

# **A Machine Learning-Based Search for Axion-Like Particles in Exotic Higgs Boson Decays to Four Photons with the ATLAS Detector**



JOHANNES GUTENBERG  
UNIVERSITÄT MAINZ

Peter Krämer

Mainz, 23.07.2025

Sofern nicht anders angegeben, steht diese Dissertation unter der Lizenz Creative Commons Namensnennung 4.0 International (CC BY 4.0).

|                    |                                                                                     |
|--------------------|-------------------------------------------------------------------------------------|
| 1. Gutachter:      | Prof. Dr. Matthias Schott                                                           |
| 2. Gutachter:      |  |
| Tag der Promotion: | 05.12.2025                                                                          |
| Erscheinungsjahr:  | 2025                                                                                |

---

# Eigenständigkeitserklärung

---

Hiermit erkläre ich,

dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel (dazu zählen auch KI-basierte Anwendungen oder Werkzeuge\*) benutzt habe. Sämtliche wörtlichen oder sinngemäßen Übernahmen und Zitate sind kenntlich gemacht und nachgewiesen.

Im Anhang „Nutzung KI-Tools“ habe ich die verwendeten KI-Tools dokumentiert.

Mit Abgabe der vorliegenden Leistung übernehme ich die Verantwortung für das eingereichte Gesamtprodukt. Ich verantworte damit auch jegliche KI-generierten Inhalte, die ich in meine Arbeit übernommen habe. Die Richtigkeit übernommener (KI-generierter) Aussagen und Inhalte habe ich nach bestem Wissen und Gewissen geprüft.

Mir ist bekannt, dass ein Verstoß gegen die genannten Punkte prüfungsrechtliche Konsequenzen hat.

---

Ort, Datum und Unterschrift





---

# Abstract

---

## Abstract

The Standard Model (SM) of particle physics, while remarkably successful, leaves several phenomena unexplained — including the strong CP problem and the anomalous magnetic moment of the muon  $((g - 2)_\mu)$ . Axion-like particles (ALPs) provide a theoretically motivated extension to the SM that could address both.

This thesis presents a search for ALPs produced in exotic Higgs boson decays ( $H \rightarrow aa$ ), with a four-photon final state ( $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ ), using LHC proton–proton collision data recorded by the ATLAS detector from 2015 to 2018. The analysis covers a broad phase space and employs advanced machine learning techniques, including convolutional neural networks, to tackle key reconstruction challenges.

No significant excess is observed, and stringent limits are set on the ALP–photon coupling and  $H \rightarrow aa$  branching ratio. These results constrain regions compatible with the  $(g - 2)_\mu$  anomaly and provide the most sensitive limits in this channel to date.

## Kurzzusammenfassung

Das Standardmodell (SM) der Teilchenphysik beschreibt fundamentale Teilchen und ihre Wechselwirkungen äußerst erfolgreich, kann jedoch zentrale Beobachtungen nicht erklären – darunter das CP-Problem der starken Wechselwirkung und die Diskrepanz im anomalen magnetischen Moment des Myons  $((g - 2)_\mu)$ . Die hypothetischen, Axionen ähnelnden (axion-like) Partikel (ALPs) stellen eine theoretisch gut motivierte Erweiterung des SM dar, die beide Probleme lösen könnte.

Diese Arbeit beschreibt eine Suche nach ALPs in exotischen Zerfällen des Higgs-Bosons ( $H \rightarrow aa$ ) mit einem charakteristischen Endzustand mit vier Photonen ( $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ ), basierend auf Proton–Proton-Kollisionsdaten, die zwischen 2015 und 2018 mit dem ATLAS-Detektor am LHC aufgenommen wurden. Die Analyse deckt einen breiten Phasenraum ab und verwendet fortgeschrittene Methoden des maschinellen Lernens, darunter Convolutional Neural Networks, um experimentelle Herausforderungen wie überlappende Photonen und Zerfälle mit versetztem Vertex zu bewältigen.

Es wird keine signifikante Abweichung gegenüber dem SM beobachtet und damit ALPs in diesem Phasenraum ausgeschlossen. Daraufhin werden Ausschlussgrenzen auf die ALP-Photon-Kopplung und die  $H \rightarrow aa$ -Zerfallswahrscheinlichkeit gesetzt. Diese schließen Parameterbereiche aus, die die  $(g - 2)_\mu$ -Anomalie erklären könnten und stellen die bislang genauesten Ergebnisse in diesem Bereich dar.



---

# Contents

---

|                                                             |            |
|-------------------------------------------------------------|------------|
| <b>Eigenständigkeitserklärung</b>                           | <b>iii</b> |
| <b>Abstract</b>                                             | <b>v</b>   |
| <b>1 Introduction</b>                                       | <b>1</b>   |
| 1.1 Introduction . . . . .                                  | 1          |
| <b>2 Theoretical Foundations</b>                            | <b>3</b>   |
| 2.1 The Standard Model of Particle Physics . . . . .        | 3          |
| 2.1.1 Fundamental interactions . . . . .                    | 3          |
| 2.1.2 Elementary particles . . . . .                        | 4          |
| 2.1.3 Fermions . . . . .                                    | 5          |
| 2.1.4 Bosons . . . . .                                      | 8          |
| 2.2 Higgs mechanism and Higgs boson . . . . .               | 10         |
| 2.2.1 Proton-Proton Collisions . . . . .                    | 14         |
| 2.3 Limitations of the Standard Model . . . . .             | 15         |
| 2.3.1 The Strong CP Problem . . . . .                       | 15         |
| 2.3.2 Dark Matter . . . . .                                 | 16         |
| 2.3.3 The Muon $(g - 2)_\mu$ Discrepancy . . . . .          | 16         |
| <b>3 Axions and Axion-Like Particles</b>                    | <b>17</b>  |
| 3.1 Theoretical Motivation . . . . .                        | 17         |
| 3.1.1 The Strong CP Problem and QCD Axions . . . . .        | 17         |
| 3.1.2 Axion-Like Particles (ALPs) . . . . .                 | 18         |
| 3.1.3 ALPs and the Muon Anomalous Magnetic Moment . . . . . | 18         |
| 3.1.4 Solar ALPs and the Primakoff Effect . . . . .         | 18         |
| 3.2 Searches for axions and ALPs . . . . .                  | 19         |
| 3.2.1 Haloscope Experiments . . . . .                       | 19         |
| 3.2.2 Helioscope Experiments . . . . .                      | 19         |
| 3.2.3 Light-Through-Wall Experiments . . . . .              | 20         |
| 3.3 Collider searches for ALPs . . . . .                    | 21         |
| 3.3.1 ALPs from exotic Higgs boson decays . . . . .         | 22         |
| <b>4 Experimental Environment</b>                           | <b>25</b>  |
| 4.1 The Large Hadron Collider . . . . .                     | 25         |

|          |                                                                                                                  |           |
|----------|------------------------------------------------------------------------------------------------------------------|-----------|
| 4.2      | The ATLAS detector                                                                                               | 26        |
| 4.2.1    | The ATLAS liquid Argon calorimeter                                                                               | 28        |
| 4.2.2    | The ATLAS Trigger                                                                                                | 28        |
| 4.2.3    | Object reconstruction in ATLAS                                                                                   | 29        |
| 4.2.4    | Muon Data Quality in the ATLAS Experiment                                                                        | 31        |
| <b>5</b> | <b>Experimental methods</b>                                                                                      | <b>35</b> |
| 5.1      | Machine Learning                                                                                                 | 35        |
| 5.1.1    | General concept of supervised learning                                                                           | 36        |
| 5.1.2    | Artificial neural networks (ANNs)                                                                                | 37        |
| 5.1.3    | Convolutional neural networks (CNNs)                                                                             | 38        |
| 5.1.4    | Common Neural Network Layers                                                                                     | 38        |
| 5.1.5    | Overfitting and regularisation                                                                                   | 42        |
| 5.1.6    | Model Evaluation and Performance Metrics                                                                         | 43        |
| 5.2      | Maximum Likelihood Fits                                                                                          | 47        |
| 5.2.1    | Likelihood Function                                                                                              | 47        |
| 5.2.2    | Maximum Likelihood Estimation                                                                                    | 47        |
| 5.2.3    | Profiling and Uncertainty Estimation                                                                             | 48        |
| 5.2.4    | The $CL_s$ Method                                                                                                | 48        |
| <b>6</b> | <b>Search for long-lived axions in <math>H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma</math> decays</b> | <b>49</b> |
| 6.1      | Analysis strategy                                                                                                | 49        |
| 6.2      | Data set and event simulation                                                                                    | 52        |
| 6.2.1    | Signal simulation                                                                                                | 52        |
| 6.2.2    | Background simulation                                                                                            | 53        |
| 6.3      | Event selection                                                                                                  | 54        |
| 6.3.1    | Trigger studies                                                                                                  | 54        |
| 6.3.2    | Preselections                                                                                                    | 56        |
| 6.4      | Photon classification                                                                                            | 57        |
| 6.4.1    | Training data                                                                                                    | 57        |
| 6.4.2    | Classifier 1: single + merged vs fake                                                                            | 58        |
| 6.4.3    | Classifier 2: single vs merged                                                                                   | 60        |
| 6.4.4    | Two photon energy reconstruction                                                                                 | 61        |
| 6.4.5    | ALP reconstruction                                                                                               | 62        |
| 6.5      | Event categorization                                                                                             | 64        |
| 6.6      | Background estimation                                                                                            | 67        |
| 6.6.1    | Background estimation for two photon categories                                                                  | 68        |
| 6.7      | Uncertainties                                                                                                    | 70        |
| 6.7.1    | Systematic uncertainties                                                                                         | 70        |
| 6.7.2    | Magnitude of uncertainties                                                                                       | 73        |
| 6.8      | Statistical analysis                                                                                             | 73        |
| 6.8.1    | Fit strategy                                                                                                     | 73        |
| 6.8.2    | Statistical analysis for $m_a = 1$ GeV and $C_{a\gamma\gamma} = 1$                                               | 76        |
| 6.9      | Results                                                                                                          | 77        |

|          |                                                                          |            |
|----------|--------------------------------------------------------------------------|------------|
| <b>7</b> | <b>Analysis optimised on promptly decaying ALPs</b>                      | <b>83</b>  |
| 7.1      | Selection and background estimation . . . . .                            | 83         |
| 7.2      | Statistical analysis and results . . . . .                               | 86         |
| <b>8</b> | <b>Merged photon identification with CNNs</b>                            | <b>89</b>  |
| 8.1      | Calorimeter data and data preparation . . . . .                          | 89         |
| 8.2      | Photon identification with multiclass CNN . . . . .                      | 90         |
| 8.2.1    | CNN architecture . . . . .                                               | 91         |
| 8.2.2    | CNN training . . . . .                                                   | 92         |
| 8.2.3    | CNN evaluation . . . . .                                                 | 92         |
| 8.2.4    | Comparison with a shower-shape based classifier . . . . .                | 94         |
| 8.2.5    | $p_T$ dependence . . . . .                                               | 94         |
| 8.3      | ALP mass regression with a CNN . . . . .                                 | 97         |
| 8.4      | Future studies . . . . .                                                 | 102        |
| <b>9</b> | <b>Conclusion</b>                                                        | <b>105</b> |
| <b>A</b> | <b>Appendix</b>                                                          | <b>107</b> |
| A.1      | Background estimation for the three and four photon categories . . . . . | 107        |
| A.2      | Uncertainties due to displaced vertices . . . . .                        | 110        |
| A.3      | Trigger studies . . . . .                                                | 114        |
| A.4      | Good Run Lists . . . . .                                                 | 115        |
| A.5      | Model Architecture and Weight Visualization . . . . .                    | 116        |
| A.5.1    | Model Architecture . . . . .                                             | 116        |
| A.5.2    | Understanding Filter Weights . . . . .                                   | 116        |
| A.6      | $p_T$ reweighting for CNNs . . . . .                                     | 116        |
| A.7      | Different $m_a$ regression results . . . . .                             | 125        |
| A.8      | Use of AI Tools . . . . .                                                | 128        |
|          | <b>Bibliography</b>                                                      | <b>129</b> |
|          | <b>List of Figures</b>                                                   | <b>143</b> |
|          | <b>List of Tables</b>                                                    | <b>151</b> |



---

# Introduction

---

## 1.1 Introduction

The Standard Model (SM) of particle physics is one of the most successful theories in science, providing a remarkably accurate description of the fundamental particles and their interactions. With the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC), the final missing piece of the SM was experimentally confirmed. However, despite its many triumphs, the SM is incomplete. It fails to account for several observed phenomena such as the strong CP problem, the nature of dark matter, neutrino masses, and the matter-antimatter asymmetry of the Universe.

One particularly intriguing discrepancy lies in the anomalous magnetic moment of the muon  $(g - 2)_\mu$ , where a persistent tension exists between the theoretical SM prediction and experimental measurements. At the same time, the strong CP problem – why QCD does not appear to violate CP symmetry despite allowing for it – remains unresolved. These puzzles motivate the search for new particles beyond the SM, including the axion, originally proposed to solve the strong CP problem, and its generalization: axion-like particles (ALPs).

ALPs are hypothetical pseudo-scalar bosons that arise in many extensions of the SM, including string theory and models of spontaneous symmetry breaking at high energies. They are characterised by a small mass and a generic coupling to photons, offering a rich phenomenology. Depending on their parameters, ALPs could not only explain the  $(g - 2)_\mu$  anomaly but also serve as portals to hidden sectors or contribute to the dark matter density.

High-energy colliders such as the LHC provide a unique environment to search for ALPs through rare and exotic decays. One promising signature is the decay of the Higgs boson into a pair of ALPs ( $H \rightarrow aa$ ), with each ALP subsequently decaying into two photons, yielding a final state of four photons ( $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ ). This channel offers sensitivity to ALPs over a broad range of masses and couplings, especially in previously unexplored regions involving displaced decays or collimated photon signatures.

This thesis presents a dedicated search for ALPs in the  $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$  final state using proton-proton collision data collected by the ATLAS detector between 2015 and 2018. The analysis covers a wide phase space and employs advanced machine learning techniques – including neural networks and convolutional architectures – to overcome experimental challenges such as identifying merged photons and handling displaced decay topologies.

The results place stringent constraints on the parameter space of ALPs, excluding regions compatible with the  $(g - 2)_\mu$  anomaly, and set the most competitive limits to date in several kinematic regimes. In addition, novel approaches to photon classification and ALP mass reconstruction are explored, offering avenues for enhanced sensitivity in future searches.

The main results of this thesis have been published in [1].

The thesis is structured as follows.

Chapter 1 introduces the broader context and motivation of the work.

Chapter 2 lays the theoretical foundation, covering the Standard Model of particle physics, including its fundamental particles and interactions, as well as the Higgs mechanism. Beyond-Standard-Model phenomena such as the muon  $(g - 2)_\mu$  anomaly and hypothetical solutions including axions and axion-like particles (ALPs) are also introduced.

Section 3.2 provides an overview of current experimental searches for ALPs, with particular emphasis on collider-based approaches.

Chapter 4 introduces the experimental environment, including the Large Hadron Collider (LHC) and the ATLAS detector, highlighting key subsystems relevant for the analysis. Object reconstruction methods used in the experiment are also described.

Chapter 5 details the methodologies employed throughout the thesis, such as machine learning techniques and statistical inference methods including likelihood fits and the  $\text{CL}_s$  method.

The core analysis is presented in Chapter 6, which outlines the analysis strategy, data sets and simulation, event selection, photon classification techniques, background estimation, treatment of systematic uncertainties, statistical interpretation, and final results.

A complementary analysis optimised for promptly decaying ALPs is described in Chapter 7, focusing on selection criteria, background estimation, and statistical interpretation.

Chapter 8 explores advanced machine learning approaches, particularly convolutional neural networks (CNNs), applied to raw calorimeter data for photon classification and ALP mass regression in single-photon topologies.

Finally, Chapter 9 concludes the thesis by summarising the main findings and outlining potential directions for future research.



---

## Theoretical Foundations

---

Physics advances through a dual approach to exploring the laws of nature. On one hand, experiments are essential for collecting data on various phenomena and quantifying them through measurement. On the other hand, theoretical models provide the framework to interpret, explain, and predict experimental results.

This chapter provides a brief overview of the theoretical foundations of particle physics, with a particular focus on the analysis presented in this thesis. Section 2.1 introduces the Standard Model of particle physics (SM), outlining its key principles and relevance to the study at hand. The concept of axions and ALPs is described in Section 3.1.

### 2.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics is a theoretical framework that describes three of the four fundamental forces, as well as all currently known fundamental particles. It has been extensively tested and is considered self-contained following the discovery of the Higgs boson [2, 3]. A comprehensive overview of the Standard Model, with a focus on the Higgs boson and its discovery, can be found in [4].

The SM is a Quantum Field Theory (QFT) that can be written as gauge theory with  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge symmetry group. The  $SU(3)_C$  symmetry defines the structure of the **strong force**, with its charge often referred to as **colour charge**. The  $SU(2)_L$  symmetry governs the **weak force**, while **electric charge** arises from a combination of the  $SU(2)_L$  and  $U(1)_Y$  symmetries. More information about QFTs can be found in [5].

#### 2.1.1 Fundamental interactions

All known physical phenomena can be described by four fundamental interactions: the gravitational, electromagnetic, weak, and strong forces. Among these, only three are currently included in the Standard Model of Particle Physics: the electromagnetic, weak, and strong interactions. Gravity, while fundamental, is not yet described within a quantum field theory and lies beyond the scope of the Standard Model.

Each of the three Standard Model interactions is associated with a specific gauge symmetry and a corresponding force mediator:

- **The electromagnetic force** is described by Quantum Electrodynamics (QED), based on a  $U(1)$  gauge symmetry. It is mediated by the massless photon, and couples to electric charge [6, 7].
- **The weak force** is described by the electroweak theory, which unifies the weak and electromagnetic forces under a combined  $SU(2)_L \times U(1)_Y$  symmetry [8]. It is mediated by the  $W^\pm$  and  $Z^0$  bosons, which are responsible for flavour-changing and neutral-current processes. These bosons are experimentally observed to be massive [9–11].
- **The strong force** is described by Quantum Chromodynamics (QCD), based on an  $SU(3)_C$  gauge symmetry. It acts on colour-charged particles (quarks) and is mediated by eight massless gluons [12, 13].

In quantum field theory, gauge invariance is a central principle that requires all force-mediating particles (gauge bosons) to be massless. This leads to a theoretical problem: the  $W$  and  $Z$  bosons are clearly massive, which would naively violate gauge invariance.

This apparent contradiction is resolved through the mechanism of **spontaneous symmetry breaking**. Within the electroweak sector, the Higgs field acquires a non-zero vacuum expectation value (VEV), breaking the gauge symmetry and giving mass to the  $W^\pm$  and  $Z^0$  bosons – while leaving the photon massless. This process, known as the **Higgs mechanism**, also gives rise to the scalar Higgs boson [14–18].

The following sections describe the elementary particles – fermions and bosons – that interact via these fundamental forces.

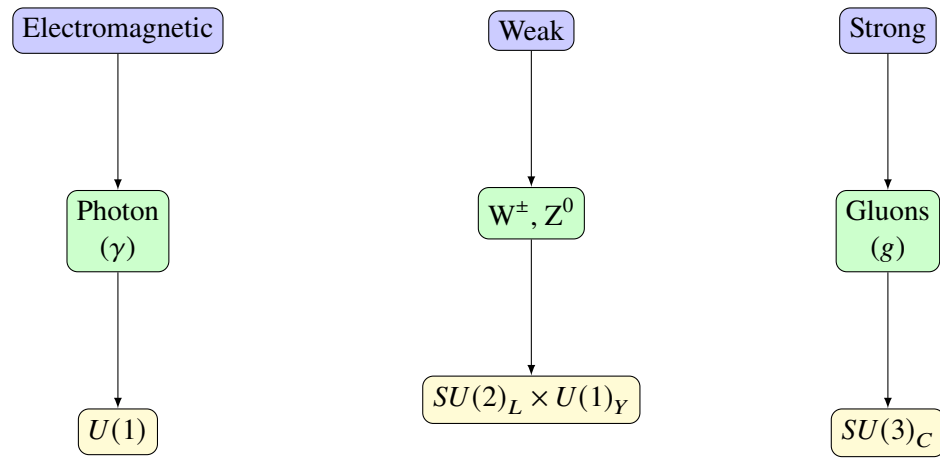


Figure 2.1: Schematic of the fundamental interactions in the Standard Model, their corresponding gauge bosons, and the underlying gauge symmetries.

### 2.1.2 Elementary particles

Elementary particles are classified into two categories based on their spin: **bosons**, which have integer spin, and **fermions**, which have half-integer spin.

Fermions can be further divided into quarks and leptons, each organized into three generations. Quarks carry colour charge as well as weak and electric charge, allowing them to interact via all fundamental

forces. Leptons, on the other hand, do not interact via the strong force, but only via the weak force and, if electrically charged, also through electromagnetic interactions.

Stable matter is only made of first generation particles, namely up and down quarks as well as electrons. Other fermions have shorter lifetimes and decay to first generation fermions subsequently.

Photons ( $\gamma$ ), gluons,  $W^\pm$  and  $Z^0$  bosons are mediators of fundamental forces, thus called gauge bosons and have a spin of one, while the Higgs boson has a spin of zero.

All elementary particles of the SM and some of their properties are shown in Fig. 2.2.

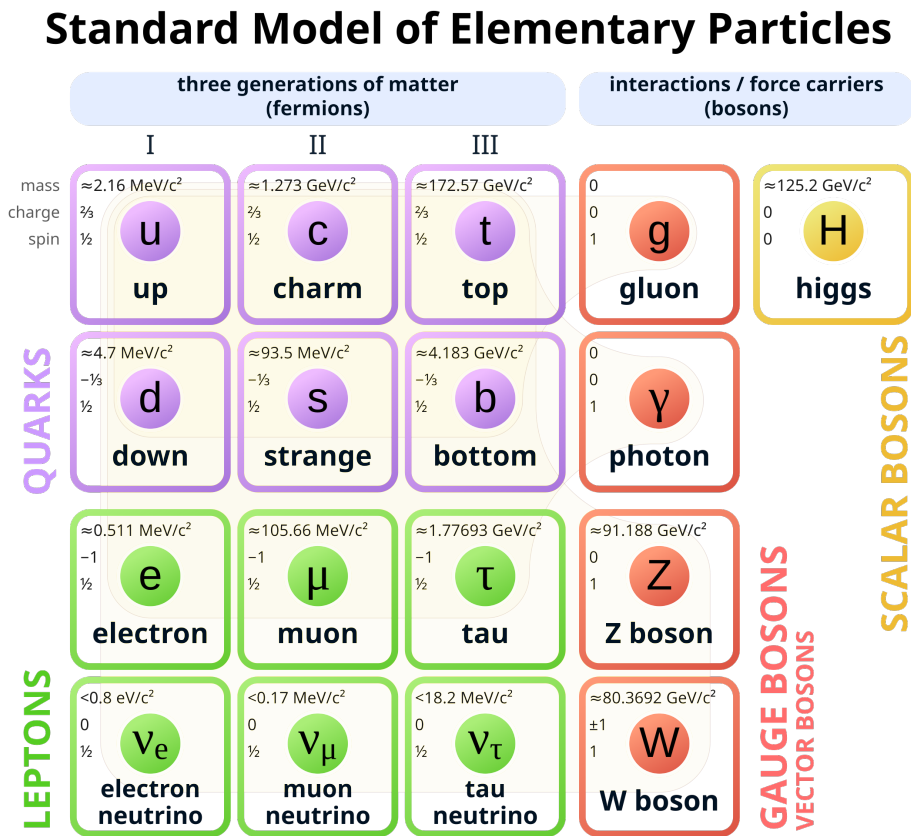


Figure 2.2: Standard Model of Elementary Particles. From the left-hand side to the right hand-side, the three generations of fermions can be seen. Quarks are coloured purple, leptons green. Further on the right-hand side, the gauge bosons are depicted in red and the scalar Higgs boson in yellow. Mass, electrical charge and spin of the particles are written in the upper left corner of each particle box. Taken from [19].

### 2.1.3 Fermions

Fermions are particles with half-integer spin that follow the Fermi-Dirac statistics, obeying the Pauli exclusion principle. They can either be fundamental particles, but also composite particles like baryons. Fundamental particles are further split into Quarks and Leptons.

## Quarks

All quarks are depicted together with their basic properties as purple squares in Fig. 2.2. The quarks in the upper row have an electric charge of  $\frac{2}{3}e$ . They are called up quark, charm quark, and top quark, with the up quark having the smallest and the top quark having the largest mass. The down quark, strange quark, and bottom quark in the lower row have an electric charge of  $-\frac{1}{3}e$ . Here, the down quark is the lightest and the bottom quark the heaviest, although all are significantly lighter than the top quark.

Quarks come in **six flavors**, grouped into **three generations**: the first comprises up and down, the second comprises charm and strange, and the third comprises top and bottom.

The quarks were not all discovered at once; each was identified through a combination of indirect signals and direct observation. The up and down quarks were first postulated by Gell-Mann and Zweig in the 1960s and later confirmed through deep inelastic scattering experiments [20]. The strange quark was identified in the 1950s to account for the properties of strange particles [21]. The charm quark was discovered in 1974 during the November revolution, when two groups independently observed a narrow  $J/\Psi$  resonance [22, 23]. The bottom quark followed in 1977 with the observation of the  $\Upsilon$  resonance [24]. The top quark was finally observed in 1995 by the CDF and DØ collaborations at Fermilab [25, 26].

For completeness, approximate quark masses are given in Table 2.2:

Table 2.1: Approximate quark masses, exact values are found in [27]

| Quark   | approximate Mass |
|---------|------------------|
| up      | 2 MeV            |
| down    | 5 MeV            |
| strange | 100 MeV          |
| charm   | 1.3 GeV          |
| bottom  | 4.2 GeV          |
| top     | 173 GeV          |

For each quark, there is a corresponding antiparticle (antiquark) with the same mass but opposite electric charge.

Additionally to their electric charge, quarks are subject to a property called colour charge. The name comes from colour theory, as there are three different colour charges (and their corresponding anti colours) which are commonly . Quarks form colour-neutral states with integer electric charges called **hadrons**. These can be classified as **baryons**, which consist of 3 quarks (like protons, composed of up, up, and down quarks, and neutrons, composed of up, down, and down quarks), or **mesons**, which consist of a quark and an antiquark.

Quarks do not exist as free particles. Before a single quark can be extracted from a colour-neutral state (leaving a free colour charge), a quark–antiquark pair is created from the vacuum, forming two colour-neutral particles. This phenomenon is known as **confinement**. As quarks are pulled apart, the strong force between them does not diminish; instead, it grows nearly linearly with their separation. This can be understood in terms of a colour flux tube or “string” forming between quarks. At a certain point, it becomes more energetic to create a quark–antiquark pair from the vacuum than to continue extending the flux tube, thereby preventing the isolation of a single quark.

The top quark is a special case because its weak decay occurs faster than the QCD confinement scale, allowing observation of its properties directly [28]. Its dominant decay mode is to a bottom quark and a  $W$  boson, governed by the structure of the Cabibbo–Kobayashi–Maskawa (CKM) matrix [29, 30]. The CKM matrix encodes the mixing between different generations of quarks in weak interactions and determines the relative strength of flavour-changing transitions. In particular, the large value of the matrix element  $V_{tb}$  explains the preference for  $t \rightarrow bW$  decays. More generally, the CKM matrix plays a central role in quark flavour physics and is essential to understanding phenomena such as CP violation within the Standard Model.

The CKM matrix is a unitary  $3 \times 3$  matrix that connects the weak eigenstates to the mass eigenstates of the quarks:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

Its entries are determined experimentally. A current approximate numerical form is [27]:

$$|V_{\text{CKM}}| \approx \begin{pmatrix} 0.974 & 0.225 & 0.0037 \\ 0.225 & 0.973 & 0.041 \\ 0.0084 & 0.040 & 0.999 \end{pmatrix}$$

This matrix reflects the strong suppression of transitions between non-neighbouring generations, such as from the top to the strange or down quark.

## Leptons

Leptons, like quarks, are fundamental fermions and appear in three generations. They are depicted in Fig. 2.2 as green boxes. The upper row shows the charged leptons – the electron, muon, and tau – each carrying an electric charge of  $-1e$ . The lower row shows the corresponding neutral leptons, known as neutrinos. Leptons are also grouped into three families, each consisting of a charged lepton and its associated neutrino. Unlike in the quark sector, lepton flavours are strongly correlated within each generation. As a result, the neutrino flavour eigenstates are named after their charged partners: electron neutrino, muon neutrino, and tau neutrino.

The charged leptons have well-defined and significantly different masses, spanning several orders of magnitude. Approximate mass values are shown below:

Table 2.2: Approximate lepton masses, exact values are found in [27]

| Lepton           | Mass      |
|------------------|-----------|
| Electron ( $e$ ) | 0.511 MeV |
| Muon ( $\mu$ )   | 105.7 MeV |
| Tau ( $\tau$ )   | 1.777 GeV |

The electron was discovered by J.J. Thomson in 1897 [31], followed by the muon in cosmic-ray experiments in 1936 [32], and the tau lepton in 1975 at SLAC [33]. Neutrinos were first postulated by Pauli [34] and later discovered by Cowan and Reines in 1956 through the detection of the electron

anti-neutrino [35]. The muon neutrino was discovered in 1962 at Brookhaven National Laboratory by observing interactions inconsistent with the electron neutrino [36]. The tau neutrino was confirmed much later, in 2000 by the DONUT experiment at Fermilab [37]. While these discoveries have established the existence of all three neutrino flavours, the **tau anti-neutrino** has not yet been directly observed, leaving it the only undiscovered elementary particle within the SM. The proposed **NaNu experiment** at CERN aims to provide the first direct observation of this elusive particle [38].

According to the original formulation of the Standard Model, neutrinos are massless. However, the observation of **neutrino oscillations** – the periodic change of neutrino flavour during propagation – provides direct evidence that at least two neutrino mass eigenstates must have non-zero masses [39, 40]. This discovery demonstrates that the Standard Model is incomplete in its original form.

Neutrino oscillations imply that the flavour eigenstates of neutrinos are quantum superpositions of mass eigenstates. This mixing is described by the **Pontecorvo–Maki–Nakagawa–Sakata (PMNS)** matrix, analogous to the CKM matrix in the quark sector. The PMNS matrix connects the flavour states ( $\nu_e, \nu_\mu, \nu_\tau$ ) to the mass eigenstates ( $\nu_1, \nu_2, \nu_3$ ):

$$U_{\text{PMNS}} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix}$$

A representative numerical approximation of the absolute values of the elements is [41]:

$$|U_{\text{PMNS}}| \approx \begin{pmatrix} 0.82 & 0.55 & 0.15 \\ 0.36 & 0.70 & 0.61 \\ 0.44 & 0.45 & 0.77 \end{pmatrix}$$

Unlike the CKM matrix, the PMNS matrix features large off-diagonal elements, reflecting the significant mixing between neutrino flavours.

The precise values of the neutrino masses and the structure of the PMNS matrix are still subjects of active research. A key effort to determine the absolute neutrino mass scale is the **KATRIN** experiment, which aims to measure the effective electron neutrino mass with sub-eV sensitivity [42].

### 2.1.4 Bosons

Bosons are particles with integer spin that obey Bose–Einstein statistics and do not follow the Pauli exclusion principle. This allows multiple bosons to occupy the same quantum state, giving rise to macroscopic quantum phenomena such as superconductivity (via Cooper pairs) and superfluidity (e.g. in  $^4\text{He}$ ).

Within the Standard Model, elementary bosons are divided into two categories: the spin-1 **gauge bosons** depicted as red boxes in Fig. 2.2, which mediate the fundamental forces, and the spin-0 **Higgs boson** shown as yellow box in Fig. 2.2, responsible for the mechanism of mass generation.

In addition to these elementary bosons, there also exist composite bosons such as mesons (bound quark–antiquark states) and quasi particles like Cooper pairs, which behave as bosons in condensed matter systems. Beyond the Standard Model, hypothetical bosons such as the **axion** have been proposed. These are typically **pseudo scalar** particles (see Section 3.1).

## Gluons

There are eight massless gluons that act as the mediators of the strong interaction. These arise as the gauge bosons of the  $SU(3)_C$  symmetry group in Quantum Chromodynamics (QCD), which has eight generators corresponding to eight distinct gluon states.

Gluons do not carry electric charge, but they do carry combinations of colour and anticolour charges. Unlike the photon in QED, gluons also couple to themselves, due to the non-Abelian nature of the  $SU(3)$  gauge group.

Although gluons are massless, the strong force they mediate is confined to short distances. This is due to the energy dependence of the strong coupling constant  $\alpha_s$ , a phenomenon known as the **running of the coupling**. At low energies (long distances), the interaction becomes strong, leading to confinement; at high energies (short distances), the coupling weakens – a key feature called **asymptotic freedom** [43, 44].

This running of the coupling constant  $\alpha_s$  is depicted in Fig. 2.3 where data points show experimental results from different experiments while the yellow band shows the QCD prediction.

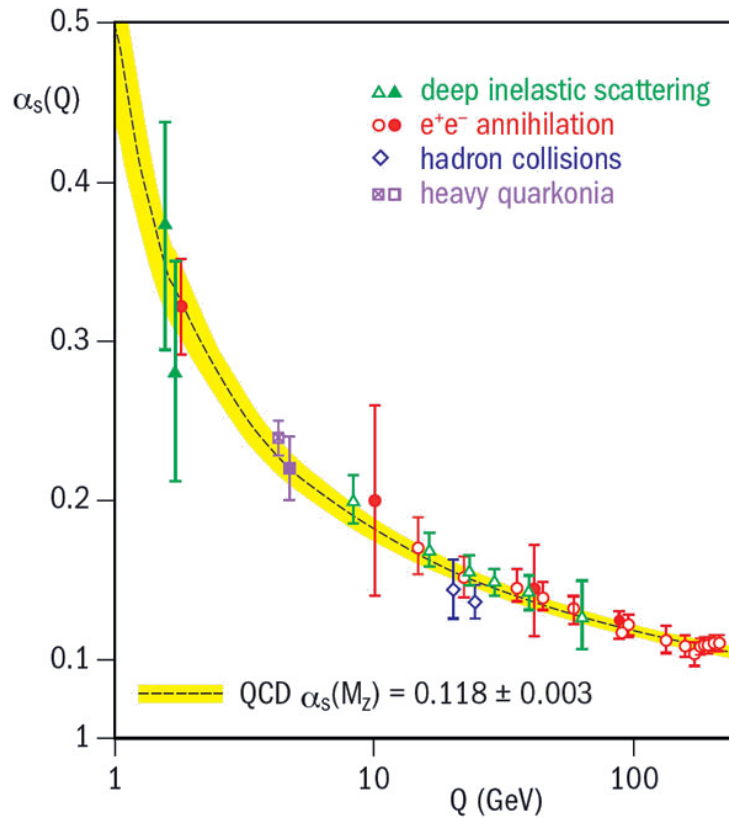


Figure 2.3: Running of the strong coupling constant  $\alpha_s(Q)$  as measured in various high-energy processes. The data points illustrate the experimentally observed decrease in  $\alpha_s$  with increasing energy scale  $Q$ , while the band shows the QCD prediction. Figure adapted from [45].

Experimental evidence for the existence of gluons was first obtained through the observation of

three-jet events in  $e^+e^-$  collisions at the PETRA collider at DESY [46].

### Photons

The photon is the mediator of the electromagnetic interaction, as described by Quantum Electrodynamics (QED). It is a massless, electrically neutral vector boson that couples to all particles with non-zero electric charge. Due to its zero mass, the photon mediates a long-range force with infinite range. It plays a central role in both classical electrodynamics and quantum field theory, and is the most well-understood of all gauge bosons. As an electrically neutral, massless gauge boson, the photon does not couple to itself – a consequence of the Abelian nature of the electromagnetic  $U(1)$  symmetry.

### W and Z Bosons

The W and Z bosons are the mediators of the weak interaction. They are massive, spin-1 vector bosons that originate from the gauge fields of the electroweak  $SU(2)_L \times U(1)_Y$  symmetry, acquiring their mass through the Higgs mechanism.

The charged W bosons ( $W^+$  and  $W^-$ ) carry electric charges of +1 and −1, respectively, and have a mass of approximately 80.4 GeV [27]. They mediate charged-current weak interactions, such as beta decay, where both flavour and electric charge change:

$$n \rightarrow p + e^- + \bar{\nu}_e.$$

These bosons couple only to left-handed fermions and right-handed antifermions, reflecting the maximal parity violation of the weak interaction. W bosons play a central role in quark flavour-changing processes governed by the CKM matrix, and are also responsible for lepton flavour transitions.

The neutral Z boson ( $Z^0$ ) mediates neutral-current weak interactions, in which the flavour of the interacting particles remains unchanged. It is electrically neutral and has a mass of approximately 91.2 GeV [27]. The Z boson couples to both left- and right-handed fermions, although with different coupling strengths. Examples of Z-mediated processes include neutrino–electron scattering and invisible decays to neutrino pairs.

At low energies, both the W and Z bosons appear as virtual particles in weak decays. At high energies, such as those accessible in particle colliders, they can be produced on-shell and observed as real particles. Their discovery in 1983 by the UA1 and UA2 experiments at CERN [9–11] marked a major confirmation of the electroweak theory.

Both bosons decay into fermion–antifermion pairs. The W boson decays to lepton–neutrino or quark–antiquark pairs, while the Z boson can decay into charged lepton pairs, neutrinos, or hadrons. Their decay widths and branching ratios are in good agreement with Standard Model predictions, and the precise measurement of these properties has been essential in testing the electroweak sector. A good summary of precision measurements with the ATLAS detector can be found in [47].

All vector bosons are summarized in Table 2.3.

## 2.2 Higgs mechanism and Higgs boson

In a gauge-invariant quantum field theory, explicit mass terms for gauge bosons are not allowed, as they violate the symmetry. This poses a problem in the electroweak sector of the Standard Model,



Table 2.3: Properties of the Standard Model vector bosons. Mass values are approximate and taken from [27].

| Boson    | Symbol   | Mass [GeV] | Force           | Charge  | Notes                                           |
|----------|----------|------------|-----------------|---------|-------------------------------------------------|
| Photon   | $\gamma$ | 0          | Electromagnetic | 0       | Massless, mediates infinite-range force         |
| W bosons | $W^\pm$  | 80.4       | Weak            | $\pm 1$ | Charged, mediates flavour-changing interactions |
| Z boson  | $Z^0$    | 91.2       | Weak            | 0       | Neutral current weak interactions               |
| Gluons   | $g$      | 0          | Strong          | Colour  | Eight massless bosons, self-interacting         |

where the W and Z bosons are experimentally known to be massive. To reconcile this, the Standard Model invokes the Higgs mechanism, a process of spontaneous symmetry breaking that preserves gauge invariance while generating masses dynamically.

ellis2013higgsphysics

The Higgs mechanism introduces a scalar field, the Higgs field, which transforms as a complex doublet under the  $SU(2)_L$  gauge group. The potential energy of the Higgs field is given by:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (2.1)$$

where  $\mu^2 < 0$  and  $\lambda > 0$ . The negative mass term leads to a non-zero vacuum expectation value (VEV) of the field:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad \text{with } v \approx 246 \text{ GeV}. \quad (2.2)$$

This VEV spontaneously breaks the electroweak symmetry:

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{em}}, \quad (2.3)$$

leaving the photon massless while giving mass to the W and Z bosons.

The resulting mass terms for the electroweak gauge bosons are:

$$m_W = \frac{1}{2} g v, \quad m_Z = \frac{1}{2} \sqrt{g^2 + g'^2} v, \quad (2.4)$$

where  $g$  and  $g'$  are the gauge couplings associated with  $SU(2)_L$  and  $U(1)_Y$ , respectively.

Fermions also acquire mass through Yukawa couplings to the Higgs field:

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L \phi \psi_R + \text{h.c.}, \quad (2.5)$$

where  $y_f$  is the Yukawa coupling for a fermion  $f$ , and  $\psi_{L,R}$  denote its left- and right-handed components. After symmetry breaking, this results in a mass term  $m_f = y_f v / \sqrt{2}$ , implying that heavier fermions couple more strongly to the Higgs field.

One physical scalar degree of freedom remains after the symmetry breaking: the Higgs boson. It is

a spin-0, CP-even neutral particle. Its mass is given by:

$$m_H = \sqrt{2\lambda} v, \quad (2.6)$$

but since  $\lambda$  is not predicted by the theory, the Higgs mass must be determined experimentally.

The Higgs boson was discovered in 2012 by the ATLAS and CMS experiments at the LHC [2, 3], with a measured mass of approximately 125 GeV. Its properties – including spin, parity, and couplings to other particles – have since been measured and found to be consistent with Standard Model predictions. Leading order Feynman diagrams for four main production modes which have been discovered at the LHC, namely gluon-gluon-Fusion (ggF), Vector-Boson-Fusion (VBF), Higgs-Strahlung (VH) and the associated production of a Higgs boson with a top quark top anti-quark pair (ttH) are depicted in Fig. 2.4. Each production cross section is written beneath the diagram.

The discovery of the Higgs boson completed the particle content of the Standard Model and provided crucial confirmation of the electroweak theory.

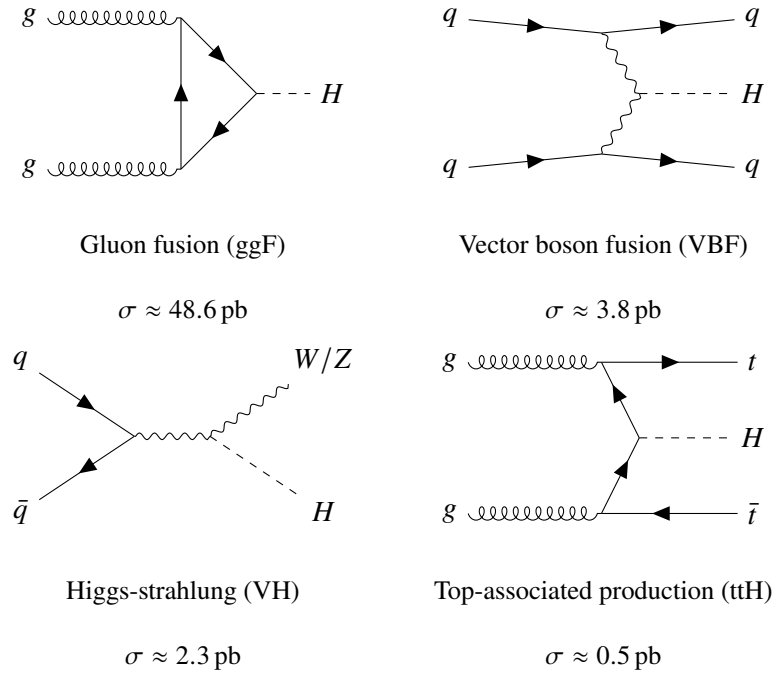


Figure 2.4: Leading Higgs boson production mechanisms at the LHC with approximate cross sections at  $\sqrt{s} = 13 \text{ TeV}$ . Values taken from [48]

Since the Higgs boson couples to all massive particles, it can, in principle, decay directly into any pair of such particles, provided the decay is kinematically allowed. For example, decays to top quarks are not possible due to the Higgs boson's mass being lower than twice the top mass. Thus, the top quark Yukawa coupling must be inferred indirectly, such as through loop-induced processes or associated production channels.

The coupling strength is proportional to the particle's mass, making decays to heavier particles more likely. This mass dependence of the couplings is illustrated in Fig. 2.5, which shows the reduced coupling strength modifiers for various particles.

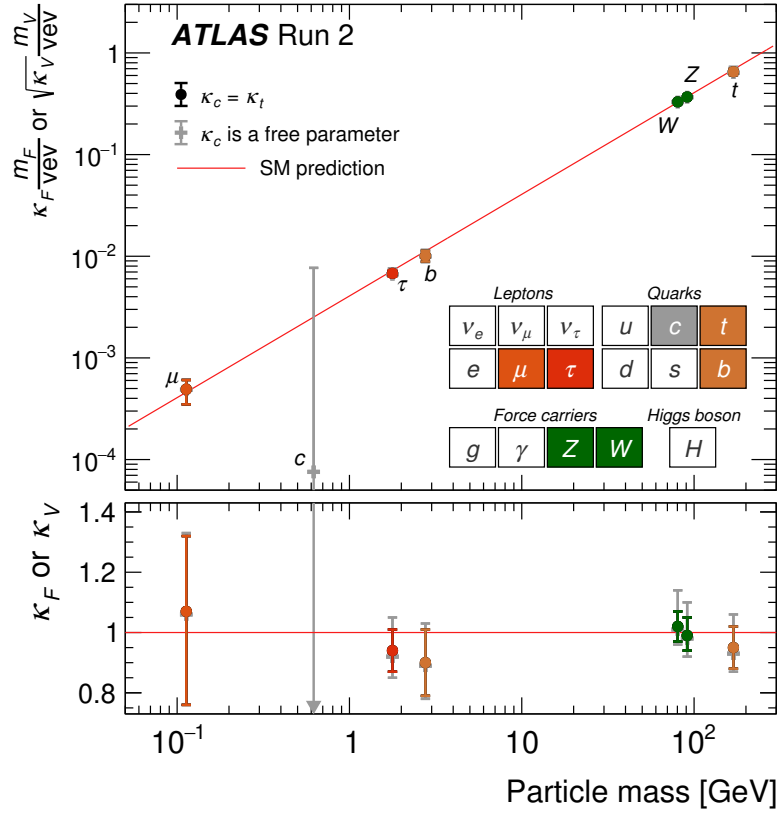
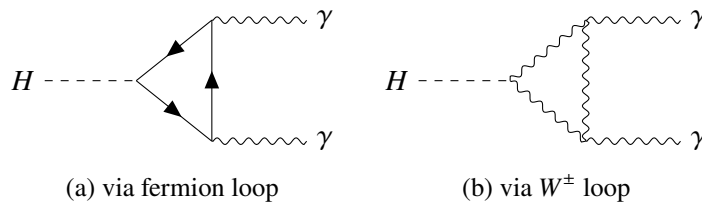


Figure 2.5: Reduced coupling strength modifiers of the Higgs boson. Taken from [49].

In addition to tree-level decays, the Higgs boson can decay into photons via loop diagrams involving virtual heavy charged particles. This loop-induced decay mode,  $H \rightarrow \gamma\gamma$ , played a central role in the Higgs boson discovery due to its clean experimental signature.

There are two leading-order Feynman diagrams that contribute to this decay, featuring a top quark loop and a  $W^\pm$  boson loop. These contributions interfere destructively and are shown in Fig. 2.6.


 Figure 2.6: Feynman diagrams for the loop-induced decay  $H \rightarrow \gamma\gamma$ . The Higgs couples to a virtual fermion (left) or  $W^\pm$  boson (right) loop, with photons radiated from the loop.

### 2.2.1 Proton-Proton Collisions

Protons are not elementary particles but are instead composite objects made up of quarks and gluons, collectively referred to as partons. In high-energy proton–proton collisions such as those at the LHC, the fundamental interactions occur not between entire protons but between their constituent partons. The interaction of interest—often termed the **hard scatter**—involves the collision of two partons with large momentum transfer, which may produce heavy particles or other phenomena accessible at TeV energy scales. However, alongside this primary interaction, additional softer processes occur between the remnants of the colliding protons. These so-called underlying event activities contribute to the final-state particle content and must be accounted for in both theoretical predictions and experimental analyses.

The probabilistic nature of finding a parton of type  $i$  (e.g. a quark or gluon) with momentum fraction  $x_i$  inside the proton is described by the parton distribution functions (PDFs), denoted by  $f_i(x_i, Q^2)$ , where  $Q$  is the characteristic energy scale of the interaction. Since partons carry only a fraction of the proton’s total momentum and their momentum fractions are not fixed, theoretical predictions for hadronic cross sections require integration over all possible parton momentum fractions. The differential cross section for a generic hadronic process is given by:

$$\sigma = \sum_{i,j} \int_0^1 dx_1 \int_0^1 dx_2 f_i(x_1, Q^2) f_j(x_2, Q^2) \hat{\sigma}_{ij}(x_1 x_2 s, Q^2),$$

where:

- $x_1$  and  $x_2$  are the momentum fractions of the incoming partons,
- $f_i(x_1, Q^2)$  and  $f_j(x_2, Q^2)$  are the PDFs for partons  $i$  and  $j$ ,
- $\hat{\sigma}_{ij}$  is the parton-level (hard-scatter) cross section for the subprocess  $ij \rightarrow X$ ,
- $s$  is the squared centre-of-mass energy of the proton–proton system.

This formula illustrates the convolution of the PDFs with the hard-scatter matrix elements. The predictive power of perturbative QCD lies in the fact that the partonic cross section  $\hat{\sigma}_{ij}$  can be calculated using Feynman diagrams, while the PDFs are extracted from experimental data and evolved according to the DGLAP equations [50–52].

Fig. 2.7 shows PDFs for  $Q^2 = 10 \text{ GeV}^2$  (left hand-side) and  $Q^2 = 10 \times 10^4 \text{ GeV}^2$ . It can be seen, that gluon interactions dominate at LHC energies.

Following the hard scatter and initial-state radiation, the outgoing partons are not observed as free particles due to colour confinement. Instead, they undergo a process known as **hadronisation**, in which they transform into colour-neutral hadrons. This transition is governed by non-perturbative QCD dynamics and is modelled using phenomenological approaches implemented in Monte Carlo event generators. The resulting hadrons then propagate through the detector, where they may decay further or interact with detector materials, forming the observable signatures used in experimental analyses.

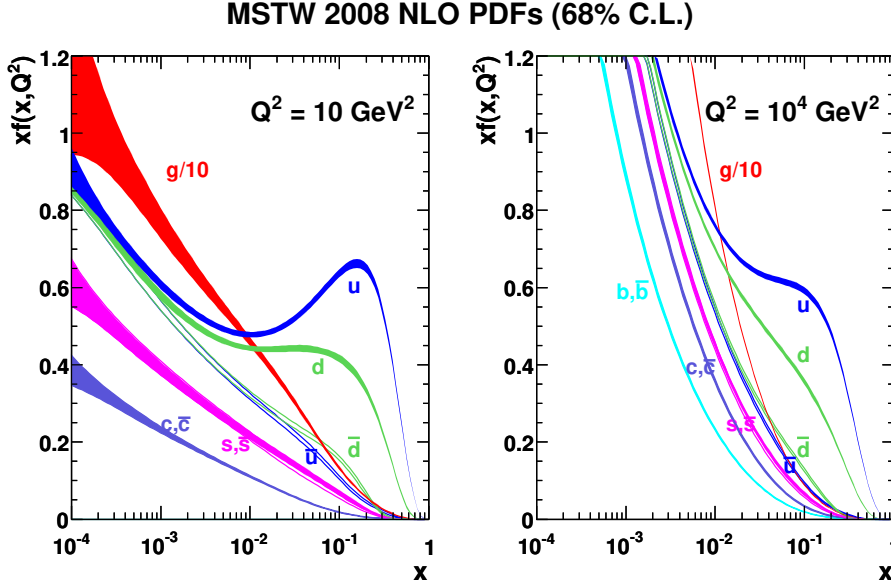


Figure 2.7: Parton distribution functions for different energies. The gluon contribution is scaled down by  $\frac{1}{10}$ . Taken from [53].

## 2.3 Limitations of the Standard Model

Despite its remarkable success in describing a wide range of phenomena, the Standard Model (SM) of particle physics is widely regarded as incomplete. While it provides a consistent quantum field theory framework for the electromagnetic, weak, and strong interactions, several key observations and theoretical issues remain unexplained within the SM.

### 2.3.1 The Strong CP Problem

One such unresolved issue is the so-called **strong CP problem**. Quantum Chromodynamics (QCD), the theory of the strong interaction, allows for a CP-violating term in the Lagrangian of the form

$$\mathcal{L}_\theta = \theta \frac{g_s^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu}, \quad (2.7)$$

where  $\theta$  is a dimensionless parameter,  $G_{\mu\nu}^a$  is the gluon field strength tensor, and  $\tilde{G}^{a\mu\nu}$  its dual. Such a term would induce a non-zero electric dipole moment (EDM) of the neutron, in contradiction with experimental bounds. Current measurements constrain the neutron EDM to be less than  $1.8 \times 10^{-26} e \text{ cm}$  [54], implying that  $|\theta| < 10^{-10}$ , a fine-tuning that is theoretically unnatural. A dynamical solution to this problem is offered by the Peccei–Quinn mechanism, which predicts the existence of a new pseudoscalar particle: the axion. This will be discussed in the following chapter (see Section 3.1).

### 2.3.2 Dark Matter

Another major shortcoming of the Standard Model is its inability to account for the dark matter that constitutes approximately 27% of the energy density of the Universe [55]. The evidence for dark matter comes from a wide range of astrophysical and cosmological observations, including galaxy rotation curves, gravitational lensing, and the cosmic microwave background. None of the known SM particles have the required properties—neutral, stable, and non-relativistic—to serve as a viable dark matter candidate. This has led to the proposal of various extensions involving new weakly interacting particles, including Weakly Interacting Massive Particles (WIMPs), sterile neutrinos, and axion-like particles (ALPs). ALPs, in particular, arise naturally in many extensions of the SM and will be explored in detail in the next chapter.

### 2.3.3 The Muon $(g - 2)_\mu$ Discrepancy

The magnetic moment of the muon is predicted to be exactly  $g_\mu = 2$  at tree level. However, higher-order loop corrections involving virtual particles lead to a small deviation from this value. This deviation is quantified by the so-called **anomalous magnetic moment** of the muon, defined as

$$a_\mu = \frac{g_\mu - 2}{2}. \quad (2.8)$$

Within the Standard Model,  $a_\mu$  can be computed with exceptional precision, incorporating contributions from QED, electroweak, and hadronic effects.

If measured with comparable precision,  $a_\mu$  provides a stringent test of the Standard Model's completeness and accuracy. Such high-precision measurements have been carried out at Brookhaven National Laboratory (BNL) and, more recently, at Fermilab, making  $a_\mu$  one of the most precisely determined observables in particle physics [56–59].

Intriguingly, the experimental results deviate from the SM prediction by several standard deviations [60–62]. This discrepancy is currently the largest among precision electroweak observables and has persisted for over two decades.

Recent advances in theoretical methods—most notably lattice QCD calculations—suggest a smaller hadronic vacuum polarisation contribution, which would reduce the SM prediction of  $a_\mu$  and thus the significance of the discrepancy [63]. Whether this tension points to new physics or a refined understanding of hadronic effects remains an open question.

# Axions and Axion-Like Particles

## 3.1 Theoretical Motivation

Axions and Axion-Like Particles (ALPs) are well-motivated extensions of the Standard Model (SM), initially proposed to resolve the strong CP problem in QCD but now also explored as candidates for dark matter, dark radiation, and possible explanations of experimental anomalies such as the muon  $g - 2$  discrepancy (cf. Section 2.3.3).

### 3.1.1 The Strong CP Problem and QCD Axions

The strong CP problem arises from a possible CP-violating term in the QCD Lagrangian,

$$\mathcal{L}_\theta = \theta \frac{g_s^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu}, \quad (3.1)$$

which would induce a neutron electric dipole moment (EDM). Experimental bounds on the EDM require  $|\theta| < 10^{-10}$  [54], implying extreme fine-tuning. A full discussion is provided in Section 2.3.

The Peccei–Quinn mechanism [64] solves this problem dynamically by promoting  $\theta$  to a field that relaxes to zero. The associated pseudo-Goldstone boson is the axion [65, 66].

### Invisible Axion Models

Early axion models were ruled out by laboratory constraints, but high-scale “invisible” axion models remain viable. Two leading frameworks are:

- **KSVZ model** [67, 68], which introduces new heavy vector-like quarks coupling to a complex scalar that breaks the PQ symmetry. The axion couples to gluons through the anomaly but has no tree-level couplings to SM fermions or photons.
- **DFSZ model** [69, 70], which extends the Higgs sector with an additional Higgs doublet. The axion couples at tree level to both fermions and photons.

