This is the accepted manuscript made available via CHORUS. The article has been published as:

First observation of the decay $B^0 \rightarrow \psi(2S)\pi^0$
V. Chobanova et al. (Belle Collaboration)
Phys. Rev. D 93, 031101 — Published 23 February 2016
DOI: 10.1103/PhysRevD.93.031101
First Observation of the Decay $B^0 \rightarrow \psi(2S)\pi^0$


(The Belle Collaboration)

1Aigirashi Muslim University, Aigir 200202
2University of the Basque Country UPV/EHU, 48080 Bilbao
3Budker Institute of Nuclear Physics SB RAS, Novosibirsk 630090
4Faculty of Mathematics and Physics, Charles University, 121 16 Prague
5Chern National Normal University, Kwejngu 660-701
6University of Cincinnati, Cincinnati, Ohio 45221
7Deutsches Elektronen-Synchrotron, 22607 Hamburg
8University of Florida, Gainesville, Florida 32611
9Justus-Liebig-Universität Gießen, 35392 Gießen
10Gifu University, Gifu 501-1193
11SOKENDAI (The Graduate University for Advanced Studies), Hayama 240-0193
12Hanyang University, Seoul 133-791
13University of Hawaii, Honolulu, Hawaii 96822
14High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801
15IKERBASQUE, Basque Foundation for Science, 48013 Bilbao
16Indian Institute of Technology Bhuvaneshwar, Satya Nagar 751007
17Indian Institute of Technology Guwahati, Assam 781039
18Indian Institute of Technology Madras, Chennai 600036
19Indiana University, Bloomington, Indiana 47408
20Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049
21Institute of High Energy Physics, Vienna 1050
22Institute for High Energy Physics, Provința 142281
23INFN - Sezione di Torino, 10125 Turin
24J. Stefan Institute, 1000 Ljubljana
25Kanagawa University, Yokohama 221-8686
26Institut für Experimentelle Kernphysik, Karlsruher Institut für Technologie, 76131 Karlsruhe
27Kennesaw State University, Kennesaw GA 30144
We report a measurement of the $B^0 \rightarrow \psi(2S)\pi^0$ branching fraction based on the full $\Upsilon(4S)$ data set of $772 \times 10^6 \bar{B}B$ pairs collected by the Belle detector at the KEKB asymmetric-energy $e^+e^-$ collider. We obtain $\mathcal{B}(B^0 \rightarrow \psi(2S)\pi^0) = (1.17 \pm 0.17(\text{stat}) \pm 0.08(\text{syst})) \times 10^{-5}$. The result has a significance of 7.2 standard deviations and is the first observation of the decay $B^0 \rightarrow \psi(2S)\pi^0$.

PACS numbers: 12.15.Hh, 13.25.Hw

Violation of the combined charge–parity symmetry (CP violation) in the Standard Model (SM) arises from a single irreducible phase in the Cabibbo–Kobayashi–Maskawa (CKM) quark-mixing matrix [1, 2]. A primary objective of the Belle experiment is to overconstrain the unitarity triangle of the CKM matrix related to $B_{u,d}$ decays. This permits a precision test of the CKM mechanism for CP violation as well as the search for effects beyond the SM. Mixing-induced CP violation in the $B$ sector has been clearly established by the Belle [3] and BaBar [4] collaborations in the $b \rightarrow c\bar{c}s$-induced decays $B^0 \rightarrow (c\bar{c})^0K^0$. 
While these decays allow access to the $CP$ violating angle $\phi_1 \equiv \arg(\langle V_{cd}V_{cb}^\ast\rangle/(\langle V_{ub}V_{ub}^\ast\rangle)$ at first order (tree), its value is prone to distortion from suppressed higher-order loop-induced (penguin) amplitudes containing different weak phases. Applying SU(3) symmetry arguments, the related $b \to c\bar{c}d$-induced channels $B^0 \to (c\bar{c})\pi^0$ can be used to quantify the shift in $\phi_1$ caused by these loop contributions [5]. Thus, this $b \to c\bar{c}d$ decay is a promising place to search for new physics effects [6]. This paper establishes the $B^0 \to \psi(2S)\pi^0$ channel, which may be used to constrain the penguin contamination in $B^0 \to \psi(2S)K^0$ in a future measurement of its time-dependent $CP$ asymmetry.

