BaBar Analysis Document #137, Version 07

Measurement of the B^0 Lifetime and Oscillation Frequency Using Exclusively Reconstructed Semileptonic Decays

Chih-Hsiang Cheng, David Kirkby*,
Chris LeClerc, T.I. Meyer, Patricia R. Burchat

August 4, 2002

Abstract

We measure the B^0 lifetime and oscillation frequency with a sample of approximately 14,000 exclusively reconstructed $B^0 \to D^{*-}\ell^+\nu_\ell$ signal events. The sample selection and background characterization are described in BABAR Analysis Document 147. We use a total of about 68,000 events in 12 signal and background control samples. The lifetime and oscillation frequency are measured simultaneously with an unbinned likelihood fit that uses, for each event, the measured decay time difference Δt, the calculated uncertainty on Δt, the signal and background probabilities, and tagging information for the other B. The parameters that are floating in the fit are the lifetime, mixing frequency, mistag probabilities, signal Δt resolution parameters, background Δt model parameters and the fraction of B^+ decays in the peak of the $m(D^*) - m(D^0)$ mass-difference distribution. The results are

$$\tau_{B^0} = (1.523\ ^{+0.024}_{-0.023} \pm 0.022)\ \text{ps}$$

and

$$\Delta m_d = (0.492 \pm 0.018 \pm 0.013)\ \text{ps}^{-1},$$

where the first error is statisitcal and the second systematic. The statisitcal correlation coefficient between τ_{B^0} and Δm_d is -0.22.

*primary contact

Contents

1 Introduction **3**

2 Sample Selection and Characterization **4**
2.1 Data Sample 4
2.2 Monte Carlo Samples 4
2.3 Event Selection 5
2.4 Background Characterization 5
2.5 Event Selection Bias 7

3 Decay-Time Measurement **9**
3.1 Measurement of Δz 9
3.1.1 $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ z position 9
3.1.2 z position of the tagging B 10
3.2 Conversion from Δz to Δt 11
3.3 Calculation of Error on Δt 12
3.4 Resolution Models 12
3.5 Resolution in Signal Monte Carlo 18

4 Flavor Tagging **29**
4.1 Tagging Categories 29
4.2 Tagging Performance 30
4.3 Correlation Between Mistag and Δt Error 31

5 Fit Method **33**
5.1 Parameter Notation 36
5.2 Signal Δt Model 38
5.2.1 Fit to Monte Carlo Signal Events 38
5.3 Combinatoric Background Δt Model 44
5.4 Peaking Background Δt Models 52
5.4.1 Continuum Background 52
5.4.2 Fake Lepton Background 56
5.4.3 Uncorrelated Lepton Background 61
5.4.4 $\mathrm{B}^+ \to \mathrm{D}^{*-}\ell^+\mathrm{X}$ 68

6 Fit Results **70**
6.1 Signal Outliers in Data 70
6.2 Results of Master Model Fit to Data 72
6.3 Correction on Statistical Errors 76
6.4 Fit Result Summary 77

7 Validation and Toy MC Studies **82**
7.1 Validation Fits to Generic Monte Carlo Samples . . . 82
7.2 Correction Based on Generic Monte Carlo Sample . . . 89
7.3 Change in Kaon Mistag Slope . . . 89
7.4 Toy Monte Carlo Studies . . . 90
7.5 Comparison with Hadronic Mixing Analysis . . . 91

8 Cross Checks **94**
8.1 Fit to MC truth . . . 94
8.2 True vs Reconstructed Flavor Tagging . . . 95
8.3 Δm_d from Δt-Shape or Flavor Information Only . . . 95
8.4 Subsamples by Tagging Category . . . 96
8.5 Subsamples by B^0 Flavor . . . 96
8.6 Subsamples of Signal MC . . . 97
8.7 Subsamples by D^0 Decay Mode . . . 97
8.8 Sensitivity to NT3 tagging information . . . 97
8.9 Varying Δt Selection Criterion . . . 99
8.10 Varying $\sigma_{\Delta t}$ Criterion . . . 100
8.11 Fixing τ_{B^0} or Δm_d to PDG2000/2002 values . . . 102
8.12 Fitting with Different Resolution Models for signal MC . . . 103
8.13 Mistag rate – $\sigma_{\Delta t}$ correlation . . . 103
8.14 Varying $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ Upper Limit . . . 103
8.15 Summary . . . 106

9 Systematic Studies **111**
9.1 Event Selection Bias and Monte Carlo Correction . . . 111
9.2 Fixed Outlier Scale . . . 111
9.3 z Scale . . . 111
9.4 PEP-II Boost Uncertainty . . . 112
9.5 B^+ Mistag Rate, Lifetime and Fraction . . . 112
9.6 SVT Alignment . . . 113
9.6.1 Estimate of bias using signal MC . . . 114
9.6.2 Consistency in Data between different alignments . . . 117
9.7 Beam Spot Position . . . 117
9.8 Background Fractions . . . 121
9.9 Fitting with Different Resolution Model . . . 124
9.10 Background Δt Models . . . 124
9.11 Dependence on Known PID Efficiencies . . . 124
9.12 Summary . . . 124

10 Closure **126**

A Filtered Generic Monte Carlo **127**

B Criteria to Remove Outliers **134**

C Correlation Between $\sigma_{\Delta t}$ and Bias in Resolution Function **137**

D Study of Δt Resolution for Signal Events **139**

E Combinatoric Background Composition and Parameter Breakdown **153**
E.1 Characterization of Combinatoric Background in Monte Carlo 153
E.2 Motivation for Combinatoric Δt Model from Monte Carlo 162
E.3 Combinatoric Δt Model Motivated by Data 165

F $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ Vertex Constraints **181**

G Extraction of τ_{B^0} and Δm_d Using Only Δt or Only Flavor Information **187**

H Correlation Coefficients between Physics Parameters and Other Floating Parameters **187**

I Additional Δt-projection Plots for Fit Result to Data **189**

1 Introduction

We describe a measurement of the B^0 lifetime and oscillation frequency with a sample of approximately 14,000 fully reconstructed $B^0 \to D^{*-}\ell^+\nu_\ell$ signal events. The lifetime τ and oscillation frequency Δm_d are measured simultaneously with an unbinned likelihood fit that uses, for each event, the measured decay time difference Δt, the calculated uncertainty on Δt, the combinatorial and peaking background probabilities, and tagging information from the rest of the event. The resulting statistical uncertainties are approximately 0.02 ps for the lifetime measurement and 0.02 ps^{-1} for the oscillation frequency. For comparison, the PDG2002 [17, 18] uncertainty on the world average is 0.016 ps for τ_{B^0} and 0.009 ps^{-1} for Δm_d. The challenge in this analysis is to reduce the systematic uncertainties so that they do not dominate the total uncertainty, which in turn means understanding the types and sizes of backgrounds, the Δt resolution function and mistag probabilities for signal, and the Δt distributions for combinatorial and peaking backgrounds.

There are several reasons why we choose to fit the B^0 lifetime and oscillation frequency simultaneously.

1. The expected precision for both τ and Δm_d is comparable to that of the PDG2002 world average. Therefore, there is no benefit from fixating on the world average measurement of either. The resolution on Δt leads to a correlation between τ and Δm_d. We will study the correlations between τ and Δm_d, as well as between τ and $\chi_d = \Delta m_d/\tau$. Our choice of physics parameters to quote as a primary result of the analysis will be based on minimizing correlations and systematic uncertainties.

2. Mixing information improves the measurement of τ in two distinct ways:

   (a) Since mixed and unmixed events have different Δt distributions, the mixing information for each event gives greater sensitivity to the Δt resolution function and a smaller statistical uncertainty on τ. Our current estimate is that the statistical uncertainty on τ decreases by about 15%.

   (b) Since $B^0\overline{B^0}$ and B^+B^- events have different mixing behaviour, we can use the Δt distributions for mixed and unmixed events to help discriminate between $B^0\overline{B^0}$ signal events and peaking B^+B^- background events.

3. The use of untagged events even for the mixing measurement increases our sensitivity to the resolution function.[1]

This document summarizes the sample selection and characterization, which is described in detail in BABAR Analysis Document 147, and reviews the algorithms used

[1] In BABAR, so-called "untagged" events are in the category `NT3`. As described later, we use the tagging information for these events.

to measure Δt and tag the flavor of the other B. Most of the document is dedicated to studies of the resolution function for Δt based on Monte Carlo and data samples, for both signal and backgrounds, a description of the unbinned likelihood fit, validation exercises, and estimates of systematic uncertainties.

2 Sample Selection and Characterization

The event selection for $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ is described in detail in BAD 147. This section summarizes the event selection and background characterization, and presents a Monte Carlo study of possible bias to the lifetime or mixing frequency introduced by the selection criteria.

2.1 Data Sample

The data sample consists of 20.6 fb^{-1} collected on-peak and 2.6 fb^{-1} collected $\approx$40 MeV below the $\Upsilon(4S)$ resonance, during 1999-2000. The data was processed with the Release 8 reconstruction software. The `BToDstarlnuLoose` skim was used.

2.2 Monte Carlo Samples

We use all available SP3 $\mathrm{B}^0\overline{\mathrm{B}^0}$, $\mathrm{B}^+\mathrm{B}^-$, $\mathrm{c}\overline{\mathrm{c}}$, and $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ signal Monte Carlo samples generated and reconstructed with release 8.8.0x. The generic Monte Carlo samples each correspond to about 15 fb^{-1} equivalent luminosity for the $\mathrm{B}^+\mathrm{B}^-$ and $\mathrm{B}^0\overline{\mathrm{B}^0}$ samples, and about 9 fb^{-1} for the $\mathrm{c}\overline{\mathrm{c}}$.

In order to increase the size of our generic Monte Carlo sample, a special "filtered" sample was produced. In this "filtered" sample, a true lepton with center-of-mass momentum greater than 1 GeV/c is required at the generator level. These samples provide an additional 10.4 fb^{-1} $\mathrm{B}^0\overline{\mathrm{B}^0}$ and 16.4 fb^{-1} $\mathrm{B}^+\mathrm{B}^-$ equivalent luminosity. See Appendix A for details.

The signal Monte Carlo samples correspond to events in which one B decays to $\mathrm{D}^{*+}\ell^-\bar{\nu}$ or $\mathrm{D}^{*-}\ell^+\nu$, with $\mathrm{D}^{*+} \to \mathrm{D}^0\pi^+$. Separate samples are created for four different D^0 decay modes. The samples consist of 424k $\mathrm{K}\pi$ (161.3 fb^{-1}), 436k $\mathrm{K}\pi\pi\pi$ (84.9 fb^{-1}), 654k $\mathrm{K}\pi\pi^0$ (68.7 fb^{-1}) and 102k $\mathrm{K_S}\pi\pi$ (55.2 fb^{-1}).

Note that since the equivalent luminosities for the various D^0 decay modes are different, the relative fractions of the different decay modes are not realistic in studies that use all of the signal Monte Carlo.

The generator value of τ is 1.548 ps and of Δm_d is 0.472 ps. In generic Monte Carlo, the ratio of mixed to unmixed events (parameterized by χ_d) is consistent with these values of τ and Δm_d. However, signal Monte Carlo samples were generated with a value of $\chi_d = 0.160$, which is inconsistent with the value of $\chi_d = 0.174$ that corresponds to the generated values of τ and Δm_d. We have corrected this effect for all studies described in this document by rejecting a precise fraction (9.5785%) of the unmixed events. (See the presentation by Tim Meyer at the April 2001 Collaboration Meeting for more details[3].)

We rely on Monte Carlo truth-matching in some proof-of-principle studies for this analysis. We often refer to the "correctly reconstructed signal MC sample" which is the set of signal MC event whose $B^0 \to D^{*-}\ell^+\nu_\ell$signal candidates pass the Monte Carlo truth matching test. This sample therefore has realistic tagging-side performance characteristics, but the best possible resolution on the reconstructed side.

2.3 Event Selection

The event selection criteria are listed in BAD 147, V4, Sec. 2. The most relevant criteria are summarized here. The D^0 candidates are reconstructed in the modes $K^-\ \pi^+$, $K^-\ \pi^+\ \pi^0$, $K_S\ \pi^+\ \pi^-$ and $K^-\ \pi^+\ \pi^-\ \pi^+$. The soft pion candidate must have total momentum less than 450 MeV and transverse momentum greater than 50 MeV, in the lab frame. The candidate D^* momentum in the CM frame must be between 0.5 and 2.5 GeV. All D^* candidates with $\delta m = m(D^*) - m(D^0)$ less than 170 MeV are used in the analysis. The candidate D^0 is constrained to have the PDG mass. The candidate lepton momentum in the CM frame is required to be greater than 1.2 GeV. For the signal sample, we require the D^* and the lepton candidate to be on opposite sides of the event and $|\cos\theta_{B^0,D^*\ell}| < 1.1$, where $\cos\theta_{B^0,D^*\ell}$ is the cosine of the angle between the direction of the B^0 and the sum of the D^* and lepton momenta calculated in the $\Upsilon(4S)$ frame. See BAD 147, V4, Sec. 4.2 for the formula for calulating $\cos\theta_{B^0,D^*\ell}$. For some background control samples, described in Sec. 2.4, we require the D^* and the lepton candidate to be on the same side of the event and $|\cos\theta_{B^0,D^*(-\ell)}| < 1.1$ (calculated with the lepton flipped by 180°). We require $|\cos\theta^*_{\rm thrust}| < 0.85$, where $\theta^*_{\rm thrust}$ is the angle between the thrust of the $D^*\ell$ candidate and that of the rest of the event. The error on the decay time difference Δt must be less than 1.4 ps. The magnitude of the decay time is required to be less than 18 ps.

We use the $B^0 \to D^{*-}\ell^+\nu_\ell$ selection criteria described in BAD 147, but emphasize here two of the final cuts that are motivated in the appendices of this document:

1. We require the number of tracks used in the tagging B vertex to be ≥ 2 to reduce the number of outliers in Δt. See Appendix B for details.

2. We require the calculated error on the decay time difference Δt to be less than 1.8 ps to improve the robustness of the Δt resolution model. See Appendix C for details.

10% of candidate signal events are rejected by the above two criteria.

2.4 Background Characterization

In order to determine the amount of different types of background events in the sample of signal candidates and to characterize the shapes of signal and background events in

δm, Δt, $\sigma_{\Delta t}$, etc., we use 12 different control samples that are enhanced in signal or a particular type of background. We also divide these control samples into subsamples according to characteristics that could affect the amount of background or the shapes of the distributions that we analyse. The total number of events in all control samples is $\approx 68,000$.

The $2 \times 2 \times 3 = 12$ signal and control samples are defined by three characteristics:

1. on-resonance data or off-resonance data (2 choices);

2. candidate lepton on the same side of event in CM as candidate D^*, or on opposite side (2 choices);

3. lepton candidate passes "very tight" electron identification criteria or "very tight" muon ID, or fails "loose" electron and muon ID (3 choices).

When appropriate, these samples are further divided into up to 30 subsamples according to the following three characteristics:

1. soft pion reconstructed in SVT only, or in the SVT and DCH (2 choices);

2. the D^0 candidate is reconstructed in the mode $\mathrm{K}\pi$ or $\mathrm{K}\pi\pi^0$ or ($\mathrm{K_S}\pi\pi$ or $\mathrm{K}\pi\pi\pi$) (3 choices).

3. the tagging B is in the lepton, kaon, NT1 or NT2 tagging category, or is not tagged at all (5 choices).

This allows subdivision into up to 360 samples. The software tools that are used for this analysis[4] allow maximum flexibility in grouping subsamples. Individual fit parameters can be shared among different sets of subsamples based on physics motivation or observations in the data.

The extraction of signal and background fractions in each of the samples is decribed fully in BAD 147. We fit the difference between the D^* candidate mass and the D^0 candidate mass, δm, to a double Gaussian plus a background function, described in BAD 147, V4, Sec. 5.1. The amount of D^* and combinatoric background in each sample is estimated with an extended maximum likelihood fit. We estimate the amount of each type of peaking background (a true D^* with either a fake lepton, an "uncorrelated" lepton from the other B in the event, or a lepton from continuum events) using the various control samples and PID efficiencies and misidentification probabilities (measured in data), and relative efficiencies between signal and uncorrelated background control samples (determined from MC).

The resulting peak yields and background fractions are listed in Table 1 for each of the 12 control samples. The δm distributions for the signal samples are shown in Fig. 1 separately for D^* e and D^* μ signal candidates. These figures correspond to the first two rows in Table 1.

Category			Peak Yield	$f_{\mathrm{cont.}}(\%)$	$f_{\mathrm{fake}}(\%)$	$f_{\mathrm{uncorr.}}(\%)$	$f_{\mathrm{comb.}}(\%)$
On	OS	e	7008 ± 91	1.53 ± 0.42	0.1678 ± 0.0042	3.14 ± 0.39	17.89 ± 0.24
		μ	6569 ± 88	2.27 ± 0.57	2.669 ± 0.067	2.85 ± 0.48	18.36 ± 0.25
		f	8770 ± 108	12.8 ± 1.3	72.4 ± 1.8	0.7 ± 1.6	31.40 ± 0.24
	SS	e	306 ± 21	0.000 ± 0.006	0.533 ± 0.039	56.9 ± 7.0	34.0 ± 1.3
		μ	299 ± 20	5.1 ± 3.6	8.89 ± 0.64	48.9 ± 8.0	34.4 ± 1.3
		f	1350 ± 45	20.4 ± 4.1	74.4 ± 5.4	3.6 ± 7.8	42.59 ± 0.61
Off	OS	e	13.6 ± 3.8	100	0	0	21.0 ± 4.5
		μ	18.9 ± 4.7	100	0	0	25.1 ± 4.3
		f	142 ± 14	100	0	0	44.2 ± 1.8
	SS	e	0.000 ± 0.002	100	0	0	100 ± 100
		μ	1.9 ± 1.4	100	0	0	38 ± 14
		f	34.9 ± 7.1	100	0	0	52.5 ± 3.6

Table 1: Peak yields and continuum, fake-lepton and uncorrelated-lepton fractions of the peak yield, and the combinatoric fraction of total events in a δm signal window for the various control samples from Data. Peak yields include the peaking backgrounds. The signal window for combinatoric background fractions is defined as 143 – 148 MeV.

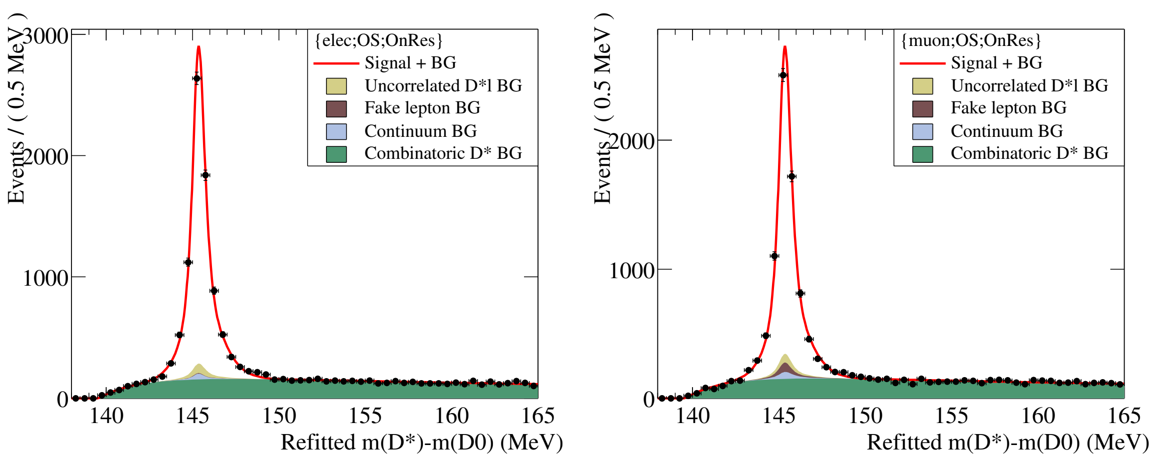

Figure 1: $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ distributions of signal samples in Data. Left: electron sample. Right: muon sample.

2.5 Event Selection Bias

We use the signal Monte Carlo samples described in Sec. 2.2 to determine whether the event selection efficiency depends on the lifetime or mixing properties of the event. To do this, we use the MC truth information to determine the difference in decay times Δt and whether the event corresponds to a mixed ($\mathrm{B}^0\mathrm{B}^0$ or $\overline{\mathrm{B}^0\mathrm{B}^0}$) or unmixed ($\mathrm{B}^0\overline{\mathrm{B}^0}$) event. We require the D* ℓ candidate to match the generated $B^0 \to D^{*-}\ell^+\nu_\ell$ candidate according to the `digi` associator. This last requirement rejects backgrounds in which tracks from both B decays are included in the D* ℓ candidate. Because of charge correlations, it is more likely that a mixed event will contribute to this background than an unmixed event. Therefore, we remove these events from the sample before determining χ_d, Δm_d and τ from truth information. [Note however that we account for different levels of backgrounds in mixed and unmixed events when fitting to reconstructed distributions in data and MC, as described in Sec. 5.]

We then use the truth information on whether the event is mixed or not to calculate the time-integrated mixing probability

$$\chi_d = \frac{N_{mix}}{N_{mix} + N_{unmix}}$$

and its corresponding error. We also do a simultaneous fit for τ and Δm_d to the true Δt distribution. (See Fig. 8 in Sec. 3.5.) In Table 2, we show the difference between the measured and true values of χ_d, τ and Δm_d, after all D* selection criteria and for the two different lepton types. With all cuts applied, the biases on τ and Δm_d are -0.7σ and -0.9σ, respectively, where σ is the Monte Carlo statistical uncertainty. The signal MC sample is more than four times larger than our data set so these biases will be even less significant in the data sample.

Characteristics of signal MC

Sample	χ_d $\times 10^{-3}$	Δm (ps^{-1})	τ (fs)
	174	0.472	1548
all cuts	-0.03 ± 1.4	-0.0012 ± 0.0013	-3.4 ± 5.1
e only	0.92 ± 2.1	-0.0021 ± 0.0018	1.1 ± 7.6
μ only	-0.89 ± 2.1	-0.0013 ± 0.0019	-10.8 ± 7.6

Table 2: Results of calculation of χ_d and simultaneous fits for Δm and τ using MC truth information in the signal model. The values in the table are the difference between the measured and true values of each parameter (generated value shown above double line).

3 Decay-Time Measurement

In this section, we describe the method used to determine the decay-time difference Δt and its error $\sigma_{\Delta t}$ for each event. We define Δt as the difference between the decay time of the fully reconstructed $B^0 \to D^{*-}\ell^+\nu_\ell$ candidate and the decay time of the inclusively reconstructed tagging B:

$$\Delta t = t_{\mathrm{reco}} - t_{\mathrm{tag}}.$$

In Sec. 3.1, we describe the determination of the z position of the $B^0 \to D^{*-}\ell^+\nu_\ell$ candidate and the tagging B. In Sec. 3.2, we describe the approximations made in the calculation of Δt from Δz, and the impact of those approximations. In Sec. 3.3, we identify the effects that are taken into account in the calculation of $\sigma_{\Delta t}$. In Appendix B, we summarize our investigations of "outliers" in Monte Carlo and how to reduce them. In Sec. 3.4, we describe our modeling of the Δt resolution function for signal, and in Sec. 3.5 we present detailed studies of the resolution function in Monte Carlo. Similar studies were done earlier with signal Monte Carlo for both hadronic and semileptonic B decay modes and are documented in BAD 130, "Vertexing performances and systematic checks with fully reconstructed B events", V3, Sec. 5.

3.1 Measurement of Δz

A complete discussion of the algorithms used to reconstruct decay vertices in BaBar can be found in BaBar Analysis Document 102; the performance of the algorithms is described in BaBar Analysis Documents 130 and 183. This section will highlight the most important aspects of vertexing as they relate to our decay-time measurement.

3.1.1 $B^0 \to D^{*-}\ell^+\nu_\ell$ z position

In the calculation of the D^* ℓ vertex position, the D^0 candidate is first fit to a common vertex. The resulting momentum and position are combined with the slow pion and lepton candidates in a vertex fit that uses as a constraint the beam position and "size" in the plane transverse to the detector z axis. The horizontal size of the beamspot is assumed to be the size that is measured by the default beamspot-finding alogorithm run in rolling calibrations. The vertical size is set to 30 μm to account for the transverse flight length of the B and the intrinsic point resolution. By including the beamspot constraint in the in this single vertex fit, we improve the RMS of the residual distribution by 30% (see Appendix F). [2] The difference between the reconstructed z position and the true z position for the D^* ℓ vertex is shown in the left-hand plot in Fig. 2.

[2] The constrained fit for the D^* ℓ candidate that properly accounts for all correlations between the D^0 and D^* and is also used in the calculation of the D^*- D^0 mass difference, as described in Sec. 4.1 of BAD 147, V5.

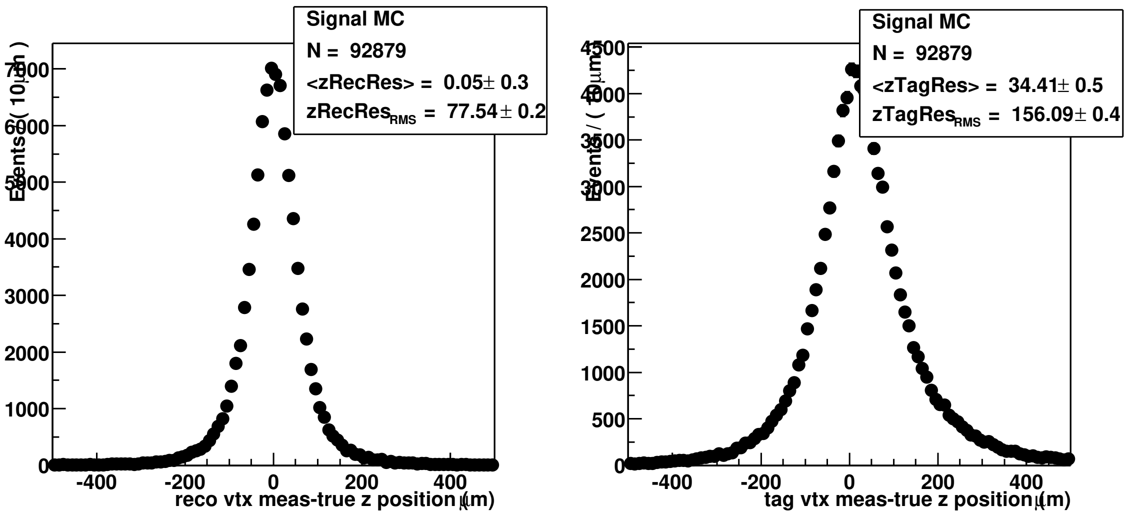

Figure 2: The difference between the reconstructed z position and the true z position for the tagging B vertex (right-hand plot) and the $D^* \ell$ vertex (left-hand plot), in μm, as measured in correctly reconstructed signal Monte Carlo. (Recall the conversion 166 μm/ps.)

3.1.2 z position of the tagging B

The package used for determining the tagging B vertex position and errors is `VtxTagBtaSelFit`, described in Sec. 4.2, BAD102, V5. The vertex is constructed from all good quality tracks in the event except the daughters of the $D^* \ell$ candidate. In order to reduce bias and tails due to long-lived particles, K^0_s and Λ candidates are used as inputs to the fit in place of their daughters. In addition, tracks that are consistent with photon conversions are excluded. The vertex is constrained to the beamspot position in the plane transverse to the detector z axis as described in the previous section. To reduce contributions from charm decay products, which bias the determination of the vertex position, the track with the largest vertex χ^2 contibution is removed if the contribution is greater than 6. This procedure is repeated until no track contributes more than 6 to the χ^2, or there is only one track left.

The difference between the reconstructed z position and the true z position for the tagging B vertex is shown in the right-hand plot in Fig. 2. Compared to the $D^* \ell$ vertex, the residual distribution for the tagging vertex is wider and has a significant bias toward positive z due to tracks from charm decays. This bias is discussed in more detail in Appendix B and Sec. 3.4.

3.2 Conversion from Δz to Δt

The decay time difference between B mesons can be calculated from the difference in decay positions along the BABAR z-axis with the following equation:

$$\Delta t = t_1 - t_2 = m_{\mathrm{B}} \left(\frac{z_1}{p_1^z} - \frac{z_2}{p_2^z} \right) ,$$

where we define the $\Upsilon(4S)$ decay point to be at $z = 0$. The decay times t_1 and t_2 are defined in the respective B rest frames, which are different due to the small momentum of the B meson in the $\Upsilon(4S)$ frame, $\langle p_B^* \rangle = \frac{1}{2}\sqrt{s - 4m_{\mathrm{B}^0}^2} \simeq 320\,\mathrm{MeV/c}$.[3] If we ignore this momentum, the decay time difference is simply

$$\Delta t = \frac{\Delta z}{\gamma \beta_z c} , \tag{1}$$

where γ is the boost factor for the $\Upsilon(4S)$ in the lab frame and β_z its velocity projected on the BABAR z-axis. In BABAR, $\beta_z\gamma \simeq 0.56$.[4]

Equation 1 ignores two effects. The most important is the finite momentum of the B meson in the $\Upsilon(4S)$ rest frame, mentioned in the previous paragraph and quantified below. Secondly, the $\Upsilon(4S)$ momentum is smeared due to the momentum spread of the colliding beams. This smearing is about 6 MeV/c, which corresponds to 0.1% smearing of $\beta\gamma$. The factor of β_z accounts for the fact that the direction of the colliding beams is tilted by 20 mrad with respect to the detector z-axis. The boost along the z-axis is reduced by 1.2 MeV/c (0.02%) due to this effect. There is also an effective boost of about 118 MeV/c in the x direction, which generates an azimuthal asymmetry, but has no direct impact on the estimation of decay time difference.

For a B meson that decays with polar angle θ^* with respect to the boost direction in the $\Upsilon(4S)$ frame, the distance Δz can be written as [5]

$$\Delta z = c\gamma\beta\gamma^*(t_1 - t_2) + c\gamma\gamma^*\beta^* \cos\theta^*(t_1 + t_2) , \tag{2}$$

where $\gamma^*\beta^*$ is the boost of the B meson in the $\Upsilon(4S)$ frame. If the acceptance does not depend on θ^*, the transformation Eq. 1 is only biased by a factor γ^*, which is approximately 1.002, since $\langle \cos\theta^* \rangle = 0$.

In general we don't know $t_1 + t_2$ event by event, but if the B decay angle θ^* is known (as is the case for fully reconstructed B's), the estimate of event-by-event Δt can be improved using the expectation value of $t_1 + t_2$. Since both t_1 and t_2 are positive, the minimum value of $t_1 + t_2$ is $|\Delta t|$. By integrating $t_1 + t_2$ from $|\Delta t|$ to infinity, we get $\langle t_1 + t_2 \rangle = \tau_{\mathrm{B}} + |\Delta t|$. Substituting this correction in Eq. 2, we get

$$\frac{\Delta z}{\gamma\beta\gamma^* c} = \Delta t + \frac{\beta^* \cos\theta^*}{\beta}(\tau_{\mathrm{B}} + |\Delta t|) .$$

---

[3] We use values used in Monte Carlo simulation: $E_{e^+} = 3.11175\,\mathrm{GeV}$, $E_{e^-} = 8.99\,\mathrm{GeV}$, $\sqrt{s} = 10.5782\,\mathrm{GeV}$ and $m_{\mathrm{B}^0} = 5.2794\,\mathrm{GeV}$.

[4] The value used in Monte Carlo is $\beta_z\gamma = 0.555596$.

In this analysis, we do not measure the B decay direction. The effect of the second term in Eq. 2 is to smear the exponential distribution of Δz. The RMS of this term is

$$c\gamma\gamma^*\beta^*\sqrt{\langle\cos^2\theta^*\rangle\,\langle(t_1+t_2)^2\rangle} \simeq 21\cdot\sqrt{6}\,\tau_{\rm B}\sqrt{\langle\cos^2\theta^*\rangle}\ (\mu{\rm m}) \tag{3}$$

where we have used $\beta^* \simeq 0.0606$, $\gamma \simeq 1.144$ and the expectation value of $\langle(t_1+t_2)^2\rangle = 6\tau_{\rm B}^2$. The angular distribution for $\Upsilon(4{\rm S}) \rightarrow {\rm B}\overline{\rm B}{}^0$ is given by $(1-\cos^2\theta^*)d(\cos\theta^*)$, which gives $\langle\cos^2\theta^*\rangle = 1/5$. The RMS of the Δz residual due to the boost approximation is then about 35 μm, which is about 0.21 ps for Δt. This result is for integration over the entire (t_1, t_2) space. For a given Δt, the expectation value of $(t_1+t_2)^2$ depends on $|\Delta t|$. We can integrate $(t_1+t_2)^2$ from $|\Delta t|$ to infinity and get

$$\left\langle(t_1+t_2)^2\right\rangle|_{\Delta t} = 2\tau_{\rm B}^2 + 2\tau_{\rm B}|\Delta t| + |\Delta t|^2\,.$$

If we further integrate $|\Delta t|$ from zero to infinity, we get the factor of $6\tau_{\rm B}^2$ that appears in Eq. 3. Therefore the smearing introduced by using the approximation $\Delta z/(\gamma\beta\gamma^* c)$ can be written as a function of $|\Delta t|$:

$$\frac{\beta^*}{\sqrt{5}\beta}\sqrt{\tau_{\rm B}^2 + (\tau_{\rm B}+|\Delta t|)^2} \simeq 0.0557\sqrt{\tau_{\rm B}^2 + (\tau_{\rm B}+|\Delta t|)^2}\,. \tag{4}$$

We can demonstrate this effect by plotting the RMS of $\Delta z_{\rm truth}/(\gamma\beta\gamma^* c) - \Delta t_{\rm truth}$ in bins of $|\Delta t_{\rm truth}|$. Figure 3 shows this plot for 80k signal MC events that pass all our final signal cuts and 100k events directly from the event generator, and a fit to a function of the form given in Eq. 4: $p_0\ \cdot\ \sqrt{p_1^2 + (p_1 + |\Delta t|)^2}$.

3.3 Calculation of Error on Δt

The calculated uncertainties on the vertex position of the ${\rm B}^0 \rightarrow {\rm D}^{*-}\ell^+\nu_\ell$ candidate and the tagging B account for uncertainties on the track parameters due to SVT and DCH hit resolution and multiple scattering, our knowledge of the beam spot size, and the average B flight length in the vertical direction. The calculated uncertainty does not account for errors in pattern recognition in tracking or errors in associating tracks with the B vertex. The calculated uncertainties will also be incorrect if our assumptions for the amount of material in the detector or the beam spot size or position are incorrect. We use parameters in the resolution model, measured with data, to account for uncertainties and biases introduced by these effects.

3.4 Resolution Models

A resolution model $\mathcal{R}$ specifies the expected distribution of residuals $\delta\Delta t \equiv \Delta t_{\rm meas} - \Delta t_{\rm true}$ for an event with measured $\Delta t_{\rm meas}$ and calculated vertex error $\sigma_{\Delta t}$. We assume that resolution is independent of $\Delta t_{\rm meas}$, but expect a strong correlation with $\sigma_{\Delta t}$; i.e., $\mathcal{R} =$

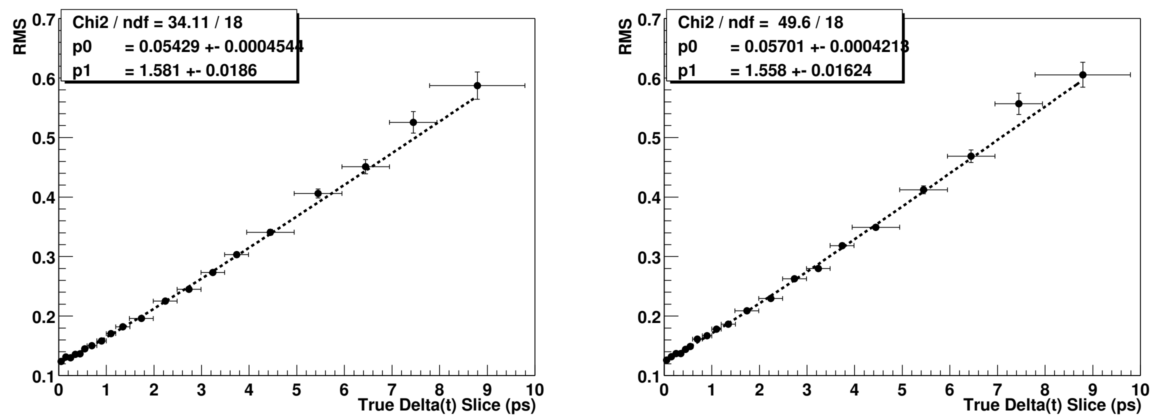

Figure 3: RMS of the distribution of residuals $\Delta z_{\mathrm{truth}}/(\gamma\beta\gamma^{*}c) - \Delta t_{\mathrm{truth}}$ in bins of $|\Delta t_{\mathrm{truth}}|$ for 80k signal MC events that pass all event slection criteria (left) and for 100k events from the event generator (right), and a fit to a function $p_0 \cdot \sqrt{p_1^2 + (p_1 + |\Delta t|)^2}$. The $\Delta z \rightarrow \Delta t$ conversion factor used here is 166.87μm/ps.

$\mathcal{R}(\delta\Delta t, \sigma_{\Delta t})$. The left-hand plot in Fig. 4 shows the distribution of $\sigma_{\Delta t}$ values observed in signal MC. For the study presented in this subsection, the cut on $\sigma_{\Delta t} < 1.8$ ps is *not* applied. The study in this subsection motivates the cut.

If the calculated errors are accurate and the Δt reconstruction is unbiased, then the distribution of pulls $\delta\Delta t/\sigma_{\Delta t}$ is a unit Gaussian and the resulting resolution function has no free parameters:

$$\mathcal{R}(\delta\Delta t, \sigma_{\Delta t}) = G(\delta\Delta t;\, 0, \sigma_{\Delta t}) \;,$$

where we define the Gaussian function

$$G(x;\, x_0, \sigma) \equiv \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\bigl(-(x - x_0)^2/(2\sigma)^2\bigr) \;. \tag{5}$$

The right-hand plot in Fig. 4 shows that the pull distribution observed in signal MC is approximately Gaussian, but is biased towards negative residuals and has an RMS of about 1.3, which indicates that the calculated errors are underestimated on average by about 30%.

In order to study these deviations from the ideal case in more detail, we divide the sample of correctly reconstructed signal MC into slices of $\sigma_{\Delta t}$ as shown in the left-hand plot in Fig. 4. Figure 5 shows the distribution of residuals $\delta\Delta t$ within each slice, and Fig. 6 shows the mean and RMS calculated directly from the residuals in each slice, as a function of each slice's midpoint $\sigma_{\Delta t}$ value. The RMS of the residuals scales with the calculated error as expected. It is clear from the upper plot in Fig. 6 that the bias also scales with the calculated error, a result that is not so obvious. We present an explanation for this scaling of the bias in App. C.

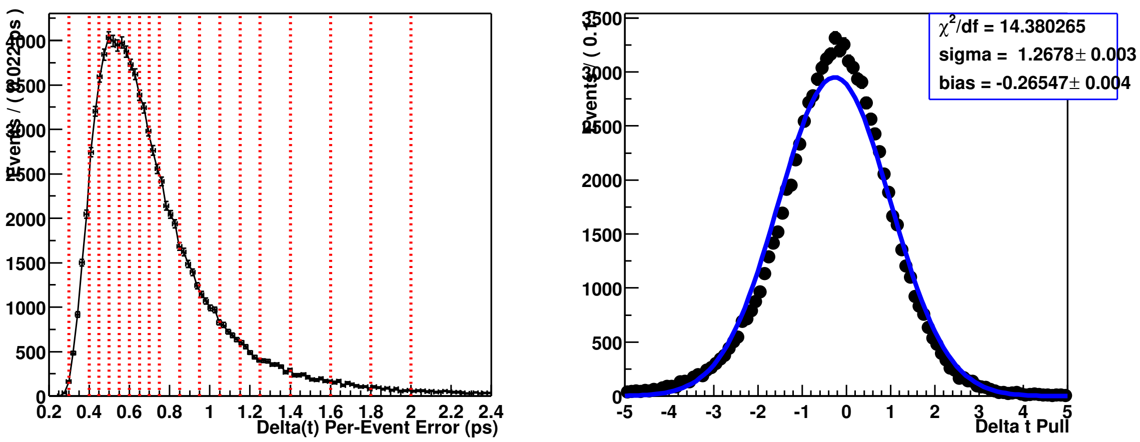

Figure 4: Distribution of calculated vertex errors $\sigma_{\Delta t}$ (left side) and pulls (right side) obtained from correctly reconstructed signal MC. Vertical dashed lines in the $\sigma_{\Delta t}$ distribution indicate the slices used to select subsamples with approximately equal calculated errors. The solid curve superimposed on the pull distribution is the result of a single Gaussian fit.

In order to accomodate these effects, we consider two different models for the "core" of the resolution function (loosely defined as $|\delta\Delta t| < 5\,\sigma_{\Delta t}$, as used in Figs. 4 and 5). The first core model consists of a double Gaussian whose biases and widths are scaled by the measured $\sigma_{\Delta t}$:

$$\mathcal{R}_{\rm G+G}(\delta\Delta t, \sigma_{\Delta t};\, b_1, b_2, s_1, s_2, f) \equiv f \cdot G(\delta\Delta t;\, b_1\sigma_{\Delta t}, s_1\sigma_{\Delta t}) + (1-f)\cdot G(\delta\Delta t;\, b_2\sigma_{\Delta t}, s_2\sigma_{\Delta t})\ .$$

This model is similar to the one used for the CP [8] and hadronic mixing [9] analyses, except for the scaling of the bias that is clearly present in signal Monte Carlo (Fig. 6).

The second model is the sum of a single Gaussian and the same Gaussian convoluted with a one-sided ($\delta\Delta t < 0$) exponential $E(\delta\Delta t;\, \kappa)$:

$$\begin{aligned}\mathcal{R}_{\rm GExp}(\delta\Delta t, \sigma_{\Delta t};\, \kappa, s, f) \equiv\ & f \cdot G(\delta\Delta t;\, 0, s\,\sigma_{\Delta t}) + \\ & (1-f)\cdot \frac{1}{2\,\kappa\sigma_{\Delta t}} \int_{-\infty}^{0} du\, \exp\big(+u/(\kappa\,\sigma_{\Delta t})\big)\cdot G(u - \delta\Delta t;\, 0, s\,\sigma_{\Delta t})\ . \quad (6)\end{aligned}$$

The integral in this definition can be performed analytically to give

$$\begin{aligned}\mathcal{R}_{\rm GExp}(\delta\Delta t, \sigma_{\Delta t};\, \kappa, s, f) =\ & f\, G(\delta\Delta t;\, 0, s\,\sigma_{\Delta t}) + \\ & (1-f)\,\frac{1}{2\,\kappa\sigma_{\Delta t}}\, \exp\left(\frac{s^2 + 2\,\delta\Delta t\cdot\kappa/\sigma_{\Delta t}}{2\kappa^2}\right)\, \mathrm{erfc}\left(\frac{s^2 + \delta\Delta t\cdot\kappa/\sigma_{\Delta t}}{\sqrt{2}\,s\,\kappa}\right)\ .\end{aligned}$$

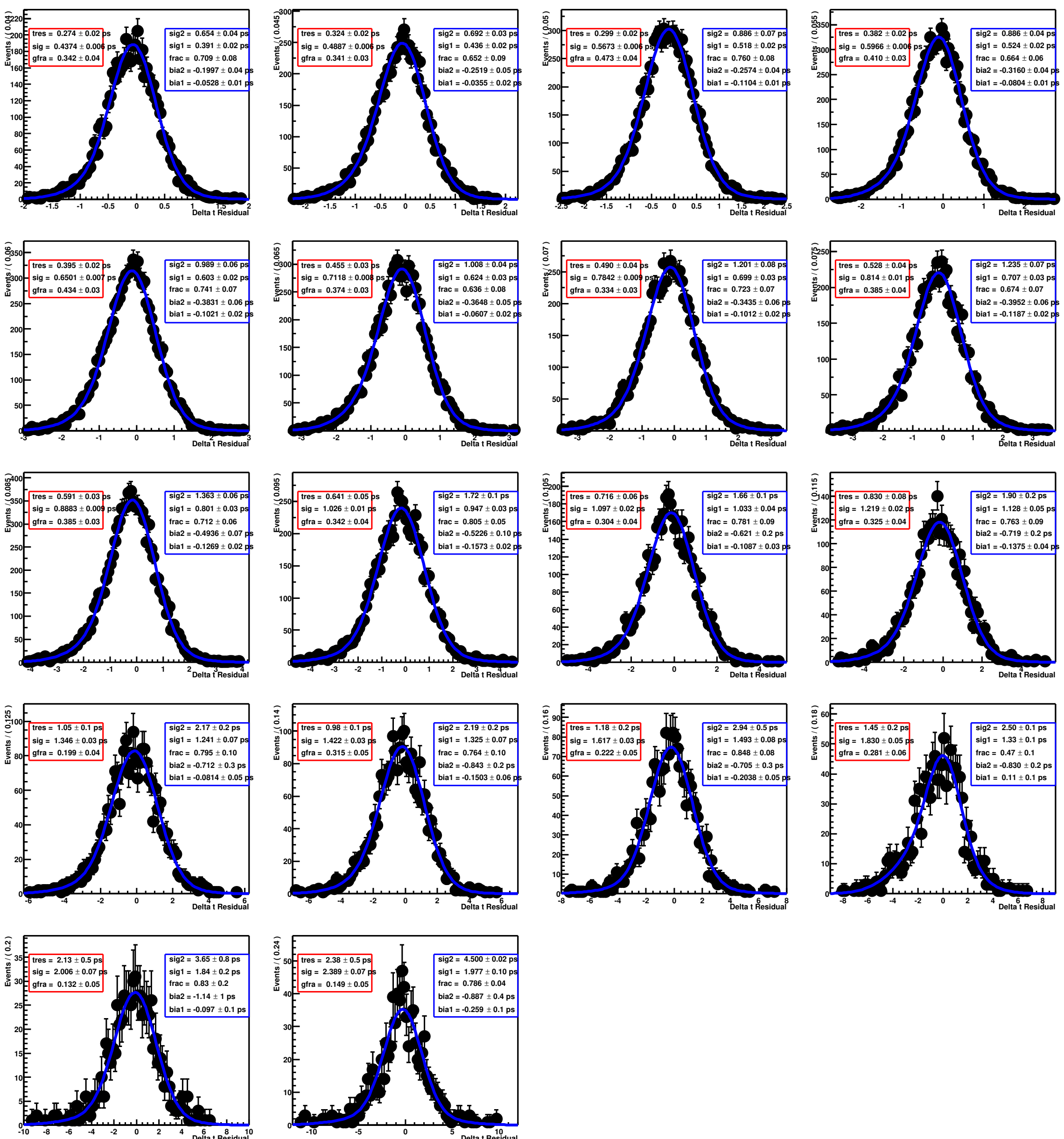

Figure 5: Distributions of residuals in signal MC for slices of $\sigma_{\Delta t}$. A G+G fit (solid curve and parameters in upper box) and GExp fit (dashed curve and parameters in lower box) are superimposed. The fits are only to the range shown, which corresponds to $|\delta\Delta t| < 5\,\sigma_{\Delta t}$.

This model was also used in the hadronic lifetime analysis[10].

Once the parameters of a resolution model are known, we can calculate the corre-

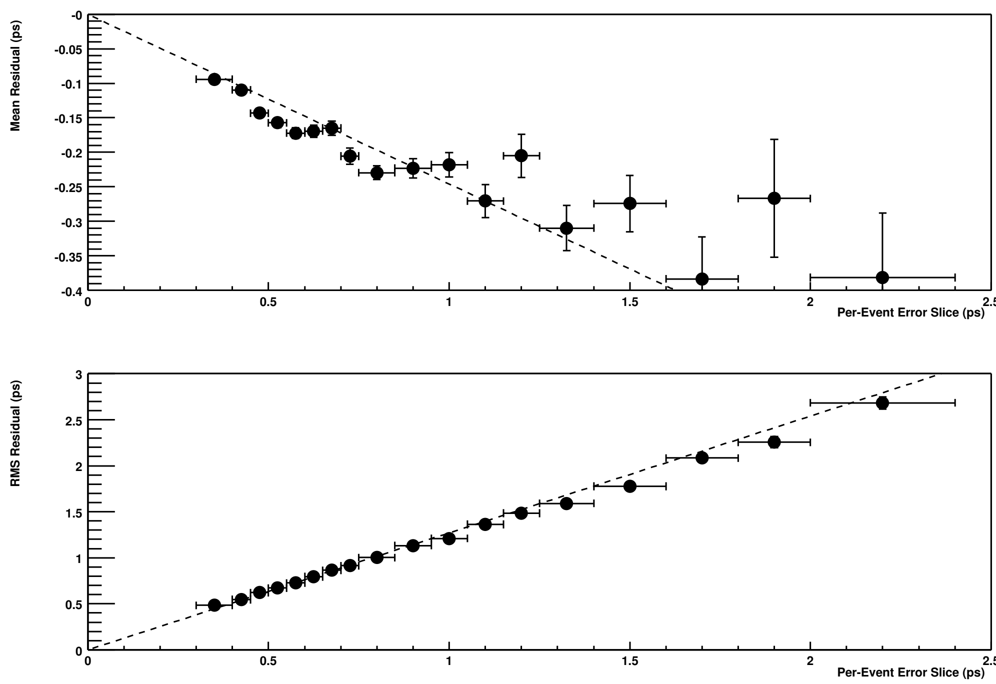

Figure 6: Mean and RMS of the residual distributions obtained from correctly reconstructed signal MC, in slices of $\sigma_{\Delta t}$. Statistics are calculated for the core region of $|\delta\Delta t| < 5\,\sigma_{\Delta t}$ only. Dashed lines indicate the slopes of -0.25 (mean) and 1.27 (RMS) obtained from the single Gaussian fit in Figure 4.

sponding moments of the residual distribution directly in terms of the moments of the per-event error distribution. For the G+G model, the first and second moments are

$$\langle \delta\Delta t\rangle_{\mathrm{G+G}} = \big(f b_1 + (1-f) b_2\big)\cdot\langle\sigma_{\Delta t}\rangle \tag{7}$$

$$\langle (\delta\Delta t)^2\rangle_{\mathrm{G+G}} = \big(f(s_1^2+b_1^2) + (1-f)(s_2^2+b_2^2)\big)\cdot\langle(\sigma_{\Delta t})^2\rangle\ , \tag{8}$$

from which we calculate the G+G RMS squared

$$\begin{aligned}(\delta\Delta t)^2_{\mathrm{RMS}} = &\big(f s_1^2 + (1-f)s_2^2\big)\cdot\langle(\sigma_{\Delta t})^2\rangle + f(1-f)\,\big(b_1-b_2\big)^2\cdot\langle\sigma_{\Delta t}\rangle^2\\ &+\big(f b_1^2 + (1-f)b_2^2\big)\cdot(\sigma_{\Delta t})^2_{\mathrm{RMS}}\ .\end{aligned}$$

For comparison, the first and second moments of the GExp model are

$$\langle \delta\Delta t\rangle_{\mathrm{GExp}} = -(1-f)\,\kappa\cdot\langle\sigma_{\Delta t}\rangle \tag{9}$$

$$\langle (\delta\Delta t)^2\rangle_{\mathrm{GExp}} = \big(s^2 + 2\,(1-f)\,\kappa\big)\cdot\langle(\sigma_{\Delta t})^2\rangle\ , \tag{10}$$

from which we calculate the GExp RMS squared

$$(\delta\Delta t)^2_{\rm RMS} = s^2 \cdot \langle(\sigma_{\Delta t})^2\rangle + (1-f^2)\,\kappa \cdot \langle\sigma_{\Delta t}\rangle^2 \\ + 2\,(1-f)\,\kappa \cdot (\sigma_{\Delta t})^2_{\rm RMS} \; .$$

Both of these models predict a mean and RMS residual that scale linearly with $\sigma_{\Delta t}$ and vanish in the limit of zero measured error. The predicted slopes for the dependence of the mean residual with the measured error are

$$f\,b_1 + (1-f)\,b_2 \quad ({\rm G+G}) \qquad , \qquad -(1-f)\,\kappa \quad ({\rm GExp}) \; ,$$

and the predicted slopes for the RMS are

$$\sqrt{f\,s_1^2 + (1-f)\,s_2^2 + f(1-f)\,(b_1-b_2)^2} \quad ({\rm G+G}) \quad , \quad \sqrt{s^2 + (1-f^2)\,\kappa} \quad ({\rm GExp}) \; .$$

Figure 7 shows the results of G+G and GExp fits to each of the slices shown in Figure 5. We observe the expected scaling of the fit parameters at least out to 1.8 ps. Since events with calculated uncertainties $\sigma_{\Delta t} > 1.8$ ps represent only about 1.5% of our sample and contribute even less to our statistical sensitivity, we apply an additional vertex quality cut of $\sigma_{\Delta t} < 1.8$ ps. This is tighter than the cut at 2.4 ps applied in the CP and hadronic mixing analyses.

The plots in the second row in Fig. 7 show once more that the bias scales with the per-event error. The explanation of this effect is given in App. C.

In addition to the core of the resolution function, we model the contribution of "outliers": candidates whose residual $\delta\Delta t$ is (on average) at least 5 times larger than their calculated error $\sigma_{\Delta t}$ (and which are therefore poorly vertexed.) Outliers represent about 1% of the candidates selected in signal Monte Carlo. We model outliers with a single Gaussian whose width and bias are scaled by the calculated error

$$\mathcal{R}_{\rm out}(\delta\Delta t, \sigma_{\Delta t};\, b_{\rm out}, s_{\rm out}) \equiv G(\delta\Delta t;\, b_{\rm out}\sigma_{\Delta t}, s_{\rm out}\sigma_{\Delta t}) \; .$$

We combine the two core and outlier models using an extra parameter $f_{\rm out}$ that specifies the relative fraction of events in the outlier contribution:

$$\mathcal{R}_{\rm G+G+G}(\delta\Delta t, \sigma_{\Delta t};\, b_1, b_2, b_{\rm out}, s_1, s_2, s_{\rm out}, f, f_{\rm out}) \equiv \\ f \cdot G(\delta\Delta t;\, b_1\sigma_{\Delta t}, s_1\sigma_{\Delta t}) + (1-f-f_{\rm out}) \cdot G(\delta\Delta t;\, b_2\sigma_{\Delta t}, s_2\sigma_{\Delta t}) + f_{\rm out}\cdot R_{\rm out} \; ,$$

and

$$\mathcal{R}_{\rm GExp+G}(\delta\Delta t, \sigma_{\Delta t};\, \kappa, s, f, b_{\rm out}, s_{\rm out}, f_{\rm out}) \equiv f \cdot G(\delta\Delta t;\, 0, s\,\sigma_{\Delta t}) + \\ (1-f-f_{\rm out}) \cdot \frac{1}{2\,\kappa\sigma_{\Delta t}} \int_{-\infty}^{0} du\, \exp\big(+u/(\kappa\,\sigma_{\Delta t})\big) \cdot G(u-\delta\Delta t;\, 0, s\,\sigma_{\Delta t}) \\ + f_{\rm out}\cdot R_{\rm out}(\delta\Delta t, \sigma_{\Delta t};\, b_{\rm out}, s_{\rm out}) \; .$$

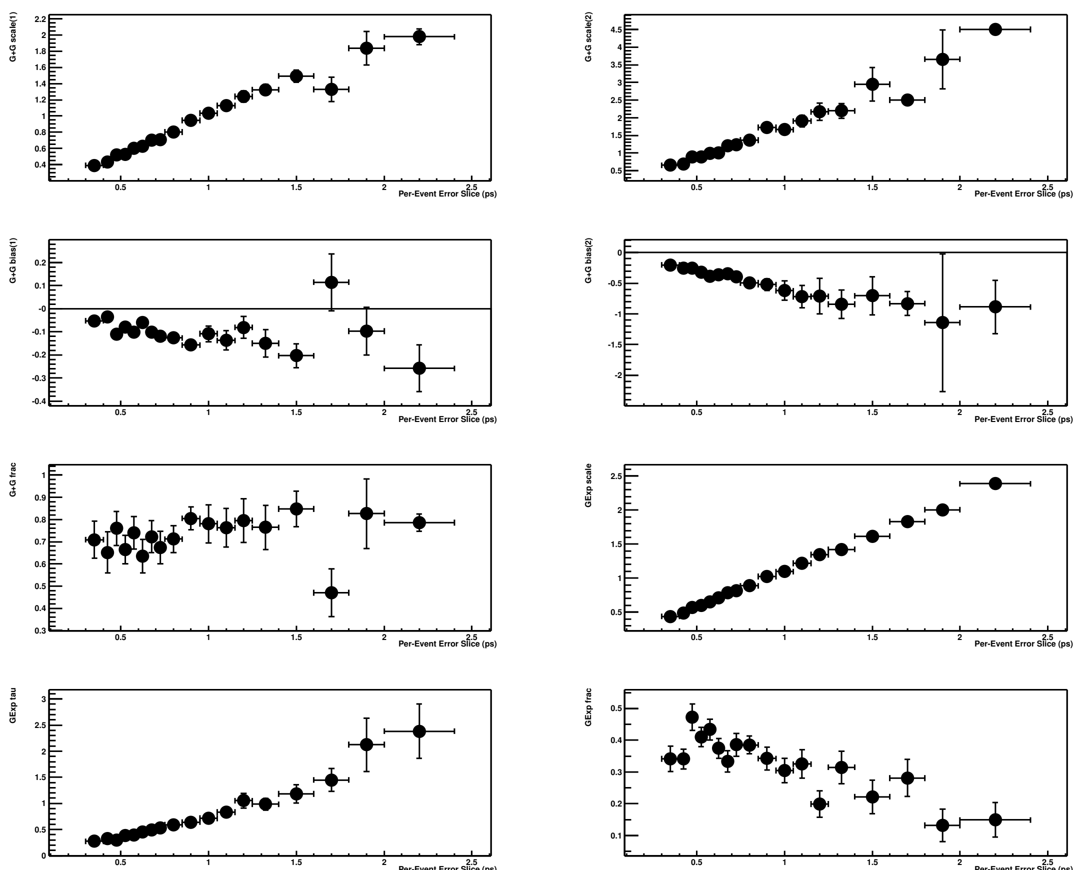

Figure 7: Parameters obtained from G+G and GExp fits to the residual distributions of Figure 5 (signal MC). The first 5 plots are the G+G parameters s_1, s_2, b_1, b_2, and f. The last 3 plots are the GExp parameters s, κ, and f.

3.5 Resolution in Signal Monte Carlo

In this section, we describe a series of fits that we perform to our sample of signal Monte Carlo (2.5–8 times the data sample, depending on the D^0 decay mode). All of the candidates used in these studies have been selected by our analysis chain and are correctly reconstructed according to the `digi` associator.

Many of the detailed figures and tables referred to in this section appear in Appendix D.

Table 3 gives the summary statistics for our signal MC sample and the various decompositions that we use for systematic studies. The mean and RMS values quoted in this table (which are also shown graphically in Figs. 66 and 67 in Appendix D) are calculated directly from the events, instead of being obtained from a fit. The quantity $\sqrt{\langle\sigma^2_{\Delta t}\rangle}$

given in these tables is the appropriate number for multiplying the scale factors in the resolution model to convert them into an RMS for the weighted sample, as described in the previous section and made explicit in Eqs. 8 and 10. The combined sample does not contain the correct branching-ratio proportions of the four D^0 decay modes we study. The numbers broken down by D^0 mode do not add up to the full sample since they do not include events in which a candidate is correctly reconstructed from the "other B". The "perfect tag" category refers to candidates in which the flavor tag is correctly assigned. The categories "SVT only" and "SVT+DCH" refer to the subsystems in which the slow pion is reconstructed.

Sample	Events	Mean $\delta\Delta t$ (ps)	RMS $\delta\Delta t$ (ps)	$\sqrt{\langle\sigma^2_{\Delta t}\rangle}$ (ps)	$f(\lvert\text{pull}\rvert > 5)$ (%)
All Combined	92879	-0.208 ± 0.004	1.071 ± 0.002	0.764	$0.71\pm$ 0.03
Electrons	42688	-0.211 ± 0.005	1.063 ± 0.004	0.758	$0.67\pm$ 0.04
Muons	40905	-0.201 ± 0.005	1.068 ± 0.004	0.763	$0.74\pm$ 0.04
SVT only	74429	-0.206 ± 0.004	1.077 ± 0.003	0.770	$0.68\pm$ 0.03
SVT+DCH	18450	-0.216 ± 0.008	1.048 ± 0.005	0.740	$0.83\pm$ 0.07
$D^0 \to K\pi$	51838	-0.208 ± 0.005	1.061 ± 0.003	0.758	$0.71\pm$ 0.04
$D^0 \to K3\pi$	21616	-0.209 ± 0.007	1.064 ± 0.005	0.761	$0.70\pm$ 0.06
$D^0 \to K\pi\pi^0$	17135	-0.201 ± 0.008	1.010 ± 0.006	0.784	$0.71\pm$ 0.06
$D^0 \to K_S\pi\pi$	2290	-0.2426 ± 0.02	1.16 ± 0.02	0.783	$0.87\pm$ 0.19
Lepton Tag	12270	-0.105 ± 0.008	0.852 ± 0.005	0.661	$0.40\pm$ 0.06
Kaon Tag	32509	-0.233 ± 0.006	1.050 ± 0.004	0.739	$0.69\pm$ 0.05
NT1 Tag	7732	-0.13 ± 0.01	0.974 ± 0.008	0.688	$0.66\pm$ 0.09
NT2 Tag	13425	-0.211 ± 0.010	1.111 ± 0.007	0.776	$0.78\pm$ 0.08
NT3 Tag	26943	-0.242 ± 0.007	1.183 ± 0.005	0.848	$0.86\pm$ 0.06
Reco'd B^0	47184	-0.210 ± 0.005	1.079 ± 0.004	0.765	$0.71\pm$ 0.04
Reco'd $\overline{B^0}$	45695	-0.205 ± 0.005	1.064 ± 0.004	0.763	$0.71\pm$ 0.04
B^0 Tag	47652	-0.216 ± 0.005	1.086 ± 0.004	0.767	$0.75\pm$ 0.04
$\overline{B^0}$ Tag	45227	-0.199 ± 0.005	1.056 ± 0.004	0.761	$0.67\pm$ 0.04
Perfect Tag	68448	-0.194 ± 0.004	1.040 ± 0.003	0.747	$0.65\pm$ 0.03
Imperfect Tag	24431	-0.246 ± 0.007	1.153 ± 0.005	0.812	$0.87\pm$ 0.06
Unmixed	60894	-0.193 ± 0.004	1.035 ± 0.003	0.751	$0.64\pm$ 0.03
Mixed	31985	-0.235 ± 0.006	1.136 ± 0.004	0.789	$0.84\pm$ 0.05

Table 3: Statistics of the semileptonic signal Monte Carlo sample, and residual and $\sigma_{\Delta t}$ distributions.

We study the resolution in signal MC using the following sequence of fits:

- fit the MC-truth Δt and Δz distributions for the lifetime,
- fit the distribution of MC residuals $\delta(\Delta t)$ and calculated uncertainty on Δt, $\sigma_{\Delta t}$, to obtain the parameters of the three resolution models G+G, G+G+G, and GExp+G, and
- fit the reconstructed Δt distribution to obtain the lifetime, using the three resolution models, and with the resolution parameters either fixed (to values obtained in the previous step) or floating in the fit.

Tables 71, 72 and 73 list the parameters of the G+G, G+G+G and GExp+G models that we obtain from fits to MC residuals using calculated per-event errors. Figures 60 and 61 show the fits of the combined sample to both residuals and pulls on linear and log scales. The GExp+G fits are represented by dashed curves in both figures. The solid curve corresponds to the G+G model in Fig. 60 and the G+G+G model in Fig. 61.

Table 4 shows the results of different lifetime fits to the signal MC sample. The first column gives the results of the fit to the distribution of true decay-time difference Δt_{true}, relative to the MC generator value of 1.548 ps. The fit to Δt_{true} reveals any bias introduced by the event selection (see Section 2.5.) Figure 8 shows the fits of pure exponentials to Δt_{true} and Δz_{true}. By comparing the lifetimes extracted from the fits to Δt_{true} and Δz_{true}, we can estimate any additional bias due to the boost approximation. However, the difference between the lifetime from these two fits is probably an overestimate of the boost effects since some of the bias will be absorbed into the resolution model. When we fit to the Δt_{true} distribution, we measure a lifetime of (1.5446 ± 0.0051) ps, which is (3.4 ± 5.1) fs below the generated value. When we fit to the Δz_{true} distribution and scale the fit parameter by the average value of $\beta\gamma\gamma^* c = 166.87$ μm/ps for the generated MC sample, we measure a lifetime of (1.5494 ± 0.0054) ps. This value is 4.8 fs higher than that extracted from Δt_{true}, which is significant, given that the events are identical.

This bias, however, is expected. The reason is the following. Fitting to a pure exponential $e^{-|x|/\tau}$ is mathematically identical to calculating the average of the value $|x|$. According to Eq. 2, for large $t_1 - t_2$ and small $t_1 + t_2$, Δz is always positive, independent of $\cos\theta^*$. The average of $|\Delta t|$ is equal to that of $|\Delta z|/\beta\gamma\gamma^* c$ for these events since $\langle\cos\theta^*\rangle = 0$. However, for small $t_1 - t_2$ and large $t_1 + t_2$, Δz can be negative even if $t_1 - t_2 > 0$. After taking the absolution value, these events will contribute in positive values, the the average over $\cos\theta^*$ will not cancel, therefore bias the average high.

To estimate this bias, we generate 100 toy experiments of $\approx$81k events simply using $e^{-t/\tau}$ and $(1 - \cos^2\theta^*)$ distributions and using Eq. 2 to calculate Δz. We then calculate the average of $|\Delta t|$ and $|\Delta z|/\beta\gamma\gamma^* c$. The histogram of the shifts is shown in Fig. 9. The blue line shows the value found in signal MC fit. The mean bias is 4.7 fs with RMS 0.76 fs. Signal MC fit is consistent with toy experiments.

We can use the distribution of residuals $\Delta z_{\mathrm{true}}/\gamma\beta\gamma^* c - \Delta t_{\mathrm{true}}$ to model the resolution function for Δz_{true} (see Eq. 2) in the fit to the lifetime. Figure 10 shows the residual distribution and its fit to three Gaussians, and the lifetime fit using the resolution function from the fit to residuals. Here we assume the resolution function does not depend on Δt_{true}. The result is (1.5352 ± 0.0055) ps, which is 5.2 fs below the value from the fit to Δt_{true}. Again the difference is significant.

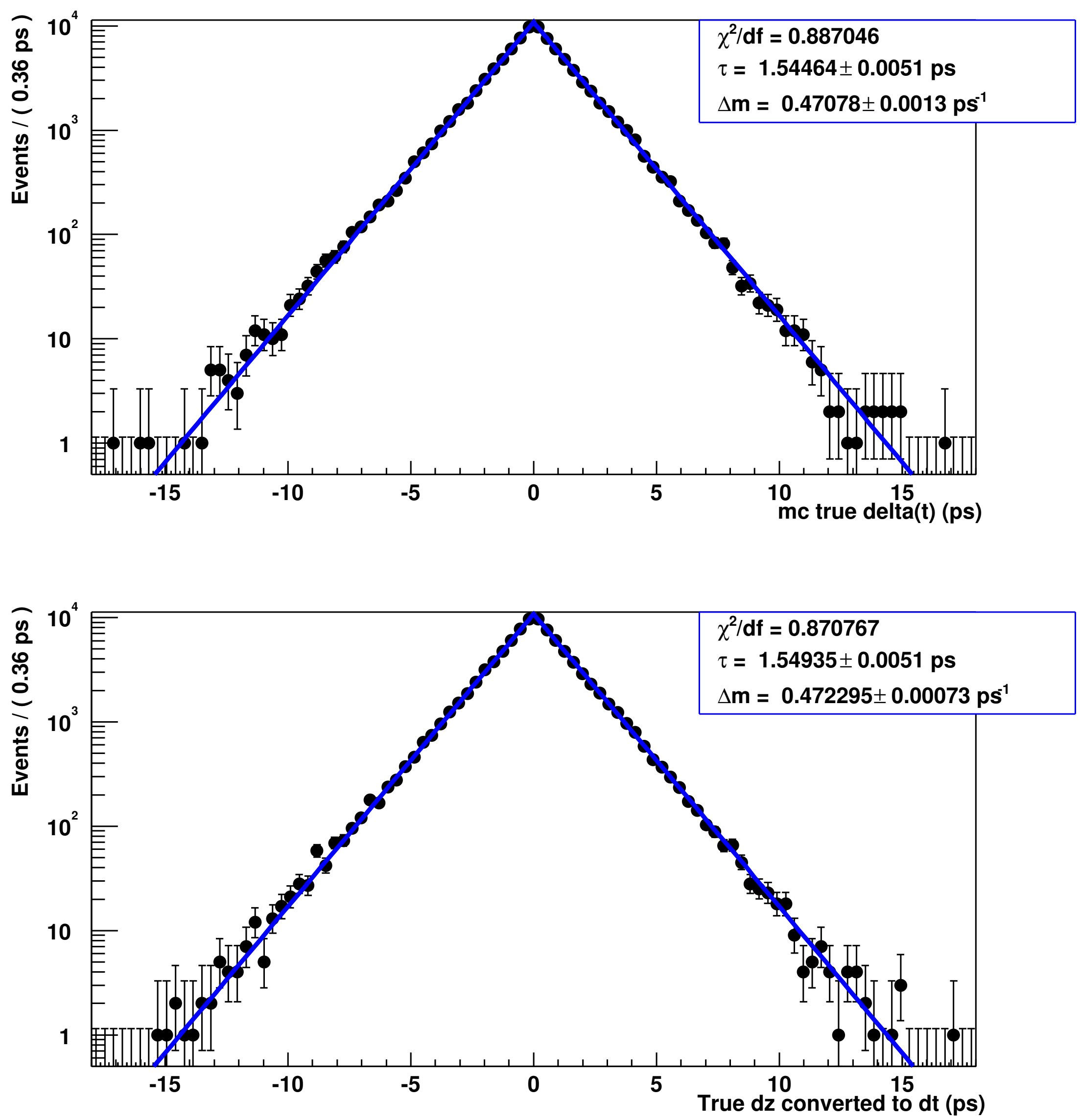

Figure 8: Fits to signal MC truth Δt and Δz distributions. See the text for details.

Sample	$\tau(\Delta t_{\rm true})$ $-\tau$(gen) (fs)	$\tau(\Delta z_{\rm true})$ $-\tau$(gen) (fs)	G+G+G, $\tau(\Delta t)-\tau(\Delta t_{\rm true})$ (fs) τ Only	τ+Resln	G+G, $\tau(\Delta t)-\tau(\Delta t_{\rm true})$ (fs) τ Only	τ+Resln	GExp+G, $\tau(\Delta t)-\tau(\Delta t_{\rm true})$ (fs) τ Only	τ+Resln
Combined	-3.4 ± 5.0	1.3 ± 5.0	-9.8 ± 5.9	2.0 ± 15.7	-13.8 ± 5.9	9.3 ± 10.6	-11.8 ± 5.9	3.4 ± 10.2
Electrons	1.1 ± 7.5	6.5 ± 7.5	-12.2 ± 8.7	-18.0 ± 7.6	-16.5 ± 8.7	-14.8 ± 13.6	-14.2 ± 8.7	-16.9 ± 14.5
Muons	-10.8 ± 7.6	-7.5 ± 7.6	-12.0 ± 8.8	4.3 ± 17.7	-15.6 ± 8.8	9.4 ± 15.8	-13.8 ± 8.8	7.2 ± 15.8
SVT only	-3.1 ± 5.7	1.0 ± 5.7	-10.7 ± 6.6	5.3 ± 11.5	-14.9 ± 6.6	12.8 ± 12.9	-12.8 ± 6.6	7.4 ± 11.3
SVT+DCH	-4.5 ± 11.4	3.0 ± 11.4	-6.0 ± 13.1	-18.2 ± 16.7	-9.6 ± 13.1	-2.0 ± 19.2	-7.6 ± 13.1	-17.1 ± 24.3
$\mathrm{D^0 \to K\pi}$	-8.4 ± 6.8	-4.1 ± 6.8	-4.9 ± 7.9	9.1 ± 12.9	-9.1 ± 7.8	13.5 ± 11.9	-6.8 ± 7.8	11.1 ± 7.2
$\mathrm{D^0 \to K3\pi}$	-2.8 ± 10.5	2.2 ± 10.5	-17.8 ± 12.1	1.9 ± 31.9	-22.7 ± 12.1	11.9 ± 34.6	-20.3 ± 12.1	-1.6 ± 48.8
$\mathrm{D^0 \to K\pi\pi^0}$	6.1 ± 11.9	11.8 ± 11.9	-13.1 ± 13.8	1.0 ± 22.8	-16.0 ± 13.8	35.4 ± 26.8	-15.1 ± 13.8	-1.4 ± 23.1
$\mathrm{D^0 \to K_S\pi\pi}$	34 ± 33	39 ± 33	$-17pm38$	-40 ± 55	-20 ± 38	-38 ± 61	-18 ± 38	-43 ± 63
Reco'd $\mathrm{B^0}$	-5.8 ± 7.1	-1.9 ± 7.1	-7.6 ± 8.2	-11.9 ± 15.3	-11.6 ± 8.2	-4.8 ± 15.1	-9.7 ± 8.2	-9.6 ± 12.6
Reco'd $\overline{\mathrm{B^0}}$	-0.8 ± 7.2	4.7 ± 7.3	-12.0 ± 8.4	13.7 ± 12.4	-16.2 ± 8.4	16.9 ± 13.0	-13.9 ± 8.4	12.8 ± 12.7
$\mathrm{B^0}$ Tag	-1.3 ± 7.1	4.7 ± 7.1	-11.1 ± 8.2	6.0 ± 21.7	-15 ± 8	9 ± 15	-13 ± 8	7 ± 14
$\overline{\mathrm{B^0}}$ Tag	-5.6 ± 7.3	-2.1 ± 7.3	-8.3 ± 8.4	1.0 ± 8.2	-12.6 ± 8.4	4.5 ± 18.2	-10.5 ± 8.4	-0.6 ± 14.3
Lepton Tag	-33 ± 14	-29 ± 14	-31 ± 15	-25 ± 16	-35 ± 15	-24 ± 15	-33 ± 15	-25 ± 13
Kaon Tag	-0.5 ± 8.6	5.0 ± 8.6	-2.4 ± 9.4	2.5 ± 5.0	-10 ± 10	8 ± 15	-8 ± 10	5.9 ± 7.9
NT1 Tag	-12 ± 17	-8 ± 17	-11 ± 20	-16 ± 23	-14 ± 20	-9 ± 31	-12 ± 20	12 ± 27
NT2 Tag	5 ± 13	9 ± 13	-24 ± 16	-38 ± 27	-29 ± 15	-27 ± 28	-27 ± 16	-38 ± 27
NT3 Tag	5 ± 9	9 ± 9	5 ± 11	47 ± 23	1 ± 11	59 ± 36	3 ± 11	48 ± 19
Perfect Tag	-4.7 ± 5.9	0.7 ± 5.93	-2.4 ± 1.4	4.7 ± 9.8	-18 ± 6.8	1.4 ± 9.9	-15.5 ± 6.8	-1.7 ± 11.9
Imperfect	0.4 ± 9.9	3.0 ± 9.9	2.0 ± 11.7	35 ± 16	-2.6 ± 12	35 ± 29	-1 ± 12	22 ± 19

Table 4: Results of fitting lifetimes to various resolution models to the (signal-side) correctly reconstructed signal MC sample. In the first column of results, we give the difference between the lifetime extracted from a fit to the true Δt distribution and the generated lifetime (1548 fs). The second column is the difference between the lifetime extracted from the true Δz distribution converted to fs using the boost approximation and the generated lifetime. In all the other columns of results, we give the difference between the lifetime extracted from a fit to reconstructed Δt and the fit to true Δt (column 1, not the generated value of 1548 fs). For the $\Delta\tau$ Only column, the parameters for the resolution function are fixed to the values extracted from the fit to residuals and errors; for the $\Delta\tau$+Resln column the resolution parameters are floating in the fit. All results have units of fs.

