Simulation Free Measurement of the B^+ Lifetime using $\mathrm{B}^+ \rightarrow \overline{D^0}\pi^+$ Decays Selected Using Displaced Tracks

Sneha Malde
St-Anne's College

Thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy at the University of Oxford

Michaelmas Term, 2008

Abstract

The lifetime of the $\mathrm{B}^{\pm}$ meson is measured using the decay channel $B^{+} \rightarrow \overline{D^{0}}\pi^{+}$. The measurement is made using approximately 1.0 fb^{-1} of Tevatron proton-anti-proton collision data at $\sqrt{s}$ =1.96 TeV collected by the CDF detector. The data were collected using impact parameter based triggers that were designed to select events with a secondary vertex. The trigger selection criteria result in data rich in a variety of B hadron decays, but intrinsically bias the lifetime distribution of the collected signal events. The traditional way to compensate for the bias is to use information from simulation. Presented here is a new method for correction of the lifetime bias using an analytical technique that uses information from the data only. This eliminates measurement uncertainty due to data and simulation agreement, ultimately resulting in a smaller systematic measurement uncertainty. The $B^{\pm}$ lifetime measurement is the first measurement using this new technique and demonstrates its potential for use in future measurements. The $B^{\pm}$ lifetime is measured to be

$$\tau(B^{\pm}) = 1.662 \pm 0.023(stat) \pm 0.015(syst)ps \tag{1}$$

For my Mum and Dad

Acknowledgments

To reach this point has not been easy and it could not have happened without the help of many people.

I would like to thank my supervisor Farrukh for taking every good and bad turn with me, providing plenty of laughter and giving me my first lessons in HEP analysis. I can't thank you enough for all the support and advice over the last 4 years. To Jonas whose work formed the basis for my own thesis - thanks for handing it on to me, and the constant source of ideas and remaining upbeat when i was at my most frustrated. Thanks to Joe Boudreau for interesting conversations, physics or otherwise and for making me part of another analysis. Thanks also to Louis, Todd and Aziz for helping in this analysis. I'd also like to thank Giovanni, Diego and Sinead for helping me find the right direction to take with this analysis which has been instrumental to its completion. And finally to Nicola who is much more to me than an office mate, someone to share frustrations and ideas with, getting me settled in Chicago and showing me the city. Of all the things i have come to associate with my PhD i will miss you the most.

I spent a fantastic two years in Chicago shaped by the people who became my friends. Thanks to Phil, Nicola, Sam, Toby, Bo, Phil Rich, Mark, Sinead, Joey, Joe, Mark, Leena, Emily, Matt, Sarah, Dean, Kirby, James, Smudge, Conor, Billy, Troy, Lester , Louise, you all for the nights out, tea breaks, brunches, trips and laughs. You have given me some lovely memories. In particular I'd like to thank Dean for the ice cream, Joey for organising so much of the fun, Sam for showing me the next level of tea appreciation, Bo for being available to talk to, Conor and Lester for making shift less boring and lots of giggles, and Toby, for just being Toby, Billy for teaching me plenty and showing me healthy ways of venting frustrations.

Thanks to Nicky - for being with me through the best and worst times of the last nine years - I've always been able to go to you for advice or your opinion. Thanks to Stefanie, it was great to be able to live in the same city as you for a year and loved the time we got to spend together. To Shima and Petrina thanks for being such wonderful friends since almost as long as i can remember, i can always count on you for conversation without physics and much much more. Thanks to Jo and Vanessa - the real doctors - your gross work stories only serve to reinforce the good points about

working in Physics.

I'd like to thank my family, especially mum and dad who have borne the brunt of my moods and to Chirag for always keeping me in my place. I'm so happy to be back around you, and you have all given me the strength to become the person i am today. A special thanks to Alpa for all her advice over the years and always having a perceptive opinion on my latest problem. I'd also like to thank my huge extended family who are too numerous to list. I have always felt blessed to be surrounded by so many people who care for me.

Last but by no means least I'd like to mention Phil, for his unending support, love and encouragement. By far the best thing about my PhD has been meeting you.

Contents

3 Event Reconstruction and Selection 51

4 Lifetime Probability Distribution Function for Signal Events 66

List of Figures

List of Tables

Introduction

In 1983 the Mac [1] and MarkII [2] collaborations discovered that the tracks of prompt leptons produced in bottom decays at e^+ e^- collisions were on average displaced from the primary interaction and thus inferred that the lifetime of the B hadrons was ∼1 ps. This result was unexpected as a recent paper in 1979 had predicted the lifetime of B hadrons to be approximately an order of magnitude smaller [3]. The implications of a long B lifetime include measurable $B\bar{B}$ mixing [4] and a lower limit on the top quark mass [5]. Experimentally the long B lifetime means that it is possible to identify high energy b quark jets by observing decay vertices and this is now an important part of some Higgs and other new physics searches.

The studies of the endpoint of the B semileptonic decay spectrum [6] [7] show that the b quark couples predominantly to the c quark, and the early lifetime measurements were used to extract $|V_{cb}|$. The extraction of $|V_{cb}|$ from lifetime measurements is now dominated by theoretical uncertainty. Now the lifetimes of individual B hadron species are used to test the Heavy Quark Expansion [8], look for evidence of new physics in the $B_s - \overline{B_s}$ system and form an important part of many time dependent analyses.

Outline of this thesis

The long lifetime of B hadrons can be used to select B-decays in hadron collider experiments as the large impact parameters of the daughter tracks can be used to select events rich in B-hadrons. The selection of events in this way will bias the lifetime distribution of these hadrons. The traditional method of accounting for this bias is to determine the aggregate effect of the impact parameter based selection on the lifetimes through detailed simulation. This thesis presents a new method in which the bias is accounted for by exact calculation. Among the advantages of the method presented in this thesis is that the systematic uncertainties are much smaller. The method has been tested and has used CDF data to make a measurement of the $B^{\pm}$ hadron using this new technique.

Chapter 1 describes the context and theoretical motivation for the measurement. Chapter 2 describes the Fermilab accelerator and the CDF detector used to collect the data used in this thesis. In chapter 3 there is a description of the how the data used for the measurement is selected. In chapters 4 and 5 there is the derivation of the probability density function(PDF) and all the parameterizations of data variables used to make the measurement. The validity of the method is discussed in chapter 6 and the fast simulation used to do this is described. The fit to data is presented in chapter 7 and is followed in chapter 8 by a discussion of the sources of systematic uncertainty. The thesis ends with the result of the measurement and concluding remarks.