In both models, the axion mass and its couplings scale inversely with the symmetry-breaking scale  $f_a$ , typically constrained to  $f_a \gtrsim 10^9$  GeV.

### 3.1.2 Axion-Like Particles (ALPs)

Many theories, especially string theory compactifications [71, 72], predict the existence of pseudo-scalar particles similar to the axion but without a fixed mass–coupling relationship. These are referred to as Axion-Like Particles (ALPs).

The low-energy effective Lagrangian for ALPs typically includes couplings to photons and fermions:

$$\mathcal{L}_{\text{ALP}} \supset \frac{1}{4} g_{a\gamma\gamma} a F_{\mu\nu} \tilde{F}^{\mu\nu} + i g_{aff} a \bar{f} \gamma_5 f + \dots \quad (3.2)$$

where  $a$  is the ALP field, and  $g_{a\gamma\gamma}, g_{aff}$  are model-dependent coupling constants.

This generality allows ALPs to appear in a variety of contexts:

- As dark matter candidates,
- As mediators in fifth-force or long-range interactions,
- As contributors to early-universe dark radiation or the Hubble tension,
- As explanations for anomalies such as the muon  $g - 2$  (see Section 3.1.3).

ALPs can couple to photons, gluons, SM fermions, or electroweak bosons, depending on the model.

### 3.1.3 ALPs and the Muon Anomalous Magnetic Moment

The long-standing discrepancy between the measured and SM-predicted value of the muon anomalous magnetic moment [57, 58] (discussed in Section 2.3.3) may be explained by the existence of light new particles. Among these, ALPs coupling to muons and photons are particularly well-motivated.

The relevant one-loop diagrams are shown in Fig. 3.1. While the two contributions interfere destructively, the diagram involving the ALP–photon coupling typically dominates if  $g_{a\gamma\gamma}$  is large enough [73, 74].

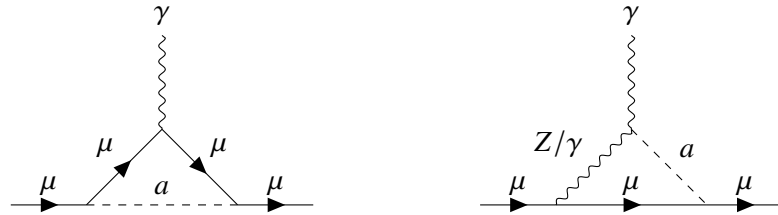


Figure 3.1: ALP contribution to the anomalous magnetic moment of the muon. Reproduced from [75].

This scenario remains viable in parts of parameter space not yet excluded by astrophysical or collider constraints.

### 3.1.4 Solar ALPs and the Primakoff Effect

Axions and ALPs can be produced in stellar environments. In particular, **solar axions** may be generated in the Sun via the **Primakoff effect**, where thermal photons convert into axions in the Coulomb field of charged particles [76]. The interaction is described by:



$$\mathcal{L}_{a\gamma\gamma} = -\frac{1}{4}g_{a\gamma\gamma} a F_{\mu\nu} \tilde{F}^{\mu\nu} = g_{a\gamma\gamma} a \mathbf{E} \cdot \mathbf{B}. \quad (3.3)$$

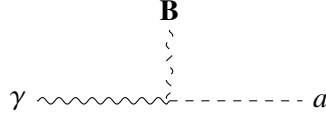


Figure 3.2: Primakoff effect: photon–axion conversion in a background magnetic field  $\mathbf{B}$ .

This process enables both solar axion detection (e.g. in helioscope experiments) and laboratory-based searches using strong magnetic fields.

## 3.2 Searches for axions and ALPs

In this chapter, a brief overview of past and current searches for axions and axion-like particles (ALPs) is given. Depending on the type of ALP and its properties, different search strategies can be employed.

### 3.2.1 Haloscope Experiments

Haloscopes are designed to search for ALPs (and axions) making up the local dark matter halo of the Milky Way. In these experiments, a high-Q microwave cavity is placed in a strong magnetic field, enabling ambient ALP dark matter to convert into detectable photons with energies corresponding to the ALP mass. These photons resonate with the cavity mode, allowing sensitive detection.

The most notable haloscope is ADMX (Axion Dark Matter eXperiment) [77], which has reached sensitivity to QCD axions in the  $\mu\text{eV}$  mass range. Other experiments like HAYSTAC [78], ORGAN McAllister:2017lkb, and MADMAX [79] are also exploring broader mass ranges and improving sensitivity. In, Mainz, the SUPAX experiment [80] contributes to this search. Haloscope searches are strongly model-dependent in that they assume ALPs constitute a significant fraction of the local dark matter density. Moreover, their sensitivity is largely confined to ultra-light ALPs (typically in the  $\mu\text{eV}$  to MeV range), far below the GeV-scale mass window explored in collider experiments.

The latest results of ADMX can be seen in Fig. 3.3. It shows 90 % C.L. upper limits on  $g_{a\gamma\gamma}$  as a function of the cavities frequency and the corresponding ALP mass. Yellow, blue and grey bands show the exclusion regions of older ADMX analyses. Dashed lines indicate the region, where invisible axion signals would be expected. A large area from 2.6  $\mu\text{eV}$  up to 4.2  $\mu\text{eV}$  can be excluded.

### 3.2.2 Helioscope Experiments

Helioscope experiments aim to detect ALPs that are produced in the Sun via the Primakoff process—where photons in the solar plasma convert into ALPs in the presence of electric fields of charged particles. These ALPs escape the solar interior and can be reconverted into photons in a strong laboratory magnetic field aligned with the incoming solar ALP flux.

The most prominent helioscope to date is the CERN Axion Solar Telescope (CAST) [82], and its successor, the International Axion Observatory (IAXO) [83], is under development. These experiments are sensitive to ALPs with masses up to a few eV and are particularly powerful at probing photon

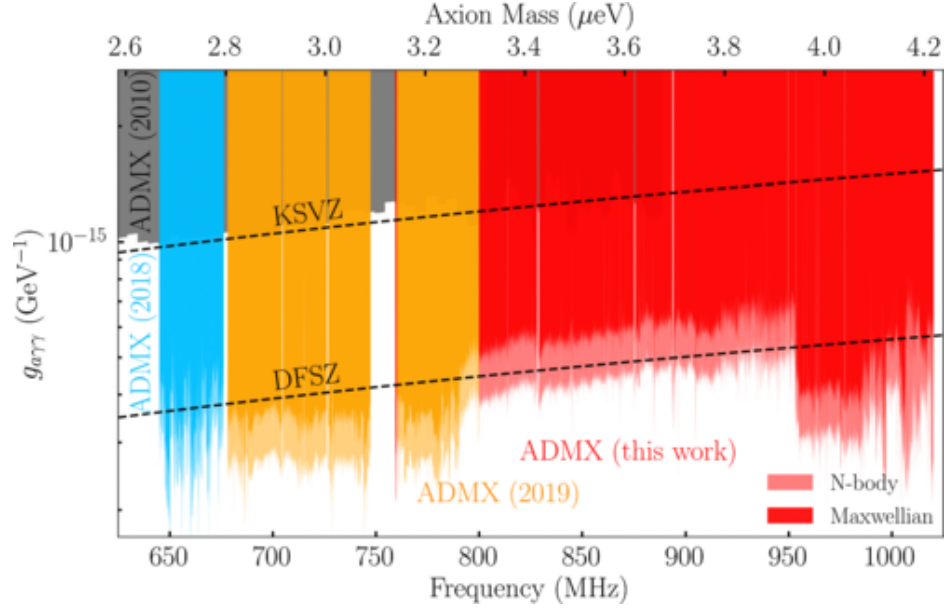


Figure 3.3: 90 % C.L. exclusion limits from latest ADMX searches for invisible axions. Taken from [81]

couplings in the range  $g_{a\gamma\gamma} \sim 10^{-11}$  to  $10^{-10} \text{ GeV}^{-1}$ . Helioscopes are relatively model-independent for ALPs that couple to photons but are limited to probing solar production mechanisms and therefore require assumptions about ALP interactions in stellar plasmas.

Fig. 3.4 shows the expected sensitivity of IAXO and babyIAXO together with regions of phase space in the  $m_a$ - $C_{a\gamma\gamma}$  plane that have already been excluded by other experiments such as CAST or ADMX.

### 3.2.3 Light-Through-Wall Experiments

Light-Through-Wall (LTW) experiments are designed to detect extremely weakly interacting particles such as ALPs by exploiting photon-ALP conversions in external magnetic fields. A laser beam is directed into a magnetic field where photons may convert into ALPs. An opaque wall blocks the photons, but ALPs, if produced, pass through and may reconvert into photons in a second magnetic region on the other side. The detection of these regenerated photons would indicate the presence of ALPs.

Major LTW experiments include ALPS (Any Light Particle Search) at DESY [85], OSQAR at CERN [86], and future upgrades like ALPS II [87]. These experiments are largely model-independent within the class of ALPs that couple to photons, as the signal relies only on this coupling and not on other interactions. However, they are most sensitive to very low-mass ALPs (typically  $m_a \lesssim \text{meV}$ ) due to coherence requirements over the magnetic field region. As such, they complement but do not overlap with the mass range probed in this thesis.

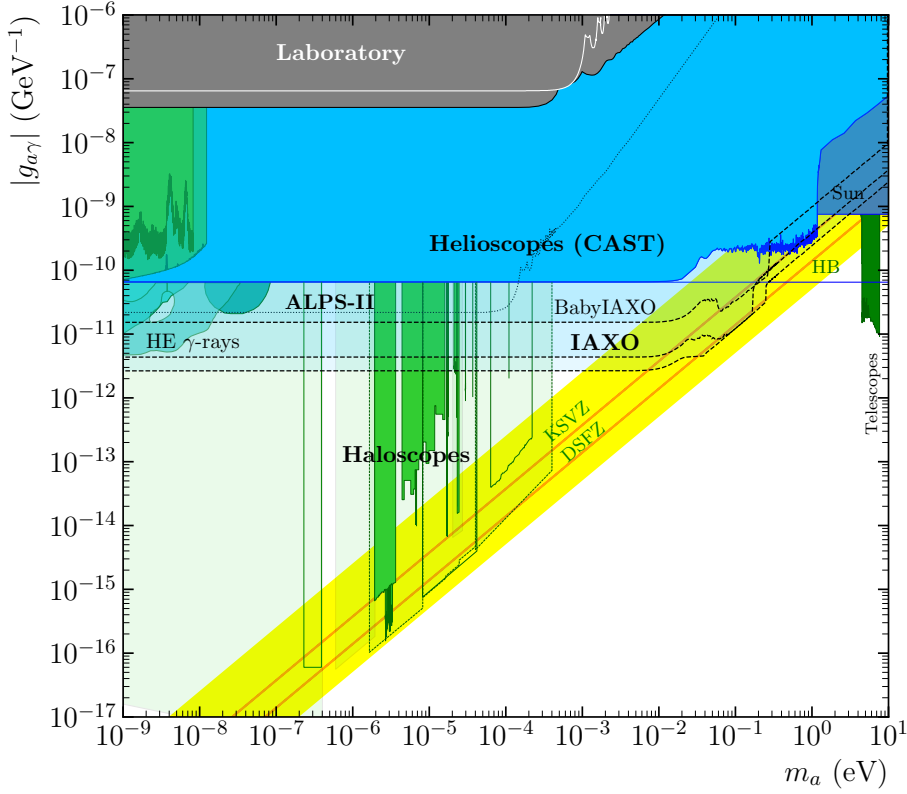


Figure 3.4: Exclusion limits of CAST together with predicted sensitivity of IAXO and babyIAXO. Taken from [84]

### 3.3 Collider searches for ALPs

Another possibility to search for ALPs is at particle colliders, where they can be probed through their effective couplings to Standard Model particles. One approach to search for long-lived ALPs at colliders involves mono-jet signatures in combination with missing transverse energy ( $E_T^{\text{miss}}$ ). However, this is not a search specific to ALPs, as similar signatures can also arise from other invisible or long-lived particles [89, 90]

Another strategy relies on possible visible decay channels of ALPs, where one can search for resonances in the corresponding invariant mass spectra. The most important ALP decay channels are shown as Feynman diagrams in Fig. 3.6. For example, the decay  $a \rightarrow Za$  is only kinematically allowed for ALPs with sufficient mass. In this case, the  $Z$ -boson would subsequently decay into fermion pairs, providing additional handles for event reconstruction.

The decay into gluons ( $a \rightarrow gg$ ) is experimentally more challenging at hadron colliders due to the large background from QCD jet production.

Another option that came up in the last years with the possibility to observe light-by-light scattering in heavy ion collisions [91–93]. The final state with only a pair of photons could as well be produced by an ALP decaying to photons. The corresponding Feynman diagrams can be seen in Fig. 3.7. The process involving an ALP will peak around the ALP-mass  $m_a$ , while the SM light-by-light scattering

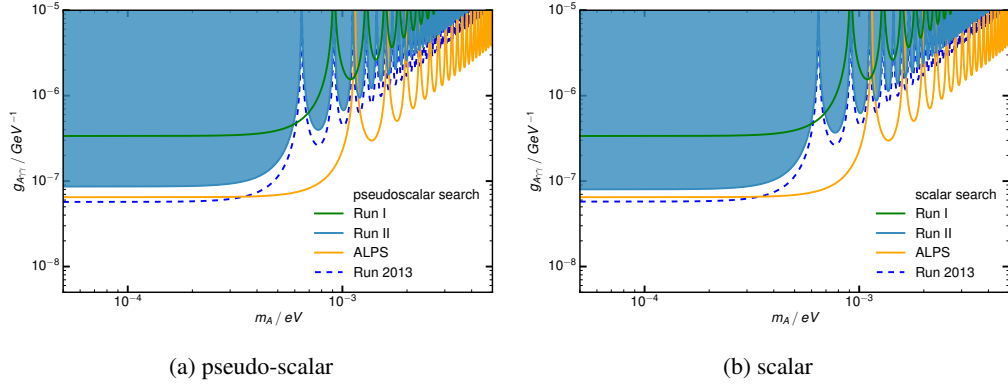


Figure 3.5: Exclusion limits for ALPS and OSQAR experiments for the search for pseudo-scalar ALPS (a) and scalar ALPS (b). Taken from [88].

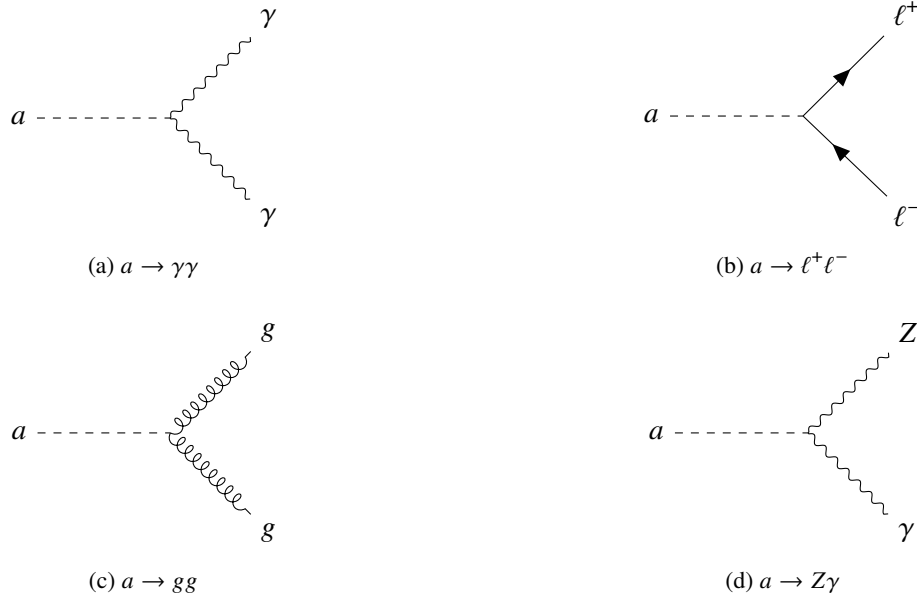


Figure 3.6: Important decay channels of ALPs depending on their mass and couplings.

will produce a smooth shape in the invariant mass spectrum. The search for ALPs in light-by-light scattering is described in [94].

### 3.3.1 ALPs from exotic Higgs boson decays

ALPs may not only be produced directly in the hard scattering, but also appear in decays of SM particles. Besides the  $Z^0$  boson which was measured extensively at LEP [95, 96], the Higgs boson is a promising candidate to search for ALPs in its decay products.

Different exotic Higgs boson decays involving ALPs have been proposed in [75]. The two most prominent ones are  $H \rightarrow Za$  and  $H \rightarrow aa$ . In principle, the ALPs can subsequently decay to many particles, e.g. Fig. 3.6 (a)-(c). Some examples can be found in [97–99]. However, this work will focus

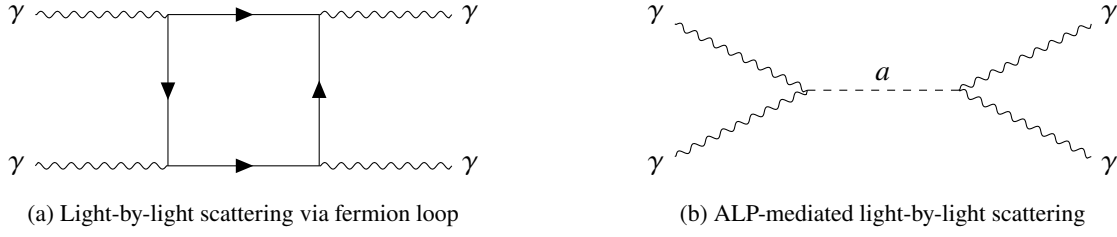


Figure 3.7: Comparison of standard and ALP-mediated light-by-light scattering diagrams

on the decay to a pair of photons  $a \rightarrow \gamma\gamma$ .

Searches for  $H \rightarrow Za$  have been performed in [100, 101]. How the latter search for promptly decaying ALPs can be extended allowing for ALPs being long-lived is described in [102].

The expected sensitivity of an  $H \rightarrow aa$  search in dependence of the ALP mass  $m_a$  and the ALP coupling to photons  $C_{a\gamma\gamma}$  can be seen in Fig. 3.8 as light green shaded area with blue borders. Experimentally excluded regions are shown as filled areas without surroundings. As a red shaded area with red surrounding, the parameter space in which ALPs could explain the  $(g-2)_\mu$  discrepancy is depicted. It can be seen, that most of the remaining  $(g-2)_\mu$  parameter space could be covered by a  $H \rightarrow aa$  search.

Searches for  $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$  have been performed by ATLAS and CMS [103, 104]. However, these searches have been performed allowing only promptly decaying ALPs and thus not sensitive to the whole shaded green area, but only for the upper part with only high coupling. The analysis presented in Chapter 6 is the first one also taking into account ALPs with larger lifetimes.

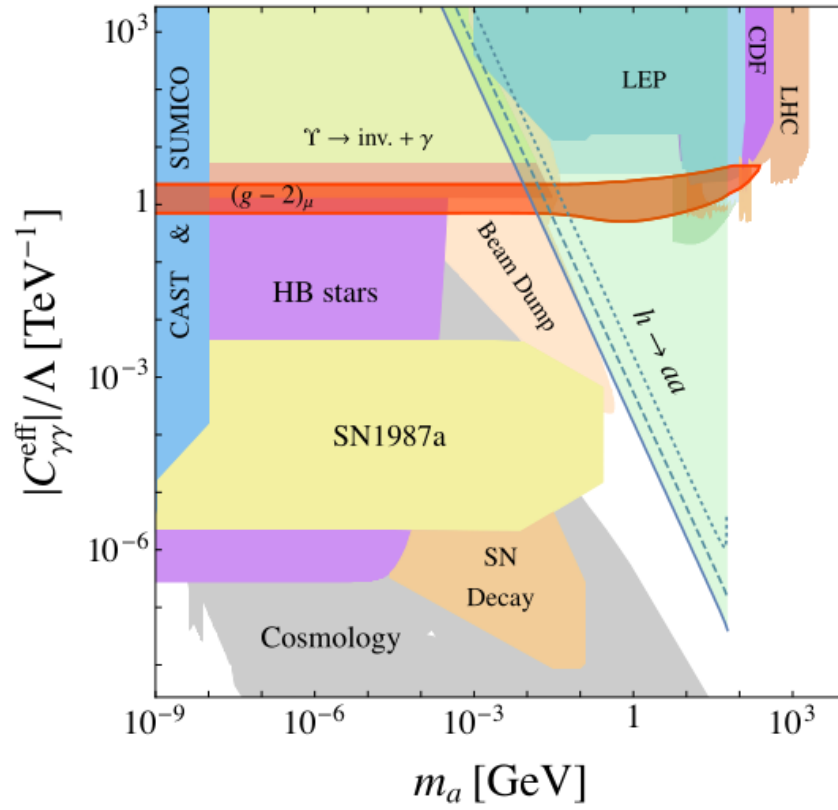


Figure 3.8: The ALP parameter space in dependence of  $m_a$  and  $C_{a\gamma\gamma}$  with experimentally excluded regions as filled areas. The shaded red area shows the area in which ALPs could explain the  $(g-2)_\mu$  discrepancy and the shaded green area that can be assessed by  $H \rightarrow aa$ . Taken from [75].

## Experimental Environment

The analysis presented in this work uses data recorded by the ATLAS detector at the Large Hadron Collider (LHC) at CERN in Switzerland. In this chapter, both the LHC and the ATLAS detector are briefly introduced. Later on, a method is developed to optimize analysis results using calorimeter information with machine learning techniques. Hence the electromagnetic Calorimeter of the ATLAS detector will be described in more detail in Section 4.2.1.

### 4.1 The Large Hadron Collider

The Large Hadron Collider is the largest and most powerful particle accelerator to date. It is based in Geneva Switzerland and France at European Organization for Nuclear Research (CERN). In the following, a brief summary of its function and design will be given using information from [105, 106] and other sources.

The LHC is a synchrotron with a circumference of 27km designed to accelerate protons and heavier ions like lead ions. However, it is not a stand-alone machine, but only the last link in CERN's accelerator chain and hadrons pass a series of other systems before being injected in the main ring. At first,  $H^-$  ions (negatively charged Hydrogen atoms with an additional electron) are accelerated to an energy of 160 MeV at a linear accelerator, the LINAC4. During the injection from LINAC4 to the Proton Synchrotron Booster (PSB), the two electrons are stripped off and only protons remain. These are further accelerated in the PSB up to 2 GeV from where they go through the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) where they reach energies of 26 GeV and 450 GeV. This is enough for the injection in the two LHC beam pipes. There the protons circulate in bunches in opposite directions being further accelerated up to energies of 6.5 TeV per beam. The beams are brought to collision in four of the interaction points where the big LHC experiments are placed.

The four main LHC experiments are ALICE [107], LHCb [108] and the two multi-purpose detectors CMS [109] and ATLAS, described in Section 4.2.

The amount of data, the experiments can record is limited by the number of collisions per time interval. This is the instantaneous luminosity. It can be calculated as

$$\mathcal{L} = f \frac{N_b \cdot N_i \cdot N_j}{4\pi\sigma_x\sigma_y} \quad (4.1)$$

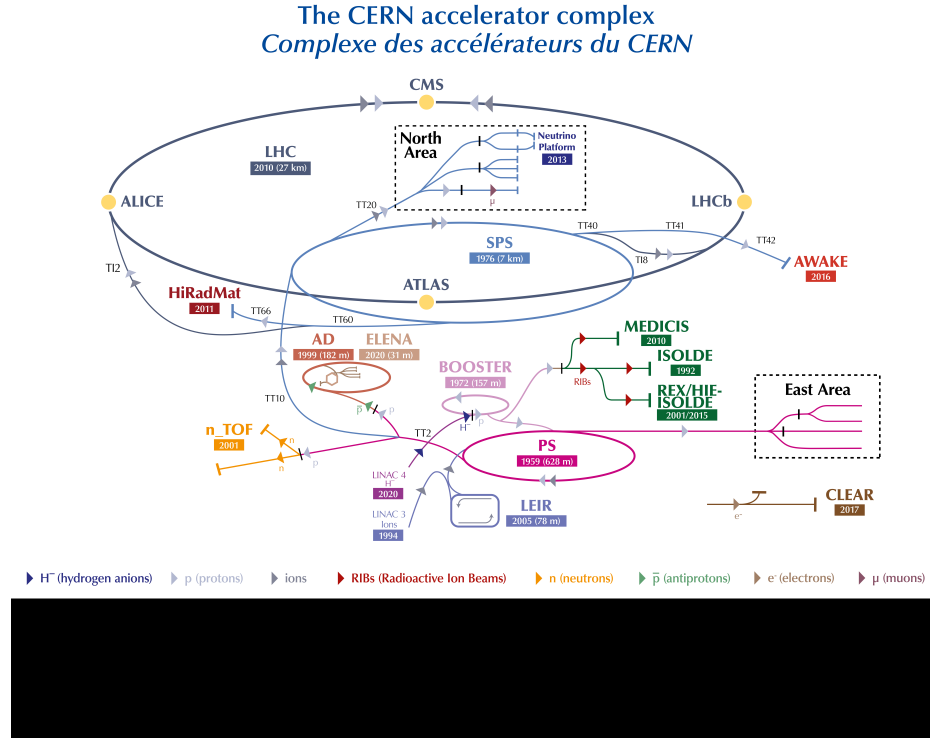


Figure 4.1: The LHC accelerator complex [110]

with the revolution frequency  $f$ , the Number of bunches  $N_b$  as well as  $N_i$ ,  $N_j$  the Numbers of protons in the bunches  $i, j$ . In the denominator, the beam profile area with the geometrical spread of a bunch in vertical ( $\sigma_y$ ) and horizontal ( $\sigma_x$ ) direction is given. It can be seen, that the instantaneous luminosity only depends on machine parameters.

## 4.2 The ATLAS detector

The ATLAS detector is one of the two big multi purpose detectors at the LHC. It is designed to be able to investigate a wide range of physics. Therefore, several detecting subsystems are arranged in cylindrical layers around the interaction point. A schematic of the ATLAS detector can be seen in Fig. 4.2.

The innermost layer is called Inner Detector [112, 113]. It consists of the Pixel Detector [114], the Semiconductor Tracker (SCT) [115] and the Transition Radiation Tracker (TRT) [1]. The Inner detector is used to measure charged particles. The subdetectors are immersed in a 2 T magnetic field which is created by a solenoid magnet [116]. With the curvatures caused by the magnetic field, it is possible to measure the energy and the sign of charged particles. Outside of the solenoid magnet, the calorimeters are placed [117]. First the liquid Argon calorimeter (lAr) [118] which is designed to measure electromagnetic interacting particles and will be described in more detail in Section 4.2.1. It is followed by the tile hadronic calorimeter (HCAL) [119] that is used to measure energies of hadrons. After that, the toroidal magnet is located. It creates a magnetic field up to 3.5 T which bends



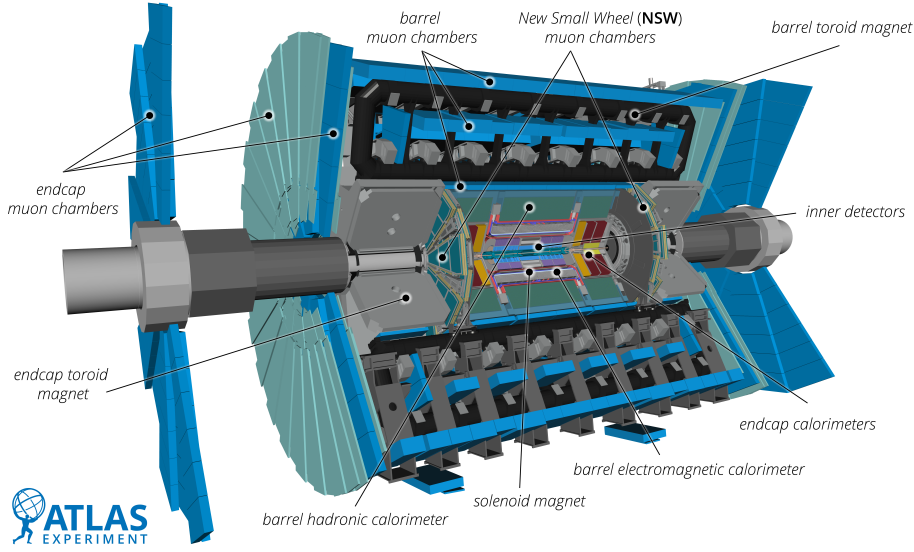


Figure 4.2: Schematic of the ATLAS detector with labelled subsystems. From the inside these are the Inner Detector, the solenoid magnet, the electromagnetic calorimeter and the hadronic calorimeter, the muon chambers and the toroidal magnet. [111]

the trajectories of muons. Muons are measured with the outermost part of the detector, the Muon Spectrometer [120].

A Cartesian coordinate system can be defined with the  $z$ -axis along the beam pipe,  $x$ -axis pointing towards the centre of the LHC ring and the  $y$ -axis vertically upwards. Using this Cartesian coordinate system, an often used observable, the transverse momentum can be expressed as

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (4.2)$$

Since the detector itself has a cylindrical shape, it can be beneficial to use a cylindrical coordinate system. Therefore, two angles are defined. The azimuthal angle  $\varphi$  ranging from  $-\pi$  to  $+\pi$ .  $\Theta$  denotes the polar angle to the beam pipe, starting from positive  $z$ . However, the polar angle is usually expressed by the rapidity

$$Y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) \quad (4.3)$$

or the pseudo rapidity

$$\eta = -\ln \left[ \tan \left( \frac{\Theta}{2} \right) \right] \quad (4.4)$$

With these quantities, the distance between two objects  $i$  and  $j$  can be written as

$$\Delta R = \sqrt{(\Delta \eta_{ij})^2 + (\Delta \varphi_{ij})^2} \quad (4.5)$$

### 4.2.1 The ATLAS liquid Argon calorimeter

The ATLAS calorimeter is designed to absorb most known particles, except muons and neutrinos. A sketch of the ATLAS calorimeter can be seen in Fig. 4.3. Electrons and photons are preliminary measured in the ECAL. The ECAL consists of alternating layers of lead as a passive absorber and liquid argon as the active medium. An incoming photon or electron interacts with the lead, initiating an electromagnetic shower of secondary electrons and photons. These ionize the liquid argon. The amount of ionization is recorded by the electronics and is proportional to the energy of the original particle.

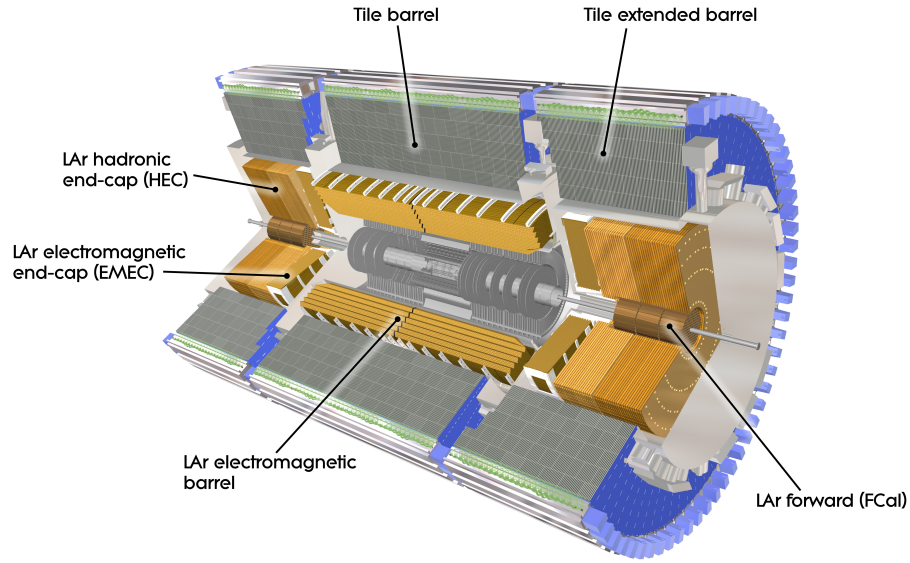


Figure 4.3: Sketch of the ATLAS calorimeter. [121]

The layers have an accordion structure. This is shown in Fig. 4.4, where a small diameter of the calorimeter at  $\eta = 0$  is depicted. The sketch also shows the dimensions of the ECAL. It consists of four layers. First a presampler to correct for the energy loss due to material of detector components in front of the ECAL. Afterwards, three sampling layers are placed. It covers a range of 24 interaction lengths  $X_0$ . It can be seen, that layer 1 has a finer granularity in  $\Delta\eta$ , whereas layers two and three have a better granularity in  $\Delta\phi$ .

### 4.2.2 The ATLAS Trigger

Since the LHC produces up to 40 million collisions per second and ATLAS is not able to save these amounts of data in the short time, it is necessary to reduce the number of events to save. It would be beneficial to preselect interesting events, while those with less energy for example are discarded.

Therefore the ATLAS trigger is used. It consists of two parts. The first is the hardware based L1 trigger. It uses data from the calorimeters and the muon system which are processed by custom electronics such as FPGAs. If the L1 trigger decides to keep the event, the complete detector data is stored in a buffer. This reduces the data-flow from 40 MHz to 100 kHz.

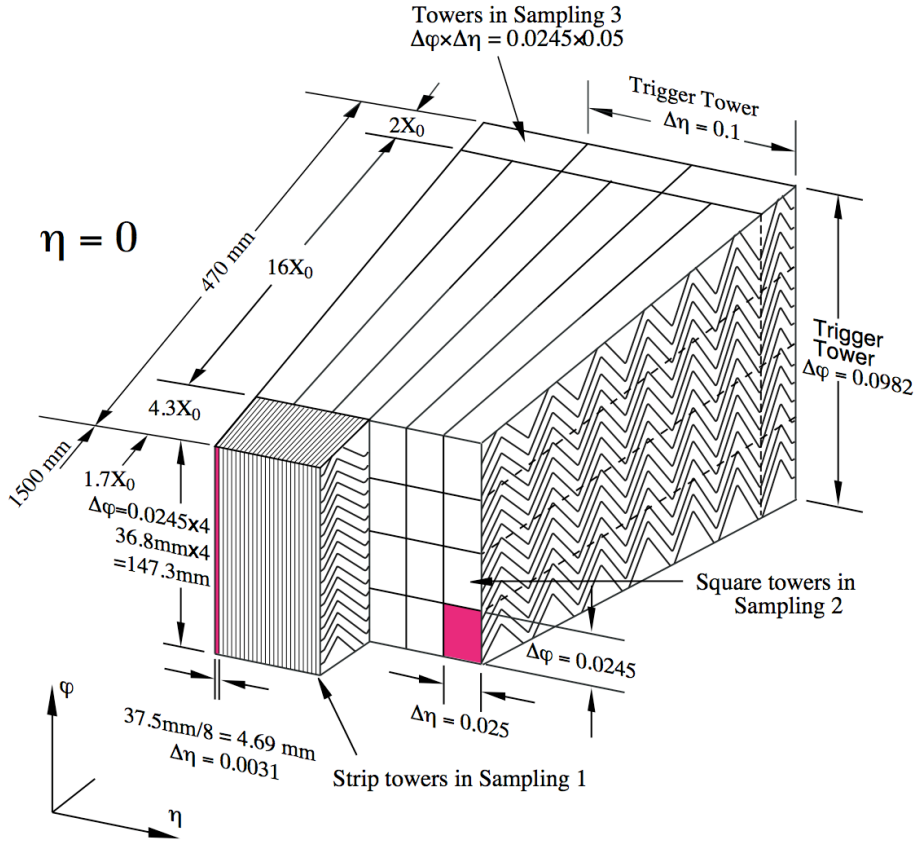


Figure 4.4: Sketch of the sampling layers of the ECAL. The accordion structure is indicated by solid zigzag lines. The dimensions of the calorimeter cells is given as well as the depth of the calorimeter in terms of  $X_0$ . [118]

Afterwards, the software based High-Level Trigger (HLT) further reduces this to 1 kHz. This is done using full event reconstruction including track reconstruction and energy clustering. Depending on the objects used in an analysis, the so called trigger menu can be chosen. E.g. an analysis searching for final states with two muons, can select only events that were selected by a trigger requiring two muons in the event.

### 4.2.3 Object reconstruction in ATLAS

In order to analyse the detector data in a meaningful way, it is necessary to transform it such that they describe physical objects as close possible. This transformation is done by several ATLAS object reconstruction methods and algorithms.

A first step is to reconstruct the tracks charged particles leave in the inner detector (and muon spectrometer in case of muons). First, these energy deposits have to be transformed to space points, i.e. the coordinates of the detector part hit by the particle has to be found out. Afterwards, the algorithm searches for seeds, which are 3 space points in different layers of the detector. These three points are connected by a helical which is extrapolated to the interaction point. If there are other points on the curvature, they are added using a Kalman filter. Due to the high number of hits in the detector, it is possible, that more than one track can be associated to a seed. In this case, this ambiguity needs

to be solved by an ambiguity-solving algorithm. Finally, the tracks are fitted again using also TRT information. The track reconstruction is described in more detail in [122].

The reconstructed tracks are used to reconstruct more complex objects, such as jets, electrons and muons. Especially the differentiation between electrons and photons is important for the analysis described later on. Therefore a good track reconstruction is essential.

### Muon reconstruction in ATLAS

Muons are reconstructed in ATLAS by combining information from the Inner Detector (ID) and the Muon Spectrometer (MS), which provides precision measurements over a wide pseudorapidity range up to  $|\eta| < 2.7$  [123]. The reconstruction strategy depends on the detector regions and the quality of the available information, and is designed to achieve high efficiency while maintaining a low fake rate.

There are several types of reconstructed muon objects:

- **Combined muons**, which are reconstructed by matching tracks independently reconstructed in the MS and ID, and refitting them together.
- **Segment-tagged muons**, where a track in the ID is matched to at least one segment in the MS.
- **Standalone muons**, reconstructed using only MS information, typically in regions outside ID coverage.
- **Calorimeter-tagged muons**, identified using energy deposits compatible with a minimum ionising particle in the calorimeter, when MS information is not available.

Combined muons form the highest-quality category and are used in most physics analyses. The reconstruction process includes track finding, fitting with a Kalman filter, and momentum calibration using cosmic rays, resonances like the  $Z$  boson and  $J/\psi$ , and beam halo events.

Muon reconstruction performance is validated using data and simulation, with efficiencies typically above 96 % in most of the detector acceptance [124]. The reconstruction is robust against high pile-up conditions and remains efficient in dense environments, such as the core of high- $p_T$  jets or boosted topologies.

### Jet Reconstruction in ATLAS

Quarks and gluons hadronise almost instantaneously after being produced, due to colour confinement. In this process, they form hadrons, which may undergo further decays, ultimately giving rise to a collimated spray of particles.

These particle showers leave distinct signatures in the ATLAS subdetectors, extending up to the hadronic calorimeter (HCAL). The resulting energy deposits can be clustered into jets using various jet clustering algorithms. Jets serve as experimental proxies for the original partons and are essential tools in collider physics.

The theoretical foundations of jets and a comprehensive overview of jet algorithms are presented in [125]. In this analysis, jets are reconstructed using the anti- $k_T$  algorithm [126], as implemented in FastJet [127], with a distance parameter of  $R = 0.4$ . This algorithm is integrated into the ATLAS software framework, and its performance, as well as the associated jet energy calibration and pile-up mitigation techniques, are extensively documented in [128, 129].

### Photon reconstruction and identification

Both electrons and photons can be reconstructed from clusters [130, 131] in the ECAL with transverse energies  $E_T > 2.5$  GeV.[132, 133] For electron candidates, this cluster can be associated with a track extrapolated to the calorimeter. The track with an extrapolation closest to the barycentre of a cluster is selected.

Photons instead are supposed to have no associated tracks in the inner detector. Hence, clusters fulfilling certain isolation criteria and no matching track are selected as unconverted photon candidates. However, a fraction of photons interacts with the detector material in front of the calorimeter producing electron-positron pairs ( $\gamma \rightarrow e^+ e^-$ ). These are called converted photon candidates. The conversion tracks have to fulfil several criteria such as stemming from a secondary vertex and having opposite charge.

However, it showed that using variable cluster sizes can be beneficial compared to clusters with fixed sizes. This is realized using so-called superclusters. The association of these superclusters to the different kinds of electromagnetic particles is not so easy any more and can variate with different energies. Hence a first energy calibration is applied to those superclusters. Afterwards an algorithm decides of what kind the supercluster is. This can either be only photon, only electron or both. The purpose is to have objects being reconstructed as photons or electrons only directly using the correct calibration or leaving the possibility of selecting to use the object as electron or as photon during the analysis applying the correct calibration on the run.

Finally the energy of all objects is recalibrated using the classification of the previous step. This corrects energy losses in the detector material. Additionally, the shower properties are derived and saved as shower-shape variables.

These shower-shape variables are used as discriminating variables to achieve photon identification (ID) [134]. There are two levels of photon ID, Tight and Loose. While only information from the second sampling layer of the ECAL and information from the HCAL are used to exploit the Loose ID, the whole granularity of the ECAL is used for the Tight ID. The relevant variables are listed in ???. Not only the higher number of discriminating variables, but also stricter selection criteria on the variables used for the Loose ID assures a much better background rejection in the Tight ID.

It should be noted that the ATLAS photon reconstruction is optimised to reconstruct prompt photons and reject others. This is done to minimise the amount of photons from hadron decays and misidentified hadrons. [134]

#### 4.2.4 Muon Data Quality in the ATLAS Experiment

Data-taking at the Large Hadron Collider (LHC) is a highly coordinated process involving both the accelerator complex and the associated experiments. The LHC team prepares the beams by accelerating bunches of protons to a pre-defined energy—most recently 6.8 TeV per beam for Run 3 operations—and bringing them into collision at interaction points where the detectors, such as ATLAS, are situated. Prior to the start of collisions, the ATLAS detector transitions into a data-taking mode: all subsystems (inner detector, calorimeters, muon spectrometer, trigger and data acquisition systems) are brought online and configured for stable beam conditions.

Once beam conditions are deemed stable, the experiment begins recording data. A single data-taking period, referred to as a **run**, typically lasts for several hours and is continuously monitored by trained shift crews operating from the ATLAS control room. Their task is to ensure that all detector systems

are functioning correctly in real time. If issues arise—such as a sub-detector failing to record data properly—corrective action must be taken immediately to minimize data loss. This online monitoring phase is critical: once collisions occur, they cannot be repeated, and any undetected failure in the detector would result in irretrievable data loss.

Following the end of each run, the recorded data are processed through several stages of reconstruction, converting raw detector signals into high-level physics objects such as tracks, jets, and muons. It is during this **offline** phase that data quality is assessed in greater detail.

To ensure that this work can be carried out with sufficient rigour, both institutes and individual collaborators are required to contribute to the development and maintenance of code and algorithms, as well as to support tasks in daily data-taking operations.

Alongside the analysis and thesis work, I contributed to the development and coordination of the muon offline data quality (DQ) efforts.

Although this analysis focuses on photons rather than muons, examples from the muon data quality domain are presented. This is because the muon system provides well-established validation techniques, which are routinely employed to assess the overall performance and reliability of the ATLAS detector during data-taking. These methods offer illustrative insight into how detector quality is monitored and understood. Similar methods are employed to ensure high-quality photon reconstruction, making use of information from both the Inner Detector and the calorimeters.

For the muon spectrometer—which plays a crucial role in many precision measurements and searches for new physics—it is essential to verify that all subcomponents are fully operational and exhibit the expected behaviour. To monitor detector performance in finer time intervals, a **run** is divided into smaller units of time (or integrated luminosity), known as Lumiblocks. The stability of detector segments can be assessed by plotting quantities such as the number of hits per Lumiblock. This is illustrated for the TGC detector and for various types of reconstructed muon tracks in Fig. 4.5.

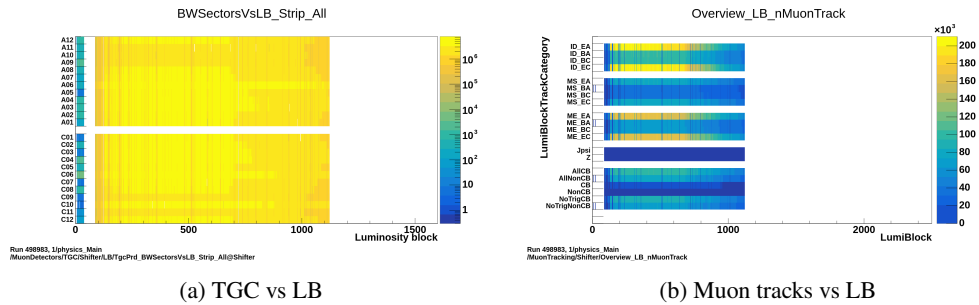


Figure 4.5: Hits in the TGC detector per Lumiblock (a) and different types of muon tracks per Lumiblock (b). These plots were generated automatically by the ATLAS offline DQ framework [135].

Building on the involvement in muon offline data quality efforts, several new algorithms were developed to probe and validate the performance of the ATLAS Muon System in Run 3. This work is particularly relevant given the major upgrades implemented during the long shutdown between Run 2 and Run 3, which included the installation of the New Small Wheel (NSW) detectors and improvements to the muon trigger system.

The newly developed tools evaluate data-taking efficiency across all muon chambers, validate the synchronisation and response of timing detectors, and monitor the reconstruction of well-known resonances such as the  $J/\psi$  meson and  $Z$  bosons, whose invariant mass peaks serve as benchmarks for



the spectrometer's calibration and alignment.

Examples of these checks are illustrated in Figs. 4.6 and 4.7. Fig. 4.6 shows the invariant mass distribution of oppositely charged muon pairs around the  $Z$  boson mass. A Gaussian fit is applied to extract the mean and standard deviation of the distribution, which are then compared to reference values. If both the mean and standard deviation lie within the expected range, the algorithm assigns a **green** status flag. Otherwise, a **yellow** or **red** flag is issued.

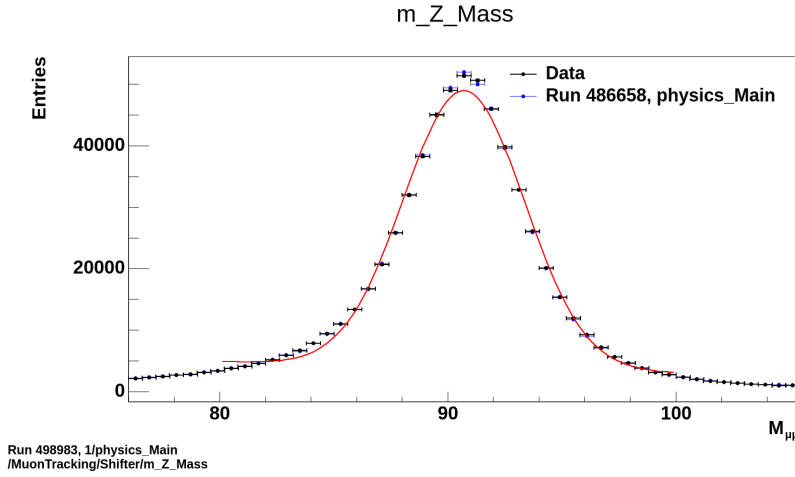
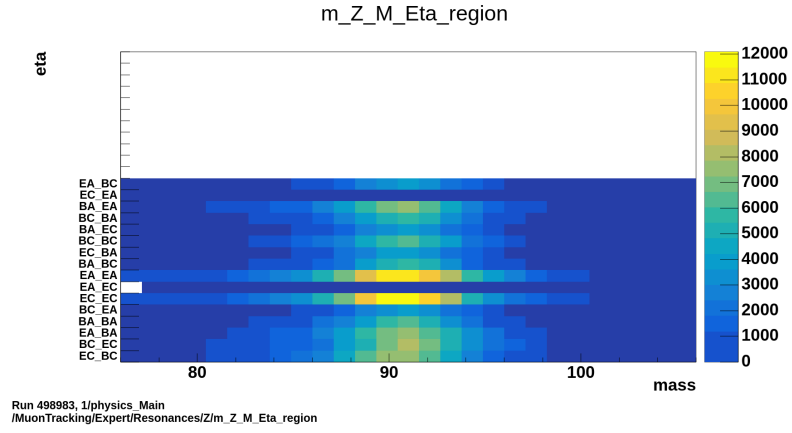


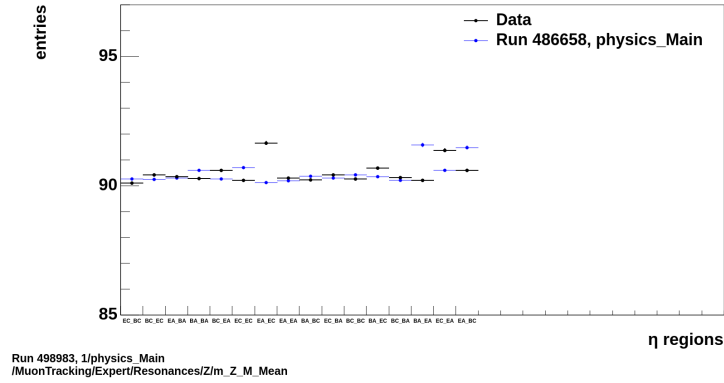
Figure 4.6: ATLAS muon offline DQ plot showing the invariant mass distribution of opposite-sign muon pairs around the  $Z$  boson mass. The black histogram represents data from the run under investigation, while the blue histogram shows a reference run. A Gaussian fit (red curve) is used to compare the mean and width of the peak to reference values. This plot was generated automatically by the ATLAS offline DQ framework [135].

If discrepancies are observed in this summary plot, more detailed views are available, showing contributions from different regions of the detector. Some examples of these are shown in Fig. 4.7.

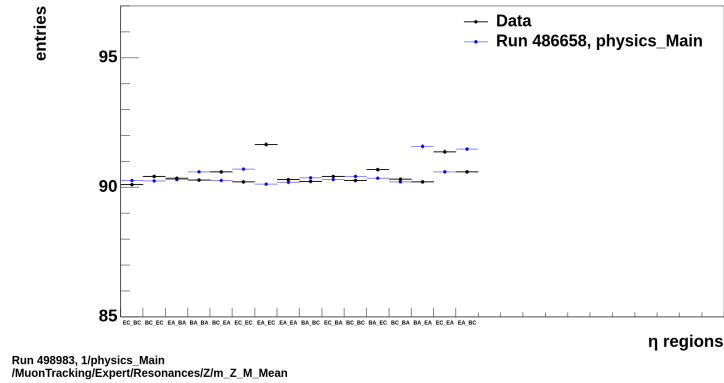
If the data passes all quality criteria—including hardware performance, trigger response, and consistency of physics observables—it is signed off by the relevant data quality teams. Only after this certification is the data deemed suitable for use in physics analyses. The reliability of this procedure is critical to ensuring the robustness and reproducibility of ATLAS physics results.



(a)  $m_Z$  per detector region  
m\_Z\_M\_Eta\_regionprojection mean



(b) mean of  $m_Z$  per det. region  
m\_Z\_M\_Eta\_regionprojection mean



(c)  $\sigma$  of  $m_Z$  per det. region

Figure 4.7: Invariant mass of opposite-sign muon pairs in different detector regions. Panel (a) shows the mass distribution by detector region, while (b) and (c) show projections of the mean and standard deviation ( $\sigma$ ) of the mass distributions. These plots were generated automatically by the ATLAS offline DQ framework [135].



---

## Experimental methods

---

Not only the experimental setups but also the experimental methods are subject to constant improvements. While a few decades ago, most analyses were done with so-called cut and count experiments, nowadays most particle physics measurements and searches profit from novel data analysis approaches such as machine learning. This chapter gives a brief overview on machine learning in general and techniques used in the later chapters of this work in particular.

### 5.1 Machine Learning

Machine learning is a branch of artificial intelligence. Instead of creating separate algorithms for different problems, the goal of machine learning is to generalise so that a single model can adapt and perform well on a variety of tasks by learning patterns from data rather than relying on explicitly programmed rules. Some early thoughts about machine learning were done beginning mid of the last century [136]. However, the main developments started in the last years with increasing computation power.

There is a wide range of machine learning algorithms, including boosted decision trees (BDTs), support vector machines (SVMs), random forests, and various types of neural networks (NNs), among others. The fundamental principle behind all these methods is that they learn from a given dataset to identify patterns and structures, enabling them to generalise and make predictions on previously unseen data. Machine learning algorithms are based on different paradigms, including unsupervised learning and reinforcement learning. Unsupervised learning works with unlabelled data, identifying patterns and structures without explicit guidance, while reinforcement learning focuses on optimizing decision-making processes, such as in automated navigation. However, in particle physics, the predominant approach is supervised learning.

In supervised learning, the training process is based on labelled data, where each input is associated with a corresponding expected output. A set of input-output pairs is referred to as a dataset, and each input consists of multiple input variables or features along with an output label. A single input-output pair is often referred to as an event or sample. However, to avoid confusion with detector-recorded events or simulated samples, this thesis will use the term **feature vector** throughout.

Depending on the type of output label, the machine learning task can be categorized as **binary classification** (with exactly two discrete labels), **multiclass classification** (with more than two discrete

labels), or **regression**, where the output is a continuous variable. This section uses information from [137–140] where especially [139] provides a good introduction to machine learning methods.

At first, the general concept of supervised learning and its terminology will be briefly explained (Section 5.1.1). Afterwards, artificial neural networks (ANNs) will be discussed Section 5.1.2 and finally the specifics of convolutional neural networks will be introduced Section 5.1.3.

### 5.1.1 General concept of supervised learning

Supervised learning is a machine learning approach where a model (a complicated function) is trained on labelled data, meaning that each input (feature vector) has a corresponding known output (label). For classification tasks, the model usually maps a feature vector  $\vec{x} = (x_1, \dots, x_n)$  of  $n$  input features to an output vector  $\vec{y} = (y_1, \dots, y_m)$  of  $m$  classes. The outputs  $y_i$  can be interpreted as a probability that the input vector belongs to class  $i$ . For binary classification, the vector is usually reduced to length one, where zero is assigned to one class (background) and one to the other (signal), leaving the output  $\hat{o}$  as probability to be signal and  $1 - \hat{o}$  as probability to be background. The goal for the model is to learn the relationship between inputs and outputs so it can make accurate predictions on unseen data. The advantage of machine learning models is their ability to exploit higher-order correlations between input variables, which would be challenging to identify manually.

Before training the model, the input data has to be prepared. For supervised learning it is crucial to have a label for each feature vector. Furthermore, all inputs need to have the same input variables. It can be beneficial, to normalise all input variables to the same range.

Typically, the dataset is divided into three parts: the training set, used to train the model; the validation set, used to evaluate performance and adjust model parameters during training; and the test set, which assesses the final trained model. In particle physics, best practices recommend avoiding further use of the training dataset whenever possible. Any additional samples used for fits or limit setting should be statistically independent to ensure unbiased results.

To quantify how much the predicted values deviate from the true output labels, loss functions are used. A common example is the Mean Squared Error (MSE), which measures the squared differences between the true and predicted values:

$$L = \sum_j^{n_x} \sum_k^{n_{classes}} (y_{jk} - \hat{y}_{jk})^2 \quad (5.1)$$

where  $\hat{y}_{jk}$  represents the true labels,  $y_{jk}$  the predicted values,  $n_x$  the number of feature vectors, and  $n_{classes}$  the number of classes.

The model's parameters are updated to minimise the loss function, a process known as training. This iterative optimization continues until the loss function converges to a stable minimum or the maximum number of iterations is reached.

During training, the model's parameters are updated after each batch, progressively reducing the loss. A complete pass over the entire dataset is called an epoch, and training typically requires multiple epochs to refine the model's performance. However, training for too many epochs can cause the model to memorize the training data instead of learning general patterns – a phenomenon known as overfitting (or overtraining). It usually arises if the number of trainable parameters is large compared to the training dataset. Overfitting results in poor generalization, meaning the model performs well on training data but struggles with unseen data.

Several strategies can be used to prevent overfitting and improve the model's ability to generalize. Training in batches is already one of them. Regularisation terms can be added to the loss function to penalise large weights [141]. Another method is to randomly disable parts of the model, which forces the model to learn more robust features. This is called dropout [142]. Furthermore, early stopping routines can be implemented easily to end the training in case there is no further improvement in the loss function [143].

