CDF/ANAL/BOTTOM/PUBLIC/9458
Version 1.0
August 7, 2008

An Updated Measurement of the *CP* Violating Phase $\beta_s^{J/\psi\phi}$

The CDF Collaboration[1]
[1] *www-cdf.fnal.gov*

We present an updated analysis of the *CP* violation parameter $\beta_s^{J/\psi\phi}$ in flavor-tagged $B_s^0 \to J/\psi\,\phi$ decays using 2.8 fb^{-1} of data collected with dedicated J/ψ triggers between February 2002 and February 2008. With the present selection, which neglects the use of dE/dx or time-of-flight information to identify the kaons from the ϕ decay, we find $\sim 3,150$ signal events. We report a two-dimensional profile likelihood in the $\beta_s^{J/\psi\phi} - \Delta\Gamma$ plane adjusted for the non-Gaussian behavior of the uncertainties. We find that the p-value at the standard model expectation, defined as the ratio of the value of the likelihood in which $\beta_s^{J/\psi\phi}$ (and $\Delta\Gamma$) is fixed in the fit to the standard model expected value relative to the value of the likelihood where $\beta_s^{J/\psi\phi}$ (and $\Delta\Gamma$) floats freely in the fit, is 7%. The p-value is the same whether $\beta_s^{J/\psi\phi}$ alone or $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma$ float in the fit. We also report the one-dimensional profile likelihood which has been adjusted in an analogous manner. Additionally, we evaluate the mean B_s^0 lifetime τ_s, the width difference of the heavy and light mass eigenstates $\Delta\Gamma$, and the transversity amplitudes $|A_0(0)|^2$, and $|A_\parallel(0)|^2$ assuming that no *CP* violation ($\beta_s^{J/\psi\phi} = 0.0$) is present and find results consistent with the previous determination of these quantities.

We present an update of the measurement of the *CP*-violating phase $\beta_s^{J/\psi\phi}$ in $B_s^0 \to J/\psi\,\phi$ decays, where $J/\psi \to \mu^+\mu^-$ and $\phi \to K^+K^-$. The previous measurement, which observed a 1.5σ deviation from the standard model, has recently been published in PRL [1]. Since the previous result, there has been considerable interest in the possible discrepancy of $\beta_s^{J/\psi\phi}$ with the standard model predicted value of $\beta_s^{J/\psi\phi} = 0.02$. This is largely due to the combination of the CDF and D0 measurements [2, 3], which indicate an intriguing deviation from the standard model. The standard model prediction is well-constrained by other CKM measurements [3], so if we indeed observe a large value of $\beta_s^{J/\psi\phi}$ it will be a clear sign of new physics [4].

It is interesting to note that the Belle and *BABAR* collaborations have observed an asymmetry between direct *CP* asymmetries of charged and neutral $B \to K\pi$ decays with 5σ significance [5, 6]. In the absence of an under-estimation of the contribution from color-suppressed tree decays, it is difficult to explain this discrepancy without some source of new physics contributing to the electroweak penguin which governs the $b \to s$ transition.In the standard model, this isospin-violating diagram should be highly suppressed, but if a new source of physics is indeed present in these transitions it may be enough to cause the different *CP* asymmetries that have been observed.. In the $B_s^0 \to J/\psi\,\phi$ decay, the $b \to s$ transition occurs through the mixing box diagram shown in Fig. 1. It is possible that new particles could enter this transition through the $b \to s$ quark transition. While there are surely a number of possible sources of new physics that might give rise to such discrepancies, George Hou predicted the presence of a t' quark with mass between ~ 300 and 1,000 GeV/c^2 in order to explain the Belle result and predicted *a priori* the observation of a large *CP*-violating phase in $B_s^0 \to J/\psi\,\phi$ decays [7, 8]. Another result of interest in the context of these measurements is the excess observed at ~ 350 GeV/c^2 in the recent t' search at CDF using 2.3 fb^{-1} of data [9]. In this direct search for a fourth generation up-type quark, a significance of less than 2σ is obtained for the discrepancy between the data and the predicted backgrounds, so that the effect, while intriguing, is presently consistent with a statistical fluctuation. A updated search with more data would also clearly be of interest, particularly if a large value of $\beta_s^{J/\psi\phi}$ persists with the addition of more data.

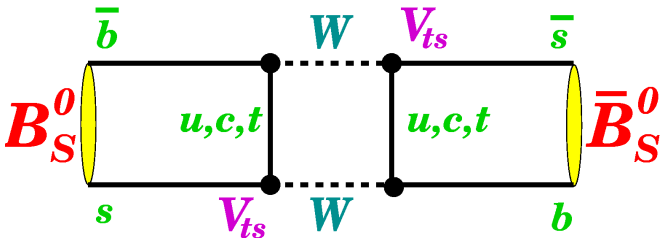

FIG. 1: Feynman diagrams for the B_s^0 mixing tree diagram. New physics, if present, would be expected to contribute to the $b \to s$ transition.

The measurement presented in this note follows closely the techniques and strategy of the previous tagged analysis of $\beta_s^{J/\psi\phi}$ [1]. The reconstructed data is selected via an artificial neural network (ANN) and the treatment of the transversity angles is also identical to the previous measurement. Because the particle identification (PID) is not yet calibrated for the entire 2.8 fb^{-1} of data, the tagging uses only opposite-side flavor tags (OST) after the first 1.35 fb^{-1} of data and no PID is used in the selection. The OST is calibrated on $B^+ \to J/\psi\,K^+$ decays separately for B^+ and B^-, while the same-side kaon tag (SSKT) is calibrated for an admixture of B_s^0 and $\bar{B}_s^0$ sample on the first half of the data using Pythia Monte Carlo, as was done in the measurement of the B_s^0 oscillation frequency [10, 11]. The likelihood fit is unchanged for the default fit in which $\beta_s^{J/\psi\phi}$ floats freely and we also perform a standard model fit where $\beta_s^{J/\psi\phi} = 0.0$ is fixed in the fit.

I. DATA SELECTION AND RECONSTRUCTION

In this update we use data corresponding to 2.8 fb^{-1} of data collected with the CDF II detector between February 2002 and February 2008. As in the previous measurement, we use data collected with the di-muon J/ψ trigger. For our final selection, we use an ANN which is identical to that used previously, except that the PID from dE/dx and time-of-flight (TOF) is not used to help identify the kaons from the ϕ decay. In Table I we list the input quantities to B_s^0 neural network, while in Table II we list the input quantities to B^+ neural network. The quantities denoted as correlation significance and significance loss are calculated based on the correlation of the variables to the classification

Index	Variable	Rank	Significance $[\sigma]$	Loss $[\sigma]$
1	training target	-	-	-
2	$\chi^2_{r-\phi}(B_s)$	9	90.4	10.1
3	$p_T(B_s^0)$	10	114.1	11.2
4	$Prob(\chi^2)(B_s^0)$	4	75.7	15.5
5	$p_T(\phi)$	14	178.7	7.8
6	$Prob(\chi^2)(\phi)$	11	0.9	10.5
7	$\lvert m_{K^+K^-}\rvert$	2	100.4	48.6
8	$p_T(K_1)$	3	174.2	26.4
9	$p_t(K_2)$	1	174.1	29.2
10	$p_T(J/\psi)$	6	25.4	17.6
11	$Prob(\chi^2)(J/\psi)$	15	24.7	0.9
12	$\lvert m_{\mu^+\mu^-} - m^{PDG}_{J/\psi}\rvert$	12	20.1	9.7
13	$\max(\mathrm{lh}(\mu^+), \mathrm{lh}(\mu^-))$	13	60.1	8.9
14	$\min(\mathrm{lh}(\mu^+), \mathrm{lh}(\mu^-))$	5	68.7	24.8
15	$\max(p_T(\mu^+), p_T(\mu^-))$	7	17.6	22.0
16	$\min(p_T(\mu^+), p_T(\mu^-))$	8	15.8	18.1

TABLE I: List of input quantities used in B_s^0 neural network. Rank sorts inputs according to their importance. The significance gives the amount of information a given variable has on its own, while the loss tells how much of separation power we would loose if we remove single variable while keeping all others. $\mathrm{lh}(\mu^+)$ is muon likelihood as used by the soft muon tagger [12].

truth. To calculate the correlation significance, one calculates for each variable the correlation to the truth defined as

$$\rho^{ti} = \frac{\frac{1}{n}\sum_{j=1}^{n}\left(x_j^t - < x^t >\right)\cdot\left(x_j^i - < x^i >\right)}{\sqrt{V[x^t]V[x^i]}},$$

where t denotes the truth, i the correlated variable, $< x^i >$ the expectation value for the given variable, $V[x^i]$ its variance and n is number of events in training sample. Then the correlation significance for a given variable is $\rho^{ti}\cdot\sqrt{n}$. For the significance loss, we first decorrelate all input quantities and than calculate the correlation between the truth and each decorrelated variable $\tilde{\rho}^{ti}$. The total correlation is then defined by

$$\rho^2_{TN} = \sum_{i=1}^{N}\tilde{\rho}^{ti2},$$

where N is the number of variables. An analogous procedure is repeated without the considered variable to calculate ρ^2_{TN-1}. The significance loss is then given by

$$\sqrt{\rho^2_{TN} - \rho^2_{TN-1}}\cdot\sqrt{n}.$$

Intuitively, the correlation significance is proportional to the amount of information which is provided by a given quantity without all others while the significance loss corresponds to the amount of information lost for the neural network if we remove the given quantity while we keep all others.

