CDF 2008
CDF/ANAL/BOTTOM/9334
18-Mar-2008

A measurement of the $B^{\pm}$ lifetime 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 B lifetimes using 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 $498.2 \pm 6.8 \pm 4.5$ μm in the decay mode $B^{\pm} \to D^0\pi^{\pm}$ with $D^0 \to K^{\mp}\pi^{\pm}$. This compares well with the PDG value of 491.1 ± 2

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. This note contains the original note with additional sections and appendices at the end addressing issues since preblessing.

Contents

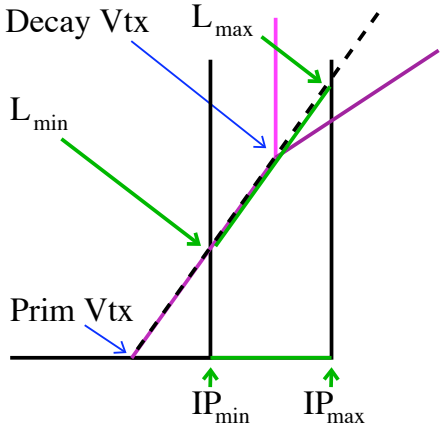
1 Introduction

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

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).



hadronic decays $B_s^0 \to D_s D_s$ [3] and $B_s^0 \to K^+K^-$, which are both CP even, and compare that with the lifetime measured in flavour specific decays like $B_s^0 \to D_s\pi$. Until 2009, significant numbers of $B_s^0 \to D_s D_s$, $B_s^0 \to K^+K^-$ and $B_s^0 \to D_s\pi$ 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

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.

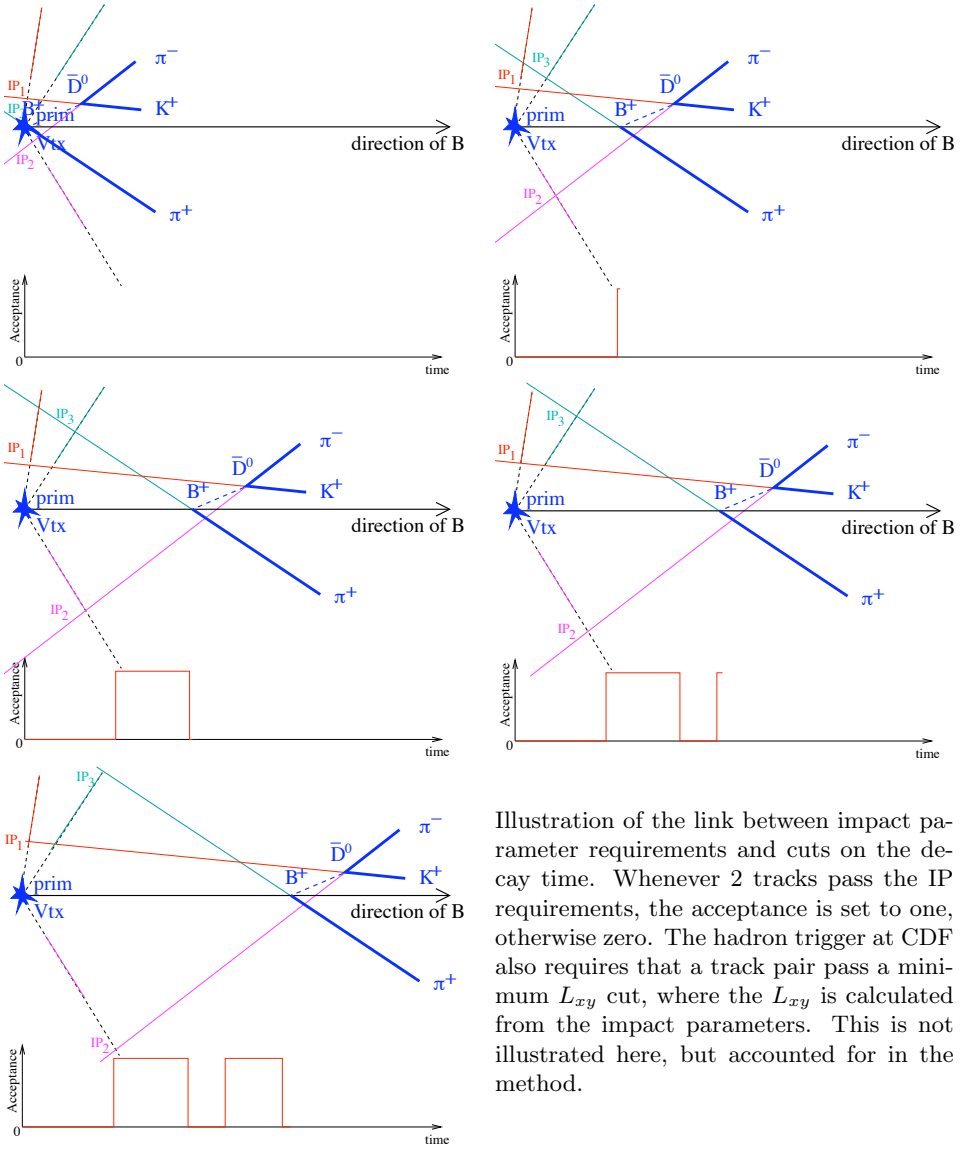


Illustration of the link between impact parameter requirements and cuts on the decay time. Whenever 2 tracks pass the IP requirements, the acceptance is set to one, otherwise zero. The hadron trigger at CDF also requires that a track pair pass a minimum L_{xy} cut, where the L_{xy} is calculated from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

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 perpendicular to them. Where an individual track passs 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 is must be between $t_{\min}$ and $t_{\max}$ and the probability that t is constrained to lie within those limits:

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

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 $\mathrm{P}(t_{\min\, i}, t_{\max\, i})$. However, $\mathrm{P}(t_{\min\, i}, t_{\max\, i})$ 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 $\mathrm{D_s}$ 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 $\log\left(\mathrm{P}(t_{\min\, i}, t_{\max\, i})\right)$ 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:

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

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 looses 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:

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

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 measurment

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 it 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.
$$A_{\rm ip}\left(t_0\right).$$

In terms of these parameters, the total probability is:

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

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

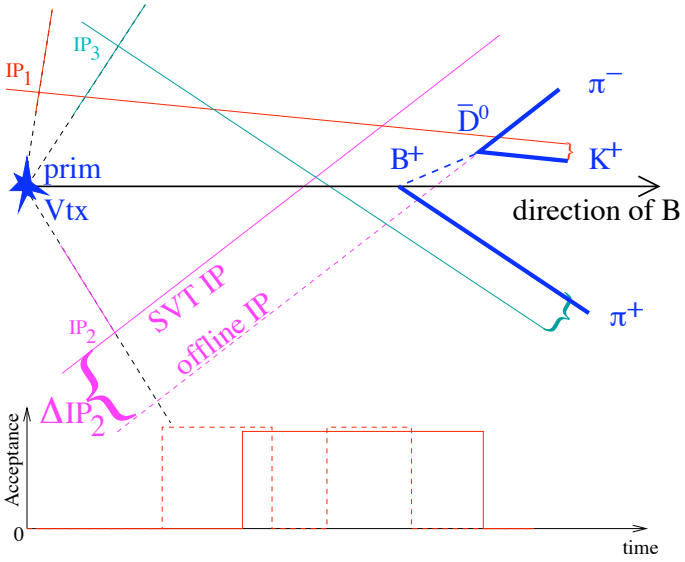
$$A_{\rm ip}\left(t_0, \ldots\right) = \sum_{\substack{i=\text{all}\\ \text{intervals}}} \left(\theta(t_0 - t_{\min\, i}) - \theta(t_0 - t_{\max\, i})\right) \tag{6}$$

where θ is the Heavidside 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.

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



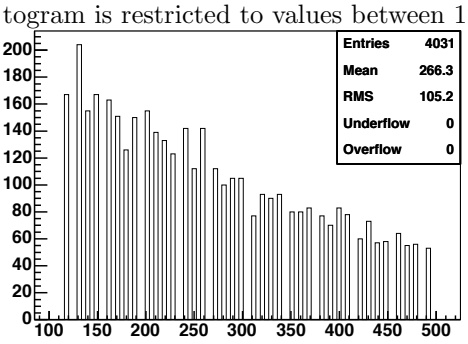
5.1 Using Δd_0

To to 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 $d_0^{\mathrm{SVT}} = d_0^{\mathrm{off}} + \Delta d_0$, 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 is 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 the 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 μm, i.e. possible values are $d_0^{\mathrm{SVT}} = 0, \pm 10\,\mu, \pm 20\,\mu, \pm 30\,\mu, \ldots$. A typical

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



d_0^{SVT} distribution in Monte Carlo, for values between $120\,\mu$ and $500\,\mu$, is shown in Figure 4. Steps of $10\,\mu$ seem small compared to the SVT resolution of about $\sim 50\,\mu$m, but ignoring this discretisation results in a significantly biassed fit result. Ignoring the effect in a fit to 15 k detailed MC events yielded a fit result of $c\tau = 448 \pm 6\,\mu$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 $d_0^{\mathrm{SVT}}(c\tau) = d_0^{\mathrm{off}}(c\tau) + \Delta d_0$, instead the result for d_0^{SVT} is rounded to the nearest $10\,\mu$m.

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

With this modification, the fit result, using the same events, is $494 \pm 7\,\mu$m, in good agreement with the input value of $496\,\mu$m. More on detailed and toy Monte Carlo studies in Section 12.

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:

$$A_{\mathrm{ip}}(t_0, \ldots) = \sum_{\substack{i=\text{all} \\ \text{intervals}}} \left(\theta(t_0 - t_{\min\, i}) - \theta(t_0 - t_{\max\, i})\right) \tag{7}$$

With this equation 5 becomes:

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

measurement. Using the frequency function

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

this can be written as

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

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:

$$P(t_0) = \frac{\frac{1}{\tau} e^{\frac{-t_0}{\tau}} \mathrm{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{\substack{i=\text{all}\\ \text{intervals}}} \left[-e^{\frac{-t}{\tau}} \mathrm{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + e^{-\frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\min\,i}}^{t=t_{\max\,i}}} \tag{11}$$

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 $B_d \to \pi\pi$, $B_s \to KK$), 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 $|d_0| < 1\,\text{mm}$. For the 2-track case, the trigger cuts ensure $|d_0| < 1\,\text{mm}$ for both tracks. Since no track with $|d_0| > 1\,\text{mm}$ 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 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.

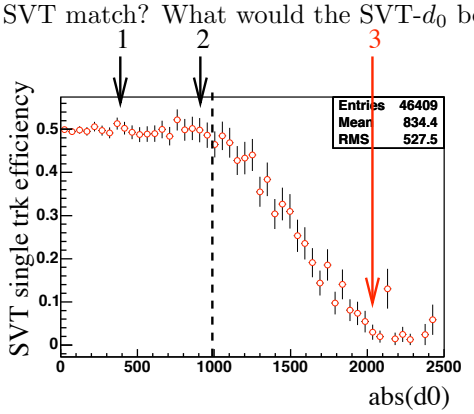
If the SVT track finding efficiency were indeed independent of $c\tau$, the following argument would save us the complicated calculation: We could simply calculate

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

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

Unfortunately, beyond $|d_0| = 1\,\text{mm}$, 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 $|d_0| > 1\,\text{mm}$ are never seen because of the very trigger requirements we are correcting for. But in a three body decay, two tracks with $0.12\,\text{mm} < |d_0| < 1\,\text{mm}$ can fire the trigger, while the third track

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



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 $< 1\,\rm mm$, 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?
- 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 $|d_0| > 1\,\rm mm$, 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 chosing 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 $|d_0| = 1\,\rm mm$. 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 $|d_0| < 1\,\rm mm$. 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 $|d_0| > 1\,\text{mm}$ 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].

Since both simple solutions outlined above are too costly in statistical precison, 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.

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

Figure 6: Δd_0 Distribution for tracks in $B_u \to D\pi$ candidates.

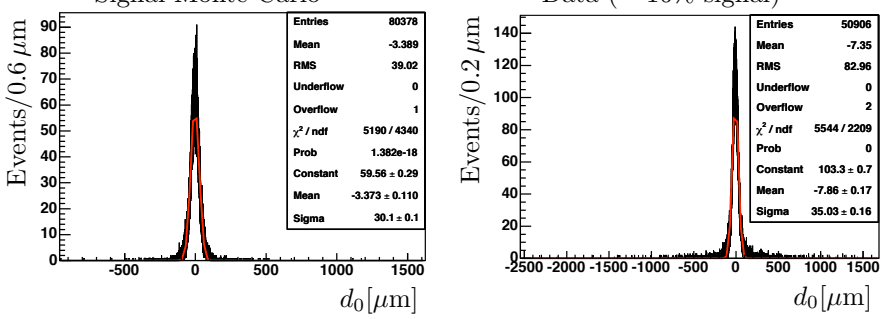


Table 1: Trigger efficiency in terms of the SVT-single track finding efficiency, for a three particle final state.

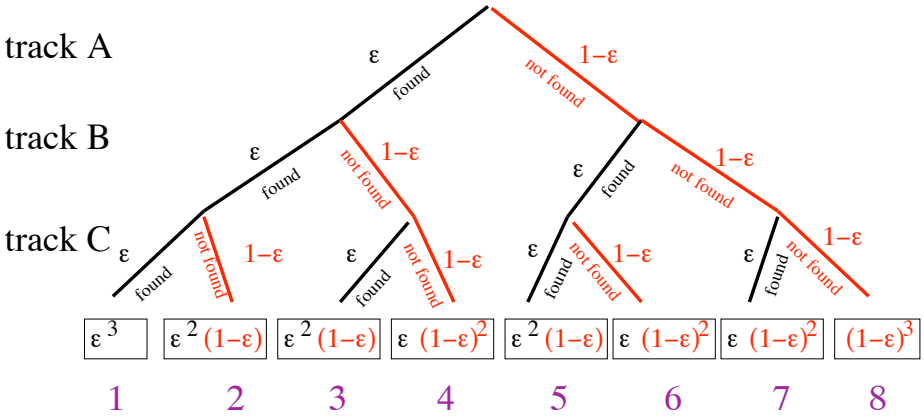
Number of track-pairs passing the trigger cuts, out of 3 tracks	Trigger efficiency in terms of ε_s
1 track pair	ε_s^2
2 track pairs	$2\varepsilon_s^2 - \varepsilon_s^3$
3 track pairs	$3\varepsilon_s^2 - 2\varepsilon_s^3$

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 $d_0^{\rm SVT}$ 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 $d_0^{\rm SVT}$ 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 $|d_0^{\rm off}| = 1\,{\rm mm}$ we describe the SVT single track finding efficiency as flat for $|d_0^{\rm off}| < 1\,{\rm mm}$ and zero elsewhere. For this to be accurate, we have to treat those tracks with $|d_0^{\rm off}| > 1\,{\rm mm}$ 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 $|d_0^{\rm off}| < 1\,{\rm mm}$, 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 effi-

Figure 7: Decision tree for the example of 3 tracks in the final states. Each track can either be found (probabilitye 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.



ciencies, 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 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 minumum 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 $|d_0^{\text{off}}| < 1\,\text{mm}$ and a minimum p_t of $2\,\text{GeV}$. The probabilites 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. Fortunalety 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 Changes in the SVT

The final issue that needs addressing is how the fit copes with data that contains periods of distinct different efficiency. The periods of distinct efficiency arise due to changes in the SVT algorithms or hardware. Over the period of data taking that results in 1fb of data there are 4 changes to the SVT. They are

- The initial running from start to 12/26/02
- On 12/26/02 begin using 4/5 logic for the silicon hits
- On 07/24/04 use new patterns and new geometry file
- On 07/22/05 installation of new AM board
- On 08/31/05 move to using 128K patterns

If we plot the efficiency in these different periods we see that the last two had little effect on the efficiency in the impact parameter range of 0-1000 microns, the only region we are interested in. While they do increase the efficiency as a function of impact parameter beyond 1000 microns it is irrelavent to this analysis as all tracks with matches above 1000 microns are deemed to have been not found. The first two changes have made significant increases to the efficiency. These results are shown in two plots below in figure 8

Now that we know there are two significant changes in the SVT efficiency we must consider whether it is necessary to fit 3 separate efficiencies based upon run number or whether a single averaged efficiency will be sufficient. To test this we generate toyMC that has sections with differing efficiency as found in data and fit the lifetime with a single floating efficiency. We generate 1000 pseudo experiments of 24K events each. We find that although there is no significant pull in the lifetime distribution the fitted gaussian is wide, figure 9. This implies that in using this single efficiency when there were infact 3 present leads to an underestimation of the statistical error by almost 20%. We try therefore instead to fit the 3 efficiencies simultaneously with an event having sensitivity to a particular parameter if its "run" falls into the correct range. We run the pull study again and find that the the fitted gaussian to the pull distribtuion is

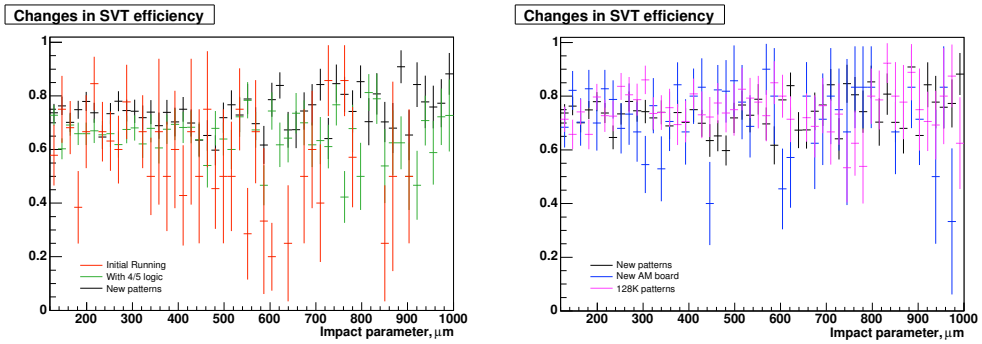


Figure 8: SVT track finding efficiency as a function of impact parameter and split into different time periods. The top plot shows the first two changes that did affect the efficiency in the range of interest. The lower plots shows the final two changes which did not affect significantly the data in the range of interest.

centred on 0 and has unit width, as shown in figure 10. We decide therefore to fit 3 efficiencies simultaneously with the lifetime. Each efficiency will be fit by events corresponding to that run period.

6.4.5 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-biassed signal sample. For multibody decays the PDF needs to 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.
- $P(\text{trigger}|trk, t_o)$ 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, $p(\text{trigger}|\varepsilon_s, t_o) = \sum_{\text{trk}} p(trk|\varepsilon_s)P(\text{trigger}|trk, t_o)$. This corresponds to the entries in Table 1. It is essentially the normalisation factor to go with $P(\text{trk}|\varepsilon_s)P(\text{trigger}|trk, t_o)$
- $\text{poly}_i(\varepsilon_s)$: Since $P(\text{trigger}|\varepsilon_s)$ is constant for t_0 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).

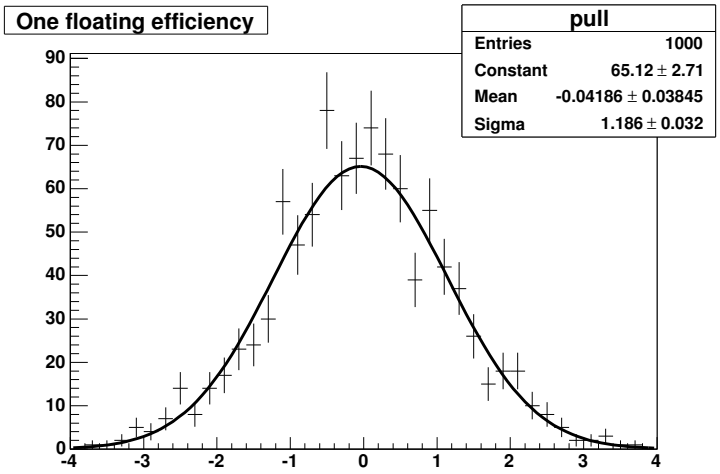


Figure 9: Pull distribution of the lifetime fit if only one floating efficiency is allowed.

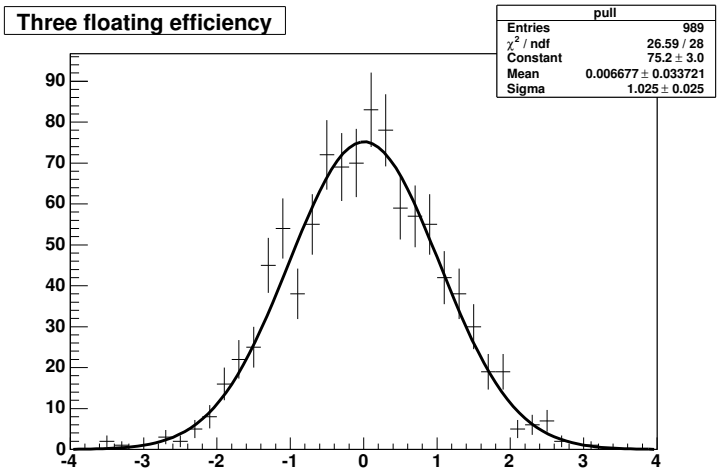


Figure 10: Pull distribution of the lifetime fit if 3 separate floating efficiencies are allowed.

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

$$
\begin{aligned}
P\left(t_0\right) &= \frac{P(\mathrm{trk}|\varepsilon_s)P(\mathrm{trigger}|\mathrm{trk},t_o)\frac{1}{\tau}e^{\frac{-t_0}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t_0}{\sigma}-\frac{\sigma}{\tau}\right)}{\sum\limits_{\text{all trk}}\sum\limits_{\substack{i=\text{all}\\ \text{intervals}}}\int\limits_{t_{\min\,\mathrm{i}}}^{t_{\max\,\mathrm{i}}} P(\mathrm{trk}|\varepsilon_s)P(\mathrm{trigger}|\mathrm{trk},t_o)\frac{1}{\tau}e^{\frac{-t_0}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t_0}{\sigma}-\frac{\sigma}{\tau}\right)\,dt_0} \\
&= \frac{P(\mathrm{trk}|\varepsilon_s)\frac{1}{\tau}e^{\frac{-t_0}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t_0}{\sigma}-\frac{\sigma}{\tau}\right)}{\sum\limits_{\substack{i=\text{all}\\ \text{intervals}}}\mathrm{poly}_i(\varepsilon_s)\left[-e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t}{\sigma}-\frac{\sigma}{\tau}\right)+\mathrm{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\min\,\mathrm{i}}}^{t=t_{\max\,\mathrm{i}}}} \qquad (12)
\end{aligned}
$$

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

6.5 Toy-MC

In order to test the basic principle, a toy-Monte Carlo simulation is used that generates isotropic $\mathrm{B_s^0} \to \mathrm{D_s}\pi$ events with a mean $\mathrm{B_s^0}$-lifetime of 1.55 ps and a mean $\mathrm{D_s}$-lifetime of 0.49 ps. The 2-D impact parameter resolution is assumed to be Gaussian with 33 μm(intrinsic) $\otimes$ 33 μm(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:

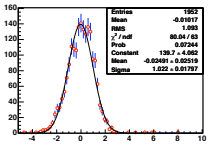
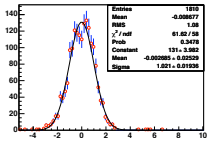
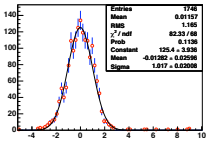
- "Standard": A simple Gaussian resolution function with $\sigma = 33\,\mu$, as described above.
- "Offset": A Gaussian resolution function with $\sigma = 33\,\mu$, with an offset of $33\,\mu$, i.e. the mean measured SVT impact parameter is $33\,\mu$ larger than the true one.
- "Exponential from hell": A positive exponential with an rms of $33\,\mu$, 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 $33\,\mu$.

None of these rather drastic biases produces a bias in the fitted lifetimes (Figure 11).

