Karlsruhe Institute of Technology

ETP-KA/2022_03

Sensitivity Study in the Search for $B^{\pm} \to K^{\pm} a$ (displaced $a \to \gamma\gamma$) Decays at Belle II

Alexander Heidelbach

Masterthesis

28th April 2022

Institute of Experimental Particle Physics (ETP)

Advisor: Prof. Dr. Torben Ferber
Coadvisor: Prof. Dr. Günter Quast

Editing time: 28th April 2021 – 28th April 2022

Karlsruhe Institute of Technology

ETP-KA/2022_03

Sensitivitätsstudie in der Suche nach $B^{\pm} \to K^{\pm} a$ (versetzt $a \to \gamma\gamma$) Zerfällen bei Belle II

Alexander Heidelbach

Masterarbeit

28. April 2022

Institut für Experimentelle Teilchenphysik (ETP)

Referent:	Prof. Dr. Torben Ferber
Korreferent:	Prof. Dr. Günter Quast

Bearbeitungszeit: 28. April 2021 – 28. April 2022

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 28. April 2022

. . . . . .

(**Alexander Heidelbach**)

Contents

1. Introduction

It has been a long road from „apple“ to „plum pudding“ to „celebrity walking through a crowd“, yet the scientific community is still not satisfied with the outcome. Developed in the late 20th century [1], the Standard Model (SM) is currently the most complete description of elementary particle physics. This theory combines three of the four known elementary forces to describe the fundamental interactions of particles and gives rise to their measured masses. Despite its great success in predicting many observations, the SM leaves the physics community with open questions, theoretical puzzles, and experimental anomalies.

Physicists around the world are developing, studying, and probing extensions of the SM to account for theoretical and experimental discrepancies, such as Dark Matter, baryon asymmetry, the hierarchy problem and the anomalous muon magnetic moment. In search for answers, scientists explore various techniques in two directions. On the one hand, heavy New Physics (NP) could arise at or above the TeV scale. Resulting particles can be probed by direct decays in high-energy colliders like the LHC [2] and future projects like the FCC [3] or indirectly in low-energy experiments like Belle II [4]. On the other hand, NP could present itself in the form of light particles with very small couplings to the SM particles. The impact of this new particles could be directly measurable in the MeV to GeV range [5] at high intensity accelerator-based experiments.

One highly motivated and popular set of models in the search for light NP are Axion Like Particles (ALPs). They arise as pseudo Nambu-Goldstone-bosons of theories with a spontaneously broken Peccei-Quinn symmetry that is anomalous under the SM gauge symmetry. An experimental sector that has yet to be thoroughly probed includes models where the ALP directly couples to the $SU(2)_L$ W gauge boson [6]. This opens up the possibility of studying flavour-changing neutral current transitions such as $b \to s$, which lead to one-loop $B^{\pm} \to K^{\pm}a$ decays. Depending on the model, ALP mass, and coupling, the ALP will decay into two photons, and therefore, give rise to the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ signature with displaced ALP decays.

This thesis performs a sensitivity study in the search for the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ signature based on Belle II Monte Carlo (MC) samples corresponding to an integrated luminosity of $100\,\mathrm{fb}^{-1}$. The analysis includes signal yield scans of the invariant di-photon mass ($M_{\gamma\gamma}$) distribution in the range of $0.175\,\mathrm{GeV}/c^2$ to $4.600\,\mathrm{GeV}/c^2$ for eight different ALP lifetimes between 0.0 cm to 50.0 cm. Finally, the resulting sensitivity is compared to a similar search for $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ on data collected by the BABAR experiment in [7].

Chapter 2 presents an overview of the theoretical foundations by shortly discussing spontaneous breaking of global symmetries and introducing chiral anomalies in the Standard Model. Subsequently, the chapter will discuss Axions as a solution to the strong CP problem and will also introduce ALP models. Chapter 3 outlines the experimental setup, such as the SuperKEKB accelerator and the Belle II experiment, and the analysis framework. Chapter 4 presents the reconstruction acceptance and preselection studies and discusses the optimized candidate selection. Chapter 5 discusses the $M_{\gamma\gamma}$ fit and scan procedure as well as the calculation of the sensitivity. Lastly, Chapter 6 concludes this thesis and gives an outlook on future steps of this analysis.

2. Theoretical Foundations

This chapter outlines the theoretical foundations for understanding NP models involving ALPs. The first discussion in Section 2.1 includes the concept of spontaneous symmetry breaking with a focus on global symmetries in quantum field theories. Subsequently, chiral anomalies are briefly introduced in Section 2.2. Finally, Section 2.3 combines the results of the first two sections in order to reach an understanding of the strong CP problem and shines light on a popular solution, the Axion, and its generalization in the form of ALP models.

2.1. Spontaneous Breaking of Global Symmetries

A fundamental concept in quantum field theories is spontaneous symmetry breaking. It is very accommodating for the description of observed natural concept patterns with effective field theories. In contrast to explicit symmetry breaking, which causes the Lagrangian never to retain the same structure at any given field configuration, spontaneous symmetry breaking leaves the Lagrangian invariant under the symmetry but in the ground state. This section summarizes the key aspects of spontaneous symmetry breaking of global symmetries and one of its substantial consequences for physics: Goldstone bosons. The following is based on the discussions in [8, 9].

Consider a theory that is characterized by a Lagrangian $\mathcal{L}[\phi_n, \partial_\mu \phi_n]$ which is a functional built only of quantum fields ϕ_n and its first derivatives. A continuous global symmetry can be expressed as a variation of the fields in the form of

$$\phi_n \to \phi_n + \alpha f_n(\phi), \tag{2.1}$$

where α is a small parameter and f_n is an arbitrary function of the quantum fields ϕ_n[1]. To leave the action insensitive to these variations, the Lagrangian should also be invariant under the transformation, which leads to the equation

$$\frac{\delta \mathcal{L}}{\delta \alpha} = \sum_n \left(\left[\frac{\partial \mathcal{L}}{\partial \phi_n} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_n)} \right] \frac{\delta \phi_n}{\delta \alpha} + \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_n)} \frac{\delta \phi_n}{\delta \alpha} \right] \right) = 0. \tag{2.2}$$

[1] An often used example is the variation in the form of $\phi \to e^{-i\alpha}\phi$ which can be expanded for small α.

Under the assumption that the fields satisfy the equation of motion, the second square bracket in Eq. (2.2) becomes the defining condition of the Noether current $\partial_\mu J^\mu = 0$, with

$$J^\mu = \sum_n \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_n)} \frac{\delta \phi_n}{\delta \alpha}. \tag{2.3}$$

To highlight the importance of this statement, consider the associated conserved charge

$$Q = \int \mathrm{d}^3 x J_0(x) = \int \mathrm{d}^3 x \sum_n \frac{\partial \mathcal{L}}{\partial \dot{\phi}_n} \frac{\delta \phi_n}{\delta \alpha}, \tag{2.4}$$

in combination with the canonically conjugated fields $\pi_n = \partial \mathcal{L} / \partial \dot{\phi}_n$ and the canonical commutator relations $[\phi_n(\vec{x}), \pi_m(\vec{y})] = i\delta^3(\vec{x} - \vec{y})\delta_{nm}$. Putting everything together, it is apparent that the conserved current generates the symmetry transformation:

$$[Q, \phi_n(\vec{y})] = -i \frac{\delta \phi_n(\vec{y})}{\delta \alpha}. \tag{2.5}$$

As a result, spontaneous symmetry breaking arises in the presence of an unstable symmetric vacuum state with $Q \left|\Omega\right\rangle_{\mathrm{sym}} = 0$ and a charged, stable vacuum with $Q \left|\Omega\right\rangle_{\mathrm{asym}} \neq 0$. Under consideration of the commutation relation between the Hamiltonian and the conserved charge $[H, Q] = 0$ and a vacuum with energy $H \left|\Omega\right\rangle = E_0 \left|\Omega\right\rangle$, it follows directly that

$$HQ \left|\Omega\right\rangle = [H, Q] \left|\Omega\right\rangle + QH \left|\Omega\right\rangle = E_0 Q \left|\Omega\right\rangle, \tag{2.6}$$

where the state $Q \left|\Omega\right\rangle$ degenerates the ground state. Now, it is always possible to find a constant field configuration ϕ_n^0 that minimizes the underlying potential of the Lagrangian

$$\left.\frac{\partial V}{\partial \phi_n}\right|_{\phi_n^0} = 0, \tag{2.7}$$

and corresponds to a new massless particle, the Goldstone boson of the spontaneously broken, continuous global symmetry. To see that the resulting particle is massless, one must consider the eigenvalues of the second coefficient in the Taylor expansion of the potential

$$m_{ij}^2 = \left.\frac{\partial^2 V}{\partial \phi_i \partial \phi_j}\right|_{\phi^0}. \tag{2.8}$$

For the sake of simplicity, let the quantum fields remain constant such that only the potential has to be invariant under the given transformation in Eq. (2.1). This simplifies the condition in Eq. (2.2) to

$$f_n(\phi) \frac{\partial V}{\partial \phi_n} = 0. \tag{2.9}$$

The next step is to take the derivative with respect to ϕ_n and set the field configuration to the one that minimizes the potential:

$$\left.\frac{\partial f_n(\phi)}{\partial \phi_m}\right|_{\phi_n^0} \left.\frac{\partial V}{\partial \phi_n}\right|_{\phi_n^0} + f_n(\phi_n^0) \left.\frac{\partial^2 V}{\partial \phi_m \partial \phi_n}\right|_{\phi_n^0} = 0. \tag{2.10}$$

The left-hand side of Eq. (2.10) vanishes due to the minimum of the potential. Therefore, the right-hand side represents an eigenvalue equation that states that in the non-trivial case ($f_n(\phi_n^0) \neq 0$), all eigenvalues, i. e. all masses, must be zero.

2.2. Chiral Anomalies in the Standard Model

Another essential aspect in the discussion of symmetries in quantum field theories is anomalies. A symmetry of a theory is called anomalous if it is broken by quantum effects, e.g. chiral anomalies, which will be the main focus of this section. Generally, anomalies in the SM can be categorized into three groups. Gauge anomalies describe anomalous symmetries which are connected to gauge bosons. Without a symmetry, the Noether current is not conserved, which leads to problems in theories where this current couples to massless bosons. Consequently, this kind of anomaly should not be present in a unitary quantum field theory. Anomalies in scale invariances are called trace anomalies. An example of this is the scale invariance of massless quantum chromodynamics (QCD), which is broken by loop effects. Lastly, global anomalies generally do not lead to inconsistencies since no particles couple to the unconserved Noether currents. A significant anomaly is baryon number violation, one of the Sakharov conditions, necessary to explain the observed matter-antimatter asymmetry in the universe. The following will explain chiral anomalies using an example and state the most important results for the next section, omitting lengthy calculations based on the discussion in [9].

An excellent example of chiral anomalies can be seen in the Lagrangian of quantum electrodynamics,

$$\begin{aligned} \mathcal{L} &= \bar{\psi}(i\partial\!\!\!/ - eA\!\!\!/ - m)\psi && (2.11)\\ &= \bar{\psi}_L(i\partial\!\!\!/ - eA\!\!\!/)\psi_L + \bar{\psi}_R(i\partial\!\!\!/ - eA\!\!\!/)\psi_R - m\bar{\psi}_L\psi_R - m\bar{\psi}_R\psi_L, && (2.12) \end{aligned}$$

where $\psi_{L/R} = \frac{1}{2}(1 \mp \gamma_5)\psi$ represent the left-/right-handed fields. In the massless limit this Lagrangian is invariant under two global symmetries of the form

$$\psi \to e^{i\alpha}\psi, \quad \psi \to e^{i\beta\gamma_5}\psi. \tag{2.13}$$

The resulting Noether currents are the vector current, which corresponds to the symmetry under which the right- and left-handed fields transform equivalently, and the axial current, which corresponds to the symmetry under which the chiral fields transform with opposite charge:

$$J^{\mu} = \bar{\psi}\gamma^{\mu}\psi, \quad J^{\mu 5} = \bar{\psi}\gamma^{\mu}\gamma^{5}\psi. \tag{2.14}$$

By taking the derivatives of these currents

$$\partial_{\mu} J^{\mu} = 0, \quad \partial_{\mu} J^{\mu 5} = 2im\bar{\psi}\gamma^{5}\psi, \tag{2.15}$$

it is apparent that classically the vector current is conserved. On the other hand, the axial current is only conserved in the massless limit. This observation can be interpreted as additional three-point loop contributions in the presence of a constant background field $F_{\mu\nu}$

$$\langle A| \, \partial_{\mu} J^{\mu 5} \, |A\rangle = -\frac{e^2}{16\pi^2} \varepsilon^{\mu\nu\alpha\beta} F_{\mu\nu} F_{\alpha\beta}. \tag{2.16}$$

Note that Eq. (2.16) is independent of the mass. Thus, in the case $m = 0$, the axial current would not vanish. However, Eq. (2.15) implies that for $m = 0$ the axial current is classically conserved. This contradiction can only be resolved if the chiral symmetry is anomalous.

This type of anomaly can be studied for symmetries involving three currents. For an arbitrary symmetry group $G_a \times G_b \times G_c$, the axial currents will look like

$$\partial^{\mu} J^{a}_{\mu} = \left(\sum_{f_L} A(R_l) - \sum_{f_R} A(R_r) \right) d^{abc} \frac{g^2}{128\pi^2} \varepsilon^{\mu\nu\alpha\beta} F^{b}_{\mu\nu} F^{c}_{\alpha\beta}, \tag{2.17}$$

where $A(R)$ is the anomaly coefficient for a fermion in representation R and d^{abc} is the anomaly constant that summarizes the contributing triangle diagrams. Furthermore, it becomes clear that global symmetries can only contribute physically in connection with two other gauge groups since the $\varepsilon^{\mu\nu\alpha\beta} F_{\mu\nu} F_{\alpha\beta}$ term must be present. An example of this structure could be the combination of a general, anomalous and spontaneously broken $U(1)_X$ and the SM $SU(3)_C$ QCD symmetry group. In a field basis where only the Goldstone boson G transforms under the global symmetry as

$$G \to G + \alpha f_G, \tag{2.18}$$

with α being the $U(1)_X$ transformation parameter and f_G the Goldstone boson decay constant, the divergence of the Noether current predicts terms in the form of

$$\partial^{\mu} J^{U(1)_X}_{\mu} = \frac{\partial \mathcal{L}}{\partial(\partial_{\mu} G)} \underbrace{\frac{\delta(\partial_{\mu} G)}{\delta\alpha}}_{=0} + \frac{\partial \mathcal{L}}{\partial G} \underbrace{\frac{\delta G}{\delta\alpha}}_{=f_G}. \tag{2.19}$$

In comparison with the resulting anomaly term

$$\partial^\mu J_\mu^{U(1)_X} = \underbrace{\left(\sum_{f_L \in 3} X_{f_L} - \sum_{f_R \in 3} X_{f_R}\right)}_{=2N} \frac{g^2}{64\pi^2} \varepsilon^{\mu\nu\alpha\beta} G^A_{\mu\nu} G^A_{\alpha\beta}, \tag{2.20}$$

where G^A are the gauge field strengths, it emerges that if a global symmetry is spontaneously broken and anomalous under some gauge symmetry, the associated Goldstone bosons couple to the gauge field strengths proportional to the anomaly. Hence, Eq. (2.21) can be added to the Lagrangian of the underlying theory, leading to a measurable, physical Goldstone boson.

$$\mathcal{L}_G = N \frac{g^2}{32\pi^2} \frac{G(x)}{f_G} \varepsilon^{\mu\nu\alpha\beta} G^A_{\mu\nu} G^A_{\alpha\beta} \tag{2.21}$$

2.3. Axions and Axion Like Particles

The SM is a widely accepted theory since it predicts and explains most observations in particle physics. Yet, it leaves the community with open questions, theoretical puzzles and experimental anomalies. Scientists build new models and additions to the SM to account for these discrepancies. This section will briefly discuss one of the theoretical puzzles, the strong CP problem, and a possible solution for it, the Axion [10]. The later part of the section will then discuss ALPs, which are very similar to the Axion except for the fact that they do not solve the strong CP problem.

As seen in the previous section, under chiral transformations of fermions in the form of $\psi \to e^{i\theta\gamma^5}\psi$, a CP and P violating term can be added to the Lagrangian,

$$\mathcal{L}_{\text{QCD}} = \theta_{\text{QCD}} \frac{g_s^2}{64\pi^2} \varepsilon^{\mu\nu\alpha\beta} G^A_{\mu\nu} G^A_{\alpha\beta}. \tag{2.22}$$

This term can be shifted under $U(1)_A$ field transformations of the form

$$u_L \to e^{i\alpha} u_L, \quad u_R \to e^{-i\alpha} u_R, \tag{2.23}$$

where u are up-type quark fields, and equivalently for down-type quarks. Eq. (2.23) leads to additional terms to the Lagrangian

$$\Delta\mathcal{L} = 2\alpha \frac{g_s^2}{64\pi^2} \varepsilon^{\mu\nu\alpha\beta} G^A_{\mu\nu} G^A_{\alpha\beta}, \tag{2.24}$$

resulting in a total additional term

$$\mathcal{L}_{\bar{\theta}} = \bar{\theta} \frac{g_s^2}{64\pi^2} \varepsilon^{\mu\nu\alpha\beta} G^A_{\mu\nu} G^A_{\alpha\beta}. \tag{2.25}$$

Here, $\bar{\theta} = \theta_{\mathrm{QCD}} + 2\alpha$ is a basis independent measure of strong CP violation, since chiral rotations only shift a phase between θ_{QCD} and α. Under consideration of the mass terms in the Lagrangian, the α phase can be expressed through the Yukawa mass matrices as $2\alpha = \arg\det M_u M_d$, which makes $\bar{\theta}$ a physical parameter under the assumption $\det M_u M_d \neq 0$. An important consequence of this additional CP violating term is the generation of a neutron electric dipole moment d_N which arises through an additional pion-neutron coupling $\bar{g}_{\pi NN}$ shown in Fig. 2.1. The comparison between the theoretical prediction for the neutron electric dipole moment in [9] with $d_N = (5.2 \times 10^{-16}\,\mathrm{cm})\bar{\theta}$ and the newest measurement in [11], which places the upper limit at $|d_N| < 1.8 \times 10^{-26}\,\mathrm{cm}$ (90% C.L.), reveals the strong CP problem: CP violation in the strong sector with $|\bar{\theta}| < 10^{-10}$ is much smaller than in the electroweak one.

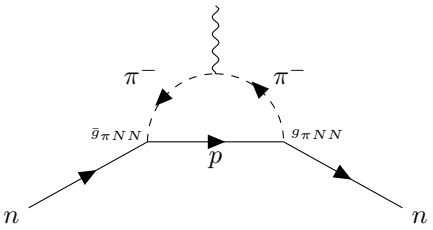


Figure 2.1.: Strong CP violation induces additional pion nucleon interactions in form of the displayed loop with both CP conserving $g_{\pi NN}$ an CP violating coupling $\bar{g}_{\pi NN}$. These interactions lead to a non-vanishing neutron electric dipole moment.

A possible solution to the strong CP problem is to eliminate the CP violating contribution by introducing a dynamical field that cancels the dependence between $\bar{\theta}$ and the neutron electric dipole moment [10, 12, 13]. This can be achieved by introducing an additional $U(1)$ – so-called „Peccei-Quinn“ (PQ) – symmetry that is spontaneously broken and anomalous under the QCD gauge group. According to Eq. (2.21), this would result in a physical field with a particle, the so-called Axion, which is associated to fluctuations around the vacuum expectation value. The Axion shifts under the PQ transformation as $a \to a + \alpha f_{\mathrm{PQ}}$, where α is the symmetry parameter and f_{PQ} corresponds to the scale where the PQ symmetry is broken. The combination of the CP violating term in Eq. (2.25) and the anomaly term from Eq. (2.21) for the $U(1)_{\mathrm{PQ}}$ symmetry results in a total contribution of

$$\mathcal{L} = \left(\bar{\theta} + 2N\frac{a(x)}{f_{\mathrm{PQ}}}\right)\frac{g_s^2}{64\pi^2}\varepsilon^{\mu\nu\alpha\beta}G^A_{\mu\nu}G^A_{\alpha\beta}. \tag{2.26}$$

In the chiral limit, $m_u = m_d$, and $f_a = f_{\mathrm{PQ}}/2N$, one can now find that the vacuum energy is dependent on the Axion and the $\bar{\theta}$ parameter,

$$V_a = F_\pi^2 m_\pi^2 \cos\left(\bar{\theta} - \frac{a(x)}{f_a}\right), \tag{2.27}$$

and is minimized for $\langle a\rangle = f_a\bar{\theta}$, which solves the strong CP problem. Expanding the vacuum energy, it is apparent that even though it is a Goldstone boson the Axion has a finite mass which can be approximated to

$$m_a \approx \frac{m_\pi F_\pi}{2 f_a} = 6.2\,\mu\text{eV}\left(\frac{10^{12}\,\text{GeV}}{f_a}\right). \tag{2.28}$$

It is crucial to notice that the relation between the decay constant and the mass defines the QCD Axion. However, a generalization of the model, in the form of an additional, spontaneously broken and anomalous PQ symmetry, can be made and is motivated by theories beyond the SM [14]. The resulting particles, namely ALPs, in general, do not fulfil this particular relation between mass and decay constant and therefore do not solve the strong CP problem. With an additional, massive pseudo-Goldstone boson with potentially weak couplings to the SM fields, many models can be built and probed. An especially interesting theory [15] arises if the ALP a couples predominantly to the electroweak SM gauge bosons:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu a)(\partial^\mu a) - \frac{1}{2}m_a^2 a^2 - \frac{c_B}{4f_a} a \varepsilon^{\mu\nu\alpha\beta} B_{\mu\nu} B_{\alpha\beta} - \frac{c_W}{4f_a} a \varepsilon^{\mu\nu\alpha\beta} W^i_{\mu\nu} W^i_{\alpha\beta}. \tag{2.29}$$

ALP couplings to SM fermions and gluons are neglected at this point since they are strongly constrained by additional flavour changing neutral current (FCNC) decays [16, 17]. This thesis focuses mainly on the effective model for ALP couplings to SM W bosons as discussed in [6]. For energies $E \ll M_W$, the coupling can be part of FCNC transitions leading to $B \to Ka$ decays as shown in Fig. 2.2. Such decays can be probed at high energy $e^+e^- \to \Upsilon(4S)$ colliders with a possible sensitivity for the ALP mass up to $m_a \leq 5\,\text{GeV}/c^2$. In the case of $m_a < 2M_W$, the ALP will decay dominantly into two photons. Therefore, the $B^\pm \to K^\pm a$, $a \to \gamma\gamma$ signature becomes very attractive to study since the SM background is minimal.

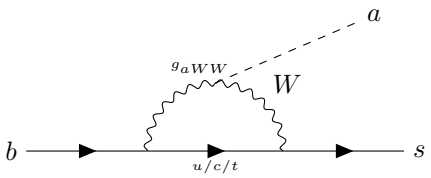


Figure 2.2.: FCNC down-type quark transitions can produce ALPs with a dominant coupling g_{aWW} to SM W bosons in loop processes.

3. Experimental Setup

The third chapter outlines the key aspects of the experimental setup for this analysis. Section 3.1 introduces the asymmetric e^+e^- collider in Tsukuba, Japan, and Section 3.2 mentions the most important components of the Belle II sub-detectors used in this analysis. Subsequently, Section 3.3 gives an overview of the software tools used in this analysis. Lastly, Section 3.4 explains the MC samples and discusses their production.

3.1. The SuperKEKB Accelerator

One possible way to search for NP is using existing colliders. While the LHC experiments use center-of-mass energies at the TeV scale for their probes, KEK, the Japanese High-Energy Accelerator Research Organisation, has operated e^+e^- accelerators at the $\Upsilon(4S)$ resonance since 1998. These so-called B factories provide an excellent environment for high precision measurements of rare decays and CP violation, which is crucial for understanding the slightest deviations from the SM through higher-order processes or new, weakly coupling particles. The current setup in Tsukuba, Japan, is SuperKEKB [18] which is an asymmetric e^+e^- collider, mainly operating at center-of-mass energy of $\sqrt{s} = 10.58\,\text{GeV}$ which is slightly above the $\Upsilon(4S)$ resonance. Fig. 3.1 shows the energy of the two beams that leads to $e^+e^- \to \Upsilon(4S)$ processes, which then decay in about 96% to $B\bar{B}$ pairs [19]. This mode produces B^+B^- pairs with a probability of about 51.4 %. Besides the mentioned $\Upsilon(4S)$ resonance, it is possible to run the experiment in the entire energy range between $\Upsilon(1S)$ and $\Upsilon(6S)$. The SuperKEKB collider targets an instantaneous luminosity of $8 \times 10^{35}\,\text{cm}^{-2}\,\text{s}^{-1}$, which is 40 times the amount of the predecessor, KEKB, and leads to an integrated luminosity target of $50\,\text{ab}^{-1}$. This goal is achieved by squeezing the beam size at the Interaction Point (IP) with additional focusing quadrupole magnets and doubling the beam currents.

3.2. The Belle II Detector

Fig. 3.1 shows the location of the Belle II detector [4] in the Tsukuba experimental hall. Like its predecessor, the Belle detector, the current detector is a general-purpose high-energy spectrometer specifically built to record outgoing data of an asymmetric e^+e^- B factory. To recover as many particle signatures as possible with a solid angle coverage in an asymmetric system, Belle II is built by a series of layers of particle detectors arranged around the IP and accounting for the boost of the Final State Particles (FSP). Fig. 3.2 shows the arrangement of the different sub-detectors, which are shortly introduced in the following.

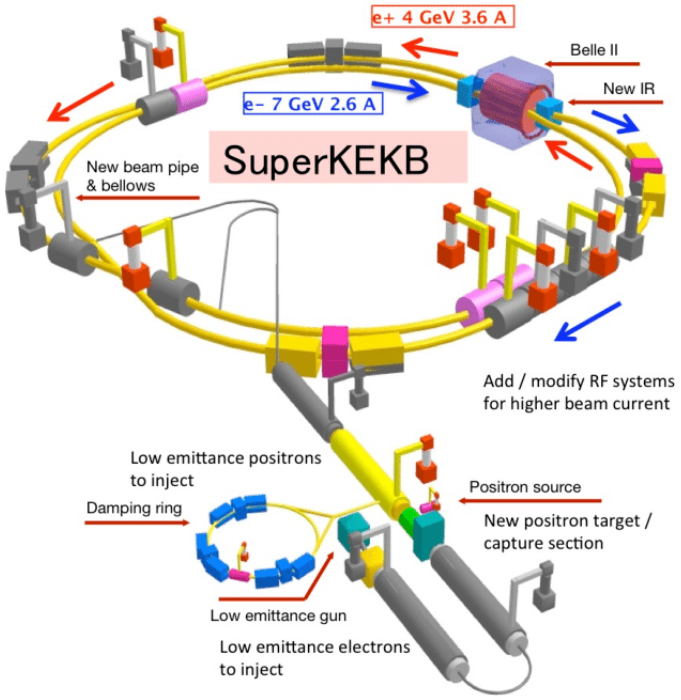


Figure 3.1.: This schematic view of the SuperKEKB collider shows the main accelerator parts. The low emittance gun produces an electron beam which is accelerated in the linac and hits the new positron target. Both beams are led to the storage system and are accelerated in opposite directions, as indicated by the coloured arrows. The figure is adapted from [20].

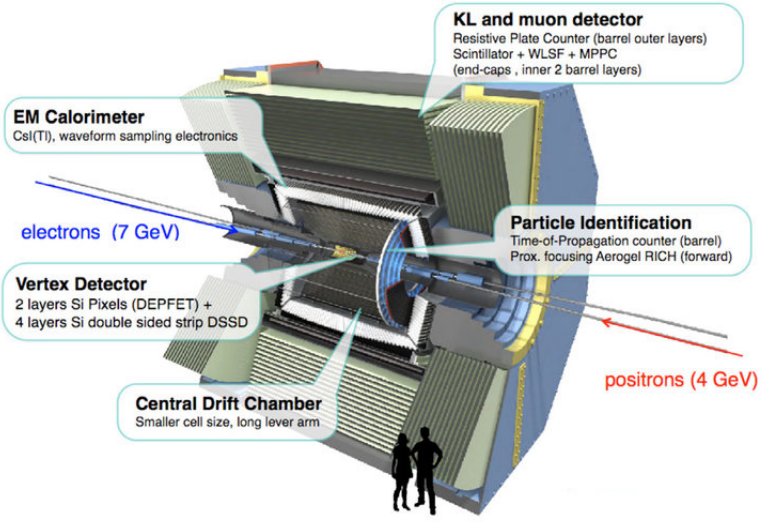


Figure 3.2.: This schematic view of the Belle II detector shows short summaries of the most commonly used detector parts. The figure is adapted from [21].

Tracking System
Belle II's tracking system consists of three different sub-detectors. The innermost tracking is performed by one layer of silicon Pixel Detectors (PXD) and four additional layers of double-sided Silicon Vertex Detectors (SVD). With a total diameter of 348 mm, they can measure the decay vertices of primary particle decays and are therefore crucial for the measurement of CP violation in mixing. The next layer is the Central Drift Chamber (CDC), a gas-filled, cylindrical wire chamber. It consists of 56 differently arranged layers of wires with a smaller distance between each layer closer to the IP. The outermost reach of the CDC is 113 cm and has a polar angle acceptance of 17° to 150°. With these properties, the CDC is ideal for measuring the trajectories and momenta of charged particles while also providing the energy loss measurement for these particles, which is crucial in their Particle Identification (PID).

Particle Identification
An essential group of sub-detectors in B factories are systems for the PID. While the CDC already provides good information with the help of the energy loss, especially for particles with a lower momentum, the main PID information for Belle II is collected in the Time-of-Propagation (TOP) detector and by the Aerogel Ring-imaging Cherenkov (ARICH) system in the forward region. The TOP detector consists of 16 modules arranged in a barrel form around the CDC. Each module is a quartz radiator in which charged particles will emit cones of Cherenkov light. These are then reflected towards the readout electronics at the end surface of the quartz bars. The likelihood for each particle type is then calculated by the (x, y) coordinates of the readout plates and precise timing information. The ARICH detector in the forward end-cap of the Belle II spectrometer uses a similar principle. The key differences are an aerogel radiator and an expansion volume between the radiator and the photon detectors where the Cherenkov photons can form rings. The backward end-cap does not provide PID information besides the CDC.

Electromagnetic Calorimeter (ECL)
The calorimeter of Belle II is the first sub-detector to measure neutral particles' properties. With this goal in mind, its main tasks are efficient photon detection with a precise energy measurement and angular position approximation. For this purpose, CsI(Tl) were chosen for the barrel region while pure CsI crystals are installed in the end caps. CsI(Tl) maintains a high photon stopping power while delivering a substantial light output for a photo-diode readout system. A single crystal has an average front face surface of 6 cm × 6 cm, a length of about 30 cm (16.1 radiation lengths) and is covered by µm-thin Teflon, aluminium and mylar foils. The roughly 3 m long barrel section consists of 6624 crystals in 29 different shapes, which are tilted towards the IP as shown in Fig. 3.3 and have an inner radius of 125 cm. In combination with the 2112 end-cap crystals in 69 shapes, the ECL has an total polar reach of $12.4° < \theta < 155.1°$ excepted for two $\sim 1°$ gaps between the barrel and the end-caps.

