

## Study of the muon decay-in-flight in the $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ decay to measure the Michel parameter $\xi'$

D. Bodrov<sup>✉</sup>, P. Pakhlov<sup>✉</sup>, I. Adachi<sup>✉</sup>, H. Aihara<sup>✉</sup>, S. Al Said<sup>✉</sup>, D. M. Asner<sup>✉</sup>, H. Atmacan<sup>✉</sup>, T. Aushev<sup>✉</sup>, R. Ayad<sup>✉</sup>, V. Babu<sup>✉</sup>, Sw. Banerjee<sup>✉</sup>, P. Behera<sup>✉</sup>, K. Belous<sup>✉</sup>, J. Bennett<sup>✉</sup>, M. Bessner<sup>✉</sup>, B. Bhuyan<sup>✉</sup>, T. Bilka<sup>✉</sup>, D. Biswas<sup>✉</sup>, A. Bobrov<sup>✉</sup>, A. Bondar<sup>✉</sup>, J. Borah<sup>✉</sup>, A. Bozek<sup>✉</sup>, M. Bračko<sup>✉</sup>, P. Branchini<sup>✉</sup>, T. E. Browder<sup>✉</sup>, A. Budano<sup>✉</sup>, M. Campajola<sup>✉</sup>, D. Červenkov<sup>✉</sup>, M.-C. Chang<sup>✉</sup>, B. G. Cheon<sup>✉</sup>, K. Chilikin<sup>✉</sup>, K. Cho<sup>✉</sup>, S.-J. Cho<sup>✉</sup>, S.-K. Choi<sup>✉</sup>, Y. Choi<sup>✉</sup>, S. Choudhury<sup>✉</sup>, D. Cinabro<sup>✉</sup>, S. Das<sup>✉</sup>, G. De Nardo<sup>✉</sup>, G. De Pietro<sup>✉</sup>, R. Dhamija<sup>✉</sup>, F. Di Capua<sup>✉</sup>, J. Dingfelder<sup>✉</sup>, Z. Doležal<sup>✉</sup>, T. V. Dong<sup>✉</sup>, D. Epifanov<sup>✉</sup>, T. Ferber<sup>✉</sup>, D. Ferlewicz<sup>✉</sup>, B. G. Fulsom<sup>✉</sup>, V. Gaur<sup>✉</sup>, A. Garmash<sup>✉</sup>, A. Giri<sup>✉</sup>, P. Goldenzweig<sup>✉</sup>, E. Graziani<sup>✉</sup>, D. Greenwald<sup>✉</sup>, T. Gu<sup>✉</sup>, Y. Guan<sup>✉</sup>, K. Gudkova<sup>✉</sup>, C. Hadjivasiliou<sup>✉</sup>, S. Halder<sup>✉</sup>, K. Hayasaka<sup>✉</sup>, H. Hayashii<sup>✉</sup>, M. T. Hedges<sup>✉</sup>, D. Herrmann<sup>✉</sup>, W.-S. Hou<sup>✉</sup>, C.-L. Hsu<sup>✉</sup>, T. Iijima<sup>✉</sup>, K. Inami<sup>✉</sup>, N. Ipsita<sup>✉</sup>, A. Ishikawa<sup>✉</sup>, R. Itoh<sup>✉</sup>, M. Iwasaki<sup>✉</sup>, W. W. Jacobs<sup>✉</sup>, E.-J. Jang<sup>✉</sup>, Q. P. Ji<sup>✉</sup>, S. Jia<sup>✉</sup>, Y. Jin<sup>✉</sup>, K. K. Joo<sup>✉</sup>, D. Kalita<sup>✉</sup>, A. B. Kaliyar<sup>✉</sup>, K. H. Kang<sup>✉</sup>, T. Kawasaki<sup>✉</sup>, C. Kiesling<sup>✉</sup>, C. H. Kim<sup>✉</sup>, D. Y. Kim<sup>✉</sup>, K.-H. Kim<sup>✉</sup>, Y.-K. Kim<sup>✉</sup>, H. Kindo<sup>✉</sup>, K. Kinoshita<sup>✉</sup>, P. Kodyš<sup>✉</sup>, A. Korobov<sup>✉</sup>, S. Korpar<sup>✉</sup>, P. Križan<sup>✉</sup>, P. Krokovny<sup>✉</sup>, T. Kuhr<sup>✉</sup>, M. Kumar<sup>✉</sup>, R. Kumar<sup>✉</sup>, K. Kumara<sup>✉</sup>, Y.-J. Kwon<sup>✉</sup>, J. S. Lange<sup>✉</sup>, S. C. Lee<sup>✉</sup>, J. Li<sup>✉</sup>, L. K. Li<sup>✉</sup>, Y. Li<sup>✉</sup>, J. Libby<sup>✉</sup>, K. Lieret<sup>✉</sup>, Y.-R. Lin<sup>✉</sup>, D. Liventsev<sup>✉</sup>, Y. Ma<sup>✉</sup>, M. Masuda<sup>✉</sup>, T. Matsuda<sup>✉</sup>, S. K. Maurya<sup>✉</sup>, F. Meier<sup>✉</sup>, M. Merola<sup>✉</sup>, F. Metzner<sup>✉</sup>, K. Miyabayashi<sup>✉</sup>, R. Mizuk<sup>✉</sup>, G. B. Mohanty<sup>✉</sup>, R. Mussa<sup>✉</sup>, I. Nakamura<sup>✉</sup>, M. Nakao<sup>✉</sup>, D. Narwal<sup>✉</sup>, Z. Natkaniec<sup>✉</sup>, A. Natochii<sup>✉</sup>, L. Nayak<sup>✉</sup>, M. Nayak<sup>✉</sup>, N. K. Nisar<sup>✉</sup>, S. Nishida<sup>✉</sup>, S. Ogawa<sup>✉</sup>, H. Ono<sup>✉</sup>, P. Oskin<sup>✉</sup>, G. Pakhlova<sup>✉</sup>, S. Pardi<sup>✉</sup>, H. Park<sup>✉</sup>, J. Park<sup>✉</sup>, S.-H. Park<sup>✉</sup>, S. Patra<sup>✉</sup>, S. Paul<sup>✉</sup>, T. K. Pedlar<sup>✉</sup>, R. Pestotnik<sup>✉</sup>, L. E. Piilonen<sup>✉</sup>, T. Podobnik<sup>✉</sup>, E. Prencipe<sup>✉</sup>, M. T. Prim<sup>✉</sup>, A. Rabusov<sup>✉</sup>, G. Russo<sup>✉</sup>, S. Sandilya<sup>✉</sup>, A. Sangal<sup>✉</sup>, L. Santelj<sup>✉</sup>, V. Savinov<sup>✉</sup>, G. Schnell<sup>✉</sup>, C. Schwanda<sup>✉</sup>, Y. Seino<sup>✉</sup>, K. Senyo<sup>✉</sup>, W. Shan<sup>✉</sup>, M. Shapkin<sup>✉</sup>, C. Sharma<sup>✉</sup>, J.-G. Shiu<sup>✉</sup>, J. B. Singh<sup>✉</sup>, A. Sokolov<sup>✉</sup>, E. Solovieva<sup>✉</sup>, M. Starič<sup>✉</sup>, Z. S. Stottler<sup>✉</sup>, M. Sumihama<sup>✉</sup>, M. Takizawa<sup>✉</sup>, U. Tamponi<sup>✉</sup>, K. Tanida<sup>✉</sup>, F. Tenchini<sup>✉</sup>, R. Tiwary<sup>✉</sup>, K. Trabelsi<sup>✉</sup>, M. Uchida<sup>✉</sup>, T. Uglov<sup>✉</sup>, Y. Unno<sup>✉</sup>, K. Uno<sup>✉</sup>, S. Uno<sup>✉</sup>, S. E. Vahsen<sup>✉</sup>, G. Varner<sup>✉</sup>, A. Vinokurova<sup>✉</sup>, A. Vossen<sup>✉</sup>, D. Wang<sup>✉</sup>, E. Wang<sup>✉</sup>, M.-Z. Wang<sup>✉</sup>, S. Watanuki<sup>✉</sup>, X. Xu<sup>✉</sup>, B. D. Yabsley<sup>✉</sup>, W. Yan<sup>✉</sup>, S. B. Yang<sup>✉</sup>, J. Yelton<sup>✉</sup>, J. H. Yin<sup>✉</sup>, C. Z. Yuan<sup>✉</sup>, Y. Yusa<sup>✉</sup>, Z. P. Zhang<sup>✉</sup>, V. Zhilich<sup>✉</sup>, and V. Zhukova<sup>✉</sup>

(Belle Collaboration)



(Received 21 March 2023; accepted 7 June 2023; published 11 July 2023)

We present the first measurement of the Michel parameter  $\xi'$  in the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay using the full data sample of  $988 \text{ fb}^{-1}$  collected by the Belle detector operating at the KEKB asymmetric energy  $e^+e^-$  collider. The method is based on the reconstruction of the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay-in-flight in the Belle central drift chamber and relies on the correlation between muon spin and its daughter electron momentum. We study the main sources of the background that can imitate the signal decay, such as kaon and pion decays-in-flight and charged particle scattering on the detector material. Highly efficient methods of their suppression are developed and applied to select 165 signal-candidate events. We obtain  $\xi' = 0.22 \pm 0.94 \pm 0.42$  where the first uncertainty is statistical, and the second one is systematic. The result is in agreement with the Standard Model prediction of  $\xi' = 1$ .

DOI: [10.1103/PhysRevD.108.012003](https://doi.org/10.1103/PhysRevD.108.012003)

### I. INTRODUCTION

In the Standard Model (SM), the  $\tau$  lepton decay proceeds through a weak charged current, whose amplitude can be approximated with high accuracy by the four-fermion

interaction with the  $V - A$  Lorentz structure. A deviation from this structure would indicate physics beyond the SM, which can be caused by an anomalous coupling of the  $W$  boson with the  $\tau$  lepton, a new gauge or charged Higgs bosons contribution, etc [1–4]. The presence of massive neutrinos can also modify experimental observables, leading to a deviation from the SM prediction [5].

The most general form of the Lorentz invariant, local, derivative-free, lepton-number-conserving four-fermion interaction Hamiltonian [6] leads to the following matrix

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). 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<sup>3</sup>.

element of the  $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$  decay ( $\ell = e$  or  $\mu$ ) written in the form of helicity projections [7–9]:

$$M = \frac{4G_F}{\sqrt{2}} \sum_{\substack{\lambda=S,V,T \\ \varepsilon,\omega=L,R}} g_{\varepsilon\omega}^\lambda \langle \bar{\ell}_\varepsilon | \Gamma^\lambda | (\nu_\ell)_\alpha \rangle \langle (\bar{\nu}_\tau)_\beta | \Gamma_\lambda | \tau_\omega \rangle, \quad (1)$$

where

$$\Gamma^S = 1, \quad \Gamma^V = \gamma^\mu, \quad \Gamma^T = \frac{i}{2\sqrt{2}} (\gamma^\mu \gamma^\nu - \gamma^\nu \gamma^\mu); \quad (2)$$

$S$ ,  $V$ , and  $T$  denote scalar, vector, and tensor interaction, respectively;  $\varepsilon, \omega = L, R$  means left- and right-handed leptons, respectively. Each set of indices  $\lambda$ ,  $\varepsilon$ , and  $\omega$  uniquely determines the neutrino handedness  $\alpha$  and  $\beta$ . The total strength of the weak interaction in Eq. (1) is given by  $G_F$ , while  $g_{\varepsilon\omega}^\lambda$  are normalized as

$$\sum_{\varepsilon,\omega=L,R} \left( \frac{1}{4} |g_{\varepsilon\omega}^S|^2 + |g_{\varepsilon\omega}^V|^2 + 3 |g_{\varepsilon\omega}^T|^2 \right) \equiv 1. \quad (3)$$

It is convenient to express the observables in the lepton decay in terms of the Michel parameters (MPs), which are bilinear combinations of the coupling constants  $g_{\varepsilon\omega}^\lambda$ . The MPs are described in detail elsewhere [10].

At present, in  $\tau$  decays four Michel parameters,  $\rho$ ,  $\eta$ ,  $\xi$ , and  $\xi\delta$ , have been measured with accuracies at the level of a few percent [11], and the obtained values  $\rho = 0.745 \pm 0.008$ ,  $\eta = 0.013 \pm 0.020$ ,  $\xi = 0.985 \pm 0.030$ , and  $\xi\delta = 0.746 \pm 0.021$  are in agreement with the SM prediction of  $\rho = 3/4$ ,  $\eta = 0$ ,  $\xi = 1$ , and  $\xi\delta = 3/4$ . These parameters describe the differential decay width, integrated over the neutrinos momenta and summed over the daughter lepton spin. Measurements of the remaining Michel parameters,  $\xi'$ ,  $\xi''$ ,  $\eta''$ ,  $\alpha'/A$ , and  $\beta'/A$ , requires knowledge of the daughter lepton polarization, and no measurements of them have yet been performed. The only exception is two parameters,  $\xi\kappa$  and  $\bar{\eta}$ , obtained in the radiative leptonic  $\tau$  decays by the Belle collaboration [12]. These parameters are related to the Michel parameters  $\xi'$  and  $\xi''$  through linear combinations with the parameters  $\xi$ ,  $\xi\delta$ , and  $\rho$ :  $\xi' = -\xi - 4\xi\kappa + 8/3\xi\delta$  and  $\xi'' = 16/3\rho - 4\bar{\eta} - 3$ . Substituting parameters  $\xi$  and  $\xi\delta$  with their SM values and  $\xi\kappa$  with the value for the radiative muonic  $\tau$  decay from Ref. [12], one obtains  $\xi' = -2.2 \pm 2.4$ . However, this measurement still suffers from very large uncertainties: physically allowed  $\xi'$  values range from  $-1$  to  $1$  (in SM, it is equal to  $1$ ).

