CDF/PHYS/BOTTOM/CDFR/10722
Version 1.0

An Updated Measurement of the CP–Violating Phase $\beta_s^{J/\psi\phi}$ using $B_s^0 \to J/\psi\,\phi$ Decays in 9.6 fb^{-1} of Data

M. Dorigo, L.Grillo,A. M. Zanetti
INFN Trieste and University of Trieste, Trieste, Italy

M. Rescigno
INFN Roma, Roma, Italy

A.Di Canto, S. Leo, G. Punzi
INFN Pisa and University of Pisa, Pisa, Italy

D.Tonelli
FNAL, Batavia, USA

Abstract

We report an updated measurement of the CP-violating phase $\beta_s^{J/\psi\phi}$ using flavor-tagged $B_s^0 \to J/\psi\,\phi$ decays in 9.6 fb^{-1} of integrated luminosity collected by the dimuon trigger, which corresponds to the full CDF Run II dataset. Using an analysis technique and tools largely inherited from the previous (5.2 fb^{-1}) analysis, we reconstruct approximately 11 000 $B_s^0 \to J/\psi\,\phi$ signal events. The opposite side tagging algorithms are calibrated using using 81 000 $B^+ \to J/\psi K^+$ decays reconstructed in the same dataset. The same side tagging algorithms are not re-calibrated and thus used in only half of our dataset. The CP-violating phase is found to be in the range $\beta_s^{J/\psi\phi} \in [-\pi/2, -1.54] \bigcup [-0.03, 0.27] \bigcup [1.29, \pi/2]$ [STAT ONLY] at the 68% confidence level, in agreement with the standard model expectation. Assuming CP conservation ($\beta_s^{J/\psi\phi}$=0.0) we also determine the mean B_s^0 lifetime, $\tau_s = 1.528 \pm 0.019$ (stat) ps; the width difference between heavy and light mass eigenstates, $\Delta\Gamma_s = 0.071 \pm 0.026$ (stat) ps^{-1}; and the transversity amplitudes, $|A_0(0)|^2 = 0.514 \pm 0.011$ (stat), $|A_\parallel(0)|^2 = 0.230 \pm 0.013$ (stat). The results are amongst the most precise from a single experiment and consistent with previous determinations and world's average results.

Contents

1 Introduction **3**
1.1 Current experimental status 4

2 Analysis Strategy **5**

3 Polarization Amplitudes and Transversity Basis Definition **6**

4 Trigger and Dataset **7**

5 Data Selection and B_s^0 Mass Reconstruction **7**
5.1 Data Pre-Selection 8
5.2 Neural Network Selection 8
5.3 Mass Distributions 10

6 Tagging **12**
6.1 Opposite Side Tagging 13
6.1.1 Data and tools 13
6.1.2 Trigger and selection requirements 13
6.1.3 The calibration sample 14
6.1.4 Tagging performance 14
6.1.5 Opposite side tagging results 16

7 MC Reweighting and Angular Efficiency **24**

8 Likelihood Function **28**
8.1 B mass PDF 28
8.2 PDF of angles and ct for signal 28
8.3 Background lifetime PDF 30
8.4 Background angular PDF 31
8.5 Lifetime error PDFs 32
8.6 Flavor Tagging PDFs 33
8.7 Symmetry of the likelihood function 33
8.8 Summary of the fit variables 34

9 Fitter validation and checks **35**
9.1 Fit dependence input values and distributions 36
9.2 Pull studies 36
9.2.1 CP conserving fit 36
9.2.2 CP violating fit 39

10 B_s^0 lifetime, decay width difference and polarization amplitudes **45**

11 Tagged Results **46**
11.1 Tagged Contour 49
11.2 Tagged Intervals 49

12 Conclusions **58**

1 Introduction

Flavor physics of quarks is considered one of the most promising sectors in which to pursue indirect searches for non-standard model particles and their couplings. Indirect searches in quark flavor have motivated the design and operation of dedicated kaon and B-factory experiments in the past two decades. And, in the near-medium term, a new generation of experiments like LHCb, SuperB, Belle-II, will keep enriching the experimental information. In spite of remarkable experimental successes, no conclusive deviation from the SM has been observed. The CKM ansatz has survived the challenging scrutiny of many diverse and precise experimental tests, confirming itself as the leading source of flavor and CP violation at the scales probed thus far. However, a few intriguing, mild discrepancies have come and go in the last years, motivating an extended and deeper exploration. The B_s^0 dynamics specifically, seems a promising field of investigation: dedicated kaon experiment and B factories have provided very stringent constraints on the presence of NP in leading (and some subleading) processes involving charged and neutral kaons and bottom mesons. But a significantly smaller amount of experimental information, comparatively, is available for strange bottom mesons. This is the chief motivation for pursuing B_s^0 physics at hadron colliders and at CDF in particular.

In the B_s^0 system two flavor eigenstates are conventionally identified based on their valence quark content: $B_s^0 = |\bar{b}s\rangle$ and $\overline{B_s^0} = |b\bar{s}\rangle$. The time evolution of this binary system is approximately governed by the Schroedinger equation,

$$i\frac{d}{dt}\begin{pmatrix}|B_s^0(t)\rangle\\ |\overline{B}_s^0(t)\rangle\end{pmatrix} = \left[\begin{pmatrix}M_0 & M_{12}\\ M_{12}^* & M_0\end{pmatrix} - \frac{i}{2}\begin{pmatrix}\Gamma_0 & \Gamma_{12}\\ \Gamma_{12}^* & \Gamma_0\end{pmatrix}\right]\begin{pmatrix}|B_s^0(t)\rangle\\ |\overline{B}_s^0(t)\rangle\end{pmatrix} = [M - \frac{i}{2}\Gamma]\begin{pmatrix}|B_s^0(t)\rangle\\ |\overline{B}_s^0(t)\rangle\end{pmatrix} \quad (1)$$

where M is the mass matrix and Γ is the decay matrix. The eigenstates of the hamiltonian, admixtures of the flavor eigenstates, are observable particles with definite mass and lifetime:

$$|B_s^H\rangle = p|B_s^0\rangle - q|\overline{B}_s^0\rangle,\ |B_s^L\rangle = p|B_s^0\rangle + q|\overline{B}_s^0\rangle,\ \text{with}\ \frac{q}{p} = \frac{V_{tb}^* V_{ts}}{V_{tb} V_{ts}^*}. \quad (2)$$

The non-coincidence between mass and flavor eigenstates produces quantum-mechanical flavor-mixing. Mixing induces flavor oscillations between the B_s^0 and $\bar{B}_s^0$ states with a frequency proportional to the mass difference of the mass eigenstates, $\Delta m_s = m_H - m_L \approx 2|M_{12}|$. In the SM, particle-antiparticle oscillations are explained in terms of second-order weak processes (box diagram, see Fig. 1, left) involving virtual massive particles that provide a transition amplitude between the B_s^0 and $\overline{B}_s^0$ states. Non-SM particles can enter this amplitude. For example a 4th-generation up-type quark (t') could compete with the SM-dominant *top* contribution, modifying the mixing "intensity", that is, the oscillation frequency — and the phase. The 2006 measurement of the mass difference Δm_s by CDF [1] showed agreement with the SM within sizable theory uncertainties. This ruled out a broad class of SM extensions and represented a prime experimental achievement. However, it left completely unconstrained the phase of B_s^0 mixing, which could also exhibit NP contributions. We approximately define as mixing phase $\phi_s = arg(-M_{12}/\Gamma_{12})$ which, rigorously, is the phase difference between mixing amplitude (M_{12}) and the amplitude of B_s^0 and $\overline{B}_s^0$ decays into common final states (Γ_{12}). If non-SM particles are present in the mixing, their couplings in general will carry a non-trivial phase that will contribute to the M_{12} phase. It is considered significantly less likely that non-SM particles could contribute to the decay amplitudes, inducing a phase on Γ_{12}. The mixing phase also impacts the decay width difference between the two mass eigenstates $\Delta\Gamma_s = \Gamma_L - \Gamma_H \sim 2|\Gamma_{12}|\cos\phi_s$.

The study of the time evolution of $B_s^0 \to J/\psi\phi$ decays is widely recognized as the most effective experimental probe of the B_s^0 mixing phase. The $J/\psi\phi$ final state is common to B_s^0 and $\overline{B}_s^0$ decays, a necessary condition for mixing-induced CP violation to occur. The mixing phase becomes observable through the interference of two amplitudes, the amplitude of direct decay and the amplitude of decay preceded by mixing (Fig. 1). What is actually observable is the phase *difference* between decay and mixing, but since the decay is dominated by a single real amplitude, the difference equals the mixing phase. The fact that the decay is strongly dominated by a single, tree-level, real amplitude is what makes the extraction of the mixing phase from this process theoretically solid. The $B_s^0 \to J/\psi\phi$ decays offer several experimental advantages as well. The decay rate is at the per mil level, which makes the collection of large samples possible in hadron collisions. All final state particles are charged, thus easier to reconstruct in hadron collisions. The fully reconstructed final state provides a strong discrimination against the background processes. This is further enhanced by the presence of two narrow intermediate resonances whose masses can be used to impose constraints to reduce background.

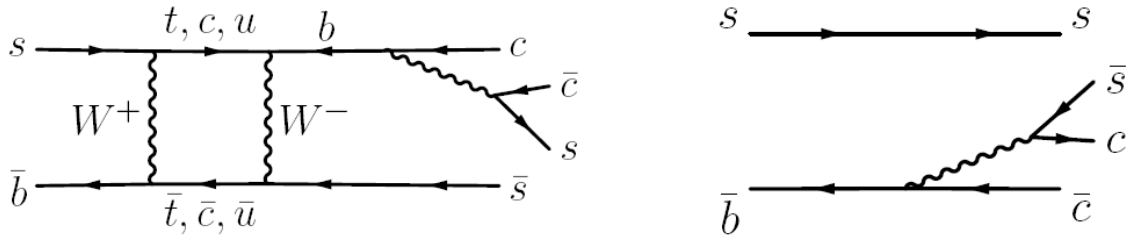


Figure 1: Feynman graph of the $B_s^0 \to J/\psi\phi$ decay with (left) and without (right) mixing.

1.1 Current experimental status

The first measurements of the CP-violating phase in $B_s^0 \to J/\psi\phi$ decays was finalized in 2008 by the CDF experiment [2]. It showed a mild, 1.5σ discrepancy from the SM. It was intriguing that the D0 experiment, few months later, found a similar, and consistent effect [3]. Indeed, the combination yielded a 2.2σ deviation from the SM [4]. This attracted some interest, further enhanced by the recent dimuon asymmetry results from the D0 collaboration [5],[6] which probe the same dynamics as $B_s^0 \to J/\psi\phi$, and report suggestive, even more significant anomalies. More recently, both the CDF and D0 collaborations updated their measurements of $B_s \to J/\psi\phi$ time-evolution. CDF used an event sample based on 5.2 fb^{-1} of integrated luminosity [7, 8], D0 on 8 fb^{-1} of integrated luminosity [9]. The results from both experiments, although consistent with the previous ones, showed an improved agreement with the SM. Also LHCb began recently to contribute, with a preliminary measurement on only 300 pb^{-1} of data [10], which appears already very competititve. Table 1 reports a summary of the current experimental status along with a comparison of key experimental parameters.

Parameter	LHCb (340 pb^{-1})	D0 (8 fb^{-1})	CDF (5.2 fb^{-1})
$\beta_s^{J/\psi\phi}$ [rad]	-0.07 ± 0.10	$0.28^{+0.18}_{-0.19}$	$\approx 0.27 \pm 0.25$
$\Delta\Gamma_s$ [ps^{-1}]	0.123 ± 0.031	$0.163^{+0.065}_{-0.064}$	0.075 ± 0.036
$\sigma_t(B_s^0)$ [fs]	≈ 50	≈ 100	≈ 90
$\sigma_m(B_s^0)$ [MeV/c^2]	≈ 7	≈ 30	≈ 10
Effective tagging power	$\approx 2.1\%$	$\approx 2\%$	$\approx 4.7\%$
Signal yield	8 300 ($t > 0.3$ ps)	5 600	6 500
S/B at peak	33/1 ($t > 0.3$ ps)	1/3	2/1

Table 1: *Summary of current experimental status and comparison of key experimental parameters. The D0 analysis uses an additional constraint from a theory assumption in the fit to the phase.*

2 Analysis Strategy

The measurement of the phase $\beta_s^{J/\psi\phi}$ relies on an analysis of the time-evolution of the $B_s^0 \to J/\psi\phi$ decay in which decays from mesons produced as B_s^0 or $\bar{B}_s^0$ are studied independently, and the CP-parity of the final state is statistically determined using angular distributions. The analysis can be dissected in four main steps:

- selection and reconstruction of the signal event sample;
- preparation of the analysis tools;
- fit to the time-evolution;
- statistical procedure to extract results and uncertainties.

For this update we follow the general analysis strategy used for earlier CDF publications. In particular the interesting physical parameters are extracted from an unbinned likelihood fit to the B_s^0 candidate mass, the angular variables in the transversity basis, the proper decay time, and flavor–tagging information as described in detail in Sec. 8. We adopt the same fitting code employed in the latest CDF measurement [7] with minimal simplifications and updated acceptance maps and other needed inputs (see Sec. 7). Signal contributions in the $B_s^0 \to J/\psi K^+K^-$ final state other than $B_s^0 \to J/\psi\phi$ signal itself are taken into account assuming an S–wave state for the KK system. Since the the K^+K^- mass m_{KK} is not used as a input to the fit, its contribution is integrated as described in [7], assuming a flat m_KK shape for the S–wave contribution and a relativistic Breit–Wigner shape with mass dependent width for the P–wave $\phi(1020)$ contribution.

As is well known the problem for $B_s^0 \to J/\psi\phi$ has several symmetries corresponding to ambiguities in the extracted physical parameters $\beta_s^{J/\psi\phi}$ and $\Delta\Gamma_s$, which are only marginally lifted by the S–wave-P–wave interference in our analysis. Half of the solutions are however eliminated using the difference in time evolution B_s^0 and $\bar{B}_s^0$. Flavor tagging, furthermore, improves the statistical behaviour of the likelihood for $B_s^0 \to J/\psi\phi$ decays in the presence of limited event samples. In the present analysis the Opposite-Side-Tagging algorithm has been recalibrated using data corresponding to the final dataset 6, however the available statistics of B_s^0 flavor specific decays from the Two Track Trigger trigger for the latest part of the data is limited, thus it was not possible to reliable calibrate the Same-Side-Kaon-Tagging algorithm [11, 1], and we employ this tagger only for the first part of the data, corresponding to 5.2fb^{-1} of integrated luminosity, where a reliable calibration is availble [12].

The complexity of the fit and the irreducible symmetries of the likelihood make the extraction of proper confidence intervals challenging from the simple fit results. A thorough work of simulation is needed to construct correct confidence regions and finally extract the results as discussed in 11.2.

3 Polarization Amplitudes and Transversity Basis Definition

Our analysis relies on the time evolution and on the kinematics of the $B_s^0 \to J/\psi\phi$ decay which is a decay of a Pseudoscalar meson (with $J^P = 0^-$), like the B, into two Vector mesons ($J \neq 0$). The $B_s^0 \to J/\psi\phi$ decay is of the $B \to VV$ type where a pseudo–scalar meson decays into two $J = 1$ mesons (Vector mesons) in the final state. In this case, there are three independent decay amplitudes, governing the probability that the B_s^0 meson decays in a state with one of the only three possible relative angular momenta: $L = 0, 1, 2$ of the final state vector mesons, in order to conserve the total initial angular momentum of the parent pseudo–scalar meson. The decay can be alternatively described in the linear polarization basis, i.e. throus the corresponding amplitudes $|A_0|$, $|A_\parallel|$, and $|A_\perp|$, and two relative strong phases of the amplitudes defined as: $\delta_\parallel = \arg(A_0 A_\parallel^*)$ and $\delta_\perp = \arg(A_0 A_\perp^*)$. This basis is particularly convenient since the $|A_0|$, $|A_\parallel|$ components lead to a CP–even final state, while $|A_\perp|$ is related to a CP–odd final state. Statistically seaparating through the angular analysis the two CP–even from the CP–odd component improves the sensitivity to the CP–violating phase, and gives also access to the so–called untagged observables arising from the interference between the two components with opposite CP parity and which survive even if the B_s^0 and $\bar{B}_s^0$ sample are summed together without flavor–tagging.

There are three angles that completely define the decay kinematics of the four particles in the final state. In our analysis we use the transversity basis, illustrated in Fig. 2, where the angles are

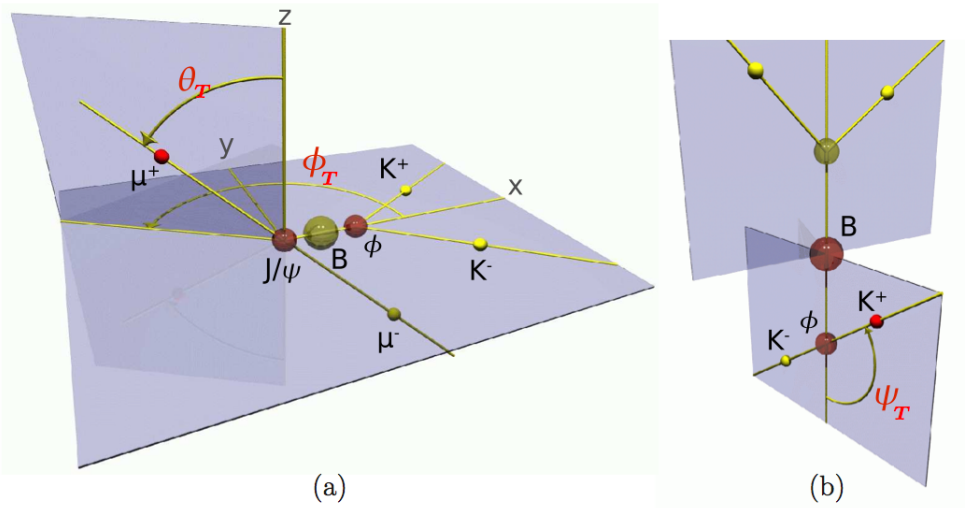


Figure 2: Transversity basis definition.

defined as follows. The first two angles are calculated in the rest frame of the J/ψ, and the third in the rest frame of the ϕ. In the rest frame of the J/ψ, the ϕ meson direction defines the x axis. The plane of K^+K^- defines the xy plane with $p_y(K^+) > 0$. From there:

- ϑ_T : in the J/ψ meson rest frame, the angle between $p(\mu^+)$ and the xy plane

- ϕ_T : in the J/ψ meson rest frame, the angle between the x axis and $p_{xy}(\mu^+)$
- ψ_T : in the ϕ meson rest frame, the angle between $p(K^+)$ and $p(J/\psi)$

For brevity and convenience sometimes the symbol $\vec{\omega} = \{\cos\vartheta_T, \phi_T, \cos\psi_T\}$ is used to refer to all the transversity angles and the transversity subscripts are sometimes dropped.

4 Trigger and Dataset

As in the previous measurements, we use data collected with the dimuon trigger [13, 14]. The analysis make use of the whole Run II dimuon sample, which corresponds to an integrated luminosity of about 9.6 fb^{-1} once good run list selection is applied. The latter is the version v45 of the B group run list without good-quality requirements for calorimeters and SVT.

Both the $B_s^0 \to J/\psi\,\phi$ and the $B \to J/\psi K$ decays are reconstructed using BStntuples of the J/ψ dataset [15]. The latter comprise the following datasets of the cdfpbnt catalog corresponding to different data acquisition periods (P): xpmmgd for P0; xpmmgh for P1-P4; xpmmhi for P5-P10; xpmmhj for P11-P13; xpmmij for P14; xpmmik for P15-P17; xpmmfp for P18-P38.

The previous analysis was performed with 5.2 fb^{-1} of data, which are P0-P25 of the above list. Difference of that dataset with respect to the same periods used in this update are the use of recently reprocessed (October 2011) Bstntuples for P18-P25 and the inclusion of runs with low silicon tracking efficiency of P18 (about 100 pb^{-1}). Events are required to explicitely fire JPSI triggers in order to be skimmed. We also thus modified the list of active JPSI trigger paths for runs more recent than P25 (for which the last trigger table is version v5 03). In our skimming (see next section) we then require in addition to those already kept in the previous analysis **??** the following trigger paths:

- JPSI-CMUP4-CMU-L2-RL100HZ-LUMI125, replacing JPSI-CMUP4-CMU-L2-DPS
- JPSI-CMUP4-CMX-L2-RL100HZ-LUMI125, replacing JPSI-CMUP4-CMX-L2-DPS
- JPSI-CMUP3-CMU-DPS, added
- JPSI-CMUP3-CMX-DPS , added

5 Data Selection and B_s^0 Mass Reconstruction

Selections of the signal events and rejection of background is implemented in a two step strategy: first the application of a set of cuts (that are called *rectangular cuts*) on different variables and then the use of a multivariate classifier implemented using an artificial Neural Network (NN). According to the first method, the value of a single variable determines if the event is kept in the data sample or not. The second method instead is able to take into account also the discriminating power of the single variables, and the correlation between the variables. The first step involves a set of loose pre-selection criteria, while in the second one a cut on the artificial Neural Network output select the final sample.

5.1 Data Pre-Selection

The first stage of event selection of the data sample used in this analysis, has been implemented making use of loose rectangular pre-selection cuts in order to reduce the sample size; afterwards the final selection is achieved by applying the NN. These cuts are defined as follows:

- Track quality: At least 10 axial and 10 stereo COT hits
  for kaon tracks + At least 3 Si hits for all tracks
- $(5.1 < M(B) < 5.6)$ GeV/c^2
- $Pt(k_1) > 0.4$ GeV/c and $Pt(k_2) > 0.4$ GeV/c
- $Pt(\phi) > 1$ GeV/c
- $(3.014 < M(J/\psi) < 3.174)$ GeV/c^2
- $(1.009 < M(\phi) < 1.028\ )$ GeV/c^2
- $\chi^2_{r\phi} < 50$
- $Pt(B) > 4$ GeV/c

The purpose of the pre-selection cuts is to eliminate most of the background events from the data sample, and at the same time to avoid rejecting signal events. In other words, one wants to improve the purity of the sample, while keeping the same efficiency. This goal is achieved by keeping cuts "loose", meaning that some contamination of background event is accepted in our sample. An additional advantage that is gained is that the obtained sample is of a significant smaller size, improving the computational speed of the subsequent stage.

5.2 Neural Network Selection

The artificial neural network used to make the final candidate selection has been constructed using the `NeuroBayes` package [16] in the context of the previous iteration of this analysis [7], and trained with simulated events as signal sample and mass sidebands data as background training sample. The training sample consisted of about $350k$ signal events and about $300k$ background events. The sidebands region is defined as in sec. 5.1. A NN combines the information from all the kinematic distributions into a single output variable, that denotes whether an event is signal-like or background-like. This output variable, O_{NN}, assumes values between -1 and 1, where events with O_{NN} close to -1 are classified as background and events with O_{NN} close to 1 as signal (see Fig. 5). A weight is assigned to each kinematic variable in input to the NN and it represents the magnitude of the variable contribution to the NN output. The weight associated to a certain variable is proportional to its discriminating power.

