Search for $B^0 \rightarrow \pi^- \tau^+ \nu_\tau$ with hadronic tagging at Belle

P. Hamer et al. (Belle Collaboration)

Phys. Rev. D 93, 032007 — Published 22 February 2016

DOI: 10.1103/PhysRevD.93.032007
Search for $B^0 \rightarrow \pi^- \tau^+ \nu_\tau$ with hadronic tagging at Belle


1 University of the Basque Country UPV/EHU, 48080 Bilbao
2 University of Bonn, 53115 Bonn
3 Basque Institute of Nuclear Physics (BAS, UPV/EHU, 48080 Bilbao
4 Faculty of Mathematics and Physics, Charles University, 112 16 Prague
5 Chonnam National University, Kwangju 660-701
6 University of Cincinnati, Cincinnati, Ohio 45221
7 Deutsches Elektronen-Synchrotron, 22607 Hamburg
8 Justus-Liebig-Universität Gießen, 35392 Gießen
9 II. Physikalisches Institut, Georg-August-Universität Göttingen, 37073 Göttingen
10 SOKENDAI (The Graduate University for Advanced Studies), Hayama 240-0193
11 Hanyang University, Seoul 133-791
12 University of Hawaii, Honolulu, Hawaii 96822
13 High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801
14 IKERBASQUE, Basque Foundation for Science, 48013 Bilbao
15 Indian Institute of Technology Bhubaneswar, Satya Nagar 751007
16 Indian Institute of Technology Guwahati, Assam 781039
17 Indian Institute of Technology Madras, Chennai 600036
18 Indiana University, Bloomington, Indiana 47408
19 Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049
20 Institute of High Energy Physics, Vienna 1050
21 Institute for High Energy Physics, Protvino 142281
22 INFN - Sezione di Torino, 10125 Torino
23 J. Stefan Institute, 1000 Ljubljana
24 Kyungpook National University, Daegu 702-701
The decay $B^0 \rightarrow \pi^- \tau^+ \nu_\tau$ is mediated by the $W^+$ boson via the $b \rightarrow \bar{u}$ transition. The transition amplitude is described by \[ f^+(q^2) \left[ 2p_\mu + \left( 1 - \frac{m_B^2 - m_{\pi}^2}{q^2} \right) q_\mu \right] + f^0(q^2) \frac{m_B^2 - m_{\pi}^2}{q^2} q_\mu, \] with $p$ and $q$ being the momentum transfers to the pion and lepton pair, respectively.

The form factors $f^+$ and $f^0$ can be computed from QCD light-cone sum rules for $q^2 < 16 \text{ GeV}^2/c^4$ and lattice QCD for $q^2 > 16 \text{ GeV}^2/c^4$. Various parametrizations exist to interpolate between the two regions. In this study, we use the parametrization introduced by Bourrely, Caprini, and Lellouch (BCL) for which can describe both form factors in $m_\tau^2 \leq q^2 \leq (m_B - m_\pi)^2$. The parameter values are taken from Ref. \[.\]

It has been stated that the differential ratio \[ \frac{d\Gamma(B \rightarrow \pi\tau\nu_\tau)}{dq^2}, \quad \ell = e, \mu \]
can be used as a test for the Standard Model (SM) as it
depends solely on the ratio of the scalar and vector form
factors $f^0/f^+$. The CKM matrix $V_{ub}$ element $|V_{ub}|$ enters
both differential branching fractions and cancels in the
ratio.

In new physics models like the two-Higgs-doublet model (2HDM) \[10, 11\], the decay $B^0 \rightarrow \pi^- \tau^+ \nu_\tau$ can also be mediated by a charged Higgs boson. Possible
contribution of a $H^+$ and other couplings in the 2HDM
and MSSM \[12, 13\], which would affect the branching
fraction and the differential ratio of branching fractions,
have been evaluated in Refs. \[2\] and \[14-17\].

The decay $B^0 \rightarrow \pi^- \tau^+ \nu_\tau$ has not been ob-
erved, nor has an upper limit on the branching fraction
been obtained. Recent results \[6\] on the
two form factors obtained from a joint fit to $(2+1)$-
flavor lattice QCD calculations and $B \rightarrow \pi \nu$ data
from Belle \[18, 19\] and BaBar \[20, 21\] result in
$B(B^0 \rightarrow \pi^- \tau^+ \nu_\tau)/B(B^0 \rightarrow \pi^\mp \ell^\pm \nu_\ell) = 0.641(17)$ and
$B(B^0 \rightarrow \pi^- \tau^+ \nu_\tau) = 9.35(38) \times 10^{-5}$ \[22\].

