

A measurement of the B^\pm and B^0 lifetime ratio
using the SVT based trigger and accounting for
the induced bias using a Monte-Carlo Free
method

Farrukh Azfar, Joseph Boudreau, Todd Huffman, Louis Lyons,
Sneha Malde, Nicola Pounder, Jonas Rademacker, Azizur
Rahaman

Abstract

We present a measurement of the ratio of charged to neutral B lifetimes using roughly 1fb^{-1} of data accumulated by the high impact parameter selection based hadronic B trigger at CDF. The problem of fitting decay time distributions is solved in a novel Monte-Carlo independent way by analytically calculating acceptances for each event from the decay geometry and known trigger selection criteria. We measure a B^\pm lifetime of $488.5 \pm 6.2 \pm 5.7 \mu\text{m}$ in the decay mode $B^\pm \rightarrow D^0 \pi^\pm$ with $D^0 \rightarrow K^\mp \pi^\pm$ and a B^0 lifetime of $454.3 \pm 6.4 \pm 5.9 \mu\text{m}$ in the decay $B^0 \rightarrow D^\mp \pi^\pm$, with $D^0 \rightarrow K^\mp \pi^\pm \pi^\pm$.

The calculated ratio of charged to neutral B lifetimes is $1.075 \pm 0.020 \pm 0.007$ and is consistent with the PDG value of 1.071 ± 0.009 . All uncertainties quoted from our analysis are statistical and systematic respectively, the PDG uncertainty combines the two categories.

This measurement is presented as a demonstration that we can pursue lifetime measurements in other B hadron decay modes selected by the hadronic B trigger at CDF.

1 Introduction

CDF is the only running experiment to be accumulating a high-statistics sample of hadronic B decays across the full spectrum of B hadrons. This note is concerned with using these data from CDF's hadronic trigger sample, for lifetime measurements.

B-hadron lifetimes, being parameters of fundamental importance in their own right, gain specific significance due to the precise predictions of Heavy Quark Expansion (HQE) [1], [2]. Precision lifetime measurements provide a testing ground for this theoretical tool that is frequently relied upon for relating experimental observables to parameters of the CKM matrix. While precise measurements exist for the types of B-hadrons produced at the B-factories, the accuracy for B_s and Λ_b lags behind the precision of the HQE calculations.

The relative width difference between the long and short lived CP eigenstate of the $B_s^0 - \bar{B}_s^0$ system is predicted to be $\frac{\Delta\Gamma_s}{\Gamma_s} \sim \mathcal{O}(10\%)$. Combined with a measurement of the mass difference between those two states, this parameter could be sensitive to new physics. The lifetime difference can be extracted by measuring the B_s lifetime in decays to pure CP eigenstates, like the fully hadronic decays $B_s^0 \rightarrow D_s D_s$ [3] and $B_s^0 \rightarrow K^+ K^-$, which are both CP even, and compare that with the lifetime measured in flavour specific decays like $B_s^0 \rightarrow D_s \pi$. Until 2007, significant numbers of $B_s^0 \rightarrow D_s D_s$, $B_s^0 \rightarrow K^+ K^-$ and $B_s^0 \rightarrow D_s \pi$ decays will only be available in the CDF hadronic B sample.

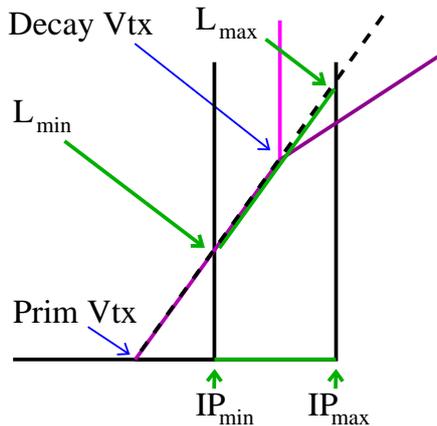
The hadronic B trigger which is so crucial for obtaining these data, biases B lifetime distribution by triggering on the impact parameter of tracks in the event. Currently, at CDF, this effect is taken into account by using a Monte Carlo simulation to calculate an efficiency function.

In this note, we present a Monte Carlo-independent method to correct for this lifetime bias. It only uses information from the measured data on which the lifetime fit is performed, only, to correct for the lifetime bias on an event-by-event basis. This eliminates some systematic problems, maximizes the use of information, and is robust against several effects that could bias the SVT acceptance.

2 The Basic idea

Taking a given event and keeping every kinematic aspect of it fixed, except for the decay time of the primary particle, an upper and a lower impact parameter cut directly translate into cuts on the decay-length and hence on the lifetime of decaying particle, as decay-length and hence on the lifetime of decaying particle, as illustrated for the case of a two-body decay and an impact parameter cut on only one track, in figure 1. A more realistic scenario is given in figure 2. The figure illustrates that, by sliding a decay tree along the direction of the B. The impact parameters correspond to the distance between the prim. vertex and the point where the backwards extensions of the tracks hit the dashed lines

Figure 1: Given the 3-momenta of all particles in the decay, the cut on the Impact parameter of the decay products translates directly into a cut on the lifetime of the primary particle. For clarity, the figure only illustrates the effect of an impact parameter cut on one of the decay products (the one going straight upwards).



perpendicular to them. Where an individual track passes the IP requirements, the corresponding perpendicular dashed line is solid (in same colour as the corresponding track). Whenever 2 tracks pass the IP requirements, the acceptance as a function of time (plotted at the bottom) is set to one, otherwise zero. The hadron trigger at CDF also requires that a track pair pass a minimum L_{xy} cut, where the L_{xy} is calculated from the impact parameters. This is not illustrated here, but accounted for in the method.

The clue is that none of those kinematics needed to translate from an impact parameter cut to a cut (or cuts) on the decay time, have themselves any dependence on the life time of the primary particle.

3 The signal Probability Density Function (PDF) ignoring measurement errors and other detector effects.

We can write the probability to find an event with decay time t as the product of the probability to find t given that t must be between t_{\min} and t_{\max} and

Figure 2: Given the 3-momenta of all particles in the decay, and the decaylengths of particles down the decay chain (here a D^0), the requirements of the hadronic trigger that two particles pass the IP cut translates into an acceptance of one or more intervals.

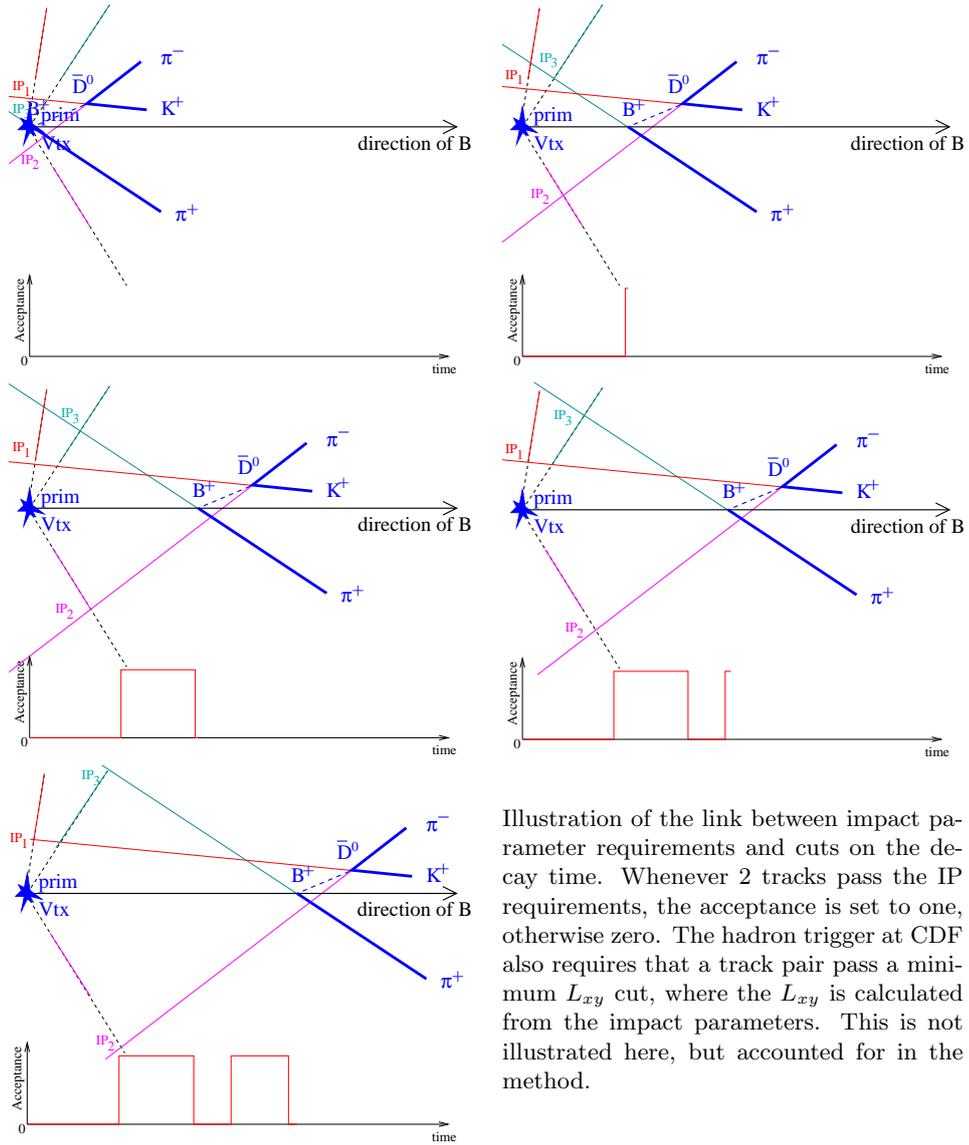


Illustration of the link between impact parameter requirements and cuts on the decay time. Whenever 2 tracks pass the IP requirements, the acceptance is set to one, otherwise zero. The hadron trigger at CDF also requires that a track pair pass a minimum L_{xy} cut, where the L_{xy} is calculated from the impact parameters. This is not illustrated here, but accounted for in the method.

the probability that t is constrained to lie within those limits:

$$\begin{aligned}
P(t) &= P(t|t \in [t_{\min}, t_{\max}]) \cdot P(t_{\min}, t_{\max}) \\
&= \frac{\frac{1}{\tau} e^{-\frac{t}{\tau}}}{\int_{t_{\min}}^{t_{\max}} \frac{1}{\tau} e^{-\frac{t'}{\tau}} dt'} \cdot P(t_{\min}, t_{\max}) \\
&= \frac{\frac{1}{\tau} e^{-\frac{t}{\tau}}}{e^{-\frac{t_{\min}}{\tau}} - e^{-\frac{t_{\max}}{\tau}}} \cdot P(t_{\min}, t_{\max}) \tag{1}
\end{aligned}$$

For a series of measurements, the probabilities for each measured time t_i can be multiplied to give the likelihood for the mean decay time τ . The limits $t_{\min i}$ and $t_{\max i}$ can be calculated easily from the kinematics of each decay. In general it will be difficult to calculate $P(t_{\min i}, t_{\max i})$. However, $P(t_{\min i}, t_{\max i})$ depends only on the impact parameter cut, and the kinematics of the decay – the momenta of the particles, and possibly the decaylengths of some long-lived particles within the decay chain, like the D_s in $B_s \rightarrow D_s^\mp \pi^\pm$ – but not on the life time of the primary itself. So in the log-likelihood, the sum over the $\log(P(t_{\min i}, t_{\max i}))$ is simply a constant that can be ignored. The total log-likelihood function for a set of N “ideal” decays (no measurement uncertainties, background, etc) is given by:

$$\begin{aligned}
\log \mathcal{L} &= -N \log(\tau) \\
&\quad - \sum_{i=1}^N \left(\frac{t_i}{\tau} + \log \left(e^{-t_{\min i}/\tau} - e^{-t_{\max i}/\tau} \right) \right) \tag{2}
\end{aligned}$$

where the index i labels the event, each of which has its measured decay time t_i and minimum and maximum decay times $t_{\min i}$ and $t_{\max i}$.

Note that the only difference to the likelihood function without an impact parameter cut is the term:

$$\log \mathcal{L}_{\text{ip}} = - \sum_{i=1}^N \log \left(e^{-t_{\min i}/\tau} - e^{-t_{\max i}/\tau} \right) \tag{3}$$

The upper lifetime cut has some dramatic effect on the precision with which the lifetime can be measured. Finding an event with lifetime t contains less information, if already restricted the range of possible values for t due to lifetime cuts. The effect is quite significant. For example, an upper lifetime cut at twice the B lifetime loses only 14% of events. However, the statistical error of the measurement is increase by a factor of 2, equivalent to a signal loss of 75%. This is discussed in more detail elsewhere [4].

For more complicated decay geometries, the impact parameter cuts on the decay products can translate into a series of disjoint time-intervals which changes the correction term to:

$$\log \mathcal{L}_{\text{ip}} = - \sum_{i=1}^N \log \left(\sum_{j=1}^{n_i} e^{-\frac{t_{\min ij}}{\tau}} - e^{-\frac{t_{\max ij}}{\tau}} \right) \quad (4)$$

where i labels the events and j labels the allowed time-intervals for each event. The likelihood function in equation 2 is derived for the ideal case that we are dealing with an exact time measurement and an exact impact parameter cut. Any real measurement will have an uncertainty on both.

4 The signal PDF for an “offline trigger”, with measurement errors

As an intermediate step, to illustrate the concepts, assume that the impact parameter cuts are applied to the offline data, only (rather than the SVT-measured quantities). Then the acceptance would still be a top-hat function (or a combination of them), but now as a function of measured decay time, rather than true decay time. Nothing would change in the illustrations 2, except that all quantities are now offline-measured quantities, and the acceptance for the event is plotted as a function of the measured proper time.

We can write the probability to measure a decay time t_0 (given the IP cut and the decay kinematics that relate the impact parameter to the time measurement for any given decay) as an integral over all true decay times t in terms of the following functions:

- The probability that a particle decays with true decay time t , given its mean life τ ,

$$\frac{1}{\tau} e^{-\frac{t}{\tau}}.$$

- The probability that, given the true decay time t and measurement uncertainty of σ_t , the measured decay time is t_0

$$\frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-t_0)^2}{2\sigma_t^2}}.$$

- The acceptance as a function of the *measured* decay time t_0 for the given decay kinematics.

$$A_{\text{ip}}(t_0).$$

In terms of these parameters, the total probability is:

$$P(t_0) = \frac{\int_0^{\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-t_0)^2}{2\sigma_t^2}} A_{\text{ip}}(t_0) dt}{\int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-t_0)^2}{2\sigma_t^2}} A_{\text{ip}}(t_0) dt dt_0} \quad (5)$$

If the impact parameter cut were applied on the offline quantities,

$$A_{\text{ip}}(t_0, \dots) = \sum_{\substack{i=\text{all} \\ \text{intervals}}} (\theta(t_0 - t_{\min i}) - \theta(t_0 - t_{\max i})) \quad (6)$$

where θ is the Heaviside function. It has this simple form because we established a direct link between the offline impact parameter and the measured lifetime.

5 The signal PDF for different online and offline quantities.

The real trigger uses fast-measured SVT quantities rather than offline quantities to cut on, thus, at first sight, destroying the one-to-one correspondence between impact parameters and $c\tau$. We will now re-establishing a direct link between the SVT-measured impact parameter and measured lifetime, and thus keep the very simple form of the acceptance function Equation 6.

5.1 Using Δd_0

To do so, we simply include the difference between the offline impact parameter and the SVT impact parameter amongst those parameters that we assume to be lifetime independent. This means that if we plot the (true) lifetime in bins of “(SVT impact parameter) minus (offline impact parameter)”, we assume that the distributions will all look the same. This is the case if the SVT impact parameter error is independent of the actual value of the impact parameter, which can be validated using the same data the fit is performed on. Figure 3 illustrates how the SVT- $d_0(c\tau)$ is calculated from the offline $d_0(c\tau)$ using $d_0^{\text{SVT}} = d_0^{\text{off}} + \Delta d_0$, assuming a constant Δd_0 . One of the advantages of this method is that we do not need to know the actual impact parameter error. The SVT resolution function can have any shape. This method can even handle systematic shifts in the impact parameter measurement of the SVT, as long as those shifts are uniform within the allowed impact parameter range, which means that this method is much more robust and requires a far less detailed understanding of the SVT performance than any Monte-Carlo based method.

5.2 The discretised SVT d_0

While the method is intrinsically insensitive to shifts and skews in the SVT- d_0 resolution function, it turns out to be surprisingly sensitive to the discretisation in the SVT-measured impact parameter. The SVT, using fast integer arithmetic to fit the track parameters, returns impact parameters in multiples of $10 \mu\text{m}$, i.e. possible values are $d_0^{\text{SVT}} = 0, \pm 10 \mu, \pm 20 \mu, \pm 30 \mu, \dots$. A typical d_0^{SVT} distribution in Monte Carlo, for values between 120μ and 500μ , is shown

Figure 3: Re-establishing the direct link between impact parameter cuts (in SVT) and measured lifetime (measured from offline data).

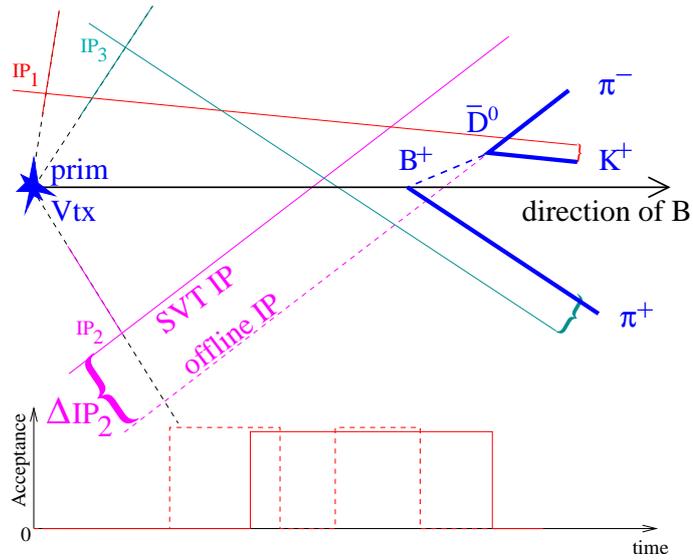
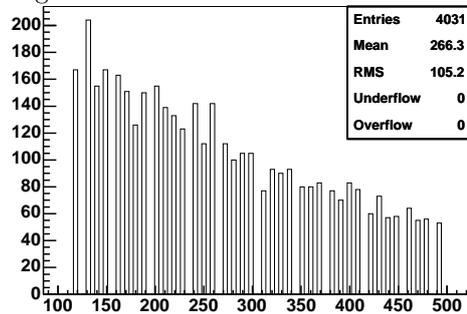


Figure 4: The SVT-calculated d_0 in Monte Carlo events. For clarity, the histogram is restricted to values between 120μ and 500μ .



in Figure 4. Steps of $10\ \mu$ seem small compared to the SVT resolution of about $\sim 50\ \mu\text{m}$, but ignoring this discretisation results in a significantly biased fit result. Ignoring the effect in a fit to 15 k detailed MC events yielded a fit result of $c\tau = 448 \pm 6\ \mu\text{m}$ for an input value of $496\ \mu$, about 10σ off. This is easy to take into account though. As the event is slid along in $c\tau$, the SVT- d_0 is not simply calculated as $d_0^{\text{SVT}}(c\tau) = d_0^{\text{off}}(c\tau) + \Delta d_0$, instead the result for d_0^{SVT} is rounded to the nearest $10\ \mu\text{m}$.

$$d_0^{\text{SVT}}(c\tau) = [\text{nearest multiple of } 10\ \mu\text{m of}] (d_0^{\text{off}}(c\tau) + \Delta d_0).$$

With this modification, the fit result, using the same events, is $494 \pm 7\ \mu\text{m}$, in good agreement with the input value of $496\ \mu\text{m}$. More on detailed and toy Monte Carlo studies in Section 11.

5.3 The full PDF with realistic SVT errors, but a flat SVT efficiency between $d_0 = 0$ and $d_0 = \infty$.

Now we have a *direct* relationship between the measured lifetime and the SVT impact parameter. So we can take the decay geometry and vary, as the only parameter, the measured lifetime by sliding the decay vertex position along the direction of the measured momentum. For each position of the decay vertex, we can calculate, from the measured decay geometry, the corresponding offline impact parameter, and get from that the impact parameter the SVT would have measured for any given measured decay time, as illustrated in figure 3.

Because of the direct correspondence between measured lifetime and SVT-measured impact parameter, the acceptance in terms of the measured time t_0 is still:

$$A_{\text{ip}}(t_0, \dots) = \sum_{\substack{i=\text{all} \\ \text{intervals}}} (\theta(t_0 - t_{\text{min } i}) - \theta(t_0 - t_{\text{max } i})) \quad (7)$$

With this equation 5 becomes:

$$P(t_0) = \frac{\int_0^\infty \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-t_0)^2}{2\sigma_t^2}} dt}{\sum_{\substack{i=\text{all} \\ \text{intervals}}} \int_{t_{\text{min } i}}^{t_{\text{max } i}} \int_0^\infty \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-t_0)^2}{2\sigma_t^2}} dt dt_0} \quad (8)$$

measurement. Using the frequency function

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \quad (9)$$

this can be written as

$$\begin{aligned}
P(t_0) &= \frac{\frac{1}{\tau} e^{-\frac{t_0}{\tau} + \frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{\substack{i=\text{all} \\ \text{intervals}}} \int_{t_{\min i}}^{t_{\max i}} \frac{1}{\tau} e^{-\frac{t}{\tau} + \frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) dt_0} \\
&= \frac{\frac{1}{\tau} e^{-\frac{t_0}{\tau} + \frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{\substack{i=\text{all} \\ \text{intervals}}} \left[-e^{-\frac{t}{\tau} + \frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + \text{F}\left(\frac{t}{\sigma}\right) \right]_{t=t_{\min i}}^{t=t_{\max i}}} \quad (10)
\end{aligned}$$

Deviding both numerator and denominator by $e^{\frac{1}{2} \frac{\sigma^2}{\tau^2}}$ makes the formula numerically robust against very large values for σ/τ , yielding very large values for $e^{\frac{1}{2} \frac{\sigma^2}{\tau^2}}$. So finally we get:

$$\boxed{
\begin{aligned}
P(t_0) &= \frac{\frac{1}{\tau} e^{-\frac{t_0}{\tau}} \text{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{\substack{i=\text{all} \\ \text{intervals}}} \left[-e^{-\frac{t}{\tau}} \text{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + e^{-\frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t}{\sigma}\right) \right]_{t=t_{\min i}}^{t=t_{\max i}}} \quad (11)
\end{aligned}
}$$

The Frequency Function F can be calculated by fast numerical algorithms, implemented for example in the cernlib function `FREQ`, a C++ translation of which is available in root as `TMath::Freq`. Therefore, with equation 10, or equation 11, we have a fully analytical formula to calculate the log-likelihood function, taking into account the SVT-based trigger.