### 5.1.2 Artificial neural networks (ANNs)

Artificial Neural Networks (ANNs) are computational models inspired by the structure and function of biological neural networks. They consist of layers of interconnected neurons (or nodes), where each neuron processes inputs using weighted connections and activation functions to produce an output. The first layer or input layer consists of one node per input variable which is directly fed into. The last layer or output layer consists of one node for binary classification or regression tasks or one node per class in a multiclassification task. All layers in between are called hidden layers. In a feed forward network, all nodes are connected to all nodes in the following layer. The connection between nodes  $i$  and  $j$  has a weight  $w_{ij}$  that can be updated. A simplistic model of an ANN with one hidden layer and three input features can be seen in Fig. 5.1.

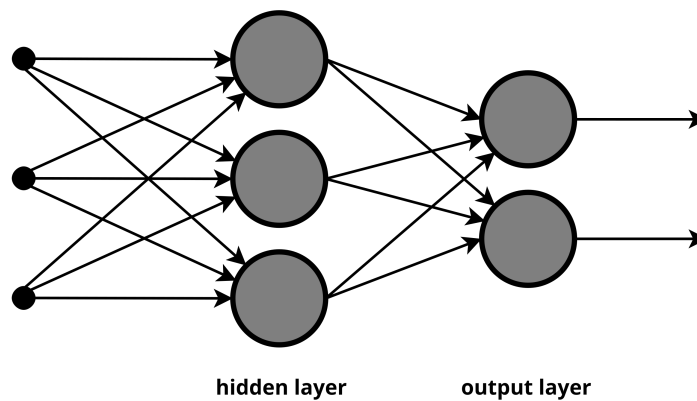


Figure 5.1: Model of an ANN with one hidden layer, three input features and two output nodes. Taken from [144]

In every node of the ANN, an output value is computed using a deterministic function

$$o = f \cdot g(x_i \cdot w_i) \quad (5.2)$$

with an arbitrary function  $f$  called *activation function*. The function  $g(x_i w_i)$  instead computes an output considering all inputs of the node. It depends on the architecture of the node itself. But in most cases it is the weighted sum

$$g(x_i w_i) = \sum_i x_i w_i. \quad (5.3)$$

Beginning from the output layer of the ANN the weights of all nodes get updated. This is known as backpropagation [145]. Having an output function in dependence of all weights  $o(w_1, \dots, w_i)$ , each

weight can be updated by subtracting the gradient of the Loss function

$$\lambda \frac{\partial L}{\partial w_i}. \quad (5.4)$$

This method is called *GradientDescent* and  $\lambda$  is a tunable hyperparameter called learning rate.

### 5.1.3 Convolutional neural networks (CNNs)

Convolutional Neural Networks (CNNs) are a specialized type of artificial neural network designed for processing structured grid-like data, such as images. Unlike traditional fully connected neural networks, CNNs efficiently capture spatial hierarchies and local patterns in data by leveraging convolutional and pooling layers.

CNNs have revolutionized fields like computer vision, enabling breakthroughs in image classification, object detection, and facial recognition. Their ability to automatically learn relevant features from raw data makes them a powerful tool for pattern recognition tasks.

The core building blocks of CNNs include:

- **Convolutional layers**, which extract local patterns and hierarchical features;
- **Pooling layers**, which reduce the spatial resolution and increase robustness;
- **Dense layers**, which interpret the learned features and generate output.

These components are explained in more detail in Section 5.1.4, including visualizations of how convolution and pooling operate.

### 5.1.4 Common Neural Network Layers

To better understand the practical construction and capabilities of neural networks – especially artificial neural networks (ANNs) and convolutional neural networks (CNNs) – it is useful to examine the typical layers that compose these architectures. While network design can vary considerably depending on the task, several layer types appear consistently due to their fundamental computational roles. Importantly, layers do not have to adhere to predefined types; depending on the use case, customized layers can be beneficial for solving problems more efficiently. Such customizations have led to the development of more advanced architectures, such as graph neural networks (GNNs) [146, 147].

#### Dense Layers

Dense layers, also known as fully connected layers, form the foundation of traditional ANNs and are commonly used at the output stage of CNNs. Each neuron in a dense layer is connected to all outputs from the previous layer. Each node has an associated bias term, which – like the connection weights – is a trainable parameter. Adjusting these parameters allows the network to learn complex, high-level interactions between features.

Given  $n_{\text{in}}$  inputs and  $n_{\text{out}}$  outputs, a dense layer introduces  $n_{\text{in}} \times n_{\text{out}}$  trainable weights and  $n_{\text{out}}$  bias parameters. This large number of parameters enables the model to capture global feature interactions but also increases the risk of overfitting (see Section 5.1.5).

## Activation Layers

Non-linear activation functions are applied after linear transformations (e.g., in dense or convolutional layers) to introduce non-linearity into the network. Without them, even deep architectures would behave like a single-layer linear model. Common choices include the rectified linear unit (ReLU), which promotes sparse activations and mitigates vanishing gradient issues, and the softmax function, typically used in the final layer for multi-class classification tasks.

In most frameworks, activation functions are specified as part of the preceding layer (e.g., a dense layer with an activation), rather than as separate layers. However, when designing custom layers, explicitly including an activation layer may be crucial.

The most common activation function can be seen in Fig. 5.2. Their definition as well as their derivatives can be seen in Table 5.1.

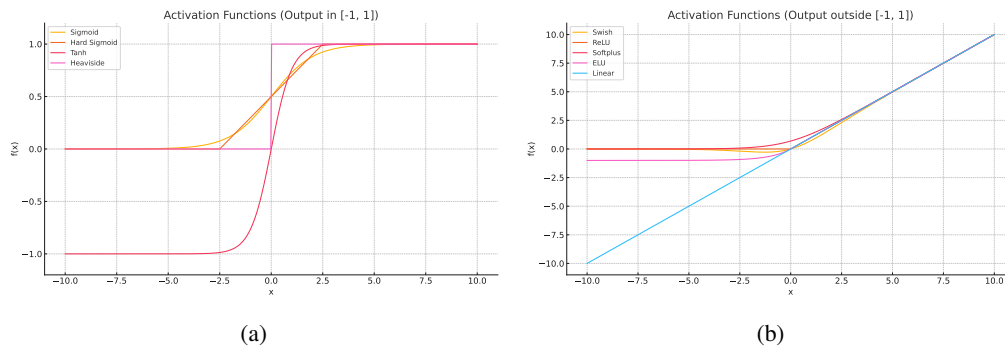


Figure 5.2: Commonly used activation functions restricted to outputs between  $-1$  and  $1$  (a) and without range restrictions (b).

Table 5.1: Commonly used activation functions with definition and first derivation

| Activation   | Definition $f(x)$                  | Derivative $f'(x)$                             |
|--------------|------------------------------------|------------------------------------------------|
| Sigmoid      | $\frac{1}{1+e^{-x}}$               | $f(x)(1 - f(x))$                               |
| Swish        | $x \cdot \text{sigmoid}(x)$        | $\sigma(x) + x \cdot \sigma(x)(1 - \sigma(x))$ |
| Hard Sigmoid | $\text{clip}(0.2x + 0.5, 0, 1)$    | 0.2 if $-2.5 < x < 2.5$ , else 0               |
| Softmax (1D) | $\frac{e^{x_i}}{\sum_j e^{x_j}}$   | Jacobian matrix (complex)                      |
| ReLU         | $\max(0, x)$                       | 1 if $x > 0$ , else 0                          |
| Tanh         | $\tanh(x)$                         | $1 - \tanh^2(x)$                               |
| Softplus     | $\log(1 + e^x)$                    | $\frac{1}{1+e^{-x}}$                           |
| Heaviside    | 1 if $x \geq 0$ , else 0           | 0 (undefined at $x = 0$ )                      |
| ELU          | $x$ if $x \geq 0$ , else $e^x - 1$ | 1 if $x \geq 0$ , else $e^x$                   |
| Linear       | $x$                                | 1                                              |

tab:methods:ML:layers:activation

**Batch Normalization Layers**

Batch normalization layers stabilize and accelerate training by normalizing the input distribution to each layer. By maintaining consistent activation statistics across training batches, they help mitigate issues such as internal covariate shift and allow for higher learning rates. These layers are typically placed between the linear transformation and the non-linear activation.

As with activation functions, batch normalization is often implemented directly within common layer types.

**Dropout Layers**

Dropout layers are introduced to reduce overfitting. They are typically placed between the output of one layer and the input of the next. Instead of introducing additional trainable parameters, dropout layers randomly disable a specified percentage of connections during each training step. This effectively samples a different subnetwork from the full architecture at every iteration. As a result, the network is encouraged to learn redundant representations and avoid relying too heavily on any single feature. During evaluation or inference, dropout is disabled, and the full network is used without dropping any connections.

**Convolutional Layers**

Convolutional layers are the defining component of convolutional neural networks (CNNs). They apply learnable filters to local regions of the input, enabling the network to extract spatial hierarchies of features such as edges, textures, and object parts.

A convolution operation can be written as:

$$Y(i, j) = \sum_m \sum_n X(i + m, j + n) \cdot W(m, n) \quad (5.5)$$

with the input values under the filter  $X(i + m, j + n)$  and  $W(m, n)$  representing filter weights. This produces a new feature map where specific patterns detected by the filter are highlighted.

Unlike dense layers, the number of trainable parameters in a convolutional layer does not depend on the size of the input. In dense layers, each input feature is connected to every node, leading to a parameter count that scales with input dimensionality. In contrast, the number of parameters in a convolutional layer is determined solely by the filter dimensions, the number of filters, and the number of input channels.

For a convolutional layer with  $n_{\text{filters}}$  filters, each of spatial size  $k_h \times k_w$  and operating on an input with  $c_{\text{in}}$  channels, the total number of trainable parameters is:

$$n_{\text{params}} = n_{\text{filters}} \cdot (k_h \cdot k_w \cdot c_{\text{in}} + 1), \quad (5.6)$$

where the additional +1 accounts for a bias term per filter.

The output spatial dimensions ( $H_{\text{out}}, W_{\text{out}}$ ) of a convolutional layer applied to an input of height  $H_{\text{in}}$  and width  $W_{\text{in}}$  are given by:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} - k_h + 2p_h}{s_h} \right\rfloor + 1, \quad W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} - k_w + 2p_w}{s_w} \right\rfloor + 1, \quad (5.7)$$

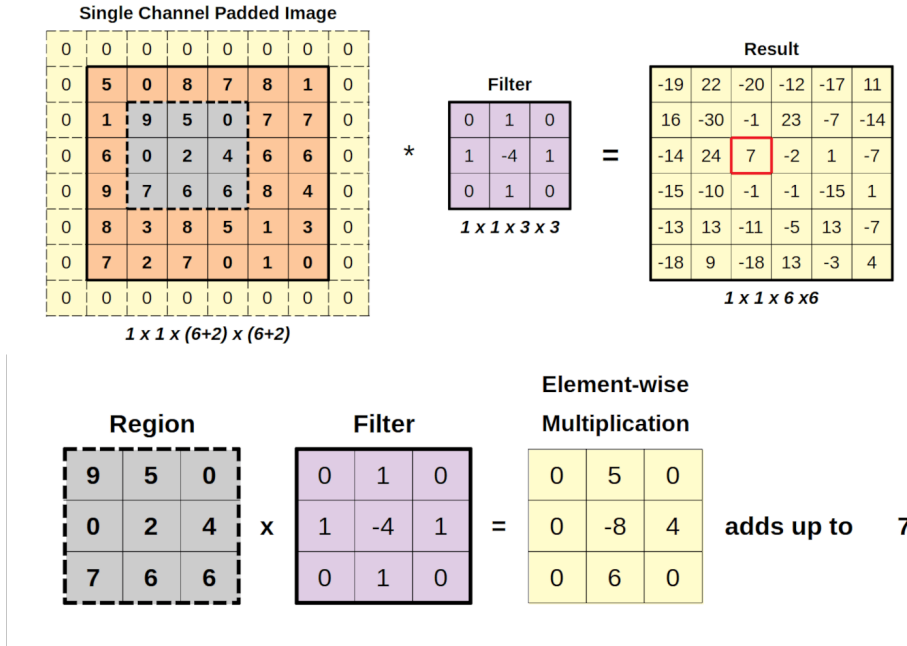


Figure 5.3: Visual representation of a convolution in a CNN. The  $3 \times 3$  filter slides over an input matrix, performing elementwise multiplication and summation to produce a new feature map. Taken from [148].

The variables in the above formula are defined as follows:  $k_h$  and  $k_w$  denote the height and width of the filter, respectively.  $p_h$  and  $p_w$  represent the amount of padding applied to the input along the vertical and horizontal dimensions.  $s_h$  and  $s_w$  are the stride values in the vertical and horizontal directions, indicating how far the filter moves during each step of the convolution.

Common padding strategies include:

- **Valid padding:** No padding is applied ( $p_h = p_w = 0$ ), so the filter only operates on positions where it fully fits inside the input. This reduces the output size.
- **Zero padding (or same padding):** The input is padded with zeros (or same number) at the borders to preserve its spatial dimensions, i.e.,  $H_{\text{out}} = H_{\text{in}}$  and  $W_{\text{out}} = W_{\text{in}}$ , when stride is 1.

In the context of particle physics, convolutional layers can be used to process detector data arranged in the  $\eta$ - $\varphi$  plane. The granularity may correspond to the physical spacing of detector cells or to a discretisation of arbitrary  $\eta$  and  $\varphi$  values. Different detector layers can be treated as separate input channels, analogous to colour channels in image data.

### Pooling Layers

As convolutional layers with  $n_{\text{filters}} > 1$  increase the number of feature maps, the dimensionality of the intermediate representations grows rapidly with network depth. While convolutional layers themselves introduce relatively few parameters compared to dense layers, stacking many of them can result in significant computational and memory demands.

To manage this growing complexity and reduce the spatial resolution of feature maps while retaining the most important information, pooling layers are commonly used. These layers perform a fixed aggregation operation over small local regions of the input, typically reducing both the height and width of the feature maps.

Pooling layers are usually inserted after convolutional layers. Their main role is to downsample the spatial dimensions, which reduces the computational cost and promotes translation invariance by summarizing local features.

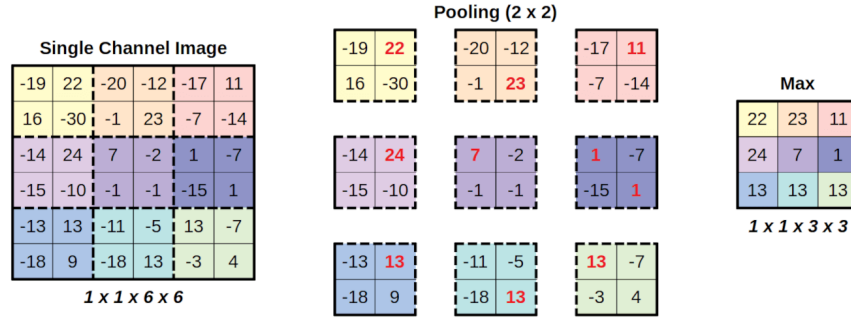


Figure 5.4: Sketch how a  $2 \times 2$  max pooling window moves across the feature map selecting the maximum value in each region. Taken from [149].

The two most common types of pooling are:

- **Max pooling:** Selects the maximum value in each local window (e.g., a  $2 \times 2$  region), emphasizing the most active features.
- **Average pooling:** Computes the mean value over each local window, leading to smoother feature representations.

Given an input feature map of size  $H_{\text{in}} \times W_{\text{in}}$ , a pooling window of size  $p_h \times p_w$ , and stride  $s$ , the output dimensions are computed as

$$H_{\text{out}} = \frac{H_{\text{in}} - p_h}{s} + 1, \quad W_{\text{out}} = \frac{W_{\text{in}} - p_w}{s} + 1. \quad (5.8)$$

Note that this formula assumes the pooling window fits exactly and no padding is applied. In practical implementations, rounding may be involved depending on the framework settings.

Pooling layers do not contain any trainable parameters, making them computationally efficient and widely used for spatial compression in convolutional neural networks.

### 5.1.5 Overfitting and regularisation

If too many free parameters (weights) are fitted, an artificial neural network (ANN) can learn to **perfectly classify** the training data, including noise and statistical fluctuations. While this may appear to indicate good performance, such a model typically fails to generalise to new, unseen data. This phenomenon is known as **overfitting** (or **overtraining**), and it results in poor predictive performance outside the training set.



Overfitting is particularly likely when the number of trainable parameters in the model is large relative to the size of the training dataset. In such cases, the model has enough capacity to memorise the training data instead of learning general patterns. Therefore, a careful balance must be struck between model complexity and the amount of training data.

Several strategies have been developed to mitigate overfitting:

- **Train/Validation/Test Split and Early Stopping:** A common approach is to split the dataset into three subsets: a **training set**, a **validation set**, and a **test set**. The training set is used to optimise the model parameters. The validation set monitors generalisation performance during training. If the training loss continues to decrease while the validation loss starts to increase, it indicates overfitting. **Early stopping** halts training at the point where validation performance begins to degrade.
- **L2 Regularisation (Weight Decay):** L2 regularisation adds a penalty term to the loss function proportional to the squared magnitude of the model weights. This discourages excessively large weights and encourages simpler models, which typically generalise better.
- **Dropout:** During training, **dropout** randomly disables a fraction of neurons in the network. This prevents the model from becoming overly reliant on particular nodes and promotes redundancy in the representation. The **dropout rate**, which specifies the fraction of dropped neurons, is a tunable hyperparameter.
- **Reducing Model Complexity:** A simple yet effective regularisation method is to **reduce the number of trainable parameters**, for example by using fewer layers or fewer neurons per layer. This limits the model's capacity, making it less likely to memorise noise in the data and more likely to capture general trends.
- **Learning Rate Tuning:** A **large learning rate** may cause the model to overshoot optimal minima and potentially learn noisy patterns, leading to instability or poor generalisation. Conversely, a **smaller learning rate** promotes more gradual updates, which can improve generalisation, especially when combined with other regularisation techniques.
- **Data Augmentation (if applicable):** In cases where increasing the dataset size is feasible, **data augmentation** – such as applying transformations to input samples – or generating **synthetic data** can effectively increase the amount of training data, thereby reducing the risk of overfitting.

Ultimately, effective regularisation requires balancing the model's capacity (determined by the number of trainable parameters) against the **effective sample size**. A general rule of thumb is that more complex models require more data to avoid overfitting. When data is limited, regularisation becomes critical to achieving good generalisation performance.

### 5.1.6 Model Evaluation and Performance Metrics

A critical component of applying machine learning in particle physics is the evaluation of model performance. Robust and interpretable performance metrics are essential to ensure that trained models generalize well and yield reliable physical insights. This section presents key evaluation techniques for binary classification, multiclass classification, and regression tasks.

## Binary Classification

In binary classification tasks – such as distinguishing between signal and background – a machine learning model outputs a continuous score indicating the likelihood that a given event belongs to the signal class. The performance of the classifier is assessed by applying thresholds to this score and analysing how predictions vary.

A first qualitative evaluation involves inspecting the classifier output distribution. Ideally, signal events should be concentrated near 1, while background events should be concentrated near 0.<sup>1</sup> A clear separation between the two distributions is indicative of good classifier performance. Conversely, if the distributions significantly overlap or are skewed toward a single edge, this may suggest suboptimal model behaviour or training deficiencies.

More quantitative performance measures can be obtained by sweeping over the classifier score threshold and counting:

- True Positive (TP): actual positive and predicted positive
- False Negative (FN): actual positive but predicted negative
- False Positive (FP): actual negative but predicted positive
- True Negative (TN): actual negative and predicted negative

From these, the True Positive Rate (TPR) and False Positive Rate (FPR) are defined as:

$$TPR = \frac{TP}{TP + FN} \quad FPR = \frac{FP}{FP + TN} \quad (5.9)$$

Fig. 5.5 shows the distribution of an exemplary classifier output for signal and background. Dashed vertical lines in red, green, and blue correspond to thresholds of 0.3, 0.5, and 0.7, respectively, and are used consistently throughout the evaluation. The right-hand plot presents the True Positive Rate (TPR) and False Positive Rate (FPR) as functions of the threshold, illustrating the trade-off between sensitivity and background rejection.

The Receiver Operating Characteristic (ROC) curve, shown in Fig. 5.6, plots the TPR against the FPR across all possible thresholds. The Area Under the Curve (AUC) serves as a scalar performance measure, with values closer to 1 indicating better separation between signal and background. Again, exemplary threshold values are marked in red, green, and blue. In addition to the ROC curve, the Significance Improvement Characteristic (SIC) curve illustrates the expected statistical significance as a function of the threshold. It is defined as:

$$SIC = \frac{TPR}{\sqrt{FPR}} \quad (5.10)$$

While the ROC curve visualizes the trade-off between TPR and FPR, the SIC curve focuses on optimizing the significance of signal detection, which is particularly useful for hypothesis testing. It helps identify the threshold that maximises sensitivity while minimizing background contamination.

---

<sup>1</sup> This can vary depending on the implementation of the classifier.

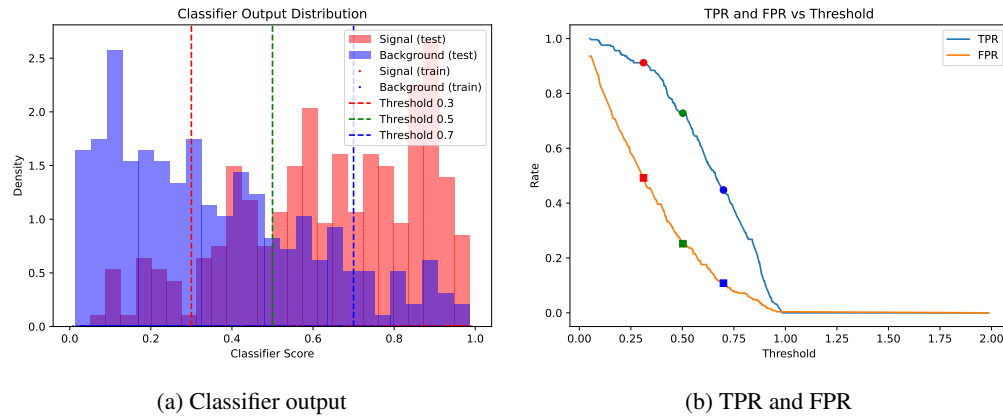


Figure 5.5: Classifier output and corresponding TPR/FPR as a function of threshold.

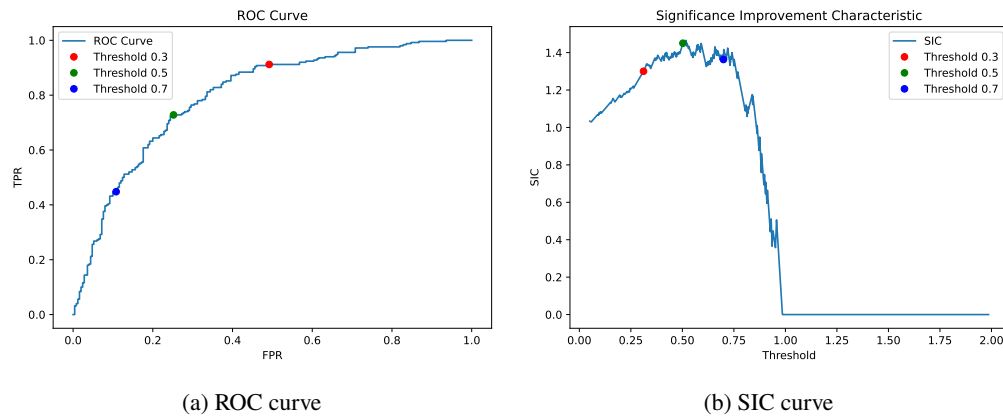


Figure 5.6: Receiver Operating Characteristic (ROC) and Significance Improvement Characteristic (SIC) curves with exemplary thresholds.

## Multiclass Classification

In multiclass classification problems the model predicts among three or more discrete classes. Evaluation techniques include confusion matrices, which visualize the correspondence between predicted and true labels, and ROC curves constructed for each class in a one-vs-rest fashion, as shown in Fig. 5.7.

## Regression

For regression tasks – such as estimating energy, momentum, or spatial positions of intermediate particles – the model predicts continuous values. Evaluation focuses on comparing predicted values to the ground truth using scatter plots and residual distributions, as shown in Fig. 5.8.

The scatter plot on the left hand-side shows the true values on the x-axis and the predicted values on the y-axis. The z-axis symbolically represents the value of an input feature. Most points cluster around the diagonal, indicating a strong predictive performance. Additionally, the colour of the scatter points

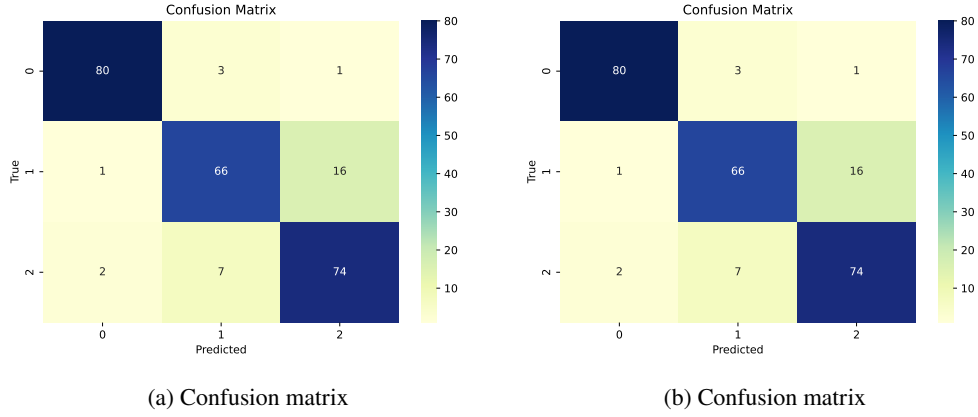


Figure 5.7: Multiclass classification evaluation using confusion matrix with truth class labels on the x-axis and predicted class labels on the y-axis as well as ROC curves for every class/output node.

suggests that, particularly near zero, the relationship between the input feature and the prediction is not entirely linear.

The residual plot on the right hand-side displays the predicted values on the x-axis and the residuals – the differences between predicted and true values – on the y-axis. Again, the z-axis encodes the value of an input feature. The residuals appear symmetrically distributed across the range of predictions, with no visible structure or systematic bias.

When processing large datasets, two-dimensional histograms may be preferable to scatter plots for clarity. Overall, strong alignment between predicted and true values, along with minimal structure in the residuals, indicates high model fidelity.

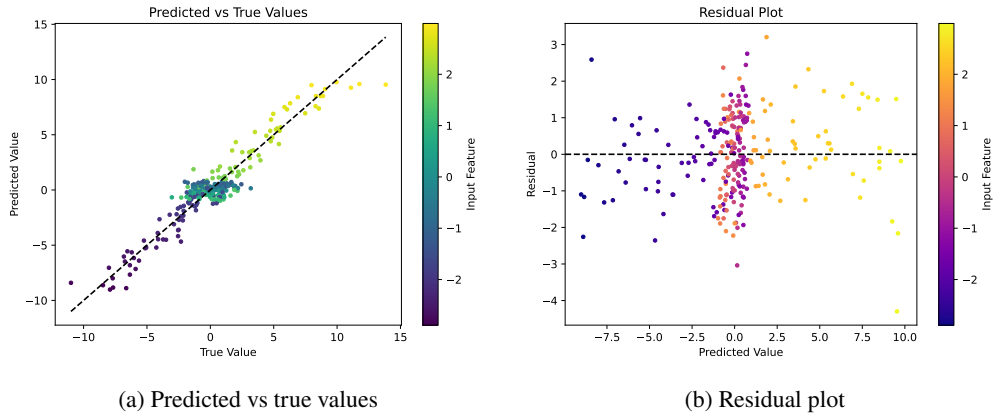


Figure 5.8: Evaluation of regression model using scatter and residual plots.

These evaluation tools form the foundation for selecting and validating machine learning models in high-energy physics, guiding reliable model deployment and statistical inference.

## 5.2 Maximum Likelihood Fits

Maximum Likelihood Estimation (MLE) is a cornerstone technique in statistical inference, widely used to estimate parameters of a model based on observed data. In particle physics, MLE enables both parameter extraction and hypothesis testing, particularly in searches for new phenomena. This section outlines the construction of likelihood functions, the MLE procedure, uncertainty estimation via profiling, and the  $CL_S$  method for setting exclusion limits.

### 5.2.1 Likelihood Function

The likelihood function  $\mathcal{L}(\theta; \mathbf{x})$  quantifies the probability of observing a dataset  $\mathbf{x}$  given a set of model parameters  $\theta$ . Typically, these parameters include:

- **Parameters of interest**, such as the signal strength  $\mu$
- **Nuisance parameters**, such as background rates or detector efficiencies

A nuisance parameter is not directly of interest but must be incorporated to obtain unbiased estimates of the parameters of interest.

**Poisson Likelihood with Auxiliary Constraints** Consider a scenario where the expected number of events is given by  $\mu s + b$ , with  $s$  representing the expected signal,  $b$  the background, and  $\mu$  the signal strength modifier. The number of observed events  $n$  follows a Poisson distribution:

$$\mathcal{L}(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)}$$

To incorporate uncertainties in  $b$ , auxiliary measurements can be included. For example, if  $b$  is constrained by a Gaussian distribution with mean  $\hat{b}$  and standard deviation  $\sigma_b$ :

$$\mathcal{L}(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \cdot \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{(b-\hat{b})^2}{2\sigma_b^2}}$$

### 5.2.2 Maximum Likelihood Estimation

MLE involves finding the parameter values that maximise the likelihood. To avoid large values during the optimisation, the log-likelihood function is used. Optimisers usually work by minimisation. Therefore, the negative log likelihood (NLL) is used in practice.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} -\log \mathcal{L}(\theta)$$

Suppose  $n = 10$  events are observed, with  $s = 5$  expected signal events, and a background estimate of  $b = 3 \pm 1$ . The likelihood is:

$$\mathcal{L}(\mu, b) = \frac{(\mu s + b)^{10}}{10!} e^{-(\mu s + b)} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(b-3)^2}{2}}$$

One can numerically maximise this expression with respect to  $\mu$  and  $b$  to obtain best-fit estimates.

### 5.2.3 Profiling and Uncertainty Estimation

To estimate uncertainties on parameters such as  $\mu$ , the profile likelihood method is employed. Here, nuisance parameters are profiled, meaning they are optimised for each fixed value of the parameter of interest:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$

where  $\hat{\hat{\theta}}$  are the profiled nuisance parameters at fixed  $\mu$ , and  $\hat{\mu}, \hat{\theta}$  are the unconditional MLE estimates.

The test statistic based on this ratio is:

$$q_\mu = -2 \log \lambda(\mu)$$

According to Wilks' theorem,  $q_\mu$  asymptotically follows a  $\chi^2$  distribution under the null hypothesis, enabling confidence interval construction.

### 5.2.4 The $CL_S$ Method

The  $CL_S$  method is a statistical procedure for setting upper limits in hypothesis testing, commonly used to determine exclusion bounds on signal models.

The method considers two hypotheses:

- $H_0$ : background-only hypothesis ( $\mu = 0$ )
- $H_1$ : signal+background hypothesis ( $\mu > 0$ )

For a given value of  $\mu$ , the observed test statistic  $q_\mu$  is compared against its distribution under both  $H_0$  and  $H_1$ . The corresponding  $p$ -values are:

$$p_\mu = P(q_\mu^\mu \geq q_\mu^{\text{obs}}), \quad p_0 = P(q_\mu^0 \geq q_\mu^{\text{obs}})$$

The  $CL_S$  value is then defined as:

$$CL_S = \frac{p_\mu}{1 - p_0}$$

If  $CL_S$  is less than  $1 - \alpha$ , the signal hypothesis is excluded at confidence level  $\alpha$ .

Suppose the test yields  $p_{\mu=1} = 0.03$  and  $p_0 = 0.5$ . Then:

$$CL_S = \frac{0.03}{1 - 0.5} = 0.06$$

This result implies that the signal hypothesis cannot be excluded at the 95% confidence level (since  $0.06 > 0.05$ ), but would be excluded at 90% confidence.

## Search for long-lived axions in $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ decays

This chapter describes the search for long-lived axion-like particles (ALPs) produced in Higgs boson decays, which subsequently decay into photons. This analysis has been published as part of [1]. The chapter begins with an overview of the analysis strategy (Section 6.1). This is followed by a description of the dataset, the simulated events, the event selection criteria, and the trigger efficiencies in Sections 6.2 and 6.3. As the standard ATLAS photon identification is inefficient for the targeted signal process, two custom photon classifiers are introduced in Section 6.4. The categorisation of events and the estimation of background contributions in each category are detailed in Sections 6.5 and 6.6. Finally, the sources of systematic uncertainties are discussed, and the methodology used to obtain the final results is presented in Sections 6.7 and 6.8. The final results are discussed in Section 6.9.

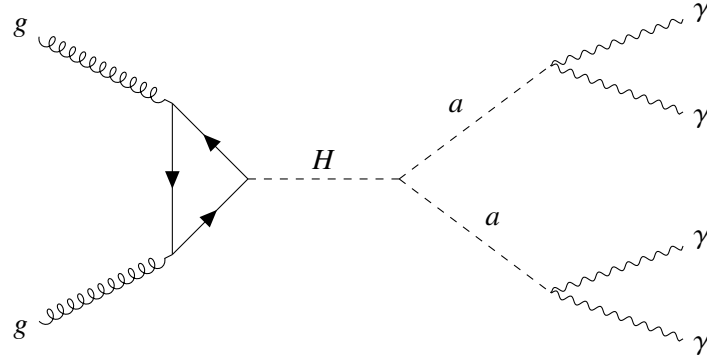
### 6.1 Analysis strategy

This analysis is developed to search for axion-like particles (ALPs) produced in Higgs boson decays, which subsequently decay into photons ( $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ ). A Feynman diagram of this process, where the Higgs boson is produced via gluon–gluon fusion, is shown in Fig. 6.1. If this process exists, it should manifest as an enhancement of events near the Higgs boson mass in the four-photon invariant mass spectrum.

The parameter space for ALP candidates in this process ranges from masses of  $100 \text{ MeV}^1$  up to half the mass of the Higgs boson. The coupling between ALPs and photons that can be probed in this search lies in the range  $10 \times 10^{-7} < C_{a\gamma\gamma} < 1$ . Depending on the ALP mass and its coupling to photons, the kinematic properties of the decay photons vary, and the analysis must be optimised for different regions of phase space.

At low ALP masses, the decay photons receive a large Lorentz boost, resulting in strong collimation of the photon pair. This can lead to overlapping electromagnetic showers in the calorimeter. A schematic illustration of two collimated photons is shown in Fig. 6.2. The reconstruction algorithms

<sup>1</sup> This lower limit is determined by the technically feasible range in the Monte Carlo generator.

Figure 6.1: Feynman diagram for  $gg \rightarrow H \rightarrow aa \rightarrow 4\gamma$ 

described in Section 4.2.3 are not designed for such topologies and are likely to reconstruct the photon pair as a single photon.

In cases of low ALP–photon coupling, the ALP can have a significant lifetime, resulting in displaced decays. Both collimation and displacement affect the event topology. If both photon pairs are reconstructed as single photons, the event appears to contain only two photons. If one photon pair is merged or one of the decay photons is not detected, the event will contain three photons. Thus, signal events may contain two, three, or four photons.

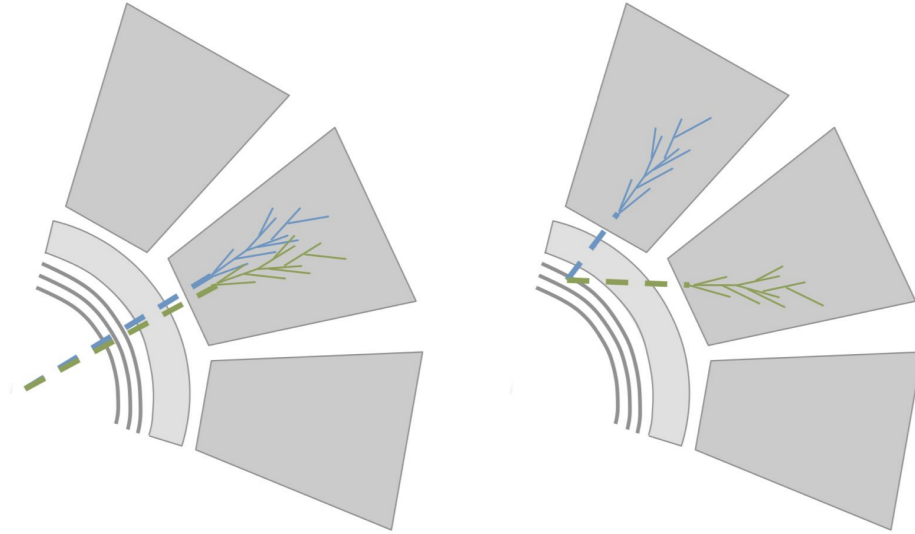


Figure 6.2: Sketch of special ALP decay topologies. On the left-hand side, two collimated photons with large Lorentz boost are shown. On the right-hand side, the ALP has a significant lifetime and decays with a displacement close to the calorimeter. Taken from [150].

This leads to several sources of background. The dominant background is the continuum production of two, three, or four photons. Additionally, there are backgrounds involving photons and jets, where a jet is misreconstructed as a photon.

As the production of high-energy photons becomes increasingly rare, background contributions are



larger in event topologies with fewer photons.

In the two-photon region, the dominant background is the continuum, but there is also a resonant contribution from Standard Model Higgs boson decays to two photons ( $H \rightarrow \gamma\gamma$ ).

The various simulated background processes for the two-photon case are shown in Fig. 6.3.

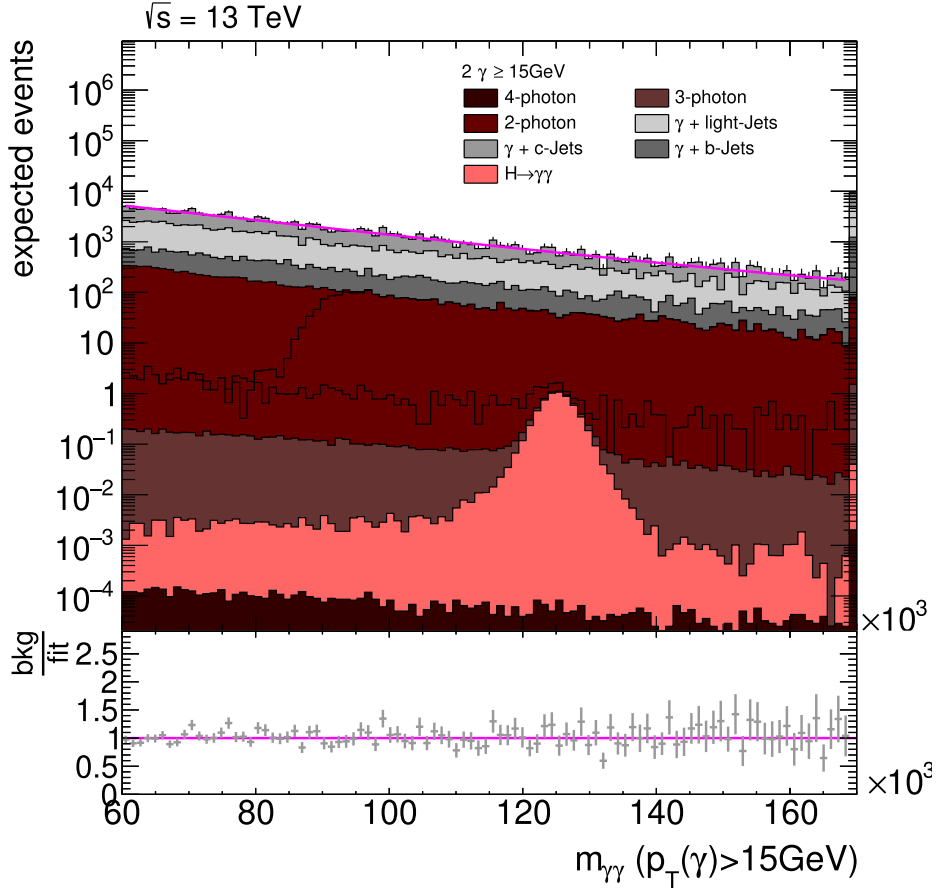


Figure 6.3: Simulated background processes in the two-photon case, requiring both photons to have at least 15 GeV. The brown filled histograms show prompt photon production, with darker colours indicating events with three and four photons. The grey histograms represent photon + jet processes, with darker shades for additional  $c$  and  $b$  jets. The light pink histogram shows the SM expectation from  $H \rightarrow \gamma\gamma$ . A simple exponential fit is shown as a pink line. The bottom panel shows the ratio of the combined background events to the fit.

To suppress both non-resonant and resonant backgrounds, machine learning techniques are employed to classify collimated photon pairs and single photons. This is discussed in more detail in Section 6.4. With the assignment of these photon types, each event can be assigned to one out of six categories. This is described in Section 6.5.

Once each event is uniquely assigned to a category, the invariant mass distribution of the relevant photons is constructed. The continuum background is estimated from data using a sideband fit, with the region around the Higgs boson mass blinded.

If at least three photons are reconstructed, the ALP mass can be determined, provided the correct photon combination is identified. To select the correct combination, additional machine learning

classifiers are used, as described in Section 6.4.5.

The reconstructed ALP mass can then serve as a further discriminant to reduce background for specific ALP mass–coupling hypotheses.

The simulated signal distribution for each ALP mass–coupling hypothesis is then fitted to data, together with the background estimate. In the absence of a significant excess, an upper limit on the signal cross section is set.

## 6.2 Data set and event simulation

This analysis uses data from  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 13$  TeV, recorded with the ATLAS detector between 2015 and 2018. The total integrated luminosity corresponds to  $(140.0 \pm 1.2) \text{ fb}^{-1}$  [151].

To ensure reliable data quality, only events recorded with all relevant subsystems of the ATLAS detector functioning nominally are considered. This requirement is enforced through the use of centrally maintained Good Run Lists (GRLs), provided by the ATLAS Data Quality group [152]. These GRLs specify valid data-taking periods and exclude runs affected by detector problems or calibration issues. The GRLs used in this analysis are listed in Table A.2.

All simulated samples are generated using the GEANT4-based [153] full ATLAS detector simulation [154] and are reconstructed with the standard ATLAS reconstruction framework.

Higgs boson production is modelled at next-to-leading order (NLO) using POWHEG-Box v2 [155–157]. The variation in selection efficiency between different production modes, such as gluon–gluon fusion (ggF) and vector boson fusion (VBF), is at the per cent level. Combined with the differences in production cross-sections, this effect is negligible; therefore, only Higgs bosons produced via ggF are simulated.<sup>2</sup>

In Section 6.2.1, the simulation and selection of the signal samples used in this analysis are described. The background samples are detailed analogously in Section 6.2.2.

### 6.2.1 Signal simulation

The simulated signal samples model Higgs boson production using POWHEG Box with the AZNLO tune [161] and the NNPDF30 parton distribution function (PDF) set [162]. The Higgs boson decays into a pair of axion-like particles (ALPs), which subsequently decay into photons. This decay chain is simulated with PYTHIA 8.212 [163], using a generic two-Higgs-doublet model (2HDM) implementation [164].

To cover the relevant phase space efficiently, samples are generated in discrete mass points with the following step sizes:

- 100 MeV steps from 0.1 GeV to 0.5 GeV,
- 0.5 GeV steps from 0.5 GeV to 5 GeV,
- 1 GeV steps for masses above 5 GeV.

<sup>2</sup> The  $H \rightarrow \gamma\gamma$  background and the results in Section 6.9 are scaled to the total Higgs boson production cross-section as predicted by next-to-next-to-next-to-leading order (NNNLO) QCD calculations, including NLO electroweak corrections [48, 158–160].

For each mass point, signal samples are produced for four values of the photon coupling  $C_{a\gamma\gamma}$ , specifically 1, 0.01,  $5 \times 10^{-4}$  and  $1 \times 10^{-5}$ . Due to their longer lifetimes, ALPs with smaller couplings may decay outside the detector acceptance. To avoid simulating events with negligible efficiency, the lowest two coupling values are only simulated for ALP masses  $m_a \geq 10 \text{ GeV}$ .

Each signal sample consists of 20 000 fully simulated events. Sample names follow the convention:

`mc16_13TeV.[DSID].PowhegPy8EG_AZNLO_ggH_H125_aXXaXX_4y_CyyY.merge.AOD.[TAG],`

where **XX** denotes the ALP mass and **Y** denotes the coupling value. The **[DSID]** is the unique dataset ID, and **[TAG]** summarises the software settings used during reconstruction.

In addition to these samples used in the final result, a number of additional signal samples were privately generated to support development and testing.

### Signal with VBF production

To evaluate the dependence of the signal efficiency on the Higgs boson production mode, samples were generated where the Higgs originates from vector boson fusion (VBF). Mass points between 0.5 GeV and 60 GeV were produced with 1 000 events per point. The selection efficiency and kinematic distributions were compared to those from gluon–gluon fusion (ggF), and found to differ by less than 1%. As a result, VBF samples were not included in the main analysis.

### Multimass sample with varying $C_{a\gamma\gamma}$

An alternative simulation strategy was explored using a single sample with varying  $C_{a\gamma\gamma}$  and a flat lifetime distribution. To ensure meaningful statistics, a generator-level filter required both ALPs to decay within the detector acceptance—specifically, with decay lengths  $l_{\text{decay}} < 1\,970 \text{ mm}$ , corresponding approximately to the end of the electromagnetic calorimeter. This preselection avoids unnecessary detector simulation and reduces computational cost. The filter efficiency varies between 40 % and 50 % depending on the mass point. Per point, 1 000 filtered events were retained for use.

Originally, this sample was intended for use with reweighting techniques to emulate different lifetime distributions corresponding to specific  $C_{a\gamma\gamma}$  values. However, the presence of two independently decaying ALPs complicates the interpretation: the event topology depends on the joint distribution of both decay lengths. As a result, a one-dimensional reweighting in lifetime is insufficient to capture the full distribution. A two-dimensional reweighting scheme in  $l_{\text{decay},1}$  and  $l_{\text{decay},2}$  would be required, but this demands a dense population of the 2D phase space, which was impractical due to simulation cost.

Consequently, the decision was made to simulate discrete  $C_{a\gamma\gamma}$  values and interpolate results between them.

The filtered multimass samples are used exclusively for training neural network classifiers, ensuring that no overlap occurs between training and evaluation datasets.

## 6.2.2 Background simulation

Continuum  $\gamma\gamma$  production (excluding interference with an  $X \rightarrow \gamma\gamma$  signal) is simulated with SHERPA 2.2.4 [165], using its default underlying-event tune. Matrix elements are calculated at next-to-leading order (NLO) in the strong coupling  $\alpha_S$  for up to one additional parton and at leading order (LO) for two

or three extra partons; the samples are merged with the Sherpa parton shower in the ME+PS@NLO scheme [166]. The NNPDF3.0 NNLO PDF set is employed.

During the early stages of the analysis, additional SM 2-, 3- and 4-photon samples (generated with SHERPA 2.2.8 and the same PDF set) were used for cross-checks; all are available at NLO accuracy in  $\alpha_S$ .

Higgs-boson backgrounds ( $H \rightarrow \gamma\gamma$ ) from ggF, VBF, VH and  $t\bar{t}H$  production are generated with POWHEGBOX v2 [155–157] interfaced to PYTHIA 8 [163] for parton showering. The PDF4LHC15 NNLO set is used together with the AZNLO tune [161].

All MC backgrounds except for the Higgs contribution serve only to validate the data-driven estimate and are not propagated into the final statistical interpretation.

To assess systematic uncertainties related to displaced vertices (see Appendix A.2), an additional dijet sample is used. It is produced with PYTHIA 8.230 at LO, matched to the parton shower, and employs the NNPDF2.3<sub>LO</sub> PDF set together with the A14 tune [167].

## 6.3 Event selection

The selection of signal-like events is performed in several steps. During data taking, real-time event selection is handled by a set of trigger algorithms, as described in Section 4.2.2. To further reduce the background, preselection criteria are applied to physics objects, primarily based on their kinematic properties. In order to categorise each event, the classification of photons is required. These steps are described in the following subsections.

### 6.3.1 Trigger studies

To determine suitable trigger configurations, various trigger chains are evaluated on simulated signal events. Since the ATLAS trigger menu has changed slightly across data-taking years, dedicated trigger selections are applied for each year.

Additionally, the kinematic properties of the signal can vary significantly depending on the ALP mass and coupling, necessitating different trigger strategies across the phase space.

The combined efficiency of all considered triggers (listed in ??) ranges from approximately 40 % to 78 %, but drops sharply for small ALP masses and couplings. In these cases, ALPs are more likely to decay with a significant displacement, possibly outside the sensitive volume of the detector, resulting in a drop in both acceptance and trigger efficiency.

To isolate the effects of trigger efficiency from overall acceptance, trigger performance is evaluated both before and after applying the preselection criteria defined in Section 6.3.2. In addition to the combined efficiency, the standalone performance of individual trigger chains is studied.

It is found that similar efficiency can be achieved using only one of two trigger chains, depending on the ALP mass. For ALP masses below 5 GeV, the most effective trigger is: HLT\_g35\_loose\_g25\_loose (or HLT\_g35\_medium\_g25\_medium\_L12EM20VH in 2017/2018).

For masses above 5 GeV, better efficiency is achieved using HLT\_2g22\_tight (or HLT\_2g22\_tight\_L12EM15VH in 2017/2018).

Since some of the considered trigger chains do not have centrally provided systematic uncertainties, assigning such uncertainties becomes more challenging, especially when multiple triggers are used simultaneously. Therefore, only the most efficient single trigger is applied in each region of phase

Table 6.1: Overview of used background samples with full names of the MC background sets

| Short         | Sample name                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ggF H         | mc16_13TeV.343981.PowhegPythia8EvtGen_NNLOPS_nnlo_30_ggH125_gamgam.recon.<br>AOD.e5607_s3126_r9364<br>mc16_13TeV.343981.PowhegPythia8EvtGen_NNLOPS_nnlo_30_ggH125_gamgam.recon.<br>AOD.e5607_s3126_r10201<br>mc16_13TeV.343981.PowhegPythia8EvtGen_NNLOPS_nnlo_30_ggH125_gamgam.recon.<br>AOD.e5607_s3126_r10724                                                                                                                                                                                                                                                                                                                                                                           |
| VBF H         | mc16_13TeV.346214.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_gamgam.recon.<br>AOD.e6970_s3126_r9364<br>mc16_13TeV.346214.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_gamgam.recon.<br>AOD.e6970_s3126_r10201<br>mc16_13TeV.346214.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_gamgam.recon.<br>AOD.e6970_s3126_r10724                                                                                                                                                                                                                                                                                                                                                                              |
| ZH            | mc16_13TeV.345319.PowhegPythia8EvtGen_NNPDF30_AZNLO_ZH125J_Hyy_Zincl_MINLO.recon.<br>AOD.e5743_s3126_r9364<br>mc16_13TeV.345319.PowhegPythia8EvtGen_NNPDF30_AZNLO_ZH125J_Hyy_Zincl_MINLO.recon.<br>AOD.e5743_s3126_r10201<br>mc16_13TeV.345319.PowhegPythia8EvtGen_NNPDF30_AZNLO_ZH125J_Hyy_Zincl_MINLO.recon.<br>AOD.e5743_s3126_r10724                                                                                                                                                                                                                                                                                                                                                   |
| WH            | mc16_13TeV.345317.PowhegPythia8EvtGen_NNPDF30_AZNLO_WmH125J_Hyy_Wincl_MINLO.recon.<br>AOD.e5734_s3126_r9364<br>mc16_13TeV.345317.PowhegPythia8EvtGen_NNPDF30_AZNLO_WmH125J_Hyy_Wincl_MINLO.recon.<br>AOD.e5734_s3126_r10201<br>mc16_13TeV.345317.PowhegPythia8EvtGen_NNPDF30_AZNLO_WmH125J_Hyy_Wincl_MINLO.recon.<br>AOD.e5734_s3126_r10724<br>mc16_13TeV.345318.PowhegPythia8EvtGen_NNPDF30_AZNLO_WpH125J_Hyy_Wincl_MINLO.recon.<br>AOD.e5734_s3126_r9364<br>mc16_13TeV.345318.PowhegPythia8EvtGen_NNPDF30_AZNLO_WpH125J_Hyy_Wincl_MINLO.recon.<br>AOD.e5734_s3126_r10201<br>mc16_13TeV.345318.PowhegPythia8EvtGen_NNPDF30_AZNLO_WpH125J_Hyy_Wincl_MINLO.recon.<br>AOD.e5734_s3126_r10724 |
| ttH           | mc16_13TeV.346525.PowhegPythia8EvtGen_A14NNPDF23_NNPDF30ME_ttH125_gamgam.recon.<br>AOD.e7488_s3126_r9364<br>mc16_13TeV.346525.PowhegPythia8EvtGen_A14NNPDF23_NNPDF30ME_ttH125_gamgam.recon.<br>AOD.e7488_s3126_r10201<br>mc16_13TeV.346525.PowhegPythia8EvtGen_A14NNPDF23_NNPDF30ME_ttH125_gamgam.recon.<br>AOD.e7488_s3126_r10724                                                                                                                                                                                                                                                                                                                                                         |
| SM 2 $\gamma$ | mc16_13TeV.364352.Sherpa_224_NNPDF30NNLO_Diphoton_myy_90_175.merge.<br>AOD.e6452_e5984_a875_r10201_r10210<br>mc16_13TeV.364351.Sherpa_224_NNPDF30NNLO_Diphoton_myy_50_90.merge.<br>AOD.e6452_e5984_a875_r10724_r10726                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| SM 3 $\gamma$ | mc16_13TeV.700007.Sh_228_yyy_01NLO.merge.<br>AOD.e7999_e7400_s3126_r10724_r10726                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| SM 4 $\gamma$ | mc16_13TeV.700009.Sh_228_yyyy.merge.<br>AOD.e7999_e7400_s3126_r10724_r10726                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| dijet         | mc16_13TeV.361024.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4W.deriv.DAOD_RPVLL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

space to ensure proper uncertainty treatment. The resulting efficiency loss—typically only a few percent—is considered acceptable.

The trigger efficiencies (as a function of the analysis categories introduced in Section 6.5) are shown for selected ALP masses in Fig. 6.4. Different trigger chains are indicated by their line styles: solid lines represent the logical **OR** of all photon triggers, dashed lines show results for HLT\_2g22\_tight, and dash-dotted lines correspond to HLT\_g35\_loose\_g25\_loose.

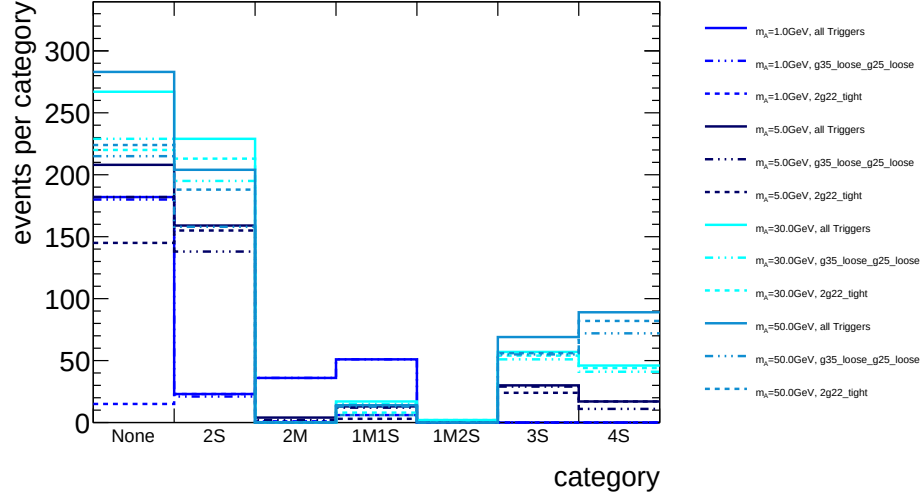


Figure 6.4: Trigger efficiencies for  $m_a = 1$  GeV (blue),  $m_a = 5$  GeV (dark blue),  $m_a = 30$  GeV (teal), and  $m_a = 50$  GeV (dark teal). Solid lines indicate the logical AND of all photon triggers, dashed lines show HLT\_2g22\_tight, and dash-dotted lines indicate HLT\_g35\_loose\_g25\_loose.

