
























































































































































































































































































































































































PREPARED FOR SUBMISSION TO JHEP

Search for lepton flavor-violating decay modes $B^0 \rightarrow K^{*0} \tau^\pm \ell^\mp$ ($\ell = e, \mu$) with hadronic B-tagging at Belle and Belle II

The Belle and Belle II Collaborations

I. Adachi , Y. Ahn , H. Aihara , N. Akopov , S. Alghamdi , M. Alhakami ,
A. Aloisio , 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 , J. Baudot , A. Baur , A. Beaubien , F. Becherer ,
J. Becker , J. V. Bennett , F. U. Bernlochner , V. Bertacchi , M. Bertemes ,
E. Bertholet , M. Bessner , S. Bettarini , B. Bhuyan , F. Bianchi , T. Bilka ,
D. Biswas , A. Bobrov , D. Bodrov , J. Borah , A. Boschetti , A. Bozek ,
M. Bračko , P. Branchini , R. A. Briere , T. E. Browder , A. Budano ,
S. Bussino , M. Campajola , L. Cao , G. Casarosa , C. Cecchi , M.-C. Chang ,
P. Cheema , B. G. Cheon , K. Chilikin , J. Chin , H.-E. Cho , K. Cho ,
S.-J. Cho , S.-K. Choi , S. Choudhury , I. Consigny , L. Corona , J. X. Cui ,
E. De La Cruz-Burelo , S. A. De La Motte , G. de Marino , G. De Pietro ,
R. de Sangro , M. Destefanis , A. Di Canto , J. Dingfelder , Z. Doležal ,
I. Domínguez Jiménez , T. V. Dong , X. Dong , M. Dorigo , G. Dujany ,
P. Ecker , D. Epifanov , R. Farkas , P. Feichtinger , T. Ferber , T. Fillinger ,
C. Finck , G. Finocchiaro , F. Forti , A. Frey , B. G. Fulsom , A. Gale ,
E. Ganiev , M. Garcia-Hernandez , R. Garg , G. Gaudino , V. Gaur ,
V. Gautam , A. Gaz , A. Gellrich , D. Ghosh , H. Ghumaryan ,
G. Giakoustidis , R. Giordano , A. Giri , P. Gironella Gironell , B. Gobbo ,
R. Godang , O. Gogota , P. Goldenzweig , W. Gradl , E. Graziani ,
D. Greenwald , K. Gudkova , I. Haide , H. Hayashii , S. Hazra , C. Hearty ,
M. T. Hedges , A. Heidelberg , G. Heine , I. Heredia de la Cruz ,
M. Hernández Villanueva , T. Higuchi , M. Hoek , M. Hohmann , P. Horak ,
C.-L. Hsu , T. Humair , T. Iijima , K. Inami , G. Inguglia , N. Ipsita ,
A. Ishikawa , R. Itoh , M. Iwasaki , P. Jackson , D. Jacobi , W. W. Jacobs ,
E.-J. Jang , Q. P. Ji , S. Jia , Y. Jin , A. Johnson , M. Kaleta , J. Kandra ,
K. H. Kang , G. Karyan , F. Keil , C. Ketter , M. Khan , C. Kiesling ,
D. Y. Kim , J.-Y. Kim , K.-H. Kim , K. Kinoshita , P. Kodyš , T. Koga 

S. Kohani , K. Kojima , A. Korobov , S. Korpar , E. Kovalenko ,
 R. Kowalewski , P. Križan , P. Krovovny , D. Kumar , R. Kumar , K. Kumara ,
 T. Kunigo , A. Kuzmin , Y.-J. Kwon , K. Lalwani , T. Lam , T. S. Lau ,
 M. Laurenza , R. Leboucher , F. R. Le Diberder , M. J. Lee , C. Lemettais ,
 P. Leo , P. M. Lewis , C. Li , H.-J. Li , L. K. Li , 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. Liu , D. Liventsev , S. Longo , T. Lueck , C. Lyu , C. Madaan ,
 M. Maggiora , S. P. Maharana , R. Maiti , G. Mancinelli , R. Manfredi ,
 E. Manoni , M. Mantovano , D. Marcantonio , S. Marcello , C. Marinas ,
 C. Martellini , A. Martens , T. Martinov , L. Massaccesi , M. Masuda ,
 K. Matsuoka , D. Matvienko , M. Maushart , J. A. McKenna ,
 Z. Mediankin Gruberová , F. Meier , D. Meleshko , M. Merola , C. Miller ,
 M. Mirra , S. Mitra , K. Miyabayashi , R. Mizuk , G. B. Mohanty ,
 S. Moneta , H.-G. Moser , R. Mussa , I. Nakamura , M. Nakao ,
 Y. Nakazawa , M. Naruki , Z. Natkaniec , A. Natochii , M. Nayak , M. Neu ,
 S. Nishida , R. Okubo , H. Ono , G. Pakhlova , S. Pardi , K. Parham ,
 H. Park , J. Park , S.-H. Park , A. Passeri , S. Patra , S. Paul , R. Pestotnik ,
 L. E. Piilonen , T. Podobnik , A. Prakash , C. Praz , S. Prell , E. Prencipe ,
 M. T. Prim , I. Prudiiev , 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 , S. Sandilya , L. Santelj , C. Santos ,
 V. Savinov , B. Scavino , C. Schmitt , J. Schmitz , S. Schneider , G. Schnell ,
 M. Schnepf , K. Schoenning , C. Schwanda , Y. Seino , A. Selce , K. Senyo ,
 J. Serrano , M. E. Sevier , C. Sfienti , W. Shan , X. D. Shi , T. Shillington ,
 J.-G. Shiu , D. Shtol , B. Shwartz , A. Sibidanov , F. Simon , J. Skorupa ,
 R. J. Sobie , M. Sobotzik , A. Soffer , A. Sokolov , E. Solovieva , S. Spataro ,
 B. Spruck , M. Starič , P. Stavroulakis , S. Stefkova , R. Stroili , Y. Sue ,
 M. Sumihama , H. Svidras , M. Takizawa , K. Tanida , F. Tenchini , F. Testa ,
 O. Tittel , R. Tiwary , E. Torassa , K. Trabelsi , F. F. Trantou , I. Tsaklidis ,
 M. Uchida , I. Ueda , T. Uglov , K. Unger , Y. Unno , K. Uno , S. Uno ,
 P. Urquijo , Y. Ushiroda , R. van Tonder , K. E. Varvell , M. Veronesi ,
 A. Vinokurova , V. S. Vismaya , L. Vitale , V. Vobbilisetti , R. Volpe ,
 A. Vossen , S. Wallner , M.-Z. Wang , A. Warburton , M. Watanabe ,
 S. Watanuki , C. Wessel , B. D. Yabsley , S. Yamada , W. Yan , S. B. Yang ,
 K. Yi , J. H. Yin , K. Yoshihara , J. Yuan , Y. Yusa , L. Zani , M. Zeyrek ,
 B. Zhang , V. Zhilich , J. S. Zhou , Q. D. Zhou , L. Zhu , R. Žlebčík 

ABSTRACT: We present the results of a search for the charged-lepton-flavor violating decays $B^0 \rightarrow K^{*0} \tau^\pm \ell^\mp$, where ℓ^\mp is either an electron or a muon. The results are based on 365 fb⁻¹ and 711 fb⁻¹ datasets collected with the Belle II and Belle detectors, respectively. We use an exclusive hadronic B -tagging technique, and search for the signal decay in the system recoiling against the fully reconstructed B meson. We find no evidence for $B^0 \rightarrow K^{*0} \tau^\pm \ell^\mp$ decays and set upper limits on the branching fractions in the range of $(2.9\text{--}6.4) \times 10^{-5}$ at 90% confidence level.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | The Belle and Belle II detectors, simulation and data samples | 2 |
| 3 | Event selection and background rejection | 3 |
| 4 | Branching fraction measurement | 6 |
| 5 | Systematic uncertainties | 8 |
| 6 | Results and conclusion | 11 |

1 Introduction

In the standard model (SM) of particle physics charged lepton flavor violating (LFV) processes are highly suppressed. They can occur only via neutrino-mixing with rates far below the current and foreseen experimental reach [1]. However, in many extensions beyond the standard model (BSM), LFV decays are enhanced since they are not protected by any fundamental symmetry (see Ref. [2] for a recent review). An observation of such decays would thus provide indisputable evidence of physics beyond the SM. In addition, in the SM, lepton flavor universality (LFU) holds, meaning that the interaction of the three generations of leptons with the gauge bosons is identical, except for differences arising from their masses [3, 4]. The recent anomalies observed in $b \rightarrow c\tau\nu$ transitions [5, 6] may nevertheless suggest deviation from LFU, which also imply LFV in many BSM scenarios. For instance, a heavy Z' boson mediator or a leptoquark-mediated transition would produce LFV [7–9]. Measurements of $b \rightarrow s$ transitions can also be used to investigate LFU deviation and LFV [10–13]. In particular, the recent excess observed by Belle II in $b \rightarrow s\nu\bar{\nu}$ transitions [14] can be described by allowing LFV [15], which could give an enhancement of the branching fractions of $B \rightarrow K\tau^\pm\ell^\pm$ up to 3×10^{-6} , where ℓ stands for e or μ . This is close to the current experimental limits and motivates further searches.