The only task remaining is to use the direct relationship between the offline-measured time and the SVT-impact parameter to find the intervals of measured times within which a given decay would be accepted. The trigger does not only cut on one impact parameter, but requires two tracks to pass the impact parameter cuts, p_t cuts and χ^2 cuts. The SVT also calculates the L_{xy} of each track pair, from the above information. We can calculate what impact parameter the SVT would have measured for each measured decay time, and know all the other SVT quantities used by the trigger. So, sliding the decay-tree up and down along the measured momentum direction, we can calculate for each position (each possible measured time), if the event would have passed the trigger or not. In practice, this is implemented as a search algorithm. The algorithm scans through all times between some absolute minimum and maximum time cut in sensibly sized steps for a first estimate. It then refines the intervals using standard iterative methods.

6 The cut-off in the SVT single track efficiency, and the absolute trigger efficiency for 2 tracks and more.

6.1 Why it doesn't matter for 2 tracks

For two particle final states (like $B_d \rightarrow \pi\pi$, $B_s \rightarrow KK$), the absolute value of the efficiency function is irrelevant, because it only changes between zero, and a constant non-zero value as we slide the event along in $c\tau$. The absolute value of that constant does not affect the fit. Note that this argument assumes that the SVT track-finding efficiency is independent of $c\tau$. This is a reasonable assumption for tracks with $|d_0| < 1$ mm. For the 2-track case, the trigger cuts ensure $|d_0| < 1$ mm for both tracks. Since no track with $|d_0| > 1$ mm enters the fit, it does not matter that the trigger efficiency does not remain constant beyond that point.

6.2 The complication for more than 2 tracks

An interesting complication arises if there is more than one track pair in the decay that could fire the trigger. As we slide the decay along, there'll be regions in $c\tau$ where one track pair is available for triggering, and others, where there are two. In general, the efficiency should be higher if two track pairs satisfy the trigger requirement, rather than only one. This is because, for a given single-track finding efficiency of the SVT (which is around 50%), the probability of finding two tracks out of three is higher than the probability of finding two out of two.

If the SVT track finding efficiency were indeed independent of $c\tau$, the following argument would save us the complicated calculation: We could simply calculate

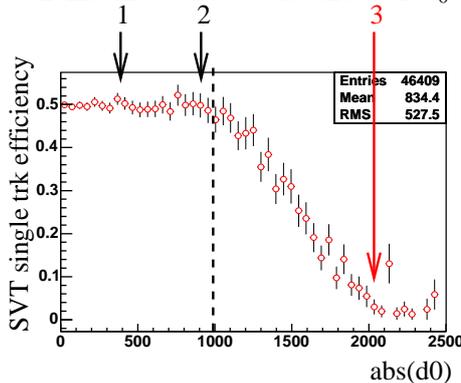
$$P(t|t \in [t_{\min}, t_{\max}])$$

given that the SVT found exactly those tracks it did. *Given* the found tracks, the trigger efficiency is either 1 or 0, no matter how many tracks are available for triggering.

Unfortunately, beyond $|d_0| = 1$ mm, the SVT efficiency is clearly not flat, instead it drops quite rapidly, as shown in figure Figure 5. This is not a problem for two body decays, because tracks with $|d_0| > 1$ mm are never seen because of the very trigger requirements we are correcting for. But in a three body decay, two tracks with $0.12 \text{ mm} < |d_0| < 1$ mm can fire the trigger, while the third track can have an impact parameter $|d_0|$ well beyond 1 mm, where the SVT single track finding efficiency is essentially 0. This track won't have an SVT-measured d_0 . As we slide the event along in $c\tau$, at some point the d_0 of the third track will be < 1 mm, and it could potentially play a role in the trigger decision. We now need two pieces of information to get the trigger efficiency at this $c\tau$:

- How likely would the track have been found by the SVT?

Figure 5: SVT single track finding efficiency as a function of $|d_0^{\text{off}}|$ in Monte Carlo, for the π_B from the B_u in $B_u \rightarrow D(K\pi_D)\pi_B$ with $p_t > 1.5$ GeV. The arrows indicate d_0^{off} values for a 3-track event. Tracks 1 and 2 have an SVT match. Track 3 hasn't. As the efficiency f_{ct} for this event is calculated for different values of $c\tau$, at some point track 3's $|d_0^{\text{off}}|$ would be below $1000 \mu\text{m}$. Would it have an SVT match? What would the SVT- d_0 be?



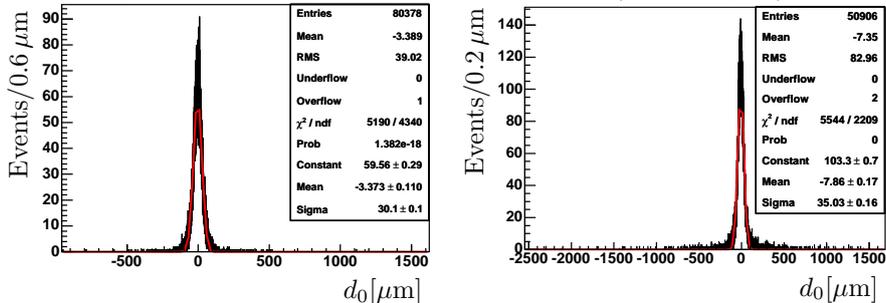
- What would the SVT-measured d_0 have been?

6.3 Easy, but expensive ways out

There are two simple ways out of this predicament, that ensure that there never is a track with $|d_0| > 1$ mm, and therefore no need answer to the two questions posed in the previous section:

- “Two track”: Treat a multibody decay like a two body decay, by declaring two tracks as “trigger tracks” and re-applying the trigger condition using these two tracks only. Note that the decision which two tracks are the trigger tracks must be made before we know if they actually fired the trigger or not – we can’t increase event numbers by simply choosing event-by-event the tracks that actually did fire the trigger. This way we are back in the same situation as for two-body decays and don’t need to worry what happens to the SVT efficiency beyond $|d_0| = 1$ mm. This solution is rather costly in statistics.
- “Fiducial cut”: Use all tracks in the trigger decision, but impose a cut requiring *all* of them to have an impact parameter $|d_0| < 1$ mm. The effect of this cut on the acceptance as a function of $c\tau$ can be calculated in the same way as that of the other impact parameter cuts. Again, no track with $|d_0| > 1$ mm affects the calculation of the trigger acceptance, and it doesn’t matter how the SVT single-track finding efficiency looks like beyond 1 mm. This solution is not very costly in the number of events, but since it reduces the width of the lifetime window, it significantly reduces the statistical power per event, due to the effect discussed in [4].

Figure 6: Δd_0 Distribution for tracks in $B_u \rightarrow D\pi$ candidates.
Signal Monte Carlo Data ($\sim 16\%$ signal)



Since both simple solutions outlined above are too costly in statistical precision, we will have to answer the above questions, how likely a track is to be found once its $|d_0|$ is below 1 mm (single track SVT efficiency), and what its impact parameter would have been.

6.4 Solving the > 2 track problem in an efficient way

So in order to fit lifetimes efficiently, we will finally have to answer the questions:

- What would the SVT-measured d_0 have been for those tracks that haven't got an SVT match?
- How likely is a track to be found by the SVT?

We will then use this to calculate the absolute value of the SVT efficiency, which will vary depending on how many tracks pass the trigger requirements at a given $c\tau$.

6.4.1 Assigning a value for the SVT- d_0 to those that haven't got one.

As we slide the event along in $c\tau$ to calculate the efficiency, we calculate the SVT d_0 at a given value for $c\tau$ from the offline d_0 , assuming that $\Delta d_0 \equiv d_0^{\text{SVT}} - d_0^{\text{off}}$ is independent of $c\tau$. To assign a value for the SVT- d_0 to those tracks that weren't actually found by the SVT (for example because their d_0 was outside the SVT acceptance), we first histogram the $\Delta d_0 \equiv d_0^{\text{SVT}} - d_0^{\text{off}}$ distribution for those tracks where this information is available. Such a histogram is shown for real data in figure Figure 6. For all tracks without an SVT d_0 , we draw a random number from this histogram, i.e. we generate a random Δd_0 according to the Δd_0 distribution found in data.

Table 1: Trigger efficiency in terms of the SVT-single track finding efficiency, for a three particle final state.

Number of track-pairs passing the trigger cuts, out of 3 tracks	Trigger efficiency in terms of ε_s
1 track pair	ε_s^2
2 track pairs	$2\varepsilon_s^2 - \varepsilon_s^3$
3 track pairs	$3\varepsilon_s^2 - 2\varepsilon_s^3$

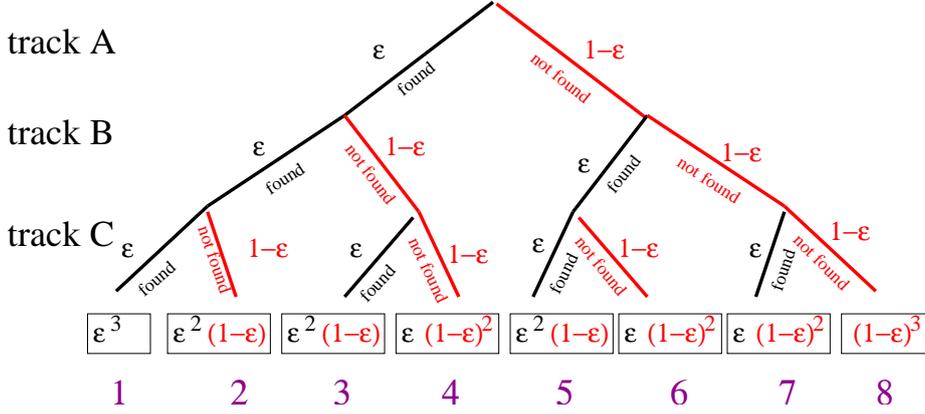
6.4.2 Absolute Trigger efficiency from the SVT single-track finding efficiency

Now that we include tracks outside the d_0 range where the SVT single track finding efficiency is flat, we cannot simply calculate the efficiency function *given* that the SVT found the exactly those tracks it did, because this condition is no longer $c\tau$ independent. Therefore the efficiency function will no longer simply be either 1 or 0. Instead it will depend on the number of tracks available for the trigger decision, and the probability of the SVT to find those tracks. In order to decide which track combinations could have fired the trigger, we need d_0^{SVT} for all tracks involved, including those which were not actually found by the SVT. For the tracks not found in the SVT we use the d_0^{SVT} values generated from random numbers and the measured Δd_0 distribution, as described in Section 6.4.1 above.

In order not to have to model the complicated turn-off curve of the SVT efficiency near $|d_0^{\text{off}}| = 1$ mm we describe the SVT single track finding efficiency as flat for $|d_0^{\text{off}}| < 1$ mm and zero elsewhere. For this to be accurate, we have to treat those tracks with $|d_0^{\text{off}}| > 1$ mm as having not been found by the SVT. With this simple form, the SVT single-track finding efficiency is described by a single parameter, the SVT single track finding efficiency for tracks with $|d_0^{\text{off}}| < 1$ mm, ε_s . At each given $c\tau$, for each given track, the SVT single track finding efficiency is either 0 or ε_s .

The total SVT efficiency is the probability that at least one track pair that satisfies the trigger requirements will be found by the SVT. This can be expressed as a polynomial in ε_s . The possible values for the SVT efficiency for the three track case are given in Table 1. For 4 or more tracks in the final states, this is a bit more complicated, for example we would need to distinguish two possible ways in which 2 track pairs could pass the trigger: the pairs could either have a track in common, or not. In the computer program calculating those efficiencies, this is handled in the most general way, allowing to calculate the total efficiency for any number of tracks and any track combination. This is achieved by generating a “decision tree” at the end of which stand all possible, mutually exclusive combinations of found and missed tracks. The probability for each such combination is calculated, where each track found contributes a factor of ε_s , and each missed track a factor of $(1 - \varepsilon_s)$. These probabilities are added up for all combinations that pass the trigger cuts. This process is illustrated for

Figure 7: Decision tree for the example of 3 tracks in the final states. Each track can either be found (probability ε_s) or not be found (probability $1 - \varepsilon_s$) by the SVT, giving $2^3 = 8$ possible combinations. The total trigger efficiency is calculated by adding up the individual probabilities of those combinations that pass the trigger cuts.



the three-track case in Figure 7, from which the results listed in table Table 1 can be read off in the following way:

- 1 pair: If for example only the track pair (A,B) passes the trigger cuts, we need to add up the probabilities for combinations 1 and 2, giving $\varepsilon_s^3 + \varepsilon_s^2(1 - \varepsilon_s) = \varepsilon_s^2$.
- 2 pairs: If (A,B) and (B,C) pass, but not (A,C) (for example because of the opposite charge requirement), the possible combinations are 1, 2, 5, giving $\varepsilon_s^3 + \varepsilon_s^2(1 - \varepsilon_s) + \varepsilon_s^2(1 - \varepsilon_s) = 2\varepsilon_s^2 - \varepsilon_s^3$.
- 3 pairs: If all three possible track pairings pass the trigger requirements (which is possible in the B.CHARM.LOWPT scenario which has no opposite charge requirement), we add up combinations 1, 2, 3, 5, giving $3\varepsilon_s^2 - 2\varepsilon_s^3$.

6.4.3 Fitting ε_s

The method described above requires the absolute value of the SVT single track finding efficiency. This is fit at the same time as the lifetime, and the other parameters of the fit. The information used to fit the single track finding efficiency is the number of tracks found in each event, relative to the minimum of 2 required to pass the trigger. For the three track case, it is the frequency of finding two tracks in the SVT versus three, for those events where all three tracks are within the SVT's reach, i.e. have $|d_0^{\text{off}}| < 1$ mm and a minimum p_t of 2 GeV. The probabilities associated with those track configurations are exactly those at the end of the decision tree in Figure 7.

It is obvious that for two body decays, there is not enough information to fit the single track finding efficiency, because in order to pass the trigger, all decays will have exactly two tracks found in the SVT. Fortunately this doesn't matter, since the single track finding efficiency is not needed for two body decays anyway, as discussed in Section 6.1.

6.4.4 The full signal PDF with realistic trigger for decays to three or more particles

For 2 body decays, the probability density function given in Equation 7 is sufficient to fit a lifetime to an SVT-biased signal sample. For multibody decays the PDF needs to be modified to take into account the above considerations. We will use the following definitions:

- $P(\text{trk}|\varepsilon_s)$: The probability to find exactly the given track combination, which corresponds to one single element at the end of the decision tree in Figure 7.
- $P(\text{trigger}|\text{trk}, t_o)$: The probability that the given track configuration fires the trigger, given the impact parameters etc calculated for the measured decay time t_o , using the sliding method. This is either 1 or 0.
- $P(\text{trigger}|\varepsilon_s, t_o)$: The probability that the trigger fires, given ε_s , but summed over all possible track combinations that could have fired the trigger, $p(\text{trigger}|\varepsilon_s, t_o) = \sum_{\text{trk}} p(\text{trk}|\varepsilon_s)P(\text{trigger}|\text{trk}, t_o)$. This corresponds to the entries in Table 1. It is essentially the normalisation factor to go with $P(\text{trk}|\varepsilon_s)P(\text{trigger}|\text{trk}, t_o)$
- $\text{poly}_i(\varepsilon_s)$: Since $P(\text{trigger}|\varepsilon_s)$ is constant for t_o within one time interval with constant track configuration, it can be replaced by $\text{poly}_i(\varepsilon_s)$, where the index i labels the time interval, and $\text{poly}_i(\varepsilon_s)$ is one of the polynomials in table Table 1 (or equivalent).

With these definitions, the PDF for a single decay can be expressed as

$$\begin{aligned}
P(t_0) &= \frac{P(\text{trk}|\varepsilon_s)P(\text{trigger}|\varepsilon_s, t_o)\frac{1}{\tau}e^{-\frac{t_0}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}}\text{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{\text{all trk}} \sum_{i=\text{all intervals}} \int_{t_{\min i}}^{t_{\max i}} P(\text{trk}|\varepsilon_s)P(\text{trigger}|\varepsilon_s, t_o)\frac{1}{\tau}e^{-\frac{t}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}}\text{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) dt} \\
&= \frac{P(\text{trk}|\varepsilon_s)\frac{1}{\tau}e^{-\frac{t_0}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}}\text{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{i=\text{all intervals}} \text{poly}_i(\varepsilon_s) \left[-e^{-\frac{t}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}}\text{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + \text{F}\left(\frac{t}{\sigma}\right) \right]_{t=t_{\min i}}^{t=t_{\max i}}} \quad (12)
\end{aligned}$$

where we omitted $P(\text{trigger}|\varepsilon_s, t_o)$ in the numerator because it is 1 for all events in the sample.

6.5 The trigger at Level 3

We must confirm that the decay tracks can pass the trigger in this analysis. When doing this we must consider the information that the trigger had and not the offline quantities that we calculate during offline.

The trigger at Level 2 uses SVT quantities which are used to confirm the trigger at Level 2 as explained previously. At Level 3 there are Level 3 tracks that are used to decide if the event satisfies the criteria. Therefore it is more accurate for us to use the Level 3 tracks to confirm the Level 3 trigger rather than the offline quantities.

Each event comes attached with a Level 3 bank which contains a list of Level 3 tracks and their properties. This object only contains tracks with $P_T > 1.2\text{GeV}$ but this does not pose a problem for us as all track in a trigger decision must have Pt greater than 2GeV. To use these tracks we must match the offline tracks to L3 ones and then use the L3 information of the matched track to see if the event passes the trigger and then use this information again to calculate the acceptance function.

The tracks are matched using the the Pt and phi for the offline tracks and finding which L3 track comes closest. To to this we look for the L3 track that has the minimum value for this chi-square like quantity shown below (equation 13)

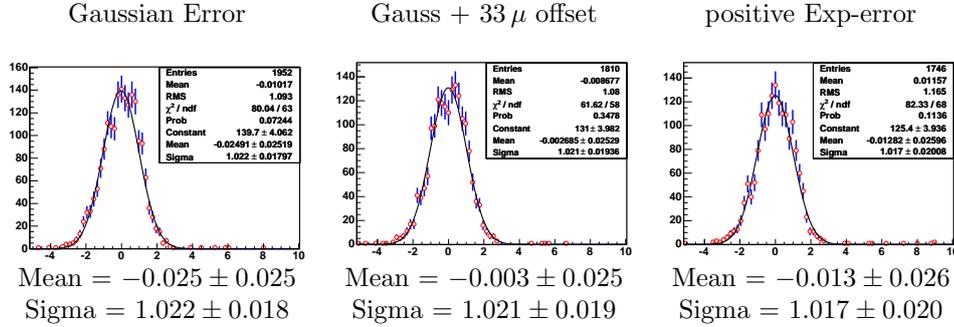
$$\chi^2 = \frac{(\text{Off}_{P_T} - L3_{P_T})^2}{\sigma^2_{L3_{P_T}}} + \frac{(\text{Off}_\phi - L3_\phi)^2}{\sigma^2_{L3_\phi}} + \quad (13)$$

6.6 Toy-MC

In order to test the basic principle, a toy-Monte Carlo simulation is used that generates isotropic $B_s^0 \rightarrow D_s \pi$ events with a mean B_s^0 -lifetime of 1.55 ps and a mean D_s -lifetime of 0.49 ps. The 2-D impact parameter resolution is assumed to be Gaussian with $33 \mu\text{m}(\text{intrinsic}) \otimes 33 \mu\text{m}(\text{beam-spot})$. The impact parameter measured in the x-y plane is required to be between 0.12 mm and 1 mm. Alternative intrinsic IP resolution functions have been tried out to demonstrate the robustness of the method against systematic effects:

- “Standard”: A simple Gaussian resolution function with $\sigma = 33 \mu$, as described above.
- “Offset”: A Gaussian resolution function with $\sigma = 33 \mu$, with an offset of 33μ , i.e. the mean measured SVT impact parameter is 33μ larger than the true one.
- “Exponential from hell”: A positive exponential with an rms of 33μ , i.e. the SVT impact parameter is always bigger than the true impact parameter, and the difference is distributed according to an exponential with a “lifetime” of 33μ .

Figure 8: Toy MC pulls $1 - 2k$ MC experiments, $0.5k$ signal evts each, $S/B = 1$, with different intrinsic IP-resolutions



None of these rather drastic biases produces a bias in the fitted lifetimes (Figure 8).

6.7 Detailed MC

The method has been tested on a Monte Carlo sample of $B_u \rightarrow D\pi$ signal events, with a detailed detector simulation, in particular a detailed simulation of the SVT and the trigger. 35 k events passed all cuts, including those imposed on the SVT. The fit result of $495 \pm 5 \mu$ compares well with the true value of 496μ . A projection of the fit to the MC data is shown in figure 9.