K_L and Muon Detector (KLM)
The last sub-detector layer, the KLM system, consists of alternating iron plates and active detector materials in the form of glass-electrode resistive plate chambers. The iron plates serve two purposes. On the one hand, they function as magnetic flux

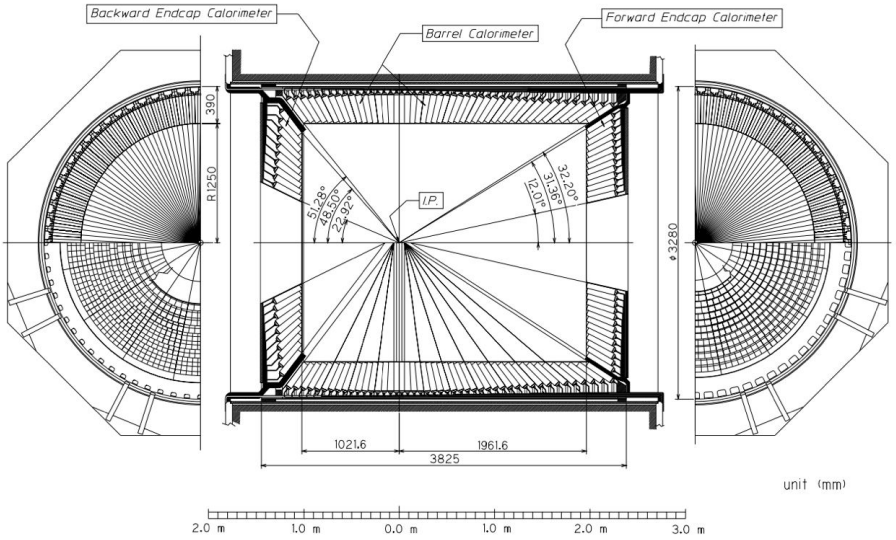


Figure 3.3.: This schematic view of the ECL shows a cross-section of the detector in one ϕ plane and the θ plane of the end-caps. All crystals are tilted towards the asymmetric IP. The figure is adapted from [22].

return for the 1.5 T solenoid, which is positioned after the ECL and on the other hand as shower material for K_L mesons which do not deposit all their energy in the ECL. Other weakly interacting particles, like muons, can be reconstructed by the energy deposition in the active detector materials.

3.3. Analysis Software

Belle II, as a high-energy physics experiment, produces a huge amount of data. To keep process structures as consistent as possible, the collaboration created the Belle II Analysis Software Framework (`basf2`) [23]. This framework creates the possibility of simulating MC events, unpacking raw detector data, and performing reconstructions such as track fitting and clustering. It also provides several tools for an efficient post-reconstruction analysis up to the level of mass-constrained fits and pre-trained MVA weights. `basf2` is accessible through `Python` interfaces via packages built by chains of efficiently programmed `C++` modules. Most commonly, the user creates a steering file that will declare a `basf2` path that stores the configured `basf2` modules in a linear order. Each module in the path is processed upon execution. This file format contains variables for each reconstructed event for further offline analysis.

Most offline analyses include extracting certain parameters by fitting statistical models to distributions of given variables. Over the years, many tools have been developed to accomplish this task. `zfit` [24] is one of the most recent developments in this area. This tool provides a minimalist `Python` interface designed for model manipulation and sampling of distributions. With the use of `TensorFlow` [25], `zfit` opens up the possibility for parallelisation and makes the procedures scalable for large datasets. Due to the specialisation

in manipulating Probability Density Functions (PDF) and the design of `zfit`, it does not include further statistical analysis methods. Therefore, the task of computing statistical limits is performed with the use of `hepstats` [26].

3.4. Analysis of Monte Carlo Samples

It has become common practice in modern physics to first study possible outcomes of an experiment via simulations before looking at measured data. This method forces the analyst to understand the detector thoroughly and not mistake technical artefacts as (new) signals. Especially the search for New Physics at high precision colliders requires a vast amount of care since the effects that make the difference for discovery are very subtle compared to Standard Model processes. The Belle II collaboration makes a big effort to streamline the process of MC production and provides most of the necessary samples for analyses. This analysis uses and produces MC samples corresponding to the so-called *early-phase 3* setup of the detector, which considers the general state of the detector in the run periods between early 2020 and 2022. For this purpose, this [27] `basf2` release is used to be comparable with the run-independent MC14 campaign.

The background of the studied signature, $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$, can be split into two groups. Firstly, irreducible, same final state backgrounds of the rare Standard Model decays $B^{\pm} \to K^{\pm}h$, $h \to \gamma\gamma$, where $h = \pi^0, \eta, \eta'$ are the neutral pseudoscalar mesons[1], contribute in the form of peaking structures at the positions of the respective $M_{\gamma\gamma}$. Secondly, due to this being an untagged analysis, any other combination of $K^{\pm}\gamma\gamma$ as FSP produces a flat background in the studied $M_{\gamma\gamma}$ range. Consequently, the following Belle II MC samples are considered background samples, while the contributions from leptonic decays $e^+e^- \to l^+l^-$ ($l = e, \mu, \tau$) are neglected[2].

Belle II MC Samples $\left(100\,\text{fb}^{-1}\right)$

Generic Samples
These events contain all decays with $b \to u, d, s, c$ transitions. The samples are most commonly split into two categories. So-called *charged* samples contain all $\Upsilon(4S) \to B^+B^-$ decays, while the *mixed* samples consist of the $\Upsilon(4S) \to B^0\bar{B}^0$ events. The decays are produced by `EvtGen` [28] which is controlled by a decay table that is filled with all considered decays, their branching ratios and the decay amplitudes.

Continuum Samples
Even though the SuperKEKB operates at a center-of-mass energy slightly above the $\Upsilon(4S)$ resonance, only about a third of all hadronic events are the desired $B\bar{B}$ decays. Therefore, the dominant background are the so-called *continuum* processes which involve decays in the form of $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$). These $q\bar{q}$ events are generated by `KKMC` [29] and the hadronisation is performed with the help of `PYTHIA8.2` [30].

[1] The contribution of η_c is neglected at this point due the small branching ratio $\mathcal{O}(10^{-7})$ [19] in the $B^{\pm} \to K^{\pm}h$, $h \to \gamma\gamma$ process.

[2] This choice was made after a private discussion with the argument that these final states do not contribute significantly. Before proceeding further with the analysis, the respective MC samples will be added.

Privately Generated Signal Samples

Depending on the ALP mass and the underlying model, the ALP can have a non-vanishing lifetime. This thesis proposes a sensitivity to couplings of long-lived ALP decays up to $c\tau_a \leq 50\,\text{cm}$. For this purpose, the primary decay of the signal signature $B^{\pm} \to K^{\pm}a$ is produced with `EvtGen` and the ALP decay $a \to \gamma\gamma$ is displaced randomly using an exponential decay function according to

$$r = -\ln\big(x_{[0,1]}\big)c\tau_a\gamma\beta, \tag{3.1}$$

where r is the displacement in the detector, $x_{[0,1]}$ an uniformly distributed random number and γ and β the particle boost factors. Beam background corresponding to the expected amount in *early-phase 3* is added. This analysis uses two sets of MC signal samples with parameters according to Table 3.1. The first set was used for general studies and the second set for computing the final results.

Table 3.1.: Overview of the ALP parameter space used in the different sets of glsmc signal samples

Set	ALP mass m_a in GeV/c^2	ALP lifetime in cm	N_{events} per sample
1	0.10, 1.00, 2.00, 3.00, 4.00, 4.50	0.00, 1.00, 10.00, 20.00, 30.00, 40.00, 50.00	5×10^3
2	[0.175, 4.60] in steps of 0.1 excluding η and η'	0.00, 1.00, 5.00, 10.00, 20.00, 30.00, 40.00	5×10^4

4. Candidate Selection

This chapter discusses the experimental attributes of the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ decay and presents the candidate selection for reconstructed events. The first discussions in Section 4.1 regard the general acceptance of the signal signature by studying the detector response to the signal FSP. Section 4.2 outlines the applied preselection on the signal decay products to ensure a good reconstruction quality. Finally, Section 4.3 discuss in detail the candidate selection with the application of the Punzi Figure of Merit (PFOM).

4.1. Reconstruction Acceptance

The first step in investigating signatures of New Physics decays at colliders is to estimate the number of detectable signal events. After the ALP events are simulated by the generators described in Section 3.4, a detector simulation with `Geant4` [31] is performed to mimic the response of the Belle II detector to the FSP. The following discussions distinguish two different levels of particle information. First, the *generator* information indicates the variable values of particles before the detector simulation, e.g. reconstructed particles will show the energy with which they were generated by the event generator and not the measured energy. Second, the *detector* level will express the reconstructed information about particles as the detector would do. In the previous example, the reconstructed energy of a particle will most commonly be lower than the generated energy due to technical limitations of the detector.

A possible metric to describe the acceptance A of events is

$$A = \frac{N^{\text{True}}}{N^{\text{MC}}}, \tag{4.1}$$

where N^{True} is the number of correctly reconstructed events and N^{MC} is the number of generated events. The definition of correctly reconstructed events is ambiguous and will differ in the following discussions. Primarily, it is derived by checking the MC PDG value ($mcPDG$) of a particle or its mother particle or a combination of both. The investigations in this section are based on the smaller set of MC signal samples, as described in Table 3.1, and therefore N^{MC} will always be set to 5000. Section 4.1.1 outlines the different reasons for the acceptance loss in the individual regions of the ALP mass and lifetime, and Section 4.1.2 briefly discusses the kaon acceptance.

4.1.1. Photon Acceptance

The displaced neutral $a \to \gamma\gamma$ decay in the $B^{\pm} \to K^{\pm}a$ event leads to a significant reconstruction acceptance loss in both low and high ALP mass regions for higher lifetimes. Fig. 4.1 shows the total acceptance of the $a \to \gamma\gamma$ decay where the correct *mcPDG* values define the correct reconstruction for the neutral mother particle, i. e. the neutral mother and the mother of both photons must match the ALP. Samples with an ALP lifetime of $c\tau_a = 0.0\,\text{cm}$ show that about 20 % to 30 % of the generated events cannot be reconstructed which is caused by either one or both photons being outside the detecot acceptance. Generally, higher lifetime samples show, dependent on the ALP mass, a decreasing detector acceptance. The dependence of acceptance on lifetime arises mainly from the ALP decays, which boost the photons. This subsection first discusses the substantial acceptance loss for light ALPs. It will then give an overview of the general decay kinematics that lead to lower detector acceptance for all considered mass- and lifetime points and close with remarks on the lifetime dependence in the higher ALP mass region.

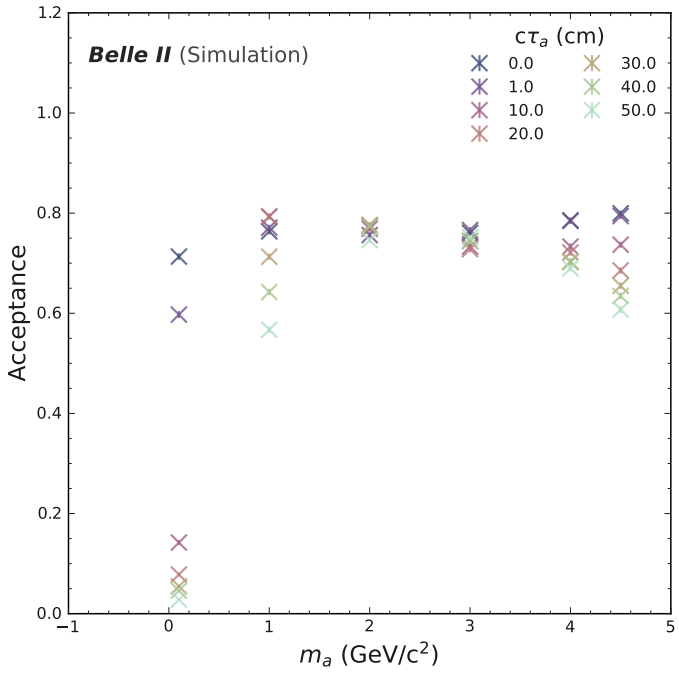
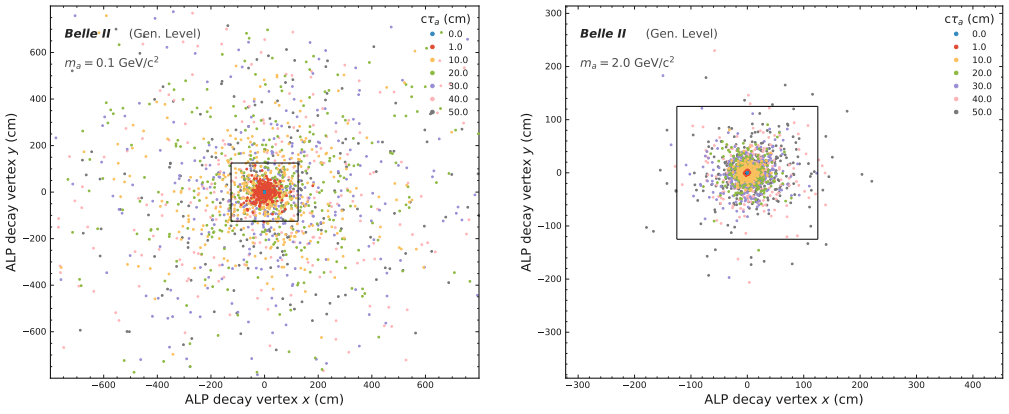


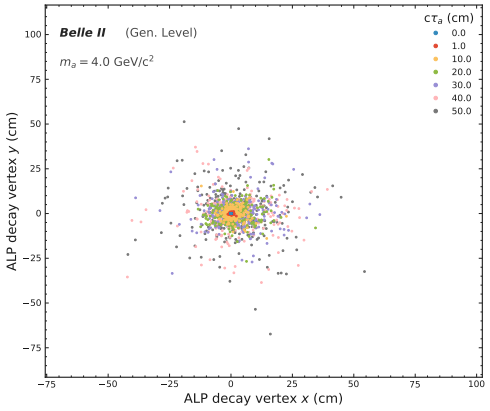
Figure 4.1.: This plot displays the $a \to \gamma\gamma$ acceptance for different ALP masses and lifetimes. Uncertainties on the acceptance calculation are of the order of $\mathcal{O}(10^{-2})$ and therefore are not visible due to the large y-scale. The acceptance shows at least 20 % loss of all generated events after the event reconstruction. Additionally, high dependence on the ALP lifetime can be observed for masses $m_a \leq 2.0\,\text{GeV}/c^2$ and $m_a \geq 4.0\,\text{GeV}/c^2$.

Due to the assumed exponential lifetime of the ALP and the high ALP boost for masses below roughly half the B meson energy, the actual position of the ALP decay can be far greater than the underlying ALP lifetime, as shown by Eq. (3.1). Fig. 4.2 illustrates the x and y coordinates of the ALP decay position extracted at the generator level. Note,

that the the inner radius of the ECL is indicated in Fig. 4.2a and Fig. 4.2b. Thus, it is evident that boosted ALPs can escape the detector before decaying into two photons and therefore cannot be detected. In the scope of this thesis, the inefficiency to these decays is not specifically treated but left to dedicated future analyses which consider invisibly decaying ALPs [32].



(a) $m_a = 0.1\,\mathrm{GeV}/c^2$ (b) $m_a = 2.0\,\mathrm{GeV}/c^2$



(c) $m_a = 4.0\,\mathrm{GeV}/c^2$

Figure 4.2.: These plots show the generated ALP's x and y coordinates for 500 events for three different m_a samples. The black square demonstrates the inner ECL radius in the barrel region if decay vertices lie beyond it. Considering the decay position of the ALP, it is apparent that low mass ALP decay beyond the ECL acceptance.

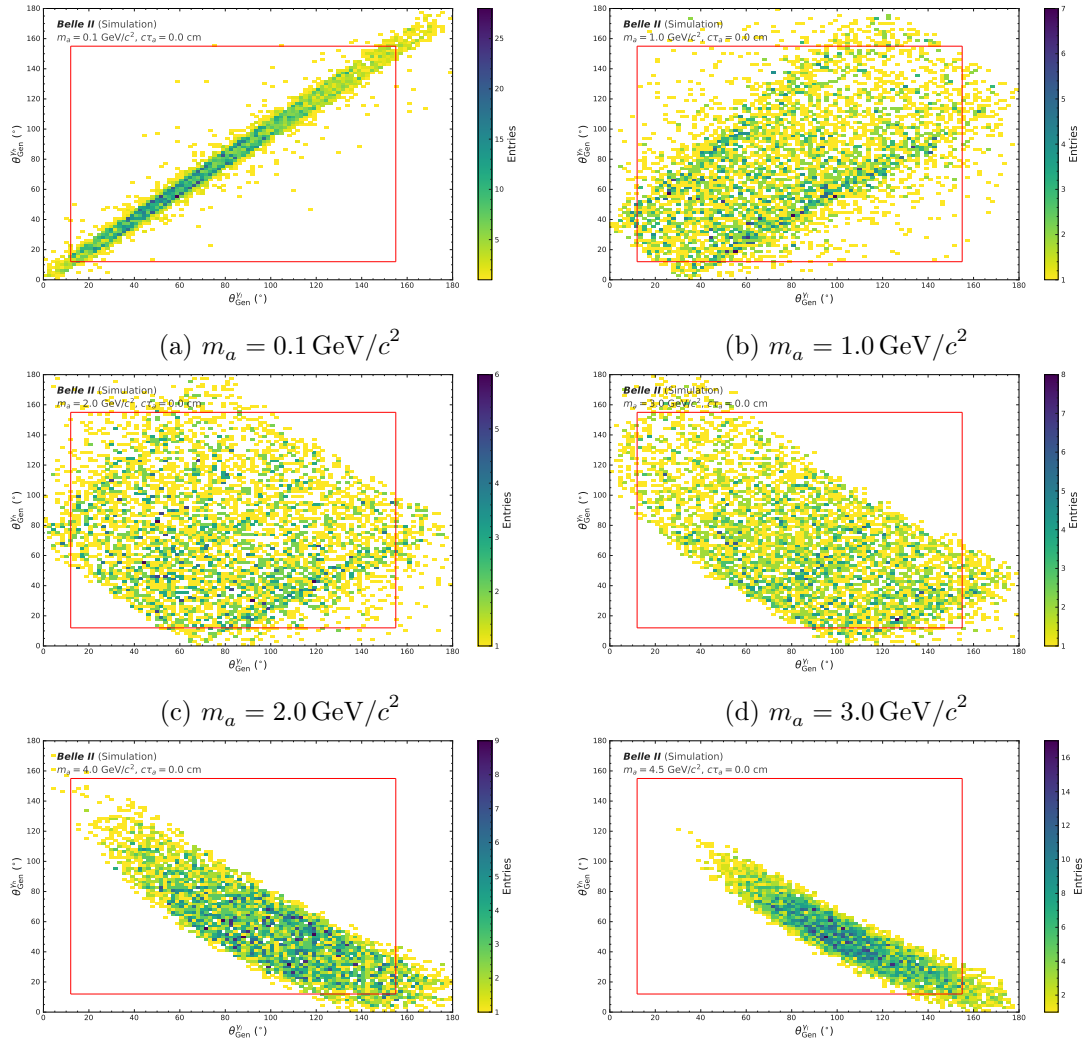
In addition to the ALPs potentially leaving the detector before decaying to measurable FSP, two more sources of acceptance loss occur for all mass and lifetime hypotheses. First, even if the photons are produced within potential ECL acceptance, they can escape by either missing the polar angle acceptance of the ECL or by propagating into the gaps of the detector. Second, either one or both photons are misreconstructed as another particle or as one photon because their opening angle was not large enough to be detected as two isolated photons. Before discussing these cases, it is important to notice the kinematics of the events.

Only the polar angle distribution will be considered to avoid overcomplicating the following discussion since most differences arise here due to the asymmetric beam. Also, both FSP are sorted by their energy and will be denoted as γ_h for the higher energetic and γ_l for the lower energetic photon. Fig. 4.3 displays the generated polar angle θ_{Gen} distributions for both photons for $c\tau_a = 0.0\,\text{cm}$. Differences for the higher lifetimes hypotheses are small at the generation level since the angular distribution is independent of the decay position. The shown distributions follow a typical two-body decay structure. For light ALPs, as in Fig. 4.3a, both photons follow the boost of the ALP and therefore propagate in a very similar direction. Photons from higher mass ALPs, Fig. 4.3f, emerge more back-to-back like in the detector since the ALP is produced approximately at rest in the B meson system, which only carries the non-trivial asymmetric boost from the beam energy difference. The back-to-back like structure can be noticed by the simultaneous low θ_{Gen} values for the γ_h, and high values for γ_l since the asymmetry of the beams boosts the particles in the forward direction. In between the two extreme cases, a superposition of both applies to the polar distribution of the photons.

Coming back to the acceptance loss, it is now apparent, by the polar ECL acceptance indicated in Fig. 4.3, that for each MC signal sample, independent of the mass and lifetime hypothesis, some photons will not be in the ECL acceptance and therefore can not be reconstructed. In the sense of missing photons due to their kinematics, other limitations are the gaps between the barrel region and the end-caps and the ability to distinguish two photons in neighbouring clusters. Fig. 4.4a shows the generation level information for the $m_a = 3.0\,\text{GeV}/c^2$ samples with $c\tau_a = 0.0\,\text{cm}$ of not correctly reconstructed events, i. e. events where the *mcPDG* information could not be matched to the correct particles. At around 31° and 131°, straight lines are visible which correspond to the gaps between different ECL regions. Fig. 4.4b shows the not correctly reconstructed events for a mass hypothesis of $m_a = 0.1\,\text{GeV}/c^2$ where the ALP and therefore the photons are strongly boosted. Some of the displayed events are in the close vicinity of the shown diagonal, which indicates the same θ_{Gen} values for both photons. Most of these are probably not reconstructed as two isolated photons due to their resulting clusters in the ECL being too close in cases where the ϕ distance is also very small. The major energy deposition is in neighbouring or even the same crystal. Regarding Fig. 4.4b, under the right circumstances, all events with a distance of equal or less than two bins, which corresponds to an angular distance of 3°[1], can be affected. This thesis does not correct the misreconstruction of close energy clusters, which is the subject of ongoing investigations.

Before concluding the photon acceptance discussion, one last region of acceptance loss in conjunction with the lifetime of an ALP exists for high ALP masses. While this region is not yet fully understood, some remarks can be made. Under consideration of the previous discussion of not correctly reconstructed events, Fig. 4.5 shows a significant difference in the density of misreconstructed events inside the ECL acceptance for the lifetimes $c\tau_a = 0.0\,\text{cm}$ and $c\tau_a = 50.0\,\text{cm}$ at $m_a = 4.5\,\text{GeV}/c^2$. In this exact case, the number of not correctly reconstructed events $c\tau_a = 50\,\text{cm}$ is about a factor two higher than for the $c\tau_a = 0.0\,\text{cm}$ hypothesis. Almost half of these events can be traced to the case that only the lower energetic photon could not be reconstructed, which means that the *mcPDG* information of

[1] 3° is roughly the mean distance between two crystals, strongly depending on the region and position in the ECL.



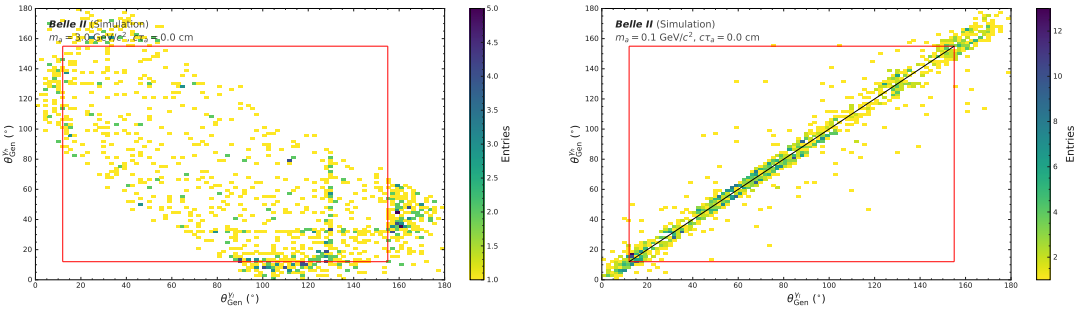
(a) $m_a = 0.1\,\mathrm{GeV}/c^2$ (b) $m_a = 1.0\,\mathrm{GeV}/c^2$

(c) $m_a = 2.0\,\mathrm{GeV}/c^2$ (d) $m_a = 3.0\,\mathrm{GeV}/c^2$

(e) $m_a = 4.0\,\mathrm{GeV}/c^2$ (f) $m_a = 4.5\,\mathrm{GeV}/c^2$

Figure 4.3.: This plot series shows the generated polar angle distributions of the ALP photons, sorted by their energy, for different m_a. The red squares indicate the polar acceptance of the ECL. On the one side, light ALPs are highly boosted, which leads to the same direction emerging of the photons. On the other side, heavy ALPs are only boosted by the asymmetric beam energy, and there the photons show a back-to-back like appearance in the detector.

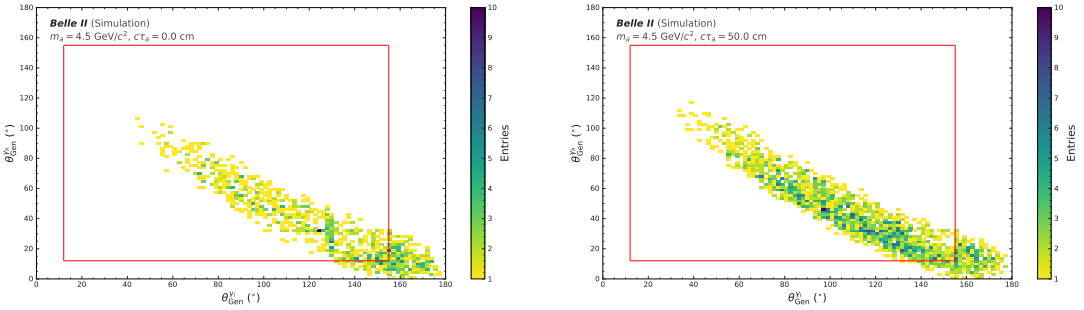
the higher energetic photon and the ALP is correct. Checking the *mcPDG* information of this particle and its mother (*genMotherPDG*) in Fig. 4.6 gives rise to the hypothesis that one of the ALP photons underwent pair-conversion in the detector, and one of its remnants can be detected. This is based on the fact that most of the incorrectly reconstructed FSP are either electrons or positrons based on their *mcPDG* information, and most mother particles are photons that are connected to the ALP. The photons and pions in Fig. 4.6b and the electrons/positrons as mothers in Fig. 4.6d could occur as other or multiple transitions



(a) $m_a = 3.0\,\mathrm{GeV}/c^2$ (b) $m_a = 0.1\,\mathrm{GeV}/c^2$

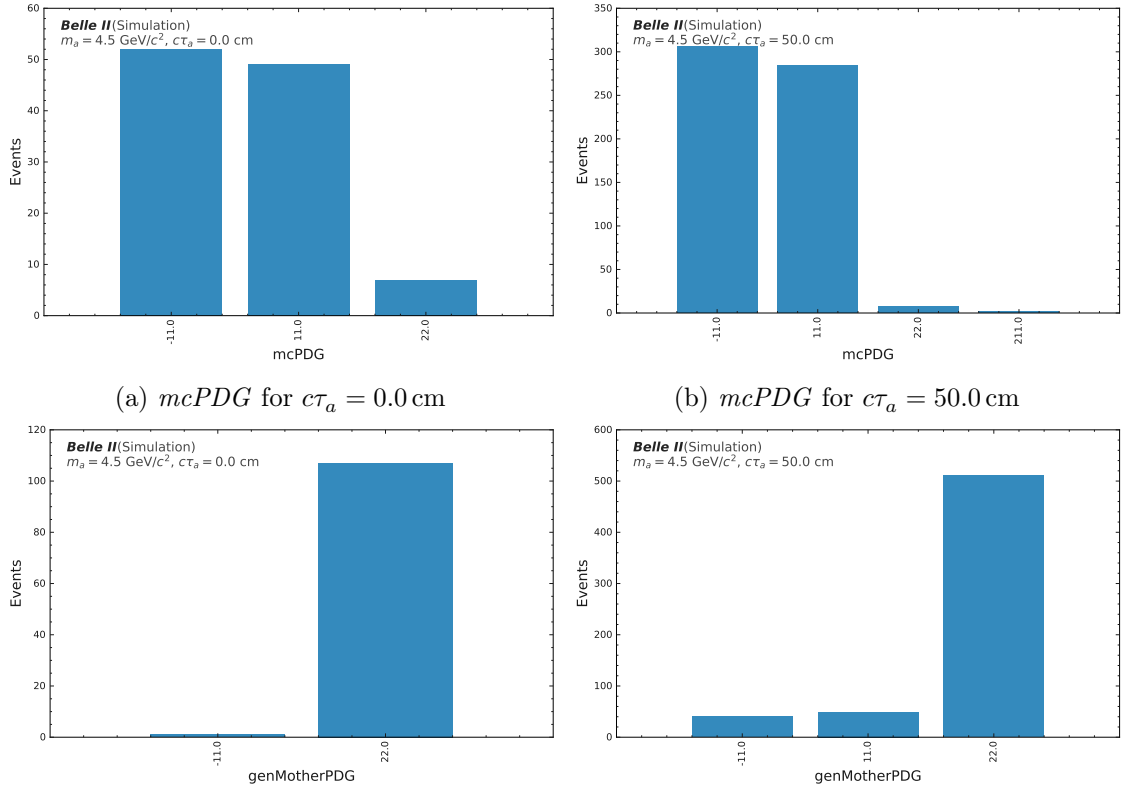
Figure 4.4.: These plots show the generated polar angle distributions of the ALP photons for misreconstructed events. Using the example of $m_a = 3.0\,\mathrm{GeV}/c^2$ and $c\tau_a = 0.0\,\mathrm{cm}$, (a) displays photons that are missed since they are beyond the ECL acceptance and in the gaps between the barrel region and the end-caps. In the case of light ALPs, (b) shows that the majority of not correctly reconstructed photons have very similar polar distribution which, in combination with a similar azimuthal distribution, can lead to indistinguishable clusters in the ECL.

of the initial photon. Other cases where the high energetic photon or both photons are misreconstructed do not occur as often as the missing lower energetic photon. Keep in mind, even if this hypothesis is true, it still would only account for half of the difference between $c\tau_a = 0.0\,\mathrm{cm}$ and $c\tau_a = 50.0\,\mathrm{cm}$. The other half is subject to ongoing investigations.



(a) $c\tau_a = 0.0\,\mathrm{cm}$ (b) $c\tau_a = 50.0\,\mathrm{cm}$

Figure 4.5.: The θ_{Gen} distributions of not correctly reconstructed photons for $m_a = 4.5\,\mathrm{GeV}/c^2$ are shown in this plots. These distributions show about twice as many events in the case of $c\tau_a = 50.0\,\mathrm{cm}$ compared to $c\tau_a = 0.0\,\mathrm{cm}$. Most of these events arise in the ECL acceptance.



(a) *mcPDG* for $c\tau_a = 0.0\,\text{cm}$

(b) *mcPDG* for $c\tau_a = 50.0\,\text{cm}$

(c) *genMotherPDG* for $c\tau_a = 0.0\,\text{cm}$

(d) *genMotherPDG* for $c\tau_a = 50.0\,\text{cm}$

Figure 4.6.: These four plots compare *mcPDG* information of the lower energetic photon in cases where it is not correctly reconstructed. The higher energetic photon and the ALP are correctly reconstructed. The number of events for the FSP own *mcPDG* value and the mother particle, denoted as *genMotherPDG*, are shown for an ALP hypothesis of $m_a = 4.5\,\text{GeV}/c^2$ with $c\tau_a = 0.0\,\text{cm}$ and $c\tau_a = 50.0\,\text{cm}$. Note that in this case, that γ_l does not indicate that the particle is necessary a photon.

4.1.2. Kaon Acceptance

In contrast to long-lived and highly boosted neutral decays to photons, the influence of the kaon in $B \to Ka$ on the acceptance is much smaller and easier to understand. With the initial goal of studying CP violation, the Belle II detector was perfectly designed for kaons. Thus, the acceptance of these particles is maximized. This subsection discusses the significant two aspects responsible for the acceptance loss shown in Fig. 4.7 where correctly reconstructed kaons are those where the *mcPDG* information about its sister particle is correct.

