Belle II Preprint 2024-012
KEK Preprint 2024-6

Prepared for submission to JHEP

Search for the lepton-flavor-violating $\tau^- \to e^\mp \ell^\pm \ell^-$ decays at Belle II

arXiv:2507.18236v1 [hep-ex] 24 Jul 2025

The Belle II Collaboration

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, 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, T. Bilka, D. Biswas, A. Bobrov, D. Bodrov, A. Bondar, G. Bonvicini, J. Borah, A. Boschetti, A. Bozek, M. Bračko, P. Branchini, N. Brenny, T. E. Browder, A. Budano, S. Bussino, Q. Campagna, M. Campajola, L. Cao, G. Casarosa, C. Cecchi, M.-C. Chang, R. Cheaib, P. Cheema, C. Chen, L. Chen, B. G. Cheon, K. Chilikin, J. Chin, K. Chirapatpimol, H.-E. Cho, K. Cho, S.-J. Cho, S.-K. Choi, S. Choudhury, J. Cochran, I. Consigny, L. Corona, J. X. Cui, E. De La Cruz-Burelo, S. A. De La Motte, G. De Nardo, G. De Pietro, R. de Sangro, M. Destefanis, S. Dey, A. Di Canto, F. Di Capua, J. Dingfelder, Z. Doležal, I. Domínguez Jiménez, T. V. Dong, X. Dong, M. Dorigo, K. Dugic, G. Dujany, P. Ecker, D. Epifanov, J. Eppelt, R. Farkas, P. Feichtinger, T. Ferber, T. Fillinger, C. Finck, G. Finocchiaro, A. Fodor, F. Forti, A. Frey, B. G. Fulsom, A. Gabrielli, A. Gale, E. Ganiev, M. Garcia-Hernandez, R. Garg, G. Gaudino, V. Gaur, V. Gautam, A. Gaz, A. Gellrich, G. Ghevondyan, D. Ghosh, H. Ghumaryan, G. Giakoustidis, R. Giordano, A. Giri, P. Gironella Gironell, A. Glazov, B. Gobbo, R. Godang, O. Gogota, P. Goldenzweig, W. Gradl, 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, G. Heine, I. Heredia de la Cruz, M. Hernández Villanueva, T. Higuchi, M. Hoek, M. Hohmann, R. Hoppe, P. Horak, C.-L. Hsu, 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, Q. P. Ji, S. Jia, Y. Jin, A. Johnson, K. K. Joo, H. Junkerkalefeld, D. Kalita, A. B. Kaliyar, J. Kandra, K. H. Kang, G. Karyan, T. Kawasaki, F. Keil, C. Ketter, M. Khan, C. Kiesling, C.-H. Kim, D. Y. Kim, J.-Y. Kim, K.-H. Kim, Y. J. Kim, Y.-K. Kim, H. Kindo, K. Kinoshita, P. Kodyš, T. Koga, S. Kohani, K. Kojima, A. Korobov, S. Korpar, E. Kovalenko, 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, R. Leboucher, F. R. Le Diberder, M. J. Lee, C. Lemettais, P. Leo, P. M. Lewis, H.-J. Li, L. K. Li, Q. M. Li, S. X. 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, G. Nazaryan, M. Neu, S. Nishida, S. Ogawa, R. Okubo, H. Ono, Y. Onuki, G. Pakhlova, A. Panta, S. Pardi, K. Parham, H. Park, J. Park, K. Park, S.-H. Park, B. Paschen, A. Passeri, S. Patra, S. Paul, T. K. Pedlar, I. Peruzzi, R. Peschke, R. Pestotnik, M. Piccolo, L. E. Piilonen, P. L. M. Podesta-Lerma, T. Podobnik, S. Pokharel, A. Prakash, C. Praz, S. Prell, E. Prencipe, M. T. Prim, S. Privalov, H. Purwar, P. Rados, G. Raeuber, S. Raiz, 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, L. Salutari, D. A. Sanders, S. Sandilya, L. Santelj, V. Savinov, B. Scavino, J. Schmitz, S. Schneider, M. Schnepf, K. Schoenning, C. Schwanda, A. J. Schwartz, Y. Seino, A. Selce, K. Senyo, J. Serrano, M. E. Sevior, C. Sfienti, W. Shan, G. 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, W. Song, S. Spataro, B. Spruck, M. Starič, P. Stavroulakis, S. Stefkova, L. Stoetzer, 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, F. F. Trantou, I. Tsaklidis, I. Ueda, T. Uglov, K. Unger, Y. Unno, K. Uno, S. Uno, P. Urquijo, Y. Ushiroda, S. E. Vahsen, R. van Tonder, K. E. Varvell, M. Veronesi, A. Vinokurova,

V. S. Vismaya, L. Vitale, V. Vobbilisetti, R. Volpe, A. Vossen, M. Wakai, S. Wallner, M.-Z. Wang, X. L. Wang, A. Warburton, M. Watanabe, S. Watanuki, C. Wessel, E. Won, X. P. Xu, B. D. Yabsley, S. Yamada, W. Yan, W. C. Yan, S. B. Yang, J. Yelton, K. Yi, J. H. Yin, K. Yoshihara, C. Z. Yuan, J. Yuan, L. Zani, F. Zeng, M. Zeyrek, B. Zhang, V. Zhilich, J. S. Zhou, Q. D. Zhou, L. Zhu, R. Žlebčík

E-mail: coll-publications@belle2.org

Abstract: We present the result of a search for the charged-lepton-flavor violating decays $\tau^- \to e^\mp \ell^\pm \ell^-$, where ℓ is a muon or an electron, using a data sample with an integrated luminosity of 428 fb^{-1} recorded by the Belle II experiment at the SuperKEKB e^+e^- collider. The selection of $e^+e^- \to \tau^+\tau^-$ events containing a signal candidate is based on an inclusive-tagging reconstruction and on a boosted decision tree to suppress background.

Upper limits on the branching fractions between 1.3 and 2.5 $\times 10^{-8}$ are set at the 90% confidence level. These results are the most stringent bounds to date for four of the modes.

Contents

1 Introduction 1

2 The Belle II detector, simulation and data samples 2

3 Candidate reconstruction 3

4 Background rejection 4

5 Fitting procedure 8

6 Systematic uncertainties 10

7 Result 11

8 Summary 13

1 Introduction

In the standard model (SM) neutrinos are assumed to be massless and charged-lepton flavor is conserved. However, this symmetry is broken at loop-level when taking into account neutrino mixing, which implies the existence of charged-lepton-flavor violation, and thus, the existence of processes such as $\mu \to e$, $\tau \to e$ and $\tau \to \mu$. In the simplest SM extension that allows for massive neutrinos, all charged-lepton-flavor-violating (LFV) amplitudes are proportional to the differences of the relevant squared neutrino masses. This results in predicted decay rates of 10^{-50} that are well below the sensitivities of any experiment [1–3]. The observation of LFV decays would thus provide indisputable evidence of physics beyond the SM.

Over the past four decades, the CLEO experiment at CESR, and the first generation B-factory experiments, BaBar at SLAC and Belle at KEK, have searched for LFV in τ lepton decays. In total, 52 LFV τ decays with neutrinoless two-body or three-body final states have been investigated [4]. Among these, $\tau^- \to \ell^-\ell^+\ell^-$ decays,[1] where $\ell = e, \mu$, and in particular $\tau^- \to \mu^-\mu^+\mu^-$, have garnered significant attention in recent years due to the potential enhancement of the branching fraction up to a value of 10^{-8} in scenarios beyond the SM [5–10]. The most stringent limit of $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 1.9 \times 10^{-8}$ at the 90% confidence level (C.L.) was obtained by Belle II [11]. The $\tau^- \to e^\mp\ell^\pm\ell^-$ decays include three final states in which lepton flavour is violated by one unit ($e^-e^+\mu^-$, $e^-e^+e^-$, and $e^-\mu^+\mu^-$) and two final states where it is violated by two units ($e^-e^-\mu^+$, $e^+\mu^-\mu^-$). These decays are enhanced in models such as Type II Seesaw [12], and are accessible to Belle II.

[1] Charge conjugation is implied throughout this paper.

In addition, they make it possible to probe the existence of axion-like particles (X) that could arise through the decays $\tau^- \to X\ell^-, X \to \ell^+\ell^-$ [13]. Searches for $\tau^- \to e^\mp\ell^\pm\ell^-$ decays were previously performed by the B-factories and the most stringent limits were obtained by Belle in the range 1.5–2.7 $\times 10^{-8}$ at the 90% C.L. using a data sample with an integrated luminosity of 782 fb^{-1} [14, 15].

We report the results of a search for the five LFV $\tau^- \to e^\mp\ell^\pm\ell^-$ decays using data collected with the Belle II detector [16] at the asymmetric-energy e^+e^- SuperKEKB collider [17] between 2019 and 2022. The data sample has an integrated luminosity of 428 fb^{-1} and was recorded at e^+e^- center-of-mass energies of 10.58 GeV (365 fb^{-1}), 10.52 GeV (43 fb^{-1}), and at various energies around 10.75 GeV (20 fb^{-1}). This is equivalent to 393 million produced τ-pairs [18].