The signal decay is reconstructed in the four one-prong
decays of the $\tau$ lepton, $\tau^- \rightarrow \ell^- \nu_\ell \nu_\tau$, with $\ell = e$ or $\mu$, $\tau^- \rightarrow \pi^- \nu_\tau$, and $\tau^- \rightarrow \rho^- \nu_\tau$, corresponding to
72\% of all $\tau$ decays \[23\]. The most powerful decay modes
are the two aforementioned hadronic $\tau$ decays and the
$\tau^- \rightarrow e^- \nu_\tau \nu_\tau$ mode, while the $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ decay mode
does not improve the final expected significance. This is
mainly due to low muon momenta in the signal decay
and the resulting low muon identification efficiency.
The result of this analysis is based on the three most powerful
$\tau$ decay modes.

II. DATA SAMPLE

The search for $B^0 \rightarrow \pi^- \tau^+ \nu_\tau$ described in this paper is
performed on the full data sample collected with the Belle
detector at the KEKB asymmetric-energy $e^+ e^-$ (3.5 on
8.0 GeV) collider \[24\], operating at the $\Upsilon(4S)$ resonance.
The data sample consists of an integrated luminosity of
711 fb$^{-1}$, which corresponds to $(771.6 \pm 10.6) \times 10^6 B \bar{B}$
pairs.

The Belle detector is a large-solid-angle magnetic spec-
trometer that consists of a silicon vertex detector (SVD),
a 50-layer central drift chamber (CDC), an array of aeros-
gel threshold Cherenkov counters (ACC), a barrel-like
arrangement of time-of-flight (TOF) scintillation coun-
ters, and an electromagnetic calorimeter (ECL) com-
prised of CsI(Tl) crystals located inside a superconduct-
ing solenoid coil that provides a 1.5 T magnetic field.
An iron flux-return yoke located outside of the coil is in-
strumented to detect $K^0_L$ mesons and to identify muons
(KLM). Two inner detector configurations were used.
A 2.0 cm beampipe and a 3-layer SVD were used for the first
sample of $152 \times 10^6 B \bar{B}$ pairs, while a 1.5 cm beampipe,
a 4-layer silicon detector and a small-cell inner drift cham-
ber were used to record the remaining $620 \times 10^6 B \bar{B}$
pairs \[25\]. The detector is described in detail in Ref. \[26\].

The study was performed as a blind analysis based on
simulated data. Monte Carlo (MC) samples were gener-
ated with EvtGen \[27\] and the detector simulation was
performed by GEANT3 \[28\]. Recorded beam background
was added to the MC samples. The expected non-beam
background is estimated using MC samples that describe
all physics processes at Belle. A resonant $\Upsilon(4S)$ event at
Belle produces a $B \bar{B}$ pair. Two samples of $b \rightarrow c$ decays
for $B^0 \bar{B}^0$ and $B^+ \bar{B}^-$ events, respectively, each contain
10 times the integrated luminosity of the data sample.
Semileptonic $b \rightarrow u$ decays are simulated in a sample contain-
ing 20 times the integrated luminosity. Rare $b \rightarrow s$
and other rare decays are described in another sample
corresponding to 50 times the integrated luminosity of the data.
Continuum $e^+ e^- \rightarrow q \bar{q}$ ($q = u, d, s, c$) was gen-
erated with PYTHIA \[29\] and included in the analysis in
an MC sample containing five times the integrated lumin-
osity of the data sample. Additionally, a high statistics
sample of $B^0 \rightarrow X_s \tau \nu$ containing $24 \times 10^6$ events
was generated with a phase-space and ISGW2 \[30\] model.

The signal MC sample is generated using BCL results
for the vector and scalar form factors \[2\]. A total of
$84 \times 10^6 B^0 \bar{B}^0$ events were generated with one meson
decaying into the signal final state and the other decaying
generically.

No constraints on the $\tau$ decay were applied. The signal
MC sample corresponds to approximately 2000 times the
expected $B(B^0 \rightarrow \pi^- \tau^+ \nu_\tau) = 9.35 \times 10^{-5}$.

