Test of lepton flavor universality in inclusive semileptonic $B$ decays at Belle II

(The Belle II Collaboration)


(The Belle II Collaboration)

Abstract

We present the measurement of the ratio of branching fractions of inclusive semileptonic B decays, $R(X_{e/\mu}) = \frac{\mathcal{B}(B \to Xe\nu)}{\mathcal{B}(B \to X\mu\nu)}$, with a hadronic tagged analysis using 189.9 fb$^{-1}$ of Belle II data. We find $R(X_{e/\mu}) = 1.033 \pm 0.010 \pm 0.020$. The errors quoted correspond to the statistical and systematic uncertainties, respectively. To the best of our knowledge, this is the most precise single test of $e-\mu$ flavor universality in semileptonic B decays and agrees with the Standard Model expectation within less than 1.5$\sigma$. This measurement paves the path to the measurement of $R(X) = \frac{\mathcal{B}(B \to X\tau\nu)}{\mathcal{B}(B \to X\ell\nu)}$ at Belle II.
1. INTRODUCTION

$B$ meson decays to tau leptons involving the $b \to c\tau\nu$ transition are a powerful probe for physics beyond the Standard Model (SM). Their branching fractions, normalized to the analogous decays to light leptons ($\ell \in \{e, \mu\}$), may be enhanced due to lepton flavor universality violating new physics \cite{1, 2}. A persistent tension of more than 3$\sigma$ between experiment and the SM has been observed in ratios of exclusive decays $R(D^{(*)}) = \frac{B(B \to D^{(*)}\tau\nu)}{B(B \to D^{(*)}\ell\nu)}$ \cite{3-10}. The ratio of branching fractions of inclusive decays $R(X) = \frac{B(B \to X\tau\nu)}{B(B \to X\ell\nu)}$ provides an alternative probe of this anomaly, but it is experimentally challenging and has not been measured since the LEP era \cite{11-15}.

In this work, we present a measurement of the inclusive ratio $R(X_{e/\mu}) = \frac{B(B \to Xe\nu)}{B(B \to X\mu\nu)}$, which tests lepton flavor universality in the light leptons sector and serves as a preliminary test of the analysis’ robustness towards the first measurement of $R(X)$ at a $B$-factory. We utilize a Belle II collision dataset corresponding to an integrated luminosity of 189.9 $fb^{-1}$.

2. THE BELLE II DETECTOR, COLLISION DATA AND SIMULATED SAMPLES

The Belle II detector \cite{16, 17} operates at the SuperKEKB asymmetric-energy electron-positron collider \cite{18} at the KEK laboratory in Tsukuba, Japan. The detector consists of several nested subsystems arranged around the beam pipe in a cylindrical geometry. The innermost subsystem is the vertex detector, which is comprised of two layers of silicon pixels (PXD) and four outer layers of silicon strip (SVD) detectors. Currently, the second pixel layer is installed in only a small part of the solid angle, whilst the remaining layers are fully operational. Most of the tracking volume consists of a small-cell drift chamber (CDC) filled with a He (50%) and $C_2H_6$ (50%) gas mixture. Moving outwards from the interaction point, a Cherenkov-light imaging and time-of-propagation (TOP) detector provides pion and kaon identification in the barrel region. To serve an analogous purpose, the forward endcap region is instrumented with a proximity-focusing, ring-imaging Cherenkov (ARICH) detector with an aerogel radiator. Further out, the electromagnetic calorimeter (ECL) provides neutral particles and electron identification. The ECL consists of a barrel and two endcap sections containing 6624 and 2112 CsI(Tl) crystals, respectively. All the subsystem thus far described are embedded in a uniform 1.5 T magnetic field from a superconducting solenoid situated outside the calorimeter. The outermost subsystem, the $K_L^0$ and muon identification (KLM) detector, consists of scintillator strips in the endcaps and the inner part of the barrel and resistive plate chambers in the outer barrel, interleaved to iron plates that serve as magnetic flux return yoke.