The LHCb experiment searched for $B^0 \rightarrow K^{*0}\tau^-\mu^+$ ($B^0 \rightarrow K^{*0}\tau^+\mu^-$) decays using a 9 fb^{-1} dataset [16] and obtained upper limits of $0.82 \text{ (} 1.0 \text{)} \times 10^{-5}$ at the 90% confidence level (C.L.). Searches for $B^+ \rightarrow K^+\tau^\pm\ell^\mp$ have been also performed by the BaBar [17] and Belle [18] experiments, setting the best upper limits in these modes between 0.6 and 2.5×10^{-5} at 90% C.L. Here, we present the first search for $B^0 \rightarrow K^{*0}\tau^\pm e^\mp$ decays and the first search for $B^0 \rightarrow K^{*0}\tau^\pm\mu^\mp$ decays at a B factory, using the combined dataset of the Belle and Belle II experiments, with integrated luminosities of 711 fb^{-1} and 365 fb^{-1} , respectively adding up to a total of 1076 fb^{-1} . Four different final states are distinguished according to the flavor of the final state lepton ℓ , and to the sign of its charge with respect to that of

the kaon from the K^{*0} : same-sign $SS\ell$ for $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)\tau^-\ell^+$ and opposite-sign $OS\ell$ for $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)\tau^+\ell^-$. Charge conjugated final states are implied through this paper.

We use a B -tagging technique to reconstruct one B meson decaying into hadronic modes and search for the signal in the system recoiling against the $K^{*0}\ell$ from the other B meson of the pair and the fully reconstructed tagged B . Combinatorial and some targeted backgrounds are rejected using a cut-based approach followed by a Boosted Decision Tree (BDT). Finally the signal is extracted from a simultaneous fit to the τ lepton recoil mass in the Belle and Belle II datasets.

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

The Belle II experiment is located at SuperKEKB [19, 20], an accelerator colliding electrons and positrons with center-of-mass energies near the $\Upsilon(4S)$ resonance. The Belle II detector [21] has a cylindrical geometry surrounding the interaction point and includes a two-layer silicon-pixel detector (PXD) surrounded by a four-layer double-sided silicon-strip detector (SVD) [22] and a 56-layer central drift chamber (CDC). These detectors reconstruct the trajectories (tracks) of charged particles and measure energy loss due to ionization in the material of the detector. Only one sixth of the second layer of the PXD had been installed for the data analyzed here. Surrounding the CDC are a time-of-propagation detector (TOP) [23] in the central region and an aerogel-based ring-imaging Cherenkov detector (ARICH) in the forward region. These detectors provide information used to identify charged particles. Surrounding the TOP and ARICH is an electromagnetic calorimeter (ECL) based on CsI(Tl) crystals providing energy and timing measurements, primarily for photons and electrons. Outside the ECL is a superconducting solenoid magnet that provides a 1.5 T axial magnetic field. The magnetic flux is returned via an iron yoke, which serves the dual purpose of also being instrumented with resistive-plate chambers and plastic scintillator modules (KLM) to detect muons, K_L^0 mesons, and neutrons. The symmetry axis of the magnet, which almost coincides with the direction of the electron beam, defines the z axis.

The Belle detector was located at the interaction point of the KEKB collider [24, 25]. It shares a similar structure to Belle II, but lacks a silicon pixel detector and plastic scintillators in the KLM, uses aerogel threshold Cherenkov counters (ACC) and a barrel-like arrangement of time-of-flight scintillation counters (TOF) for particle identification. Vertexing and tracking are performed using the Belle SVD and CDC. A detailed description of the Belle detector can be found in Ref. [26, 27].

This analysis uses the 711 fb^{-1} Belle dataset corresponding to $(771.6 \pm 10.6) \times 10^6$ $\Upsilon(4S)$ events, and the dataset collected by Belle II during the first data taking period, corresponding to 365 fb^{-1} or $(387.1 \pm 5.6) \times 10^6$ $\Upsilon(4S)$ events.

Monte Carlo (MC) simulated events are used to optimize the signal selection, to improve background rejection, to model the signal and measure its efficiency as well as to estimate the systematic uncertainties. The signal $B^0 \rightarrow K^{*0}\tau^\pm\ell^\mp$ channels are modeled using a uniform three-body phase space model; 20 million events are produced for each de-

68 cay mode. Simulated samples that reproduce the composition of Belle and Belle II events,
 69 including $B\bar{B}$ and $e^+e^- \rightarrow q\bar{q}$ continuum (where q indicates an u , d , s or c quark) back-
 70 grounds, and equivalent to approximately four times the data luminosity are used to inves-
 71 tigate the sample composition and validate the analysis before examining the signal region
 72 in data. Simulated events are generated using the KKMC generator for quark-antiquark pro-
 73 duction from e^+e^- collisions [28], the PYTHIA8 (PYTHIA6 for Belle) generator for hadroniza-
 74 tion [29, 30], the EvtGen software package and the PYTHIA generator for the decay of the
 75 generated hadrons [31], the PHOTOS package for the final state radiation (FSR) [32] and
 76 the Geant4 (Geant3 for Belle) software package for the detector response [33, 34]. The
 77 simulation includes beam-induced background data overlay [35]. The data and the MC
 78 simulations are processed using the Belle II analysis software (basf2) [36, 37] and Belle
 79 data and MC are converted from the Belle analysis software (basf) [38] format into the
 80 Belle II format for basf2 compatibility using the B2BII framework [39].

81 **3 Event selection and background rejection**

82 Events are selected by a hardware trigger targeting $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ events, based on the
 83 charged-particle multiplicity and total energy, in order to suppress low-multiplicity events
 84 and beam-related background. We reconstruct the selected events using a B -tagging ap-
 85 proach. We reconstruct one of the B meson in the pair, called B_{tag} , in exclusive hadronic
 86 decays, using the hadronic Full Event Interpretation (FEI) B -tagging algorithm [40]. FEI is
 87 a machine-learning based algorithm developed for B -tagged analyses at Belle and Belle II.
 88 It reconstructs B meson candidates from exclusive decays, using a hierarchical approach
 89 starting from reconstructed charged and neutral particles in the detector. We then recon-
 90 struct the second B meson of the event, called B_{sig} , in our signal channel $B^0 \rightarrow K^{*0}\tau^\pm\ell^\mp$
 91 from the tracks left after B_{tag} reconstruction. Since the B_{sig} final state contains at least
 92 one neutrino coming from the τ decay, its kinematics cannot be fully determined. How-
 93 ever, the B_{sig} can be constrained by exploiting the knowledge of the $\Upsilon(4S)$ initial state
 94 and information on the fully-reconstructed B_{tag} .

95 The B_{tag} candidates are selected by requiring at least three tracks, three ECL energy
 96 deposits (clusters) in the event, and a visible energy in the center-of-mass frame of at least
 97 4 GeV. These tracks are required to have a transverse impact parameter $d_0 < 0.5$ cm, a
 98 longitudinal impact parameter $|z_0| < 2$ cm and a transverse momentum $p_T > 0.1$ GeV/ c .
 99 The clusters are required to be in the angular acceptance of the CDC (polar angle from 17°
 100 to 150°) with energies larger than 0.1 GeV. The B_{tag} must have a beam-energy-constrained
 101 mass $M_{\text{bc}} = \sqrt{(E_{\text{beam}}/c^2)^2 - (p_{B_{\text{tag}}}/c)^2}$ larger than 5.27 GeV/ c^2 and an energy difference
 102 $\Delta E = E_{B_{\text{tag}}} - E_{\text{beam}}$ in the range $-0.15 < \Delta E < 0.1$ GeV. Here, E_{beam} , $E_{B_{\text{tag}}}$ and $p_{B_{\text{tag}}}$ are
 103 the beam energy, and the energy and momentum of the B_{tag} candidate in the e^+e^- center
 104 of mass frame. Each reconstructed B_{tag} is assigned a multivariate classifier output, \mathcal{P}_{FEI} ,
 105 ranging from zero to one and corresponding to candidates identified as background-like and
 106 signal-like respectively. We require the B_{tag} candidate to have $\mathcal{P}_{\text{FEI}} > 0.001$. If multiple
 107 B_{tag} candidates are reconstructed in an event, the one with the highest \mathcal{P}_{FEI} is retained.
 108 After this selection, the B_{tag} purity is approximately 45% (40%) for Belle II (Belle).

