ETP-KA/2025-20

Search for Axion-Like Particles Decaying to Two Photons at Belle II

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Alexander Heidelbach

Tag der mündlichen Prüfung:	28 November 2025
Referent:	Prof. Dr. Torben Ferber
Korreferent:	Prof. Dr. Markus Klute

Disclaimer

Data analyses in high-energy physics, such as the measurement presented in this doctoral thesis, are a collaborative effort. The SuperKEKB particle accelerator, which provides the particle beams essential for all studies at Belle II, was built and is operated and maintained by the SuperKEKB accelerator group. The Belle II detector was built, is maintained, and is operated by the Belle II collaboration. The Belle II collaboration also creates the simulated and recorded datasets and maintains the computing infrastructure necessary to process them. The software environment necessary for studies with Belle II data plays an important role and was created and is maintained by the collaboration. I have been a part of the Belle II collaboration since 2021 and performed all studies detailed in this thesis except for the following:

- The development of the `MadGraph5 aMC@NLO` model used for the generation of the signal samples (Section 3.4.1)
- The determination of correction factors and the associated systematic uncertainties for:
  - The recorded data to simulation photon efficiency correction in Section 8.1
  - The photon energy bias correction in recorded data in Section 8.2

Furthermore, for teaching purposes, I have worked in close collaboration with my student Frederik Schmitt on the analysis pipeline. While I provided the original idea, implementation and structure of the analysis code pipeline, including the development of the plotting and fitting framework, my student and I have extended the pipeline to create a common toolset for our closely related work. While attributing individual contributions in detail is not straightforward, the anaylsis codes' Git repository provides a transparent record of authorship and development. All results, plots, fits and their intepretation that I present in this thesis are my own making.

This thesis incorporates the use of Artificial Intelligence (AI) tools to help with grammatical or stylistic improvement of text, and program code creation.

Grammarly* is utilised throughout the thesis for spell and grammar checks, as well as for paraphrasing individual, selected sentences to improve clarity and precision in academic writing. I have approved all suggested changes.

ChatGPT† is used to aid the development of C++ and Python code, in particular code restructuring and optimisation that do not constitute the core scientific work of this thesis. I have approved and tested all suggestions to provide robust and reliable results.

*Grammarly: An AI writing assistant. See https://app.grammarly.com/ (Access Date: 2024-09-13).

†ChatGPT: A virtual AI assistant based on large language models. See https://openai.com/chatgpt/ (Access Date: 2024-09-13).

Contents

1 Introduction **1**

2 Theoretical Foundations **3**

2.1 The QCD Axion 3

2.2 Axion-Like Particles 5

2.3 Signature at Belle II 7

2.4 Existing Limits 10

3 The Experimental Setup **13**

3.1 The SuperKEKB Accelerator 13

3.2 The Belle II Detector 14

3.2.1 Tracking Detectors 15

3.2.2 Particle Identification Detectors 16

3.2.3 Electromagnetic Calorimeter 17

3.2.4 K_L^0 and Muon Detector 18

3.2.5 Trigger System 19

3.3 Datasets 20

3.3.1 Signal Samples 20

3.3.2 Background Samples 21

3.3.3 Recorded Dataset 24

3.4 Datasets 24

3.4.1 Signal Samples 24

3.4.2 Background Samples 25

3.4.3 Recorded Dataset 28

4 Reconstruction **29**

4.1 Event Reconstruction 30

4.1.1 Photon and Event Candidates 30

4.1.2 Best Candidate Selection 31

4.1.3 Kinematic Fit 32

4.1.4 Summary 32

4.2 Event Selection 34

4.2.1 Number of Clean Tracks . . . 34
4.2.2 Pair Conversion Veto . . . 36
4.2.3 Additional Photons . . . 39
4.2.4 π^0 Veto . . . 41
4.2.5 Summary . . . 45
4.3 Candidate Selection . . . 47
4.3.1 ALP Helicity . . . 47
4.3.2 Photon Energy . . . 50
4.3.3 Summary . . . 52
4.4 MVA Selection . . . 53
4.4.1 Training Procedure . . . 54
4.4.2 Kinematic Selection . . . 57
4.4.3 Merged Photon Selection . . . 62
4.5 Summary . . . 74

5 Signal Efficiency 77
5.1 Trigger Efficiency . . . 77
5.1.1 L1 Trigger . . . 78
5.1.2 HLT Trigger . . . 86
5.2 Run-Dependent Signal Efficiency . . . 88

6 Signal and Background Parametrisation 93
6.1 Signal Parametrisation . . . 93
6.2 Background Parametrisation . . . 97

7 Control Regions 101
7.1 Inverted π^0 Veto . . . 102
7.2 η Region . . . 108
7.3 ω Region . . . 111

8 Corrections and Systematic Uncertainties 113
8.1 Photon Efficiency Correction . . . 113
8.2 Energy Bias Correction . . . 116
8.3 Photon Resolution . . . 117
8.4 Trigger Efficiency . . . 121
8.5 Run-Dependent Interpolation . . . 121
8.6 Selection . . . 123
8.7 ISR Modelling . . . 128
8.8 Signal Shape . . . 129
8.9 Luminosity . . . 130
8.10 Summary . . . 131

9 Signal Extraction **133**
9.1 Fit Model . . . 134
9.2 Significance . . . 137
9.3 Upper Limit . . . 139
9.4 Model Dependent Sensitivity . . . 140
9.5 Validation . . . 141
9.5.1 Look-Elsewhere Effect . . . 142
9.5.2 Fit Bias . . . 144
9.5.3 Fit Linearity and Coverage . . . 145
9.5.4 Impact of Systematic Uncertainties . . . 150

10 Results **151**
10.1 Distributions in Recorded Data . . . 151
10.2 Trigger Efficiency . . . 154
10.3 Model-Independent Result . . . 155
10.4 Model-Dependent Result . . . 159

11 Conclusion **161**

Bibliography **165**

Appendix A Feature for MVA-based Selection **174**
A.1 Features for Kinematic Selection . . . 174
A.2 Features for Merged Photon Selection . . . 177

Appendix B Trigger Bit Definitions **179**

Appendix C $q\bar{q}$ Final States in the Control Regions **181**

Appendix D Features for Kinematic Selection in the Control Regions **183**

Acronyms **189**

Chapter 1

Introduction

"If you wish to make an apple pie from scratch, you must first invent the universe."
— Carl Sagan

Yet, what are the apples for the pie made of? And how does the universe need to look like, if one wants to bake a pie? Scientists have been trying to answer these questions for centuries.

In recent decades, physicists have answered many important questions by developing the Standard Model of particle physics (SM) [1–7] and the Lambda cold dark matter (ΛCDM) model. The SM theory explains that the apple is made of a zoo of elementary particles, their anti-particles, and the forces between them. Thirteen years ago, with the discovery of the Higgs boson [8, 9], physicists found the last SM piece which explains how these particles acquire their mass. Yet to this day, with our theories, we cannot explain how the universe created the apples or where all the anti-apples have gone. While I cannot answer these questions in the thesis I present here, I aim to show how my peers and I continue in the quest towards the missing explanations through the search for new physics beyond the SM.

Efforts for this search are intensifying in every possible aspect. At the high energy frontier, not only is the Large Hadron Collider (LHC) [10] being continuously upgraded [11] and improved, but also the next generation collider [12] is in preparation to explore the possibility of even heavier particles and deviations in the SM. At the cosmic frontier, detectors are designed to be sensitive to the smallest signals, enabling them to probe even the slightest interactions between the SM and dark matter (DM) [13], and thereby bringing us one step closer to understanding the formation of our universe. And finally, at the high intensity frontier, fixed-target experiments, forward facilities and lepton colliders, such as Belle II [14], diversify their programs [15, 16] to search not only for the rarest SM processes, which shed light into the matter anti-matter asymmetry, but also for weakly interactive massive particles (WIMPs) and other extensions of the SM.

In this thesis, I present a search for axion-like particles (ALPs), a specific type of WIMPs, at the Belle II experiment. ALPs arise in many extensions of the SM [17] due their fundamental properties. Motivated by the quantum chromodynamics (QCD) axion [18–21] but without the strict relationship of mass and coupling, ALPs provide a wide open field for new physics searches. They are pseudoscalar particles with potentially low mass and feeble interaction with the SM. ALPs with masses at the GeV scale connect to several open problems in particle physics [22–24] and can also serve as a mediator between the SM and DM [25]. In this work, I focus on ALPs, which couple to the neutral electroweak sector, specifically to photons. I search for the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process via a significant peak in the reconstructed diphoton mass distribution in the data collected by Belle II from 2019 to 2022. Finally, I do not find any statistically significant evidence for the signature and thus place new constraints on the cross section of the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process and the model-dependent ALP-photon coupling. The limits that I place on the ALP-photon coupling are the most stringent at the GeV scale parameter space today.

In Chapter 2, I begin with a brief overview of the theoretical foundations of axions and ALPs. I explain the three-photon signature in the Belle II detector, outline the analysis strategy and give an overview of the existing constraints on the ALP-photon coupling. Following, Chapter 3 describes the experimental setup by highlighting the most important parts of the accelerator and detector facilities, and outlining the simulated and recorded datasets used to develop and execute the analysis. In Chapter 4, I discuss the reconstruction and selection procedure of the searched-for signature and extend the discussion to the trigger and signal efficiency in Chapter 5. I then define the parameterisations of the signal and background distributions in the reconstructed diphoton mass in Chapter 6. In Chapter 7, I show studies in the control regions, which are used to derive some of the corrections and systematic uncertainties that I discuss in Chapter 8. In Chapter 9, I explain the signal extraction method and discuss the statistical evaluation. Finally, I show the results obtained in recorded data in Chapter 10 and conclude the search in Chapter 11.

Chapter 2

Theoretical Foundations

In this chapter, I present an overview of the theoretical foundations relevant to this work. I start by introducing the strong *CP* problem in quantum chromodynamics (QCD) and the proposed solution via the QCD axion in Section 2.1. Building upon this, I discuss axion-like particles (ALPs) as generalisations of the QCD axion in Section 2.2, focusing on their couplings to photons and production mechanisms at electron-positron colliders. In Section 2.3, I describe the experimental signature of ALPs at the Belle II experiment and outline the analysis strategy. Finally, in Section 2.4, I summarise the current experimental constraints on ALPs.

2.1 The QCD Axion

The Standard Model of particle physics (SM) is a widely accepted theory, as it successfully predicts and explains most observations in particle physics. Nevertheless, it leaves several open questions, theoretical puzzles, and experimental anomalies unresolved. To address these issues, the scientific community has developed various extensions and modifications to the SM. In this section, I briefly discuss one of these puzzles - the strong *CP* problem - and a proposed solution, the QCD axion [18–21]. I then use some of the fundamental properties of the QCD axion to motivate later the search for ALPs. I emphasise, however, that this represents only one possible motivation for ALPs as extensions to the SM. For a more cosmological perspective on ALPs, please refer to [17].

An anomalous symmetry is one for which the classical action remains invariant, while the corresponding quantum theory does not. In the case of a chiral rotation of the quark fields, $\psi \to e^{i\gamma_5\theta}\psi$, the action itself is invariant, but the path integral measure is not [26]. This anomaly introduces a *CP*-violating term in the QCD Lagrangian, which can be written as

$$\mathcal{L}_\theta = \theta \frac{g_s^2}{32\pi^2} G^a_{\mu\nu}\tilde{G}^{a,\mu\nu}, \tag{2.1}$$

where θ is the chiral rotational phase, g_s is the strong coupling constant, $G^a_{\mu\nu}$ is the gluon field strength tensor, and $\tilde{G}^{a\mu\nu}$ is its dual. By rotating the quark fields into the mass basis $\psi \to e^{i\alpha}\psi$, the *CP* violating term can be shifted to express a physical, observable parameter as

$$\mathcal{L}_{\bar{\theta}} = \bar{\theta}\frac{g_s^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^{a,\mu\nu}, \tag{2.2}$$

where $\bar{\theta} = \theta + 2\alpha = \theta - \arg\det Y_u Y_d$. Here, Y_u and Y_d denote the up- and down-type quark Yukawa matrices, respectively. In this formulation, any chiral rotation of the quark fields leaves $\bar{\theta}$ invariant. Thus, $\bar{\theta}$, commonly referred to as the strong *CP* phase, is a physical parameter that can, in principle, be measured experimentally.

A direct consequence of a non-zero $\bar{\theta}$ is the existence of a permanent electric dipole moment of the neutron, d_N. Theoretical calculations [27] based on pion-loop contributions predict the neutron dipole moment to be on the order of

$$d_N = 5.2 \times 10^{-16}\,\mathrm{e\,cm} \cdot \bar{\theta}. \tag{2.3}$$

However, experimental measurements [28] place an Upper Limit (UL) on the neutron electric dipole moment of

$$d_N < 1.8 \times 10^{-26}\,\mathrm{e\,cm}. \tag{2.4}$$

This result implies that the *CP*-violating phase must satisfy $\bar{\theta} < 10^{-10}$, which is significantly smaller than the observed *CP* violation in the electroweak sector [29]. This discrepancy is known as the strong *CP* problem.

The most prominent solution to the strong *CP* problem is the QCD axion [18–21]. This solution proposes the existence of a new global chiral symmetry, $U(1)_{\mathrm{PQ}}$, which is spontaneously broken at an energy scale f_{PQ}. Due to the spontaneous breaking combined with the chiral anomaly, a new physical field $a'(x)$ emerges in the same form of the anomaly in the Lagrangian, providing a massive pseudo-Nambu-Goldstone boson known as the QCD axion. Under the chiral rotation of this new field, the strong *CP* violating term can be transformed as

$$\mathcal{L}_{\bar{\theta}} = \left(\bar{\theta} + \frac{a'}{f_{a'}}\right)\frac{g_s^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^{a,\mu\nu}, \tag{2.5}$$

where $f_{a'}$ is the QCD axion decay constant, which is proportional to the symmetry breaking scale f_{PQ}. In this formulation, the QCD axion field dynamically relaxes to a vacuum expectation value of $\langle a' \rangle = -\bar{\theta} f_{a'}$, effectively cancelling the strong *CP* violating term. In the chiral limit where the up and down quark masses are much smaller than the QCD scale, the relationship

$$m_{a'} f_{a'} \approx m_\pi f_\pi, \tag{2.6}$$

where f_π and m_π are the pion decay constant and mass, respectively, follows [30].

Although Eq. (2.6) provides a specific relationship between the mass and coupling of the QCD axion, many extensions provide a wide range of possible values for these parameters [1]. The most commonly used benchmark models are the KSVZ [31, 32] and DFSZ [33, 34] models. In the former model, new, heavy quarks with a $U(1)_{\mathrm{PQ}}$ charge are introduced, while in the latter model, the SM Higgs doublet is extended, and fermions carry the $U(1)_{\mathrm{PQ}}$ charge. In many of these models, the QCD axion is predicted to be extremely light and weakly coupled to the SM particles, making it an excellent candidate for dark matter (DM) or other cosmological applications [17].

2.2 Axion-Like Particles

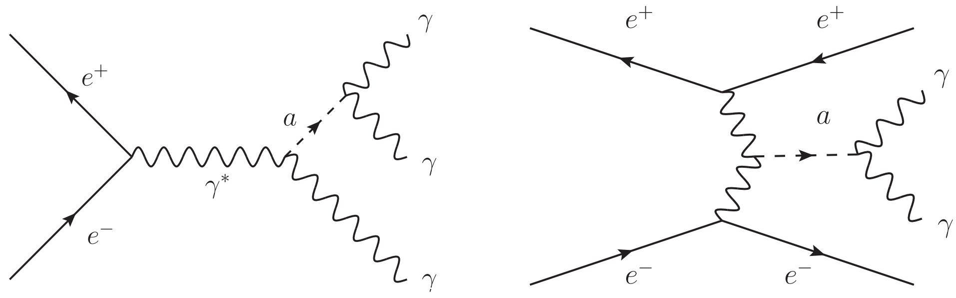
ALPs are a generalisation of the QCD axion discussed in the previous section. The key difference is that ALP models do not necessarily solve the strong *CP* problem, hence, relieving the strict relationship between the mass and coupling strength present in QCD axion models (see Eq. (2.6)). In any other aspect, ALPs share the same properties: Originating from a spontaneously broken, global, anomalous chiral symmetry, they are pseudo-Nambu-Goldstone bosons that couple weakly to the SM. One might even argue that the ALP is naturally motivated, as the SM is missing a fundamental pseudo-scalar particle. Moreover, with in principle free mass m_a and decay constant f_a parameters, ALPs can be accommodated in a wide range of theoretical models. Particularly interesting is the case where the f_a scale is sufficiently higher than the electroweak scale. In this case, not only the lack of new particles at the Large Hadron Collider (LHC) is explained, but also the low ALP mass and feeble interaction with the SM particles explains why ALPs have not yet been observed [35].

While ALPs with masses smaller than the MeV/c^2 scale have a plethora of applications in astrophysics and cosmology [36], I focus in this section on ALP masses in the MeV/c^2 to GeV/c^2 range, as this is the range where Belle II is sensitive to. In this mass range, ALPs can contribute to a wide range of open questions below and at the electroweak scale or serve as a mediator to the dark sector [25]. In this work, I focus on ALPs that couple dominantly to electroweak bosons, specifically photons. Excluding the coupling to fermions or gluons in the following discussion is, on the one hand, motivated by tight experimental constraints on these couplings, and on the other hand, they would not contribute to this work's signature on the tree-level. The strong constraints mainly arise from the fact that such couplings would enhance flavour-changing neutral current processes, which are probed by searches for rare decays [37, 38] and await further input from searches for long-lived signatures.

The relevant interaction terms for an ALP coupling to the electroweak gauge bosons can, according to [25], be written as

$$\mathcal{L}_{\mathrm{ALP}} = \frac{1}{2}\partial^\mu a \partial_\mu a - \frac{1}{2}m_a^2 a^2 + \frac{c_B}{4f_a} a B_{\mu\nu}\tilde{B}^{\mu\nu} + \frac{c_W}{4f_a} a W^i_{\mu\nu}\tilde{W}^{i,\mu\nu}, \tag{2.7}$$

where $B_{\mu\nu}$ and $W^i_{\mu\nu}$ are the field strength tensors of the $U(1)_Y$ and $SU(2)_L$ gauge fields, respectively, and $\tilde{B}^{\mu\nu}$ and $\tilde{W}^{i,\mu\nu}$ are their duals. The dimensionless coefficients c_B and c_W



(a) ALP-strahlung process (b) VBF process

Figure 2.1: Leading-order ALP production mechanisms in electron-positron collisions: ALP-strahlung process (left) and vector-boson fusion (VBF) process (right). Taken from [25].

parameterise the coupling strengths of the ALP to the hypercharge and weak isospin gauge fields, respectively. After electroweak symmetry breaking, the couplings of an ALP to the electroweak gauge bosons can be expressed as

$$\mathcal{L} \propto -\frac{g_{a\gamma\gamma}}{4} aF_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{g_{a\gamma Z}}{4} aF_{\mu\nu}\tilde{Z}^{\mu\nu} - \frac{g_{aZZ}}{4} aZ_{\mu\nu}\tilde{Z}^{\mu\nu} - \frac{g_{aWW}}{4} aW^{+}_{\mu\nu}\tilde{W}^{-,\mu\nu}, \tag{2.8}$$

where $F_{\mu\nu}$, $Z_{\mu\nu}$, and $W^{\pm}_{\mu\nu}$ are the field strength tensors of the photon, Z boson, and $W^{\pm}$ bosons, respectively, and $\tilde{F}^{\mu\nu}$, $\tilde{Z}^{\mu\nu}$, and $\tilde{W}^{\pm\,\mu\nu}$ are their duals. For this work, the most relevant coupling is the one to photons, $g_{a\gamma\gamma}$, which can be expressed as

$$g_{a\gamma\gamma} = \frac{c_B \cos^2\theta_W + c_W \sin^2\theta_W}{f_a}, \tag{2.9}$$

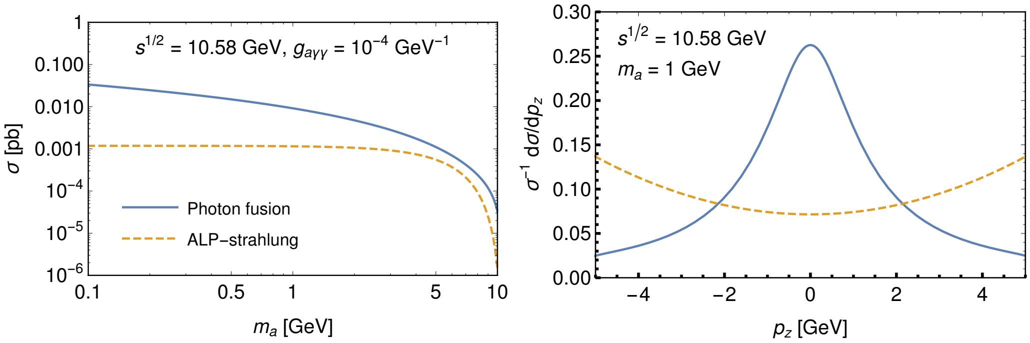
where θ_W is the Weinberg angle. The $a\gamma Z$ coupling $g_{a\gamma Z}$ can be expressed similarly as

$$g_{a\gamma Z} = \frac{\sin(2\theta_W)(c_W - c_B)}{f_a}. \tag{2.10}$$

In principle, as long as c_B and c_W are treated as independent parameters, $g_{a\gamma\gamma}$ and $g_{a\gamma Z}$ are independent as well. However, depending on future experimental results, *e.g.* in $b \to sa$ transitions [39], it might be possible that $c_B \gg c_W$ could be favoured, which would lead to the relationship where $g_{a\gamma\gamma} \sim -g_{a\gamma Z}$. Cases in which $c_B \approx c_W$ would instead lead to a suppressed $g_{a\gamma Z}$ coupling.

In the presence of ALP-photon couplings, the ALP can be produced in electron-positron collisions via two leading-order processes. Figure 2.1a shows the ALP-strahlung process, $e^+e^- \to \gamma a$, where the ALP is radiated off a photon. Alternatively, the ALP can be produced via the vector-boson fusion (VBF) process, $e^+e^- \to e^+e^-a$, as shown in Figure 2.1b, where the ALP is produced via the fusion of two virtual bosons emitted from the incoming electron and positron. In both cases, the ALP subsequently decays into two photons.

The decay $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ is the primary focus of this analysis, as it can provide a



(a) Integrated cross section (b) Differential cross section

Figure 2.2: Comparison of the integrated cross sections (left) and differential cross sections with respect to the longitudinal momentum p_z (right) for the ALP-strahlung (yellow) and VBF (blue) processes. Taken from [25].

clean experimental signature at Belle II (see Section 2.3). The differential cross section for the ALP-strahlung process can be expressed as [40]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta} = \frac{g_{a\gamma\gamma}^2 \alpha}{128} (3 + \cos 2\theta) \left(1 - \frac{m_a^2}{s}\right)^3, \tag{2.11}$$

where α is the fine-structure constant, s is the centre-of-mass energy squared, and θ is the photon angle in the Centre of Mass (CM) frame. The mild angular dependence gives the recoiling photon a sizeable transverse momentum, making the full final state detection promising. Contrary, in the case of the VBF process, the ALP is produced with a high transversal momentum, while the outgoing electron and positron are scattered at very low angles and typically escape detection along the beam pipe. Figure 2.2a shows a comparison of the total cross sections for both processes as a function of the ALP mass at the CM energy $\sqrt{s} = 10.58\,\mathrm{GeV}$ with a coupling of $g_{a\gamma\gamma} = 1 \times 10^{-4}\,\mathrm{GeV}^{-1}$. The cross section of the VBF process exceeds that of the ALP-strahlung process, specifically for lower ALP masses. Figure 2.2b compares the processes' differential cross section with respect to the longitudinal momentum p_z of the ALP. Due to the lower scattering angle of the outgoing leptons, the ALP produced in the VBF process is experimentally challenging to analyse. Requiring the full reconstruction of the final state results in very high signal efficiency losses. Reconstructing only the ALP candidate and missing energy would require precise knowledge of the detector performance and Bhabha scattering $e^+e^- \to e^+e^-$ at low angles.

2.3 Signature at Belle II

As previously mentioned, in this work I focus on the ALP-strahlung production process, $e^+e^- \to \gamma a$, $a \to \gamma\gamma$, where the ALP is produced in association with a photon and subsequently decays into two photons. In this section, I give a brief overview of the

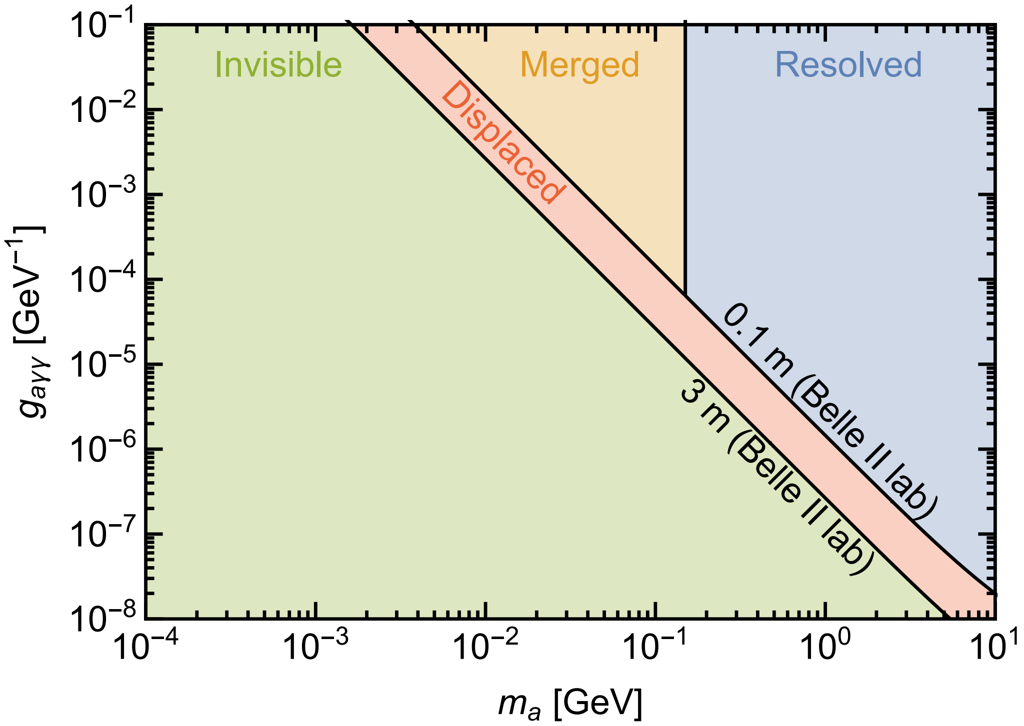


Figure 2.3: Illustration of the four different experimental signatures of an ALP at Belle II depending on its mass m_a and coupling to photons $g_{a\gamma\gamma}$. Taken from [25].

experimental signature of this process at Belle II and thereby present the analysis strategy. The detailed description of the Belle II detector follows in Chapter 3.

Through the interaction with photons in the previously discussed model, the ALP lifetime $\tau_a = \Gamma_a^{-1}$ can be expressed as

$$\tau_a = \frac{64\pi}{g_{a\gamma\gamma}^2 m_a^3}. \tag{2.12}$$

For a given set of ALP mass and coupling parameters, the ALP decay width Γ_a can be very small, thus making the ALP long-lived. In the context of the Belle II detector, m_a and $g_{a\gamma\gamma}$ determine four different experimental signatures. Figure 2.3 summarises the experimental signatures depending on the ALP mass and coupling to photons. Sharp transitions in this figure separate the different regimes, but in reality, these transitions are smooth and depend on the detector geometry and performance.

The signature depends mainly on whether the ALP decays promptly or is long-lived, and how boosted the ALP is at production. The four signatures are:

- **Invisible**
  Very small couplings in combination with low masses lead to long-lived ALPs that escape the detector before decaying. In this case, the experimental signature consists of a single photon from the ALP-strahlung process and missing energy from the

undetected ALP. The search for the invisible signature is equivalent to a search for ALP decaying to DM or a search for long-lived dark photons [16, 41].

- **Displaced**
  For slightly larger couplings, the ALP can decay within the detector volume but at a measurable distance from the interaction point. With sufficiently low mass, the ALP can have a high boost, leading to collimated decay photons that can be reconstructed as a single photon cluster. This signature would be identical to the *merged* signature discussed below. WIth hiher masses, the ALP has a low boost, leading to two well-separated displaced photon clusters. However, reconstructing displaced photons significantly reduces the photon resolution as the Belle II detector has a single Electromagnetic Calorimeter (ECL) layer and no other mechanisms to reconstruct the displaced vertex of the neutral ALP.

- **Merged**
  Mainly for very low ALP masses, the two photons from the ALP decay are highly collimated due to the high boost of the ALP. Depending on the beam background conditions (see Section 3.4) and the detector region, the reconstruction algorithms are not able to resolve the two photons separately. The resulting signature consists of two photon clusters, one from the initial state radiation and one merged cluster from the ALP decay. Another challenging aspect of this signature is that the ALP mass, at which this effect occurs more frequently, is close to the π^0 mass. The challenge being distinguishing contributions from ALP and π^0 as they are identical in their quantum numbers.

- **Resolved**
  For larger couplings and/or higher masses, the ALP decays promptly, and the two photons from the ALP decay are well-separated and can be reconstructed individually. This signature leads to a final state with three resolved photon clusters in the ECL, which can be fully reconstructed.

In this work, I reconstruct events with three well-resolved photons with little to no missing energy, and search for a peak in the invariant diphoton mass $M_{\gamma\gamma}$ which forms the ALP candidate. I perform the search in the region $M_{\gamma\gamma} \in [0.17, 9.80]\,\mathrm{GeV}/c^2$, thus mostly evading the challenges of the merged and invisible signatures. By requiring three resolved photons, the merged case presents mainly just as a reconstruction inefficiency for low ALP masses. I don't explicitly consider the displaced signature in this analysis, as the final sensitivity shows in Fig. 9.3, the integrated luminosity of Belle II is not yet enough to become sensitive to this regime.

2.4 Existing Limits

In the MeV/c^2–GeV/c^2 mass range and for photon couplings $g_{a\gamma\gamma} < 1 \times 10^{-5}\,\mathrm{GeV}^{-1}$, the strongest limits are provided by collider-based experiments. The different limits on the ALP-photon coupling can be broadly categorised into five groups depending on the production mechanism and experimental signature. Light ALP masses below or slightly above 100 MeV/c^2 are mainly constrained by searches for invisible final states in $e^+e^- \to \gamma$+invisible processes at colliders, fixed-target experiments, and forward facilities. At higher masses, roughly above 5 GeV/c^2, heavy-ion collisions provide the most stringent bounds through searches for the VBF signature. In the intermediate mass range, the best limits arise from ALP-strahlung production with prompt $a \to \gamma\gamma$ decays at electron-positron colliders and fixed-target experiments. In the following, I briefly summarise the current constraints in these regimes. Whenever I mention limits from reinterpretations, I refer to the results presented in [25].

In beam-dump experiments, ALPs can be produced via the Primakoff process, where a high-energy photon scatters off a nucleus and produces an ALP through its interaction with the nuclear electromagnetic field [42]. Reinterpretations of existing results from electron-beam experiments such as SLAC E141 [43], SLAC E137 [44], and PrimEx [45] provide limits at low ALP masses, especially around the π^0 mass. Reinterpretations from proton beam-dump experiments results from CHARM [46] and NuCal [47, 48] extend these limits up to 100 MeV/c^2 for very small photon couplings. A dedicated search for the ALP-photon coupling by the NA64 experiment [49] has, until recently, set the most stringent limits in this region, reaching couplings around $g_{a\gamma\gamma} \sim 10^{-4}\,\mathrm{GeV}^{-1}$.

These NA64 limits have since been mostly surpassed by a search from the FASER experiment [50], which extended the sensitivity to slightly higher masses and lower couplings.

For the high-mass regime above 5 GeV/c^2, heavy-ion collisions at the LHC provide the leading constraints, particularly from ATLAS [51] and CMS [52].

Reinterpretations of invisible final-state searches at electron-positron colliders include the mono-photon search at LEP [53, 54], the radiative Υ decay $\Upsilon(3S) \to \gamma$+invisible at *BABAR* [55], and the invisible dark-photon search at *BABAR* [56]. Dedicated searches for the $a \to \gamma\gamma$ signature at electron-positron colliders have been performed by BESIII [57, 58] and Belle II [59]. The BESIII analyses studied radiative J/ψ decays, $J/\psi \to \gamma a$, $a \to \gamma\gamma$, at CM energies of $\psi(3686)$ and J/ψ, while the Belle II search - the predecessor of this work - probed the ALP-strahlung production mechanism. However, that analysis was performed with a much smaller dataset of only 445 pb^{-1}.

Figure 2.4 summarises the existing UL for the ALP mass and photon coupling $g_{a\gamma\gamma}$ phase space from the previously discussed searches.

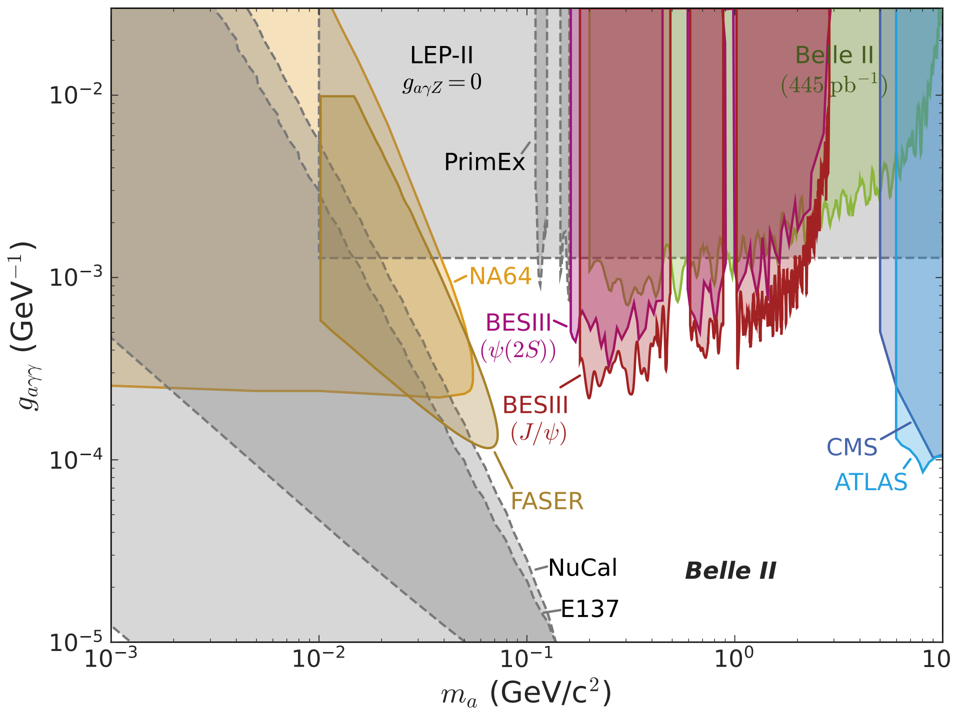


Figure 2.4: Existing UL for the ALP mass and photon coupling $g_{a\gamma\gamma}$ phase space. Grey shaded areas indicate excluded regions from reinterpretations. Colored shaded areas indicate excluded regions from dedicated searches. The references for each limit are provided in the text.

Chapter 3

The Experimental Setup

In this chapter, I describe the experimental setup for this thesis. First, Sections 3.1 and 3.2 provide a brief overview of the SuperKEKB accelerator [60] and the Belle II detector [14]. Finally, Section 3.4 describes the simulation of the datasets used in this analysis to obtain and optimise the selection procedure, as well as the recorded data which is used to perform the final measurement.

3.1 The SuperKEKB Accelerator

The Belle II experiment's main purpose is to record events from electron-positron collisions with a CM energy at the $\Upsilon(4S)$ resonance. These initial particle beams for the interactions are provided by the SuperKEKB accelerator located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. Figure 3.1 shows a schematic view of the accelerator with all its different components. Unlike hadron colliders such as the LHC, SuperKEKB collides elementary particles, providing precise knowledge of the initial collision, which is a key advantage for analyses.

A photo-cathode radio frequency gun is used to produce the low-emittance, high-current electron beam. For the positron beam, a thermionic gun produces high-current electrons, which generate positrons from a tungsten target; the positrons are then collected and focused by a flux concentrator. The electron and positron beams are then accelerated in a linear accelerator up to their desired energy of 7 GeV for electrons and 4 GeV for positrons. During the acceleration process, the positron beam is passed through a damping ring where the emittance is reduced before being sent further along the accelerator. Electrons are stored in the High-Energy Ring (HER) and positrons in the Low-Energy Ring (LER). Due to the linear accelerator design, continuous top-up injection is possible. The asymmetric beam energies result in a CM energy of $\sqrt{s} \approx 10.58\,\text{GeV}$, corresponding to the $\Upsilon(4S)$ resonance, which decays in about 96 % of the cases into $B\overline{B}$ pairs [1]. Nevertheless, the accelerator also runs on other beam energies to allow for studies of other Υ resonances or even off-resonance. For this work, off-resonance runs below the $\Upsilon(4S)$ resonance are

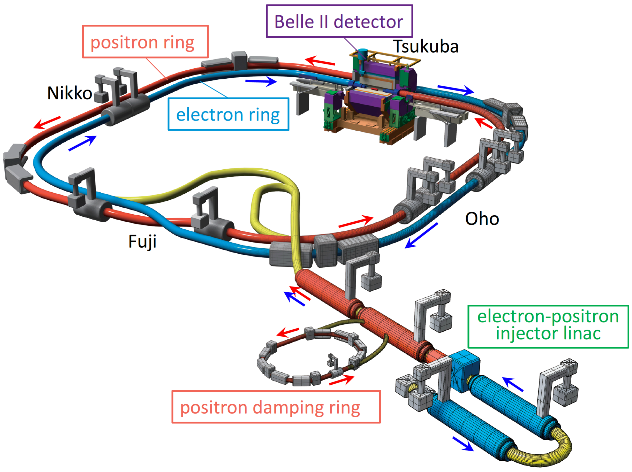


Figure 3.1: SuperKEKB accelerator structure schematic overview. Taken from [60]. SuperKEKB accelerator structure schematic overview. Taken from [60].

included. The energy asymmetry induces a Lorentz boost of $\beta\gamma = 0.28$ in the direction of the electron beam, which is crucial for time-dependent *CP* violation studies.

The current Belle II program aims to collect a total integrated luminosity of $\mathcal{L} = 50\,\mathrm{ab}^{-1}$ by the beginning of the next decade [61]. To achieve this goal, currently, an instantaneous luminosity of $L = 6.5 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [62] is foreseen, which is about a factor 30 larger than the instantaneous luminosity reached by its predecessor KEKB. The, at SuperKEKB implemented, nano-beam scheme* enables these immensely high instantaneous luminosities by colliding sufficiently thin bunches at a large horizontal crossing angle. For the dataset used in this work, the target instantaneous luminosity was not reached; however, the world record instantaneous luminosity of $L = 4.71 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ was achieved in June 2022 [64].

3.2 The Belle II Detector

Figure 3.1 shows the Belle II detector located in the Tsukuba experimental hall. Like its predecessor, the Belle detector, Belle II is a general-purpose, 4π, high-energy spectrometer specifically built to record events from the asymmetric e^+e^- collision. To recover as many particle signatures as possible with a solid angle coverage, the Belle II detector is built by a series of layers of particle detectors arranged asymmetrically around the Interaction Point (IP). Each subdetector is purpose-built to enable the reconstruction and identification of distinct particle types. Figure 3.2 depicts a schematic overview of the Belle II detector.

The following sections summarise the main subdetectors of the Belle II experiment, with emphasis on components particularly relevant to this analysis. A comprehensive description of the detector is given in [14].

*Initially proposed in the SuperB conceptual design report [63].

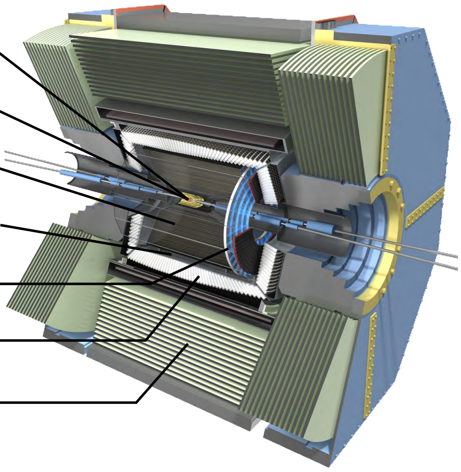


Figure 3.2: Belle II detector schematic overview with annotated subdetectors. Taken from [65].

For this and all the following chapters, the Belle II coordinate system is defined as follows:

- The origin is at the nominal IP.
- The z-axis points in the direction of the electron beam.
- The x-axis is defined horizontally, pointing outward from the centre of the storage ring.
- The y-axis points vertically to the top.
- The polar angle θ is defined in the range from 0 to π, where $\theta = 0$ describes the forward direction.
- The azimuthal angle ϕ is defined from $-\pi$ to π in the transverse x-y plane and $y = 0$ being $\phi = 0$.

3.2.1 Tracking Detectors

For this work, the tracking system's importance lies in the ability to reject charged particles, as a fully neutral final state is expected.

The Belle II tracking system comprises three subdetectors designed to measure the trajectories and vertex positions of charged particles. From these trajectories, the particle momenta can be determined using the 1.5 T magnetic field generated by the superconducting solenoid that encloses most of the Belle II detector. Two vertex detectors, the Pixel Detector (PXD) [66, 67] and the Silicon Vertex Detector (SVD) [68] cover the innermost region close to the IP. Due to the high detector material granularity, these detectors provide precise spatial information about the particle vertex, which is crucial for measurements of short-lived particles targeting *e.g.* CP violation. The Central Drift Chamber (CDC) [69, 70] extends the tracking system to larger radii and constitutes most of the inner detector volume. While it offers lower spatial resolution compared to the vertex detectors, the CDC provides a large active volume for track reconstruction and enables measurement of the particle energy loss for particle identification.

The core principle of the three tracking detectors is the same, however utilised with different technologies. Charged particles traverse material and produce a current measured by the detector components, interpreted as a detector hit. Reconstruction algorithms subsequently combine these hits to form particle trajectories, commonly referred to as tracks.

Vertex Detectors

The PXD is the innermost vertex detector, located directly around the beam pipe. It consists of two pixelated silicon Depleted Field Effect Transistor (DEPFET) layers. With radii of 1.4 cm and 2.2 cm, it provides high spatial resolution measurements close to the IP. For the data-taking period during which the data analysed in this thesis were recorded, only one-sixth of the second PXD layer was installed.

Surrounding the PXD, the SVD provides additional spatial measurements at larger radii. In contrast to the pixel detector, the SVD uses silicon strip sensors to measure particle positions. These silicon strips are arranged cylindrically around the beam pipe at radii of 3.9 cm, 8.0 cm, 10.4 cm, and 13.5 cm. With a slight slant of the last three layers in the forward direction, the SVD angular coverage reaches a total range of 17 to 150°.

Central Drift Chamber

The CDC is the main tracking detector used to measure particle momentum and provide particle identification information. It is a gas-filled drift chamber with a very low material budget to reduce the impact of multiple scattering on traversing particles. Inside the chamber, 56 layers of sense and field wires are arranged cylindrically around the beam axis. To provide three-dimensional position measurements, the wires are arranged in axial and stereo layers that are tilted with respect to the beam axis. Similar to the SVD, the CDC covers an angular range of 17 to 150°. Charged particles traversing the gas mixture of 50 % helium and 50 % methane ionise the gas, producing an avalanche of electrons that is collected at the wires.

3.2.2 Particle Identification Detectors

In contrast to current hadron collider experiments, Belle II includes two dedicated particle identification detectors: the Time-Of-Propagation counter (TOP) and the Aerogel Ring-Imaging Cherenkov detector (ARICH), [71]. Although all Belle II subdetectors except the

PXD provide some level of particle identification (PID) information, the two dedicated PID detectors significantly enhance the overall performance, particularly for separating charged pions and kaons. Similar to the tracking detectors, the PID detectors can only measure charged particles. For this work, particle identification does not play a major role, as the flavour of neutral particles cannot be directly detected in these detectors.

The core principle of both PID detectors is based on the Cherenkov effect. Charged particles traversing a medium with a velocity greater than the speed of light in that medium emit Cherenkov photons, which are subsequently detected.

Time-Of-Propagation Counter

The TOP is located just outside the CDC wall, directly in front of the ECL in the barrel region. This detector consists of 16 quartz bars arranged parallel to the z-axis around the IP. The Cherenkov photons emitted in the quartz bars are reflected internally within the bars. At one end of each bar, a mirror reflects the photons, while at the opposite end, photomultiplier tubes collect them. From the position and timing of the detected photons, the characteristic Cherenkov angle can be reconstructed, which then contributes to the particle identification.

Aerogel Ring-Imaging Cherenkov Detector

The ARICH is located in the forward endcap region of the Belle II detector, just in front of the ECL. In contrast to the TOP detector, the ARICH measures the Cherenkov angle directly from the cone of emitted Cherenkov photons. To observe this cone, charged particles first traverse a radiator material where they emit Cherenkov photons. These photons then propagate through a proximity gap before reaching hybrid avalanche photon detectors arranged in a circular pattern. To improve the resolution of the Cherenkov rings formed by the photons, the ARICH employs a two-layer aerogel radiator with different refractive indices.

3.2.3 Electromagnetic Calorimeter

The ECL [72] is the most important subdetector for this work, as it is used to reconstruct and measure the energy of the photons in the final state. In general, the ECL is designed to measure the energy and position of incoming particles, most importantly electrons, photons, and neutral hadrons. In addition to the energy measurement, the calorimeter also provides PID information to distinguish electrons from muons and charged hadrons based on their ionisation energy loss.

The ECL is located just outside the TOP, ARICH, and CDC detectors, inside the superconducting solenoid. This subdetector consists of 8736 thallium-doped caesium iodide crystals, each with a length of 30 cm, corresponding to 16.1 radiation lengths. The crystals are arranged in a barrel and two endcaps in the forward and backward directions, covering an angular acceptance of $12.4^\circ < \theta < 155.1^\circ$, with two small gaps between $31.4^\circ < \theta < 32.2^\circ$ and $128.7^\circ < \theta < 130.7^\circ$. To enhance the detection efficiency, the crystals have irregular shapes, resulting in a projective geometry pointing towards a region near to the IP.

As the name suggests, the ECL is optimised to measure electromagnetic showers. Photons and electrons entering the crystals initiate electromagnetic cascades through pair conversion, Compton scattering, or the photoelectric effect for photons, and bremsstrahlung for electrons. The stochastic branching of particles in the shower produces a cascade of lower-energy electrons, positrons, and photons, which deposit their energy by ionising the crystal material. When the ionised atoms de-excite, scintillation light is emitted and collected by two photodiodes attached to the rear side of each crystal via an 1 mm acrylite plate. To maximise the light collection efficiency, each crystal is wrapped in reflective Teflon foil and covered with two thin layers of aluminium and Mylar. The photons detected by the photodiodes are analogously summed and processed by the so-called Shaper Digital Signal Processor and Collector modules. During the processing, the shape, amplitude and time of the electronic signal are determined. Depending on the task, the energy depositions in the crystals are clustered to infer the incoming particle properties further. In the case of the later explained trigger system, the crystal energies are first summed in a 4×4 grid, so-called ECL trigger cells, and then clustered for fast online processing. For the offline reconstruction, the individual crystals are clustered with region-of-interest algorithms to form so-called ECL clusters. From these clusters, particle candidates are built, and their shower region is determined. These shower regions are particularly insightful in determining the shower shape to separate single-photon showers from showers produced by multiple particles. In this work, unless stated otherwise, when I refer to the energy or spatial position of photon candidates, I mean the corresponding ECL cluster properties.

As a homogeneous calorimeter, and owing to the high light yield and short radiation length of the CsI(Tl) crystals, the ECL provides excellent energy resolution for photons. For photon energies above 1 GeV, the calorimeter achieves a relative energy resolution of about 1.25 to 2.25 %, depending on the photon energy.

3.2.4 $\mathrm{K_L^0}$ and Muon Detector

The $\mathrm{K_L^0}$ and muon detector (KLM) is the outermost sub-detector of Belle II. It is located outside the superconducting solenoid and interleaved with the iron plates that form the magnetic flux return yoke. This subdetector is crucial for identifying minimal ionising particles such as muons and long-lived neutral kaons ($\mathrm{K_L^0}$) that traverse the detector. The KLM consists of alternating layers of iron absorber plates and active detection material. The active material consists of either resistive plate chambers or scintillator strips coupled to silicon photomultipliers. For this work, the KLM is of least importance as the photons in the final state need to be detected primarily by the ECL for a precise energy measurement.

3.2.5 Trigger System

With the target instantaneous luminosity of $L = 6 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, the collision rate of the most frequent process — Bhabha scattering $e^+e^- \to e^+e^-(\gamma)$ — exceeds 40 kHz. However, due to bandwidth limitations, the maximum rate at which events can be recorded is restricted to 10 kHz. To reduce the event rate to this manageable level, a two-level trigger system is employed at Belle II. First, a hardware-based trigger, referred to as the Level 1 trigger (L1 trigger), decreases the rate to about 30 kHz. Subsequently, a software-based High Level Trigger (HLT) further reduces the rate to the storable maximum of 10 kHz. Both trigger systems are designed to select events of potential physics interest while rejecting the most frequent background processes, such as the aforementioned Bhabha scattering and $e^+e^- \to \gamma\gamma(\gamma)$.

The latter process, however, is of particular importance for this work, as events from very light or very heavy ALPs events produce similar topologies — back-to-back high-energy clusters in the ECL. At Belle II, these events are retained due to the introduction of a dedicated trigger selection designed to reject $e^+e^- \to e^+e^-(\gamma)$ while keeping $e^+e^- \to \gamma\gamma(\gamma)$ events: the `ggsel` trigger line. This dedicated trigger line is essential for maintaining high signal efficiency for both high- and low-mass ALP events, particularly after the full integration of the Bhabha veto during data taking.

In this work, I present the first physics analysis that utilises this dedicated trigger line. A detailed description of the `ggsel` trigger line and its availability during the data-taking period is provided in Section 5.1.

Level 1 Trigger

The L1 trigger is the hardware trigger making the initial decision on whether an event is of interest based on level detector information. Built from field-programmable gate arrays (FPGAs), the L1 trigger combines signals from the CDC, ECL and KLM that enter the global decision logic (GDL). Input signals pass through the logic implemented on the FPGAs, resulting in multiple bit-wise decisions per event. From these bit-wise decisions, commonly referred to as trigger bits, trigger lines are constructed that are eventually evaluated by the GDL to take the final decision for an event to pass or be rejected.

High Level Trigger

As previously mentioned, the HLT is a software-based trigger system that further reduces the event rate to the storable maximum. This trigger system is operating on CPUs and runs the Belle II reconstruction software. In contrast to the L1 trigger, the HLT takes information from all subdetectors besides the PXD into account. Similar to the L1 trigger, several trigger lines are formed; however, these decisions are based on fully reconstructed objects and event properties.

3.3 Datasets

As it is typical for so-called "blind analysis", I first optimise the selection procedure and test the signal extraction method on simulated samples. Once all the steps in the analysis are set, only then can I use recorded data for the evaluation of the final results. This analysis technique reduces the so-called experimenter bias. In this section, I describe the event samples that I use in this analysis, including the signal and background simulations, as well as the data collected from the Belle II experiment. I elaborate on the details of the generation and simulation of the signal Monte Carlo (MC) samples for different ALP mass hypotheses and run conditions. I then describe the centrally produced background MC samples used to estimate the background contributions. Finally, I outline the recorded data for the final result evaluation.

The simulation of the processes and the detector response, as well as the reconstruction, is performed using the Belle II Analysis Software Framework (basf2) [73, 74].

3.3.1 Signal Samples

I generate the signal MC samples for each ALP mass hypothesis and different run conditions. For this, I simulated with `MadGraph5 aMC@NLO` [75] the events which I use to optimise the selection as well as to estimate the signal efficiency and shape. During the simulation, I set the beam energy to match the targeted run, leaving out the beam asymmetry since basf2 applies the boost before simulation. The ALP width is calculated automatically by `MadGraph5 aMC@NLO`, as I do not observe numerical instabilities. I include Initial State Radiation (ISR) effects in the generation with the `MadGraph5 aMC@NLO` ISR option [76], which accounts for momentum changes in the final state, although no additional photons are generated.

Subsequently, I pass the generated events to basf2 for the detector simulation with Geant4 [77] and reconstruction. To account for different run conditions and remain consistent with the centrally produced background, I apply the run- and experiment-based settings of the central simulation campaign and overlay the so-called beam background, which is provided by the collaboration and taken from recorded data. The different sources of beam background and their impact on the detector and reconstruction are discussed in detail in [41, 78]. In short summary for this work, the beam background increases with increasing instantaneous luminosity and leads to additional energy depositions in the ECL, which increases the number of photon candidates and can reduce the energy resolution of true photon candidates. Finally, the events are boosted to account for the asymmetric beam energies, and the primary vertex is smeared.

I choose seven runs to monitor the effect of different beam background conditions. Six of these are $\Upsilon(4S)$ resonance runs, and one is a pure $\Upsilon(4S)$ off-resonance run. These runs represent different time periods with comparable luminosities but varying beam backgrounds.

Table 3.1: Summary of different experiment and run combinations used to produce the signal MC samples.

Experiment	**Run**	$\sqrt{s}$	$N_{\mathrm{OOT}}^{\mathrm{Crys.}}$
22	30	$\Upsilon(4S)$	51.8
22	468	$\Upsilon(4S)$	102.2
24	985	$\Upsilon(4S)$	48.3
24	2176	$\Upsilon(4S)$	115.5
25	343	$\Upsilon(4S)$ off-resonance	123.3
26	898	$\Upsilon(4S)$	175.4
26	1485	$\Upsilon(4S)$	294.6

To quantify these differences, I use the mean number of out-of-time ECL crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$*. Table 3.4 summarises the chosen runs and experiments along with the corresponding beam background measurements.

3.3.2 Background Samples

I use the background MC samples to train the selection multivariate analysis (MVA), to check the overall MC/data agreement and to estimate the background shape. For this work, I use centrally produced run-dependent MC samples for the non-peaking background components and specifically requested samples for the peaking background components.

Table 3.5 summarises the non-peaking background components, including all available processes. Among these, radiative diphoton events $e^+e^- \to \gamma\gamma(\gamma)$ contribute the most to the background due to their large cross section. Other significant contributions come from the $e^+e^- \to q\bar{q}$ process †, specifically from light quarks such as $e^+e^- \to u\bar{u}$ through the process $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$, and from the radiative Bhabha process $e^+e^- \to e^+e^-(\gamma)$, which requires two tracks to be missed during reconstruction. A variety of generators are used for the production of these MC samples, including BabaYaga@NLO [79], KKMC [80], Pythia8 [81], AAFH [82], PHOKHARA [83], TAUOLA [84], KoralW [85], and EvtGen [86]. During the analysis of these events, I weight them according to the number of streams, which corresponds to a normalisation to the integrated luminosity. The number of streams corresponds to how many more or fewer events were produced for a given process, *e.g.* a stream of two means that twice as many events are produced for the given luminosity.

The peaking background processes consist of SM processes with the same final state $e^+e^- \to h^0\gamma \to \gamma\gamma\gamma$, corresponding to the pseudo-scalar mesons $h^0 \in [\pi^0, \eta, \eta']$ and their radiative equivalents. Since the form factors of the mesons favour small momentum transfer,

*The exact definition of $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ is in Section 5.2.

†During the studies of the control regions, which I present in Chapter 7, I found that the simulation of the $e^+e^- \to q\bar{q}$ process leads to unphysical final states, so these events are removed from consideration.

Table 3.2: Overview of the centrally produced background processes.

Process	**Generator**	**Streams**
$e^+e^- \to \gamma\gamma(\gamma)$	BabaYaga@NLO	2
$e^+e^- \to q\overline{q}$ $(q \in [u,d,s,c])$	KKMC + Pythia8	4
$e^+e^- \to e^+e^-(\gamma)$	BabaYaga@NLO	0.1
$e^+e^- \to e^+e^-e^+e^-$	AAFH	1
$e^+e^- \to e^+e^-\mu^+\mu^-$	AAFH	1
$e^+e^- \to hh\gamma$		1
$e^+e^- \to \pi^+\pi^-\gamma$	PHOKHARA	
$e^+e^- \to \pi^+\pi^-\pi^0\gamma$	PHOKHARA	
$e^+e^- \to K^+K^-\gamma$	PHOKHARA	
$e^+e^- \to K^0_\mathrm{S}K^0_\mathrm{L}\gamma$	PHOKHARA	
$e^+e^- \to \ell^+\ell^-XX$		1
$e^+e^- \to e^+e^-\tau^+\tau^-$	AAFH + TAUOLA	
$e^+e^- \to e^+e^-\pi^+\pi^-$	TREPS	
$e^+e^- \to e^+e^-K^+K^-$	TREPS	
$e^+e^- \to e^+e^-p\overline{p}$	TREPS	
$e^+e^- \to \mu^+\mu^-\mu^+\mu^-$	AAFH	
$e^+e^- \to \mu^+\mu^-\tau^+\tau^-$	AAFH + TAUOLA	
$e^+e^- \to \tau^+\tau^-\tau^+\tau^-$	KoralW + TAUOLA	
$e^+e^- \to \mu^+\mu^-(\gamma)$	KKMC	1
$e^+e^- \to \tau^+\tau^-(\gamma)$	KKMC + TAUOLA	4
$e^+e^- \to B\overline{B}$	EvtGen + Pythia8	4

$e^+e^- \to h^0\gamma(\gamma)$ has a significantly higher cross section than $e^+e^- \to h^0\gamma$. This effect does not apply to pointlike ALPs. Consequently, missing a low-energy photon or reconstructing the two hadronic photons as one, due to the potentially large boost and resulting small photon opening angle, produces the same final state as the signal decay. Specifically, in the case of merged photons, this leads to a broadly peaking contribution of $e^+e^- \to \omega\gamma \to P\gamma\gamma$ around the ω mass window if the neutral pseudoscalar particle is reconstructed as a single photon candidate. The PHOKHARA generator is used to simulate a total of six different decays. Table 3.6 summarises the settings applied in the `PhokharaInput` module of basf2 and provides an overview of the calculated cross sections.

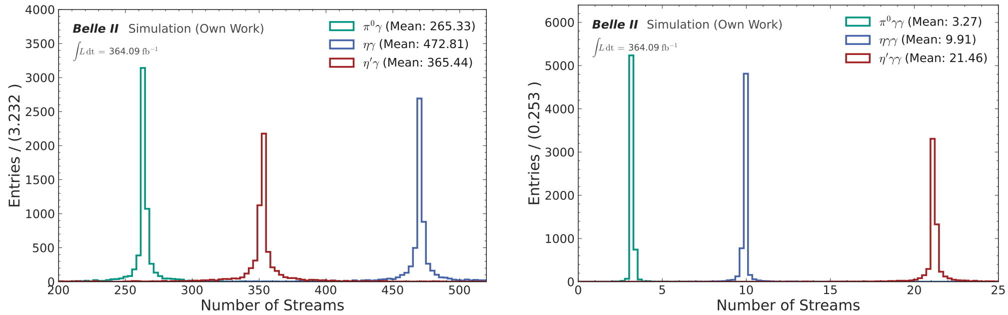
Additional generation settings that are applied equally to all events are that the minimal photon energy is at least 1 % of the CM energy and that the final state particle scattering angle is between 10 to 160°.

For the central production of the peaking background components and their radiative partners, I request more events than would be produced based on the calculated cross sections. However, during central production, various technical factors reduce the number

Table 3.3: Summary of the additionally produced peaking background components and their radiative processes. `FinalState` and `LO` correspond to the PHOKHARA specific settings used to produce these processes. The cross section is calculated by PHOKHARA. N corresponds to the number of requested events and $\hat{N}_{\text{Stream}}$ to the mean number of resulting streams from Fig. 3.4.

Process	`FinalState`	`LO`	**cross section** [nb]	N	$\hat{N}_{\text{Stream}}$
$e^+e^- \rightarrow \pi^0\gamma$	13	-1	$(7.260 \pm 0.015) \times 10^{-6}$	2.4×10^6	265.33
$e^+e^- \rightarrow \pi^0\gamma\gamma$	13	0	$(1.682 \pm 0.002) \times 10^{-3}$	7.0×10^6	3.27
$e^+e^- \rightarrow \eta\gamma$	14	-1	$(4.091 \pm 0.008) \times 10^{-6}$	2.4×10^6	472.81
$e^+e^- \rightarrow \eta\gamma\gamma$	14	0	$(5.567 \pm 0.018) \times 10^{-4}$	7.0×10^6	9.91
$e^+e^- \rightarrow \eta'\gamma$	15	-1	$(5.441 \pm 0.011) \times 10^{-6}$	2.4×10^6	365.44
$e^+e^- \rightarrow \eta'\gamma\gamma$	15	0	$(2.602 \pm 0.008) \times 10^{-4}$	7.0×10^6	21.46

of events that can actually be generated. These factors include the availability of beam background overlays or extremely noisy events that could make the simulation unstable. Since the reduction of events is run-dependent but small compared to the requested number, I weight each event according to the number of generated events per run. Specifically, for each run, I calculate the expected number of events based on the run's luminosity and the PHOKHARA cross section, and then compare this to the number of events actually generated after the reduction. The ratio of the expected to the generated number of events defines the so-called number of streams (equivalent to Table 3.5), which is used to weight each event. Figure 3.4 shows the distributions of the calculated number of streams per run per process. Since the distributions peak sharply with little spread, I use the mean number of streams for all centrally generated events.



(a) Processes: $e^+e^- \to h^0\gamma$

(b) Processes: $e^+e^- \to h^0\gamma\gamma$

Figure 3.3: Background process and run-dependent distributions of the number of streams. One entry corresponds to one run. The legend indicates the mean of distribution, which is eventually used to weight each process's event.

3.3.3 Recorded Dataset

For this analysis, I use the full Run 1 dataset with runs at the CM energies of $\Upsilon(4S)$ and $\Upsilon(4S)$ off resonance. These runs were recorded in the time period of 2019-2022. The analysis targets a combined integrated luminosity of $(408.110 \pm 1.712)\,\mathrm{fb}^{-1}$ [87].

3.4 Datasets

As it is typical for so-called "blind analysis", I first optimise the selection procedure and test the signal extraction method on simulated samples. Once all the steps in the analysis are set, only then can I use recorded data for the evaluation of the final results. This analysis technique reduces the so-called experimenter bias. In this section, I describe the event samples that I use in this analysis, including the signal and background simulations, as well as the data collected from the Belle II experiment. I elaborate on the details of the generation and simulation of the signal MC samples for different ALP mass hypotheses and run conditions. I then describe the centrally produced background MC samples used to estimate the background contributions. Finally, I outline the recorded data for the final result evaluation.

The simulation of the processes and the detector response, as well as the reconstruction, is performed using the Belle II Analysis Software Framework (basf2) [73, 74].

3.4.1 Signal Samples

I generate the signal MC samples for each ALP mass hypothesis and different run conditions. For this, I simulated with `MadGraph5 aMC@NLO` [75] the events which I use to optimise the selection as well as to estimate the signal efficiency and shape. During the simulation, I set the beam energy to match the targeted run, leaving out the beam asymmetry since basf2 applies the boost before simulation. The ALP width is calculated automatically by

Table 3.4: Summary of different experiment and run combinations used to produce the signal MC samples.

Experiment	**Run**	$\sqrt{s}$	$N_{\mathrm{OOT}}^{\mathrm{Crys.}}$
22	30	$\Upsilon(4S)$	51.8
22	468	$\Upsilon(4S)$	102.2
24	985	$\Upsilon(4S)$	48.3
24	2176	$\Upsilon(4S)$	115.5
25	343	$\Upsilon(4S)$ off-resonance	123.3
26	898	$\Upsilon(4S)$	175.4
26	1485	$\Upsilon(4S)$	294.6

`MadGraph5 aMC@NLO`, as I do not observe numerical instabilities. I include ISR effects in the generation with the `MadGraph5 aMC@NLO` ISR option [76], which accounts for momentum changes in the final state, although no additional photons are generated.

Subsequently, I pass the generated events to basf2 for the detector simulation with Geant4 [77] and reconstruction. To account for different run conditions and remain consistent with the centrally produced background, I apply the run- and experiment-based settings of the central simulation campaign and overlay the so-called beam background, which is provided by the collaboration and taken from recorded data. The different sources of beam background and their impact on the detector and reconstruction are discussed in detail in [41, 78]. In short summary for this work, the beam background increases with increasing instantaneous luminosity and leads to additional energy depositions in the ECL, which increases the number of photon candidates and can reduce the energy resolution of true photon candidates. Finally, the events are boosted to account for the asymmetric beam energies, and the primary vertex is smeared.

I choose seven runs to monitor the effect of different beam background conditions. Six of these are $\Upsilon(4S)$ resonance runs, and one is a pure $\Upsilon(4S)$ off-resonance run. These runs represent different time periods with comparable luminosities but varying beam backgrounds. To quantify these differences, I use the mean number of out-of-time ECL crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$*. Table 3.4 summarises the chosen runs and experiments along with the corresponding beam background measurements.

3.4.2 Background Samples

I use the background MC samples to train the selection MVA, to check the overall MC/data agreement and to estimate the background shape. For this work, I use centrally produced run-dependent MC samples for the non-peaking background components and specifically requested samples for the peaking background components.

Table 3.5 summarises the non-peaking background components, including all available

*The exact definition of $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ is in Section 5.2.

Table 3.5: Overview of the centrally produced background processes.

Process	Generator	Streams
$e^+e^- \to \gamma\gamma(\gamma)$	BabaYaga@NLO	2
$e^+e^- \to q\overline{q}$ $(q \in [u,d,s,c])$	KKMC + Pythia8	4
$e^+e^- \to e^+e^-(\gamma)$	BabaYaga@NLO	0.1
$e^+e^- \to e^+e^-e^+e^-$	AAFH	1
$e^+e^- \to e^+e^-\mu^+\mu^-$	AAFH	1
$e^+e^- \to hh\gamma$		1
$e^+e^- \to \pi^+\pi^-\gamma$	PHOKHARA	
$e^+e^- \to \pi^+\pi^-\pi^0\gamma$	PHOKHARA	
$e^+e^- \to K^+K^-\gamma$	PHOKHARA	
$e^+e^- \to K^0_{\mathrm{S}}K^0_{\mathrm{L}}\gamma$	PHOKHARA	
$e^+e^- \to \ell^+\ell^-XX$		1
$e^+e^- \to e^+e^-\tau^+\tau^-$	AAFH + TAUOLA	
$e^+e^- \to e^+e^-\pi^+\pi^-$	TREPS	
$e^+e^- \to e^+e^-K^+K^-$	TREPS	
$e^+e^- \to e^+e^-p\overline{p}$	TREPS	
$e^+e^- \to \mu^+\mu^-\mu^+\mu^-$	AAFH	
$e^+e^- \to \mu^+\mu^-\tau^+\tau^-$	AAFH + TAUOLA	
$e^+e^- \to \tau^+\tau^-\tau^+\tau^-$	KoralW + TAUOLA	
$e^+e^- \to \mu^+\mu^-(\gamma)$	KKMC	1
$e^+e^- \to \tau^+\tau^-(\gamma)$	KKMC + TAUOLA	4
$e^+e^- \to B\overline{B}$	EvtGen + Pythia8	4

processes. Among these, radiative diphoton events $e^+e^- \to \gamma\gamma(\gamma)$ contribute the most to the background due to their large cross section. Other significant contributions come from the $e^+e^- \to q\overline{q}$ process *, specifically from light quarks such as $e^+e^- \to u\overline{u}$ through the process $e^+e^- \to q\overline{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$, and from the radiative Bhabha process $e^+e^- \to e^+e^-(\gamma)$, which requires two tracks to be missed during reconstruction. A variety of generators are used for the production of these MC samples, including BabaYaga@NLO [79], KKMC [80], Pythia8 [81], AAFH [82], PHOKHARA [83], TAUOLA [84], KoralW [85], and EvtGen [86]. During the analysis of these events, I weight them according to the number of streams, which corresponds to a normalisation to the integrated luminosity. The number of streams corresponds to how many more or fewer events were produced for a given process, *e.g.* a stream of two means that twice as many events are produced for the given luminosity.

The peaking background processes consist of SM processes with the same final state $e^+e^- \to h^0\gamma \to \gamma\gamma\gamma$, corresponding to the pseudo-scalar mesons $h^0 \in [\pi^0, \eta, \eta']$ and their

*During the studies of the control regions, which I present in Chapter 7, I found that the simulation of the $e^+e^- \to q\overline{q}$ process leads to unphysical final states, so these events are removed from consideration.

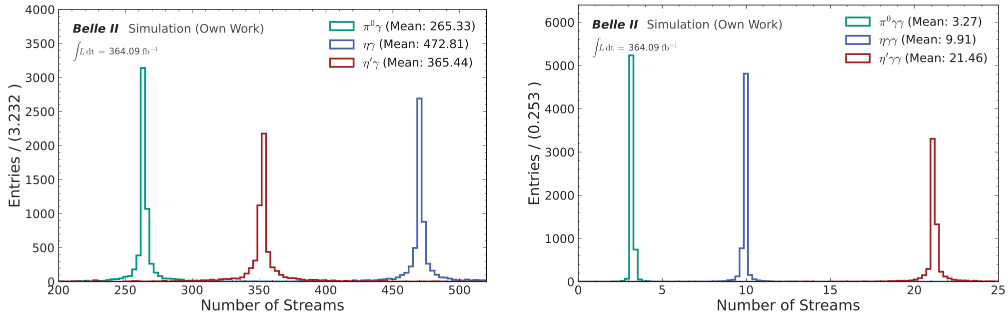
Table 3.6: Summary of the additionally produced peaking background components and their radiative processes. `FinalState` and `LO` correspond to the PHOKHARA specific settings used to produce these processes. The cross section is calculated by PHOKHARA. N corresponds to the number of requested events and $\hat{N}_{\text{Stream}}$ to the mean number of resulting streams from Fig. 3.4.