III. EVENT SELECTION

The complete reconstruction of the $B$ meson decay into
the signal final state ($B_{\text{sig}}$) is not possible due to the
presence of at least two neutrinos. However, since the initial
state of the $e^+ e^-$ collision is completely defined by the
momenta of the colliding leptons, we can constrain the
signal side by fully reconstructing the other $B$ meson
($B_{\text{tag}}$) in hadronic decay modes. Tracks and clusters in
the event that are not assigned to the $B_{\text{tag}}$ after the suc-
cessful reconstruction are assumed to originate from $B_{\text{sig}}$.

A. Tag side

This analysis uses the Belle hadronic full-
reconstruction algorithm \[31\] based on the artificial
Neural Network package NeuroBayes \[32\]. Neural net-
works were trained to reconstruct $B^0$ and $B^+$ candidates
from a total of 1104 decay channels.

Additional event shape variables are added to sup-
press continuum events. $B$ mesons of resonant events are
nearly at rest in the center-of-mass (c.m.) frame, leading
to a spherical distribution of their decay products. Con-
tinuum events, on the other hand, produce back-to-back jets
due the large available kinetic energy. Useful ob-
servables that differentiate between the two event types
are the thrust axis of the $B_{\text{tag}}$ meson \[33\] and modified
Fox-Wolfram moments \[^{34}\]. For this analysis, the thrust axis and the second modified Fox-Wolfram moment are included in the neural network for the full hadronic reconstruction. If the algorithm does not succeed in reconstructing a \(B^0\) candidate, the event is discarded.

Differences in the full reconstruction efficiency between MC and data, depending on the network output and \(B_{\text{tag}}\) reconstruction channel, are observed \[^{19}\] but depend on the tag-side reconstruction only; a correction factor is determined from charmed semileptonic decays of the signal-side \(B\) meson.

The beam-energy-constrained mass, 
\[
M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - (\vec{p}_{B_{\text{tag}}}c)^2 / c^2},
\]
is required to be greater than 5.27 GeV/c\(^2\), where \(E_{\text{beam}}\) and \(\vec{p}_{B_{\text{tag}}}\) denote the beam energy and reconstructed three-momentum, respectively, of the \(B_{\text{tag}}\), evaluated in the \(e^+e^-\) c.m. frame. With this requirement and the correction factor applied, we estimate a reconstruction efficiency of 0.3\% from the signal MC sample, which is in very good agreement with the reconstruction efficiency of \(B^0\) mesons in the Belle data sample \[^{31}\]. The neural network output, \(\phi_{\text{tag}}^{\text{cs}} \in [0, 1]\), is a continuous variable whose high (low) values correspond to candidates which are likely (unlikely) to be a true \(B\) meson. It is used at a later selection stage, as described below. The distributions of \(M_{\text{bc}}\) and \(\ln(\phi_{\text{tag}}^{\text{cs}})\) for the three reconstruction channels are shown in Fig. \[^{11}\] where the green (solid) contribution shows the dominant charmed \(B \rightarrow X_c\) background decays. No further requirements are applied to the \(M_{\text{bc}}\) distributions, while \(M_{\text{bc}} > 5.27\) GeV/c\(^2\) is required for the \(\ln(\phi_{\text{tag}}^{\text{cs}})\) distributions.

### B. Signal side

Only one-prong decays of the \(\tau\) lepton are considered in this search. For a correctly reconstructed \(B_{\text{tag}}\), there should be exactly two remaining oppositely-charged tracks in the detector. Additionally, the event should contain undetected (missing) momentum. Since the initial state of the \(e^+e^-\) collisions is given by the four-momenta of the colliding leptons, the undetected momentum can be measured. The missing momentum is defined as

\[
p_{\text{miss}} = 2p_{\text{beam}} - \vec{p}_{B_{\text{tag}}} - \vec{p}_{\text{vis}},
\]

where \(2p_{\text{beam}} = p_{e^+} + p_{e^-}\) is twice the beam momentum and \(B_{\text{vis}}\) denotes the visible part of the \(B_{\text{sig}}\) meson. Tracks with low transverse momentum \(p_t\) can curl in the solenoidal field and be detected as two tracks with opposite charge. Any two tracks with \(p_t < 275\) MeV/c with an angle between the momentum vectors, calculated at the points closest to the interaction point, below 15\(^{\circ}\) and total momentum difference less than 100 MeV/c are therefore counted as a single track. We reduce the number of poor quality tracks by requiring that \(|dr| < 2.0\) cm and \(|dz| < 4.0\) cm, where \(|dz|\) and \(|dr|\) are the distances of closest approach of a track to the interaction point along the \(z\)-axis and in the transverse plane, respectively.

