Search for lepton-flavor-violating $\tau^- \to \ell^- K_s^0$ decays at Belle and Belle II

The Belle and Belle II collaborations

E-mail: coll-publications@belle2.org, laura.zani@roma3.infn.it

ABSTRACT: We present the results of a search for charged-lepton-flavor violating decays $\tau^- \to \ell^- K_S^0$,[1] where ℓ^- is either an electron or a muon. We combine e^+e^- data samples recorded by the Belle II experiment at the SuperKEKB collider ($428\,\mathrm{fb}^{-1}$) with samples recorded by the Belle experiment at the KEKB collider ($980\,\mathrm{fb}^{-1}$) to obtain a sample of 1.3 billion $e^+e^- \to \tau^+\tau^-$ events. We observe 0 and 1 events and set 90% confidence level upper limits of 0.8×10^{-8} and 1.2×10^{-8} on the branching fractions of the decay modes $\tau^- \to e^- K_S^0$ and $\tau^- \to \mu^- K_S^0$, respectively. These are the most stringent upper limits to date.

KEYWORDS: Beyond Standard Model, Tau Physics, e^+-e^- Experiments, Flavour Physics

ARXIV EPRINT: 2504.15745

---

[1]Charge conjugation is implied throughout this document.

Contents

1 Introduction

The conservation of lepton flavor in charged-lepton decays is built into the Standard Model (SM) at tree level. The observation of neutrino oscillations implies that neutrinos have mass, so charged-lepton-flavor violating (LFV) decays can occur via internal loops, manifesting in processes such as $\mu \to e$, $\tau \to e$, and $\tau \to \mu$ conversions. All LFV amplitudes are suppressed by the squared ratio of the neutrino mass to the W-boson mass $(m_\nu/m_W)^2$. Consequently, predicted branching fractions are of the order of 10^{-50} [1–3], well below the sensitivities of current experiments. The detection of charged LFV decays would unequivocally indicate physics beyond the SM. Due to the large mass of the τ lepton, a wide range of possible LFV τ decays can be probed. In particular, in an effective field theory approach, the decays $\tau \to \ell M$, where M is a meson, can be used to constrain different types of operators [4–6], for example the two-lepton- and two-quark-operators that can arise in leptoquark models [7].

Over the past four decades, experiments such as CLEO at CESR and the first-generation B-factory experiments, BaBar at SLAC and Belle at KEK, have tested LFV τ-lepton decays [8]. A total of 52 LFV τ decay modes involving neutrinoless two-body and three-body final states have been investigated. The most stringent limits on the decays $\tau^- \to \ell^- K_S^0$, where $\ell = e, \mu$,

were obtained by the Belle collaboration, which found branching fraction upper limits at 90 % confidence level (C.L.) of 2.6×10^{-8} for the electron mode and 2.3×10^{-8} for the muon mode using a $671 \, \text{fb}^{-1}$ data sample, or 617 million $e^+e^- \to \tau^+\tau^-$ events [9].

We present the results of a search for the LFV decays $\tau^- \to \ell^- K_S^0$ using a combined data sample of 1.3 billion $e^+e^- \to \tau^+\tau^-$ events recorded with the Belle ($980 \, \text{fb}^{-1}$) and Belle II ($428 \, \text{fb}^{-1}$) detectors at the asymmetric-energy e^+e^- KEKB and SuperKEKB colliders [10, 11]. Candidate $\tau^- \to \ell^- K_S^0$ decays are selected in events where the second τ (*tag*) is reconstructed in a one-prong (single charged track) topology. The background rejection is optimized separately for Belle and Belle II datasets, and for each tag-side category, electronic, muonic or pionic. The optimization, which does not use data in the kinematic region where signal events are expected to peak, places loose requirements on kinematic and global-event variables and refines the selection using a boosted decision tree (BDT). The expected background yield is obtained from a fit to the reconstructed τ-mass sidebands, and an upper limit on the branching fraction is obtained using a frequentist method. The rest of this paper is structured as follows. Section 2 provides an overview of the Belle (II) detector and the data samples utilized. Section 3 outlines the candidate reconstruction and selection process, and section 4 addresses the systematic uncertainties. Section 5 presents the branching fraction measurement and limit computation, and section 6 summarizes the results.

2 The Belle and Belle II detectors, simulation and data samples

The Belle II experiment is located at SuperKEKB [11], which collides electrons and positrons near the $\Upsilon(4S)$ resonance. The Belle II detector [12] has a cylindrical geometry and includes a two-layer silicon-pixel detector (PXD) surrounded by a four-layer double-sided silicon-strip detector (SVD) [13] and a 56-layer central drift chamber (CDC). These detectors reconstruct the trajectories (tracks) of charged particles and provide energy loss measurements. Only one sixth of the second layer of the PXD was installed for the data analyzed here. Surrounding the CDC are a time-of-propagation detector (TOP) [14] in the central region and an aerogel-based ring-imaging Cherenkov detector (ARICH) in the forward region, which corresponds to the electron beam direction. These detectors provide information for identifying charged particles. Surrounding the TOP and ARICH is an electromagnetic calorimeter (ECL) based on CsI(Tl) crystals that provides energy and timing measurements, primarily for photons and electrons. Outside the ECL is a superconducting solenoid magnet that provides a 1.5 T axial field. Its flux return is instrumented with resistive-plate chambers and plastic scintillator modules to detect muons, K_L^0 mesons, and neutrons. The symmetry axis of the magnet, which is almost coincident with the direction of the electron beam, is used to define the z axis. The Belle detector was located at the interaction point of the KEKB collider [15]. It shares a similar structure to Belle II but lacks a silicon pixel detector and uses aerogel threshold Cherenkov counters (ACC) and a barrel-like arrangement of time-of-flight scintillation counters (TOF) for particle identification. Charged particle trajectories are reconstructed using the Belle SVD and CDC. A detailed description of the Belle detector can be found in ref. [10]. The Belle dataset used in this search was recorded between 2000 and 2010, and comprises $711 \, \text{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance, $121 \, \text{fb}^{-1}$ at the $\Upsilon(5S)$ resonance, $89 \, \text{fb}^{-1}$ recorded 60 MeV below the $\Upsilon(4S)$ (off-resonance), $28 \, \text{fb}^{-1}$ in energy scans above the $\Upsilon(4S)$ resonance, and the

remainder at and near the $\Upsilon(1S, 2S, 3S)$ resonances. The Belle II dataset, recorded between 2019 and 2022, includes 366 fb^{-1} at the $\Upsilon(4S)$, 42 fb^{-1} 60 MeV below it, and 19 fb^{-1} from a scan around a center-of-mass energy of 10.75 GeV. The resulting combined data sample has an integrated luminosity of 1408 fb^{-1}, corresponding to 1.3 billion $e^+e^- \to \tau^+\tau^-$ events [16].

Monte-Carlo (MC) simulated events are used to optimize the selection and background rejection and to measure the signal efficiency. To study the signal process in Belle (Belle II), we use 400 thousand (1 million) $e^+e^- \to \tau^+\tau^-(\gamma)$ events, where both initial- and final-state photon radiation is included. The signal tau decays via a phase space model to an electron or muon and a K_S^0 and the other tau decays to a SM-allowed decay. The potential background processes studied using simulation include $e^+e^- \to q\bar{q}$ events, where q indicates a u, d, c, or s quark; $e^+e^- \to b\bar{b}$ events; $e^+e^- \to \ell^+\ell^-(\gamma)$, where $\ell = e, \mu$ or τ; the two-photon processes $e^+e^- \to e^+e^-+$hadrons, where the hadrons can be h^+h^- ($h = \pi, K$ or proton) or the fragmentation products from a $q\bar{q}$ pair; and four-lepton processes: $e^+e^- \to e^+e^-e^+e^-$, $\mu^+\mu^-\mu^+\mu^-$, $\mu^+\mu^-e^+e^-$, $e^+e^-\tau^+\tau^-$, $\mu^+\mu^-\tau^+\tau^-$. A full list of simulated processes used for Belle and Belle II can be found in table 5 in appendix A. We simulate the $e^+e^- \to \tau^+\tau^-(\gamma)$ process using the KKMC generator [17], with subsequent tau decays simulated by the TAUOLA [18] package, with final state radiation (FSR) added by the PHOTOS [19] package. KKMC is also used to simulate $\mu^+\mu^-(\gamma)$ and $q\bar{q}$ production. Fragmentation of $q\bar{q}$ pairs is simulated using the PYTHIA [20] package. For the production and decay of $e^+e^- \to b\bar{b}$ events, we use PYTHIA interfaced with the EvtGen [21] generator. Different versions of the same generator packages are deployed for Belle and Belle II, which results in different samples for the simulated background processes for the two experiments and therefore different starting points for selection optimization. In Belle (Belle II) the BHLUMI [22] (BabaYaga@NLO [23–27]) generator is used to simulate $e^+e^- \to e^+e^-(\gamma)$ events. Two-photon processes are simulated using the AAFH [28–30] and TREPS [31] packages.

The Belle II analysis software framework (basf2) [32, 33] uses the GEANT4 [34] package to simulate the detector response to particles traversing the active volume. For the simulation of the detector response in Belle, GEANT3 [35] is used. Belle collision and simulation data are converted into the Belle II format for basf2 compatibility using the B2BII framework [36]. The online event selection (hardware trigger) for Belle and Belle II data is based on the energy deposits (clusters) and their topologies in the ECL, or on an independent trigger selection based on the number of charged particles reconstructed in the CDC. Most of the events are selected by requiring a total ECL energy larger than 1 GeV and a topology incompatible with Bhabha events.

3 Event selection and background rejection

3.1 Candidate reconstruction and event selection

We search for $e^+e^- \to \tau^+\tau^-$ events where one tau decays into the LFV channel $\tau^- \to \ell^- K_S^0$ and the other into a one-prong final state. In the center-of-mass (c.m.) frame of the electron-positron collision, the τ leptons are emitted in opposite directions, with the decay products of each τ confined to opposite hemispheres. The partition into hemispheres is delineated by the plane perpendicular to the estimated flight axis of the τ pair, which is determined

experimentally as the direction $\hat{\mathbf{t}}$ that maximizes the thrust value:

$$T = \max_{\hat{\mathbf{t}}} \left(\frac{\sum_i \left| \mathbf{p}^*_i \cdot \hat{\mathbf{t}} \right|}{\sum_i |\mathbf{p}^*_i|} \right), \tag{3.1}$$

where $\mathbf{p}^*_i$ represents the momentum of the final state particle i in the e^+e^- c.m. frame [37, 38], including both charged and neutral particles. Quantities in the e^+e^- c.m. frame are marked with an asterisk throughout this paper.