109 The B_{sig} candidates are selected by combining three tracks, corresponding to a recon-
 110 structed $K^{*0}(K^+\pi^-)$ with a charged lepton (e, μ). The τ lepton is not explicitly recon-
 111 structed, but up to three additional tracks are permitted to be present in the event. The
 112 kaon, pion and lepton candidates are identified as tracks not used in the B_{tag} reconstruc-
 113 tion, with $d_0 < 0.5$ cm, $|z_0| < 5$ cm, in the CDC angular acceptance, and for hadrons only,
 114 to have at least 20 hits in the CDC (in the following we refer to tracks satisfying these
 115 requirements as “good tracks”). For both Belle and Belle II, kaons and pions are identified
 116 requiring an identification likelihood ratio $\mathcal{P}_K, \mathcal{P}_\pi > 0.6$ while electrons and muons can-
 117 didates must satisfy $\mathcal{P}_e, \mathcal{P}_\mu > 0.9$. The hadron identification likelihood uses information
 118 from the ACC, CDC, and TOF for Belle. For Belle II, information from all subdetectors
 119 except the PXD and SVD is used, resulting in a kaon identification efficiency of 85% (88%
 120 for Belle) for a pion fake rate of 5% (8% for Belle) and a pion identification efficiency of
 121 90% (91% for Belle) for a kaon fake rate of 7% (6% for Belle) at the particle identification
 122 working points $\mathcal{P}_K, \mathcal{P}_\pi > 0.6$. For Belle, electrons are identified using the information from
 123 the ECL, CDC and ACC, and information from the KLM only is used for muon identifi-
 124 cation. Using the likelihood ratio requirement $\mathcal{P}_e(\mathcal{P}_\mu) > 0.9$, the lepton identification has
 125 an efficiency of 92% (89%) and a pion fake rate of 0.3% (1.4%) for electrons (muons with
 126 momentum larger than 0.6 GeV/c). For Belle II, the electron identification uses a boosted-
 127 decision-tree (BDT) classifier trained with information from all sub-detectors except the
 128 PXD and SVD and the muon identification uses information from all sub-detectors except
 129 the PXD and SVD. At the particle identification working point $\mathcal{P}_e(\mathcal{P}_\mu) > 0.9$, the electron
 130 (muon) identification has an efficiency of 86% (89%) and a pion mis-identification rate of
 131 0.4% (7%).

132 To recover electron candidates with bremsstrahlung, we accept photons having min-
 133 imum energies of 50 MeV within a 50 mrad angle of an electron track. We reconstruct
 134 $K^{*0} \rightarrow K^+\pi^-$ candidates combining a kaon and a pion of opposite charge. The kaon,
 135 pion and lepton are fitted to a common vertex, and their kinematic information is updated
 136 according to the fit result. The K^{*0} candidate should have an invariant mass in the range
 137 $0.842 < M(K^+\pi^-) < 0.942$ GeV/ c^2 . The presence of a τ lepton is inferred from the
 138 presence of a single good track t_τ with a charge opposite to that of the primary lepton.
 139 This track is not used in the signal kinematic reconstruction and is only used to reduce the
 140 background contamination.

141 After the event is properly reconstructed we can proceed to the signal selection. In
 142 order to reject background, we use properties of the rest-of-event (ROE) that correspond
 143 to good tracks and photons not used in the reconstruction of the B_{sig} (i.e. $K^{*0}\ell$ system)
 144 candidate, the B_{tag} and the t_τ track. Photons are reconstructed from ECL clusters within
 145 the CDC acceptance and not associated with any tracks. Photon candidates must satisfy
 146 additional requirements, described in Ref. [41], to reduce photons from beam background.
 147 We select events with at most two tracks in the ROE, in order to retain 3-prong decays of
 148 the τ lepton, or potential signal candidate in which the B_{tag} is not correctly reconstructed,
 149 leading to partner- B daughters tracks falling in the ROE. For Belle, we also require the
 150 total charge of the ROE to be zero. This requirement is not applied to Belle II due to its
 151 smaller data sample, to preserve enough events for a BDT training and the fit. At this

stage, multiple candidates due to the different possible combinations of tracks reconstructed as t_τ or ROE tracks are kept. Finally, to reduce contamination from $e^+e^- \rightarrow q\bar{q}$ events, we require an event sphericity larger than 0.2 [42] and the absolute value of cosine of the angle between the thrust axis of B_{tag} and that of particles not used in B_{tag} to be less than 0.9.

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

$$\vec{p}_\tau = -\vec{p}_{B_{\text{tag}}} - \vec{p}_{K^*0} - \vec{p}_\ell, \quad E_\tau = E_{\text{beam}} - E_{K^*0} - E_\ell. \quad (3.1)$$

The τ mass is then reconstructed as

$$M_\tau = [m_B^2 + M^2(K^*0\ell) - 2(E_{\text{beam}}E_{K^*0\ell} + |\vec{p}_{B_{\text{tag}}}||\vec{p}_{K^*0\ell}|\cos\theta)]^{1/2}. \quad (3.2)$$

Here, m_B is the known B^0 mass [43]; $M(K^*0\ell)$, $E_{K^*0\ell}$, $\vec{p}_{K^*0\ell}$ are the mass, energy, and momentum of the system composed of the K^*0 and ℓ , respectively; θ is the angle between $\vec{p}_{B_{\text{tag}}}$ and $\vec{p}_{K^*0\ell}$. All the above quantities are defined with respect to the center-of-mass frame. Candidates having a reconstructed M_τ outside the range $[1.0, 2.5]$ GeV/ c^2 are discarded and the signal region is defined as $[1.65, 1.90]$ GeV/ c^2 . To avoid biases, we do not examine the signal region until the analysis strategy is fixed.

Background can arise from $B^0 \rightarrow K^*0 J/\psi$ decays when the J/ψ products are reconstructed as the ℓt_τ pair. They are removed requiring the ℓt_τ invariant mass to be outside the range $[3.05, 3.15]$ GeV/ c^2 . The background for $SS\ell$ modes ($B^0 \rightarrow K^*0 \tau^- \ell^+$) is mainly due to semileptonic B decays such as $B \rightarrow D\ell\nu$, with $D \rightarrow K\pi\pi$. These events are vetoed by removing candidates with $M(K^*0 t_\tau)$ in the range $[1.83, 1.90]$ GeV/ c^2 . The main backgrounds in $OS\ell$ modes ($B^0 \rightarrow K^*0 \tau^+ \ell^-$) are from $B \rightarrow DX$ decays where the D meson decays semileptonically as $D \rightarrow K^*0 \ell\nu$. Hadronic $B \rightarrow DX$ decays with $D \rightarrow K\pi\pi$ can be reconstructed as $OS\mu$ signal if a pion is misreconstructed as a muon. A veto is thus applied for candidates with $M(K^*0\ell)$ in the range $[1.83, 1.90]$ GeV/ c^2 . In addition, according to the simulation, $q\bar{q}$ processes amount for 7 to 24% of the background in $SS\ell$ modes and 24 to 56% of the background in $OS\ell$ modes after applying the above selection criteria. For each mode, there is a slightly larger $q\bar{q}$ contribution in Belle II than in Belle.

To reduce the remaining background, eight BDTs are trained separately for each signal mode, and for Belle and Belle II data using the **fastBDT** library [40]. The training is performed using simulated signal and $q\bar{q}$ and $B\bar{B}$ processes for the background. For each BDT, between twelve and fourteen input variables are chosen from a common set of fifteen variables, removing the ones that do not improve the performance. The variables comprise quantities related to the signal B , with the $M(K^*0 t_\tau)$ and $M(K^*0\ell)$ invariant masses, the energy of the lepton and track from τ , the χ^2 probability of the $K^*0\ell$ vertex and its distance with respect to the interaction point in the transverse plane. The BDTs also use event shape variables such as the sphericity and the modified Fox-Wolfram moments [44] to suppress the $q\bar{q}$ background. In addition, quantities characterizing the ROE are also considered: the number of tracks and clusters, the total cluster energy and the ROE momentum. The

requirements applied to the BDT outputs are optimized using the figure of merit defined as $\frac{\epsilon}{\alpha/2 + \sqrt{N}}$ [45], where ϵ is the signal efficiency and N is the number of background events evaluated in the M_τ range $[1.70, 1.85]$ GeV/ c^2 , which corresponds to approximately 90% signal coverage. We set α to 3, corresponding to 3σ signal significance.

After the BDT selection the average multiplicity is 1.1 in both MC and data samples. In events with more than one candidate, one is selected randomly.

The final reconstruction and selection signal efficiencies are given in Table 1, while the two-dimensional final efficiencies as function of the squared invariant masses $M^2(K^{*0}t_\tau)$ and $M^2(K^{*0}\ell)$ are shown in Fig. 1. A uniform Dalitz distribution of the $K^{*0}\tau\ell$ system is assumed in signal generation. The efficiencies are determined on simulated samples taking into account known mismodeling affecting the particle identification and B_{tag} reconstruction, detailed in Sec. 5. The efficiency difference between Belle and Belle II is mainly due to the BDT selection. In average, the optimal cut leads to a higher background rejection and thus a lower efficiency for Belle compared to Belle II. The differences in the efficiency maps between $OS\ell$ and $SS\ell$ modes are due to the different background composition. In particular, $OS\ell$ modes are polluted by $B \rightarrow DX$ background where the D decays semileptonically, which peaks at low values of $M(K^{*0}\ell)$. Those events are thus suppressed by the BDT, giving a lower efficiency in that region for $OS\ell$ final states.