The following variables are used as input to the NN:

- $\chi^2_{r\phi}$ - the χ^2 of the two dimensional vertex fit in the transverse plane.
- $P(\chi^2, p)$ - χ^2 probability for the three dimensional vertex fit.
- $p_T(p)$ - Momentum component transverse to the beam direction for particle p.

- $LL_{\mu\mu}(p)$ - Value for a likelihood based quantity used for muon identification. [17]
- $LL_K(p)$ - Value of likelihood based discriminant for particle identification. It is constructed based on the dE/dx and TOF informations. [18]

They are listed in order of decreasing discriminating power and relevance to the final discriminant: the transverse momentum p_T of the ϕ meson, the kaon likelihood [18] based on TOF and dE/dx information, the muon likelihood [17] for the J/ψ muon daughters, $\chi^2_{r\phi}$ for the B_s decay vertex reconstruction, the transverse momentum p_T of the B_s meson, and the probabilities to reconstruct vertices from the B_s, ϕ, and J/ψ candidates. The muon and kaon likelihoods are quantities used for particle identification. The algorithm determining the muon likelihood is described in Ref. [17]. The kaon likelihood [18] is a combined discriminant constructed from the kaon track specific energy loss, dE/dx, and its time-of-flight information. Both likelihood variables have been calibrated on large data control samples.

Since the NN that we are using is the same of the previous iteration, we first of all need to understand whether the NN needs to be trained with background taken from the new dataset and signal with MC events properly rewighted or not. We achieved this purpose by making a comparison between the kinematic variables of the first $\sim 5.2 fb^{-1}$ of data and the remaining ones. The comparison between the kinematic variables of the first $\sim 5.2 fb^{-1}$ of data and the remaining ones, allows us to understand if and how much how much the data have changed over the time: e. g. the relative abundances of events given by a trigger path instead of another one can have changed during the time. This can happen for instance because of modifications to the trigger table (collection of all the trigger paths). Since each trigger path implements a set of requirements, often involving kinematic variables, the distribution of those variables can change depending on the used data sample. Furthermore the data comparison for different running periods can spot problems in the data sample sidebands subtracted distributions [1] have been compared at this stage and the data were required to satisfy the following set of rectangular cuts:

- $P_t(\mu_1) > 1.5 \;\; GeV/c$and $P_t(\mu_2) > 1.5 \;\; GeV/c$
- $P_t(k_1) > 0.6 \;\; GeV/c$ and $P_t(k_2) > 0.6 \;\; GeV/c$
- B vertex Prob > 0.001
- $P_t(B) > 5 \;\; GeV/c$

in substitution of the NN selection, since the NN performance is what we want to be probe. This comparison between the first $\sim 5.2 fb^{-1}$ of data and the remaining ones has been made for all the NN input variables, in order to have a first information on whether these variables have a similar behavior in the two periods, justifying the NN use without the need to be trained again, maybe with a different MC for the signal region (or the same but weighted in such a way to recover the kinematic variables distributions of the data sample) and with the new sidebands. These plots can be found in Fig. 3 and Fig. 4. Another important check is the comparison between the NN outputs (see Fig. 5). The comparison in terms of the NN output has been done for the sideband subtracted sample, and for the sidebands separately. The statistical test used to quantify the goodness of the agreement between the two dataset is the Kolmogorov test, in both for the kinematic variables and the NN output. The distributions are in good agreement, this implying that the NN does not need to be trained again.

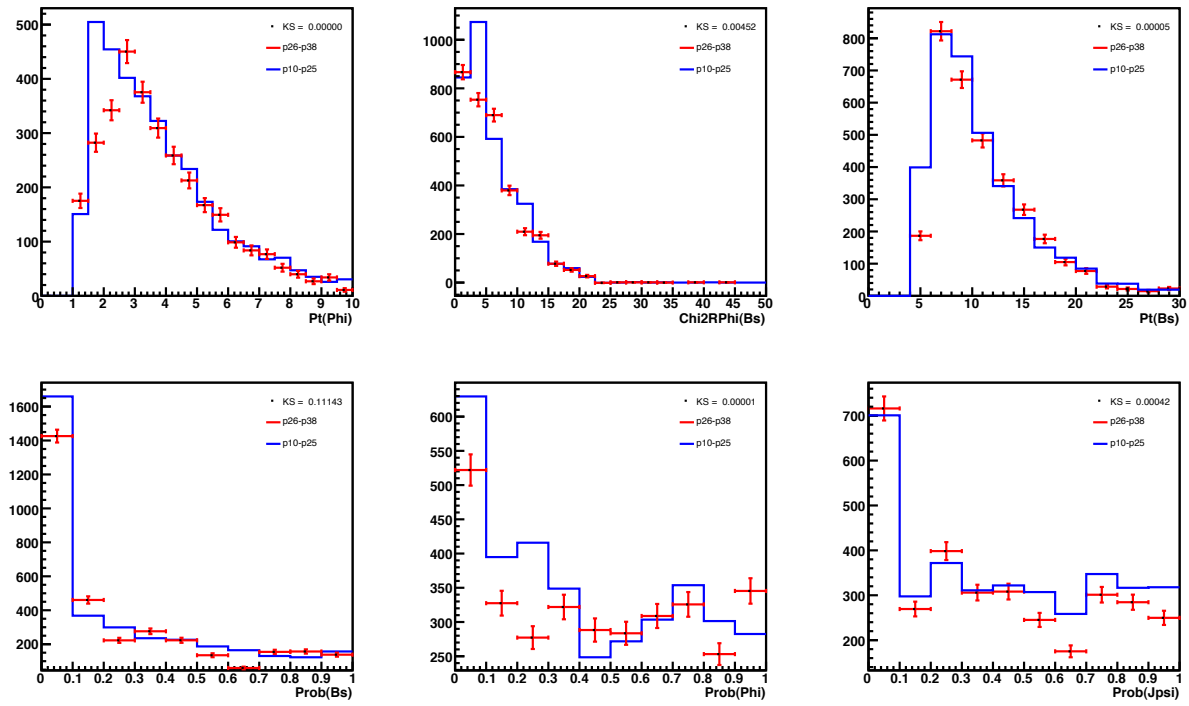


Figure 3: NN input variables (kinematic varaibles): comparison between the first $5.2fb^{-1}$ of data (BLUE line) and the full dataset (RED points). First raw from left to right $p_T(\phi)$, $\chi^2_{r\phi}$, $p_T(B_s)$, second raw $Prob(B_s), Prob(\phi), Prob(J/\psi)$.

Once the NN have been trained, the value for the cut on O_{NN} needs to be chosen. As in the previous analysis iteration the cut was optimized by maximizing the sensitivity to β_s in terms of its statistical error. This has been done by investigating the size of the statistical errors on β_s in different samples of pseudo experiments [7].

5.3 Mass Distributions

After the data selection described, the invariant mass distribution $m_{\mu^+\mu^-K^+K^-}$ for the $B_s \to J/\psi\phi$ is obtained (see Fig. 6). The final data sample obtained has mainly three components:

Signal: the final yield obtained for $B_s \to J/\psi\phi$ decay events is of 10953 ± 111 events. This value is obtained by a binned likelihood fit on the mass histogram. The function used to model the signal is a single Gaussian.

Combinatorial background: these events are random combinations of charged tracks that satisfy accidentally the selection requirements, as well as events with real J/ψ reconstructed together with two random charged tracks. They produce a continuous invariant mass distribution and a smooth slowing decreasing distribution in the signal region is expected. It is the main source of background in our analysis.

[1] see sec: 5.1

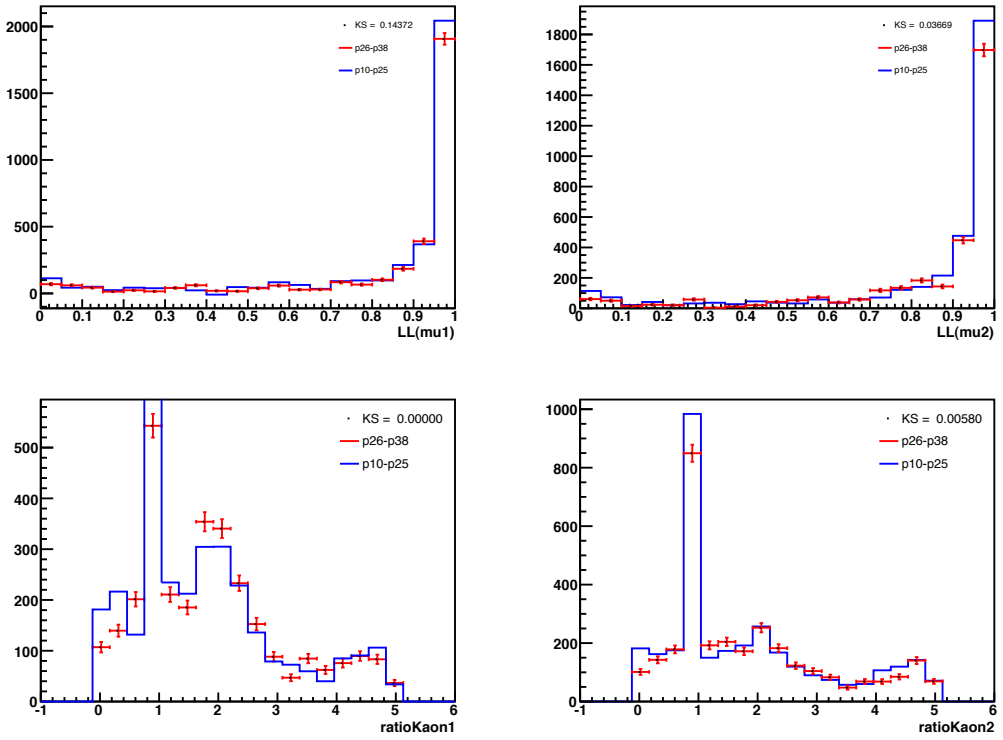


Figure 4: NN input variables (identification varibles): comparison between the first $5.2fb^{-1}$ of data (BLUE line) and the full dataset (RED points). First raw from left to right $LL_\mu(\mu_1), LL_\mu(\mu_2)$, second raw $LL_K(K_1), LL_K(K_2)$.

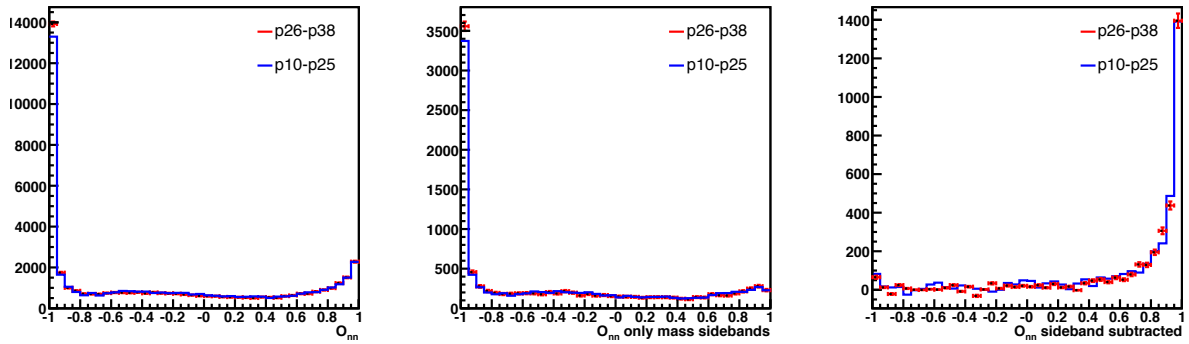


Figure 5: NN output variable: comparison between the first $5.2fb^{-1}$ of data (BLUE line) and the full dataset (RED points).Top left the B_s mass distribution with a fit of mass overlayed. Top right the NN output variable (O_{NN}) for both signal (event with $O_{NN} \sim 1$) and background events ($O_{NN} \sim 1$). Bottom left: O_{NN} for the sidebands events. Bottom right: t O_{NN} signal events only (sideband subtracted).

Physics background: in our data sample there can be some contamination from $B_d \to J/\psi K^* \to [\mu^+\mu^-][K^\pm\pi^\mp]$ decay events mis- reconstructed as $B_s \to J/\psi\phi$ decays (defined as B^0 cross-feed); it occurs when in the reconstruction the daughter tracks of the K^* are assumed to be two kaons and an incorrect invariant mass is computed. In this analysis there is a systematic

error that accounts for this affect (see sec. **??**). The fraction f of B^d cross-feed events in the B_s sample have been calculated to be (1.6±0.6) into the signal sample of the previous analysis iteration. To estimate this fraction, production fractions of the B_s and B_d mesons need to be known as their relative decay rates to $J/\psi\phi$ and $J/\psi K^*$, respectively, and the efficiency of each type of event passing the final selection criteria established under the $B_s \to J/\psi\phi$ hypothesis. Both the production fractions and the branching fractions are taken from Ref. [19]. The efficiencies can be estimated using simulation, with both $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$ modes reconstructed as $B_s \to J/\psi\phi$ decay. Eventually the fraction of B_d cross-feed is obtained as:

$$f(B_d \; in \; B_s \; sample) = \frac{f(\bar{b} \to B_d)\mathcal{B}(B_d \to J/\psi K^*)\varepsilon(B_d)}{f(\bar{b} \to B_s)\mathcal{B}(B_s \to J/\psi\phi)\varepsilon(B_s)}. \quad (3)$$

Another additional contributions from S-wave K^+K^- under the ϕ peak in $B_s \to J/\psi\phi$ decay can contribute up to few percents of the total rate. a normalized probability density for the decay $B_s \to J/\psi K^+K^-$ (kaons in an S-wave state) has been added to the likelihood function. These kaons can either be a non resonant pair of kaons, or the decay products of a scaler particle, the $f_0(980)$. In that case the final state of $B_s \to J/\psi f_0(980)$ can be only in S- wave, since the decay is a $P \to VS$ decay. We account for these two contributions by adding a normalized probability density to the likelihood in the full fit determining β_s, $\Delta\Gamma$ and all the parameters that we want to measure (see sec. **??**).

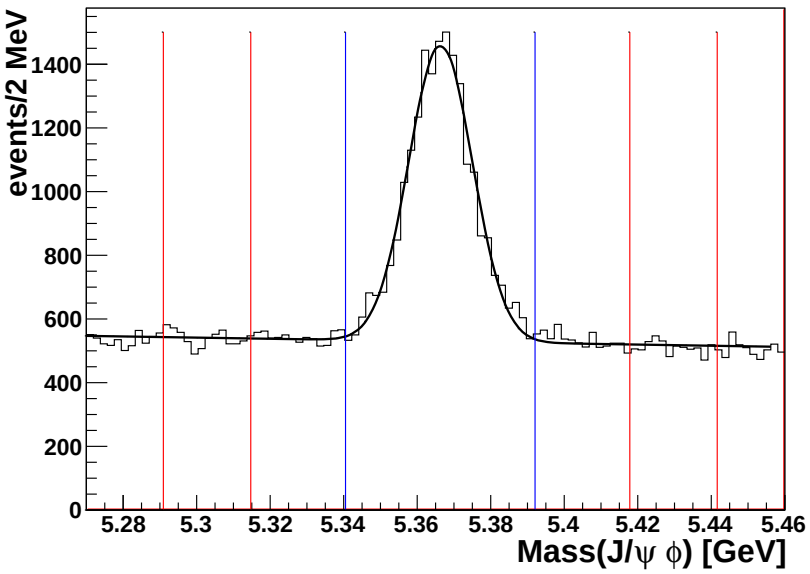


Figure 6: Mass distribution of our final data sample: the events satisfying preselection requirements + NN. BLUE lines show the signal region, and RED lines the sidebands.

6 Tagging

Flavor tagging algorithms are of crucial importance in enhancing the sensitivity of B_s^0 mixing analyses. They exploit the properties of b quark hadronization and decay to infer the flavor content

(b quark or $\bar{b}$ quark) of the signal bottom meson at production. Two classes of flavur tagging algorithms are used at CDF, Same Side and Opposite Side Taggers (SST,OST). The SST algorithms deduce the production flavor of the signal bottom meson by exploiting charge-flavor correlations of the neighboring particles produced during its fragmentation. The OST algorithms exploit pair $p\bar{p} \to b\bar{b}$ production using information from the "other" bottom hadron in the event (the "opposite side" B). The flavor-specific semi-leptonic decay may be used to identify the flavor of the opposite-side B hadron at decay time. Similarly, the net charge of the jet identified as coming from the opposite-side bottom hadron fragmentation and decay can provide information on its quark content. From this information the flavor of the candidate B at production can be inferred, assuming that the opposite-side hadron was not a neutral meson that undergo oscillation.

In the following two subsections we detail the calibration and use of OST and SST algorithms in this analysis.

6.1 Opposite Side Tagging

Different OST algorithms have been developed in CDF, using semimuonic [20] and semielectronic [21] decays, and jet [22] or opposite side kaon charge [23]. A NN combination of them has been validated and used in the previous measurement of the B_s^0 mixing phase. In this update, we use the same tool, after updating the extraction of the scale factors to the full run II data sample. This is achieved by applying the OST algorithms to a large sample of fully reconstructed $B^+ \to J/\psi K^+$ decays (charge conjugates implied everywhere). The tagging decision and associated dilution of the algorithm is then compared to the actual b quark content of the meson at decay time, which is known from the charge of the kaon. This is also the flavor at production since charged B do not oscillate.

6.1.1 Data and tools

We reconstruct exclusive $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ decays in the same dimuon trigger data used for the B_s^0 mixing phase analysis. These are skimmed from Bstntuples datasets xpmmgd, xpmmgh, xpmmhi, xpmmhj, xpmmij, xpmmik, and xpmmfp collected throughout Run II and corresponding to an integrated luminosity of $\mathcal{L} \approx 10$ fb^{-1} (trigger prescaling not included). The goodrun list v44 is used. The selection is applied in two stages. After event reconstruction the events are subject to a loose preselection aimed at speeding up downstream processing. Then the preselected data are fed into an artificial Neural Network [16] to achieve improved background suppression. The following subsections describe the event reconstruction and selection processes used to obtain the analysis sample.

6.1.2 Trigger and selection requirements

The dimuon sample is selected based on trigger requirements at Level 1 and Level 3. To pass Level 1, the event must have two muon stubs [24] either both in the CMU, or one in the CMU and one in the CMX. The muon stubs have to be matched to an XFT track of $p_T > 1.5$ (CMU) or $p_T > 2.0$ (CMX). The Level 2 and 3 triggers further impose that the pair of muons must have opposite charge. Additionally a maximum z_0 separation of $|z_0(1) - z_0(2)| < 5$ cm is required. The dimuon mass is selected to be in the range $2.7 < m_{\mu\mu} < 4$ GeV/c^2. Some loose selection requirements are applied to reduce the sample size:

- $5.16 < M(J/\psi K^+) < 5.40$ GeV/c^2.
- ≥ 3 axial hits per track in the silicon detector for muons.
- $\geq$ axial hits per track in the silicon detector for the kaon.
- successful XFT-muons match.
- $0.0 \leq ct \leq 0.1$ μm .

where $ct = \frac{ML_{xy}}{P_T}$, M being the known B^0 meson mass and L_{xy} the transverse decay length of the B candidate projected onto its reconstructed transverse momentum.
To improve signal-to-background separation, we reused the NN training and optimization discussed in [25, 26], selecting events with a NN output ≥ 0.8 which had been shown to maximizes the $S/\sqrt{S+B}$ figure. In addition to the NN requirement, we impose also a threshold on the decay length at 60 μm. This rejects a large part of the combinatorial background while preserving about 85% of the signal. As it will be shown in the following, the scale factors obtained with thsi additional requirement exhibit an increased consistency between B^+ and B^- sample, allowing the use of a single, common scale factor.

6.1.3 The calibration sample

The resulting $B^+ \to J/\psi K^+$ sample is shown in fig. 7. A simple gaussian fit over a linear background finds approximately 40 000 B^- decays and 41 000 B^+ decays. The signal yield is consistent with what we expected from the previous iteration. Central mass values and widths are consistent as well.