The collision data used in this analysis were collected at a center-of-mass (CM) energy of $\sqrt{s} = 10.58$ GeV, corresponding to the mass of the $\Upsilon(4S)$ resonance. The energies of the electron and positron beams are 7 GeV and 4 GeV, respectively, resulting in a boost of $\beta\gamma = 0.28$ of the CM frame relative to the laboratory frame. In addition, 18 $fb^{-1}$ of off-resonance collision data, collected 60 MeV below the $\Upsilon(4S)$ resonance, is used to model background from $e^+e^-$ continuum processes, i.e. $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$ and $c\bar{c}$. Due to the $e^+e^- \to q\bar{q}$ cross section dependency on $1/s$, the off-resonant data yield is scaled by a factor...
We use Monte Carlo simulation to produce the signal and the remaining background model templates, and to calculate reconstruction efficiencies and detector acceptance. Physics simulation is generated with the EvtGen [19] software. The detector simulation is performed with GEANT4 [20]. The final state radiation of photons from charged stable particles is simulated using the PHOTOS [21] software. After this step, simulated events are overlaid with real, randomly triggered beam-induced background events from collision data to improve their modelling, and are digitized. They are subsequently reconstructed and analyzed in the same fashion as the collision data with the open-source Belle II Analysis Software Framework, basf2 [22].

The $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^+B^-(B^0\overline{B}^0)$ samples contain semileptonic and most hadronic $B$ decays, namely the signal $B \rightarrow X\ell\nu$ events as well as any backgrounds arising from $B$ decays. The inclusive signal model mostly consists of the sum of several well-measured exclusive decays: $B \rightarrow D\ell^+\nu_\ell$, $B \rightarrow D^*\ell^+\nu_\ell$, $B \rightarrow D^{**}\ell^+\nu_\ell$. In the latter, $D^{**}$ collectively indicates the excited charm states $D_0^*, D_1, D_1', D_2^*$, whose masses and widths are taken from [23]. The $B \rightarrow D^{(*)}\ell^+\nu_\ell$ decays are modeled with the BGL [24–26] form factor parametrization. The modeling of $B \rightarrow D^{**}\ell^+\nu_\ell$ decays relies on form factors based on the heavy quark effective field theory BLR model [27, 28]. Semileptonic $B$ decays into the non-resonant final states $B \rightarrow D^{(*)}\pi\pi\ell^+\nu_\ell$, $B \rightarrow D^{(*)}\eta\ell^+\nu_\ell$ are used to fill the remaining “gap” between the sum of individual branching ratios of exclusive decays, $B \rightarrow D^{(*)}\ell^+\nu_\ell$ and $B \rightarrow D^{**}\ell^+\nu_\ell$, and the measured total $B$ meson decay width as from [23]. We henceforth indicate such decays as gap modes. They are included in dedicated simulated samples that use intermediate, broad $D^{**}$ resonances whose decays are modeled with BLR.

3. ANALYSIS DESCRIPTION

We reconstruct $\Upsilon(4S)$ events by first reconstructing one $B$ meson candidate in a fully hadronic decay. We then identify a light lepton candidate and assign all other reconstructed particle candidates to the $X$ system. The lepton and $X$ candidates together constitute the signal $B$ meson candidate, $B_{\text{sig}}$.

We use the Full Event Interpretation (FEI) algorithm [29] to tag the hadronically decaying $B$ meson in an event, $B_{\text{tag}}$. Starting from reconstructed charged and neutral particles ($e^\pm, \mu^\pm, \pi^\pm, K^\pm, p, d, \gamma$), the FEI combines them into a set of intermediate states ($J/\psi, K_S^0, D^+, D^{*+}D_s^-, D_s^{*+}, \Lambda, \Sigma^+$) and assigns a probability of correctly reconstructing each candidate using boosted decision trees. These particles are then further combined to form $B$ meson candidates, whose signal probability is assigned via the same algorithm, with each subsequent step of the combination relying on the probability of the intermediate states in the chain, as well as on updated information from kinematic fits. The FEI eventually reconstructs more than 100 hadronic $B$ decays. We use three variables to maximize the purity of the $B_{\text{tag}}$ selection: the beam-constrained mass $M_{bc} = \sqrt{(\sqrt{s}/2)^2 - |\vec{p}_B|^2}$, the energy difference $\Delta E = E_B^* - \sqrt{s}/2$, and the binary score produced by the FEI to classify $B$ mesons, $P_{\text{FEI}}$. The asterisk indicates quantities expressed in the CM frame. We select $B_{\text{tag}}$ candidates with $M_{bc} \in [5.2725 \text{ GeV}, 5.285 \text{ GeV}]$ and $\Delta E \in [-0.15 \text{ GeV}, 0.1 \text{ GeV}]$ and