Additionally, in order to guarantee good vertex resolution, we require all four daughter tracks from the B_s^0 decay to have at least three $r-\phi$ silicon hits. The distributions of the ANN output for signal and background from the B^+ and B_s^0 training samples are shown in Fig. 2.

We choose our selection value of the ANN output variable that optimizes the standard figure-of-merit (FOM) $S/\sqrt{S+B}$, as was done previously. While we have investigated alternate optimal values of ANN selection in order to maximize sensitivity to $\beta_s^{J/\psi\phi}$ with toy Monte Carlo studies, we have not found any significant improvement in sensitivity to $\beta_s^{J/\psi\phi}$ relative to the standard FOM. Consequently, we choose to use the usual FOM for the optimization of the neural network selection variable. Optimization plots for the B^+ and B_s^0 are shown in Fig. 3, where both signal and background are determined from a single Gaussian fit to the signal region, shown in Fig. 4. We choose to apply a cut on the ANN output variable at 0.8 for the B^+ signal and at 0.6 for the B_s^0 signal. With these selections we find $33,868 \pm 194$ B^+ signal candidates and $3,166 \pm 56$ B_s^0 signal candidates. The B_s^0 yield is lower by $\approx$ 600 candidates than if an ANN with PID was used in the selection. Additionally, it might be possible to loosen the cut on the ANN output variable when PID is included, as previous studies have indicated that the optimal value of the selection variable is lower when PID information is included. Consequently, when calibrated PID is available we stand to gain between ~ 600 and $1,000$ $B_s^0 \to J/\psi\,\phi$ candidates in the same 2.8 fb^{-1} of data.

Index	Variable	Rank	Significance $[\sigma]$	Loss $[\sigma]$
1	Target	-	-	-
2	$\lvert d_0(B^+)\rvert$	6	39.7	12.7
3	$Prob(\chi^2)(B^+)$	3	50.4	13.3
4	$p_T(J/\psi)$	4	106.1	14.9
5	$p_T(K^+)$	1	340.1	246.2
6	$helA(\mu_1)$	8	139.0	7.1
7	$\lvert\eta(K^+)\rvert$	10	144.6	6.0
8	$p_T(B^+)$	2	176.6	54.7
9	$M(J/\psi)$	5	33.6	14.1
10	$Prob(\chi^2)(J\psi)$	9	13.1	6.6
11	$\max(p_T(\mu^+), p_T(\mu^-))$	11	90.6	0.3
12	$\min(p_T(\mu^+), p_T(\mu^-))$	7	65.3	10.1

TABLE II: List of input quantities used in B^+ neural network. Rank sorts inputs according to their importance. Significance gives amount of information given variable has on its own, while loss tells us how much of separation power we would loose if we remove single variable while keeping all others.

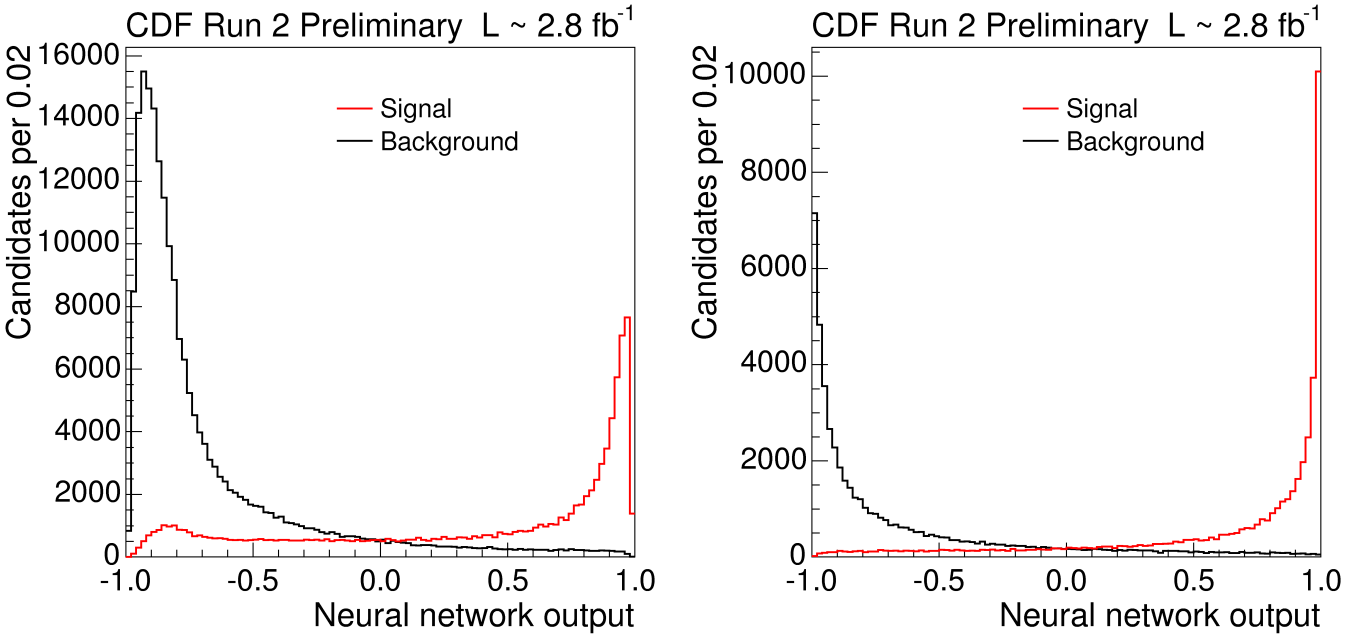

FIG. 2: Distribution of signal and background in ANN output variable for B^+ (left) and B_s^0 (right).

II. TRANSVERSITY ANGLES

As in the previous untagged and tagged measurements of $\beta_s^{J/\psi\phi}$, this measurement makes use of the transversity angles $\vec{\rho} = (\cos\theta_T, \phi_T, \cos\psi)$ defined in [13], in order to separate the CP-even and CP-odd components of the $J/\psi\,\phi$ final state. In addition to the straight-forward theoretical prediction for the behavior of the signal, we must take into account the detector effects on the predicted distribution of the transversity angles and we must model the transversity angle distributions of the background. Realistic $B_s^0 \to J/\psi\,\phi$ Monte Carlo, generated according to a phase space decay model, is used to determine the detector sculpting of the angles due to the non-hermeticity of the CDF II detector. The three-dimensional distribution of the angles are fit with expansions of the Legendre polynomials and spherical harmonics, as has been done in the previous measurement [1], to describe the efficiencies in the transversity angles. These efficiencies are then applied to the signal probability distribution function (PDF) in the maximum likelihood fit. The background transversity angles, determined from the B_s^0 mass sidebands, are fit with an empirical model that is found to describe the data.

