Search for Lepton-Flavor-Violating Decay Modes $B^0 \to K^0_S \tau^{\pm} \ell^{\mp}$ with Hadronic B Tagging at Belle and Belle II

I. Adachi, K. Adamczyk, L. Aggarwal, H. Ahmed, H. Aihara, N. Akopov, M. Alhakami, A. Aloisio, N. Althubiti, N. Anh Ky, D. M. Asner, H. Atmacan, 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, V. Bhardwaj, B. Bhuyan, F. Bianchi, D. Biswas, A. Bobrov, D. Bodrov, A. Bolz, A. Boschetti, A. Bozek, M. Bračko, P. Branchini, R. A. Briere, T. E. Browder, A. Budano, S. Bussino, Q. Campagna, M. Campajola, L. Cao, G. Casarosa, C. Cecchi, J. Cerasoli, M.-C. Chang, P. Chang, R. Cheaib, P. Cheema, B. G. Cheon, K. Chilikin, K. Chirapatpimol, H.-E. Cho, K. Cho, S.-J. Cho, S.-K. Choi, S. Choudhury, J. Cochran, 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, R. Dhamija, A. Di Canto, F. Di Capua, J. Dingfelder, Z. Doležal, I. Domínguez Jiménez, T. V. Dong, X. Dong, M. Dorigo, D. Dossett, S. Dubey, K. Dugic, G. Dujany, P. Ecker, P. Feichtinger, T. Ferber, T. Fillinger, C. Finck, G. Finocchiaro, A. Fodor, F. Forti, B. G. Fulsom, A. Gabrielli, E. Ganiev, M. Garcia-Hernandez, R. Garg, G. Gaudino, V. Gaur, 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, O. Gogota, P. Goldenzweig, W. Gradl, S. Granderath, E. Graziani, D. Greenwald, Z. Gruberová, Y. Guan, K. Gudkova, I. Haide, Y. Han, T. Hara, C. Harris, 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, D. Jacobi, W. W. Jacobs, D. E. Jaffe, E.-J. Jang, Q. P. Ji, S. Jia, Y. Jin, A. Johnson, K. K. Joo, H. Junkerkalefeld, M. Kaleta, A. B. Kaliyar, J. Kandra, K. H. Kang, S. Kang, G. Karyan, T. Kawasaki, F. Keil, C. Ketter, C. Kiesling, C.-H. Kim, D. Y. Kim, J.-Y. Kim, K.-H. Kim, Y.-K. Kim, 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, R. Kumar, K. Kumara, T. Kunigo, A. Kuzmin, Y.-J. Kwon, S. Lacaprara, Y.-T. Lai, K. Lalwani, T. Lam, L. Lanceri, 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, 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, T. Luo, C. Lyu, Y. Ma, C. Madaan, M. Maggiora, S. P. Maharana, R. Maiti, G. Mancinelli, R. Manfredi, E. Manoni, M. Mantovano, 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, G. B. Mohanty, S. Mondal, S. Moneta, H.-G. Moser, R. Mussa, I. Nakamura, K. R. Nakamura, M. Nakao, Y. Nakazawa, M. Naruki, Z. Natkaniec, A. Natochii, M. Nayak, G. Nazaryan, M. Neu, S. Nishida, S. Ogawa, H. Ono, Y. Onuki, F. Otani, G. Pakhlova, E. Paoloni, S. Pardi, K. Parham, H. Park, J. Park, K. Park, S.-H. Park, B. Paschen, A. Passeri, S. Patra, T. K. Pedlar, I. Peruzzi, R. Peschke, R. Pestotnik, M. Piccolo, L. E. Piilonen, P. L. M. Podesta-Lerma, T. Podobnik, S. Pokharel, C. Praz, S. Prell, E. Prencipe, M. T. Prim, I. Prudiiev, H. Purwar, P. Rados, G. Raeuber, S. Raiz, N. Rauls, K. Ravindran, J. U. Rehman, M. Reif, S. Reiter, M. Remnev, L. Reuter, D. Ricalde Herrmann, I. Ripp-Baudot, G. Rizzo, S. H. Robertson, M. Roehrken, J. M. Roney, A. Rostomyan, N. Rout, D. A. Sanders, S. Sandilya, L. Santelj, V. Savinov, B. Scavino, S. Schneider, G. Schnell, C. Schwanda, Y. Seino, A. Selce, K. Senyo, J. Serrano, M. E. Sevior, C. Sfienti, W. Shan, C. Sharma, 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, S. Spataro, B. Spruck, W. Song, M. Starič, P. Stavroulakis, S. Stefkova, R. Stroili, J. Strube, Y. Sue, M. Sumihama, K. Sumisawa, W. Sutcliffe, 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, M. Wakai, S. Wallner, M.-Z. Wang, Z. Wang, A. Warburton, M. Watanabe, S. Watanuki, C. Wessel, J. Wiechczynski, E. Won, X. P. Xu, B. D. Yabsley, S. Yamada, W. Yan, J. Yelton, J. H. Yin, K. Yoshihara, C. Z. Yuan, J. Yuan, L. Zani, F. Zeng, B. Zhang, J. S. Zhou, Q. D. Zhou, L. Zhu, V. I. Zhukova, and R. Žlebčík

(Belle and Belle II Collaborations)

(Received 4 January 2025; revised 13 April 2025; accepted 30 May 2025; published 25 July 2025; corrected 17 July 2026)

We present the first search for the lepton-flavor-violating decay modes $B^0 \to K_S^0 \tau^\pm \ell^\mp$ ($\ell = \mu, e$) using the 711 and 365 fb^{-1} data samples recorded by the Belle and Belle II detectors, respectively. We use a hadronic B-tagging technique to fully reconstruct a B meson and search for signal decays in the system recoiling against the tagged meson, considering τ decays to either light leptons, one charged hadron, or one charged hadron and a neutral pion. We find no evidence for $B^0 \to K_S^0 \tau^\pm \ell^\mp$ decays and set 90% confidence level upper limits on the branching fractions in the range of $[0.8, 3.6] \times 10^{-5}$.