Chapter 1

Theoretical Motivation

This chapter gives an overview of the Standard Model and concentrates on the theoretical predictions of B hadron lifetimes. Towards the end of the chapter there is an overview of current experimental measurements of B hadron lifetimes

1.1 The Standard Model of Particle Physics

The Standard Model of particle physics is one of the most successful theories in describing the particles in the universe and their interactions, and there has been good agreement between empirical observations and theoretical predictions. A series of theoretical developments and experimental measurements in the 1960's and 1970's led to the combined theory that describes 3 of the 4 forces of nature. In the Standard Model there are three types of fundamental particles: quarks, leptons and the force carriers [9]. The force carriers have integer spin 1 and are called gauge bosons, while the quarks and leptons have spin 1/2 and are collectively known as fermions. The fermions interact with each other through the exchange of the force carriers.

1.1.1 The Fundamental Particles and Interactions

The masses, charges and names of the fundamental fermions are given in table 1.1.1. The quarks can be ordered into three generations where each pair in the generation have similar properties to another generation except for their mass. The leptons can be similarly ordered. Each particle has an anti-particle, which is a particle that behaves and has the same mass as its corresponding particle but has opposite electric charge and opposite direction of magnetic moment.

Free quarks have not been observed, and the quarks come in bound states. In addition to electric charge, quarks also carry a color charge that comes in three colors: red, green, blue (and the corresponding anti-colors). In the bound state the object must be colorless, which means either a particle has three quarks, one with each color, or a quark and anti-quark pair where a particular color and anti-color is present [10]. Three-quarks states are called baryons, and the quark anti-quark state is a meson. Baryons represent the bulk of matter that we observe in our universe, as the nuclei of atoms contain protons and neutrons, which are baryons composed of the first generation quarks. Leptons, unlike quarks, do exist as free particles, and the first generation lepton, the electron, is also abundant in the universe as it is a constituent of all atoms.

In theoretical physics, interactions are dictated by symmetry principles. The Standard Model is described by a quantum field theory that is invariant under the local gauge transformations in $SU(3) \otimes SU(2) \otimes U(1)$. The three gauge symmetries encompassed in the transformation give rise to the three types of interaction described by the Standard Model; electromagnetic, weak and strong [12].

The strong force is mediated by the gluon and keeps the quarks together in hadrons,

Quark Properties			
Generation	Symbol	Charge	Mass, MeV
I	u	$+\frac{2}{3}$	1.5-3.3
	d	$-\frac{1}{3}$	3.5-6.0
II	c	$+\frac{2}{3}$	1270^{+70}_{-110}
	s	$-\frac{1}{3}$	104^{+26}_{-34}
III	t	$+\frac{2}{3}$	$(171.2\pm2.1)\times10^3$
	b	$-\frac{1}{3}$	$(4.2^{+0.17}_{-0.07})\times10^3$
Lepton Properties			
I	e	-1	0.51
	ν_e	0	<0.002, 95% C.L
II	μ	-1	105.7
	ν_μ	0	<0.19, 90% C.L
III	τ	-1	1776.8±0.2
	ν_τ	0	<18.2, 95% C.L

Table 1.1: The properties of quarks and leptons [11]. Only an upper limit on the neutrino masses can be set.

and the hadrons together in nuclei. The confinement of quarks in colorless states shows that the strong force is short range, of order the size of a nucleus. There are 8 gluons, each carrying a different color and anti-color charge. As the leptons do not carry any color charge, the gluons do not couple to them. The $SU(3)$ symmetry gives rise to quantum chromodynamics (QCD) which is the quantum field theory behind the strong force.

The electromagnetic force is mediated by the photon which couples to all charged particles. The electromagnetic force binds electrons to nuclei and is therefore the basis of chemistry and is the most common interaction observed in everyday life. The range of the force is infinite although its strength drops as $\frac{1}{r^2}$.

The weak nuclear force covers all interactions which change quark flavors, and the force is felt by quarks and leptons. The force carriers are the massive W and Z bosons, and it is a very short range force, smaller than the size of a nucleus. The Glashow-Salam-Weinberg [13] [14] model unifies the electromagnetic and weak forces within the same theoretical framework via the $SU(2) \otimes SU(1)$ symmetry. A summary of the properties of the force carriers is given in table 1.1.1.

Gauge Boson Properties			
Force	Carrier	Mass, GeV	Relative Strength
Strong	Gluon (g)	0	1
Electromagnetic	Photon (γ)	$<1\times10^{-27}$	10^{-2}
Weak	$W^{\pm}$ Z^0	80.398 ± 0.025 91.1876 ± 0.0021	10^{-7}

Table 1.2: The properties of the gauge bosons [11].

1.1.2 The Weak Couplings of Quarks

The coupling of the W boson to leptons measured through muon decay and to quarks measured though nuclear β decay are observed to be slightly different. This leads to the suggestion that universality between the weak couplings to quarks and leptons could be achieved if the weak interaction quark eigenstates are not the same as the mass or strong eigenstates [15]. Even before the discovery of the charm, bottom or top quark, this rotation between weak and strong eigenstates had been formulated into the Cabibbo-Kobayashi-Maskawa mixing matrix, where the strong and weak(primed) eigenstates are related by [16]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.1}$$

Using this matrix the transition amplitude from a b-quark to a c-quark is V_{cb}. A common parameterization of the CKM matrix is the Wolfenstein parameterization that takes into account the order of magnitude differences between elements of the matrix in terms of $\lambda \equiv \sin(\theta_C) = 0.22$ where θ_C is the Cabibbo angle. The parameterization up to $\mathcal{O}(\lambda^3)$ is also shown. The parameters A, ρ, η are all real and of order 1. The strongest couplings are referred to as "Cabibbo allowed" and the weaker ones as "Cabibbo suppressed". For example, a decay of a W to a u and d quark is Cabibbo allowed, while the decay of a W to a u and s quark is the Cabibbo suppressed decay. The Standard Model does not predict any of the couplings themselves; it only states that the matrix must be unitary.

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(\rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.2}$$

Current experimental determinations of the values of the matrix are given in equation 1.3. From the experimental precision currently available there is no indication that this matrix is not unitary. The values given in equation 1.3 are taken from [11].

$$\begin{pmatrix} 0.97419 \pm .0.00022 & 0.2257 \pm 0.0010 & 0.00359 \pm 0.00016 \\ 0.2256 \pm 0.0010 & 0.97334 \pm 0.00023 & 0.0415^{+0.0010}_{-0.0011} \\ 0.00874^{+0.00026}_{-0.00037} & 0.0407 \pm 0.0010 & 0.999133^{+0.000044}_{-0.000043} \end{pmatrix} \tag{1.3}$$

1.2 The Theory of B Hadron Lifetimes

B hadrons are unstable and decay to other particles through the weak interaction. The lifetime of a particle is determined by the sum of the different interaction processes that can occur to various final states. In the next three sections the theory of B hadron lifetimes is described. This starts with the spectator model and progresses to a discussion of phenomenological effects that will cause differences between the B hadron species. It ends with a description of the Heavy Quark Expansion (HQE), the analytical tool used to predict the lifetime of B hadrons in comparison to the B^0 lifetime.