In this paper, we present the first direct measurement of the Michel parameter  $\xi'$  in the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay. This parameter determines the longitudinal polarization of muons  $P_L$  and enters the term of the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$

differential decay width that does not depend on the  $\tau$  lepton polarization. The parameter  $\xi'$  is written in terms of the coupling constants  $g_{\varepsilon\omega}^\lambda$  as

$$\xi' = 1 - 2 \sum_{\omega=L,R} \left( \frac{1}{4} |g_{R\omega}^S|^2 + |g_{R\omega}^V|^2 + 3 |g_{R\omega}^T|^2 \right). \quad (4)$$

Thus, a measurement of  $\xi'$  provides the necessary information required to calculate the probability of an unpolarized  $\tau$  lepton to decay to a right-handed muon:  $Q_R^\mu = (1 - \xi')/2$ . This paper is accompanied by a Letter in Physical Review Letters [13].

## II. METHOD

### A. Differential decay width

The method of the muon polarization measurement is based on the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay reconstruction since the electron momentum in the muon rest frame correlates with the muon spin. Initially, the idea was suggested in Ref. [14], where it was proposed to use stopped muons. Recently, it was proposed to use the muon decay-in-flight (kink) in the tracking system of the detector to measure  $\xi'$  in the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay in a future experiment at the Super Charm-Tau Factory [15,16]. In this paper, we rely on the adaptation of this method for the application at the  $B$ -factories from Ref. [17].

The differential decay width of the cascade decay  $\tau^- \rightarrow \mu^- (\rightarrow e^- \bar{\nu}_e \nu_\mu) \bar{\nu}_\mu \nu_\tau$  obtained in Ref. [17] follows

$$\begin{aligned} \frac{d^3\Gamma}{dx dy d\cos\theta_e} &= \mathcal{B}_{\mu \rightarrow e\nu} \frac{12\Gamma_{\tau \rightarrow \mu\nu\nu}}{1 - 3x_0^2} y^2 \sqrt{x^2 - x_0^2} \\ &\times [(3 - 2y)F_{IS}(x) + (2y - 1)F_{IP}(x) \cos\theta_e]. \end{aligned} \quad (5)$$

Here  $\Gamma_{\tau \rightarrow \mu\nu\nu}$  is the partial width of the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay;  $\mathcal{B}_{\mu \rightarrow e\nu}$  is the branching fraction of the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay;  $x = E_\mu/W_{\mu\tau}$  is the reduced muon energy in the  $\tau$  rest frame [ $W_{\mu\tau} = (m_\mu^2 + m_\tau^2)/(2m_\tau)$  is the maximum muon energy];  $x_0 = m_\mu/W_{\mu\tau}$  is the reduced muon mass; and  $y = 2E_e/m_\mu$  is the ratio of the electron energy to its maximum value in the muon rest frame. Functions  $F_{IS}(x)$  and  $F_{IP}(x)$  are expressed in terms of Michel parameters and depend only on  $x$ :

$$\begin{aligned} F_{IS}(x) &= x(1 - x) + \frac{2}{9}\rho(4x^2 - 3x - x_0^2) + \eta x_0(1 - x), \\ F_{IP}(x) &= \frac{1}{54} \sqrt{x^2 - x_0^2} \left[ -9\xi' \left( 2x - 3 + \frac{x_0^2}{2} \right) \right. \\ &\quad \left. + 4\xi \left( \delta - \frac{3}{4} \right) \left( 4x - 3 - \frac{x_0^2}{2} \right) \right]. \end{aligned} \quad (6)$$

<sup>1</sup>Charge conjugation is implied throughout the paper unless otherwise indicated.

Since  $\rho$ ,  $\eta$ ,  $\xi$ , and  $\xi\delta$  are measured very precisely, we fix their values to the SM expectations<sup>2</sup>; thus, only  $\xi'$  in Eqs. (5) and (6) is to be determined.

The variable  $\theta_e$  is the angle between  $\vec{n}_\mu$  and  $\vec{n}'_e$ , where  $\vec{n}_\mu$  is the direction opposite to the  $\tau$  lepton momentum in the muon rest frame at the muon production vertex, and  $\vec{n}'_e$  is the direction of the electron in the muon rest frame at the muon decay vertex. The former vector is represented in the conventional coordinate system introduced in Ref. [17] as  $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$  while  $\vec{n}'_e$  is represented in the coordinate system obtained from the initial one by rotation through an angle  $\phi$  (the muon momentum angle of rotation in the magnetic field of the Belle detector before the decay). The procedure of the coordinate system rotation and  $\theta_e$  calculation is explained in detail in Ref. [17].

The angle  $\theta_e$  has a simple physical meaning when the muon decays immediately at the production vertex: this is the angle between mother and daughter charged leptons in the muon rest frame. Once the muon propagates in the magnetic field of the detector, its momentum in the laboratory frame and spin in the muon rest frame are rotated through the same angle  $\phi$  (assuming  $g_\mu - 2 \approx 0$  without loss of precision [11]). The rotation of the coordinate system in each event is designed to compensate for the effect of the magnetic field, bringing the event to the case of the instantaneous muon decay.

## B. $\tau$ lepton momentum reconstruction

For the  $\xi'$  measurement, a knowledge of the  $\tau$  lepton momentum is essential. While the  $\tau$  energy is known from the beam energy (up to the initial-state radiation), it is not feasible to reconstruct the true direction of the  $\tau$  momentum due to neutrinos in the final state. However, it is possible to find the region where the  $\tau$  lepton momentum is directed using the second (tagging)  $\tau$  lepton in the event [18]. The method is based on the kinematics of the  $\tau^+\tau^-$  pair production and decay in the center-of-mass (c.m.) frame.

For hadronic modes of the tagging  $\tau$ , the angle between  $\tau$  lepton and daughter hadron momenta is the following:

$$\cos\psi = \frac{2E_\tau E_h - m_\tau^2 - m_h^2}{2p_\tau p_h}. \quad (7)$$

Here  $E_\tau$  and  $p_\tau$  are the  $\tau$  lepton energy and momentum magnitude in the c.m. frame;  $E_h$ ,  $p_h$ , and  $m_h$  are the hadron system energy, momentum magnitude, and invariant mass, respectively. We use all one-prong and three-prong modes of the tagging  $\tau$ , including  $\tau^+ \rightarrow \ell^+ \nu_\ell \bar{\nu}_\tau$  ( $\ell = e$  or  $\mu$ ); however, we treat the leptonic mode as a hadronic one for simplification.

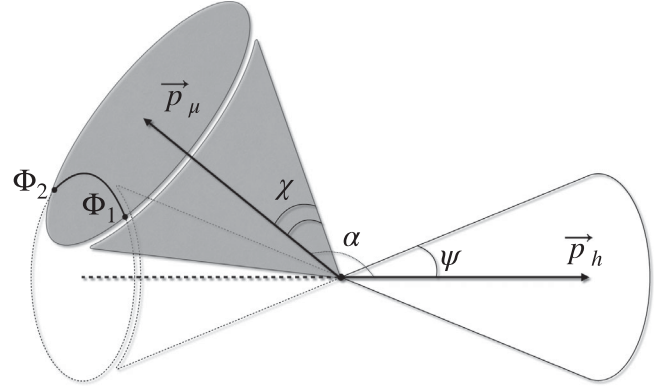


FIG. 1. Geometric interpretation of the  $\tau$  lepton momentum reconstruction.

For the signal  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay, the angle between  $\tau$  lepton and daughter muon momenta is restricted to

$$\frac{2E_\tau E_\mu - m_\tau^2 - m_\mu^2}{2p_\tau p_\mu} \leq \cos\chi \leq \frac{E_\tau E_\mu - m_\tau m_\mu}{p_\tau p_\mu}. \quad (8)$$

Here,  $E_\mu$  and  $p_\mu$  are the muon energy and momentum in the c.m. frame.

Thus, the true  $\tau^+\tau^-$  pair production axis lies on the generatrix of a cone with an apex angle of  $2\psi$  and inside a cone with an apex angle of  $2\chi$  (see Fig. 1). This restricts the region of possible  $\tau$  lepton directions to an arc  $(\Phi_1, \Phi_2)$ , where  $\Phi_1$  and  $\Phi_2$  are defined as follows

$$\begin{aligned} \Phi_1 &= \pi + \arcsin \frac{\cos\psi \cos\alpha + \cos\chi}{\sin\psi \sin\alpha}, \\ \Phi_2 &= 2\pi - \arcsin \frac{\cos\psi \cos\alpha + \cos\chi}{\sin\psi \sin\alpha}. \end{aligned} \quad (9)$$

Here  $\alpha$  is an angle between  $\vec{p}_\mu$  and  $\vec{p}_h$ . For simplicity, we use the average value of  $\Phi = (\Phi_1 + \Phi_2)/2 = -\pi/2$  instead of averaging Eq. (5) over  $(\Phi_1, \Phi_2)$ . This approximation has a negligible impact on the  $\xi'$  measurement: the increase of the statistical uncertainty is less than 0.01.

## C. Decay vertex reconstruction

Track reconstruction at Belle is optimized for long-lived particles that originate close to the interaction point of the beams (IP) and does not contain dedicated algorithms for identifying charged particle decays in flight. However, tracks that do not point to IP are also reconstructed with a considerable efficiency. In our case, the muon track is reconstructed first, and it may absorb some hits produced by the daughter electron, thereby smearing the muon momentum resolution. The remaining hits are used to reconstruct the electron track.

We define the decay vertex as a point of the closest approach of muon and electron tracks in the region of their endpoint and starting point, respectively.

<sup>2</sup>It is checked that this assumption has a negligible effect.

### III. THE DATA SAMPLE AND THE BELLE DETECTOR

This analysis is based on a data sample taken at or near the  $\Upsilon(1S)$ ,  $\Upsilon(2S)$ ,  $\Upsilon(3S)$ ,  $\Upsilon(4S)$ , and  $\Upsilon(5S)$  resonances with an integrated luminosity of  $988 \text{ fb}^{-1}$  corresponding to about  $912 \times 10^6 \tau^+\tau^-$  pairs [19]. The data are collected with the Belle detector [20] at the KEKB asymmetric-energy  $e^+e^-$  collider [21,22].

The Belle detector is a large-solid-angle magnetic spectrometer that consists of a silicon vertex detector (SVD), a 50-layer central drift chamber (CDC), an array of aerogel threshold Cherenkov counters (ACC), a barrel-like arrangement of time-of-flight scintillation counters (TOF), and an electromagnetic calorimeter comprised of CsI(Tl) crystals (ECL) located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux-return located outside of the coil is instrumented to detect  $K_L^0$  mesons and to identify muons (KLM).

The most critical subdetector for this study is CDC [23]. It has the following dimensions: the length is 2400 mm, and the inner and outer radii are 83 and 874 mm, respectively. This size is large enough to reliably reconstruct both daughter electron and mother muon tracks.

To study the background processes, optimize the selection criteria, and obtain the fit function, signal and background Monte Carlo (MC) samples are used.

A signal MC sample of  $e^+e^- \rightarrow \tau^+\tau^-$  with the following  $\tau^- \rightarrow \mu^- (\rightarrow e^- \bar{\nu}_e \nu_\mu) \bar{\nu}_\mu \nu_\tau$  cascade decay is  $\sim 50$  times larger than the data. The production and subsequent decay of  $\tau^+\tau^-$  pairs are generated with KKMC [24] and TAUOLA [25,26] generators, respectively, and decay products are propagated by GEANT3 [27] to simulate the detector response. The  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay is also generated by GEANT3, assuming muons are unpolarized as if  $\xi' = 0$ . To speed up the signal MC sample generation, we reduce the muon lifetime in GEANT3 by 100. This procedure is justified and only slightly biases the distribution of the muon decay length because the CDC size is much smaller than the average flight distance of the muon from the  $\tau$  decay, which is of the order of a kilometer. We evaluate the effect of this reduction of the muon lifetime in the MC generation and quote a systematic uncertainty associated with it.