7 Including Background

7.1 Introduction

So far, we have described a Monte Carlo-free method to correct for the trigger bias, that works on signal data alone. Including background makes the situation considerably more complicated, because the basic trick we applied doesn't quite work anymore. In our PDF, we calculate the probability to find a lifetime *given* the efficiency function A_{trig} , calculated from the decay kinematics that translate the trigger cuts into different lifetime cuts event by event. The argument was that those kinematics do not themselves depend on the lifetime, and the corresponding term in the PDF can be ignored. Mathematically: In the expression

$$P(c\tau, \text{kin}) = P(c\tau|\text{kin})P(a) \quad (14)$$

we can ignore $P(\text{kin})$ because it is a simple factor and $\frac{d}{d\tau}P(\text{kin}) = 0$. However, if we add background, the full expression is (where $P(s)$ is the signal probability and $P(b) = 1 - P(s)$ the background probability)

$$P(c\tau, A_{\text{trig}}) = P(s)P(c\tau|A_{\text{trig}}, s)P(A_{\text{trig}}|s) + P(b)P(c\tau|A_{\text{trig}}, b)P(A_{\text{trig}}|b) \quad (15)$$

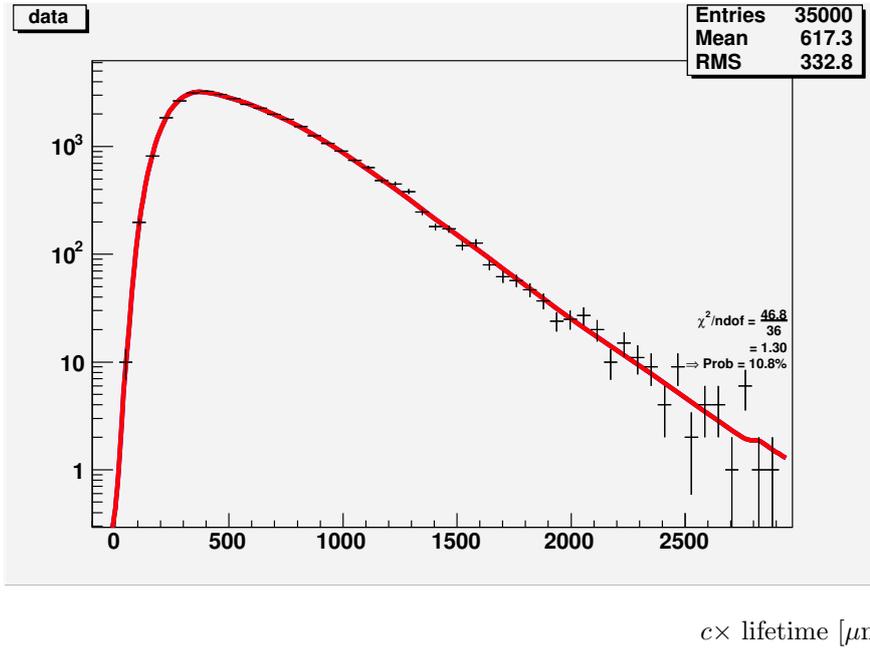


Figure 9: Monte Carlo-independent lifetime fit (line) to 35k simulated $B^+ \rightarrow D\pi$ events (crosses), subject to the impact parameter trigger at CDF, using a detailed detector simulation. MC-input: $c\tau = 496 \mu\text{m}$. Fit result: $c\tau = 495 \pm 5 \mu\text{m}$.

Now the efficiency-function terms, $P(A_{\text{trig}}|s)$ and $P(A_{\text{trig}}|b)$, only factor out if they are the same for signal and background. If they are different for signal and background, ignoring these factors is equivalent to getting the signal fraction wrong in the fit, which is more obvious if we re-write Equation 15 as

$$\begin{aligned}
 P(c\tau, A_{\text{trig}}) &= P(s)P(A_{\text{trig}}|s)P(c\tau|A_{\text{trig}}, s) + P(b)P(A_{\text{trig}}|b)P(c\tau|A_{\text{trig}}, b) \\
 &= \{P(s|A_{\text{trig}})P(c\tau|A_{\text{trig}}, s) + P(b|A_{\text{trig}})P(c\tau|A_{\text{trig}}, b)\} P(A_{\text{trig}})
 \end{aligned}
 \tag{16}$$

So we can either, as in Equation 16, fit the probability to find a given efficiency function, or at least, as in Equation 17 calculate an event-by-event signal probability based on the efficiency function. The last term in Equation 17, describing the total probability to get the given efficiency function (whether it's signal or background), does indeed factor out and can be ignored, but if we ignore the kinematics altogether, we will get the event-by-event signal fractions wrong and hence the wrong fit result.

The same problem shows up for anything that changes our PDF event-by-event, be it the event-by-event efficiency functions, or event-by-event lifetime errors. The latter is the example used by Giovanni Punzi when he discusses this effect in [6].

7.2 The full likelihood with everything

Now that things are getting more complicated, it is worth starting from scratch, deriving the exact expression for the probability density function from first principles. We'll use the following notation:

- $P(A)$ “probability of A”
- $P(\bar{A})$ “probability of not A”
- $P(A, B)$ “probability of A and B”
- $P(A \text{ or } B)$ “probability of A or B”
- $P(A|B)$ “probability of A given B”

Some rules of manipulating probabilities: In the following, we'll basic rules of manipulating probabilities. Here's a reminder:

- A and B

$$P(A, B) = P(A)P(B|A) \quad (18)$$

- ... which leads to Bayes' theorem

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} \quad (19)$$

- A or B.

$$P(A \text{ or } B) = P(A) + P(B) - P(A, B) \quad (20)$$

We include the following measured quantities in our fit:

- The measured lifetime, t_o .
- The error estimate for the lifetime σ_t .
- The efficiency function A_{trig} , calculated from the decay kinematics and the trigger cuts.
- The mass, m .
- The track-configuration observed, trk . Basically how often we find, say, three tracks in the SVT compared to two. Used to fit the SVT's single track efficiency.

Since we only have triggered events, we want to calculate the probability of making these measurements, *given* the event passed the trigger:

$$P(t_o, \sigma_t, m, trk, A_{\text{trig}}|\text{trigger}) \quad (21)$$

It is important at this point, to distinguish between the probability of finding an acceptance function, $P(A_{\text{trig}})$, and the probability that the trigger triggers,

$P(\text{trigger})$. $P(A_{\text{trig}})$ depends on the decay kinematics only, it is simply the probability to find an event where the decay kinematics translate the trigger cuts to the given efficiency function. $P(\text{trigger}|A_{\text{trig}})$ is the probability that a decay with these kinematics passes the trigger. This includes integrating over all other quantities (decay times, masses, track configurations), for the given acceptance function. It is basically the denominator in 12. $P(\text{trigger})$ is the same, except that it now also requires the integration over all acceptance functions, i.e. this is what you'd calculate using an average acceptance function, for example derived from Monte Carlo. The difference between $P(A_{\text{trig}})$ on one side, and $P(\text{trigger})$, $P(\text{trigger}|A_{\text{trig}})$ on the other, is important, because $P(\text{trigger})$ and $P(\text{trigger}|A_{\text{trig}})$ depend on the mean lifetime, while $P(A_{\text{trig}})$ doesn't.

Now we have background, we separate Equation 21 into a signal and a background part. We use the letters s and b for signal and background. Using Equation 20, $P(s, b) = 0$, and $P(\text{sig}) = 1$:

$$P(t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) = P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) + P(b, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) \quad (22)$$

In the following, we focus on the first term on the right hand side in Equation 22, only:

$$P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) \quad (23)$$

The results for

$$P(b, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) \quad (24)$$

will be analogous.

So far, nothing has happened. Using Equation 19 on Equation 23 gives:

$$P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) = \frac{P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}) P(\text{trigger}|s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}})}{P(\text{trigger})} \quad (25)$$

Note that $P(\text{trigger}|s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}})$ is either 1 or 0, because the trigger decision is completely determined by the efficiency function, the decay time, and which tracks have actually been found by the SVT. The denominator in Equation 25 is the probability that the trigger fires - we would rather re-write this in terms of the event-by-event probability that the trigger fires *given* the acceptance function, $P(\text{trigger}|A_{\text{trig}})$. The lifetime error σ_t also enters the calculation of $P(\text{trigger}|A_{\text{trig}})$, and if we want to use the event-by-event lifetime error as in the numerator, we better find an expression in terms of $P(\text{trigger}|\sigma_t, A_{\text{trig}})$. And finally, it is easier to calculate this denominator for signal and background separately, so the aim is to find an expression in terms of $P(\text{trigger}|\sigma_t, A_{\text{trig}}, s)$. Using Bayes' theorem (19), we find for the denominator in Equation 25:

$$P(\text{trigger}) = P(\text{trigger}|A_{\text{trig}}, \sigma_t, s) \frac{P(A_{\text{trig}}, \sigma_t, s)}{P(A_{\text{trig}}, \sigma_t, s|\text{trigger})} \quad (26)$$

The left-hand term in the numerator of Equation 25 can be written as

$$P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}) = P(s, \sigma_t, A_{\text{trig}}) P(t_o, m, \text{trk}|s, \sigma_t, A_{\text{trig}}) \quad (27)$$

Putting these together (note the cancellation of $P(A_{\text{trig}}, \sigma_t, s)$), and abbreviating $P(\text{trigger}|s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}})$ as $P(\text{trigger}|\text{all})$, we get:

$$\begin{aligned} & P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) \\ &= \frac{P(t_o, m, \text{trk}|s, \sigma_t, A_{\text{trig}})P(A_{\text{trig}}, \sigma_t, s|\text{trigger})P(\text{trigger}|\text{all})}{P(\text{trigger}|A_{\text{trig}}, \sigma_t, s)} \end{aligned} \quad (28)$$

The first term in the numerator can be split further

$$\begin{aligned} & P(t_o, m, \text{trk}|s, \sigma_t, A_{\text{trig}}) \\ &= P(t_o|s, \sigma_t, A_{\text{trig}})P(\text{trk}|t_o, s, \sigma_t, A_{\text{trig}})P(m|\text{trk}, t_o, s, \sigma_t, A_{\text{trig}}) \end{aligned} \quad (29)$$

So far, we only used basic rules of manipulating probabilities, nothing else. Now we make some sensible assumptions:

- t_o , the measured lifetime, is independent of the A_{trig} , i.e. the decay kinematics. Remember that A_{trig} is all about decay kinematics, so $P(t_o|A_{\text{trig}})$ is not the probability of measuring t_o given the trigger, but it is the probability of finding t_o given the decay kinematics that translate trigger cuts into lifetime cuts, *before the trigger is applied*.
- trk , the number of tracks found by the SVT, is independent of σ_t , and A_{trig} . Again, remember that A_{trig} is all about decay kinematics, so $P(\text{trk}|A_{\text{trig}})$ is not the probability to find k out of n tracks in the SVT given the trigger, but it is the probability of finding k out of n tracks in the SVT given the decay kinematics that translate trigger cuts into lifetime cuts.
- m , the reconstructed mass, is independent of trk , t_o , σ_t , A_{trig} .

With this we get:

$$P(t_o, m, \text{trk}|s, \sigma_t, A_{\text{trig}}) = P(t_o|s, \sigma_t, A_{\text{trig}})P(\text{trk}|t_o, s)P(m|s) \quad (30)$$

so our PDF is now:

$$\begin{aligned} & P(s, t_o, \sigma_t, m, \text{trk}, A_{\text{trig}}|\text{trigger}) \\ &= \frac{P(t_o|s, \sigma_t)P(\text{trk}|t_o, s)P(m|s)P(A_{\text{trig}}, \sigma_t, s|\text{trigger})P(\text{trigger}|\text{all})}{P(\text{trigger}|A_{\text{trig}}, \sigma_t, s)} \end{aligned} \quad (31)$$

Finally, we'll have to deal with the second but last term in the numerator, $P(A_{\text{trig}}, \sigma_t, s|\text{trigger})$. Note that the condition “|trigger” ensures that we need to look only at quantities as they are distributed *after* the trigger, which is also all we have access to. There are several different ways in which this term could be disentangled

1. $P(s|\text{trigger})P(\sigma_t|s, \text{trigger})P(A_{\text{trig}}|s, \text{trigger})$, where the first term is simply the overall signal fraction after the trigger, i.e. in the data we see. The other terms fit the σ_t and A_{trig} distribution (how to fit a distribution of acceptance functions is the subject of an entire section later on). Here we assume that A_{trig} and σ_t are independent.

2. $P(A_{\text{trig}}, \sigma_t | \text{trigger})P(s|A_{\text{trig}}, \sigma_t, \text{trigger})$. Here we can ignore the first term, as it does not depend on the parameters we are interested in, and it is the same for signal and background. The second term is a signal fraction as a function of A_{trig} and σ_t . While there are other possible disentanglements this turns out to be the default solution that we choose. Since the acceptance function and σ_t are both quantities computed from Lxy or σ_{Lxy} divided by P_T , there will be a correlation between these two terms. We cannot further disentangle the second term without having to model this correlation as well as any other terms that arise. So we keep the one term $P(s|A_{\text{trig}}, \sigma_t, \text{trigger})$. We use fisher discriminants to model this term, but this is complicated enough to deserve its own section 8.

With this, the final version of the signal-part of the total PDF is:

$$\begin{aligned}
& P(s, t_o, \sigma_t, m, trk, A_{\text{trig}} | \text{trigger}) \\
&= P(s | A_{\text{trig}}, \text{trigger}) \times \\
& \quad \frac{P(t_o | s, \sigma_t) P(\sigma_t | s, \text{trigger}) P(m | s) P(trk | t_o, s) P(\text{trigger} | \text{all})}{P(\text{trigger} | A_{\text{trig}}, \sigma_t, s)}
\end{aligned} \tag{32}$$

The different terms in the PDF are listed below. Where background has a different model this is also described:

$P(t_o | s, \sigma_t)$ This is the probability of measuring a lifetime t_o , given a lifetime error σ_t , for signal events with a mean lifetime τ . It is given by

$$P(t_o | s, \sigma_t) = \frac{1}{\tau} e^{-\frac{t_o}{\tau} + \frac{1}{2} \frac{\sigma_t^2}{\tau^2}} \text{F} \left(\frac{t_o}{\sigma} - \frac{\sigma}{\tau} \right) \tag{33}$$

$P(t_o | b, \sigma_t)$ The background lifetime component is modelled as a sum of events coming from 3 “lifetimes” τ_i each with their own relative weight. These are left to float in the final fit.

$P(m | s)$ The signal mass distribution is by a 2 Gaussians with different means and widths. Further details of this mass model and why it is chosen are described in another section 13.

$P(m | b)$ The background mass distribution is a first order polynomial. Reasons for this choice are detailed in section 13.

$P(trk | t_o, s)$ The probability to find the given track configuration in the SVT, expressed in terms of the SVT’s single track finding efficiency, ε_s . It is given by

$$P(trk | t_o, s) = P(\text{SVT found } k \text{ tracks out of } n | t_o, s) = \varepsilon_s^k (1 - \varepsilon_s)^{(n-k)} \tag{34}$$

where k is the number of tracks found in the SVT, and n is the number of tracks available to be found; n is smaller or equal to the total number of tracks in the final state. A track is “available” if it has $|d_0^{\text{off}}| < 1$ mm and

$p_t > 2 \text{ GeV}$. Note that this is not the usual “ n over k ” expression, because we are not asking for the probability that *any* k out of n tracks are found, but that those specific k tracks that have SVT matches are found, and the others not.

$P(\text{trigger}|\text{all})$ The probability that the trigger fires, given all measured quantities. This is simply one or zero:

$$P(\text{trigger}|\text{all}) = \begin{cases} 1 & \text{if event passes trigger cuts} \\ 0 & \text{else} \end{cases} \quad (35)$$

So its value is 1 for all events in the sample and could be omitted. It is however useful a term to keep in mind if one wants to calculate the PDF for values of, say, $c\tau$, not actually found in the event, for example if one wants to integrate the expression.

$P(\text{trigger}|A_{\text{trig}}, \sigma_t, s)$ The probability that the trigger fires, given the efficiency function A_{trig} and the event-by-event error estimate σ_t . It’s the normalisation factor. It’s given by:

$$\begin{aligned} & \sum_{\text{all trk}} \sum_{\substack{i=\text{all} \\ \text{intervals}}} \int_{t_{\min i}}^{t_{\max i}} P(\text{trk}|t_o, s) P(\text{trigger}|\text{all}) \frac{1}{\tau} e^{-\frac{t_0}{\tau} + \frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F} \left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau} \right) dt_0 \\ & = \sum_{\substack{i=\text{all} \\ \text{intervals}}} \text{poly}_i(\varepsilon_s) \left[-e^{-\frac{t}{\tau} + \frac{1}{2} \frac{\sigma^2}{\tau^2}} \text{F} \left(\frac{t}{\sigma} - \frac{\sigma}{\tau} \right) + \text{F} \left(\frac{t}{\sigma} \right) \right]_{t=t_{\min i}}^{t=t_{\max i}} \quad (36) \end{aligned}$$

$P(s|A_{\text{trig}}, \sigma_t, \text{trigger})$ The signal fraction as a function of the acceptance function. This is tricky because it involves a function of a function, rather than the usual function of a parameter. The strategy we follow is, to characterise the the the acceptance function with a single number and then evaluate signal/background probabilities as a function of this number. In doing so, we need to make sure that

$$P(s|\text{Number}(A_{\text{trig}})) \approx P(s|A_{\text{trig}}) \quad (37)$$

to a good-enough approximation. This means that the characteristic number must be chosen such as to minimise the information loss in the process $(A_{\text{trig}}) \rightarrow \text{number}$, in terms of the signal-ness or background-ness of the acceptance function. A relatively simple number to calculate that is very good at separating signal from background, i.e. at minimising the information loss regarding the signal-ness or background-ness of the acceptance function, is the Fisher discriminant. In the following we describe how we associate a Fisher discriminant to each acceptance function, using data only, and then how we use this to calculate $P(s|A_{\text{trig}}, \text{trigger})$. We test the method by fitting a mixture of MC signal and background from data.

8 Fisher Discriminant to calculate $P(s|acc, \sigma_{c\tau})$

$P(s|acc, \sigma_{c\tau})$: The more complicated term to deal with is the acceptance function. The use of fisher discriminants is introduced to deal with this term. Let

us first consider the term $P(s|acc)$, and then later show the simple extension to $P(s|acc, \sigma_{c\tau})$

8.1 The use of Fisher Discriminants

If our acceptance function were characterizable using a set of variables then a Fisher Discriminant Analysis could be used to separate signal and background in the way required above. Our acceptance functions typically have the form of top hat functions over intervals of $c\tau$, so we can sample the height of the function at N points, creating a vector v_i with N entries. Each entry is a variable describing the shape of the acceptance function at a given value of $c\tau$. In the Fisher Discriminant Analysis we find the N -component projection vector w and use it to form the scalar product $w.v_i$ for every event. We name the scalar product $w.v_i$ the Fisher scalar and the distribution of the fisher scalar is parameterized to separate signal and background in much the same way as for invariant mass or any other kinematic variable. A detailed description follows.

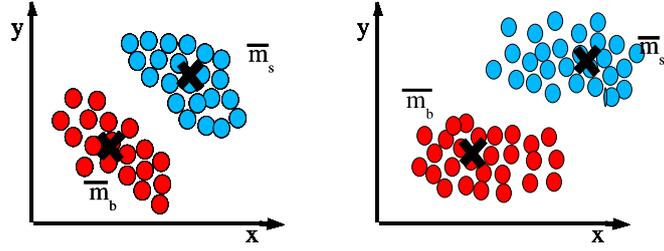
8.2 Basics of Fisher Linear Discriminant Analysis

Imagine two classes of events, eg signal and background with their own distributions of variable x and y as shown in figures 10(a), 10(b) and 10(c). The means of variable x and y for each distribution are shown as the points $\overline{m}_s = \frac{\overline{x}_s}{\overline{y}_s}$ and $\overline{m}_b = \frac{\overline{x}_b}{\overline{y}_b}$. We are looking for a linear direction w on which to project these events such that value of the projected point along w provides the best discriminator between signal and background. From the diagrams we can conclude that the best projection direction is one where the distance between the projected means of each class of event is large while the spread around each mean remains small.

Firstly we consider the square of the separation of projected means of signal and background events along the projection direction. This is given in equation 38, and gives the definition for the matrix we refer to as S_M .

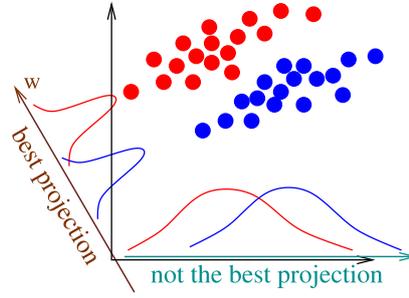
$$\begin{aligned} (\langle w|\overline{m}_s \rangle - \langle w|\overline{m}_b \rangle)^2 &= \langle w|(\overline{m}_s - \overline{m}_b) \rangle \langle (\overline{m}_s - \overline{m}_b)|w \rangle \\ &= \langle w|S_M|w \rangle \end{aligned} \quad (38)$$

Secondly we consider the square of the spread of the signal events around the projected mean, $Scat_{sig}^2$ which leads to the definition of the matrix S_{sig} as shown in 39, where $p_i = (\frac{x_i}{y_i})$. There is a similar expression for the background events,



(a) The best separation here is most likely along the means

(b) Here the separation along the y axis is better than along the means



(c)

Figure 10: This illustrates that it is necessary to take the means and spread of each variable in finding the direction of best separation

S_{bg} .

$$\begin{aligned}
 Scat_{sig}^2 &= \sum_{All\ signal\ events} (\langle w|p_i \rangle - \langle w|\bar{m}_s \rangle)^2 \\
 &= \sum_{All\ signal\ events} \langle w|(p_i - \bar{m}_s) \rangle \langle (p_i - \bar{m}_s)|w \rangle \\
 &= \sum_{All\ signal\ events} \langle w|S_{sig}|w \rangle
 \end{aligned} \tag{39}$$

It is clear that the best projection direction is one in which the means of the two types of events fall far apart but simultaneously the spread is small. This is Fisher's criterion and is expressed mathematically as finding the w for which $J(w)$ is maximized, where $J(w)$ is given below 40.

$$J(w) = \frac{\langle w|S_M|w \rangle}{\langle w|(S_{sig} + S_{bg})|w \rangle} = \frac{\langle w|S_M|w \rangle}{\langle w|(S_w)|w \rangle} \tag{40}$$

From equations 41 and 42 we find that by maximizing this condition we are

left simply with an eigenvalue equation. Furthermore using the definition of the matrix S_M we can simplify the equation and remove the need to find the actual eigenvalues and just use the inverse of S_w and the vector $(\overline{m_s} - \overline{m_b})$ to find the vector w .

$$\nabla w(J(w)) = \frac{2S_M|w\rangle}{\langle w|S_w|w\rangle} - \frac{\langle w|S_M|w\rangle}{\langle w|S_w|w\rangle} \cdot \frac{2S_w|w\rangle}{\langle w|S_w|w\rangle} = 0 \quad (41)$$

$$\begin{aligned} S_M|w\rangle - \lambda S_w|w\rangle &= 0 \\ S_M|w\rangle &= \lambda S_w|w\rangle \\ S_w^{-1}S_M|w\rangle &= \lambda|w\rangle \end{aligned} \quad (42)$$

Using the definition of S_M from equation 38 we can rewrite $S_M|w\rangle$ in the following way.

$$\begin{aligned} S_M|w\rangle &= |(\overline{m_s} - \overline{m_b})\rangle \langle (\overline{m_s} - \overline{m_b})|w\rangle \\ S_M|w\rangle &\propto |(\overline{m_s} - \overline{m_b})\rangle \end{aligned} \quad (43)$$

If we insert this into equation 41 we see that it is not necessary to solve for the eigenvalues and that all we require to find $|w\rangle$ is the inverse of S_w and $|(\overline{m_s} - \overline{m_b})\rangle$.

$$S_w^{-1}|(\overline{m_s} - \overline{m_b})\rangle \propto |w\rangle \quad (44)$$

The value of the discriminating variable is given by the scalar product of the event vector (x_i, y_i) in this case and w , and this fisher scalar variable is the best one for distinguishing between the two classes of events.

