Belle II preprint: 2025-017
KEK preprint: 2025-16

Prepared for submission to JHEP

Search for an Axion-Like Particle in $B \to K^{(*)} a(\to \gamma\gamma)$ Decays at Belle

The Belle and Belle II Collaborations

I. Adachi, L. Aggarwal, H. Ahmed, Y. Ahn, H. Aihara, N. Akopov, S. Alghamdi, M. Alhakami, A. Aloisio, N. Althubiti, K. Amos, M. Angelsmark, N. Anh Ky, C. Antonioli, D. M. Asner, H. Atmacan, T. Aushev, V. Aushev, M. Aversano, R. Ayad, V. Babu, H. Bae, N. K. Baghel, S. Bahinipati, P. Bambade, Sw. Banerjee, S. Bansal, M. Barrett, M. Bartl, J. Baudot, A. Baur, A. Beaubien, F. Becherer, J. Becker, J. V. Bennett, F. U. Bernlochner, V. Bertacchi, M. Bertemes, E. Bertholet, M. Bessner, S. Bettarini, B. Bhuyan, F. Bianchi, T. Bilka, D. Biswas, A. Bobrov, D. Bodrov, A. Bondar, G. Bonvicini, J. Borah, A. Boschetti, A. Bozek, M. Bračko, P. Branchini, T. E. Browder, A. Budano, S. Bussino, Q. Campagna, M. Campajola, L. Cao, G. Casarosa, C. Cecchi, M.-C. Chang, P. Cheema, L. Chen, B. G. Cheon, K. Chilikin, J. Chin, K. Chirapatpimol, H.-E. Cho, K. Cho, S.-J. Cho, S.-K. Choi, S. Choudhury, J. Cochran, I. Consigny, L. Corona, J. X. Cui, E. De La Cruz-Burelo, S. A. De La Motte, G. De Nardo, G. De Pietro, R. de Sangro, M. Destefanis, S. Dey, A. Di Canto, J. Dingfelder, Z. Doležal, I. Domínguez Jiménez, T. V. Dong, K. Dugic, G. Dujany, P. Ecker, J. Eppelt, R. Farkas, P. Feichtinger, T. Ferber, T. Fillinger, C. Finck, G. Finocchiaro, A. Fodor, F. Forti, A. Frey, B. G. Fulsom, A. Gabrielli, A. Gale, M. Garcia-Hernandez, R. Garg, G. Gaudino, V. Gaur, V. Gautam, A. Gaz, A. Gellrich, G. Ghevondyan, D. Ghosh, H. Ghumaryan, G. Giakoustidis, R. Giordano, A. Giri, P. Gironella Gironell, B. Gobbo, R. Godang, O. Gogota, P. Goldenzweig, E. Graziani, D. Greenwald, Z. Gruberová, Y. Guan, K. Gudkova, I. Haide, Y. Han, T. Hara, K. Hayasaka, H. Hayashii, S. Hazra, C. Hearty, M. T. Hedges, G. Heine, I. Heredia de la Cruz, M. Hernández Villanueva, T. Higuchi, M. Hoek, M. Hohmann, R. Hoppe, P. Horak, C.-L. Hsu, T. Humair, T. Iijima, K. Inami, N. Ipsita, A. Ishikawa, R. Itoh, M. Iwasaki, P. Jackson, W. W. Jacobs, E.-J. Jang, Q. P. Ji, S. Jia, Y. Jin,

A. Johnson, K. K. Joo, H. Junkerkalefeld, J. Kandra, K. H. Kang,
G. Karyan, T. Kawasaki, F. Keil, C. Ketter, M. Khan, C. Kiesling,
C. Kim, C.-H. Kim, D. Y. Kim, J.-Y. Kim, K.-H. Kim, Y. J. Kim,
Y.-K. Kim, H. Kindo, K. Kinoshita, P. Kodyš, T. Koga, S. Kohani,
K. Kojima, A. Korobov, S. Korpar, E. Kovalenko, P. Križan, P. Krokovny,
T. Kuhr, Y. Kulii, D. Kumar, R. Kumar, K. Kumara, T. Kunigo,
A. Kuzmin, Y.-J. Kwon, S. Lacaprara, K. Lalwani, T. Lam, J. S. Lange,
T. S. Lau, M. Laurenza, R. Leboucher, F. R. Le Diberder, M. J. Lee,
C. Lemettais, P. Leo, P. M. Lewis, C. Li, H.-J. Li, L. K. Li, Q. M. Li,
W. Z. Li, Y. Li, Y. B. Li, Y. P. Liao, J. Libby, J. Lin, S. Lin, M. H. Liu,
Q. Y. Liu, Y. Liu, Z. Q. Liu, D. Liventsev, S. Longo, T. Lueck, C. Lyu,
Y. Ma, C. Madaan, M. Maggiora, S. P. Maharana, R. Maiti,
G. Mancinelli, R. Manfredi, E. Manoni, M. Mantovano, D. Marcantonio,
S. Marcello, C. Marinas, C. Martellini, A. Martens, T. Martinov,
L. Massaccesi, M. Masuda, D. Matvienko, S. K. Maurya, M. Maushart,
J. A. McKenna, R. Mehta, F. Meier, D. Meleshko, M. Merola, C. Miller,
M. Mirra, S. Mitra, K. Miyabayashi, H. Miyake, R. Mizuk,
G. B. Mohanty, S. Mondal, S. Moneta, A. L. Moreira de Carvalho,
H.-G. Moser, R. Mussa, I. Nakamura, M. Nakao, Y. Nakazawa,
M. Naruki, Z. Natkaniec, A. Natochii, M. Nayak, M. Neu, M. Niiyama,
S. Nishida, S. Ogawa, H. Ono, G. Pakhlova, S. Pardi, K. Parham,
H. Park, J. Park, K. Park, S.-H. Park, B. Paschen, A. Passeri, S. Patra,
S. Paul, T. K. Pedlar, I. Peruzzi, R. Peschke, R. Pestotnik, M. Piccolo,
L. E. Piilonen, P. L. M. Podesta-Lerma, T. Podobnik, S. Pokharel,
A. Prakash, C. Praz, S. Prell, E. Prencipe, M. T. Prim, S. Privalov,
H. Purwar, P. Rados, G. Raeuber, S. Raiz, K. Ravindran, J. U. Rehman,
M. Reif, S. Reiter, M. Remnev, L. Reuter, D. Ricalde Herrmann,
I. Ripp-Baudot, G. Rizzo, S. H. Robertson, J. M. Roney, A. Rostomyan,
N. Rout, L. Salutari, D. A. Sanders, S. Sandilya, L. Santelj, V. Savinov,
B. Scavino, J. Schmitz, S. Schneider, G. Schnell, M. Schnepf,
C. Schwanda, Y. Seino, A. Selce, K. Senyo, J. Serrano, M. E. Sevior,
C. Sfienti, W. Shan, G. Sharma, C. P. Shen, X. D. Shi, T. Shillington,
T. Shimasaki, J.-G. Shiu, D. Shtol, A. Sibidanov, F. Simon, J. B. Singh,
J. Skorupa, M. Sobotzik, A. Soffer, A. Sokolov, E. Solovieva, S. Spataro,
B. Spruck, M. Starič, P. Stavroulakis, S. Stefkova, R. Stroili, Y. Sue,
M. Sumihama, K. Sumisawa, N. Suwonjandee, H. Svidras, M. Takizawa,
U. Tamponi, K. Tanida, F. Tenchini, F. Testa, A. Thaller, O. Tittel,
R. Tiwary, E. Torassa, K. Trabelsi, F. F. Trantou, I. Tsaklidis, I. Ueda,
T. Uglov, K. Unger, Y. Unno, K. Uno, S. Uno, P. Urquijo, Y. Ushiroda,
S. E. Vahsen, R. van Tonder, K. E. Varvell, M. Veronesi, A. Vinokurova,
V. S. Vismaya, L. Vitale, V. Vobbilisetti, R. Volpe, A. Vossen, E. Waheed,
M. Wakai, S. Wallner, M.-Z. Wang, A. Warburton, S. Watanuki,
C. Wessel, E. Won, X. P. Xu, B. D. Yabsley, S. Yamada, W. Yan,

W. C. Yan, S. B. Yang, J. Yelton, K. Yi, J. H. Yin, K. Yoshihara, C. Z. Yuan, J. Yuan, L. Zani, F. Zeng, M. Zeyrek, B. Zhang, V. Zhilich, J. S. Zhou, Q. D. Zhou, L. Zhu, R. Žlebčík

ABSTRACT:

We report a search for an axion-like particle a in $B \to K^{(*)}a$ decays using data collected with the Belle detector at the KEKB asymmetric energy electron-positron collider. The search is based on a $711\,\mathrm{fb}^{-1}$ data sample collected at the $\Upsilon(4S)$ resonance energy, corresponding to a sample of 772×10^6 $\Upsilon(4S)$ events. In this study, we search for the decay of the axion-like particle into a pair of photons, $a \to \gamma\gamma$. We scan the two-photon invariant mass in the range $0.16\,\mathrm{GeV} - 4.50\,\mathrm{GeV}$ for the K modes and $0.16\,\mathrm{GeV} - 4.20\,\mathrm{GeV}$ for the K^* modes. No significant signal is observed in any of the modes, and 90% confidence level upper limits are established on the coupling to the W boson, g_{aW}, as a function of a mass. The limits range from $3 \times 10^{-6}\,\mathrm{GeV}^{-1}$ to $3 \times 10^{-5}\,\mathrm{GeV}^{-1}$, improving the current constraints on g_{aW} by a factor of two over the most stringent previous experimental results.

Contents

1 Introduction

The Peccei-Quinn (PQ) theory introduces axions as a solution to the strong-CP problem, positioning them as promising candidates for dark matter in extensions to the standard model (SM) [1–3]. Other extensions to the SM introduce axion-like particles (ALPs) a, which share the quantum numbers of axions, but, unlike axions, have uncorrelated masses and couplings. While the PQ axion is expected to have a mass below $\mathcal{O}(1\,\text{MeV})$ inversely proportional to its decay constant f_a ($m_a \approx 6.3\,\text{eV} \cdot 10^6\ \text{GeV}/f_a$), ALPs encompass a broader mass spectrum. ALPs have the potential to address fundamental problems [4–6]. In particular, they can act as dark matter mediators through the axion portal at the $\mathcal{O}(\text{ GeV})$ scale [7].

Recent years have seen a surge of interest in ALPs between the MeV and GeV mass range [8–17]. The ALP couplings to photons, leptons, and gluons have been extensively studied by collider and beam dump experiments [18–24]. In contrast, the coupling to $W^\pm$ bosons remains relatively uncharted [25]. The most recent study of ALP coupling to $W^\pm$ bosons was conducted by the BaBar experiment using the $B^+ \to K^+ a(\to \gamma\gamma)$ decay [26]. In this paper, we work in natural units ($\hbar = c = 1$), and charge-conjugate modes are implicitly included. Although the ALP is assumed to primarily couple to the W boson, gauge-boson mixing induces a coupling to photons, resulting in a nearly 100% branching fraction for the $a \to \gamma\gamma$ decay in the analysed mass range, $m_a < M_W$.