The result presented in this paper is based on the final $\Upsilon(4S)$ data sample, containing $772 \times 10^6 BB$ pairs collected with the Belle detector at the KEKB asymmetric-energy $e^+e^-$ (3.5 on 8 GeV) collider [7]. At the $\Upsilon(4S)$ resonance, corresponding to a center-of-mass energy $\sqrt{s} = 10.58$ GeV, the $BB$ pairs are produced with a Lorentz boost $\beta\gamma = 0.425$ nearly along the $+z$ direction, which is opposite the positron beam direction.

The Belle detector is a large-solid-angle magnetic spectrometer that consists of a silicon vertex detector (SVD), a 50-layer central drift chamber (CDC), an array of aerogel threshold Cherenkov counters (ACC), a barrel-like arrangement of time-of-flight scintillation counters (TOF), and an electromagnetic calorimeter (ECL) comprising CsI(Tl) crystals located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux-return yoke located outside of the coil is instrumented to detect $K^0_L$ mesons and to identify muons (KLM). The detector is described in detail elsewhere [8].

Two inner detector configurations were used: A 2.0 cm radius beampipe and a three-layer silicon vertex detector (SVD1) were used for the first sample of $152 \times 10^6 BB$ pairs, while a 1.5 cm radius beampipe, a four-layer silicon vertex detector (SVD2), and a small-cell inner drift chamber were used to record the remaining $620 \times 10^6 BB$ pairs [9]. Simulated $B$ decay Monte Carlo (MC) events are generated by EvtGen [10], in which final-state radiation is described with PHOTOS [11]. We use the GEANT3 [12] toolkit to model the interaction of the generated particles with the detector and its response in order to determine the detector acceptance.

We reconstruct $\psi(2S)$ candidates in the $t^+t^-$ decay channels ($\ell = e, \mu$, referred to as leptonic hereinafter, and the $J/\psi\pi^+\pi^-$ decay channel, referred to as hadronic. All charged tracks are identified using a loose requirement on the distance of closest approach with respect to the interaction point along the beam direction of under 5.0 cm and in the transverse plane of under 1.5 cm. The $J/\psi$ candidates are reconstructed from $t^+t^-$ pairs. Electron tracks are identified by a combination of $dE/dx$ in the CDC, shower shape and position in the ECL, light yield in the ACC, and $E/p$, where $E$ is the energy deposited in the ECL and $p$ is the momentum measured by the SVD and the CDC. To account for radiative energy losses in the $e^+e^-$ decays, we include the bremsstrahlung photons ($\gamma$) that are in a cone with an opening angle of 50 mrad around the $e^+$ ($e^-$) tracks so that the reconstructed $J/\psi$ or $\psi(2S)$ candidate is denoted as $e^+e^-(\gamma)$. For muon tracks, the identification is based on track penetration depth and hit scatter in the KLM.