While the diagram illustrates the technique for 2 variables only, the mathematics is general and hence we can extend this to any number of variables and the matrices S_M or S_w are just expanded to $n \times n$ square matrices, and the vectors w etc grow to length n too.

8.3 Using the Fisher Scalar Distribution to calculate signal probability

Imagine that we could quantify the acceptance functions as a set of variables. Then we could use fisher discriminant analysis to find the direction of best separation. One way in which we could do this is to turn the acceptance function into a column vector v_i . A detailed description on how this is achieved follows in the next section. Each row of the vector would be a different variable treated similar to x and y as above. We find the vector \mathbf{w} as above and then $\mathbf{w} \cdot \mathbf{v}$ would be a discriminating variable, that we could use to give us the probabilities $P(s|A_{trig}, trigger)$ and $P(b|A_{trig}, trigger)$. We call the variable $\mathbf{w} \cdot \mathbf{v}$ the fisher scalar.

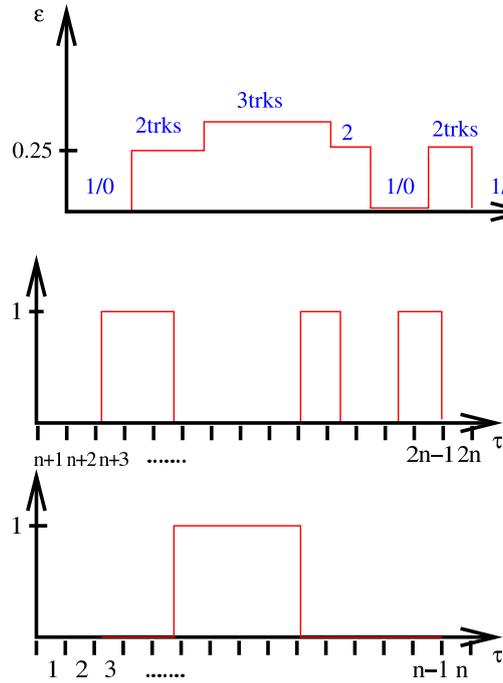


Figure 11: Splitting the acceptance function into the sum of its parts

8.3.1 Acceptance function \rightarrow Vector

Our acceptance functions are like a series of top hats added together. We can draw them in a histogram by plotting the acceptance probability as a function of $c\tau$ for every event. A typical acceptance function may look like the one in figure 11 where the differing heights are regions in which there are 2,3 (or more) tracks with IP in the region where it could play a part in the trigger decision. We can write this one histogram as the sum of the histogram that contains the sections where there are only two tracks and the histogram that contains the section where there were are three tracks and so on. The heights of the acceptance function are set to one and each histogram is multiplied by the relevant probability polynomial as described in section 6.4.2. This is illustrated in figure 11. So for each event there will be t types of histogram where t maybe 1,2 (or more) depending on the number of tracks in the final state of the decay.

The histograms are binned finely over a very large range (-500 to 10,000 microns) to ensure all parts of the acceptance function are included. We then find the minimum and maximum bin over the whole dataset, and then rebin each histogram into a smaller number of n bins over the new range. Typically this number is 20 for 3 track decays and 10 for 4 track decays. The height of the first histogram in each of the n bins provide the values for the first n bins of the vector. We then move onto the next histogram and the height of its bins provide the next n entries into the column vector. We arrived at the choice of 20 and 10 through testing on MC signal and background mixes. To try and preserve as

much information as possible it is desirable to use an increased number of bins. However we have found that using too many bins has caused problems during the inversion of a matrix. One eigenvalue can become numerically close to zero and this stops accurate inversion of the matrix. We found that 20 bins was a good choice for 3 track decays; in tests it did not appear to cause a lifetime shift, nor did we encounter errors during matrix inversion. For a 4 track decay the total acceptance function is split over more histograms and if we used 20 bins we found that we were running into these inversion problems. We found that reducing the number of bins used to describe the acceptance histograms to 10 was a sensible choice; again there were no indications of significant shift in lifetime and the numerical problems were removed. While we have chosen 20 and 10 to be our default number of bins per acceptance histogram for 3 and 4 track decays respectively, we do realise that this choice is a little arbitrary and so we include this choice as a systematic. This choice of bins actually makes the vector sizes for 3 and four track decays the same. There is only one further change made to this vector which will be explained in a later section.

Now that we have each acceptance function as an acceptance vector, v_i , for each event, and it is of length $n*t$. As there are $n*t$ variables the matrices S_M and S_w are of dimension $n*t \times n*t$. We can now consider how to find these matrices and the vector $|\overline{m}_s - \overline{m}_b \rangle$ so that we can find the vector w and hence find the fisher scalar for every event.

Since each row is treated as an independent variable, we can simply add one further variable, $\sigma_{c\tau}$. Now that $\sigma_{c\tau}$ is part of the overall vector, the fisher direction takes into account both the Acceptance function and $\sigma_{c\tau}$. Furthermore as they are both in the same calculation of fisher direction any correlation between the two quantities is accounted for.

8.3.2 Extracting ($|\overline{m}_s - \overline{m}_b \rangle$) from the dataset

If we had a sample of events which we knew, a priori, were signal and another that we knew were background, making this vector is a trivial exercise. However we can use a mass only fit to the data to define two regions; a sideband region and a signal region. We can assume that all the events in the sideband region are background events and that this background is typical of all the background in the sample. We can find $|\overline{m}_b \rangle$ by simply summing all the acceptance vectors in the sideband region and dividing by the number of events in this region.

The equivalent vector for events in the signal region is called $|\overline{m}_r \rangle$ which we can write as equation 45 where we know f_s from the mass fit. f_s is the fraction of signal in the signal region. The vector ($|\overline{m}_r - \overline{m}_b \rangle$) is proportional to ($|\overline{m}_s - \overline{m}_b \rangle$). This is our vector for the difference between the means.

$$\begin{aligned} \overline{m}_r &= \frac{\sum_{\text{events in signal region}} |v_i \rangle}{\sum_{\text{events in signal region}} 1} \\ |\overline{m}_r \rangle &= f_s |\overline{m}_s \rangle + (1 - f_s) |\overline{m}_b \rangle \\ |\overline{m}_r - \overline{m}_b \rangle &= f_s (|\overline{m}_s \rangle - |\overline{m}_b \rangle) \end{aligned} \tag{45}$$

We may find that some of the variables in the vector $|\overline{m}_s - \overline{m}_b\rangle$ have value 0, which means that the variable in that entry can provide little discriminating power. Keeping these variables in the vector turns out to cause problems during matrix inversion and so we truncate ($|\overline{m}_s - \overline{m}_b\rangle$) by removing rows where the entry in ($|\overline{m}_s - \overline{m}_b\rangle$) is 0. We also remove the corresponding rows from the individual acceptance vector v_i so that all the vectors have the same dimension.

8.3.3 Finding S_W

We wish to find S_w which can be written as 46 where v_s and v_b are the acceptance vectors of pure signal and pure background events respectively. The definition is taken from equations 39 and 40.

$$S_w = \sum_{\substack{\text{Signal} \\ \text{events}}} |(v_s - \overline{m}_s)\rangle\langle (v_s - \overline{m}_s)| + \sum_{\substack{\text{Background} \\ \text{events}}} |(v_b - \overline{m}_s)\rangle\langle (v_b - \overline{m}_s)| \quad (46)$$

Again we consider the signal region and background region.

We can calculate the matrix called S_{bk} as given in equation 47 trivially as we already have \overline{m}_b .

$$S_{bk} = \sum_{\substack{\text{Sideband} \\ \text{events}}} |(v_i - \overline{m}_b)\rangle\langle (v_i - \overline{m}_b)| \quad (47)$$

We can also calculate the matrix called S_{bkassg} given in equation 48. We calculate $|\overline{m}_s\rangle$ as we know the value of f_s from the mass fit and we know $f_s^*|\overline{m}_s - \overline{m}_b\rangle$ and $|\overline{m}_b\rangle$.

$$S_{bassig} = \sum_{\substack{\text{Sideband} \\ \text{events}}} |(v_i - \overline{m}_s)\rangle\langle (v_i - \overline{m}_s)| \quad (48)$$

The matrix $S(\text{sg} + \text{b})$ is calculated in 49.

$$\begin{aligned} S_{sg+b} &= \sum_{\substack{\text{Signal} \\ \text{region} \\ \text{events}}} |(v_i - \overline{m}_s)\rangle\langle (v_i - \overline{m}_s)| \\ &= \sum_{\substack{\text{Signal} \\ \text{events in signal} \\ \text{region}}} |(v_s - \overline{m}_s)\rangle\langle (v_s - \overline{m}_s)| \\ &+ \sum_{\substack{\text{background} \\ \text{events in signal} \\ \text{region}}} |(v_b - \overline{m}_s)\rangle\langle (v_b - \overline{m}_s)| \end{aligned} \quad (49)$$

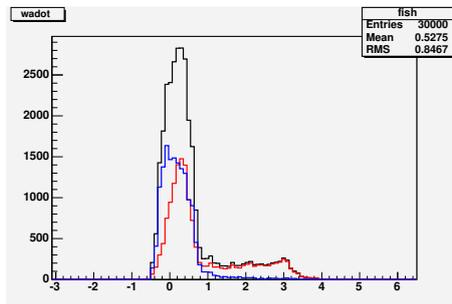


Figure 12: The distribution of fisher scalar is shown here from Monte Carlo is red and that from background is blue. It is clear that their distributions are different.

We can calculate the 3 matrices in equations 47, 48 and 49 by considering the signal fraction under the peak, events in the signal region and events in the background sideband. We can combine these matrices together to give us S_W as shown in 50.

$$S_w = S_{sg+B} - \frac{N^0 \text{ Background in sideband}}{N^0 \text{ Background in signal region}} \times S_{bassig} + \frac{N^0 \text{ Background in dataset}}{N^0 \text{ Background in sideband region}} \times S_{bk} \quad (50)$$

In practice, we optimize the procedure slightly by using the event-by-event signal probability derived from the mass fit, i.e. we use the information that events near the center of the B mass peak have a larger signal probability than those at the edges of our signal window.

8.4 Using the fisher variable to get signal probability

To verify the procedure, we apply it to a mix of signal MC events and background from data. The background events from data come from the upper sideband. In order to make the fit realistic, the reconstructed masses of the background events are shifted so that they lie underneath the signal mass peak. For the purpose of defining the Fisher direction and calculating the Fisher discriminant, the data-MC mix was used like any other data sample, the information which event came from data and which from MC was not used. We see that the two classes of events are separated, the blue events coming from background and the red from signal. Dividing the signal by the total we can see the distribution of signal fraction as a function of signal.

We model this distribution using the Lagrange interpolating polynomials. Their parameters are the value $P(s|\text{Fisher} - \text{scalar})$ for certain values of the Fisher scalar, which are then smoothly interpolated - for details see [5].

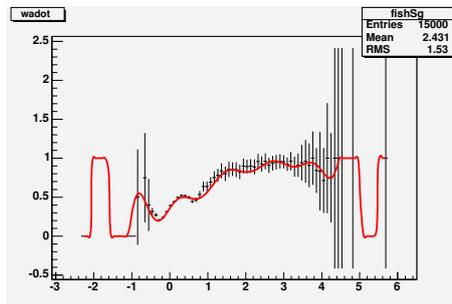


Figure 13: The data points show the truth signal fraction in the fit. The red line shows the fitted function. The data points are shown to demonstrate that the correct signal fraction has been found. The fit itself does not know which events are signal or background so does not know the truth, yet manages to match it well.

We can fit for the height of this function at regular intervals of the fisher variables. The distribution is binned and the signal fraction in each bin is a fit parameter. An example of this function after the fit is performed is shown in Figure 13. The fit parameters are constrained to lie between 0 and 1 as this is a probability distribution function. The fit did not know which event were signal or background. As we see from the figure the fit matches the truth distribution very well. The data is binned as a function of fisher scalar, and the signal fraction in each bin is a fit parameter. Again we use Monte Carlo and background data mixed together to find how many bins the data needs to be split into so that the distribution is well modelled. By construction the higher the number of bins that are used, the higher the degree of polynomial that Lagrange Interpolating Polynomials uses to fit the distribution. To fit the truth well we found we had to use about 15 - 20 bins, and pick 18 as a default. In doing this we do introduce some fluctuations at the end of the distribution but as there are so few events in these regions we do not expect this to cause any pull in the best fit lifetime result. The choice in number of bins (and therefore order of polynomial) is tested as a systematic error.

We fit each bin for the signal fraction, and the advantage of using the Lagrange Interpolating Polynomials is that the probability changes smoothly across the bin instead of jumping at the bin edge. The disadvantage is that in the tails of the fisher scalar distribution where the statistics in each bin are low the function is poorly behaved as it is not pinned down well. This is seen in 13. To improve this we make a small change to start and end the interpolating polynomial in the region of high statistics and use a single bin either side to fit the tails. After implementing this in a few test MC fits and Data we saw no shift in the fit result, which is not surprising seeing as there were very few events in the tails. However the plot is more pleasing and the overall function to describe the signal fraction as a function of fisher scalar and $\sigma_{c\tau}$ is well behaved. The figure 14 shows this better behaved behaviour.

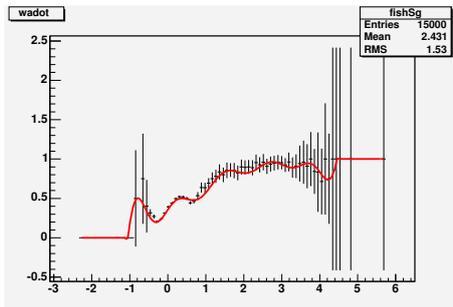


Figure 14: The data points show the truth signal fraction in the fit. The tails are now modelled by a single bin which gives a better behaved fit function although there was no change in the lifetime result..

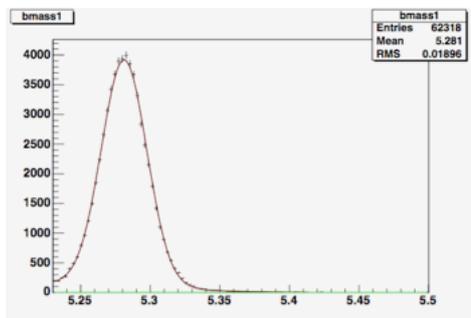


Figure 15: We find that realistic MC is fit well by 2 Gaussians

9 Modelling the Mass

In this section we give some detail to the mass model for signal and background. The mass fit is important for two reasons. Firstly it is present in the full Likelihood expression and is a significant discriminant between signal and background. Secondly the initial step in the fit is to do a mass only fit so that we can make the fisher vector. The mass distribution is an important part of making this fisher vector.

We discuss in detail the B0 mode, but the same model is used for both modes

All we need to do is model the shape for signal and the shape for background so there is a good fit to the data. In both Bu and B0 there are 3 classes of events that we classify as signal. These are the main peak, some events where one or more photons were radiated and also some presence of the cabibbo suppressed mode; B to DK. We treat all these three types of events as signal, and do not distinguish between them. We have Monte Carlo that contains $B \rightarrow D\pi$ and $B \rightarrow D\pi(n\gamma)$ and find that over our fit range of 5.23 to 5.5 the signal mass is well fit by 2 Gaussians where the means and widths are allowed to float. 9

In our data the contribution of events that radiated photons may be different

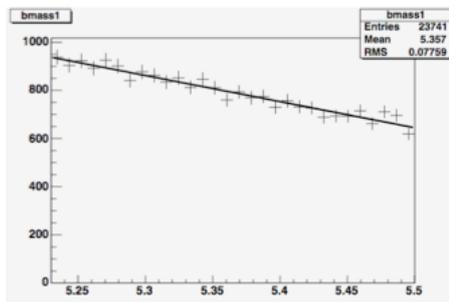


Figure 16: The wrong sign distributions looks like a first order polynomial

and furthermore there is the cabibbo suppressed mode. We try the same two Gaussian model for data, expecting the floating parameters to adjust themselves for these differences.

The shape of the combinatoric background should have the same shape as the wrong sign combination of $B' \rightarrow D - \pi^-$. We examined the mass distribution of this reconstructed data over the fit range 5.23 to 5.5. It is well described by a first order polynomial. We only look at the wrong sign to decide upon a sensible shape for our model. We do not fix the slope in our data from the wrong sign distribution.

Putting this all together

$$P(m|s) = f_1 \times \text{Gauss}(m|m_1, \sigma_1) + (1 - f_1) \times \text{Gauss}(m|m_2, \sigma_2) P(b|s) = 1 + \beta m \quad (51)$$

The functions Gauss and the polynomial have been normalized over the restricted mass range 5.23 -5.5GeV. So in total there six parameters introduced by this model. They are $m_1, m_2, \sigma_1, \sigma_2, f_1, \beta$ In the initial mass fit there is an extra parameter f_s that fits the fraction of signal and background. In the final lifetime fit the signal fraction is taken care of by the term $P(s|Acc, \sigma_t)$

We fit the mass parameters alone in the initial mass fit and then hold them constant in the time fit. We find that this model fits the data well. Plots are shown in the result section.

The extraction of the lifetime is done in a two step fit. The mass distribution is fit first, the results of this fit are used to weight signal-like and background-like acceptance functions which are then used in the calculation of the Fisher discriminant. The second part of the fit uses the Fisher scalar distribution and the fully reconstructed proper-decay length to discriminate signal from background. There are seven parameters in the initial mass fit, there are two gaussians for the signal, hence there is a mean and width for each gaussian, a fraction of events falling in the first gaussian and a first order polynomial which describes the background, a requirement of normalization reduces the number of background parameters to 1 and there is a fraction of events that are signal. As described in section 8.4 once the Fisher scalar is calculated from the mass fit the signal

fraction from the mass fit is no longer used, rather the Fisher scalar distribution is used to determine the probability of each event being signal. It should be clear to the reader that the uncertainties in the mass fit are propagated to the second part of the fit. A description of this follows:

The mass fit parameters are highly correlated with one another and so varying the parameters by $\pm 1\sigma$ from the first part of the fit and noting the shift in lifetime in the second part of the fit is an incomplete treatment. It is necessary to take the correlations into account and so the full error matrix of the first part of the fit is used. The row and column corresponding to the signal fraction is removed from this error propagation since the signal fraction from the mass fit is not used explicitly in the second part of the fit. The rotation matrix that diagonalizes the error matrix is found, this rotation matrix represents the transformation from the original set of mass fit parameters to their uncorrelated linear combinations (eigenvectors of the error matrix). Errors derived from the diagonal error matrix of the mass fit represent the uncertainty on the uncorrelated linear combinations, and these are transformed back using the rotation matrix to produce appropriately correlated variations on the mass fit parameters. The second part of the fit is repeated with the mass fit parameter varied, it should be clear to the reader that 6 such variations are performed and the shifts are added in quadrature to the error in best fit lifetime from the second part of the fit.

As this procedure requires 6 further iterations of the lifetime fit for every fit result we only carry out this procedure for the final two results, namely the 1fb^{-1} lifetime for the charged and neutral B meson. We do find that the error due to the mass fit is small, it is shown in the results section.

10 Analysis cuts for $B^\pm \rightarrow D^0\pi^\pm$, $D^0 \rightarrow K^\mp\pi^\pm$ and for the $B^0 \rightarrow D^\mp\pi^\pm$, $D^0 \rightarrow K^\mp\pi^\pm\pi^\pm$ decay modes

In this section we define the analysis cuts used to select $B^\pm \rightarrow D^0\pi^\pm$ and $B^0 \rightarrow D^\mp\pi^\pm$ decays. The same cuts are used for testing Monte Carlo and analysing data. When there has been reason to depart from this for testing Monte-Carlo we have mentioned this explicitly.

10.1 Track Quality Cuts

The following cuts are applied to all tracks from all modes:

- Each track has transverse momentum P_T greater than 0.35 Gev.
- Each track is required to have hits in a minimum of 5 COT axial super-layers and 5 COT stereo super-layers.
- Each track is required to have hits in a minimum of 3 silicon R- Φ layers.

- Each individual track is required to have an $\eta < 2$

We use the xbh0d, xbh0h, and xbh0i datasets which are fed from the hadronic B trigger. We begin by reconstructing a charged or a neutral D and then combining it with a candidate track with a pion mass hypothesis to form a B candidate. Selection cuts are applied on the D s and the fully reconstructed B s. The final reconstructed quantities are obtained from the AC++ wrapped CTVMFT vertexing program, using version 6.1.4 of CDF software and pass 17 of the alignment. All information from L00 of the Silicon detector is dropped.

The selection cuts themselves are detailed in the following subsections.

10.2 Selection cuts for the $B^\pm \rightarrow D^0 \pi^\pm$, with $D^0 \rightarrow K^\mp \pi^\pm$

We begin reconstructing D^0 candidates in the $K^\mp \pi^\pm$ mode by combining all opposite track combinations assigning them the mass of a K and π .