6.1.4 Tagging performance

The outcome of the tagging algorithm at CDF is a tag decision, ξ, an integer variable that can assume values ± 1 or 0. A value of $\xi = -1$ ($\xi = 1$) implies that the B meson was tagged as being a B_s^0 ($\overline{B_s^0}$) at production. The necessary information may not be available in every event to make a flavor decision. When the tagger is unable to reach a decision the value is $\xi = 0$. Two empirical parameters characterize the performance of flavor tagging:

$$\varepsilon = \frac{N_{tagged}}{N_{untagged} + N_{tagged}} \qquad \text{and} \qquad D = \frac{N_R - N_W}{N_R + N_W}. \tag{4}$$

The fraction of events for which a decision is made is called the tagging efficiency, ε. The dilution D, where N_R (N_W) is the number of correctly (wrongly) tagged events, quantifies the mistagging rate. Kaon, pion, muon or electron misidentification, or use of tagging information that is unrelated to the $b\bar{b}$ vertex can lead to a mistag. The product εD^2, where D is the dilution averaged over the studied sample, characterize the overall performance of flavor tagging algorithms providing an idea of the effective reduction in signal sample size when a correct tagging decision is needed. To use tagging information in our fit we need to characterize the tagging algorithms performance. For each decay we compare the true flavor (as indicated by the kaon charge) with the flavor identified by the OST algorithm. Indeed, for each event the tagging algorithm provides a tag (b or $\bar{b}$) and a predicted dilution that quantifies the reliability of such tag. The algorithms were designed and developed at the beginning of Run II using semileptonic B decays, where the charge of the lepton

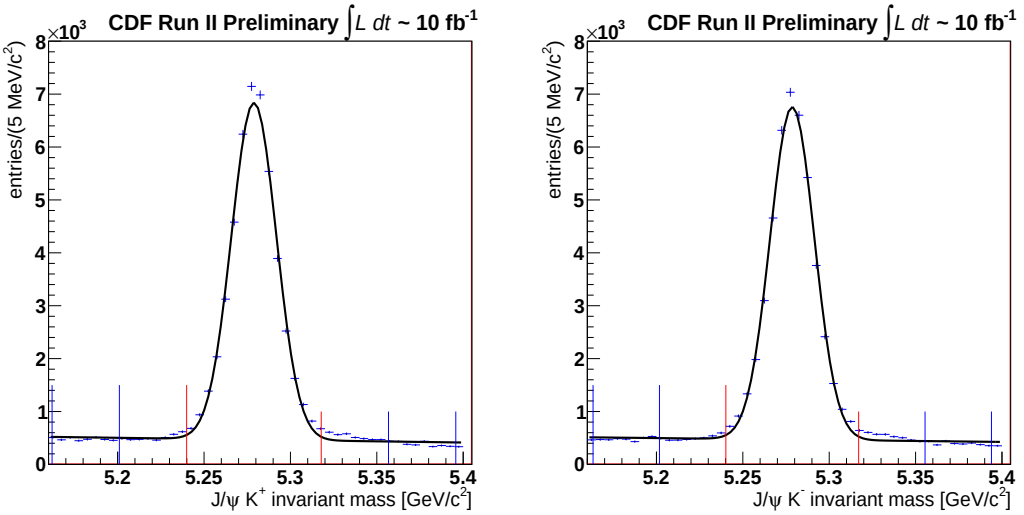


Figure 7: *$J/\psi K^+$ Mass distribution (left) and $J/\psi K^+$ Mass distribution (right) for the full Run II data sample*

identified the flavor of the parent B hadron. However the lepton could come from a mixed B meson or from a sequential $b \to c \to \ell$ decay. These, and other subtler effect yield mistags. Hence, the performance of the algorithm in the current B^+ sample could slightly deviate from the performance as predicted in semileptonic decays. Such deviations are modeled through a "scale factor" correction (S), which should be extracted from data. This is done by comparing the known dilution (since we know the B^+ flavor exactly) with the dilution predicted by the algorithm. We divide the sample in independent subsamples according to their predicted dilution. For each bin of predicted dilution we count the number of right (wrong) tags to extract the actual dilution. Then we graph the actual dilution as a function of the predicted one (Fig. 10) to determine the scale factor. All dilutions distributions are background subtracted using mass sidebands. The scale factor is determined as the slope of the straight line fits of figures Fig. 10. For the entire dataset, we use two scale factors for the opposite side tagger, one for the B^+ and one for the B^-, in order to allow for any asymmetry in the tagging algorithms. We find $S_D^+ = 1.09 \pm 0.05$ and $S_D^- = 1.08 \pm 0.05$ respectively with a total average dilution of $\overline{D} = 6.88 \pm 0.03\%$. Since the calibrated values of the scale factor for the B^+ and B^- are approximately equal we use the average of the two in the fit and the spread as a systematic uncertainty. As a check, we determine separately the scale factors for period up to 25 (Fig. 11) and period 26 to 38 (Fig.**??**) to identify any major drift in performance. Table 6.1.4 shows the OST dilution scale factors in different parts of the data

We determine OST tagging efficiency and dilution for different periods of data and summarize the results in table 11. As an additional cross-check, we determine scale factors and efficiencies in periods of data with approximately similar statistics (~ 1.7 fb^{-1} each), to ensure stability and

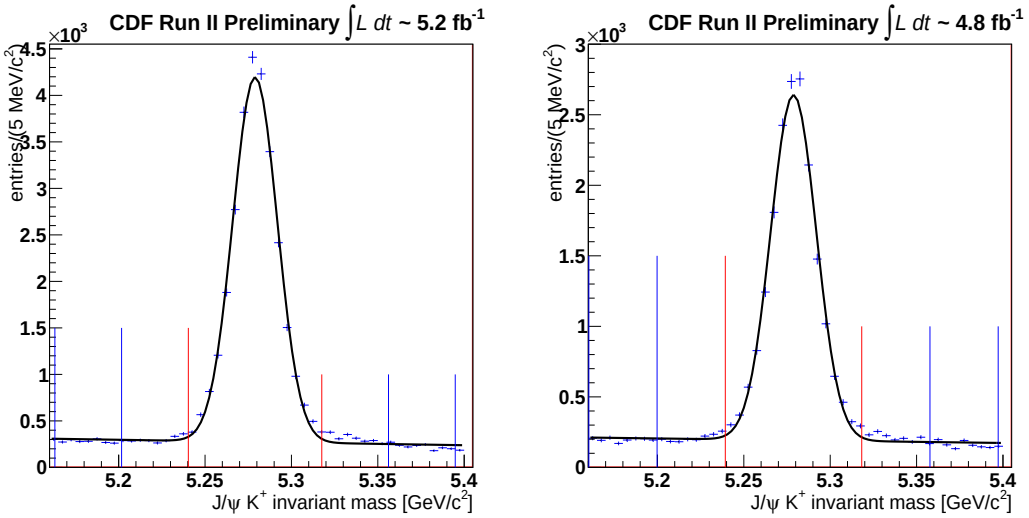


Figure 8: *Mass distributions of $J/\psi K^+$ for data collected up to period 25 (left) and for data collected in P26-38 (right).*

Scale Factor	PRD (0–5.2 fb^{-1})	0–5.2 fb^{-1}	5.2–10 fb^{-1}	0–10 fb^{-1}
$S_D{}^+$	0.93 ± 0.09	1.09 ± 0.06	1.08 ± 0.08	1.09 ± 0.05
$S_D{}^-$	1.12 ± 0.10	1.06 ± 0.07	1.10 ± 0.08	1.08 ± 0.05
ε	$94.3 \pm 0.3\%$	$93.8 \pm 0.1\%$	$91.2 \pm 0.2\%$	$92.8 \pm 0.1\%$
$\overline{D}$	$6.9 \pm 0.1\%$	$6.84 \pm 0.04\%$	$6.93 \pm 0.05\%$	$6.88 \pm 0.03\%$
$\sqrt{\overline{D^2}}$	$11.5 \pm 0.02\%$	$11.26 \pm 0.08\%$	$11.36 \pm 0.10\%$	$11.30 \pm 0.06\%$
εD^2	1.2%	$1.19 \pm 0.02\%$	$1.18 \pm 0.02\%$	$1.18 \pm 0.01\%$

Table 2: *OST performance for B^+ and B^- in different parts of the data, compared with the analysis submitted to PRD.*

consistence throughout all parts of the data. Fig. 13 shows as the scale factors are stable through data but a sensible degradation in efficiency is observed.

6.1.5 Opposite side tagging results

The measured scale factor in ~ 10 fb^{-1} of data is $S_D = 1.09 \pm XXX$, with a Tagging efficiency of $92.8 \pm 0.1\%$ and a mean predicted dilution of $6.88 \pm 0.03\%$. The total effective tagging power is $\varepsilon D^2 = 1.30 \pm 0.05\%$.

subsectionSame Side Tagging

The SST used in this analysis was originally developed for the CDF mixing measurement [11, 1].

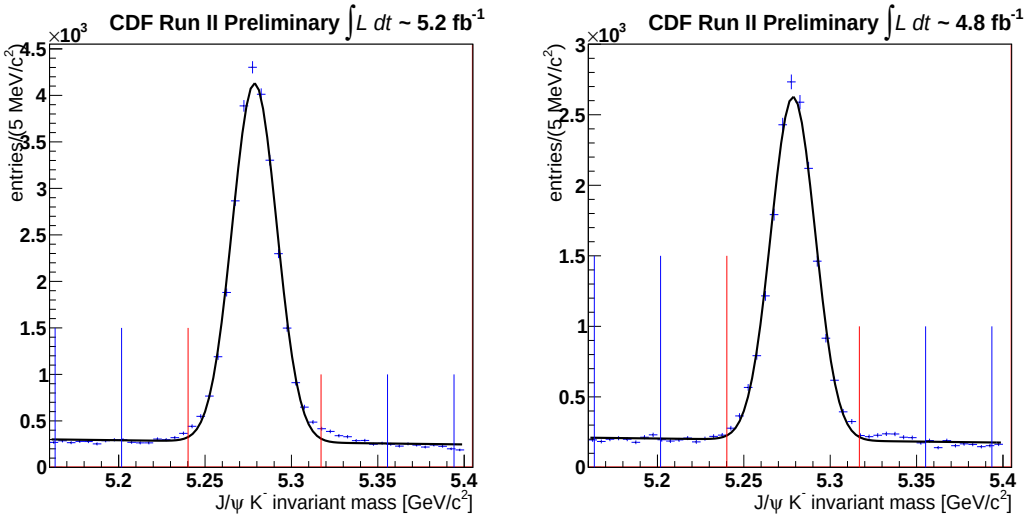


Figure 9: *Mass distributions of $J/\psi K^-$ for data collected up to period 25 (left) and for data collected in P26-38 (right).*

It is calibrated on 5.2 fb^{-1} of data by repeating the measurement of the B_s^0 mixing frequency and extracting the dilution. The calibration uses four channels collected using the displaced vertex trigger:

- $B_s^0 \to D_s^- \pi^+, D_s^- \to \phi\pi^-, \phi \to K^+K^-$;
- $B_s^0 \to D_s^- \pi^+, D_s^- \to K^*K^-, K^* \to K^+\pi^-$;
- $B_s^0 \to D_s^- \pi^+, D_s^- \to \pi^+\pi^+\pi^-$;
- $B_s^0 \to D_s^- \pi^+\pi^+\pi^-, D_s^- \to \phi\pi^-, \phi \to K^+K^-$.

where about half of the signal events come from the first channel. We have not yet extended the calibration of the SST to the full 10 fb^{-1} data sample. This is in part due to the marginal increase in calibration sample statistics with respect to the 5.2 fb^{-1} analysis. Figure 16 compares the mass distribution of the $B_s^0 \to D_s^- \pi^+, D_s^- \to \phi\pi^-, \phi \to K^+K^-$ signal for samples corresponding to 5.2 fb^{-1} and 7.2 fb^{-1}. The 10% increase in sample statistics against a 40% increase in nominal integrated luminosity shows that the displaced-tracks trigger, which select the above hadronic decays, is severely suppressed in the latest part of the sample. For the time being, we use the calibration (tab. **??**) obtained with 5.2 fb^{-1} of data [12] and, accordingly, use the information from this tagger only in the first half of our sample. This conservative choice prevents potential problems arising from drift of the tagger performances as a function of time. We are still evaluating if a re-calibration

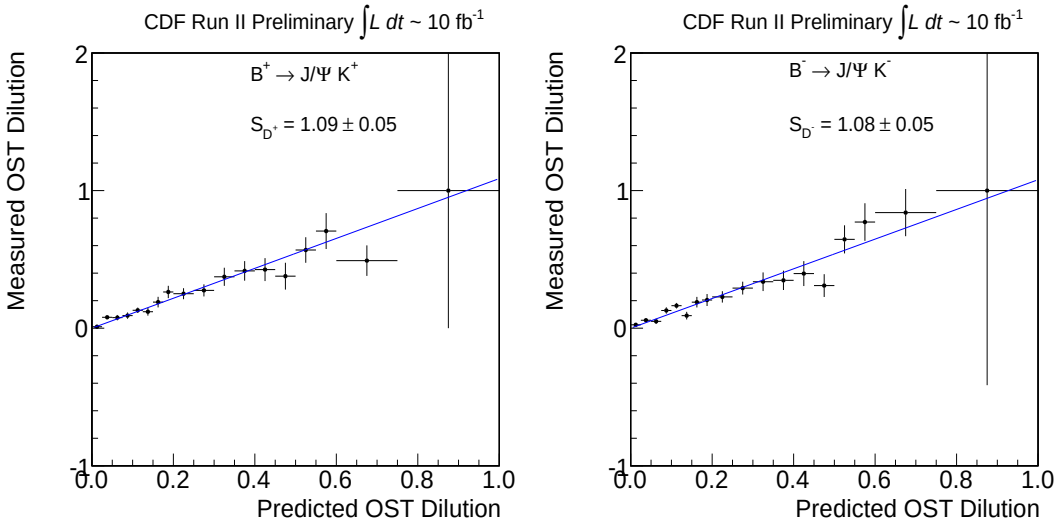


Figure 10: *Measured dilution as a function of predicted dilution for B^+ (left) and B^- (right) for all data through period 38.*

of the SST on the entire sample is worth. Monte Carlo studies show an average 12% worsening of the expected $\beta_s^{J/\psi\phi}$ uncertainty if the SST is not used in the second half of the data.

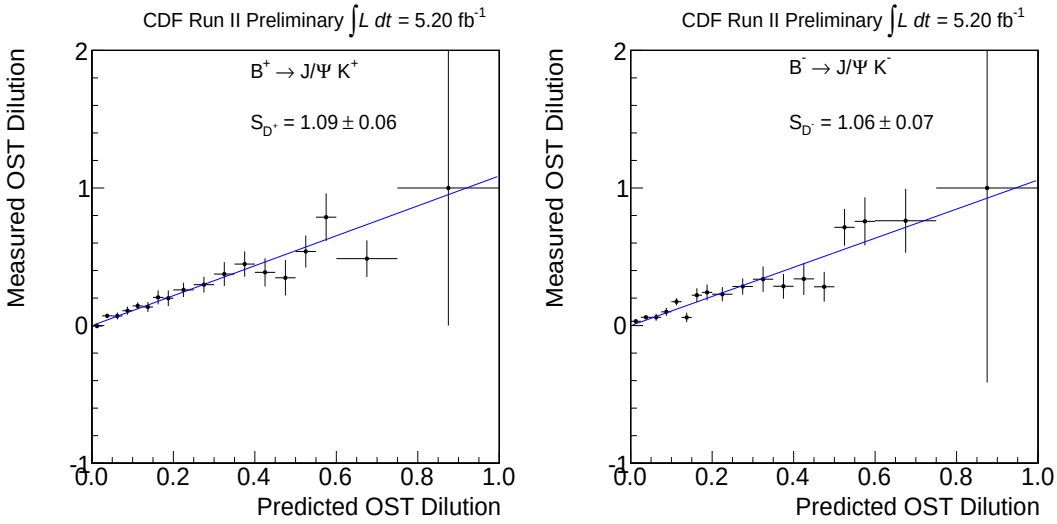


Figure 11: *Measured dilution as a function of predicted dilution for B^+ (left) and B^- (right) for all data through period 25.*

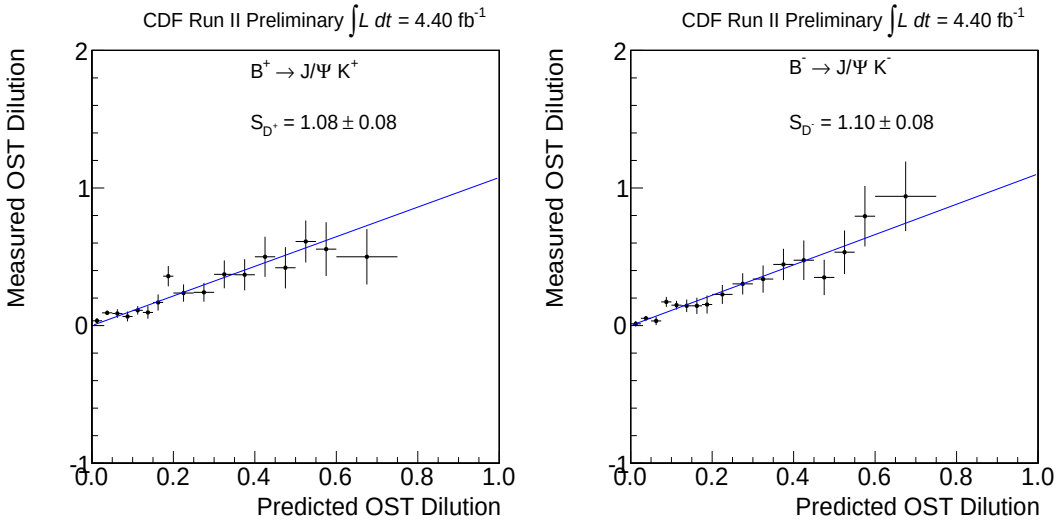


Figure 12: *Measured dilution as a function of predicted dilution for B^+ (left) and B^- (right) for data between period 26 and period 38.*

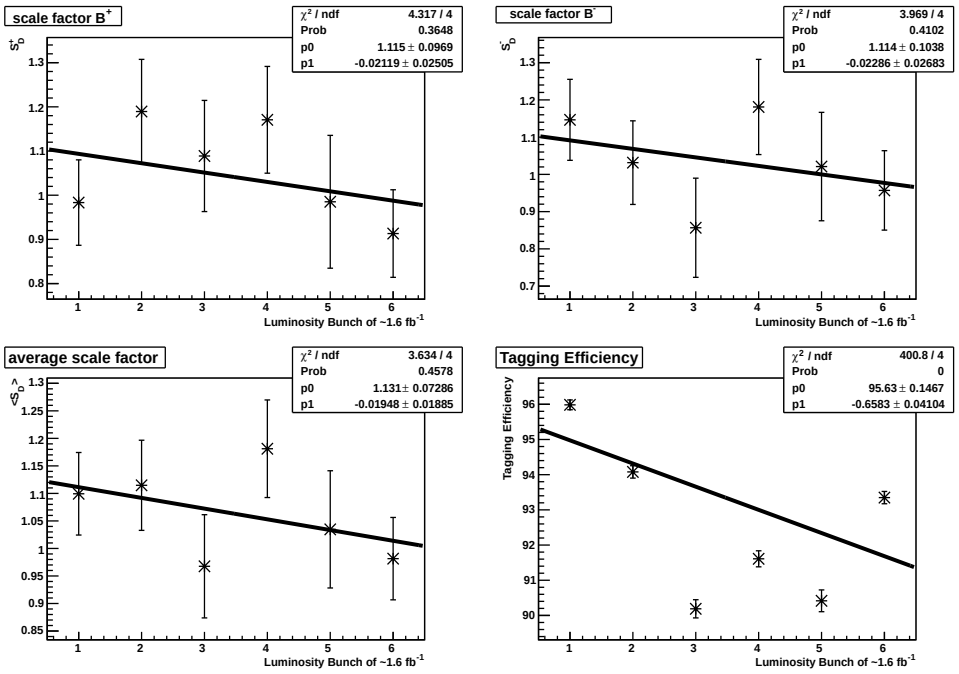


Figure 13: *OST dilution scale factors for B^+ (top, left), B^- (top, right) and an average of the two (bottom, left) and Efficiency (bottom, right) in different parts of the data. Each point in the x-axis correspond to* $\sim 1.6 fb{-}1$ *af data*

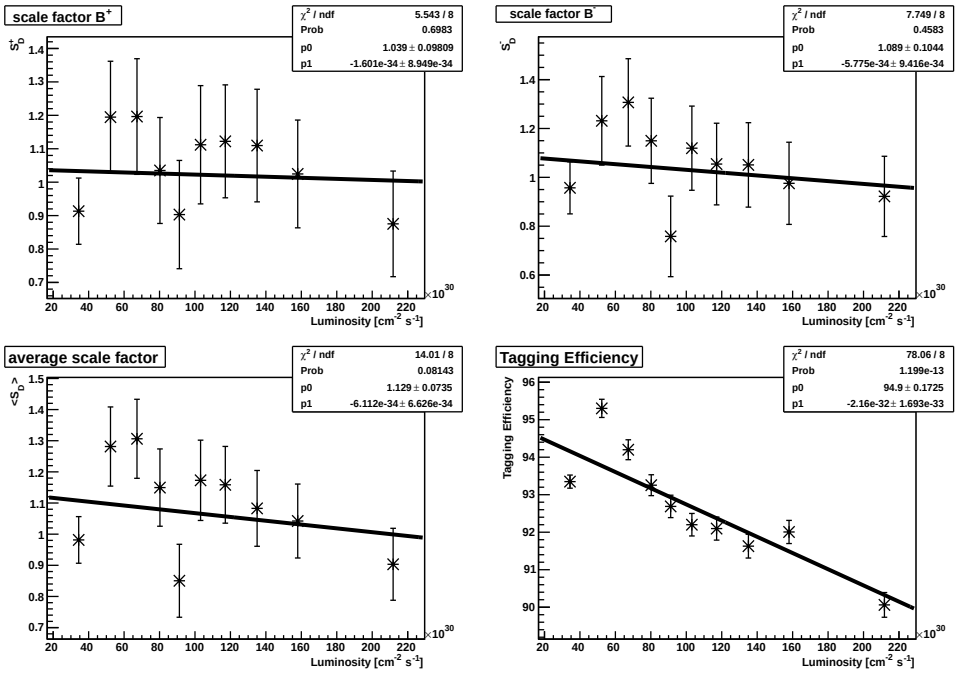


Figure 14: *OST dilution scale factors for B^+ (top, left), B^- (top, right), average them (bottom, left) and tagging efficiency (bottom, right) in instantaneous Luminosity Bins. Each x-axis bin contains roughly the same number of events (~ 55000)*

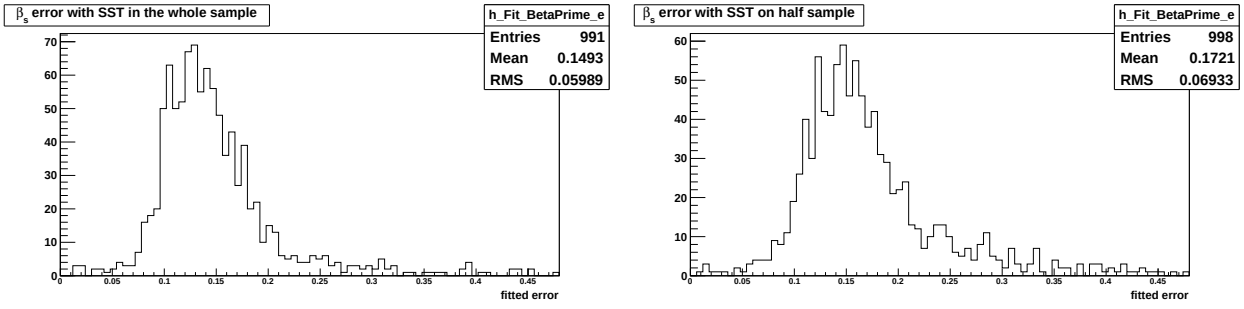


Figure 15: Comparison between β_s errors for toy with SST up 38 (left) and SST up 25 (right)

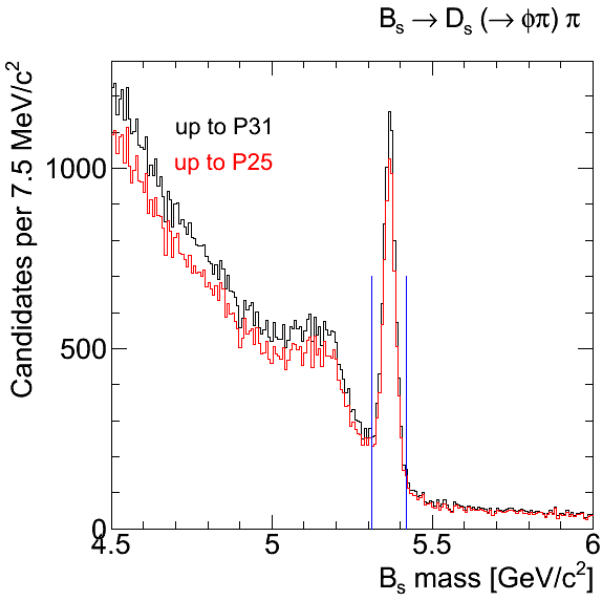


Figure 16: Mass distribution of the $B_s^0 \to D_s^- \pi^+, D_s^- \to \phi\pi^-, \phi \to K^+K^-$ in $7.2 fb^{-1}$ of CDF RunII data

7 MC Reweighting and Angular Efficiency

In this analysis, simulation of B production and decay processes and of the subsequent detector response is used to determine the detector sculpting of the angles due to the non-hermeticity of the CDFII detector.
In order to achieve this purpose, a Monte Carlo event sample obtained exploiting a *phase-space* decay model of `EVTGEN` has been used . The fact that all spins of the particles in the final state are averaged yields flat distributions in the angular variables whose acceptance we wish to study. it is necessary that the simulated events have the same characteristics of the data sample, for this reason they are fed to the same pre-selection cuts and NN selection that has been used for the data (see Sec. 5.1, 5.2).