The remaining columns of Table 4 show the results of lifetime fits to the reconstructed Δt distributions. The three pairs of columns show the results of fits to the G+G+G, G+G and GExp+G models. The first column of each pair is the result of a fit to the lifetime only, with the resolution model parameters fixed at the values given in Table 72 and 73. The second column is a fit to the same data but with both the lifetime and the resolution parameters floating. Figure 62 shows the Δt distributions for these fits to the full signal MC sample on linear (left-hand plots) and log scales (right-hand plots). The solid curves are the G+G+G fits and the dashed curves are GExp+G fits. The upper two plots are fits to the lifetime only, and the bottom two plots are the full fits to lifetime and the resolution parameters.

Figures 64 and 65 show the lifetimes extracted from the fits to the full signal MC samples with the three resolution models. There is good agreement in the changes in the lifetime observed with different subsamples. We will use all three models when fitting data as a consistency check and to estimate systematic uncertainties.

As we can see in Table 4, fitting the lifetime using the resolution from residual fits gives significantly lower lifetimes than the simultaneous fit to lifetime and resolution function. This implies that our resolution model may not be totally accurate. We already know that the RMS of the residual $\Delta z_{\mathrm{true}} - \Delta t_{\mathrm{true}}$ depends on $|\Delta t|$. We can check the significance of this dependence at the level of reconstructed Δt. The top plot in Fig. 11 shows the RMS of the pull of reconstructed Δt in bins of true Δt. We can clearly see the dependence on Δt_{true}. The middle plot shows the same sort of distribution except that Δt_{meas} is replaced by $\Delta z_{\mathrm{true}}/\gamma\beta\gamma^* c$ in the calculation of the RMS of the pull. The bottom plot is the bin-by-bin difference in quadrature between the top two histograms. The fact that the dependence of the pull RMS on Δt_{true} almost disappears after removing the effect due to boost in the conversion from Δz to Δt suggests that the boost approximation is the major source of the dependence of resolution on Δt_{true}. Since this dependence can be fully understood, in principle we can modify our model to accommodate this effect.

We study the effect of not modeling this Δt_{true} dependence of resolution with toy Monte Carlo. We generate a set of datasets with the same number of events as in the full signal Monte Carlo. The pull of Δt is smeared by double-Gaussian or GExp resolution models, whose RMS varies as a function of Δt_{true} according to the shape found in signal Monte Carlo (first plot in Fig. 11 [5].) We then use a simple double-Gaussian or GExp model, ignoring Δt_{true} dependence, to fit the residual, and τ, Δm_d, and mistag fraction with and without fixing the resolution to the fits to the residual. The comparisons of the fit results to the fits to the truth Δt distribution are shown in Fig. 12 and 13. The mean values of bias on lifetime are listed in Table 5. In the table we also compare them to the bias in fits to the full signal Monte Carlo [12].

Some conclusions from the toy studies...

[5] We fit this plot with $p_0\sqrt{(p_1 + |\Delta t_{\mathrm{true}}|)^2 + p_2^2}$. The numbers used in the toy MC are $p_0 = 0.0737$, $p_1 = 4.45$, $p_2 = 16$.

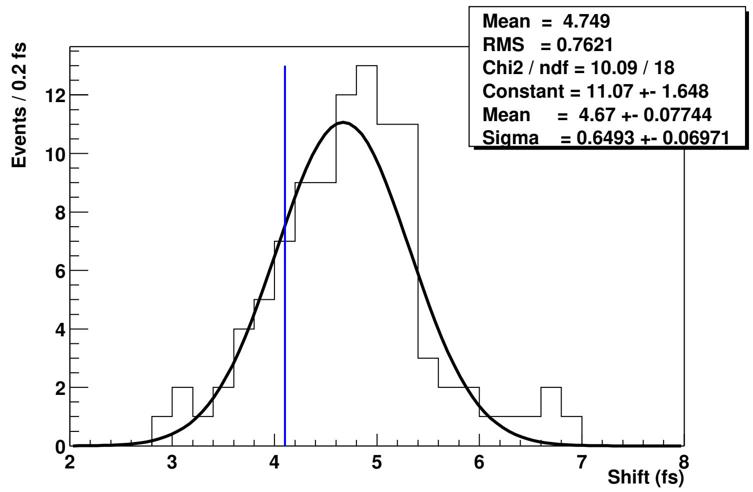

Figure 9: Shift of $\langle|\Delta z|/\beta\gamma\gamma^{*}c\rangle$ from $\langle|\Delta t|\rangle$ with 100 toy experiments. Blue line indicates the value found in signal MC fit.

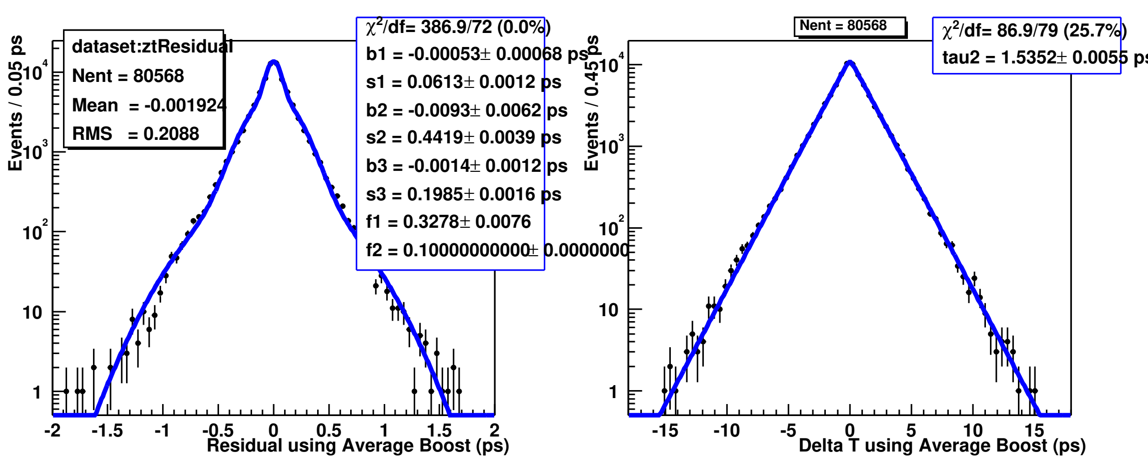

Figure 10: Fits to the signal MC residual (left) of $\Delta z_{\text{truth}}/\gamma\beta\gamma^{*}c - \Delta t_{\text{truth}}$ and to $\Delta z_{\text{truth}}/\gamma\beta\gamma^{*}c$ with the resolution fixed to the fit to residual to extract lifetime.

	GExp Toy	G+G Toy	GExp+G Full MC	G+G+G Full MC
Fix Resol.	-4.623	-5.514	-7.9 ± 6.3	-5.8 ± 6.3
Float Resol.	4.205	-2.382	7.5 ± 9.1	3.2 ± 9.4

Table 5: Biases on τ using toy MC (shown in Fig.12 and 13), and full fits shown in [12].

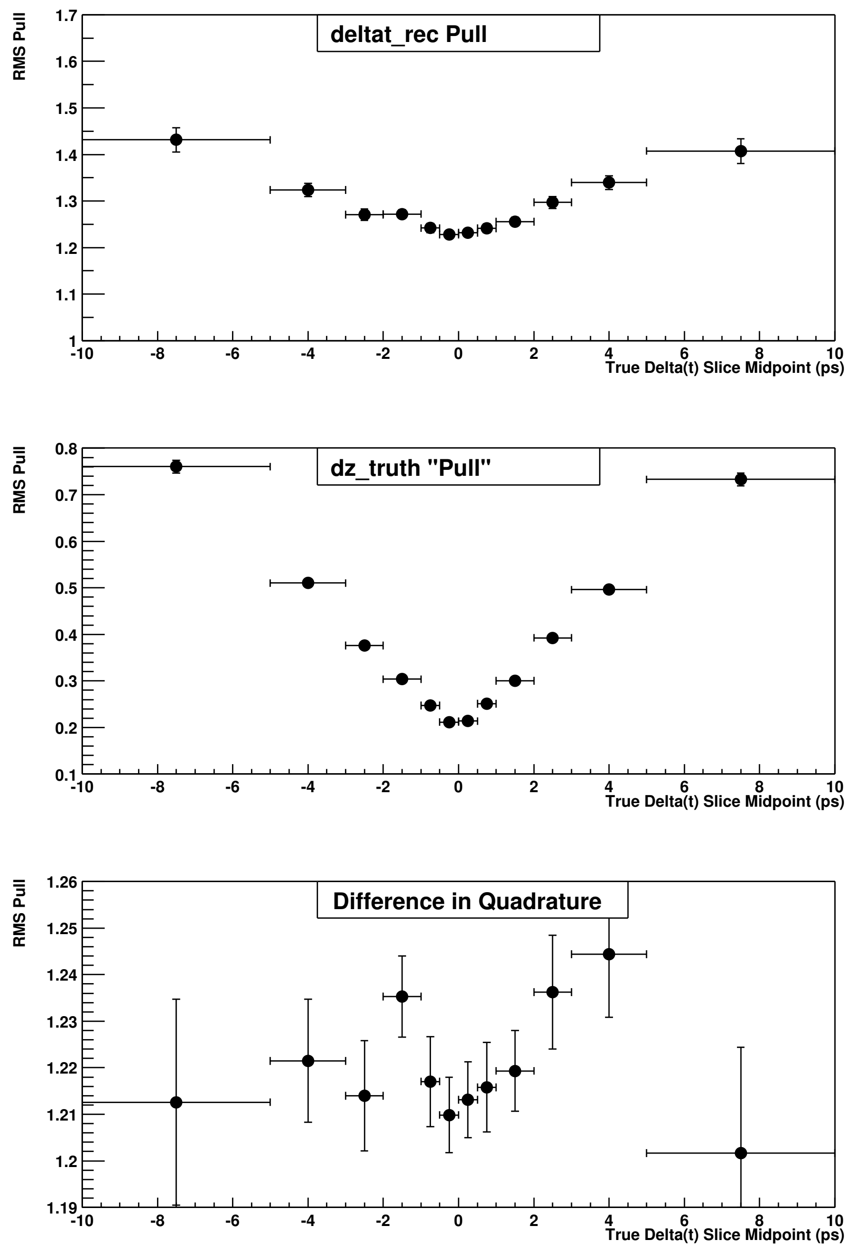

Figure 11: Top: the RMS of the pull $((\Delta t - \Delta t_{\text{true}})/\sigma_{\Delta t})$ in bins of Δt_{true} for signal MC. Middle: same plot with Δt replaced by $\Delta z_{\text{true}}/\gamma\beta\gamma^* c$ in the calculation of the RMS of the pull. Bottom: bin-by-bin difference in quadrature between the top two histograms. Note the factor of 10 difference in the vertical scales of the top and bottom histograms.

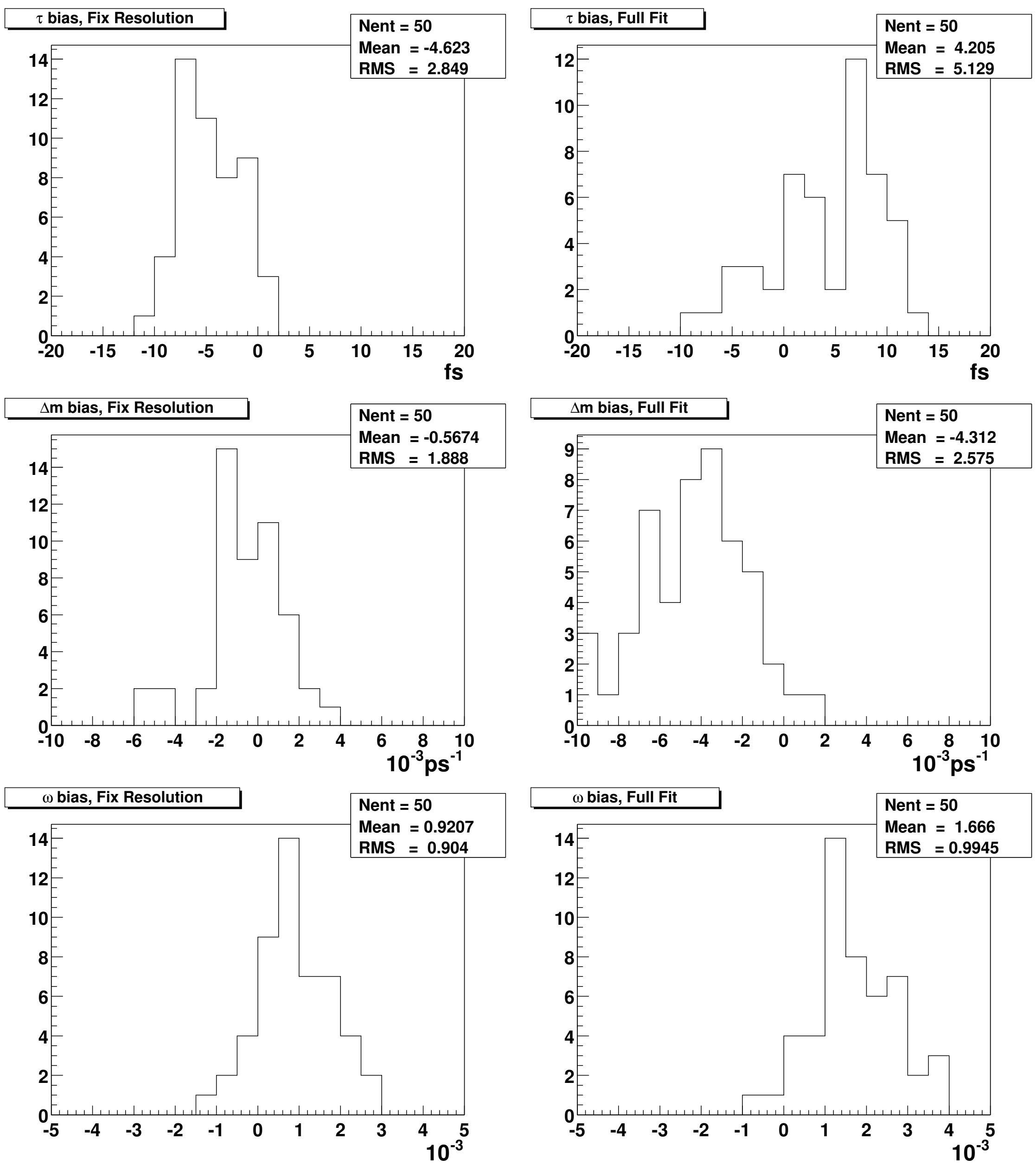

Figure 12: Biases due to ignoring the resolution dependence on Δt_{true} with toy Monte Carlo study, using GExp model.

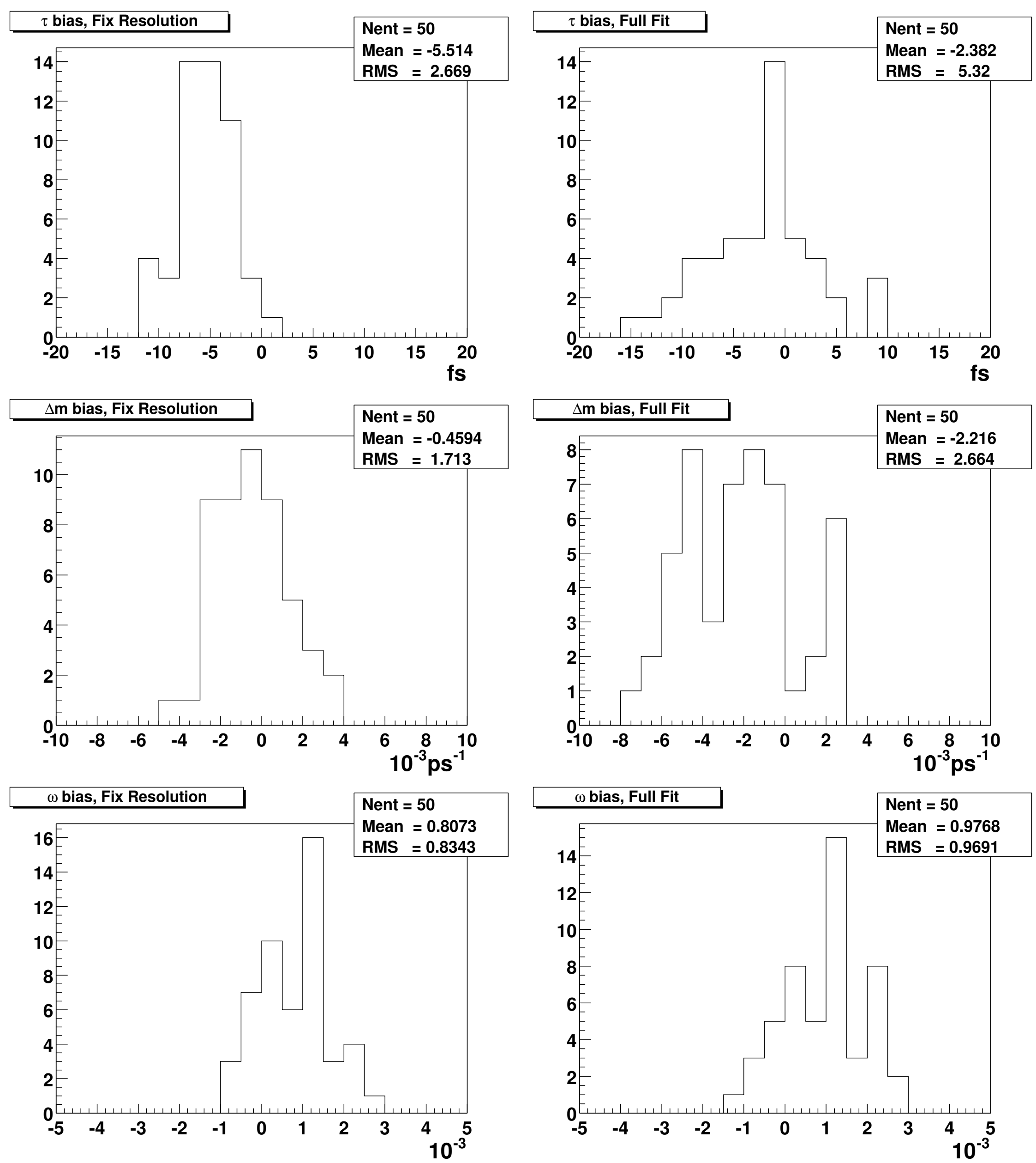

Figure 13: Biases due to ignoring the resolution dependence on Δt_{true} with toy Monte Carlo study, using double-Gaussian model.

4 Flavor Tagging

Flavor tagging is a critical element for the oscillation frequency measurement, since it identifies the mixing state (mixed or unmixed) for each candidate. It can also be used to increase the sensitivity of the lifetime measurement since some tagging categories are purer than others, and to decrease bias if the resolution parameters and δm shapes are different for backgrounds in different tagging categories. *(Refer to DK's talk.)*

Once the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate is found, all remaining tracks are used to determine the flavor of the other B. The tagging package `BTaggingTools` uses these tracks and returns a tag flavor and tagging category.

4.1 Tagging Categories

Events can fall into one of five different categories:

- `lepton`: This is the tagging category with the lowest mistag fraction. It uses the correlation between the charge of the primary lepton and the flavor of the b quark. A momentum cut of 1.1 GeV helps to reduce contamination due to secondary leptons from semileptonic charm decays.
- `kaon`: Uses the correlation between the charge of the kaon and the b quark flavor in the most probable decay chain $\mathrm{b} \to \mathrm{c} \to \mathrm{s}$.
- `NT1`: Events for which the absolute value of the NetTagger neural net output (described below) is greater than 0.5.
- `NT2`: Events for which the absolute value of the neural net output is between 0.2 and 0.5.
- `NT3`: Events for which the absolute value of the neural net output is less than 0.2.

The order of tagging categories in the above list corresponds to the hierarchy of tagging categories. If an event has a lepton tag and no conflicting kaon tag information, it will fall into the `lepton` category. Events with no lepton information but a kaon tag will fall into the `kaon` category. All other events (those with either no lepton and kaon tags or conflicting lepton and kaon tags) are passed to a neural network algorithm called `NetTagger`, which uses other features that can be used to determine the flavor of the tagging B. These include soft pions from D^* decays, high momentum primary leptons that are not selected by the electron or muon identification algorithms, lower momentum primary leptons, and charged kaons that are not selected by the kaon identification alogorithm. For a complete description of these tagging algorithms see BABAR Analysis Document 119.

The categories described above are the same as the "Elba Tagger", with the addition of NT3. The real benefit of NT3 comes from the additional events used to determine resolution parameters in the fit. The actual tagging power of NT3 is very small.

4.2 Tagging Performance

The tagging performance for each tagging category is characterized by an effective tagging efficiency $Q_i = \epsilon_i(1-2w_i)^2$, where ϵ_i is the fraction of events associated to the tagging category i and w_i is the mistag fraction – the probability of incorrectly assigning the opposite tag to an event of this category. The total effective tagging efficiency is the sum over categories: $Q = \sum Q_i$.

The tagging algorithm was optimized, and the neural networks were trained, on Monte Carlo simulations only. However, the performance can be measured directly from the data. The tagging efficiencies in signal Monte Carlo, extracted from Table 3, and in data, extracted from Table 6 are given in Table 7.

Category	Peak Yield	f_{fake}	$f_{\text{continuum}}$	$f_{\text{uncorr.}}$
LTag	1625 ± 42	1.4%	0%	3.0%
KTag	5244 ± 79	1.6%	2.8%	2.8%
NT1	1238 ± 38	1.5%	1.1%	4.3%
NT2	1995 ± 49	1.5%	2.5%	3.4%
NT3	4335 ± 73	1.5%	1.4%	2.1%

Table 6: Peak yields and peaking background fractions for tagging categories from Data.

Category	MC Raw Numbers	ϵ (%)	Data Raw Numbers	ϵ (%)
`Kaon`	28707/81773	35.1±0.2	4865/13538	35.9
`Lepton`	10824/81773	13.2±0.1	1554/13538	11.5
`NT1`	6791/81773	8.3±0.1	1152/13538	8.5
`NT2`	11850/81773	14.5±0.1	1847/13538	13.6
`NT3`	23601/81773	28.9±0.2	4120/13538	30.4
Total	81773/81773	100	13538/13538	100

Table 7: Tagging efficiencies for each tagging category. The efficiencies in signal Monte Carlo and Data are extracted from the sample sizes in Tables 3 and 6.

The effect of mistags is to reduce the amplitude of the oscillation in the mixing asymmetry

$$A(\Delta t) = \frac{N_{unmix}(\Delta t) - N_{mix}(\Delta t)}{N_{unmix}(\Delta t) + N_{mix}(\Delta t)}$$

by the dilution factor $(1-2w_i)$. The simultaneous fit to the mixing frequency and its amplitude allows us to determine both Δm_d and the mistag rates w_i in each tagging category.

In BAD 119, the tagging performance is documented for the event samples used for the $\sin 2\beta$ analysis and for the mixing analysis based on fully-reconstructed hadronic events, and for a sample of semileptonic events used with a time-integrated technique. We do not expect the tagging performance to be the same for the analysis described in this document because we require two or more tracks to be used in the vertex for the tagging B.

The mistag rates from the final fit to data are given in Table 29, to signal Monte Carlo in Table 29 and to full generic Monte Carlo in Table 34.

4.3 Correlation Between Mistag and Δt Error

A strong correlation between the mistag fraction and $\sigma_{\Delta t}$ for kaon tagging category is found. As shown in Fig. 14 for signal Monte Carlo sample, mistag fraction for KTag increases almost linearly for small $\sigma_{\Delta t}$ and levels off at larger $\sigma_{\Delta t}$, while mistag fractions for other categories are fairly independent of $\sigma_{\Delta t}$. The source of this dependence for kaon category is documented elsewhere [11]. This correlation is mostly due to the differences of decay dynamics between decays that produce right-sign and wrong-sign kaons, e.g., from double charm decays or wrong-sign kaon from D^+. The momentum spectrum of charged tracks from decays that produce wrong-sign kaons are softer. This contributes to larger errors in tag-side vertexing. Therefore, the fraction of wrong-sign decays for events with larger $\sigma_{\Delta t}$, is larger than for those with smaller $\sigma_{\Delta t}$.

In order to optimize the sensitivity to physics quantities, we will account for this correlation by letting the mistag fraction in kaon category be a linear function of $\sigma_{\Delta t}$ in our signal PDF, and let the slope and interception be free parameters.

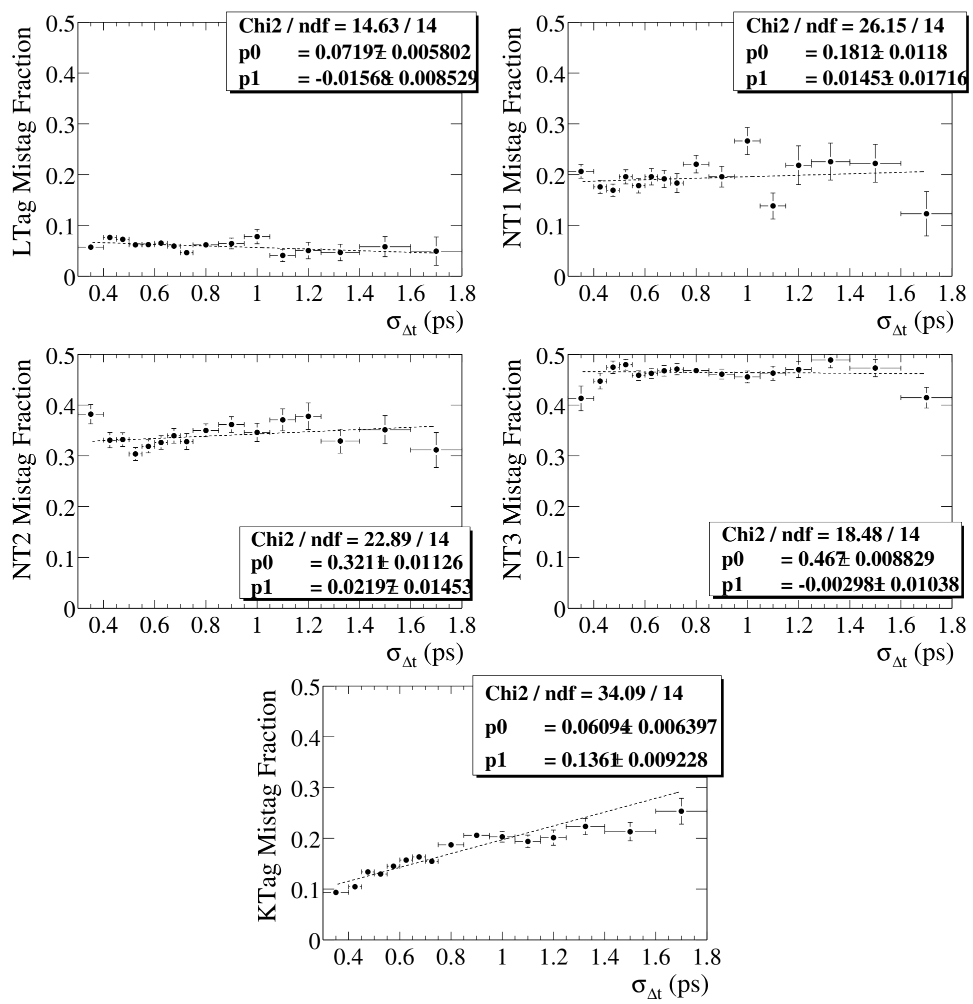

Figure 14: Mistag fractions in slices of $\sigma_{\Delta t}$ for five tagging categories for true signal events in signal Monte Carlo sample. The values are calculated by counting the numbers of incorrectly tagged events by comparing the observed mixing status and MC truth. The plots are fitted with a straight line. Only KTag shows a significant slope.

5 Fit Method

We perform an unbinned fit simultaneously to events in each of the 12 signal and control samples (indexed by s) that are further subdivided into 30 subsamples (indexed by c)[6] using a likelihood

$$\mathcal{L} = \prod_{s=1}^{12} \prod_{c=1}^{30} \prod_{k=1}^{N(s,c)} P_{s,c}(\vec{x}_k\,; \vec{p}) \;, \tag{11}$$

where k indexes the $N(s,c)$ events $\vec{x}_k$ in each of the 360 subsamples. The probability $P(\vec{x}\,;\vec{p})$ of observing an event $\vec{x} = (\delta m, \Delta t, \sigma_{\Delta t}, g, s, c)$ is calculated as a function of the parameters $\vec{p} = (f_{s,c}^{(comb)}, \vec{p}_{s,c}^{(comb)}, \vec{p}_{c}^{(peak)}, \vec{q}_{s,c}^{(comb)}, f_{s,c,1}^{(pkg)}, f_{s,c,2}^{(pkg)}, f_{s,c,3}^{(pkg)}, \vec{q}_{s,c,1}^{(pkg)}, \vec{q}_{s,c,2}^{(pkg)}, \vec{q}_{s,c,3}^{(pkg)}, \vec{q}_{c}^{(sig)})$ as

$$\begin{aligned} &P(\delta m, \Delta t, \sigma_{\Delta t}, g, s, c\,;\vec{p}) = \\ &f_{s,c}^{(comb)} \cdot \mathcal{F}^{(comb)}(\delta m\,; \vec{p}_{s,c}^{(comb)}) \cdot \mathcal{G}^{(comb)}(\Delta t, \sigma_{\Delta t}, g\,; \vec{q}_{s,c}^{(comb)}) + \left(1 - f_{s,c}^{(comb)}\right) \cdot \mathcal{F}^{(peak)}(\delta m\,; \vec{p}_{c}^{(peak)}) \cdot \\ &\quad \left[\sum_{j=1}^{3} f_{s,c,j}^{(pkg)} \cdot \mathcal{G}_j^{(pkg)}(\Delta t, \sigma_{\Delta t}, g\,; \vec{q}_{s,c,j}^{(pkg)}) + \left(1 - \sum_{j=1}^{3} f_{s,c,j}^{(pkg)}\right) \cdot \mathcal{G}^{(sig)}(\Delta t, \sigma_{\Delta t}, g\,; \vec{q}_{c}^{(sig)})\right] \;, \end{aligned} \tag{12}$$

where j indexes the 3 sources of peaking background. The index g identifies whether the event is tagged as mixed or unmixed. By allowing different effective mistag rates for apparently mixed or unmixed events in the background functions $\mathcal{G}^{(comb)}$ and $\mathcal{G}^{(pkg)}$, we accomodate the different levels of backgrounds in mixed and unmixed, shown in Table 14 on p. 41 in BAD 147, Version 4. Functions labeled with $\mathcal{F}$ describe the probability of observing a particular value of δm while functions labeled with $\mathcal{G}$ give probabilities for values of Δt and $\sigma_{\Delta t}$ in category g. Parameters labeled with f describe the relative contributions of different types of events. Parameters labeled with $\vec{p}$ describe the shape of a δm distribution, and those labeled with $\vec{q}$ describe a Δt, $\sigma_{\Delta t}$ shape. Refer to Tables 8 and 9 for more detailed descriptions of the functions and parameters appearing in the likelihood. Note that we make explicit assumptions that the δm peak shape (parameterized by $\vec{p}_{c}^{(peak)}$) and the signal Δt,$\sigma_{\Delta t}$ shape (parameterized by $\vec{q}_{c}^{(sig)}$) depend only on the subsample index c. The first of these assumptions is supported by data, and simplifies the analysis of peaking background contributions. The second assumption reflects our expectation that the Δt distribution of signal events does not depend on whether they are selected in the signal sample or appear as a background in a control sample.

The ultimate aim of the fit is to obtain the B^0 lifetime and mixing parameters, which by construction are common to all sets of signal parameters $\vec{q}_{c}^{(sig)}$. Most of the statistical power for determining these parameters comes from the signal sample, although the fake and uncorrelated background control samples also contribute due to their signal content.

[6] See Sec. 2.4 for a description of the 12 control samples and 30 subsamples.

Function	Description
P	Overall probability to observe an event
$\mathcal{F}^{(comb)}$	Probability of observing δm for a combinatoric BG event
$\mathcal{F}^{(peak)}$	Probability of observing δm for an event with a real D*
$\mathcal{G}^{(comb)}$	Probability of observing Δt, $\sigma_{\Delta t}$ for a combinatoric BG event
$\mathcal{G}_1^{(pkg)}$	Probability of observing Δt, $\sigma_{\Delta t}$ for a continuum BG event
$\mathcal{G}_2^{(pkg)}$	Probability of observing Δt, $\sigma_{\Delta t}$ for a fake-lepton BG event
$\mathcal{G}_3^{(pkg)}$	Probability of observing Δt, $\sigma_{\Delta t}$ for an uncorrelated D$^*\ell$ BG event
$\mathcal{G}^{(sig)}$	Probability of observing Δt, $\sigma_{\Delta t}$ for a signal event

Table 8: Description of the functions appearing in the likelihood equation 11. Functions labeled with $\mathcal{F}$ calculate a probability of observing a particular value of δm while functions labeled with $\mathcal{G}$ give probabilities for values of Δt and $\sigma_{\Delta t}$.

Parameter	Description
$f_{s,c}^{(comb)}$	Fraction of combinatoric events in subsample (s, c) (normalized to all subsample events in the δm window)
$f_{s,c,j}^{(pkg)}$	Fraction of peaking BG events of type j in subsample (s, c) (normalized to subsample events in the δm peak)
$\vec{p}_{s,c}^{(comb)}$	Combinatoric BG δm shape parameters for subsample (s, c)
$\vec{p}_c^{(peak)}$	δm Peak shape parameters for subsample (c)
$\vec{q}_{s,c}^{(comb)}$	Combinatoric BG Δt,$\sigma_{\Delta t}$ shape parameters for subsample (s, c)
$\vec{q}_{s,c,j}^{(pkg)}$	Peaking BG type-p Δt,$\sigma_{\Delta t}$ shape parameters for subsample (s, c)
$\vec{q}_c^{(sig)}$	Signal Δt,$\sigma_{\Delta t}$ shape parameters for subsample (c)

Table 9: Description of the parameters appearing in the likelihood equation 11. Parameters labeled with f describe the relative contributions of different types of events. Parameters labeled with $\vec{p}$ describe the shape of a δm distribution, and those labeled with $\vec{q}$ describe a Δt, $\sigma_{\Delta t}$ shape.

In principle, we would like to fit with all parameters in Table 9 floating since this approach provides the most straightforward and reliable estimate of the systematic errors due to the assumed values of all secondary parameters. In practice, we are limited by the time required for a fit with many free parameters. There is also a technical limitation that we cannot float the peaking background fractions $f_{s,c,j}^{(pkg)}$ since their values are calculated in terms of other parameters as well as external inputs (various efficiencies and the on-off peak scale factor). Our approach is to bootstrap the full fit with a sequence of fits using reduced likelihood functions, as outlined in Table 10. As a final step, we float as many parameters as possible, in addition to floating sets of secondary parameters one at a time together with the primary signal parameters, in a simultaneous fit.

Step	Output	Description
1	$\vec{p}_{s,c}^{\,(comb)}, \vec{p}_{c}^{\,(peak)}$	simultaneous δm shape fit to all events
2	$f_{s,c}^{(comb)}$	δm yield fits to each (s, c) subsample
3	$f_{s,c,p}^{(pkg)}$	calculation of peaking contributions using external inputs
4	$\vec{q}_{s,c}^{\,(comb)}$	simultaneous Δt, $\sigma_{\Delta t}$ shape fit to all sideband events
5		pre-calculation of per-event probabilities $\mathcal{F}^{(peak)}$, $\mathcal{F}^{(comb)}$ and $\mathcal{G}^{(comb)}$
6	$\vec{q}_{c}^{\,(sig)}$	simultaneous Δt, $\sigma_{\Delta t}$ shape fit to signal peak events
7	$\vec{q}_{s,c,1}^{\,(pkg)}$	simultaneous Δt, $\sigma_{\Delta t}$ shape fit to off-resonance peak events
8	$\vec{q}_{s,c,2}^{\,(pkg)}$	simultaneous Δt, $\sigma_{\Delta t}$ shape fit to on-resonance,OS,fake peak events
9	$\vec{q}_{s,c,3}^{\,(pkg)}$	simultaneous Δt, $\sigma_{\Delta t}$ shape fit to on-resonance,SS,lepton peak events
10	$\vec{q}_{c}^{\,(sig)}, \vec{q}_{s,c,p}^{\,(pkg)}$	simultaneous Δt, $\sigma_{\Delta t}$ shape fit to all peak events (starting from background shapes from previous steps)

Table 10: Description of the steps used to bootstrap the full fit.

The first three steps are described in BAD 147 and summarized in Sec. 2.4. In the following sections, we describe the remaining steps in more detail. There are two goals in each step:

1. Find a model that describes the Δt distribution for each type of event: signal, combinatoric background, and the three types of backgrounds that peak in the D*-D mass difference distribution (a true D* with either a fake lepton, an "uncorrelated" lepton from the other B in the event, or a lepton from continuum events). To establish a model, we use Monte Carlo samples that have been selected to correspond to only one type of signal or background event based on Monte Carlo truth information. These samples are used to determine the Δt model and the categories of events (*e.g.,* tagging category, fake or real lepton) that can share each of the parameters in the model. The fitting software (`RooFit`) allows any subset of parameters to be shared among any subset of categories. The goal is to minimize the number of different parameters while still providing an adequate description of the Δt distribution. This sharing of a particular set of parameters among a particular set of categories is referred to as "parameter splitting" or "parameter breakdown" in the following sections, depending on who wrote the section.

2. Find the starting values for the background parameters by fitting to each of the background-enhanced control samples in data, using the model (and splitting of parameters) determined in the previous step. Since these background control samples are not pure, we start with the purest control sample (combinatoric background events from the δm sideband) and move on to less pure control samples, always using the models established from earlier steps to describe the Δt distribution of the contamination from other backgrounds. We also test this method of fitting to impure background control samples with generic Monte Carlo samples.

Note that the result of the above two steps is a Δt model for each type of event and a set of starting values for all parameters in the fit. However, when we do the final fit, we fit all signal and control samples simultaneously ($\approx$68k events), leaving essentially all parameters free in the fit (over 70 free parameters).

5.1 Parameter Notation

The final model contains about 70 parameters. Some parameters are split according to different combinations of categories. To avoid confusion, we summarize the notation used in this note for parameters in the Δt models. In general, a parameter is denoted by

$$p_{\mathrm{Cat}}^{\mathrm{label}},$$

where the subscript represents the category values (if any) to which this parameter is restricted, and the superscript is an extra label for distinguishing different roles of a certain type of parameter in a model, if necessary.

Recall the motivation for "splitting" a parameter: in some cases, we expect different groups of events to behave according to the same Δt model, but with different parameter values. By "splitting" a parameter according to a certain candidate property, we assign the parameter a domain of context or relevance that is the subset of data possessing a certain value of that property, and create additional instances of the parameter for the other subsets of the data with different values for that property. For instance, we often "split" resolution model parameters by tagging category. This means that each tagging category will have identical resolution model structure, but that certain parameters of that model will be unique to the LTag or KTag events, etc.

For the signal physics model ($G^{(\mathrm{sig})}$ in Sec. 5.2),

- τ_{B^0} and Δm_d represent B^0 lifetime and mixing frequency, respectively.
- The mistag fraction is noted as ω and is split by tagging category (LTag, KTag and NTx are short-handed as "LT", "KT", and "Nx") :
    - ω_{LT}, ω_{KT}, ω_{N1}, ω_{N2}, ω_{N3} are the mistag probabilities averaged over B and $\overline{\mathrm{B}^0}$.

– Mistag for KTag is a linear function of $\sigma_{\Delta t}$: $\omega_{\mathrm{KT}} = m_{\mathrm{KT}} \cdot \sigma_{\Delta t} + \omega_{\mathrm{KT}}^{\mathrm{offset}}$. The slope m_{KT} is always given in units of ps^{-1}.

– $\Delta\omega$ is the difference in mistag probabilities for B and $\overline{\mathrm{B^0}}$ (see next section).

- $f_{\mathrm{B}^\pm}$ represents the charged B fraction in signal events (Sec. 5.4.4).

For each of the background Δt models, we use a "physics model" convoluted with a "resolution function" to describe the Δt distribution of the background events. Although for background events, the true Δt and resolution are not well-defined, we still use the names "physics model" and "resolution function" just to refer to the mathematical form of the Δt models.

A background Δt model consists of a prompt term plus either a lifetime term or an oscillating term, or a prompt term by itself. These three terms can be written as

$$\begin{aligned}
\mathcal{G}_{\mathrm{phys}}^{\mathrm{pmt}}(\Delta t, g) &= (1/2) \cdot \delta(\Delta t) \cdot \big(1 + g \cdot (1 - \omega^{\mathrm{pmt}})\big) , \\
\mathcal{G}_{\mathrm{phys}}^{\mathrm{life}}(\Delta t, g) &= (1/4) \cdot \exp(-|\Delta t|/\tau^{\mathrm{bg}}) \cdot \big(1 + g \cdot (1 - \omega^{\mathrm{life}})\big) , \\
\mathcal{G}_{\mathrm{phys}}^{\mathrm{osc}}(\Delta t, g) &= (1/4) \cdot \exp(-|\Delta t|/\tau^{\mathrm{bg}}) \cdot \big(1 + g \cdot (1 - \omega^{\mathrm{osc}}) \cos \Delta m^{\mathrm{bg}} \Delta t\big) ,
\end{aligned}$$

where $\delta(\Delta t)$ is a δ-function, g is a sign representing mixing status, i.e., $g = +1\,(-1)$ for unmixed (mixed) events, and τ^{bg} and Δm^{bg} are effective lifetime and mixing frequency for this particular background. The combined physics model for a background model, then, is typically one of the following:

- $f^{\mathrm{life}} \cdot \mathcal{G}_{\mathrm{phys}}^{\mathrm{life}} + (1 - f^{\mathrm{life}}) \cdot \mathcal{G}_{\mathrm{phys}}^{\mathrm{pmt}}$
- $f^{\mathrm{osc}} \cdot \mathcal{G}_{\mathrm{phys}}^{\mathrm{osc}} + (1 - f^{\mathrm{osc}}) \cdot \mathcal{G}_{\mathrm{phys}}^{\mathrm{pmt}}$
- $\mathcal{G}_{\mathrm{phys}}^{\mathrm{pmt}}$

We use three different resolution functions for Δt models, GExp+G, G+G+G and G+G functions. The basic ingredients are a Gaussian, G (Eq. 5, Sec. 3.4) and the sum of a single Gaussian and the same Gaussian convoluted with a one-sided ($\delta\Delta t < 0$) exponential, $G \otimes E$ (Eq. 6, Sec. 3.4).

$$\begin{aligned}
&\mathcal{R}_{\mathrm{G+G}}(\delta\Delta t, \sigma_{\Delta t}; b^1, s^1, f^1, b^{\mathrm{out}}, s^{\mathrm{out}}) \\
&= f^1 \cdot G(b^1\sigma_{\Delta t}, s^1\sigma_{\Delta t}) + (1 - f^1) \cdot G(b^{\mathrm{out}}\sigma_{\Delta t}, s^{\mathrm{out}}\sigma_{\Delta t}) , \\
&\mathcal{R}_{\mathrm{G+G+G}}(\delta\Delta t, \sigma_{\Delta t}; b^1, s^1, f^1, b^2, s^2, b^{\mathrm{out}}, s^{\mathrm{out}}, f^{\mathrm{out}}) \\
&= f^1 \cdot G(b^1\sigma_{\Delta t}, s^1\sigma_{\Delta t}) + (1 - f^1 - f^{\mathrm{out}}) \cdot G(b^2\sigma_{\Delta t}, s^2\sigma_{\Delta t}) + f^{\mathrm{out}} \cdot G(b^{\mathrm{out}}, s^{\mathrm{out}}) , \\
&\mathcal{R}_{\mathrm{GExp+G}}(\delta\Delta t, \sigma_{\Delta t}; s^1, \kappa, f^1, b^{\mathrm{out}}, s^{\mathrm{out}}, f^{\mathrm{out}}) \\
&= f^1 \cdot G(0, s^1\sigma_{\Delta t}) + (1 - f^1 - f^{\mathrm{out}}) \cdot G(0, s^1\sigma_{\Delta t}) \otimes E(\kappa\sigma_{\Delta t}) + f^{\mathrm{out}} \cdot G(b^{\mathrm{out}}, s^{\mathrm{out}}) ,
\end{aligned}$$

where the dependent variable $\delta\Delta t$ in G and E is not shown explicitly. The function E is defined as $E(x;a) = \exp^{x/a}$. (Note that we explicitly do not scale the bias and width parameters of the third term, the outlier Gaussian, when present. This is because it addresses physics events which are so poorly misreconstrcuted (mistaken, actually), that the vertex error $\sigma_{\Delta t}$ has little to do with the true uncertainty.)

The same type of function may be used for different backgrounds, therefore the same symbol will represent different parameters in different background models. In principle we should add another label to distinguish them, but this will make the notation very cumbersome. Fortunately we rarely need to put different backgrounds in the same context. So we will not add this extra label unless it is necessary for clarity.

Finally, we typically represent the change in a parameter due to a variation in the fit method, data sample or some other procedure as $\delta(\text{par}) = \text{par}^{\text{changed}} - \text{par}^{\text{unchanged}}$.

5.2 Signal Δt Model

We use the general model described in BAD 188 [2]. For a given tagging category c, the PDF consists of a physics model and a resolution function:

$$\mathcal{G}_c^{(\text{sig})}(\Delta t, \sigma_{\Delta t}, g\,; \vec{q}_c^{(\text{sig})}) = \left\{ \frac{1}{4\tau_{\text{B}^0}} e^{-|\Delta t|/\tau_{\text{B}^0}} (1 + g(1-2\omega_c)\cos(\Delta m_d \Delta t)) \right\} \otimes \mathcal{R}(\delta\Delta t, \sigma_{\Delta t}; \vec{q_r}) \,,$$

where $\mathcal{R}$ is a resolution function, which can be different for different event categories, and g is $+1$ (-1) for unmixed (mixed) events. The physics model shown in the above equation has seven parameters, Δm_d, τ_{B^0} and mistag fraction ω_c for each of the five tagging categories. To account for the correlation between the kaon mistag and $\sigma_{\Delta t}$, we allow the mistag in kaon category to vary as a linear function of $\sigma_{\Delta t}$ (see Sec. 4.3):

$$\omega_{\text{KT}} = m_{\text{KT}} \cdot \sigma_{\Delta t} + \omega_{\text{KT}}^{\text{offset}} \,.$$

In addition, we allow the mistag fractions for B^0 tags and $\overline{\text{B}^0}$ tags to be different. We define $\Delta\omega = \omega_{\text{B}^0} - \omega_{\overline{\text{B}^0}}$ and $\omega = (\omega_{\text{B}^0} + \omega_{\overline{\text{B}^0}})/2$, so that

$$\omega_{\text{B}^0/\overline{\text{B}^0}} = \omega \pm \frac{1}{2}\Delta\omega \,.$$

In the end there are 13 free parameters in the physics model.

For the resolution function, depending on the model and how parameters are either shared or not shared ("split") among event categories, the number of parameters can range from five to more than 20. (Note that we often fix one or both of the outlier shape parameters, b^{out} and/or s^{out}.)

5.2.1 Fit to Monte Carlo Signal Events

In this section, we fit the model to correctly reconstructed B^0 signal MC events in the signal sample and all control samples. We include signal events that end up in background control samples because in the final fit to data, all signal events in the signal and control samples are described by the signal PDF. We study the GExp plus Gaussian outlier (GExp+G) model and the triple Gaussian model (G+G+G) and how parameters should be split for different event categories. The bias, width and decay constant (in GExp+G model) in the resolution models are all scaled by $\sigma_{\Delta t}$.

The number of signal MC events that are correctly reconstructed is 92879. We first fit the true Δt and tagging information and count the mistag fraction. The result is shown in Table 11.

τ_{B^0} (ps)	Δm_d (ps^{-1})	ω_{LT} (%)	ω_{KT} (%)	ω_{N1} (%)	ω_{N2} (%)	ω_{N3} (%)
1.5446 ± 0.0051	0.4708 ± 0.0013	6.3	15.7	19.2	33.8	46.5

Table 11: Results for τ_{B^0} and Δm_d as fit to the true Δt and tagging information from signal MC. Mistag fractions are calculated from the number of wrongly tagged events according to the MC truth.

When different categories of events appear to have different Δt resolution, we allow the parameters in the resolution functions to be different. resolution functions for them. The mean and rms of Δt residual distributions vary significantly among tagging categories (see Table 3 and Table 71 in Appendix D). To account for these differences, we could let the following parameters be different in each tagging categories: for the GExp+G model, the core Gaussian fraction (f^1), decay constant (κ) and GExp+G scale factor (s^1); for the G+G+G model, the core Gaussian fraction (f^1), bias (b^1) and width scale factor (s^1). To avoid making models that are too complicated, at most two of the three parameters listed above will be split according to tagging categories.

All parameters, unless specified otherwise, are floating in the fits shown below.

Table 12 shows the split parameters of three fits with GExp+G model whose (f^1, κ), (f^1, s^1), or (κ, s^1) are split by tagging category. The Δm_d and τ_{B^0} from these three fits are shown in Table 13. The results of Δm_d and τ_{B^0} are all consistent with the fit to true Δt. We find that the scale factors s^1 are consistent across the tagging categories. We also find that when we split (f^1, κ), the κ are consistent across tagging categories while the f^1 are quite different for different tagging categories. We therefore decide that the default resolution model for signal is the GExp+G model with the core Gaussian fraction f^1 split by tagging category. This resolution model has 10 parameters. All $13 + 10$ parameter of the default fit are shown in Table 14.

We test the default model by fitting to signal events in the generic Monte Carlo sample. The number of signal events is 17288, which is close to the total signal yield in data. We also fit to five independent signal samples of 17288 events each from signal Monte Carlo sample. Out of these six fits, three fail to converge. We notice that the correlations among the three parameters describing the outlier Gaussian are large and that the outlier parameters are also highly correlated with some other resolution parameters. We therefore

GExp+G Model					
split	$f^1_{\rm LT}$	$f^1_{\rm KT}$	$f^1_{\rm N1}$	$f^1_{\rm N2}$	$f^1_{\rm N3}$
(f^1, κ)	0.791 ± 0.062	0.625 ± 0.045	0.852 ± 0.054	0.739 ± 0.055	0.708 ± 0.034
(f^1, s^1)	0.824 ± 0.030	0.647 ± 0.034	0.845 ± 0.036	0.709 ± 0.034	0.689 ± 0.030
split	$s^1_{\rm LT}$	$s^1_{\rm KT}$	$s^1_{\rm N1}$	$s^1_{\rm N2}$	$s^1_{\rm N3}$
(κ, s^1)	1.041 ± 0.050	0.982 ± 0.034	1.071 ± 0.063	0.975 ± 0.050	1.009 ± 0.035
(f^1, s^1)	1.003 ± 0.047	0.988 ± 0.033	1.033 ± 0.062	0.970 ± 0.048	1.010 ± 0.033
split	$\kappa_{\rm LT}$	$\kappa_{\rm KT}$	$\kappa_{\rm N1}$	$\kappa_{\rm N2}$	$\kappa_{\rm N3}$
(κ, s^1)	0.50 ± 0.12	1.02 ± 0.13	0.41 ± 0.14	0.88 ± 0.16	0.94 ± 0.14
(f^1, κ)	0.87 ± 0.24	0.96 ± 0.12	1.12 ± 0.36	1.13 ± 0.24	1.11 ± 0.13

Table 12: The values of the fit parameters for signal MC candidates that are split among tagging categories for fits to the GExp+G model with three different parameter splittings.

	Δm_d (ps^{-1})	$\tau_{\rm B^0}$ (ps)
(f^1, κ)	0.4716 ± 0.0044	1.5525 ± 0.0088
(f^1, s^1)	0.4725 ± 0.0044	1.5528 ± 0.0086
(κ, s^1)	0.4723 ± 0.0045	1.5534 ± 0.0088

Table 13: The Δm_d and $\tau_{\rm B^0}$ from three fits to signal MC with different parameter splittings for the GExp+G model.

decide to fix the outlier scale factor s^{out} to a constant value obtained from large signal Monte Carlo sample fit (Table 14). We will vary the outlier width to determine the systematic uncertainty due to fixing it. Later, we will also decide to fix the outlier bias, $b^{\rm out}$, because we cannot reliably extract this value from Data from a single fit. We warn the reader to read carefully the caption of each table to determine if outlier parameters have been fixed.

The Δt distributions for signal events in the entire signal Monte Carlo are shown in Fig. 15- 16 for mixed and unmixed events in each of the five tagging categories, along with the projected fit with the fixed outlier scale factor. The mixing asymmetry as a function of Δt is also shown in this figure.

The fits with a G+G+G resolution model are quite unstable if all parameters float. Therefore, we fix the outlier scale factor $s^{\rm out}$ to 8. We split one the following sets of parameters among tagging categories in each fit: b^1, f^1, (f^1, s^1), (f^1, b^1), or (b^1, s^1). In these five fits, only fits with parameters b^1, or (f^1, s^1) split converge within the imposed parameter limits. The values of the split parameters for these fits are shown in Table 15, and the values of Δm_d and $\tau_{\rm B^0}$ are shown in Table 16. Although the fit with f^1 split among tagging categories fails to converge, we show the values of the parameters in the

GExp+G			
par.	value	par.	value
Δm_d	0.4714 ± 0.0043 ps^{-1}	$f^1_{\rm LT}$	0.841 ± 0.035
$\tau_{\rm B^0}$	1.5523 ± 0.0096 ps	$f^1_{\rm KT}$	0.686 ± 0.024
$\omega_{\rm LT}$	0.0561 ± 0.0035	$f^1_{\rm N1}$	0.853 ± 0.029
$\omega_{\rm N1}$	0.1840 ± 0.0061	$f^1_{\rm N2}$	0.740 ± 0.026
$\omega_{\rm N2}$	0.3392 ± 0.0054	$f^1_{\rm N3}$	0.719 ± 0.022
$\omega_{\rm N3}$	0.4617 ± 0.0040	κ	1.16 ± 0.07
$\omega^{\rm offset}_{\rm KT}$	0.0672 ± 0.0085	s^1	1.001 ± 0.021
$m_{\rm KT}$	0.128 ± 0.012 ps^{-1}	$b^{\rm out}$	-1.680 ± 1.05
$\Delta\omega_{\rm LT}$	0.0044 ± 0.0058	$f^{\rm out}$	0.0044 ± 0.0022
$\Delta\omega_{\rm KT}$	-0.01403 ± 0.0045	$s^{\rm out}$	5.44 ± 0.86
$\Delta\omega_{\rm N1}$	0.0120 ± 0.0098		
$\Delta\omega_{\rm N2}$	-0.04155 ± 0.0083		
$\Delta\omega_{\rm N3}$	-0.07418 ± 0.0061		

Table 14: All 13+10 parameters of default fit to signal event in signal Monta Carlo sample with GExp+G model, in which the core Gaussian fraction is split by tagging category. For this particular set of results, we do not hold either of the outlier shape parameters to be constant, since the full power of truth-matched signal MC statistics is sufficient for direction measurement.

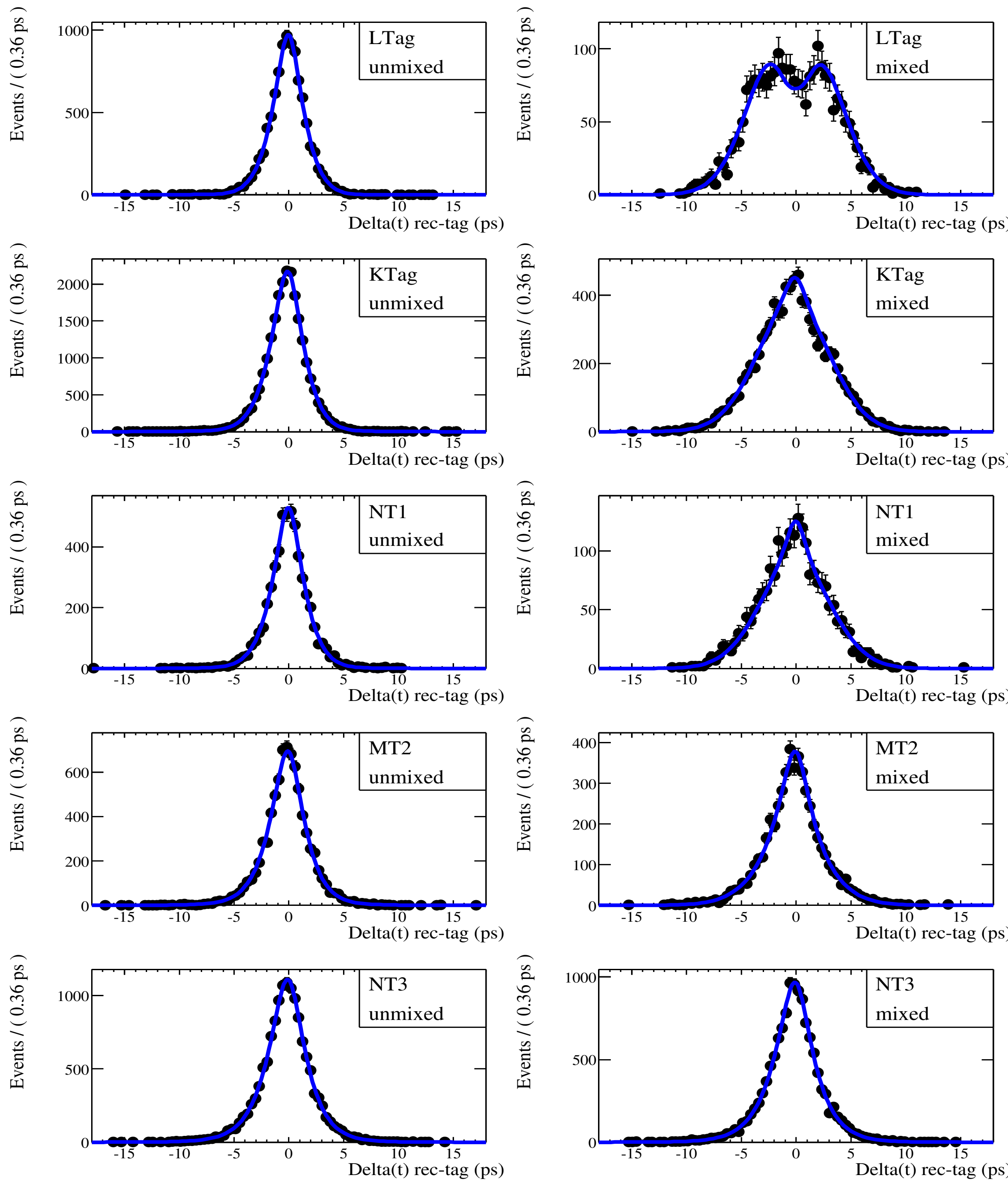

Figure 15: Δt distributions for unmixed and mixed events for the signal Monte Carlo sample for the five tagging categories. The curves are projections of the signal model from the GExp+G fit to signal Monte Carlo.

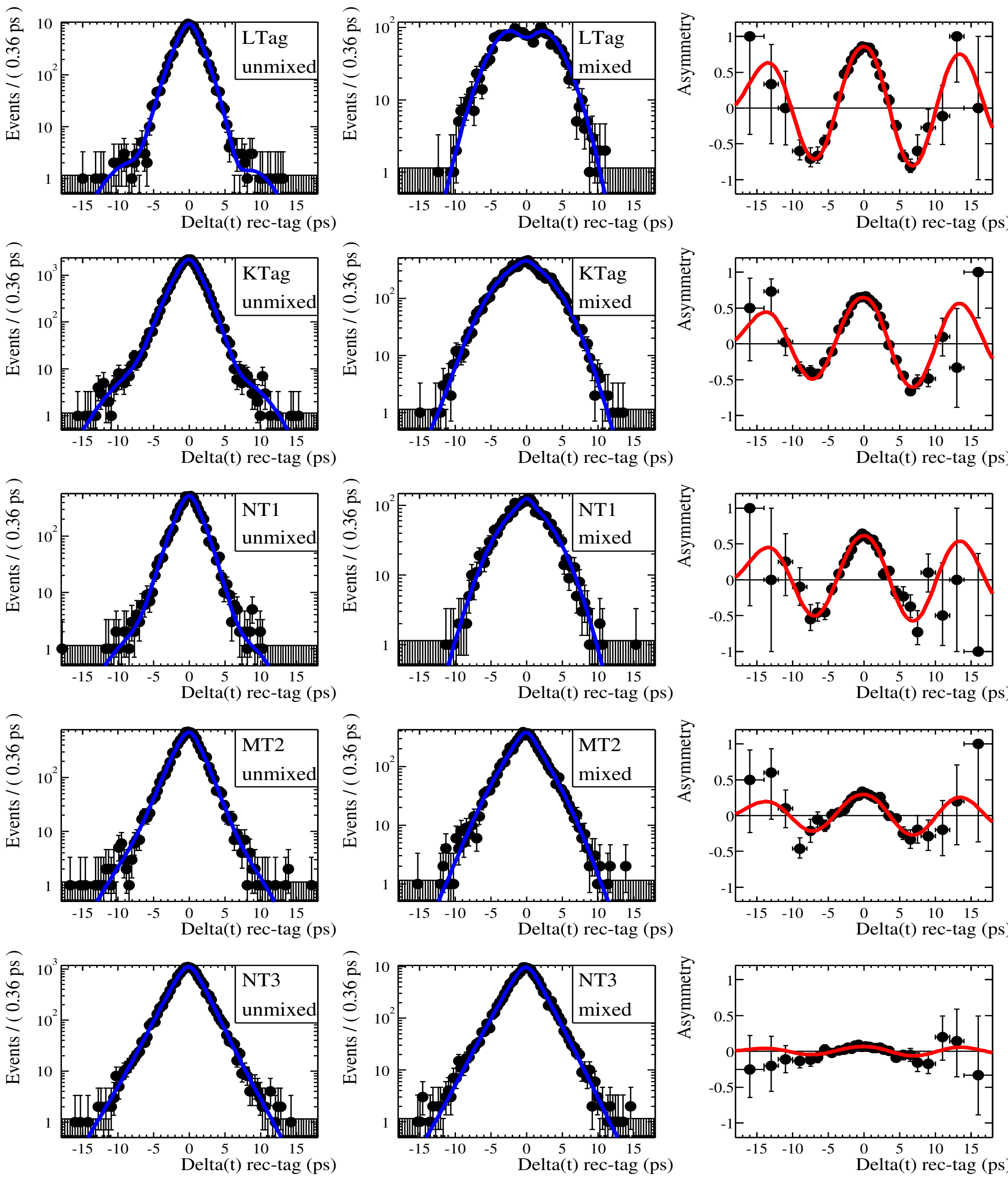

Figure 16: Δt distributions for unmixed and mixed events on a log scale, and the asymmetry plots for the signal Monte Carlo sample for the five tagging categories. The curves are projections of the signal model from the GExp+G fit to signal Monte Carlo.

tables as well.

It appears that there is no need to split s^1, and the fit with b^1 split among tagging categories is more reliable than the fit with f^1 split. Therefore the G+G+G model with core bias (b^1) split by tagging category will be the alternative model to the default GExp+G model when calculating the systematic uncertainty due to choice of resolution model (see Sec. 8.12).

G+G+G Model					
split	f^1_{LT}	f^1_{KT}	f^1_{N1}	f^1_{N2}	f^1_{N3}
f^1	0.952 ± 0.023	0.860 ± 0.025	0.956 ± 0.022	0.893 ± 0.024	0.876 ± 0.023
(f^1, s^1)	0.938 ± 0.033	0.790 ± 0.041	0.947 ± 0.034	0.838 ± 0.039	0.818 ± 0.037
split	b^1_{LT}	b^1_{KT}	b^1_{N1}	b^1_{N2}	b^1_{N3}
b^1	-0.067 ± 0.038	-0.245 ± 0.029	-0.044 ± 0.044	-0.175 ± 0.035	-0.193 ± 0.029
split	s^1_{LT}	s^1_{KT}	s^1_{N1}	s^1_{N2}	s^1_{N3}
(f^1, s^1)	1.061 ± 0.059	0.959 ± 0.047	1.086 ± 0.070	0.965 ± 0.062	0.990 ± 0.046

Table 15: The parameters that are split in fits to signal MC with different parameter splitting with G+G+G model.

G+G+G	Δm_d (ps^{-1})	τ_{B^0} (ps)
f^1	0.4708 ± 0.0043	1.5541 ± 0.0088
b^1	0.4747 ± 0.0044	1.5505 ± 0.0089
(f^1, s^1)	0.4720 ± 0.0044	1.5524 ± 0.0089

Table 16: Δm_d and τ_{B^0} from fits to signal MC with three different parameter splittings among the tagging categories, with the G+G+G model.

5.3 Combinatoric Background Δt Model

The dominant background in the $\mathrm{D}^*\ell$ data sample is combinatoric D^* candidates, as described in BAD 147, and summarized in Sec. 2. Appendix E gives a detailed look at the composition of combinatoric background events. The appendix also builds up a model for describing combinatoric events and explores different ways of breaking down the model parameters by various categories to create the best description of the combinatoric background. The ultimate model given in Appendix E is a prompt plus mixing term convoluted with a double Gaussian resolution function:

$$\begin{aligned}\mathcal{G}^{(\mathrm{comb})} =& [f^{\mathrm{osc}} \cdot \mathcal{G}^{\mathrm{osc}}_{\mathrm{phys}}(\Delta t, g; \omega^{\mathrm{pmt}}) + (1 - f^{\mathrm{osc}}) \cdot \mathcal{G}^{\mathrm{pmt}}_{\mathrm{phys}}(\Delta t, g; \tau^{\mathrm{comb}}, \Delta m^{\mathrm{comb}}, \omega^{\mathrm{osc}})] \otimes \\ & \mathcal{R}_{\mathrm{G+G}}(\delta\Delta t, \sigma_{\Delta t}; b^1, s^1, f^1, b^{\mathrm{out}}, s^{\mathrm{out}})\end{aligned} \tag{13}$$

G+G+G Model			
par.	value	par.	value
Δm_d	0.4747 ± 0.0044	b^1_{LT}	-0.067 ± 0.038
τ_{B^0}	1.5505 ± 0.0089	b^1_{KT}	-0.245 ± 0.029
ω_{LT}	0.0548 ± 0.0035	b^1_{N1}	-0.043 ± 0.044
ω_{N1}	0.1830 ± 0.0061	b^1_{N2}	-0.175 ± 0.035
ω_{N2}	0.3387 ± 0.0054	b^1_{N3}	-0.193 ± 0.029
ω_{N3}	0.4616 ± 0.0040	s^1	1.051 ± 0.032
$\omega^{\mathrm{offset}}_{\mathrm{KT}}$	0.0680 ± 0.0085	f^1	0.906 ± 0.026
m_{KT}	0.126 ± 0.012	b^2	-1.477 ± 0.32
$\Delta\omega_{\mathrm{LT}}$	0.0041 ± 0.0058	s^2	2.40 ± 0.26
$\Delta\omega_{\mathrm{KT}}$	-0.0140 ± 0.0045	b^{out}	-2.72 ± 1.6
$\Delta\omega_{\mathrm{N1}}$	0.0121 ± 0.0098	f^{out}	0.0038 ± 0.0015
$\Delta\omega_{\mathrm{N2}}$	-0.0415 ± 0.0083	s^{out}	8
$\Delta\omega_{\mathrm{N3}}$	-0.0742 ± 0.0061		

Table 17: All 13+11 parameters from a fit to signal MC events with the G+G+G model, in which the core Gaussian bias is split by tagging category.

The total number of parameters is ten. The final division of parameters among categories is shown in Table 18. The terse and slightly obscure notation used in the table is explained in the caption.

A very important question is whether combinatoric events in the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ sideband region have the same Δt distribution as combinatoric events in the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ peak region, Figure 69 in Appendix E shows that the fraction of $\mathrm{B}^0\overline{\mathrm{B}^0}$ declines very slowly with increasing values of $m(\mathrm{D}^*) - m(\mathrm{D}^0)$. One might expect that the effective oscillation frequecy could change as a function of $m(\mathrm{D}^*) - m(\mathrm{D}^0)$.

In Fig. 17, we show the fitted value of Δm^{comb} in slices of $m(\mathrm{D}^*) - m(\mathrm{D}^0)$. The value of Δm^{comb} may have a small dependence on $m(\mathrm{D}^*) - m(\mathrm{D}^0)$, but the data points are also consistent with no slope. The final decision is not to fit for any dependence of Δm^{comb} on $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ in the combinatoric background sample since the effect is small and hard to measure and we are already bordering on overparameterizing the background PDF as it is.

Table 19 shows the measured values of all parameters in the combinatoric background PDF for data and for MC. The combinatoric events in MC are identified by using the Monte Carlo truth associator to pick only misreconstructed D^* events. For MC, the parameters are shown separately for all combinatoric background events and for δm sideband events. The δm sideband is defined as $m(\mathrm{D}^*) - m(\mathrm{D}^0) > 150\,\mathrm{MeV}$ for MC and data.

Figures 18 and 19 show the Δt distribution and the projected PDF for several different categories for the truth-associated combinatoric events in generic MC.

Parameter	Group	# pars
b^1	(LTag, NT1, NT2) $\oplus$ (KTag, NT3)	2
s^1	($\text{OffRes}_{\text{or}}$SS) $\oplus$ (OnRes and OS)	2
ω^{osc}	tagCatGrp$\otimes$(D$^*\ell$ $\oplus$ D*-fake)	10
f^{osc}	[OffRes] $\oplus$ [SS] $\oplus$ TwoLept $\oplus$ OneLept $\oplus$ NoLept	5
Δm^{comb}	shared by all categories	1
τ^{comb}		1
ω^{pmt}		1
f^1		1
b		1
s^{out}		1

Table 18: The final parameter breakdown by categories. Most breakdown categories have been previously defined. The TwoLept category is for events with a lepton tag and a D$^*\ell$ candidate (*i.e.,* not a D*-fake candidate. OneLept is an event with either a D$^*\ell$ candidate and no lepton tag, or a D*-fake candidate and a lepton tag NoLept is an event with a D*-fake candidate and no lepton tag. For the parameter shown at the beginning of each row, the entire group of categories to the left of the $\oplus$ sign share a common parameter and the entire group of categories to the right of the $\oplus$ sign share another common parameter. A *otimes* sign indicates that there is a different parameter for each subcategory of events defined by the categories to the left and the right of the $\otimes$ sign. For example, tagCatGrp$\otimes$(D$^*\ell$ $\oplus$ D*-fake) in the third row means that the ω_{osc} parameter is allowed to be different for the 10 subcategories of events defined by the five tagging categories and the two classes D$^*\ell$ or D*-fake. Finally, a set of square brackets around an attribute indicates that all events with that attribute share a parameter, with attributes to the left taking precedence. For example, the notation in the fourth line indicates that f_{osc} has one common value for all off-resonance samples, one common value for all on-resonance same-side samples, and one common value for each of the three on-resonance opposite-side samples with two, one or zero identified leptons. Got that?

In this model, the bias and width of the resolution are scaled by $\sigma_{\Delta t}$. We justify this model by fitting to the sideband data in slices of $\sigma_{\Delta t}$ with combinatoric Δt model. We still maintain the scaling in the model. If the bias and width of the resolution do *not* scale by $\sigma_{\Delta t}$, we should see the scale factors decrease as $\sigma_{\Delta t}$ increases. If the model is adequate, the scale factors should be consistent to each other for different $\sigma_{\Delta t}$ slices. The result is shown in Table. 20

The scale factors are consistent with each other. The scale factor for width changes by only less than 20%, although the mean of $\sigma_{\Delta t}$ changes by about factor of three. This study shows that scaling the bias and with is a adequate approach for combinatoric model.

