1. Introduction

In the Standard Model (SM), the purely leptonic decay $B^- \rightarrow \tau^- \bar{\nu}_\tau$ proceeds via annihilation of $b$ and $\tau$ quarks to a $W^-$ boson. It provides a direct determination of the product of the $B$ meson decay constant $f_B$ and the magnitude of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element $|V_{ub}|$. The branching fraction is given by

$$B(B^- \rightarrow \tau^- \bar{\nu}_\tau) = \frac{G_F^2 m_B m_\tau^2}{8\pi} \left(1 - \frac{m_\tau^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B,$$

where $G_F$ is the Fermi coupling constant, $m_B$ and $m_\tau$ are the $B$ and $\tau$ masses, respectively, and $\tau_B$ is the $B^-$ lifetime [1].

Physics beyond the SM, such as supersymmetry or two-Higgs doublet models, could modify $B(B^- \rightarrow \tau^- \bar{\nu}_\tau)$ through the introduction of a charged Higgs boson [2].

Purely leptonic $B$ decays have not been observed in past experiments. The most stringent upper limit on $B^0 \rightarrow \tau^- \bar{\nu}_\tau$ comes from the BaBar experiment: $B(B^0 \rightarrow \tau^- \bar{\nu}_\tau) < 2.6 \times 10^{-4}$ (90\% C.L.) [3].

The possibility to study decays of $B_s$ at very high luminosity $e^+e^-$ colliders running at the energy of the $\Upsilon(5S)$ resonance has been discussed in several theoretical papers [4, 5]. The first data at the $\Upsilon(5S)$ were taken many years ago at CESR [6–8], but the collected data sample was not enough to observe a $B_s$ signal. In 2003, the CLEO experiment collected 0.42 fb$^{-1}$ at the $\Upsilon(5S)$ and observed some evidence for $B_s$ meson production in both inclusive and exclusive modes. However, simple calculations assuming an approximate $SU(3)$ symmetry indicate that many interesting $B_s$ measurements require a data sample of at least 20 pb$^{-1}$, which can be collected by $B$ Factories in the future. To test the experimental feasibility of such measurements, a data sample of 1.86 fb$^{-1}$ was recently taken with the Belle detector at the center-of-mass (CM) energy corresponding to the mass of the $\Upsilon(5S)$ resonance. This data sample is more than four times larger than the CLEO dataset at the $\Upsilon(5S)$.

The Belle detector is a large-solid-angle magnetic spectrometer consisting of a silicon vertex detector, a 50-layer central drift chamber (CDC), a system of aerogel threshold Čerenkov counters (ACC), time-of-flight scintillation counters (TOF), and an electromagnetic calorimeter comprised of CsI(Tl) crystals (ECL) located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux-return located outside of the coil is instrumented to identify $K_1^{0}$ and muons. The detector is described in detail elsewhere [9].

2. Evidence of the Purely Leptonic Decay $B^- \rightarrow \tau^- \bar{\nu}_\tau$

We use a 414 fb$^{-1}$ data sample containing $447 \times 10^6$ $B$ meson pairs collected with the Belle detector at the KEKB asymmetric-energy $e^+e^-$ (3.5 on 8 GeV) collider operating at the $\Upsilon(4S)$ resonance ($\sqrt{s} = 10.58$ GeV).

We use a detailed Monte Carlo (MC) simulation, which fully describes the detector geometry and response based on GEANT [10], to determine the signal selection efficiency and to study the background. In order to reproduce effects of beam background, data taken with random triggers for each run period are overlaid on simulated events. The $B^- \rightarrow \tau^- \bar{\nu}_\tau$ signal decay is generated by the EvtGen package [11]. To model the background from $e^+e^- \rightarrow B\overline{B}$ and continuum $q\overline{q}$ ($q = u, d, s, c$) production processes, large $B\overline{B}$ and $q\overline{q}$ MC samples corresponding to about twice the data sample are used. We also use MC samples for rare $B$ decay processes, such as charmless hadronic, radiative, electroweak decays and $b \rightarrow u$ semileptonic decays.

We fully reconstruct one of the $B$ mesons in the event, referred to hereafter as the tag side ($B_{\text{tag}}$), and compare properties of the remaining particle(s), referred to as the signal side ($B_{\text{sig}}$), to those expected for signal and background. The method allows us to suppress strongly the combinatorial background from both $B\overline{B}$ and continuum events. In order to avoid exp-
perimental bias, the signal region in data is not looked at until the event selection criteria are finalized.