We designate the signal hemisphere as the one containing the $\tau^- \to \ell^- K_S^0$ decay candidate, formed by combining a lepton with a K_S^0. We reconstruct $K_S^0 \to \pi^+\pi^-$ candidates using two oppositely charged tracks assumed to be pions. The invariant mass of the K_S^0 candidates must satisfy $0.45 < M_{\pi^+\pi^-} < 0.55$ GeV/c^2. The pions are then fit to a common vertex, and the significance of the distance between the pion vertex and the interaction point, defined as the flight length divided by its uncertainty, must exceed 3.

Lepton candidates, as well as all remaining charged particles not used for K_S^0 reconstruction, are required to have a transverse momentum $p_T > 0.1$ GeV/c^2 and must originate within 3 cm along the z axis and 1 cm in the transverse plane from the e^+e^- interaction point. Charged particles not identified as electrons or muons are assumed to be pions. Muons in Belle II are identified using the discriminator $\mathcal{P}_\mu = \mathcal{L}_\mu/(\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d)$, where the likelihoods $\mathcal{L}_i$ for each charged-particle hypothesis ($i = e, \mu, \pi, K$, proton or deuteron) combine particle-identification information from CDC, TOP, ARICH, ECL, and KLM sub-detectors. For Belle II electrons, the output of a classifier based on a BDT, $\mathcal{P}_e$, is used. The BDT incorporates likelihoods from individual subdetectors and supplementary ECL observables, including variables that are sensitive to shower development [39]. For Belle II, we retain electrons and muons with $\mathcal{P}_{e,\mu} > 0.95$, which correspond to identification efficiencies of 96.3 % for electrons and 91.4 % for muons. The corresponding pion misidentification probabilities in Belle II are 0.3 % and 2.9%, respectively. The muon identification in Belle uses information from the KLM and extrapolated tracks to form a likelihood-based discriminator $\mathcal{P}'_\mu = \mathcal{L}_\mu/(\mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K)$, [40]. Electrons in Belle are identified using a likelihood ratio $\mathcal{P}'_e = \mathcal{L}_e/(\mathcal{L}_\mu + \mathcal{L}_{non-e})$ [41] based on information from the CDC, ACC and ECL. We retain candidates with $\mathcal{P}'_{e,\mu} > 0.9$, resulting in identification efficiencies in Belle of 93.5 % for electrons and 88.6 % for muons and corresponding pion misidentification rates of 0.5 % for electrons and 2.7 % for muons.

The vertices of τ and K_S^0 candidate are fitted with the TreeFitter tool [42], which updates the momenta of the reconstructed parent particles and the vertex positions of the individual tracks in the fit. Candidates with a successfully converged vertex fit result and a K_S^0 candidate mass $0.45 < M_{\pi^+\pi^-} < 0.55$ GeV/c^2 are selected.

The tag tau is reconstructed as an electron, muon or pion in the hemisphere opposite to the signal tau candidate. The total number of tracks in the event must be four, and their assigned charges must sum to zero.

In addition to the signal and tag tau reconstruction, all charged and neutral particles in the events are used to compute event-based observables that can be used to reduce the backgrounds. Photons are reconstructed from ECL clusters within the CDC acceptance with no tracks in the vicinity of the cluster. For π^0 reconstruction, photons that leave an energy

		Belle II		Belle	
		mean	δ	mean	δ
eK_S^0	$M(\mathrm{MeV}/c^2)$	1776.79 ± 0.36	11.91 ± 0.82	1777.39 ± 1.62	12.41 ± 0.16
	$\Delta E(\mathrm{MeV})$	-2.25 ± 0.20	49.43 ± 1.12	-1.25 ± 0.22	50.68 ± 2.57
μK_S^0	$M(\mathrm{MeV}/c^2)$	1777.15 ± 0.22	8.55 ± 0.95	1777.75 ± 0.80	10.56 ± 3.07
	$\Delta E(\mathrm{MeV})$	-1.05 ± 0.17	40.87 ± 0.91	0.55 ± 2.39	45.64 ± 2.51

Table 1. Fitted means and resolutions for $M(\ell K_S^0)$ and $\Delta E(\ell K_S^0)$ for both signal channels on Belle II and Belle simulation.

deposit of at least 0.1 GeV are combined in pairs and required to have an invariant mass in the range $0.115 < M_{\gamma\gamma} < 0.152$ GeV/c^2, which corresponds to approximately ± 2.5 units of resolution about the known π^0 mass [43]. Photons that contribute to the reconstructed π^0 candidates, as well as all photons with energies exceeding 0.1 GeV and all tracks, are used to calculate variables related to event kinematics, such as the missing momentum, missing mass, or the thrust axis.

A dedicated correction for electron bremsstrahlung energy loss is applied during reconstruction. All photons with energies $E_\gamma > 0.02$ GeV within a cone of 0.05 (0.15) radians around the direction of the electron momentum for Belle (Belle II) data are added to the electron energy.

Since the $\tau^- \to \ell^- K_S^0$ decay is a neutrinoless process, the invariant mass $M(\ell K_S^0)$ of the reconstructed τ decay products should coincide with the τ lepton mass [43], except for decays affected by final state radiation (FSR). In the c.m. system, the τ energy E_τ^* should be half of the $e^+ e^-$ energy, $\sqrt{s}/2$, apart from corrections due to initial state radiation (ISR) from the $e^\pm$ beams and FSR. Thus, the energy difference $\Delta E(\ell K_S^0) = E_\tau^* - \sqrt{s}/2$ should be near zero. The distribution of signal candidates in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane (see figure 1) is broadened by detector resolution and radiative effects. Initial state photon emission yields a tail towards lower $\Delta E(\ell K_S^0)$ values, while photons from FSR create a diagonal band, primarily at lower $M(\ell K_S^0)$ and $\Delta E(\ell K_S^0)$ values.

The analysis is performed in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane. We define several rectangular regions for use in optimizing the selection. These rectangular boxes are centered around the expected signal peak in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane, with side lengths proportional to their corresponding resolutions, δ. For each variable, δ is approximated as the standard deviation of the sum of two Gaussians and a Crystal Ball function [44] fitted to the simulated signal distribution. The fitted means and resolutions are shown in table 1.

Only events that lie within a rectangular region of width $\pm 20\delta$ are retained for the optimization. The final yield extraction is carried out within an elliptical signal region (SR) whose orientation and size is optimized for signal efficiency and background rejection. The optimized widths used for the final signal yields are 2δ for both semi-axes, for all channels, except for the Belle II electron channel, whose major semi-axis is 3δ wide (described in section 3.2). Events falling inside this elliptical SR in the data are masked during the selection optimization to avoid experimental bias.
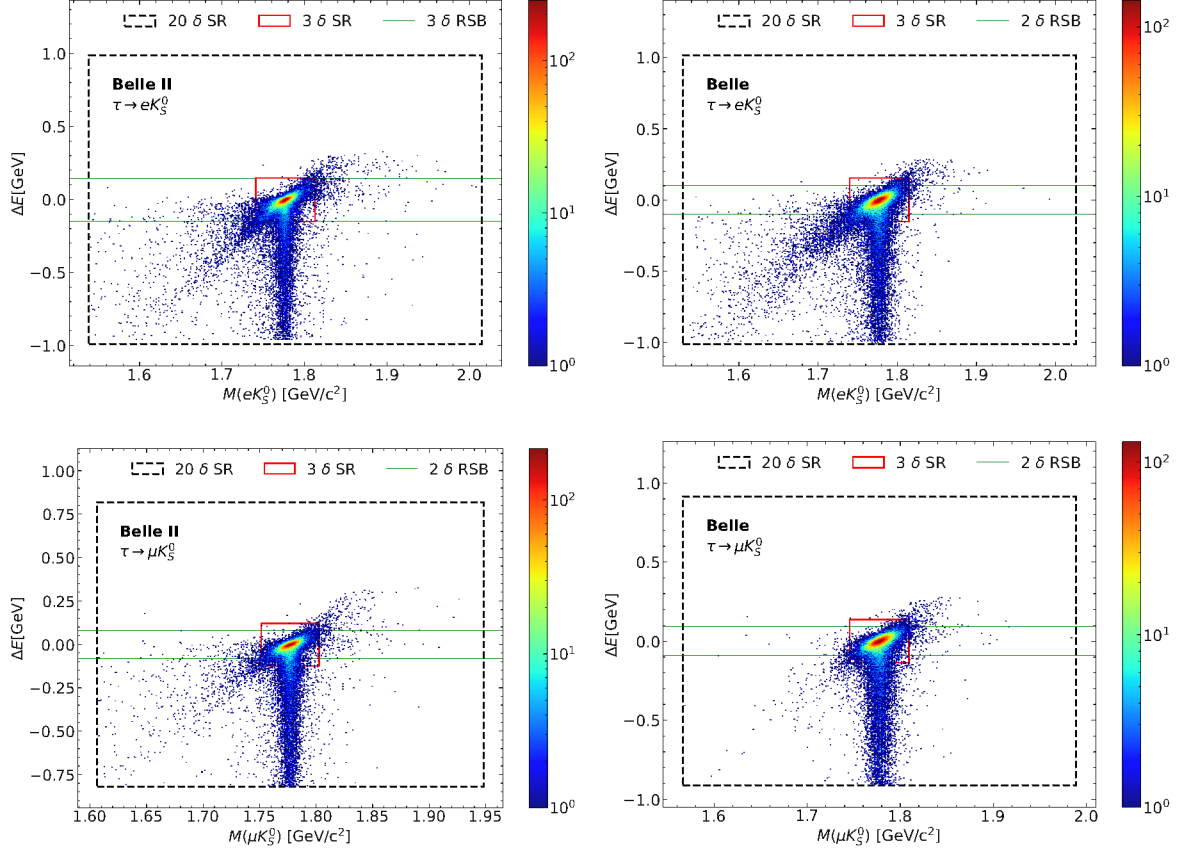
Figure 1. Distributions of simulated $\tau^- \to e^- K_S^0$ events for Belle II (left) and $\tau^- \to e^- K_S^0$ for Belle (right) in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane. The lower row shows the $\tau^- \to \mu^- K_S^0$ channel. The large, dashed black box represents the $20\,\delta$ region, the small red rectangle is the SR and the green bands define the RSB.