6.6 Detailed MC

The method has been tested on a Monte Carlo sample of $\mathrm{B_u} \to \mathrm{D}\pi$ 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 $496\,\mu$. A projection of the fit to the MC data is shown in figure 12.

Figure 11: Toy MC pulls $1-2k$ MC experiments, $0.5k$ signal evts each, $S/B=1$, with different intrinsic IP-resolutions

Gaussian Error	Gauss + $33\,\mu$ offset	positive Exp-error
		
Mean $= -0.025 \pm 0.025$	Mean $= -0.003 \pm 0.025$	Mean $= -0.013 \pm 0.026$
Sigma $= 1.022 \pm 0.018$	Sigma $= 1.021 \pm 0.019$	Sigma $= 1.017 \pm 0.020$

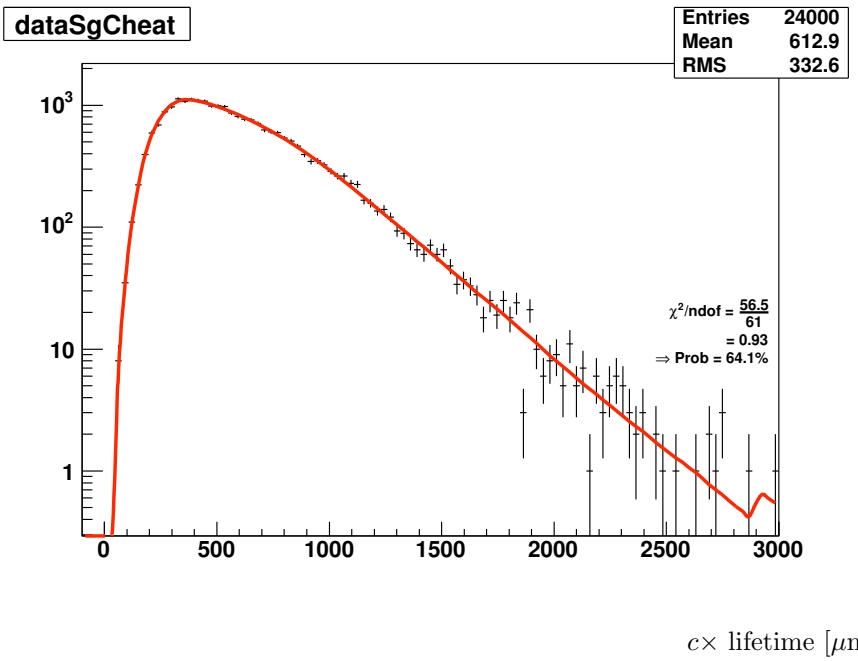


Figure 12: Monte Carlo-independent lifetime fit (line) to 24k simulated $\mathrm{B}^+ \to \mathrm{D}\pi$ events (crosses), subject to the impact parameter trigger at CDF, using a detailed detector simulation. MC-input: $c\tau = 496\,\mu$m. Fit result: $c\tau = 491.4 \pm 5\,\mu$m.

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 efficienty 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

$$P(c\tau, \text{kin}) = P(c\tau|\text{kin})P(\text{a}) \tag{13}$$

we can ignore $P(\text{kin})$ because it is a simple factor and $\frac{\text{d}}{\text{d}\tau}P(\text{kin}) = 0$. 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)

$$P(c\tau, A_{\text{trig}}) = P(s)P(c\tau|A_{\text{trig}}, s)P(A_{\text{trig}}|s) + P(b)P(c\tau|A_{\text{trig}}, b)P(A_{\text{trig}}|b) \tag{14}$$

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 14 as

$$\begin{aligned} P(c\tau, A_{\text{trig}}) &= P(s)P(A_{\text{trig}}|s)P(c\tau|A_{\text{trig}}, s) + P(b)P(A_{\text{trig}}|b)P(c\tau|A_{\text{trig}}, b) & (15)\\ &= \{P(s|A_{\text{trig}})P(c\tau|A_{\text{trig}}, s) + P(b|A_{\text{trig}})P(c\tau|A_{\text{trig}}, b)\}\, P(A_{\text{trig}}) & (16) \end{aligned}$$

So we can either, as in Equation 15, fit the probability to find a given efficiency function, or at least, as in Equation 16 calculate an event-by-event signal probability based on the efficiency function. The last term in Equation 16, 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 alltogether, 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 ex ample 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"
- $P(AorB)$ "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
$$P(A, B) = P(A)P(B|A) \tag{17}$$
- ... which leads to Bayes' theorem
$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} \tag{18}$$
- A or B.
$$P(A\text{or}B) = P(A) + P(B) - P(A, B) \tag{19}$$

We include the following measured quantities in our fit:

- The measured lifetime, t_o.
- 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.
- The measured lifetime error is NOT used as it is correlated to other quantities resulting in problems described later. Instead we use a average resolution for all events.

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

$$P\left(t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \tag{20}$$

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 accepance 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 20 into a signal and a background part. We use the letters s and b for signal and background. Using Equation 19, $P(s,b) = 0$, and $P(s\text{or}b) = 1$:

$$\begin{aligned} P\left(t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) &= P\left(s, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \\ &+ P\left(b, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \end{aligned} \tag{21}$$

In the following, we focus on the first term on the right hand side in Equation 21, only:

$$P\left(s, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \tag{22}$$

The results for

$$P\left(b, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \tag{23}$$

will be analogous.

So far, nothing has happened. Using Equation 18 on Equation 22 gives:

$$\begin{aligned} &P\left(s, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \\ &= \frac{P\left(s, t_o, m, trk, A_{\text{trig}}\right) P\left(\text{trigger}|s, t_o, m, trk, A_{\text{trig}}\right)}{P(\text{trigger})} \end{aligned} \tag{24}$$

Note that $P\left(\text{trigger}|s, t_o, m, trk, A_{\text{trig}}\right)$ 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 24 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, $P(\text{trigger}|A_{\text{trig}})$. 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 $P(\text{trigger}|A_{\text{trig}}, s)$. Using Bayes' theorm (18), we find for the denominator in Equation 24:

$$P(\text{trigger}) = P(\text{trigger}|A_{\text{trig}}, s)\frac{P(A_{\text{trig}}, s)}{P(A_{\text{trig}}, s|\text{trigger})} \tag{25}$$

The left-hand term in the numerator of Equation 24 can be written as

$$P\left(s, t_o, m, trk, A_{\text{trig}}\right) = P\left(s, A_{\text{trig}}\right) P\left(t_o, m, trk|s, A_{\text{trig}}\right) \tag{26}$$

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

$$\begin{aligned} &P\left(s, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \\ &= \frac{P\left(t_o, m, trk|s, A_{\text{trig}}\right) P(A_{\text{trig}}, s|\text{trigger}) P\left(\text{trigger}|\text{all}\right)}{P(\text{trigger}|A_{\text{trig}}, s)} \end{aligned} \tag{27}$$

The first term in the numerator can be split further

$$\begin{aligned} & P\left(t_o, m, trk|s, A_{\text{trig}}\right) \\ & \quad = \quad P\left(t_o|s, A_{\text{trig}}\right) P\left(trk|t_o, s, A_{\text{trig}}\right) P\left(m|trk, t_o, s, A_{\text{trig}}\right) \end{aligned} \tag{28}$$

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 $P(trk|A_{\text{trig}})$ 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 independend of *trk*, t_o, A_{trig}.

With this we get:

$$P\left(t_o, m, trk|s, A_{\text{trig}}\right) \quad = \quad P\left(t_o|s, A_{\text{trig}}\right) P\left(trk|t_o, s\right) P\left(m|s\right) \tag{29}$$

so our PDF is now:

$$\begin{aligned} & P\left(s, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \\ & \quad = \quad \frac{P\left(t_o|s,\right) P\left(trk|t_o, s\right) P\left(m|s\right) P(A_{\text{trig}}, s|\text{trigger}) P\left(\text{trigger}|\text{all}\right)}{P(\text{trigger}|A_{\text{trig}}, s)} \end{aligned} \tag{30}$$

Finally, we'll have to deal with the second but last term in the numerator, $P(A_{\text{trig}}, s|\text{trigger})$. Note that the condition "|trigger" ensures that we need to look only at quantities as the 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. $P(s|\text{trigger})P(A_{\text{trig}}|s, \text{trigger})$, where the first term is simply the overall signal fraction after the trigger, i.e. in the data we see. The other term fits the A_{trig} distribution (how to fit a distribution of acceptance functions is the subject of an entire section later on).
2. $P(A_{\text{trig}}, |\text{trigger})P(s|A_{\text{trig}}, \text{trigger})$. Here we can ignore the first term, as it does not depend on the parameters we are interersted 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. 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:

$$\begin{aligned} & P\left(s, t_o, m, trk, A_{\text{trig}}|\text{trigger}\right) \\ = \;& P\left(s|A_{\text{trig}}, \text{trigger}\right) \times \\ & \frac{P\left(t_o|s,\right) P\left(m|s\right) P\left(trk|t_o, s\right) P\left(\text{trigger}|\text{all}\right)}{P(\text{trigger}|A_{\text{trig}}, s)} \end{aligned} \tag{31}$$

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

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

$$P\left(t_o|s, \sigma_t\right) = \frac{1}{\tau} e^{\frac{-t_0}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right) \tag{32}$$

$P\left(t_o|b, \sigma_t\right)$ The background lifetime model is discussed in its own section **??**.

$P\left(m|s\right)$ 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 10.

$P\left(m|b\right)$ The background mass distribution is a first order polynomial. Reasons for this choice are detailed in section 10.

$P\left(trk|t_o, s\right)$ 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

$$P\left(trk|t_o, s\right) = P\left(\text{SVT found } k \text{ tracks out of } n|t_o, s\right) = \varepsilon_s^k \left(1 - \varepsilon_s\right)^{(n-k)} \tag{33}$$

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 $|d_0^{\text{off}}| < 1\,\text{mm}$ 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.

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

$$P\left(\text{trigger}|\text{all}\right) = \begin{cases} 1 & \text{if event passes trigger cuts} \\ 0 & \text{else} \end{cases} \tag{34}$$

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.

$P(\text{trigger}|A_{\text{trig}}, s)$ The probability that the trigger fires, given the efficiency function A_{trig}It's the normalisation factor. It's given by:

$$\sum_{\text{all trk}} \sum_{\substack{i=\text{all} \\ \text{intervals}}} \int_{t_{\text{min i}}}^{t_{\text{max i}}} P\left(trk|t_o, s\right) P\left(\text{trigger}|\text{all}\right) \frac{1}{\tau} e^{\frac{-t_0}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right) dt_0$$

$$= \sum_{\substack{i=\text{all} \\ \text{intervals}}} \text{poly}_i(\varepsilon_s) \left[-e^{\frac{-t}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + \text{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\text{min i}}}^{t=t_{\text{max i}}} \quad (35)$$

$P\left(s|A_{\text{trig}}, \text{trigger}\right)$ 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

$$P(s|\text{Number}(\text{A}_{\text{trig}})) \approx P(s|\text{A}_{\text{trig}}) \quad (36)$$

to a good-enough approximation. This means that the characteristic number must be chosen such as to minimise the information loss in the process $(\text{A}_{\text{trig}}) \rightarrow$ number, in terms of the signal-ness or background-ness of the acceptance function. A relatively simple number to calulate 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 $P\left(s|A_{\text{trig}}, \text{trigger}\right)$.

8 Fisher Discriminant to calculate $P(s|acc)$

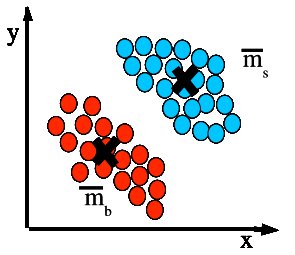
$P(s|acc)$: The more complicated term to deal with is the acceptance function. The use of fisher discriminants is introduced to deal with this term.

8.1 The use of Fisher Discriminants

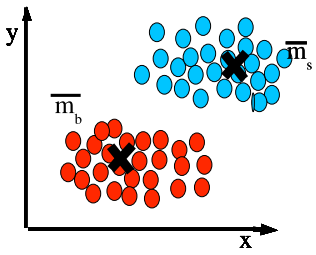
If our acceptance function were characterizable using a set of variables then a Fisher Discriminant Analysis could be used to separate signal and background in the way required above. Our acceptance functions typically have the form of top hat functions over intervals of $c\tau$, so we can sample the 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

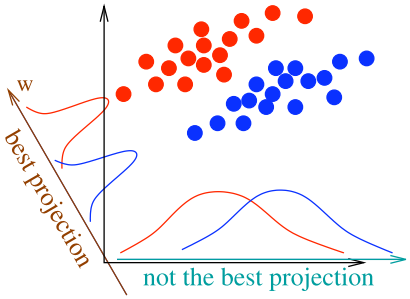
Imagine two classes of events, eg signal and background with their own distributions of variable x and y as shown in figures 13(a), 13(b) and 13(c). The means of variable x and y for each distribution are shown as the points $\overline{m_s} = (\overline{x_s}, \overline{y_s})$ and $\overline{m_b} = (\overline{x_b}, \overline{y_b})$. 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.



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

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 37, and gives the definition for the matrix we refer to as S_M.

$$\begin{aligned}(< w|\overline{m_s} > - < w|\overline{m_b} >)^2 &=< w|(\overline{m_s} - \overline{m_b}) >< (\overline{m_s} - \overline{m_b})|w > \\ &=< w|S_M|w >\end{aligned} \tag{37}$$

Secondly we consider the square of the spread of the signal events around the projected mean, $Scat^2_{sig}$ which leads to the definition of the matrix S_{sig} as shown

in 38, where p_i =$((x_i, y_i))$.. There is a similar expression for the background events, S_{bg}.

$$\begin{aligned} Scat^2_{sig} &= \sum_{All\ signal\ events} (< w|p_i > - < w|\overline{m_s} >)^2 \\ &= \sum_{All\ signal\ events} < w|(p_i - \overline{m_s}) >< (p_i - \overline{m_s})|w > \\ &= \sum_{All\ signal\ events} < w|S_{sig}|w > \end{aligned} \tag{38}$$

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 39.

$$J(w) = \frac{< w|S_M|w >}{< w|(S_{sig} + S_{bg})|w >} = \frac{< w|S_M|w >}{< w|(S_w)|w >} \tag{39}$$

From equations 40 and 41 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.

$$\nabla w(J(w)) = \frac{2S_M|w >}{< w|S_w|w >} - \frac{< w|S_M|w >}{< w|S_w|w >} . \frac{2S_w|w >}{< w|S_w|w >} = 0 \tag{40}$$

$$\begin{aligned} S_M|w > -\lambda S_w|w > &= 0 \\ S_M|w > &= \lambda S_w|w > \\ S_w^{-1} S_M|w > &= \lambda|w > \end{aligned} \tag{41}$$

Using the definition of S_M from equation 37 we can rewrite $S_M|w >$ in the following way.

$$\begin{aligned} S_M|w > &= |(\overline{m_s} - \overline{m_b}) >< (\overline{m_s} - \overline{m_b})|w > \\ S_M|w > &\propto |(\overline{m_s} - \overline{m_b}) > \end{aligned} \tag{42}$$

If we insert this into equation 40 we see that it is not necessary to solve for the eigenvalues and that all we require to find $|w >$ is the inverse of S_W and $|(\overline{m_s} - \overline{m_b} >$.

$$S_w^{-1}|(\overline{m_s} - \overline{m_b}) > \propto |w > \tag{43}$$

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 $\mathbf{w}.\mathbf{v}$ 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 $\mathbf{w}.\mathbf{v}$ the fisher scalar.

8.3.1 Acceptance function → 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 14 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. This is illustrated in figure 14. 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 reason the acceptance function in split into regions of different heights is to allow the acceptance function to be independent of the efficiency for the fisher discriminant part of the analysis.

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. The height of the 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 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

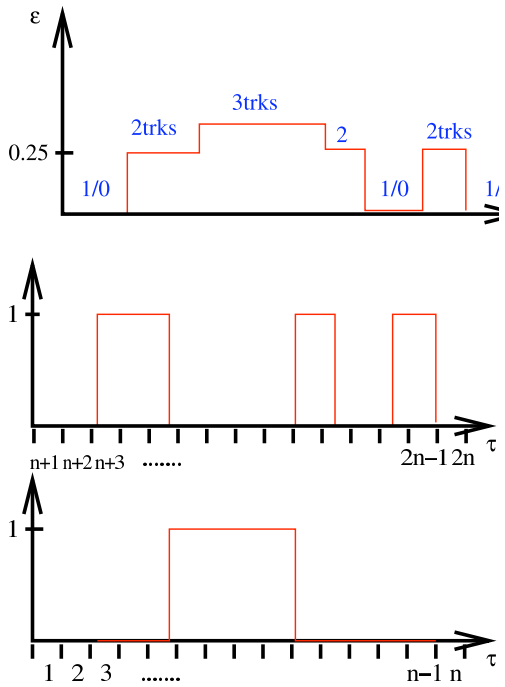


Figure 14: Splitting the acceptance function into the sum of its parts.

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. 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 $|\overline{m_s}$ - $\overline{m_b} >$ so that we can find the vector w and hence find the fisher scalar for every event.

8.3.2 Extracting ($|\overline{m_s}$ - $\overline{m_b} >$) 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 $|\overline{m_b} >$ 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 $|\overline{m_r} >$ which we

can write as equation 44 where we know f_s from the mass fit.f_s is the fraction of signal in the signal region. The vector ($|\overline{m_r}$ - $\overline{m_b}>$) is proportional to ($|\overline{m_s}$ - $\overline{m_b}>$). This is our vector for the difference between the means.

$$\begin{aligned} \overline{m_r} &= \frac{\sum_{\substack{events\ in \\ signal\ region}} |v_i>}{\sum_{\substack{events\ in \\ signal\ region}} 1} \\ |\overline{m_r}> &= f_s|\overline{m_s}> + (1-f_s)|\overline{m_b}> \\ |\overline{m_r} - \overline{m_b}> &= f_s|(\overline{m_s}> - |\overline{m_b}>) \end{aligned} \tag{44}$$

We may find that some of the variables in the vector $|\overline{m_s}$ - $\overline{m_b}>$ 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 ($|\overline{m_s}$ - $\overline{m_b}>$) by removing rows where the entry in ($|\overline{m_s}$ - $\overline{m_b}>$) 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 45 where v_s and v_b are the acceptance vectors of pure signal and pure background events respectively. The definition is taken from equations 38 and 39.

$$S_w = \sum_{\substack{Signal \\ events}} |(v_s - \overline{m_s})><(v_s> - \overline{m_s})| + \sum_{\substack{Background \\ events}} |(v_b - \overline{m_s})><(v_b - \overline{m_s})| \tag{45}$$

Again we consider the signal region and background region.

We can calculate the matrix called S_{bk} as given in equation 46 trivially as we already have $\overline{m_b}$.

$$S_{bk} = \sum_{\substack{Sideband \\ events}} |(v_i - \overline{m_b})><(v_i - \overline{m_b})| \tag{46}$$

We can also calculate the matrix called S_{bkassg} given in equation 47. We calculate $|\overline{m_s}>$ as we know the value of f_s from the mass fit and we know f_s*$|(\overline{m_s}$-$\overline{m_b})>$ and $|\overline{m_b}>$.

$$S_{bassig} = \sum_{\substack{Sideband \\ events}} |(v_i - \overline{m_s})><(v_i - \overline{m_s})| \tag{47}$$

The matrix S(sg +b) is calulated in 48.

$$\begin{aligned} S_{sg+b} &= \sum_{\substack{Signal\\ region\\ events}} |(v_i - \overline{m_s}) >< (v_i - \overline{m_s})| \\ &= \sum_{\substack{Signal\\ events\ insignal\\ region}} |(v_s - \overline{m_s}) >< (v_s - \overline{m_s})| \\ &+ \sum_{\substack{background\\ events\ in\ signal\\ region}} |(v_b - \overline{m_s}) >< (v_b - \overline{m_s})| \end{aligned} \tag{48}$$

We can calculate the 3 matrices in equations 46, 47 and 48 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 49.

$$\begin{aligned} Sw = S_{sg+B} \quad &- \quad \frac{\text{N}^0 \text{ Background in sideband}}{\text{N}^0 \text{ Background in signal region}} \times S_{bassig} \\ &+ \quad \frac{\text{N}^0 \text{ Background in dataset}}{\text{N}^0 \text{ Background in sideband region}} \times S_{bk} \end{aligned} \tag{49}$$

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 our toy MC, details of which come later. For the purpose of defining the Fisher direction and calculating the Fisher discriminant, the toy mc was used like any other data sample, the information which event came from signal and which from background. 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].

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

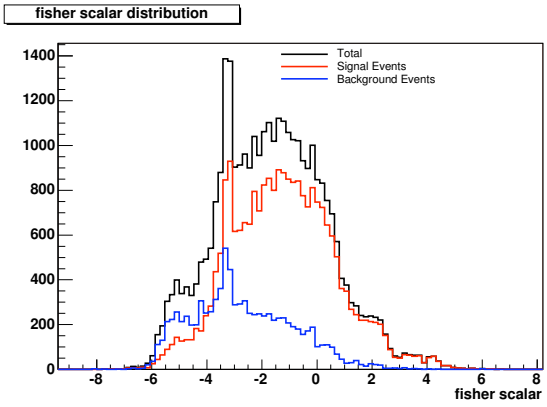


Figure 15: The distribution of fisher scalar is shown here from signal Monte Carlo is red and that from background is blue. It is clear that their distributions are different.

Figure 16. 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 use about 15 - 20 bins, and pick 15 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. 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. 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. This results in an effective 13th order polynomial. The choice of number of bins and the order of the polynomial is tested in crosschecks and we note here that as we reduce the order of polynomial from 13 to 9 there is no change in the fitted region. We conclude that this choice gives a sufficiency fit.

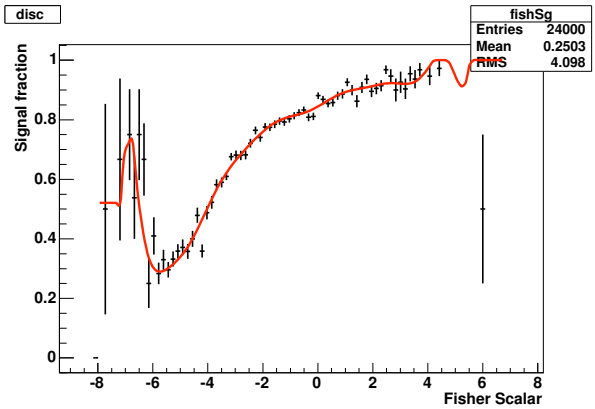


Figure 16: The data points show the truth signal fraction in the fit. The red line shows the fitted function. The data points are shown to demonstrate that the correct signal fraction has been found. The fit itself does not know which events are signal or background so does not know the truth, yet manages to match it well.

9 Using an overall, average $\sigma_{c\tau}$, motivation, verification and choice

.

At this point we make a small diversion to explain the choice of resolution for this analysis and also to explain why it is not possible to use the event by event measured lifetime error.

Since the calculated uncertainty of each events proper decay length $c\tau$, $sigma_{c\tau}$ is correlated with the event's acceptance function we have to take this into consideration before extracting a lifetime from a fit. This is illustrated by the following plots of the event by event resolution

Formerly we had attempted to deal with this correlation by including the σ_{ctau} into the vector of quantities from which a Fisher scalar discriminant was to be extracted. The hope was that this would decorrelate the appropriate variables and provide a scalar distribution that would allow us to determine the probability of an event being signal given a particular $\sigma_{c\tau}$ *and* acceptance function. We have since determined that this procedure has potential pitfalls:

1. This method of accounting for correlation works best for strong and linear correlations, which is not the case as illustrated in Fig 17

   We also note the following:

   i. Ignoring the different background and signal acceptance function distributions biases the lifetimes to a lower value.