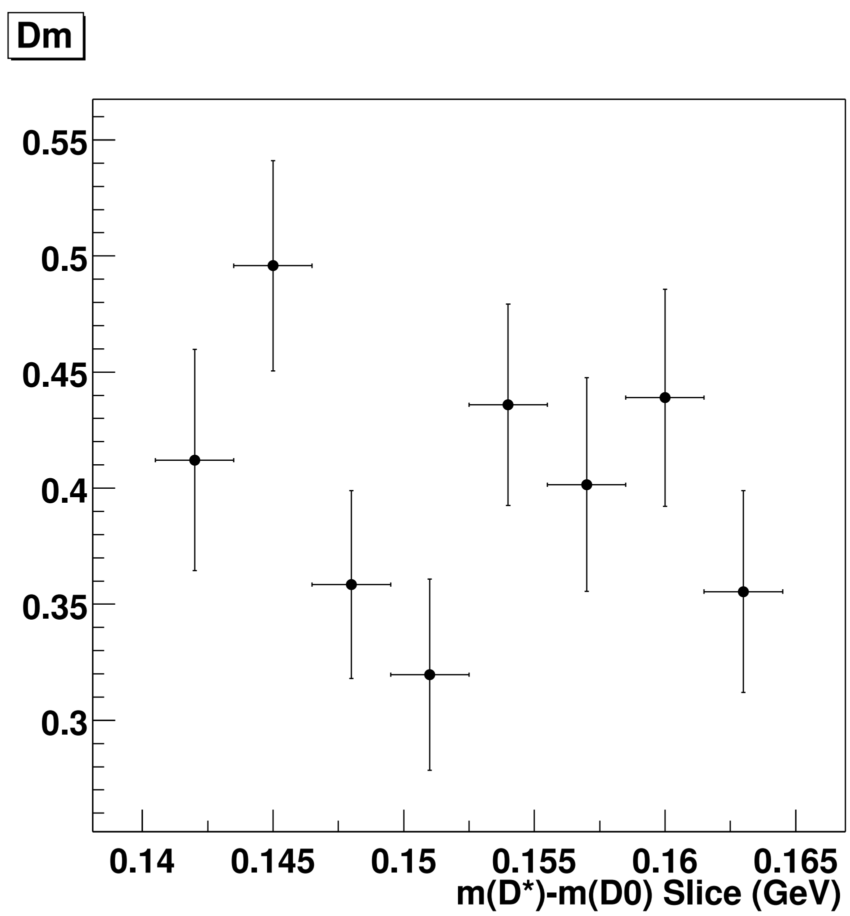

Figure 17: The measured value of Δm^{comb} in slices of $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ from fits to generic MC using the model described in the text.

par.	generic comb.	generic sideband	data sideband
$f^{\mathrm{osc}}_{\mathrm{TwoLept}}$	1.000	1.000	1.000
$f^{\mathrm{osc}}_{\mathrm{OneLept}}$	0.887 ± 0.019	0.864 ± 0.023	0.880 ± 0.023
$f^{\mathrm{osc}}_{\mathrm{NoLept}}$	0.670 ± 0.018	0.645 ± 0.022	0.506 ± 0.020
$f^{\mathrm{osc}}_{\mathrm{OffRes}}$	0.159 ± 0.056	0.129 ± 0.058	0.105 ± 0.036
$f^{\mathrm{osc}}_{\mathrm{SS}}$	0.485 ± 0.026	0.519 ± 0.031	0.409 ± 0.026
b^{1}_{LNT1NT2}	-0.064 ± 0.021	-0.061 ± 0.025	-0.044 ± 0.023
$b^{1}_{\mathrm{K-NT3}}$	-0.154 ± 0.012	-0.150 ± 0.015	-0.114 ± 0.014
s^{1}_{OnOS}	1.146 ± 0.024	1.153 ± 0.027	1.262 ± 0.023
s^{1}_{OffSS}	1.247 ± 0.030	1.204 ± 0.035	1.334 ± 0.028
$\omega^{\mathrm{osc}}_{\mathrm{LT;Fake}}$	0.406 ± 0.023	0.392 ± 0.028	0.460 ± 0.030
$\omega^{\mathrm{osc}}_{\mathrm{KT;Fake}}$	0.192 ± 0.014	0.205 ± 0.017	0.147 ± 0.020
$\omega^{\mathrm{osc}}_{\mathrm{N1;Fake}}$	0.386 ± 0.027	0.379 ± 0.032	0.451 ± 0.036
$\omega^{\mathrm{osc}}_{\mathrm{N2;Fake}}$	0.455 ± 0.018	0.471 ± 0.022	0.512 ± 0.027
$\omega^{\mathrm{osc}}_{\mathrm{N3;Fake}}$	0.519 ± 0.014	0.548 ± 0.018	0.605 ± 0.022
$\omega^{\mathrm{osc}}_{\mathrm{LT;Lept}}$	0.243 ± 0.024	0.236 ± 0.029	0.152 ± 0.028
$\omega^{\mathrm{osc}}_{\mathrm{KT;Lept}}$	0.268 ± 0.010	0.271 ± 0.013	0.310 ± 0.012
$\omega^{\mathrm{osc}}_{\mathrm{N1;Lept}}$	0.276 ± 0.024	0.304 ± 0.030	0.382 ± 0.030
$\omega^{\mathrm{osc}}_{\mathrm{N2;Lept}}$	0.426 ± 0.016	0.446 ± 0.019	0.411 ± 0.019
$\omega^{\mathrm{osc}}_{\mathrm{N3;Lept}}$	0.488 ± 0.011	0.502 ± 0.013	0.489 ± 0.013
Δm^{comb} (ps^{-1})	0.403 ± 0.021	0.385 ± 0.025	0.390 ± 0.024
τ^{comb} (ps)	1.257 ± 0.021	1.279 ± 0.026	1.263 ± 0.027
ω^{pmt}	0.432 ± 0.012	0.414 ± 0.015	0.391 ± 0.010
f^{1}	0.9809 ± 0.0052	0.9790 ± 0.0068	0.9855 ± 0.0056
b^{out}	-1.673 ± 0.508	-1.72 ± 0.56	-1.56 ± 0.71
s^{out}	6.583 ± 0.693	6.12 ± 0.87	6.9 ± 1.1

Table 19: Parameters from a fit using $G^{(comb)}$ to each of combinatoric events in generic MC, $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ sideband events in generic MC, and to $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ sideband events in Data. The outlier scale and bias were fixed to globally determined quantities. The term $f^{\mathrm{osc}}_{\mathrm{TwoLept}}$ is extremely close to 1 in all fits and was fixed to 1 to avoid problems with the fitted value being equal to the parameter limit.

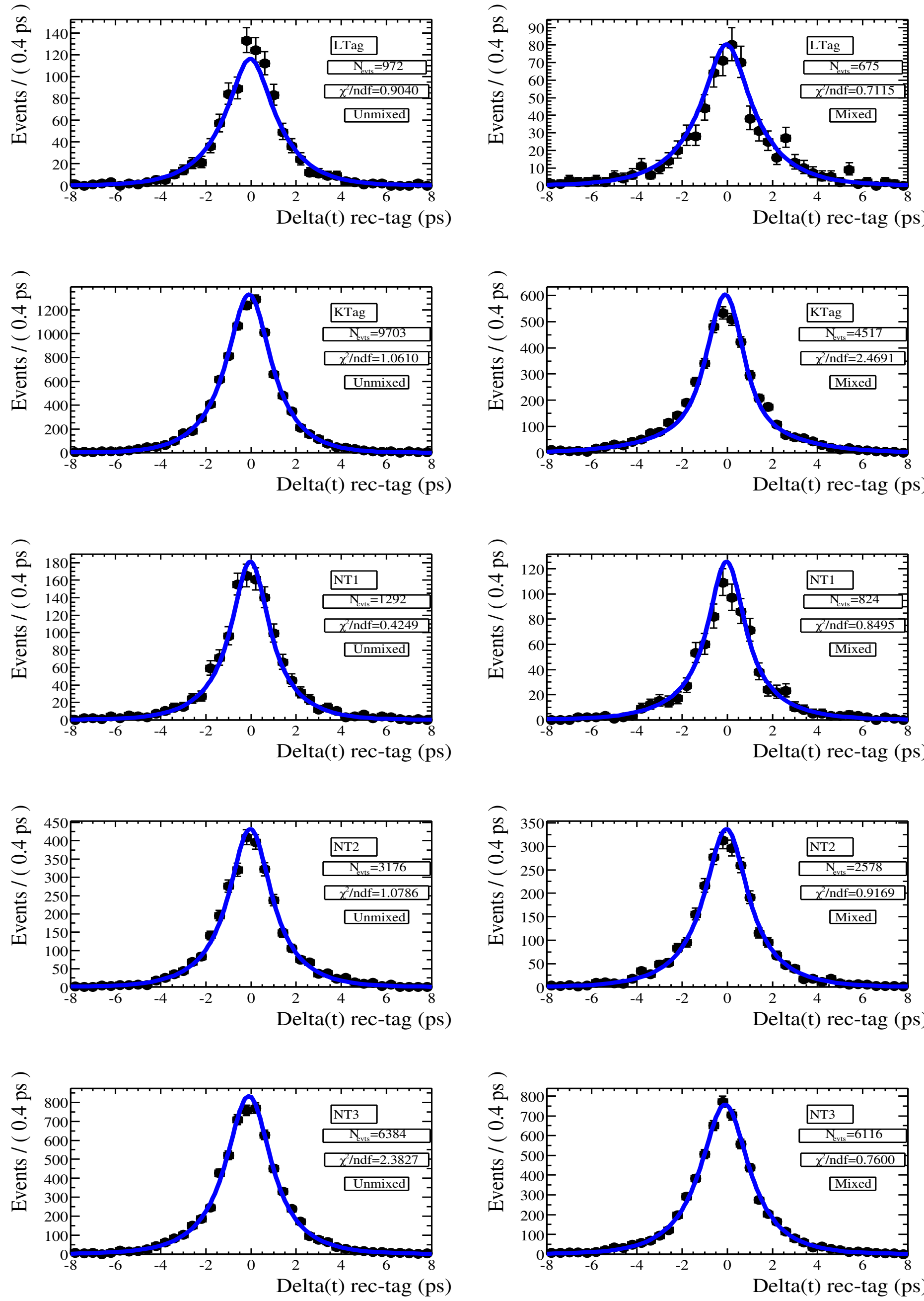

Figure 18: The Δt distribution and the projected PDF for MC combinatoric background events in the five different tagging categories, for unmixed (left column) and mixed (right column) samples. The events are identified in the generic $B^0\overline{B^0}$ and B^+B^- Monte Carlo with the truth associator.

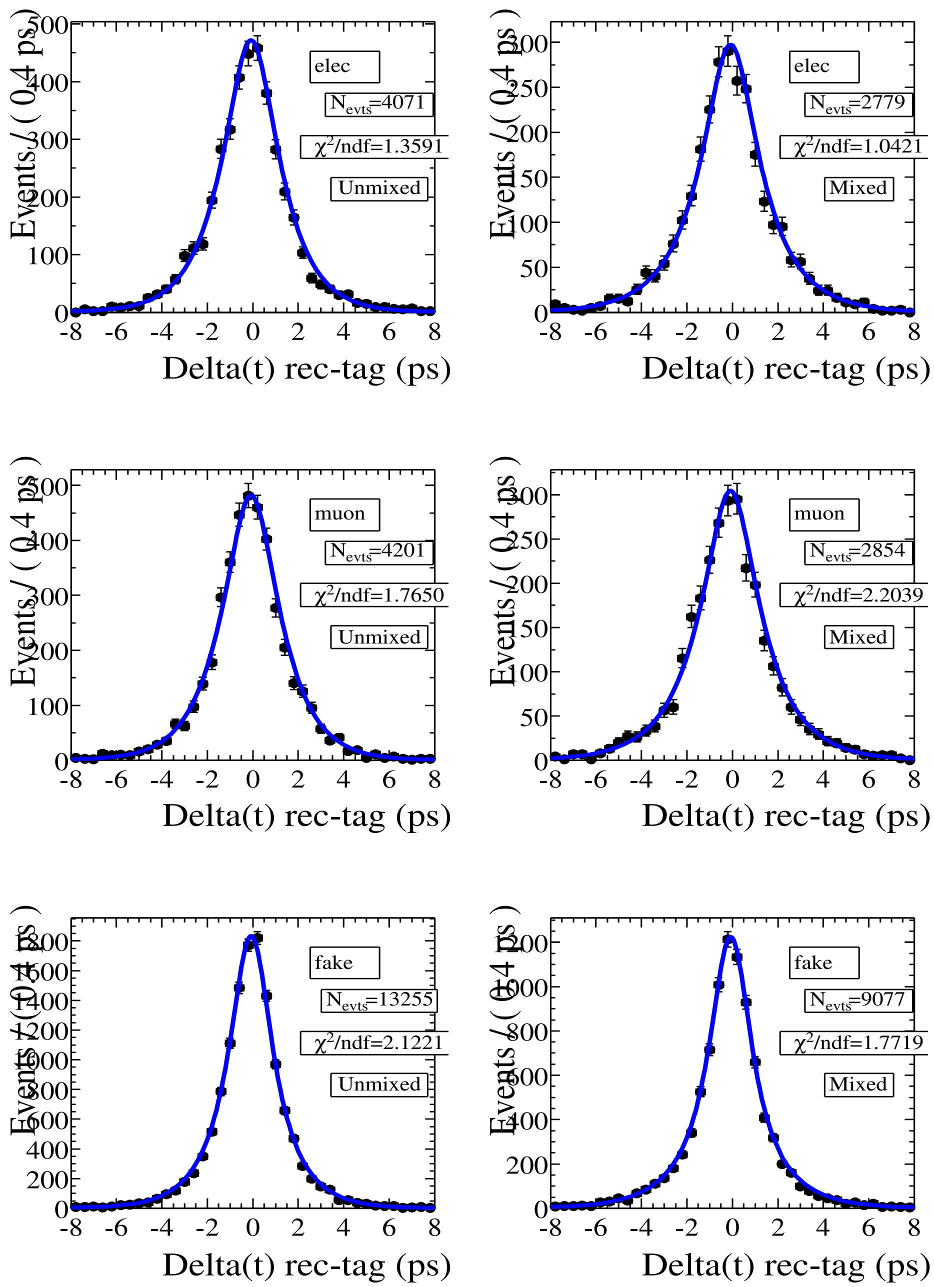

Figure 19: The Δt distribution and the projected PDF for MC combinatoric background events for $\mathrm{D}^* - e$, $\mathrm{D}^* - \mu$ and $\mathrm{D}^* - fake$ candidates, for unmixed (left column) and mixed (right column) samples.

$\sigma_{\Delta t}$ slice	s^1_{OnOS}	s^1_{OffSS}	b^1_{LNT1NT2}	$b^1_{\mathrm{K-NT3}}$
0.0 – 0.5 ps	1.393 ± 0.065	1.439 ± 0.075	-0.065 ± 0.058	-0.088 ± 0.042
0.5 – 0.7 ps	1.25 ± 0.038	1.369 ± 0.048	-0.053 ± 0.039	-0.103 ± 0.023
0.7 – 1.0 ps	1.298 ± 0.046	1.235 ± 0.047	-0.029 ± 0.042	-0.131 ± 0.023
1.0 – 1.8 ps	1.171 ± 0.056	1.267 ± 0.067	-0.013 ± 0.056	-0.115 ± 0.031

Table 20: Fit results for scaled bias and width for the combinatoric Δt model fit to Data in slices of $\sigma_{\Delta t}$. The values are statistically consistent, indicating the scaling the bias and width per $\sigma_{\Delta t}$ is supported.

5.4 Peaking Background Δt Models

There are three types of backgrounds that peak in the D*-D mass difference distribution: a true D* with either a fake lepton, an "uncorrelated" lepton from the other B in the event, or a lepton from continuum events. For each type of peaking background, we determine the Δt model (including appropriate sharing of parameters) with pure Monte Carlo samples. We then determine the starting values for the $\vec{q}^{(pkg)}_{s,c,p}$ parameters with control samples in data enhanced in each type of peaking background. In the approximation that the subsamples are mostly pure, we can fit directly for the starting values of the $\vec{q}^{(pkg)}_{s,c,p}$ parameters, assuming only contributions from the peaking mode and combinatoric background. See Table 1 in Sec. 2.4 of this document or Figs. 20 and 21 on pp. 45-46 in BAD 147, version 6, for the detailed composition of each type of background control sample.

5.4.1 Continuum Background

All $c\overline{c}$ events that have a correctly reconstructed D* are defined as continuum peaking background, independent of whether the associated lepton candidate is a real lepton or a fake. The generic $c\overline{c}$ Monte Carlo is used to identify the appropriate Δt model and sharing of parameters. The off-resonance control samples in data are used to determine starting values for the final combined fit.

Events with a real D* from continuum $c\overline{c}$ production should have vanishing Δt in the case of perfect reconstruction. This suggests the choice of the following model for the Δt distribution of these events:

$$\mathcal{G}^{(\mathrm{cont})} = \mathcal{G}^{\mathrm{pmt}}_{\mathrm{phys}}(\Delta t, g; \omega^{\mathrm{pmt}}) \otimes \mathcal{R}_{\mathrm{G+G}}(\delta\Delta t, \sigma_{\Delta t}; b^1, s^1, f^1, b^{\mathrm{out}}, s^{\mathrm{out}}) \ ,$$

in which we have used a double-Gaussian resolution function. Dependence on the flavor tagging information is included to accomodate any differences in the amount of mixed and unmixed events in the sample.

To study the Δt model for continuum events, we use the $c\overline{c}$ Monte Carlo sample. The combinatoric events are identified by Monte Carlo truth association and removed from this sample. The remaining sample contains 1443 events with $143 < m(\mathrm{D}^*) - m(\mathrm{D}^0) < 148\,\mathrm{MeV}$. By fitting to this sample with different sharing of parameters across categories, we find that the "mistag fraction" for kaon tagged events is significantly different from the mistag fraction for other tagging categories. We also find that the core Gaussian bias for OS and SS events is significantly different. For this model, we obtain the parameter values shown in Table 21 when we fit to the $c\overline{c}$ Monte Carlo sample with a real reconstructed D*. We use these values as starting values for the continuum peaking background in subsequent fits to Monte Carlo samples.

In data, the background control samples for the continuum peaking background are the off-resonance samples shown in the lower half of Table 1 in Sec. 2.4. These control samples contain combinatoric background as well as contiuum peaking background. We use the combinatoric Δt model described in Sec. 5.3, and the parameters determined from a fit to the sideband events ($m(\mathrm{D}^*) - m(\mathrm{D}^0) > 150\,\mathrm{MeV}$, including both on- and off-resonance), shown in Table 19.

We do the exercise of fitting to a combination of combinatoric background and continuum peaking background in both data and Monte Carlo. For the Monte Carlo sample, the $c\overline{c}$ sample is artificially divided into two samples with the proportion of numbers of events being 1:7 (similar to the ratio of off- to on-resonance data). The smaller Monte Carlo sample is artificially marked as off-resonance. The off-resonance sample sizes are 563 and 1228 for the Monte Carlo and data samples, respectively. Both the data and Monte Carlo samples are analysed to extract $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ shapes and background fractions as described in BAD 147. In each case, the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ sideband events are used to obtain the parameters of the Δt model for combinatoric background. We fix all parameters except the continuum peaking background Δt model. The results are shown in Table 21. The parameters from the fit to data (right-hand column) will be used as starting values in the final full fit to data.

Parameter	pure $c\overline{c}$ D^*	"Off-Res" MC	Off-Res data
$\omega^{\mathrm{pmt}}_{\mathrm{LNTx}}$	0.483 ± 0.018	0.491 ± 0.050	0.445 ± 0.055
$\omega^{\mathrm{pmt}}_{\mathrm{KT}}$	0.117 ± 0.013	0.064 ± 0.033	0.094 ± 0.034
b^1_{OS}	0.015 ± 0.039	-0.048 ± 0.15	-0.098 ± 0.15
b^1_{SS}	-0.177 ± 0.072	0.01 ± 0.41	-0.597 ± 0.28
s^1	1.145 ± 0.037	1.07 ± 0.25	1.16 ± 0.16
f^1	0.930 ± 0.023	0.69 ± 0.28	0.83 ± 0.11
s^{out}	3.28 ± 0.40	2.16 ± 0.54	3.01 ± 0.65
b^{out}	-0.11 ± 0.38	0.20 ± 0.48	0.44 ± 0.96

Table 21: $G^{(\mathrm{cont})}$ fit parameters for continuum D^* candidates from $c\overline{c}$ Monte Carlo with a real D^*, and combined fits to combinatoric background and continuum peaking background in "off-resonance" $c\overline{c}$ Monte Carlo (defined in text) and data.

In Figs. 20 and 21, we show the Δt distributions and the projected PDF's for the pure $c\overline{c}$ D^* Monte Carlo sample, for the five tagging categories, and for mixed and unmixed events.

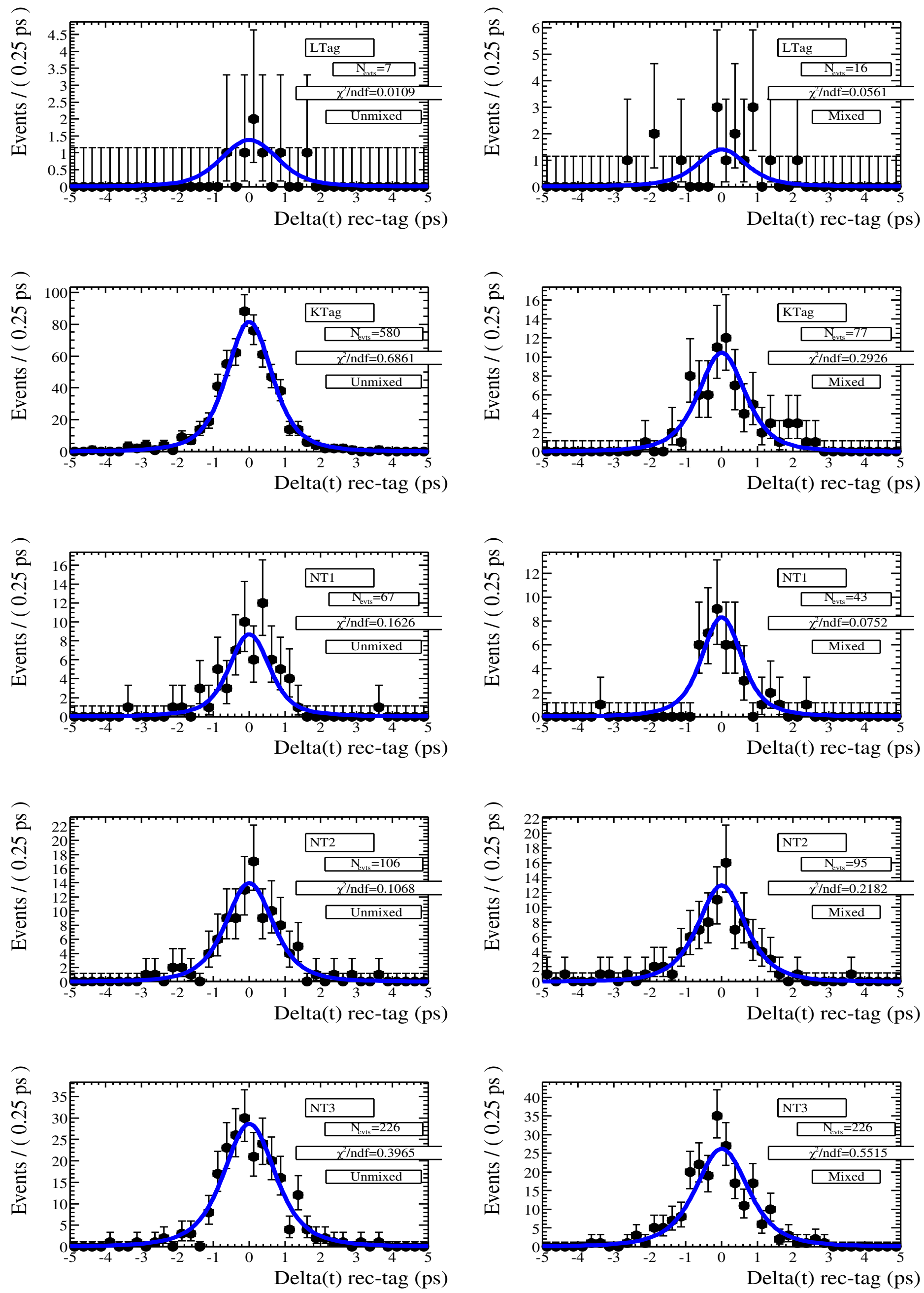

Figure 20: The Δt distributions for pure $c\bar{c}$ D^* Monte Carlo sample and the projected continuum peaking background PDF, for the five tagging categories, and for mixed and unmixed events.

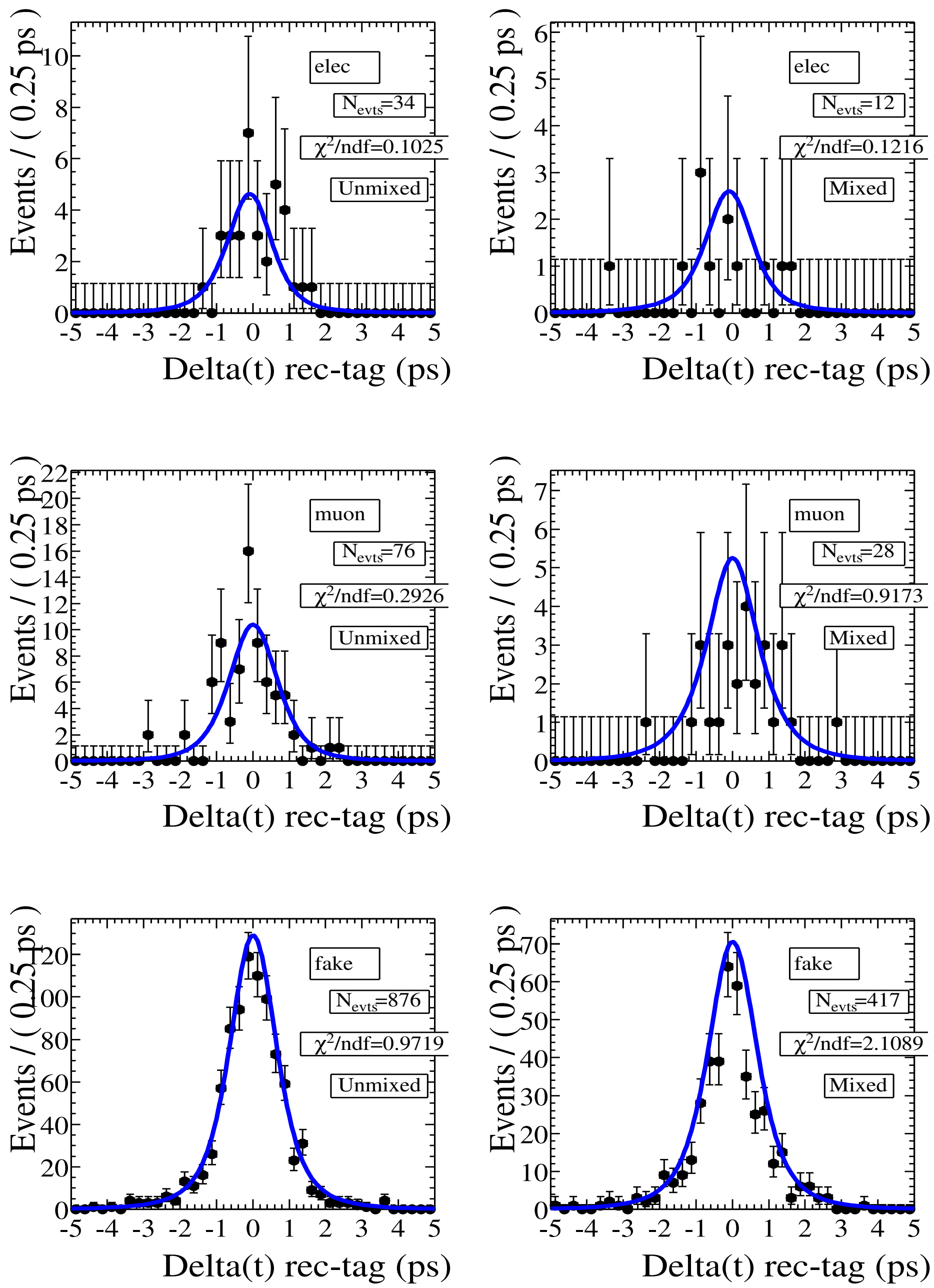

Figure 21: The Δt distributions for pure $c\bar{c}$ D^* Monte Carlo sample and the projected continuum peaking background PDF, for the three lepton-candidate categories, and for mixed and unmixed events.

5.4.2 Fake Lepton Background

We follow a similar procedure to that described in Sec. 5.4.1 to estimate the Δt model parameters, $\vec{q}^{(pkg)}_{s,c,p=2}$, for fake lepton backgrounds.

We first study the Δt structure in the generic $B^0\overline{B^0}$ and B^+B^- Monte Carlo sample by selecting events in which the D^* is correctly reconstructed but the lepton candidate is misidentified. There are 12929 events in this sample.

The PDF used to model the fake lepton backgrounds is the sum of a prompt and mixing term convoluted with a double Gaussian resolution model. This is the same model used for the combinatoric backgrounds, shown in Eq. 13. The parameter breakdown for the fake-lepton peaking background is shown in Table 22. It differs only slightly from the combinatoric background.

Parameter	Group	# pars
b^1	$(\mathrm{L_{or}NT1_{or}NT2Tag}) \oplus (\mathrm{Kaon_{or}NT3})$	2
ω^{pmt}	$(\mathrm{L_{or}NT1_{or}NT2Tag}) \oplus (\mathrm{K_{or}NT3})$	2
ω^{osc}	tagCatGrp	5
Δm^{fake}	common to all	1
τ^{fake}		1
f^{osc}		1
s^1		1
f^1		1
b^{out}		1
s^{out}		1

Table 22: The final parameter breakdown by categories for the fake-lepton peaking background, $G^{(\mathrm{fake})}$. The symbolic notation used in this table is described in the caption of Table 18.

The total number of parameters is 16, six of which are resolution parameters. The result of a fit to the pure fake-lepton sample in generic Monte Carlo is shown in the first column in Table 23 under "MC matched fake".

The goodness of fit can be judged from Figs. 22 and 23. The projected PDF and the Δt distributions are shown for 10 different categories (five tagging categories, mixed or unmixed) in Fig. 22 and six different categories (three lepton types, mixed or unmixed) in Fig. 23. The PDF fits the data well in each category, even though the PDF parameters are not intrinsically broken down into these particular categories.

Again, we perform exercises on generic MC and data control samples to extract fake lepton Δt parameters. The control sample used here is the On-Resonance, opposite-side (OS), fake-lepton control sample. (See Table 1, {On;OS;fake}.) For the data sample, the δm peak region is 72% pure D^*-hadron events, with about 13% contamination from continuum events and 14% from real signal. We fit the sideband events in this control sample to determine the combinatoric Δt model. We then take the continuum model parameters determined in Sec. 5.4.1 to describe the continuum events in this control sample. We ignore the uncorrelated background, which is less than 1% on average. The remaining portion is the signal, including a charged B contribution. We use the signal model described in Sec. 5.4.4 for the neutral and charged B. The parameters are determined from a fit to signal Monte Carlo sample (Sec. 5.2.1). We determine the charged B fraction in this {On;OS;fake} generic Monte Carlo sample using truth association, and fix the charged B fraction at this value (5.9 %) in the fits. All parameters in combinatoric, continuum and signal models are fixed. Since we fix the combinatoric model, we fit to the peak region only (143–148 MeV). The fit results for the generic MC sample and for data are shown in the middle two columns in Table 23. We can see that a couple of parameters reach the limits that we imposed.

In an attempt to avoid having parameters at their limits, we reduce the number of free parameters in the fit. We let the models for fake-lepton peaking background and for combinatoric background share common parameters for the outlier Gaussion: bias ($b^{\rm out}$), scale factor ($s^{\rm out}$) and fraction (f^1). Since this model involves free parameters for combinatoric events, we now fit the entire $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ range. The results for the generic MC sample and for data are shown in the last two columns in Table 23. Now none of the parameters from the fit to data are at their limits. Keep in mind that these parameter values will be used as starting values in the final fit, but will float as many parameters as possible in the final fit.

		Control sample combined $G^{(fake)}$ fits			
				Share outliers with comb. bg.	
Parameter	MC matched fake	Generic MC	Data	Generic MC	Data
b^1_{LNT1NT2}	-0.056 ± 0.036	-0.047 ± 0.050	-0.188 ± 0.074	-0.0467 ± 0.049	-0.178 ± 0.075
$b^1_{\mathrm{K-NT3}}$	-0.219 ± 0.026	-0.252 ± 0.040	-0.260 ± 0.054	-0.2575 ± 0.039	-0.2493 ± 0.055
$\omega^{\mathrm{pmt}}_{\mathrm{LNT1NT2}}$	0.507 ± 0.074	0.49 ± 0.13	0.58 ± 0.20	0.43 ± 0.11	0.57 ± 0.17
$\omega^{\mathrm{pmt}}_{\mathrm{K-NT3}}$	0.256 ± 0.056	0.294 ± 0.098	0.63 ± 0.20	0.268 ± 0.085	0.57 ± 0.16
$\omega^{\mathrm{osc}}_{\mathrm{LT}}$	0.182 ± 0.021	0.107 ± 0.026	0.173 ± 0.040	0.119 ± 0.024	0.165 ± 0.042
$\omega^{\mathrm{osc}}_{\mathrm{KT}}$	0.182 ± 0.015	0.180 ± 0.020	0.125 ± 0.030	0.188 ± 0.019	0.125 ± 0.031
$\omega^{\mathrm{osc}}_{\mathrm{N1}}$	0.293 ± 0.024	0.245 ± 0.030	0.291 ± 0.045	0.256 ± 0.029	0.286 ± 0.046
$\omega^{\mathrm{osc}}_{\mathrm{N2}}$	0.401 ± 0.022	0.379 ± 0.029	0.391 ± 0.040	0.389 ± 0.028	0.389 ± 0.041
$\omega^{\mathrm{osc}}_{\mathrm{N3}}$	0.540 ± 0.018	0.497 ± 0.023	0.484 ± 0.035	0.500 ± 0.023	0.493 ± 0.036
Δm^{fake} (ps^{-1})	0.384 ± 0.019	0.398 ± 0.023	0.424 ± 0.027	0.375 ± 0.019	0.435 ± 0.031
τ^{fake} (ps)	1.387 ± 0.029	1.437 ± 0.041	1.484 ± 0.049	1.495 ± 0.031	1.512 ± 0.043
f^{osc}	0.865 ± 0.021	0.894 ± 0.027	0.891 ± 0.038	0.886 ± 0.026	0.877 ± 0.038
s^1	1.076 ± 0.049	1.109 ± 0.071	1.321 ± 0.096	1.115 ± 0.070	1.256 ± 0.096
f^1	0.9741 ± 0.0073	0.964 ± 0.013	0.9832 ± 0.0058	0.9863 ± 0.0027	0.9539 ± 0.0073
b^{out}	-2.13 ± 0.86	-1.49 ± 0.85	$-4.00000 \pm 0.00061^*$	-2.77 ± 0.85	-0.806 ± 0.28
s^{out}	7.86 ± 0.83	7.26 ± 1.00	$10.0 \pm 1.9^*$	8.58 ± 0.83	4.77 ± 0.39

*at limit

Table 23: Fit parameters for fake-lepton model, $G^{(\mathrm{fake})}$. Left column: a fit to a pure fake-lepton sample from generic $\mathrm{B}^+\mathrm{B}^- + \mathrm{B}^0\overline{\mathrm{B}^0}$ Monte Carlo. Central two columns: fits to the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ peak region of the {On;OS;fake} control samples with a combined model, for generic Monte Carlo and data. Right two columns: fits to the full $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ region for the {On;OS;fake} control samples with a combined model where the outlier Gaussian shares parameters between the fake lepton and combinatoric models.

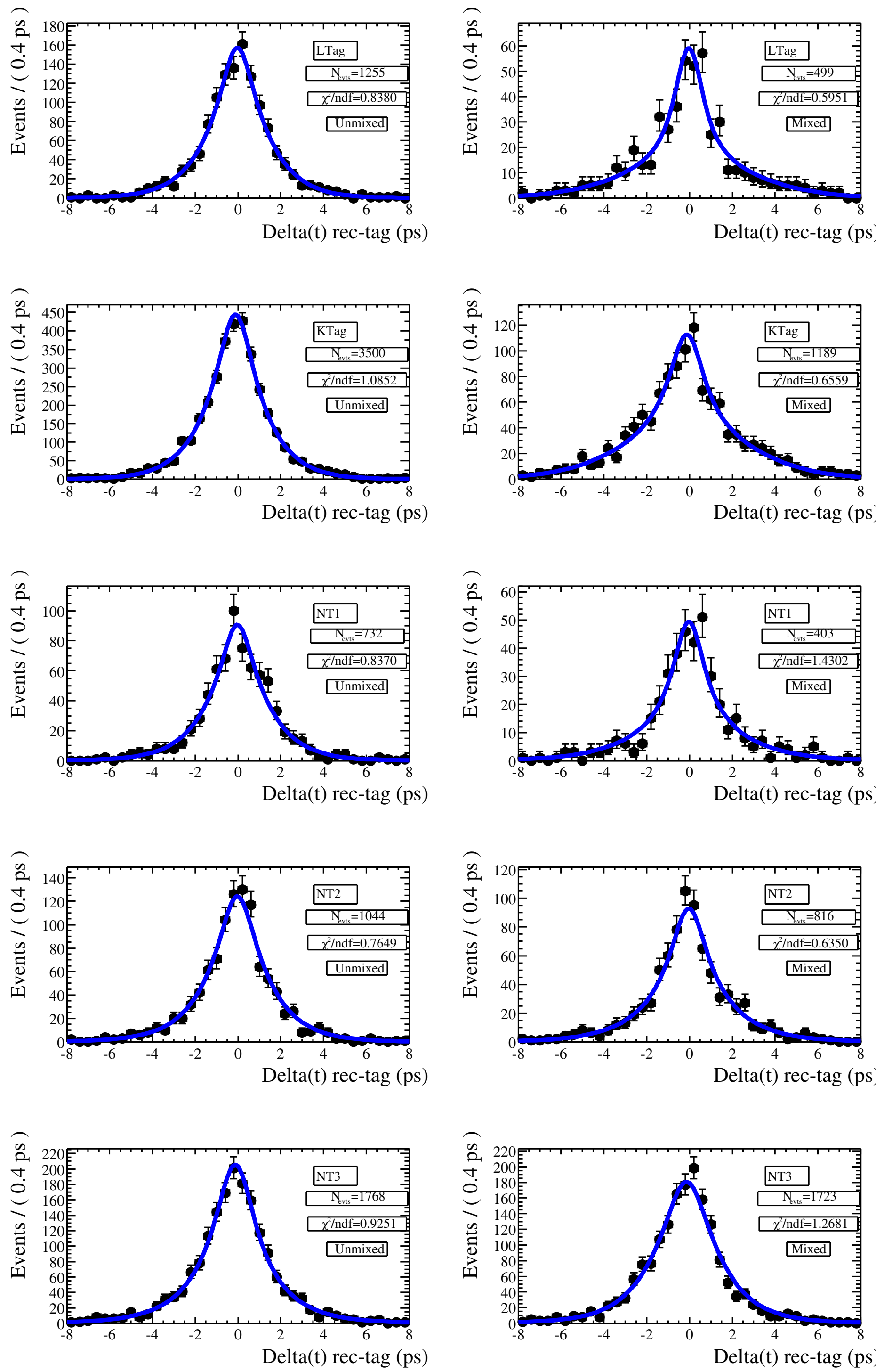

Figure 22: The fake-lepton Δt distributions and $G^{(\mathrm{fake})}$ PDF projections onto Δt for the five tagging categories, for unmixed (left) and mixed (right) events in generic MC.

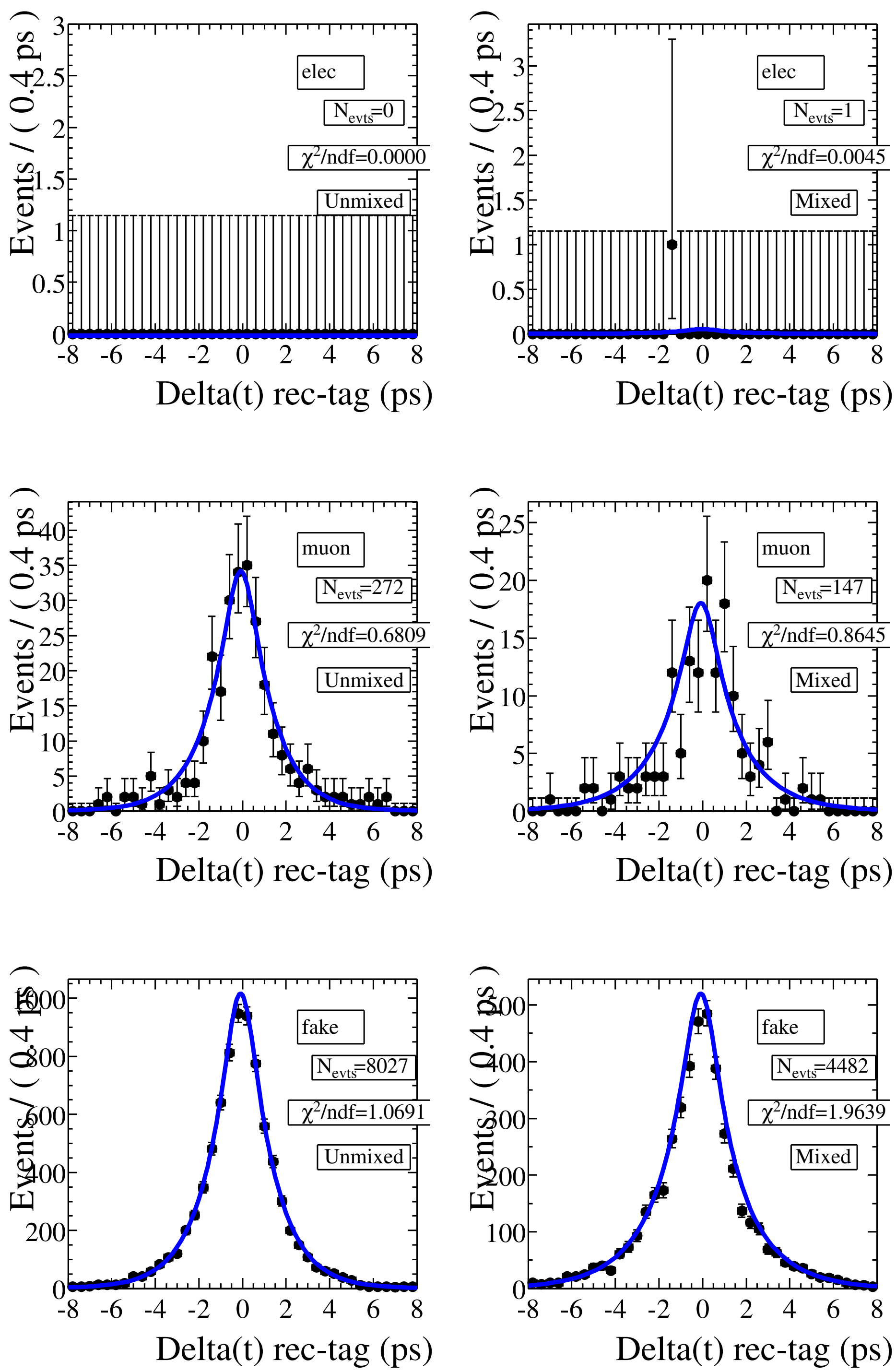

Figure 23: The fake-lepton Δt distributions and projected PDF for the three types of lepton candidates in generic MC, for unmixed (left) and mixed (right) events.

5.4.3 Uncorrelated Lepton Background

The uncorrelated-lepton background is defined as events in which a real D^* is reconstructed in a $\mathrm{B}^0\overline{\mathrm{B}^0}$ or $\mathrm{B}^+\mathrm{B}^-$ event but the lepton candidate is from the other B or from a cascade decay. We use the same-side lepton control samples in on-resonance data to measure the Δt, $\sigma_{\Delta t}$ behavior of uncorrelated lepton backgrounds. See the second group of three rows (On, SS) in Table 1 in Sec. 2.4 for the yields and background fractions in this control sample. This control sample of about 600 events is only 30% pure due to significant contributions from true signal and combinatoric background. Therefore, we first establish the Δt model by studying a pure sample of misreconstructed generic $\overline{\mathrm{B}^0}$ and B^+ Monte Carlo candidates. The sample uses truth association to identify good D^*'s matched with an uncorrelated lepton candidate (*i.e.,* same side cascade lepton or lepton from the other B). This Monte Carlo sample consists of 913 events, 27% of which have a B^+ parent. Table 24 shows the composition of this sample.

Category	Value	Population
`Dmode`	Kpi	328
	K3pi	294
	Kpipi0	291
`tagCat`	LTag	119
	KTag	393
	NT1	64
	NT2	102
	NT3	235
`leptID`	elec	441
	muon	360
	fake	112
`angCut`	OS	483
	SS	430

Table 24: Composition of the pure uncorrelated-lepton sample selected from generic B Monte Carlo.

Physics considerations suggest two features of the uncorrelated sample. First, we expect the reconstructed Δt to systematically underestimate the true Δt value since "swapping" a lepton with the other B will generally dilute the separation between the two B candidate. We might also expect that events with small Δt separation might have a higher probability of being misreconstructed as an uncorrelated lepton candidate because it is "easier" to associate the lepton candidate with either vertex and still pass the D^*-ℓ vertexing requirements.

Secondly, we might expect truly mixed events to be more likely to suffer uncorrelated lepton association during reconstruction due to the like-charge correlation between the ℓ

on the reconstructed side and a ℓ on the tagging side. (That is, in order to successfully match a lepton candidate from the other B with a good D^* on the reconstruction side, the ℓ must have the correct (opposite) charge with respect to the D^* candidate.) After "taking" the ℓ candidate from the other B, the flavor identification algorithm may mistake the flavor of the other B, so we expect the enhancement of mixed candidates to be strongest at the Monte Carlo truth level, with a slight dilution at the reconstructed level. Therefore, we expect the lepton tagging category to exhibit this relative enhancement in mixed events to the greatest extent.

We observe these effects to varying degrees in the data. Figure 24 shows the Δt distribution from Monte Carlo truth for the pure correlated sample for true mixed and unmixed events with the expected true distribution overlaid. From here we see that not only is the number of mixed events significantly enhanced in the sample, but that unmixed events with small true Δt are preferentially selected. *(Not so obvious that this is true, from the figure alone.)*

In Table 25, we show χ_d, the ratio of the number of mixed events to the total number of events, for each of the five tagging categories, based on the Monte Carlo truth information and based on the measured tag. The high values of χ_d shown in this table support the expectation for enrichment of mixed states in the uncorrelated sample. To fully identify the effect, we consider the subset of the pure uncorrelated-lepton sample from $B^0\overline{B^0}$ (candidates from B^+ parents are completely unmixed at the truth level, of course). The effective χ_d from Monte Carlo truth counting is 0.548, significantly higher than 0.174 as computed for the entire $B^0\overline{B^0}$ Monte Carlo sample. Note, however, that after flavor reconstruction, the effective χ_d value increases slightly to 0.586. This is because the uncorrelated lepton candidate is not always the primary lepton, and therefore truly unmixed events have a small probability of being mistakenly reconstructed as mixed.

These observations suggest that while the time information is inconsistent with a good measure of mixing, tagging can still retain substantial information about flavor content. A simplified Δt model allowing for a mixing term fitted to this sample yields an effective mixing frequency consistent with zero. We therefore model the probability density function for uncorrelated-lepton candidates with a prompt term and a lifetime term. The result of a fit to this Monte Carlo sample is shown in Table 26. The Δt distribution for the sample and the projected PDF are shown in Fig 25.

We now use this model to fit the uncorrelated-lepton background control samples in data and Monte Carlo. In this fit, the combinatoric-background parameters are from a fit to sideband events in the entire data set; the continuum parameters are from a fit to the off-resonance control sample; the fake parameters are from a fit to {On;OS;fake} control sample, in which the outlier terms are shared with combinatoric model; the signal model are from a fit to the large pure signal MC sample. The outlier terms in the uncorrelated lepton model are shared with fake and combinatoric Δt models. The charged B fraction in the signal model is fixed at 1.8%, which is obtained from the truth information in the generic Monte Carlo control sample. Only the uncorrelated-lepton parameters and shared

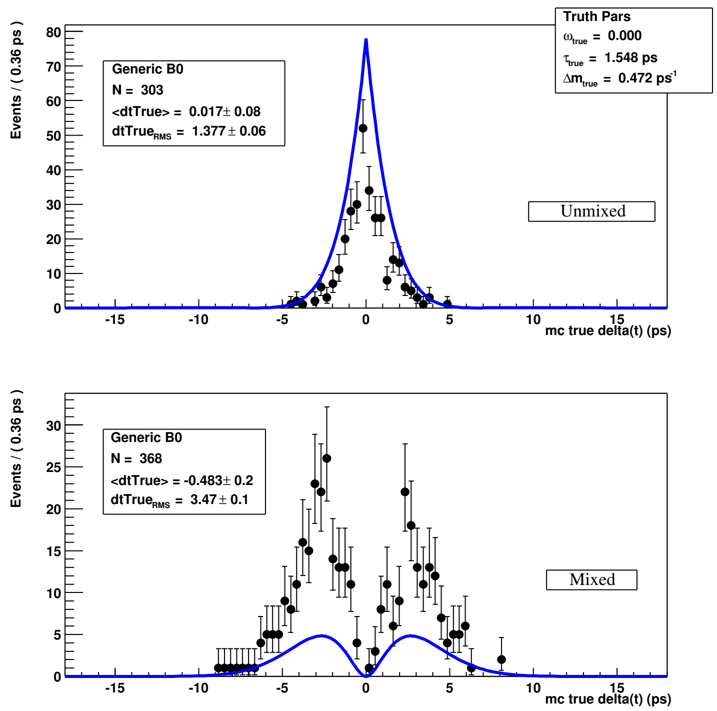

Figure 24: Distribution of true Δt from generic MC for pure uncorelated-lepton candidates in the sample compared to the expected true signal distribution, shown separately for mixed and unmixed events. The relative normalization of the curves reflects the expected number of mixed and unmixed events in an unbiased sample of neutral B pairs. The unmixed plot (upper) shows the preferential selection for "smaller" Δt candidates. The mixed plot (lower) illustrates the inconsistent mixing information in the sample, as evidenced by the consistency of Δt shape between the sample and curve, but inconsistency by counting, shown by the poor relative normalization.

outlier parameters are free in the fit. The fit function is

$$\begin{aligned}\mathcal{G}^{(\mathrm{unco})} =& [f^{\mathrm{life}} \cdot \mathcal{G}^{\mathrm{life}}_{\mathrm{phys}}(\Delta t, g; \omega^{\mathrm{pmt}}) + (1 - f^{\mathrm{life}}) \cdot \mathcal{G}^{\mathrm{pmt}}_{\mathrm{phys}}(\Delta t, g; \tau^{\mathrm{unco}}, \Delta m^{\mathrm{unco}}, \omega^{\mathrm{life}})] \otimes \\ & \mathcal{R}_{\mathrm{G+G}}(\delta\Delta t, \sigma_{\Delta t}; b^1, s^1, f^1, b^{\mathrm{out}}, s^{\mathrm{out}}) \end{aligned} \tag{14}$$

We fit this model to the full $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ range. The results of the fits to generic MC and data are shown in Table 26.

We test the performance of this model by projecting the Δt distribution for different subsamples of the data. Figures 26 and 27 show that the model is generally consistent in subsamples separated by tagging category or type of lepton candidate.

Category	Tagging Category	Effective χ_d
MC truth tag	LTag	0.68
	KTag	0.53
	NT1	0.59
	NT2	0.43
	NT3	0.55
Measured tag	LTag	0.75
	KTag	0.62
	NT1	0.59
	NT2	0.55
	NT3	0.45

Table 25: Ratio of the number of mixed events to the total number of events (χ_d) for the pure uncorrelated lepton sample from generic $B^0\overline{B^0}$ Monte Carlo, before and after reconstruction. Note that the lepton tagging category has the highest fractional population of mixed events, consistent with expectation.

Parameter	MC matched	Generic MC	Data
$\omega^{\rm life}_{\rm LT}$	0.765 ± 0.074	0.98 ± 0.24	1.000 ± 0.020
$\omega^{\rm life}_{\rm KNT}$	0.550 ± 0.045	0.68 ± 0.18	0.68 ± 0.15
$\omega^{\rm pmt}$	0.415 ± 0.083	0.591 ± 0.060	0.59 ± 0.10
$f^{\rm life}$	0.66 ± 0.11	0.24 ± 0.11	0.41 ± 0.25
$\tau^{\rm unco}$ (ps)	1.22 ± 0.14	2.06 ± 0.59	1.54 ± 0.71
b^1	-0.329 ± 0.069	-0.33 ± 0.15	-0.00 ± 0.32
s^1	1.37 ± 0.11	1.26 ± 0.16	1.44 ± 0.30
f^1	0.965 ± 0.030	0.825 ± 0.058	0.56 ± 0.18
$b^{\rm out}$	-1.4 ± 1.6	-0.66 ± 0.31	-0.46 ± 0.18
$s^{\rm out}$	5.7 ± 1.5	2.60 ± 0.40	1.99 ± 0.28

Table 26: Fit parameters for uncorrelated lepton model, $G^{(\rm unco)}$. Left column: a fit to a pure uncorrelated lepton sample in B^+B^- + $B^0\overline{B^0}$ Monte Carlo sample; mid and right columns: fits to the full region of {On;SS;lepton} control sample for generic MC sample and data, respectively, where outlier terms ($b^{\rm out}$,$b^{\rm out}$ and f^1) are shared among combinatoric, fake and uncorrelated background models.

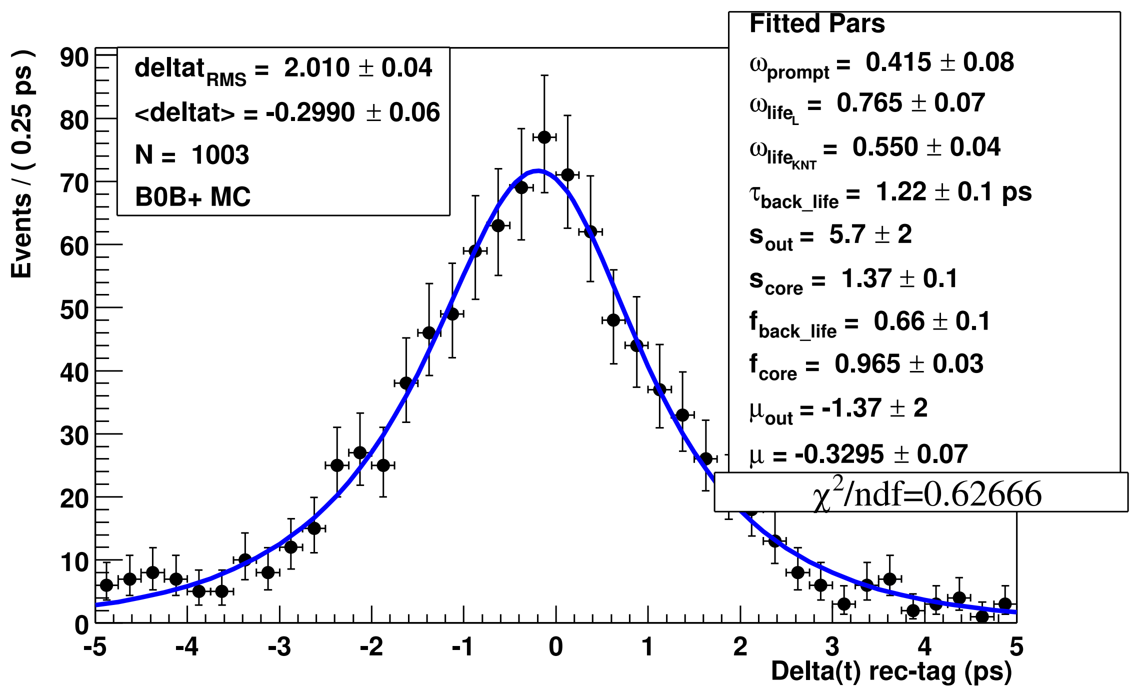

Figure 25: Δt projection of the simultaneous fit to the pure uncorrelated lepton sample from generic $B^0\overline{B^0}$ and B^+B^- Monte Carlo.

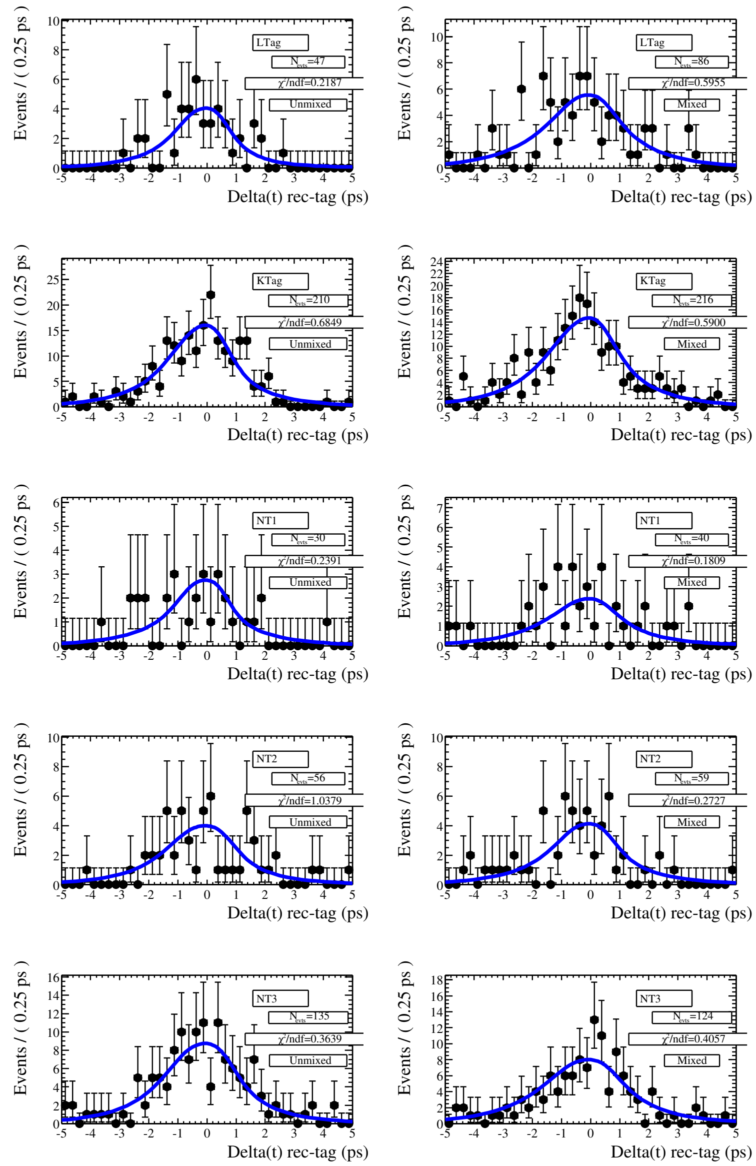

Figure 26: Δt projection of the simultaneous fit to the pure uncorrelated lepton sample from generic $B^0\overline{B^0}$ and B^+B^- Monte Carlo for various subsamples separated by tagging category and mixing status.

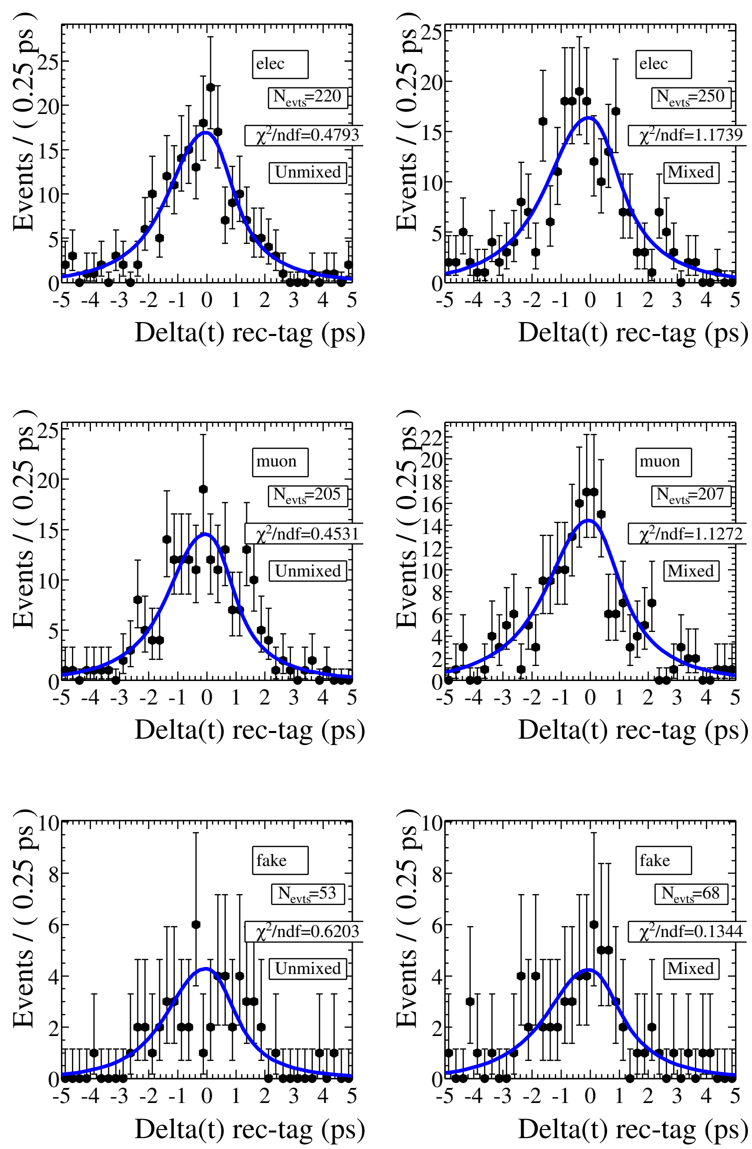

Figure 27: Δt projection of the simultaneous fit to the pure uncorrelated lepton sample in generic $B^0\overline{B^0}$ and B^+B^- Monte Carlo for various subsamples separated by type of lepton candidate and mixing status.

5.4.4 $B^+ \to D^{*-}\ell^+X$

The charged B fraction found for the $B^0 \to D^{*-}\ell^+\nu_\ell$ sample described in the $\sin 2\beta$-mixing Phys. Rev. D journal paper [14] is $(4.5+-0.3+-2.2)\%$. This result was determined from a fit to the $\cos\theta_{B^0,D^*\ell}$ distribution in data, with shapes taken from Monte Carlo. Since charged B's do not exhibit mixing behavior, our strategy is to use the Δt and tagging information to discriminate these charged B peaking background events from neutral B signal events, in the simultaneous fit to τ and Δm_d. We choose to use the same resolution model as for the signal since the decay dynamics are very similar. The signal model, after adding a charged B background term, becomes

$$\begin{aligned}\mathcal{G}_c^{(\text{sig})}(\Delta t, \sigma_{\Delta t}, g\,; \vec{q}_c^{(\text{sig})}) = [&\frac{1}{4\tau_{B^0}} e^{-|\Delta t|/\tau_{B^0}}(1 + g(1-2\omega_{B^0}^c)\cos(\Delta m_d \Delta t)) + \\ &\frac{1}{4\tau_{B^+}} e^{-|\Delta t|/\tau_{B^+}}(1 + g(1-2\omega_{B^+}^c)\} \otimes \mathcal{R}(\delta\Delta t, \sigma_{\Delta t}; \vec{q_r})\,,\end{aligned}$$

where $\omega_{B^0}^c$ ($\omega_{B^+}^c$) is the mistag fraction for netural (charged) B for a given tagging category c.

Given that the ratio of the charged B to neutral B lifetime is close to 1 and only 5% of the events are charged B, we will not have sufficient sensitivity to distinguish the lifetimes in the fit. Additionally, since we are aiming to measure the B^0 lifetime, we parameterize the B^+ time-dependent term in terms of the ratio of τ_{B^+}/τ_{B^0}, and fix this ratio to the PDG2002 world average. Later, we vary the ratio by the PDG errors to estimate a systematic uncertainty.

We also need to fix either ω_{B^+} or the ratio $\omega_{B^+}/\omega_{B^0}$ because there are only two independent parameters among ω_{B^+}, ω_{B^0} and the charged B fraction (f_{B^+}) [16]. The reason is that, given the fact that the two lifetimes are so close, the time evolution of mixed and unmixed events can be expressed approximately as

$$\begin{aligned} U &\propto (1 + f_{B^\pm}D_{B^\pm})/2 + ((1-f_{B^\pm})D_{B^0}\cos\Delta m_d\Delta t)/2 \\ M &\propto (1 - f_{B^\pm}D_{B^\pm})/2 - ((1-f_{B^\pm})D_{B^0}\cos\Delta m_d\Delta t)/2\,,\end{aligned}$$

where D is the dilution $1-2\omega$. Only two quantities, $f_{B^\pm}D_B^\pm$ and $(1-f_{B^\pm})D_B^0$ are measurable. Therefore, in our fit, we fix the mistag ratio as well as the lifetime ratio.

We use this model to fit a sample of neutral and charged B signal candidates in the generic Monte Carlo, in which 17290 events are real signal and 952 are $B^+ \to D^{*-}\ell^+X$ events, and there are no other backgrounds. The charged B fraction in this mixture is 5.2%. The lifetime ratio τ_{B^+}/τ_{B^+} and mistag ratio $\omega_{B^+}/\omega_{B^0}$ are set to be 1.069 and 0.97, respectively. The lifetime ratio is based on the PDG2000 value, which is the value used in the event generator of the Monte Carlo samples. The mistag ratio is calculated from the overall mistag fractions for $B^0\overline{B^0}$ and B^+B^- samples by event counting based on the truth. The fit results for Δm_d, τ_{B^0} and the charged B fraction are shown in Table 27, along with the results of the fit to B^0 signal only, and the fit with charged B fraction fixed at the true value. In all fits, the resolution function and mistag fractions are free parameters.

Signal Δt fits to (truth-matched) Signal & Generic Monte Carlo				
parameter	Sig MC(*)	Gen $B^0\overline{B^0}$	Gen $B^0\overline{B^0}$, B^+B^-	Gen $B^0\overline{B^0}$,B^+B^-
Δm_d (ps^{-1})	0.4715 ± 0.00427	0.4640 ± 0.0101	0.4630 ± 0.0106	0.4617 ± 0.0132
τ_{B^0} (ps)	1.5523 ± 0.00915	1.5749 ± 0.0227	1.5666 ± 0.0221	1.5656 ± 0.0227
f^{out}	0.0044 ± 0.0015	-0.0012 ± 0.0027	-0.0005 ± 0.0029	-0.00024 ± 0.0034
f_{B^+}	0	0	0.0521	0.047 ± 0.030

Table 27: $G^{(sig)}$ fits to MC truth-matched signal $B^0\overline{B^0}$ and B^+ from Generic MC (we include the result for the fit to signal MC for reference). We show the fitted results (with full statistical errors) for Δm_d, τ_{B^0}, f_{B^+}, and the signal outlier fraction f^{out}. The third column of values represents a fit to truth-matched Generic MC events where we fix f_{B^+} to its true value and hold it constant in the fit; in the fourth column we float this parameter and extract it from the fit. In all the fits, resolution function and mistag fractions are floating, except for the two signal outlier model shape parameters. (*) Note that these signal MC fit results differ slightly from ones quoted elsewhere in this document; these results are the ones that are most relevant to this table since they incorporate the final fixed values for both of the outlier shape parameters, etc.

The fitted charged B fraction is consistent with the "true" value. The generic MC sample is significantly smaller than the signal MC sample, but the fitted lifetime and mixing values appear consistent with the generator values at the 1σ level. We note however, that in this truth-matched sample of Generic MC events, we are unable to measure a non-zero outlier fraction. The correlation between f^{out} and τ_{B^0} is what drives the measured lifetime higher in the Generic MC fits. This "systematic" failure is reflected in the "statistical" error which incorporates these correlations. Upon closer inspection, we see the error on Δm_d inflates when we float f_{B^+}, as expected, and because of the positive correlation, an underestimate of f_{B^+} also causes an underestimate of Δm_d, as seen by comparing columns 2,3 and 4 of the table. We conclude that while we are unlucky with our Generic MC sample in that we cannot reliably measure the signal outlier fraction, the fit results are still correctly described by their statistical errors.

6 Fit Results

6.1 Signal Outliers in Data

The outlier portion of the resolution function is intended to describe the Δt distribution for candidates whose residual is not consistent with the calculated error. These outliers are typically due to gross errors in tagside vertex reconstruction, as discussed in Appendix B. Since there is little reason to expect the "mean" and "width" of the signal outlier Gaussian to scale with the per-event error, the outlier resolution function is parameterized as a Gaussian in $\delta(\Delta t)$.

We found that the fit to data is not very robust when all three parameters for the outlier Gaussian of the resolution function are allowed to float in the fit. In particular, the best fit value for the bias of the outlier function is always at the lower end of the allowed range, even when the lower limit is set at -18 ps! This is likely due to the negative tail in the Δt distribution for events in the signal sample; see Fig. 31. In addition, when all outlier parameters are free, many fits fail to converge or have an error matrix that is not positive definite. Therefore, we repeated the full fit with 36 different *fixed* values of the outlier bias and outlier width, covering a wide range for each parameter (-1 to -10 ps in bias, 4 to 11 ps in width).

Figure 28 shows the negative-ln-likelihood surface as a function of outlier bias and width. The negative-ln-likelihood has been offset so that the minimum value over the plot is zero. For fixed width, the minimum in negative-ln-likelihood always occurs at the most negative bias. For fixed bias, the likelihood curve as a function of resolution becomes broader as the bias becomes more negative. Note that there is a local minimum at small bias. Also note that the change in negative-ln-likelihood is less than 0.5 for most of the values of bias and width scanned in this study.

We also studied the use of GExp model for the outlier Gaussian, and several "extremely wide" Gaussian models (widths of order 15 ps) to explore flat/uniform distributions of outliers in residual space, but we observed no significant changes in the fit results for the two physics parameters.

A scatterplot of the resulting values of Δm_d and τ_{B^0} is shown in Fig. 29. Note that the spread in values of τ_{B^0} and Δm_d is small compared to the statistical uncertainty on each quantity. This is consistent with the fact that the change in ln-likelihood is small as the outlier bias and width are varied (see Fig. 28).

In order to choose an outlier bias and width for the default fit to data, we select a point near the middle of the cluster of points in the τ_{B^0} versus Δm_d scatterplot (Fig. 29) and use the corresponding bias and width. We end up using a bias of -5 ps and a width of 6 ps.

We assign half the full spread of lifetime and mixing results as the systematic uncertainty due to the outlier parameters, as described in Sec. 9.2. The resulting systematic uncertainties are 0.005 ps for τ_{B^0} and 0.001 ps^{-1} for Δm_d.

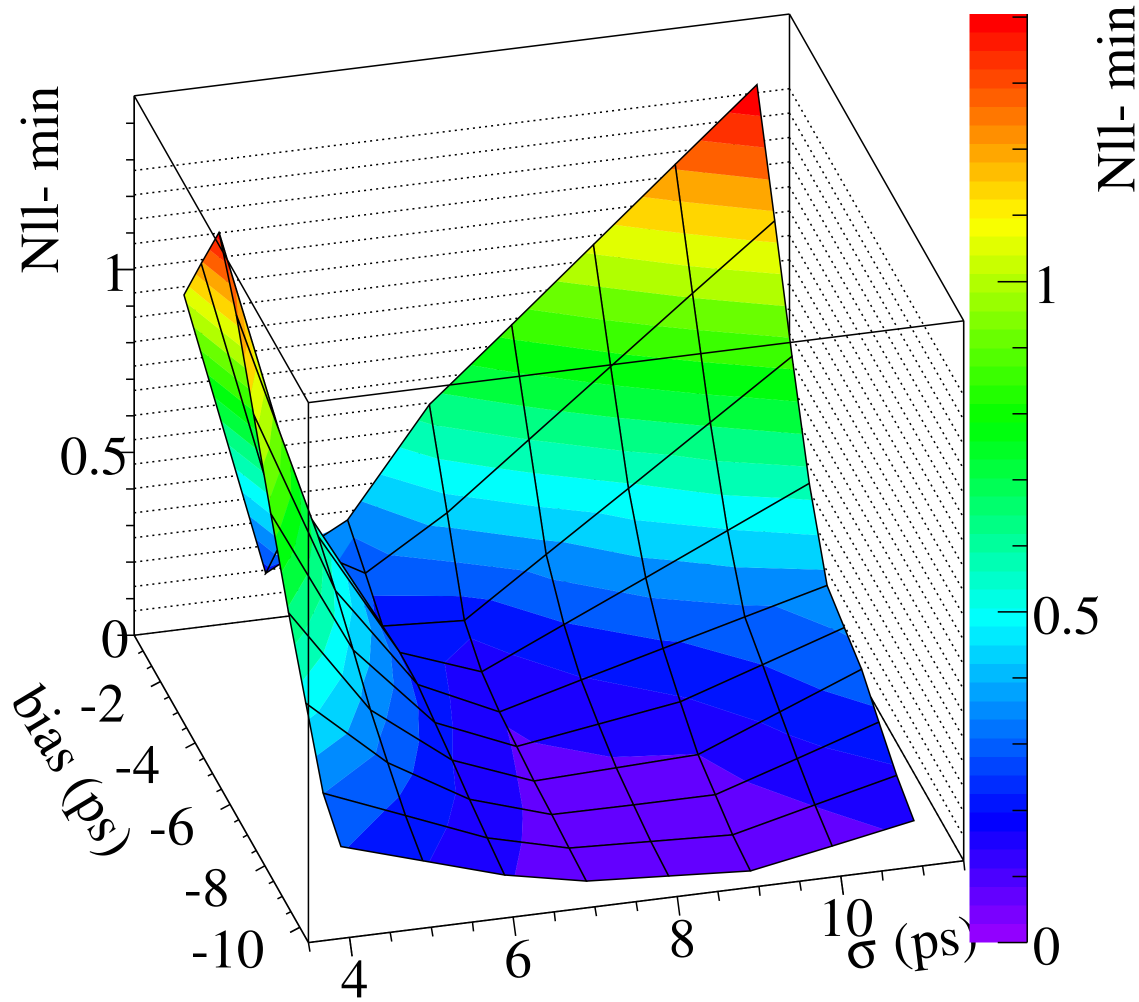

Figure 28: The negative-ln-likelihood surface in the space of outlier bias and outlier width for the outlier Gaussian of the signal resolution model in Data. The vertical scale has been offset so that minimum is at 0. The surface is constructed from 36 fits to the Data with different fixed values of the signal outlier parameters.