To explain the general acceptance loss for kaons in all ALP mass and lifetime signal MC samples, it is sufficient to consider the kinematics of the kaons. The following discussed properties are independent of the assumed ALP lifetime, and therefore all differences regarding these are random. The kaon's angular distributions do not change significantly

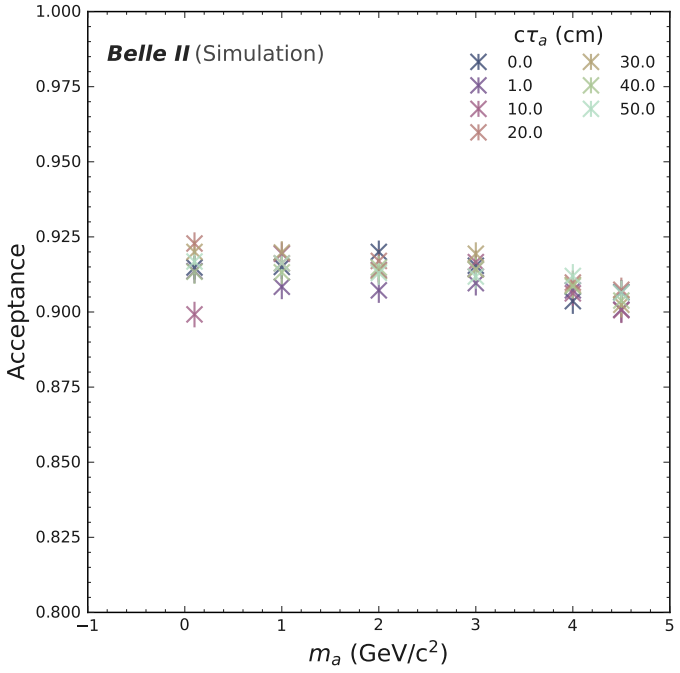
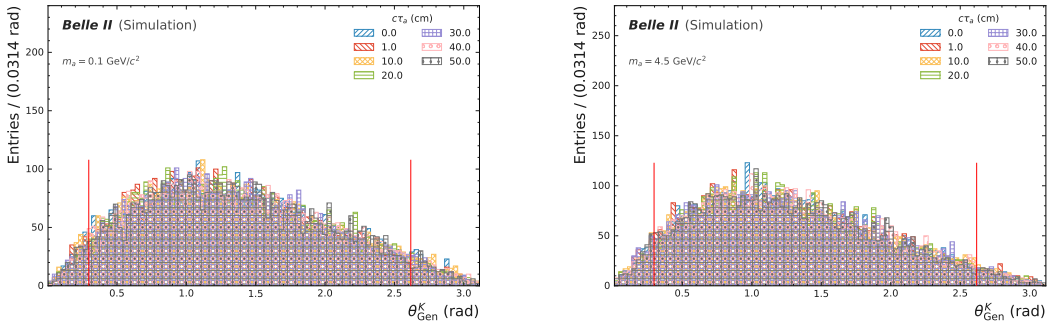


Figure 4.7.: This plots displays the kaon acceptance. It shows a loss in the range of 8 % to 10 % over all signal MC samples. Additionally, a decrease for high ALP masses is observed.

for different ALP masses. The kaon is mostly boosted according to the asymmetric beam energies, which results in a slight forward boost. Therefore, most of the kaons will be inside the CDC acceptance, which is the limiting factor for the general reconstruction efficiency of the kaon. Later in Section 4.2, this thought will be revisited and the PID information considered. Fig. 4.8 shows the polar angular distribution of the kaon for the smallest and highest considered masses and all corresponding lifetimes. Events contributing to the acceptance loss will mostly be outside the indicated CDC acceptance.

One noticeable aspect of the kaon acceptance loss for samples with ALP masses $m_a \geq 4.0\,\mathrm{GeV}/c^2$ in Fig. 4.7. The momentum the kaon can accumulate in the decay of the $B^{\pm}$ meson explains this behaviour. Simplistically, the $B^{\pm}$ provides roughly half of the center-of-mass energy, which is slightly more than 5 GeV. Most of this energy is used to produce a heavy ALP, which leaves some kaons with insufficient transverse momentum to leave the CDC, where they could be identified as kaon. The inability to leave the CDC is caused by the strong magnetic field surrounding the inner detector parts, which causes the charged particle to turn in a circular motion before reaching the outer detector regions.



(a) $m_a = 0.1\,\text{GeV}/c^2$ (b) $m_a = 4.5\,\text{GeV}/c^2$

Figure 4.8.: The generated kaon polar distributions is shown in these plots for MC samples with an ALP mass of $m_a = 0.1\,\text{GeV}/c^2$ and $m_a = 4.5\,\text{GeV}/c^2$. The kaon is mostly forward boosted in the detector, with minor differences between different ALP mass hypotheses. The CDC polar angle aceptance is indicated by red lines. Events outside the CDC acceptance will not be reconstructed due to the lack of track and PID information.

4.2. Preselection of the Final State Particles

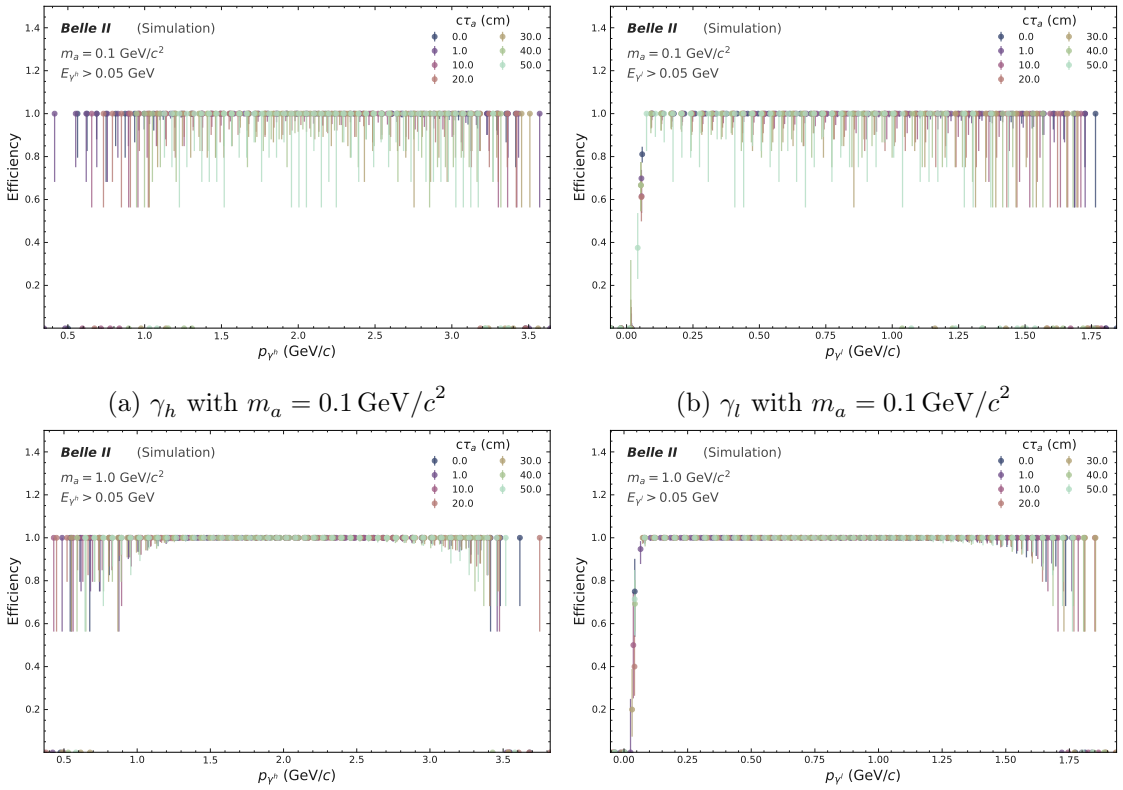
The next step in the analysis preparation is to reduce the number of reconstructed events. Only the number of correctly reconstructed events in the previous subsection is of interest. In reality, the used Mini Data Summary Tables (mDST) files carry a lot more data since every particle combination in the detector that matches the signal FSP signature is reconstructed for each event. Hence, each event has many possible candidates, especially if the FSP consists of photons, which are commonly mimicked by the beam background. To increase the reconstruction purity while maintaining high efficiency, the Belle II community provides recommendations for empiric preselections. This subsection discusses the selections for the FSP based on the same reconstructions for the neutral $a \to \gamma\gamma$ decay and the kaon as in Section 4.1. To study the influence of different selections on the signal events, the efficiency ϵ is used

$$\epsilon = \frac{N^{\text{Pass}}}{N^{\text{Reco}}}, \tag{4.2}$$

where N^{Reco} is the number of correctly reconstructed events and N^{Pass} is the number of correctly reconstructed events that pass the selection. This efficiency is evaluated in bins of the measured momentum for the respective FSP. Therefore, the higher energetic and lower energetic photon are discussed separately. The momentum distributions for these particles are displayed in Appendix C.1.

The first variables of interest consist of the energy of the reconstructed photons. Beam background processes lead to multiple, mostly low energetic energy depositions in ECL crystals produced by secondary interactions with no charged track association. In contrast

to photons from the beam background, the deposited energy of the reconstructed ALP photons is well known and, in most of the signal samples, higher than the empiric threshold of $E_\gamma > 50\,\text{MeV}$ in other analyses. Fig. 4.9a and Fig. 4.9c show that even for small ALP masses the higher energetic photon nearly always has events with an energy $E_{\gamma^h} > 0.5\,\text{GeV}$. Only a negligible amount of events in Fig. 4.9b and Fig. 4.9d do not pass the energy selection for ALPs with $m_a \leq 1.0\,\text{GeV}/c^2$. All other mass and lifetime samples do not show an efficiency loss.



(a) γ_h with $m_a = 0.1\,\text{GeV}/c^2$

(b) γ_l with $m_a = 0.1\,\text{GeV}/c^2$

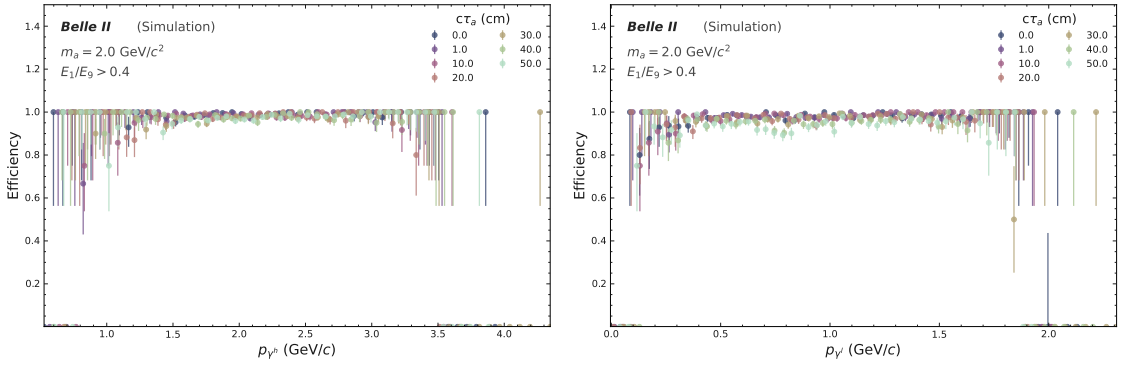
(c) γ_h with $m_a = 1.0\,\text{GeV}/c^2$

(d) γ_l with $m_a = 1.0\,\text{GeV}/c^2$

Figure 4.9.: These plots show the photon efficiency for light ALP MC samples in bins of the photon momenta with the applied selection $E_{\gamma^l} > 0.05\,\text{GeV}$. This selection does not effect the efficiency of the higher energetic photons since they show an minimal energy of $E_{\gamma^h} > 0.5\,\text{GeV}$. However, an efficiency loss is observable in low momentum bins for lower energetic photons.

Another commonly used variable in the preselection of photons is the ratio E_1/E_9, where E_1 is the central crystal energy, the crystal with the highest energy deposition for the reconstructed particle and E_9, the energy of the 3×3 crystals around the central crystal. As previously mentioned, beam background can induce multiple hits in ECL crystals. In the case of a beam background photon, E_1/E_9 would be small since those particles do not hit the face of the crystal but from a direction not originating in the IP and therefore do not carry substantial transverse momentum. The selection of $E_1/E_9 > 0.4$ can exclude such events. However, in the case of displaced ALPs, photons from signal events also do

not necessarily originate from the IP; therefore, this selection will consistently lead to lower efficiency. Fig. 4.10 summarizes the total effect for the $m_a = 2.0\,\mathrm{GeV}/c^2$ hypothesis; other mass hypotheses show similar efficiency losses. The majority of the higher energetic photons shows an efficiency loss up to 5 % in some bins with only small differences for different ALP lifetimes. While the efficiency loss of lower energetic photons from ALPs with smaller lifetimes is comparable to the one for higher energetic photons, events with higher lifetimes show a total loss of up to 10 % in some bins.

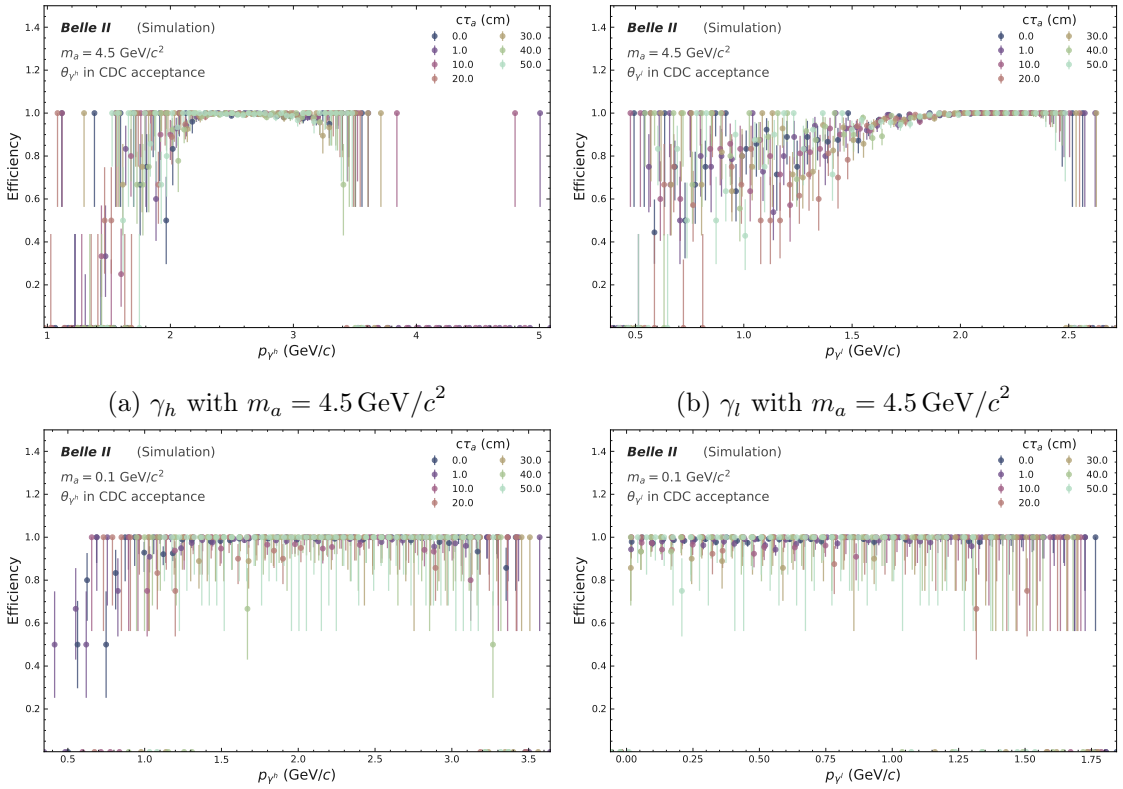


(a) γ_h for $m_a = 2.0\,\mathrm{GeV}/c^2$ (b) γ_l for $m_a = 2.0\,\mathrm{GeV}/c^2$

Figure 4.10.: These plots show the photon efficiency for the $m_a = 2.0\,\mathrm{GeV}/c^2$ ALP hypotheses in bins of the photon momenta with the applied selection $E_1/E_9 > 0.4$. A minor efficiency loss can be observed for both photons in all momentum bins. This loss increases for lower energetic photons in samples with higher ALP lifetimes.

An essential part of the photon reconstruction is that no track can be associated with the energy deposition in the ECL. A selection on the reconstructed polar angle θ_γ can be applied to exclude charged particles from outside the tracking acceptance. More precisely, the photons need to be inside the so-called CDC acceptance, where charged particles can be tracked, which requires the reconstructed polar angle to be $17° < \theta_\gamma < 150°$. Obviously, with Fig. 4.3 in mind, this selection will affect all signal samples and nearly equally for both photons. Fig. 4.11 displays the effect of the selection on the samples with an ALP mass of $m_a = 4.5\,\mathrm{GeV}/c^2$, where the photons are mostly back-to-back scattered in the detector, and $m_a = 0.1\,\mathrm{GeV}/c^2$, where both photons are highly boosted. Other ALP mass samples show comparable results between these two cases.

Fig. 4.12 summarizes the preselection of the photons by displaying the total acceptance under consideration of all selections. In comparison to the acceptance before applying the preselection in Fig. 4.1, only a minor loss can be ascertained. However, the number of reconstructed events is reduced to a third of its original quantity.



(a) γ_h with $m_a = 4.5\,\mathrm{GeV}/c^2$

(b) γ_l with $m_a = 4.5\,\mathrm{GeV}/c^2$

(c) γ_h with $m_a = 0.1\,\mathrm{GeV}/c^2$

(d) γ_l with $m_a = 0.1\,\mathrm{GeV}/c^2$

Figure 4.11.: These four plots display the efficiency loss in bins of the photon momenta for high and low ALP mass samples with an applied CDC acceptance selection. Under consideration of the photon polar distribution, the efficiency shows the expected result for photons that are boosted in regions with no CDC information.

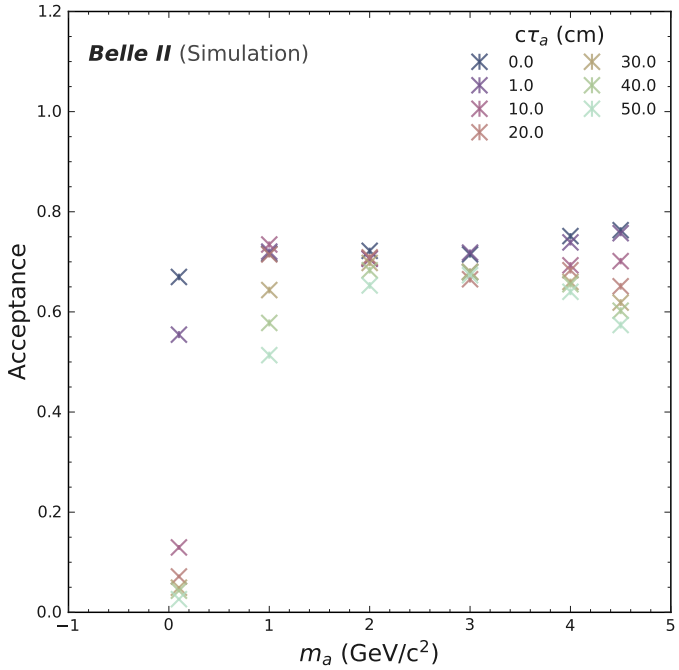
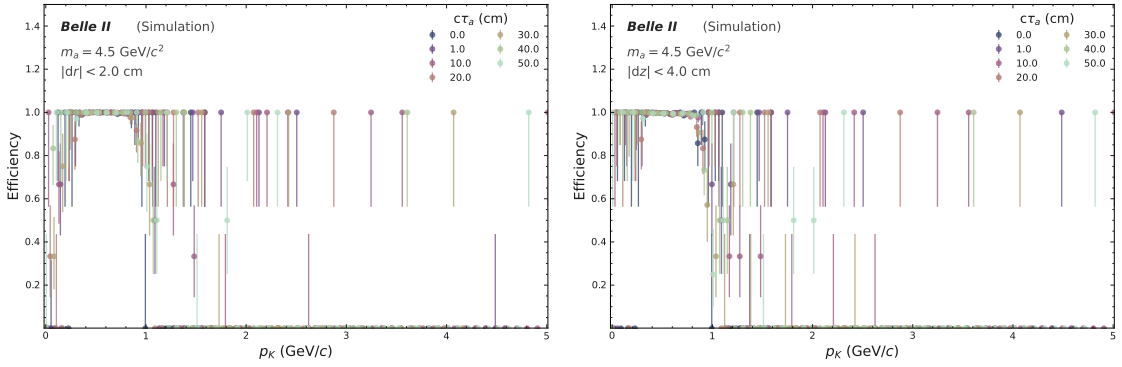


Figure 4.12.: This graphic displays the total $a \to \gamma\gamma$ acceptance with the applied preselection. In comparison to the reconstruction acceptance with no selections, the photon acceptance decreases by about 10 % for all signal MC samples.

Most of the following kaon preselection targets a good track quality. Additionally, this subsection will conclude the definition of PID_K and a small discussion concerning its selection strength. To fully understand the upcoming plots, one needs to visualize again the reconstructed kaon momentum given in Appendix C.1. In contrast to the efficiency binned in the photon momentum, the reconstructed kaon momentum has more outliers, and the plots will look nosier than the previous ones. However, relevant changes can only be observed in momentum bins where the majority of the events are situated in Fig. C.3.

The good track quality selection for kaons can be divided into two groups. On the one side, charged particle tracks are required to originate close to the IP for good vertex fits. This criterion is not of utmost importance for this analysis, where no vertex fit is performed. However, the selection can still be applied since the $B^{\pm} \to K^{\pm} a$ signature generates the kaon at the $B^{\pm}$ decay position. On the other side, the same argument for the polar angle distribution from the previous discussion applies to the kaon. To ensure reasonable PID information and momentum reconstruction, the kaon can be required to be inside the CDC acceptance and have a minimal number of hits in the CDC ($N_{\text{CDC Hits}}$). The distance to the IP is mainly measured by the polar coordinates. Thus, the hadron expert group recommends a transverse distance of $|\mathrm{d}r| < 2\,\text{cm}$ and along the beam direction of $|\mathrm{d}z| < 4\,\text{cm}$. The most significant influence of these selections can be observed for events with high ALP masses, in which case some kaons do not have enough momentum to leave well detectable traces. Fig. 4.13 shows that for $m_a = 4.5\,\text{GeV}/c^2$ some momentum bins display an efficiency loss in the high and low momentum region. These lost events will be most commonly outside the CDC acceptance and, therefore, are not good candidates for the analysis. Other mass points show similar effects in their respective momentum bins.

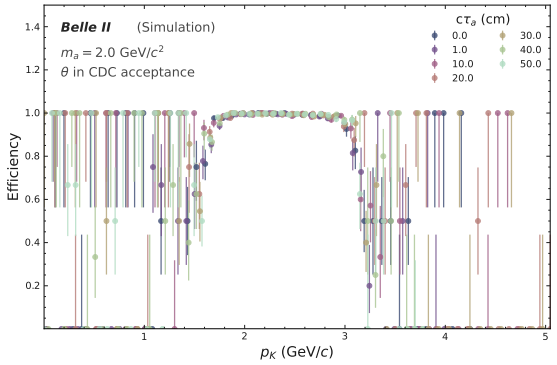


(a) $|\mathrm{d}r|$ selection for $m_a = 4.5\,\text{GeV}/c^2$ (b) $|\mathrm{d}z|$ selection for $m_a = 4.5\,\text{GeV}/c^2$

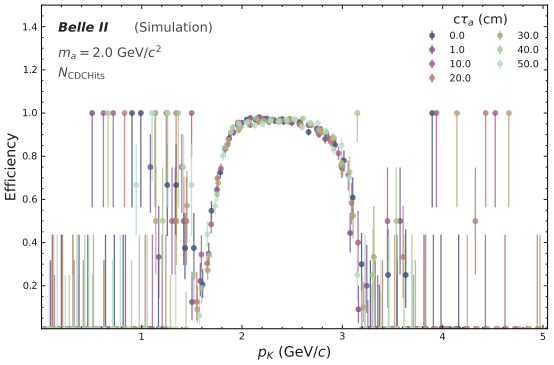
Figure 4.13.: These plots show the kaon efficiency in bins of its momentum under consideration of the IP selections for $|\mathrm{d}r|$ and $|\mathrm{d}z|$. An efficiency loss is observable in weakly populated momentum regions.

Concerning the second part of ensuring a good track quality, standard selections are the criteria that the charged particles are inside the polar angle CDC acceptance and a minimal number of registered wire hits $N_{\text{CDC Hits}} > 20$ can be associated to the particle's track. Fig. 4.14a shows the expected efficiency loss for $m_a = 2.0\,\text{GeV}/c^2$ in high and low kaon momentum bins where the particle propagated either mainly in the forward or backward

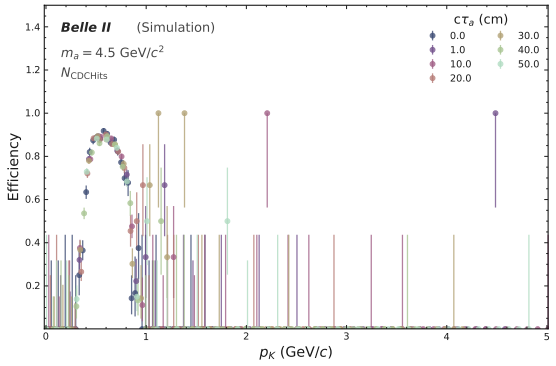
direction and therefore was never inside the CDC acceptance. Simultaneously, Fig. 4.14b displays for the same ALP mass hypothesis the influence of the $N_{\text{CDC Hits}}$ selection where the impact is more visible on the events in the outer momentum regions. Most probably, these kaons propagated through the CDC but left it before leaving the required amount of hits. In the inner momentum regions, a loss of about 5 % per momentum bin can be observed. This loss will be maximal for kaons with a lower momentum, as in Fig. 4.14c, due to the inability to ionise the gas inside the CDC.



(a) θ_K selection for $m_a = 2.0\,\text{GeV}/c^2$



(b) $N_{\text{CDC Hits}}$ selection for $m_a = 2.0\,\text{GeV}/c^2$



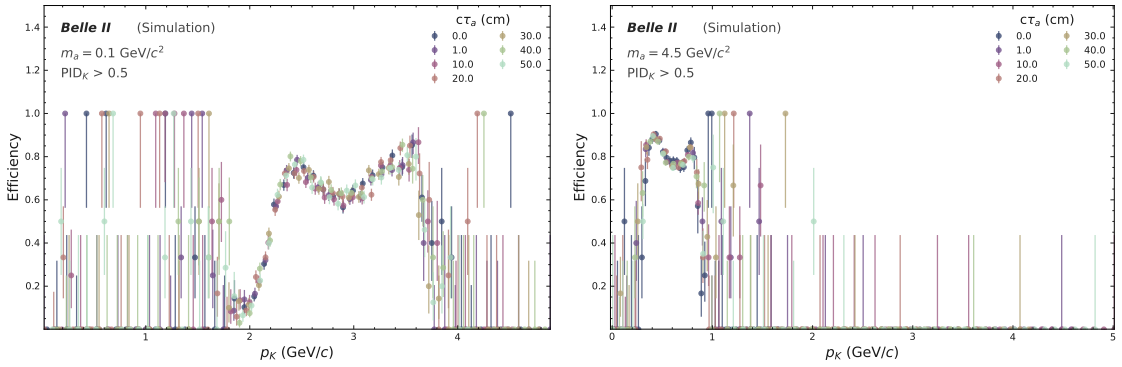
(c) $N_{\text{CDC Hits}}$ selection for $m_a = 4.5\,\text{GeV}/c^2$

Figure 4.14.: This plot collection shows the effect of CDC acceptance and $N_{\text{CDC Hits}}$ selections on the reconstructed kaon efficiency for various ALP mass samples. The CDC acceptance selection results in an efficiency loss for kaons strongly boosted in the forward or backward direction of the detector. This loss is more pronounced for the $N_{\text{CDC Hits}}$ selection since it excludes events with a lowe transversal momentum.

The last commonly used preselection variable is the so-called kaon ID (PID_K). This quantity is a measure of the PID probability of a reconstructed particle. It is calculated via the likelihood ratios of being the desired particle and the sum of likelihoods for all particles. In the case of the kaon, the ratio is

$$\text{PID}_K = \frac{\mathcal{L}_K}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}, \tag{4.3}$$

where for a given track, the likelihood assuming the kaon mass hypothesis is compared to that of electrons, muons, pion, protons and deuterons. Usually, the likelihoods are extracted from the collected information of all sub-detectors. However, at this time, the SVD is excluded from the calculation due to the lack of SVD PDFs and their validation for all particles. Commonly, if an analysis aims for high purity and has high statistics samples, a selection of $\mathrm{PID}_K > 0.5$ is the typical choice. Applying this selection on the reconstructed kaons in this analysis leads to a maximal loss of up to 40 % for central momentum bins in the case of samples with $m_a = 0.1\,\mathrm{GeV}/c^2$ and about 30 % for higher mass, e. g. $m_a = 4.0\,\mathrm{GeV}/c^2$. Fig. 4.15 shows the efficiency loss for low and high ALP mass samples.



(a) $m_a = 0.1\,\mathrm{GeV}/c^2$ – $\mathrm{PID}_K > 0.5$ (b) $m_a = 4.5\,\mathrm{GeV}/c^2$ – $\mathrm{PID}_K > 0.5$

Figure 4.15.: These plots display the efficiency loss in bins of the kaon momentum for the $\mathrm{PID}_K > 0.5$ selection. As explained in detail explained in the corresponding text, the efficiency loss is highly correlated with the detector region the kaon can propagate to and the kaon's momentum.

The efficiency loss per momentum bin creates visible structures. It can be best explained with the help of a low ALP mass sample, e. g. Fig. 4.15a, respectively Fig. 4.16 which shows the median PID_K of all correctly reconstructed kaons in bins of the reconstructed polar angle θ_K and the momentum p_K. Starting from the low momentum kaons, the particles propagate mostly into the end-cap of the detector, where no additional PID information can be collected due to the absence of PID detectors in this region. At this point, most of the PID information is accumulated by the CDC and only a small fraction by the TOP detector. More PID information can be collected with rising momentum since more kaons hit the barrel region. At around $p_K \approx 2.4\,\mathrm{GeV}/c$ nearly all kaons hit the TOP detector and the PID information is satisfyingly good. However, higher momentum causes the PID information to fall until reaching a local minimum before it rises again. The later rise is induced by kaons with a high momentum which are mostly propagating into the forward direction where the ARICH can provide good PID information. This region is clearly visible in Fig. 4.16 for $p_K \gtrsim 3.0\,\mathrm{GeV}/c$ and $\theta_K \lesssim 30°$; excluding the last θ_K bin. In between the two PID peaks is a region where specific combinations of momentum and direction of the kaon lead to Cherenkov cones inside the TOP quartz bars that do not reflect fully inside the bars, which leads to a smaller number of detectable photons and therefore for less PID information provided by the TOP. With this effect in mind, this analysis chooses

a significantly smaller PID_K selection of $\mathrm{PID}_K > 0.002$ to accumulate as many kaons as possible and simultaneously exclude events with negligible PID information, leading to a significant reduction of reconstructed signatures. Fig. 4.17 displays the remaining efficiency loss for the high and low ALP mass samples.