The $B_{\text{tag}}$ candidates are reconstructed in the following decay modes: $B^+ \rightarrow \overline{D}^{(*)0}\pi^+$, $D^{(*)0}\overline{\rho}^+$, $\overline{D}^{(*)0}\mu^+$, and $\overline{D}^{(*)0}D_s^{(*)+}$. The $D$ mesons are reconstructed as $D^0 \rightarrow K^+\pi^-$, $K^+\pi^-\pi^0$, $K^+\pi^-\pi^+\pi^-$, $K^0_S\pi^+$, $K^0_S\pi^-\pi^+\pi^0$ and $K^-K^+$, and the $D_s^+$ mesons are reconstructed as $D_s^+ \rightarrow K_S^0K^+$ and $K^+K^−\pi^+$. The $\overline{D}^{*0}$ and $D_s^{*+}$ mesons are reconstructed in $\overline{D}^{*0} \rightarrow D^{0}\pi^0, D^{0}\gamma$, and $D_s^{*+} \rightarrow D_s^{+}\gamma$ modes. The selection of $B_{\text{tag}}$ candidates is based on the beam-constrained mass $M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - p_B^2}$ and the energy difference $\Delta E \equiv E_B - E_{\text{beam}}$. Here, $E_B$ and $p_B$ are the reconstructed energy and momentum of the $B_{\text{tag}}$ candidate in the $e^+e^−$ center-of-mass system, and $E_{\text{beam}}$ is the beam energy in the CM frame. The selection criteria for $B_{\text{tag}}$ are defined as $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $−80 \text{ MeV} < \Delta E < 60 \text{ MeV}$. If an event has multiple $B_{\text{tag}}$ candidates, we choose the one having the smallest $\chi^2$ based on deviations from the nominal values of $\Delta E$, the $D$ candidate mass, and the $D^* - D$ mass difference if applicable.

In the events where a $B_{\text{tag}}$ is reconstructed, we search for decays of $B_{\text{sig}}$ into a $\tau$ and a neutrino. Candidate events are required to have one or three charged track(s) on the signal side with the total charge being opposite to that of $B_{\text{tag}}$. The $\tau$ lepton is identified in the five decay modes, $\mu^-\bar{\nu}_\mu\nu_\tau$, $e^-\bar{\nu}_e\nu_\tau$, $\pi^-\nu_\tau$, $\pi^-\pi^+\nu_\tau$, and $\pi^-\pi^+\pi^−\nu_\tau$, which taken together correspond to 81% of all $\tau$ decays [1]. The muon, electron and charged pion candidates are selected based on information from particle identification devices. The leptons are selected with requirements that have efficiencies greater than 90% for both muons and electrons in the momentum region above 1.2 GeV/$c$, and misidentification rates of less than 0.2%/1.5% for electrons (muons) in the same momentum region. Kaon candidates are rejected for all charged tracks on the signal side. The $\pi^0$ candidates are reconstructed by requiring the invariant mass of two $\gamma$‘s to satisfy $|M_{\gamma\gamma} - m_{\pi^0}| < 20 \text{ MeV}/c^2$. For all modes except $\tau^- \rightarrow \pi^-\pi^0\nu_\tau$, we reject events with $\pi^0$ mesons on the signal side. All the selection criteria have been optimized to achieve the highest sensitivity in MC.

The most powerful variable for separating signal and background is the remaining energy in the ECL, denoted as $E_{\text{ECL}}$, which is sum of the energy of photons that are not associated with either the $B_{\text{tag}}$ or the $\pi^0$ candidate from the $\tau^- \rightarrow \pi^-\pi^0\nu_\tau$ decay. For signal events, $E_{\text{ECL}}$ must be either zero or small value arising from beam background hits, therefore, signal events peak at low $E_{\text{ECL}}$. On the other hand, background events are distributed toward higher $E_{\text{ECL}}$ due to the contribution from additional neutral clusters.