The Monte Carlo sample that is here used has previously been used for the previous iterations of the same analysis, and for the untagged measurement of $\Delta\Gamma$ and β_s [27] and it corresponds to data conditions of the first ~ 1 fb^{-1} of data. The quality of our determination of the transversity angle sculpting depends on the agreement between our data and the generated MC in variables that affect the angular decay features of the J/ψ and the ϕ. Previously it has be seen ([27], [18], [7]) a disagreement in the $p_T(B_s)$ spectrum between this MC sample and the data, and the same effect has been observed when considering our data sample (see Fig. 19a). Since the p_T spectra can affect the distributions of the transversity angles, the agreement between data and MC has been investigated in several variables. Since part of the original reweighting in Ref [18], [7] is depending on trigger prescales, the reweighting needs to be done in order to match with the current dataset.

In order to weight the MC events according to the data sample, it is necessary to use a only-signal or sideband-subtracted data sample. The side-band subtraction is needed in this comparison, because the MC data reproduce only the signal events.

The MC reweighting procedure used involves three steps: the first takes into account the different trigger path mixture that characterizes our dataset; the second accounts for the agreement in the $p_T(B_s)$ spectrum, and the purpose of the third step is to account for the combined effect of both the different trigger paths mixture and and the $p_T(B_s)$ spectrum.

Trigger path mixture: the candidates are first of all split into two groups, depending whether the candidate triggers with a CMU-CMU or a CMU-CMX muon pair. After this, each of the two classes is split in three classes defined as:

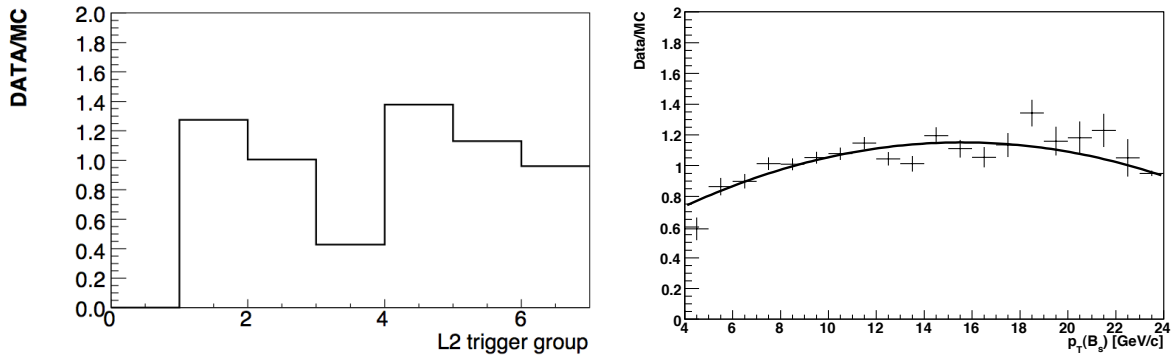
- Both muons have $p_T > 3$ GeV/c
- Both muons have $p_T > 2$ GeV/c and at least one muon has $p_T > 3$ GeV/c
- all events left, not falling in the previous two classes

In this way 6 different classes have been obtained and their fraction in the simulated events has to be adjust in order to match with the current data sample. This classes can be considered at first approximation mutual exclusive and such that their union gives the whole data sample. Fig 17a shows the weights found for the six classes. A weight is obtained by making the ratio between the number of real events belonging to one of the six classes defined above over the number of the simulated events belonging to the same class. The high of the first three bins of the histograms represents the weights for the three CMU-CMU trigger classes and the other bins involve the CMU-CMX triggers.

The $p_T(B_s)$ distribution is compared between data and MC events after the first step of reweighting procedure ("trigger classes"). The number of the simulated events has previously

been normalized to the number of the signal events in the data and the $p_T(B_s)$ region considered extends from 4 GeV/c to 24 GeV/c. This ratio has been fitted with a second order polynomial (see Fig 17b). That function will then be used to reweight the MC events.

Combined effect of trigger path admixture and $p_T(B_s)$ distribution: the weight factor associated to each simulated event accounts simultaneously for the weight factor associated to the class at which the events belongs and a factor computed using the second order polynomial used to fit the ratio of the two $p_T(B_s)$ distributions (see the previous stage of the MC reweighting procedure). Fig. 18 and Fig. 19 show the good agreement between the data sample and the so reweighted MC events both for the "trigger classes" composition and for the $p_T(B_s)$ distributions.

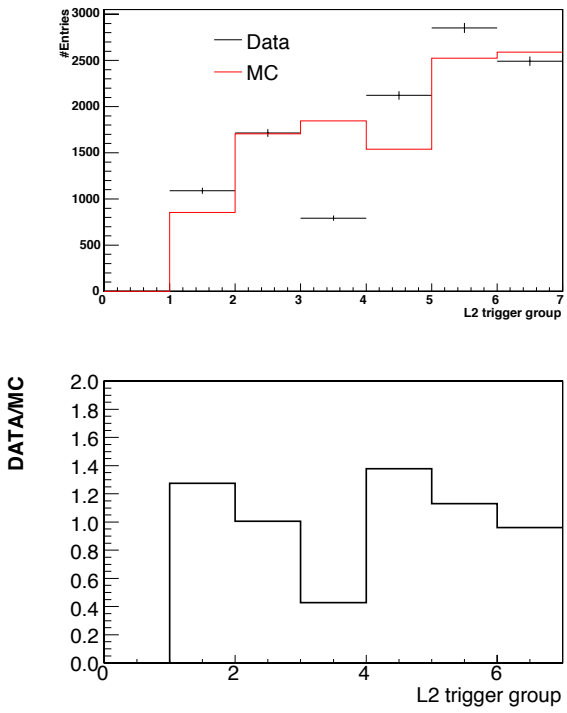


(a) Trigger group reweighting (stage 1)

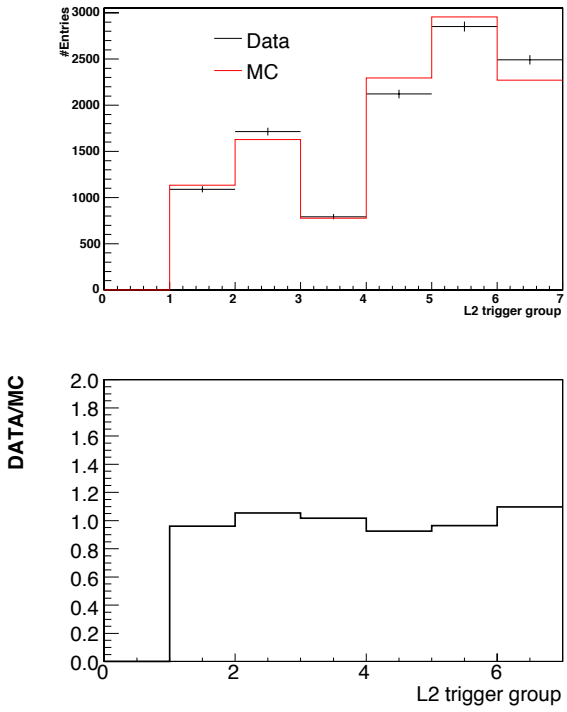
(b) $p_T(B_s)$ reweighting (stage 2)

Figure 17: MC weights. In 17a the weight according to trigger group is discrete (one different weight is associate to each trigger group defined in sec. 7. In 17b the weight according to the $p_T(B_s)$ distribution is according to the continuous function that fits the distribution in the plot, as a result for each different value of $p_T(B_s)$ there is a different weight.

The reweighted MC events are used to fit the detector efficiency with a an expansion of real spherical harmonics for the (ϑ, ϕ) angles, where spherical harmonic each term is expanded as a function of a Legendre polynomial used to fit ψ [13].

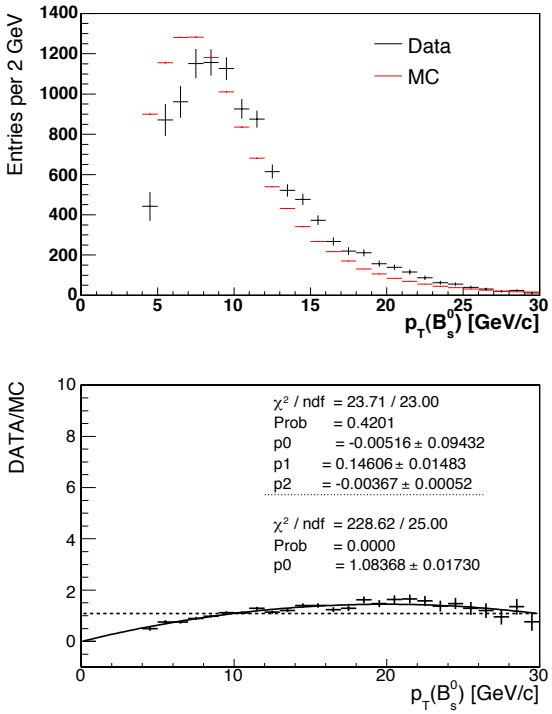


(a) MC not reweighted

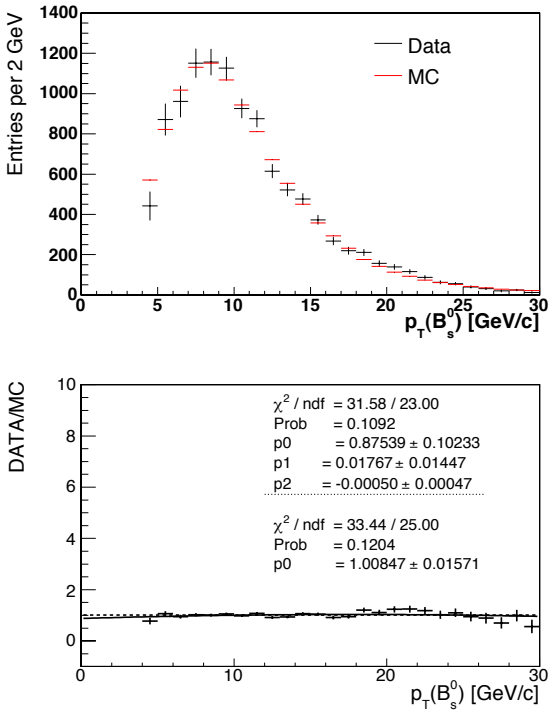


(b) MC after the 3 reweighting stages

Figure 18: Comparison between data and MC events samples: trigges classes.



(a) MC not reweighted



(b) MC after the 3 reweighting stages

Figure 19: Comparison between data and MC events samples: $p_T(B_s)$ distribution.

8 Likelihood Function

A very short description of the unbinned likelihood is given here. Since we inherited the fit used in the last iteration we will not give much details that can be found in [7]. The purpose is to provide a bief summary such as a grounding for the discussion of the results.

The PDF for an event is made of two main part, the signal and the background model. Each part is normalized to one and it is decomposed in the products of PDFs when it is appropriate to treat event variables as independent. For instance, this is a reasonable choice for all the background PDFs (see later on). We consider the following event variables: the mass of B candidates and its error (m and σ_m), the proper decay time and its error (ct and σ_{ct}), the flavor tag with its predicted diluton -both OST and SSKT- (ξ and $\mathcal{D}$) and finally the three transversity angles $\omega = (\cos\Psi, \cos\Theta, \Phi)$. Considering f_s the signal fraction in the sample, the entire PDF for an event is written as:

$$\begin{aligned} \text{PDF} = & f_s P_s(m,\sigma_m) T(ct,\omega,|\sigma_{ct},\xi,\mathcal{D}) P_s(\sigma_{ct}) P_s(\xi) P_s(\mathcal{D}) + \\ & (1-f_s) P_b(m) P_b(ct|\sigma_{ct}) P_b(\sigma_{ct}) P_b(\omega) P_b(\xi) P_b(\mathcal{D}). \end{aligned} \tag{5}$$

In the following subsections each component of the full PDF is described.

8.1 B mass PDF

The signal mass distribution $P_s(m,\sigma_m)$ is modeled by a single gaussian function with central value M, fixed to the PDG value, smeared with the event–by–event mass resolution (σ_m) scaled using a scale factor (s_m) to account for a general mis–estimation on the mass errors. The background mass model $P_b(m)$ is a first order polynomial function, and it models the dominant combinatorial background. Other source of background from physics decays (such as $B^0 \to J/\psi K^\star$) are neglected since was previously found that constitute a negligible fraction of the sample (about 1.6%) [7].

8.2 PDF of angles and ct for signal

The PDF of the signal describing the distribution of kaons and muons angles and the distribution of the ct of B candidates is not separable and it is modeled by the differential decay rate as function of the transversity angles and proper decay time, $d^4\Lambda/(d^3\omega\, dt)$.

Sculpting of the angular distributions caused by non–hermicity of the detector and selection criteria must be taken into account as well as the resolution on the measured ct of the event. The former is assumed independent of ct and is modeled by a multiplicative term $A(\omega)$ representing the angular acceptance. That is parametrized by an expansion in spherical harmonics and Legendre polynomial and it is described in detail section 7.

The ct resolution caused a smearing of the function describing the time evolution in the differential decay rate. Therefore, the exponential functions describing the decay and the oscillating sin/cos functions for the B_s mixing probability must be convolved with the ct resolution function. The latter is empirically parameterized with a sum of two gaussians whose parameters are extracted from the fit of the prompt ct–background and thus it will be described in sect. 8.3. Once the analytical form of the resolution function is given, the smeared terms properly normalized replace the time evolution functions of the decay rate. Therefore, the PDF becomes $P(\omega, ct|\sigma_{ct}) = A(\omega) d^4\Lambda/(d^3\omega\, dt) \otimes \mathcal{R}(ct|\sigma_{ct})$ and is properly normalized following prescriptions in [28].

The analytical form of the differential decay rate $d^4\Lambda/(d^3\omega\, dt)$ for $B_s^0 \to J/\psi\,\phi$ can be found in literature [29]. A compact formalism for its implementation was developed by authors of previous

iteration and it is here adopted [28]. The inclusion of the a potential S–wave contamination from background decays where the kaon pairs come from decays of the scalar $f_0(980)$ resonance was also worked out and included in the last analysis. We retain the choice of use the likelihood fit incorporating this S–wave component as the central analyis fit, even if the fraction of S–wave is found to be very tiny (see [7]). The differential decay rate for an initially produced B_s^0 meson which decay into a J/ψ and a K^+K^- pair, where the latter are decay products of the vector ϕ resonace ($\mathrm{P_{wave}}$) or the scalar $f_0(980)$ resonance ($\mathrm{S_{wave}}$), can be summarized as follows:

$$\frac{d^4\Lambda}{d^3\omega dt} \propto (1-\mathrm{f_S})|\mathrm{P_{wave}}|^2 + \mathrm{f_S}|\mathrm{S_{wave}}|^2 + 2\sqrt{(1-\mathrm{f_S})\mathrm{f_S}}\Re\mathfrak{e}(\mathrm{P_{wave}S^{\star}_{wave}}), \tag{6}$$

where $\mathrm{f_S}$ is the fraction of $B_s^0 \to J/\psi f_0(980)$ decays in our sample; $|\mathrm{P_{wave}}(\omega,\mathrm{t})|^2$ is the amplitude for $B_s^0 \to J/\psi\,\phi$ decomposed in the polarization amplitudes A_0, $A_\parallel$ and $A_\perp$ [29]; $|\mathrm{S_{wave}}|^2$ is the amplitude for the $B_s^0 \to J/\psi f_0(980)$ component and finally $\Re\mathfrak{e}(\mathrm{P_{wave}S^{\star}_{wave}})$ is the interference term. Explicity written, $|\mathrm{P_{wave}}(\omega,\mathrm{t})|^2$ has six components (3 squared moduli and 3 interference terms of polarization amplitudes), $|\mathrm{P_{wave}}(\omega,t)|^2 = \sum_{i=1}^{6} K_i(t)g_i(\omega)$; since $B_s^0 \to J/\psi f_0(980)$ is a pseudoscalar to vector-scalar decay, the decay amplitudes is given only by one term, $|\mathrm{S_{wave}}(\omega,t)|^2 = K_7(t)g_7(\omega)$; the iterference comprises 3 terms, since the $\mathrm{S_{wave}}$ component interferes with each of the polarization state of $B_s^0 \to J/\psi\,\phi$, therefore $\Re\mathfrak{e}(\mathrm{P_{wave}S^{\star}_{wave}})(\omega,t) = \mathcal{I}\sum_{i=8}^{9} K_i(t)g_i(\omega)$. The factor $\mathcal{I}$ represent an overlap integral between the ϕ and f_0 mass propagators once the dependence of the total decay rate on the invariant mass of the kaon pair in integrated in the range [1.009,1.028] GeV/c^2. In tab. 3 the ten $K_i(t)$ and $g_i(\omega)$ terms are reported. The mixing frequency Δm entering $K_i(t)$ has been gaussian costrained in the fit to the measured value [1] and its uncertainty is taken as standard deviation of the gaussian.

	$K_i(t)$	$g_i(\omega)$	CP
1	$\lvert A_0\rvert^2\mathcal{O}^+(t)$	$4\cos^2\Psi(1-\sin^2\Theta\cos^2\Phi)$	even
2	$\lvert A_\parallel\rvert^2\mathcal{O}^+(t)$	$\sin^2\Psi(1-\sin^2\Theta\sin^2\Phi)$	even
3	$\lvert A_\perp\rvert^2\mathcal{O}^-(t)$	$\sin^2\Psi\sin^2\Theta$	odd
4	$\lvert A_\parallel\rvert\lvert A_\perp\rvert\mathcal{E}_I(t,\delta_\perp-\delta_\parallel)$	$-\sin^2\Psi\sin 2\Theta\sin\Phi$	mix
5	$\lvert A_\parallel\rvert\lvert A_0\rvert\cos\delta_\parallel\mathcal{O}^+(t)$	$\frac{1}{\sqrt{2}}\sin 2\Psi\sin^2\Theta\sin 2\Phi$	even
6	$\lvert A_\perp\rvert\lvert A_0\rvert\mathcal{E}_I(t,\delta_\perp)$	$\frac{1}{\sqrt{2}}\sin 2\Psi\sin 2\Theta\cos\Phi$	mix
7	$\mathcal{O}^-(t)$	$2(1-\sin^2\Theta\cos^2\Phi)$	odd
8	$\lvert A_\parallel\rvert\mathcal{E}_R(t,\delta_\parallel-\delta_S)$	$2\cos\Psi(1-\sin^2\Theta\cos^2\Phi)$	mix
9	$\lvert A_\perp\rvert\sin(\delta_\perp-\delta_S)\mathcal{O}^-(t)$	$\frac{1}{\sqrt{2}}2\sin\Psi\sin^2\Theta\sin 2\Phi$	odd
10	$\lvert A_0\rvert\mathcal{E}_R(t,-\delta_S)$	$\frac{1}{\sqrt{2}}2\sin\Psi\sin 2\Theta\cos\Phi$	mix

Table 3: Time and angular dependeces of the $B_s^0 \to J/\psi K^+K^-$ decay rate. We have defined $\mathcal{O}^{\pm}(t) = e^{-\Gamma t}(\cosh\frac{\Delta\Gamma t}{2} \mp \cos 2\beta_s\sinh\frac{\Delta\Gamma t}{2} \pm \sin 2\beta_s\sin\Delta mt)$, $\mathcal{E}_I(t,\alpha) = e^{-\Gamma t}(\sin\alpha\cos\Delta mt - \cos\alpha\cos 2\beta_s\sin\Delta mt - \cos\alpha\sin 2\beta_s\sinh\frac{\Delta\Gamma t}{2})$, $\mathcal{E}_R(t,\alpha) = e^{-\Gamma t}(\cos\alpha\cos\Delta mt - \sin\alpha\cos 2\beta_s\sin\Delta mt - \sin\alpha\sin 2\beta_s\sinh\frac{\Delta\Gamma t}{2})$. The amplitude A_0 is taken real, and $delta_S$ is the strong phase of the $\mathrm{S_{wave}}$ amplitude.

The decay rate for a initially $\bar{B}_s^0$ meson is sligthly different because of the change of sign of cos/sin function in above formulae. Therefore, we must consider a different PDF ($\bar{P}(\omega,ct|\sigma_{ct})$) for it and introduce the flavor tag decision ξ to chose between the two probabilities. We need to include the two tagging algorithms indicated here with index 1 for the OST and with index 2 for the SSKT. Each dilution $\mathcal{D}$ of the tagging decision is multiplied by a scale factor s_i to account for some mis–estimation of the algorithm. This scale factors are separately extracted from a dedicated

calibration of both taggers (see Sect. **??**). Then, in the likelihood each scale factor is free to float within a gaussian constrain which has the results of the calibration and its error as central value and standard deviation respectively. Two main differences are adopted with respect to the previous analysis:

1. we use a single scale factor of OST dilutions for both tagging decisions (B or $\bar{B}$) instead of using two separte scale factors. This choice is motived by the result of the updated calibration of the OST which gives the same scale factor for the $B^+ \to J/\Psi K^+$ and $B^- \to J/\Psi K^-$ decays (see sect. **??**).

2. We used the OST in the whole dataset, while the SSKT is used for the first 5.2 fb^{-1} only because of the lack of its calibration in the second part of data (see sect. **??**). We've studied the impact of the inclusion of the SSKT in the entire dataset using pseudo–experiments: an improvement at most of 10% is found on the statistical resolution of β_s when considering the same tagging performances in the second part of the sample. Since it is not clear the actual improvement in total resolution when including a potential systematic uncertainty associated to latter assumption, we decide to take a conservative choice and not use the SSKT in the second part of the data.

Finally the PDF that includes all the terms described in this section is:

$$\begin{aligned} T(\omega, ct|\sigma_{ct}, \mathcal{D}_1, \mathcal{D}_2, \xi_1, \xi_2) = & \frac{1+\xi_1 s_1 \mathcal{D}_1}{1+|\xi_1|}\frac{1+\xi_2 s_2 \mathcal{D}_2}{1+|\xi_2|} P(\omega, ct|\sigma_{ct}) \\ & + \frac{1-\xi_1 s_1 \mathcal{D}_1}{1+|\xi_1|}\frac{1-\xi_2 s_2 \mathcal{D}_2}{1+|\xi_2|} \bar{P}(\omega, ct|\sigma_{ct}). \end{aligned} \tag{7}$$

8.3 Background lifetime PDF

Usually, three components are identified in the ct distribution of background events:

- a prompt peak which is most of the combinatorial background events, that are expected to have no significant lifetime;

- two positive exponentials used to describe the longer lived background events;

- a negative exponential is needed to account for those background events that present a negative decay length in the vertex reconstruction.

The prompt peak of the background has a relevant role since allows to determine the resolution function in ct. The resolution function has been modeled using two gaussians, as follows:

$$\mathcal{R}(ct|\sigma_{ct}) = f_1 G_1(ct, s_{ct1} \cdot \sigma_{ct}) + (1-f_1) G_2(ct, s_{ct2} \cdot \sigma_{ct}), \tag{8}$$

where f_1 is the fraction of the first gaussian, and s_{ct1} and s_{ct2} represent the scale factors of the event by event ct-error, σ_{ct}, that are free to float in the fit. In fig. 20 we report the ct distribution of background along with fit projection overlaid.