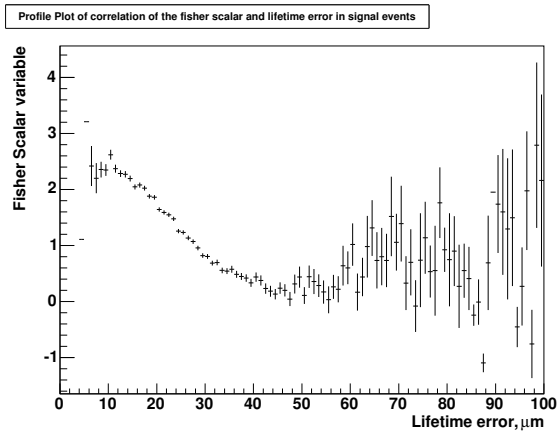
   ii. Ignoring the different background and signal $\sigma_{c\tau}$ distributions biases the lifetimes high, see CDFNOTE 8524 and others.

2. Given i and ii, we can never be sure if these effects have been accounted for separately or have simply canceled. The use of such a technique in a mode with different topology would have a different effect.

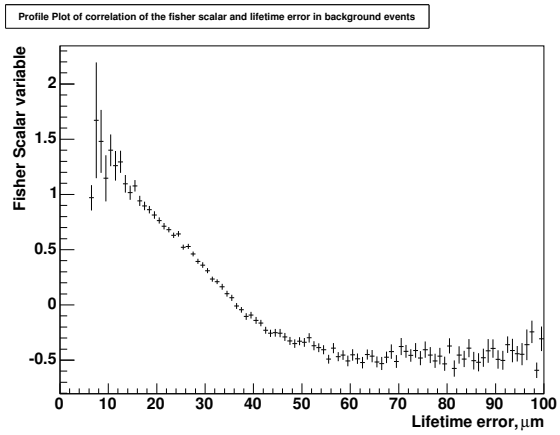
In the light of the above we propose the use of an overall $\sigma c\tau$ in place of using the event by event quantity. We demonstrate that this has an effect at the submicron level on the best fit lifetimes. We also determine what an appropriate value for such an $\sigma_{c\tau}$ would be using data and background subtraction however the reader will note the choice hardly matters as demonstrated below.

9.1 Demonstrating the effect of an overall $\sigma_{c\tau}$

To demonstrate the effect of using an overall resolution rather than the event by event quantity we take a sample of roughly 80,000 Monte-Carlo events for the charged decay mode and fit this sample using the event by event and several different overall $\sigma_{c\tau}$ the results are tabulated for each mode below.



(a) Signal events



(b) Background events

Figure 17: Profile Plots illustrating the correlation between fisher scalar and lifetime error for signal and background events. Realistic Monte Carlo was used for the signal and upper sideband was used for the Background. The fisher direction was calculated using the events from these 2 samples.

Table 2: 80000 realistic MC $B^{\pm} \to D^0\pi^{\pm}$ truth lifetime 497 μm

Configuration	Best Fit ±Error
Event by Event $\sigma_{c\tau}$	494.149 ± 3.083 μm
$\sigma_{c\tau}$=2 μm (fixed)	494.158 ± 3.083 μm
$\sigma_{c\tau}$=21 μm (fixed)	494.158 ± 3.083 μm
$\sigma_{c\tau}$=26 μm (fixed)	494.156 ± 3.083 μm
$\sigma_{c\tau}$=32 μm (fixed)	494.159 ± 3.083 μm
$\sigma_{c\tau}$=100 μm (fixed)	490.694 ± 3.058 μm

Table 3: 20000 realistic MC $B^{\pm} \to D^0\pi^{\pm}$ truth lifetime 497 μm

Configuration	Best Fit ±Error
Event by Event $\sigma_{c\tau}$	496.116 ± 6.070 μm
$\sigma_{c\tau}$=2 μm (fixed)	496.128 ± 6.070 μm
$\sigma_{c\tau}$=21 μm (fixed)	496.127 ± 6.070 μm
$\sigma_{c\tau}$=26 μm (fixed)	496.125 ± 6.070 μm
$\sigma_{c\tau}$=32 μm (fixed)	496.118 ± 6.070 μm
$\sigma_{c\tau}$=100 μm (fixed)	492.585 ± 6.019 μm

We use the whole 80,000 events and then a lower statistics piece of 20,000 events using overall resolutions of 2, 21, 26 32, 100 microns. Given that in data the average resolution is 26 μm for the $B_u^{\pm}$ the last data point is hardly neccesary. It seems that applying of order 25 μm is fine as a choice and the negligible shift in best fit lifetime shows us that given the L_{xy} cut the current CDF resolution is about as good as we would want.

We see that only when the overall resolution is changed to a drastic value of 100 μm is there a shift ≈4 μm micron in the lifetime.

9.2 Understanding the effect: Why it doesn't matter what overall resolution we choose.

To understand why these different choices make very little difference let us begin by writing down the expression for the probability density function in proper decay time:

$$P(t_0) = \frac{\frac{1}{\tau} e^{\frac{-t_0}{\tau}} \mathrm{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum\limits_{\substack{i=\text{all} \\ \text{intervals}}} \left[-e^{\frac{-t}{\tau}} \mathrm{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + e^{-\frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\text{min i}}}^{t=t_{\text{max i}}}} \tag{50}$$

recall that we apply an L_{xy} cuts of 350 μm for our analyses, and the average of

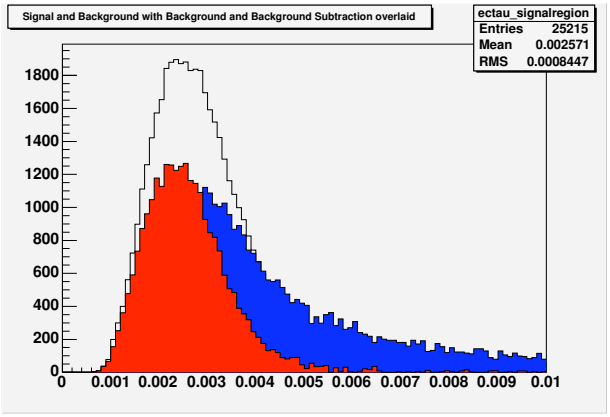


Figure 18: $B_u^{\pm}$ $\sigma_{c\tau}$ distribution red is signal, blue is background, unfilled is total.

the $c\tau$ distribution is 600 μm, for large arguments the Ffunction tends to 1, and the behaviour of the probability density functions is determined by the exponential terms. We can see that we expect very small changes in the likelihood for $\mid \frac{t}{\tau} \mid >> \frac{\sigma^2}{t^2}$ which we expect is always the case.

To summarize if the lifetime resolution is good, then it is nearly the same as having perfect resolution, (we can see that making a drastic change to the overall resoultion from 21 μm to 2mm hardly changes the best fit lifetime) this simply that the usual resolution is already very close to the limiting case of a perfect resolution. We see that the answer begins to shift at the 1 % level only when an unrealistic resolution of 100 μm is applied to all events.

9.3 Determining a reasonable average $\sigma_{c\tau}$ for the $B_u^{\pm}$ sample

Finally we make a choice of a specific overall resolution to apply to the mode. We take our final data sample after all cuts and look at the background subtracted sample for the $\sigma c\tau$ distribution. The overall and background subtracted subtracted $\sigma c\tau$ distributions are shown below. The overall resolutions derived are 25.7 μm for the B. Recall that we hardly expect variations in these particular choices to make a difference.

Finally we have made the choice of 26 μm for the B_u^{pm} when we do analyse data.

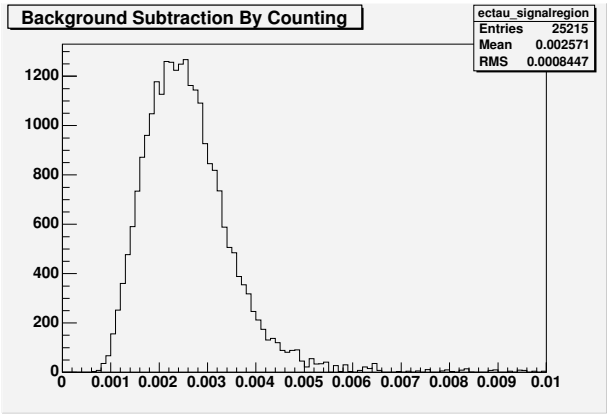


Figure 19: $B_u^{\pm}$ $\sigma_{c\tau}$ distribution of signal only: Average 25.7 μm .

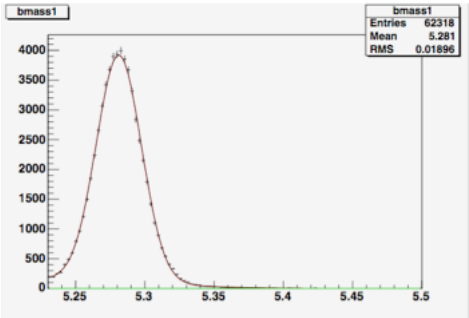


Figure 20: We find that realistic MC is fit well by 2 Gaussians.

10 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.

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 Bu 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 $B \to D\pi$ and $B \to D\pi(n\gamma)$ 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. 20

In our data the contribution of events that radiated photons may be different 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 $'B' \to D - \pi-$. 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. The wrong sign distribution is shown in Figure 21.

Putting this all together

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

The functions Gauss and the polynomial have been normalized over the re-

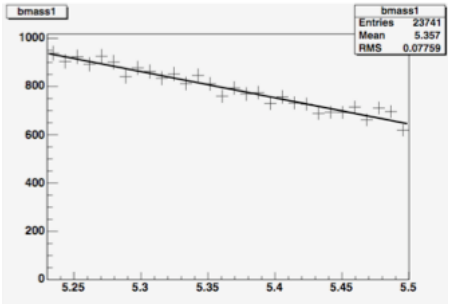


Figure 21: The wrong sign distributions looks like a first order polynomial.

stricted 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.

11 Modelling the Background Lifetime

We are not interested in the physical meaning of the background lifetime distribution and we postulate a general PDF $y(t)$. This is the distribution of lifetimes in the background before the trigger. This includes all detector resolution effects so our function does not depend on the measured uncertainty on lifetime, σ_t. The probability of measuring a lifetime, t, given that an event is background and given the acceptance function, calculated for that event, is

$$P(t|b, A_{\text{trig}}) = \begin{cases} \frac{y(t)}{\int Acceptance \mathrm{d}t} & \text{for } t_{\min} \leq t \leq t_{\max} \\ 0 & \text{for all other } t. \end{cases} \tag{52}$$

$y(t)$ has been normalised such that t lies within the acceptance function. Note that this function has no physical meaning and we don't require one for background.

We parameterise the background by fitting the height, $y(t)$, at different lifetimes, t_j, and interpolating between these points using exponential functions. So

$$y(t) = e^{a_j + \left(\frac{a_{j+1}-a_j}{t_{j+1}-t_j}\right)(t-t_j)} \qquad \text{for } t_j \leq t \leq t_{j+1} \tag{53}$$

where the a_j are constants to be found. We are free to choose the number, n, and spacing of the points t_j. We would like to use as few parameters as are needed to describe the distribution so we space the points most tightly at low lifetimes where the distribution of lifetime varies most rapidly and have fewer

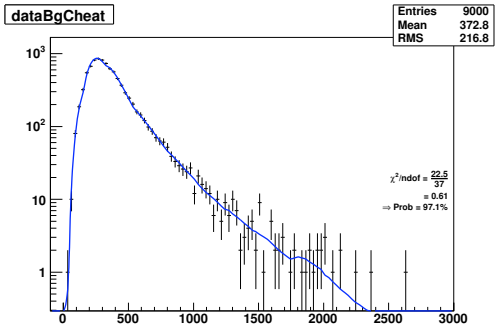


Figure 22: A fit to uppersideband of using the interpolated model for lifetime

points at large lifetimes where the distribution varies less. We distribute t_j logarithmically according to

$$t_j = t_{\min} + \left(-c + e^{\left(ln(c) + \frac{ln(c+b)-ln(c)}{n} \cdot j\right)}\right) \cdot \frac{(t_{\max} - t_{\min})}{b}, \tag{54}$$

where c and b are constants which affect the scale of the logarithmic spacing. For a given c (we choose $c = 1$) decreasing b spaces t_j more equally. We chose $c = 1$ and $b = 4$ as this gives a good fit to the background in the upper sideband. We use 10 parameters in the fit. Using more does little to affect the goodness of fit but does alter the fit stability. Using these parameters the t_j are found at 0, 146.9, 322.6, 532.7, 783.9, 1084.3, 1443.5, 1873.1, 2386.7, 3001. The parameterisation is tested using the background in the uppersideband and is shown in figure 22.

12 Validation of Method using toy Monte Carlo

This section details the studies that have been done and their methods to validate the Monte-Carlo Free technique. In an ideal scenario we would like to have generated enough realistic Monte Carlo for signal and background to be able to run 1000 independent pseudoexperiments and check for any residual biases. However this involves generating 24M events that pass all analysis and trigger cuts for each signal mode and 9M background events using pythia as the background is assumed to be combinatoric. There is not the time nor resource to be able to do this.

We have at our disposal approximately 65K realistic MC events for B+ and 20K background events from the upper sideband. One way to boost the sample size would be to bootstrap the events whereby for each psuedo experiment 24K sg and 9K bkg are chosen from the parent samples with the possibility of choosing the same event more than once available. However while this method of constructing toy experiments is useful in assessing the statistical error calculation it cannot be used to determine any residual bias of the method or fitter.

Imagine that the 65K sample has a truth input of 496 microns and a fitted

lifetime of 498+/-2 microns. This one sample on its own is deviant by 1 sigma which would happen relatively often. If we then use this sample to bootstrap from we will find that the mean fit result is also close to 498 simply because the parent sample is too. However this does not imply that the mean bias of the method is 2 microns. It is unclear how to interpret results using a bootstrap method. To use the background sample this way we also have to reassign the mass and in the process of doing this run the chance of breaking correlations that aren't taken into account.

Another method is to take the realistic MC and the upper sideband shifted down and perform the fit that way. While this method doesn't have the problems associated with bootstrapping we only have enough for 2 samples and it is not possible to make any conculsions from this on the performance of the method.

We have turned therefore to toyMC as our main validation tool. The toy is described below.

12.1 The Toy

The toy used is not simply a toy that generates the fit quantities from the pdf. Infact for this method it is not possible to create such a simple toy, the reason being that the lifetime probability pdf changes from one event to the next based on the event's acceptance function. We do not have a way of generating stand alone acceptance functions. Instead this toy generates the whole decay chain. A toy for this method would require knowlegde not only of the fit parameters but also all the track momenta, track impact parameters, a modelling of the trigger etc. In an attempt to do this we must stress that we cannot expect to create a toy that matches all features of the data, as this would, in effect, have to be as detailed as B Generator or Pythia. We are merely trying to achieve a close approximation to the data so that for example the mean momentum in toy and data are similar or that the acceptance functions in the toy roughly match those found in data.

12.1.1 Generating Singal Events

Let us take the B^+ as an example. Four quantities are independently generated. Described below is the "standard toy for signal". Studies using variants of this are detailed later in this section.

- The B Lifetime - This is generated from an (unbiased) exponential, using the current PDG lifetime as mean $\tau = 491.1\mu m$, smeared by a gaussian of width 26 microns
- The D Liftime - This is also generated from an exponential with mean lifetime taken from the PDG
- The B Mass - for the basic toy this is a simple Gaussian as we are more interested in the ability of the method to remove lifetime biases. The mean

mass is the PDG mass and we use a width of 18MeV. A more complicated mass model is added once we are satisfied that the method can unbias the effect of the trigger.

- The B transverse momentum. We use histograms from realistic B Generator Monte Carlo as a basis for the B_{PT}. However as the B Generator MC has already been triggered once, using that distribution for our toy MC will result in a bias towards accepting events with a higher momentum. We find that using the B Generator spectrum x 0.72 gives a final momentum spectrum closest to what we find in data.

With the 4 quantities above we have sufficient information to generate the entire decay. Firstly a momentum direction is chosen. We choose a direction that is uniform in ϕ and flat in η. The momentum vector for our particle is $(cos\phi * sin\theta, sin\phi * sin\theta, cos\theta) * B_P$. We calculate the Lxy and B vertex of the particle in the lab frame. In the rest frame of the B we create a decay resulting in a D^0 and π^+. In the rest frame these will have equal and opposite momenta and the decay will be isotropic. We transform these quantities to the lab frame to calulate all the track parameters requrired. The decay of the $D^0 \rightarrow K + \pi$ is done in exactly the same way. We now have all the kinematic quantities required. The SVT single track efficiency for signal over 1fb^{-1} is approximately 0.75. So at random we give 75% of the tracks an "SVT Match". Only these tracks can be used by the trigger. In addition the tracks that are given an SVT Match also have their SVT impact parameter discretised to the nearest $10\mu m$ as would have been done had this been a real event.We have also now modelled the SVT single track effieciency which would be crucial in any toy validating the method for more than a two body decay. Finally the event, with all the kinematic information is passed through the trigger simulation and all analysis cuts are also applied to the event at this stage. There are three trigger paths (Low-Pt, Scenario-A (Medium-Pt) and Scenario-C(High-Pt)). We ensure that the final sample has a similar trigger mix to that found in data as this is important to make the acceptance function distributions similar. For events that pass all cuts (trigger and analysis) we can now construct the acceptance function in the same way that we would for real data. Now that we have the acceptance fucntion, mass and b liftime we have all the quantities to required for a fit.

12.1.2 Generating Background Events

Background is generated primarily in the same way. The mass spectrum for the basic toy is flat. The B lifetime comes from the lifetime fit to the uppersideband. The resulting function is the distribution of lifetimes before the trigger and so we use this to generate the distribution of lifetimes for background. For the D lifetime we simply draw from the histogram of the same quantity found in the uppersideband. Momentum distributions were based on exponentials that were scaled until they resulted in a distribution similar to that found in data. Recall once again that the aim was only to find a close approximation and not an exact match. The single track SVT effieciency is lower for background, partly

due to an increased fraction of tracks with only 3 hits in teh silicon. Therefore in background generation the SVT effieciency is set to 0.65.

Kinematically one key difference between signal and background is the momentum distribution of the pion from the B. In signal this tends to be of high momentum, whereas in the sideband this is generally of low momentum. To make the background toy kinematics better reflect the uppersideband we reject a large fraction of events in the background that have a higher momentum. Once this had been done we find that the agreement between the toy signal MC and the sideband subtracted signal region and the toy background and the sideband was good enough and have choosen to establish this as the standard toy.

12.2 Agreement between Standard Toy and Data

The most important plots to compare are those of the acceptance function. If we can get the toy to match the data for this variable it will give a Punzi bias of approximately the right size and direction. Moreover if there is agreement between the acceptance function distributions it implies that all the kinematics are in broad agreement as the acceptance function is calculated using a combination of Impact parameters, momenta, opening angles etc. For B+ the acceptance function is split into three parts as described elsewhere. These three parts are where there is one track pair available for the trigger, two track pairs and three track pairs. For the data and the toy we plot the average acceptance function in each of these three categories. The comparison for transverse momentum is also shown.

The agreement between toy signal and the sideband subtracted signal region is very good considering that the detail of the simulation is minimal in comparison to the cdfSim. The agreement between toy background and the sideband is less good but still sufficient and broadly matching.

12.3 Validation of method

We start with the validation of the signal only pdf. Our data sample contains approx 24K signal events so we generate 1000 sets of 24K events and perform a lifetime fit upon them. The pull distribution (Fit lifetime - lifetime error)/error on fit is a unit gaussian. This demonstrates the validity of the signal only pdf, in particular the method of dealing with tracks that are initially outside the trigger impact parameter cuts. This is shown in Figure 25.

Secondly we consider the addition of background. We add to the signal events 9K background events and perform the fit in two configurations. Firstly with the full pdf and secondly omitting the part that deals with the Punzi effect due to acceptance function to show that this toy really does contain such a bias. We find that there is such a bias. In figure 26 the pull is biased and this corresponds to a shift of -4.7$\pm$0.24 microns. When we take into account the Punzi effect with the full likelihood we find that we can correct for this bias and

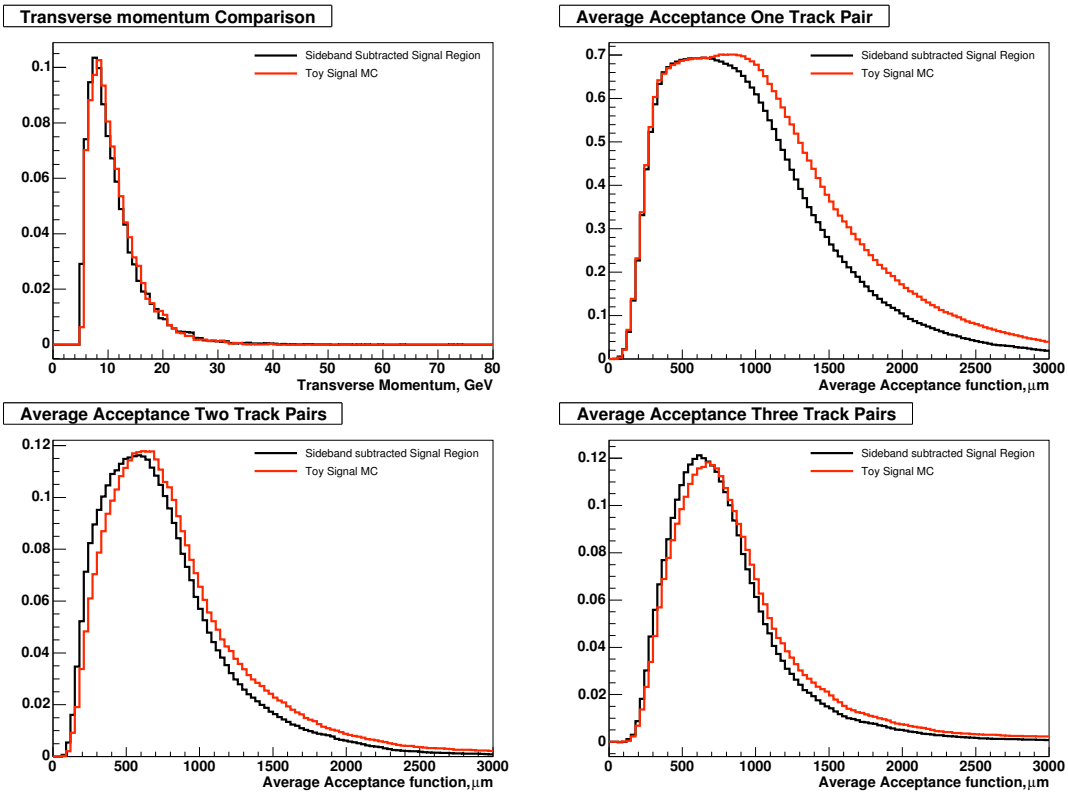


Figure 23: Comparison plots between the sideband subtracted signal region in the B plus data and the toy signal generator. From the similarities of the acceptance function it is possible to see that the kinematic distributions of the toy broadly match those found in data.

measure only a very small shift in the lifetime of $0.22 \pm 0.21 \mu m$. The plots in figure 27 demonstrates that this method is capable of working.

This toy represents only the default signal and background generation. To test the robustness of the method we try 4 further scenarios. Firstly we soften and harden the momentum spectrum of the background, and use these varied backgrounds to perform a pull study and check there is no bias. The different momentum spectra are shown in Figure 28 with the resulting pull distributions. We observe no significant bias in the lifetime measurement.

We also vary the input signal lifetime by $\pm 50 \mu m$ and check that the method can work with different lifetimes. The pull distributions and results are shown in 29 and again there is no significant bias observed.

It should be noted that each of these changes in pt or lifetime will change the acceptance functions and this in term will change the fisher scalar distribution. Despite this the fitting methods prove to be robust against such changes.