\[ c_{\text{off-res}} = \left( \frac{\sqrt{s_{\text{off-res}}}}{\sqrt{s_{\Upsilon(4S)}}} \right)^2 = 0.98. \]
$P_{\text{FEI}} > 0.1$. In case multiple $B_{\text{tag}}$ candidates pass these selections in an event, we choose the one with the highest value of $P_{\text{FEI}}$. We use scaled off-resonant data to describe continuum backgrounds. In this dataset, the total event energy is decreased by a small fraction compared to collisions at the $T(4S)$ resonance. Assuming that the momenta and energies of the final state particles are also decreased by this same fraction, we scale them in the off-resonant sample by a factor of $1/\sqrt{c_{\text{off-res}}} = 1.02$. This scaling improves the description of the continuum backgrounds, particularly in $M_{\text{bc}}$ and $\Delta E$.

We reconstruct signal lepton candidates from the remaining tracks after the $B_{\text{tag}}$ reconstruction. They are required to have an impact parameter consistent with the interaction point in radius ($dr < 1\, \text{cm}$) and along the beam axis ($|dz| < 3\, \text{cm}$), and to lie within the CDC polar angle acceptance. Muon candidates are required to have $p_T > 0.4\, \text{GeV}$ and are identified by means of a likelihood ratio discriminator. This discriminator combines information from likelihood functions defined in each sub-detector for each charged stable particle hypothesis $i \in \{e, \mu, \pi, K, p, d\}$:

$$
\text{PID}_\mu = \frac{\mathcal{L}_\mu}{\sum_i \mathcal{L}_i}, \quad \mathcal{L}_i = \prod_d \mathcal{L}_i^d,
$$

where $D = \{\text{CDC}, \text{TOP}, \text{ARICH}, \text{ECL}, \text{KLM}\}$. The value of $\text{PID}_\mu$ is required to be above 0.95.

We correct the four-momenta of electron candidates for Bremsstrahlung radiation by adding calorimeter clusters not matched to any track that are found within a cone centered on the electron track’s momentum vector. The opening angle of this cone depends on the momentum magnitude: 0.1368, 0.0737, and 0.0632 radians for $p \leq 0.6\, \text{GeV}$, $p \in (0.6, 1.0]\, \text{GeV}$, and $p > 1.0\, \text{GeV}$, respectively. Electrons are then required to have a transverse momentum of $p_T > 0.3\, \text{GeV}$ and are identified by means of a multi-class boosted decision tree classifier that exploits several ECL cluster observables in combination with likelihood ratios from the other Belle II sub-systems, cf. Eq. 1. The classifier cuts are tuned in a three-dimensional grid of lab-frame momentum, polar angle and charge bins to achieve a uniform 80% identification efficiency. To further suppress anti-proton fakes, electron candidates are demanded to fulfill $\text{PID}_p < 0.9$.

The lepton identification efficiency and the pion, kaon to lepton mis-identification probabilities in the simulation are corrected in a data-driven way by means of dedicated calibration channels. The corrections are measured in bins of lab-frame momentum, polar angle and charge of the lepton candidate. The electron identification efficiency correction factors are derived from a combination of $J/\psi \rightarrow e^+e^-$, $e^+e^- \rightarrow e^+e^- (\gamma)$ and $e^+e^- \rightarrow (e^+e^-)e^+e^-$ events. For muon identification, events from $J/\psi \rightarrow \mu^+\mu^-$, $e^+e^- \rightarrow \mu^+\mu^- \gamma$ and $e^+e^- \rightarrow (e^+e^-)\mu^+\mu^-$ processes are used. In the barrel region, the correction factors for simulated electrons range between 0.96 and 1.01 with a total per-electron uncertainty found in the 0.1-2% interval. This uncertainty is dominated by the observed difference between the calibration channels in some of the bins. For muons, the correction factors range between 0.90 and 1.02, and the achieved precision is of the same order as for the electrons’ case.