We impose asymmetric requirements on the $J/\psi$ and $\psi(2S)$ masses due to energy leakage in the ECL and bremsstrahlung. The invariant masses of the $J/\psi$ candidates must fulfill $M_{e^+e^-}(\gamma) - m_{J/\psi} \in (-0.150, +0.036)$ GeV/$c^2$ or $M_{\mu^+\mu^-} - m_{J/\psi} \in (-0.060, +0.036)$ GeV/$c^2$, where $m_{J/\psi}$ denotes the world-average $J/\psi$ mass [13], and $M_{e^+e^-}(\gamma)$ and $M_{\mu^+\mu^-}$ are the reconstructed invariant masses of the $e^+e^-(\gamma)$ and $\mu^+\mu^-$ candidates, respectively. For the $\psi(2S)$, the invariant masses must fulfill $M_{e^+e^-}(\gamma) - m_{\psi(2S)} \in (-0.150, +0.036)$ GeV/$c^2$ or $M_{\mu^+\mu^-} - m_{\psi(2S)} \in (-0.060, +0.036)$ GeV/$c^2$, where $m_{\psi(2S)}$ denotes the world-average $\psi(2S)$ mass [13]. For the $\psi(2S) \to J/\psi\pi^+\pi^-$ candidates, $\Delta M \equiv M_{e^+e^-}(\gamma) - M_{\mu^+\mu^-}(\gamma)$ must fulfill $\Delta M \in (0.550, 0.600)$ GeV/$c^2$. To reduce background particle combinations in this channel, we select $\pi^+\pi^-$ pairs with an invariant mass above a loose threshold of 400 MeV/$c^2$. Using information obtained from the CDC, ACC, and TOF, these pion candidates are also required to be inconsistent with the kaon mass hypothesis. This requirement retains 99.8% of the pion candidates, while 5% of kaons are falsely identified as pions. To improve the $B$ meson mass resolution, we apply a vertex- and mass-constrained kinematic fit to the $J/\psi$ and $\psi(2S)$ candidates. We assign each candidate its nominal mass and require that its charged daughters originate from the same vertex.

Photons are identified as isolated ECL clusters that are not matched to any charged particle track. To suppress combinatorial background, the photons are required to have energies above 50 MeV if in the ECL barrel or above 100 MeV if in the ECL endcaps, where the barrel region covers the polar angle range $32^\circ < \theta < 130^\circ$ and the endcap regions cover the polar angle ranges $12^\circ < \theta < 32^\circ$ and $130^\circ < \theta < 157^\circ$. Two $\gamma$ candidates are combined to form a $\pi^0$ candidate that must satisfy $M_{\gamma\gamma} - m_{\pi^0} \in (-17, 15)$ MeV/$c^2$, where $m_{\pi^0}$ is the world-average mass of the $\pi^0$ [13]. This corresponds to about three times the experimental resolution. The four-momenta of retained candidates are then adjusted in a mass-constrained fit wherein the parent mass is constrained to $m_{\pi^0}$.

We combine the $\psi(2S)$ and $\pi^0$ to form a neutral $B$ meson. The $B$ candidates are identified using two kinematic variables: a modified beam-energy-constrained mass,
and the energy difference \( \Delta E \equiv E_B - E_{\text{beam}} \), where \( \vec{p} \) denotes 3-momentum and \( E_{\text{beam}} \) the beam energy, all evaluated in the \( Y(4S) \) center-of-mass system. This definition of \( M'_{bc} \) is preferred over the standard form used at the B factories as it exhibits a lower correlation with \( \Delta E \) when \( \pi^0 \) is present in the final state.

A significant background arises from \( e^+e^- \rightarrow q\bar{q} \) (where \( q = u, d, s, c \)) continuum events. To suppress it, we construct the ratio of second- to zeroth-order Fox–Wolfram moments [14], \( R_2 = H_2/H_0 \), which ranges between zero (spherical) and one (jet-like). A loose requirement of less than 0.5 is applied. This removes around 50% of all continuum background with a negligible loss of signal efficiency.

On average, 1.13 \( B^0 \) candidates are reconstructed per event and 11.6% of all events have more than one candidate. In a multi-candidate event, we choose the \( B^0 \) with the lowest \( \chi_{\text{mass}}^2 \equiv (M_{\text{Rec}} - m)^2/\sigma_{\text{Rec}}^2 \) per daughter particle with a reconstructed mass \( M_{\text{Rec}} \), a nominal mass \( m \) and a mass resolution \( \sigma_{\text{Rec}} \). For the leptonic channels, \( \chi_{\text{mass}}^2 \equiv (\chi^2_{(2S)} + \chi^2_{\psi})/2 \). For the hadronic channels, \( \chi_{\text{mass}}^2 \equiv (\chi^2_{\psi} + \chi^2_{3m} + \chi^2_{4m})/3 \), where \( \chi_{3m}^2 \) is defined similarly except that the reconstructed and nominal mass differences between \( \psi(2S) \) and \( J/\psi \) are used in place of \( M_{\text{Rec}} \) and \( m \), respectively. According to MC simulation, this procedure has a 75% success rate when more than one \( B \) candidate is reconstructed and the correct \( B \) is in the list. After this best-candidate selection, the detection efficiency, including a correction for the difference between data and MC in the particle identification and including the daughter branching fraction uncertainties and the selection criteria uncertainties, is \( (0.43 \pm 0.02)\% \) for the leptonic channels and \( (0.52 \pm 0.02)\% \) for the hadronic. Approximately 0.5% (10%) of the signal candidates are misreconstructed in the leptonic (hadronic) channels.

