Evidence for $B^- \rightarrow D_s^+ K^- \ell^- \bar{\nu}_\ell$ and search for $B^- \rightarrow D_s^+ K^- \ell^- \bar{\nu}_\ell$

J. Stypula, M. Rozanska, I. Adachi, K. Adamczyk, H. Alihara, D. M. Asner, T. Aushev, A. M. Bakich, V. Bhardwaj, B. Bhuyan, M. Bischofberger, A. Bondar, G. Bonvicini, A. Bozek, M. Bračko, T. E. Browder, M.-C. Chang, P. Chang, V. Chekelian, A. Chen, P. Chen, B. G. Cheon, R. Chistov, I.-S. Cho, Y. Choi, D. Dalenko, M. Danilov, J. Dingfelder, Z. Dolegal, A. Drutskoy, S. Eidelman, E. Fast, V. Gaur, N. Gabyshov, R. Gillard, M. Goh, B. Golob, J. Haba, K. Hayasaka, Y. Hayashii, Y. Horii, Y. Hoshi, W.-S. Hou, B. Y. Hsiung, H. J. Hyun, T. Iijima, K. Inami, A. Ishikawa, R. Itoh, M. Iwabuchi, Y. Iwasaki, T. Julius, J. H. Kang, P. Kapusta, T. Kawasaki, H. Kichimi, C. Kiesling, H. J. Kim, J. B. Kim, J. H. Kim, K. T. Kim, Y. J. Kim, K. Kinoshita, B. R. Ko, P. Kodyš, S. Korpar, R. T. Kouzes, P. Križan, P. Krokovny, T. Kuhr, T. Kumita, A. Kuzmin, Y.-J. Kwon, S.-H. Lee, J. Li, Y. Li, J. Libby, C. Liu, Y. Liu, Z. Q. Liu, D. Liventsev, R. Louvot, K. Miyabayashi, H. Miyata, Y. Miyazaki, R. Mizuk, G. B. Mohanty, A. Moll, N. Muramatsu, E. Nakano, M. Nakao, Z. Natkaniec, C. Ng, S. Nishida, K. Nishimura, O. Nitoh, T. Nozaki, O. Ogawa, T. Ohshima, S. Okuno, L. S. Olsen, Y. Onuki, P. Pakhlov, G. Pakhlova, C. Schwanda, K. Senyo, O. Seon, M. E. Sevior, M. Shapkin, C. P. Shen, T.-A. Shibata, J.-G. Shin, B. Shwert, A. Sibidanov, F. Simon, P. Smerkol, Y.-S. Sohn, A. Sokolov, E. Soloviev, S. Stanic, M. Starić, M. Sumihama, T. Sumiyoshi, Y. Teramoto, S. T. Sumo, P. Urquijo, Y. Usoskin, P. Vanhoefer, G. Varner, K. E. Varvell, V. Vorobyev, P. Wang, X. L. Wang, M. Watanabe, Y. Watanabe, J. H. Wetterich, M. Williams, E. Won, B. D. Yabsley, H. Yamamoto, Y. Yamashita, Z. P. Zhang, V. Zhilich, V. Zhulanov, A. Zupanc, The Belle Collaboration