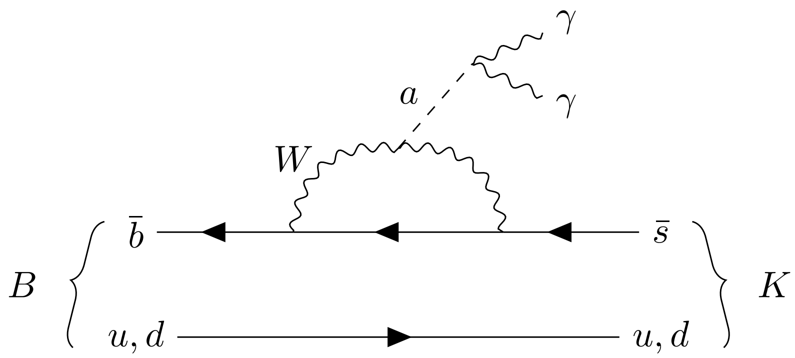

Figure 1: The Feynman diagram for a $B \to K^{(*)} a(\to \gamma\gamma)$ decay.

The coupling of the ALP to $W^\pm$ bosons is described by the Lagrangian [25],

$$\mathcal{L} = -\frac{g_{aW}}{4} a W_{\mu\nu} \tilde{W}^{\mu\nu}, \tag{1.1}$$

where a is the ALP field, g_{aW} is the coupling strength of the a to W bosons, and $W_{\mu\nu}$ is the gauge boson field strength for a W boson, with dual tensor $\tilde{W}^{\mu\nu} = \epsilon^{\mu\nu\rho\sigma} V_{\rho\sigma}/2$. The branching fraction of the process $B \to K^{(*)} a(\to \gamma\gamma)$, which is shown in Fig. 1, depends quadratically on g_{aW} (see the details in the Appendix A), reflecting the fact that the ALP decay width is given by $\Gamma = g_{aW}^2 m_a^3 \sin^4\theta_W/64\pi$, where θ_W is the weak mixing angle [26].

In this paper, we report a search for ALPs in $B \to K^{(*)} a(\to \gamma\gamma)$ decays using four kaon modes, K_S^0, K^+, K^{*0}, and K^{*+}. The data were collected with the Belle detector [27] at the energy-asymmetric e^+e^- KEKB collider [28] at a centre-of-mass energy of 10.58 GeV. This search is based on a data set of 772 ± 11 million $\Upsilon(4S)$ mesons, corresponding to an integrated luminosity of 711 fb^{-1}. Our study has better sensitivity than the previous BaBar study, leveraging the higher total integrated luminosity of the Belle experiment, as well as additional kaon modes. The ALP mass hypotheses for the K^+ and K_S^0 modes range from 0.16 GeV to 4.50 GeV, while for the K^{*+} and K^{*0} modes the range is 0.16 GeV to 4.20 GeV. This analysis also probes regions of the parameter space in which the lifetime of the ALP is not negligible. For decays that occur at a displaced vertex, the photon energy and the ALP mass resolution are affected, reducing the signal efficiency. Once we obtain the results separately for each mode, we combine the four kaon modes using a simultaneous fit to improve the constraint on g_{aW}.

2 The Belle detector and simulation

The Belle detector is a general purpose detector, described in detail in [27]. It has a cylindrical symmetry around the beam line, with the z-axis being defined as the direction opposite to the positron beam. The detector consists of six subdetectors: a silicon vertex detector (SVD) for precise vertex determination, a central drift chamber (CDC) for reconstruction of charged particle trajectories (tracks) and for measuring their momentum, an aerogel Cherenkov counter (ACC) and a time-of-flight scintillation counter (TOF) for particle identification, an electromagnetic calorimeter (ECL) for photon detection, photon energy measurement, and electron identification, surrounded by a 1.5 T superconducting solenoid, and resistive plate chambers installed in the flux-return yoke to detect K_L^0 and μ (KLM). The ECL, which is crucial for this analysis, consists of 8736 CsI(Tl) crystals with a nearly projective geometry covering the polar angle range of $12° < \theta < 157°$.

To mitigate possible biases in the analysis, we establish the event selection and the search method using simulation, and we validate these with control modes and off-resonance data before the experimental data are examined. The signal $B \to K^{(*)}a(\to \gamma\gamma)$ processes are simulated using the EvtGen generator [29]. We generate 60 samples for K modes, with a masses ranging from 0.16 GeV to 4.50 GeV, and 56 samples for K^* modes, with a masses ranging from 0.16 GeV to 4.20 GeV. The long-lived a sample is simulated with the ALP lifetime $c\tau$ varying from 10 mm to 500 mm, where τ is the proper lifetime. In addition, we simulate a sample with uniformly distributed diphoton invariant mass in the entire mass range from 0.01 GeV to 4.78 GeV for K modes, and 0.01 GeV to 4.28 GeV for K^* modes with 10 MeV intervals. We simulate the background processes $\Upsilon(4S) \to B\overline{B}$ and $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) using the EvtGen [29] and PYTHIA [30] generators. Final-state radiation is simulated using PHOTOS [31]. The detector response is simulated with GEANT3 [32]. Both experimental data and simulated events are converted to the Belle II format using B2BII [33] and then analysed using the Belle II analysis software framework [34, 35].

3 Selection of signal events

The charged particles in $B \to K^{(*)}a(\to \gamma\gamma)$ decays, which originate from one of the four kaon modes K^+, K^{*0}, K^{*+}, and K_S^0, are required to have a point of closest approach to the interaction point (IP) of less than 4 cm in the z-direction and less than 3 cm in the radial direction. The charged kaons (K^+) are identified using a likelihood ratio $\mathcal{P}$, which compares two particle hypothesis i and j, $\mathcal{P}(i : j) = \mathcal{L}_i/(\mathcal{L}_i + \mathcal{L}_j)$. The likelihoods $\mathcal{L}$ are calculated based on the Cherenkov photon yield in the ACC, the energy-loss measurements in the CDC, and the time-of-flight information from the TOF. To select K^+, we require $\mathcal{P}(K : \pi) > 0.6$ and $\mathcal{P}(K : p) > 0.4$. The K^* candidates are reconstructed from the decay modes $K^{*0} \to K^+\pi^-$ and $K^{*+} \to K_S^0\pi^+$ within the invariant mass range of 0.8 GeV to 1.0 GeV. To select charged pions $\pi^{\pm}$ we require $\mathcal{P}(\pi : K) > 0.4$ and $\mathcal{P}(\pi : p) > 0.7$, with the likelihood ratio definitions being the same as for the charged kaon. The identification efficiencies for pions and kaons are 96.9% and 85.3%, respectively, with misidentification rates for pions as kaons and vice versa of 10.3% and 1.3%, respectively, while the misidentification rates for protons are negligible. The $K_S^0 \to \pi^+\pi^-$ candidates are selected based on an artificial neural network ranking [36], applied to two oppositely charged particles that both have a pion mass hypothesis. The dipion invariant mass is required to be within a range of 20 MeV around the nominal mass of the K_S^0 [37]. This range corresponds to a mass range of about 5σ, where σ is the K_S^0 mass resolution. The neural network is trained with 13 variables. Detailed information is provided in Ref. [38]. Particles with high lepton probability are excluded from our analysis. For the electron identification, we primarily use the information from the ECL along with other subdetectors. The electron likelihood ratio is defined as $\mathcal{P}(e) = \mathcal{L}_e/(\mathcal{L}_e + \mathcal{L}_{\bar{e}})$, where $\mathcal{L}_e$ is the electron likelihood and $\mathcal{L}_{\bar{e}}$ is the non-electron likelihood, a product of likelihoods from the ACC, CDC, and ECL. For muon identification, the likelihood is calculated using the information from the KLM. The likelihood ratio is defined as $\mathcal{P}(\mu) = \mathcal{L}_\mu/(\mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K)$, where the value $\mathcal{L}_\mu$ is determined based on whether the charged particle has an associated KLM signature. We require $\mathcal{P}(e) < 0.9$ and $\mathcal{P}(\mu) < 0.9$ to veto leptons.

The ALP candidates are reconstructed from the $a \to \gamma\gamma$ decay. Photons are identified from ECL energy deposits that are not associated with reconstructed charged particles. To suppress the contribution from photons that originate from beam background, each photon candidate is required to have a minimum energy that depends on the ECL region: $E_\gamma > 50\,\text{MeV}$ in the barrel ($32.2^\circ < \theta < 128.7^\circ$), $E_\gamma > 100\,\text{MeV}$ for the forward endcap ($12.0^\circ < \theta < 31.4^\circ$), and $E_\gamma > 150\,\text{MeV}$ for the backward endcap ($131.5^\circ < \theta < 157.1^\circ$), where θ is the polar angle of the photon candidate in the laboratory frame.

The signal B candidates are then formed by combining an ALP candidate with a $K^{(*)}$ candidate. We select B candidates using two kinematic variables: the beam-constrained mass $M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - p_B^2}$ and the energy difference $\Delta E = E_B - E_{\text{beam}}$, where p_B and E_B are the momentum and energy of the B candidate in the centre-of-mass (c.m.) frame, and E_{beam} is the beam energy in the c.m. frame. We require $M_{\text{bc}} > 5.27\,\text{GeV}$ and $-0.2\,\text{GeV} < \Delta E < 0.1\,\text{GeV}$. A vertex fit is applied to the selected B candidates, constraining the photon and kaon candidates to originate from the IP, and the total invariant mass to the nominal B meson mass. After B meson reconstruction, about 25% of events have multiple signal candidates. For such events, we select the candidate with the smallest $|\Delta E|$. According to Monte Carlo simulations, approximately 85% of the candidates are correctly reconstructed after the candidate selection.

A series of fast boosted decision tree (BDT) classifiers [39] are used after event selection to suppress the background processes. To train the BDT classifiers, simulations with uniformly distributed diphoton invariant mass, $M_{\gamma\gamma}$, are used. This ensures that the classifiers do not learn or exploit the signal peak position in their background rejection strategy, and consequently avoids bias in the search for ALPs at a specific mass hypothesis. Since the distributions of input variables differ significantly between the low and high mass regions, the data above and below $1\,\text{GeV}$ are treated separately to enhance classification power. The training of all classifiers is based on simulated samples that are statistically independent of those used to develop and validate the fitting strategy. This separation ensures an unbiased optimization of the background suppression procedure before its application to the unblinded data.

We apply two continuum suppression (CS) classifiers to suppress the dominant background process, $e^+e^- \to q\bar{q}$. The first classifier, $CS1$, uses 10 variables. These are the ratio of zeroth to second Fox-Wolfram moments (R_2) [40], several modified Fox-Wolfram moments [41, 42], the cosine of the angle between the signal B-thrust and the rest-of-event (ROE) thrust axes, the cosine of the angle between the signal B-thrust and the z axes, and the magnitude of the ROE thrust. The thrust axis is defined as the axis that maximizes the sum of the projected momenta of all particles [43]. The signal B thrust axis is calculated using particles in the signal candidate, while the ROE thrust axis is determined from the charged particles and photons not used in the signal candidate reconstruction. The second classifier, $CS2$, is trained using events with $CS1 > 0.1$ and utilizes the event sphericity and aplanarity (which is a linear combination of the sphericity eigenvalue and 3/2 of the third sphericity eigenvalue[44]), the sum of the absolute value of the momenta of the particles moving along or against the thrust axis, harmonic moments (coefficients of the spherical harmonic event expansion around the thrust axis), the energy asymmetry between the two photons, modified Fox-Wolfram moments, and number of ALP candidates per event. The most discriminating variables in both classifiers are the cosine of the angle between the signal B-thrust and ROE-thrust axes for the low ALP mass hypothesis and the reduced Fox-Wolfram $R2$ for the high ALP mass hypothesis.