1.2.1 The Spectator Model

The simplest theory of B hadron lifetimes is the spectator model. In this theory the b-quark is much heavier than the other quarks and its mass becomes irrelevant to the non perturbative dynamics of QCD, i.e, the assumption is made that the interactions between the heavy quark and the lighter valence and sea quarks can be ignored [17].

In the case of the B^+ meson, the u quark is assumed not to interact with the b quark, and so the Feynman diagram for the weak decay of the b quark can be compared to the diagram for the weak decay of the muon. The Feynman diagrams are shown in figure 1.1. The Feynman diagrams can be used to calculate the particle width which is inversely proportional to the the particle lifetime.

The muon width has been analytically calculated and for the first order process is given in equation 1.4,

$$\Gamma_{\mu} = \frac{m_{\mu}^5 G_F^2}{192\pi^3} \tag{1.4}$$

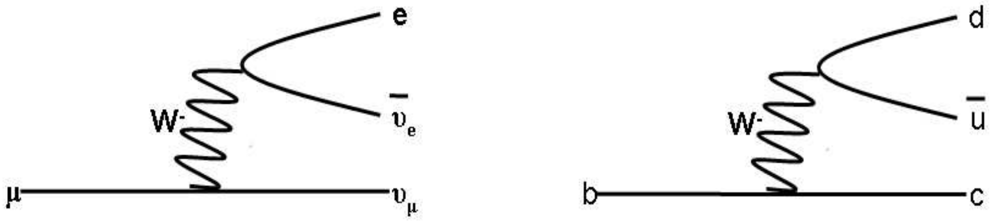


Figure 1.1: diagram showing the decay of a muon and the decay of a free b quark

where m_μ is the muon mass and G_F is the Fermi point coupling. This is the strength of the four fermion interaction, which was the old explanation for the weak force. In the modern theory G_F is defined as

$$G_F = \frac{\sqrt{2}}{8}\frac{g^2}{m_W^2} \tag{1.5}$$

where g is the coupling constant of the weak interaction and m_W is the mass of the W boson. Precision measurements of the muon lifetime have determined the value of G_F.

Assuming the spectator model, the calculation of the Feynman diagram of the B meson width and the muon width are very similar. The changes to be made are to replace the muon mass with the b quark mass. A factor of $|V_{cb}|^2$ will also arise, as this accounts for the difference in coupling between a b and c quark and muon and neutrino. (The b quark can also decay to a u quark although this occurs infrequently and is ignored in the current discussion.) Phase space must also be considered. The muon is the lightest unstable particle and can only decay into an electron and neutrinos. The phase space for the b-quark decay is approximately 9 times larger as the W can decay to any of the three lepton generations and the two quark generations ud' and cs'. The phase space for quarks is multiplied by a factor of three to take into account the different color charges possible. In the spectator model the b meson lifetime is given by

$$\Gamma_b = \frac{m_b^5 G_F^2 |V_{cb}|^2}{192\pi^3} \times 9 \tag{1.6}$$

where m_b is the b quark mass.

Comparing the two equations 1.4 and 1.6, it is possible to predict the B hadron lifetime in terms of the muon lifetime, which is measured as 2.2 μs [11].

$$\tau_b = \tau_\mu \times (\frac{m_\mu}{m_b})^5 \times \frac{1}{9|V_{cb}|^2} \tag{1.7}$$

Using 4.2 GeV [11] as the mass of the b quark and $|V_{cb}| = 0.0404 - 0.0425$, the lifetime of B hadrons is estimated between 1.3 and 1.5 ps, which is of the order of observed lifetimes.

In the spectator model all the b hadrons would have the same lifetime, as the lifetime would depend only on the b-quark mass. The theory can also be extended to the charm sector where larger statistic samples were available at the time of the initial postulation of the spectator model. However, the ratio $\frac{\tau_{D^+}}{\tau_{D^0}} \sim 2.5$ clearly showed that the interactions of the light quark cannot be ignored.

1.2.2 Spectator Effects

If the B$^-$ and B^0 meson are considered, it is possible to see how the light quarks contribute to give the B$^-$ a longer lifetime in comparison to B^0.

The diagram in figure 1.2 shows the effect of Pauli interference. The B$^-$ meson has two ways to decay to the same final state. There is interference between these two diagrams. This interference turns out to be destructive when calculated and therefore leads to a shortening of the width. The same situation for the B^0 shows two different

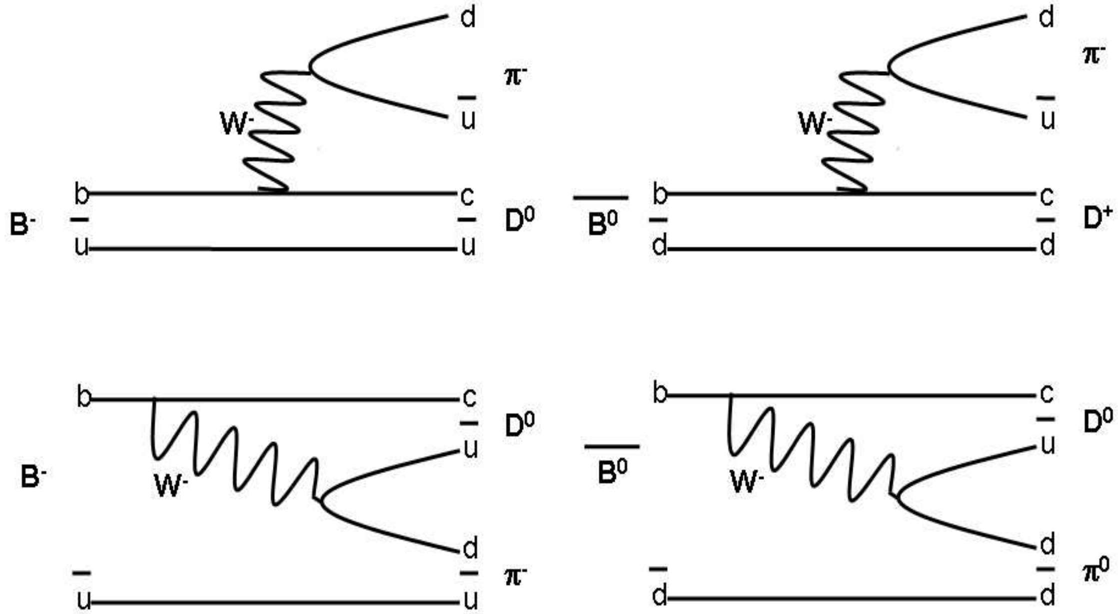


Figure 1.2: This diagram shows the effect of Pauli interference. The B^0 decay has two different final states while the B^- decay has the same final states, and the two processes destructively interfere.

final states that do not interfere. This leads to a lengthening of the B^- lifetime in relation to the B^0.