It can be seen that the efficiency loss between using all triggers combined and only the selected trigger is small. In particular, for  $m_a = 1$  GeV, the curves are identical in the 2M and 1M1S categories.

### 6.3.2 Preselections

Selected events must satisfy the trigger conditions and contain two, three, or four reconstructed photons.

On the object level, additional kinematic and geometric requirements are imposed. Photons must have a transverse momentum  $p_T$  greater than 15 GeV and lie within the pseudorapidity range  $|\eta| < 2.37$ , excluding the transition region between the barrel and endcap calorimeters ( $1.37 \leq |\eta| \leq 1.52$ ).

To ensure operation in the plateau region of the trigger efficiency curve, the leading and subleading photons are required to exceed the respective  $p_T$  thresholds of the chosen HLT trigger.

Photons originating from the signal process are typically isolated in the calorimeter, whereas background photons (e.g., from jets) are often surrounded by additional energy deposits. To suppress such non-isolated backgrounds, a calorimeter-based isolation requirement is applied:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2 \quad (6.1)$$

A photon is considered isolated if the scalar sum of the transverse momenta of tracks in this cone is

less than 5% of the photon's transverse momentum:

$$\frac{p_T^{\text{cone20}}}{p_T^\gamma} < 0.05 \quad (6.2)$$

A summary of all preselection criteria is provided in Table 6.2.

Table 6.2: Analysis pre-selections by  $m_a$  category

| selection                    | $m_a \leq 5 \text{ GeV}$                         | $m_a > 5 \text{ GeV}$     |
|------------------------------|--------------------------------------------------|---------------------------|
| trigger 2015/2016            | HLT_g35_loose                                    | HLT_2g22                  |
|                              | _g25_loose                                       | _tight                    |
| trigger 2017/2018            | HLT_g35_medium_g25                               | HLT_2g22                  |
|                              | _medium_L12EM20VH                                | _tight_L12EM15VHI         |
| number of photons            | 2 or 3 or 4                                      |                           |
| $p_T(\gamma_1)$              | $p_T \geq 35 \text{ GeV}$                        | $p_T \geq 22 \text{ GeV}$ |
| $p_T(\gamma_2)$              | $p_T \geq 25 \text{ GeV}$                        | $p_T \geq 22 \text{ GeV}$ |
| $p_T(\gamma_{\text{other}})$ | $p_T \geq 15 \text{ GeV}$                        |                           |
| pseudo-rapidity              | $ \eta  \leq 2.37$ and $!(1.37 <  \eta  < 1.52)$ |                           |
| isolation                    | $\frac{p_T^{\text{cone20}}}{p_T^\gamma} < 0.05$  |                           |

## 6.4 Photon classification

As mentioned previously, the standard ATLAS photon identification is not designed to efficiently recognise merged photon signatures. Consequently, many such signatures are either misidentified or not reconstructed at all. To address this, a dedicated machine learning-based approach is employed.

Because the standard identification rejects a large fraction of merged photon candidates, a two-step classification strategy is adopted. In the first step, photons are separated from background objects such as hadrons misidentified as photons. This step corresponds to a standard photon identification task. In the second step, the photons passing this initial selection are further classified as either single-photon signatures or merged-photon signatures.

Both classification stages are implemented using neural networks based on the TMVA package [168]. The architecture of these neural networks is described in Sections 6.4.2 and 6.4.3, and the electromagnetic shower variables used as input features are detailed in Section 6.4.1.

### 6.4.1 Training data

Photons interacting in the electromagnetic calorimeter produce electromagnetic showers. The resulting energy deposits in the calorimeter cells are measured, and from these, more abstract quantities—known as shower shape variables—are computed. These variables exhibit different distributions for real photons compared to other objects (e.g., hadrons or jets), and are thus used to discriminate real photons from background. Based on these variables, each photon candidate is assigned an identification (ID) level, such as Loose or Tight, indicating the likelihood of being a genuine photon.

The set of shower shape variables used in this analysis is listed in Table 6.3. This table, along with further details on photon ID and its associated efficiencies, is also described in [169].

Three datasets are used for the training. The first contains merged photons from simulated  $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$  events, the second consists of single photons from simulated  $H \rightarrow \gamma\gamma$  events, and the third contains other photon candidates—typically not originating from prompt photons—labelled as ‘fake’.

Simulated photons are matched to generator-level (truth) photons such that each single reconstructed photon corresponds to exactly one truth photon, and each merged photon corresponds to two truth photons.

Fake photon candidates are selected from data by requiring events that either pass the analysis triggers or dedicated single-photon triggers. Events must contain at least two photons with an invariant mass above 60 GeV, excluding the Higgs mass window between 115 GeV and 135 GeV. Each photon is required to satisfy  $p_T > 15$  GeV.

To ensure consistent  $p_T$ – $\eta$  distributions across all training samples, a two-dimensional histogram is created for the dataset with the fewest training photons (i.e., the fewest input feature vectors as defined in Section 5.1.1). The feature vectors from the other datasets are then filled into histograms with identical binning, but with the restriction that no bin exceeds the population of the corresponding bin in the reference histogram. Any excess entries are discarded. As a result, all datasets are matched in both event count and  $p_T$ – $\eta$  distribution, eliminating the need for reweighting and avoiding potential biases. This bin-wise rejection is applied consistently across all datasets using only the bin counts of the reference sample, ensuring statistical equivalence in phase space coverage.

Each dataset is then randomly split into training, validation, and test subsets. An overview of the number of feature vectors in each subset is given in Table 6.4.

Differences in the shower shape variables between the three sample types are illustrated in Fig. 6.5. The variables  $f_1$ ,  $R_\Phi$ ,  $\Delta E$ , and  $E_{\text{ratio}}$  are shown for fake photons (grey), single photons (blue), and merged photons (purple). Notably, fake photons show distinct distributions compared to both single and merged photons.

#### 6.4.2 Classifier 1: single + merged vs fake

The first classifier is designed to distinguish fake photons from genuine ones, i.e. both single and merged photons. It is implemented as a neural network using the TMVA framework [168]. The architecture consists of an input layer with one node per input variable, followed by two fully connected hidden layers with 7 and 5 nodes, respectively, and a single output node. All nodes employ a sigmoid activation function.

Fake photons are treated as background and assigned a target output of 0, while both single and merged photons are treated as signal and assigned a target output of 1. The network is trained on the ten shower shape variables described previously, with all inputs scaled to the same numerical range. This standardisation improves both training stability and performance.

Training is stopped after 600 epochs, at which point no further improvement is observed. The output distribution of the trained classifier is shown in Fig. 6.6. The output for fake photons (yellow line) peaks near zero, while both single (light blue) and merged (dark blue) photons yield outputs clustered near one. The close agreement between the single and merged photon output distributions demonstrates that the classifier does not introduce any significant bias between these two signal types. Overall, a clear separation between real and fake photon candidates is achieved.



Table 6.3: Explanation of shower shape variables. The last two columns indicate if the variable is used for the Tight or Loose photon ID. Recreated from [169].

| Category         | Description                                                                                                                                         | Name               | Tight | Loose |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-------|-------|
| Acceptance       | $ \eta  < 2.37$ , excluding the region $1.37 \leq  \eta  < 1.52$                                                                                    | –                  | ✓     | ✓     |
| Hadronic leakage | Ratio of $E_T$ in the first layer of the hadronic calorimeter to $E_T$ of the EM cluster (used for $ \eta  < 0.8$ or $ \eta  > 1.52$ )              | $R_{\text{had}_1}$ | ✓     | ✓     |
|                  | Ratio of $E_T$ in the hadronic calorimeter to $E_T$ of the EM cluster (used for $0.8 <  \eta  < 1.37$ )                                             | $R_{\text{had}}$   | ✓     | ✓     |
| EM middle layer  | Energy in $3 \times 7 \eta \times \phi$ cells over energy in $7 \times 7$ cells centered on cluster                                                 | $R_\eta$           | ✓     | ✓     |
|                  | Lateral width:<br>$\sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i}\right)^2}$ calculated in a $3 \times 5$ window | $w_{\eta^2}$       | ✓     | ✓     |
|                  | Energy in $3 \times 3 \eta \times \phi$ cells over energy in $3 \times 7$ cells                                                                     | $R_\phi$           | ✓     |       |
| EM strip layer   | Lateral width:<br>$\sqrt{\frac{\sum E_i (i - i_{\text{max}})^2}{\sum E_i}},$ where $i$ runs over $3 \times 2$ strips                                | $w_s^3$            | ✓     |       |
|                  | Total lateral width over $20 \times 2 \eta \times \phi$ strip window                                                                                | $w_s^{\text{tot}}$ | ✓     |       |
|                  | Energy outside the 3 central strips but within 7 strips divided by central energy                                                                   | $f_{\text{side}}$  | ✓     |       |
|                  | Difference between energy of the second maximum and minimum between maxima                                                                          | $\Delta E_s$       | ✓     |       |
|                  | Ratio of energy difference between max and secondary max to their sum                                                                               | $E_{\text{ratio}}$ | ✓     |       |
|                  | Ratio of energy in first layer to total EM cluster energy                                                                                           | $f_1$              | ✓     |       |

Table 6.4: Training samples and number of feature vectors in each subset.

| Sample | Source                                        | # Training | # Validation | # Test |
|--------|-----------------------------------------------|------------|--------------|--------|
| single | $H \rightarrow \gamma\gamma$ , prompt photons | 20000      | 2000         | 2000   |
| merged | $H \rightarrow aa$ , flat lifetime            | 20000      | 2000         | 2000   |
| fake   | Data                                          | 20000      | 2000         | 2000   |

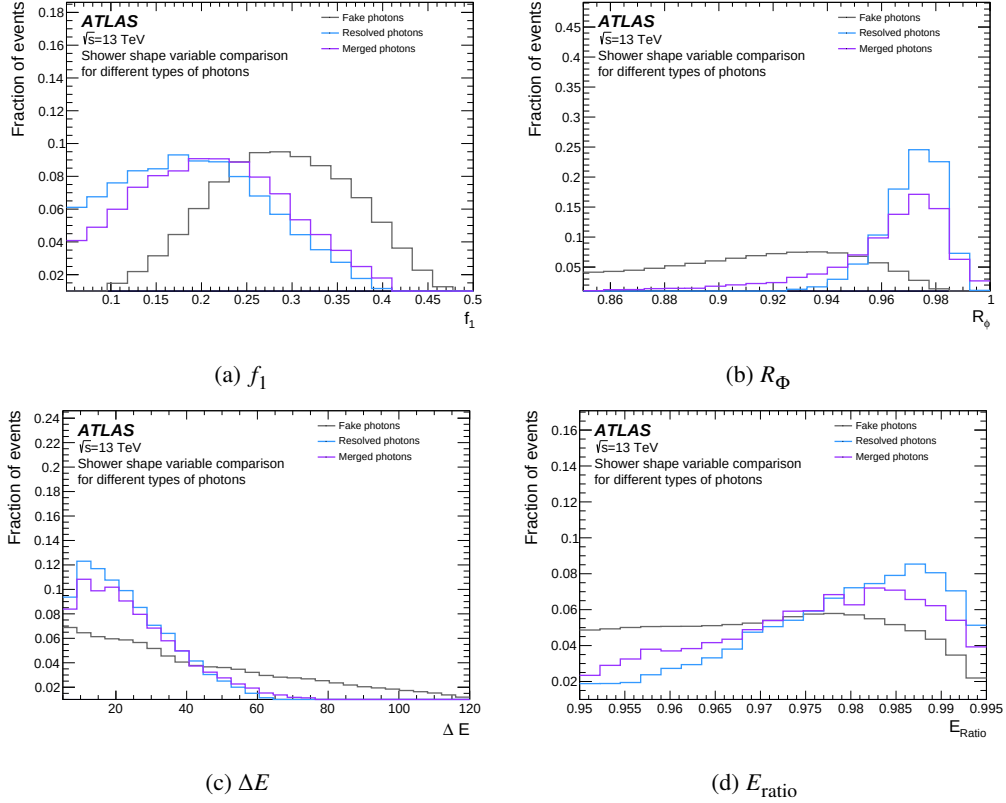


Figure 6.5: Shower shape variables for fake photons (grey), single photons (blue), and merged photons (purple). Taken from [1].

### 6.4.3 Classifier 2: single vs merged

The second classifier is trained to distinguish between single and merged photon signatures. It uses the same neural network architecture as the first classifier: 10 input nodes corresponding to the shower shape variables, two fully connected hidden layers with 7 and 5 nodes, and a single output node. All nodes use a sigmoid activation function.

As in the previous case, the ten shower shape variables are scaled to the same numerical range prior to training to ensure stable learning and balanced contributions from all inputs. In this classifier, single photons are treated as background and assigned a target output value of 0, while merged photons are considered signal with a target value of 1. Training is stopped after 600 epochs due to a lack of further improvement in performance.

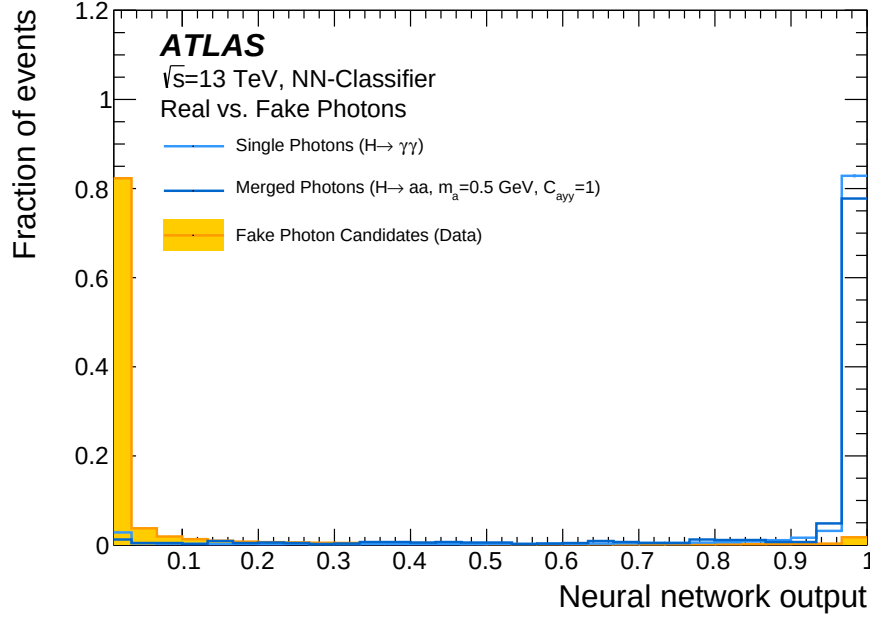


Figure 6.6: Output of the classifier distinguishing fake photons (yellow) from single photons (light blue) and merged photons (dark blue). Taken from [1].

The classifier outputs for single photons from a  $H \rightarrow \gamma\gamma$  sample and merged photons from two distinct  $H \rightarrow aa$  samples are shown in Fig. 6.7. The merged photon outputs (blue histograms) peak near one, reflecting the classifier’s high confidence in identifying these as merged signatures. In contrast, the distribution for single photons (yellow) is skewed toward zero, indicating successful separation.

However, compared to the first classifier, the output distributions for single and merged photons exhibit more overlap, indicating somewhat reduced discriminating power. Despite this, the classifier achieves a single-photon rejection rate of approximately 70 % and a selection efficiency for merged photons of around 80 %, which is sufficient for the purposes of the analysis. Additionally, the similarity between the outputs for the two merged photon samples demonstrates that the classifier behaves consistently across different ALP kinematic configurations.

It is important to note that none of the photons used during the training of the classifiers are used in the subsequent analysis. This is ensured by applying additional selection criteria to the fake photon sample and by using completely separate simulated datasets for the single and merged photon training. As a result, the training and evaluation phases remain statistically independent, avoiding any risk of bias in the analysis.

#### 6.4.4 Two photon energy reconstruction

Collinear photon signatures that are not reconstructed as two separate photon candidates typically exhibit poor energy resolution. In such cases, a significant portion of the energy from one of the photons is assigned to the main reconstructed photon candidate, while the remaining fraction is

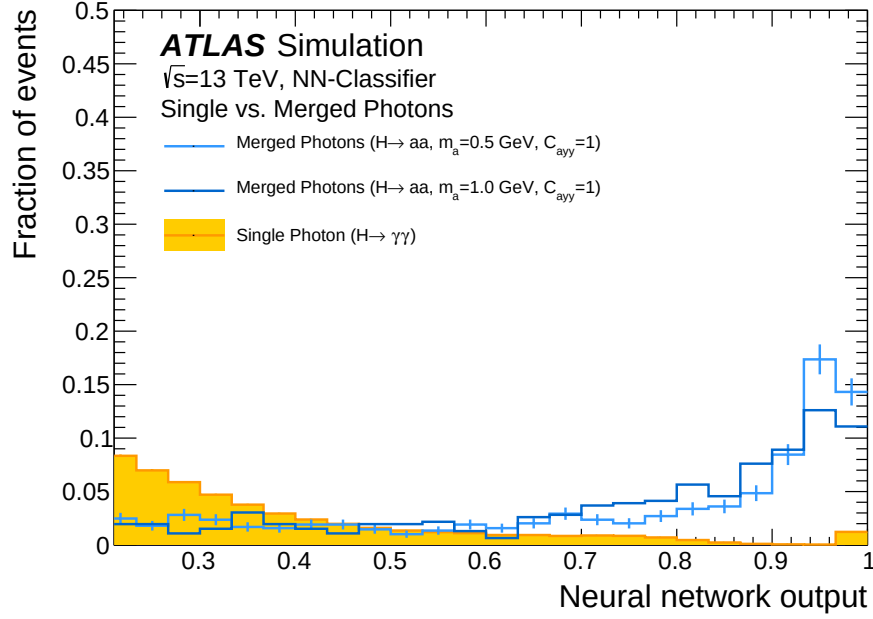


Figure 6.7: Output of the classifier distinguishing between single photons (yellow) and merged photons (blue). Two merged photon distributions are shown from different ALP samples. Taken from [1].

deposited in nearby calorimeter topo-clusters that are not associated with that candidate. This effect is a primary reason why calorimeter-based isolation criteria are not applied in this analysis; instead, only track-based isolation requirements are used to maintain robustness against this topology.

The degraded energy resolution from these merged signatures introduces substantial tails in the reconstructed Higgs boson mass distribution, particularly in the two-photon categories, as shown in Fig. 6.8. To address this issue, an improved energy reconstruction method is employed: energy deposits from nearby topo-clusters within a cone of radius  $\Delta R = 0.2$  around each photon candidate are added to its total energy. The impact of this correction on the Higgs boson mass reconstruction is illustrated in Fig. 6.8, where a notable suppression of the mass distribution tails is observed—most prominently in the two-single category.

This improvement is expected, as merged photon signatures (typically identified via shower shape variables) concentrate most of their energy in a single calorimeter cluster, whereas single photons may benefit more from the inclusion of adjacent energy deposits.

#### 6.4.5 ALP reconstruction

In case the ALP decay photons are not merged, the ALP could be reconstructed by selecting the two decay photons of the same ALP. This could be possible for events with three or four photons. To select the correct combination of photons again Neural Networks are used.

This time all the training data comes from simulated  $H \rightarrow aa$  samples with multiple masses but flat lifetime distribution. Reconstructed kinematic properties, i.e. the differences in  $p_T$ ,  $\eta$  and  $\phi$  as well as the invariant two-photon masses are used as input variables. In the three photon case

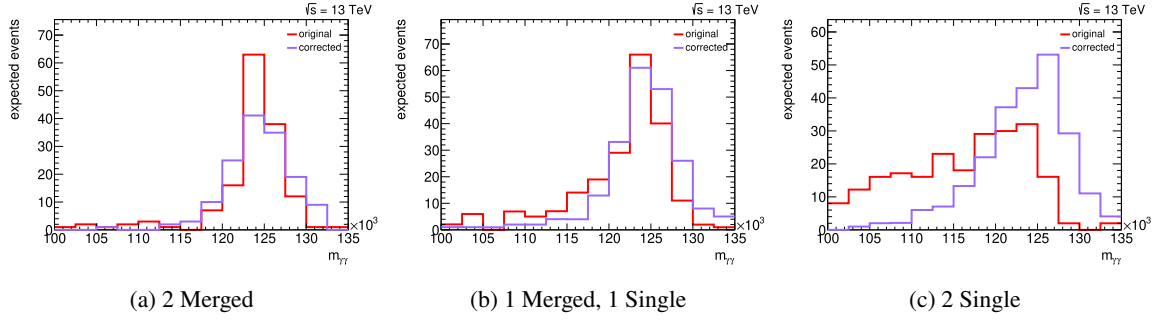


Figure 6.8: Reconstructed Higgs boson mass distributions for  $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$  events with  $m_a < 3$  GeV in the two-photon categories, comparing the nominal photon energy reconstruction to the corrected method including nearby topo-cluster energy. Left: two merged photon signatures. Middle: one merged and one single photon. Right: two single photons. Taken from [150]

these are  $(\Delta p_T^{12}, \Delta p_T^{23}, \Delta p_T^{13})$ ,  $(\Delta \eta^{12}, \Delta \eta^{23}, \Delta \eta^{13})$ ,  $(\Delta \phi^{12}, \Delta \phi^{23}, \Delta \phi^{13})$  and  $(m_{inv}^{12}, m_{inv}^{23}, m_{inv}^{13})$ . For four photons there are accordingly more variables.

All possible combinations of photons are build. If the combination is correct, the combination is used as a signal feature and an output value of one is assigned. If the combination is not correct, it is used as background feature and an output value of zero is assigned.

For three photons, this means that the photon indexes 1 and 2 are assigned to the photons coming from the same ALP decay and the other gets index 3. In other words,  $\Delta p_T^{12}$ ,  $\Delta \eta^{12}$ ,  $\Delta \phi^{12}$  and  $m_{inv}^{12}$  correspond to the correct combination, while all others describe wrong combinations.

To evaluate the classifier on data, all possible photon combination are build for the three and four photon category events. In case of a three photon event, this would be the tuples  $(\gamma_1, \gamma_2, \gamma_3)$ ,  $(\gamma_2, \gamma_3, \gamma_1)$  and  $(\gamma_1, \gamma_3, \gamma_2)$ . For each of these permutations, the classifier is evaluated. It is not necessary to find a selection criteria based on a minimal classifier output, but the combination with the highest classifier output is considered the correct combination.

If for example the combination  $(\gamma_2, \gamma_3, \gamma_1)$  gets the highest output value, it is known that the first two used photons would be the ones coming from the same ALP. In this example case, these would be photon two and photon three. Now the fourvectors of the two photons can be recombined to the fourvector of the ALP and thus, interesting ALP parameters like the mass can be computed.

This mass will be referred to as the reconstructed ALP mass in this chapter.

In Fig. 6.9 the invariant masses of the reconstructed ALPs from the combination with highest classifier output for three and four photons are shown. The blue histogram is created from the best combinations of a  $m_a = 5$  GeV sample with different couplings. The violet shows reconstructed ALP masses for  $m_a = 20$  GeV samples and the magenta histograms for  $m_a = 40$  GeV. The flat yellow histograms show the invariant mass distributions for the best photon combination in simulated multiphoton samples. It can be seen, that there is no bias towards a specific mass introduced by the classifiers.

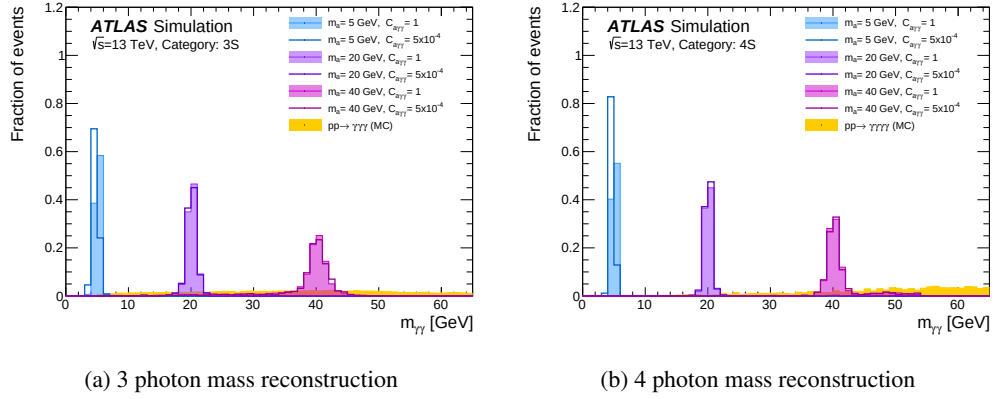


Figure 6.9: Reconstructed ALP mass for different simulated samples. The blue histograms show  $m_a = 5$  GeV samples with different coupling, violet  $m_a = 20$  GeV and magenta  $m_a = 40$  GeV samples. The shaded histograms show samples with coupling  $C_{a\gamma\gamma} = 1$  and the line histograms samples with coupling  $C_{a\gamma\gamma} = 5 \times 10^{-4}$ . The yellow histogram shows the invariant mass of the best photon combination for simulated multiphoton samples. [1]

## 6.5 Event categorization

There are six possible final-state configurations expected in the analysis. These configurations form the basis for categorising events, with each event assigned to exactly one of the categories listed in ??.

The categorisation begins by applying the photon classifiers described in Section 6.4 to each photon in the event. A photon is considered **merged** if it satisfies two conditions: its output score exceeds 0.98 in the first classifier (which separates real from fake photons), and is greater than 0.5 in the second classifier (which distinguishes single from merged photon signatures).

Once all photons in an event have been classified, the event is assigned to the first category whose criteria it satisfies, following the order below:

1. **Four Single Photons (4S):** The event contains exactly four photons passing the Loose ID criteria, with at least one also satisfying the Tight ID requirement (see Section 4.2.3).
2. **One Merged and Two Single Photons (1M2S):** If not classified as 4S, the event is checked for exactly one merged photon and exactly two additional Loose ID photons.
3. **Three Single Photons (3S):** Events with exactly three Loose ID photons, of which at least two also pass Tight ID, fall into this category.
4. **Two Merged Photons (2M):** Events with exactly two merged photons and no additional Loose ID or Tight ID photons.
5. **One Merged and One Single Photon (1M1S):** One merged photon and exactly one Loose ID photon are required.
6. **Two Single Photons (2S):** Exactly two photons must be present, both of which must pass Tight ID.

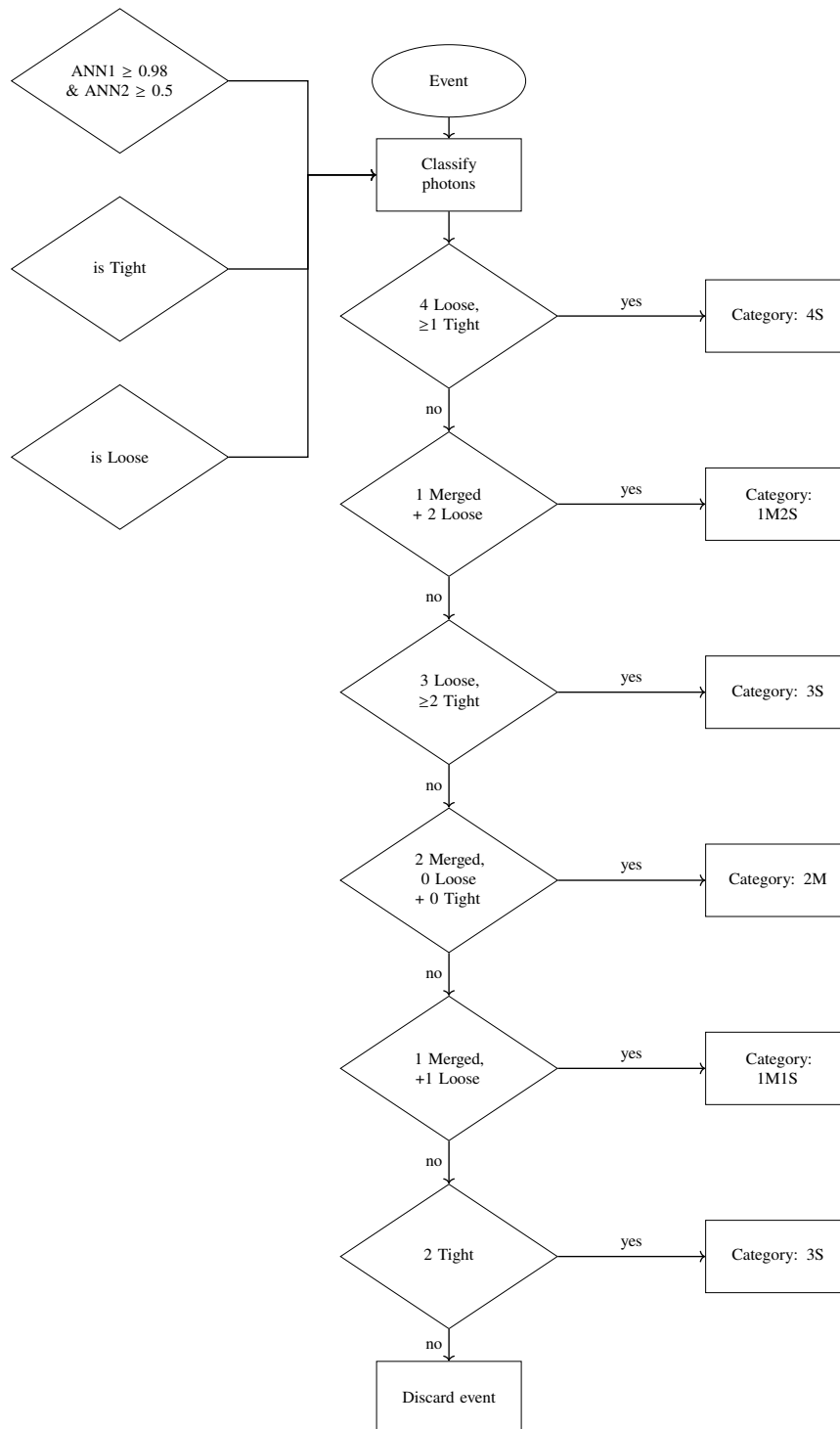


Figure 6.10: Schematic overview of the event categorisation process. Each photon in the event is first evaluated for classification as merged (based on ANN outputs), tight, or loose. The event is then sequentially checked against the six predefined categories: 4S, 1M2S, 3S, 2M, 1M1S, and 2S. If none of the category criteria are satisfied, the event is discarded.

To visualise the categorisation logic described above, a schematic overview of the decision sequence is shown in Section 6.5. Each event is first subjected to a classification of its photons into merged, tight, or loose types, using the outputs of the trained artificial neural networks. Based on these classifications, events are sequentially checked against the six predefined categories (4S, 1M2S, 3S, 2M, 1M1S, 2S). The first matching condition determines the event category; if none apply, the event is discarded.

The categorisation efficiency for selected ALP mass and coupling combinations is shown in Fig. 6.11. Different colours represent different ALP masses, while line styles distinguish between coupling strengths. Specifically, blue histograms correspond to  $m_a = 1$  GeV, dark blue to  $m_a = 5$  GeV, teal to  $m_a = 30$  GeV, and dark teal to  $m_a = 50$  GeV. Solid lines indicate  $C_{a\gamma\gamma} = 1$ , dashed lines  $C_{a\gamma\gamma} = 0.01$ , and dash-dotted lines  $C_{a\gamma\gamma} = 5 \times 10^{-4}$ .

For lower masses, events are more likely to fall into the two-photon categories, whereas higher masses tend to populate the three- and four-photon categories. A significant fraction of events remains unclassified and is therefore discarded. This loss is primarily due to the strict selection criteria applied during categorisation. Relaxing these criteria could improve signal efficiency, but would also reduce the rejection power against background events.

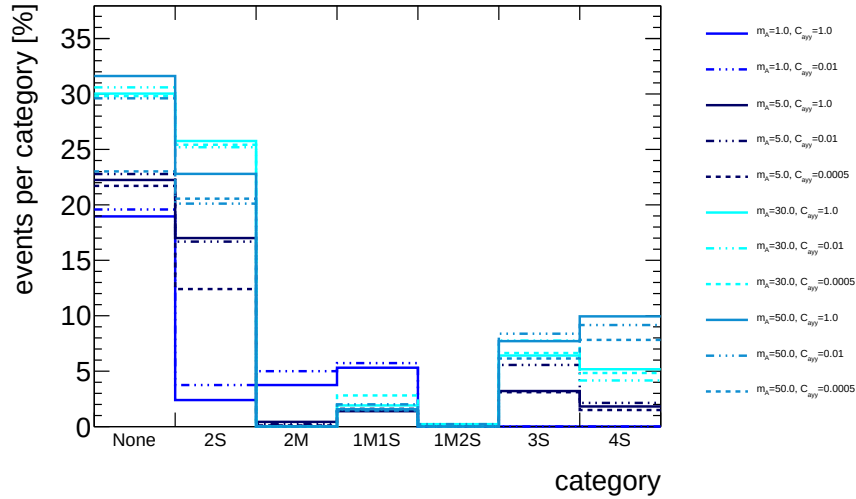


Figure 6.11: Categorization efficiency in percent for different ALP masses and couplings. Blue histograms show the categories for  $m_a = 1$  GeV, dark blue for  $m_a = 5$  GeV, teal for  $m_a = 30$  GeV and darker teal for  $m_a = 50$  GeV. Different line styles correspond to different couplings: solid lines for  $C_{a\gamma\gamma} = 1$ , dashed lines for  $C_{a\gamma\gamma} = 0.01$ , and dash-dotted lines for  $C_{a\gamma\gamma} = 5 \times 10^{-4}$ .

The cut-flow and number of observed events for selected signal samples are summarized in Table 6.5. As the ALP mass increases, the fraction of events with more than two reconstructed photons rises accordingly. This reflects the reduced boost of the ALP decay products at higher masses, leading to less collimated photons and a higher likelihood of resolving them as separate objects. As a result, more events populate categories involving three or four single photons, while the relative importance of merged-photon categories decreases. This trend is consistent with the categorization efficiencies shown in Fig. 6.11, and supports the classification strategy based on photon topology.



Table 6.5: Cut-flow and number of observed events for selected signal samples with  $C_{a\gamma\gamma} = 1$  in the mass window  $110 \text{ GeV} < m_a < 130 \text{ GeV}$ . Recreated from [150]

|                                                                                      | $m_a = 1 \text{ GeV}$ | $m_a = 2 \text{ GeV}$ | $m_a = 5 \text{ GeV}$ | $m_a = 15 \text{ GeV}$ | $m_a = 30 \text{ GeV}$ | $m_a = 45 \text{ GeV}$ |
|--------------------------------------------------------------------------------------|-----------------------|-----------------------|-----------------------|------------------------|------------------------|------------------------|
| <b>Cut-flow</b>                                                                      |                       |                       |                       |                        |                        |                        |
| All events                                                                           | 50716                 | 50264                 | 49278                 | 47445                  | 45975                  | 46783                  |
| Triggered events                                                                     | 10916                 | 7411                  | 21971                 | 25665                  | 25712                  | 27119                  |
| $\geq 2$ photon candidates                                                           | 10916                 | 7411                  | 21971                 | 25665                  | 25712                  | 27119                  |
| $\geq 3$ photon candidates                                                           | 1004                  | 2027                  | 13582                 | 15654                  | 15560                  | 17996                  |
| $\geq 4$ photon candidates                                                           | 126                   | 225                   | 3584                  | 4163                   | 4134                   | 5662                   |
| <b>Event categories (All events)</b>                                                 |                       |                       |                       |                        |                        |                        |
| 2 Single                                                                             | 1315                  | 2494                  | 9295                  | 10623                  | 10762                  | 10191                  |
| 2 Merged                                                                             | 1947                  | 331                   | 44                    | 26                     | 31                     | 25                     |
| 1 Merged, 1 Single                                                                   | 1904                  | 1002                  | 607                   | 663                    | 706                    | 716                    |
| 1 Merged, 2 Single                                                                   | 0                     | 1                     | 2                     | 4                      | 10                     | 15                     |
| 3 Single                                                                             | 12                    | 143                   | 2174                  | 2835                   | 2808                   | 2815                   |
| 4 Single                                                                             | 0                     | 25                    | 1278                  | 1900                   | 1767                   | 2890                   |
| <b>Event categories (<math>110 \text{ GeV} &lt; m_a &lt; 130 \text{ GeV}</math>)</b> |                       |                       |                       |                        |                        |                        |
| 2 Single                                                                             | 1247                  | 2202                  | 4983                  | 1988                   | 771                    | 43                     |
| 2 Merged                                                                             | 1835                  | 193                   | 14                    | 1                      | 1                      | 0                      |
| 1 Merged, 1 Single                                                                   | 1813                  | 835                   | 354                   | 125                    | 51                     | 5                      |
| 1 Merged, 2 Single                                                                   | 0                     | 0                     | 1                     | 2                      | 1                      | 1                      |
| 3 Single                                                                             | 12                    | 122                   | 1431                  | 1643                   | 1313                   | 803                    |
| 4 Single                                                                             | 0                     | 23                    | 1273                  | 1887                   | 1753                   | 2853                   |

## 6.6 Background estimation

After categorising each event, invariant mass distributions are constructed for every category. In the three- and four-photon categories, events are further binned according to the reconstructed ALP mass. The objective is to perform maximum likelihood fits for each  $m_a$ – $C_{a\gamma\gamma}$  pair using the corresponding invariant mass histogram.

To identify potential excesses, it is essential to estimate the background contribution in the signal region of each histogram. Across all categories, a large, continuous background of unknown composition is expected. In the two-photon categories, an additional contribution from  $H \rightarrow \gamma\gamma$  is anticipated, exhibiting the same invariant mass shape as the signal processes.

The continuous background is estimated in a fully data-driven manner using sideband fits, following ATLAS recommendations [170]. In each invariant mass histogram, the signal region—defined as the region around the Higgs boson mass with a signal contribution exceeding 95 %—is blinded to avoid bias.

The remaining sideband regions are fitted using three different functions:

1. **Nominal function:** The function providing the best description of the sideband data (according to the  $\chi^2$  value) is chosen as the main background estimate.
2. **Alternative fit range:** The same functional form as the nominal function is used, but with a slightly modified mass range. The difference from the nominal fit is taken as a systematic uncertainty on the background estimate.

3. **Alternative functional form:** A different type of function is used to model the background. The resulting deviation is used to evaluate the spurious signal uncertainty (see Section 6.7.1).

The functions and their respective fit and signal regions for all categories and  $m_a^{\text{reco}}$  bins are summarised in Table 6.6.

Table 6.6: Fit functions, fit ranges and range of signal region for all categories and  $m_a^{\text{reco}}$  bins. The first two columns show the category and the  $m_a^{\text{reco}}$  bin, the third the  $m_a$  range considered to define fitting and blinding ranges. Afterwards, the nominal fit function (nom fkt) and the alternative fit function (alt fkt) are listed. The last columns show the nominal (nom) fitting range, the alternative (alt) range and the blinded signal region.

| Category | $m_a^{\text{reco}}$ bin | $m_a$ range      | nom fkt | alt fkt     | nom range | alt range | sig region |
|----------|-------------------------|------------------|---------|-------------|-----------|-----------|------------|
| 2S       | -                       | $0 < m_a < 62$   | Landau  | Exp + Pol-2 | 100-150   | 105-145   | 115-135    |
| 1M1S     | -                       | $0 < m_a < 62$   | Landau  | Exp + Pol-2 | 100-150   | 105-145   | 115-135    |
| 2M       | -                       | $0 < m_a < 62$   | Pol-2   | Pol-3       | 100-150   | 105-145   | 115-135    |
| 3S       | I.                      | $3.5 < m_a < 10$ | Pol-3   | Pol-4       | 80-150    | 85-145    | 105-130    |
| 3S       | II.                     | $10 < m_a < 25$  | Pol-3   | Pol-4       | 80-150    | 85-145    | 100-125    |
| 3S       | III.                    | $25 < m_a < 40$  | Pol-3   | Pol-4       | 80-150    | 85-145    | 100-125    |
| 4S       | I.-IV.                  | $3.5 < m_a < 62$ | Pol-2   | Pol-3       | 105-145   | 100-150   | 120-130    |

The background estimation method is validated using both simulated control samples and dedicated data control regions.

For the two-photon categories, a control region in data is defined by inverting the isolation criterion, thereby selecting events that are unlikely to be signal. In the three- and four-photon categories, control regions are constructed for each  $m_a$  bin by combining the remaining three  $m_a^{\text{reco}}$  regions.

To validate the method in simulation, multiphoton samples containing only single photons are used. These represent an extreme case of the expected background composition. In contrast, the data control region—created via inverted isolation—is dominated by misidentified hadronic activity.

In reality, the background is expected to be a mixture of both scenarios. Since the method yields consistent results in both validation cases, it can be concluded that the background estimation strategy is robust and appropriate for this analysis.

A detailed discussion of the two-photon categories is provided in Section 6.6.1. The analysis of the three- and four-photon categories is discussed in [171], and summarised again in Appendix A.1.

### 6.6.1 Background estimation for two photon categories

Fig. 6.12 shows the background estimation for the two-photon categories. The black dots represent data. Only the data points outside the vertical grey dashed lines—corresponding to the blinded signal region—are used in the sideband fits. The dashed blue line shows the best (nominal) sideband fit, while the dashed green line shows the alternative fit. The blue histogram, with the patterned band, represents the estimated background together with the systematic uncertainty derived from the fit range variation.

A purple histogram shows the shape of a representative  $H \rightarrow aa$  signal sample, scaled arbitrarily for better visibility. The  $H \rightarrow \gamma\gamma$  contribution, scaled to the Standard Model expectation, is indicated by the yellow histogram, though its size renders it nearly invisible. The ratio plots show the data-to-background ratio, with the systematic uncertainty displayed as a shaded band.

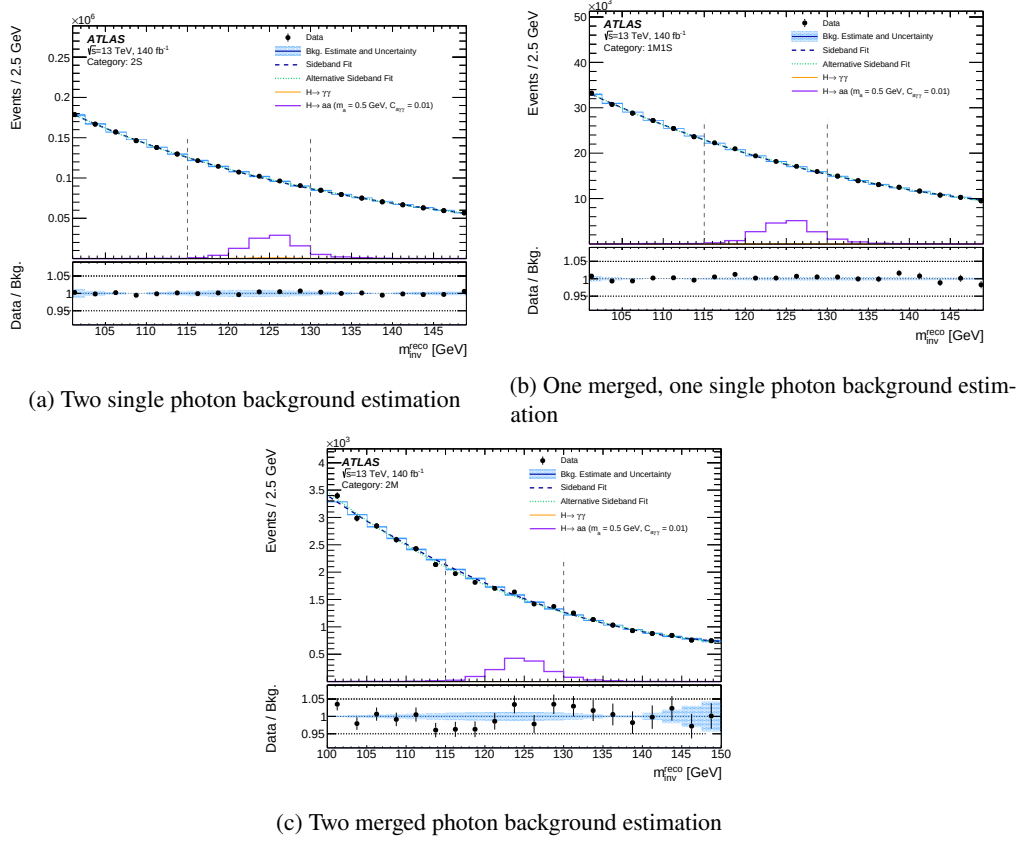


Figure 6.12: Background estimation for the two-photon categories. Data are shown as black points. The region between the vertical dashed lines is excluded from the fit and corresponds to the blinded signal region. The blue dashed line represents the nominal sideband fit, while the green dashed line corresponds to the alternative fit. The blue histogram with a patterned band indicates the background estimation and its associated systematic uncertainty. The yellow line shows the  $H \rightarrow \gamma\gamma$  contribution, scaled to the Standard Model expectation, and the purple line represents an arbitrarily scaled signal sample for visual clarity. [1]

In the 2S (two single photon) category, the main fit function is a Landau function, which was also used in the Higgs boson discovery [3]. It is fitted over the range 100–150 GeV, excluding the Higgs mass window:

$$f(x)^{(100-150)} \approx N_0 \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}(x + e^x)\right) \quad (6.3)$$

The alternative fit uses an exponential plus a second-order polynomial in the same mass range:

$$f(x)^{(100-150)} = N_0 \cdot \exp(px) + N_1 + ax^2 + bx \quad (6.4)$$

To evaluate the systematic uncertainty, the nominal function is re-fitted with a reduced mass range of 105–145 GeV.

In the 1M1S (one merged, one single photon) category, the same functional forms and procedure are used, as the distribution is found to be similar.

For the 2M (two merged photons) category, the background shape differs significantly due to its

strong dependence on the neural network classifier. Here, the nominal function is a second-order polynomial, and the alternative is a third-order polynomial.

It is evident that the method successfully reproduces the data distribution, even within the blinded signal region after unblinding. The ratio plots confirm that the prediction closely follows the data in the 2S and 1M1S categories. While the 2M category shows minor fluctuations, the agreement remains well within the assigned uncertainties.

## 6.7 Uncertainties

To determine the  $H \rightarrow aa$  production cross-section, a profile likelihood fit is performed for each mass–coupling point using the invariant mass distributions discussed in the previous section. In the absence of a significant excess, an upper limit on the production cross-section is instead derived.

The complete statistical procedure is outlined in Section 6.8.1. This analysis requires the incorporation of various systematic uncertainties, which are detailed in Section 6.7.1 and summarised in Table 6.7. An illustrative example of the full procedure is presented in Section 6.8.2, and the final results are reported in Section 6.9.

### 6.7.1 Systematic uncertainties

To obtain meaningful results, it is essential to consider various sources of uncertainty. Statistical uncertainties arise from fluctuations due to limited sample sizes and must be accounted for. In addition, systematic uncertainties—originating from both experimental and theoretical sources—must also be taken into consideration.

Systematic uncertainties can broadly be categorised into two types: **experimental** and **theoretical**. Experimental uncertainties result from incomplete knowledge of the detector environment, such as calibration, alignment, and resolution. Theoretical uncertainties, on the other hand, are introduced by limitations in the modelling of the underlying physics processes.

The following subsections outline the main sources of systematic uncertainties. Special attention is given to those related to the custom photon identification (Section 6.7.1) and to background modelling (Section 6.7.1). A novel treatment of displaced vertex uncertainties—developed for the first time within ATLAS—is discussed in detail in [171], with a summary provided in Appendix A.2. The remaining subsections describe standard experimental uncertainties.

#### Uncertainty on Luminosity

The results presented in Section 6.9 are normalised to the integrated luminosity of the dataset. Since the luminosity is not precisely known, it introduces a systematic uncertainty. This analysis uses the full ATLAS Run-2 dataset, corresponding to an integrated luminosity of  $\mathcal{L} = 140.1 \text{ fb}^{-1}$ , with a relative uncertainty of 0.83 %. This value is derived using the ATLAS luminosity calibration procedure described in [151].

#### Uncertainty Due to Pileup

At each bunch crossing of the LHC beams, around 50 proton–proton collisions occur simultaneously. This leads to additional tracks and energy deposits in the detector unrelated to the primary hard-scatter

process—a phenomenon known as **pileup**.

Pileup is modelled in the simulation, but the distribution of the average number of interactions per bunch crossing,  $\langle\mu\rangle$ , may differ between data and simulation. To match this distribution, a reweighting is applied. Systematic variations of this reweighting are provided by ATLAS and propagated through the analysis.

The associated uncertainty ranges from 1.3 % in the two-single-photon category up to 3.6 % in categories with more than two photons. This increase is primarily due to the purely track-based isolation criteria. Nevertheless, no significant distortion of the invariant mass shape is observed.

### Trigger Uncertainty

Trigger efficiencies must be accurately modelled to ensure consistency between data and simulation. This is achieved by applying scale factors to simulated events to match the trigger performance observed in data. These scale factors and their systematic variations are provided centrally by ATLAS and are typically parametrised as functions of relevant kinematic variables, such as photon transverse momentum and pseudorapidity.

Including these corrections is essential for precise modelling of selection efficiencies and, consequently, accurate signal and background predictions. The magnitude of the uncertainty varies by category and is typically at the percent level. In the 2S category, however, it reaches up to 7.4 % [172].

### Photon Scale and Resolution Uncertainties

Accurate photon energy reconstruction requires detector calibration, which introduces systematic uncertainties. These are accounted for via photon energy scale and resolution uncertainties, provided centrally by ATLAS.

The simplest recommended scheme—with two uncorrelated nuisance parameters—is used, and the associated variations are propagated through the full analysis. For clarity, the contributions from photon scale, resolution, identification, and isolation are summarised collectively as “Standard Photon” uncertainties in Table 6.7. All contributions are generally at the percent level.

### Uncertainties Due to Photon Identification and Isolation

Photon identification in ATLAS is based on the shower shape variables introduced in Section 6.4.1. To align the simulation with data, a shower shape “fudging” tool is used to adjust the distributions. Additionally, ATLAS provides scale factors for standard identification criteria, along with systematic variations, which are applied event-by-event and propagated through the analysis.

Isolation-related uncertainties are treated similarly. Both calorimeter and track-based isolation criteria must be scaled appropriately, and systematic variations of the corresponding scale factors are included in the uncertainty budget.

### Theoretical Uncertainties

Theoretical models inherently introduce uncertainties, primarily due to missing higher-order corrections. These are estimated by varying the renormalisation and factorisation scales, and recalculating the cross-sections for all simulated signal samples.

The maximum observed deviation is below 1 %, which is conservatively applied as the associated uncertainty across all samples.

Additional theoretical uncertainties stem from incomplete knowledge of the proton’s partonic structure and the strong coupling constant  $\alpha_s$ . These are estimated using common prescriptions: the uncertainty on  $\alpha_s$  is assessed by varying its value within recommended bounds. The combined impact of these uncertainties is also below 1 %.

Finally, the uncertainty on the Higgs boson branching ratio into two photons, relevant for the  $H \rightarrow \gamma\gamma$  contribution, is taken to be 5.7 %, following [48].

### Uncertainties from Neural Network Classifiers

Machine learning techniques, such as the neural network classifiers used for custom photon identification, also introduce systematic uncertainties. Quantifying these uncertainties remains the subject of ongoing discussion and investigation within the high-energy physics community.

One common approach to estimating classifier-related uncertainties is to compare its performance on data and simulation. However, this is often challenging due to the lack of truth labels in data. Moreover, the composition of data samples is typically not fully understood, and discrepancies between data and simulation must be minimised to ensure that any derived uncertainty accurately reflects true differences in classifier behaviour. It is also essential to avoid evaluating uncertainties using the same events that were employed during training or analysis.

A viable strategy is to use electrons, which exhibit nearly identical shower shapes to photons. In particular,  $Z \rightarrow ee$  decays provide a clean and well-understood control sample for comparing classifier outputs between data and simulation. These events consist of prompt, well-separated electrons, enabling a reliable assessment of the classifier’s performance on single-photon-like objects in both data and Monte Carlo.

Good agreement is observed between data and simulation in the classifier output distributions. Due to the symmetric construction of the classifier, the efficiency for single photons is the same as that for fake or merged photons by design. The observed differences between data and simulation are propagated as systematic uncertainties, leading to impacts of up to 15 % in the 2M and 1M1S categories.

### Uncertainties from Background Modelling

The background estimation method described in Section 6.6 is also subject to systematic uncertainties. To quantify this, the fitting range of the sideband model is reduced and the fit is repeated. The resulting uncertainty is defined as the bin-by-bin difference between the nominal and the alternative fit functions. Reducing the fit range provides additional flexibility, often improving the agreement with the data in the sidebands and reducing potential bias in signal extraction. However, if the model becomes overly flexible, it risks absorbing parts of a potential signal. To account for this, an additional uncertainty—referred to as the **fit bias** or **spurious signal** uncertainty—is introduced, following the ATLAS recommendations in [170].

This uncertainty quantifies the degree to which a signal could be absorbed under a background-only hypothesis due to imperfect modelling of the background shape.

To evaluate it, a template fit is performed on pseudo-data containing the expected number of events in the signal region. The alternative fit function introduced earlier is used to generate the pseudo-data.

For each mass–coupling pair, a full signal-plus-background fit—including all systematic uncertainties—is performed independently in each category. The aim of the fit is to determine the signal strength, denoted by  $\langle\mu\rangle$ . The spurious signal in the background-only case is defined as

$$s_{\text{spur}}^{\mu=0} = \langle\mu_{\text{fit}}\rangle, \quad (6.5)$$

3

This value is typically sufficient to define the spurious signal uncertainty, which is implemented as a normalisation uncertainty on the signal yield with magnitude  $s_{\text{spur}}^{\mu=0}$ .

To verify the stability of the fit, the procedure is repeated using a signal-plus-background hypothesis. In this test, a signal corresponding to  $\mu = 1$  is injected by adding all simulated signal events to the pseudo-data with unit weight. The fit is then expected to return  $\langle\mu\rangle = 1$ , and the deviation defines a second spurious signal estimate:

$$s_{\text{spur}}^{\mu=1} = \langle\mu_{\text{fit}}\rangle - 1. \quad (6.6)$$

The two estimates typically yield comparable results. To remain conservative, the larger of the two values is adopted as the spurious signal uncertainty in each category.

The largest observed impact occurs in the 4S category, where the spurious signal uncertainty reaches up to 0.4 %.

### 6.7.2 Magnitude of uncertainties

The magnitudes of all relevant uncertainties are summarised in Table 6.7. Photon scale and resolution uncertainties are grouped together with photon identification and isolation uncertainties under the label **Standard photon**. The **Custom photon** category includes uncertainties associated with the neural network classifiers and with the treatment of displaced vertices. Additionally, uncertainties due to pile-up, trigger modelling, and theoretical inputs are listed. The statistical uncertainty, arising from the finite number of simulated events, is also shown. The final column displays the total combined uncertainty for each category.

## 6.8 Statistical analysis

To measure the  $H \rightarrow aa$  production cross-section in the presence of an excess, or to set an upper limit otherwise, a profile likelihood fit is performed using the invariant mass spectra introduced in Section 6.6. The fitting strategy is detailed in Section 6.8.1. Subsequently, a full example fit is shown in Section 6.8.2.

### 6.8.1 Fit strategy

Studies have shown that the sensitivity of a profile likelihood fit for different mass–coupling pairs is primarily driven by the category with the highest signal-to-background ratio  $\left(\frac{S}{\sqrt{B}}\right)$ . Nonetheless, simultaneously fitting multiple categories can help constrain background contributions more effectively. However, fitting all categories for each mass–coupling pair is computationally expensive, and including categories with few signal events may compromise the stability of the fit.

<sup>3</sup> The superscript  $\mu = 0$  indicates that no signal is injected, i.e. the fit is performed under the background-only hypothesis.