We conclude therefore that the method does remove the bias induced by the

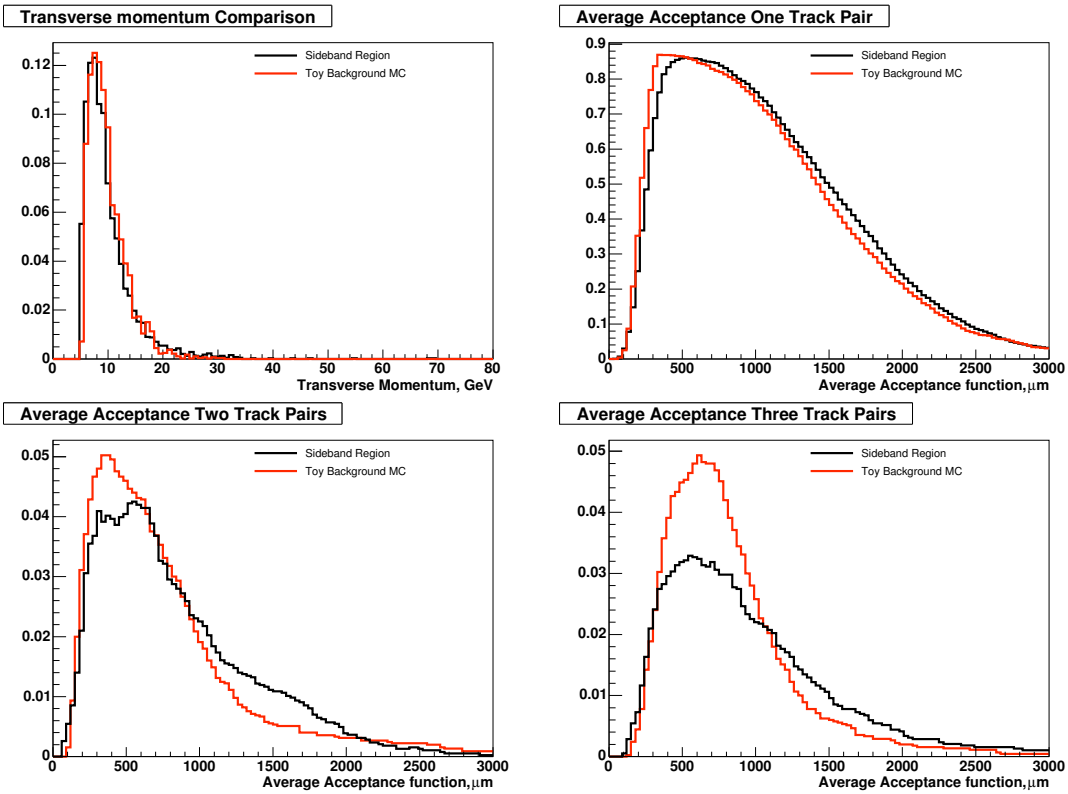


Figure 24: Comparison plots between the sideband region in the B plus data and the toy background generator. From the similarities of the acceptance function it is possible to see that the kinematic distributions of the toy broadly match those found in data.

SVT based trigger.

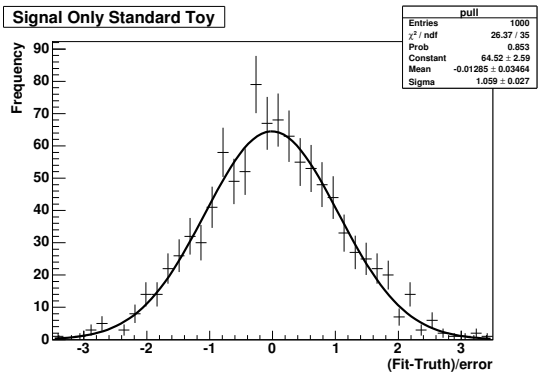


Figure 25: Pull distribution of signal only pseudo experiments.

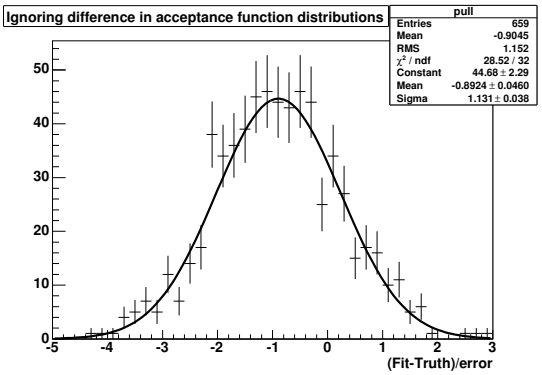


Figure 26: Pull of lifetime fit when ignoring the Punzi effect

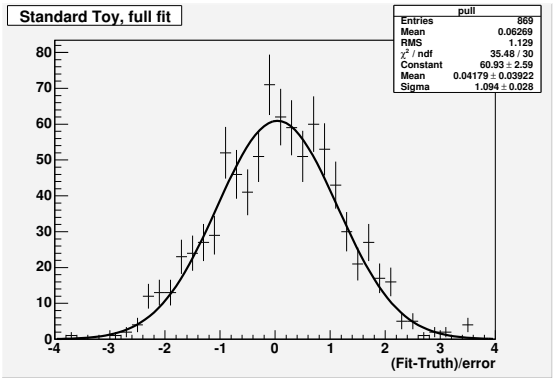


Figure 27: Pull of lifetime fit when accounting for the Punzi effect.

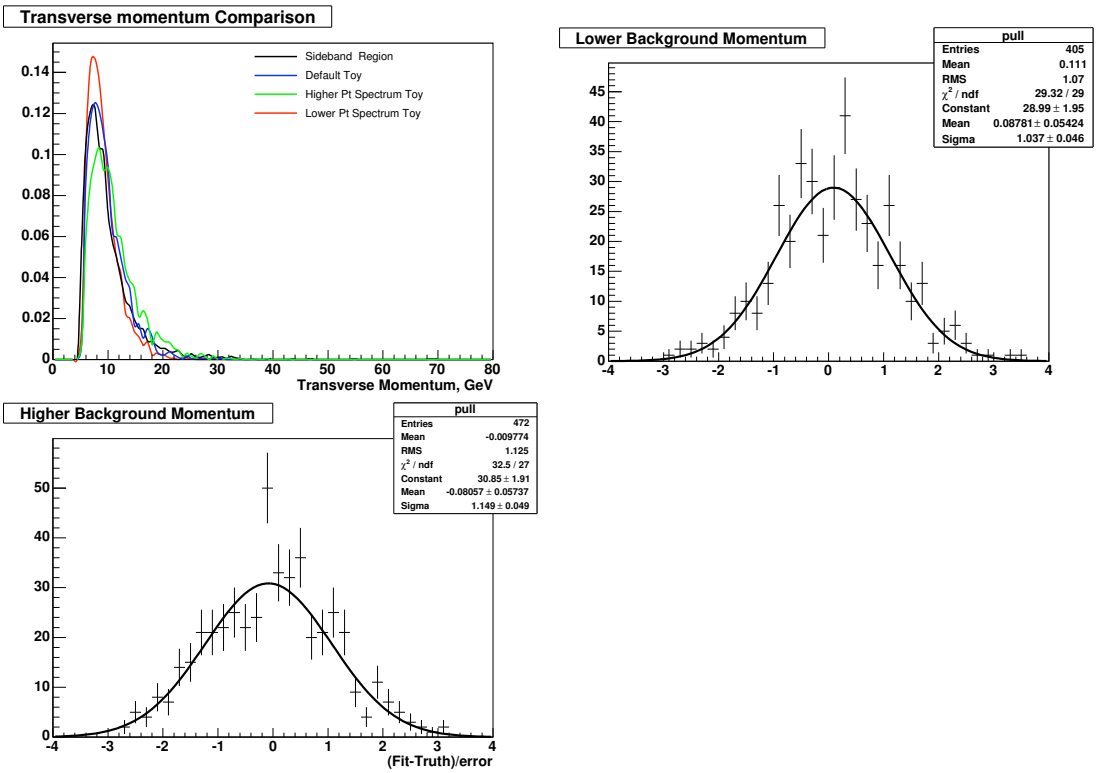


Figure 28: The plots shows the different Pt spectrums used for the background variants. The pull plots are also shown and are unit gaussians

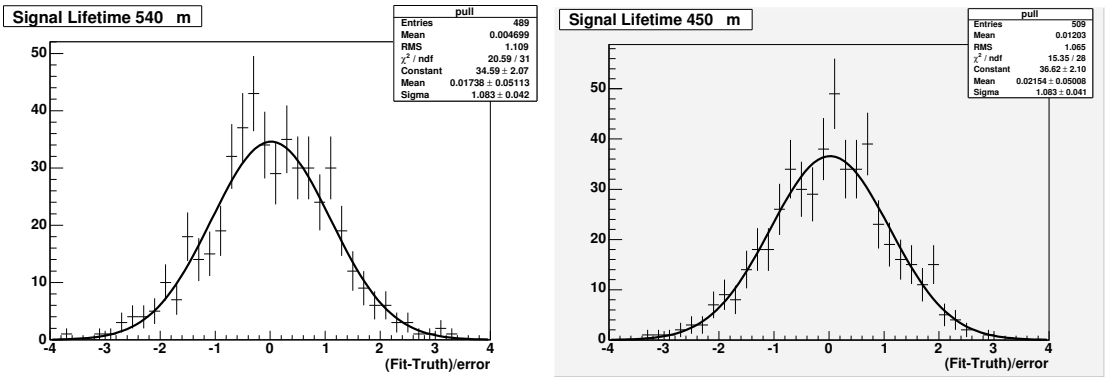


Figure 29: Pull of lifetime using the input lifetimes of $450\mu m$ and $540\mu m$. The pulls are unit Gaussian.

13 Analysis cuts for $B^{\pm} \rightarrow D^0\pi^{\pm}$, $D^0 \rightarrow K^{\mp}\pi^{\pm}$

In this section we define the analysis cuts used to select $B^{\pm} \rightarrow D^0\pi^{\pm}$. 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.

13.1 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 datsets 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 Ds and the fully reconstructed Bs. The final reconstructed quantities are obtained from the AC++ wrappered 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.

13.2 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 transverse flight distance of the D in the direction of its $\vec{P_T}$ ($D_{L_{xy}}$)is $> -100\mu$m and is $<$ than 1 cm.
- 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 $\geq$2.4 GeV.
- The scalar sum of $\pi^{\mp}$ and K $^{\pm}$ transverse momenta is $\geq$ 2.4GeV
- The K$^{\pm}$ and $\pi^{\mp}$ P_Ts are each individually $\geq$ 0.4GeV

Next we loop over all tracks in the event with charge opposite to the $\pi^{\mp}$ 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 $> 350\mu$m and $<$ than 1 cm
- The candidate B vertex $\chi^2 < 15$
- The P_T of the $\pi^{\pm}$ from the $B^{\pm}$ is $\geq$ 1 GeV.
- The impact parameter of the B with respect to the beam spot is $< 80\mu$m.
- The angular separation in ϕ between the B and its π daughter is $\leq$ 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 $z0$ within 5cm of each other.
- The $B^{\pm}$ transverse momentum (P_T) $\geq$5.5
- The scalar sum of all B daughter charged tracks P_Ts is $\geq$ 5.0
- The calculated uncertainty of the proper decay time ($\times c$) of the B, $\sigma_{c\tau}$ is less than 100 μ m

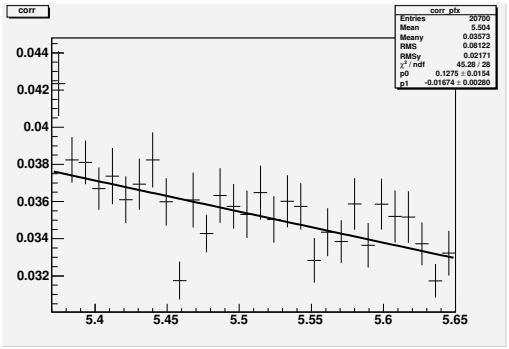


Figure 30: A profile plot of mass vs lifetime in the sideband. As the mass is increased the average lifetime falls.

14 Systematic Studies

14.1 Systematic error due to Mass Lifetime Correlation

One assumption that is present in the pdf is that there is no correlation between mass and lifetime. While this is true for signal it is not true for background. We can see this by looking at the profile plot in Figure 30 of the upper sideband for these two quantites over a mass range of 270MeV which is the range in our datafit. To assign a systematic for ignoring this correlation we introduce the correlation into they toy by rejecting events such that the final correlation matches that found in data. We then perform a pull study using this toy. We find that there is a shift on the mean of the pull of 0.42 $\pm$0.05. This corresponds to a lifetime shift of 2.2 microns and we assign that as a systematic to ignoring the correlation in the pdf.

This figure turns out to be one of the leading systematic errors of this analysis and further improvements in the future would include accounting for a correlation in the pdf.

14.2 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 1000μm 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 region sample from $B_u^{\pm} \rightarrow D^0\pi^{\pm}$ is used and a description of the evaluation of the systematic error now follows.

14.2.1 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 for this study 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 $P_T \geq 2$ GeV

iii . and track $\mid \eta \mid < 1.1$

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

$$\epsilon_{SVT}(d_0) = \frac{N^{SVT}(Ax-hits \geq 3 : P_T \geq 2GeV, \mid \eta \mid < 1.1)}{N^{OFF}(Ax-hits \geq 3 : P_T \geq 2GeV, \mid \eta \mid < 1.1)} \tag{55}$$

We fit this expression to the function :

$$\epsilon_{SVT} = p_0 \times erfc(\frac{d_0 - p_1}{p_2}) \tag{56}$$

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.From figure 31 we see that the efficiency is almost flat but starts to drop just before 1000 microns.

14.2.2 Determination of the systematic error

To ascertain the error the assumption of flatness causes we turn to our toy monte carlo. We generate signal events using an efficiency determined by the fit of the distribution above, but fit with the standard one floating efficiency. We test this on 1000 samples of 24K events and find that the mean shift in the fitted lifetime is -1.9 $\pm$0.2 microns and assign this as a systematic.

14.3 An approach to evaluating a systematic error due to Silicon Misalignment

In the appendix we have 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 results in a large statistical error on the measurement itself. We do not wish to accept this large (5.0 micron) statistical error as a systematic all indications suggest that this would be a large overestimate.

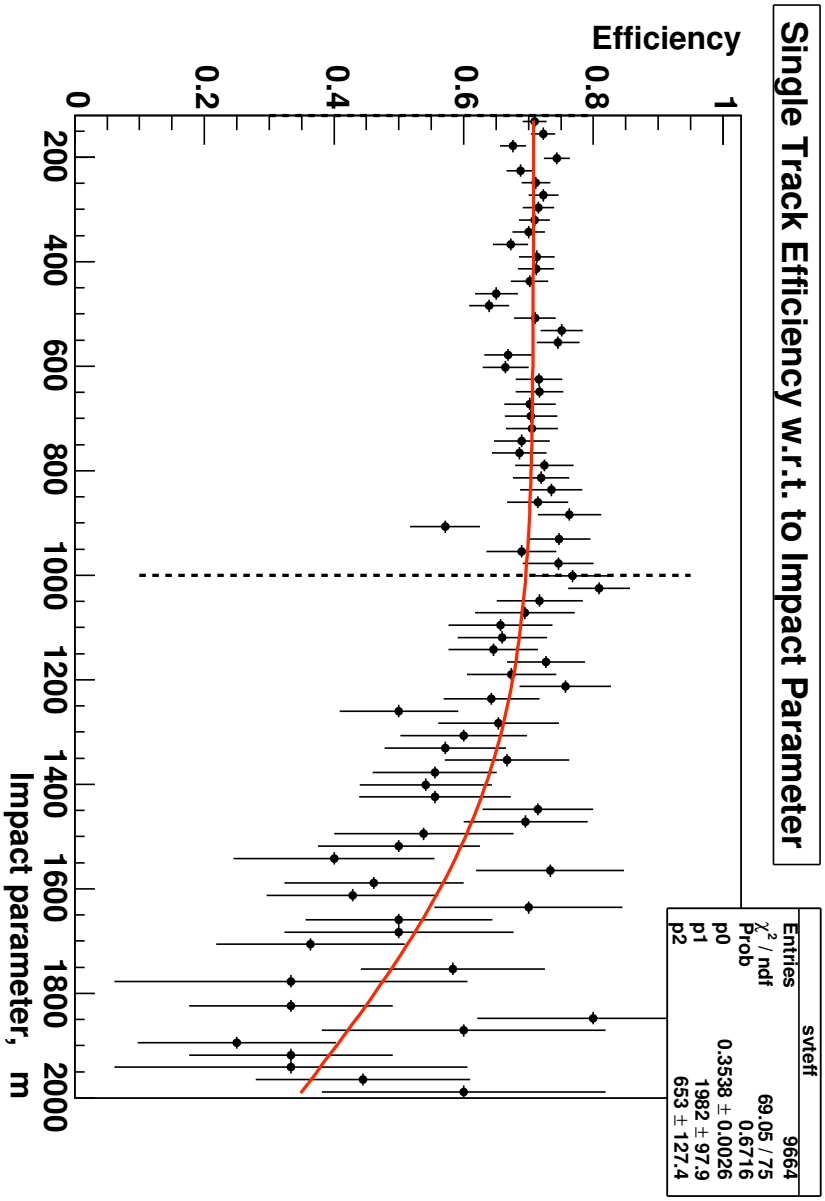


Figure 31: Single track efficiency of the SVT.

Instead we turn to an estimate made by toy MC. By using geometry we calculate the effect of radially moving all Silicon wafers. The new impact parameter is give by:

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

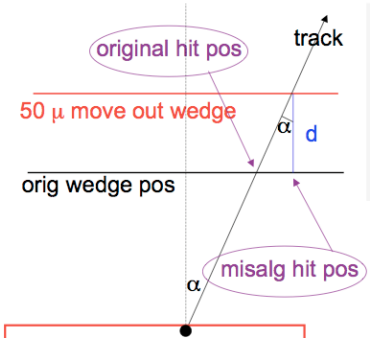


Figure 32: Diagram to show how hit are assumed to be at the wrong place if the position of the silicon layer is not at the official position.

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. The effect of this shift is shown in Figure 32 where we can see that the change in position of silicon layer causes the hit to be assumed to be at a different place from the true point.

We turn to CDF note 6551 that did a study using unbiased decays and 4 different alignment scenarios. They found that a 50 micron move out of silicon layers caused the largest shift in lifetime and that overall they quoted a conservative error of $\pm 1\mu m$ as a systematic due to alignment.

We use our toy generator that has been configured to shift impact parameters and decay vertices assuming the alternate alignment, in two cases. Firstly we generate "unbiased" decays by removing all trigger and analysis lifetime based cuts. This will provide a crosscheck to the study using realistic MC and the Jpsi trigger. Secondly we apply the effect of the silicon shift to our decays using that have lifetime based trigger cuts applied to them. We generate 1000 samples of 24K events and look at the pull of the fitted lifetime.

14.3.1 Results

We find that for the unbiased events we find a shift of 0.97 ± 0.1 . This is consistent with the findings of the realistic MC study using the Jpsi trigger. For events passing the two track trigger we find a smaller shift of 0.46 ± 0.2 which is smaller. We assign a systematic of 0.5 microns due to silicon misalignment in this method.

14.3.2 Interpretation

There is one partial explanation for this reduction in systematic error. If we condsider our simple likelihood without measurement errors and assume that the events have a single top hat bounded by tmin and tmax as the acceptance function we have:

$$P = \frac{\frac{1}{\tau} e^{\frac{-t}{\tau}}}{e^{\frac{-t_{\min}}{\tau}} - e^{\frac{-t_{\max}}{\tau}}} \tag{57}$$

A change in alignment alters t, tmin and tmax. If the change in each is the same then there is no change to the probability. If all events were to behave like this we would expect an alignment systematic of 0. To get an estimate for how often this happens we look at a sample of events in toy triggered by the tracks from the D. This gives a simple acceptance function. We then look at those that would pass both alignment scenarios and look for events that have changes in t, tmin and tmax within 5 microns of each other. We find that 10% of the events in this sample fall into this category. Other combinations of changes in tmin. tmax and time would also serve to reduce the change in the likelihood in comparison to a sample passing a trigger with no lifetime cuts (and hence only changes in t). While there are also combinations that would serve to alter likelihood more heavily, given the results of the two studies and a demonstration that the toy study does observe effects similar to those observed in realistic MC we conclude that these happen less often.

14.4 Fitter Bias

To assess any bias of the fitter itself we generate the toy in the same way as described in the validation section. The one difference is to add the more complicated mass model of 2 gaussians for signal and a first order polynomial for background. Again we generate 1000 samples of events and find a small shift in the mean of the pull of 0.04 ± 0.04. This corresponds to a shift in lifetime of 0.2 microns, we use this as the overall bias of the fitter.

As the mass parameters are fixed this leads to a underestimation of the statistical error. We use the width of the pull distribution to assign a scale factor by which the statistical error of the data fit must be multiplied. The gaussian fitted to the pull distibution has a width of 1.10 ± 0.03 and so we assign this as a scale factor to the data fit.

14.5 Background parameterisation systematic

The form of the background lifetime parameterisation is fairly arbitrary and it is useful to test how robust the parameterisation is to fitting different distributions that have broadly the same features. In the toy studies presented so far including background the lifetime was generated using an exponential interpolating function similar to the used to fit the data.

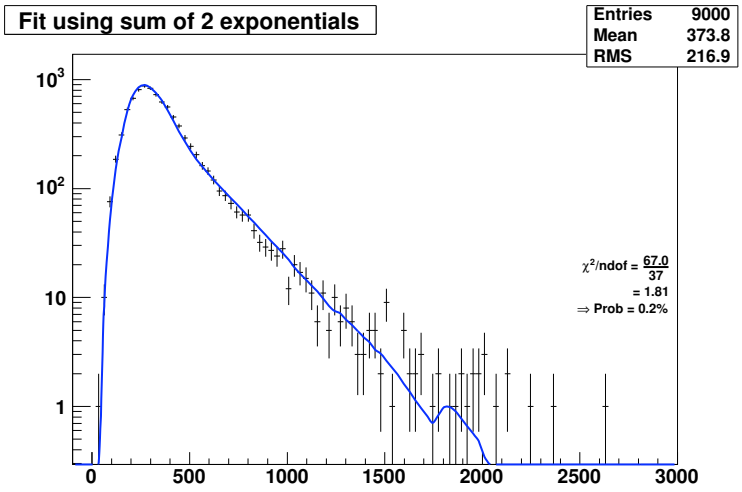


Figure 33: Fit to background using a sum of two exponentials.

We can also fit the upper sideband to a sum of 2 exponentials although the fit is of a poor quality it does at least broadly mimic the shape of the data. This is shown in figure 33. We use the parameters obtained in this fit to generate a different toy background and repeat the toy study measuring the signal lifetime in 1000 samples using our interpolating model for background. We find the the mean of the pull is shifted from zero and corresponds to a shift of 0.8 ± 0.2 microns and assign this as a systematic due to the background parameterisation.

14.6 Systematic Error due to Resolution Function

In this analysis we have assumed that our lifetime errors are gaussian and to model the effect of these we have convoluted our exponential decay function with a single gaussian model of constant width as described in section 9, this single gaussian is then our resolution function. In the same section we have also demonstrated that the choice of using a single average $\sigma_{c\tau}$ in place of the event-by-event quantity is typically less than a tenth of a micrometer.

Several lifetime and related analysis at CDF have found that the resolution function is in fact composed of 3 gaussians. These analyses use an event by event $\sigma_{c\tau}$ and so the derived resolution function is has essentially three scale factors for the event by event $\sigma_{c\tau}$ of the three different categories of event. A fit is made of the distribution of the differences in lifetime of the prompt events divided by the individual event by event $\sigma_{c\tau}$ to 2 or three gaussians. If we were able to fit this distribution to a single gaussian of unit width and center at zero, this would mean that a single gaussian models our resolution function very well and also that our errors at CDF are correctly calculated. The need for more than one scale factor in the fits we describe tells us that neither of these assumptions is true can be seen in CDF note 8524.