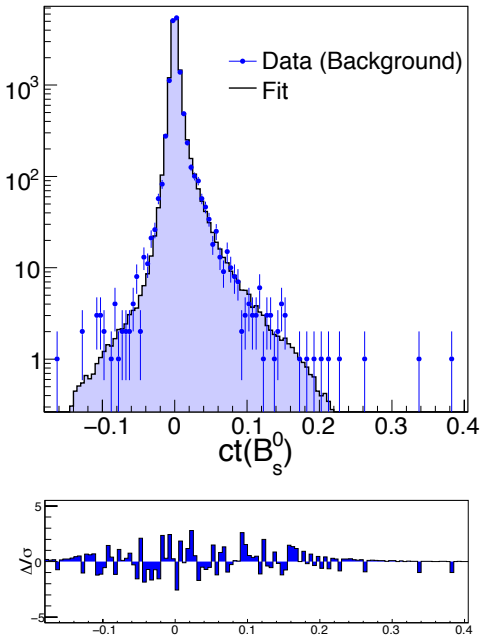


Figure 20: ct distribution of background along with fit projection overlaid

8.4 Background angular PDF

The background PDF of the transversity angles is parametrized empirically from data of the sidebands of B mass distribution. Each transversity angle distribution is reasonably considered uncorrelated with respect to the other two angles. For this reason each angle distribution is modeled separately: $P(\omega) = P(\cos\Theta)P(\Phi)P(\cos\Psi)$. Moreover, the background angular distributions are assumed to not depend on ct. We have checked this is a fairly good approximation at least for event with $ct > -0.006$ micron (see fig. 21). As for previous analysis we consider $f(\cos\Theta) \propto 1 - a\cos^2(\Theta)$ and $f(\Phi) \propto 1 + b\cos(2\Phi)$ (where a and b are fit parameters), while we adopt a flat distribution for $\cos(\Psi)$ instad of a parabolic shape analogous to the $\cos\Theta$ description. In fig. 22 the projection of fit to the sidebands events is shown.

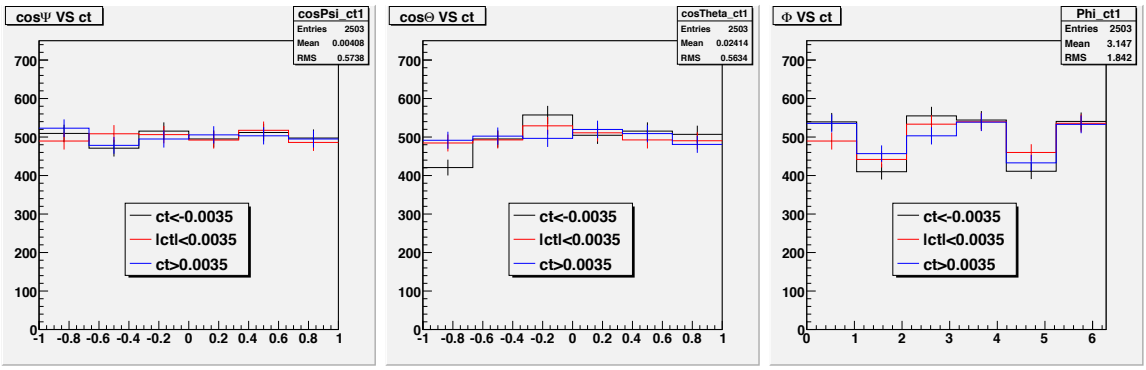


Figure 21: Background angular distributions in slice of ct.

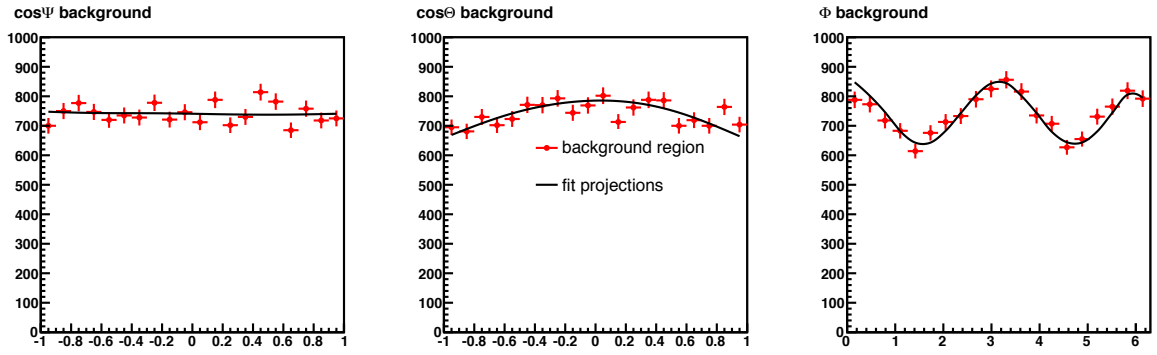


Figure 22: Angular distributions of background events. The fit to $\cos\Psi$ with $1 - d\cos^2(\Psi)$ gives $d = 0.003 \pm 0.016$, therefore we removed that parameter in the global fit.

8.5 Lifetime error PDFs

For the decay time error, the PDF has been built using Gamma functions as follows:

$$P(\sigma_{ct}) = f_P \frac{(\sigma_{ct})^{a_1} e^{\frac{\sigma_{ct}}{b_1}}}{(b_1)^{a_1+1}\Gamma(a_1+1)} + (1-f_P)\frac{(\sigma_{ct})^{a_2} e^{\frac{\sigma_{ct}}{b_2}}}{(b_2)^{a_2+1}\Gamma(a_2+1)}, \tag{9}$$

where a_1, b_1, a_2, b_2 define the mean and the width of respectively the first and the second distribution, and f_P define the fraction of the first distribution. Both the background and the signal PDF has the same form of eq. 9 with two different set of parameters. These parameters are found with a preliminary lifetime–only fit on the data. Parameters determined with this method are then used as input in the full likelihood used for the complete analysis. Distributions of decay time errors with fit projections overlaid are shown in fig. 23.

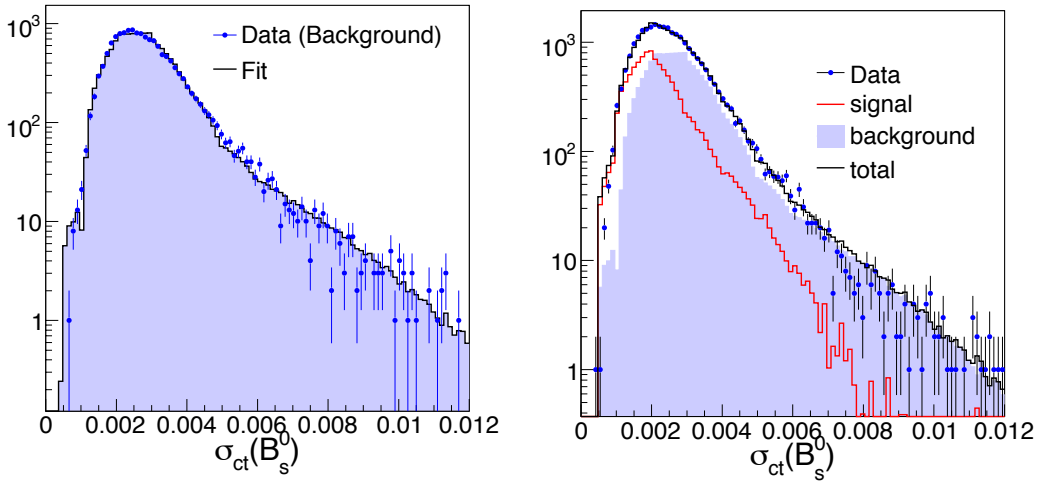


Figure 23: Distribution of σ_{ct} in signal (left) and background (right) regions.

8.6 Flavor Tagging PDFs

The PDF that accounts for the tag decision takes into account the efficiencies ε_i of the two taggers (see sect. **??**) is:

$$P(\xi) \equiv P(\xi_1)P(\xi_2) = \sum_{j=1}^{2} \varepsilon_j \cdot \delta(|\xi_j| - 1) + (1 - \sum_{j=1}^{2} \varepsilon_j) \cdot \delta(\sum_{j=1}^{2} \xi - 0) \tag{10}$$

The PDF of the dilution ($P_i(\mathcal{D})$) is modeled with a template that consists of an histogram, taken from the data itself. Separate histograms are produced for the signal and the background, different histograms are produced for different taggers. The signal histograms are produced using background subtracted data; the background dilution histograms are complementary produced used mass sidebands region data. These distributions are reported in Fig. 25 and **??** for respectively signal and background and for both OST and SSKT.

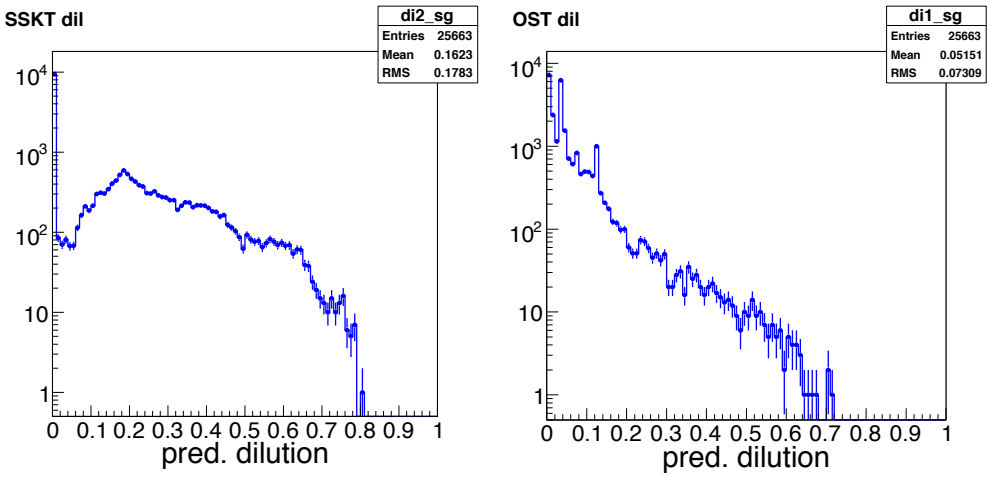


Figure 24: Dilution histograms for signal: OST on the right, SSKT on the left.

8.7 Symmetry of the likelihood function

The PDF in eq. 5 is invariant under the simultaneous transformations:

$$\begin{cases} \beta_s \to \pi/2 - \beta_s \\ \Delta\Gamma \to -\Delta\Gamma \\ \delta_{\|} \to 2\pi - \delta_{\|} \\ \delta_{\perp} \to \pi - \delta_{\perp} \end{cases}$$

when the S–wave component is neglected. That leads to a two–fold ambiguity for the minimum of the likelihood. Considering also the contribution of S–wave, an approximated symmetry still holds when along with above transformations the strong phase of S–wave amplitude changes in $\delta_S \to \pi - \delta_S$. The latter asymmetry shoul be broken as larger as the fraction of S–wave in the sample, since the asymmetric shape of the integral $\mathcal{I}$ in eq. 6 around the ϕ pole due to the relativisic Breit Wigner function of its mass propagator. We have numerically checked the discrete symmetries

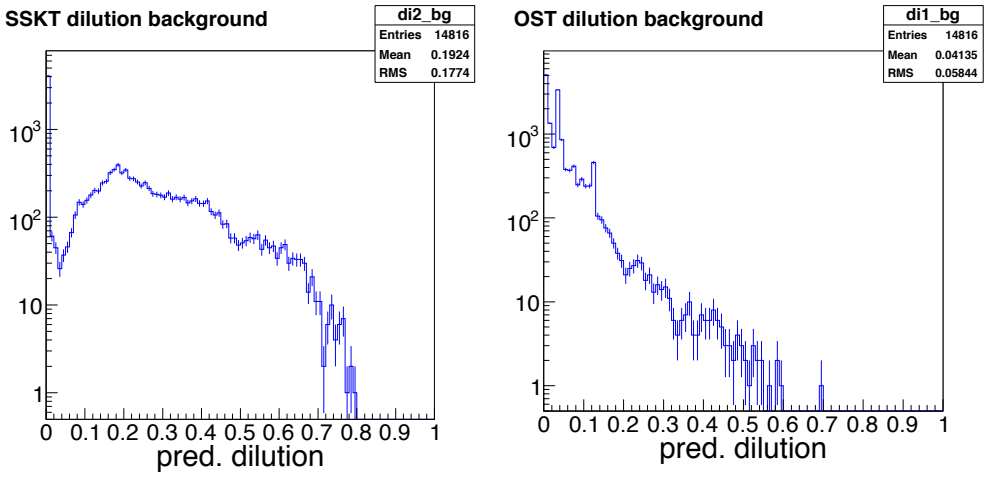


Figure 25: Dilution histograms for background: OST on the right, SSKT on the left.

of our likelihood sampling some points of the space parameters. When considering a fraction of S–wave as the value found in our data ($\approx 1\%$), we can treat the likelihood symmetric under the above transformations.

In the case of the likelihood without flavor tagging, the symmetry under the transformation written above holds when all the four parameters transform simultaneously, but also when they transform separately for:

$$\begin{cases} \beta_s \to \pi/2 - \beta_s \\ \Delta\Gamma \to -\Delta\Gamma \end{cases}$$

and

$$\begin{cases} \delta_{\parallel} \to 2\pi - \delta_{\parallel} \\ \delta_{\perp} \to \pi - \delta_{\perp} \end{cases}$$

therefore a four–fold ambiguity is present for the minima of the untagged likelihood. Moreover, it is known that an additional complication arises in the untagged fit from the fact that the strong phase $\delta_{\perp}$ appears always in a product with $\sin 2\beta_s$. As a result, in case of tiny CP violation there is no sensitivity to $\delta_{\perp}$, and the fit tends to bias the result as by increasing the CP violation to gain sensitivity on $\delta_{\perp}$ as an additional parameter available to describe the statistical fluctuations.

8.8 Summary of the fit variables

At this point it could be useful to summarize the parameters entering the maximum likelihood fit. Table 4 lists all of them along with a very short description. In this table are not listed the parameters used to model the error lifetime PDF, since they are not floating in the full fit, but they are determined with a previous lifetime–only fit as it was previously described.

Parameter	Description
β_s	β_s CP -violating phase
$\Delta\Gamma$	$\Gamma_L - \Gamma_H$
$\alpha_\perp$	CP odd fraction
$\alpha_\parallel$	fraction in CP even states
$\delta_\perp$	$\arg(A_\perp A_0)$
$\delta_\parallel$	$\arg(A_\parallel A_0)$
$c\tau$	$\frac{1}{\Gamma_s} = \frac{2}{\Gamma_L+\Gamma_H}$
A_{SW}	fraction of S-wave KK component in the signal
δ_{SW}	phase of S-wave component
Δm_s	B_s^0 mixing frequency
f_s	Signal fraction
s_m	Mass error scale factor
m	B_s mass [GeV/c^2]
p_1	mass background slope
$s_{c\tau 1}$	lifetime error scale factor 1
$s_{c\tau 2}$	lifetime error scale factor 2
f_{sf_1}	fraction of first lifetime error scale factor
f_p	fraction of prompt background
f_-	fraction of bkg which decays with λ_-
f_{++}	fraction of bkg which decays with λ_{++}
λ_-	Effective bkg lifetime, neg. component
λ_+	Effective bkg lifetime, pos. component 1
λ_{++}	Effective bkg lifetime, pos. component 2
ϕ_1	parameter in bkg fit to Φ
$cos(\vartheta)_1$	parameter in bkg fit to cos(Θ)
$S_D(OST)$	OST dilution scale factor
$S_D(SST)$	SST dilution scale factor
$\varepsilon_b(OST)$	OST tagging efficiency for background
$\varepsilon_b(SST)$	SST tagging efficiency for background
$A^+(OST)$	OST background positive tag asymmetry
$A^+(SST)$	SST background positive tag asymmetry
$\varepsilon_s(OST)$	OST tagging efficiency for signal
$\varepsilon_s(SST)$	SST tagging efficiency for signal

Table 4: Fit parameters
RED: parameters of the time and angles PDF; PURPLE: mass PDF; BLUE: lifetime PDF; ORANGE: ct model for the background; YELLOW: angular PDF for the background; PINK: tagging parameters.

9 Fitter validation and checks

The techniques used to verify the good behavior of the fit include pull studies, the probe on the sensibility of the fit towards small changed in the inputs values and distributions, and the examination. The consistency check with the previous analysis is carried out dividing the dataset in two subsamples and fitting them separately.

9.1 Fit dependence input values and distributions

We first perform a test of the sensitivity to small changes in the input parameters and distributions of our fitter. Some of the parameters of the PDFs of the Likelihood need to be fed in the fitter as input. Those inputs have been updated in order to perform the fit with the new data sample. The parameters used to quantify the detector sculpting effects (see sec. **??**), the parameters describing the σ_{ct} PDF (see sec. 8.5) , the parameters describing the background angular PDF (see sec. 8.4) and the tagging dilution scale factors are here indicated as "input parameters", while with the term "input distributions" we refer to the tagging dilution histograms, that are used in the fit as PDF for the tagging component (see sec. 6), and histogram describing the detector acceptance. In order to probe the fit sensitivity to small variations of the inputs, two different fits have been performed with input files where all the input parameters and distributions are defined. In the first case we use the updated inputs of our final fit and in the other case the inputs of the previous iteration of the analysis. Tab. 5 and Fig. 26 summarize the obtained results using a subset of the final statistics corresponding to 8.4 fb^{-1} of integrated luminosity. The NOT updated input colum refer to the same tuning used for the latest CDF measurement using 5.2fb^{-1} of integrated luminosity **??**.

Parameter	NOT updated input	updated input
$c\tau$	0.04604 ± 0.00063	0.04580 ± 0.00062
$\Delta\Gamma$	0.062± 0.028	0.063 ± 0.029
$\alpha_\perp$	0.272± 0.011	0.272 ± 0.012
$\alpha_\parallel$	0.308 ± 0.012	0.311 ± 0.012
$\phi_\perp$	2.877 ± 0.522	2.949 ± 0.612

Table 5: Fit parameters results in case of input updated or not.

9.2 Pull studies

We investigate the fit consistency probing whether biases are present in values of the fitted parameters. About 1000 pseudo experiments have been generated with the same statistics of data ($\sim$ 55000 events) and we look at the parameters distributions of all pseudo experiments.

9.2.1 CP conserving fit

Let us firstly examine the pull distribution of the parameters that we want to measure: $\Delta\Gamma$, $c\tau$, $\alpha_\parallel$, α_{CPOdd}, and the strong phases $\delta_\perp$ and $\delta_\parallel$ (see Fig. 27). The plots show a good behavior for all the quantities, apart from $\delta_\parallel$. The pulls for $\delta_\parallel$ show a non-Gaussian behavior and thus it is not possible to quote a value for that parameter with unbiased errors. The reason why the pull distribution for $\delta_\parallel$ is so badly behaving, could arise from the likelihood symmetries. For $\delta_\parallel$ there is a symmetry for reflection around π; for values close to π, probably the fit cannot always clearly determine the value between the two symmetric cases. and has the tendency to return the boundary value, π, as the fitted value for $\delta_\parallel$, since it precisely in the middle of the two possible solutions.

It is useful to look also at the pull distribution for the S-wave amplitude A_{SW} and its phase with respect to the P-wave δ_s (see Fig. 28), focusing in particular on the S-wave amplitude. The S-wave amplitude parameter A_{SW} is left free to float in the full fit, but within the boundary at [0,1] to

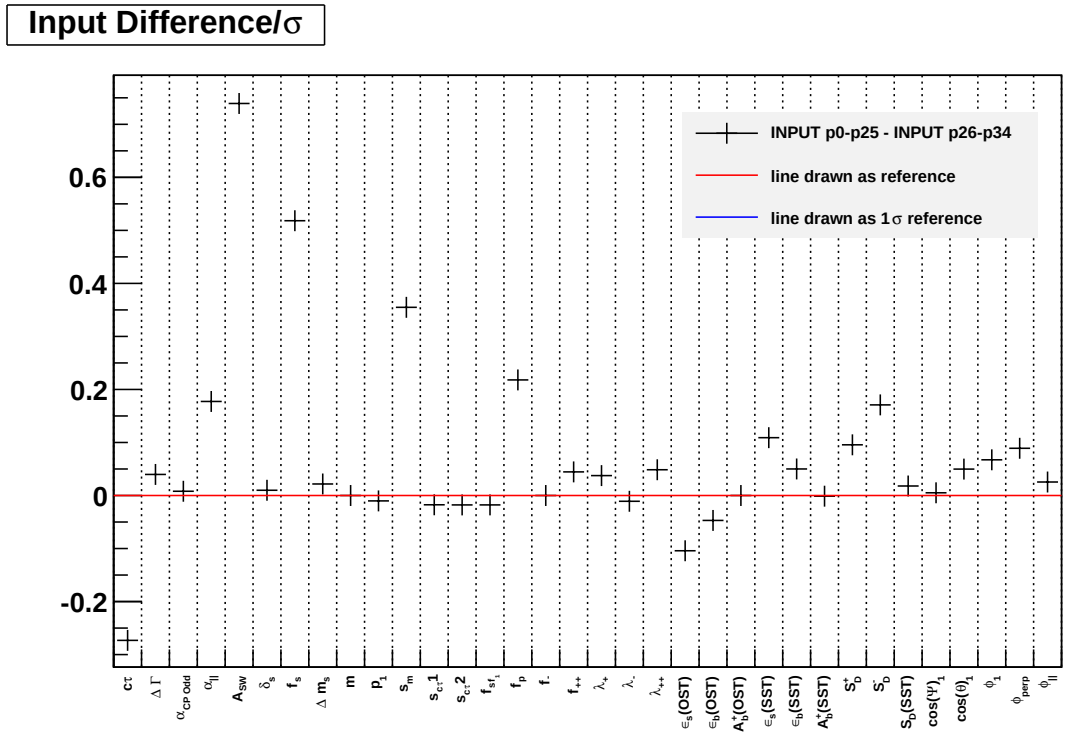


Figure 26: Comparison between fit results obtained with the updated fit inputs or with the inputs of the previous analysis iteration.

prevent A_{SW} from assuming negative values that are not allowed in our likelihood parametrization where the square root of the S-wave amplitude enters. The value obtained for the A_{SW} by the fit to data is small, around $2\% \pm 2\%$, i.e. very close to the boundary. It has been noticed that when β_s is left floating, sometimes the fit shows convergence problems, because the minimizer(`MINUIT`) has been stuck at the limit for A_{SW} (see Ref. [?]). The strategy to avoid it is to restart the minimization from the local minimum of the likelihood, moving the starting point of A_{SW}. This strategy has been applied for the fit on data, but not for the pull studies. This explains the strange behavior of the pull distribution for A_{SW} (Fig. 29c). As can be seen from Fig. 29a slightly more than half of the times the fit finds a value for A_{SW} that is either zero or in the interval between 0.0 and 0.02. The residual plot in Fig. 29b shows how the boundary at $A_{SW} = 0$ is responsible of the values found for A_{SW} and therefore the pull distribution behavior (Fig. 29c).
To support the hypothesis that the boundary was causing the bad behavior of the A_{SW} pull distribution, another set of pseudo experiments has been generated, with the generation value of $A_{SW} = 25\%$ far away from the boundary. We expect a gaussian distribution centered on 0.25 for the values of A_{SW} fitted on the pseudo experiment, and a Normal distribution for the pulls. Fig 30 shows the obtained distribution which is in agreement with the expectations.