Table 1. Final signal efficiencies after all selection described in sec. 3 for each signal mode and experiment. The values contain the corrections developed to take into account known data mismodeling that affect the particle identification and the B_{tag} reconstruction. All values are in percent.

| | OSe | SSe | $OS\mu$ | $SS\mu$ |
|----------|-------|-------|---------|---------|
| Belle | 0.046 | 0.038 | 0.052 | 0.024 |
| Belle II | 0.075 | 0.056 | 0.060 | 0.051 |

4 Branching fraction measurement

The signal branching fractions are obtained from unbinned extended maximum likelihood fits to the recoil M_τ distributions in the defined fit range $[1.3, 2.3]$ GeV/ c^2 , simultaneously for Belle and Belle II data. For each channel, the likelihood is expressed as

$$\mathcal{L}(M_\tau) = \frac{e^{-(n_{\text{sig}} + n_{\text{bg}})}}{N!} \prod_{i=1}^N (n_{\text{sig}} \cdot \mathcal{P}_{\text{sig}}(M_\tau^i) + n_{\text{bg}} \cdot \mathcal{P}_{\text{bg}}(M_\tau^i)) \cdot \prod_s \text{Gauss}(s, \sigma_s), \quad (4.1)$$

where $n_{\text{sig}}(n_{\text{bg}})$ and $\mathcal{P}_{\text{sig}}(\mathcal{P}_{\text{bg}})$ are the number of events and probability density function (PDF) for the signal (background) and N is the total number of events. The Gaussian terms account for the systematic uncertainties σ_s on the sources s detailed in sec. 5. One likelihood is defined per data sample Belle and Belle II. We express the number of signal events in the dataset $\text{exp} = \text{Belle}, \text{Belle II}$ as:

$$n_{\text{sig}}^{\text{exp}} = \mathcal{B}(B^0 \rightarrow K^{*0}\tau\ell) \times 2 \times \epsilon^{\text{exp}} \times f^{00} \times N_{\Upsilon(4S)}^{\text{exp}} \times \mathcal{B}(K^{*0} \rightarrow K^+\pi^-), \quad (4.2)$$

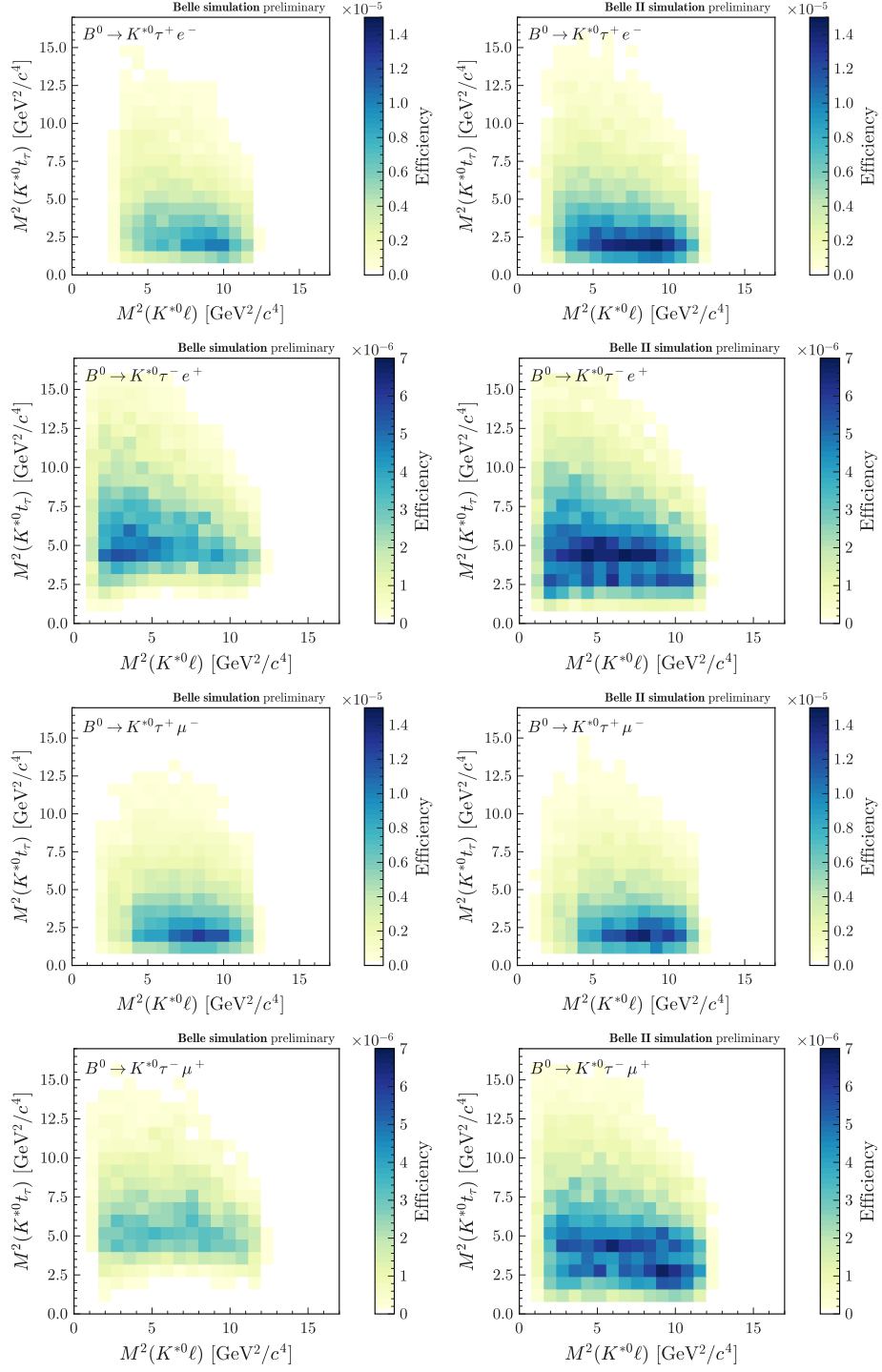


Figure 1. Final signal efficiencies after all selection described in sec. 3 as a function of the kinematic variables $M^2(K^{*0}\ell)$ and $M^2(K^{*0}t_\tau)$ for Belle (left) and Belle II (right). From top to bottom: OSe , SSe , $OS\mu$, $SS\mu$.

221 where ϵ^{exp} is the signal efficiency given in Table 1, $f^{00} = 0.4861 \pm 0.0080$ the fraction of
 222 $\Upsilon(4S)$ decaying into $B^0\bar{B}^0$ pairs [6] and $N_{\Upsilon(4S)}^{\text{exp}}$ is the number of produced $\Upsilon(4S)$ mesons,

223 $N_{\Upsilon(4S)}^{\text{Belle}} = (771.6 \pm 10.6) \times 10^6$ and $N_{\Upsilon(4S)}^{\text{Belle II}} = (387.1 \pm 5.6) \times 10^6$ and $\mathcal{B}(K^{*0} \rightarrow K^+ \pi^-) = 2/3$
 224 assuming isospin conservation. The signal branching fraction $\mathcal{B}(B^0 \rightarrow K^{*0} \tau \ell)$ is a common
 225 parameter of the simultaneous fit.

226 The signal PDF is modeled using a Johnson function [46], defined as

$$227 \quad \mathcal{P}_{\text{sig}}(M_\tau, \mu, \lambda, \gamma, \delta) = \frac{\delta}{\lambda \sqrt{2\pi}} \frac{1}{\sqrt{1 + (\frac{M_\tau - \mu}{\lambda})^2}} \exp \left[-\frac{1}{2} \left(\gamma + \delta \sinh^{-1} \left(\frac{M_\tau - \mu}{\lambda} \right) \right)^2 \right] \quad (4.3)$$

228 where μ is the mean of the Gaussian component, λ its width, γ the distortion of the
 229 distribution to the left/right, and δ the strength of the Gaussian-like component. The
 230 parameters describing the signal shape are fixed to the values obtained from the fit to the
 231 simulated samples. Background events have a smooth distribution in the M_τ fit region and
 232 are analytically modeled using a second-order polynomial whose coefficients are left free
 233 to vary. To validate the fit before the unblinding of the signal region, pseudo-experiments
 234 generated from the data sidebands $M_\tau \in [1.30, 1.65] \cup [1.90, 2.30] \text{ GeV}/c^2$ are used. Signal
 235 is injected with various branching fraction values and the dataset is fitted with the PDFs
 236 described above. No biases are found in these studies.

237 5 Systematic uncertainties

238 There are systematic uncertainties in the branching fraction measurements due to the
 239 determination of the signal efficiencies, the PDF modeling and the external inputs.

240 The efficiency of the requirement on \mathcal{P}_{FEI} is calibrated using the control channel $B^0 \rightarrow$
 241 $D^- \pi^+$. After reconstructing the B_{tag} and the charged pion, we search for the D meson
 242 resonance in the recoil mass of the $B_{\text{tag}} \pi$ system. Calibration factors, defined as weights
 243 to account for data/MC differences, are obtained by comparing the yields in data and
 244 simulation, for each B_{tag} decay mode. In Belle II, the inclusive semileptonic decay $B \rightarrow$
 245 $X \ell \nu$ is also used, and the calibration factors are combined with those from the hadronic
 246 control channel [47]. The signal efficiencies are corrected using those calibration factors
 247 and their associated uncertainties are taken into account as systematic uncertainties.