An example of the  $e^+e^- \rightarrow \tau^+\tau^-$  MC event display with the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  kink in the CDC is presented in Fig. 2. The  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay that occurred in the central volume of the CDC is clearly observed as a kinked track due to the change of the trajectory curvature since the daughter electron from the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay has a smaller momentum in the laboratory frame compared to the mother muon. Both electron and muon trajectories are reconstructed as separate tracks by the Belle track reconstruction algorithm.

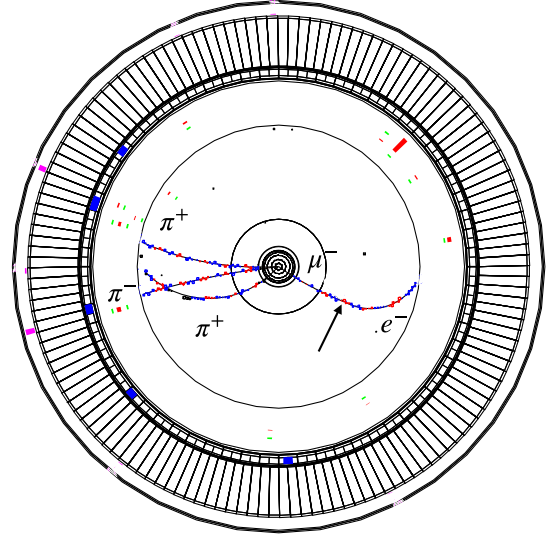


FIG. 2. Event display of a MC event  $e^+e^- \rightarrow \tau^+\tau^- \rightarrow (\pi^+\pi^-\pi^+\bar{\nu}_\tau)(\mu^-\bar{\nu}_\mu\nu_\tau)$  with  $\mu^- \rightarrow e^-\bar{\nu}_e\nu_\mu$  decay in the CDC (the arrow points to the decay vertex). The Belle detector, without the KLM, is shown projected onto  $x$ - $y$  plane.

The background consists of  $\tau^+\tau^-$ -pair events without a  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay and non- $\tau^+\tau^-$ -pair events. The MC sample for the former contribution is generated the same way as the signal, with an exception of the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay generation step. The non- $\tau^+\tau^-$ -pair background consists of the dimuon  $e^+e^- \rightarrow \mu^+\mu^-$  process,  $e^+e^- \rightarrow q\bar{q}$  ( $q = u, d, s$ , and  $c$ ) continuum and  $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$  events, two-photon mediated processes ( $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ ,  $e^+e^-q\bar{q}$ , where  $\ell = e, \mu$  and  $q = u, d, s$ , and  $c$ ), and Bhabha scattering generated with KKMC, EvtGen [28], AAFH [29], and BHLUMI [30] generators, respectively. Final-state radiation is simulated using the PHOTOS [31] package for all charged final-state particles.

A list of the background MC samples is presented in Table I with the ratio of the number of generated events  $N_{\text{MC}}^{\text{gen}}$  to the expected number of corresponding events in data (product of the integrated luminosity  $\mathcal{L}_{\text{data}}^{\text{int}}$  and the process cross section  $\sigma_{\text{proc}}$ ).

TABLE I. Background MC samples with their size.

Processes	$N_{\text{MC}}^{\text{gen}} / (\mathcal{L}_{\text{data}}^{\text{int}} \sigma_{\text{proc}})$
$e^+e^- \rightarrow \tau^+\tau^-$ background	4.5
$e^+e^- \rightarrow \mu^+\mu^-$	4.4
$e^+e^- \rightarrow q\bar{q}$ ( $q = u, d, s, c$ )	5.8
$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$	10.2
$e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ ( $\ell = e, \mu$ )	6.9
$e^+e^- \rightarrow e^+e^-q\bar{q}$ ( $q = u, d$ )	7.5
$e^+e^- \rightarrow e^+e^-q\bar{q}$ ( $q = s, c$ )	8.1
Bhabha scattering	0.2



#### IV. EVENT SELECTION

The event selection is performed in three steps. The first step is the preselection of candidates in  $\tau^+\tau^-$  events with the  $\tau^+\tau^-$ -pair decay topology of interest. The second step is dedicated to the kink candidate selection. In the last step, we apply the BDT (boosted decision tree classifier) machine learning algorithm [32,33] to select signal event candidates and suppress the kink background.

##### A. Preselection

In the first step,  $\tau^+\tau^-$ -pair event candidates are required to pass the preliminary selection criteria. They are used to select the  $\tau^+\tau^-$ -pair decay topology and suppress the contribution from Bhabha scattering,  $e^+e^- \rightarrow \mu^+\mu^-$ , two-photon production,  $e^+e^- \rightarrow q\bar{q}$  ( $q = u, d, s, \text{ or } c$ ), and  $B\bar{B}$  events.

In the c.m. frame, the event is divided by the plane perpendicular to the thrust vector  $\vec{n}_T$  into two hemispheres. The vector  $\vec{n}_T$  is defined as follows

$$T = \max_{\vec{n}_T} \frac{\sum_i |\vec{p}_i \cdot \vec{n}_T|}{\sum_i |\vec{p}_i|}. \quad (10)$$

Here  $\vec{p}_i$  is the momentum of the  $i$ th track; the summation is over all tracks in the event. The signal hemisphere is determined by the muon candidate momentum direction. The complementary one is called tagging hemisphere.

In the present analysis, the decay mode of the second  $\tau$  lepton is not important. Therefore, our selection includes only the information about the event topology formed by charged tracks from the IP. In the signal hemisphere, we require only one track from the IP with the impact parameters in the  $r\phi$ -plane and along the  $z$ -axis (the direction opposite to the  $e^+$  beam) to be  $dr_{\text{sig}} < 2$  cm and  $|dz_{\text{sig}}| < 4$  cm, respectively. We also require one secondary electron candidate track in the signal hemisphere; however, at this step, the parameters of this track are not used. As the  $\tau$  lepton decays dominantly into one or three charged tracks in the final state, in the tagging hemisphere, we require one (topology 1–1) or three (topology 1–3) charged tracks from the IP with their impact parameters to be  $dr_{\text{tag}} < 0.5$  cm and  $|dz_{\text{tag}}| < 2$  cm. The total charge of the event is required to be zero.

Some events may contain photons, for example, from  $\pi^0$ s. They are selected with the energy requirement  $E_\gamma > 50$  MeV. In the signal hemisphere, the maximum photon energy and the total sum of the photon energies are limited to be less than 300 MeV and 400 MeV, respectively, and the  $\pi^0$  candidates (a combination of two photons with  $|M(\gamma\gamma) - m_{\pi^0}| < 15$  MeV/ $c^2$ , corresponding to approximately  $\pm 3\sigma$  window in the resolution) are vetoed.

For topology 1–1, the primary backgrounds are Bhabha scattering, two-photon interactions,  $e^+e^- \rightarrow \mu^+\mu^-$ , and  $e^+e^- \rightarrow q\bar{q}$  ( $q = u, d, s, \text{ or } c$ ). For topology 1–3, the

main background is  $e^+e^- \rightarrow q\bar{q}$  ( $q = u, d, s, \text{ or } c$ ). To suppress the contribution of these processes, additional requirements are used. They are based on the fact that the  $e^+e^- \rightarrow \tau^+\tau^-$  events with the  $\tau^+\tau^-$ -pair subsequent decay are characterized by a large missing energy ( $E_{\text{miss}}$ ) and missing momentum ( $\vec{p}_{\text{miss}}$ ) due to undetected neutrinos. The missing four-momentum ( $\vec{p}_{\text{miss}}, E_{\text{miss}}$ ) is defined as follows

$$P_{\text{miss}} = P_{\text{beam}}^* - P_{\text{trk(IP)}}^* - P_\gamma^*, \quad (11)$$

where  $P_{\text{beam}}^*$  is the beam four-momentum in the c.m. frame,  $P_{\text{trk(IP)}}^*$  is a sum of four-momenta of all tracks from the  $\tau^+\tau^-$ -pair in the c.m. frame, and  $P_\gamma^*$  is a sum of four-momenta of all photons in the c.m. frame. Another feature of  $\tau^+\tau^-$  events is a nearly uniform distribution of  $\cos \theta_{\text{miss}}$ , where  $\theta_{\text{miss}}$  is the angle between the  $\vec{p}_{\text{miss}}$  and  $z$ -axis.

Thus, we apply the requirements on the missing mass ( $m_{\text{miss}}^2 = P_{\text{miss}}^2$ )  $1 \text{ GeV}/c^2 < m_{\text{miss}} < 7 \text{ GeV}/c^2$ , missing angle  $\pi/6 < \theta_{\text{miss}} < 5\pi/6$ , thrust magnitude  $0.85 < T < 0.99$ , and invariant mass of the tag-side tracks  $m_{\text{tk}}^{\text{tag}} < 1.8 \text{ GeV}/c^2$ .

To suppress the remaining Bhabha scattering contribution, we apply an electron veto for the tag-side track for topology 1–1 using identification based on the information from the CDC, ACC, and ECL [34]. We require its likelihood ratio  $\mathcal{R}(e/x) = \mathcal{L}_e/(\mathcal{L}_e + \mathcal{L}_x)$  to be less than 0.4, where  $\mathcal{L}_e$  and  $\mathcal{L}_x$  are the likelihood values of the track for the electron and non-electron hypotheses, respectively. This requirement rejects about 80% of events with an electron on the tag side.

##### B. Kink selection

In this subsection, we describe the preselection of candidates for events with the  $\mu^- \rightarrow e^-\bar{\nu}_e\nu_\mu$  decay in the CDC. As mentioned above, the daughter electron track originating from the muon decay in the signal hemisphere is required to infer the muon polarization. To suppress random combinations with tracks from IP, we require the electron candidate impact parameter in the  $r\phi$ -plane to be  $dr_e > 4$  cm. To reconstruct the  $\mu^- \rightarrow e^-\bar{\nu}_e\nu_\mu$  decay inside the CDC, both the muon track and the electron track have to be reconstructed, leaving enough hits in the tracker. The last point of the muon track and the first point of the electron track must be inside the CDC, detached at least 10 cm from its walls.

The track helix is parametrized by five parameters, whose determination requires at least five hits in the CDC. It is also important to discard fake tracks; thus, it is required for the total number of the CDC hits to be larger than 7 for the electron candidates and larger than 10 for the muon candidates. Both tracks from the  $\mu^- \rightarrow e^-\bar{\nu}_e\nu_\mu$  decay are shorter than the average track of the nondecayed particle from IP; therefore, we require the number of their

CDC hits to be less than 40. Since the decayed muon does not leave the drift chamber, the absence of associated hits in the outer TOF, ECL, and KLM systems is required. The electron tracks originate outside the SVD and are stopped in the ECL; therefore, for them, we veto signals from the SVD and KLM systems.

Finally, the distance between the muon and electron tracks at the decay vertex is required to be less than 5 cm. This requirement is loose enough to keep almost 100% of kink events while rejecting random combinations of tracks.

The overwhelming majority of events that passed these selection criteria have the form of a track kink. One of these processes is  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$ , and the rest are backgrounds, which mimic the signal. They are light meson decays ( $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$ ,  $K^- \rightarrow \mu^- \bar{\nu}_\mu$ ,  $K^- \rightarrow \pi^0 \mu^- \bar{\nu}_\mu$ ,  $K^- \rightarrow \pi^0 e^- \bar{\nu}_e$ ,  $K^- \rightarrow \pi^- \pi^0$ ,  $K^- \rightarrow \pi^- \pi^+ \pi^-$ ,  $K^- \rightarrow \pi^- \pi^0 \pi^0$ ), and electron scattering, muon scattering, and hadron scattering. In Table II, the signal and the main background processes are listed with their relative contributions. About 20% of pion decay events, 30% of kaon decay events, and 30% of hadron scattering events come from  $e^+ e^- \rightarrow q \bar{q}$ , while all other events are mainly from  $e^+ e^- \rightarrow \tau^+ \tau^-$ .

The kinks, formed by a decay-in-flight, are characterized by daughter particle kinematics in the mother particle rest frame determined by the momentum magnitude and emission angle. These two variables are only defined for the correct pair of mass hypotheses assigned to the tracks, e.g., for  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$ , they are electron and muon mass hypotheses assigned to the daughter and mother particles, respectively. To indicate which pair is used in the particular case, we introduce the following notation:  $p_{p_1 p_2}$  and  $\theta_{p_1 p_2}$  mean the daughter particle momentum and emission angle in the mother particle rest frame with  $p_1$  and  $p_2$  mass hypotheses assigned to the daughter and mother tracks, respectively. Here we measure the daughter particle emission angle from the direction of the mother particle in the laboratory frame because this angle determines the efficiency to reconstruct a decay-in-flight. The efficiency to reconstruct the daughter track from a kink has a maximum for the daughter particles emitted perpendicular to the mother particle direction,

TABLE II. Relative contribution of the signal and background processes after the kink selection and before applying the BDT requirement.

Type	Contribution (%)
$\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$	3.2
$\pi^- \rightarrow \mu^- \bar{\nu}_\mu$	22.4
$K^- \rightarrow \pi^- \pi^0$	3.3
$K^- \rightarrow 3 \text{ body}$	4.6
$K^- \rightarrow \mu^- \bar{\nu}_\mu$	45.9
$e$ -scattering	9.5
$\mu$ -scattering	1.1
Hadron scattering	10.0

while it drops for daughter particle emitted along the muon direction.