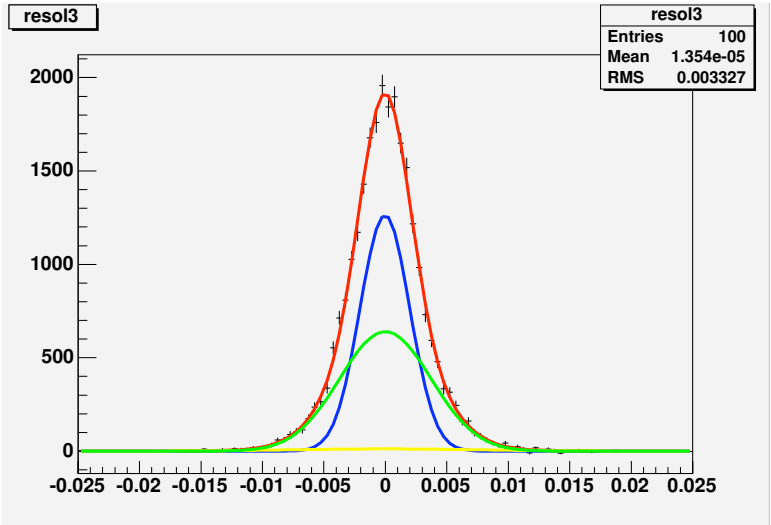


Figure 34: Resolution function for the $B^{\pm}$ decay mode.

In order to determine a resolution function for our data we use the scale factors for the different gaussian components of the resolution functions for the $B^{\pm} \to D^0\pi^{pm}$, in table 1 page 14 CDF note 7500. We take each event by event sigma in our data sample using each $\sigma_{c\tau}$ calculate the resolution by: $\sigma_{res} = \sum(f_1.G(0, s_1.\sigma_{c\tau}) + (1 - f1).G(0, s_2.\sigma_{c\tau})$, where f_i and s_i are the fractions and scale factors for each resolution function component and G represents a Gaussian, and the first argument is its mean (zero). We then make a background subtracted plot of σ_{res} and fit it to three Gaussians. Note that since we do not use event by event errors ($\sigma_{c\tau}$), we cannot use the scale factors, but we can use these to weight each error and then smear them according to how we believe the scale factors should effect each error on average. We then obtain resolution functions composed of the sum of gaussians of different resolutions and different fractions. We summarize this for $B^{\pm}$ in the table below. Note the resolution functions is assumed to be:
$f(x) = f_1.G(x, \sigma_1) + (1 - f_1).(f_2.G(x, \sigma_2) + (1 - f_2).G(x, \sigma_3))$ and if $g(t)$ defines the pure signal lifetime distribution then the convolution to be performed to obtain a lifetime probability density function would be: $\int g(t).f(x - t)dt$

Mode	σ_1	σ_2	σ_3	f_1	f_2
$B^{\pm} \to D^0\pi^{\pm}$	20. μm	37.2 μm	74.8μm	0.51	0.96

and below we show the fits to the σ_{res} distributions for the $B^{\pm}$ mode.

The next step is to determine the systematic uncertainty due to using a single overall width when in fact the resolution function is composed of 3 gaussians and 3 widths. We generate toy monte-carlo, smear the lifetimes according to the generated resolution functions and then fit first with an overall $\sigma_{c\tau}$ fixed at 26

μm and then using the function appropriate for the mode in the convolution. We quote the difference between the two measurements as a systematic uncertainty due to assigning a single width instead of the whole resolution function. The results are summarized in the table below.

Mode	Input lifetime	Fixed σ	Full Resolution Function	Systematic Uncertainty
$B^{\pm} \to D^0\pi^{\pm}$	491.4 μm	487.25 $\pm$ 3.71 μm	486.98 $\pm$ 3.71	0.27 μm

14.7 Inclusion of the Cabibbo supressed mode

The signal region contains events that are actually mis-reconstructed $B \to DK$ events. These events have the same true lifetime as our signal events but as the mass assignment for one track is incorrect the mass and P_T are shifted and this causes a change on the determined $c\tau$. The first order is to estimate what proportion of the signal classed events are in fact the cabibbo supressed decays. We use the toy generater with an input mean mass of 5.276 and width of 20MeV to generate $B \to DK$ events and then "reconstruct" the events using a pion for a kaon. Then we apply the lower lifetime cut of 5.23 and find that only 33% of events that are of the supressed mode will pass the mass cut. Then we use the branching fraction of $B \to DK$ and $B \to D\pi$ from the PDG to determine that only 3% of events in the sample are expected to be cabibbo supressed decay mode.

To estimate the systematic we generate 1000 samples of 24K events where 3% of them were generated as $B \to DK$ and fit the lifetime and observe the pull on the mean lifetime. This is found to be -0.0173 ± 0.03 and therefore we assign a negligible systematic due to including the cabibbo supressed mode. This is unsurprising as they constiute a small percentage of the sample.

14.8 Some CrossChecks

In this section we outline some crosschecks on the choices of parameterisation and on some of the systematics. We start with establishing whether the fitter is biased in anyway to samples from differing luminosity periods etc. We measure the lifetime of the 0d data alone and that of the 0h+0i dataset. We find that the 0d dataset gives a lifetime of 488.3 $\pm$11.9 and the combined 0h and 0i has a lifetime of 502.2$\pm$8.8 . These two lifetimes are not statistically different. We also check that the fit results are consistent if the data is split into two randomly. For sample 1 we get 498.9$\pm$10. and for sample 2 we get 498.8$\pm$9.6 As there will be variations in fisher direction between sample 1,2 and the total and all the lifetimes are consistent this gives confidence that the actual direction of fisher scalar doesn't matter.

We also consider the parameterisation of the fisher variable. The default choice is to use an effective 13th order interpolating polynomial. Using one toy sample we reduce the number of polynomial to see if there is any effect. We remove 4 orders. The plots of these fits are shown in Figure 35. Both are giving a reasonable fit. The difference in lifetime between these fits is negligble (¿0.1

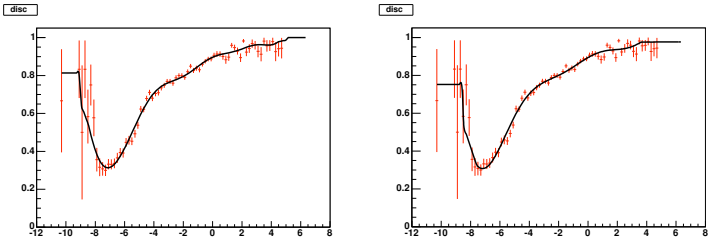


Figure 35: The Left plot shows a fit using a 13th order polynomial and the right plot shows a fit using a 9th order polynomial. There was no change in lifetime between the two.

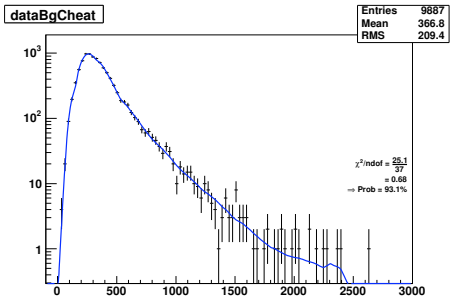


Figure 36: A fit to the upper sideband over the range 5.5 - 5.8 GeV.

microns). We cross check this behaviour in the data sample and find again a neglible change in the lifetime. We also observe an increase in the statistical error by 0.1 microns. With this study we conclude that we could have used a lower order polynomial, although it makes little difference to the final result.

We also try to crosscheck some of the systematics on data and show that the shift in measured lifetime is consistent with the prediction from toy.

For the Background parameterisation there is some choice as the the spacing and the number of fit points. We change the number of fit points by ± 1 and change the spacing variable to try for more bunched or more equal spacing. Using these alternatives no fit shows a difference from the default by more than 0.4 microns. This shows that the default choices are in an optimum region. We know that choosing something very different from the defualt choice leads to a poorer quality of fit of the upper sideband only. We do not use the 2 exponential model to fit the data as it is known to have a very poor fit to the upper sideband. This study has not shown a need to re-evaluate our systematic of 0.8 microns.

We also consider how well the default parmeterisation continues to fit the sideband when we consider upper sideband in a higher mass window. We take events that lie between 5.5 and 5.8GeV as this provides statistics similar to the data sample and find that the fit continues to fit well. The fit is shown in Figure 36

We also test the resolution function. We apply the alternate resolution function

to data and see that the lifetime shifts by -0.2 microns. This is consistent with the quoted systematic of $\pm 0.3 \mu m$.

14.9 Summary of Systematic Errors

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

Source	Assigned Error μm
Misalignment of Silicon	0.5
SVT single track Efficiency	1.9
Correlations in Background	2.2
Background Parameterisation	0.8
Resolution	0.3
General Fitter Bias	0.2
Suppressed Mode inclusion	negligible

We add these uncertainties for the charged B lifetime in quadrature and obtain $\pm$ 3.1 μm.

15 Results from the Data fit

The fit is performed in two step in the same way as the toy studies. First a mass fit that has 7 parameters. Using this we can do the necessary sideband subtraction for the fisher discriminant and calculate this variable for every event. Then we perform the lifetime fit. There are a total of 30 floating parameters: 10 for background lifetime, 13 for P(S—Acc) 3 efficiencies for signal and 3 efficiencies for background and finally the mean signal lifetime itself.

With the pdf that has now been tested on Toy Monte Carlo we turn our attention to data and perform the lifetime fit. The mass and lifetime projections are shown and demonstrate that the model is a good fit to the data. The projections are shown in Figure 39.

We find that the lifetime for B+ in the decay to Dπ is

$$c\tau = 498.2 \pm 6.8(stat) \pm 3.1(syst) \tag{58}$$

In Conclusion we have shown that using a Monte Carlo free method of correcting for lifetime bias is a viable method for lifetime measurement. In a quick comparison to the Monte Carlo based efficiency curve we find that our systematic error is lower and that there is a small increase in the statistical error of less than 20%. A direct comparison is difficult due to differing analysis cuts leading to different signal yields. The MC contains information not available to our fitter and this is why the MC-free method returns a higher statistical error.

The real benefit of this method will come when it is used on modes where the kinematics of the decay are not fully present in the CDF Monte Carlo or where

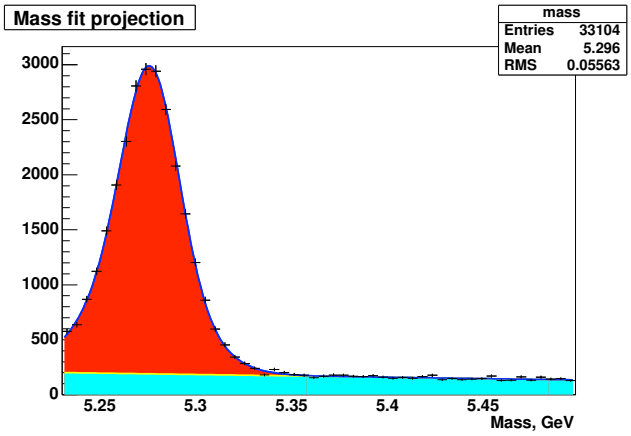


Figure 37: Projection of mass fit on data. The two separate components signal(red) and Background (blue) are also shown

the detailed MC does not reproduce data for example the movement of the beam spot. For example in modes where there is dalitz interference present in the decay of a daughter particle or polarisation issues which are not added in the Monte Carlo. The Monte-Carlo free method corrects for these effects via the acceptance function calculation exactly and for modes that are otherwise dominated by systematic errors of this kind will see a significant reduction in systematic error in comparison to the MC-efficiency curve analysis.

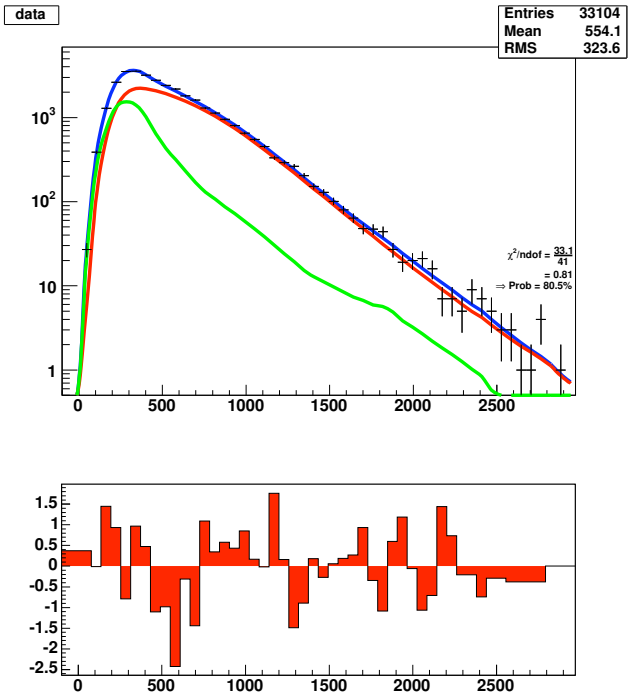


Figure 38: The lifetime projection. Blue represents the total while signal and background are represented by red and green respectively.

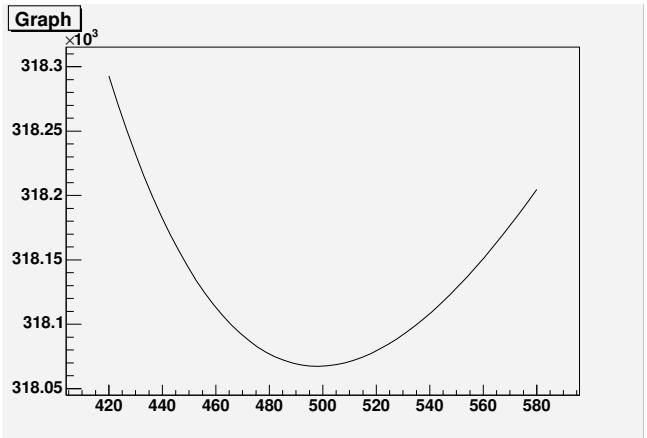


Figure 39: The Likelihood function.

16 Update to Systematic Errors after Pre blessing

At preblessing a few questions were asked on the systematic errors. The answers and further studies are collected here. In addition we have moved from using the mean of the fitted pull distribution to calculate the systematic error and instead use the mean of the residual distribution. The plots of the residual distributions are found in Appendix 2. This has resulted in a ± 0.01 change in some of the systematic error.

16.1 Systematic errors due to SVT single track efficiency

The assumption in this analysis that has the most effect on the lifetime will be the assumption that the SVT single track finding efficiency is flat. It is particularly important for variables which are indirectly related to lifetime. For expample imagine an event that has a given set of kinematics that would fire the trigger should it decay at any point between $500\mu m$ and $1500\mu m$. The method assumes that the event is equally likely to be triggered at any time provided it decays within these times. However in general the IP of the tracks will be higher if the event decays at $1500\mu m$ than if it decays earlier. If the SVT is less likely to find tracks with IP= 900 microns than it is to find tracks with IP=200 microns then the assumption that the event is equally likely to fire the trigger no longer holds.

We look therefore at the deviation from flat due to impact parameter and track transverse momentum. We also look at the single track efficiency as a function of Eta, although we expect this not to be as large a problem as the eta of a track is not related to the lifetime.

16.1.1 Variation in effieciency as a function of IP

We had already investigate the effect of the non-flat effieciency as a function of impact parameter by fitting a curve to the efficiency, generating toy using that efficiency and then using 1000 pseudoexperiments to assign a systematic. We now look at 1 sigma deviations from the fitted curve and test 2 further curves that are more curved than the one originally used. The curves are the blue and green ones in 40.

For the two curves we generated 1000 samples of toy using the new efficiency parameters. The green curve caused the most deviation of -3.1 microns. To be conservative we use this as systematic error.

16.1.2 Variation in effieciency as a function of Track Pt

We also consider now the variation in efficiency as a function of track transverse momentum. The plot from data is shown in 41, and the efficiency is fit to

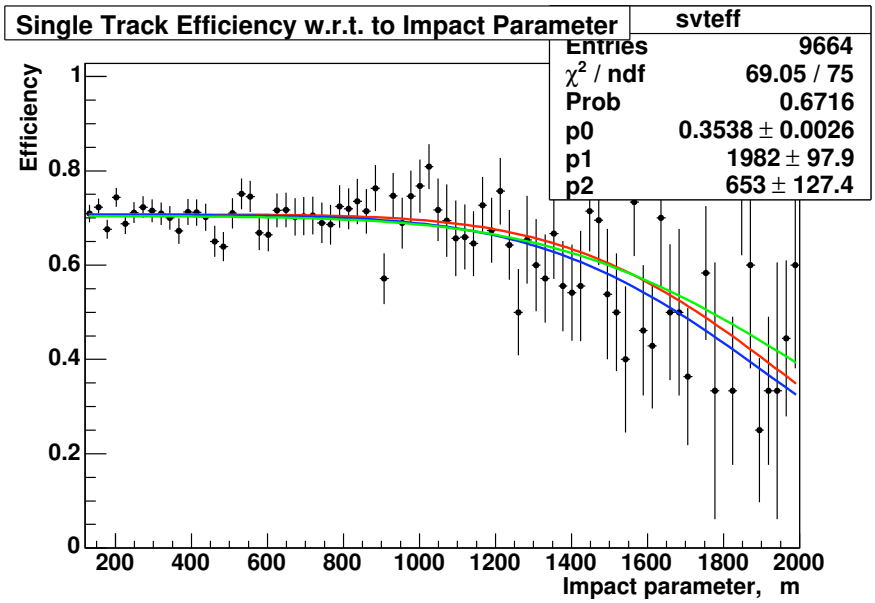


Figure 40: Single track efficiency of the SVT. The red curve shows the orignial fit while the blue and green curves are 1 sigma deviations of the fit that are more curved than the original fit result.

a 3rd order polynomial. We generate toy with this efficiency distribution and fit a mean shift in 1000 samples of 1.8 microns. while the curve for transverse momentum is more deviated from flat than that of impact parameter there is less direct correlation between track pt and lifetime than there is for impact parameter and lifetime.

16.1.3 Variation in effieciency as a function of Track Eta

We also consider variation of the efficiency with track eta. We see that the effieciency is flat out to approximately an eta of 0.8 after which it falls rapidly, becoming close to 0 by an eta of 1.1, 42. We generate toy with this shape and find from 1000 experiments a small deviation of 0.3 microns from the truth input lifetime.

16.1.4 Other Considerations

It is prudent also to consider the variation of the determined systematic error as a function of lifetime. These studies have been carried out using an input lifetime of 491.1 microns. We repeat the SVT flatness as a fucntion of IP using the red curve for an input lifetime of 540 microns. We find an increase in the systmatic error of only 0.1 microns. Given that the B+ lifetime is not as larger

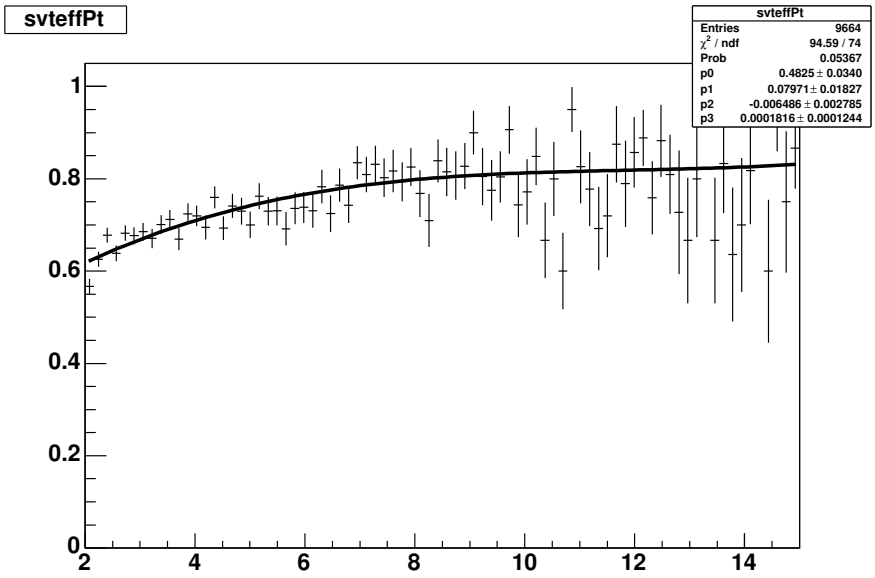


Figure 41: Single track efficiency of the SVT as a function of track transverse momentum

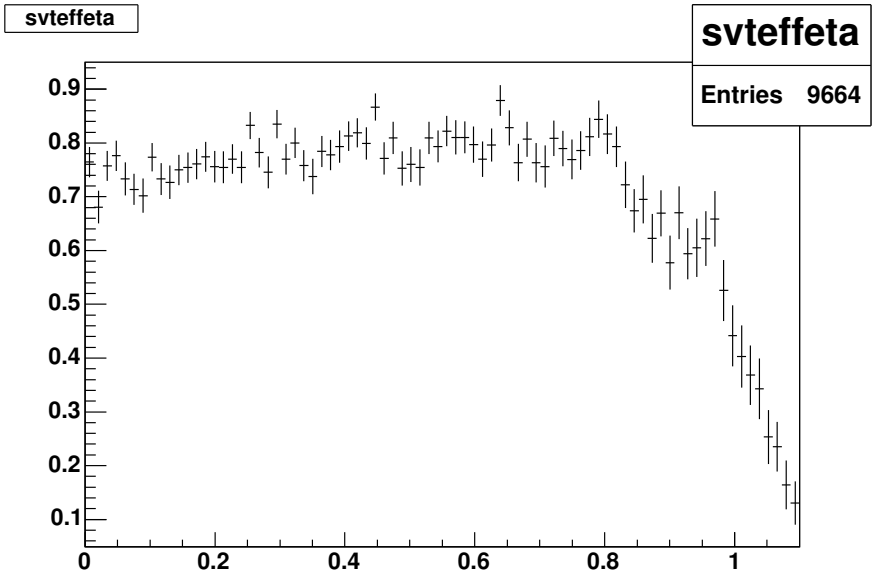


Figure 42: Single track efficiency of the SVT as a function of track eta

as 540 microns we conclude that any reasonable variation in the lifetime of B+ will not increase the above systematics.

16.2 Cross Checking the Alignment Systematic

Previously we had considered a radial movement out of all the silicon layers. This time instead we leave the inner and outer layers fixed and move the middle two layers out by 50 microns. The toy is modified to produce "hits" on all four layers. Due to the misalignment these hits no longer lie on a straight line, and we use least squares technique to fit for a straight line. We caluclate from this the new track impact parameter and new track phi, and then find the new vertex point. We run 1000 toy samples of this and find that the bias in lifetime due to this alignment scenario is only 0.3 microns which is smaller than that found by moving all layers. We therefore continue to use the previous systematic.

16.3 Mass and Lifetime Correlation

We have increased the number of pseudo experiments for this systematics to 1000 which was not possible earlier due to CPU constraints. While the systematic from the residual does increase from 2.4 to 2.5 microns this is within the error or the error on the systematic. The mass lifetime correlation in background is one of the larger systematic errors. In principle it is possible to invent a function that will take into account the correlation and reduce the systematic error, however this is not trivial. We leave this as an improvement that could be made to a future analysis. The mass and lifetime are not correlated for signal events, and a plots demonstrating this has been added to the main body of the note.

16.4 Final Systematic Errors

We present here a summary of all sources of updated systematic error for the $B^{\pm}$ decay mode.

Source	Assigned Error μm
Non-flat single track Efficiency wrt to impact parameter	3.1
Non-flat single track Efficiency wrt to track P_T	1.8
Correlations in Background	2.5
Background Parameterisation	0.8
Resolution	0.3
General Fitter Bias	0.4
Silicon Alignment	0.4
Non-flat single track Efficiency wrt to Eta	0.3

We add these uncertainties for the charged B lifetime in quadrature and obtain $\pm$ 4.5 μm.