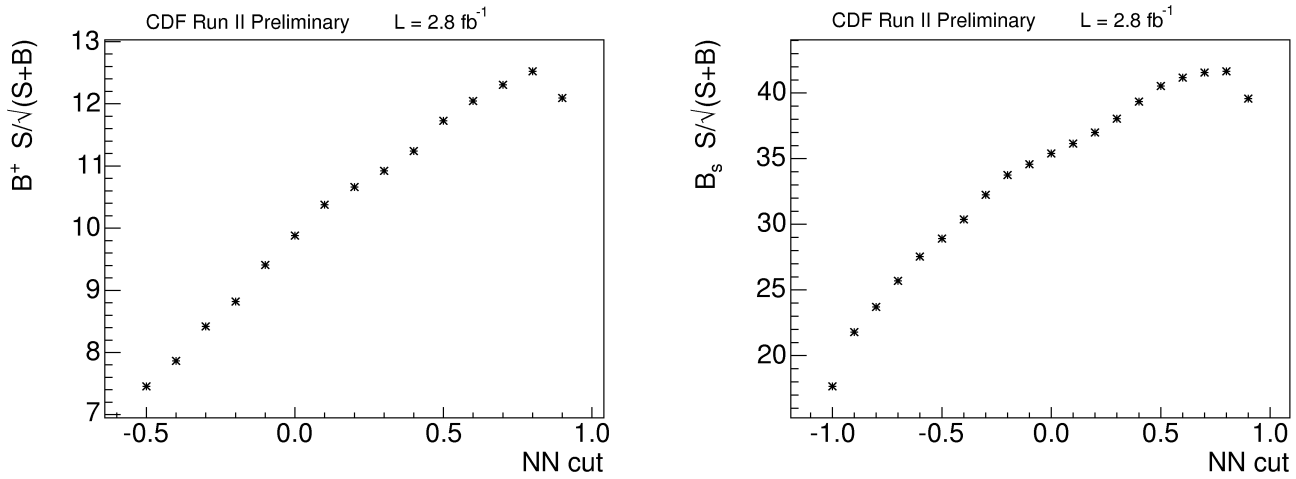

FIG. 3: Optimization of neural network selection for $B^+ \to J/\psi\, K^+$ decays (left) and $B_s^0 \to J/\psi\, \phi$ decays (right). The neural network for the B_s^0 does not use PID for identification of the kaons.

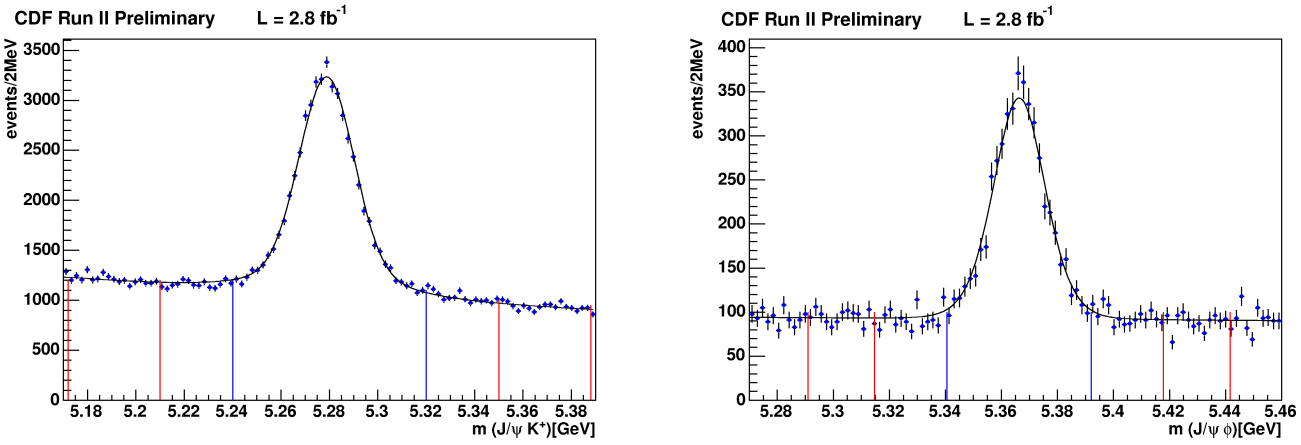

FIG. 4: Invariant mass distributions of $\mu^+\mu^- K^+$ (left) and $\mu^+\mu^- K^+ K^-$ (right). We find $33,868 \pm 194$ $B^+ \to J/\psi\, K^+$ events and $3,166 \pm 56$ $B_s^0 \to J/\psi\, \phi$ events. The neural network selection variable does not use PID for identification of the kaons.

III. FLAVOR TAGGING

In this measurement, we use flavor tagging to distinguish the flavor of the B_s^0 at production (*i.e.* whether a B_s^0 or $\bar{B}_s^0$ was produced) in order to maximize our sensitivity to the *CP*-violating phase $\beta_s^{J/\psi\phi}$. Flavor tagging at CDF takes advantage of the fact that b quarks are produced in quark-antiquark pairs at the Tevatron. Consequently, we can determine information about the flavor of the B_s^0 at production by looking at tracks associated with the hadronization of the $b/\bar{b}$ quark that produced the $\bar{B}_s^0/B_s^0$ we observe (called same-side tags), or by looking at the decay products of B hadrons produced by the other $\bar{b}/b$ quark in the event (called opposite-side tags.)

Previously we have used a hierarchical combination of various OST, namely the soft-muon tag (SMT) [12], the soft-electron tag (SET) [14], and the jet-charge tag (JQT) [15]. We now use an ANN combination of the OST for more optimal tagging power. After the OST decision is made, it is combined with the SSKT [11] as per the usual combination of two independent flavor tags [10]. Although a new tagger which will combine both same-side and opposite-side tagging information into a single tag decision is under development [16], this new flavor tag is not yet ready, as it requires the updated PID calibrations.

Scale Factor	02/02 - 09/06	11/06 - 02/08
$S^+_{\mathcal{D}}$	0.90 ± 0.09	1.11 ± 0.11
$S^-_{\mathcal{D}}$	1.10 ± 0.09	1.07 ± 0.11

TABLE III: OST dilution scale factors measured separately in the B^+ and B^- samples for different periods of data.

A. Opposite-side Flavor Tagging

The ANN OST used in this measurement improves the tagging power of the OST by $\sim 20\%$, from $\epsilon\mathcal{D}^2 \approx 1.5\%$ to 1.8% [10]. We use a symmetrizing procedure in the ANN OST to remove any asymmetry between particle and anti-particle. In order to check this, we compare the measured dilution ($\mathcal{D} = \frac{RS-WS}{RS+WS}$) in $B^+ \to J/\psi\, K^+$ data as a function of the predicted dilution, which is included event-per-event in the un-binned maximum likelihood fit. These distributions give us information about two things: (1) whether an asymmetry between particle and anti-particle exists and (2) a rough measure of the dilution scale factor needed to translate the predicted dilution from the ℓ+SVT trigger sample in which it was determined to the present di-muon J/ψ trigger. Distributions of measured vs. predicted dilutions in B^+ can be seen in Fig. 5 for data taken before and after September, 2006, while the same distributions can be seen for B^- in Fig. 6. Figure 7 shows the combined sample for data before and after the first 1.35 fb^{-1} of data. We find from these distributions that we expect some difference between the dilution scale factors before and after the first 1.35 fb^{-1} of data.

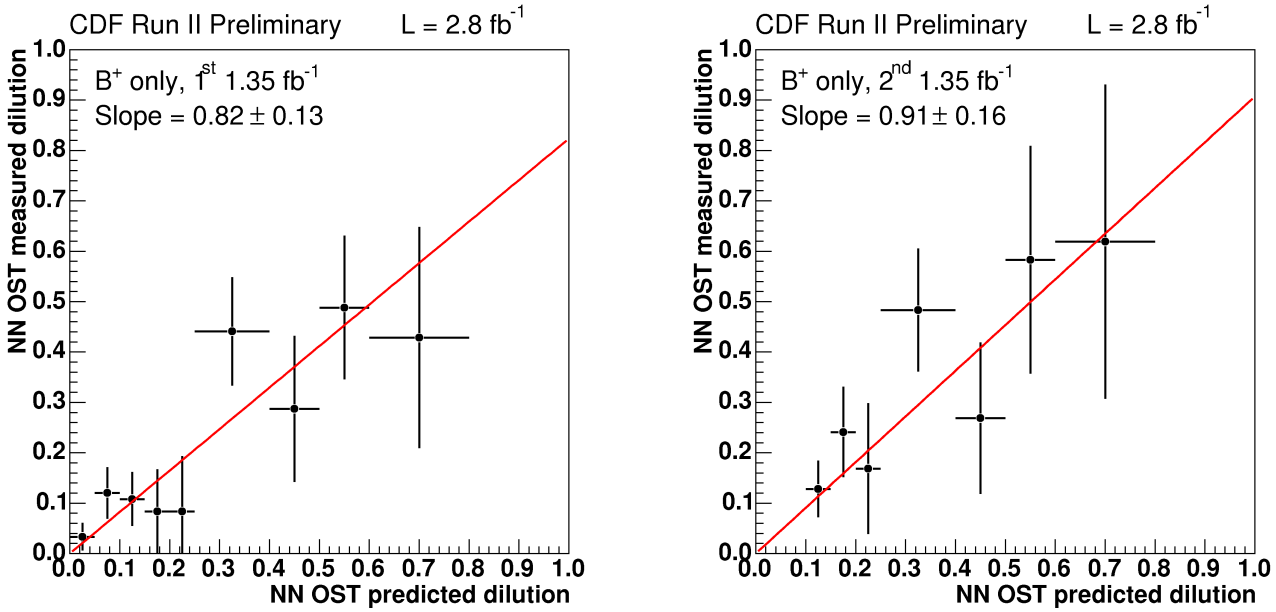

FIG. 5: Comparison of measured dilution as a function of predicted dilution in B^+ data for the first 1.35 fb^{-1} of data (left) and the second 1.4 fb^{-1} of data (right).