1 University of Bonn, Bonn
2 Budker Institute of Nuclear Physics SB RAS and Novosibirsk State University, Novosibirsk 630090
3 Faculty of Mathematics and Physics, Charles University, Prague
4 University of Cincinnati, Cincinnati, Ohio 45221
5 Department of Physics, Fu Jen Catholic University, Taipei
6 Gifu University, Gifu
7 Hankyung University, Seoul
8 University of Hawaii, Honolulu, Hawaii 96822
9 High Energy Accelerator Research Organization (KEK), Tsukuba
10 Institute of Technology Guwahati, Guwahati
11 Institute of Technology Madras, Madras
12 Institute of High Energy Physics, Chinese Academy of Sciences, Beijing
13 Institute of High Energy Physics, Vienna
14 Institute of High Energy Physics, Protvino
15 Institute for Theoretical and Experimental Physics, Moscow
16 J. Stefan Institute, Ljubljana
17 Kanagawa University, Yokohama
18 Institut für Experimentelle Kernphysik, Karlsruher Institut für Technologie, Karlsruhe
19 Korea Institute of Science and Technology Information, Daejeon
20 Korea University, Seoul
21 Kyungpook National University, Taegu
22 École Polytechnique Fédérale de Lausanne (EPFL), Lausanne
23 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana
24 Luther College, Decorah, Iowa 52101
25 University of Maribor, Maribor
26 Max-Planck-Institut für Physik, München
27 University of Melbourne, School of Physics, Victoria 3010
28 Graduate School of Science, Nagoya University, Nagoya
29 Kobayashi-Maskawa Institute, Nagoya University, Nagoya
30 Nara Women’s University, Nara
We report measurements of the decays $B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}$ in a data sample containing $657 \times 10^{6}$ $B\bar{B}$ pairs collected with the Belle detector at the KEKB asymmetric-energy $e^{+}e^{-}$ collider. We observe a signal with a significance of 6σ for the combined $D_{s}$ and $D_{s}^{*}$ modes and find the first evidence of the $B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}$ decay with a significance of 3.4σ. We measure the following branching fractions: $B(B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}) = (0.30 \pm 0.09) \times 10^{-3}$ and $B(B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}) = (0.59 \pm 0.15) \times 10^{-3}$ and set an upper limit $B(B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}) < 0.56 \times 10^{-3}$ at the 90% confidence level. We also present the first measurement of the $D_{s}^{(*)} K^{-}$ invariant mass distribution in these decays, which is dominated by a prominent peak around 2.6 GeV/c$^2$.

PACS numbers: 13.20.He, 14.40.Nd

Semileptonic $B$ decays play a key role in testing the Standard Model (SM) and in the understanding of heavy quark dynamics. In particular, they are used to determine the weak mixing parameters $|V_{qb}|$ ($q = c, u$), complementing the measurements of $CP$ asymmetries used to verify the Cabibbo-Kobayashi-Maskawa (CKM) mechanism of the SM [1]. The tension at the level of 2 standard deviations ($\sigma$) between the values of $|V_{qb}|$ extracted from inclusive and exclusive $B$ decays [2], as well as some discrepancies between measurements and theoretical expectations for semileptonic $B$ decays to excited charmed mesons, may indicate problems in the theoretical tools or in the interpretation of the experimental results.

Semileptonic $B$ decays to final states containing a $D_{s}^{(*)+} K$ system [3] provide information about the poorly known region of hadronic masses above 2.46 GeV/c$^2$, covering radially excited $D$ meson states [3]. Further exploration of this region may help solving some puzzles in semileptonic $B$ decays [3]. Recently, BaBar reported an observation of $B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}$ (which did not distinguish between the $D_{s}$ and $D_{s}^{*}$ final states) with a branching fraction of $B(B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}) = (6.13^{+0.04}_{-0.03}) \times 10^{-4}$ [4].

In this paper, we present measurements of $B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}$ and $B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}$ decays using a data sample containing $657 \times 10^{6}$ $B\bar{B}$ pairs that were collected with the Belle detector at the KEKB asymmetric-energy $e^{+}e^{-}$ collider [7] operating at the $\Upsilon(4S)$ resonance (center-of-mass energy $\sqrt{s} = 10.58$ GeV). The Belle detector is a large-solid-angle magnetic spectrometer consisting of a silicon vertex detector, a 50-layer central drift chamber, a system of aerogel Cherenkov counters, time-of-flight scintillation counters and an electromagnetic calorimeter comprised of CsI(Tl) crystals located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux return located outside the coil is instrumented to identify $K_{S}^{0}$ mesons and muons. A detailed description of the detector can be found in Ref. [8]. We use Monte Carlo (MC) simulations to estimate signal efficiencies and background contributions. Large signal samples of $B^{-} \rightarrow D_{s}^{(*)+} K^{-}\ell^{-}\bar{\nu}_{\ell}$ decays are generated with the EvtGen package [9], using
a phase space model and the ISGW2 model \[10\] including the resonances that can decay to \(D_s^*(s)K\). Radiative effects are modeled by PHOTOS \[11\]. MC samples equivalent to about ten (six) times the accumulated data are used to evaluate the background from \(BB\) (continuum \(qq\) where \(q = u, d, s, c\) events).