DOI: 10.1103/yzvh-v3hm

Recent anomalies observed in semileptonic B decays, particularly in transitions like $b \to c\tau\nu$ [1], may indicate deviations from lepton flavor universality. According to this principle, the three generations of leptons are expected to interact identically with gauge bosons, except for mass differences. Any experimental deviation from the standard model (SM) predictions would suggest the existence of new heavy particles, such as leptoquarks or Z' bosons [2–4], with preferential couplings to third-generation leptons. Recently, Belle II reported a $b \to s\nu\bar{\nu}$ excess and obtained the branching fraction $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (2.3 \pm 0.7) \times 10^{-5}$ [5], which is 2.7 standard deviations (σ) larger than the SM expectation. If confirmed, this would not only be physics beyond the standard model (BSM), but could also reflect off-diagonal couplings between leptons of different flavors [2,6]. These BSM couplings can enhance the rate of lepton-flavor-violating (LFV) B decays to observable values. Reference [6], starting from the $B^+ \to K^+\nu\bar{\nu}$ excess, predicts an enhancement of $\mathcal{B}(B \to K\tau^\pm\mu^\mp)$ to $[2, 3] \times 10^{-6}$, which is close to the current experimental sensitivity. The BSM models [2–4] predict significant enhancements in the branching fractions of processes such as $b \to s\tau\ell$, with the effect being enhanced due to the coupling of the third-generation b quark with the heaviest lepton, τ. This results in a notable increase in the branching fractions of $B \to K\tau^\pm\ell^\mp(\ell = \mu, e)$ decays and provides a potential experimental window into BSM physics.

BABAR performed the first LFV searches in $B^+ \to K^+\tau^\pm\ell^\mp$ modes and set upper limits (ULs) on their branching fractions in the range of $[1.5, 4.5] \times 10^{-5}$ at 90% confidence level (CL) [7]. Belle provided the most stringent UL on the $B^+ \to K^+\tau^+\mu^-$ decay of 6×10^{-6} at 90% CL using 772×10^6 $B\bar{B}$ pairs [8], significantly better than LHCb's limit of 3.9×10^{-5}, obtained with 9 fb^{-1} of pp collision data [9]. Moreover, LHCb set ULs on $B^0 \to K^{*0}\tau^\pm\mu^\mp$ decays in the range of $[0.8, 1.0] \times 10^{-5}$ at 90% CL [10].

In this Letter, we present the first search for the LFV decays $B^0 \to K_S^0 \tau^\pm \ell^\mp (\ell = \mu, e)$. We employ a hadronic B-tagging technique, and then use the recoil mass to reconstruct the mass of the τ. Our results are based on a combined analysis of 772×10^6 $B\bar{B}$ pairs (711 fb^{-1}) from Belle and 387×10^6 $B\bar{B}$ pairs (365 fb^{-1}) from Belle II (2019–2022). The advantage of using K_S^0 over K^+ and K^{*0} in $B \to K\tau\ell$ decays is that its clean $K_S^0 \to \pi^+\pi^-$ signature eliminates backgrounds from $K - \pi$ misidentification, demonstrating the unique strengths of Belle and Belle II's hermetic detectors.

Belle operated at the KEKB asymmetric-energy collider with electron (positron) beam energies of 8.0(3.5) GeV [11]. Belle II operates at the successor, SuperKEKB, designed to deliver 40 times higher instantaneous luminosity than KEKB, with electron (positron) beam energies of 7(4) GeV [12]. The Belle II detector [13] is an upgraded version of the Belle detector [14], including a vertex detector (VXD), composed of two inner layers of pixel detectors and four outer layers of double-sided strip detectors, a central drift chamber (CDC), a time-of-propagation detector in the central detector volume and an aerogel ring-imaging Cherenkov detector in the forward region, and an electromagnetic calorimeter (ECL). All these subdetectors are located inside the same solenoid as in Belle, with a K_L^0-Muon detector (KLM) instrumented in the yoke.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP^3.

The analysis procedure is first developed using simulation and validated with real-data control samples ($B^0 \to D^- \pi$, $B^0 \to D_s^+ D^-$) before being applied to collision data. EvtGen [15] is used to model B decays with final-state radiation simulated by PHOTOS [16]. The $B^0 \to K_S^0 \tau^\pm \ell^\mp$ signal channels are modeled using an uniform three-body phase-space model. The KKMC [17] and PYTHIA [18] packages are used to simulate the $e^+e^- \to q\bar{q}$ continuum ($q = u, d, s, c$). The detector responses are modeled by GEANT3 [19] for Belle and GEANT4 [20] for Belle II. We use the Belle II analysis software framework (basf2) [21] to reconstruct the events for both Belle and Belle II data. The Belle data are converted to the Belle II format for basf2 compatibility using the B2BII framework [22].