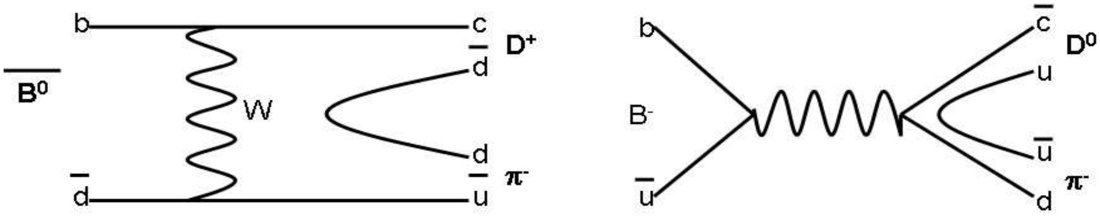


Figure 1.3: This diagram shows the effect of Weak interference or annihilation. It can proceed at the Cabibbo allowed level for the B^0 decay but only at Cabibbo suppressed level for the B^- decay.

The diagram in figure 1.3 shows the weak exchange Feynman diagrams that the B mesons can have. The weak exchange occurs at the Cabibbo allowed level for the

B^0 but only at Cabibbo suppressed level for the B^- meson, and so this also lengthens the B^- lifetime in comparison to the B^0. The Pauli interference is expected to be a more dominant effect. Finally, consideration of the weak exchange in meson and baryon systems leads to the inference that baryon lifetimes must be slightly shorter. This is because weak exchange in mesons is helicity suppressed, but this is not the case in baryons where there are two light spectator quarks. Thus these phenomenological considerations lead to the following prediction of the hierarchy of B hadron lifetimes:

$$\tau(B^-) > \tau(B^0)\ \tau(B_s) > \tau(\Lambda_b) \tag{1.8}$$

1.2.3 Heavy Quark Expansion

Phenomenological considerations can only predict the lifetime hierarchy. For quantitative predictions a more rigourous approach is required. The analytical calculation of the lifetime difference between different mesons and baryons is done in the framework of the Heavy Quark Expansion (HQE). This is a mature theory based on QCD. The expansion can be expressed through an operator product expansion (OPE) [18] and in terms of Λ_{QCD}/m_b, where Λ_{QCD} is the QCD confinement scale. Λ_{QCD} ≈400 MeV is the typical energy where QCD becomes non perturbative. The expression is rapidly convergent for B hadrons and is given by:

$$\Gamma = \frac{G_F^2 m_b^5}{192\pi^3} \sum_i (|V_{q_i b}|^2 \cdot [A_{0_i} + A_{2_i}(\frac{\Lambda_{QCD}}{m_b})^2 + A_{3_i}(\frac{\Lambda_{QCD}}{m_b})^3]) \tag{1.9}$$

where the constants A_0, A_2, A_3 can be written in terms of Wilson coefficients and the expectation values of local operators. Wilson coefficients can be calculated by perturbation theory, and the expectation values of local operators can be inferred from symmetry arguments, QCD sum rules, lattice studies and quark models [8]. The sum over quarks includes the possibility of the b quark decay to a u quark.

At lowest order in $\frac{\Lambda_{QCD}}{m_b}$, the expression returns to the spectator model and the lifetime for all B hadrons is the same. The first splitting between hadrons occurs at order $(\frac{\Lambda_{QCD}}{m_b})^2$, which describes the helicity suppression in mesons. It is at this order that differences between the baryons and mesons appear. There are no differences in meson lifetimes at this order. At order $(\frac{\Lambda_{QCD}}{m_b})^3$, the differences describe the Pauli interaction and the weak interference in mesons. For baryons this third order correction turns out to be larger than the second order calculation due to phase space considerations.

The HQE predicts the ratio of the B hadron species with respect to the B^0 lifetime, and these are given in table 1.2.3.

Lifetime Ratio	Theoretical prediction	Experimental Average (PDG 2008)
$\tau(B^+)/\tau(B^0)$	1.06 ± 0.02	1.071 ± 0.009
$\tau(B_s)/\tau(B^0)$	1.00 ± 0.01	0.939 ± 0.021
$\tau(\Lambda_b)/\tau(B^0)$	0.88±0.05	0.904 ± 0.032

Table 1.3: The theoretical predictions of the lifetime ratios of B hadrons. The theoretical predictions are from [19], [20] and [21]. The experimental averages are from [11] current at 15th Jan 2008 and are discussed further in section 1.4.

1.3 Motivation for Lifetime Measurement

One of the most important reasons for making lifetime measurements is to test the Heavy Quark Expansion. This theory is rooted in QCD, and lifetimes are a good place to test the theoretical and experimental agreement. Precise measurement is therefore very important to detect small disagreements between the theoretical prediction and experimental measurement. Precision is achieved by using samples containing large signal yields to reduce statistical error and a method of measurement that will give a small systematic uncertainty in the measurement. Lifetimes in the B_s system are also interesting as they can be used to measure the lifetime difference between the two CP eigenstates. In flavor specific decay modes, such as $B_s \rightarrow D_s\pi$, the lifetime is used to measure the mean lifetime, and decay modes such as $B_s \rightarrow KK$ or $B_s \rightarrow D_sD_s$ are almost entirely CP even and can be used to measure the lifetime difference between the 2 B_s lifetimes.

The CDF hadronic trigger and similar triggers at LHC-b can collect a range of B hadron decays. After large amounts of data become available, the data from these triggers can be used to make a measurement with a small statistical error. However these triggers will select preferentially on B hadrons that have lived longer, and this results in a bias in the lifetime distribution of the selected events. This can be corrected through the use of information from simulation. However, this can result in large systematic uncertainty which depends on how well the simulated data and simulated performance of the trigger match the actual data. This thesis presents an alternate method to correct for the selection bias and uses an analytical approach that takes information from only the data, with no use of information from simulation. The aim of this new technique is to have a significantly smaller measurement uncertainty. This method can be applied to data at CDF and the data that will be collected at LHC-b. However, before using this method to improve measurement on the B_s, Λ_b or even a make a measurement of the Σ_b lifetime, the method must be shown to work. To do this it is best to measure a lifetime in a high statistics decay mode, compare it with the world average, and carry out extensive studies to demonstrate the robustness of the method. This is the work contained in this thesis. The B^+ lifetime has been chosen as there is a large sample available at CDF. The motivation behind this is primarily a demonstration of the validity of the method, as well as the physics result obtained.

1.4 Current Experimental Measurements

A selection of the latest measurements of the charged and neutral B meson are shown in figures 1.4 and 1.5.(The CDF preliminary result highlighted in red is not included in the world average in the figures. These measurements come from analysis of B decay reconstructed in decay modes J/ΨX, where X is either K^+, K^* or K_{short}.) The most accurate measurements have come from the B factories due to the large numbers of decays available. At the B factories, a number of different decay channels have been used to measure lifetimes and the results combined. Their measurements are now limited by the systematic error.