A major background arises from π^0 mesons produced in $\Upsilon(4S) \to B\bar{B}$ or $e^+e^- \to q\bar{q}$ processes, decaying into $\gamma\gamma$. To suppress these, we first calculate identification variables by combining one photon, γ_a, from the reconstructed ALP candidate with any other photon, γ_b, in the ROE. For each pair of photons (γ_a, γ_b), we perform three binary classification tests, each based on different assumptions about their source:

(1) We compare a pair in which γ_a is from a true ALP signal and γ_b is from the ROE in a signal event, against a pair from the same π^0 decay in a background event;

(2) We compare a pair in which γ_a is from a true ALP signal and γ_b is from the ROE in a signal event, against a pair where γ_a and γ_b are from different processes in a background event;

(3) We compare a pair in which both γ_a and γ_b come from different processes in a background event, against a pair where both photons come from the same particle decay, such as a π^0, also in a background event.

The tests are carried out with BDTs trained with multiple variables from Monte Carlo simulated events such as the ratio of energies E_9 and E_{25} in the inner 3×3 and 5×5 crystals around the central crystal (E_9/E_{25}), the energy of the most energetic crystal in the ECL cluster, the sum of weights of all crystals in an ECL cluster, the photon energy, as well as kinetic properties of the diphoton system such as its mass, energy, opening angle, transverse momentum, and energy asymmetry. Among the various possible partner photons γ_b, the one that yields the highest π^0-like score in each test is selected. These three individual scores are then combined into a single score, which is used to calculate the probabilities, $P_{\pi^0}(\gamma_1)$ and $P_{\pi^0}(\gamma_2)$, for each signal candidate photon, γ_1 and γ_2, to originate from a π^0.

For each ALP a mass range and kaon mode, selection criteria on the four BDT classifier scores, $CS1$, $CS2$, $P_{\pi^0}(\gamma_1)$ and $P_{\pi^0}(\gamma_2)$, are optimized using the Punzi figure of merit (PFM) [45]. A four-dimensional grid search with a step size of $\mathcal{O}(10^{-3})$ is performed, selecting the points with the highest PFM value as the minimum acceptance values of the BDT score. The BDT selections have average signal efficiencies of 53.4%, 62.0%, 94.0% and 98.6%, with average background events rejection rate of 95.4%, 31.1%, 27.8% and 11.1%, respectively.

An additional background contribution comes from the process $B \to X_s\gamma$. To suppress these events, we employ a BDT classifier trained with six variables, separately for the low and high ALP mass regions. These variables include the E_9/E_{25} ratio of each photon, the energy of the most energetic crystal in the ECL cluster, and two helicity angles. The first angle is between the photon momentum and the direction opposite to the B momentum in the a rest frame, and second angle is between the $K^{(*)}$ momentum and the direction opposite to the c.m. system in the B rest frame. The most powerful discriminating variables for all mass hypotheses are the first helicity angle and the energy of the most energetic photon in the c.m. frame. The BDT for $X_s\gamma$ suppression achieves an average signal efficiency of 76.6%, with a background rejection rate of 53.0%. Distributions of the BDT classifier scores can be found in Appendix B.

After all selection requirements, the signal selection efficiency varies across different kaon modes. The K^+ mode reaches the highest efficiency, ranging from 8% and 10% in the low mass region, and with more variation in the high mass region. The efficiency ranges from a minimum of 7.5% at the lowest point and peaks around 16% near an ALP mass of 3 GeV. The K_S^0 mode follows with slightly lower efficiency between 6% and 11%. The K^{*0} mode shows further reduced efficiency between 3% and 8%. The K^{*+} mode has the lowest efficiency, around 2% across the entire mass range.

4 Simulation calibration

We use the off-resonance data to evaluate corrections to the simulation by comparing experimental data with simulated distributions. The off-resonance data are collected 60 MeV below the $\Upsilon(4S)$ resonance energy, where e^+e^- collisions produce all processes except for $B\overline{B}$ pair production. There are two types of discrepancies found between the experimental data and the simulation: the number of background events, and the shape of the event kinematic and topology variables. For the former, we derive the ratios between the off-resonance data and simulation for each ALP mass range as weights, which are then applied to both the off-resonance and on-resonance continuum simulations. For the latter, we adopt a data-driven method [46]. A BDT classifier is trained to distinguish the differences between the off-resonance data and simulation, using the same variables applied for continuum suppression. A correction $w(i) = p(i)/(1-p(i))$ is derived for each simulated event i from the classifier output $p(i)$, where $0 < p(i) < 1$, and applied as a "continuum shape correction" weight. The resulting on-resonance $M_{\gamma\gamma}$ distributions for low and high ALP mass regions in the $B^+ \to K^+a$ decay are shown in Fig. 2.

5 Signal extraction and validations

The signal yield N_{Sig} and its standard deviation σ_{Sig} are obtained from an unbinned maximum likelihood fit to the $M_{\gamma\gamma}$ distribution. We use a double-sided Crystal Ball function [47] to model the signal $M_{\gamma\gamma}$ distribution. The non peaking background in the $M_{\gamma\gamma}$ distribution is parametrized using a second-order polynomial. The peaking background

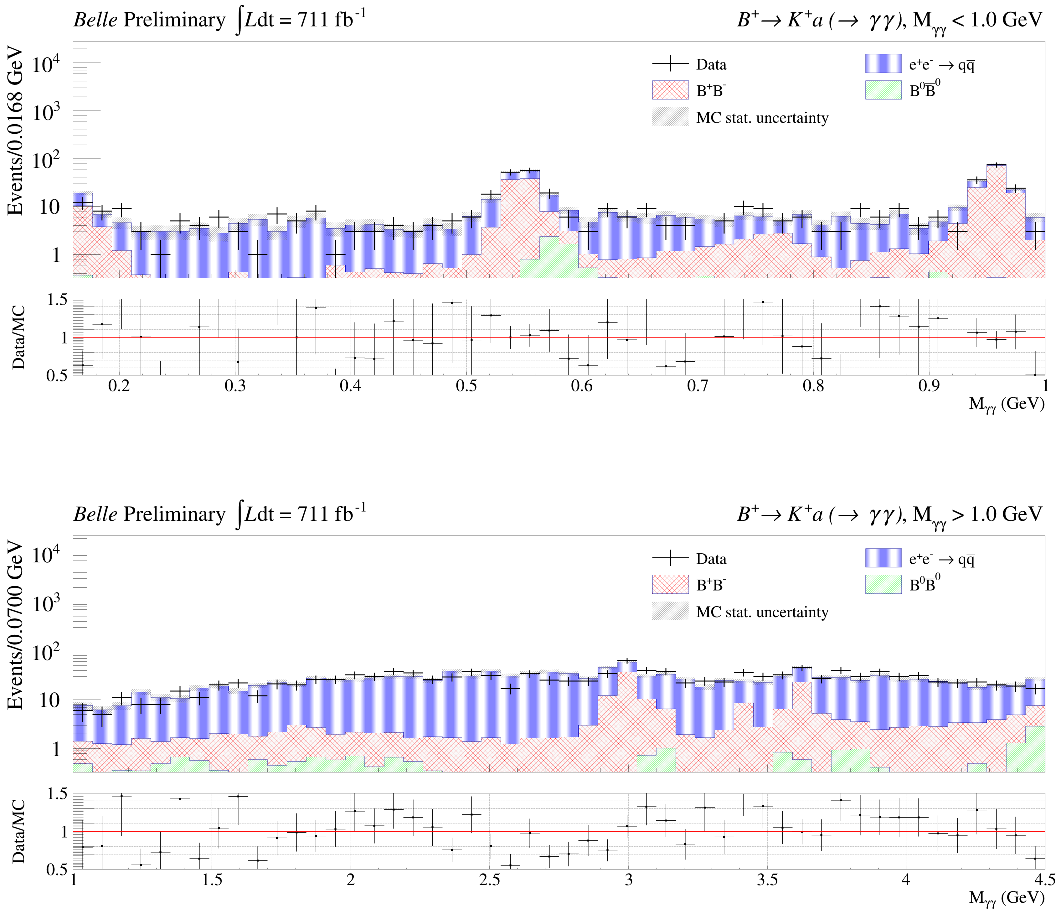

Figure 2: Diphoton invariant mass distribution of ALP candidates in $B^+ \to K^+a$ decay, overlaid with simulated background contributions from $e^+e^- \to q\overline{q}$ (blue vertical hatched), $e^+e^- \to \Upsilon(4S) \to B^+B^-$ (red cross-hatched), and $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ (green diagonal hatched) normalized to the experimental data luminosity, with all weights applied.

components from $h \to \gamma\gamma$ ($h = \pi^0$, η, or η') are modelled with double-sided Crystal Ball functions.

We perform a mass scan with a step size equal to the high-side mass resolution parameter from the Crystal Ball signal, $\sigma^R_{\gamma\gamma}$. The latter varies with the ALP mass, ranging from 7.8 MeV at $m_a = 0.160$ GeV to 19.4 MeV at $m_a = 1.9$ GeV, and decreasing to 17.9 MeV at $m_a = 4.5$ GeV. Each fit range extends over an $M_{\gamma\gamma}$ interval with a width of $9 \times (\sigma^R_{\gamma\gamma} + \sigma^L_{\gamma\gamma})$, where $\sigma^L_{\gamma\gamma}$ is the low-side Crystal Ball resolution parameter. The signal peak shape parameters depend on the ALP mass, and are derived from the corresponding signal samples. The shape parameters and position of the peaking background are fixed based on values obtained from the simulation. The combinatorial background parameters are floating in the fit, along with the signal and peaking background normalization.