In each $B\bar{B}$ pair, if one B meson, $B_{\rm sig}$, decays to a final state involving neutrinos, such as that searched for in the present work, it cannot be fully reconstructed as neutrinos escape the detector. However, the presence of missing energy and momentum can be inferred from the other $\bar{B}$ meson. This process is called tagging, with the other $\bar{B}$ meson referred to as $B_{\rm tag}$. We require a fully reconstructed $B_{\rm tag}$ using the hadronic full-event-interpretation (FEI) algorithm [23], a machine-learning based algorithm developed for B-tagging analyses at Belle and Belle II. Each reconstructed $B_{\rm tag}$ candidate is assigned a multivariate classifier output, $\mathcal{P}_{\rm FEI}$, ranging from zero (background-like) to one (signal-like). To constrain the $B_{\rm tag}$ kinematics, we require the beam-energy-constrained mass $M_{\rm bc} = \sqrt{(E_{\rm beam}/c^2)^2 - (\vec{p}_{B_{\rm tag}}/c)^2} > 5.27\ {\rm GeV}/c^2$, and the energy difference $\Delta E = E_{B_{\rm tag}} - E_{\rm beam}$ to satisfy $-0.15 < \Delta E < 0.1$ GeV. Here, $E_{\rm beam}$, $E_{B_{\rm tag}}$, and $\vec{p}_{B_{\rm tag}}$ denote the beam energy, and the energy and momentum of the $B_{\rm tag}$ candidate in the e^+e^- center-of-mass frame, respectively. If multiple $B_{\rm tag}$ candidates are reconstructed in a single event, the one with the highest $\mathcal{P}_{\rm FEI}$ is chosen, and candidates satisfying $\mathcal{P}_{\rm FEI} > 0.001$ are retained. Using these criteria, the B tagging has an average efficiency of 0.59% and a purity of 44%, corresponding to 6.8×10^6 tagged B^0 mesons. Here, purity is the ratio of reconstructed or expected signal events to total events in the signal region. Combining the visible particles from $B_{\rm sig}$ with the fully reconstructed $B_{\rm tag}$, we can kinematically constrain the undetected neutrinos as possible products of a τ decay.

Tracks and clusters not associated to $B_{\rm tag}$ are used to reconstruct the $B_{\rm sig}$, whose flavor is assumed to be opposite to that of the $B_{\rm tag}$ candidate. The BSM couplings between $b\tau$ and $b\ell$, or between $s\tau$ and $s\ell$, can be different, leading to an asymmetric differential decay rate between $b \to s\tau^+\ell^-$ and $b \to s\tau^-\ell^+$ [see Eq. (9) in Ref. [2]]. To address this and the different background characteristics, we distinguish signal channels by the (fractional) charge of the b quark and the primary lepton (ℓ): same-sign SS_ℓ ($B^0 \to K_S^0 \tau^- \ell^+$) and opposite-sign OS_ℓ ($B^0 \to K_S^0 \tau^+ \ell^-$). Because of B^0-$\bar{B}^0$ mixing, a small fraction (χ_d) [24] of SS_ℓ decays are classified as OS_ℓ decays, and vice versa. The τ candidates are reconstructed via $\tau \to e\nu\bar{\nu}, \mu\nu\bar{\nu}, \pi\nu$, or $\rho(\to\pi\pi^0)\nu$, covering over 70% of τ decays [24]. Signal channels are formed by combining a K_S^0, a ℓ, and a τ prong ($t_\tau = e/\mu/\pi/\rho$).

We reconstruct K_S^0 candidates from a pair of oppositely charged tracks assumed to be pions, with a common vertex. We use a standard momentum-binned K_S^0 selection that includes requirements on the K_S^0 flight information [25]. The K_S^0 reconstruction achieves over 98% purity, and most background events also contain true K_S^0 candidates.

To select ℓ and t_τ charged track ($e/\mu/\pi/\rho \to \pi$), we require the transverse (d_0) and longitudinal (z_0) projection of the distance of closest approach to the origin to be less than 0.5 cm and 5.0 cm to reduce misreconstructed or spurious tracks from beam-induced background. At least 20 hits in the CDC are required. For Belle, we use information from the KLM only to identify muon candidates, while for Belle II, we use information from all subdetectors except the VXD. Muon candidates are required to have momenta greater than $0.6\ {\rm GeV}/c$ to sufficiently penetrate the KLM. This selection has an efficiency of 89% with a pion misidentification rate lower than 2.5% for both samples [26]. Electrons are required to have momenta greater than $0.3\ {\rm GeV}/c$ to lie in the acceptance of the ECL. For Belle, electrons are identified using the information from the ECL, CDC, and aerogel threshold Cerenkov counter, respectively. For Belle II, the electron identification uses a boosted-decision-tree (BDT) classifier trained with information from all subdetectors except the VXD. The electron identification has an efficiency of 92% (86%) and a pion misidentification rate below 0.3% (0.5%) for Belle (Belle II) [27]. To recover electron candidates with bremsstrahlung, we accept photons with a minimum energy of 50 MeV that are within a 50 mrad angle of an electron track. Pion candidates for t_τ reconstruction are selected using particle identification (PID) likelihoods using information from the aerogel threshold Cerenkov counter, CDC, and time-of-flight scintillation counters for Belle. For Belle II, information from all the subdetectors except the VXD is used. This achieves a pion identification efficiency of 85% (83%) and a kaon misidentification rate of 6% (8%) for Belle (Belle II).

The $\tau \to \rho(\to\pi\pi^0)\nu$ mode, the most probable τ decay channel ($\mathcal{B} > 20\%$), has never been used in $B \to K\tau\ell$ analyses [7–9]. Understanding this mode is crucial for improving the kinematic properties of t_τ and enhancing background rejection strategies. However, reconstruction is challenging due to contamination from misidentified π^0 candidates formed from clusters associated with hadronic interactions or photons from beam-related background. To reconstruct the clean $\tau \to \rho\nu$ mode, photons are selected with energies above 50(60), 100(75), and 150(100) MeV in

the barrel, forward, and backward end caps for Belle (Belle II). We developed classifiers to suppress these backgrounds using BDTs. The common cluster features are energy, polar angle (relative to the beam pipe), lateral energy distribution [28], distance between the cluster and its nearest track, and fraction of cluster energy detected in the central part. For Belle II classifiers, we include additional features: the time difference between the collision and reconstructed cluster, outputs of classifiers using eleven Zernike moments [29], and identifiers for electromagnetic or hadronic showers using pulse shape discrimination [30]. For Belle, additional features include azimuthal angle, number of crystals in the cluster, and energy in the most energetic crystal. On average, the classifiers reduce backgrounds by 90% with a π^0 efficiency of 70%, including the selection $0.125 < M_{\pi^0} < 0.145\ \mathrm{GeV}/c^2$. The ρ candidates are selected with $0.60 < M_{\pi\pi^0} < 0.94\ \mathrm{GeV}/c^2$. If multiple candidates exist (occurring in 7% of events), one is chosen randomly.