In the present study, we use three pairs  $(p_1, p_2)$ :  $(e, \mu)$ ,  $(\pi, K)$ , and  $(\mu, \pi)$ . For these mass hypotheses, we plot  $p_{p_1 p_2}$  and  $\cos \theta_{p_1 p_2}$  distributions in Fig. 3. A good agreement between the MC simulation (filled histograms) and the data (points with errors) is observed.

Both Table II and Fig. 3 show that the largest contribution to the background comes from the pion and kaon two-body decays. These processes are characterized by a peak in the momentum distribution of the daughter particle in the rest frame of the decayed one, which is clearly observed for the  $K^- \rightarrow \mu^- \bar{\nu}_\mu$  and  $K^- \rightarrow \pi^- \pi^0$  decays in Fig. 3(c) and for the  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$  decay in Fig. 3(e), proving the correctness of the applied kink selection procedure. Before the selection based on the BDT, the signal contribution is small and hardly visible in Fig. 3. The signal shape has no sharp structures due to a three-body decay, and it is further smeared in the variables calculated with the wrong pair of mass hypotheses.

We define the signal region for  $p_{e\mu} < 70 \text{ MeV}/c$ . As can be seen from Fig. 3(a), the largest background contribution to this region is from the pion decay and electron scattering. However, in the region for the  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$  decay,  $p_{\mu\pi} < 100 \text{ MeV}/c$ , there are almost no  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  events [see Fig. 3(e)], which makes it possible to effectively suppress the  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$  background, as well as a significant part of the electron scattering events.

### C. BDT based signal selection

To further suppress the background, we apply the BDT machine learning (ML) classification algorithm. To separate the signal from the background, we select twelve features based on the physics of the background processes. The first two features are  $p_{\mu\pi}$  and  $p_{\pi K}$ . The next group of five features is responsible for the particle identification (PID) of muon and electron candidates. They are defined as likelihood ratios  $\mathcal{R}(\ell/x) = \mathcal{L}_\ell / (\mathcal{L}_\ell + \mathcal{L}_x)$ , where  $\ell = \mu$  or  $e$ , and  $\mathcal{L}_\ell$  and  $\mathcal{L}_x$  are the likelihood values of the track for the muon (electron) and non-muon (non-electron) hypotheses, respectively. For muon candidates, we use PID based on the  $dE/dx$  losses inside the CDC against electron, pion, kaon, and proton hypotheses ( $x = e, \pi, K, \text{ and } p$ , respectively). For electron candidates, PID is based on the  $dE/dx$  losses inside the CDC and ECL information; here only  $\mathcal{R}(e/\mu)$  is used. Two more features are related to the decay vertex; they are the  $z$ -coordinate of the last point of the mother particle track and the distance between the daughter and mother tracks at the decay vertex. Finally, to suppress the residual contribution from  $e^+ e^- \rightarrow q \bar{q}$ , we use  $m_{\text{miss}}$ ,  $\cos \theta_{\text{miss}}$ , and thrust magnitude as separation variables.

Although the  $\cos \theta_{p_1 p_2}$  variables show a good separation power (see Fig. 3), we do not use them in the BDT because they are strongly correlated with  $\cos \theta_e$  (the main variable

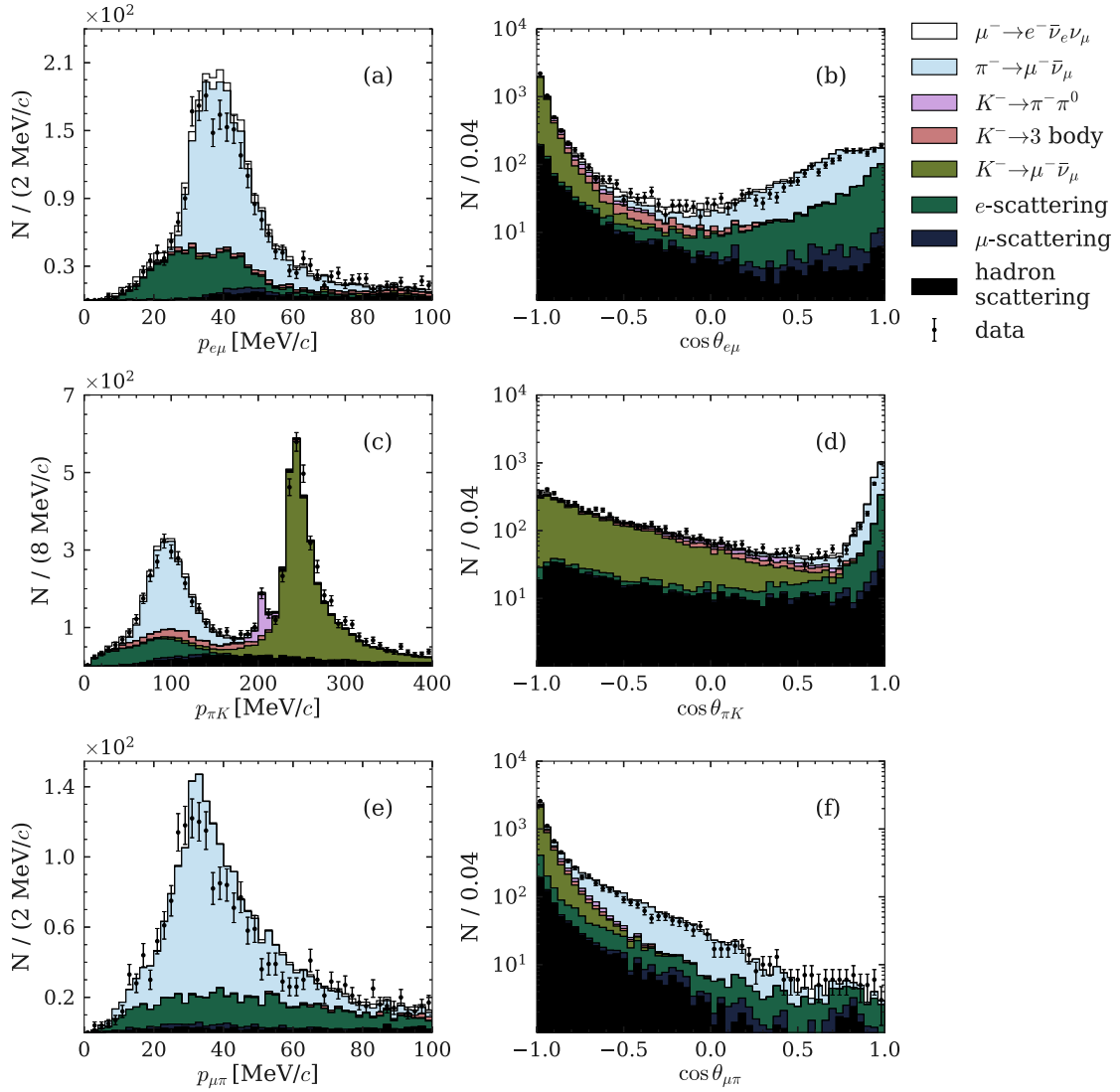


FIG. 3. Momentum  $p_{p_1 p_2}$  and angular  $\cos \theta_{p_1 p_2}$  distributions for the daughter particle (the mass hypothesis  $p_1$ ) in the mother particle (the mass hypothesis  $p_2$ ) rest frame. (a)  $p_{e\mu}$ , (b)  $\cos \theta_{e\mu}$ , (c)  $p_{\pi K}$ , (d)  $\cos \theta_{\pi K}$ , (e)  $p_{\mu\pi}$ , and (f)  $\cos \theta_{\mu\pi}$ .

to fit  $\xi'$ ) and, therefore, bias the  $\xi'$  measurement with poorly controlled systematics.

The distribution of the BDT output variable  $O_{\text{BDT}}$  is shown in Fig. 4 for signal and background for training and test samples. The optimal selection of  $O_{\text{BDT}} > 0.0979$  is obtained by maximizing the ratio  $N_{\text{sig}} / \sqrt{N_{\text{sig}} + N_{\text{bckg}}}$ , where  $N_{\text{sig}}$  is the number of selected signal events, and  $N_{\text{bckg}}$  is the number of selected background events. The obtained signal selection efficiency is  $\varepsilon_{\text{sig}} \approx 80\%$ , while the background is suppressed by a factor of fifty.

To illustrate the performance of BDT, we plot the electron candidate momentum in the muon rest frame shown in Fig. 5. The absence of the Belle track

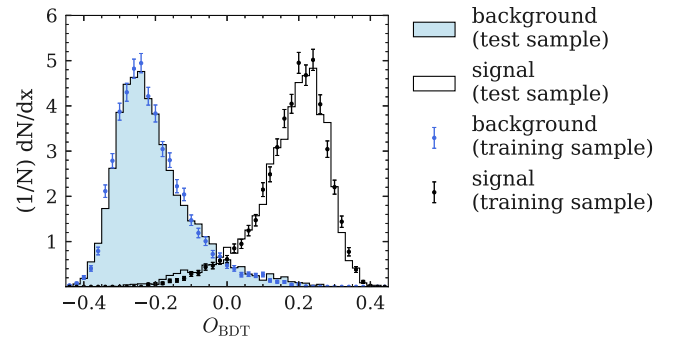


FIG. 4.  $O_{\text{BDT}}$  distribution for the signal and background (training and test samples).

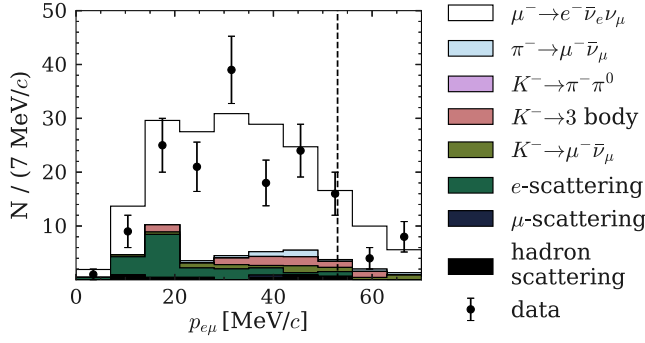


FIG. 5. Momentum distribution for the electron and muon mass hypotheses for the daughter and mother particles, respectively. The dashed line shows the 53 MeV/c threshold.

reconstruction algorithm optimization for the kink events leads to a wide tail above the kinematic threshold of 53 MeV/c in the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay. The relative contribution of the signal and background processes after the BDT application is listed in Table III. About 6% of the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decays come from the non- $\tau^+ \tau^-$  events.

Finally, the number of the reconstructed signal  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decays and background events are estimated from the MC simulation to be  $139 \pm 2$  and  $50 \pm 5$ , respectively, where the uncertainty is due to the limited size of the MC samples. In the data, 165 signal-candidate events pass all the applied selection criteria.

## V. BACKGROUND STUDY

In the present study, the background suppression and determination of the fit function are based on the MC simulation; thus, it is important to control the differences between the MC samples and the data and take them into account as systematic uncertainties. Therefore, we conduct a study of background processes in the data and the MC simulation using large pure samples with different types of kink candidates (pion and kaon decays, hadron and electron scattering).

Light meson decays are selected in two ways. The first method is based on the BDT described in the previous section, where we mark  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$  or  $K^- \rightarrow \mu^- \bar{\nu}_\mu$  decay as

TABLE III. Relative contribution of the signal and background processes after the BDT application.

Type	Contribution (%)
$\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$	77.8
$\pi^- \rightarrow \mu^- \bar{\nu}_\mu$	2.2
$K^- \rightarrow 3 \text{ body}$	4.3
$K^- \rightarrow \mu^- \bar{\nu}_\mu$	3.7
$e$ -scattering	9.6
$\mu$ -scattering	0.2
Hadron scattering	2.2

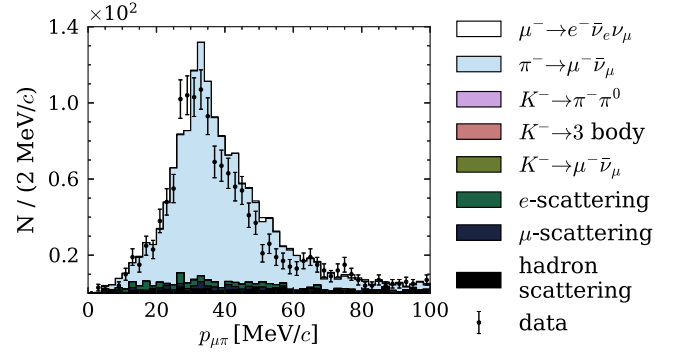


FIG. 6.  $p_{\mu\pi}$  distribution for the  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$  event candidates selected with the BDT from the  $\tau$  sample.

a signal. The samples obtained in this way have a purity close to unity.