Due to the peaking background from π^0, η and η' decays, masses below 0.160 GeV, in

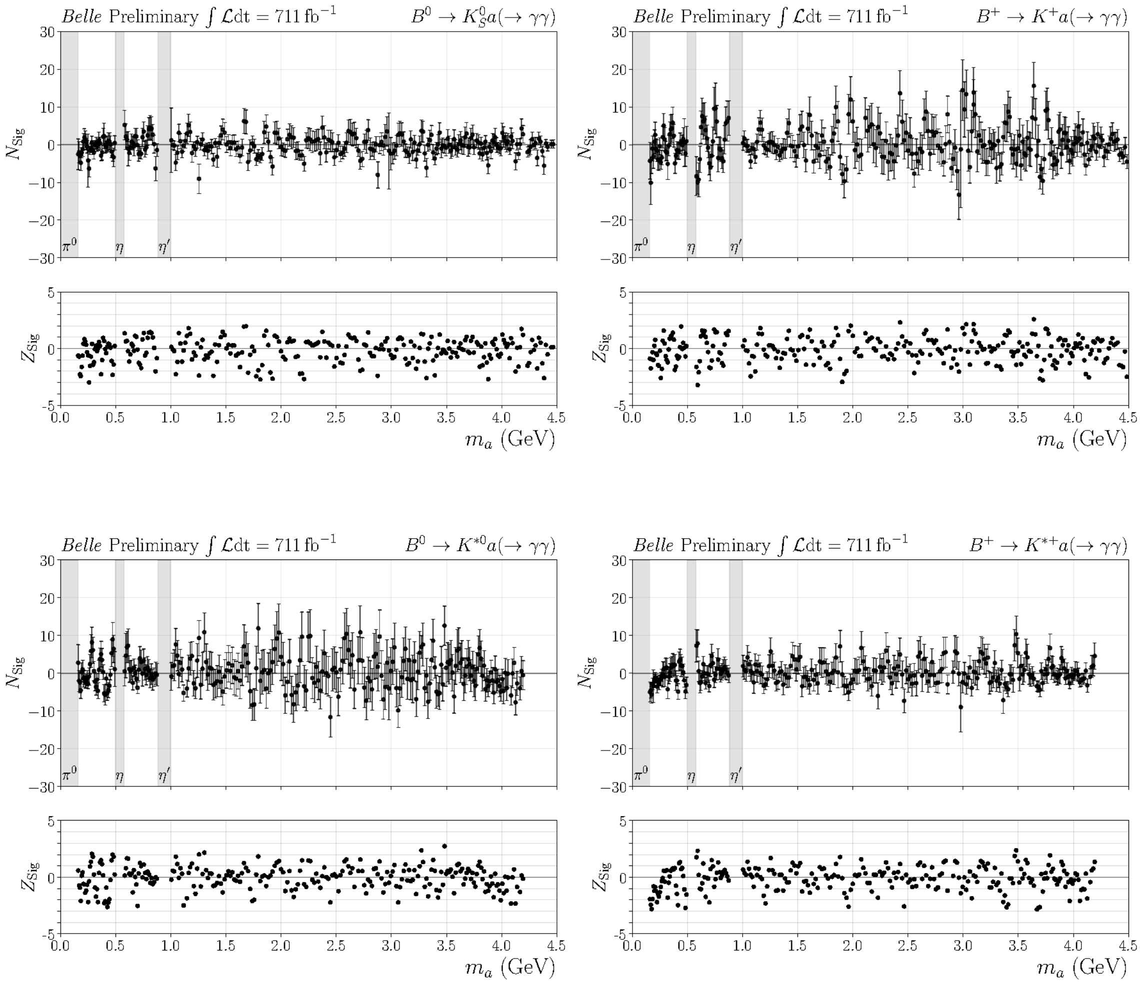

Figure 3: Extracted signal yield (N_{Sig}) (top) and the significance level (Z_{Sig}) (bottom) for four kaon modes. The grey bands are the excluded regions corresponding to the π^0, η and η' mass regions.

the ranges 0.497 GeV–0.578 GeV and 0.938–0.997 GeV, are excluded from the ALP mass scan. The latter two correspond to $\pm 3\sigma$ mass resolution windows centered on the η and η', respectively. Note that despite these exclusions, the tails of these peaking background components extend into the fitting regions. The ALP mass ranges close to the kinematic limit, from 4.50 GeV to 4.78 GeV for K and from 4.20 GeV to 4.38 GeV for K^*, are excluded from our analysis due to low signal efficiency and insufficient background population for reliable signal extraction.

To validate the signal extraction method, we measure the branching fraction of $B \to Kh(\to \gamma\gamma)$ using the fitting procedure described above. The branching fractions of the $B^+ \to K^+\eta$ and $B^+ \to K^+\eta'$ decays are measured to be $(0.95 \pm 0.27) \times 10^{-6}$ and $(1.79 \pm 0.45) \times 10^{-6}$, respectively, consistent with both the previous Belle results [48] and world

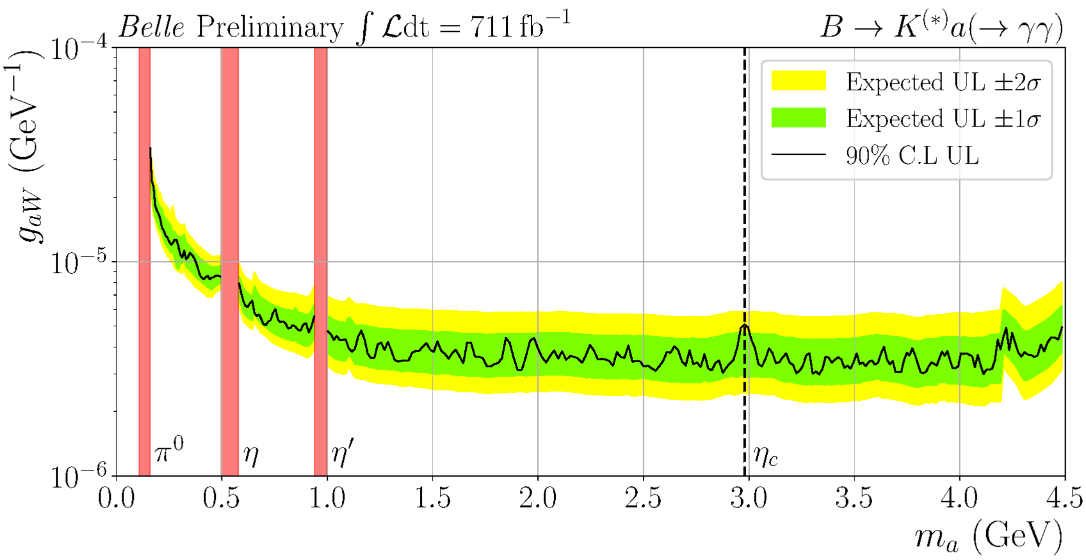

Figure 4: 90% CL upper limits on the coupling g_{aW} as a function of the ALP mass obtained with the CLs method with simultaneous fit to the four kaon modes. The green and yellow bands are the ± 1 and ± 2 standard deviation ranges, respectively, for the expected upper limits in the background only model. The red bands are the excluded π^0, η and η' mass regions. The vertical dashed line indicates the nominal η_c mass. Systematic uncertainties are included in the figure.

average values [37].

In addition, a toy Monte Carlo (ToyMC) study [49] — a simplified, fast simulation in which observables are sampled from probability density functions (p.d.f.) — is carried out to evaluate the fitting bias and signal sensitivity. Ten thousand pseudo-datasets are generated for each ALP mass hypothesis across the four kaon modes using the fitted p.d.f. and the Poisson distribution of signal and background yield obtained from the fit. The result of fits on pseudo-data from the ToyMC has a small negative bias of $4.2\% \times \sigma_{\mathrm{Sig}}$. This is applied as a correction factor to the signal yields in data.

For all ALP mass hypotheses and kaon modes, the decay yield of $B \to K^{(*)}a(\to \gamma\gamma)$ is measured under the assumption that all signal events originate from the prompt decay of the ALP. However, the ALP can be long-lived, and the displaced vertex reduces the signal efficiency. The primary reason for the reduced signal efficiency is the assumption that the photons originate from the IP. The calculated opening angle between the two photons, which enters the ALP mass calculation, is systematically smaller than the true value, which produces a low-side tail on the reconstructed mass distribution. The fit using Crystal Ball parameters from prompt decays systematically underestimates the number of signal events for displaced vertices, effectively reducing the efficiency. To quantify this effect, signal processes with lifetimes $c\tau$ of 10 mm, 50 mm, 100 mm, 200 mm, 300 mm, 400 mm and 500 mm are generated. The decrease in the signal efficiency is modelled as

$$\frac{\varepsilon_{\rm Sig}(c\tau)}{\varepsilon_{\rm Sig}(0)} = re^{a_1 c\tau} + (1-r)e^{a_2 c\tau}, \tag{5.1}$$

where $\varepsilon_{\rm Sig}(0)$ and $\varepsilon_{\rm Sig}(c\tau)$ are the reconstruction efficiencies of prompt and long-lived ALPs, a_1, a_2 and r are floating parameters obtained from fits to the long-lived ALP simulation results for each ALP mass hypothesis. The resulting functions are incorporated into the upper limit calculation. Consequently, ALP signals with relatively low mass have a longer lifetime and lower signal efficiency, leading to less stringent limits on g_{aW}.

The significance is evaluated as $Z_{\rm Sig} = \sqrt{2(L_{\rm s+b} - L_{\rm b})}$, where $L_{\rm s+b}$ and $L_{\rm b}$ are the negative log-likelihoods of the fits with and without signal hypothesis, respectively. The largest observed positive local significance is 2.74σ at $m_a = 3.482\,\text{GeV}$ in the K^{*0} mode. This local significance corresponds to 1.89σ global significance, after including the look-elsewhere effect [50]. As shown in Fig. 3, no significant excess over background is observed and we set 90% confidence level (CL) upper limits on the coupling g_{aW} using the CLs method [51, 52]. For each kaon mode and ALP mass hypothesis, we obtain the branching fractions of $B \to K^{(*)}a(\to \gamma\gamma)$ decays as

$$\mathcal{B}(B \to K^{(*)}a(\to \gamma\gamma)) = \frac{N_{\rm Sig}}{(2 \times N_{\Upsilon(4S)} \times f^x \times \varepsilon_{\rm Sig})}, \tag{5.2}$$

where $N_{\Upsilon(4S)}$ is the total number of $\Upsilon(4S)$ mesons, f^x is the production fraction of $B\overline{B}$ pairs for the neutral mode, f^{00}, or the charged mode, f^{+-} [53], and $\varepsilon_{\rm Sig}$ is the signal efficiency. For each ALP mass hypothesis, a simultaneous fit is performed on four kaon modes to obtain g_{aW}. Figure 4 shows the resulting limit on the coupling constant g_{aW} as a function of mass. The expected limits obtained with the background only hypothesis are also shown as green and yellow bands.

6 Systematic uncertainties

The systematic uncertainties that affect the extraction of g_{aW} can be classified into two main categories: those originating from prompt decay signal and from long-lived ALP efficiency (Table 1).

The uncertainty on the prompt decay signal is derived from our most precise control sample, $B^0 \to K^{*0}\eta$. The obtained branching fraction $\mathcal{B}(B^0 \to K^{*0}\eta) \times \mathcal{B}(\eta \to \gamma\gamma) = (6.8 \pm 1.3) \times 10^{-6}$ is in good agreement with the world-average value of $(6.3 \pm 0.4) \times 10^{-6}$ [37]. We therefore use the 21% uncertainty on the measurement as a conservative estimate of the systematic uncertainty. The indented items listed under the prompt decay signal in Table 1 represent numerically estimable components that affect the prompt decay signal calculation. Other sources not explicitly listed in the table, such as MVA selection efficiency, could not be individually determined and are thus incorporated into the total prompt decay signal alongside the indented items. Consequently, the quadratic sum of the individual components differs from the overall prompt decay signal value.

Table 1: Summary of the systematic uncertainties (%). The "Prompt decay signal " term is derived from a control sample: it includes the indented items that follow, as well as additional effects (see the text for details).