Table 6.7: Overview of relative systematic uncertainties for different  $m_a$  and  $C_{a\gamma\gamma}$  in all relevant categories. Taken from [1].

| Category   | $m_a$<br>GeV | $C_{a\gamma\gamma}$<br>$\text{TeV}^{-1}$ | MC-Stat<br>[%] | Standard<br>Photon [%] | Custom<br>Photon [%] | Trigger<br>[%] | Pile-Up<br>[%] | Modelling/<br>Theory [%] | Combined<br>[%] |
|------------|--------------|------------------------------------------|----------------|------------------------|----------------------|----------------|----------------|--------------------------|-----------------|
| 2S         | 0.5          | 1                                        | 2.12           | 2.5                    | 2.0                  | 7.4            | 1.3            | 3.3                      | 9.06            |
| 1S1M       | 0.5          | 1                                        | 1.64           | 2.1                    | 5.0                  | 2.5            | 2.1            | 3.3                      | 7.32            |
| 2M         | 0.5          | 1                                        | 1.39           | 2.4                    | 10.3                 | 2.4            | 2.3            | 3.3                      | 11.65           |
| 2S         | 1            | 1                                        | 2.77           | 2.6                    | 1.5                  | 6.4            | 1.3            | 3.3                      | 8.38            |
| 1S1M       | 1            | 1                                        | 2.29           | 2.1                    | 5.0                  | 2.6            | 2.1            | 3.3                      | 7.53            |
| 2M         | 1            | 1                                        | 2.27           | 2.4                    | 10.3                 | 2.5            | 2.3            | 3.3                      | 11.81           |
| 2S         | 3            | 1                                        | 1.35           | 2.9                    | 1.1                  | 6.5            | 1.3            | 3.3                      | 8.14            |
| 1S1M       | 3            | 1                                        | 3.82           | 2.2                    | 5.0                  | 1.9            | 2.1            | 3.3                      | 7.96            |
| 2M         | 3            | 1                                        | 14.0           | 2.2                    | 10.3                 | 1.8            | 2.3            | 3.3                      | 18.07           |
| Prompt     |              |                                          |                |                        |                      |                |                |                          |                 |
| 4S         | 5            | 1                                        | 3.03           | 9.86                   | 0.0                  | 2.7            | 3.4            | 2.1                      | 11.39           |
| 4S         | 10           | 1                                        | 2.36           | 8.8                    | 0.0                  | 3.2            | 3.4            | 2.1                      | 10.45           |
| 4S         | 20           | 1                                        | 2.45           | 8.6                    | 0.0                  | 2.5            | 3.4            | 2.1                      | 10.11           |
| 4S         | 30           | 1                                        | 2.54           | 8.1                    | 0.0                  | 1.3            | 3.4            | 2.1                      | 9.47            |
| 4S         | 40           | 1                                        | 2.34           | 8.4                    | 0.0                  | 1.4            | 3.4            | 2.1                      | 9.69            |
| 4S         | 50           | 1                                        | 1.83           | 8.7                    | 0.0                  | 1.5            | 3.4            | 2.1                      | 9.86            |
| Long-Lived |              |                                          |                |                        |                      |                |                |                          |                 |
| 4S         | 10           | $10^{-5}$                                | 37.8           | 6.496                  | 22.2                 | 0.4155         | 3.6            | 3.3                      | 44.59           |
| 4S         | 20           | $10^{-5}$                                | 3.39           | 7.11                   | 27.17                | 2.603          | 3.6            | 3.3                      | 28.82           |
| 4S         | 30           | $10^{-5}$                                | 2.55           | 8.388                  | 9.884                | 1.321          | 3.6            | 3.3                      | 14.15           |
| 4S         | 40           | $10^{-5}$                                | 2.25           | 8.701                  | 3.753                | 2.183          | 3.6            | 3.3                      | 11.11           |
| 4S         | 50           | $10^{-5}$                                | 1.65           | 8.185                  | 3.259                | 1.816          | 3.6            | 3.3                      | 10.37           |

Given the diverse kinematic properties of the different ALP scenarios, events predominantly fall either into the two-photon categories or the three- and four-photon categories. Therefore, for each mass–coupling pair, the two categories with the best  $\frac{S}{\sqrt{B}}$  (provided they contain at least 10 signal events) are selected for the fit.

For very low ALP masses ( $m_a \leq 0.5 \text{ GeV}$ ), the photons are highly collimated and are often reconstructed as a single photon, making the 2S category particularly relevant. Thus, all three two-photon categories are included in the fit in this case. Conversely, for high masses ( $m_a > 40 \text{ GeV}$ ), the signal shape in the 3S category becomes broad, complicating proper blinding for the background fit. Therefore, only the 4S category is used in this region.

The categories used in the fit for different mass ranges are summarised in Table 6.8.

The likelihood function is defined as follows:



Table 6.8: Categories used in the fit.

| mass range                                  | categories                            |
|---------------------------------------------|---------------------------------------|
| $m_a \leq 0.5 \text{ GeV}$                  | 2 Merged, 1 Merged 1 Single, 2 Single |
| $0.5 \text{ GeV} < m_a < 3.5 \text{ GeV}$   | 2 Merged, 1 Merged 1 Single           |
| $3.5 \text{ GeV} \leq m_a < 40 \text{ GeV}$ | 3 Single, 4 Single                    |
| $m_a > 40 \text{ GeV}$                      | 4 Single                              |

$$\mathcal{L} = \prod_c \left( \text{Pois}(n_c | N_c(\boldsymbol{\theta})) \cdot \prod_{i=1}^{n_{\text{bins}}^c} f_c(n_i, m_{\text{inv}}^{\text{reco},i}, \boldsymbol{\theta}) \right) \cdot G(\boldsymbol{\theta}). \quad (6.7)$$

Here,  $n_{\text{bins}}^c$  denotes the number of bins,  $N_c$  the expected, and  $n_c$  the observed number of events in category  $c$ . The probability density function (PDF)  $f_c$  is estimated from simulation and evaluated using the observed number of events  $n_i$  in bin  $i$  and the nuisance parameters (NPs)  $\boldsymbol{\theta}$ . Systematic uncertainties are modelled via the nuisance parameters and constrained by the functions  $G(\boldsymbol{\theta})$ . All normalisation uncertainties are grouped into a single nuisance parameter. The other uncertainties and their associated constraint terms are listed in Table 6.9.

Table 6.9: Name, short description and type of systematic uncertainties with corresponding constraint terms. The **type** column specifies whether the uncertainty affects only the normalisation (**norm**), both shape and normalisation (**shape**), or is a bin-by-bin statistical uncertainty (**stat**). **Global scale factor** refers to a single normalisation factor affecting all signal and background yields.

| name             | description                                                   | type                | constraint term       |
|------------------|---------------------------------------------------------------|---------------------|-----------------------|
| lumi             | Luminosity                                                    | global scale factor | Gaus                  |
| sig_norm         | Signal normalisation                                          | norm                | Gaus                  |
| PhotonScale      | Photon energy scale                                           | shape               | Gaus                  |
| PhotonResolution | Photon energy resolution                                      | shape               | Gaus                  |
| Trigger          | Trigger efficiency                                            | shape               | Gaus                  |
| uncorr_siguncrt  | Statistical uncertainty (signal MC)                           | stat                | $\prod_i \text{Gaus}$ |
| uncorr_bkguncrt  | Statistical uncertainty (background)                          | stat                | $\prod_i \text{Gaus}$ |
| hyy_xs           | $\text{H} \rightarrow \gamma\gamma$ cross-section uncertainty | norm                | Gaus                  |
| bkg_est_sys2     | Background estimation (sideband fit)                          | shape               | Gaus                  |
| spurious_signal  | Spurious signal uncertainty                                   | shape               | Gaus                  |

The fit is implemented using the pyhf framework [173, 174].

### $\text{H} \rightarrow \gamma\gamma$ contribution in two-photon categories

The  $\text{H} \rightarrow \gamma\gamma$  contribution in the two-photon categories cannot be reliably estimated using the sideband method described in Section 6.6.

Several strategies exist to address this: One approach is to introduce a second parameter of interest (POI) for the  $\text{H} \rightarrow \gamma\gamma$  process, allowing the fit to scale the Higgs branching fraction freely. However,

excluding the 2S category in this approach can result in unphysical solutions due to the strong anti-correlation between the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow aa$  components in the mass spectrum. For example, a combined signal yield of zero could result from a positive  $H \rightarrow aa$  and an equally negative  $H \rightarrow \gamma\gamma$  scale factor. While a simultaneous fit across multiple categories mitigates this issue for most ALP scenarios (due to their differing contributions), the problem persists for low-mass ALPs whose events primarily populate the 2S category.

An alternative method fixes the  $H \rightarrow \gamma\gamma$  branching fraction to the Standard Model prediction and allows it to vary within its experimental uncertainty. This approach is motivated if the sensitivity to  $H \rightarrow \gamma\gamma$  is comparable to previous measurements. To test this, several Asimov pseudo-data studies are performed:

- A single-POI fit to the  $H \rightarrow \gamma\gamma$  process is conducted in the 2S category to determine the branching fraction.
- Simultaneous fits using two POIs are performed in combinations of the two-photon categories, fixing the  $H \rightarrow aa$  POI to zero and one, respectively.

The resulting measurements agree well with Standard Model expectations. Therefore, this strategy is adopted in the final analysis.

### 6.8.2 Statistical analysis for $m_a = 1$ GeV and $C_{a\gamma\gamma} = 1$

In this section, the fit and its results are discussed using the example of  $m_a = 1$  GeV and  $C_{a\gamma\gamma} = 1$ . At this mass, ALPs are highly boosted, causing a large fraction of decay photons to merge. The event yields for signal, background, and  $H \rightarrow \gamma\gamma$  are shown in Table 6.10. As expected, a large number of  $H \rightarrow aa$  events fall into the 2M and 1M1S categories. Due to lower background in the 2M category, the  $S/\sqrt{B}$  ratio is more favourable there. It should also be noted that the majority of  $H \rightarrow \gamma\gamma$  events are categorised as 2S, as expected.

Table 6.10: Simulated signal ( $H \rightarrow aa$ ) and simulated  $H \rightarrow \gamma\gamma$  background events with background events from data-driven estimation and  $S/\sqrt{B}$ .

| Category       | signal events | background events | $H \rightarrow \gamma\gamma$ events | $\frac{s}{\sqrt{B}}$ |
|----------------|---------------|-------------------|-------------------------------------|----------------------|
| 2Merged        | 1935.49       | 32466.43          | 25.15                               | 10.742               |
| 1Merged1Single | 1900.32       | 380228.94         | 1496.12                             | 3.0818               |
| 2Single        | 1307.52       | 2007919.96        | 24528.53                            | 0.92273              |
| 3Single        | 12.23         | 10767.32          | 0                                   | 0.11791              |
| 4Single        | 0.0           | 46.78             | 0                                   | 0.0011697            |
| 1Merged2Single | 0.0           | 0.01              | 0                                   | 0.11402              |

As described earlier, in this case the 2M and 1M1S categories are fitted simultaneously using a freely floating POI to scale the  $H \rightarrow aa$  contribution. The  $H \rightarrow \gamma\gamma$  contribution is fixed to its Standard Model cross-section. The invariant photon mass distributions for these categories, including nominal and varied signal shapes, are shown in Fig. 6.13. The continuous background is shown in green, and the nominal signal distribution is overlaid as a solid red histogram, scaled arbitrarily for better visibility (the scale factor is indicated in the legend). Up and down variations for photon scale, resolution, and

trigger uncertainties are shown as dashed lines. The ratio plots underneath display the deviation of each systematic variation from the nominal shape.

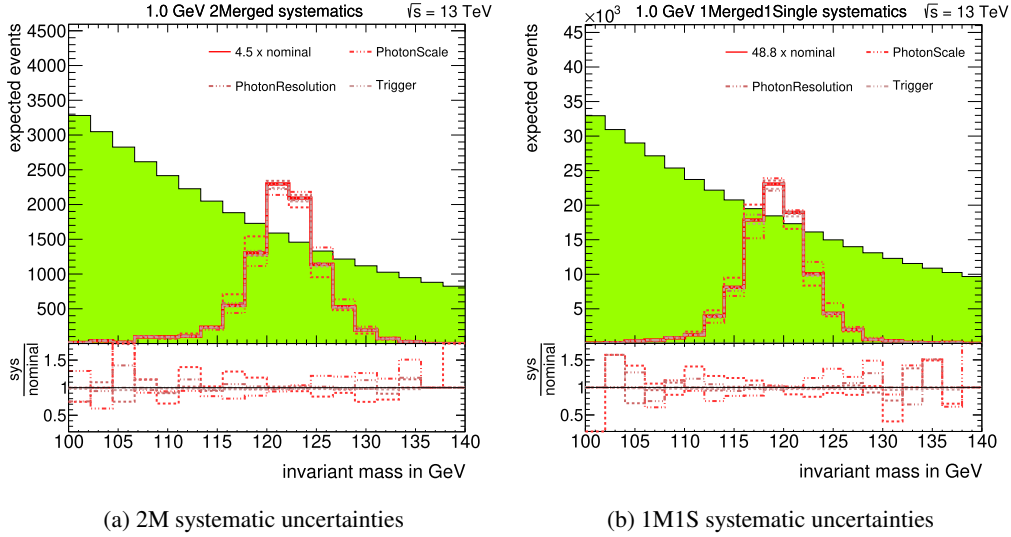


Figure 6.13: Background distribution (green) with nominal signal shape in solid red and systematic variations as dashed lines. Left: 2 Merged (a); Right: 1 Merged 1 Single (b). Lower panels show ratios of systematic variations to the nominal signal shape.

Fig. 6.14 shows the post-fit distribution corresponding to the best-fit value of  $\mu$ . The green histogram represents the post-fit background with statistical uncertainties indicated as a grey band. The black dots correspond to the data. No  $H \rightarrow aa$  or  $H \rightarrow \gamma\gamma$  excess is observed.

The pull plot in Fig. 6.15 shows the post-fit values of the nuisance parameters. Each dot represents the deviation of the parameter from its nominal value (in units of its standard deviation), while the bars indicate post-fit uncertainties. A value near zero implies little to no pull from the nominal value. Most nuisance parameters show negligible pull and are unconstrained. Only the nuisance parameters associated with the spurious signal are constrained, which is expected since their shape closely resembles that of the signal. This effect is also visible in the correlation matrix shown in Fig. 6.16, where a strong anti-correlation between the spurious signal nuisance parameters and  $\mu$  is observed. Otherwise, only weak correlations exist between parameters.

In Fig. 6.17, the upper limits on the signal strength  $\mu$  are shown. These are obtained by scanning fixed values of  $\mu$  and computing the corresponding  $CL_s$  confidence levels. The scan is stopped once the curves intersect the 5% line. As expected, the upper limits from the statistical-only fit are significantly lower than those including systematics.

## 6.9 Results

Across all fitted mass–coupling hypotheses, the fit results exhibit similar behaviour: no significant deviations from the background-only hypothesis are observed. As shown in the example in Section 6.8.2, the nuisance parameters remain largely unconstrained, and no excess is present in the data.

Upper limits at 95% confidence level (CL) on the branching fraction  $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$  are

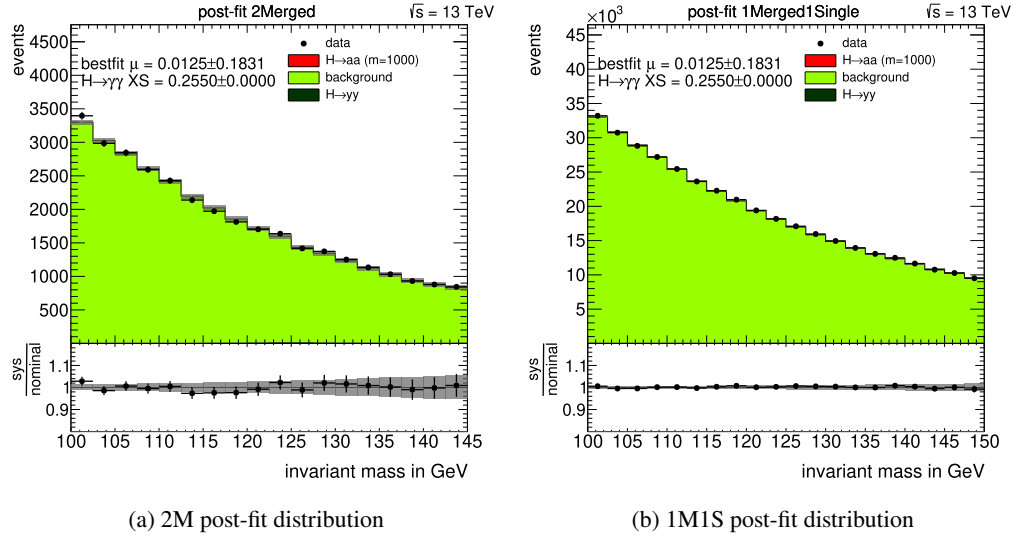


Figure 6.14: Post-fit distributions for 2 Merged (a) and 1 Merged 1 Single (b) categories for  $m_a = 1$  GeV,  $C_{a\gamma\gamma} = 1$ . The filled green histogram shows the fitted background, the grey band indicates its uncertainty, and the black points represent the data.

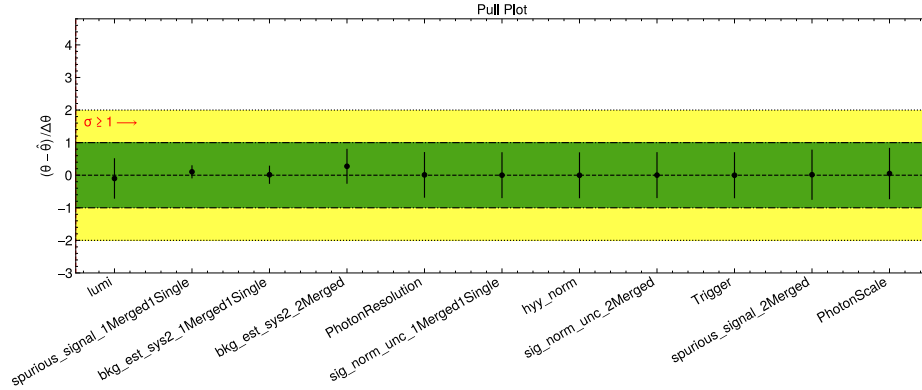


Figure 6.15: Pull plot of the nuisance parameters from the signal-plus-background fit. Each row corresponds to a nuisance parameter. The central dot shows the post-fit value (in standard deviations), and the bar represents its uncertainty.

derived for every mass–coupling pair using the  $CL_S$  method [175]. These upper limits are presented as a function of  $m_a$  for each probed value of  $C_{a\gamma\gamma}$ . Fig. 6.18 shows the results for four tested couplings:  $C_{a\gamma\gamma} = 1$ ,  $C_{a\gamma\gamma} = 0.01$ ,  $C_{a\gamma\gamma} = 5 \times 10^{-4}$ , and  $C_{a\gamma\gamma} = 10^{-5}$ .

All limits are interpolated and mildly smoothed between neighbouring mass points. The solid black lines indicate the observed limits, while the dashed lines represent the expected limits. The green and yellow bands correspond to the  $\pm 1\sigma$  and  $\pm 2\sigma$  variations around the expected limits, respectively.

It can be seen that the limits for masses below  $m_a < 3.5$  GeV are weaker compared to those at higher masses. This is attributed to increased sensitivity at higher masses due to reduced background

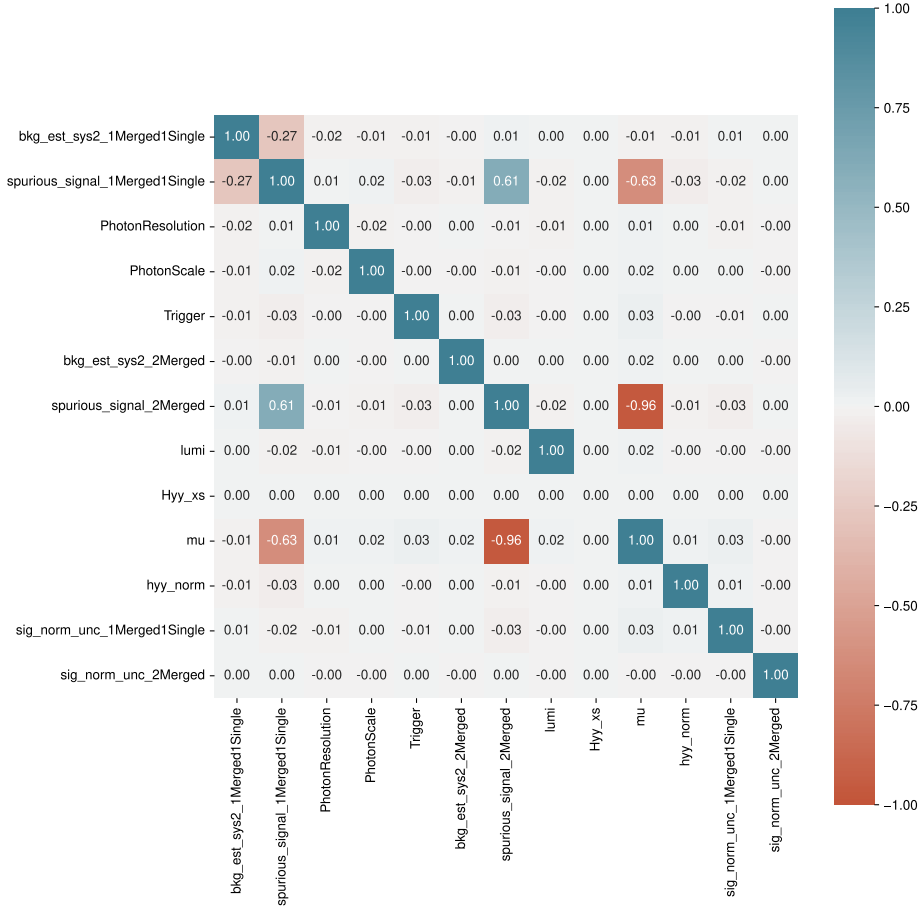


Figure 6.16: Correlation matrix of the fit parameters. Strong anti-correlation between spurious signal nuisance parameters and  $\mu$  is observed.

in the 4S category. For  $C_{a\gamma\gamma} = 10^{-5}$ , the limits for  $m_a > 10$  GeV and  $m_a > 25$  GeV are relatively uniform, ranging between  $2 \times 10^{-4}$  and  $2 \times 10^{-5}$ .

Another clear trend is that, for a given mass, lower coupling values lead to weaker limits. This loss in sensitivity arises from the longer ALP lifetimes associated with smaller couplings, causing decays to occur outside the detector acceptance. In these regions, some mass points are missing due to insufficient experimental sensitivity.

A more detailed view of the low-mass region is shown in Fig. 6.19, which provides a zoom-in for masses between 0.4 GeV and 5.5 GeV and couplings  $C_{a\gamma\gamma} = 1$  and  $C_{a\gamma\gamma} = 0.01$ . The limits in this region range between  $3 \times 10^{-2}$  and  $10^{-4}$ . Sensitivity is lowest around 3 GeV, corresponding to the transition between merged and resolved photon categories where background levels are higher.

The branching fraction limits can be translated into exclusion limits on the ALP–photon coupling, following the procedure in [75]. This assumes that all ALPs decay into photons and that the effective Higgs–ALP coupling is  $C_{aH}^{\text{eff}}/\Lambda^2 = 1 \text{ TeV}^{-2}$ . The resulting exclusion is displayed in Fig. 6.20, which shows the ALP mass  $m_a$  on the x-axis and the coupling  $C_{a\gamma\gamma}$  on the y-axis.

The shaded blue region corresponds to the parameter space excluded by this analysis. It is evident

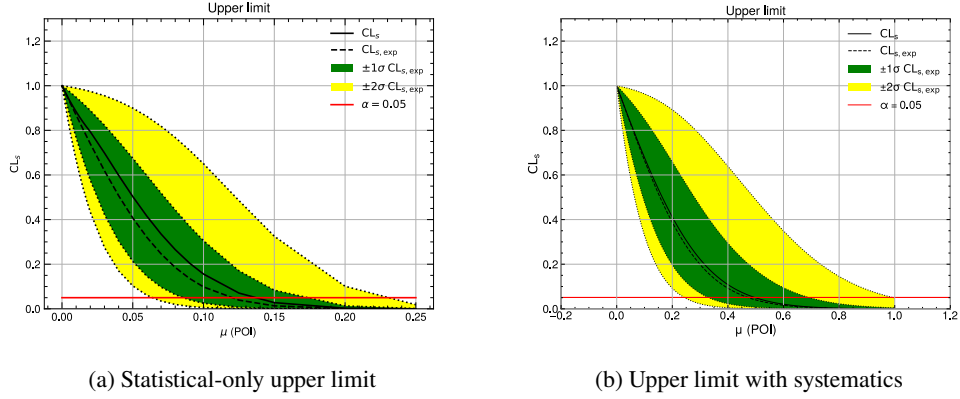


Figure 6.17: Upper limits on the signal strength  $\mu$  from fits including only statistical uncertainties (a) and all systematic uncertainties (b). The observed limit is shown where the red curve intersects the solid black line. The expected limit is indicated by the dashed line, with the green and yellow bands representing the  $1\sigma$  and  $2\sigma$  uncertainties.

that the excluded region significantly reduces the parameter space available to ALP models that could explain the  $(g - 2)_\mu$  anomaly, indicated by the orange band centred around  $C_{a\gamma\gamma} = 1$ . The solid blue line shows the exclusion assuming  $C_{aH} = 1$ , while the dashed line corresponds to  $C_{aH} = 0.1$ .

For comparison, previously excluded regions are also displayed: the green and purple areas indicate exclusions from light-by-light scattering measurements by CMS [92] and ATLAS [176], respectively. Additional exclusions from beam dump experiments, supernova SN1987a, and cosmological observations are taken from [177].

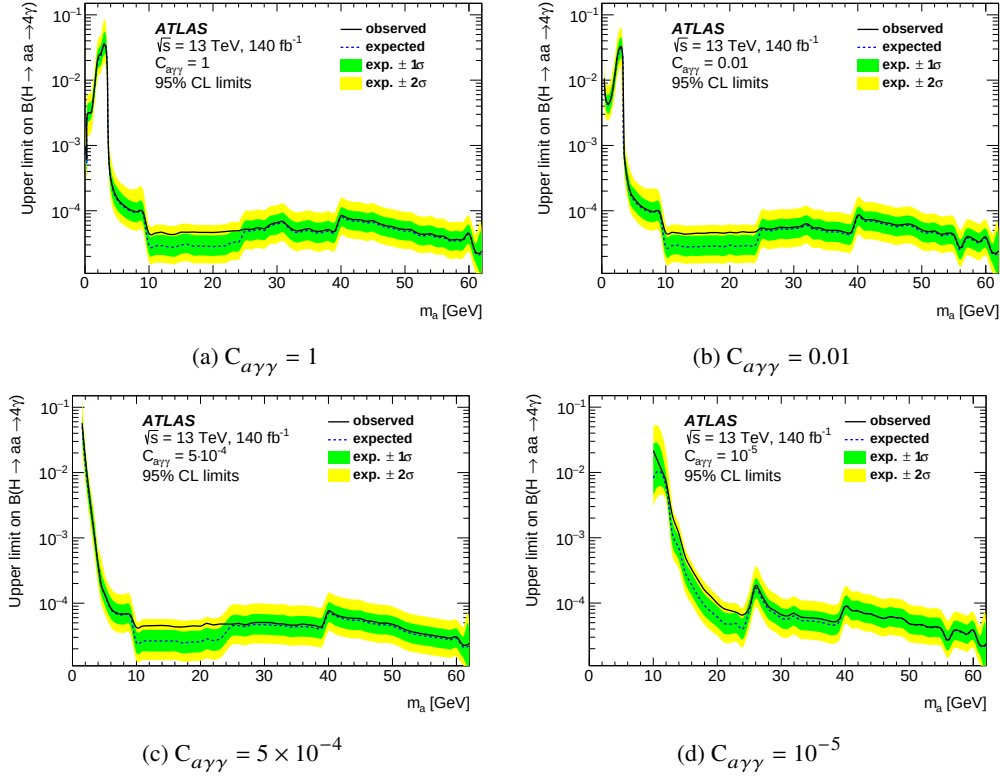


Figure 6.18: Upper limits at 95 % CL on the branching fraction  $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$  for different ALP–photon couplings. Observed limits are shown as solid black lines; expected limits as dashed lines with  $\pm 1\sigma$  (green) and  $\pm 2\sigma$  (yellow) uncertainty bands [1].

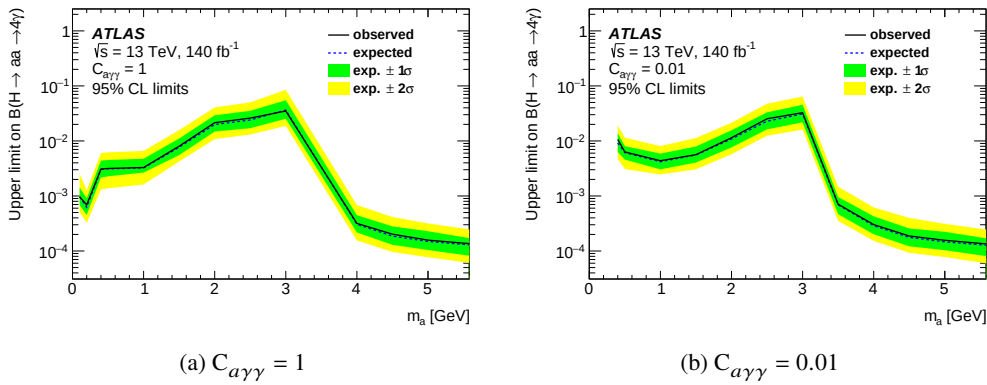


Figure 6.19: Upper limits at 95 % CL on  $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$  for  $C_{a\gamma\gamma} = 1$  and  $C_{a\gamma\gamma} = 0.01$ , zoomed in on the low-mass region  $0.4 < m_a < 5.5$  GeV [1].

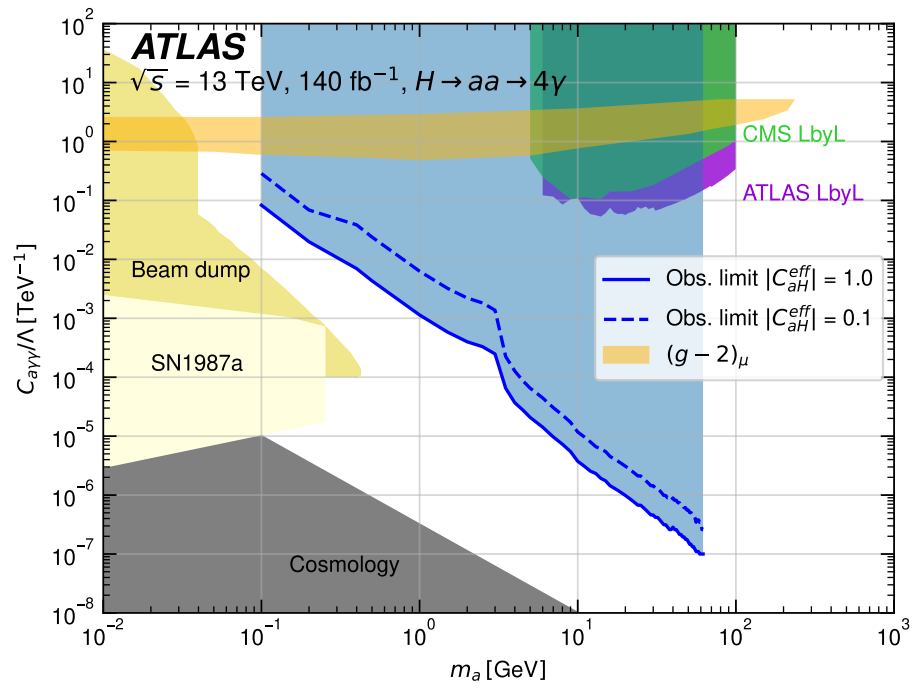


Figure 6.20: Two-dimensional exclusion plot in the  $m_a$ - $C_{a\gamma\gamma}$  plane. The blue shaded area is excluded by this analysis, while the orange band shows the region where ALPs could explain the  $(g-2)_\mu$  anomaly. [1]



## Analysis optimised on promptly decaying ALPs

The high performance of the photon association classifier introduced in Section 6.4.5 enables an alternative analysis approach optimized for promptly decaying ALPs, achieving an almost background-free selection through stricter criteria. This alternative analysis approach is also published with the main analysis in [1].

### 7.1 Selection and background estimation

ALPs with high coupling to photons and a mass  $m_a > 5$  GeV are mostly promptly decaying and well isolated, which makes them perfectly suitable to be assessed by stricter selection with common methods like the tight photon ID. This has the advantage to profit from well known uncertainties and their handling. The sensitivity for prompt photons, is almost completely driven by the 4S category, so it is decided to focus on this category for the explicit prompt photon analysis. To improve the  $\frac{S}{\sqrt{B}}$  ratio, an additional selection criterion is set up on top of the ones described in Section 6.3.2, namely having at least three photons fulfilling the tight photon ID. This leads to a reduction of background events by a factor of 10, while only 20 % of signal gets cut away.

Having so low background, the sideband fit method is not working any more. Thus another methodology has to be found. Using the reconstructed ALP mass, each event can be filled in a two dimensional plane with the invariant 4-photon invariant mass  $m_{4\gamma}$  on the x-axis and the reconstructed  $m_a$  on the y-axis.

In this plane, signal regions can be defined around the Higgs bosons mass and the true value of  $m_a$ . Signal regions are defined as  $m_a - \text{stepsize} \leq m_a^{\text{reco}} \leq m_a + \text{stepsize}$  and  $120 \text{ GeV} \leq m_{4\gamma} \leq 130 \text{ GeV}$ . The size of the step size in  $m_{4\gamma}$  is defined by the resolution of the reconstructed ALP mass. Step sizes increase with  $m_a$ . The used step sizes are shown in Table 7.1.

Assuming the plane is populated uniform, the background can be estimated by defining an area around the signal region as control region, count the events in there and calculate the expected amount of events in the signal region from the ratio of the areas. The control region is defined similar to the signal region, just scaling the stepsize by 1.5. To get a handle for systematic uncertainties of this approach, another control region is introduced by scaling the stepsize with 2.5. The signal region is excluded from both control regions. The definition of signal and control regions in  $m_a^{\text{reco}}$  and  $m_{4\gamma}$  are shown in Table 7.2

Table 7.1: Definition of the stepsize that is used to define signal and control regions for different  $m_a$ .

| $m_a$ range                                | stepsize |
|--------------------------------------------|----------|
| $5 \text{ GeV} < m_a \leq 25 \text{ GeV}$  | 1 GeV    |
| $25 \text{ GeV} < m_a \leq 40 \text{ GeV}$ | 2 GeV    |
| $40 \text{ GeV} < m_a \leq 50 \text{ GeV}$ | 3 GeV    |
| $50 \text{ GeV} < m_a \leq 55 \text{ GeV}$ | 5 GeV    |
| $55 \text{ GeV} < m_a \leq 62 \text{ GeV}$ | 8 GeV    |

Table 7.2: Definition of signal and control regions for the prompt photon analysis. The signal region is excluded from both control regions.

| region                 | size in $m_a^{\text{reco}}$                                                             | size in $m_{4\gamma}$                             |
|------------------------|-----------------------------------------------------------------------------------------|---------------------------------------------------|
| Signal region (SR)     | $m_a - \text{stepsize} < m_a^{\text{reco}} < m_a + \text{stepsize}$                     | $120 \text{ GeV} < m_{4\gamma} < 130 \text{ GeV}$ |
| Control region 1 (CR1) | $m_a - 1.5 \cdot \text{stepsize} < m_a^{\text{reco}} < m_a + 1.5 \cdot \text{stepsize}$ | $110 \text{ GeV} < m_{4\gamma} < 140 \text{ GeV}$ |
| Control region 2 (CR2) | $m_a - 2.5 \cdot \text{stepsize} < m_a^{\text{reco}} < m_a + 2.5 \cdot \text{stepsize}$ | $100 \text{ GeV} < m_{4\gamma} < 150 \text{ GeV}$ |

This method is validated using simulated  $pp \rightarrow 4\gamma$  events. The background estimation for  $m_a = 40 \text{ GeV}$  is shown in Section 7.1. The left-handside plot (a) shows signal and control regions with data events as black dots. The plotrange corresponds exactly to the range of CR2. CR1 is depicted by a red dashed rectangle. The signal region is excluded from both control regions. The background estimate colour coded and written in the SRs rectangle is calculated by applying a transfer factor to the number of events in CR1. The transfer factor is the ratio of areas of SR and CR1. It counts

$$\frac{N_{SR}}{N_{CR1}} = \frac{A_{SR}}{A_{CR1}} \quad (7.1)$$

where  $N_{SR}$  denote the number of events in the signal region, i.e. the background estimate.  $N_{CR1}$  is the number of events in CR1 that can be counted. The alternative background estimate is done in the same way using CR2 instead of CR1. The plot on the right-handside (b) shows the number of events in CR1 and CR2 (bins 1 and 2), the background estimate and the alternative background estimate (bins 3 and 4). Bins 5 and 6 show the statistical and systematic uncertainties where the latter is the difference of background estimation and alternative background estimation. The transfer factors are the ration of areas  $\frac{A_{SR}}{A_{CR1}}$  being shown in bins 7 and 8. Finally the last bin shows the actual number of events in the signal region. It can be seen that the background estimate is reasonable.

The complete two dimensional planes for the simulated  $pp \rightarrow 4\gamma$  (a) and data (b) are shown in Fig. 7.2. All simulated (a) and measured (b) events that fulfil the selection criteria are shown in a 2D plane with  $m_{4\gamma}$  on the x-axis and  $m_a^{\text{reco}}$  on the y-axis. The control regions for  $m_a = 40 \text{ GeV}$ , shown as blue shaded rectangle with the SR in its middle and  $m_a = 10 \text{ GeV}$ , shown as purple shaded rectangle with the SR in its middle are also included. Signal candidates for one of the two exemplary mass hypotheses are depicted as filled black squares, while background candidates are depicted as empty black squares.

In contrast to the analysis aiming for long-lived ALPs, an individual background estimation is done for every mass hypothesis in this search for promptly decaying ALPs. In Fig. 7.3, the background

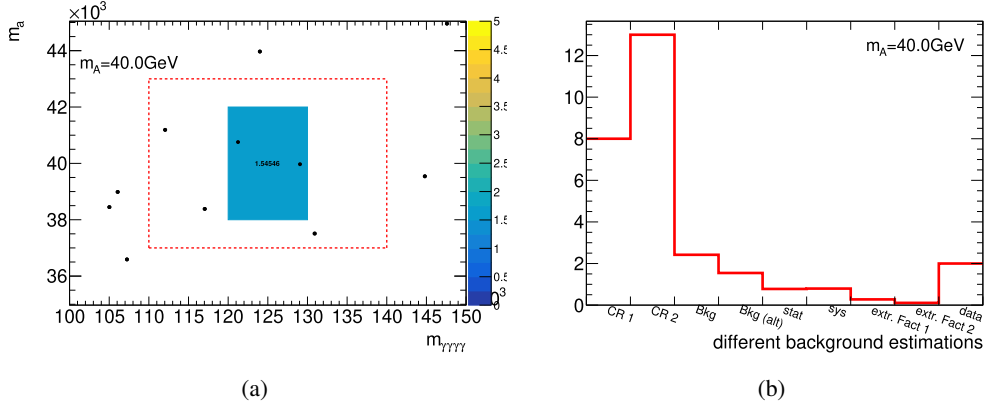


Figure 7.1: 2D background estimation for  $m_a = 40$  GeV. The centre rectangle depicts the signal region with background estimation. CR1 is inside the red dashed rectangle excluding the SR. CR2 includes the whole histogram range excluding the SR. The black dots show simulated events. In the right-handside plot, the yields in CR1 and CR2 are shown, as well as the transfer factors which are used to calculate the background estimate and the alternative background estimate. The difference between background estimate and alternative estimate is used as systematic uncertainty. Statistical uncertainty and number of simulated events in the SR are shown in separate bins.

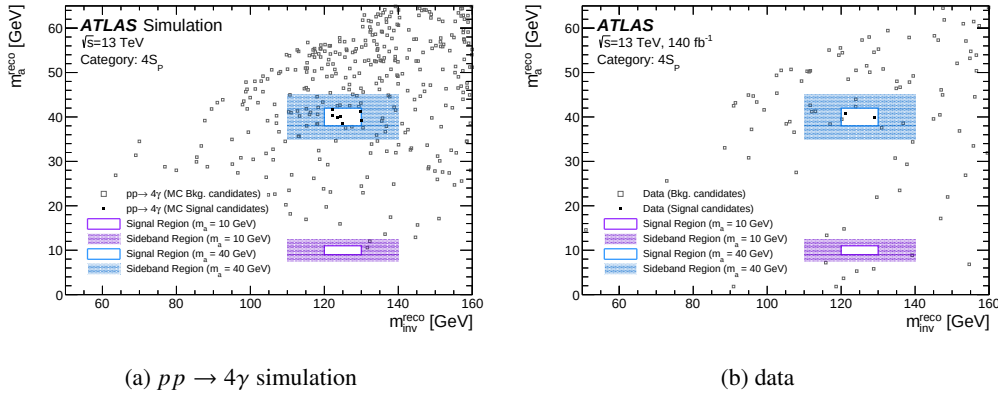


Figure 7.2: All simulated (a) and measured (b) events fulfilling selection criteria in a 2D plane with  $m_{4\gamma}$  on x-axis and  $m_a^{\text{reco}}$  on y-axis. Signal and control regions for  $m_a = 10$  GeV and  $m_a = 40$  GeV are shown exemplary. Signal candidates in the SR of the two examples are filled black squares while background candidates are depicted as empty squares. Control regions are shown as coloured rectangles with the SR surrounded by a solid coloured line in their middle. [1]

estimation is compared to data for several different ALP mass hypotheses over the whole mass range. The comparison for the categories of the search for long-lived ALPs is also included. The background estimation is shown as blue histogram with systematic uncertainties as blue shaded area. In the two photon categories of the search for long-lived ALPs on the left-hand side of the plot, the SM expectation for  $H \rightarrow \gamma\gamma$  is included. The black dots correspond to data. Overall, a good agreement between background estimation and data can be seen.

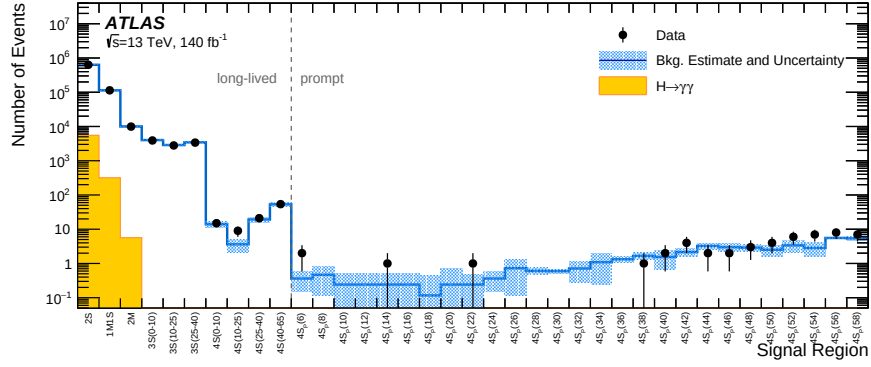


Figure 7.3: Comparison of background estimation and data for categories in the search for long-lived ALPs as well as for different mass hypotheses in the search for promptly decaying ALPs. Background estimations are shown as histogram with a solid blue line and blue shaded areas for systematic uncertainties. Data is shown as black dots.  $H \rightarrow \gamma\gamma$  scaled to the SM expectation are added in the two photon categories as filled yellow histogram.[1]

## 7.2 Statistical analysis and results

In contrast to the search for long-lived ALPs, in the search for promptly decaying ALPs, the whole signal region is treated and fitted as a single bin. This means in particular, that uncertainties don't have an effect on the shape of the distribution. Thus, all uncertainties introduced in Section 6.7.1 are combined to a single normalisation uncertainty together with the systematic uncertainty from the background estimation.

Since only few events contribute to the maximum likelihood fit in the search for promptly decaying ALPs, the asymptotic limit setting might not be accurate. Therefore the limits computed with an asymptotic fit are compared to limits computed using toy data sets. The results for asymptotic fits can be seen in Table 7.3, while the results for fits with toy datasets are shown in Table 7.4. Remarkable differences can be seen, so the toy based limit setting with 20k toys is used to set limits on promptly decaying ALPs. These limits can be seen in Fig. 7.4.

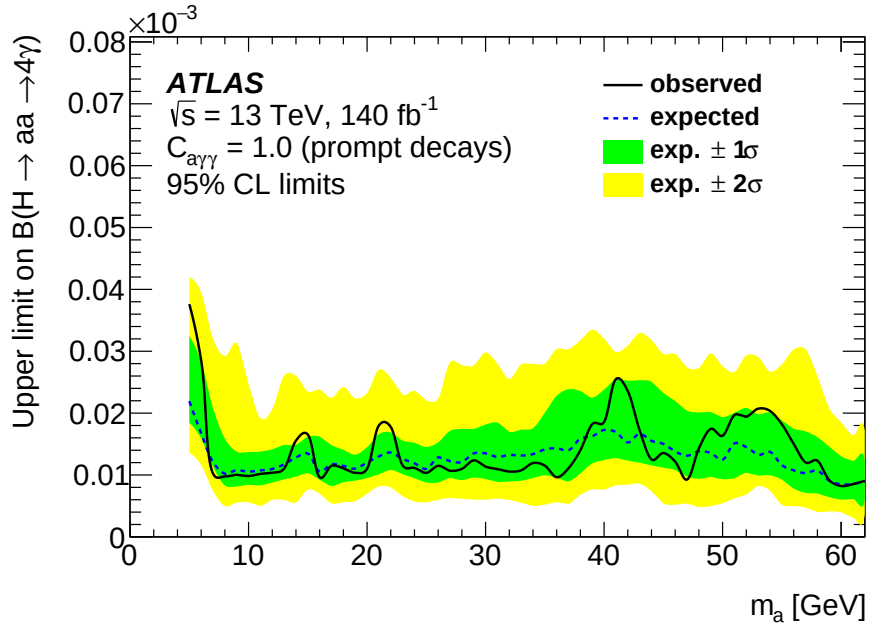
Table 7.3: Limit setting using asymptotic fits.

| $m_a$ GeV | obs   | $-2\sigma$ | $-1\sigma$ | exp   | $+1\sigma$ | $+2\sigma$ |
|-----------|-------|------------|------------|-------|------------|------------|
| 27.0      | 0.069 | 0.032      | 0.049      | 0.083 | 0.151      | 0.314      |
| 27.5      | 0.069 | 0.034      | 0.051      | 0.085 | 0.155      | 0.320      |
| 28.0      | 0.071 | 0.036      | 0.054      | 0.090 | 0.164      | 0.343      |
| 32.5      | 0.070 | 0.038      | 0.057      | 0.092 | 0.166      | 0.343      |
| 33.0      | 0.065 | 0.033      | 0.049      | 0.082 | 0.146      | 0.269      |
| 37.0      | 0.066 | 0.045      | 0.063      | 0.096 | 0.164      | 0.310      |
| 37.5      | 0.090 | 0.108      | 0.124      | 0.157 | 0.221      | 0.418      |
| 42.5      | 0.185 | 0.069      | 0.097      | 0.144 | 0.221      | 0.425      |

The limits on  $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$  found in the search optimised on promptly decaying ALPs are mostly flat at  $\mathcal{B} < 2 \times 10^{-5}$  for masses  $10 \text{ GeV} < m_a < 62 \text{ GeV}$ . Compared to the limits on

Table 7.4: Toybased limit setting using 20k toys.

| $m_a$ GeV | obs   | $-2\sigma$ | $-1\sigma$ | exp   | $+1\sigma$ | $+2\sigma$ |
|-----------|-------|------------|------------|-------|------------|------------|
| 27.0      | 0.093 | 0.085      | 0.089      | 0.096 | 0.121      | 0.183      |
| 27.5      | 0.093 | 0.076      | 0.088      | 0.096 | 0.126      | 0.187      |
| 28.0      | 0.093 | 0.085      | 0.089      | 0.096 | 0.132      | 0.201      |
| 32.5      | 0.093 | 0.083      | 0.090      | 0.095 | 0.134      | 0.295      |
| 33.0      | 0.091 | 0.082      | 0.088      | 0.093 | 0.119      | 0.190      |
| 37.0      | 0.086 | 0.069      | 0.081      | 0.092 | 0.168      | 0.491      |
| 37.5      | 0.084 | 0.072      | 0.079      | 0.090 | 0.207      | 0.491      |
| 42.5      | 0.206 | 0.186      | 0.189      | 0.196 | 0.243      | 0.389      |

Figure 7.4: Upper limits on the branching fraction  $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$  at 95 % CL for promptly decaying ALPs.[1]

$\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$  for long-lived ALPs shown in Section 6.9, the limits for promptly decaying ALPs are more sensitive, as expected.



## Merged photon identification with CNNs

As seen in the results in Section 6.9, the limits for lower mass ALPs are less sensitive. This could probably be improved if the identification of merged photons could be improved. The classifiers used for merged photon identification in [1] are trained using shower-shape variables as input. These are already derived from detector information. During the derivation of the shower shape-variables, dimensions are boiled down and information is lost. If the detector information could directly be used to train the classifier, the merged photon identification could probably be improved significantly.

In section 8.1, an impression on the detector data will be given. Furthermore it will be described what preparation steps are needed to make the data suitable for machine learning using CNNs. Afterwards a CNN classifier to distinguish between single, merged and fake photons will be developed in Section 8.2. Finally it will be explored if the ALP mass can be assessed using CNNs in Section 8.3.

### 8.1 Calorimeter data and data preparation

As described in Section 4.2.1, the electromagnetic calorimeter consists of four layers. The presampler and three layers of sampling calorimeter. Each of the layers has a different granularity. Also the granularity and detector coverage differs with  $\eta$ . Both, different detector geometry and a different angle of incident in the calorimeter are challenging for comparability. Therefore, all studies are done in the central (barrel) region with an  $|\eta| < 1.4$  only.

The granularity in the barrel of the ECal can be found in Table 8.1.

Table 8.1: Granularity of the barrel ECal. Taken from [118]

| Layer      | $\eta$ granularity | $\phi$ granularity |
|------------|--------------------|--------------------|
| Presampler | 0.025              | 0.1                |
| Sampling 1 | 0.003              | 0.1                |
| Sampling 2 | 0.025              | 0.025              |
| Sampling 3 | 0.05               | 0.025              |

Once the data is read, one gets a large array with entries for all calorimeter cells. This data has

to be processed before it can be used for training. First of all, the data is on an event basis, but the goal is to build a classifier on photon level. Therefore, photon candidates are taken and the cells in each layer around the photons  $\eta$  and  $\phi$  are selected. If the position in the detector is kept, this might bias the classifier. Moreover, it is much more challenging having still all calorimeter cells as input with a lot being zero and only few being filled. To avoid this, all photons are centred, such that their centre-of-mass (in this case rather the centre-of-energy) is moved in the  $\eta$ - $\phi$ -plane, such that it is centred at (0,0). All other calorimeter cells have to be moved in the same way, such that every cell still has the same neighbouring cells. This leads to a large 2D plane with energy deposits in the central region and zeros further away. It is tested how large the  $\eta$ - $\phi$  distance of a photon is. All cells with zero entries outside this distance are cut away. After that, four 2D histograms – one for each calorimeter layer – remain with the photon centred at (0,0). But still, the 2D histograms for different layers are differently sized and thus cannot be used to train a CNN. In order to get equally sized input pictures for the training, the input histograms are rebinned such that the ones with coarser granularity are rebinned and extended by adding new bins in the same  $\eta$ - $\phi$  range with the same energy entries as in the large bin. Doing this, input pictures with  $130 \times 20 \times 4$  pixels are created where the  $\times 4$  denotes the four calorimeter layers.

As described in Section 4.2.3, photons can interact with the material in the inner detector and convert to an electron-positron pair. These photon conversions can be reconstructed and treated like regular photons in analyses. However, the electron-positron pair may mimic a merged photon. Therefore, all photon conversions are excluded from further analysis.

After these selections and data preparations, the energy deposit of a photon can be displayed for every layer. A simulated fake photon (i.e. a photon in a simulated  $H \rightarrow aa$  sample without an associated truth photon) can be seen in Fig. 8.1. It can be observed that there is one prong of energy deposition across all layers. Additionally, the energy is widely spread over multiple cells around the energy centre. In comparison to that, a single photon (displayed in Fig. 8.2) has a much narrower energy distribution and only few neighbouring cells measure something. Otherwise, single photons also show only one main spot of energy deposition. This is fundamentally different for a merged photon in Fig. 8.3, where the prongs of energy deposition from the two overlapping photons are clearly visible. Also the energy is wider spread than in the single photon example.

In the comparison by eye, clear differences between the three exemplary photons can be seen. The goal of the following analysis is to check if and how well these can be utilized by a CNN to distinguish between the three kinds of photons.

## 8.2 Photon identification with multiclass CNN

After preparing the detector data such that individual photons can be compared, a CNN classifier shall be trained to distinguish between merged, single and fake photons. In contrast to the classifiers introduced in Section 6.4 only a single classifier shall be used. Instead of having two binary classifiers applied after another, only one classifier doing multiclassification is trained. To be able to do multiclassification, the single outputnode for binary classification has to be replaced by one output node for each class. In this case three output nodes, one for fake photons, one for single photons and one for merged photons are needed.



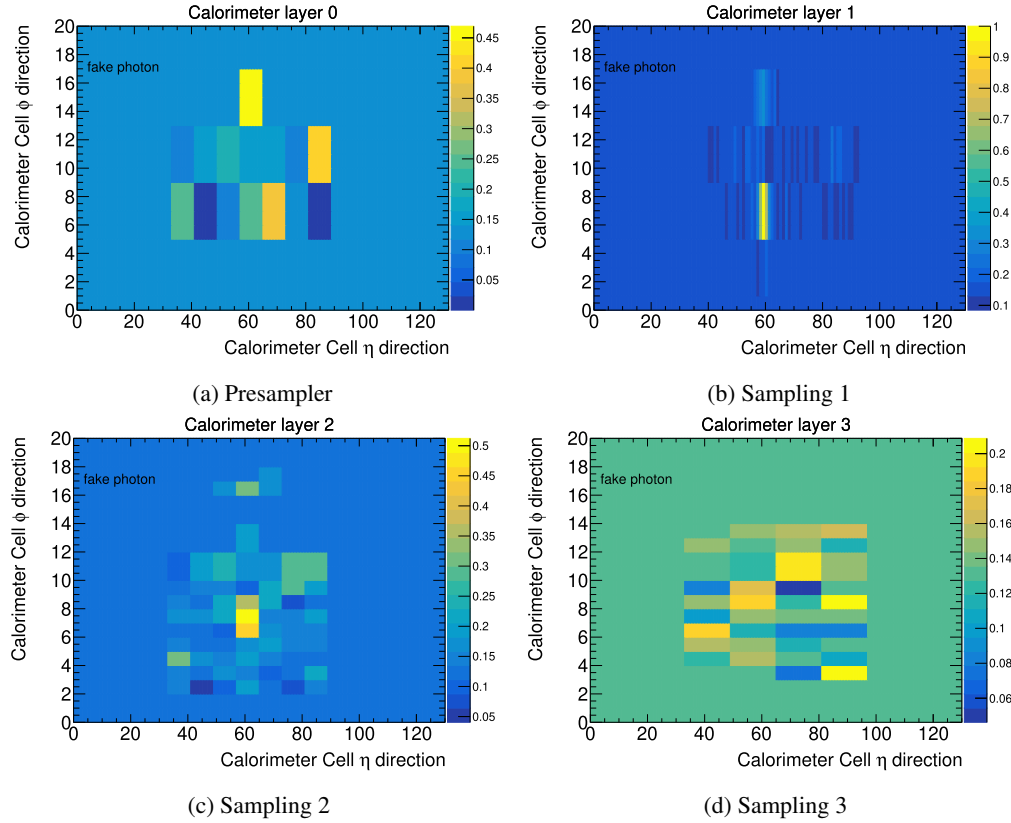


Figure 8.1: A fake photon in all four calorimeter layers. The energy deposits are wide spread over many calorimeter cells, but only one main prong can be seen.