Process	`FinalState`	`LO`	**cross section** [nb]	N	$\hat{N}_{\text{Stream}}$
$e^+e^- \to \pi^0\gamma$	13	-1	$(7.260 \pm 0.015) \times 10^{-6}$	2.4×10^6	265.33
$e^+e^- \to \pi^0\gamma\gamma$	13	0	$(1.682 \pm 0.002) \times 10^{-3}$	7.0×10^6	3.27
$e^+e^- \to \eta\gamma$	14	-1	$(4.091 \pm 0.008) \times 10^{-6}$	2.4×10^6	472.81
$e^+e^- \to \eta\gamma\gamma$	14	0	$(5.567 \pm 0.018) \times 10^{-4}$	7.0×10^6	9.91
$e^+e^- \to \eta'\gamma$	15	-1	$(5.441 \pm 0.011) \times 10^{-6}$	2.4×10^6	365.44
$e^+e^- \to \eta'\gamma\gamma$	15	0	$(2.602 \pm 0.008) \times 10^{-4}$	7.0×10^6	21.46

radiative equivalents. Since the form factors of the mesons favour small momentum transfer, $e^+e^- \to h^0\gamma(\gamma)$ has a significantly higher cross section than $e^+e^- \to h^0\gamma$. This effect does not apply to pointlike ALPs. Consequently, missing a low-energy photon or reconstructing the two hadronic photons as one, due to the potentially large boost and resulting small photon opening angle, produces the same final state as the signal decay. Specifically, in the case of merged photons, this leads to a broadly peaking contribution of $e^+e^- \to \omega\gamma \to P\gamma\gamma$ around the ω mass window if the neutral pseudoscalar particle is reconstructed as a single photon candidate. The PHOKHARA generator is used to simulate a total of six different decays. Table 3.6 summarises the settings applied in the `PhokharaInput` module of basf2 and provides an overview of the calculated cross sections.

Additional generation settings that are applied equally to all events are that the minimal photon energy is at least 1 % of the CM energy and that the final state particle scattering angle is between 10 to 160°.

For the central production of the peaking background components and their radiative partners, I request more events than would be produced based on the calculated cross sections. However, during central production, various technical factors reduce the number of events that can actually be generated. These factors include the availability of beam background overlays or extremely noisy events that could make the simulation unstable. Since the reduction of events is run-dependent but small compared to the requested number, I weight each event according to the number of generated events per run. Specifically, for each run, I calculate the expected number of events based on the run's luminosity and the PHOKHARA cross section, and then compare this to the number of events actually



(a) Processes: $e^+e^- \to h^0\gamma$ (b) Processes: $e^+e^- \to h^0\gamma\gamma$

Figure 3.4: Background process and run-dependent distributions of the number of streams. One entry corresponds to one run. The legend indicates the mean of distribution, which is eventually used to weight each process's event.

generated after the reduction. The ratio of the expected to the generated number of events defines the so-called number of streams (equivalent to Table 3.5), which is used to weight each event. Figure 3.4 shows the distributions of the calculated number of streams per run per process. Since the distributions peak sharply with little spread, I use the mean number of streams for all centrally generated events.

3.4.3 Recorded Dataset

For this analysis, I use the full Run 1 dataset with runs at the CM energies of $\Upsilon(4S)$ and $\Upsilon(4S)$ off resonance. These runs were recorded in the time period of 2019-2022. The analysis targets a combined integrated luminosity of $(408.110 \pm 1.712)\,\mathrm{fb}^{-1}$ [87].

Chapter 4

Reconstruction

In this chapter, I discuss the entire event selection process and its implications. Beginning with Section 4.1, I explain the event reconstruction, including the evaluation of the photon and event candidate selections, the best candidate selection, and the kinematic fit. Section 4.2 describes the event and candidate selection, which I apply after the reconstruction and before the MVA-based selection. In Chapter 7, I use the inversion of one of these selections as a control region. Finally, Section 4.4 discusses in detail the implementation of the MVAs and the two selection stages I apply before the signal extraction. These selection stages rely on different MVAs trained to select events based on the event kinematics and to address the additional challenge of photon candidates originating from merged photons[*]. Each section concludes with a summary listing all selections applied within that section.

For this and the following chapters, I use the following conventions:

- I define truth-matched signal events as events in which all three photon candidates and the ALP candidate are primary particles on generator level. Furthermore, I require that two photon candidates originate from the simulated ALP by matching the particle code of the generated mother and the candidate to the ALP.

- Since exactly three photon candidates are expected in the final state of $e^+e^- \to \gamma a$, $a \to \gamma\gamma$, it is necessary to distinguish the photon candidates. To resolve this ambiguity, unless otherwise noted, I discuss the photon candidates sorted by their energy. Accordingly, I denote the highest-energy photon in an event as γ^h, the middle-energy photon as γ^m, the lowest-energy photon as γ^l, and, if present, the photon with the next highest energy in the event as γ^4.

- Unless otherwise noted, all discussed quantities are defined in the laboratory frame.

[*]This selection was originally motivated by the $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\text{ISR}}$ background, which produces a large number of merged photon candidates. However, as I show in the control region studies in Chapter 7, this process is physically forbidden, and all contributions from $e^+e^- \to q\bar{q}$ processes are therefore removed from consideration. Consequently, the merged photon classifier selection is omitted from the final analysis.

4.1 Event Reconstruction

In the basf2 reconstruction procedure, all detector responses, in either simulation or recorded data, are evaluated, and a plethora of reconstruction and calibration algorithms are run. These algorithms form from the detector responses, so-called particle candidates, which I, as an analyst, combine to form event candidates from high-level, reconstructed quantities. In this section, I summarise all the selections that I apply during the event reconstruction. The event reconstruction is split into three parts: photon and event candidate selection, best candidate selection, and kinematic fit and its implications.

4.1.1 Photon and Event Candidates

I select photon candidates based on a minimal energy, their acceptance in the CDC, and their cluster timing. The minimal energy selection is $E_\gamma > 100\,\mathrm{MeV}$, which vetoes very-low-energy photon candidates, that are likely to be beam background or radiative photons. The CDC acceptance is defined as $17^\circ < \theta_\gamma < 150^\circ$. This selection ensures the possibility of vetoing tracks in the event, which could be associated with energy depositions in the ECL. Lastly, mainly to combat beam background and non-signal photon candidates, I require the cluster timing to be consistent with the ECL average event time. I first require `clusterErrorTiming` $< 10^6\,\mathrm{ns}$ which excludes failed timing fits. After all calibrations, photons from relevant physical process have cluster timing consistent with 0 ns. However, deviations can occur if the event's t_0 time is not determined by the ECL. For beam-induced background or particles with bad timing fits, the cluster timing may significantly differ from 0 ns. To further ensure reliable timing fits, I apply the selection $|T_{\mathrm{Cluster}}| < 200\,\mathrm{ns}$, following a recommendation from the Belle II collaboration.

For the event candidates, I require that the invariant mass of the three-photon system is within $80\,\% \cdot \sqrt{s} < M_{\gamma\gamma\gamma} < 105\,\% \cdot \sqrt{s}$ since I am searching for events with exactly three high-energy photons. Furthermore, to select event candidates where all three photons likely originate from the same interaction, I require that the three-photon candidates are simultaneous. I ensure the simultaneity of the three photon candidates by requiring that the maximum weighted difference between their reconstructed cluster times and the ECL average event time remains within the selection threshold. This maximum weighted distance $\Delta T^{\max}$ is defined as

$$\Delta T^{\max} = \max\left(\frac{|T_i - \hat{T}|}{\Delta T_i^{99}}\right), \tag{4.1}$$

where T_i is the timing of the photons, ΔT_i^{99} is the calibrated timing resolution of the crystal with the highest deposited energy, and

$$\hat{T} = \sum_j \frac{T_j / \left(\Delta T_j^{99}\right)^2}{1/\left(\Delta T_j^{99}\right)^2} \tag{4.2}$$

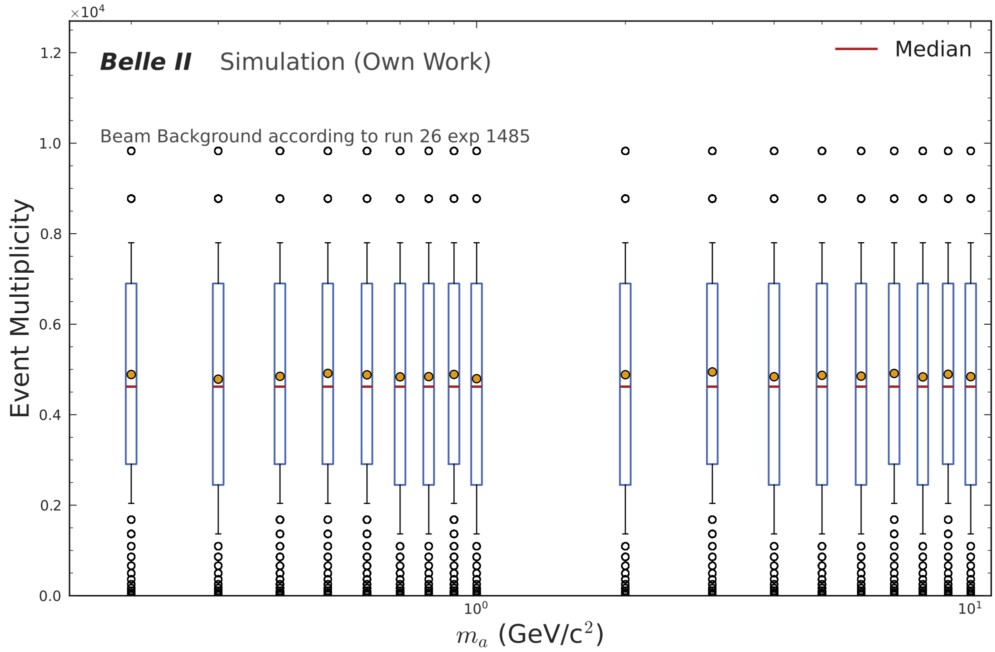
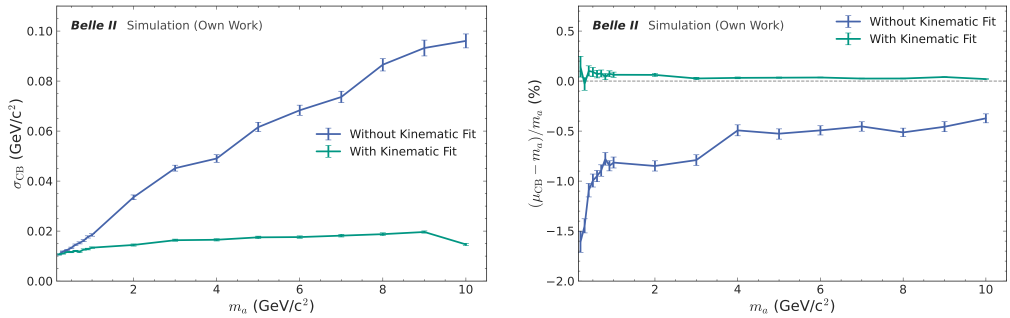


Figure 4.1: Event multiplicity distributions per simulated signal mass sample. The red line shows the median, the orange circle the mean, and the blue box depicts the 68 % quantile.

is the weighted average ECL time. This distance ensures that the three photons in the event arrive at the ECL in the same time window. On the contrary, beam background photons happen continuously; hence, they would lead to larger weighted distance values. I choose a selection threshold of two for the weighted distance, based on the distribution observed in MC signal simulations.

4.1.2 Best Candidate Selection

The multiplicity of each event is driven by the number of reconstructed photons in the event. Higher beam backgrounds lead to more photon candidates, and hence, the multiplicity drastically rises. For example, in experiment 26 run 1485, the average number of reconstructed photon candidates per event passing the previous selection in Section 4.1.1 is around 30, which leads to 4060 different combinations of three photons. Figure 4.1 displays the multiplicity distribution per event and mass point for beam background conditions corresponding to the previously mentioned run. To reduce the multiplicity, I apply a best candidate selection. It consists of ranking the reconstructed photons by their energy and only taking the three most energetic photons into account for the following stages of the event reconstruction. I observed no impact on the signal efficiency. The best candidate selection results in a multiplicity of three candidates per event since the three highest energetic photons can still be freely interchanged to form the ALP candidate.



(a) Signal Width (b) Signal Peak Position

Figure 4.2: Comparison of the width (left) and peak (right) position of the signal $M_{\gamma\gamma}$ distribution for events with and without the kinematic fit. Solid lines are just guiding the eye, but do not contain any information. The width is calculated as the fit parameter $\sigma^{\rm CB}$ parameter of the signal $M_{\gamma\gamma}$ distribution. The position is approximated by the $\mu^{\rm CB}$ parameter. Both quantities are explained in Section 6.1.

4.1.3 Kinematic Fit

To achieve a more accurate reconstruction of the three-photon final state and enable a uniform treatment of all mass hypotheses, I employ a kinematic fit of the event. By constraining the final state momenta of the three photons to the initial state, I gain a significant improvement in the invariant diphoton mass resolution $M_{\gamma\gamma}$. For the kinematic fit, I use the basf2 implementation of `OrcaKinFit` [88]. The improvement not only produces significantly narrower $M_{\gamma\gamma}$ peaks but also reduces the strong dependence on $M_{\gamma\gamma}$, which allows me to use $M_{\gamma\gamma}$ as a fit variable for all ALP masses. Figure 4.2a compares the width of the signal distribution in $M_{\gamma\gamma}$ for the same events with and without the kinematic fit. An additional effect of the kinematic fit, depicted in Fig. 4.2b, is the shift of the invariant mass peak central position towards the simulated ALP mass. This reduces the bias significantly, hence correcting the ALP mass m_a and the peak position in $M_{\gamma\gamma}$ is unnecessary. The calculation of the width and peak position through fitting is explained in Section 6.1.

4.1.4 Summary

The final reconstruction selection, which I explained in the previous subsections, consists of the following selections:

- **Photon Candidates**
  - $E_\gamma > 100\,\mathrm{MeV}$
  - $17^\circ < \theta_\gamma < 150^\circ$
  - $|T_{\rm Cluster}| < 200\,\mathrm{ns}$
  - `clusterErrorTiming` $< 10^6\,\mathrm{ns}$

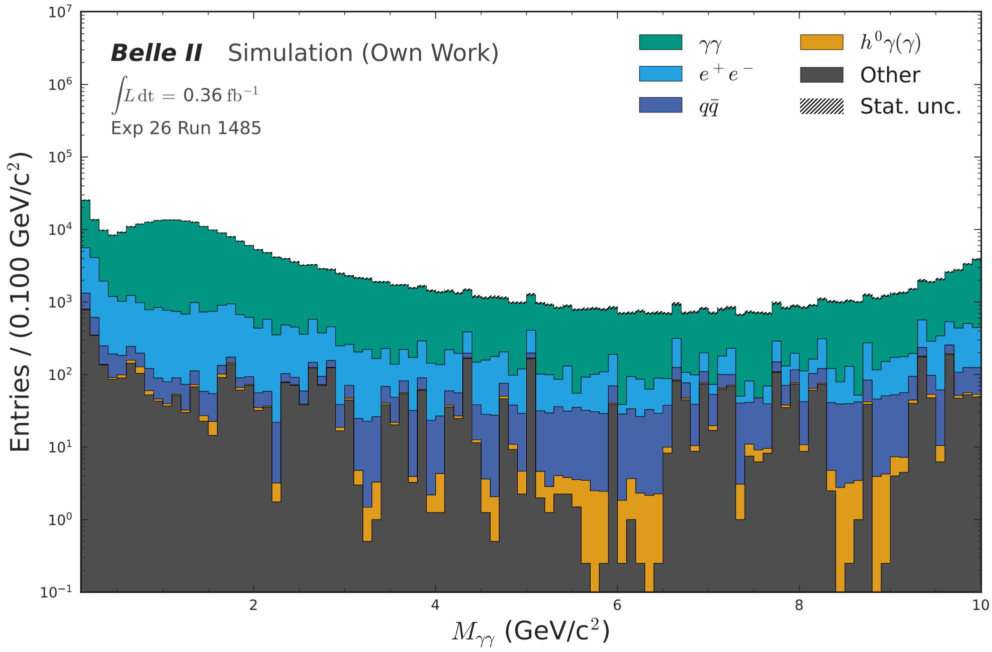


Figure 4.3: Distribution of the invariant diphoton mass, $M_{\gamma\gamma}$, after the application of the reconstruction criteria. The distribution is shown for experiment 26 run 1485.

- **Event Candidate $e^+e^- \to \gamma a$**
  - Three highest energetic photons in the event
  - $80\,\% \cdot \sqrt{s} < M_{\gamma\gamma\gamma} < 105\,\% \cdot \sqrt{s}$
  - Simultaneity of the three-photon candidates

Figure 4.3 displays the distribution of $M_{\gamma\gamma}$ for the background events in experiment 26 run 1485 after the application of the previously mentioned reconstruction criteria.

Figure 4.4 displays the reconstruction efficiency for the presented selection criteria for beam background conditions according to the $\Upsilon(4S)$ runs mentioned in Table 3.4. I define the signal efficiency as the ratio of the reconstructed and selected events to the number of generated events.

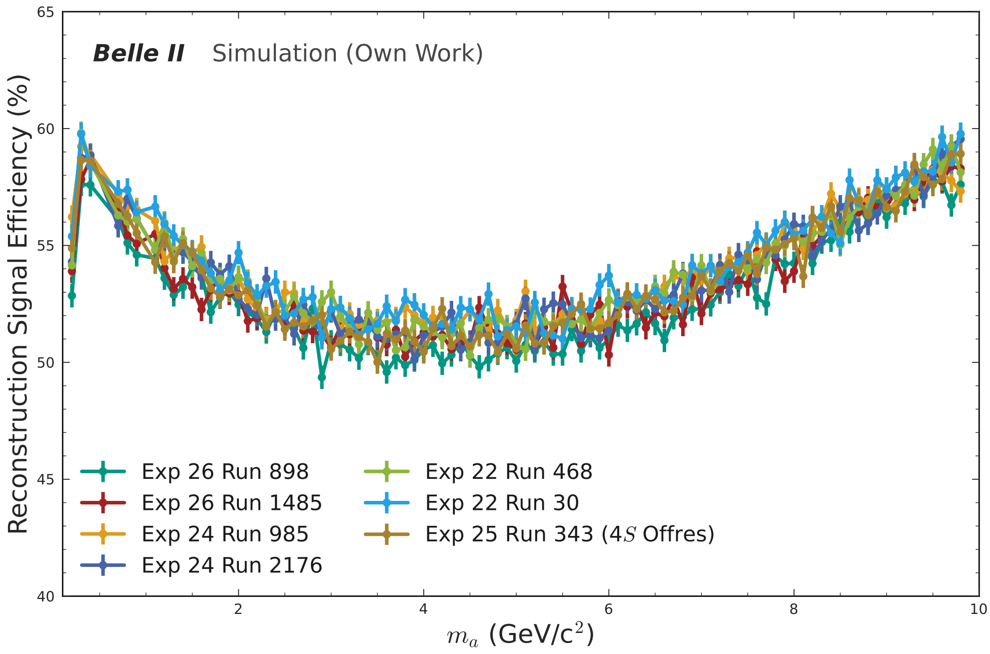


Figure 4.4: Reconstruction efficiency for the presented selection criteria for beam background conditions. The efficiency is shown in different colours for simulations with different beam background conditions.

4.2 Event Selection

After applying the minimal quality selections described in Section 4.1 and reconstructing the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ signature, I introduce a set of refined event selections before studying event candidates. In this section, I describe the additional criteria that I apply to the reconstructed events. These selections are designed to remove events that are not relevant to the analysis, either because they would not occur under ideal reconstruction conditions or because they can mimic signal-like features under specific conditions. In particular, I apply the following requirements: the rejection of events with good tracks, vetoing of photon pair conversions, removal of events containing a fourth photon, and vetoing of events in which a π^0 could form the ALP candidate.

4.2.1 Number of Clean Tracks

Reconstructing a fully neutral final state of three photons does not yield any tracks in the event. However, due to the presence of beam background and physical or technical limitations of cluster-track matching, some events contain tracks. To ensure that I only reconstruct exactly three good photons, I aim for a selection which guarantees the absence of good tracks.

Good tracks are driven by the ability to find and fit the track hits. A collaboration-wide recommendation uses the distance of a reconstructed track to the IP as a first-order measure

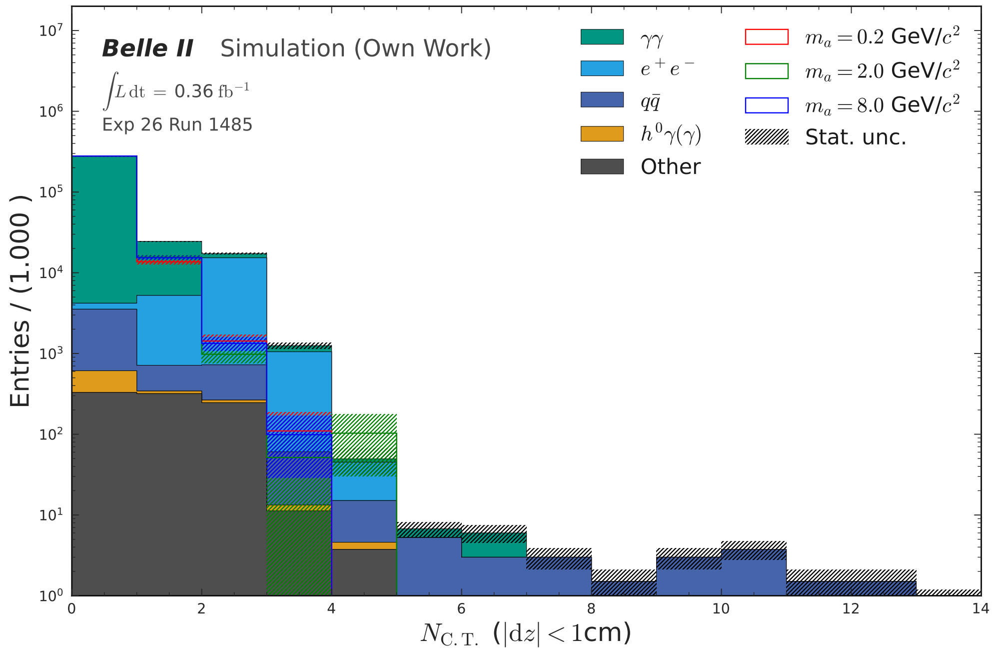


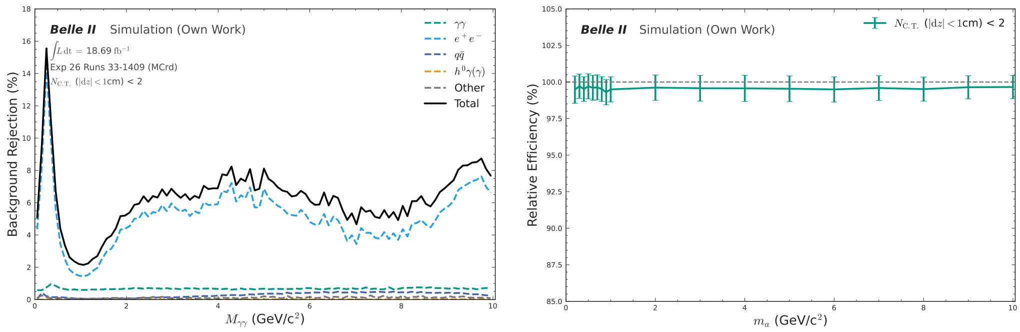
Figure 4.5: Distribution of the number of clean tracks $N_{\mathrm{C.T.}}$ in the ROE for the simulated background events and three example signal ALP masses. The tracks in the rest of the event are selected by $\mathrm{d}r < 1\,\mathrm{cm}$, $|\mathrm{d}z| < 1\,\mathrm{cm}$ and θ in CDC acceptance.

of good track definition. For the following study, I study the number of clean tracks $N_{\mathrm{C.T.}}$, which is the number of tracks in the event that pass a given quality selection. I select everything that passes a track quality selection based on the radial distance to the IP $\mathrm{d}r$ and the distance in z-direction $|\mathrm{d}z|$. Finally, I study the number of clean tracks in the event where clean tracks are defined by the tracks which pass the rest-of-event (ROE) mask, assuming a charged pion hypothesis.

At the time of writing, the recommendation suggests selecting good tracks with $\mathrm{d}r < 1\,\mathrm{cm}$, $|\mathrm{d}z| < 3\,\mathrm{cm}$ and θ in CDC acceptance. Due to significant disagreement in data and simulation, a warning is issued for selections of the number of CDC hits N_{CDC}. Figure 4.5 shows the distribution of the number of clean tracks $N_{\mathrm{C.T.}}$ with a tighter selection on $|\mathrm{d}z| < 1\,\mathrm{cm}$. I choose the tighter selection because it shows more signal events in the selection region.

As mentioned at the beginning of this section, beam background is one of the components of additional tracks in the event. Hence, any selection of the number of clean tracks introduces a dependence on the beam background conditions, which directly influences the signal efficiency. To avoid introducing a strong background condition dependence, I choose the selection $N_{\mathrm{C.T.}} < 2$. Figure 4.6a shows the background rejection as function of $M_{\gamma\gamma}$ relative to Fig. 4.3. The background rejection is defined as

$$\text{Background Rejection} = 1 - \frac{N(\text{After Selection})}{N(\text{Before Selection})}. \tag{4.3}$$



(a) Background rejection (b) Signal efficiency

Figure 4.6: Background rejection (left) and relative signal efficiency (right) for the number of clean tracks selection. The background rejection is highest for the $e^+e^- \to e^+e^-(\gamma)$ background process. The signal efficiency is not significantly affected by the selection.

Additionally, the plot displays the contributions of the different background processes to the total background rejection, which are shown normalised to the total background rejection. To calculate the number of background events, I use a test set of 18.69 fb^{-1} simulated events. I bin the $M_{\gamma\gamma}$ distribution in 100 bins between 0 to 10 GeV and calculate the number of background events in each bin with and without the presented selection, which in this case is the selection on $N_{\mathrm{C.T.}}$. As expected, the $N_{\mathrm{C.T.}}$ selection is most impactful for the $e^+e^- \to e^+e^-(\gamma)$ background events, since these produce in the majority of events at least two good tracks in the event. Figure 4.6b shows that the signal efficiency is not significantly affected by the $N_{\mathrm{C.T.}}$ selection. Here, I calculate the efficiency relative to the reconstruction efficiency in Fig. 4.4.

4.2.2 Pair Conversion Veto

Photons with energies above 1 MeV can undergo pair production when interacting with material, creating an electron-positron pair. In the Belle II detector, this process can occur in any mechanical component. The no-good-track selection removes most additional tracks from this process. However, if the tracks cannot be found, *e.g.* the pair conversion occurs significantly displaced from the IP, processes with two high-energetic photons of which one produces an e^+e^- pair result in a signal-like signature.

For this work, pair production in the outer CDC wall in $e^+e^- \to \gamma\gamma(\gamma)$ processes is a major background source in the low $M_{\gamma\gamma}$ region. The magnetic field separates the electron and positron outside of the tracking acceptance, leading to two potentially high-energy, well-resolved clusters without associated tracks. This mimics ALP photons with a small opening angle. Figure 4.7 shows the $M_{\gamma\gamma}$ distribution for $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$, where a clear concentration of events is visible at low $M_{\gamma\gamma}$ values.

I veto events with very small photon opening angles to reduce background from pair conversions. Figure 4.8 shows the θ and ϕ differences between photon pairs in $e^+e^- \to \gamma\gamma(\gamma)$

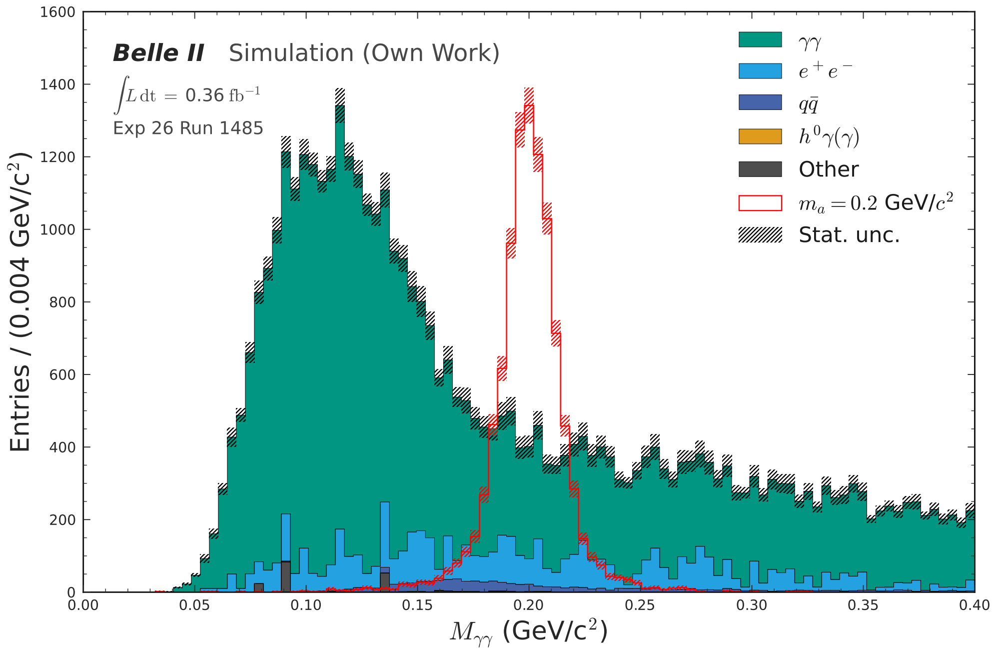


Figure 4.7: Distribution of $M_{\gamma\gamma}$ for values below 0.4 GeV for the simulated background events and one low signal ALP mass. The largest contribution originates from the $e^+e^- \to \gamma\gamma(\gamma)$ background process, which consists of pair conversion in the outer CDC wall.

events for the following photon pairs: high- and middle-energetic photons, high- and low-energetic photons, and middle- and low-energetic photons. The ϕ difference is calculated as

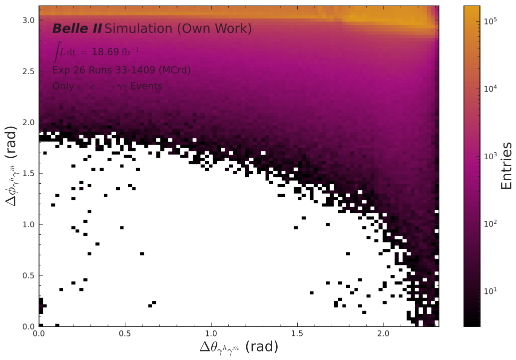
$$\Delta\phi_{\gamma^i\gamma^j} = \left|(\phi_{\gamma^i} - \phi_{\gamma^j} + \pi)\mathrm{mod}(2\pi) - \pi\right|, \tag{4.4}$$

ensuring rotational invariance, where i, j refere to one of the photon candidates.

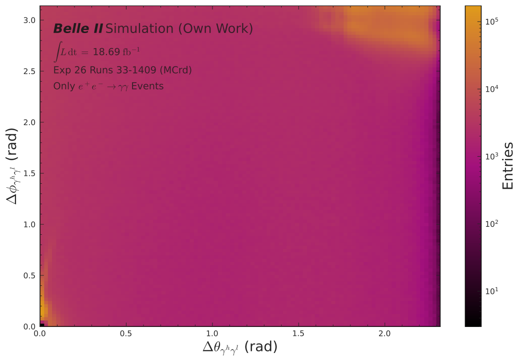
It is evident that for the high- and low-energy photon pairs, as well as the middle- and low-energy photon pairs, several events have an extremely low θ difference and low ϕ difference. Figure 4.9 displays a zoomed-in version of Fig. 4.8 for the angular differences, which are part of the selection. Such events align with the hypothesis of pair conversions occurring in the outer CDC wall. In this scenario, the initial photon remains unaffected by the magnetic field, resulting in negligible θ differences between the ECL clusters, while the produced lepton pair is deflected in the transversal plane.

For signal ALP photon daughters, the distributions of the angular difference depend strongly on the ALP mass hypothesis. Lower ALP mass events exhibit a strong correlation between the θ and ϕ differences. As can be seen in Fig. 4.10, signal events for low ALP mass events are distributed in a quarter circle for the high- and low-energy photon pairs, and the middle- and low-energy photon pairs. Higher ALP mass events display similar correlations, but with higher values for the angular differences.

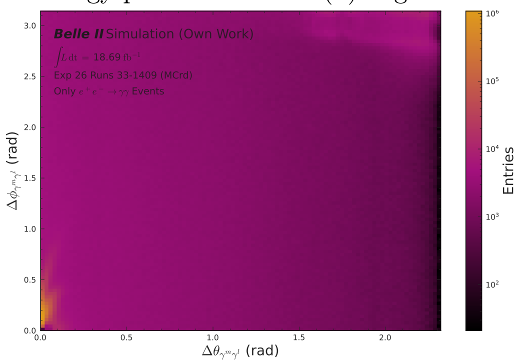
Any selection on the opening angles inevitably removes signal events. Consequently, my



(a) High- and middle-energy photons

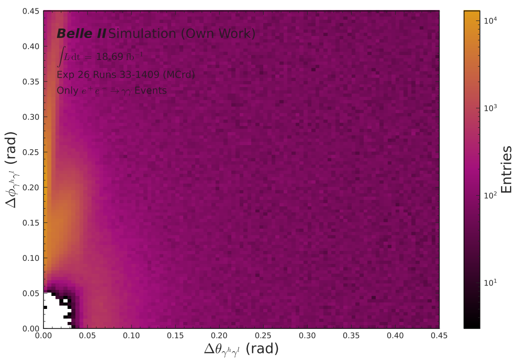


(b) High- and low-energy photons

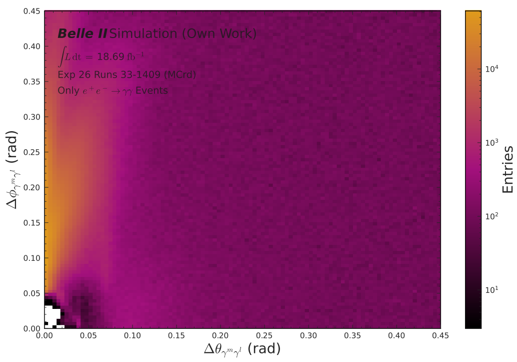


(c) Middle- and low-energy photons

Figure 4.8: 2D distribution of the angular differences between the photon pairs for the $e^+e^- \to \gamma\gamma(\gamma)$ background process.



(a) High- and low-energy photons



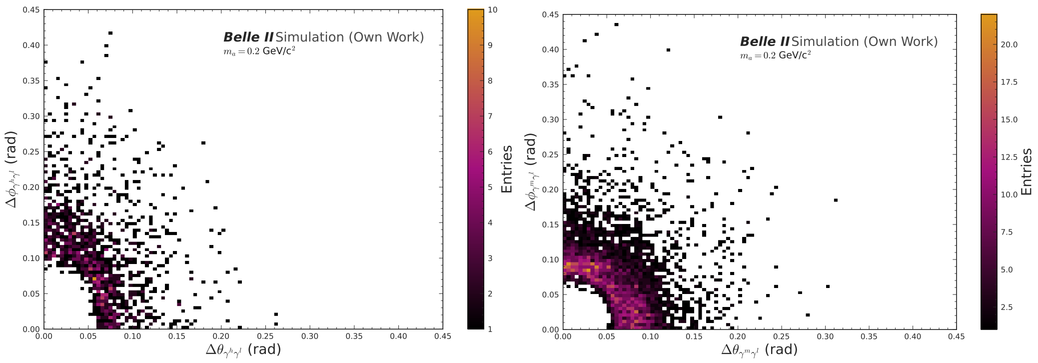
(b) Middle- and low-energy photons

Figure 4.9: 2D distribution of the angular differences between the photon pairs for the $e^+e^- \to \gamma\gamma(\gamma)$ background process, zoomed in on the selection region.

target is to remove only the cluster of background events with very small angular differences. Therefore, I make the following selection

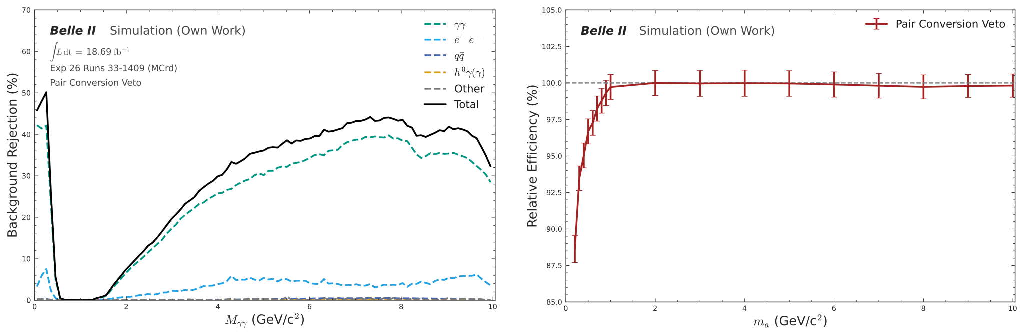
$$\Delta\theta > 0.014 \text{ rad} \ \& \ \Delta\phi > 0.4 \text{ rad}, \tag{4.5}$$

for the angular differences of the high- and low-energy photon pairs, as well as the middle- and low-energy photon pairs.



(a) High- and low-energy photons (b) Middle- and low-energy photons

Figure 4.10: 2D distribution of the angular differences between the photon pairs for the signal ALP mass hypothesis of $m_a = 0.4\,\mathrm{GeV}/c^2$ zoomed in on the selection region.



(a) Background rejection (b) Signal efficiency

Figure 4.11: Background rejection (left) and signal efficiency (right) for the pair conversion veto. The background rejection is highest for the $e^+e^- \to \gamma\gamma(\gamma)$ background process, specifically, for low $M_{\gamma\gamma}$ values. The signal efficiency is significantly affected by the selection also for $M_{\gamma\gamma}$ values.

Figure 4.11a displays the background rejection, which clearly shows a strong rejection for events with low $M_{\gamma\gamma}$ values of up to 50 %. The largest component of the rejected background events is the $e^+e^- \to \gamma\gamma(\gamma)$ background, which is expected since these events consist of high-energy photons possibly interacting with the detector material. The relative signal efficiency is shown in Fig. 4.11b where I observe a loss of up to 12 %.

4.2.3 Additional Photons

Similar to the absence of a good track selection from Section 4.2.1, this selection targets the clean-up of events not consistent with the three-photon signature. In signal events, it can be expected that three high-energy, well-resolved photons are present. As described in Section 4.1.1, the pre-kinematic fit $M_{\gamma\gamma\gamma}$ selection ensures the presence of these high-energetic photons. However, it also allows for additional photon candidates with energies of around 2 GeV in the entire event. Processes with four photons in the final state or processes

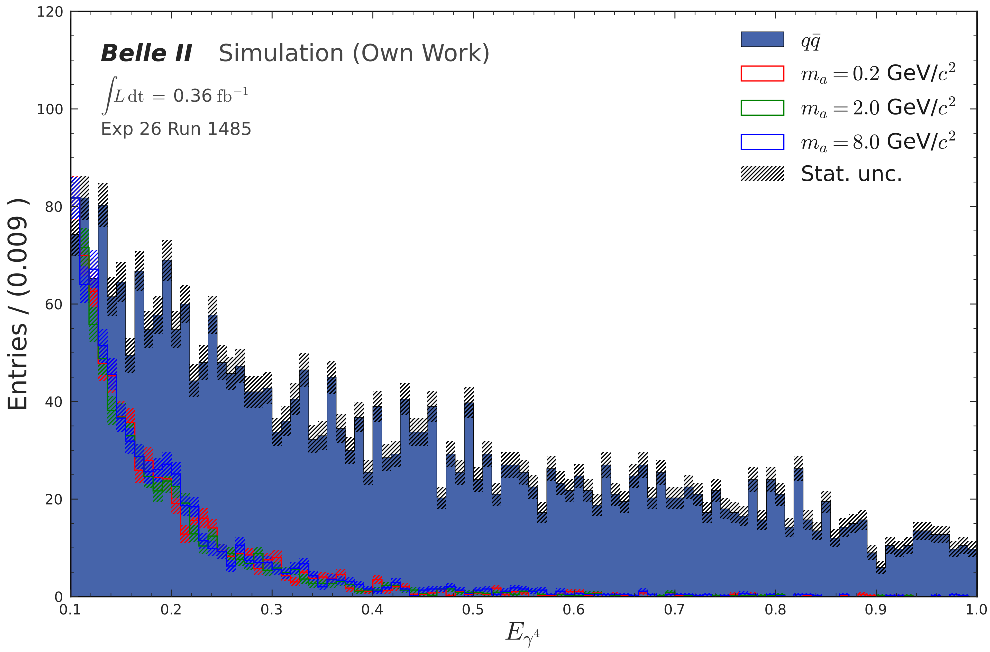


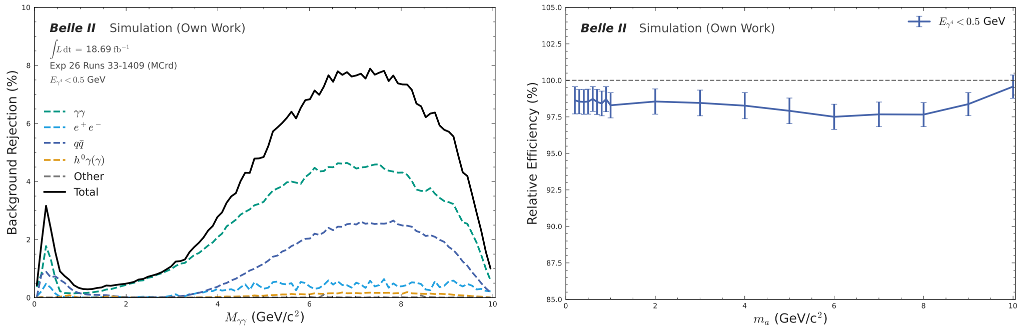
Figure 4.12: Distribution of the energy of the fourth-highest energetic photon for the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ background processes and three example signal ALP masses. The signal distributions show an exponentially decreasing distribution with only very few events above 0.5 GeV.

with additional ISR or Final State Radiation (FSR) photons can lead to events where three of the photons pass all previous requirements due to a low-energy fourth photon. A prominent example of processes with four photons is the decay $e^+e^- \to q\bar{q} \to \pi^0\pi^0$, which will be discussed in detail in the following section.

To reject events with additional photons, I study the fourth-highest energetic photon in the reconstruction. Figure 4.12 displays an example distribution of the fourth photon energy E_{γ^4} for simulated $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ events.

Due to beam background producing clusters with non-zero energy and without associated tracks, a full exclusion of a fourth photon candidate is not feasible without a significant signal efficiency loss. As Fig. 4.12 displays, for reconstructed signal events with beam background conditions corresponding to experiment 26 run 1485, signal events also show a significant number of events with a fourth photon candidate. Since the energy distribution shows events for a run with one of the highest beam background conditions, I conservatively choose to apply a selection of $E_{\gamma^4} < 0.5\,\mathrm{GeV}$.

Figure 4.13a shows the background rejection. Here, I observe a strong rejection of the $e^+e^- \to \gamma\gamma(\gamma)$ background, which is explained by the high number of events for this background process. Four photons in the final state can be expected from either additional beam background photon candidates or higher-order processes. The second largest component is the expected $e^+e^- \to q\bar{q}$ events. Figure 4.13b shows the relative signal



(a) Background rejection (b) Signal efficiency

Figure 4.13: Background rejection (left) and signal efficiency (right) for the fourth photon energy selection. The background rejection is highest for the $e^+e^- \to \gamma\gamma(\gamma)$ background process, and second highest for the $e^+e^- \to q\bar{q}$ background process. The signal efficiency is affected by the selection with a loss of about 2 % for all ALP masses.

efficiency with a loss of about 2 % equally for all ALP masses.

4.2.4 π^0 Veto

This pre-selection targets the removal of events where a π^0 would constitute the ALP candidate. As I pointed out earlier, one of the background sources, specifically in the low ALP mass region, is the decay $e^+e^- \to q\bar{q} \to \pi^0\pi^0$. Although the selection that I apply for the removal of additional photons from Section 4.2.3 removes a significant amount of these events, some events remain. Figure 4.14 displays the $M_{\gamma\gamma\gamma}$ distribution for the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ background for low ALP mass events. The histogram clearly shows one seemingly peaking structure around $M_{\gamma\gamma} \approx 0.18\,\mathrm{GeV}/c^2$ and another not-so-prominent structure around $M_{\gamma\gamma} \approx 0.25\,\mathrm{GeV}/c^2$. From generator information, I inferred that the left peak is dominated by events for which both photons originate from the same π^0. Events in the right bump show that the ALP candidate is built by photons originating from different π^0 in the decay.

Figure 4.15 summarizes the $M_{\gamma\gamma}$ distribution for the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ background for the different components. The reason for the peaks showing up in the $M_{\gamma\gamma}$ distribution with higher masses than the π^0 mass is the kinematic fit: Out of four photons in the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ event, I select the three highest energetic photons, leaving out a low-energy photon. The kinematic fit shifts the three-photon four momenta to the beam four momentum, thus, over-correcting the $M_{\gamma\gamma}$ mass due to the absence of the fourth photon in the event reconstruction. Additionally, for the events where the photons share the same grandmother, I select the higher energetic photons from each of the π^0, also leading to higher $M_{\gamma\gamma}$ values.

Because I do not want any π^0s building the ALP candidate, I apply a π^0 veto on the entire event. For this veto, I calculate the invariant mass of each photon pair before the kinematic fit - three invariant masses if there are three photons in the event or six invariant

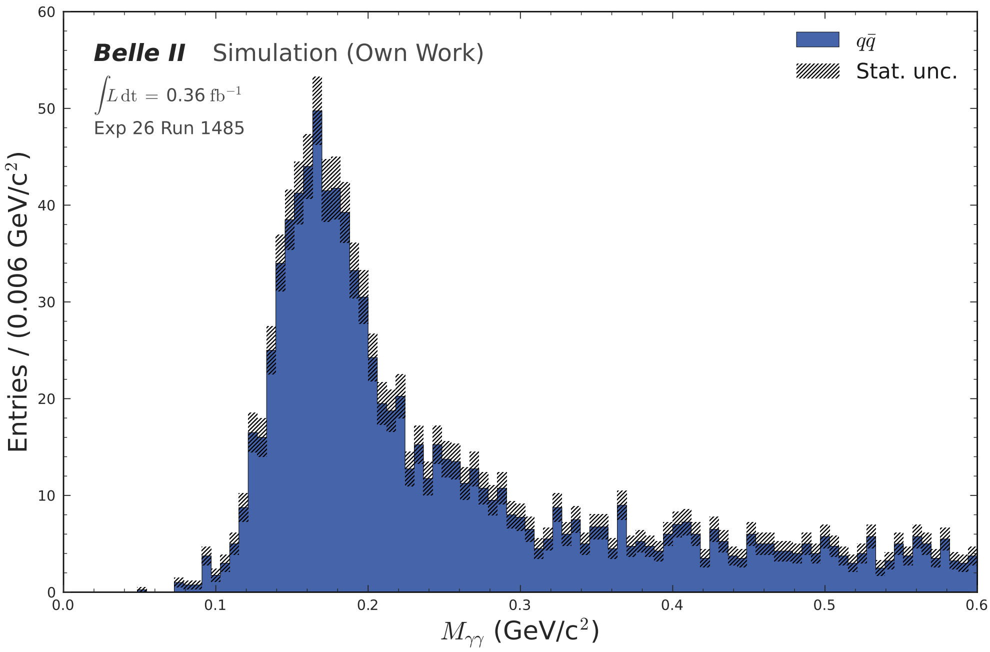


Figure 4.14: Distribution of $M_{\gamma\gamma}$ for the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ background processes. The distribution shows a clear peak at the π^0 mass and a second bump at slightly higher $M_{\gamma\gamma}$ values around $M_{\gamma\gamma} \approx 0.25\,\mathrm{GeV}/c^2$.

masses if there are four photons in the event. Figure 4.16 shows the distribution for the low-energy and fourth photon pairs, where a clear peak is visible at the π^0 mass. If any of these invariant mass combinations is below $M_{\gamma\gamma} \approx 0.16\,\mathrm{GeV}/c^2$, the entire event is discarded.

Figure 4.17a displays the background rejection for the π^0 veto. Again, the $e^+e^- \to \gamma\gamma(\gamma)$ background is the most affected background process. However, this is only true considering that random low-energetic photon combinations lead to low $M_{\gamma\gamma}$ values as well. Furthermore, as can be seen in Fig. 4.18, a substantial number of events that are affected by the π^0 veto are also affected by the pair conversion veto. The relative signal efficiency loss is about 5 % roughly equally for all ALP mass events.

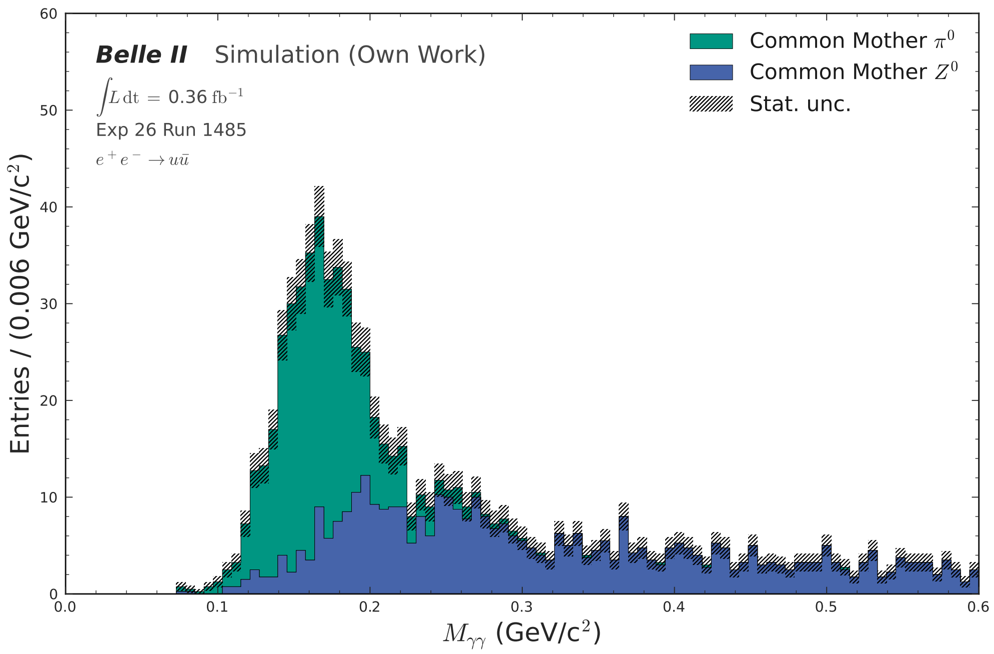


Figure 4.15: Distribution of $M_{\gamma\gamma}$ for the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ background processes. The background processes are separated by the common mother of the photons. The common mother (green) refers to the ALP candidate being built by two photons from the same π^0, while the other component (blue) refers to the case where the photons originate from different π^0.

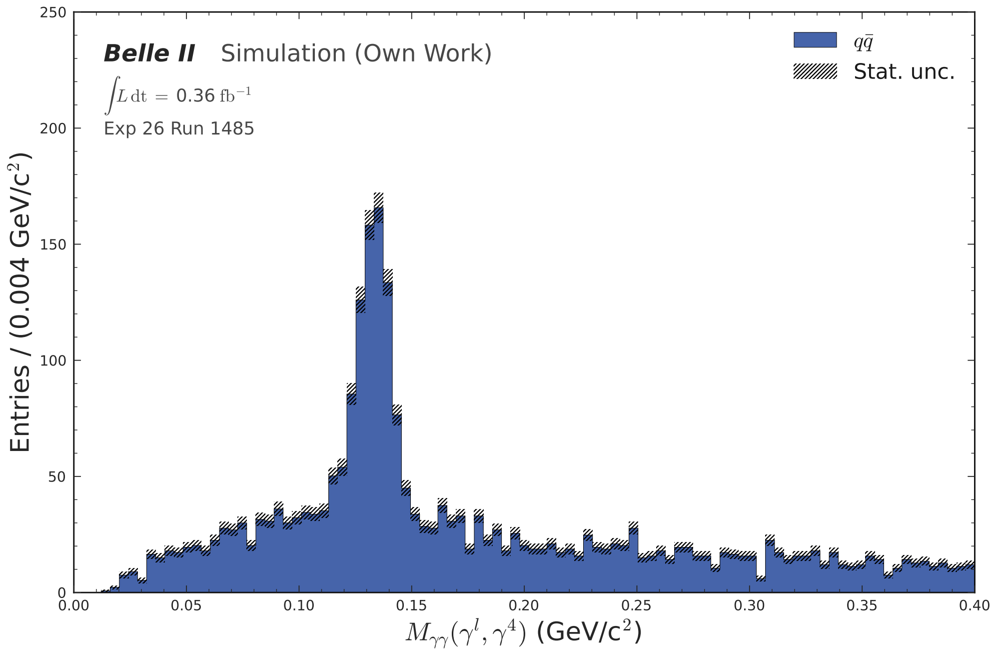
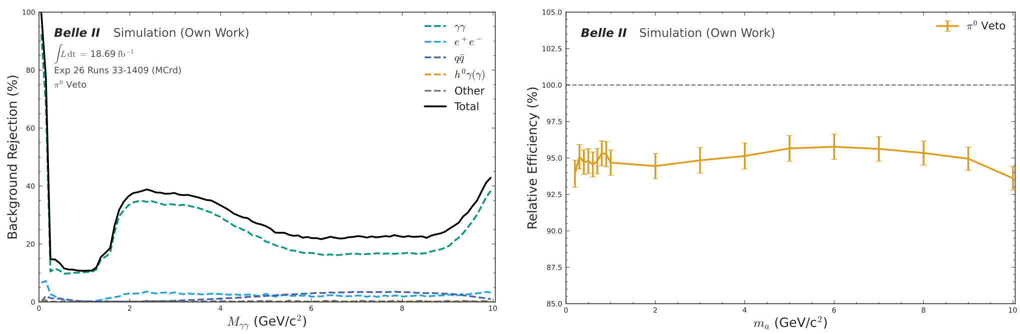


Figure 4.16: Distribution of the invariant mass of the low-energy and fourth photon pairs for the $e^+e^- \to q\bar{q} \to \pi^0\pi^0$ background processes. The distribution shows a clear peak at the π^0 mass.



(a) Background rejection

(b) Signal efficiency

Figure 4.17: Background rejection (left) and signal efficiency (right) for the π^0 veto. This selection displays an almost 100 % background rejection for very low $M_{\gamma\gamma}$ values.

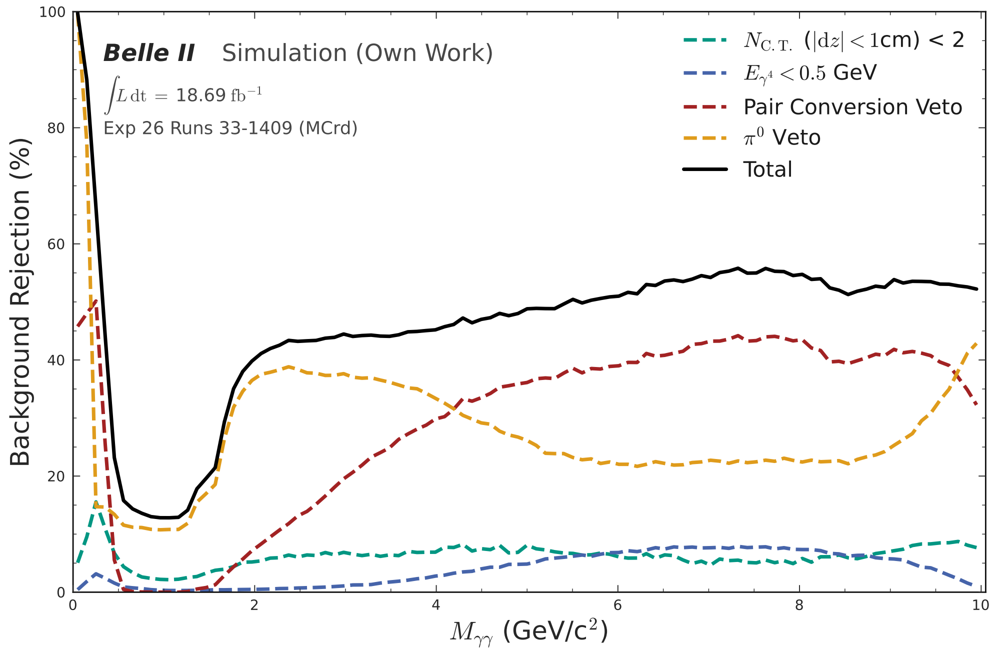


Figure 4.18: Summary of the background rejection for all selections. The total background rejection is displayed, as well as the contributions of the different background processes.

4.2.5 Summary

To summarise, the four selections that I apply after the reconstruction and before the candidate-based selection are as follows:

- $N_{\mathrm{C.T.}} < 2$
- $E_{\gamma^4} < 0.5\,\mathrm{GeV}$
- $\Delta\theta_{h,l} > 0.014$ rad & $\Delta\phi_{h,l} > 0.4$ rad & $\Delta\theta_{m,l} > 0.014$ rad & $\Delta\phi_{m,l} > 0.4$ rad
- $M_{\gamma^i,\gamma^j} > 0.16\,\mathrm{GeV}/c^2$ for $i,j \in [h,m,l,4]$

Figure 4.18 summarises the total background rejection for all discussed selections. Due to the π^0 and the pair conversion vetos, the background is almost entirely rejected for very low $M_{\gamma\gamma}$ values. For $M_{\gamma\gamma}$ values between roughly $M_{\gamma\gamma} \in [0.5, 2.0]\,\mathrm{GeV}/c^2$, the background rejection is lowest with only up to 15 %. For $M_{\gamma\gamma} \in [2.0, 10.0]\,\mathrm{GeV}/c^2$, the background rejection is again increasing with a maximum of roughly 45 %, dominated by the π^0 and pair conversion vetos.

Figure 4.19 displays the total background distribution in $M_{\gamma\gamma}$ for the entire Run 1 dataset after applying the final event reconstruction and the event selection.

Figure 4.20 summarises the effect of the selections on the total signal efficiency. In contrast to the background rejection, most selections are additive, meaning they do not affect the same signal events. The highest efficiency loss originates from the π^0 veto.

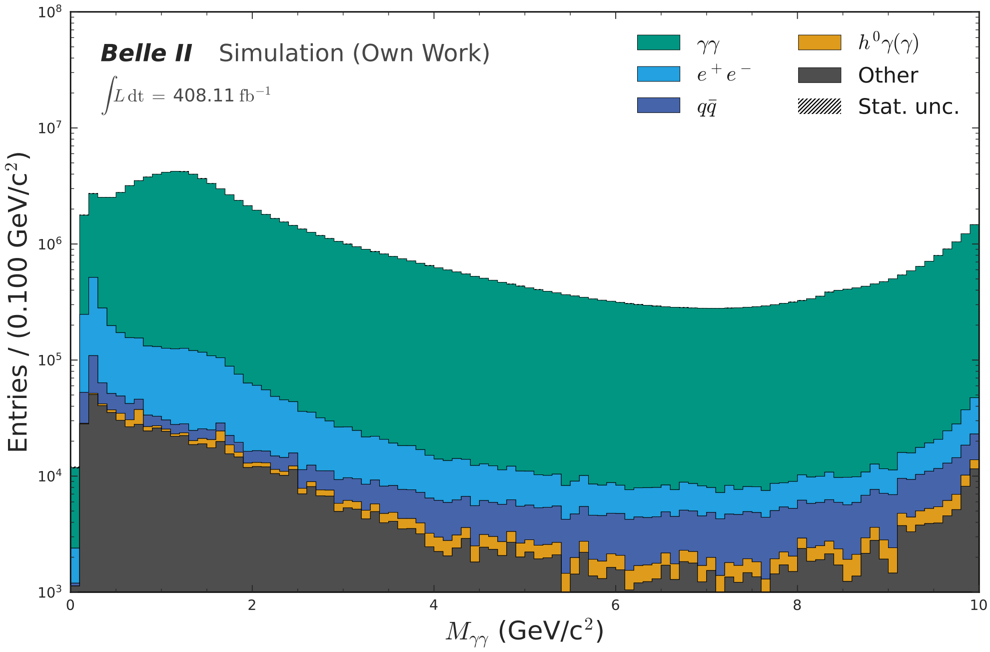


Figure 4.19: Distribution of $M_{\gamma\gamma}$ after applying the final event reconstruction and the event selection.

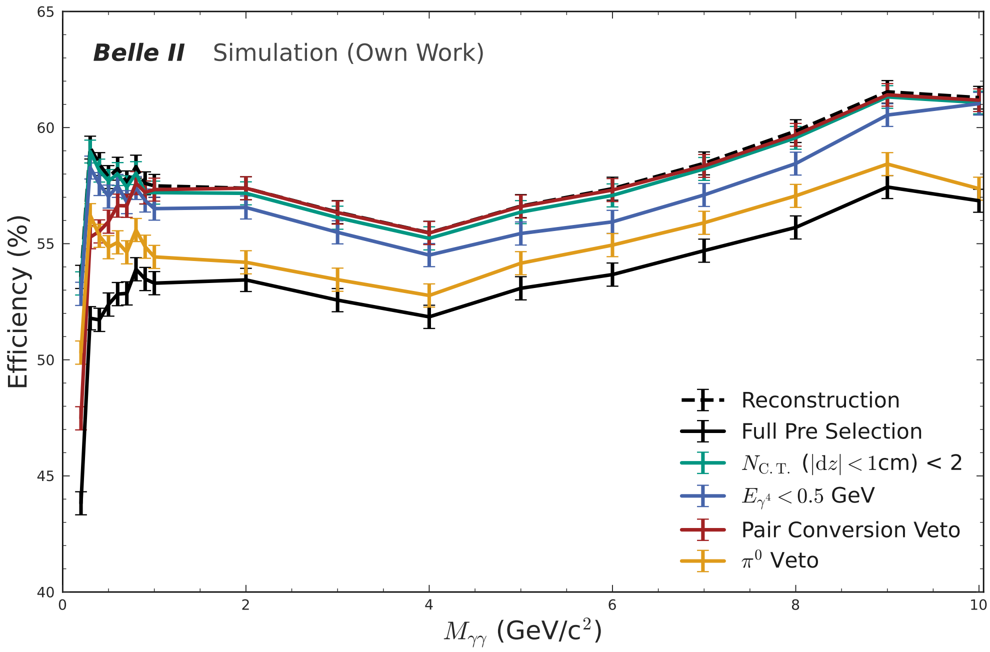


Figure 4.20: Summary of the signal efficiency for all selections. The total signal efficiency is displayed, as well as the contributions of the different selections. The π^0 veto shows the largest signal efficiency loss.

4.3 Candidate Selection

Due to the combinatorial nature of the searched-for signature, the selection in this section aims to reduce the candidate multiplicity. The candidate-based selection targets the helicity angle $|\cos\lambda|$ of the ALP candidate, which is a strong discriminator between two-body decays and other processes. Given the correlation of the helicity angle and the signal event properties, in preparation for the MVA-based selection, I add another selection on the photon energy at the end of this section. In this section, I describe the selection applied to the event candidates passing the event selection in Section 4.2.

4.3.1 ALP Helicity

The helicity angle λ is a commonly used variable in the context of decay properties. This angle provides exclusive information on the angular distribution of the decay products, polarisation, and spin correlations. In two-body decays $A \to B + C$ with $B \to X + Y$, the helicity angle of B is defined in the rest frame of B as the angle between the momentum vector of X $\mathbf{p}_X^B$ and the momentum of A in the same frame $\mathbf{p}_A^B$:

$$\cos\lambda = \frac{\mathbf{p}_X^B \cdot \mathbf{p}_A^B}{|\mathbf{p}_X^B||\mathbf{p}_A^B|}. \tag{4.6}$$

For true two-body decays, as for the $e^+e^- \to \gamma a$, the helicity angle of the ALP is distributed uniformly. Conversely, scattering processes like $e^+e^- \to \gamma\gamma(\gamma)$ will exhibit a strong tendency towards $\cos\lambda = \pm 1$ due to their collinear emission in the process.

Figure 4.21 shows the helicity angle distribution for the ALP candidate for the background and truth-matched signal processes. One essential aspect to keep in mind is that for the background contributions, the plot displays three entries per event. In the simplest case where the ISR photon is emitted collinear to one of the back-to-back leading-order photons, $|\cos\lambda|$ for two candidates is close to 1, where the ALP candidate is built by opposite side photons. Subsequently, for the case where the ALP candidate is built with the same side photons, $|\cos\lambda|$ is uniformly distributed.

To suppress candidates which are likely built from true $e^+e^- \to \gamma\gamma(\gamma)$ processes, I apply a selection on the ALP helicity angle. A selection of this kind can only be applied at the candidate level. The reason for this is the event shape of the unmatched signal process candidates, which is shown by Fig. 4.22. Truth-matched signal events are affected by the selection only through their uniform distribution in $|\cos\lambda|$. Nevertheless, depending on the topological distribution of the photons, the other signal candidates in the same event can exhibit a helicity angle close to $|\cos\lambda| = 1$. Events from very low and high ALP mass regions are especially affected due to their back-to-back topology. In the majority of the background scattering processes, with a tight enough selection, the number of candidates per event can be reduced to one. Entire events can only be excluded if all candidates fall

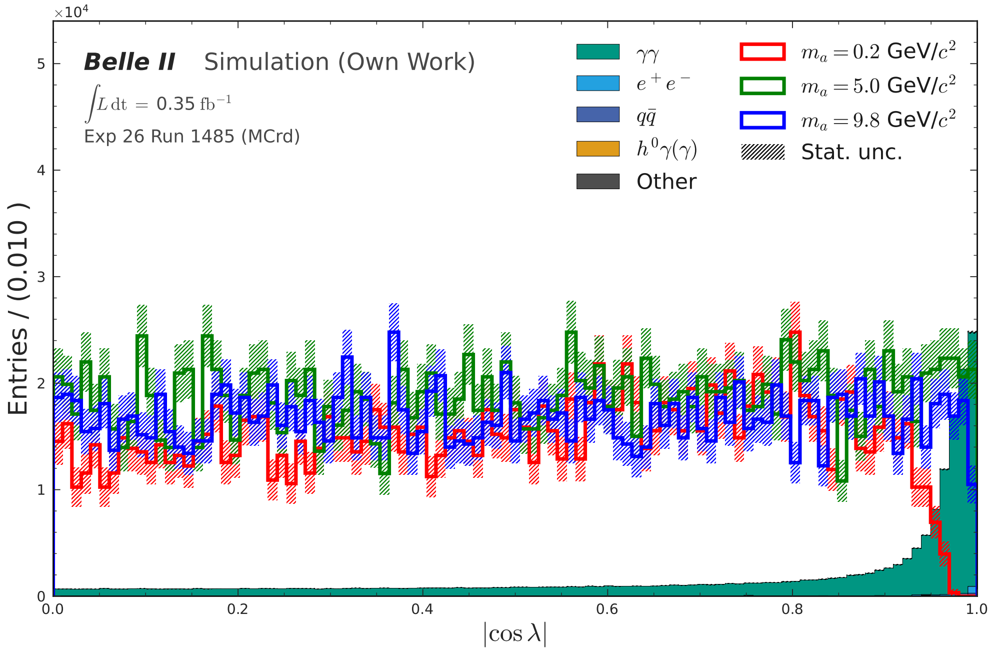


Figure 4.21: Distribution of $|\cos\lambda|$ for the simulated background events and three example signal ALP masses. The background distribution corresponds to the simulation based on detector conditions corresponding to experiment 26 run 1485. For the signal samples, a run with the same conditions is used, and only truth-matched events are considered. All events shown have passed the selections from Section 4.2.

outside the selection window.

To determine a common selection point in $|\cos\lambda|$ across all mass hypotheses, I use the Punzi Figure of Merit (PFOM) [89], defined as

$$\mathrm{PFOM} = \frac{\varepsilon}{\frac{a}{2} + \sqrt{B}}, \tag{4.7}$$

where ε denotes the signal efficiency, B is the number of background events, and a represents the target significance (in units of σ) for a one-sided hypothesis test. In this analysis, I choose $a = 5$, corresponding to a discovery-level significance. I calculate the PFOM for each ALP mass in the range of $m_a \in [0.16, 10]\,\mathrm{GeV}/c^2$ in steps of $0.01\,\mathrm{GeV}/c^2$ in the $\pm 10\sigma^{\mathrm{CB}}$ (see Section 6.1) signal window per sample mass. I then average across all ALP masses to get a mean PFOM as a function of the selection on $|\cos\lambda|$ and choose the maximum PFOM as the selection point.