The signal candidates are selected with an inclusive-tagging method that was used for the first time by the Belle II collaboration for the 3μ final state [11]. The background rejection strategy is based on a preselection followed by a requirement on the output of boosted decision trees that are trained on control regions in data. The branching fractions are extracted from a fit to the invariant masses of the three-lepton signal candidates.

2 The Belle II detector, simulation and data samples

The Belle II detector consists of several subdetectors arranged in a cylindrical structure around the e^+e^- interaction point [16]. Charged-particle trajectories (tracks) are reconstructed using a two-layer silicon-pixel detector, surrounded by a four-layer double-sided silicon-strip detector and a central drift chamber (CDC). Only 15% of the second pixel layer was installed when the data were collected. The CDC also provides dE/dx energy-loss measurements for particle identification. Outside the CDC is a time-of-propagation (TOP) detector and an aerogel ring-imaging Cherenkov (ARICH) detector, both of which cover the barrel and forward endcap regions, respectively. The forward region is by definition aligned with the electron beam direction. An electromagnetic calorimeter (ECL), divided into forward endcap, barrel, and backward endcap regions, fills the remaining volume inside a 1.5 T superconducting solenoid and is used to reconstruct photons and identify electrons. A K_L^0 and muon detection system (KLM) based on resistive-plate chambers and plastic scintillator modules is installed in the iron flux return of the solenoid. The z axis in the laboratory frame is defined along the detector solenoid axis, with the positive direction along the electron beam. The polar angle θ and the transverse plane are defined relative to this axis.

Monte-Carlo (MC) simulated events are used to estimate the selection efficiency and optimize the selection. We use 10 million $e^+e^- \to \tau^+\tau^-$ events, in which one τ decays to three leptons following a phase space model, and the other has a SM decay according to the branching fractions from Ref. [19]. The 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 $\ell = e, \mu$; $e^+e^- \to e^+e^-h^+h^-$ events, where h indicates a pion, kaon, or proton; and four-lepton processes: $e^+e^- \to e^+e^-e^+e^-, \mu^+\mu^-\mu^+\mu^-$, $\mu^+\mu^-e^+e^-, e^+e^-\tau^+\tau^-, \mu^+\mu^-\tau^+\tau^-$. The $e^+e^- \to \tau^+\tau^-$ process is generated using the

KKMC generator [20]. The τ decays are simulated by the Belle II version of the TAUOLA generator [21] and their final-state radiated photons by the PHOTOS package [22]. We use KKMC to simulate $\mu^+\mu^-(\gamma)$ and $q\bar{q}$ production; the PYTHIA program [23] for the fragmentation of the $q\bar{q}$ pair; the EvtGen package [24], interfaced to PYTHIA, for the production of $e^+e^- \to B\bar{B}$ events and decays of produced hadrons. PHOTOS is also used by EvtGen and KKMC to simulate final state radiations. We use the BabaYaga@NLO generator [25–29] for $e^+e^- \to e^+e^-(\gamma)$ events; and the AAFH program [30–32] and the TREPS generator [33] for the production of non-radiative four-leptons and $e^+e^-h^+h^-$ final states. The size of the simulated samples for $e^+e^- \to \tau^+\tau^-$ and $e^+e^- \to q\bar{q}$ events is equivalent to an integrated luminosity of 8 ab^{-1}, while it ranges between 100 fb^{-1} and 2 ab^{-1} for the other processes. In particular, the equivalent luminosity of the $e^+e^- \to e^+e^-e^+e^-$ and $\mu^+\mu^-e^+e^-$ samples is 200 fb^{-1}, and that of $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$, $e^+e^-\tau^+\tau^-$ and $\mu^+\mu^-\tau^+\tau^-$ is 2 ab^{-1}.

The events are selected by a hardware trigger that is based on the energy deposits (clusters) and their topologies in the ECL. Most of the events are selected requiring a total energy in the ECL larger than 1 GeV and a topology that is incompatible with Bhabha events. The trigger efficiency on reconstructed signal candidates, as defined in the following Section, is about 95%. Trigger lines based on tracks reconstructed in the CDC are also used to obtain systematic uncertainties from control samples. The Belle II analysis software [34, 35] uses the GEANT4 [36] package to simulate the response of the detector to the passage of the particles and also provides a simulation of the trigger selection algorithms.

3 Candidate reconstruction

In the e^+e^- center-of-mass (c.m.) frame, τ leptons are produced in opposite directions, with the decay products of one τ isolated from those of the other τ and contained in opposite hemispheres. The boundary between the hemispheres is experimentally defined by the plane perpendicular to the vector $\hat{\mathbf{t}}$ that maximizes the thrust value (T):

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

where $\mathbf{p}_i^*$ is the 3-momentum of final-state particle i in the e^+e^- c.m. frame[2] [37, 38]. The sum is over both charged and neutral particles.

Signal $\tau^- \to e^\mp \ell^\pm \ell^-$ candidates are reconstructed by combining one electron and two other lepton candidates with a total charge of ± 1, belonging to the same hemisphere. The displacement of the leptons from the average interaction point must be less than 3 cm along the z axis and less than 1 cm in the transverse plane. The signal τ vertex is obtained by the TreeFitter tool [39], which also provides refitted trajectories of the final state particles. Muon candidates are identified using the criterion $\mathcal{P}_\mu = \mathcal{L}_\mu/(\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d) > 0.5$ where the likelihoods $\mathcal{L}_i$ for each charged-particle hypothesis $i = e, \mu, \pi, K$,proton (p),

[2] In this paper quantities marked with an asterisk are calculated in the e^+e^- c.m. frame

deuteron (d) combine particle identification information from the CDC, TOP, ARICH, ECL, and KLM subdetectors.

Electron candidates are identified using a boosted decision tree classifier trained to separate electrons from all other charged particles [40]. Inputs to the classifier are the likelihoods from each sub-detector, as well as additional ECL observables, such as variables characterising the cluster's spatial structure. We use the output of the classifier, $\mathcal{P}_e$, as a discriminator for electron identification, requiring $\mathcal{P}_e > 0.5$. The electron four-momentum is corrected for energy loss due to bremsstrahlung by adding back the energies of photons reconstructed within a cone of 8.6 degrees around the initial direction of the electron and with an energy greater than 20 MeV. Photons are reconstructed from ECL clusters within the CDC acceptance ($17° < \theta < 150°$) and not associated with any tracks.

We search for signal events in the two-dimensional plane consisting of $M_{e\ell\ell}$ and $\Delta E_{e\ell\ell}$. The invariant mass, $M_{e\ell\ell}$, is determined from the three charged particles in the decay. The energy difference, $\Delta E_{e\ell\ell}$, is the difference between the energy of the three leptons in the c.m. system and half the beam energy: $\Delta E_{e\ell\ell} = E^*_{e\ell\ell} - \sqrt{s}/2$.

The $\tau^- \to e^\mp \ell^\pm \ell^-$ decays are neutrinoless processes, hence the invariant mass $M_{e\ell\ell}$ should be consistent with the mass of the τ lepton while the energy difference $\Delta E_{e\ell\ell}$ should be close to zero. The signal peak in the ($M_{e\ell\ell}$, $\Delta E_{e\ell\ell}$) two-dimensional distribution is broadened by detector resolution and radiative effects. The radiation of photons from the initial state (ISR) leads to a tail at low values of $\Delta E_{e\ell\ell}$ while final state radiation (FSR) produces a tail at low values for $M_{e\ell\ell}$ and $\Delta E_{e\ell\ell}$. Several rectangular regions are defined in the ($M_{e\ell\ell}$, $\Delta E_{e\ell\ell}$) plane. Blind regions, in which data are hidden until the analysis strategy is fixed in order to mitigate experimental bias, have boundaries defined to retain 99% of signal candidates ignoring events from the ISR tails. The fit regions, defined in Section 4, contain the events that will be used in the fit to obtain the branching fractions. Events outside the blind region with $1.4 < M_{e\ell\ell} < 2.0\,\mathrm{GeV}/c^2$ and $-1.0 < \Delta E_{e\ell\ell} < 0.5$ GeV are used to further refine the selection as explained in Section 4. The regions for the case of $\tau^- \to e^- e^+ e^-$ decays are shown in Fig. 1 together with the signal Monte Carlo events.