Most of these systematics errors relate to features in the data that the method is not yet sophisticated enough to accommodate. Therefore crosschecking these errors in data is not possible as the effect is already present and we cannot take it away. We can however check the systematic error due to using different

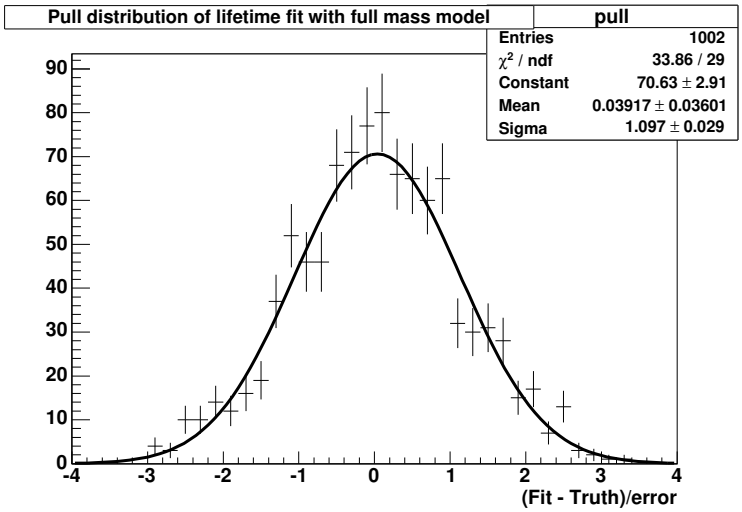


Figure 43: Pull distribution of standard Toy

functions. We use the alternate resolution function in a data fit and find that the fit results increases by 0.2 microns which is consistent with the quoted error. For background paramterisation we alter the spacing and number of parameters in the fit and again find that deviations in best fit lifetime are small and less than 0.8 microns.

17 Answers to other questions from Pre blessing

17.1 Scaling of the Statistical Error

We had previously stated that fixing the mass parameters caused an underestimation of the statistical error. This is wrong. The width of the pull is wide at 1.10±0.03 as seen in figure 43. We run toy using a simple mass model and keep the mass free, this still results in a pull distribution that is wide at 1.07 ±0.03 and is shown in figure 44. This indicates that fixing the mass fit is certainly not the cause of the error underestimation. In an attempt to discover where this underestimation is coming from we have tried fixing both the efficiency and the fisher scalar parameters to their truth values but continue to see a pull distribution that is slightly wide. Any difference we see in parabolic errors calculated by Minos and the error returned by Migrad is of order 0.05 microns, which is not large enough to account for a wider pull. In light of this we conclude that we don't fully understand why pull distribution is wide but that the underestimation of statistical error is small, and so we continue to scale the statistical error by ten percent.

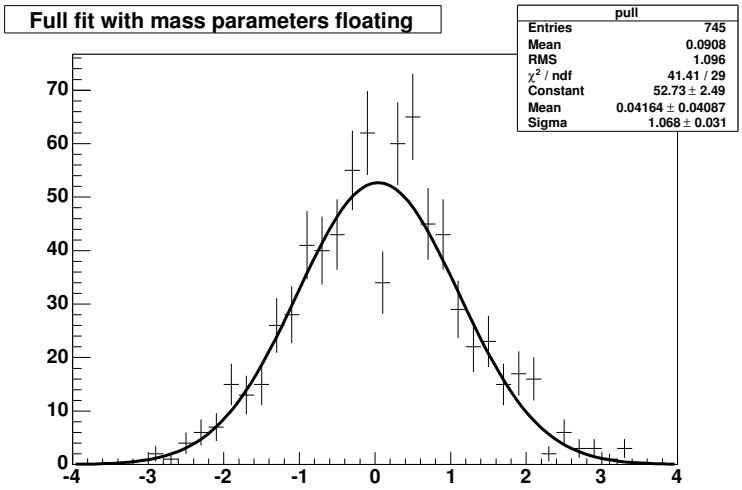


Figure 44: Pull distribution of toy using a floating simple mass model

17.2 Fisher Scalar distributions and plots for the fit to data

It would be useful to have similar plots for data as those we make for MC. In MC we have a plot of signal fraction as a function of fisherscalar, where we can plot the truth signal fraction in each bin as we have access to the truth information. In data this is obviously not the case, however another option would be to split the data in bins of fisher scalar and perform a mass fit to the events in each bin. From the mass fit we can obtain a signal fraction for that bin of fisher scalar which we can then compare to the value of signal fraction as a function of fisher scalar as returned by the fit using the lagrange interpolating polynomial function.

We have done this and is displayed in figure 45. In the tails there are not enough events to do a mass fit. The fisher scalar distribution itself is shown in figure 46.

17.3 Fits to realistic Monte Carlo

We have additional statistics in B-¿D pi realistic monte carlo for both the charged and neutral modes. We fit for B+ with an input lifetime of 496 75K events with best fit lifetime 493.3±3.2 microns and for B0 with an input lifetime of 464 microns and 71K events a lifetime of 467.8±2.8. The projections of these fits are shown in figures 47, 48.

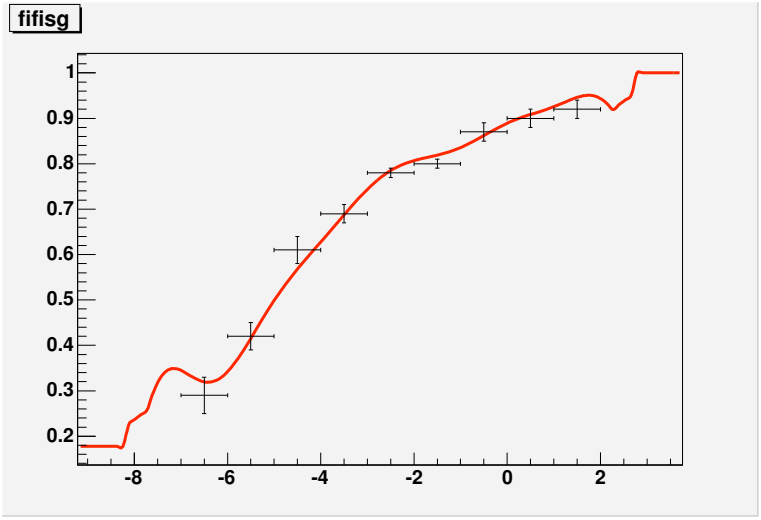


Figure 45: The red curve depicts the signal fraction as a function of fisher scalar as determind by the fit to data. the black points give the signal fraction as determined independently by mass fits.

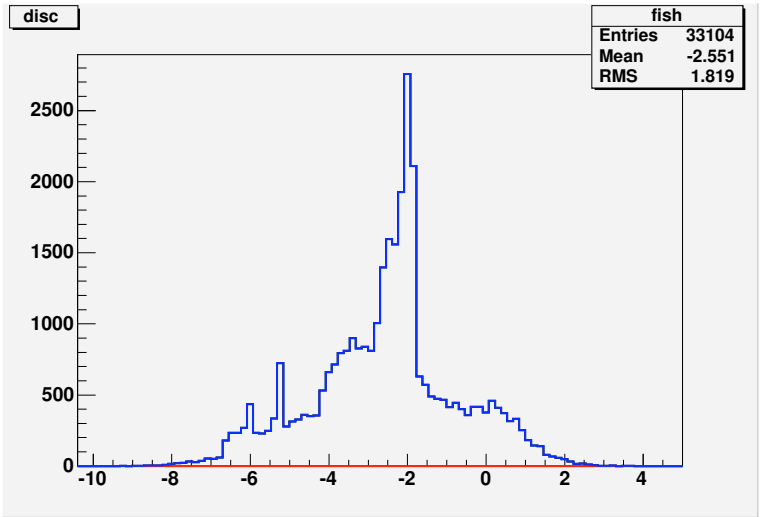


Figure 46: The fisher scalar distribution in data.

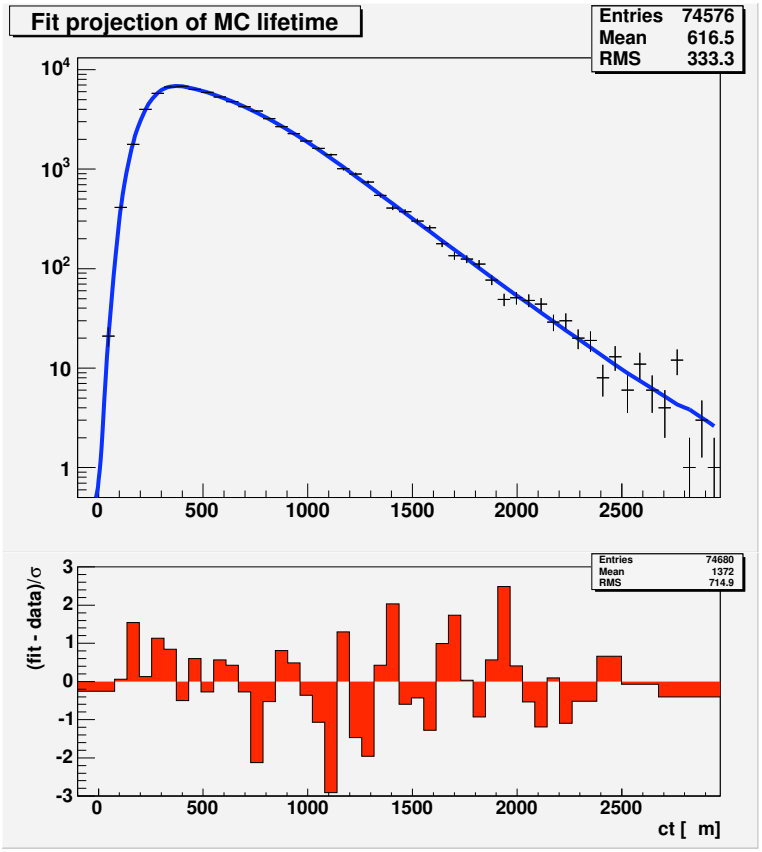


Figure 47: Bu realistic MC fit

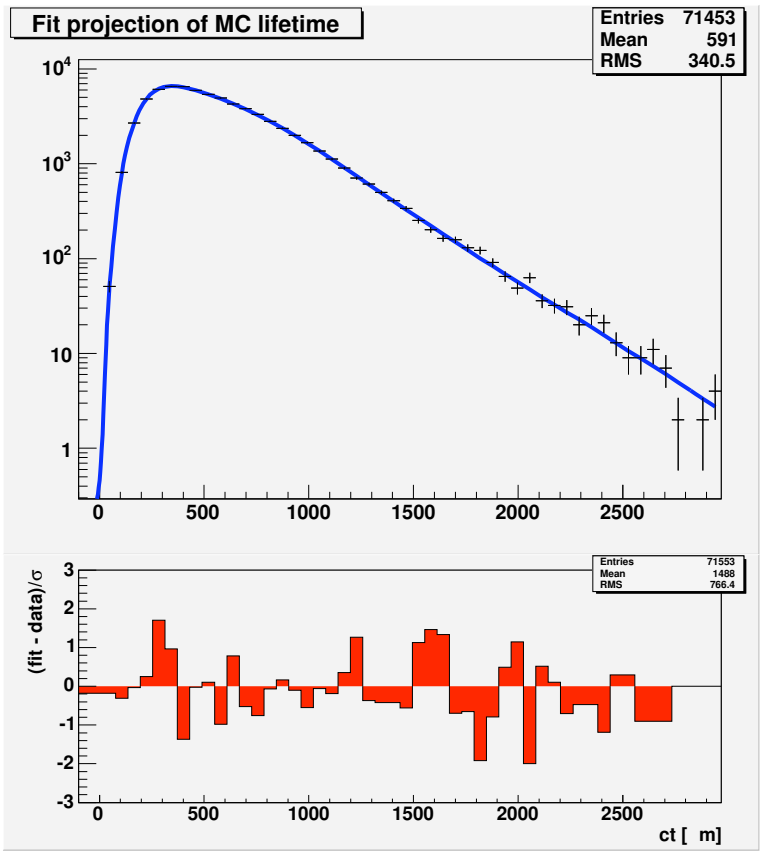


Figure 48: Bu realistic MC fit

17.4 Agreement between Toy and Data

There remains a small disagreement between the standard toy and data. It must be stressed that no information from the toy mc is used in the data fit, it is used simply for validation and assesment of systematic errors. Furthermore the toy has been tested in a variety of configurations with small variations on the standard and extreme variations and the method has been robust against these all. CDF generator momentum spectrums have been used and we have tried further tweaking of parameters however the differences remain. It is unrealistic to expect better agreement without full detector simulation, therefore we no longer wish to pursue a better agreement of toy and data.

17.5 Calculation of Acceptance

For tracks that do not have SVT matches we draw delta from a histogram where delta is the difference between the SVT and offline impact parameter. How do we know that the distribution is valid for unmatched tracks. We can never know this but have the expectation that tracks are not found simply because the SVT algorithm does not that the time or patterns to pick up all tracks that are subsequently found offline. This is clear from the study of efficiency as a function of IP.

Having said this we do want to explore how sensistive we are to the details of this histogram and try two tests on data. The first is to shift the histogram by 10 microns so that tracks with no match are shifted in IP causing a change to all accpetance functions that contained an unmatched track. The second change was the increase the value of delta by 10% which means that all unmatched tracks have in general a larger differece between their offline and "svt" impact parameters. In both cases we find a shift of 0.1 microns indicating that there is little sensitivity to this plot. Any large problems relating to this would have been picked up in the realistic MC fits.

17.6 Affect of neglecting difference in accpetance distributions in data

We run a fit on data ignoring the difference between signal and backround accpetance functions. In practice this means instead of fitting the signal fraction as a function of fisher scalar we simply fit an overall signal fraction. We find a lifetime of 489.7 $\pm$ 6.4 microns. This is approx 8.5 microns different to the fit result. There are examples of fits to toyMC that exhibit a similar difference though the average difference in toy is 5 microns. This result simply demonstrates the need to take into account the distribution of acceptance functions.

18 Appendix 1: Full Fit results

The next pages detail the fit results.

The mass fit has 7 parameters. There is the mean mass of two gaussians, their two widths and the fraction of the first gaussian. The remaining two parameters are the slope of the background and the overall signal fraction. The overall signal fraction is fit as Nb/Ns where Nb and Ns are the numbers of background and signal events respectively.

The mass fit parameters are as follows:

0	massG1	5.27607e+00	3.96198e-04	8.66936e-07	8.67397e-01
1	WidthG1	1.44530e-02	1.06187e-03	-4.28272e-06	-2.44704e+00
2	massG2	5.26991e+00	2.59145e-03	9.73308e-06	6.57738e-01
3	WidthG2	2.64230e-02	3.45552e-03	-1.17986e-05	-2.54983e-01
4	FracG2	4.38550e-01	1.35813e-01	1.10956e-03	-1.42748e-02
5	polyBg	-1.65742e-01	3.73389e-03	-2.94891e-04	-1.10570e-02
6	FracBtoS	3.69421e-01	1.28983e-02	6.27225e-06	4.83176e-02

The error matrix is

	0	1	2	3	4
0	1.57e-07	-2.88e-07	5.694e-07	-9.587e-07	3.783e-05
1	-2.88e-07	1.128e-06	-2.574e-06	3.431e-06	-0.0001432
2	5.694e-07	-2.574e-06	6.716e-06	-8.597e-06	0.0003443
3	-9.587e-07	3.431e-06	-8.597e-06	1.194e-05	-0.000463
4	3.783e-05	-0.0001432	0.0003443	-0.000463	0.01893
5	5.004e-07	-1.363e-06	3.801e-06	-6.194e-06	0.000184
6	2.037e-06	-5.32e-06	1.43e-05	-2.401e-05	0.0007267

	5	6
0	5.004e-07	2.037e-06
1	-1.363e-06	-5.32e-06
2	3.801e-06	1.43e-05
3	-6.194e-06	-2.401e-05
4	0.000184	0.0007267
5	1.401e-05	3.997e-05
6	3.997e-05	0.0001664

The correlation matrix is

	0	1	2	3	4
0	1	-0.6846	0.5546	-0.7003	0.6942
1	-0.6846	1	-0.9355	0.9349	-0.9806
2	0.5546	-0.9355	1	-0.9601	0.9658
3	-0.7003	0.9349	-0.9601	1	-0.974
4	0.6942	-0.9806	0.9658	-0.974	1
5	0.3374	-0.3428	0.3919	-0.4789	0.3573
6	0.3986	-0.3884	0.4277	-0.5386	0.4095

	5	6
0	0.3374	0.3986
1	-0.3428	-0.3884
2	0.3919	0.4277
3	-0.4789	-0.5386
4	0.3573	0.4095
5	1	0.8279
6	0.8279	1

The lifetime fit consists of 30 parameters. There is 1 mean lifetime. There are 3 different periods of efficiency and we fit one parameter in each for signal and background. There are then 10 parameters to describe the background lifetime distribution and finally 13 parameters to describe the signal fraction as a function of fisher scalar.

The parameters and their errors are given below:

0	ctau	4.98192e+02	6.18302e+00
1	effsg1	4.87819e-01	3.28771e-02
2	effbg1	5.07594e-01	6.62107e-02
3	effsg2	6.56376e-01	8.48179e-03
4	effbg2	5.22103e-01	1.86643e-02
5	effsg3	7.25423e-01	5.57023e-03
6	effbg3	5.59604e-01	1.74698e-02
7	ctBg1	1.05000e+01	3.51666e-01
8	ctBg2	7.07979e+00	5.85253e-02
9	ctBg3	4.76888e+00	3.68506e-02
10	ctBg4	2.71620e+00	4.37419e-02
11	ctBg5	1.27828e+00	6.67679e-02
12	ctBg6	1.28060e-01	1.00457e-01
13	ctBg7	-1.19225e+00	1.86124e-01
14	ctbg8	-1.94059e+00	2.92104e-01
15	ctBg9	-2.78179e+00	4.69809e-01
16	ctBg10	-7.15913e+00	2.62929e+00
17	fish1	1.78634e-01	7.63403e-02
18	fish2	2.59323e-01	6.30126e-02
19	fish3	3.28393e-01	2.34602e-02
20	fish4	3.79042e-01	1.31397e-02
21	fish5	5.38973e-01	1.08355e-02
22	fish6	6.60332e-01	7.79486e-03
23	fish7	7.68602e-01	5.54558e-03
24	fish8	8.13567e-01	5.11646e-03
25	fish9	8.49147e-01	7.60666e-03
26	fish10	9.00413e-01	7.18338e-03
27	fish11	9.36690e-01	1.02055e-02
28	fish12	9.19232e-01	3.61318e-02
29	fish13	1.00000e-00	4.80377e-02

The error matrix follows. The numbers refer to fit parameters as above.

	0	1	2	3	4
0	38.7	-0.0006851	0.002309	0.0001657	0.001966
1	-0.0006851	0.001082	-0.0006823	6.779e-09	-1.594e-07
2	0.002309	-0.0006823	0.007169	-5.794e-08	6.894e-07
3	0.0001657	6.779e-09	-5.794e-08	7.467e-05	-5.282e-05
4	0.001966	-1.594e-07	6.894e-07	-5.282e-05	0.0007453
5	0.0002425	2.038e-08	-1.235e-07	3.43e-08	-6.984e-09
6	0.001983	-2.672e-07	1.131e-06	-4.82e-08	7.479e-07
7	-0.1208	-1.211e-05	2.723e-05	-1.358e-06	-7.869e-05

8	-0.1041	1.401e-05	-5.409e-05	-3.114e-06	-1.342e-05
9	-0.1015	-8.889e-08	1.518e-05	-2.73e-06	1.259e-05
10	-0.1057	6.388e-06	-9.303e-06	-1.285e-06	1.426e-05
11	-0.1244	-1.529e-06	2.687e-05	7.271e-07	5.076e-06
12	-0.1512	2.786e-05	-0.0001049	-1.556e-06	1.494e-05
13	-0.2029	1.815e-05	-6.923e-05	-4.417e-06	2.493e-05
14	-0.2374	5.479e-06	-7.768e-06	4.869e-06	-2.983e-05
15	-0.2882	3.326e-05	-0.0001385	-1.002e-05	4.579e-05
16	-0.3549	-0.0002314	0.001075	5.638e-06	-3.954e-05
17	0.001054	-5.588e-07	1.998e-06	4.161e-07	-2.035e-06
18	-0.005169	1.212e-06	-4.655e-06	-8.063e-07	1.357e-06
19	-0.004498	-5.259e-07	2.568e-06	7.601e-07	-7.339e-06
20	-0.006362	2.215e-06	-8.768e-06	3.387e-07	-4.965e-06
21	-0.005463	-1.277e-07	1.571e-06	1.162e-07	-3.912e-06
22	-0.004417	1.565e-06	-6.304e-06	-3.225e-07	-1.558e-06
23	-0.003118	1.075e-06	-4.678e-06	-1.718e-07	-1.669e-06
24	-0.002201	-7.522e-07	4.51e-06	-1.094e-07	-1.755e-06
25	-0.001981	-1.765e-07	2.345e-06	-6.08e-07	5.097e-07
26	-0.001591	-5.637e-07	3.5e-06	-8e-07	1.628e-06
27	-0.001101	7.421e-07	-2.587e-06	-7.265e-07	2.39e-06
28	-0.0005338	-9.856e-06	4.413e-05	-1.28e-06	2.43e-06
29	-1.965e-09	3.789e-12	-1.887e-11	9.445e-12	-3.622e-11

	5	6	7	8	9
0	0.0002425	0.001983	-0.1208	-0.1041	-0.1015
1	2.038e-08	-2.672e-07	-1.211e-05	1.401e-05	-8.889e-08
2	-1.235e-07	1.131e-06	2.723e-05	-5.409e-05	1.518e-05
3	3.43e-08	-4.82e-08	-1.358e-06	-3.114e-06	-2.73e-06
4	-6.984e-09	7.479e-07	-7.869e-05	-1.342e-05	1.259e-05
5	3.114e-05	-2.281e-05	-8.159e-06	-6.172e-06	-3.076e-06
6	-2.281e-05	0.0003454	-4.897e-05	8.895e-07	1.434e-05
7	-8.159e-06	-4.897e-05	0.4317	0.3059	0.308
8	-6.172e-06	8.895e-07	0.3059	0.3099	0.3067
9	-3.076e-06	1.434e-05	0.308	0.3067	0.3086
10	-1.867e-06	1.762e-05	0.3068	0.3069	0.3073
11	1.056e-06	3.037e-06	0.3075	0.307	0.3081
12	-2.018e-06	1.928e-05	0.3085	0.3083	0.3091
13	2.953e-07	-3.554e-07	0.3104	0.3101	0.3111
14	-3.763e-06	1.97e-05	0.3122	0.312	0.3128
15	-1.186e-06	-5.2e-07	0.3171	0.3168	0.3177
16	-3.311e-05	0.0002049	0.3421	0.3418	0.3427
17	7.609e-08	7.322e-07	-2.784e-05	-0.0001199	-0.0001171
18	-2.345e-06	5.97e-06	0.0002844	8.74e-05	3.598e-05
19	4.498e-07	-5.996e-06	0.0001398	7.263e-05	2.308e-05
20	1.065e-07	-4.797e-06	9.102e-05	6.793e-05	4.114e-05
21	-1.69e-07	-2.604e-06	0.0001375	5.92e-05	3.732e-05
22	-3.091e-07	-2.314e-06	0.0001198	5.663e-05	4.458e-05
23	-2.023e-07	-1.881e-06	4.048e-05	3.403e-05	2.15e-05
24	-3.528e-07	-9.665e-07	3.34e-05	2.244e-05	1.691e-05
25	-6.166e-07	-4.873e-07	4.048e-05	2.597e-05	1.535e-05
26	-7.745e-07	7.71e-07	1.095e-05	1.579e-05	9.053e-06
27	-8.852e-07	2.474e-06	6.582e-06	4.479e-06	8.274e-06
28	-3.832e-06	1.846e-05	4.931e-06	6.807e-06	1.9e-05
29	1.728e-11	-8.261e-11	1.477e-10	1.471e-10	1.736e-10