In the analysis, we use charged tracks with impact parameters that are consistent with an origin at the beam spot and have transverse momenta above 50 MeV/c. Masses are assigned using information from particle identification subsystems. The efficiency for kaon (pion) identification ranges from 84% to 98% (92% to 94%) depending on the track momentum with a pion (kaon) misidentification probability of about 8% (16%). Electrons and muons are selected with an efficiency of about 90% and a misidentification rate below 0.2% (e) and 1.4% (\(\mu\)). The momenta of particles identified as electrons are corrected for bremsstrahlung by adding photons within a 50 mrad cone around the charged particle’s trajectory.

\(D_s^+\) candidates are reconstructed in the cleanest decay chain: \(D_s^+ \rightarrow \phi\pi^+, \phi \rightarrow K^+K^-\) \((2.32 \pm 0.14\%\) product branching fraction) and subjected to a vertex fit. We accept candidates in the invariant mass range of 1.934 GeV/c\(^2\) < \(M_{D_s}\) < 2.003 GeV/c\(^2\), and define the signal window within ±14 MeV/c\(^2\) around the world average \(D_s\) mass \(12\). The width of this window corresponds to \(4\sigma\) of the reconstructed \(D_s\) mass, using the resolution determined from control samples in data (mentioned later). The regions outside the signal window are considered as \(M_{D_s}\) sidebands. \(D_s^+\) candidates are combined with photons with an energy \(E_\gamma\) > 125 MeV to form \(D_{s}^{*+}\) candidates, subjected to a mass constrained vertex fit. Throughout this paper, all kinematic variables are defined in the \(\Upsilon(4S)\) rest frame, unless otherwise stated. \(D_s^+\) candidates with an invariant mass in the range of 2.079 GeV/c\(^2\) < \(M_{D_s}\) < 2.155 GeV/c\(^2\) are accepted for further analysis. The signal window is defined as 2.087 GeV/c\(^2\) < \(M_{D_s}\) < 2.137 GeV/c\(^2\) (3.7\sigma in \(M_{D_s}\)). Signal candidates for the decays considered here \(B_{\text{sig}}\) are formed by combining a negatively charged kaon and lepton (\(e\) or \(\mu\)) with a \(D_s^{*+}\) candidate. In the case of multiple \(B_{\text{sig}}\) candidates \(22\%\) of events after final selection requirements have multiple \(B_{\text{sig}}\) candidates), the one with the greatest confidence level of the vertex fit is chosen. Events with accepted \(D_s^{*+}K^-\bar{\ell}^-\) candidates \(D_s^+\) sample are removed from the set of \(D_s^{*+}K^-\bar{\ell}^-\) candidates \(D_s\) sample). Another charge configuration, \(D_s^{*+}K^+\bar{\ell}^-\), populated by decays of the type \(B \rightarrow D_s^{*+}\bar{D}^*(s), \bar{D} \rightarrow \ell^-\bar{\nu}_\ell K^+X\), is used as a control sample.

Signal events are identified using the variable \(X_{\text{mis}}\), introduced in Ref. \[13\] and defined as: \(X_{\text{mis}} \equiv (E_{\text{beam}} - E_{D_s\ell\bar{\nu}})/\sqrt{E_{\text{beam}}^2 - m_{B^+}^2}\), where \(E_{\text{beam}}\) is the beam energy, \(E_{D_s\ell\bar{\nu}}\) and \(p_{D_s\ell\bar{\nu}}\) denote the total energy and momentum of the \(D_s\ell\bar{\nu}\) system, respectively, and \(m_{B^+}\) is the nominal \(B^+\) mass. For decays with at most one massless invisible particle, as expected for the signal, \(X_{\text{mis}}\) takes values in the range of \([-1, 1]\), defined as the signal region, while the background has a much broader distribution. \(X_{\text{mis}}\) is calculated with the four-momentum of the \(D_s\) both in the \(D_s\) and \(D_s^*\) samples, causing a small shift of \(X_{\text{mis}}\) toward higher values for the \(D_s^*\) case due to the additional low-energy photon. With this definition, the \(X_{\text{mis}}\) distribution is more robust against imperfect modeling of photon spectra in MC and simplifies the signal extraction.