The \( B^0 \rightarrow \psi(2S)\pi^0 \) branching fraction, \( B(B^0 \rightarrow \psi(2S)\pi^0) \), is extracted from an unbinned extended maximum likelihood fit to \( M'_{bc} \) and \( \Delta E \). The following categories are considered in the event model: correctly-reconstructed signal, misreconstructed signal, other \( b \rightarrow (c\bar{c})q \) transitions, and combinatorial background. Unless otherwise stated, the probability density function (PDF) is the product of PDFs for each observable, \( P_{\psi}(M'_{bc}, \Delta E) \equiv P_{\psi}(M'_{bc}) \times P_{\psi}(\Delta E) \), in each \( \psi(2S) \) decay mode, \( m \), and in each category, \( c \).

We study the distributions of both signal components – correctly reconstructed and misreconstructed – using an MC sample that contains only \( B^0 \rightarrow \psi(2S)\pi^0 \) events. We define a correctly-reconstructed event as one in which all charged tracks are correctly associated with the signal \( B \) meson. For such events, we fit the distributions of the fit observables in the \( \psi(2S) \rightarrow e^+e^- \) and \( \psi(2S) \rightarrow J/\psi(e^+e^-)\pi^+\pi^- \) to be similar. The distributions in the \( \psi(2S) \rightarrow \mu^+\mu^- \) and \( \psi(2S) \rightarrow J/\psi(\mu^+\mu^-)\pi^+\pi^- \) modes are also alike. Thus, we divide the signal MC into an electron and a muon component and model these separately. The \( M'_{bc} \) PDF for both modes consists of a Crystal Ball (CB) function [15], \( C \), combined with an ARGUS distribution [16], \( A \), which additionally accounts for the tail towards lower \( M'_{bc} \) values due to the photon and electron energy leakage in the ECL. Due to a correlation between \( M'_{bc} \) and \( \Delta E \), we parametrize the \( M'_{bc} \) PDF in terms of \( \Delta E \),

\[
P_{\psi}(M'_{bc}|\Delta E) \equiv (f^m + \rho_1^m \Delta E^2)C(M'_{bc}; \alpha, \mu_{M'_{bc}}^m, \sigma_{M'_{bc}}^m) + \rho_2^m g^m(\Delta E),
\]

where \( \sigma_{M'_{bc}}^m \) and \( \alpha \) are correlation factors obtained from \( B^0 \rightarrow J/\psi K^+K^- \) control sample; \( \rho_1^m \) and \( \rho_2^m \) are correlation factors and \( g^m(\Delta E) \) are functions in \( \Delta E \) determined from MC: \( g^{e^+e^-} = \Delta E^2 \) for the electron component and \( g^{\mu^+\mu^-} = |\Delta E| \) for the muon component. For both types of correctly reconstructed signal events, the \( \Delta E \) PDF is the combination of a CB distribution and a sum of Chebyshev polynomials up to the first order,

\[
P_{\psi}(\Delta E) \equiv f^m C(\Delta E; \alpha_{\Delta E}, \mu_{\Delta E}^m, \sigma_{\Delta E}^m, \rho_{\Delta E}^m) + (1 - f^m)(1 + c^m |\Delta E|),
\]

where \( \sigma_{\Delta E}^m, \rho_{\Delta E}^m, \mu_{\Delta E}^m, f^m, c^m \) are obtained from MC, while \( \rho_{\Delta E} \) and \( \sigma_{\Delta E} \) are correction factors obtained from the control sample.