### 8.2.1 CNN architecture

The used CNN is implemented with Keras & Tensorflow [178] and has a rather simple structure. During the data preparation, training or evaluation also some feature of the python packages scikit-learn [179], numpy [180] and pandas [181] are used. It consists of a convolutional layer with 32 filters and a kernel size of  $(3 \times 3)$ . A 'relu' activation function is used. Directly after the convolutional layer, a max pooling layer is implemented. It has a kernel size of  $(2 \times 2)$  and is used to reduce dimensions. After that another convolutional layer with 64 filters, a kernel size of  $(2 \times 2)$  and 'relu' activation function is used. The convolutional layer is again followed by a max pooling layer with a pool size of  $(3 \times 3)$ . The third convolutional layer has 128 filters, a  $(3 \times 3)$  kernel size and 'relu' activation and is followed by the next max pooling layer of size  $(2 \times 2)$ . After these layers, a 'flatten' layer is used to get a 1D array as input for three fully connected dense layers with 100, 25 and 10 nodes and 'relu' activation. Dense layers use a dropout of 0.1 to prevent overfitting. Finally the output layer with its three nodes and sigmoid activation is implemented.

For the convolutional layers a build-in batchnormalisation is used. This ensures, the mean of the output is kept close to zero and the standard deviation close to one. This should improve the stability and performance of the classifier. To avoid problems if input size and kernel size are not compatible, the padding='same' option is used. The CNN uses a categorical crossentropy loss function.

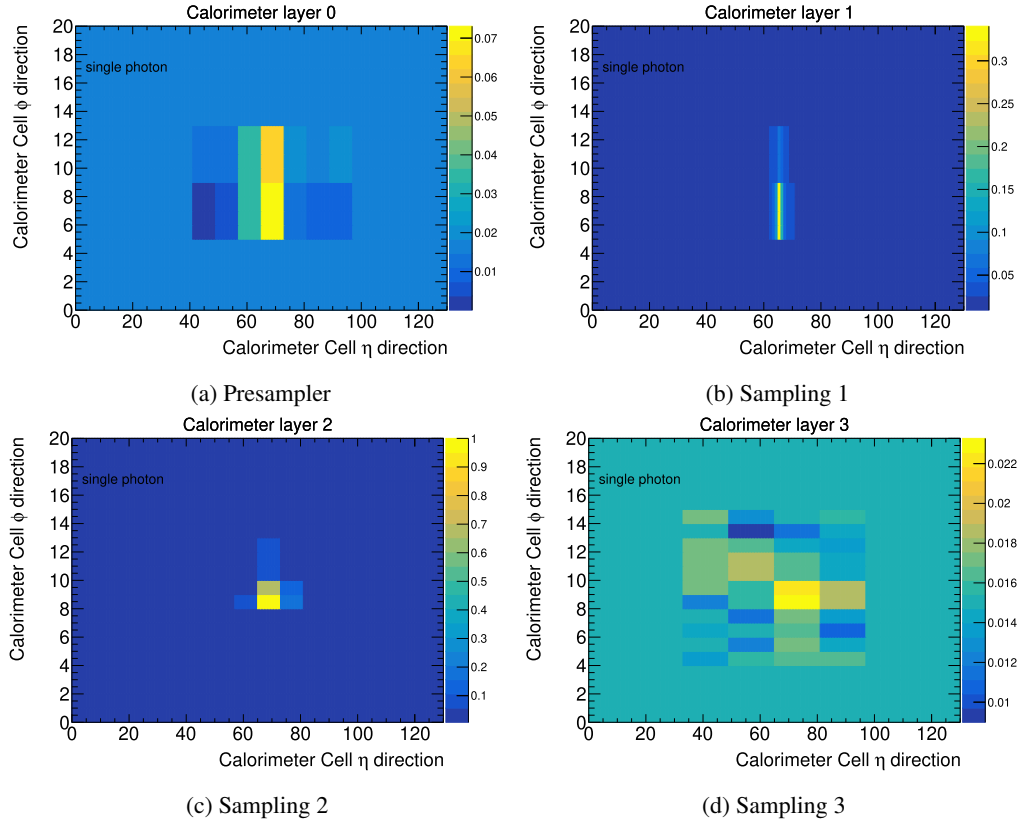


Figure 8.2: A single photon in all four calorimeter layers. Only one prong can be seen and the energy deposit is very narrow, only few neighbouring cells are hit.

### 8.2.2 CNN training

Before starting the training, photons are selected randomly such that there are equal amounts of merged, single and fake photons. Afterwards the energy of each photon is normalized. On one hand, this improves and stabilizes the training performance. On the other hand possible biases due to different energies could be avoided. Afterwards, the events are split in a training, a validation and a test set.

To monitor how well the training is progressing, several measures are observed. The most important are the value of the loss function and the categorical accuracy. The value of the loss function computed for the validation set is also used as an early stopping criterion. If the loss value is not improved for five epochs, the training is stopped. The loss function (a) and the categorical accuracy (b) for each training epoch are shown in Fig. 8.4. The value of the loss function reduces with more epochs for both, training and validation set while the categorical accuracy improves. This is exactly the behaviour that was aimed for.

### 8.2.3 CNN evaluation

To assess the effectiveness of the training, several evaluation criteria can be considered. As a first criterion, the output values in each of the output nodes can be regarded. These are shown in Figs. 8.5 to 8.7. Photons from the training set are depicted with dots, while photons from the test set are shown

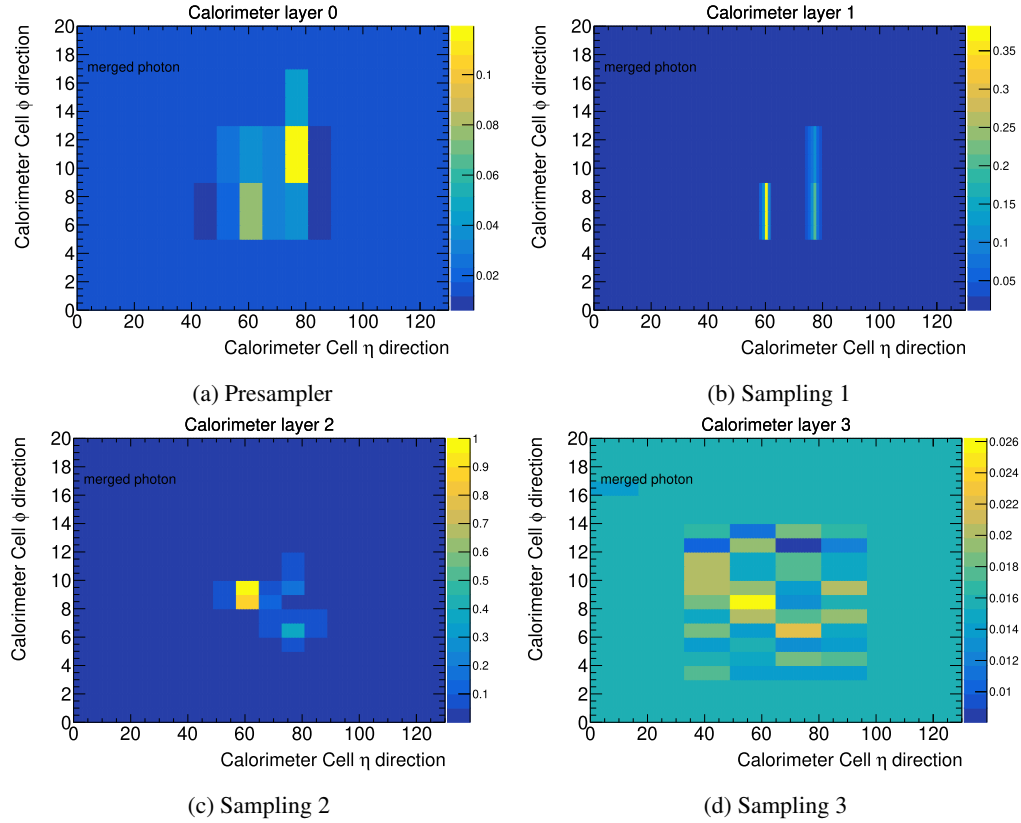


Figure 8.3: A merged photon in all four calorimeter layers. Two clear prongs can be seen and the energy deposits are spread across multiple calorimeter cells.

as filled histogram. Fake photons are coloured blue, single photons are yellow and merged photons red. A good separation is observed. In each node, the output for the corresponding photon class is shifted towards one, while the output for the other two classes is shifted towards zero. In the ratio plot, it can be seen, that the shape for training and test photons of each class is similar and a good closure is observed. If there would be differences between the output shapes for training and test, this might be a hint for overfitting.

For binary classification, the receiver-operating-characteristic (ROC curve) is a popular metric. The definition of a ROC curve for multiclassification is not so easy, but separate ROC curves for each output node are shown in Fig. 8.8. Not only the curve itself, but also the area under the curve (ROC-AUC) is a good measure to check and compare the performance of a classifier. In this case, the ROC curves are very promising with a ROC-AUC of 99 %.

Not only the shape of the output distribution is of interest, but in particular if the predicted class is correct. To get the class prediction out of the values, for each photon, the node with the highest output is selected and the class of that node is assigned to the photon. How well the multiclassification works can be visualised with a confusion matrix, where the truth class is on the x-axis and the predicted class on the y-axis. The confusion matrices for the training and the test sample are shown in Fig. 8.9. For a perfect training, only the diagonal elements are filled with 100 %. It can be seen, that the results come actually quite close.

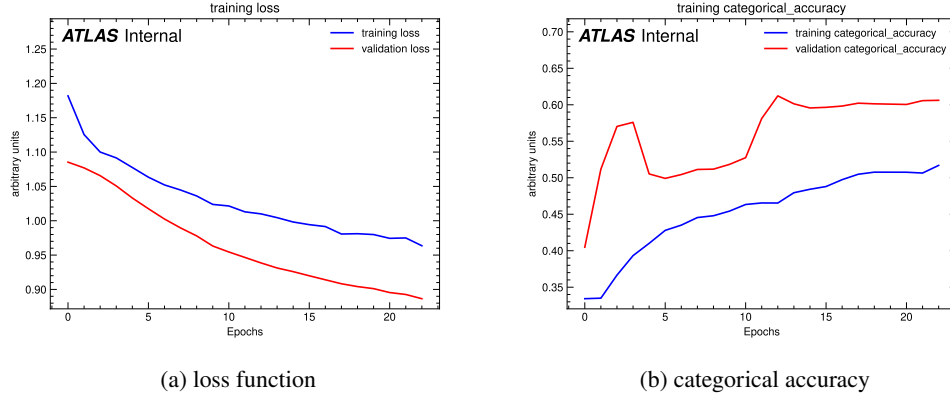


Figure 8.4: Loss function and categorical accuracy per epoch for training set in blue and for the validation set in red.

### 8.2.4 Comparison with a shower-shape based classifier

Since the multiclass CNN classifier cannot be directly compared to the binary classifiers of the long-lived ALP analysis, a multiclass ANN classifier is trained on shower-shape variables. The setup is quite similar to the one of the CNN, but instead of a CNN, an ANN is used and instead of the calorimeter cell information, the shower-shape variables are used.

Evaluating the ANN classifier, the closure plots for the output nodes in Fig. 8.10 show worse results than the CNN. Even tho the closure is quite good. Especially single and fake photons are not well separated by the ANN. Their shapes are very similar in every of the three plots.

### 8.2.5 $p_T$ dependence

As described in Section 6.4, the classification performance of the CNN is further investigated as a function of the photon transverse momentum ( $p_T$ ). Instead of modifying the training sample to equalize the  $p_T$  distributions across classes, a weighting scheme is applied to each feature vector during training.

The  $p_T$  range is divided into five bins. For each bin  $i$ , a weight is initially assigned as

$$w_i = \frac{n_i^{\text{cls}}}{n_i^{\text{all}}},$$

where  $n_i^{\text{cls}}$  denotes the number of feature vectors for a given class, and  $n_i^{\text{all}}$  is the total number of feature vectors in that bin. Each weight is then normalized by the average weight across all bins, such that the mean of the weights equals one. This normalization helps maintain training stability and performance.

Figure 8.11 shows the classification efficiency for each class as a function of  $p_T$ . The efficiency for

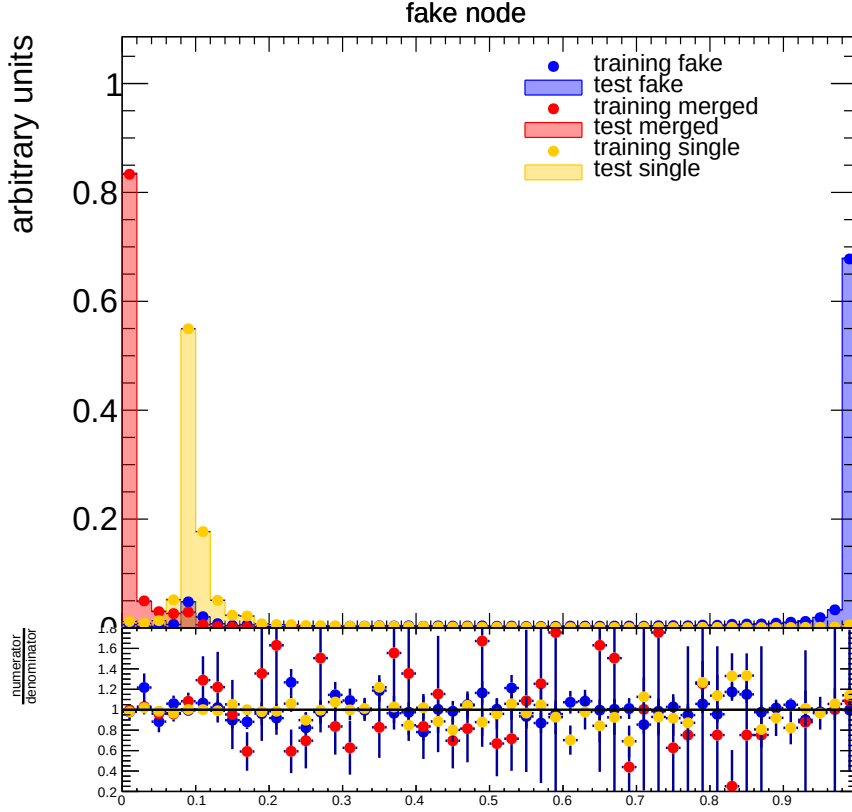


Figure 8.5: Classifier output in the fake photon output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.

class cls in each  $p_T$ -bin is defined as:

$$\text{efficiency}_{\text{cls}} = \frac{\sum_{i=1}^{N_{\text{evts}}} \left[ (\text{pred}_i = \text{cls}) \wedge (\text{true}_i = \text{cls}) \right]}{\sum_{i=1}^{N_{\text{evts}}} (\text{true}_i = \text{cls})} \quad (8.1)$$

That is, the number of correctly classified events for a given class divided by the total number of true events in that class.

The red, blue, and yellow curves represent the efficiencies for merged, fake, and single photons, respectively. The left plot shows the performance using equally sized training samples per class, but without applying  $p_T$ -based weights. The efficiency for merged photons remains high and relatively flat across the full  $p_T$  range.

The right plot applies the  $p_T$  reweighting described above. While the efficiency for merged photons

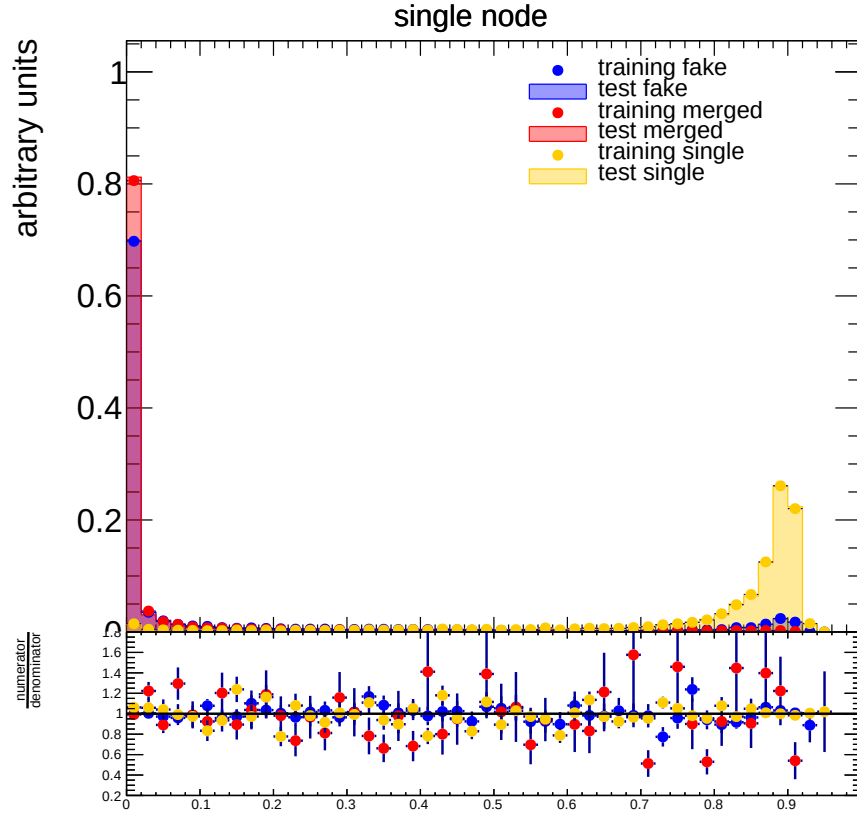


Figure 8.6: Classifier output in the single photon output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.

remains high, the performance for single and fake photons degrades and becomes increasingly  $p_T$  dependent. This behavior is contrary to expectations: the reweighting was intended to equalize  $p_T$  dependence, but instead appears to introduce additional variation.

These tests have been repeated with different setups. For example, photons with  $p_T$  below 15 GeV were excluded, or all photons passing the selection criteria were used – leading to approximately 900 k single, 300 k fake, and only 33 k merged photons. In such cases, the efficiency for single photons was notably higher, while that for fake and merged photons was lower. Despite the reweighting scheme equalizing the total sum of weights per class, the model shows a preference for classifying photons as single. Once again, the efficiency shows increased  $p_T$  dependence when reweighting is applied.

Plots corresponding to these additional tests are included in the appendix (see Appendix A.6).

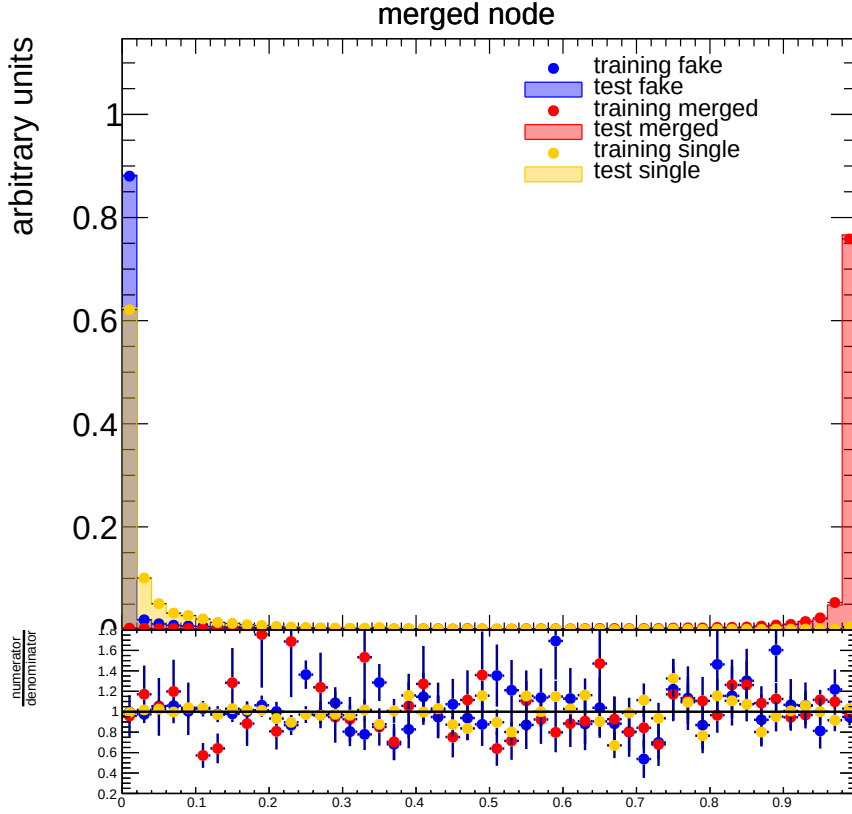


Figure 8.7: Classifier output in the merged photon output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.

### 8.3 ALP mass regression with a CNN

A major advantage of the three- and four-photon categories in the search for long-lived ALPs is that these allow for full ALP reconstruction, making the mass  $m_a$  accessible. The reconstructed mass  $m_a^{\text{reco}}$  can be used as a powerful discriminant to suppress background significantly.

If a similar approach were applicable to the two-photon categories, the sensitivity – particularly in the low-mass region – could potentially be improved. However, in these events only a single photon is reconstructed per ALP, making it impossible to directly reconstruct the ALP mass from its decay products.

To address this, a regression-based approach using machine learning can be considered. Instead of classifying events into discrete categories, a regression CNN can be trained to predict a continuous variable – in this case, the ALP mass  $m_a$ . In principle, such a network could infer information related to the opening angle or the energy sharing between the two photons that are too closely spaced to be resolved separately. These learned features could then allow an estimate of  $m_a$ .

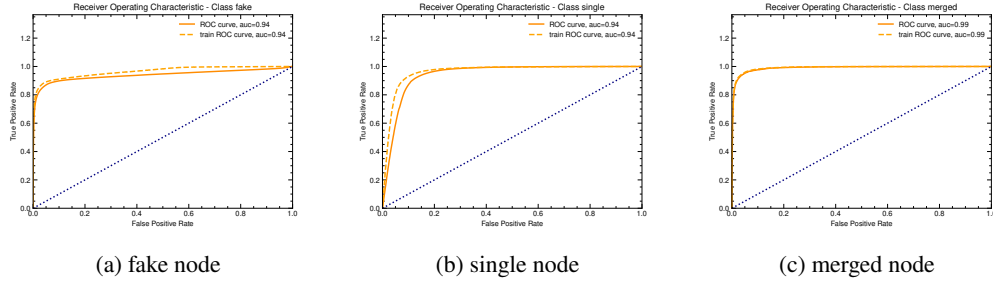


Figure 8.8: ROC curves for every output node.

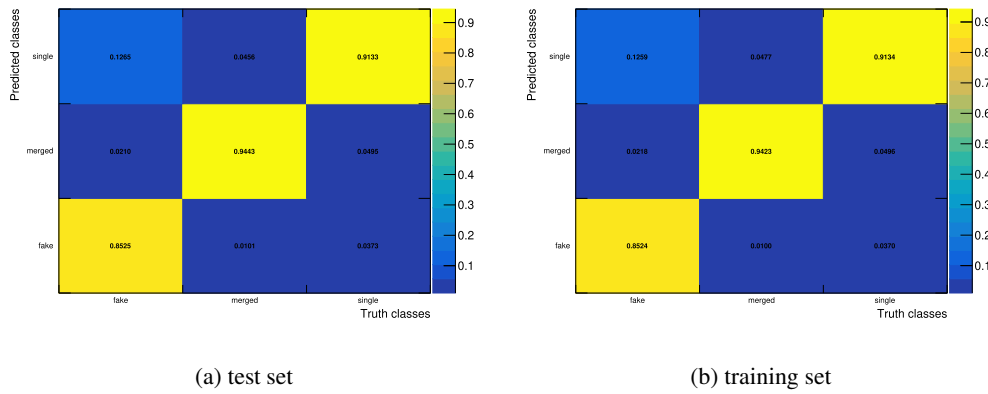


Figure 8.9: Normalized confusion matrices with the truth classes on the x-axis and the predicted classes on the y-axis.

For this regression task, the same CNN architecture described in Section 8.2.1 and illustrated in Fig. A.7 is employed. The input consists of calorimeter information of merged photon signatures, preprocessed as described in Section 8.1. The target is the true ALP mass  $m_a$  associated with the merged photon, normalized during training to improve convergence. The predicted values are later rescaled to the original mass range.

Fig. 8.12 shows a two-dimensional response matrix for the trained regression CNN. The y-axis corresponds to the true ALP mass, while the x-axis shows the predicted values  $m_a^{\text{reco}}$ . For better visualization, each row is normalized such that the total response per true mass is unity. Ideally, all entries would lie on the diagonal, indicating perfect predictions. Deviations from the diagonal reveal the CNN's resolution, as well as potential biases.

It can be seen that most predictions cluster around the diagonal, with no significant systematic offset – indicating minimal bias. Overall, the performance appears promising.

Fig. 8.13 shows the regression output distributions for selected ALP mass values. Each legend entry includes the mean and RMS of the predicted values. The means closely match the true values, while the RMS increases with mass, reflecting a degradation in resolution at higher  $m_a$ .

The mean and RMS values (with RMS shown as error bars) for this training are also depicted in Fig. 8.14 (a). On the right-hand side (b), the residuals are shown. In an ideal training, the first plot should exhibit a line through the origin with a slope of 1, and the second should display a flat line at 0.



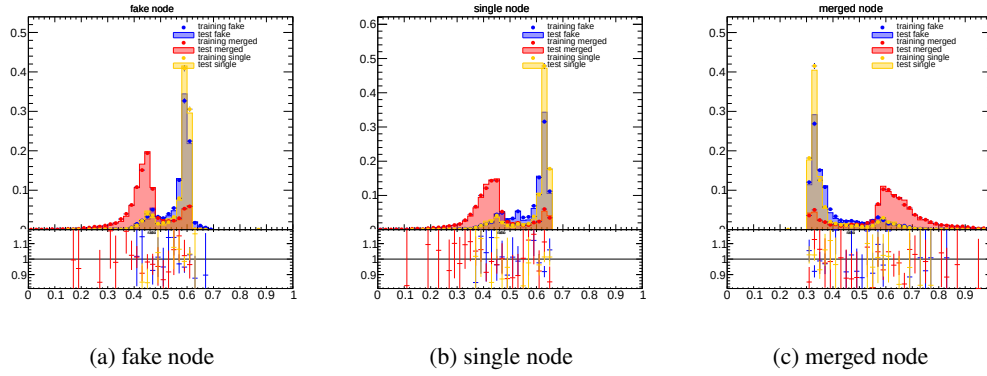


Figure 8.10: Classifier output in each output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.

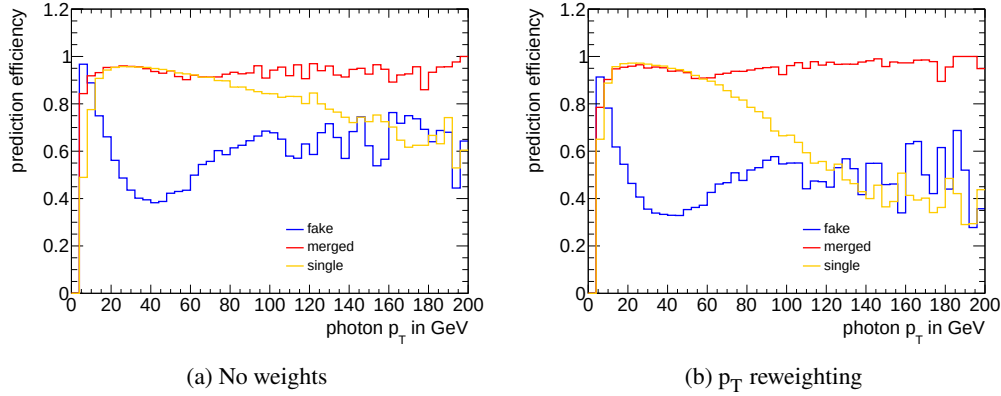


Figure 8.11: Classification efficiency per class as a function of photon  $p_T$ . Panel (a) uses equal training sample sizes without  $p_T$ -based weights. Panel (b) shows results after applying  $p_T$  reweighting.

However, a clear upward trend is visible in the residual plot. This is fitted with a linear function

$$f(x) = a + b \cdot x. \quad (8.2)$$

Using the fit results, it may be possible to improve the regression output by applying the correction

$$\text{prediction}_{\text{corrected}} = \text{prediction} + a + b \cdot \text{truth}. \quad (8.3)$$

This can be done during evaluation on data where the truth value for each entry is known. To enable application on unlabelled data, the formula must be slightly modified to

$$\text{prediction}_{\text{corrected}} = \text{prediction} + a + b \cdot \text{prediction}. \quad (8.4)$$

However, this approach carries the risk of amplifying existing biases.

Fig. 8.15 compares both correction methods for labelled data. The plot on the left-hand side shows

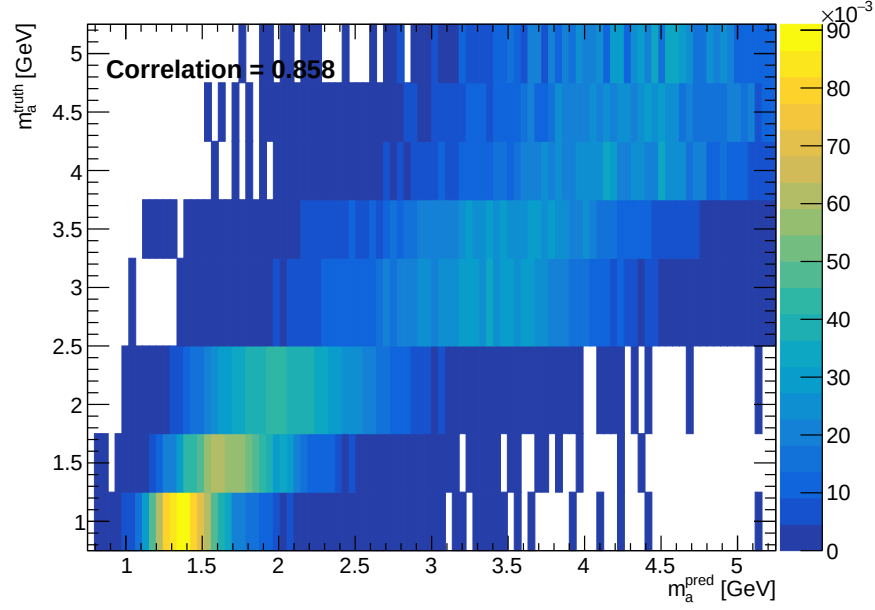


Figure 8.12: Two-dimensional response matrix showing the correlation between the CNN-predicted (x-axis) and true values (y-axis) of the ALP mass  $m_a$ . For better visualisation, each row is normalised to 1 (i.e. for each ALP mass the same amount of entries is shown). A diagonal line indicates perfect agreement; deviations from this line reflect the regression performance, including bias and resolution across the mass range.

a comparison between the uncorrected results (mean with RMS as error bars) and the results after applying corrections based on the true ALP mass or the predicted value. It can be seen that the correction based on the true value generally performs better. The correction using the predicted value consistently lies between the other two curves. This suggests that, in this particular case, applying such a correction cautiously may be justified. The right-hand side plot (b) displays the regression matrix after applying the correction based solely on the predicted value.

At first glance, it may seem surprising that the model does not automatically learn an appropriate correction during training. However, the CNN is trained to minimise the average prediction error, and not specifically to produce unbiased predictions across the entire target range. As a result, for regression tasks with a limited and sharply bounded range (such as in this case, where negative masses are unphysical), it may be more beneficial in terms of loss minimisation to shift extreme predictions toward the centre of the allowed region.

Moreover, the training data are not uniformly distributed across the target range. The total number of training events sorted by  $m_a$  can be found in the appendix in Table A.3.

Basic reweighting strategies have been tested, but due to the sparsity of data near the edges, no significant improvements were observed. Additionally, the training data are spaced in discrete steps of  $m_a = 0.5$  GeV, which may be too coarse to enable the network to learn a smooth regression function. Evidence for this can be seen in the residual plot: for example, there is no training data for ALP masses around  $m_a = 2.5$  GeV, which may lead the network to systematically underestimate predictions at  $m_a = 3$  GeV and overestimate them at  $m_a = 2$  GeV.

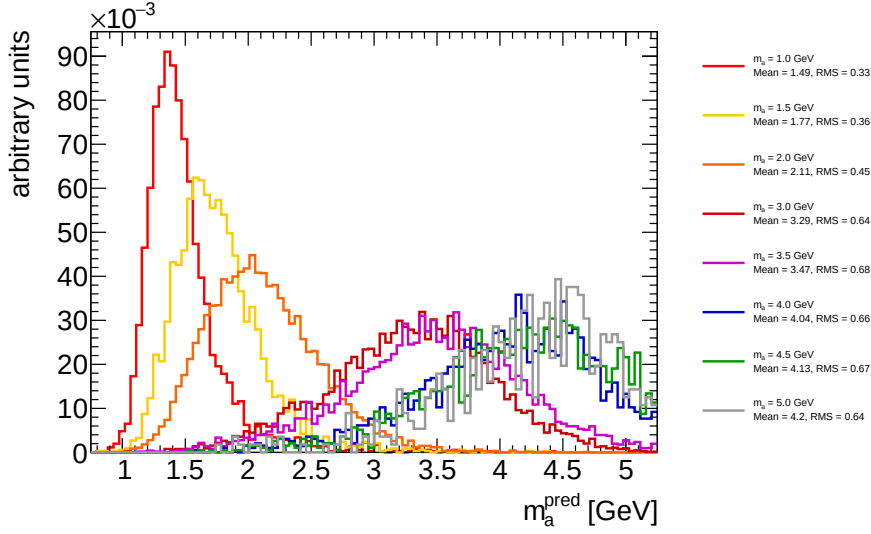


Figure 8.13: Regression CNN output for different masses. Each distribution is labeled with its mean and RMS.

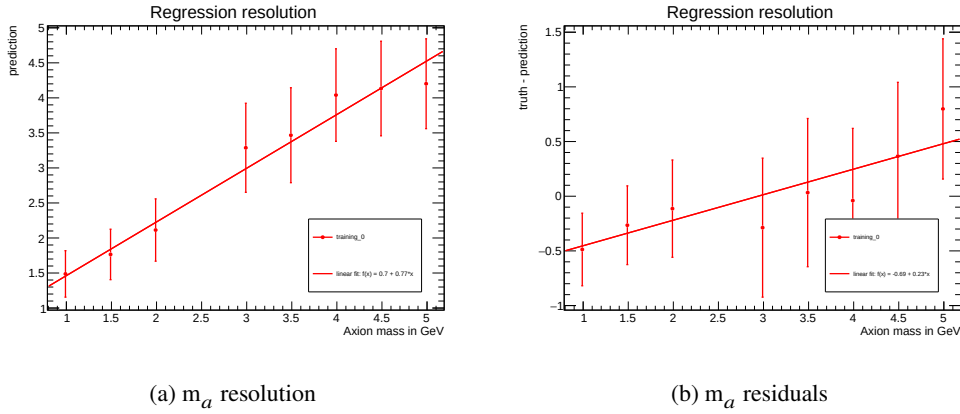


Figure 8.14: Mean and RMS of the regression results for different truth ALP masses (a) and the residuals of  $\text{Mean} - m_a^{\text{truth}}$  for different truth masses (b). A linear fit was performed on (b).

One possible improvement would be to retrain the regression on a dataset with more merged photons and finer spacing between mass points. Alternatively, regularisation techniques could be explored, although initial tests have not shown significant gains. A more targeted approach may involve modifying the loss function or introducing an adversarial component that penalises global biases.

Across multiple trainings, a consistent trend is observed: the predicted mass range tends to contract slightly – low true masses are often overestimated, and high masses underestimated. This behaviour likely reflects the CNN’s difficulty in learning extreme values due to limited statistics and its tendency to favour predictions near the centre of the training distribution. The RMS of the CNN output increases with true ALP mass, indicating degraded regression resolution at higher masses. This is expected, as the decay photons become more widely separated, increasing the variability in shower morphology. At low masses, by contrast, the photons are highly collimated and often resemble single-photon showers, which limits the network’s ability to extract mass-dependent features. Despite these limitations,

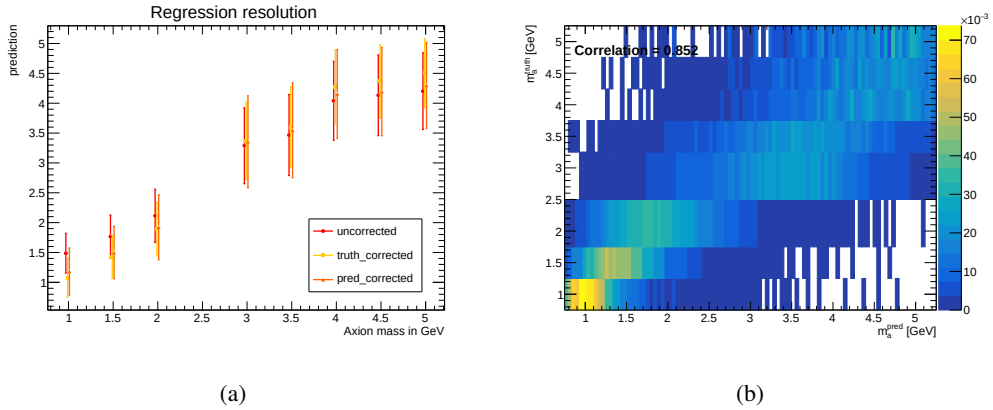


Figure 8.15: Comparison of correction methods and the uncorrected regression result (a). (b) shows the regression matrix with the correction based on the prediction itself applied. It can be seen, that the correlation becomes slightly better compared to the original figure.

within the most relevant mass window for merged photon signatures,  $1 \text{ GeV} \leq m_a \leq 5 \text{ GeV}$ , the CNN performs reliably: biases remain typically below 0.5 GeV, and RMS values stay below 1 GeV.

To better interpret the CNN output in presence of background, Fig. 8.16 compares the output distributions for fake photons, single photons, and merged photons. These samples use the same MC dataset employed for training the multiclass CNN.

Merged photons exhibit a broad distribution extending up to 8 GeV. Single photons, shown in dark brown, produce a sharp peak at low predicted masses, consistent with expectations: at very low  $m_a$ , the two decay photons are highly collimated and mimic a single-photon signature. Fake photons (light brown) yield a broader, lower-mass distribution. This may reflect less distinct shower shapes, lacking the double-prong structure associated with merged decays, and thus limiting the CNN’s ability to infer an opening angle.

### Notes on training and stability

While the multiclass CNN showed stable performance across different initializations and hyperparameters, the regression task is more sensitive to training conditions. Results from various runs with fixed settings but different seeds are discussed in Appendix A.7. Further optimization of architecture and hyperparameters may be necessary, or alternatively, one may train an ensemble of models and select the one best matching the desired performance criteria.

## 8.4 Future studies

This section presented convolutional neural networks (CNNs) trained on low-level calorimeter data from the ATLAS detector to classify electromagnetic signatures as single, merged, or fake photons. In addition, a regression-based CNN was developed to estimate the ALP mass  $m_a$  in the two-photon categories, where direct reconstruction is not feasible.

The multiclass CNN significantly outperforms a comparable artificial neural network (ANN) in distinguishing photon topologies. Likewise, the ALP mass regression yields promising results: the

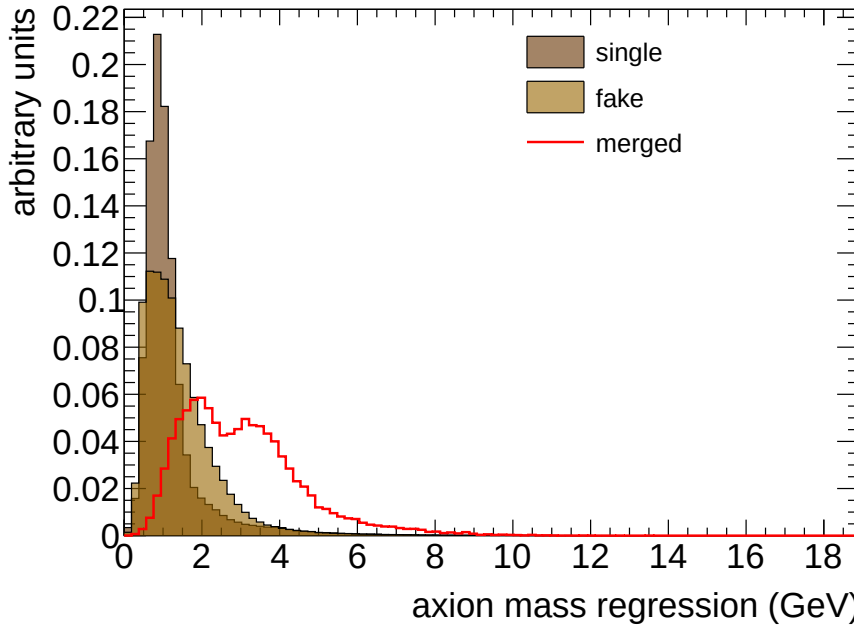


Figure 8.16: Comparison between regression results for fake (light brown), single (dark brown), and merged (red line) photons. The CNN predicts higher mass values primarily for merged photon signatures.

predicted values show good correlation with the truth, with only minor biases and a mass-dependent resolution.

The  $m_a$  regression output enables further background reduction in the merged-photon category by splitting the invariant mass spectrum into bins of predicted mass,  $m_a^{\text{regr}}$ . While this technique is less applicable to the other two-photon categories, derived variables such as the difference between photon-level  $m_a^{\text{regr}}$  estimates could offer additional discriminating power and improve sensitivity.

In an effort to improve classification stability across different photon  $p_T$  ranges, various  $p_T$  reweighting strategies were investigated. However, reweighting did not yield the expected gains; instead, it often degraded performance, particularly for fake and single photons, and introduced undesired  $p_T$  dependencies. These effects could stem from increased training instability, insufficient sample statistics in rare bins, or overcompensation in the loss function.

At present, the most robust and effective training strategy appears to be the simplest: selecting a balanced subset of training events such that each class has the same number of input samples. This baseline approach provides stable performance across the full  $p_T$  range without introducing additional bias.

Future improvements may be achieved by further optimizing the CNN architecture and hyperparameters. Beyond traditional CNNs, advanced models such as graph neural networks (GNNs) could be explored to better capture spatial correlations in the calorimeter. In addition, alternative weighting techniques such as focal loss, stratified sampling, or adaptive batch construction could be tested to address imbalances in  $p_T$  or class distribution more effectively.

Another interesting avenue is to extend the regression task to estimate the decay displacement of long-lived ALPs in addition to their mass. This would be more challenging than mass regression,

as it requires disentangling kinematic effects and learning from more complex spatial and temporal patterns. A possible strategy could involve training separate CNNs for different mass hypotheses or designing a two-dimensional regression targeting both mass and displacement simultaneously.

So far, the training has been restricted to photons in the central region of the detector. In the high- $\eta$  regions, such as the endcaps and especially the transition region between the barrel and endcap, both the kinematic configuration and detector geometry are significantly different. These effects are expected to alter the photon shower signatures in the calorimeter substantially.

As a result, alternative classification strategies may be required. One straightforward approach would be to divide the data into  $\eta$  regions and train separate CNNs (or other classifiers) for each region. However, this requires careful validation and sufficient training statistics in each bin to avoid overfitting or bias.

Additionally, all events involving photon conversions have so far been excluded from the training. When photons convert to  $e^+e^-$  pairs within the inner detector, track information becomes available and could be leveraged by the model. This could potentially help locate the ALP decay vertex or directly identify merged photon signatures – particularly in cases where both photons convert.

However, a CNN trained solely on calorimeter data may be insufficient to classify such events accurately, even with a model specialized on converted photons. Integrating tracking information into the input could improve classification and enable robust handling of conversion events.

It is also essential to quantify uncertainties associated with CNN predictions. Techniques such as Bayesian neural networks, dropout-based uncertainty estimation, or ensemble methods could help to improve the methodology introduced in Section 6.7.1.

So far, all results are based exclusively on simulated data. As a crucial next step, the models should be validated using control and validation regions in real collision data to assess their robustness and performance in realistic conditions.

---

## Conclusion

---

After many decades of extensive research, the discovery of the Higgs boson in 2012 by CMS and ATLAS marked the final missing piece of the Standard Model (SM) of particle physics. However, despite its tremendous success, the SM cannot explain several observed phenomena.

One elegant extension of the SM involves the axion or, more generally, axion-like particles (ALPs). These hypothetical pseudo-scalar particles could resolve the strong CP problem – avoiding fine-tuning in QCD – as well as account for the long-standing discrepancy between the measured and predicted anomalous magnetic moment of the muon  $((g - 2)_\mu)$ .

This thesis presents a search for ALPs in the exotic decay of the Higgs boson into a pair of ALPs, with a final state of four photons ( $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ ). This channel provides sensitivity across a wide phase space, with ALP masses ranging from a few MeV up to half the Higgs boson mass, and covering over five orders of magnitude in the ALP-photon coupling strength. Notably, this includes much of the region that could explain the  $(g - 2)_\mu$  anomaly.

The broad phase space results in diverse kinematic signatures. For low ALP masses, the decay photons are highly boosted and often reconstructed as a single photon due to their collimation. For low couplings  $C_{a\gamma\gamma}$ , the ALPs acquire a significant lifetime, causing the decay to photons to occur at displaced vertices.

This analysis uses data collected with the ATLAS detector at the LHC from 2015 to 2018. A dedicated neural network classifier was trained to distinguish between merged and single (fake) photons. Events are categorized into six topologies based on the photon reconstruction: 2 single photons, 1 single and 1 merged photon, 2 merged photons, 3 single photons, 2 single + 1 merged, and 4 single photons. The background in each category is estimated using a fully data-driven sideband fit method.

The invariant mass spectrum of simulated signal events at discrete  $(m_a, C_{a\gamma\gamma})$  points, together with the estimated background in the most sensitive categories, is analyzed using maximum likelihood fits. A full suite of systematic uncertainties is included, such as spurious signal contributions in the background model and custom uncertainties related to the photon classifier and the reconstruction of displaced decays. This constitutes the first analysis targeting non-prompt ALPs with small couplings.

No significant excess is observed across the  $(m_a, C_{a\gamma\gamma})$  phase space. Consequently, upper limits are set using the  $CL_S$  method. These are translated into exclusion regions in the  $(m_a, C_{a\gamma\gamma})$  plane, excluding a substantial portion of the parameter space in which ALPs could explain the  $(g - 2)_\mu$  discrepancy.

A complementary search focusing exclusively on promptly decaying ALPs is also performed, using stricter selection criteria. Artificial neural networks (ANNs) are employed to assign the correct combination of decay photons to the ALP. With the correct combination of decay photons, the ALP mass can be calculated in events with more than three photons. The decay photons are combined to form  $m_a^{\text{reco}}$  and  $m_H^{\text{reco}}$ , and a two-dimensional background estimation approach is developed. This leads to the most stringent limits on the  $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$  channel to date.

In the final part of this thesis, novel machine learning techniques are explored. Convolutional Neural Networks (CNNs), well-suited for spatially structured data, are trained on electromagnetic calorimeter readout information. The calorimeter energy deposits in all 4 layers around each photon are transformed into  $(130 \times 20 \times 4)$  images and used to classify photon types. The CNN outperforms a comparable ANN trained on shower-shape variables.

Additionally, merged-photon input data are used to train another CNN aimed at regressing the ALP mass from a single merged-photon signature. This method shows promising results, enabling finer  $m_a$  resolution not only in the 3S and 4S categories but also in the two photon categories. This could help to significantly reduce the background for some ALP masses.

Finally, potential avenues for further improving future ALP searches – both in methodology and experimental sensitivity – are discussed.

Recent lattice QCD calculations have substantially reduced the discrepancy between theoretical predictions and experimental measurements of the muon’s anomalous magnetic moment. Moreover, the analysis presented in this thesis yields no further evidence supporting the last remaining axion model capable of explaining this anomaly. It can therefore be concluded that the prospects for discovering new physics in this specific channel are limited.



## Appendix

### A.1 Background estimation for the three and four photon categories

The background estimation for the 3S and 4S categories follows the same approach. As mentioned earlier, at least one ALP can be reconstructed in these categories and thus a reconstructed ALP mass can be calculated. Consequently, the invariant mass spectrum can be divided into bins of the reconstructed ALP mass. These bins are

- $3.5 \text{ GeV} < m_a \leq 10 \text{ GeV}$
- $10 \text{ GeV} < m_a \leq 25 \text{ GeV}$
- $25 \text{ GeV} < m_a \leq 40 \text{ GeV}$
- $40 \text{ GeV} < m_a \leq 62 \text{ GeV}$

The signal regions in the invariant mass spectra are defined such that they contain at least 80 % of each signal distribution. This criterion ensures sufficiently wide sideband regions with minimal signal contribution. For all 4S histograms, the blinded signal region spans  $120 < m_{4\gamma} < 130 \text{ GeV}$ .

In the 3S categories, the signal distributions are broader and shifted towards lower invariant masses. This behaviour is expected as one of the photons is often outside the detector acceptance, resulting in missing energy and a reduced reconstructed mass. The effect becomes more pronounced for larger reconstructed ALP masses.

In the mass range  $40 \text{ GeV} < m_a < 62 \text{ GeV}$ , defining suitable sideband regions for the 3S category is particularly challenging due to significant signal contributions and potential contamination from Z boson decays. Sensitivity studies were performed using maximum likelihood fits on pseudo-data, considering different fitting ranges to estimate the background in the 3S category and including only statistical uncertainties. The pseudo-data was generated using the alternative background model. Tests were conducted using only the  $4S_{(40-62 \text{ GeV})}$  category, only the  $3S_{(40-62 \text{ GeV})}$  category, and both combined. The results clearly indicate that the analysis sensitivity in this mass range is dominated by the 4S category. Consequently, the 3S category is excluded from the analysis for this range of  $m_a$ .

The defined signal regions, along with the nominal and alternative sideband fitting functions, are summarized in Table 6.6. For all cases in the 4S category, the fit using a reduced mass range yielded better results than the nominal fit range; thus, the reduced range is adopted as the new nominal. To

estimate systematic uncertainties, rather than further reducing the range, the original (previously nominal) fit range is used as an alternative.

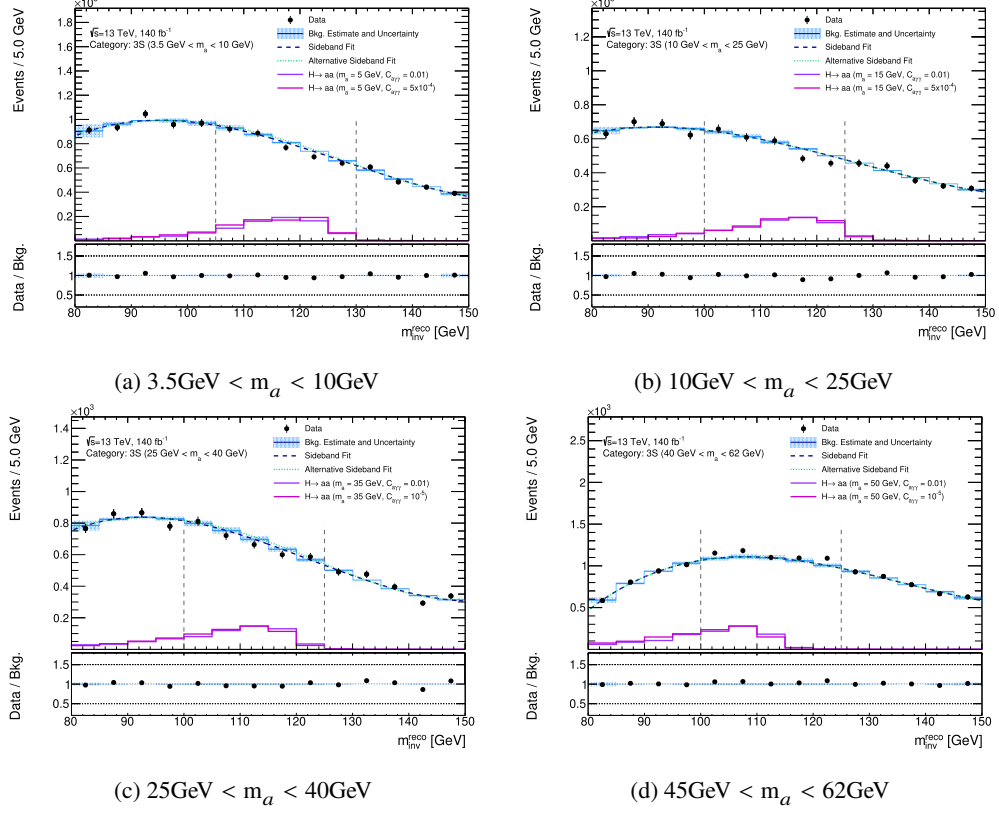


Figure A.1: Invariant mass distribution of the reconstructed four photon system in four bins of reconstructed ALP mass. The blue dashed line shows the nominal sideband fit function. The estimated background is shown as blue histogram with blue shaded areas depicting systematic uncertainties obtained from a sideband fit with reduced range. The green dotted line shows the alternative sideband fit function. Purple and violet histograms show different signal shapes. The black dots show the measured data.

Fig. A.1 shows the background estimation in the 3S category. Blue histograms with shaded blue areas show the background estimate which is taken from the polynomial function of third order depicted as blue dashed line with systematic uncertainties from a sideband fit with reduced range. The green dotted line shows the alternative sideband fit. It can be seen, that the functions describe the data well and the systematic uncertainties are relatively small. The purple and violet histograms show different signal hypothesis. It is visible, that the shape gets broader and shifted to lower masses for higher  $m_a$ . This results in a large fraction of signal events in the  $m_a > 40\text{GeV}$  case leaking into the sideband region. This makes a proper background estimation challenging. To avoid a bias on the background estimation, it is decided to not use the 3S category for  $m_a > 40\text{GeV}$ .

The invariant mass plots for the 4S background estimation can be seen in Fig. A.2 Blue histograms show again the background estimate together with systematic uncertainties obtained by the sideband fit with reduced range as shaded areas. The nominal sideband fitting function is depicted as dashed blue line, whereas the alternative sideband fitting function is shown as dotted green line. It can be

## A.1 Background estimation for the three and four photon categories

seen that the nominal function fits the data within statistical uncertainties. However, the alternative function shows remarkable differences which are taken into account as spurious signal uncertainty. Purple and violet histograms show the expected shape for different signal hypothesis. It can be seen that most of these histograms are located within the signal range between the two dashed vertical lines. Data is depicted with black dots. The data in the signal region is added after the background estimation was completed and the analysis unblinded. It can be seen that the background estimation can describe the data in both, the sideband and the signal region.