Pion-to-lepton mis-identification rates are corrected from $K_S^0 \rightarrow \pi^+\pi^-$ and $e^+e^- \rightarrow \tau^\pm(1P)\tau^\mp(3P)$ events, whereas $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ events are used to measure the kaon-to-lepton mis-identification probability corrections. The $\pi \rightarrow e (\mu)$ mis-identification rate corrections range between 2 and 8 (0.5 and 1.5); uncertainties are in the range 20-
70% (5-20%), with the lowest precision reached in the $p > 2.0 \text{GeV}/c$ region, outside of the
kinematic acceptance of the high statistics $K_S^0$ sample. The $K \rightarrow \pi$ scale factors are found
between 0 and 10, although with uncertainties of order 200% or more due to the limited
size of the $D^{\ast +} \rightarrow D^0(\rightarrow K^- \pi^+)\pi^+$ sample for the selected, very tight, working point. The
$K \rightarrow \mu$ correction factors are in the range between 0.9 and 2, with a precision of order
20-30%, also limited by the calibration sample statistics.

All tracks and neutral ECL clusters remaining in the event that pass certain quality tests
are then combined to form the $X$ system. Clusters must be found more than 30 cm away
from the nearest track to suppress contamination of fake neutral clusters from hadronic
interactions, and are required to have energies greater than 0.04, 0.055, and 0.09 GeV in the
forward, barrel, and backward regions of the calorimeter, respectively. Tracks are required to
have impact parameters consistent with the interaction point ($dr < 2 \text{cm}$ and $|dz| < 4 \text{cm}$),
be in the CDC polar angle acceptance, and have at least one hit in the CDC. Mass hypotheses
are assigned to each track by checking particle identification criteria in a specific sequence and
assigning the hypothesis of the first satisfied criterion. The sequence (criteria) are: electron
(same as signal), muons (same as signal), kaons (PID $K > 0.6$), protons (PID $p > 0.5$),
dereuterons (PID $d > 0.5$) and pions (all remaining tracks).

To suppress continuum, we train a boosted decision tree on a $B^+B^-, B^0\bar{B}^0$ simulated
sample and a continuum background sample from off-resonant data, using 21 event-shape
variables which are well-modeled in the simulation. These variables are built from track
candidates that pass the same selection criteria as the ones used for the $X$ system recon-
struction. The cut on the classifier score is chosen to be $P_{CS} > 0.2$.

The $X$ system contains a large variety of different charged and neutral final state particles
arising from a diverse set of decay chains. This rich structure is difficult to model in the
simulation, therefore we do not consider any kinematic observable associated to it for the
measurement. Instead, the lepton momentum in the $B_{\text{sig}}$ CM frame, $p_{\ell}^*$, is used to extract
the signal yield from a one-dimensional fit in a background-depleted region of phase space,
since it only depends on the modeling of the leptonic part of the $B$ decay. This observable
is also advantageous to suppress contributions from $B \rightarrow X \tau \nu$ events. In the fit, we define
three templates for each lepton flavor: one for $B \rightarrow X \nu$, one for the continuum background,
and one for the collective remaining backgrounds, which are mostly events with hadrons mis-
identified as leptons (“fake” leptons) and candidates with true leptons originating mainly
from decays of charmed hadronic $B$ decays (“secondary” leptons). The electron and muon
channel are fitted simultaneously so that any correlations between uncertainties affecting
both are taken into account.

Figure[1] shows the pre-fit distribution of $p_{\ell}^*$ over its full range, showing the different model
templates considered. To suppress backgrounds, events with $p_{\ell}^* < 1.3 \text{GeV}$ are rejected. The
measurement of the ratio between electron and muon events is independent from possible
overall normalization differences between data and simulation due to FEI tagging efficiency
discrepancies. Thus, in the figures the templates based on simulation are further scaled to
match the data normalization, on top of the integrated luminosity scaling. The $B \rightarrow \tau \nu
template scaling factors are derived in the signal region itself. The global scale factor for the
fake and secondary leptons background templates is extracted from a fit to data in dedicated
background-enriched control regions where the requirement on the lepton charge and the
charge of the $B_{\text{tag}}$ inferred from its flavor ($\Upsilon(4S) \rightarrow B_{\text{tag}}^{+0}, B_{\text{sig}} \rightarrow X\ell^{\ast} \nu + \text{c.c.}$) is reversed.
The remaining contamination of $B \rightarrow X \ell \nu$ events is due to $B^0$ flavour mixing and incorrect
FIG. 1. The pre-fit distributions of the lepton momentum in the $B_{\text{sig}}$ CM frame, $p_{\ell}^*$, for the electron (left) and muon (right) channel in the wrong charge (top) and correct charge (bottom) case. In the fit, $B \to X_u \ell^+ \nu_\ell$, $B \to D^{(*)} \ell \nu$ and other $B \to X_c \ell^+ \nu_\ell$ decays are summarized in the $B \to X \ell \nu$ template. Fake leptons, secondary leptons, other $B\bar{B}$ and the leftover $B \to X \tau \nu$ contributions are summarized as backgrounds.