Lifetime measurements of the B_s meson are from the Tevatron and LEP and are shown in figure 1.6. These measurements all come from flavor specific modes which yields the average B_s lifetime.

Lifetime measurements of the Λ_b baryon have raised much interest in recent years. For a long time the experimental measurements resulted in a ratio of $\tau(\Lambda_b)/\tau(B^0)$ lower than the theoretical prediction. In 2006, the CDF measurement using the $J/\Psi\Lambda$ decay mode measured a lifetime that was 3σ higher than the world average [24]. In 2008, the CDF preliminary result using the decay mode $\Lambda_c\pi$ measured a lifetime that falls between the other CDF result and the older measurements, and produces a lifetime ratio that is consistent with the theoretical predictions [25]. Further measurements, such as those achievable by the method in this thesis are awaited. Selected lifetime measurements of the Λ_b are given in figure 1.7.

To date there are no lifetime measurements of the Σ_b or the recently discovered Ω_b [11]. The lifetime of a mixture of Ξ_b^0 and Ξ_b^- is measured to be $1.42^{+0.28}_{-0.24}$ ps [11]. Once a sufficient sample size is obtained, measurements of all of these hadrons can be

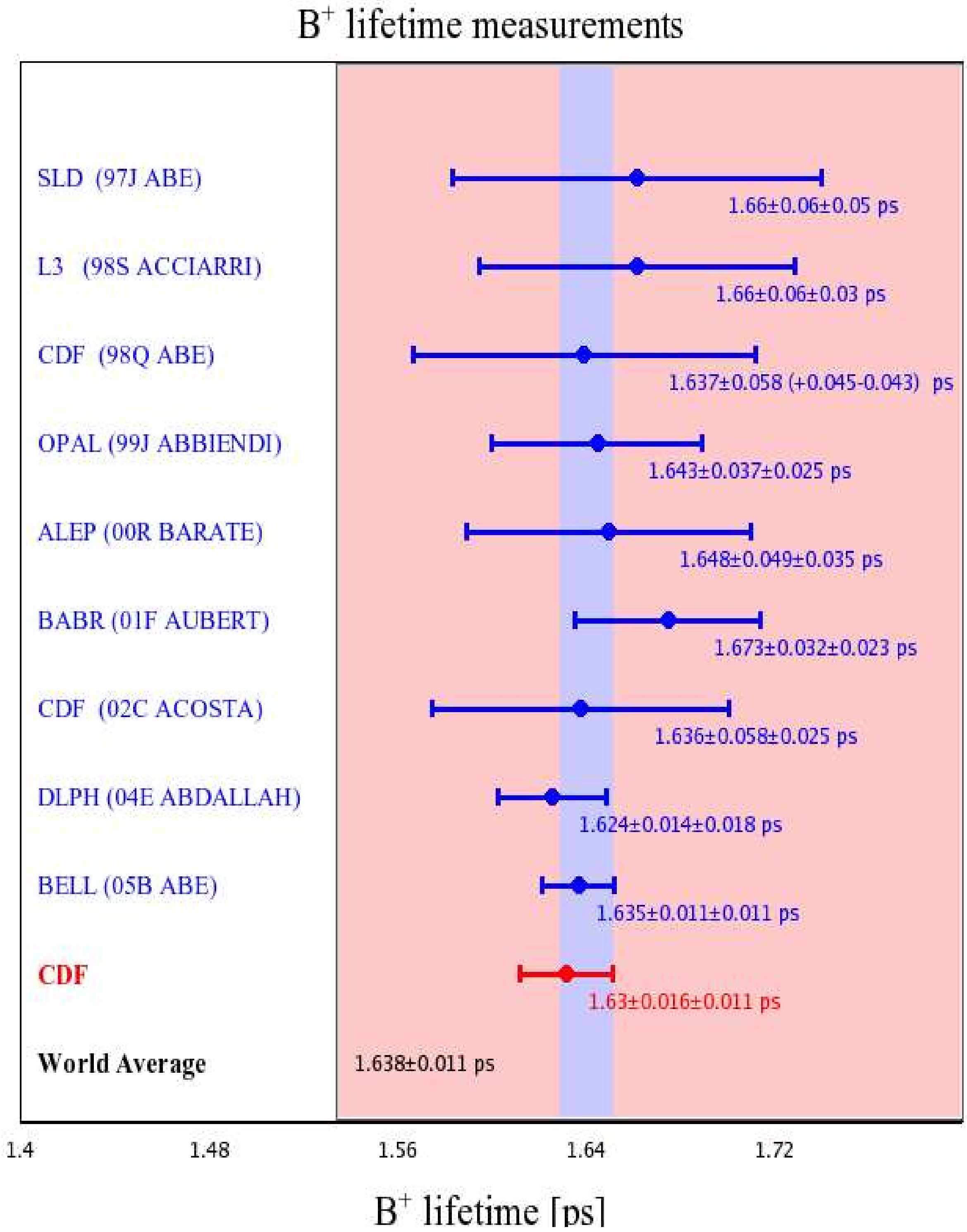


Figure 1.4: Selected lifetime measurements of the charged B meson [22]. The CDF preliminary result highlighted in red is not included in the world average in the above figure. This is the recent CDF measurement of the lifetime using the decay mode $B^- \to J/\Psi K^-$.