The following cuts are then applied on D^0 candidates assumed to decay in the mode: $D^0 \rightarrow K^\mp \pi^\pm$

- Oppositely charged track pairs are assigned the mass of the K or π .
- The raw mass of the D^0 must lie between 1.81 and 1.92
- The angular separation in ϕ between the K candidate and the flight path of the D^0 is ≤ 1.5 radians.
- The D daughters lie in a cone defined by $\Delta R = \sqrt{(\Delta\eta)^2 + \Delta\phi^2} < 2$
- The transverse momentum of the D^0 is ≥ 2.4 GeV.
- The scalar sum of π^\mp and K^\pm transverse momenta is ≥ 2.4 GeV
- The K^\pm and π^\mp P_T s are each individually ≥ 0.4 GeV

Next we loop over all tracks in the event with charge opposite to the π^\mp from the D^0 that are not its daughters and assigning them the mass of a π and constrain the 3 tracks to a common vertex this is our B^\pm candidate on which the following selection criteria are applied:

- The reconstructed B mass lies between 5.23 and 5.5 GeV
- The transverse flight distance of the B in the direction of its \vec{P}_T (L_{xy}) is $> 350\mu\text{m}$ and $<$ than 1 cm
- The candidate B vertex $\chi^2 < 15$
- The P_T of the π^\pm from the B^\pm is ≥ 1 GeV.
- The impact parameter of the B with respect to the beam spot is $< 80\mu\text{m}$.

- The angular separation in ϕ between the B and its π daughter is ≤ 3.0 radians
- The momenta of the D and π from the B lie within a cone defined by $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} < 2$
- All the B daughters have a z_0 within 5cm of each other.
- The B^\pm transverse momentum (P_T) ≥ 5.5
- The scalar sum of all B daughter charged tracks P_{TS} is ≥ 5.0
- The calculated uncertainty of the proper decay time ($\times c$) of the B , $\sigma_{c\tau}$ is less than $100 \mu\text{m}$

10.3 Selection cuts for $B^0 \rightarrow D^\mp \pi^\pm$, with $D \rightarrow K^\pm \pi^\mp \pi^\mp$

We begin reconstructing D^\pm candidates in the $K^\mp \pi^\pm \pi^\pm$ mode by combining all 3 track combinations with two of like charge and one with charge opposite to this. The two like charge tracks are assigned the π mass and the third track the mass of the K .

The following selection criteria are then applied to the candidate D^\pm :

- The D^\pm raw mass (before vertex constraining) lies between $1.81 >$ and < 1.92 .
- The angular separation in ϕ between the D flight path and the K daughter is ≤ 1.5 radians.
- The D daughters lie in a cone defined by $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} < 2$.
- The transverse momentum of the D , is $\geq 2.4\text{GeV}$

Once the D^\pm candidates are reconstructed, a fourth track with charge opposite to that of the π daughters of the D^\pm is combined and constrained to a common vertex along with all the daughters. This is the candidate B^0 , the following selection criteria are applied:

- The mass of the B^0 lies between 5.23 and 5.5 GeV .
- The L_{xy} of the B^0 is $> 350\mu\text{m}$ and $< 1\text{cm}$.
- The B^0 vertex χ^2 is < 30 .
- The P_T of pion from $B \geq 1 \text{ GeV}$.
- Impact Parameter of $B < 80\mu\text{m}$.
- The calculated error of the B s proper decay length $\sigma_{c\tau} < 100\mu\text{m}$.
- The angular separation in ϕ between the B^0 and its daughter pion is ≤ 3.0 radians.

- The momenta of the D and π from the B lie within a cone defined by $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} < 2$
- All the B daughters have $z0$ s within $5cm$ of each other.
- The P_T of the B is ≥ 5.5

In addition to the criteria listed above we note that the B^0 sample is contaminated by the decay modes $\Lambda_B \rightarrow \Lambda_C X$ and $B_s \rightarrow D_s X$, these are vetoed and their removal is accomplished by the following criteria:

Contamination from $\Lambda_B \rightarrow \Lambda_C X$ decays occurs when we assume a proton from a $\Lambda_C \rightarrow p\pi K$ decay is a pion from a D^\pm decay. To remove Λ_B s we remove the daughter Λ_C : we reconstruct the D as if it were a Λ_C by assuming the pion is proton. We look at both combinations that are possible and if the final mass is within 25MeV of the Λ_C mass we remove the candidate from the sample.

Contamination from $B_s \rightarrow D_s X$ decays occurs when we mis-reconstruct a $D_s^\pm \rightarrow \phi\pi^\pm$ ($\phi \rightarrow K^+K^-$) as a $D^\pm \rightarrow \pi^\pm\pi^\pm K^\mp$ by mis-assigning track masses.

We remove D_s^\pm by rejecting candidate D^\pm whose oppositely charged daughter tracks are consistent with coming from a ϕ with $\phi \rightarrow K^+K^-$.

We remove those candidates where an oppositely charged daughter track pair of the D^\pm yields an invariant mass within 10MeV of the ϕ mass when each daughter is assigned a kaon mass.

11 Testing the technique: Data Monte-Carlo Mix

We have already demonstrated that the method of removing bias works in signal only Monte Carlo and we have detailed the full PDF necessary to extract a lifetime from real data in section 2

In this section we demonstrate that using the PDF derived from our technique in a multivariate log likelihood minimization extracts the correct lifetime when we mix signal only Monte-Carlo with background from real data. We do this for each decay mode for which we have results using the data.

All the results quoted for Bu are for 15000 signal-only Monte-Carlo mixed with 15000 background events from data. The B0 fits in this section were done with 1000 signal from Monte Carlo and 20000 events background events.

The real-data background events come from the upper sideband of the reconstructed decay and the masses of the events are reassigned so that they are distributed over the full mass range that we expect to perform our fit over. We believe that data in the upper sideband is a close approximation to the background underneath the peak, as background should consist mainly of combinatorics and analysis cuts are chosen to remove physics background.

For the Bu and B0 decays we have taken events in the mass region 5.37 and 6 GeV to be the upper sideband. These events have their mass reassigned randomly into the fit region 5.23-5.5 GeV. The masses are distributed according to a linear polynomial of negative slope.

The MC and data are mixed together to make one sample and again the overall signal to background ratio is chosen so that it is realistic of the mode we are testing.

The fit is performed over the mixed sample, and the fit result is compared to that of the same Monte Carlo sample only for a consistency check. We have found very good agreement between the results of our fit on signal-only Monte-Carlo and for a mixture of data-background and signal only Monte-Carlo. Only once we have performed this test do we move on to looking at data-only.

The results of these tests now follow.

11.1 Data Monte-Carlo Results: $B^\pm \rightarrow D^0\pi^\pm$, $D^0 \rightarrow K^\mp\pi^\pm$

We begin by fitting a sample of signal only Monte-Carlo for a mode. The sample contains 15000 $B^\pm \rightarrow D^0\pi^\pm$ events. We obtain an answer of $494.7 \pm 7.2\mu\text{m}$, the Monte-Carlo was generated with a lifetime of $496 \mu\text{m}$ in this mode.

We then mix this sample with 8 different background samples and fit the lifetime using a full likelihood describing background as well as signal. The results are tabulated below. In terms of the σ deviation we quote it is calculated as the deviation from the fit with MC only. Adding the background increases the statistical error on the lifetime. We use the increase in the statistical error as the size of the error against which the σ deviation is calculated. What we want to see in this test is if there is some problem with the method that shifts the lifetime from the original background free lifetime. We see that the shifts are small. A more accurate systematic is calculated using a pull study.

Sample	Lifetime with Background (μm)	Lifetime from Signal Only (same) (μm)	σ deviation
1	499.1 ± 8.1	494.7 ± 7.2	1.18
2	498.1 ± 8.1	494.7 ± 7.2	0.91
3	490.4 ± 8.1	494.7 ± 7.2	-1.15
4	496.4 ± 8.0	494.7 ± 7.2	0.48
5	492.9 ± 8.0	494.7 ± 7.2	-0.51
6	499.1 ± 7.9	494.7 ± 7.2	1.35
7	497.5 ± 8.0	494.7 ± 7.2	0.80
8	501.7 ± 8.3	494.7 ± 7.2	1.62

The Cabbibo suppressed mode $B^\pm \rightarrow D^0K^\pm$ is not taken into account for this study. We point out to the reader that the Monte-Carlo signal only sample best fit lifetime is less than 1σ away from the generated lifetime and that mixing in background and fitting signal and background still retains this good agreement

11.2 Data Monte-Carlo Results: $B^0 \rightarrow D^\mp \pi^\pm, D^0 \rightarrow K^\mp \pi^\pm \pi^\pm$

Once again we begin by fitting a sample of signal only Monte-Carlo for a mode, with a sample of 10000 $B^0 \rightarrow D^\mp \pi^\pm, D^0 \rightarrow K^\mp \pi^\pm \pi^\pm$ events. We obtain an answer of $465 \pm 5 \mu\text{m}$, the Monte-Carlo was generated with a lifetime of $464 \mu\text{m}$ in this mode.

We then mix this sample with 3 different background samples and fit the lifetime using a full likelihood describing background as well as signal. The results are tabulated below.

Sample	Lifetime with Background (μm)	Lifetime from Signal Only (same) (μm)	σ deviation
1	478.8 ± 8.6	465.3 ± 7.5	3.2
2	457.4 ± 7.9	465.3 ± 7.5	-3.18
3	457.1 ± 7.8	465.3 ± 7.5	-3.8

The Cabbibo suppressed mode $B^0 \rightarrow D^\pm K^\mp$ is not taken into account for this study. In real data this B^0 mode has background under the peak from $\Lambda_B \rightarrow \Lambda_C X$ and $B_s^0 \rightarrow D_s^\pm \pi^\mp$, the removal of these modes via the appropriate vetoes is described in section 10.

11.3 Pull Study

While Table 11.1 and 11.2 show that the fit work in a realistic situation with signal and background present there is of course some deviation from the truth. To investigate this further and to find out if there is a systematic shift induced by this method we carry out a pull study.

We would like to make this study as realistic as possible, especially in trying to match the distributions of acceptance functions of signal and background as they would be in real data. We would also like to run our pull study over samples where the sample size is similar to what we see in Data, roughly 20,000 signal events to 30,000 background events in the mass range 5.23 - 5.5 GeV.

To get a good estimate of any systematic shift in best fit lifetime we would like to carry out approximately 400 different tests. This however would require 8M Monte Carlo events that pass our selection cuts, which we do not have. To overcome this we employ a method similar to the bootstrap method.

We have at our disposal 70K MC signal events that pass our analysis cuts and 125K events from the upper sideband of our B_0 reconstruction. We take 20K Monte Carlo events at random from the large MC sample. The same event can be chosen more than once. We would also like to make the mass distribution of these signal events look more like what we see in data and so reassign masses to the chosen MC events according to the double Gaussian model. The parameters for this Gaussian model are taken from the B_0 data mass fit.

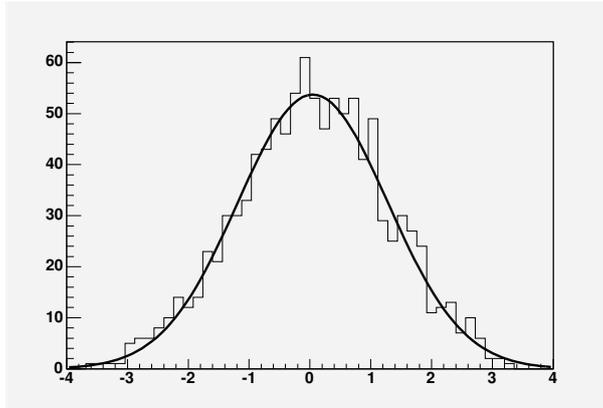


Figure 17: A pull distribution of the best fit lifetime mean= 0.048 ± 0.039 width= 1.20 ± 0.03

We add to this chosen MC sample 30K background events. These are taken at random from the sample of upper sideband and once again the same event can be chosen more than once. The masses of the background events also need to be reassigned so that they lie along the full mass range of the fit. The masses are reassigned according to a linear polynomial of negative slope and once again the slope we use to generate these new masses is similar to what we see in data.

We mix the chosen MC and background events and run them through the fitting procedure. The fitter does not know which events are signal and background. We look at the pull of the lifetime from the best fit value ($464\mu\text{m}$), and fit for a Gaussian.

We have just over 1000 sample fits. The plot for pulls is shown in 17. The fit for the mean and width of this Gaussian are 0.048 ± 0.04 and 1.20 ± 0.03 respectively. While the width of the fitted gaussian is greater than one, we must remember that the statistical errors used here have been underestimated as we have not taken into account the statistical error from the mass fit. In the B0 we find that this causes an increase in statistical error of 12 percent, and so we are not concerned that the fit here shows a slightly wide gaussian. The mean of the distribution is 0.048 ± 0.04 which is consistent with 0. For the purposes of assigning a systematic in lifetime we need to multiply the mean of the distribution by the error. The mean error is just under 6 microns and so for as systematic we estimate $0.3 \pm 0.2\mu\text{m}$. To be conservative we apply a systematic of $0.3\mu\text{m}$.

12 Fits to Data

We present in this section our results for the charged and neutral lifetimes. Statistical errors only are quoted. The following section discusses our systematic errors, and following that we show the combined statistical and systematic error

along with the charge to neutral lifetime ratio.

12.1 $B_u^- \rightarrow D^0 \pi^-$

The analysis cuts and the data used for this measurement are described in section 10. We perform the lifetime fit and obtain the B^+ lifetime as $488.5 \pm 6.2 \mu m$. The statistical error quoted here includes the propagated error from the mass parameters as described in 13. The statistical error propagated from the mass fit is $1.6 \mu m$ which is added in quadrature to $6.0 \mu m$ from the time fit to give the total statistical error as 6.2. We find a signal yield of 25180 ± 340 and show the mass and lifetime projections in figure 18. We also show the distributions of fisher scalar and the fitted function for signal fraction as a function of fisher scalar. While there are some fluctuations in the extremes of the distribution we must note that there are also few events in this region. These plots are shown in 19.

The best fit parameters for the B^\pm decay are presented here. We have assumed that the 18 order Lagrange interpolating polynomial fit to the Fisher scalar is of less interest to the reader than the numbers describing the mass and lifetime signal and background parameters, the Fisher scalar fit is shown in the appendix 17.1 for the interested reader. Figure 19 shows the fit parameters displayed pictorially. In table 12.1 we show the results of the fit of the mass distribution displayed in figure 18. The parameters $M_1, M_2, \sigma_{M1}, \sigma_{M2}, f_1$, relate to the signal the terms are the masses, widths of the two gaussians and the fraction of events in the first gaussian respectively. The parameters f_s and m_{bck} are the slope of the mass background and the fraction of events that are signal respectively.

Parameter	Best Fit value \pm Error
M_1	5.271 ± 0.003
M_2	5.277 ± 0.001
σ_{M1}	0.0246 ± 0.0041
σ_{M2}	0.0140 ± 0.0019
f_1	0.528 ± 0.223
f_s	0.509 ± 0.007
m_{bck}	-0.1686 ± 0.0016

Next we tabulate the lifetime results, we tabulate a best fit B^\pm lifetime λ_{B^\pm} , the first and second background lifetimes λ_1 and λ_2 and the weight of the first lifetime f_1 , and the weight of the prompt component f_{prompt} . We also include here the fit values for the single track SVT efficiency for signal and background, E_{signal} and E_{Bkg} respectively.

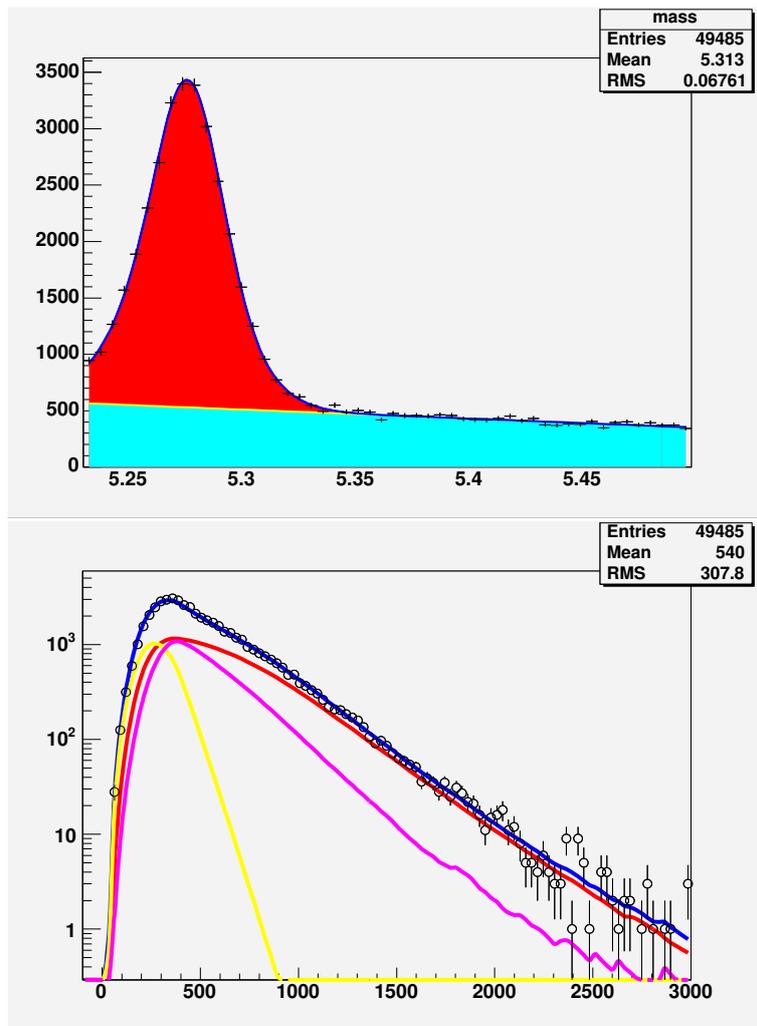


Figure 18: The mass and lifetime projections for Bu, in 1fb^{-1} of data. The lifetime is $488.5 \pm 6.2\mu\text{m}$. In the lifetime projection the blue line shows the total $c\tau$ projection, while the red is the signal component and the other colours show the background contributions.

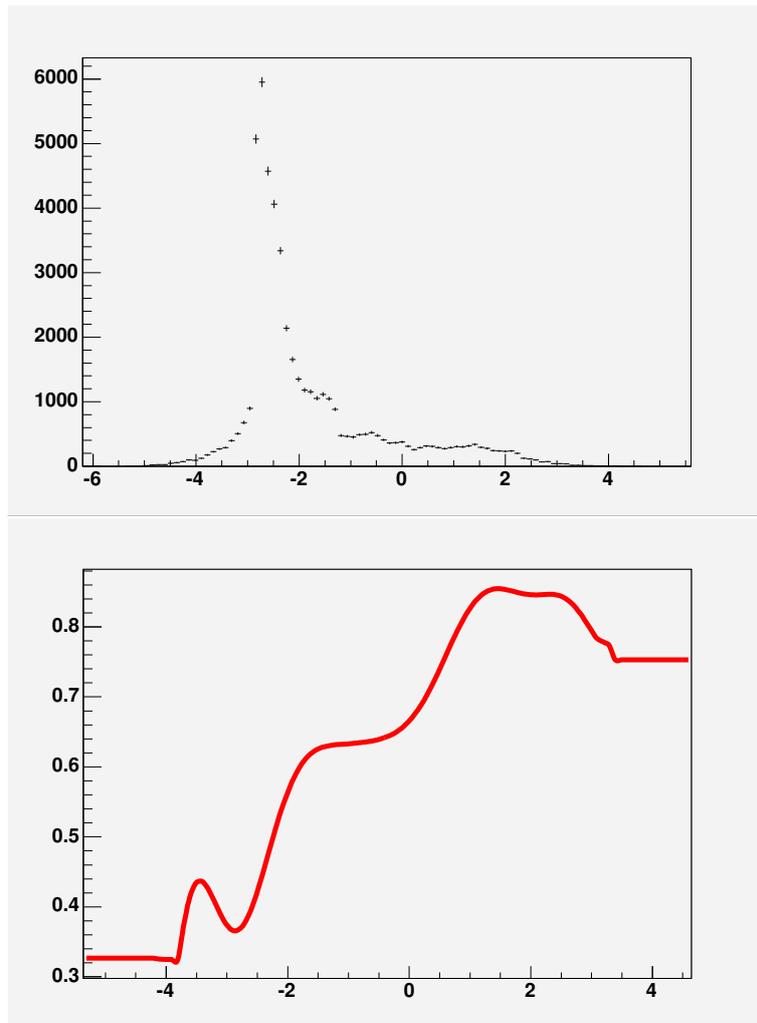


Figure 19: The overall distribution of fisher scalar is shown at the top. The lower plot shows the Lagrange Interpolating Polynomial which is the best fit to signal fraction as a function of fisher scalar.

Parameter	Best Fit value \pm Error
λ_{B^\pm}	488.5 ± 6.0
λ_1	63.2 ± 2.4
λ_2	219.6 ± 4.7
f_1	8.24 ± 0.51
f_{prompt}	$2 \times 10^{-7} \pm 0.02$
E_{Signal}	0.695 ± 0.005
E_{bkg}	0.594 ± 0.010

No signal fraction is presented in this table, this is because the signal fraction is fit in each of 18 bins of the Fisher scalar and is presented in the appendix.

12.2 $B^0 \rightarrow D^- \pi^+$

The analysis cuts and the data used for this measurement are described in section 10. We perform the lifetime fit and obtain the B^+ lifetime as $454.3 \pm 6.4 \mu m$. The statistical error quoted here includes the propagated error from the mass parameters as described in 13. The statistical error propagated from the mass fit is $2.8 \mu m$ which is added in quadrature to $5.7 \mu m$ from the time fit to give the total statistical error as 6.4. We find a signal yield of 20400 ± 430 and show the mass and lifetime projections in figure 20. We also show the distributions of fisher scalar and the fitted function for signal fraction as a function of fisher scalar. While there are some fluctuations in the extremes of the distribution we must note that there are also few events in this region. These plots are shown in 21.