	10	11	12	13	14
0	-0.1057	-0.1244	-0.1512	-0.2029	-0.2374
1	6.388e-06	-1.529e-06	2.786e-05	1.815e-05	5.479e-06
2	-9.303e-06	2.687e-05	-0.0001049	-6.923e-05	-7.768e-06
3	-1.285e-06	7.271e-07	-1.556e-06	-4.417e-06	4.869e-06
4	1.426e-05	5.076e-06	1.494e-05	2.493e-05	-2.983e-05
5	-1.867e-06	1.056e-06	-2.018e-06	2.953e-07	-3.763e-06
6	1.762e-05	3.037e-06	1.928e-05	-3.554e-07	1.97e-05
7	0.3068	0.3075	0.3085	0.3104	0.3122
8	0.3069	0.307	0.3083	0.3101	0.312
9	0.3073	0.3081	0.3091	0.3111	0.3128
10	0.3094	0.3073	0.3096	0.3111	0.3131
11	0.3073	0.3139	0.3082	0.3133	0.3141
12	0.3096	0.3082	0.3263	0.3088	0.3196
13	0.3111	0.3133	0.3088	0.3623	0.3071
14	0.3131	0.3141	0.3196	0.3071	0.4217
15	0.3178	0.3195	0.3209	0.339	0.287
16	0.3432	0.3444	0.3501	0.3306	0.5117
17	-0.000116	-0.0001598	-8.816e-05	-0.0005304	-0.0002307
18	-3.407e-06	-3.746e-05	2.183e-05	-5.402e-05	-0.0002185
19	3.91e-06	-2.307e-06	-9.274e-05	-4.867e-05	-3.89e-05
20	9.776e-06	-4.164e-06	-3.251e-06	-1.19e-07	-6.088e-05
21	2.984e-05	1.207e-05	1.196e-05	-2.794e-05	2.41e-05
22	3.395e-05	3.148e-05	1.515e-05	2.498e-06	5.099e-05
23	1.243e-05	1.532e-05	1.082e-05	1.84e-05	2.319e-05
24	1.106e-05	9.012e-06	9.218e-06	1.109e-05	1.775e-05
25	1.292e-05	1.115e-05	7.376e-06	1.159e-05	2.102e-05
26	7.546e-06	9.283e-06	4.817e-06	1.133e-05	1.616e-05
27	6.456e-06	4.768e-06	5.365e-06	1.163e-05	9.065e-06
28	2.051e-05	-1.001e-06	-1.223e-05	2.375e-05	1.275e-07
29	2.908e-11	4.001e-10	5.379e-10	2.633e-10	7.381e-10

	15	16	17	18	19
0	-0.2882	-0.3549	0.001054	-0.005169	-0.004498
1	3.326e-05	-0.0002314	-5.588e-07	1.212e-06	-5.259e-07
2	-0.0001385	0.001075	1.998e-06	-4.655e-06	2.568e-06
3	-1.002e-05	5.638e-06	4.161e-07	-8.063e-07	7.601e-07
4	4.579e-05	-3.954e-05	-2.035e-06	1.357e-06	-7.339e-06
5	-1.186e-06	-3.311e-05	7.609e-08	-2.345e-06	4.498e-07
6	-5.2e-07	0.0002049	7.322e-07	5.97e-06	-5.996e-06
7	0.3171	0.3421	-2.784e-05	0.0002844	0.0001398
8	0.3168	0.3418	-0.0001199	8.74e-05	7.263e-05
9	0.3177	0.3427	-0.0001171	3.598e-05	2.308e-05
10	0.3178	0.3432	-0.000116	-3.407e-06	3.91e-06
11	0.3195	0.3444	-0.0001598	-3.746e-05	-2.307e-06
12	0.3209	0.3501	-8.816e-05	2.183e-05	-9.274e-05
13	0.339	0.3306	-0.0005304	-5.402e-05	-4.867e-05
14	0.287	0.5117	-0.0002307	-0.0002185	-3.89e-05
15	0.6876	-0.2944	-0.0004735	-5.495e-05	6.308e-05
16	-0.2944	11.01	0.000153	-6.871e-05	-8.72e-05
17	-0.0004735	0.000153	0.01333	-0.001277	0.0001237
18	-5.495e-05	-6.871e-05	-0.001277	0.004519	-0.0003179

19	6.308e-05	-8.72e-05	0.0001237	-0.0003179	0.0005879
20	1.56e-05	-0.0001197	-7.266e-05	0.0001619	-5.362e-05
21	3.079e-05	-0.0001602	4.956e-05	-0.0001038	2.36e-05
22	2.483e-06	0.0001512	-1.177e-05	3.106e-05	2.166e-06
23	1.619e-05	5.081e-05	-4.634e-07	-1.5e-06	-3.956e-06
24	2.159e-05	5.965e-05	-6.769e-06	1.623e-05	-6.413e-08
25	1.587e-05	6.377e-05	-1.094e-05	1.966e-05	-7.593e-06
26	1.982e-05	1.282e-06	1.164e-05	-2.092e-05	7.312e-06
27	1.12e-05	3.72e-05	-3.279e-06	4.83e-06	-3.342e-06
28	-7.517e-07	0.0001167	-0.0001273	0.000248	-5.489e-05
29	8.275e-11	1.173e-09	-3.717e-10	7.204e-10	-1.495e-10

	20	21	22	23	24
0	-0.006362	-0.005463	-0.004417	-0.003118	-0.002201
1	2.215e-06	-1.277e-07	1.565e-06	1.075e-06	-7.522e-07
2	-8.768e-06	1.571e-06	-6.304e-06	-4.678e-06	4.51e-06
3	3.387e-07	1.162e-07	-3.225e-07	-1.718e-07	-1.094e-07
4	-4.965e-06	-3.912e-06	-1.558e-06	-1.669e-06	-1.755e-06
5	1.065e-07	-1.69e-07	-3.091e-07	-2.023e-07	-3.528e-07
6	-4.797e-06	-2.604e-06	-2.314e-06	-1.881e-06	-9.665e-07
7	9.102e-05	0.0001375	0.0001198	4.048e-05	3.34e-05
8	6.793e-05	5.92e-05	5.663e-05	3.403e-05	2.244e-05
9	4.114e-05	3.732e-05	4.458e-05	2.15e-05	1.691e-05
10	9.776e-06	2.984e-05	3.395e-05	1.243e-05	1.106e-05
11	-4.164e-06	1.207e-05	3.148e-05	1.532e-05	9.012e-06
12	-3.251e-06	1.196e-05	1.515e-05	1.082e-05	9.218e-06
13	-1.19e-07	-2.794e-05	2.498e-06	1.84e-05	1.109e-05
14	-6.088e-05	2.41e-05	5.099e-05	2.319e-05	1.775e-05
15	1.56e-05	3.079e-05	2.483e-06	1.619e-05	2.159e-05
16	-0.0001197	-0.0001602	0.0001512	5.081e-05	5.965e-05
17	-7.266e-05	4.956e-05	-1.177e-05	-4.634e-07	-6.769e-06
18	0.0001619	-0.0001038	3.106e-05	-1.5e-06	1.623e-05
19	-5.362e-05	2.36e-05	2.166e-06	-3.956e-06	-6.413e-08
20	0.0001955	1.443e-05	-1.009e-05	7.923e-06	-2.644e-06
21	1.443e-05	0.0001435	1.914e-05	-9.632e-06	3.603e-06
22	-1.009e-05	1.914e-05	7.075e-05	7.325e-06	1.305e-06
23	7.923e-06	-9.632e-06	7.325e-06	3.637e-05	-5.375e-06
24	-2.644e-06	3.603e-06	1.305e-06	-5.375e-06	3.345e-05
25	5.546e-06	-7.174e-06	6.507e-06	-4.634e-06	1.268e-05
26	-2.749e-06	5.352e-06	-4.533e-06	4.154e-06	-6.092e-06
27	2.806e-06	-3.71e-06	1.682e-06	-1.008e-06	2.135e-06
28	6.421e-06	-2.275e-05	3.17e-05	-3.371e-05	2.867e-05
29	5.127e-12	-5.103e-11	8.439e-11	-9.373e-11	7.271e-11

	25	26	27	28	29
0	-0.001981	-0.001591	-0.001101	-0.0005338	-1.965e-09
1	-1.765e-07	-5.637e-07	7.421e-07	-9.856e-06	3.789e-12
2	2.345e-06	3.5e-06	-2.587e-06	4.413e-05	-1.887e-11
3	-6.08e-07	-8e-07	-7.265e-07	-1.28e-06	9.445e-12
4	5.097e-07	1.628e-06	2.39e-06	2.43e-06	-3.622e-11
5	-6.166e-07	-7.745e-07	-8.852e-07	-3.832e-06	1.728e-11
6	-4.873e-07	7.71e-07	2.474e-06	1.846e-05	-8.261e-11
7	4.048e-05	1.095e-05	6.582e-06	4.931e-06	1.477e-10

8	2.597e-05	1.579e-05	4.479e-06	6.807e-06	1.471e-10
9	1.535e-05	9.053e-06	8.274e-06	1.9e-05	1.736e-10
10	1.292e-05	7.546e-06	6.456e-06	2.051e-05	2.908e-11
11	1.115e-05	9.283e-06	4.768e-06	-1.001e-06	4.001e-10
12	7.376e-06	4.817e-06	5.365e-06	-1.223e-05	5.379e-10
13	1.159e-05	1.133e-05	1.163e-05	2.375e-05	2.633e-10
14	2.102e-05	1.616e-05	9.065e-06	1.275e-07	7.381e-10
15	1.587e-05	1.982e-05	1.12e-05	-7.517e-07	8.275e-11
16	6.377e-05	1.282e-06	3.72e-05	0.0001167	1.173e-09
17	-1.094e-05	1.164e-05	-3.279e-06	-0.0001273	-3.717e-10
18	1.966e-05	-2.092e-05	4.83e-06	0.000248	7.204e-10
19	-7.593e-06	7.312e-06	-3.342e-06	-5.489e-05	-1.495e-10
20	5.546e-06	-2.749e-06	2.806e-06	6.421e-06	5.127e-12
21	-7.174e-06	5.352e-06	-3.71e-06	-2.275e-05	-5.103e-11
22	6.507e-06	-4.533e-06	1.682e-06	3.17e-05	8.439e-11
23	-4.634e-06	4.154e-06	-1.008e-06	-3.371e-05	-9.373e-11
24	1.268e-05	-6.092e-06	2.135e-06	2.867e-05	7.271e-11
25	6.457e-05	5.621e-06	1.509e-06	-5.911e-05	-1.731e-10
26	5.621e-06	5.951e-05	-8.77e-06	5.7e-05	1.698e-10
27	1.509e-06	-8.77e-06	0.0001272	-0.0001492	-3.593e-10
28	-5.911e-05	5.7e-05	-0.0001492	0.001877	1.997e-09
29	-1.731e-10	1.698e-10	-3.593e-10	1.997e-09	1.36e-11

The Correlation matrix follows:

	0	1	2	3	4
0	1	-0.003348	0.004384	0.003083	0.01158
1	-0.003348	1	-0.245	2.385e-05	-0.0001775
2	0.004384	-0.245	1	-7.919e-05	0.0002982
3	0.003083	2.385e-05	-7.919e-05	1	-0.2239
4	0.01158	-0.0001775	0.0002982	-0.2239	1
5	0.006986	0.000111	-0.0002614	0.0007113	-4.584e-05
6	0.01715	-0.0004372	0.0007186	-0.0003002	0.001474
7	-0.02956	-0.0005603	0.0004896	-0.0002392	-0.004387
8	-0.03006	0.0007652	-0.001147	-0.0006473	-0.0008833
9	-0.02938	-4.865e-06	0.0003227	-0.0005687	0.0008301
10	-0.03056	0.0003492	-0.0001975	-0.0002673	0.0009392
11	-0.03569	-8.298e-05	0.0005664	0.0001502	0.0003319
12	-0.04255	0.001483	-0.00217	-0.0003153	0.0009579
13	-0.0542	0.0009168	-0.001358	-0.0008492	0.001517
14	-0.05877	0.0002565	-0.0001413	0.0008678	-0.001683
15	-0.05587	0.001219	-0.001972	-0.001399	0.002023
16	-0.01719	-0.00212	0.003827	0.0001966	-0.0004365
17	0.001468	-0.0001472	0.0002045	0.0004172	-0.0006458
18	-0.01236	0.000548	-0.0008179	-0.001388	0.0007393
19	-0.02982	-0.0006594	0.001251	0.003628	-0.01109
20	-0.07314	0.004816	-0.007406	0.002804	-0.01301
21	-0.0733	-0.0003241	0.001549	0.001123	-0.01196
22	-0.08441	0.005657	-0.008852	-0.004437	-0.006785
23	-0.0831	0.005418	-0.009161	-0.003296	-0.01014
24	-0.06118	-0.003954	0.00921	-0.00219	-0.01112
25	-0.03963	-0.0006678	0.003446	-0.008756	0.002324
26	-0.03316	-0.002222	0.005359	-0.012	0.007731

27	-0.01568	0.002	-0.002708	-0.007453	0.007762
28	-0.00198	-0.006916	0.01203	-0.003418	0.002055
29	-8.563e-05	3.124e-05	-6.042e-05	0.0002964	-0.0003597

	5	6	7	8	9
0	0.006986	0.01715	-0.02956	-0.03006	-0.02938
1	0.000111	-0.0004372	-0.0005603	0.0007652	-4.865e-06
2	-0.0002614	0.0007186	0.0004896	-0.001147	0.0003227
3	0.0007113	-0.0003002	-0.0002392	-0.0006473	-0.0005687
4	-4.584e-05	0.001474	-0.004387	-0.0008833	0.0008301
5	1	-0.22	-0.002225	-0.001987	-0.0009923
6	-0.22	1	-0.004011	8.598e-05	0.001389
7	-0.002225	-0.004011	1	0.8364	0.8438
8	-0.001987	8.598e-05	0.8364	1	0.9919
9	-0.0009923	0.001389	0.8438	0.9919	1
10	-0.0006016	0.001705	0.8395	0.9913	0.9944
11	0.0003377	0.0002917	0.8354	0.9843	0.9898
12	-0.000633	0.001816	0.822	0.9696	0.974
13	8.791e-05	-3.177e-05	0.785	0.9256	0.9303
14	-0.001038	0.001632	0.7318	0.863	0.8671
15	-0.0002562	-3.374e-05	0.5821	0.6863	0.6897
16	-0.001788	0.003322	0.1569	0.185	0.1859
17	0.0001181	0.0003413	-0.0003671	-0.001866	-0.001827
18	-0.006252	0.004779	0.00644	0.002336	0.0009635
19	0.003324	-0.01331	0.008775	0.00538	0.001714
20	0.001364	-0.01846	0.009908	0.008728	0.005296
21	-0.002528	-0.0117	0.01747	0.008876	0.005607
22	-0.006585	-0.0148	0.02168	0.01209	0.009541
23	-0.00601	-0.01678	0.01022	0.01014	0.006417
24	-0.01093	-0.008993	0.008791	0.006972	0.005264
25	-0.01375	-0.003263	0.007668	0.005806	0.003438
26	-0.01799	0.005378	0.00216	0.003678	0.002112
27	-0.01406	0.0118	0.0008881	0.0007134	0.00132
28	-0.01585	0.02293	0.0001732	0.0002823	0.0007893
29	0.0008393	-0.001205	6.097e-05	7.165e-05	8.472e-05

	10	11	12	13	14
0	-0.03056	-0.03569	-0.04255	-0.0542	-0.05877
1	0.0003492	-8.298e-05	0.001483	0.0009168	0.0002565
2	-0.0001975	0.0005664	-0.00217	-0.001358	-0.0001413
3	-0.0002673	0.0001502	-0.0003153	-0.0008492	0.0008678
4	0.0009392	0.0003319	0.0009579	0.001517	-0.001683
5	-0.0006016	0.0003377	-0.000633	8.791e-05	-0.001038
6	0.001705	0.0002917	0.001816	-3.177e-05	0.001632
7	0.8395	0.8354	0.822	0.785	0.7318
8	0.9913	0.9843	0.9696	0.9256	0.863
9	0.9944	0.9898	0.974	0.9303	0.8671
10	1	0.9863	0.9745	0.9292	0.8669
11	0.9863	1	0.9631	0.9291	0.8632
12	0.9745	0.9631	1	0.8981	0.8616
13	0.9292	0.9291	0.8981	1	0.7857
14	0.8669	0.8632	0.8616	0.7857	1

15	0.6891	0.6878	0.6775	0.6791	0.5329
16	0.1859	0.1853	0.1847	0.1655	0.2375
17	-0.001807	-0.002472	-0.001337	-0.007634	-0.003078
18	-9.112e-05	-0.0009947	0.0005686	-0.001335	-0.005006
19	0.0002899	-0.0001699	-0.006696	-0.003335	-0.00247
20	0.001257	-0.0005316	-0.000407	-1.413e-05	-0.006704
21	0.004478	0.001799	0.001748	-0.003875	0.003098
22	0.007258	0.006681	0.003154	0.0004934	0.009335
23	0.003707	0.004533	0.003142	0.00507	0.005921
24	0.003439	0.002781	0.00279	0.003187	0.004726
25	0.00289	0.002477	0.001607	0.002396	0.004028
26	0.001759	0.002148	0.001093	0.00244	0.003226
27	0.001029	0.0007545	0.0008327	0.001713	0.001237
28	0.0008513	-4.122e-05	-0.0004942	0.0009109	4.531e-06
29	1.417e-05	0.0001936	0.0002553	0.0001186	0.0003081

	15	16	17	18	19
0	-0.05587	-0.01719	0.001468	-0.01236	-0.02982
1	0.001219	-0.00212	-0.0001472	0.000548	-0.0006594
2	-0.001972	0.003827	0.0002045	-0.0008179	0.001251
3	-0.001399	0.0001966	0.0004172	-0.001388	0.003628
4	0.002023	-0.0004365	-0.0006458	0.0007393	-0.01109
5	-0.0002562	-0.001788	0.0001181	-0.006252	0.003324
6	-3.374e-05	0.003322	0.0003413	0.004779	-0.01331
7	0.5821	0.1569	-0.0003671	0.00644	0.008775
8	0.6863	0.185	-0.001866	0.002336	0.00538
9	0.6897	0.1859	-0.001827	0.0009635	0.001714
10	0.6891	0.1859	-0.001807	-9.112e-05	0.0002899
11	0.6878	0.1853	-0.002472	-0.0009947	-0.0001699
12	0.6775	0.1847	-0.001337	0.0005686	-0.006696
13	0.6791	0.1655	-0.007634	-0.001335	-0.003335
14	0.5329	0.2375	-0.003078	-0.005006	-0.00247
15	1	-0.107	-0.004947	-0.0009857	0.003137
16	-0.107	1	0.0003995	-0.000308	-0.001084
17	-0.004947	0.0003995	1	-0.1646	0.04418
18	-0.0009857	-0.000308	-0.1646	1	-0.195
19	0.003137	-0.001084	0.04418	-0.195	1
20	0.001346	-0.00258	-0.04502	0.1722	-0.1582
21	0.003099	-0.00403	0.03584	-0.1289	0.08123
22	0.000356	0.005418	-0.01212	0.05494	0.01062
23	0.003237	0.002539	-0.0006657	-0.003701	-0.02705
24	0.004503	0.003108	-0.01014	0.04175	-0.0004573
25	0.002381	0.002391	-0.01179	0.03639	-0.03897
26	0.003098	5.009e-05	0.01307	-0.04035	0.03909
27	0.001197	0.0009937	-0.002518	0.00637	-0.01222
28	-2.092e-05	0.000812	-0.02545	0.08514	-0.05225
29	2.706e-05	9.584e-05	-0.000873	0.002906	-0.001671

	20	21	22	23	24
0	-0.07314	-0.0733	-0.08441	-0.0831	-0.06118
1	0.004816	-0.0003241	0.005657	0.005418	-0.003954
2	-0.007406	0.001549	-0.008852	-0.009161	0.00921
3	0.002804	0.001123	-0.004437	-0.003296	-0.00219

4	-0.01301	-0.01196	-0.006785	-0.01014	-0.01112
5	0.001364	-0.002528	-0.006585	-0.00601	-0.01093
6	-0.01846	-0.0117	-0.0148	-0.01678	-0.008993
7	0.009908	0.01747	0.02168	0.01022	0.008791
8	0.008728	0.008876	0.01209	0.01014	0.006972
9	0.005296	0.005607	0.009541	0.006417	0.005264
10	0.001257	0.004478	0.007258	0.003707	0.003439
11	-0.0005316	0.001799	0.006681	0.004533	0.002781
12	-0.000407	0.001748	0.003154	0.003142	0.00279
13	-1.413e-05	-0.003875	0.0004934	0.00507	0.003187
14	-0.006704	0.003098	0.009335	0.005921	0.004726
15	0.001346	0.003099	0.000356	0.003237	0.004503
16	-0.00258	-0.00403	0.005418	0.002539	0.003108
17	-0.04502	0.03584	-0.01212	-0.0006657	-0.01014
18	0.1722	-0.1289	0.05494	-0.003701	0.04175
19	-0.1582	0.08123	0.01062	-0.02705	-0.0004573
20	1	0.08614	-0.08578	0.09397	-0.0327
21	0.08614	1	0.1899	-0.1333	0.052
22	-0.08578	0.1899	1	0.1444	0.02682
23	0.09397	-0.1333	0.1444	1	-0.1541
24	-0.0327	0.052	0.02682	-0.1541	1
25	0.04937	-0.07452	0.09628	-0.09564	0.2729
26	-0.02549	0.0579	-0.06986	0.0893	-0.1365
27	0.01779	-0.02745	0.01773	-0.01482	0.03273
28	0.0106	-0.04382	0.08699	-0.129	0.1144
29	9.941e-05	-0.001155	0.00272	-0.004214	0.003409

	25	26	27	28	29
0	-0.03963	-0.03316	-0.01568	-0.00198	-8.563e-05
1	-0.0006678	-0.002222	0.002	-0.006916	3.124e-05
2	0.003446	0.005359	-0.002708	0.01203	-6.042e-05
3	-0.008756	-0.012	-0.007453	-0.003418	0.0002964
4	0.002324	0.007731	0.007762	0.002055	-0.0003597
5	-0.01375	-0.01799	-0.01406	-0.01585	0.0008393
6	-0.003263	0.005378	0.0118	0.02293	-0.001205
7	0.007668	0.00216	0.0008881	0.0001732	6.097e-05
8	0.005806	0.003678	0.0007134	0.0002823	7.165e-05
9	0.003438	0.002112	0.00132	0.0007893	8.472e-05
10	0.00289	0.001759	0.001029	0.0008513	1.417e-05
11	0.002477	0.002148	0.0007545	-4.122e-05	0.0001936
12	0.001607	0.001093	0.0008327	-0.0004942	0.0002553
13	0.002396	0.00244	0.001713	0.0009109	0.0001186
14	0.004028	0.003226	0.001237	4.531e-06	0.0003081
15	0.002381	0.003098	0.001197	-2.092e-05	2.706e-05
16	0.002391	5.009e-05	0.0009937	0.000812	9.584e-05
17	-0.01179	0.01307	-0.002518	-0.02545	-0.000873
18	0.03639	-0.04035	0.00637	0.08514	0.002906
19	-0.03897	0.03909	-0.01222	-0.05225	-0.001671
20	0.04937	-0.02549	0.01779	0.0106	9.941e-05
21	-0.07452	0.0579	-0.02745	-0.04382	-0.001155
22	0.09628	-0.06986	0.01773	0.08699	0.00272
23	-0.09564	0.0893	-0.01482	-0.129	-0.004214
24	0.2729	-0.1365	0.03273	0.1144	0.003409
25	1	0.09068	0.01665	-0.1698	-0.00584
26	0.09068	1	-0.1008	0.1705	0.005967

27	0.01665	-0.1008	1	-0.3053	-0.008637
28	-0.1698	0.1705	-0.3053	1	0.0125
29	-0.00584	0.005967	-0.008637	0.0125	1

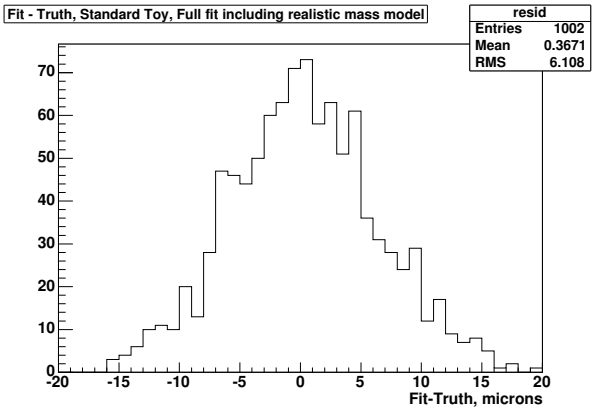


Figure 49: GeneralFitterBias. This is the residual of plot of figure 43 Systematic error assigned = 0.4 microns

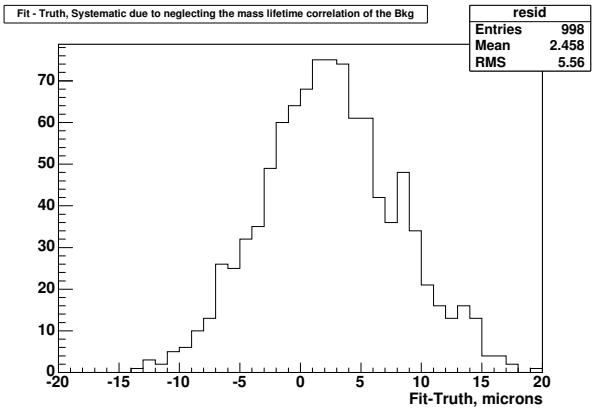


Figure 50: This is the residual of the bkg mass lifetime correlation pull study. Assigned systmatic error is 2.5 microns

19 Appendix 2: Systematic Study residual Plots

Contained in this section are the residual plots of Fit-Truth for the pull studies of Toy MC that resulted in a systematic error.