We omit the misreconstructed signal component in the leptonic decay modes due to its insignificant contribution. Each of the two hadronic modes is modeled with a separate two-dimensional histogram in \( M'_{bc} - \Delta E \). The major background contribution originates from \( b \rightarrow (c\bar{c})q \) decays other than the signal. We study this component from an MC sample containing all known \( b \rightarrow (c\bar{c})q \) decays. Since the two leptonic channels have similar distributions, as do the two hadronic channels, we divide the \( b \rightarrow (c\bar{c})q \) background events into a leptonic and a hadronic subsample. We model each of these with a two-dimensional \( M'_{bc} - \Delta E \) histogram.
refer to these as combinatorial background. We study their distributions from $\Upsilon(4S)$ data in the dilepton and $\Delta M$ sidebands. The $J/\psi$ sideband is defined as $M_{J/\psi} \in (2.60, 2.80)$ $\cup$ $(3.20, 3.40)$ GeV/$c^2$, the $\psi(2S)$ sideband as $M_{\psi(2S)} \in (3.45, 3.53)$ $\cup$ $(3.80, 3.90)$ GeV/$c^2$, and the $\Delta M$ sideband as $\Delta M \in (0.49, 0.53)$ $\cup$ $(0.64, 0.68)$ GeV/$c^2$.

In all sidebands, the $M_{bc}'$ PDF is an ARGUS distribution. In the leptonic sidebands, we model the $\Delta E$ combinatorial background distribution with a sum of Chebyshev polynomials up to the first order. The combinatorial $\Delta E$ PDF in the $\Delta M$ sideband is a sum of Chebyshev polynomials up to the second order. We verify that the models in the lower and upper sidebands are in agreement and thus the combined model provides a reliable description of the events in the signal region.

The total extended likelihood is given by

$$\mathcal{L} = \prod_{m} \frac{e^{-\sum_{i=1}^{c} N_{i}^{c} N_{i}^{m}}} {N_{m}!} \prod_{i=1}^{c} N_{i}^{m} \mathcal{P}_{c}^{m}(M_{bc}', \Delta E^{i}),$$

where $i$ indexes the events, $c$ the categories and $m$ the decay modes.

The $B^0 \rightarrow \psi(2S)\pi^0$ branching fraction is a free parameter in the fit to the data and is obtained by transforming the signal yields according to

$$N_{\text{Sig}}^{m} = B(B^0 \rightarrow \psi(2S)\pi^0) N_{B\bar{B}} \epsilon_{\text{Sig}}^{m},$$

where $N_{B\bar{B}}$ is the number of $B\bar{B}$ pairs collected by the Belle detector and $\epsilon_{\text{Sig}}^{m}$ is the detection efficiency, including daughter branching fractions for each subcategory. The misreconstructed-signal yields are fixed from MC relative to the two hadronic-mode signal yields. Only the muonic hadronic mode's yield is free in the $c\bar{c}$ background category, while the yields of the three remaining decay modes are fixed from MC relative to it. The four combinatorial-background yields are free.

We study the fit performance using pseudo-experiments in a linearity test covering the region of the expected branching fraction. There is no bias in experiments where the events are generated according to the total PDF. However, a bias at the level of 10% of the statistical error tending towards higher values is observed in experiments generated by selecting random events from the MC samples that have passed the full selection. This indicates that the bias is not due to a low signal yield but rather to imperfections in the modeling of correlations. We apply a fit correction of the full bias and consider half the correction as a systematic uncertainty.

The contribution of peaking background that originates from decays to the same final state as the signal is studied in the $J/\psi$, $\psi(2S)$ and $\Delta m$ sidebands. We define the combinatorial background as non-peaking in $M_{bc}'$ and $\Delta E$, while we assume that a potential peaking background has the same shape as the correctly reconstructed signal. Using the combinatorial background and the signal PDFs in a common fit to the sidebands, we extract two yields: one for the combinatorial background and the other for the peaking background. The peaking-background yield is consistent with zero for all modes except for the muonic signal mode in the $\Delta M$ sideband, where it has a statistical significance of 3.7$\sigma$.