The best fit parameters for the B^0 decay are presented here in a similar fashion to the B^\pm . Once again we have assumed that the 18 order Lagrange interpolating polynomial fit to the Fisher scalar is of less interest to the reader than the numbers describing the mass and lifetime signal and background parameters, the Fisher scalar fit is shown in the appendix 17.1 for the interested reader. In table 12.2 we show the results of the fit of the mass distribution displayed in figure 20. The parameters M_1 , M_2 , σ_{M1} , σ_{M2} , f_1 , relate to the signal the terms are the masses, widths of the two gaussians and the fraction of events in the first gaussian respectively. The parameters f_s and m_{bck} are the slope of the mass background and the fraction of events that are signal respectively. We also include here the fit values for the single track SVT efficiency for signal and background, E_{signal} and E_{Bkg} respectively.

Parameter	Best Fit value \pm Error
M_1	5.269 ± 0.003
M_2	5.276 ± 0.003
σ_{M1}	0.0323 ± 0.0049
σ_{M2}	0.0143 ± 0.001
f_1	0.471 ± 0.096
f_s	0.4113 ± 0.0083
m_{bck}	-0.1662 ± 0.0020

Next we tabulate the lifetime results, we tabulate a best fit B^0 lifetime λ_{B^0} , the

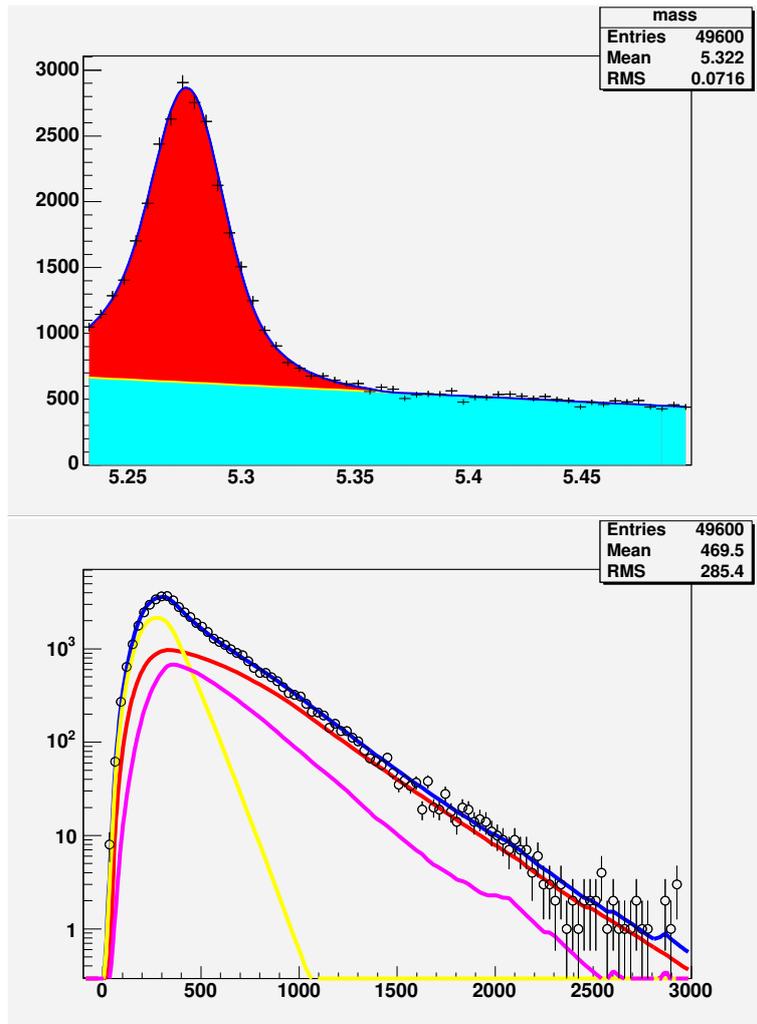


Figure 20: The mass and lifetime projections for B^0 , in $1fb^{-1}$ of data. The lifetime is $454.3 \pm 6.4\mu m$. In the lifetime projection the blue line shows the total $c\tau$ projection, while the red is the signal component and the other colours show the background contributions.

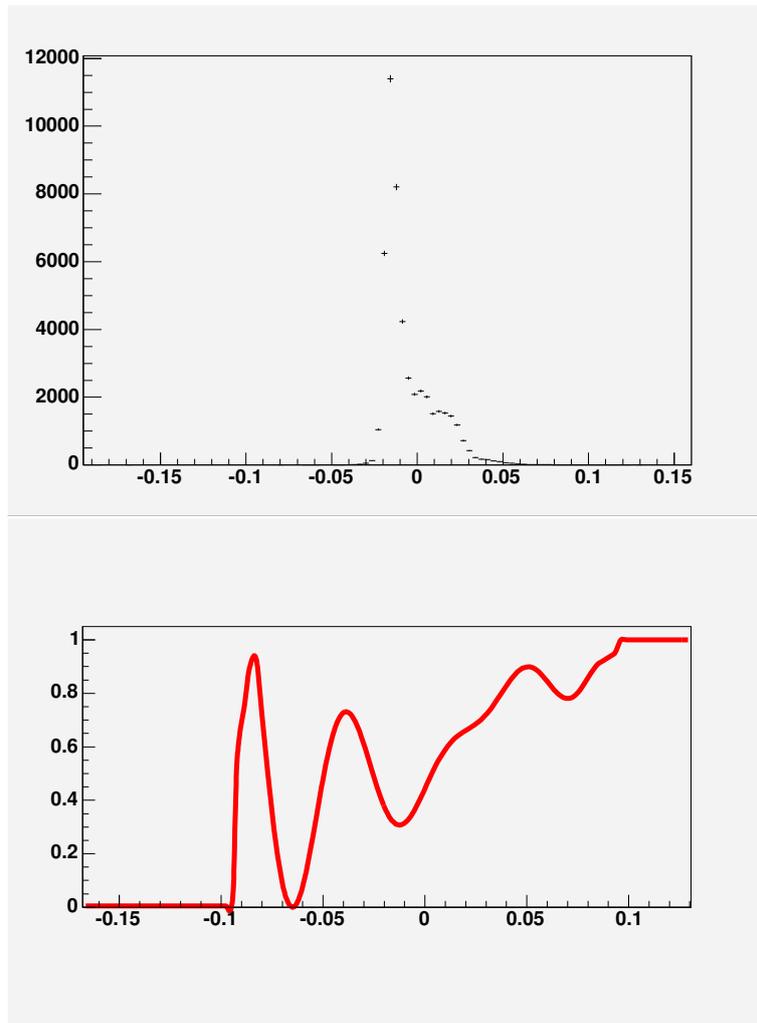


Figure 21: The overall distribution of fisher scalar is shown at the top. The lower plot shows the Lagrange Interpolating Polynomial which is the best fit to signal fraction as a function of fisher scalar.

the first and second background lifetimes λ_1 and λ_2 and the weight of the first lifetime f_1 , and the weight of the prompt component f_{prompt} .

Parameter	Best Fit value \pm Error
λ_{B^0}	454.3 ± 5.7
λ_1	73.9 ± 1.5
λ_2	224 ± 6
f_1	16.3 ± 1.1
f_{prompt}	$2x10^{-5} \pm 0.005$
E_{Signal}	0.734 ± 0.004
E_{bkg}	0.668 ± 0.005

No signal fraction is presented in this table, this is because the signal fraction is fit in each of 18 bins of the Fisher scalar and this fit presented in the appendix 17.1

13 Comparison of Fit projection and Data

In this section we present a comparison of the fit projection to the data. We start by looking at the fit projection of Monte Carlo signal only in high statistics. This is shown in 22 along with the residuals between the fit projection and ctau distribution in the Monte Carlo. Overall the residuals are small and show that the acceptance function technique is correcting for the bias imposed by the trigger. Figure 23 shows the fit projection and residuals for a typical Monte Carlo signal plus upper sideband mix. It shows that there are some large residuals which are coming from the background model. In this analysis we calculate the acceptance function for an event which is the points along its path of travel at which the event could have been accepted by the trigger. For real B mesons this has a real and tangible meaning, however for an event that was in fact combinatoric background the meaning of the acceptance function is less clear. We have also found that if we wish to use event by event acceptance functions for the signal PDF we must also use them for the background. However this makes background $c\tau$ distribution difficult to model perfectly as each event comes with its own acceptance function. We model the background as a sum of exponentials with the τ and weight of these components allowed to float in the fit. Figure 24 shows the residuals for the fit to Data for the B0. We see that the residuals follow the same pattern as that for the realistic Monte Carlo plus sideband mix. By adding extra background components we do see a small improvement in the residuals but no change in the lifetime result. From the realistic pull study we see no bias coming inherently from the method. For all the fits in the pull study we can compute a chi-squared type quantity which is the sum of the squares of the residuals divided by the number of bins. A histogram of these chi-square type quantities is shown in figure 25. The same number is computed for the fit to Data and it is 2.01 which lies just to the left of the center of the distribution of chi squares from the pull study. There does not appear to be anything in the Data that is not also in the Pull study. Hence although the background model isn't perfect we conclude that it is good enough to allow the extraction of the signal lifetime which is the quantity of interest.

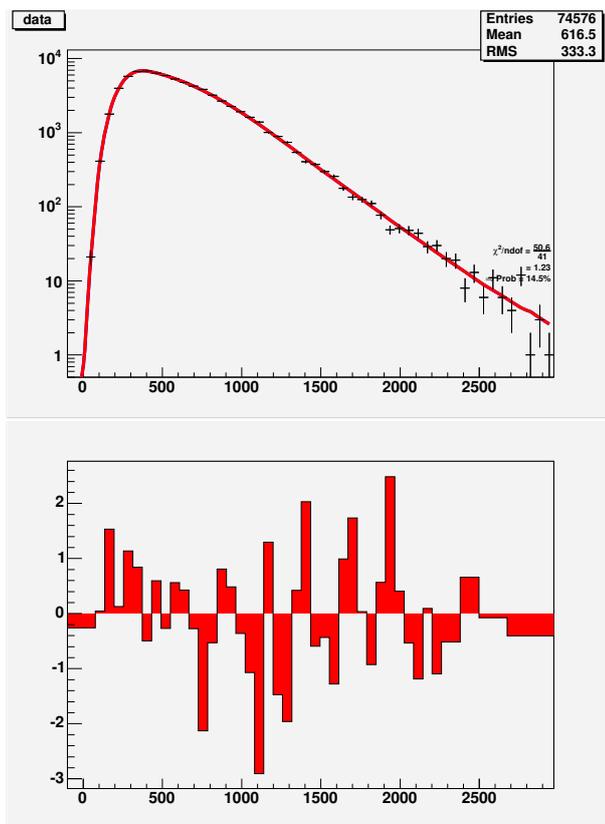


Figure 22: 70K events, MC signal only fit

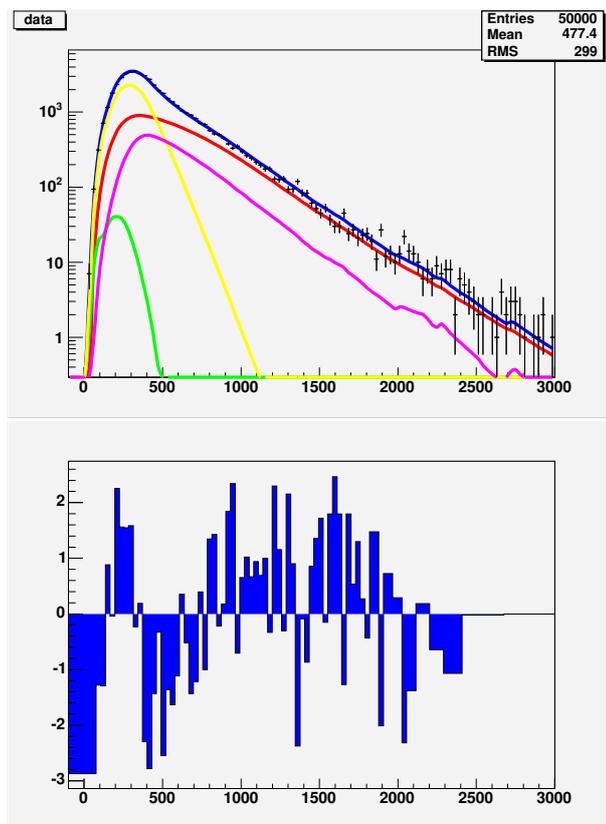


Figure 23: A typical Lifetime projection from a MC + upper sideband mix

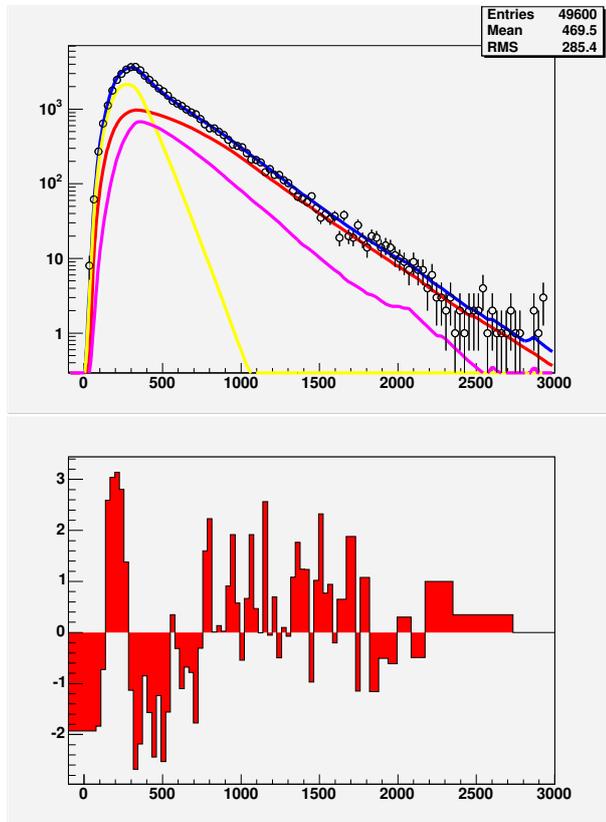


Figure 24: The lifetime projection for B0 Data fit

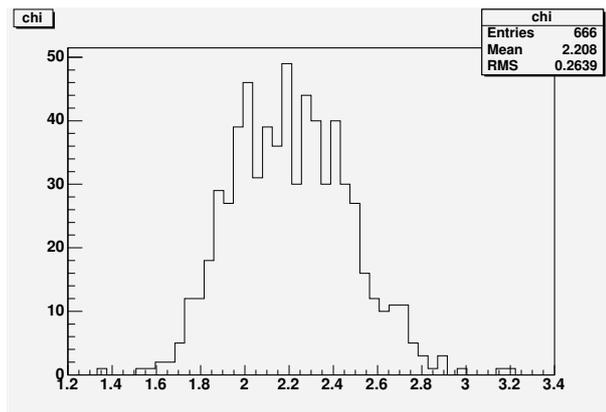


Figure 25: The distribution of the chi-squared type quantity from the fits performed in the pull study. For our fit to Data the corresponding quantity of 2.01 is consistent with the above distribution

14 Systematic Errors

14.1 Systematic Due to Silicon Misalignment

The alignment group quotes the systematic error on the alignment of the silicon system as being $50\mu\text{m}$. The usual way in which the effect of a misaligned detector has upon a lifetime result is to simulate the misaligned detector. One simulates with one alignment table and estimates track parameters using another, observing the effect upon the fitted lifetime.

What others have done Since the selection of events in dimuon channels is largely independent of alignment (apart from very loose cuts like χ^2), any differences due to selection are assumed to be statistical fluctuations and these are zeroed by fixing the selection and varying only the alignment constants. Event selection is performed using one alignment table; then, the tracking fits are redone using a different table. The wafer misalignments produce hit slewing which propagates to tracks, then to vertex positions, to proper decay times and ultimately to the fitted B lifetime. Since track refits can be performed in the analysis step, the entire systematic study can be conducted without even a rerun of production.

Why we cannot do that. The effect that a misaligned detector has upon the lifetime measurement extracted in this analysis, and the treatment we use to estimate it, differs from previous analyses in several ways.

- The *selection* of events by the hadronic B trigger is affected by the alignment. This is not the case in analyses in which lifetime distributions are unbiased by the trigger.
- Estimating the effect requires that the misalignments be introduced into the offline tracking but also into the SVT.
- Since now the selection of events changes, the samples used to estimate the lifetime before and after the shift vary slightly. If the samples are not 100% correlated, then any shift in the central value induced by the offline tracking must be considered to have a statistical error coming from the sample difference.

A simple thought experiment serves to illuminate the last issue. If the wafer positions are changed, some events will enter or leave the sample. In case the events that are gained/lost all come from the front edge of the acceptance, we would say that the alignment was affecting the lifetime. In case the events are gained/lost at random, without regard to their lifetime we would see a different measured lifetime for sure but we would have to say that the change in the measured value was due to statistical fluctuation.

One could approach this problem by determining the degree to which the two measurements were statistically correlated. Our approach is to determine, through a procedure we describe below, the change in fitted lifetime value due to alignment together with its statistical error. This will be quoted as $\delta \pm \sigma$.

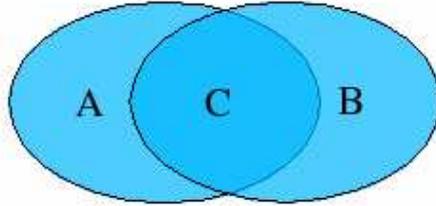


Figure 26: Sample 1 overlaps with Sample 2. We consider Sample A (unique to selection 1), Sample B (unique to selection 2) and Sample C (common to both selections).

Needless to say, the procedures by which a misaligned detector is introduced into the SVT is a non trivial task both for us and for the CDF analysis farm.

14.2 The Statistical Error in the Alignment Shift

Two measurements of a quantity performed with overlapping samples are expected to show differences due to statistical fluctuations. Here, we estimate the size of those differences (“ σ ”, in the discussion above). Let’s call the first set of events “Sample 1” and the second set of events “Sample 2”. Furthermore, we call their intersection “Sample C” (for common); those are the events common to both selections.

Furthermore, we call the events unique to the first selection “Sample A” and those unique to the second selection “Sample B”. The event selections can be visualized with Venn diagrams, as if Fig. 26.

Suppose that unbinned maximum likelihood fits to Samples 1 and 2 return measured values $x_1 \pm \sigma_1$ and $x_2 \pm \sigma_2$. These measurements can be each be considered as weighted average results. $x_1 \pm \sigma_1$ could be obtained as a weighted average of an estimate of x within sample A and an estimate of x within sample C. We’ll denote these estimates as $x_A \pm \sigma_A$ and $x_B \pm \sigma_B$. The weight w of a the estimate is defined as

$$w = \frac{1}{\sigma^2}$$

Then,

$$x_1 = \frac{w_A x_A + w_C x_C}{w_A + w_C} \tag{52}$$

$$w_1 = w_A + w_C \tag{53}$$

And

$$x_2 = \frac{w_B x_B + w_C x_C}{w_B + w_C} \quad (54)$$

$$w_2 = w_B + w_C \quad (55)$$

This composition is useful because, unlike samples 1 and 2, samples A, B, and C are disjoint so they are statistically independent. Notice that we do not claim that the estimate of x within samples A, B, or C is physically meaningful. We merely are stating that they relate to the measurements within samples 1 and 2 as stated above. (E.G: the average height of students in a class is certainly the average of the average of the short students and the average of the tall students). For this reason we shall refrain from calling estimates of x in samples A, B, or C as *measurements*; we refer to *estimates* of x in samples A, B, and C; and *measurements* of x in samples 1 and 2.

We are interested in the difference $\Delta = x_1 - x_2$ between measurement 1 and measurement 2. We can write it in terms of the estimates within samples A, B, and C in the following way:

$$\Delta = x_1 - x_2 = \quad (56)$$

$$\frac{w_A x_A + w_C x_C}{w_A + w_C} - \frac{w_B x_B + w_C x_C}{w_B + w_C} \quad (57)$$

The advantage in doing so is that we know the degree of statistical correlation between x_A , x_B , and x_C is zero, whereas we do not know the degree of correlation between x_1 and x_2 at all.

Our goal is to determine the σ_Δ , the error on the measurement difference. It can be got from a straight propagation of errors using the above expression. We compute

$$\sigma_\Delta^2 = \sum_{i=A,B,C} \left(\frac{\partial \Delta}{\partial x_i} \sigma_i^2 \right) \quad (58)$$

Carrying out the algebra, we obtain:

$$\sigma_\Delta^2 = \frac{w_A}{(w_A + w_C)^2} + \frac{w_B}{(w_B + w_C)^2} + \frac{w_C (w_B - w_A)^2}{(w_A + w_C)^2 (w_B + w_C)^2} \quad (59)$$

Two limiting cases are of interest. When the events do not overlap at all, $w_C = 0$, and one sees easily that $\sigma_\Delta^2 = \sigma_A^2 + \sigma_B^2$. When all of the events are common $w_A = w_B = 0$ and $\sigma_\Delta = 0$. These are precisely what one expects. We shall use this formula below to obtain the statistical error on the alignment shift.

Sample	Size (events)	τ (default) (μm)	τ (out) (μm)	σ_τ (μm)
A (default only)	26.4K	443		6
B (Move out)	28.2K		438	6
C (Common)	38.0K	533	537	7

Table 2: Raw lifetime estimates within samples (A,B,C). Sample A consists of events selected with default alignment, only; sample B of events selected with wafers moved out; and sample C of common events.

14.3 Estimate of the Alignment Systematic.

The lifetime shift has been evaluated by shifting the wafers outwards by 50 μm . We simulated $B^+ \rightarrow D^0\pi^+$, simulating the detector and the SVT trigger, reconstructing the events and applying the lifetime estimators to selected events. The input lifetime was 496 μm . The procedure to simulate the misaligned SVT is detailed in section 16.

The events which are common to each selection are about 60% of the selection. The raw numbers are shown in Table 14.3. Samples A and B contain events lying at the edge of the acceptance, so the low value of the lifetime estimate for these subsamples is expected. Sample C is depleted in such events so the high lifetime seen in that subsample is also expected. Sample 1, which is AUC, gives a measurement of 495 ± 5 while sample 2, BUC, gives 493 ± 5 , so we observe a downward shift of two microns.