The $E_{\text{ECL}}$ signal region is optimized for each $\tau$ decay mode based on the MC simulation, and is defined by $E_{\text{ECL}} < 0.2 \text{ GeV}$ for the $\mu^-\bar{\nu}_\mu\nu_\tau$, $e^-\bar{\nu}_e\nu_\tau$ and $\pi^-\nu_\tau$ modes, and $E_{\text{ECL}} < 0.3 \text{ GeV}$ for the $\pi^-\pi^0\nu_\tau$ and $\pi^-\pi^+\pi^-\nu_\tau$ modes. The $E_{\text{ECL}}$ sideband region is defined by 0.4 GeV < $E_{\text{ECL}}$ < 1.2 GeV for the $\mu^-\bar{\nu}_\mu\nu_\tau$, $e^-\bar{\nu}_e\nu_\tau$ and $\pi^-\nu_\tau$ modes, and by 0.45 GeV < $E_{\text{ECL}}$ < 1.2 GeV for the $\pi^-\pi^0\nu_\tau$ and $\pi^-\pi^+\pi^-\nu_\tau$ modes. Table I shows the number of events found in the sideband region for data ($N_{\text{side}}$) and for the background MC simulation ($N_{\text{MC}}$) scaled to the equivalent integrated luminosity in data. Their good agreement for each $\tau$ decay mode indicates the validity of the background MC simulation. Table I also shows the number of the background MC events in the signal region ($N_{\text{sig}}$). In order to validate the $E_{\text{ECL}}$ simulation, we use a control sample of events (double tagged events), where the $B_{\text{tag}}$ is fully reconstructed as described above and $B_{\text{sig}}$ is reconstructed in the decay chain, $B^- \rightarrow D^{*0}\ell^−\bar{\nu}$ ($D^{*0} \rightarrow D^0\pi^0$), followed by $D^0 \rightarrow K^-\pi^+$ or $K^-\pi^−\pi^+\pi^+$ where $\ell$ is a muon or electron. The sources affecting the $E_{\text{ECL}}$ distribution in the control sample are similar to those affecting the $E_{\text{ECL}}$ distribution in the signal MC simulation. Figure 1 shows the $E_{\text{ECL}}$ distribution in the control sample for data and the MC simulation scaled to equivalent integrated luminosity in data. Their agreement demonstrates the validity of the $E_{\text{ECL}}$ simulation in the signal MC.

After finalizing the signal selection criteria, the sig-

![Figure 1](https://example.com/figure1.png)

Figure 1: $E_{\text{ECL}}$ distribution for the control sample of doubly tagged events, where one $B$ is fully reconstructed and the other is reconstructed as $B^- \rightarrow D^{*0}\ell^−\bar{\nu}$. The dots with errors indicate the data. The solid histogram represents the distribution as deduced background from $B\overline{B}$ MC ($B^+B^- + B^0\overline{B}^0$), and the dashed histogram shows the contribution from $B^0\overline{B}^0$ events.
where $E_i$ is the $E_{ECL}$ in the $i$th event, $N$ is the total number of events in the data, and $n_s$ and $n_b$ are the signal yield and background yield to be determined by the fit. To combine likelihood functions of the five decay modes, we multiply the likelihood functions to produce the combined likelihood ($L_{\text{com}} = \prod_{i=1}^{5} L_i$). The results are listed in Table I. The number of signal events in the signal region deduced from the fit ($N_s$) is $21.2^{+5.7}_{-7}$ when all $\tau$ decay modes are combined. Table I also gives the number of background events in the signal region deduced from the fit ($N_b$), which is consistent with the expectation from the background MC simulation ($N_{b,\text{MC}}$).

The branching fractions are calculated as $B = N_s / (2 \cdot N_{B^+B^-})$ where $N_{B^+B^-}$ is the number of $\Upsilon(4S) \to B^+B^-$ events, assumed to be half of the number of produced $B$ meson pairs. The efficiency is defined as $\varepsilon = \varepsilon_{\text{tag}} \times \varepsilon_{\text{sel}}$, where $\varepsilon_{\text{tag}}$ is the tag reconstruction efficiency for events with $B^- \to \tau^- \bar{\nu}_\tau$ decays on the signal side, determined by MC to be $0.136 \pm 0.001(\text{stat})\%$, and $\varepsilon_{\text{sel}}$ is the event selection efficiency listed in Table I, as determined by the ratio of the number of events surviving all of the selection criteria including the $\tau$ decay branching fractions over the number of fully reconstructed $B^{\pm}$. When all $\tau$ decay modes are combined we obtain a branching fraction of $(1.06^{+0.34}_{-0.28}) \times 10^{-4}$. The branching fraction for each $\tau$ decay mode is consistent within error as shown in Table I.

Systematic errors for the measured branching fraction are associated with the uncertainties in the number of $B^+B^-$, signal yields and efficiencies. The total fractional uncertainty of the combined measurement is $^{+24.0}_{-25.5}\%$, and we measure the branching fraction to be

$$B(B^- \to \tau^- \bar{\nu}_\tau) = (1.06^{+0.34}_{-0.28}(\text{stat})^{+0.22}_{-0.25}(\text{syst})) \times 10^{-4}.$$

The significance is 4.0$\sigma$ when all $\tau$ decay modes are combined, where the significance is defined as $\Sigma = \sqrt{-2 \ln (L_0 / L_{\text{max}})}$, where $L_{\text{max}}$ and $L_0$ denote the maximum likelihood value and likelihood value obtained assuming zero signal events, respectively. Here the likelihood function from the fit is convolved with a Gaussian systematic error function in order to include the systematic uncertainty in the signal yield.