Figure 4.23 shows the efficiency and number of background events for different selections of $|\cos\lambda|$. For the signal efficiency, I see an almost linear decline of the efficiency across all ALP masses simultaneously. For the number of background events, I see the highest impact in candidate reduction around $m_a \approx 1\,\mathrm{GeV}/c^2$ with a relatively loose selection on $|\cos\lambda|$. Between $m_a \in [5, 9]\,\mathrm{GeV}/c^2$, the number of background events is only significantly

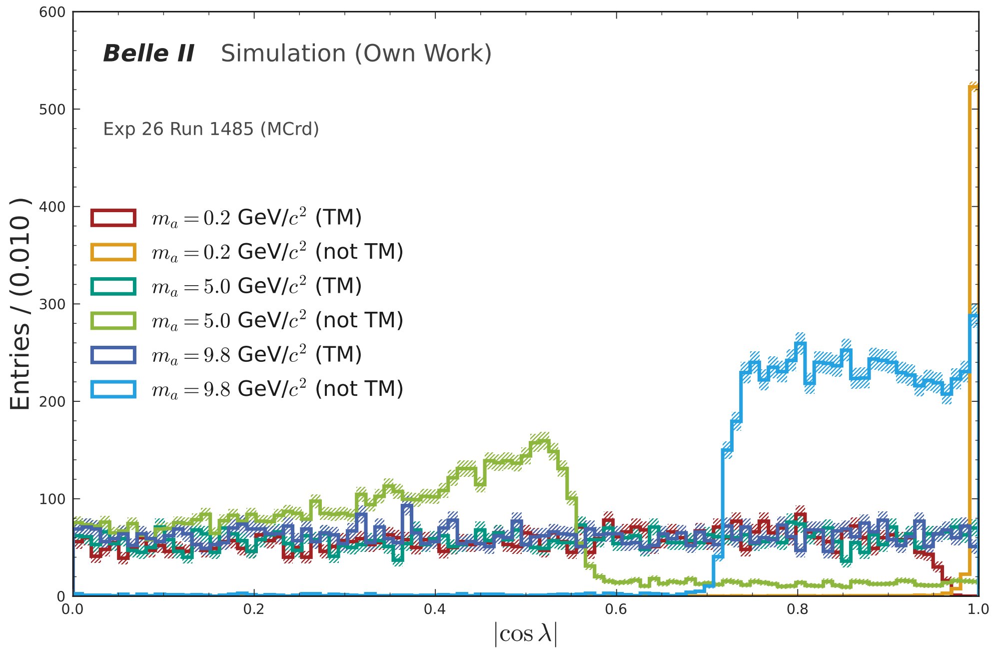
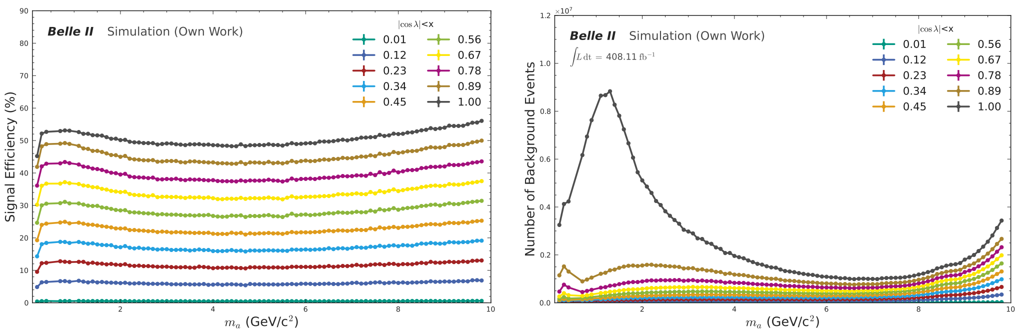


Figure 4.22: Distribution of $|\cos\lambda|$ for three example signal ALP masses. The colour family indicates different ALP masses. Darker colours correspond to truth-matched (TM) candidates, while lighter colours correspond to unmatched candidates. The unmatched distributions are scaled down to 5 % for $m_a = 0.2\,\mathrm{GeV}/c^2$ and 50 % for the other ALP masses for better visibility. All events shown have passed the selections from Section 4.2.



(a) Signal efficiency

(b) Number of background events

Figure 4.23: Signal efficiency (left) and number of background events (right) for different selections of $|\cos\lambda|$. The signal efficiency and number of background events are calculated in a $\pm10\sigma^{\mathrm{CB}}$ window per sample mass on which the training was performed. For the signal events, all six $\Upsilon(4S)$ runs in Table 3.4 are used. For the number of background events calculation, the Run 1 dataset for $\Upsilon(4S)$ and off-resonance $\Upsilon(4S)$ runs is used. The different colored lines correspond to different $|\cos\lambda|$ selections. All events shown have passed the selections from Section 4.2.

reduced after a tight selection.

Figure 4.24 shows the PFOM as a function of $|\cos\lambda|$ for example ALP masses and the

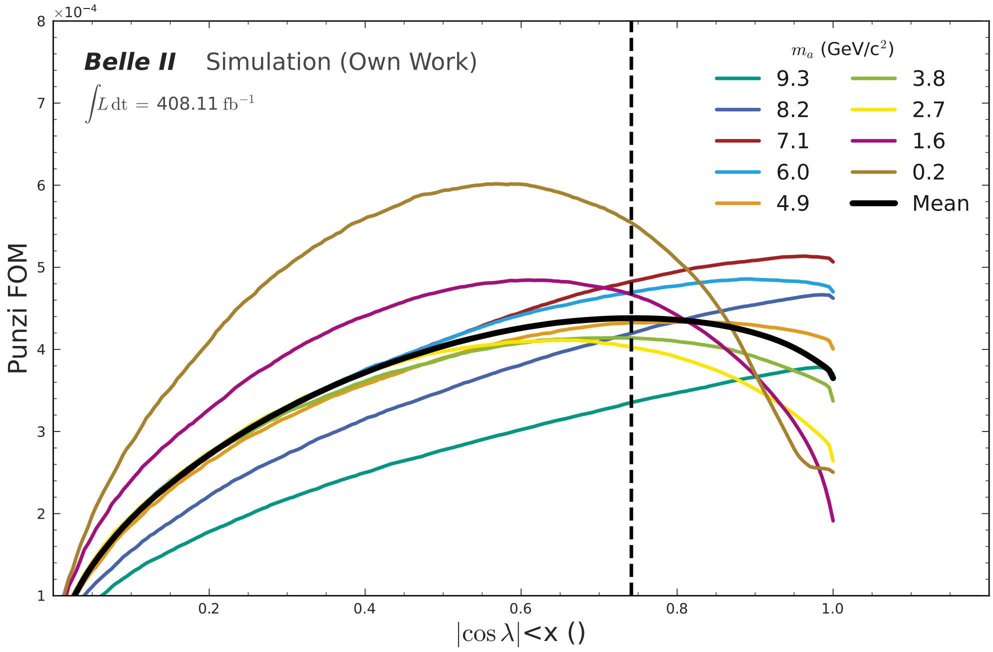
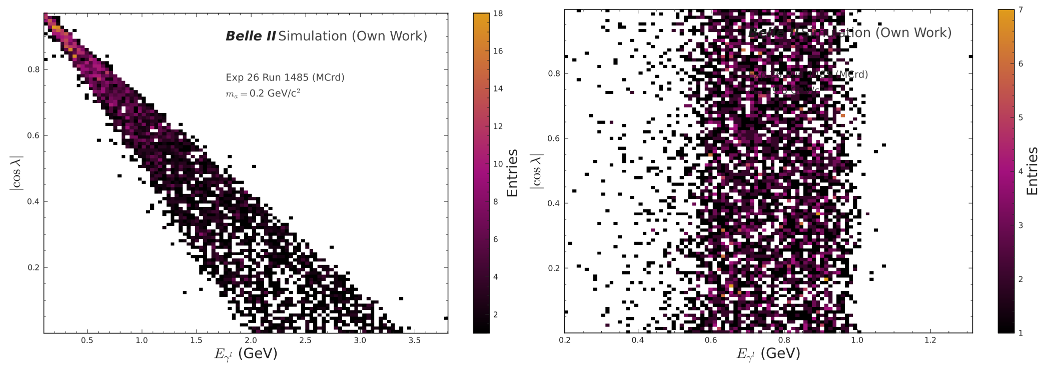


Figure 4.24: PFOM as a function of $|\cos\lambda|$ for few example ALP masses and the mean PFOM across all ALP masses. For the number of background events and the signal efficiencies, the same samples are used as in Fig. 4.23. Differently colored lines correspond to different ALP masses. The black line corresponds to the mean PFOM across all ALP masses, and the dotted line indicates the working point of $|\cos\lambda| < 0.74$.

mean PFOM across all ALP masses. It is visible that lower ALP masses profit from a tighter selection in $|\cos\lambda|$ while higher ALP masses benefit from a looser selection. To compromise between these two extremes, I choose a working point that balances the PFOM across all ALP masses by calculating the mean PFOM. According to the presented scan, a selection of $|\cos\lambda| < 0.74$ is chosen as the working point. Applying this selection on $|\cos\lambda|$ to the event candidates, the candidate multiplicity is reduced from 3 to on average 1.24 candidates per event for the background events.

4.3.2 Photon Energy

After selecting on the ALP candidate helicity angle, a tighter selection on the photon energy becomes effective. Before applying the helicity angle selection, a minimal photon energy cut does not improve the sensitivity, since the energy distribution of the lowest-energetic photon is broad for low ALP mass hypotheses. However, for low ALP masses, the lowest-energetic photon often originates from the ALP decay due to energy–momentum conservation. In this case, as shown in Fig. 4.25a, the helicity angle of the ALP candidate and the energy of the lowest-energetic photon, E_{γ^l}, are fully correlated. For higher ALP masses, this correlation disappears (see Fig. 4.25b), as the lowest-energetic photon is typically the recoiling photon in $e^+e^- \to \gamma a$, because most of the total collision energy is used to produce the ALP.



(a) $m_a = 0.2\,\mathrm{GeV}/c^2$ (b) $m_a = 9.8\,\mathrm{GeV}/c^2$

Figure 4.25: 2D distributions of the energy of the lowest-energetic photon in the event E_{γ^l} versus the helicity angle $|\cos\lambda|$ of the ALP candidate for two example ALP masses. The signal samples show only truth-matched events simulated with detector conditions corresponding to experiment 26 run 1485.

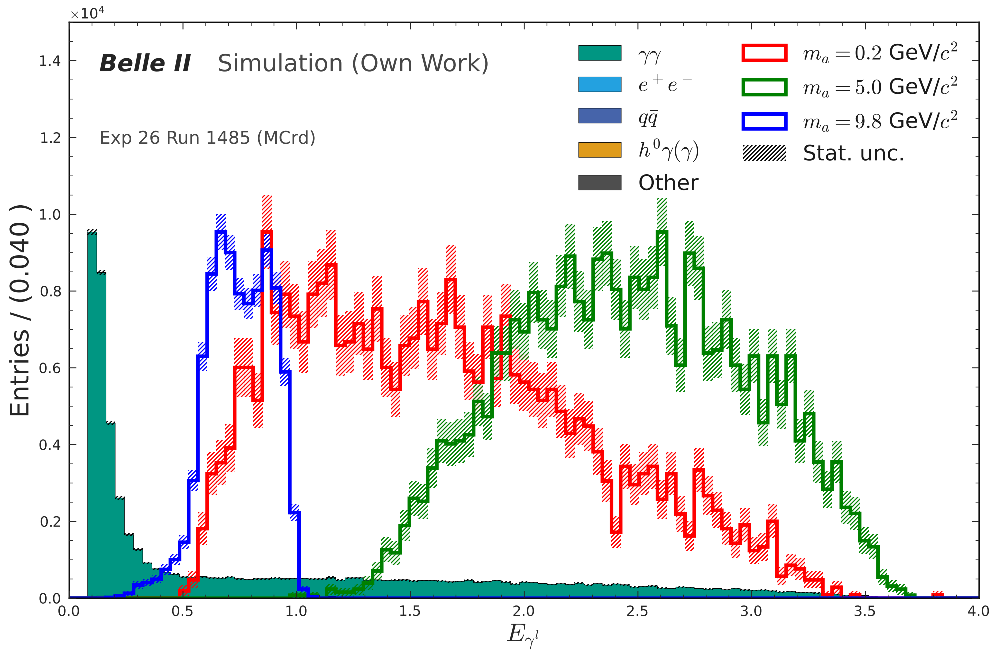


Figure 4.26: Distribution of E_{γ^l} for the simulated background events and three example signal ALP masses. The background distribution corresponds to the simulation based on detector conditions corresponding to experiment 26 run 1485. For the signal samples, a run with the same conditions is used, and only truth-matched events are considered. All events shown have passed the selections from Sections 4.2 and 4.3.1.

Figure 4.26 displays the distribution of the energy of the lowest-energetic photon in the event E_{γ^l} for the background and truth-matched signal processes. To further reduce the background contributions, I apply a selection on the lowest-energetic photon energy of $E_{\gamma^l} > 0.5\,\mathrm{GeV}$.

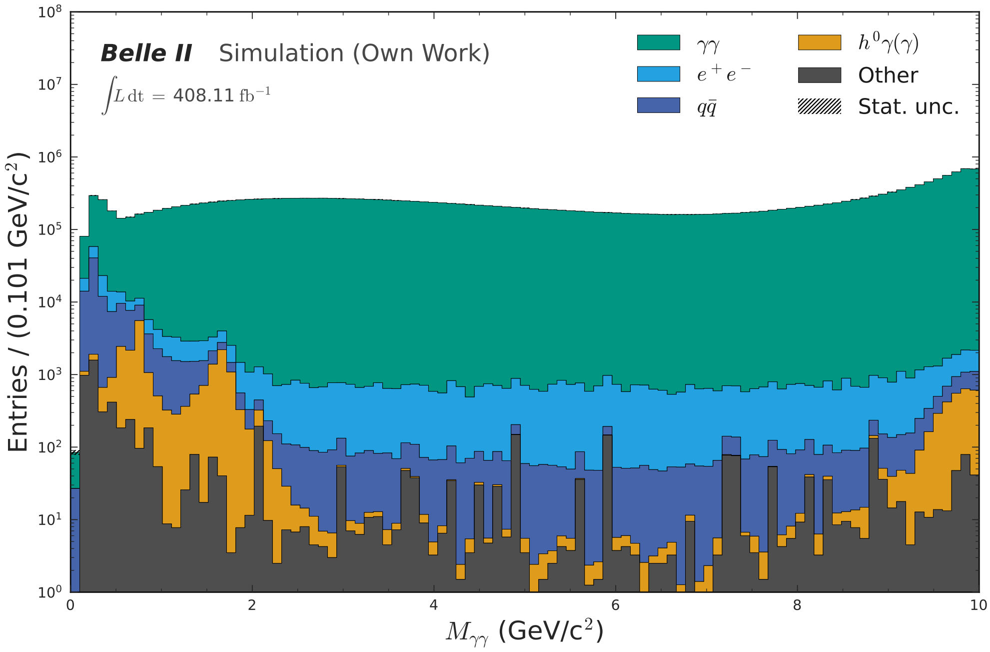


Figure 4.27: Distribution of $M_{\gamma\gamma}$ after applying the event reconstruction, the event selection and the candidate selection.

4.3.3 Summary

To summarise, the candidate-based selection focuses on the helicity angle $|\cos\lambda|$ of the ALP candidate and the resulting selection on the energy of the lowest-energetic photon in the event E_{γ^l}:

- $|\cos\lambda| < 0.74$
- $E_{\gamma^l} > 0.5\,\mathrm{GeV}$

The helicity angle selection reduces the average event multiplicity from strictly 3 candidates per event to, on average, 1.24 candidates per event. Figure 4.27 displays the total background distribution in $M_{\gamma\gamma}$ for the entire Run 1 dataset after applying the event reconstruction, the event selection and the here presented candidate selection.

Figure 4.28 summarises the effect of the selections on the total signal efficiency.

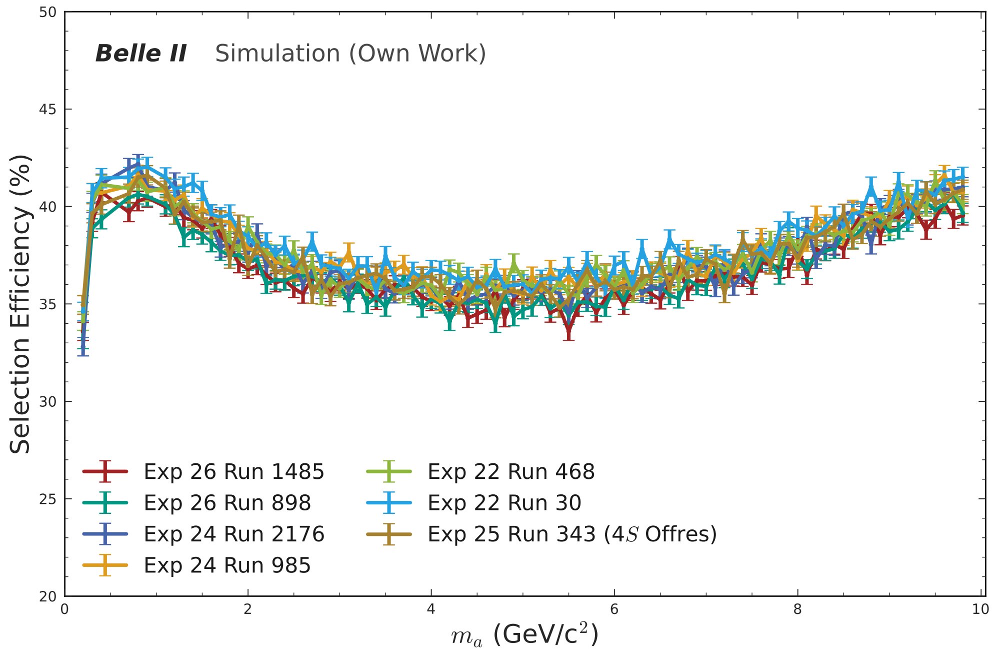


Figure 4.28: Signal efficiency for the ALP mass points after the event reconstruction, the event selection and the candidate selection for several runs with different beam background conditions. The efficiency is shown in different colours for simulations with different beam background conditions.

4.4 MVA Selection

To further increase the sensitivity of the search, I apply an MVA-based selection to the reconstructed events that remain after all previous selections. As MVA techniques are an integral part of modern high-energy physics analyses, I assume a basic familiarity with their principles and applications. The event selection that I present in this section is based on a simple neural network (NN) architecture — the Feedforward Neural Network (FNN), also known as a Multilayer Perceptron (MLP). This architecture is commonly used to learn compact representations of input parameters through hidden layers, non-linear activation functions, and backpropagation [90]. Furthermore, the FNN is a supervised learning method, meaning that the network is trained using labelled input data.

The goal of the selection is to define a common selection applicable to all ALP masses simultaneously, ideally enabling a scan over any ALP mass of interest, even if that specific mass was not included in the training. Moreover, this strategy simplifies the performance study of the selection, as it avoids sharp transitions between distinct selection regions. To achieve this goal, the signal training samples always consist of a variety of different ALP masses. In the following sections, I describe the general training procedure and the two training stages: one targeting the overall kinematic separation between signal and background events, and another addressing the specific challenge of classifying merged photons.

4.4.1 Training Procedure

The MVA training procedure is divided into three parts: preparation of the training sample, training of the FNN, and evaluation of the classifier output. The overall procedure does not change for either of the two classification stages; only the input training sample, the features, and the architecture differ. For the preparation and training of the FNN, I use the PyTorch [91] framework as well as the Scikit-learn [92] library. To simplify the logging and tracking of the training process as well as the hyperparameter scans, I use the Weights & Biases [93] framework.

Input Samples

The training samples consist of different experiment and run combinations for the background and signal samples, as well as for the different classifier stages. As previously mentioned, I target a single selection per stage for all signal masses at once. Subsequently, the signal training samples always consist of a variety of equally distributed ALP masses across the entire analysis range. To balance the training samples, I add weights to the background and signal events, and randomly sample according to the actual number of events that are in the training sample. The specific experiment and run combinations, as well as the ALP masses and the sampling rate, depend on the stage of the selection and will be described in the respective sections.

Background Weights

I label any event from the simulated processes in Tables 3.5 and 3.6 as background. Subsequently, I weight the events according to the process cross section. More precisely, for a given luminosity, I determine the weight by the product of the number of events in the training sample and the inverse number of available streams. For the $e^+e^- \to h^0\gamma(\gamma)$ processes, I use the average number of streams which were produced (see Fig. 3.4). Lastly, I normalise all the event weights to one. This weighting ensures that each process is correctly represented in relation to the other processes.

Signal Weights

For the training, I use only the truth-matched events, and do not consider the combinatorial signal events. To account for different reconstruction and pre-selection efficiencies, I weight the signal events according to the number of events per ALP mass in the training sample. Subsequently, I normalise all the signal event weights to 1 with respect to the background event weights. In the second stage of the selection, I apply an additional weight to the signal events according to the ALP mass. I adjust each signal event weight w_s

$$w'_s = w_s e^{-m_a \alpha}, \tag{4.8}$$

where α becomes a free parameter in the training of the FNN and is chosen based on the resulting performance. I introduce the additional weight to give lower masses a higher importance in the training.

Input Features

The input features depend on the stage of the selection and are described in the respective sections. To account for the different ranges of the features, I normalise the input features to a range of $[0, 1]$. If an event has a feature with no value, as for the fourth photon features, I set the value to -1. For all used features, I sort the photon candidates by their energy in the event as described at the beginning of this chapter, thus reducing the number of three-photon candidates to one per event if not already reduced by the truth matching in the signal samples.

Architecture

I employ a classical FNN for this analysis. The task is to classify events as either signal or background, so the network is designed to output a single probability P^{FNN}: 0 for background and 1 for signal. During the hyperparameter optimisation, I explore different architectures to find the best-performing one. All architectures start with a number of hidden nodes equal to the number of input features N_{feature} and end with a single output node. The number of hidden layers N_{layer}, the maximum number of hidden nodes $N_{\mathrm{nodes}}^{\mathrm{max}}$, and the activation function are open to optimisation. For architecture layouts, I test a range of different layer configurations:

- **Constant**: The number of hidden nodes is constant across all layers, starting with N_{feature} and raised to $N_{\mathrm{nodes}}^{\mathrm{max}}$ until the last layer.
- **Linear Encoder**: The number of hidden nodes is raised in the second layer to $N_{\mathrm{nodes}}^{\mathrm{max}}$ and then reduced linearly until the last layer.
- **Exponential Encoder**: The number of hidden nodes is raised in the second layer to $N_{\mathrm{nodes}}^{\mathrm{max}}$ and then reduced exponentially until the last layer.
- **Decoder**: The number of hidden nodes is linearly increased until $N_{\mathrm{nodes}}^{\mathrm{max}}$ before the last layer.
- **"Bottleneck"**: Combination of encoder and decoder architecture, in this order.
- **"Diamond"**: Combination of decoder and encoder architecture, in this order.

Figure 4.29 visualises the different architectures for 40 N_{feature}, a maximal depth of $N_{\mathrm{nodes}}^{\mathrm{max}} = 128$ and a number of layers of $N_{\mathrm{layer}} = 15$. Additionally, I introduce batch normalisation between the second and the next-to-last layer as well as a 20 % dropout until the next-to-last layer.

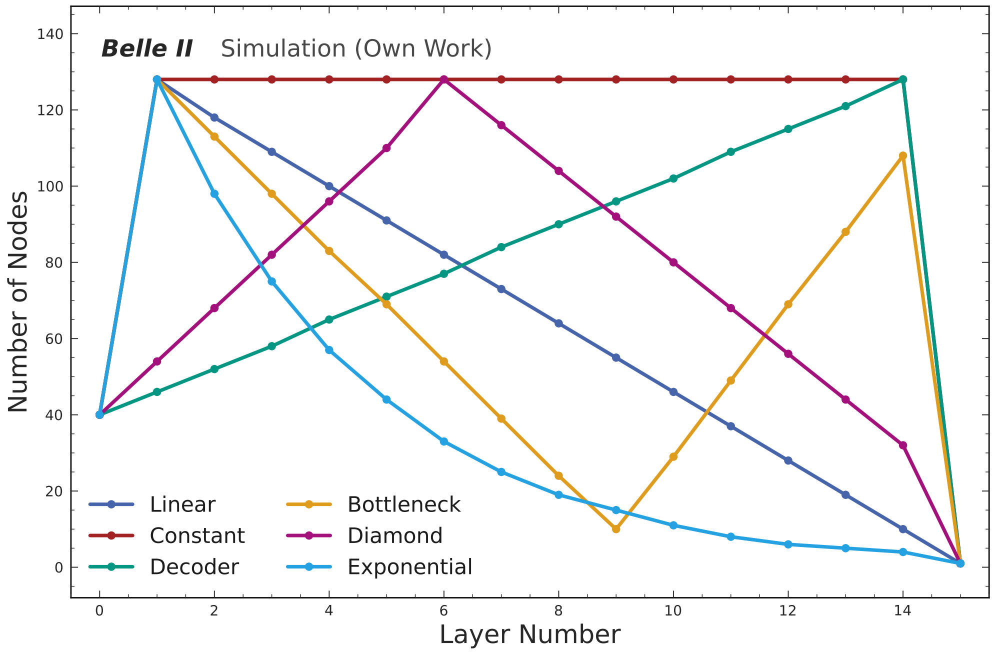


Figure 4.29: Show of different architectures for the FNN used during the hyperparameter optimisation in the selection training. Each colour represents a different architecture layout where the number of hidden nodes is shown on the y-axis and the layer number on the x-axis.

Loss Function

As the classification targets a binary classification of the events, I use the Binary Cross-Entropy (BCE) loss. To account for the different event weights, I multiply the loss by the weight before summing over all events and finally dividing the loss-sum by the sum of the weights. The overall training performance in terms of gradient stabilisation and improved generalisation is enhanced by batching the training samples. The batch size as well as the number of epochs are also part of the hyperparameter optimisation.

Learning Rate

To ensure consistent training improvement over the total training time, I use a learning rate scheduler. The training starts with a learning rate of 5×10^{-3} and is halved if the loss is not reduced by a factor of 1×10^{-4} for 30 epochs.

Evaluation

I track the general training performance by the training loss as a function of the epoch. To spot general overtraining, I split the input samples into training, validation and test samples. For both stages, I split the samples by random sampling into 60 % for training, 30 % for validation and 10 % for testing. I define a training as successful if the validation loss does not significantly diverge from the training loss. In addition to the loss, during the hyperparameter optimisation, I study the signal efficiency ε for each training for a given selection of the classifier output to estimate the generalisation performance for all signal masses. The goal for the signal efficiency is not to introduce any strong fluctuations between different signal mass points. If two trainings show similar loss and equally good performance for the signal efficiency shape, I choose the training with the higher Area Under Curve (AUC) of the receiver operating characteristic (ROC) curve as the best performing one.

4.4.2 Kinematic Selection

The first stage of the selection targets the removal of events that are easily distinguishable from signal events purely because of their kinematic properties. For this purpose, the input features that I use to train the classifier are based on functions built from the energy and spatial coordinates of the photon candidates in the event.

The features* that are presented to the network are the following:

- **Energy** of the photons: E_{γ^h}, E_{γ^m}, E_{γ^l}
- **Energy product** $E^*_{i,j} = E_{\gamma^i} \cdot E_{\gamma^j}$ of the photons: $E^*_{h,m}$, $E^*_{h,l}$, $E^*_{m,l}$
- **Polar angle difference** $\Delta\theta_{i,j} = |\theta_{\gamma^i} - \theta_{\gamma^j}|$ of the photons: $\Delta\theta_{h,m}$, $\Delta\theta_{h,l}$, $\Delta\theta_{m,l}$
- **Azimuthal angle difference** $\Delta\phi_{i,j} = (\left|\phi_{\gamma^i} - \phi_{\gamma^j}\right| + \pi)\mathrm{mod}2\pi - \pi$ of the photons: $\Delta\phi_{h,m}$, $\Delta\phi_{h,l}$, $\Delta\phi_{m,l}$
- **Opening angle** $\Delta\alpha_{i,j} = \cos\theta_{\gamma^i} \cdot \cos\theta_{\gamma^j} + \sin\theta_{\gamma^i} \cdot \sin\theta_{\gamma^j} \cdot \left(\cos\phi_{\gamma^i} - \cos\phi_{\gamma^j}\right)$ of the photons: $\Delta\alpha_{h,m}$, $\Delta\alpha_{h,l}$, $\Delta\alpha_{m,l}$
- **Sphericity** of the event, where the spericity $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ is defined as the linear combination of the sphericity eigenvalues λ_i of the sphericity tensor

  $$S^{\alpha\beta} = \frac{\sum_{i=1} p_i^\alpha p_i^\beta}{\sum_{i=1} p_i^2},$$

  where p_i is the three-momentum of the i-th particle in the event and α, β are the spatial components. The momenta are the reconstructed momenta of the photon candidates in the event and not the cluster properties.
- **Fox-Wolfram moment ratio** of the second and zero order, defined as $R_2 = \frac{H_2}{H_0}$, where the Fox-Wolfram moments [94] are defined as

  $$H_l = \frac{4\pi}{2l} \sum_{m=-l}^{l} \left| \sum_{i=1}^{N} Y_l^m(\theta_i, \phi_i) \frac{|\mathbf{p}_i|}{\sqrt{s}} \right|^2 = \sum_{i,j=1}^{N} \frac{|\mathbf{p}_i||\mathbf{p}_j|}{s} P_l(\cos\Omega_{ij}),$$

*If not otherwise stated, all features are based on the cluster property.

where Y_l^m are the spherical harmonics, s is the invariant mass of the event and Ω_{ij} is the opening angle between the two particles i and j. The momenta and opening angles are the reconstructed quantities of the photon candidates in the event and not the cluster properties.

- **Absolute three-photon momentum** in the CM system: $|\mathbf{p}^*_{\gamma\gamma\gamma}|$
- **Invariant three-photon mass** of the event: $M_{\gamma\gamma\gamma}$

All the feature distributions are displayed in Section A.1.

I train this selection on ALP masses in the range of $m_a \in [0.1, 10]\,\mathrm{GeV}/c^2$ in steps of $0.1\,\mathrm{GeV}/c^2$. To account for different beam background conditions, for the signal samples, I train on the combination of all runs in Table 3.4. For the background events, I use experiments 22, 24, 25, and 26. To balance the training, I sample 40 % of the signal events and 6.74 % of the background events, which leads to an approximately equal number of signal and background events of around 1×10^6.

For the hyperparameter optimisation, I perform a random grid search with 100 iterations over the predefined hyperparameter space. The search space is spanned by the hyperparameters mentioned in Section 4.4.1. Concrete values for the search are:

- **Number of layers** $N_{\mathrm{layer}} \in [3, 20]$ uniformly sampled,
- **Epochs** $N_{\mathrm{epoch}} \in [200, 500]$ uniformly sampled,
- **Number of maximal nodes** $N^{\mathrm{max}}_{\mathrm{nodes}} \in [2^2, 2^9]$ quantized $\log_2$ uniformly sampled,
- **Batch size** $N_{\mathrm{batch}} \in [2^6, 2^{16}]$ quantized $\log_2$ uniformly sampled,
- **Activation function** choice of `ELU`, `ReLU`, `Tanh`, `Sigmoid`, `LeakyReLU`,
- **Architecture** choice of *Constant*, *Linear Encoder*, *Exponential Encoder*, *Decoder*, *“Bottleneck”*, *“Diamond”*.

The hyperparameter scan shows that the overall training is stable for a wide range of hyperparameters. With batch normalisation and dropout, the training is robust against overfitting, leading to only very few hyperparameter combinations that lead to overfitting. Generally, the final loss is higher for the `Sigmoid` activation function, while the `ELU` activation function leads to the lowest final losses for training. Further observations include the preference for fewer N_{layer} and more $N^{\mathrm{max}}_{\mathrm{nodes}}$ per layer, as well as the preference for the *Linear Encoder* or *Constant* architectures. Judging by the smooth efficiency transitions across different ALP masses, the good training, and the validation loss agreement, the best performing hyperparameters are: $N_{\mathrm{layer}} = 3$; $N_{\mathrm{epoch}} = 338$; $N^{\mathrm{max}}_{\mathrm{nodes}} = 56$; $N_{\mathrm{batch}} = 2278$; `ReLU` activation function; *Constant* architecture.

Figure 4.30 displays the training and validation loss for the best-performing hyperparameters. The training and validation losses show similar development over the epochs. An

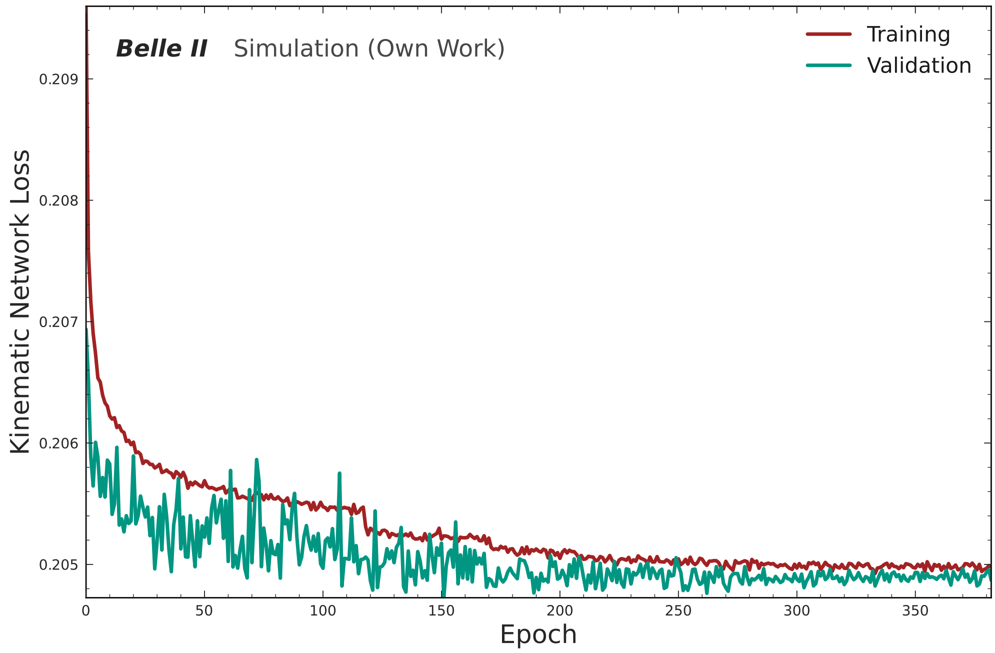


Figure 4.30: Training (red) and validation (green) loss for the kinematic classifier network as a function of the epochs. The hyperparameters used for this particular training are $N_{\text{layer}} = 3$, $N_{\text{epoch}} = 338$, $N_{\text{nodes}}^{\text{max}} = 56$, $N_{\text{batch}} = 2278$, `ReLU` activation function, and *Constant* architecture.

absolute difference in their values can be explained by the dropout, which is only applied during training and not during validation. Sharp drops in the training loss are caused by drops in the learning rates by the learning rate scheduler.

Figure 4.31 shows the distributions of the classifier output $P_{\text{Kin}}^{\text{FNN}}$ for the different background components as well as a few signal samples with different masses. The classifier output signal distributions are scaled to the same yield as the total background yield. From the distributions, I can deduce that very low ALP masses are well separated from the $e^+e^- \to \gamma\gamma(\gamma)$ background, but not as much as compared to events from $e^+e^- \to q\bar{q}$ or $e^+e^- \to h^0\gamma(\gamma)$ processes. For the background from $e^+e^- \to q\bar{q}$ processes, I train a dedicated FNN as explained in the next section. ALP masses in the intermediate range show a slightly less distinguished $P_{\text{Kin}}^{\text{FNN}}$ distribution compared to the lower ALP masses, yet still show a good separation from the background with a tight selection on the classifier output. For the higher ALP masses, the $P_{\text{Kin}}^{\text{FNN}}$ distribution shows less separation from the background, which is expected due to the kinematic similarities of signal and background events in this mass range. Events from high ALP masses tend to be topologically similar to background events from $e^+e^- \to \gamma\gamma(\gamma)$ processes, where most of the energy is used to produce the ALP. Subsequently, the ALP decays almost at rest, which leads to a back-to-back topology of the two photons from the ALP decay. The recoiling photon then has lower energy, which mimics the ISR photon in the $e^+e^- \to \gamma\gamma(\gamma)$ process.

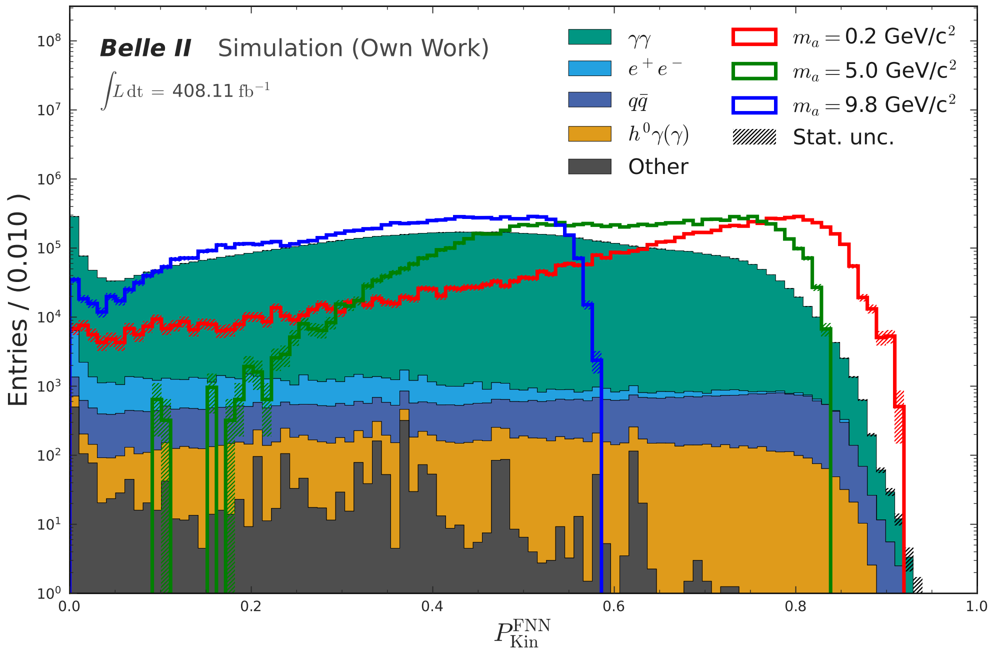
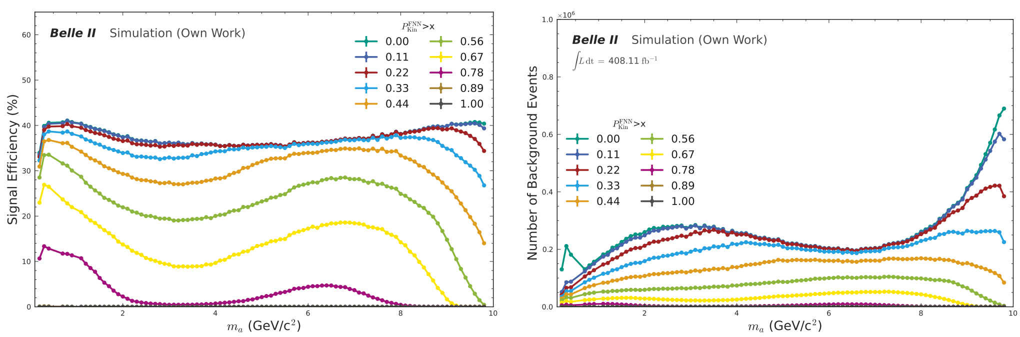


Figure 4.31: Normalised distributions of the classifier output $P_{\mathrm{Kin}}^{\mathrm{FNN}}$ for different background components and signal samples for $m_a = 0.2\,\mathrm{GeV}/c^2$, $m_a = 5.0\,\mathrm{GeV}/c^2$, and $m_a = 9.8\,\mathrm{GeV}/c^2$. The background events used for this plot correspond to the entire Run 1 dataset used in this analysis. For the signal events, all runs from Table 3.4 are used.



(a) Signal efficiency

(b) Number of background events

Figure 4.32: Signal efficiency (left) and number of background events (right) for different selections of $P_{\mathrm{Kin}}^{\mathrm{FNN}}$. The signal efficiency and number of background events are calculated in a $\pm 10\sigma^{\mathrm{CB}}$ window per sample mass on which the training was performed. For the signal events, all runs in Table 3.4 are used. For the number of background events calculation, the Run 1 dataset for $\Upsilon(4S)$ and off-resonance $\Upsilon(4S)$ runs is used. The different colored lines correspond to different $P_{\mathrm{Kin}}^{\mathrm{FNN}}$ selections.

Figure 4.32 shows the efficiency and number of background events for different selections of $P_{\mathrm{Kin}}^{\mathrm{FNN}}$. For both quantities, I calculate the values in a $\pm 10\sigma^{\mathrm{CB}}$ (see Section 6.1) window per sample mass on which the training was performed. It is visible that towards the kinematically allowed $M_{\gamma\gamma}$ range, the signal efficiency and the number of background drops

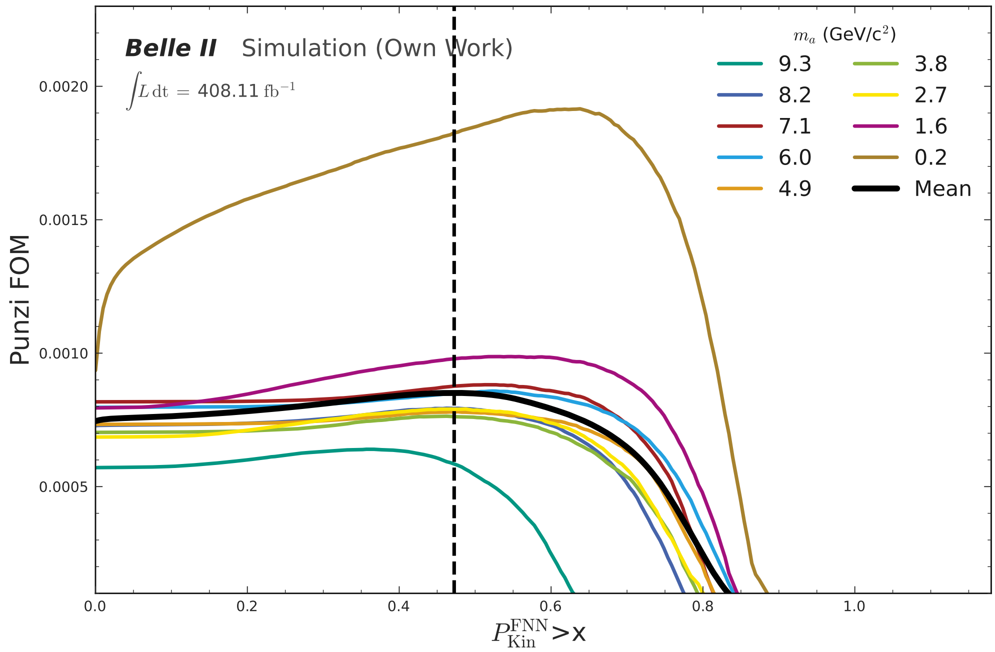


Figure 4.33: PFOM as a function of $P_{\text{Kin}}^{\text{FNN}}$ for few example ALP masses and the mean PFOM across all ALP masses. For the number of background events and the signal efficiencies, the same samples are used as in Fig. 4.32. Differently colored lines correspond to different ALP masses. The black line corresponds to the mean PFOM across all ALP masses, and the dotted line indicates the working point of $P_{\text{Kin}}^{\text{FNN}} > 0.47$.

towards lower values. Since the evaluation on the $m_a = 0.2\,\text{GeV}/c^2$ sample looks in line with the general trend of the higher ALP masses and the fact that I will exclude anything below $m_a \leq 0.17\,\text{GeV}/c^2$, I can conclude that the selection is well-suited for the ALP mass range of interest. For the number of background events, I see the highest impact towards the edges of the scan $M_{\gamma\gamma}$ range, while for the intermediate $M_{\gamma\gamma}$ range, the number of background events is only significantly reduced after a tight selection. Together with the high signal efficiency in this region, I can conclude that for these intermediate ALP masses, the selection is very difficult due to the kinematic similarities to the background, namely three topologically even distributed photons.

To calculate the working point for the selection across all masses, I use the PFOM from Eq. (4.7). I calculate the PFOM for each ALP mass in the range of $m_a \in [0.16, 10]\,\text{GeV}/c^2$ in steps of $0.01\,\text{GeV}/c^2$. Figure 4.33 shows the PFOM as a function of $P_{\text{Kin}}^{\text{FNN}}$ for few example ALP masses as well as the mean PFOM across all ALP masses. The ideal working point for the selection is the one with the highest PFOM. As can be seen for the example ALP masses, the ideal working point is different for each ALP mass. With a single selection across all ALP masses in mind, I choose the mean PFOM maximum, which lies at a working point of $P_{\text{Kin}}^{\text{FNN}} > 0.47$. This selection lies within the plateau of the PFOM for all ALP masses and is therefore a good compromise between the different ALP masses.

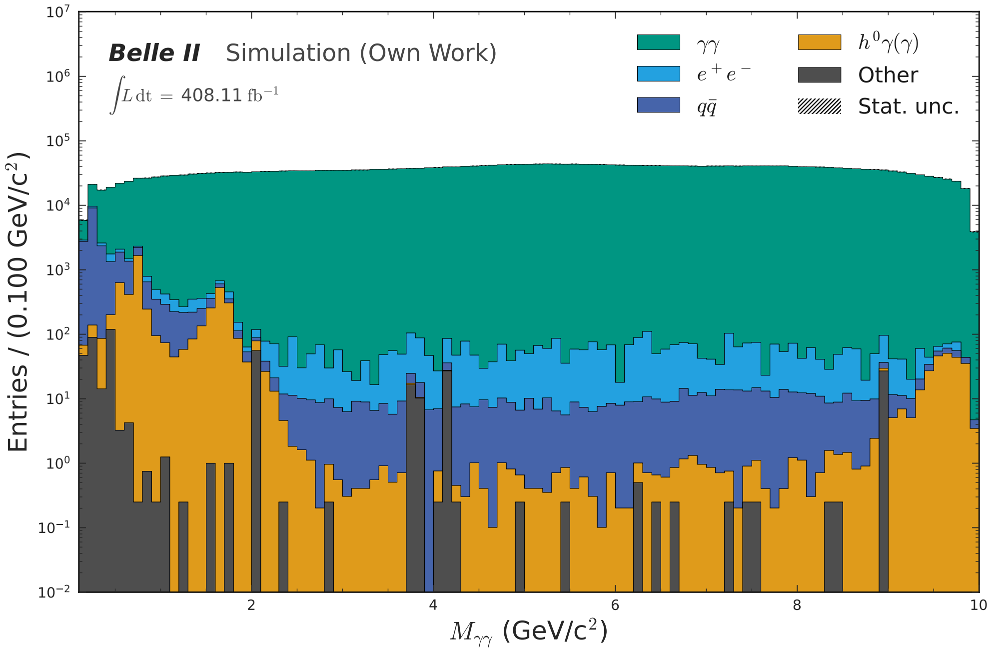


Figure 4.34: Distribution of $M_{\gamma\gamma}$ after applying the final event reconstruction, pre-selection and the kinematic selection with the working point of $P_{\text{Kin}}^{\text{FNN}} > 0.47$.

Finally, Fig. 4.34 displays the total background distribution in $M_{\gamma\gamma}$ for the entire Run 1 dataset after applying the final event reconstruction, pre-selection and the kinematic selection.

4.4.3 Merged Photon Selection

The selection I describe in this section was originally motivated by simulated $e^+e^- \to q\bar{q}$ processes, which produce a large number of merged photon candidates. However, as I demonstrate in the control region studies in Chapter 7, these simulated processes are physically forbidden. Consequently, I exclude all $e^+e^- \to q\bar{q}$ processes from the analysis, rendering this selection redundant, as no improvement in the PFOM is observed once these processes are removed. Therefore, this selection is omitted from the final event selection used for signal extraction. While the motivation for this selection was ultimately shown to be unphysical, this section is included for completeness and to document the performed studies and their implications.

The main reason for including a second MVA based selection is the low ALP mass region. Events from $e^+e^- \to q\bar{q}$ show a significant, yet widely double-peaking distribution slightly above the π^0 mass as displayed in Fig. 4.35. In fact, similar to the discussion in Section 4.2.4, the main background source in this region is the process $e^+e^- \to q\bar{q}$ dominated by the contributions from $u\bar{u}$. The events of interest are those that survive the

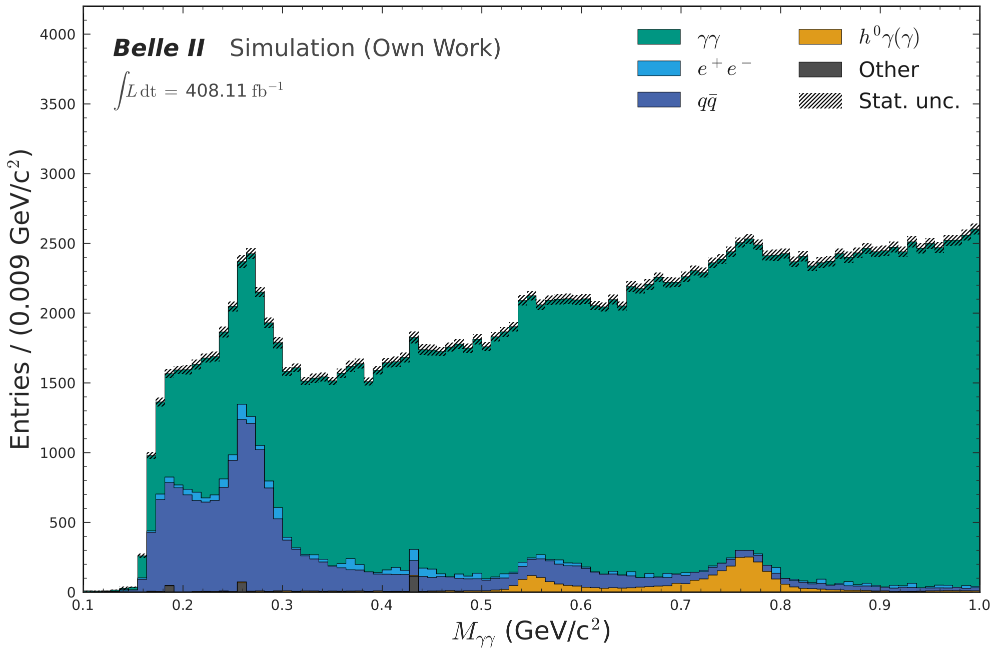


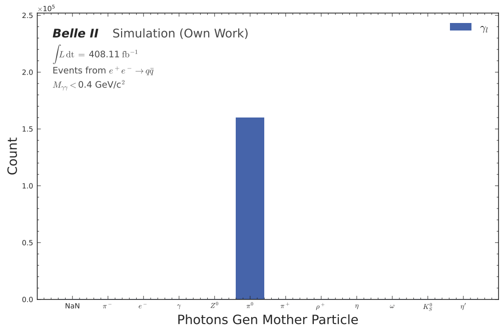
Figure 4.35: Distribution of $M_{\gamma\gamma}$ for $M_{\gamma\gamma} \leq 1.0\,\mathrm{GeV}/c^2$ for the Run 1 dataset of $\Upsilon(4S)$ and off-resonance $\Upsilon(4S)$ runs. The $e^+e^- \to q\bar{q}$ process shows a clear double-peaking structure around $M_{\gamma\gamma} = 0.20\,\mathrm{GeV}/c^2$ and $M_{\gamma\gamma} = 0.25\,\mathrm{GeV}/c^2$.

pre-selection in Section 4.2 (specifically the π^0 veto) and the first MVA-based classification stage from the previous section. In this section, I first discuss the underlying properties of the events of interest and introduce the reconstruction of photon candidates from merged photons. Finally, I present the second MVA selection and the final working point.

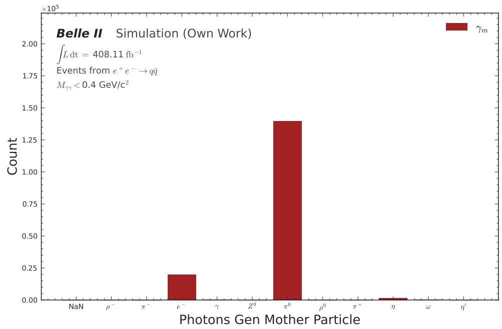
Most of the surviving events originate from the $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$ process, where an additional ISR photon is present in the final state, as can be seen by the distribution of the mother on the generator-level for the three photon candidates in Fig. 4.36. In these plots, I only show the events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$ and the most dominant contributions. Loosening the selection to higher values of $M_{\gamma\gamma}$ shows that the photons also originate from the other neutral hadron processes. Since the following discussion is equivalent for these states, I focus on the $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$ process. Furthermore, most of the higher-energetic photon candidates are matched to originate from an e^-, which is a strong indication that the photon candidate is the ISR photon. The other two photon candidates are matched to originate from a π^0.

As Fig. 4.37 shows, the overall kinematic distribution of the events of interest is dominated by the process where the high-energy ISR photon recoils against the rest of the decay products. The recoil is evident from the strong negative correlation between θ_{γ^h} and θ_{γ^m}, along with a simultaneous positive correlation between θ_{γ^l} and θ_{γ^m}.

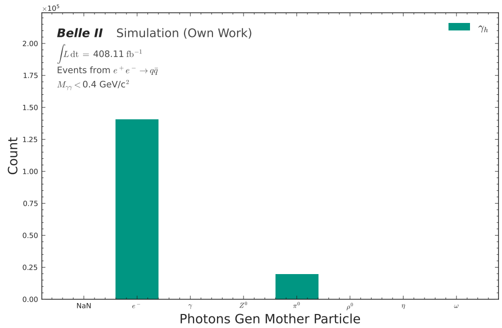
Tables 4.1 and 4.2 summarise all possible final states that could be present in the event after all previous selections. Table 4.1 lists the cases where no merging of two photons is



(a) Low-energetic photon candidates

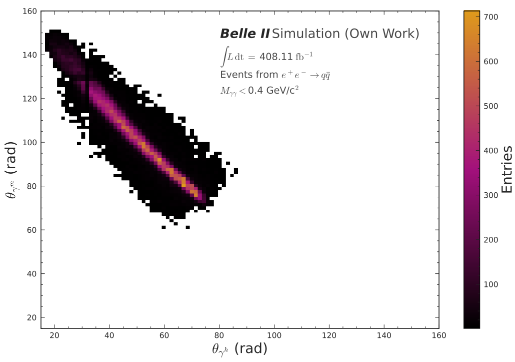


(b) Middle-energetic photon candidates

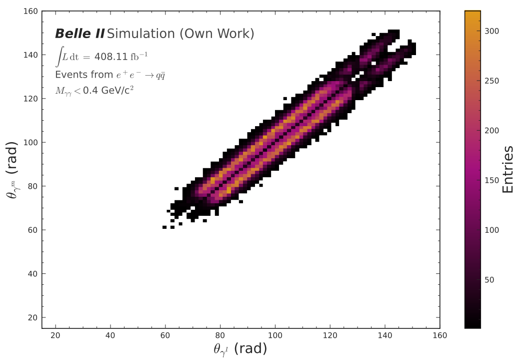


(c) High-energetic photon candidates

Figure 4.36: Distributions of the mother particle code on the generator-level for the three photon candidates for $e^+e^- \to q\bar{q}$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$. In the vast majority of the events, the high-energetic photon is an ISR photon. The middle- and low-energetic photon candidates are mostly matched to the π^0 decay products.



(a) High- and middle-energetic photon candidates



(b) Middle- and low-energetic photon candidates

Figure 4.37: 2D distributions of the photon candidates' polar angles for the $e^+e^- \to q\bar{q}$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$. The high-energetic photon candidate shows to be mostly in the forward direction, while the middle- and low-energetic photon candidates are oriented towards the backwards region. The strong negative correlation between θ_{γ^h} and θ_{γ^m} and the strong correlation between θ_{γ^l} and θ_{γ^m} are a clear indication of the recoil of the high-energetic photon against the rest of the decay products.

Table 4.1: Overview of the possible final states of the $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$ process for the case where no two photons are reconstructed as a merged cluster. Each row represents the decay $e^+e^- \to \gamma^{\mathrm{ISR}}\pi_1^0\pi_2^0$, where the two π^0 photons are denoted as π_1^0 and π_2^0, while the photons originating from these π^0 photons are denoted as $\gamma_1^{\pi^0}$, $\gamma_2^{\pi^0}$, $\gamma_3^{\pi^0}$, and $\gamma_4^{\pi^0}$. Any interchangeability of the indices is implied.

π_1^0	π_2^0	Number of reconstructed photons in the final state
$\gamma_1^{\pi^0} + \gamma_2^{\pi^0}$	$\gamma_3^{\pi^0} + \gamma_4^{\pi^0}$	5
$\gamma_1^{\pi^0} + \gamma_2^{\pi^0}$	$\gamma_3^{\pi^0}$ + Missing $\gamma_4^{\pi^0}$	4
$\gamma_1^{\pi^0} + \gamma_2^{\pi^0}$	Missing $\gamma_3^{\pi^0}$ + Missing $\gamma_4^{\pi^0}$	3
$\gamma_1^{\pi^0}$ + Missing $\gamma_2^{\pi^0}$	$\gamma_3^{\pi^0}$ + Missing $\gamma_4^{\pi^0}$	3

present, while Table 4.2 lists the cases where merging of the photons is present. Merging of photons is defined as the case where two photons are reconstructed as one single photon candidate due to their proximity in the ECL. The tables classify the photon candidates according to the number of true photons in the final state. I differentiate between photons that are correctly reconstructed, those that are present* but not reconstructed (denoted as "missing"), and those that are merged during reconstruction. Furthermore, the tables show only the photon candidates originating from the π^0 decays, while assuming the presence of the recoiling ISR photon. For clarity, I denote the entire process $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$ as $e^+e^- \to \gamma^{\mathrm{ISR}}\pi_1^0\pi_2^0$. Here, the indices have no physical meaning and only serve to distinguish the two π^0 candidates in the final state. Similarly, I denote the photons originating from the first π_1^0 as $\gamma_1^{\pi^0}$ and $\gamma_2^{\pi^0}$, and those from the second π_2^0 as $\gamma_3^{\pi^0}$ and $\gamma_4^{\pi^0}$. Any interchangeability is implied; for example, if I discuss events where one photon is missing, it does not matter which of the two photons is missing, and I mark only one of the four π^0 photons as missing.

All the cases listed in Table 4.1 are less relevant for the low ALP candidate reconstruction since these events most likely do not pass all the previous selections (specifically the $M_{\gamma\gamma\gamma}$ selection). The events from this table that pass the selection can then be further split into categories where more than three photons can be correctly reconstructed or not. In the first two cases in Table 4.1, the fourth most energetic photon could be reconstructed, and the event must pass the $M_{\gamma\gamma\gamma}$ selection for the three highest energetic photons, the π^0 veto, and the E_{γ^4} selection to mimic an ALP event, making these events the least problematic. For the third case, where two photons from the same π^0 are missing, only the E_{γ^4} selection is easily passed. The event still needs to pass the π^0 veto for $\gamma_1^{\pi^0}$ and $\gamma_2^{\pi^0}$ as well as the $M_{\gamma\gamma\gamma}$ selection to mimic an ALP event. Finally, the case where from each π^0 one photon is missing, the π^0 veto is easily passed, contributing a potentially problematic ALP candidate through the $M_{\gamma_1^{\pi^0},\gamma_3^{\pi^0}}$ invariant mass.

Reconstructed events with merged photons are potentially more problematic for the

*These photons are either outside of the acceptance or are not matched to a photon on generator-level.

Table 4.2: Overview of the possible final states of the $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$ process for the case where two photons are reconstructed as a merged cluster. Each row represents the decay $e^+e^- \to \gamma^{\mathrm{ISR}}\pi^0_1\pi^0_2$, where the two π^0 photons are denoted as π^0_1 and π^0_2, while the photons originating from these π^0 photons are denoted as $\gamma_1^{\pi^0}$, $\gamma_2^{\pi^0}$, $\gamma_3^{\pi^0}$, and $\gamma_4^{\pi^0}$. Any interchangeability of the indices is implied.

π^0_1	π^0_2	Number of reconstructed photons in the final state
Merged $\gamma_1^{\pi^0} + \gamma_2^{\pi^0}$	$\gamma_3^{\pi^0} + \gamma_4^{\pi^0}$	4
Merged $\gamma_1^{\pi^0} + \gamma_2^{\pi^0}$	$\gamma_3^{\pi^0}$ + Missing $\gamma_4^{\pi^0}$	3
Merged $\gamma_1^{\pi^0} + \gamma_2^{\pi^0}$	Merged $\gamma_3^{\pi^0} + \gamma_4^{\pi^0}$	3

low ALP candidate reconstruction. Invariant diphoton masses from merged photons can mimic the signal by presenting the events as if they originated from slightly higher energetic photons. Table 4.2 summarises all possible cases where merging of photons is present. While the first case, one reconstructed merged photon and two well-resolved π^0 photons, is again less problematic, since these events are potentially removed by the previous selections (specifically the E_{γ^4} or $M_{\gamma\gamma\gamma}$ selections or the π^0 veto), the second and third cases are more problematic. In the second case, where two photons are again merged but one of the other π^0 photons is missing, only the $M_{\gamma\gamma\gamma}$ selection can potentially remove the event. However, if the event passes all previous selections, the $M_{\gamma_{1,2}^{\pi^0},\gamma_3^{\pi^0}}$ invariant mass mimics an ALP candidate. Finally, the last case, where all four photons are merged, is the most problematic since the $M_{\gamma_{1,2}^{\pi^0},\gamma_{3,4}^{\pi^0}}$ invariant mass mimics an ALP candidate and no previous selection can remove the event.

To briefly summarise the discussion so far: events from the $e^+e^- \to q\bar{q} \to \pi^0\pi^0\gamma^{\mathrm{ISR}}$ process that mimic low ALP mass candidates fall into two categories. The first consists of events that pass the π^0 veto and the E_{γ^4} selection and satisfy the $M_{\gamma\gamma\gamma}$ selection due to specific kinematic configurations, such as low-energy missing photons. Another possibility is events with merged π^0 photons, which naturally pass the selections. The first type of events poses a problem because tightening these selection criteria would significantly reduce the signal efficiency. However, as I discuss later, these events do not exhibit a pronounced peak in the $M_{\gamma\gamma}$ distribution, aside from possible distortions due to the kinematic fit. The second type of events is more challenging because they can form a potential ALP candidate, which peaks around $M_{\gamma\gamma} \approx 0.25\,\mathrm{GeV}/c^2$. Nevertheless, the contribution from merged photons can be suppressed by applying selections based on ECL shower shape observables, which will be discussed in the following.

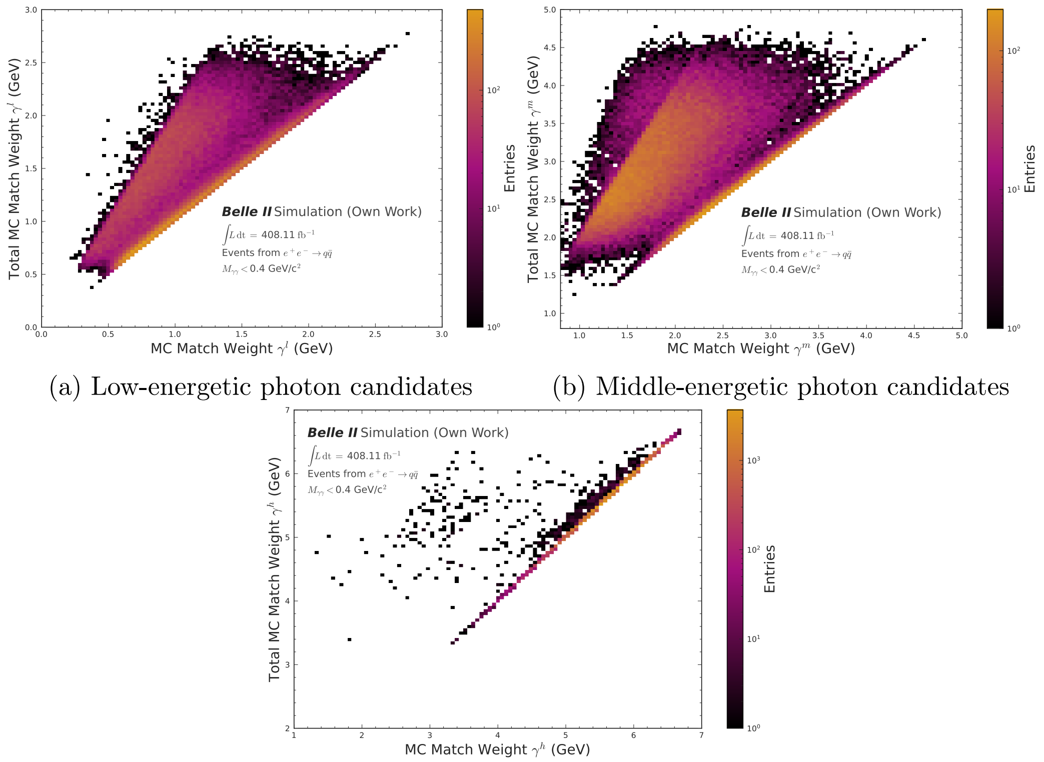
To build a classifier which can separate merged photons, I first define a quantity which separates merged photon candidates from well-resolved photons. For this, I make use of the so-called *energy dependent MC matching* in basf2. To briefly summarise, the matching method builds a relation between MC particles to the ECL clusters. Each cluster is related to an ECL shower, which in turn has a weight assigned to the corresponding ECL crystals. One crystal can be related to multiple showers through a fractional weight. Furthermore, each crystal has then an energy-weighted relation to one or more MC particles. The so-called *MC Match Weight* is then the weighted relation of the cluster to none, or multiple MC particles. If one cluster is related to more than one MC particle, the *MC Match Weight* is the highest weight. In the case of a good isolated photon, the *MC Match Weight* is close to the energy of the photon and the total *MC Match Weight*, which is the sum of all weights of the corresponding cluster. If multiple MC particles are related to the cluster, the *MC Match Weight* is lower than the energy of the photon and the total *MC Match Weight*.

Figure 4.38c displays the correlation of the *MC Match Weight* and the total *MC Match Weight* for the higher-energetic photon candidates in the events of interest. The strong correlation of the two variables shows that most of the photon candidates are well isolated, as expected by the kinematic configuration discussed earlier. For the middle- and the low-energetic photon candidates, the correlation is different and split in two different regions as can be seen in Figs. 4.38a and 4.38b. For ratios of the *MC Match Weight* and the total weight above 0.9, the photon candidates are well isolated, similar to the case of the high-energetic photon candidates. Below ratios of 0.9, the weight assigned to the MC particle is less than the total weight assigned to all MC particles related to the cluster, indicating that the photon candidates are either merged or not well isolated.

Figure 4.39 summarizes the previous discussion by displaying the *MC Match Weight* ratios for the middle- and low-energetic photon candidates. In the plot, two lines separate the events into four regions. For the lower left quadrant, where both the *MC Match Weight* ratios are below 0.9, the photon candidates are most probably merged photons. These events build up the majority of the background contribution in the low ALP mass region. The upper left and lower right quadrants are the regions where one of the two photon candidates is merged. Quantitatively, the middle-energetic photon candidate consists of merged photons about as often as both photon candidates. The case where only the low-energetic photon candidate is from merged photons is about a quarter as frequent as the case for only the middle-energetic photon candidate. Lastly, the upper right quadrant is the region where both photon candidates are well isolated. These events, as discussed earlier, are difficult to distinguish from signal events, but are flat in $M_{\gamma\gamma}$.

For the classification of well-isolated photons *vs.* photon candidates from merged photons, I make use of the shower shape variables [95]. Overlapping electromagnetic showers can be distinguished to a certain degree by these variables. Figure 4.40 shows the distribution of one of the Zernike moments $|Z_{40}|$ for the middle-energetic photon candidates in the events of interest. Events with a weight ratio below 0.9 are labelled as merged photons; otherwise, they are labelled as not merged. The distribution shows a clear separation of the two classes; however, the histogram is normalised to unity and does not display the actual proportions.

As clearly visible in the distribution of $|Z_{40}|$, a simple selection on one value will always



(a) Low-energetic photon candidates (b) Middle-energetic photon candidates

(c) High-energetic photon candidates

Figure 4.38: 2D distributions of the *MC Match Weight* and the total *MC Match Weight* for the three photon candidates in the $e^+e^- \to q\bar{q}$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$. The high-energetic photon displays an *MC Match Weight* ratio of close to one, indicating that the photon candidate is well isolated. The middle- and low-energetic photon candidates show enhanced regions where the ratio between *MC Match Weight* and the total *MC Match Weight* is below one, indicating that the photon candidates are either merged or not well isolated.

lead to a significant loss of signal efficiency. Even when combining different shower shape variables and selecting in a 2D space, the loss of signal efficiency is still significant. Therefore, I use an MVA classifier which can separate the two classes.

The features that are presented to the network are the following:

- **Cluster Lateral Energy Distribution** of the photons: $S^{\mathrm{LAT}}_{\gamma^m}$, $S^{\mathrm{LAT}}_{\gamma^l}$. The lateral energy distribution S^{LAT} is defined as

$$S^{\mathrm{LAT}} = \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^{n} w_i E_i r_i^2}, \tag{4.9}$$

  where E_i denotes the crystal energies sorted ascending by energy, w_i the shower to crystal weights, and r_i the distance of the crystal to the shower centre projected to a plane perpendicular to the shower axis.

- **Cluster Second Moment** of the photons: $S^{\mathrm{2nd}}_{\gamma^m}$, $S^{\mathrm{2nd}}_{\gamma^l}$. The second moment is

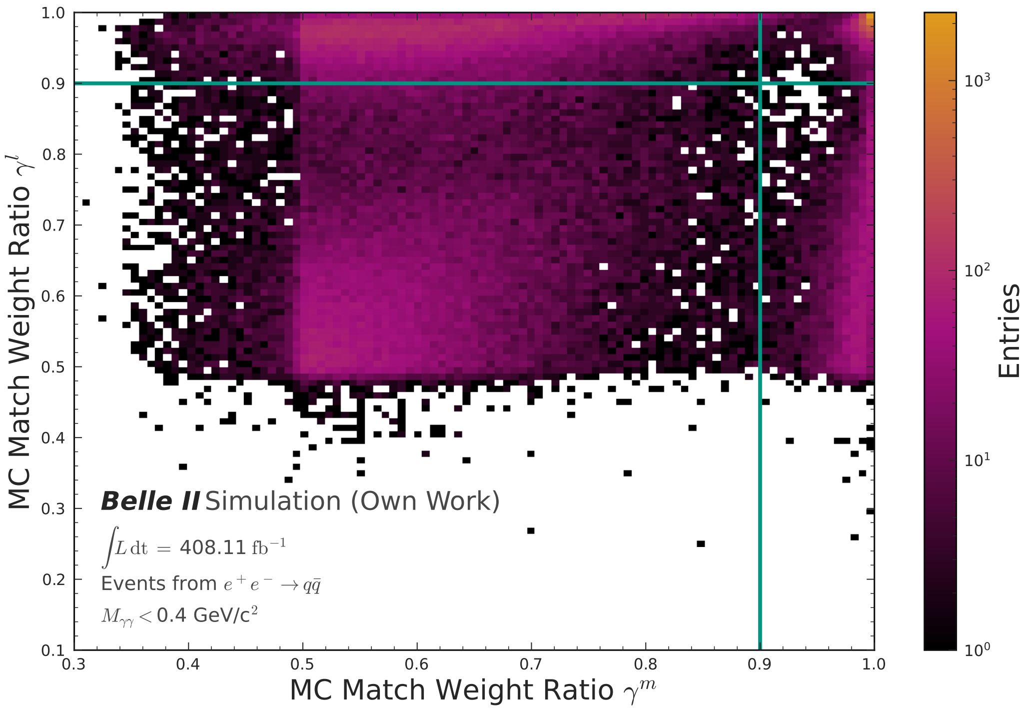


Figure 4.39: 2D distribution of the *MC Match Weight* ratios for the middle- and low-energetic photon candidates in the $e^+e^- \to q\bar{q}$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$. Green lines indicate the split into four quadrants. In the case that both ratios are above 0.9, the middle- and low-energetic photon candidates are well isolated as represented by the upper right quadrant. The other quadrants represent different configurations of merged and well-isolated photon candidates.

defined as

$$S^{2\mathrm{nd}} = \frac{1}{S_0(\theta)} \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{\sum_{i=2}^{n} w_i E_i}, \tag{4.10}$$

where $S_0(\theta)$ is the normalization factor that normalizes $S^{2\mathrm{nd}}$ to one for isolated photons.