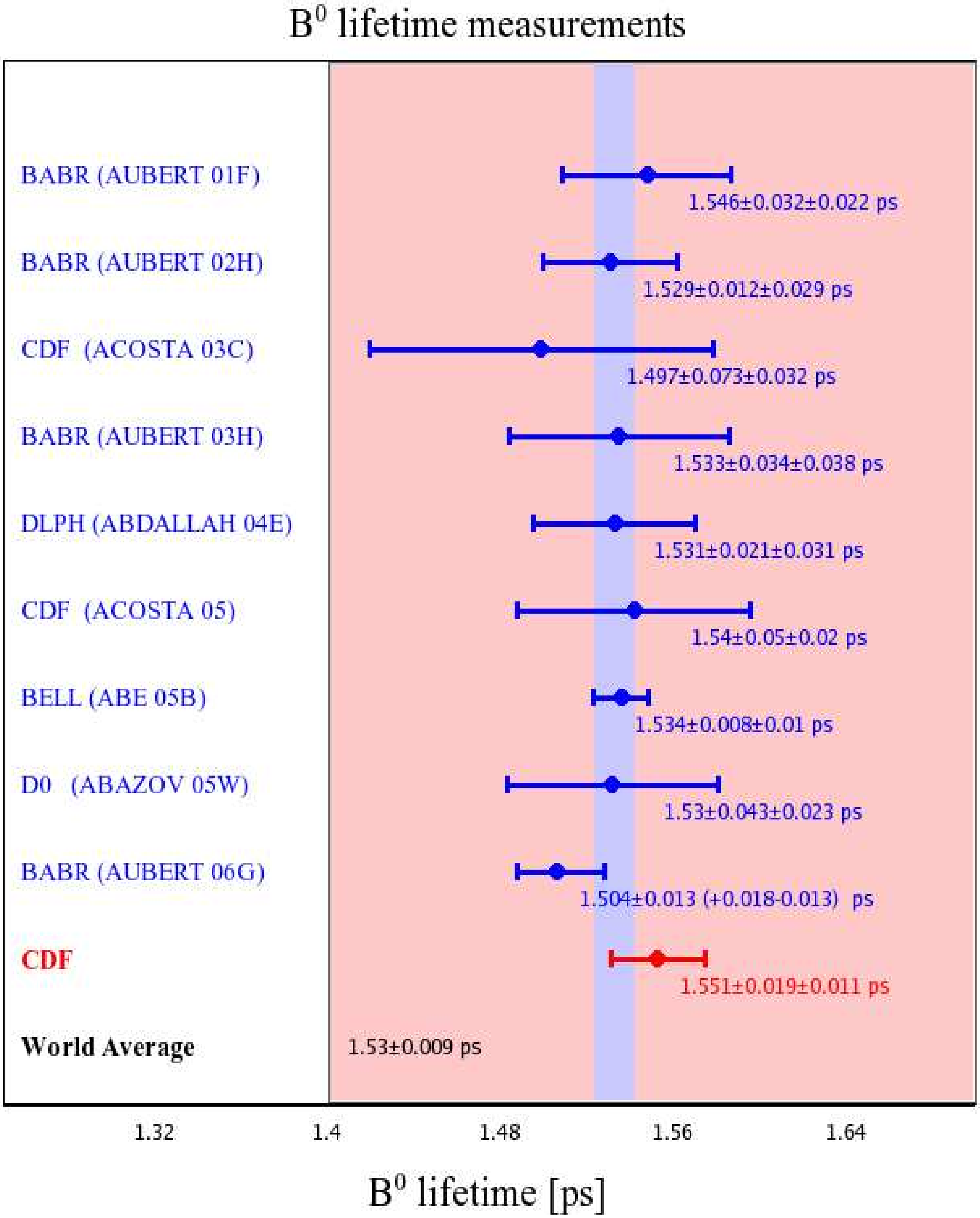


Figure 1.5: Selected lifetime measurements of the neutral B meson [22]. The CDF preliminary result highlighted in red is not included in the world average in the above figure. This is the recent CDF measurement of the lifetime using the decay mode $B^0 \rightarrow J/\Psi K^{0*}$ and $B^0 \rightarrow J/\Psi K_{short}$.

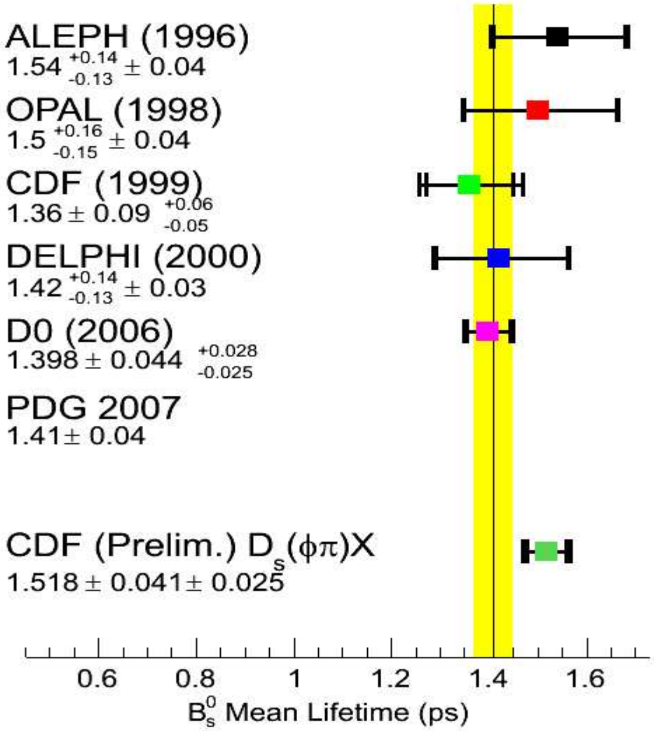


Figure 1.6: Selected lifetime measurements of the neutral B_s meson [23]. The CDF preliminary result is not included in the world average in the above figure.

carried out using the technique presented in this thesis. Theoretical predictions will also become more precise in the future, and these baryons will provide a rich testing ground for the HQE.

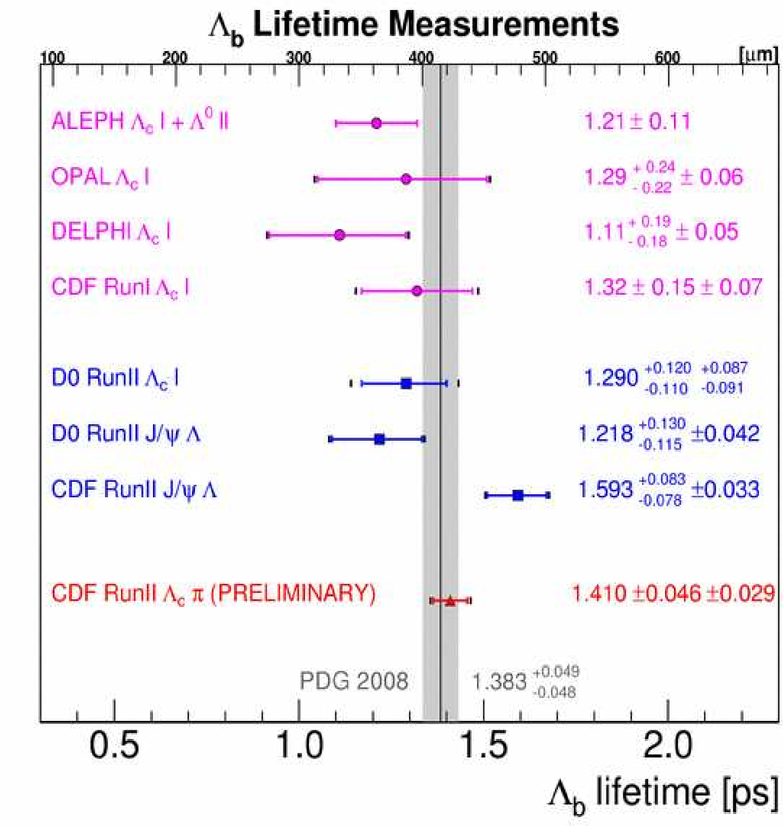


Figure 1.7: Selected lifetime measurements of the Λ_b baryon [26]. The CDF preliminary result highlighted in red is not included in the world average in the above figure.