To summarize, we fix the bias and width in the outlier Gaussian of the resolution function to -5 ps and 6 ps, respectively. We note that it is not unusual to fix outlier parameters in lifetime and mixing fits in BABAR– no other lifetime and mixing analysis in BABAR has allowed the outlier parameters to be free. (In fact, this detailed studied of the likelihood surface in the outlier shape parameter space is most likely a first.)

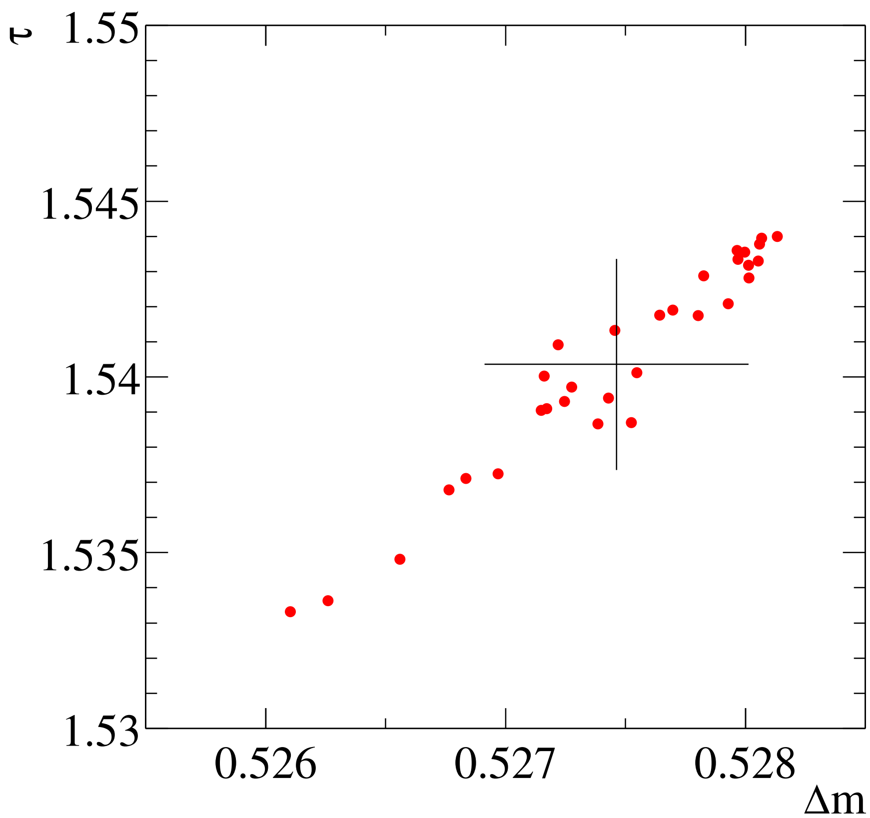

Figure 29: Scatter plot of the set of (blinded) Δm_d, τ_{B^0} values obtained from data for 36 fits with different fixed values of the outlier bias and width in the range -1 to -10 ps for bias and 4 to 11 ps for width. The cross indicates the mean and rms of the Δm_d, τ_{B^0} distribution. Note that the range of the τ_{B^0} (y) axis is slightly less than one statistical error bar, and that the range of the Δm_d (x) axis is less than 1/5 of a statistical error bar.

6.2 Results of Master Model Fit to Data

In this section, the baseline result for the Master Model fit to Data is shown. The total number of floating parameters is 72: 22 in signal model, 24 in combinatoric model and 26 in peaking background models. The ratio of B^+ to B^0 lifetimes is fixed at the central value from PDG2002 [17], 1.083 ± 0.017. The ratios of B^+ to B^0 mistag fractions are fixed at the central values extracted from BAD119 [6]. They are summarized in Table 28.

The fitted signal Δt model parameters are shown in Table 29, the combinatroic Δt model parameters in Table 30, and the three peaking background Δt model parameters

Category	ratio
LTag	0.54 ± 0.10
KTag	0.68 ± 0.05
NT1	0.99 ± 0.12
NT2	1.05 ± 0.07
NT3	$1.12 \pm 0.12^{*}$

Table 28: B^+ to B^0 mistag fraction ratios extracted from BAD119 [6], except the NT3 category, which is not available. We calculate the ratio for NT3 from generic Monte Carlo sample.

in Table 31. The one-sigma contour for τ_{B^0} *vs.* Δm_d is shown in Fig. 30. The correlation coefficient is $\rho(\Delta m_d, \tau_{B^0}) = -0.22$. This number along with other large correlations between τ_{B^0} or Δm_d and a few other parameters are shown in Table 32. A complete list of correlation coefficients between all parameters and (τ_{B^0}, Δm_d) can be found in Appendix H. Additional Δt-projection plots for various subsamples of the Data can be found in Appendix I.

Figure 31 and 32 show the Δt and the asymmetry distributions of unmixed and mixed events in signal sample {On;OS;e or μ }, and the projections of the model.

Final Fit to Data, Signal Model Parameters					
parameter	value	parameter	value	parameter	value
Δm_d (ps^{-1})	$0.xxx \pm 0.018$	$f_{B^\pm}$	0.082 ± 0.029	s^1	1.201 ± 0.063
τ (ps)	$1.yyy \pm 0.023$	ω_{LT}	0.071 ± 0.015	κ	0.86 ± 0.17
-	-	ω_{KT}^{offset}	0.002 ± 0.024	${f^1}_{LT}$	0.72 ± 0.10
-	-	m_{KT}	0.229 ± 0.036	f^1_{KT}	0.609 ± 0.088
-	-	ω_{N1}	0.212 ± 0.020	${f^1}_{N1}$	0.69 ± 0.13
-	-	ω_{N2}	0.384 ± 0.018	${f^1}_{N2}$	0.70 ± 0.10
-	-	ω_{N3}	0.456 ± 0.012	${f^1}_{N3}$	0.723 ± 0.078
-	-	$\Delta\omega_{LT}$	-0.001 ± 0.022	f^{out}	0.0027 ± 0.0017
-	-	$\Delta\omega_{KT}$	-0.024 ± 0.015	b^{out} (ps)	-5.000
-	-	$\Delta\omega_{N1}$	-0.098 ± 0.032	s^{out} (ps)	6.000
-	-	$\Delta\omega_{N2}$	-0.112 ± 0.028	-	-
-	-	$\Delta\omega_{N3}$	-0.023 ± 0.019	-	-

Table 29: Default fit result for Data — signal model parameters.

The charged B fraction in data from this final fit is $(8.2 \pm 2.9)\%$. Although the selection criteria are not exactly the same, it is interesting to compare this result with the

Final Fit to Data, Combinatoric Δt Model Parameters			
parameter	value	parameter	value
Δm^{comb}	0.422 ± 0.020	$b^1{}_{\mathrm{LNT12}}$	-0.056 ± 0.021
τ^{comb}	1.234 ± 0.024	$b^1{}_{\mathrm{KNT3}}$	-0.104 ± 0.012
$\omega^{\mathrm{osc}}_{\mathrm{Fake;LTag}}$	0.434 ± 0.026	$f^{\mathrm{osc}}_{\mathrm{Off}}$	0.099 ± 0.035
$\omega^{\mathrm{osc}}_{\mathrm{Fake;KTag}}$	0.128 ± 0.017	$f^{\mathrm{osc}}_{\mathrm{SS}}$	0.385 ± 0.023
$\omega^{\mathrm{osc}}_{\mathrm{Fake;NT1}}$	0.443 ± 0.032	$f^{\mathrm{osc}}_{\mathrm{NoLept}}$	0.526 ± 0.019
$\omega^{\mathrm{osc}}_{\mathrm{Fake;NT2}}$	0.507 ± 0.024	$f^{\mathrm{osc}}_{\mathrm{OneLept}}$	0.894 ± 0.022
$\omega^{\mathrm{osc}}_{\mathrm{Fake;NT3}}$	0.589 ± 0.020	$f^{\mathrm{osc}}_{\mathrm{TwoLept}}$	1.000
$\omega^{\mathrm{osc}}_{\mathrm{Lept;LTag}}$	0.156 ± 0.023	$s^1{}_{\mathrm{OffSS}}$	1.337 ± 0.024
$\omega^{\mathrm{osc}}_{\mathrm{Lept;KTag}}$	0.297 ± 0.010	$s^1{}_{\mathrm{On;OS}}$	1.259 ± 0.021
$\omega^{\mathrm{osc}}_{\mathrm{Lept;NT1}}$	0.377 ± 0.025	f^1	0.9666 ± 0.0071
$\omega^{\mathrm{osc}}_{\mathrm{Lept;NT2}}$	0.401 ± 0.016	b^{out}	-0.98 ± 0.24
$\omega^{\mathrm{osc}}_{\mathrm{Lept;NT3}}$	0.491 ± 0.011	s^{out}	4.60 ± 0.42
ω^{pmt}	0.3960 ± 0.0093	-	-

Table 30: Default fit result for Data — Δt model parameters for combanitoric background, $G^{(\mathrm{comb})}$.

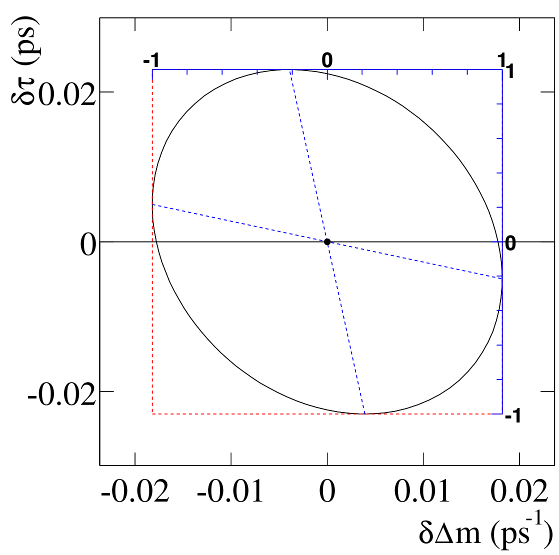

Figure 30: The one-sigma contour for τ_{B^0} *vs.* Δm_d from the full fit to Data. The correlation coefficient is -0.22.

charged B fraction found for the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ sample described in the $\sin 2\beta$-mixing Phys. Rev. D journal paper: $(4.5 \pm 0.3 \pm 2.2)\%$. This result was determined from a fit to

Final Fit to Data, Peaking BG Model Parameters					
cont. par.	value	fake par.	value	unco. par.	value
$\omega_{\rm KT}$	0.083 ± 0.028	$\Delta m^{\rm fake}$	0.444 ± 0.031	$\tau^{\rm unco}$	1.07 ± 0.17
$\omega_{\rm LNT}$	0.457 ± 0.040	$\tau^{\rm fake}$	1.416 ± 0.060	$\omega^{\rm life}_{\rm KNT}$	0.71 ± 0.13
$b^1{}_{\rm OS}$	0.04 ± 0.11	$\omega^{\rm osc}_{\rm LT}$	0.248 ± 0.038	$\omega^{\rm life}_{\rm LT}$	0.9998
$b^1{}_{\rm SS}$	-0.08 ± 0.13	$\omega^{\rm osc}_{\rm KT}$	0.142 ± 0.029	$\omega^{\rm pmt}$	0.00000 ± 0.00064
s^1	1.300 ± 0.090	$\omega^{\rm osc}_{\rm NT1}$	0.342 ± 0.044	$f^{\rm life}$	0.877 ± 0.085
f^1	0.909 ± 0.040	$\omega^{\rm osc}_{\rm NT2}$	0.377 ± 0.039	b^1	-0.30 ± 0.18
-	-	$\omega^{\rm osc}_{\rm NT3}$	0.512 ± 0.042	f^1	1.000
-	-	$\omega^{\rm pmt}_{\rm KNT3}$	0.46 ± 0.14	s^1	1.34 ± 0.24
-	-	$\omega^{\rm pmt}_{\rm LNT12}$	0.64 ± 0.15	-	-
-	-	$f^{\rm osc}$	0.852 ± 0.046	-	-
-	-	$b^1{}_{\rm KNT3}$	-0.208 ± 0.062	-	-
-	-	$b^1{}_{\rm LNT12}$	-0.078 ± 0.065	-	-
-	-	f^1	0.944 ± 0.023	-	-
-	-	s^1	1.159 ± 0.091	-	-

Table 31: Default fit result for Data — Δt model parameters for peaking backgrounds. The outlier bias ($\mu_{\rm out}$) and scale factor ($\sigma_{\rm out}$) for these peaking background are shared with combinatoric background parameters.

Δm_d global correlation	0.74
$\tau_{\rm B^0}$ global correlation	0.69
$\rho(\Delta m_d, \tau_{\rm B^0})$	-0.22
$\rho(\Delta m_d, f_{\rm B^\pm})$	0.58
$\rho(\tau_{\rm B^0}, \sigma^1_{\rm sig})$	-0.49
$\rho(\tau_{\rm B^0}, f^{\rm out}_{\rm sig})$	-0.26

Table 32: Global correlation coefficients for Δm_d and $\tau_{\rm B^0}$ from the full fit to Data. Correlation coefficients for pairs of key parameters.

the $\cos\theta_{\rm B^0,D^*\ell}$ distribution in data, with shapes taken from Monte Carlo. Given that the errors on the two results are not correlated, the difference is $(3.7 \pm 3.7)\%$. Therefore, the two results are consistent.

Since we float many parameters in the model, it is interesting to see how the errors on $\tau_{\rm B^0}$ and Δm_d, and their correlation change when different parameters are free in the fit, or fixed to their best value. We perform a series of fits, fixing all parameters at the values obtained from the default fit, except (a) Δm_d and $\tau_{\rm B^0}$, (b) Δm_d, $\tau_{\rm B^0}$ and all mistag fractions in signal model, (c) Δm_d, $\tau_{\rm B^0}$ and $f_{\rm B^\pm}$, (d) Δm_d, $\tau_{\rm B^0}$, $f_{\rm B^\pm}$ and all mistag fractions in signal model, (e) all parameters in signal Δt model. The resulting errors on Δm_d, τ_{B^0} and their correlations are shown in Table 33. The one-sigma contours for these fits and the default fit are shown in Fig. 33.

We can see that the error on τ_{B^0} changes very little until we float the signal resolution function. Floating the background parameters adds a very small contribution to the error. The contribution from the charged B fraction and mistag fractions to τ_{B^0} error is negligible. On the other hand, the charged B fraction changes the error on Δm_d the most. The contributions from floating the mistag fractions, resolution functions and background models are relatively small.

floating pars	$\sigma(\Delta m_d)$ (ps^{-1})	$\sigma(\tau_{\mathrm{B}^0})$ (ps)	$\rho(\Delta m_d, \tau_{\mathrm{B}^0})$
Δm_d,τ_{B^0}	0.0124	0.0166	−0.072
Δm_d,τ_{B^0},mistag	0.0132	0.0166	−0.057
Δm_d,τ_{B^0},$f_{\mathrm{B}^\pm}$	0.0163	0.0169	−0.161
Δm_d,τ_{B^0},$f_{\mathrm{B}^\pm}$,mistag	0.0174	0.0169	−0.149
all signal Δt	0.0178	0.0229	−0.224
default fit	0.0182	0.0230	−0.216

Table 33: Comparison of errors on Δm_d and τ_{B^0} and their correlation coefficients in full fits to Data in which different sets of parameters are floating.

6.3 Correction on Statistical Errors

The statistical errors reported by MINUIT are estimated from the curvature of the negative log-likelihood function at it minimum, and are assumed to be symmetric with respect to the minimum (HESSE errors). However, the log-likelihood function may not be symmetric and the real error (deviation from the minimum so that log-likelihood changes by 0.5 unit) may not be the same as that estimated from the curvature. To estimate the error at higher accuracy, we scan the log-likelihood function on the one-sigma contour on the Δm_d-τ_{B^0} plane predicted by HESSE, i.e., fix Δm_d and τ_{B^0} at values on the one-sigma contour and repeat fits with all other parameters floating. If the HESSE errors are correct, the log-likelihood should be greater than the minimum by exactly 0.5 unit.

The resulting values of eight points on the one-sigma contour are shown in Fig. 34. The percentages shown in parenthesis are how much the error should increase in that particular direction if a parabolic curve is assumed. The largest differece from 0.5 is in the positive τ_{B^0} direction. We correct the positive τ_{B^0} statistical error by 6.5%. For Δm_d and negative τ_{B^0} errors, the correction is negligible.

6.4 Fit Result Summary

The fit results and statistical errors for lifetime and B^0 mixing frequency, and their statistical correlation are:

$$\begin{aligned}
\Delta m_d &= 0.512 \pm 0.018 \text{ ps}^{-1} \\
\tau_{\mathrm{B}^0} &= 1.545 \,^{+0.024}_{-0.023} \text{ ps} \\
\rho(\Delta m_d, \tau_{\mathrm{B}^0}) &= -0.22
\end{aligned}$$

These numbers are before any correction, which is discussed in Sec. 7.2.

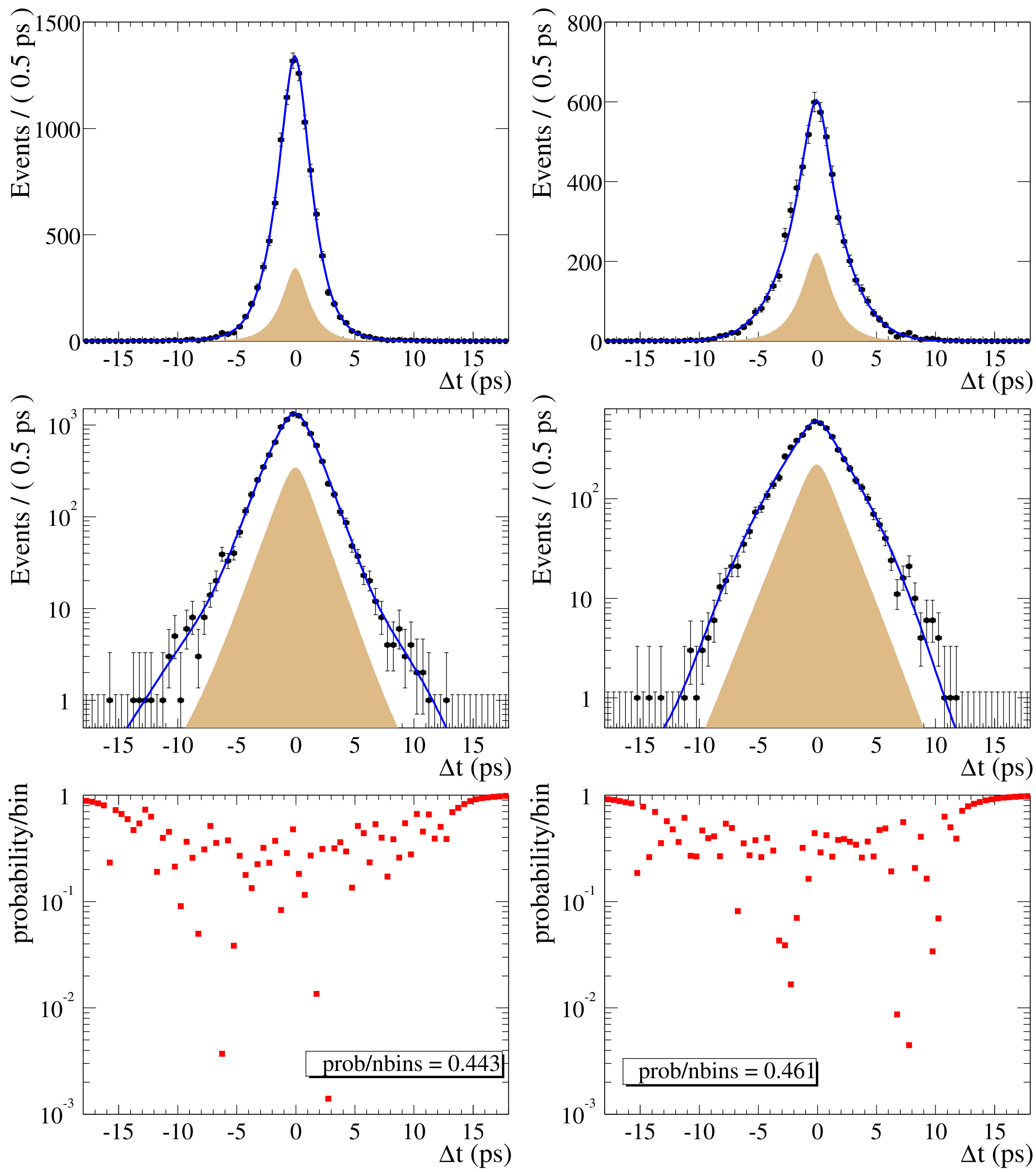

Figure 31: The Δt distribution of unmixed and mixed events in signal sample {On;OS;e or μ } of Data within the peak region $143 < m(\mathrm{D}^*) - m(\mathrm{D}^0) < 148$ MeV, and the projection of the model. The left hand plots are for unmixed events, the right for mixed events; the middle row is simply a log-y plot, The superimposed shaded areas on the upper plots show the background contribution to the distributions. The bottom row is the Poisson probability of observing n events or smaller (larger) if n is smaller (larger) than the expected value, i.e., $y = \sum_{i=0}^{n} P(i;\mu)$ for $n < \mu$ and $y = \sum_{i=n}^{\infty} P(i;\mu)$ for $n > \mu$.

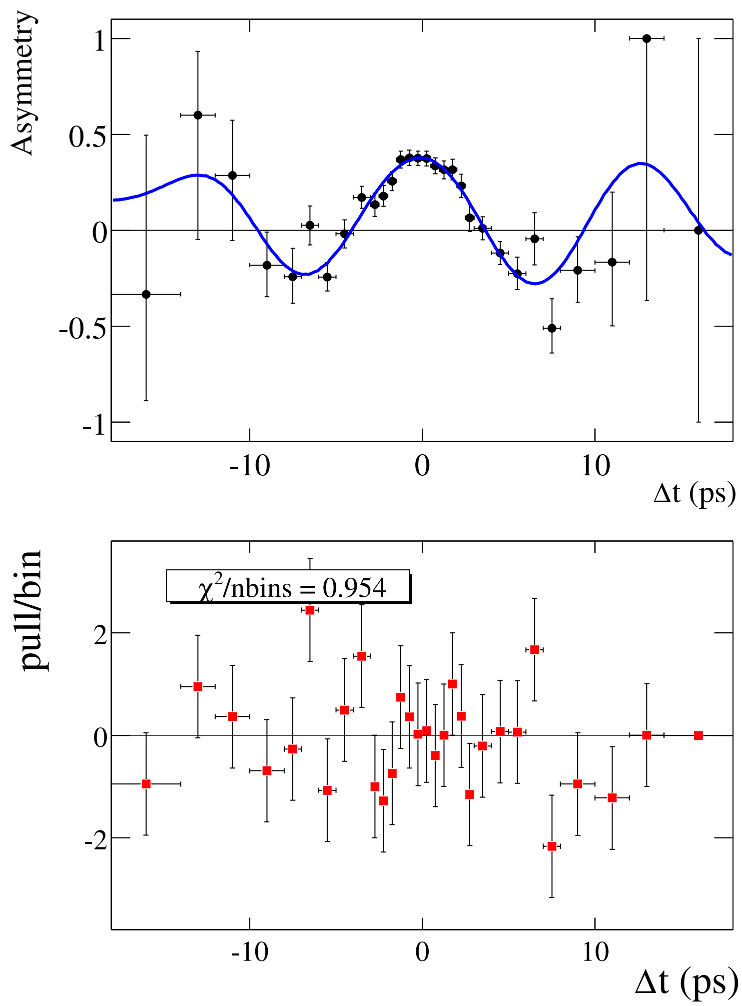

Figure 32: The asymmetry plot for signal sample {On;OS;e or μ } of Data within the peak region $143 < m(\mathrm{D}^{*}) - m(\mathrm{D}^{0}) < 148$ MeV, and the projection of the model fit result from Data. The lower plot shows the bin-by-bin difference normalized by the error estimated from data on each bin.

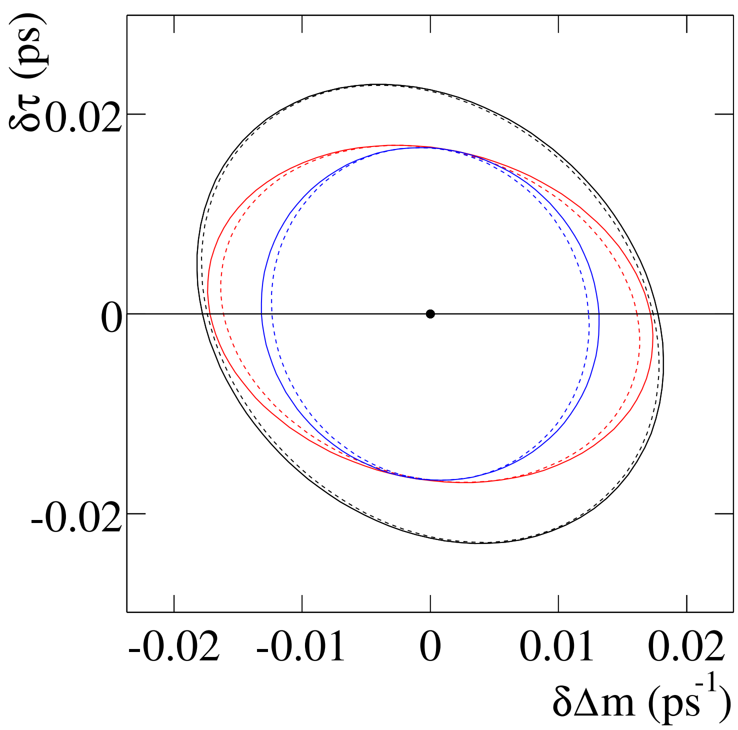

Figure 33: Comparison of one-sigma likelihood contours in the Δm_d- τ_{B^0} plane for fits in which different sets of parameters are floating. From the innermost to the outermost ellipse, the floating parameters are (Δm_d, τ_{B^0}), (Δm_d, τ_{B^0}, mistag fractions), (Δm_d, τ_{B^0}, $f_{B^\pm}$), (Δm_d, τ_{B^0}, f_{B^+}, mistag fractions), all signal Δt parameters, and the default fit (72 floating parameters).

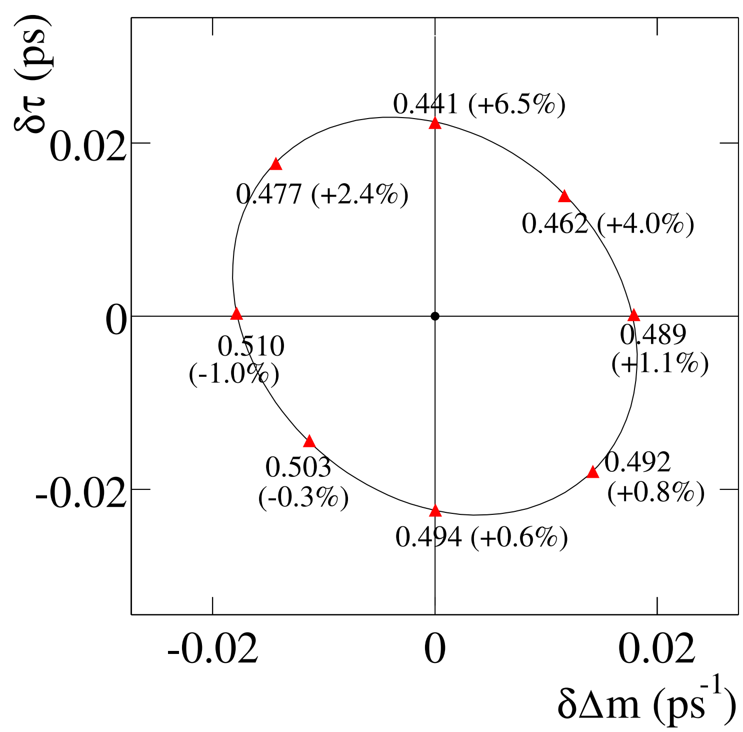

Figure 34: The lig-likelihood difference between points on one-sigma contours and the minimum. The percentages shown in parenthesis are how much the error should increase in that particular direction if a parabolic curve is assumed.

7 Validation and Toy MC Studies

7.1 Validation Fits to Generic Monte Carlo Samples

In this section, we describe fits to generic Monte Carlo samples with complete fitting procedures. The sample is a mixture of 14.5 $\text{fb}^{-1}$of $B^0\overline{B^0}$, 16.4 $\text{fb}^{-1}$of B^+B^- and 7.6 $\text{fb}^{-1}$of $c\overline{c}$ events. Because the event selection efficiency for Monte Carlo events is higher than that for data, we have more events (72677) selected from generic MC than from data (67904). Since we do not have off-resonance Monte Carlo sample, we artifically divide the $c\overline{c}$ sample into two samples; one has seven times as many events as the other, and the smaller sample is marked as off-resonance events.

For this validation exercise, we perform a complete fit to generic Monte Carlo. The total number of floating parameters is 72: 22 in signal model, 24 in combinatoric model and 26 in peaking background model. The ratio of B^+ to B^0 lifetimes is fixed at their generated value, $r = 1.069$, and the ratios of B^+ to B^0 mistag fractions are fixed at the values from MC truth counting (Table 56).

We conduct a study similar to that described in Sec. 6.1 to determine the best values for the signal outlier model shape parameters, b^{out} and s^{out}. We cannot expect the population of outliers in Generic MC to be identical to those in true Data, so we undertake the full "scan" of fixed outlier shape parameters as before. As suggested in Sec. 5.4.4, we appear to be unlucky with Generic MC in that the preferred value of the outlier fraction is close to zero and slightly negative.

The likelihood surface for Generic MC in the $b^{\text{out}}-s^{\text{out}}$ space is much different than that for Data. All of f the 33 different fits used for the scan in outlier shape space, resulted in an essentially zero outlier fraction, and a large f_{B^+}. For comparison, we include in Fig. 35- 37 some summary information from the full set of fits. We observe that since the outlier fraction is essentially zero, the minimum likelihood value (and fit result) is mostly insensitive to the details of the outlier shape. It is only at the extreme values of the outlier shape parameters ($b^{\text{out}} = -8, s^{\text{out}} = 10$) ps that the likelihood starts changing dramatically – this is where the outlier fraction is becoming increasingly negative, about -0.3%.

We choose the fixed values for the Generic MC MasterModel for two reasons

- The signal MC fits using just the signal Δt model preferred outlier shape parameters ($b^{\text{out}} = -1.68 \pm 0.99$, $s^{\text{out}} = 5.45 \pm 0.86$) ps.
- The weight of the (Δm_d,τ_{B^0}) distribution due to variations in the outlier parameters is centered near $(b^{\text{out}}, s^{\text{out}}) = (-2, 6)$ and $(0, 8)$ ps. (See Fig. 37.)

We prefer to use the set of fixed integer values closest to the fit results obtained from the high statistics signal MC sample, so we choose, for Generic MC, to fix the signal outlier shape parameters to be $(b^{\text{out}}, s^{\text{out}}) = (-2, 6)$. (We note in passing that the other preferred value (8,0) is what other time dependent analyses have assumed for their outlier models.)

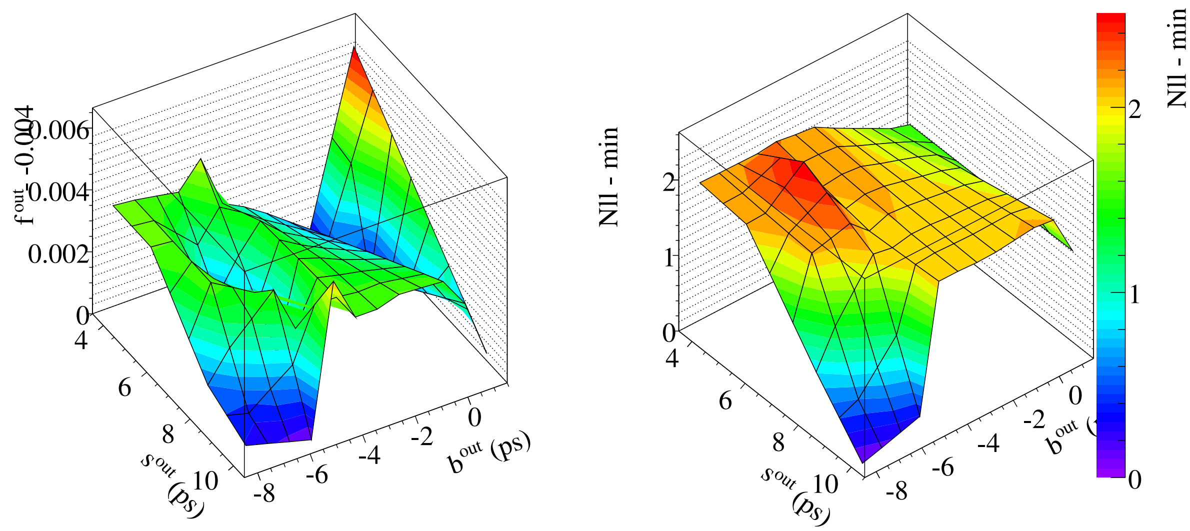

Figure 35: The negative-ln-likelihood surface (right) and f^{out} surface (left) in the space of bias and outlier width of the outlier signal resolution model, as fitted in Generic MC. The vertical scale has been offset so that minimum is at 0 (note that the minimum for f^{out} is at -0.004, so the surface has been moved by -0.004 units). The surface is constructed from 33 fits to Generic MC with different fixed values of the signal outlier parameters. This surface is essentially flat until the edges where the extreme values of the model make it unstable, *i.e.* the fitted outlier fraction becomes significantly less than 0. We do not observe a physical, preferred minimum.

The signal Δt model parameters are shown in Table 34, the combinatoric Δt model parameters in Table 36, and the three peaking background Δt model parameters in Table 37. Table 35 compares the fit results for Generic MC to the true values from the event generator, and to the best-fit results from the signal fit to the truth-matched Generic MC sample.

These fit results bear some discussion, particularly the signal parameters. Several observations are in order:

- The observed outlier fraction is low by, perhaps, about 1σ.
- The measured Δm_d value is 1σ high, the measured τ_{B^0} value is 1.4σ too high, each with respect to the true values from the generator level. They are 1.5σ and 0.58σ too high (respectively) compared to the values from the truth-matched sample of Generic B^+B^- and B^+.
- The outlier fraction is still consistent with zero, as originally observed in the fits to only only truth-matched candidates from Generic MC.

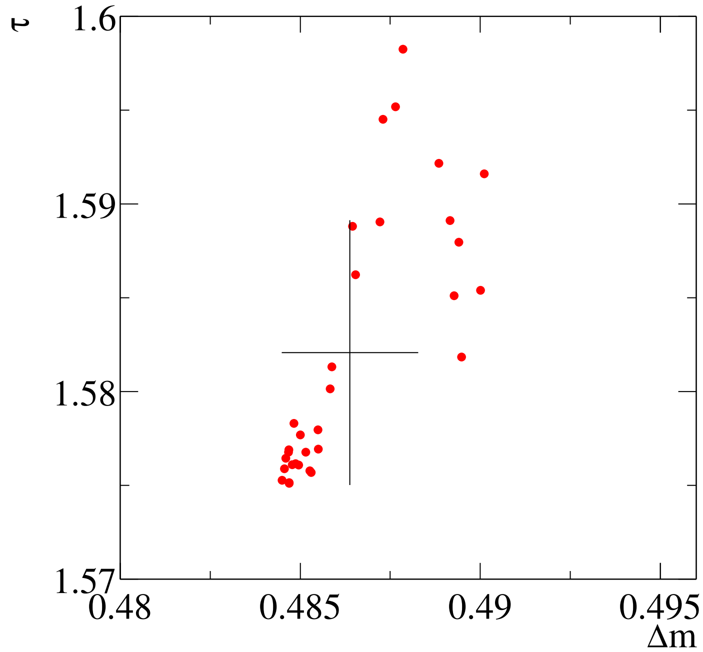

Figure 36: Scatter plot of the set of Δm_d, τ_{B^0} values obtained from data for 33 fits with different fixed values of the outlier bias and width in the range 1 to -7 ps for bias and 4 to 10 ps for width. The cross indicates the mean and rms of the Δm_d, τ_{B^0} distribution. Recall that the statistical error from the full fit is 0.023 ps for τ_{B^0} and 0.014ps^{-1} for Δm_d for the Generic MC fit. We observe, then, that the variation in τ_{B^0} and Δm_d due different assumptions about the fixed values of the signal outlier shape is comparatively minor.

- The charged B fraction, f_{B+}, is significantly higher, by about 1.6σ with respect to the true value, or 1.65σ with respect to the result from the truth-matched fit.
- The observed slope for the linear relation between m_{KT} and $\sigma_{\Delta t}$ is significantly higher, by 2.1σ. Additionally, all the mistag rates seems perturbed by at least 1σ, and this is an underestimate because the truth-match signal events are a subset of the full Generic MC data sample.

The large value of f_{B+} is surprising, in particular since the result from the full Master-Model fit should be highly correlated with the result from the fit to only the truth-matched signal events. We suspect that the our Generic MC sample has an unusually small population of outliers in the large Δt tails which degrades our ability to fit for their overall fraction. This inability, in turn, weakens the signal resolution model which convolutes the charged and neutral B time structures together.

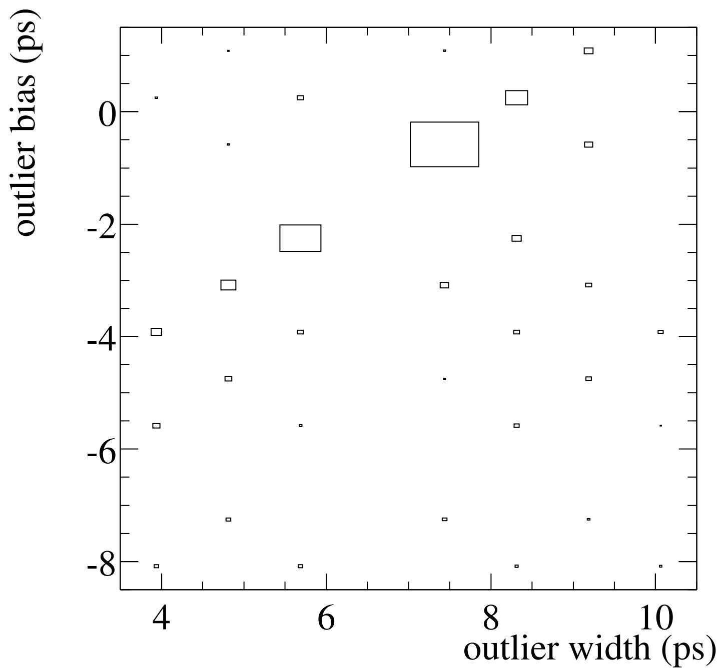

Figure 37: Distribution of full MasterModel fit results for fits to Generic MC. The grid represents the space of different assumptions about the two outlier shape parameters, and the weight assigned to each point represents the proximity of the (Δm_d,τ_{B^0}) fit results to the mean and rms mentioned in the previous plot. From this distribution, we see that the grid point (-2,6) and (8,0) are preferred as being sets of outlier shape parameters that yield Δm_d, τ_{B^0} fit results most near the center of their distribution.

To investigate this claim, we observe the correlation coefficients in the final MasterModel Generic MC result shown in Table 39. We see that an upward fluctuation in $f_{\mathrm{B}+}$ by 1.6σ would induce a downward flucctuation in f^{out} of about 0.5σ, and cause a positive change in Δm_d of about 0.8σ, The low f^{out} could induce an increase in τ_{B^0} of about 0.4σ. These shifts are not quite sufficient to explain the parameters we observe in the full fit result, however. We do, see, though, that if we fix $f_{\mathrm{B}+}$ to lower values, we can recover performance, as illustrated in Table 38.

Full Fit to Generic MC, Signal Model Parameters					
parameter	value	parameter	value	parameter	value
Δm_d (ps^{-1})	0.486 ± 0.015	$f_{\mathrm{B}^{\pm}}$	0.095 ± 0.028	s^1	0.981 ± 0.064
τ (ps)	1.580 ± 0.024	ω_{LT}	0.039 ± 0.011	κ	1.19 ± 0.19
-	-	$\omega_{\mathrm{KT}}^{\mathrm{offset}}$	0.015 ± 0.027	${f^1}_{\mathrm{LT}}$	0.884 ± 0.055
-	-	m_{KT}	0.200 ± 0.039	f^1_{KT}	0.679 ± 0.060
-	-	ω_{N1}	0.180 ± 0.018	${f^1}_{\mathrm{N1}}$	0.670 ± 0.092
-	-	ω_{N2}	0.352 ± 0.015	${f^1}_{\mathrm{N2}}$	0.708 ± 0.073
-	-	ω_{N3}	0.445 ± 0.011	${f^1}_{\mathrm{N3}}$	0.743 ± 0.054
-	-	$\Delta\omega_{\mathrm{LT}}$	-0.024 ± 0.018	f^{out}	-0.0019 ± 0.0022
-	-	$\Delta\omega_{\mathrm{KT}}$	-0.026 ± 0.014	b^{out}	-2.00
-	-	$\Delta\omega_{\mathrm{N1}}$	0.012 ± 0.029	s^{out}	6.00
-	-	$\Delta\omega_{\mathrm{N2}}$	-0.004 ± 0.024	-	-
-	-	$\Delta\omega_{\mathrm{N3}}$	-0.096 ± 0.018	-	-

Table 34: Full fit to generic MC — signal model parameters. (Recall that s^{out} and b^{out} for the signal resolution model are fixed to constant values.)

Full Fit to Generic MC, Comparing Signal Model Pars					
$\delta(par)$	value	$\delta(par)$	value	$\delta(par)$	value
Δm_d (ps^{-1})	0.014 ± 0.015	$f_{\mathrm{B}^{\pm}}$	0.043 ± 0.028	s^1	0.026 ± 0.064
τ (ps)	0.032 ± 0.024	ω_{LT}	-0.015 ± 0.011	κ	0.041 ± 0.19
-	-	$\omega_{\mathrm{KT}}^{\mathrm{offset}}$	-0.062 ± 0.027	${f^1}_{\mathrm{LT}}$	0.040 ± 0.055
-	-	m_{KT}	0.081 ± 0.039	f^1_{KT}	-0.023 ± 0.060
-	-	ω_{N1}	0.041 ± 0.018	${f^1}_{\mathrm{N1}}$	-0.067 ± 0.092
-	-	ω_{N2}	0.041 ± 0.015	${f^1}_{\mathrm{N2}}$	-0.057 ± 0.073
-	-	ω_{N3}	-0.029 ± 0.011	${f^1}_{\mathrm{N3}}$	0.011 ± 0.054
-	-	$\Delta\omega_{\mathrm{LT}}$	-0.015 ± 0.018	f^{out}	0.0024 ± 0.0022
-	-	$\Delta\omega_{\mathrm{KT}}$	-0.020 ± 0.014	b^{out}	0.32
-	-	$\Delta\omega_{\mathrm{N1}}$	0.041 ± 0.029	s^{out}	0.55
-	-	$\Delta\omega_{\mathrm{N2}}$	0.041 ± 0.024	-	-
-	-	$\Delta\omega_{\mathrm{N3}}$	-0.029 ± 0.018	-	-

Table 35: Full fit to generic MC — comparison of signal model results to generator truth (for Δm_d, τ_{B^0}, and f_{B^+}) and truth-matched generic MC results for everything else. (Recall thats^{out} and b^{out} for the signal resolution model are fixed to constant values.) Errors listed are the error on the Gen MC fit result, the appropriate σ for measuring degree of difference.

Full Fit to Generic MC, Combinatoric Δt Model Parameters			
parameter	value	parameter	value
Δm_{comb}	0.394 ± 0.018	$b^1{}_{\mathrm{LNT12}}$	-0.060 ± 0.022
τ_{comb}	1.275 ± 0.020	$b^1{}_{\mathrm{KNT3}}$	-0.159 ± 0.012
$\omega^{\mathrm{osc}}_{\mathrm{Fake;LTag}}$	0.367 ± 0.023	$f^{\mathrm{osc}}_{\mathrm{Off}}$	0.150 ± 0.053
$\omega^{\mathrm{osc}}_{\mathrm{Fake;KTag}}$	0.194 ± 0.013	$f^{\mathrm{osc}}_{\mathrm{SS}}$	0.469 ± 0.025
$\omega^{\mathrm{osc}}_{\mathrm{Fake;NT1}}$	0.381 ± 0.027	$f^{\mathrm{osc}}_{\mathrm{NoLept}}$	0.686 ± 0.018
$\omega^{\mathrm{osc}}_{\mathrm{Fake;NT2}}$	0.444 ± 0.017	$f^{\mathrm{osc}}_{\mathrm{OneLept}}$	0.900 ± 0.018
$\omega^{\mathrm{osc}}_{\mathrm{Fake;NT3}}$	0.533 ± 0.014	$f^{\mathrm{osc}}_{\mathrm{TwoLept}}$	1.000
$\omega^{\mathrm{osc}}_{\mathrm{Lept;LTag}}$	0.215 ± 0.023	$s^1{}_{\mathrm{OffSS}}$	1.242 ± 0.030
$\omega^{\mathrm{osc}}_{\mathrm{Lept;KTag}}$	0.258 ± 0.010	$s^1{}_{\mathrm{On;OS}}$	1.129 ± 0.025
$\omega^{\mathrm{osc}}_{\mathrm{Lept;NT1}}$	0.269 ± 0.024	f^1	0.9837 ± 0.0037
$\omega^{\mathrm{osc}}_{\mathrm{Lept;NT2}}$	0.426 ± 0.016	b^{out}	-1.57 ± 0.42
$\omega^{\mathrm{osc}}_{\mathrm{Lept;NT3}}$	0.494 ± 0.011	s^{out}	6.80 ± 0.53
ω^{pmt}	0.427 ± 0.012	-	-

Table 36: Full fit to generic MC — combintoric Δt model parameters.

Full Fit to Generic MC, Peaking BG Model Parameters					
cont. par.	value	fake par.	value	unco. par.	value
ω_{KT}	0.076 ± 0.029	Δm^{fake}	0.405 ± 0.023	τ^{unco}	1.68 ± 0.35
ω_{LNT}	0.503 ± 0.037	τ^{fake}	1.386 ± 0.038	$\omega^{\mathrm{life}}_{\mathrm{KNT}}$	0.67 ± 0.13
$b^1{}_{\mathrm{OS}}$	-0.022 ± 0.104	$\omega^{\mathrm{osc}}_{\mathrm{LT}}$	0.192 ± 0.025	$\omega^{\mathrm{life}}_{\mathrm{LT}}$	0.999
$b^1{}_{\mathrm{SS}}$	-0.092 ± 0.111	$\omega^{\mathrm{osc}}_{\mathrm{KT}}$	0.180 ± 0.020	ω^{pmt}	0.58 ± 0.07
s^1	1.286 ± 0.076	$\omega^{\mathrm{osc}}_{\mathrm{NT1}}$	0.278 ± 0.029	f^{life}	0.37 ± 0.12
f^1	0.963 ± 0.024	$\omega^{\mathrm{osc}}_{\mathrm{NT2}}$	0.381 ± 0.028	b^1	-0.44 ± 0.13
-	-	$\omega^{\mathrm{osc}}_{\mathrm{NT3}}$	0.508 ± 0.024	f^1	0.993
-	-	$\omega^{\mathrm{pmt}}_{\mathrm{KNT3}}$	0.252 ± 0.086	s^1	1.33 ± 0.17
-	-	$\omega^{\mathrm{pmt}}_{\mathrm{LNT12}}$	0.471 ± 0.106	-	-
-	-	f^{osc}	0.881 ± 0.027	-	-
-	-	$b^1{}_{\mathrm{KNT3}}$	-0.216 ± 0.041	-	-
-	-	$b^1{}_{\mathrm{LNT12}}$	-0.043 ± 0.045	-	-
-	-	f^1	0.965 ± 0.010	-	-
-	-	s^1	1.062 ± 0.068	-	-

Table 37: Full fit to generic MC — peaking background Δt model parameters. The outlier bias (b^{out}) and scale factor (s^{out}) for these peaking background are shared with combinatoric background parameters.

Full fit results to Generic MC w/fixed f_{B^+}						
f_{B^+}	$\delta(f_{B^+})$	Δm_d	$\delta(\Delta m_d)$	τ_{B^0}	$\delta(\tau_{B^0})$	f^{out}
0.0474	-0.0047	0.4718	-0.0002	1.5718	0.0238	-0.00008
0.0521	0.0000	0.4732	0.0012	1.5728	0.0248	-0.0003
0.0600	0.0079	0.4756	0.0036	1.5744	0.0264	-0.0006
0.0700	0.0179	0.4786	0.0066	1.5764	0.0284	-0.0011
0.0951	0.0430	0.4856	0.0136	1.5806	0.0326	-0.0019

Table 38: Selected results from full MasterModel fits to Generic MC with fixed f_{B^+} to several different values. We directly observe the strong correlations between these four parameters. The columns labelled $\delta()$ reflect the change with respect to the value from MC truth. Recall that the statistical error on Δm_d from these fits is around 0.015 ps^{-1} and about 0.024 ps for τ_{B^0}.

Corr. coeff's from full fit to Gen MC			
pars	value	pars	value
$\rho(\Delta m_d, \tau_{B^0})$	-0.162	$\rho(m_{KT}, \omega_{KT}^{offset})$	-0.944
$\rho(f^{out}, \Delta m_d)$	-0.169	$\rho(f_{B^+}, \Delta m_d)$	0.552
$\rho(f^{out}, \tau_{B^0})$	-0.402	$\rho(f_{B^+}, \tau_{B^0})$	0.162
$\rho(f^{out}, m_{KT})$	-0.018	$\rho(f_{B^+}, m_{KT})$	0.086
$\rho(f^{out}, \omega_{KT}^{offset})$	0.004	$\rho(f_{B^+}, \omega_{KT}^{offset})$	-0.060
$\rho(f^{out}, f_{B^+})$	-0.333		

Table 39: Correlation coefficients from baseline MasterModel fit to Generic MC. (f_{B^+} is correlated with all the mistag rates at only the few % level.)

7.2 Correction Based on Generic Monte Carlo Sample

There appears to be a bias between the fit to signal events and the full fit to signal plus background events for generic Monte Carlo sample (see Table 40). The biases are of the same order as the statistical error on data. If we simply assign the generic MC statistical error as a systematic on this bias, this systematic will be as large as the statistical error on data.

Because lack of generic MC, instead of purely fitting to generic MC with full model and quote its statistical error as a systematic error, we investigate the possible bias with two steps: fit to a much larger signal MC sample with signal model only, and compare fits between signal model fit and full fits to generic MC. The hope is that we can increase the sensitivity of generic MC fits through the connection to the signal model fit to larger signal MC. In other words, because the accuracy of large signal MC fits is quite high and the uncertainty of the change from signal model fit to full fit (to generic MC) should be smaller than the statistical error in full fit alone (due to overlapped signal events), the combined error should in principle be smaller than the statistical error in full fit alone.

Unfortunately, it is difficult to estimate the error on the change from the signal model fit to the full fit, and the change itself looks uncomfortably large. In order to increase sensitivity, we include the filtered generic Monte Carlo (see Appendix A). Although the generic + filtered MC have slightly different background composition from only generic MC, the dominant background (combinatoric) are similar. More over, the biases are consistent between fits to generic MC sample only and fits to generic + filtered MC sample. Therefore, we decided to use the latter.

The fit to signal events with signal model has been shown to be unbiased with large statistics of signal Monte Carlo sample (Table 14 in Sec. 5.2.1). Therefore we test the bias of the full model based on the difference between the fit to signal events with with signal model and the full fit to signal and background events.

The bias is $+0.022$ ps for τ_{B^0} and $+0.020$ ps^{-1} for Δm_d. We apply a correction for these biases. For the systematic error due to the statistical uncertainty on this correction, we conservatively apply the full statistical error of the fit to signal+background, i.e., ± 0.018 ps for τ_{B^0} and ± 0.012 ps^{-1} for Δm_d.

7.3 Change in Kaon Mistag Slope

The kaon mistag is modeled as a linear function of $\sigma_{\Delta t}$ in signal model. The slope from the fit to signal only in generic MC is 12%, while the slope from the full fit to signal and background events is $20 \pm 4\%$. The change is due to the fact that we ignore the similar effect in background events, i.e., kaon mistag in background model is flat. We can confirm that the kaon mistag in background events do increase as $\sigma_{\Delta t}$ increases by calculating the observed *mixed* events in background in slices of $\sigma_{\Delta t}$. Kaon tagged events have a slope similar to kaon tagged signal events, while events in other tagging categories do not have a non-zero slope.

Fits to Generic Monte Carlo Sample					
Type	τ_{B^0}	change(10^{-3})	Δm_d	change(10^{-3})	change w.r.t.
generator	1.548	–	0.472	–	-
truth fit	1.538 ± 0.011	-10 ± 11	0.469 ± 0.005	-3 ± 5	generator
signal model	1.566 ± 0.023	$+28 \pm 20$	0.462 ± 0.013	-7 ± 12	truth fit
full model	1.580 ± 0.024	$+14 \pm 6$	0.486 ± 0.015	$+24 \pm 6$	signal model
Fits to Generic + Filtered Monte Carlo Sample					
Type	τ_{B^0}	change(10^{-3})	Δm_d	change(10^{-3})	change w.r.t.
truth fit	1.556 ± 0.009	$+8 \pm 9$	0.470 ± 0.004	-2 ± 4	generator
signal model	1.554 ± 0.016	-2 ± 13	0.459 ± 0.011	-12 ± 10	truth fit
full model	1.576 ± 0.018	$+22 \pm 9$	0.479 ± 0.012	$+20 \pm 5$	signal model

Table 40: Fit results and biases for generic Monte Carlo sample and for generic + filtered Monte Carlo sample. The errors on changes are calcuated as difference in quadrature between two errors. This way of estimating error on the change may NOT be adequate.

The correlation coefficients between this slope and (Δm_d, τ_{B^0}) are very small, such that fixing this slope at 12% does not change τ_{B^0} and Δm_d by more than 0.001 ps or ps^{-1}. Therefore, we can safely ignore the slope for backgrounds.

7.4 Toy Monte Carlo Studies

We test the fitting procedure and the validity of the full model with toy Monte Carlo samples. We generate a set of toy Monte Carlo datasets of exactly the same number of events as data. The central values of the parameters are from the result of the default fit to data. (Note that the toy MC generator is fed the blinded physics parameters from data. These parameters are internally unblinded to generate the toy MC events. The fit parameters from the toy MC are then blinded with the same blinding string as data.) The toy generator goes through each event in data and chooses a set of parameters corresponding to that particular event based on its categoreies (`svtDch`, `Dmode`, `tagCat`, `angCut`, `onOffRes`, `leptID`, etc.), calcuates the signal and background probabilities according to its $m(D^*) - m(D^0)$ value, and scales the resolution function based on $\sigma_{\Delta t}$. The variables that are generated/randomized are Δt, mixing state and tag-side B flavor.

An ensemble of 87 Monte Carlo data sets are generated and fitted with the full model. The distributions of the fit values of Δm_d, τ_{B^0} and $f_{B^\pm}$ are shown in Fig. 38. The mean and RMSs of Δm_d, τ_{B^0} and $f_{B^\pm}$ are:

$$\begin{array}{rll} \Delta m_d & \text{mean} = 0.5280 \pm 0.0018\ \text{ps}^{-1} & \text{RMS} = 0.0171 \pm 0.0013\ \text{ps}^{-1} \\ \tau_{B^0} & \text{mean} = 1.5390 \pm 0.0024\ \text{ps} & \text{RMS} = 0.0223 \pm 0.0017\ \text{ps} \\ f_{B^\pm} & \text{mean} = 0.076 \pm 0.004 & \text{RMS} = 0.034 \pm 0.003\ , \end{array}$$

and should be compared with the central value and statistical error on the (blind) fit values from data:

$$\begin{aligned} \Delta m_d &= 0.527 \pm 0.018\ \text{ps}^{-1}\ \text{(blind)} \\ \tau_{B^0} &= 1.539 \pm 0.023\ \text{ps (blind)} \\ f_{B^\pm} &= 0.082 \pm 0.029\ . \end{aligned}$$

The central values of Δm_d, τ_{B^0} and $f_{B^\pm}$ from the fit to data (used to generate the toy MC samples) are shown as vertical dashed lines in Fig. 38. Both means and RMS's are consistent between fits to toy MC and data. The correlation coefficient calculated from the Δm_d-τ_{B^0} distribution is -0.21, which is in a good agreement with the fit to data, -0.22.

The distribution of minimized negative log-likelihood (minNll) is shown in Fig. 39. The number of toy MC samples that result in a smaller (better) value of minNll is 17 (out of 87).

7.5 Comparison with Hadronic Mixing Analysis

BABAR has recently published a measurement of Δm_d based on fully reconstructed hadronic final states in 30/fb of data. In this section, we compare the statistical errors on mixing from the hadronic and semileptonic analyses.

In the hadronic analysis, ≈6k *tagged* events are used. Events in the NT3 category are not used. Our sample consists of about 14k events, including those in the NT3 category, which do not have much sensitivity to mixing. Therefore, to compare the statistical errors on mixing, we should compare the 6k tagged hadronic signal candidates to the number of semileptonic candidates not including NT3, which is about 9k. We then expect the ratio of statistical errors in the two analyses to be ≈ 0.8 when the same parameters are free in the hadronic and semileptonic fits. Table 41 shows the ratio of statistical errors on Δm_d when equivalent parameters are free in the fit. The observed ratio of statistical errors is in the range 0.76 to 0.91, depending on which parameters are free in the fit, in reasonable agreement with the simple expectation of 0.8.

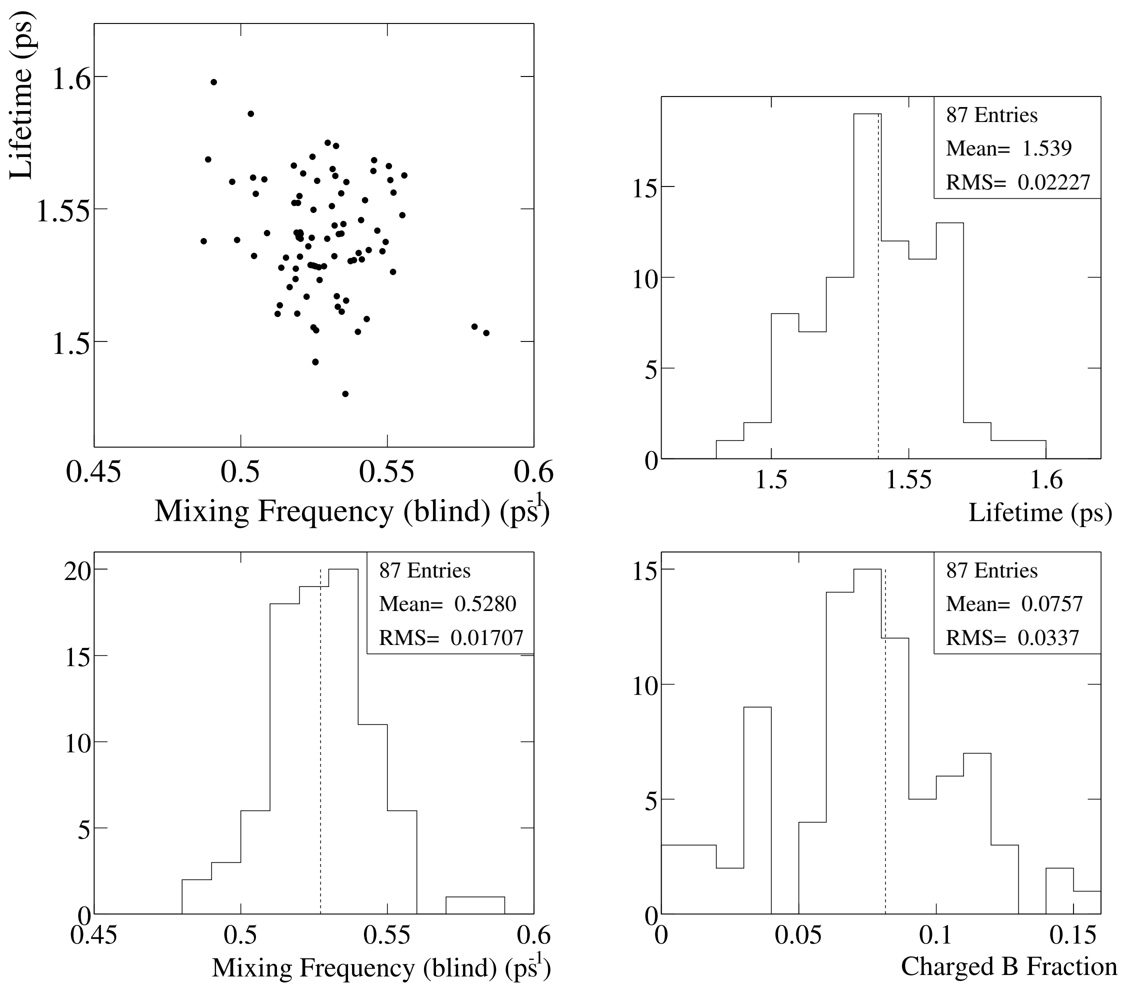

Figure 38: The distributions of τ_{B^0}, Δm_d and $f_{B^\pm}$ from 87 toy Monte Carlo sample fits. The values from the default fit to data, which were used to generate the toy MC samples, are indicated by vertical dashed lines.

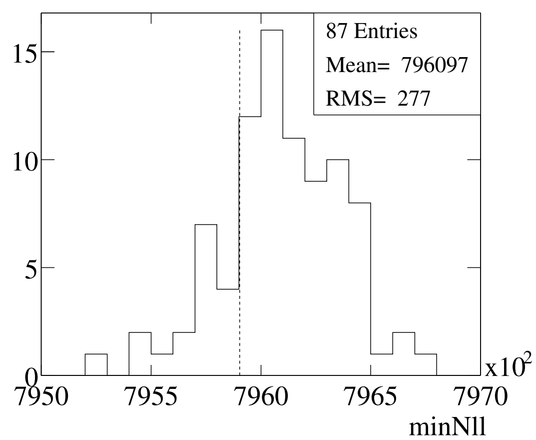

Figure 39: The distribution of minimized negative log-likelihood (minNll) from 87 toy Monte Carlo sample fits. The value from default fit to data is indicated by the vertical dashed lines. The number of fits that result in a smaller (better) value of minNll is 17.

Fit model	Stat error on Δm_d		Ratio
	Hadronic analysis	This analysis	
Float only Δm_d	0.0139	0.0122	0.88
Float only Δm_d, ω	0.0147	0.0129	0.88
Float everything, except τ_{B^0} and f_{B^+}	0.0162	0.0147	0.91
Float everything, except τ_{B^0}	0.0232	0.0177	0.76

Table 41: Comparison of statistical errors on Δm_d in this analysis and the hadronic mixing analysis when similar parameters are free in the fits. Given the ratio of the number of signal candidates in each sample, we expect the ratio of statistical errors to be about 0.8.

8 Cross Checks

A large number of consistency checks have been performed to ensure the validity of the results. This section describes the various cross checks performed using both Monte Carlo and real data. Many of the tables in this section compare fit results to a standard or default fit, the baseline. The standard fit is defined as the full fit to Data shown in Tables 29, 30 and 31 in Sec. 6 using the Master Model, or the full fit to Monte Carlo shown in the last column of Table 53. For ease of reading, most tables in this section will show the difference between the parameters from a cross-check fit and the standard fit parameters. If all the differences are essentially zero for a group of parameters, this will be noted with a single line entry in the table.

8.1 Fit to MC truth

Sample	generator value	MCtruth Δt fit	MCtruth Δz fit	signal model fit
Δm_d (ps^{-1})	0.472	0.4708±0.0013	0.4723±0.0007	0.4715±0.0043
τ_{B^0} (ps)	1.548	1.5446±0.0051	1.5494±0.0051	1.5523±0.0091
ω_{LT}	-	0.063	0.063	0.056±0.003
ω_{KT}^{offset}	-	0.157	0.157	0.067±0.008
m_{KT}	-	-	-	0.127±0.012
ω_{N1}	-	0.192	0.192	0.184±0.006
ω_{N2}	-	0.338	0.338	0.339±0.005
ω_{N3}	-	0.465	0.465	0.462±0.004

Table 42: Fitting for τ_{B^0}, mistag rate and Δm_d using signal MC candidates. The MC truth fits use truth information from all correctly reconstructed events in the signal MC sample. The standard fit uses reconstruction information from all correctly reconstructed signal MC events and fixes the outlier bias and width. The MC truth mistag rates are calculated by comparing the reconstructed mixing status with the truth mixing status for all events. The standard fit uses a GExp+G resolution model, the resolution fit parameters are the same as shown in Table 53.

The fit to true values of Δt and Δz were already described in Sec. 5.2.1. The results are summarized in Table 42. For the MC truth Δt fits, the value of τ_{B^0} deviates from the generator value of 1.548 ps by -0.0030±0.0051 ps (about -0.6σ) while the value of Δm_d deviates from the generator value of 0.472 ps^{-1} by -0.0014±0.0012 ps^{-1} (about -1.2σ). For the MC truth Δz fits, the value of τ_{B^0} deviates from the generator value by 0.0014±0.0051 ps (about 0.3σ) while the value of Δm_d deviates from the generator value by 0.0004±0.0012 ps^{-1} (about 0.3σ). A plot of τ_{B^0} fit is shown in Fig. 8 using MC truth Δt and Δz information.

8.2 True vs Reconstructed Flavor Tagging

One consistency check that can be performed is to use MC truth to determine mistag rates and compare this to values determined from the full fit to signal events. The value of χ_d can also be extracted and compared to the generator value. The χ_d comparison was already performed in the event selection (see Sec. 2.5). The counting results are summarized in Table 42.

The MC truth fit uses both true Δt and true tagging information. The mistag rates are calculated by comparing the recostructed mixing status with the true mixing status. A simple counting of incorrect mixing status events and total number of events yields the MC truth mistag rates. The last column is the "standard" fit which used both reconstructed Δt and tagging information. Only the last fit uses a resolution model, since the other fits have perfect Δt information. Also, the kaon mistag dependence on $\sigma_{\Delta t}$ is only valid for the last fit, the other fits don't have per-event-errors.

8.3 Δm_d from Δt-Shape or Flavor Information Only

In this analysis, we use both Δt-shape information and flavor-tagging information simultaneously to extract Δm_d, and τ_{B^0}. An important consistency check is to consider what value of Δm_d (and what statistical error) we might measure using only the time information or only the flavor tagging information (*i.e.,* mixing status) of each event. (By "flavor information," we refer to the relative fraction of the observed number of mixed and unmixed events; *i.e.,* a "time-integrated" mixing measurement.) We use the full signal Monte Carlo sample in this study to establish the highest precision for our conclusions.

To estimate Δm_d using only Δt-shape information (and the mixing status of the events, but not their relative number), we take advantage of one of the design features of **RooFit** – the lack of a strong distinction between parameters and dependents of a pdf at contruction time. The **RooFit** pdf only becomes useful when it is assigned to a dataset, which in turn it uses to distinguish which parameters are dependent variables (i.e. "input" variables, having a value for each event) and which are parameters (to be varied as part of the maximum likelihood fitting procedure). Our standard signal Δt model is associated with datasets containing both Δt and mixing status (tag) information. We eliminate the mixing information from the dataset as a dependent variable in the fit, and then can fit for Δm_d (and even τ_{B^0} and the resolution model if we choose) using only the Δt information. To achieve an answer most comparable with the the flavor-only result, we fix all parameters in the Δt-only fit to be constant except for Δm_d. For signal Monte Carlo, we then extract $\Delta m_d^{(shape-only)} = 0.47627 \pm 0.00487\ \mathrm{ps}^{-1}$, which can be compared with the full fit result of $\Delta m_d = 0.47257 \pm 0.00357$ (when only floating Δm_d).

To estimate Δm_d using only flavor information, we can naively caclulate it using χ_d as measured in signal Monte Carlo which amount to simply counting mixed and unmixed events. This yields $\chi_d = 0.174 \rightarrow \Delta m_d = 0.4705$ using the measured τ_{B^0} and in the absence of mistag rates. An alternative technique is to return the full signal fit, but fix the resolution function to a very, very wide gaussian (width > 18 ps) so that all Δt information is significantly smeared, and re-fit for Δm_d only. In this case, using the mistag rates obtained from the full fit, we obtain $\Delta m_d^{(flavor-only)} = 0.47361 \pm 0.00533\ \mathrm{ps}^{-1}$.

We note that the two results are consistent with the full fit, and both achieve a comparable statistical error, indicating that both shape and flavor information are important in obtaining the most precise value of Δm_d from the fit. (Note, for instance, the statistical error on the simultaneous fit is the quadrature sum of the inverses of the independent errors, i.e. $1/\sigma^2_{tot} = 1/\sigma^2_{shape-only} + 1/\sigma^2_{flavor-only}$.) (See Appendix G for additional remarks.)

8.4 Subsamples by Tagging Category

This fit was done for each tagging category, a common outlier and core scale factor are used. The mistag rate, core bias, τ_{B^0} and Δm_d are fit inside each tagging category. The mistag rate is made up of two parts, a constant offset, ω^{offset}, and a scale factor which multiplies the per-event-error, m. Thus $\omega = \omega^{offset} + m \cdot \sigma_{\Delta t}$. The scale factor is non-zero only for the kaon tagging category. The results are consistent across the five tagging categories, as shown in Table 43 for signal MC and Table 44 for the full fit to Data. The values of τ_{B^0} and Δm_d for each tagging category are consistent with the baseline results. (Comparing quadrature difference of statistical errors is not quite correct due to different character of this fit.)

8.5 Subsamples by B^0 Flavor

The sample was split into two subsamples, according to the reconstructed flavor of the signal B^0,$\overline{\mathrm{B}^0}$ and by tag side B^0,$\overline{\mathrm{B}^0}$ as a consistency check. The results are shown in Table 45 for signal MC, and in Table 46 for Data. For Data, we re-computed the set of per-event (signal and background) probabilties for the two subsamples to maximize our sensititvity to differences in reconstruction between B^0,$\overline{\mathrm{B}^0}$. Because we greatly on mixing time structure to separate the charged from neutral B candidates in the sample, these flavor-separated subsample fits have diminished precision. In some cases, listed in Table 46, we were forced to fix some of the signal model parameters to their baseline result as determined by the fit to the full sample in order to extract reliable results.

The differences in mistag rates for the Rec B^0, Rec $\overline{\mathrm{B}^0}$ categories are consistent with the Δ Dilutions measured in the standard fit. The differences in the Tag B^0, Tag $\overline{\mathrm{B}^0}$ samples come from the fact that most events are unmixed, thus there is a strong correlation between an event having a Tag B^0 and a Rec $\overline{\mathrm{B}^0}$. The correlation isn't perfect, so the difference in mistag rates isn't as large. The values of the two physics parameters appear to be consistent across the subsamples.

Signal Δt fit (GExp+G) to signal MC					
par.	LTag	KTag	NT1	NT2	NT3
$\delta(\Delta m_d)$	-0.004 ± 0.007	0.000 ± 0.006	0.011 ± 0.013	0.000 ± 0.019	0.012 ± 0.049
$\delta(\tau_{\mathrm{B}^0})$	-0.038 ± 0.016	-0.000 ± 0.012	0.007 ± 0.021	-0.000 ± 0.017	0.029 ± 0.014
$\delta(\Delta\omega)$	0.000 ± 0.005	-0.000 ± 0.006	-0.000 ± 0.010	0.000 ± 0.008	0.000 ± 0.006
$\delta(\omega^{\mathrm{offset}})$	0.001 ± 0.004	0.000 ± 0.009	-0.002 ± 0.006	0.000 ± 0.006	-0.000 ± 0.004
$\delta(m_{\mathrm{KT}})$		0.000 ± 0.012			
$\delta(f^1)$	-0.013 ± 0.025	-0.008 ± 0.024	-0.006 ± 0.031	-0.007 ± 0.027	0.003 ± 0.022
$\delta(\kappa)$			-0.024 ± 0.072		
$\delta(s^1)$			-0.006 ± 0.002		
$\delta(f^{\mathrm{out}})$			-0.000 ± 0.001		

Table 43: One signal model fit to correctly reconstructed signal Monte Carlo allowing the Δm_d and τ_{B^0} parameters to be different for each tagging category. The table shows the difference in the fit results with respect to the baseline result which uses the combined statistical power of all tagging categories. The fit is done with the GExp+G resolution model. The units for Δm_d and τ_{B^0} are ps^{-1} and ps. The lower section of the table shows parameters which are split on tagging category and are common for the whole sample; there we report the shift in central values as well. Because the fits are different, we quote the statistical error on the new fit result, which is a liberal estimate of the error.

8.6 Subsamples of Signal MC

The signal MC sample was split into four data size subsamples and each sample was fitted. The results are shown in Table 47. Unfortunately we don't have sufficient Monte Carlo data to equal ten or more data size samples to produce a reasonable histogram - however this is all we have. In the future more Monte Carlo data (hopefully generic) will be availible so a more complete check can be made.

8.7 Subsamples by D^0 Decay Mode

The fill fit to Data was performed separately for each D^0 decay mode group. Each `Dmode` category has roughly the same number of peak $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ events, although the Kπ mode is the cleanest due to the much reduced combinatoric backgrounds in the three-track final state signal mode. In Table 48 we compare the final fit values for each subsample of the Data as grouped by `Dmode`, and find no significant variation between subsamples.