- **Cluster Neighbour Energy Ratio** where the ratio is built between the central crystal and its neighbours in a 3×3 grid of the photons: $E^1/E^9_{\gamma^m}$, $E^1/E^9_{\gamma^l}$.
- **Cluster Zernike Moments** of the photons: $|Z_{40}|_{\gamma^m}$, $|Z_{51}|_{\gamma^m}$, $|Z_{40}|_{\gamma^l}$, $|Z_{51}|_{\gamma^l}$. The Zernike moments are defined as

$$|Z_{nm}| = \frac{n+1}{\pi} \frac{1}{\sum_i w_i E_i} \left| \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} w_i E_i \right|, \tag{4.11}$$

where i runs over all crystals associated with the shower, $R_{nm}(\rho_i)$ is the Zernike polynomial of order n, ρ_i is the radial distance of the crystal to the shower centre projected to a plane perpendicular to the shower axis, and α_i is its polar angle.

All the feature distributions are displayed in Section A.2.

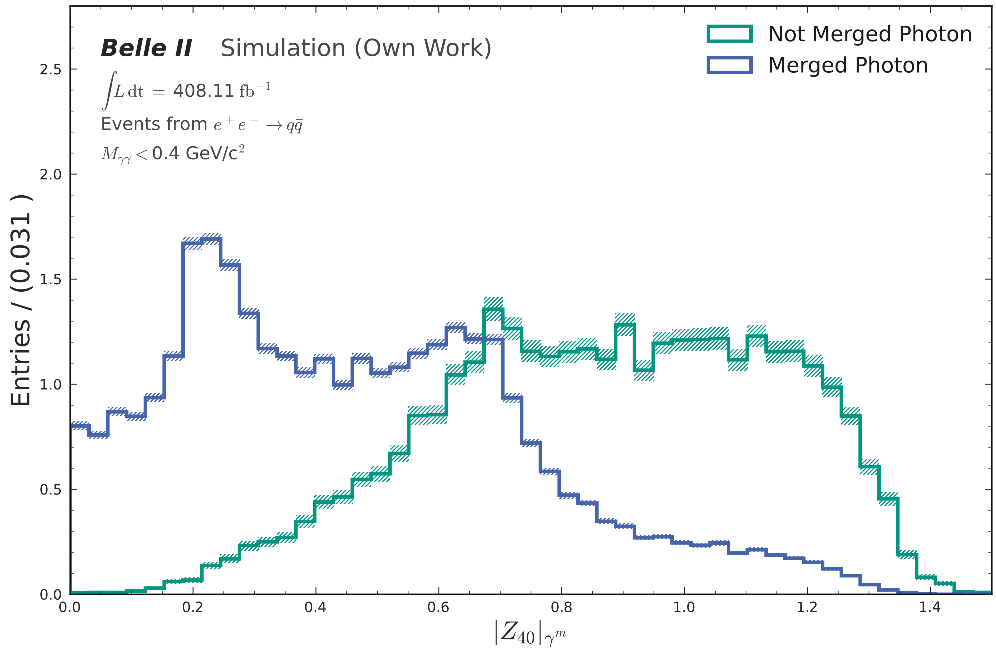


Figure 4.40: Distribution of $|Z_{40}|$ for the middle-energetic photon candidates in the $e^+e^- \to q\bar{q}$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$. The events are split into two classes, one with a *MC Match Weight* ratio below 0.9 (blue), indicating that the photon candidate is from merged photons, and one with a ratio above 0.9 (green), indicating that the photon candidate is well isolated.

I train this selection on ALP masses in the range of $m_a \in [0.1, 1]\,\mathrm{GeV}/c^2$ in steps of $0.01\,\mathrm{GeV}/c^2$ and in the range of $m_a \in [1, 10]\,\mathrm{GeV}/c^2$ in steps of $0.1\,\mathrm{GeV}/c^2$. To account for different beam background conditions, I train on the combination of all runs in Table 3.4. Additionally, as I describe in Section 4.4.1, I adjust the signal event weights with $\alpha = 0.4$ to shift the attention of the network towards lower ALP masses. For the background events, I only use $e^+e^- \to q\bar{q}$ events in which at least one of the middle- or low-energetic photon candidates has an *MC Match Weight* below 0.9 (see Fig. 4.39 for illustration of this selection). To balance the training, I sample only 3.77 % of the signal events, resulting in approximately 150,000 signal and background events each.

I perform the hyperparameter optimisation in the same way as for the first MVA classifier in Section 4.4.2. Overall, the grid search results are similar to the first stage. Furthermore, a low number of hidden layers and a higher number of neurons per layer show positive correlations with good training results. The network with the lowest final loss, and no significant divergence of validation and training loss, has the following hyperparameters: $N_{\mathrm{layer}} = 4$; $N_{\mathrm{epoch}} = 232$; $N_{\mathrm{nodes}}^{\mathrm{max}} = 110$; $N_{\mathrm{batch}} = 4468$; `ReLU` activation function; *Linear* architecture.

Figure 4.41 displays the training and validation loss for the best-performing hyperparameters. Similar to Fig. 4.30, the training shows good convergence behaviour, with no

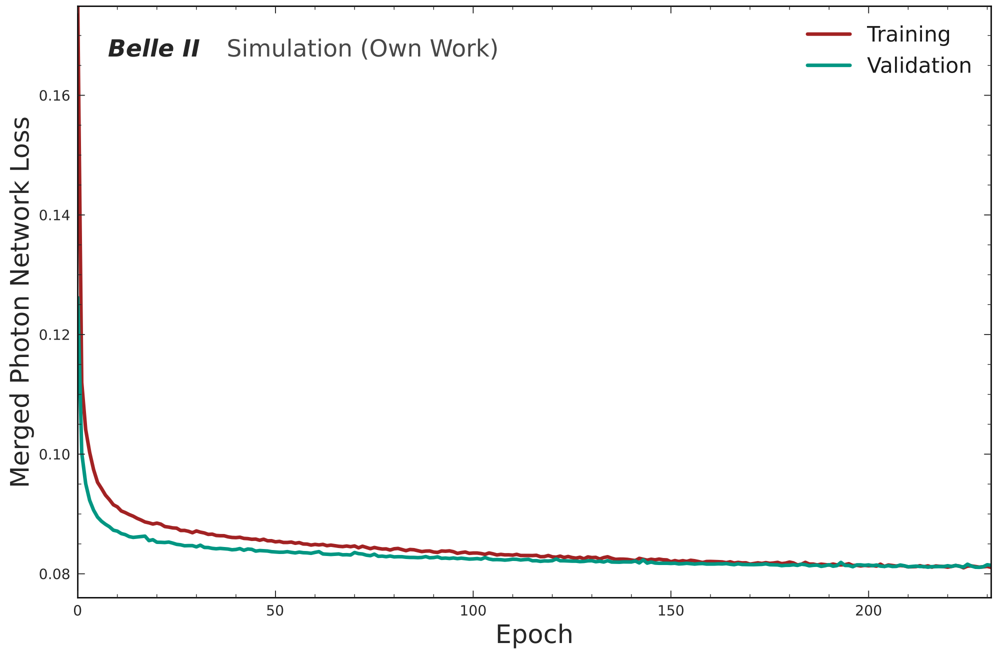


Figure 4.41: Training (red) and validation (green) loss for the kinematic classifier network as a function of the epochs. The hyperparameters used for this particular training are $N_{\text{layer}} = 4$; $N_{\text{epoch}} = 232$; $N^{\text{max}}_{\text{nodes}} = 110$; $N_{\text{batch}} = 4468$; `ReLU` activation function, and *Linear* architecture.

significant divergence between training and validation loss.

For the evaluation of the classifier, I rely on the same event type as the training has seen, namely the events where either the middle- or low-energetic photon candidates show signs of originating from merged photons. Figure 4.42 shows the normalised distributions of the classifier output for the different background components, as well as a few signal samples with different masses.

Figure 4.43 shows the efficiency and number of background events for different selections of $P^{\text{FNN}}_{\text{MP}}$. For both quantities, I calculate the values in the same way as described in Section 4.4.2, using a $\pm 10\sigma^{\text{CB}}$ (see Section 6.1) window per sample mass on which the training was performed. For the classifier selection–dependent number of background events shown in Fig. 4.43b, I observe the expected effect consistent with the goal of the selection: most of the events rejected by tightening the selection originate from the very low ALP mass region. The signal efficiency in Fig. 4.43a shows a rather constant efficiency loss with tightening of the selection. This behaviour is expected since the signal events do not contain any additional photons that could result in a photon candidate from merged photons.

For the working point determination, I rely again on the PFOM. Figure 4.44 displays the PFOM as a function of the $P^{\text{FNN}}_{\text{MP}}$ selection for some example ALP masses and the mean over all PFOM values for each simulated mass. Since the main problem is a signal mimicking structure in the low $M_{\gamma\gamma}$ region, I calculate a weighted mean of the PFOM to

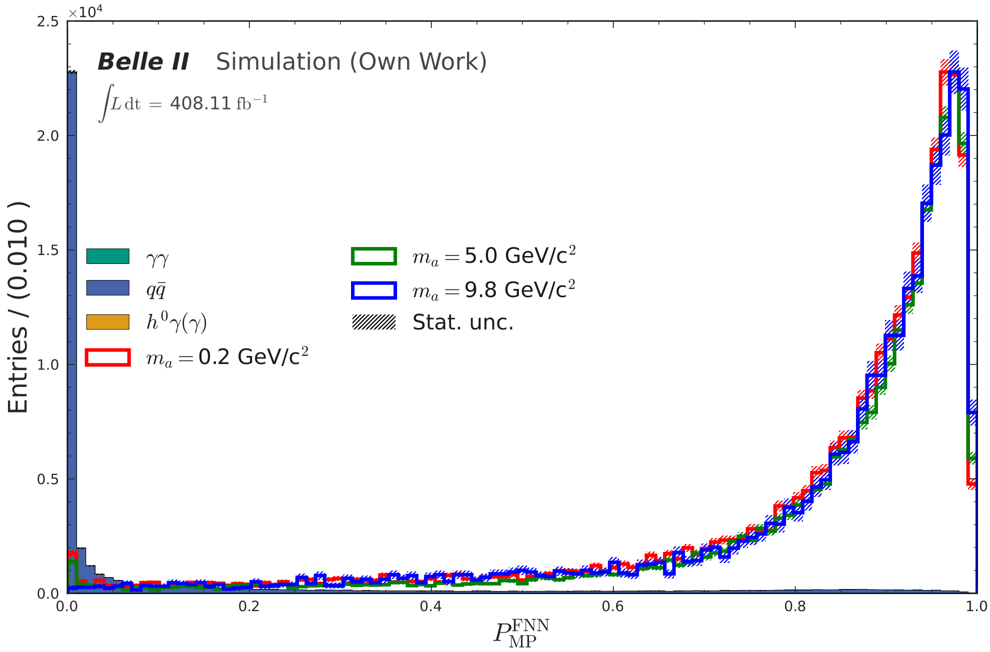
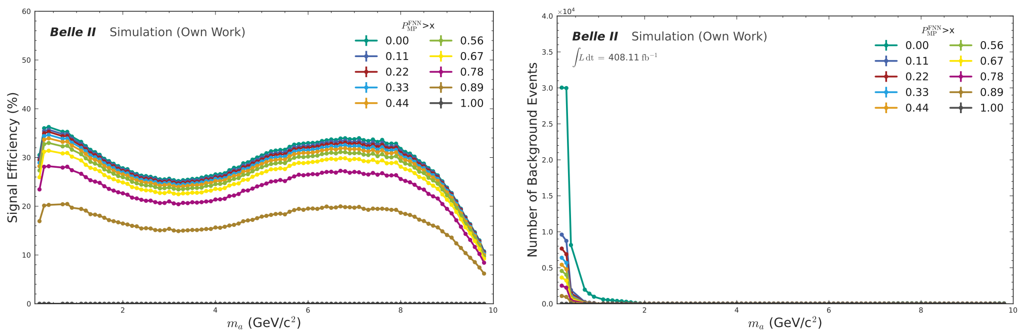


Figure 4.42: Normalised distributions of the classifier output for different background components and signal samples for $m_a = 0.2\,\mathrm{GeV}/c^2$, $m_a = 5.0\,\mathrm{GeV}/c^2$, and $m_a = 9.8\,\mathrm{GeV}/c^2$. The background events used for this plot only show $e^+e^- \to u\bar{u}$ events where either the middle- or low-energetic photon candidates show signs of originating from merged photons – the same events that were used for the training.



(a) Signal efficiency (b) Number of background events

Figure 4.43: Signal efficiency (left) and number of background events (right) for different selections of $P_{\mathrm{MP}}^{\mathrm{FNN}}$. The signal efficiency and number of background events are calculated in a $\pm 10\sigma^{\mathrm{CB}}$ window per sample mass on which the training was performed. For the signal events, all runs in Table 3.4 are used. For the number of background events calculation, the Run 1 dataset for $\Upsilon(4S)$ and off-resonance $\Upsilon(4S)$ runs is used. However, only $e^+e^- \to q\bar{q}$ events where either the middle- or low-energetic photon candidates show signs of originating from merged photons are used. The different colored lines correspond to different $P_{\mathrm{MP}}^{\mathrm{FNN}}$ selections.

determine the working point. For each ALP mass' PFOM value, I apply the same weight which I apply during the training $\alpha = 0.4$. In contrast to the working point determination of

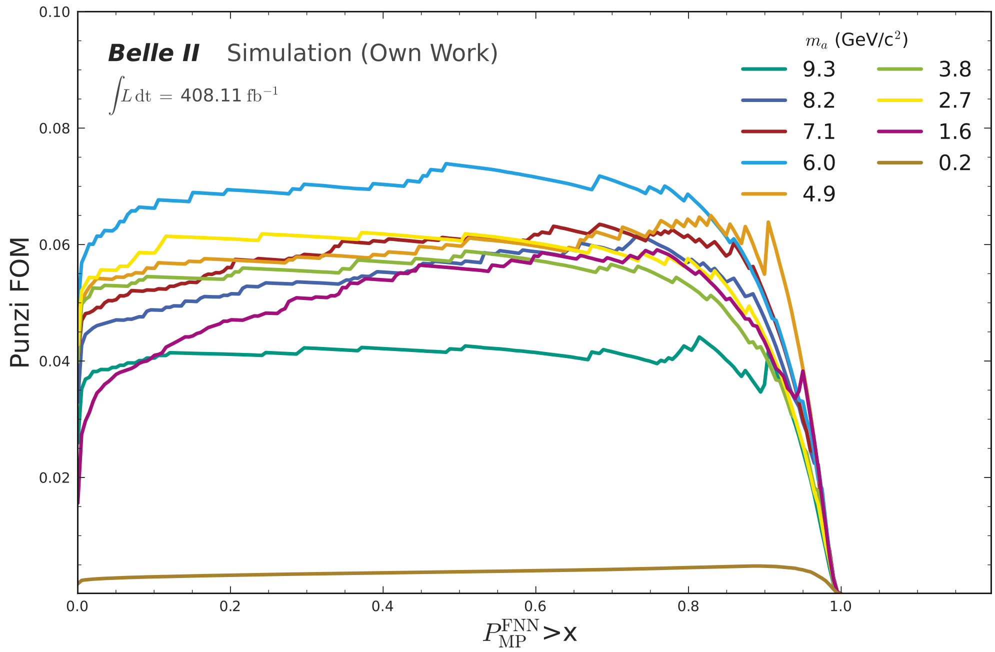
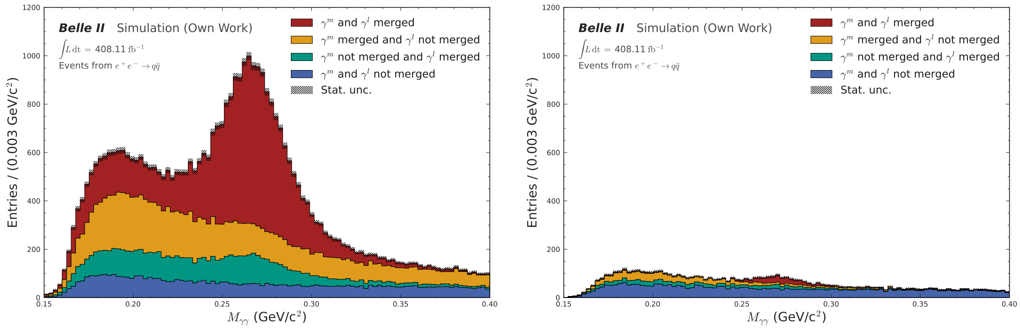


Figure 4.44: PFOM as a function of $P_{\rm MP}^{\rm FNN}$ for few example ALP masses and the mean PFOM across all ALP masses. For the number of background events and the signal efficiencies, the same samples are used as in Figure 4.43. Differently colored lines correspond to different ALP masses.

the first stage, I do not see clear maxima of the PFOM for different masses. The maximum of the weighted mean over all ALP masses suggests using $P_{\rm MP}^{\rm FNN} > 0.65$. Nevertheless, to reduce the seemingly peaking contribution, specifically around $M_{\gamma\gamma} \approx 0.27\,\mathrm{GeV}/c^2$, I choose a tighter working point of $P_{\rm MP}^{\rm FNN} > 0.80$.

With the application of this selection, I manage to remove the majority of the events that are discussed in this section. Figure 4.45 compares the $M_{\gamma\gamma}$ distribution of the $e^+e^- \to q\bar{q}$ events before and after the selection, divided into the groups where either one or both or none of the middle- or low-energetic photon candidates are made up by possibly merged photons. While the events with merged photons are now subdominant compared to the events with well-isolated photons, the events with merged photons still contribute to the low $M_{\gamma\gamma}$ region with peaking substructures. However, as Fig. 4.46 shows, with the help of the selection, I can suppress the contribution from merged photons far enough to be dominated by the $e^+e^- \to \gamma\gamma(\gamma)$ events.



(a) Pre $P_{\mathrm{MP}}^{\mathrm{FNN}}$ selection (b) Post $P_{\mathrm{MP}}^{\mathrm{FNN}}$ selection

Figure 4.45: Distribution of $M_{\gamma\gamma}$ for the $e^+e^- \to u\bar{u}$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$ before (left) and after (right) the merged photon selection. The events are split into four categories: events where both the middle- and low-energetic photon candidates are well isolated (blue), events where only one of the two photon candidates is from merged photons (yellow for γ^m and green for γ^l), and events where both photon candidates are from merged photons (red).

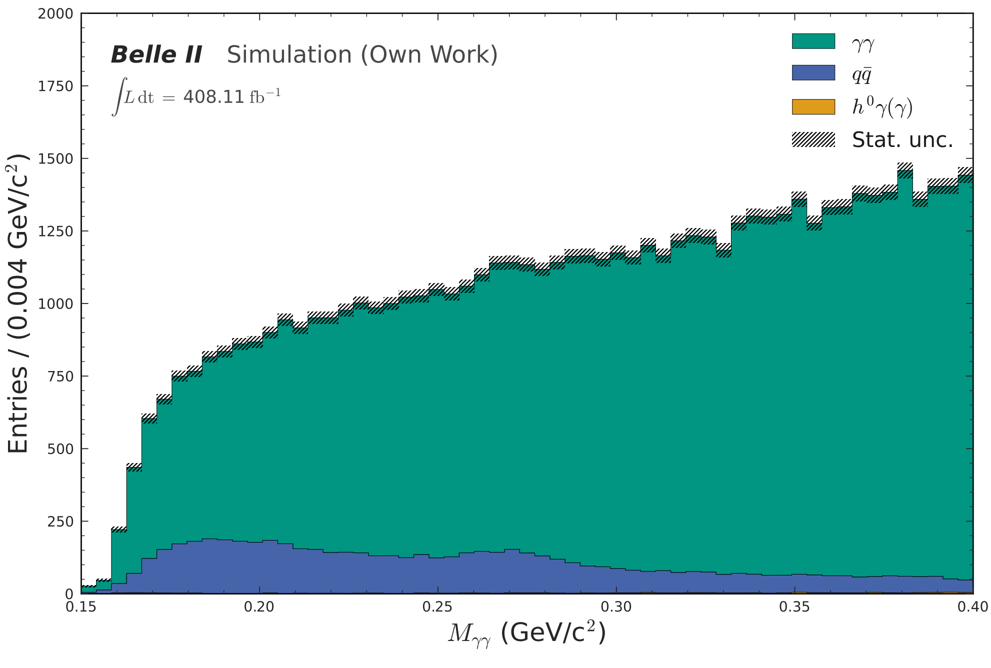


Figure 4.46: Distribution of $M_{\gamma\gamma}$ for the $e^+e^- \to \gamma\gamma(\gamma)$, $e^+e^- \to u\bar{u}$ and $e^+e^- \to h^0\gamma(\gamma)$ events with $M_{\gamma\gamma} \leq 0.4\,\mathrm{GeV}/c^2$ after the merged photon selection.

4.5 Summary

Table 4.3 summarises all the selections that I apply to the reconstructed events. As I mentioned in **??**, the $e^+e^- \to q\bar{q}$ contribution is excluded from consideration, and the merged photon classifier selection is omitted.

Figure 4.47 displays the distribution of $M_{\gamma\gamma}$ for background events passing selections in

Table 4.3: Summary of event reconstruction, event and candidate selection, and MVA selection criteria.

Category	**Description**	**Selection**
Candidate	Photon Candidate	· $E_\gamma > 500\,\mathrm{MeV}$ · $17^\circ < \theta_\gamma < 150^\circ$ · $\lvert T_{\mathrm{Cluster}}\rvert < 200\,\mathrm{ns}$ · `clusterErrorTiming` $< 10^6$ ns
	Event Candidate	· Three highest energetic photons in the event · $80\,\% \cdot \sqrt{s} < M_{\gamma\gamma\gamma} < 105\,\% \cdot \sqrt{s}$ · Simultaneity of the three-photon candidates
	ALP Candidate	· $\lvert\cos\lambda\rvert < 0.74$
Event	Number of Clean Tracks	· $N_{\mathrm{C.T.}} < 2$
	Additional Photons	· $E_{\gamma^4} < 0.5\,\mathrm{GeV}$
	Pair Conversion Veto	· $\Delta\theta_{h,l} > 0.014$ rad · $\Delta\phi_{h,l} > 0.4$ rad · $\Delta\theta_{m,l} > 0.014$ rad · $\Delta\phi_{m,l} > 0.4$ rad
	π^0 veto	· $M_{\gamma^i,\gamma^j} > 0.16\,\mathrm{GeV}/c^2$ for $i,j \in [h,m,l,4]$
MVA	Kinematic Classifier	· $P^{\mathrm{FNN}}_{\mathrm{Kin}} > 0.47$

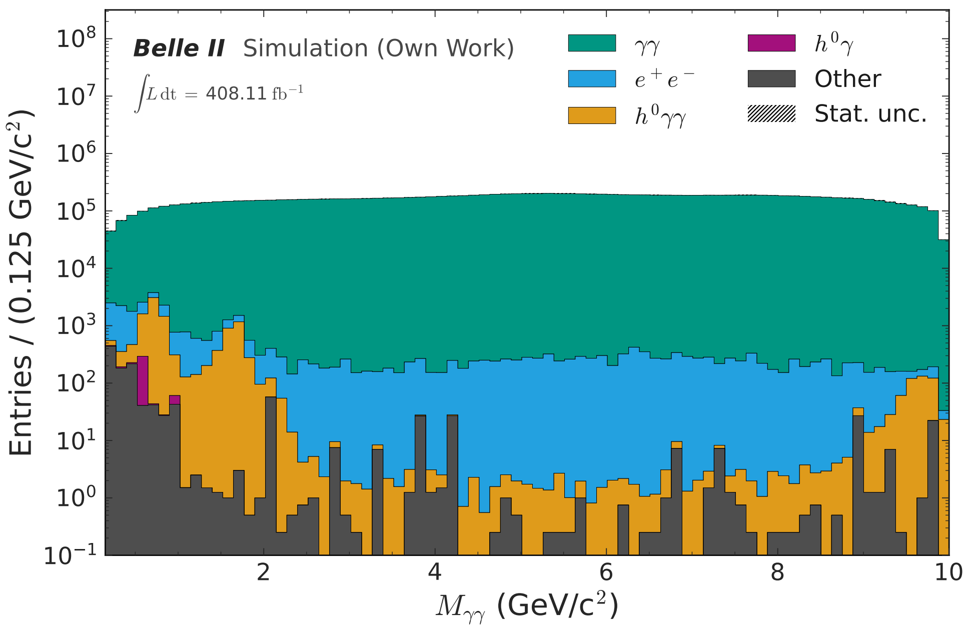


Figure 4.47: Distribution of $M_{\gamma\gamma}$ after applying the final selection criteria.

Table 4.3 for the entire Run 1 dataset.

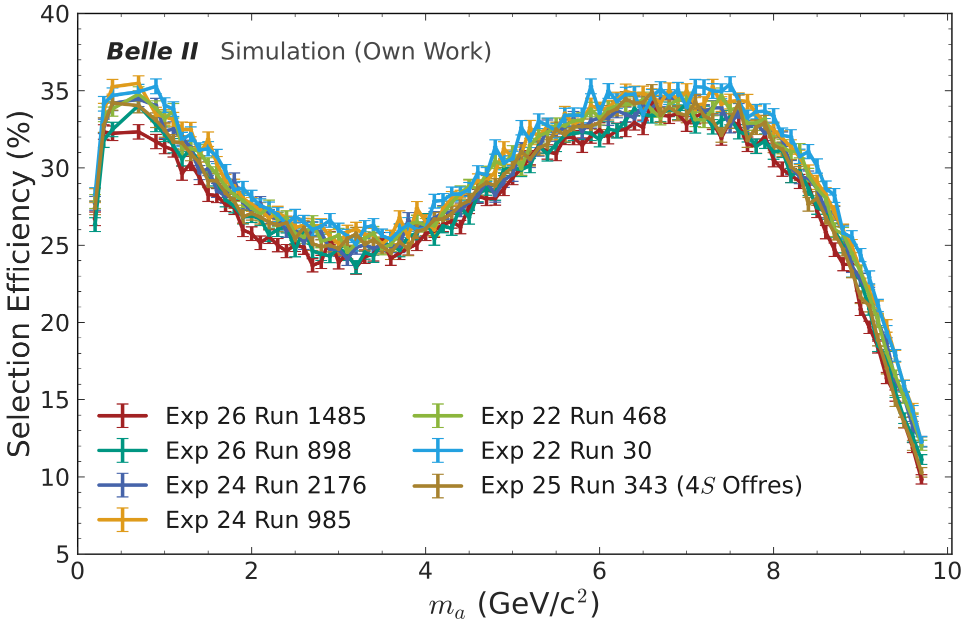


Figure 4.48: Signal efficiency for the ALP mass points after all selections for several runs with different beam background conditions. The efficiency is shown in different colours for simulations with different beam background conditions.

Finally, Figure 4.48 shows the signal efficiency for the ALP mass points after the same selections for several runs with different beam background conditions according to Table 3.4.

Chapter 5

Signal Efficiency

In this chapter, I discuss the signal efficiency of the analysis after the application of the event reconstruction and selection from Chapter 4. Section 5.1 describes first the L1 trigger bits used in this analysis and evaluates their trigger efficiency. Furthermore, I present the method to measure the trigger efficiency and assert the systematic uncertainty. Lastly, I briefly discuss the HLT selection efficiency. Section 5.2 discusses the signal efficiency as a function of beam background conditions. Finally, I present an interpolation method to obtain the signal efficiency for the entire Run 1 dataset. All simulated events in this section are additionally weighted using the MC/data photon efficiency correction, as I describe in Section 8.1.

5.1 Trigger Efficiency

During the data taking of Run 1, the Belle II detector was operated with a fully enabled trigger system. Given the high cross section of Bhabha scattering, most trigger lines contributing to the L1 trigger decision include a so-called Bhabha veto. This veto rejects any events that share the same signature as Bhabha events based purely on the information of the ECL trigger. I provide a more detailed description of the Bhabha veto in the following. For very light or heavy ALPs, the event topology may be extremely similar to Bhabha scattering events – high energetic back-to-back clusters in the ECL. With this in mind, a specific trigger line was implemented to keep events where no tracks are in the vicinity of the ECL clusters: `ggsel`. Furthermore, event topologies that pass the Bhabha veto can fully rely on the other ECL trigger lines. In this section, I discuss the overall trigger efficiency of this analysis for the events passing the total trigger selection. First, I define an ALP mass-independent combination of L1 trigger lines and outline the differences between simulation and data. I then present the method to validate the trigger efficiency and assert a systematic uncertainty. Finally, I briefly discuss the HLT efficiency.

5.1.1 L1 Trigger

The event topology plays a crucial role not only in the selection but also in trigger decision. On the one hand, very light ALPs can result in an event topology where the ALP and the recoil photon are produced almost back-to-back in the lab frame. With the high boost of the ALP, the photons produced by the ALP decay are colinear in the lab frame. Subsequently, the probability is very high that the ALP photons are not reconstructed as two separate clusters due to the reduced granularity of the ECL trigger cells. On the other hand, very heavy ALPs will take up most of the energy-momentum of the beam; hence, the ALP decays almost at rest in the lab frame. This leads to a similar event topology, where two high-energy clusters are produced in the ECL, which are back-to-back in the CM frame, with the addition of a low-energy cluster from the recoiling photon. In both cases, the event topology is very similar to Bhabha scattering events, and therefore, the Bhabha veto is triggered. For all other ALP masses in between, the event topology is more isotropic; hence, the Bhabha veto is not triggered, and other ECL based trigger lines can be used.

With the event topology in mind, to trigger on both classes of events, the ideal trigger line would be the combination of `ggsel` for the Bhabha-like events and the high-energy cluster trigger line `hie`. However, since `ggsel` is a relatively new trigger line, it is only available for the data taken in experiment 14 and later. To ensure full Run 1 coverage, I require the 3D Bhabha veto, represented by the trigger line `bha3d`, to be triggered alongside `hie` and `ggsel`. The resulting combination of L1 trigger lines for this analysis is therefore defined as: `hie|bha3d|ggsel`. For experiments before 14, the trigger line is reduced to `hie|bha3d` due to the unavailability of `ggsel`. For experiments after 14, the trigger line `bha3d` is kept for the trigger decision; however, the trigger line is prescaled* by the L1 trigger implementation for most experiments. Shortly summarising, the trigger lines that I use in this analysis are defined as:

- `hie`: Total ECL energy greater than 1 GeV in the ECL barrel and part of the forward endcap (`ehigh`) and the 3D Bhabha veto (`!ecl_3dbha`),
- `ggsel`: 3D Bhabha-like event with back-to-back clusters in the ECL with both cluster energies above 3 GeV and one of these clusters above 4.5 GeV (`ecl_3dbha`). Additionally, not exactly one and not two tracks in the vicinity of the clusters (`!trkbha1` & `!trkbha2`) and at least one of the clusters in the CDC barrel acceptance (`bha_intrk`),
- `bha3d`: 3D Bhabha-like event, as described above. (`ecl_3dbha`)

Table B.1 describes in detail the trigger lines used, their definitions, their availability and the prescale factor. The corresponding input trigger bit definitions are listed in Table B.2.

* N prescaled trigger lines keep only every Nth event

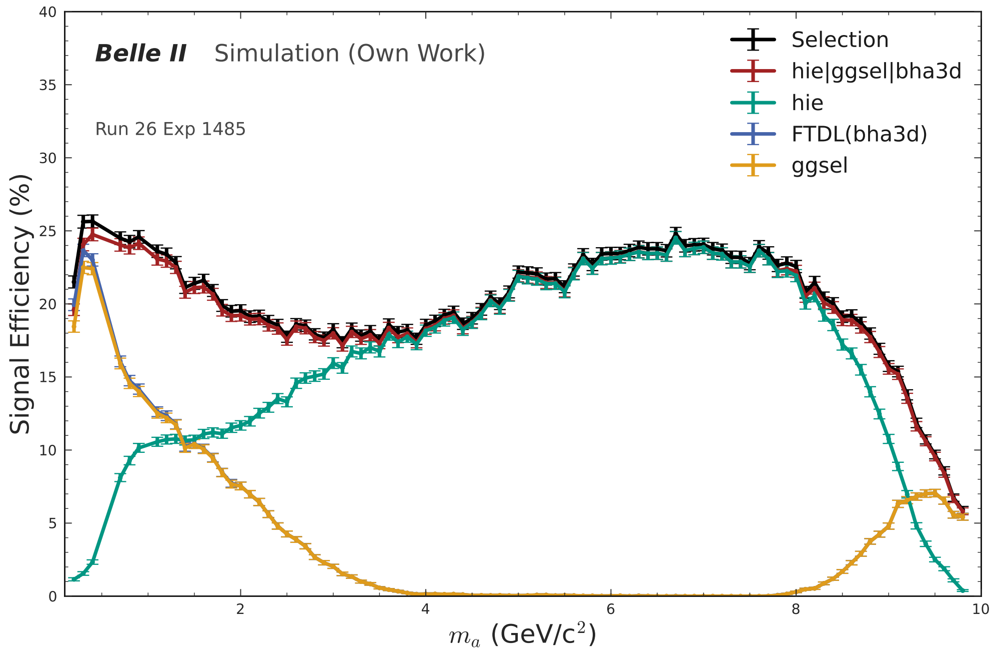
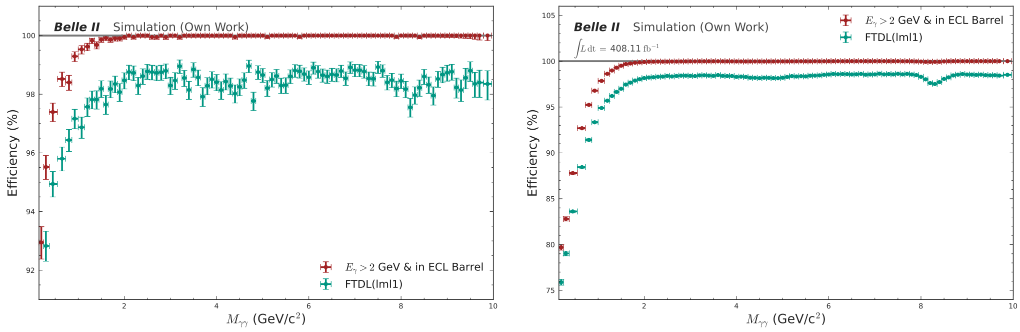


Figure 5.1: Signal efficiency of the ALP signal samples with and without the discussed trigger lines for beam background conditions corresponding to experiment 26 Run 1485. Black shows the final selection efficiency from **??** without any trigger. Yellow shows the signal efficiency when only triggering with `ggsel`, same is done for `hie` in green and FTDL `bha3d` in blue. Red displays the combined trigger efficiency of the combination of trigger lines `hie`|`bha3d`|`ggsel`.

Figure 5.1 displays a comparison of the signal efficiency with and without application of the discussed trigger lines. For `bha3d`, the Final Trigger Decision Logic (FTDL) bit is shown, which is the non-prescaled version of the `bha3d` trigger line. The previously mentioned trigger behaviour is visible when looking at the applications of the separate trigger lines. Starting from low ALP masses, `ggsel` is nearly always the most efficient trigger line while `hie` does not contribute to the efficiency, due to the Bhabha veto. With higher ALP masses, the efficiency of `ggsel` decreases, and `hie` becomes the most efficient trigger line. At around $m_a \approx 8\,\mathrm{GeV}/c^2$, the `ggsel` trigger efficiency begins to increase again, simultaneously the `hie` trigger efficiency decreases. This exact intertwining of the mutually exclusive trigger lines is the reason for defining the combined trigger line and the subsequent trigger validation strategy.

Trigger Efficiency Validation

Validating the trigger efficiency is usually done in one of two ways. The first evaluates the efficiency using sideband data, measuring it in a region where the signal is absent but the background is similar. However, due to the specific selections applied in the `ggsel`



(a) Signal (b) Background

Figure 5.2: Efficiency comparison between the triggered FTDL `lml1` line and a selection where each photon candidate is required to have at least 2 GeV of energy in the ECL barrel. The efficiency is shown for signal events (left) and background events (right) that pass the selection. The trigger efficiency is calculated for the signal MC samples with beam background conditions corresponding to experiment 26 Run 1485.

trigger line, no sideband region* accurately reflects the event kinematics expected in the signal region. The π^0 veto, pair conversion veto, and E_{γ^4} pre-selections specifically target very low $M_{\gamma\gamma}$ regions with distinct kinematics, making their sidebands unsuitable for this purpose. Furthermore, the number of clean tracks $N_{\mathrm{C.T.}}$ selection is also not viable due to the presence of `!trkbhaX` components in `ggsel`. Given that this analysis is not statistically limited, I instead measure the trigger efficiency directly on data. In the following, I present the method to measure the trigger efficiency and assert a systematic uncertainty.

To measure the trigger efficiency, I apply the commonly known tag and probe method. Originally, this method measures the trigger efficiency by calculating the efficiency of all events that pass the trigger line selection $N(\text{Probe})$ normalised to the number of events that are triggered by an independent trigger line $N(\text{Tag})$. In short, the trigger efficiency is defined as:

$$\varepsilon_{\text{trigger}} = \frac{N(\text{Probe \& Tag})}{N(\text{Tag})}. \tag{5.1}$$

Nevertheless, due to the combination of ECL- and CDC-based trigger information for `ggsel`, there is formally no orthogonal tag line†. Therefore, I choose a tag trigger line that is sensitive to the entire probe trigger line. A suitable candidate is the low multiplicity trigger line `lml1` which selects events that have at least one trigger cluster with an energy greater than 2 GeV in the ECL barrel region. The exact definition and the prescale are also presented in Chapter B. Higher coverage of the ECL barrel region is not possible due to the high prescales applied to the low multiplicity lines that select on ECL trigger clusters in the ECL endcaps. Figure 5.2 displays the FTDL `lml1` efficiency for the signal and background events that pass the selection. At the time of writing, there is a bug in ECL-based trigger lines in the simulation, which does not consider all ECL trigger clusters

*The sideband regions that are defined by inverting the pre-selections from Section 4.2.
†KLM based trigger decisions are not applicable for this analysis.

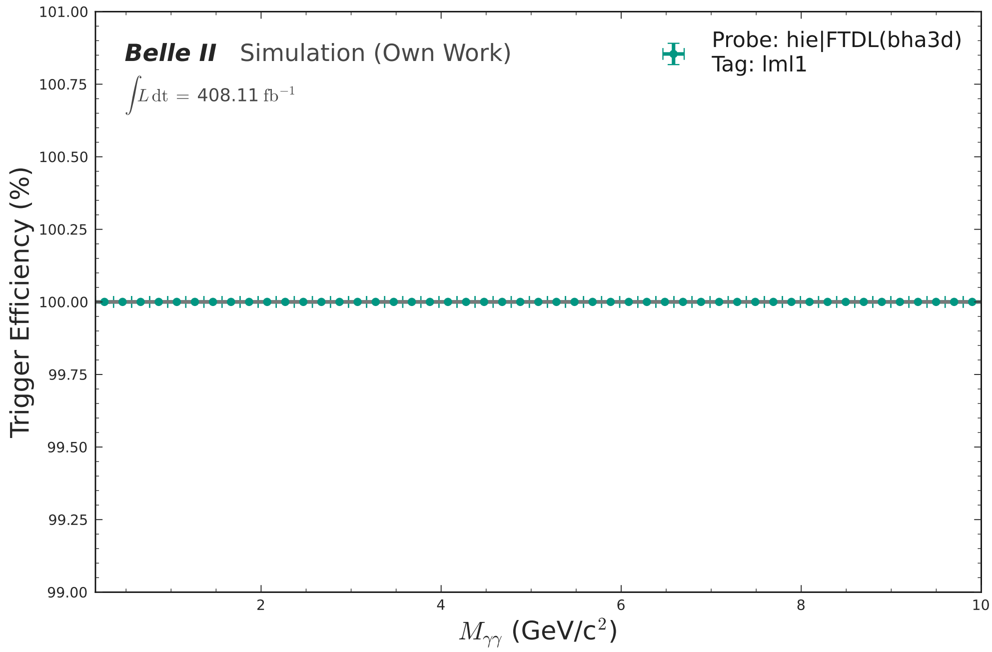


Figure 5.3: Trigger efficiency of `hie`|FTDL(`bha3d`) as a function of $M_{\gamma\gamma}$ on the simulated background events. Numerically, the trigger efficiency is 100 % for all $M_{\gamma\gamma}$ values.

in specific evaluations. This causes a trigger efficiency loss of about 1 to 2 %. To illustrate the difference, Fig. 5.2 displays two selections: one for the `lml1` trigger line and one where I mimic `lml1` by requiring each photon candidate to be in the ECL barrel region and have more than 2 GeV energy. As can be seen in the plots, the overall sensitivity of the trigger line is suitable for the trigger efficiency validation. In the low $M_{\gamma\gamma}$ region, the efficiency drops to about 93 % for signal events and about 80 % for background events due to the event topology discussed above. Around $M_{\gamma\gamma} \approx 8\,\mathrm{GeV}/c^2$ is another yet significantly smaller drop in efficiency of about 3 %.

To showcase the ideal scenario, Figure 5.3 displays the trigger efficiency of the trigger line `hie`|FTDL(`bha3d`) on the simulated background events that pass all selections from Chapter 4. This means that the energy selection of `ehigh` is enough to have a high trigger efficiency. In addition to that, I can also infer that no additional selection is strictly required to ensure that `hie` triggers past its activation point around 1 GeV. To ensure a trigger efficiency close to 100 %, I add a selection on a proxy variable for the energy that goes into the trigger decision of `hie`. For this, similar to [96], I calculate the ECL energy E_{ECL} in the event as the sum of uncorrected cluster energies for all photon and track candidates that are within the `hie` acceptance and have at least 100 MeV of energy. Figure 5.4 displays the distribution of E_{ECL} for the signal and background events that pass the selection. To all the previously discussed selections in Chapter 4, I add a requirement on E_{ECL} to be in the range of 4 to 13 GeV. This selection has no implications for the signal efficiency.

Figure 5.5 displays the trigger efficiency of the full trigger line `hie`|`ggsel`|`bha3d` on the

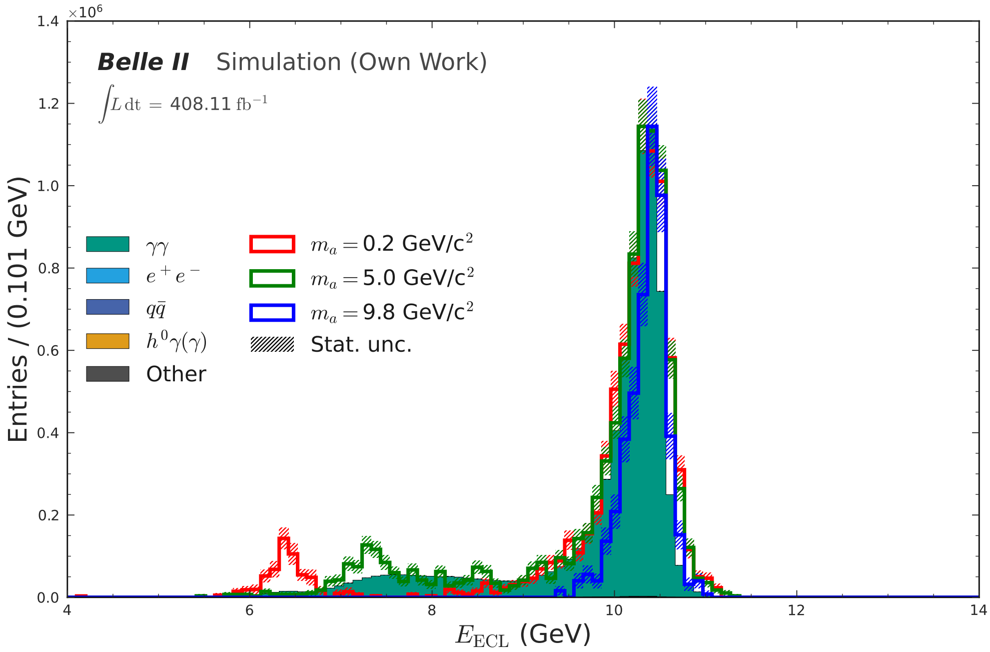


Figure 5.4: Distribution of the ECL energy E_{ECL} for a few example signal samples and the background events. For the signal MC samples, I show the simulation with beam background conditions corresponding to experiment 26 Run 1485.

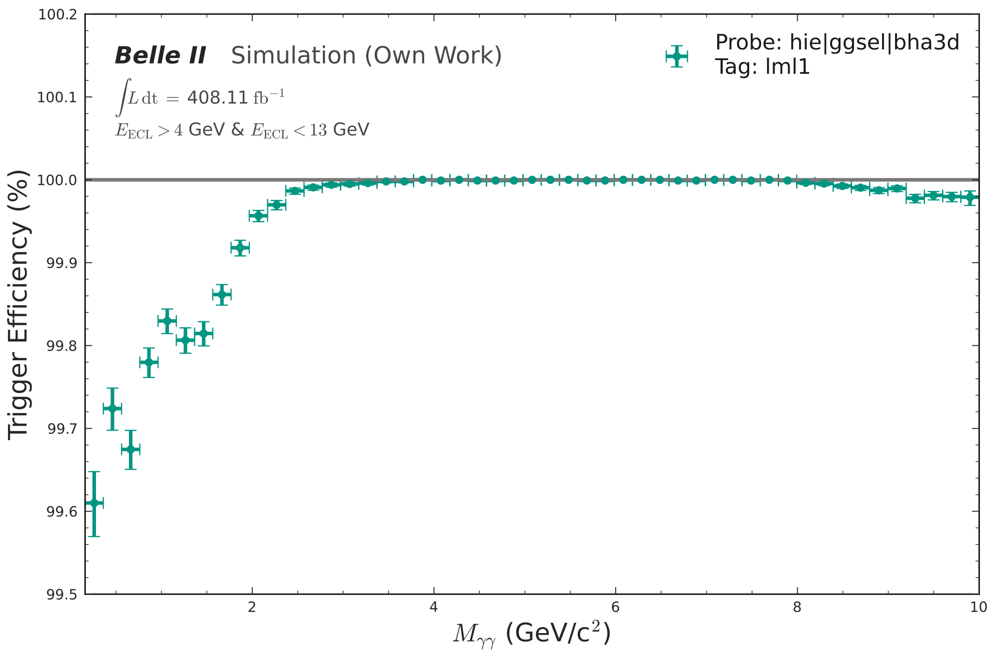


Figure 5.5: Trigger efficiency of `hie|ggsel|bha3d` with respect to `lml1` as a function of $M_{\gamma\gamma}$ on the simulated background events.

simulated background events. It is clear to see that without any additional selection, the trigger efficiency is very high for all $M_{\gamma\gamma}$ values – showing only a loss of about 0.4 % at the lowest $M_{\gamma\gamma}$ values. However, to further optimise the trigger efficiency, I study the additional input bits of the `ggsel` trigger line.

The main difference between the `ggsel` trigger line and the `bha3d` trigger line is due to the input bits `!trkbha1`, `!trkbha2` and `bha_intrk`. At the time of writing, there is a bug in the definition of `bha_intrk` in the trigger simulation. In contrast to what is implemented on the trigger hardware, the check for at least one of the clusters being in the CDC barrel region requires, in simulation, that both clusters be in this acceptance. Consequently, `ggsel` is in simulation not sensitive to the backward ECL endcap region and requires two clusters in the ECL barrel region plus the first trigger cell ring in the ECL forward endcap. Together with the already high trigger efficiency shown in Fig. 5.5, I estimate this effect to be negligible and expect only possibly a small increase in trigger efficiency on data in the very low or high $M_{\gamma\gamma}$ region.

The trigger lines `trkbhaX` are dependent on the number of tracks that the trigger can find. With the $N_{\mathrm{C.T.}}$ pre-selection, I specifically exclude events with more than one clean track in the event. However, reconstructed clean tracks and tracks found by the CDC trigger system are not necessarily the same. Allowing one clean track in the event, `!trkbha1` is the main reason for trigger efficiency loss in `ggsel`, `!trkbha2` should be less often the cause for not triggering `ggsel`, but is not entirely negligible. To increase the trigger efficiency in this regard, I apply an additional selection on the minimal cluster track distance minC2TD. This variable measures the shortest 3D distance between the extrapolated track and the associated ECL cluster of the photon candidate. Figure 5.6 displays the same trigger efficiency as in Fig. 5.5 but with additional example selections on minC2TD. Stricter selections on minC2TD lead to a higher trigger efficiency in the regions where `ggsel` is the main trigger line.

Since selections on minC2TD also affect the signal efficiency, I perform a PFOM (see Equation (4.7)) scan to find the optimal selection. Figure 5.7 displays the number of background events, the signal efficiency and the resulting PFOM. For the number of background events, the effect of the minC2TD is almost negligible. The signal efficiency starts to become significantly affected for selections of minC2TD $\gtrsim 100\,\mathrm{cm}$, equally across the entire ALP mass spectrum. As can be seen in Figure 5.7c, the PFOM scan does not show a clear optimum. The only trend that I observe is that the PFOM decreases slightly for selections of minC2TD $\gtrsim 100\,\mathrm{cm}$. For a conservative selection, I choose minC2TD $> 50\,\mathrm{cm}$.

To summarise the L1 trigger efficiency validation, I have shown that the trigger efficiency, with `lml1` as tag trigger line, of the full trigger line `hie|ggsel|bha3d` is high for all $M_{\gamma\gamma}$ values. I add to all previous selections an additional selection on the ECL energy E_{ECL} to be in the range of 4 to 13 GeV and minC2TD $> 50\,\mathrm{cm}$. The resulting trigger efficiency is shown in Figure 5.8. Above $M_{\gamma\gamma} \geq 2.0\,\mathrm{GeV}/c^2$, the trigger efficiency is mainly stable on the signal samples and shows only very small fluctuations on the simulated background. In both cases, the fluctuations are below 0.2 % and can be considered negligible. Below

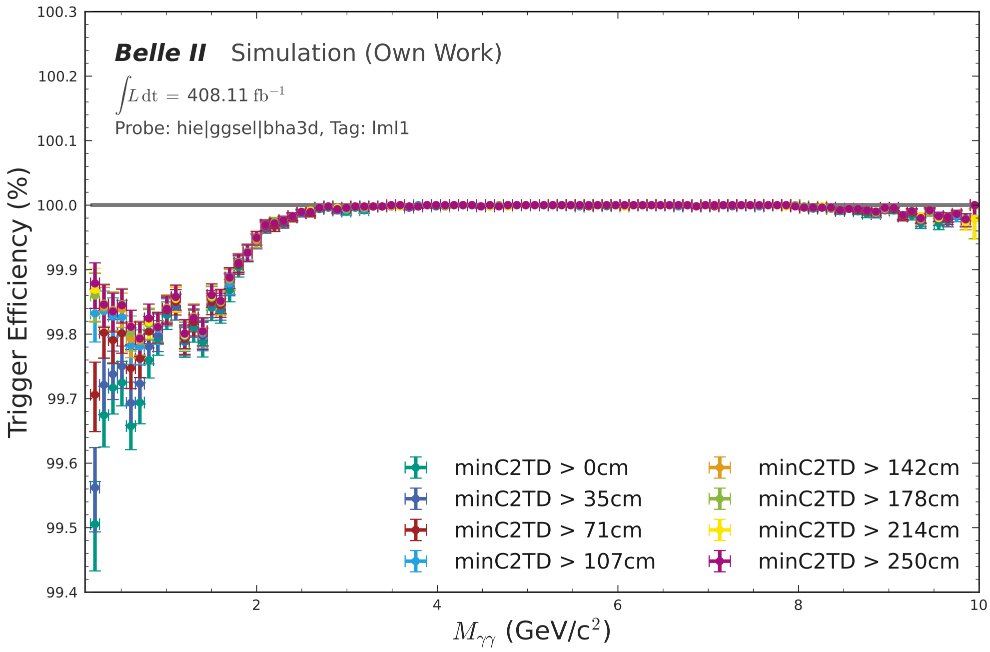
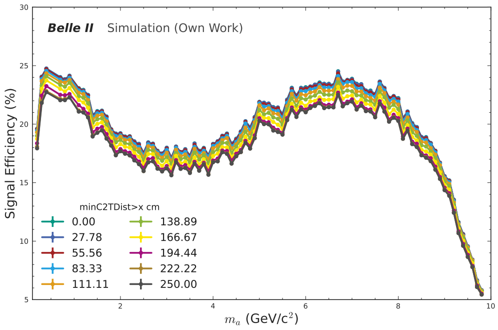
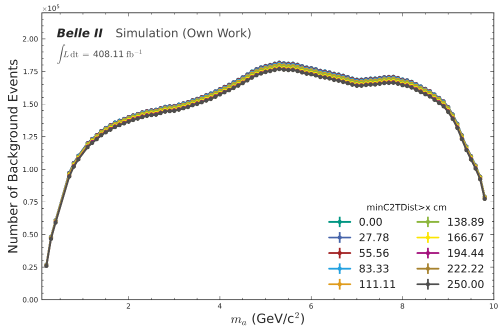


Figure 5.6: Trigger efficiency of `hie|ggsel|bha3d` as a function of $M_{\gamma\gamma}$ on the simulated background events. Additional lines in different colours show the trigger efficiency with additional selections on the minimal cluster track distance minC2TD.

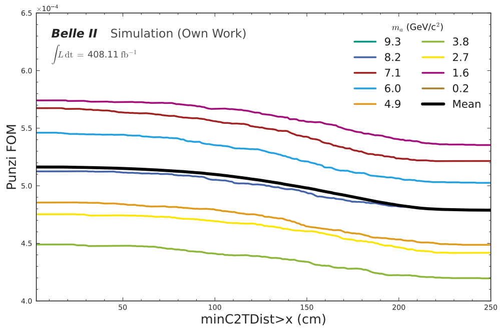
$M_{\gamma\gamma} \leq 1.0\,\mathrm{GeV}/c^2$, the trigger efficiency shows inefficiencies of the order of 0.4 %.



(a) Signal efficiency

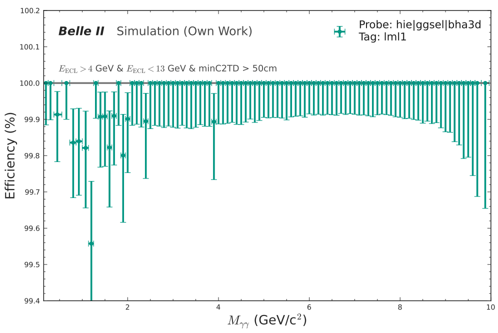


(b) Background

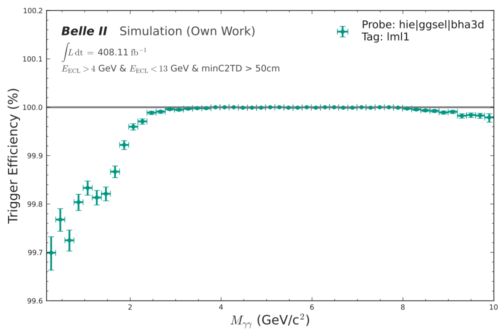


(c) PFOM

Figure 5.7: Signal efficiency (top left), number of background events (top right) for different selections of minC2TD, and the resulting PFOM (bottom). The signal efficiency and number of background events are calculated in a $\pm 10\sigma^{\mathrm{CB}}$ (see Section 6.1) window per sample mass on which the training was performed. For the signal events, samples with beam background conditions according to experiment 26 Run 1485 are used. For the number of background events calculation, the Run 1 dataset for $\Upsilon(4S)$ and off-resonance $\Upsilon(4S)$ runs is used. The different colored lines in the top plots correspond to different minC2TD selections. For the PFOM in the bottom plot, different colored lines show the result for example ALP mass samples.



(a) Signal efficiency



(b) Background

Figure 5.8: Trigger efficiency of `hie|ggsel|bha3d` as a function of $M_{\gamma\gamma}$ on the simulated background events and signal events with beam background conditions corresponding to experiment 26 run 1485. An additional selection on the ECL energy E_{ECL} to be in the range of 4 to 13 GeV and minC2TD > 50 cm is applied.

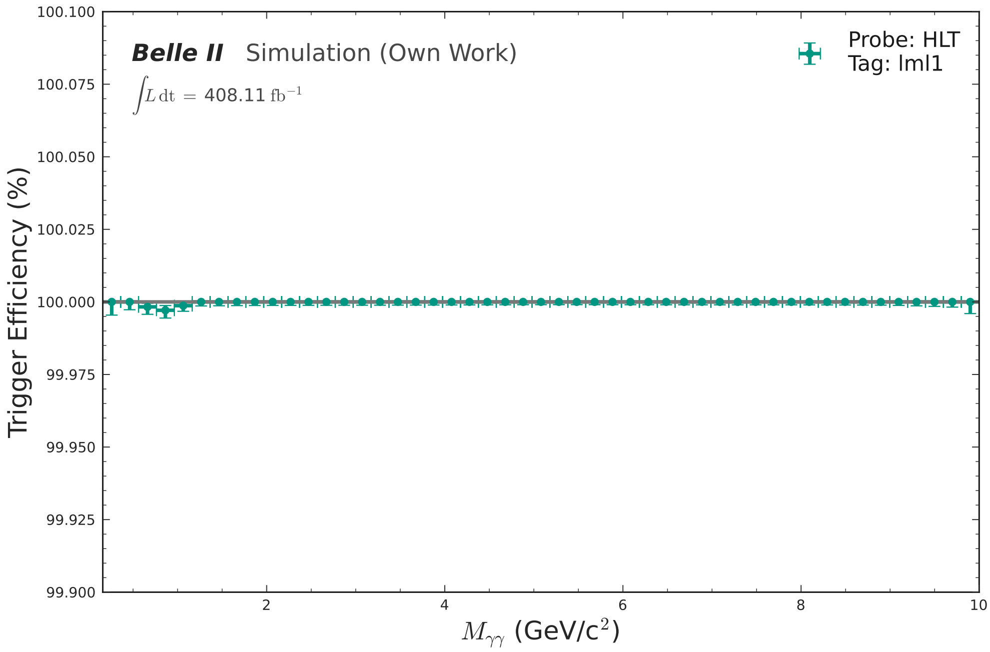


Figure 5.9: Trigger efficiency of the final HLT decision with respect to `lml1` as a function of $M_{\gamma\gamma}$ on the simulated background events.

5.1.2 HLT Trigger

In this section, I extend the trigger efficiency validation to the HLT since recorded events that enter the signal extraction must pass the L1 trigger and HLT. For the validation, the method discussed in Section 5.1.1 is executed on data in Section 10.2; the only addition to the previous discussion is the requirement of the positive final decisions of the HLT. First, I show that the previously discussed validation method is valid for the final trigger selection. Then, I estimate the HLT efficiency with respect to the L1 trigger based trigger line `hie|ggsel|bha3d`. Finally, I outline the method to determine the systematic uncertainty of the trigger efficiency. Throughout this section, I exclusively study these events that passed the selections from Chapter 4 and the additional trigger efficiency selections from Section 5.1.1.

The main reason why I do not need to adjust the validation method for the HLT triggered events is that I use `lml1` as a tag trigger line. Figure 5.9 shows that the `lml1` trigger line is fully efficient for the HLT decision within the statistical uncertainty.

However, during the signal extraction, I do not use `lml1` trigger line. Inefficiencies of the L1 trigger lines `hie|ggsel|bha3d` are therefore not directly visible in the trigger efficiency validation. For the signal samples, I observe only a minor inefficiency of about 0.5 % in the low $M_{\gamma\gamma}$ in the HLT efficiency in simulation. I attribute the inefficiency to the ECL barrel acceptance inefficiency shown in Fig. 5.2a. Figure 5.10 displays the pure HLT efficiency and the HLT efficiency with respect to the L1 trigger line `hie|ggsel|bha3d` for the signal

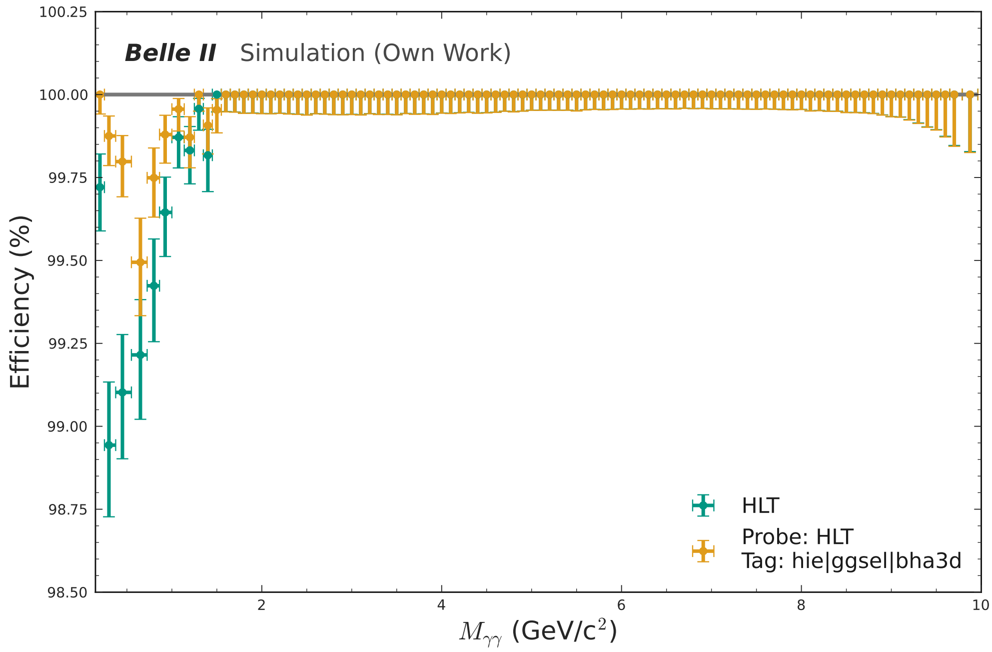


Figure 5.10: Trigger efficiency of the final HLT decision with respect to `lml1` as a function of $M_{\gamma\gamma}$ on the simulated signal events with beam background conditions corresponding to experiment 26 Run 1485. The pure HLT efficiency (green) and the HLT efficiency with respect to the L1 trigger line `hie|ggsel|bha3d` (orange) are shown.

samples. Already, the pure HLT efficiency shows a high efficiency with an inefficiency of up to 1.25 % for very low ALP masses. Taking into account the nearly 100 % efficiency of the `lml1` trigger line, I conclude that the trigger efficiency validation method can be performed as described in Section 5.1.1.

Figure 5.11 displays the final trigger efficiency, which is equivalent to Fig. 5.8b; the only difference is the requirement of the final decision of the HLT. For each ALP scan mass point, I calculate the trigger efficiency in the respective fit window in $M_{\gamma\gamma}$ and assign the statistical uncertainty of the efficiency calculation as systematic uncertainty on the signal efficiency for this scan point.

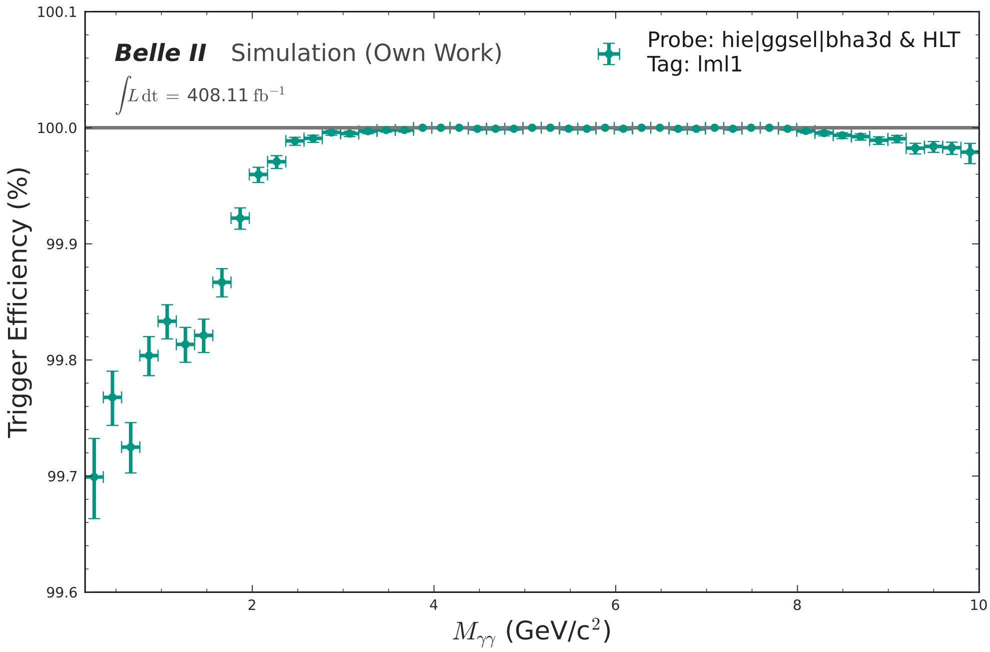


Figure 5.11: Trigger efficiency of (`hie`|`ggsel`|`bha3d`) with respect to `lml1` for HLT triggered events as a function of $M_{\gamma\gamma}$ on the simulated background events.

5.2 Run-Dependent Signal Efficiency

The final selection efficiencies displayed in **??** show slightly different efficiencies for simulations with different beam background conditions. Some differences are expected, simply due to the run dependency of *e.g.* the photon energy resolution, which factors into the width of the $M_{\gamma\gamma}$ distribution. To effectively discuss the run dependency during this sub-section, I will first introduce a metric to quantify beam background effects. Finally, I will discuss how the run dependency will be accounted for. For the entirety of this sub-section, I apply the selection from Chapter 4 and the previously described trigger selection from Section 5.1.

Because beam background is an umbrella term for a variety of different processes (see Section 3.4.1), it is not straightforward to quantify the beam background conditions. However, certain event properties have been identified that show a strong correlation with the beam background conditions. One of these properties is the number of out-of-time crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ in the event. The $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ is defined as the number of crystals in the event that have an absolute timing difference to the event time larger than 110 ns and have more than 7 MeV of energy deposited. Figure 5.12 displayes the luminosity weighted mean of the number of out-of-time crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ as a function of the continious run number. For the data recorded by the Belle II detector, the $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ is directly extracted without any prior selection to ensure no bias on the quantity. I then weight the $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ by the number of events in the run to account for the different integrated luminosities. The mean over all runs is $N_{\mathrm{OOT}}^{\mathrm{Crys.}} = 101.33$.

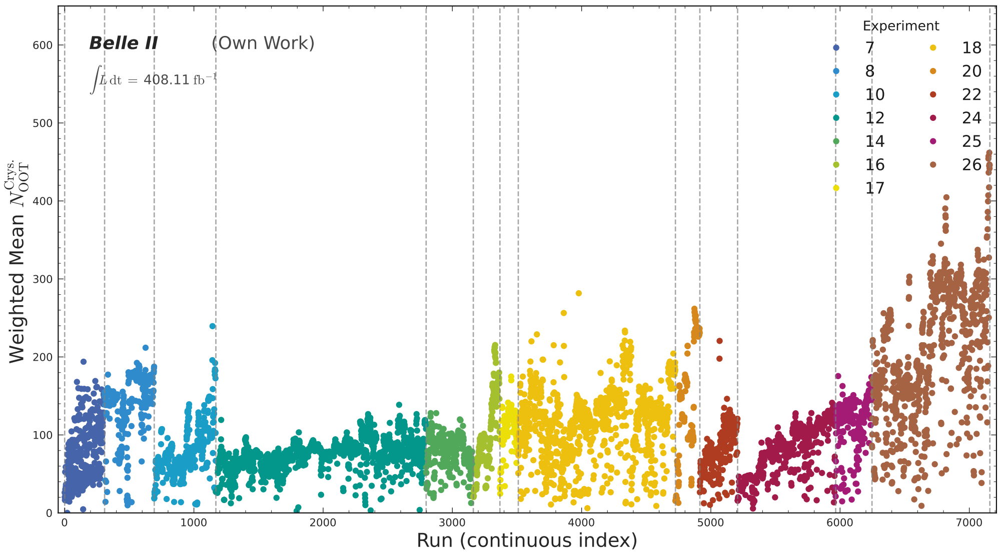
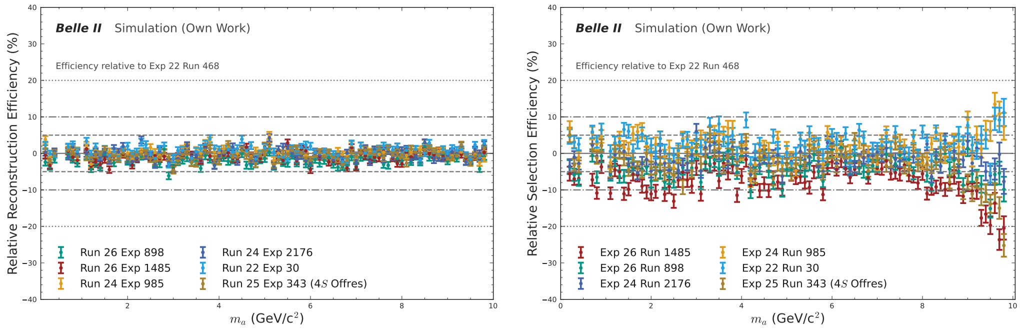


Figure 5.12: Luminosity weighted mean of the number of out-of-time crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ as a function of the continuous run number. The weight is based on the number of events in the run. Each colour represents a different experiment, and vertical lines indicate the experiment boundaries.