4 Background rejection

The background rejection for each decay mode is performed in two steps: first using pre-selection requirements to remove the main contributions, and second using the output of a Boosted Decision Tree (BDT). The background rejection primarily relies on global event variables, which are calculated using all tracks and photons in the event satisfying the following requirements aimed at suppressing beam-related backgrounds. Tracks must be displaced from the average interaction point by less than 3 cm along the z axis and less than 1 cm in the transverse plane and are assumed to be pions. Neutral pions are obtained from the combination of two photons with energy greater than 0.1 GeV, having an invariant mass within $0.115 < M_{\gamma\gamma} < 0.152$ GeV/c^2, which corresponds to a range of approximately ± 2.5 times the experimental resolution around the known π^0 mass [41]. Photons not used in π^0 reconstruction must have an energy greater than 0.2 GeV. These selected tracks and photons are used to define variables related to the kinematic properties of the event such as

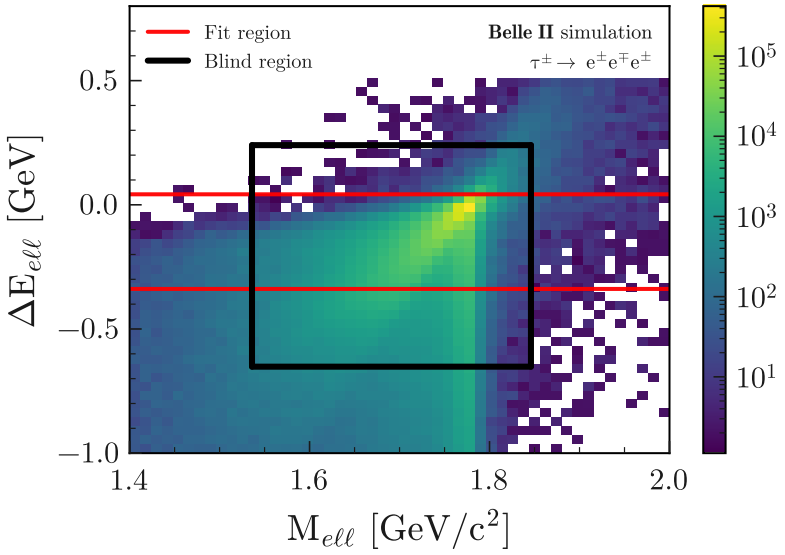

Figure 1. The signal distribution in the $(M_{e\ell\ell}, \Delta E_{e\ell\ell})$ plane for MC simulated $\tau^- \to e^-e^+e^-$ decays. The blind region is the region inside the black lines, while the signal fit region is the one between the red lines. The distributions for the other modes are given in the Additional Material section.

the thrust axis, the missing momentum, defined as the difference between the momentum of the initial e^+e^- system and that of all reconstructed tracks and photons in the event, and the corresponding missing mass.

We also use properties of the rest of the event (ROE) that correspond to the tracks and photons not used in the reconstruction of the signal decays. The ROE is built from the tracks with an additional requirement that their transverse momenta be larger than 0.075 GeV/c. The mass of each ROE particle is assigned on the basis of the most likely particle identification hypothesis. In addition, particles in the ROE are required to be within the CDC angular acceptance, and the total charge of the ROE and signal tracks is required to be 0.

One of the main background components is from radiative dilepton and four-lepton events (low-multiplicity backgrounds). In addition, background can arise from $e^+e^- \to q\overline{q}$ events, where pions are misidentified as muons. To suppress backgrounds from events with misidentified leptons, we tighten the lepton identification criteria requiring at least one electron and one muon with $\mathcal{P}_{e,\mu} > 0.9$. For the $\tau^- \to e^-e^+e^-$ decay, all three electrons must have $\mathcal{P}_e > 0.9$. For some modes, the comparison between data and simulation outside the blind region shows a large data excess, as illustrated in Fig. 2 for the $\tau^- \to e^-\mu^+e^-$ decay. This excess is consistent with low-multiplicity backgrounds such as four-lepton final state processes with initial and final state radiation, which are not simulated. Processes with converted photons are rejected using a minimal threshold on the invariant mass of opposite lepton pairs computed assuming an electron mass hypothesis, $M^{\gamma}_{\ell_1\ell_2}$. This threshold is set to 50 MeV/c^2 for the $\tau^- \to e^-e^+e^-$ decay and 25 MeV/c^2 for the other decays. In addition, the total visible energy in the c.m., $E^*_{\rm vis}$, is required to be less than 10.4 GeV, and

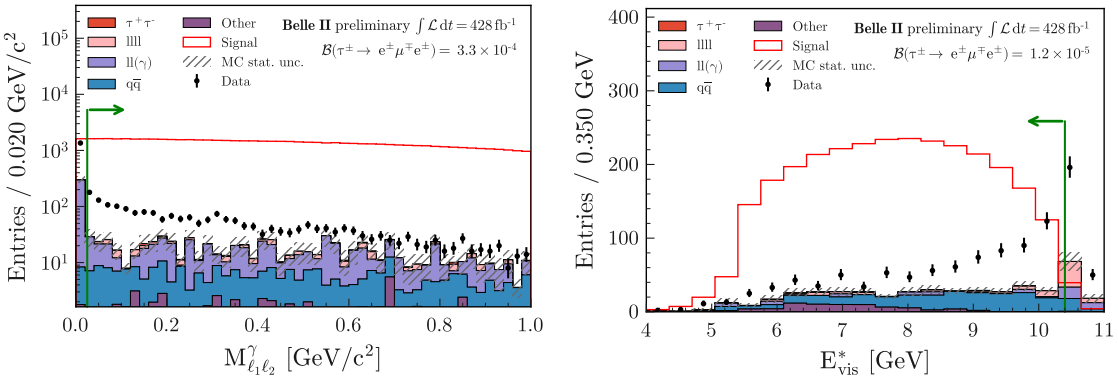

Figure 2. Distribution of the invariant mass of one of the e-μ systems (left) and the visible energy in the c.m. (right) for the $\tau^- \to e^-\mu^+e^-$ mode. The green arrows correspond to the applied selections. The selection on the invariant mass of the lepton pairs is applied for the visible energy distribution. The various simulated background processes are shown as a stack of color-filled histograms, with statistical uncertainties displayed as hatched areas. The signal, not blinded, is shown as a red histogram with branching fraction values given on the plots.

the magnitude of the thrust vector is required to be less than 0.97. Finally, a requirement on the polar angle of the missing momentum in the c.m. is set, $0.3 < \theta^*_{\text{miss}} < 2.7\,\text{rad}$, to discard two-photon events with missing energy along the beam axis.

The agreement between data and simulated events after the preselection outside the blind region is shown in Fig. 3 for the $M_{e\ell\ell}$ and $\Delta E_{e\ell\ell}$ variables. An excess of events is still observed in data. Those candidates mostly correspond to four-track background events in which the particle not used in the signal τ reconstruction has a high probability to be an electron. In addition, those events have a high thrust value and a missing momentum pointing at the boundaries of the polar angular acceptance in the c.m. frame. Therefore, they are consistent with low-multiplicity backgrounds such as four-lepton final state processes with initial and final state radiation. As the simulated events do not provide a reliable description of the candidates observed in the sidebands, we define a data-driven strategy based on a BDT classifier to control the remaining background. In order to avoid biases, it is important that the data events used to train the BDTs are not used later in the fit to obtain the branching fractions. We define the fit region, whose events are excluded from the BDT training, as a slice in $\Delta E_{e\ell\ell}$ chosen to retain 90% of signal candidates satisfying the preselection requirements. The BDT is trained on 15,000 simulated signal events and data events that are outside the fit region and blind region. The boundaries of the fit region as well as the number of data events used in the BDT training are given in Table 1.

For each final state, the BDT input variables are chosen among a common set of 32 variables related to three distinct categories. Variables that are highly correlated both for signal and background, with a linear correlation coefficient larger than 0.85, or that have low discriminating power are removed. The first category consists of variables associated with the signal τ, such as the polar angle of each lepton and ordered energies, the invari-

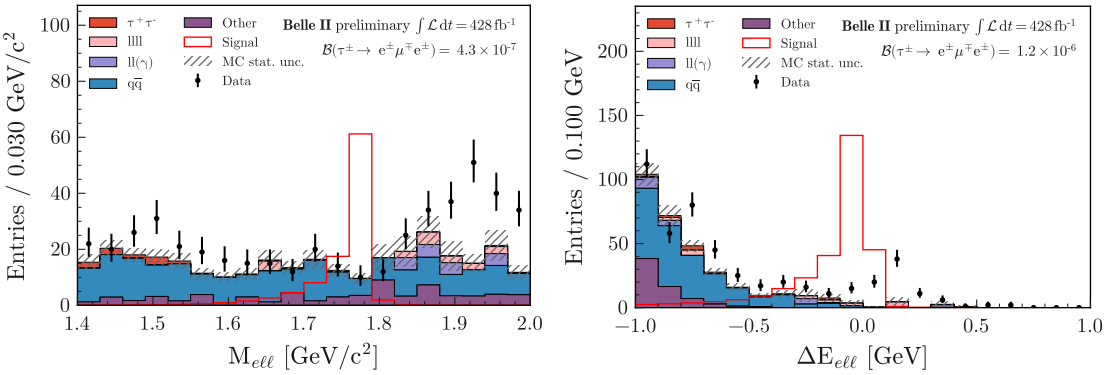

Figure 3. $M_{e\ell\ell}$ (left) and $\Delta E_{e\ell\ell}$ (right) distributions for the $\tau^- \to e^-\mu^+e^-$ mode for data and simulation outside the blind region after the preselection. The various simulated background processes are shown as a stack of color-filled histograms, with statistical uncertainties displayed as hatched areas. The signal, not blinded, is shown as a red histogram with branching fraction values given on the plots.

ant masses of the three lepton-pair combinations, the flight time of the τ divided by its uncertainty, the τ azimuthal angle and the cosine of its polar angle. The second category involves variables related to the ROE properties such as its mass, defined as the mass of the four-vector resulting from the sum of all reconstructed objects forming the ROE, its energy, the total energy of the clusters in the ROE and the number of neutral clusters in the ROE. We also use a categorical variable based on the numbers of electrons, muons and pions in the ROE, that assigns a distinct value to the standard model decay modes of the non-signal τ. The third category of input variables comprises the thrust value, the cosine of the polar angle between the $\tau^- \to e^\mp\ell^\pm\ell^-$ momentum and the thrust axis, the total visible energy in the event, the numbers of photons and the total photon energy in the event, and variables related to the missing momentum of the event. The latter include its magnitude and polar angle in the c.m. frame, the angle between the missing momentum and each of the three leptons, and the square of the invariant missing mass. This category also comprises variables related to the event shape such as the reduced Fox Wolfram moment R_2 [42] and the first three CLEO cones, which characterize the energy flux in a cone with opening angle of 10, 20 or 30 degrees around the thrust axis [43].