Chapter 2

Accelerator and Detector

The data used in this thesis was collected using the Collider Detector at Fermilab (CDF) detector at the Tevatron. The Tevatron is currently the world's highest energy superconducting synchrotron. Protons and anti-protons accelerated to 980 GeV each are collided with each other at two collision points, one of which is the location of the CDF detector and the other is at the location of the D0 detector site. The following chapter gives a description of the accelerator complex that produces these collisions and the CDF detector that records these collisions.

2.1 The Tevatron Accelerator

A full description of all the components used to create the $p\overline{p}$ collisions can be found in [27]. A diagram of the Fermilab Accelerator Chain is shown in Figure 2.1. The following sections give an overview on how the protons and anti-protons are produced and accelerated.

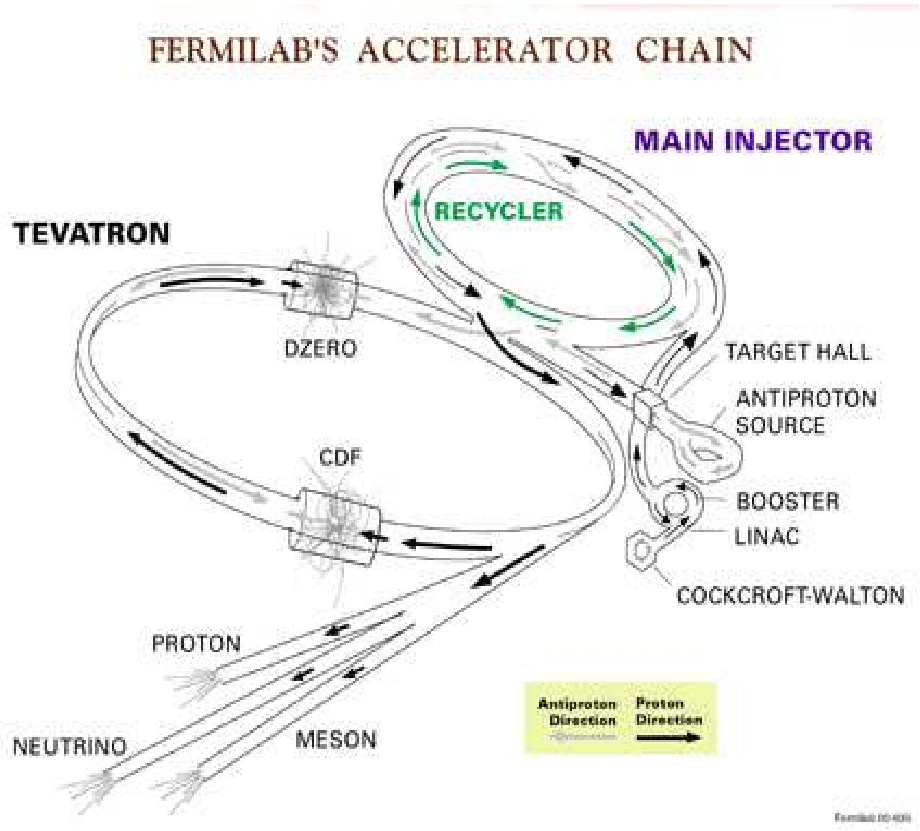


Figure 2.1: The Accelerator chain at Fermilab [28].

2.1.1 Proton Production

The protons are accelerated to 980 GeV in a multistage process. The process starts with hydrogen gas, ionized by electrical discharges to form H$^-$ ions. The Cockcroft Walton pre accelerator accelerates the ions to 750 KeV using a diode capacitor voltage multiplier. The H$^-$ ions are separated from the other particles using a magnetic field. The process of separation and accumulation occurs every 66 ms which segments the beam into bunches. From here the H$^-$ bunches are moved to the Linac, a 150 m linear accelerator that receives the H$^-$ ions. Using radio frequency (RF) cavities the ions are accelerated to 400 MeV. The beam is also focused in the plane transverse to the direction of their motion using quadrapole magnets. The ions are passed through a thin carbon foil which strips them of their electrons leaving a beam of protons which is injected into the booster, a 75 m radius synchrotron. Using 18 RF cavities the protons reach an energy of 8 GeV and are grouped into 84 bunches each carrying approximately 6×10^6 protons.

From the booster the proton bunches are moved to the Main Injector, a circular synchrotron of 1 km in diameter, which can accelerate the protons to 120 GeV or 150 GeV. The 150 GeV protons are used for injection into the Tevatron and the 120GeV protons are used to create anti-protons.

2.1.2 Anti-proton Production and Accumulation

More complex than proton production is anti-proton production and accumulation, as there is no ready source of anti-protons. To create anti-protons, 120GeV protons from the Main Injector are collided with a fixed nickel alloy target. Through the interaction $p + p \rightarrow p + p + X$, many types of secondary particles are produced including anti-protons. Approximately 20 anti-protons are produced for every 10^6 protons incident on the nickel target. The secondary particles are focused into a parallel beam using a cylindrical lithium lens. Then a pulsed dipole magnet is used to separate particles of negative charge and momentum of approximately 8 GeV. At this stage there is a bunched beam of anti protons with a large momentum spread. The beam is transferred to the debuncher, a triangular synchrotron, to make a continuous beam. By applying the technique of bunch rotation [27] the beam is turned from one with narrow time distribution and large energy spread to one of narrow energy spread and wide time distribution, i.e., a monochromatic, continuous beam. The narrow energy spread is essential to maintain efficient transfer of the beam to subsequent accelerators. Stochastic cooling processes [29] are applied to cool the transverse energy, which reduces the spread in the momentum spectrum. Anti-protons are stored in the accumulator, another 8 GeV synchrotron housed in the same tunnel as the debuncher. Periodically anti-protons are moved from the accumulator to the recycler, which can store 5×10^{12} anti-protons. The recycler is housed in the same tunnel as the main injector. The anti-protons are further cooled in the recycler using stochastic cooling and electron cooling [30]. It takes between 10-20 hours to store enough anti-protons for collisions in the Tevatron. The anti-protons are moved to the Main Injector for acceleration to 150 GeV before entering the Tevatron.