Using the formulae derived in this note, we determine the shift to be -2 ± 5 μm . The direction of the observed shift is important because the lifetime goes down when the wafers are moved out. While more aggressive estimates of the alignment systematic error might be justified in this situation, we choose the conservative approach and quote the systematic error due to alignment as ± 5 μm .

14.4 Another approach to evaluating a systematic error due to Silicon Misalignment

We have already described how full detector and trigger simulation MC can be used for evaluating a possible systematic error in lifetime due to a misalignment of the detector ???. This technique is computationally intensive and we have describe here a method to evaluate the same error using a toy Monte-Carlo technique.

A toy Monte Carlo of approximately 1 million simulated $B_s^0 \rightarrow D_s^\pm\pi^\mp$ are used with lifetimes of 465 (B_s) microns and 147 microns (D_s) and a P_T distribution based on realistic BGEN CDF Monte-Carlo with full detector and trigger simulation is used (this is the only input from full detector simulation). The D_s^\pm decays to $\pi^\pm\phi$ and $\phi \rightarrow K^+K^-$. The choice of decay is arbitrary and has a similar topology to $B^0 \rightarrow D^\pm\pi^\mp$.

By using geometry we calculate the effect of radially moving all Silicon wafers.

The new impact parameter is give by:

$$d_{0shifted} = d_{0true} + R \cdot \sin(\phi_w)$$

in this expression $d_{0shifted}$ is the impact parameter recalculated due to the radial movement of the wafers, d_{0true} is the true, generated impact parameter of the track. The angle ϕ_w is the track ϕ_0 but measured from the bisector of a wedge to the origin of the co-ordinate system (also assumed to be the beam spot), thus ϕ_w lies between $\pm \frac{\alpha_w}{2}$, here α_w is defined as the angle subtended by a silicon wedge at the center of the co-ordinate system. Finally R is the radial shift, in or out. The secondary vertex positions are all re-calculated analytically taking the shifted impact parameters into account.

No effect of the alignment on the individual track ϕ_0 is assumed. The decays are required to pass the L2 B.CHARM (medium P_T) trigger path cuts.

We move the detector out by 50 microns and fit a lifetime of 468+- 0.75 microns, the shift is 3 ± 0.75 microns and we assign a sytematic uncertainty of 3 μm due to a possible detector residual misalignment, this estimate is conservative.

14.5 Systematic error due to the single-track efficiency of the SVT

The MC-free method assumes that the single-track finding efficiency of the SVT is flat between for $0 < d_0 < 1000 \mu\text{m}$ where d_0 is the impact parameter measured by the SVT. However we observe that there is some deviation from the flat hypothesis. To estimate the effect of the deviation on a lifetime measurement we reject events according to the deviation and estimate a systematic error.

A signal only sample from $B_u^\pm \rightarrow D^0 \pi^\pm$ Monte Carlo is used and a description of the evaluation of the systematic error now follows.

14.5.1 Determining the single track finding efficiency of the SVT

The single track efficiency of the SVT is found by dividing the number of tracks found by the SVT by the number of tracks found by the offline as a function of the offline impact parameter d_0 .

Offline tracks from 5.1.3 Monte Carlo of the decay $B_u^\pm \rightarrow D^0 \pi^\pm$ are selected according to to the following criteria:

- i . The number of Silicon hits in R-phi (Ax-hits) should be ≥ 3
- ii . The track transverse momentum $P_T \geq 2 \text{ GeV}$
- iii . and track $|\eta| \leq 1.1$

The final expression for SVT single track efficiency is given by the expression:

$$\epsilon_{SVT}(d_0) = \frac{N^{SVT}(Ax - hits \geq 3 : P_T \geq 2GeV, |\eta| < 1.1)}{N^{OFF}(Ax - hits \geq 3 : P_T \geq 2GeV, |\eta| < 1.1)} \quad (60)$$

We fit this expression to the function :

$$\epsilon_{SVT} = p_0 \times \text{erfc}\left(\frac{d_0 - p_1}{p_2}\right) \quad (61)$$

where erfc is the complementary error function, d_0 is the offline impact parameter and p_0 , p_1 and p_2 are free parameters. The best fit shown in 27 gives values for the free parameters of $p_0 = 0.261134 \pm 0.0516650$, $p_1 = 0.156018e \pm 0.0180816$, and $p_2 = 0.0428187 \pm 0.028783$.

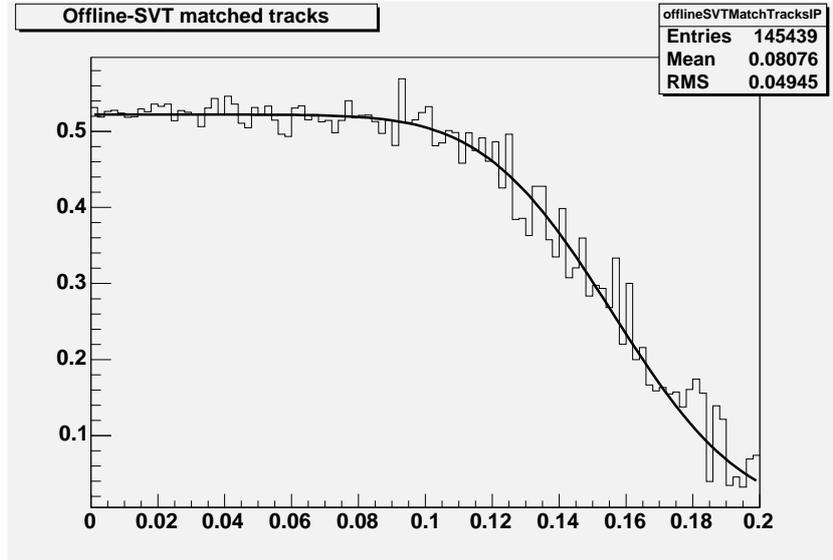


Figure 27: Single track efficiency of the SVT

14.5.2 Determination of the systematic error

Using the parameterization described in the previous subsection we proceed to calculate the systematic error in the lifetime measurement. The first step is to measure the lifetime of the signal sample assuming-as the MC-free method does-that the single track finding efficiency of the SVT is flat. Using 13670 events we obtain a best fit lifetime of $484 \pm 7.58 \mu\text{m}$. Let us call this sample of events sample A, with best fit lifetime and error: $\bar{x}_A \sigma_A$ respectively .

Next we do the fit whilst randomly rejection events using the efficiency function: using the impact parameter d_0 of each track we evaluate the efficiency function

$\epsilon_{SVT}(d_0)$ and generate a uniform deviate between 0 and 1. If this deviate lies between the difference between the y-intercept of this curve and its value at d_0 we reject it. This gives us a slightly smaller sample of 13631 events, yeielding a best fit lifetime of $482 \pm 7.50 \mu\text{m}$. Let us call this sample B with best fit lifetime and error: $\bar{x}_B \sigma_B$ respectively.

Let us now designate the rejected sample of 39 events by R and the best fit lifetime and error in this sample by $\bar{x}_R \sigma_R$ respectively. Note that we have not fit a lifetime in such a small sample but the utility of this designation will become apparent shortly.

The shift in best fit lifetimes of samples A and B is not the systematic error since the two samples A and B are highly correlated. We need to know the uncertainty of this shift before we are able to estimate what the systematic error is. In order to do this an approach similar to that used for calculating the systematic error due to misalignment of the detector which contains highly correlated samples of events as well. We begin by writing down the expression for the uncertainty on the shift Δ (uncertainty is denoted by σ_Δ):

$$\sigma_\Delta^2 = \left(\frac{\partial\Delta}{\partial\bar{x}_A}\right)^2 \cdot (\sigma_A)^2 + \left(\frac{\partial\Delta}{\partial\bar{x}_B}\right)^2 \cdot (\sigma_B)^2 + \left(\frac{\partial\Delta}{\partial\bar{x}_A}\right)\left(\frac{\partial\Delta}{\partial\bar{x}_B}\right) \cdot Cov(\bar{x}_A, \bar{x}_B) \quad (62)$$

we cannot determine the correlation $Cov(\bar{x}_A, \bar{x}_B)$ so we write down the expression for the shift $\bar{x}_A - \bar{x}_B$ in terms of the two uncorrelated samples R and B denoting their weights by $w_R = \frac{1}{\sigma_R^2}$ and $w_B = \frac{1}{\sigma_B^2}$ respectively:

$$\Delta = \bar{x}_A - \bar{x}_B = \frac{w_B \cdot \bar{x}_B + w_R \cdot \bar{x}_R}{w_B + w_R} - \bar{x}_B \quad (63)$$

from this expression we can derive the uncertainty:

$$\sigma_\Delta = \frac{w_R}{w_B + w_R} \times \sqrt{w_R^{-1} + w_B^{-1}} \quad (64)$$

We can estimate w_R by using the relative size of sample R compared to sample B, using this and the above expression we obtain $\sigma_\Delta = 0.4\mu\text{m}$ as the uncertainty of the shift Δ ($2\mu\text{m}$). To be conservative we assign a systematic error of 2 microns due to our parameterization of the SVT single-track finding efficiency.

14.6 Systematic Error due to Resolution Function

The Resolution function is the model we use to describe the distribution of the $\sigma_{c\tau}$ errors. Our model uses a Gaussian and this is convoluted with the exponential for lifetime. We have to consider that this description is not perfect. From the exclusive B lifetimes analysis using Jpsi modes (CDF Note 8524) it is found that the errors are not distributed according to a Gaussian but there

exists a longer tail. On average they find that the tail is modelled by a Gaussian of width 3 times the main Gaussian and comprises a fraction of 10%.

To test what effect this may have in our model we use our signal toy generator, and generate gaussian distributed errors but 10% of the time we increase them by a factor of 3. The events from the toy simulation are then fit using only our single gaussian resolution model. For 1M events we see a lifetime shift from the input truth as $+1.6 \pm 0.7$. We therefore assign a systematic of 1.6 microns as the systematic due to modelling the resolution errors by a single gaussian.

14.7 Systematic Error due to binning of Acceptance Functions

As we have described earlier in section 8 in order to differentiate acceptance functions of signal from background, we use the Fisher discriminant technique, in order to obtain the matrix of weights we have to transform each calculated acceptance function into a vector by binning it. The heights of various bins contain the information used to make the discrimination possible.

The acceptance functions have different shapes depending on the number of tracks with which it is possible to form a trigger. The more tracks in a final state the more possible combinations that can pass the trigger, and hence the more structure the acceptance function has, this is described in section 8.3.1. The choice of the number of bins given a particular decay is discussed in section 8.3.1. We have found that we can vary the number of bins and still discriminate between signal and background using this technique. To investigate whether there is any systematic effect due to a choice of the number of bins we fit the data and vary the number of bins in both the B^0 and B^\pm samples. We record the resulting shifts in the best fit lifetime and estimate a systematic uncertainty.

A table for both decays is given below. For the B^\pm we have quoted the final results from 20 bins and 10 bins for the B^0 . All shifts recorded are relative to these measurements.

Decay Mode	Acceptance Function Bins	Best Fit Lifetime (μm)	Shift (μm)	$\delta_{systematic}$ (μm)
$B^\pm \rightarrow D^0 \pi^\mp$	20	488.5 \pm 6.1 (quoted)	0	± 1.3
	15	487.2 \pm 6.1	+0.7	
	17	489.2 \pm 6.1	-1.3	
$B^0 \rightarrow D^\pm \pi^\mp$	10	454.3 \pm 5.7 (quoted)	0	± 1.7
	9	456.0 \pm 5.7	+1.7	
	11	454.3 \pm 5.7	0	

In this table the word ‘‘quoted’’ indicates that this is the final result and all shifts are calculated from this number.

To be conservative we assign a systematic error of $\pm 1.3 \mu\text{m}$ for the B^\pm and $\pm 1.7 \mu\text{m}$ for the B^0 , these correspond to the largest shifts that we see in varying the binning of our acceptance functions.

14.8 Systematic error due to using acceptance functions and using fisher discriminants to find the signal probability for each event

This systematic error can be thought of as the shift introduced by the method of unbiasing the data itself. We take the systematic error found by doing the pull study. As described in section 11.3 we assign a systematic of 0.9 microns.

14.9 Systematic error due to fixing the scale factor of calculated proper lifetime error

The Monte-Carlo free method of measuring hadronic lifetimes uses event by event calculated errors $\sigma_{c\tau}$ on the proper decay length $c\tau$ as do all other lifetime analyses at CDF. These are returned from the C++ wrapped vertex fitting program CTVMFT. From Monte-Carlo pull distributions we observe a scale factor of approximately 1.1 ± 0.03 , which means that the errors returned from CTVMFT have to be scaled by an appropriate amount.

The L_{xy} cuts made in this analysis (section 10) leave very little of a prompt component in background and this is not enough to allow a parameter describing the scaling of $\sigma_{c\tau}$ as is customary in lifetime analysis that have sufficient prompt background. For this reason all event by event errors are scaled by a factor of 1.1.

In order to estimate a possible systematic error due to this fixing we change the scale factor to 1.2 and 1.3 and remeasure the B^\pm and B^0 lifetimes, the shifts are then used to estimate a systematic error due to the fixing of the scale factor.

We tabulate our results below and ask the reader to note that our assignment of systematic error is conservative.

B Meson	$\sigma_{c\tau}$ scale	Fit Lifetime	Shift (μm)	$\delta_{systematic}$ (μm)
B^\pm	1.1	488.5 ± 6.0	0 (quoted)	± 0.5
	1.2	488.2 ± 6.1	-0.3	
	1.3	488.0 ± 6.0	-0.5	
B^0	1.1	454.3 ± 5.8	0 (quoted)	± 0.9
	1.2	453.9 ± 5.8	-0.4	
	1.3	453.6 ± 5.8	-0.9	

In this table the word “quoted” indicates that this is the final result and all shifts are calculated from this number.

We assign a systematic error of $\pm 0.3 \mu\text{m}$ to the B^\pm lifetime and $\pm 0.4 \mu\text{m}$ to the B^0 lifetime due to possible variation in scaling of the event by event proper decay length errors. We choose the shifts seen by a scale factor of 1.2 since even this is far from what we expect from Monte Carlo and using the shifts from a scale factor of 1.3 would be an overestimate.

14.10 Systematic Error due to Background Parameterization

We model our background in lifetime as the sum of a prompt component—described by a Gaussian with a width equal to the event by event error and two exponentially decaying long lived components convoluted with an event by event error. The prompt component fraction is allowed to vary as part of the fit and comes out to be consistent with zero.

To evaluate a systematic uncertainty associated with a possible misparameterization of the background we refit our data with three long lived components instead of two for both decays. The results are tabulated below. The only shift we see is for the B^\pm .

Decay Mode	Background lifetimes	Best Fit Lifetime (μm)	Shift (μm)	$\delta_{\text{systematic}}$ (μm)
$B^\pm \rightarrow D^0 \pi^\mp$	2	488.5 ± 6.1 (quoted)	0	
	3	487.4 ± 6.1	-1.1	± 1.1
$B^0 \rightarrow D^\pm \pi^\mp$	2	454.3 ± 5.7 (quoted)	0	
	3	454.3 ± 5.8	0	± 1.1

In this table the word “quoted” indicates that this is the final result and all shifts are calculated from this number.

To be conservative we assign a systematic error of $\pm 1.1 \mu\text{m}$ for both the decay modes despite the fact that the only observed shift is in the B^\pm decay.

14.11 Systematic Error due to fit of Fisher Scalar distribution

A description of the Fisher scalar or Fisher discriminant distribution and its motivation is given elsewhere [8]. This scalar quantity is extracted from the vector of acceptance function bin heights and allows discrimination between signal and background-like acceptance functions. The Fisher scalar distribution is fit using interpolating Lagrange polynomials this is described in section 8.4. The fit of the scalar allows an extraction of $P(s|\text{Fisher scalar})$ the probability that an event is signal given a value of the Fisher scalar.

From Monte-Carlo data-mixes, we know that if we bin the Fisher scalar by 15-20 bins we can get $P(s|\text{FisherScalar})$. See figure 13, the function found matches the truth very well. For our data fits we have used 18 bins. We would like to remind the reader that the order of Lagrange interpolating polynomial needed to fit N data points is exactly N , this being an intrinsic quality of these polynomials.

To see if there is a systematic error associated with our choice of number of bins for the Fisher scalar distribution, we change this to 15 and 20 and see if there is any shift in lifetime and estimate a systematic error accordingly. The

final results quoted in this note have the number of bins set to 18 and the shifts quoted in the table are from these.

Decay Mode	Number of Bins	Best Fit Lifetime (μm)	Shift (μm)	$\delta_{systematic}$ (μm)
$B^\pm \rightarrow D^0\pi^\pm$	18	488.5 \pm 6.1 (quoted)	0	± 0.1
	15	488.5 \pm 6.1	0	
	20	488.4 \pm 6.1	0.1	
$B^0 \rightarrow D^\pm\pi^\mp$	18	454.3 \pm 5.7 (quoted)	0	± 0.1
	15	454.4 \pm 5.7	0.1	
	20	454.3 \pm 5.8	0	

In this table the word “quoted” indicates that this is the final result and all shifts are calculated from this number.

We assign a systematic error of $\pm 0.1 \mu\text{m}$ for both the decay modes.

14.12 Systematic Error due to Inclusion of part of the FSR tail and the Cabibbo suppressed modes

In this subsection we examine whether there is any shift in lifetime caused by the inclusion of the FSR tail and the Cabibbo suppressed modes as signal. We do not separate these modes from the main signal peak. Any systematic we expect to get from treating these the same as the main decay are expected to be very small indeed. The reasoning behind this is that firstly there is a lower mass cut at 5.23 GeV which cuts out most of the contribution of the FSR tail and a large portion of the cabibbo suppressed decay. Secondly these events are still the charged or the neutral meson and so the lifetime of these events will be the same as that of the main peak. The only difference is that the mass has been underestimated in the case of a Cabibbo suppressed decay since a Kaon has been reconstructed as a Pion and in the case of the FSR contribution both the mass and the transverse momentum are slightly underestimated since the photons have not been reconstructed. The differences in any distribution will have to be small otherwise the event will not pass the mass cuts. And finally, since these events lie in the tail of the mass distribution they are not weighted as heavily as signal as events lying in the centre of the main peak. A very small change will occur in the lifetime distribution due to this as the $c\tau$ is given by $\frac{L_{xy}xM_B}{P_T}$ and we use the events’ reconstructed mass.

To investigate any change this might lead to we use our Toy Monte Carlo simulator to generate events with the mass distribution that we see in Data. This means that there are more lighter events than heavier and so we may expect to see some slightly smaller $c\tau$ that may lead to a lower lifetime. We generate 1M toy events and measure the shift from the truth as 0.4 ± 0.7 . We assign a systematic of $0.7\mu\text{m}$.

14.13 Summary of Systematic Errors

The sources of error are: due to possible misalignment of the SVT, choice of scale factor for the proper decay length errors, a possible error due to our choice of the number of acceptance function bins, a possible error due to our choice of the number of bins for the Fisher scalar, a possible error due to the assumption that the SVT single track efficiency is flat, a possible error due to the background lifetime parameterization and finally a possible error due to the fitting method. These have been evaluated separately for the B^\pm and B^0 decay modes.

Although we have presented systematic errors due to a residual misalignment of the detector from two different methods, one involving a full detector simulation 14.1 and from a toy monte-carlo study 14.4, we note that the former study is statistics limited and the latter is not. We therefore cite the systematic error from the latter (14.4).

We present here a summary of all sources of systematic error for the B^\pm decay mode first.

The results for the B^\pm mode are presented below:

Source	Assigned Error μm
Misalignment	3.0
SVT single track Efficiency	2.0
Resolution Function	1.6
Fisher Vector	1.3
Background Parameterisation	1.1
Method Bias	0.3
Suppressed Mode inclusion	0.7
Resolution scale factor	0.3
Fisher discriminant model	0.1

We add these uncertainties for the charged B lifetime in quadrature and obtain $\pm 4.4 \mu\text{m}$.

The systematic errors for the B^0 mode are presented follow:

Source	Assigned Error μm
Misalignment	3.0
SVT single track Efficiency	2.0
Resolution Function	1.6
Fisher Vector	1.7
Background Parameterisation	1.1
Method Bias	0.3
Suppressed Mode inclusion	0.7
Resolution scale factor	0.4
Fisher discriminant model	0.1

again we add these uncertainties for the neutral B lifetime in quadrature and obtain $\pm 4.5 \mu\text{m}$.

15 Lifetime Ratio and Summary

In Summary we have made the following two measurements:

$$c\tau(B^\pm) = 488.5 \pm 6.2(stat) \pm 4.4(syst)\mu m \quad (65)$$

$$c\tau(B^0) = 454.3 \pm 6.4(stat) \pm 4.5(syst)\mu m \quad (66)$$

Theoretical prediction of lifetimes are calculated in terms of ratios so it is interesting to make the same calculation here. In taking the ratio we would like to propagate the error correctly and take into account that a number of systematics are correlated for each mode, such as the alignment. To calculate the error on the ratio we use the simple formula that calculates the error on y , if y is a function of x_a, x_b

$$\sigma_y^2 = \left[\frac{dy}{dx_a} \quad \frac{dy}{dx_b} \right] \times (\text{Error matrix between } x_a, x_b) \times \begin{bmatrix} \frac{dy}{dx_a} \\ \frac{dy}{dx_b} \end{bmatrix} \quad (67)$$

In calculating the error matrix for systematic errors between the two measurements we have assumed that the alignment error, the SVT efficiency and the bias of the method will be the same for both modes.

We find

$$\frac{c\tau(B^\pm)}{c\tau(B^0)} = 1.075 \pm 0.020(stat) \pm 0.008(syst). \quad (68)$$

Acknowledgements

We offer our thanks to Marjorie Shapiro, Amanda Deisher and Hung Chung Fang for access to their stripped data and Monte Carlo. To Guillermo Gomez-Ceballos for helping us recover lost trigger bits and to Johannes Muelmenstadt, Hung Chung Fang and Amanda Deisher again for their advice on mass model parameterization. To Donatella Lucchesi again for data samples. Thanks to Ivan Furic and Manfred Paulini for advice and consultations on pull distributions, to Stefano Giagu for help with the SVT matching code. Also many thanks to our mascot the Moose to continue to inspire us.