The rest-of-event (ROE) consists of the tracks and clusters not used in B_{tag} and B_{sig}. We select events without any track in the ROE having $|d_0| < 10$ cm and $|z_0| < 20$ cm. The ROE clusters are required to satisfy the same selection criteria as the photons used for π^0 reconstruction. If the t_τ candidate is identified with multiple particle hypotheses, we assign a single one according to the following priority order (based on the purity of the modes): muon, electron, and pion. We require the $\tau \to \pi\nu$ mode to have no additional π^0 candidate in the ROE to avoid double counting with the $\tau \to \rho\nu$ mode.

The B_{sig} momentum is equal in magnitude and opposite in direction to that of B_{tag}, $\vec{p}_{B_{\mathrm{tag}}}$, and the B_{sig} energy is equal to E_{beam} in the center-of-mass frame. Therefore, the τ momentum and energy are given by

$$\vec{p}_\tau = -\vec{p}_{B_{\mathrm{tag}}} - \vec{p}_{K_S^0} - \vec{p}_\ell,$$
$$E_\tau = E_{\mathrm{beam}} - E_{K_S^0} - E_\ell,$$

from which we reconstruct the recoil M_τ according to Eq. (1). The signal yields are then extracted from M_τ, as signal events peak at the known τ mass [24], while the background remains flat without any peaking structures in the simulation. When t_τ is μ or e (i.e., when there are two leptons in B_{sig}), it is possible to form both SS_ℓ and OS_ℓ candidates, but this does not bias the signal yield as misassigned candidates do not peak in M_τ,

$$M_{\mathrm{recoil}} = M_\tau = \Big[m_B^2 + M_{K_S^0\ell}^2 - 2\Big(E_{\mathrm{beam}}E_{K_S^0\ell} + |\vec{p}_{B_{\mathrm{tag}}}||\vec{p}_{K_S^0\ell}|\cos\theta\Big)\Big]^{\frac{1}{2}}. \quad (1)$$

Here, m_B is the known B^0 mass [24]; $M_{K_S^0\ell}$, $E_{K_S^0\ell}$, $\vec{p}_{K_S^0\ell}$ are the mass, energy, and momentum of the system composed of the K_S^0 and ℓ, respectively; θ is the angle between $\vec{p}_{B_{\mathrm{tag}}}$ and $\vec{p}_{K_S^0\ell}$. The M_τ resolution is approximately $25\ \mathrm{MeV}/c^2$ for both Belle and Belle II simulations.

The background after the preselection is mainly decays arising from $b \to c$ transitions. Events with $B^0 \to D^{(*)-}(\to K_S^0 t^- X)\ell^+\nu$ decays, where the reconstructed primary lepton originates from semileptonic B^0 decay and a track t^- from $D^{(*)-}$ is misinterpreted as coming from the τ decay, with X representing any other particles, constitute the dominant source of background in SS_ℓ modes. If $B^0 \to D^{(*)-}\ell^+\nu$ undergoes a flavor transformation due to B^0-$\bar{B}^0$ mixing, this results in a change in the signs of the final states ($\bar{B}^0 \to D^{(*)+}\ell^-\bar{\nu}$) and provides the appropriate sign configuration for the final state in OS_ℓ modes. Additionally, the reconstruction of the primary lepton from B_{sig} tends to favor a higher momentum lepton originating from the B^0 meson. Consequently, semileptonic B decays are the primary background in both SS_ℓ and OS_ℓ modes. As the K_S^0 and t_τ come from a D meson, we require the invariant mass of K_S^0 and t_τ to be greater than $1.91\ \mathrm{GeV}/c^2$, i.e., greater than the D meson mass accounting for resolution. For $\tau \to \rho\nu$ decay, we require $M_{K_S^0\rho}$ to be greater than $2.1\ \mathrm{GeV}/c^2$ due to the poor mass resolution resulting from the presence of a π^0 candidate. Because $B^0 \to K_S^0 J/\psi(\to \ell^+\ell^-)$ background can pass our selection criteria, events in the range $3.00 < M_{t_\tau\ell} < 3.14\ \mathrm{GeV}/c^2$ are rejected when ℓ is an electron (muon) for the $\tau \to e(\mu)\nu\bar{\nu}$ mode. To suppress the photon conversion background, we require $M_{t_\tau\ell}$, in this case $M_{e^+e^-}$, to be greater than $0.15\ \mathrm{GeV}/c^2$.

Continuum $q\bar{q}$ events can be distinguished from $B\bar{B}$ events by exploiting their difference in event topologies. We use sphericity-related variables [31] and require the cosine of the angle between the thrust axis of B_{tag} and the other particles not used in B_{tag} ($\cos\theta_{\mathrm{T}}$) to be less than 0.9. These selections reduce the $q\bar{q}$ background by 86% and retain 88% of the signal for both samples.

After the above selection criteria, the background consists of charm meson semileptonic decays with a K_S^0 and ℓ in their final state, other $B\bar{B}$ decays, and $q\bar{q}$ events. For each signal mode, BDT classifiers are trained with 11 features to suppress the residual backgrounds. These features include $M_{K_S^0\ell}$, which helps to suppress the background from charm meson semileptonic decays, the sum of ECL cluster energies in the ROE, energies of the ℓ and t_τ, event-shape variables, and modified Fox-Wolfram moments [32], which help to reduce the $q\bar{q}$ background. The same BDT output criterion are used for Belle and Belle II because of the similar performance, determined using figure of merit $\epsilon_{\mathrm{sig}}/[(a/2) + \sqrt{N_{\mathrm{bkg}}}]$ [33]. Here, ϵ_{sig} is the signal efficiency, N_{bkg} is the expected background yield, and $a = 3$ represents the target significance in terms of standard deviations. The BDT selection results in an average signal

efficiency of about 75% and rejects 90% of the remaining background.