3. Results from the $\Upsilon(5S)$ Engineering Run

We use a data sample of 1.86 fb$^{-1}$ taken at the $\Upsilon(5S)$ energy of $\sim 10869$ MeV. The experimental conditions of data taking at $\Upsilon(5S)$ are identical to that for $\Upsilon(4S)$ or continuum running. The data sample of 3.67 fb$^{-1}$ taken in the continuum at an energy of 60 MeV below the $\Upsilon(4S)$ was also used in this analysis for comparison.
The $B_s$ mesons are produced at the $\Upsilon(5S)$ through the intermediate $B_s\overline{B}_s$, $B_s^*\overline{B}_s$, $B_s^*\overline{B}_s^*$ or $B_s^*\overline{B}_s^*$ pair production channels, where $B_s$ decays to $B_s\gamma$. These intermediate channels can be distinguished kinematically and their production ratios can be obtained from the reconstruction of exclusive $B_s$ decays. To improve the statistical significance of our exclusive $B_s$ signal, we combined the six modes $B_s \rightarrow D_s^+\pi^-$, $B_s \rightarrow D_s^+\pi^-$, $B_s \rightarrow D_s^+\rho^-$, $B_s \rightarrow D_s^+\rho^-$, $B_s \rightarrow J/\psi\phi$ and $B_s \rightarrow J/\psi\eta$, which have large reconstruction efficiencies and are described by unsuppressed conventional tree diagrams.

Six conventional $B_s$ decays to $D_s^+\pi^-$, $D_s^+\rho^-$, $D_s^+\rho^-$, $J/\psi\phi$ and $J/\psi\eta$ final states are reconstructed. The signals can be observed using two variables: the energy difference $\Delta E = E_{BC} - E_{CM}$ and beam-constrained mass $M_{bc} = \sqrt{(E_{BC}^2 - p_{BC}^2)}$. The $B_s$ mesons are produced at the $\Upsilon(5S)$ energy via the intermediate $e^+e^-\rightarrow B_s^*\overline{B}_s^*$ channels, with $B_s^* \rightarrow B_s\gamma$. The $B_s$ signal regions in $M_{bc}$ and $\Delta E$ are separated for different intermediate channels.

After all selections, the dominant background is from $e^+e^-\rightarrow q\bar{q}$ continuum events ($q = u, d, s, or c$). The distribution of data in $M_{bc}$ and $\Delta E$ for the $B_s \rightarrow D_s^+\pi^-$ decay mode is shown in Figure 3a. Three $D_s^+$ decay modes, $\phi\pi^+$, $K^{*0}K^+$ and $K^{*0}K^+$, are used to reconstruct $B_s$ candidates. Nine events are observed within the $B_s$ signal ellipsoidal region corresponding to $B_s^*\overline{B}_s^*$ pair production channel. Only one event is observed in the $B_s$ signal region for $B_s^*\overline{B}_s + B_s\overline{B}_s^*$ channels, and no events are observed for $B_s^*\overline{B}_s^*$ channel. Background outside the signal regions is small and corresponds to 0.1 event for any of three signal regions. The inclusive studies at the $\Upsilon(5S)$ found that $92,000 \pm 7,900 \pm 23,500 B_s^{(*)}\overline{B}_s^{(*)}$ pairs are contained within that $1.86 \text{ fb}^{-1} \ \Upsilon(5S)$ data sample. Using this value, we measure the branching fraction to be $B(B_s \rightarrow D_s^+\pi^-) = (0.65 \pm 0.21 \pm 0.19)\%$.

The $M_{bc}$ and $\Delta E$ scatter plots are also obtained for the $B_s \rightarrow D_s^+\pi^-$ (Figure 3b) and $B_s \rightarrow D_s^{(*)}\rho^-(c)$ (Figure 3c) decay modes. Again, three $D_s^+$ decay modes, $\phi\pi^+$, $K^{*0}K^+$ and $K^{*0}K^+$, are used to reconstruct $B_s$ candidates. The numbers of events within the $B_s$ signal region for the $B_s^*\overline{B}_s^*$ pair production channel are 4 for $B_s \rightarrow D_s^+\pi^-\rightarrow B_s\gamma$ and 7 for $B_s \rightarrow D_s^{(*)}\rho^-\rightarrow B_s\gamma$ decay.