Another check is to verify whether the mean error associated by the fit to a given parameter in the toy studies, is comparable with the error obtained from the fit on data. This comparison has been made for the physics parameters we are interested in, and is reported in Tab. 6, together with the parameters describing the pull distributions (mean and width). For the physical quantities that we want to measure in this thesis, the pull distributions show that the values found by the

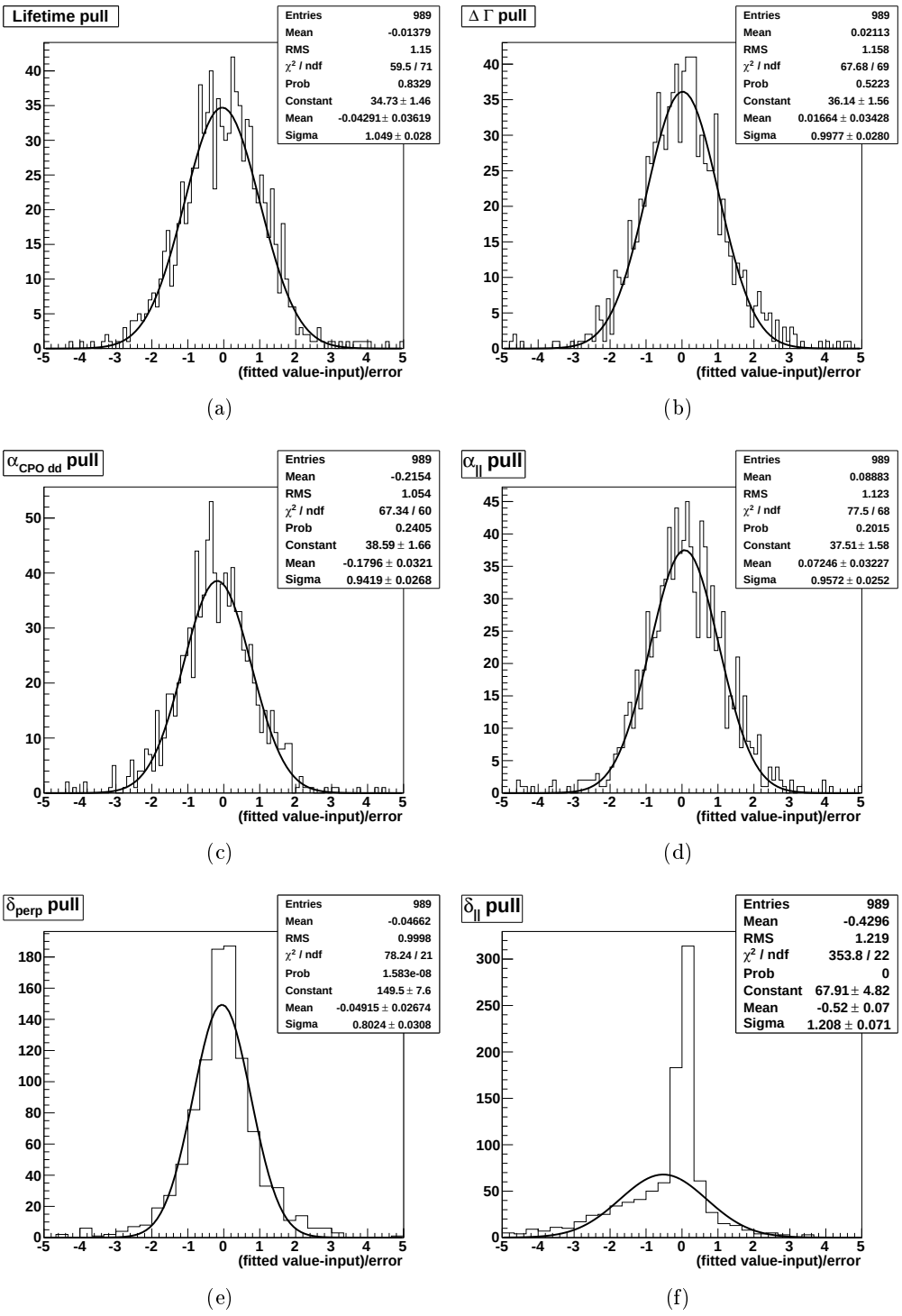
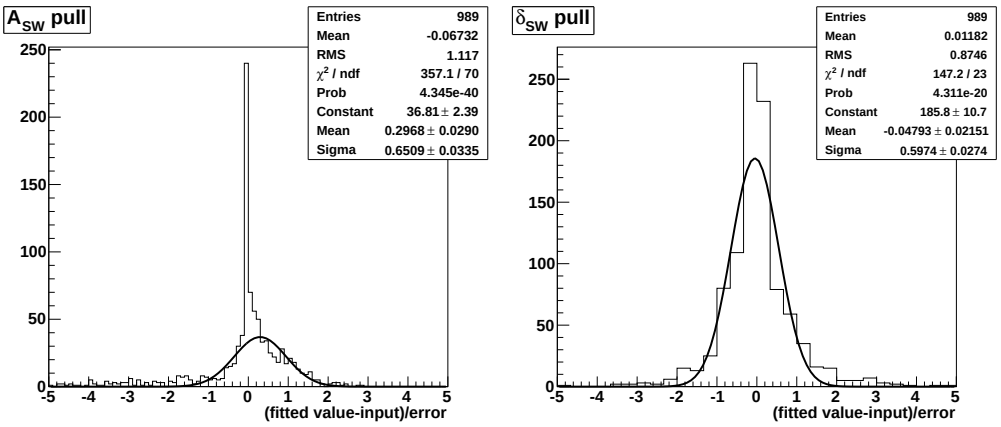


Figure 27: Pull distributions for the main physics parameters.

fit are reliable, and the comparison between the statistical error in the data fit with the average error from the pseudo experiment shows a satisfactory agreement, with the exception of $\delta_\perp$. In this



(a) (b)

Figure 28: Pull distributions for the S-wave amplitude A_{SW} and it phase relative to the P-wave δ_s.

case the mean error is much larger than the error assigned by the fit on data; this behavior can be understood by looking at the likelihood profile in the neighborhood of the fit global minimum. Considering a large number of pseudo experiments, it can happen that the fit for some pseudo-experiments converges in correspondence of a minimum that is not the generation value, due to the proximity of the two minima. This enlarges the mean error.

Parameter	Pull mean	Pull σ	Mean Error	Fit Error
$c\tau$	- 0.043±0.036	1.049±0.028	0.00061	0.00062
$\Delta\Gamma$	0.016±0.034	0.998±0.028	0.028	0.029
α_{CPOdd}	-0.180±0.032	0.942±0.028	0.012	0.012
$\alpha_{\|\|}$	0.072±0.032	0.957±0.027	0.013	0.012
$\phi_\perp$	-0.049±0.027	0.802±0.030	1.354	0.612

Table 6: Mean and σ of the pull distribution; variable mean error and in the last column the fitted parameter error.

In Fig. 31 we show mean and width of the pull distributions for all the parameters appearing in the full likelihood function.

9.2.2 CP violating fit

We performed studies with pseudo-experiments generated at $\beta_s^{J/\psi\phi} = 0.11$ (the central value found in our data fit on the whole datasample). Figure 32 shows the pull distribution for the main phisics parameters. When $\beta_s^{J/\psi\phi}$ is allowed to float significant bias is present on $\Delta\Gamma$ (about 200 times the bias found in the CP-conseving studies). Since in the last iteration significant biases were found for the mixing phase, particular attention regreats figure 33. It shows no significant bias probably related to the more statistic used. Besides, same considerations as in the CP-conserving case hold for $\delta_{||}$ A_{SW} and δ_{SW}.

Following the same scheme adopted in describing the CP-Conserving fit, in table 7 we report a

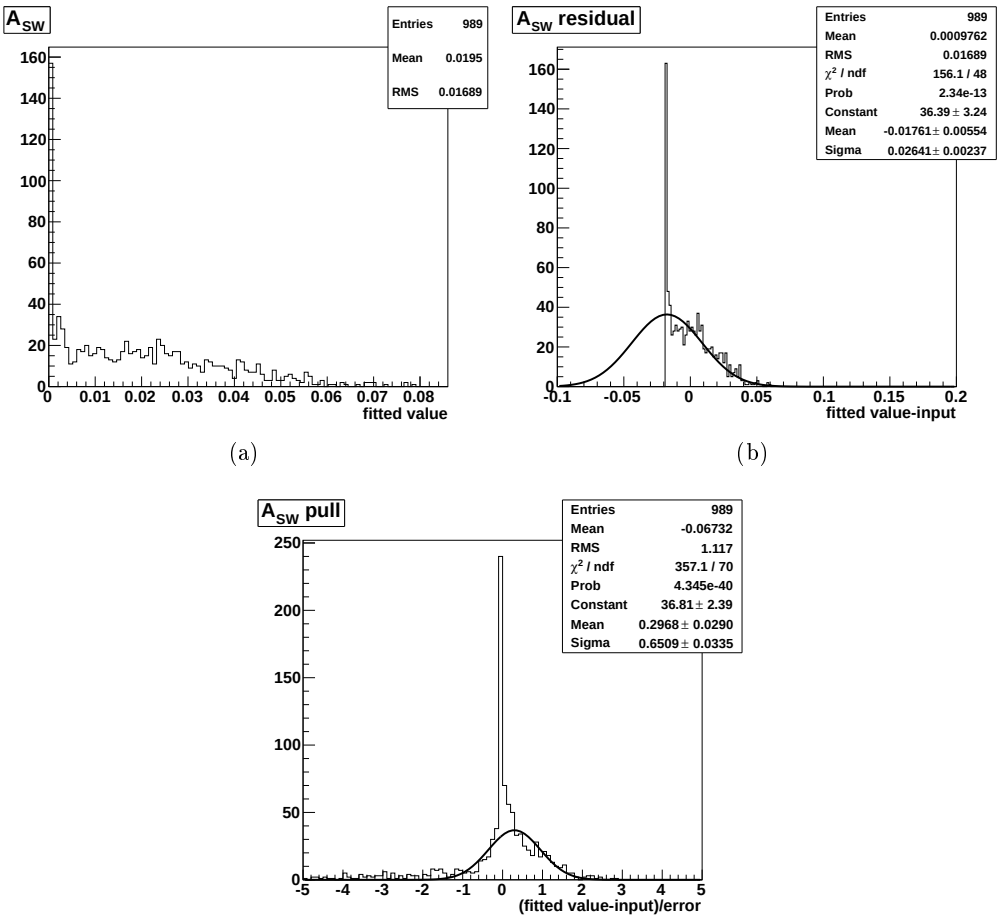


Figure 29: Value, residual and pull distributions for the S-wave amplitude A_{SW}; pseudo experiments generated with the A_{SW} value of 1.8 % as obtained from the CP conserving fit on data.

comparison between the mean error on a given parameter in the toy and the error obtained fitting data. Again, the considerations made in the CP-conserving case are still applicable here.

Parameter	Pull mean	Pull σ	Mean Error	Fit Error
$\beta_s^{J/\psi\phi}$	0.046 $\pm$0.033	0.984$\pm$0.032	0.197	0.123
$c\tau$	-0.086$\pm$0.030	0.914$\pm$0.025	0.00069	0.00058
$\Delta\Gamma$	0.026$\pm$0.032	0.965$\pm$0.028	0.038	0.028
α_{CPOdd}	-0.197$\pm$0.032	0.949$\pm$0.028	0.012	0.011
$\alpha_{\parallel}$	0.170$\pm$0.032	0.963$\pm$0.027	0.013	0.011
$\phi_{\perp}$	-0.020$\pm$0.019	0.537$\pm$0.026	1.701	0.739

Table 7: Mean and σ of the $10fb^{-1}$ CP-violating fit pull distribution; variable mean error and in the last column the fitted parameter error.

Finally, in Fig. 34 we show mean and width of the pull distributions for all the parameters

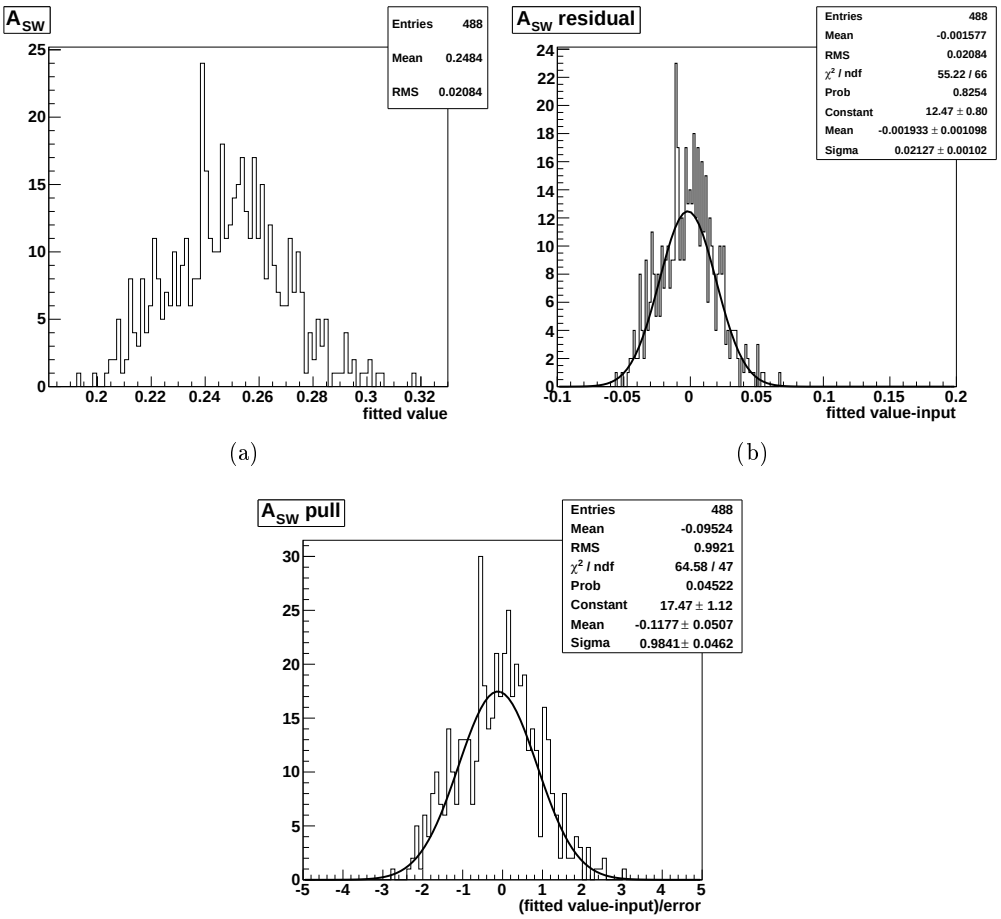


Figure 30: Value, residual and pull distributions for the S-wave amplitude A_{SW}; pseudo experiments generated with the A_{SW} value of 25 %

appearing in the full likelihood function.

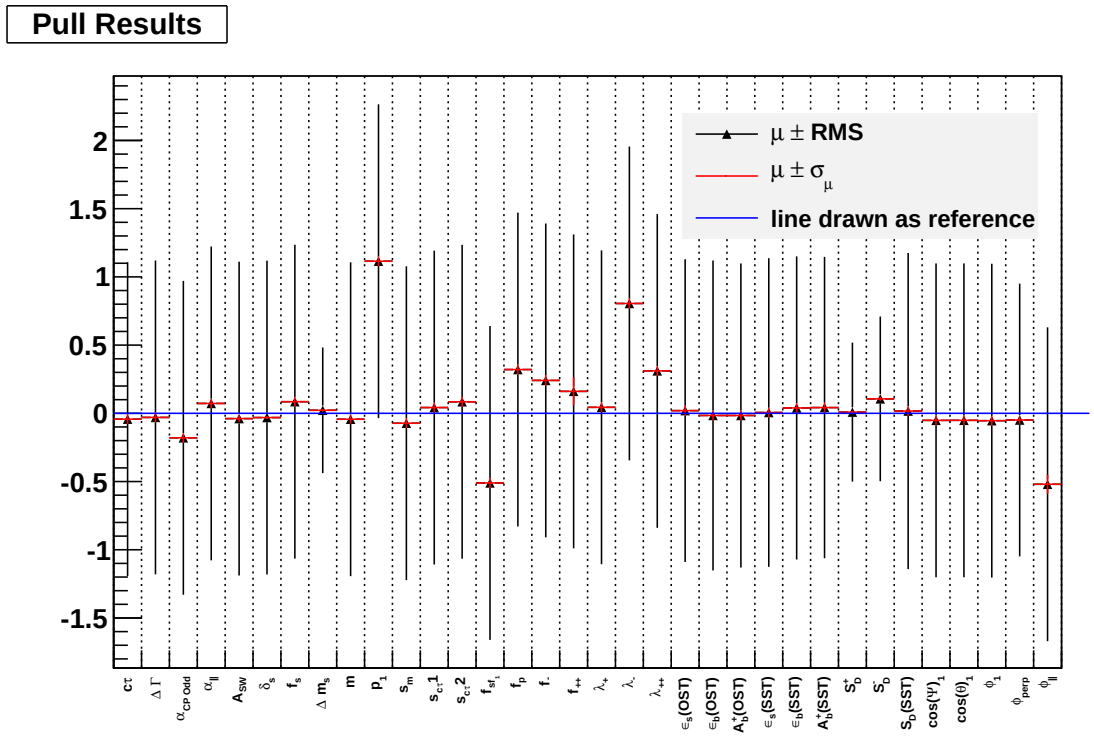


Figure 31: Summary of the pull distributions for all the variables of the likelihood function.

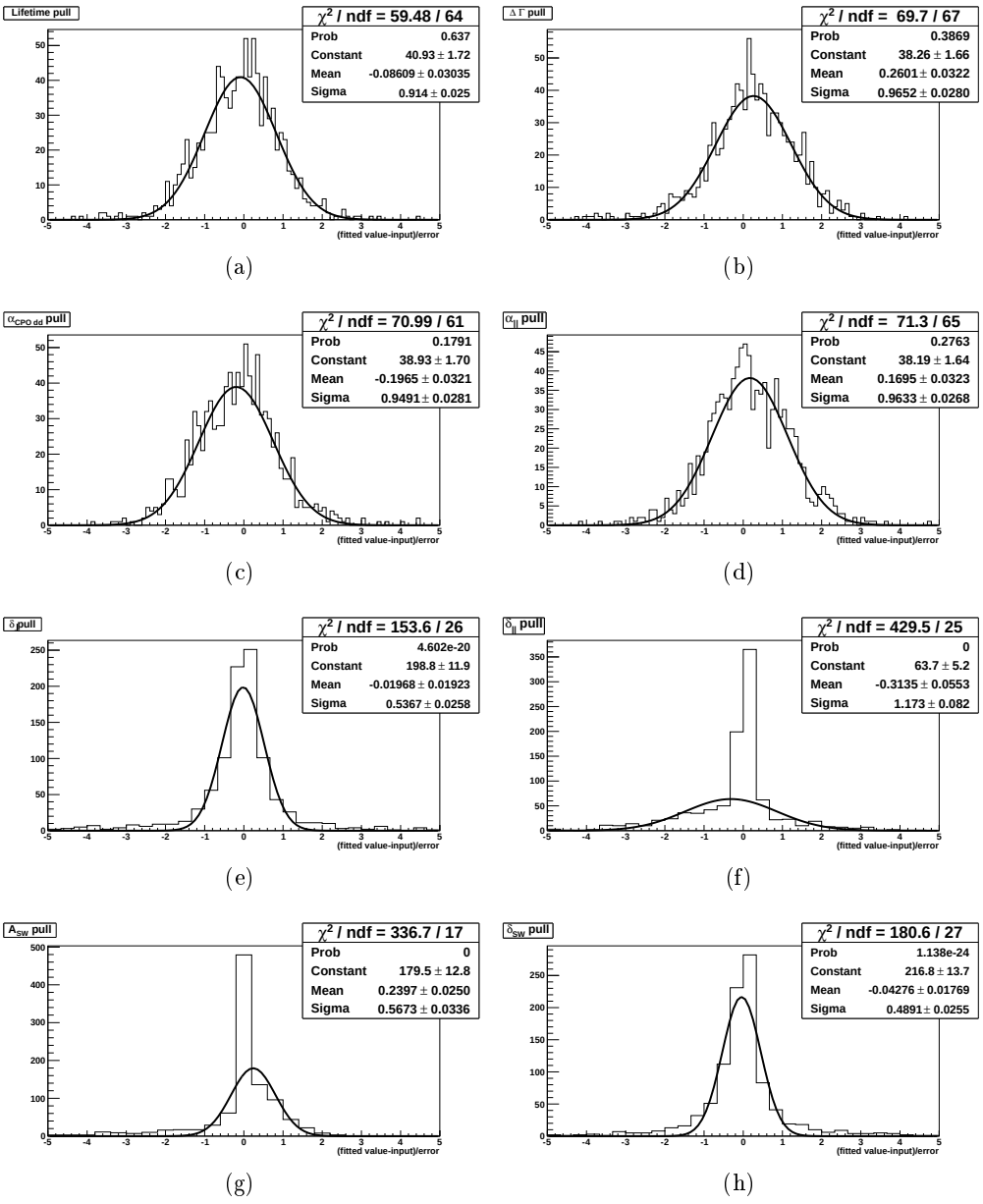


Figure 32: $10fb^{-1}$ CP-violating fit Pull distributions for the main physics parameters.