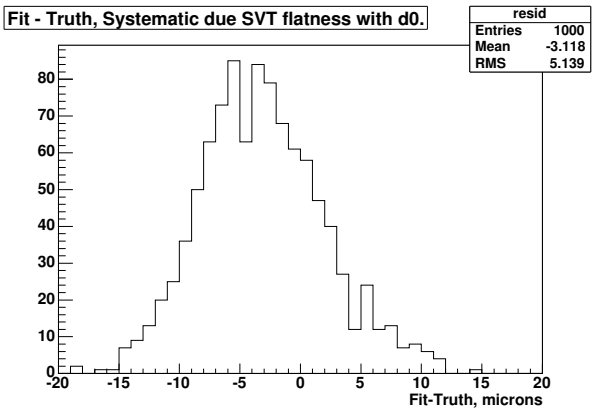


Figure 51: This is the residual of study using the worst case scenario of single track efficiency as a function of IP. Assigned error 3.1 microns

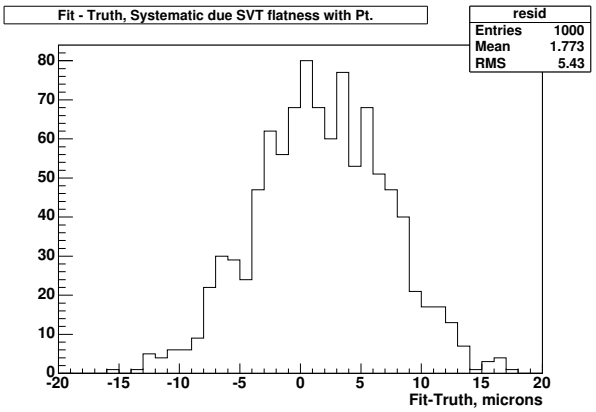


Figure 52: This is the residual of study using investigating the single track efficiency as a function of track Pt. Assigned error 1.8microns

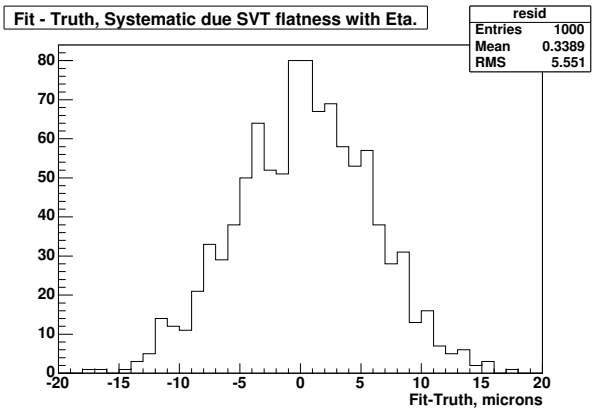


Figure 53: This is the residual of study investigating single track efficiency as a function of track eta. Assigned error 0.3 microns

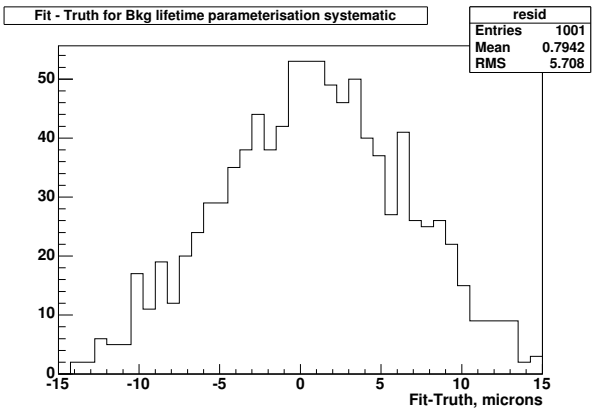


Figure 54: This is the residual of study using the an alternate background generation but the standard fit function. Assigned error 0.8 microns

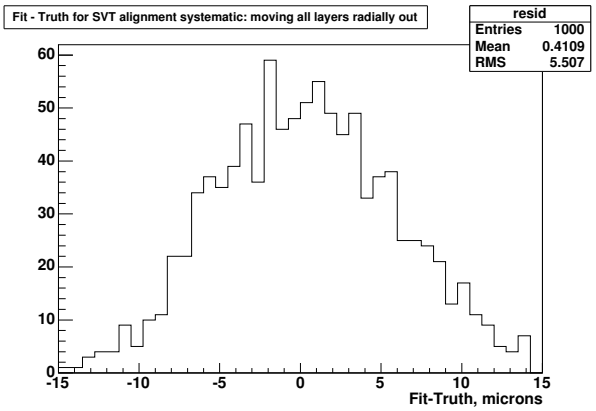


Figure 55: This is the residual of study for silicon alignment. Assigned error 0.4 microns

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 Phystat 2003. SLAC

20 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 μm 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

	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 4: Beamspot shifts induced by misalignment of silicon wafers.

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. 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.

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

The misaligments 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 beampsot. 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 ehanced 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 4 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 beasmpot 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 displacment divided by the wafer position, or (50 μ) / 2.5 (cm) = 0.2%. The D_0 vs ϕ_0 plots using both default beam positions and the apparent beam position as determined from our fits are shown in Figs. 56,

57, and 58

We noticed another effect: the 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 aproximately 30 microns, peak-to-peak. This effect. which is a D_0 distortion due to the collective shift of wafers, is in fact larger that 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.

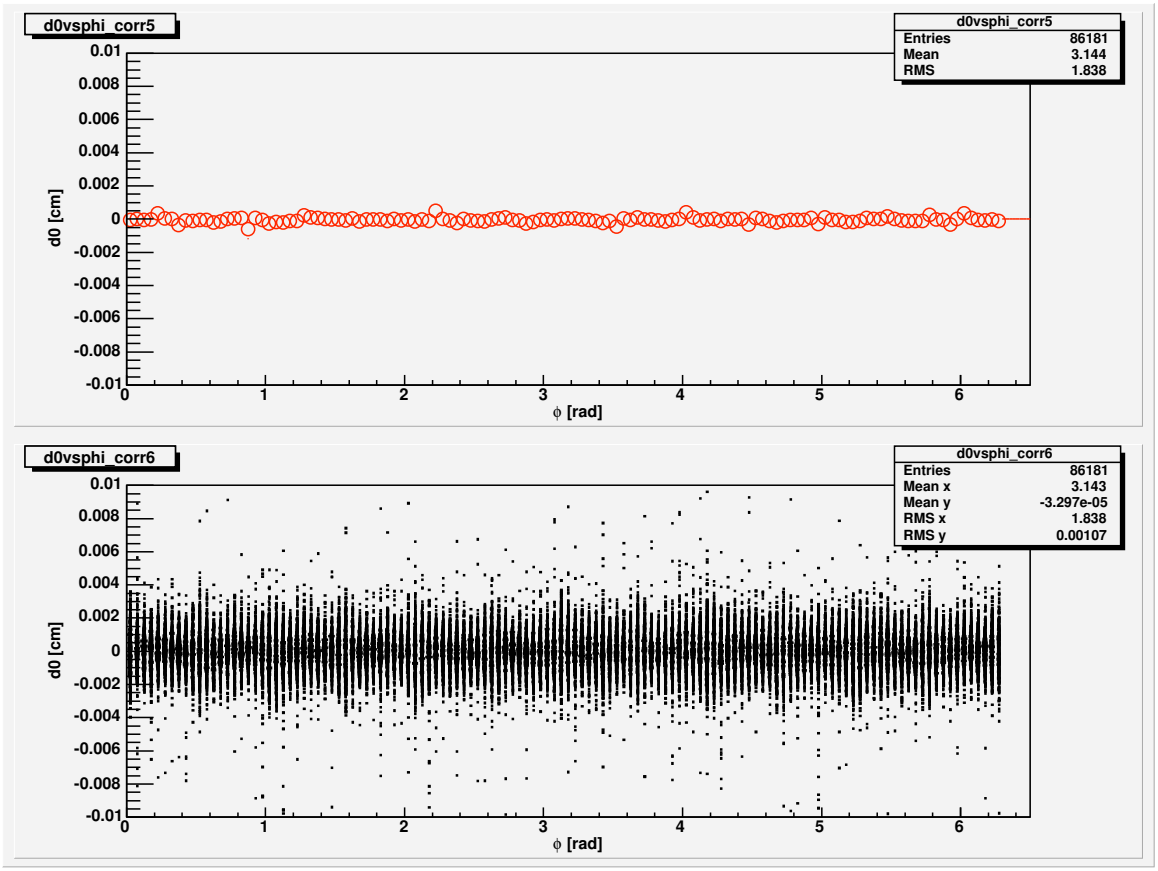


Figure 56: D_0 vs. ϕ_0 plot after correction for the beamspot. Events are single muons at 50 GeV. Events are simulated with tracks at their default positions and reconstructed in the same way.

We can also make the plot of the SVT D_0 vs ϕ_0, using SVTD banks after the

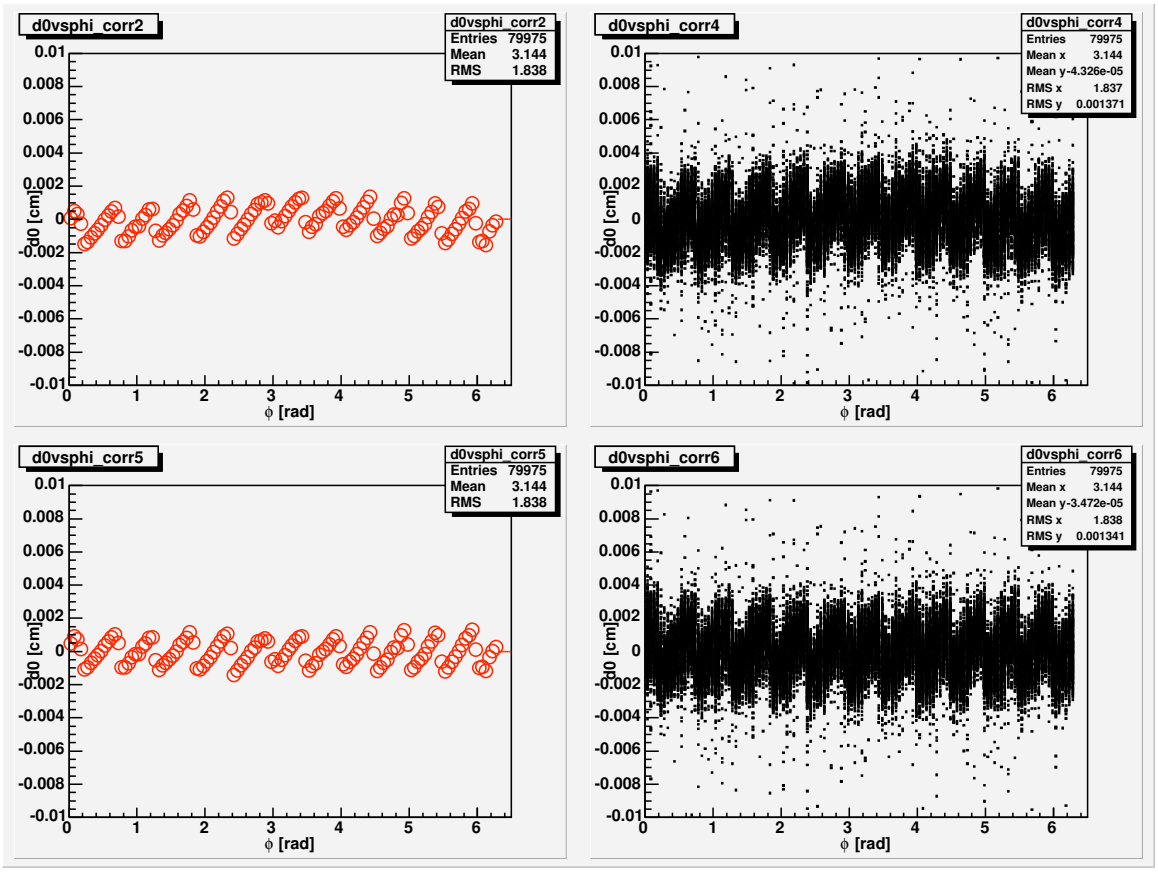


Figure 57: D_0 vs. ϕ_0 plot after correction for the beamspot. Events are single muons at 50 GeV. Events are simulated with tracks at their default positions and reconstructed with wafers moved out, by 50 μ m. Top: the default beamspot position is used. Bottom: misaligned beamspot position is used.

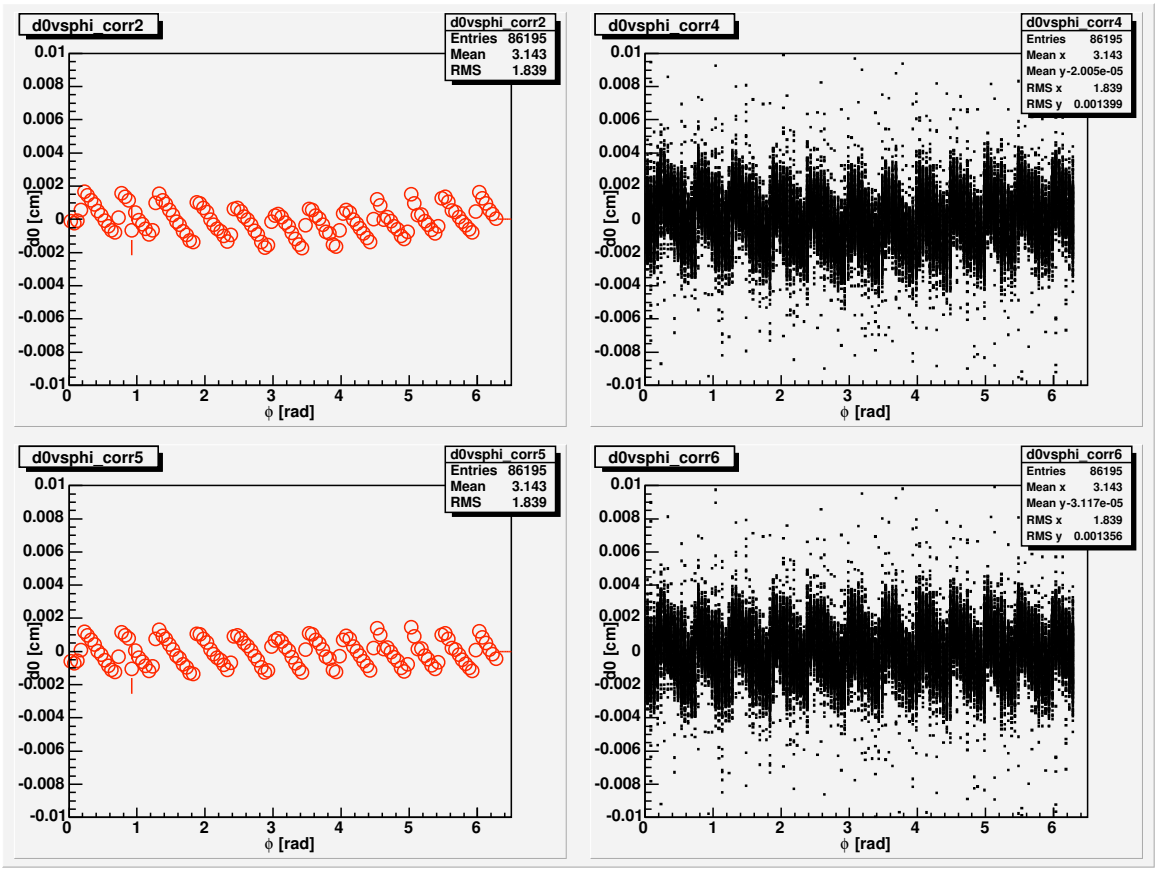


Figure 58: D_0 vs. ϕ_0 plot after correction for the beamspot. Events are single muons at 50 GeV. Events are simulated with tracks at their default positions and reconstructed with wafers moved in, by 50 μ m. Top: the default beamspot position is used: Bottom: misaligned beamspot position is used.

beamspots determined by our procedure have been loaded into SVTSIM. This is shown in Fig 59. 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.

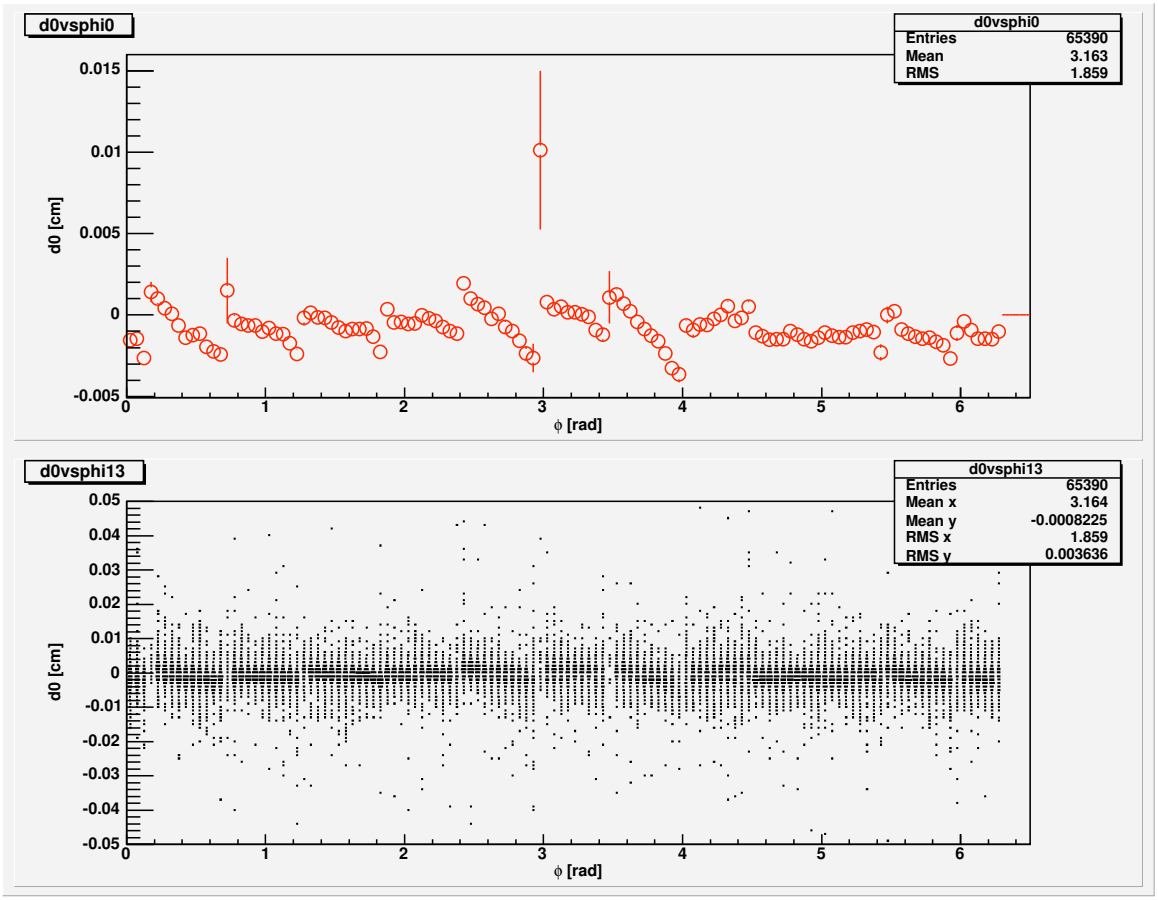


Figure 59: D_0 vs. ϕ_0 plot for the SVT tracks. The apparent beamspot position as determined from our fits are fed to the SVT.

20.1 Systematic Due to Silicon Misalignment

The alignment group quotes the systematic error on the alignment of the silicon system as being 50μm. 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 aligment (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.
- 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 $\delta \pm \sigma$.

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.

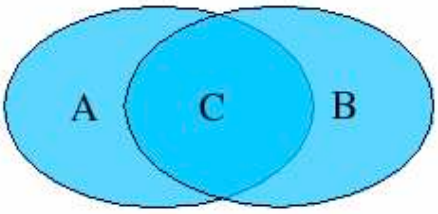


Figure 60: Sample 1 ovelaps with Sample 2. We consider Sample A (unique to selection 1), Sample B (unique to selection 2) and Sample C (common to both selections).

20.2 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. 60.

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,

$$x_1 = \frac{w_A x_A + w_C x_C}{w_A + w_C} \tag{59}$$

$$w_1 = w_A + w_C \tag{60}$$

And

$$x_2 = \frac{w_B x_B + w_C x_C}{w_B + w_C} \tag{61}$$

$$w_2 = w_B + w_C \tag{62}$$

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:

$$\Delta = x_1 - x_2 \quad = \tag{63}$$

$$\frac{w_A x_A + w_C x_C}{w_A + w_C} - \frac{w_B x_B + w_C x_C}{w_B + w_C} \tag{64}$$

The advantage in doing so is that we know the degree of statistical correlation between x_A, x_B, and x_C is zero, wheras 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

$$\sigma_\Delta^2 \quad = \quad \sum_{i=A,B,C} \left(\frac{\partial \Delta}{\partial x_i}^2 \sigma_i^2 \right) \tag{65}$$

Carrying out the algebra, we obtain:

$$\sigma_\Delta^2 \quad = \quad \frac{w_A}{(w_A + w_C)^2} + \frac{w_B}{(w_B + w_C)^2} + \frac{w_C (w_B - w_A)^2}{(w_A + w_C)^2 (w_B + w_C)^2} \tag{66}$$

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 5: 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.

20.3 Estimate of the Alignment Systematic.

The lifetime shift has been evaluated by shifting the wafers outwards by 50 μm. We simulated $B^+ \to D^0\pi^+$, 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 20.

The events which are common to each selection are about 60% of the selection. The raw numbers are shown in Table 20.3. 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 A$\cup$C, gives a measurement of 495 $\pm$ 5 while sample 2, B$\cup$C, gives 493$\pm$5, so we observe a downward shift of two microns.

Using the formulae derived in this note, we determine the shift to be -2 $\pm$ 5 μm. This is large and dominated by statistical uncertainty. Given the sizes of the other systematics we wish to pursue a more agressive approach.

Acknowledgements

We give our thanks to Marjorie Shapiro, Amanda Deisher and Hung Chung Fang for access to their stripped data and Monte Carlo. To Guillelmo 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. To Stefano Giagu for help with the SVT matching code. Also to Stefano Torre for help in understanding changes in the SVT. We would also like to thank numerous sub-convenors and convenors including Giovanni Punzi, Manfred Paulini, Diego Tonelli, Gavril Giurgiu and Sinead Farrington for their useful advice. Also many thanks to our mascot the Moose to continue to inspire us.