A sideband region (SB), corresponding to the events falling within the $\pm 20\,\delta$ rectangle but outside the $\pm 3\,\delta$ rectangle in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane, is defined to validate the background rejection. The final number of expected background events is extracted from a fit to the data in the reduced sideband (RSB), defined as the region within $\pm 3\delta$ ($\pm 2\delta$) of the $\Delta E(\ell K_S^0)$ peak for the Belle II electron mode (the other signal modes). A larger RSB is chosen for the Belle II electron mode since it has lower expected background levels than the other modes. The data yields in the SB and RSB are consistent with the expected yields in simulation. These regions are shown for the Belle and Belle II electron and muon channels in figure 1.

After the candidate reconstruction, the most important backgrounds for the electron channel arise from low-multiplicity processes such as $e^+e^- \to e^+e^-(\gamma)$, $e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$, $e^+e^-h^+h^-$ and $\mu^+\mu^-(\gamma)$, while the background for the muon channel is dominated by $e^+e^- \to q\bar{q}$ processes.

We first apply a set of selection requirements (preselection) to remove obvious background events and then a dedicated BDT-based selection requirement to further suppress remaining background. The preselection requirements are chosen by comparing the simulated signal

distributions with the background distributions obtained from the sideband regions in data and in simulation, separately for Belle and Belle II and for each tag-side decay type. As mentioned previously in section 2, the Belle and Belle II experiments use different MC simulations. As a consequence, selection optimization results in different preselection requirements and differences in input features for the BDT selections, as detailed in section 3.2. In order to remove low-multiplicity events with converted photons, we impose a minimum requirement on the K_S^0 mass reconstructed using the electron mass hypothesis for the pions, $M^{ee}(K_S^0)$. Since low-multiplicity events consist only of detectable particles, these events can be discarded using variables related to the missing energy and momentum, the tag-side mass and ΔE. The projections of the lepton and tag-side track momentum on the z axis are also used, as the background from low-multiplicity events tends to be aligned more closely with the beam axis than are signal events. Continuum background events are rejected using event shape properties like the thrust value, the numbers of photons and neutral pions, or characteristics of the decay dynamics, like the energy of the lepton. The full set of preselections is shown in appendices B and C for the electron and muon channels, respectively, comparing the different requirements applied for Belle and Belle II. The distributions of the pair-converted invariant mass $M^{ee}(K_S^0)$ (upper row) and event thrust (lower row) are shown before (left) and after (right) applying the preselection requirements in figure 2. Hereafter, all plots sum over the three tag-side categories for each signal channel.

We show the $M(\ell K_S^0)$ and $\Delta E(\ell K_S^0)$ distributions for the four different categories after preselection in figure 3. The overall signal efficiencies in the $20\,\delta$ region after reconstruction and preselection obtained from Belle (Belle II) simulation are 13.6 (13.9)% and 17.2 (16.6)% for the electron and muon channels, respectively. These values are obtained after correcting the simulation to account for the mismodelling of the detector response, which affects the lepton identification.

3.2 BDT-based background rejection

To reduce the residual background contamination, primarily due to $e^+e^- \to q\bar{q}$ processes, a BDT is trained using the XGBoost library [45] on simulated signal and background samples. The BDT incorporates 34 variables (listed in appendix D) pertaining to both signal- and tag-side τ kinematics, K_S^0 and track kinematics, event shape properties, and variables associated with photons and neutral pions in the event.

The first set of variables includes the τ properties: the beam-constrained mass of the tag-side τ, the invariant mass hypothesis for the track in the tag side, as well as the invariant mass of the tag-side particles (the track and any tag-side photons), while for the signal τ it consists of the transverse momentum (p_T), flight time and distance and their respective measurement uncertainties. Additionally, it includes the angle between the lepton and the K_S^0. Variables related to the K_S^0 candidate are the flight time and distance with their uncertainties, the momentum and energy in the c.m. system, and two hypothetical invariant masses computed under the assumption that either the first or second daughter is a proton, to avoid contamination from other long-lived particles decays as $\Lambda^0 \to \pi^- p^+$. Track-related variables consist of the lepton momentum and the ranked p^T values of the three signal-side tracks, all in the c.m. system. The event characteristics included are the missing mass, missing energy,
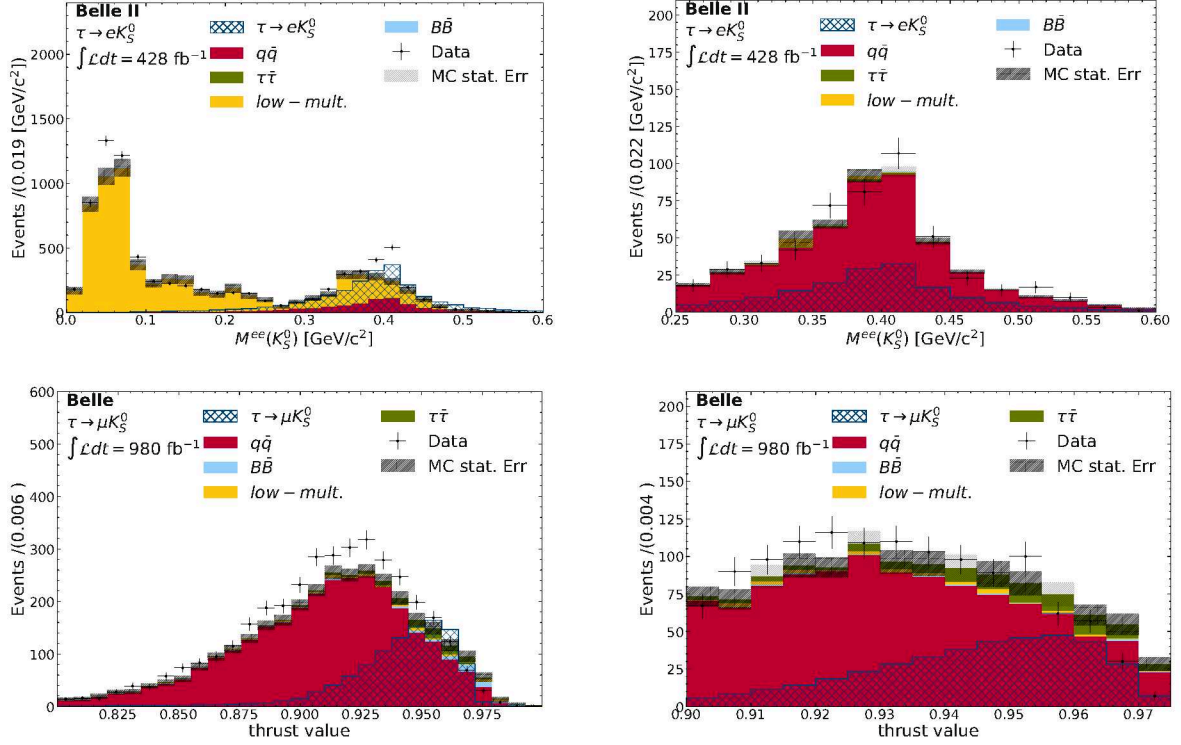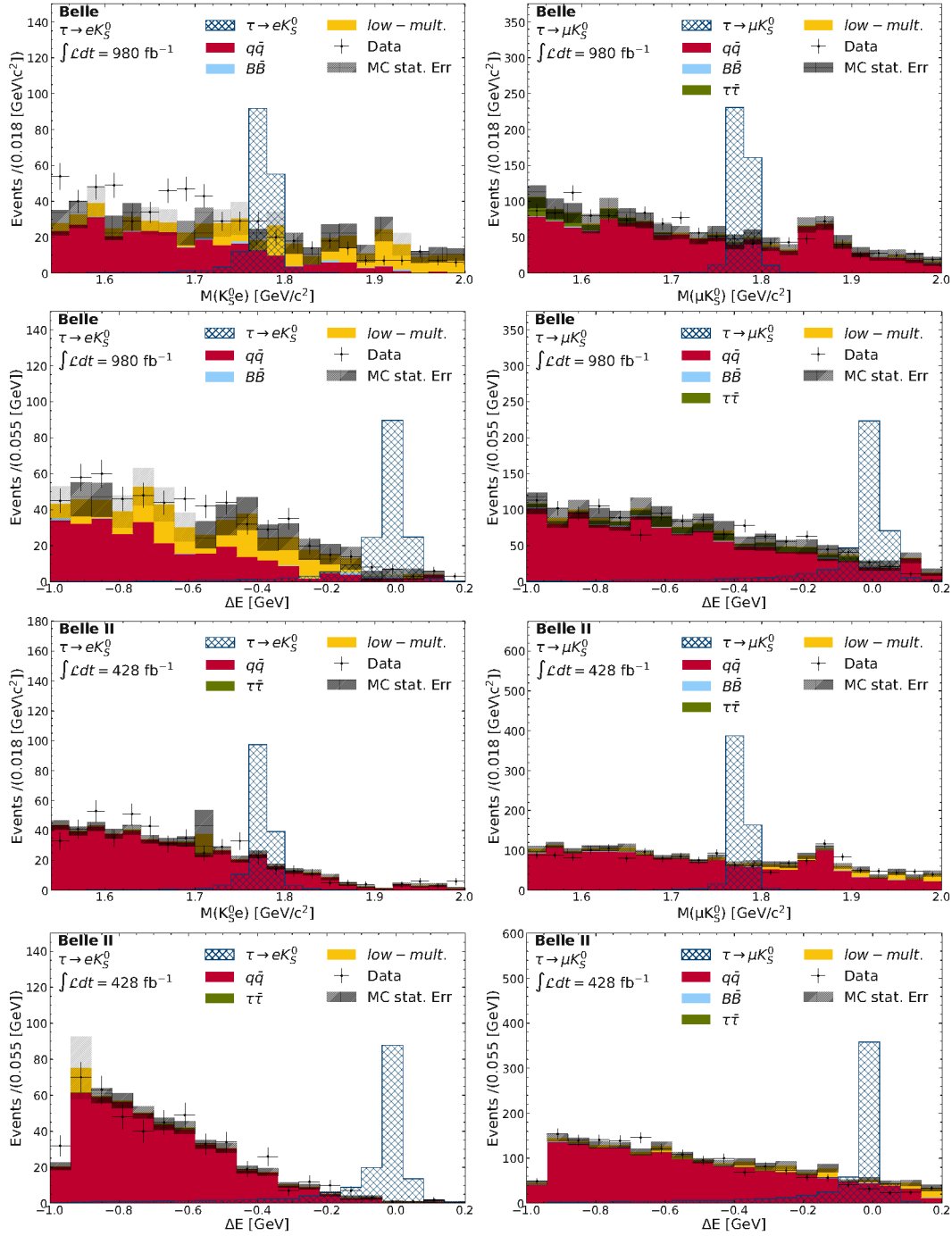
Figure 2. Comparison between data (black points with error bars) and background simulation (solid-filled histograms) in the sidebands before (left) and after (right) pre-selections are applied to reject low-multiplicity events. The signal in the 20δ region is shown as a blue hatched histogram with an arbitrary scale. The statistical uncertainties are displayed in light gray shading. The upper row displays the K_S^0 mass reconstructed using the electron mass hypothesis for the pions, $M^{ee}(K_S^0)$, for the Belle II electron channel. The lower row displays the thrust value for the Belle muon channel.