The distributions of  $p_{\mu\pi}$  for the selected  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$  sample and  $p_{\mu K}$  for the selected  $K^- \rightarrow \mu^- \bar{\nu}_\mu$  sample are shown in Figs. 6 and 7(a), respectively. In the former plot, the  $p_{\mu\pi}$  distribution in the MC sample is shifted to the higher muon momentum compared to the data. This effect is related to the imperfection of the track reconstruction algorithm in the case of a kink and is especially pronounced in the pion decay due to the low energy release. In Fig. 7(a), we observe that the muon momentum peak has a larger width in the data indicating a better resolution in the MC simulation, while kaon momentum distribution in the

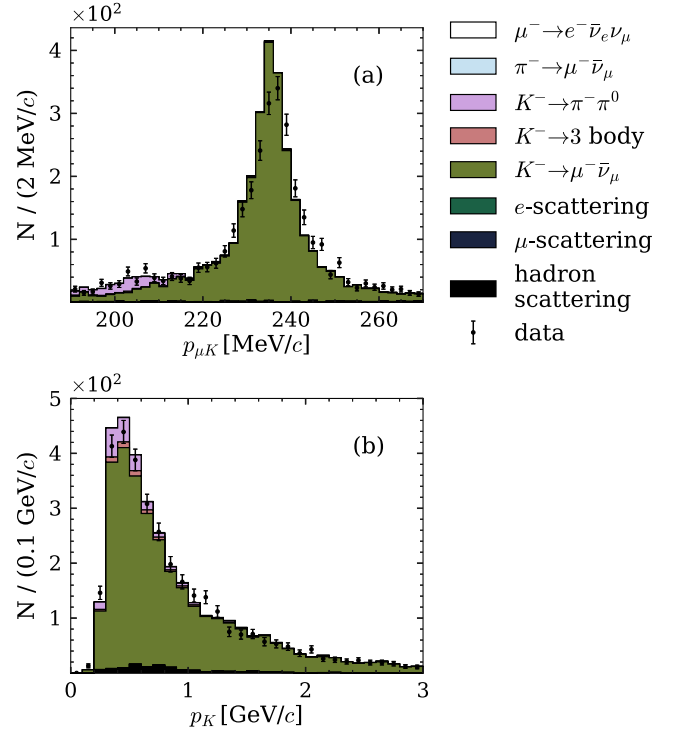


FIG. 7. (a)  $p_{\mu K}$  and (b)  $p_K$  distributions for the  $K^- \rightarrow \mu^- \bar{\nu}_\mu$  event candidates selected with the BDT from the  $\tau$  sample.



laboratory frame  $p_K$  plotted in Fig. 7(b) shows an agreement between the MC simulation and the data within statistical uncertainties of both samples.

It is also important to control systematics caused by the BDT application; thus, background processes have to be studied in samples obtained without ML algorithms. To obtain a high-purity kink sample without BDT, we use the decay chain  $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$  since there is a large sample of  $D^{*+}$  mesons collected by the Belle detector, and both decays,  $D^{*+} \rightarrow D^0\pi^+$  and  $D^0 \rightarrow K^-\pi^+$ , are well-studied. In these events, it is possible to reconstruct the light meson decay-in-flight and tag the kink type. A detailed description of the  $D^{*+}$  sample selection is given in Appendix A.

For the  $\pi^- \rightarrow \mu^-\bar{\nu}_\mu$  kinks selected from the  $D^{*+}$  sample, we plot the  $p_{\mu\pi}$  distribution in Fig. 8(a). It is similar to Fig. 6, although the statistics are several times smaller.

Concerning kaon kinks from the  $D^{*+}$  sample, they include both kaon two-body and three-body decays and a large number of events with hadron scattering. To illustrate the abundance of selected processes, we plot the  $p_{\pi K}$  distribution in Fig. 8(b). Here we observe a relatively large contribution of  $K^- \rightarrow \pi^-\pi^0$  decays compared to Fig. 3(c), where such events are suppressed by requirements for photons and  $\pi^0$ s. For this kaon decay mode, we confirm the agreement between the MC simulation and the data within statistical uncertainties. For  $K^- \rightarrow \mu^-\bar{\nu}_\mu$  decays, we observe a discrepancy; therefore, to study this process in more detail, we plot the  $p_{\mu K}$

distribution in Fig. 8(c). Here the discrepancy in the muon momentum peak width for  $K^- \rightarrow \mu^-\bar{\nu}_\mu$  decays is observed to be similar to one in Fig. 7(a) for the  $\tau$  sample and thus confirms this to be a systematic effect.

Kaon kinks from the  $D^{*+}$  sample also include hadron scattering events [e.g., Fig. 8(b)]. This process is typical for slow hadrons, as can be seen from Fig. 8(d), where  $p_K$  is plotted. For the first two bins with data, we observe a significant discrepancy between the data and MC samples, while an agreement is observed for kaon kinks from the  $\tau$  sample [Fig. 7(b)]. Another confirmation that the MC simulation does not reproduce hadron scattering is an underestimation of the events number in the hadron scattering region observed in Fig. 8(b). The difference between the MC simulation and the data is expected since this process is not perfectly described by GEANT3. For larger  $p_K$ , the MC simulation reproduces the data within statistical uncertainties for both  $D^{*+}$  and  $\tau$  samples.

The electron scattering process makes a significant contribution to the background. The study of this process is based on the sample obtained from the  $\gamma$ -conversion on the detector material in the IP vicinity. The selection of the  $\gamma$ -conversion sample is described in detail in Appendix B. To illustrate the electron scattering process, we use the same pair of mass hypotheses as in the fit of the  $\mu^- \rightarrow e^-\bar{\nu}_e\nu_\mu$  process. The  $p_{e\mu}$  distribution is shown in Fig. 9. A discrepancy between the MC and data samples is observed in the shape of the electron spectrum and taken into account in the systematics.

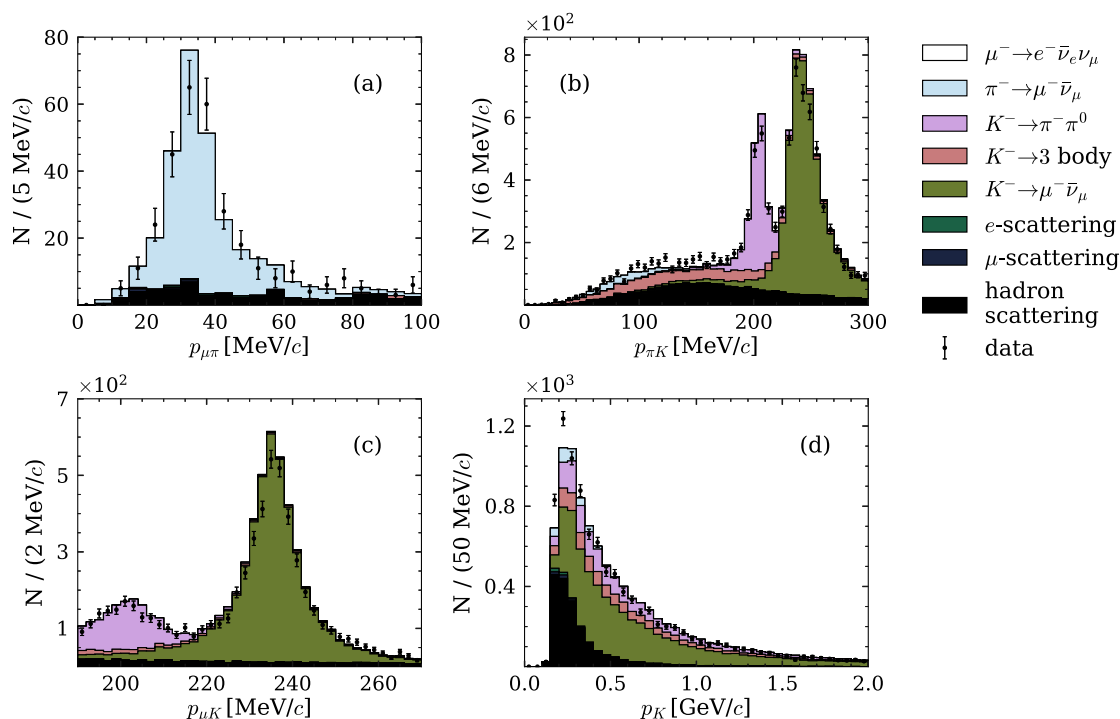


FIG. 8.  $D^{*+}$  sample. (a)  $p_{\mu\pi}$  distribution for pion kinks; (b)  $p_{\pi K}$ , (c)  $p_{\mu K}$ , and (d)  $p_K$  distributions for kaon kinks.

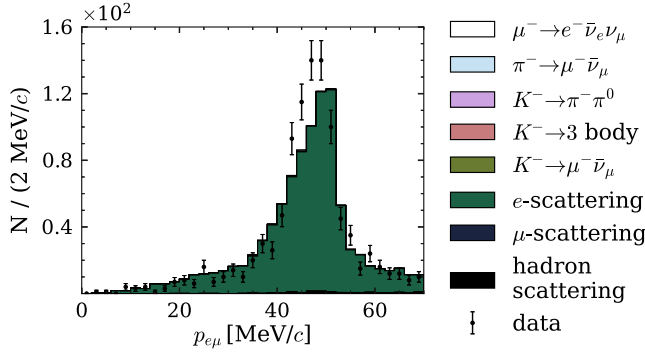


FIG. 9.  $p_{e\mu}$  distribution for the electron scattering kinks selected from the  $\gamma$ -conversion sample.

In conclusion, a complete study of the main background processes is done. All observed discrepancies are taken into account as systematic uncertainties. In addition, the discussed samples also provide important information about secondary and primary tracks in events that contain kinks for all main particle types except primary muon. This information is also used in systematics estimation, as it is described in the corresponding section below.

## VI. FIT FUNCTION AND FIT RESULT

According to Eq. (5), the term proportional to  $\xi'$  depends on  $\cos \theta_e$ ,  $y$ , and  $x$ . Since the dependence on  $x$  is very weak,<sup>3</sup> we can integrate over it without loss of sensitivity. In contrast, the dependence on  $y$  is strong, and integration over it dramatically decreases the sensitivity to  $\xi'$ ; thus, we perform a two-dimensional (2D) fit on the  $(y, \cos \theta_e) \equiv (y, c)$  distribution using an unbinned maximum-likelihood method.

The likelihood function is

$$\mathcal{L} = \prod_{i=1}^n \mathcal{P}(y_i, c_i; \xi'), \quad (12)$$

where  $n$  denotes the number of data events, and  $\mathcal{P}(y, c; \xi')$  is a probability density function (PDF)

$$\mathcal{P}(y, c; \xi') = p \mathcal{P}_{\text{sig}}(y, c; \xi') + (1 - p) \mathcal{P}_{\text{bckg}}(y, c). \quad (13)$$

Here  $p = N_{\text{sig}} / (N_{\text{sig}} + N_{\text{bckg}}) = 0.74$  is the signal purity, the ratio of the number of signal events  $N_{\text{sig}}$  to the total number of events  $N_{\text{sig}} + N_{\text{bckg}}$ ,  $\mathcal{P}_{\text{sig}}(y, c; \xi')$  is a signal PDF, and  $\mathcal{P}_{\text{bckg}}(y, c)$  is a background PDF. The signal purity and both PDFs are obtained from the MC simulation.

The signal PDF can be determined from the theoretical PDF  $\mathcal{P}_{\text{th}}(\tilde{y}, \tilde{c}, \mathbf{z}; \xi')$  by applying efficiency

corrections and performing a convolution with the detector resolution:

$$\begin{aligned} \mathcal{P}_{\text{sig}}(y, c; \xi') &= \frac{1}{N(\xi')} \int \mathcal{P}_{\text{th}}(\tilde{y}, \tilde{c}, \mathbf{z}; \xi') \\ &\quad \times \eta(\tilde{y}, \tilde{c}, \mathbf{z}) g(y, c, \tilde{y}, \tilde{c}, \mathbf{z}) d\mathbf{z} d\tilde{y} d\tilde{c}, \\ N(\xi') &= \int \mathcal{P}_{\text{th}}(\tilde{y}, \tilde{c}, \mathbf{z}; \xi') \eta(\tilde{y}, \tilde{c}, \mathbf{z}) \\ &\quad \times g(y, c, \tilde{y}, \tilde{c}, \mathbf{z}) d\mathbf{z} d\tilde{y} d\tilde{c} dy dc. \end{aligned} \quad (14)$$

Here  $(\tilde{y}, \tilde{c}) \equiv (\tilde{y}, \cos \tilde{\theta}_e)$  are “true” physical quantities of our interest, and  $\mathbf{z}$  is a vector of the rest of the true physical variables not used in the fit. Functions  $\eta(\tilde{y}, \tilde{c}, \mathbf{z})$  and  $g(y, c, \tilde{y}, \tilde{c}, \mathbf{z})$  are the efficiency and resolution functions, respectively. The theoretical PDF,  $\mathcal{P}_{\text{th}}(\tilde{y}, \tilde{c}, \mathbf{z}; \xi')$ , is obtained from the differential decay width given by Eq. (5).