While we could use the slopes of the lines in Figs. 5-7 to obtain the calibration of the OST predicted dilutions, which are determined from a dataset with a higher p_T spectrum than the data used in this measurement, we improve our uncertainty on the calibration by performing a simultaneous un-binned maximum likelihood fit of the mass and lifetime in $B^+ \to J/\psi\, K^+$ decays. We fit for separate predicted dilution calibration scale factors for B^+ ($S^+_{\mathcal{D}}$) and B^- ($S^-_{\mathcal{D}}$), which are listed in Table III. Previously we averaged the two calibration scale factors and took the difference from the mean to the individual scale factors as a systematic uncertainty. In this measurement, however, we choose to include the two scale factors separately in the likelihood fit of the $B^0_s \to J/\psi\, \phi$ decays. We also choose to use two different sets of dilution scale factors for data taken before and after September, 2006, which correspond to the first 1.35 fb^{-1} of data and the second 1.4 fb^{-1} of data, respectively.

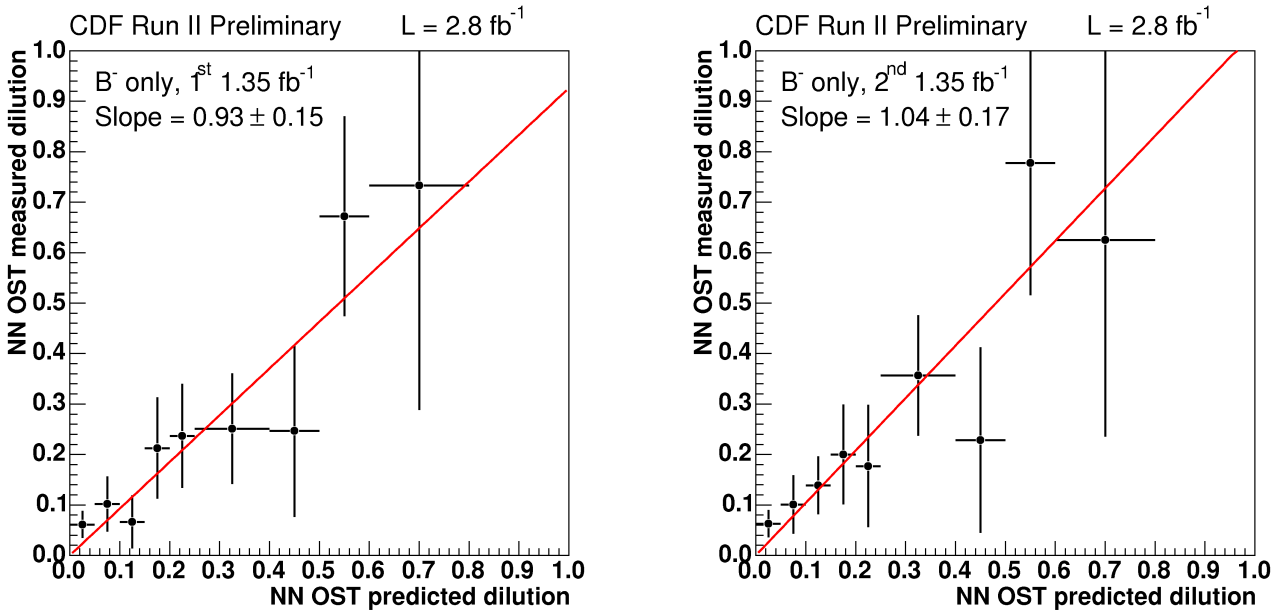

FIG. 6: Comparison of measured dilution as a function of predicted dilution in B^- data for the first 1.35 fb^{-1} of data (left) and the second 1.4 fb^{-1} of data (right).

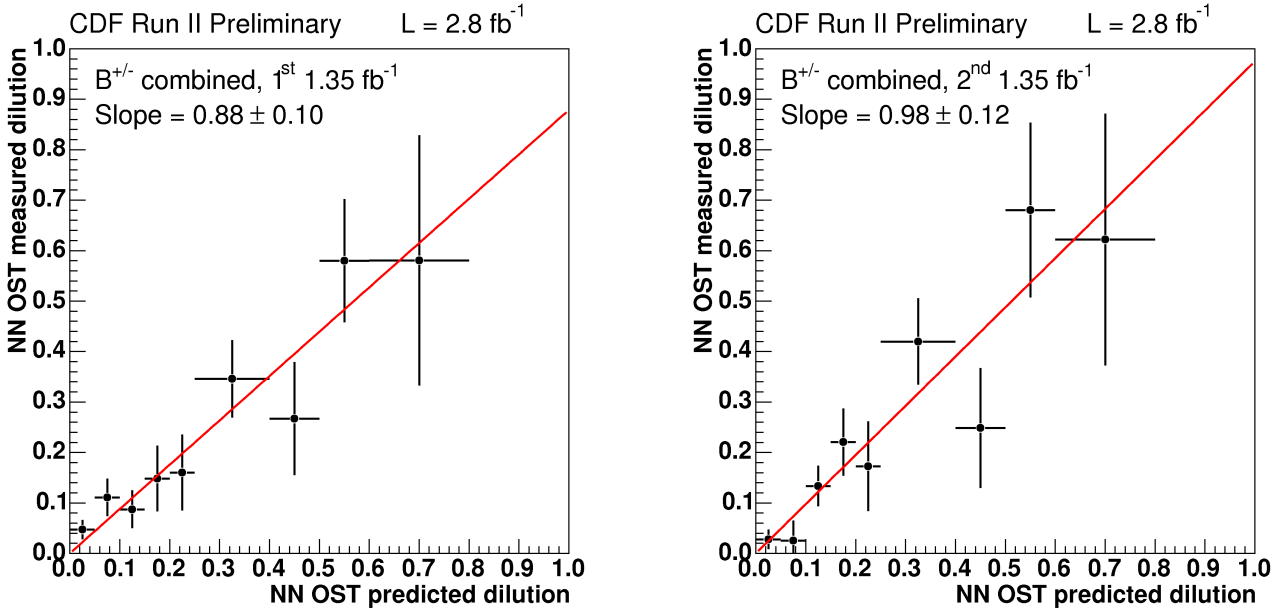

FIG. 7: Comparison of measured dilution as a function of predicted dilution in combined B^+/B^- data for the first 1.35 fb^{-1} of data (left) and the second 1.4 fb^{-1} of data (right).

B. Same-side Flavor Tagging

Since calibrations for the PID for the second half of the data set are not yet ready, we only include the SSKT for data taken before September, 2006. We again use an ANN tagging algorithm for the SSKT [11] and use the dilution scale factors determined for the B_s^0 mixing measurement [10], listed in Table IV.

Scale Factor	First 360 pb^{-1}	Second 400 pb^{-1}	Third 210 pb^{-1}
$S_{\mathcal{D}}$	$0.992^{+0.107}_{-0.143}$	$0.959^{0.108}_{-0.144}$	$0.950^{+0.108}_{-0.144}$

TABLE IV: SSKT predicted dilution scale factors corresponding to different data periods.