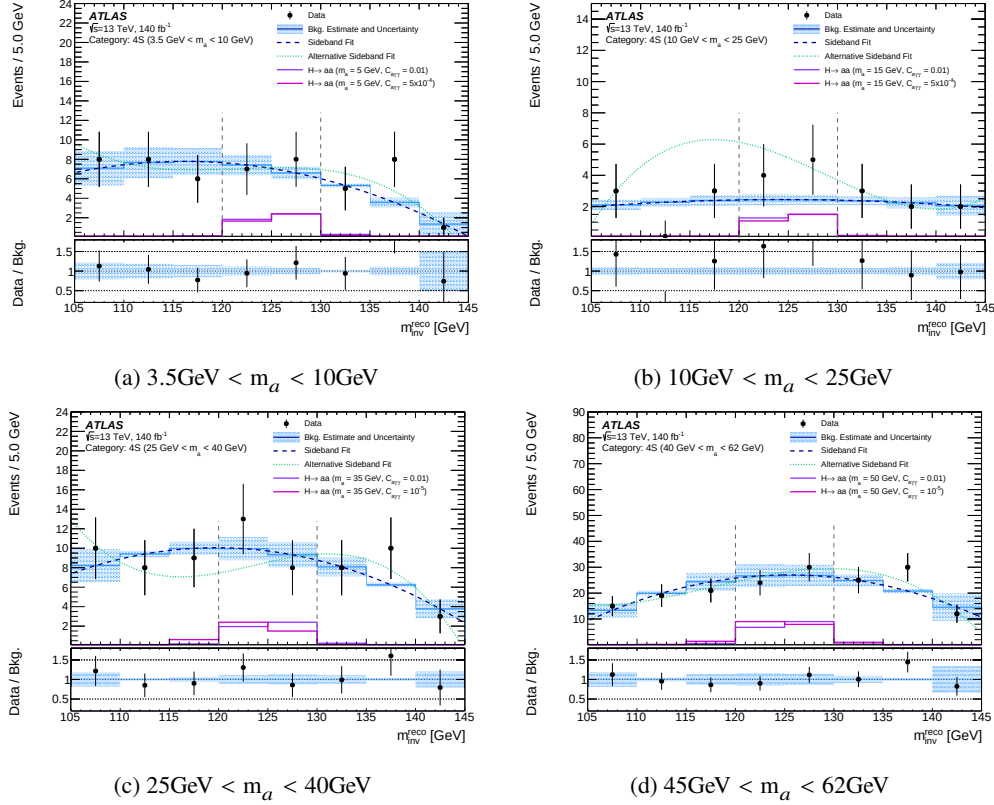


Figure A.2: Invariant mass distribution of the reconstructed four photon system in four bins of reconstructed ALP mass. The blue dashed line shows the nominal sideband fit function. The estimated background is shown as blue histogram with blue shaded areas depicting systematic uncertainties obtained from a sideband fit with reduced range. The green dotted line shows the alternative sideband fit function. Purple and violet histograms show different signal shapes. The black dots show the actual measured data. [1]

## A.2 Uncertainties due to displaced vertices

In this analysis, long-lived ALPs decaying to photons are considered for the first time. This introduces significant challenges, not only for the analysis strategy but also for the estimation of systematic uncertainties. While standard uncertainties based on recommendations are sufficient for analyses searching for prompt ALPs, they are inadequate for events involving ALPs that decay with significant displacement. Hence new uncertainty models for long-lived ALPs need to be developed.

Usually, when introducing systematic uncertainties, a reasonable approach is to compare simulated events with measured data. For a decay of ALPs to photons, it is not possible to find the decay vertex in reconstructed variables, since photons do not leave tracks in the detector. It would be beneficial to have a decay with similar displacement involving charged particles to profit from tracking. Electrons would be an obvious choice for particles creating tracks while providing similar electromagnetic showers than photons. However, electrons are nearly always produced promptly.

A way to get similar signatures is to use long-lived baryons (mainly kaons) and associate the tracks stemming from the displaced vertex to calorimeter clusters. For these calorimeter clusters, shower-shape variables can be used similarly to shower-shape variables for photons. This can be done in simulation as well as in data. However, the composition of processes in the data is not well understood, making it challenging to obtain a suitable simulated sample.

Under the assumption, that the difference in shower-shapes between simulation and data for prompt decays can be accounted for by reweighting, any remaining discrepancy between simulation and data at displaced decays can now be assumed to be caused by a potential mismodelling of the detector response due to the displacement.

First, all events are split into three parts. A near regime that consists of tracks with a distance  $20 \text{ mm} < z_0$  and  $1 \text{ mm} < d_0$ . A medium regime containing tracks with a distance between  $20 \text{ mm} < z_0 < 500 \text{ mm}$  and  $1 \text{ mm} < d_0 < 80 \text{ mm}$ . And finally a far regime with tracks with a distance above  $z_0 > 500 \text{ mm}$  and  $d_0 > 80 \text{ mm}$ .

For each of these regions, the ratio between data and simulation is calculated for all shower-shape variables. The ratio in the near regime is used to correct the simulation in all regimes. This leads (by construction) to a perfect agreement between data and simulation in the near regime. Fig. A.3 shows exemplary the shower-shape  $R_\phi$  in all regions. The black dots show data, the grey histograms show the simulated data and the blue histograms show simulated data reweighted to the data in the near region. Mean and RMS for the  $\frac{\text{data}}{\text{simulation}}$  distributions of all shower-shape variables are calculated for the medium and far regimes. A linear interpolation with respect to the displacement distance is performed. The ratio as well as the interpolation for the  $R_\phi$  variable can be seen in Fig. A.4.

In a next step, the shower-shape values in simulated events can be modified using this interpolation and the true vertex distance. With these modified shower-shape variables, the efficiency of the customized photon identification as well as the ATLAS photon identification are re-evaluated. This is shown for the two custom classifiers in Fig. A.6 and ATLAS Loose (Tight) ID in Appendix A.2. The top row shows the classifier efficiency using nominal input variables (black) as well as modified input variables (blue). Below, the ratio between nominal and fudged classifier efficiencies is shown for both of the two classifiers. To get a smooth function for all decay lengths, this is fitted with a linear function  $f_{fit}(x)$ . Up and Down variations are build by assigning additional event weights of  $1 \pm f_{fit}(r_{decay})$  to each simulated signal event according to the truth decay length  $r_{decay}$ .

This leads to uncertainties around 4 % for classifier 1 and uncertainties up to 15 % for classifier 2. The uncertainties on Loose ID range up to 24 % for long displacements while the uncertainties on the

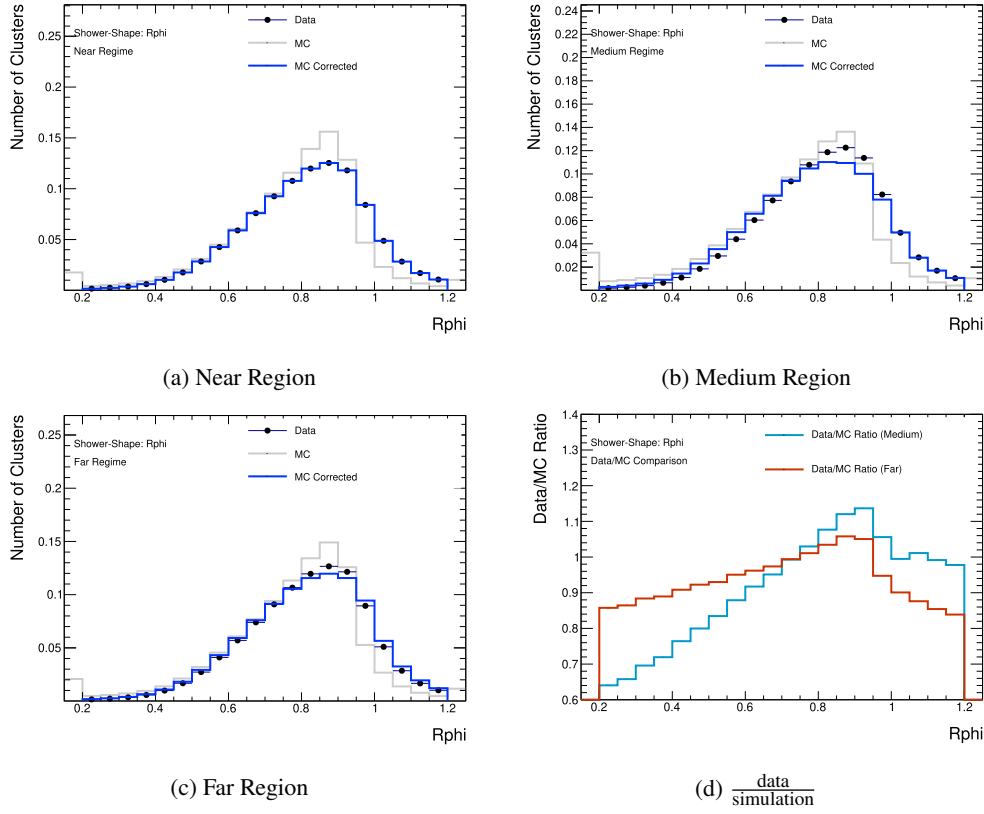


Figure A.3: Shower-shape  $R_\phi$  in for data and simulation in the close, medium and far region. The black dots show data while the grey histogram shows simulation. The blue histogram shows the simulated data reweighted to the data distribution in the near region. [150]

Tight ID reach 22 %.

A more detailed description and further improvements of this methodology can be found in [171].

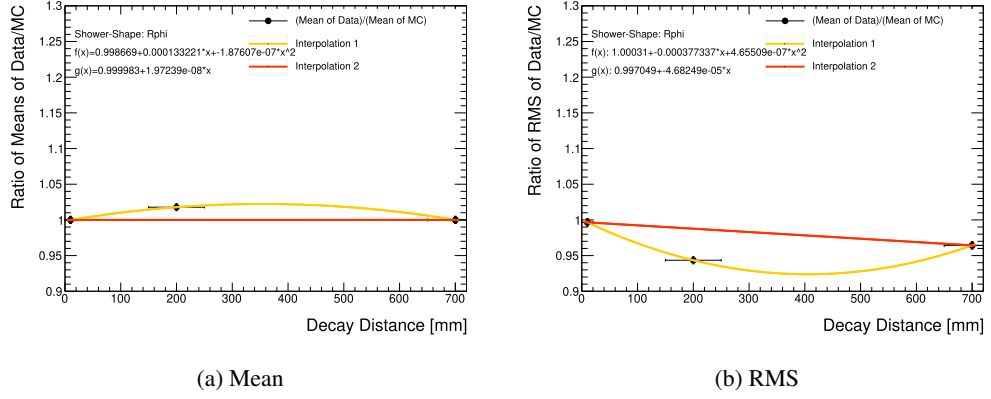


Figure A.4: Mean and RMS of the  $R_\phi$  shower-shape with interpolation. [150]

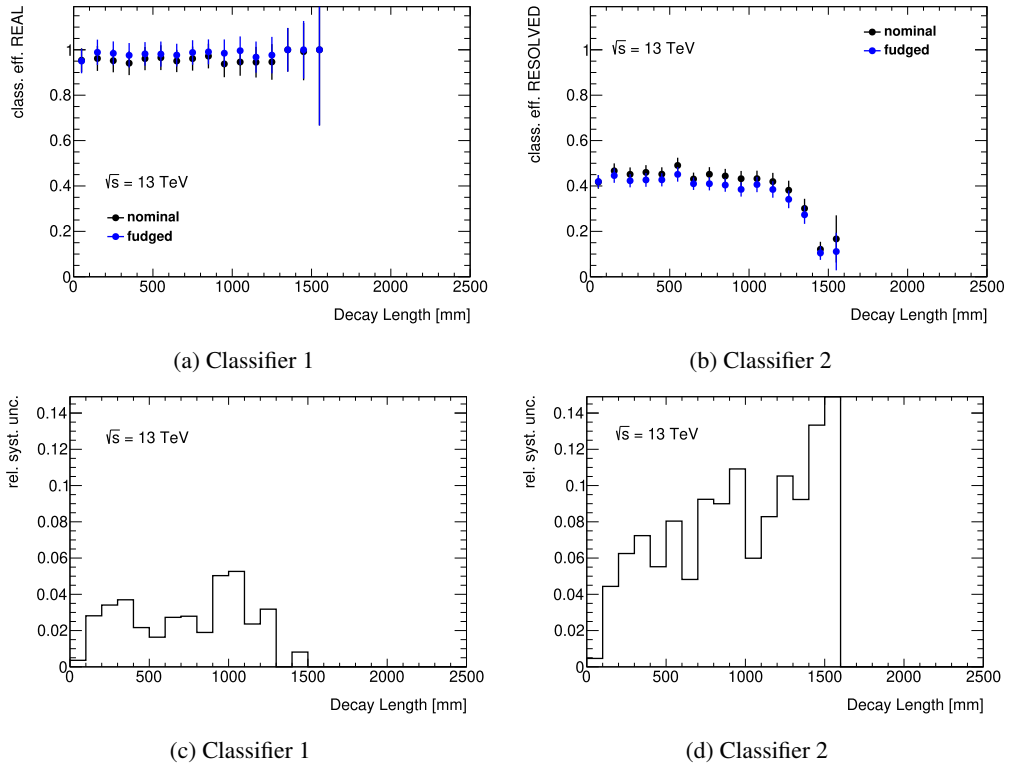


Figure A.5: Classifier efficiencies for nominal (black) and fudged (blue) input variables with respect to the decay length of ALPs. The bottom row shows the ratio between fudged and nominal classifier efficiencies. [150]

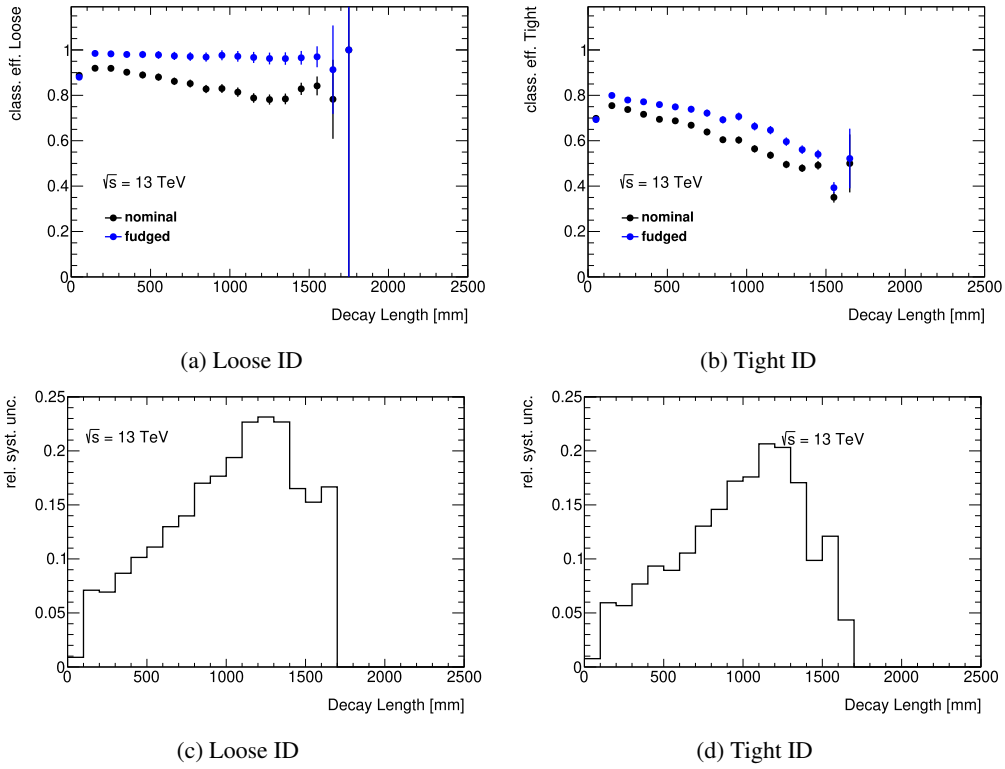


Figure A.6: Efficiencies of the ATLAS photon identification for nominal (black) and fudged (blue) input variables with respect to the decay length of ALPs. The bottom row shows the ratio between fudged and nominal ID efficiencies. This is used as symmetric uncertainty. [150]

### A.3 Trigger studies

?? lists all triggers. Studies have been performed using all triggers independently, all combined requiring one of all triggers to fire (with an 'or').

Table A.1: List of studied trigger chains. Marked are the 2 trigger chains with the highest signal efficiency per year. Out of those only the trigger chain with the largest signal efficiency is used. Taken from [150]

| Trigger chain                         | 2018 | 2017 | 2016 | 2015 |
|---------------------------------------|------|------|------|------|
| HLT_g35_loose_g25_loose               |      |      | X    | X    |
| HLT_g35_medium_g25_medium             |      |      |      |      |
| HLT_g35_medium_g25_medium_L12EM20VH   | X    | X    |      |      |
| HLT_2g25_tight_L12EM20VH              |      |      |      |      |
| HLT_2g22_tight_L12EM15VHI             | X    | X    |      |      |
| HLT_2g22_tight                        |      |      | X    | X    |
| HLT_2g20_tight_icalovloose_L12EM15VHI |      |      |      |      |
| HLT_3g20_loose                        |      |      |      |      |



## A.4 Good Run Lists

Table A.2: Good Run Lists used when processing data samples.

| Year | GRL                                                                                                                      |
|------|--------------------------------------------------------------------------------------------------------------------------|
| 2015 | data15_13TeV.periodAllYear_<br>DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_<br>All_Good_25ns.xml                     |
| 2016 | data16_13TeV.periodAllYear_<br>DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_<br>All_Good_25ns.xml          |
| 2017 | data17_13TeV.periodAllYear_<br>DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_<br>All_Good_25ns_TriggerNo17e33prim.xml  |
| 2018 | data18_13TeV.periodAllYear_<br>DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_<br>All_Good_25ns_TriggerNo17e33prim.xml |

## A.5 Model Architecture and Weight Visualization

To better understand the inner workings of the trained model, we examine both its architecture Fig. A.7 and the learned weights of its layers. This includes the structure of the network as well as visualizations of weight matrices from dense layers Figs. A.11, A.12, A.13 and A.14 and convolutional filters from each Conv2D layer Figs. A.8, A.9 and A.10.

### A.5.1 Model Architecture

### A.5.2 Understanding Filter Weights

Each filter in a Conv2D layer acts as a local pattern detector. These filters are small kernels (e.g.,  $3 \times 3 \times \text{channels}$ ) whose weights are learned during training. They slide across the input feature map and respond to specific spatial patterns. The values of these filter weights reflect their sensitivity to various input features:

- **High-magnitude weights** indicate strong influence; they emphasize certain features.
- **Near-zero weights** suggest insensitivity to those input positions (possibly unused).
- **Structured patterns** in filters (e.g., edge-like or checkerboard patterns) reveal that the network is extracting meaningful features like edges, textures, or shapes.
- **Smooth or uniform filters** might detect broader, low-frequency features or reflect insufficient training.

Early Conv2D layers typically learn low-level features (e.g., edges), while deeper ones learn more abstract patterns. Blank or noisy filters may indicate underused capacity, redundancy, or convergence issues.

## A.6 $p_T$ reweighting for CNNs

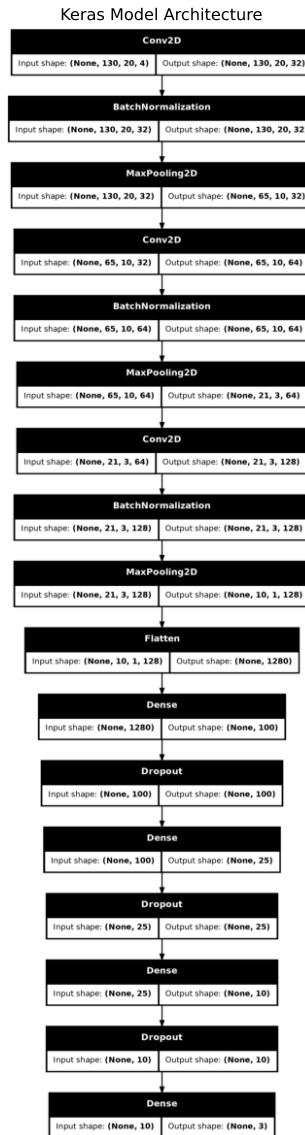


Figure A.7: The full Keras model architecture visualized. The model follows a sequential design, with three Conv2D-BatchNormalization-MaxPooling blocks for feature extraction, followed by a Flatten layer and a series of Dense layers for classification. The final layer uses a softmax activation to output probabilities over three classes.

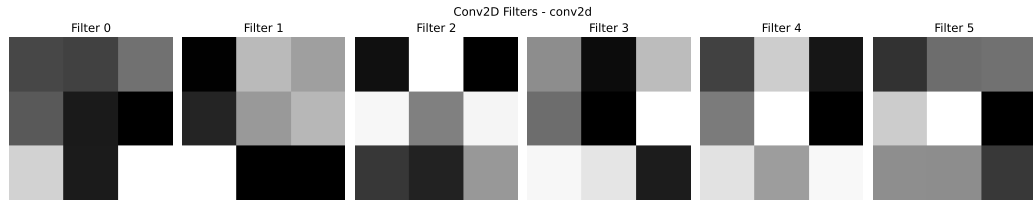


Figure A.8: Filters from the first Conv2D layer. These filters generally display high-frequency patterns with sharp transitions, such as edges and gradients, which are typical in early convolutional layers. The contrast and symmetry observed in filters like 0 and 1 suggest that they are acting as edge detectors in specific orientations. Others, like filter 4, appear more diffuse, capturing broader spatial patterns.

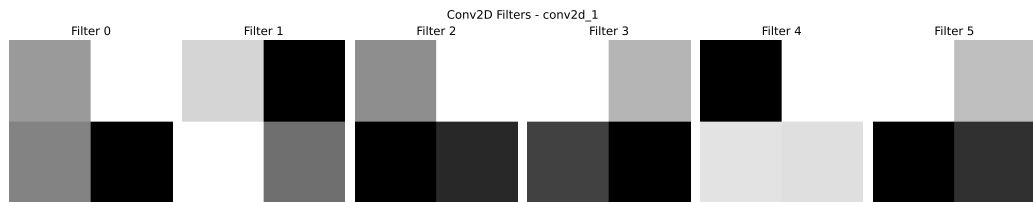


Figure A.9: Filters from the second Conv2D layer (conv2d\_1). These filters show smoother and more complex structures compared to the first layer. The patterns are less edge-like and more texture-based, indicating that the network is starting to integrate and abstract information from the earlier spatial features. Filters 2 and 5 show a strong diagonal gradient, potentially detecting orientation-based textures.

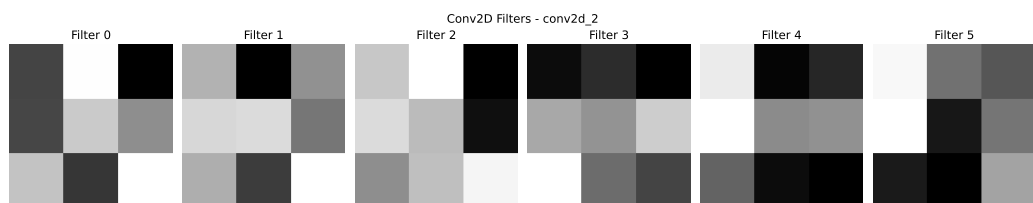


Figure A.10: Filters from the third Conv2D layer (conv2d\_2). These filters exhibit more diffuse and irregular patterns, which is typical for deeper layers focusing on abstract, class-specific patterns. Some filters (e.g., 0 and 3) are less visually structured, which may suggest underutilization or redundancy, while others (e.g., filter 4) still preserve directional patterns.

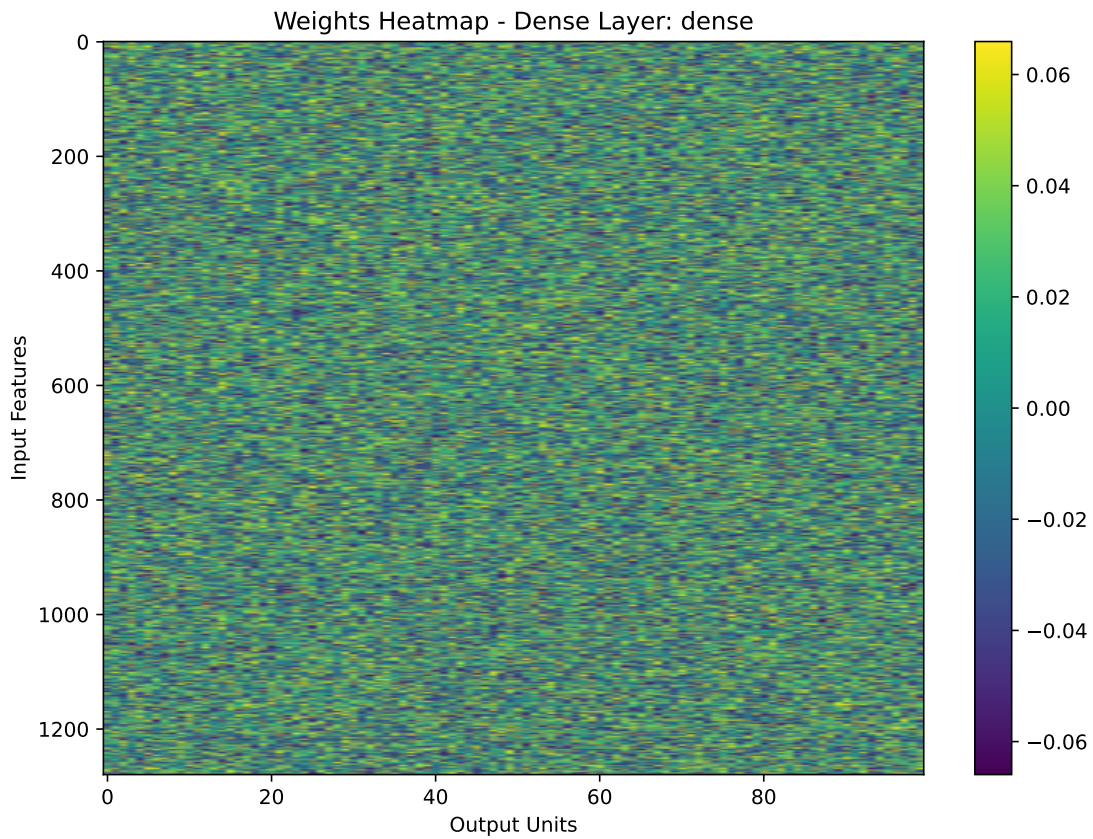


Figure A.11: Heatmap of the first Dense layer (dense) weights, connecting 1280 features to 100 output neurons. The weights are mostly low-magnitude and centered around zero, with no extreme outliers. This indicates balanced initialization and controlled learning, likely aided by dropout and normalization.

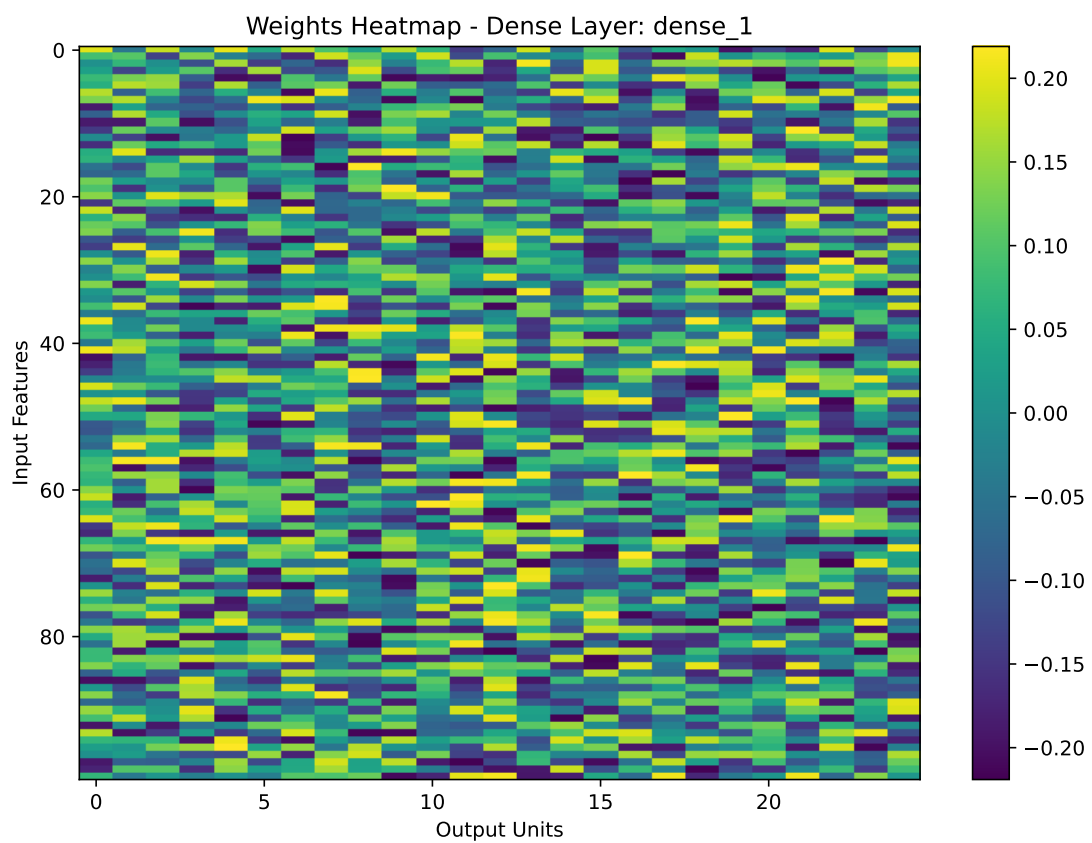


Figure A.12: Weights of Dense layer dense\_1 (100 to 25 units). Compared to the first layer, some input features dominate more clearly, suggesting that certain neurons specialize in specific patterns from earlier layers.

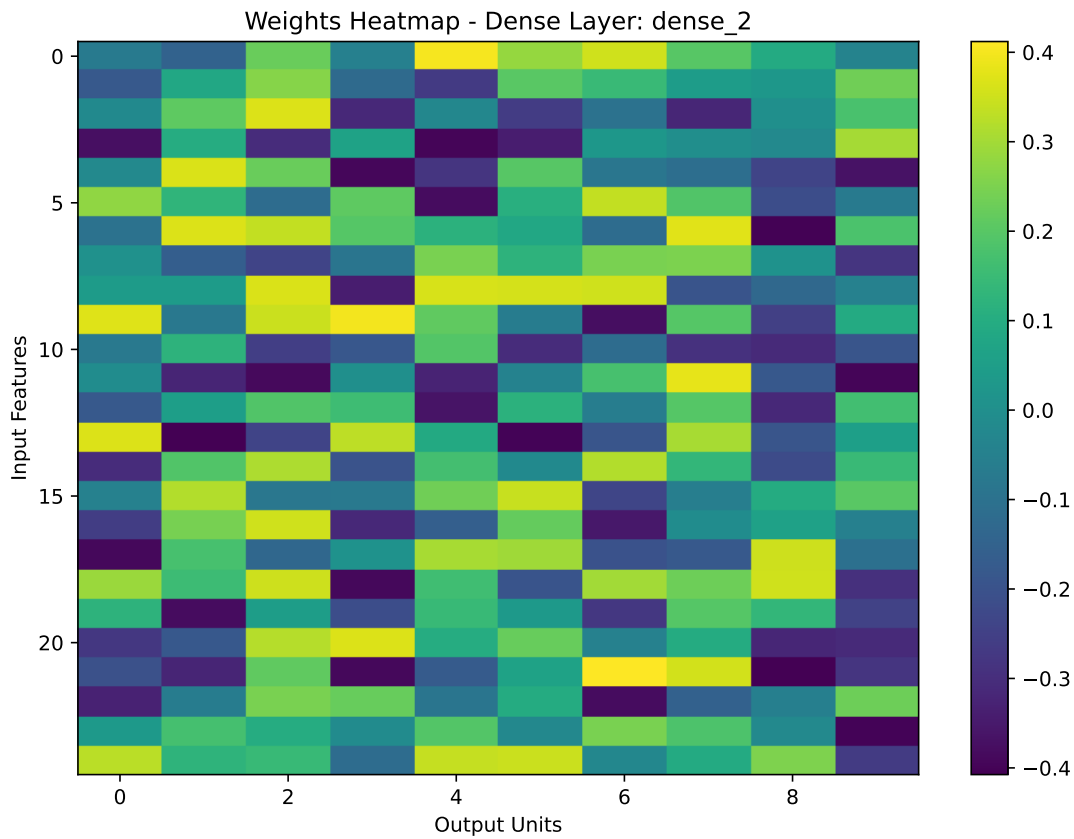


Figure A.13: Weight matrix for Dense layer dense\_2 (25 to 10 units). A smaller network width results in more sparsity and sharp contrasts, which may reflect feature compression or focused decision paths leading into the final classifier.

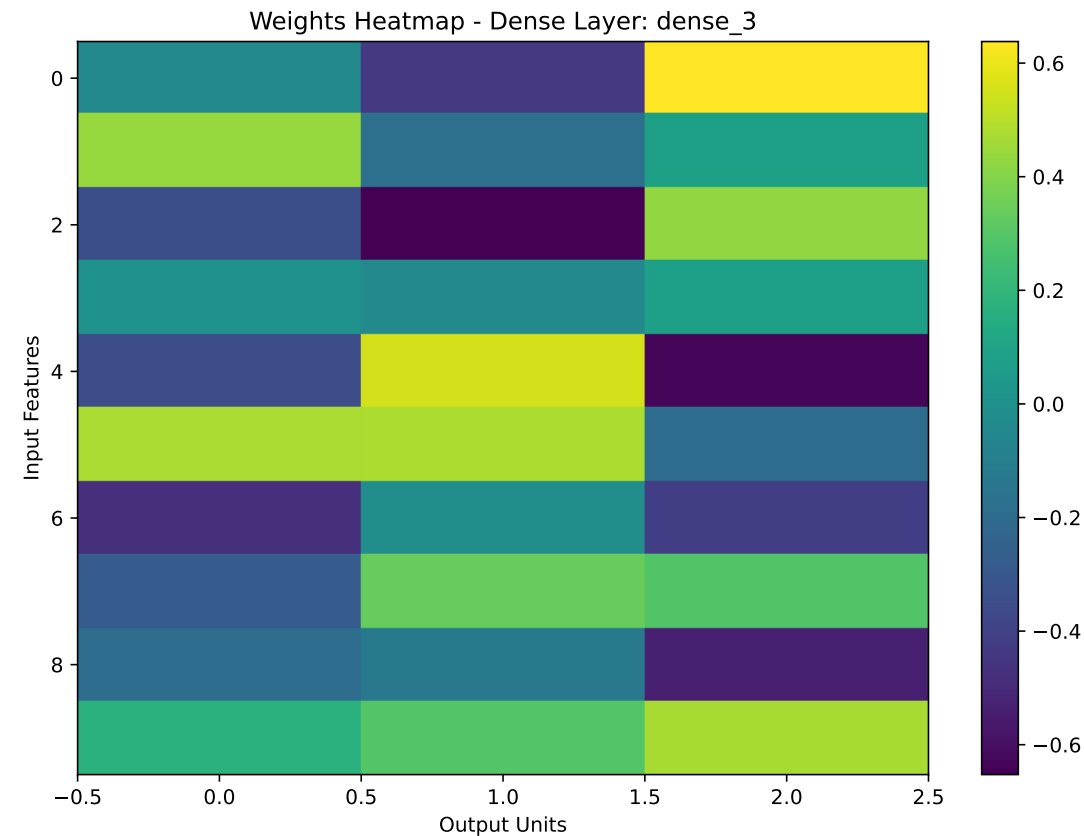


Figure A.14: Weights for the output layer dense\_3, connecting 10 neurons to 3 softmax outputs. The matrix suggests each class is influenced by a different subset of the penultimate layer, as indicated by column-specific patterns in the heatmap.



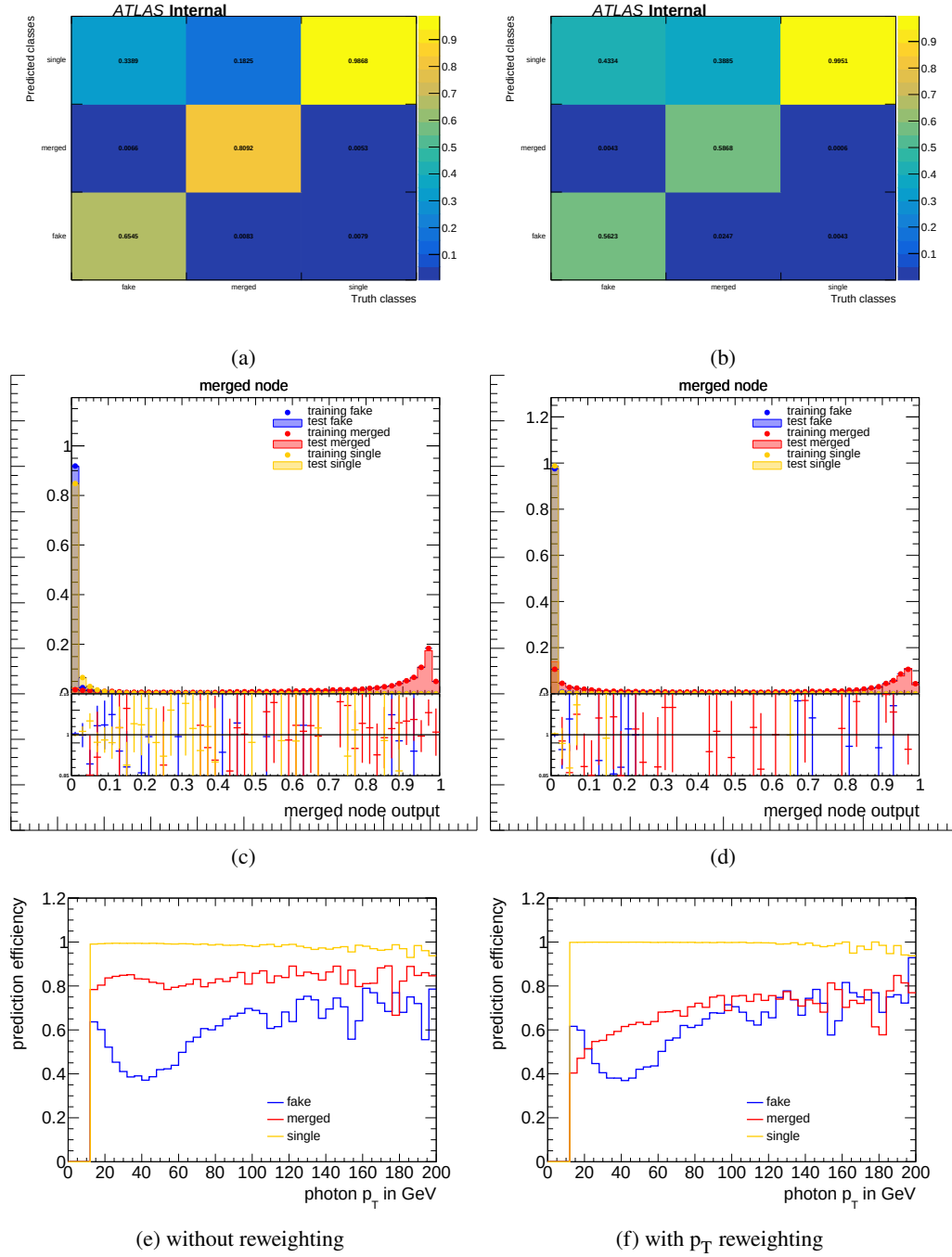


Figure A.15: Results for all photons with  $p_T > 15$  GeV. The plots on the left hand-side result from a training with a weight of 1 for each input, whereas the plots are from a training that used  $p_T$  reweighting

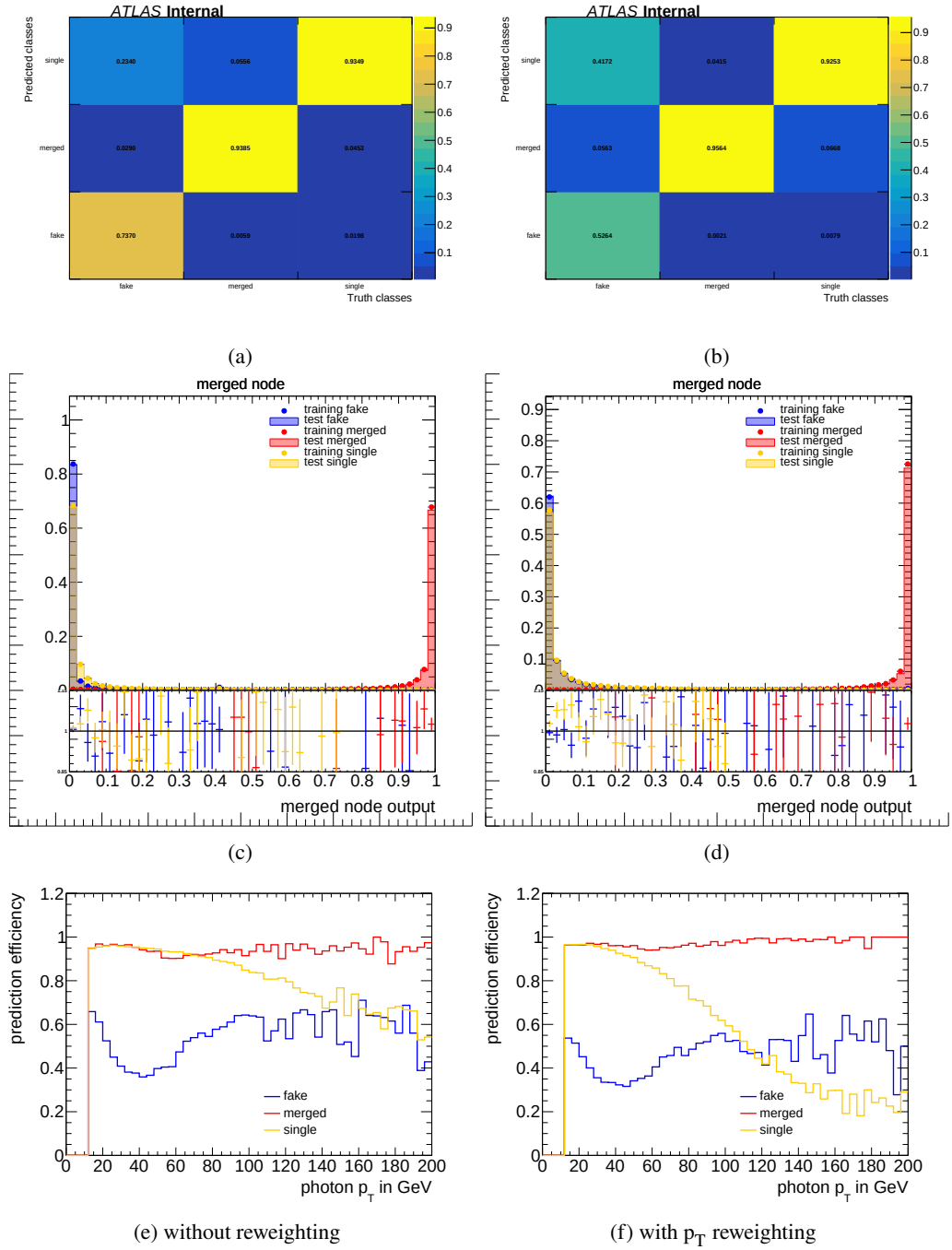


Figure A.16: Results a training with equal amounts of photons with  $p_T > 15$  GeV. The plots on the left hand-side result from a training with a weight of 1 for each input, whereas the plots are from a training that used  $p_T$  reweighting

## A.7 Different $m_a$ regression results

Table A.3: Number of merged photons as a function of the ALP mass  $m_a$ .

| $m_a$ [GeV]  | Merged photons |
|--------------|----------------|
| 0.5          | 21             |
| 1.0          | 3527           |
| 1.5          | 3333           |
| 2.0          | 9335           |
| 3.0          | 2403           |
| 3.5          | 8416           |
| 4.0          | 1463           |
| 4.5          | 2799           |
| 5.0          | 1309           |
| 6.0          | 709            |
| 7.0          | 720            |
| 8.0          | 315            |
| <b>Total</b> | <b>34350</b>   |

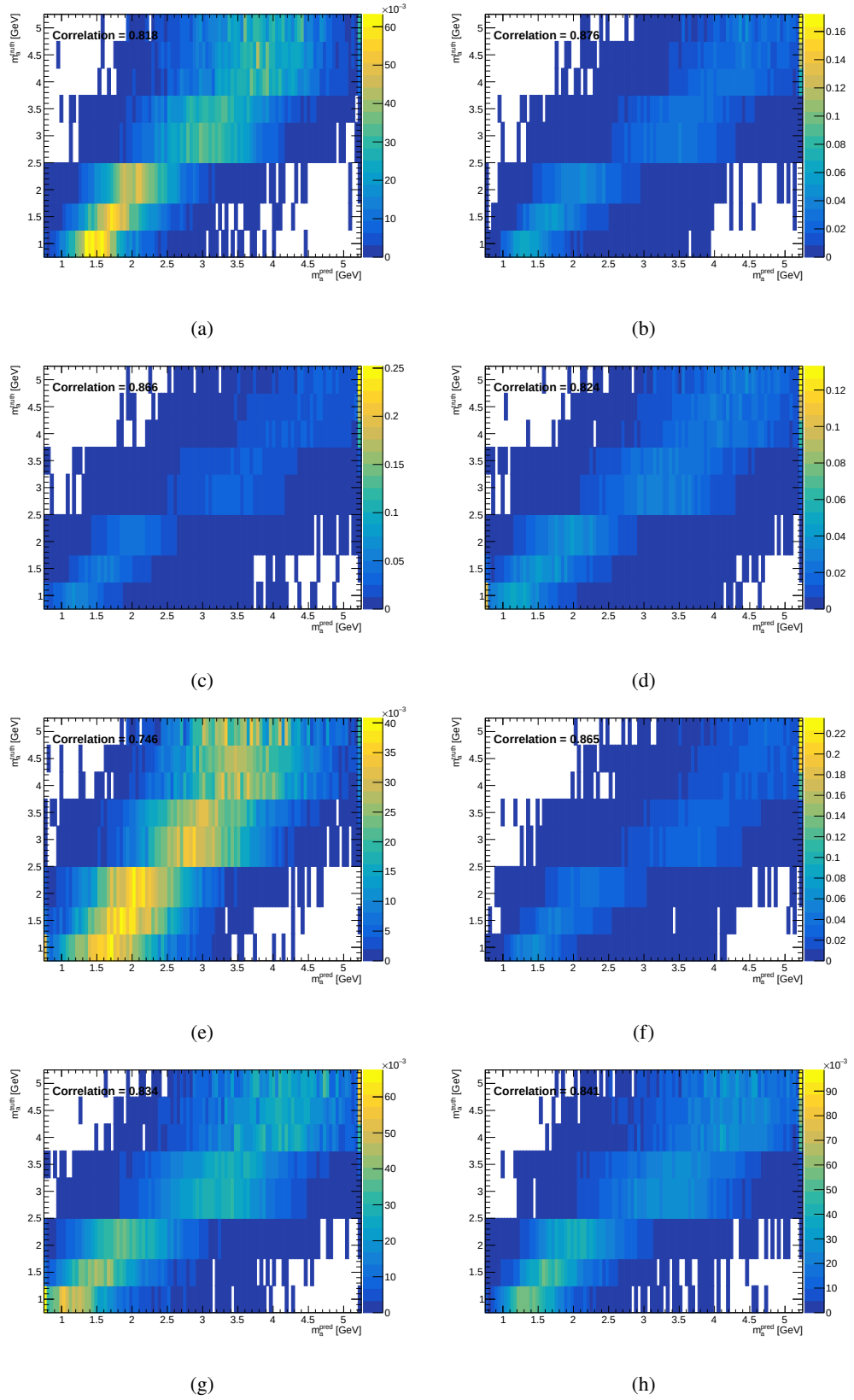


Figure A.17: 2D responses for different  $m_a$  regression trainings with similar settings

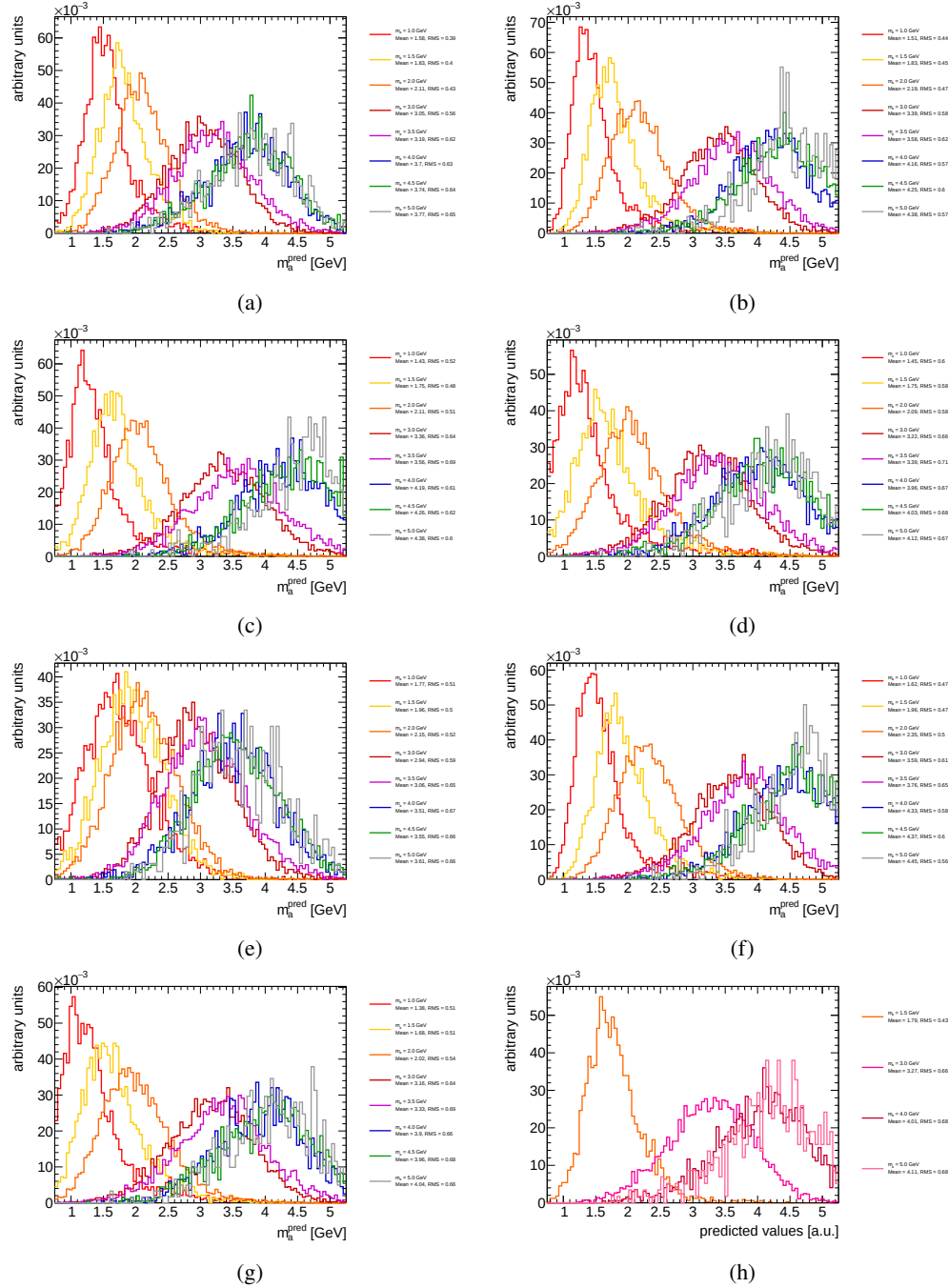


Figure A.18: 1D results for different  $m_a$  regression trainings with similar settings. Mean and RMS are shown for different truth  $m_a$  in the legend.

**A.8 Use of AI Tools**

Appendix: Use of AI Tools

| AI Tool | Used for                                                             | Reason                                                                   | When / Where                                                             |
|---------|----------------------------------------------------------------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------|
| ChatGPT | Correcting text for consistently British English                     | To improve linguistic quality and academic style                         | Throughout the thesis, especially transitions and summaries              |
| ChatGPT | Generating and refining $\LaTeX$ code (e.g. tables, method sections) | Efficient formatting and technically correct output                      | While writing methods and presenting results                             |
| ChatGPT | Python code for ML model evaluation and example plots                | Quick prototyping and validation of visualisation and evaluation scripts | During development of CNN regression and classification evaluation tools |
| ChatGPT | Reviewing drafts for consistency, clarity, and tone                  | Ensure logical coherence and scientific readability                      | At the end of each chapter draft                                         |

---

## Bibliography

---

- [1] ATLAS Collaboration, *Search for short- and long-lived axion-like particles in  $H \rightarrow aa \rightarrow 4\gamma$  decays with the ATLAS experiment at the LHC*, *Eur. Phys. J. C* **84** (2024) 742, arXiv: [2312.03306 \[hep-ex\]](#) (cit. on pp. 2, 49, 60–62, 64, 69, 74, 81–83, 85–87, 89, 109).
- [2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](#) (cit. on pp. 3, 12).
- [3] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv: [1207.7214 \[hep-ex\]](#) (cit. on pp. 3, 12, 69).
- [4] R. Wolf, *The Higgs Boson Discovery at the Large Hadron Collider*, Springer, 2015 (cit. on p. 3).
- [5] F. Dyson, *Dyson Quantenfeldtheorie: Die weltbekannte Einführung von einem der Väter der QED*, Berlin, Heidelberg: Springer Spektrum, 2014, ISBN: 978-3-642-37677-1, URL: <https://link.springer.com/book/10.1007/978-3-642-37678-8> (cit. on p. 3).
- [6] R. P. Feynman, *The Theory of Positrons*, 1949, URL: <https://link.aps.org/doi/10.1103/PhysRev.76.749> (cit. on p. 4).
- [7] J. Schwinger, *Quantum Electrodynamics in a Covariant Formulation*, 1948, URL: <https://link.aps.org/doi/10.1103/PhysRev.74.1439> (cit. on p. 4).
- [8] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, 1959, URL: <https://doi.org/10.1007/BF02726525> (cit. on p. 4).
- [9] The UA1 Collaboration, *Experimental observation of isolated large transverse energy electrons with associated missing energy at  $\sqrt{s}=540$  GeV*, 1983, URL: [https://doi.org/10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2) (cit. on pp. 4, 10).
- [10] The UA2 Collaboration, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, 1983, URL: [https://doi.org/10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2) (cit. on pp. 4, 10).
- [11] The UA1 Collaboration, *Experimental observation of lepton pairs of invariant mass around 95 GeV/c<sup>2</sup> at the CERN SPS collider*, 1985, URL: <http://cds.cern.ch/record/163857> (cit. on pp. 4, 10).

- [12] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, (1961),  
URL: <https://www.osti.gov/servlets/purl/4008239/> (cit. on p. 4).
- [13] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, 1973,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.30.1346> (cit. on p. 4).
- [14] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, 1964,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on p. 4).
- [15] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, 1964,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 4).
- [16] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble,  
*Global Conservation Laws and Massless Particles*, 1964,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on p. 4).
- [17] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, 1960,  
URL: <https://link.aps.org/doi/10.1103/PhysRev.117.648> (cit. on p. 4).
- [18] J. Goldstone, *Field theories with « Superconductor » solutions*, 1961,  
URL: <https://doi.org/10.1007/BF02812722> (cit. on p. 4).
- [19] *Standard Model of Elementary Particles*,  
[https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg),  
Image licensed under CC BY-SA 4.0 (cit. on p. 5).
- [20] J. I. Friedman, H. W. Kendall and R. E. Taylor,  
*Observed Behavior of Highly Inelastic Electron-Proton Scattering*,  
*Phys. Rev. Lett.* **23** (1969) 935 (cit. on p. 6).
- [21] M. Gell-Mann, *Isotopic Spin and Strange Particles*, *Phys. Rev.* **92** (1953) 833 (cit. on p. 6).
- [22] J. Aubert and et al., *Experimental observation of a heavy particle J*,  
*Phys. Rev. Lett.* **33** (1974) 1404 (cit. on p. 6).
- [23] J. Augustin and et al., *Discovery of a Narrow Resonance in  $e^+e^-$  Annihilation*,  
*Phys. Rev. Lett.* **33** (1974) 1406 (cit. on p. 6).
- [24] S. Herb and et al.,  
*Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*,  
*Phys. Rev. Lett.* **39** (1977) 252 (cit. on p. 6).
- [25] F. Abe and et al. (CDF),  
*Observation of Top Quark Production in  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.8$  TeV*,  
*Phys. Rev. Lett.* **74** (1995) 2626 (cit. on p. 6).
- [26] DØ Collaboration, *Observation of the Top Quark*, *Phys. Rev. Lett.* **74** (1995) 2632,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.74.2632> (cit. on p. 6).
- [27] S. Navas et al., *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001  
(cit. on pp. 6, 7, 10, 11).
- [28] I. I. Bigi, Y. L. Dokshitzer, V. Khoze, J. Kühn and P. M. Zerwas,  
*Production and decay properties of ultra-heavy quarks*, *Phys. Lett. B* **181** (1986) 157,  
URL: <https://cds.cern.ch/record/171099> (cit. on p. 7).