Table 1. Definition of the $\Delta E_{e\ell\ell}$ boundaries for the fit region and corresponding background yields N_{bg} used to train the BDTs.

	$\Delta E_{e\ell\ell}^{\text{fit,low}}$ [GeV]	$\Delta E_{e\ell\ell}^{\text{fit,high}}$ [GeV]	N_{bg}
$e^-e^+e^-$	−0.34	0.04	768
$e^-e^+\mu^-$	−0.29	0.05	991
$e^-\mu^+e^-$	−0.30	0.03	452
$\mu^-\mu^+e^-$	−0.21	0.03	1,471
$\mu^-e^+\mu^-$	−0.21	0.03	625

80% of the training samples are used to train the BDTs with the XGBoost library [44], while the remaining 20% are used for validation, control of overtraining, and optimization of the parameters with the `Optuna` library [45]. Figure 4 shows the distribution of the BDT score, s_{BDT}, in the sidebands, defined as the fit region outside the blind region, for data and simulation. The requirements applied on the BDT scores are chosen by scanning the score in the range [0.5,1] with a step of 0.05, and selecting the one that gives the best expected upper limit, following the procedure explained in Section 7. For each mode, Table 2 shows the nominal cut on the BDT score, the absolute signal efficiency, as well as the number of remaining events in the data sidebands. Possible correlations between the BDT score and $M_{e\ell\ell}$ have been checked in the validation samples and background simulation and are found to be negligible.

Table 2. Nominal cut on the BDT score s_{BDT} for each mode, with the corresponding number of remaining background events N_{SB} in the data sidebands and total signal efficiencies ϵ_{sig}. The uncertainty on the signal efficiencies is due to the size of the MC sample.

	s_{BDT}	N_{SB}	ϵ_{sig}
$e^-e^+e^-$	0.95	3	$(15.0 \pm 0.1)\%$
$e^-e^+\mu^-$	0.8	6	$(20.4 \pm 0.1)\%$
$e^-\mu^+e^-$	0.5	6	$(23.5 \pm 0.1)\%$
$\mu^-\mu^+e^-$	0.8	12	$(20.1 \pm 0.1)\%$
$\mu^-e^+\mu^-$	0.5	4	$(24.1 \pm 0.1)\%$

5 Fitting procedure

We extract the branching fractions using unbinned maximum likelihood fits to the three-lepton invariant masses. The likelihood is expressed as:

$$\mathcal{L} = \frac{e^{-(n_{\mathrm{sig}}+n_{\mathrm{bg}})}}{N!} \prod_{i=1}^{N} (n_{\mathrm{sig}} \times \mathcal{P}_{\mathrm{sig}}(M^i_{e\ell\ell}) + n_{\mathrm{bg}} \times \mathcal{P}_{\mathrm{bg}}(M^i_{e\ell\ell})), \tag{5.1}$$

where $n_{\mathrm{sig}}(n_{\mathrm{bg}})$ and $\mathcal{P}_{\mathrm{sig}}(\mathcal{P}_{\mathrm{bg}})$ are the number of events and probability density function (PDF) for the signal (background) and N is the total number of events. We express the number of signal events as $n_{\mathrm{sig}} = 2 \times \mathcal{B}(\tau^- \to e^{\mp}\ell^{\pm}\ell^-) \times \epsilon_{\mathrm{sig}} \times \sigma_{\tau\tau} \times L$, where ϵ_{sig} is the signal efficiency, L is the integrated luminosity, and $\sigma_{\tau\tau}$ is the tau-pair production cross section.

The signal PDF is an asymmetric double-sided crystal ball (CB) function [46], which accounts for the long FSR tail at low mass, as described in Section 3. The parameters of the signal PDF are obtained by fitting the mass distribution of simulated signal events. For the background, an exponential function is used:

$$\mathcal{P}_{\mathrm{bg}}(M_{e\ell\ell}) = A \cdot \exp(C_{\mathrm{bg}} \cdot M_{e\ell\ell}), \tag{5.2}$$

where the constant A normalizes the background PDF to unity in the fit range, between 1.4 and 2.0 GeV/c^2.

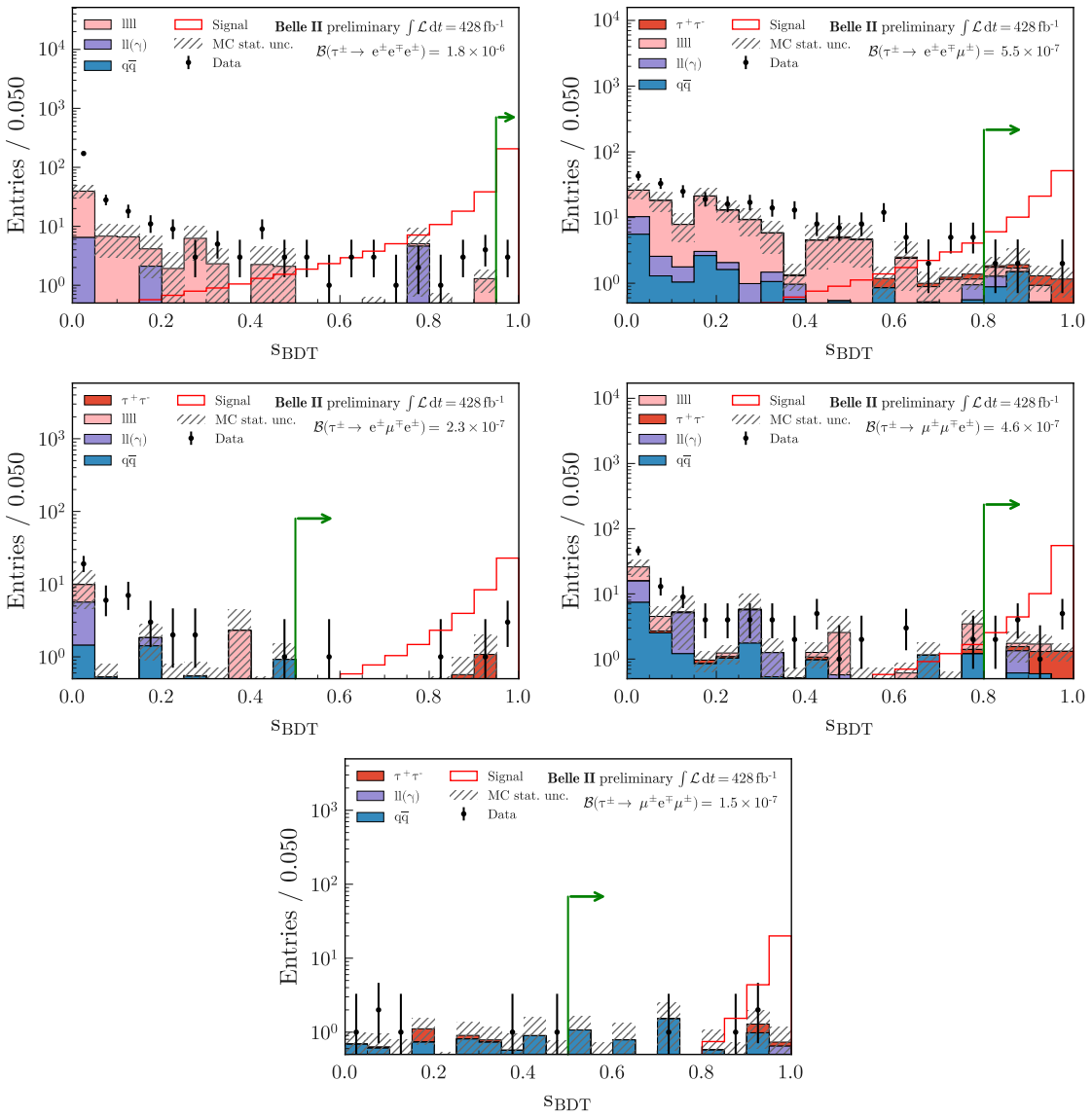

Figure 4. BDT score distribution in the sideband region after the preselection. The green arrows indicate the selection criteria applied to the BDT score. The various simulated background processes are shown as a stack of color-filled histograms, with statistical uncertainties displayed as hatched areas. The signal in the full fitting region is shown as a red histogram with branching fraction values given on the plots.