The purity in the simulations is similar in Belle and Belle II for every channel after applying all selection criteria, so the two datasets are merged. To extract the signal yield, we do a single unbinned-maximum-likelihood fit to the M_τ distribution of the combined dataset. The probability density function (PDF) used to model the M_τ signal distribution is a Johnson function [34]. The parameters that describe the signal shape are fixed to the values obtained from the fit to the simulated samples. Background events have a smooth distribution in the M_τ fit region, modeled using a second-order polynomial. To validate the fitting procedure, we generate large ensembles of simulated experiments, in which the M_τ distributions are produced from the PDFs used for fitting. Comparisons of the simulated and measured signal yields indicate no obvious bias.

The $B^0 \to D^- \pi^+$ sample is used to calibrate $B_{\rm tag}$ efficiency. We reconstruct $B_{\rm tag}$ using the FEI and a high momentum π^+, then compute $M_{\rm recoil}$ to observe the D signal using Eq. (1). The yield ratio $\mathcal{R}_{\rm FEI} = N_{\rm data}/N_{\rm simulation} = 0.74 \pm 0.04 (0.81 \pm 0.04)$ is taken as the calibration factor for the $B_{\rm tag}$ efficiency in Belle (Belle II).

The $B^0 \to D_s^+ D^-$ sample is used to calibrate the signal PDF and BDT selections. The D^- mass is reconstructed similarly to the τ mass in the signal decays. We reconstruct neutral $B_{\rm tag}$ candidates using the FEI algorithm and D_s^+, sharing the same momentum range as $K_S^0 \ell$, through the decays $\phi(\to K^+K^-)\pi^+$ and $K_S^0 K^+$ and a charged track from D as t_τ. The distribution of the mass recoiling against the $B_{\rm tag}$ and D_s^+ system is shown in Fig. 1 and clear signals for D^- and D^{*-} are visible. The $B^0 \to D_s^+ D^{*-}$ component is also fitted but we do not use it due to its lower purity. The $B^0 \to D_s^+ D^-$ signal PDF is modeled by the Johnson function and the parameters are fixed in the fits, except the mean, while the background is described by an exponential function with floating yield and shape parameters. We introduce a scale factor f to account for data-simulation difference on the width from signal simulation. The f ratio, determined to be 1.04 ± 0.15, is used as a correction factor for the width. The value of $\mathcal{B}(B^0 \to D_s^+ D^-)$ is measured to be $(10.1 \pm 1.2) \times 10^{-3}$, consistent with the world-average value [24] within 2σ, and serves as a closure test of the entire analysis chain. To validate the BDT performance, we apply it to $B^0 \to D_s^+ D^-$ events. The efficiency is derived from $B^0 \to D_s^+ D^-$ yields before and after BDT selection using $M_{\rm recoil}$ fits. The data-simulation efficiency ratios ($\mathcal{R}_{\rm BDT}$) are 0.93 ± 0.17, 0.96 ± 0.16, 0.92 ± 0.16, and 0.96 ± 0.18 for the OS_μ, SS_μ, OS_e, and SS_e modes, respectively.

Figure 2 shows the M_τ fits to data for $B^0 \to K_S^0 \tau^\pm \ell^\mp$ decays. There is no significant signal in any of the fit

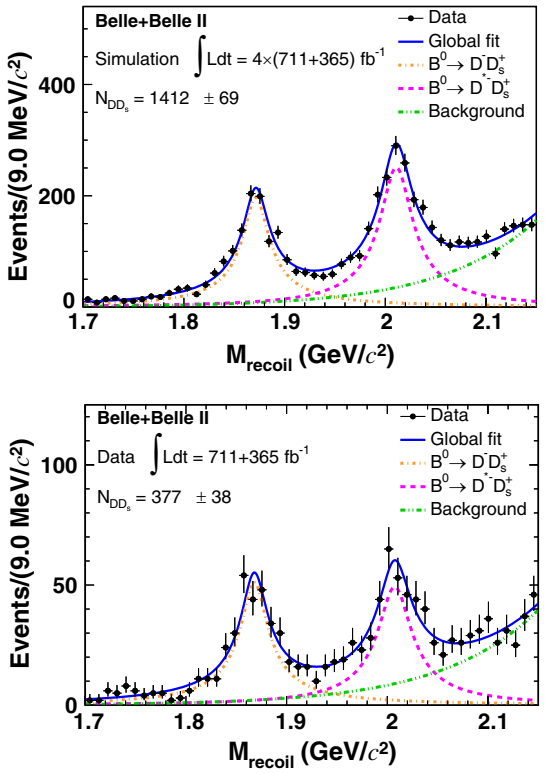


FIG. 1. Fit to the recoil mass of the $B_{\rm tag}$ and D_s^+ system for simulation (upper) and data (lower) with the combined Belle and Belle II samples.

channels. We translate the number of observed events $N_{\rm sig}$ into a branching fraction $\mathcal{B}$ using the expression

$$\mathcal{B} = \frac{N_{\rm sig}}{\epsilon \times 2N_{B\bar{B}} \times (1 + f_{+-}/f_{00})^{-1}}, \tag{2}$$

where ϵ is the efficiency after $\mathcal{R}_{\rm BDT}$ and $\mathcal{R}_{\rm FEI}$ calibrations. The efficiency also includes the branching fractions of K_S^0, τ, ρ, π^0, and the effect of B^0-$\bar{B}^0$ mixing (i.e., signal loss in mixed events) in the simulation. In the case where the true branching fractions are zero, the resulting estimates are unbiased. We use $N_{B\bar{B}} = 1159 \times 10^6$, which is the total number of $B\bar{B}$ pairs for the combined datasets, and $f_{+-}/f_{00} = 1.052 \pm 0.031$, which is the ratio of $\mathcal{B}[\Upsilon(4S) \to B^+B^-]$ to $\mathcal{B}[\Upsilon(4S) \to B^0\bar{B}^0]$ [1].