missing momentum and missing transverse momentum of the event in the c.m. frame. We also use the cosine of the angle between the missing momentum and the tag-side track, and between the missing momentum and the signal-side lepton, as well as the visible energy in the c.m. and the thrust value. The last category of variables contributing to the BDT classifier includes the number of photons and their total energy on the signal- and tag-sides, along with the total photon energy in the event and the number of π^0 candidates.

Four BDTs are separately trained, one for each of the two final states in Belle and Belle II, on samples corresponding to 1/3 of the Belle (Belle II) simulation statistics within the 20δ region for the background and about 12000 simulated signal events. The BDT parameters are optimized using the `Optuna` library [46], which minimizes the logarithmic loss function. The BDT is validated using an independent data sample, ensuring that the signal retention and background rejection rates remain comparable to those of the training sample, thus mitigating the risk of overfitting. Both the training and validation samples are merged into a combined sample. This combined sample is then used to optimize simultaneously the size of the elliptical SR in units of δ and the BDT classifier requirement for each tag-side category using the Punzi figure of merit. This figure of merit [47] is defined as $\frac{\varepsilon_{\ell K_S^0}}{\alpha/2+\sqrt{B}}$, where $\varepsilon_{\ell K_S^0}$ is the signal efficiency and B is the number of background events in the elliptical SR region, respectively.

Figure 3. Comparison between data (black points with error bars) and background simulation (solid-filled histograms) in the sideband regions after preselections for the $\Delta E(\ell K_S^0)$ and $M(\ell K_S^0)$ distributions, for both Belle (first and second rows) and Belle II (third and fourth rows) data sets. Only backgrounds with a non-null residual contribution after the selection is applied are shown in the plots. The signal in the 20δ region is shown as a blue hatched histogram with an arbitrary scale. The statistical uncertainties are displayed in light gray shading. The left column shows the distributions for electron channels and the right column for the muon channels.
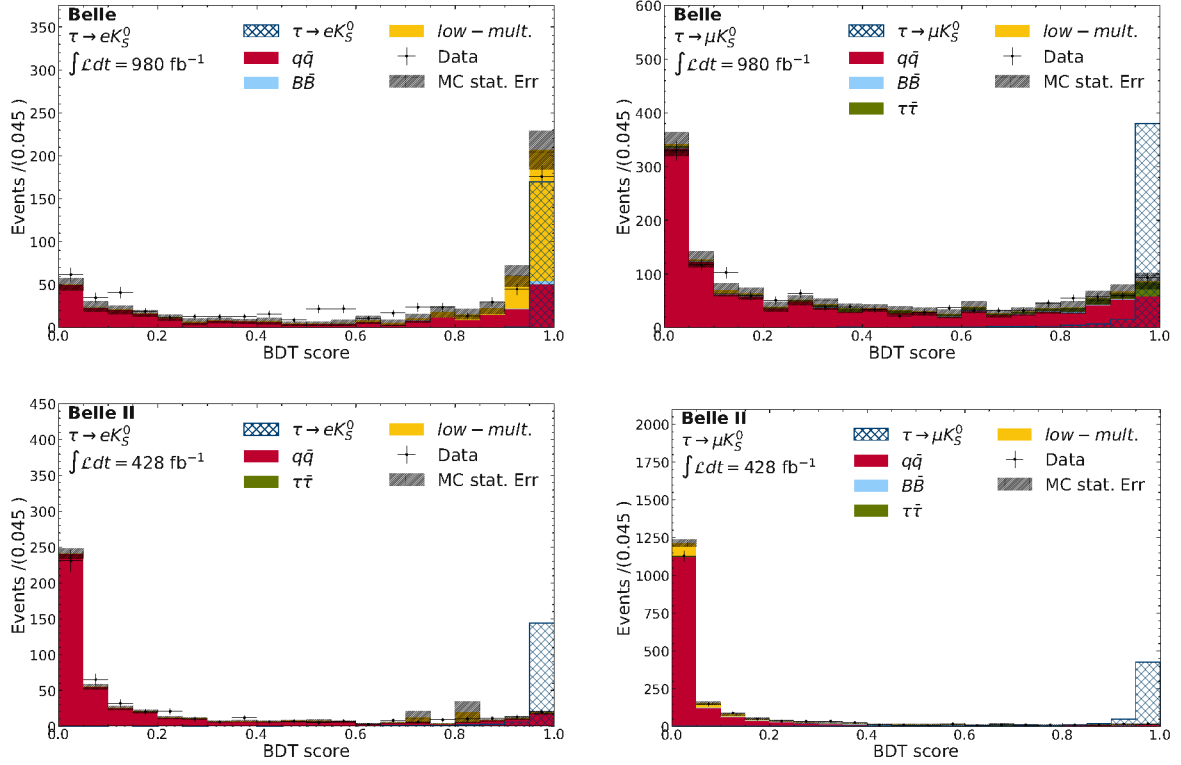
Figure 4. Comparison between data (black points with error bars) and simulation (solid-filled histograms) for the BDT output distribution for events in the sideband region after preselection. The signal is shown as a blue hatched histogram with arbitrary scale. The statistical uncertainties are displayed as gray shaded areas. Top row shows the electron (left) and muon (right) channel for Belle, the bottom row shows electron (left) and muon (right) channel for Belle II.

We set α to 3, corresponding to 3σ signal significance. As noted in section 3.1, the input features and requirements for the BDT classifiers differ for each channel and experiment since the working points for each BDT are different, due to the different MC simulations used in Belle and Belle II for the optimization of the pre-selections that are applied before the BDT training. Because the dominant background source depends on the identity of the single charged track reconstructed in the opposite (tag) hemisphere, the BDT working point is optimised independently for each tag category — electron-tag (e−tag), muon-tag (μ−tag) and pion-tag (π−tag) — and for each experiment (Belle, Belle II). The optimisation maximises the Punzi figure-of-merit in the 3σ signal ellipse. The resulting thresholds therefore span the full BDT output range. For example, in Belle II, the μ-tag sample for the electron channel is already background-free after the pre-selection, so the optimum is applying no further BDT; in Belle data the π−tag samples still contain large continuum contamination for both electron and muon signal channels and require a tighter selection for a BDT output > 0.99; intermediate values (0.5–0.95) are found for the remaining categories (signal channel, tag-side identity and experiments). These tag-specific working points equalise the residual background across categories while retaining, on average, $(10 \pm 1)\%$ of the signal after all selections. The summed BDT distribution over all three tag categories are displayed for sideband data and

	eK_S^0			μK_S^0		
	$\epsilon_{eK_S^0}$ [%]	N_{RSB}^{data}	N_{RSB}^{MC}	$\epsilon_{\mu K_S^0}$ [%]	N_{RSB}^{data}	N_{RSB}^{MC}
Belle	10.4	6	$8.3^{+4.0}_{-2.8}$	10.2	8	$9.4^{+4.2}_{-3.0}$
Belle II	10.1	6	$4.1^{+3.2}_{-1.9}$	10.2	5	$10.5^{+4.3}_{-3.2}$

Table 2. Signal efficiencies in the elliptical SR, observed yields in data and expected yields in simulation in the RSB for the electron and muon channels in Belle and Belle II data.

simulation in figure 4. The BDT rejects 87% (72%) of the background events in the electron channel according to Belle (Belle II) simulation, and 96% (89%) of the background events in the muon channel. The overall signal efficiencies after the BDT requirements obtained from Belle (Belle II) simulation in the signal region are 10.4% (10.1%) and 10.2% (10.2%) for the electron and muon channels, respectively. Those efficiencies represent an improvement with respect to the previous Belle analysis between 38 and 45%.

3.3 Simulation validation and expected background

The agreement between data and simulation is checked using the SB region, defined as the black dashed rectangle in figure 1 without the SR box, and the RSB region (green bands in figure 1).

The outputs of the BDTs in the SB region for all channels are shown in figure 4, with the simulated signal overlaid as a hatched blue histogram. The signal efficiency and the yields in data and simulation in the RSB are given in table 2. Differences in the background components between the Belle and Belle II electron channels arise from different simulated backgrounds (see table 5 in appendix A) and preselections.

To estimate the number of expected events in the signal region after applying the BDT selection, we perform an extended unbinned maximum likelihood fit to the signal tau invariant mass distribution of the events retained in the data RSB region. We use an exponential function to model the background shape, $f(m) = N_{bkg}e^{Cm}$, where the free parameters in the fit are the shape parameter C and the background normalization N_{bkg} in the range $1.5 < M_{lK_S^0} < 2.0$ GeV/c^2. The results for each channel and experiment are shown in figure 5. The expected yields N_{exp} in the elliptical SR are obtained by integrating the fitted functions over the $\pm 3\delta$ intervals in $M_{lK_S^0}$ based on signal resolutions as defined in table 1 and approximated by multiplying the integral N_{bkg}^{SR} by the ratio of the elliptical and rectangular areas, f_{ell}, so that $N_{exp} = f_{ell}N_{bkg}^{SR}$. The inputs and the resulting yields with their statistical uncertainties propagated from the fits are listed in table 3.