The parameters allowed to vary in the fit are $\mathcal{B}(\tau^- \to e^\mp \ell^\pm \ell^-)$, which is constrained to be positive, $n_{\rm bg}$, and $C_{\rm bg}$. The fit is validated using pseudo-experiments with different branching ratio values and exponent coefficients obtained from fits to the data sidebands. Due to the constraint of non-negative $\mathcal{B}(\tau^- \to e^\mp \ell^\pm \ell^-)$, a small positive bias is observed, which results in a conservative limit.

6 Systematic uncertainties

One category of systematic uncertainties arises from differences between experimental data and simulation due to possible mismodeling in the generation and reconstruction of the simulated samples, and affects the signal efficiency uncertainty. We take into account the systematic uncertainty associated with the corrections to the simulated muon and electron identification efficiencies, derived from auxiliary measurements in data using $J/\psi \to \ell^+\ell^-$, $e^+e^- \to \ell^+\ell^-\gamma$, and $e^+e^- \to e^+e^-\ell^+\ell^-$ 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 within 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 in the range 1.0 to 1.8%.

The difference between data and simulation in track-reconstruction efficiency is measured in $e^+e^- \to \tau^+\tau^-$ events, selecting $\tau^- \to e^-\overline{\nu}_e\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ decays. Good agreement is observed within the associated uncertainty of 0.24% per track, resulting in a systematic uncertainty of 1.0% for the $\tau^- \to e^\mp \ell^\pm \ell^-$ decay.

The agreement between data and simulation for the trigger efficiency is evaluated using the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample, which is reconstructed in the same way as the signal decays, replacing the final-state leptons with pions having $\mathcal{P}_\pi > 0.9$. Additional requirements on event variables are used to remove backgrounds that mainly originate from $q\bar{q}$ events, and candidates in the range $0.5 < M_{3\pi} < 1.7\,\mathrm{GeV}/c^2$ and $-1 < \Delta E_{3\pi} < 0$ GeV are selected. In data, the trigger efficiency is computed using independent trigger selections: the efficiency of the ECL-based trigger selection is obtained using events triggered by the CDC. The agreement between data and simulation efficiencies is within 0.5%. The possible bias coming from this method is tested on simulated events. A difference of 0.5% is found with respect to the absolute efficiency.

The $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample is used to obtain a systematic uncertainty on the BDT selection. The same BDTs that were trained for the signal decays are applied to these events and a requirement is chosen on the outputs in order to have the same relative efficiency on $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ events as on the signal decays. In each mode, the difference between the efficiency in data and simulation for the chosen BDT requirement, between 0.4 and 2.5%, is considered as a systematic uncertainty. The same control sample is used to estimate a systematic uncertainty due to potential mismodeling of ISR effects that would affect the efficiency of the requirement on $\Delta E_{e\ell\ell}$. Since the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ ΔE distribution is much broader than the signal one and has negative values because of the missing neutrino, we cannot use the same requirement that was applied for the LFV decays.

Table 3. Summary of the systematic uncertainties affecting $\mathcal{B}(\tau^- \to e^\mp \ell^\pm \ell^-)$. The first five sources affect the signal efficiencies and are added in quadrature to get the total uncertainty on the signal efficiencies σ_ϵ^{tot}. The signal efficiency and the last two sources, the luminosity L and the tau pair cross section $\sigma_{\tau\tau}$, directly affect the branching fraction.

source	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
LID	1.0%	1.2%	1.3%	1.7%	1.8%
tracking	1.0%	1.0%	1.0%	1.0%	1.0%
trigger	0.7%	0.7%	0.7%	0.7%	0.7%
BDT	0.7%	2.5%	0.7%	1.5%	0.4%
ISR	3.4%	3.4%	3.4%	3.4%	3.4%
σ_ϵ^{tot}	3.9%	4.6%	3.9%	4.3%	4.1%
L	0.5%	0.5%	0.5%	0.5%	0.5%
$\sigma_{\tau\tau}$	0.3%	0.3%	0.3%	0.3%	0.3%

Instead, we compute the relative variation of the efficiency between data and simulation scanning different requirement values between -1.1 and -0.1 GeV. The average, which is found to be 3.4%, is taken as the systematic uncertainty.

Since the effect of FSR depends on the final state particles, we cannot rely on the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control channel. The approach adopted consists of artificially inflating the signal PDF tail at low mass by 5%. The value is chosen according to a similar procedure followed in Ref. [47]. The new PDF parameters are used as default in the fit to the data and no additional systematic uncertainty is assigned.

The systematic uncertainty in the integrated luminosity L is measured using samples of Bhabha, diphoton and dimuon events [48]. The relative uncertainty is 0.5%.

Finally, we also assign an uncertainty of 0.003 nb on the τ-pair production cross section, as evaluated in Ref. [18].

A summary of the systematic uncertainties is given in Table 3.

7 Result

The fits to the $M_{e\ell\ell}$ variable for the five $\tau^- \to e^\mp \ell^\pm \ell^-$ decays are shown in Fig. 5 and the results are summarized in Table 4. This table also lists the expected number of events obtained from the fit to the sidebands ($N_{\rm exp}$) and the number of observed events $N_{\rm obs}$.

As no signal is found, we compute 90% C.L. upper limits on the $\tau^- \to e^\mp \ell^\pm \ell^-$ branching fractions. We estimate the upper limit using the modified frequentist CL$_s$ [49, 50] method implemented in the RooStat framework. We generate 10,000 pseudo-experiments for 10 branching fraction values distributed uniformly in the range $[0, 3] \times 10^{-8}$. The generation uses the likelihood from Eq. 5.1, with the fitted value of $C_{\rm bg}$, and the observed number of events, $N_{\rm obs}$. Systematic uncertainties on $C_{\rm bg}$, $\epsilon_{\rm sig}$, L, and $\sigma_{\tau\tau}$ are accounted for by applying Gaussian constraints using the values described in Section 6. The expected limits are computed using input values obtained from fits to the data sidebands, assuming a number of observed events in the signal region equal to that extrapolated from the sidebands.

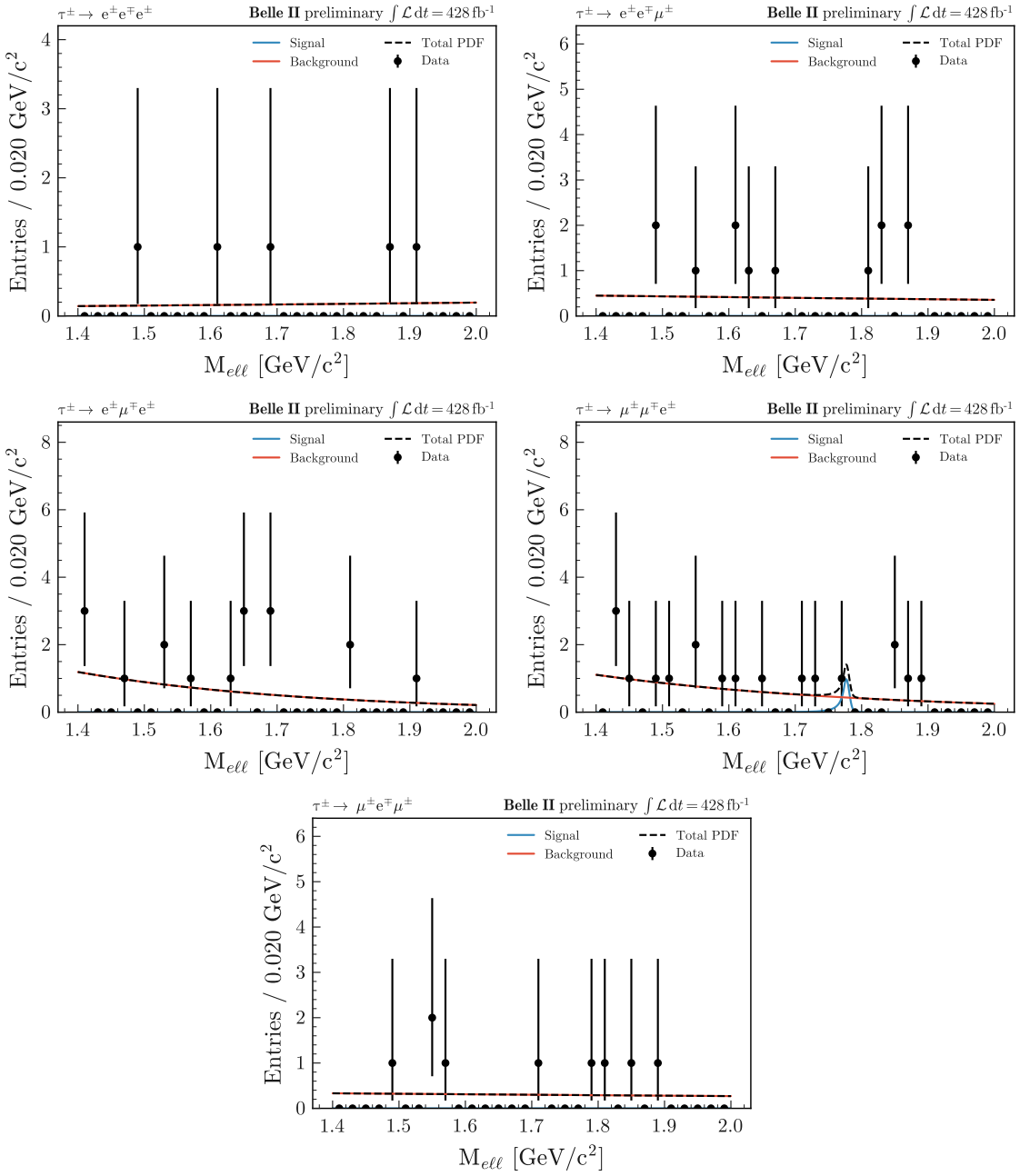

Figure 5. Distribution of data events and fits to the $M_{e\ell\ell}$ variable. The background and signal components of the PDF are shown in red and blue, respectively, while the black dashed line is the total fitted PDF.