We extrapolate the expected peaking background yield into the signal region and subtract the obtained value from the signal yield obtained from the data.

We determine the $M_{bc}'$ and $\Delta E$ signal model correction factors from a control sample with a similar decay topology, $B^+ \rightarrow J/\psi K^{*+}$, where the $K^{*+}$ candidates are reconstructed from a $K^+$ and a $\pi^0$ candidate. To ensure a high momentum of the $\pi^0$, replicating the kinematic conditions of $B^0 \rightarrow \psi(2S)\pi^0$, we require the angle between the $\pi^0$ momentum vector and the vector opposite the $B$ flight direction in the $K^{*+}$ rest frame to be smaller than 1.5 rad. For the $J/\psi$ and $\pi^0$ candidates, we use the same selection criteria as for the $B^0 \rightarrow \psi(2S)\pi^0$ mode. Only $K^{*+}$ candidates fulfilling $M_{K^{*+}\pi^0} \in (0.793, 0.990)$ GeV/$c^2$ are retained. Using a model similar to $B^0 \rightarrow \psi(2S)\pi^0$ for the control sample, we obtain a $B^+ \rightarrow J/\psi K^{*+}$ signal yield of 3081 $\pm$ 71 events and the signal correction factors from the fit to the data.

We determine the $M_{bc}'$ and $\Delta E$ signal model correction factors from a control sample with a similar decay topology, $B^+ \rightarrow J/\psi K^{*+}$, where the $K^{*+}$ candidates are reconstructed from a $K^+$ and a $\pi^0$ candidate. To ensure a high momentum of the $\pi^0$, replicating the kinematic conditions of $B^0 \rightarrow \psi(2S)\pi^0$, we require the angle between the $\pi^0$ momentum vector and the vector opposite the $B$ flight direction in the $K^{*+}$ rest frame to be smaller than 1.5 rad. For the $J/\psi$ and $\pi^0$ candidates, we use the same selection criteria as for the $B^0 \rightarrow \psi(2S)\pi^0$ mode. Only $K^{*+}$ candidates fulfilling $M_{K^{*+}\pi^0} \in (0.793, 0.990)$ GeV/$c^2$ are retained. Using a model similar to $B^0 \rightarrow \psi(2S)\pi^0$ for the control sample, we obtain a $B^+ \rightarrow J/\psi K^{*+}$ signal yield of 3081 $\pm$ 71 events and the signal correction factors from the fit to the data.

From the fit to the data containing 1090 $B^0 \rightarrow \psi(2S)\pi^0$ candidates, we obtain the bias-corrected branching fraction

$$B(B^0 \rightarrow \psi(2S)\pi^0) = (1.17 \pm 0.17) \times 10^{-5}.$$ (6)