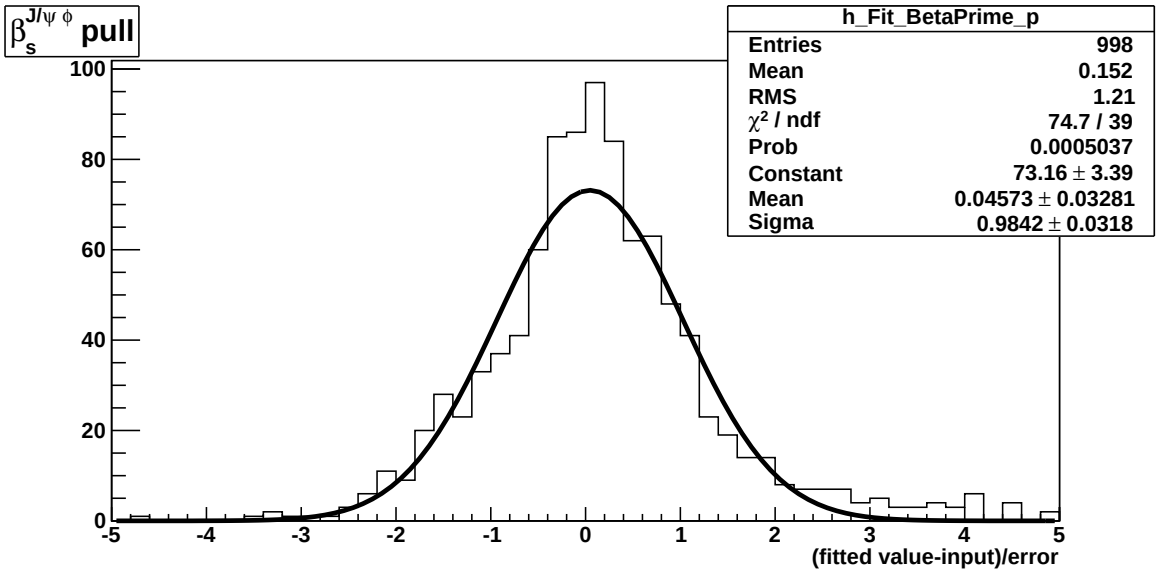


Figure 33: $\beta_s^{J/\psi\phi}$ Pull distribution

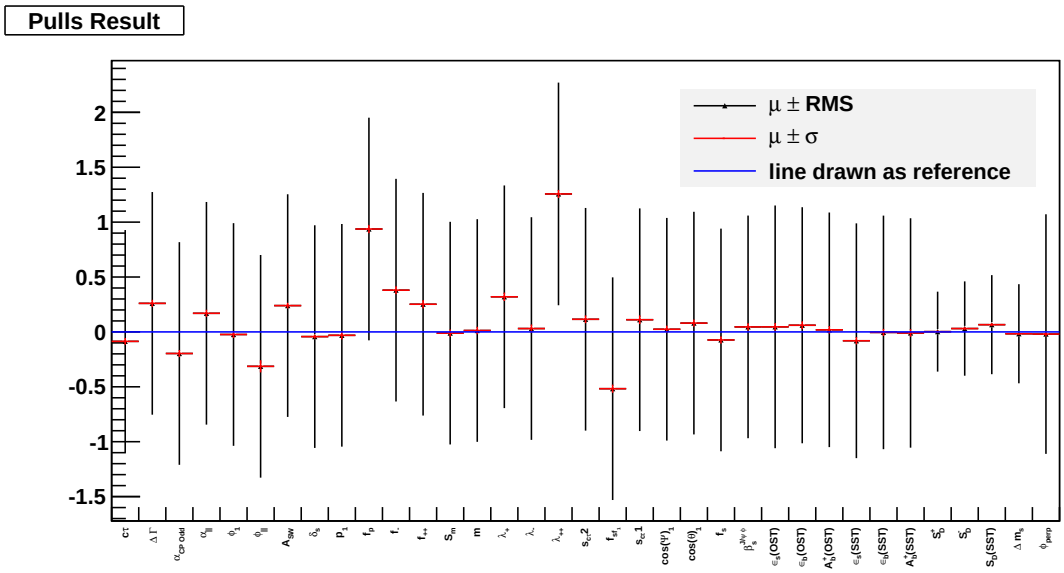


Figure 34: Summary of the pull distributions for all the variables of the likelihood function for the CP violating fit. Pseudo-experiments generated with the same statistics of the whole sample (63623 events)

10 B_s^0 lifetime, decay width difference and polarization amplitudes

In this section we report the fit results for the physics observables in the case where the CP–violating phase $\beta_s^{J/\psi\phi}$ is fixed to its SM value. We compare results from the previous CDF analysis [7] on a dataset corresponding to 5.2fb^{-1} of integrated luminosity, the same result obtained on the equivalent dataset but obtained with newer BStntuples and updated good run list, and the result corresponding to the full CDF Run II dataset in table 8. We observe very good consistency for all the results and observe a reduction in statistical precision in line with the increased statistics used in the present analysis.

Table 8: Summary of the $B_s^0 \to J/\psi\phi$ fit with $\beta_s^{J/\psi\phi}$ fixed to its SM value.

Observable	CDF 10053	5.2 fb^{-1} result	Complete Sample
$\Delta\Gamma_s$ $[ps^{-1}]$	$0.071X \pm 0.036$	$0.075 \pm 0.03X$	0.071 ± 0.026
$\alpha_\perp$	0.267 ± 0.015	0.26 ± 0.014	0.277 ± 0.011
$\alpha_\parallel$	0.309 ± 0.01	0.306 ± 0.015	0.319 ± 0.011
$\delta_\perp$ [rad]	$--$	2.95 ± 0.64	2.81 ± 0.68
$\delta_\parallel$ [rad]	2.93 ± 0.32	$--$	3.08 ± 0.35
$c\tau$ $[\mu m]$	458 ± 8	$459.X \pm 7.5$	458.6 ± 5.8
A_{SW}	0.010 ± 0.026	$--$	0.010 ± 0.026
f_s	0.181 ± 0.0024	$--$	0.1721 ± 0.0017

The fit projections in the case of $\beta_s^{J/\psi\phi}$ fixed to its SM value are reported in... (sill mising plots), and show an overall agreement of background and signal distribution with parametrization from the fit.

As in previous analyses we choose the parameter estimation for $\delta_\parallel$ is biased and we choose not to report its point estimate. A similar consideration holds for the fraction of S–wave contribution in the mass region $1.09 < m_{KK} < 1.28$. Systematic uncertainties will derived following the strategy already followed in earlier analysis, considering variation of the fit model and systematic shift related to SVX II alignment.

Preliminary results for the SM fit using the full dataset are:

$$\begin{aligned} c\tau &= 458.6 \pm 5.8\,[\mu m], \\ \Delta\Gamma_s &= 0.071 \pm 0.026\,[\mathrm{p}s^{-1}], \\ |A_\parallel(0)|^2 &= 0.230 \pm 0.013, \\ |A_0(0)|^2 &= 0.514 \pm 0.011, \\ \delta_\perp &= 2.81 \pm 0.68, \end{aligned} \tag{11}$$

where the quoted uncertainty is only statistical for the time being. These results are in good agreement with results from ohter experiments and with theoretical preditions within uncertainties.

Figure 35: ct projection for the SM fit

11 Tagged Results

The projections of the CP–violating fit are shown in fig. 36 for the transversity angles distributions, in fig. 37 for the ct distribution of the B candidates, while in fig. 23 are reported the σ_{ct} distributions. The results of the fit are reported in tab. **??**. They are in very well agreement with estimates by CP–conserving fit.

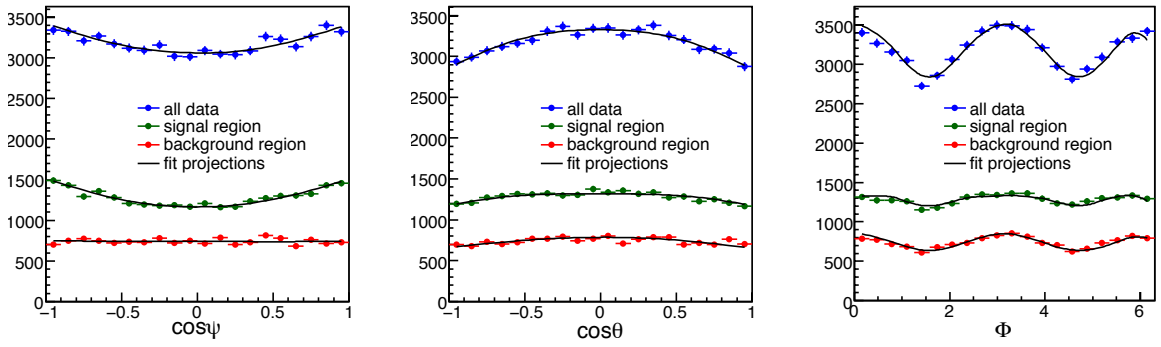


Figure 36: Angular distributions with fit projections overlaid.

Unfortunately, the pathologies observed in the pull studies and experience from the previous iterations of this analysis show that we cannot reliably use the results of the CP–violating fit to

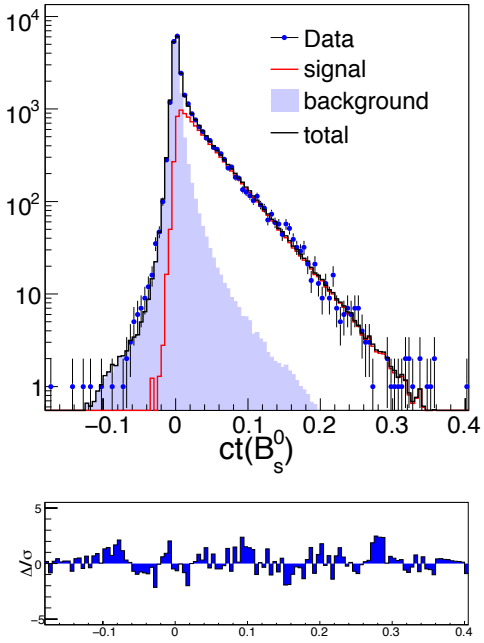


Figure 37: *ct* distribution of background along with fit projection overlaid

quote point estimates (i.e. central value ± uncertainty) for the parameters of interest. The maximum likelihood estimators show significant biases that depend on the true values of the parameters thus making it challenging any attempt at bias corrections. In addition, the estimated uncertainties are unlikely to represent actual confidence regions with the desired level of confidence. This phenomenology has been tracked to originate from a combination of the complications due to likelihood symmetries, which introduce multiple, equivalent solutions, the sensitivity to some parameters depending sensibly on the estimated values of others, and the current data size being still insufficient to approximate the asymptotic regime. As is customary in these cases, we abandon the point estimates by resorting to the Neyman construction of a fully frequentist confidence region. Being the outcome of this measurement crucially sensitive to the possible presence of non-SM physics, we believe it's a scientifically good way to operate to make sure that any evaluation of compatibility with the SM is done with the most rigorous and solid method. In any frequentist confidence region construction an arbitrariness is associated to the choice of the ordering algorithm, i.e. the procedure chosen to accumulate regions of the parameters space until the desired confidence level is attained.

Constructing correct and informative confidence regions from highly multi-dimensional likelihoods is challenging. In our case, determining the full 32-dimensional confidence space is computationally prohibitive. More importantly, the choice of the ordering algorithm is non-trivial: one needs to avoid that the projection of the region onto the $(\beta_s, \Delta\Gamma_s)$ subspace of interest includes most, if not all, of the allowed values, thus yielding a useless result. Using a standard procedure, we choose to replace the likelihood, $L(\beta_s, \Delta\Gamma_s, \vec{\nu})$ with the profile likelihood, $L_p(\beta_s, \Delta\Gamma_s, \hat{\vec{\nu}})$. For every point in the $(\beta_s, \Delta\Gamma_s)$ plane, $\hat{\vec{\nu}}$ are the values of nuisance parameters that maximize the likelihood. The profile-likelihood ratio $-2\Delta\ln(L_p)$ is then used as a χ^2 variable to derive confidence regions in the two-dimensional space $(\beta_s, \Delta\Gamma_s)$. However, the simulation shows that the observed distribution of $-2\Delta\ln(L_p)$ deviates from the $\chi^2(2)$ one. Specifically, the resulting confidence regions contain true

Parameter		Description	Full Dataset
MsFs			0.167523 ± 0.00262896
MsMass			5.36637 ± 0.000147382
MsSigma			$0.00868373 \pm 0.00015376$
MsSlope			-1.8103 ± 0.421389
ASWave	A_{SW}	fraction of $S-wave$ KK component in signal	0.009011 ± 0.028276
AlphaCPOdd	$\alpha_\perp$	CP odd fraction	0.277145 ± 0.010713
AlphaPara	$\alpha_\parallel$	fraction in CP even states	0.309873 ± 0.011472
BGmslope	p_1	mass background slope	-1.888269 ± 0.419769
BetaPrime	$\beta_s{}^{J/\psi\phi}$	CP asymmetry parameter	0.112602 ± 0.123043
DScale	S_D(OST)	OST dilution scale factor +	1.088967 ± 0.049877
DScale21	S_D(SST)	SST dilution scale factor	0.863540 ± 0.183537
DeltaS	δ_{SW}	Relative phase of $S-wave$ KK component	1.225665 ± 1.105874
EffBkg	ε_b(OST)	OST tagging efficiency for background	0.758875 ± 0.001897
EffBkg2	ε_b(SST)	SST tagging efficiency for background	0.724188 ± 0.002001
EffBkgp	A^+(OST)	OST background positive tag asymmetry	0.497566 ± 0.002553
EffBkgp2	A^+(SST)	SST background positive tag asymmetry	0.495643 ± 0.002599
EffSig	ε_s(OST)	OST tagging efficiency for signal	0.929345 ± 0.002971
EffSig2	ε_s(SST)	SST tagging efficiency for signal	0.520942 ± 0.005264
Fgauss	f_p	Prompt fraction of background	0.883513 ± 0.003871
Fm	f_-	Fraction of bkg which decays w/λ_-	0.209293 ± 0.037229
Fpp	f_{++}	Fraction of bkg which decays w/λ_{++}	0.716578 ± 0.039212
FracSig	f_s	Signal Fraction	0.172141 ± 0.001747
Lambdam	λ_-	Effective background lifetime, neg. comp.	0.040342 ± 0.003190
Lambdap	λ_+	Effective background lifetime, pos. comp. 1	0.043800 ± 0.003727
Lambdapp	λ_{++}	Effective background lifetime, pos. comp. 2	0.013346 ± 0.001014
Lifetime	$c\tau$	average of $c\tau_H$ and $c\tau_L$	0.045820 ± 0.000584
MassScl	s_m	Mass error scale factor	1.727770 ± 0.016735
Nevents	N_{events}	Number of signal events	$10952.154811 \pm 111.148470$
Phi1	ϕ_1	First parameter in bkg fit to ϕ	0.144261 ± 0.006262
PhiPara	$\phi_\parallel$	arg($A_\parallel A_0$) asymmetry parameter	3.089784 ± 0.379158
PhiPerp	$\phi_\perp$	arg($A_\perp A_0$) asymmetry parameter	2.767288 ± 0.728939
ScaleFac	$s_{c\tau}1$	Lifetime error scale factor 1	1.307769 ± 0.012070
ScaleFac2	$s_{c\tau}2$	Lifetime error scale factor 2	3.342624 ± 0.127260
ScaleFrac	f_{sf_1}	fraction of 1st lifetime error scale factor	0.851428 ± 0.010281
cosPsi1	$\cos(\psi)_1$	First parameter in bkg fit to cos(ψ)	0.169611 ± 0.013019
deltaG	$\Delta\Gamma$	CP asymmetry parameter [ps^{-1}]	0.071177 ± 0.026734
deltaM	Δm_s	$B^0{}_s$ mixing frequency	17.728119 ± 0.127649

Table 9: *Flavor tagged fit results with β_s floating in the whole CDF RunII dataset*

values of the parameters with lower probability than the nominal confidence level (C.L.) because the observed $-2\Delta\ln(L_p)$ distribution has longer tails than a χ^2. In addition, the $-2\Delta\ln(L_p)$ distribution appears to depend on the true values of the nuisance parameters, which are unknown. We use therefore the simulation of a large number of pseudoexperiments to derive the actual distribution of $-2\Delta\ln(L_p)$. The effect of systematic uncertainties is accounted for by randomly sampling a limited number of points in the space of all nuisance parameters and using the most conservative of the resulting profile-likelihood ratio distributions to calculate the final confidence level.

We first fit the data with all parameters floating. Then, for each point in a $XX \times XX$ grid on the $(\beta_s\Delta\Gamma_s)$ plane we fit the data by floating all parameters but β_s and $\Delta\Gamma_s$, which are fixed to the values corresponding to the probed point. Twice the negative difference between the logarithms of

the likelihood obtained in each of the two steps provide a profile-likelihood ratio value $(\beta_s\Delta\Gamma_s)$ for each of the points in the $(\beta_s\Delta\Gamma_s)$ plane. To map the observed values of $-2\Delta\ln(L_p)$ into confidence levels, we need to compare them with the expected distribution of $-2\Delta\ln(L_p)$. We obtain these distributions by generating 16 ensembles of 1000 pseudoexperiments each. In each ensemble, the true values of β_s and $\Delta\Gamma_s$ correspond to the probed point, while the true values of the nuisance parameters are a random sampling from an hypercube centered at their best fit values in data, with side corresponding to 10 standard deviation. As we do not know the true values for these parameters we ensure coverage over a wide range of possible values, but always within their physically allowed range. Profile-likelihood ratios are determined for each of these pseudoexperiment exactly as for data. The ensemble giving the broadest $-2\Delta\ln(L_p)$ distribution is chosen. For each point in the $(\beta_s\Delta\Gamma_s)$ grid, we calculate the p-value as the fraction of pseudoexperiments from this ensemble in which a $-2\Delta\ln(L_p)$ value as large or larger than in data is observed. The $(\beta_s\Delta\Gamma_s)$ region where the p-value is larger than $1-x$ forms the x% CL region. In practice we observe that the $-2\Delta\ln(L_p)$ distribution is fairly independent of the value of $(\beta_s\Delta\Gamma_s)$ probed, so we don't need to generate pseudoexperiments for each $(\beta_s\Delta\Gamma_s)$ point. It suffices to compare the $-2\Delta\ln(L_p)$ observed in data for each point to just the $-2\Delta\ln(L_p)$ distribution generated in a single point. Because the main goal of this analysis is to quantify compatibility of our data with the SM, we choose the SM value $(\beta_s = 0.02, \Delta\Gamma_s = 0.090)$ to generate the reference $-2\Delta\ln(L_p)$ distribution. An idea of the deviation of the observed profile-likelihood ratio distribution from the expected $\chi^2(2)$ distribution is shown in Fig.XX. Including the coverage adjustment and the effect of systematics uncertainties we need to change the value of $-2\Delta\ln(L_p)$ by approximately CC units so that projections on the $(\beta_s\Delta\Gamma_s)$ plain contains the true values with 95% CL, compared with the nominal value of 5.99.

11.1 Tagged Contour

We present the likelihood–ratio (LLR) profile as confidence regions in the β_s–$\Delta\Gamma$ plane in fig. 41. Because of the approximate symmetries of the likelihood, it is knwon that MINUIT can fall sometimes in a local minimum during the minimization of the likelihood function. Therefore, we compute two LLR profiles for each point of the plane: the first has starting points $\Delta\Gamma > 0$ and $\delta_{\|} < \pi$ (0 and π are symmetry points for those parameters), the second has reversed starting points $\Delta\Gamma < 0$ and $\delta_{\|} > \pi$. This should allow to have the global minima at least in one of the two profiles, without imposing any boxing on the fitting parameters. Thus, for each point of the plane, we choose the deepest minima found in the two profiles as the right minima to have. In fig. 39 we present also the comparison with the LLR profile done for the fit without the S–wave component of the Likelihood.

Following the procedure described in sect. **??**, the map between p–value (1 -C.L.) as a function of LLR is evaluated with a thousand of pseudoexperiments, generated at the SM point. That is reported in fig. 40. We found that a LLR of 2.75 must be set to guarantee the correct coverage for a 68% probability C.L., while a LLR of 7.07 corresponds to 95% C.L. The adjusted confidence regions are reported in fig. 41. The SM point has a p–value of 0.87 which corresponds to a discrepancy of our data less than 1σ.

11.2 Tagged Intervals

The LLR profile is evaluated also as a function of β_s only by the same method previously described. It is reported in fig. 42 when no adjustement for coverage is applied. In fig. 43 the LLR profile

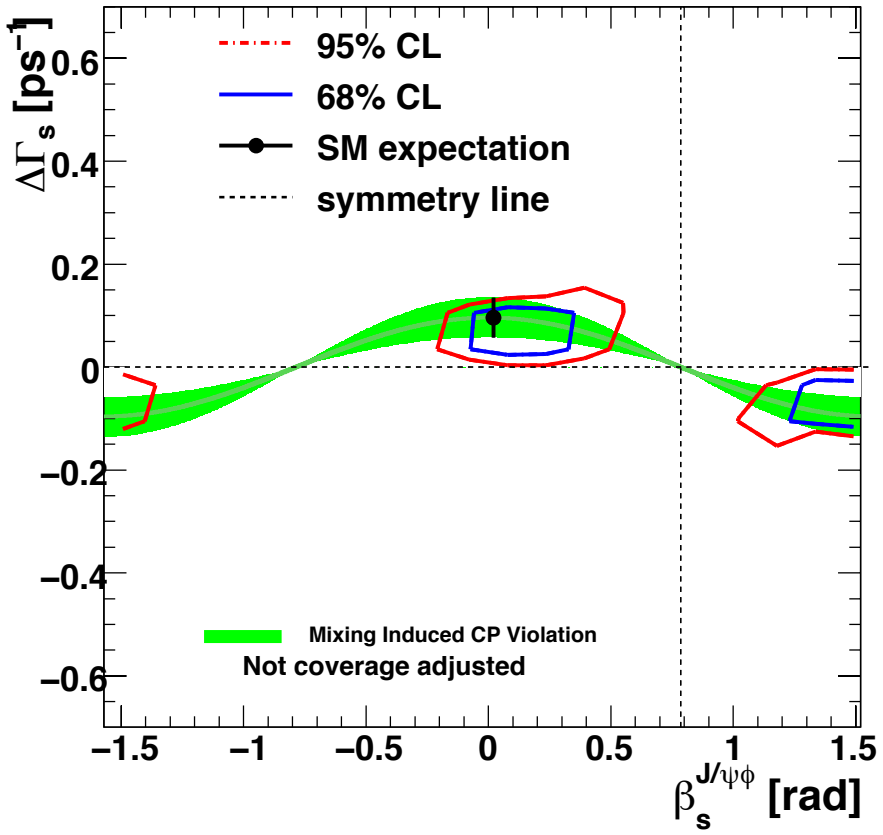


Figure 38: LLR profile as confidence regions in the β_s–$\Delta\Gamma$ plane before any adjustement.

for the fit without the S–wave is shown. Following analogous procedure as for the two-dimensional case, we built the map between LLRs and p–values to ensure right coverage properties to our C.L. intervals using a thousand of pseudo experiments (see fig. 44). After the adjustment, the new intersection with the LLR profiles corresponding to a 68% C.L. interval and to a 95% C.L. interval are reported in fig. 45. We found $\beta_s^{J/\psi\phi} \in [-\pi/2, -1.54] \bigcup [-0.03, 0.27] \bigcup [1.29, \pi/2]$ at 68% C.L., while $\beta_s^{J/\psi\phi} \in [-\pi/2, -1.38] \bigcup [-0.19, 0.47] \bigcup [1.10, \pi/2]$ at 95% C.L.

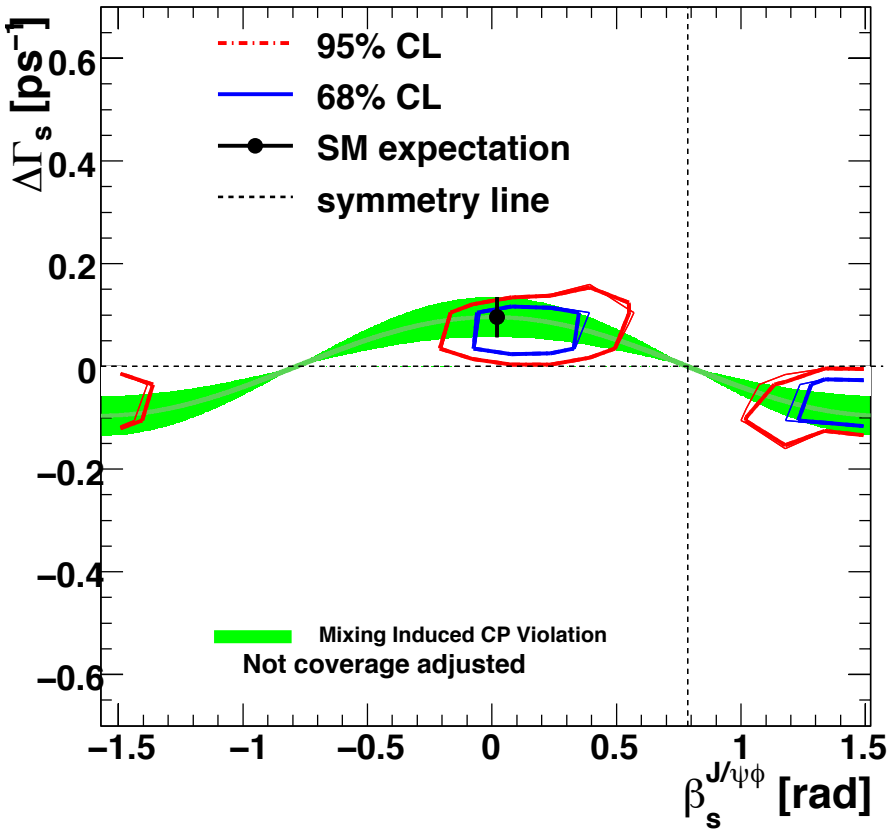


Figure 39: LLR profile as confidence regions in the β_s–$\Delta\Gamma$ plane for the fit without S–wave (light line), compared with the default fit (bold line). No adjustement is applied.