8.8 Sensitivity to NT3 tagging information

Using the tagging information for events in the NT3 category allows us to treat all events on an equal footing for the lifetime and mixing measurement. Since the Q value for

	Full Δt fit (GExp+G) to Data				
par.	LTag	KTag	NT1	NT2	NT3
$\delta(\Delta m_d)$	-0.047 ± 0.026	0.016 ± 0.022	0.036 ± 0.045	0.012 ± 0.088	0.009 ± 0.183
$\delta(\tau_{B^0})$	0.011 ± 0.048	0.007 ± 0.032	-0.003 ± 0.059	0.023 ± 0.050	-0.066 ± 0.036
$\delta(\Delta\omega)$	-0.003 ± 0.004	0.000 ± 0.003	-0.002 ± 0.005	0.001 ± 0.008	0.000 ± 0.003
$\delta(\omega^{\text{offset}})$	0.007 ± 0.004	-0.005 ± 0.003	-0.008 ± 0.007	-0.001 ± 0.006	-0.001 ± 0.003
$\delta(m_{\text{KT}})$		0.000 ± 0.005			
$\delta(f^1)$	0.125 ± 0.077	0.113 ± 0.073	0.100 ± 0.097	0.100 ± 0.076	0.066 ± 0.064
$\delta(\kappa)$			-0.027 ± 0.079		
$\delta(s^1)$			0.009 ± 0.012		
$\delta(f^{\text{out}})$			-0.002 ± 0.001		

Table 44: Master Model fits to the full Dataset using separate Δm_d and τ_{B^0} parameters for each tagging category. The table shows the difference in the fit results with respect to the baseline result which uses the combined statistical power of all tagging categories to measure the physics parameters. The fit is done with the GExp+G resolution model. The units for Δm_d and τ_{B^0} are ps^{-1} and ps. The errors shown are the errors from the fit. The lower section of the table shows parameters which are already split on tagging category and are common for the whole sample; there we report the shift in central values as well.

this category is estimated to be about 0.3 %, it has only a small impact on the mixing measurement. This analysis is currently the only analysis in BABAR that uses tagging information from the NT3 category, which is called the "NoTag" category in other analyses. Since the NT3 tagging performance is not well verified in BABAR, we also perform a fit to Data in which we assume the tagging power of NT3 is 0. This is a non-trivial assumption – we are, in effect, assuming a value of χ_d for the NT3 tagged events when we do this; *i.e. we fix the mistag probability at 50% rather than fitting for it.*

We fix the NT3 mistag rates (ω_{NT3}) in signal and background Δt models to be 0.5. In a few parameters in background models, NT3 is grouped with other categories. We maintain the original parameter grouping, without separating NT3 from them, and still let these parameters float. Therefore, in this fit, we are not totally fixing NT3 tagging information, but very close.

The change in parameters are all very minor. When the NT3 tagging information is *not* used, the lifetime increases by 0.0022 ps and the statistical error increases by 0.002 ps larger (in quadrature); the change in Δm_d is $+0.0008$ ps^{-1} with an error that is larger by 0.0036 ps^{-1} (in quadrature). We therefore conclude that extracting NT3 mistag rates from data is warranted, and is actually slightly beneficial. (Clearly, including them in a lifetime-only fit is helpful.)

	Signal Δt fits (GExp+G) to signal MC subsamples			
	Reconstruction side		Tag side	
	Rec B^0	Rec $\overline{\mathrm{B}^0}$	Tag B^0	Tag $\overline{\mathrm{B}^0}$
$\delta(\Delta m_d)$ (ps^{-1})	-0.002 ± 0.005	0.002 ± 0.004	0.007 ± 0.004	-0.006 ± 0.005
$\delta(\tau_{\mathrm{B}^0})$ (ps)	-0.009 ± 0.010	0.010 ± 0.009	0.008 ± 0.008	-0.009 ± 0.009
$\delta(\omega_{\mathrm{LT}})$	0.001 ± 0.004	-0.002 ± 0.003	0.002 ± 0.004	-0.003 ± 0.003
$\delta(\omega_{\mathrm{KT}}^{\mathrm{offset}})$	-0.012 ± 0.009	0.012 ± 0.008	-0.007 ± 0.009	0.007 ± 0.008
$\delta(m_{\mathrm{KT}})$	0.019 ± 0.013	-0.019 ± 0.012	0.013 ± 0.013	-0.013 ± 0.012
$\delta(\omega_{\mathrm{N1}})$	0.005 ± 0.006	-0.006 ± 0.006	0.013 ± 0.006	-0.014 ± 0.006
$\delta(\omega_{\mathrm{N2}})$	-0.002 ± 0.005	0.002 ± 0.005	0.007 ± 0.006	-0.006 ± 0.005
$\delta(\omega_{\mathrm{N3}})$	0.002 ± 0.004	-0.002 ± 0.004	0.012 ± 0.004	-0.010 ± 0.004
	All Resolution Parameters Consistent			

Table 45: Signal Δt fits on subsamples of correctly reconstructed signal MC, separated by tagged or reconstructed flavor of the B^0, reconstructed B^0,$\overline{\mathrm{B}^0}$ or tag side B^0,$\overline{\mathrm{B}^0}$. We show the difference in central values with respect to the nominal fit result, and the errors shown are the uncorrelated errors – the difference in qudrature between the individual fit result and the nominal one (with fixed delta mistags). The fits use the standard GExp+G resolution function. Because of the division of the sample by B^0 flavor, we hold the $\Delta\omega$ mistag rates constant in the fit. The results are consistent between the flavor subsamples.

8.9 Varying Δt Selection Criterion

In this section, we determine the effect of the choice of Δt range to use in the fit. For the default fits to data and Monte Carlo, we use the full range used in the production of the ntuples: ± 18 ps. Table 49 shows the effect of using a smaller $|\Delta t|$ range for signal fits to signal Monte Carlo, while Table 50 shows the effect on the fits to data. In these fits, we fix the outlier parameters (bias, width and fraction) to the values determined from the fit to the full ± 18 ps Δt range. Note that the relative fraction of outliers is defined over an infinite Δt range so it makes sense to hold it contant while changing the Δt range for the fit. We are testing the sensitive of the physics results to the events at large Δt, not our ability to extract outlier parameters when the range of Δt is reduced.

The number of events gained or lost is very small around the cut at 18 ps. There are a total of 16 events lost when moving from $|\Delta t\ | <$18 ps to $|\Delta t\ | <$14 ps in the data sample, and 114 events for the very tight cut on $|\Delta t| < 10$ ps.

The mistag rates and resolution parameters are insensitive to the Δt selection criterion. The values of τ_{B^0} and Δm_d change very little compared to the statistical errors. The changes are consistent with 0.

	Full Δt fits (GExp+G) to Data			
	Reconstruction side		Tag side	
	RecB^0	Rec$\overline{B}^0(*)$	TagB^0	Tag$\overline{B}^0(\#)$
$\delta(\Delta m_d)$ (ps^{-1})	0.026 ± 0.018	-0.041 ± 0.018	-0.045 ± 0.010	0.001 ± 0.015
$\delta(\tau_{\mathrm{B}^0})$ (ps)	-0.047 ± 0.025	0.041 ± 0.023	-0.015 ± 0.023	0.019 ± 0.025
f^{out}	0.005 ± 0.002	0.001 ± 0.019	0.003 ± 0.002	0.001 ± 0.003
$\delta(\omega_{\mathrm{LT}})$	-0.011 ± 0.017	0.024 ± 0.012	-0.010 ± 0.011	0.028 ± 0.013
$\delta(\omega_{\mathrm{KT}}^{\mathrm{offset}})$	0.024 ± 0.031	0.001 ± 0.011	-0.001 ± 0.029	0.006 ± 0.029
$\delta(m_{\mathrm{KT}})$	-0.049 ± 0.046	0.043 ± 0.053	0.004 ± 0.043	-0.007 ± 0.042
$\delta(\omega_{\mathrm{N1}})$	-0.003 ± 0.023	0.014 ± 0.016	-0.015 ± 0.019	0.017 ± 0.021
$\delta(\omega_{\mathrm{N2}})$	-0.017 ± 0.022	0.027 ± 0.016	0.022 ± 0.020	-0.013 ± 0.015
$\delta(\omega_{\mathrm{N3}})$	-0.004 ± 0.012	0.011 ± 0.016	0.023 ± 0.011	-0.020 ± 0.012
	Others Parameters Suppressed			

Table 46: Fitting for τ_{B^0}, mistag rate and Δm_d broken down by reconstructed B^0,$\overline{\mathrm{B}^0}$ and tag side B^0,$\overline{\mathrm{B}^0}$. The fits use the full data sample and a GExp+G resolution function in the MasterModel. The values listed are the change in central value with respect to the nominal fit. The errors are the difference in quadrature between the subsample and full sample fit results, with the $\Delta\omega_{\mathrm{i}}$ parameters fixed to the nominal results, and therefore take into account the correlations due to the common events in both samples. The fit results from the Rec $\overline{B}^0(*)$ subsample did not provide a positive-definite covariance matrix (in part b/c the preferred outlier fraction is so small), but we belive the error estimates to be approximately correct. Additionally, as mentioned earlier, our ability to precisely extract the charged B fraction $f_{\mathrm{B}+}$ is limited. For the Tag $\overline{B}^0(\#)$ fit result, we were forced to fix $f_{\mathrm{B}+}$ to be constant and equal to the nominal fit result to guarantee convergence. The errors listed take this into account when comparing to the baseline result.

8.10 Varying $\sigma_{\Delta t}$ Criterion

The value of the $\sigma_{\Delta t}$ cut was determined in Sec. 3.4 based on the fact that resolution bias parameters scaled linearly with $\sigma_{\Delta t}$ up to a value of 1.8 ps for both resolution models (G+G+G and GExp+G). The events with a $\sigma_{\Delta t}$ larger than 1.8 ps represent less than ≈1.5% of the total events in our sample, plus they have the largest uncertainty in the measurement of Δt, so removing these events has a minimal impact on the statistical uncertainty of our measurements. The maximum value of $\sigma_{\Delta t}$ is 2.4 ps based on a cut made in DstarlnuUser ntuples. Table 51 shows how parameters change with different $\sigma_{\Delta t}$ cuts. Once again the values of τ_{B^0} and Δm_d are consistent across the different samples. Table 52 shows the change in final fit results for the tighter $\sigma_{\Delta t}$ cut of 1.4 ps. Recall that the outlier shape is fixed, although we float the overall fraction. As we select harder on $\sigma_{\Delta t}$, the overall population of outliers decreases, although the physics results are relatively

Difference from default fit — Signal Δt fits (GExp+G) to signal MC				
	Sample1	Sample2	Sample3	Sample4
$\delta(\Delta m_d)$ (ps^{-1})	0.008 ± 0.008	-0.006 ± 0.008	-0.011 ± 0.007	-0.002 ± 0.007
$\delta(\tau_{\mathrm{B}^0})$ (ps)	-0.011 ± 0.015	-0.008 ± 0.014	0.035 ± 0.016	-0.015 ± 0.015
$\delta(\omega_{\mathrm{LT}})$	-0.010 ± 0.006	-0.002 ± 0.006	0.009 ± 0.006	0.001 ± 0.006
$\delta(\omega_{\mathrm{KT}}^{\mathrm{offset}})$	0.014 ± 0.015	-0.004 ± 0.015	0.001 ± 0.015	-0.009 ± 0.015
$\delta(m_{\mathrm{KT}})$	-0.013 ± 0.022	0.0015 ± 0.021	-0.004 ± 0.021	0.012 ± 0.021
$\delta(\omega_{\mathrm{N1}})$	-0.001 ± 0.011	0.0085 ± 0.011	0.020 ± 0.011	-0.018 ± 0.010
$\delta(\omega_{\mathrm{N2}})$	-0.008 ± 0.010	-0.0085 ± 0.009	0.001 ± 0.009	0.006 ± 0.009
$\delta(\omega_{\mathrm{N3}})$	-0.011 ± 0.007	-0.0017 ± 0.007	0.020 ± 0.007	-0.006 ± 0.007
$\delta(\Delta\omega_{\mathrm{LT}})$	-0.008 ± 0.010	0.014 ± 0.010	-0.001 ± 0.010	-0.005 ± 0.010
$\delta(\Delta\omega_{\mathrm{KT}})$	-0.004 ± 0.008	-0.006 ± 0.008	0.006 ± 0.008	0.004 ± 0.008
$\delta(\Delta\omega_{\mathrm{N1}})$	0.005 ± 0.017	-0.015 ± 0.017	-0.011 ± 0.017	0.018 ± 0.016
$\delta(\Delta\omega_{\mathrm{N2}})$	-0.004 ± 0.015	0.000 ± 0.014	0.007 ± 0.015	-0.003 ± 0.014
$\delta(\Delta\omega_{\mathrm{N3}})$	0.004 ± 0.010	-0.014 ± 0.011	0.006 ± 0.011	0.003 ± 0.011
	Res parameters also very consistent			

Table 47: The signal MC sample split into four data size samples. After splitting the sample, only correctly reconstructed events are used in the four fits.

	Full Δt fits (GExp+G) to Data		
Parameter	$\mathrm{D}^0 \to \mathrm{K}\pi$	$\mathrm{D}^0 \to \mathrm{K}\pi\pi\pi$	$\mathrm{D}^0 \to \mathrm{K}\pi\pi^0/\mathrm{K_S}\pi\pi$
N_{events}	14523	33621	19760
$\delta(\Delta m_d)$	0.034 ± 0.024	-0.037 ± 0.026	0.019 ± 0.025
$\delta(\tau_{\mathrm{B}^0})$	-0.035 ± 0.027	0.017 ± 0.039	0.016 ± 0.032
$\delta(f^{\mathrm{out}})$	0.001 ± 0.002	0.000 ± 0.003	-0.002 ± 0.002
$\delta(f_{\mathrm{B}^\pm})$	0.043 ± 0.035	-0.022 ± 0.048	-0.025 ± 0.036

Table 48: Difference in fit results between MasterModel fits to the D^0 subsamples and the baseline fit result. Errors listed are the difference in quadrature. We observe no statistically significant difference between the combined D^0 fit to all Data and the separate subsample fit.

robust. We also did a full fit to Data with $\sigma_{\Delta t} < 1.4$ ps in which we fixed the outlier fraction of the signal model (f^{out}) to the value from a fit to Data with $\sigma_{\Delta t} < 1.8$ ps. In that case, the shift in the fitted value of τ_{B^0} is about twice that shown in Table 52, but is still not statistically significant. the one-sigma level.

Signal Δt (GExp+G) fits to signal MC				
$\|\Delta t\|$ cut (ps)	18 (ref)	$\delta(16)$	$\delta(14)$	$\delta(10)$
Δm_d (ps^{-1})	0.4715 ± 0.0043	0.00013 ± 0.0013	0.0001 ± 0.0001	0.0022 ± 0.0010
τ_{B^0} (ps)	1.5523 ± 0.0091	-0.0004 ± 0.0028	-0.0023 ± 0.0012	-0.0264 ± 0.0019

Table 49: The effect of changing the Δt cut on τ_{B^0}, and Δm_d. The fits use the GExp+G resolution model and correctly reconstructed events in our large signal MC sample compared to the reference selection of $|\Delta t| < 18ps$. The resolution and mistag parameters are consistent and so are not shown. Note,however, that we fix the outlier fraction to be constant in the subsample fits since thje ability to measure the outlier frwction degrades rapidly with tighter Δt cuts. The errors reported are the difference in quadrature between the baseline result with fixed signal outlier model and these fit results, thereby correctly accounting for the largely correlated event samples.

Full fits to Data, fixed outliers		
$\|\Delta t\|$ cut (ps)	10	14
$\delta(\Delta m_d)$ (ps^{-1})	-0.0102 ± 0.0188	0.0016 ± 0.0020
$\delta(\tau_{\mathrm{B}^0})$ (ps)	0.0381 ± 0.0467	-0.0023 ± 0.0047

Table 50: The effect of changing the Δt cut on τ_{B^0}, and Δm_d. The fits use the GExp+G resolution model on the full data sample, and since the physics parameters are blinded, we report the shift with respect to nominal fit result. The parameters of the signal outlier model (bias, width and fraction) are fixed in these fits.

Signal Δt fits (GExp+G) to signal MC					
$\sigma_{\Delta t}$ cut (ps)	1.4	1.6	1.8	2.0	2.2
$\delta(\Delta m_d)$ (ps^{-1})	0.0001 ± 0.0009	0.0000 ± 0.0009	0	0.0005 ± 0.0003	0.0001 ± 0.0008
$\delta(\tau_{\mathrm{B}^0})$ (ps)	-0.0001 ± 0.0023	0.0004 ± 0.0018	0	-0.0004 ± 0.0022	0.0013 ± 0.0029

Table 51: The effect of changing the $\sigma_{\Delta t}$ cut on τ_{B^0} and Δm_d. The fits use the GExp+G resolution model and correctly reconstructed events in the signal MC sample. Resolution and mistag parameters are consistent across each fit.

8.11 Fixing τ_{B^0} or Δm_d to PDG2000/2002 values

Table 53 compares the values of τ_{B^0}, Δm_d when they are both allowed to float and when one or the other is fixed to its corrsspnding PDG2000 value (also the value used in the Monte Carlo decay file). The fit is done using only correctly reconstructed events in the signal MC sample. The effect of fixing τ_{B^0} or Δm_d is very small. For the equivalent test

Full Δt fits (GExp+G) to Data	
$\sigma_{\Delta t}$ cut (ps)	1.4
$\delta(\Delta m_d)$ (ps^{-1})	0.0012 ± 0.0069
$\delta(\tau_{\mathrm{B}^0})$ (ps)	-0.0046 ± 0.0082
$\delta(f_{\mathrm{B}^\pm})$	-0.004 ± 0.007
$\delta(f^{\mathrm{out}})$	-0.0013 ± 0.0003

Table 52: The effect on τ_{B^0} and Δm_d of changing the $\sigma_{\Delta t}$ cut from the nominal 1.8 ps to 1.4 ps. The fits use the GExp+G resolution model for Data. Mistag parameters are consistent across each fit, but suppressed for clarity. The significant change in outlier population is plausible since the wayward outliers are likely removed by the tighter $\sigma_{\Delta t}$ cut, and thus the outlier population descreases. The central values reported are the change with respect to the baseline value, and the errors are the quadrature difference of statistical errors from the fits.

on Data, we mark the PDG2002 values on the correlation plot (Fig. 30).

8.12 Fitting with Different Resolution Models for signal MC

The fitted parameters for the three different resolution models are shown in Table 54. They all do a good job of determining Δm_d, the G+G model has more trouble with determining τ_{B^0}, but this is expected since the G+G model can't model the asymmetrical bias and outlier events simultaneously. The G+G+G and GExp+G models have more freedom to capture these features. The G+G+G and GExp+G are the only serious choices for resolution models for this measurement.

8.13 Mistag rate – $\sigma_{\Delta t}$ correlation

This section looks at how sensitive the values of Δm_d and τ_{B^0} are to using the wrong value for the kaon $\sigma_{\Delta t}$ slope. The reason for considering a non-zero slope was shown in the tagging chapter. The kaon mistag rate is the only mistag rate which shows a scaling behavior with $\sigma_{\Delta t}$. The value of the slope is varied from -2σ to $+2\sigma$. The results are shown in Table 55.

The value of $\omega_{\mathrm{KT}}^{\mathrm{offset}}$ changes as we would expect since the average Kaon mistag rate should stay roughly constant (Kaon mistag = $\omega_{\mathrm{KT}}^{\mathrm{offset}} + m_{\mathrm{KT}} \cdot \sigma_{\Delta t}$). All other parameters are stable and consistent with the standard fit parameters.

8.14 Varying $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ Upper Limit

The parameters for combintoric background are determined primarily by events in $m(\mathrm{D}^*)-m(\mathrm{D}^0)$ sideband. As discussed in App. E, the events in the sideband are pretty similar

Signal Δt (GExp+G) fits to signal MC			
	$\delta(fix\tau_{B^0})$	$\delta(fix\Delta m_d)$	fit τ_{B^0}, Δm_d
Δm_d (ps^{-1})	0.0007 ± 0.00034	0.472	0.4715 ± 0
τ_{B^0} (ps)	1.548	-0.0004 ± 0.0006	1.5523 ± 0.0091
Mistag Rates			
ω_{LT}	$-0.00016 \pm 1.0e-4$	$-9.4e-05 \pm 3.7e-5$	0.0561 ± 0.0035
ω_{KT}^{offset}	$-0.00021 \pm 1.9e-4$	$-4.2e-05 \pm 2.1e-4$	0.0672 ± 0.0084
m_{KT}	$8.0e-05 \pm 2.7e-4$	$-8.0e-05 \pm 3.1e-4$	0.1275 ± 0.0121
ω_{N1}	$-0.00015 \pm 1.8e-4$	$-9.0e-05 \pm 1.3e-4$	0.1840 ± 0.0061
ω_{N2}	$-6.0e-05 \pm 1.6e-4$	$-5.0e-05 \pm 1.4e-4$	0.3391 ± 0.0054
ω_{N3}	$0 \pm 1.2e-4$	$-1.0e-05 \pm 1.1e-4$	0.4617 ± 0.0040
Delta Mistag Rates			
$\Delta\omega_{LT}$	$-9.9e-06 \pm 6.8e-5$	$3.5e-06 \pm 9.0e-5$	0.0044 ± 0.0058
$\Delta\omega_{KT}$	$-4.0e-06 \pm 3.0e-5$	$-4.2e-05 \pm 3.0e-5$	-0.0140 ± 0.0045
$\Delta\omega_{N1}$	$5.0e-06 \pm 4.4e-5$	$-6.0e-06 \pm 6.3e-5$	0.0120 ± 0.0098
$\Delta\omega_{N2}$	$-4.6e-06 \pm 1.0e-7$	$1.6e-06 \pm 1.0e-7$	-0.0415 ± 0.0083
$\Delta\omega_{N3}$	$2.0e-07 \pm 1.0e-7$	$1.0e-07 \pm 1.0e-7$	-0.0742 ± 0.0061
Resolution Parameters			
κ	0.0053 ± 0.0023	0.0008 ± 0.0017	1.1598 ± 0.0680
s^1	0.0058 ± 0.0025	0.0011 ± 0.0030	1.0017 ± 0.0216
f_{LT}^1	0.00037 ± 0.0024	-0.00031 ± 0.0010	0.8412 ± 0.0230
f_{KT}^1	0.0012 ± 0.0029	$9.0e-05 \pm 0.0011$	0.6864 ± 0.0219
f_{N1}^1	$7e-05 \pm 0.0030$	-0.00026 ± 0.0012	0.8532 ± 0.0289
f_{N2}^1	0.00085 ± 0.0029	0.00018 ± 0.0010	0.7397 ± 0.0248
f_{N3}^1	0.00093 ± 0.0026	0.00026 ± 0.0009	0.7188 ± 0.0202
f^{out}	0.00031 ± 0.0001	$2.1e-05 \pm 7.7e-5$	0.0044 ± 0.0015
b^{out}	-	-	-1.68
s^{out}	-	-	5.44

Table 53: The effect of fixing τ_{B^0} and Δm_d to the generator values compared to the baseline result. These fits use all correctly reconstructed signal MC events. All fits use the GExp+G resolution model; the right most column is the "standard" fit to which all other fits are generally compared. The outlier width is fixed to a value of 5.44 ps for all fits, and the width to -1.68 ps. The errors are statistical, and represent the difference in quadrature between the nominal result w/fixed τ_{B^0} or Δm_d and the result from the corresponding fixed generator value fit.

to the events in the peak region. We cross check the potential difference of events in different region of $m(D^*) - m(D^0)$ space by varying the upper limit and repeat the fitting

Signal Δt fits to signal MC			
	GExp+G	G+G+G	G+G
Δm_d (ps^{-1})	0.4715 ± 0.0043	0.0032 ± 0.0014	0.0030 ± 0.0009
τ_{B^0} (ps)	1.5523 ± 0.0091	-0.0017 ± 0.0007	0.0050 ± 0.0025
ω_{LT}	0.0561 ± 0.0035	-0.0014 ± 0.0004	-0.0011 ± 0.0004
ω_{KT}^{offset}	0.0672 ± 0.0084	-0.0008 ± 0.0015	0.0015 ± 0.0016
m_{KT}	0.1275 ± 0.0121	0 ± 0.0025	-0.0027 ± 0.0025
ω_{N1}	0.1840 ± 0.0061	-0.0010 ± 0.0006	-0.0007 ± 0.0006
ω_{N2}	0.3392 ± 0.0054	-0.0005 ± 0.0005	-0.0005 ± 0.0004
ω_{N3}	0.4617 ± 0.0040	-0.0001 ± 0.0003	-0.0002 ± 0.0003

Table 54: Comparison of results with three different Δt resolution models: GExp+G, G+G+G and G+G. The fits used all correctly reconstructed signal MC events. The outlier width scale factor for the GExp+G fit was fixed at 5.44 ps, and the outlier bias set to -1.68 ps. In the G+G+G fit, the width of the outlier was fixed to 8 ps and the width of the wide Gaussian was set to 2 ps. (The fit had trouble with just the outlier width fixed.) All parameters in the G+G model were floating. The G+G+G and G+G fit results are shown with reference to the GExp+G baseline result. The errors listed are the difference in quadrature between the baseline and the new result, in an effort to isolate the uncorrelated error.

Difference from default fit — Signal Δt fits (GExp+G) to signal MC				
$m_{KT} =$	0.100	0.112	0.137	0.150
$\delta(\Delta m_d)$ (ps^{-1})	0.0003 ± 0.0043	0.0001 ± 0.0043	0.0000 ± 0.0043	0.0000 ± 0.0043
$\delta(\tau_{B^0})$ (ps)	0.0002 ± 0.0087	0.0001 ± 0.0087	0.0000 ± 0.0087	0.0000 ± 0.0087
$\delta(\omega_{KT}^{offset})$	0.017 ± 0.003	0.009 ± 0.003	-0.007 ± 0.003	-0.015 ± 0.003
	All Resolution & Mistag Parameters Consistent			

Table 55: Fitting for τ_{B^0}, mistag rate and Δm_d while changing the value of the dependence of the mistag rate on Δt error for the kaon tagging category (m_{KT} in units of ps^{-1}). The sample is all correctly reconstructed signal MC events.

procedure. The default fit uses $m(D^*) - m(D^0) < 165$ MeV. It is moved to 162.5, 160.0, 157.5, and 155.0 MeV. The lifetime changes by $+0.6$, -1.2, $+1.0$, $+0.4$ fs, and Δm_d by $+0.7$, $+0.5$, -0.5, -0.2×10^{-3} ps^{-1}, respectively. There is no systematic effect due to the choice of the sideband upper limit. (The change on estimated error is very small. Some have smaller error with less events. Therefore we do not attemp to calculate the error on the changes on τ_{B^0} and Δm_d. But since all the changes are much smaller than most systematic uncertainties, it is safe to discard them even if they were statistically significant.)

8.15 Summary

Figures 40–41 summarize the results of this set of cross-checks on the Monte Carlo signal sample. Figures 42–43 summarize the results of this set of cross-checks on the Data sample.

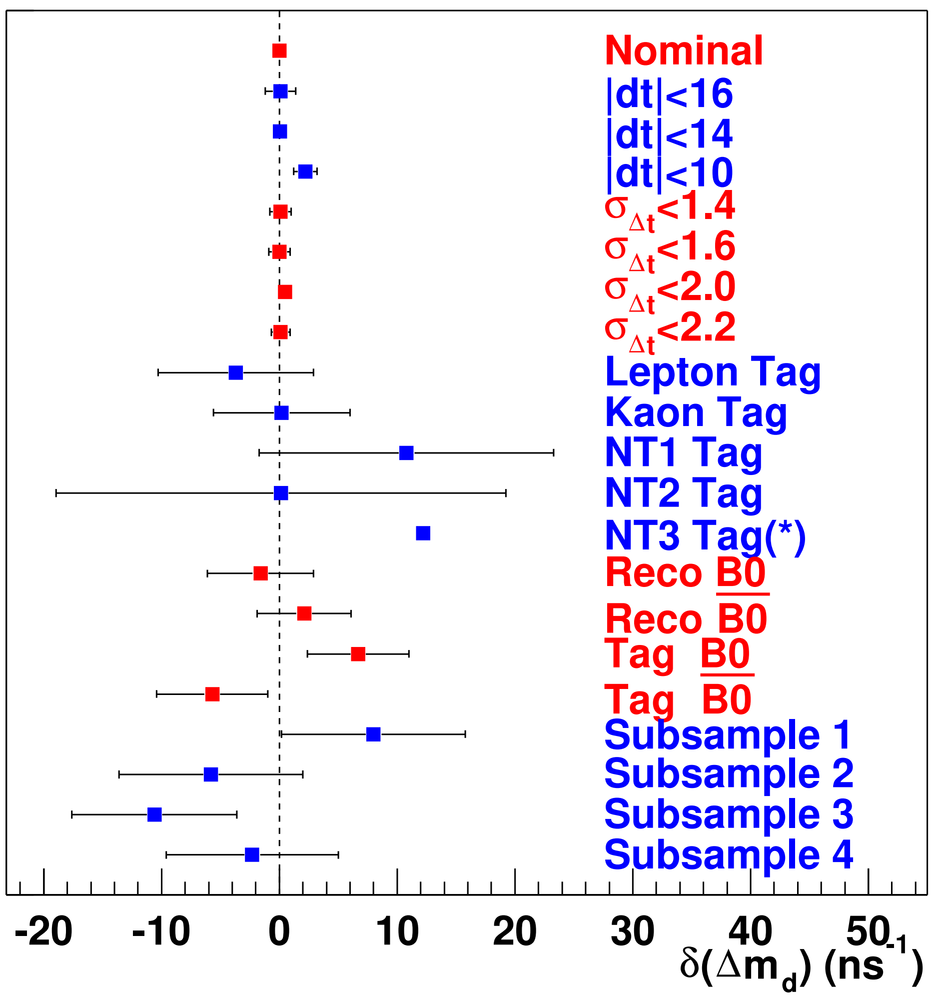

Figure 40: These plots show the Δm_d differences from the standard fit for various fits to the Monte Carlo signal sample. The units of the plots are ns^{-1}. The errors shown on the plots are the sample and subsample errors subtracted in quadrature as appropriate. The lines marked with (*) have had the very large erros suppressed.

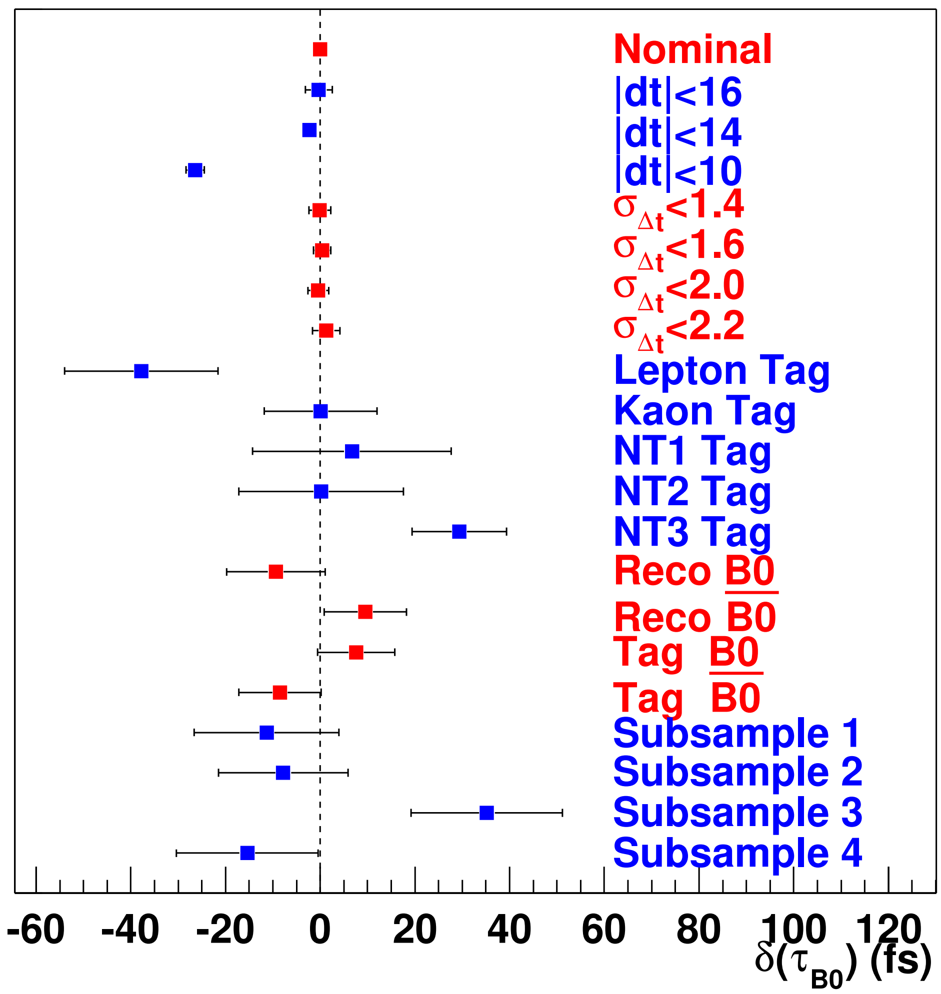

Figure 41: These plots show the τ_{B^0} differences from the standard fit for various fits to the Monte Carlo signal sample. The units of the plots are fs. The errors shown on the plots are the sample and subsample errors subtracted in quadrature.

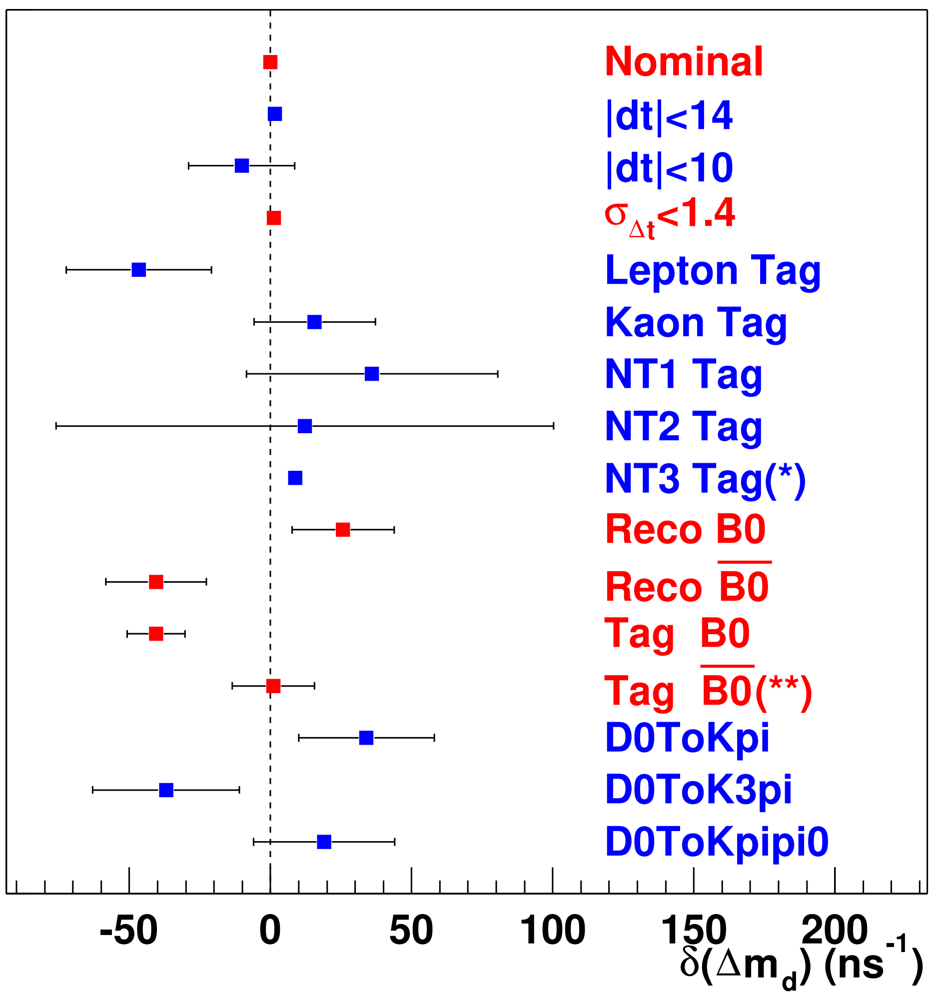

Figure 42: These plots show the Δm_d differences from the standard fit for various fits to the Data sample. The units of the plots are ns^{-1}. The errors shown on the plots are the sample and subsample errors subtracted in quadrature. The lines marked with an (*) indicate extremely large errors, which are suppressed. The line marked with (**) represents a fit that only converged when f_{B+} was fixed to the nominal value.

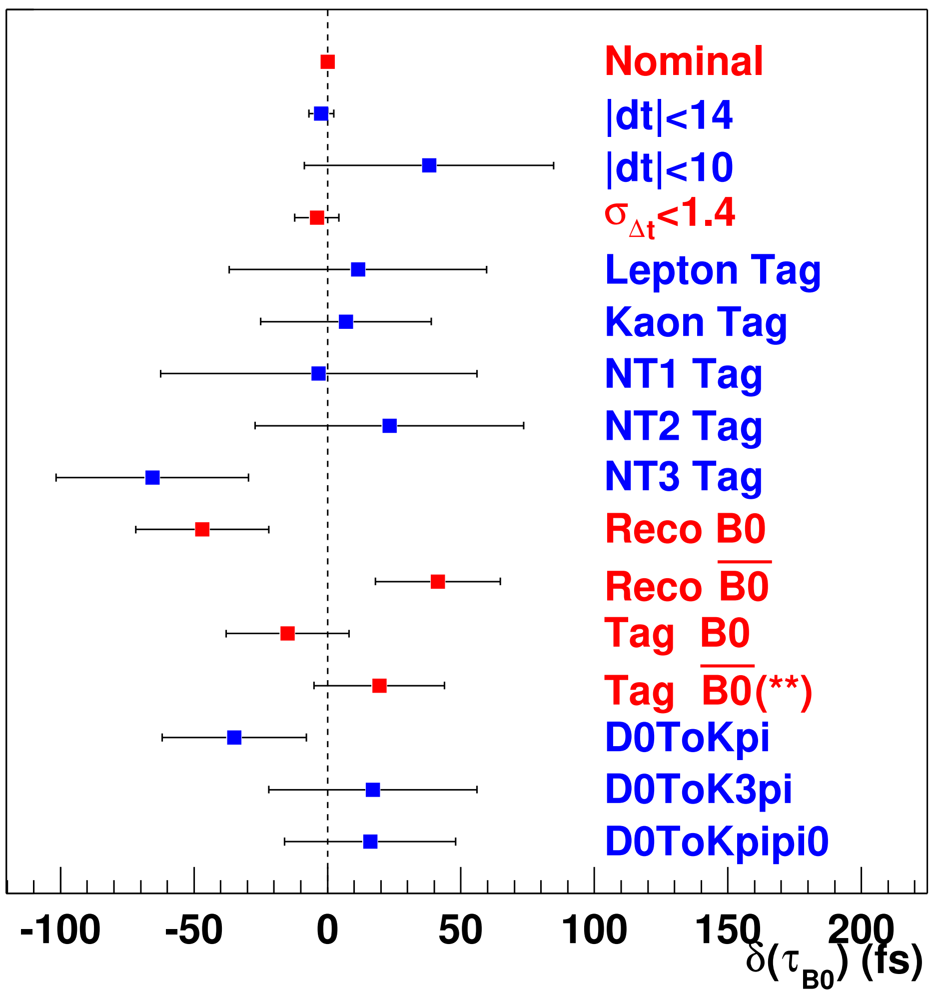

Figure 43: These plots show the τ_{B^0} differences from the standard fit for various fits to the Data sample. The units of the plots are fs. The errors shown on the plots are the sample and subsample errors subtracted in quadrature. The line marked with (**) represents a fit that only converged when f_{B^+} was fixed to the nominal value.

9 Systematic Studies

This section investigates the size and character of systematic errors in the measurement of τ_{B^0} and Δm_d. The studies make use of both Monte Carlo and real Data, and both the signal Δt model and full MasterModel (including background, $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ probabilities, etc).

9.1 Event Selection Bias and Monte Carlo Correction

We check the event selection bias by fitting to the truth Δt in signal Monte Carlo sample using theoretical distribution, as shown in Fig. 8 in Sec. 3.5. There is a slightly negative bias for both mixing and lifetime. And we also expect a small positive bias in lifetime due to ignoring resolution dependence on true Δt, as discussed in Sec. 3.5. However, the result from the fit to signal Monte Carlo events with default signal Δt model, as shown in Table 14 in Sec. 5.2.1 ($\Delta m_d = 0.4726 \pm 0.0045$ ps^{-1}, $\tau_{\mathrm{B}^0} = 1.5529 \pm 0.0090$ ps), does not show any significant deviation from the generated values ($\Delta m_d = 0.472$ ps^{-1}, $\tau_{\mathrm{B}^0} = 1.548$ ps).

However, the full fit to generic Monte Carlo does show a bias. As discussed in Sec. 7.2, we apply a correction of -0.022 ps for τ_{B^0} and -0.020 ps^{-1} for Δm_d and conservatively apply the full statistical error of the fit to signal+background, i.e., ± 0.018 ps for τ_{B^0} and ± 0.012 ps^{-1} for Δm_d.

This uncertainty has already covered the possible event selection bias and the small bias due to ignoring the B momentum at the $\Upsilon(4\mathrm{S})$ frame (see Sec. 3.4 for details).

9.2 Fixed Outlier Scale

As described in Sec. 6.1, we repeated the full fit with 36 different fixed values of the outlier bias and outlier width, covering a wide range for each parameter (-1 to -10 ps in bias, 4 to 11 ps in width). A scatterplot of the resulting values of Δm_d and τ_{B^0} is shown in Fig. 29. We assign half the full spread of lifetime and mixing results as the systematic uncertainty due to the outlier parameters. The resulting systematic uncertainties are ± 0.005 ps for τ_{B^0} and ± 0.001 ps^{-1} for Δm_d.

9.3 z Scale

The z scale systematic error is common to many analyses which measure the time difference between B^0 decays. The difference in the z position is used to determine Δt for an event, so a bias in the z scale will directly lead to a bias in the measured Δt. A conservative estimate of the z scale uncertainty from the fully reconstructed hadronic group is less than $\pm 0.4\%$ [9]. Our measurement of the time difference is very similar so we will adopt this estimate. The error on Δm_d is less than ± 0.002 ps^{-1} while the error on τ_{B^0} is ± 0.006 ps.

9.4 PEP-II Boost Uncertainty

Since Δt is directly proportional to the measured average PEP-II boost, the errors on τ_{B^0} and Δm_d are also directly related to the uncertainty on the boost. We use the standard 0.1 % prescription [20] for the systematic uncertainties, i.e., $\pm 0.0005\ \mathrm{ps}^{-1}$ for Δm_d and ± 0.0015 ps for τ_{B^0}.

9.5 B^+ Mistag Rate, Lifetime and Fraction

We fix the ratio of the $\mathrm{B}^+/\mathrm{B}^0$ lifetimes and mistag ratios in the final, full fit as explained previously. In this study, we conduct a short test to look for systematic changes in our fit results when we fix the ratios to different values using signal Monte Carlo, and the change in the results from the fit to Data when varying these values within their known errors.

For this study, we use the 20/fb sample of generic B^0 and B^+ Monte Carlo, select only the properly reconstructed signal candidates (including those from B^+ parents), and perform a fit using the signal Δt model (GExp+G resolution) with the mistag ratio fixed to three different values, and similarly for the lifetime ratio. The nominal values for these constants is shown in Tables 56-57, and we vary the parameters by one σ to measure the associated systematic error. We observe results as indicated in Table 58. (The values for Data are obtained from the BTagging folks.)

Nominal $\mathrm{B}^+/\mathrm{B}^0$ Ratios for MC	
$\tau_{\mathrm{B}^+}/\tau_{\mathrm{B}^0}$	1.0690 ± 0.029
$\omega_{\mathrm{LT}}^{\mathrm{B}^+}/\omega_{\mathrm{LT}}^{\mathrm{B}^0}$	1.210 ± 0.46
$\omega_{\mathrm{KT}}^{\mathrm{B}^+}/\omega_{\mathrm{KT}}^{\mathrm{B}^0}$	0.724 ± 0.12
$\omega_{\mathrm{N1}}^{\mathrm{B}^+}/\omega_{\mathrm{N1}}^{\mathrm{B}^0}$	0.740 ± 0.26
$\omega_{\mathrm{N2}}^{\mathrm{B}^+}/\omega_{\mathrm{N2}}^{\mathrm{B}^0}$	0.942 ± 0.17
$\omega_{\mathrm{N3}}^{\mathrm{B}^+}/\omega_{\mathrm{N3}}^{\mathrm{B}^0}$	1.121 ± 0.12

Table 56: Values for ratios of B^+ to B^0 parameters for use in the signal Δt model which describes candidates from both charged and neutral parents. The mistag rates were determined directly from Monte Carlo truth by counting.

In the MasterModel fit to Data, the B^+ to B^0 lifetime ratio is fixed. We vary this ratio up and down by one sigma according to PDG2002 lifetime ratio [17] and repeat the full fit. The mixing and lifetime results are shown in Fig. 59. We assign the systematic errors to be half of the observed difference.

The B^+ to B^0 mistag ratios for all five tagging categories are also fixed. We move all five mistag ratios up and down by one sigma according to Table 28 and repeat the full fit. The results are shown in Table 60. Again we assign half of the difference as the systematic due to fixing mistag ratios.

Nominal B^+/B^0 Ratios for Data	
τ_{B^+}/τ_{B^0}	1.083 ± 0.017
$\omega_{LT}^{B^+}/\omega_{LT}^{B^0}$	0.54 ± 0.10
$\omega_{KT}^{B^+}/\omega_{KT}^{B^0}$	0.68 ± 0.05
$\omega_{N1}^{B^+}/\omega_{N1}^{B^0}$	0.99 ± 0.12
$\omega_{N2}^{B^+}/\omega_{N2}^{B^0}$	1.05 ± 0.07
$\omega_{N3}^{B^+}/\omega_{N3}^{B^0}$	1.12 ± 0.12

Table 57: Values for ratios of B^+ to B^0 parameters for use in the signal Δt model which describes candidates from both charged and neutral parents. These values for performance on Run-1 data were obtained from BAD199, and BTaggingTools studies, except NT3 category, which are taken from Monte Carlo truth counting.

Signal Δt fits to generic MC				
Parameter	$(\tau_{B^+}/\tau_{B^0}) + 1\sigma$	$(\tau_{B^+}/\tau_{B^0}) - 1\sigma$	$(\omega_{B^+}/\omega_{B^0}) + 1\sigma$	$(\omega_{B^+}/\omega_{B^0}) - 1\sigma$
$\delta(\Delta m_d)$ (ps^{-1})	-0.0004	0.0004	0.0007	-0.0005
$\delta(\tau_{B^0})$ (ps)	-0.0007	0.0008	0.0001	-0.0001

Table 58: Fitted results from generic MC for the two physics parameters for diufferent values of the fixed charged B properties.

Full Δt fits to Data			
Parameter	$(\tau_{B^+}/\tau_{B^0}) + 1\sigma$	$(\tau_{B^+}/\tau_{B^0}) - 1\sigma$	diff./2
$\delta(\Delta m_d)$ (ps^{-1})	-0.00002	0.0006	0.0003
$\delta(\tau_{B^0})$ (ps)	-0.0016	0.0021	0.0019
$\delta(f_{B^+})$	-0.0025	0.0029	

Table 59: Fitted results from Data for the two physics parameters for different values of the fixed B^+ to B^0 lifetime ratio. All other fit parameters were nearly identical.

9.6 SVT Alignment

This analysis relies on the precision vertexing possible with the combined reoslution of the SVT and DCH. However, the technique is most sensitive to internal misalignments of the SVT, i.e. to the relative positions of the wafers and strips which report individual hits per track. To estimate this effect, we study the fit results using two different alignment scenarios with signal Monte Carlo, and then consider the consistency of two large subsamples of the Run-1 Data using different SVT alignment sets "D" and "E".

Full Δt fits to Data			
	$(\omega_{B^+}/\omega_{B^0}) + 1\sigma$	$(\omega_{B^+}/\omega_{B^0}) - 1\sigma$	diff./2
$\delta(\Delta m_d)$ (ps^{-1})	-0.00007	-0.00023	0.00008
$\delta(\tau_{B^0})$ (ps)	-0.0002	0.0004	0.0003
$\delta(f_{B^+})$	-0.0020	-0.0026	

Table 60: Fitted results from Data for the two physics parameters for diufferent values of the fixed B^+ to B^0 mistag rate ratio.

9.6.1 Estimate of bias using signal MC

We form a signal Monte Carlo cocktail of 20 fb^{-1} by combining SP3 simulted events (in luminosity proportion) for the four different D^0 decay modes for the three largest months of Conditions, Jul2000, Aug2000, and Oct2000 (these three months dominate the luminosity weight of the data sample as well). The cocktail created with this recipe has an approximate composition as shown in Table 61.

D^0 Mode	Evts in 20/fb
D0ToKpi	52k
D0ToK3i	102k
D0ToKpipi0	188k
D0ToKspipi	36k

Table 61: Numbers of events per D^0 decay mode in the signal Monte Carlo cocktail with luminosity weight totalling 20^{-1}.

An SVT alignment scenario consists of a set of transformations that are applied to the set of individual wafers and modules. The alignment scenario can be used to describe the difference between one state of SVT internal alignment and another. So-called "misalignment" occurs when the particle hits and tracks are reconstructed using incorrect knowledge of the relative positions of the sensors. "Perfect" alignment is therefore the default case, where tracks are reconstructed from hits with completely accurate knowledge of their relative positions.

The SP3 Monte Carlo default scenario uses so-called "perfect" alignment, i.e. it uses the same alignment scenario for event generation and event reconstruction. In the real detector, however, we do not have the luxury of perfect knowledge of the sensors' locations when the event was "generated," or rather, detected. Through a variety of sophisticated fittings algorithms using dimuon events, we derive an alignment scenario for the real detector which describes the "true" positions of the sensors with respect to the nominal positions. During Run-1 and Run-2 data-taking, the SVT internal alignment algorithm was substantially improvied, leading to several different procedures for constructing alignment scenarios and some data was processed with each of these alignment scenarios, or "sets."

To estimate the effect of the systematic uncertainties inherent in the estimated positions of the SVT wafers in the real detector, we "misalign" the signal Monte Carlo cocktail (described above) to match the alignment scenarios used in the real data processing and reconstruction. We then look for biases in the parameter values after the full signal fit to the "misaligned" signal Monte Carlo. The key point is that the data sample is reconstructed with imperfect knowledge of the SVT's internal alignment; to mimic the effect of that uncertainty, we perturb the SVT in signal Monte Carlo and the reconstruct the data.

To achieve this in practice, we follow a recipe suggested by the Tracking Group. We assume that alignment set "L" is perfect for Run-1 data, which we then associate with the default perfect alignment of SP3 Monte Carlo. We then assume the alignment sets used in Release-8 processing (28% set "D" and 56% set "E") are additively related to this near-perfect "L" alignment. We derive the alignment transformations "diffDL" and "diffEL," which we claim are reasonable descriptions of the changes in internal SVT alignment (with respect to nominal) necessary to emulate performance under alignment sets "D" and "E". (i.e. We degrade the internal alignment with respect to the alignment used for event generation to match the scenario used for reconstruction. The degradation is chosen to be consistent with the two different misalignments used for data-taking/processing.) We then apply these alignment transformations to the raw Monte Carlo data, and rerun the entire event processing chain from Bear (reconstruction) to ascii file production (input to the signal fits). We report here on a study of three different datasets, each corresponding to reconstruction, event selection, and signal Δt fitting under a different alignment scenario for the 20 fb^{-1} signal Monte Carlo cocktail:

- Default (nominal SP3),
- diffEL,
- diffDL.

In the full signal fit, we use Monte Carlo truth information to determine the mixing status of each candidate. This improves the statistics used in the study, and focuses our attention on the changes in Δt due to the degraded alignment scenarios. (This also eliminates changes in the fitted physics parameters due to correlations between reconstructed flavor and vertexing. All three datasets were fitted successfully with the full signal model, and yield consistent, less precise conclusions.)

By comparing the best-fit values for τ_{B^0} and Δm_d between the nominal "perfect" alignement and the degraded models of real detector alignment, we estimate the uncertainty in the result from data due to the uncertainty about the true internal alignment of the SVT. Table 62 summarizes the change in physics and resolution parameters in the two (mis)alignment scenarios diffDL and diffEL with respect to the default case.

	Effect of misalignment on signal Δt fits to MC		
Parameter	Nominal Value	Δ diffDL	Δ diffEL
Δm_d	0.477 ± 0.00516	-0.00316	-0.00292
τ_{B^0}	1.53 ± 0.0155	0.0057	0.0055
κ	0.812 ± 0.131	0.0452	0.00591
s^1	0.969 ± 0.0423	0.0569	0.108
f^1_{KT}	0.507 ± 0.0767	-0.0213	0.0159
f^1_{LT}	0.687 ± 0.0812	-0.0254	0.00904
f^1_{N1}	0.831 ± 0.0839	-0.0997	0.0589
f^1_{N2}	0.555 ± 0.084	0.0059	-0.0105
f^1_{N3}	0.566 ± 0.069	-0.05	0.00131
b^{out}	-0.440 ± 0.551	0.112	0.368
f^{out}	0.0291 ± 0.00498	0.000435	0.00319

Table 62: Parameter values and their change from nominal values for a full signal Δt fit (with Monte Carlo truth tagging information) to the signal Monte Carlo cocktail in different SVT alignment scenarios: the nominal scenario typical of SP3, misalignment scenario diffDL, and misalignment scenario diffEL.

We observe that the central value of Δm_d decreases in both misalignment scenarios, and that τ_{B^0} increases. In the case of Δm_d, the change is nearly comparable to the statistical precision of the nominal fit, indicating that misalignment is probably an important systematic uncertainty. (Note: The statistical error on Δm_d increases from ± 0.00516 in the perfect alignment case to ± 0.00530 in the misalignment scenario. This might suggest that the misalignment itself contributed ± 0.001 to the error on Δm_d, while the resolution model abasorbed other effects.)

Based on these results, we assign the following systematic uncertainty to our measurements of Δm_d and τ_{B^0}:

- $\delta(\tau_{B^0}) = -0.0056$ ps
- $\delta(\Delta m_d) = -0.0030$ ps^{-1}

With the understanding that the alignment transformations diffDL and diffEL distort the SVT alignment from nominal (perfect) to imperfect, we recognize that our estimate of Δm_d (τ_{B^0}) from Run-1 data (using the imperfect alignment scenarios setD and setE) may be biased low (high).

We might consider applying a correction to our final results, but retain the magnitude of the correction as an estimate of the systematic error. This

	Full Δt fits to Data	
Parameter	Set "D"	Set "E"
$\delta(\Delta m_d)$	-0.0202 ± 0.0449	-0.0001 ± 0.0308
$\delta(\tau_{B^0})$	-0.0342 ± 0.0434	-0.0083 ± 0.0227
$\delta(f^{\rm out})$	0.0034 ± 0.0035	-0.000088 ± 0.0024
$\delta(\kappa)$	-0.254 ± 0.442	-0.133 ± 0.304

Table 63: Results of the full fit to subsamples of the Data, as grouped by SVT alignment set (here, sets "D" and "E"). We show the difference in central values as compared to the baseline fit, but quote the statistical error on the subsample fit result, as we wish to compare the results from Sets D and E with each other.

decision will be made after we better determine the approximate size of the overall error, and after consultation with the review committee.

9.6.2 Consistency in Data between different alignments

For this study, we separate the Run-1 Data sample into three subsamples according to the SVT alignment scenario used in event processing and reconstruction: set "D" (29%), set "E" (53%), and "other" (18%). (We perform the grouping by matching event run numbers and conditions information.) We fit the full model to each of the "D" and "E" subsamples, independently, including the background $m(D^*) - m(D^0)$ analysis for evaluation of background composition/fractions. Although the statistical power of the test is poor, we find results as shown in Table 63. The fitted central values for the key parameters are statistically consistent between these two subsets of the Data.

9.7 Beam Spot Position

The beamspot is used as a constraint in both the tag-side and reconstructed-side vertices. We investigate the dependence of the fitted values of τ and Δm_d on the assumed beamspot position by using the entire signal Monte Carlo and varying the beamspot position in y direction, where it most important.

We use beamspot information returned from the **eventInfo** $\rightarrow$ **beamSpotBFlight()** method, which incorporates a 25 μm width due to the flight length of the B candidate. This 25 μm is combined with the roughly 10 μm intrinsice width of the beamspot in the y-direction to form an effective spread of 30 μm. For this systematic study we examine two scenarios: one where we systematically shift the beamspot position (via the **BtaLoad-BeamSpot::offsetY** parameter), and the other where we smear the position randomly according to a gaussian of fixed width (via the **BtaLoadBeamSpot::errorYPos** tcl parameter). Tables 64- 65 describe the effect we observe in the final parameter values after performing the full fit (including resolution functions) on the signal sample. For these results, we use the GExp resolution model exclusively. (Note: For this study to be generally valid, we assume that detector effects and event selection are symmetric in the y-direction.)

Change in fit results to signal MC w/ beamspot variations					
		Beamspot Variation [μm]			
Parameter	nominal err	shift 10	shift 30	shift 80	shift/smear 20
Δm_d	± 0.00435	-0.00054	-0.00129	-0.00202	0.00008
τ_{B^0}	± 0.00878	-0.0042	-0.0001	0.0012	-0.0024
mistag rates	All values very consistent				
κ	± 0.0847	-0.004	-0.0125	-0.0853	0.0307
f^1_{KT}	± 0.0304	-0.00371	-0.0106	-0.0214	0.0149
f^1_{LT}	± 0.0278	0.00203	0.00343	0.00606	0.0167
f^1_{N1}	± 0.034	0.00443	0.00566	0.0005	0.0224
f^1_{N2}	± 0.0312	0.00372	-0.00967	-0.0162	0.0169
f^1_{N3}	± 0.0271	-0.00274	-0.00909	-0.0177	0.00023
b^{out}	± 0.699	-0.124	-0.476	-0.821	-0.269
f^{out}	± 0.00318	-0.000225	-0.00101	0.000076	-0.000683
s^1	± 0.0227	0.0159	0.019	0.0997	0.0203

Table 64: Difference in final parameter values from the full signal fit to signal Monte Carlo samples for different variations applied to the beamspot position, as compared to a fit to the nominal sample. Here we shift the beamspot position systematically before applying reco- and tag-side vertexing. (The first column of numbers lists the statistical errors on the fitted parameters for the 20^{-1} nominal signal Monte Carlo sample.

We note that in particular, the lifetime τ_{B^0} is very robust with respect to movements of the beamspot position. The resolution function model adapts easily even to the most significant smearing and offset of the beamspot position. The mixing frequency, $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ appears more sensitive, and we note that the measured dilution parameters change slightly as we systematically disturb the beamspot. To investigate this further, we perform another series of singal Δt fits to the different Monte Carlo samples where we use Monte Carlo truth information to determine flavor tagging status. This increases our statistical precision by nearly a factor of two , and allows us to write another (abbreviated) table, Table 66.

By studying the z-residuals of each vertex (reco and tag) separately, we also find that this uncertainty in the beamspot constraint affects each vertex about equally. From figures similar to Fig. 2 for each of the transformed-beamspot datasets, we write into Table 67, which records the effect of the beamspot transformation on the z-residual distribution's width and bias.

Change in fit results to signal MC w/ beamspot variations				
	Beamspot Variation [μm]			
Parameter	smear 20	smear 40	smear 80	shift/smear 80
Δm_d	0.00053	-0.00022	-0.00295	-0.00577
τ_{B^0}	-0.005	-0.0036	-0.0029	0.0024
mistag rates	All values consistent.			
f^1_{KT}	0.00685	0.0104	-0.0126	-0.0271
f^1_{LT}	0.00078	0.00036	-0.0197	-0.0108
f^1_{N1}	0.0158	0.00839	0.0124	-0.00667
f^1_{N2}	-0.00742	0.0107	-0.0265	-0.0328
f^1_{N3}	0.00065	0.00945	-0.0107	-0.0189
b^{out}	0.383	0.303	0.0398	-0.461
f^{out}	-0.000712	-0.000468	0.000463	0.00187
s^1	0.0248	0.0314	0.0947	0.146

Table 65: Difference in final parameter values from the full signal fit to the signal Monte Carlo samples for different variations applied to the beamspot position. Here we smear the beamspot position randomly by sampling from a gaussian of the indicated width, before applying reco- and tag-side vertexing.

	Fits to signal MC using truth-tagging					
Parameter	Nominal Value	shift 10	shift80	smear 20	smear 80	shift/smear 80
Δm_d	0.472 ± 0.00216	0.00022	-0.00106	0.00008	-0.00026	-0.00262
τ_{B^0}	1.54 ± 0.00686	-0.0017	0.0045	-0.0022	-0.0012	0.0053
κ	0.928 ± 0.0668	-0.0266	-0.0482	-0.00771	-0.00666	-0.0341
s^1	1.02 ± 0.0182	0.0117	0.085	0.0177	0.0874	0.14
b^{out}	-1.48 ± 0.357	-0.151	-0.405	0.192	-0.0836	-0.52

Table 66: Fit results for select parameters from signal Δt fits to Monte Carlo samples with a shifted and/or smeared beamspot constraint. As described in the text, Monte Carlo truth information was used to determine the flavor status of each candidates, thereby improving the statistical precision on the other physics parameters. The first column of numbers shows the fit results to the nominal singal Monte Carlo sample and their statistical error, the other columns show the change in central value.

Based on these observations, and knowing that (a) the difference in estimated beamspot position between different algorithm is than 10 μm and (b) that average resolution of the beamspot is about 30 μm, we estimate the systematic error due to uncertainties in the beamspot position to be

- $\delta(\tau_{B^0}) = \pm 0.005$ ps

Charactersitcs of z-vtx residuals from signal MC					
Variation	Size	z_{reco} Bias	z_{reco} RMS	z_{tag} Bias	z_{tag} RMS
nominal	0	0.05 ± 0.3	77.5 ± 0.2	34.4 ± 0.5	156.1 ± 0.4
shift	10	0.07	77.4	34.6	156.8
shift	80	1.92	85.5	36.7	162.2
smear	20	0.09	77.9	34.1	156.8
smear	80	0.03	85.0	34.8	161.6

Table 67: Shape parameters for tag- and reco- vertex z-residual distributions of correctly reconstructed signal Monte Carlo under several different beamspot transformation scenarios. [All units are μm.] As expected, systematic offsers of the beamspot increase the bias and RMS of the z-reisdual for each vertex separately, while smearing only degrades the rms of the residuals.

- $\delta(\Delta m_d) = \pm 0.001$ ps^{-1}

9.8 Background Fractions

We report here on a study which examines the sensitivity of τ_{B^0} and Δm_d to our uncertainty in the knowledge of the per-event signal probability, or more precisely, the per-event signal and four background probabilities (combinatoric, peaking fakes, peaking uncorrelated leptons, and peaking continuum). Recall that these per-event probabilities are computed using the $m(D^*) - m(D^0)$ value and mass-difference spectrum fits in a separate, independent stage. Essentially, we are addressing the question of "How do we account for the fact that our per-event probability estimates are based on fit results (the $m(D^*) - m(D^0)$ spectrum fits) that have statistical errors?" Because of the sophistication of the $m(D^*) - m(D^0)$ fits, as described in detail in [1], we use a novel technique to properly understand the dependence.

The per-event probabilities use two key ingredients: fit results that describe the peak and sideband lineshapes for $m(D^*) - m(D^0)$ in 360 sub-categories of the data, and background fractions which essentially count the fraction of peak and sideband events in each of the sub-categories (using the lineshape fit results). The fits and subsequent computations that produce this set of values (which are held fixed in the full Δt fit because we expect Δt and $m(D^*) - m(D^0)$ to be independent variables) introduce many correlations between them (due to the simultaneous nature of the result). To understand the effect of the statistical error on the calculations needed for per-event probabilities, we therefore perturb this set of values jointly and randomly, taking into account the individual errors on each result as well as the correlation between any two results (via the covariance matrix). We then have a "new" set of values from which to compute per-event probabilities in the full Δt fit.

We examine the distribution of fitted τ_{B^0} and Δm_d for 111 different perturbations of the parameters used to compute the per-event probabilities. In order to optimize our use of computing resources, (and possible be conservative on our estimate of the error), we fix the other 70 parameters in the full fit to the baseline values. These 2-parameter (τ_{B^0} and Δm_d) run quickly, and allows us to gather sufficient statistics for a conclusion. We also ran full Δt fits to the Data for 20 different random perturbations to verify that the 2-parameter fit results were not an inappropriate estimate. Our results are shown in Fig. 44 and Tab. 68. When floating only the two physics parameters, we achieve a statistical error from the fit of ± 0.012 ps^{-1} on Δm_d and ± 0.017 ps on τ_{B^0}.

In this study, we performed two separate perturbations to affect the per-event probability calculations (varying the $m(D^*) - m(D^0)$ shape fit values, and varying the yields fit results used to compute the background fractions). We also treated each perturbation separately, and found that for the 111 fits floating only the 2-physics parameters, the dominant source of spread in the fitted values was due to the perturbation of the background fractions. For instance, the rms of the distribution of the fitted Δm_d values from just the shape-fit perturbation was small, about 0.00055 ps^{-1} whereas the rms of the Δm_d values fitted using just the perturbed background fractions was 0.0033 ps^{-1}.

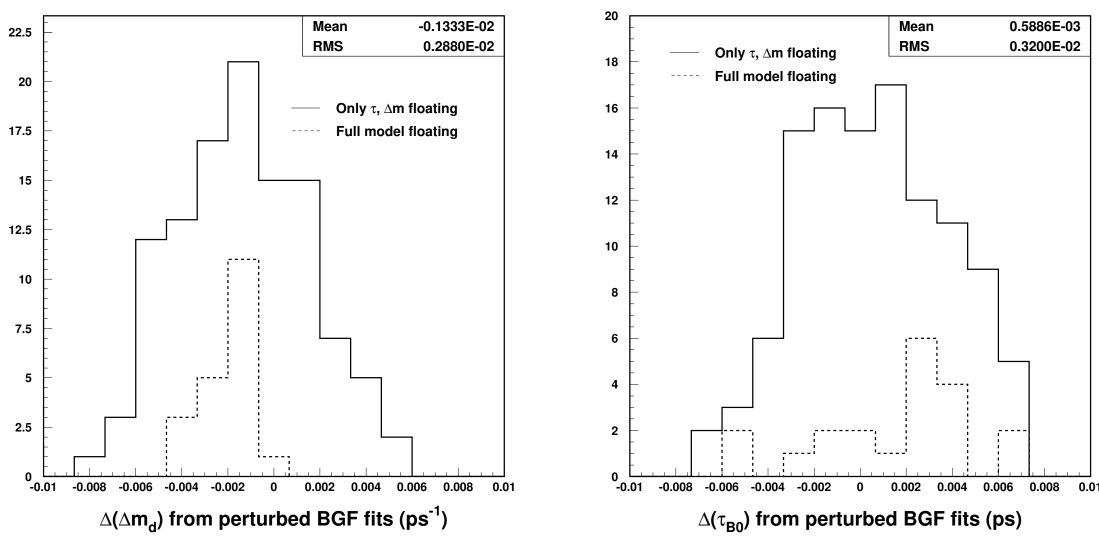

Figure 44: Histograms of the spread in fitted results (wrt the baseline blinded value) from Data using 111 different statistical variations of the input $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ parameters and floating only τ_{B^0} and Δm_d in the final Δt fits. The superimposed shaded histograms show the distribution of fit results for the 20 fits where the entire model was floating in the fit.

Full fits to Data, floating only Δm_d, τ_{B^0}		
	$\delta(\Delta m_d)$ (ps^{-1})	$\delta(\tau_{\mathrm{B}^0})$ (ps)
Mean	-0.0013 ± 0.0003	-0.0006 ± 0.0003
RMS	0.0029 ± 0.0002	0.0032 ± 0.0002

Table 68: Characteristics of the distribution of fitted results for Δm_d and τ_{B^0} (with respect to the baseline) from the 111 fits with perturbed background fraction when floating only the two physics parameters.

(Likewise for τ_{B^0}, the numbers were 0.0015 ps and 0.0034 ps, respectively.) This suggests that the statistical uncertainty on the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ shape fits is negligible compared to the statistical uncertainty on the background fractions due to statistical fluctuations of the Data populations in each of the 360 categories.

parameters in a full fit to Data. Although the precision is poor, the rms of the distributions of Δm_d and τ_{B^0} values are consistent with the 111 two-parameter fits, and the bias with respect to the baseline fit is very small. The fitted Δm_d values from the 20 different fits using perturbed background parameters yield an average shift (wrt the nominal fit) of -0.002 with an rms of 0.0012 ps^{-1}. Likewise for τ_{B^0}, the distribution of fitted values have an average of 0.0016 and an rms of 0.0033 ps. (The error on these rms values is roughly 20%.) This increases our confidence in assigning the systematic errors as listed at the end of this section.

Finally, we examine the correlation between τ_{B^0} and Δm_d as a function of the background fraction perturbations. Fig. 45 shows the empirical lack of correlation between τ_{B^0} and Δm_d from the 111 different perturbed background fraction results. Using the 20 full fits, however, we can examine the measured correlation coefficient between the two physics parameters. Unfortunately, only 10 of the fits yielded a good measure of the error matrix. From these fits, we find that the average correlation between τ_{B^0} and Δm_d is -0.213 ± 0.003 with an rms of 0.011.

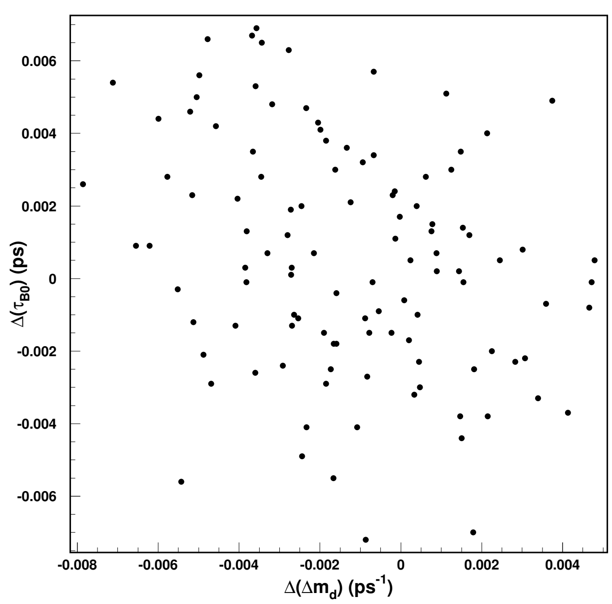

Figure 45: Scatter plot of pairs of (τ_{B^0},Δm_d) values (wrt baseline fit) from the 111 2-parameter floating Master fits to the Data, each using a different set of perturbed background fractions in the fit. We observe no induced correlation on τ_{B^0} and Δm_d due to varying the $m(D^*) - m(D^0)$ fit results and background fractions.

Based on these results, and recognizing that ~100 tests give us a fractional precision of order 7% on the rms of the distribution, we assign the following systematic errors:

- $\delta(\Delta m_d) = \pm 0.0029$ ps^{-1}
- $\delta(\tau_{B^0}) = \pm 0.0032$ ps.

9.9 Fitting with Different Resolution Model

We fit to the data using an alternative resolution funciton for signal Δt model, a triple-gaussian (G+G+G). The outlier gaussian shape is kept the same, i.e., $\sigma_{\rm out} = 6$ ps and $\mathrm{b}_{\rm out} = -5$ ps. The biases and widths of the other gaussians are scaled by $\sigma_{\Delta t}$. The number of free parameters is 74, two more than that of default fit.

In comparison with the default fit, the lifetime shifts by $+0.0034$ ps and Δm_d shifts by -0.0009 ps^{-1}. The minimized negative log-likelihood is 795904.4, (for reference, the value for the default fit is 795903.8.).

9.10 Background Δt Models

Although we end up floating most of the parameters in the Δt distributions for the backgrounds, we could have a bias in our result due to the choice of models. Since the largest background contribution by far is from combinatotic events, we focus on that Δt model only. An important feature of the Δt model we've selected is its effort to accomodate the mixing structure of the combinatoric background. As a suitable variation, we consider modelling the combinatoric background Δt and flavor distribution with a pure lifetime-based model. Comparing the results from a full fit to data gives us an estimate of our systematc error due to the choice of model for the largest background.

The fit result with this combinatoric model shows that the lifetime shifts by $+0.0063$ ps and Δm_d shifts by -0.0012 ps^{-1}. It is worth noticing that the The minimized negative log-likelihood is 795996.3, which is much larger than the default fit value, 795903.8. The charged B fraction, $f_{\mathrm{B}^\pm}$, also changes from 8% to 6%. Although this is an extreme test for exposing our sensitivity to assumptions about the functional form of the background Δt dsitributions, we use the size of this shift as an estimate of our systematic uncertainty.

9.11 Dependence on Known PID Efficiencies

Here we show that our sensitivity to knowledge of the BABAR PID efficiencies (via the background fractions) is less than the statistical error on the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ fit results, and is therefore negligible.

This requires a bit of technical bookkeeping, but is expected to be entirely negligible. If the review committee insists, we'll undertake it.

9.12 Summary

We summarize the results of this section in Table 69, which lists each source of systematic uncertainty and our evaluation of its contribution to the error on τ_{B^0} and Δm_d, and the sum in quadrature.

Source	$\delta(\Delta m_d)$ (ps^{-1})	$\delta(\tau_{\mathrm{B}^0})$ (ps)	Note
Selection and fit correction	± 0.0123	± 0.0178	Stat error on generic MC fit
z scale	± 0.0020	± 0.0060	0.4% recipe
PEP-II boost	± 0.0005	± 0.0015	0.1% recipe
SVT alignment	± 0.0030	± 0.0056	diffDL, diffEL recipes with sig MC
Beamspot position	± 0.0010	± 0.0050	Move/smear BS in sig MC
Bkg / signal prob.	± 0.0029	± 0.0032	Vary $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ fits
Bkg Δt models	± 0.0012	± 0.0063	Vary comb. bkg model
Fixed $\mathrm{B}^+/\mathrm{B}^0$ lifetime ratio	∓ 0.0003	± 0.0019	Vary lifetime ratio by $\pm 1\sigma$ (PDG2002)
Fixed $\mathrm{B}^+/\mathrm{B}^0$ mistag ratio	∓ 0.0001	∓ 0.0003	Vary mistag ratios by $\pm 1\sigma$ (BAD119)
Fixed signal outlier shape	± 0.0010	± 0.0054	Vary outlier width & bias
Signal resolution model	± 0.0009	± 0.0034	G+G+G vs GExp+G
Total syst. error	± 0.013	± 0.022	

Table 69: Summary of systematic uncertainties on the two physics parameters, τ_{B^0} and Δm_d.

10 Closure

On the advice of the review committee, and with the full powers and privilege vested within us by the BABAR collaboration, we unblinded the fit results on 05 July, 2002. On the Run-1 Data sample using Release 8 processing, we measure the following properties of B^0 time evolution with a sample of approximately 14,000 exclusively reconstructed $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ signal events,

- $\Delta m_d = 0.492 \pm 0.018 \pm 0.013\ \mathrm{ps}^{-1}$,
- $\tau_{B^0} = 1.523\ ^{+0.024}_{-0.023} \pm 0.022$ ps
- statistical correlation $\rho(\Delta m_d, \tau_{B^0}) = -0.22$.