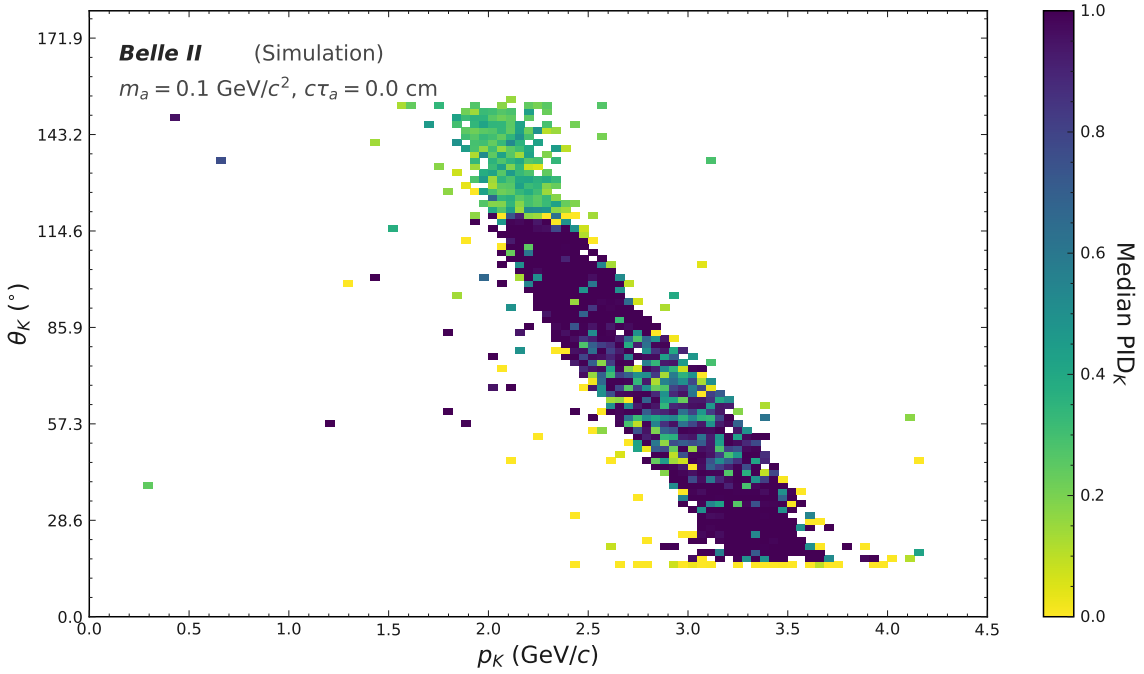
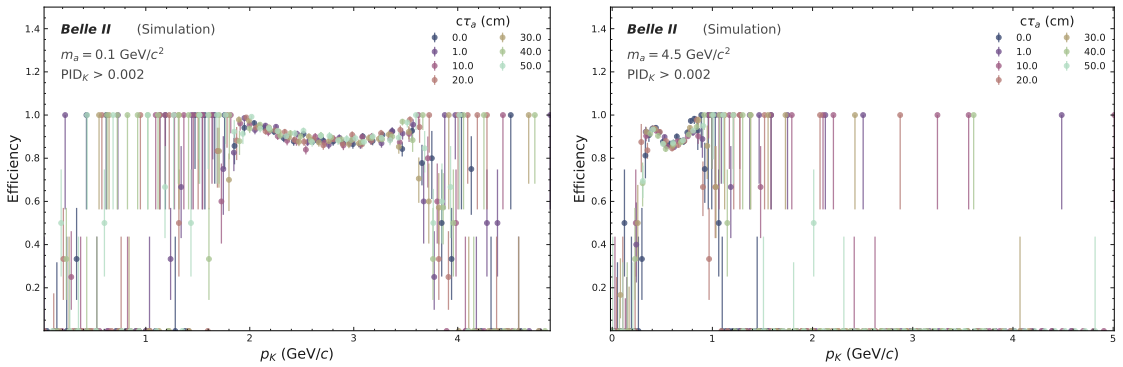


Figure 4.16.: This 2D histogram displays the median PID_K value in dependence of the kaon momentum and polar angle distribution. The PID_K value is significantly lower in the backward direction and for specific steering angle, momentum combinations.

Finally, Fig. 4.18 summarizes the preselection of the kaons, and therefore the FSP, by displaying the total acceptance under consideration of all selections. In comparison to the acceptance before applying the preselection in Fig. 4.7, a more significant acceptance loss can be observed than previously for the photons, especially for the high ALP mass samples. Nevertheless, all applied kaon selections are empirically probed and lead to a good purity and efficiency ratio while reducing the number of reconstructed particles by about 80 %.



(a) $m_a = 0.1\,\mathrm{GeV}/c^2$ – $\mathrm{PID}_K > 0.002$ (b) $m_a = 4.5\,\mathrm{GeV}/c^2$ – $\mathrm{PID}_K > 0.002$

Figure 4.17.: These plots display the efficiency loss in bins of the kaon momentum for the $\mathrm{PID}_K > 0.002$ selection. This selection leads to a significantly smaller efficiency loss but comes with the risk of reconstructing misidentified charged particles.

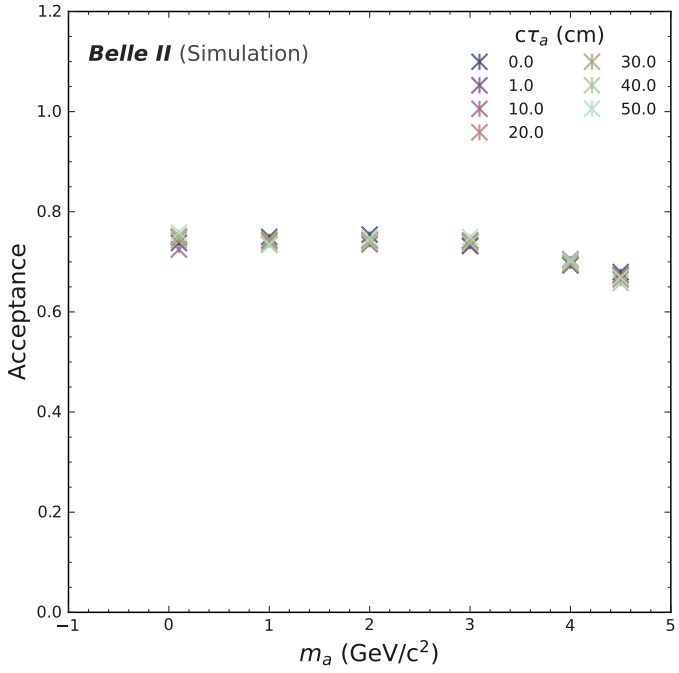


Figure 4.18.: This plot shows the total kaon acceptance with the applied preselection. In comparison to the reconstruction acceptance with no selections, the kaon acceptance decreases by about 15 % to 20 % for all signal MC samples with an increasing loss for higher ALP masses.

4.3. Selection Optimization

In contrast to SM precision analyses, where the candidate selection can be based on a-priori expectations, a search for NP phenomena ideally needs to be open to the parameters of the

new theory, e. g. mass and lifetime. [33] proposes an axiomatic approach in the search of a signal with unknown parameters. The main goal of this approach is to determine sets of selections that have maximal discovery or limit setting potential in counting experiments. This analysis uses the PFOM to optimize selections for every signal MC sample independently and shows smooth transitions of the PFOM in between samples. Later, the signal extraction in Chapter 5 interpolates the selections and performs a bump hunt in $M_{\gamma\gamma}$. Section 4.3.1 outlines the considered variables which will be used for selecting signal events. Subsequently, Section 4.3.2 describes the signal shape in detail and demonstrates a possible way to calculate the signal range. Finally, Section 4.3.3 performs the optimization of selections with the help of the PFOM.

The following introduction of the PFOM is based on the works of [33]. The general idea is based on the assumption that the search for a new signal is a simple counting experiment, where n is the Poisson distributed number of observed events, which is determined by the expectation of the number of background events B and signal events S_m, where m represents the dependency on the signal model parameters. In the case of H_0 being the hypothesis of no new signal that is either accepted or rejected in favour of the hypothesis H_m for a new signal with parameters m, the number of observed events is distributed according to

$$p(n|H_0) = \frac{B^n}{n!}e^{-B}, \tag{4.4}$$

$$p(n|H_m) = \frac{(B+S_M)^n}{n!}e^{-(B+S_m)}. \tag{4.5}$$

It is easy to see that the relevant condition for measuring a non-zero S_m is an observation $n > n_{\min}$, where $n_{\min}$ is determined by B and the significance level α. In the case of no discovery, the limit will be set by the power of the test in dependence on the model parameters m and a in advance chosen Confidence Level (CL)

$$1 - \beta_\alpha(m) \geq CL. \tag{4.6}$$

This inequality can be translated, together with the condition to observe $n > n_{\min}$, in case of present signal, to the condition $S_m \geq S_{\min}$, where $S_{\min}$ is characterized by the significance α, the false-negative probability β and the number of background events B. Furthermore, [33] proposes to parametrize $S_{\min}$ in the form of

$$S_{\min} = a\sqrt{B} + b\sqrt{B + S_{\min}}, \tag{4.7}$$

where a and b are the number of standard deviations corresponding to Gaussian tests at α and β. Together, with the idea that B and S_m both depend on a selection of variables t, it is the general task to find, for a set of parameters m, the selection t that maximizes the sensitivity of the experiment. In other words, after solving Eq. (4.7) for $S_{\min}$, the best t is determined by solving for each m the inequality

$$S_m(t) \geq \frac{b^2}{2} + a\sqrt{B(t)} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}. \tag{4.8}$$

In the simple case, where the selection efficiency $\epsilon(t)$ is independent of the signal parameters m, such that the number of signal events can be rewritten as $S_m(t) = \epsilon(t)L\sigma_m$, where L is the integrated luminosity, and σ_m the cross section of new processes, a minimal detectable cross section will be the defining quantity in the search for an optimal selection t. Optimizing Eq. (4.8) for the maximal sensitivity (minimal detectable cross section) leads to the PFOM which maximum will give the best selection t. With the choice of $a = b$, the PFOM is an elementary quantity:

$$\text{PFOM} = \frac{\epsilon(t)}{\frac{a}{2} + \sqrt{B(t)}}. \tag{4.9}$$

This analysis will maximize Eq. (4.9) for a set of selections on experimental observables for each signal MC sample independently. The general procedure on one signal sample is to maximize the PFOM for one variable, apply this selection and move on to optimising the PFOM of the following variable.

4.3.1. Observables and Final Preselections

Optimizing the PFOM, as it is done in this analysis, is a so-called cut-based approach in the candidate selection. Each variable will have one or two selections that will determine which variable range is sensitive to the signal. Obviously, this approach has a strict limitation which is the correlation of different variables. Furthermore, not every variable will have the same signal selection strength. Thus, the final selection depends on the chosen variables and the order in which their respective optimal selections were determined. Within the scope of this thesis, several different observables were inspected, mainly variables that are in close relation to the ones chosen in [7]. However, the final choice of observables in this analysis is reduced to only five observables, where all of them are strongly physically motivated with a high separation power. The relatively small number of variables, which contribute to the selection, is justified by the strength the selection can have. After applying the first few optimized selections, the number of events can be so low that all following selections will be optimized based on single random events. At this point, the selections become biased towards fluctuations. Additionally, the number of background events is too small such that further analysis steps cannot be performed reliably. This aspect is also discussed in Section 5.2. This subsection will describe the observables used for the PFOM optimization, examine their correlations and close with remarks on final preselections on the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ events.

B meson kinematics

Physics analysis at B-factories, like Belle II, possess the advantage of a clear initial state which is the $B\bar{B}$ pair, respectively the B meson as mother particle to the analyzed decay tree. This well-known prerequisite permits a unique way of selecting signal candidates over background. The B-physics community adapted two observables, the beam-energy constrained mass M_{bc} and the energy difference ΔE, from simple principles. The reconstructed

energy of all FSP in a B decay E_B has to correspond to the initial energy of the beam E_{Beam} which is about half the center-of-mass energy. Consequently, the difference

$$\Delta E = E_B^* - E_{\text{Beam}}^*, \tag{4.10}$$

where the star denotes the center-of-mass frame, peaks around zero for events originating from B decays. 4.19a displays the ΔE distribution for correctly reconstructed $B^\pm \to K^\pm a$, $a \to \gamma\gamma$ events with an ALP mass of $m_a = 4.5\,\text{GeV}/c^2$. Due to the high energy photons, that induce shower leakage in the ECL crystals, the ΔE distribution shows a left tail which is less dominant for lower ALP mass samples. Continuum background events, as shown in Fig. 4.19c for continuum and generic samples[2] corresponding to $100\,\text{fb}^{-1}$, are uniformly distributed with a rising slope to the left side. The second observable is motivated by the invariant mass definition for the B mesons but substitutes the B energy with the energy of the beam

$$M_{\text{bc}} = \sqrt{E_{\text{Beam}}^{*2} - \vec{p}_B^{*2}}, \tag{4.11}$$

where $\vec{p}_B$ is the momentum of the B meson. Fig. 4.19b displays M_{bc} for correctly reconstructed signal events and shows a peak at the position of the B meson mass, which is absent in the case of continuum background in Fig. 4.19d. The distribution of M_{bc} gets wider for signal samples with a higher lifetime due to the worse $\vec{p}_B$ resolution caused by the photons. Section 4.3.2 discusses this in detail.

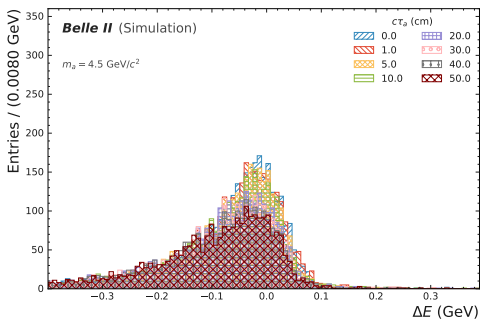
Continuum Suppression

As discussed in Section 3.4 and visible in Fig. 4.19, the so-called continuum background is a major background source in the reconstruction of B decays. A key difference between $e^+e^- \to q\bar{q}$ and $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ events is the overall event topology which allows separating the events. While the $B\bar{B}$ pair is produced nearly at rest, only with a non-trivial boost according to the asymmetric beam-energies, the continuum events form a back-to-back jet-like structure in the center-of-mass frame, as illustrated in Fig. 4.20. This analysis uses two observables to distinguish between the signal $B^\pm$ decay and the continuum background. One of these variables is the normalized second Fox-Wolfram moment R_2. Fox-Wolfram moments H_k, which are defined as

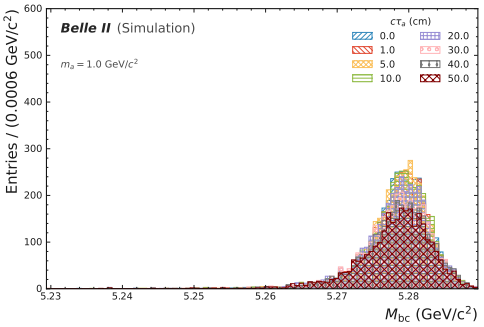
$$H_k = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{E_{\text{Event}}^2} P_k(\cos\theta_{i,j}), \quad R_k = \frac{H_k}{H_0}, \tag{4.12}$$

where $\vec{p}_i$ denotes the momentum of the i-th particle in the event, P_i is the i-th Legendre polynomial and E_{Event} the energy of the event, describe the spatial distribution of particles in an event. Back-to-back, two jet-like structures will take on R_2 values closer to 1 as in Fig. 4.21c. In contrast, spherical events will take on lower values as displayed in Fig. 4.21a for $m_a = 3.0\,\text{GeV}/c^2$ which has a topological evenly distributed event shape. Higher and

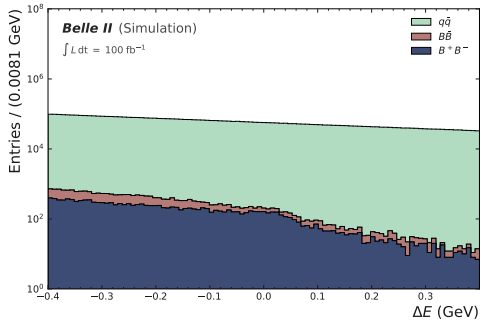
[2] These samples are reconstructed with very loose selections added to the preselection on M_{bc} and ΔE which will be explained later in this subsection.



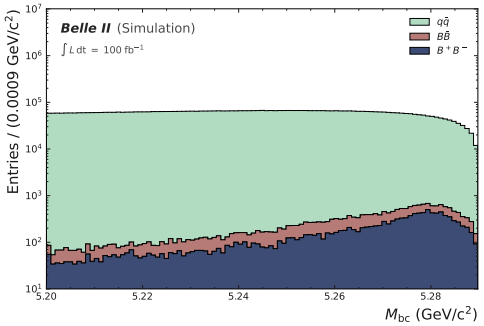
(a) ΔE for $m_a = 4.5\,\mathrm{GeV}/c^2$



(b) M_{bc} for $m_a = 2.0\,\mathrm{GeV}/c^2$



(c) ΔE Background



(d) M_{bc} Background

Figure 4.19.: These four plots display the distributions of M_{bc} and ΔE for reconstructed events of exemplary signal samples (top) and of continuum and generic background (bottom).

lower ALP masses take on higher R_2 values due to the boost of the photons. The other event shape variable in the selection is the angle $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ between the thrust vectors of the reconstructed B meson and the Rest of Event (ROE), which is the name for all remaining tracks and energy depositions in this event. The thrust vector $\vec{T}$ points into the direction for which the sum of the momenta $\vec{p}_i$ projections of all particles in the event is maximized. Events that produced $B\bar{B}$ pairs will have a low momentum in the center-of-mass frame, which results in an even distribution of $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$, as shown in Fig. 4.21b. Continuum events, in Fig. 4.21d, with a strong back-to-back like structure will peak for high $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ values.

Photon Energy

The last selection variable is the energy of the lower energetic photon. For background events, this observable clearly peaks at 0, as displayed by Fig. 4.22c. Signal events will have a broader distribution of this energy at a position dependent on the ALP mass of the sample. In the case of low mass samples, as displayed in Fig. 4.22a for $m_a = 0.1\,\mathrm{GeV}/c^2$, the separation power is low. However, for higher ALP masses, i. e. $m_a \geq 1.0\,\mathrm{GeV}/c^2$, the efficiency loss per E_{γ_l} bin for rising E_{γ_l} is lower than the $\sqrt{B}$ loss per E_{γ_l} bin which results in a good separation power. Fig. 4.22b illustrates the separation potential in the case of

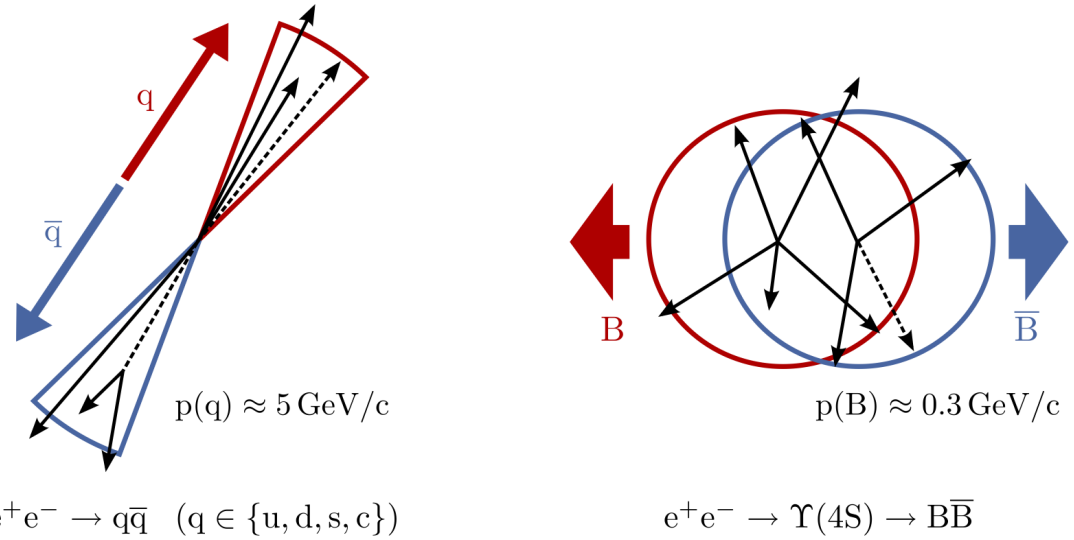


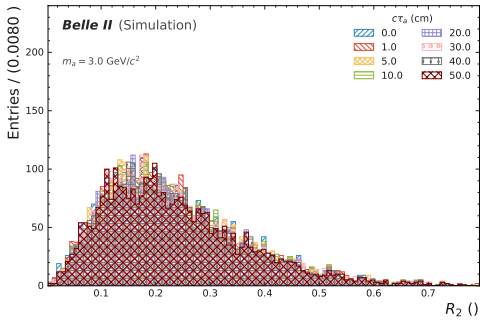
Figure 4.20.: This graphic displays the different event topology for continuum and $B\bar{B}$ events. Figure adapted from [34].

$m_a = 3.0\,\mathrm{GeV}/c^2$ which is easier to see. Section 4.3.3 discusses this topic in detail. The choice of E_{γ^l} over E_{γ^h} is due to discussed separation power. With the preselection and the following small M_{bc} and ΔE selection, for both background and signal, higher energetic photon have an energy of at least $E_\gamma > 1.0\,\mathrm{GeV}$ and very similar distributions which causes the PFOM to be monotonically decreasing.

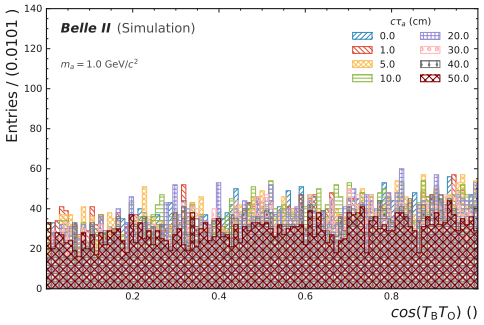
Correlations

As explained in the introduction to this section, the general workflow of the PFOM optimization on one MC sample is to determine the maximal PFOM for a variable selection, apply this selection to the sample and then proceed with the following variable. This linear procedure is susceptible to the correlations of the chosen variables and, therefore, on the order of their selection optimization. Practically, by approaching minimal background, only the values of the selections for two correlated variables will change in the case of interchanging their order in the optimization process. After applying both selections, the resulting signal efficiency or the number of background events should be equal. A profound understanding of the correlations is necessary to understand this outcome fully.

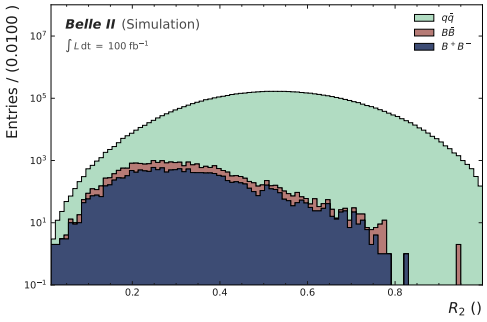
The observables M_{bc} and ΔE are naturally correlated due to common inputs such as the energy/momentum of the photons in the final state. Especially when effects such as shower leakage or spatial reconstruction of the photons dominate the distributions of M_{bc} and ΔE, a clear selection becomes more challenging. In general, events with low energetic photons and no displacement, as shown for correctly reconstructed events in Fig. 4.23a, will lead to an only moderate correlation, which in turn gets more apparent for high ALP mass samples with larger displacement, as in Fig. 4.23b. Since most of the events in Fig. 4.23 are distributed in close vicinity, a tight selection for both variables, which is preferred by the maximal PFOM, will ideally not influence the final signal efficiency and the number of background events.



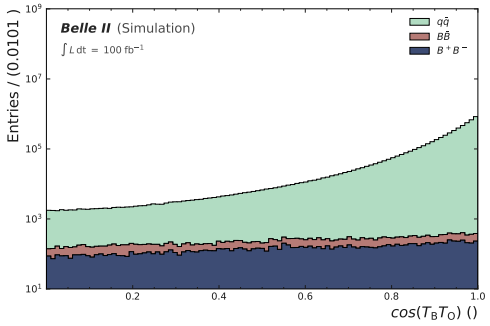
(a) R_2 for $m_a = 3.0\,\mathrm{GeV}/c^2$



(b) $\cos(T_\mathrm{B}T_\mathrm{O})$ for $m_a = 1.0\,\mathrm{GeV}/c^2$



(c) R_2 Background



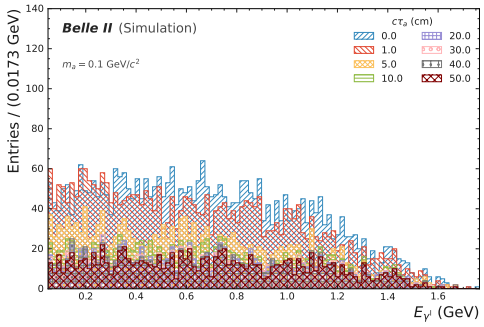
(d) $\cos(T_\mathrm{B}T_\mathrm{O})$ Background

Figure 4.21.: These four plots display the distributions of R_2 and $\cos(T_\mathrm{B}T_\mathrm{O})$ for reconstructed events of exemplary signal samples and of continuum and generic background.

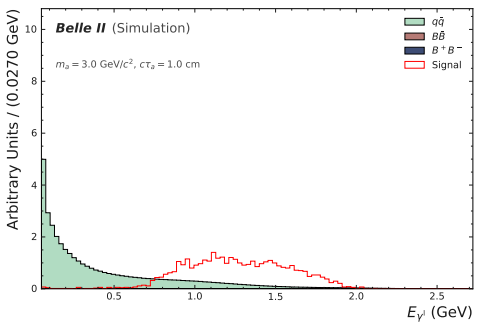
Two highly correlated observables are the ones to suppress the continuum background. Both R_2 and $\cos(T_\mathrm{B}T_\mathrm{O})$ variables express topological quantities of an event that connects them naturally. As previously explained, high and low ALP mass samples will share common values for both variables and therefore show similar correlation, as shown for the low mass case in Fig. 4.24a. For samples with masses in between the outer cases, where the events are mostly spherical, the correlation shown in Fig. 4.24b is slightly but not significantly different. In both cases, a selection on either observable will heavily influence the optimal selection value for the following variable. However, since the selections will affect the same events, the result of the optimized selection is ideally equal.

All other combinations of the here shown observables show insignificant correlations. The energy of the lower energetic photon is slightly correlated with M_bc, ΔE and R_2 since it is the input parameter for all three variables. The most noticeable accidental correlation exists for very spherical events, e. g. $m_a = 3.0\,\mathrm{GeV}/c^2$ in Fig. 4.25, for high values of the energy for the lower energetic photon. In this case, all three FSP have nearly identical energies and, therefore, will spread evenly in the detector, which leads to a low R_2 value.

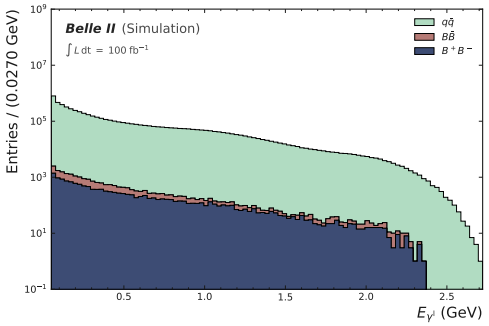
Final Preselection



(a) E_{γ^l} for $m_a = 3.0\,\mathrm{GeV}/c^2$

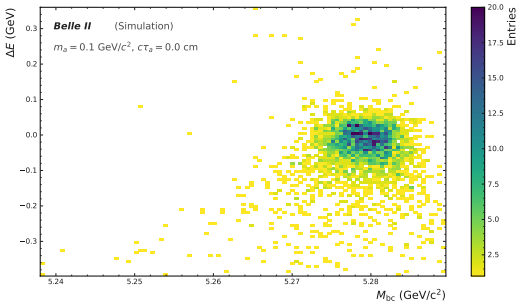


(b) E_{γ^l} for $m_a = 1.0\,\mathrm{GeV}/c^2$

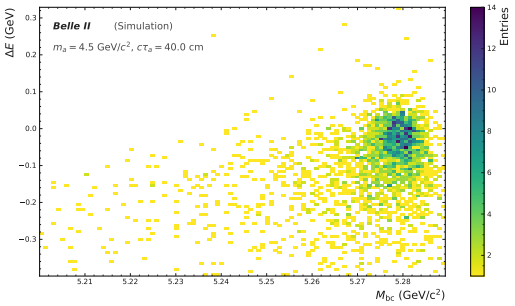


(c) E_{γ^l} Background

Figure 4.22.: These three plots display the distributions of E_{γ^l} for reconstructed events of exemplary signal samples, of continuum and generic background and the combination of both.

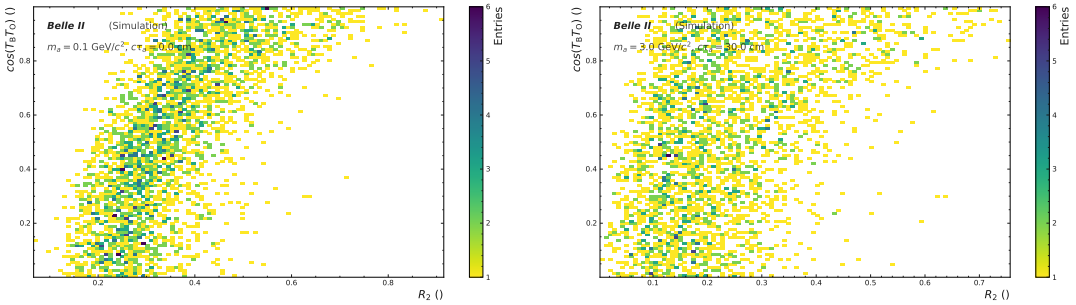


(a) $m_a = 0.1\,\mathrm{GeV}/c^2$, $c\tau_a = 0.0\,\mathrm{cm}$



(b) $m_a = 4.5\,\mathrm{GeV}/c^2$, $c\tau_a = 40.0\,\mathrm{cm}$

Figure 4.23.: These 2D histograms demonstrate the distributions of ΔE over M_{bc} for correctly reconstructed events in two signal samples. The plot in (a) displays a light and promptly decaying and (b) a heavy and long-lived ALP sample. The correlations of both variables increases with higher ALP mass and lifetime.



(a) $m_a = 0.1\,\mathrm{GeV}/c^2$, $c\tau_a = 0.0\,\mathrm{cm}$ (b) $m_a = 3.0\,\mathrm{GeV}/c^2$, $c\tau_a = 30.0\,\mathrm{cm}$

Figure 4.24.: These 2D histograms demonstrate the distributions of $\cos(T_\mathrm{B}T_\mathrm{O})$ over R_2 for correctly reconstructed events in two signal samples. The plot in (a) displays an ALP sample with a less spherical event topology while (b) presents a more spherical case. In both cases, the observables are highly correlated.

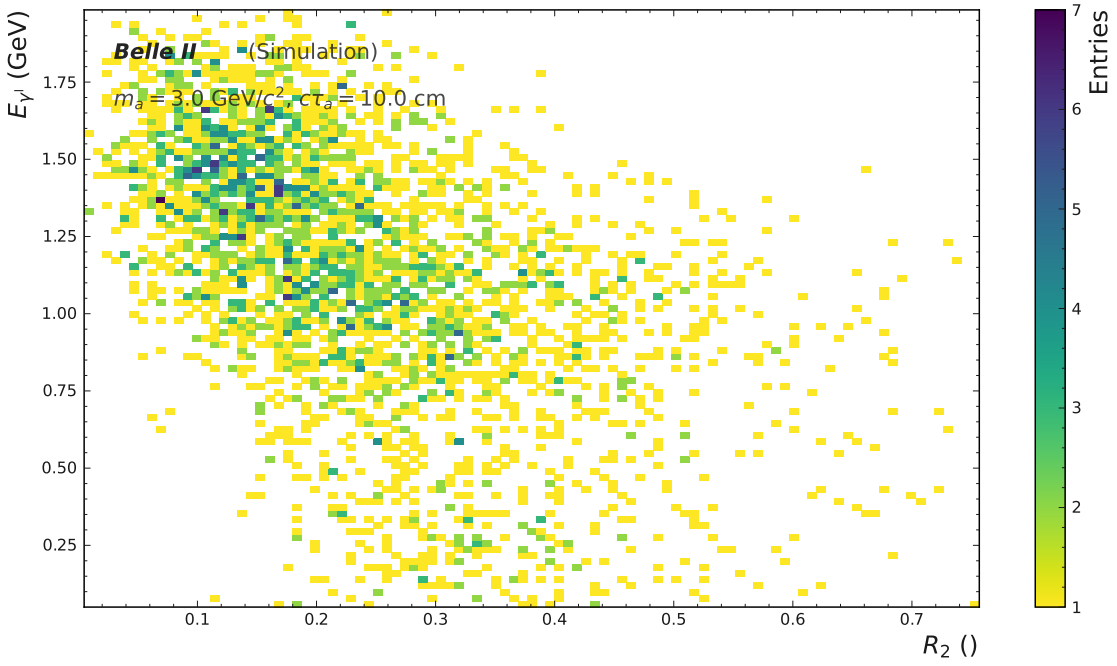


Figure 4.25.: This 2D histogram shows the distributions of E_{γ^l} over R_2 for correctly reconstructed events in the $m_a = 3.0\,\mathrm{GeV}/c^2$ and $c\tau_a = 10.0\,\mathrm{cm}$ sample. The displayed correlation is caused by the nearly perfectly spherical event topology in cases where both photons and the kaon share equal energies.