Particles not assigned to the \(B_{\text{sig}}\) are used to reconstruct the tagging side of the event \(B_{\text{tag}}\). Exploiting the information given by \(B_{\text{tag}}\) allows for background suppression without assumptions on the (unknown) signal dynamics. We require zero total event charge as well as a negatively charged lepton with a momentum above 0.5 GeV/c on the tagging side. This reduces the main background, where a \(D_s^*\) produced in a decay of the type \(B \rightarrow D_s^{*+}D^*(s)\) is combined with a lepton and a kaon from the subsequent \(D\) decay in a semileptonic decay \(B \rightarrow \ell^-\bar{\nu}_\ell D^{(*)}X\) of the accompanying \(B\) meson. Further improvement of the sensitivity is achieved with two tagging side variables \(M_{\text{tag}}^c \equiv (E_{\text{tag}} - E_{\text{tag}}^f)^2 - (\vec{p}_{\text{tag}} - \vec{p}_{\text{tag}}^f)^2\) and \(X_{\text{tag}} \equiv (E_{\text{beam}} - E_{\text{tag}} - |\vec{p}_{\text{tag}}|)/\sqrt{E_{\text{beam}}^2 - m_{B^+}^2}\), where \(E_{\text{tag}}\) and \(\vec{p}_{\text{tag}}\) denote the total energy and momentum of all reconstructed particles not assigned to \(B_{\text{sig}}\), and \(E_{\text{tag}}^f\) and \(\vec{p}_{\text{tag}}^f\) represent the energy and momentum of the prompt tagging lepton. Here \(M_{\text{tag}}^c\) represents the inclusively reconstructed mass of the hadronic system produced in the \(B_{\text{tag}}\) decay and \(X_{\text{tag}}\) is the tagging side equivalent of \(X_{\text{mis}}\). The \(M_{\text{tag}}^c\) and \(X_{\text{tag}}\) distributions for signal and background are shown in Fig. \[1\].

In this blind analysis, the selection criteria for \(X_{\text{tag}}\) and \(M_{\text{tag}}^c\) are optimized for the \(D_s\) mode by maximizing the expected statistical significance, \(N_S/\sqrt{N_S + N_B}\), where \(N_S\) (\(N_B\)) is the predicted number of signal (background) events in the \((X_{\text{mis}}, M_{D_s})\) signal window. This optimization is carried out for signal branching fractions \(B(B^- \rightarrow D_s^{*+}K^-\ell^-\bar{\nu})\) in the range of \((0.25 - 0.50) \times 10^{-3}\).
and yields similar optimal selection criteria for the whole range, namely $-2 < X_{tag} < 3$ and $M_{tag}^c < 2.4 \text{ GeV}/c^2$. $N_{f}$ is evaluated considering two background categories in the $D_s$ sample: “true $D_s$” background with correctly reconstructed $D_s^+$, described by the MC scaled to the integrated luminosity in data, and a “fake $D_s$” component, where random track combinations are misreconstructed as $D_s^+$, which is evaluated from the $M_{D_s}$ sidebands. In the $D_s^+$ sample, the background with true $D_s$ is split into two parts: “true $D_s^+$” with properly reconstructed $D_s^+$ and “fake $D_s^+$”, where a true $D_s^+$ is combined with a random photon candidate. The background model is tested using distributions in the sideband regions $X_{mis} < -1$ and $X_{mis} > 1$.

The $X_{mis}$ and $M_{D_s}$ distributions in data are shown in Fig. 2. Figure 3 shows the invariant mass distribution of the $D_s^+K^-\pi^-$ system, $M_{D_sK}$, for the combined $D_s$ and $D_s^*$ samples in the signal window and in the $X_{mis}$ sidebands. Superimposed histograms represent the expected backgrounds. While the background model describes the experimental $M_{D_sK}$ distribution well in the $X_{mis}$ sidebands, a clear excess over the expected background is seen in the signal region. The $M_{D_sK}$ distribution in the signal window is dominated by a prominent peak at $\approx 2.6$ GeV/$c^2$, similarly to that observed in $B^- \rightarrow D_s^+K^-\pi^-$ decays [13].