248 We take into account the systematic uncertainty associated with the corrections to the
 249 simulated lepton-identification efficiencies, derived from auxiliary measurements in data
 250 using $J/\psi \rightarrow \mu^+ \mu^-$, $e^+ e^- \rightarrow \ell^+ \ell^- \gamma$, and $e^+ e^- \rightarrow e^+ e^- \mu^+ \mu^-$ events. These corrections
 251 are obtained as functions of track momentum, polar angle and charge, and applied to
 252 events reconstructed from simulation. The systematic uncertainty is obtained by varying
 253 the correction within their uncertainties and estimating the impact of these variations on
 254 the selection efficiency. A similar method is employed for systematic uncertainty due to
 255 hadron identification, using the $D^{*+} \rightarrow D^0 (\rightarrow K^- \pi^+) \pi^+$ decays.

256 The efficiency of the requirements on the BDT outputs is evaluated using the $B^0 \rightarrow$
 257 $D^- D_s^+ (\rightarrow K^{*0} K^+ / \phi \pi^+)$ control sample, with $K^{*0} \rightarrow K^+ \pi^-$ and $\phi \rightarrow K^+ K^-$. Here,
 258 the B_{tag} is reconstructed in a hadronic channel using the FEI algorithm and the D_s^+ is
 259 used as a proxy for the $K^{*0} \ell$ system. The t_τ track is obtained by selecting a random
 260 track from the D^- with the correct charge, while other D^- decay products are treated

261 as the missing energy. The selection criteria applied to the control channel are identical
 262 to the signal channels when relevant (B_{tag} , hadron identification, K^{*0} mass window, ROE
 263 selection). In addition, the $K^{*0}K/\phi\pi$ system is selected to be within 20 MeV/ c^2 of the
 264 D_s mass, and in case of $D_s \rightarrow \phi\pi$, the $\phi(\rightarrow KK)$ invariant mass should be within 20
 265 MeV/ c^2 around the ϕ mass. The recoil mass of the D^- meson is evaluated in the same
 266 way as M_τ for the signal channels. The D^- recoil mass distributions are shown in Fig. 2
 267 for simulation and data for the two datasets. A component corresponding to the D^{*-} is
 268 also clearly visible. The uncertainty related to the BDT requirements is obtained by fitting
 269 the D^- and D^{*-} yields using a Johnson PDF for the signal and a second-order polynomial
 270 for the background, before and after applying the BDT requirement. The parameters of
 271 the Johnson function are fixed to values from fits to the simulation while the background
 272 parameters are allowed to float. Since the control sample has different properties than
 273 the signal LFV channels, especially for the ROE, the BDT distributions are expected to
 274 differ. The BDT corresponding to each signal channel is applied to the simulated control
 275 channel, and the requirement on the BDT output is set such that its efficiency be the same
 276 as the efficiency of the optimized BDT requirement on the nominal channel. The BDT is
 277 also applied on a data control sample and events are selected according to the requirement
 278 on the BDT score determined for the control channel. The data to MC efficiency ratio
 279 $\mathcal{R}_{\epsilon_{\text{BDT}}}^{\text{data/MC}}$ is measured and the assigned uncertainty is symmetrized so that it covers 68.3%
 280 of the area of a Gaussian function with mean $1 - \mathcal{R}_{\epsilon_{\text{BDT}}}^{\text{data/MC}}$ and standard deviation equal to
 281 the statistical uncertainty on the ratio. The corresponding uncertainties on the efficiencies
 282 are within 18% and 34%. Since the data/MC efficiency ratios are compatible with one, no
 283 correction is applied to the efficiency and only the systematic uncertainty is considered.

284 The difference between data and simulation in the track-reconstruction efficiency in
 285 Belle II is measured in $e^+e^- \rightarrow \tau^+\tau^-$ events with $\tau^- \rightarrow e^-\bar{\nu}_e\nu_\tau$ and $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ to
 286 yield a 0.24% uncertainty per track. For Belle, a 0.35% per-track uncertainty is assigned
 287 using a control sample of $D^{*+} \rightarrow \pi^+D^0$, $D^0 \rightarrow \pi^+\pi^-K_S^0$ decays. The uncertainty coming
 288 from the limited size of the simulated signal samples is negligible.

289 The systematic uncertainties related to the signal PDF are obtained from the control
 290 sample $B^0 \rightarrow D^-D_s^+(\rightarrow K^{*0}K^+/\phi\pi^+)$, with $K^{*0} \rightarrow K^+\pi^-$ and $\phi \rightarrow K^+K^-$. The data
 291 are fitted with the nominal PDF, allowing a shift of the mean value $\mu' = \mu + \delta m$ where μ is
 292 fixed to the value fitted in the control channel signal MC and δm is free to vary. The largest
 293 value between δm and its error is taken as uncertainty on the signal mean, leading to 0.1%
 294 (0.2%) uncertainty for Belle (Belle II). For the parameter λ , corresponding to the width of
 295 the signal distribution, we define $\lambda' = \lambda f$ where λ is fixed to the value fitted in the control
 296 channel signal MC and f is a free scaling factor. The maximum of $1 - f$ and the error on
 297 f is taken as the systematic error on this parameter, leading to 21% (59%) uncertainty for
 298 Belle (Belle II). The relatively large resulting uncertainty on λ is mainly due to the limited
 299 size of the control sample. The uncertainties on the signal PDF parameters then translate
 300 into a variation of the branching fraction, as given in Table 2. The tail parameters γ and
 301 δ are very sensitive to the background shape. Since no signal is expected, we do not assign
 302 any systematic uncertainty on these parameters, and only a dedicated systematic on the
 303 background description is estimated.

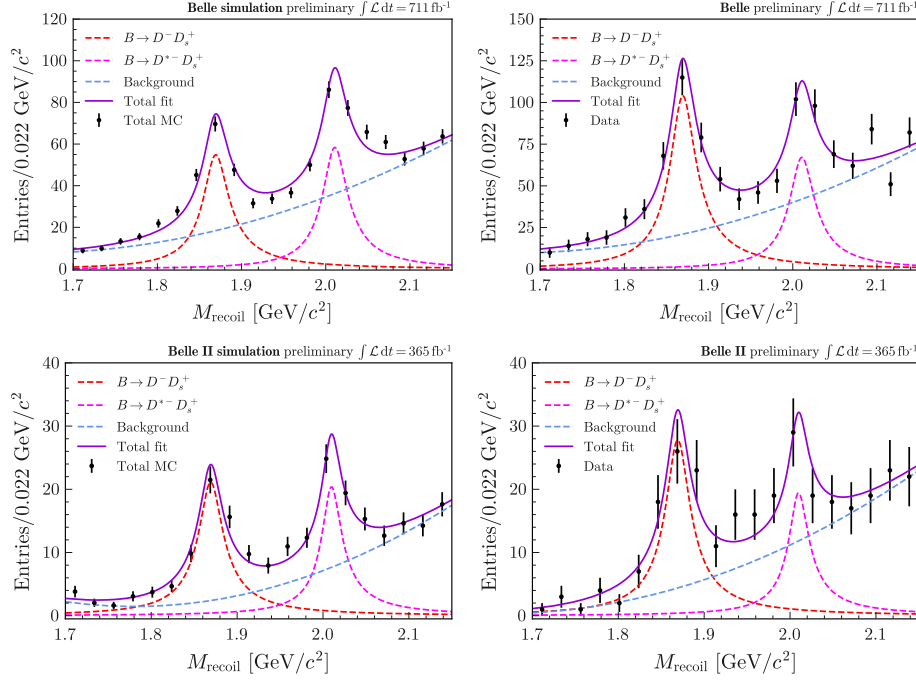


Figure 2. Fits to the recoil mass of the B_{tag} and D_s^+ system in the control channel $B^0 \rightarrow D^- D_s^+$ for simulation (left) and data (right). The top plots correspond to Belle data, while the bottom ones are for Belle II data.

To assign an uncertainty due to the choice of the background PDF, we fit the data with a third-order polynomial. The variation on the fitted branching fraction with respect to the nominal PDF is taken as a systematic uncertainty.

External inputs to the signal branching fractions determination also introduce systematic uncertainties. Those are related to the number of produced $\Upsilon(4S)$ mesons, $N_{\Upsilon(4S)}^{\text{Belle}} = (771.6 \pm 10.6) \times 10^6$ and $N_{\Upsilon(4S)}^{\text{Belle II}} = (387.1 \pm 5.6) \times 10^6$, and the fraction of $\Upsilon(4S)$ decaying into $B^0 \bar{B}^0$ pairs, $f^{00} = (48.61 \pm 0.80)\%$ [6]. The uncertainties on the branching fraction $\mathcal{B}(K^{*0} \rightarrow K^+ \pi^-)$ and the branching fraction of τ to one and three prongs are negligible.

A summary of the systematic uncertainties is given in Table 2.