All the reconstructed events in this subsection and the following chapter have additional preselections common for B-decay analysis. Most of the following selections are empirically approved by the community and are subject to minimal examples in `basf2`. The first selections targets the obvious background events which did not originate from correctly

reconstructed B-decays: $M_{\mathrm{bc}} > 5.2\,\mathrm{GeV}$ and $|\Delta E| < 0.4$. These selections will exclude mostly continuum and misidentified, combinatorial background. The second set of selections targets the previously mentioned ROE. In the case of this analysis, the ROE contains the other B meson in the $\Upsilon(4S)$ decay and additional tracks and clusters from beam-induced or other background processes. So-called ROE masks, which are a group of selections, exclude events in the ROE that do not pass the selection. The here used ROE mask contains the selections $N_{\mathrm{CDC\ Hits}} > 0$, $p^* \leq 3.2\,\mathrm{GeV}/c$ and $p > 0.05\,\mathrm{GeV}/c$. By applying the preselection from Section 4.2 and the M_{bc} and ΔE selections, the used ROE mask does not have a significant influence on reconstructed events. However, the mask is necessary to calculate the continuum suppression observables, which use the ROE and are therefore also applied to the reconstructed events. Fig. 4.26 displays the total acceptance of the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ signal signature. The total acceptance loss is a non-trivial combination of the loss by the preselection on the photons in Fig. 4.12, which dominates the loss for different lifetime samples, the kaons in Fig. 4.18, which reduces the acceptance, especially for heavy ALP samples, and the here discussed preselection on the $B^{\pm}$ decay.

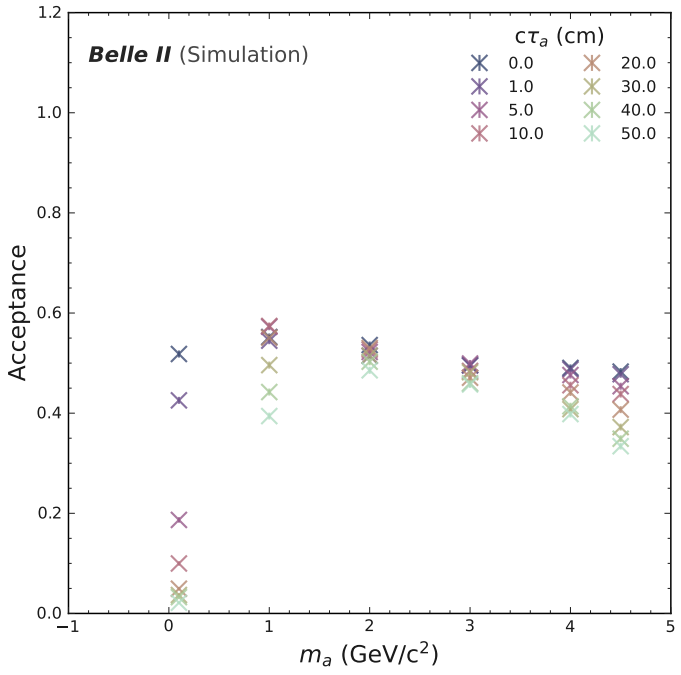


Figure 4.26.: This plot shows the total $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ acceptance with the applied preselection for the photons, the kaon, M_{bc} and ΔE. The signal signature displays a total acceptance of 50 % to 60 % for prompt ALP mass samples. Additionally, the acceptance decreases with higher ALP lifetime samples as discussed for the photon acceptance.

4.3.2. Signal Range

To perform the PFOM optimization as efficiently as possible, choosing a region in which the signal efficiency is already enriched is advantageous. One possible choice is selecting a signal

range in $M_{\gamma\gamma}$ and only maximising the PFOM on events inside this range. The selection and definition of such a range are ambiguous, yet the optimization outcome should ideally not depend on it. Due to the photons in the final state of the neutral $a \to \gamma\gamma$ decay, the $M_{\gamma\gamma}$ picks up a sizeable left tail for displaced ALP hypotheses which makes the definition of a practical signal range challenging. Firstly, this subsection explains the appearance of the left tail in $M_{\gamma\gamma}$ for long-lived ALPs. Secondly, it will discuss a commonly used method to correct momentum misreconstruction of neutral particles in the final state. Lastly, the signal range calculation is discussed.

Energy E_i and momentum $\vec{p}_i$ are the necessary parameters to calculate the invariant mass m_{inv} of a set of particles via

$$m_{\text{inv}}^2 = \left(\sum_i \frac{E_i}{c^2}\right)^2 - \left|\sum_i \frac{\vec{p}_i}{c}\right|^2 . \tag{4.13}$$

In the case of calculating the invariant mass of photons, which are massless, the momentum vector $\vec{p}_\gamma$ is determined by the energy E_γ and the spatial detection in the detector. The only way to measure these quantities with the Belle II detector is by the one crystal layered ECL. Thus, the observables are the reconstructed cluster energy E_{Reco} and cluster spherical coordinates $(\theta_{\text{Reco}}, \phi_{\text{Reco}})$. Since the deposited energy directly reconstructs the cluster energy in the ECL cluster, it does not contribute significantly to the shape of $M_{\gamma\gamma}$. Displaced photons are likely not to hit the face side of crystals, which leads to potentially larger ECL clusters. However, the energy spread into several crystals caused by not-frontal hits is negligible due to the number of radiation lengths per crystal. The difference between reconstructed and generated energy of a photon is mostly very low and tends to reconstruct the energy lower than the generated energy. This can be seen for the lower energetic photon in Fig. 4.27 for correctly reconstructed events[3] in the $m_a = 3.0\,\text{GeV}/c^2$, $c\tau_a = 50.0\,\text{cm}$ ALP sample. Other mass and lifetime samples show similar results with smaller differences compared to lower energetic and slightly larger differences to higher energetic photons.

The reconstruction of the momentum vector is significantly less precise. Due to the absence of a tracking system for neutral particles, such as multiple calorimeter layers, it is unfeasible to measure the accurate direction the particles originated from. Each crystal has a known position relative to the $(0, 0, 0)$ point of the detector, which is combined with the actual IP position to determine the spherical coordinates. In the case of photons, which arise from displaced ALP decays, this method will always lead to wrong values for the coordinates as displayed for a planar example decay in Fig. 4.28.

In total, three different cases exist for the reconstruction of θ_γ and ϕ_γ and are summarized in Fig. 4.29 for the lower energetic photon. Promptly decaying ALPs are correctly reconstructed since the method of calculating the coordinates is applicable. In the detector, evenly distributed displaced photons have the largest over- and underestimation in reconstructing the spherical coordinates. Photons from heavy ALPs are mostly back-to-back oriented. Therefore the reconstruction will lead to smaller values of θ since the photons hit the detector at a flatter angle than anticipated.

[3] These events correspond to the reconstruction in Section 4.1 and Section 4.2 with applied photon preselection.

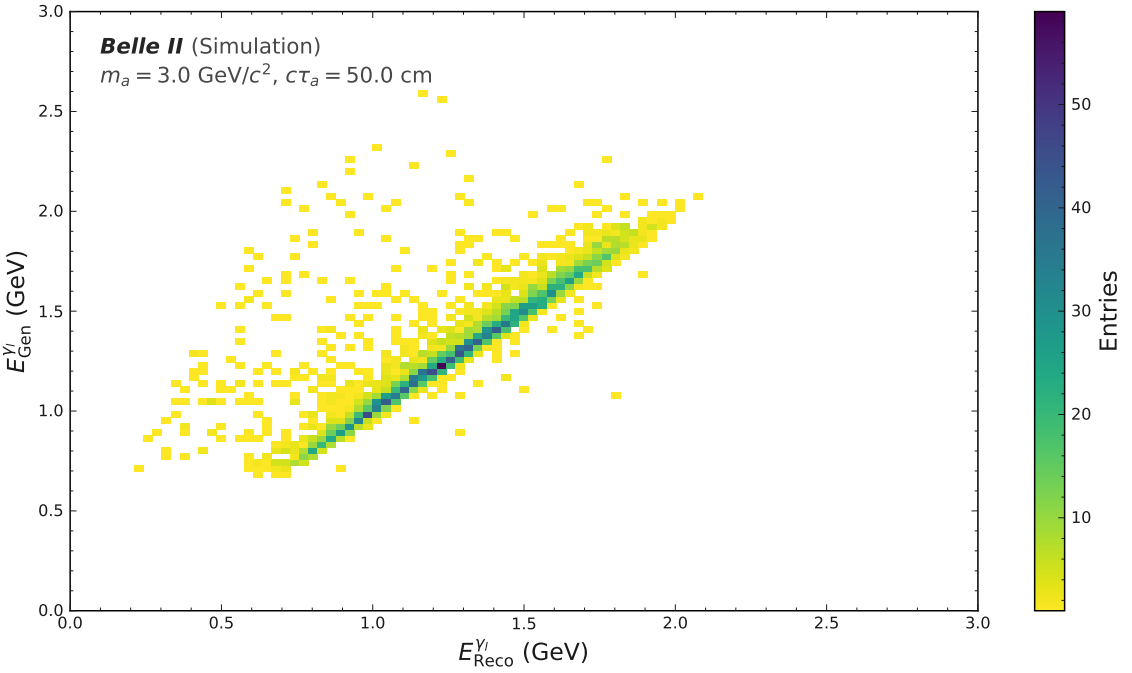


Figure 4.27.: This 2D histogram displays the distributions of generated over reconstructed E_{γ^l} for correctly reconstructed events in the $m_a = 3.0\,\text{GeV}/c^2$ and $c\tau_a = 50.0\,\text{cm}$ sample. The comparisons shows a good agreement between both energy distributions with a slight tendency to reconstruct the energy of the photon to lower values.

Promptly decaying ALPs with a good spherical reconstruction of the photons momentum lead to an approximately Gaussian-like distribution of $M_{\gamma\gamma}$, as shown in Fig. 4.30a, with a slight shift of the peak to the left. The peak shift appears due to the photon energy calibration being performed on a single photon. ALPs with non-zero lifetimes, i. e. a displaced photon production vertex, will broaden the distribution of the invariant mass in the form of a left tail which is more significant for higher displacements. A common analytical description for the shape in Fig. 4.30b is a double-sided Crystal Ball distribution which is defined as

$$f(x;\mu,\sigma,\alpha_l,n_l,\alpha_r,n_r) = \begin{cases} A_l\left(B_l - \frac{x-\mu}{\sigma}\right)^{-n_l} & \text{for } \frac{x-\mu}{\sigma} < -\alpha_l, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \text{for } -\alpha_l \leq \frac{x-\mu}{\sigma} \leq \alpha_r, \\ A_r\left(B_r - \frac{x-\mu}{\sigma}\right)^{-n_r} & \text{for } \frac{x-\mu}{\sigma} > \alpha_r, \end{cases} \tag{4.14}$$

$$\text{with} \quad A_{l/r} = \left(\frac{n_{l/r}}{\alpha_{l/r}}\right)^{-n_{l/r}} e^{-\frac{\alpha^2_{l/r}}{2}},$$

$$B_{l/r} = \frac{n_{l/r}}{\alpha_{l/r}} - \alpha_{l/r},$$

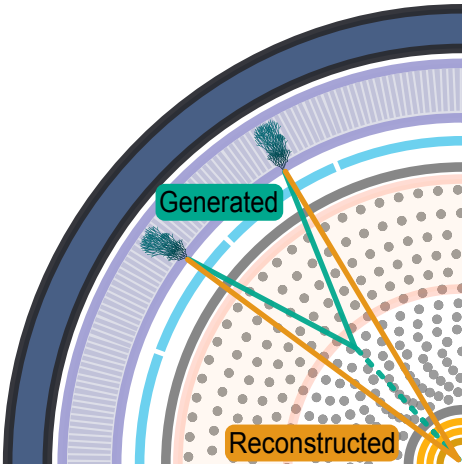
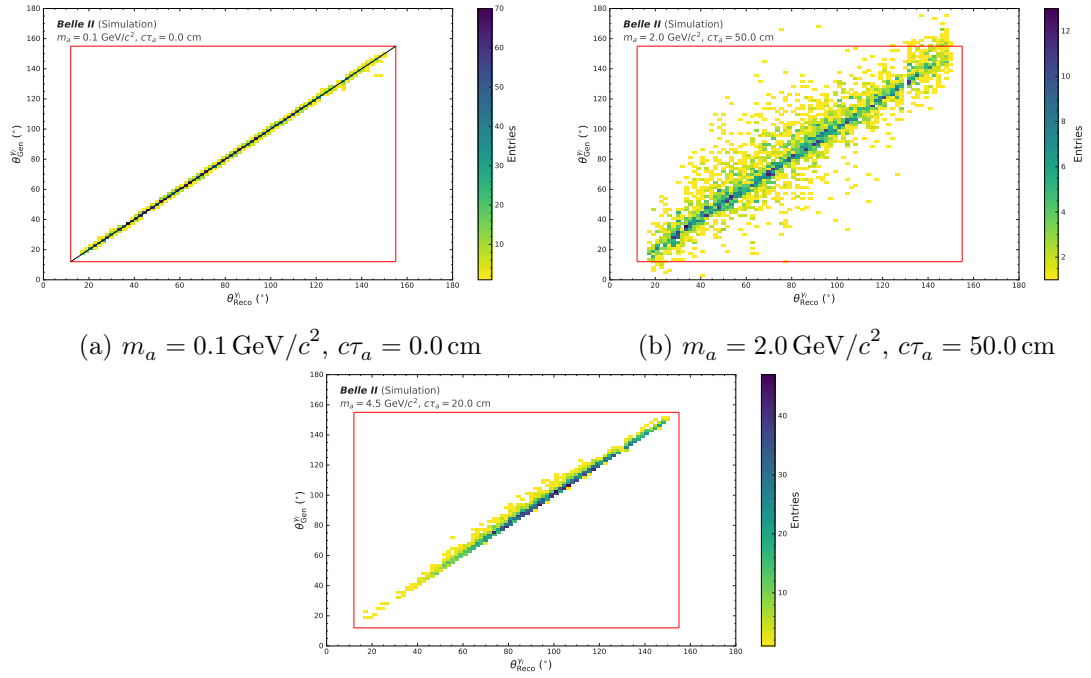


Figure 4.28.: This graphic illustrates the inner part of the detector and the generated and reconstructed event topology of the $a \to \gamma\gamma$ decay. The displayed sub-detector components are the silicon trackers (orange), CDC (grey), ARICH (pink), TOP (light-blue), ECL (purple) and the solenoid (dark-blue). Due to the inability to track neutral particles in the detector, the angular distribution of displaced neutral decays is always misreconstructed.

where μ and σ correspond to the mean and standard deviation of the Gaussian peak structure, $\alpha_{l/r} > 0$ determines the position of the transition between Gaussian and exponential distribution and $n_{l/r} \in \mathbb{N}_+$ are the power of the exponential tails.

The mass constrained fit is commonly used to correct for momentum resolution effects in B decays with neutral particles in the final state. This fit is a special case of the global decay chain vertex fitting [35] algorithms developed at Belle II. The general idea behind the proposed method is to fit the entire decay tree, in this case, the $B^\pm \to K^\pm\gamma\gamma$, by minimizing the distance between the measurement and a given hypothesis under consideration of the measurement uncertainty. This distance can be formulated as a set of equations that express the desired constraint depending on the measured four-vectors of all FSP and intermediate decays. In the case of $B^\pm \to K^\pm\gamma\gamma$, with only one charged track, the only option is to constrain the FSP positions and momenta to the nominal mass of the $B^\pm$ meson. The fit will adjust the FSP momentum-four-vectors to reduce the distance between the reconstructed $B^\pm$ four-vector scalar and the PDG value of the $B^\pm$ mass. Updating the four-momentum vector of the photons with the mass constrained fit reduces the tails of $M_{\gamma\gamma}$ distributions. Heavy ALP mass samples achieve the best fit results due to the photons contributing significantly to the momentum and energy. Therefore, effects on the $M_{\gamma\gamma}$ distributions from displaced ALP decays depend strongly on the ALP mass. For example, samples with an ALP mass of $m_a = 2.0\,\mathrm{GeV}/c^2$ show a significant narrowing of $M_{\gamma\gamma}$ in Fig. 4.31a for prompt decays but only a minor effect on highly displaced samples in Fig. 4.31b due to the large discrepancy for the reconstruction of the momentum direction. This changes significantly for samples with a mass of $m_a = 4.0\,\mathrm{GeV}/c^2$. Here, also the displaced samples in Fig. 4.31c show excellent results by the mass constrained fit.

Applying mass constrained fits on the $B^\pm$ reconstructions can significantly reduce the width



(a) $m_a = 0.1\,\text{GeV}/c^2$, $c\tau_a = 0.0\,\text{cm}$

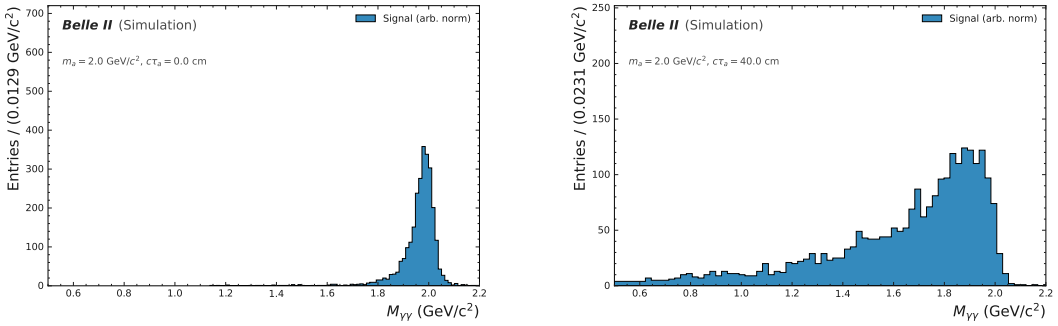
(b) $m_a = 2.0\,\text{GeV}/c^2$, $c\tau_a = 50.0\,\text{cm}$

(c) $m_a = 4.5\,\text{GeV}/c^2$, $c\tau_a = 20.0\,\text{cm}$

Figure 4.29.: These 2D histograms demonstrate the distributions of generated over reconstructed θ_{γ_l} for correctly reconstructed events in three different cases. The plot in (a) displays the correct reconstruction for promptly decaying ALPs. Contrary, the plots in (b) and (c) show the misreconstruction of θ_{γ_l} for displaced ALP events with an uniform polar distribution and a back-to-back like $a \to \gamma\gamma$ decay.

and asymmetry of $M_{\gamma\gamma}$. However, the definition of a signal range in $M_{\gamma\gamma}$ is only reliable if it can account for the remaining asymmetry. The signal range definition presented here is based on the double-sided Crystal Ball distribution since it was chosen to represent the asymmetric shape best. First, an unbinned maximum likelihood fit is performed to extract the double-sided Crystal Ball PDF from the $M_{\gamma\gamma}$ events. Second, the extracted μ parameter marks the center of the PDF, which is split in half at the position of μ. Lastly, strongly inspired by the general Interquartile Range (IQR) definition, the right bound of the signal range is calculated by integrating the PDF from the μ value to the $M_{\gamma\gamma}$ value right of it, which contains 99 % of the area enclosed by the right part of the PDF; analogously for the calculation of the left bound. Fig. 4.32a displays the result of this procedure for a MC signal sample with an ALP mass of $m_a = 3.21\,\text{GeV}/c^2$ and a lifetime of $c\tau_a = 50\,\text{cm}$ from the second set[4] of signal MC samples. The majority of the correctly reconstructed events lie within the asymmetric bounds of the so-called $\text{IQR}_{(1,99)}$ signal range. One limitation to the presented method of signal range calculation is the agreement between the $M_{\gamma\gamma}$ shape and the double-sided Crystal Ball PDF, which is strongly influenced by the number of not reconstructed events. In the case of light ALPs with a high boost, i. e. ALPs leaving

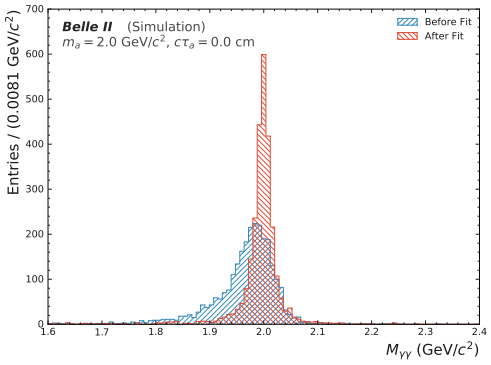
[4] Refer to Table 3.1.



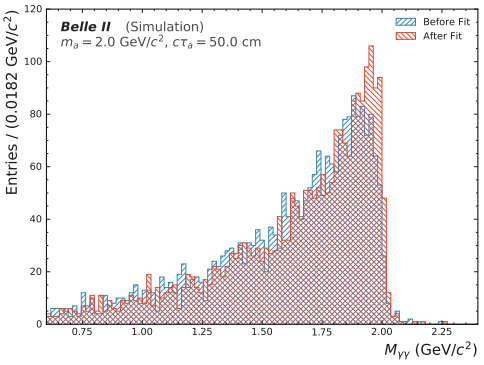
(a) $m_a = 2.0\,\mathrm{GeV}/c^2$, $c\tau_a = 0.0\,\mathrm{cm}$ (b) $m_a = 2.0\,\mathrm{GeV}/c^2$, $c\tau_a = 40.0\,\mathrm{cm}$

Figure 4.30.: These plots present the $M_{\gamma\gamma}$ distributions for two different ALP lifetimes samples with a mass of $m_a = 2.0\,\mathrm{GeV}/c^2$. Samples with a low ALP lifetime show a clear Gaussian-like peak, while high lifetime samples shown a long left tail in the $M_{\gamma\gamma}$ distribution due to the larger spatial reconstruction discrepancy.

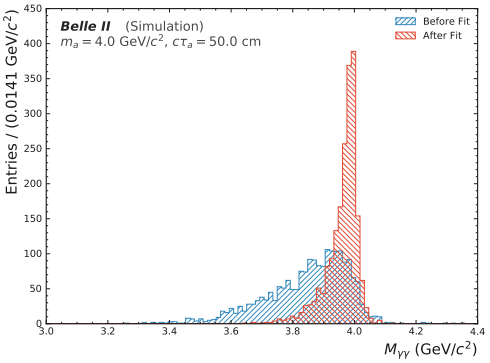
the detector before decaying, the $M_{\gamma\gamma}$ shape can differ significantly from the double-sided Crystal Ball description, as can be seen in Fig. 4.32b. To reliably fit the $M_{\gamma\gamma}$ shape for all mass and lifetime hypotheses, the number of free parameters for the double-sided Crystal Ball distribution is reduced to four by fixing both n parameters. This method is a common technique since the α, and n parameters are highly correlated as both adjust the exponential tails. The parameters are set to empirically determined values of $n_l = 7$ respectively $n_r = 2$.



(a) $m_a = 2.0\,\mathrm{GeV}/c^2$, $c\tau_a = 0.0\,\mathrm{cm}$

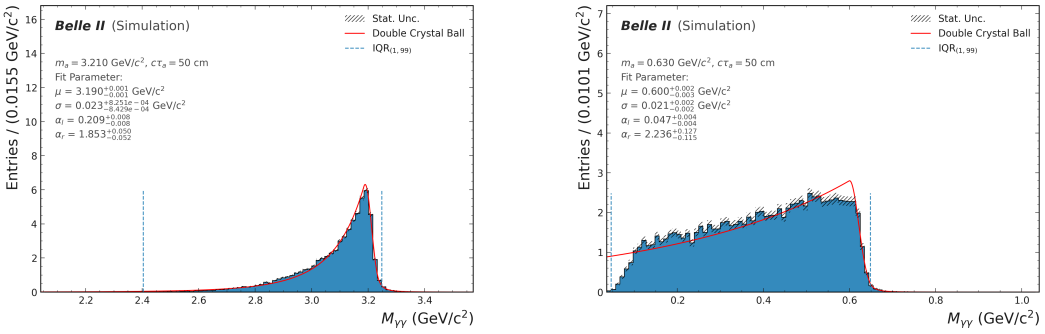


(b) $m_a = 2.0\,\mathrm{GeV}/c^2$, $c\tau_a = 50.0\,\mathrm{cm}$



(c) $m_a = 4.0\,\mathrm{GeV}/c^2$, $c\tau_a = 50.0\,\mathrm{cm}$

Figure 4.31.: These plots display the $M_{\gamma\gamma}$ distributions before and after the mass constrained fit for three different signal MC samples. The mass constrained fit influences the $M_{\gamma\gamma}$ shape more with higher ALP mass as can be seen in comparison of the $m_a = 2.0\,\mathrm{GeV}/c^2$ samples in the plots in (a) respectively (b) and (c).



(a) $m_a = 3.21\,\mathrm{GeV}/c^2$, $c\tau_a = 50.0\,\mathrm{cm}$ (b) $m_a = 0.63\,\mathrm{GeV}/c^2$, $c\tau_a = 50.0\,\mathrm{cm}$

Figure 4.32.: These plots present the $M_{\gamma\gamma}$ distributions, a double-sided Crystal Ball model fit for these events and the signal range calculation for two highly displaced ALP signal samples. High ALP mass samples show good agreement with the analytical description of the double-sided Crystal Ball PDF, while the low ALP mass samples display deviations from it. These deviations arise through ALP events that do not decay in the detector acceptance.

4.3.3. Punzi Figure of Merit

Determining a good set of selections with the help of the PFOM is a typical optimization problem. The goal is to use a set of observables to find the selection that excludes the most background while keeping the signal efficiency as high as possible. This makes the optimization an iterative process that must be adapted to the requirements. The PFOM itself determines two critical aspects in applying the following algorithm. One of the aspects is the direct consequence of the $\frac{a}{2}$ term in the denominator in Eq. (4.9). Consequently, the PFOM can be maximal for zero background events, leading to selections that exclude a large portion of the signal before the signal efficiency decreases significantly. Another aspect is the previously mentioned selection power of an observable, which is proportional to $\epsilon/\sqrt{B}$. This is a useful insight since it can help find good observables for the optimization. For example, if the number of signal events and background events decreases linearly with a tighter selection of an observable, then the PFOM will not have a distinguished maximum. As a result, the used observable can be discarded. Additionally, a profound understanding of the used observables is advantageous for interpreting the suggested selections and possibly adapting them. This subsection will first explain the used algorithm to find a set of selections, then show the intermediate steps between different observables for an example MC sample and conclude with the candidate selection results.

The following algorithm is used to determine the candidate selection based on a set of observables on each ALP mass and lifetime MC sample. The selection variables are not mass constrain fitted.

1. Apply signal range selection determined by the asymmetric $\mathrm{IQR}_{(1,99)}$ bounds based on the mass constrain fitted $M_{\gamma\gamma}$ distribution
2. Exclusion of the irreducible background regions for π^0, η, η' according to Table 4.1
3. Calculate PFOM in bins of a given observable range with the number of standard deviations set to $a = 5$ for discovery and the efficiency ϵ definition

$$\epsilon = \frac{N_{\mathrm{Signal}}}{N_{\mathrm{MC}}}, \tag{4.15}$$

   where N_{Signal} corresponds to the number of events in the MC signal sample and N_{MC} is the number of generated events
4. Choose observable selection value at the position of the PFOM maximum
5. Check if suggested observable selection violates preferred physically motivated limits, if so, set selection value to preferred value
6. Apply selection on MC sample and start at 3. for next variable

The displayed algorithm is applied to the observables in the order from top to bottom in Table 4.2. As previously stated, the order of the observables plays a crucial role in the optimisation result. Interchanging the order of highly correlated variables can change the selection values but ideally leads to similar results. However, interchanging not

Table 4.1.: This table displays the excluded $M_{\gamma\gamma}$ ranges for the particles that form the irreducible background in the $B^{\pm} \to K^{\pm}\gamma\gamma$ decay.

Particle	Excluded $M_{\gamma\gamma}$ range in GeV/c^2
π^0	[0.100, 0.175]
η	[0.450, 0.630]
η'	[0.910, 1.010]

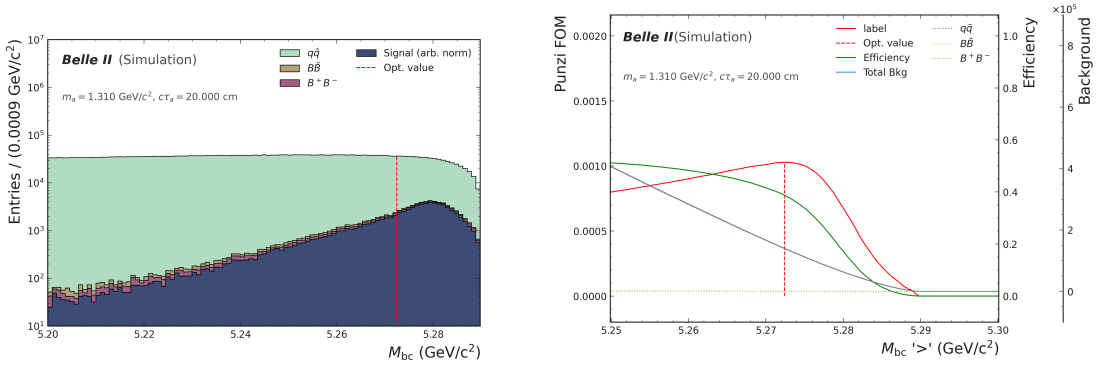
Table 4.2.: This table outlines the order of the observables in the PFOM algorithm. Additionally, it displays the scanned range which is split into 500 selection steps.

Observable	Range of Bins	Comparison Operator
M_{bc}	[5.25, 5.30]GeV/c^2	$>$
ΔE	[0.00, 0.10]GeV	$<$
ΔE	[-0.40, 0.00]GeV	$>$
R_2	[0.00, 1.00]	$<$
$\cos(T_{\mathrm{B}}T_{\mathrm{O}})$	[0.00, 1.00]	$<$
E_{γ^l}	[0.00, 2.50]GeV	$>$

correlated variables can lead to different excluded events and therefore have an impact on the resulting signal efficiency and the number of background events. To better control this difference, it is practical to select correlated variables directly after one another and not mix between highly correlated groups. For this analysis, different orders have been tested, and ultimately, this order has been selected due to a good performance and a reasonable physical motivation. The following example shows the PFOM calculation steps for each variable in the $m_a = 1.31\,\mathrm{GeV}/c^2$, $c\tau_a = 20\,\mathrm{cm}$ MC sample.

A key motivation in the choice of the observable order is to first exclude as many events as possible without reducing the signal efficiency drastically. For this reason, motivated by the discussion in Section 4.3.1, a selection on M_{bc} begins with the exclusion targeted on events that likely do not originate from a B meson decay. Fig. 4.33b shows a good separation power by a clear peak in the PFOM which looses, on the one hand, about 30 % signal efficiency, whereas, on the hand, it excludes about 60 % of the background which is also indicated by Fig. 4.33a.

The following selection variable, ΔE, follows the same motivation as the M_{bc} selection. Additionally, selections on ΔE exclude misidentified kaons since the particle hypothesis is included in the energy calculation of the particle. Since the selection on ΔE is split into two parts, for values smaller and larger than zero, the first selection will always show a low separation power, as seen in Fig. 4.34b, due to the background events on the other side of $\Delta E = 0\,\mathrm{GeV}$. As a result, the second selection will always seem to have a more significant impact, displayed by the clear PFOM peak in Fig. 4.34d. Interchanging the order of these selections has no significant effect on the fully optimized selection. A physically motivated



(a) M_{bc} Distribution (b) M_{bc} PFOM

Figure 4.33.: These two plots display the PFOM, signal efficiency and the number of background events distribution and indicate the best selection in the observable distribution for M_{bc} in one signal MC sample. The selection for M_{bc} shows a clear optimum in the PFOM.