Source	K_S^0 mode	K^+ mode	K^{*0} mode	K^{*+} mode
Prompt decay signal		21.0		
$\vdash$ Continuum shape correction [54]		4.1		
$\vdash$ Photon-detection efficiency [55]		4.0		
$\vdash$ K^+ identification efficiency	-	3.6	3.6	-
$\vdash$ f^{00} or f^{+-} [53]	1.7	2.1	1.7	2.1
$\vdash$ K_S^0 reconstruction efficiency [56]	1.6	-	-	1.6
$\vdash$ $N_{\Upsilon(4S)}$ [57]		1.4		
$\vdash$ Tracking efficiency [55]	0.7	0.4	0.7	1.1
Long-lived ALP efficiency		7.4		
Total		22.3		

The systematic uncertainty of the long-lived ALP signal efficiency constraint is derived from $K_S^0 \to \pi^0\pi^0$ decays reconstructed in $D^{*+} \to D^0(\to K_S^0\pi^+\pi^-)\pi^+$ events. Using the same method used for long-lived ALP decays (Eq. 5.1), we construct the signal efficiency function for $K_S^0 \to \pi^0\pi^0$ and determine the expected mass resolution from this function and prompt decaying K_S^0 simulation. The $M(\pi^0\pi^0)$ mass resolution is extracted via p.d.f. fitting in both data and simulation, following the same approach used in ALP signal extraction. The difference between the obtained values and the expectation is (6.0 ± 4.4)%, leading to a conservative estimation of the systematic uncertainty of 7.4% for the long-lived ALP efficiency calculation. The efficiency for long-lived ALP signals decreases as the ALP mass increases, since higher masses correspond to shorter lifetimes. This effect becomes negligibly small for masses above 2.0 GeV. However, as our analysis is dominated by statistical uncertainties, and since mass-dependent refinements to systematic uncertainties would not significantly impact the final results, we have applied the most conservative systematic uncertainty estimate across the entire mass range in our calculations.

The total systematic uncertainty is obtained by summing the prompt decay signal and long-lived ALP efficiency uncertainties in quadrature. The resulting systematic uncertainty is found to be 22.3% for all kaon modes and ALP mass hypotheses. The resulting systematic uncertainty is included in the upper limit calculation procedure by convolving an appropriate Gaussian function with the signal likelihood.

7 Result and conclusion

We report a search for an axion-like particle in $B \to K^{(*)}a(\to \gamma\gamma)$ decays using a 711 fb^{-1} data sample collected by the Belle experiment at the KEKB e^+e^- collider at a centre-of-mass energy of 10.58 GeV. We search for the decay of the axion-like particle into a pair of photons, $a \to \gamma\gamma$, and explore four kaon modes, K_S^0, K^+, K^{*0} and K^{*+}. We scan the two-photon invariant mass in the range of 0.16 GeV–4.50 GeV for the K modes and 0.16 GeV–4.20 GeV for the K^* modes. No significant signal is observed in any of the modes. Figure 5 shows the resulting 90% confidence level upper limits on the g_{aW} coupling

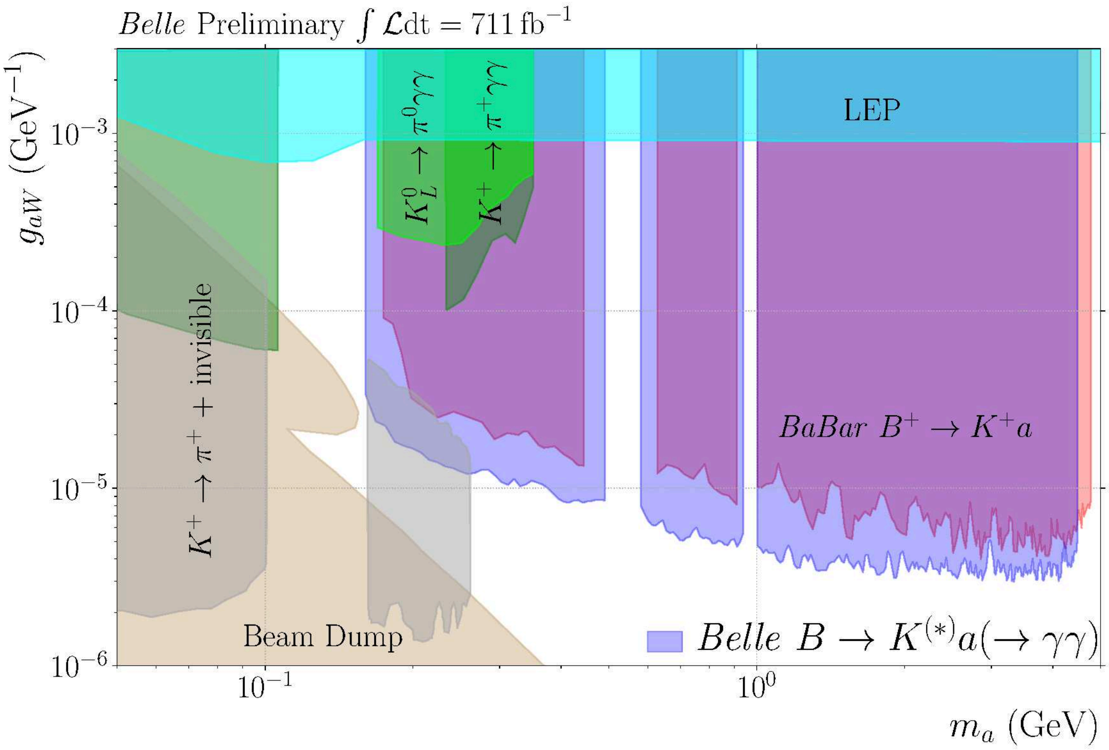

Figure 5: The 90% CL upper limits on the coupling g_{aW} from a simultaneous fit to the four $B \to K^{(*)}a$ modes as a function of the ALP mass, compared with existing constraints [25, 26, 58, 59].

as a function of m_a, derived from the combination of four kaon modes. The limits are $3 \times 10^{-6}\,\text{GeV}^{-1}$ for the ALP mass hypotheses above 2.0 GeV, increasing to $3 \times 10^{-5}\,\text{GeV}^{-1}$ at the lowest ALP mass. This trend is due to the increase in the lifetime, which leads to a lower signal efficiency. Figure 5 also shows the constraints derived from the NA62 $K^+ \to \pi^+ +$ invisible search [58]. Based on the methodology presented in Ref. [59] we reinterpret the NA62 results on a dark scalar decaying to SM particles as limits on ALPs. The constraints on the coupling of the axion-like particle to electroweak gauge bosons g_{aW} are improved by a factor of two compared to the most stringent previous experimental results.[26]

This work, based on data collected using the Belle detector, which was operated until June 2010, was supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, the Japan Society for the Promotion of Science (JSPS), and the Tau-Lepton Physics Research Center of Nagoya University; the Australian Research Council including grants DP210101900, DP210102831, DE220100462, LE210100098, LE230100085; Austrian Federal Ministry of Education, Science and Research (FWF) and FWF Austrian Science Fund No. P 31361-N36; National Key R&D Program of China under Contract No. 2022YFA1601903, National Natural Science Foundation of China and research grants No. 11575017, No. 11761141009, No. 11705209, No. 11975076, No. 12135005, No. 12150004, No. 12161141008, and No. 12175041, and Shandong Provincial Natural Science Foundation Project ZR2022JQ02; the Czech Science Foundation Grant No. 22-18469S; Horizon 2020 ERC Advanced Grant No. 884719 and ERC Starting Grant No. 947006 "InterLeptons" (European Union); the Carl Zeiss Foundation, the Deutsche Forschungsgemeinschaft, the Excellence Cluster Universe, and the VolkswagenStiftung; the Department of Atomic Energy (Project Identification No. RTI 4002), the Department of Science and Technology of India, and the UPES (India) SEED finding programs Nos. UPES/R&D-SEED-INFRA/17052023/01 and UPES/R&D-SOE/20062022/06; the Istituto Nazionale di Fisica Nucleare of Italy; National Research Foundation (NRF) of Korea Grant Nos. 2016R1-D1A1B02012900, 2018R1A2B3003643, 2018R1A6A1A06024970, RS202200197659, 2019R1-I1A3A01058933, 2021R1A6A1A03043957, 2021R1F1A1060423, 2021R1F1A1064008, 2022R1-A2C1003993; Radiation Science Research Institute, Foreign Large-size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/GLORIAD; the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation and the HSE University Basic Research Program, Moscow; University of Tabuk research grants S-1440-0321, S-0256-1438, and S-0280-1439 (Saudi Arabia); the Slovenian Research Agency Grant Nos. J1-50010 and P1-0135; Ikerbasque, Basque Foundation for Science, and the State Agency for Research of the Spanish Ministry of Science and Innovation through Grant No. PID2022-136510NB-C33 (Spain); the Swiss National Science Foundation; the Ministry of Education and the National Science and Technology Council of Taiwan; and the United States Department of Energy and the National Science Foundation. These acknowledgements are not to be interpreted as an endorsement of any statement made by any of our institutes, funding agencies, governments, or their representatives. We thank the KEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid; and the KEK computer group and the Pacific Northwest National Laboratory (PNNL) Environmental Molecular Sciences Laboratory (EMSL) computing group for strong computing support; and the National Institute of Informatics, and Science Information NETwork 6 (SINET6) for valuable network support.

References

[1] R.D. Peccei and H.R. Quinn, CP *conservation in the presence of pseudoparticles*, *Phys. Rev. Lett.* **38** (1977) 1440.

[2] S. Weinberg, *A new light boson?*, *Phys. Rev. Lett.* **40** (1978) 223.

[3] F. Wilczek, *Problem of strong p and t invariance in the presence of instantons*, *Phys. Rev. Lett.* **40** (1978) 279.

[4] J. Preskill, M.B. Wise and F. Wilczek, *Cosmology of the Invisible Axion*, *Phys. Lett. B* **120** (1983) 127.

[5] L.F. Abbott and P. Sikivie, *A Cosmological Bound on the Invisible Axion*, *Phys. Lett. B* **120** (1983) 133.

[6] M. Dine and W. Fischler, *The Not So Harmless Axion*, *Phys. Lett. B* **120** (1983) 137.

[7] Y. Nomura and J. Thaler, *Dark matter through the axion portal*, *Phys. Rev. D* **79** (2009) 075008.

[8] K. Mimasu and V. Sanz, *ALPs at Colliders*, *JHEP* **06** (2015) 173 [arXiv:1409.4792].

[9] M.J. Dolan, F. Kahlhoefer, C. McCabe and K. Schmidt-Hoberg, *A taste of dark matter: Flavour constraints on pseudoscalar mediators*, *JHEP* **03** (2015) 171 [arXiv:1412.5174].

[10] J. Jaeckel and M. Spannowsky, *Probing MeV to 90 GeV axion-like particles with LEP and LHC*, *Phys. Lett. B* **753** (2016) 482 [arXiv:1509.00476].