IV. UN-BINNED MAXIMUM LIKELIHOOD FIT

An unbinned maximum likelihood fit is performed to extract the parameters of interest, $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma$, plus additional parameters (referred to as "nuisance parameters") that include the signal fraction f_s, the mean B_s^0 width $\Gamma \equiv (\Gamma_L + \Gamma_H)/2$, the mixing frequency Δm_s, the magnitudes of the polarization amplitudes $|A_0|^2$, $|A_\parallel|^2$, and $|A_\perp|^2$, and the strong phases $\delta_\parallel \equiv \arg(A_\parallel^* A_0)$ and $\delta_\perp \equiv \arg(A_\perp^* A_0)$. The fit uses information on the reconstructed B_s^0 candidate mass m and its uncertainty σ_m, the B_s^0 candidate proper decay time t and its uncertainty σ_t, the transversity angles $\vec{\rho} = \{\cos\theta_T, \phi_T, \cos\psi_T\}$, and tag information $\mathcal{D}$ and ξ, where $\mathcal{D}$ is the event-specific dilution and $\xi = \{-1, 0, +1\}$ is the tag decision, in which $+1$ corresponds to a candidate tagged as B_s^0, -1 to a $\bar{B}_s^0$, and 0 to an untagged candidate. The single-event likelihood is described in terms of signal (P_s) and background (P_b) probability distribution functions (PDFs) as

$$\begin{aligned} & f_s P_s(m|\sigma_m) P_s(t, \vec{\rho}, \vec{\xi}|\vec{\mathcal{D}}, \sigma_t) P_s(\sigma_t) P_s(\mathcal{D}) \\ & +(1 - f_s) P_b(m) P_b(t|\sigma_t) P_b(\vec{\rho}) P_b(\sigma_t) P_b(\vec{\mathcal{D}}). \end{aligned} \quad (1)$$

The signal mass PDF $P_s(m|\sigma_m)$ is parameterized as a single Gaussian with a standard deviation determined separately for each candidate, while the background mass PDF, $P_b(m)$, is parameterized as a first order polynomial. The distributions of the decay time uncertainty and the event-specific dilution are observed to be different in signal and background, so we include their PDFs explicitly in the likelihood. The signal PDFs $P_s(\sigma_t)$ and $P_s(\vec{\mathcal{D}})$ are determined from sideband-subtracted data distributions, while the background PDFs $P_b(\sigma_t)$ and $P_b(\vec{\mathcal{D}})$ are determined from the $J/\psi\phi$ invariant mass sidebands. The PDFs of the decay time uncertainties, $P_s(\sigma_t)$ and $P_b(\sigma_t)$, are described with a sum of Gamma function distributions, while the dilution PDFs $P_s(\vec{\mathcal{D}})$ and $P_b(\vec{\mathcal{D}})$ are included as histograms that have been extracted from data.

The tagged un-binned maximum likelihood fit is nearly identical to that used previously [1], with the exception of the OST scale factors. As discussed in Section III, we now use separate scale factors for positive and negative tags. This slightly modifies the expressions for our decay time signal PDF $P_s(t, \vec{\rho}, \vec{\xi}|\vec{\mathcal{D}})$, which is expanded in terms of probability for B_s^0 ($P(t, \vec{\rho})$) and $\bar{B}_s^0$ ($\bar{P}(t, \vec{\rho})$),

$$\begin{aligned} P_s(t, \vec{\rho}, \vec{\xi}|\vec{\mathcal{D}}) = & \left(\frac{1 + \xi_1 S^+_{\mathcal{D},1} \mathcal{D}_1}{1 + |\xi_1|} \times \frac{1 + \xi_2 S_{\mathcal{D},2} \mathcal{D}_2}{1 + |\xi_s|} \right) P(t, \vec{\rho}) \\ & + \left(\frac{1 - \xi_1 S^-_{\mathcal{D},1} \mathcal{D}_1}{1 + |\xi_1|} \times \frac{1 - \xi_2 S_{\mathcal{D},2} \mathcal{D}_2}{1 + |\xi_2|} \right) \bar{P}(t, \vec{\rho}). \end{aligned} \quad (2)$$

where $\{\xi_1, \mathcal{D}_1, S^\pm_{\mathcal{D},1}\}$ correspond to the tag decsion, predicted dilution, and diltuion scale factors of the OST and $\{\xi_2, \mathcal{D}_2, S_{\mathcal{D},2}\}$ correspond to the tag decision, predicted dilution, and scale factor of the SSKT. The time and angular probabilities for B_s^0 can be expressed as

$$\begin{aligned} P(t, \vec{\rho}) \; \propto \; & |A_0|^2 \mathcal{T}_+ f_1(\vec{\rho}) + |A_\parallel|^2 \mathcal{T}_+ f_2(\vec{\rho}) \\ & + \; |A_\perp|^2 \mathcal{T}_- f_3(\vec{\rho}) + |A_\parallel||A_\perp| \mathcal{U}_+ f_4(\vec{\rho}) \\ & + \; |A_0||A_\parallel| \cos(\delta_\parallel) \mathcal{T}_+ f_5(\vec{\rho}) \\ & + \; |A_0||A_\perp| \mathcal{V}_+ f_6(\vec{\rho}). \end{aligned} \quad (3)$$

The functions $f_1(\vec{\rho}) \ldots f_6(\vec{\rho})$ are defined as

$$\begin{aligned}
f_1(\vec{\rho}) &= 2\cos^2\psi_T(1-\sin^2\theta_T\cos^2\phi_T) \\
f_2(\vec{\rho}) &= \sin^2\psi_T(1-\sin^2\theta_T\sin^2\phi_T) \\
f_3(\vec{\rho}) &= \sin^2\psi_T\sin^2\theta_T \\
f_4(\vec{\rho}) &= -\sin^2\psi_T\sin 2\theta_T\sin\phi_T \\
f_5(\vec{\rho}) &= \frac{1}{\sqrt{2}}\sin 2\psi_T\sin^2\theta_T\sin 2\phi_T \\
f_6(\vec{\rho}) &= \frac{1}{\sqrt{2}}\sin 2\psi_T\sin 2\theta_T\cos\phi_T,
\end{aligned}$$

The probability $\bar{P}$ for $\bar{B}^0_s$ is obtained by substituting $\mathcal{U}_+ \to \mathcal{U}_-$ and $\mathcal{V}_+ \to \mathcal{V}_-$. The time-dependent term $\mathcal{T}_\pm$ is defined as

$$\begin{aligned}
\mathcal{T}_\pm = e^{-\Gamma t} \times\ & [\cosh(\Delta\Gamma t/2) \mp \cos(2\beta_s)\sinh(\Delta\Gamma t/2) \\
& \mp \eta\sin(2\beta_s)\sin(\Delta m_s t)]\,,
\end{aligned}$$

where $\eta = +1$ for P and -1 for $\bar{P}$. The other time-dependent terms are defined as

$$\begin{aligned}
\mathcal{U}_\pm = \pm e^{-\Gamma t} \times\ & [\sin(\delta_\perp - \delta_\parallel)\cos(\Delta m_s t) \\
& - \cos(\delta_\perp - \delta_\parallel)\cos(2\beta_s)\sin(\Delta m_s t) \\
& \pm \cos(\delta_\perp - \delta_\parallel)\sin(2\beta_s)\sinh(\Delta\Gamma t/2)]\,, \\
\mathcal{V}_\pm = \pm e^{-\Gamma t} \times\ & [\sin(\delta_\perp)\cos(\Delta m_s t) \\
& - \cos(\delta_\perp)\cos(2\beta_s)\sin(\Delta m_s t) \\
& \pm \cos(\delta_\perp)\sin(2\beta_s)\sinh(\Delta\Gamma t/2)]\,.
\end{aligned}$$

As has been done previously [1], we model the decay time background PDF $P_b(t|\sigma_t)$ with a delta function at $t = 0$, one and two exponentials with negative slope for $t < 0$ and $t > 0$, respectively, all of which are convolved with the Gaussian resolution function. The background angular PDFs are factorized, $P_b(\vec{\rho}) = P_b(\cos\theta_T)P_b(\varphi_T)P_b(\cos\psi_T)$, and are obtained using B^0_s mass sidebands events.

The signal region ct fit projections for signal and background are shown in Fig. 8. Angular fit projections for signal and background are shown in Fig. 9.