Electron identification \[^{53}\] is performed by calculating a likelihood ratio using the matching of charged tracks with the shower position in the ECL, the shower shape, the ratio of the energy deposited in the ECL and the measured momentum, the energy loss \(dE/dx\) in the CDC, and the Cherenkov light production in the ACC. Muon identification \[^{50}\] is also done by evaluating a likelihood ratio. Clusters in the KLM are matched to charged tracks by extrapolation. For matched tracks, the difference between expected and measured penetration depth and the transverse deviation of all KLM hits associated with the track are used in this likelihood ratio. For a charged track not identified as a lepton, a kaon veto is applied using a likelihood ratio that discriminates between kaons and pions \[^{57}\]. The ratio is formed from the energy loss \(dE/dx\) in the CDC, flight time information from the TOF, and photon yield in the ACC. All remaining tracks are identified as a pion. Neutral pions are reconstructed from pairs of photons. The absolute difference between the invariant mass of the \(\pi^0\) candidate and the nominal \(\pi^0\) mass, normalized to its uncertainty, must be below 3.0. Photons are required to have energies in the laboratory frame greater than 50 MeV for the ECL barrel and 100(150) MeV for the forward (backward) endcap. Neutral pion candidates with at least one photon being used in the tag-side full reconstruction are discarded.

Events are required to have exactly two oppositely charged particles within the allowed impact parameter range, with one additional track allowed outside the range. At least one charged pion is required. If the event contains two charged pions and neutral pion candidates, we search for \(\rho^\pm\) candidates. The charged pion with the lower momentum in the c.m. frame is combined with each neutral pion candidate and a mass vertex fit is performed. A pair that can be successfully fitted with \(\chi^2 < 20\) is accepted as a charged \(\rho^\pm\) meson if its invariant mass is between 625 and 925 MeV/c\(^2\). If multiple candidates are found, the candidate with a mass closest to the nominal \(\rho^\mp\) mass \[^{23}\] is selected. Due to the broad \(\rho^\pm\) mass range, not all \(\tau^- \rightarrow \rho^- \nu_\tau\) events are correctly reconstructed. These events contain two oppositely charged pions and a neutral pion in the final state and are misidentified in the \(\tau^- \rightarrow \pi^- \nu_\tau\) channel. Each event is reconstructed in one of the four reconstruction channels. In many \(\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau\) events, the momentum of the muon is too low to reach the KLM and thus is not identified as a muon. In most of these cases, the muon is identified as a pion so that the event is placed in the \(\tau^- \rightarrow \pi^- \nu_\tau\) sample.

Since \(K_L\) mesons do not completely deposit their energy in the detector, charm \(B\) decays with subsequent decays \(D \rightarrow K_L\pi\) or \(D \rightarrow K_L\nu_\tau\) have the signal’s missing-momentum signature. A \(K_L\) candidate is identified as a cluster in the KLM without an associated
charged track. An ECL cluster without an associated charged track is associated with the K_L cluster in the KLM if it lies along the flight path extrapolated from the interaction point to the KLM cluster. As described below, the extra energy in the ECL is used to determine the signal yield. Therefore, only events with a K_L without energy deposition in the ECL are vetoed.

C. Extra energy

We extract the signal yield from a fit to the distribution of the energy deposited in the ECL (E_{ECL}) by particles not used in the full reconstruction or by the two remaining charged signal tracks. To reduce background, the aforementioned photon energy requirements are applied. For signal decays, there is no additional energy deposition, so E_{ECL} peaks strongly at zero. Misreconstructions of B_{tag} lead to a small tail towards higher energy depositions for true signal events. In contrast, most background decays exhibit non-vanishing extra energy due to the presence of additional neutral particles.