[11] B. Döbrich, J. Jaeckel, F. Kahlhoefer, A. Ringwald and K. Schmidt-Hoberg, *ALPtraum: ALP production in proton beam dump experiments*, *JHEP* **02** (2016) 018 [arXiv:1512.03069].

[12] S. Knapen, T. Lin, H.K. Lou and T. Melia, *Searching for Axionlike Particles with Ultraperipheral Heavy-Ion Collisions*, *Phys. Rev. Lett.* **118** (2017) 171801 [arXiv:1607.06083].

[13] 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*, *JHEP* **12** (2017) 094 [arXiv:1709.00009].

[14] I. Brivio, M.B. Gavela, L. Merlo, K. Mimasu, J.M. No, R. del Rey et al., *ALPs Effective Field Theory and Collider Signatures*, *Eur. Phys. J. C* **77** (2017) 572 [arXiv:1701.05379].

[15] M. Bauer, M. Neubert and A. Thamm, *LHC as an Axion Factory: Probing an Axion Explanation for $(g-2)_\mu$ with Exotic Higgs Decays*, *Phys. Rev. Lett.* **119** (2017) 031802 [arXiv:1704.08207].

[16] K. Choi, S.H. Im, C.B. Park and S. Yun, *Minimal Flavor Violation with Axion-like Particles*, *JHEP* **11** (2017) 070 [arXiv:1708.00021].

[17] M. Bauer, M. Neubert and A. Thamm, *Collider Probes of Axion-Like Particles*, *JHEP* **12** (2017) 044 [arXiv:1708.00443].

[18] CHARM collaboration, *Search for Axion Like Particle Production in 400-GeV Proton - Copper Interactions*, *Phys. Lett. B* **157** (1985) 458.

[19] E.M. Riordan et al., *A Search for Short Lived Axions in an Electron Beam Dump Experiment*, *Phys. Rev. Lett.* **59** (1987) 755.

[20] J.D. Bjorken, S. Ecklund, W.R. Nelson, A. Abashian, C. Church, B. Lu et al., *Search for Neutral Metastable Penetrating Particles Produced in the SLAC Beam Dump*, *Phys. Rev. D* **38** (1988) 3375.

[21] J. Blumlein et al., *Limits on neutral light scalar and pseudoscalar particles in a proton beam dump experiment*, *Z. Phys. C* **51** (1991) 341.

[22] OPAL collaboration, *Multiphoton production in e^+e^- collisions at $\sqrt{s}$ = 181 GeV to 209 GeV*, *Eur. Phys. J. C* **26** (2003) 331 [arXiv:hep-ex/0210016].

[23] D. Aloni, Y. Soreq and M. Williams, *Coupling QCD-Scale Axionlike Particles to Gluons*, *Phys. Rev. Lett.* **123** (2019) 031803 [arXiv:1811.03474].

[24] Belle II collaboration, *Search for Axion-Like Particles produced in e^+e^- collisions at Belle II*, *Phys. Rev. Lett.* **125** (2020) 161806 [arXiv:2007.13071].

[25] E. Izaguirre, T. Lin and B. Shuve, *Searching for axionlike particles in flavor-changing neutral current processes*, *Phys. Rev. Lett.* **118** (2017) 111802.

[26] BaBar collaboration, *Search for an Axionlike Particle in B Meson Decays*, Phys. Rev. Lett. **128** (2022) 131802 [arXiv:2111.01800].

[27] Belle collaboration, *The Belle Detector*, Nucl. Instrum. Meth. A **479** (2002) 117.

[28] S. Kurokawa and E. Kikutani, *Overview of the KEKB accelerators*, Nucl. Instrum. Meth. A **499** (2003) 1.

[29] D.J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[30] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159 [arXiv:1410.3012].

[31] E. Barberio, B. van Eijk and Z. Wcas, *PHOTOS: A universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. **66** (1991) 115.

[32] R. Brun, F. Bruyant, M. Maire, A.C. McPherson and P. Zanarini, *GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version*, CERN, Geneva (9, 1987).

[33] M. Gelb et al., *B2BII: Data Conversion from Belle to Belle II*, Comput. Softw. Big Sci. **2** (2018) 9 [arXiv:1810.00019].

[34] Belle II Framework Software Group, *The Belle II Core Software*, Comput. Softw. Big Sci. **3** (2019) 1 [arXiv:1809.04299].

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

[36] Belle collaboration, *Measurement of branching fractions of $\Lambda_c^+ \to pK_S^0K_S^0$ and $\Lambda_c^+ \to pK_S^0\eta$ at Belle*, Phys. Rev. D **107** (2023) 032004 [arXiv:2210.01995].

[37] Particle Data Group, *Review of particle physics*, Phys. Rev. D **110** (2024) 030001.

[38] H. Nakano, *Search for new physics by a time-dependent CP violation analysis of the decay $B \to K_S\eta\gamma$ using the Belle detector*, Ph.D. thesis, Tohoku U., 2015.

[39] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, Comput. Softw. Big Sci. **1** (2017) 2.

[40] G.C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978) 1581.

[41] Belle collaboration, *Evidence for $B^0 \to \pi^0\pi^0$*, Phys. Rev. Lett. **91** (2003) 261801 [arXiv:hep-ex/0308040].

[42] Belle collaboration, *Observation of Cabibbo suppressed $B \to D^{(*)}K^-$ decays at Belle*, Phys. Rev. Lett. **87** (2001) 111801 [arXiv:hep-ex/0104051].

[43] Ed. A. J. Bevan, B. Golob, Th. Mannel, S. Prell, and B. D. Yabsley, *The physics of the B factories*, Eur. Phys. J. **C74** (2014) 3026 [arXiv:1406.6311].

[44] J.D. Bjorken and S.J. Brodsky, *Statistical Model for Electron-Positron Annihilation into Hadrons*, Phys. Rev. D **1** (1970) 1416.

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

[46] D. Martschei, M. Feindt, S. Honc and J. Wagner-Kuhr, *Advanced event reweighting using multivariate analysis*, J. Phys. Conf. Ser. **368** (2012) 012028.

[47] J. Gaiser, *Charmonium spectroscopy from radiative decays of the J/ψ and ψ'*, Ph.D. thesis, Stanford University, 1982.

[48] Belle collaboration, *Search for resonant $B^{\pm} \to K^{\pm}\ h \to K^{\pm}\gamma\gamma$ Decays at Belle*, Phys. Lett. B **662** (2008) 323 [arXiv:hep-ex/0608037].

[49] W. Verkerke and D.P. Kirkby, *The RooFit toolkit for data modeling*, vol. C0303241, p. MOLT007, 2003 [arXiv:physics/0306116].

[50] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, Eur. Phys. J. C **70** (2010) 525 [arXiv:1005.1891].

[51] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554 [arXiv:1007.1727].

[52] A.L. Read, *Presentation of search results: The CL_s technique*, J. Phys. G **28** (2002) 2693.

[53] S. Banerjee et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023*, arXiv:2411.18639.

[54] Belle collaboration, *Search for $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging*, Phys. Rev. D **101** (2020) 032007 [arXiv:1911.03186].

[55] Belle collaboration, *Search for the radiative penguin decays $B^0 \to K^0_S K^0_S \gamma$ in the Belle experiment*, Phys. Rev. D **106** (2022) 012006 [arXiv:2203.05320].

[56] N. Dash et al., *Search for CP Violation and Measurement of the Branching Fraction in the Decay $D^0 \to K^0_S K^0_S$*, Phys. Rev. Lett. **119** (2017) 171801 [arXiv:1705.05966].

[57] Belle collaboration, *Study of Exclusive $B \to X_u \ell\nu$ Decays and Extraction of $|V_{ub}|$ using Full Reconstruction Tagging at the Belle Experiment*, Phys. Rev. D **88** (2013) 032005 [arXiv:1306.2781].

[58] NA62 collaboration, *Measurement of the very rare $K^+ \to \pi^+\nu\overline{\nu}$ decay*, JHEP **06** (2021) 093 [arXiv:2103.15389].

[59] E. Goudzovski et al., *New physics searches at kaon and hyperon factories*, Rept. Prog. Phys. **86** (2023) 016201 [arXiv:2201.07805].

[60] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited*, Phys. Rev. D **71** (2005) 014029 [arXiv:hep-ph/0412079].

[61] P. Ball and R. Zwicky, *New results on $B \to \pi, K, \eta$ decay formfactors from light-cone sum rules*, Phys. Rev. D **71** (2005) 014015 [arXiv:hep-ph/0406232].

A Decay width of $B \to K^{(*)}a(\to \gamma\gamma)$ and coupling strength g_{aW}

The $B \to K^{(*)}a(\to \gamma\gamma)$ decay width is given by

$$\begin{aligned}\Gamma(B \to Ka) &= \frac{M_B^3}{64\pi}|g_{abs}|^2(1-\frac{M_K^2}{M_B^2})^2 f_0^2(m_a^2)\lambda_{Ka}^{1/2},\\ \Gamma(B \to K^*a) &= \frac{M_B^3}{64\pi}|g_{abs}|^2 A_0^2(m_a^2)\lambda_{K^*a}^{3/2},\end{aligned} \tag{A.1}$$

where

$$\begin{aligned}g_{abs} &= -\frac{3\sqrt{2}G_F M_W^2 g_{aW}}{16\pi^2}\sum_{\alpha=c,t} V_{\alpha b}V_{\alpha s}^* f(\frac{M_\alpha^2}{M_W^2}),\\ f(x) &\equiv \frac{x[1+x(\log x-1)]}{(1-x)^2},\\ f_0(m_a^2) &= \frac{0.330}{1-m_a^2/37.46},\\ A_0(m_a^2) &= \frac{1.364}{1-m_a^2/27.88}-\frac{0.990}{1-m_a^2/36.78},\\ \lambda_{K^{(*)}a} &= (1-\frac{(m_a+M_{K^{(*)}})^2}{M_B^2})(1-\frac{(m_a-M_{K^{(*)}})^2}{M_B^2}).\end{aligned} \tag{A.2}$$

Here, M_B, M_K and M_W are the masses of B meson, kaon and W boson, respectively, while m_a is the ALP mass. The functions $f_0(q)$ and $A_0(q)$ are the form factors from the hadronic matrix elements [60, 61], G_F is the Fermi constant, and $V_{\alpha b}$ and $V_{\alpha s}$ with $\alpha = c, t$ are the corresponding Cabibbo-Kobayashi-Maskawa (CKM) matrix elements.

B Figures

The following includes additional figures.