4 Systematic uncertainties

Systematic uncertainties can affect the branching fraction measurements through the modeling of the background used to extract N_{exp}, possible differences between experimental data and simulation that could impact the signal efficiency, and external inputs, such as the data luminosity and the tau pair cross-section.
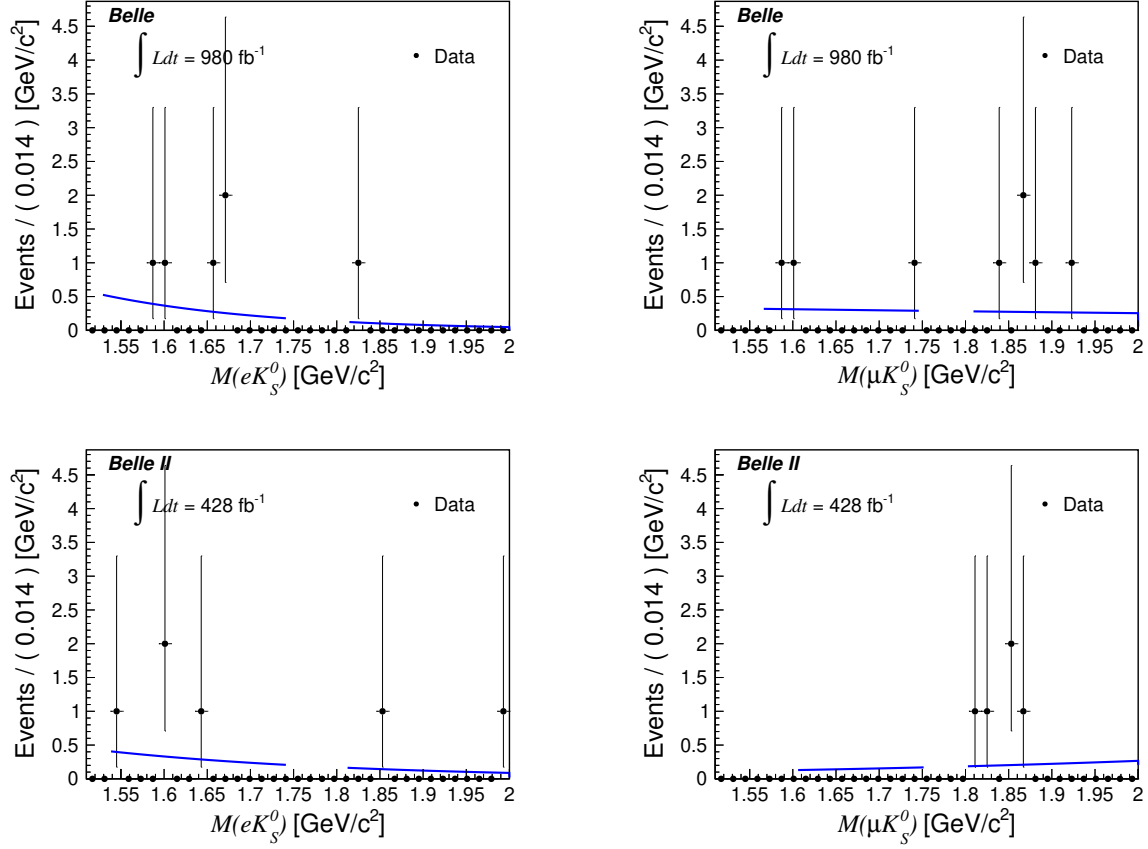
Figure 5. Fits (blue solid line) to selected events (points with error bars) in the RSB region as function of $M(\ell K_S^0)$. Top row shows the electron (left) and muon (right) channels for Belle, the bottom row shows the electron (left) and muon (right) channels for Belle II.

		$N_{\mathrm{bkg}}^{\mathrm{SR}}$	f_{ell}	N_{exp}
eK_S^0	Belle	$0.78^{+0.36}_{-0.28}$	0.557	$0.43^{+0.20}_{-0.16}$
	Belle II	$0.94^{+0.44}_{-0.33}$	0.453	$0.42^{+0.20}_{-0.15}$
μK_S^0	Belle	$1.29^{+0.51}_{-0.40}$	0.554	$0.71^{+0.28}_{-0.22}$
	Belle II	$0.65^{+0.34}_{-0.25}$	0.555	$0.36^{+0.19}_{-0.14}$

Table 3. Number of background events in the signal regions $N_{\mathrm{bkg}}^{\mathrm{SR}}$ as integrated from fits to Belle and Belle II data RSB regions, scaling factors f_{ell} and the number of expected background yields in the elliptical SR for the electron (top) and muon (bottom) channels. The uncertainties are statistical only.

The background function used in the RSB fits was modified from the exponential model to a linear function to test the impact of the assumed background shape on the background yield estimate. The differences between the yields obtained with the two functions is small compared to the statistical uncertainty of the fitted yields and has a negligible effect on the expected upper limits. Therefore, no systematic uncertainty in N_{exp} is assigned for the background modeling.

We take into account the systematic uncertainty associated with the corrections to the simulated lepton-identification efficiencies, derived from auxiliary measurements in data using $J/\psi \to \mu^+\mu^-$, $e^+e^- \to \ell^+\ell^-\gamma$, and $e^+e^- \to e^+e^-\mu^+\mu^-$ events. These corrections are obtained as functions of momentum, polar angle and charge, and applied to events reconstructed from simulation. The systematic uncertainty is obtained by varying the corrections by their statistical and systematic uncertainties and estimating the impact of these variations on the selection efficiency. Adding the statistical and systematic variations in quadrature, the result is a relative uncertainty in the signal efficiency of 2.35(2.41)% for the electron (muon) channel for Belle and 0.72(1.34)% for the electron (muon) channel for Belle II.

In Belle II the difference between data and simulation in the track-reconstruction efficiency is measured in $e^+e^- \to \tau^+\tau^-$ events with $\tau^- \to e^-\nu_e\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ to yield a 0.24% uncertainty per track for a total 0.96% relative uncertainty. For Belle, a 0.35% per-track uncertainty is assigned using $D^{*+} \to \pi^+D^0, D^0 \to \pi^+\pi^-K_S^0$ decays, resulting in a total relative signal efficiency uncertainty of 1.4%.

For the Belle II experiment, we use triggers provided by the ECL and CDC sub-detectors. In data, the trigger efficiency is evaluated using independent trigger selections: the efficiency of the ECL-based trigger selection is obtained using events triggered by the CDC, while the efficiency of the CDC-based trigger selection is evaluated using events passing the ECL trigger requirements. The level of agreement between data and simulation efficiencies is 0.5% for the ECL trigger selection and 4.3% for the CDC trigger selection. Given that the efficiency of trigger selections based on the ECL only is 88%, the weighted average of the data-simulation efficiency differences is computed to be 0.68%, which is taken as the systematic uncertainty. For Belle, we use a trigger efficiency uncertainty of 0.9% from ref. [48].

To correct for differences in K_S^0 reconstruction between data and simulation for Belle II, we compare K_S^0 yields in data and MC using the decay $\tau^- \to K_S^0\pi^-\nu$. The K_S^0 yields are obtained from fits to the $\pi^+\pi^-$ invariant mass in ten bins of the flight distance. The yield ratio is then fitted with a linear function, which is used to reweight the reconstructed K_S^0 in signal MC to evaluate the systematic uncertainty. Values of 5.96% and 5.31% are obtained for the electron and muon channels respectively. For Belle, $D^{*+} \to \pi^+D^0, D^0 \to \pi^+\pi^-K_S^0$ decays are used, leading to a correction of the simulated signal efficiency by a factor of 0.9789 per K_S^0 candidate and a 0.73 % contribution to the systematic uncertainty.

We obtain a systematic uncertainty on the signal efficiency in the $20\,\delta$ plane due to the BDT selection by applying the BDTs trained for the electron and muon channels to the reconstructed standard model decay $\tau^- \to K_S^0\pi^-\nu$. Employing the same BDT requirement as optimized for the respective signal channels, we calculate BDT selection efficiencies on both simulation and data. Their relative difference is used as a systematic uncertainty, giving 1.49% (1.59%) and 5.06% (5.37%) for the electron and muon channels, respectively, for Belle (Belle II).

Quantity	Source	eK_S^0 [%]		μK_S^0 [%]	
		Belle	Belle II	Belle	Belle II
$\epsilon_{lK_S^0}$	Lepton identification	2.35	0.72	2.41	1.34
	Tracking efficiency	1.40	0.96	1.4	0.96
	Trigger efficiency	0.90	0.68	0.9	0.68
	K_S^0 efficiency	0.73	5.96	0.73	5.31
	BDT efficiency	1.49	1.59	5.06	5.37
	Signal region	$+3.15$ -5.75	$+3.22$ -4.51	$+2.98$ -5.22	$+2.63$ -4.23
$\mathcal{L}$	Luminosity	1.4	0.5	1.4	0.5
$\sigma_{\tau\tau}$	Tau pair cross-section	0.3	0.3	0.3	0.3

Table 4. Relative systematic uncertainties in % for the Belle and Belle II datasets.

To assess a systematic uncertainty due to potential mismodeling of the signal resolutions in $\Delta E(\ell K_S^0)$ and $M(\ell K_S^0)$ variables, the size of the elliptical SR, whose values are given in table 1, is varied by $\pm 1\delta$ and the corresponding change in signal efficiency is used as systematic uncertainty. The check is performed on the final selected signal samples, where the signal peaks are mostly contained in the 1δ-wide SR and the signal leakage of non-Gaussian tails parametrized by the δ resolutions is below 6% for all channels. We find relative systematic uncertainties of $(+3.15, -5.75)\%$ and $(+2.98, -5.22)\%$ for the Belle electron and muon channels, respectively. The corresponding uncertainties for Belle II are $(+3.22, -4.51)\%$ and $(+2.63, -4.23)\%$.