Systematic uncertainties are taken into account in the upper limit measurement by applying a Gaussian constraint to each of the parameters of the branching fraction. The Gaussian constraint uses the nominal value of the parameter as the mean and for the standard deviation the corresponding systematic uncertainty from Table 2. The systematic uncertainty on the background PDF is estimated directly from the branching fraction, which is a free parameter of the fit and thus cannot be Gaussian constrained. For that reason, the corresponding systematic uncertainty is added to the final upper limit value. For the fit parameters that have a common uncertainty for Belle and Belle II (f^{00} , $\mathcal{B}(K^{*0} \rightarrow K^+ \pi^-)$), a single parameter with the appropriate uncertainty is used in the simultaneous fit. The other systematic uncertainty sources are assumed to be uncorrelated between Belle and Belle II.

Table 2. Summary of the systematic uncertainties. The upper part of the table are systematic uncertainties applied to the efficiency, which are combined in the “Total efficiency” line. The uncertainties expressed as percentages are multiplicative and applied to the corresponding parameter via a Gaussian constraint in the fit. The background PDF uncertainties are additive and directly applied to the branching fraction and upper limit.

| Source | Belle | | | | Belle II | | | |
|---|--------|--------|----------|----------|----------|--------|----------|----------|
| | OS_e | SS_e | OS_μ | SS_μ | OS_e | SS_e | OS_μ | SS_μ |
| FEI efficiency [%] | 4.9 | 4.9 | 4.9 | 4.9 | 6.2 | 6.1 | 6.1 | 6.2 |
| Lepton ID efficiency [%] | 2.0 | 2.4 | 2.2 | 2.2 | 0.7 | 1.1 | 0.7 | 0.6 |
| Hadron ID efficiency [%] | 1.9 | 2.0 | 1.9 | 2.0 | 3.7 | 3.7 | 3.6 | 3.7 |
| BDT efficiency [%] | 27 | 21 | 18 | 23 | 29 | 31 | 34 | 31 |
| Tracking efficiency [%] | 1.4 | | | | 1.1 | | | |
| Total efficiency [%] | 27.6 | 21.8 | 18.9 | 23.7 | 29.8 | 31.8 | 34.7 | 31.7 |
| Signal PDF μ [%] | 0.1 | | | | 0.2 | | | |
| Signal PDF λ [%] | 21 | | | | 59 | | | |
| $N_{\Upsilon(4S)}$ [%] | 1.4 | | | | 1.6 | | | |
| f^{00} [%] | 0.8 | | | | | | | |
| Background PDF ($\times 10^{-5}$) | 0.11 | 0.28 | 0.09 | 0.02 | 0.11 | 0.28 | 0.09 | 0.02 |
| Total impact on UL ($\times 10^{-5}$) | 0.3 | 0.9 | 0.4 | 0.5 | 0.3 | 0.9 | 0.4 | 0.5 |

6 Results and conclusion

The fit results are shown in Fig. 3 with the corresponding branching fractions (\mathcal{B}^{fit}) given in Table 4, where the uncertainties on the fitted branching fractions contain both the statistical and the systematic components. The fitted number of signal and background in Belle and Belle II are shown in Table 3. The values for the number of signal are extracted according to expression (4.2) from the efficiencies displayed in Table 1 and the fitted branching fractions shown in Table 4. All fits are validated with pseudo-experiments and no bias is observed. Since no signal is observed, we set upper limits on the branching fraction using the CLs asymptotic method [48, 49]. The observed upper limits ($\mathcal{B}_{\text{obs}}^{\text{UL}}$) at 90% C.L. are given in Table 4 together with the expected ones ($\mathcal{B}_{\text{exp}}^{\text{UL}}$).

The expected limits are computed from a fit to the data sidebands, assuming a number of observed events in the signal region equal to that extrapolated from the sidebands.

The upper limits on the branching fraction at 90% C.L. are:

$$\begin{aligned}
\mathcal{B}(B^0 \rightarrow K^{*0} \tau^+ e^-) &< 2.9 \times 10^{-5}, \\
\mathcal{B}(B^0 \rightarrow K^{*0} \tau^- e^+) &< 6.4 \times 10^{-5}, \\
\mathcal{B}(B^0 \rightarrow K^{*0} \tau^+ \mu^-) &< 4.2 \times 10^{-5}, \\
\mathcal{B}(B^0 \rightarrow K^{*0} \tau^- \mu^+) &< 5.6 \times 10^{-5}.
\end{aligned}$$

These results are the most stringent upper limits to date for the electron modes. We also report the first search for $B^0 \rightarrow K^{*0} \tau \ell$ LFV decays at $e^+ e^-$ B factories.

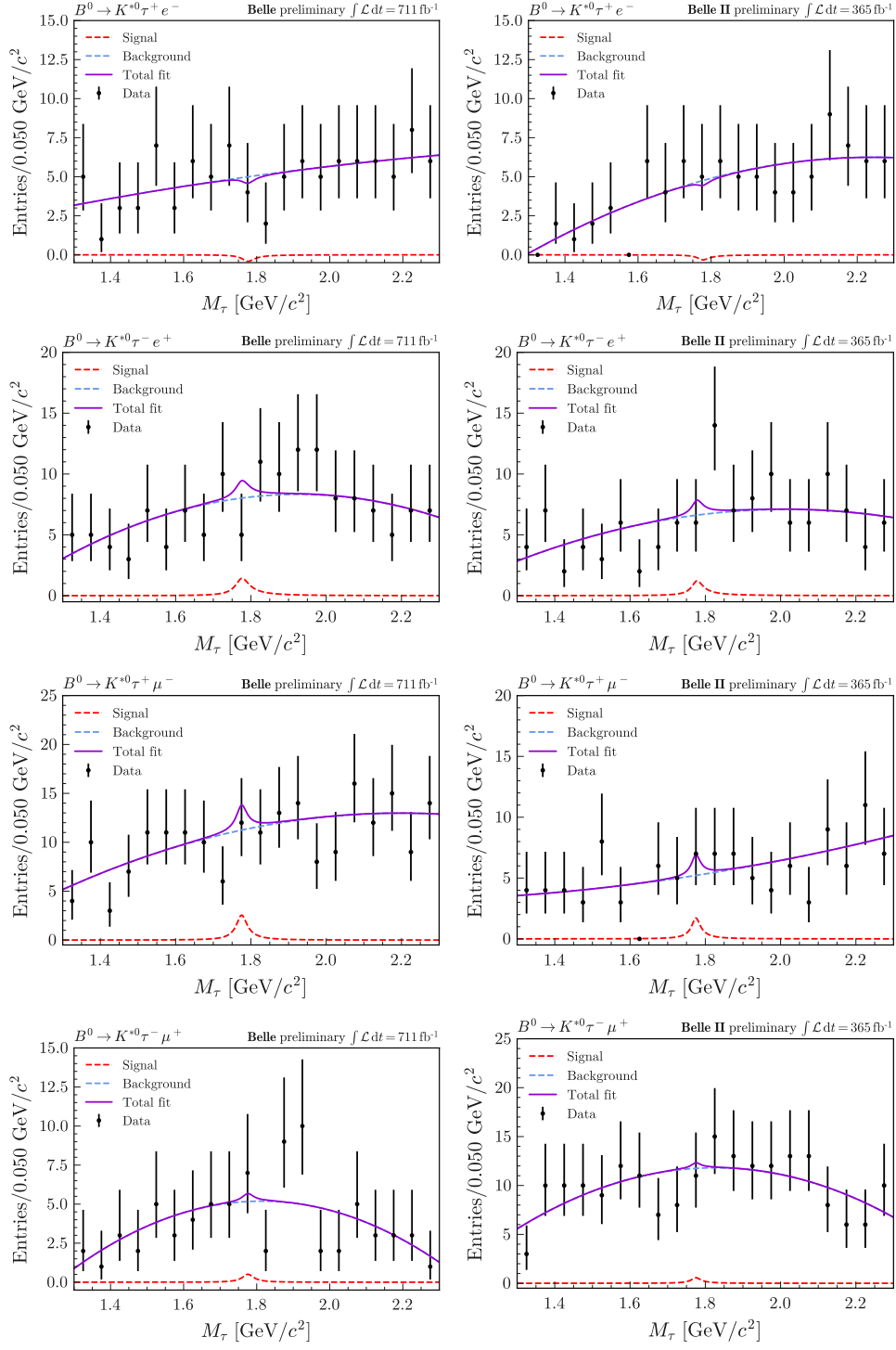


Figure 3. The M_τ distribution and results of the simultaneous fits for Belle (left) and Belle II (right). The black dots with error bars show the data, the red dash-dotted curve shows the signal component, the blue dashed curve shows the background component, and the purple solid curve shows the global fit. From top to bottom: OSe , SSe , $OS\mu$, $SS\mu$.