To quantify the dependency of the signal efficiency on the run conditions, I normalise the signal efficiency to the simulated run whose $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ value is closest to the overall run average. Using this reference, I find that only the reconstruction efficiency (see Fig. 5.13a[*]) fluctuates by up to 5 % when normalised to the efficiency of the MC signal samples with beam background conditions matching experiment 22 run 468 ($N_{\mathrm{OOT}}^{\mathrm{Crys.}} = 115.5$) which is the run with $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ closest to the absolute average of the whole Run 1 dataset. After the selection, the efficiency becomes slightly more run-dependent and fluctuates, depending on the beam background conditions, within 10 % as shown in Fig. 5.13b. Furthermore, starting at around $m_a \approx 9\,\mathrm{GeV}/c^2$, the relative selection efficiencies show greater divergence from the normalising efficiencies where the absolute efficiencies increase for runs with lower mean $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ values and decrease for runs with higher mean $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ values.

To account for differences in the run-dependent signal efficiency, I calculate the selection efficiency for all runs as a function of the $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ for all scan masses. I then fit a linear function to the selection efficiency. For the final determination of the signal efficiency, I evaluate the linear function at the mean $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ of the Run 1 dataset and assert an additional uncertainty on the efficiency based on the uncertainty of the fit. Figure 5.14 displays the procedure for three different ALP mass points. For the efficiency values, I assume the mean of the asymmetric statistical uncertainty as the y uncertainty and $\sqrt{N_{\mathrm{OOT}}^{\mathrm{Crys.}}}$ as the x uncertainty. The overall fit uncertainty is calculated as the quadratic sum of the

[*]These are the same efficiencies as in Fig. 4.4.



(a) Reconstruction efficiency (b) Selection efficiency

Figure 5.13: Relative signal efficiencies for different run conditions. The signal efficiencies are normalised to the signal efficiency of the signal MC samples with beam background conditions corresponding to experiment 22 run 468 ($N_{\mathrm{OOT}}^{\mathrm{Crys.}} = 102.2$). Different colours for simulations with different beam background conditions.

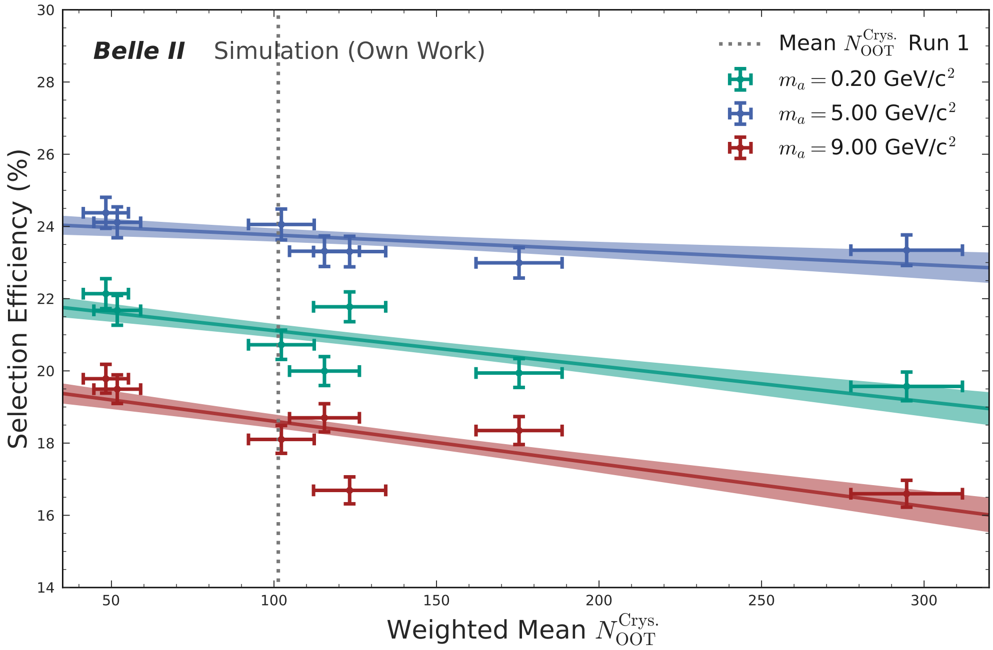


Figure 5.14: Selection efficiency as a function of the mean number of out-of-time crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ for three different ALP mass points. The solid lines show the linear fit to the selection efficiency, and the filled areas show the uncertainty of the fit. The grey dashed line indicates the mean $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ of the Run 1 dataset.

uncertainties of the y and x uncertainties and their covariances:

$$\varepsilon\left(N_{\mathrm{OOT}}^{\mathrm{Crys.}}\right) = aN_{\mathrm{OOT}}^{\mathrm{Crys.}} + b, \tag{5.2}$$

$$\delta_{\varepsilon(N_{\mathrm{OOT}}^{\mathrm{Crys.}})} = \sqrt{\left(N_{\mathrm{OOT}}^{\mathrm{Crys.}}\delta_a\right)^2 + \delta_b^2 + 2N_{\mathrm{OOT}}^{\mathrm{Crys.}}\delta_{a,b}}, \tag{5.3}$$

where a and b are the fit parameters, δ_a and δ_b their uncertainties, and $\delta_{a,b}$ the covariance between the two parameters.

Chapter 6

Signal and Background Parametrisation

In this chapter, I describe the parameterisation of the signal and background in the invariant diphoton mass observable, $M_{\gamma\gamma}$. For the signal extraction (see Chapter 9), I search for a significant excess of events in the $M_{\gamma\gamma}$ spectrum above the SM background expectation in a given ALP mass fit window. The parametrisation of the signal and background is therefore crucial to obtain a reliable estimate of the background and the significance of the signal. First, Section 6.1 describes the parametrisation of the signal with the peaking and combinatorial components. Additionally, this section derives the width of the signal σ^{CB}, which is used throughout this work. Then, Section 6.2 describes the parametrisation of the background with a polynomial function. The fits displayed here and in the following chapters are performed with the fitting framework `zfit` [97]. All fits in this chapter are unbinned, minimum Negative Log-Likelihood (NLL) fits.

6.1 Signal Parametrisation

Due to the unavoidable combinatorial contribution of the interchangeable photon candidates that build the diphoton invariant mass, the signal is parametrised by two components. The first component describes the peak of the signal, when both photon candidates in $M_{\gamma\gamma}$ are the decay products of the ALP decay. This component shows a clear peak at the position of the ALP mass. The second component describes the combinatorial contribution in the cases where one of the photon candidates is the recoiling photon in $e^+e^- \to \gamma a$, $a \to \gamma\gamma$. This contribution is heavily dependent on the ALP mass.

The total probability density function (PDF) of the signal parametrisation is defined as the sum of the two components:

$$S(M_{\gamma\gamma}; f, \boldsymbol{\theta}_p, \boldsymbol{\theta}_c) = f \cdot P(M_{\gamma\gamma}; \boldsymbol{\theta}_p) + (1 - f) \cdot C(M_{\gamma\gamma}; \boldsymbol{\theta}_c), \tag{6.1}$$

where $P(M_{\gamma\gamma};\boldsymbol{\theta}_p)$ is the peaking component, $C(M_{\gamma\gamma};\boldsymbol{\theta}_c)$ is the combinatorial component, and f is the fraction of the peak component. The fraction f and the shape parameters $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_c$ are nuisance parameters since they will be fixed based on the signal MC samples, as explained in the following.

For the peaking component, I use the so-called Double Sided Crystal Ball (DSCB) function, which is an extension of the Crystal Ball (CB) function [98, 99]. This extension adds the same power law tail to the right side of the peak of CB that is present on the left side. The total description of the peaking component is therefore given by:

$$P(M_{\gamma\gamma};\boldsymbol{\theta}_p) = \begin{cases} A_l \cdot \left(B_l - \dfrac{x-\mu}{\sigma}\right)^{-n_l}, & \text{for } \dfrac{x-\mu}{\sigma} < -\alpha_l \\ \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } -\alpha_l \leq \dfrac{x-\mu}{\sigma} \leq \alpha_r \,, \\ A_r \cdot \left(B_r + \dfrac{x-\mu}{\sigma}\right)^{-n_r}, & \text{for } \dfrac{x-\mu}{\sigma} > \alpha_r \end{cases} \tag{6.2}$$

with

$$\begin{aligned} \boldsymbol{\theta}_p &= \{\mu, \sigma, n_l, n_r, \alpha_l, \alpha_r\}\,, \\ A_{l/r} &= \left(\frac{n_{l/r}}{|\alpha_{l/r}|}\right)^{n_{l/r}} \cdot \exp\left(-\frac{|\alpha_{l/r}|^2}{2}\right), \text{ and} \\ B_{l/r} &= \frac{n_{l/r}}{|\alpha_{l/r}|} - |\alpha_{l/r}|. \end{aligned}$$

To stabilise the fits, I set the parameters n_l and n_r to the fixed values 3.0 and 2.0, respectively. These values are empirically motivated and well-suited for distributions with slightly larger left tails, which is the case for the ALP $M_{\gamma\gamma}$ signal distribution. Nevertheless, the exact choice is not of big importance, as the n and α parameters are highly correlated and can pick up variations in the shape of the peak.

For the combinatorial component, I use an orthogonal polynomial of first order. The general polynomial PDF is describe as:

$$C(M_{\gamma\gamma};\boldsymbol{\theta}_c) = \sum_{i=0}^{n} c_i \cdot T_i(M_{\gamma\gamma}), \tag{6.3}$$

where $T_i(M_{\gamma\gamma}) = \cos(i \arccos(M_{\gamma\gamma}))$ is the i-th order Chebyshev polynomial and $\boldsymbol{\theta}_c = \{c_0, c_1, \ldots, c_{n_c}\}$ are the coefficients of the polynomial. These polynomials are orthogonal, and the coefficients tend towards small values when the explicit order of the polynomial is not contributing to the total PDF. Due to the normalisation of the PDF, the coefficient c_0 is set to 1.0.

Figure 6.1 displays the fitted parameters of the signal parametrisation for different example ALP mass point samples with varying beam background conditions. The μ^{CB}

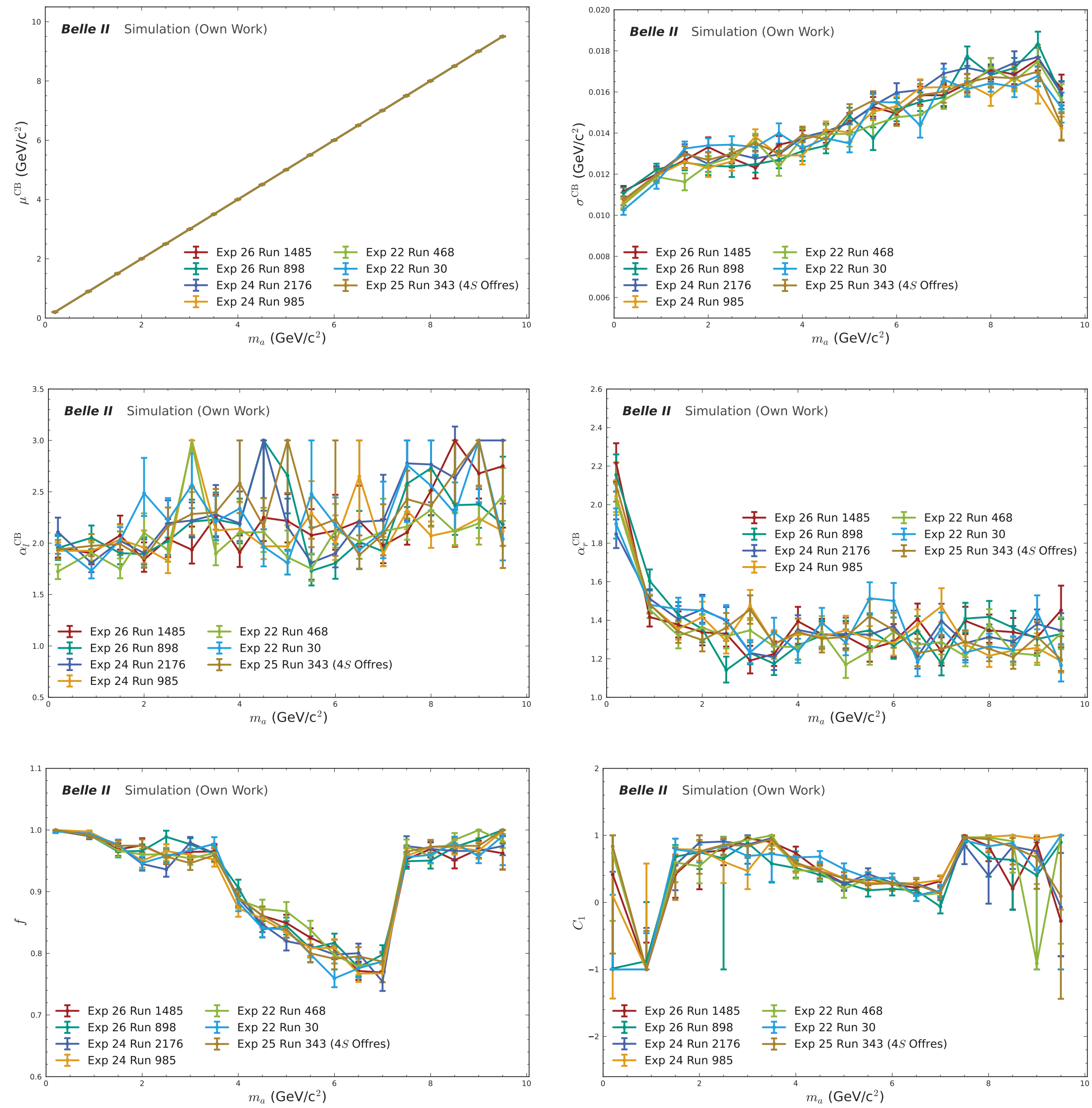


Figure 6.1: Signal PDF parameters as a function of ALP mass. Different colours represent different beam background conditions corresponding to the runs displayed in 3.4.

values follow the expected linear behaviour with the ALP mass; no bias in the relationship of the ALP mass and the μ^{CB} is observed. Similarly, the σ^{CB} values show an increasing trend for higher ALP masses, as is expected since higher ALP masses require higher photon energies. Consequently, higher photon energies lead to worse energy resolutions, resulting in broader $M_{\gamma\gamma}$ distributions. The only exception is the case of very high ALP masses, where the resolution of the recoiling photon becomes worse, leading to overcorrections from the kinematic fit. Furthermore, the shape of the fraction parameter f is expected to show the influence of the combinatorial contribution. Very low ALP masses show a very high fraction of the peak component, since the combinatorial contribution produces only very high values of $M_{\gamma\gamma}$. Similarly, very high ALP masses have very low $M_{\gamma\gamma}$ values in the combinatorial contribution. Only for very specific ALP masses, the fraction of the combinatorial contribution is higher when also the combinatorial $M_{\gamma\gamma}$ values are close to the ALP mass. The highest observed combinatorial contribution is about 25 % in the fit window for the ALP mass of $m_a \approx 7\,\mathrm{GeV}/c^2$. Following the trend of the fraction parameter f, the c_1 parameter has sensible and sizable values only in regions where the fraction is significantly different from 1. In regions where the fraction is close to 1, the c_1 parameter is not well constrained and shows large fluctuations. However, this is of no concern since the combinatorial contribution PDF is not contributing to the total PDF.

Since no significant differences in the fitted parameters are observed for different beam background conditions, I use the same values to parameterise the signal across the entire Run 1 dataset. Figure 6.2 displays the $M_{\gamma\gamma}$ distributions with the fitted signal parametrisation for a few example ALP masses.

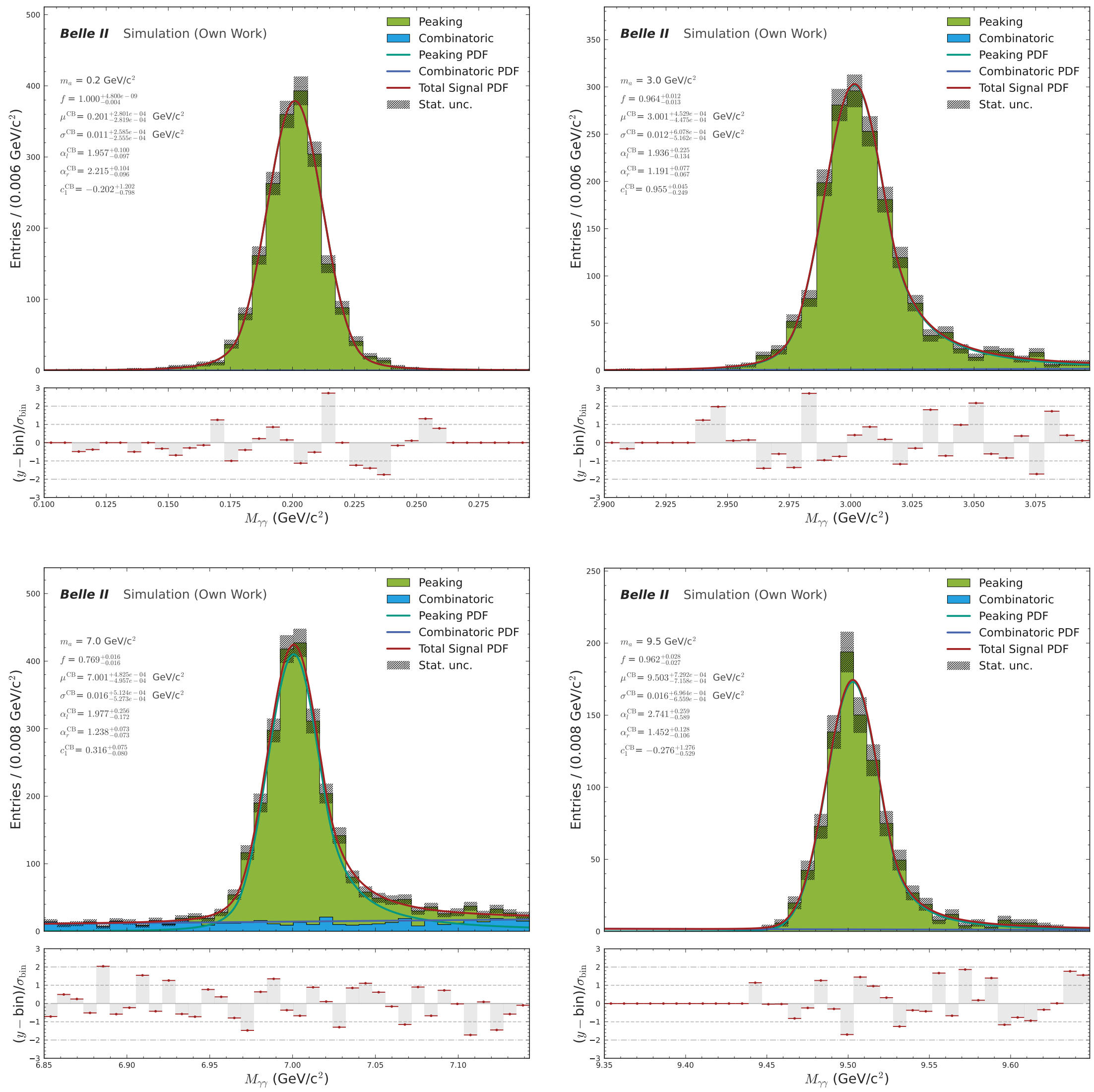


Figure 6.2: Distributions of the $M_{\gamma\gamma}$ observable for signal events with the fitted signal parametrisation for example ALP masses. Each histogram shows the peaking component (may-green), the combinatorial component (cyan), the fitted peaking PDF (green), the fitted combinatorial PDF (blue), and the total signal PDF (red).

6.2 Background Parametrisation

As I outline in Section 3.4.2, the background consists of peaking and non-peaking components across the $M_{\gamma\gamma}$ spectrum. For the peaking components, I exclude the resonant regions corresponding to $e^+e^- \to h^0\gamma(\gamma)$ processes in the π^0 and η regions, as well as the ω region. As shown in Chapter 7, the MC/data agreement in these resonant regions is good, particularly when considering the PHOKHARA-based simulations. Since no significant yield is expected in simulation from the $e^+e^- \to \eta\gamma$ process, this region is included in the signal extraction. Table 6.1 summarises the $M_{\gamma\gamma}$ regions that are excluded from the scan

Table 6.1: $M_{\gamma\gamma}$ regions excluded from the signal extraction due to peaking background processes. $M_{\gamma\gamma}^{\text{min}}$ and $M_{\gamma\gamma}^{\text{max}}$ correspond to the lower and upper bounds of the excluded regions, respectively.

Process	$M_{\gamma\gamma}^{\text{min}}$ [GeV/c^2]	$M_{\gamma\gamma}^{\text{max}}$ [GeV/c^2]
$e^+e^- \to \pi^0\gamma$	0.09	0.16
$e^+e^- \to \eta\gamma$	0.45	0.63
$e^+e^- \to \omega\gamma$	0.72	0.81

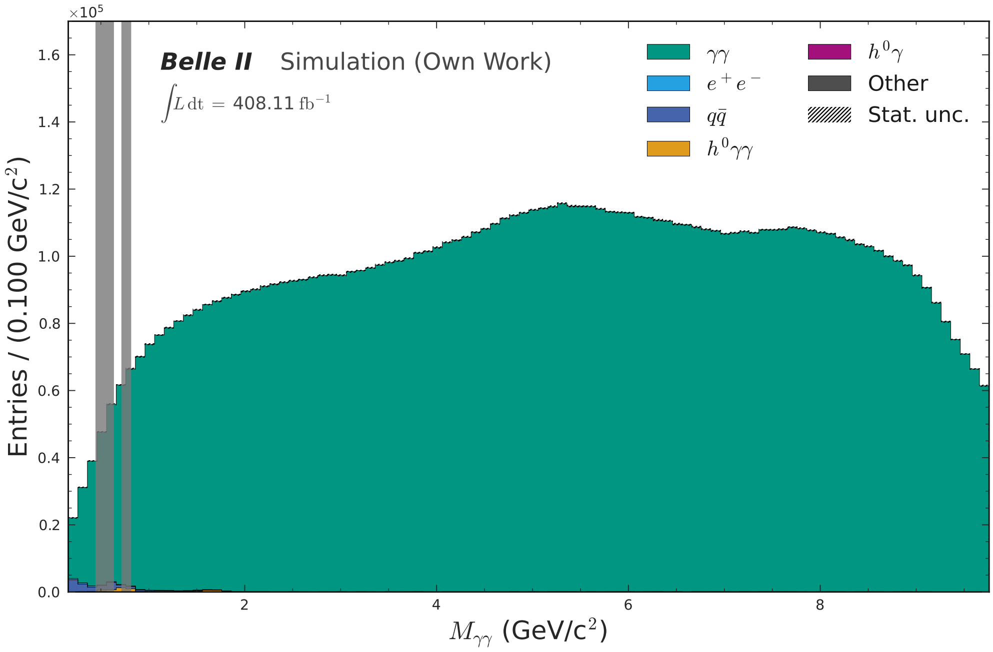


Figure 6.3: Distribution of $M_{\gamma\gamma}$ for the background processes that go into the signal extraction. These events have passed the selection described in Chapter 4, the trigger selection and are HLT and L1 trigger triggered as described in Section 5.1. Grey bars symbolise the $M_{\gamma\gamma}$ regions excluded from the signal extraction due to peaking background processes $e^+e^- \to h^0\gamma$.

and partially used for the control region studies.

The remaining background is mainly dominated by the $e^+e^- \to \gamma\gamma(\gamma)$ process, which can be seen in Fig. 6.3. Due to the large dominance of the $e^+e^- \to \gamma\gamma(\gamma)$ process and its smooth shape, I parametrise the background locally in each ALP mass fit window with a polynomial function. Similar to Eq. (6.3) I define the background PDF as

$$B(M_{\gamma\gamma};\boldsymbol{\theta}_b) = \sum_{i=0}^{n} b_i \cdot T_i(M_{\gamma\gamma}),, \tag{6.4}$$

where T_i are the Chebyshev polynomials and $\boldsymbol{\theta}_b = \{b_1, \ldots, b_{n_b}\}$ are the coefficients of the polynomial.

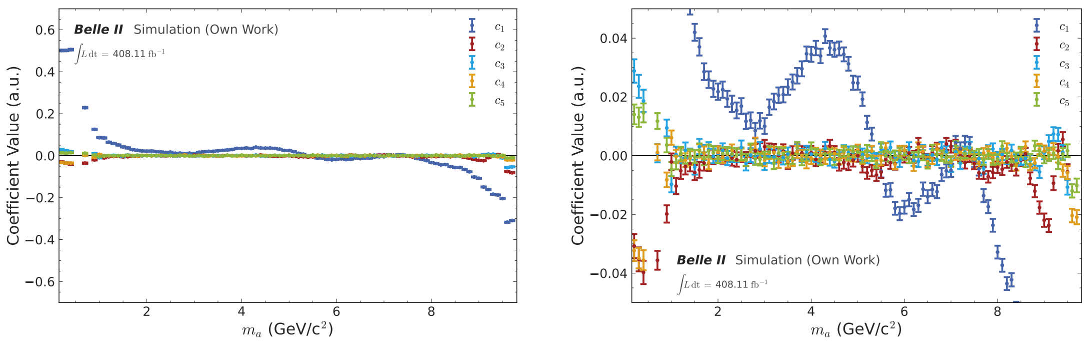


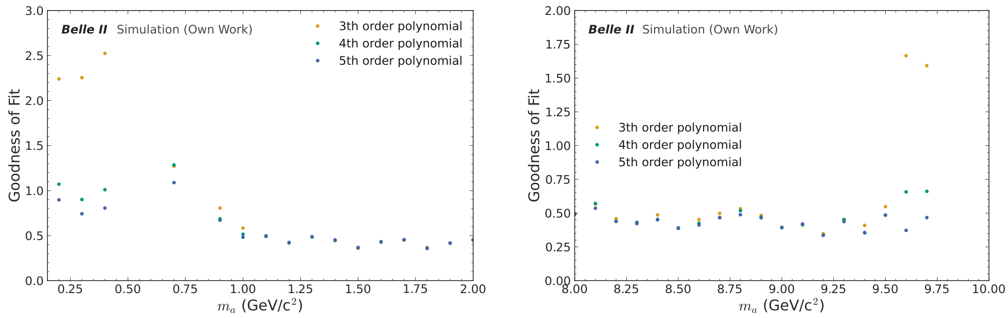
Figure 6.4: Distribution of the background fit coefficients for a fifth-order polynomial. The left plot shows the full range of the coefficients, while the right plot zooms into the range $[-0.5, 0.5]$ to better visualise the distributions of the higher-order coefficients.

To estimate the optimal order of the polynomial based on simulated events, I perform background-only fits in all $\pm 20\sigma^{\mathrm{CB}}$ ALP mass windows. The background events are subject to the same selection criteria as those used for the signal extraction described in Chapter 9. Fits are evaluated for polynomial orders of three, four, and five. Figure 6.4 shows the distributions of the fitted coefficients for a fifth-order polynomial. When using orthogonal polynomials, coefficients that do not contribute significantly are expected to be compatible with zero. From the coefficient distributions, I observe that in the range $m_a \in [1, 8.5]\,\mathrm{GeV}/c^2$, the first-order coefficient is the most significant, while the second-order coefficient is mostly inconsistent with zero, and higher-order terms are compatible with zero. In the remaining mass ranges, the higher-order coefficients also become significant.

To estimate which polynomial order is sufficient to describe the background based on simulation, I study the goodness of fit for different polynomial orders. Since no well-defined estimator exists for unbinned fits, I approximate the χ^2/ndf approach by binning the background events into histograms with 100 bins and comparing the fitted model to the binned data. The number of degrees of freedom is estimated as the number of bins minus the number of free fit parameters. The χ^2 values are then calculated as

$$\chi^2 = \sum_{i}^{N_{\mathrm{bins}}} \frac{(N_i^{\mathrm{hist}} - N_i^{\mathrm{fit}})^2}{N_i^{\mathrm{fit}}}, \tag{6.5}$$

where N_i^{hist} is the number of entries in bin i of the histogram, and N_i^{fit} is the number of entries in bin i of the fitted function, normalised to the total number of entries in the histogram. Figure 6.5 shows the distribution of the χ^2/ndf values for the different polynomial orders in the low- and high-ALP mass hypothesis regions. Based on the goodness-of-fit values being close to one, I conclude that the third-order polynomial is insufficient. The fourth- and fifth-order polynomials show comparable performance, with the fourth order performing slightly better; therefore, I choose the fourth-order polynomial



(a) $m_a \leq 1\,\mathrm{GeV}/c^2$ (b) $m_a \geq 8.5\,\mathrm{GeV}/c^2$

Figure 6.5: Distribution of the goodness of fit values for different polynomial orders in the low (left) and high (right) ALP mass hypothesis regions.

as the nominal background parametrisation in the low- and high-ALP mass regions. To summarise, the polynomial orders used in the different ALP mass regions are:

- fourth-order polynomial for $m_a \leq 1\,\mathrm{GeV}/c^2$,
- second-order polynomial for $m_a \in [1, 8.5]\,\mathrm{GeV}/c^2$,
- fourth-order polynomial for $m_a \geq 8.5\,\mathrm{GeV}/c^2$.

Chapter 7

Control Regions

Typically, in blind particle physics analyses, the agreement between simulation and data in the signal region remains unknown until the unblinding procedure, in order to avoid experimenter bias. For this reason, control regions are defined to validate the simulation in regions that are kinematically similar to the signal region. In this work, a complete coverage of the kinematically equivalent phase space is not possible, as only a few fundamental processes exhibit the $e^+e^- \to \gamma\gamma(\gamma)$ signature without simultaneously being part of the signal region. Consequently, I define two strategies to validate the analysis strategy.

First, I invert the π^0-veto described in Section 4.2.4. Physically, this is the most well-motivated control region obtained from inverted selections, since other inverted selections primarily probe detector effects rather than underlying physics processes. In Section 7.1 I discuss the observations in this control region, including the overall MC/data agreement, the $e^+e^- \to q\overline{q}$ and PHOKHARA-based simulations, the L1 trigger simulation, and the impact of the kinematic fit at low $M_{\gamma\gamma}$ values.

Second, I use the irreducible background regions from $e^+e^- \to \eta\gamma$ and $e^+e^- \to \omega\gamma \to \pi^0\gamma\gamma$, which are excluded from the signal extraction scan, as control regions to validate the simulation under the full selection. I summarise the observations in Section 7.2 and Section 7.3 in these control regions, which primarily focus on the overall MC/data agreement and the validation of the trigger efficiency determination.

I define no dedicated control region for the $e^+e^- \to \eta'\gamma$ process since it is possible to include it in the signal extraction scan. As I show in the following chapters, the agreement between data and simulation is well established — particularly for the PHOKHARA-based simulations — and no significant contributions from the $e^+e^- \to \eta'\gamma$ process are expected according to the simulation.

Throughout this chapter, I weight all simulated events according to the MC/data photon efficiency correction that I discuss in Section 8.1, and all recorded data events according to the photon energy bias correction in Section 8.2.

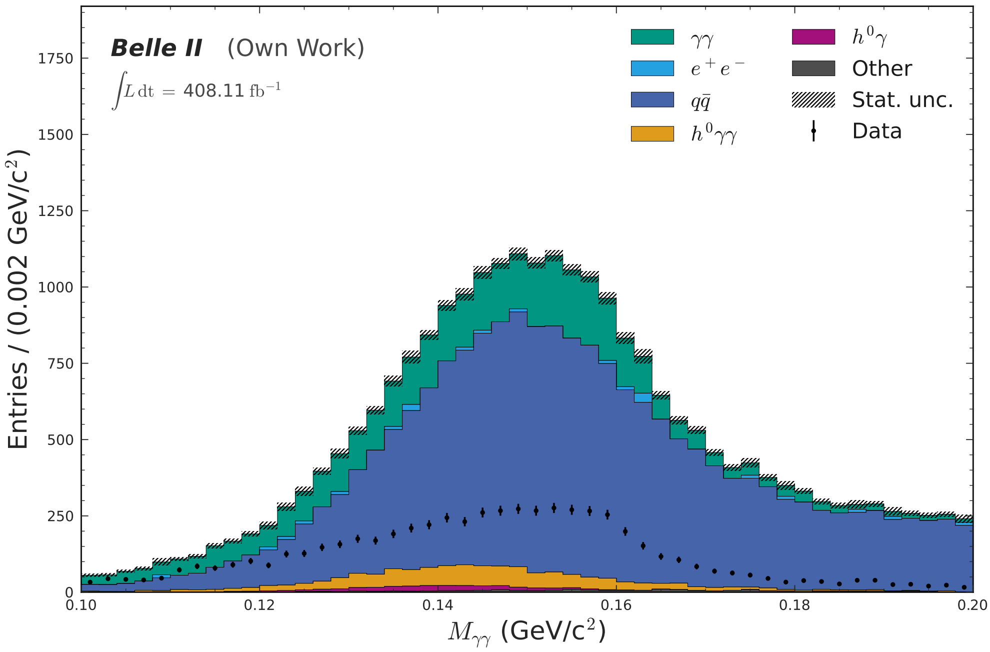


Figure 7.1: Distribution of the invariant diphoton mass, $M_{\gamma\gamma}$, in the inverted π^0-veto control region. Filled, coloured histograms represent the expectations from simulated events, while black error bars indicate the recorded data. The distribution includes events that pass all selection criteria, as well as the L1 trigger and HLT selections. Simulated events are additionally weighted according to the photon efficiency correction, and recorded data events are weighted according to the photon energy bias correction.

7.1 Inverted π^0 Veto

In this control region, I invert the π^0-veto described in Section 4.2.4 to define a selection orthogonal to that used for the signal extraction. I study the events presented here in the range $M_{\gamma\gamma} \in [0.1, 0.2]\,\mathrm{GeV}/c^2$, which should, in principle, be enriched with π^0 events. Nevertheless, the following discussion shows that the contribution from $e^+e^- \to q\bar{q}$ is significantly overestimated in the simulation, while the dominant background contribution is estimated to originate from $e^+e^- \to \gamma\gamma(\gamma)$ and $e^+e^- \to \pi^0\gamma(\gamma)$. Consequently, I adjust the selection strategy to exclude selections that specifically target the reduction of $e^+e^- \to q\bar{q}$ contributions. Furthermore, I observe significant discrepancies between data and simulation in this control region, primarily due to the L1 trigger simulation. However, no corrective action is deemed necessary, as the impact on the signal region is estimated to be negligible. Finally, I discuss the impact of the kinematic fit at low $M_{\gamma\gamma}$ values.

Events from $e^+e^- \to q\bar{q}$

Figure 7.1 shows the $M_{\gamma\gamma}$ distribution in the inverted π^0-veto control region. All events shown here pass the full selection described in Chapter 4, except for the π^0-veto, which

Table 7.1: Overview of the most frequent final states in the $e^+e^- \to q\overline{q}$ simulation that pass the selection criteria and fall within the inverted π^0-veto control region. The fractions are calculated with respect to the total number of $e^+e^- \to q\overline{q}$ events that satisfy the selection. Final states contributing less than 0.1 % are not shown. The final states are identified using the generator-level information provided by the basf2 framework. Additional photons originating from ISR or FSR are not included in this overview.

Final State	**Fraction** (%)
$\pi^0\pi^0$	93.48
$\pi^0\pi^0\pi^0$	1.78
$\pi^\pm\rho^\mp$	1.32
$\pi^+\pi^-\pi^0$	1.31
$\eta\pi^0$	0.66
$\pi^+\pi^-\pi^0\pi^0$	0.30
$\pi^\pm\pi^0\rho^\mp$	0.30
$\omega\pi^0$	0.18
$\pi^0\pi^0\pi^0\pi^0$	0.12
$\eta\pi^0\pi^0$	0.11

is inverted. Additionally, the L1 trigger and HLT selections are applied as outlined in Chapter 5, and the event weights are adjusted as described in Chapter 8. The seemingly largest contribution in this control region originates from $e^+e^- \to q\overline{q}$. The distribution of events in recorded data does not agree with the simulation expectation, both in shape and normalisation.

Upon closer examination of the simulation process and output for the $e^+e^- \to q\overline{q}$ events, it becomes evident that the MC generation does not accurately reproduce the underlying physical conditions. As I outline in Table 3.5, the $e^+e^- \to q\overline{q}$ events, particularly those involving lighter quarks, are generated using KKMC. The KKMC output, which contains the $q\overline{q}$ final state, is subsequently processed with Pythia8 to simulate the hadronisation process. During this stage, information about the parity and angular momentum of the intermediate state is not passed to Pythia8. Consequently, Pythia8 fragments the $q\overline{q}$ pair into hadrons without enforcing the corresponding physical conservation laws. This behaviour leads to the generation of physically forbidden processes, such as $e^+e^- \to q\overline{q} \to \pi^0\pi^0$, which, at leading order, violates charge conjugation (C) and parity (P) conservation. The e^+e^--annihilation carries the total quantum numbers $J^{CP} = 1^{--}$. Final states with two pseudoscalars, which are their own anti-particles, such as $\pi^0\pi^0$, must have a symmetric spatial wavefunction under exchange. This requires even orbital angular momentum ($L = J = 0, 2, \dots$), which leads to $C = +1$ and $P = +1$. Therefore, the quantum numbers of the initial and final states cannot match, making the transition forbidden. Table 7.1 lists the most frequent final states in the $e^+e^- \to q\overline{q}$ simulation that satisfy the same selection criteria as described above. The most common final state is $\pi^0\pi^0$, which accounts for 93.48 % of all $e^+e^- \to q\overline{q}$ events in this control region and is physically forbidden. Other

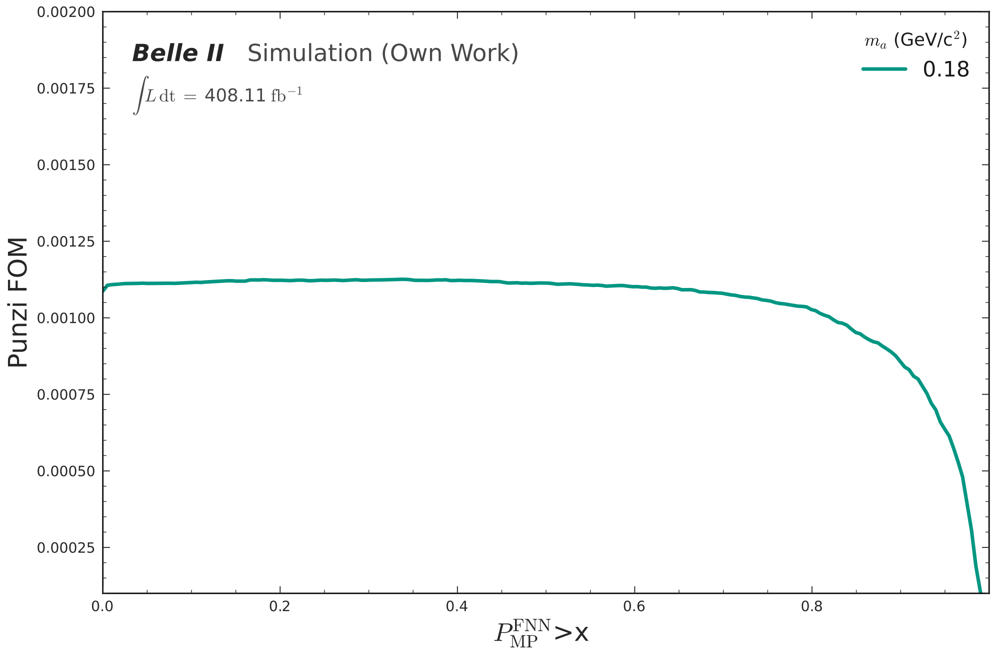


Figure 7.2: PFOM as a function of $P_{\mathrm{MP}}^{\mathrm{FNN}}$ for one example ALP mass with $m_a = 0.18\,\mathrm{GeV}/c^2$. For the number of background events and the signal efficiencies, the same samples are used as in Figure 4.43.

listed final states consisting solely of pseudoscalar particles are likewise forbidden under the same conservation law. Final states containing fermions, however, are physically allowed. Nevertheless, given their small contribution and the overall agreement between simulation and recorded data, as shown in Figure 7.3, I neglect these processes.

Considering the simulation details presented above, and the qualitatively similar picture observed in the other two control regions, I exclude all events from $e^+e^- \to q\bar{q}$ processes from further analysis. A direct consequence of this decision is the removal of the merged photon classifier selection (see Section 4.4.3) from the overall event selection. This is because the merged photon classifier is trained exclusively on events originating from the excluded $e^+e^- \to q\bar{q}$ processes. Furthermore, as shown in Fig. 7.2 for a low ALP mass example, applying a selection on the merged photon classifier output does not improve the sensitivity according to the PFOM when the remaining background processes are considered.

Figure 7.3 shows the $M_{\gamma\gamma}$ distribution in the inverted π^0-veto control region, excluding the simulation contributions from $e^+e^- \to q\bar{q}$ processes and omitting the merged photon classifier selection. The remaining simulated processes describe the recorded data well in most regions in shape and normalisation. In particular, the PHOKHARA-based simulation of $e^+e^- \to h^0\gamma(\gamma)$ provides a significant contribution to the overall distribution, indicating its importance and accuracy in the background modelling. Nevertheless, a noticeable discrepancy remains between the simulation and the recorded data, especially in the very low $M_{\gamma\gamma}$ region.

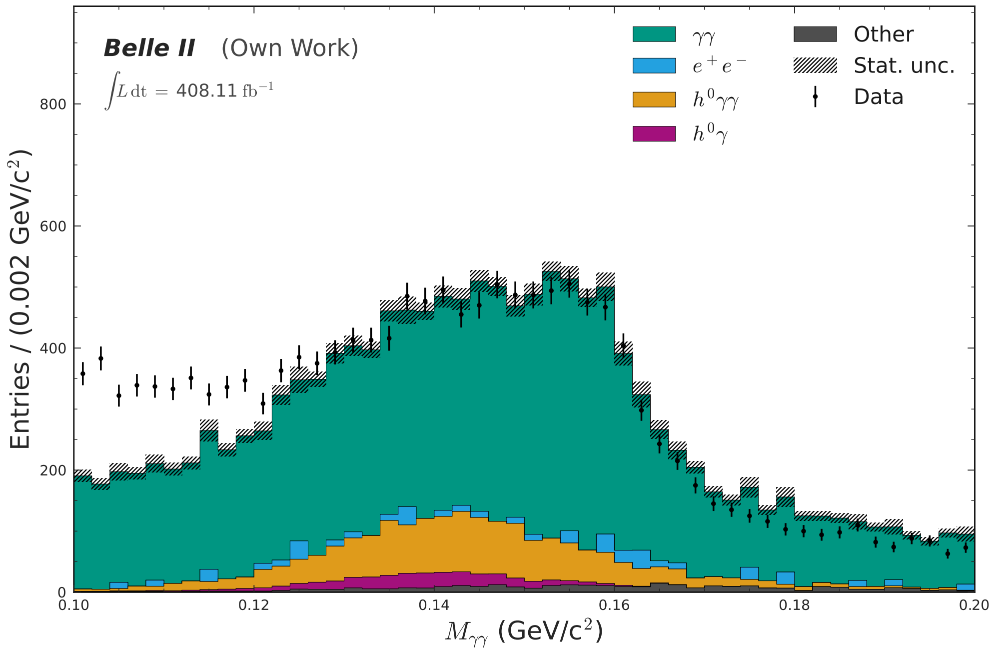


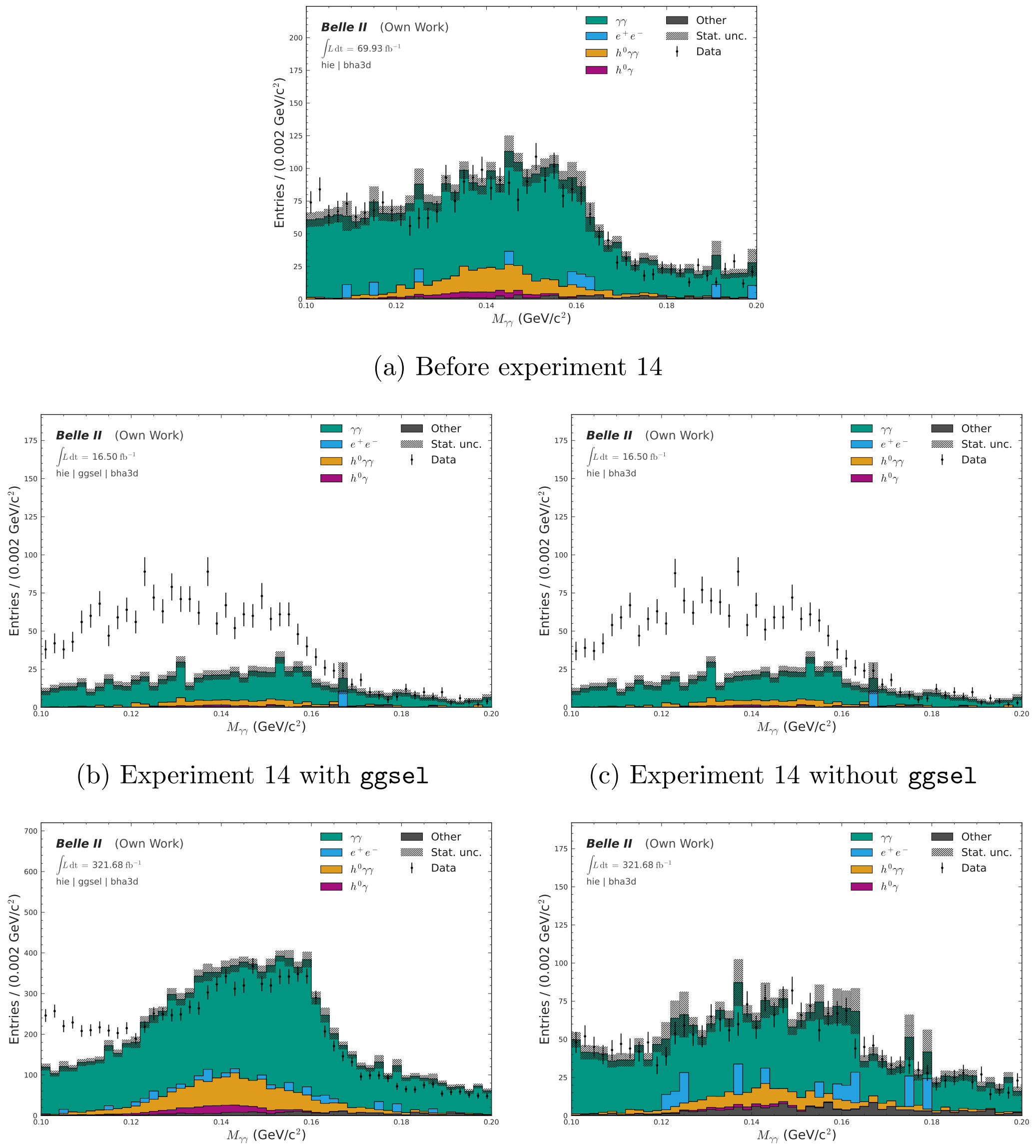
Figure 7.3: Distribution of the invariant diphoton mass, $M_{\gamma\gamma}$, in the inverted π^0-veto control region. The shown events are identical to those in Figure 7.1, except that all contributions from $e^+e^- \to q\bar{q}$ processes are excluded, and the merged photon classifier selection is omitted.

L1 Trigger Simulation

I attribute the discrepancies observed in Fig. 7.3 to the L1 trigger simulation. Although the exact cause of these discrepancies remains unidentified, a comparison of MC/data agreement across different experiments suggests that the most likely source is the simulation of `ggsel`, which results in fewer events passing the selection for MC than for recorded data. For the following discussion, I apply all selections, including the HLT selection but excluding the L1 trigger selection.

As I mention in Section 5.1.1, before experiment 14, `ggsel` did not exist and `bha3d` had a prescale factor of one. For events from this period, the agreement between recorded data and simulation is very good, as Fig. 7.4a shows. Starting with experiment 14, when `ggsel` was introduced, the MC/data agreement deteriorates significantly. Figures 7.4b and 7.4c show the $M_{\gamma\gamma}$ distributions for experiment 14, with and without the `ggsel` selection, respectively. For this experiment, the two distributions are almost identical, even though the L1 trigger lines had partially different prescale factors or were occasionally turned off entirely. I list the exact prescale factors and trigger line availabilities in Table B.1. After experiment 14, `ggsel` is fully available, and `bha3d` is prescaled by a factor of 100 starting from experiment 17. During this period, the disagreement between recorded data and simulation persists only for the `ggsel` selection, as Fig. 7.4d shows. Figure 7.4e shows that the disagreement disappears when `ggsel` is not applied. The documentation of the

L1 trigger line implementation for these time periods does not indicate any deviations from what is propagated to the simulation within basf2. Consequently, the most likely explanation is that the L1 trigger hardware implementation deviates from the documented configuration at the time of writing. Since these discrepancies occur only in the very low $M_{\gamma\gamma}$ region — well separated from the signal region — and because the trigger efficiency is derived directly from recorded data, I take no further action to address this disagreement.



(a) Before experiment 14

(b) Experiment 14 with `ggsel`

(c) Experiment 14 without `ggsel`

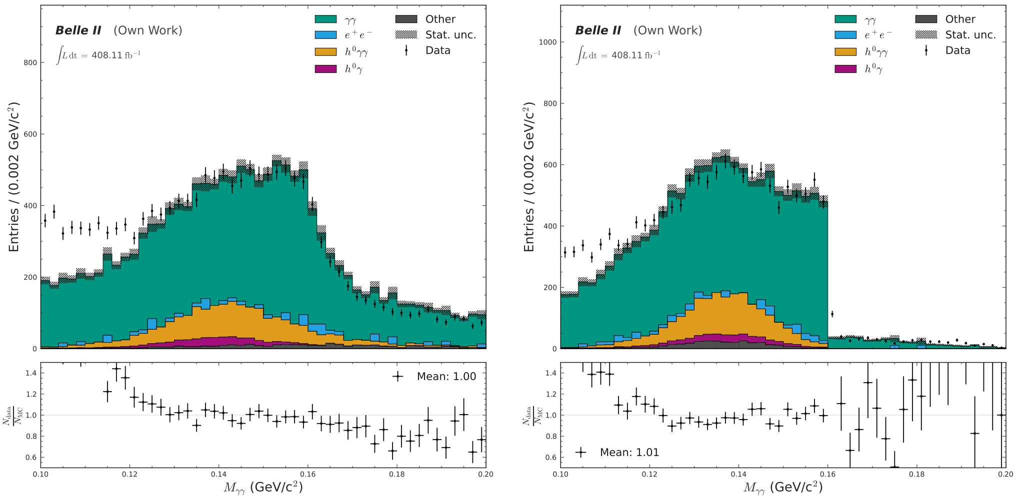
(d) After experiment 14 with `ggsel`

(e) After experiment 14 without `ggsel`

Figure 7.4: Distribution of the invariant diphoton mass, $M_{\gamma\gamma}$, in the inverted π^0-veto control region. The shown events are identical to those in Figure 7.3, but split for different experiment periods and L1 trigger selections, as indicated in the subfigure captions. Plot (a) shows the distribution for all events recorded before experiment 14, when `ggsel` was not yet available. Plots (b) and (c) show the distributions for events recorded during experiment 14, with and without the `ggsel` selection, respectively. Plots (d) and (e) show the distributions for events recorded after experiment 14, with and without the `ggsel` selection, respectively.

Kinematic Fit

As I discuss in Section 4.1.3, I apply a kinematic fit to all candidates to constrain the four-momenta of the final-state particles to the initial-state four-momentum. This fit significantly improves the $M_{\gamma\gamma}$ resolution and reduces the bias in the reconstructed $M_{\gamma\gamma}$



(a) With kinematic fit (b) Without kinematic fit

Figure 7.5: Distributions of the invariant diphoton mass, $M_{\gamma\gamma}$, in the inverted π^0-veto control region with (left) and without (right) the kinematic fit applied. The shown events are identical to those in Figure 7.3. Both histograms display the ratio between the recorded data and the simulation in the lower part of the plots.

value. Figure 7.5 compares the $M_{\gamma\gamma}$ distributions in this control region with and without the kinematic fit applied. For both distributions, I apply the same selection, including the L1 trigger and HLT selections and the exclusion of the merged photon classifier selection. I observe the same discrepancies between recorded data and simulation at low $M_{\gamma\gamma}$ values in both cases, indicating that the kinematic fit neither introduces nor mitigates these discrepancies.

7.2 η Region

In this and the following section, I demonstrate the agreement between recorded data and simulation for the irreducible background control regions, which are excluded from the signal extraction scan. In contrast to the previous section, I apply the full selection in these control regions, including the π^0-veto and the trigger selections. However, I exclude events from $e^+e^- \to q\bar{q}$ simulations, and omit the merged photon classifier selection. The η region is defined by the invariant diphoton mass window $M_{\gamma\gamma} \in [0.45, 0.63]\,\mathrm{GeV}/c^2$. This section primarily focuses on the overall MC/data agreement and on validating the trigger efficiency estimation.

Figure 7.6 shows the $M_{\gamma\gamma}$ distribution in the η control region. The overall agreement between recorded data and simulation is good, with the recorded data yield exhibiting an almost constant offset of approximately 3 %. I observe a similarly small but consistent discrepancy between recorded data and simulation across all kinematic distributions, as shown in Chapter D, which I use in the training of the kinematic classifier described in

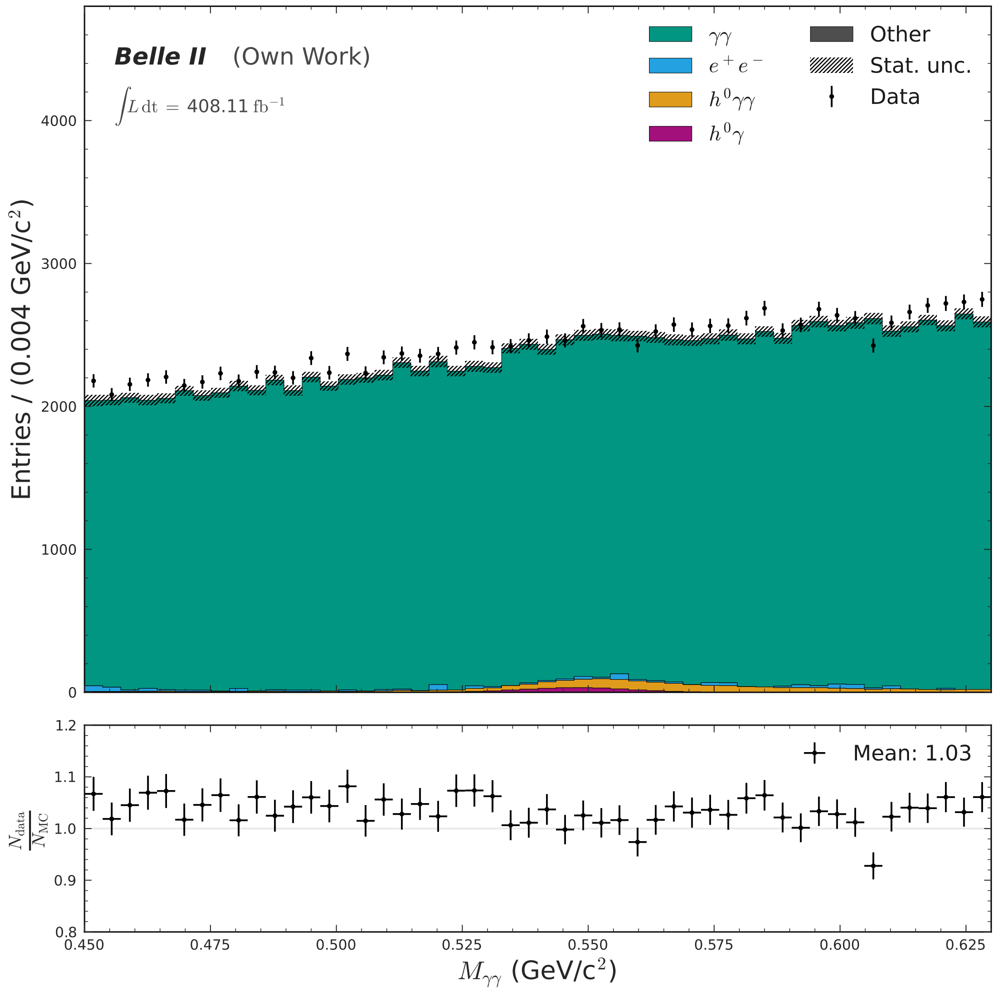
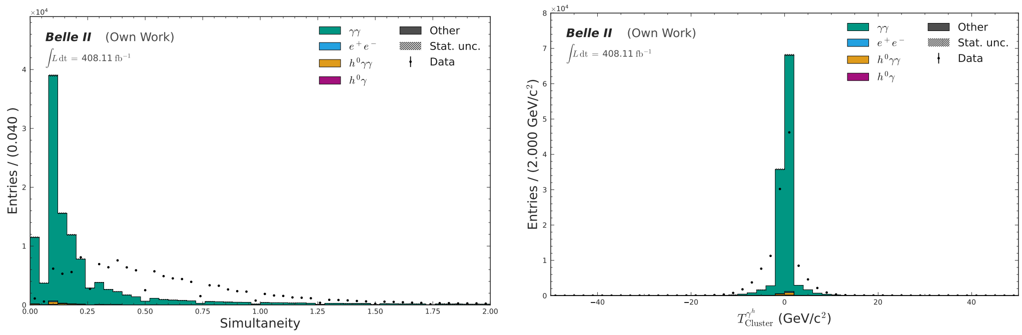


Figure 7.6: Distribution of the invariant diphoton mass, $M_{\gamma\gamma}$, in the η control region. The distribution includes events that pass all selection criteria, as well as the L1 trigger and HLT selections. All contributions from $e^+e^- \to q\bar{q}$ processes are excluded, and the merged photon classifier selection is omitted.

Section 4.4.2. Therefore, this discrepancy is not introduced by the kinematic classifier selection, as the classifier responds similarly to both simulation and recorded data. I discuss the distributions of the kinematic classifier, as well as the resulting systematic uncertainty due to the MC/data agreement and the applied selection, in Section 8.6.

I observe one exception to the otherwise consistent MC/data agreement in variables related to the cluster timing information of the photon candidates. However, these discrepancies are expected and well understood. Figure 7.7 shows two examples of such observables: the event simultaneity (defined in Section 4.1.1), and the cluster timing of the low-energy photon candidate in the event. The primary cause of the differences in cluster timing observables is the imperfect simulation of the ECL crystals. Variations in crystal geometry lead to differences in scintillation light propagation times. Yet, the unique shapes



(a) Event simultaneity (b) Cluster Timing γ^l

Figure 7.7: Distributions of cluster timing related observables, event simultaneity (left) and cluster timing of γ^l (right), in the η control region. The distributions include the same events as Fig. 7.6.

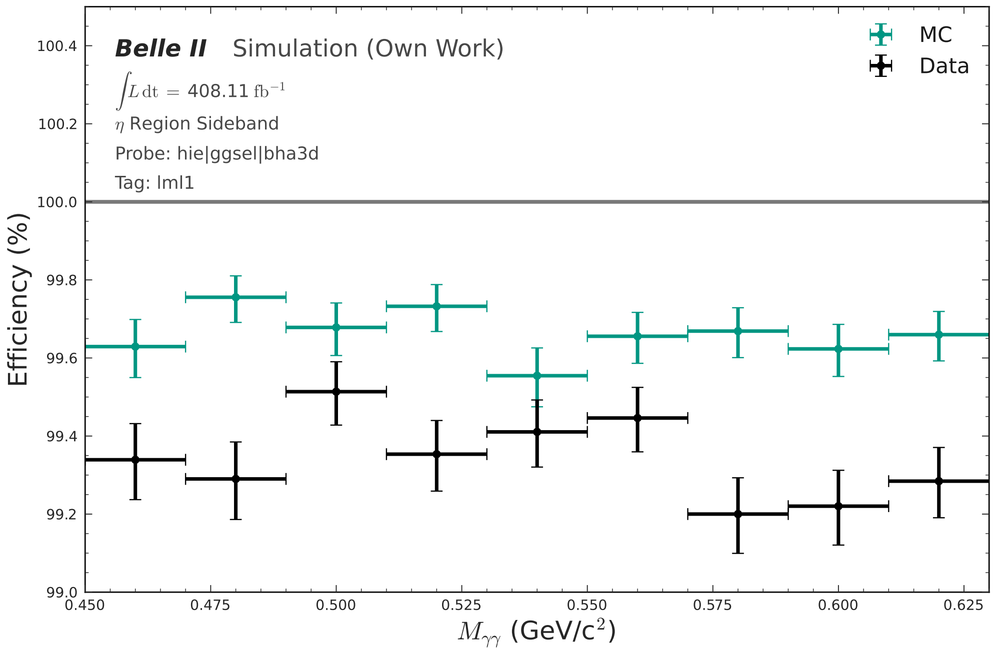


Figure 7.8: Trigger efficiency of (`hie`|`ggsel`|`bha3d`) with respect to `lml1` for HLT triggered events as a function of $M_{\gamma\gamma}$ in the η control region.

of individual crystals are not fully represented in the simulation.

As I describe in Sections 5.1.1 and 5.1.2, I determine the trigger efficiency using a data-driven method based on the tag-and-probe technique. Before unblinding, I test the trigger efficiency validation in the η and ω control regions, as these regions are kinematically closest to the signal region. Fig. 7.8 shows the trigger efficiency determination in the η control region for both data and simulation. The overall agreement between data and simulation is good, with absolute deviations of only about 0.5 %, well within the empirical expectation. Most importantly, the trigger efficiency can be reliably determined from

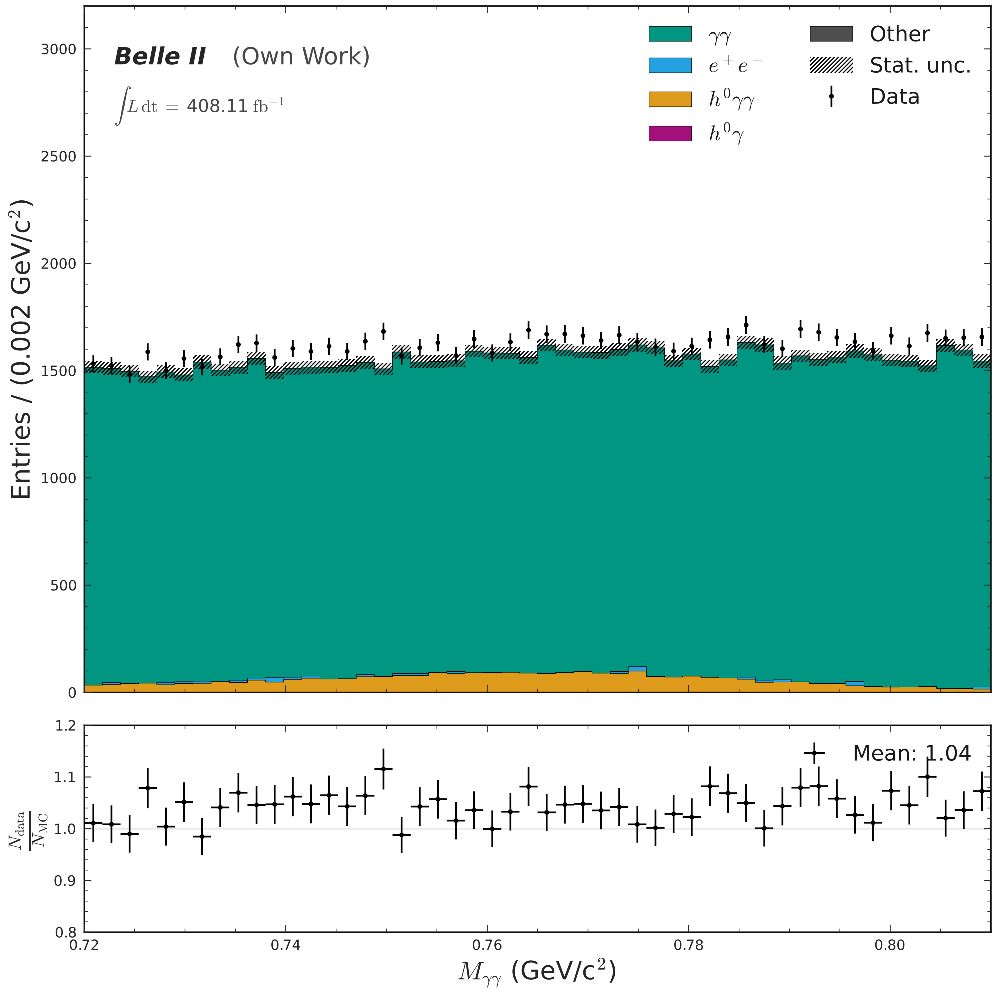


Figure 7.9: Distribution of the invariant diphoton mass, $M_{\gamma\gamma}$, in the ω control region. The distribution includes events that pass all selection criteria, as well as the L1 trigger and HLT selections. All contributions from $e^+e^- \to q\bar{q}$ processes are excluded, and the merged photon classifier selection is omitted.

recorded data in both control regions.

7.3 ω Region

Qualitatively, the observations in the ω control region are similar to those in the η control region described in Section 7.2. I obtain the following results in the same manner as for the η region, with the only difference being the $M_{\gamma\gamma}$ window defining the control region, which is set to $M_{\gamma\gamma} \in [0.72, 0.81]\,\mathrm{GeV}/c^2$ for the ω region.

Figure 7.9 shows the $M_{\gamma\gamma}$ distribution in the ω control region. As in the η region, the overall agreement between recorded data and simulation is good and exhibits a consistent offset; however, the offset amounts to approximately 4 % on average. Chapter D shows the distributions of the kinematic classifier input variables. I discuss the kinematic classifier

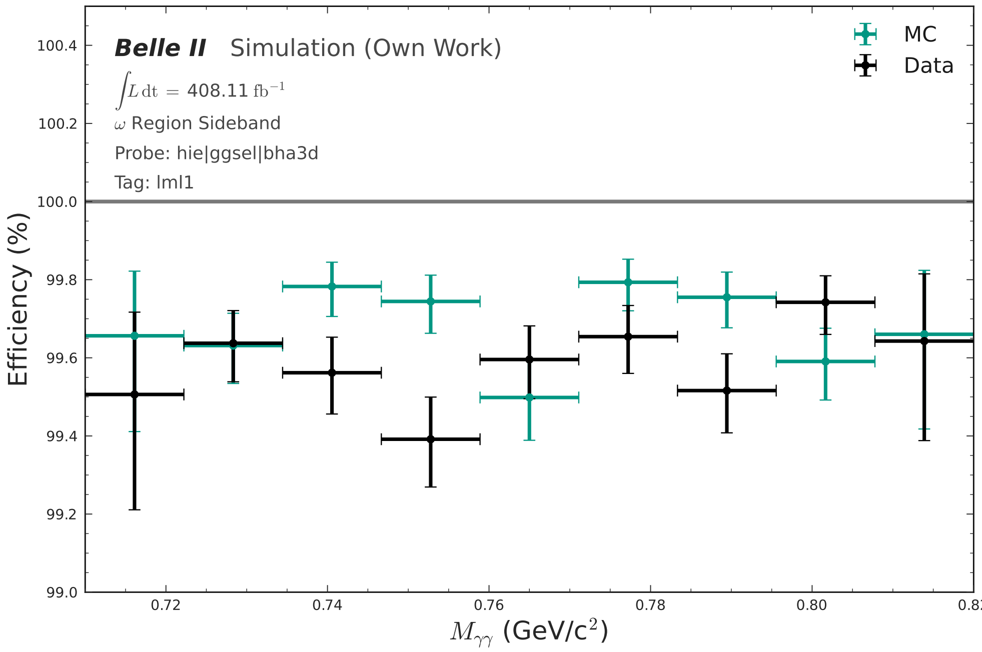


Figure 7.10: Trigger efficiency of (`hie`|`ggsel`|`bha3d`) with respect to `lml1` for HLT triggered events as a function of $M_{\gamma\gamma}$ in the ω control region.

output distribution and the corresponding systematic uncertainty due to the selection in Section 8.6.

Figure 7.10 shows the trigger efficiency determination in the ω control region for both data and simulation. The overall agreement between data and simulation is good and exhibits even better MC/data agreement than in the η control region (see Fig. 7.8).

Chapter 8

Corrections and Systematic Uncertainties

In this chapter, I summarise the corrections and systematic uncertainties considered in this analysis. The corrections introduce additional photon candidate weights aimed at improving the MC/data agreement. I propagate each photon candidate weight multiplicatively to the total event weight. The candidate weights, as well as their variations used to evaluate the systematic uncertainties, are provided centrally by the collaboration. In Sections 8.1 and 8.2, I discuss the impact of the corrections on the MC/data photon efficiency agreement and on the photon energy bias. In the following sections, I summarise the considered systematic uncertainties and their effects. Some of the corrections and systematic uncertainties can only be estimated using recorded data.

8.1 Photon Efficiency Correction

The photon efficiency correction that I present in this section is based on the work in [100]. In essence, the correction is derived from a comparison of the photon detection efficiency ε_γ in MC and data, using radiative muon pair events $e^+e^- \to \mu^+\mu^-\gamma$. The photon detection efficiency is defined as

$$\varepsilon_\gamma(p_{\text{recoil}}, \theta_{\text{recoil}}, \phi_{\text{recoil}}) = \frac{N(\text{Matched})}{N(\text{Selected})}, \tag{8.1}$$

where $p_{\text{recoil}}, \theta_{\text{recoil}}$, and ϕ_{recoil} are the momentum and the spatial coordinates of the dimuon recoil system, $N(\text{Selected})$ is the number of radiative muon pair events passing the selection, and $N(\text{Matched})$ is the number of selected events where the recoil momentum of the dimuon system could be matched with a photon candidate cluster in the ECL in the same direction and adequate energy. The included selections are:

- Exactly two oppositely charged tracks in the event,

- clusters associated with the tracks having less than 300 MeV energy and less than 80 % of the track momentum deposited as energy,
- tracks originating from the IP*,
- tracks having $p \geq 1\,\mathrm{GeV}/c$,
- the total invariant mass of the two-track recoil system has $m^2_{\mathrm{recoil}} < 2\,\mathrm{GeV}^2/c^4$,
- the total recoil momentum is $p_{\mathrm{recoil}} > 0.2\,\mathrm{GeV}/c$,
- and the photon candidate cluster has $E_\gamma > 75\,\mathrm{MeV}$ in the CM frame and an absolute cluster timing within $\pm 200\,\mathrm{ns}$.

The matching of the recoil momentum to a photon candidate cluster is done by requiring the 3D spatial position of the photon candidate to be within 0.1 rad of the recoil momentum direction and to satisfy $0.5 < E_\gamma/p_{\mathrm{recoil}} < 1.2$.