Both efficiency and resolution functions are too complicated to express in analytic form; thus, it is almost impossible to calculate the  $\mathcal{P}_{\text{sig}}(y, c; \xi')$  function given in Eq. (14) analytically. Fortunately, the theoretical PDF is linear in  $\xi'$ ; therefore, we rewrite it as follows

$$\mathcal{P}_{\text{th}}(\tilde{y}, \tilde{c}, \mathbf{z}; \xi') = A(\tilde{y}, \tilde{c}, \mathbf{z}) + \xi' B(\tilde{y}, \tilde{c}, \mathbf{z}). \quad (15)$$

Using this form, we rewrite Eq. (14):

$$\mathcal{P}_{\text{sig}}(y, c; \xi') = \frac{\bar{A}(y, c) + \xi' \bar{B}(y, c)}{\bar{A} + \xi' \bar{B}}, \quad (16)$$

where

$$\begin{aligned} \bar{A}(y, c) &= \int A(\tilde{y}, \tilde{c}, \mathbf{z}) \eta(\tilde{y}, \tilde{c}, \mathbf{z}) g(y, c, \tilde{y}, \tilde{c}, \mathbf{z}) d\mathbf{z} d\tilde{y} d\tilde{c}, \\ \bar{B}(y, c) &= \int B(\tilde{y}, \tilde{c}, \mathbf{z}) \eta(\tilde{y}, \tilde{c}, \mathbf{z}) g(y, c, \tilde{y}, \tilde{c}, \mathbf{z}) d\mathbf{z} d\tilde{y} d\tilde{c}, \\ \bar{A} &= \int \bar{A}(y, c) dy dc, \quad \bar{B} = \int \bar{B}(y, c) dy dc. \end{aligned} \quad (17)$$

In this study, the dependence of the signal PDF normalization on  $\xi'$  is negligible as  $\bar{A} \gg \bar{B}$ .

To calculate  $\bar{A}(y, c) / \bar{A}$  and  $\bar{B}(y, c) / \bar{A}$ , we use two MC samples generated with  $\xi' = 1$  and  $\xi' = -1$ . Their distributions in  $(y, c)$  are determined exactly by the following PDFs:  $\mathcal{P}_{+1}(y, c) = \mathcal{P}_{\text{sig}}(y, c; +1)$  and  $\mathcal{P}_{-1}(y, c) = \mathcal{P}_{\text{sig}}(y, c; -1)$ , respectively, providing

$$\begin{aligned} \mathcal{P}_{\text{sig}}(y, c; \xi') &= \frac{1}{2} \{ \mathcal{P}_{+1}(y, c) + \mathcal{P}_{-1}(y, c) \\ &\quad + \xi' [ \mathcal{P}_{+1}(y, c) - \mathcal{P}_{-1}(y, c) ] \}. \end{aligned} \quad (18)$$

All the PDFs can be obtained in the form of 2D histograms of  $(y, c)$  or in the form of smooth functions describing the distributions in  $(y, c)$ . Since the signal MC sample statistics is large, 2D histograms of  $(y, c)$  can already be considered as almost smooth functions (there are no statistically significant fluctuations), which can be used

<sup>3</sup>If  $x_0^2$  is neglected in Eq. (5), the dependence on  $x$  factorizes.

in the fit without loss of accuracy. Thus, for simplicity and naturalness, we obtain  $\mathcal{P}_{\pm 1}(y, c)$  in the form of the 2D histogram of  $10 \times 10$  bins with an interpolation of the intermediate values. Alternatively, we use a smooth function to evaluate the systematic uncertainty as it is described in Sec. VIII D.

In contrast to the signal, the background MC sample has modest statistics, and there is no feasibility to increase it. Therefore,  $\mathcal{P}_{\text{bckg}}(y, c)$  is obtained from the approximation of a  $6 \times 6$ -bin histogram of  $(y, c)$  distribution by a smooth parametric function so that  $\chi^2/\text{n.d.f.} \approx 1$ .

The fit procedure is tested on ensembles of 1000 statistically independent simulated samples of the size expected in the data with eleven  $\xi'$  seed values from  $-1$  to  $1$  in steps of  $0.2$ , and no statistically significant biases are observed.

Finally, the fit to the data yielded  $\xi' = 0.22 \pm 0.94$ . The projections of the data and the fit function onto the  $y$  and  $\cos\theta_e$  axes are shown in Fig. 10. The variation of the  $-2[\ln\mathcal{L}(\xi') - \ln\mathcal{L}(\xi'_{\text{fit}})]$  as a function of the  $\xi'$  value is shown in Fig. 11. The  $\xi' = -1$  scenario is more than one standard deviation away from the measured  $\xi'$  value.

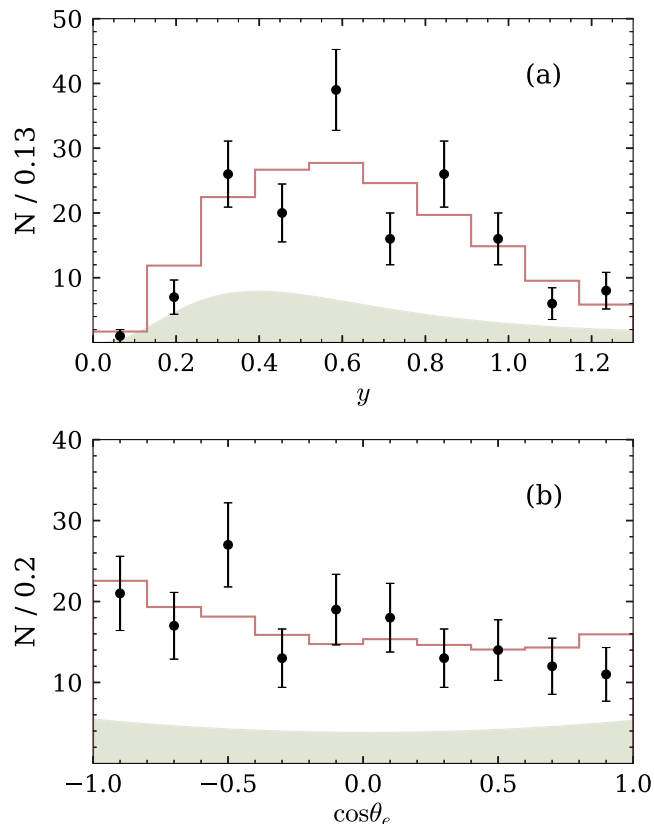


FIG. 10. The fit of the data with  $\xi' = 0.22 \pm 0.94$ . Points with errors correspond to the data, the solid line histogram corresponds to the fit function, and the shadowed area corresponds to the background function. (a) projection onto  $y$ , (b) projection onto  $\cos\theta_e$ .

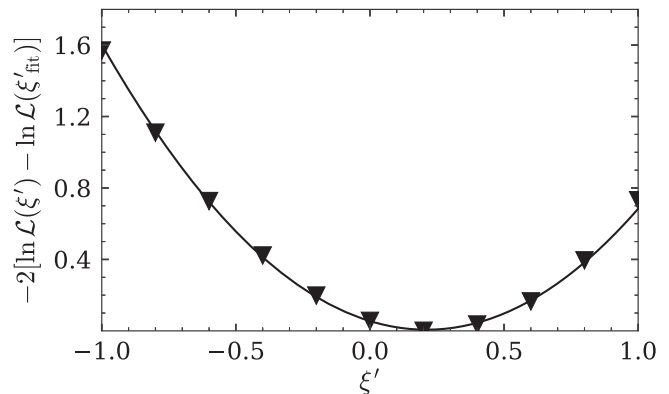


FIG. 11. The variation of the  $-2[\ln\mathcal{L}(\xi') - \ln\mathcal{L}(\xi'_{\text{fit}})]$  as a function of the  $\xi'$  value.

For a more detailed illustration of the fit, we plot three slices in  $y$  ( $0 < y < 0.52$ ,  $0.52 < y < 0.78$ , and  $0.78 < y < 1.3$ ) projected onto  $\cos\theta_e$  (Fig. 12). In addition to the fit function with  $\xi' = 0.22$ , we show fit functions with  $\xi' = -1$  (dashed) and with  $\xi' = 1$  (dash-dotted). For  $y < 0.52$ , there is almost no sensitivity to  $\xi'$ , while for  $0.78 < y < 1.3$ , the sensitivity is maximum. This behavior is expected from the theoretical function given by Eq. (5). The total  $\chi^2$  for the fit projections shown in Fig. 12 is 31 with  $\text{n.d.f.} = 29$ , demonstrating that the fit describes the data well. The total  $\chi^2$  for the projections shown in Fig. 12 for the function with  $\xi' = -1$  is 37 and for the function with  $\xi' = 1$  is 30.

## VII. SYSTEMATIC UNCERTAINTIES

The systematic uncertainties are taken into account by assuming the most conservative approach. To estimate them, we generate for each source of the systematics an ensemble of 1000 toy MC samples with eleven  $\xi'$  seed values from  $-1$  to  $1$  in steps of  $0.2$  and the same statistics as estimated from the signal and background MC samples. Each sample is generated according to the 2D distribution in  $y$  and  $\cos\theta_e$  obtained from variation of the signal and background distributions within the expected uncertainties (observed discrepancies between the MC simulation and the data described in the previous sections). Then all samples are fitted with the default PDF function, and the average of obtained  $\xi'$  values over 1000 samples is calculated. The maximum difference between these mean values and the default one is taken as a systematic uncertainty.

We also estimate the systematic uncertainties from the data by varying the PDF functions used in the fit to obtain the difference between a new  $\xi'$  value and the default one. We use this method as a crosscheck since it is less robust for systematics evaluation and always gives a lower value compared to the estimation from toy MC samples.

In this study, we distinguish four main categories of the systematic error sources: “background,” “PID in BDT,” “signal,” and “fit procedure.”

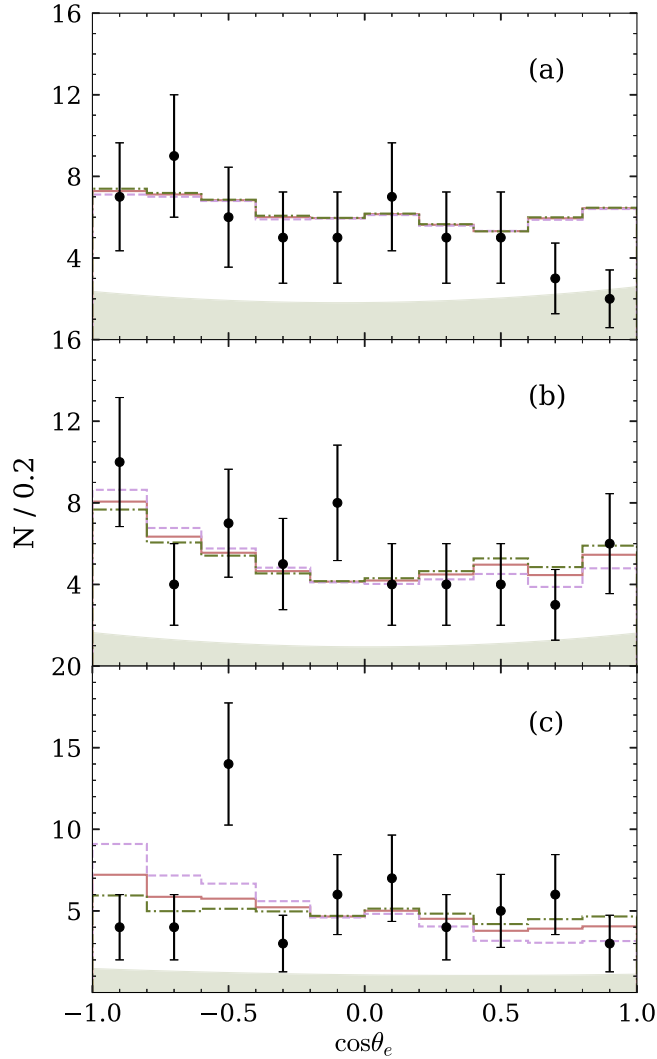


FIG. 12. Projection onto  $\cos\theta_e$  for slices in  $y$ : (a)  $0 < y < 0.52$ , (b)  $0.52 < y < 0.78$ , and (c)  $0.78 < y < 1.3$ . Points with errors correspond to the data, the solid line corresponds to the  $\xi' = 0.22$  fit function, the dashed line corresponds to the  $\xi' = -1$  fit function, the dash-dotted line corresponds to the  $\xi' = 1$  fit function, and the shadowed area corresponds to the background function.

### A. Background systematics

This category of systematic errors includes the uncertainties in the expected background fraction of each type used in the fit PDF as well as the particular background shape.

The signal purity  $p$  is obtained from the MC simulation with the statistical uncertainty of 0.02 induced by the limited number of signal and background MC events. While the signal MC sample is generated with large statistics, the size of the background MC sample is moderate, thus making a major contribution. The observed discrepancies between the data and simulation lead to an additional systematical uncertainty of 0.02 in the purity value. The variation of  $p$  within the combined error results in the systematic uncertainty of 0.10.