Table 4. Number of expected and observed events, fitted value of $C_{\rm bg}$ and branching fractions, and expected and observed upper limits at 90% C.L. The subscript exp refers to expected values obtained from the fit to sidebands only.

	$N_{\rm exp}$	$N_{\rm obs}$	$C_{\rm bg}$	$\mathcal{B}$ (10^{-8})	$\mathcal{B}^{UL}_{\rm exp}$ (10^{-8})	$\mathcal{B}^{UL}_{\rm obs}$ (10^{-8})
$e^-e^+e^-$	$6.1^{+4.3}_{-2.9}$	5	$0.52^{+2.64}_{-2.60}$	0	2.7	2.5
$e^-e^+\mu^-$	$12.1^{+5.7}_{-4.3}$	12	$-0.40^{+1.67}_{-1.68}$	0	2.1	1.6
$e^-\mu^+e^-$	$10.5^{+5.3}_{-4.3}$	17	$-2.90^{+1.48}_{-1.54}$	0	1.7	1.6
$\mu^-\mu^+e^-$	$20.7^{+6.6}_{-5.5}$	18	$-2.50^{+1.45}_{-1.52}$	$0.48^{+0.90}_{-0.48}$	1.6	2.4
$\mu^-e^+\mu^-$	$7.5^{+4.5}_{-3.2}$	9	$-0.34^{+1.93}_{-1.94}$	0	1.4	1.3

The expected and observed limits are given in Table 4 and are more stringent than previous searches for all modes except the $\tau^- \to e^-\mu^+e^-$ final state.

8 Summary

We present a search for the LFV decays $\tau^- \to e^\mp\ell^\pm\ell^-$ using a 428 fb^{-1} data sample collected by the Belle II experiment. Using an inclusive-tagging reconstruction with a BDT-based selection, the efficiencies are higher by factors between 2 and 3.3 than those in the most recent Belle analysis [14] for an expected number of background events compatible with zero. No significant signal is found and we compute the upper limits at 90% C.L. The bounds obtained, between 1.3 and 2.5 $\times 10^{-8}$, are the most stringent to date for all modes except $\tau^- \to e^-\mu^+e^-$.

This work, based on data collected using the Belle II detector, which was built and commissioned prior to March 2019, 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 (FWF) Grants DOI: 10.55776/P34529, DOI: 10.55776/J4731, DOI: 10.55776/J4625, DOI: 10.55776/M3153, and DOI: 10.55776/PAT1836324, 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. 2024YFA1610503, and No. 2024YFA1610504 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, Regional funds of EU/MEYS: OPJAK FORTE CZ.02.01.01/00/22_008/0004632 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 No. ANR-21-CE31-0009 (France); BMFTR, 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. 2021R1-F1A-1064008, No. 2022R1-A2C-1003993, No. 2022R1-A2C-1092335, No. RS-2016-NR017151, No. RS-2018-NR031074, No. RS-2021-NR060129, No. RS-2023-00208693, No. RS-2024-00354342 and No. RS-2025-02219521, 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 (K25L2M2C3) 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 Researchers Supporting Project number (RSPD2025R873), King Saud University, Riyadh, Saudi Arabia; Slovenian Research Agency and Research Grants No. J1-50010 and No. P1-0135; Ikerbasque, Basque Foundation for Science, State Agency for Research of the Spanish Ministry of Science and Innovation through Grant No. PID2022-136510NB-C33, Spain, Agencia Estatal de Investigacion, Spain Grant No. RYC2020-029875-I and Generalitat Valenciana, Spain Grant No. CIDEGENT/2018/020; 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.

These acknowledgements are not to be interpreted as an endorsement of any statement made by any of our institutes, funding agencies, governments, or their representatives.

We thank the 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, and the University of Victoria.

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 [2104.04494].

[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 [1807.06050].

[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 [1912.09862].

[4] Heavy Flavor Averaging Group collaboration, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, *Eur. Phys. J. C* **81** (2021) 226 [1909.12524].

[5] A. Abada, J. Kriewald and A.M. Teixeira, *On the role of leptonic CPV phases in cLFV observables*, *Eur. Phys. J. C* **81** (2021) 1016 [2107.06313].

[6] M. Raidal et al., *Flavour physics of leptons and dipole moments*, *Eur. Phys. J. C* **57** (2008) 13 [0801.1826].

[7] A.M. Teixeira, *Theoretical aspects of charged lepton flavour violation*, *J. Phys. Conf. Ser.* **888** (2017) 012029 [1612.05561].

[8] F. Feruglio, P. Paradisi and A. Pattori, *On the importance of electroweak corrections for B anomalies*, *JHEP* **09** (2017) 061 [1705.00929].

[9] B.M. Dassinger, T. Feldmann, T. Mannel and S. Turczyk, *Model-independent analysis of lepton flavour violating tau decays*, *JHEP* **10** (2007) 039 [0707.0988].

[10] P. Paradisi, *Higgs-mediated $\tau \to \mu$ and $\tau \to e$ transitions in II Higgs doublet model and supersymmetry*, *JHEP* **02** (2006) 050 [hep-ph/0508054].

[11] Belle II collaboration, *Search for lepton-flavor-violating $\tau^- \to \mu^-\mu^+\mu^-$ decays at Belle II*, *JHEP* **09** (2024) 062 [2405.07386].

[12] M. Ardu, S. Davidson and S. Lavignac, *Constraining new physics models from $\mu \to e$ observables in bottom-up EFT*, *Eur. Phys. J. C* **84** (2024) 458 [2401.06214].

[13] K. Cheung, A. Soffer, Z.S. Wang and Y.-H. Wu, *Probing charged lepton flavor violation with axion-like particles at Belle II*, *JHEP* **11** (2021) 218 [2108.11094].

[14] Belle collaboration, *Search for Lepton Flavor Violating Tau Decays into Three Leptons with 719 Million Produced $\tau^+\tau^-$ Pairs*, *Phys. Lett. B* **687** (2010) 139 [1001.3221].

[15] BaBar collaboration, *Limits on tau lepton-flavor violating decays in three charged leptons*, *Phys. Rev. D* **81** (2010) 111101 [1002.4550].

[16] Belle II collaboration, *Belle II technical design report*, 1011.0352.

[17] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*, *Nucl. Instrum. Meth.A* **907** (2018) 188 [1809.01958].

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

[19] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.

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

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

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

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

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

[25] G. Balossini, C.M. Carloni Calame, G. Montagna, O. Nicrosini and F. Piccinini, *Matching perturbative and parton shower corrections to Bhabha process at flavour factories*, *Nucl. Phys.B* **758** (2006) 227.

[26] G. Balossini, C. Bignamini, C.M.C. Calame, G. Montagna, O. Nicrosini and F. Piccinini, *Photon pair production at flavour factories with per mille accuracy*, *Phys. Lett.B* **663** (2008) 209.

[27] C.M. Carloni Calame, G. Montagna, O. Nicrosini and F. Piccinini, *The BABAYAGA event generator*, *Nucl. Phys. B Proc. Suppl.* **131** (2004) 48.

[28] C.M. Carloni Calame, *An improved parton shower algorithm in QED*, *Phys. Lett.B* **520** (2001) 16.

[29] C.M. Carloni Calame, C. Lunardini, G. Montagna, O. Nicrosini and F. Piccinini, *Large angle Bhabha scattering and luminosity at flavor factories*, *Nucl. Phys.B* **584** (2000) 459.

[30] F. Berends, P. Daverveldt and R. Kleiss, *Radiative corrections to the process* $e^+e^- \to e^+e^-\mu^+\mu^-$, *Nucl. Phys.B* **253** (1985) 421.

[31] F. Berends, P. Daverveldt and R. Kleiss, *Complete lowest-order calculations for four-lepton final states in electron-positron collisions*, *Nucl. Phys.B* **253** (1985) 441.