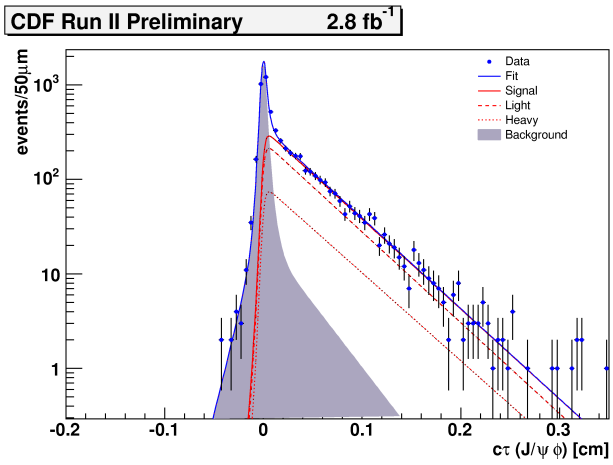

FIG. 8: Tagged B^0_s ct likelihood fit projections for the signal region.

To obtain the final result, the two-dimensional profile likelihood with all parameters floating is adjusted based on the distribution of p-values, defined as the ratio of the value of the likelihood in which $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma$ are fixed in the fit to particular values of $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma$ relative to the value of the likelihood in which $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma$ float freely in the fit, calculated for 10,000 default pseudo-experiments generated at the standard model point of $\beta_s^{J/\psi\phi} = 0.02$ and

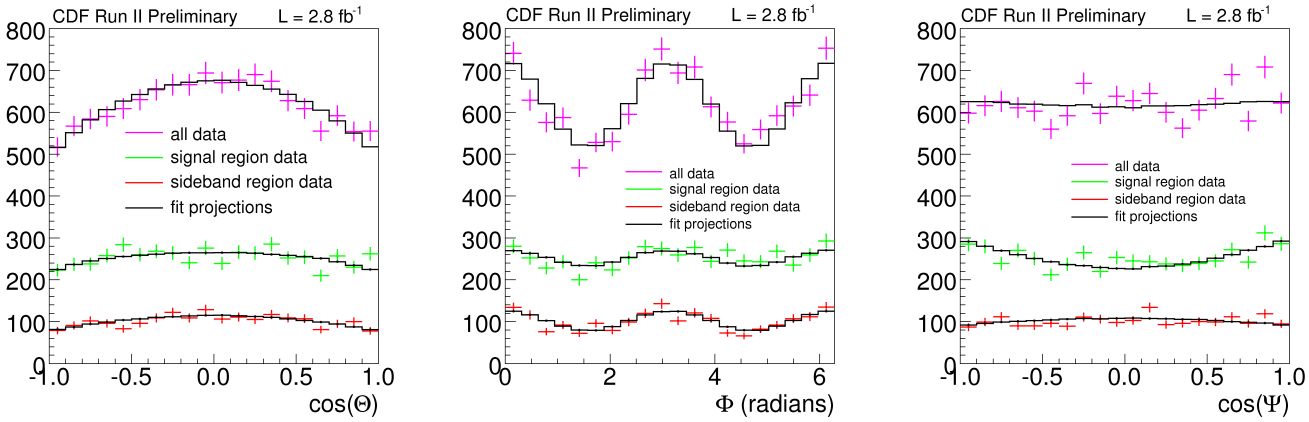

FIG. 9: Transversity angle projections in the signal region for $\cos\theta_T$ (left), ϕ_T (center), and $\cos\psi_T$ (right).

$\Delta\Gamma = 0.096$ ps^{-1} and 5,300 pseudo-experiments conducted in sixteen "alternate universes" in which the nominal values of all nuisance parameters have been varied randomly by $\pm 5\sigma$. We then choose the most conservative adjustment and p-value for the final result. This method of profile adjustment is used by HFAG in their combinations of results on $\beta_s^{J/\psi\phi}$ [3]. The adjustment curves for the profile likelihood are shown in Fig. 10. The final, adjusted two-dimensional confidence regions in $\beta_s^{J/\psi\phi} - \Delta\Gamma$ plane are seen in Fig. 11. We find that the p-value at the standard model expectation is 7%. The new result is compared with the previous, published result [1] in Fig. 12. Our expected sensitivity to $\beta_s^{J/\psi\phi} = 0.40$ for different amounts of integrated luminosity is shown in Fig. 13.

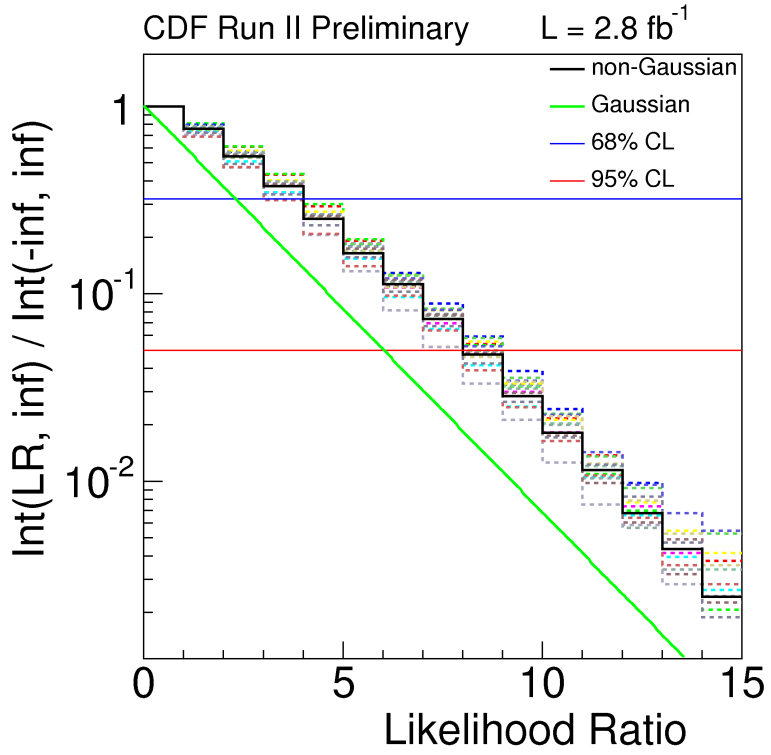

FIG. 10: Distribution of p-values for the two-dimensional profile likelihood with the p-values for all of the "alternate universes" (colored histograms) overlaid on that from the "default" universe (black histogram).

We have also determined the one-dimensional profile likelihood for $\beta_s^{J/\psi\phi}$. This result is also adjusted based on a one-dimensional mapping calculated from p-values at the standard model point $\beta_s^{J/\psi\phi} = 0.02$ using 10,000 default

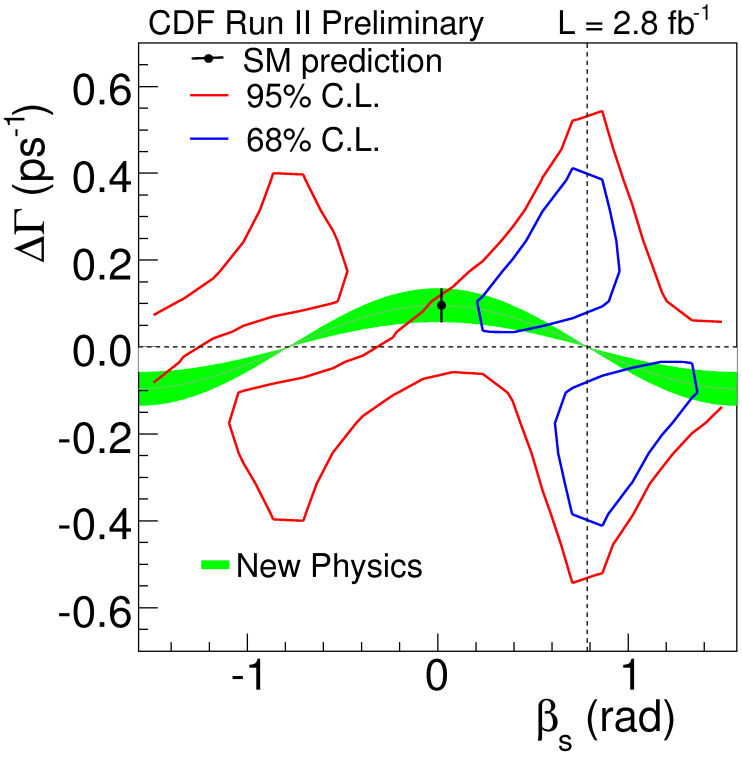

FIG. 11: Adjusted two-dimensional profile likelihood of $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma$ in 2.8 fb^{-1} of data. The standard model point is indicated by the black point with error bars. The p-value at the standard model point is 7%.