We obtain ULs on the signal yields using pseudoexperiments. They are generated using background and signal PDFs for different values of the signal branching fractions, performing 10 000 fits for each value. We then define $N_{\rm sig}^{\rm UL}$ at 90% CL as the signal yield for which 10% of the experiments have fit yields less than the observed $N_{\rm sig}$ in data. Systematic uncertainties are included by smearing the $N_{\rm sig}$ distribution obtained from the pseudoexperiments with the fractional systematic uncertainty, which has an effect of less than 1% on the mean $N_{\rm sig}$. The ULs on the branching fractions $\mathcal{B}^{\rm UL}$ are then obtained from $N_{\rm sig}^{\rm UL}$ using Eq. (2).

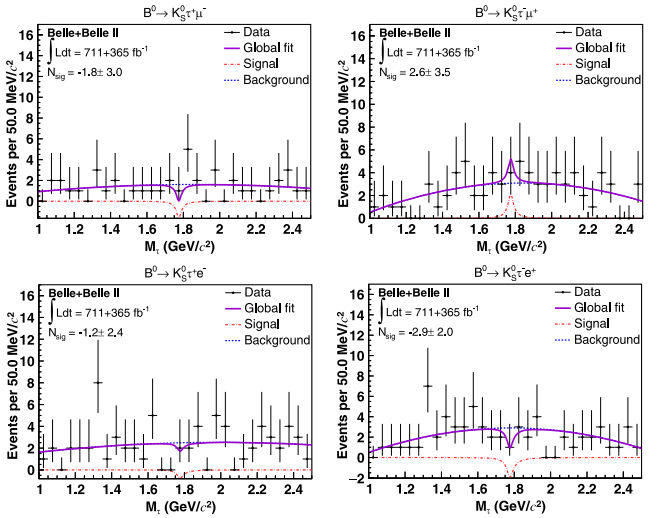


FIG. 2. The M_τ distributions and fits for the combined Belle and Belle II datasets. The black dots with error bars show the data, the red dash-dotted curve shows the signal component, the blue dashed curve shows the background component, and the purple solid curve shows the global fit.

Our UL estimation is conservative because we only reconstruct B^0-$\bar{B}^0$ events, not B^0-B^0 and $\bar{B}^0$-$\bar{B}^0$ events, and we include B^0-$\bar{B}^0$ mixing in the evaluation of the efficiency [Eq. (2)]. Table I summarizes the efficiency, fit results, and observed ULs at 90% CL for the four channels. Expected ULs, derived from the no-signal assumption, are in the range $[2.1, 2.2] \times 10^{-5}$.

The primary source of systematic uncertainty arises from the BDT selections, which is 16%–18%, based on the uncertainty in $\mathcal{R}_{\rm BDT}$ using the $B^0 \to D_s^+ D^-$ sample. Using the same sample, the uncertainty from the signal PDF is 15%. This includes the uncertainties in width (uncertainty of the width correction factor), mean (deviation from nominal D mass in the data fit), skewness, and Gaussian component strength of the Johnson function, estimated using a new PDF reweighted by mode-dependent calibration factors for the dominant B-tagging modes. The uncertainty in the $B_{\rm tag}$ efficiency is taken from the uncertainty of $\mathcal{R}_{\rm FEI}$ (4%). The small difference (0.8%–1.6%) in the validation of the fitting procedure is treated as the associated uncertainty. The uncertainty in K_S^0 reconstruction is estimated to be 1.1% using a $D^{*+} \to \pi^+ D^0$, $D^0 \to K_S^0 \pi^+ \pi^-$ sample. The Belle PID uncertainties are evaluated using $J/\psi \to \ell^+ \ell^-$ and $D^{*+} \to D^0(\to K^- \pi^+)\pi^+$ samples to be 0.3%, 0.4%, and 1.0% for muons, electrons, and pions, respectively. The Belle II PID uncertainties for muon, electron, and pion are 0.5%, 1.0%, and 1.0%, respectively, which are obtained using the samples described in Ref. [35]. The uncertainty from the π^0 reconstruction is 1.3% using $B^+ \to K^{*+}(\to K^+ \pi^0) J/\psi$ and $D^{*-} \to \bar{D}^0(\to K^+ \pi^- \pi^0)\pi^-$ samples. The uncertainty for the requirement that there is no additional π^0 candidate in the ROE in the $\tau \to \pi\nu$ mode is 1.0% using $B_{\rm tag} B(\to K_S^0 J/\psi)$ events. The uncertainties arising from $N_{B\bar{B}}$, f_{+-}/f_{00}, and the branching fractions of K_S^0, τ, ρ, and π^0 decays [24] are 1.1%, 1.5%, and 0.7%, respectively. For sources with different systematic uncertainties in Belle and Belle II, we calculate the total multiplicative values by weighting the individual uncertainties according to the integrated luminosities of the two samples. The total systematic uncertainties are 24%, 22%, 23%, and 24% for OS_μ, SS_μ, OS_e, and SS_e modes, respectively.

TABLE I. Efficiencies (ϵ), signal yields ($N_{\rm sig}$) of the data fit, central value of the branching fractions, and the observed $\mathcal{B}^{\rm UL}$ at 90% CL. The first uncertainty of the central value is statistical and the second is systematic.