[32] F. Berends, P. Daverveldt and R. Kleiss, *Monte Carlo simulation of two-photon processes: II: Complete lowest order calculations for four-lepton production processes in electron-positron collisions*, *Comp. Phys. Commun.* **40** (1986) 285.

[33] S. Uehara, *TREPS: A Monte-Carlo event generator for two-photon processes at* e^+e^- *colliders using an equivalent photon approximation*, 1310.0157.

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

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

[36] GEANT4 collaboration, *GEANT4: A simulation toolkit*, *Nucl.Instrum.Meth.A* **506** (2003) 250.

[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.

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

[39] Belle II analysis software Group collaboration, *Global decay chain vertex fitting at Belle II*, *Nucl. Instrum. Meth. A* **976** (2020) 164269 [1901.11198].

[40] 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.

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

[42] G.C. Fox and S. Wolfram, *Observables for the analysis of event shapes in e^+e^- annihilation and other processes*, *Phys. Rev. Lett.* **41** (1978) 1581.

[43] CLEO collaboration, *Search for exclusive charmless hadronic B decays*, *Phys. Rev. D* **53** (1996) 1039 [hep-ex/9508004].

[44] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, Aug., 2016, DOI.

[45] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, *Optuna: A Next-generation Hyperparameter Optimization Framework*, 1907.10902.

[46] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Υ' and Υ resonances*, Ph.D. thesis, Cracow, INP, 1986.

[47] Belle II collaboration, *Test of light-lepton universality in τ decays with the Belle II experiment*, *JHEP* **08** (2024) 205.

[48] Belle II collaboration, *Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*, 2407.00965.

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

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

Additional Material

The ($M_{e\ell\ell}$,$\Delta E_{e\ell\ell}$) plane for the signal, the lepton pair invariant mass and visible energy of the event, and the $M_{e\ell\ell}$ and $\Delta E_{e\ell\ell}$ distributions, for the modes omitted throughout the paper, are shown in Fig. 6, Fig. 7, and Fig. 8, respectively. The signal efficiency as a function of the two-dimensional plane defined by the mass squared of the opposite-charge lepton pairs is provided in Fig. 9.

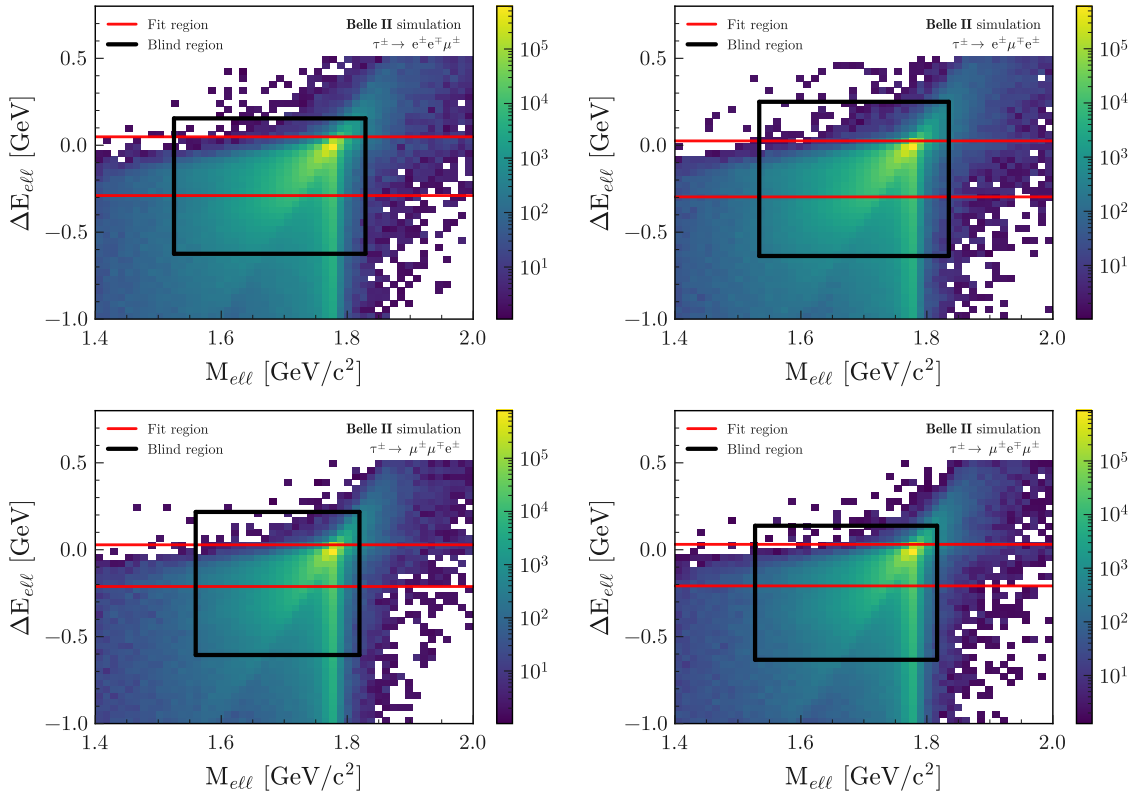

Figure 6. Signal distribution in the $(M_{e\ell\ell}, \Delta E_{e\ell\ell})$ plane for MC simulated $\tau^- \to e^\mp \ell^\pm \ell^-$ decays other than $\tau^- \to e^- \mu^+ e^-$. The blind region is the region inside the black lines, while the signal fit region is the one between the red lines.

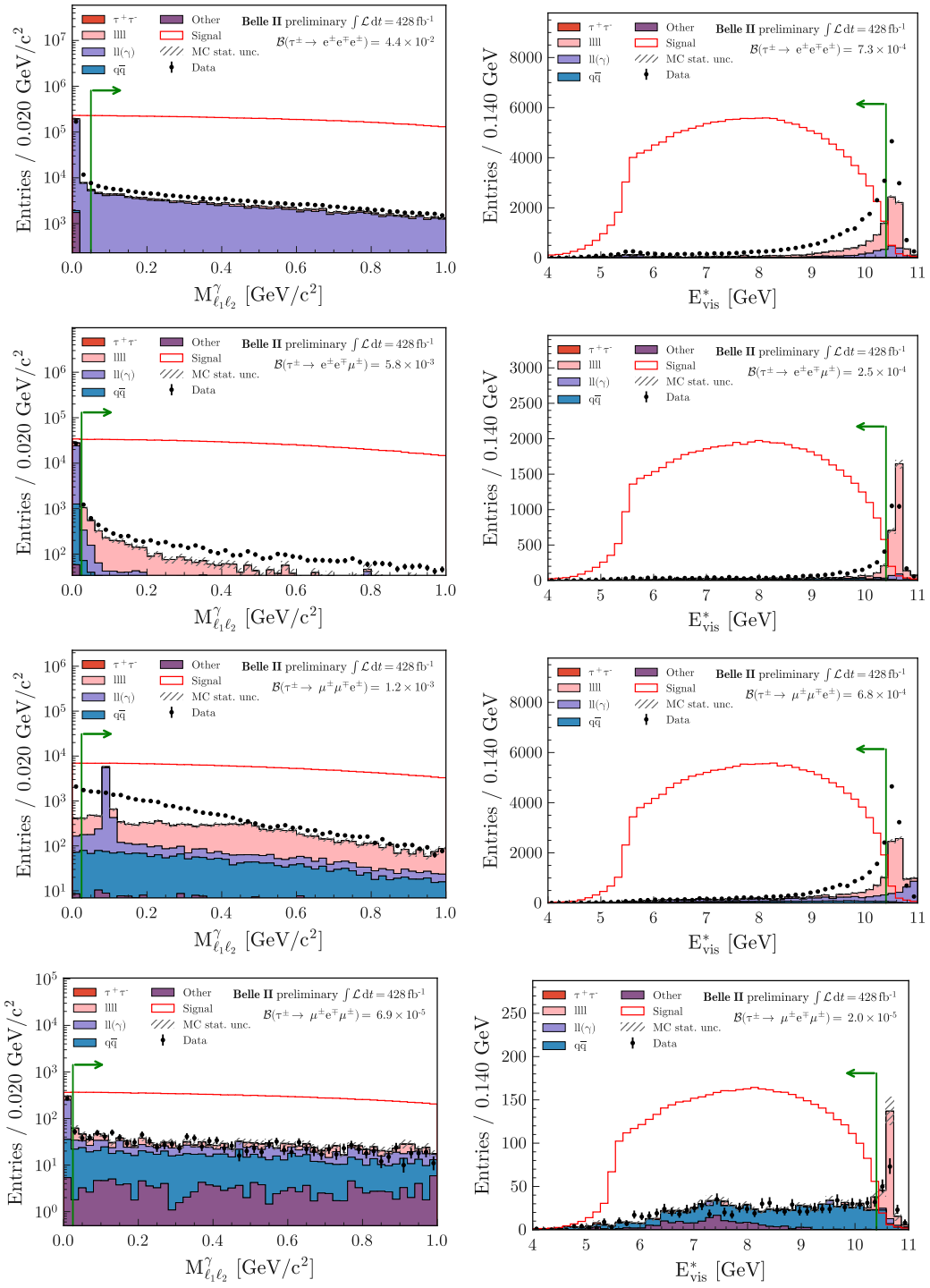

Figure 7. Distribution of the invariant mass $M^{\gamma}_{\ell_1\ell_2}$ (left) and the visible energy in the c.m. (right) for the $\tau^- \to e^-e^+e^-, e^-e^+\mu^-, \mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$ modes. The green arrows correspond to the applied selections. The various simulated background processes are shown as a stack of color-filled histograms, with statistical uncertainties displayed as hatched areas. The signal, not blinded, is shown as a red histogram with branching fraction values given on the plots.