Correction photon candidate weights w_{eff} are derived as the ratio of the photon detection efficiency in data and MC:

$$w_{\mathrm{eff}}(p_{\mathrm{recoil}}, \theta_{\mathrm{recoil}}, \phi_{\mathrm{recoil}}) = \frac{\varepsilon_\gamma^{\mathrm{data}}(p_{\mathrm{recoil}}, \theta_{\mathrm{recoil}}, \phi_{\mathrm{recoil}})}{\varepsilon_\gamma^{\mathrm{MC}}(p_{\mathrm{recoil}}, \theta_{\mathrm{recoil}}, \phi_{\mathrm{recoil}})}. \tag{8.2}$$

The total event weight is then determined by multiplying the photon candidate weights and applying them to the MC events. For the background MC events, I multiply the total efficiency correction weight by the luminosity normalisation weights. As the impact of the overall correction on the kinematic variables is in total negligible for the selection procedure, I apply this correction for any step after Chapter 4. For the signal MC events specifically, I already consider the efficiency correction for the run-dependent signal efficiency interpolation in Section 5.2.

Figure 8.1 displays the relative photon efficiency correction as a function of ALP mass. I calculate the signal efficiency for each scanned ALP mass point with and without the photon efficiency correction applied after the interpolation for the different beam background conditions and show the relative difference. Moreover, I select all events in a $\pm 20\sigma^{\mathrm{CB}}$ (see Section 6.1) window around the ALP mass, which corresponds to the later signal extraction fit window.

To estimate the systematic uncertainty, the correction weight is provided with upwards and downwards fluctuations based on the statistical fluctuations of w_{eff}. I apply these variation weights to the signal MC samples and re-evaluated the run-dependent signal efficiency interpolation. Figure 8.2 shows the absolute relative variation of the signal efficiency due to the application of the statistical uncertainties of the photon efficiency correction. The upwards variations always have a stronger absolute influence on the signal

*For this selection, no exact definition is given in [100]. Typical collaboration selections are $\mathrm{d}r < 1\,\mathrm{cm}$ and $|\mathrm{d}z| < 3\,\mathrm{cm}$.

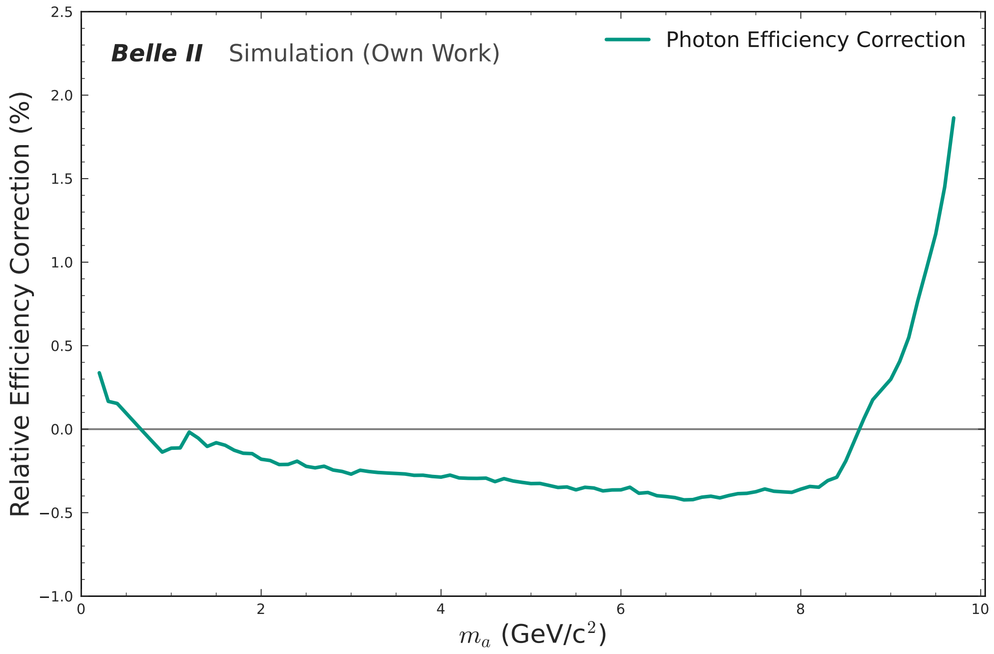


Figure 8.1: Relative correction of the signal efficiency due to the MC/data photon efficiency correction as a function of the ALP mass. The efficiencies with and without corrections are determined in a $\pm 20\sigma^{\mathrm{CB}}$ signal window.

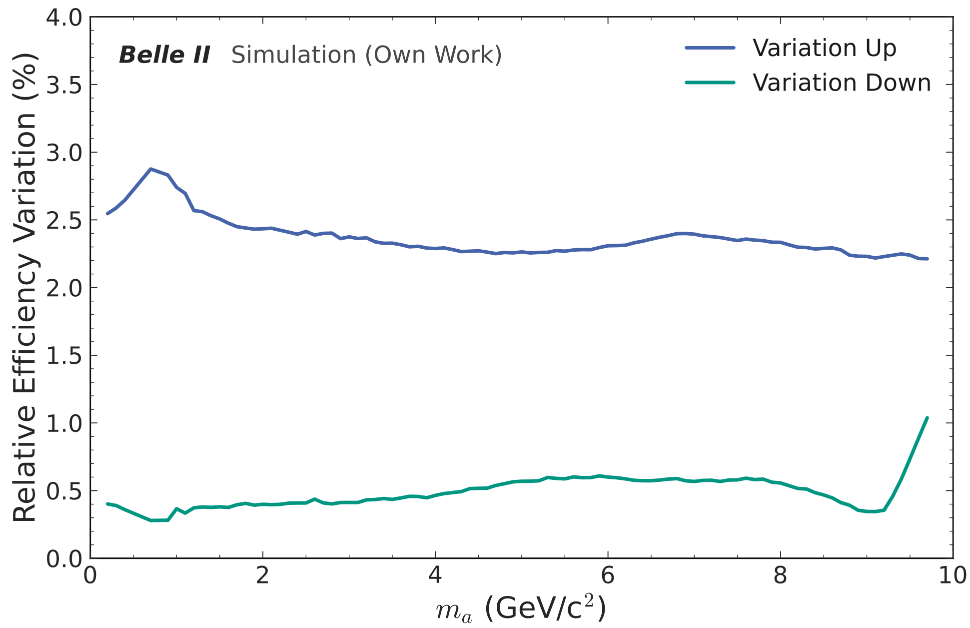


Figure 8.2: Absolute relative variation of the signal efficiency due to the MC/data photon efficiency correction upwards (blue) and downwards (green) weight variations as a function of the ALP mass. The efficiencies are determined in a $\pm 20\sigma^{\mathrm{CB}}$ signal window.

efficiency compared to the downwards variation. However, the upwards variations agree well with the expected variation, which is roughly 0.8 % per photon. For a conservative estimation, I chose always the higher absolute relative efficiency variation, which is the upwards variation, and propagate it as systematic uncertainty on the corrected signal efficiency.

8.2 Energy Bias Correction

The photon energy bias correction that I present in this section is based on the work in [101]. In essence, the energy bias correction is determined via the MC and data comparison of the peaking $M_{\gamma\gamma}$ position in symmetric $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ decays as a function of the photon energy. Particularly, the event weights w_{bias} are derived as the ratio of the $M_{\gamma\gamma}$ peak position in data and MC:

$$w_{\text{bias}} = \frac{\hat{M}_{\gamma\gamma}^{\text{data}}}{\hat{M}_{\gamma\gamma}^{\text{MC}}}, \tag{8.3}$$

where $\hat{M}_{\gamma\gamma}^{\text{data}}$ and $\hat{M}_{\gamma\gamma}^{\text{MC}}$ are the fitted $M_{\gamma\gamma}$ peak positions in data and MC, respectively, in bins of the photon energy E_γ. The events that enter the fit pass the following selection:

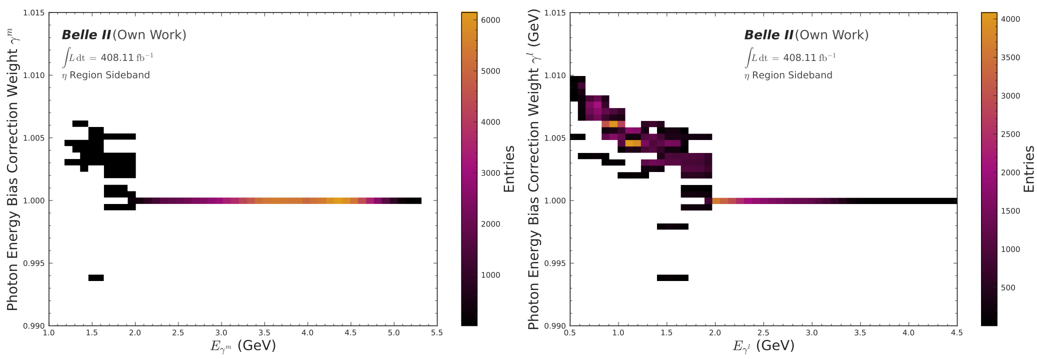
- Photon candidates that pass the *loose*[*] requirement and are located within the ECL barrel or endcap regions,
- $M_{\gamma\gamma} \in [0.08, 0.7]\,\text{GeV}/c^2$,
- and that the photon candidate energies E_{γ^1}, E_{γ^2} are symmetric, meaning that

$$\left|\frac{E_{\gamma^1} - E_{\gamma^2}}{E_{\gamma^1}}\right| < 0.05. \tag{8.4}$$

The final photon candidate weights w_{bias} are determined by fitting the $M_{\gamma\gamma}$ distribution for three different topological cases, and then the total average is determined. The three topological cases are: the photon pair has no ECL region restriction, both photons are in the ECL barrel, and one photon is in the forward endcap and the other in the barrel region. Similar to the photon efficiency correction, I determine the total event weight by multiplying the photon candidate weights. However, I apply the energy bias correction only to events from recorded data.

Considering the selection, the photon candidate weights are restricted to photon energies $E_\gamma < 2.0\,\text{GeV}$. Consequently, the photon energy bias correction has a negligible effect on the events passing this analysis's selection, due to the generally high photon candidate

[*]The exact definition of the *loose* requirement can no longer be retraced. From the general context of [101] and assuming that the requirements have not changed significantly over time, the selections are estimated to be that the θ angle of the photon candidates is in the CDC acceptance, the timing fit did not fail, the photon candidate cluster energy ratio passes $E_1/E_9 > 0.4$ or $E > 75\,\text{MeV}$.



(a) Middle-energy photon candidate (b) Low-energy photon candidate

Figure 8.3: 2D distributions of the photon energy bias correction weight and the cluster energy for the middle- (left) and low-energetic (right) photon candidates. The distributions are based on the events from the recorded data in the η control region.

energies. Figure 8.3 displays the photon candidate weights as a function of the photon candidate energy for the middle-energy and low-energy photon candidates in the event – for the high-energy photon candidates, no weight is given due to their high energy. The weights that I show here are taken from the η sideband selection described in Section 7.2. Generally, the weights rarely exceed a 1 % correction. Due to the negligible impact of the energy bias correction on the final event selection, I do not assign a systematic uncertainty due to the energy bias correction.

8.3 Photon Resolution

As the simulation can not represent the detector perfectly, differences in the photon candidate energy and position resolutions can arise. Different resolutions can impact the performance of the kinematic fit, which in turn can affect the signal shape of the post-kinematic fit $M_{\gamma\gamma}$ distribution. I briefly outline how I estimate the resolution in simulation and data, and how I evaluate the impact of potential differences on the signal efficiency.

The estimation of the photon energy resolution is motivated by the work in [59]. In short summary, I use radiative dimuon events $e^+e^- \to \mu^+\mu^-\gamma$ to estimate the photon energy resolution in both simulation and recorded data. In a perfect scenario, due to the energy and momentum conservation, the photon energy $E(\gamma)$ needs to be equivalent to the recoil momentum of the dimuon system $p_{\text{recoil}}(\mu\mu)$. This equivalence is broken in the real world due to detector resolution effects and additional radiation. Particularly, the difference arises due to the resolution of the photon energy, the tracking resolution of the muon tracks, and additional ISR and FSR. The total parametrisation can be expressed as

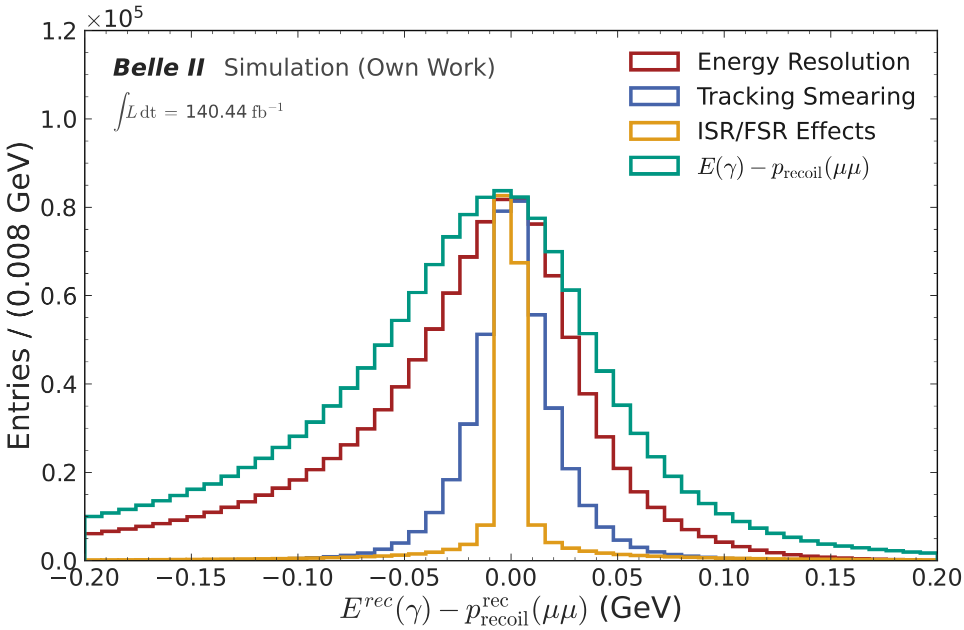


Figure 8.4: Distribution of the difference between the reconstructed photon energy and the recoil momentum of the dimuon system $E^{\mathrm{rec}}(\gamma) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)$ (green) for simulated radiative dimuon events. The individual components from Eq. (8.5) are shown as well: the photon energy resolution (blue), the tracking resolution (orange), and the additional radiation (red) and are normalised to the total distribution.

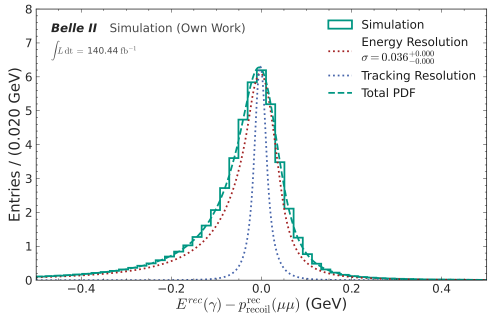
$$\begin{aligned} E^{\mathrm{rec}}(\gamma) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu) = &\Big(E^{\mathrm{rec}}(\gamma) - E^{\mathrm{true}}(\gamma)\Big) \otimes \\ &\Big(p^{\mathrm{true}}_{\mathrm{recoil}}(\mu\mu) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)\Big) \otimes \\ &\Big(E^{\mathrm{true}}(\gamma) - p^{\mathrm{true}}_{\mathrm{recoil}}(\mu\mu)\Big)\,, \end{aligned} \tag{8.5}$$

where the first term on the right-hand side represents the photon energy resolution, the second term the tracking resolution of the muon tracks, and the third term the additional radiation. In this notation, "rec" refers to reconstructed quantities and "true" to the MC truth information. All effects are folded (denoted by $\otimes$) together in the observed distribution of $E^{\mathrm{rec}}(\gamma) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)$. Figure 8.4 shows the distribution of $E^{\mathrm{rec}}(\gamma) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)$ for simulated dimuon events with photon energies $E^{\mathrm{rec}}(\gamma) > 1\,\mathrm{GeV}$ and the photon in the ECL barrel. In principle, this distribution needs to be studied in bins of the photon energy and angles $\theta(\gamma)$, $\phi(\gamma)$ to capture potential dependencies of the resolution. Furthermore, the tracking resolution would need to be determined in recorded data and the ISR contribution would need to be modelled accurately in simulation. Additionally, all the effects would need to be studied as a function of the beam background conditions. However, for this analysis, I perform a simplified study to estimate potential differences in the photon energy resolution between simulation and recorded data by assuming that the tracking smearing is well modelled in simulation and the ISR contribution is negligible in comparison to the resolution effects.

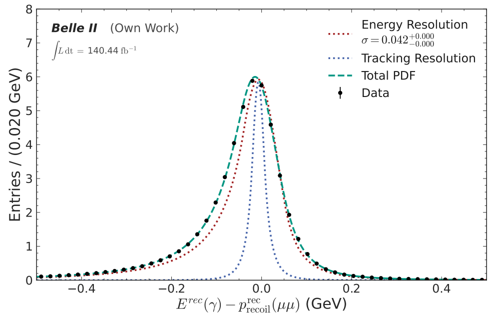
To estimate the photon resolutions, I reconstruct dimuon events from experiments 24 and 26 in both simulation and recorded data. These experiments have the highest recorded beam background conditions in the Run 1 dataset (see Fig. 5.12), hence providing a conservative estimate of potential resolution differences. For the reconstruction, I apply the selections:

- **μ-Candidate**
  - Momentum $p > 1\,\mathrm{GeV}/c$,
  - Originating from the IP ($\mathrm{d}r < 0.5\,\mathrm{cm}$, $\mathrm{d}z < 2\,\mathrm{cm}$),
  - θ in the CDC acceptance ($17^\circ < \theta < 150^\circ$),
  - More than 0 hits in the CDC,
  - Cluster energy $0.05\,\mathrm{GeV} < E < 0.25\,\mathrm{GeV}$,
- **Photon-Candidate**
  - Energy $E_\gamma > 1\,\mathrm{GeV}$,
  - In the ECL barrel ($32.2^\circ < \theta < 128.7^\circ$),
  - Absolute cluster timing within $|T| < 200\,\mathrm{ns}$,
  - At least 1.5 cluster hits in the ECL,
  - Separated more than 50 cm frm the closest track
- **Event-Candidate**
  - Invariant dimuon mass $M_{\mu\mu} > 1\,\mathrm{GeV}/c^2$,
  - invariant $\mu^+\mu^-\gamma$ mass $M_{\mu\mu\gamma} < 10.58\,\mathrm{GeV}/c^2$,
  - $-0.1 < \theta^{\mathrm{rec}}(\gamma) - \theta^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu) < 0.1$,
  - $-0.1 < \phi^{\mathrm{rec}}(\gamma) - \phi^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu) < 0.1$, and
  - $0.5 < \frac{E^{\mathrm{rec}}(\gamma)}{p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)} < 1.5$.

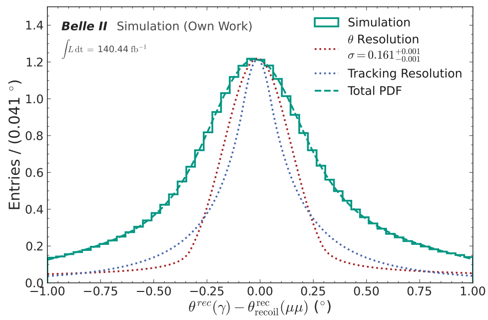
For the estimation of the tracking smearing, I fit the distribution of $p^{\mathrm{true}}_{\mathrm{recoil}}(\mu\mu) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)$ in simulated dimuon events. Since the distribution has no physically motivated shape, I use a DSCB PDF to model the distribution. For the estimation of the photon energy resolution, I fit the distribution of $E^{\mathrm{rec}}(\gamma) - p^{\mathrm{rec}}_{\mathrm{recoil}}(\mu\mu)$ in both simulation and recorded data using the convolution of the fixed tracking smearing PDF and a floating DSCB PDF to model the photon energy resolution. The width σ of the photon energy resolution PDF describes the photon energy resolution. Figures 8.5a and 8.5b show the fits determining the energy resolution in simulation and data.



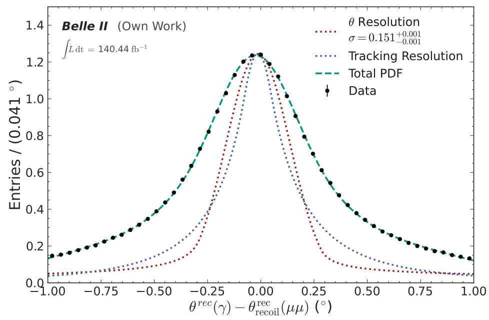
(a) Energy resolution simulation



(b) Energy resolution data



(c) θ resolution simulation

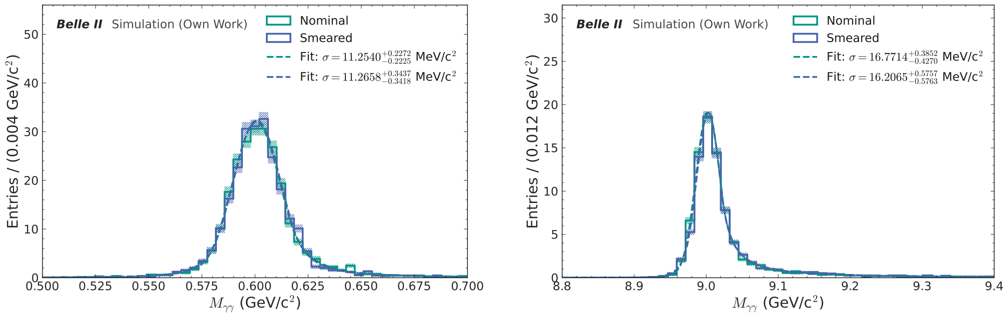


(d) θ resolution data

Figure 8.5: Distributions and fits to determine the photon energy and θ angle resolutions in simulation and recorded data. The top row shows the determination of the energy resolution in simulation (left) and recorded data (right). The bottom row shows the determination of the θ angle resolution in simulation (left) and recorded data (right). The solid green histograms show the simulated events, and the black error bars show the recorded data events. The solid green line shows the total fit PDF, the dotted red line shows the photon energy or θ resolution component, and the dotted blue line shows the fixed tracking resolution component.

Taking the width difference between simulation and recorded data, I find that the photon energy resolution in recorded data is lower by about 17 % compared to the simulation. The resolution in the θ angle of the photon candidate is found to be slightly higher in recorded data by about 9 % compared to simulation. Figures 8.5c and 8.5d show the fits determining the θ resolution in simulation and data, the ϕ resolution exhibits the same results as θ. Accounting for the assumptions made in this simplified study, the resolution for θ and ϕ can be deemed equal in simulation and recorded data.

To account for the different photon energy resolution in simulation and recorded data, I smear the photon energy in simulated signal events and determine the width of the signal in $M_{\gamma\gamma}$ after the kinematic fit. Figure 8.6 shows the comparison of the signal widths before and after smearing for two example ALP mass points. For the signal samples $m_a = 0.6\,\mathrm{GeV}/c^2$, the kinematic fit does not change the $M_{\gamma\gamma}$ distribution significantly compared to the pre-kinematic fit distribution. For the signal samples $m_a = 9.0\,\mathrm{GeV}/c^2$, the kinematic fit improves the resolution significantly. In both cases, the smearing of the



(a) $m_a = 0.6\,\mathrm{GeV}/c^2$ (b) $m_a = 9.0\,\mathrm{GeV}/c^2$

Figure 8.6: Distribution of the post-kinematic fit $M_{\gamma\gamma}$ for two example ALP mass points before (green) and after (blue) smearing the photon energy in simulated signal events to account for the 17 % difference between energy resolutions in simulation and recorded data. The solid lines show the fit PDF to determine the signal width.

photon energy changes the signal width only marginally within the statistical uncertainty of the width parameters. In the case of the higher mass point, the smearing even leads to a slightly lower signal width due to the kinematic fit compensating for the worse photon energy resolution. Since resolution differences have a negligible impact on the signal width after the kinematic fit, I do not assign a systematic uncertainty to the signal efficiency due to potential differences in the photon energy resolution between simulation and recorded data.

8.4 Trigger Efficiency

I determine the trigger efficiency using a data-driven method based on the tag-and-probe technique described in Sections 5.1.1 and 5.1.2. I estimate the systematic uncertainty of the trigger efficiency by the statistical uncertainty of the efficiency calculation on data for each ALP scan mass point. Section 10.2 summerises the results of the trigger efficiency determination and Fig. 10.4 shows the stastical uncertainty.

8.5 Run-Dependent Interpolation

Section 5.2 describes the interpolation procedure to account for the approximately linear dependence of the signal efficiency on the mean number of out-of-time crystals $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$. I evaluate the signal efficiency at the mean $N_{\mathrm{OOT}}^{\mathrm{Crys.}} = 102.2$ of the Run 1 dataset. For this, I propagate the uncertainty of the linear fit, given in Eq. (5.3), as a systematic uncertainty on the signal efficiency.

For every ALP mass point, I apply the interpolation procedure and extract the uncertainty relative to the total interpolated signal efficiency. Figure 8.7 shows the relative uncertainty as function of the scan ALP mass. Overall, the relative uncertainty is stable

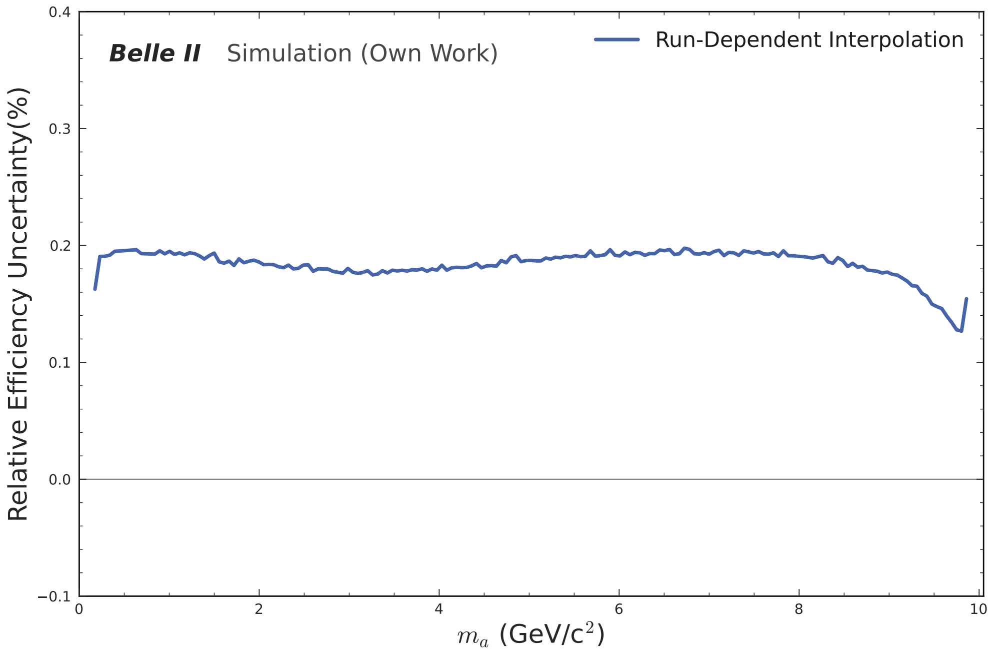


Figure 8.7: Statistical uncertainty of the run-dependent signal efficiency interpolation in Section 5.2 relative to the interpolated efficiency as a function of the ALP mass.

around 0.2 % with slight fluctuations towards very low and high ALP masses.

8.6 Selection

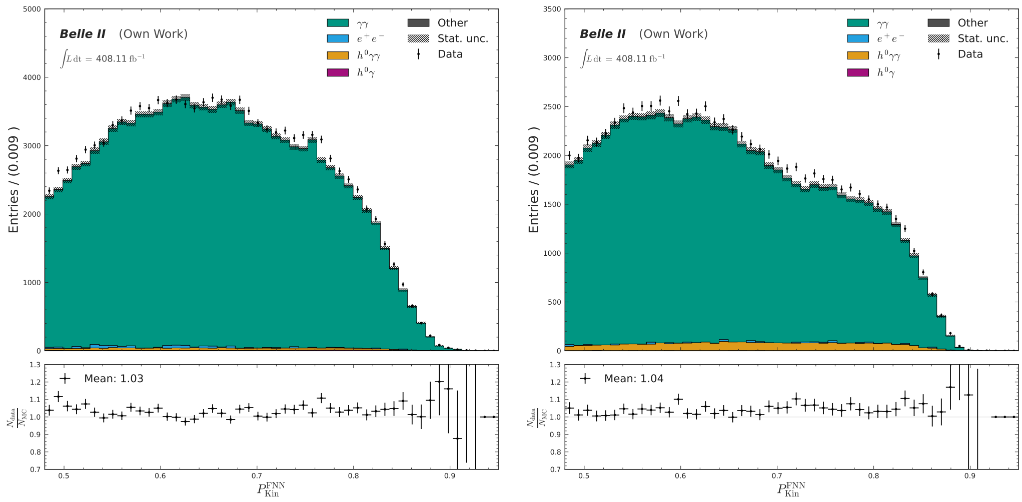
Differences in the simulation with respect to the recorded data arise through imperfect modelling of the underlying physical model as well as the detector response. In this and the following sections, I summarise the systematic uncertainties considered to address these differences. Particularly, in this section, I discuss the overall uncertainty I estimate due to the event selection. To simplify the discussion, the selection procedure can be divided into two main parts: the selections before the MVA-based classifiers and the selections based on the classifiers.

The selections before the classifiers consist of basic quality criteria on the photon and event candidates, as well as vetoes to suppress specific background processes. These selections are either loose enough to account for known differences between simulation and data, such as cluster timing, or they target well-known physical processes, such as the π^0 or pair conversion vetoes. The only exception is the selection on the energy of the possible fourth photon candidate in the event, which is affected by the ISR modelling of the signal process as implicitly considered in the following section.

Selections based on the classifiers are driven by overall modelling of the input variables, and the effect is summarised by the classifier output. Ideally, one would need to evaluate the response of the classifier in data for the signal process, which is not possible due to this analysis being a search for new physics. Instead, I evaluate the classifier response in data and simulation in the η and ω control regions defined in Chapter 7. I choose these control regions as they are able to represent best the events passing all selections, while being excluded from the signal extraction scan. For the final uncertainty estimation, I focus on the classifier input and output distributions of the kinematic classifier described in Section 4.4.2. I do not consider the merged photon classifier selection for any of the following studies due to its removal from the selection chain (see Section 7.1). In addition to the selections from Chapter 4, I also require the events to pass the L1 trigger and HLT selections as described in Sections 5.1.1 and 5.1.2.

The distributions for the kinematic classifier input variables in the η and ω control regions are shown in Chapter D. Overall, the agreement between simulation and data is equally good as the agreement shown for $M_{\gamma\gamma}$ in Sections 7.2 and 7.3. The same consistent picture is present for the kinematic classifier output distributions in Fig. 8.8. I apply a total systematic uncertainty of $4\,\%$ to the signal efficiency based on the larger overall weighted mean MC/data ratio of the control regions.

To make sure that a signal-like event selection is represented by the control regions, I also evaluate the effect of the kinematic classifier on signal-like distributions. For this study, I use events generated with an ALP mass of $m_a = 0.6\,\mathrm{GeV}/c^2$, which is in the middle of the η control region and is never part of any selection optimisation due to the exclusion of the η mass region. The simulated background events and the recorded data in this study correspond to the same events used in Section 7.2. To mimic the signal behaviour, I

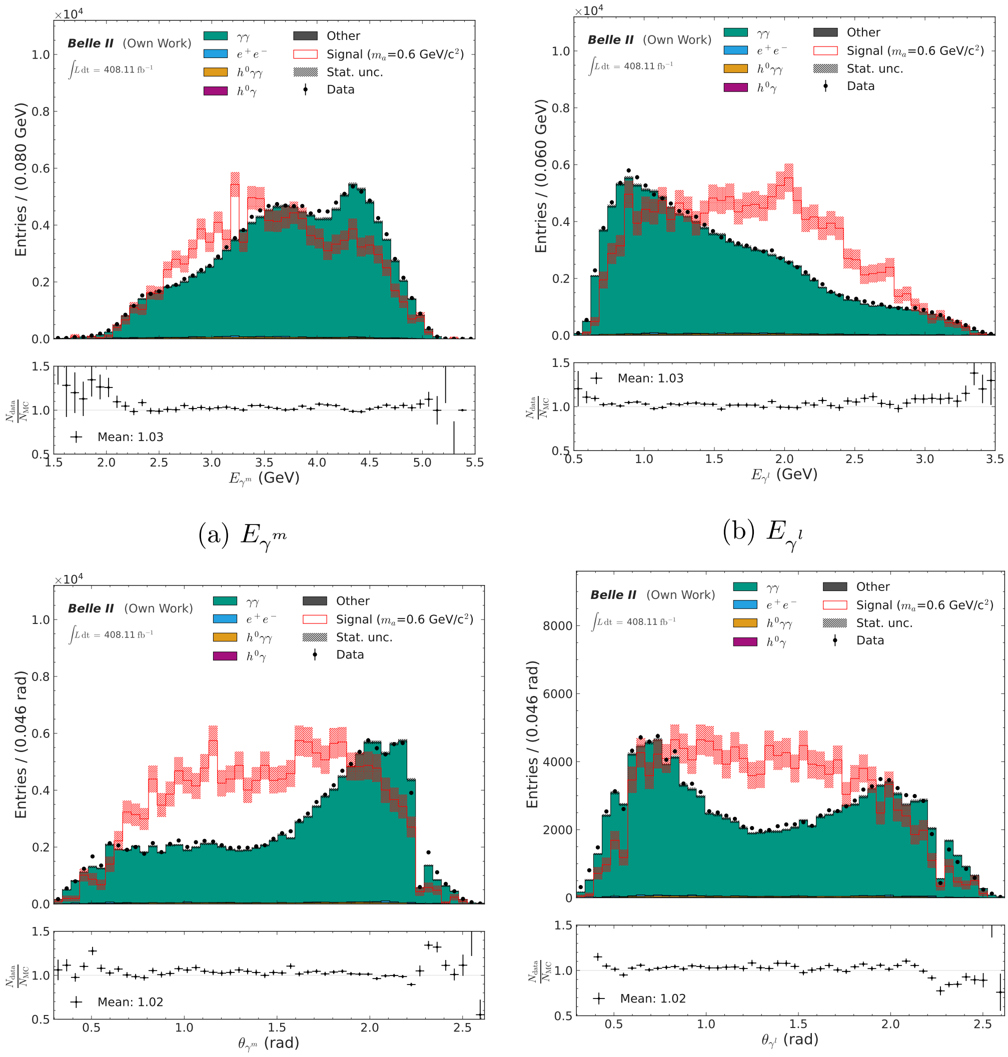


(a) η control region (b) ω control region

Figure 8.8: Distribution of the kinematic classifier output $P_{\mathrm{Kin}}^{\mathrm{FNN}}$ in the η (left) and ω (right) control regions. Filled, coloured histograms correspond to the expectation from the simulated events in these regions, and black error bars correspond to the recorded data. Both histograms display the ratio between the recorded data and the simulation in the lower part of the plots. The distributions are shown for events that pass all selections, excluding the merged photon classifier, as well as the L1 trigger and HLT selections. Simulated events are weighted additionally according to the photon efficiency correction, and data events are weighted according to the photon energy bias correction.

choose four variables that define the signal characteristics, calculate a weight which shapes the simulated background distributions to the signal distributions, and apply this weight to both the simulated background and recorded data events. The four chosen variables are the energy of the middle- and low-energy photon candidates, E_{γ^m} and E_{γ^l}, and the θ angles of the middle- and low-energy photon candidates, θ_{γ^m} and θ_{γ^l}. I choose these variables, as at this ALP mass hypothesis, the ALP candidate is mainly formed by the middle- and low-energy photon candidates while high-energy photon candidate corresponds to the recoiling photon. Figure 8.9 shows the distributions of the chosen variables.

To determine the weights in four dimensions, I estimate the PDF in the four-dimensional space using a Gaussian kernel density estimation [102, 103] for the simulated background and signal events separately. The resulting weights are then given by the ratio of the two PDFs. Figure 8.10 shows the distributions of the four variables after applying the weights to both the simulated background and recorded data events. Overall, the agreement between the background simulation and recorded data remains good after applying the signal-like weights. Finally, Figure 8.11 shows the comparison of the kinematic classifier output before and after applying the signal-like weights. Also here, the agreement between simulation and data remains good after applying the signal-like weights and the distribution of the simulated background and recorded data events follows closely the distribution of the simulated signal events. Consequently, I do not assign an additional systematic uncertainty



(a) E_{γ^m}

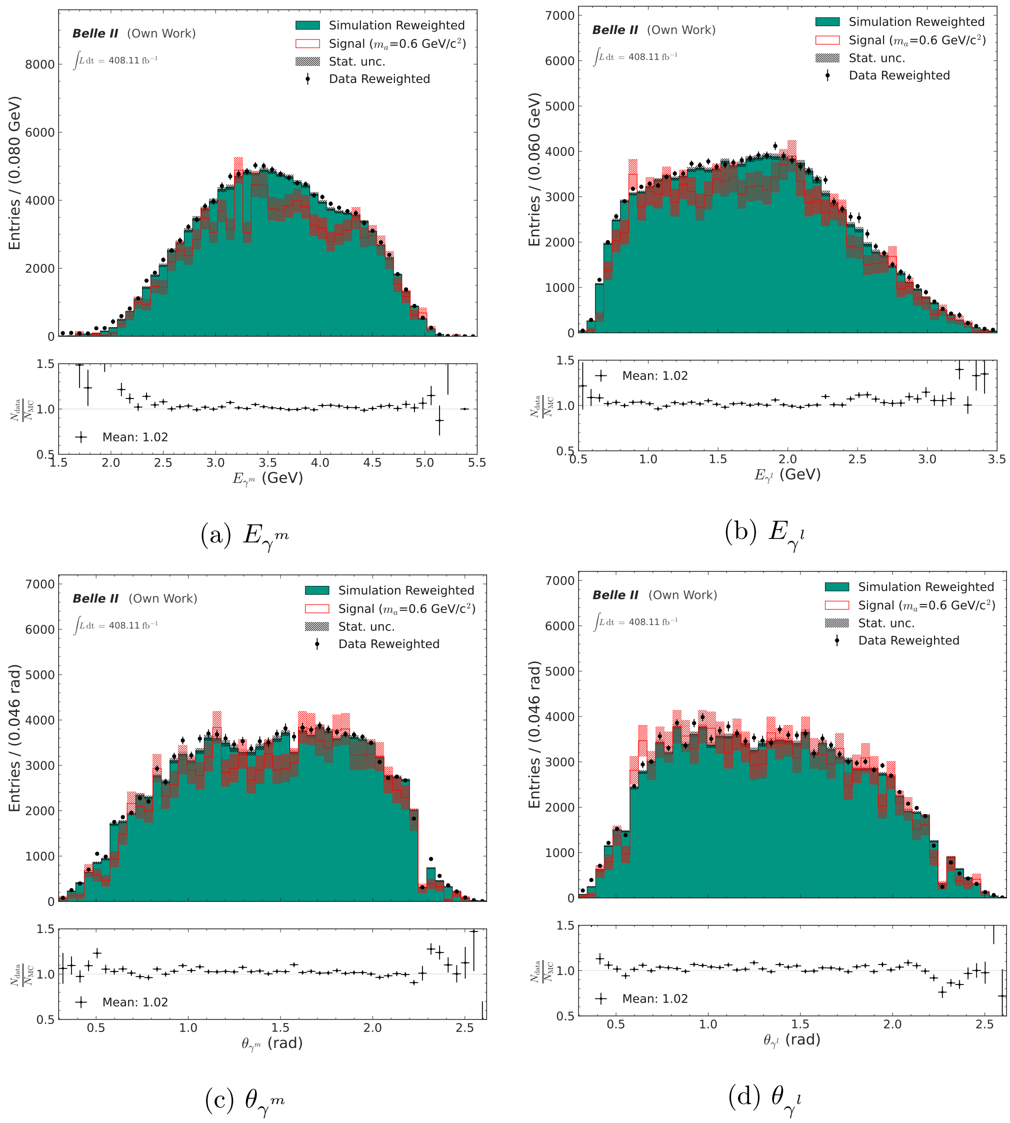
(b) E_{γ^l}

(c) θ_{γ^m}

(d) θ_{γ^l}

Figure 8.9: Distribution of the energy and θ of the middle- and low-energy photon candidate in the η control region. Filled, coloured histograms correspond to the expectation from the simulated events in these regions, and black error bars correspond to the recorded data. The red unfilled histograms correspond to the expectation from the simulated signal events with an ALP mass of $m_a = 0.6\,\mathrm{GeV}/c^2$.

to the signal efficiency due to the event selection based on this study. The only assigned uncertainty remains the 4 % uncertainty based on the overall MC/data ratio in the control regions.



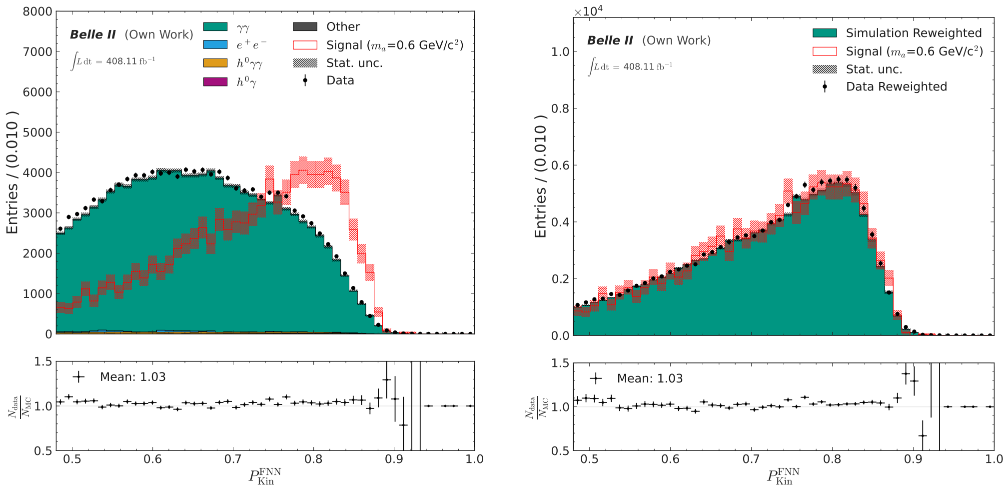
(a) E_{γ^m}

(b) E_{γ^l}

(c) θ_{γ^m}

(d) θ_{γ^l}

Figure 8.10: Distribution of the energy and θ of the middle- and low-energy photon candidate in the η control region. The same events as in Figure 8.9 are shown, but with additional weights applied to both the simulated background and recorded data events to mimic signal-like distributions.



(a) Before reweighting

(b) After reweighting

Figure 8.11: Distribution of the kinematic classifier out $P_{\text{Kin}}^{\text{FNN}}$ in the η control region. The same events as in Figures 8.9 and 8.10 are shown. The left plot shows the distribution before applying the signal-like weights, and the right plot shows the distribution after applying the weights.

8.7 ISR Modelling

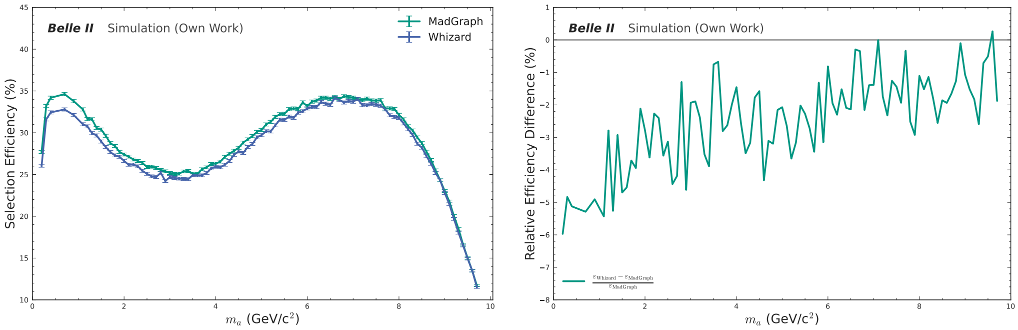
As I outline in Section 3.4.1, the signal MC samples are generated with the `MadGraph5 aMC@NLO` ISR option to account for momentum changes in the final state. However, no additional photons are generated by `MadGraph5 aMC@NLO`, and only the next-to-leading order effects are estimated by the radiative corrections. In this section, I estimate the uncertainty related to the ISR modelling of the signal process.

The simplification by the `MadGraph5 aMC@NLO` radiative corrections can lead to two aspects that need to be considered as systematic uncertainties under the assumption that the final state kinematics is correct at the first order. Firstly, since no real fourth, radiative photon enters the detector simulation, the effect of an additional photon on the event selection cannot be considered. A direct impact could be expected for the discussion of the fourth photon energy selection in Section 4.2.3. Here, the fourth photon energy distribution for signal events is only non-zero due to beam background effects. In the presence of real, hard ISR photons, the distribution could be altered and lead to a higher signal efficiency loss. However, the effect is expected to be small since most high-energy ISR photons are emitted collinear to the beam axis and thus escape detection. Secondly, since the `MadGraph5 aMC@NLO` radiative corrections only account for next-to-leading order effects, higher-order corrections could also impact the overall event kinematics. Different kinematics would impact the entire selection chain, mainly the distributions of the kinematic classifier input variables and thus the classifier output.

To estimate the overall effect of the ISR modelling, I compare the signal efficiency for events that are generated with `MadGraph5 aMC@NLO` and those generated with `WHIZARD` [104, 105]. Other than `MadGraph5 aMC@NLO`, `WHIZARD` includes the emission of additional photons in the event generation by always producing two additional photons in the final state. All final state particle four-momenta, including the additional photons, are adjusted in the presence of real, hard ISR photon emission. If for one event, no ISR photons are emitted, the additional photons will have numerically negligible momentum. For a fair comparison, I use the same theory model as input for both generators.

As I previously pointed out, changes due to the ISR modelling can impact the entire selection chain. Consequently, I compare the signal efficiency for the two different generators after the full event selection. The full event selection includes all event and candidate selections as described in Chapter 4, excluding the merged classifier selection as described in Section 7.1. Furthermore, I also require the events to pass the L1 trigger and HLT selections as described in Sections 5.1.1 and 5.1.2 and apply the MC corrections from Section 8.1. To ensure a fair comparison, I generate the `WHIZARD` samples with the same run and experiment conditions as the `MadGraph5 aMC@NLO` samples and interpolate the signal efficiency for different beam background conditions.

Figure 8.12a compares the signal efficiency for the two different generators as a function of the ALP mass. Generally, the signal efficiencies show the same ALP mass dependence,



(a) Signal Efficiency Comparison (b) Relative Efficiency Difference

Figure 8.12: The left plot compares the signal efficiencies after the run-dependent interpolation for the `MadGraph5 aMC@NLO` (green) and `WHIZARD` (blue) generators. The right plot shows the relative difference in the efficiencies.

which excludes large differences in the kinematic distributions. However, the `WHIZARD` efficiencies are consistently lower than the `MadGraph5 aMC@NLO` efficiencies, specifically for low ALP masses where the maximum absolute difference can reach up to 3 %. Stronger differences for low ALP masses are expected since the phase space for additional photon emission is larger, as not the majority of the energy-momentum is used to produce the ALP. Figure 8.12b displays the relative efficiency difference between the two generators for a few selected ALP masses. As expected by the absolute signal efficiency comparison, the relative differences are largest for low ALP masses with up to 6 % difference and decrease towards higher masses. For the systematic uncertainty estimation, I calculate the relative difference for all ALP mass scan points and propagate the difference to the signal efficiency.

8.8 Signal Shape

As a New Physics (NP) process cannot be studied on recorded data before it is found in recorded data, there is no other option than to rely on the MC simulation to estimate the signal shape. In this section, I estimate the systematic uncertainty on the signal shape. The signal parametrisation consists of two components: the main peaking component and the combinatorial component, which is mainly relevant for ALP mass hypotheses roughly in the range of $m_a \in [3.5, 8.0]\,\mathrm{GeV}/c^2$.

For the peaking component, I use the DSCB PDF to describe the peak and its tails. Due to the asymmetric shape of the $M_{\gamma\gamma}$ peaking signal distribution with tails on both sides, no other PDF, such as a Gaussian, double Gaussian or CB, is feasible to consider. Hence, I estimate the systematic uncertainty by studying the fit uncertainties of the DSCB parameters that arise from the discussion in Section 6.1. For the combinatorial component, I use a first-order Chebyshev polynomial with a free parameter to determine the fraction between the peaking and non-peaking components. Due to the overall small contribution of the combinatorial component for most ALP mass hypotheses, I do not assign a systematic

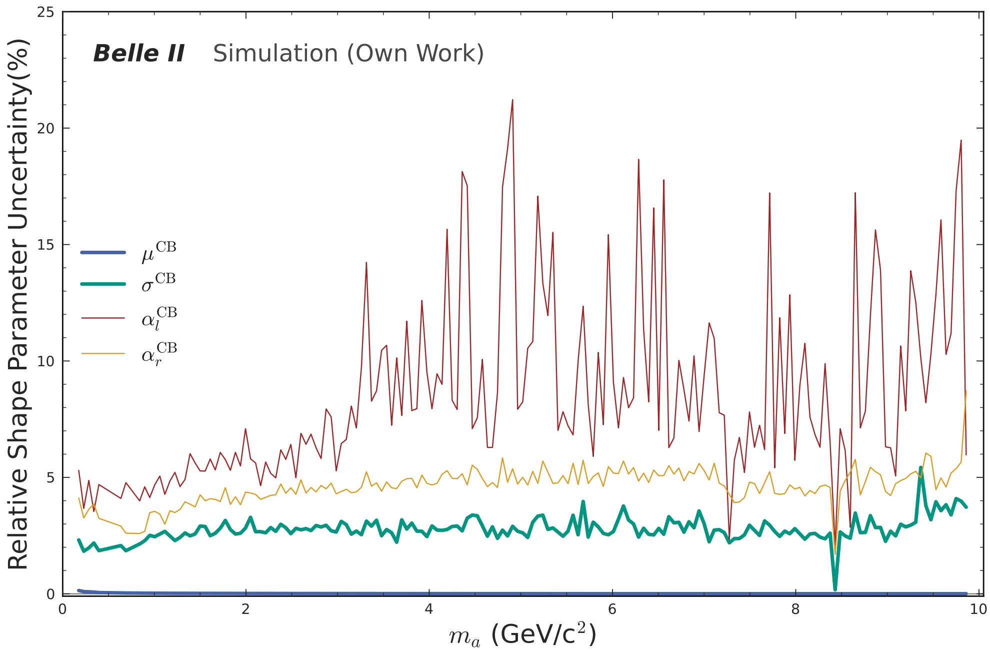


Figure 8.13: Peaking signal PDF parameter fit uncertainties relative to the fitted central parameter value. The red and yellow lines show the relative uncertainty for the α^{CB} parameters, which are neglected as systematic uncertainty for the signal extraction. The blue line corresponds to the μ^{CB} relative uncertainty, and the green line to the σ^{CB} relative uncertainty.

uncertainty to this component, since any potential differences is be dominated by the large yield of the background distribution.

Figure 8.13 displays the fit uncertanties of the DSCB parameters as a function of the ALP mass relative to the parameter size. The uncertainties of the μ^{CB} parameter are mostly on the order of 0.01 % across the considered mass range, consistent with the expectation from Fig. 6.1. For the σ^{CB} parameter, the uncertainties are typically in the range of 2 to 3 %, with a few outliers. The relative statistical uncertainties of the α parameters are about 5 % for α_r^{CB} and, on average, 10 % for α_l^{CB}. However, due to the strong correlation with the σ^{CB} parameter, larger relative uncertainty variations in the α parameters are not of concern. I propagate these uncertainties as systematic uncertainties to the signal shape during the signal extraction.

8.9 Luminosity

This section points out the systematic uncertainty considered to account for the uncertainty in the integrated luminosity measurement, which is provided by the collaboration [87]. For this analysis, I combine the integrated luminosities of the $\Upsilon(4S)$ resonance and the off-resonance data taking periods as described in Section 3.4.3. The total integrated luminosity amounts to

$$\int \mathcal{L}\,\mathrm{d}t = (408.11 \pm 1.71)\,\mathrm{fb}^{-1}. \tag{8.6}$$

I propagate the uncertainty of $1.71\,\mathrm{fb}^{-1}$ ($0.42\,\%$ relative systematic uncertainty) as a systematic uncertainty to the signal extraction.

8.10 Summary

For this analysis, I consider the corrections based on the MC/data photon efficiency agreement and the photon energy bias, which are provided centrally by the collaboration. I apply both corrections to the MC simulation in case of the photon efficiency correction and to the recorded data in case of the energy bias correction.

Figure 8.14 summarises the total systematic uncertainty on the signal efficiency as a function of the ALP mass. I obtain the total uncertainty by adding the individual contributions discussed in the previous sections in quadrature. I neglect the uncertainty of the energy bias correction due to its low value. The systematic uncertainties for the trigger efficiency and the run-dependent interpolation are the subdominant contributions across the whole ALP mass range. The dominant contributions arise from the photon efficiency correction, the selection efficiency, and the ISR modelling. For the low ALP mass region, the uncertainty is driven by the ISR modelling, while for the high mass region, the uncertainty is dominated by the selection efficiency. In total, the systematic uncertainty on the signal efficiency ranges from about $8\,\%$ for low ALP masses to about $5\,\%$ for high ALP masses.

Figure 8.13 summarises the total systematic uncertainty on the peaking signal shape parameters as a function of the ALP mass. The uncertainty for the combinatorial component is not considered, as any almost constant offset will be overshadowed by the large background yield, which is determined directly from the recorded data.

The total relative uncertainty of the integrated luminosity of $0.42\,\%$ is propagated as a systematic uncertainty on the luminosity.

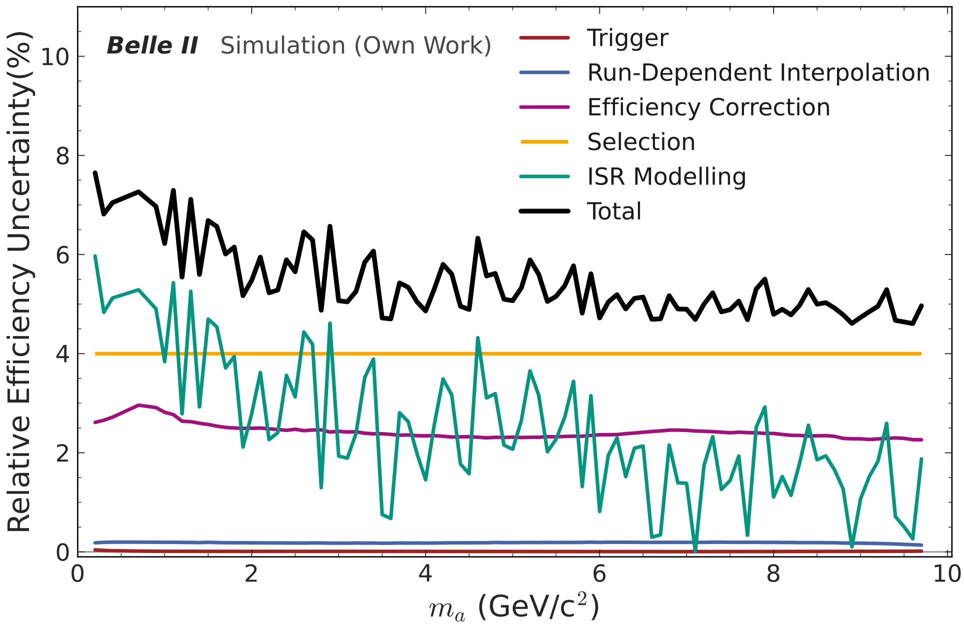


Figure 8.14: Summary of the systematic uncertainty estimation relative to the signal efficiency. The total systematic uncertainty (black) corresponds to the quadratic sum of the different considered systematic uncertainty components: Trigger efficiency estimation (red), run-dependent interpolation (blue), selection (yellow), and ISR modelling (green).

Chapter 9

Signal Extraction

In this chapter, I describe the signal extraction procedure for the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process. I perform the signal extraction by fitting the $M_{\gamma\gamma}$ spectrum at each tested ALP mass point to search for a significant excess over the SM background. To identify a potential excess, a combined PDF - composed of the signal and background parameterisations described in Chapter 6 - is fitted to the $M_{\gamma\gamma}$ distribution within a window of $\pm 20\sigma^{\mathrm{CB}}$ around the tested ALP mass, where σ^{CB} denotes the width of the peaking component of the signal PDF. From this fit, I determine the signal yield and the significance of any observed excess. In the absence of a significant excess, I set global significance greater than 3σ, an UL on the production cross section.

To ensure that no potential signal-peaking structure is missed, I perform the fits in steps of half the minimal signal width $5\,\mathrm{MeV}/c^2$. The total scanned range in the invariant diphoton mass $M_{\gamma\gamma}$ starts above the π^0 mass at $M_{\gamma\gamma} = 0.17\,\mathrm{GeV}/c^2$ and extends up to $M_{\gamma\gamma} = 9.8\,\mathrm{GeV}/c^2$, excluding the peaking regions of the π^0, η, and ω mesons, as I describe in Section 6.2. As previously mentioned, I perform each fit within a range of $\pm 20\sigma^{\mathrm{CB}}$ around the tested ALP mass. To ensure stable fits near the excluded regions, the fit ranges include the blinded $M_{\gamma\gamma}$ intervals; however, no signal significance is extracted within these regions. As I demonstrate in the control region studies in Chapter 7, the impact of peaking background processes within the fit range is negligible, as they are entirely dominated by events from the $e^+e^- \to \gamma\gamma(\gamma)$ process. I use these excluded regions to stabilise the background PDF.

First, Section 9.1 describes the combined signal and background PDF used for each scan. This section also details the determination of the fixed and free parameters in the fit, as well as the treatment of systematic uncertainties. Next, in Section 9.2 I explain the procedure for determining the significance of any observed excess. Section 9.3 outlines the method used to derive an UL on the cross section in the absence of a significant excess, and Section 9.4 places the resulting UL in the context of model-dependent exclusions. Finally, in Section 9.5 I describe the validation of the signal extraction procedure, including the treatment of the look-elsewhere effect, fit bias, fit injection, and fit coverage studies.

All events presented in this chapter pass the selection described in Chapter 4, excluding the merged photon classifier, which I discuss in Chapter 7. In addition, all events satisfy the L1 trigger and HLT selections described in Section 5.1 and are weighted according to either the MC/data photon efficiency correction for simulated events or the photon energy bias correction for recorded data. Both corrections are described in Chapter 8. I perform the fits, extract the significance and ULs, and validate the results using the `zfit` and HEPSTATS [106] Python packages. As in Chapter 6, all fits presented in this chapter are unbinned minimum-NLL fits.

9.1 Fit Model

I construct the fit model for the signal extraction by combining the signal and background parameterisations described in Chapter 6. To directly extract the cross section for the signal process $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ and to facilitate easier access to the signal efficiency for the treatment of systematic uncertainties, I add extended yield parameters to the PDFs. Therefore, the final PDF reads as

$$
\begin{aligned}
F(M_{\gamma\gamma}; \sigma, N_{\text{bkg}}, \boldsymbol{\theta}_b) &= N_{\text{sig}}(\sigma) \cdot S(M_{\gamma\gamma}) + N_{\text{bkg}} \cdot B(M_{\gamma\gamma}; \boldsymbol{\theta}_b), \quad (9.1)\\
&= \mathcal{L}\varepsilon\sigma \cdot (f \cdot P(M_{\gamma\gamma}) + (1-f) \cdot C(M_{\gamma\gamma})) + N_{\text{bkg}} \cdot B(M_{\gamma\gamma}; \boldsymbol{\theta}_b), \quad (9.2)
\end{aligned}
$$

where $N_{\text{sig}} = \mathcal{L}\varepsilon\sigma$ is the extended yield parameter for the signal, N_{bkg} is the extended yield parameter for the background, $S(M_{\gamma\gamma})$ is the signal PDF with peaking and combinatoric components, $B(M_{\gamma\gamma}; \boldsymbol{\theta}_b)$ is the background PDF with polynomial coefficients $\boldsymbol{\theta}_b$, $P(M_{\gamma\gamma})$, $C(M_{\gamma\gamma})$ are the peaking and combinatoric components of the signal PDF, and f is the ration between these components, respectively.

Signal Parameters

The integrated luminosity $\mathcal{L}$ corresponds to the total integrated luminosity collected during Run 1 for all datasets taken at the $\Upsilon(4S)$ resonance and $\Upsilon(4S)$ off-resonance energies. I determine the signal efficiency ε individually for each ALP mass scan point, based on the final fitted efficiency described in Section 5.2. I fix the efficiency and the integrated luminosity fit parameters to their pre-determined values in the fit and add a common nuisance parameter to account for their associated systematic uncertainties.

The cross section σ, corresponding to the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process, is a free parameter in the fit. I extract it separately for each ALP mass point, and it represents the only free parameter entering the signal yield N_{sig}, making it the sole free parameter in the signal PDF $S(M_{\gamma\gamma})$. I do not constrain the cross section σ to be positive in the fit, allowing for negative fluctuations. Although negative signal yields are unphysical, this approach enables a more accurate estimation of the fit significance, as discussed in Section 9.2. Negative cross sections are ultimately interpreted as the absence of a signal.

I fix all other parameters of the signal PDF in Eq. (6.1), namely f, $\boldsymbol{\theta}_p$, and $\boldsymbol{\theta}_c$, to the values determined from the MC samples. For the peaking component, I treat the mean μ^{CB}, width σ^{CB}, α_l^{CB}, and α_r^{CB} as nuisance parameters to account for systematic uncertainties in the signal shape. The central values of the signal PDF correspond to those shown in Fig. 6.1, derived under the beam background conditions of experiment 22, run 468, which has the $N_{\mathrm{OOT}}^{\mathrm{Crys.}}$ value closest to the average of the Run 1 dataset.

Background Parameters

The background PDF consists of the extended background yield N_{bkg} and the polynomial function $B(M_{\gamma\gamma};\boldsymbol{\theta}_b)$, as defined in Eq. (6.4). The constraints that I apply to the free fit parameters are that the total background yield N_{bkg} must be positive, and the polynomial coefficients $\boldsymbol{\theta}_b$ are restricted to the range -2 to 2 based on the values estimated in Figure 6.4.

Fit Procedure

For studies based on simulated events, I perform the signal extraction fits in two stages. First, I fit the background-only hypothesis to the $M_{\gamma\gamma}$ distribution of all background MC events. This initial fit determines the starting values for the background parameters, N_{bkg} and $\boldsymbol{\theta}_b$. Next, I sample the background distribution from the fitted background PDF to generate pseudo-samples. For each fit, I sample the background yield N_{bkg} from a Poisson distribution with a mean equal to the fitted background yield from the initial fit. I then fit the sampled events with the full signal-plus-background PDF, including the free background PDF parameters, as described previously. This strategy ensures that the statistical uncertainty of the weighted background MC sample does not bias the fit results. For the final fits to recorded data, I perform only the full hypothesis fit, using the actual recorded data as input.

Figure 9.1 shows four example fits of the final fit model to the $M_{\gamma\gamma}$ distribution for different ALP mass hypotheses. These examples illustrate the typical background shapes observed across the entire $M_{\gamma\gamma}$ spectrum. Fits at very low ALP masses exhibit a rising $M_{\gamma\gamma}$ spectrum. The fit for a ALP mass of $m_a \approx 0.65\,\mathrm{GeV}/c^2$ represents a typical case between the excluded regions of the peaking background processes. Here, the background shape is stabilised by including the excluded regions in the fit, due to the strong dominance of the $e^+e^- \to \gamma\gamma(\gamma)$ process. For mid-range ALP masses, the background $M_{\gamma\gamma}$ distribution is almost entirely dominated by the $e^+e^- \to \gamma\gamma(\gamma)$ process and appears nearly flat, as reflected by the small values of the orthogonal polynomial coefficients. Finally, the fit for a ALP mass of $m_a \approx 9.50\,\mathrm{GeV}/c^2$ shows a decreasing $M_{\gamma\gamma}$ spectrum towards the kinematic allowed region.

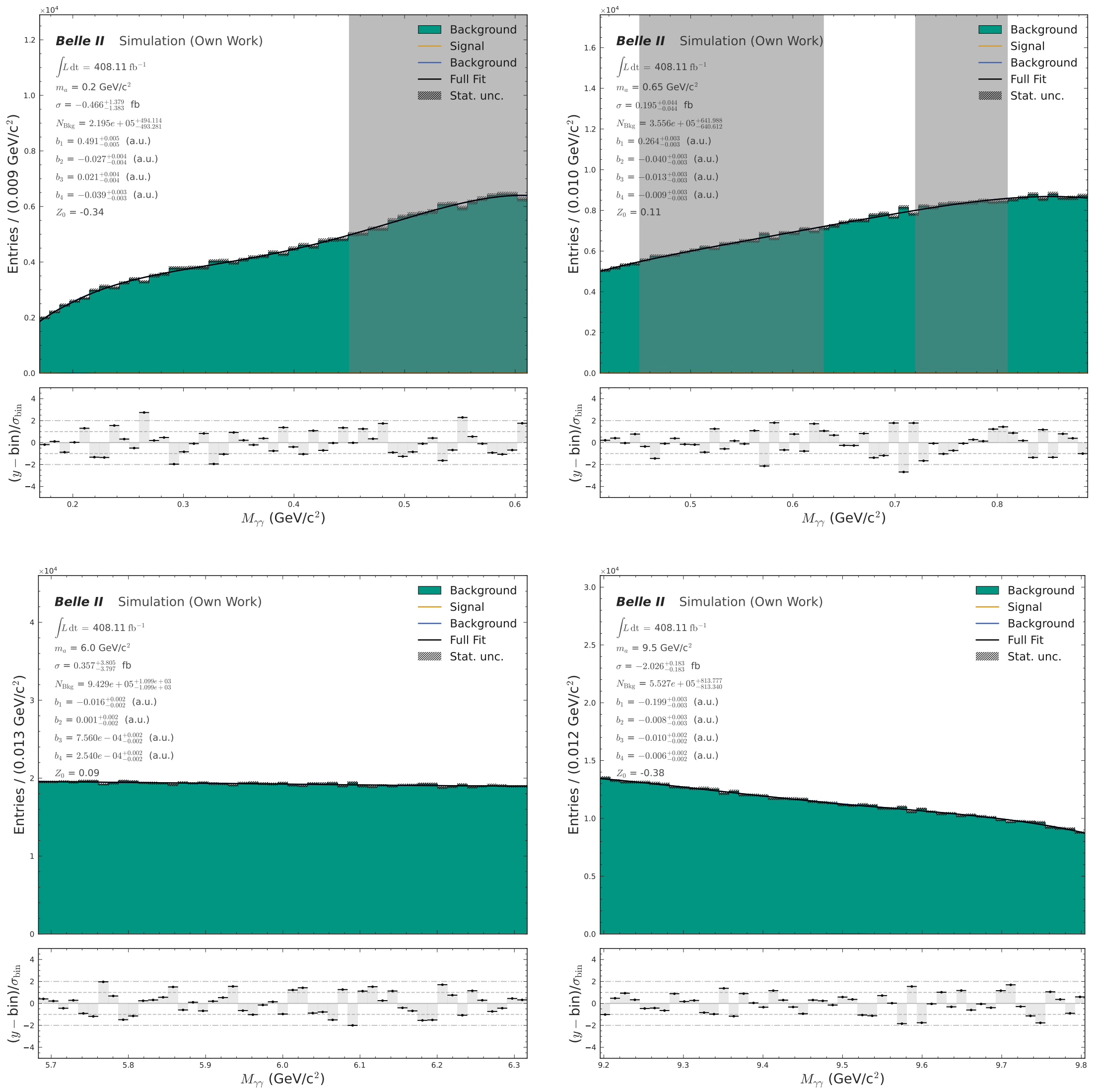


Figure 9.1: Distributions of $M_{\gamma\gamma}$ for sampled background events, fitted with the final combined signal and background PDF for four example ALP mass hypotheses. The filled histogram shows the background events sampled from the background-only fit to simulated events. Solid lines indicate the fitted PDF components: the total background PDF (blue), the combined signal peaking combinatorial component (orange), and the total PDF (black). Grey shaded regions indicate the excluded areas around the η and ω masses.

Treatment of Systematic Uncertainties

I treat systematic uncertainties by applying constraints on the nuisance parameters of the fit model. To account for these, I extend the likelihood of the fit model by Gaussian constraints on the nuisance parameters. The total likelihood $\mathscr{L}$ then becomes

$$\mathscr{L}(M_{\gamma\gamma}; \sigma, N_{\mathrm{bkg}}, \boldsymbol{\theta}_b) = \prod_{i=1}^{n} F(M_{\gamma\gamma}^{i}; \sigma, N_{\mathrm{bkg}}, \boldsymbol{\theta}_b) \cdot \prod_{j=1}^{m} \mathcal{G}(\theta_j; \mu_j, \sigma_j), \tag{9.3}$$

where $F(M^i_{\gamma\gamma};\sigma, N_{\text{bkg}}, \boldsymbol{\theta}_b)$ is the fit model described above, $\mathcal{G}(\theta_j;\mu_j,\sigma_j)$ is the Gaussian constraint on the nuisance parameter θ_j with mean μ_j and standard deviation σ_j, and n and m denote the number of data points and nuisance parameters, respectively. The nuisance parameters θ_j are not of direct interest in the fit but are necessary to account for variations arising from prior knowledge.

For the shape parameters of the signal PDF, I apply Gaussian constraints directly to the mean μ^{CB}, σ^{CB}, α_l^{CB}, and α_r^{CB} of the peaking component. The mean values μ_j correspond to the nominal values determined from the simulation studies in Section 6.1, and the standard deviations σ_j of the Gaussian constraints are summarised in Section 8.8. For the integrated luminosity $\mathcal{L}$ and signal efficiency ε, which enter the signal yield, I parametrise the yield factors using a combined nuisance parameter:

$$N_{\text{sig}}(\sigma,\delta) = \mathcal{L}\cdot\varepsilon\cdot\sigma\cdot(1+\delta), \tag{9.4}$$

where δ is a nuisance parameter with a Gaussian constraint centred at zero and a standard deviation corresponding to the total relative uncertainty of the product $\mathcal{L}\cdot\varepsilon$. This parametrisation stabilises the numerical treatment of the uncertainty in the fit. I calculate the total uncertainty σ_δ as the quadratic sum of the total relative efficiency uncertainty, shown in Fig. 8.14, and the relative uncertainty of the integrated luminosity (see Section 8.9).