The PDF shape and relative contribution of each type of the background processes are the sources of the systematics because the MC simulation does not reproduce data perfectly. The main background contamination comes from the kaon and pion decays, electron and hadron scattering. For each of them, we conducted a small dedicated study described in Sec. V. Prepared pure background samples allow us to observe discrepancies between the MC simulation and the data in both normalization and shape. To take these discrepancies into account, we reweight each type of the background MC sample based on particular kinematical characteristics. For the reweighted background sample, we obtain new values of the background smooth function parameters. The estimated systematic uncertainties are the following: 0.05 for  $\pi^- \rightarrow \mu^- \bar{\nu}_\mu$ , 0.06 for  $K^- \rightarrow 3$  body, 0.05 for  $K^- \rightarrow \mu^- \bar{\nu}_\mu$ , 0.10 for  $e$ -scattering, and 0.11 for hadron scattering. The statistical uncertainties of the parameters of the background PDF do not have much impact on the shape and have already been taken into account in the signal purity systematics.

The combined background systematic uncertainty is 0.20.

### B. PID in BDT systematics

To estimate the effects of PID usage in the BDT, we take advantage of the availability of various tagged kinks in the data selected without BDT application. This systematic uncertainty contains two separate contributions: PID uncertainties of primary muons and daughter electrons.

The PID uncertainty of the daughter track is easier to analyze since we have tagged secondary electrons from the electron scattering ( $\gamma$ -conversion sample), muons from kaon decays ( $D^{*+}$  sample), and pions from kaon decays ( $D^{*+}$  sample). We reweight the  $\mathcal{R}(e/\mu)$  distribution for both the signal and background according to the weights obtained from the corresponding sample and then apply a new PDF for toy MC sample generation. The obtained systematic uncertainty is 0.13.

To identify muons, we use PID with all four pairs of particle hypotheses (muon against electron, pion, kaon, or proton) in the BDT. To simplify, we evaluate the systematics of PID for all of them separately. Although they are correlated, the separate analysis only increases the systematic uncertainty.

For all kink mother particle types except for the muon, we have a clean sample providing the corresponding PID distribution in the data (electrons from the  $\gamma$ -conversion sample, kaons and pions from the  $D^{*+}$  sample). Muons do not have a suitable sample; therefore, we treat them as pions instead. This replacement is justified since we use only  $dE/dx$  losses, and they are almost the same for the muon and pion mass hypotheses.

We reweight the  $\mathcal{R}(\mu/x)$  distribution ( $x = e, \pi, K, \text{ or } p$ ) for both the signal and background and then apply a new PDF for toy MC sample generation. The obtained



uncertainties are 0.13 for  $\mathcal{R}(\mu/e)$ , 0.09 for  $\mathcal{R}(\mu/\pi)$ , 0.10 for  $\mathcal{R}(\mu/K)$ , and 0.06 for  $\mathcal{R}(\mu/p)$ .

The combined PID in BDT systematic uncertainty is 0.24.

### C. Signal PDF systematics

Here, we study all sources of systematic uncertainties related to the signal PDF  $\mathcal{P}_{\text{sig}}$ . These include signal reconstruction efficiency depending on the muon laboratory momentum  $p_\mu$  and the electron emission angle in the muon rest frame  $\theta_{e\mu}$ , electron momentum resolution in the muon rest frame, and also the systematics of the signal MC sample generation method.

The systematic uncertainty due to the discrepancy in reconstruction efficiency between the data and the MC simulation consists of two different contributions: one is related to the trigger efficiency of the selected topology, and the other is due to the kink reconstruction efficiency. The trigger efficiency uncertainty results in a small discrepancy between primary muon momentum distributions in the MC and data samples. We obtain weights for the estimation from the sample of  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decays without  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  kink. This systematic uncertainty is 0.08.

The reconstruction efficiency also strongly depends on the electron emission angle  $\theta_{e\mu}$ . We control this effect using the largest selected background sample of  $K^- \rightarrow \mu^- \bar{\nu}_\mu$  decays (from  $D^{*+}$  decays). Since kaons are pseudoscalars, their decay angular distribution is uniform. Therefore, after reconstruction, the  $\cos \theta_{\mu K}$  distribution represents the kink reconstruction efficiency. Applying weights from the kaon decay to our signal, we estimate the systematic uncertainty to be 0.09.

To take into account the systematics of the electron momentum resolution in the muon rest frame, we also exploit the  $K^- \rightarrow \mu^- \bar{\nu}_\mu$  decay sample. We use kaons selected from the  $\tau$  sample since here we observe a discrepancy in  $p_{\mu K}$  resolution slightly larger compared to kaon kinks from the  $D^{*+}$  sample. We estimate systematic uncertainty to be 0.03.

The systematic uncertainty induced by the muon lifetime reduction for the signal MC sample generation is 0.07. It is estimated by comparing the signal MC sample to the ten times smaller MC sample generated with the default muon lifetime.

The combined signal PDF systematic uncertainty is 0.14.

### D. Fit procedure systematics

To estimate the systematic uncertainty of the fit procedure, we compare the fit results of the Michel parameter  $\xi'$  for two different  $\mathcal{P}_{\text{sig}}$ . The first one is the default signal PDF in the form of a histogram. The second one is obtained from the MC ( $y, \cos \theta_e$ ) distribution by smoothing a  $16 \times 16$ -bin 2D histogram with a parametric function. We estimate the systematic uncertainty of this source to be 0.25.

In addition, we check the difference in the result by varying the bin width of the default  $\mathcal{P}_{\text{sig}}$ . It is impossible to vary the bin size much since with too fine binning, a few empty bins appear, leading to a bias of the fit, while with too rough binning, the sensitivity suffers due to the fitting function sharpness in some regions. Thus,  $8 \times 8$  and  $12 \times 12$  net is used to check the variation. The obtained difference is small and does not exceed 0.13.

In conclusion, we consider 0.25 as a systematic uncertainty of the fit procedure.

### E. Systematics summary

Finally, the combined overall systematic uncertainty of the Michel parameter  $\xi'$  measurement is estimated to be  $\sigma_{\xi'} = 0.42$ . In Table IV, we summarize the results of the systematic uncertainty estimation for all sources.

Systematic uncertainty is significantly smaller than the statistical one in this analysis, demonstrating the potential of the applied method. The method allows for a significant improvement in accuracy in the near future in already working experiments or those being under development. Thus, the task of control of systematic uncertainty with increasing statistics is worth considering.

A qualitative consideration that the statistical uncertainty will dominate the systematic one in similar analyses in the near future experiments is discussed in detail in Ref. [17]. This analysis confirms in practice the validity of that conclusion: most of the systematic sources are controlled

TABLE IV. Sources of the systematic uncertainties of the Michel parameter  $\xi'$  measurement (absolute values).

Source	Uncertainty
Background	
Purity ( $p$ )	0.10
$\pi^- \rightarrow \mu^- \bar{\nu}_\mu$ MC	0.05
$K^- \rightarrow 3$ body MC	0.06
$K^- \rightarrow \mu^- \bar{\nu}_\mu$ MC	0.05
$e$ -scattering MC	0.10
Hadron scattering MC	0.11
PID in BDT	
$\mathcal{R}(e/\mu)$	0.13
$\mathcal{R}(\mu/e)$	0.13
$\mathcal{R}(\mu/\pi)$	0.09
$\mathcal{R}(\mu/K)$	0.10
$\mathcal{R}(\mu/p)$	0.06
Signal PDF	
$p_\mu$ efficiency	0.08
$\cos \theta_{e\mu}$ efficiency	0.09
$p_{e\mu}$ resolution	0.03
Signal MC generation	0.07
Fit procedure	
Fit function	0.25
Total	0.42

with large independent samples, and no limiting factors for further improvements in accuracy have yet been observed.

### VIII. RESULT

We measure the Michel parameter  $\xi'$  to be

$$\xi' = 0.22 \pm 0.94 \pm 0.42, \quad (19)$$

where the first uncertainty is statistical and the second one is systematic. This result is consistent with the Standard Model prediction of  $\xi'_{\text{SM}} = 1$ . The combined uncertainty,  $\sigma_{\xi'} = 1.03$ , is more than two times smaller compared to the previous Belle result  $\xi' = -2.2 \pm 2.4$  calculated from  $\xi\kappa$  obtained in the study of the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau \gamma$  decay [12].

Based on the gained experience, it is possible to improve the result in the near future in the Belle II experiment [35], taking into account the upgraded detector with an enlarged CDC and the implementation of improved tracking algorithms. In particular, the kink reconstruction algorithm implementation will provide a better momentum resolution, which is important for both background suppression and sensitivity increase (smeared by the resolution otherwise).

### IX. CONCLUSION

In summary, we report the first direct measurement of the Michel parameter  $\xi'$  in the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay with the full Belle data sample using the  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  decay-in-flight in the Belle drift chamber. The obtained value of  $\xi' = 0.22 \pm 0.94 \pm 0.42$ , where the first uncertainty is statistical and the second one is systematic, is in agreement with the Standard Model prediction  $\xi'_{\text{SM}} = 1$ .

### ACKNOWLEDGMENTS

This work, based on data collected using the Belle detector, which was operated until June 2010, was supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, the Japan Society for the Promotion of Science (JSPS), and the Tau-Lepton Physics Research Center of Nagoya University; the Australian Research Council including Grants No. DP210101900, No. DP210102831, No. DE220100462, No. LE210100098, No. LE230100085; Austrian Federal Ministry of Education, Science and Research (FWF) and FWF Austrian Science Fund No. P 31361-N36; the National Natural Science Foundation of China under Contracts No. 11675166, No. 11705209; No. 11975076; No. 12135005; No. 12175041; No. 12161141008; Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS), Grant No. QYZDJ-SSW-SLH011; Project ZR2022JQ02 supported by Shandong Provincial Natural Science Foundation; the Ministry of Education, Youth and Sports of the Czech Republic under Contract

No. LTT17020; the Czech Science Foundation Grant No. 22-18469S; Horizon 2020 ERC Advanced Grant No. 884719 and ERC Starting Grant No. 947006 “InterLeptons” (European Union); the Carl Zeiss Foundation, the Deutsche Forschungsgemeinschaft, the Excellence Cluster Universe, and the VolkswagenStiftung; the Department of Atomic Energy (Project Identification No. RTI 4002) and the Department of Science and Technology of India; the Istituto Nazionale di Fisica Nucleare of Italy; National Research Foundation (NRF) of Korea Grant No. 2016R1D1A1B02012900, No. 2018R1A2B3003643, No. 2018R1A6A1A06024970, No. RS202200197659, No. 2019R1I1A3A01058933, No. 2021R1A6A1A03043957, No. 2021R1F1A1060423, No. 2021R1F1A1064008, No. 2022R1A2C1003993; Radiation Science Research Institute, Foreign Large-size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/GLORIAD; the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation, Agreement 14.W03.31.0026, and the HSE University Basic Research Program, Moscow; University of Tabuk research Grants No. S-1440-0321, No. S-0256-1438, and No. S-0280-1439 (Saudi Arabia); the Slovenian Research Agency Grants No. J1-9124 and No. P1-0135; Ikerbasque, Basque Foundation for Science, Spain; the Swiss National Science Foundation; the Ministry of Education and the Ministry of Science and Technology of Taiwan; and the United States Department of Energy and the National Science Foundation. These acknowledgements are not to be interpreted as an endorsement of any statement made by any of our institutes, funding agencies, governments, or their representatives. We thank the KEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid; and the KEK computer group and the Pacific Northwest National Laboratory (PNNL) Environmental Molecular Sciences Laboratory (EMSL) computing group for strong computing support; and the National Institute of Informatics, and Science Information NETwork 6 (SINET6) for valuable network support.

### APPENDIX A: KINK EVENTS SELECTION IN THE DECAY CHAIN $D^{*+} \rightarrow D^0(\rightarrow K^- \pi^+) \pi^+$

We reconstruct  $D^{*+}$  candidates in the decay chain  $D^{*+} \rightarrow D^0(\rightarrow K^- \pi^+) \pi^+$ . The following selection criteria on the  $K^- \pi^+$  and  $K^- \pi^+ \pi^+$  invariant masses are used:  $1.82 \text{ GeV}/c^2 < M(K^- \pi^+) < 1.9 \text{ GeV}/c^2$  and  $|M(K^- \pi^+ \pi^+) - M(K^- \pi^+) + M^{\text{PDG}}(D^0) - M^{\text{PDG}}(D^{*+})| < 3 \text{ MeV}/c^2$ , providing a large sample of  $D^{*+}$  candidates. The momentum of the  $D^{*+}$  candidates in the c.m. frame is limited at  $p_{D^{*+}} > 2.3 \text{ GeV}/c$  since our MC simulation of  $e^+ e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$  does not reproduce  $K^- \pi^+ \pi^+$  invariant mass distribution

well for both the  $D^{*+}$  signal and combinatorial background. For larger momentum,  $D^{*+}$  are produced in the continuum, and our MC simulation of  $e^+e^- \rightarrow q\bar{q}$  describes the combinatorial background well. However, there is a discrepancy between the data and MC samples in the  $D^{*+}$  peak since the effects of the  $c$ -quark fragmentation were not properly accounted for in the MC simulation. The fragmentation is based mainly on the momentum spectrum; therefore, we reweight the MC sample with a real  $D^{*+} \rightarrow D^0\pi^+$  decay in bins of its momentum. The following procedure is used: the  $M(K^-\pi^+\pi^+)$  distribution in data is fitted in bins of  $p_{D^{*+}}$ , and the number of  $D^{*+}$  mesons is obtained. After that, we determine the weight for the MC event with real  $D^{*+}$  as  $w(p_{D^{*+}}) = N_{D^{*+}}^{\text{data}}(p_{D^{*+}})/N_{D^{*+}}^{\text{MC}}(p_{D^{*+}})$ . We perform this procedure with  $D^{*+}$  candidates reconstructed before any kink selection.