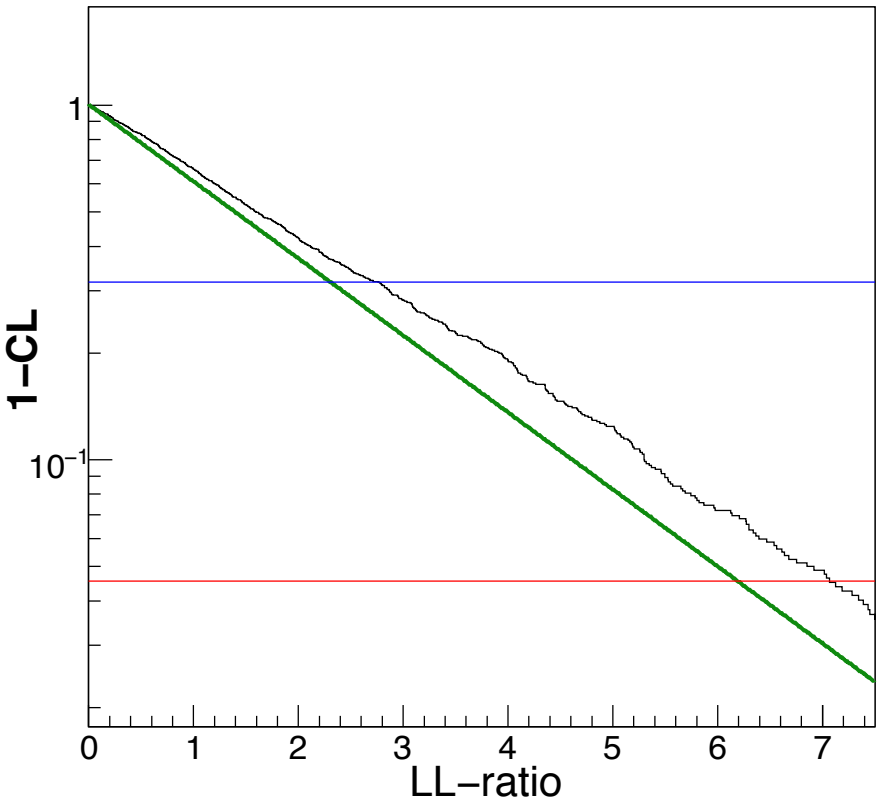


Figure 40: Map between p–value (1 -C.L.) as a function of LLR for the coverage adjustment of β_s–$\Delta\Gamma$ confidence regions.

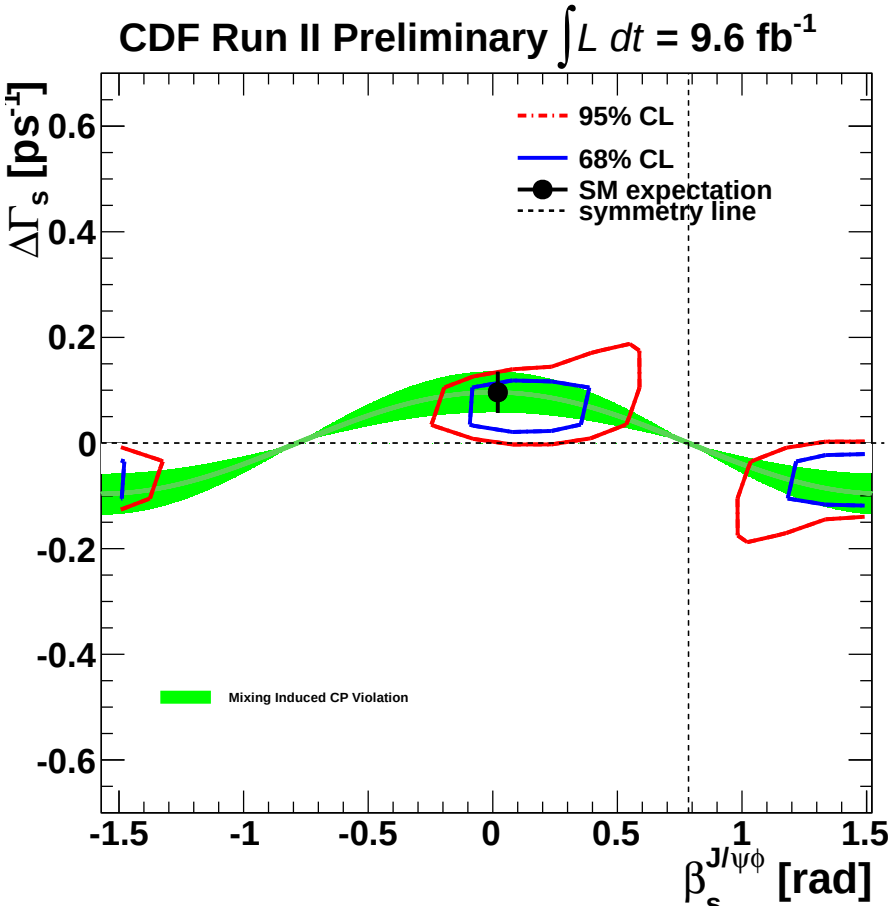


Figure 41: Coverage–adjusted confidence regions in the β_s–$\Delta\Gamma$ plane

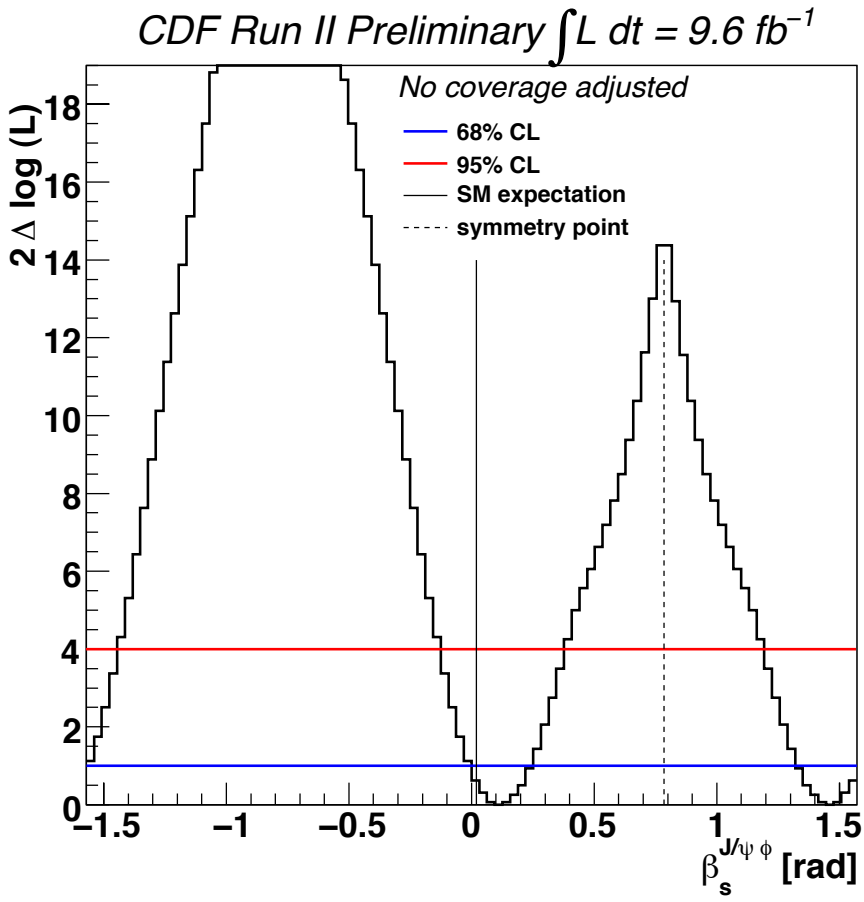


Figure 42: LLR profile for β_s before any adjustement.

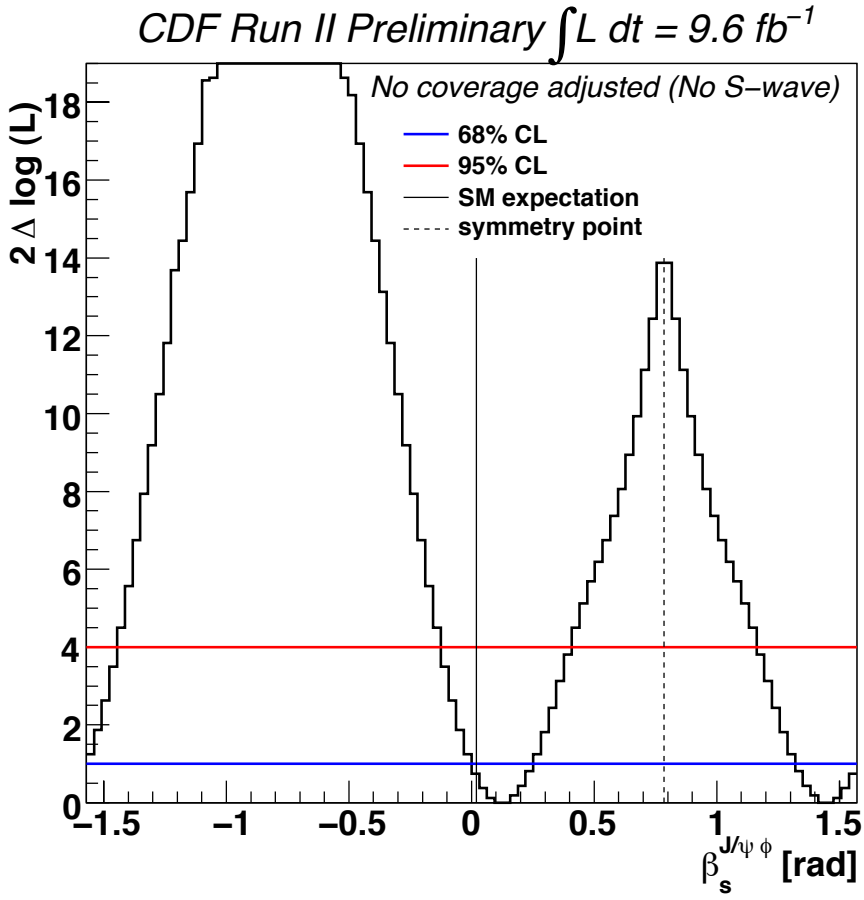


Figure 43: LLR profile for β_s for the fit without S–wave.

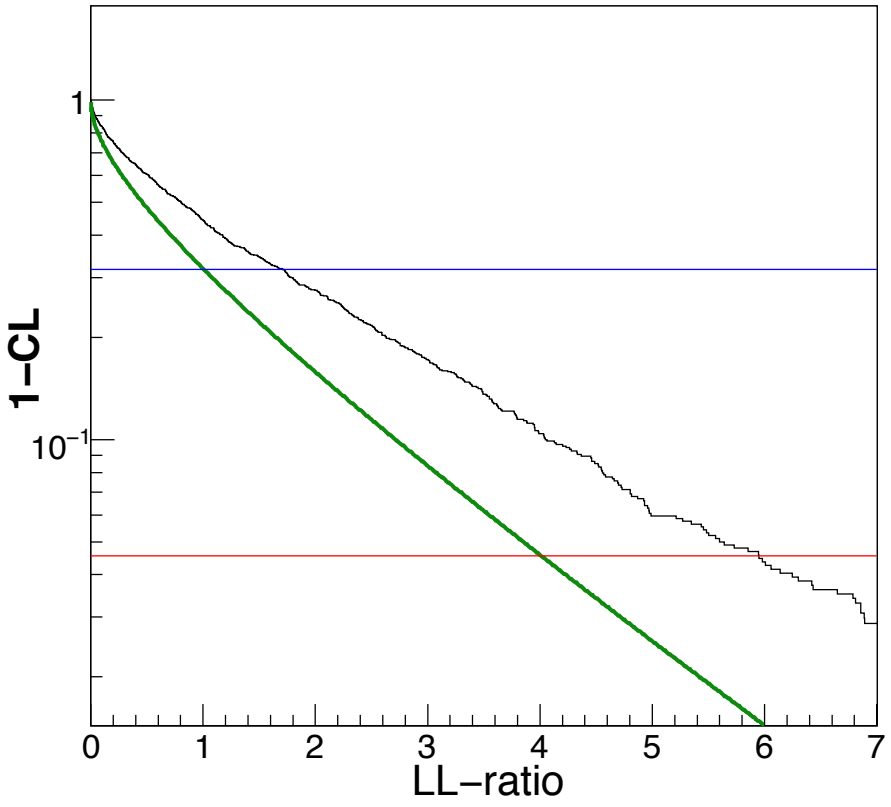


Figure 44: Map between p–value (1 -C.L.) as a function of LLR for the coverage adjustment of β_s confidence intervals.

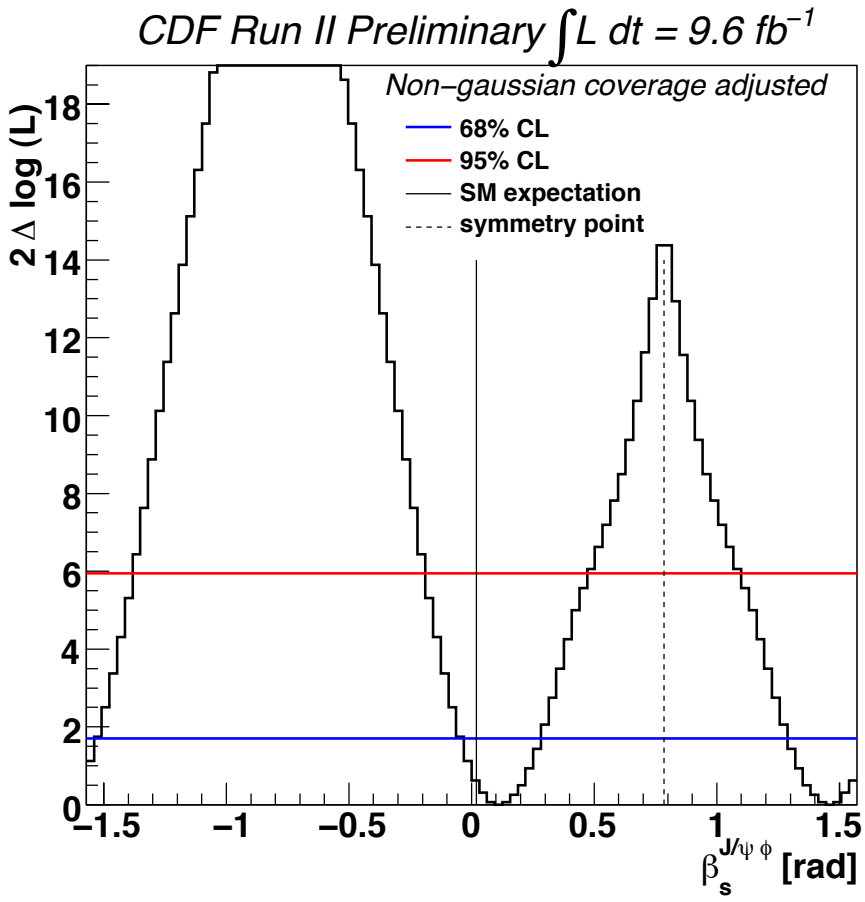


Figure 45: Coverage–adjusted confidence regions in the β_s–$\Delta\Gamma$ plane

12 Conclusions

We reported an updated measurement of the CP-violating phase $\beta_s^{J/\psi\phi}$ using flavor-tagged $B_s^0 \to J/\psi\,\phi$ decays in 9.6 fb^{-1} of integrated luminosity collected by the dimuon trigger, which corresponds to the full CDF Run II dataset. Using an analysis technique and tools largely inherited from the previous (5.2 fb^{-1}) analysis, we reconstruct approximately 11 000 $B_s^0 \to J/\psi\,\phi$ signal events. The opposite side tagging algorithms are calibrated using using 81 000 $B^+ \to J/\psi K^+$ decays reconstructed in the same dataset. The same side tagging algorithms are not re-calibrated and thus used in only half of our dataset. The CP-violating phase is found to be in the range $\beta_s^{J/\psi\phi} \in [-\pi/2, -1.54] \bigcup [-0.03, 0.27] \bigcup [1.29, \pi/2]$ [STAT ONLY] at the 68% confidence level, in agreement with the standard model expectation. Assuming CP conservation ($\beta_s^{J/\psi\phi}$=0.0) we also determine the mean B_s^0 lifetime, $\tau_s = 1.528 \pm 0.019$ (stat) ps; the width difference between heavy and light mass eigenstates, $\Delta\Gamma_s = 0.071 \pm 0.026$ (stat) ps^{-1}; and the transversity amplitudes, $|A_0(0)|^2 = 0.514 \pm 0.011$ (stat), $|A_\parallel(0)|^2 = 0.230 \pm 0.013$ (stat). The results are amongst the most precise from a single experiment and consistent with previous determinations and world's average results.

Acknowledgments

We express our special gratitude to all the authors of the previous iterations of this analysis, and in particular Louise Oakes, Elisa Pueschel, Joe Boudreau, Farrukh Azfar, Manfred Paulini, G. Giurgiu, K. Gibson, Thomas Kuhr, Michal Kreps. Without their generous help the update of this analysis in such a timely fashion would never had been possible.

References

[1] A. Abulencia et al. Observation of B_s^0-$\bar{B}_s^0$ oscillations. *Phys. Rev. Lett.*, 97:242003, 2006.

[2] T. Aaltonen et al. First Flavor-Tagged Determination of Bounds on Mixing- Induced CP Violation in $B_s \to J/\psi\phi$ Decays. *Phys. Rev. Lett.*, 100:161802, 2008.

[3] V.M. Abazov et al. Measurement of B_s^0 mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$. *Phys.Rev.Lett.*, 101:241801, 2008.

[4] Manfred Paulini. Properties of Heavy B Hadrons. *Int.J.Mod.Phys.*, A24:4413–4435, 2009.

[5] Victor Mukhamedovich Abazov et al. Evidence for an anomalous like-sign dimuon charge asymmetry. *Phys.Rev.Lett.*, 105:081801, 2010.

[6] Victor Mukhamedovich Abazov et al. Measurement of the anomalous like-sign dimuon charge asymmetry with 9 fb$^-$1 of p pbar collisions. *Phys.Rev.*, D84:052007, 2011.

[7] F. Azfar *et al.*, "An Update of the Measurement of the CP-Violating Phase β_s using Bs to $J/\psi\phi$ Decays", CDF note 10053.

[8] CDF Collaboration, "Measurement of the CP–violating Phase $\beta_s^{J/\psi\phi}$ in $B_s^0 \to J/\psi\phi$ Decays with the CDF II Detector",CDF note 10586, to be submitted to PRD.

[9] Victor Mukhamedovich Abazov et al. Measurement of the CP-violating phase $\phi_s^{J/\psi\phi}$ using the flavor-tagged decay $B_s^0 \to J/\psi\phi$ in 8 fb^{-1} of $p\overline{p}$ collisions. 2011. Long author list - awaiting processing.

[10] LGCb Collaboration, Report No. LHCb–CONF–2011-049 (2011).

[11] Bs Mixing Group, "Optimization of the Same Side Kaon tagging algorithm combining PID and kinematic variables", CDF note 8344.

[12] M. Feindt, T. Kuhr, M. Kreps, J. Morlock, and A. Schmidt, "Public note for the calibration of the same side kaon tagger using Bs mixing", *CDF Note 10108*, 2006.

[13] J. Boudreau, J.P. Fernandez, K. Gibson, G. Giurgiu, G. Gomez-Ceballos, L. Labarga, C. Liu, P. Maksimovic, M. Paulini, "Measurement of the CP Violation Paramter beta' in Bs to Jpsi Phi", CDF note 8960.

[14] J. Boudreau, J.P Fernandez, I. Furic, K. Gibson, G.P. di Giovanni, G. Giurgiu, G. Gomez-Ceballos, L. Labarga, Ch. Liu, Kh. Makhoul, P. Maksimovic, M. Paulini, Ch. Paus, A. Savoy-Navarro, "Study of the time evolution of flavor-tagged Bs mesons flavor-tagged Bs mesons", CDF note 9090.

[15] http://www-cdf.fnal.gov/tiki/tiki-index.php?page=BStntuple.Status.

[16] M. Feindt, "A Neural Bayesian Estimator for Conditional Probability Densities", arXiv:physics/0402093.

[17] Gavril Giurgiu. B flavor tagging calibration and search for b(s) oscillations in semileptonic decays with the cdf detector at fermilab. *Ph.D. thesis, Carnegie Mellon Univ., FERMILAB-THESIS-2005-41*, 2005.

[18] Michael Milnik. Measurement of the lifetime difference and cp-violating phase in $B_s \to J/\psi\phi$ decays. *Ph.D. thesis, FERMILAB-THESIS-2007-38*, 2007.

[19] K. Nakamura *et al.* Review of Particle Physics. *J. Phys. G*, 37:075201, 2010.

[20] G. Giurgiu *et al.*, "Updated Likelihood Muon tagger", *CDF Note 7644*, 2005.

[21] V. Tiwari *et al.*, "Likelihood Based Electron Tagging", *CDF Note 7121*, 2004.

[22] G. Bauer *et al.*, "Improved Jet charge Tagger for summer conferences 2004", *CDF Note 7131*, 2004.

[23] G. Salamanna *et al.*, "Opposite Side Kaon Tagging", *CDF Note 8179*, 2006.

[24] R. Blair et al. (CDF-II Collaboration), *FERMILAB-Pub-96/360-E.*

[25] F. Azfar *et al.*, "An Updated Measurement the CP Violating Phase β_s using $B_s \to J/\psi\phi$ Decays", *CDF Note 9395*, 2008.

[26] 2008. The CDF Collaboration, "An Updated Measurement of the CP Violating Phase β_s using $B_s^0 \to J/\psi\phi$ Decays", CDF Public Note 9458.

[27] M. Feindt, M. Kreps, T. Kuhr, M. Milnik, "Measurement of DeltaGamma and deltaphi in untagged Bs → Jpsi phi and Bd → Jpsi K* decays", CDF note 8753.

[28] F. Azfar, J. Boudreau, N. Bousson, J.P. Fernandez, K. Gibson, et al. Formulae for the Analysis of the Flavor-Tagged Decay $B_s^0 \to$ Jpsi phi. *JHEP*, 1011:158, 2010.

[29] Isard Dunietz, Robert Fleischer, and Ulrich Nierste. In pursuit of new physics with b_s decays. *Phys. Rev. D*, 63:114015, 2001.