9.2 Significance

To estimate the significance of a fitted signal yield, I use the profile likelihood ratio $\lambda(\sigma)$ test statistic in the asymptotic limit [107]. The test statistic q_σ is defined as

$$q(\sigma) = -2\ln\lambda(\sigma) = -2\ln\left(\frac{\mathscr{L}(\sigma,\hat{\hat{N}}_{\text{bkg}},\hat{\hat{\boldsymbol{\theta}}}_b)}{\mathscr{L}(\hat{\sigma},\hat{N}_{\text{bkg}},\hat{\boldsymbol{\theta}}_b)}\right) \tag{9.5}$$

where N_{bkg} and $\boldsymbol{\theta}_b$ correspond to the background yield and polynomial coefficients as described above, respectively. One hat denotes their best fitted values, and two hats denote their best fitted values under the null hypothesis. For the null hypothesis, I define that the signal yield N_{sig} and therefore cross section σ is zero, while the alternative hypothesis is that the signal yield corresponds to the measured cross section $\hat{\sigma}$. Classically, the p-value of the test statistic q_0 is determined by comparing it to the distribution of the test statistic:

$$p(q_0) = \int_{q_0}^{\infty} f(q|\sigma=0,\hat{\hat{N}}_{\text{bkg}},\hat{\hat{\boldsymbol{\theta}}}_b)\mathrm{d}q, \tag{9.6}$$

where $f(q|\sigma=0,\hat{\hat{N}}_{\text{bkg}},\hat{\hat{\boldsymbol{\theta}}}_b)$ is the distribution of the test statistic under the null hypothesis. Under the assumption that the number of events per fit is large enough, the distribution of the test statistic q_0 is asymptotically χ^2 distributed with one degree of freedom. Considering a negative signal yield, hence performing a two-sided test, the p-value can then be

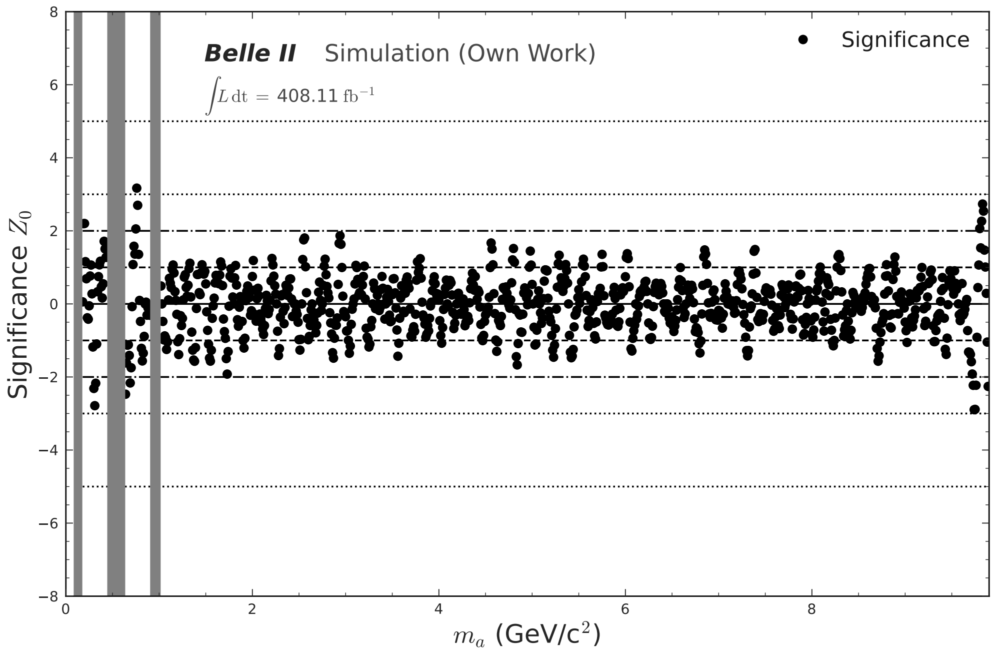


Figure 9.2: Significance of the excess for each scanned ALP mass point in simulation.

approximated as

$$p(q_0) = 2\left(1 - \Phi\left(\sqrt{q_0}\right)\right), \tag{9.7}$$

where Φ is the cumulative distribution function of the standard normal distribution. The factor of 2 accounts for the two-sided coverage probability. The significance Z_0 of the excess can then be determined as

$$Z_0 = \Phi^{-1}\left(1 - \frac{p(q_0)}{2}\right) = \sqrt{q_0} = \sqrt{\mathscr{L}'(\sigma = \hat{\sigma}, \hat{N}_{\mathrm{bkg}}, \hat{\boldsymbol{\theta}}_b) - \mathscr{L}'(\sigma = 0, \hat{\hat{N}}_{\mathrm{bkg}}, \hat{\hat{\boldsymbol{\theta}}}_b)}, \tag{9.8}$$

where $\mathscr{L}'(\sigma = \hat{\sigma}, \hat{N}_{\mathrm{bkg}}, \hat{\boldsymbol{\theta}}_b)$ is the minimum NLL of the fit and $\mathscr{L}'(\sigma = 0, \hat{\hat{N}}_{\mathrm{bkg}}, \hat{\hat{\boldsymbol{\theta}}}_b)$ is the minimum NLL of the fit under the null hypothesis, which is created from a so-called Asimov dataset. For the Asimov dataset, I evaluate the PDF defined in Eq. (9.1) in N_{A} bins and fit with the signal cross section σ fixed to zero. To account for negative signal yields, I multiply the resulting significance by the sign of the fitted signal yield.

Figure 9.2 shows the significance Z_0 of the excess for each scanned ALP mass point in simulation. The distribution of the significance is consistent with expectations, where most of the significances are consistent with zero, as no signal is present in the simulation. Fluctuations in the background lead to some local significances between $Z_0 = 0$ and $Z_0 = 1$, as well as few with $Z_0 > 2$. I calculate the here shown significance within $N_{\mathrm{A}} = 400$ Asimov bins across the $\pm 20\sigma^{\mathrm{CB}}$ fit window around the ALP mass.

9.3 Upper Limit

If no significant excess is found, I determine an UL on the cross section. The UL calculation follows the same thoughts as the significance calculation in Section 9.2. Instead of testing the null hypothesis of a zero signal yield, the UL calculation tests the signal hypothesis of $\hat{\sigma}$ against a range of alternative hypotheses of $\hat{\hat{\sigma}} \geq \hat{\sigma}$. The classical test statistic is then defined as

$$q_{\hat{\hat{\sigma}}} = \begin{cases} -2\ln\left(\frac{\mathscr{L}(\sigma=\hat{\hat{\sigma}},\hat{\hat{N}}_{\mathrm{bkg}},\hat{\hat{\boldsymbol{\theta}}}_b)}{\mathscr{L}(\sigma=\hat{\sigma},\hat{N}_{\mathrm{bkg}},\hat{\boldsymbol{\theta}}_b)}\right) & ;\ \hat{\hat{\sigma}} \geq \hat{\sigma}, \\ 0 & ;\ \text{otherwise.} \end{cases} \tag{9.9}$$

The corresponding p-value can then be determined as

$$p(q_{\hat{\hat{\sigma}}}) = \int_{q_{\hat{\hat{\sigma}}}}^{\infty} f(q|\sigma = \hat{\hat{\sigma}}, \hat{\hat{N}}_{\mathrm{bkg}}, \hat{\hat{\boldsymbol{\theta}}}_b)\mathrm{d}q. \tag{9.10}$$

For this analysis, I determine the UL at a 95 % Confidence Level (CL), which corresponds to a p-value of 0.05. To estimate which cross section $\hat{\hat{\sigma}}$ corresponds to this p-value, I perform a scan of different σ values. For each scanned cross section $\hat{\hat{\sigma}}$, I determine the test statistic $q_{\hat{\hat{\sigma}}}$ in the asymptotic limit and calculate the

$$\mathrm{CL_s} = \frac{p(q_{\hat{\hat{\sigma}}})}{1 - p_0} \tag{9.11}$$

value. Finally, I determine σ^{UL} as the largest cross section $\hat{\hat{\sigma}}$ for which $\mathrm{CL_s} \leq 0.05$ by interpolating the scanned $p(q_{\hat{\hat{\sigma}}})$ values.

For the UL calculation, I set the scan $\hat{\hat{\sigma}}$ between 0 fb and $\max(\hat{\sigma}, 50\,\mathrm{fb})$ in 200 steps and $N_{\mathrm{A}} = 200$ Asimov bins. Figure 9.3 shows the calculated UL on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section for each scanned ALP mass point on the entire Run 1 simulation. Due to the lower background, the UL is stronger in the lower ALP mass regions. For ALP masses between $m_a \geq 2.0\,\mathrm{GeV}/c^2$ and $m_a \leq 9.0\,\mathrm{GeV}/c^2$, the UL is constant, similar to the distributions of the number of background events and the signal efficiency. Finally, for high ALP mass hypotheses, the UL decreases due to the decreasing signal efficiency.

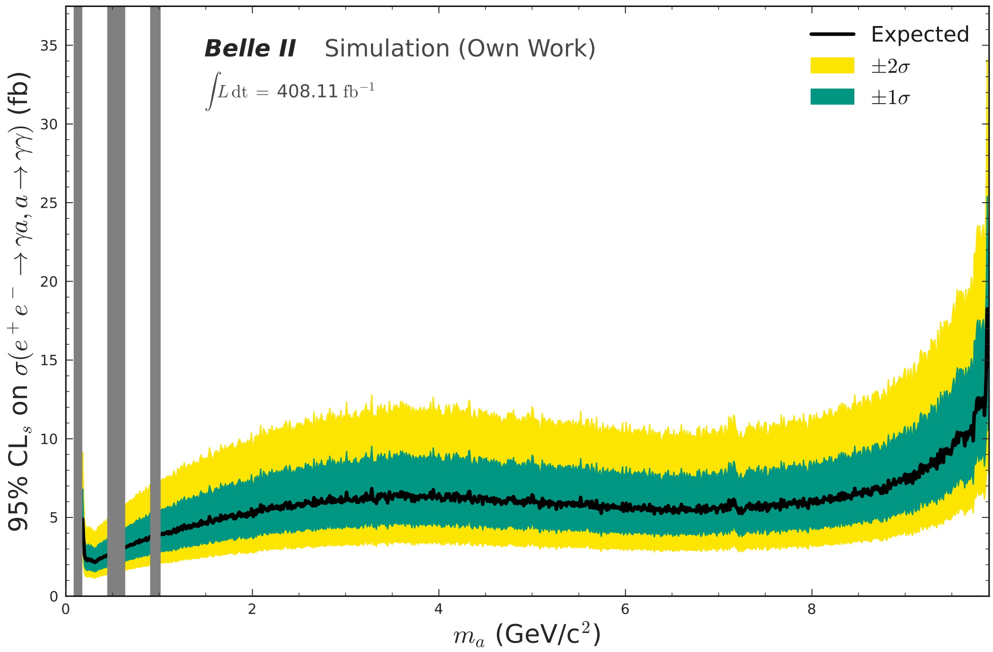


Figure 9.3: 95 %CL$_\mathrm{s}$ upper limit on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section for each scanned ALP mass point in simulation.

9.4 Model Dependent Sensitivity

The upper limit on the cross section σ for $e^+e^- \to \gamma a$, $a \to \gamma\gamma$, which I calculate in the previous section Section 9.3, can be translated into an UL on the ALP-photon coupling $g_{a\gamma\gamma}$. The re-calculation to a coupling limit presented here follows the calculations in [25], making the coupling limit model dependent. To determine the full cross section, I integrated Eq. (2.11) over the full solid angle:

$$\sigma = \frac{g_{a\gamma\gamma}^2 \alpha}{24} \left(1 - \frac{m_a^2}{s}\right)^3 . \tag{9.12}$$

Taking the results from the previous section, more specifically the 95 %CL$_\mathrm{s}$ UL on the cross section σ^{UL} from Fig. 9.3, I determine the UL on the ALP-photon coupling $g_{a\gamma\gamma}$ from Eq. (9.12). Figure 9.4 shows the UL on the ALP-photon coupling $g_{a\gamma\gamma}$ for each scanned ALP mass point. The here shown sensitivity estimates an improvement of the current best limit by BESIII in the region of low ALP masses by a factor of four. For ALP mass $m_a \in [1.0, 5.0]\,\mathrm{GeV}/c^2$, the limit improves the current best limit by a factor of 9.

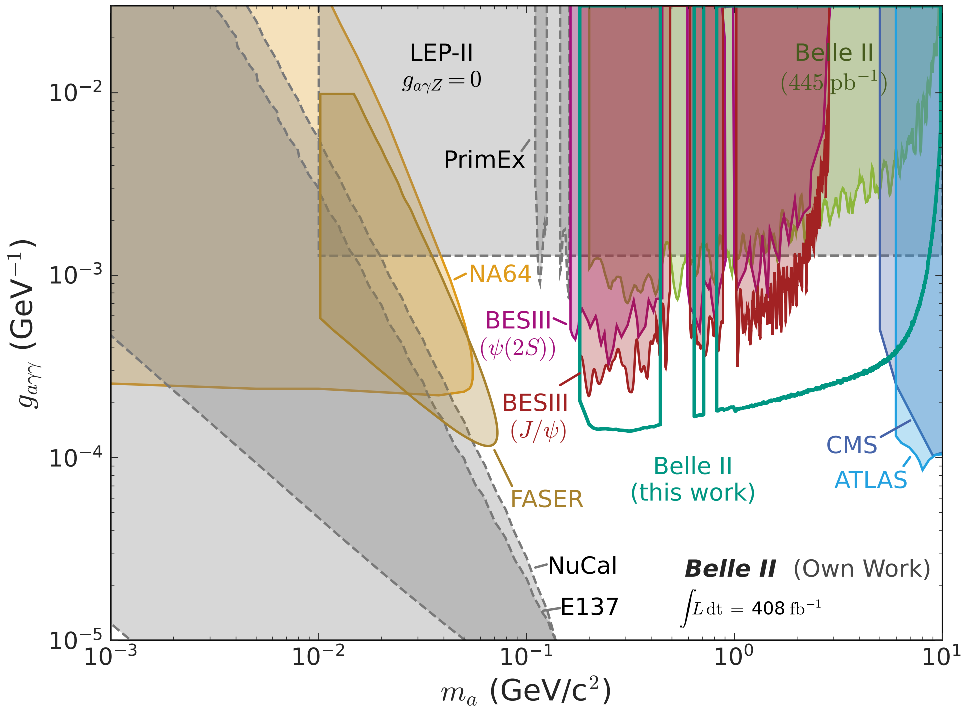


Figure 9.4: 95 %$\mathrm{CL_s}$ UL on the ALP-photon coupling $g_{a\gamma\gamma}$ for each scanned ALP mass point. The sensitivity of this analysis is shown as a solid red line. All other $g_{a\gamma\gamma}$ limits are taken from Figure 2.4.

9.5 Validation

For the validation of the extracted ULs, I evaluate and apply the following steps. I evaluate the treatment of the look-elsewhere effect. In addition, I assess the fit bias, the fit linearity and coverage. Finally, I perform a short study of the impact of different systematic uncertainties on the UL extraction in.

The setup for the validation studies follows the general fit procedure described in Section 9.1 for fits on simulated events. First, I fit the background-only hypothesis to the $M_{\gamma\gamma}$ distribution of all background MC events to determine the background PDF. Then, I sample the background distribution from the fitted background PDF. I sample the background yield N_{bkg} from a Poisson distribution with a mean equal to the fitted background yield from the initial fit. For the treatment of the look-elsewhere effect, the fit bias study, and the systematic uncertainty study, I do not inject signal. In cases where a signal injection is required, I sample the signal yield N_{sig} also from a Poisson distribution with a mean equal to the injected signal yield N_{inj}. In principle, the injected signal yield can be chosen arbitrarily. However, to ensure that the injected signal yield is statistically significant, I approximate the injected yield N_{inj} for a given target significance Z_0^{inj} as follows: For a given ALP mass m_a, the significance can be approximated as

$$Z_0 \approx \frac{N_{\text{sig}}(m_a)}{\sqrt{N_{\text{sig}}(m_a) + N_{\text{bkg}}^{\pm w\sigma}(m_a)}}, \tag{9.13}$$

where $N_{\text{bkg}}^{\pm w\sigma}(m_a)$ is the number of background events within $\pm w\sigma^{\text{CB}}$ around the ALP mass m_a For these tests, I choose $w = 3$. Solving Equation (9.13) for N_{sig} leads to

$$N_{\text{sig}}(m_a) = N_{\text{inj}}(m_a) = \frac{1}{2}\left({Z_0^{\text{inj}}}^2 + \sqrt{{Z_0^{\text{inj}}}^4 + 4N_{\text{bkg}}^{\pm 3\sigma}(m_a){Z_0^{\text{inj}}}^2} \right). \tag{9.14}$$

Since the determination of N_{inj} is only an approximation, the actual achieved significance Z_0^{inj} may differ from the fitted significance and is not evaluated in the context of the fit validation.

9.5.1 Look-Elsewhere Effect

In searches involving scans over multiple parameters, such as the ALP mass in this analysis, the probability of observing a significant excess due to random fluctuations increases with the number of independent tests performed. This correlation is known as the look-elsewhere effect and must be taken into account when interpreting the significance of any observed excess. A common approach to address the look-elsewhere effect is to adjust the local p-value of an observed excess to account for the number of independent tests, yielding a global p-value. Following the prescription in [108], I determine a so-called trial factor, which quantifies the ratio between the probability of observing an excess at any given scan point and the probability of observing it anywhere within the entire scan range.

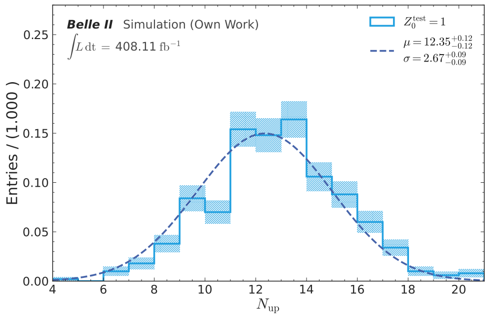
Treating each ALP mass hypothesis as an independent test, I estimate the trial factor from the number of upcrossings $N_{\text{up}}(Z_0^{\text{test}})$ of the extracted local significance Z_0^{local} above a chosen threshold Z_0^{test}. In the asymptotic limit for a single scan parameter, the relationship between the local and global p-values is given by

$$p_{\text{global}} = p_{\text{local}} + \hat{N}_{\text{up}}(Z_0^{\text{test}})e^{-({Z_0^{\text{local}}}^2 - {Z_0^{\text{test}}}^2)/2}, \tag{9.15}$$

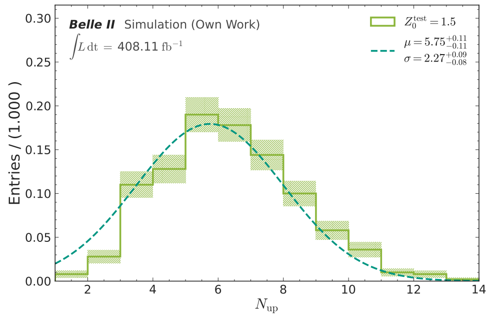
where $\hat{N}_{\text{up}}$ is the mean number of upcrossings at the threshold Z_0^{test}.

To determine $\hat{N}_{\text{up}}(Z_0^{\text{test}})$, I perform a series of signal hypothesis scans on background-only $M_{\gamma\gamma}$ distributions sampled from the fitted background PDF. For each scan point, I extract the local significance Z_0^{local} and count the number of upcrossings per scan. I repeat this procedure 500 times, counting the number of upcrossings above the thresholds $Z_0^{\text{test}} \in \{1, 1.5, 2\}$. Figure 9.5 shows the resulting distributions of the number of upcrossings for the different thresholds across all scans. I obtain the mean number of upcrossings $\hat{N}_{\text{up}}(Z_0^{\text{test}})$ by fitting these distributions with Gaussian PDFs.

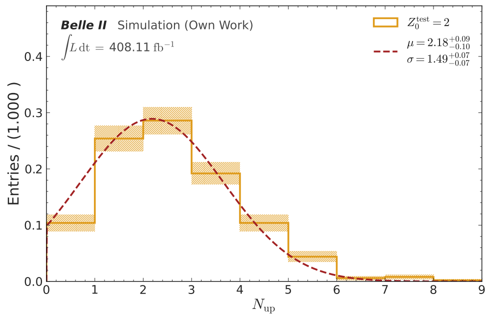
Following, I use the mean number of upcrossings to convert the local p-value of any



(a) $Z_0^{\mathrm{test}} = 1$



(b) $Z_0^{\mathrm{test}} = 1.5$



(c) $Z_0^{\mathrm{test}} = 2$

Figure 9.5: Distribution of the number of upcrossings for different thresholds Z_0^{test} across 500 background-only scans. Dashed lines indicate the fitted Gaussian PDFs.

observed excess into a global p-value. Figure 9.6 shows the resulting global significance Z_0^{global} as a function of the threshold Z_0^{test} for different observed local significances Z_0^{local}. Since I do not observe significant variation in the global significance for different threshold choices, I adopt $Z_0^{\mathrm{test}} = 1$ for the final treatment of the look-elsewhere effect. For a potential observed local significance of $Z_0^{\mathrm{local}} = 5$, the corresponding global significance is reduced to $Z_0^{\mathrm{global}} = 3.78$ and $Z_0^{\mathrm{local}} = 3$ to $Z_0^{\mathrm{global}} = 0.75$.

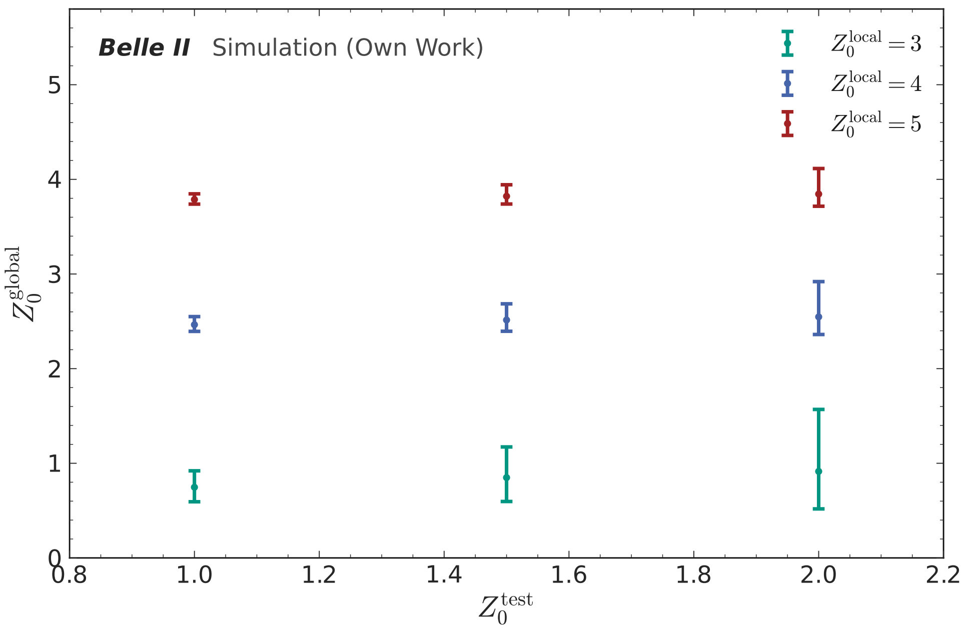


Figure 9.6: Global significance Z_0^{global} as a function of the threshold Z_0^{test} for different observed local significances Z_0^{local}.

9.5.2 Fit Bias

To test for a possible fit bias towards positive or negative significances, I examine the distribution of the observed local significances Z_0^{local} across all mass points. For this study, I use the same set of significances that I use to determine the look-elsewhere effect in the previous section.

In the ideal case, the distribution of Z_0^{local} should follow a Gaussian PDF with a mean of zero and a standard deviation of one. Figure 9.7 shows the observed distribution of significances together with the result of a Gaussian fit. The mean being consistent with zero indicates no bias towards positive or negative significances, while the standard deviation being consistent with one confirms that the estimated uncertainties cover the expected range.

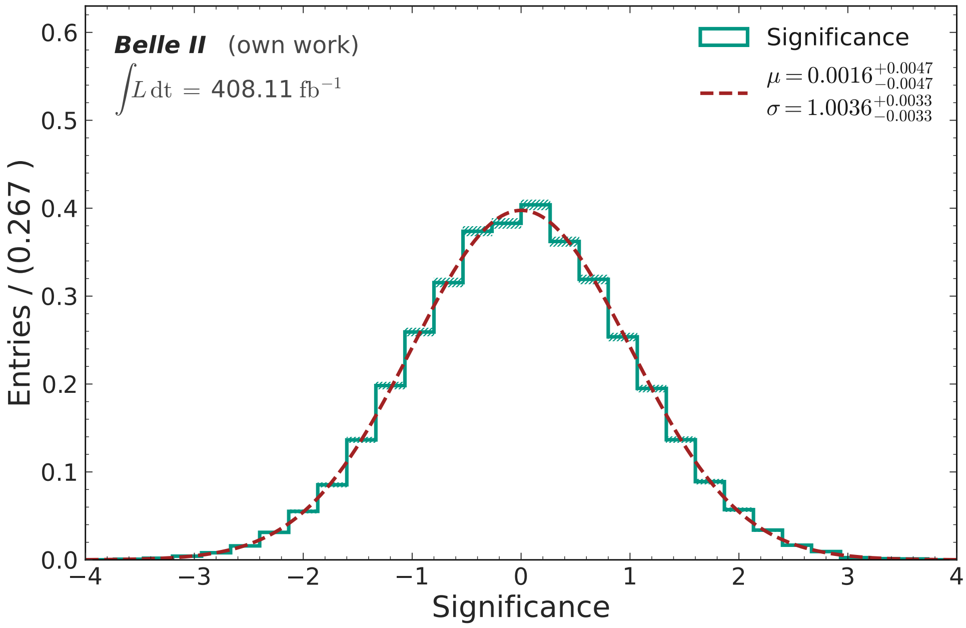


Figure 9.7: Significance distribution (green histogram) across all mass points. The red line shows a Gaussian fit to the distribution, which has a mean of 0.0016 ± 0.0047 and a standard deviation of 1.0036 ± 0.0033.

9.5.3 Fit Linearity and Coverage

I evaluate the linearity and coverage of the fit model using a series of signal injection tests. For each ALP mass hypothesis and injection significance Z_0^{inj}, I inject a signal yield N_{inj} corresponding to the approximation in Eq. (9.14) into the sampled background events. I repeat each injection 500 times per ALP mass and Z_0^{inj} value, with the background and signal yields sampled from Poisson distributions as described previously. The injection tests are restricted to ALP masses within the scan range, using a step size of $0.1\,\text{GeV}/c^2$ and injection significances of $Z_0^{\text{inj}} \in \{0, 1, 2, 3, 4, 5\}$. Figure 9.8 displays four example $M_{\gamma\gamma}$ distributions with injected signal events at different ALP masses and injection significances.

I evaluate the fit linearity and coverage by studying the pull distribution of the fitted signal yield:

$$P_{\text{inj}} = \frac{\hat{N}_{\text{inj}} - \langle N_{\text{inj}} \rangle}{\sigma_{\hat{N}_{\text{inj}}}}, \tag{9.16}$$

where $\hat{N}_{\text{inj}}$ is the fitted signal yield, $\langle N_{\text{inj}} \rangle$ is the mean of the injected signal yield, and $\sigma_{\hat{N}_{\text{inj}}}$ is the uncertainty of the fitted signal yield. In the ideal case, the pull distribution P_{inj} follows a standard normal distribution, *i.e.* it has a mean of zero and a standard deviation of one. Figure 9.9 shows an example pull distribution for the ALP mass hypothesis $m_a = 2.1\,\text{GeV}/c^2$ and an injection significance of $Z_0^{\text{inj}} = 3$.

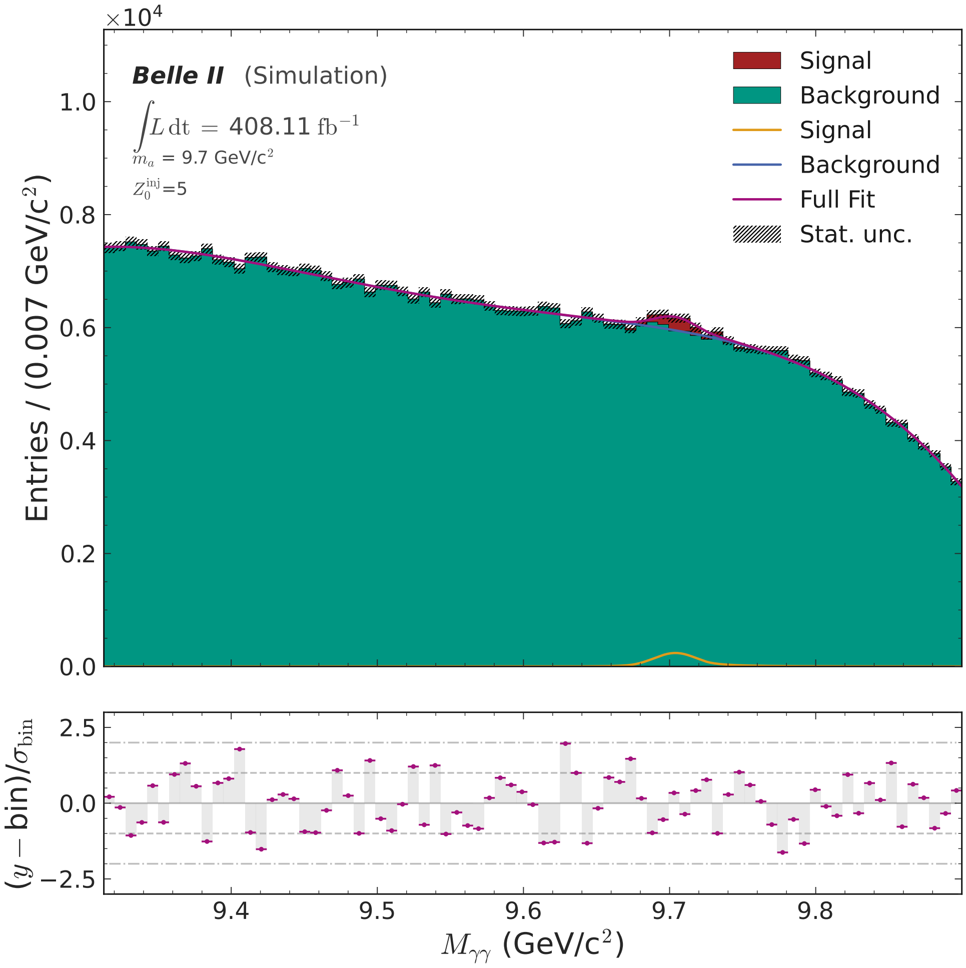


Figure 9.8: Distribution of the $M_{\gamma\gamma}$ observable for sampled background events and $Z_0^{\mathrm{inj}} = 5$ injected signal events, fitted with the final combined signal and background PDF for the ALP mass hypothesis $m_a = 9.7\,\mathrm{GeV}/c^2$. The filled histogram shows the background events sampled from the background-only fit to simulated events (green) and the injected signal (red). Solid lines indicate the fitted PDF components: the total background PDF (blue), the combined signal peaking combinatorial component (orange), and the total PDF (black).

Figures 9.10 and 9.11 show the mean $\mu^{P_{\mathrm{inj}}}$ and standard deviation $\sigma^{P_{\mathrm{inj}}}$ of the pull distributions for each injected ALP mass and each injection significance.

As evident from the pull distributions, the results are nearly constant within statistical uncertainties across the entire ALP mass range. To estimate the overall mean and standard deviation of the pull distributions, I perform constant fits to the pull means and standard deviations obtained from all injection studies. The largest observed deviation of the pull mean is $\frac{\hat{\mu}^{P_{\mathrm{inj}}}}{\mu^{P_{\mathrm{inj}}}} = (1.6 \pm 0.5)\,\%$ for an injection significance of $Z_0^{\mathrm{inj}} = 1$, while the greatest deviation from the ideal standard deviation of one is $\frac{\hat{\sigma}^{P_{\mathrm{inj}}}}{\sigma^{P_{\mathrm{inj}}}} = (2.7 \pm 0.3)\,\%$ for $Z_0^{\mathrm{inj}} = 4$. Given the overall small deviations from the expected values for $\mu^{P_{\mathrm{inj}}}$, I conclude that the

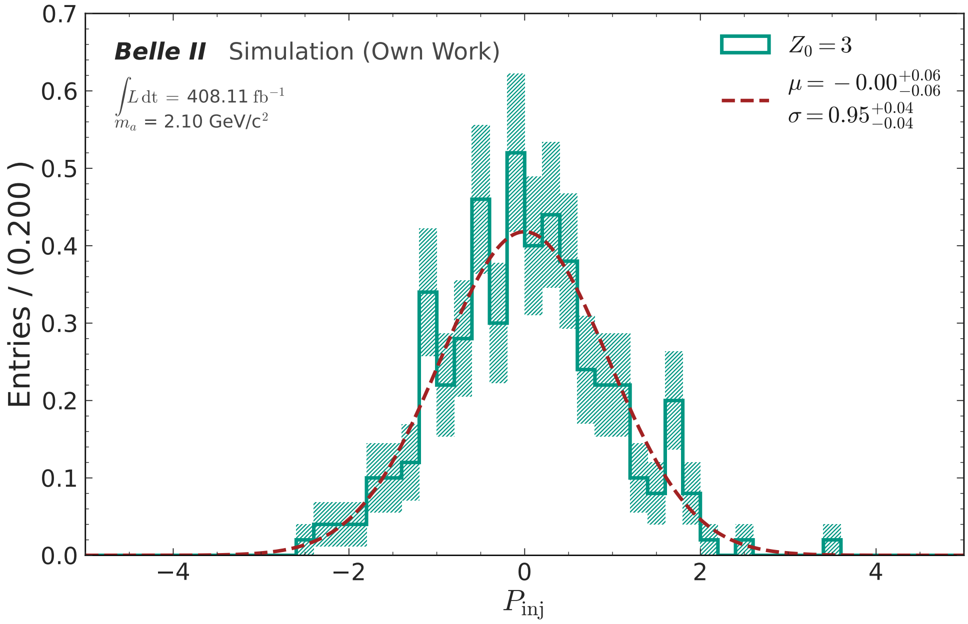


Figure 9.9: Distribution of the pull P_{inj} distribution for $Z_0^{\mathrm{inj}} = 3$ injected signal events, fitted with the final combined signal and background PDF for the ALP mass hypothesis $m_a = 2.1\,\mathrm{GeV}/c^2$. The red line shows a Gaussian fit to the distribution, which has a mean of 0.00 ± 0.06 and a standard deviation of 0.95 ± 0.04.

fit model demonstrates the expected linear behaviour. The slight underestimation of the standard deviation $\sigma^{P_{\mathrm{inj}}}$ indicates a minor overestimation of the uncertainties on the fitted signal yield. Nevertheless, the observed deviations are small enough to ensure sufficient coverage of the uncertainties.

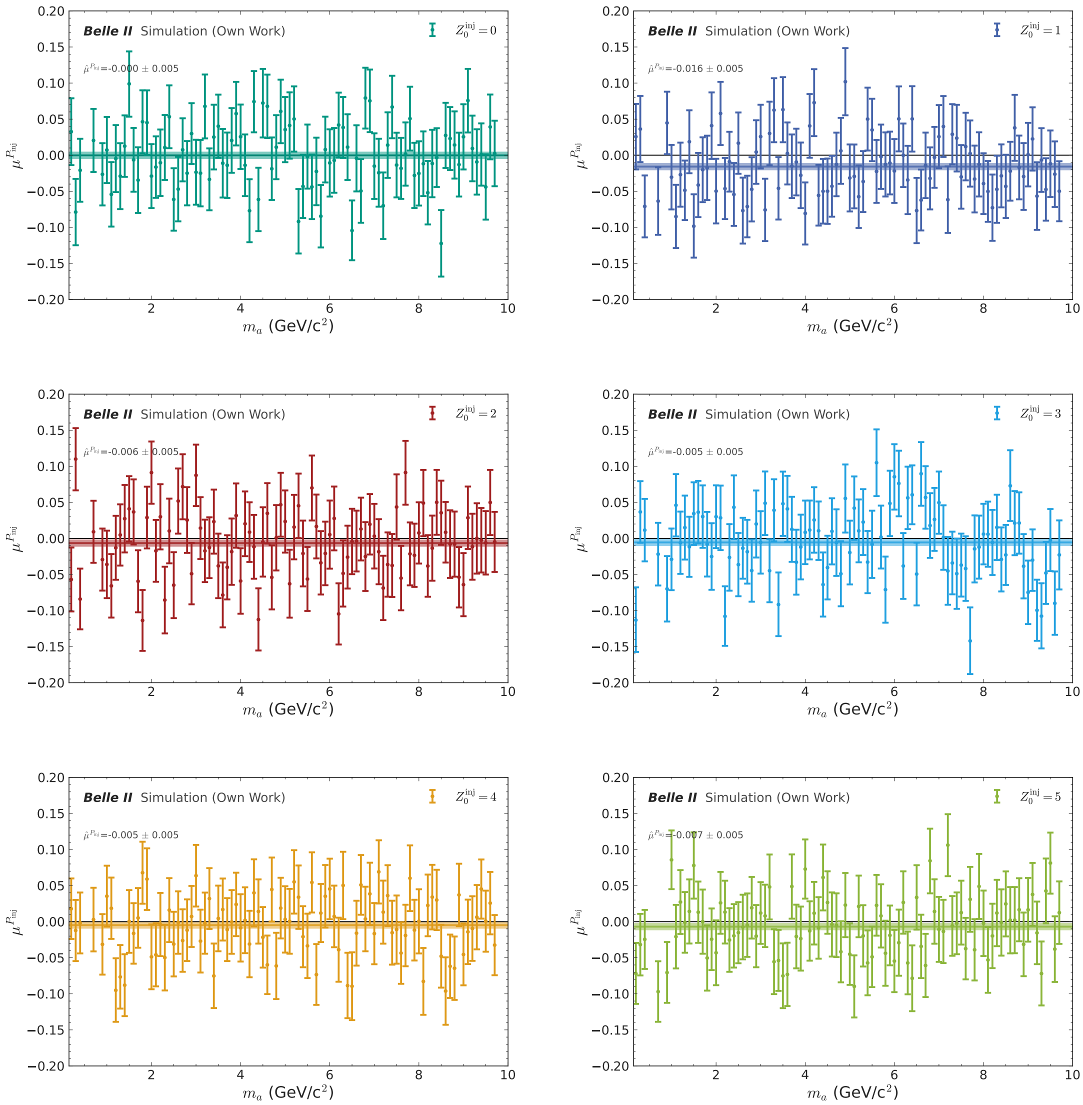


Figure 9.10: Distribution of the $\mu^{P_{\mathrm{inj}}}$ as a function of ALP mass hypothesis for different injection significances $Z_0^{\mathrm{inj}} \in \{0, 1, 2, 3, 4, 5\}$. Error bars indicate the central fit values and statistical uncertainties of the P_{inj} distributions. Each point corresponds to one injection study with 500 repetitions. Solid lines indicate constant fits to the distributions. Colored areas represent the statistical uncertainty of the constant fit.

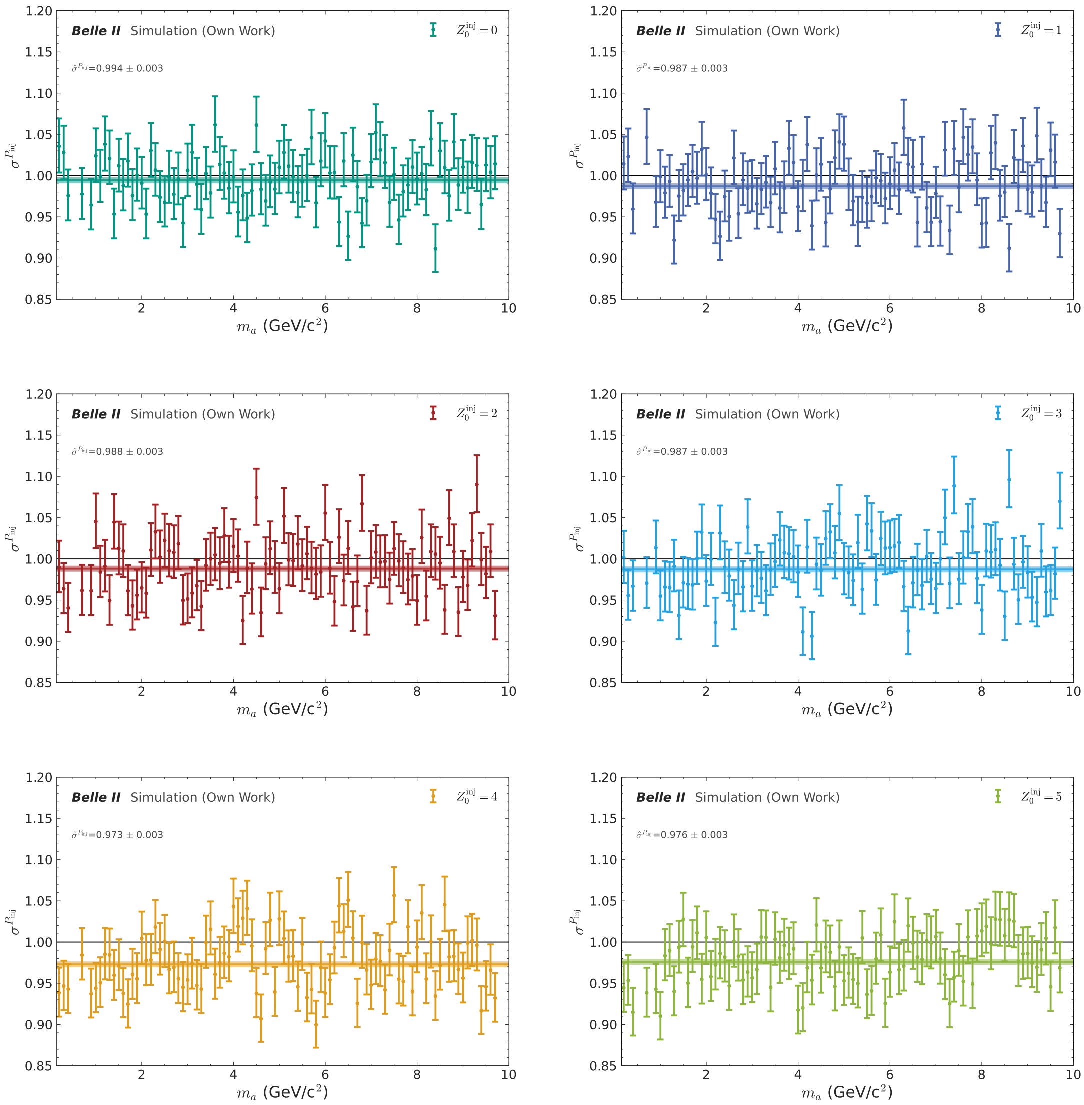


Figure 9.11: Distribution of the $\sigma^{P_{\text{inj}}}$ as function of ALP mass hypothesis for different injection significances $Z_0^{\text{inj}} \in \{0, 1, 2, 3, 4, 5\}$. Error bars indicate the central fit values and statistical uncertainties of the P_{inj} distributions. Each point corresponds to one injection study with 500 repetitions. Solid lines indicate constant fits to the distributions. Colored areas represent the statistical uncertainty of the constant fit.

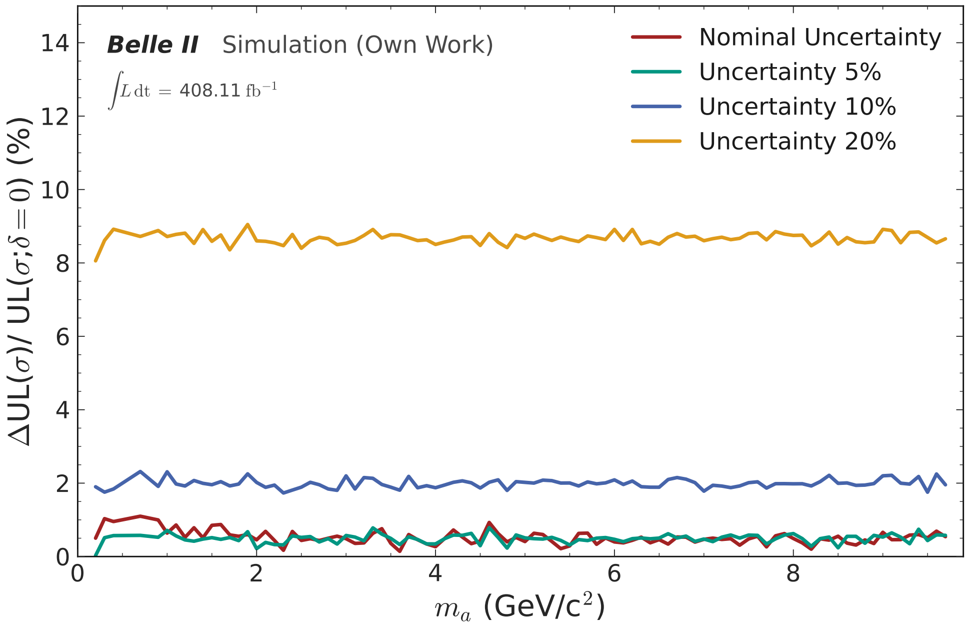


Figure 9.12: 95 %CL$_s$ upper limit comparison on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section for each scanned ALP mass point. The upper limits are normalised to the case without systematic uncertainties for different assumed total uncertainties on the signal yield. Compared levels are 5 % (green), 10 % (blue), and 20 % (orange), as well as the nominal uncertainty (red) on the signal yield.

9.5.4 Impact of Systematic Uncertainties

This section briefly discusses the impact of different levels of systematic uncertainties on the UL extraction. For the study of the impact of the systematic uncertainties on the UL, I follow the general fit procedure described in Section 9.1, introducing variations to account for different assumed values of systematic uncertainties.

Since the extracted signal yield is the dominant factor in the UL calculation, I focus on the effect of the systematic uncertainty represented by the nuisance parameter δ in the signal yield (see Eq. (9.4)), while systematic uncertainties on the signal shape are neglected. Figure 9.12 shows the expected 95 % CL upper limits on the cross section and the model-dependent $g_{a\gamma\gamma}$ coupling, normalised to the case without systematic uncertainties, as a function of the ALP mass for different total uncertainty levels on the signal yield. For these determinations, I calculate the UL once without systematic uncertainties and then with total uncertainties of 5, 10, and 20 %, as well as with the nominal uncertainty on the signal yield. As can be seen, the impact of systematic uncertainties on the extracted UL values is relatively small, with an increase of approximately 2 % for a total systematic uncertainty of 10 %. Even for a total uncertainty of 20 %, which would constitute an extreme case, the increase in the UL is only about 9 %. Consequently, the study indicates that the analysis is dominated by the integrated luminosity.

Chapter 10

Results

After solidifying the selection procedure (Chapter 4), determining the trigger efficiency measurement and signal efficiency (Chapter 5), studying the impact of corrections and systematic uncertainties (Chapter 8), and defining and testing the signal extraction (Chapter 9), I present the results obtained with the full Run 1 Belle II dataset at $\Upsilon(4S)$ resonance and $\Upsilon(4S)$ off-resonance CM energies. The total integrated luminosity for this search corresponds to $\int \mathcal{L}\mathrm{d}t = 408.11\,\mathrm{fb}^{-1}$.

First, I show the most relevant distributions of the fully unblinded dataset in Section 10.1. Then, I present the results of the trigger efficiency measurement in Section 10.2. Finally, I show the results of the signal extraction in the ALP mass region $m_a \in [0.17, 0.98]\,\mathrm{GeV}/c^2$, including the determination of the significance and the placement of ULs at 95 % CL on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section and the ALP–photon coupling $g_{a\gamma\gamma}$.

10.1 Distributions in Recorded Data

In this section, I present key distributions of recorded data and simulation in the unblinded signal extraction scan region. Figure 10.1 shows the $M_{\gamma\gamma}$ distribution in the full scan range of $M_{\gamma\gamma} \in [0.17, 9.80]\,\mathrm{GeV}/c^2$ including the control regions. Overall, the data and simulation agreement is the same that I observe in the η and ω control regions in Sections 7.2 and 7.3, respectively. The weighted mean relative difference between the data and simulation yields in the full scan range is 3 %.

Nevertheless, deviations between recorded data and simulation are visible in the low and high $M_{\gamma\gamma}$ regions. Figure 10.2 shows the invariant diphoton mass distribution in the regions where the data and simulation agreement is worse compared to the overall agreement. I do not take any additional measures to improve the data and simulation agreement in these regions. Eventually, for the shape and yield of the background PDF, I do not rely on the simulation but determine them directly in the recorded data. Additionally, as I discuss in Section 8.6, I assign a systematic uncertainty to the signal yield to account for the normalisation differences at the order of 4 %.

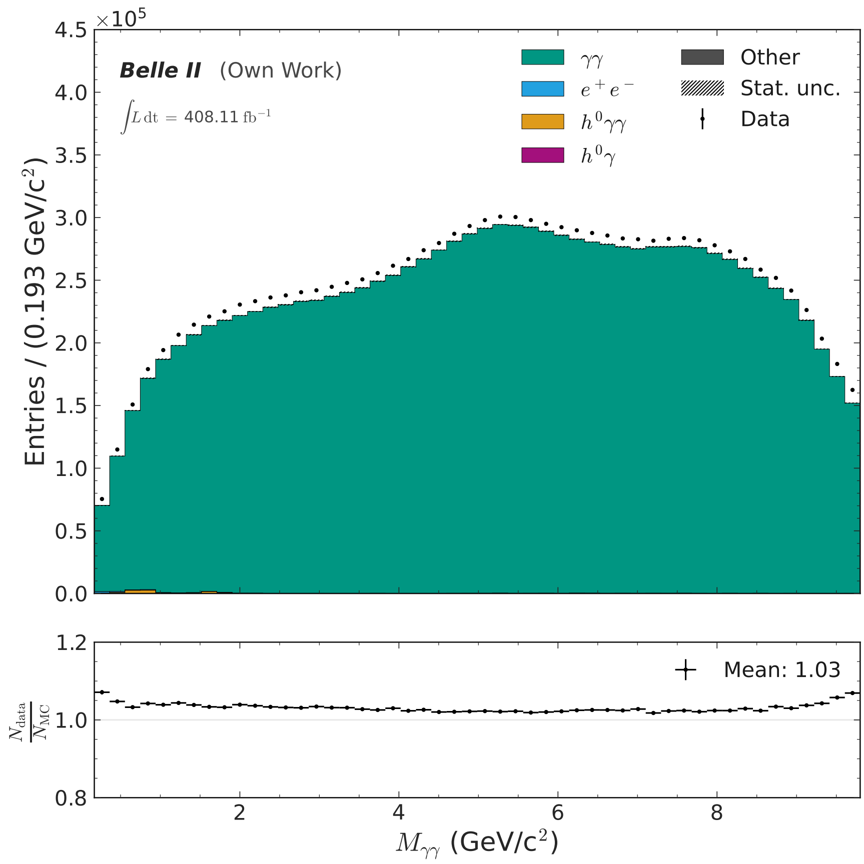
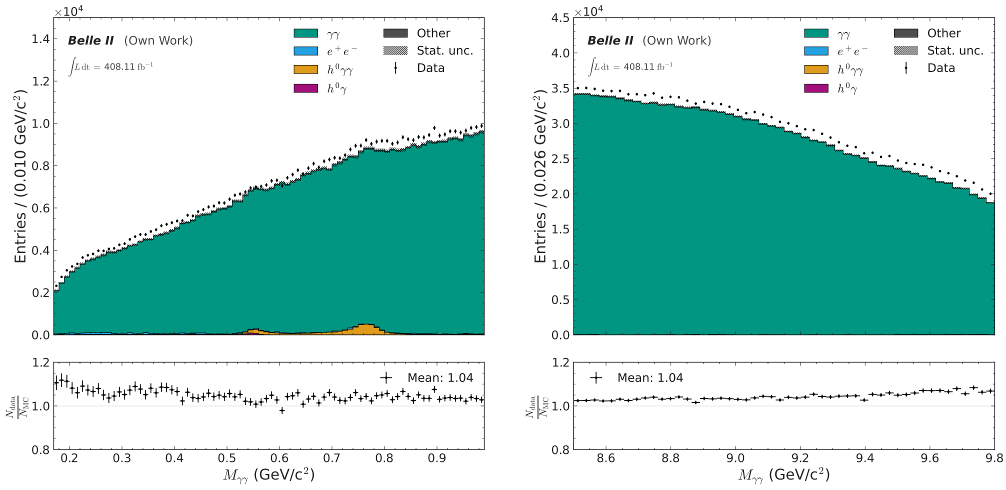


Figure 10.1: Distribution of the invariant diphoton mass $M_{\gamma\gamma}$ in the unblinded signal extraction scan region. The distribution includes events that pass all selection criteria, as well as the L1 trigger and HLT selections. The bottom panel shows the ratio of the recorded data to the total simulation yield.

Lastly, Fig. 10.3 shows the distribution of the kinematic classifier output $P_{\mathrm{Kin}}^{\mathrm{FNN}}$ in recorded data and simulation in the unblinded signal extraction scan region. The kinematic classifier output shows a similar level of agreement between recorded data and simulation as observed in the invariant diphoton mass distribution. Specifically, the shapes of the distributions agree well, which signals that no discrepancies are introduced by the kinematic classifier selection.



(a) $M_{\gamma\gamma} \in [0.17, 1.00]\ \mathrm{GeV}/c^2$ (b) $M_{\gamma\gamma} \in [8.50, 9.80]\ \mathrm{GeV}/c^2$

Figure 10.2: Distributions of the invariant diphoton mass, $M_{\gamma\gamma}$, in the unblinded signal extraction scan region. These are the same distributions as in Fig. 10.1, but zoomed into the low-mass region (left) and high-mass region (right).

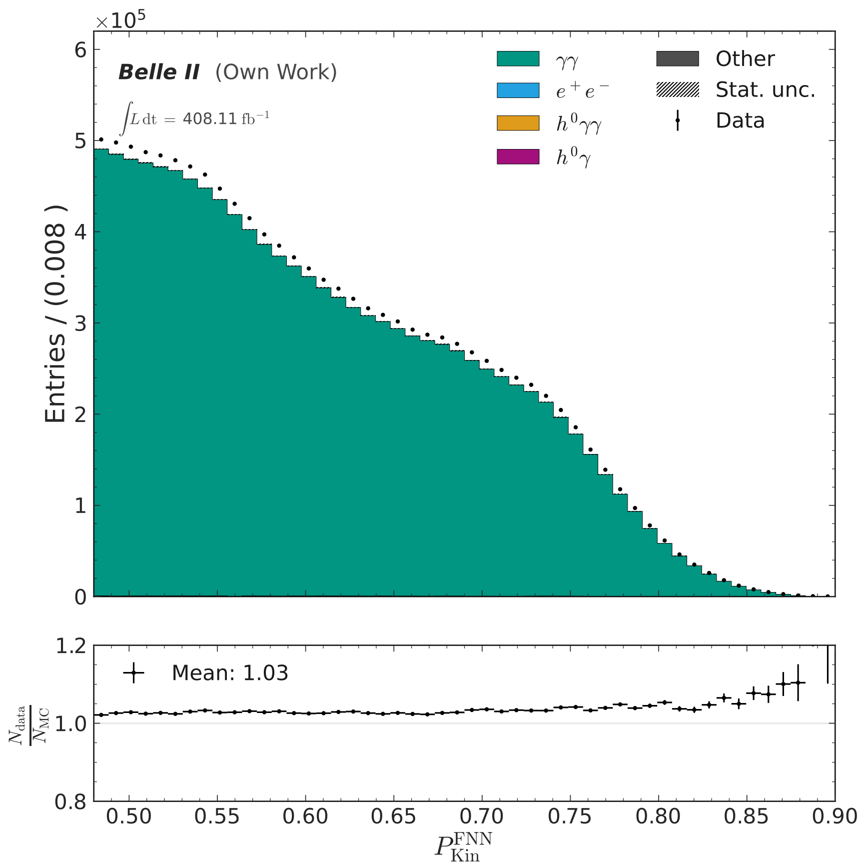


Figure 10.3: Distribution of the kinematic classifier output $P_{\mathrm{Kin}}^{\mathrm{FNN}}$ in the unblinded signal extraction scan region.

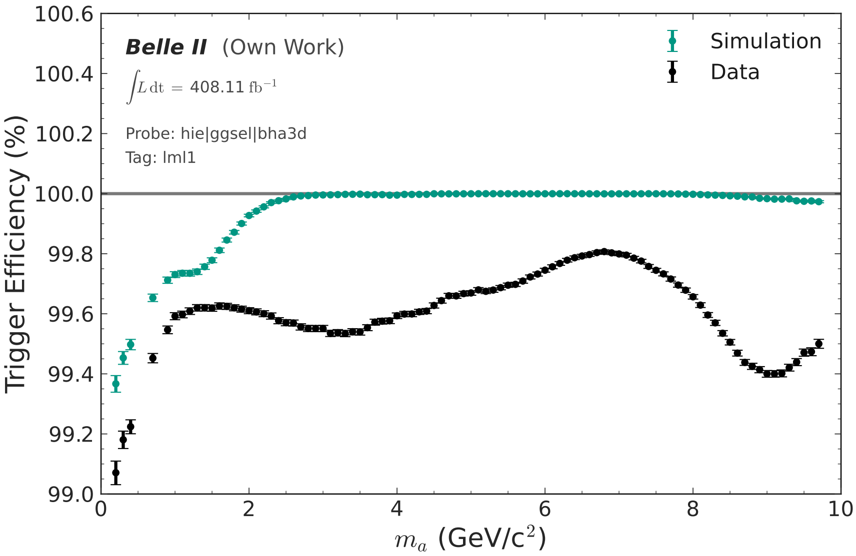


Figure 10.4: Trigger efficiency of (`hie`|`ggsel`|`bha3d`) with respect to `lml1` for HLT triggered events as a function of selected m_a in the signal extraction scan region.

10.2 Trigger Efficiency

As I discuss in Sections 5.1 and 8.4, I measure the trigger efficiency directly on the recorded data in the signal extraction scan region. The main reasons for this includes the unavailabilty of control regions that are sensitive to the full kinematic phase space of the signal, and the interplay between the high-energy and back-to-back photon L1 trigger lines in the final trigger selection. Moreover, several discrepancies in the L1 trigger simulation make a data-driven approach necessary.

To briefly summarise the method: I use a tag-and-probe technique to measure the efficiency of the full L1 trigger selection. As there is no orthogonal trigger line to the L1 trigger selection `hie`|`ggsel`|`bha3d`, I use the `lml1` trigger line as the tag, which is the most sensitive remaining trigger line. The exact definitions, prescales, and run-dependent availability of the mentioned trigger lines are given in Chapter B. For the trigger efficiency measurement, all events pass the general event selection (see Chapter 4) and the HLT selection. I outline in Section 5.1.2 that I do not expect any significant inefficiencies due to the HLT selection for the selected events.

Figure 10.4 shows the trigger efficiency of the full L1 trigger selection `hie`|`ggsel`|`bha3d` with respect to `lml1` as a function of $M_{\gamma\gamma}$ in the signal extraction scan region for simulated and recorded events. For the determination of the trigger efficiency, I calculate the efficiency for each scan ALP mass point in the corresponding $\pm 20\sigma^{\mathrm{CB}}$ fit window (see Chapter 9). The trigger efficiency in simulation is consistently higher than in recorded data by about 0.1 to 0.6 % across the full $M_{\gamma\gamma}$ range. However, the difference is not of concern, as it is very small compared to the overall trigger efficiency of more than 99 %. The largest discrepancies arise in the high ALP mass region around $M_{\gamma\gamma} \approx 8\,\mathrm{GeV}/c^2$, which is the region where the sensitivity of `ggsel` overtakes that of `hie`.

Finally, I apply the measured trigger efficiency from recorded data to correct the interpolated signal efficiency. The effect of this correction on the final signal efficiency is small, with a maximum relative change of less than 1 % at low ALP masses. As I determine the trigger efficiency directly in recorded data, I assign the statistical uncertainty of the efficiency measurement as a systematic uncertainty on the signal efficiency for each scan mass point, as discussed in Section 8.4. However, due to the high statistics of recorded events in the signal extraction scan region, the resulting systematic uncertainty is small, with a maximum of 0.04 % at low ALP masses.

10.3 Model-Independent Result

I perform the signal extraction as described in Chapter 9 for each ALP mass hypothesis in the range of $m_a \in [0.17, 9.80]\,\mathrm{GeV}/c^2$ in steps of $5\,\mathrm{MeV}/c^2$. For each mass hypothesis, I fit the total PDF Eq. (9.1) to the recorded data in the corresponding $\pm 20\sigma^{\mathrm{CB}}$ fit window around the scan mass point. The signal PDF parameters are fixed to the values obtained from the signal simulation, as described in Chapter 6, while the background PDF parameters and the signal and background yields are left free in the fit. After each fit, I determine the significance Z_0 of the signal hypothesis, as described in Section 9.2. In the case of no significant excess above a global significance of 3, I calculate the 95 % CL UL on the signal cross section σ as outlined in Section 9.3.

Figure 10.5 shows the local significance of the excess for each scanned ALP mass point in recorded data. Overall, the significances are consistent with the background-only hypothesis and the expectation from the simulation. The only major difference to Fig. 9.2 is the presence of slight fluctuations in the low ALP mass region, which I attribute to the overall higher fluctuations in this region (see Fig. 10.2a). I observe the highest local significance at a mass hypothesis of $m_a = 0.22\,\mathrm{GeV}/c^2$, with a value of $Z_0^{\mathrm{local}} = 3.3$ which, after considering the look-elsewhere effect as described in Section 9.5.1, corresponds to the global significance $Z_0^{\mathrm{global}} = 1.4$. Figure 10.6 shows the combined fit to the $M_{\gamma\gamma}$ distribution at this mass hypothesis.

As I do not observe any significant excess above the global significance threshold, I proceed to calculate the 95 % CL UL on the signal cross section σ for each scanned ALP mass point. Figure 10.7 shows the calculated UL on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section for each scanned ALP mass point in recorded data. For the majority of the ALP mass hypotheses, the ULs are consistent with the sensitivity expectation from the simulation (see Fig. 9.3). In total, I place 95 % CL ULs between $\sigma = 1\,\mathrm{fb}$ and $\sigma = 20\,\mathrm{fb}$ on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section for ALP masses between $m_a \in [0.17, 9.80]\,\mathrm{GeV}/c^2$.

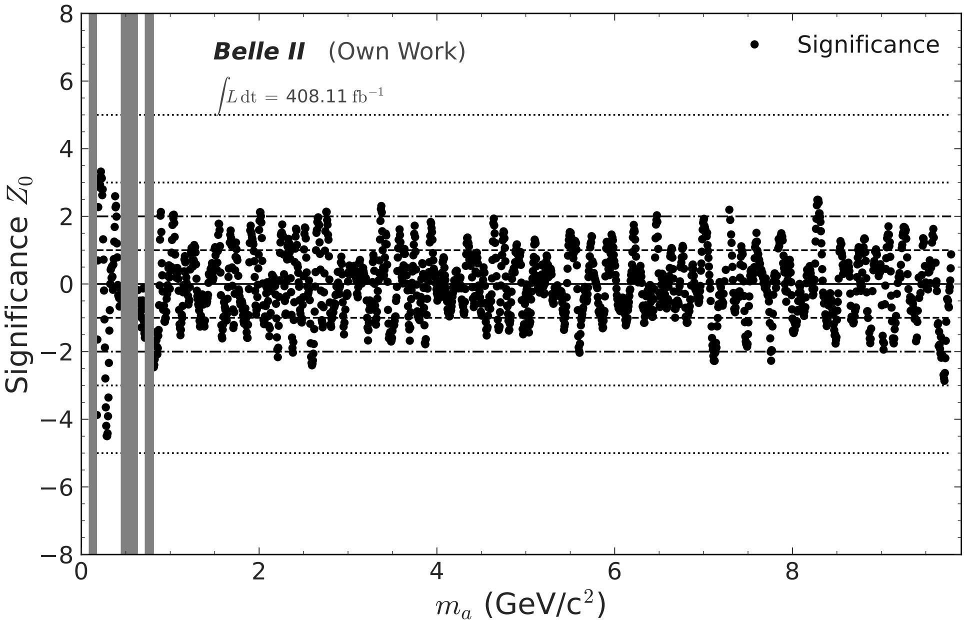


Figure 10.5: Significance of the excess for each scanned ALP mass point in recorded data.

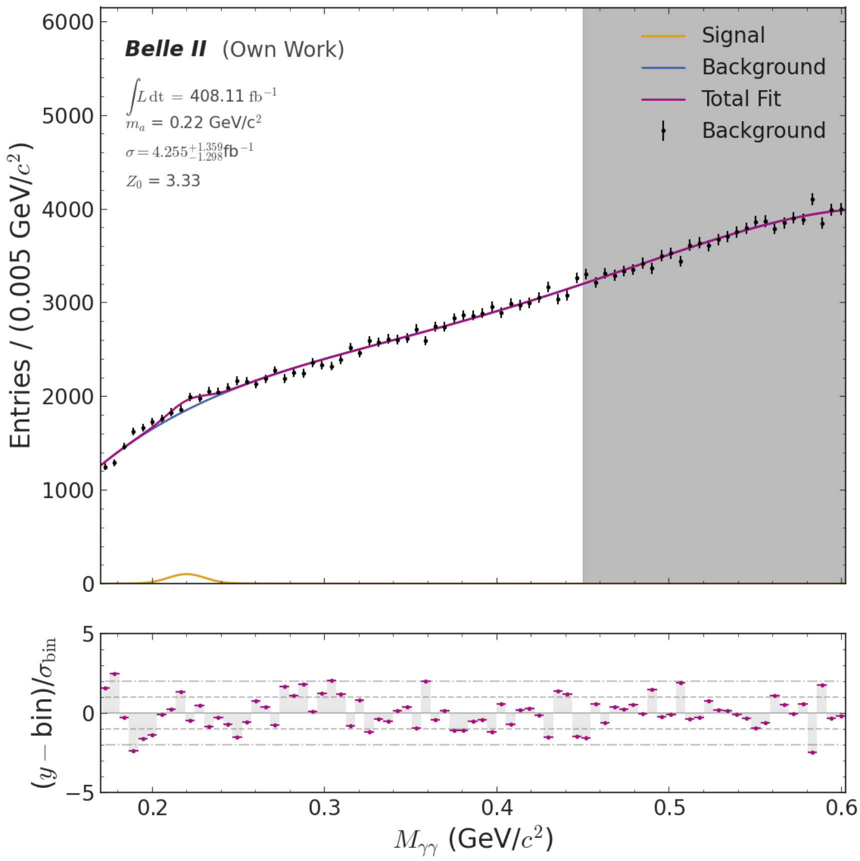


Figure 10.6: Distribution of $M_{\gamma\gamma}$ for recorded data, fitted with the final combined signal and background PDF for the ALP mass hypotheses $m_a = 0.22\,\mathrm{GeV}/c^2$ with the highest observed local significance of $Z_0^{\mathrm{local}} = 3.3$. Solid lines indicate the fitted PDF components: the total background PDF (blue), the combined signal peaking combinatorial component (orange), and the total PDF (black). The grey shaded regions indicate the excluded η mass region.

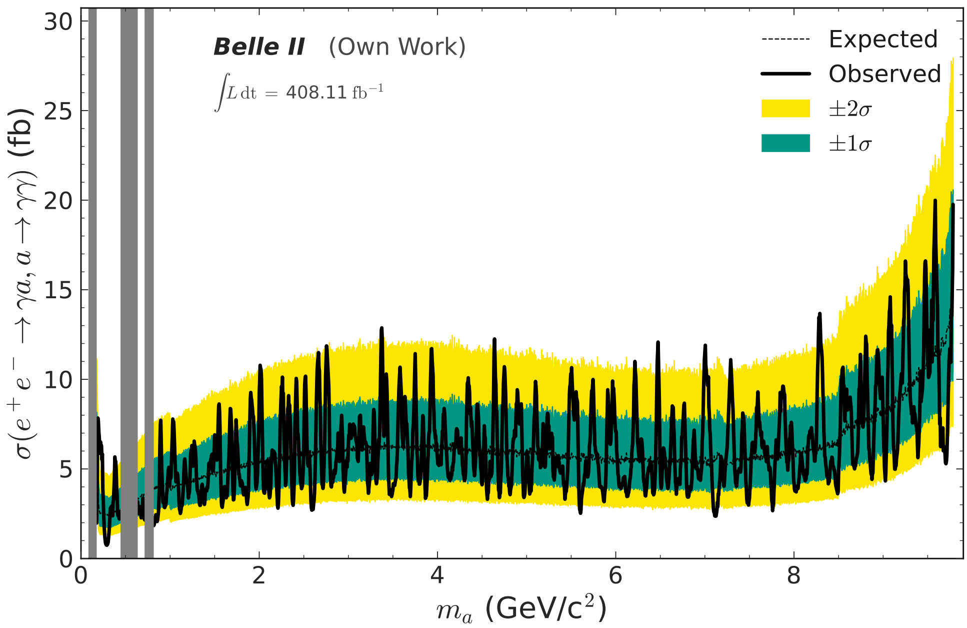


Figure 10.7: 95 % CL UL on the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ cross section for each scanned ALP mass point in recorded data.