			$\mathcal{B}(10^{-5})$	
Channels	$\epsilon(10^{-4})$	$N_{\rm sig}$	Central value	UL
$B^0 \to K_S^0 \tau^+ \mu^-$	1.7	-1.8 ± 3.0	$-1.0 \pm 1.6 \pm 0.2$	1.1
$B^0 \to K_S^0 \tau^- \mu^+$	2.1	2.6 ± 3.5	$1.1 \pm 1.6 \pm 0.3$	3.6
$B^0 \to K_S^0 \tau^+ e^-$	2.0	-1.2 ± 2.4	$-0.5 \pm 1.1 \pm 0.1$	1.5
$B^0 \to K_S^0 \tau^- e^+$	2.1	-2.9 ± 2.0	$-1.2 \pm 0.9 \pm 0.3$	0.8

In summary, we have searched for $B^0 \to K_S^0 \tau^\pm \ell^\mp$ for the first time using Belle and Belle II datasets. This is also the first direct search for LFV in B decays using the Belle II dataset. The ULs on the branching fractions at 90% CL are

$$\mathcal{B}(B^0 \to K_S^0 \tau^+ \mu^-) < 1.1 \times 10^{-5},$$
$$\mathcal{B}(B^0 \to K_S^0 \tau^- \mu^+) < 3.6 \times 10^{-5},$$
$$\mathcal{B}(B^0 \to K_S^0 \tau^+ e^-) < 1.5 \times 10^{-5},$$
$$\mathcal{B}(B^0 \to K_S^0 \tau^- e^+) < 0.8 \times 10^{-5}.$$

The results for $B^0 \to K_S^0 \tau^\pm e^\mp$ are the most stringent ULs on $b \to s\tau e$ transitions, and those for $B^0 \to K_S^0 \tau^\pm \mu^\mp$ are among the best limits on $b \to s\tau\mu$ transitions [7–10] achieved to date. These results are approaching the potential BSM enhancement level of $\mathcal{O}(10^{-6})$ implied by $B^+ \to K^+ \nu\bar{\nu}$ result [5]. Additionally, we provide the selection efficiency of the phase-space model as a function of $(M^2_{\tau\ell}, M^2_{K_S^0\ell})$ in Supplemental Material [36] to allow these results to be reinterpreted in specific BSM models.

Acknowledgments—This work, based on data collected using the Belle II detector, which was built and commissioned prior to March 2019, and data collected using the Belle detector, which was operated until June 2010, was supported by Higher Education and Science Committee of the Republic of Armenia Grant No. 23LCG-1C011; Australian Research Council and Research Grants No. DP200101792, No. DP210101900, No. DP210102831, No. DE220100462,

No. LE210100098, and No. LE230100085; Austrian Federal Ministry of Education, Science and Research, Austrian Science Fund No. P 34529, No. J 4731, No. J 4625, and No. M 3153, and Horizon 2020 ERC Starting Grant No. 947006 "InterLeptons"; Natural Sciences and Engineering Research Council of Canada, Compute Canada and CANARIE; 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, No. 12475093, and No. 12175041, and Shandong Provincial Natural Science Foundation Project ZR2022JQ02; the Czech Science Foundation Grant No. 22-18469S and Charles University Grant Agency Project No. 246122; European Research Council, Seventh Framework PIEF-GA-2013-622527, Horizon 2020 ERC-Advanced Grants No. 267104 and No. 884719, Horizon 2020 ERC-Consolidator Grant No. 819127, Horizon 2020 Marie Sklodowska-Curie Grant Agreement No. 700525 "NIOBE" and No. 101026516, and Horizon 2020 Marie Sklodowska-Curie RISE project JENNIFER2 Grant Agreement No. 822070 (European grants); L'Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) du CNRS and L'Agence Nationale de la Recherche (ANR) under grant ANR-21-CE31-0009 (France); BMBF, DFG, HGF, MPG, and AvH Foundation (Germany); Department of Atomic Energy under Project Identification No. RTI 4002, Department of Science and Technology, and UPES SEED funding programs No. UPES/R&D-SEED-INFRA/17052023/01 and No. UPES/R&D-SOE/20062022/06 (India); Israel Science Foundation Grant No. 2476/17, U.S.-Israel Binational Science Foundation Grant No. 2016113, and Israel Ministry of Science Grant No. 3-16543; Istituto Nazionale di Fisica Nucleare and the Research Grants BELLE2, and the ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union—NextGenerationEU; Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research Grants No. 16H03968, No. 16H03993, No. 16H06492, No. 16K05323, No. 17H01133, No. 17H05405, No. 18K03621, No. 18H03710, No. 18H05226, No. 19H00682, No. 20H05850, No. 20H05858, No. 22H00144, No. 22K14056, No. 22K21347, No. 23H05433, No. 26220706, and No. 26400255, and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan; National Research Foundation (NRF) of Korea Grants No. 2016R1-D1A1B-02012900, No. 2018R1-A6A1A-06024970, No. 2021R1-A6A1A-03043957, No. 2021R1-F1A-1060423, No. 2021R1-F1A-1064008, No. 2022R1-A2C-1003993, No. 2022R1-A2C-1092335, No. RS-2023-00208693, No. RS-2024-00354342 and No. RS-2022-00197659, Radiation Science Research Institute, Foreign Large-Size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center, the Korea Institute of Science and Technology Information (K24L2M1C4) and KREONET/GLORIAD; Universiti Malaya RU grant, Akademi Sains Malaysia, and Ministry of Education Malaysia; Frontiers of Science Program Contracts No. FOINS-296, No. CB-221329, No. CB-236394, No. CB-254409, and No. CB-180023, and SEP-CINVESTAV Research Grant No. 237 (Mexico); 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 No. S-0256-1438 and No. S-0280-1439 (Saudi Arabia), and King Saud University, Riyadh, Researchers Supporting Project No. RSPD2024R873 (Saudi Arabia); Slovenian Research Agency and Research Grants No. J1-9124 and No. P1-0135; Ikerbasque, Basque Foundation for Science, the State Agency for Research of the Spanish Ministry of Science and Innovation through Grant No. PID2022-136510NB-C33, Agencia Estatal de Investigacion, Spain Grant No. RYC2020-029875-I and Generalitat Valenciana, Spain Grant No. CIDEGENT/2018/020; the Swiss National Science Foundation; The Knut and Alice Wallenberg Foundation (Sweden), Contracts No. 2021.0174 and No. 2021.0299; National Science and Technology Council, and Ministry of Education (Taiwan); Thailand Center of Excellence in Physics; TUBITAK ULAKBIM (Turkey); National Research Foundation of Ukraine, Project No. 2020.02/0257, and Ministry of Education and Science of Ukraine; the U.S. National Science Foundation and Research Grants No. PHY-1913789 and No. PHY-2111604, and the U.S. Department of Energy and Research Awards No. DE-AC06-76RLO1830, No. DE-SC0007983, No. DE-SC0009824, No. DE-SC0009973, No. DE-SC0010007, No. DE-SC0010073, No. DE-SC0010118, No. DE-SC0010504, No. DE-SC0011784, No. DE-SC0012704, No. DE-SC0019230, No. DE-SC0021274, No. DE-SC0021616, No. DE-SC0022350, No. DE-SC0023470; and the Vietnam Academy of Science and Technology (VAST) under Grants No. NVCC.05.12/22-23 and No. DL0000.02/24-25. We thank the SuperKEKB team for delivering high-luminosity collisions; the KEK cryogenics group for the efficient operation of the detector solenoid magnet and IBBelle on site; the KEK Computer Research Center for on-site computing support; the NII for SINET6 network support; and the raw-data centers hosted by BNL, DESY, GridKa, IN2P3, INFN, PNNL/EMSL, and the University of Victoria.