The $B \to X \ell \nu$ yields are left to float freely in the fit, while the continuum background yields are constrained with a Gaussian term according to the number of events measured in off-resonant data. The other background yields are also left free to float, using the normalization factors and associated uncertainties obtained from the fits in the wrong charge control regions to constrain the fit.
The statistical and systematic uncertainties are incorporated into the fit via nuisance parameters, with one parameter per template and $p^*_k$ bin. This modifies the PDFs according to:

$$p^k_i \rightarrow \frac{p^k_i (1 + \epsilon^k_i \theta^k_i)}{\sum_i p^k_i (1 + \epsilon^k_i \theta^k_i)}.$$  \hfill (2)

Here $\epsilon^k_i$ denotes the total systematic error of the $i$-th bin in the $k$-th template. Further, $\theta^k_i$ denotes the nuisance parameter. The expected number of events in each bin is then given by:

$$\nu^\text{expected}_i = \sum_k \nu^k_i \frac{p^k_i (1 + \epsilon^k_i \theta^k_i)}{\sum_i p^k_i (1 + \epsilon^k_i \theta^k_i)}.$$ \hfill (3)

The total covariance matrix $C_\theta$ of a given template $k$ is constructed by summing up the covariance matrices of all individual uncertainty sources. The total negative log likelihood function that is minimized is given by:

$$-\log(\mathcal{L}) = -\log \prod_i P(\nu^\text{observed}_i, \nu^\text{expected}_i) + \frac{\theta^T C_\theta^{-1} \theta + (\bar{\mathbf{f}} - \bar{\mathbf{f}}_{\text{constraint}})^T C_\mathbf{f}^{-1} (\mathbf{f} - \mathbf{f}_{\text{constraint}})}{2}.$$ \hfill (4)

Here $\nu^\text{observed}_i$ is the observed number of events in a given bin $i$ and $P(\nu^\text{observed}_i, \nu^\text{expected}_i)$ is the Poisson function. The term $\theta^T C_\theta^{-1} \theta$ constrains the nuisance parameters. Further, $(\bar{\mathbf{f}} - \bar{\mathbf{f}}_{\text{constraint}})^T C_\mathbf{f}^{-1} (\mathbf{f} - \mathbf{f}_{\text{constraint}})$ constrains the continuum template normalization.

### 4. SYSTEMATIC UNCERTAINTIES

The electron and muon identification and fake rate corrections with their uncertainties were explained in detail previously. Other particles faking leptons for which data-driven corrections are not available, such as anti-protons faking positrons, account for no more than 10% of the total lepton fakes composition. Therefore we assign a 100% uncertainty to their mis-identification rate. We propagate lepton identification efficiency uncertainties via 200 weights that are randomly generated from Gaussian variations using the uncertainties and covariance of the efficiency corrections. We consider corrections within the same ($p, \theta$) bin to be fully correlated for the same efficiency / fake rate type, for other factors of the same type we assume that only the systematic uncertainties are correlated.