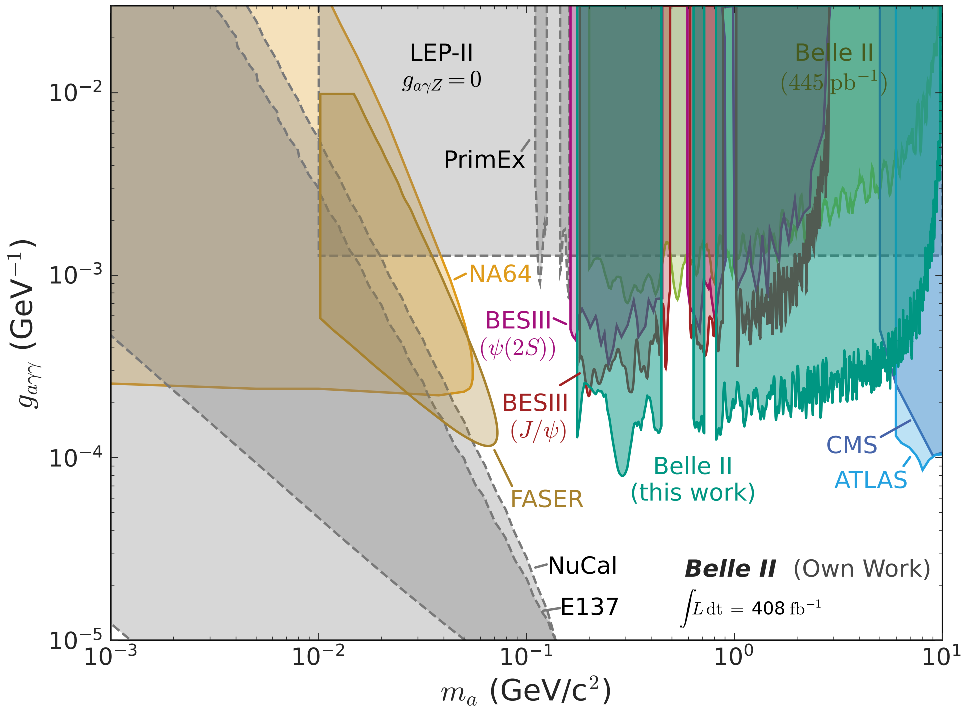


Figure 10.8: 95 % CL UL on the ALP-photon coupling $g_{a\gamma\gamma}$ for each scanned ALP mass point. The constraints of this work are shown as a solid red area. All other $g_{a\gamma\gamma}$ limits are taken from Figure 2.4.

10.4 Model-Dependent Result

Finally, I interpret the extracted 95 % CL ULs on the signal cross section σ in terms of the ALP coupling strength to photons, $g_{a\gamma\gamma}$. For this, I follow the same procedure as I describe in Section 9.4, using Eq. (9.12) to relate the cross section σ to the coupling strength $g_{a\gamma\gamma}$ for each scanned ALP mass point.

Figure 10.8 shows the 95 % CL UL on the ALP-photon coupling $g_{a\gamma\gamma}$. For ALP mass hypotheses between $m_a \geq 5.00\,\mathrm{GeV}/c^2$ and $m_a \leq 9.80\,\mathrm{GeV}/c^2$, the ULs that I place are weaker than the heavy-ion collision results from ATLAS [51] and CMS [52]. For ALP mass hypotheses between $m_a \in [0.17, 5.00]\,\mathrm{GeV}/c^2$, I place 95 % CL UL at the order of $g_{a\gamma\gamma} < 2 \times 10^{-4}\,\mathrm{GeV}^{-1}$ excluding the regions of the η and ω resonance mass regions. Thereby, I improve almost every constraint existing today by a factor of two in this mass region. The only exception is the UL that I set around the region of $m_a \approx 0.22\,\mathrm{GeV}/c^2$, which is slightly weaker than the existing limit by the BESIII measurement at the J/ψ CM energy [58]. In exactly this mass hypothesis region, I observe the highest local significance. The strongest improvement of the existing limits is at $m_a = 0.29\,\mathrm{GeV}/c^2$ at the order of $g_{a\gamma\gamma} < 8 \times 10^{-5}\,\mathrm{GeV}^{-1}$.

Chapter 11

Conclusion

In this thesis, I present a search for ALPs at the Belle II experiment. More specifically, I search for the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process via a significant peak in the reconstructed diphoton mass $M_{\gamma\gamma}$ distribution. The recorded data consists of the full Belle II Run 1 dataset, including runs at the $\Upsilon(4S)$ resonance and $\Upsilon(4S)$ off-resonance, amounting to a total integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 408.11\,\mathrm{fb}^{-1}$.

In comparison to previous work [59], I not only extend the search for roughly 1000 times more integrated luminosity, but also update the analysis strategy to increase the search sensitivity. Particularly, I employ a kinematic fit on the four-momenta of the final-state particles to constrain them to the initial state, which significantly improves the resolution of $M_{\gamma\gamma}$. Furthermore, I introduce an MVA-based classifier in the selection to separate the SM background from the signal signature based on their kinematic properties. In addition, I find a L1 trigger selection that allows a full and highly efficient sensitivity to the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process, even though it partially mimics the vetoed Bhabha scattering $e^+e^- \to e^+e^-(\gamma)$ signature.

I do not find any statistically significant evidence for the peaking $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process and therefore place constraints on the cross section and the model-dependent ALP-photon coupling $g_{a\gamma\gamma}$. For ALP masses between $0.17\,\mathrm{GeV}/c^2$ and $5.00\,\mathrm{GeV}/c^2$, I set the most stringent limits on the ALP-photon coupling, reaching values on the order of $g_{a\gamma\gamma} < 2 \times 10^{-4}\,\mathrm{GeV}^{-1}$ and thus improving upon current limits by a factor of two or more. The only exception is around $m_a \approx 0.22\,\mathrm{GeV}/c^2$, where I observe the highest local significance; in this region, the constraints are comparable to those from the BESIII measurement [58].

Today, the search for the $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process at Belle II is limited by the total integrated luminosity. However, in terms of analysis performance, improvements can be made alongside the run-dependent measurement of photon property resolutions and the determination of L1 trigger efficiency. Calibrated photon resolutions could improve the agreement between recorded data and simulation and eventually enable a selection on the kinematic fit quality, which is a strong separator between true three-photon and other signatures. An alternative approach would be to search for signatures including charged particles, similar to the BESIII searches, thereby improving the kinematic fit quality through a more precise tracking resolution compared to the resolution in fully neutral final states. Improvements in the trigger efficiency study are necessary, particularly to achieve a better understanding of the trigger performance for events with high energetic ECL clusters outside the ECL barrel region and nothing else. The increased sensitivity in this detector region depends on the availability of the relevant trigger lines to perform such studies.

With increasing integrated luminosity at Belle II, the search for $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ will become sensitive to the ALP parameter region in which the ALPs have a non-zero lifetime. This will result in a displaced, neutral decay signature, posing new challenges for the analysis. The limited photon position resolution will significantly impact the sensitivity. A possible approach to adress the displaced decays is to use the full kinematics of the event and to constrain the ALP decay direction by the recoiling photon.

Finally, to further constrain the m_a–$g_{a\gamma\gamma}$ plane, the challenging signature of the vector-boson-fusion process $e^+e^- \to e^+e^-a$, $a \to \gamma\gamma$ could be studied in more detail. While the fully reconstructed final state suffers from low efficiency due to the low scattering angle of the outgoing leptons, the study of signatures with missing four-momentum could utilise the higher cross section in comparison to the ALP-strahlung production process. The key challenges that I see for these future studies are the performance of the Belle II detector at high beam background levels, which result in the signatures of two photons, possibly a single track in the endcaps and missing four-momentum. In addition to the experimental challenges, dedicated studies are needed to better understand the background level of scattering processes at low angles.

To summarise this thesis, I present a search for $e^+e^- \to \gamma a$, $a \to \gamma\gamma$ process at the Belle II experiment and find no statistically significant excess above the SM expectation. Subsequently, I set world-leading constraints on the cross section and the model-dependent ALP-photon coupling for ALP masses between $0.17\,\mathrm{GeV}/c^2$ and $5.00\,\mathrm{GeV}/c^2$.

Danksagung

I would like to take this opportunity to thank some people who contributed to the success of this work.

First of all, I would like to thank Prof. Dr. Torben Ferber for taking on the role of supervisor. His advice has been invaluable, and I have enjoyed the numerous discussions, debates and bets we had during the course of this work. He has an excellent understanding of the physics and technical aspects of the analysis and has always been able to provide guidance when I encountered problems, while also giving me the freedom to shape the work according to my own ideas.

I also want to thank Prof. Dr. Markus Klute for taking on the role of co-supervisor and for his openness and time for questions of any kind.

Furthermore, I would like to hand out a very special thanks to Dr. Giacomo De Pietro, who enabled me the most to do this work. He was not only a great mentor and supervisor, but also a friend and a source of inspiration for me. Ever since I met him, he has been a great support and has always been there for me, both professionally and personally. Especially shortly after my first year, when my initial analysis was no longer feasible, he helped me to find a new direction and motivated me continuously until I found the motivation to work on the new topic. I hope that we have the opportunity to keep working together for a long time, and I am looking forward to our future collaborations.

To my colleagues at the ETP and in the Belle II collaboration, I would like to thank you for creating an atmosphere in which I felt very comfortable and never lost the fun of physics. Dr. Patrick Ecker and Dr. Jonas Eppelt played a special role in this regard, as they were always available to discuss topics of any kind (truly any kind), which enriched me not only in the context of the work. I further extend this list to Dr. Moritz Bauer, Dr. Isabel Haide and Dr. Lea Reuter, who have been there at every step of the way, have helped me discuss the work and made the time off work a little bit more enjoyable. To my former students Frederik Schmitt and Kylian Schmidt, I hope that they have learned just a tiny fraction as much as I have learned from working with them, and I wish them all the best for their future. Last but not least, I would like to thank all my colleagues for their diversity, which has helped me to develop new perspectives on the work (and sometimes beyond).

Abseits des beruflichen Umfelds möchte ich mich bei all meinen Freund:innen bedanken, die es mir ermöglicht haben, auch einmal abzuschalten und damit den nötigen Ausgleich zu schaffen, ohne den diese Arbeit sicherlich nicht möglich gewesen wäre.

Zuallerletzt geht der größte Dank an meine Familie, insbesondere an meine Eltern, die ihr Leben umgeworfen haben, um meiner Schwester und mir so ein großartiges Leben zu ermöglichen, und an meine Partnerin Jasmin, die immer an meiner Seite stand, mich immer unterstützt und motiviert hat und mein Leben mit Liebe gefüllt hat.

Bibliography

[1] **Particle Data Group**, S. Navas et al. “Review of Particle Physics”. In: *Phys. Rev. D* 110 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[2] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[3] S. Weinberg. “A model of leptons”. In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[4] A. Salam and J. C. Ward. “Weak and electromagnetic interactions”. In: *Il Nuovo Cimento* 11 (1959), pp. 568–577. DOI: 10.1007/BF02726525.

[5] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[6] P. W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[8] **CMS Collaboration**, S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[9] **ATLAS Collaboration**, G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[10] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1.

[11] G. Apollinari et al. “Chapter 1: High Luminosity Large Hadron Collider HL-LHC”. In: arXiv (2015). arXiv: 1705.08830.

[12] R. G. Suarez. “The Future Circular Collider (FCC) at CERN”. In: *PoS Proc. Sci.* DISCRETE2020-2021 (2022), p. 009. DOI: 10.22323/1.405.0009.

[13] A. S. Chou et al. “Snowmass Cosmic Frontier Report”. In: arXiv (2022). arXiv: 2211.09978 [hep-ex].

[14] **Belle II Collaboration**, T. Abe et al. "Belle II Technical Design Report". In: arXiv (2010). arXiv: 1011.0352 [physics.ins-det].

[15] **Physics Beyond Collider Study Group**, R. Alemany Fernández et al. "Summary Report of the Physics Beyond Colliders Study at CERN". In: arXiv (2025). arXiv: 2505.00947 [hep-ex].

[16] **Belle II Collaboration**, L. Aggarwal et al. "Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond". In: arXiv (2022). arXiv: 2207.06307 [hep-ex].

[17] J. Jaeckel and A. Ringwald. "The Low-Energy Frontier of Particle Physics". In: *Annu. Rev. Nucl. Part. Sci.* 60 (2010), pp. 405–437. DOI: 10.1146/annurev.nucl.012809.104433.

[18] R. D. Peccei and H. R. Quinn. "CP Conservation in the Presence of Pseudoparticles". In: *Phys. Rev. Lett.* 38 (25 1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440.

[19] R. D. Peccei and H. R. Quinn. "Constraints imposed by CP conservation in the presence of pseudoparticles". In: *Phys. Rev. D* 16 (6 1977), pp. 1791–1797. DOI: 10.1103/PhysRevD.16.1791.

[20] S. Weinberg. "A new light boson?" In: *Phys. Rev. Lett.* 40 (4 1978), pp. 223–226. DOI: 10.1103/PhysRevLett.40.223.

[21] F. Wilczek. "Problem of Strong P and T Invariance in the Presence of Instantons". In: *Phys. Rev. Lett.* 40 (5 1978), pp. 279–282. DOI: 10.1103/PhysRevLett.40.279.

[22] J. Preskill, M. B. Wise, and F. Wilczek. "Cosmology of the invisible axion". In: *Phys. Lett. B* 120 (1983), pp. 127–132. DOI: 10.1016/0370-2693(83)90637-8.

[23] L. Abbott and P. Sikivie. "A cosmological bound on the invisible axion". In: *Phys. Lett. B* 120 (1983), pp. 133–136. DOI: 10.1016/0370-2693(83)90638-X.

[24] M. Dine and W. Fischler. "The not-so-harmless axion". In: *Phys. Lett. B* 120 (1983), pp. 137–141. DOI: 10.1016/0370-2693(83)90639-1.

[25] M. J. Dolan et al. "Revised constraints and Belle II sensitivity for visible and invisible axion-like particles". In: *J. High Energy Phys.* 12 (2017). [Erratum: J. High Energy Phys. 03, 190 (2021)], p. 094. DOI: 10.1007/JHEP12(2017)094.

[26] M. D. Schwartz. *Quantum Field Theory and the Standard Model.* Cambridge University Press, 2014.

[27] R. Crewther et al. "Chiral estimate of the electric dipole moment of the neutron in quantum chromodynamics". In: *Phys. Lett. B* 88 (1979), pp. 123–127. DOI: 10.1016/0370-2693(79)90128-X.

[28] C. Abel et al. “Measurement of the permanent electric dipole moment of the neutron”. In: *Phys. Rev. Lett.* 124 (8 2020), p. 081803. DOI: 10.1103/PhysRevLett.124.081803.

[29] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[30] R. Crewther. “Chirality selection rules and the U(1) problem”. In: *Phys. Lett. B* 70 (1977), pp. 349–354. DOI: 10.1016/0370-2693(77)90675-X.

[31] J. E. Kim. “Weak-Interaction Singlet and Strong CP Invariance”. In: *Phys. Rev. Lett.* 43 (2 1979), pp. 103–107. DOI: 10.1103/PhysRevLett.43.103.

[32] M. Shifman, A. Vainshtein, and V. Zakharov. “Can confinement ensure natural CP invariance of strong interactions?” In: *Nucl. Phys. B* 166 (1980), pp. 493–506. DOI: 10.1016/0550-3213(80)90209-6.

[33] A. Zhitnitskij. “On possible suppression of the axion-hadron interactions (In Russian)”. In: *Yad. Fiz.* 31 (1980), pp. 497–504.

[34] M. Dine, W. Fischler, and M. Srednicki. “A simple solution to the strong CP problem with a harmless axion”. In: *Phys. Lett. B* 104 (1981), pp. 199–202. DOI: 10.1016/0370-2693(81)90590-6.

[35] A. Biekötter and K. Mimasu. “Axions and Axion-like particles: collider searches”. In: arXiv (2025). arXiv: 2508.19358 [hep-ph].

[36] D. Cadamuro and J. Redondo. “Cosmological bounds on pseudo Nambu-Goldstone bosons”. In: *J. Cosmol. Astropart. Phys.* 02 (2012), p. 032. DOI: 10.1088/1475-7516/2012/02/032.

[37] M. J. Dolan et al. “A taste of dark matter: flavour constraints on pseudoscalar mediators”. In: *J. High Energy Phys.* 03 (2015). [Erratum: J. High Energy Phys. 07, 103 (2015)], p. 171. DOI: 10.1007/JHEP03(2015)171.

[38] K. Choi et al. “Minimal flavor violation with axion-like particles”. In: *J. High Energy Phys.* 11 (2017), p. 070. DOI: 10.1007/JHEP11(2017)070.

[39] E. Izaguirre, T. Lin, and B. Shuve. “Searching for Axionlike Particles in Flavor-Changing Neutral Current Processes”. In: *Phys. Rev. Lett.* 118 (11 2017), p. 111802. DOI: 10.1103/PhysRevLett.118.111802.

[40] W. J. Marciano et al. “Contributions of axionlike particles to lepton dipole moments”. In: *Phys. Rev. D* 94 (11 2016), p. 115033. DOI: 10.1103/PhysRevD.94.115033.

[41] W. Altmannshofer et al. “The Belle II Physics Book”. In: *Prog. Theor. Exp. Phys.* 2019 (2019). Ed. by E. Kou and P. Urquijo. [Erratum: Prog. Theor. Exp. Phys. 2020, 029201 (2020)], p. 123C01. DOI: 10.1093/ptep/ptz106.

[42] Y. S. Tsai. “Axion bremsstrahlung by an electron beam”. In: *Phys. Rev. D* 34 (5 1986), pp. 1326–1331. DOI: 10.1103/PhysRevD.34.1326.

[43] E. M. Riordan et al. “Search for short-lived axions in an electron-beam-dump experiment”. In: *Phys. Rev. Lett.* 59 (7 1987), pp. 755–758. DOI: 10.1103/PhysRevLett.59.755.

[44] J. D. Bjorken et al. “Search for neutral metastable penetrating particles produced in the SLAC beam dump”. In: *Phys. Rev. D* 38 (11 1988), pp. 3375–3386. DOI: 10.1103/PhysRevD.38.3375.

[45] D. Aloni et al. “Photoproduction of Axionlike Particles”. In: *Phys. Rev. Lett.* 123 (7 2019), p. 071801. DOI: 10.1103/PhysRevLett.123.071801.

[46] F. Bergsma and J. others. “Search for axion-like particle production in 400 gev proton-copper interactions”. In: *Phys. Lett. B* 157 (1985), pp. 458–462. DOI: 10.1016/0370-2693(85)90400-9.

[47] J. Blümlein et al. “Limits on neutral light scalar and pseudoscalar particles in a proton beam dump experiment”. In: *Z. Phys. C: Part. Fields* 51 (1991), pp. 341–350. DOI: 10.1007/BF01548556.

[48] J. Blümlein et al. “Limits on the mass of light (pseudo)scalar particles from bethe–heitler e^+e^- and $\mu^+\mu^-$ pair production in a proton–iron beam dump experiment”. In: *Int. J. Mod. Phys. A* 7 (1992), pp. 3835–3849. DOI: 10.1142/S0217751X9200171X.

[49] **NA64 Collaboration**, D. Banerjee et al. “Search for Axionlike and Scalar Particles with the NA64 Experiment”. In: *Phys. Rev. Lett.* 125 (2020), p. 081801. DOI: 10.1103/PhysRevLett.125.081801. eprint: 2005.02710.

[50] **FASER Collaboration**, R. Mammen Abraham et al. “Shining light on the dark sector: search for axion-like particles and other new physics in photonic final states with FASER”. In: *J. High Energy Phys.* 01 (2025), p. 199. DOI: 10.1007/JHEP01(2025)199.

[51] **ATLAS Collaboration**, G. Aad et al. “Measurement of light-by-light scattering and search for axion-like particles with 2.2 nb^{-1} of Pb+Pb data with the ATLAS detector”. In: *J. High Energy Phys.* 03 (2021). [Erratum: J. High Energy Phys. 11, 050 (2021)], p. 243. DOI: 10.1007/JHEP11(2021)050.

[52] **CMS Collaboration**, A. M. Sirunyan et al. “Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at $\sqrt{s_{\text{NN}}} =$ 5.02 TeV”. In: *Phys. Lett. B* 797 (2019), p. 134826. DOI: 10.1016/j.physletb.2019.134826.

[53] S. Knapen et al. “Searching for Axionlike Particles with Ultraperipheral Heavy-Ion Collisions”. In: *Phys. Rev. Lett.* 118 (17 2017), p. 171801. DOI: 10.1103/PhysRevLett.118.171801.

[54] G. Abbiendi and T. O. Collaboration. "Multi-photon production in e^+e^- collisions at $\sqrt{s} = 181$–209 GeV". In: *Eur. Phys. J. C* 26 (2003), pp. 331–344. DOI: 10.1140/epjc/s2002-01074-5.

[55] **BaBar Collaboration**, P. del Amo Sanchez et al. "Search for Production of Invisible Final States in Single-Photon Decays of $\Upsilon(1S)$". In: *Phys. Rev. Lett.* 107 (2 2011), p. 021804. DOI: 10.1103/PhysRevLett.107.021804.

[56] **BaBar Collaboration**, J. P. Lees et al. "Search for Invisible Decays of a Dark Photon Produced in e^+e^- Collisions at BaBar". In: *Phys. Rev. Lett.* 119 (2017), p. 131804. DOI: 10.1103/PhysRevLett.119.131804. arXiv: 1702.03327 [hep-ex].

[57] **BESIII Collaboration**, M. Ablikim et al. "Search for an axion-like particle in radiative J/ψ decays". In: *Phys. Lett. B* 838 (2023), p. 137698. DOI: 10.1016/j.physletb.2023.137698.

[58] **BESIII Collaboration**, M. Ablikim et al. "Search for diphoton decays of an axionlike particle in radiative J/ψ decays". In: *Phys. Rev. D* 110 (2024), p. L031101. DOI: 10.1103/PhysRevD.110.L031101.

[59] **Belle II Collaboration**, F. Abudinén et al. "Search for Axionlike Particles Produced in e^+e^- Collisions at Belle II". In: *Phys. Rev. Lett.* 125 (2020), p. 161806. DOI: 10.1103/PhysRevLett.125.161806.

[60] **SuperKEKB Accelerator Team**, K. Akai, K. Furukawa, and H. Koiso. "SuperKEKB Collider". In: *Nucl. Instrum. Methods Phys. Res. A* 907 (2018), p. 188. DOI: 10.1016/j.nima.2018.08.017.

[61] **Belle II Upgrades Working Group**, H. Aihara et al. *The Belle II Detector Upgrades Framework Conceptual Design Report.* 2024. DOI: 10.48550/arXiv.2406.19421. arXiv: 2406.19421 [hep-ex].

[62] **Belle II Collaboration**, F. Forti. "Snowmass Whitepaper: The Belle II Detector Upgrade Program". In: arXiv (2022). arXiv: 2203.11349 [hep-ex].

[63] **SuperB Collaboration**, M. Bona et al. "SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report". In: arXiv (2007). arXiv: 0709.0451 [hep-ex].

[64] D. Zhou et al. "Luminosity performance of SuperKEKB". In: *J. Inst.* 19 (2024), T02002. DOI: 10.1088/1748-0221/19/02/T02002.

[65] M. Bauer. "Measuring the Branching Fraction of $B \to \rho \ell \nu_\ell$ Decays with the Belle II Experiment". PhD thesis. Karlsruher Institut für Technologie (KIT), 2023. 161 pp. DOI: 10.5445/IR/1000165627.

[66] **Belle II DEPFET and PXD Collaboration**, H. Ye et al. "Commissioning and Performance of the Belle II Pixel Detector". In: *Nucl. Instrum. Methods Phys. Res. A* 987 (2021), p. 164875. DOI: 10.1016/j.nima.2020.164875.

[67] **Belle II DEPFET and PXD Collaboration**, P. Ahlburg et al. "The New and Complete Belle II DEPFET Pixel Detector: Commissioning and Previous Operational Experience". In: *Nucl. Instrum. Methods Phys. Res. A* 1068 (2024), p. 169763. DOI: 10.1016/j.nima.2024.169763.

[68] **Belle II SVD Collaboration**, K. Adamczyk et al. "The design, construction, operation and performance of the Belle II silicon vertex detector". In: *J. Inst.* 17 (2022), P11042. DOI: 10.1088/1748-0221/17/11/P11042.

[69] N. Taniguchi. "Central Drift Chamber for Belle-II". In: *J. Inst.* 12 (2017), p. C06014. DOI: 10.1088/1748-0221/12/06/C06014.

[70] T. V. Dong et al. "Calibration and Alignment of the Belle II Central Drift Chamber". In: *Nucl. Instrum. Methods Phys. Res. A* 930 (2019), pp. 132–141. DOI: 10.1016/j.nima.2019.03.072.

[71] P. Križan. "Particle identification at Belle II". In: *J. Inst.* 9 (2014), p. C07018. DOI: 10.1088/1748-0221/9/07/C07018.

[72] A. Kuzmin. "Electromagnetic calorimeter of Belle II". In: *Nucl. Instrum. Methods Phys. Res. A* 958 (2020), p. 162235. DOI: 10.1016/j.nima.2019.05.076.

[73] **Belle II Framework Software Group**, T. Kuhr et al. "The Belle II Core Software". In: *Comput. Softw. Big Sci.* 3 (2019), p. 1. DOI: 10.1007/s41781-018-0017-9.

[74] Belle II collaboration. *Belle II Analysis Software Framework (basf2).* https://doi.org/10.5281/zenodo.5574115.

[75] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *J. High Energy Phys.* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079.

[76] S. Frixione et al. "Lepton collisions in MadGraph5_aMC@NLO". In: arXiv (2021). arXiv: 2108.10261 [hep-ph].

[77] **Geant4 Collaboration**, S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nucl. Instrum. Methods Phys. Res. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[78] A. Natochii et al. "Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider". In: *Nucl. Instrum. Methods Phys. Res. A* 1055 (2023), p. 168550. DOI: 10.1016/j.nima.2023.168550.

[79] G. Balossini et al. "Photon pair production at flavour factories with per mille accuracy". In: *Phys. Lett. B* 663 (2008), pp. 209–213. DOI: 10.1016/j.physletb.2008.04.007.

[80] S. Jadach, B. F. L. Ward, and Z. Wąs. "The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions". In: *Comput. Phys. Commun.* 130 (2000), p. 260. DOI: 10.1016/S0010-4655(00)00048-5.

[81] T. Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024.

[82] F. Berends, P. Daverveldt, and R. Kleiss. "Complete lowest-order calculations for four-lepton final states in electron-positron collisions". In: *Nucl. Phys. B* 253 (1985), pp. 441–463. DOI: 10.1016/0550-3213(85)90541-3.

[83] H. Czyż, M. Gunia, and J. H. Kühn. "Simulation of electron-positron annihilation into hadrons with the event generator phokhara". In: *J. High Energy Phys.* 08 (2013), p. 110. DOI: 10.1007/JHEP08(2013)110.

[84] S. Jadach, J. H. Kuhn, and Z. Wąs. "TAUOLA: A library of Monte Carlo programs to simulate decays of polarized tau leptons". In: *Comput. Phys. Commun.* 64 (1990), p. 275. DOI: 10.1016/0010-4655(91)90038-M.

[85] S. Jadach et al. "The Monte Carlo program KoralW version 1.51 and the concurrent Monte Carlo KoralWYFSWW3 with all background graphs and first-order corrections to W-pair production". In: *Comput. Phys. Commun.* 140 (2001), pp. 475–512. DOI: 10.1016/S0010-4655(01)00296-X.

[86] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Methods Phys. Res. A* 462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[87] **Belle II Collaboration**, I. Adachi et al. "Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment". In: *Chin. Phys. C* 49 (2025), p. 013001. DOI: 10.1088/1674-1137/ad806c.

[88] T. Ferber. *OrcaKinFit: Kinematic Fitting for Belle II.* Internal Note: BELLE2-NOTE-PH-2017-002. 2017. URL: https://docs.belle2.org/pub_data/documents/435/.

[89] G. Punzi. "Sensitivity of searches for new signals and its optimization". In: *eConf* C030908 (2003), MODT002. arXiv: physics/0308063.

[90] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (1986), pp. 533–536. DOI: 10.1038/323533a0.

[91] A. Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: arXiv (2019). arXiv: 1912.01703 [cs.LG].

[92] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830. DOI: 10.5555/1953048.2078195.

[93] L. Biewald. *Experiment tracking with weights and biases.* Software available from wandb.com. 2020. URL: https://www.wandb.com/.

[94] G. C. Fox and S. Wolfram. "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes". In: *Phys. Rev. Lett.* 41 (23 1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581.

[95] A. Hershenhorn. “Search for axion like particles with the BaBar detector and photon hadron separation using Zernike moments at Belle II”. PhD thesis. University of British Columbia, 2021. DOI: 10.14288/1.0398055.

[96] G. De Pietro et al. *Search for Inelastic Dark Matter produced in association with a Dark Higgs.* Internal Note: BELLE2-NOTE-PH-2024-019. 2024. URL: https://docs.belle2.org/pub_data/documents/42/.

[97] J. Eschle et al. “zfit: Scalable pythonic fitting”. In: *SoftwareX* 11 (2020), p. 100508. DOI: 10.1016/j.softx.2020.100508.

[98] J. Gaiser. “Charmonium spectroscopy from radiative decays of the J/ψ and ψ'”. PhD thesis. Stanford University, 1982.

[99] T. Skwarnicki. “A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances”. PhD thesis. Cracow, INP, 1986.

[100] L. Cao and K. Tackmann. *Calibration of photon detection efficiency of the Belle II electromagnetic calorimeter using radiative muon pair events.* Internal Note: BELLE2-NOTE-TE-2024-002. 2024. URL: https://docs.belle2.org/pub_data/documents/471/.

[101] Q. Ji, S. Jia, and C. Shen. *Energy resolution and bias with $E_\gamma < 2.0$ GeV using symmetric decays of $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ at Belle II.* Internal Note: BELLE2-NOTE-PH-2021-021. 2021. URL: https://docs.belle2.org/pub_data/documents/230/.

[102] D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization.* New York: John Wiley & Sons, Inc., 1992. DOI: 10.1002/9780470316849.

[103] B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Vol. 26. Monographs on Statistics and Applied Probability. London: Chapman and Hall, 1986.

[104] W. Kilian, T. Ohl, and J. Reuter. “WHIZARD: Simulating Multi-Particle Processes at LHC and ILC”. In: *Eur. Phys. J. C* 71 (2011), p. 1742. DOI: 10.1140/epjc/s10052-011-1742-y.

[105] M. Moretti, T. Ohl, and J. Reuter. “O’Mega: An Optimizing Matrix Element Generator”. In: arXiv (2001). arXiv: hep-ph/0102195.

[106] Scikit-HEP Project. *hepstats: Statistical tools for HEP analysis.* https://github.com/scikit-hep/hepstats. 2024.

[107] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0.

[108] E. Gross and O. Vitells. “Trial factors for the look elsewhere effect in high energy physics”. In: *Eur. Phys. J. C* 70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8.

Appendices

Appendix A

Feature for MVA-based Selection

A.1 Features for Kinematic Selection

In this appendix, all the feature distributions presented to the MVA classifier discussed in Section 4.4.2 are displayed. The distributions are for all the different background components considered and the signal samples for $m_a = 0.2\,\mathrm{GeV}/c^2$, $m_a = 5.0\,\mathrm{GeV}/c^2$, and $m_a = 9.8\,\mathrm{GeV}/c^2$.

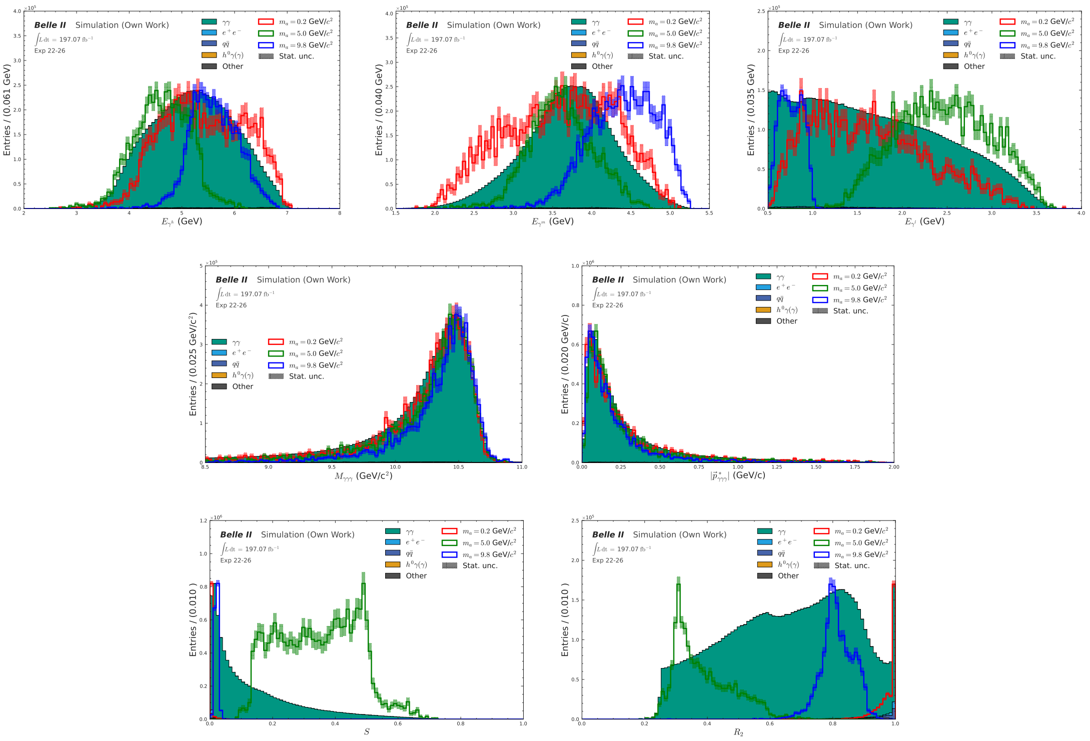


Figure A.1: Distributions of the input features before normalisation for the first stage of the MVA-based selection in Section 4.4.2. Filled and stacked histograms display the total background contribution for experiments 22, 24, 25, and 26. Unfilled, outlined histograms show the variable distribution for three example signal MC samples, simulated with beam background conditions corresponding to experiment 26, run 1485. The input features shown in the first row are the energy of the photon candidate clusters E_{γ^i}. The left plot shows the distributions for the high-energetic photon, middle plot for the middle-energetic photon, and right plot for the low-energetic photon. The second and third rows display the invariant mass distributions for the photon pairs $M_{\gamma\gamma\gamma}$, the total CM momentum magnitude $|\mathbf{p}^*_{\gamma\gamma\gamma}|$, the sphericity S, and R_2.

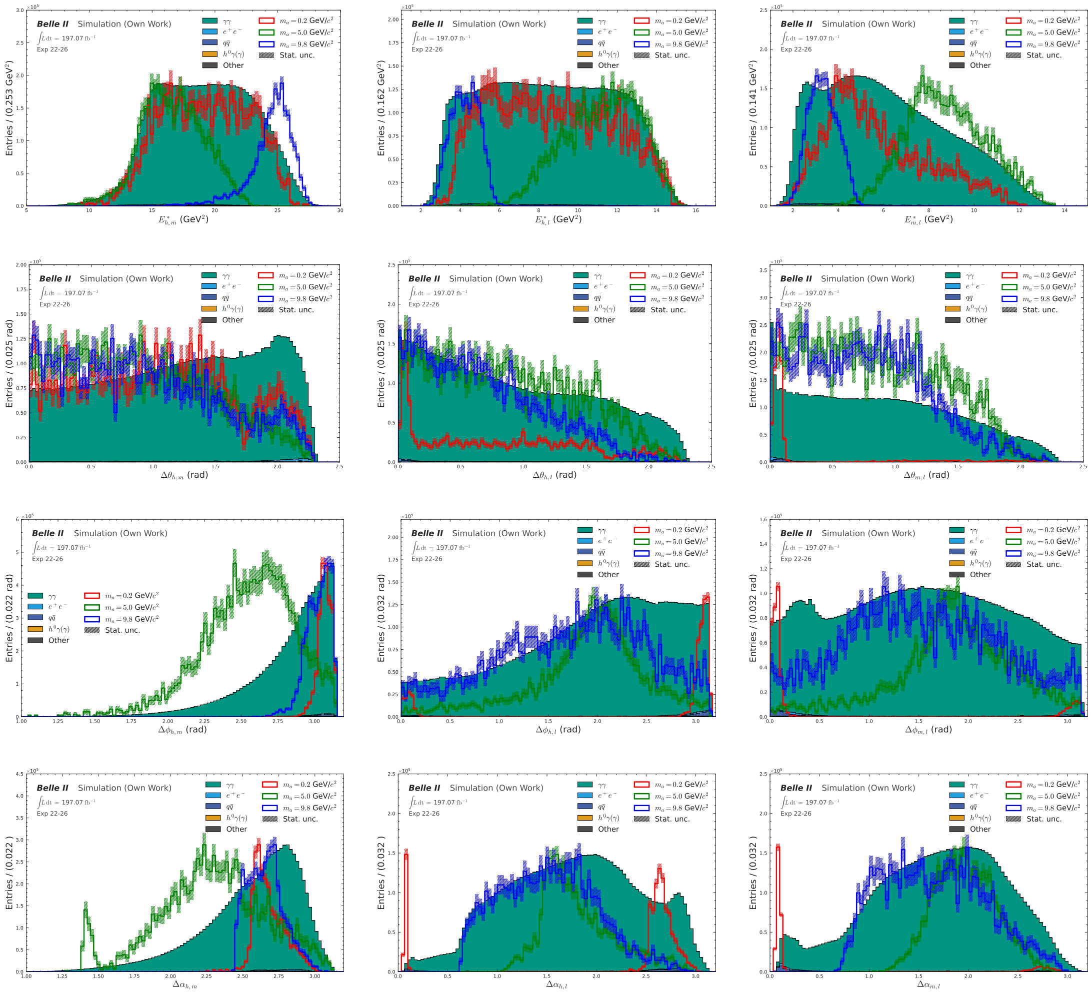


Figure A.2: Distributions of the input features before normalisation for the first stage of the MVA-based selection in Section 4.4.2. Filled and stacked histograms display the total background contribution for experiment 26, run 2485. Unfilled, outlined histograms show the variable distribution for three example signal MC samples, simulated with beam background conditions corresponding to the same run. The input features shown here are the energy product $E^*_{i,j}$, the polar $\Delta\theta_{i,j}$, the azimuthal $\Delta\phi_{i,j}$ angle difference, and the opening angle $\Delta\alpha_{i,j}$. Left plots show the distributions for the combinations of the high- and middle-energetic photons, middle plots for the high- and low-energetic photons, and right plots for the middle- and low-energetic photons.

A.2 Features for Merged Photon Selection

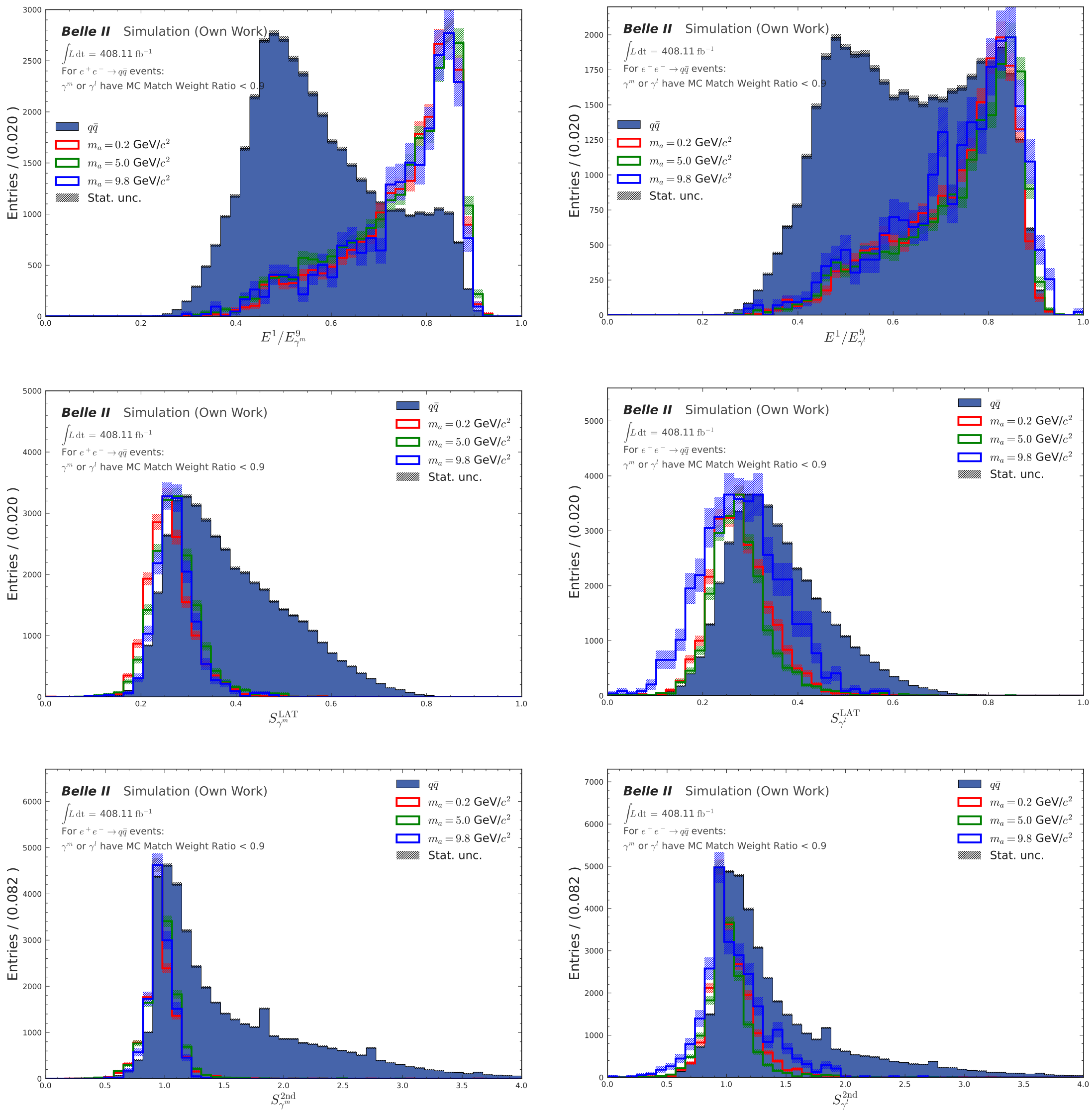


Figure A.3: Distributions of the input features before normalisation for the second stage of the MVA-based selection in Section 4.4.3. Filled and stacked histograms display the $e^+e^- \to u\overline{u}$ background contribution for the $\Upsilon(4S)$ runs. Unfilled, outlined histograms show the variable distribution for three example signal MC samples, simulated with beam background conditions corresponding to experiment 26, run 1485. The input features shown here are the cluster neighbour energy ratio $E^1/E^9_{\gamma^i}$, the cluster lateral energy distribution $S^{\mathrm{LAT}}_{\gamma^i}$, and the cluster second moment $S^{\mathrm{2nd}}_{\gamma^i}$. Left plots show the distributions for the middle-energetic photon, and right plots for the low-energetic photon.

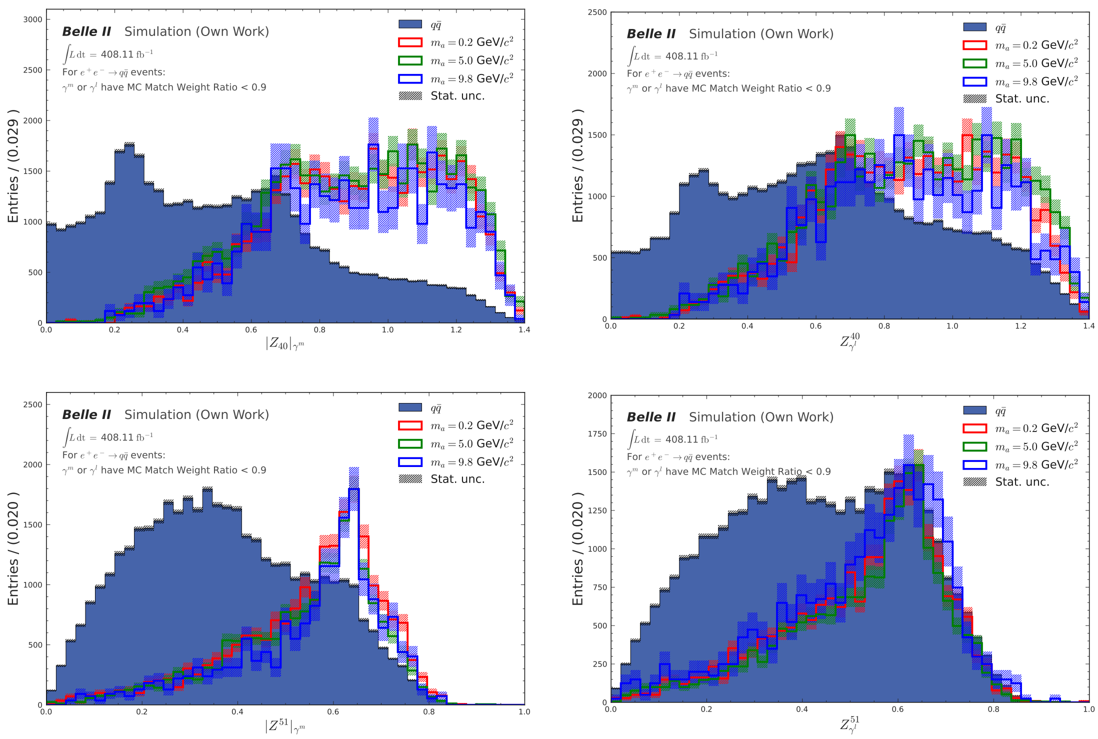


Figure A.4: Distributions of the input features before normalisation for the second stage of the MVA-based selection in Section 4.4.3. Filled and stacked histograms display the $e^+e^- \to u\bar{u}$ background contribution for the $\Upsilon(4S)$ runs. Unfilled, outlined histograms show the variable distribution for three example signal MC samples, simulated with beam background conditions corresponding to experiment 26, run 1485. The input features shown here are the cluster Zernike moments $|Z_{40}|_{\gamma^i}$ and $|Z_{51}|_{\gamma^i}$. Left plots show the distributions for the middle-energetic photon, and right plots for the low-energetic photon.

Appendix B

Trigger Bit Definitions

Table B.1: Definition of the output trigger lines used in this analysis, their availability and the prescale factor. The prescale factor N represents the saving of events only after every Nth occurence. The information presented here is based on the official Belle II trigger documentation.

Line	Definition	Availability	Prescale
`hie`	`ehigh` & `!bha_veto` & `!veto`	experiment 7-26	1
`ggsel`	`ecl_3dbha` &	experiment 7-12	0
	`!trkbha1` &	experiment 14	
	`!trkbha2` &	run 694-784	0
	`bha_intrk`	run 785-1752	1
		run 1753-1772	10
		run 1773-1971	1
		run 1972-2040	0
		run 2041-2135	1
		experiment 16-26	1
`bha3d`	`ecl_3dbha` & `!veto`	experiment 7-12	1
		experiment 14	
		run 694-1751	1
		run 1752-1953	10
		run 1954-2135	1
		experiment 16-17	1
		experiment 18-26	100
`lml1`	`ecl_lml_1` & `!veto`	experiment 7-12	1
		experiment 14	
		run 694-1752	1
		run 1753-1953	10
		run 1954-2135	1
		experiment 16-17	1
		experiment 18-26	2

Table B.2: Definition of the input bits used for the output trigger lines. The input bits are defined in the Belle II trigger documentation. Additions to the definitions are made by me based on the trigger simulation code in basf2.

Input Bit	**Definition**
`ehigh`	Total energy in the ECL greater 1 GeV with θID = 2-15
`bha_veto`	For experiment 7-9 `bha_veto` = `ecl_bha` For experiment 10-26 `bha_veto` = `ecl_3dbha`
`ecl_3dbha`	$165^\circ < \sum \theta_{\mathrm{CM}} < 190^\circ$ & $160^\circ < \Delta\phi_{\mathrm{CM}} < 200^\circ$ & $E(\mathrm{CL1}) > 3\,\mathrm{GeV}$ & $E(\mathrm{CL2}) > 3\,\mathrm{GeV}$ & $(E(\mathrm{CL1}) > 4.5\,\mathrm{GeV} \mid E(\mathrm{CL2}) > 4.5\,\mathrm{GeV})$ ($\sum \theta_{\mathrm{CM}}$ sum of polar angles of two cluster $\Delta\phi_{\mathrm{CM}}$ difference of azimuthal angles of two cluster $E(\mathrm{CLX})$ are the energies of the two clusters in the CM frame)
`ecl_bha`	Belle 2D Bhabha veto, similar to `ecl_3dbha` but with discret sums of ECL trigger cell ϕ rings and different energy thresholds
`trkbhaX`	Number of X 3D Bhabha clusters matches number of tracks in distance of at most two trigger cells in ϕID
`bha_intrk`	At least one of 3D Bhabha clusters is with θID = 3-15
`ecl_lml_1`	One cluster with more than 2 GeV with θID = 4-14
`veto`	SuperKEKB Injection veto

Appendix C

$q\bar{q}$ Final States in the Control Regions

Table C.1: Overview of the most frequent final states in the $e^+e^- \to q\bar{q}$ simulation that pass the selection criteria and fall within the η control region. The fractions are calculated with respect to the total number of $e^+e^- \to q\bar{q}$ events that satisfy the selection. Final states contributing less than 0.1 % are not shown. The final states are identified using the `MCDecayString` information provided by the basf2 framework. Additional photons originating from ISR or FSR are not included in this overview.

Final State	**Fraction** (%)
$\eta\pi^0$	49.27
$\pi^0\pi^0$	43.45
$\eta\pi^0\pi^0$	2.86
$\omega\pi^0$	1.35
$\eta\pi^+\pi^-$	0.73
$\eta\pi^+\pi^-\pi^0$	0.36
$\eta\pi^0\pi^0\pi^0$	0.27
$\pi^0\pi^0\pi^0$	0.24
$\eta\eta$	0.24
$\eta\eta\pi^0$	0.15
$\eta\pi^\pm\rho^\mp$	0.15

Table C.2: Overview of the most frequent final states in the $e^+e^- \to q\bar{q}$ simulation that pass the selection criteria and fall within the ω control region. The fractions are calculated with respect to the total number of $e^+e^- \to q\bar{q}$ events that satisfy the selection. Final states contributing less than 0.1 % are not shown. The final states are identified using the `MCDecayString` information provided by the basf2 framework. Additional photons originating from ISR or FSR are not included in this overview.

Final State	**Fraction** (%)
$\pi^0\pi^0$	69.95
$\eta\pi^0$	16.40
$\omega\pi^0$	9.21
$\omega\pi^0\pi^0$	1.06
$\pi^0\pi^0\pi^0$	0.78
$\eta\eta$	0.56
$\eta\pi^0\pi^0$	0.28
$\eta\pi^+\pi^-$	0.22
$\eta\omega$	0.19
$\pi^0\pi^0\pi^0\pi^0$	0.15
$\eta\pi^0\pi^0\pi^0$	0.15
$\omega\pi^0\pi^0\pi^0$	0.11
$\eta'\pi^0$	0.11

Appendix D

Features for Kinematic Selection in the Control Regions

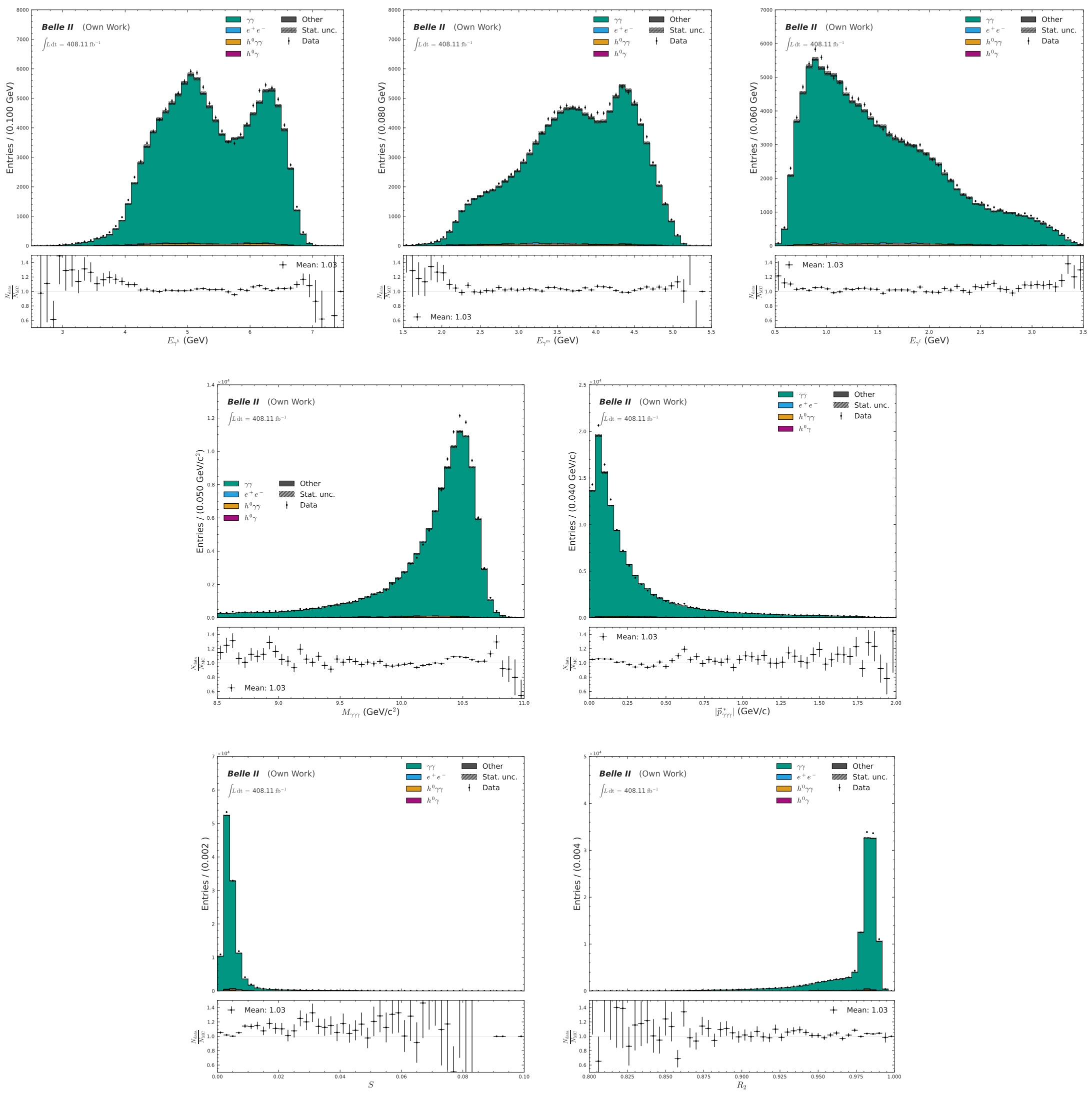


Figure D.1: Distributions of the input features for the first stage of the MVA-based selection. Filled and stacked histograms display the total background contribution for Run 1 simulation in the η control region. Black errorbar dots display the corresponding data. Same features are shown as in Figure A.1.

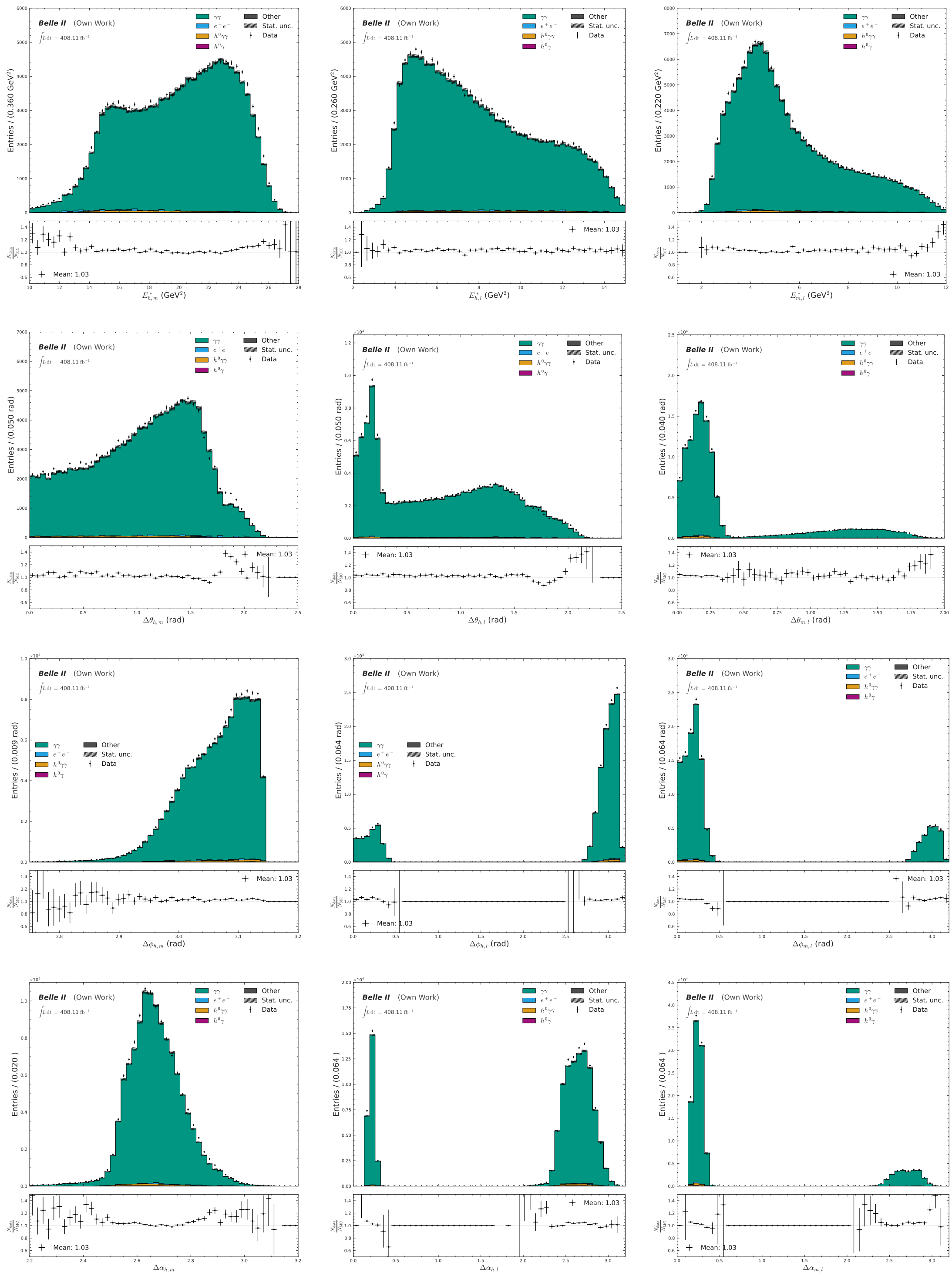


Figure D.2: Distributions of the input features for the first stage of the MVA-based selection. Filled and stacked histograms display the total background contribution for Run 1 simulation in the η control region. Black errorbar dots display the corresponding data. Same features are shown as in Figure A.2.

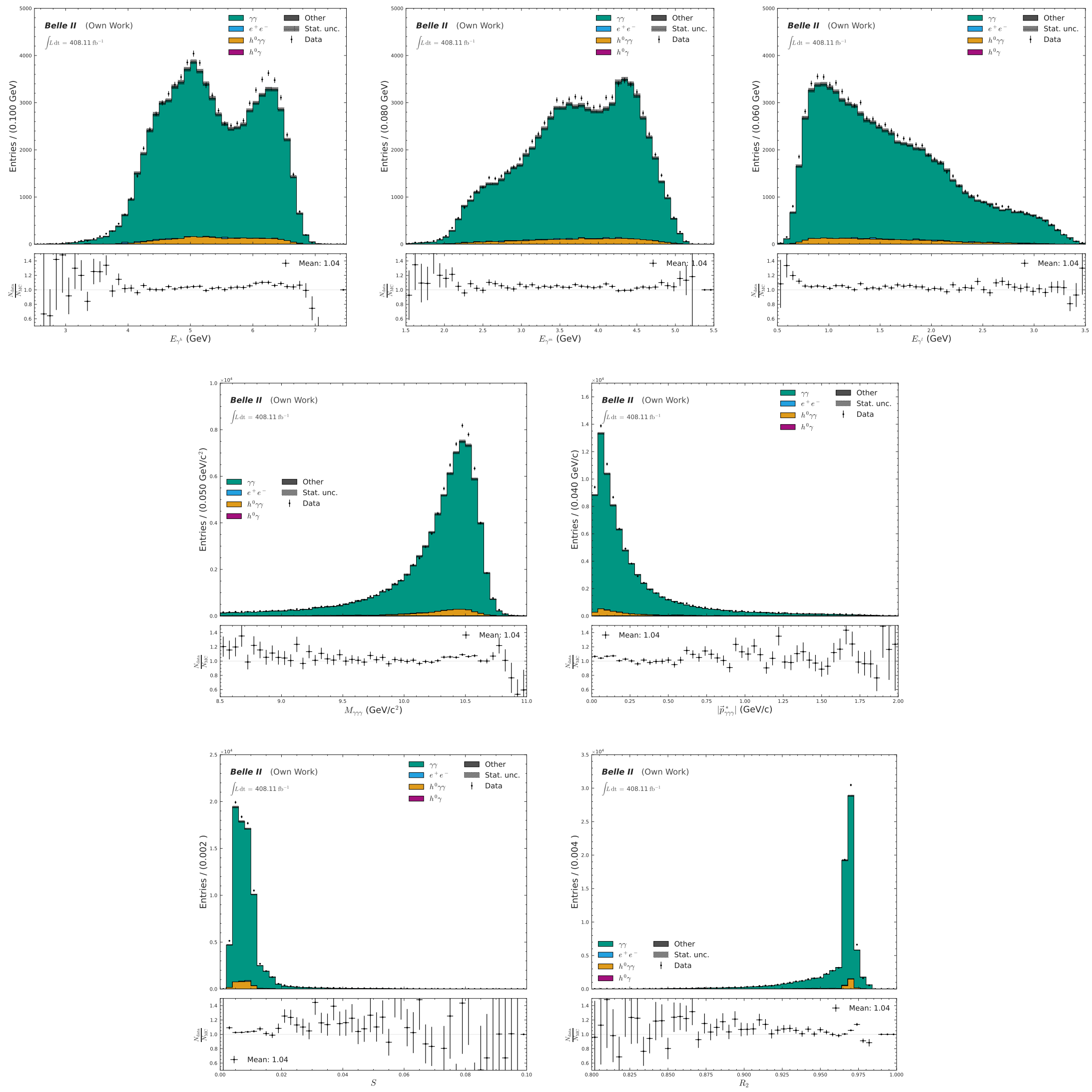


Figure D.3: Distributions of the input features for the first stage of the MVA-based selection. Filled and stacked histograms display the total background contribution for Run 1 simulation in the ω control region. Black errorbar dots display the corresponding data. Same features are shown as in Figure A.1.

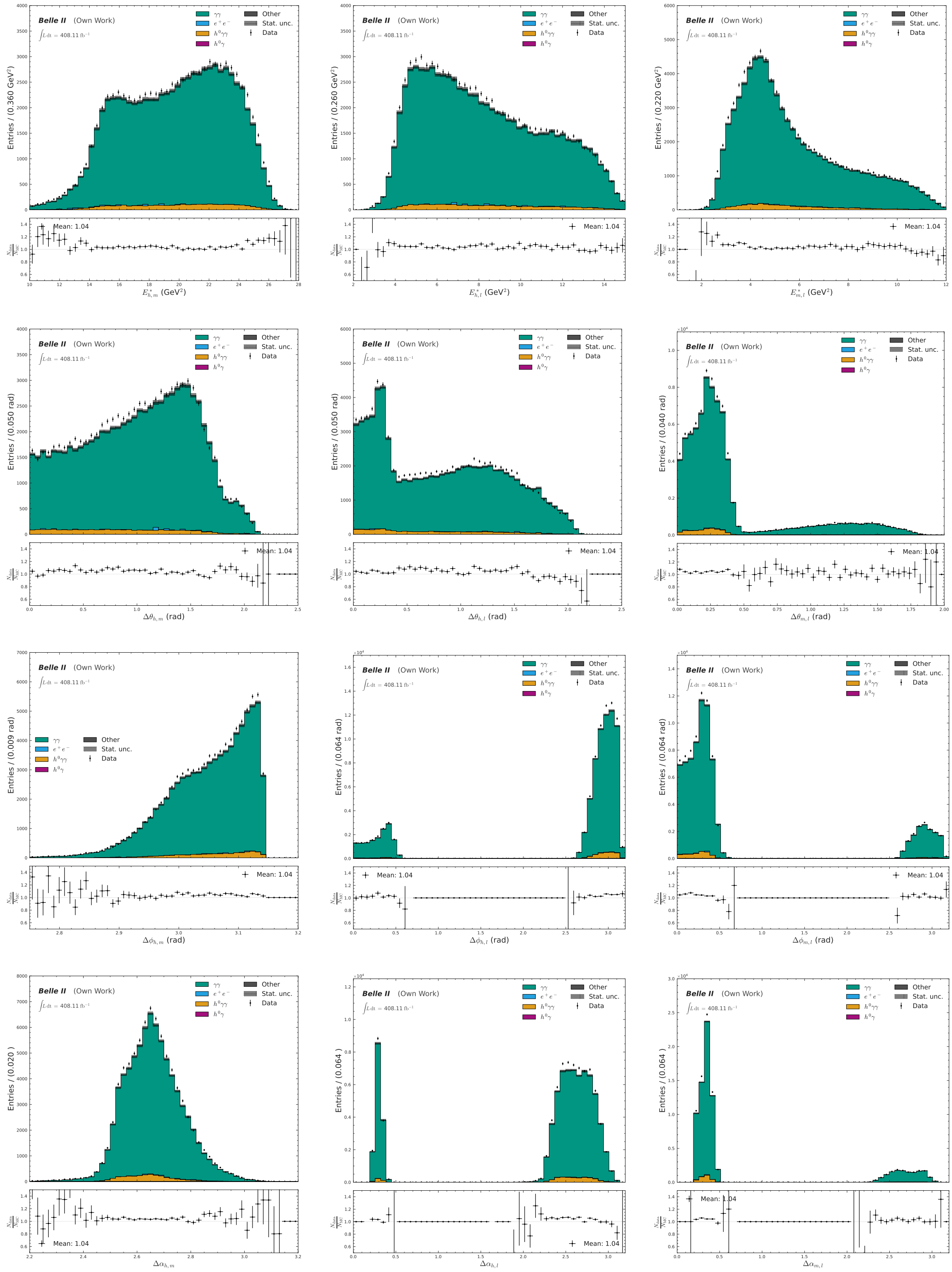


Figure D.4: Distributions of the input features for the first stage of the MVA-based selection. Filled and stacked histograms display the total background contribution for Run 1 simulation in the ω control region. Black errorbar dots display the corresponding data. Same features are shown as in Figure A.2.

Acronyms

ΛCDM Lambda cold dark matter. 1

ALP axion-like particle. 1–3, 5–11, 19–21, 25, 27, 28, 30–33, 35–39, 43–50, 54, 55, 57, 58, 61–63, 66–69, 71, 72, 74–77, 81, 83, 85, 87, 88, 91–98, 102, 112, 113, 118–123, 126–129, 131–134, 136–140, 143–149, 152–157, 159, 160

ARICH Aerogel Ring-Imaging Cherenkov detector. 16, 17

AUC Area Under Curve. 53

basf2 Belle II Analysis Software Framework. 20, 22, 26, 28, 62, 101, 104, 176–178

BCE Binary Cross-Entropy. 52

CB Crystal Ball. 92, 127

CDC Central Drift Chamber. 16, 17, 19, 26, 31–33, 76, 78, 81, 114, 117

CL Confidence Level. 137, 148, 149, 153, 156, 157

CM Centre of Mass. 7, 10, 13, 22, 24, 54, 76, 112, 149, 157, 171, 176

DEPFET Depleted Field Effect Transistor. 16

DM dark matter. 1, 2, 5, 9

DSCB Double Sided Crystal Ball. 92, 117, 127, 128

ECL Electromagnetic Calorimeter. 9, 17–20, 26, 27, 33, 61–63, 75, 76, 78–81, 83, 84, 107, 111, 114, 116, 117, 160, 176

FNN Feedforward Neural Network. 49–52, 55

FPGA field-programmable gate array. 19

FSR Final State Radiation. 36, 101, 115, 177, 178

FTDL Final Trigger Decision Logic. 77, 78

GDL global decision logic. 19

HER High-Energy Ring. 13

HLT High Level Trigger. 19, 75, 84–86, 96, 100, 101, 103, 106–110, 121, 122, 126, 132, 150, 152

IP Interaction Point. 14–18, 30–32, 112, 117

ISR Initial State Radiation. 20, 36, 43, 55, 59–61, 101, 115–117, 121, 126, 129, 130, 177, 178

KEK High Energy Accelerator Research Organization. 13

KLM K^0_L and muon detector. 18, 19, 78

L1 trigger Level 1 trigger. 19, 75, 76, 81, 84, 85, 96, 99–101, 103–107, 109, 121, 122, 126, 132, 150, 152, 159

LER Low-Energy Ring. 13

LHC Large Hadron Collider. 1, 5, 10, 13

MadGraph5 MadGraph5_aMC@NLO. *Glossary:* MadGraph5_aMC@NLO

MC Monte Carlo. 20, 21, 27, 62, 63, 75, 78, 80, 87, 88, 92, 95, 99, 101, 103, 106, 107, 110–114, 116, 121, 123, 126, 127, 129, 132, 133, 139, 171–174

MLP Multilayer Perceptron. 49

MVA multivariate analysis. 21, 25, 43, 49, 50, 58, 59, 64, 66, 73, 121, 159, 171–174, 180–183

NLL Negative Log-Likelihood. 91, 132, 136

NN neural network. 49

NP New Physics. 127

PDF probability density function. 91–96, 117–119, 122, 127, 128, 131–136, 139–142, 144, 145, 149, 153, 155

PFOM Punzi Figure of Merit. 44–46, 57, 58, 67–69, 81, 83, 102

PID particle identification. 16, 17

POCA Point of closest approach. *Glossary:* POCA

PXD Pixel Detector. 15, 16, 19

QCD quantum chromodynamics. 1, 3–5

ROC receiver operating characteristic. 53

ROE rest-of-event. 31

SM Standard Model of particle physics. 1–3, 5, 21, 91, 131, 159, 160

SVD Silicon Vertex Detector. 15, 16

TOP Time-Of-Propagation counter. 16, 17

UL Upper Limit. 4, 10, 11, 131, 132, 137–139, 148, 149, 153, 156, 157

WIMP weakly interactive massive particle. 1