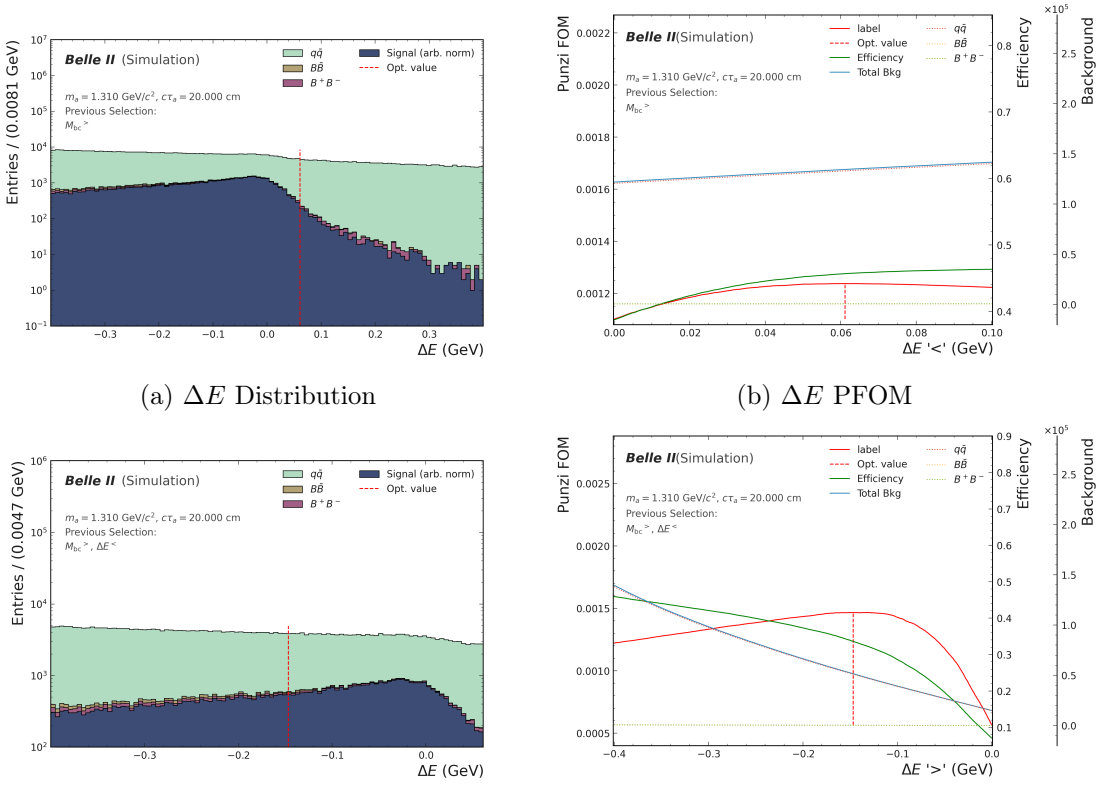
reason to choose the displayed order is that more background events occur for negative ΔE values.

Even after reducing the number of background events by 90 %, the continuum background still dominates the reconstructed events, as can be seen in Fig. 4.34c. For this reason, the next selection optimization observables are for continuum suppression. Here again, a change of the order between R_2 and $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ does not significantly change the final selection result due to excluding the same events. The first selection on one of the continuum suppression variables reduces the number of background events drastically due to the extremely high selection power displayed in Fig. 4.35b. Simultaneously, about half the signal events are also excluded, resulting from the PFOM denominator. Consequently, a hard limit is put on the selection for R_2, which does not allow for selections with $R_2 < 0.2$[5]. For the displayed case in Fig. 4.35c, nearly no events are left to be excluded by a $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ selection, which is also visible in the plateau in Fig. 4.35d. However, $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ is not excluded from the optimization to possibly exclude events in samples with less spherical ALP decays. As an outlook, the continuum suppression can be improved by the use of boosted decision trees trained specifically for the continuum background rejection on a larger set of continuum suppression variables. Additionally, advanced machine learning methods could be used further to improve the continuum suppression, so-called Deep Continuum Suppression, as shown in [36].

The last candidate selection variable is the energy of the lower energetic photon. It shows a good separation power in Fig. 4.36b since the minimal energy of the photons is determined by the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ kinematics. However, E_{γ^l} is chosen last due to the steep linear reduction in signal efficiency, which always leads to a significant loss of signal events.

Fig. 4.37 shows the final selection results for all ALP mass and lifetime samples. Most

[5]This value is the result of private discussions with a continuum suppression expert and justified that even for perfectly reconstructed spherical events this would drastically reduce the number of signal events.



(a) ΔE Distribution

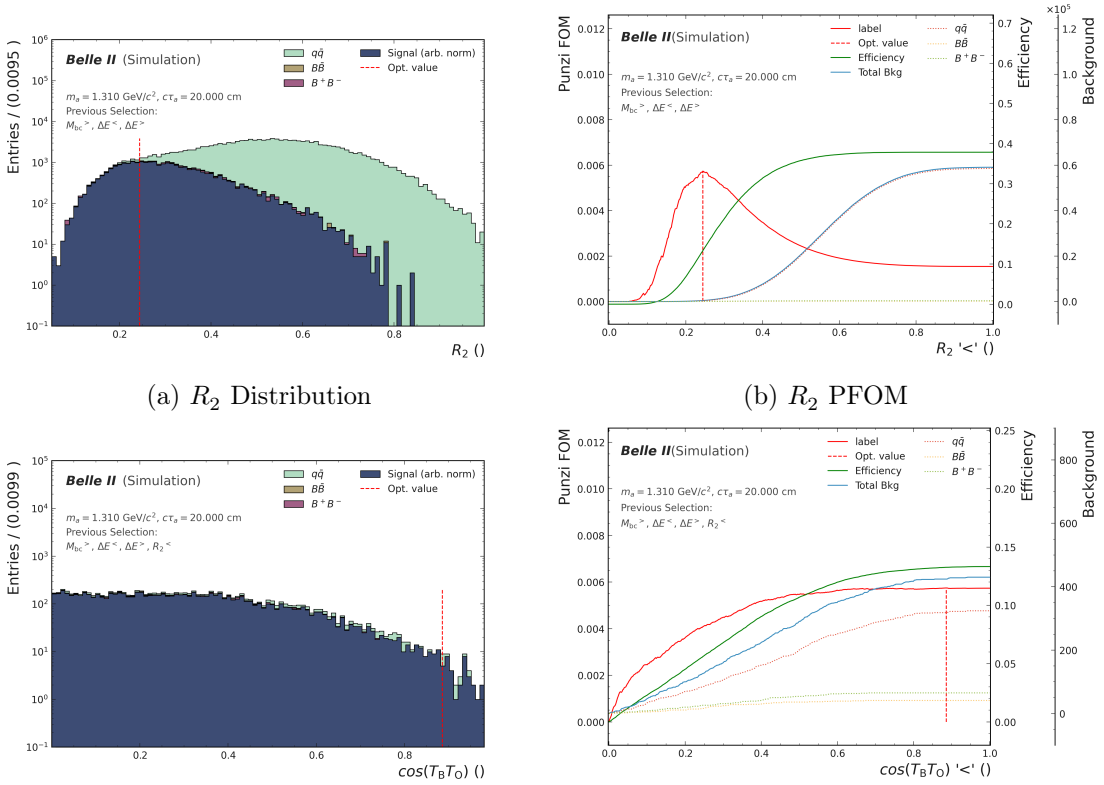
(b) ΔE PFOM

(c) ΔE Distribution

(d) ΔE PFOM

Figure 4.34.: These four plots display the PFOM, signal efficiency and the number of background events distribution and indicate the best selection in the observable distributions for negative and positive values in ΔE. The selection for negative ΔE values displays a clear optimum in the PFOM due to the previous selection on the positive values. Therefore, an interchange of the order of the selection leads to similar results.

of the selections are within the expectations, which can be deduced from Section 4.3.1 and the previous example. The optimal selection for M_{bc} tends to lower values for higher lifetimes due to the broader M_{bc} distribution caused by the misreconstructed photon spatial coordinates . The selections are approximately identical for positive values of ΔE throughout all mass and lifetime points. In contrast, negative values require a loser selection with rising lifetime due to the calibration of the photon energy. R_2 selections are dependent on the mass of the ALP sample. High and low ALP mass samples cause higher R_2 values due to the boost of the photons. Thus, the selection for these samples is looser. The selection for more spherical events is tighter and does not go below $R_2 < 0.2$ as required, resulting in a very loose $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ selection since most of the continuum events are already excluded. In ALP mass regions where the R_2 selection is looser, the $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ selection is tighter due to remaining continuum events. Lastly, optimal E_{γ^l} selections are slightly above the kinematic allowed minimum photon energy with respect to the ALP mass. Fig. 4.37e shows isolated outliers which are caused by the absence of sufficient events in the sample



(a) R_2 Distribution

(b) R_2 PFOM

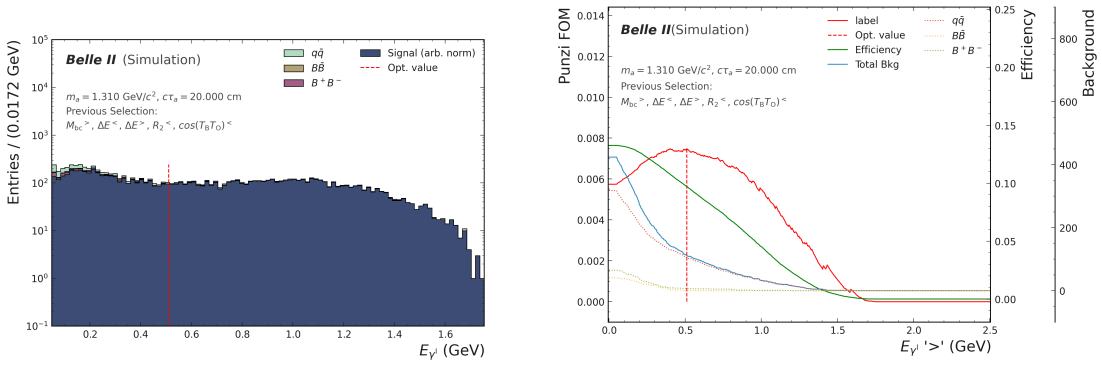
(c) $\cos(T_B T_O)$ Distribution

(d) $\cos(T_B T_O)$ PFOM

Figure 4.35.: These four plots display the PFOM, signal efficiency and the number of background events distribution and indicate the best selection in the observable distributions for R_2 and $\cos(T_B T_O)$. The selection for R_2 shows a good separation power between signal and background events with a high loss in signal efficiency.

after applying the previous selections.

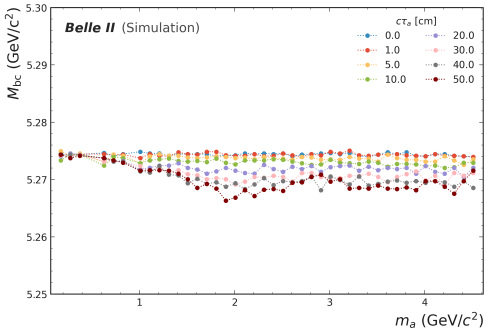
Fig. 4.38 summarizes the final candidate selection results by displaying the overall signal efficiency and the number of background events for all ALP mass and lifetime samples. Generally, the efficiency mirrors the shape of the acceptance in Fig. 4.26 with lower values for light and heavy ALP samples. Contrary to the acceptance, the separation between lifetimes is more apparent for the signal efficiency, which can be explained by the different M_{bc} and ΔE selections for different lifetime samples. These selections are looser for higher lifetimes, yet, to achieve a comparable signal efficiency for samples with lower lifetimes, they should be even looser. Loosening the M_{bc} and ΔE selections would lead to more background events, which is not the goal of this candidate selection. Fig. 4.38b displays the expected number of background events in the signal range for $100\,\mathrm{fb}^{-1}$, considering only continuum and generic background. The number of background events increases with the lifetime hypothesis of the MC samples, which is expected due to the significant larger signal ranges. In conclusion, the optimization of the PFOM shows a good background rejection for promptly decaying ALPs while maintaining a signal efficiency between 10 % to 20 %



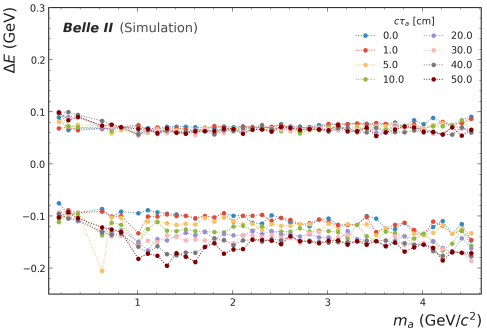
(a) E_{γ^l} Distribution (b) E_{γ^l} PFOM

Figure 4.36.: These two plots display the PFOM, signal efficiency and the number of background events distribution and indicate the best selection in the observable distribution for E_{γ^l}. The selection for E_{γ^l} shows a good separation power and reduces the number of background events drastically.

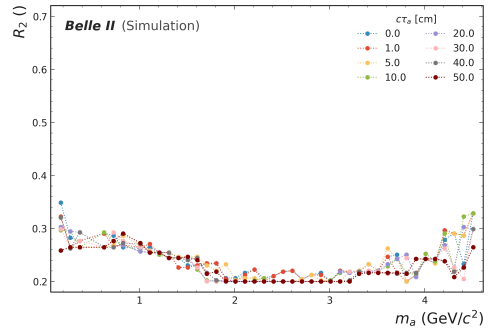
over the whole ALP mass range. The signal efficiency decreases for higher lifetimes, e. g. 1 % to 10 % in the case of $c\tau_a = 50\,\text{cm}$ where the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ kinematics causes low signal efficiency. Improvements could be possibly achieved by constructing not yet considered observables with a clearer separation power between signal and background events. Especially interesting is the implementation of an MVA based continuum suppression to keep more signal events while still excluding the dominating continuum background. Furthermore, [37] proposes an MVA based ansatz for the use of the PFOM where the figure of merit acts as loss function and generalises well for untrained mass hypotheses. This method should be tested for the case of different lifetime hypotheses.



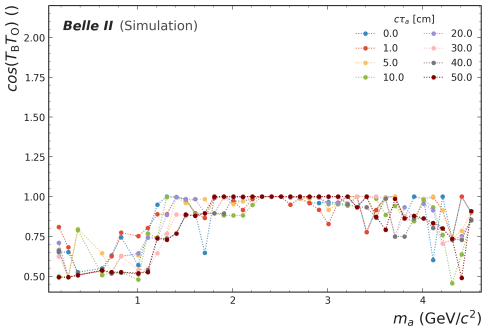
(a) M_{bc}



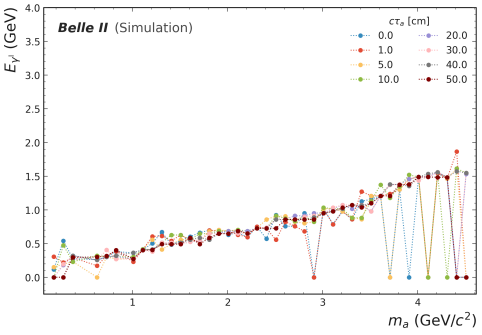
(b) ΔE



(c) R_2

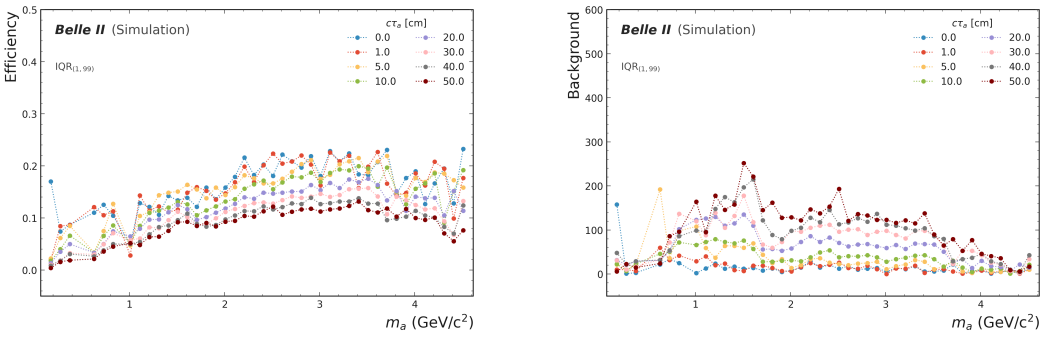


(d) $\cos(T_{\mathrm{B}} T_{\mathrm{O}})$



(e) E_{γ^l}

Figure 4.37.: This plot series presents the final selection values for all observables in all signal MC samples. Excluding some outliers, nearly selections are within physical expectations.



(a) Efficiency (b) Background

Figure 4.38.: These two plots display the resulting signal efficiency (a) and number of background events (b) for all signal MC samples after applying the previously shown selections.

5. Signal Extraction

The last analysis chapter discusses the scan of $M_{\gamma\gamma}$ and shows exclusion limits for the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ branching ratio in case no signal is found. Section 5.1 explains the intermediate steps between optimizing the selection with PFOM on MC and performing a thin stepped scan of $M_{\gamma\gamma}$. Subsequently, Section 5.2 describes the signal + background fit on the continuum and generic MC samples. Finally, as preparation for the case of no significant signal yield excess in data, Section 5.3 will show the expected limit on the branching ratio $\mathrm{Br}(B \to Ka) \times \mathrm{Br}(a \to \gamma\gamma)$, which rejects the signal + background hypothesis.

5.1. Scan Preparation

To ensure that no signal events are overlooked when scanning the invariant mass $M_{\gamma\gamma}$, the distance between neighbouring mass hypotheses must be minimal. Since the width of the signal event distribution is not clearly defined, this analysis uses half the smallest Gaussian standard deviation of all signal shapes. As a result, in the case of New Physics events at any position in $M_{\gamma\gamma}$, they will be scanned, even if they lie between two mass hypotheses. Producing enough signal MC samples to fulfil the mass hypothesis distance requirements is impractical. Thus, the signal shape and region, the optimal candidate selection and the anticipated signal efficiency must be calculated by interpolating the acquired results from signal MC samples. The interpolation is performed over all ALP mass samples and separately for each lifetime hypothesis. Section 5.1.1 discusses the interpolation results for the fit range and the optimized selection values. These interpolation results are then used to recalculate and interpolate the signal efficiency and the $M_{\gamma\gamma}$ signal shape parameters in Section 5.1.2.

The interpolation of the results of Chapter 4 is not subject to any specifications. The primary motivation of the methods used here is to create a smooth transition between the different mass points. For this reason, if needed, any strong outliers are first removed from the results to be interpolated. Polynomials then interpolate the remaining results. For the explanation of the two intermediate interpolation steps, $\{y_i\}$ represents the result to be interpolated and $\{x_i\}$ the generated ALP mass.

Outliers in $\{y_i\}$ are determined by the distance to their nearest neighbors. For each y_i, the distances $d_{j,i}$ to its neighbours, $j = i + 1$ and $j = i - 1$, is calculated by

$$d_{j,i} = \frac{|y_j - y_i|}{\Delta y}, \tag{5.1}$$

where Δy corresponds to the median of the absolute differences between the values of $\{y_i\}$. If both distances $d_{j,i}$ are larger than a number f, the result y_i is considered an outlier and therefore not interpolated. For the first two and last two y_i in $\{y_i\}$, the distance to the nearest two neighbours is calculated, and if either of these distances is larger than f, the y_i are considered outliers. The number f is empirically determined for each set of interpolations. For the here shown interpolations that require the exclusions of outliers, a value of $f = 30$ is chosen.

The interpolation of a polynomial model to the remaining $\{y_i\}$ is performed by minimizing an unscaled χ^2-function

$$\chi^2 = \sum_i \left(P(x_i, \{c_j\}) - y_i\right)^2, \tag{5.2}$$

where $P(x_i, \{c_j\})$ corresponds to the polynomial model and $\{c_j\}$ are the set of coefficients for the polynomial terms. The following interpolations uses Chebyshev polynomials $P(x_i, \{c_j\}) = \sum_j c_j T_j(x_i)$ of the first kind which are defined as

$$T_n(\cos\theta) = \cos(n\theta). \tag{5.3}$$

The benefit of Chebyshev polynomials is their orthonormal property which decorrelates the coefficients in the polynomial model. As a result, the coefficient values are stable for different orders of polynomials. Adding polynomials that do not benefit in the expectation description will ideally not lead to coefficients significantly different from zero.

5.1.1. Fit Range and Candidate Selection

Section 4.3.2 calculates an asymmetric signal range $\mathrm{IQR}_{(1,99)}$ to enclose 99 % of the signal events in the $M_{\gamma\gamma}$ distribution. To perform a scan of $M_{\gamma\gamma}$ for signal events, each signal hypothesis fit requires a similar definition for the fitted range in $M_{\gamma\gamma}$. This so-called fit range needs to include the full signal range and enough background events to determine a background PDF. For this reason, the asymmetric fit range $\mathrm{IQR}_{(1,99)} \times 10$ is defined as ten times the signal range. More precisely, the upper bound of $\mathrm{IQR}_{(1,99)} \times 10$ is ten times the difference between the upper value of the signal range and the corresponding double-sided Crystal Ball parameter μ added to μ; analogously for the lower bound. In total, the upper and lower bounds of the fit range are:

$$\max(\mathrm{IQR}_{(1,99)} \times 10) = \mu + (\max(\mathrm{IQR}_{(1,99)}) - \mu) \times 10, \tag{5.4}$$

$$\min(\mathrm{IQR}_{(1,99)} \times 10) = \mu - (\mu - \min(\mathrm{IQR}_{(1,99)})) \times 10. \tag{5.5}$$

The fit range is calculated for each signal MC sample individually. For the interpolation, no outlier detection is applied. However, lower bounds of $\mathrm{IQR}_{(1,99)} \times 10$ are set to zero if they are calculated to be negative values in $M_{\gamma\gamma}$. Fig. 5.1 shows the interpolation between different ALP mass hypotheses with a lifetime of $c\tau_a = 5.0\,\mathrm{cm}$. The upper fit range bound

is interpolated by a polynomial of second order. First-order polynomials are primarily sufficient for all lifetime samples, yet, the second-order is added to account for a slight bending trend towards higher ALP masses. A fourth-order polynomial is interpolated to account for the distributions of the lower fit range bounds, which are dependent on the lifetime of the samples. The choice of a fourth-order polynomial is mainly motivated by higher lower bound values towards higher ALP masses which are caused by the mass constrained fit. Other lifetime samples show similar results to Fig. 5.1. Deviations between the originally calculated $\mathrm{IQR}_{(1,99)} \times 10$ values and the interpolated values are not significant for two reasons. First, the loss of signal events is improbable due to the large range. Second, it is practical to cut the fit range off at the bounds of the irreducible background, which is explained in Section 5.2.

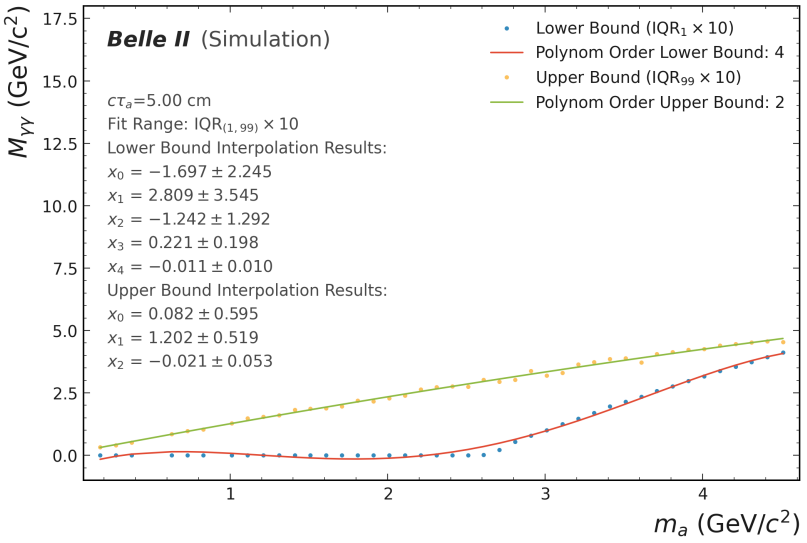
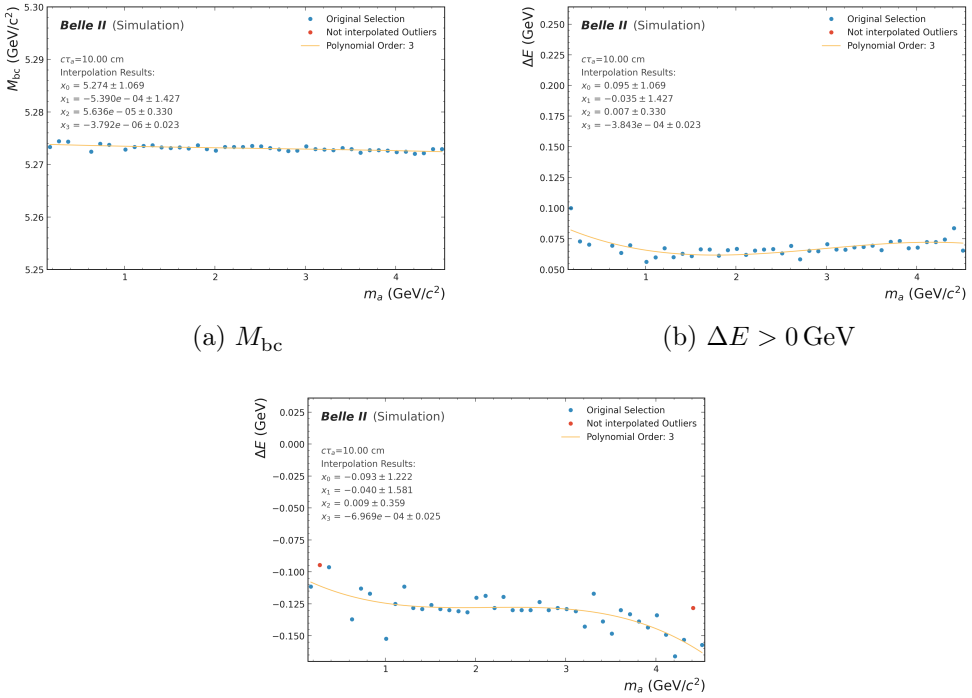


Figure 5.1.: This plot displays the interpolation of the fit range $\mathrm{IQR}_{(1,99)} \times 10$ over all ALP mass samples with a lifetime of $c\tau_a = 5.0\,\mathrm{cm}$. Upper fit range bounds are interpolated by a polynomial of the second order and lower bounds by a fourth-order polynomial. The lower bounds only can have minimal value of zero.

The interpolation of the optimized selections in Section 4.3.3 uses the outlier exclusion to make the interpolation less susceptible to selections that do not fit to their neighbouring mass hypothesis as in Fig. 4.37e. The interpolations use polynomials of the third order to account for different selection distributions for all observables. Using the $c\tau_a = 10.0\,\mathrm{cm}$ as an example, Fig. 5.2 displays the interpolation results for the selections of M_{bc} and ΔE. Therefore, the selections and their interpolations can be very tight, resulting in only very few background events due to the high separation power between signal and background events. As a result, for all lifetime samples, the selections for M_{bc} and ΔE are not set to the interpolated values, but rather $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ and $-0.15\,\mathrm{GeV} < \Delta E < 0.10\,\mathrm{GeV}$. These selections allow for a few more background events while not influencing the signal

events significantly. The largest impact, regarding the signal event exclusion, can be observed by selections for $\Delta E < 0\,\text{GeV}$ on high ALP mass samples with lifetimes $c\tau_a \geq 10.0\,\text{cm}$ as Fig. 5.2c implies.

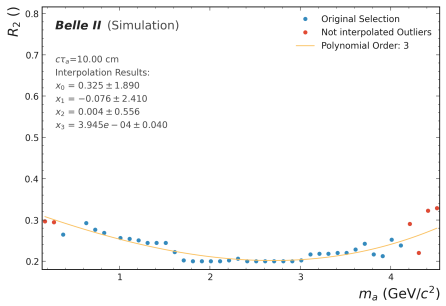


(a) M_{bc} (b) $\Delta E > 0\,\text{GeV}$

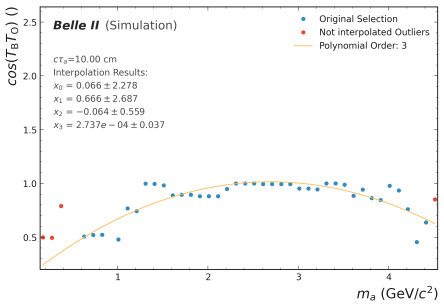
(c) $\Delta E < 0\,\text{GeV}$

Figure 5.2.: These three plots display the third-order polynomial interpolation of the M_{bc} and ΔE selections over all ALP masses for samples with a lifetime of $c\tau_a = 10.0\,\text{cm}$. For the further analysis, these interpolations ar enot taken into account, rather the selections are set to $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$ and $-0.15\,\text{GeV} < \Delta E < 0.10\,\text{GeV}$.

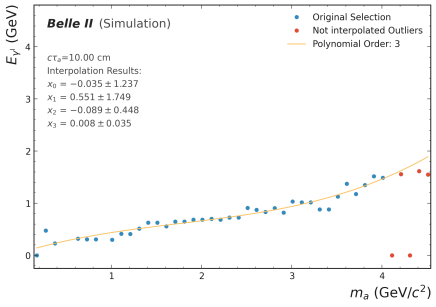
Fig. 5.3 shows the interpolation results for the remaining selection variables R_2, $\cos(T_{\text{B}}T_{\text{O}})$ and E_{γ^l}. Here, the interpolated values will be used for the following discussion. Differences between the optimized selections and interpolated results can significantly impact the signal efficiency in cases where the interpolated selection is tighter. In the here presented example for samples with an ALP lifetime of $c\tau_a = 10.0\,\text{cm}$, this occurs twice. Firstly, for higher ALP masses, the interpolated R_2 selection is tighter in Fig. 5.3a, which occurs only for the $c\tau_a = 10.0\,\text{cm}$ samples. Secondly, Fig. 5.3b shows a tighter selection on $\cos(T_{\text{B}}T_{\text{O}})$ for lower ALP masses, which can be observed for all signal MC samples. Outliers cause tighter selections in the respective ALP mass regions in the displayed cases. However, these results are still used in the following analysis since no better agreement on the interpolation parameters could be found that simultaneously kept the interpolation for all ALP mass and lifetime samples as coherent. Possible improvements can be achieved by a higher density of signal MC samples in the high and low ALP mass regions, allowing for better interpolations.



(a) R_2



(b) $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$



(c) E_{γ^l}

Figure 5.3.: These three plots display the third-order polynomial interpolation of the R_2, $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ and E_{γ^l} selections over all ALP masses for samples with a lifetime of $c\tau_a = 10.0\,\mathrm{cm}$. The interpolations for R_2 and $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$ a significantly tighter selection in high respectively low ALP mass regions than calculated with the help of the PFOM.

5.1.2. Signal Efficiency and Signal Shape

Setting a limit on the branching ratio of decays requires the knowledge of the corresponding signal efficiency. Furthermore, in this analysis, the signal efficiency is a function of the New Physics model parameters or rather their respective selection, as explained in Section 4.3. Therefore, the efficiency is recalculated for each signal MC sample according to the interpolated selections in the interpolated fit range $\mathrm{IQR}_{(1,99)} \times 10$.

Fig. 5.4 shows the recalculated signal efficiency values in the fit range for all signal MC samples. In comparison to Fig. 4.38, the distributions of the efficiency values for the same ALP lifetime samples show a smooth transition between different ALP masses. The previously mentioned tighter selection on R_2, caused by the interpolation, for $c\tau_a = 10.0\,\mathrm{cm}$ shows the expected loss of efficiency for higher ALP mass samples in comparison to the $c\tau_a = 5.0\,\mathrm{cm}$ and $c\tau_a = 20.0\,\mathrm{cm}$. For the interpolation of the signal efficiencies, fourth-order polynomials are chosen. Fig. 5.5 displays the interpolated signal efficiency for the $c\tau_a = 20.0\,\mathrm{cm}$ samples. Here, differences between interpolation and the calculated values should be as small as possible since the interpolated values are directly used for the branching ratio calculation. Due to the smooth transitions between the ALP mass

samples for $m_a \geq 1.01\,\mathrm{GeV}/c^2$, the differences are minimal. For smaller ALP masses, the differences are larger. More MC samples between the ALP mass hypotheses could improve the interpolation.