The scatter plot in $M_{bc}$ for the $B_s \rightarrow J/\psi\phi$ and $B_s \rightarrow J/\psi\eta$ decays is shown in Figure 3d. One of the observed candidates is reconstructed in the $J/\psi\rightarrow \mu^+\mu^-$ mode and one in the $J/\psi\rightarrow e^+e^-$ mode. These two candidates correspond roughly to a branching fraction of $\sim 1 \times 10^{-3}$, in agreement with expectations.

The $B_s$ and $B_s^*$ masses can be extracted from the $M_{bc}$ fits in the $B_s^*\overline{B}_s^*$ channel. The $M_{bc}$ distribution for this channel (Figure 4a) is obtained choosing candidates within the $-0.08 < \Delta E < -0.02$ MeV range. The distribution, shown in Figure 4a, is fitted by the sum of a Gaussian to describe the signal and the ARGUS function to describe the background. The fit
yields the mass value \( M(B_s^*) = 5418 \pm 1 \text{ MeV}/c^2 \). The observed width of the \( B_s \) signal is \( 3.6 \pm 0.6 \text{ MeV}/c^2 \) and agrees with the value obtained from the MC simulation. Using events from the \( B_s^* \bar{B}_s^* \) channel we can obtain also the \( B_s \) mass (Figure 4b). The distribution shown in Figure 4b is fitted to the sum of a Gaussian and the ARGUS function. The fit yields the \( B_s \) mass \( M(B_s) = 5370 \pm 1 \pm 3 \text{ MeV}/c^2 \) and width \( \sigma(B_s) = 3.6 \pm 0.6 \text{ MeV}/c^2 \).

Additionally, we searched for several \( B_s \) rare decays for the first time: the penguin decay \( B_s \to K^+K^- \), the electromagnetic penguin decay \( B_s \to \phi\gamma \), and the intrinsic penguin decay \( B_s \to \gamma\gamma \). We also searched for the tree decay \( B_s \to D_s^{(*)+}D_s^{(*)-} \), which is not yet observed and is of special interest because the \( D_s^{(*)+}D_s^{(*)-} \) states are expected to be dominantly CP eigenstates. Although the branching fractions for these decays are expected to be too low for observation in this analysis, we obtained upper limits (Table II).

![Figure 4: The \( B_s^* \) (a) and \( B_s \) (b) mass distributions for events within the \(-0.08 < \Delta E < -0.02 \text{ MeV} \) interval, corresponding to the \( B_s^* \bar{B}_s^* \) channel.](image)

| Decay mode | Yield events | Background events | Eff. (%) | upper limit \( \times 10^{-4} \) |
|------------|-------------|-------------------|---------|-----------------|
| \( B_s \to K^+K^- \) | 2 | 0.14 | 9.5 | 3.4 |
| \( B_s \to \phi\gamma \) | 1 | 0.15 | 5.9 | 4.1 |
| \( B_s \to \gamma\gamma \) | 0 | 0.5 | 20.0 | 0.56 |
| \( B_s \to D_s^{(*)+}D_s^{(*)-} \) | 0 | 0.02 | 0.020 | 710 |
| \( B_s \to D_s^{(*)+}D_s^{(*)-} \) | 1 | 0.01 | 0.0090 | 1270 |

Table II. The number of events in the signal region, estimated background events and upper limits on the 90% confidence level for \( B_s \to K^+K^- \), \( B_s \to \phi\gamma \), \( B_s \to \gamma\gamma \) and \( B_s \to D_s^{(*)+}D_s^{(*)-} \) decay modes.

**Acknowledgments**

The author wish to thank the KEKB accelerator group for the excellent operation of the KEKB accelerator.

**References**

[1] S. Eidelman et al. (Particle Data Group), Phys. Lett. B 592, 1 (2004).
[2] W. S. Hou, Phys. Rev. D 48, 2342 (1993).
[3] B. Aubert (BABAR Collaboration), Phys. Rev. D 73, 057101 (2006).
[4] A. F. Falk and A. A. Petrov, Phys. Rev. Lett. 85, 252 (2000).
[5] D. Atwood and A. Soni, Phys. Lett. B 533, 37 (2002).
[6] D. M. J. Lovelock et al., Phys. Rev. Lett. B 533, 377 (1985).
[7] D. Besson et al. (CLEO Collaboration), Phys. Rev. Lett. 54, 381 (1985).
[8] J. Lee-Franzini et al., Phys. Rev. Lett. 65 2947 (1990).
[9] A. Abashian et al. (Belle Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A 479, 117 (2002).
[10] R. Brun et al., GEANT3.21, CERN Report DD/EE/84-1 (1984).
[11] See the EvtGen package home page, http://www.slac.stanford.edu/˜lange/EvtGen/.