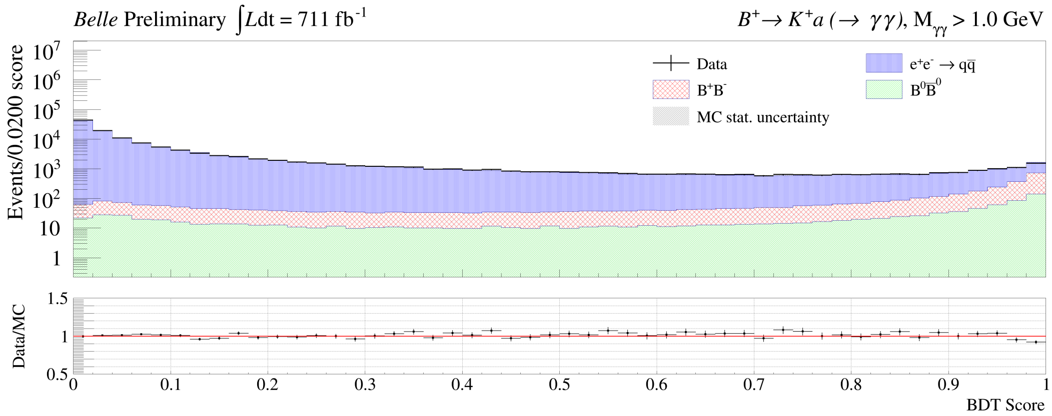

Figure 6: Distributions of the first continuum suppression BDT classifier scores from experimental data (black points with error bars) for the $B^+ \to K^+a(\to \gamma\gamma)$ decay with $m_a > 1.0\,\mathrm{GeV}/c^2$ along with simulated background contributions from $e^+e^- \to q\bar{q}$ (blue vertical hatched), $e^+e^- \to \Upsilon(4S) \to B^+B^-$ (red cross-hatched), and $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ (green diagonal hatched) normalized to the experimental data luminosity. All correction weights are applied to the continuum events.

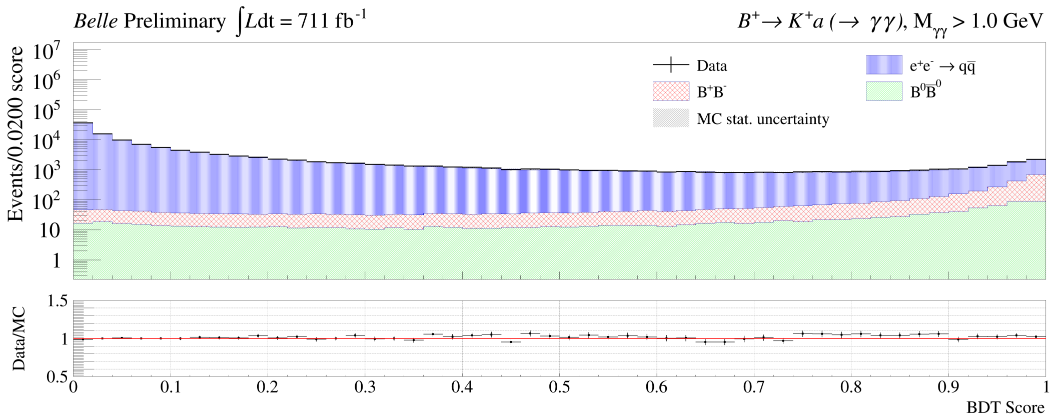

Figure 7: Distributions of the second continuum suppression BDT classifier scores. The colour convention used in this histogram is identical to that in Fig. 6

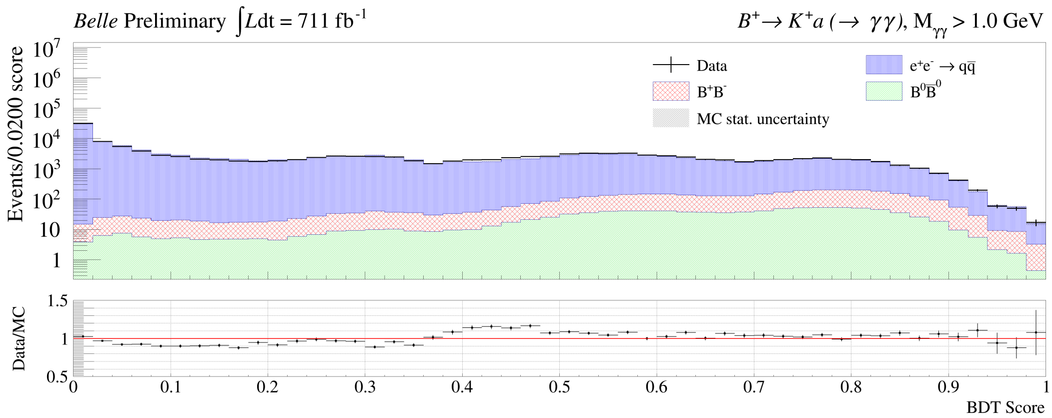

Figure 8: Distributions of $P_{\pi^0}(\gamma_1)$ BDT classifier scores. The colour convention used in this histogram is identical to that in Fig. 6

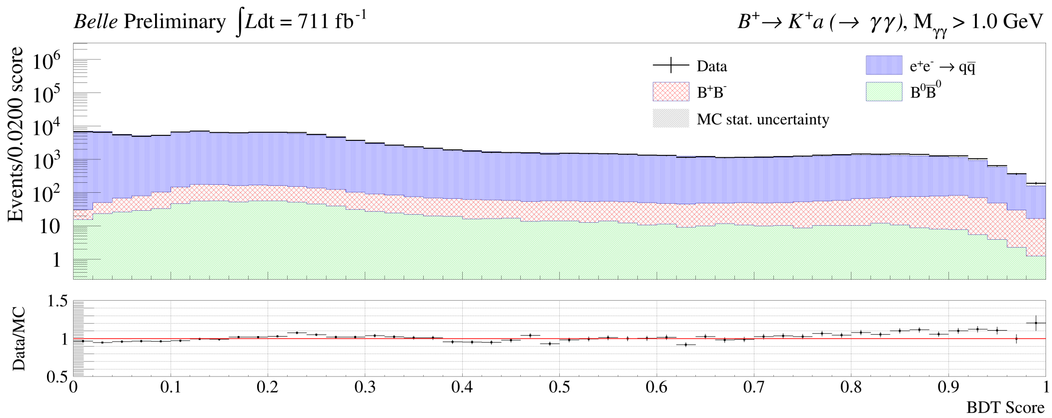

Figure 9: Distributions of $P_{\pi^0}(\gamma_2)$BDT classifier scores. The colour convention used in this histogram is identical to that in Fig. 6

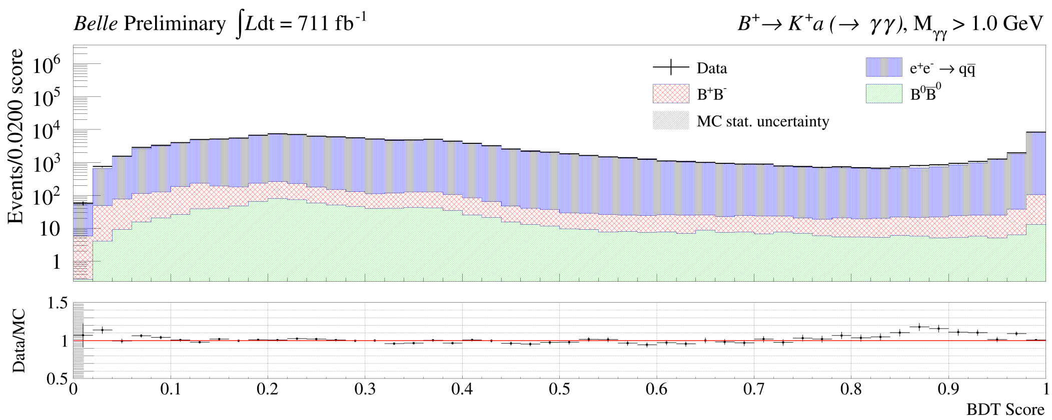

Figure 10: Distributions of $X_s\gamma$ suppression BDT classifier scores. The colour convention used in this histogram is identical to that in Fig. 6

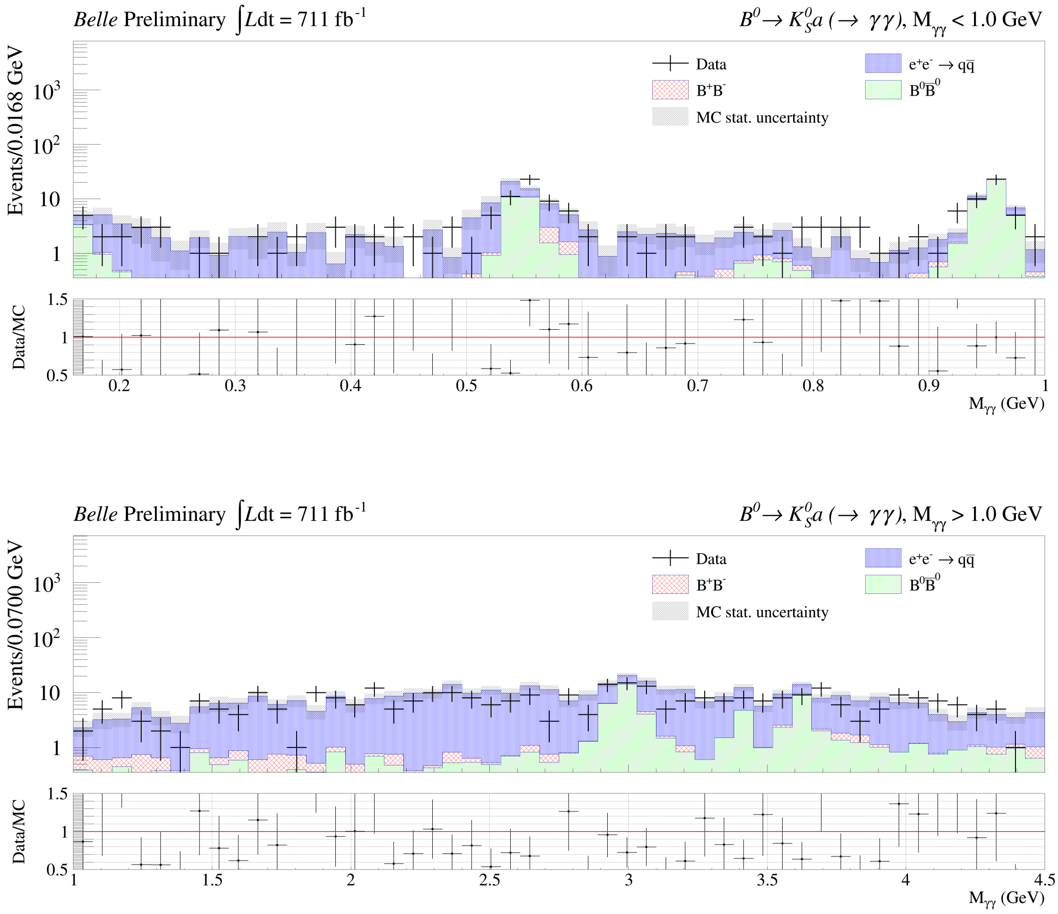

Figure 11: Diphoton invariant mass distribution of ALP candidates in $B^0 \to K_S^0 a$ decay, The colour convention used in this histogram is identical to that in Fig. 2