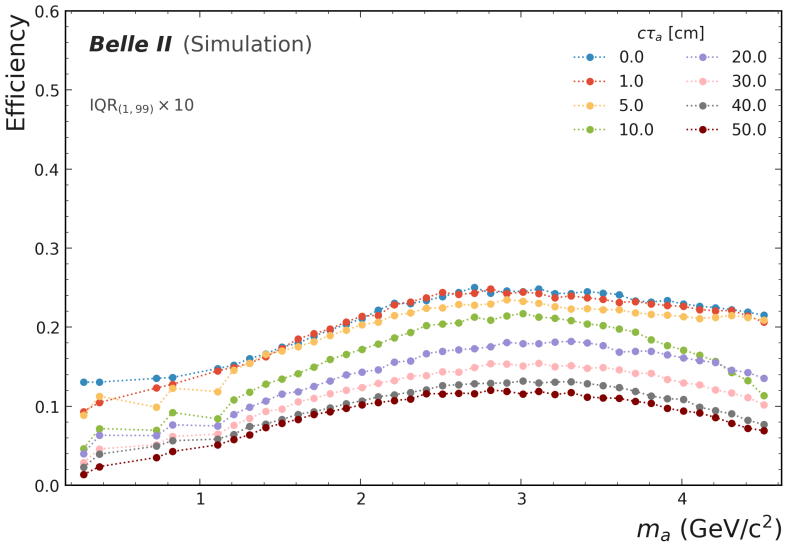


Figure 5.4.: This plot shows the recalculated signal efficiency for all ALP signal MC samples in $\mathrm{IQR}_{(1,99)} \times 10$ after the application of the interpolated selections. The efficiency shows smooth transitions between neighbouring ALP mass samples and a decrease with increasing ALP lifetime.

The last step in preparing the $M_{\gamma\gamma}$ scan is determining the expected signal shape for each signal hypothesis. As already stated in Section 4.3.2, an analytical description of the signal shape can be achieved by the parameters of the double-sided Crystal Ball distribution. To determine the parameters, the interpolated signal selection is applied to every signal MC sample, and the $M_{\gamma\gamma}$ distribution is fitted in the respective fit range $\mathrm{IQR}_{(1,99)} \times 10$. Subsequently, the parameters are interpolated between the different ALP mass hypotheses. The fit procedure is similar to the signal range extraction in Section 4.3.2: use of an extended double-sided Crystal Ball model to extract the signal yield, and an unbinned maximum likelihood loss. The signal yield does not influence the shape parameter determination. Fig. 5.6 displays the $M_{\gamma\gamma}$ fit for the $m_a = 1.51\,\mathrm{GeV}/c^2$, $c\tau_a = 30.0\,\mathrm{cm}$ MC sample. The fitted signal yield is converted to the $B^{\pm} \to K^{\pm}\gamma\gamma$ cross section by assuming a luminosity of $100\,\mathrm{fb}^{-1}$ and the signal efficiency is set to the interpolated value. Consequently, the signal efficiency is slightly overestimated since the resulting cross section should yield 500 fb (50'000 generated signal events).

Fig. 5.7 summarizes the fitted double-sided Crystal Ball parameters. The μ parameter differs slightly from the underlying ALP mass hypotheses with a small bias towards smaller values due to the shift towards smaller values of the $M_{\gamma\gamma}$ signal peak. Later stages of the analysis, outside the scope of this thesis, will correct this bias with linearity tests. The σ and α_l parameters generally show smooth transitions between ALP mass hypotheses, with only

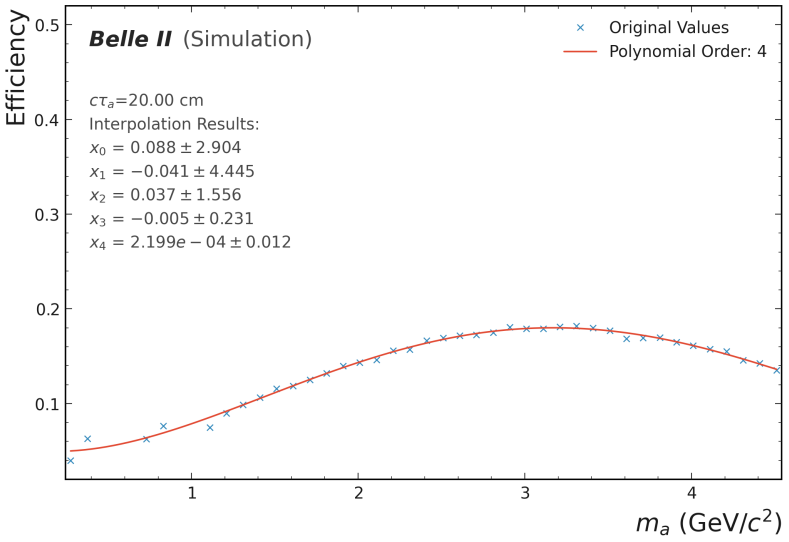


Figure 5.5.: This figure presents the interpolation of the signal efficiency for all signal samples with an ALP lifetime of $c\tau_a = 20.0\,\text{cm}$. A fourth-order polynomial is used for the interpolation.

a small number of outliers that can be excluded before the interpolations. Contrary to that, α_r shows chaotic distributions, especially for ALP masses $m_a \leq 3.0\,\text{GeV}/c^2$. No specific outlier exclusion can be performed for this parameter. As a result, instead of interpolating the α_r parameter for each lifetime separately, the median overall mass samples will be used for the signal hypotheses. This will result in a slightly wrong signal shape. However, no other procedure will lead to better results due to the high correlation between the α and σ parameters. For example, suppose the α_r parameter is set to a fixed value, like it is done with the n parameters, and not fitted. In that case, one or both of the remaining parameters will have a chaotic distribution since they are determining the shape. Finally, the μ, σ and α_l parameters are interpolated, after excluding the outliers, for each ALP lifetime hypothesis with polynomials of the second order.

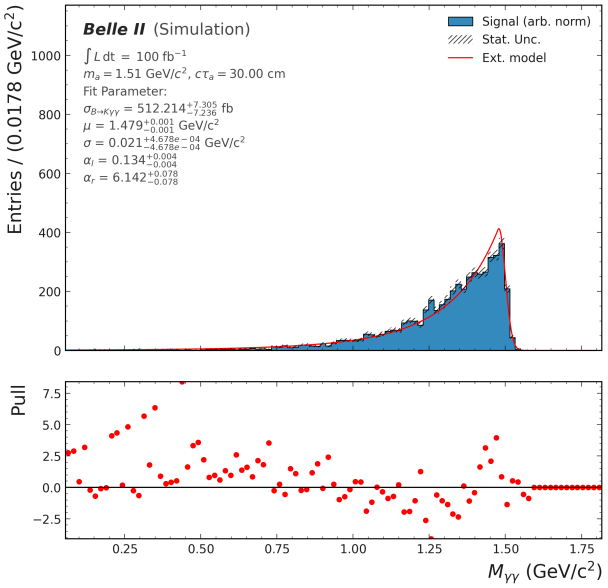
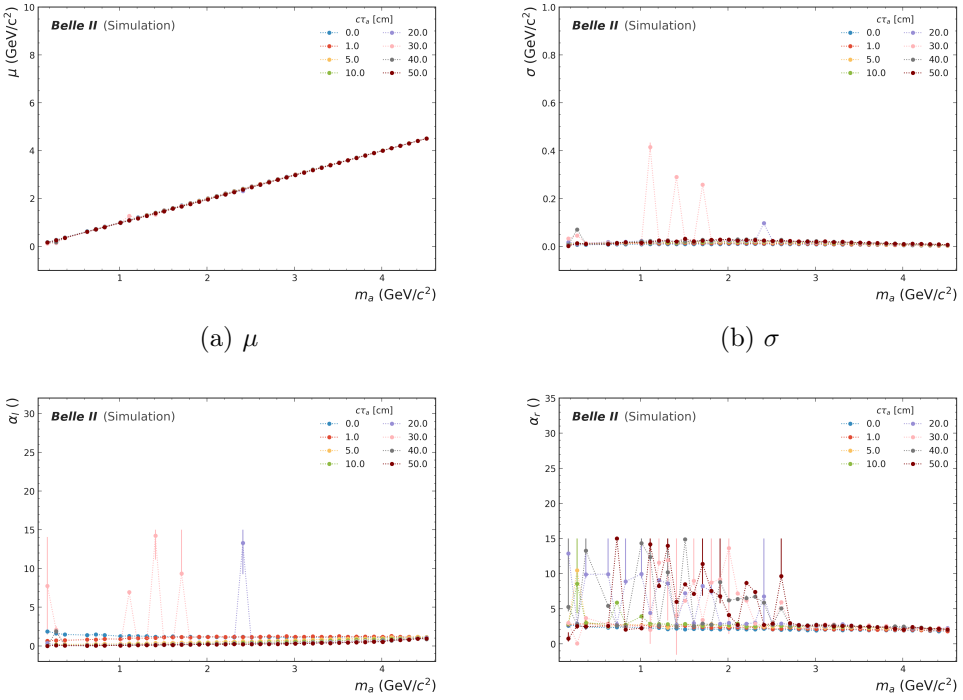


Figure 5.6.: This plot displays the $M_{\gamma\gamma}$ distribution for the $m_a = 1.51\,\mathrm{GeV}/c^2$, $c\tau_a = 30.0\,\mathrm{cm}$ signal sample after the application of the interpolated selection. The $M_{\gamma\gamma}$ is fitted by an extended double-sided Crystal Ball model and the pull is calculated and displayed in the bottom part of the plot.



(a) μ (b) σ

(c) α_l (d) α_r

Figure 5.7.: These four plots present the fitted double-sided Crystal Ball parameters for all signal MC samples. The parameters μ, σ and α_l are interpolated by second-order polynomials for the following analysis. Due to the chaotic distribution of α_r, the median of each lifetime is used for further calculations.

5.2. Invariant Di-Photon Mass Fits

The last major task in this thesis is the extraction of a signal yield for the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ signature over the considered $M_{\gamma\gamma}$ range. For this purpose, all previous results are combined to form signal + background ALP mass hypotheses in steps of $2.5\,\mathrm{MeV}/c^2$ for eight different lifetimes. These ALP hypotheses are fitted on the continuum and generic MC samples. This subsection explains the $M_{\gamma\gamma}$ scan procedure and discusses the fit strategy.

The first step of the $M_{\gamma\gamma}$ scan is the determination of the different signal + background hypotheses. For this purpose, the $M_{\gamma\gamma}$ interval $0.175\,\mathrm{GeV}/c^2$ to $4.600\,\mathrm{GeV}/c^2$ is split in equidistant steps of $2.5\,\mathrm{MeV}/c^2$, which corresponds to the smallest interpolated σ double-sided Crystal Ball parameter. These scan points in $M_{\gamma\gamma}$ are used to calculate the interpolated values for the fit range $\mathrm{IQR}_{(1,99)} \times 10$, the signal shape parameters μ, σ, α_r, α_l, and the selection values for M_{bc}, ΔE, R_2, $\cos(T_{\mathrm{B}}T_{\mathrm{O}})$, and E_{γ^l}. Due to the small shift of μ to lower values in m_a, the μ parameter is used in the next step to determine if a hypothesis is inside the exclusion regions of π^0, η, η'. Subsequently, all hypotheses for which the μ parameter is inside the intervals in Table 4.1 are excluded. This procedure is performed for every lifetime hypothesis separately.

The next step is the fit of all ALP hypotheses on the $100\,\mathrm{fb}^{-1}$ continuum and generic MC samples. Therefore, a signal + background PDF is formed by the sum of the extended signal and background PDFs for each scan point. On the one hand, the signal PDF consists of the double-sided Crystal Ball PDF $S(M_{\gamma\gamma})$, with fixed parameters corresponding to the interpolated values, and the signal yield N_{sig}. On the other hand, Chebyshev polynomials are used to estimate the background PDF $B(M_{\gamma\gamma};\{x_i\})$ in combination with the yield N_{bkg}. Polynomials of the third-order are used to model the background distribution. This choice is motivated by the observation that in fits, where only the background PDF is used, the first-order coefficient is the most dominant. For some scan points, higher-order polynomial coefficients are significantly different from zero, especially in the low and high ALP mass regions, while in between these regions, they are mostly compatible with zero. The reason for the non-zero contribution in weakly populated regions is that additional polynomial orders only lead to overfitting of the events. Therefore, third-order polynomials are chosen to compromise between enough orders to model regions with only a small number of background events and to keep the fits as stable as possible for all ALP mass and lifetime hypotheses. In total, each scan point will be defined by the extended PDF

$$P(M_{\gamma\gamma}; N_{\mathrm{sig}}, N_{\mathrm{bkg}}, x_1, x_2, x_3) = N_{\mathrm{sig}}S(M_{\gamma\gamma}) + N_{\mathrm{bkg}}B(M_{\gamma\gamma}; x_1, x_2, x_3), \tag{5.6}$$

where $N_{\mathrm{sig}}, N_{\mathrm{bkg}}, x_1, x_2$ and x_3 are the free fit parameters. The only non-technical limitation on the fit parameters is that the signal yield is required to be positive. Furthermore, an unbinned negative log-likelihood is used as a loss function. Due to the left tail of the signal $M_{\gamma\gamma}$ distribution, the fit range can span over one or several exclusion limits, resulting in unstable fits in cases where single events beyond the exclusion lead to an insufficient background estimation. This fit behaviour is favoured by the low number of background events and the use of the unbinned likelihood that has no information gain from regions with no events. As a result, the fit region of ALP hypotheses is redefined in dependence of

the double-sided Crystal Ball μ value. If the μ value is between two exclusion regions, the fit region is set to the bounds of the exclusion regions. If the μ value is on the right side of all exclusions in $M_{\gamma\gamma}$, the interpolated $\mathrm{IQR}_{(1,99)} \times 10$ is used for the fit region, where the lower bound is set to the upper η' exclusion bound if it is less than the exclusion bound. To ensure stable fits for this fit region determination, the π^0, η, η' bounds are slightly reduced, as displayed in Table 5.1, in comparison to the previously shown exclusion regions. The number of additional events arising from the irreducible background is still small enough to be considered by the background PDF. In Section 5.3, the full intervals in Table 4.1 are excluded. More stable fits could be achieved by using larger MC samples, where the appearance of single events is less common. Additionally, with larger samples, using a binned likelihood loss function may become more practical. In this case, the loss function considers weakly populated or empty bins in $M_{\gamma\gamma}$.

Table 5.1.: This table displays the excluded $M_{\gamma\gamma}$ ranges for the particles that form the irreducible background in the $B^{\pm} \to K^{\pm}\gamma\gamma$ decay. In contrast to the previously shown exclusion regions, these are slightly smaller to ensure better fit stability.

Particle	Excluded $M_{\gamma\gamma}$ range in GeV/c^2
π^0	[0.100, 0.155]
η	[0.470, 0.610]
η'	[0.930, 9.990]

Each extracted signal yield $\hat{N}_{\mathrm{sig}}$ in the scan of $M_{\gamma\gamma}$ is tested against a zero background hypothesis to determine the significance Z_0 with which the so-called null hypothesis can be rejected. In the case of no significant excess, which corresponds to five sigmas, an upper limit is calculated for the signal + background hypothesis. The limit calculation is explained in Section 5.3. The significance calculation in this thesis uses the test statistic q_0 for discoveries of a positive signal yield [38]. In this case, the test statistic is defined as

$$q_0 = \begin{cases} -2\ln \frac{\mathcal{L}\left(N_{\mathrm{sig}}=0,\hat{\theta}'\right)}{\mathcal{L}\left(N_{\mathrm{sig}}=\hat{N}_{\mathrm{sig}},\hat{\theta}\right)} & \text{for } \hat{N}_{\mathrm{sig}} \geq 0, \\ 0 & \text{for } \hat{N}_{\mathrm{sig}} < 0, \end{cases} \tag{5.7}$$

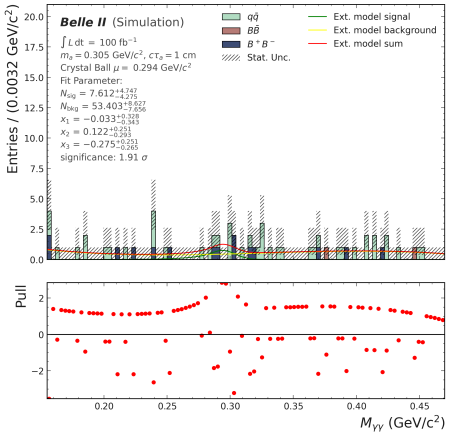
where $\hat{\theta}$ corresponds to the best estimators for the remaining free parameters at the extracted signal yield $\hat{N}_{\mathrm{sig}}$ and similar $\hat{\theta}'$ to best estimators for the hypothesis of $N_{\mathrm{sig}} = 0$. In the asymptotic approximation [38], where the test statistic q_0 follows a χ^2 distribution for one degree of freedom, the p_0 value simply calculates as $p_0 = 1 - \Phi(\sqrt{q_0})$, where Φ is the cumulative Gaussian distribution function. As a result, the significance Z_0 is calculated as

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0}. \tag{5.8}$$

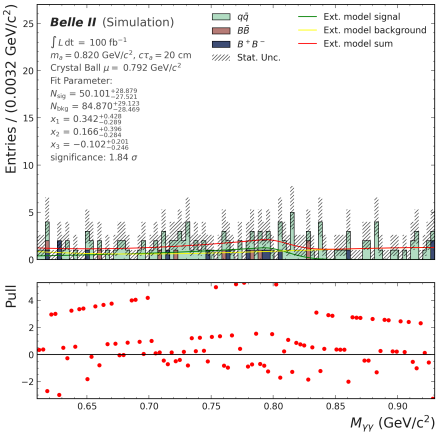
Fig. 5.8 displays examples of the most common fit results with non-zero signal yield. High and low ALP mass regions show very few background events in the case of $100\,\mathrm{fb}^{-1}$ samples,

which results in weakly populated fit regions and a challenging background estimation with polynomials. Fig. 5.8b displays a fit in the region between the η and η' exclusion for a signal hypothesis with a high ALP lifetime. In these cases, upward fluctuations in the significance may arise more often due to the long left tailed signal shape that, apart from the peak in $M_{\gamma\gamma}$, mimics the background structure. Lastly, Fig. 5.8c shows a fit result in between the high and low ALP mass regions where the number of background events allows for good background estimation. Additionally, these signal samples show the highest local significance with 3.01σ over all ALP hypotheses.

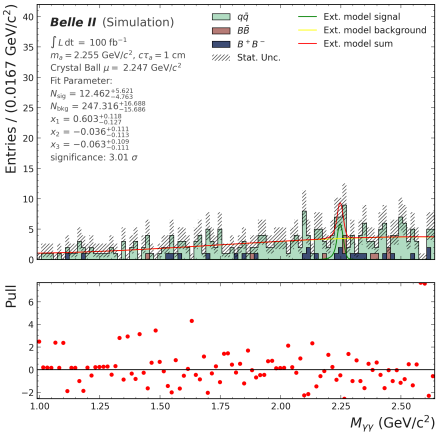
Fig. 5.9 shows the extracted fit results for the example of the $c\tau_a = 20\,\mathrm{cm}$ hypotheses. White gaps between the scan point results indicate that the fit for these scan points failed and, therefore, are not considered in the further analysis. The number of failing fits increases for scans of higher lifetimes but is small enough to be neglected. As discussed for Fig. 5.8b, it is clear to see that for low ALP masses, the fitted signal yield fluctuates upwards, as in Fig. 5.9a, for higher lifetimes due to the wide signal shape. For higher ALP masses, where more background events occur in combination with a broad fit region and a narrow signal shape, the frequency of upwards fluctuations in the signal yield decreases due to the mass constrained fit. Fig. 5.9c displays that the first-order polynomial coefficients strongly fluctuate in between the exclusion regions. However, they are mostly significantly different from zero for ALP masses $m_a \geq 1.01\,\mathrm{GeV}/c^2$. In contrast to that, both higher-order polynomial coefficients are comparable with zero for $m_a \geq 1.01\,\mathrm{GeV}/c^2$, except for the high ALP mass region, where the number of events is very low. Other ALP lifetime scans show comparable fit results with smaller extracted signal yields for lower lifetimes and vice versa.



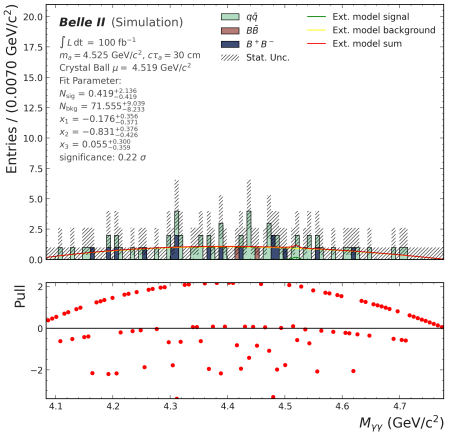
(a) $m_a = 0.305\,\mathrm{GeV}/c^2$, $c\tau_a = 1.0\,\mathrm{cm}$



(b) $m_a = 0.820\,\mathrm{GeV}/c^2$, $c\tau_a = 20.0\,\mathrm{cm}$

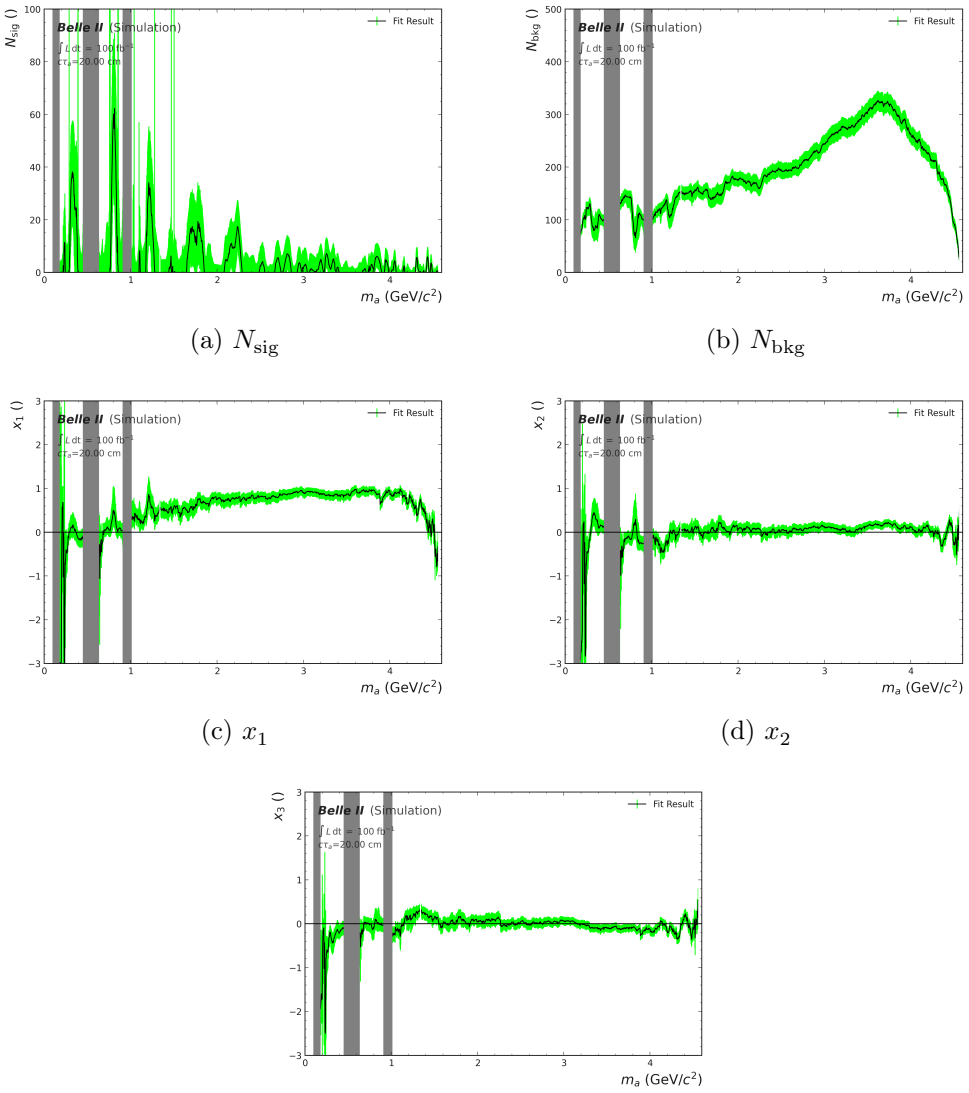


(c) $m_a = 2.255\,\mathrm{GeV}/c^2$, $c\tau_a = 1.0\,\mathrm{cm}$



(d) $m_a = 4.525\,\mathrm{GeV}/c^2$, $c\tau_a = 30.0\,\mathrm{cm}$

Figure 5.8.: This figure displays four different $M_{\gamma\gamma}$ signal + background fits on background MC samples. The signal hypotheses PDFs are described in the text. Plots (a) and (d) illustrate fits in the high and low ALP mass regions where the $M_{\gamma\gamma}$ fit regions are only weakly populated. Additionally, plot (b) shows a fit in between the η and η' exclusion region, where the small fit window leads to the effect that the signal shape of hypotheses with high ALP lifetimes mimics the background distribution. Lastly, plot (c) presents a fit with a large fit region and sufficient background event population for a good background shape estimation.



(a) N_{sig} (b) N_{bkg}

(c) x_1 (d) x_2

(e) x_3

Figure 5.9.: This plot series displays the signal + background fit parameter results for all ALP masses in the $c\tau_a = 20\,\text{cm}$ scan. The fit uncertainties are indicated by green bars. The signal yield in (a) shows that it is mostly overestimated in between the exclusion regions due to very wide signal shape. With some exceptions in the low and high ALP mass regions, the polynomial coefficient fit results display that the second- and third-order coefficients are mostly compatible with zero.

5.3. Sensitivity and Limits

In the case that the scan in $M_{\gamma\gamma}$ on real data does not show any significant excess, as in Section 5.2, an upper limit is calculated on the signal yield N_{sig}. This upper limit indicates the value of N_{sig} for which the signal + background hypothesis still would be rejected for a given CL. Similar to the significance calculation, the likelihood ratio is used to compute the test statistic $q_{N_{\text{sig}}}$ [38] for the upper limit calculation

$$q_{N_{\text{sig}}} = \begin{cases} -2\ln\frac{\mathcal{L}\left(N_{\text{sig}},\hat{\theta}'\right)}{\mathcal{L}\left(N_{\text{sig}}=\hat{N}_{\text{sig}},\hat{\theta}\right)} & \text{for } \hat{N}_{\text{sig}} \leq N_{\text{sig}}, \\ 0 & \text{for } \hat{N}_{\text{sig}} > N_{\text{sig}}. \end{cases} \tag{5.9}$$

The upper limit is the N_{sig} value that solves $p_{N_{\text{sig}}} = 1 - \text{CL}$, where in this thesis the CL is set to 90 %. The p-value is again calculated in the asymptotic approximation $p_{N_{\text{sig}}} = 1 - \Phi\left(\sqrt{q_{N_{\text{sig}}}}\right)$ and interpolated in an array of different N_{sig}. Furthermore, the sensitivity in the upper limit calculation is the expected (median) significance with which the N_{sig} at the CL is rejected in favour of the background-only hypothesis. Practically, [38] proposes in the asymptotic approximation, where $Z_{N_{\text{sig}}} = \sqrt{q_{N_{\text{sig}}}}$, that the median of the test statistic $q_{N_{\text{sig}}}$ can be approximated by its Asimov value $q_{N_{\text{sig}},A}$. The Asimov data used in this case is defined as the evaluation of the underlying PDF in its definition range, i. e. the fit range defined for the fit, such that its best estimators correspond to the estimators of the true fit. Hence, $q_{N_{\text{sig}},A}$ is calculated according to Eq. (5.9), substituting the likelihood $\mathcal{L}$ by the likelihood evaluated for the Asimov data.

For better comparison with similar analyses, the extracted expected and observed upper limits are recalculated to the branching ratios of the signal signature $\text{Br}(B \to Ka) \times \text{Br}(a \to \gamma\gamma)$. The branching fraction for this process is defined as

$$\text{Br}(B \to Ka) \times \text{Br}(a \to \gamma\gamma) = \frac{N_{\text{sig}}}{2N_{B\bar{B}}\text{Br}(\Upsilon(4S) \to B^+B^-)\epsilon_{\text{sig}}}, \tag{5.10}$$

where $N_{B\bar{B}}$ is the number of $B\bar{B}$ events corresponding to the analyzed sample, the $\Upsilon(4S) \to B^+B^-$ branching ratio is 51.4 %, as stated in Section 3.1, and ϵ_{sig} denotes the interpolated signal efficiency for the respective signal hypothesis.

Fig. 5.10 displays the observed upper limits as well as the expected result and its error bands for the scans with an ALP lifetime hypothesis of $c\tau_a = 0.0\,\text{cm}$. The upper limit is distributed uniformly for all scanned ALP mass points, where the ALP mass is approximated by the interpolated double-sided Crystal Ball parameter μ. A linearity test can correct deviations due to this approximation. In the vicinity of the π^0, η, η' exclusion borders, the limits are slightly weaker in comparison to the neighbouring scan points. These weaker limits originate from challenging fits, where the fit region excludes events that are in the signal range. Extending the fit region of signal hypotheses wider than Table 5.1 suggests leads to more irreducible background events in the fit region, which makes the approximation of a polynomial background model insufficient. A possible improvement could be achieved by additionally modelling the irreducible background. Furthermore, for high ALP mass

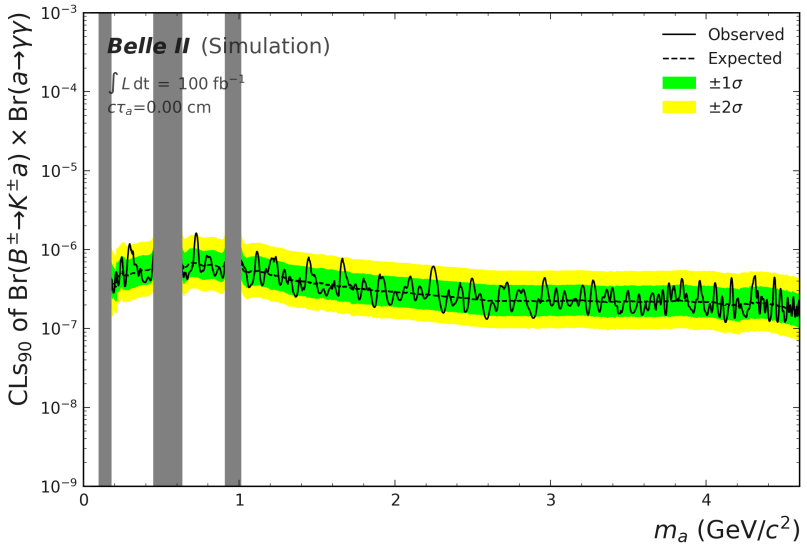


Figure 5.10.: This plot displays the calculated expected and observed CLs_{90} upper limits on the branching ratio $\mathrm{Br}(B \to Ka) \times \mathrm{Br}(a \to \gamma\gamma)$ for all ALP masses with a lifetime of $c\tau_a = 0.0\,\mathrm{cm}$. The $M_{\gamma\gamma}$ scan shows a sensitivity between 3×10^{-7} and 7×10^{-7} for the branching ratio. The observed upper limits fluctuate strongly for high and low ALP masses and are slightly weaker in the vicinity of the π^0, η, η' exclusion regions.

hypotheses, the frequency of observed upper limit fluctuations increases due to narrow signal shapes caused by the mass constrained fit.