References

- [1] N. Urals, [arXiv:hep-ph/9804275].
- [2] M. A. Shifman, [arXiv:hep-ph/0009131].

- [3] K. Anikeev *et al.*, [arXiv:hep-ph/0201071].
- [4] J. Rademacker, Nucl. Instrum. Meth. A **570** (2007) 525
<http://dx.doi.org/10.1016/j.nima.2006.09.090>
 [arXiv:hep-ex/0502042].
- [5] Archer, Branden and Weisstein, Eric W. "Lagrange Interpolating Polynomial." From MathWorld—A Wolfram Web Resource.
<http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html>
- [6] Giovanni Punzi, physics/0401045, published in conference proceedings Physstat 2003. SLAC

16 Appendix A: The Simulation of the Misaligned SVT

We simulate events with wafers in their default position, and then simulate a misalignment by introducing wafer slewing *both* in the track reconstruction and in the SVT. Unfortunately this means rerunning the simulation of the SVT. The simulation, including GEANT hit production, is carried out using wafers in their default position (TABLE 160045 1 GOOD). Then $\pm 50 \mu\text{m}$ wafer shifts are introduced into the SVT and into the Track reconstruction by using the tables TABLE 160047 1 TEST and TABLE 160047 2 TEST). To introduce these constants into the track fits is trivial: the proper alignment table is specified in the .tcl file.

To introduce the constants into the SVT trigger is much more involved. The first step is to distill the SVX geometry into a set of constants summarizing wafer position. This happens within a special procedure (makergeo.csh) developed and maintained by the SVT group. Makergeo is a script which runs an AC++-based program that can be steered through .tcl files and the (mis)alignment tables are introduced at that point. The output file containing the desired information is given the .rgeo extension.

The SVT track fit operates by taking four hit positions plus two XFT parameters (ϕ_0 and curvature c), forming a vector of input parameters and applying a linear transformation to those parameters. The constants used in this transformation are determined using linear regression to simulated tracks. The simulation, which is not to be confused with CDF's full detector simulation, simply generates particles across the detector acceptance in order to determine the linear relationship between track parameters and hit positions. In order to obtain "misaligned" constants, these tracks need to be recreated and the regression repeated. The procedure to do this is called corrgen; it appears to live only outside CDF's version-controlled code management system, but can be obtained through the SVT group. The output of "corrgen" is a file with the ".fcon" extension, containing fit constants.

Finally, the new fit constants are introduced into the simulation via "mapset" files; these contain pointers to the new .fcon files created in the previous step.

	Database	Default	Shift Out	Shift In
Beam-x (μm)	-1973.8	-1973.8	-1973.0	-1974.8
Beam-y (μm)	5152.8	5153.0	5148.8	5157.4
dx/dz (mr)	0.5598	0.5598	0.5605	0.5595
dy/dz (mr)	0.1739	0.1739	0.1738	0.1744

Table 3: Beamspot shifts induced by misalignment of silicon wafers.

This file is edited by hand, and the new file is introduced to the SVT simulation.

The alignment table used to generate the new SVT constants is presented to the reconstruction procedures, specifically the track fits, via the .tcl file.

16.0.1 How the Beamspot Changes when the Silicon Detector is Misaligned.

The misalignments we considered (a 50 micron displacement of all wafers inwards and outwards) produce a collective effect on the beamspot position. The collective motion of such detectors induces an apparent shift of the beamspot. The plot of $d_0 vs \phi$, used to obtain the beam spot position, changes amplitude when the wafers move out or in. We take account of this effect by re-doing the beamspot measurement, introducing the modified beamspot both into the SVT simulation and into the event reconstruction.

The study was performed using stiff muons (50 GeV) in order to obtain high impact-parameter resolution for each event and thereby enhance the statistical power of the events we generated. The statistical power was further enhanced by artificially shrinking the lateral size of the beamspot to one micron. The beamspot was fit to default alignment and to the two misaligned configurations, using an unbinned maximum likelihood fit.

Table 3 shows the fitted beamspot positions. In addition to the beamspot positions and slopes determined from our procedure, we include in the table the numbers coming from the database; these are to be compared with the beamspot position we determine for the default position. The discrepancy is at the submicron level. This gives us confidence that the values we extract for the apparent beamspot position in the misaligned detector is also accurate. When the wafers are shifted out (in) by 50 microns, the apparent displacement of the beamspot from the center of the detector decreases (increases) by about 4-5 microns or 0.1%. A crude scaling argument would predict that the effect would be less than the wafer displacement divided by the wafer position, or $(50 \mu) / 2.5 \text{ (cm)} = 0.2\%$. The $D_0 vs \phi_0$ plots using both default beam positions and the apparent beam position as determined from our fits are shown in Figs. 28, 29, and 30

We noticed another effect: the resolution deteriorates when the wafers are moved. This increases the apparent size of the beam. This effect vanishes at the center of the wafer and becomes more pronounced on each side of the wafer. The overall size of the effect on D_0 is approximately 30 microns, peak-to-peak.

This effect, which is a D_0 distortion due to the collective shift of wafers, is in fact larger than the overall shift of the beamspot. It has a significant effect on the event selection, migrating events in and out of the acceptance.

In principle, one could hope to use the observed flatness of CDF's beamspot to put an upper limit on the amount of distortion in the real SVX; in practice however we interpret the alignment group's "50 micron" prescription as a characterization of magnitude of possible alignment effects, so, we consider the two cases we study (50 μm in and 50 μm out) as our benchmark worst case scenarios.

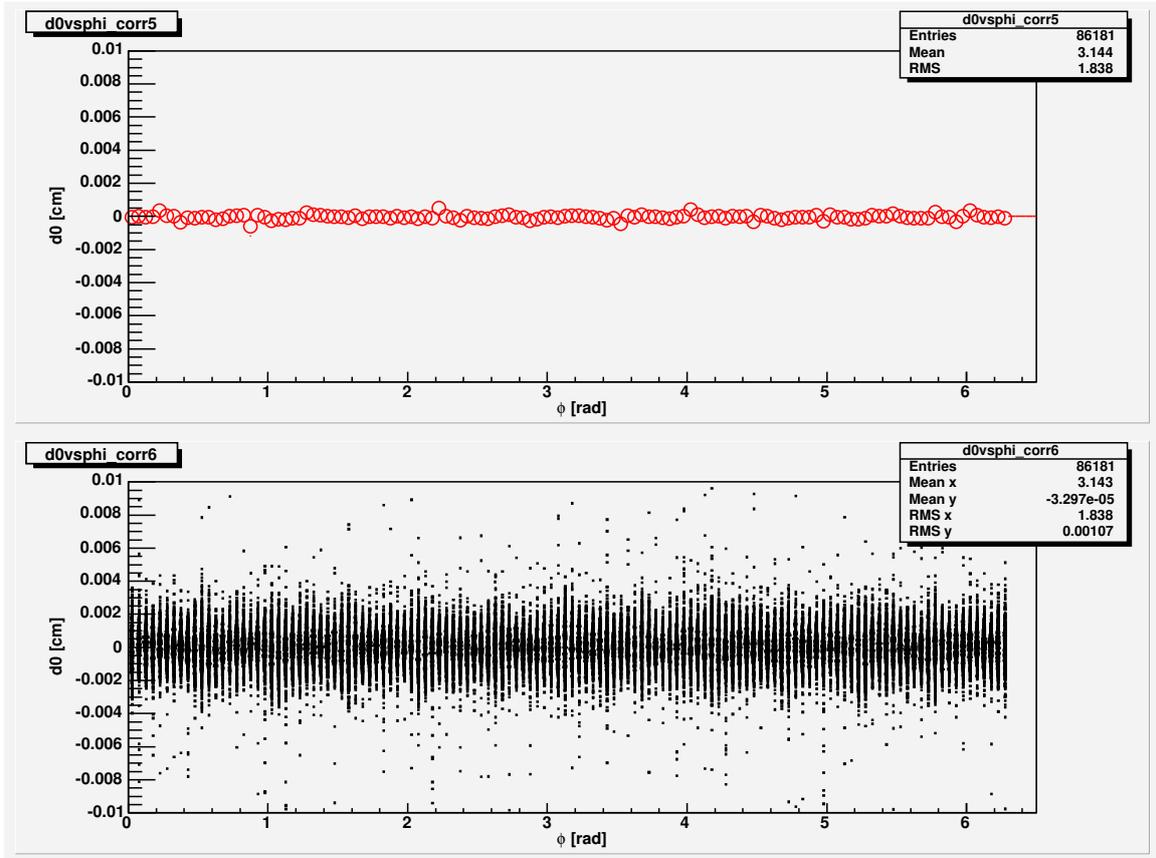


Figure 28: D_0 vs. ϕ_0 plot after correction for the beamspot. Events are single muons at 50 GeV. Events are simulated with tracks at their default positions and reconstructed in the same way.

We can also make the plot of the SVT D_0 vs ϕ_0 , using SVTD banks after the beamspots determined by our procedure have been loaded into SVTSIM. This is shown in Fig 31. This plot is sculpted by the efficiency of the hit-finding and track-finding, and it is difficult to draw conclusions from this plot. However it does appear to rule out large shifts in the SVT due a mismatch between the alignment table and the beamspot numbers fed to svtsim.

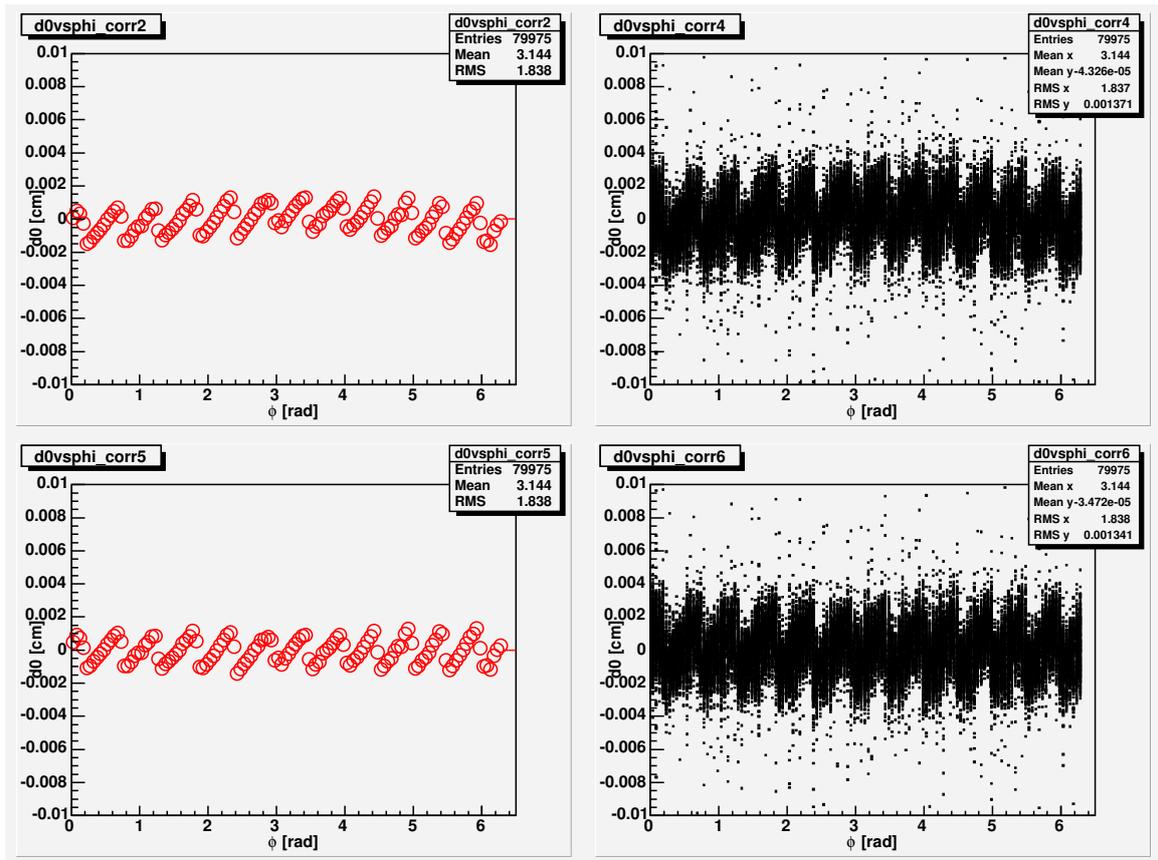


Figure 29: D_0 vs. ϕ_0 plot after correction for the beamspot. Events are single muons at 50 GeV. Events are simulated with tracks at their default positions and reconstructed with wafers moved out, by 50 μ m. Top: the default beamspot position is used. Bottom: misaligned beamspot position is used.

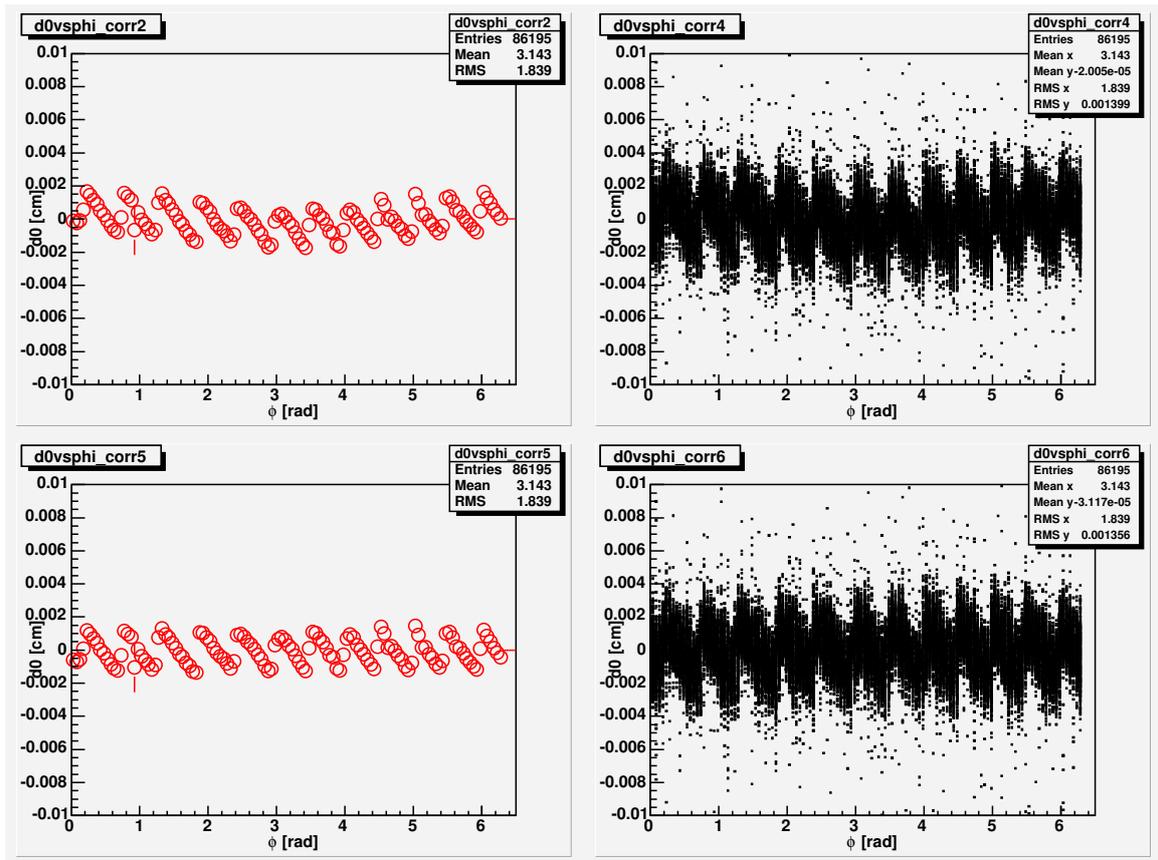


Figure 30: D_0 vs. ϕ_0 plot after correction for the beamspot. Events are single muons at 50 GeV. Events are simulated with tracks at their default positions and reconstructed with wafers moved in, by $50 \mu\text{m}$. Top: the default beamspot position is used: Bottom: misaligned beamspot position is used.

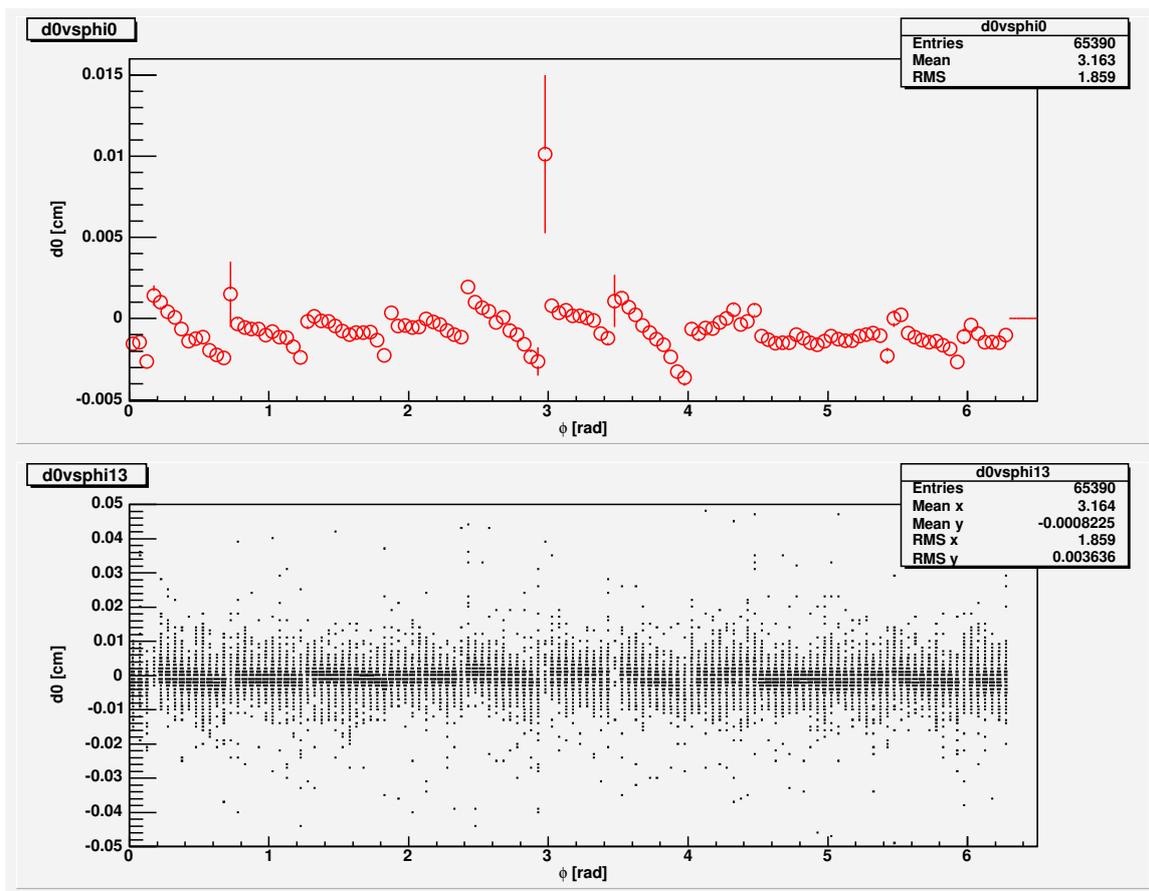


Figure 31: D_0 vs. ϕ_0 plot for the SVT tracks. The apparent beamspot position as determined from our fits are fed to the SVT.

17 Appendix B: Best fit values of the Fisher Scalar

17.1 Table of best fit values for Fisher Scalar

The tables of values for the fit of the Fisher scalar for each B decay mode are presented here. The Fisher scalar is binned by 18 and is fit with an 18th order interpolating Lagrange polynomial. The fit parameters are labelled fb_i since they represent the fraction of signal in each bin of Fisher scalar. Due to a lack of statistics in the extreme bins (different for each B mode) some of the fb_i s are the same, these bins are forced to have the same value.

The results for the B^\pm mode are presented first, please note that for this mode $fb_0 = fb_1 = fb_2$ and $fb_{15} = fb_{16} = fb_{17}$, the table follows:

Parameter	Best Fit value \pm Error
fb_0	0.3028 ± 0.0330
fb_1	0.3028 ± 0.0330
fb_2	0.3028 ± 0.0330
fb_3	0.3878 ± 0.0219
fb_4	0.3843 ± 0.0101
fb_5	0.4097 ± 0.0046
fb_6	0.5769 ± 0.0071
fb_7	0.6368 ± 0.0088
fb_8	0.6260 ± 0.0109
fb_9	0.6686 ± 0.0129
fb_{10}	0.7553 ± 0.0122
fb_{11}	0.8402 ± 0.0111
fb_{12}	0.8554 ± 0.0121
fb_{13}	0.8323 ± 0.0189
fb_{14}	0.8096 ± 0.0292
fb_{15}	0.0714 ± 0.0836
fb_{16}	0.7140 ± 0.0836
fb_{17}	0.7140 ± 0.0836

For the B^0 the results are presented here, for this mode we have $fb_1 = fb_2 = fb_3 = fb_4 = fb_0$ and $fb_{15} = fb_{16} = fb_{17}$, the table follows:

Parameter	Best Fit value \pm Error
fb_0	-0.0005 ± 0.0972
fb_1	-0.0005 ± 0.0972
fb_2	-0.0005 ± 0.0972
fb_3	-0.0005 ± 0.0972
fb_4	-0.0005 ± 0.0972
fb_5	0.9990 ± 0.0791
fb_6	-0.0002 ± 0.0171
fb_7	0.5578 ± 0.0496
fb_8	0.5518 ± 0.0154
fb_9	0.3172 ± 0.0042
fb_{10}	0.5839 ± 0.0066
fb_{11}	0.7002 ± 0.0099
fb_{12}	0.8940 ± 0.0171
fb_{13}	0.7663 ± 0.0452
fb_{14}	0.9436 ± 0.0668
fb_{15}	-0.9999 ± 0.0466
fb_{16}	-0.9999 ± 0.0466
fb_{17}	-0.9999 ± 0.0466