We update the branching fractions of the $B \to D^{(*)} \ell \nu$ processes to the latest values provided by the HFLAV group [31], combining the results of neutral and charged $B$ mesons assuming isospin symmetry. For $B \to D^{**} \ell^+ \nu_\ell$ decays, not all of the possible final states have been measured yet, thus their total branching fraction is unknown. We estimate it by extrapolating from existing measurements to the missing $D^{**}$ final state decays, assuming isospin symmetry. Among the gap modes, only $B \to D^{(*)} \pi \pi \ell \nu$ is measured in one charge configuration [32]. This measurement is extrapolated to the other charge configurations to estimate its total branching fraction. The remaining gap modes, $B \to D^{(*)} \eta \ell \nu$, are assigned with a 100% branching fraction uncertainty. All branching fraction uncertainties are propagated to the fit in the form of 1$\sigma$ event weight variations.
To estimate the effects of the uncertainties on the form factor parameters, we rotate each form factor parameter set into its eigenbasis and extract the \( \pm 1\sigma \) variation of each eigenvalue. These parameter variations lead to changes in the shape of the templates that we incorporate as event weights calculated as

\[
  w_i = \frac{\Gamma_{MC}}{\Gamma_{\text{new}}} \cdot \frac{d\Gamma_{\text{new}}}{d\Gamma_{\text{MC}}},
\]

with \( d\Gamma \) denoting the differential decay rate and \( \Gamma \) the total decay rate of the form factor model with the given parameters. We perform these calculations with the HAMMER software package \[33\].

We consider tracking efficiency uncertainties by assigning a flat uncertainty of 0.3% per track in the \( X \) system. Their effect emerged to be insignificant. Furthermore, we tested the effect of uncertainties of the secondary \( X_c \) decays in the \( X \) system and found it to be negligible due to the comparatively low momentum of leptons in these decays.

5. RESULTS AND DISCUSSION

Figure 2 shows the pre-fit and post-fit distributions of \( p^*_\ell \) for both signal and control regions. Table I summarizes the extracted relative yield factors in the the wrong charge control mode. The \( B \rightarrow X\ell\nu \) pre- and post-fit yield ratios agree with unity. We therefore consider the scaling factors of the background templates to be reliable.

<table>
<thead>
<tr>
<th>Template</th>
<th>Electron channel</th>
<th>Muon channel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-fit yield</td>
<td>Post-fit yield</td>
</tr>
<tr>
<td>Continuum</td>
<td>251</td>
<td>240 ± 51</td>
</tr>
<tr>
<td>Background</td>
<td>1736</td>
<td>2115 ± 238</td>
</tr>
<tr>
<td>( X\ell\nu )</td>
<td>5235</td>
<td>5143 ± 242</td>
</tr>
</tbody>
</table>

TABLE I. The pre-fit and post-fit yields of the different templates in the wrong charge channel. The relative factor between these yields is presented and the extracted factor for the background template is highlighted.

The fit yields in the signal regions are outlined in Table II.

<table>
<thead>
<tr>
<th>Template</th>
<th>Electron channel</th>
<th>Muon channel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asimov fit</td>
<td>Fitted yield</td>
</tr>
<tr>
<td>Continuum</td>
<td>450 ± 68</td>
<td>460 ± 65</td>
</tr>
<tr>
<td>Fakes &amp; other</td>
<td>1533 ± 167</td>
<td>1467 ± 166</td>
</tr>
<tr>
<td>( Xe\nu )</td>
<td>48786 ± 286</td>
<td>48034 ± 286</td>
</tr>
</tbody>
</table>

TABLE II. The nominal yield values from the model expectation and the yields resulting from the fit are presented for both lepton signal regions.
FIG. 2. The pre-fit (left) and post-fit (right) distributions of the lepton momentum in the $B_{\text{sig}}$ frame $p_t^*$ in the wrong charge (top) and correct charge (bottom) case. The electron and muon templates are fitted simultaneously in 10 $p_t^*$ bins with a width of 100 MeV each covering a $p_t^*$ range from 1.3 GeV to 2.3 GeV. The last bin of each lepton flavor is extended to account for any higher momenta.

From these, $R(X_e/\mu)$ can be calculated as:

$$R(X_e/\mu) = \frac{\epsilon_{X\mu\nu} \cdot N_{Xe\nu}}{\epsilon_{X\mu\nu} \cdot N_{X\mu\nu}} \quad \text{with} \quad \epsilon_{X\ell\nu} = \frac{N_{\text{sel}}^\ell \cdot \epsilon_{\text{Data/MC}}}{2 \cdot N_{B\bar{B}} \cdot B(B \to X\ell\nu)}$$ (6)