The signal yields are extracted from a simultaneous, extended unbinned maximum likelihood fit to the $D_s$ and $D_s^*$ samples, consisting of 2175 and 396 events, respectively. The $D_s$ and $D_s^*$ samples are fitted in two $(X_{mis},M_{D_s})$ and three $(X_{mis},M_{D_s},M_{D_s^*})$ dimensions, respectively. The likelihood function is constructed as follows:

$$
\mathcal{L} = \frac{e^{-(\sum_k N_k + \sum_{k'} N_{k'}^{\nu})} \prod_{i=1}^{N} \left( \sum_k N_k \mathbb{P}_k(x_i,y_i) \right)}{\prod_{i=1}^{N} \left( \sum_{k'} N_{k'}^{\nu} \mathbb{P}_{k'}^{\nu}(x_i,y_i,z_{i}) \right)},
$$

where $x_i,y_i,z_i$ denote $X_{mis},M_{D_s}$ and $M_{D_s^*}$ in the $i$th event, and $N_k$ denotes the total number of events in the $D_s^{(*)}$ data sample. The index $k$ ($k'$) runs over the signal and background components in the $D_s$ ($D_s^*$) sample; $N_k^{(*)}$ and $\mathbb{P}_k^{(*)}$ denote the number of events and the probability density functions (PDF) for each component, respectively. In the $D_s$ sample, we consider two signal components coming from the decay $B^- \rightarrow D_s^+K^-\ell^+\nu_\ell$ and from the decay $B^- \rightarrow D_s^{*+}K^-\ell^-\nu_\ell$ if a photon from the $D_s^{*+}$ has been missed. In the $D_s^*$ sample, we distinguish three signal components: one coming from the $B^- \rightarrow D_s^{*+}K^-\ell^-\nu_\ell$ mode, where the $D_s$ meson is associated with a random photon, and two from the $B^- \rightarrow D_s^{*+}K^-\ell^-\bar{\nu}_\ell$ mode, with true and fake $D_s^*$ defined similarly to the background case discussed above. The coefficients $N_k^{(*)}$ for the signal components are expressed as the products $N_k^{(*)} = N_{D_s^{(*)}} f_k^{(*)}$, where $N_{D_s^{(*)}}$ denotes the total number of signal events in the $B^- \rightarrow D_s^{(*)+}K^-\ell^-\nu_\ell$ modes. The coefficients $f_k^{(*)}$ (listed in Table I) represent the signal fraction reconstructed in each component and are evaluated from the signal MC.

| Signal component $k$ | Sample | $f_k^{(*)}$ |
|----------------------|--------|-------------|
| $B^- \rightarrow D_s^+K^-\ell^+\nu_\ell$ | $D_s^+$ | (84 ± 1)% |
| $B^- \rightarrow D_s^+K^-\ell^-\nu_\ell$ | $D_s^*$ | (16 ± 1)% |
| $B^- \rightarrow D_s^{*+}K^-\ell^-\nu_\ell$ | $D_s^*$ with true $D_s^*$ | (21 ± 1)% |
| $B^- \rightarrow D_s^{*+}K^-\ell^-\bar{\nu}_\ell$ | $D_s^*$ with fake $D_s^*$ | (13 ± 1)% |

TABLE I: The coefficients $f_k^{(*)}$, representing the signal fraction reconstructed in each component, evaluated from the signal MC.