For further event selection, we require one of the  $D^0$  daughter tracks to pass the kink selection criteria described in Sec. IV B. The second track is identified using the information from the CDC, TOF, and ACC combined to form likelihood  $\mathcal{L}_i$  ( $i = \pi$  or  $K$ ). To select the pion (kaon) kink, we require  $\mathcal{R}(K/\pi) = \mathcal{L}_K/(\mathcal{L}_K + \mathcal{L}_\pi) > 0.6$  [ $\mathcal{R}(\pi/K) > 0.6$ ] for  $K^-$  ( $\pi^+$ ) from the  $D^0$  meson.

To illustrate the result of the selection, we plot the  $K^-\pi^+$  invariant mass for pion and kaon kinks in Figs. 13(a) and 13(b), respectively. As can be seen, each sample consists of the corresponding kinks, as well as hadron scattering events.

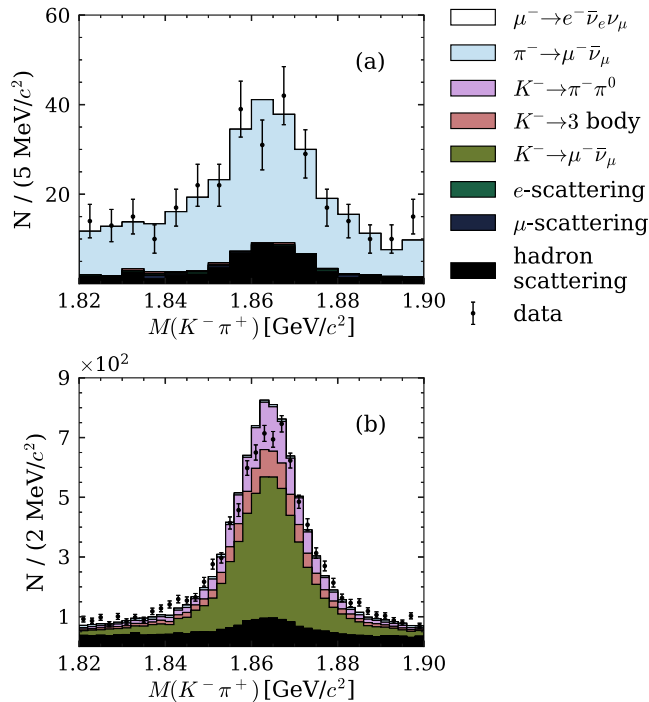


FIG. 13.  $K^-\pi^+$  invariant mass for (a)  $\pi^+$  kink candidates and (b)  $K^-$  kink candidates.

## APPENDIX B: KINK EVENTS SELECTION IN THE $\gamma$ -CONVERSION PROCESS

We select  $\gamma$ -conversion events from 1–1 and 1–3 topology  $\tau^+\tau^-$  pairs sample prepared according to the preselection criteria described in Sec. IV A. Although this preselection limits available statistics, the kink reconstruction efficiency here is similar to one in the main analysis.

The conversion is reconstructed on the one-track side from two oppositely charged tracks. To suppress background from other  $V$ -shaped processes like  $K_S^0$  decay, the invariant mass of  $e^+e^-$  pair  $m_{e^+e^-}$  is required to be less than  $40 \text{ MeV}/c^2$ . Since  $\gamma$ -conversion occurs on the detector material, the radius of the conversion vertex in the  $r\phi$ -plane has to be larger than 2 cm. To suppress a random combination of the tracks, the distance between two tracks in projection onto the  $z$ -axis is required to be less than 5 cm. The daughter electron is reconstructed as a kink with the selection criteria described in Sec. IV B. Finally, using the identification of the daughter positron  $\mathcal{R}(e/x) > 0.8$ , we obtain a clean sample of identified electron scattering events.

To illustrate the result of the described procedure, we plot the  $e^+e^-$ -pair invariant mass in Fig. 14(a) and the radius of the conversion vertex in the  $r\phi$ -plane in Fig. 14(b). The localization of the  $m_{e^+e^-}$  in the zero region is as expected. In the distribution of the  $\gamma$ -conversion vertex, the SVD structure is clearly observed. The selected sample consists of pure electron scattering events.

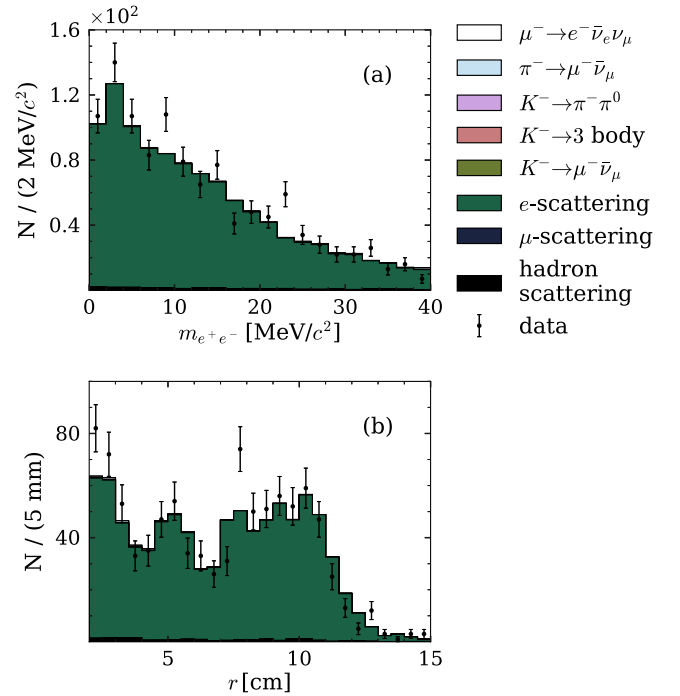


FIG. 14. (a) invariant mass  $m_{e^+e^-}$  and (b) radius of the conversion vertex for the selected  $\gamma$ -conversion events, where one of the electron is reconstructed as an electron scattering kink.

- [1] D. Bryman, V. Cirigliano, A. Crivellin, and G. Inguglia, Testing lepton flavor universality with pion, kaon, tau, and beta decays, *Annu. Rev. Nucl. Part. Sci.* **72**, 69 (2022).
- [2] P. Herczeg, On muon decay in left-right symmetric electroweak models, *Phys. Rev. D* **34**, 3449 (1986).
- [3] M. Krawczyk and D. Temes, Large 2HDM(II) one-loop corrections in leptonic tau decays, *Eur. Phys. J. C* **44**, 435 (2005).
- [4] E. J. Chun and J. Kim, Leptonic precision test of leptophilic two-Higgs-doublet model, *J. High Energy Phys.* **07** (2016) 110.
- [5] J. M. Márquez, G. L. Castro, and P. Roig, Michel parameters in the presence of massive Dirac and Majorana neutrinos, *J. High Energy Phys.* **11** (2022) 117.
- [6] L. Michel, Interaction between four half spin particles and the decay of the  $\mu$  meson, *Proc. Phys. Soc. London Sect. A* **63**, 514 (1950).
- [7] F. Scheck, *Leptons, Hadrons and Nuclei* (North-Holland, Amsterdam, 1983).
- [8] K. Mursula and F. Scheck, Analysis of leptonic charged weak interactions, *Nucl. Phys.* **B253**, 189 (1985).
- [9] W. Fetscher, H. J. Gerber, and K. F. Johnson, Muon decay: Complete determination of the interaction and comparison with the standard model, *Phys. Lett. B* **173**, 102 (1986).
- [10] W. Fetscher and H. J. Gerber, Precision measurements in muon and tau decays, *Adv. Ser. Dir. High Energy Phys.* **14**, 657 (1995).
- [11] R. L. Workman (Particle Data Group), Review of particle physics, *Prog. Theor. Exp. Phys.* **2022**, 083C01 (2022).
- [12] N. Shimizu *et al.* (Belle Collaboration), Measurement of the tau Michel parameters  $\bar{\eta}$  and  $\xi\kappa$  in the radiative leptonic decay  $\tau^- \rightarrow \ell^- \nu_\tau \bar{\nu}_\ell \gamma$ , *Prog. Theor. Exp. Phys.* **2018**, 023C01 (2018).
- [13] D. Bodrov *et al.* (Belle Collaboration), companion Letter, First Measurement of the Michel Parameter  $\xi'$  in the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  Decay at Belle, *Phys. Rev. Lett.* **131**, 021801 (2023).
- [14] W. Fetscher, Leptonic tau decays: How to determine the Lorentz structure of the charged leptonic weak interaction by experiment, *Phys. Rev. D* **42**, 1544 (1990).
- [15] D. A. Bodrov, Measurement of the Michel parameter  $\xi'$  of the leptonic  $\tau$ -decay at the future super charm-tau factory, *Bull. Lebedev Phys. Inst.* **48**, 18 (2021).
- [16] D. A. Bodrov, Measurement of the Michel parameter  $\xi'$  in tau-lepton decays at the super charm-tau factory, *Phys. At. Nucl.* **84**, 212 (2021).
- [17] D. Bodrov and P. Pakhlov, A new method for the measurement of the Michel parameters that describe the daughter muon polarization in the  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  decay, *J. High Energy Phys.* **10** (2022) 035.
- [18] J. H. Kühn, Tau kinematics from impact parameters, *Phys. Lett. B* **313**, 458 (1993).
- [19] J. Brodzicka *et al.* (Belle Collaboration), Physics achievements from the Belle experiment, *Prog. Theor. Exp. Phys.* **2012**, 04D001 (2012).
- [20] A. Abashian *et al.* (Belle Collaboration), The Belle detector, *Nucl. Instrum. Methods Phys. Res., Sect. A* **479**, 117 (2002).
- [21] S. Kurokawa and E. Kikutani, Overview of the KEKB accelerators, *Nucl. Instrum. Methods Phys. Res., Sect. A* **499**, 1 (2003), and other papers included in this Volume.
- [22] T. Abe *et al.*, Achievements of KEKB, *Prog. Theor. Exp. Phys.* **2013**, 03A001 (2013) and references therein.
- [23] H. Hirano *et al.*, A high resolution cylindrical drift chamber for the KEK B factory, *Nucl. Instrum. Methods Phys. Res., Sect. A* **455**, 294 (2000).
- [24] S. Jadach, B. F. L. Ward, and Z. Was, The precision Monte Carlo event generator K K for two fermion final states in  $e^+ e^-$  collisions, *Comput. Phys. Commun.* **130**, 260 (2000).
- [25] Z. Was, TAUOLA the library for tau lepton decay, and KKMC/KORALB/KORALZ/... status report, *Nucl. Phys. B, Proc. Suppl.* **98**, 96 (2001).
- [26] S. Jadach, Z. Was, R. Decker, and J. H. Kühn, The tau decay library TAUOLA: Version 2.4, *Comput. Phys. Commun.* **76**, 361 (1993).
- [27] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zancarini, GEANT3, Report No. CERN-DD-EE-84-1, 1987.
- [28] D. J. Lange, The EvtGen particle decay simulation package, *Nucl. Instrum. Methods Phys. Res., Sect. A* **462**, 152 (2001).
- [29] F. A. Berends, P. H. Daverveldt, and R. Kleiss, Monte Carlo simulation of two photon processes. 2. Complete lowest order calculations for four lepton production processes in electron positron collisions, *Comput. Phys. Commun.* **40**, 285 (1986).
- [30] S. Jadach, E. Richter-Was, B. F. L. Ward, and Z. Was, Monte Carlo program BHLUMI 2.01 for Bhabha scattering at low angles with Yennie-Frautschi-Suura exponentiation, *Comput. Phys. Commun.* **70**, 305 (1992).
- [31] E. Barberio, B. van Eijk, and Z. Was, PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays, *Comput. Phys. Commun.* **66**, 115 (1991).
- [32] Y. Freund and R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* **55**, 119 (1997).
- [33] A. Hocker *et al.*, TMVA—Toolkit for multivariate data analysis, *arXiv:physics/0703039*.
- [34] E. Nakano, Belle PID, *Nucl. Instrum. Methods Phys. Res., Sect. A* **494**, 402 (2002).
- [35] E. Kou *et al.* (Belle-II Collaboration), The Belle II physics book, *Prog. Theor. Exp. Phys.* **2019**, 123C01 (2019); **2020**, 029201(E) (2020).