Here, $N_{B\bar{B}}$ is the number of $B\bar{B}$ events produced in the simulation. $N_{\text{sel}}^\ell$ is the number of $B \to X\ell\nu$ events that pass the full selection. The introduced data-to-MC scaling factors to
account for discrepancies in the modeling of FEI tagging efficiencies $\epsilon_{\text{Data/MC}}$ are excluded from $N_{\text{sel}}^{\ell}$. As they are independent of the signal-side lepton flavor, they cancel in the $R(X_{e/\mu})$ ratio. The statistical and systematic uncertainties of $N_{\text{sel}}^{\ell}$, and thus on the efficiencies, are considered with their correlations across the electron and muon channels, and are propagated to the uncertainty of $R(X_{e/\mu})$ together with the uncertainties and correlations of the post-fit yields $N_{X\ell\nu}$.

We find a $R(X_{e/\mu})$ value of:

$$R(X_{e/\mu}) = 1.033 \pm 0.010 \pm 0.020$$  \hspace{1cm} (7)

To the best of our knowledge, this is the most precise single test of $e - \mu$ flavor universality in semileptonic $B$ decays and it is consistent with unity within $1.5\sigma$. The Standard Model predicts a small excess of electrons of the order of $O(10^{-3})$ in exclusive ratios $[34, 35]$. Thus, a theoretical prediction of the inclusive ratio will agree with our measurement even better. Our results also agree within 0.6$\sigma$ with a previous measurement documented from Belle $[36]$ in exclusive $B \to D^*\ell\nu$ decays.

The impact of each systematic uncertainty on the result is estimated by performing Asimov fits to the model templates. Each systematic uncertainty is turned on and off, and the resulting uncertainties are compared to the total ones when including all systematics. Assuming that they add up in quadrature, their relative importance can be inferred. This is summarized in Table III.

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Rel. unc. of $R(X_{e/\mu})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>lepton identification eff.</td>
<td>1.8%</td>
</tr>
<tr>
<td>$X_e\ell\nu$ BF</td>
<td>0.1%</td>
</tr>
<tr>
<td>$X_e\ell\nu$ FF</td>
<td>0.2%</td>
</tr>
<tr>
<td>Statistical</td>
<td>1.0%</td>
</tr>
<tr>
<td>Total</td>
<td>2.2%</td>
</tr>
</tbody>
</table>

TABLE III. The $R(X_{e/\mu})$ uncertainties caused by the most important systematic uncertainties relative to the $R(X_{e/\mu})$ value.

The main source of uncertainty are the lepton identification efficiency uncertainties, which account for 1.8% of the total uncertainty. While branching fraction (BF) and form factor (FF) uncertainties are non-negligible for the total $B \to X\ell\nu$ yields, they cancel out in the $R(X_{e/\mu})$ ratio and therefore do not contribute. Statistical uncertainties are found to be relatively small.

Stemming from this result, the path is paved to expand our attention to the low lepton momentum range in future. In that scenario, with a multi-dimensional fit that also takes the $X$ system properties into account, a measurement of:

$$R(X) = \frac{\mathcal{B}(B \to X\tau\nu)}{\mathcal{B}(B \to X\ell\nu)}$$  \hspace{1cm} (8)
would become possible to probe the measured discrepancies in $R(D^{(*)})$ in an orthogonal way. With the size of the dataset analyzed for this work, we expect the precision on $R(X)$ to be still dominated by the statistical uncertainty, which is predicted to be of roughly 13%. 
6. ACKNOWLEDGEMENTS

We thank the SuperKEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid and the KEK computer group for on-site computing support.

[11] ALEPH collaboration, R. Barate, et al., ALEPH, *Measurements of $BR (b \to \tau^- \bar{\nu}_\tau X)$ and $BR (b \to \tau^- \bar{\nu}_\tau D^{*+} X)$ and upper limits on $B^- (B^- \to \tau^- \bar{\nu}_\tau)$ and $B \to s \bar{\nu} \nu$,* Eur. Phys. J. C19 (2001) 213–227, arXiv:hep-ex/0010022 [hep-ex]
[12] DELPHI collaboration, P. Abreu, et al., DELPHI, *Upper limit for the decay $B^- \to \tau^{-}$ anti-neutrino ($\tau$) and measurement of the $b \to \tau$ anti-neutrino ($\tau$) $X$ branching ratio*, Phys. Lett. B496 (2000) 43–58


