS detector," Tech. Rep. ATL-COM-PHYS-2018-1527, CERN, Geneva, Nov 2018.

[179] ATLAS Collaboration, "Combined measurements of higgs boson production and decay using up to 80 fb- 1 of proton-proton collision data at s= 13 tev collected with the atlas experiment," *Physical Review D*, vol. 101, no. 1, p. 012002, 2020.

[180] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.

[181] ATLAS Collaboration, "Search for high-mass dilepton resonances in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector," *Physical Review D*, vol. 90, no. 5, p. 052005, 2014.

[182] G. Ordonez Sanz, *Muon identification in the ATLAS calorimeters*. PhD thesis, Nijmegen U., 2009. Presented on 12 Jun 2009.

[183] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger, "The new atlas track reconstruction (newt)," in *Journal of Physics: Conference Series*, vol. 119, p. 032014, IOP Publishing, 2008.

[184] M. C. Aleksa, W. P. Cleland, Y. T. Enari, M. V. Fincke-Keeler, L. C. Hervas, F. B. Lanni, S. O. Majewski, C. V. Marino, and I. L. Wingerter-Seez, "ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report," tech. rep., Sep 2013. Final version presented to December 2013 LHCC.

[185] LHC Experiments Committee, *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS, Geneva: CERN, 1996.

[186] F. Carminati, G. Khattak, M. Pierini, S. Vallecor-safa, and A. Farbin, "Calorimetry with deep learning: particle classification, energy regression, and simulation for high-energy physics," in *NIPS*, 2017.

[187] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[188] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[189] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[190] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[191] L. Sifre and S. Mallat, "Rigid-motion scattering for image classification," *Ph. D. thesis*, 2014.

[192] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261*, 2016.

[193] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[194] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[195] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[196] H. Bethe, "Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie," *Annalen der Physik*, vol. 397, no. 3, pp. 325–400, 1930.

[197] Y. Li, T. Tsang, C. Thorn, X. Qian, M. Diwan, J. Joshi, S. Kettell, W. Morse, T. Rao, J. Stewart, *et al.*, "Measurement of longitudinal electron diffusion in liquid argon," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 816, pp. 160–170, 2016.

[198] D. R. Lide, *CRC handbook of chemistry and physics*, vol. 85. CRC press, 2004.

[199] R. D. Ball, S. Carrazza, L. Del Debbio, S. Forte, J. Gao, N. Hartland, J. Huston, P. Nadolsky, J. Rojo, D. Stump, *et al.*, "Parton distribution benchmarking with lhc data," *Journal of High Energy Physics*, vol. 2013, no. 4, pp. 1–46, 2013.

[200] S. Goria, G. Passarino, and D. Rosco, "The Higgs-boson lineshape," *Nuclear Physics B*, vol. 864, no. 3, pp. 530–579, 2012.

[201] M. Spira, A. Djouadi, D. Graudenz, and R. Zerwas, "Higgs boson production at the lhc," *Nuclear Physics B*, vol. 453, no. 1-2, pp. 17–82, 1995.

[202] J. Bijnens, P. Eerola, M. Maul, A. Månsson, and T. Sjöstrand, "QCD signatures of narrow graviton resonances in hadron colliders," *Physics Letters B*, vol. 503, no. 3-4, pp. 341–348, 2001.

[203] J. M. Campbell and R. K. Ellis, "Update on vector boson pair production at hadron colliders," *Physical Review D*, vol. 60, no. 11, p. 113006, 1999.