We compare our measurement with the world average and other experiments. The world averages for both lifetime and mixing are dominated by recent measurements in B-Factories (BABAR and Belle). Therefore we do not show the comparison with earlier experiments. The comparisons are shown in Fig. 46. Our measurements are consistent with the world average and other experiments, and the uncertanities are still dominated be statistical errors and Monte Carlo statistics.

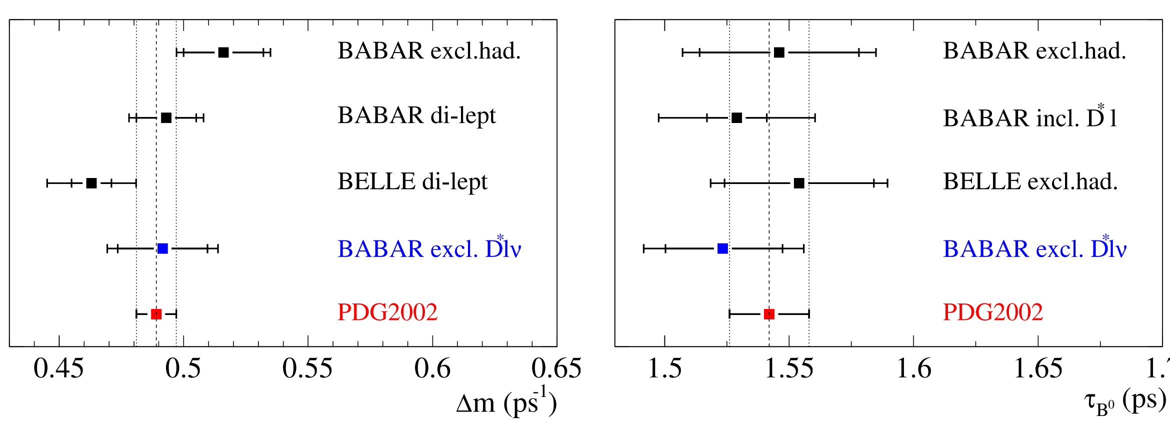

Figure 46: Comparison of this measurement with the world average and other measurements made in BABAR and Belle. The measurements described in this note are not included in the world average. Left: mixing frequency: world average (PDG2002 [18]), BABAR exclusive hadronic measurement [24], BABAR di-lepton measurement [23], and Belle di-lepton measurement [25]; right: lifetime: world average (PDG2002 [17]), BABAR exclusive hadronic measurement [21], BABAR inclusive $D^*\ \ell$ measurement [22], and Belle exclusive hadronic measurement [26].

A Filtered Generic Monte Carlo

Lack of generic Monte Carlo sample is a major concern in our analysis. To increase our generic MC sample size, an additional Monte Carlo sample was generated in release 8.8.1a, using a generator level filter to accept only those events containing a true lepton with center-of-mass momentum $p* > 1.0$ GeV (see ProdDecayFiles/filterInclusiveHardLepton.tcl). This filter is 100% efficient for the signal mode (since we require that the identified lepton candidate have $p* > 1.2$ GeV), and rejects 67% of the generic $B^0\overline{B^0}$ or B^+B^- events. The result is that each reconstructed event in the filtered sample is equivalent to about three true generic events. We use this additional sample of filtered generic events in combination with the existing unfiltered generic sample to form a total Monte Carlo sample with statistics roughly equivalent to the data. Table 70 lists the explicit sizes of the filtered generic samples.

Mode	Number Evts	Filter Eff.	Effective Lumi
$B^0\overline{B^0}$	1.883M	33.061%	10.36/fb
B^+B^-	2.956M	32.717%	16.42/fb

Table 70: Summary statistics for the filtered generic Monte Carlo sample.

Ntuple and ascii file production for the filtered generic Monte Carlo was based on the same analysis-7 release[1], although the following changes were made:

- Update to PDT V00-03-17: The SP3 production release 8.8.1a uses an updated tag of the PDT package, which is required to successfully access truth information.
- Update to BTaggingTools V00-01-28: The new tag fixes known problems with tagging bitmaps in the tagging blocks of the ntuple. This information is not relevant for the analysis, but provides additional insight for systematic studies.
- Minor modifications to NetTagger/TrackFarm.cc to improve compatibility between Linux and Solaris 2.6 based batch processing. (i.e. re-arrange floating point operations to minimize round-off errors)

The new exectuable was tested on previously processed 8.8.0x generic Monte Carlo and produced identical results.

We expect the filtered generic MC to have the same yields per unit luminosity as the unfiltered generic MC for the signal sample, defined as events with a D^* candidate and an opposite-side lepton candidate. However, we do not expect the filtered yields to match the unfiltered yields for the fake control samples because of the removal of all events that do not have a true high-momentum lepton at the generator level. Figure 47 compares the number of events per fb^{-1} in the entire selected sample (left-hand plots) and in the signal sample only (D* candidate plus opposite-side lepton candidate) (right-hand plots) separated according to the true identity of the D* and lepton candidate determined with the Monte Carlo truth associator. The different categories are described in the caption. As expected, the filtered sample has a lower yield in the fake lepton category (2) and the combinatoric D* category (7). The yields in the other categories are consistent for the filtered and unfiltered samples.

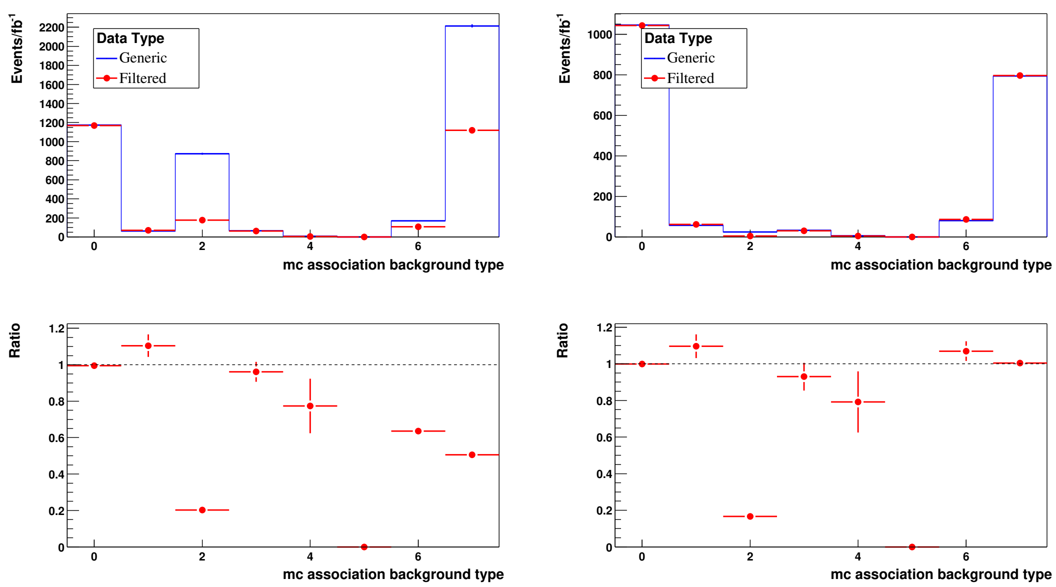

Figure 47: Number of selected events per fb^{-1} in the entire selected sample (signal and control samples; left-hand plots) and in the signal sample only (D* candidate plus opposite-side lepton candidate; right-hand plots) separated according to the true identity of the D* and lepton candidate determined with the Monte Carlo truth associator: $0 = \mathrm{B}^0$ signal, $1 = \mathrm{B}^+$ signal, 2= fake lepton, 3= uncorrelated lepton, 4= cascade lepton, 5= continuum, 6= "semigood" $\mathrm{K}\pi\pi^0$, 7= combinatoric. Yields for unfiltered generic MC are shown in blue and for filtered MC in red.

Figure 48 compares the $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ shapes in the signal samples for the true signal category and true combinatoric D* category. The shapes are consistent between the filtered and unfiltered MC samples.

Figure 49 compares the shape of the Δt_{true} distribution for all selected events in the combined signal and control samples, in the filtered and unfiltered MC. The means of the distributions are consistent with each other(and consistent with 0). However, the RMS is $(0.050 + -0.013)$ ps ($\approx 2.4\%$) larger for the filtered MC.

We next compare the Δt_{reco} distributions of the filtered and unfiltered MC for different subsets of selected events. In Fig. 50, we show the Δt_{reco} distribution for the fake lepton

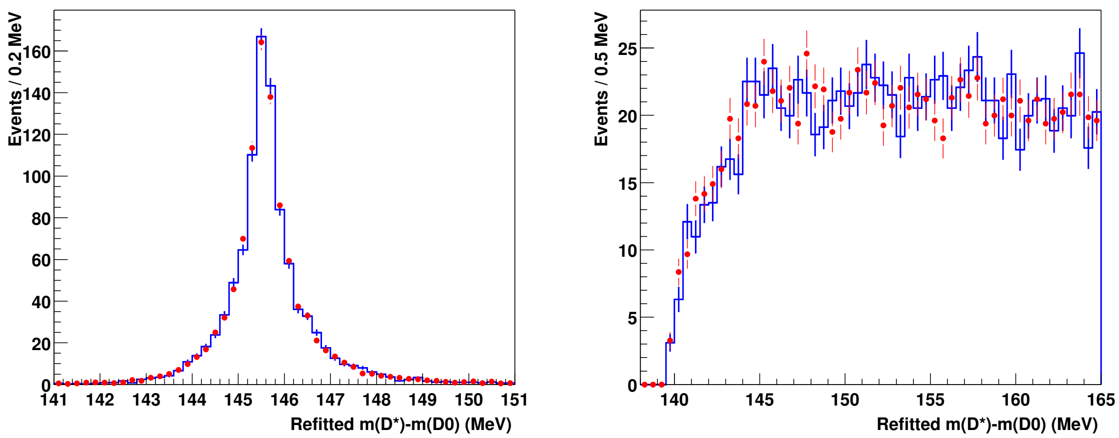

Figure 48: $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ shape for signal sample in 8.8.0x generic (blue histogram) and 8.8.1a filtered sample (red points). Left plot is for the correctly reconstructed signal. Right plot is for misreconstructed D^* sample. All histograms are normalized to the same area (1000 events).

control sample. The RMS is $(0.31 + -0.04)$ ps ($\approx 13\%$) smaller for the filtered MC. We would not expect the fake lepton Δt_{reco} distributions to be the same for the two samples since, in the filtered sample, every event contains a true high momentum lepton, which can affect both vertex resolution and tagging.

In Figure 51, we show the Δt_{reco} distribution for the signal sample (D^* candidate and opposite-sign lepton candidate). The RMS is only $(0.045 + -0.019)$ ps larger for the filtered MC and the means are consistent.

Finally, in Fig. 52, we show the Δt_{reco} distribution for events in the signal sample that are categorized as combinatoric D^* from $\mathrm{B}^0\overline{\mathrm{B}^0}$ events (left-hand plots) and true D^* from $\mathrm{B}^+\mathrm{B}^-$ events (right-hand plots) according to the MC truth associator.

We conclude, therefore, that the filtered generic Monte Carlo samples are directly comparable to the unfiltered sample for the signal sample, including the contribution from peaking backgrounds and combinatoric D^* backgrounds. However, the filtered sample does not represent the fake control sample in yields or in the Δt_{reco} shape or tagging properties. For more details, see the presentations by Tim Meyer and Chih-hsiang Cheng at the August 2, 2001, Lifetime and Mixing meeting.[7]

[7] http://www.slac.stanford.edu/BFROOT/www/Physics/Analysis/AWG/BBMixingHadr/Meetings/02Aug2001/

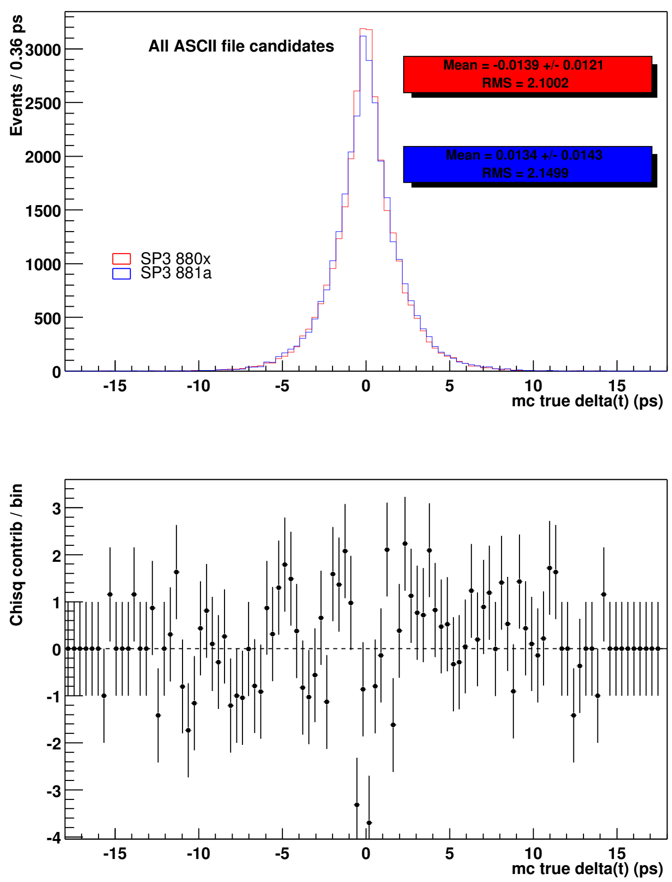

Figure 49: Comparison of the Δt_{true} distribution for combined signal and control samples for unfiltered (8.8.0x) and filtered (8.8.1a) generic $B^0\overline{B^0}$ MC, normalized to the same area. The lower plot shows the chi-square contribution per bin between the two shapes.

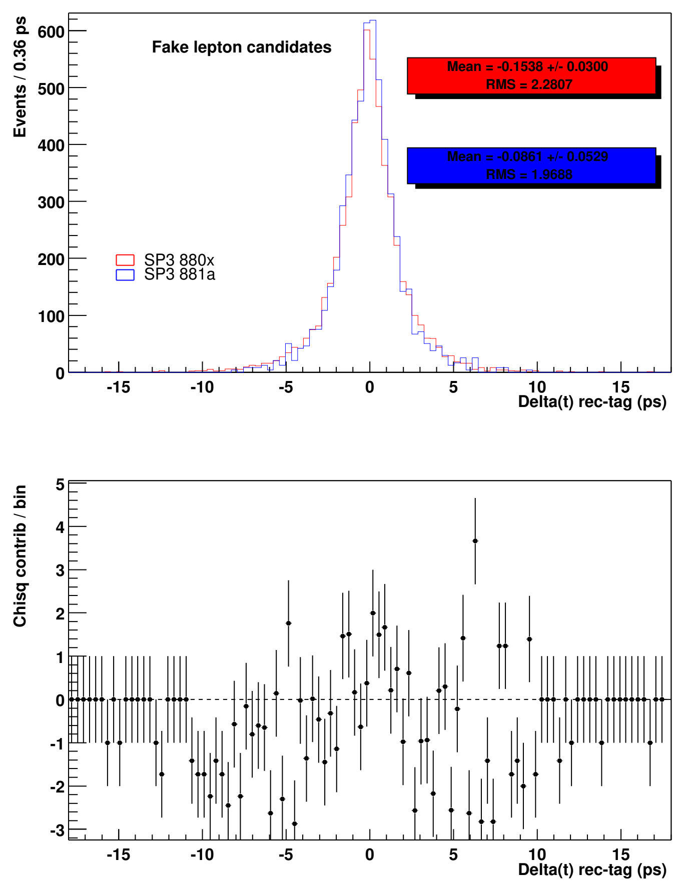

Figure 50: Comparison of Δt_{reco} distribution for the fake lepton control samples in filtered (8.8.1a; blue) and unfiltered (8.8.0x; red) $B^0\overline{B^0}$ generic MC. The spectra are normalized to the same area.

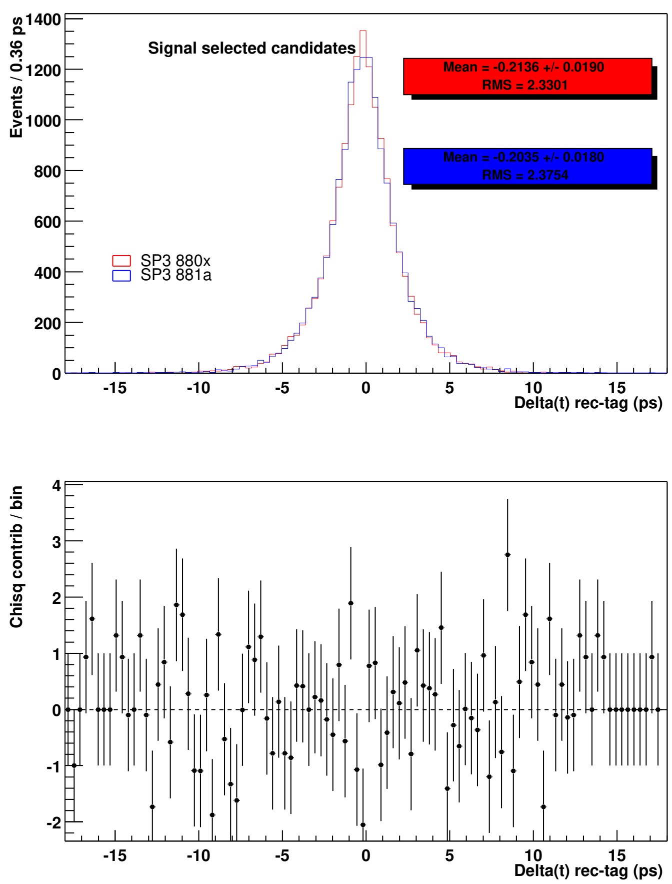

Figure 51: Comparison of the Δt_{reco} distribution for the signal sample (D^* candidate and opposite-sign lepton candidate), for filtered (8.8.1a; blue) and unfiltered (8.8.0x; red) generic MC samples.

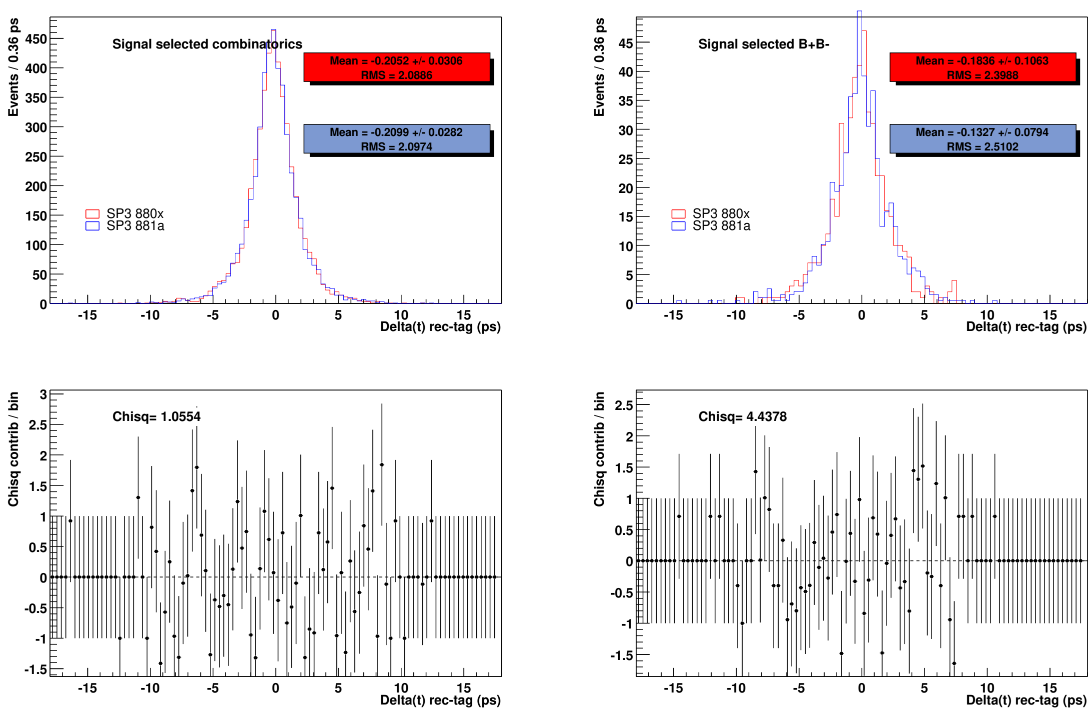

Figure 52: Comparison of the Δt_{reco} distribution for events in the signal sample that are categorized as combinatoric D* from $B^0\overline{B^0}$ events (left-hand plots) and true D* from B^+B^- events (right-hand plots) according to the MC truth associator, for filtered (8.8.1a; blue) and unfiltered (8.8.0x; red) generic MC samples.

B Criteria to Remove Outliers

In this section, we characterize "outliers" in Δt and investigate methods to reject them. For the studies presented in this subsection, the cut on $\sigma_{\Delta t}$ at 1.8 ps is *not* applied. We define an outlier as an event for which $|\delta\Delta t/\sigma_{\Delta t}| > 5$, where $\delta\Delta t = \Delta t_{\rm reco} - \Delta t_{\rm true}$ is the residual. Of course, this definition of outliers is only operative for Monte Carlo events, where we know $\Delta t_{\rm true}$. In Fig. 53 we show the distribution of calculated uncertainties $\sigma_{\Delta t}$ for signal Monte Carlo events with $|\delta\Delta t/\sigma_{\Delta t}| < 5$ (left-hand plot) and $|\delta\Delta t/\sigma_{\Delta t}| > 5$ (right-hand plot). The distribution of errors is not very different for the two classes of events.

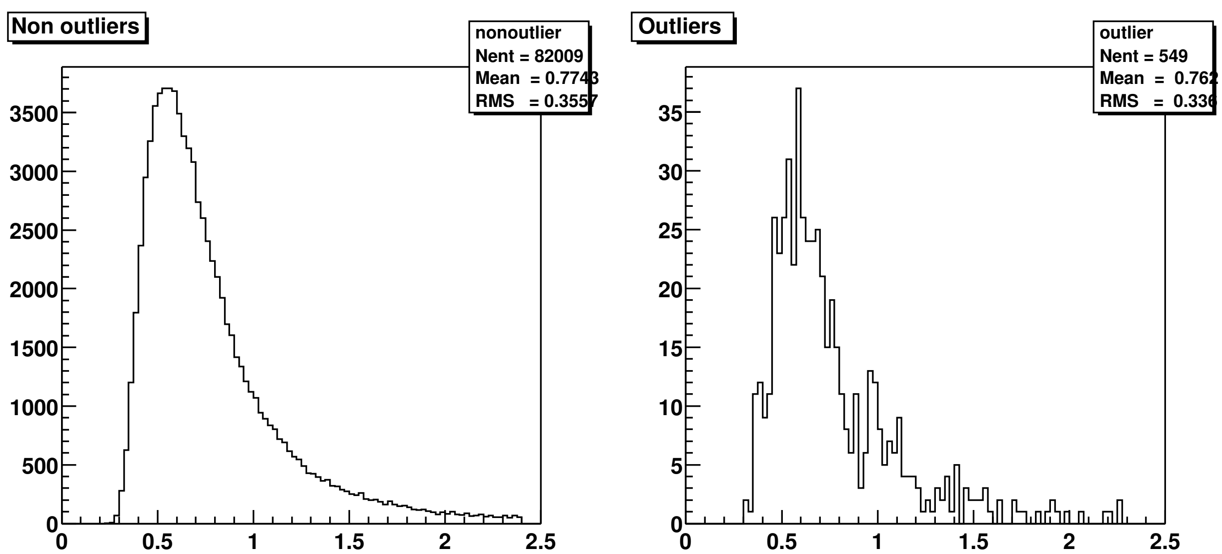

Figure 53: Distribution of calculated event errors $\sigma_{\Delta t}$ in signal Monte Carlo events with $|\delta\Delta t/\sigma_{\Delta t}| < 5$ (left-hand plot) and $|\delta\Delta t/\sigma_{\Delta t}| > 5$ (right-hand plot).

To understand the source of outliers, reconstructed tracks in Monte Carlo signal events with $|\delta\Delta t/\sigma_{\Delta t}| > 5$ were matched to the generated tracks. Individual events were studied to find the source of the large $\delta\Delta t$. Three types of failures were identified.

- Lepton from other B: The D* is correctly reconstructed but the lepton candidate is from the other B.
- Both B's misreconstructed: The event is completely misreconstructed. A mix of charged tracks from both B's is used in each vertex.
- Other B misreconstructed: The $B^0 \to D^{*-}\ell^+\nu_\ell$ candidate is reconstructed correctly but the other B vertex is misreconstructed either because the daughter charm vertex

is found, or a mixture of primary B tracks and daughter charm tracks is used in the vertex.

The first two sources of outliers will tend to decrease $|\Delta t|$ but will not bias the average value of Δt. The third source will lead to a negative bias in Δt since $\Delta t = t_{\rm reco} - t_{\rm tag}$ and $t_{\rm tag}$ will be biased in a positive direction.

In Fig. 54, we show three distributions for signal Monte Carlo events with $|\delta\Delta t/\sigma_{\Delta t}| > 5$ (outliers), in which the tagging B decayed to a D^0, D^+, D_s, Λ_c, Ξ_c or an Ω_c. In each plot, the horizontal axis corresponds to the flight length of the charm particle in the z direction divided by $\beta\gamma c$; i.e., it is the effective error in the decay time of the tagging B that would result from assigning the charm vertex to the tagging B. The vertical axis is the residual for the reconstructed tagging B decay time: $t_{\rm tag}^{\rm meas} - t_{\rm tag}^{\rm true}$. The left plot is for all outliers, the center plot for outliers in which the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate is perfectly reconstructed, and the right-hand plot for outliers in which the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate is not reconstructed correctly. The diagonal bands correspond to events in which the charm decay products dominate the tagging B vertex. The vertical band corresponds to events in which the charm decay products do not dominate the tagging B vertex but the tagging vertex is misreconstructed in some other way. Comparing the density of points in the center and right-hand plots, we see that only a small fraction of outliers are due to misreconstructed $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidates.

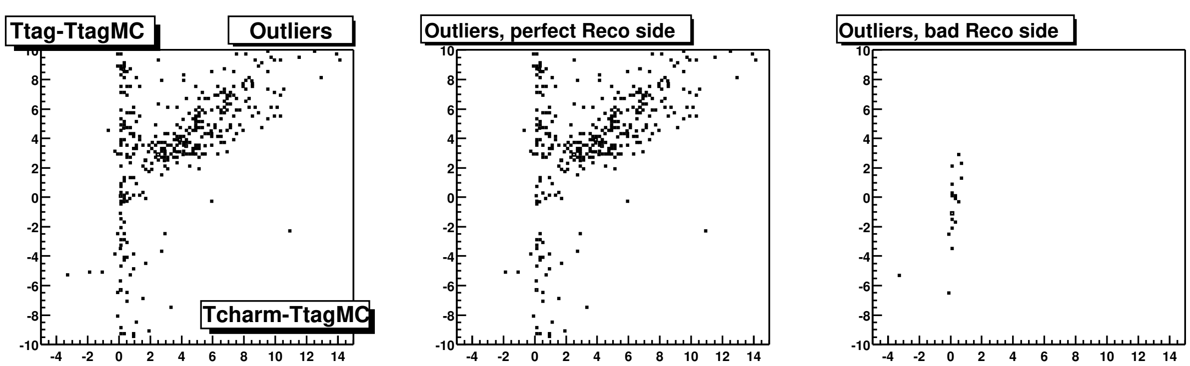

Figure 54: Scatter plots of outliers in signal Monte Carlo events in which the tagging B decayed to a charm particle. In each plot, the horizontal axis corresponds to the flight length of the charm particle in the z direction divided by $\beta\gamma c$. The vertical axis is the residual for the reconstructed tagging B decay time: $t_{\rm tag}^{\rm meas} - t_{\rm tag}^{\rm true}$. The units for both axes are ps. The left plot is for all outliers, the center plot for outliers in which the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate is perfectly reconstructed, and the right-hand plot for outliers in which the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate is not reconstructed correctly.

A number of quantities were studied to determine if they had any rejection power for outliers. The calculated error on Δt is not useful, by definition. The only quantities that

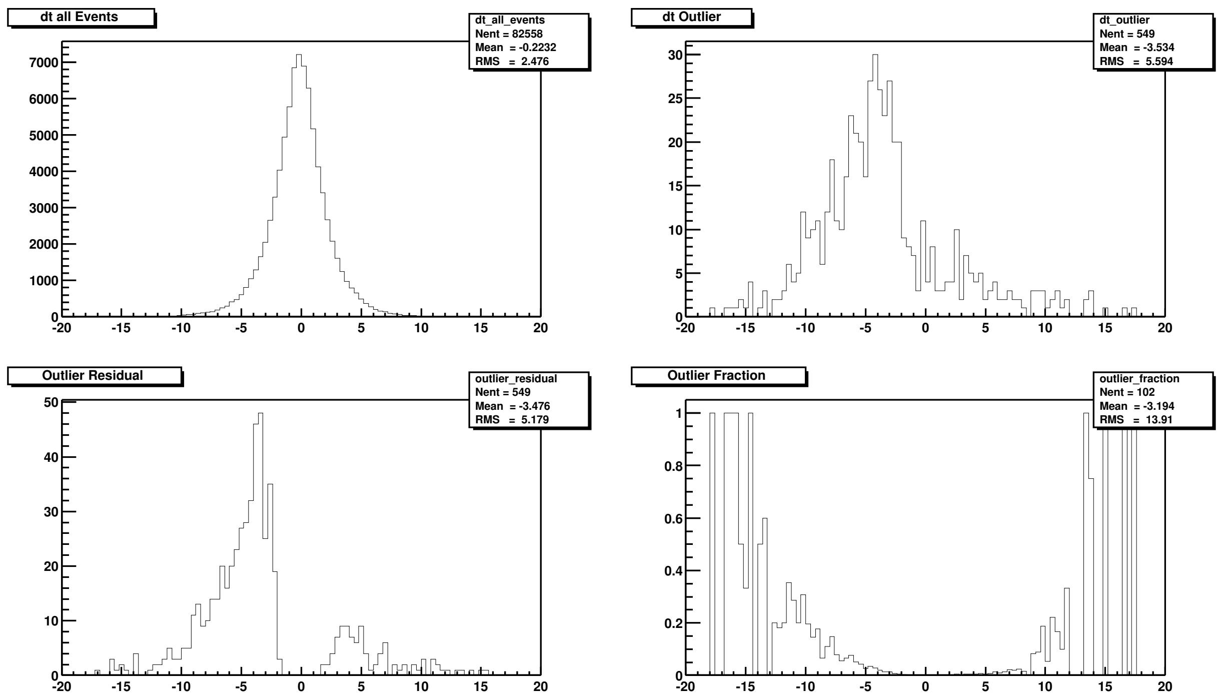

Figure 55: Distribution of Δt for all signal Monte Carlo $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate events (top left) and for outliers (top right). The bottom left plot shows the distribution of residuals $\delta\Delta t = \Delta t_{\mathrm{reco}} - \Delta t_{\mathrm{true}}$ for outliers. The bottom right plot shows the fraction of events that are outliers, as a function of Δt.

were found to discriminate against outliers are Δt itself, the number of tracks used in the tagging B vertex, and the tagging category.

In the top plots in Fig. 55, we show the distribution of Δt for all signal Monte Carlo $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidate events and for outliers. The distribution for outliers is significantly broader as expected. The third plot shows the distribution of residuals $\delta\Delta t = \Delta t_{\mathrm{reco}} - \Delta t_{\mathrm{true}}$ for outliers. The final plot shows the fraction of events that are outliers, as a function of Δt. Although a very significant fraction of events at large Δt are outliers, this is partly due to the definition of outliers ($|\delta\Delta t/\sigma_{\Delta t}| > 5$) combined with the earlier cut on maximum Δt error: $\sigma_{\Delta t} < 2.4$ ps. We have decided not to change the existing cut on $|\Delta t| < 18$ ps. Instead, we will study the effect of varying the Δt cut on the final physics results to estimate systematic uncertainties.

In Fig. 56, we show three distributions of the number of tracks used in the tagging B vertex for signal Monte Carlo. The first plot is the distribution for all events, and the second for outlier events. The third plot shows the fraction of events that are outliers in each bin. (For these plots, we make no requirement on the number of tracks used in the tagging B vertex; it is these plots that motivate the requirement of at least two tracks.) The probability of being an outlier decreases as the number of tracks in the tagging B vertex increases. While only about 6% of all events have one track in the tagging B vertex, 28% of outliers are in this one-track category. About 4% of events with one track in the tagging B vertex are outliers. We have decided to reject events with only one track in the tagging B vertex in order to reduce potential systematic uncertainties due to outliers at the cost of a slightly larger statistical uncertainty.

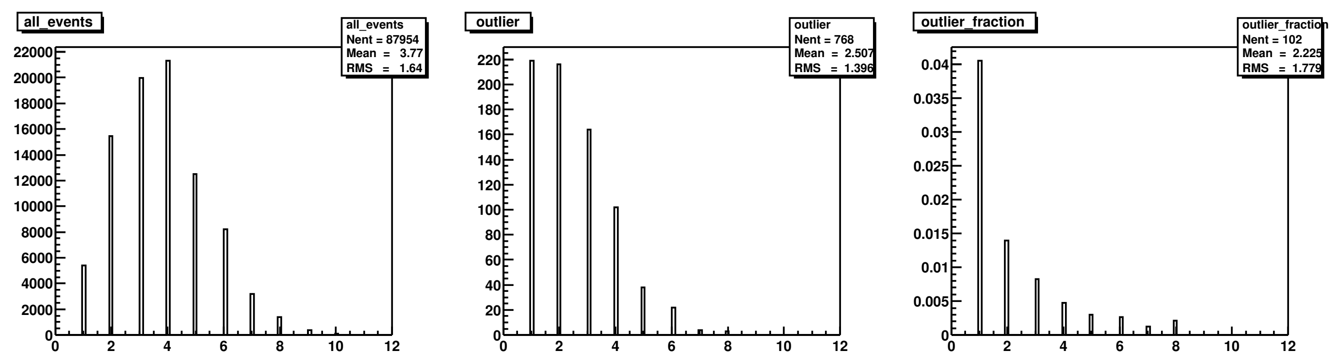

Figure 56: Distributions of the number of tracks used in the tagging B vertex for signal Monte Carlo. The first plot is the distribution for all events, and the second for outlier events. The third plot shows the fraction of events that are outliers in each bin.

We also considered the fraction of outliers in each tagging category (see Sec. 4) for the events in the one-track category. The one-track `lepton` and `NT1` categories have lower outlier fractions than the `kaon` and `NT2` categories, yet higher outlier fractions than the events with two or more tracks in the tagging B vertex. Hence, we reject all events in the one-track category, independent of tagging category.

C Correlation Between $\sigma_{\Delta t}$ and Bias in Resolution Function

As noted above, the upper plot in Fig. 6 and the plots in the second row in Fig. 7 show that the bias of the resolution function scales with the per-event error. The explanation for this is the fact that the vertex error ellipse for the D decay products is oriented with its major axis along the D flight direction, leading to a correlation between the D flight direction and the per-event D^0 error. D mesons that have a flight direction perpendicular to the z direction in the laboratory frame will have the best resolution and will introduce the least bias in a measurement of the z position of the parent B meson. D mesons that travel forward in the laboratory will have poorer resolution and will introduce a larger bias in the measurement of z_{tag}. Jan Stark demonstrated this effect with a toy MC study presented at the June 2001 Collaboration meeting and documented in Sec. 7, BAD 254, V1 (Vertexing Supporting Document for Summer 2001 Conferences).

To verify that this is the correct explanation for the scaling of the bias with per-event error, we study the correlations among D meson direction, $\sigma_{\Delta t}$ and Δt bias.

From signal Monte Carlo sample, we select corrected reconstructed candidates on reco side using Monte Carlo truth association and also require the reco-tag thrust angle cuts and the convergence of D^0. We trace the decay tree for all tracks used for tag-vertexing and look for long-lived charm (D^+, D^0, D_s^+ and Λ_c) and τ. There can be up to three such long-lived particles in the decay tree. For the simplicity of this study, for each event, only the information of the first charm-tau found in the event is recorded. Events with no charm-tau are also selected. Since the residual and error on Δt are dominated by the tag-side, we only focus on the tag-side.

The correlation between Δt mean residual and $\sigma_{\Delta t}$ as seen in Fig. 6 can be reproduced at the tag-side. Figure 57 shows the mean tag-side vertex residual (scaled by $\gamma\beta c$) in slices of $\sigma_{\Delta t}$, for events with one and zero charm-tau particle. Clearly, the correlation in question is due to charm-tau on tag-side, since the events without charm-tau have no bias.

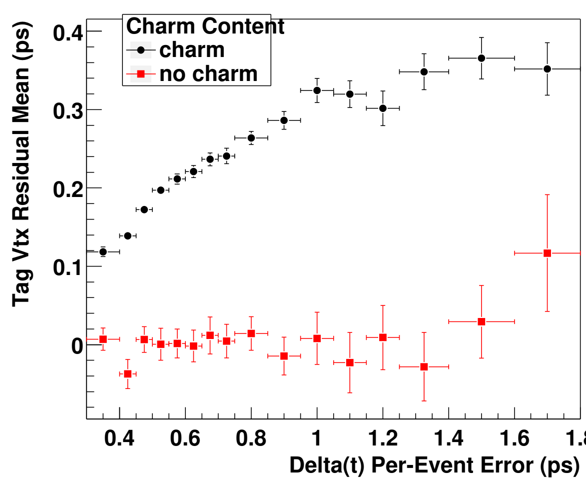

Figure 57: Mean residual of tag-side vertex (scaled by $\gamma\beta c$) in slices of $\sigma_{\Delta t}$ for signal MC events in which at least one track used for tag-vertexing is produced from charm-tau (black) and for events in which no track used for tag-vertexing is produced from charm-tau (red)

As described in BAD 254, the mean residual and the mean $\sigma_{\Delta t}$ are highly correlated with charm-tau flight direction (See upper two plots in Fig. 59.) However, the minimum of $\sigma_{\Delta t}$ is not at $\cos\theta_{\mathrm{D}} = 0$ as one would have expected by charm vertex topology. The fact that the minimum shifts towards positive $\cos\theta_{\mathrm{D}}$ is largely due to the correlation between

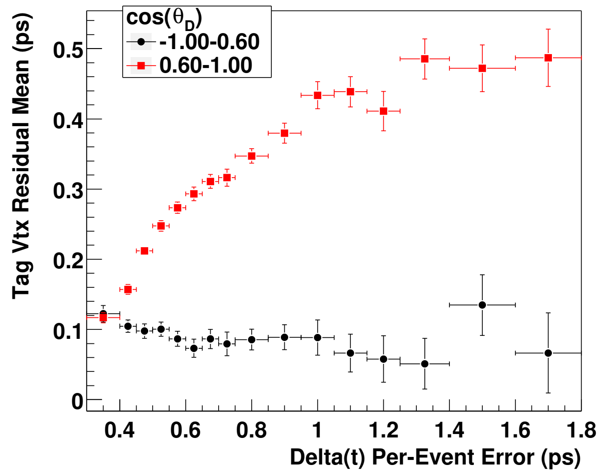

Figure 58: Mean residual of tag vertex for $\cos\theta_{\mathrm{D}} < 0.6$ and $\cos\theta_{\mathrm{D}} > 0.6$ using signal MC events.

the tracking error and the momentum of that track. The charm-tau momentum in the lab frame is highly correlated with its flight direction due to the boost of the center-of-mass frame (lower-left plot in Fig. 59). In general the higher momentum a track has, the smaller its error is, therefore, the error ellipse of the charm-tau vertex is smaller for high momentum (forward direction) charm-tau, since they tend to produce higher momentum tracks. This argument is supported by lower-right plot in Fig. 59, in which only events with $1.45\,\mathrm{GeV} < p_{\mathrm{D}} < 1.50\,\mathrm{GeV}$ are plotted. The minimum shifts back closer to zero.

We then divide the sample into two sets with $\cos\theta_{\mathrm{D}} < 0.6$ and $\cos\theta_{\mathrm{D}} > 0.6$, and plot their residual mean and $\sigma_{\Delta t}$ correlation (see Fig. 58.). These two sets have opposite behavior, but the one with $\cos\theta_{\mathrm{D}} > 0.6$ dominates since it has more events and stronger correlation. In our analysis, we cannot distinguish these two classes of events. We can only model their averaged behavior.

D Study of Δt Resolution for Signal Events

This appendix contains many of the plots and tables detailing the resolution study for signal events. This appendix includes

- Tables of resolution parameters for GExp+G, 3G and 2G fits.
- Plots comparing the resolution parameters for GExp+G, 3G and 2G fits.
- Lifetime plots using fixed and floating resolution parameters.
- Lifetime comparisons based on resolution models.

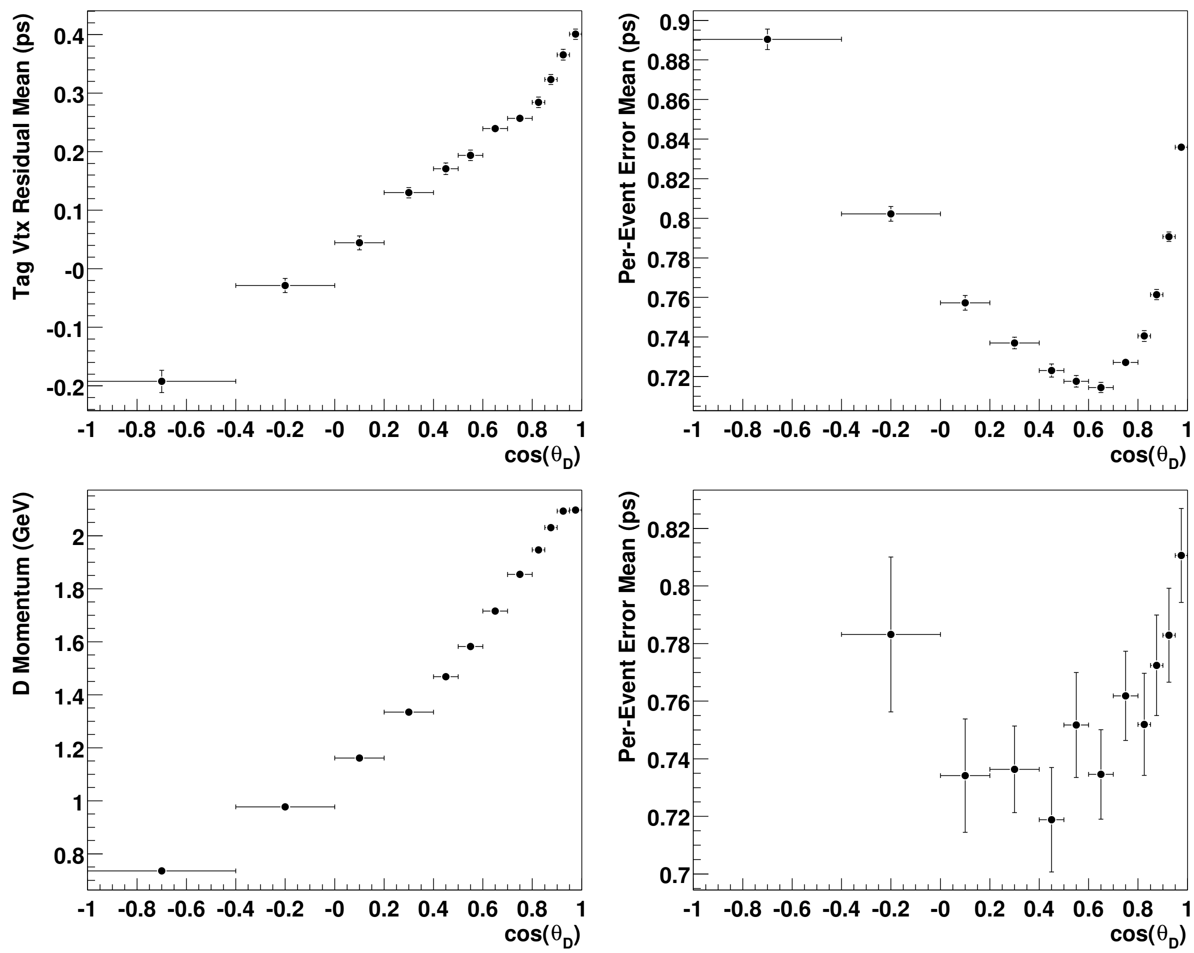

Figure 59: Various quantities in slices of $\cos\theta_{\rm D}$ where $\theta_{\rm D}$ is the polar angle of charm-tau (from which at least one track used for tag-vertexing is produced from) for signal MC events. Upper left: mean residual of tag vertex; upper right: mean $\sigma_{\Delta t}$; lower left: mean lab momentum of charm-tau; lower right: same as upper right, but only for $1.45\,{\rm GeV} < p_{\rm D} < 1.50\,{\rm GeV}$.

Figures 66 and 67 compare these values obtained from the fit with those calculated directly from the residuals. The fit systematically underestimates the mean bias and overestimates the RMS resolution. There are several possible explanations for this:

- The fit mean and RMS are integrated over $(-\infty,+\infty)$ while the residuals in data are truncated by the $|\Delta z| < 3$ mm cut. This should have litte effect on the mean but will increase the RMS calculated for the fit.

- The fit weights each event according to its calculated Δt error, but the statistics calculated directly on the data are unweighted.

- The fit assumes an equal number of events at large positive and negative residual since the outlier component is unbiased. Figure 61 actually shows an excess of events at large negative residual, as compared with the fit.

Sample	b_1	s_1	b_{out}	s_{out}	f_1
All Combined	-0.23740	1.1843	-1.7593	3.880	0.9586
	$\pm$0.004	$\pm$0.004	$\pm$0.08	$\pm$0.07	$\pm$0.002
Electrons	-0.24383	1.1972	-1.917	4.22	0.9645
	$\pm$0.006	$\pm$0.006	$\pm$0.1	$\pm$0.1	$\pm$0.002
Muons	-0.23241	1.1774	-1.645	3.677	0.9550
	$\pm$0.007	$\pm$0.006	$\pm$0.1	$\pm$0.09	$\pm$0.003
SVT only	-0.23664	1.1856	-1.7287	3.900	0.9597
	$\pm$0.005	$\pm$0.004	$\pm$0.09	$\pm$0.08	$\pm$0.002
SVT+DCH	-0.24045	1.1790	-1.868	3.80	0.9540
	$\pm$0.010	$\pm$0.009	$\pm$0.2	$\pm$0.1	$\pm$0.004
$\mathrm{D}^0 \rightarrow \mathrm{K}\pi$	-0.23528	1.1819	-1.863	3.926	0.9603
	$\pm$0.006	$\pm$0.005	$\pm$0.1	$\pm$0.09	$\pm$0.002
$\mathrm{D}^0 \rightarrow \mathrm{K}3\pi$	-0.23618	1.1817	-1.522	3.51	0.9491
	$\pm$0.009	$\pm$0.009	$\pm$0.1	$\pm$0.1	$\pm$0.004
$\mathrm{D}^0 \rightarrow \mathrm{K}\pi\pi^0$	-0.2373	1.1878	-1.531	3.92	0.9577
	$\pm$0.01	$\pm$0.010	$\pm$0.2	$\pm$0.2	$\pm$0.004
$\mathrm{D}^0 \rightarrow \mathrm{K_S}\pi\pi$	-0.2811	1.191	-3.767	5.46	0.9764
	$\pm$0.03	$\pm$0.02	$\pm$1.0	$\pm$0.6	$\pm$0.005
Reco'd B^0	-0.24565	1.1943	-1.891	4.22	0.9644
	$\pm$0.006	$\pm$0.005	$\pm$0.1	$\pm$0.1	$\pm$0.002
Reco'd $\overline{\mathrm{B}^0}$	-0.22752	1.1715	-1.6177	3.535	0.9503
	$\pm$0.006	$\pm$0.006	$\pm$0.10	$\pm$0.08	$\pm$0.003
B^0 Tag	-0.24393	1.1862	-1.944	4.064	0.9607
	$\pm$0.006	$\pm$0.005	$\pm$0.1	$\pm$0.10	$\pm$0.002
$\overline{\mathrm{B}^0}$ Tag	-0.23007	1.1809	-1.559	3.657	0.9551
	$\pm$0.006	$\pm$0.006	$\pm$0.1	$\pm$0.09	$\pm$0.003
Lepton Tag	-0.1359	1.137	-1.252	3.19	0.9616
	$\pm$0.01	$\pm$0.01	$\pm$0.2	$\pm$0.2	$\pm$0.006
Kaon Tag	-0.28711	1.2182	-1.810	3.76	0.9593
	$\pm$0.008	$\pm$0.007	$\pm$0.1	$\pm$0.1	$\pm$0.003
NT1 Tag	-0.1865	1.201	-1.882	5.39	0.9766
	$\pm$0.01	$\pm$0.01	$\pm$0.4	$\pm$0.4	$\pm$0.003
NT2 Tag	-0.2441	1.185	-1.632	4.36	0.9618
	$\pm$0.01	$\pm$0.01	$\pm$0.2	$\pm$0.2	$\pm$0.004
NT3 Tag	-0.23448	1.1528	-1.815	3.67	0.9458
	$\pm$0.008	$\pm$0.008	$\pm$0.1	$\pm$0.1	$\pm$0.004
Perfect Tag	-0.22491	1.1834	-1.7366	3.737	0.9595
	$\pm$0.005	$\pm$0.005	$\pm$0.09	$\pm$0.08	$\pm$0.002
Imperfect Tag	-0.27145	1.1852	-1.808	4.19	0.9551
	$\pm$0.008	$\pm$0.008	$\pm$0.2	$\pm$0.1	$\pm$0.003
Unmixed	-0.22461	1.1764	-1.855	3.814	0.9623
	$\pm$0.005	$\pm$0.005	$\pm$0.1	$\pm$0.08	$\pm$0.002
Mixed	-0.26235	1.2006	-1.632	3.99	0.9524
	$\pm$0.008	$\pm$0.007	$\pm$0.1	$\pm$0.1	$\pm$0.003

Table 71: Results of fitting a G+G resolution model to $\delta\Delta t$ and $\sigma_{\Delta t}$ for signal MC.

Sample	b_1	s_1	b_2	s_2	b_{out}	s_{out}	f_1	f_{out}
Combined	-0.17196	1.0803	-0.8119	1.957	-2.854	6.69	0.824	0.00808
	±0.007	±0.009	±0.04	±0.05	±0.3	±0.3	±0.01	±0.0009
Electrons	-0.17354	1.085	-0.8375	1.989	-3.548	8.11	0.827	0.00594
	±0.010	±0.01	±0.06	±0.06	±0.6	±0.6	±0.02	±0.0008
Muons	-0.1510	1.047	-0.6377	1.730	-2.251	5.22	0.750	0.0146
	±0.01	±0.02	±0.06	±0.07	±0.3	±0.3	±0.04	±0.002
SVT only	-0.16796	1.074	-0.7575	1.903	-2.751	6.51	0.810	0.00865
	±0.008	±0.01	±0.05	±0.05	±0.3	±0.3	±0.02	±0.0009
SVT+DCH	-0.1832	1.093	-1.005	2.138	-3.338	7.54	0.855	0.0060
	±0.01	±0.02	±0.1	±0.10	±0.9	±0.9	±0.02	±0.002
$D^0 \to K\pi$	-0.1680	1.078	-0.8210	1.942	-3.075	6.55	0.826	0.0084
	±0.01	±0.01	±0.07	±0.08	±0.4	±0.4	±0.02	±0.001
$D^0 \to K3\pi$	-0.1795	1.096	-0.8337	1.977	-2.123	6.06	0.831	0.0090
	±0.01	±0.02	±0.10	±0.10	±0.5	±0.5	±0.03	±0.002
$D^0 \to K\pi\pi^0$	-0.1656	1.059	-0.7108	1.916	-2.382	7.14	0.789	0.0074
	±0.02	±0.02	±0.08	±0.08	±0.8	±0.7	±0.03	±0.002
$D^0 \to K_S\pi\pi$	-0.1922	1.082	-1.112	1.92	-4.00000	8.1	0.86	0.0089
	±0.06	±0.07	±0.5	±0.4	±0.001	±2	±0.1	±0.005
Reco'd B^0	-0.19250	1.108	-0.9551	2.158	-3.358	8.39	0.868	0.00532
	±0.009	±0.01	±0.07	±0.08	±0.7	±0.7	±0.01	±0.0010
Reco'd $\overline{B^0}$	-0.1560	1.058	-0.7260	1.847	-2.539	5.67	0.787	0.0108
	±0.01	±0.01	±0.05	±0.06	±0.3	±0.3	±0.02	±0.001
B^0 Tag	-0.18037	1.079	-0.8327	2.010	-3.659	7.15	0.830	0.0073
	±0.010	±0.01	±0.06	±0.07	±0.5	±0.5	±0.02	±0.001
$\overline{B^0}$ Tag	-0.1632	1.083	-0.7992	1.908	-2.111	6.17	0.819	0.0091
	±0.01	±0.01	±0.06	±0.06	±0.4	±0.4	±0.02	±0.001
Lepton Tag	-0.1006	1.063	-0.479	1.76	-2.081	4.63	0.833	0.0094
	±0.02	±0.03	±0.1	±0.1	±0.6	±0.5	±0.05	±0.003
Kaon Tag	-0.2030	1.099	-0.8316	1.878	-2.803	6.18	0.794	0.0086
	±0.01	±0.02	±0.08	±0.08	±0.5	±0.5	±0.03	±0.002
NT1 Tag	-0.1518	1.143	-1.234	2.64	-0.38	16.9	0.930	0.0021
	±0.02	±0.02	±0.2	±0.2	±6	±7	±0.02	±0.001
NT2 Tag	-0.1981	1.105	-0.951	2.24	-2.00	8.9	0.875	0.0057
	±0.01	±0.02	±0.1	±0.1	±1	±1	±0.02	±0.002
NT3 Tag	-0.1556	1.020	-0.7329	1.852	-3.189	5.86	0.765	0.0122
	±0.01	±0.02	±0.07	±0.07	±0.5	±0.4	±0.03	±0.002
Perfect Tag	-0.16523	1.082	-0.7441	1.923	-2.984	6.12	0.822	0.0083
	±0.008	±0.01	±0.05	±0.06	±0.4	±0.3	±0.02	±0.001
Imperfect Tag	-0.1879	1.067	-0.9471	1.996	-2.430	7.89	0.814	0.0084
	±0.01	±0.02	±0.08	±0.09	±0.6	±0.7	±0.02	±0.002
Unmixed	-0.16743	1.080	-0.7889	1.958	-3.378	6.48	0.838	0.0071
	±0.008	±0.01	±0.06	±0.07	±0.4	±0.4	±0.02	±0.001
Mixed	-0.1781	1.078	-0.8396	1.935	-2.104	6.82	0.792	0.0102
	±0.01	±0.02	±0.07	±0.07	±0.5	±0.5	±0.03	±0.002

Table 72: Results of fitting a G+G+G resolution model to $\delta\Delta t$ and $\sigma_{\Delta t}$ for signal MC.

Sample	s_1	f_1	κ	s_{out}	b_{out}	f_{out}
All Combined	1.0963	0.6614	0.849	5.45	-1.472	0.0159
	±0.004	±0.009	±0.02	±0.2	±0.2	±0.001
Electrons	1.1030	0.673	0.890	6.47	-1.607	0.0115
	±0.006	±0.01	±0.03	±0.3	±0.3	±0.001
Muons	1.0939	0.641	0.779	4.72	-1.428	0.0217
	±0.006	±0.02	±0.03	±0.2	±0.2	±0.002
SVT only	1.0981	0.658	0.832	5.41	-1.432	0.0161
	±0.004	±0.01	±0.02	±0.2	±0.2	±0.001
SVT+DCH	1.0897	0.674	0.922	5.66	-1.615	0.0148
	±0.009	±0.02	±0.05	±0.4	±0.4	±0.003
$\mathrm{D}^0 \to \mathrm{K}\pi$	1.0922	0.664	0.846	5.47	-1.658	0.0158
	±0.005	±0.01	±0.03	±0.2	±0.2	±0.002
$\mathrm{D}^0 \to \mathrm{K}3\pi$	1.1044	0.651	0.835	4.89	-1.165	0.0185
	±0.008	±0.02	±0.04	±0.3	±0.3	±0.003
$\mathrm{D}^0 \to \mathrm{K}\pi\pi^0$	1.0996	0.664	0.853	5.49	-1.000	0.0163
	±0.009	±0.02	±0.04	±0.4	±0.4	±0.003
$\mathrm{D}^0 \to \mathrm{K_S}\pi\pi$	1.062	0.634	0.88	7.9	-3.0000	0.0109
	±0.02	±0.05	±0.1	±2	±0.01	±0.004
Reco'd B^0	1.0997	0.663	0.864	6.09	-1.552	0.0133
	±0.005	±0.01	±0.03	±0.3	±0.3	±0.001
Reco'd $\overline{\mathrm{B}^0}$	1.0916	0.657	0.827	4.79	-1.385	0.0200
	±0.006	±0.01	±0.03	±0.2	±0.2	±0.002
B^0 Tag	1.0925	0.657	0.856	5.70	-1.765	0.0157
	±0.005	±0.01	±0.03	±0.2	±0.2	±0.001
$\overline{\mathrm{B}^0}$ Tag	1.0997	0.665	0.840	5.12	-1.157	0.0166
	±0.006	±0.01	±0.03	±0.2	±0.2	±0.002
Lepton Tag	1.094	0.756	0.690	3.90	-1.122	0.0197
	±0.01	±0.03	±0.08	±0.3	±0.3	±0.005
Kaon Tag	1.1154	0.585	0.809	5.25	-1.533	0.0157
	±0.007	±0.02	±0.03	±0.3	±0.3	±0.002
NT1 Tag	1.118	0.786	1.036	8.9	-0.44	0.0069
	±0.01	±0.02	±0.09	±1	±1	±0.002
NT2 Tag	1.0945	0.665	0.864	6.14	-1.009	0.0152
	±0.010	±0.02	±0.05	±0.5	±0.5	±0.003
NT3 Tag	1.0648	0.654	0.868	4.95	-1.667	0.0232
	±0.007	±0.02	±0.04	±0.2	±0.2	±0.003
Perfect Tag	1.1025	0.6661	0.811	5.08	-1.560	0.0167
	±0.004	±0.009	±0.02	±0.2	±0.2	±0.001
Imperfect Tag	1.0775	0.642	0.933	6.30	-1.163	0.0152
	±0.007	±0.02	±0.03	±0.4	±0.4	±0.002
Unmixed	1.0940	0.672	0.819	5.31	-1.756	0.0148
	±0.005	±0.01	±0.03	±0.2	±0.2	±0.001
Mixed	1.1012	0.639	0.893	5.61	-1.050	0.0182
	±0.007	±0.01	±0.03	±0.3	±0.3	±0.002

Table 73: Results of fitting a GExp+G resolution model to $\delta\Delta t$ and $\sigma_{\Delta t}$ for signal MC.

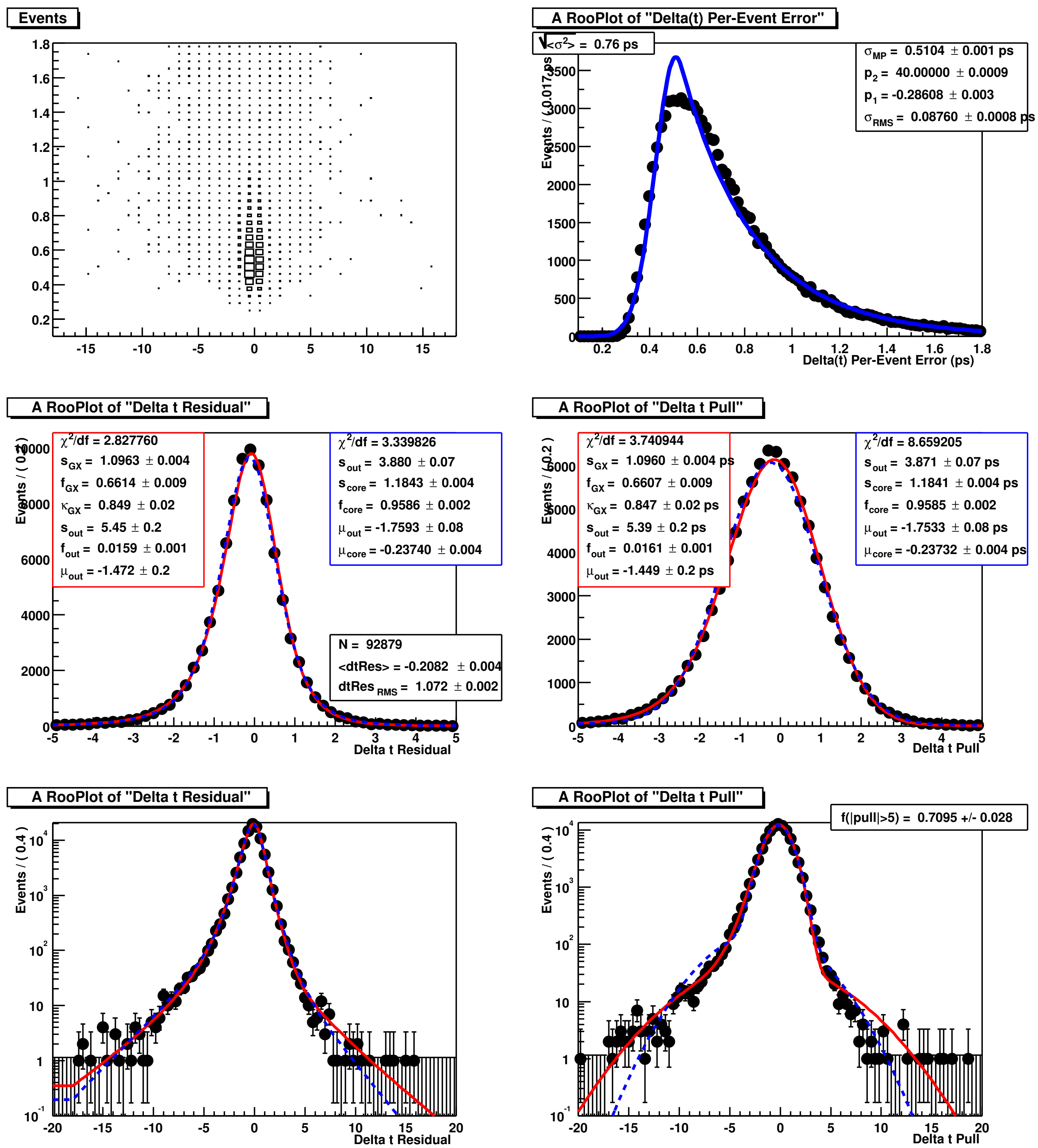

Figure 60: Fits to signal MC residuals and errors, for G+G model (solid curve and parameters in right-hand box) and Gexp+G model (dashed curve and parameters in left-hand box).

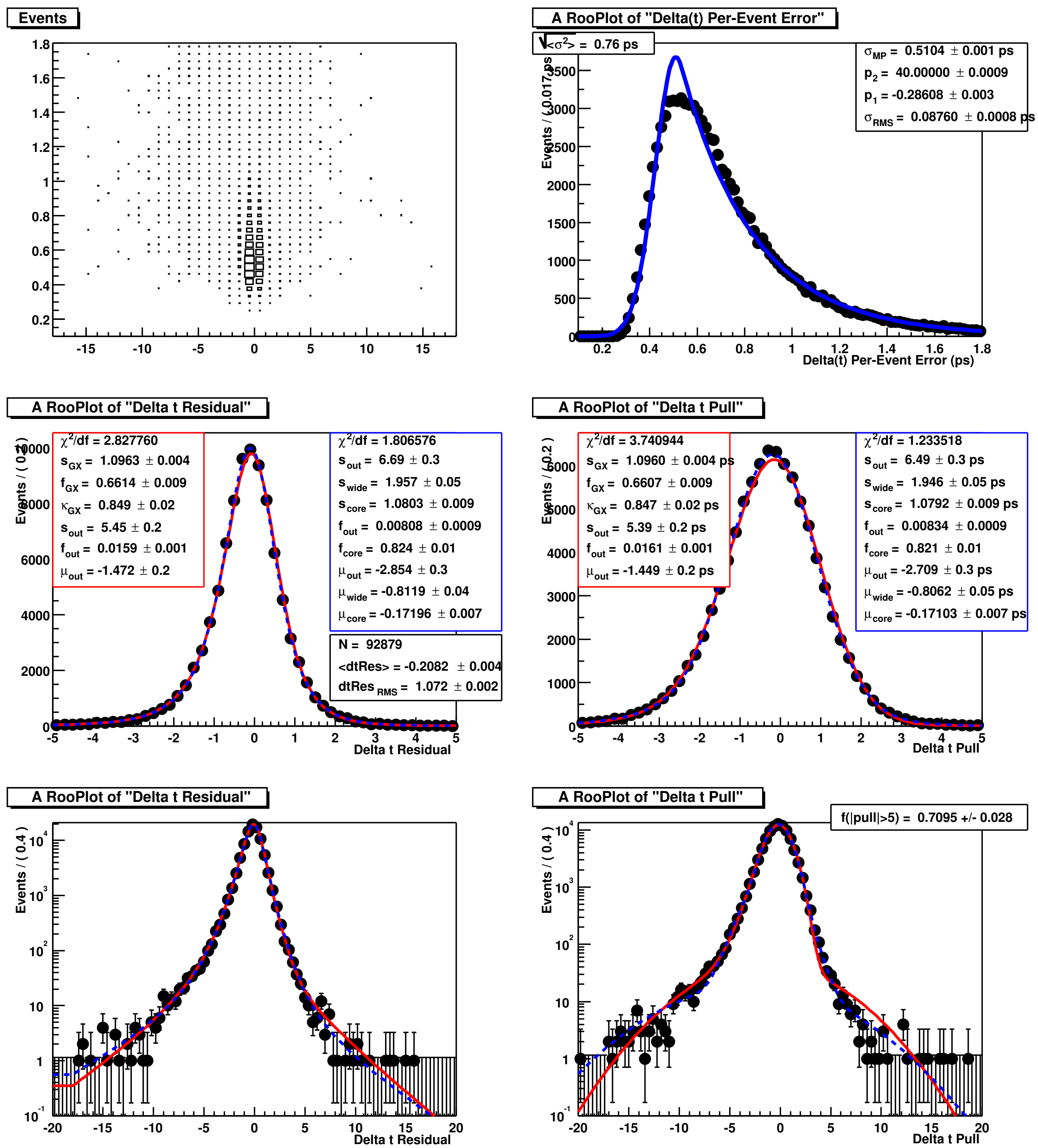

Figure 61: Fits to signal MC residuals and errors, for G+G+G model (solid curve and parameters in right-hand box) and Gexp+G model (dashed curve and parameters in left-hand box).

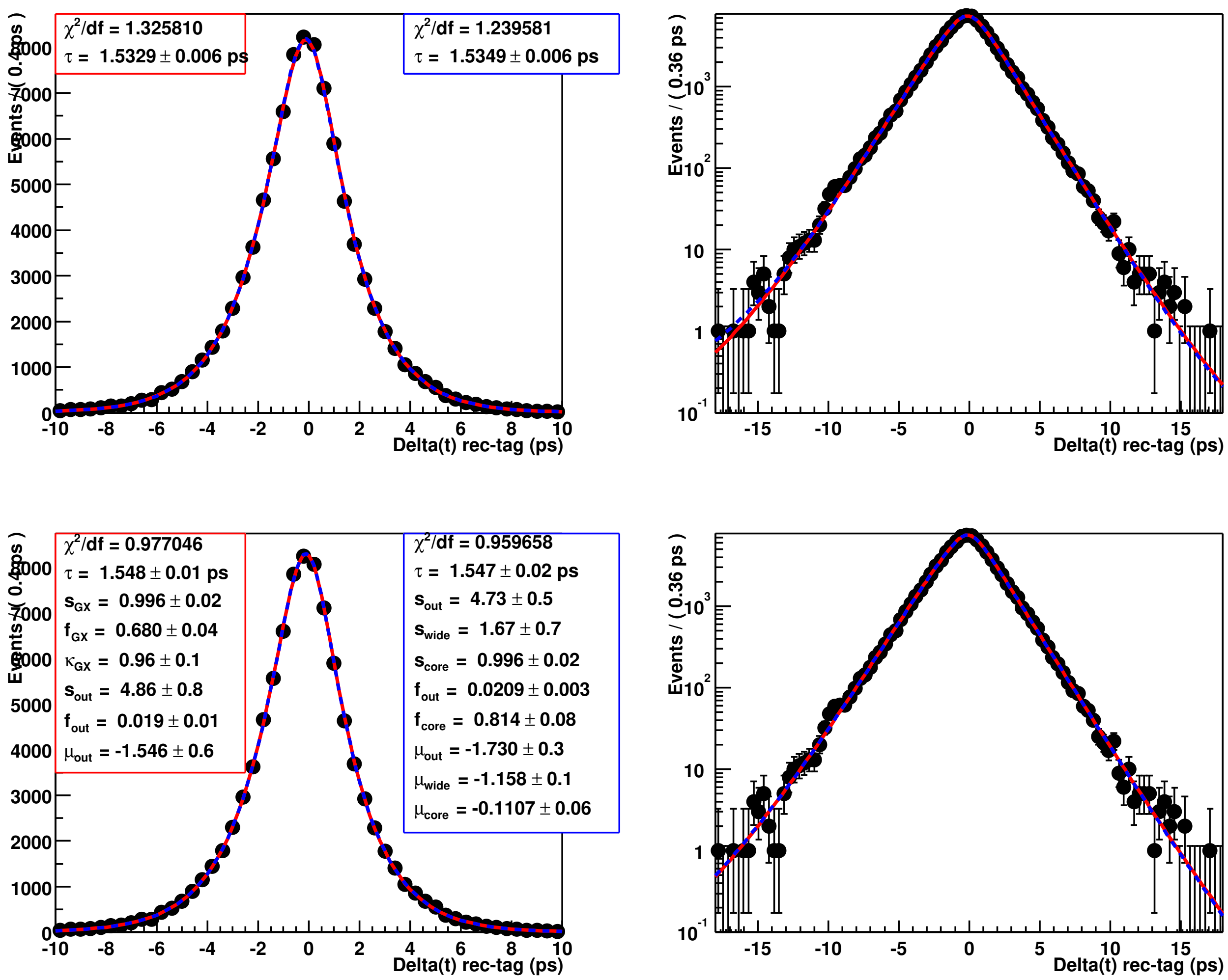

Figure 62: Lifetime fits to reconstructed Δt in signal MC, for G+G+G model (solid curves and parameters in right-hand box) and GExp+G model (dashed curves and parameters in left-hand box). In the top plots, the lifetime is the only free parameter; in the bottom plots, the resolution parameters are fit as well.

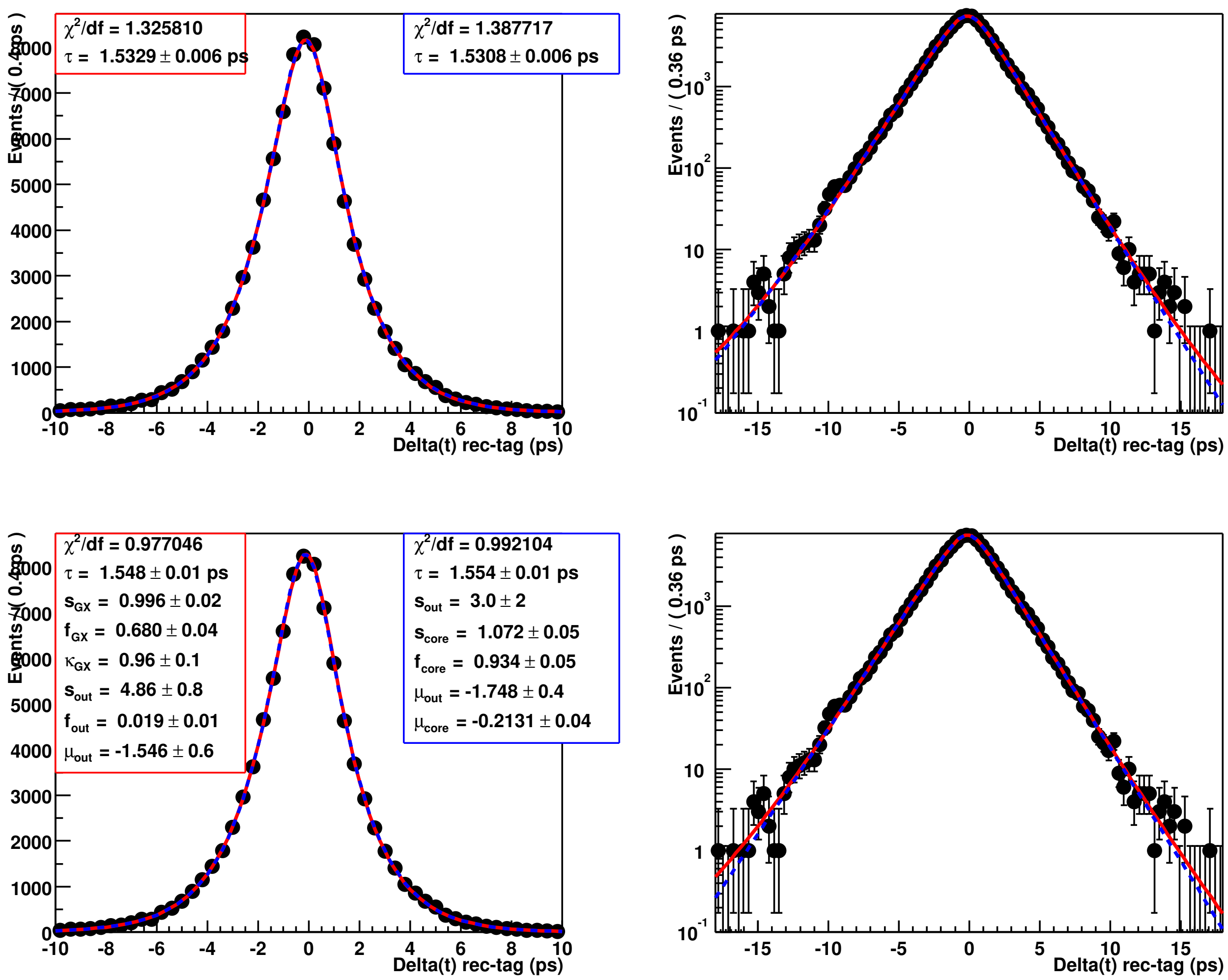

Figure 63: Lifetime fits to reconstructed Δt in signal MC, for G+G model (solid curves and parameters in right-hand box) and GExp+G model (dashed curves and parameters in left-hand box). In the top plots, the lifetime is the only free parameter; in the bottom plots, the resolution parameters are fit as well.