The luminosities are determined independently using Bhabha and diphoton events. The differences between these determinations are taken as systematic uncertainties, yielding an average uncertainty of 0.5% for Belle II and 1.4% for Belle [49, 50]. The uncertainty on the production cross-section of tau pairs is evaluated in ref. [16] to be 0.003 nb.

In table 4, systematic uncertainties for Belle and Belle II analyses are summarized.

5 Result

The distribution of events in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane is shown in figure 6 for the $428\,\mathrm{fb}^{-1}$ Belle II and $980\,\mathrm{fb}^{-1}$ Belle data samples. When examining the signal region in data (unboxing), we observe 0(0) events in the Belle II electron (muon) channel and 0(1) events for the Belle electron (muon) channel in the signal region.

The $\tau^- \to \ell^- K_S^0$ branching fractions are obtained from the number of signal events N_{sig}, the signal efficiencies $\varepsilon_{\ell K_s^0}$ and the number of tau leptons produced N_τ:

$$\mathcal{B}(\tau^- \to \ell^- K_S^0) = \frac{N_{\mathrm{sig}}}{N_\tau \times \varepsilon_{\ell K_s^0}} = \frac{N_{\mathrm{obs}} - N_{\mathrm{exp}}}{\mathcal{L} \times 2\sigma_{\tau\tau} \times \varepsilon_{\ell K_s^0}}, \tag{5.1}$$

where $\mathcal{L}$ is the integrated luminosity of the data sets, N_{exp} is the number of expected background events and N_{obs} is the number of observed events. The τ-pair production cross-section $\sigma_{\tau\tau}$, determined from the weighted average of the cross-sections at the different
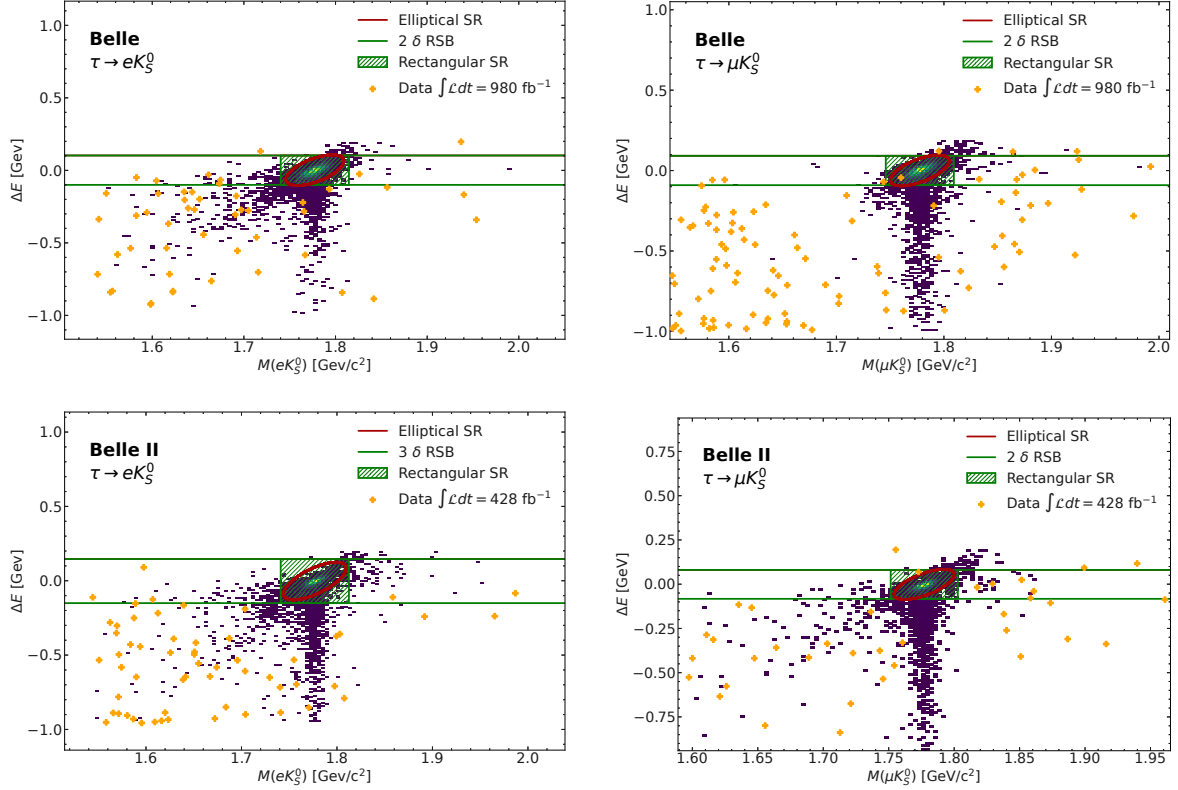
Figure 6. Scatter plots of selected events in the $(M(\ell K_S^0), \Delta E(\ell K_S^0))$ plane for signal simulation (violet) and data (orange). The elliptical SR is shown in red, the rectangular SR as a hatched green area and the RSB is indicated as green horizontal lines. Plots show distributions for the electron (left) and muon (right) channels for Belle (upper-row) and Belle II (lower-row). For the muon mode in Belle one data event is observed in the SR.

center-of-mass energies at which the data were taken, is 0.919 ± 0.003 nb for Belle II data and 0.916 ± 0.003 nb for Belle data.

As we do not observe any significant excess above the expected background within the signal region, we calculate 90 % C.L. upper limits on the $\tau^- \to \ell^- K_S^0$ branching fractions using the CL_s method [51, 52] in a frequentist approach implemented in the pyhf library [53, 54].

To determine the expected limit sensitivity, we generate 10000 pseudo-experiments at 50 points uniformly distributed in the branching ratio range of $(0-4) \times 10^{-8}$ in two bins, one for each experiment, each with their respective signal efficiencies and expected background yields. The total statistical and systematic uncertainties affecting each experimental input, as discussed in section 4, are combined in quadrature.

Figure 7 displays the CL_s curves computed as a function of the branching fractions for the combined Belle and Belle II datasets for the $\tau^- \to e^- K_S^0$ and $\tau^- \to \mu^- K_S^0$ decays. The dashed black line represents the expected CL_s, while the green and yellow bands show the $\pm 1\sigma$ and $\pm 2\sigma$ contours, respectively.

The expected limits, assuming an observed number of events consistent with the background estimation in table 3, are 0.9×10^{-8} and 1.2×10^{-8} at 90 % C.L. for the electron and muon channel, respectively. The observed limits in data after unboxing are 0.8×10^{-8} and 1.2×10^{-8} at 90 % C.L. for the electron and muon channels, respectively.
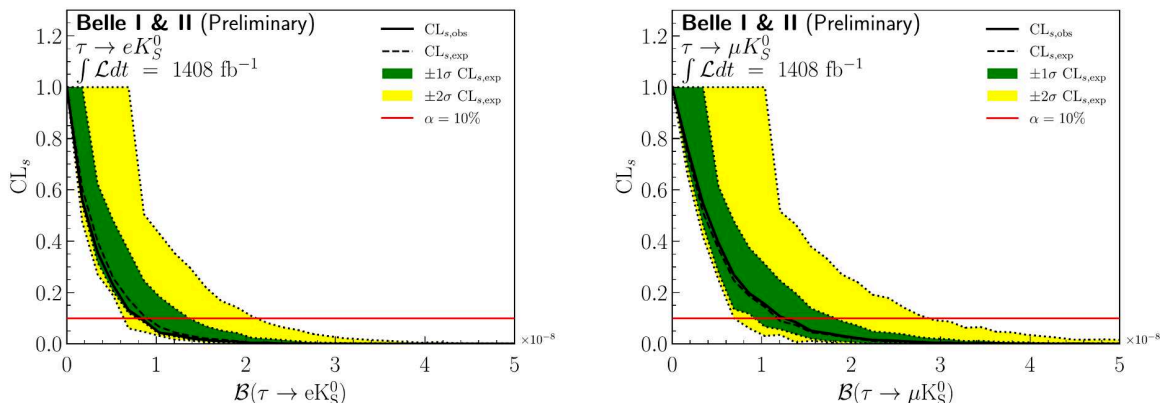
Figure 7. Observed (solid black curve) and expected (dashed black curve) CL_s as a function of the assumed branching fractions for $\tau^- \to e^- K_S^0$ (left) and $\tau^- \to \mu^- K_S^0$ (right) decays. The red line corresponds to the 90 % C.L.

6 Summary

We present a search for the LFV decays $\tau^- \to e^- K_S^0$ and $\tau^- \to \mu^- K_S^0$ using $428\,\mathrm{fb}^{-1}$ of data collected by the Belle II experiment and $980\,\mathrm{fb}^{-1}$ of data collected by the Belle experiment, which is the world's largest tau pair data set. A set of dedicated Boosted Decision Tree classifiers are used to discriminate signal decays from background processes. The signal yield is determined in the plane of the reconstructed τ mass and the difference between the reconstructed and expected τ energy, two variables in which the signal decays peak. We observe 0(0) events for the electron channel and 0(1) events for the muon channel in the signal region for Belle II(Belle) and thus set 90 % C.L. upper limits on both channels computed in a frequentist approach. The observed limits at 90% C.L. are $0.8{\times}10^{-8}$ for the electron and $1.2{\times}10^{-8}$ for the muon channel, 3.3 and 1.9 times more stringent, respectively, than the previous best limits of $2.6{\times}10^{-8}$ and $2.3{\times}10^{-8}$ [9].

A   Simulated MC samples in Belle and Belle II

MC sample	Belle	Belle II
$e^+e^- \to \tau^+\tau^-$	5.79	7
$e^+e^- \to \tau^+\tau^-\tau^+\tau^-$		10
$e^+e^- \to \mu^+\mu^-(\gamma)$	4.685	1
$e^+e^- \to \mu^+\mu^-\mu^+\mu^-$		2
$e^+e^- \to \mu^+\mu^-\tau^+\tau^-$		2
$e^+e^- \to e^+e^-(\gamma)$	0.52	0.12
$e^+e^- \to e^+e^-\mu^+\mu^-$	4.685	0.2
$e^+e^- \to e^+e^-\tau^+\tau^-$		2
$e^+e^- \to e^+e^-e^+e^-$	4.685	0.2
$e^+e^- \to e^+e^-u\bar{u}$	5.79	
$e^+e^- \to e^+e^-d\bar{d}$	5.79	
$e^+e^- \to e^+e^-s\bar{s}$	5.79	
$e^+e^- \to e^+e^-c\bar{c}$	5.79	
$e^+e^- \to e^+e^-\pi^+\pi^-$		1
$e^+e^- \to e^+e^-K^+K^-$		2
$e^+e^- \to e^+e^-p\bar{p}$		2
$e^+e^- \to q\bar{q}$	4.155	7
$B\bar{B}$	6.952	1

Table 5. MC samples simulated in Belle and Belle II (ab^{-1}).