- 
- [29] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (1963) 531 (cit. on p. 7).
- [30] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652 (cit. on p. 7).
- [31] J. J. Thomson, *Cathode Rays*, *Philos. Mag.* **44** (1897) 293 (cit. on p. 7).
- [32] S. H. Neddermeyer and C. D. Anderson, *Note on the Nature of Cosmic-Ray Particles*, *Phys. Rev.* **51** (1937) 884 (cit. on p. 7).
- [33] M. L. Perl et al., *Evidence for Anomalous Lepton Production in  $e+e^-$  Annihilation*, *Phys. Rev. Lett.* **35** (1975) 1489 (cit. on p. 7).
- [34] W. Pauli, “Pauli letter collection”, Typed copy, URL: <https://cds.cern.ch/record/83282> (cit. on p. 7).
- [35] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse and A. D. McGuire, *Detection of the Free Neutrino: A Confirmation*, *Science* **124** (1956) 103 (cit. on p. 8).
- [36] G. Danby et al., *Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos*, *Phys. Rev. Lett.* **9** (1962) 36 (cit. on p. 8).
- [37] D. Collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) 218 (cit. on p. 8).
- [38] N. Collaboration, *NaNu: A proposal to detect tau anti-neutrinos at CERN*, <https://cds.cern.ch/record/2888846>, CERN-SPSC-2023-017, SPSC-P-374, 2023 (cit. on p. 8).
- [39] S.-K. Collaboration, *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 (cit. on p. 8).
- [40] S. Collaboration, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **89** (2002) 011301 (cit. on p. 8).
- [41] C. Giganti, S. Lavignac and M. Zito, *Neutrino oscillations: The rise of the PMNS paradigm*, *Progress in Particle and Nuclear Physics* **98** (2018) 1, ISSN: 0146-6410, URL: <http://dx.doi.org/10.1016/j.pnpnp.2017.10.001> (cit. on p. 8).
- [42] KATRIN Collaboration, *Direct neutrino-mass measurement with sub-eV sensitivity using KATRIN*, *Nat. Phys.* **18** (2022) 160 (cit. on p. 8).
- [43] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343 (cit. on p. 9).
- [44] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346 (cit. on p. 9).
- [45] CERN Courier, *The history of QCD: Running coupling*, CERN Courier (2004), <https://cerncourier.com/a/the-history-of-qcd/> (cit. on p. 9).

- [46] TASSO Collaboration,  
*Evidence for gluon bremsstrahlung in  $e^+e^-$  annihilations at high energies*,  
*Physics Letters B* **86** (1979) 418, ISSN: 0370-2693,  
URL: <https://www.sciencedirect.com/science/article/pii/0370269379908694>  
(cit. on p. 10).
- [47] L. Adam, *Precision Matters: Measurement of the W boson mass and width with the ATLAS detector at a centre-of-mass energy of 7 TeV and the activity of the Underlying Event in Z boson events at a centre-of-mass energy of 13 TeV*, Online resource,  
Dissertation: Johannes Gutenberg-Universität Mainz, 2021 211,  
URL: <https://nbn-resolving.org/urn:nbn:de:hebis:77-openscience-35e0df0e-f38d-4a86-b008-536d29dc86a35> (cit. on p. 10).
- [48] D. de Florian et al.,  
*Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*,  
(2016), arXiv: 1610.07922 [hep-ph] (cit. on pp. 12, 52, 72).
- [49] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52, arXiv: 2207.00092 [hep-ex]  
(cit. on p. 13).
- [50] V. Gribov and L. Lipatov, *Deep inelastic electron scattering in perturbation theory*,  
*Physics Letters B* **37** (1971) 78, ISSN: 0370-2693,  
URL: <https://www.sciencedirect.com/science/article/pii/0370269371905764>  
(cit. on p. 14).
- [51] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*,  
*Nucl. Phys. B* **126** (1977) 298 (cit. on p. 14).
- [52] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and  $e^+e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics*,  
*Sov. Phys. JETP* **46** (1977) 641 (cit. on p. 14).
- [53] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*,  
*Eur. Phys. J. C* **63** (2009) 189, arXiv: 0901.0002 [hep-ph] (cit. on p. 15).
- [54] C. A. et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*,  
*Phys. Rev. Lett.* **124** (2020) 081803, arXiv: 2001.11966 [hep-ex] (cit. on pp. 15, 17).
- [55] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*,  
*Astron. Astrophys.* **641** (2020) A6, arXiv: 1807.06209 [astro-ph.CO] (cit. on p. 16).
- [56] G. W. e. a. Bennett,  
*Final report of the E821 muon anomalous magnetic moment measurement at BNL*,  
*Phys. Rev. D* **73** (7 2006) 072003,  
URL: <https://link.aps.org/doi/10.1103/PhysRevD.73.072003> (cit. on p. 16).
- [57] Muon  $g - 2$  Collaboration,  
*Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*,  
*Phys. Rev. Lett.* **126** (14 2021) 141801,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.126.141801>  
(cit. on pp. 16, 18).

- 
- [58] The Muon  $g - 2$  Collaboration,  
*Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm*,  
*Phys. Rev. Lett.* **131** (16 2023) 161802,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.131.161802>  
(cit. on pp. 16, 18).
- [59] Muon  $g - 2$  Collaboration, *Detailed report on the measurement of the positive muon anomalous magnetic moment to 0.20 ppm*, *Phys. Rev. D* **110** (3 2024) 032009,  
URL: <https://link.aps.org/doi/10.1103/PhysRevD.110.032009> (cit. on p. 16).
- [60] A. Kurz et al., *Electron contribution to the muon anomalous magnetic moment at four loops*,  
*Phys. Rev. D* **93** (5 2016) 053017,  
URL: <https://link.aps.org/doi/10.1103/PhysRevD.93.053017> (cit. on p. 16).
- [61] T. Aoyama, M. Hayakawa, T. Kinoshita and M. Nio,  
*Revised Value of the Eighth-Order Contribution to the Electron  $g - 2$* ,  
*Phys. Rev. Lett.* **99** (11 2007) 110406,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.99.110406> (cit. on p. 16).
- [62] T. Kinoshita and M. Nio, *Tenth-order QED contribution to the lepton  $g - 2$ : Evaluation of dominant  $\alpha^5$  terms of muon  $g - 2$* , *Phys. Rev. D* **73** (5 2006) 053007,  
URL: <https://link.aps.org/doi/10.1103/PhysRevD.73.053007> (cit. on p. 16).
- [63] S. Borsanyi et al.,  
*Leading hadronic contribution to the muon magnetic moment from lattice QCD*,  
*Nature* **593** (2021) 51 (cit. on p. 16).
- [64] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Pseudoparticles*,  
*Phys. Rev. Lett.* **38** (25 1977) 1440,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.38.1440> (cit. on p. 17).
- [65] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223 (cit. on p. 17).
- [66] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*,  
*Phys. Rev. Lett.* **40** (1978) 279 (cit. on p. 17).
- [67] J. E. Kim, *Weak Interaction Singlet and Strong CP Invariance*, *Phys. Rev. Lett.* **43** (1979) 103  
(cit. on p. 17).
- [68] M. Shifman, A. Vainshtein and V. Zakharov,  
*Can Confinement Ensure Natural CP Invariance of Strong Interactions?*,  
*Nucl. Phys. B* **166** (1980) 493 (cit. on p. 17).
- [69] A. Zhitnitsky, *On Possible Suppression of the Axion Hadron Interactions. (In Russian)*,  
*Sov. J. Nucl. Phys.* **31** (1980) 260 (cit. on p. 17).
- [70] M. Dine, W. Fischler and M. Srednicki,  
*A Simple Solution to the Strong CP Problem with a Harmless Axion*,  
*Phys. Lett. B* **104** (1981) 199 (cit. on p. 17).
- [71] P. Svrcek and E. Witten, *Axions in string theory*, *JHEP* **06** (2006) 051,  
eprint: [hep-th/0605206](https://arxiv.org/abs/hep-th/0605206) (cit. on p. 18).
- [72] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper and J. March-Russell,  
*String Axiverse*, *Phys. Rev. D* **81** (2010) 123530, eprint: [0905.4720](https://arxiv.org/abs/0905.4720) (cit. on p. 18).

- [73] W. J. Marciano, A. Masiero, P. Paradisi and M. Passera, *Contributions of axionlike particles to lepton dipole moments*, *Phys. Rev. D* **94** (2016) 115033, arXiv: [1607.01022 \[hep-ph\]](#) (cit. on p. 18).
- [74] M. Bauer, M. Neubert and A. Thamm, *Axion-Like Particles: Precision Constraints and Open Questions*, *Phys. Rev. D* **100** (2019) 055018, arXiv: [1906.11846 \[hep-ph\]](#) (cit. on p. 18).
- [75] M. Bauer, M. Neubert and A. Thamm, *Collider probes of axion-like particles*, *Journal of High Energy Physics* **2017** (2017), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP12\(2017\)044](http://dx.doi.org/10.1007/JHEP12(2017)044) (cit. on pp. 18, 22, 24, 79).
- [76] H. Primakoff, *Photo-Production of Neutral Mesons in Nuclear Electric Fields and the Mean Life of the Neutral Meson*, *Phys. Rev.* **81** (5 1951) 899, URL: <https://link.aps.org/doi/10.1103/PhysRev.81.899> (cit. on p. 18).
- [77] R. Khatriwada et al., *Axion Dark Matter Experiment: Detailed design and operations*, *Review of Scientific Instruments* **92** (2021) 124502, ISSN: 0034-6748, eprint: [https://pubs.aip.org/aip/rsi/article-pdf/doi/10.1063/5.0037857/16136265/124502\\_1\\_online.pdf](https://pubs.aip.org/aip/rsi/article-pdf/doi/10.1063/5.0037857/16136265/124502_1_online.pdf), URL: <https://doi.org/10.1063/5.0037857> (cit. on p. 19).
- [78] B. M. Brubaker et al., *A Search for Invisible Axion Dark Matter with the Axion Dark Matter eXperiment*, *Phys. Rev. Lett.* **118** (2017) 061302, arXiv: [1607.07410 \[hep-ex\]](#) (cit. on p. 19).
- [79] The MADMAX Working Group, *A new experimental approach to probe QCD axion dark matter in the mass range above 40  $\mu\text{eV}$* , *Eur. Phys. J. C* **79** (2019) 186, arXiv: [1707.08145 \[hep-ex\]](#) (cit. on p. 19).
- [80] T. Schneemann, K. Schmieden and M. Schott, *SUPAX - A Superconducting Axion Search Experiment*, *PoS EPS-HEP2023* (2024) 140 (cit. on p. 19).
- [81] ADMX Collaboration, *Search for Invisible Axion Dark Matter in the 3.3 – 4.2  $\mu\text{eV}$  Mass Range*, *Phys. Rev. Lett.* **127** (26 2021) 261803, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.127.261803> (cit. on p. 20).
- [82] CAST Collaboration, *First Results from the CERN Axion Solar Telescope*, *Phys. Rev. Lett.* **94** (12 2005) 121301, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.94.121301> (cit. on p. 19).
- [83] E. Armengaud et al., *Conceptual design of the International Axion Observatory (IAXO)*, *Journal of Instrumentation* **9** (2014) T05002, URL: <https://dx.doi.org/10.1088/1748-0221/9/05/T05002> (cit. on p. 19).
- [84] IAXO Collaboration, *Conceptual design of BabyIAXO, the intermediate stage towards the International Axion Observatory*, *J. High Energy Phys.* **2021** (2021) 137, arXiv: [2010.12076 \[physics.ins-det\]](#) (cit. on p. 21).
- [85] K. Ehret et al., *New ALPS Results on Hidden-Sector Lightweights*, *Phys. Lett. B* **689** (2010) 149, arXiv: [1004.1313 \[hep-ex\]](#) (cit. on p. 20).

- 
- [86] OSQAR Collaboration, *Axion search with the OSQAR photon regeneration experiment*, *Phys. Rev. D* **78** (2008) 092003, arXiv: 0712.3362 [hep-ex] (cit. on p. 20).
- [87] R. Bähre et al., *Any light particle search II — Technical Design Report*, *Journal of Instrumentation* **8** (2013) T09001, URL: <https://dx.doi.org/10.1088/1748-0221/8/09/T09001> (cit. on p. 20).
- [88] OSQAR Collaboration, *New exclusion limits on scalar and pseudoscalar axionlike particles from light shining through a wall*, *Phys. Rev. D* **92** (2015) 092002, arXiv: 1506.08082 [hep-ex] (cit. on p. 22).
- [89] ATLAS Collaboration, *Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 299, arXiv: 1502.01518 [hep-ex] (cit. on p. 21), Erratum: *Eur. Phys. J. C* **75** (2015) 408.
- [90] ATLAS Collaboration, *Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector*, *Phys. Rev. D* **94** (2016) 032005, arXiv: 1604.07773 [hep-ex] (cit. on p. 21).
- [91] ATLAS Collaboration, *Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC*, *Nature Phys.* **13** (2017) 852, arXiv: 1702.01625 [hep-ex] (cit. on p. 21).
- [92] CMS Collaboration, *Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV*, *Phys. Lett. B* **797** (2019) 134826, arXiv: 1810.04602 [hep-ex] (cit. on pp. 21, 80).
- [93] ATLAS Collaboration, *Measurement of light-by-light scattering and search for axion-like particles with  $2.2\text{ nb}^{-1}$  of Pb+Pb data with the ATLAS detector*, *JHEP* **03** (2021) 243, arXiv: 2008.05355 [hep-ex] (cit. on p. 21), Erratum: *JHEP* **11** (2021) 050.
- [94] F. Neuhaus, *Search for axion-like particles in light-by-light scattering with the ATLAS detector*, English, Dissertation: Johannes Gutenberg-Universität Mainz, 2021, URL: <https://openscience.ub.uni-mainz.de/items/0a36f6cf-af88-4019-8ec9-c0c79e642045> (cit. on p. 22).
- [95] L3 Collaboration, *Search for anomalous  $Z \rightarrow \gamma\gamma\gamma$  events at LEP*, *Physics Letters B* **345** (1995) 609, ISSN: 0370-2693, URL: <https://www.sciencedirect.com/science/article/pii/037026939501612T> (cit. on p. 22).
- [96] Delphi Collaboration, *The reaction  $e^+e^- \rightarrow \gamma\gamma\gamma$  at Z0 energies*, *Physics Letters B* **268** (1991) 296, ISSN: 0370-2693, URL: <https://www.sciencedirect.com/science/article/pii/037026939190819C> (cit. on p. 22).
- [97] ATLAS Collaboration, *Search for Higgs boson decays into a pair of pseudoscalar particles in the  $b\bar{b}\mu\mu$  final state with the ATLAS detector in pp collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **105** (2022) 012006, arXiv: 2110.00313 [hep-ex] (cit. on p. 22).



- [98] ATLAS Collaboration, *Search for Higgs boson decays into pairs of light (pseudo)scalar particles in the  $\gamma\gamma jj$  final state in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **782** (2018) 750, arXiv: 1803.11145 [hep-ex] (cit. on p. 22).
- [99] CMS Collaboration, *Search for new neutral Higgs bosons through the  $H \rightarrow ZA \rightarrow \ell^+ \ell^- b\bar{b}$  process in  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **03** (2020) 055, arXiv: 1911.03781 [hep-ex] (cit. on p. 22).
- [100] CMS Collaboration, *Search for an exotic decay of the Higgs boson into a Z boson and a pseudoscalar particle in proton–proton collisions at  $\sqrt{s} = 13$  TeV*, (2023), arXiv: 2311.00130 [hep-ex] (cit. on p. 23).
- [101] ATLAS Collaboration, *Search for the decay of the Higgs boson to a Z boson and a light pseudoscalar particle decaying to two photons*, (2023), arXiv: 2312.01942 [hep-ex] (cit. on p. 23).
- [102] G. Lamprinoudis, *Search for the decay of the Higgs boson to a Z boson and a long-lived axion-like particle decaying to two photons with the ATLAS Detector*, PhD thesis: Johannes Gutenberg-Universität Mainz, 2025, URL: <https://ubmz.hds.hebis.de/Record/HEB528809571> (cit. on p. 23).
- [103] CMS Collaboration, *Search for exotic Higgs boson decays  $H \rightarrow \mathcal{A}\mathcal{A} \rightarrow 4\gamma$  with events containing two merged diphotons in proton–proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. Lett.* **131** (2023) 101801, arXiv: 2209.06197 [hep-ex] (cit. on p. 23).
- [104] CMS Collaboration, *Search for the exotic decay of the Higgs boson into two light pseudoscalars with four photons in the final state in proton–proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **07** (2023) 148, arXiv: 2208.01469 [hep-ex] (cit. on p. 23).
- [105] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001 (cit. on p. 25).
- [106] O. S. Brüning et al., *LHC Design Report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2004, URL: <https://cds.cern.ch/record/782076> (cit. on p. 25).
- [107] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002 (cit. on p. 25).
- [108] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005 (cit. on p. 25).
- [109] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004 (cit. on p. 25).
- [110] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, (2022), General Photo, URL: <https://cds.cern.ch/record/2800984> (cit. on p. 26).
- [111] R. M. Bianchi and A. Collaboration, “ATLAS experiment schematic or layout illustration”, General Photo, 2022, URL: <https://cds.cern.ch/record/2837191> (cit. on p. 27).
- [112] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 1*, ATLAS-TDR-4; CERN-LHCC-97-016, 1997, URL: <https://cds.cern.ch/record/331063> (cit. on p. 26).
- [113] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 2*, ATLAS-TDR-5; CERN-LHCC-97-017, 1997, URL: <https://cds.cern.ch/record/331064> (cit. on p. 26).

- 
- [114] ATLAS Collaboration, *ATLAS Pixel Detector: Technical Design Report*, ATLAS-TDR-11; CERN-LHCC-98-013, 1998, URL: <https://cds.cern.ch/record/381263> (cit. on p. 26).
- [115] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker in LHC Run 2*, *JINST* **17** (2022) P01013, arXiv: 2109.02591 [physics.ins-det] (cit. on p. 26).
- [116] ATLAS Collaboration, *ATLAS Central Solenoid: Magnet Project Technical Design Report, Volume 4*, ATLAS-TDR-9; CERN-LHCC-97-021, CERN, 1997, URL: <https://cds.cern.ch/record/331067> (cit. on p. 26).
- [117] ATLAS Collaboration, *ATLAS Calorimeter Performance: Technical Design Report*, ATLAS-TDR-1; CERN-LHCC-96-040, 1996, URL: <https://cds.cern.ch/record/331059> (cit. on p. 26).
- [118] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, 1996, URL: <https://cds.cern.ch/record/331061> (cit. on pp. 26, 29, 89).
- [119] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, ATLAS-TDR-3; CERN-LHCC-96-042, 1996, URL: <https://cds.cern.ch/record/331062> (cit. on p. 26).
- [120] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: <https://cds.cern.ch/record/331068> (cit. on p. 27).
- [121] J. Pequeno, “Computer Generated image of the ATLAS calorimeter”, 2008, URL: <https://cds.cern.ch/record/1095927> (cit. on p. 28).
- [122] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 30).
- [123] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 30).
- [124] A. E. Pinto Pinoargote, *Results from muon reconstruction performance with ATLAS at Run-3*, tech. rep., CERN, 2023, URL: <https://cds.cern.ch/record/2870384> (cit. on p. 30).
- [125] G. P. Salam, *Towards jetography*, 2010, URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6> (cit. on p. 30).
- [126] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$ -jet clustering algorithm*, 2008, URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063> (cit. on p. 30).
- [127] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, 2012, URL: <https://doi.org/10.1140/epjc/s10052-012-1896-2> (cit. on p. 30).
- [128] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Rev. D* **96** (2017) 072002, arXiv: 1703.09665 [hep-ex] (cit. on p. 30).

- [129] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 30).
- [130] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 31).
- [131] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: <https://cds.cern.ch/record/1099735> (cit. on p. 31).
- [132] ATLAS Collaboration, *Electron and photon reconstruction and identification in ATLAS: expected performance at high energy and results at 900 GeV*, ATLAS-CONF-2010-005, 2010, URL: <https://cds.cern.ch/record/1273197> (cit. on p. 31).
- [133] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, *JINST* **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on p. 31).
- [134] ATLAS Collaboration, *Photon identification in 2015 ATLAS data*, ATL-PHYS-PUB-2016-014, 2016, URL: <https://cds.cern.ch/record/2203125> (cit. on p. 31).
- [135] ATLAS Collaboration, *ATLAS Computing: Athena Framework*, <https://atlas-software.docs.cern.ch/>, CERN, 2023 (cit. on pp. 32–34).
- [136] F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, *Psychological Review* **65** (1958) 386 (cit. on p. 35).
- [137] K. El Morabit, *Measurement of  $t\bar{t}H(H \rightarrow b\bar{b})$  production in the semi-leptonic  $t\bar{t}$  decay channel at the CMS Experiment*, PhD thesis: Karlsruhe Institute of Technology (KIT), 2021 (cit. on p. 36).
- [138] P. Krämer, *Measurement of  $t\bar{t}H$  Simplified Template Cross Sections in the  $H \rightarrow b\bar{b}$  channel at the CMS experiment*, MA thesis: Karlsruhe Institute of Technology (KIT), 2020 (cit. on p. 36).
- [139] T. Hastie, *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, ed. by R. Tibshirani and J. H. Friedman, 2009 (cit. on p. 36).
- [140] B. Brickwedde, *Measurement of the differential Drell-Yan production cross-section and application of deep convolutional neural networks on event images in the context of pileup mitigation*, Online resource, Dissertation: Johannes Gutenberg-Universität Mainz, 2020 202, URL: <https://doi.org/10.25358/openscience-5122> (cit. on p. 36).
- [141] R. Tibshirani, *Regression Shrinkage and Selection Via the Lasso*, *Journal of the Royal Statistical Society: Series B (Methodological)* **58** (2018) 267, ISSN: 0035-9246, URL: <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x> (cit. on p. 37).



- 
- [142] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15** (2014) 1929, URL: <http://jmlr.org/papers/v15/srivastava14a.html> (cit. on p. 37).
- [143] L. Prechelt, “Early Stopping — But When?”, *Neural Networks: Tricks of the Trade: Second Edition*, ed. by G. Montavon, G. B. Orr and K.-R. Müller, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012 53, ISBN: 978-3-642-35289-8, URL: [https://doi.org/10.1007/978-3-642-35289-8\\_5](https://doi.org/10.1007/978-3-642-35289-8_5) (cit. on p. 37).
- [144] *Multi-Layer Neural Network*, Accessed 25.1.2020, URL: [https://commons.wikimedia.org/wiki/File:Multi-Layer\\_Neural\\_Network-Vector.svg](https://commons.wikimedia.org/wiki/File:Multi-Layer_Neural_Network-Vector.svg) (cit. on p. 37).
- [145] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, *Nature* **323** (1986) 533, ISSN: 1476-4687, URL: <https://doi.org/10.1038/323533a0> (cit. on p. 37).
- [146] T. N. Kipf and M. Welling, *Semi-Supervised Classification with Graph Convolutional Networks*, 2017, arXiv: [1609.02907](https://arxiv.org/abs/1609.02907) [cs.LG], URL: <https://arxiv.org/abs/1609.02907> (cit. on p. 38).
- [147] L. Wu, P. Cui, J. Pei and L. Zhao, eds., *Graph Neural Networks: Foundations, Frontiers, and Applications*, Singapore: Springer Nature Singapore, 2022, ISBN: 978-981-16-6053-5 (print), 978-981-16-6054-2 (e-book), URL: <https://link.springer.com/10.1007/978-981-16-6054-2> (visited on 27/07/2022) (cit. on p. 38).
- [148] Wikipedia contributors, *Convolutional Neural Network, Convolution Worked Example*, Accessed: 2025-04-01, 2023, URL: [https://en.wikipedia.org/wiki/File:Convolutional\\_neural\\_network%2C\\_convolution\\_worked\\_example.png](https://en.wikipedia.org/wiki/File:Convolutional_neural_network%2C_convolution_worked_example.png) (cit. on p. 41).
- [149] Wikipedia contributors, *Convolutional Neural Network, Max Pooling Example*, Accessed: 2025-04-01, 2023, URL: [https://en.wikipedia.org/wiki/File:Convolutional\\_neural\\_network%2C\\_maxpooling.png](https://en.wikipedia.org/wiki/File:Convolutional_neural_network%2C_maxpooling.png) (cit. on p. 42).
- [150] K. Schmieden, *Search for short- and long-lived axion-like particles in  $H \rightarrow aa \rightarrow 4\gamma$  decays with the ATLAS experiment at the LHC*, tech. rep., Restricted to ATLAS members. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2023-040>: CERN, 2023, URL: <https://cds.cern.ch/record/2865101> (cit. on pp. 50, 63, 67, 111–114).
- [151] ATLAS Collaboration, *Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **83** (2023) 982, arXiv: [2212.09379](https://arxiv.org/abs/2212.09379) [hep-ex] (cit. on pp. 52, 70).

- [152] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, *JINST* **15** (2020) P04003, arXiv: 1911.04632 [[physics.ins-det](#)] (cit. on p. 52).
- [153] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 (cit. on p. 52).
- [154] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, arXiv: 1005.4568 [[physics.ins-det](#)] (cit. on p. 52).
- [155] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, arXiv: [hep-ph/0409146](#) (cit. on pp. 52, 54).
- [156] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, arXiv: 0709.2092 [[hep-ph](#)] (cit. on pp. 52, 54).
- [157] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, arXiv: 1002.2581 [[hep-ph](#)] (cit. on pp. 52, 54).
- [158] C. Anastasiou et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, *JHEP* **05** (2016) 058, arXiv: 1602.00695 [[hep-ph](#)] (cit. on p. 52).
- [159] S. Actis, G. Passarino, C. Sturm and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, *Phys. Lett. B* **670** (2008) 12, arXiv: 0809.1301 [[hep-ph](#)] (cit. on p. 52).
- [160] C. Anastasiou, R. Boughezal and F. Petriello, *Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion*, *Journal of High Energy Physics* **2009** (2009) 003, URL: <https://dx.doi.org/10.1088/1126-6708/2009/04/003> (cit. on p. 52).
- [161] ATLAS Collaboration, *Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **09** (2014) 145, arXiv: 1406.3660 [[hep-ex](#)] (cit. on pp. 52, 54).
- [162] NNPDF Collaboration, R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: 1207.1303 [[hep-ph](#)] (cit. on p. 52).
- [163] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159, arXiv: 1410.3012 [[hep-ph](#)] (cit. on pp. 52, 54).
- [164] B. Grzadkowski, H. E. Haber, O. M. Ogreid and P. Osland, *Heavy Higgs boson decays in the alignment limit of the 2HDM*, *Journal of High Energy Physics* **2018** (2018) 56, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP12\(2018\)056](https://doi.org/10.1007/JHEP12(2018)056) (cit. on p. 52).
- [165] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, arXiv: 1905.09127 [[hep-ph](#)] (cit. on p. 53).
- [166] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, arXiv: 0903.1219 [[hep-ph](#)] (cit. on p. 54).

- 
- [167] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419> (cit. on p. 54).
- [168] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039) [[physics.data-an](https://arxiv.org/archive/hep)] (cit. on pp. 57, 58).
- [169] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, *Eur. Phys. J. C* **79** (2019) 205, arXiv: [1810.05087](https://arxiv.org/abs/1810.05087) [[hep-ex](https://arxiv.org/archive/hep)] (cit. on pp. 58, 59).
- [170] ATLAS Collaboration, *Recommendations for the Modeling of Smooth Backgrounds*, ATL-PHYS-PUB-2020-028, 2020, URL: <https://cds.cern.ch/record/2743717> (cit. on pp. 67, 72).
- [171] O. Vujinović, PhD Thesis in preparation, Johannes Gutenberg Universität Mainz, 2025 (cit. on pp. 68, 70, 111).
- [172] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *Eur. Phys. J. C* **80** (2020) 47, arXiv: [1909.00761](https://arxiv.org/abs/1909.00761) [[hep-ex](https://arxiv.org/archive/hep)] (cit. on p. 71).
- [173] L. Heinrich, M. Feickert and G. Stark, *pyhf: v0.7.6*, version 0.7.6, <https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6>, URL: <https://doi.org/10.5281/zenodo.1169739> (cit. on p. 75).
- [174] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: pure-Python implementation of HistFactory statistical models*, *Journal of Open Source Software* **6** (2021) 2823, URL: <https://doi.org/10.21105/joss.02823> (cit. on p. 75).
- [175] A. L. Read, *Presentation of search results: the  $CL_S$  technique*, *J. Phys. G* **28** (2002) 2693 (cit. on p. 78).
- [176] ATLAS Collaboration, *Measurement of light-by-light scattering and search for axion-like particles with  $2.2\text{ nb}^{-1}$  of Pb+Pb data with the ATLAS detector*, *JHEP* **03** (2021) 243, arXiv: [2008.05355](https://arxiv.org/abs/2008.05355) [[hep-ex](https://arxiv.org/archive/hep)], [Erratum: *JHEP* **11** (2021) 050] (cit. on p. 80).
- [177] J. Jaeckel and M. Spannowsky, *Probing MeV to 90 GeV axion-like particles with LEP and LHC*, *Physics Letters B* **753** (2016) 482, ISSN: 0370-2693, URL: <http://dx.doi.org/10.1016/j.physletb.2015.12.037> (cit. on p. 80).
- [178] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from [tensorflow.org](https://www.tensorflow.org/), 2015, URL: <https://www.tensorflow.org/> (cit. on p. 91).
- [179] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825, URL: <http://jmlr.org/papers/v12/pedregosa11a.html> (cit. on p. 91).
- [180] C. R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357, ISSN: 1476-4687, URL: <http://dx.doi.org/10.1038/s41586-020-2649-2> (cit. on p. 91).
- [181] W. McKinney, *Data Structures for Statistical Computing in Python*, (2010) 56, ed. by S. van der Walt and J. Millman (cit. on p. 91).



---

## List of Figures

---

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                    |    |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Schematic of the fundamental interactions in the Standard Model, their corresponding gauge bosons, and the underlying gauge symmetries. . . . .                                                                                                                                                                                                                                                                                    | 4  |
| 2.2 | Standard Model of Elementary Particles. From the left-hand side to the right hand-side, the three generations of fermions can be seen. Quarks are coloured purple, leptons green. Further on the right-hand side, the gauge bosons are depicted in red and the scalar Higgs boson in yellow. Mass, electrical charge and spin of the particles are written in the upper left corner of each particle box. Taken from [19]. . . . . | 5  |
| 2.3 | Running of the strong coupling constant $\alpha_s(Q)$ as measured in various high-energy processes. The data points illustrate the experimentally observed decrease in $\alpha_s$ with increasing energy scale $Q$ , while the band shows the QCD prediction. Figure adapted from [45]. . . . .                                                                                                                                    | 9  |
| 2.4 | Leading Higgs boson production mechanisms at the LHC with approximate cross sections at $\sqrt{s} = 13$ TeV. Values taken from [48] . . . . .                                                                                                                                                                                                                                                                                      | 12 |
| 2.5 | Reduced coupling strength modifiers of the Higgs boson. Taken from [49]. . . . .                                                                                                                                                                                                                                                                                                                                                   | 13 |
| 2.6 | Feynman diagrams for the loop-induced decay $H \rightarrow \gamma\gamma$ . The Higgs couples to a virtual fermion (left) or $W^\pm$ boson (right) loop, with photons radiated from the loop. .                                                                                                                                                                                                                                     | 13 |
| 2.7 | Parton distribution functions for different energies. The gluon contribution is scaled down by $\frac{1}{10}$ . Taken from [53]. . . . .                                                                                                                                                                                                                                                                                           | 15 |
| 3.1 | ALP contribution to the anomalous magnetic moment of the muon. Reproduced from [75]. . . . .                                                                                                                                                                                                                                                                                                                                       | 18 |
| 3.2 | Primakoff effect: photon–axion conversion in a background magnetic field $\mathbf{B}$ . . . . .                                                                                                                                                                                                                                                                                                                                    | 19 |
| 3.3 | 90 % C.L. exclusion limits from latest ADMX searches for invisible axions. Taken from [81] . . . . .                                                                                                                                                                                                                                                                                                                               | 20 |
| 3.4 | Exclusion limits of CAST together with predicted sensitivity of IAXO and babyIAXO. Taken from [84] . . . . .                                                                                                                                                                                                                                                                                                                       | 21 |
| 3.5 | Exclusion limits for ALPS and OSQAR experiments for the search for pseudo-scalar ALPS (a) and scalar ALPS (b). Taken from [88]. . . . .                                                                                                                                                                                                                                                                                            | 22 |
| 3.6 | Important decay channels of ALPs depending on their mass and couplings. . . . .                                                                                                                                                                                                                                                                                                                                                    | 22 |
| 3.7 | Comparison of standard and ALP-mediated light-by-light scattering diagrams . . . . .                                                                                                                                                                                                                                                                                                                                               | 23 |
| 3.8 | The ALP parameter space in dependence of $m_a$ and $C_{a\gamma\gamma}$ with experimentally excluded regions as filled areas. The shaded red area shows the area in which ALPs could explain the $(g - 2)_\mu$ discrepancy and the shaded green area that can be assessed by $H \rightarrow aa$ . Taken from [75]. . . . .                                                                                                          | 24 |

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                          |    |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1 | The LHC accelerator complex [110] . . . . .                                                                                                                                                                                                                                                                                                                                                                                              | 26 |
| 4.2 | Schematic of the ATLAS detector with labelled subsystems. From the inside these are the Inner Detector, the solenoid magnet, the electromagnetic calorimeter and the hadronic calorimeter, the muon chambers and the toroidal magnet. [111] . . . . .                                                                                                                                                                                    | 27 |
| 4.3 | Sketch of the ATLAS calorimeter. [121] . . . . .                                                                                                                                                                                                                                                                                                                                                                                         | 28 |
| 4.4 | Sketch of the sampling layers of the ECAL. The accordion structure is indicated by solid zigzag lines. The dimensions of the calorimeter cells is given as well as the depth of the calorimeter in terms of $X_0$ . [118] . . . . .                                                                                                                                                                                                      | 29 |
| 4.5 | Hits in the TGC detector per Lumiblock (a) and different types of muon tracks per Lumiblock (b). These plots were generated automatically by the ATLAS offline DQ framework [135]. . . . .                                                                                                                                                                                                                                               | 32 |
| 4.6 | ATLAS muon offline DQ plot showing the invariant mass distribution of opposite-sign muon pairs around the $Z$ boson mass. The black histogram represents data from the run under investigation, while the blue histogram shows a reference run. A Gaussian fit (red curve) is used to compare the mean and width of the peak to reference values. This plot was generated automatically by the ATLAS offline DQ framework [135]. . . . . | 33 |
| 4.7 | Invariant mass of opposite-sign muon pairs in different detector regions. Panel (a) shows the mass distribution by detector region, while (b) and (c) show projections of the mean and standard deviation ( $\sigma$ ) of the mass distributions. These plots were generated automatically by the ATLAS offline DQ framework [135]. . . . .                                                                                              | 34 |
| 5.1 | Model of an ANN with one hidden layer, three input features and two output nodes. Taken from [144] . . . . .                                                                                                                                                                                                                                                                                                                             | 37 |
| 5.2 | Commonly used activation functions restricted to outputs between $-1$ and $1$ (a) and without range restrictions (b). . . . .                                                                                                                                                                                                                                                                                                            | 39 |
| 5.3 | Visual representation of a convolution in a CNN. The $3 \times 3$ filter slides over an input matrix, performing elementwise multiplication and summation to produce a new feature map. Taken from [148]. . . . .                                                                                                                                                                                                                        | 41 |
| 5.4 | Sketch how a $2 \times 2$ max pooling window moves across the feature map selecting the maximum value in each region. Taken from [149]. . . . .                                                                                                                                                                                                                                                                                          | 42 |
| 5.5 | Classifier output and corresponding TPR/FPR as a function of threshold. . . . .                                                                                                                                                                                                                                                                                                                                                          | 45 |
| 5.6 | Receiver Operating Characteristic (ROC) and Significance Improvement Characteristic (SIC) curves with exemplary thresholds. . . . .                                                                                                                                                                                                                                                                                                      | 45 |
| 5.7 | Multiclass classification evaluation using confusion matrix with truth class labels on the x-axis and predicted class labels on the y-axis as well as ROC curves for every class/output node. . . . .                                                                                                                                                                                                                                    | 46 |
| 5.8 | Evaluation of regression model using scatter and residual plots. . . . .                                                                                                                                                                                                                                                                                                                                                                 | 46 |
| 6.1 | Feynman diagram for $gg \rightarrow H \rightarrow aa \rightarrow 4\gamma$ . . . . .                                                                                                                                                                                                                                                                                                                                                      | 50 |
| 6.2 | Sketch of special ALP decay topologies. On the left-hand side, two collimated photons with large Lorentz boost are shown. On the right-hand side, the ALP has a significant lifetime and decays with a displacement close to the calorimeter. Taken from [150]. . . . .                                                                                                                                                                  | 50 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |    |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 6.3  | Simulated background processes in the two-photon case, requiring both photons to have at least 15 GeV. The brown filled histograms show prompt photon production, with darker colours indicating events with three and four photons. The grey histograms represent photon + jet processes, with darker shades for additional $c$ and $b$ jets. The light pink histogram shows the SM expectation from $H \rightarrow \gamma\gamma$ . A simple exponential fit is shown as a pink line. The bottom panel shows the ratio of the combined background events to the fit. . . . . | 51 |
| 6.4  | Trigger efficiencies for $m_a = 1$ GeV (blue), $m_a = 5$ GeV (dark blue), $m_a = 30$ GeV (teal), and $m_a = 50$ GeV (dark teal). Solid lines indicate the logical AND of all photon triggers, dashed lines show HLT_2g22_tight, and dash-dotted lines indicate HLT_g35_loose_g25_loose. . . . .                                                                                                                                                                                                                                                                               | 56 |
| 6.5  | Shower shape variables for fake photons (grey), single photons (blue), and merged photons (purple). Taken from [1]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 60 |
| 6.6  | Output of the classifier distinguishing fake photons (yellow) from single photons (light blue) and merged photons (dark blue). Taken from [1]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                        | 61 |
| 6.7  | Output of the classifier distinguishing between single photons (yellow) and merged photons (blue). Two merged photon distributions are shown from different ALP samples. Taken from [1]. . . . .                                                                                                                                                                                                                                                                                                                                                                              | 62 |
| 6.8  | Reconstructed Higgs boson mass distributions for $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ events with $m_a < 3$ GeV in the two-photon categories, comparing the nominal photon energy reconstruction to the corrected method including nearby topo-cluster energy. Left: two merged photon signatures. Middle: one merged and one single photon. Right: two single photons. Taken from [150] . . . . .                                                                                                                                                         | 63 |
| 6.9  | Reconstructed ALP mass for different simulated samples. The blue histograms show $m_a = 5$ GeV samples with different coupling, violet $m_a = 20$ GeV and magenta $m_a = 40$ GeV samples. The shaded histograms show samples with coupling $C_{a\gamma\gamma} = 1$ and the line histograms samples with coupling $C_{a\gamma\gamma} = 5 \times 10^{-4}$ . The yellow histogram shows the invariant mass of the best photon combination for simulated multiphoton samples. [1] . . . . .                                                                                       | 64 |
| 6.10 | Schematic overview of the event categorisation process. Each photon in the event is first evaluated for classification as merged (based on ANN outputs), tight, or loose. The event is then sequentially checked against the six predefined categories: 4S, 1M2S, 3S, 2M, 1M1S, and 2S. If none of the category criteria are satisfied, the event is discarded. . . . .                                                                                                                                                                                                       | 65 |
| 6.11 | Categorization efficiency in percent for different ALP masses and couplings. Blue histograms show the categories for $m_a = 1$ GeV, dark blue for $m_a = 5$ GeV, teal for $m_a = 30$ GeV and darker teal for $m_a = 50$ GeV. Different line styles correspond to different couplings: solid lines for $C_{a\gamma\gamma} = 1$ , dashed lines for $C_{a\gamma\gamma} = 0.01$ , and dash-dotted lines for $C_{a\gamma\gamma} = 5 \times 10^{-4}$ . . . . .                                                                                                                      | 66 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 6.12 | Background estimation for the two-photon categories. Data are shown as black points. The region between the vertical dashed lines is excluded from the fit and corresponds to the blinded signal region. The blue dashed line represents the nominal sideband fit, while the green dashed line corresponds to the alternative fit. The blue histogram with a patterned band indicates the background estimation and its associated systematic uncertainty. The yellow line shows the $H \rightarrow \gamma\gamma$ contribution, scaled to the Standard Model expectation, and the purple line represents an arbitrarily scaled signal sample for visual clarity. [1]              | 69 |
| 6.13 | Background distribution (green) with nominal signal shape in solid red and systematic variations as dashed lines. Left: 2 Merged (a); Right: 1 Merged 1 Single (b). Lower panels show ratios of systematic variations to the nominal signal shape.                                                                                                                                                                                                                                                                                                                                                                                                                                | 77 |
| 6.14 | Post-fit distributions for 2 Merged (a) and 1 Merged 1 Single (b) categories for $m_a = 1$ GeV, $C_{a\gamma\gamma} = 1$ . The filled green histogram shows the fitted background, the grey band indicates its uncertainty, and the black points represent the data.                                                                                                                                                                                                                                                                                                                                                                                                               | 78 |
| 6.15 | Pull plot of the nuisance parameters from the signal-plus-background fit. Each row corresponds to a nuisance parameter. The central dot shows the post-fit value (in standard deviations), and the bar represents its uncertainty.                                                                                                                                                                                                                                                                                                                                                                                                                                                | 78 |
| 6.16 | Correlation matrix of the fit parameters. Strong anti-correlation between spurious signal nuisance parameters and $\mu$ is observed.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 79 |
| 6.17 | Upper limits on the signal strength $\mu$ from fits including only statistical uncertainties (a) and all systematic uncertainties (b). The observed limit is shown where the red curve intersects the solid black line. The expected limit is indicated by the dashed line, with the green and yellow bands representing the $1\sigma$ and $2\sigma$ uncertainties.                                                                                                                                                                                                                                                                                                               | 80 |
| 6.18 | Upper limits at 95 % CL on the branching fraction $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$ for different ALP-photon couplings. Observed limits are shown as solid black lines; expected limits as dashed lines with $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) uncertainty bands [1].                                                                                                                                                                                                                                                                                                                                                           | 81 |
| 6.19 | Upper limits at 95 % CL on $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$ for $C_{a\gamma\gamma} = 1$ and $C_{a\gamma\gamma} = 0.01$ , zoomed in on the low-mass region $0.4 < m_a < 5.5$ GeV [1].                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 81 |
| 6.20 | Two-dimensional exclusion plot in the $m_a$ - $C_{a\gamma\gamma}$ plane. The blue shaded area is excluded by this analysis, while the orange band shows the region where ALPs could explain the $(g - 2)_\mu$ anomaly. [1]                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 82 |
| 7.1  | 2D background estimation for $m_a = 40$ GeV. The centre rectangle depicts the signal region with background estimation. CR1 is inside the red dashed rectangle excluding the SR. CR2 includes the whole histogram range excluding the SR. The black dots show simulated events. In the right-handside plot, the yields in CR1 and CR2 are shown, as well as the transfer factors which are used to calculate the background estimate and the alternative background estimate. The difference between background estimate and alternative estimate is used as systematic uncertainty. Statistical uncertainty and number of simulated events in the SR are shown in separate bins. | 85 |



|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 7.2 | All simulated (a) and measured (b) events fulfilling selection criteria in a 2D plane with $m_{4\gamma}$ on x-axis and $m_a^{\text{reco}}$ on y-axis. Signal and control regions for $m_a = 10 \text{ GeV}$ and $m_a = 40 \text{ GeV}$ are shown exemplary. Signal candidates in the SR of the two examples are filled black squares while background candidates are depicted as empty squares. Control regions are shown as coloured rectangles with the SR surrounded by a solid coloured line in their middle. [1] | 85 |
| 7.3 | Comparison of background estimation and data for categories in the search for long-lived ALPs as well as for different mass hypotheses in the search for promptly decaying ALPs. Background estimations are shown as histogram with a solid blue line and blue shaded areas for systematic uncertainties. Data is shown as black dots. $H \rightarrow \gamma\gamma$ scaled to the SM expectation are added in the two photon categories as filled yellow histogram.[1]                                                | 86 |
| 7.4 | Upper limits on the branching fraction $\mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$ at 95 % CL for promptly decaying ALPs.[1]                                                                                                                                                                                                                                                                                                                                                                 | 87 |
| 8.1 | A fake photon in all four calorimeter layers. The energy deposits are wide spread over many calorimeter cells, but only one main prong can be seen.                                                                                                                                                                                                                                                                                                                                                                   | 91 |
| 8.2 | A single photon in all four calorimeter layers. Only one prong can be seen and the energy deposit is very narrow, only few neighbouring cells are hit.                                                                                                                                                                                                                                                                                                                                                                | 92 |
| 8.3 | A merged photon in all four calorimeter layers. Two clear prongs can be seen and the energy deposits are spread across multiple calorimeter cells.                                                                                                                                                                                                                                                                                                                                                                    | 93 |
| 8.4 | Loss function and categorical accuracy per epoch for training set in blue and for the validation set in red.                                                                                                                                                                                                                                                                                                                                                                                                          | 94 |
| 8.5 | Classifier output in the fake photon output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.                                                                                                                                               | 95 |
| 8.6 | Classifier output in the single photon output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.                                                                                                                                             | 96 |
| 8.7 | Classifier output in the merged photon output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed.                                                                                                                                             | 97 |
| 8.8 | ROC curves for every output node.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 98 |
| 8.9 | Normalized confusion matrices with the truth classes on the x-axis and the predicted classes on the y-axis.                                                                                                                                                                                                                                                                                                                                                                                                           | 98 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |     |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 8.10 | Classifier output in each output node for fake photons in blue, for single photons in yellow and for merged photons in red. Photons from the training set are depicted with dots while photons from the test set are shown as shaded histograms. The ratio plot at the bottom shows the ratio between training and test set. A good closure can be observed. . . . .                                                                                                                                                               | 99  |
| 8.11 | Classification efficiency per class as a function of photon $p_T$ . Panel (a) uses equal training sample sizes without $p_T$ -based weights. Panel (b) shows results after applying $p_T$ reweighting. . . . .                                                                                                                                                                                                                                                                                                                     | 99  |
| 8.12 | Two-dimensional response matrix showing the correlation between the CNN-predicted (x-axis) and true values (y-axis) of the ALP mass $m_a$ . For better visualisation, each row is normalised to 1 (i.e. for each ALP mass the same amount of entries is shown). A diagonal line indicates perfect agreement; deviations from this line reflect the regression performance, including bias and resolution across the mass range. . . . .                                                                                            | 100 |
| 8.13 | Regression CNN output for different masses. Each distribution is labeled with its mean and RMS. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                            | 101 |
| 8.14 | Mean and RMS of the regression results for different truth ALP masses (a) and the residuals of $\text{Mean} - m_a^{\text{truth}}$ for different truth masses (b). A linear fit was performed on (b). . . . .                                                                                                                                                                                                                                                                                                                       | 101 |
| 8.15 | Comparison of correction methods and the uncorrected regression result (a). (b) shows the regression matrix with the correction based on the prediction itself applied. It can be seen, that the correlation becomes slightly better compared to the original figure. . . . .                                                                                                                                                                                                                                                      | 102 |
| 8.16 | Comparison between regression results for fake (light brown), single (dark brown), and merged (red line) photons. The CNN predicts higher mass values primarily for merged photon signatures. . . . .                                                                                                                                                                                                                                                                                                                              | 103 |
| A.1  | Invariant mass distribution of the reconstructed four photon system in four bins of reconstructed ALP mass. The blue dashed line shows the nominal sideband fit function. The estimated background is shown as blue histogram with blue shaded areas depicting systematic uncertainties obtained from a sideband fit with reduced range. The green dotted line shows the alternative sideband fit function. Purple and violet histograms show different signal shapes. The black dots show the measured data. . . . .              | 108 |
| A.2  | Invariant mass distribution of the reconstructed four photon system in four bins of reconstructed ALP mass. The blue dashed line shows the nominal sideband fit function. The estimated background is shown as blue histogram with blue shaded areas depicting systematic uncertainties obtained from a sideband fit with reduced range. The green dotted line shows the alternative sideband fit function. Purple and violet histograms show different signal shapes. The black dots show the actual measured data. [1] . . . . . | 109 |
| A.3  | Shower-shape $R_\varphi$ in for data and simulation in the close, medium and far region. The black dots show data while the grey histogram shows simulation. The blue histogram shows the simulated data reweighted to the data distribution in the near region. [150] . . . . .                                                                                                                                                                                                                                                   | 111 |
| A.4  | Mean and RMS of the $R_\varphi$ shower-shape with interpolation. [150] . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 112 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                        |     |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| A.5  | Classifier efficiencies for nominal (black) and fudged (blue) input variables with respect to the decay length of ALPs. The bottom row shows the ratio between fudged and nominal classifier efficiencies. [150]                                                                                                                                                                                                       | 112 |
| A.6  | Efficiencies of the ATLAS photon identification for nominal (black) and fudged (blue) input variables with respect to the decay length of ALPs. The bottom row shows the ratio between fudged and nominal ID efficiencies. This is used as symmetric uncertainty. [150]                                                                                                                                                | 113 |
| A.7  | The full Keras model architecture visualized. The model follows a sequential design, with three Conv2D-BatchNormalization-MaxPooling blocks for feature extraction, followed by a Flatten layer and a series of Dense layers for classification. The final layer uses a softmax activation to output probabilities over three classes.                                                                                 | 117 |
| A.8  | Filters from the first Conv2D layer. These filters generally display high-frequency patterns with sharp transitions, such as edges and gradients, which are typical in early convolutional layers. The contrast and symmetry observed in filters like 0 and 1 suggest that they are acting as edge detectors in specific orientations. Others, like filter 4, appear more diffuse, capturing broader spatial patterns. | 118 |
| A.9  | Filters from the second Conv2D layer (conv2d_1). These filters show smoother and more complex structures compared to the first layer. The patterns are less edge-like and more texture-based, indicating that the network is starting to integrate and abstract information from the earlier spatial features. Filters 2 and 5 show a strong diagonal gradient, potentially detecting orientation-based textures.      | 118 |
| A.10 | Filters from the third Conv2D layer (conv2d_2). These filters exhibit more diffuse and irregular patterns, which is typical for deeper layers focusing on abstract, class-specific patterns. Some filters (e.g., 0 and 3) are less visually structured, which may suggest underutilization or redundancy, while others (e.g., filter 4) still preserve directional patterns.                                           | 118 |
| A.11 | Heatmap of the first Dense layer (dense) weights, connecting 1280 features to 100 output neurons. The weights are mostly low-magnitude and centered around zero, with no extreme outliers. This indicates balanced initialization and controlled learning, likely aided by dropout and normalization.                                                                                                                  | 119 |
| A.12 | Weights of Dense layer dense_1 (100 to 25 units). Compared to the first layer, some input features dominate more clearly, suggesting that certain neurons specialize in specific patterns from earlier layers.                                                                                                                                                                                                         | 120 |
| A.13 | Weight matrix for Dense layer dense_2 (25 to 10 units). A smaller network width results in more sparsity and sharp contrasts, which may reflect feature compression or focused decision paths leading into the final classifier.                                                                                                                                                                                       | 121 |
| A.14 | Weights for the output layer dense_3, connecting 10 neurons to 3 softmax outputs. The matrix suggests each class is influenced by a different subset of the penultimate layer, as indicated by column-specific patterns in the heatmap.                                                                                                                                                                                | 122 |
| A.15 | Results for all photons with $p_T > 15$ GeV. The plots on the left hand-side result from a training with a weight of 1 for each input, whereas the plots are from a training that used $p_T$ reweighting                                                                                                                                                                                                               | 123 |
| A.16 | Results a training with equal amounts of photons with $p_T > 15$ GeV. The plots on the left hand-side result from a training with a weight of 1 for each input, whereas the plots are from a training that used $p_T$ reweighting                                                                                                                                                                                      | 124 |

## List of Figures

---

|                                                                                                                                                            |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| A.17 2D responses for different $m_a$ regression trainings with similar settings . . . . .                                                                 | 126 |
| A.18 1D results for different $m_a$ regression trainings with similar settings. Mean and RMS<br>are shown for different truth $m_a$ in the legend. . . . . | 127 |

---

## List of Tables

---

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1  | Approximate quark masses, exact values are found in [27] . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                 | 6  |
| 2.2  | Approximate lepton masses, exact values are found in [27] . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                | 7  |
| 2.3  | Properties of the Standard Model vector bosons. Mass values are approximate and taken from [27]. . . . .                                                                                                                                                                                                                                                                                                                                                                           | 11 |
| 5.1  | Commonly used activation functions with definition and first derivation . . . . .                                                                                                                                                                                                                                                                                                                                                                                                  | 39 |
| 6.1  | Overview of used background samples with full names of the MC background sets .                                                                                                                                                                                                                                                                                                                                                                                                    | 55 |
| 6.2  | Analysis pre-selections by $m_a$ category . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                | 57 |
| 6.3  | Explanation of shower shape variables. The last two columns indicate if the variable is used for the Tight or Loose photon ID. Recreated from [169]. . . . .                                                                                                                                                                                                                                                                                                                       | 59 |
| 6.4  | Training samples and number of feature vectors in each subset. . . . .                                                                                                                                                                                                                                                                                                                                                                                                             | 60 |
| 6.5  | Cut-flow and number of observed events for selected signal samples with $C_{a\gamma\gamma} = 1$ in the mass window $110 \text{ GeV} < m_a < 130 \text{ GeV}$ . Recreated from [150] . . . . .                                                                                                                                                                                                                                                                                      | 67 |
| 6.6  | Fit functions, fit ranges and range of signal region for all categories and $m_a^{\text{reco}}$ bins. The first two columns show the category and the $m_a^{\text{reco}}$ bin, the third the $m_a$ range considered to define fitting and blinding ranges. Afterwards, the nominal fit function (nom fkt) and the alternative fit function (alt fkt) are listed. The last columns show the nominal (nom) fitting range, the alternative (alt) range and the blinded signal region. | 68 |
| 6.7  | Overview of relative systematic uncertainties for different $m_a$ and $C_{a\gamma\gamma}$ in all relevant categories. Taken from [1]. . . . .                                                                                                                                                                                                                                                                                                                                      | 74 |
| 6.8  | Categories used in the fit. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                | 75 |
| 6.9  | Name, short description and type of systematic uncertainties with corresponding constraint terms. The <b>type</b> column specifies whether the uncertainty affects only the normalisation ( <b>norm</b> ), both shape and normalisation ( <b>shape</b> ), or is a bin-by-bin statistical uncertainty ( <b>stat</b> ). <b>Global scale factor</b> refers to a single normalisation factor affecting all signal and background yields. . . . .                                       | 75 |
| 6.10 | Simulated signal ( $H \rightarrow aa$ ) and simulated $H \rightarrow \gamma\gamma$ background events with background events from data-driven estimation and $S/\sqrt{B}$ . . . . .                                                                                                                                                                                                                                                                                                 | 76 |
| 7.1  | Definition of the stepsize that is used to define signal and control regions for different $m_a$ . . . . .                                                                                                                                                                                                                                                                                                                                                                         | 84 |
| 7.2  | Definition of signal and control regions for the prompt photon analysis. The signal region is excluded from both control regions. . . . .                                                                                                                                                                                                                                                                                                                                          | 84 |
| 7.3  | Limit setting using asymptotic fits. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                       | 86 |

|     |                                                                                                                                                                                                                         |     |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 7.4 | Toybased limit setting using 20k toys. . . . .                                                                                                                                                                          | 87  |
| 8.1 | Granularity of the barrel ECal. Taken from [118] . . . . .                                                                                                                                                              | 89  |
| A.1 | List of studied trigger chains. Marked are the 2 trigger chains with the highest signal efficiency per year. Out of those only the trigger chain with the largest signal efficiency is used. Taken from [150] . . . . . | 114 |
| A.2 | Good Run Lists used when processing data samples. . . . .                                                                                                                                                               | 115 |
| A.3 | Number of merged photons as a function of the ALP mass $m_a$ . . . . .                                                                                                                                                  | 125 |