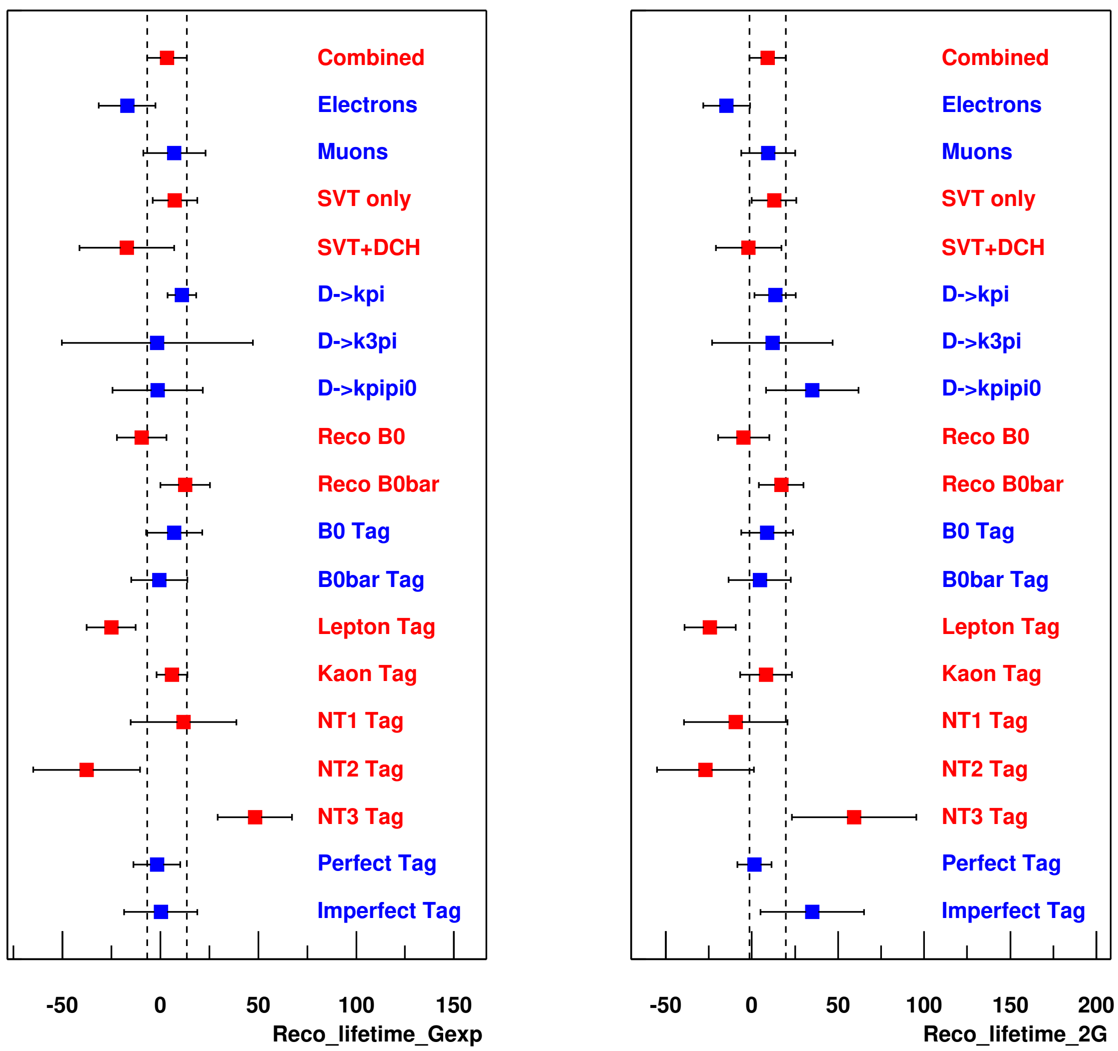

Figure 64: Comparison of lifetimes fitted with the GExp+G (left-hand side) or G+G (right-hand side) resolution models using correctly reconstructed signal B^0 MC events. Both the lifetime and the resolution model parameters are free in the fit. The displayed values are (fitted value − MC truth) in ps.

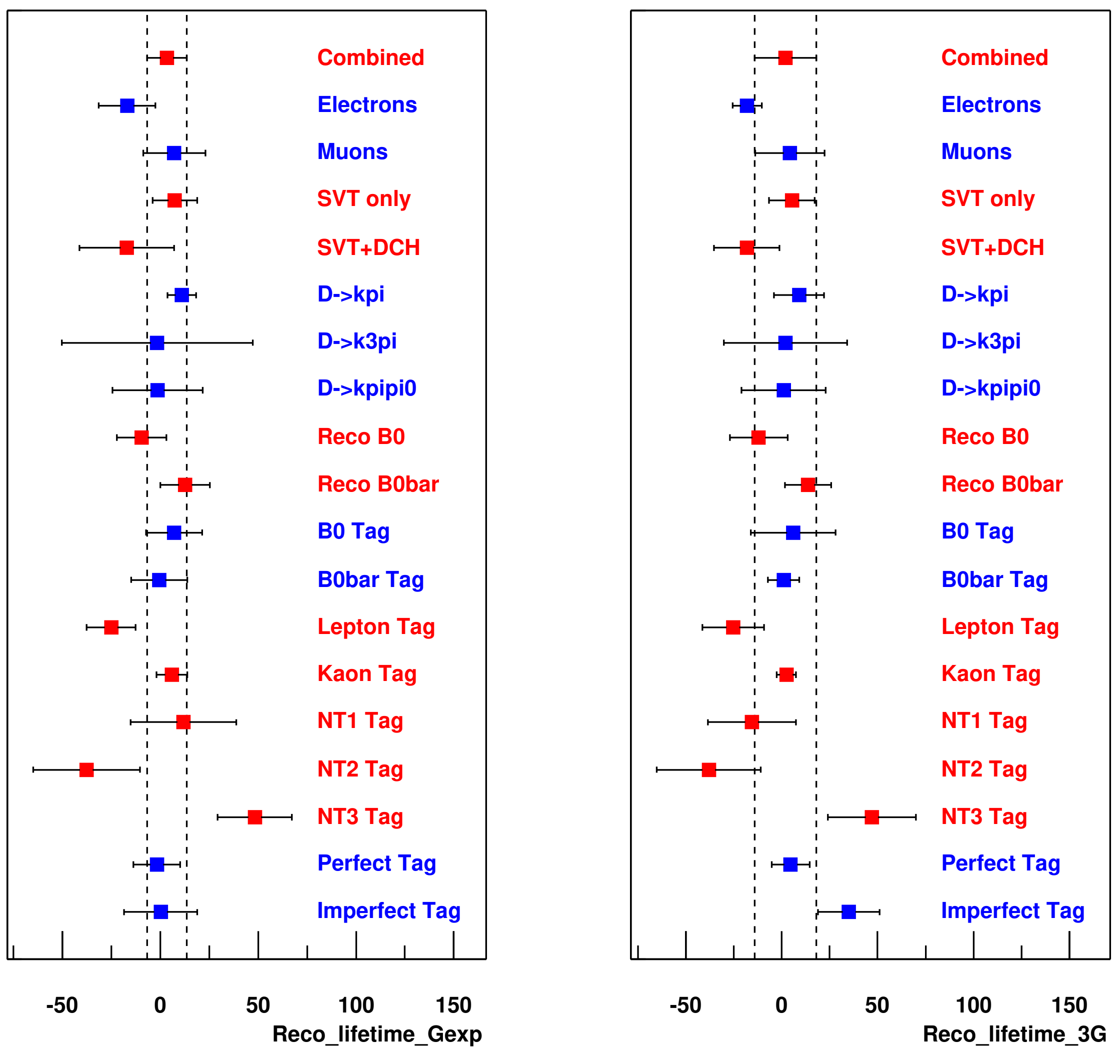

Figure 65: Comparison of lifetimes fitted with the GExp+G (left-hand side) or G+G+G (right-hand side) resolution models using correctly reconstructed signal B^0 MC events. Both the lifetime and the resolution model parameters are free in the fit. The displayed values are (fitted value − MC truth) in ps.

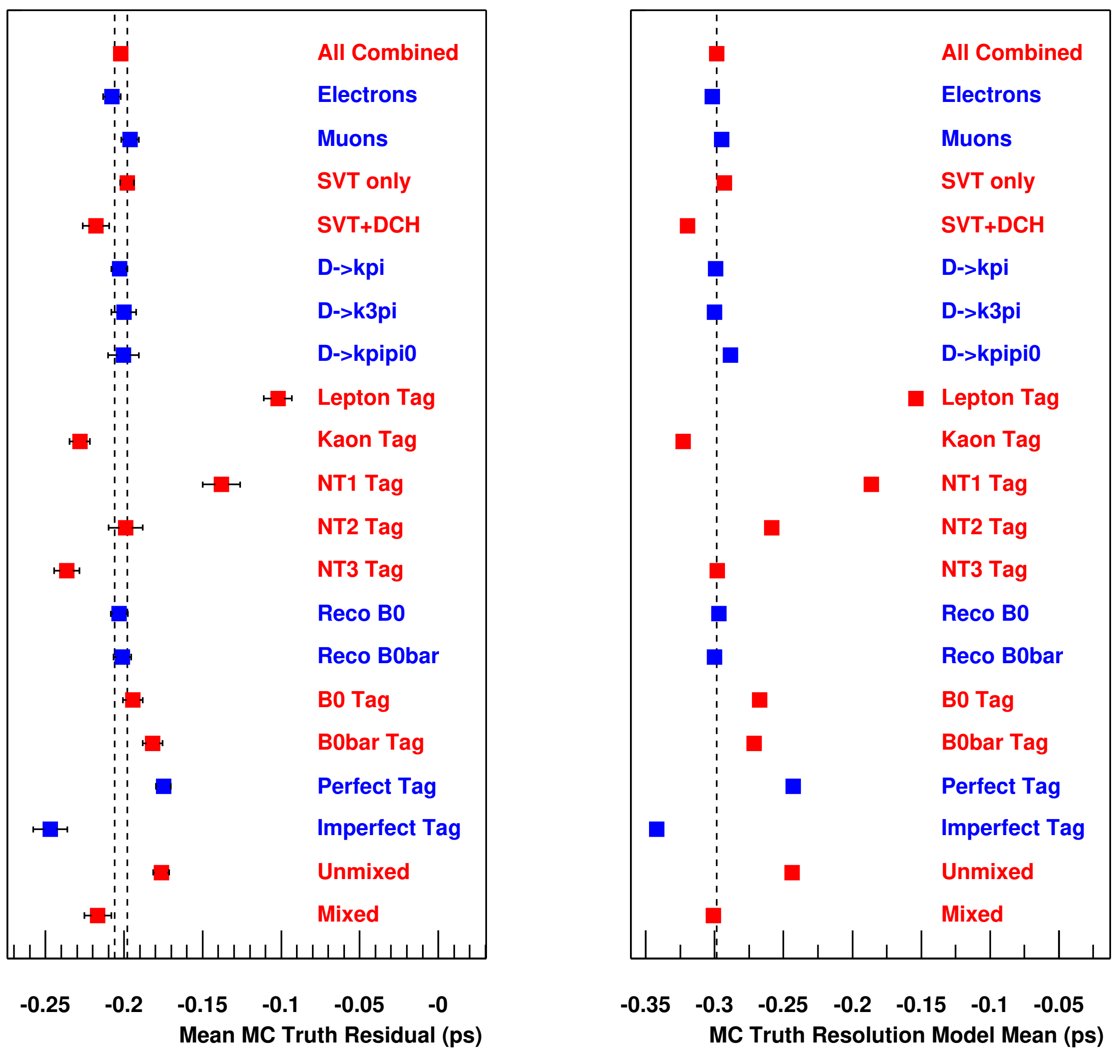

Figure 66: Comparison of mean residuals in the signal MC sample, calculated directly from MC truth (left-hand side) or calculated from models fit to MC truth (right-hand side).

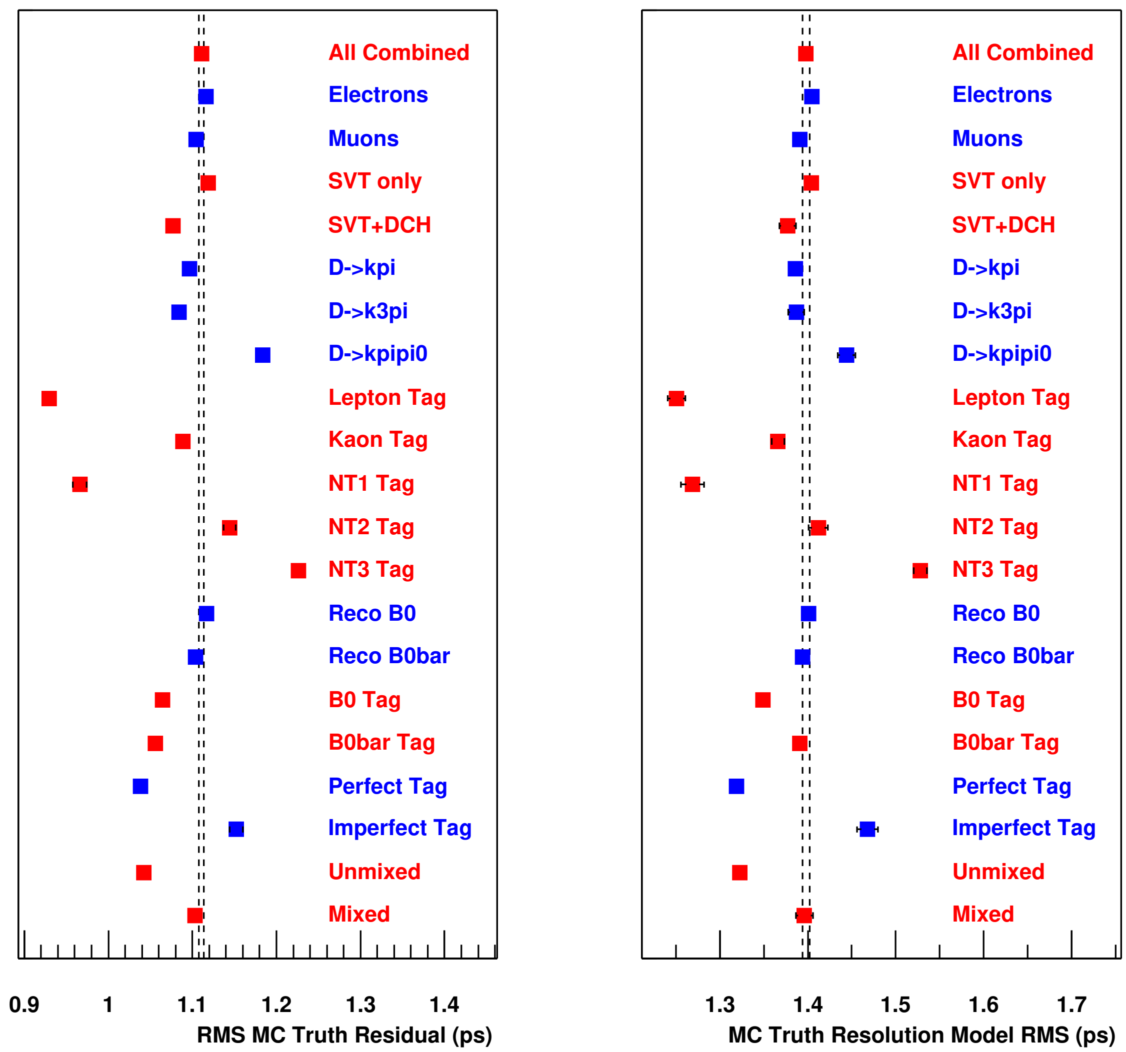

Figure 67: Comparison of RMS of residuals in the signal MC sample, calculated directly from MC truth (left-hand side) or calculated from models fit to MC truth (right-hand side).

E Combinatoric Background Composition and Parameter Breakdown

In this section, we use Monte Carlo samples to characterize the types of combinatoric backgrounds and their Δt and $\sigma_{\Delta t}$ distributions, and to determine whether the $\delta m \equiv m(\mathrm{D}^*) - m(\mathrm{D}^0)$ sideband adequately represents combinatoric background under the signal peak. We then use data in the δm sideband to determine the best resolution model.

E.1 Characterization of Combinatoric Background in Monte Carlo

In Fig. 68, we show the δm spectrum for data and for generic $\mathrm{B}^0\overline{\mathrm{B}^0}$, $\mathrm{B}^+\mathrm{B}^-$ and $\mathrm{c}\overline{\mathrm{c}}$ Monte Carlo, with correctly reconstructed $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ signal events removed from the generic $\mathrm{B}^0\overline{\mathrm{B}^0}$ MC sample except in the range 155 – 165 MeV. The relative normalization of the three generic MC components is set according to the equivalent luminosity of the three samples. The normalization of the sum of the three MC components is set so that the number of events in the range 155 to 165 MeV is equal in data and MC. The shape of the MC δm sideband is consistent with that of data.

We use Monte Carlo truth association to consider the possible sources of the combinatoric candidates. One might expect the leading source of combinatoric candidates to be misreconstructed signal. In the right hand plot of Fig. 68, we show the δm spectrum for all combinatoric backgrounds from generic $\mathrm{B}^0\overline{\mathrm{B}^0}$ Monte Carlo, and for misreconstructed $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ signal. Approximately 25% of the generic $\mathrm{B}^0\overline{\mathrm{B}^0}$ combinatoric background events are due to errors in properly reconstructing $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ signal. The other 75% of the combinatoric candidates are "cross-feed" from other decay modes that are mistakenly reconstructed as $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ signal.

Although the $\mathrm{B}^0\overline{\mathrm{B}^0}$ contribution to combinatoric candidates is relatively smooth across the δm spectrum, we are more concerned with the combination of $\mathrm{B}^+\mathrm{B}^-$, $\mathrm{c}\overline{\mathrm{c}}$, and $\mathrm{B}^0\overline{\mathrm{B}^0}$ backgrounds. In Table 74, we give the relative fraction of each source of combinatoric background in MC in the sideband region, defined as $155 < \delta m < 165$ MeV, and in the signal region, defined as $143 < \delta m < 148$ MeV. Over 60% of the events are due to $\mathrm{B}^0\overline{\mathrm{B}^0}$ events. The second largest source of combinatoric backgrounds is $\mathrm{B}^+\mathrm{B}^-$. Only a few percent of events are due to continuum $\mathrm{c}\overline{\mathrm{c}}$ events. In Fig. 69, we show how the relative fraction of each type of background varies as a function of δm. We observe that the change in $\mathrm{B}^+\mathrm{B}^-$ and $\mathrm{B}^0\overline{\mathrm{B}^0}$ relative fractions are nearly in magntidue and opposite in sign, about 0.7% / MeV.

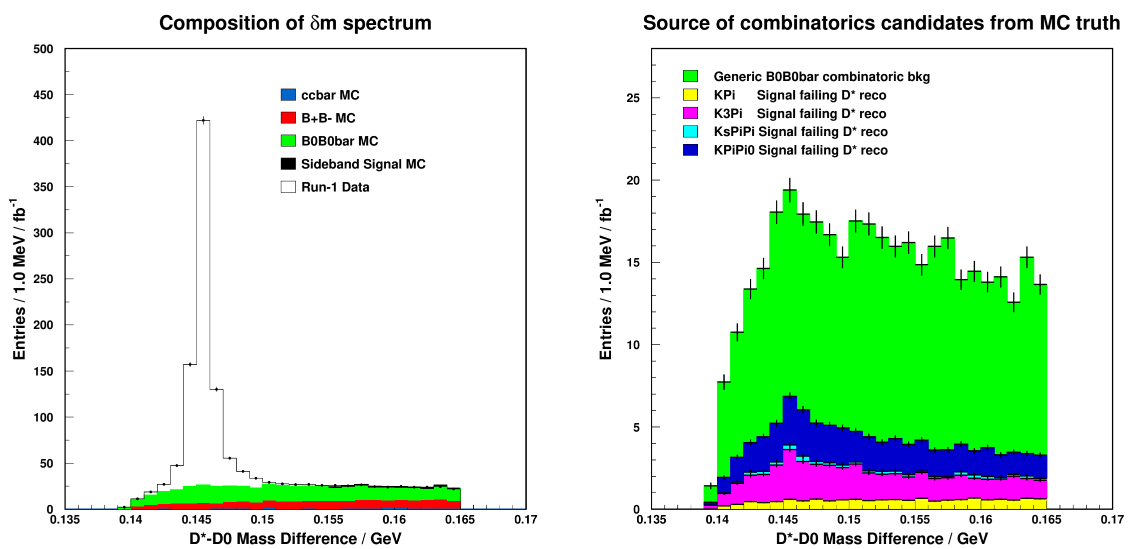

Figure 68: $m(\mathrm{D}^*) - m(\mathrm{D}^0)$ spectrum for data and for generic $\mathrm{B}^0\overline{\mathrm{B}^0}$, $\mathrm{B}^+\mathrm{B}^-$, and $\mathrm{c}\overline{\mathrm{c}}$ Monte Carlo combinatoric background. The left-hand plot compares the sideband Monte Carlo and data spectrum. Although small, the portion of correctly reconstructed signal that contributes to the sideband is included in the range 155 to 165 MeV. The normalization of the sum of the three MC components is set so that the number of events in the range 155 to 165 MeV is equal in data and MC. The right-hand figure compares the sources of combinatoric candidates using Monte Carlo truth association. The sum of the combinatoric candidates from the 4 signal modes represent only 25% of all those found in $\mathrm{B}^0\overline{\mathrm{B}^0}$ generic Monte Carlo.

Source	Fraction under peak	Fraction in sideband
$\mathrm{B}^+\mathrm{B}^-$	0.216	0.376
$\mathrm{B}^0\overline{\mathrm{B}^0}$	0.760	0.598
$\mathrm{c}\overline{\mathrm{c}}$	0.024	0.026

Table 74: Relative composition of combinatoric backgrounds in the signal region ($143 < \delta m < 148$ MeV) and in the sideband region ($155 < \delta m < 165$ MeV) using generic MC.

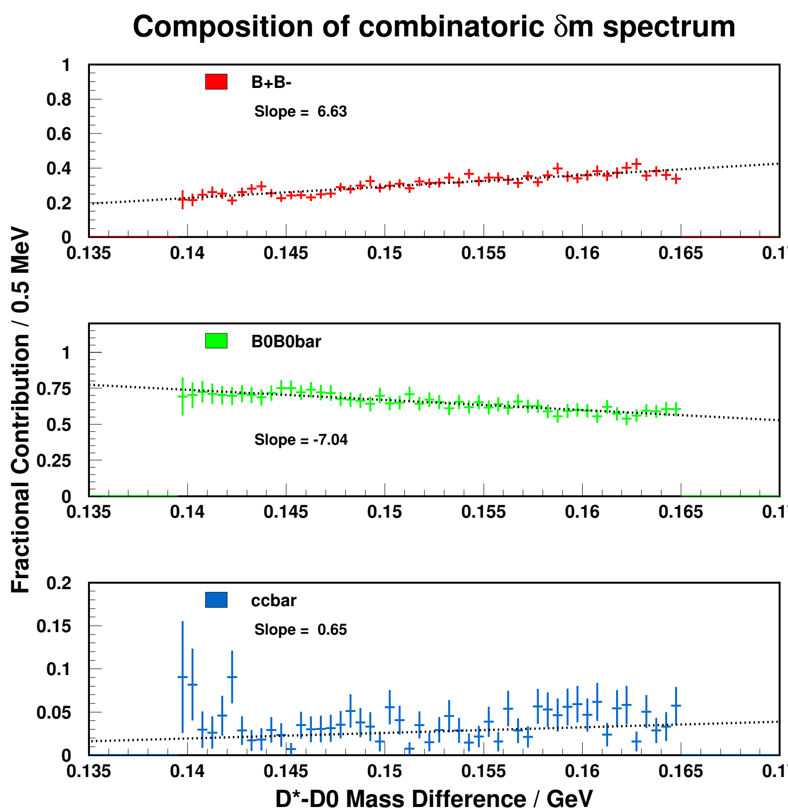

Figure 69: Relative composition of the combinatoric δm spectrum from Monte Carlo. Note that $\mathrm{B}^0\overline{\mathrm{B}^0}$ dominates everywhere, although the relative compositon between $\mathrm{B}^0\overline{\mathrm{B}^0}$ and $\mathrm{B}^+\mathrm{B}^-$ changes by 15% from the signal region to the far sideband.

Since the relative amount of $\mathrm{B^0\overline{B^0}}$ and $\mathrm{B^+B^-}$ combinatoric background changes as a function of δm (see Fig. 69), we might expect the Δt distribution to be different in the δm sideband and under the peak, if the Δt distributions are different for $\mathrm{B^0\overline{B^0}}$ and $\mathrm{B^+B^-}$ combinatoric backgrounds. Therefore, we study the Δt distributions separately for generic $\mathrm{B^0\overline{B^0}}$ and $\mathrm{B^+B^-}$ Monte Carlo events that pass all select criteria and are not associated with a true $\mathrm{B^0} \to \mathrm{D^{*-}}\ell^+\nu_\ell$. In Fig. 70, we show the Δt distributions for candidates in the δm sideband region, for combinatorial background candidates in generic MC and for data, normalized to the same area. The data and Monte Carlo shapes compare well (for these δm sideband candidates) which gives us confidence that the Monte Carlo will be accurate for combinatoric candidates in the signal region.

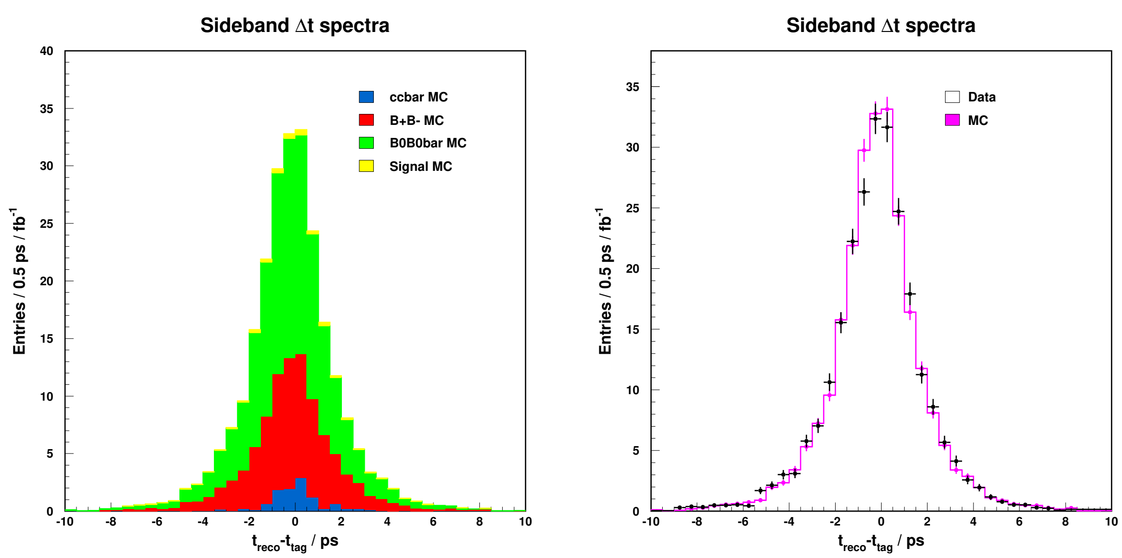

Figure 70: Distribution of Δt for candidates that pass signal selection criteria in the δm sideband region from $\mathrm{B^0\overline{B^0}}$, $\mathrm{B^+B^-}$, and $\mathrm{c\overline{c}}$ Monte Carlo. The left plot shows the composition of the MC combinatoric candidates as a function of Δt. The left-hand plot shows good shape agreement between sideband data events and the sum of the combinatorial MC contributions.

In Fig. 71, we compare the Δt distribution for MC combinatorial background events in the sideband and peak regions. The distributions are normalized to the same area. With the existing sample size, the distributions are statistically consistent.

In Fig. 72, we compare the distributions of the error on Δt ($\sigma_{\Delta t}$) for data in the δm sideband region and for MC combinatoric background events in the sideband region and the signal region. All three distributions appear statistically consistent, despite apparent poor normalization between Monte Carlo and data.

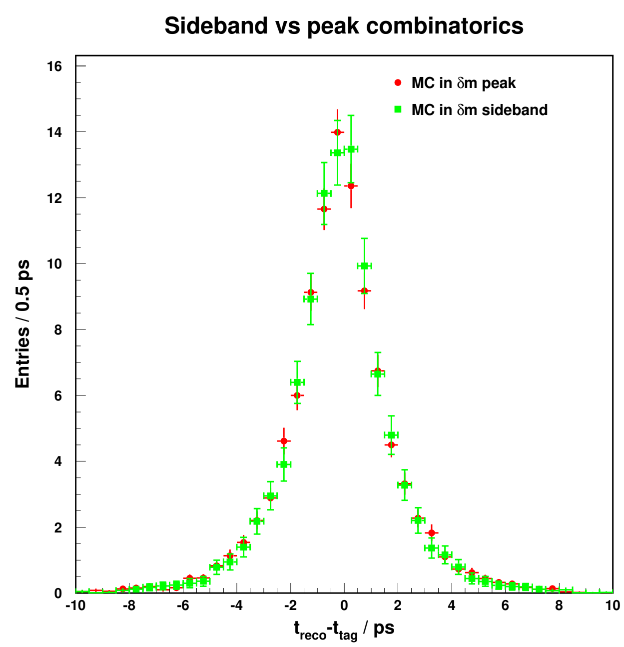

Figure 71: Comparison of Δt distributions for MC combinatorial background events in the δm peak and and the δm sideband regions. The two spectra are statistically consistent.

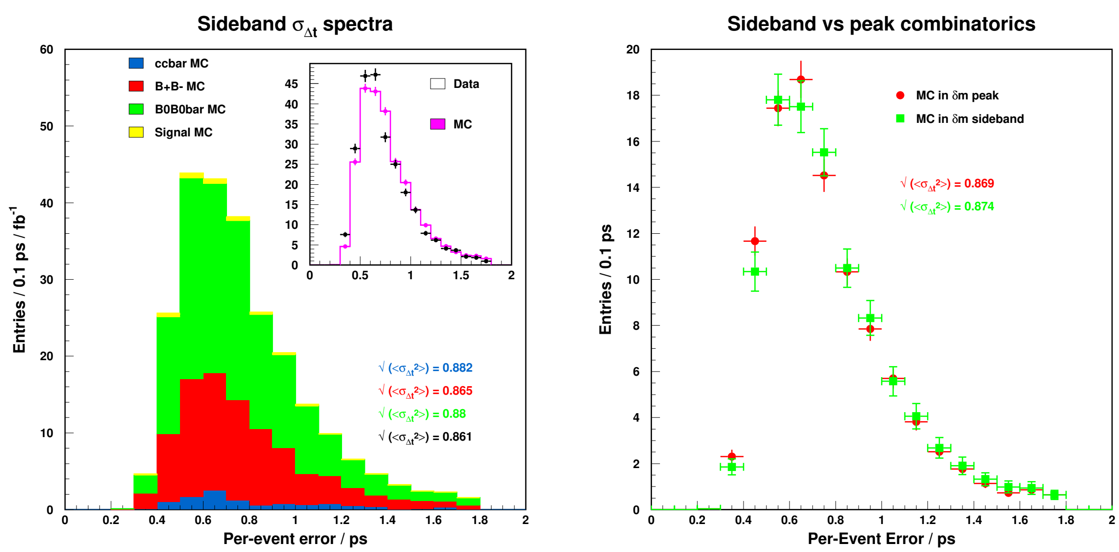

Figure 72: Distribution of the error on Δt ($\sigma_{\Delta t}$) for $\mathrm{B}^0\overline{\mathrm{B}^0}$, $\mathrm{B}^+\mathrm{B}^-$, and $\mathrm{c}\overline{\mathrm{c}}$ Monte Carlo combinatoric backgrounds in the δm sideband (left-hand plot). Distributions for data and MC in the δm sideband, normalized to the same area, are compared in the inset. Distributions for MC sideband and MC signal region, normalized to the same area, are compared in the right-hand plot.

We next compare the time dependence of the asymmetry between the number of unmixed and mixed events,

$$A(t) = \frac{N_{unmix}(t) - N_{mix}(t)}{N_{unmix}(t) + N_{mix}(t)} ,$$

for events in the δm sideband region in data and MC (Fig. 73), and for events in the sideband and signal regions in Monte Carlo (Fig. 74). Data and MC combinatoric candidates in the sideband region exhibit consistent time dependent asymmetries. The visible oscillation is due to generic $\mathrm{B^0\overline{B^0}}$ candidates (see Figure 75) that are reconstructed well enough to pass the signal selection criteria and also maintain enough information to faithfully represent their time-dependent mixing. Although the relative fraction of $\mathrm{B^0\overline{B^0}}$ and $\mathrm{B^+B^-}$ is different in the MC sideband and signal regions, the time-dependent asymmetries are statistically consistent between the two regions.

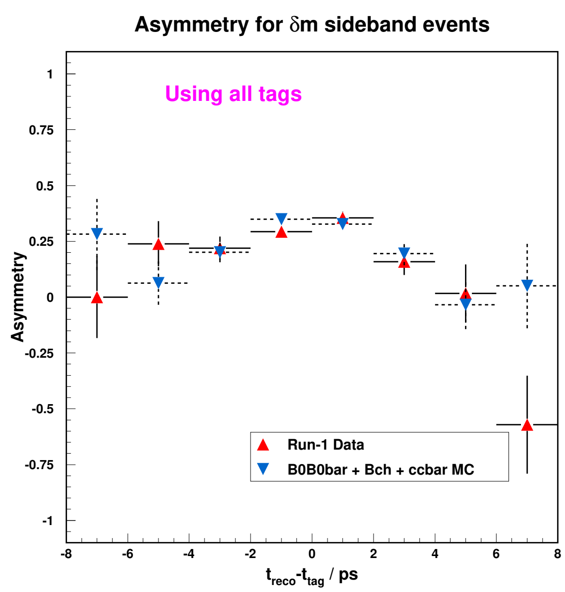

Figure 73: Binned time-dependent asymmetry between the number of unmixed and mixed events for Data candidates in the δm sideband and Monte Carlo combinatoric candidates in the sideband. They are consistent.

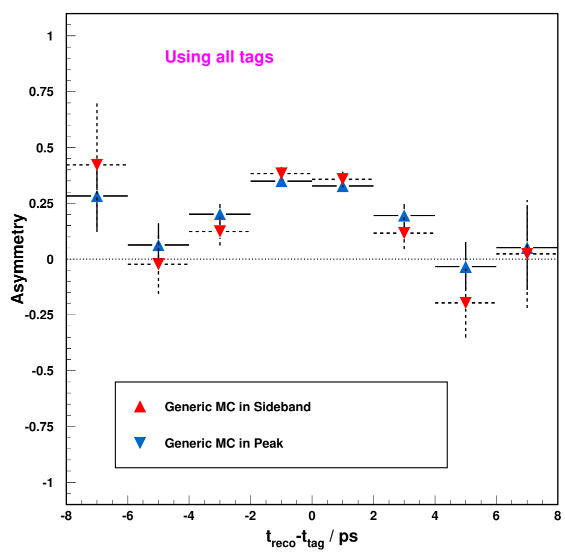

Figure 74: Comparison of time-dependent asymmetries between Monte Carlo candidates in the δm sideband region and under the signal peak. They are consistent.

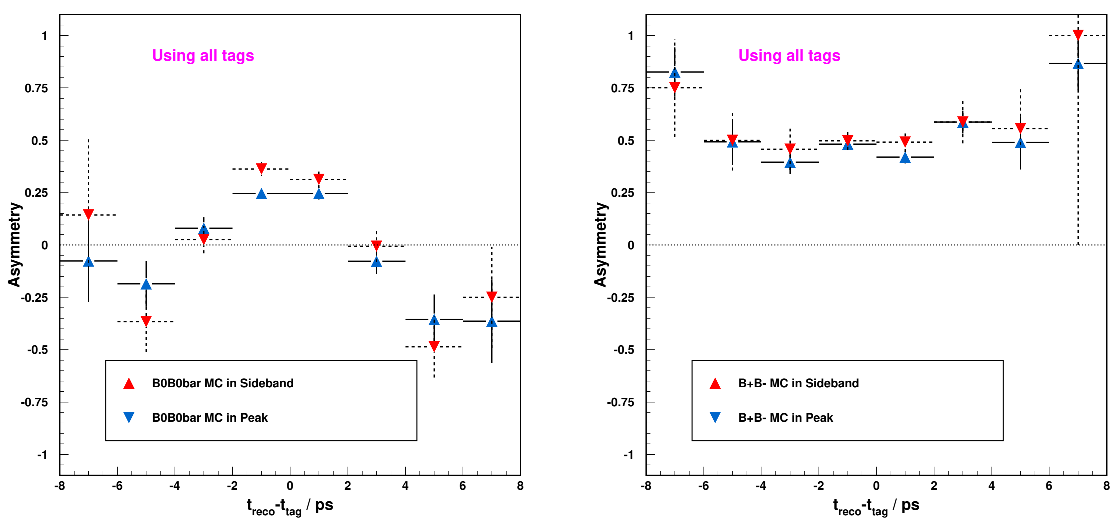

Figure 75: Comparison of time-dependent asymmetries between Monte Carlo candidates in the sideband region and under the signal peak by combinatoric source. We observe the expected oscillation in the $B^0\overline{B^0}$ candidates due to mixing, while the asymmetry for B^+B^- combinatoric candidates is mostly “flat” because we expect no mixing. A non-zero mistag fraction is responsible for the offset from 0.5.

E.2 Motivation for Combinatoric Δt Model from Monte Carlo

We now develop a model for the Δt spectrum of combinatoric background candidates, motivated by basic physics assumptions and the observations from Data and Monte Carlo discussed in the previous section. We model the reconstructed Δt distribution for combinatoric background as the sum of a component with zero lifetime ("prompt" component) and a component with a two-sided exponential decay, convoluted with a single Gaussian resolution model:

$$(f\delta(\Delta t) + (1-f)\exp(-|\Delta t|/\tau_{\mathrm{comb}})) \otimes G(\Delta t; \mu\sigma_{\Delta t}, s\sigma_{\Delta t}) \; .$$

The fit parameters are the scaled resolution bias (b), the resolution scale factor (s), the lifetime of the exponential (τ_{comb}), and the fraction of events in the zero-lifetime component (f_{pmt}). A prompt component is motivated by the observation that a significant fraction of combinatoric events combine tracks from both true B decays in forming a signal candidate, which results in a failure to separate the reconstructed vertices. Figure 76 shows the results of fits to this model for the $\mathrm{B}^0\overline{\mathrm{B}^0}$ and $\mathrm{B}^+\mathrm{B}^-$ MC samples. The fit parameters for the $\mathrm{B}^0\overline{\mathrm{B}^0}$ and $\mathrm{B}^+\mathrm{B}^-$ combinatoric backgrounds, shown in the boxes in Fig. 76, are statistically consistent with each other.

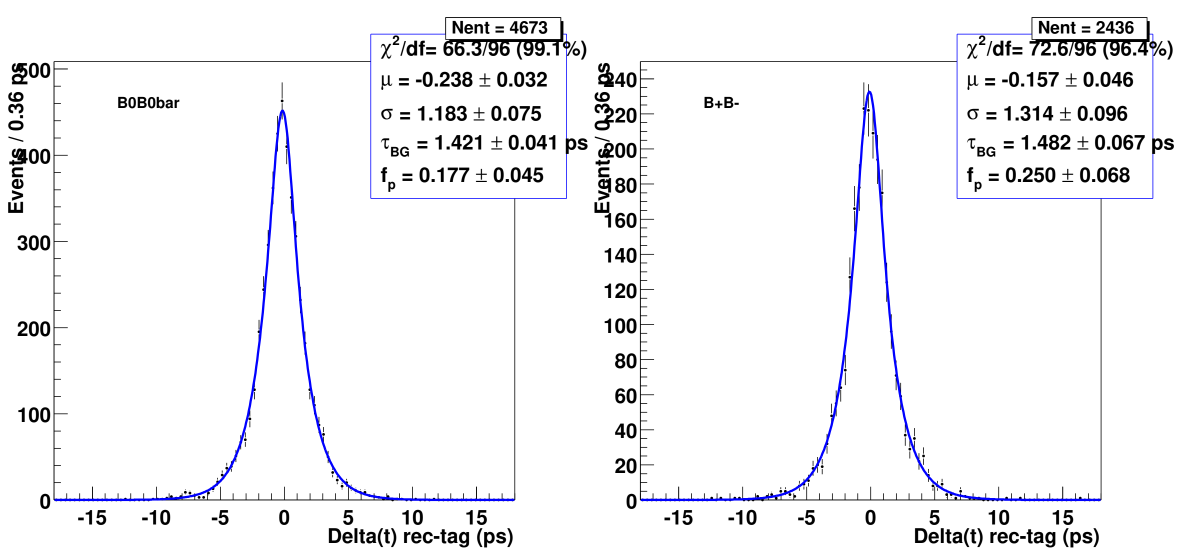

Figure 76: Reconstructed Δt distribution for $\mathrm{B}^0\overline{\mathrm{B}^0}$ (right) and $\mathrm{B}^+\mathrm{B}^-$ (left) combinatoric background events in generic MC. The curves and the parameters listed in the boxes correspond to a fit to the sum of a zero-lifetime component and a two-sided exponential distribution, convoluted with a single Gaussian resolution function. The fit parameters are the resolution bias (b), the resolution scale factor (s), the lifetime of the exponential (τ_{comb}), and the fraction of events in the zero-lifetime component (f_{pmt}).

As Figure 76 suggests, this pure lifetime Δt model does not distinguish $B^0\overline{B^0}$ candidates from B^+B^- as the fit parameters are statistically consistent with one another. A valuable cross-check is the comparison of a fit to sideband events in the combined generic MC samples and those in the data. Figure 77 illustrates this comparison, and indicates that at the precision of the existing generic Monte Carlo sample, fits using this pure lifetime Δt model are consistent between data and MC.

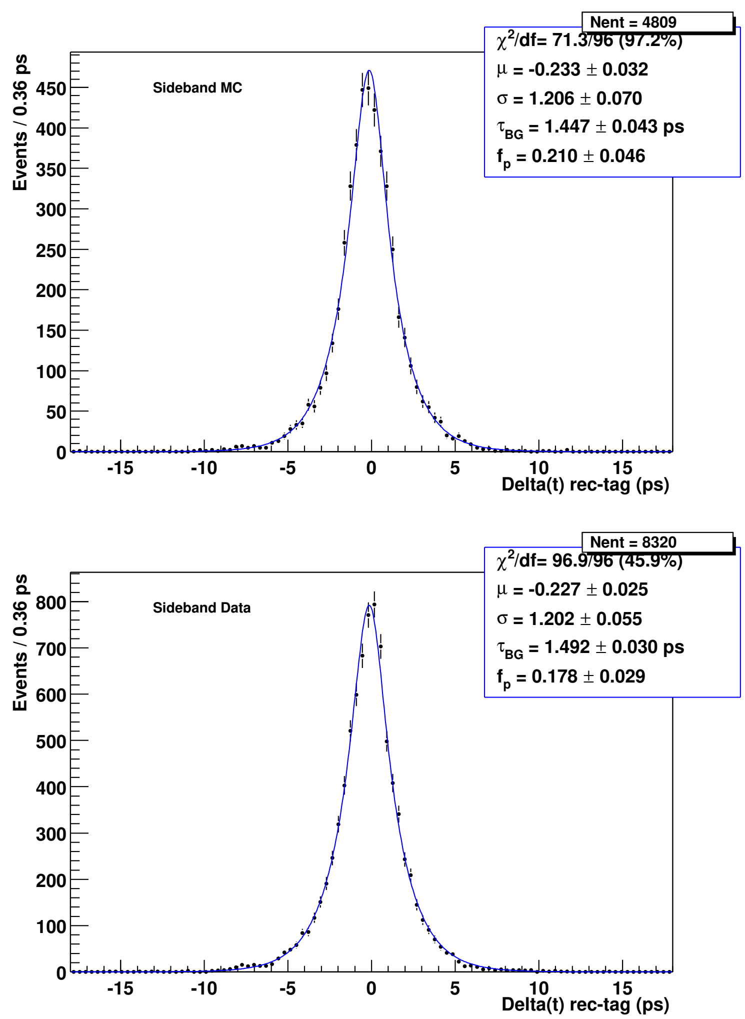

Figure 77: Comparison of a fit to the Δt spectrum of sideband events in a combined $B^0\overline{B^0}$ and B^+B^- generic MC sample to Data. The curves and the parameters listed in the boxes correspond to a fit to the sum of a zero-lifetime component and a two-sided exponential distribution, convoluted with a single Gaussian resolution function. The fits parameters are consistent.

As an additional check, we divide the events according to three intervals in δm, each 9 MeV wide, and repeat the above fit for each sample of about 1000 events. Figure 78 summarizes the results and indicates, again, that the Δt distribution of combinatoric background events does not vary significantly with the δm variable. Even with only four fit parameters, some are highly correlated, such as the scale factor s in the resolution function and the prompt fraction $f^{\rm pmt}$. We find that each parameter's dependence on δm is consistent with zero, as indicated in Figure 78.

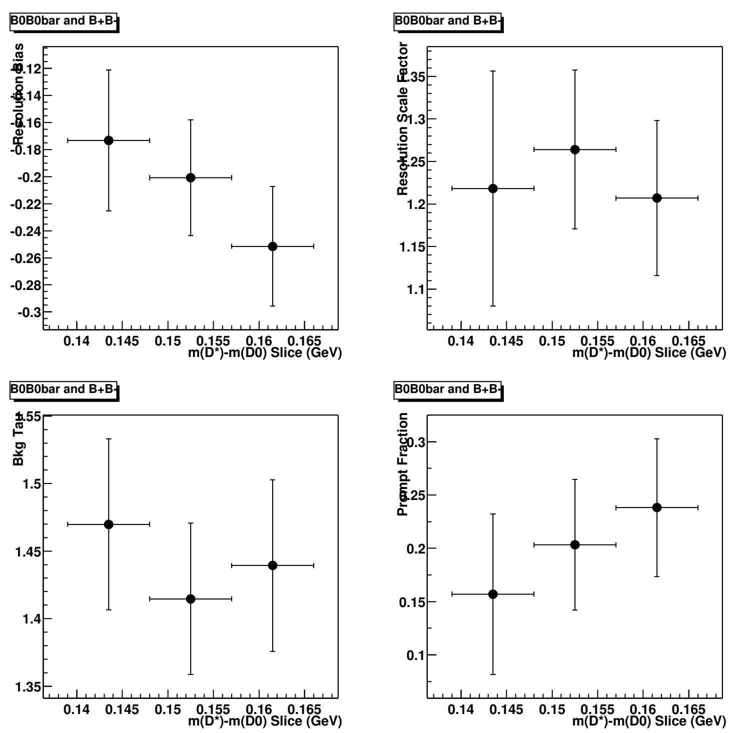

Figure 78: Results of fits to Δt distributions for combinatoric $\mathrm{B}^0\overline{\mathrm{B}^0}$ and $\mathrm{B}^+\mathrm{B}^-$ generic MC events in three slices of δm. The fit parameters are the resolution bias (b), the resolution scale factor (s), the lifetime of the exponential $(\tau_{\rm comb})$, and the fraction of events in the zero-lifetime component $(f_{\rm pmt})$. With the present statistics, the fit parameters appear consistent in each bin of δm.

Although the Δt distribution is not significantly different for $\mathrm{B}^0\overline{\mathrm{B}^0}$ and $\mathrm{B}^+\mathrm{B}^-$ combinatoric background, we do not expect this to be the case when the sample is divided according to whether the event is mixed or unmixed. There is an additional category when the other B is not tagged. In Fig. 79, we show the Δt distribution for mixed and unmixed events separately, for $\mathrm{B}^0\overline{\mathrm{B}^0}$ and $\mathrm{B}^+\mathrm{B}^-$ combinatoric background. The bottom plots show the asymmetry

$$A = \frac{N_{unmix} - N_{mix}}{N_{unmix} + N_{mix}}$$

for the two samples. As expected, there is an oscillating asymmetry in the $\mathrm{B^0\overline{B^0}}$ combinatoric background and a nearly constant asymmetry in the $\mathrm{B^+B^-}$ sample due to mistagged events. We add mixing and mistag information to the Δt fit with the following model:

$$\begin{aligned}\mathcal{G}^{(\mathrm{comb})} = & [f^{\mathrm{pmt}} \cdot \delta(\Delta t)(1 + g \cdot (1 - 2\omega_{\mathrm{pmt}})) + \\ & (1 - f)\exp(-|\Delta t|/\tau_{\mathrm{comb}})(1 + g \cdot (1 - 2\omega_{\mathrm{osc}})\cos(\Delta m_{\mathrm{comb}}\Delta t))] \otimes \\ & G(\Delta t; b\sigma_{\Delta t}, s\sigma_{\Delta t})\end{aligned} \tag{15}$$

where g takes on integer values -1, 0 and +1 for mixed, untagged and unmixed events, respectively. The Δt physics model has been expanded to include three new parameters: the mistag fractions ω_{pmt} and ω_{osc} and the mixing frequency Δm_{comb}. Figure 79 shows the results of the fit to the generic MC samples. Note that Δm_{comb} is significantly lower for the $\mathrm{B^+B^-}$ sample as expected, with a correspondingly higher mistag fraction ω_{osc}. The values of b, s, τ_{comb} and f_{pmt} are almost identical to those extracted from the fit to the combined mixed, unmixed and tagged sample, shown in Fig. 76.

The comparison to data is repeated as before to confirm that the sample of $\mathrm{B^0\overline{B^0}}$ and $\mathrm{B^+B^-}$ sideband combinatorics accurately models the structure shown in the sideband data. Figure 80 confirms this by showing consistent fit parameters between MC and data with this mixing Δt model. This conclusion reduces concerns raised by Figure 73 about the validity of using Monte Carlo to understand combinatorics in the data.

We also performed the above fit in three intervals of δm to determine whether the combinatoric "mixing" behavior is the same in the δm sideband and under the signal peak.

Fig. 81 shows the results of fits to the subset of combinatoric MC data in each δm slice. The parameters do not show any statistically significant dependence on δm, although there are still some significant correlations. Note that the value of Δm_d is limited to be greater than 0 in the fit. Fitting the $\mathrm{B^0\overline{B^0}}$ and $\mathrm{B^+B^-}$ samples separately yielded consistent results.

Our preliminary conclusion is that combinatoric background events from the δm sideband do exhibit similar Δt behavior as those in the δm signal region, at the level of present MC statistics. We anticipate the arrival and successful distillation of another 10-12 fb^{-1} within the next few weeks, which should provide additional discriminating power.

E.3 Combinatoric Δt Model Motivated by Data

In order to study the Δt distribution of combinatoric background in the data, we fit the Run-1 δm sideband data ($155 < \delta m < 165$ MeV) to the same simple model described above for Monte Carlo — a prompt component plus a two-sided exponential decay, convoluted with a single Gaussian resolution model:

$$(f\delta(\Delta t) + (1 - f)\exp(-|\Delta t|/\tau_{\mathrm{comb}})) \otimes G(\Delta t; b\sigma_{\Delta t}, s\sigma_{\Delta t}) \ .$$

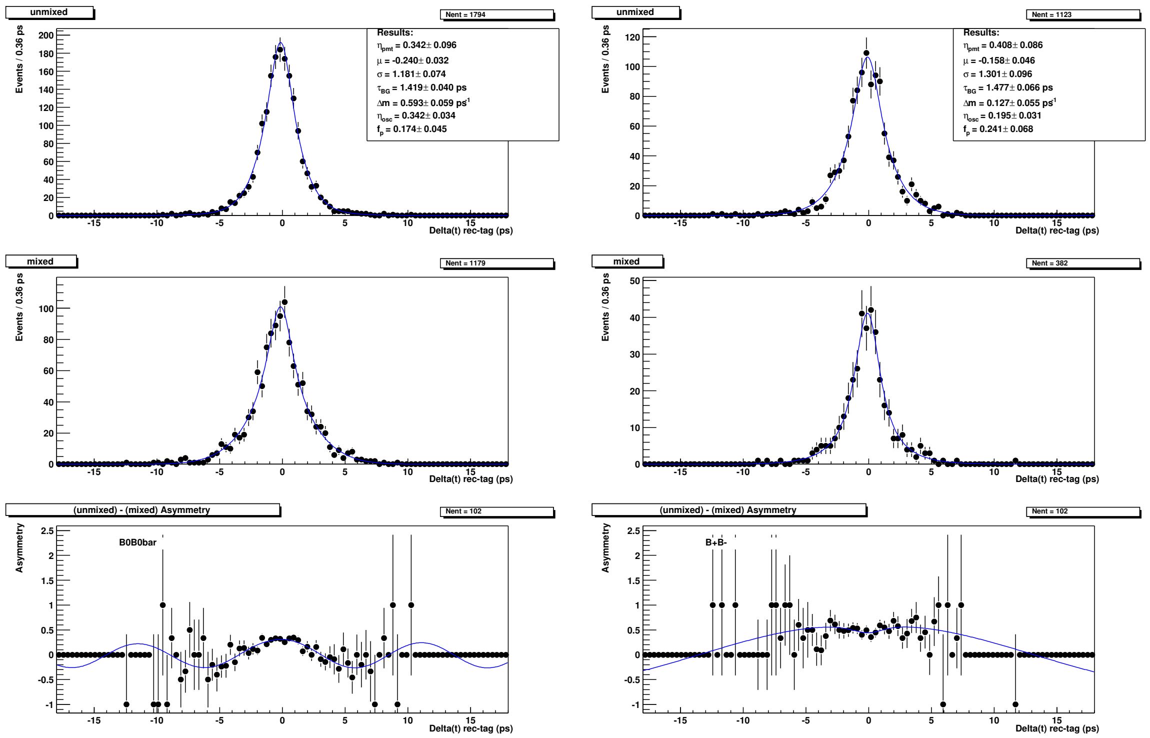

Figure 79: Distribution of Δt for unmixed (top plots) and mixed (center plots) events, and the asymmetry $A = (N_{unmix} - N_{mix})/(N_{unmix} + N_{mix})$ as a function of Δt (bottom plots), for $B^0\overline{B^0}$ (left-hand plots) and B^+B^- (right-hand plots) MC combinatoric background. The curves and the parameters in the boxes correspond to a fit to the model described in the text.

In Table 75 and Figs. 82 and 83, we show the results of fitting events in the δm sideband in data for different subsamples that might reveal different Δt structure. We use the subsamples used in the characterization of signal and background shapes and fractions, discussed in BAD 147[1] and Sec. 2, and two additional divisions according to tagging information (mixed, unmixed and untagged) and sideband ranges (155-157.5 MeV, 157.5-160 MeV, 160-162.5 MeV and 162.5-165 MeV). The most significant differences between fit parameters in different subsamples are

- the prompt fraction, resolution bias and resolution scale factor depend on whether the lepton candidate is an electron or muon, or is a fake lepton;
- the prompt fraction iis $\approx 50\%$ for on-resonance subsamples and $\approx 100\%$ for off-resonance subsamples;
- the prompt fraction is lower for the opposite-side sample compared to the same-side

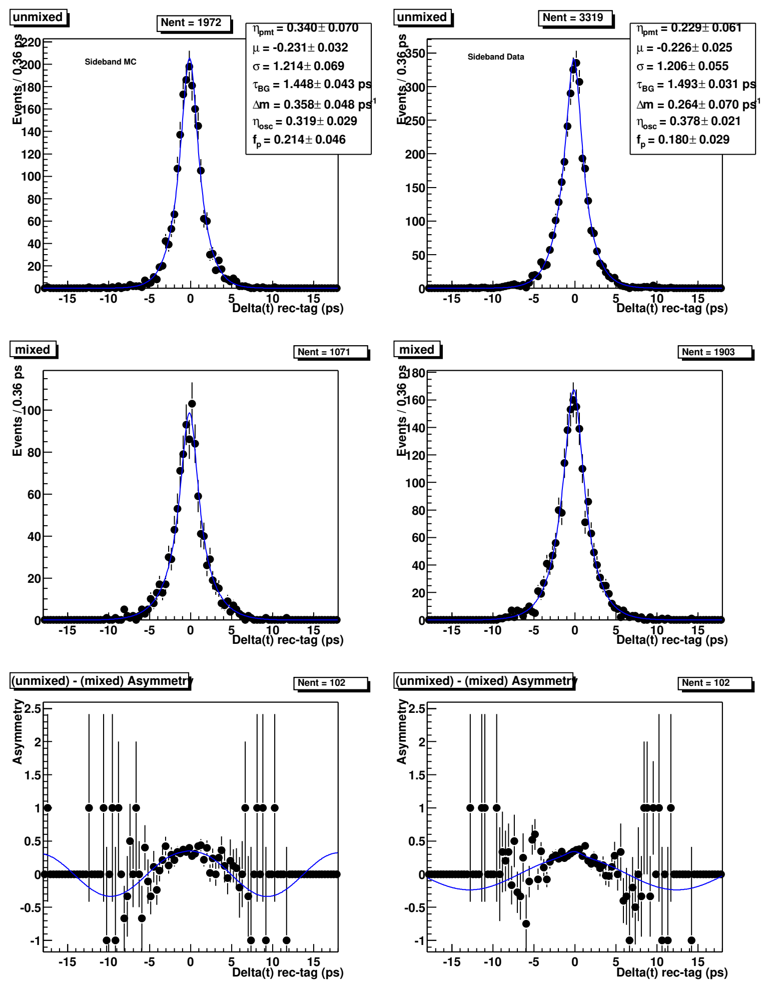

Figure 80: Comparison of a fit to the Δt spectrum of sideband events in a combined $B^0\overline{B^0}$ and B^+B^- generic MC sample to Data. The model includes an oscillating term to account for mixing. The fits parameters are consistent, despite the seemingly different shapes in the bottom asymmetry plot.

sample.

We also study the sideband data using a more complicated model that is the same as Eq. 15 with the addition of an outlier Gaussian in the resolution function:

$$\begin{aligned}\mathcal{G}^{(\mathrm{comb})} \quad = \quad & [f^{\mathrm{pmt}}\delta(\Delta t)(1+g\cdot(1-2\omega_{\mathrm{pmt}}))+ \\ & (1-f)\exp(-|\Delta t|/\tau_{\mathrm{comb}})(1+g\cdot(1-2\omega_{\mathrm{osc}})\cos(\Delta m_{\mathrm{comb}}\Delta t))]\otimes \\ & ((1-f_{\mathrm{out}})G(\Delta t; b\sigma_{\Delta t}, s\sigma_{\Delta t})+f_{\mathrm{out}}G(\Delta t; b_{\mathrm{out}}\sigma_{\Delta t}, s_{\mathrm{out}}\sigma_{\Delta t}))\end{aligned} \tag{16}$$

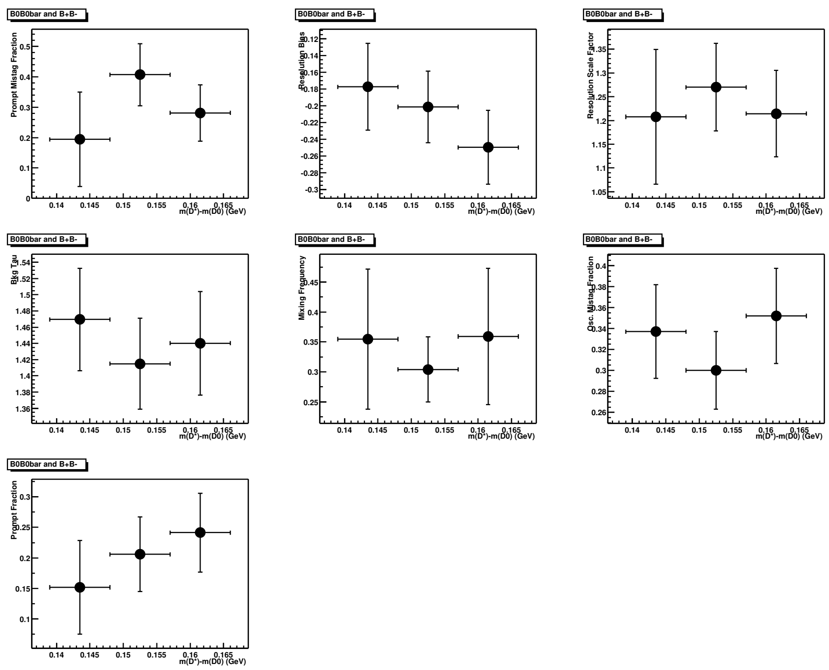

Figure 81: Results of fits to Δt distributions for combinied $B^0\overline{B^0}$ and B^+B^- combinatorics from generic MC in three intervals of δm (horizontal axis). The fit parameters are described in the text.

The total number of parameters is ten. Subsamples divided by different categories are fitted separately. The outlier Gaussian bias and width are fixed at values obtained from the fit to the whole data sample in sideband. All other eight parameters are free (Δm, $\omega_{\rm pmt}$ and $\omega_{\rm osc}$ are set to zero in the NoTag category).

The result of the fit to the whole sideband data is shown in Figure 84 and the parameters are listed in Table 76. The comparison of parameters from different subsamples are shown in Figs. 85–88.

We can take advantage of different characteristics of different subsamples by splitting up some of the ten parameters for different subsamples while keeping remaining parameters common across subsamples. By looking at results from individual fits and several attempts of simultaneous fits to all sideband samples with various ways of splitting up parameters, we come to a choice for each parameter. Table 77 shows the choices of splitting up parameters.

There are up to 49 free parameters if we break down all ten parameters. In practice, we choose a few different models, each of which uses a different set of free parameters. We then identify one of the models as our baseline model, and use the others for systematic studies. The choices of models are listed in Table 78.

Category	$\tau_{\mathrm{comb}}(ps)$	Scaled bias b	Scale factor s	f_{pmt}
unmixed	1.445 ± 0.046	-0.143 ± 0.022	1.364 ± 0.035	0.492 ± 0.031
mixed	1.624 ± 0.063	-0.184 ± 0.030	1.358 ± 0.046	0.469 ± 0.036
untagged	1.600 ± 0.060	-0.172 ± 0.022	1.364 ± 0.034	0.555 ± 0.032
SVT	1.570 ± 0.036	-0.170 ± 0.016	1.349 ± 0.024	0.498 ± 0.021
DCH	1.403 ± 0.062	-0.139 ± 0.029	1.387 ± 0.047	0.518 ± 0.041
LTag	1.60 ± 0.11	-0.120 ± 0.092	1.33 ± 0.15	0.238 ± 0.086
KTag	1.521 ± 0.051	-0.201 ± 0.023	1.395 ± 0.036	0.508 ± 0.031
NT1	1.458 ± 0.095	-0.085 ± 0.051	1.185 ± 0.089	0.414 ± 0.068
NT2	1.524 ± 0.086	-0.100 ± 0.034	1.401 ± 0.052	0.561 ± 0.047
NoTag	1.600 ± 0.060	-0.172 ± 0.022	1.364 ± 0.034	0.555 ± 0.032
B0DstlnuKpi	1.504 ± 0.070	-0.205 ± 0.040	1.233 ± 0.069	0.384 ± 0.049
B0DstlnuK3pi	1.548 ± 0.046	-0.157 ± 0.018	1.405 ± 0.027	0.560 ± 0.025
B0DstlnuKpipi0	1.602 ± 0.061	-0.176 ± 0.029	1.352 ± 0.047	0.455 ± 0.038
B0DstlnuKspipi	1.226 ± 0.085	-0.080 ± 0.050	1.01 ± 0.12	0.348 ± 0.095
elec	1.513 ± 0.050	-0.280 ± 0.038	1.178 ± 0.076	0.187 ± 0.046
muon	1.482 ± 0.052	-0.249 ± 0.037	1.233 ± 0.070	0.224 ± 0.046
fake	1.586 ± 0.050	-0.114 ± 0.016	1.402 ± 0.023	0.650 ± 0.021
onUps4S	1.531 ± 0.031	-0.165 ± 0.014	1.350 ± 0.022	0.488 ± 0.019
offUps4S	1.56 ± 0.56	-0.114 ± 0.066	1.476 ± 0.065	0.926 ± 0.049
oppositeSide	1.517 ± 0.033	-0.167 ± 0.016	1.346 ± 0.026	0.465 ± 0.021
sameSide	1.86 ± 0.13	-0.157 ± 0.028	1.451 ± 0.036	0.753 ± 0.033
SideBand1	1.557 ± 0.059	-0.181 ± 0.029	1.331 ± 0.047	0.447 ± 0.038
SideBand2	1.675 ± 0.072	-0.144 ± 0.027	1.356 ± 0.041	0.568 ± 0.034
SideBand3	1.454 ± 0.065	-0.127 ± 0.027	1.416 ± 0.041	0.534 ± 0.040
SideBand4	1.459 ± 0.057	-0.207 ± 0.027	1.314 ± 0.044	0.459 ± 0.039

Table 75: Fit parameters for $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidates in the Data δm sideband ($155 < \delta m < 165$ MeV). The Δt distribution is fit to the sum of a prompt component and an exponential function, convoluted with a single Gaussian resolution model, as described in the text. The last four rows (SideBand1 - 4) are defined as 2.5 MeV slices from 155, 157.5, 160 and 162.5 MeV, respectively. The same information is displayed graphically in Figs. 82 and 83.

The results of the baseline fit are shown in Table 79. In Figure 89, we compare the four parameters that are broken down. The Δt distribution is shown in Figure 90.

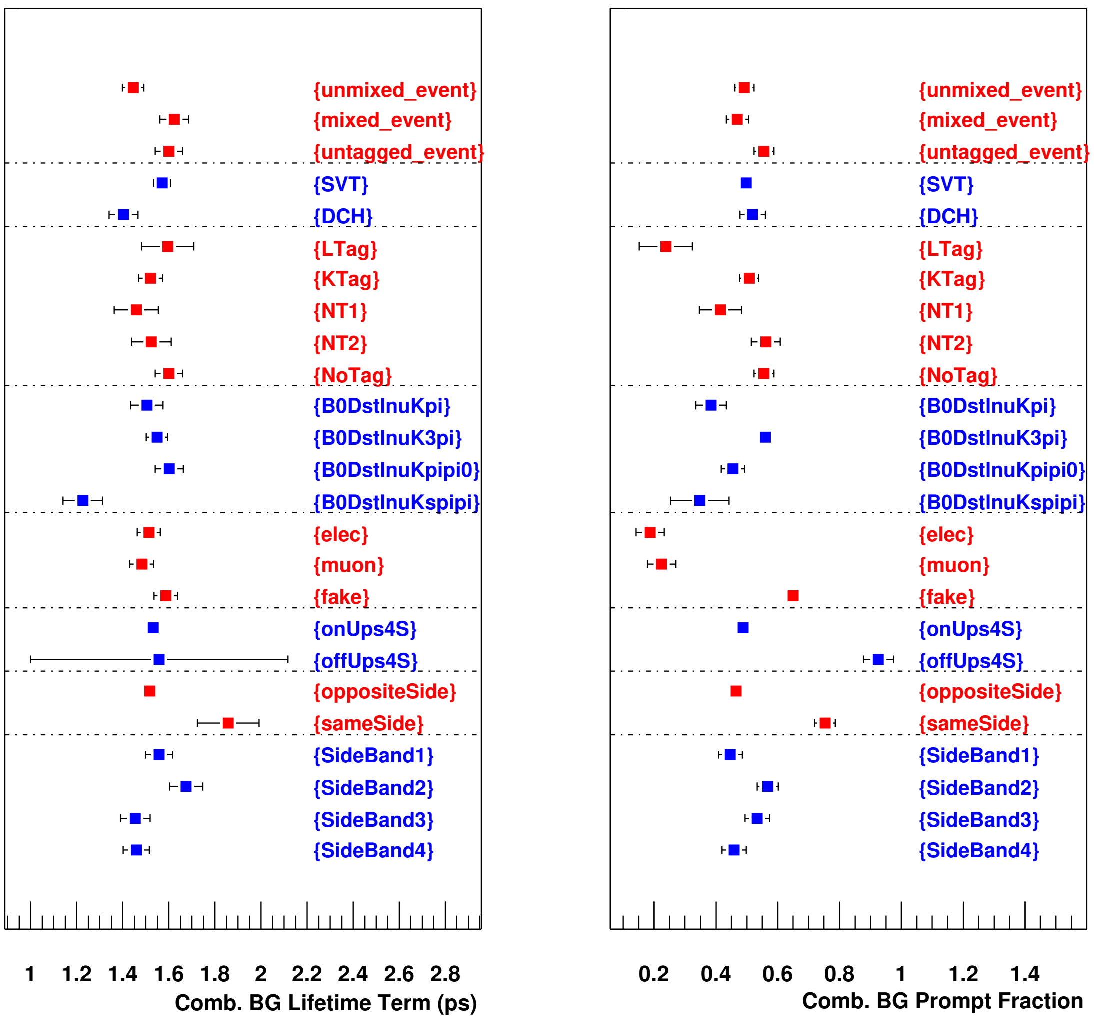

Figure 82: Parameters from Δt fit to events in δm sideband in Data. Left, $\tau_{\rm comb}$; right, $f^{\rm pmt}$. The same information is given in tabular form in Table 75.

$f^{\rm pmt}$	$\tau_{\rm comb}({\rm ps})$	$\Delta m_{\rm comb}({\rm ps}^{-1})$	$\omega_{\rm osc}$	$\omega_{\rm pmt}$
0.425 ± 0.024	1.301 ± 0.037	0.288 ± 0.043	0.343 ± 0.019	0.354 ± 0.019

b_1	s_1	$f_{\rm out}$	$b_{\rm out}$	$s_{\rm out}$
-0.158 ± 0.014	1.308 ± 0.025	0.0098 ± 0.0033	-2.7 ± 1.2	9.0 ± 1.3

Table 76: Parameters of Eq. 16 fit to the whole sideband in Data.

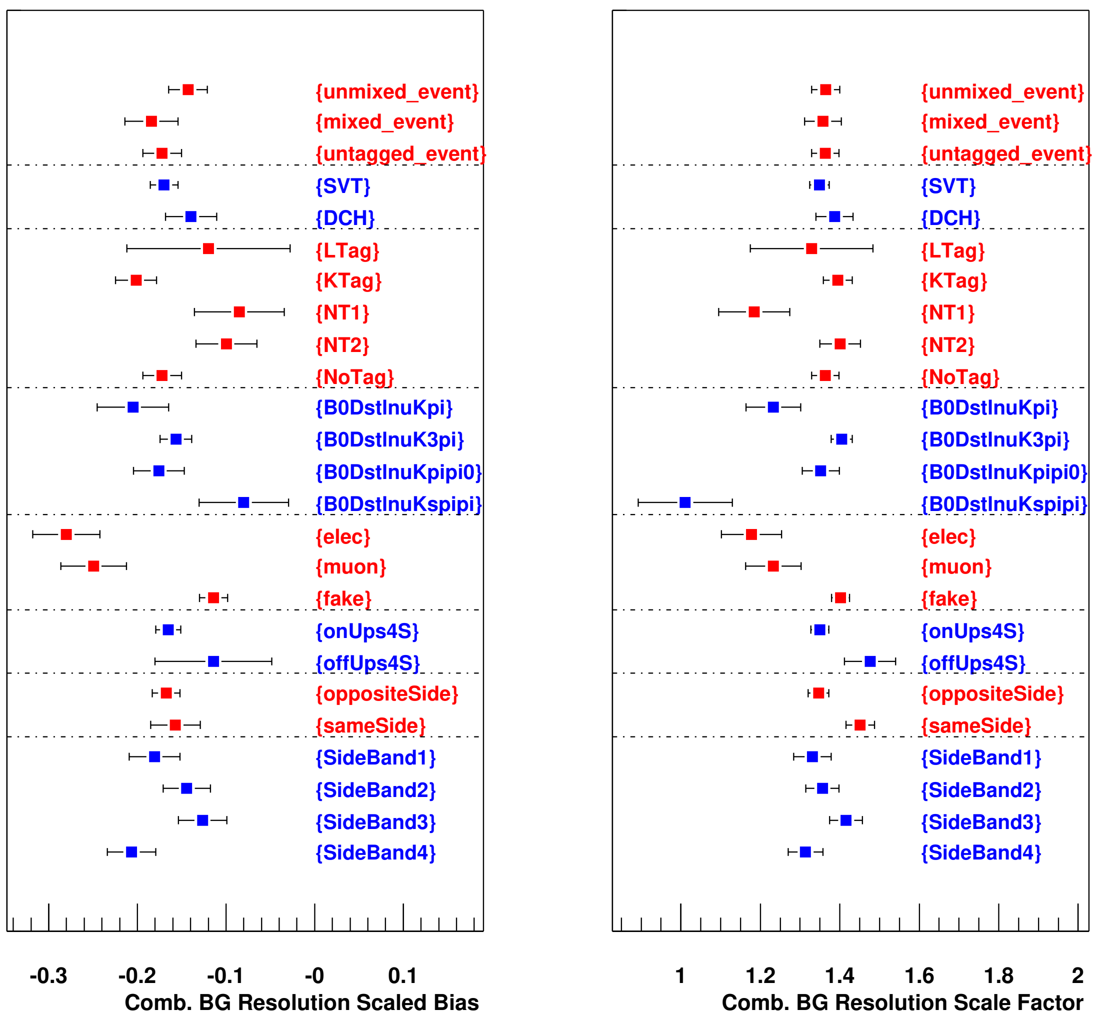

Figure 83: Parameters from Δt fit to events in δm sideband in Data. Left, resolution scaled bias; right, resolution scale factor.