B   Preselections for $\tau \to e K_S^0$ decay

Here we list the different requirements applied to the Belle and Belle II data as pre-selections for the electron channel, $\tau \to e K_S^0$. Angle requirements are expressed in unit of radians.

Belle Selections	Belle II Selections
Electron-tag	
$\cos(p_{\mathrm{miss}}^*, p_{\mathrm{tag}}^*) > 0$ $N_{\pi^0}^{\mathrm{tot}} = 0$ $\Delta E_{\mathrm{tag}} < -1\,\mathrm{GeV}$ $M^{ee}(K_S^0) > 0.2\,\mathrm{GeV/c}^2$ $M_{\mathrm{tag}}(e\gamma) < 7\,\mathrm{GeV/c}^2$	$0.3 < \theta_{\mathrm{miss}} < 2.7$ $0.49 < M_{\pi^+\pi^-} < 0.505\,\mathrm{GeV/c}^2$ $\cos(\theta_{K_S} - \theta_{\hat{t}}) > 0.8$ $E_{K_s} > 1\,\mathrm{GeV}$ $|p_{z\mathrm{tag}}^*| < 2.5\,\mathrm{GeV/c}$
Muon-tag	
$N_{\pi^0}^{\mathrm{tag}} = 0$ $M_{\mathrm{tag}}(e\gamma) < 0.2\,\mathrm{GeV/c}^2$	$N_{\pi^0}^{\mathrm{tag}} = 0$ $p_{T,\mathrm{third}}^* > 0.1\,\mathrm{GeV/c}$
Pion-tag	
$M^{ee}(K_S^0) > 0.2\,\mathrm{GeV/c}^2$	$-3 < p_z^*(\ell) < 3\,\mathrm{GeV/c}$
All tags	
$p_{\mathrm{miss}} > (-4 \times M_{\mathrm{miss}}^2 - 1)$ $p_{\mathrm{miss}} > (1.3 \times M_{\mathrm{miss}}^2 - 0.8)$ $N_\gamma^{\mathrm{tot}} < 3$	$M^{ee}(K_S^0) > 0.25\,\mathrm{GeV/c}^2$ $\cos(\theta_{\mathrm{miss}}^*, \theta_{\mathrm{tag}}^*) > 0$ $\Delta E_{\mathrm{tag}} < 1\,\mathrm{GeV}$ $0.85 < thrust < 0.98$

C Preselections for $\tau \to \mu K_S^0$ decay

Here we list the different requirements applied to the Belle and Belle II data as pre-selections for the muon channel, $\tau \to \mu K_S^0$. Angle requirements are expressed in unit of radians.

Belle Selections	Belle II Selections
Electron-tag	
• $N_\gamma^{\mathrm{tag}} < 2$	• $E_\ell > 0.5\,\mathrm{GeV}$ • $N_{\pi^0}^{\mathrm{tot}} = 0$ • $\chi_{POCA}(\tau) < 10$
Muon-tag	
• $M^{ee}(K_S^0) > 0.2\,\mathrm{GeV/c^2}$ • $\cos(p_{\mathrm{miss}}^*, p_{\mathrm{tag}}^*) > 0$ • $M_{\mathrm{bc}}^{\mathrm{tag}} > 3\,\mathrm{GeV/c^2}$ • $N_{\pi^0}^{\mathrm{tot}} = 0$	
Pion-tag	
• $N_{\pi^0}^{\mathrm{sig}} = 0$	• $E_\ell > 0.8\,\mathrm{GeV}$ • $\theta(\ell, K_s) < 1$ (angle between lepton and K_S^0) • $p_{T,\mathrm{sub}}^* > 0.3\,\mathrm{GeV/c}$ • $\cos(\theta_{\mathrm{miss}}^* - \theta_{\mathrm{tag}}^*) > 0$
All tags	
• $N_{\pi^0}^{\mathrm{tot}} < 2$ • $E_{\mathrm{tag}}^* < 5\,\mathrm{GeV}$ (energy of the tag track in the c.m.) • $0.9 < thrust < 0.975$	

D BDT input variables

Signal tau variables
Transverse momentum of largest $p_{T,lead}$, middle $p_{T,sub}$ and lowest $p_{T,third}$ signal-side track in c.m.
Signal tau p_T
Signal tau flight time and its uncertainty
Signal tau flight distance and its uncertainty
Angle between K_S^0 and lepton
K_S^0 mass calculated with proton mass hypothesis for first pion
K_S^0 mass calculated with proton mass hypothesis for second pion
Energy of K_S^0
Momentum of K_S^0 in c.m.
K_S^0 Error of the flight distance
K_S^0 flight distance
K_S^0 Error of flight Time
K_S^0 flight time
Momentum of lepton in c.m.
Tag tau variables
Mass of tag-side tau
Beam constrained mass of tag-side tau
Invariant mass of tag side track and photons
Event based variables
Transverse missing momentum of event in c.m.
Visible energy of event in c.m.
Missing mass squared of event
Thrust value
Total energy of photons in event
Number of photons on the tag side
Total energy of photons on tag side
Number of neutral pions
Missing momentum of event in c.m.
Missing energy of event in c.m.
Cosine of the angle between missing momentum and tag-side track in c.m.
Cosine of the angle between missing momentum and signal-side lepton
Number of photons in the event

Table 6. Variables used for BDT training. Ordering follows the feature importance in the training.

References

[1] T. Li, M.A. Schmidt, C.-Y. Yao and M. Yuan, *Charged lepton flavor violation in light of the muon magnetic moment anomaly and colliders*, *Eur. Phys. J. C* **81** (2021) 811 [`arXiv:2104.04494`] [InSPIRE].

[2] G. Hernández-Tomé, G. López Castro and P. Roig, *Flavor violating leptonic decays of τ and μ leptons in the Standard Model with massive neutrinos*, *Eur. Phys. J. C* **79** (2019) 84 [*Erratum ibid.* **80** (2020) 438] [`arXiv:1807.06050`] [InSPIRE].

[3] P. Blackstone, M. Fael and E. Passemar, *$\tau \to \mu\mu\mu$ at a rate of one out of 10^{14} tau decays?*, *Eur. Phys. J. C* **80** (2020) 506 [`arXiv:1912.09862`] [InSPIRE].

[4] A.A. Petrov and D.V. Zhuridov, *Lepton flavor-violating transitions in effective field theory and gluonic operators*, *Phys. Rev. D* **89** (2014) 033005 [`arXiv:1308.6561`] [InSPIRE].

[5] T. Husek, K. Monsalvez-Pozo and J. Portoles, *Lepton-flavour violation in hadronic tau decays and $\mu - \tau$ conversion in nuclei*, *JHEP* **01** (2021) 059 [`arXiv:2009.10428`] [InSPIRE].

[6] T. Husek, K. Monsalvez-Pozo and J. Portoles, *Constraints on leptoquarks from lepton-flavour-violating tau-lepton processes*, *JHEP* **04** (2022) 165 [`arXiv:2111.06872`] [InSPIRE].

[7] M. Carpentier and S. Davidson, *Constraints on two-lepton, two quark operators*, *Eur. Phys. J. C* **70** (2010) 1071 [`arXiv:1008.0280`] [InSPIRE].

[8] HEAVY FLAVOR AVERAGING GROUP (HFLAV) collaboration, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023*, `arXiv:2411.18639` [InSPIRE].

[9] BELLE collaboration, *Search for Lepton Flavor Violating τ^- Decays into $\ell^- K_s^0$ and $\ell^- K_s^0 K_s^0$*, *Phys. Lett. B* **692** (2010) 4 [`arXiv:1003.1183`] [InSPIRE].

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

[11] SUPERKEKB collaboration, *SuperKEKB Collider*, *Nucl. Instrum. Meth. A* **907** (2018) 188 [`arXiv:1809.01958`] [InSPIRE].

[12] BELLE-II collaboration, *Belle II Technical Design Report*, `arXiv:1011.0352` [InSPIRE].

[13] BELLE-II SVD collaboration, *The design, construction, operation and performance of the Belle II silicon vertex detector*, 2022 *JINST* **17** P11042 [`arXiv:2201.09824`] [InSPIRE].

[14] D. Kotchetkov et al., *Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector*, *Nucl. Instrum. Meth. A* **941** (2019) 162342 [`arXiv:1804.10782`] [InSPIRE].

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

[16] S. Banerjee, B. Pietrzyk, J.M. Roney and Z. Was, *Tau and muon pair production cross-sections in electron-positron annihilations at $\sqrt{s}$ = 10.58-GeV*, *Phys. Rev. D* **77** (2008) 054012 [arXiv:0706.3235] [INSPIRE].

[17] S. Jadach, B.F.L. Ward and Z. Was, *The Precision Monte Carlo event generator K K for two fermion final states in e^+e^- collisions*, *Comput. Phys. Commun.* **130** (2000) 260 [hep-ph/9912214] [INSPIRE].

[18] S. Jadach, J.H. Kuhn and Z. Was, *TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons*, *Comput. Phys. Commun.* **64** (1990) 275 [INSPIRE].

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

[20] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[21] D.J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152 [INSPIRE].

[22] S. Jadach, E. Richter-Was, B.F.L. Ward and Z. Was, *Monte Carlo program BHLUMI-2.01 for Bhabha scattering at low angles with Yennie-Frautschi-Suura exponentiation*, *Comput. Phys. Commun.* **70** (1992) 305 [INSPIRE].

[23] G. Balossini et al., *Matching perturbative and parton shower corrections to Bhabha process at flavour factories*, *Nucl. Phys. B* **758** (2006) 227 [hep-ph/0607181] [INSPIRE].