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.

Data availability—The data that support the findings of this Letter are not publicly available. The data are available from the authors upon reasonable request.

[1] Sw. Banerjee *et al.* (Heavy Flavor Averaging Group), arXiv:2411.18639.
[2] D. Bećirević, O. Sumensari, and R. Zukanovich Funchal, Eur. Phys. J. C **76**, 134 (2016).
[3] D. Bećirević, F. Jaffredo, J. P. Pinheiro, and O. Sumensari, Phys. Rev. D **110**, 075004 (2024).
[4] R. Mohanta, Eur. Phys. J. Spec. Top. **233**, 2105 (2024).
[5] I. Adachi *et al.* (Belle II Collaboration), Phys. Rev. D **109**, 112006 (2024).
[6] L. Allwicher, D. Bečirević, G. Piazza, S. Rosauro-Alcaraz, and O. Sumensari, Phys. Lett. B **848**, 138411 (2024).
[7] J. P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **86**, 012004 (2012).
[8] S. Watanuki *et al.* (Belle Collaboration), Phys. Rev. Lett. **130**, 261802 (2023).
[9] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 06 (2020) 129.
[10] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 06 (2023) 143.
[11] S. Kurokawa and E. Kikutani, Nucl. Instrum. Methods Phys. Res., Sect. A **499**, 1 (2003); T. Abe *et al.*, Prog. Theor. Exp. Phys. **2013**, 03A001 (2013).
[12] K. Akai, K. Furukawa, and H. Koiso, Nucl. Instrum. Methods Phys. Res., Sect. A **907**, 188 (2018).
[13] T. Abe *et al.* (Belle II Collaboration), arXiv:1011.0352.
[14] A. Abashian *et al.* (Belle Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **479**, 117 (2002).
[15] D. J. Lange, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001).
[16] E. Barberio and Z. Was, Comput. Phys. Commun. **79**, 291 (1994).
[17] B. Ward, S. Jadach, and Z. Was, Nucl. Phys. B, Proc. Suppl. **116**, 73 (2003).
[18] T. Sjöstrand, S. Mrenna, and P. Skands, Comput. Phys. Commun. **178**, 852 (2008).
[19] R. Brun *et al.*, Report No. CERN-DD-EE-84-01, 1987.
[20] S. Agostinelli *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).
[21] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun (Belle II Collaboration), Comput. Software Big Sci. **3**, 1 (2019).
[22] M. Gelb *et al.*, Comput. Software Big Sci. **2**, 9 (2018).
[23] T. Keck *et al.*, Comput. Software Big Sci. **1**, 6 (2019).
[24] S. Navas *et al.* (Particle Data Group), Phys. Rev. D **110**, 030001 (2024).
[25] E. Won *et al.* (Belle Collaboration), Phys. Rev. D **80**, 111101(R) (2009).
[26] E. Nakano, Nucl. Instrum. Methods Phys. Res., Sect. A **494**, 402 (2002).
[27] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, Nucl. Instrum. Methods Phys. Res., Sect. A **485**, 490 (2002).
[28] D. N. Brown, J. Ilic, and G. B. Mohanty, Nucl. Instrum. Methods Phys. Res., Sect. A **592**, 254 (2008).
[29] A. Khotanzad and Y. Hong, IEEE Trans. Pattern Anal. Mach. Intell. **12**, 489 (1990).
[30] S. Longo *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **982**, 164562 (2020).
[31] J. D. Bjorken and S. J. Brodsky, Phys. Rev. D **1**, 1416 (1970).
[32] S. H. Lee *et al.* (Belle Collaboration), Phys. Rev. Lett. **91**, 261801 (2003).
[33] G. Punzi, eConf **C030908**, MODT002 (2003).
[34] N. L. Johnson, Biometrika **36**, 149 (1949).
[35] L. Aggarwal *et al.* (Belle II Collaboration), Phys. Rev. Lett. **131**, 051804 (2023).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/yzvh-v3hm for the selection efficiency of the phase-space model as a function of $(M^2_{\tau\ell}, M^2_{K^0_S\ell})$ in the Supplemental Material, to allow these results to be reinterpreted in specific BSM models.

Correction: Unnecessary labels have been removed from the figures.