The coefficients $N_k^{(*)}$ for background components with fake $D_s$ are evaluated from the $M_{D_s}$ sidebands in data and are fixed in the fit. The two- (three-) dimensional PDF is parameterized as the product of two (three) one-dimensional PDFs for each variable. The validity of this parameterization has been checked with MC by examining the correlation between $X_{mis}$ and $M_{D_s}$, which has been found negligible. The components with true $D_s^*$ are parameterized as a sum of two Gaussian functions in $M_{D_s}$ or as a single Gaussian function in $M_{D_s^*}$, with means set to the world average $D_s^{(*)}$ mass values [12] and with the remaining parameters fixed from fits to control samples in data. The components with fake $D_s^{(*)}$ are parameterized as linear functions in $M_{D_s^{(*)}}$. The $X_{mis}$ distribution of the signal components is modeled with two line shapes, one describing the two components of the $B^- \rightarrow D_s^{*+}K^-\ell^-\nu_\ell$ mode and the other one describing the three components of the $B^- \rightarrow D_s^{*+}K^-\ell^-\bar{\nu}_\ell$ decay. They are parameterized using the function $C e^{-((X_{mis}-\mu)/\sigma)^n} e^{-\alpha(X_{mis}-\mu)}$, where $C$ is a normalization coefficient and the parameters $\mu, \sigma, \alpha$ and the integer parameter $n$ are fixed from fits to the signal MC samples. The $X_{mis}$ distributions of the background components are parameterized as bifurcated Gaussian functions with parameters fixed from the simulated $BB$ events with generic $B$ decays (true $D_s$) or from the $M_{D_s}$ sidebands in data (fake $D_s$). The free parameters in the fit are the two signal yields $N_{D_s^{(*)}}$, the three background yields $N_{k}^{(*)}$ of the components with true $D_s$, and the coefficients of polynomials that describe the distributions in $M_{D_s^{(*)}}$ for the fake $D_s$ components. The range of the fit is as shown in Fig. 3. The signal yields extracted from the fit are $84 \pm 24$ events for the decay $B^- \rightarrow D_s^+K^-\ell^+\nu_\ell$ and $41 \pm 22$ events for the decay $B^- \rightarrow D_s^{*+}K^-\ell^-\nu_\ell$ with statistical significances of 3.9$\sigma$ and 1.9$\sigma$, respectively. The significance is defined as $\Sigma = \sqrt{-2 \ln(\mathcal{L}_0/\mathcal{L}_{\text{max}})}$, where $\mathcal{L}_{\text{max}}$ and $\mathcal{L}_0$ denote the maximum likelihood value and the likelihood value for the zero signal hypothesis, respectively. The fit results are summarized in Table II and the fit projections in $X_{mis}$ and $M_{D_s}$, are shown in Fig. 3. The fitted signal yields are used to compute the branching
FIG. 2: Distributions (from left to right) in $X_{\text{mis}}$ and $M_{D_{s}}$ in the $D_{s}$ sample (top), and $X_{\text{mis}}, M_{D_{s}}$, and $M_{D_{s}^{*}}$ in the $D_{s}^{*}$ sample (bottom). Points with error bars are the data, and lines show the fit projections. Each variable is shown in the signal region of the other variable(s). For the $D_{s}$ sample the lines represent (from bottom to top) the fitted background components with fake (red dashed) and true $D_{s}$ (red solid), and the signal contributions from the $D_{s}^{*}$ (blue dashed) and $D_{s}$ (blue solid) modes. For the $D_{s}^{*}$ sample the lines (from bottom to top) represent the fitted background components with fake $D_{s}$ (red dashed), fake $D_{s}^{*}$ (red dotted), true $D_{s}^{*}$ (red solid), and the signal contributions from the $D_{s}$ mode (blue dashed), the $D_{s}^{*}$ mode with fake $D_{s}^{*}$ (blue dotted), and with true $D_{s}^{*}$ (blue solid). The fitted contributions are superimposed additively.

FIG. 3: The invariant mass distribution of $D_{s}^{+}K^{-}$ for the combined $D_{s}$ and $D_{s}^{*}$ samples in the signal window (left) and in the $X_{\text{mis}}$ sidebands (right). The full (blank) histograms show the expected background contribution from fake (true) $D_{s}$. The histograms are superimposed additively.

TABLE II: Signal yields ($N_{D_{s}^{(*)}}$), reconstruction efficiencies ($\epsilon^{(*)}$), statistical significances ($\Sigma$) and branching fractions ($B$). The errors on the signal yields are statistical, while for the branching fractions both statistical (first) and systematic (second) errors are provided. The correlation coefficient between $N_{D_{s}}$ and $N_{D_{s}^{*}}$ equals $-66\%$.

| Mode               | $N_{D_{s}^{(*)}}$ | $\epsilon^{(*)}$ [%] | $\Sigma$ | $B \times 10^{-4}$ |
|--------------------|-------------------|-----------------------|----------|-------------------|
| $B \to D_{s}^{+}K^{-}\ell^{-}\bar{\nu}_{\ell}$ | 84 ± 24 | 1.78 | 3.9 | 0.30 ± 0.09$^{+0.17}_{-0.08}$ |
| $B \to D_{s}^{*+}K^{-}\ell^{-}\bar{\nu}_{\ell}$ | 41 ± 22 | 0.85 | 1.9 | 0.29 ± 0.16$^{+0.10}_{-0.09}$ |