The branching fraction corresponds to 85 $\pm$ 12 signal events, of which $38 \pm 8$ are leptonic and $47 \pm 9$ are hadronic, 628 $\pm$ 65 events originate from other $b \rightarrow (c\bar{c})q$ decays and 377 $\pm$ 103 events belong to the combinatorial background. All uncertainties here are statistical. Fit projections to the data are shown in Fig. 1.
The dominant systematic uncertainty arises from the π0 reconstruction and is evaluated by comparing data-MC differences in the yield ratios between $\eta' \rightarrow \pi^0\pi^0\pi^0$ and $\eta' \rightarrow \pi^+\pi^-\pi^0$. We also consider the systematic uncertainties originating from the knowledge of the $\psi(2S)$ and $J/\psi$ decay branching fractions used to calculate the efficiency. We apply the percentage error on their world averages [13] as a systematic uncertainty. The electron and muon identification efficiency uncertainties were obtained from separate Belle studies of the two-photon processes $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ and of $J/\psi \rightarrow \ell^+\ell^-$, where $\ell = e, \mu$. The uncertainty in the reconstruction efficiency due to the hadron identification is determined using $D^+ \rightarrow D^0[K^-\pi^+]\pi^+$ decays, where the hadron identity is unambiguously determined by its charge. The uncertainty due to the tracking efficiency is calculated by comparing data-MC differences in the reconstruction efficiencies of $D^{\pm} \rightarrow D^0[K^0_S\{\pi^+\pi^-\}]^{\pm}$ decays. The hadron, electron and muon identification and tracking uncertainties are weighted by the reconstruction efficiencies of the corresponding $B$ decay modes. The misreconstructed signal uncertainty is obtained by varying the misreconstructed fraction by ±20% of its value, which is a conservative estimate. The parametric and nonparametric shapes describing the background are varied within their uncertainties. For nonparametric shapes (i.e., histograms), we modify the histogram PDFs bin by bin according to a Poisson distribution and extract the branching fraction from a fit to the data. We perform 300 tests with such modified histogram PDFs and take the width of the resulting Gaussian branching-fraction distribution as a systematic uncertainty. We find that the decay $B^0 \rightarrow \psi(2S)K^0_S[\pi^0\pi^0]$ peaks in the signal region of $M'_{bc}$. The $B^0 \rightarrow \psi(2S)K^0_S[\pi^0\pi^0]$ yield in the $b \rightarrow (c\bar{c})q$ background sample is varied by the uncertainty of its world average branching fraction and the resulting difference in the $B^0 \rightarrow \psi(2S)\pi^0$ branching fraction is taken as a systematic uncertainty. The number of peaking background events obtained from the sideband study is varied by one standard deviation ($\sigma$), and the difference in the branching fraction is assigned as a systematic uncertainty. The same approach is used for the $M'_{bc}$ and $\Delta E$ correction factors. Half the branching-fraction fit bias obtained from pseudo-experiments is taken as an additional systematic uncertainty. The total systematic uncertainty is 6.5% of the $B^0 \rightarrow \psi(2S)\pi^0$ branching fraction.

We perform a likelihood scan to obtain the statistical significance of our branching fraction measurement. We convolve the $\mathcal{L}$ distribution with a Gaussian with a zero mean and a width equal to the systematic uncertainty. The change in the $-2\log\mathcal{L}$ distribution as a function of the branching fraction is shown in Fig. 2. The statistical significance of 7.2σ is determined from $\sqrt{-2\Delta\log\mathcal{L}}$, where $\Delta\log\mathcal{L}$ is the likelihood difference between zero and the observed branching fraction. This includes the systematic uncertainties.

In summary, we report a measurement of the $B^0 \rightarrow \psi(2S)\pi^0$ branching fraction based on the full Belle data set collected at the $\Upsilon(4S)$ resonance. We obtain $\mathcal{B}(B^0 \rightarrow \psi(2S)\pi^0) = (1.17 \pm 0.17 (\text{stat}) \pm 0.08 (\text{syst})) \times 10^{-5}$. Our results are consistent with the naïve expectation that the $B^0 \rightarrow \psi(2S)\pi^0$ to $B^0 \rightarrow J/\psi K^0_S$ branching fraction ratio should be similar to the $B^0 \rightarrow J/\psi\pi^0$ to $B^0 \rightarrow J/\psi K^0_S$ ratio. The $\mathcal{B}(B^0 \rightarrow \psi(2S)\pi^0)$ result has a
significance of $7.2\sigma$, which indicates the first observation of the decay $B^0 \rightarrow \psi(2S)\pi^0$.

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 (Australia); FWF (Austria); NSFC and CCEPP (China); MSMT (Czechia); CZF, DFG, and VS (Germany); DST (India); INFN (Italy); MOE, MSIP, NRF, BK21Plus, WCU and RSRI (Korea); MNiSW and NCN (Poland); MES and RFAAE (Russia); ARRS (Slovenia); IKERBASQUE and UPV/EHU (Spain); SNSF (Switzerland); NSC and MOE (Taiwan); and DOE and NSF (USA).