D. Boosted decision trees

Final event selection uses requirements on three variables: ln(o_{tag})^2; missing mass squared (M_{miss}^2 = p_{miss}^2); and the output of the boosted decision tree (BDT). For each τ reconstruction mode, one BDT is trained using the TMVA framework. All use different input variables, background training samples, and BDT growth parameters. The signal training sample consists of 3 × 10^7 events out of the complete signal MC sample. To improve the training, events are required to have E_{ECL} < 1 GeV and B_{tag} is required to have a quality of ln(o_{tag}^2) > −7. One additional track outside the impact parameter requirement is allowed.

Another 3 × 10^7 signal events are used for performance tests of the BDT for receiver-operation characteristics (ROC) calculation and overtraining evaluation.

The input variables of the BDT used in the \( \tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau \) selection are the magnitude of the three-momenta of the pion and electron, the squared lepton-pair momentum transfer \( q^2 \), \( M_{miss}^2 \), and different combinations of all available four-momenta. The momentum transfer can be calculated using the fact that both B mesons are at rest in the c.m. frame, which implies \( p_{B_{tag}} = \bar{p}_{B_{tag}} \) and \( q = p_{B_{tag}} - p_{\tau} \). Due to the low efficiency of the full reconstruction, an additional signal sample is used with 2 × 10^7 \( \tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau \) events on the signal side. The
background training sample consists of charmed \( B^0 \) decays and \( B^0 \to X_u \ell \nu \ell \) decays. The input variables are linearly decorrelated before their use in the training.

The background sample used in the \( \tau^- \to \pi^- \nu_\tau \) selection BDT contains \( b \to c \) decays and semileptonic \( b \to u \) decays of \( B^0 \) mesons. Principal-component analysis (PCA) \(^{35}\) is applied to the input variables. PCA is an orthogonal transformation which rotates a sample of data points such that the variability along the new axis is maximized. In this way, the variables are decorrelated.

The input variables are \( M_{\text{miss}}^2 \), the missing energy, \( q^2 \), the absolute three-momentum of \( B_{\text{vis}} \) in the c.m. frame, combinations of available four-momenta, and the number of unused neutral pions in the event.

The BDT training for the \( \tau^- \to \rho^- \nu_\tau \) selection uses the same sample size of \( b \to c \) decays, but not semileptonic \( b \to u \) decays. The correlation of the input variables is again reduced by a PCA transformation. The variables used in the training are \( M_{\text{miss}}^2 \), the missing energy, \( q^2 \), and combinations of the available four-momenta in the decay.

The performance of the three final BDTs is shown in form of the ROC curves in Fig. 2.

![Background rejection versus Signal efficiency](image)

**FIG. 2:** Background rejection versus signal efficiency determined on the testing sample, from Belle simulation.

### E. Final selection

The final selection criteria are determined from MC samples by maximizing individually the expected significance of each single \( \tau \) reconstruction mode. We perform a scan over three variables simultaneously to obtain the optimal selection: \( \ln(o^\tau_{\text{tag}}), M_{\text{miss}}^2 \), and the BDT output. We require \( \ln(o^\tau_{\text{tag}}) > -7 \) for the leptonic \( \tau \) reconstruction and \( \ln(o^\tau_{\text{tag}}) > -5 \) for the hadronic \( \tau \) reconstruction, as shown in Fig I. A minimum requirement on the \( M_{\text{miss}}^2 \) is applied to reject semileptonic \( B \to \pi \ell \nu \ell \) events, which have the same final state as the signal decay: since no energy is deposited in the ECL, decays of this type peak at zero extra energy. Also, as there is only a single neutrino in these decays, \( M_{\text{miss}}^2 \) peaks at zero, unlike the case for signal decays, which contain at least two neutrinos. We require \( M_{\text{miss}}^2 \) to be greater than \( 2.2 \text{ GeV}/c^2 \) in the electron channel, \( 0.0 \text{ GeV}/c^2 \) in the pion channel, and \( 0.6 \text{ GeV}/c^2 \) in the \( \rho \) channel. The BDT output is the last variable used in the scan. The expected significance is calculated as \( \sqrt{-2 \ln(L_{\text{obs}}/L_{\text{ex}})} \); the likelihood is given by \( L_k = \prod_i \text{Poisson}(n_{\text{obs}}|n_{\text{sig}} + k \cdot n_{\text{sig}}) \), where \( n_{\text{obs}} = n_{\text{tag}} + n_{\text{sig}} \) is the best estimate from the MC samples.