fractions with the formula: $B(B^{-} \to D_{s}^{(*)+}K^{-}\ell^{-}\bar{\nu}_{\ell}) = N_{D_{s}^{(*)}}/(2N_{B^{+}B^{-}}\epsilon^{(*)}B_{\text{int}})$, where $N_{B^{+}B^{-}}$ is the number of $B^{+}B^{-}$ pairs in data, $\epsilon^{(*)}$ denotes the reconstruction efficiency of the signal decay chain and $B_{\text{int}}$ is the product of intermediate branching fractions set to their world average values [12]. The reconstruction efficiency is expressed as $\epsilon^{(*)} = \epsilon_{PS}^{(*)}\Delta\epsilon_{\text{cor}}^{(*)}$, where $\epsilon_{PS}^{(*)}$ is the efficiency calculated from the signal MC with the phase space model and $\Delta\epsilon_{\text{cor}}^{(*)} = 1.20$ (0.57) corrects for the difference between the data and the phase space distribution. It is calculated as a function of the effective masses of the two-body subsystems $D_{s}^{+}K^{-}$, $D_{s}^{*+}\ell^{-}$, and $K^{-}\ell^{-}$ and averaged using the experimentally obtained distributions. We obtain $B(B^{-} \to D_{s}^{*+}K^{-}\ell^{-}\bar{\nu}_{\ell}) = (0.30 \pm 0.09) \times 10^{-3}$ and $B(B^{-} \to D_{s}^{+}K^{-}\ell^{-}\bar{\nu}_{\ell}) = (0.29 \pm 0.16) \times 10^{-3}$.

The dominant systematic uncertainty on the signal yield is due to the parameterization of the $X_{\text{mis}}$ dependence of the signal and found to be $\pm 2\% (\pm 6\%)$ events for the $D_{s}(D_{s}^{*})$ mode. It is evaluated by refitting the data with the parameters $\mu$, $\sigma$, and $\alpha$ allowed to float, and by changing the integer parameter $n$ by $\pm 1$. Uncertainties in modeling the $X_{\text{mis}}$ distributions of the background components containing true $D_{s}$ are evaluated to be $\pm 1\% (\pm 5\%)$ events from fits with the background shape parameters varied by $\pm 1\sigma$, taking into account correlations between the parameters. We also repeat the fits with the param-
meters, whose values are determined from data (and which are fixed in the nominal fit), floating. The resulting uncertainty is \( \pm 2\% \) (\( \pm 1\% \)) events. The effect of an imperfect estimation of the relative contributions of the signal components is determined to be \( \pm 1 \) (\( \pm 1\% \)) from fits with the parameters \( f_k^{(*)} \) varied by \( \pm 1\sigma \) and taking into account a \( \pm 3\% \) uncertainty on the photon reconstruction efficiency. The above uncertainties are summed in quadrature to obtain the total systematic uncertainty of the signal yield of \( \pm 2\% \) (\( \pm 1\% \)) events for the \( D_s \) (\( D_s^* \)) modes. We include the effect of these uncertainties on the significance of the observed signals by convolving the likelihood function obtained in the fit with a Gaussian systematic error distribution. The significance of the signal in the \( B^+ \to D_s^0 K^- \ell^- \bar{\nu}_\ell \) (\( B^- \to D_s^{*+} K^- \ell^- \bar{\nu}_\ell \)) mode, after including systematic uncertainties, is \( 3.4\sigma \) (1.8\( \sigma \)).

In a similar way, we obtain a significance of \( \sigma \) for the combined \( B^+ \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell \) modes from the 2-dimensional \( (X_{\text{min}}, M_{D_s}) \) fit for the combined \( D_s \) and \( D_s^* \) samples. The much higher significance for the combined modes compared to the individual modes is due to the large cross-feed between the \( D_s \) and the \( D_s^* \) modes.

The uncertainty on the branching fractions, except for the systematic uncertainty of the signal yield, is evaluated to be 23.2\% for each signal mode. It includes uncertainties in charged track reconstruction efficiency (6.6\%), particle identification efficiency (3.9\%), intermediate branching fractions (6.1\%), number of \( B^+B^- \) pairs (1.5\%) and the reconstruction efficiency correction \( \Delta \epsilon_{\text{cor}} \) (21\%).