Using the example of $c\tau_a = 20.0\,\mathrm{cm}$, Fig. 5.11 shows the calculated upper limit for scans with higher ALP lifetime hypotheses. In comparison to the $c\tau_a = 0.0\,\mathrm{cm}$ $M_{\gamma\gamma}$ scan, the sensitivity is generally lower. Additionally, two significant differences are apparent in the distribution of the limit. Firstly, the sensitivity decreases significantly between excluded irreducible background regions and the right side of the η' exclusion. This decrease could not be resolved in the scope of this thesis and should definitely be subject to further investigations. Comparing the reduction of the sensitivity to Fig. 5.5, which displays the efficiency of the $c\tau_a = 20.0\,\mathrm{cm}$ samples, the same structure can be observed in between the exclusion regions, which makes the signal efficiency anti-correlated to the limit. A possible starting point is to increase the number of signal MC samples in between the exclusions, to understand the efficiency distribution, and to scan $M_{\gamma\gamma}$ on different MC samples with possibly more statistically significant numbers of events. Secondly, also in between the excluded π^0, η, η' regions, the expected upper limit calculations show an approximately oscillating structure. This structure could either be caused by the same effect that is responsible for the first observation, or it could be a limitation of the asymptotic approximation. The latter hypothesis is based on the early stages of the analysis, where an insufficient amount of Asimov data evaluation points caused the same structure in other ALP mass regions. However, the present structure of the expected limit between

the excluded regions could not be resolved by additional Asimov evaluation points. A frequentistic upper limit calculation with the help of toy MC samples should be performed in a future analysis to verify the expected upper limit.

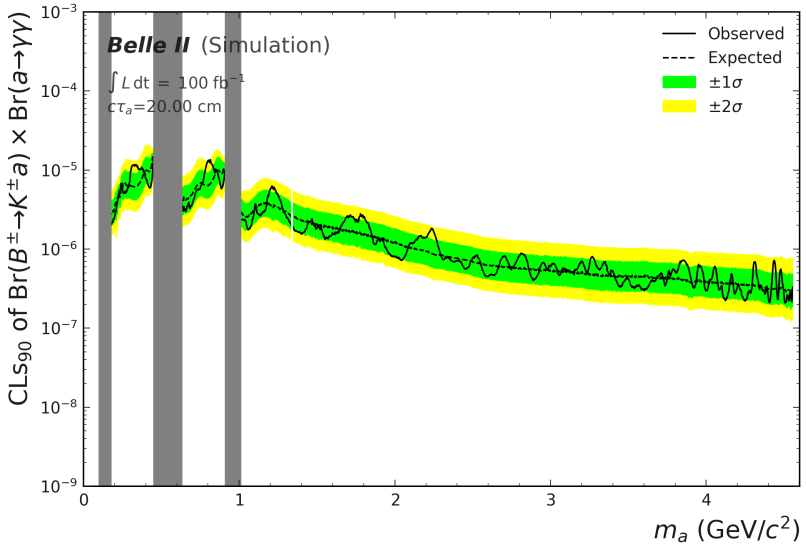


Figure 5.11.: This plot displays the calculated expected and observed CLs_{90} upper limits on the branching ratio $\mathrm{Br}(B \to Ka) \times \mathrm{Br}(a \to \gamma\gamma)$ for all ALP masses with a lifetime of $c\tau_a = 20.0\,\mathrm{cm}$. White gaps in between neighbouring scan points indicate unsuccessful fits. Compared to the $c\tau_a = 0.0\,\mathrm{cm}$ scan, this scan shows an over all weaker limit. Additionally, the upper limit shows for masses $m_a \leq 1.2\,\mathrm{GeV}/c^2$ yet to be explained structures in both the expected and observed upper limit.

Fig. 5.12 summarizes the calculated observed upper limits of $\mathrm{Br}(B \to Ka) \times \mathrm{Br}(a \to \gamma\gamma)$ for all ALP lifetime and mass hypotheses. As previously seen, the upper limit sensitivity decreases with higher lifetime. The distribution of fluctuations for different lifetimes is very similar, indicating that the selection and signal shape interpolations are coherent for both ALP parameters. In addition to the previously discussed observations on the upper limit, Fig. 5.12 displays one more limitation of the upper limit calculation. The scans for very light ALP masses in $c\tau_a = 1.0\,\mathrm{cm}$ and very heavy ALP masses in $c\tau_a = 30.0\,\mathrm{cm}$ and $c\tau_a = 40.0\,\mathrm{cm}$ show very high sensitivity in comparison to neighbouring scan points. All three scan regions have in common that the number of events is very low $N_{\mathrm{events}} \sim \mathcal{O}(10)$ which indicates that the asymptotic approximation for upper limit calculation could be invalid in these cases. Toy MC studies should verify the upper limit calculation in very weakly populated regions in a future analysis. Additionally, in weakly populated areas, a counting experiment approach could be more convenient than the proposed model in Eq. (5.6).

The following steps should be performed to finalize the analysis. Several linearity tests need to be carried out to investigate the fit and interpolation procedure. This subsection

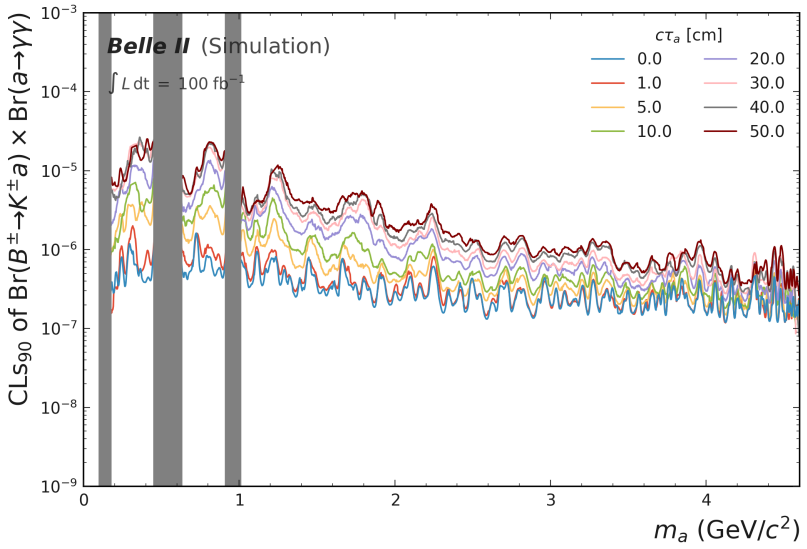


Figure 5.12.: This plot summarizes the observed CLs_{90} upper limits on the branching ratio $\mathrm{Br}(B \to Ka) \times \mathrm{Br}(a \to \gamma\gamma)$ for all scanned ALP masses and lifetimes. The scans indicate upper limits of about 5×10^{-7} for promptly decaying and about 4×10^{-6} to 2×10^{-5} for long-lived ALPs on the branching ratio.

used the interpolated double-sided Crystal Ball μ parameter equivalently to the ALP mass. Considering that the interpolated μ parameter is slightly shifted to lower values in $M_{\gamma\gamma}$ biases the prediction for the ALP mass. A linearity test for the signal yield extraction needs to be carried out concerning the fit quality. Especially the weakly populated regions could show small deviations between sampled and extracted signal yield. Furthermore, more signal MC samples between the π^0, η, η' regions are necessary to study these regions more intensely and improve the quality of the interpolations. The asymptotic approximation, which is used in this analysis, to calculate significances and upper limits, should be tested by toy MC studies specifically in weakly populated regions. Additionally, systematic uncertainties need to be assessed and included into the fit procedure. These uncertainties will consider the signal efficiency and the signal and background shapes. Uncertainties for the signal efficiency, which include the track reconstruction, PID, luminosity and ECL calibration, are provided by the Belle II collaboration. To estimate the uncertainties on the signal and background shape, the $M_{\gamma\gamma}$ fits need to be performed under the assumption of different PDF models and the difference to the here extracted yields calculated. Lastly, the local significance needs to be recalculated regarding the *look-elsewhere effect* [39].

Under consideration of the final steps that need to be performed to conclude the sensitivity study of the $B^\pm \to K^\pm a$, $a \to \gamma\gamma$ signature on Belle II MC samples, the analysis can be compared to [7]. This analysis expanded the long-lived ALP search from the lifetimes $c\tau_a \leq 10.0\,\mathrm{cm}$ to $c\tau_a \leq 50.0\,\mathrm{cm}$. Furthermore, this long-lived ALP search extended the investigated ALP mass from $m_a \leq 2.5\,\mathrm{GeV}/c^2$ to the full sensitive range of $m_a \leq 4.6\,\mathrm{GeV}/c^2$. In direct comparison to the analyzed data corresponding to an integrated luminosity of

$424\,\mathrm{fb}^{-1}$ in [7], this analysis presents a comparable sensitivity on $100\,\mathrm{fb}^{-1}$ MC samples. Extending this analysis on an identical data volume should lead to similar results.

6. Conclusion and Outlook

This thesis presents a sensitivity study in the search for the NP signature $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ considering prompt and long-lived ALPs. The analysis probes signal yields in the $M_{\gamma\gamma}$ distribution in the sensitive mass range $m_a \in [0.175, 4.600]\,\mathrm{GeV}/c^2$, excluding irreducible background regions, for eight different ALP lifetimes $c\tau_a \leq 50\,\mathrm{cm}$. In preparation for an analysis on Belle II data, upper limits on the branching ratio are estimated on Belle II MC samples equivalent to $100\,\mathrm{fb}^{-1}$.

A long-lived ALP in the $B^{\pm} \to K^{\pm}a$, $a \to \gamma\gamma$ decay results in displaced neutral decays with photons in the final state. Intending to understand the final state, this analysis performs an acceptance study for the $a \to \gamma\gamma$ decay and outlines the limitations of the search. For the case of light ALPs, constraints on the acceptance arise from the ALP's high boost. First, promptly decaying ALPs lead to highly boosted photons that can be misreconstructed due to inseparable energy depositions in the ECL clusters. Second, long-lived ALPs predominantly decay outside the detector due to the assumption of an exponential lifetime. Both topics; reconstruction of overlapping ECL clusters and invisibly decaying ALPs, are subject to ongoing investigations. In case of heavy ALPs, an acceptance loss for long-lived ALPs is observed. However, this loss is not fully understood and should be investigated in the next stage of the analysis.

In order to perform a candidate selection, this thesis optimizes a cut-based selection by maximizing the PFOM for a set of observables. The optimization procedure is performed independently for signal MC samples of all ALP mass and lifetime parameters. Final selection values can be smoothly interpolated between the simulated ALP mass points for fixed ALP lifetimes to perform the scans in $M_{\gamma\gamma}$. The application of the PFOM algorithm results in a signal efficiency of 10 % to 20 % for promptly decaying ALPs and up to 10 % for ALPs with a lifetime of $c\tau_a = 50\,\mathrm{cm}$. Simultaneously, the number of background events is reduced to less than 50 events for the prompt case and below 300 events for the highly displaced samples. Extensions to this analysis should be based on more signal MC samples for mass regions between the π^0, η, η' exclusion regions due to their challenging interpolation. Apart from more signal samples, improvements could be achieved by applying an MVA-based continuum suppression. Additionally, [37] proposes an MVA-based approach to the PFOM optimization that could simplify the procedure by a well-performing generalisation to untrained mass hypotheses.

To build a variety of signal ALP hypotheses in the sensitive $M_{\gamma\gamma}$ range, the candidate selection and the expected signal efficiency and shape are interpolated for masses without

simulated MC samples. Subsequently, for each ALP lifetime, a scan of $M_{\gamma\gamma}$ is performed in steps of $2.5\,\mathrm{MeV}/c^2$. The signal is modelled by an extended double-sided Crystal Ball distribution with fixed parameters, while an extended third-order Chebyshev polynomial model is chosen to approximate the background. In order to extract the signal yield, the sum of the extended PDFs is fitted with the unbinned maximum likelihood method on Belle II continuum and generic MC samples equivalent to an integrated luminosity of $100\,\mathrm{fb}^{-1}$. Additionally, an expected and observed upper limit is calculated in the asymptotic approximation for each scan point, which concludes this thesis.

To finalize the analysis, the fit procedure needs to be validated by linearity tests, and the asymptotic limit calculation needs to be approved by a frequentistic approach. Furthermore, fits in weakly populated regions in $M_{\gamma\gamma}$ could be better investigated with a counting experiment approach. To be comparable to similar analyses, systematic uncertainties need to be assessed and the *look-elsewhere effect* has to be considered. Additionally, leptonic background samples $e^+e^- \to l^+l^-$ should be included and the $M_{\gamma\gamma}$ scan performed on the $500\,\mathrm{fb}^{-1}$ MC samples.

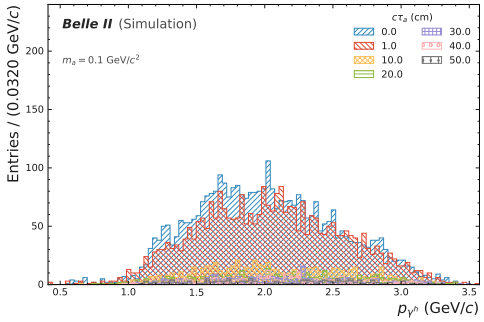
Lastly, keeping the finalizing steps in mind, the thesis results can be compared to the search performed on the BABAR data set of $424\,\mathrm{fb}^{-1}$ [7]. Considering the difference in the used MC sample statistics, the approach developed in this thesis achieves a comparable result, which will lead to a similar sensitivity in the case of an equally large data set. Additionally, this thesis expands the search for long-lived ALPs from $m_a \leq 2.5\,\mathrm{GeV}/c^2$ to the full sensitive range of $m_a \leq 4.6\,\mathrm{GeV}/c^2$ and for the ALP lifetime from $c\tau_a \leq 10.0\,\mathrm{cm}$ to $c\tau_a \leq 50.0\,\mathrm{cm}$. The presented analysis estimates CLs_{90} upper limits on the branching ratio $\mathrm{Br}(B \to Ka) \times \mathrm{Br}(a \to \gamma\gamma)$ in the range of 5×10^{-7} for promptly decaying and about 4×10^{-6} to 2×10^{-5} for long-lived ALPs.

A. List of Figures

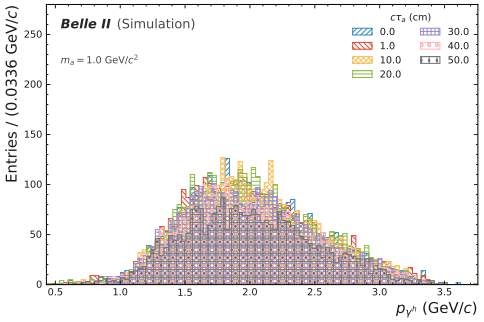
B. List of Tables

C. Appendix

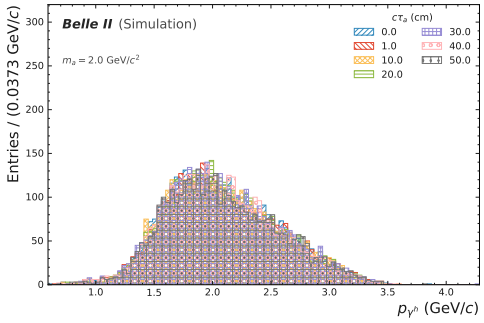
C.1. Reconstructed Momenta for the Finale State Particles



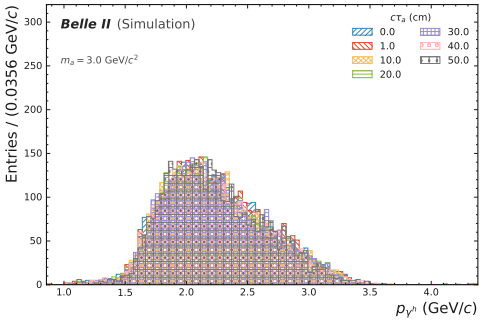
(a) $m_a = 0.1\,\mathrm{GeV}/c^2$



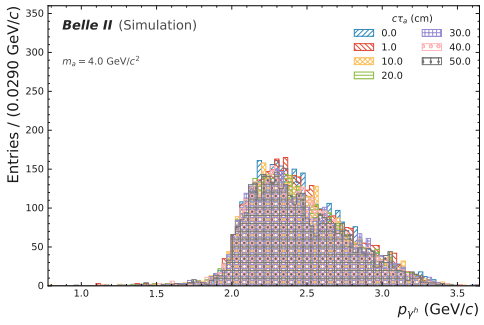
(b) $m_a = 1.0\,\mathrm{GeV}/c^2$



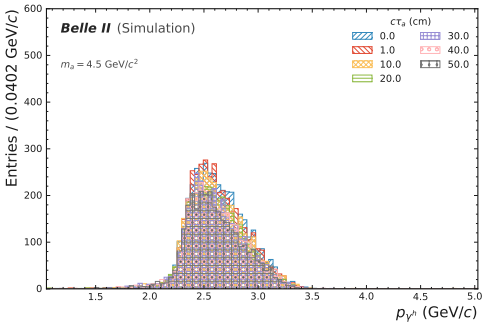
(c) $m_a = 2.0\,\mathrm{GeV}/c^2$



(d) $m_a = 3.0\,\mathrm{GeV}/c^2$

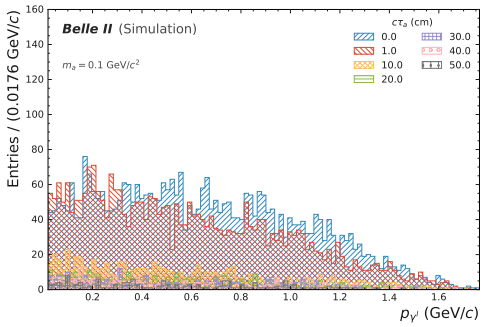


(e) $m_a = 4.0\,\mathrm{GeV}/c^2$

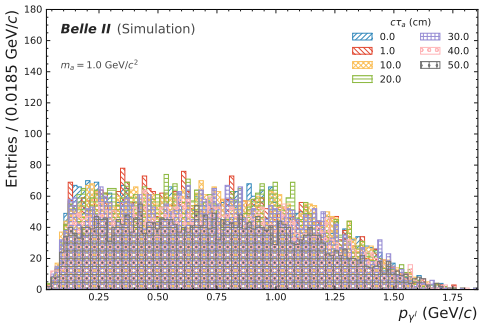


(f) $m_a = 4.5\,\mathrm{GeV}/c^2$

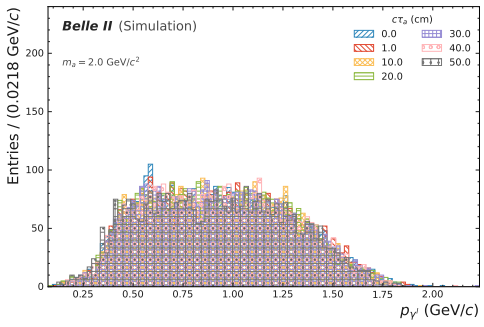
Figure C.1.: This plot series displays the reconstructed momenta of γ_h in correctly reconstructed $a \to \gamma\gamma$ events in the signal samples.



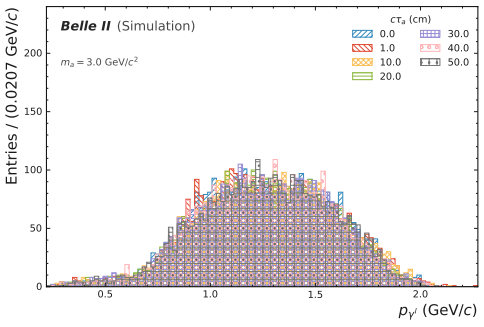
(a) $m_a = 0.1\,\mathrm{GeV}/c^2$



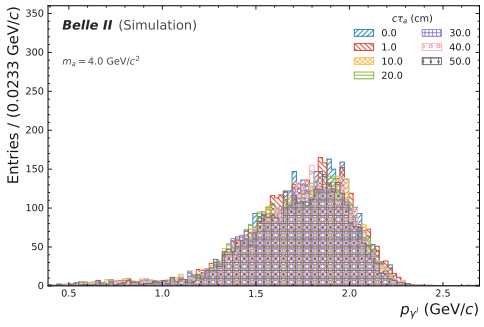
(b) $m_a = 1.0\,\mathrm{GeV}/c^2$



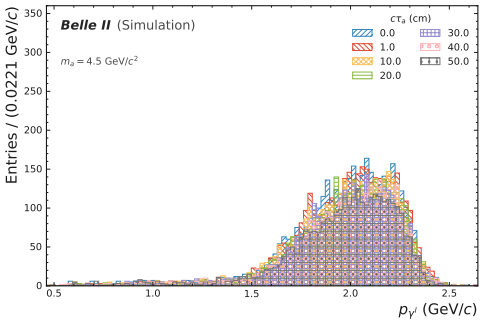
(c) $m_a = 2.0\,\mathrm{GeV}/c^2$



(d) $m_a = 3.0\,\mathrm{GeV}/c^2$

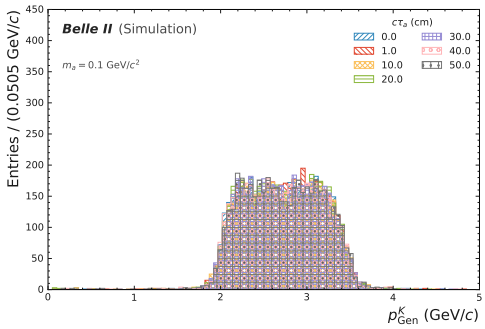


(e) $m_a = 4.0\,\mathrm{GeV}/c^2$

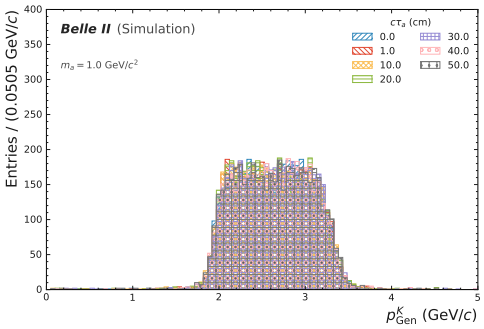


(f) $m_a = 4.5\,\mathrm{GeV}/c^2$

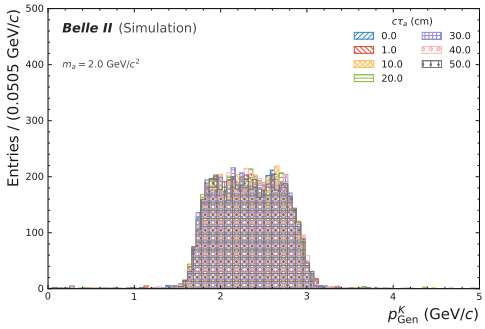
Figure C.2.: This plot series displays the reconstructed momenta of γ_l in correctly reconstructed $a \to \gamma\gamma$ events in the signal samples.



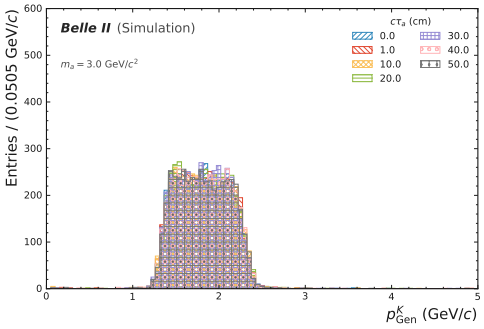
(a) $m_a = 0.1\,\text{GeV}/c^2$



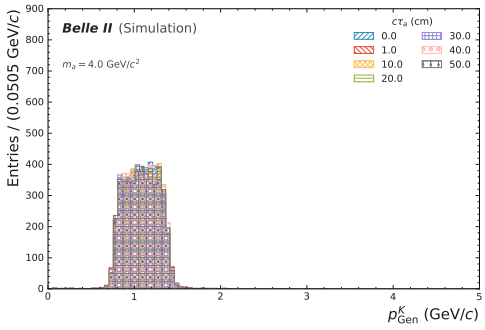
(b) $m_a = 1.0\,\text{GeV}/c^2$



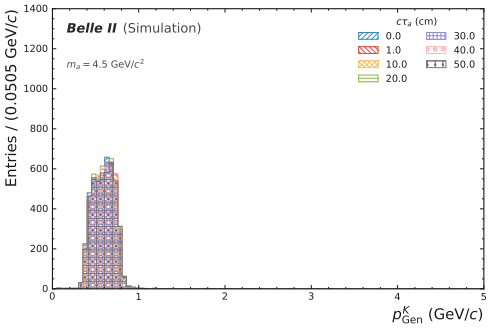
(c) $m_a = 2.0\,\text{GeV}/c^2$



(d) $m_a = 3.0\,\text{GeV}/c^2$



(e) $m_a = 4.0\,\text{GeV}/c^2$



(f) $m_a = 4.5\,\text{GeV}/c^2$

Figure C.3.: This plot series displays the reconstructed momenta of correctly reconstructed kaons in the signal samples.

Bibliography

[1] R. Oerter, *The theory of almost everything: The standard model, the unsung triumph of modern physics*. 2006.

[2] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08001–S08001. https://doi.org/10.1088/1748-0221/3/08/s08001.

[3] "Cern accelerating science." https://fcc-cdr.web.cern.ch/.

[4] T. Abe *et al.*, "Belle ii technical design report," 2010. https://arxiv.org/abs/1011.0352.

[5] G. Lanfranchi, M. Pospelov, and P. Schuster, "The search for feebly interacting particles," *Annual Review of Nuclear and Particle Science* **71** no. 1, (Sep, 2021) 279–313. https://doi.org/10.1146%2Fannurev-nucl-102419-055056.

[6] E. Izaguirre, T. Lin, and B. Shuve, "Searching for axionlike particles in flavor-changing neutral current processes," *Physical Review Letters* **118** no. 11, (Mar, 2017) . https://doi.org/10.1103%2Fphysrevlett.118.111802.

[7] The BABAR Collaboration, "Search for an axion-like particle in b meson decays," 2021. https://arxiv.org/abs/2111.01800.

[8] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Westview Press, 1995. Reading, USA: Addison-Wesley (1995) 842 p.

[9] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[10] S. Weinberg, "A new light boson?," *Phys. Rev. Lett.* **40** (Jan, 1978) 223–226. https://link.aps.org/doi/10.1103/PhysRevLett.40.223.

[11] C. Abel *et al.*, "Measurement of the permanent electric dipole moment of the neutron," *Phys. Rev. Lett.* **124** (Feb, 2020) 081803. https://link.aps.org/doi/10.1103/PhysRevLett.124.081803.

[12] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of pseudoparticles," *Phys. Rev. Lett.* **38** (Jun, 1977) 1440–1443. https://link.aps.org/doi/10.1103/PhysRevLett.38.1440.

[13] F. Wilczek, "Problem of strong p and t invariance in the presence of instantons," *Phys. Rev. Lett.* **40** (Jan, 1978) 279–282. https://link.aps.org/doi/10.1103/PhysRevLett.40.279.

[14] J. Jaeckel and A. Ringwald, "The low-energy frontier of particle physics," *Annual Review of Nuclear and Particle Science* **60** no. 1, (2010) 405–437. https://doi.org/10.1146/annurev.nucl.012809.104433.

[15] M. J. Dolan, T. Ferber, C. Hearty, F. Kahlhoefer, and K. Schmidt-Hoberg, "Revised constraints and belle II sensitivity for visible and invisible axion-like particles," *Journal of High Energy Physics* **2017** no. 12, (Dec, 2017) . https://doi.org/10.1007%2Fjhep12%282017%29094.

[16] M. J. Dolan, F. Kahlhoefer, C. McCabe, and K. Schmidt-Hoberg, "A taste of dark matter: flavour constraints on pseudoscalar mediators," *Journal of High Energy Physics* **2015** no. 3, (Mar, 2015) . https://doi.org/10.1007%2Fjhep03%282015%29171.

[17] K. Choi, S. H. Im, C. B. Park, and S. Yun, "Minimal flavor violation with axion-like particles," *Journal of High Energy Physics* **2017** no. 11, (Nov, 2017) . https://doi.org/10.1007%2Fjhep11%282017%29070.

[18] K. Akai, K. Furukawa, and H. Koiso, "SuperKEKB collider," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **907** (Nov, 2018) 188–199. https://doi.org/10.1016%2Fj.nima.2018.08.017.

[19] **Particle Data Group**, P. Zyla *et al.*, "Review of Particle Physics," *PTEP* **2020** no. 8, (2020) 083C01. and 2021 update.

[20] I. Heredia de la Cruz, "The belle ii experiment: fundamental physics at the flavor frontier," *Journal of Physics: Conference Series* **761** (09, 2016) .

[21] D. Matvienko, "The belle ii experiment: status and physics program," *EPJ Web of Conferences* **191** (01, 2018) 02010.

[22] H. Ikeda, *Development of the CsI(Tl) Calorimeter for the Measurement of CP Violation at KEK B-Factroy*. PhD thesis, Nara Women's University, 1999.

[23] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, "The belle ii core software," *Computing and Software for Big Science* **3** no. 1, (Nov, 2018) 1. https://doi.org/10.1007/s41781-018-0017-9.

[24] J. Eschle, A. Puig Navarro, R. Silva Coutinho, and N. Serra, "zfit: Scalable pythonic fitting," *SoftwareX* **11** (2020) 100508. https://www.sciencedirect.com/science/article/pii/S2352711019303851.

[25] T. Developers, "Tensorflow," Feb., 2022. https://doi.org/10.5281/zenodo.5949169. Specific TensorFlow versions can be found in the "Versions" list on the right side of this page.
See the full list of authors on GitHub.

[26] M. Marinangeli, "Api reference." https://scikit-hep.org/hepstats/api/index.html.

[27] B. I. Collaboration, "Belle ii software release-05-02-18." https://software.belle2.org/release-05-02-18/index.html.

[28] A. Ryd, D. Lange, N. Kuznetsova, S. Versille, M. Rotondo, D. P. Kirkby, F. K. Wuerthwein, and A. Ishikawa, "EvtGen: A Monte Carlo Generator for B-Physics,".

[29] S. Jadach, B. Ward, and Z. Wacs, "The precision monte carlo event generator for two-fermion final states in collisions," *Computer Physics Communications* **130** no. 3, (Aug, 2000) 260–325. https://doi.org/10.1016%2Fs0010-4655%2800%2900048-5.

[30] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Computer Physics Communications* **191** (Jun, 2015) 159–177. https://doi.org/10.1016%2Fj.cpc.2015.01.024.

[31] S. Agostinelli *et al.*, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250–303. https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[32] T. Ferber, A. Filimonova, R. Schäfer, and S. Westhoff, "Displaced or invisible? alps from b decays at belle ii," 2022. https://arxiv.org/abs/2201.06580.

[33] G. Punzi, "Sensitivity of searches for new signals and its optimization," 2003. https://arxiv.org/abs/physics/0308063.

[34] M. Röhrken, "Time-dependent cp violation measurements : Analyses of neutral b meson to double-charm decays at the japanese belle experiment," 2014. https://swbplus.bsz-bw.de/bsz39581362xcov.jpghttps://doi.org/10.1007/978-3-319-00726-7.

[35] J.-F. Krohn *et al.*, "Global decay chain vertex fitting at belle ii," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **976** (2020) 164269. https://www.sciencedirect.com/science/article/pii/S0168900220306653.

[36] L. Sowa, "Deep Continuum Suppression with Predictive Uncertainties at the Belle II Experiment," Master's thesis, Karlsruhe Institute of Technology (KIT), 2021.

[37] F. Abudinén *et al.*, "Punzi-loss:," *The European Physical Journal C* **82** no. 2, (Feb, 2022) . https://doi.org/10.1140%2Fepjc%2Fs10052-022-10070-0.

[38] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C* **71** no. 2, (Feb, 2011) . https://doi.org/10.1140%2Fepjc%2Fs10052-011-1554-0.

[39] E. Gross and O. Vitells, "Trial factors for the look elsewhere effect in high energy physics," *The European Physical Journal C* **70** no. 1-2, (Oct, 2010) 525–530. https://doi.org/10.1140%2Fepjc%2Fs10052-010-1470-8.