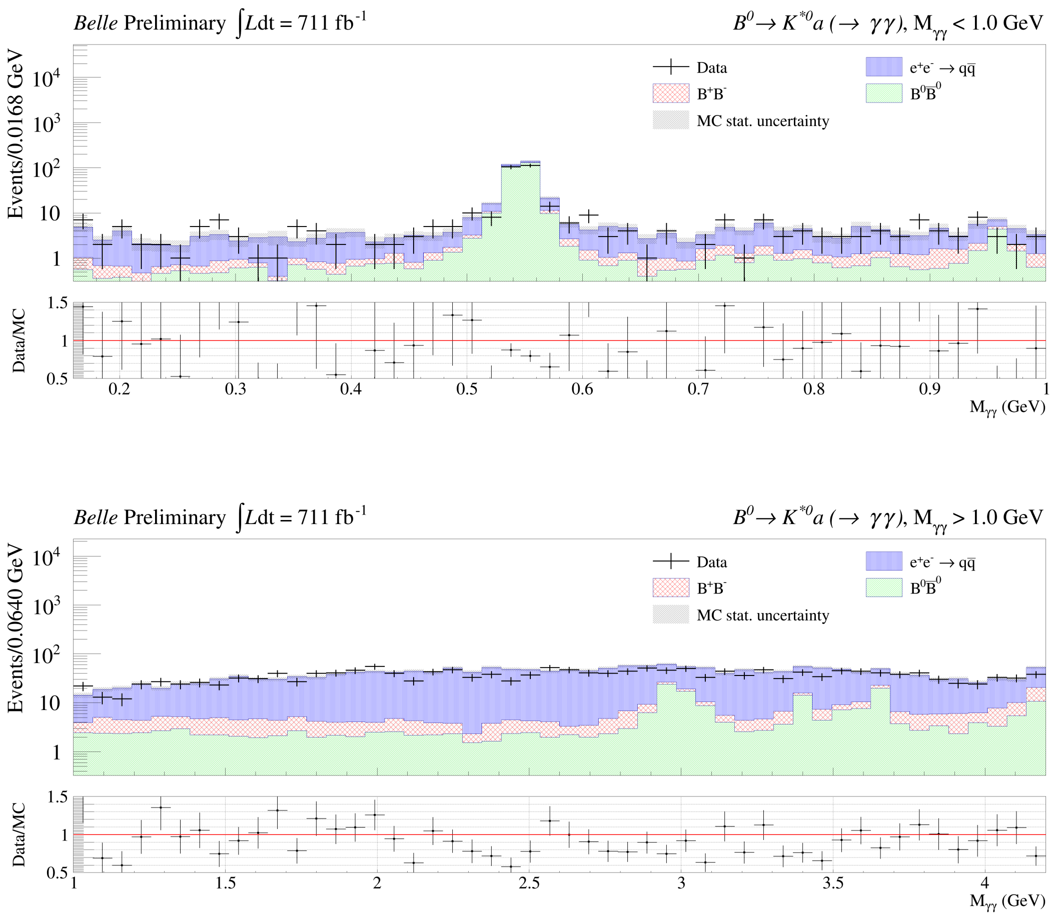

Figure 12: Diphoton invariant mass distribution of ALP candidates in $B^0 \to K^{*0}a$ decay, The colour convention used in this histogram is identical to that in Fig. 2

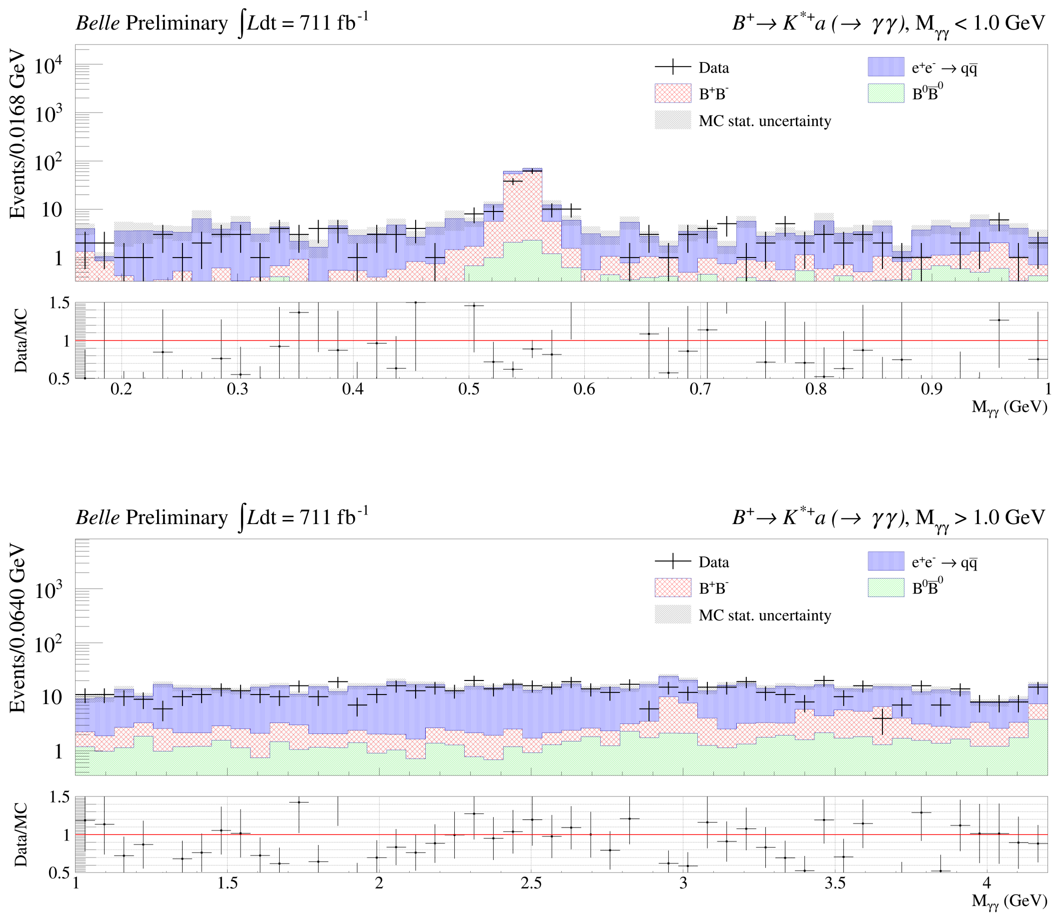

Figure 13: Diphoton invariant mass distribution of ALP candidates in $B^+ \to K^{*+}a$ decay, The colour convention used in this histogram is identical to that in Fig. 2

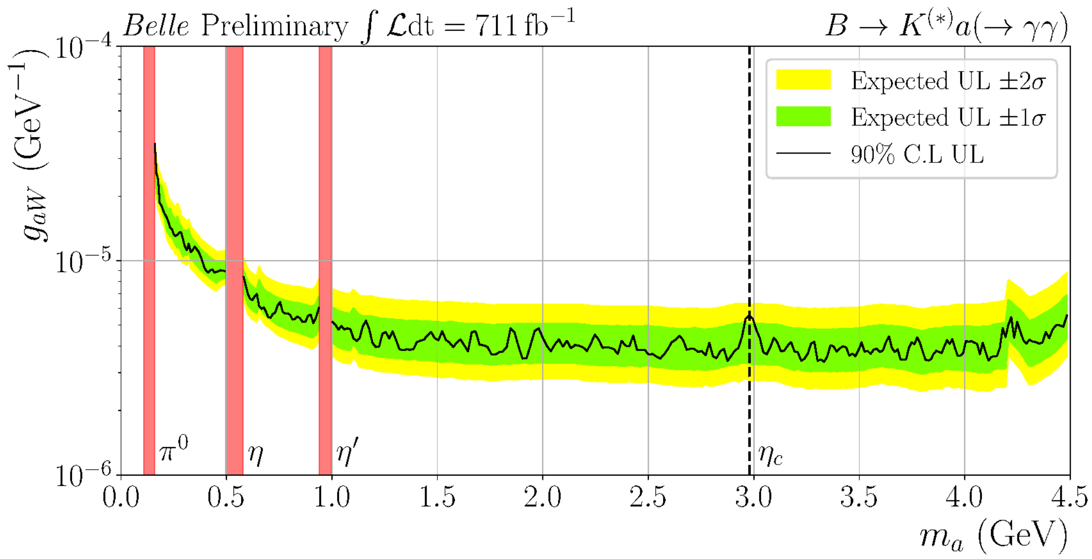

Figure 14: The 95% CL upper limits on the coupling g_{aW} as a function of the ALP mass obtained with CLs method with simultaneous fit. The green and yellow bands are ± 1 and ± 2 standard deviation ranges, respectively, of expected upper limit of background only model. The red bands are the excluded π^0, η and η' mass regions. The vertical dashed line indicates the nominal η_c mass.

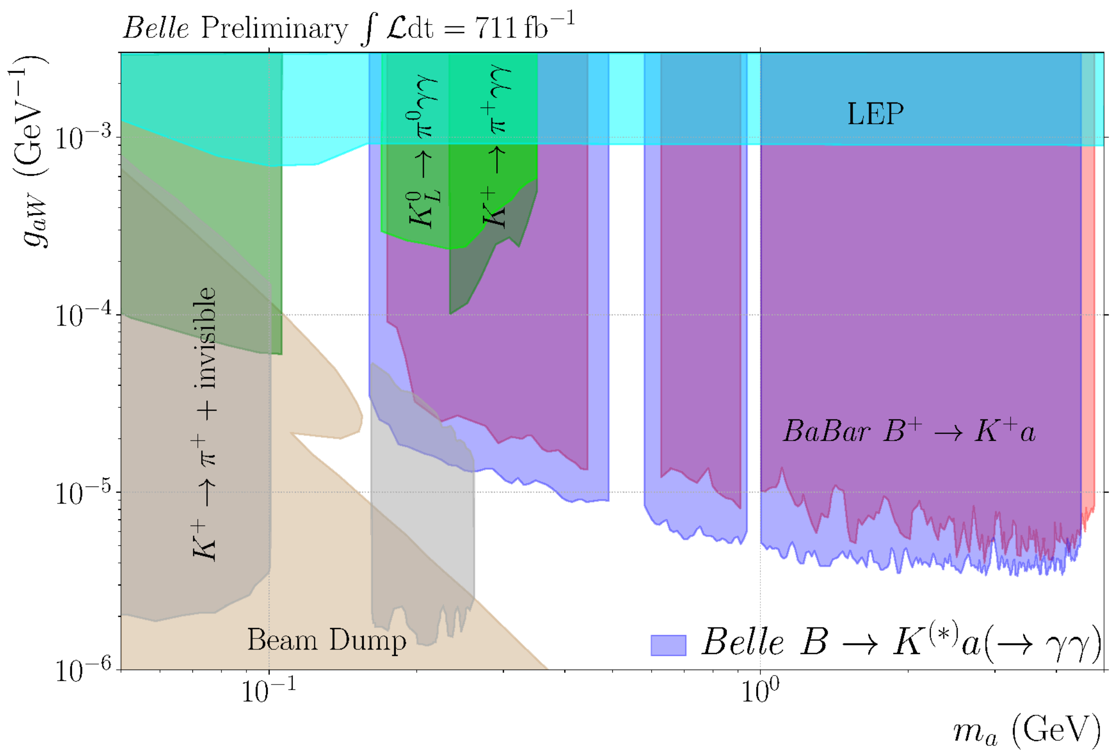

Figure 15: The 95% CL upper limits on the coupling g_{aW} from a simultaneous fit to the four $B \to K^{(*)}a$ modes as a function of the ALP mass, compared with existing constraints.