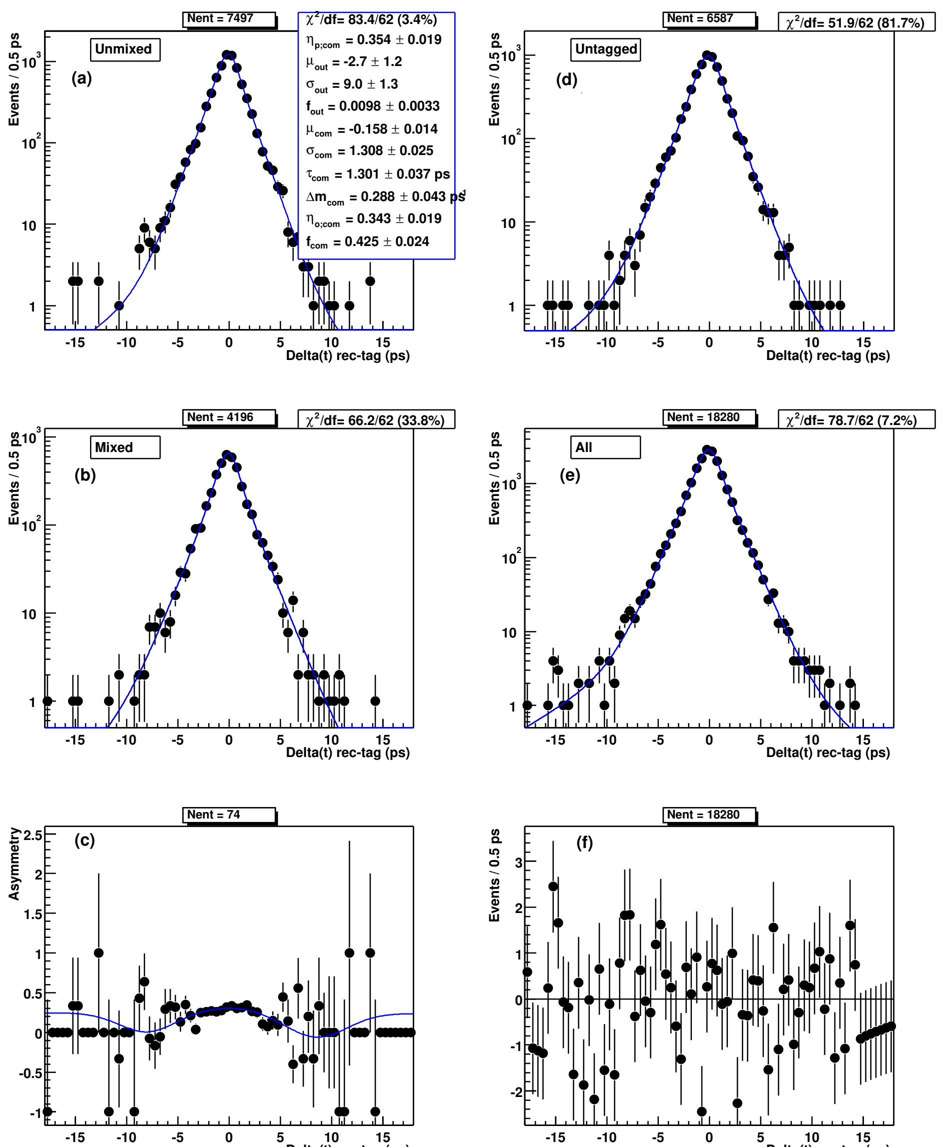

Figure 84: Δt distribution for events in δm sideband in Data. Fit results are superimposed on each plot. The fit parameters are given in the box in the first plot and in Table 76. (a) Unmixed events. (b) Mixed events. (c) Asymmetry. (d) Untagged events. (e) All events. (f) Bin-by-bin (signed) χ^2 contribution of plot (e).

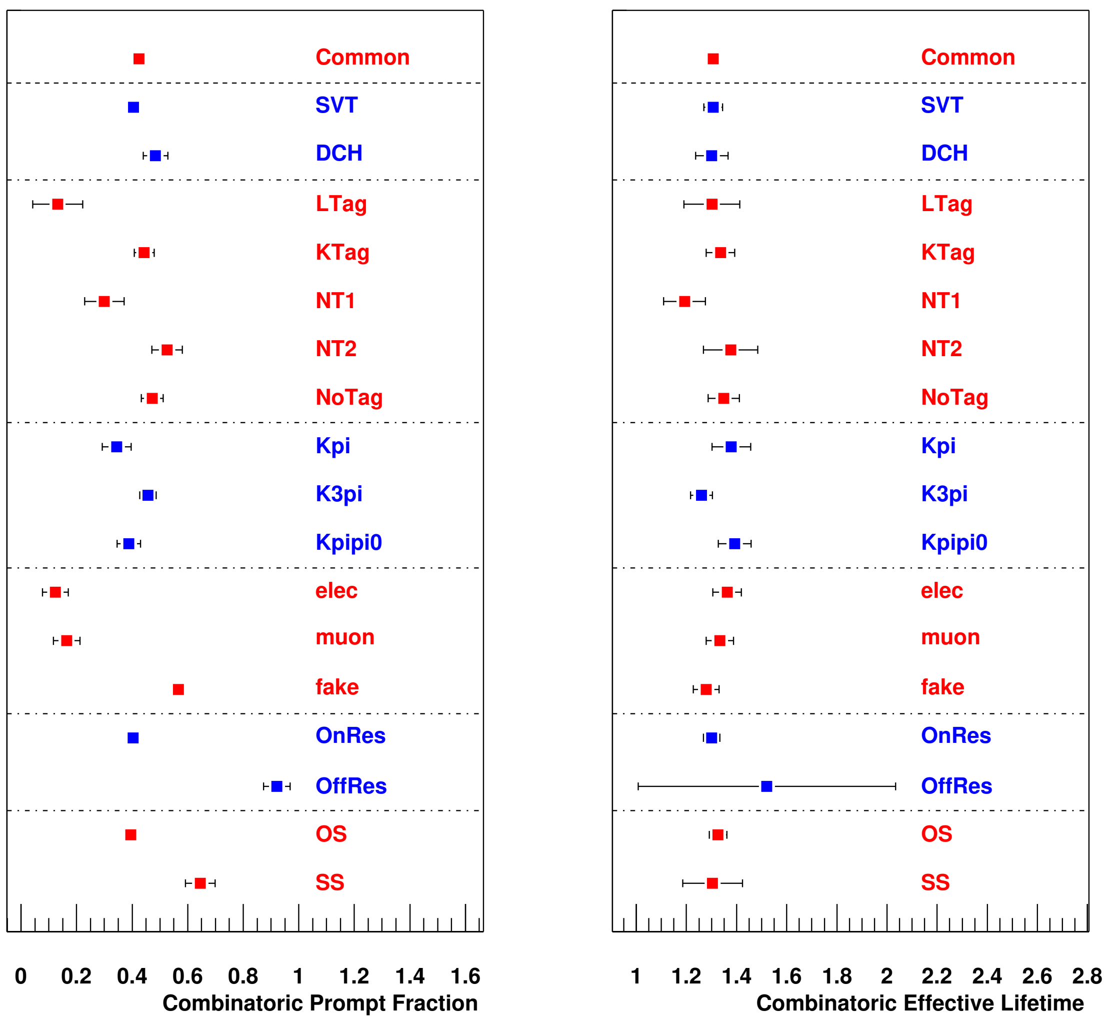

Figure 85: Fit parameters for the sideband in Data, for the mixing model. Left, prompt fraction; right, lifetime.

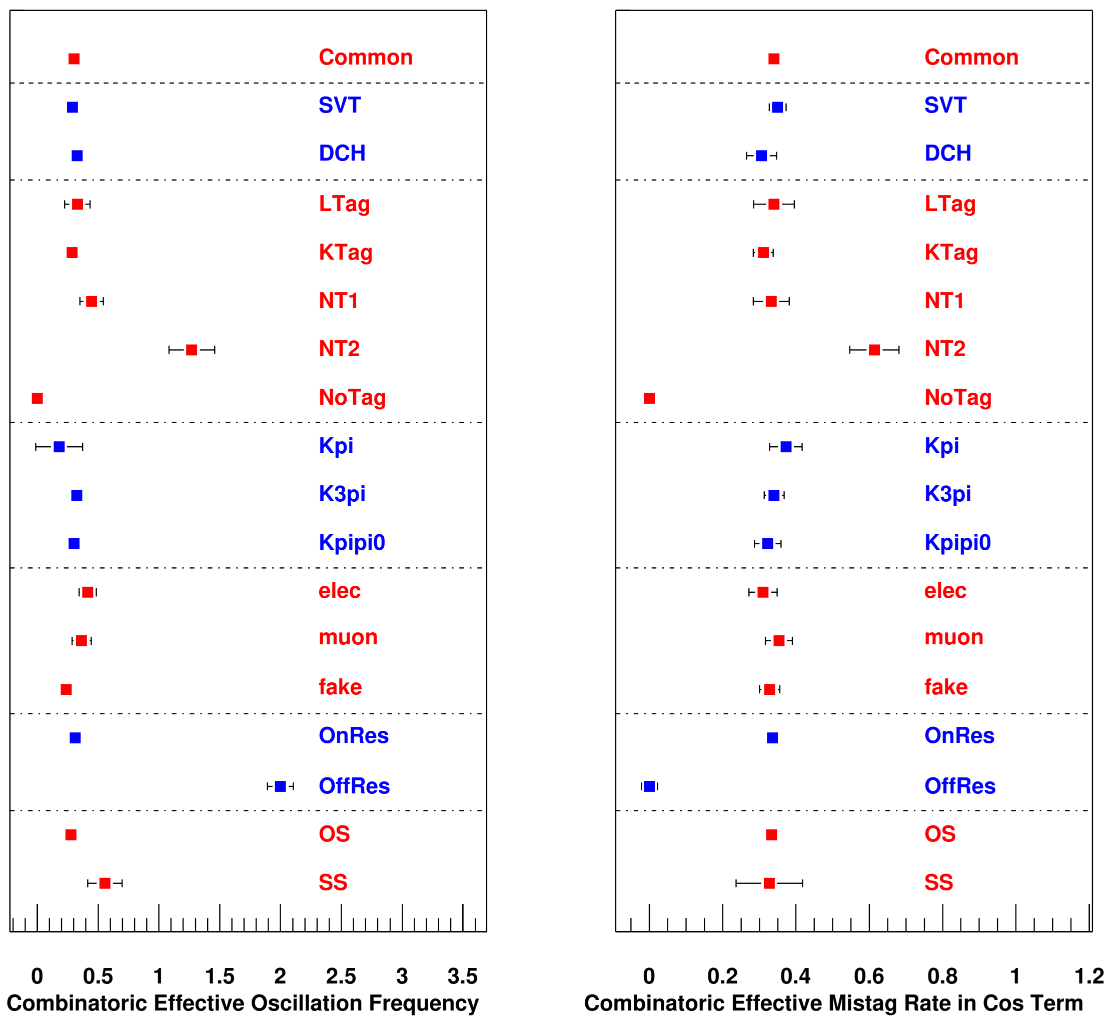

Figure 86: Fit parameters for the sideband in Data, for the mixing model. Left, effective Δm; right, effective mistag ω_o in the oscillation term.

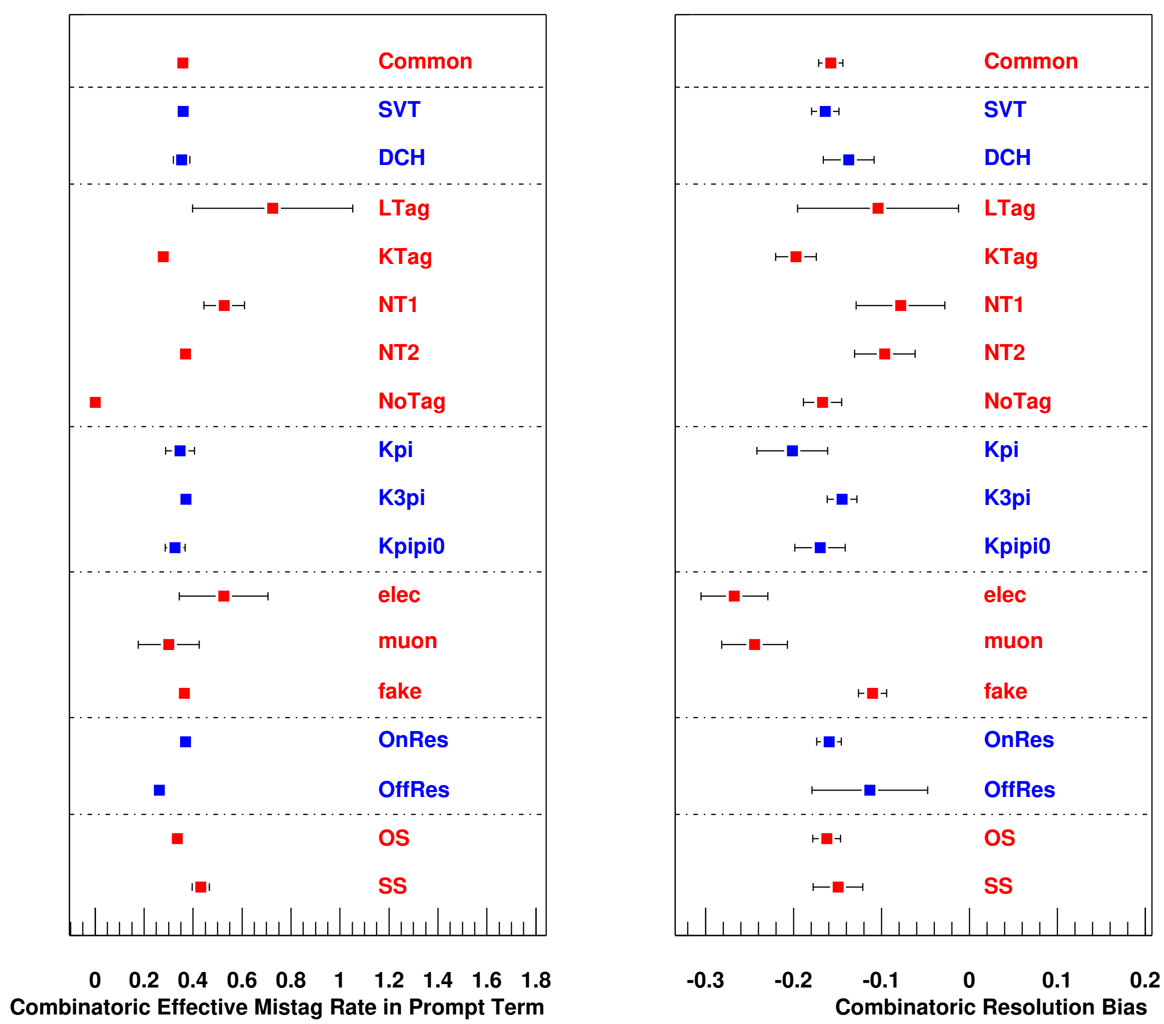

Figure 87: Fit parameters for the sideband in Data, for the mixing model. Left, effective mistag ω_p in the prompt term; right, resolution bais in core Gaussian.

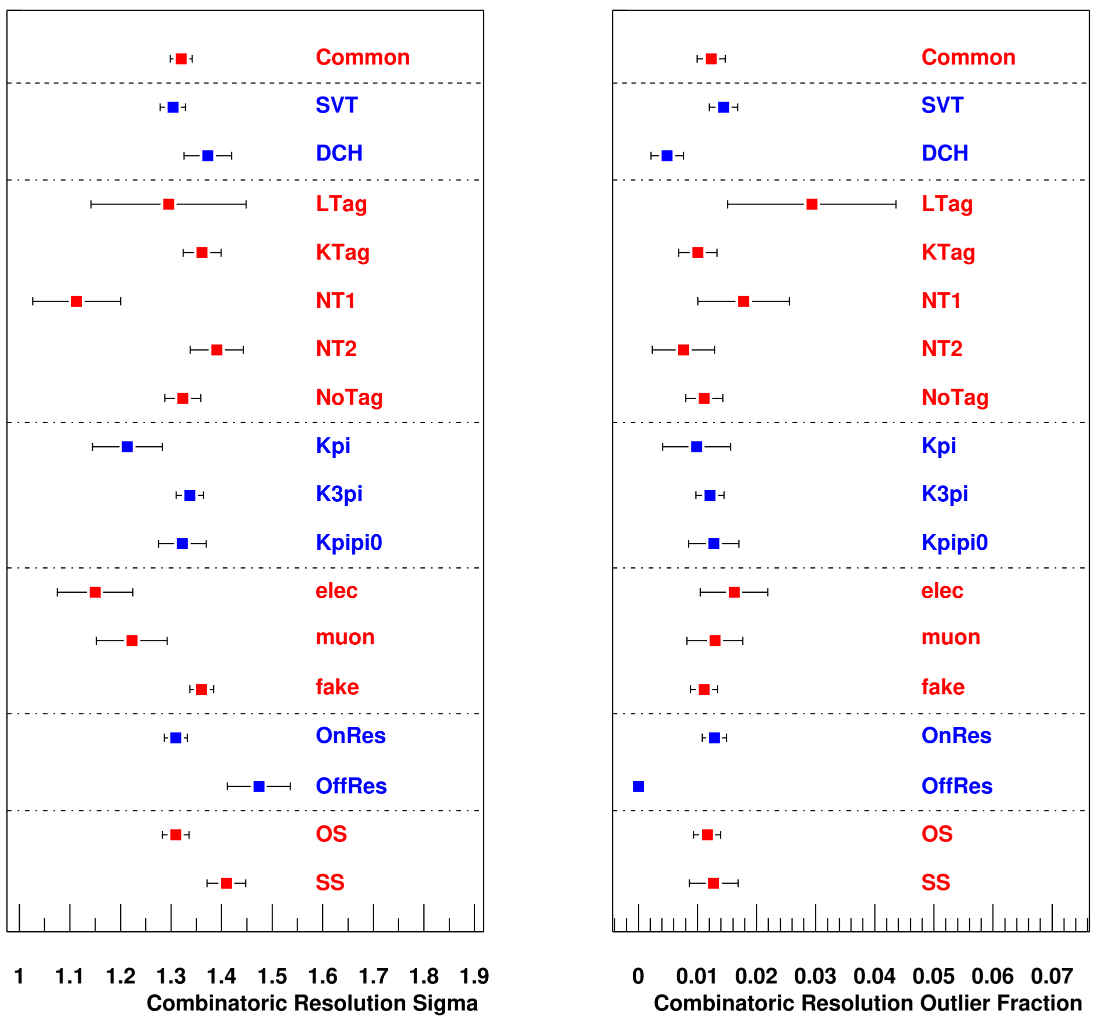

Figure 88: Fit parameters for the sideband in Data, for the mixing model. Left, resolution scale factor in core Gaussian; right, outlier Gaussian fraction.

Parameter	Group	# pars
$f_{\rm pmt}$	OffRes $\oplus$ SS $\oplus$ TwoLpet $\oplus$ OneLept $\oplus$ NoLept	5
$\tau_{\rm comb}$	SS $\oplus$ SVT $\oplus$ DCH	3
$\Delta m_{\rm comb}$	OffRes $\oplus$ SS $\oplus$ tagCatGrp$\otimes$(Lept $\oplus$Fake)	10
$\omega_{\rm osc}$	OffRes $\oplus$ SS $\oplus$ tagCatGrp$\otimes$(Lept $\oplus$Fake)	10
$\omega_{\rm pmt}$	OffRes $\oplus$ SS $\oplus$ tagCatGrp$\otimes$(Lept $\oplus$Fake)	10
b_1	(L$_{\rm or}$NT1$_{\rm or}$NT2Tag $\oplus$ K$_{\rm or}$NoTag) $\otimes$ (Lept $\oplus$ Fake)	4
s_1	(OffRes$_{\rm or}$SS) $\oplus$ Lept $\oplus$ Fake	3
$f_{\rm out}$	svtDch	2
$b_{\rm out}$	common to all	1
$s_{\rm out}$	common to all	1

Table 77: Possible choices for using different parameters for particular subsets of combinatoric background events in Δt modeling. Refer to Tables 1 and 8 of BAD147 for definitions of categories. Two/One/No Lept in the first row refer to the number of identified leptons in the event (reco-side and tag-side). For Δm, $\omega_{\rm osc}$ and $\omega_{\rm prompt}$, tagCatGrp does not include NoTag sample. The symbolic notation used in this table and other tables in this Appendix is described in the caption of Table 18.

Model	Parameters to split up	# pars
baseline	$f_{\rm pmt}$, $\omega_{\rm osc}$, b_1, s_1	28
variation-1	$f_{\rm pmt}$, $\Delta m_{\rm comb}$, b_1, s_1	28
variation-2	$f_{\rm pmt}$, $\Delta m_{\rm comb}$, $\omega_{\rm osc}$	32
variation-3	$f_{\rm pmt}$, $\Delta m_{\rm comb}$, $\omega_{\rm pmt}$	32
variation-4	$f_{\rm pmt}$, $\tau_{\rm comb}$	16
......	.....	..

Table 78: Combinatoric Δt models.

$f^{\rm pmt}_{\rm TwoLept}$	$0.00000000 \pm 0.00000033$
$f^{\rm pmt}_{\rm OneLept}$	0.145 ± 0.034
$f^{\rm pmt}_{\rm NoLept}$	0.537 ± 0.025
$f^{\rm pmt}_{\rm OffRes}$	0.893 ± 0.042
$f^{\rm pmt}_{\rm SS}$	0.634 ± 0.034
$\tau_{\rm comb}$ (ps)	1.317 ± 0.034
$\Delta m_{\rm comb}$ (ps^{-1})	0.298 ± 0.031
$\omega^{\rm osc}_{\rm LTag;Lept}$	0.204 ± 0.035
$\omega^{\rm osc}_{\rm OffRes}$	0.04 ± 0.17
$\omega^{\rm osc}_{\rm SS}$	0.521 ± 0.042
$\omega^{\rm osc}_{\rm KTag;Lept}$	0.310 ± 0.015
$\omega^{\rm osc}_{\rm NT1;Lept}$	0.388 ± 0.037
$\omega^{\rm osc}_{\rm NT2;Lept}$	0.425 ± 0.023
$\omega^{\rm osc}_{\rm NoTag}$	0
$\omega^{\rm osc}_{\rm LTag;Fake}$	0.408 ± 0.037
$\omega^{\rm osc}_{\rm KTag;Fake}$	0.174 ± 0.027
$\omega^{\rm osc}_{\rm NT1;Fake}$	0.398 ± 0.047
$\omega^{\rm osc}_{\rm NT2;Fake}$	0.533 ± 0.036
$\omega^{\rm pmt}$	0.350 ± 0.015
$b^1_{\rm LNT1NT2;Lept}$	-0.191 ± 0.058
$b^1_{\rm K-NoTag;Lept}$	-0.312 ± 0.032
$b^1_{\rm LNT1NT2;Fake}$	-0.058 ± 0.032
$b^1_{\rm K-NoTag;Fake}$	-0.125 ± 0.019
$s^1_{\rm Lept}$	1.250 ± 0.060
$s^1_{\rm Fake}$	1.317 ± 0.030
$s^1_{\rm OffSS}$	1.408 ± 0.035
$f_{\rm out}$	0.0092 ± 0.0031
$b_{\rm out}$	-2.3 ± 1.3
$s_{\rm out}$	9.1 ± 1.4

Table 79: Parameters from a simultaneous fit to sideband Data. Prompt fraction, mistag rate in cosine term and the core Gaussian bias and scale factor are broken down. The other six parameters are common for all sideband events.

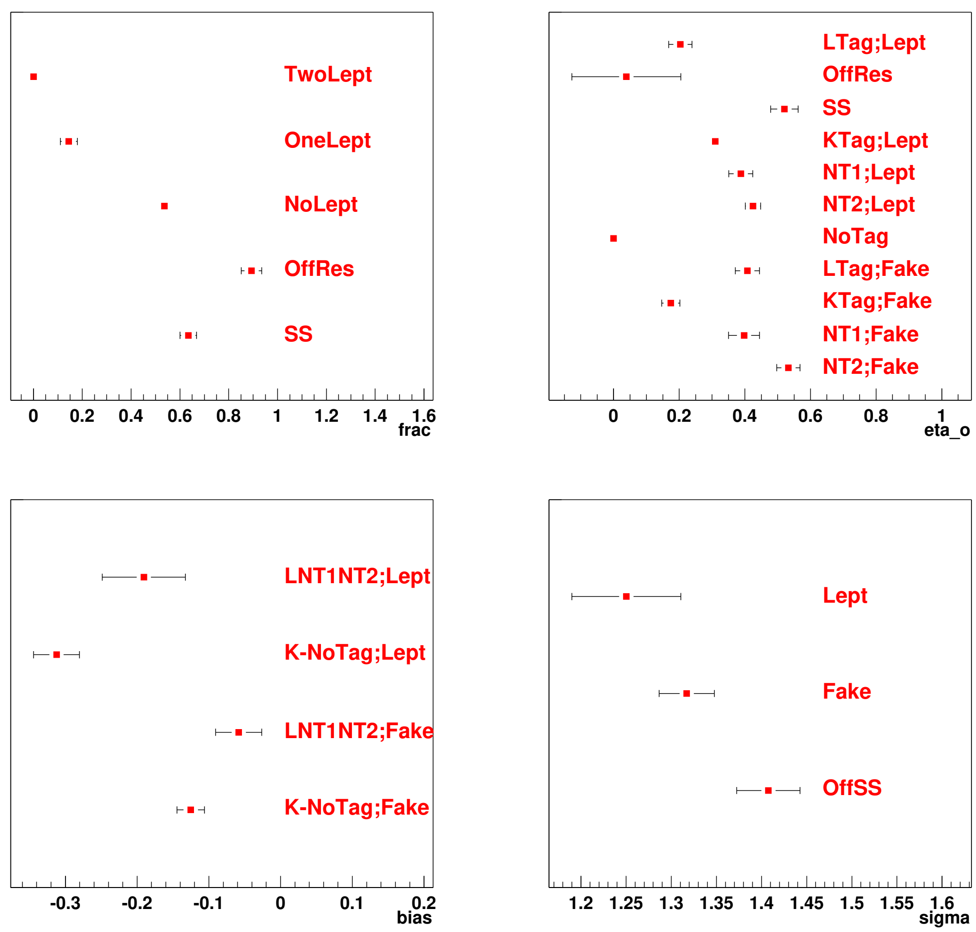

Figure 89: Comparison of parameters for sideband Data fit. Prompt fraction, mistag rate in cosine term and the core Gaussian bias and scale factor, as listed in Table 79

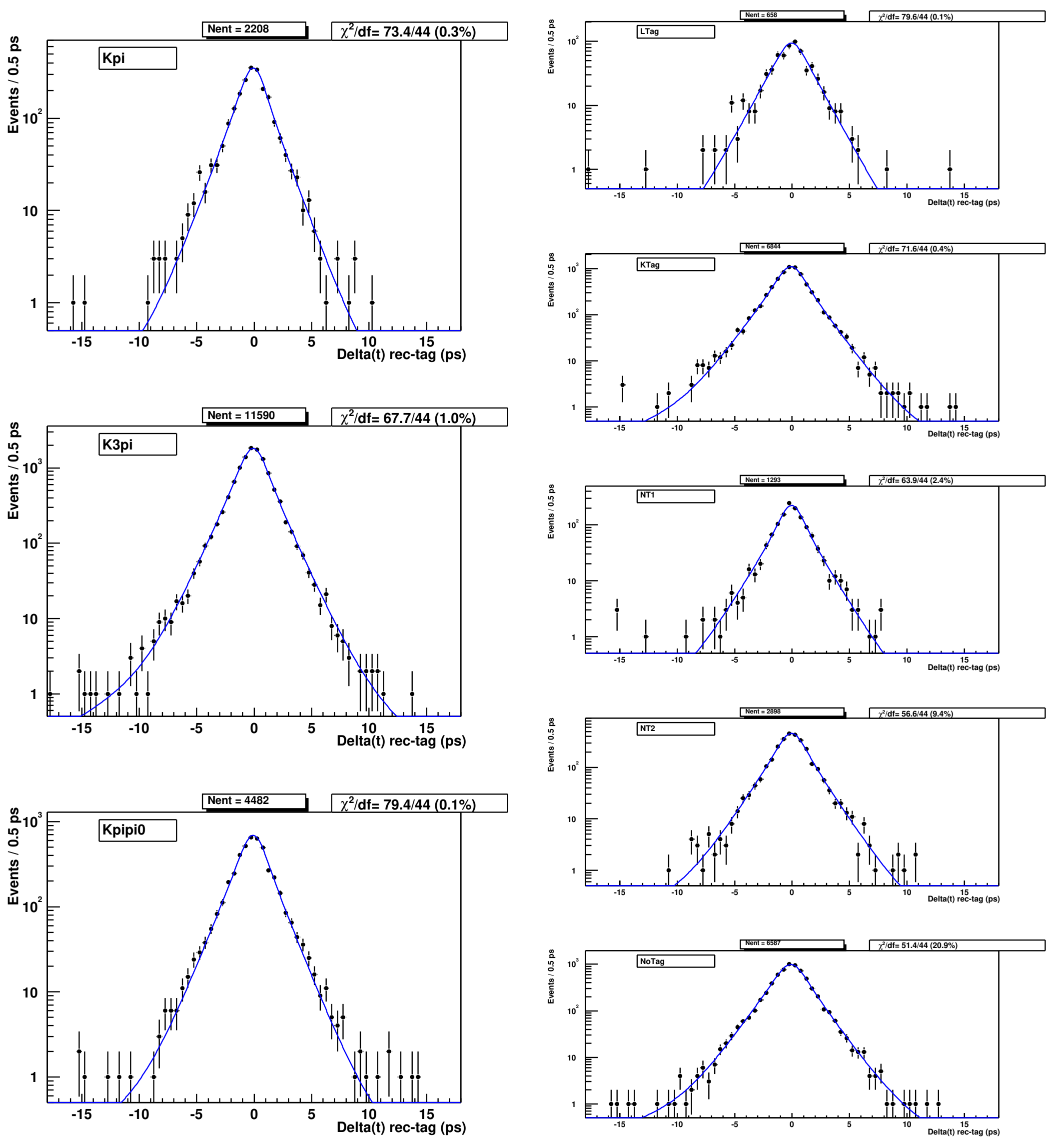

Figure 90: Sideband Data fits with parameters shown in Table 79. Data shown here are split either by D^0 decay mode (left) or by tagging categories (right).

F $B^0 \to D^{*-}\ell^+\nu_\ell$ Vertex Constraints

Although the resolution on the z-separation of the two B vertices is dominated by the tag-side vertex, the reconstructed $B^0 \to D^{*-}\ell^+\nu_\ell$ candidate does contribute. This analysis uses a novel constrained vertex fitting technique that treats the ℓ, beamspot, D^* and its daughters simultaneously to find the best vertex. The algorithm explicitly incorporates a flight distance of the daughter D^0, correctly considers the correlated errors in the fit, and provides an improved estimate of $m(D^*) - m(D^0)$ and the $D^*\ell$ vertex of the B. (The source code for this strategy is primarily located in the **VtxFitter/VtxB0_DstarlnuAlgorithm** class.)

We show here that including the beamspot constraint in the simultaneous fit for the $D*\ell\nu$ vertex improves the average resolution, and directly contributes an improvement in the Δt residuals, while introducing negligible correlations between the reco- and tag-side vertex estimates. For this study, we use $D^0 \to K\pi$ signal Monte Carlo and four different reco-side vertexing algorithms.

- (raw) - The basic **GeoKin** vertexing of the D^* and ℓ candidates as used in **CompositionTools**. No explicit treatment of the daughter D^0 flight length.
- (D^0, ℓ) - Employs the **VtxB0_DstarlnuAlgorithm** which forces a separation between the (D^*,ℓ) and the D^0 decay vertex in the simultaneous refitting of the entire D^* tree and the ℓ.
- (D^0, bs) - Same as above, except that the lepton is not used in the vertex fit; rather the beamspot is applied as constraint on the D^* decay vertex.
- (D^0, ℓ, bs) - Same as above, except the lepton candidate is also considered in the vertex fit.

Naively, we expect the most constrained refitting, the (D^0, ℓ, bs) to perform the best on average. Because these algorithms adjust the D^* 4-momentum and position, we also might expect to see a dependence of the the $\delta m = m(D^*) - m(D^0)$ mass difference resolution on the the choice of algorithm.

The technical implementation of this study involved producing new ntuples from the SP3 $D^0 \to K\pi$ signal Monte Carlo kanga files which contained four different estimates (corresponding to the four different reco-side vertexng algorithms) of each candidate's refitted δm and $z_{(D^0,\ell)}$ and corresponding error estimates. To maintain the integrity of the study, we only compare candidates that pass the ascii file cuts with all four estimators. This is because of the slight difference observed in selection efficiency as shown in Table 80. (It is primarily the $\sigma_{\Delta t}$ cut and convergence of the refitting that causes this difference.)

The initial expectation of a correlation between the δm resolution, the reco-side vertex algorithm, and angular features of the candidates is not supported in the dataset due to challenging correlations. (We define the the resolution of δm for an algorithm to be the

Reco Vtx Algorithm	Candidates Selected for ascii file	% Change
raw	84780	0.0
(D^*, ℓ)	83347	-1.7
(D^*, bs)	81810	-3.5
(D^*, ℓ, bs)	85433	+0.8

Table 80: Crude comparison of the efficiency of the different reco-side vertexing algorithms for signal MC events

combined width of a double gaussian fit to the peak of the mass difference distribution.) For instance, one might expect the power of the (D^*, bs) algorithm to depend on the azimuthal angle of the soft pion (from the D^* decay) so that when the soft pion is in the azimuthally "horizontal direction" it will be most affected by the tight y-constraint imposed by the beamspot. (Alternatively, we might think that the resolution (improvement with respect to the (raw) algorithm) of the (D^*, ℓ) algorihm might depend on the lab opening angle between the D^* and ℓ candidates. Due to a variety of correlations, however, these expectations are not borne out in data (for instance, the momentum of the soft pion appears to be empirically correlated with the CMS opening angle of the D^*-ℓ candidate due to shared kinematics, etc.

We conclude that the three refitting algorithms improve the resolution of the δm estimate, but that based on this alone, we cannot distinguish between these three. (See [1] for comments on the improved δm resolution due to just the refitting.) Figures 91- 94 illustrate our results.

We ultimately wish to choose the the reco-side vertexing algorithm that optimizes our Δt resolution, and in that dimension, we can draw distinctions between the three refitting algoirhtms. For instance, the distribution of per-event estimates for this sample is different for each of the algorithms (see Figure 93). We now focus only on correctly reconstructed signal candidates in this Monte Carlo sample, and again the (D^*, ℓ, bs) algorithm is most efficient. We compare the measured z-position of the reco-side vertex to that from Monte Carlo truth association to form a z-residual on just the reco-side. This distribution ultimately describes the Δt performance of the different algorithms. Figure 94 show these residual distributions, and the combined (D^*, ℓ, bs) algorithm is singled out as the most accurate and most precise estimate of the z-position of the D^*-ℓ candidate. Pull distributions are equivalent, although harder to distinguish between algorithms because each returns a very good estimate of its error.

An important consideration is whether the use of the beamspot constraint separately to the reco-side vertex introduces an undesirable correlation with the tag-side vertex (which also uses the beamspot in the inclusive vertex fit). To evaluate this, we measure the correlation coefficient between measured $z_{D^*-\ell}$ and z_{tag} and compute the combined $\sigma_{\delta z}$ including the individual vertex resolutions. For the sample in question, we obtain the

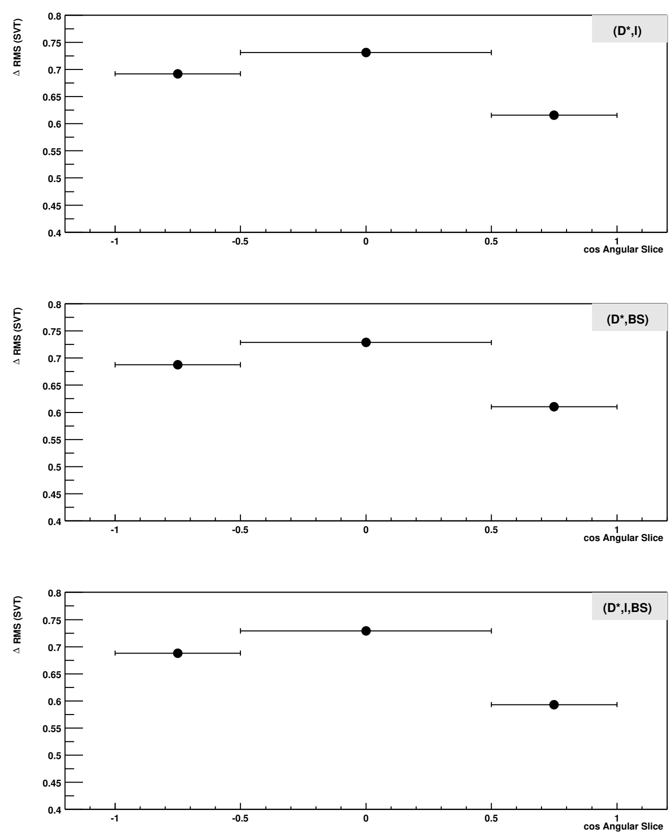

Figure 91: Each of these three figures shows the improvement in δm resolution for the reco-side vertexing algorithm shown in the upper right corner as compared to the (raw) algorithm. (The sample of candidates in each plot is identical, except for the reco-side vertex quantities, and is drawn from signal MC.) The x-axis indexes the slice in angular region of the lab quantity, $\cos\theta_{\mathrm{D}^*,\ell}$: {[-1,-0.5],(-0.5,0.5],(0.5,1],}. If indeed the resolution of the (D^*, ℓ) vertexing algorithm were dependent on the lab opening angle of the D^*-ℓ candidate, the upper plot would show the largest improvement in the center bin, while the other two plots would exhibit little structure.

results shown in Table 81, which indicate that the change in correlation between the two vertices is negligible compared to the improvement in δz resolution. (In this sample, we measure the average resolution of the z_{tag} estimate to be 130.9 μm.)

A final cross-check uses the residual distribution in a fit to determine a triple gaussian resolution function. As expected, the (D^*, ℓ, bs) algorithm returns the best RMS resolution on Δt of 0.746 ps, a 33% improvement over the (raw) algorithm. In conclusion, we select the multiply constrained simultaneous refitting algorithm (D^*, ℓ, bs) for measuring the z-

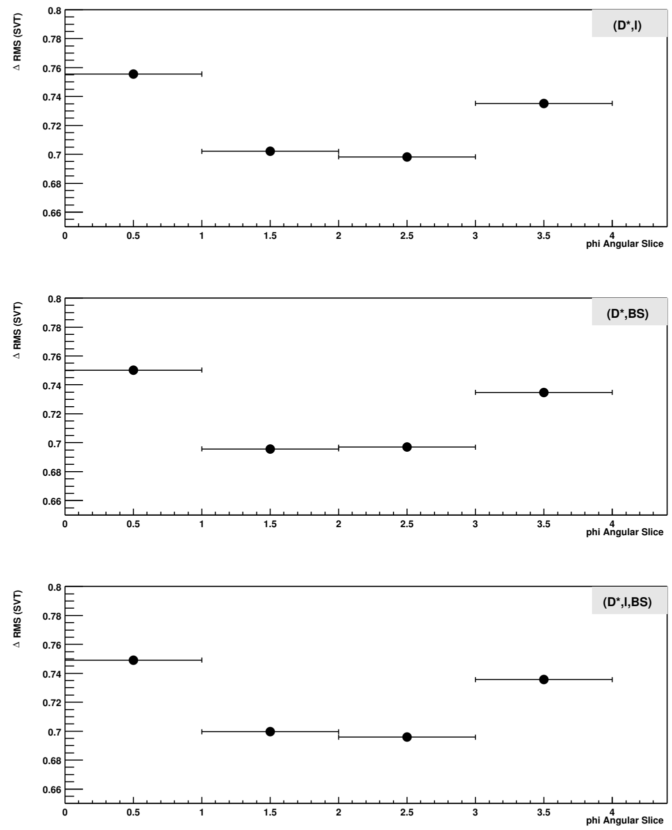

Figure 92: Each of these three figures shows the improvement in δm resolution for the reco-side vertexing algorithm shown in the upper right corner as compared to the (raw) algorithm. (The sample of candidates in each plot is identical, except for the reco-side vertex quantities, and is drawn from signal MC.) The x-axis indexes the slice in angular region of the lab quantity, $\phi_{\pi_{soft}}$: $\{[-\pi/2,\pi/2),[\pi/2,3\pi/2),[3\pi/2,-3\pi/2), [-3\pi/2,-\pi/2)\}$ or more simply {right, up, left, down}. If indeed the resolution of the (D^*, bs) vertexing algorithm were dependent on the azimuthal angle of the D*-ℓ candidate's soft pion, the middle plot would show the largest improvement in the first and third bins, while the other two plots would exhibit little structure.

vertex of the $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidates since it offers the best resolution with negligible change in the correlation between the tag- and reco-side vertices.

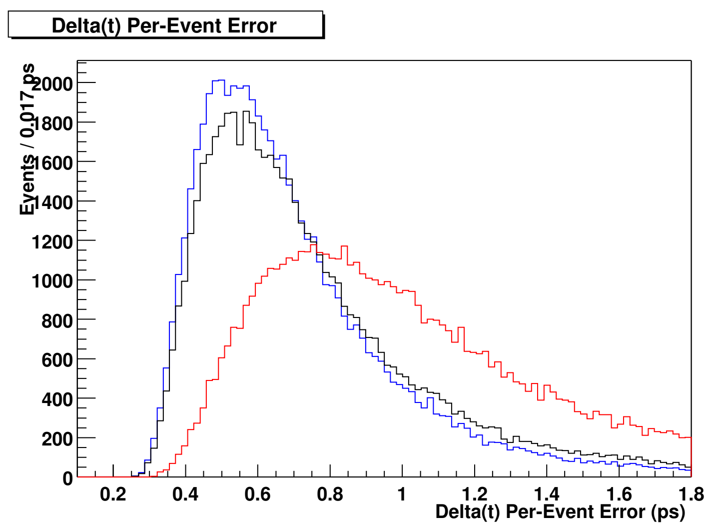

Figure 93: Distribution of per-event Δt error for the signal Monte Carlo sample described in the text. The blue curve corresponds to the $(\mathrm{D}^{*}, \ell, bs)$ algorithm, the black to the (D^{*}, ℓ), and the red to the (D^{*}, bs).

Reco Vtx Algorithm	$\rho(z_{\mathrm{D}^{*}-\ell}, z_{tag})$	RMS $\sigma_{\delta z}$ [μm]
raw	0.00276	155.4
(D^{*}, ℓ)	0.00211	155.5
(D^{*}, bs)	-0.00341	192.4
$(\mathrm{D}^{*}, \ell, bs)$	-0.00449	147.3

Table 81: Comparison of correlation between z_{tag} and $z_{\mathrm{D}^{*}-\ell}$ for the different reco-side vertexing algorithms, and resulting RMS of $\sigma_{\delta z}$ using signal MC candidates.

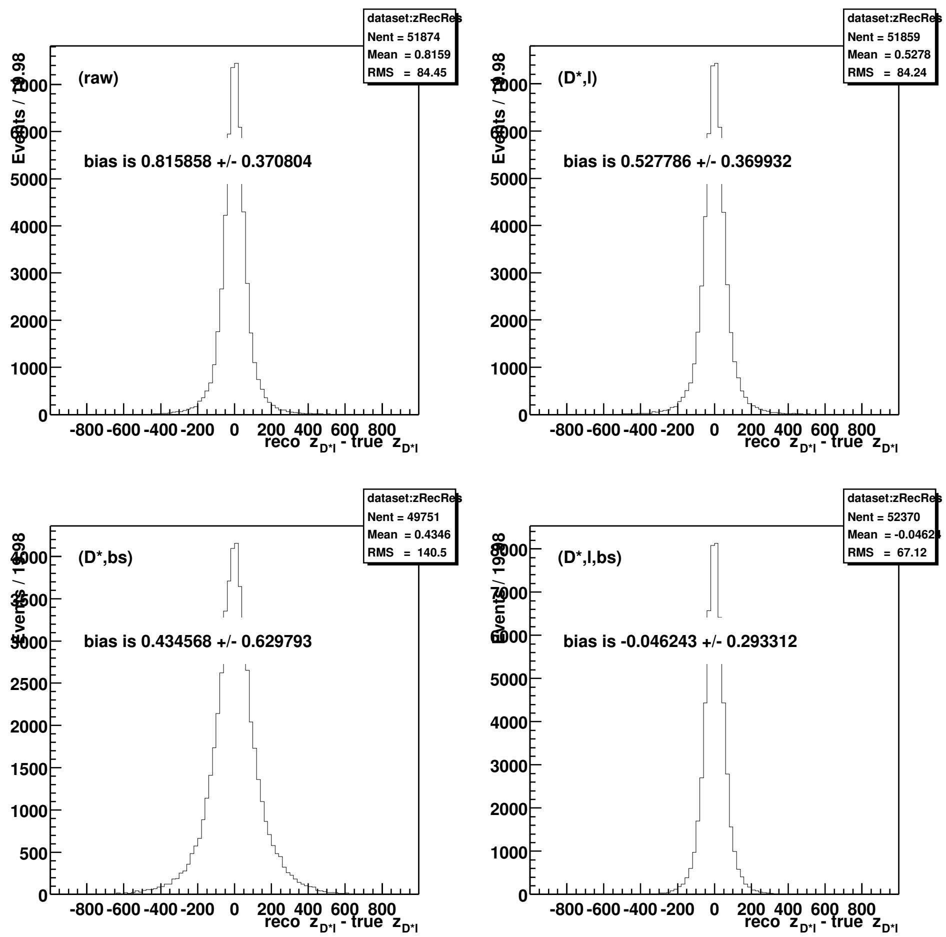

Figure 94: Resisual distributions for the z-position of correctly reconstructed $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ candidates in the signal Monte Carlo sample described in the text. Each of the four plots corresponds to a different reco-side vertexing algorithm used to estimate the z-vertex.

G Extraction of τ_{B^0} and Δm_d Using Only Δt or Only Flavor Information

As discussed in Section 8.3, we have studied the utility of Δt and flavor information when extracting the physics parameters τ_{B^0} and Δm_d. We present here a few additional tables for those more interested.

After building the Δt-shape-only pdf, we can fit for τ_{B^0}, Δm_d, and resolution model parameters either simultaneously on mixed and unmixed events or on each sub-sample separately. In order to optimize convergence of the Δt-shape-only models, we fixed the mistag rates at the values obtained from the full fit and set them constant. Table 82 illustrates the fitted central values and their statistical errors for τ_{B^0} and Δm_d for this set of Δt-shape-only fits. We note, in particular, that the measured value of τ_{B^0} is more precise in the mixed event sample despite the smaller number of events primarily due to the increased "temporal separation" between oscillation and decay time dependence.

	Δt-shape only fits to signal MC	
Sample	Δm_d [ps^{-1}]	τ_{B^0} [ps]
Unmixed evts	0.47653 ± 0.00711	1.5508 ± 0.0160
Mixed evts	0.47217 ± 0.00903	1.5522 ± 0.0121
All evts	0.47685 ± 0.00520	1.5568 ± 0.0091

Table 82: Fitted values from a full signal Δt-shape fit to different samples of signal MC. The mistag rates were held constant in these fits.

We can conclude from the central values that the mixed and unmixed event subsamples yield consistent Δt-shape-only estimates of Δm_d and τ_{B^0} at better than 0.5σ level.

We also perform a check similar to that described in Section 8.3, where we fix all parameters in the fit to those values obtained from the full Δt fit to the full signal MC sample, except for Δm_d. We then extract the two values, $\Delta m_d^{(shape-only)}(mixed) = 0.47867 \pm 0.00595$ and $\Delta m_d^{(shape-only)}(unmixed) = 0.47147 \pm 0.00847$ ps^{-1}, which are consistent at the 1.5σ level, and when combined via a weighted average yield a value of $\Delta m_d^{(shape-only)}(est) = 0.47444 \pm 0.00486$ ps^{-1} which has nearly the same error as that from the simultaneous fit and a consistent central value.

H Correlation Coefficients between Physics Parameters and Other Floating Parameters

Correlation coefficients reported by MINUIT:

```
Parameter              G_sigB0DM     G_sigB0TAU
```

G_combB1_{KaonNT3}	0.0032	0.0082
G_combB1_{L_NT12}	0.0017	0.0045
G_combB0	0.0036	-0.0227
G_combDm	-0.0440	0.0117
G_combF1	-0.0007	-0.0270
G_combFracOsc_{NoLept}	-0.0189	0.0297
G_combFracOsc_{OffRes}	-0.0061	0.0097
G_combFracOsc_{OneLept}	-0.0383	0.0284
G_combFracOsc_{SS}	-0.0104	0.0326
G_combS1_{OffOrSS}	0.0040	-0.0195
G_combS1_{OnAndOS}	0.0011	-0.0043
G_combS0	-0.0306	-0.0092
G_combTauOsc	0.0427	-0.0793
G_combWOsc_{Fake;KTag}	0.0163	0.0057
G_combWOsc_{Fake;LTag}	0.0061	-0.0007
G_combWOsc_{Fake;NT1}	0.0071	-0.0049
G_combWOsc_{Fake;NT2}	0.0094	-0.0103
G_combWOsc_{Fake;NT3}	0.0134	-0.0185
G_combWOsc_{Lepton;KTag}	0.0022	-0.0024
G_combWOsc_{Lepton;LTag}	-0.0081	-0.0008
G_combWOsc_{Lepton;NT1}	0.0008	-0.0004
G_combWOsc_{Lepton;NT2}	0.0050	-0.0026
G_combWOsc_{Lepton;NT3}	0.0088	-0.0104
G_combWPrompt	-0.0157	0.0119
G_contB1_OS	0.0018	0.0057
G_contB1_SS	0.0013	0.0013
G_contF1	-0.0109	0.0160
G_contS1	-0.0149	0.0218
G_contWPrompt_{Kaon}	0.0126	0.0021
G_contWPrompt_{L_NT123}	0.0046	-0.0099
G_fakeB1_{KaonNT3}	-0.0045	0.0018
G_fakeB1_{L_NT12}	-0.0016	0.0044
G_fakeDm	-0.0805	0.0152
G_fakeF1	0.0296	-0.0432
G_fakeFracOsc	-0.0122	0.0243
G_fakeS1	-0.0004	0.0018
G_fakeTauOsc	0.0403	-0.0912
G_fakeWOsc_KTag	-0.0110	-0.0106
G_fakeWOsc_LTag	-0.0201	-0.0002
G_fakeWOsc_NT1	-0.0045	-0.0090

G_fakeWOsc_NT2	-0.0011	-0.0047
G_fakeWOsc_NT3	-0.0099	-0.0173
G_fakeWPrompt_{KaonNT3}	0.0063	0.0228
G_fakeWPrompt_{L_NT12}	-0.0017	0.0229
G_sigB0DM	1.0000	-0.2159
G_sigB0DMTG_KTag	-0.0262	0.0017
G_sigB0DMTG_LTag	0.0461	-0.0050
G_sigB0DMTG_NT1	-0.0642	0.0012
G_sigB0DMTG_NT2	-0.0637	-0.0025
G_sigB0DMTG_NT3	-0.0129	0.0031
G_sigB0F0	-0.0561	-0.2648
G_sigB0MTGO_KTag	-0.1224	0.0192
G_sigB0MTGO_LTag	-0.1797	0.0718
G_sigB0MTGO_NT1	-0.1346	0.0239
G_sigB0MTGO_NT2	-0.0886	0.0146
G_sigB0MTGO_NT3	-0.1081	0.0130
G_sigB0MTGS_KTag	0.0287	0.0080
G_sigB0TAU	-0.2159	1.0000
G_sigB0XD	0.0438	-0.1460
G_sigB0XF_KTag	0.0192	-0.1077
G_sigB0XF_LTag	0.0069	-0.0873
G_sigB0XF_NT1	0.0038	-0.0508
G_sigB0XF_NT2	0.0343	-0.0568
G_sigB0XF_NT3	0.0421	-0.0833
G_sigB0XS	0.1780	-0.4878
G_sigBpmF	0.5799	0.0102
G_uncoB1	-0.0090	-0.0084
G_uncoFracLife	0.0542	-0.0452
G_uncoS1	-0.0817	0.1229
G_uncoTauLife	0.0151	-0.2089
G_uncoWLife_KNT	-0.1118	0.0807
G_uncoWPrompt	0.1063	-0.0936

I Additional Δt-projection Plots for Fit Result to Data

In this section we provide additional plots illustrating the performance of the final full fit result to Data. We select subsamples of the Data by requiring the per-event probabiltiies (as determined from the δm analysis) to meet certain selection criteria, and then projecting the full model onto that subspace. We find that the model describes background samples selected in this way very well.

We first select a subsample of the data that is 80% pure in signal, and show the results in Fig. 95.

We can also select a sample of events enriched in combinatoric bakcgrounds, equivalent to using the δm sideband region. In Fig. 96 we show the fit results projected onto a 99.5% pure combiantoric background sample of the Data.

We finally select a subsample of events that is 60% pure fake leptons, based on a minimum required fake-lepton per-event probability of 0.4. We show the projection of the model onto this subsample in Fig. 98 and the asymmetry in Fig. 99.

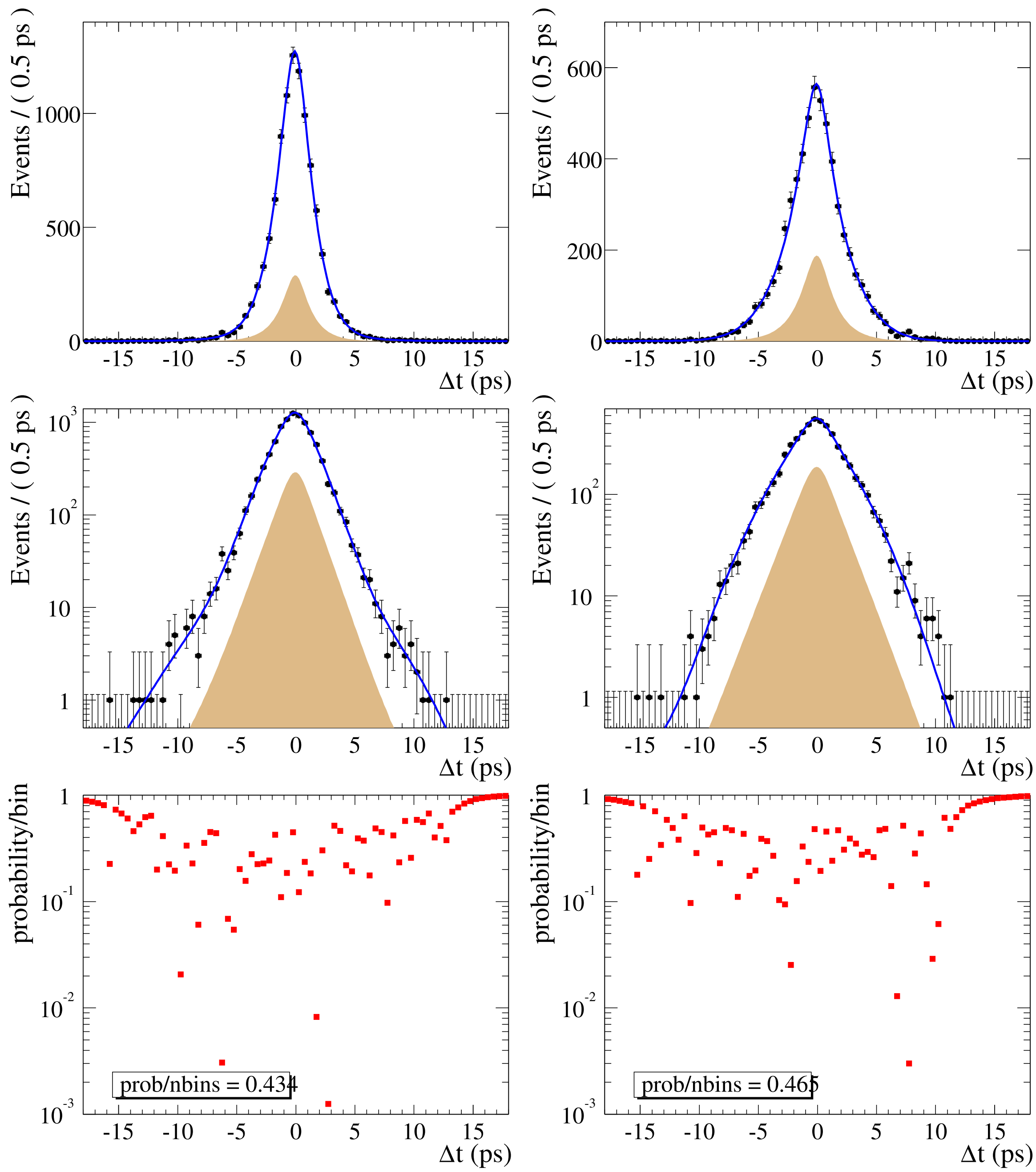

Figure 95: The Δt distribution of unmixed and mixed events in a sub-sample of the Data that is 80% pure in signal (as selected by a cut on the per-event signal probability of greater than 0.4). The points are from Data, the curve is the projection of the full PDF. The superimposed shaded areas on the upper plots show the background contribution to the distributions. The left hand plots are for unmixed events, the right for mixed events; the middle row is simply a log-y plot, The bottom row is the Poisson probability of observing n events or smaller (larger) if n is smaller (larger) than the expected value, i.e., $y = \sum_{i=0}^{n} P(i;\mu)$ for $n < \mu$ and $y = \sum_{i=n}^{\infty} P(i;\mu)$ for $n > \mu$.

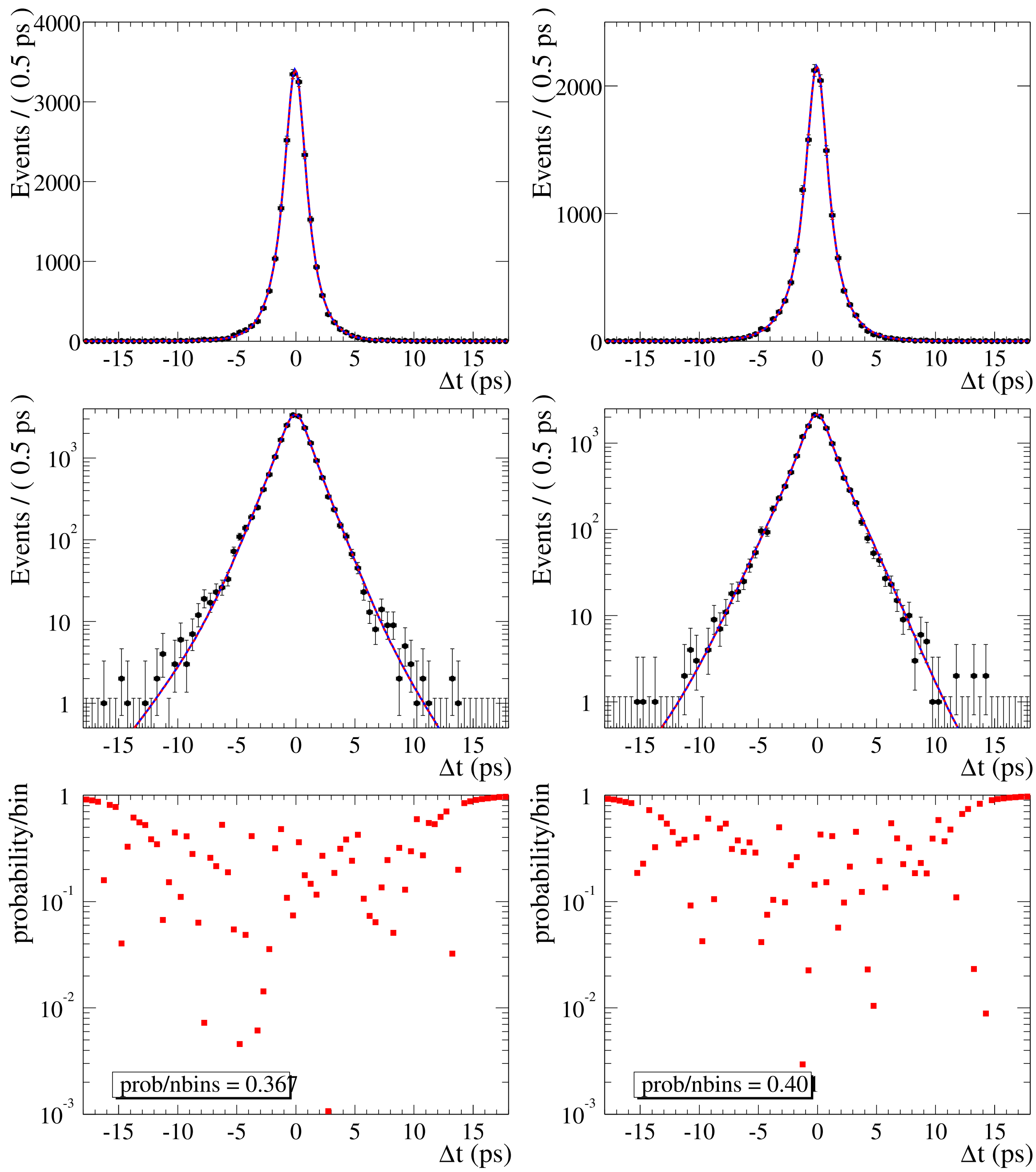

Figure 96: The Δt distribution of unmixed and mixed events in a sub-sample of the Data that is 99.5% pure in combinatorics (as selected by a cut on the per-event combinatoric probability of greater then 0.9). The points are from Data, the smooth (blue) curve is the projection of the full PDF, and the dashed (red) curve is the contribution from the background portion of the full PDF. (In this very pure sample, the curves are identical.) The left hand plots are for unmixed events, the right for mixed events; the middle row is simply a log-y plot, The bottom row is the Poisson probability of observing n events or smaller (larger) if n is smaller (larger) than the expected value, i.e., $y = \sum_{i=0}^{n} P(i;\mu)$ for $n < \mu$ and $y = \sum_{i=n}^{\infty} P(i;\mu)$ for $n > \mu$.

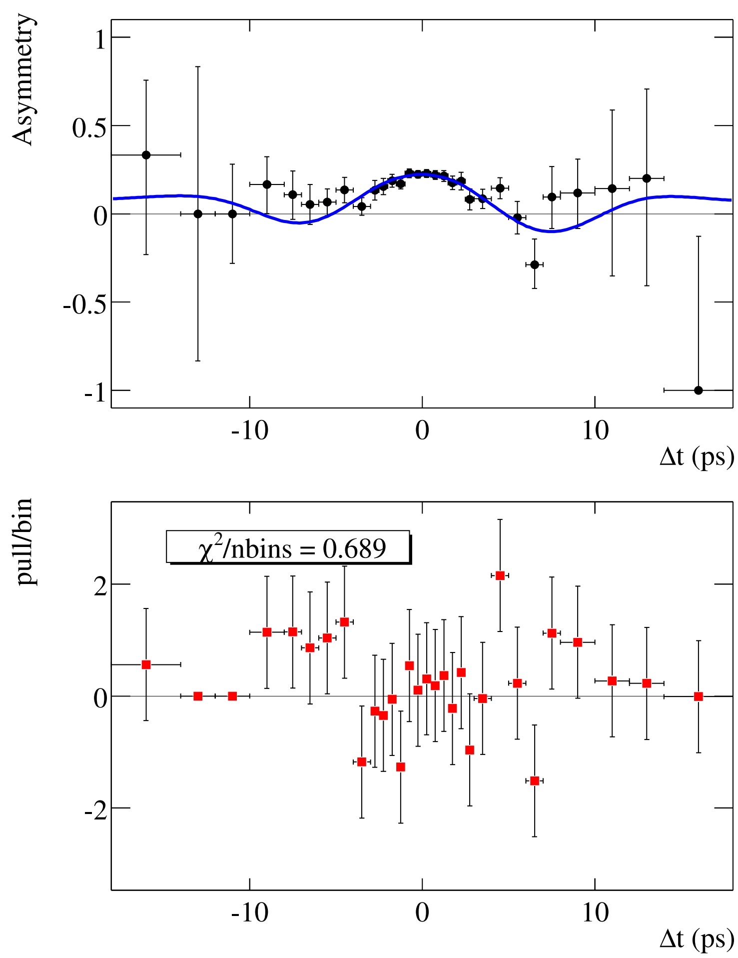

Figure 97: Mixing asymmetry for the 99.5% pure combinatoric subsample of the Data. The lower plot shows the bin-by-bin pull when comparing the model projection to the data distribution.

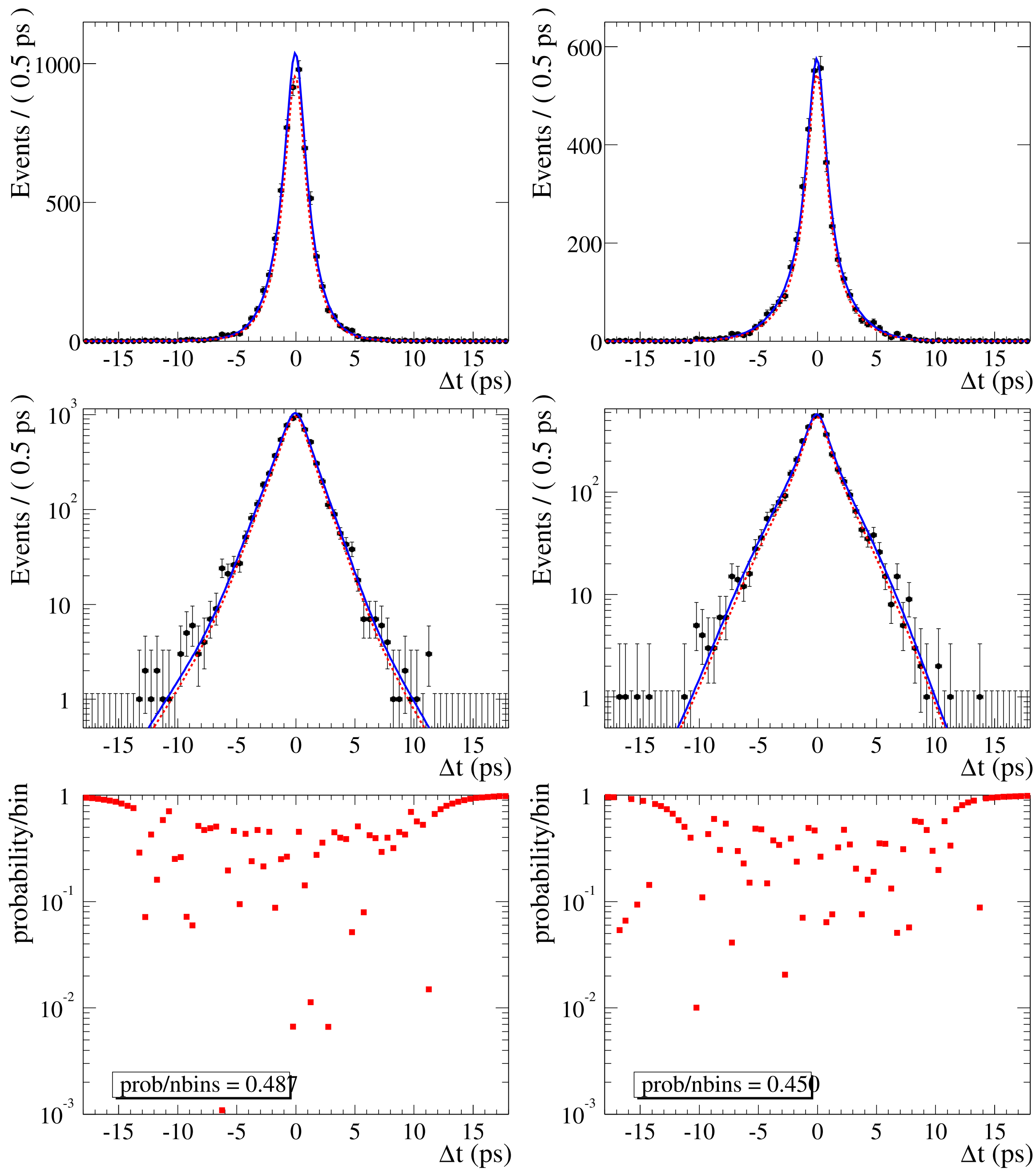

Figure 98: The Δt distribution of unmixed and mixed events in a sub-sample of the Data that is 60.1% pure in fake lepton candidates (as selected by a cut on the per-event fake lepton probability). The points are from Data, the smooth curve is the projection of the full PDF and the dashed (red) curve is the contribution from the backgrounds, inclusding combinatorics. The left hand plots are for unmixed events, the right for mixed events; the middle row is simply a log-y plot, The bottom row is the Poisson probability of observing n events or smaller (larger) if n is smaller (larger) than the expected value, i.e., $y = \sum_{i=0}^{n} P(i;\mu)$ for $n < \mu$ and $y = \sum_{i=n}^{\infty} P(i;\mu)$ for $n > \mu$.

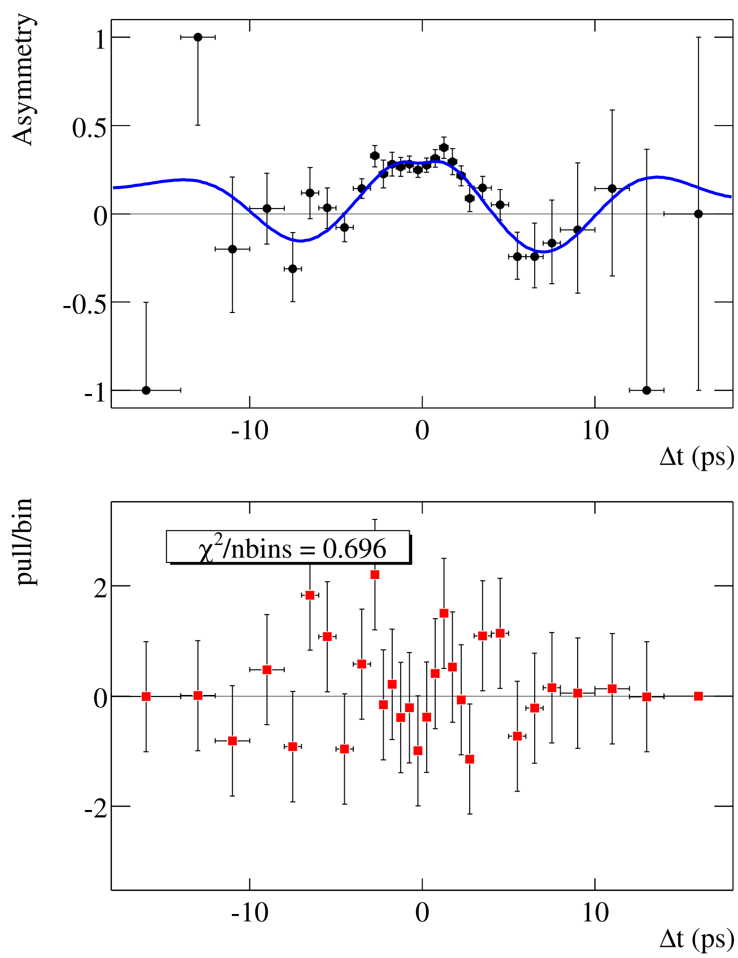

Figure 99: Mixing asymmetry for the 60.1% pure fake-lepton subsample of the Data. The lower plot shows the bin-by-bin pull when comparing the model projection to the data distribution.

References

[1] C. Cheng, D. Kirkby, C. LeClerc, T.I. Meyer, "Semileptonic B Sample for Lifetime and Mixing Analyses", BABAR Analysis Document 147, internal supporting doucment.

[2] B. Cahn, D. Kirkby, "A General Model for Neutral B Decay Time Distributions", BABAR Analysis Document 188.

[3] T.I. Meyer, presentation at April 2001 Collaboration Meeting, Lifetime and Mixing session, Tuesday, April 24, 2001.

[4] D. Kirkby *et al.,* "A User's Guide to the RooFitTools Package for Unbinned Maximum Likelihood Fitting", BABAR Analysis Document 18.

[5] F. Martinez-Vidal *et al.,* "The BaBar Vertexing", BABAR Analysis Document 102.

[6] J. Beringer *et al.,* "B Tagging in BaBar: Status for the sin2β Journal Publication", BABAR Analysis Document 119.

[7] N. Roe *et al.,* "A measurement $B^0\overline{B^0}$ oscillation frequency using reconstructed B^0 decays", BABAR Analysis Document 81, contributed paper to ICHEP2000.

[8] D. Lange *et al.,* "Supporting document for the Run 1 sin2beta analysis", BABAR Analysis Document 115.

[9] G. Raven *et al.,* "Measurement of B^0 mixing using fully reconstructed hadronic B^0 decays", BABAR Analysis Document 125, internal supporting document.

[10] J. Stark *et al.,* "Measurements of the charged and neutral B meson lifetimes using fully reconstructed B decays", BABAR Analysis Document 144, internal supporting document.

[11] C. Cheng, P. Burchat "Correlation between mistag of kaon tag and calculated error on Δt", BABAR Analysis Document 349, internal supporting document.

[12] See page 10 of the talk given by Chih-hsiang Cheng at June 2001 collaboration meeting. http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtgs/2001/detJun2001/Tues4a/cheng.pdf

[13] PDG 2000

[14] BABAR-PUB-01/03, "A Study of Time-Dependent CP-Violating Asymmetries and Flavor Oscillations in Neutral B Decays at the $\Upsilon(4S)$", submitted to Phys. Rev. D.

[15] C. Cheng, "Charged B fraction fits in $D^*\ell$ sample", http://www.slac.stanford.edu/ chcheng/internal/BmixtureFit.ps

[16] R. Cahn, Hypernews post,
http://babar-hn.slac.stanford.edu:5090/HyperNews/get/Bmixing/330/2/1.html

[17] LEP B lifetime working group,
http://claires.home.cern.ch/ claires/lepblife_page2.html#Av

[18] LEP B oscillation working group,
http://lepbosc.web.cern.ch/LEPBOSC/combined_results/PDG_2002/

[19] J. Beringer, *et al.,* "B Tagging in BaBar: Status for the sin2beta Winter Conference results", BABAR Analysis Document 119x.

[20] A. Soffer, *et al.,* "Measuring the PEP-II Boost", BABAR Analysis Document 14.

[21] BABAR Collaboration (B. Aubert et al.), Phys.Rev.Lett.87, 2001

[22] BABAR Collaboration (B. Aubert et al.), Phys.Rev.Lett.89, 2002

[23] BABAR Collaboration (B. Aubert et al.), Phys.Rev.Lett.88, 2002

[24] BABAR Collaboration (B. Aubert et al.), Phys.Rev.Lett.88, 2002

[25] Belle Collaboration (K. Abe et al.), Phys.Rev.Lett.86, 2001

[26] Belle Collaboration (K. Abe et al.), Phys.Rev.Lett.88, 2002