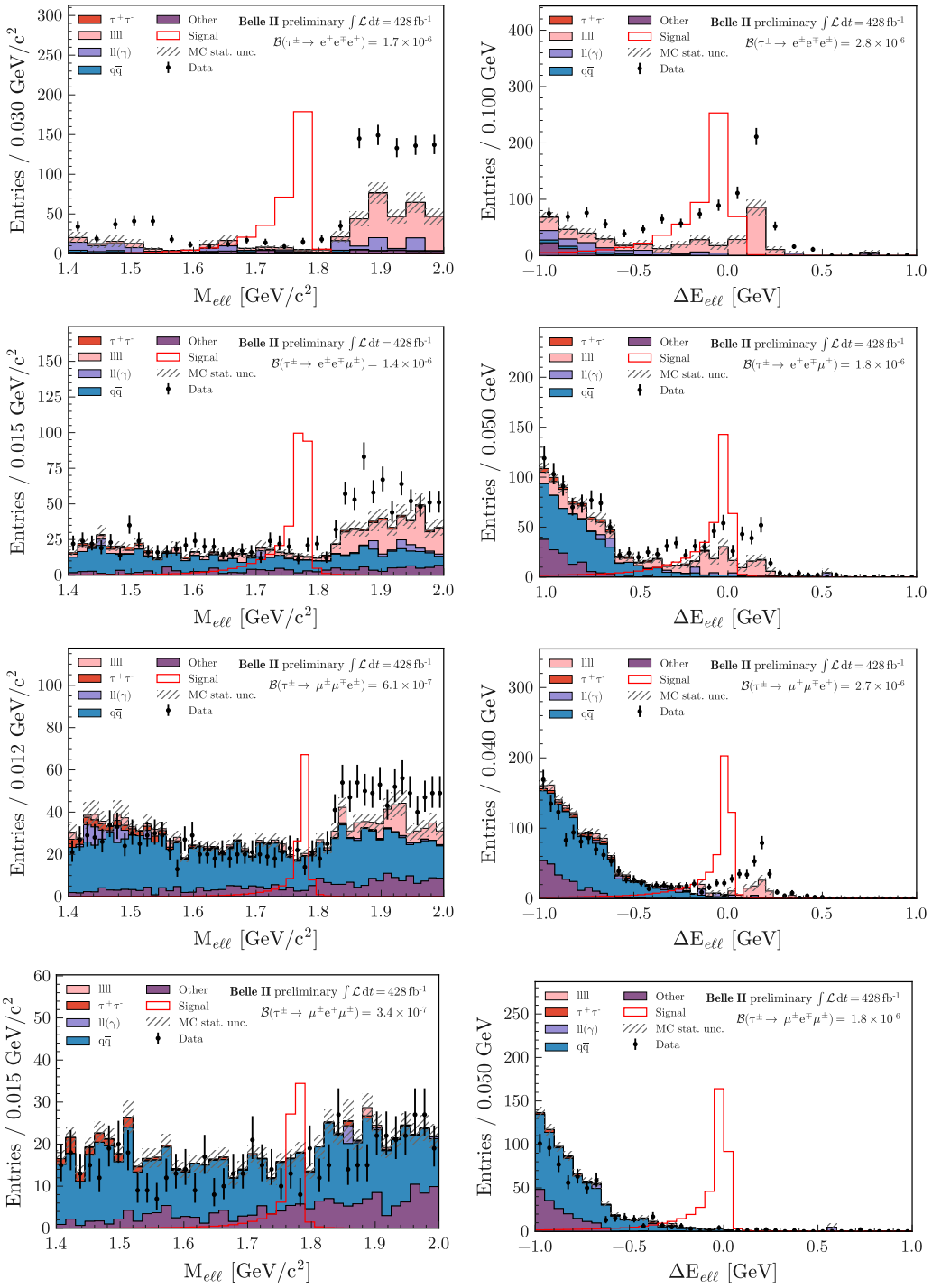

Figure 8. $M_{e\ell\ell}$ (left) and $\Delta E_{e\ell\ell}$ (right) distribution for the $\tau^- \to e^-e^+e^-, e^-e^+\mu^-, \mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$ modes for data and simulation outside the blind region after the preselection. The various simulated background processes are shown as a stack of color-filled histograms, with statistical uncertainties displayed as hatched areas. The signal, not blinded, is shown as a red histogram with branching fraction values given on the plots.

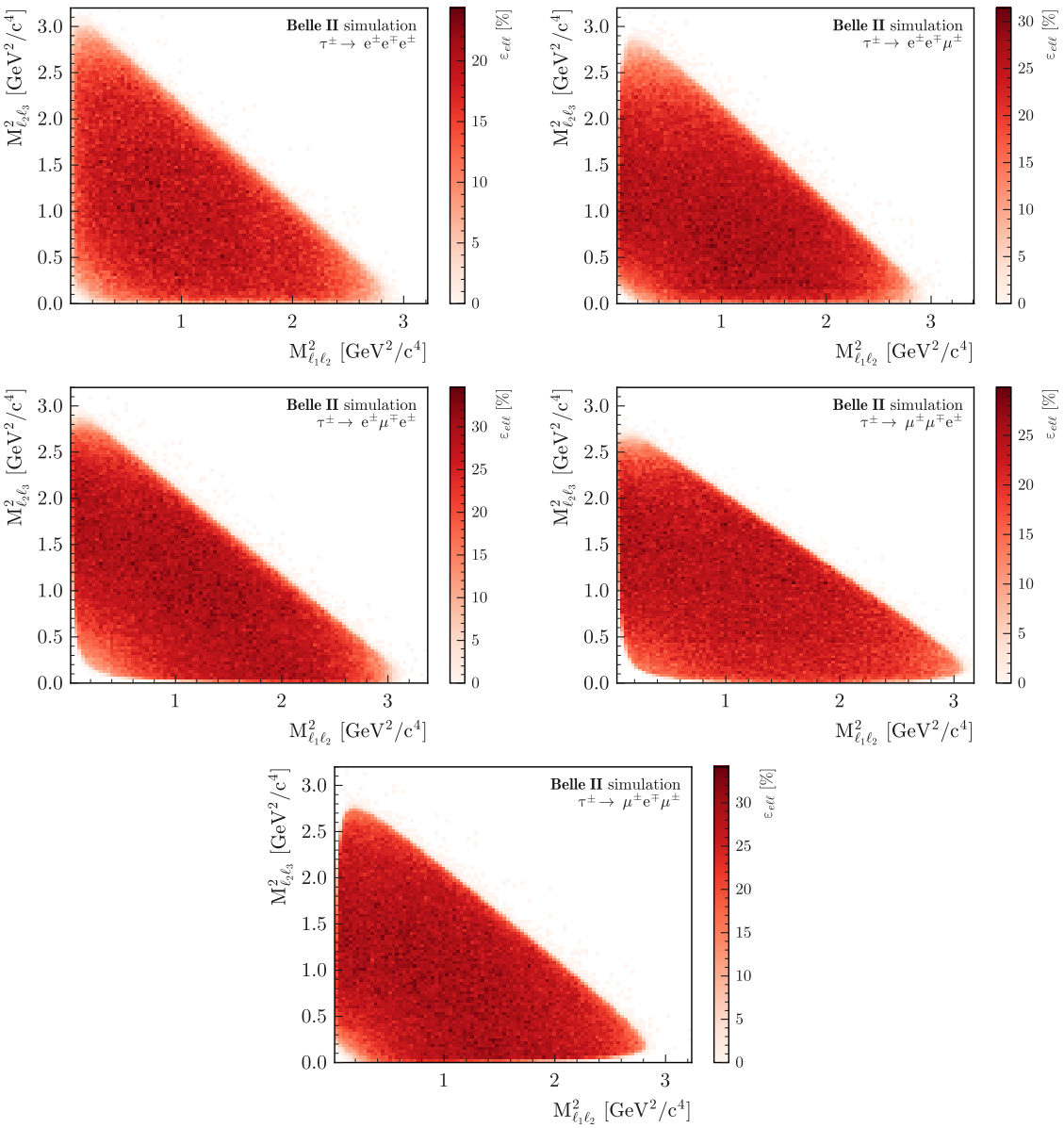

Figure 9. Signal efficiency as a function of the two-dimensional plane defined by the mass squared of the opposite-charge lepton pairs. The indices (1,2,3) follow the ordering of the decay chains ($\tau \to \ell_1\ell_2\ell_3$).