The efficiency of the final selection is determined from MC to be \( 4.57 \times 10^{-4} \). The dominant reconstructed \( \tau \) decay modes and their relative occurrences are listed in Table I.

**TABLE I:** Signal reconstruction by \( \tau \) decay modes. Percentages are obtained from signal MC and sum to 100%.

<table>
<thead>
<tr>
<th>( \tau^- ) decay</th>
<th>Relative Occurrence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho^- \nu_\tau )</td>
<td>29.54</td>
</tr>
<tr>
<td>( e^- \bar{\nu}<em>e \nu</em>\tau )</td>
<td>29.43</td>
</tr>
<tr>
<td>( \pi^- \nu_\tau )</td>
<td>16.70</td>
</tr>
<tr>
<td>( \mu^- \bar{\nu}<em>\mu \nu</em>\tau )</td>
<td>13.21</td>
</tr>
<tr>
<td>( a_1^- \nu_\tau )</td>
<td>8.72</td>
</tr>
<tr>
<td>other</td>
<td>2.4</td>
</tr>
</tbody>
</table>

The dominant background in the low-\( E_{\text{ECL}} \) region arises from \( B^0 \to D^{(*)}\ell \nu_\ell \) and \( B^0 \to D^{(*)}\rho \) decays with a subsequent decay of \( D \to K_L \pi \). The \( K_L \) is undetected in these cases and the resulting decay signature resembles that of the signal. No explicit selection is available to further suppress decays of this type.

### IV. SYSTEMATIC UNCERTAINTIES

In the computation of the significance level and upper limit, systematic uncertainties are included in the likelihood as nuisance parameters. The likelihood is built from probability density functions (PDFs) determined from MC predictions of each background sample, as described in Section V. All systematic uncertainties are assumed to be Gaussian-distributed and are evaluated at one standard deviation (\( \sigma \)).

Uncertainties of the particle identification and of the correction factor needed for the full reconstruction efficiency are included as a flat effect over all bins in \( E_{\text{ECL}} \). All other uncertainties are included in a bin-by-bin fashion. A constant uncertainty of 0.35% has been determined for each charged track with \( p_t > 0.2 \text{ GeV}/c \). Tracks below that threshold have to be treated differently depending on the track momentum \(^{35}\). The uncertainty on the number of produced \( B \)-meson pairs is 1.4%. The uncertainty due to the \( K_L \) veto is determined by vary-
ing the $K_L$ efficiency by its uncertainty. The branching
fractions of the dominant backgrounds are varied by
their errors stated in Ref. 23 to determine the effect
on the MC prediction. The uncertainty on the correc-
tion factor of the tagside reconstruction is determined in
Ref. 19 and applied to the samples. The discrepancy
between inclusive and exclusive $|V_{ub}|$ measurements has
been included as a flat but asymmetric uncertainty in
the $B \to X_u \ell \nu$ sample of $^{+5}_{-15} \%$. An uncertainty of
$\pm 10\%$ is applied to the branching fractions in the MC
sample of rare $b \to s$ and other rare $B$ decays. Addi-
tionally, decays of type $B \to X_u \tau \nu$ are present in the
final event selection. The contribution to the $E_{\text{CL}}$ dis-
tribution is evaluated from the MC sample assuming a
$\mathcal{B}(B^{0} \to \rho^{\pm} \tau^{-} \nu_{\tau}) = 1.5 \times 10^{-4}$ and found to be small; a
relative uncertainty of $\pm 50\%$ is applied. Statistical un-
certainties in the PDF shape due to finite MC sample
size are included in a way similar to the approach by
Barlow and Beeston 40. Instead of using one Poisson
constraint per background sample per bin per $\tau$ decay
channel, only one constraint term per bin per channel is
used. The uncertainty introduced by this approximation is
negligible for bins with non-vanishing content and re-
duces the amount of computation time needed. Instead
of the finite MC uncertainty, the fit error is included as
a systematic uncertainty for the dominant $b \to c$ contribu-
tion. The theoretical uncertainties of the signal form
factors $f^+$ and $f^0$ are included by generating additional
signal MC with one form factor fixed and the other varied
by its $1\sigma$ uncertainty. The relative uncertainties deter-
mined in this way are combined into a single uncertainty
estimate. The systematic uncertainties due to the track-
ing efficiency and particle identification affect only the
overall efficiency and are only included in the calculation
of the upper limit. The relative effect on the branching
fraction is determined by repeatedly fitting modified
PDFs to data. The PDFs are modified by replacing each
background contribution with the respective contribution
where the systematic effect is applied. For each system-
atic uncertainty, two fits are performed for the positive
and negative deviation. The maximum, absolute devia-
tion is quoted in Table IX.