Table 3. Fitted values of the number of signal (N_{sig}) and background (N_{bkg}) in the Belle and Belle II datasets. The value for N_{sig} is extracted from the efficiency and fitted branching fraction.

| | | OSe | SSe | $OS\mu$ | $SS\mu$ |
|----------|------------------|-----------------|------------------|------------------|------------------|
| Belle | N_{sig} | -0.6 ± 3.3 | 2.2 ± 5.2 | 2.8 ± 4.7 | 0.6 ± 3.2 |
| | N_{bkg} | 99.5 ± 10.4 | 140.0 ± 12.7 | 207.4 ± 15.0 | 76.4 ± 9.3 |
| Belle II | N_{sig} | -0.5 ± 2.8 | 1.7 ± 3.9 | 1.6 ± 2.7 | 0.6 ± 3.3 |
| | N_{bkg} | 86.5 ± 9.8 | 120.1 ± 11.8 | 109.1 ± 10.9 | 198.4 ± 14.5 |

Table 4. Measured branching fractions and observed (expected) upper limits at 90% CL on the four $B^0 \rightarrow K^{*0}\tau^\pm\ell^\pm$ decays.

| Decay | $\mathcal{B}^{\text{fit}} (\times 10^{-5})$ | $\mathcal{B}_{\text{obs(exp)}}^{\text{UL}} (\times 10^{-5})$ |
|--|---|--|
| $OSe: B^0 \rightarrow K^{*0}\tau^+e^-$ | -0.24 ± 1.46 | 2.9 (2.8) |
| $SSe: B^0 \rightarrow K^{*0}\tau^-e^+$ | 1.17 ± 2.77 | 6.4 (4.4) |
| $OS\mu: B^0 \rightarrow K^{*0}\tau^+\mu^-$ | 1.07 ± 1.80 | 4.2 (3.0) |
| $SS\mu: B^0 \rightarrow K^{*0}\tau^-\mu^+$ | 0.48 ± 2.61 | 5.6 (5.5) |

This work, based on data collected using the Belle II detector, which was built and
 commissioned prior to March 2019, and data collected using the Belle detector, which was
 operated until June 2010, was supported by Higher Education and Science Committee of
 the Republic of Armenia Grant No. 23LCG-1C011; Australian Research Council and Re-
 search 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 Hori-
 zon 2020 ERC Starting Grant No. 947006 “InterLeptons”; Natural Sciences and Engi-
 neering Research Council of Canada, Compute Canada and CANARIE; National Key
 R&D Program of China under Contract No. 2022YFA1601903, National Natural Science
 Foundation of China and Research Grants No. 11575017, No. 11761141009, No. 11705209,
 No. 11975076, No. 12135005, No. 12150004, No. 12161141008, No. 12475093, and No. 12175041,
 and Shandong Provincial Natural Science Foundation Project ZR2022JQ02; the Czech
 Science Foundation Grant No. 22-18469S and Charles University Grant Agency project
 No. 246122; European Research Council, Seventh Framework PIEF-GA-2013-622527, Hori-
 zon 2020 ERC-Advanced Grants No. 267104 and No. 884719, Horizon 2020 ERC-Consolidator
 Grant No. 819127, Horizon 2020 Marie Skłodowska-Curie Grant Agreement No. 700525
 “NIOBE” and No. 101026516, and Horizon 2020 Marie Skłodowska-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 Na-
 tionale de la Recherche (ANR) under Grant No. ANR-21-CE31-0009 (France); BMBF,
 DFG, HGF, MPG, and AvH Foundation (Germany); Department of Atomic Energy un-
 der 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 Bi-
 national 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; Na-
 tional Research Foundation (NRF) of Korea Grants No. 2016R1-D1A1B-02012900, No. 2018R1-
 A6A1A-06024970, No. 2021R1-A6A1A-03043957, No. 2021R1-F1A-1060423, No. 2021R1-
 F1A-1064008, No. 2022R1-A2C-1003993, No. 2022R1-A2C-1092335, No. RS-2023-00208693,
 No. RS-2024-00354342 and No. RS-2022-00197659, Radiation Science Research Institute,
 Foreign Large-Size Research Facility Application Supporting project, the Global Science
 Experimental Data Hub Center, the Korea Institute of Science and Technology Information
 (K24L2M1C4) and KREONET/GLORIAD; Universiti Malaya RU grant, Akademi Sains

Malaysia, and Ministry of Education Malaysia; Frontiers of Science Program Contracts No. FOINS-296, No. CB-221329, No. CB-236394, No. CB-254409, and No. CB-180023, and SEP-CINVESTAV Research Grant No. 237 (Mexico); the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation and the HSE University Basic Research Program, Moscow; University of Tabuk Research Grants No. S-0256-1438 and No. S-0280-1439 (Saudi Arabia), and Researchers Supporting Project number (RSPD2025R873), King Saud University, Riyadh, Saudi Arabia; Slovenian Research Agency and Research Grants No. J1-9124 and No. P1-0135; Ikerbasque, Basque Foundation for Science, the State Agency for Research of the Spanish Ministry of Science and Innovation through Grant No. PID2022-136510NB-C33, Agencia Estatal de Investigacion, Spain Grant No. RYC2020-029875-I and Generalitat Valenciana, Spain Grant No. CIDEAGENT/2018/020; the Swiss National Science Foundation; The Knut and Alice Wallenberg Foundation (Sweden), Contracts No. 2021.0174 and No. 2021.0299; National Science and Technology Council, and Ministry of Education (Taiwan); Thailand Center of Excellence in Physics; TUBITAK ULAKBIM (Turkey); National Research Foundation of Ukraine, Project No. 2020.02/0257, and Ministry of Education and Science of Ukraine; the U.S. National Science Foundation and Research Grants No. PHY-1913789 and No. PHY-2111604, and the U.S. Department of Energy and Research Awards No. DE-AC06-76RLO1830, No. DE-SC0007983, No. DE-SC0009824, No. DE-SC0009973, No. DE-SC0010007, No. DE-SC0010073, No. DE-SC0010118, No. DE-SC0010504, No. DE-SC0011784, No. DE-SC0012704, No. DE-SC0019230, No. DE-SC0021274, No. DE-SC0021616, No. DE-SC0022350, No. DE-SC0023470; and the Vietnam Academy of Science and Technology (VAST) under Grants No. NVCC.05.12/22-23 and No. DL0000.02/24-25.

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 IBelle on site; the KEK Computer Research Center for on-site computing support; the NII for SINET6 network support; and the raw-data centers hosted by BNL, DESY, GridKa, IN2P3, INFN, PNNL/EMSL, and the University of Victoria.

References

- [1] L. Calibbi and G. Signorelli, *Charged Lepton Flavour Violation: An Experimental and Theoretical Introduction*, *Riv. Nuovo Cim.* **41** (2018) 71.
- [2] S. Davidson, B. Echenard, R.H. Bernstein, J. Heeck and D.G. Hitlin, *Charged Lepton Flavor Violation*, [2209.00142](#).
- [3] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (1967) 1264.
- [4] S.L. Glashow, J. Iliopoulos and L. Maiani, *Weak interactions with lepton-hadron symmetry*, *Phys. Rev. D* **2** (1970) 1285.
- [5] LHCb collaboration, *Measurement of the Ratios of Branching Fractions $\mathcal{R}(D^*)$ and $\mathcal{R}(D^0)$* , *Phys. Rev. Lett* **131** (2023) .

- [6] HEAVY FLAVOR AVERAGING GROUP (HFLAV), *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2023*, [2411.18639](#).
- [7] G. Hiller, D. Loose and K. Schönwald, *Leptoquark flavor patterns and B decay anomalies*, *JHEP* **12** (2016) 027.
- [8] A. Angelescu, D. Bečirević, D.A. Faroughy, F. Jaffredo and O. Sumensari, *Single leptoquark solutions to the B -physics anomalies*, *Phys. Rev. D* **104** (2021) 055017.
- [9] A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski and J. Rosiek, *Lepton-flavor violating B decays in generic Z' models*, *Phys. Rev. D* **92** (2015) 054013.
- [10] D. Bečirević, O. Sumensari and R. Zukanovich Funchal, *Lepton flavor violation in exclusive $b \rightarrow s$ decays*, *Eur. Phys. J. C* **76** (2016) 134 [[1602.00881](#)].
- [11] D. Bečirević, F. Jaffredo, J.P. Pinheiro and O. Sumensari, *Lepton flavor violation in exclusive $b \rightarrow d\ell_i\ell_j$ and $b \rightarrow s\ell_i\ell_j$ decay modes*, [2407.19060](#).
- [12] LHCb collaboration, *Amplitude analysis of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay*, *Phys. Rev. Lett.* **132** (2024) 131801.
- [13] LHCb collaboration, *Determination of short- and long-distance contributions in $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays*, *Phys. Rev. D* **109** (2024) 052009.
- [14] BELLE II collaboration, *Evidence for $B^+ \rightarrow K^+\nu\bar{\nu}$ decays*, *Phys. Rev. D* **109** (2024) 112006.
- [15] L. Allwicher, D. Bečirević, G. Piazza, S. Rosauero-Alcaraz and O. Sumensari, *Understanding the first measurement of $\mathcal{B}(B \rightarrow K\nu\bar{\nu})$* , *Physics Letters B* **848** (2024) 138411.
- [16] LHCb collaboration, *Search for the lepton-flavour violating decays $B^0 \rightarrow K^{*0}\tau^\pm\mu^\mp$* , *JHEP* **06** (2023) 143 [[2209.09846](#)].
- [17] BABAR collaboration, *Search for the decay modes $B^\pm \rightarrow h^\pm\tau\ell$* , *Phys. Rev. D* **86** (2012) 012004.
- [18] BELLE collaboration, *Search for the Lepton Flavor Violating Decays $B^+ \rightarrow K^+\tau^\pm\ell^\mp$ ($\ell = e, \mu$) at Belle*, *Phys. Rev. Lett.* **130** (2023) 261802.
- [19] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*, *Nucl. Instrum. Meth.* **A907** (2018) 188 [[1809.01958](#)].
- [20] Y. Ohnishi, T. Abe, T. Adachi, K. Akai, Y. Arimoto, K. Ebihara et al., *Accelerator design at superkekb*, *Progress of Theoretical and Experimental Physics* **2013** (2013) 03A011.
- [21] BELLE II collaboration, *Belle II technical design report*, [1011.0352](#).
- [22] BELLE II SVD collaboration, *The design, construction, operation and performance of the Belle II silicon vertex detector*, *JINST* **17** (2022) P11042 [[2201.09824](#)].
- [23] D. Kotchetkov et al., *Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector*, *Nucl. Instrum. Meth.* **A941** (2019) 162342 [[1804.10782](#)].
- [24] S. Kurokawa and E. Kikutani, *Overview of the KEKB accelerators*, *Nucl. Instrum. Meth.* **A499** (2003) 1.
- [25] T. Abe et al., *Achievements of KEKB*, *PTEP* **2013** (2013) 03A001.
- [26] A. Abashian et al., *The Belle detector*, *Nucl. Instrum. Meth.* **A479** (2002) 117.
- [27] BELLE collaboration, *Physics Achievements from the Belle Experiment*, *PTEP* **2012** (2012) 04D001 [[1212.5342](#)].