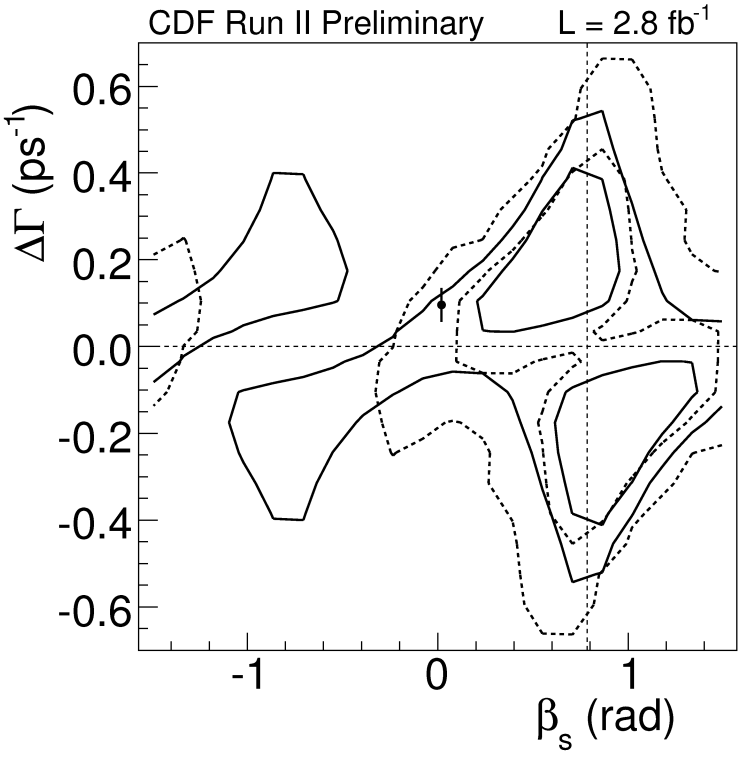

FIG. 12: Comparison between the new result (solid lines) and the previously published result (dashed lines). The standard model values is indicated by the black point with error bars.

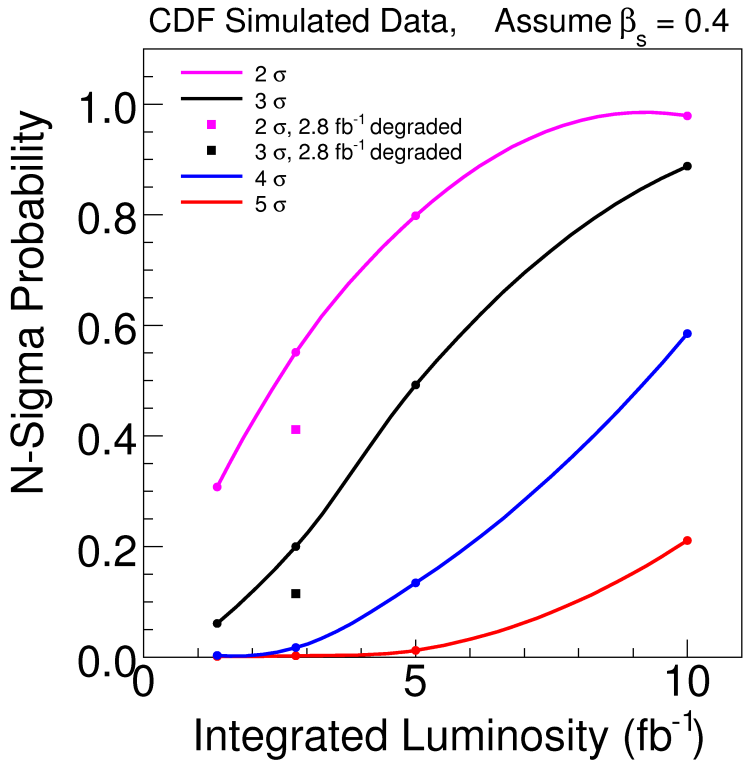

FIG. 13: Expected sensitivity to $\beta_s^{J/\psi\phi} = 0.40$ for 1.35 fb^{-1}, 2.8 fb^{-1}(present analysis and with expected analysis improvements), 5 fb^{-1}, and 10 fb^{-1} of integrated luminosity .

pseudo-experiments and 4,100 pseudo-experiments in each of sixteen "alternate universes". The adjustment curves for the one-dimensional profile likelihood are shown in Fig. 14 and the final adjusted profiles are shown in Fig. 15. The final one-dimensional range is $\beta_s^{J/\psi\phi} \in [0.28, 1.29]$ at the 68% confidence level. The p-value at the standard model point is again 7%.

V. LIKELIHOOD FIT WITH STANDARD MODEL CONSTRAINTS

We also measure the inverse of the mean decay width τ_s, the decay width difference $\Delta\Gamma$ and the transversity amplitudes $|A_0(0)|^2$ and $|A_\parallel(0)|^2$ (where $|A_\perp(0)|^2 \equiv 1 - |A_0(0)|^2 - |A_\parallel(0)|^2$) with the standard model constraint that $\beta_s^{J/\psi\phi} = 0.0$. Previously, this was done in the context of the measurement of $\Delta\Gamma$ and $\beta_s^{J/\psi\phi}$ without flavor tagging [17]. While we have considered using the tagged fit for with $\beta_s^{J/\psi\phi}$ fixed to 0.0, we find that biases persist in the fit which prevent us from making statements about the point estimates of these quantities. However, we have again studied point estimates for the fit to $B_s^0 \to J/\psi\,\phi$ which does not make use of flavor tagging and find that the pull distributions in this case are consistent with unit pulls.

We assign systematic uncertainties from studies in which we generate systematic variations of our default pseudo-experiments, which correspond to our default fit model, and then fit as we normally would. In each systematic variation we assign the systematic uncertainty due to a particular effect using the difference in the mean of the distribution of fit results from the systematic variation and the value with which the pseudo-experiment was generated. In all cases, the initial random seed used to generate pseudo-experiments is identical to minimize statistical fluctuations as much as possible.

The largest contributions to the systematic uncertainties arises from the knowledge of the resolution model. The modeling of the background lifetime is the most significant source of systematic uncertainties in the determination of both $\Delta\Gamma$ and the lifetime. The transversity amplitudes are primarily affected by the knowledge of the signal angular efficiencies. The systematic uncertainties assigned are listed in Table V.

- Alignment of SVX

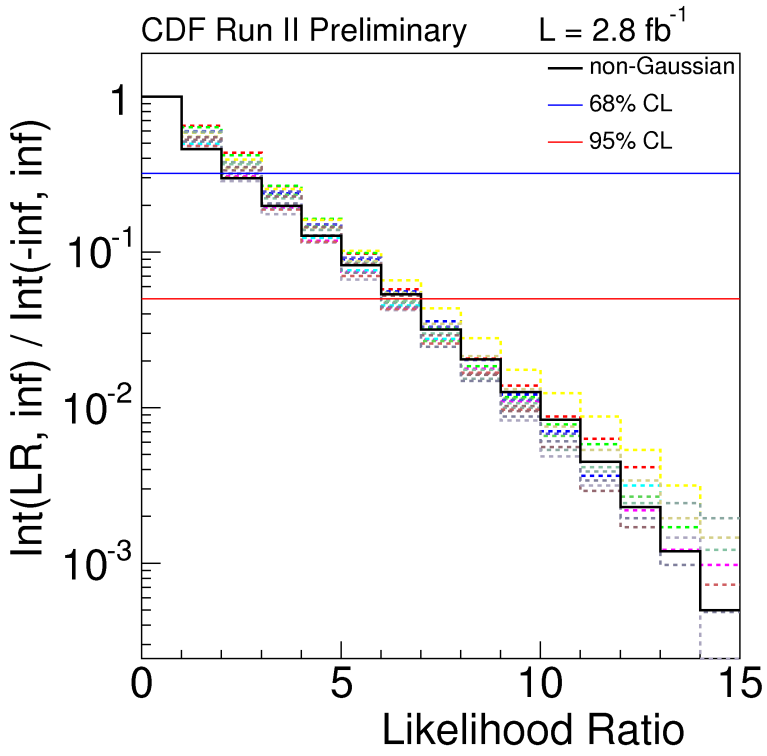

FIG. 14: Distribution of p-values for the one-dimensional profile likelihood with the p-values for all of the "alternate universes" (colored histograms) overlaid on that from the "default" universe (black histogram).