\section*{V. RESULT}

A binned maximum likelihood fit is performed to $E_{\text{CL}}$
in bins of 0.15 GeV. Due to similar shapes in the back-
ground predictions, all background contributions except
for the dominant $b \to c$ transitions are fixed to the MC
prediction. Possible errors introduced by this approach
are accounted for as systematic uncertainties. The fit
is performed simultaneously in all three reconstruction
modes. The signal strength parameter $\mu$ is constrained
between the three modes while the background contribu-
tions of the three reconstruction modes are floating pa-
rameters. The fit result of the $B^0 \to X_c$ background con-
tribution agrees well with the prediction obtained from
the MC sample. The signal strength has been chosen
such that $\mu = 1.0$ corresponds to $\mathcal{B}(B^{0} \to \pi^{\pm} \tau^{\mp} \nu_{\tau}) = 1.0 \times 10^{-4}$. We obtain a best fit of $\mu = 1.52 \pm 0.72$, corre-
sponding to $51.9 \pm 24.3$ signal events. The fit results by $\tau$ reconstruction mode are listed in Table II. The $E_{\text{CL}}$

table II: Effects of the systematic uncertainties on
the branching fraction.

<table>
<thead>
<tr>
<th>Source</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle ID</td>
<td>2.4</td>
</tr>
<tr>
<td>Track efficiency</td>
<td>0.7</td>
</tr>
<tr>
<td>$N(B\bar{B})$</td>
<td>1.4</td>
</tr>
<tr>
<td>$K_L$ veto</td>
<td>3.2</td>
</tr>
<tr>
<td>BG $B$</td>
<td>2.8</td>
</tr>
<tr>
<td>Tagside</td>
<td>4.6</td>
</tr>
<tr>
<td>$</td>
<td>V_{ub}</td>
</tr>
<tr>
<td>Rare processes</td>
<td>2.0</td>
</tr>
<tr>
<td>$B \to X_u\tau\nu$</td>
<td>2.2</td>
</tr>
<tr>
<td>Background fit</td>
<td>0.2</td>
</tr>
<tr>
<td>Signal model</td>
<td>1.8</td>
</tr>
<tr>
<td>Total</td>
<td>8.3</td>
</tr>
</tbody>
</table>

TABLE III: Fit results for signal yield. Total and split
by $\tau$ reconstruction mode.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Signal Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e$</td>
<td>13.2 $\pm$ 6.2</td>
</tr>
<tr>
<td>$\pi$</td>
<td>30.6 $\pm$ 14.3</td>
</tr>
<tr>
<td>$\rho$</td>
<td>8.1 $\pm$ 3.8</td>
</tr>
<tr>
<td>Total</td>
<td>51.9 $\pm$ 24.3</td>
</tr>
</tbody>
</table>

distribution and fit results are shown in Fig. 3.

The significance of the measurement is obtained from
a pseudo MC study. A test statistic based on the profile
likelihood ratio is used. The likelihood is built in bins of
0.15 GeV in $E_{\text{CL}}$. The binned likelihood is given by

$$
\mathcal{L} = \prod_{c} \prod_{b} \text{Poisson}(n_{cb}|\nu_{cb}) \cdot \prod_{p \in \mathcal{P}} f_p(a_p|\alpha_p),
$$

where the indices $c$ and $b$ label the reconstruction chan-
nel and bin in $E_{\text{CL}}$, respectively, and $\mathcal{P}$ denotes
the set of systematic uncertainties $p$ that are included as
nuisance parameters $\alpha_p$ in the calculation of the number
of expected events $\nu_{cb}$ per channel per bin. The nuisance
parameters are parametrized as a relative effect on the
nominal template prediction, assumed to be Gaussian-
distributed with the nominal value being the global ob-
servable $a_p$. The number of events in the background-
only hypothesis is determined from MC simulation and
a fit to data for the dominant $b \to c$ background. The
likelihood is constructed using the HistFactory tool in
the RooStats package 41, 42.