- [28] S. Jadach, B. 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](#)].
- [29] 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](#)].
- [30] T. Sjöstrand, S. Mrenna and P. Skands, *Pythia 6.4 physics and manual*, *JHEP* **05** (2006) 026.
- [31] D.J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152.
- [32] 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.
- [33] GEANT4 collaboration, *GEANT4: A simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.
- [34] R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson et al., *GEANT Detector Description and Simulation Tool*, .
- [35] Z. Liptak, A. Paladino, L. Santelj, J. Schueler, S. Stefkova, H. Tanigawa et al., *Measurements of beam backgrounds in SuperKEKB Phase 2*, *Nucl. Instrum. Meth.* **A1040** (2022) 167168.
- [36] BELLE II FRAMEWORK SOFTWARE GROUP, *The Belle II Core Software*, *Comput. Softw. Big Sci.* **3** (2019) 1 [[1809.04299](#)].
- [37] BELLE II collaboration, “Belle II Analysis Software Framework (basf2).” <https://doi.org/10.5281/zenodo.5574115>.
- [38] N. Katayama, R. Itoh, A. Manabe and T. Sasaki, *Belle computing model*, *Comput. Phys. Commun.* **110** (1998) 22–25.
- [39] M. Gelb et al., *B2BII: Data Conversion from Belle to Belle II*, *Comput. Softw. Big Sci.* **2** (2018) 9 [[1810.00019](#)].
- [40] T. Keck et al., *The Full Event Interpretation*, *Comput. Softw. Big Sci.* **3** (2019) 6 [[1807.08680](#)].
- [41] BELLE AND BELLE II collaborations, *Search for lepton flavor-violating decay modes $B^0 \rightarrow K_S^0 \tau^\pm \ell^\mp$ ($\ell = \mu, e$) with hadronic B-tagging at Belle and Belle II*, **2412.16470**.
- [42] J.D. Bjorken and S.J. Brodsky, *Statistical Model for electron-Positron Annihilation Into Hadrons*, *Phys. Rev. D* **1** (1970) 1416.
- [43] PARTICLE DATA GROUP, *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001.
- [44] BELLE collaboration, *Evidence for $B^0 \rightarrow \pi^0 \pi^0$* , *Phys. Rev. Lett.* **91** (2003) 261801 [[hep-ex/0308040](#)].
- [45] G. Punzi, *Sensitivity of searches for new signals and its optimization*, *eConf* **C030908** (2003) MODT002 [[physics/0308063](#)].
- [46] N.L. Johnson, *Systems of frequency curves generated by methods of translation*, *Biometrika* **36** (1949) 149.
- [47] BELLE II collaboration, *A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \rightarrow X \ell \nu$ decays*, **2008.06096**.
- [48] A.L. Read, *Presentation of search results: The $CL(s)$ technique*, *J. Phys. G* **28** (2002) 2693.

[49] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011) .

Additional Material

The BDT score distributions of MC simulations, signal MC and sidebands data ($M_\tau \in [1.0, 1.65 \cup 1.9, 2.5] \text{ GeV}/c^2$) of the four modes for Belle and Belle II are shown in Figure 4.

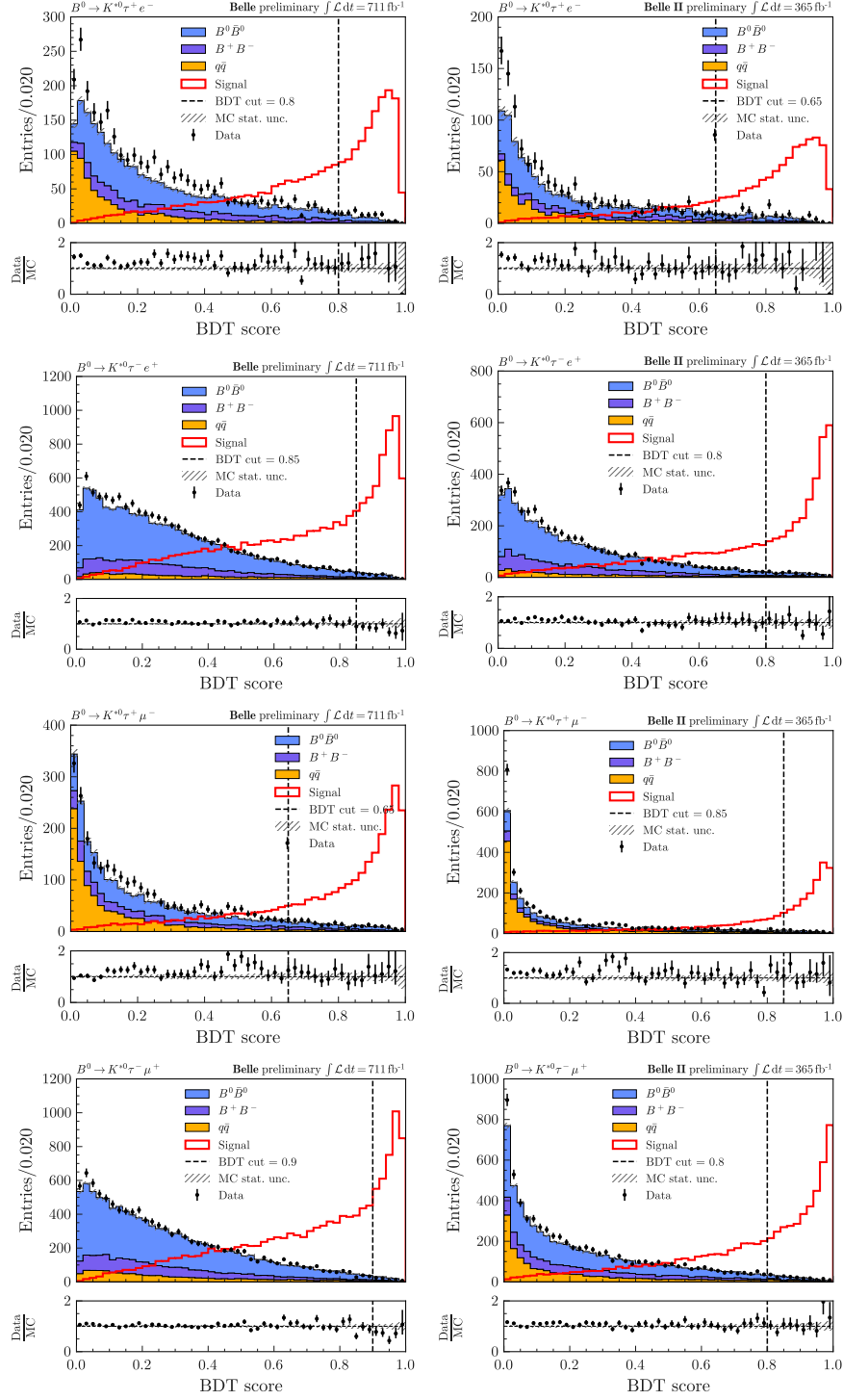


Figure 4. BDT distributions in generic MC simulations (stacked histograms), data sidebands $M_\tau \in [1.0, 1.65] \cup [1.9, 2.5] \text{ GeV}/c^2$ (black points) and signal MC (red histogram) for Belle (left) and Belle II (right) datasets. The generic MC are corrected for known data/MC mismodelling and scaled to the data luminosity, and signal MC is scaled to the same area as data. The dashed line indicates the value of the BDT cut applied. From top to bottom: OSe , SSe , $OS\mu$, $SS\mu$.