[24] G. Balossini et al., *Photon pair production at flavour factories with per mille accuracy*, *Phys. Lett. B* **663** (2008) 209 [arXiv:0801.3360] [INSPIRE].

[25] C.M. Carloni Calame, G. Montagna, O. Nicrosini and F. Piccinini, *The BABAYAGA event generator*, *Nucl. Phys. B Proc. Suppl.* **131** (2004) 48 [hep-ph/0312014] [INSPIRE].

[26] C.M. Carloni Calame, *An improved parton shower algorithm in QED*, *Phys. Lett. B* **520** (2001) 16 [hep-ph/0103117] [INSPIRE].

[27] C.M. Carloni Calame et al., *Large angle Bhabha scattering and luminosity at flavor factories*, *Nucl. Phys. B* **584** (2000) 459 [hep-ph/0003268] [INSPIRE].

[28] F.A. Berends, P.H. Daverveldt and R. Kleiss, *Radiative Corrections to the Process $e^+e^- \to e^+e^-\mu^+\mu^-$*, *Nucl. Phys. B* **253** (1985) 421 [INSPIRE].

[29] F.A. Berends, P.H. Daverveldt and R. Kleiss, *Complete Lowest Order Calculations for Four Lepton Final States in electron-Positron Collisions*, *Nucl. Phys. B* **253** (1985) 441 [INSPIRE].

[30] F.A. Berends, P.H. Daverveldt and R. Kleiss, *Monte Carlo Simulation of Two Photon Processes. 2. Complete Lowest Order Calculations for Four Lepton Production Processes in electron Positron Collisions*, *Comput. Phys. Commun.* **40** (1986) 285 [INSPIRE].

[31] S. Uehara, *TREPS: A Monte-Carlo Event Generator for Two-photon Processes at e^+e^- Colliders using an Equivalent Photon Approximation*, arXiv:1310.0157 [INSPIRE].

[32] BELLE-II FRAMEWORK SOFTWARE GROUP collaboration, *The Belle II Core Software*, *Comput. Softw. Big Sci.* **3** (2019) 1 [arXiv:1809.04299] [INSPIRE].

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

[34] GEANT4 collaboration, *GEANT4 - A Simulation Toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 [INSPIRE].

[35] R. Brun et al., *GEANT3*, CERN-DD-EE-84-1 (1987),
https://cds.cern.ch/record/1119728/files/CERN-DD-EE-84-1.pdf [INSPIRE].

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

[37] S. Brandt, C. Peyrou, R. Sosnowski and A. Wroblewski, *The principal axis of jets. An attempt to
analyze high-energy collisions as two-body processes*, *Phys. Lett.* **12** (1964) 57 [INSPIRE].

[38] E. Farhi, *A QCD Test for Jets*, *Phys. Rev. Lett.* **39** (1977) 1587 [INSPIRE].

[39] M. Milesi, J. Tan and P. Urquijo, *Lepton identification in Belle II using observables from the
electromagnetic calorimeter and precision trackers*, *EPJ Web Conf.* **245** (2020) 06023 [INSPIRE].

[40] A. Abashian et al., *Muon identification in the Belle experiment at KEKB*, *Nucl. Instrum. Meth.
A* **491** (2002) 69 [INSPIRE].

[41] K. Hanagaki et al., *Electron identification in Belle*, *Nucl. Instrum. Meth. A* **485** (2002) 490
[hep-ex/0108044] [INSPIRE].

[42] BELLE-II ANALYSIS SOFTWARE GROUP collaboration, *Global decay chain vertex fitting at Belle
II*, *Nucl. Instrum. Meth. A* **976** (2020) 164269 [arXiv:1901.11198] [INSPIRE].

[43] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01
[INSPIRE].

[44] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and
Upsilon resonances*, Ph.D. thesis, Cracow, INP, Poland (1986) [INSPIRE].

[45] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, arXiv:1603.02754
[DOI:10.1145/2939672.2939785] [INSPIRE].

[46] T. Akiba et al., *Optuna: A Next-generation Hyperparameter Optimization Framework*,
arXiv:1907.10902 [INSPIRE].

[47] G. Punzi, *Sensitivity of searches for new signals and its optimization*, *eConf C* **030908** (2003)
MODT002 [physics/0308063] [INSPIRE].

[48] BELLE collaboration, *Search for lepton-flavor-violating τ decays into a lepton and a vector meson
using the full Belle data sample*, *JHEP* **06** (2023) 118 [arXiv:2301.03768] [INSPIRE].

[49] BELLE-II collaboration, *Measurement of the integrated luminosity of data samples collected
during 2019-2022 by the Belle II experiment*, *Chin. Phys. C* **49** (2025) 013001
[arXiv:2407.00965] [INSPIRE].

[50] BELLE collaboration, *Physics Achievements from the Belle Experiment*, *PTEP* **2012** (2012)
04D001 [arXiv:1212.5342] [INSPIRE].

[51] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl.
Instrum. Meth. A* **434** (1999) 435 [hep-ex/9902006] [INSPIRE].

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

[53] L. Heinrich, M. Feickert and G. Stark, *pyhf: v0.7.6*, DOI:10.5281/zenodo.1169739.

[54] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: pure-Python implementation of
HistFactory statistical models*, *J. Open Source Softw.* **6** (2021) 2823 [INSPIRE].

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 ⓘ, V. Aushev ⓘ, M. Aversano ⓘ, R. Ayad ⓘ, V. Babu ⓘ,
N. K. Baghel ⓘ, S. Bahinipati ⓘ, P. Bambade ⓘ, Sw. Banerjee ⓘ, 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 ⓘ, D. Biswas ⓘ, A. Bobrov ⓘ, D. Bodrov ⓘ, A. Bondar ⓘ, G. Bonvicini ⓘ,
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 ⓘ, R. Cheaib ⓘ, P. Cheema ⓘ, 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 Marino ⓘ, 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 ⓘ, M. Dorigo ⓘ, K. Dugic ⓘ, G. Dujany ⓘ, P. Ecker ⓘ, D. Epifanov ⓘ, J. Eppelt ⓘ,
P. Feichtinger ⓘ, T. Ferber ⓘ, T. Fillinger ⓘ, C. Finck ⓘ, G. Finocchiaro ⓘ, A. Fodor ⓘ, F. Forti ⓘ,
B. G. Fulsom ⓘ, A. Gabrielli ⓘ, A. Gale ⓘ, M. Garcia-Hernandez ⓘ, 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 ⓘ, A. Glazov ⓘ, B. Gobbo ⓘ,
R. Godang ⓘ, P. Goldenzweig ⓘ, W. Gradl ⓘ, 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 ⓘ, A. Heidelbach ⓘ, 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 ⓘ, G. Inguglia ⓘ, N. Ipsita ⓘ, A. Ishikawa ⓘ,
R. Itoh ⓘ, M. Iwasaki ⓘ, P. Jackson ⓘ, D. Jacobi ⓘ, W. W. Jacobs ⓘ, D. E. Jaffe ⓘ, E.-J. Jang ⓘ,
S. Jia ⓘ, Y. Jin ⓘ, A. Johnson ⓘ, K. K. Joo ⓘ, H. Junkerkalefeld ⓘ, A. B. Kaliyar ⓘ, J. Kandra ⓘ,
K. H. Kang ⓘ, G. Karyan ⓘ, T. Kawasaki ⓘ, F. Keil ⓘ, C. Ketter ⓘ, C. Kiesling ⓘ, C.-H. Kim ⓘ,
D. Y. Kim ⓘ, J.-Y. Kim ⓘ, K.-H. Kim ⓘ, Y. J. Kim ⓘ, H. Kindo ⓘ, K. Kinoshita ⓘ, P. Kodyš ⓘ,
T. Koga ⓘ, S. Kohani ⓘ, K. Kojima ⓘ, A. Korobov ⓘ, S. Korpar ⓘ, E. Kovalenko ⓘ,
R. Kowalewski ⓘ, P. Križan ⓘ, P. Krokovny ⓘ, T. Kuhr ⓘ, Y. Kulii ⓘ, D. Kumar ⓘ, J. Kumar ⓘ,
R. Kumar ⓘ, K. Kumara ⓘ, T. Kunigo ⓘ, A. Kuzmin ⓘ, Y.-J. Kwon ⓘ, S. Lacaprara ⓘ,
K. Lalwani ⓘ, T. Lam ⓘ, L. Lanceri ⓘ, J. S. Lange ⓘ, T. S. Lau ⓘ, M. Laurenza ⓘ, K. Lautenbach ⓘ,
R. Leboucher ⓘ, F. R. Le Diberder ⓘ, M. J. Lee ⓘ, C. Lemettais ⓘ, P. Leo ⓘ, 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 ⓘ,
V. Lisovskyi ⓘ, 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 ⓘ, A. Martini ⓘ, 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 ⓘ, I. Nakamura ⓘ, M. Nakao ⓘ, Y. Nakazawa ⓘ,

M. Naruki , Z. Natkaniec , A. Natochii , M. Nayak , M. Neu , S. Nishida , S. Ogawa ,
R. Okubo , H. Ono , Y. Onuki , G. Pakhlova , S. Pardi , K. Parham , 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 , V. Raj , 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 ,
D. A. Sanders , S. Sandilya , L. Santelj , V. Savinov , B. Scavino , J. Schmitz ,
S. Schneider , G. Schnell , 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 ,
R. J. Sobie , M. Sobotzik , A. Soffer , A. Sokolov , E. Solovieva , W. Song , S. Spataro ,
B. Spruck , M. Starič , P. Stavroulakis , S. Stefkova , R. Stroili , Y. Sue , M. Sumihama ,
K. Sumisawa , N. Suwonjandee , H. Svidras , M. Takahashi , M. Takizawa , U. Tamponi ,
K. Tanida , F. Tenchini , A. Thaller , O. Tittel , R. Tiwary , E. Torassa , K. Trabelsi ,
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 , M. Wakai , S. Wallner ,
M.-Z. Wang , A. Warburton , M. Watanabe , S. Watanuki , C. Wessel , E. Won ,
X. P. Xu , B. D. Yabsley , S. Yamada , W. Yan , S. B. Yang , J. Yelton , 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

https://belle.kek.jp/
https://www.belle2.org/