The distribution of the test statistic is obtained by
pseudo-experiments. A full frequentist approach is used
in both the computations of the significance level and
FIG. 3: Distributions of $E_{ECL}$ in the three $\tau$ reconstruction modes. The signal and $b \to c$ contributions are scaled according to the fit result.

the upper limit. First, the likelihood is fitted to data to obtain the maximum likelihood estimates (MLEs) of all nuisance parameters on data. In each pseudo-experiment generation, the nuisance parameters are fixed to their respective MLE. In the subsequent maximization of the likelihood, the nuisance parameters are free parameters. The global observables are randomized in each pseudo-experiment.

Using pseudo-experiments, the $p$-value of the background-only hypothesis for data is determined and the significance level $Z$ is computed in terms of standard deviations as

$$Z = \Phi^{-1} (1 - p),$$

where $\Phi^{-1}$ is the cumulative distribution function of the standard normal Gaussian.

We observe a signal significance of $2.8\sigma$, not including systematic uncertainties in the calculation. Including all relevant systematic effects results in a significance of $2.4\sigma$. For this result, the test statistic has been computed on 10 000 background-only pseudo-experiments.

Given the level of significance of these results, we invert the hypothesis test and compute an upper limit on the branching fraction. Pseudo-experiments are generated for different signal strength parameters for both signal-plus-background and background-only hypotheses in order to obtain $CL_{s+b}$ and $CL_b$, respectively. The upper limit is then computed using $CL_s = CL_{s+b}/CL_b$ [43], where a scan over reasonable signal strength parameter values is performed. At each step, 10 000 pseudo-experiments have been evaluated for both hypotheses.

At the 90% confidence level, we obtain an upper limit of $B (B^0 \to \pi^- \tau^+ \nu_\tau) < 2.5 \times 10^{-4}$. The upper limit at the 95% confidence level has been computed to $B (B^0 \to \pi^- \tau^+ \nu_\tau) < 2.8 \times 10^{-4}$. This result is the first result on $B (B^0 \to \pi^- \tau^+ \nu_\tau)$ and is in good agreement with the SM prediction.
ACKNOWLEDGMENTS

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, the National Institute of Informatics, and the PNLL/EMSL computing group for valuable computing and SINET4 network support. We acknowledge support from 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; Austrian Science Fund under Grant No. P 22742-N16 and P 26794-N20; the National Natural Science Foundation of China under Contracts No. 10575109, No. 10775142, No. 10875115, No. 11175187, and No. 11475187; the Chinese Academy of Science Center for Excellence in Particle Physics; the Ministry of Education, Youth and Sports of the Czech Republic under Contract No. LG14034; the Carl Zeiss Foundation, the Deutsche Forschungsgemeinschaft and the VolkswagenStiftung; the Department of Science and Technology of India; and the Istituto Nazionale di Fisica Nucleare of Italy; the WCU program of the Ministry of Education, National Research Foundation (NRF) of Korea Grants No. 2011-0029457, No. 2012-0008143, No. 2012R1A1A2008330, No. 2013R1A1A3007772, No. 2014R1A2A2A0105586, No. 2014R1A2A2A01002734, No. 2015R1A2A2A01003280, No. 2015H1A2A1033649; the Basic Research Lab program under NRF Grant No. KRF-2011-0020333, Center for Korean J-PARC Users, No. NRF-2013K1A3A7A06056592; the Brain Korea 21-Plus program and Radiation Science Research Institute; the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Education and Science of the Russian Federation and the Russian Foundation for Basic Research; the Slovenian Research Agency; the Basque Foundation for Science (IKERBASQUE) and the Euskal Herriko Unibertsitatea (UPV/EHU) under program UFI 11/55 (Spain); the Swiss National Science Foundation; the National Science Council and the Ministry of Education of Taiwan; and the U.S. Department of Energy and the National Science Foundation. This work is supported by a Grant-in-Aid from MEXT for Science Research in a Priority Area (“New Development of Flavor Physics”) and from JSPS for Creative Scientific Research (“Evolution of Tau-lepton Physics”).

[1] Throughout this paper, the inclusion of the charge-conjugate mode process is implied unless otherwise stated.