The largest uncertainty, due to \( \Delta \epsilon_{\text{cor}} \), is determined by calculating \( \Delta \epsilon_{\text{cor}} \) in 10000 toy MC experiments. The width of a Gaussian function fitted to the obtained efficiencies is taken as systematic uncertainty. The uncertainties due to the intermediate branching fractions are taken from the errors quoted in [12]. Combining all uncertainties, we obtain \( B(B^+ \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell) = (0.30 \pm 0.09 \text{(stat)} \pm 0.11 \text{(syst)}) \times 10^{-3} \), \( B(B^- \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell) = (0.29 \pm 0.16 \text{(stat)} \pm 0.15 \text{(syst)}) \times 10^{-3} \) and \( B(B^- \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell) = (0.59 \pm 0.12 \text{(stat)} \pm 0.15 \text{(syst)}) \times 10^{-3} \) for the combined modes obtained in a similar way, taking correlations into account. Since the significance in the \( D_s^* \) mode does not exceed 3\( \sigma \), we set an upper limit of \( B(B^- \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell) < 0.56 \times 10^{-3} \) at the 90\% confidence level, using the likelihood integration method.

In conclusion, we find evidence for the decay \( B^- \to D_s^0 K^- \ell^- \bar{\nu}_\ell \) with a significance of 3.4\( \sigma \) and measure \( B(B^- \to D_s^0 K^- \ell^- \bar{\nu}_\ell) = (0.30 \pm 0.09 \text{(stat)} \pm 0.11 \text{(syst)}) \times 10^{-3} \). The combined \( B^- \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell \) decay modes are observed with a significance of 6\( \sigma \) to be \( B(B^- \to D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell) = (0.59 \pm 0.12 \text{(stat)} \pm 0.15 \text{(syst)}) \times 10^{-3} \). The branching fraction results are consistent with the measurement of BaBar [6]. We also present the first measurement of the \( D_s^0 K^- \) invariant mass distribution, which is dominated by a prominent peak around 2.6 GeV/c\(^2\), possibly from excited \( D \) mesons decays.

We thank the KEKB group for excellent operation of the accelerator; the KEK cryogenics group for efficient solenoid operations; and the KEK computer group, the NII, and PNNL/EMSL for valuable computing and SINET4 network support. We acknowledge support from MEXT, JSPS and Nagoya’s TLPRC (Japan); ARC and DIISR (Australia); NSFC (China); MSMT (Czechia); DST (India); INFN (Italy); MEST, NRF, GSDC of KISTI, and WCU (Korea); MNiSW and NCN (Poland); MES and RFAEE (Russia); ARRS (Slovenia); SNSF (Switzerland); NSC and MOE (Taiwan); and DOE and NSF (USA).

[1] N. Cabibbo, Phys. Rev. Lett. 10, 531 (1963); M. Kobayashi and T. Maskawa, Prog. Theor. Phys. 49, 652 (1973).
[2] D. Asner et al. (Heavy Flavor Averaging Group), arXiv:1010.1589 [hep-ex].
[3] Throughout this paper, the inclusion of the charge-conjugate decay mode is implied.
[4] P. del Amo Sanchez et al. (BaBar Collaboration), Phys. Rev. D 82, 111101(R) (2010).
[5] F. U. Bernlochner, Z. Ligeti, S. Turczyk, arXiv:1202.1834 [hep-ph].
[6] P. Amo del Sanchez et al. (BaBar Collaboration), Phys. Rev. Lett. 107, 041804 (2011).
[7] S. Kurokawa and E. Kikutani, Nucl. Instr. and Meth. A 499, 1 (2003), and other papers included in this volume.
[8] A. Abashian et al. (Belle Collaboration), Nucl. Instr. and Meth. A 479, 117 (2002).
[9] D. J. Lange, Nucl. Instr. and Meth. A 462, 152 (2001).
[10] D. Scora and N. Isgur, Phys. Rev. D 52, 2783 (1995).
[11] E. Barberio and Z. Wys, Comput. Phys. Commun. 79, 291 (1994).
[12] K. Nakamura et al. (Particle Data Group), J. Phys. G 37, 075021 (2010) and 2011 partial update for the 2012 edition.
[13] A. Matyja et al. (Belle Collaboration), Phys. Rev. Lett. 99, 191807 (2007).
[14] B. Aubert et al. (BaBar Collaboration), Phys. Rev. Lett. 100, 171803 (2008); J. Wiechczynski et al. (Belle Collaboration), Phys. Rev. D 80, 052005 (2009).