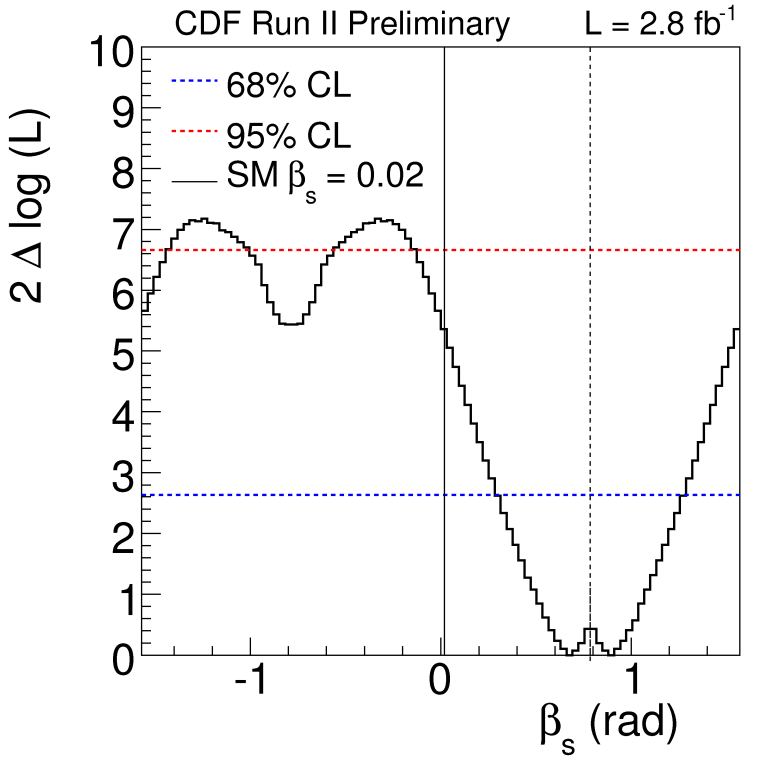

FIG. 15: Coverage-adjusted one-dimensional profile likelihood for $\beta_s^{J/\psi\phi}$ in 2.8 fb^{-1} of data. The p-value at the standard model point is 7%.

The alignment of the SVX affects the measurement of $\beta_s^{J/\psi\phi}$ by altering the length scale of the SVX detector. We follow the previous work [18] and assign a systematic of 2 μm to the lifetime and evaluate the effect on the rest of the parameters resulting from random 2 μm shift in the generated lifetime.

- Resolution scale factors

  We consider the systematic uncertainty due to alternate models of the resolution. In order to do so, we fit the data with a two Gaussian resolution function for signal and the prompt background, with a single Gaussian resolution function applied for all other background tails. Although both Gaussians are allowed to float in the fit, we observe that the fit prefers a single Gaussian for the signal and prompt background resolution and the fraction of the second Gaussian returned by a fit to data is consistent with zero. In order to assess a systematic effect, we generate with a two Gaussian resolution function in toy MC, where the fraction of the second Gaussian is 17% with an uncertainty in the fit to the data of 5%. The toy is generated with a second scale factor of 2.46, which is the value determined from the fit to data, while the first scale factor is 1.19. We then fit the toy MC with the default single Gaussian resolution model to determine the systematic uncertainty.

- Signal angular efficiency

  Systematics due to the modeling of the signal angular efficiencies are determined by generating toy MC from the 3D histograms we find in the realistic Monte Carlo from the previous analysis and fitting with our default model, in which the efficiencies are parameterized by spherical harmonics and Legendre polynomials and which have been updated for the new data.

- Background angular distributions

  We assign a systematic uncertainty for the background angular distributions by generating toy MC with histogram from sideband region of data and fitting with the parameterization we use in data.

- B^0 reconstructed as B_s^0

  We find that from a large realistic Monte Carlo sample of $B^0 \to J/\psi\, K^{*0}$ events generated according to phase space and reconstructed as $B_s^0 \to J/\psi\, \phi$, 4.8% survive the final selection requirements. We assign a systematic uncertainty due to neglecting these events using pseudo-experiments with angular amplitudes and strong phases generated using the values previously determined at CDF [19].

- Signal mass fit model

  We examine the effect of the mass fit model by generating toy MC with a double Gaussian model with two mass uncertainty scale factors. The parameters for generating the toy are determined by fitting data with a double Gaussian and two mass uncertainty scale factors. To determine the systematic effect, we then fit the toy with our default single Gaussian model.

- Background lifetime fit model

  Our background lifetime fit model is evaluated by generating toy MC according to the lifetime distribution from the sidebands. The default fit model is used to fit the toy and assign the systematic uncertainty.

Systematics	$\Delta\Gamma$ [ps^{-1}]	$c\tau_s$ [μm]	$\lvert A_\parallel(0)\rvert^2$	$\lvert A_0(0)\rvert^2$
Signal efficiency	0.003	0.8	0.007	0.007
Mass model	0.003	0.8	0.002	0.002
Resolution model	0.006	1.4	0.001	0.001
Background lifetime model	0.006	0.2	0.001	0.001
Background angular distribution	0.004	0.9	0.001	0.001
$B^0 \to J/\psi\, K^{*0}$ cross-feed	0.002	0.2	0.002	0.002
SVX alignment	0.003	2.0	0.000	0.002
Total	**0.011**	**3**	**0.007**	**0.008**

TABLE V: List of systematic uncertainties assigned.

Therefore, assuming that $\beta_s^{J/\psi\phi} = 0.0$, we find

$$\begin{aligned} c\tau_s &= 459 \pm 12\ (\text{stat.}) \pm 3\ (\text{syst.})\ \mu\text{m} \\ \Delta\Gamma &= 0.02 \pm 0.05\ (\text{stat.}) \pm 0.01\ (\text{syst.})\ \mu\text{m} \\ |A_\parallel(0)|^2 &= 0.241 \pm 0.019\ (\text{stat}) \pm 0.007\ (\text{syst}) \\ |A_0(0)|^2 &= 0.508 \pm 0.024\ (\text{stat}) \pm 0.008\ (\text{syst}) \end{aligned} \tag{4}$$

which are consistent with previous determinations of these quantities [17].

[1] T. Aaltonen et al. (CDF Collaboration), Phys. Rev. Lett. **100**, 161802 (2008).
[2] M. Bona et al. (UTfit Collaboration) (2008), 0803.0659.
[3] http://www.slac.stanford.edu/xorg/hfag/.
[4] I. Dunietz, R. Fleischer, and U. Nierste, Phys. Rev. D **63**, 114015 (2001).
[5] The Belle Collaboration, Nature, **452**, 332 (2008).
[6] M. E. Peskin, Nature **452**, 293 (2008).
[7] W.-S. Hou, M. Nagashima, and A. Soddu, Phys. Rev. D **76**, 016004 (2007).
[8] W.-S. Hou, M. Nagashima, and A. Soddu, Phys. Rev. Lett. **95**, 141601 (2005).
[9] J. Conway, D. Cox, R. Erbacher, A. Ivanov, W. Johnson, A. Lister, T. Schwarz, "Search for Heavy top t'-¿Wq in Lepton plus Jet Events in 2.3 fb-1", CDF note 9209.
[10] A. Abulencia et al. (CDF Collaboration), Phys. Rev. Lett. **97**, 242003 (2006), and references therein.
[11] A. Belloni, Ph.D. thesis, MIT, FERMILAB-THESIS-2007-36 (2007).
[12] G. Giurgiu, Ph.D. thesis, Carnegie Mellon Univ., FERMILAB-THESIS-2005-41 (2005).
[13] A. S. Dighe, I. Dunietz, and R. Fleischer, Eur. Phys. J. C **6**, 647 (1999).
[14] V. Tiwari, Ph.D. thesis, Carnegie Mellon Univ., FERMILAB-THESIS-2007-09 (2007).
[15] C. Lecci, Ph.D. thesis, Univ. of Karlsruhe, FERMILAB-THESIS-2005-89 (2005).
[16] Andreas Schmidt, "Combined B Flavour Tagging with Artificial Neural Networks", http://www-ekp.physik.uni-karlsruhe.de/pub/web/thesis/iekp-ka2006-11.pdf.
[17] T. Aaltonen et al. (CDF Collaboration), Phys. Rev. Lett. **100**, 121803 (2008).
[18] K. Anikeev, Ph.D. thesis, MIT, FERMILAB-THESIS-2004-12 (2004).
[19] http://www-cdf.fnal.gov/physics/new/bottom/070830.blessed-BdPsiKS.