2.1.3 Tevatron Injection and Luminosity

The Tevatron ring is a 1 km radius superconducting synchrotron. It houses 774 dipole magnets for steering and 216 quadrapole magnets for beam focusing. The magnets are made of Nb-Sn and produce a peak field of 4.5 T which allows the particles to be accelerated from their 150 GeV injection energy to 980 GeV, the energy used for collisions. The protons and the anti-protons share the same beam pipe and each is forced into its own helical closed orbit. The result is two strands that have sufficient transverse separation to prevent collisions outside the desired interaction point. There are 36 bunches each of protons and anti-protons. The protons are injected first. They are accelerated to 150GeV in the Main Injector and then coalesced into a single bunch and injected into the Tevatron. Each of the 36 bunches contain 3×10^{11} protons. The 36 bunches are separated into 3 trains, each train containing 12 bunches, and within them the bunches are separated by 396 ns. There are abort gaps between the trains. To abort the beam a special magnet is turned on which will divert the beam into a designated target. The abort gap allows the special magnet to reach the correct field without any of the beam being affected while the field is increasing. Once the protons are injected anti-protons are taken from the recycler and accumulator and accelerated to 150 GeV in the Main Injector. They are injected in sets of 4 bunches until there are 36 each of about 6×10^{10} anti-protons. Their spacing mirrors that of the protons. High luminosity collisions take place at the 2 designated interaction points (the CDF and D0 experiments) 1.7 million times a second.

The collider performance is measured by the luminosity. The luminosity is given by

$$L = \frac{N_p N_{\overline{p}} B f_0}{4\pi\sigma^2} \tag{2.1}$$

where N is the number of protons or anti-protons in a bunch, B is the number of bunches, f_0 is the collision frequency and σ is the area of the beam cross section. Therefore to increase the luminosity of a collider either the beam width must be reduced or the number of colliding particles increased. The peak instantaneous luminosity is shown over the period of Run II in Figure 2.2. Run II is the data taking period that began in 2001 and is expected to continue until the end of 2010. The large increase in peak luminosity is due to improvements in the storage and and transfer of anti-protons. This remains the limiting factor in reaching further luminosity milestones.

For any given store the luminosity decreases exponentially over time as particles are lost in collisions and beam interactions. In good running conditions the store is removed after 15-20 hrs and a new set of protons and anti protons which have accumulated in the other parts of the accelerator in that time are injected. The integrated luminosity per week and over Run II is shown in Figure 2.3.

At the time of writing the Tevatron has delivered 4.93 fb^{-1} of data, of which 4.04 fb^{-1} has been recorded to tape. There is a natural inefficiency in collecting data due to deadtime and failures. The data in this thesis corresponds to the first 1.0 fb^{-1} of Run II data which was collected between February 2002 and March 2006.

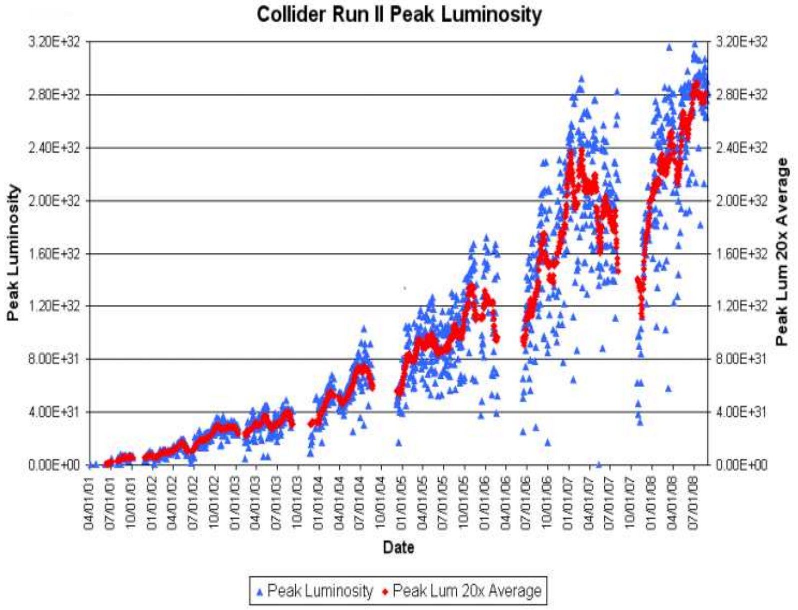


Figure 2.2: The peak luminosity of a store over the period of Run II data [31].

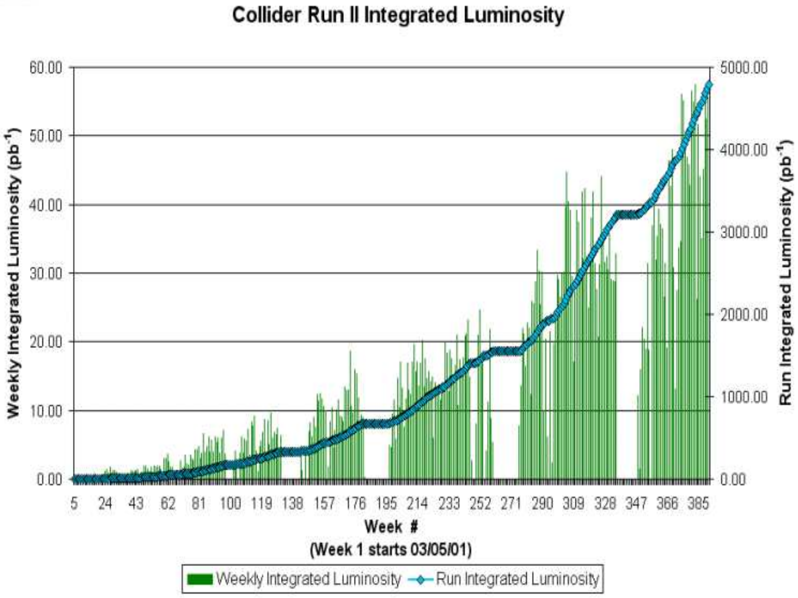


Figure 2.3: The integrated luminosity of the Tevatron [31].

2.2 The CDF Detector

2.2.1 Overview

The CDF Run II detector is a large forward-backward symmetric detector situated at one of the collision points on the Tevatron ring. The detector is designed to be multi-purpose so that it can be used for a range of high energy physics analyses. The data from this detector has been used for precision measurements of W mass and B_s mixing as well as searches for Higgs, Super Symmetry and other New Physics. To be useful for such a wide range of physics the detector has many components to measure momentum and energy of particles produced in the collisions. A description of the CDF Run II detector can be found at [32]. An elevation of the CDF detector is shown in Figure 2.4. From inside to out, radially, the detector components are the Silicon Tracker, Drift Chamber Tracker, Time of Flight Detector, the Solenoid, the calorimeters and the muon chambers. There are also luminosity counters at the forward and backward ends of the detector, close to the beam pipe.

Not all the detector components are used to analyse the data in this thesis. The tracking systems which are used extensively are described in more detail. The components pertinent to this analysis are the Silicon vertex detector (SVX) and the Central outer tracker(COT). The silicon vertex tracker(SVT), the extremely fast tracker(XFT) and the three levels of the trigger system were also used.

2.2.2 The Coordinate System

Three coordinate systems can be used when describing and using the CDF-II detector. The origin of all 3 is defined at the B0 beam interaction point, which is geometrically

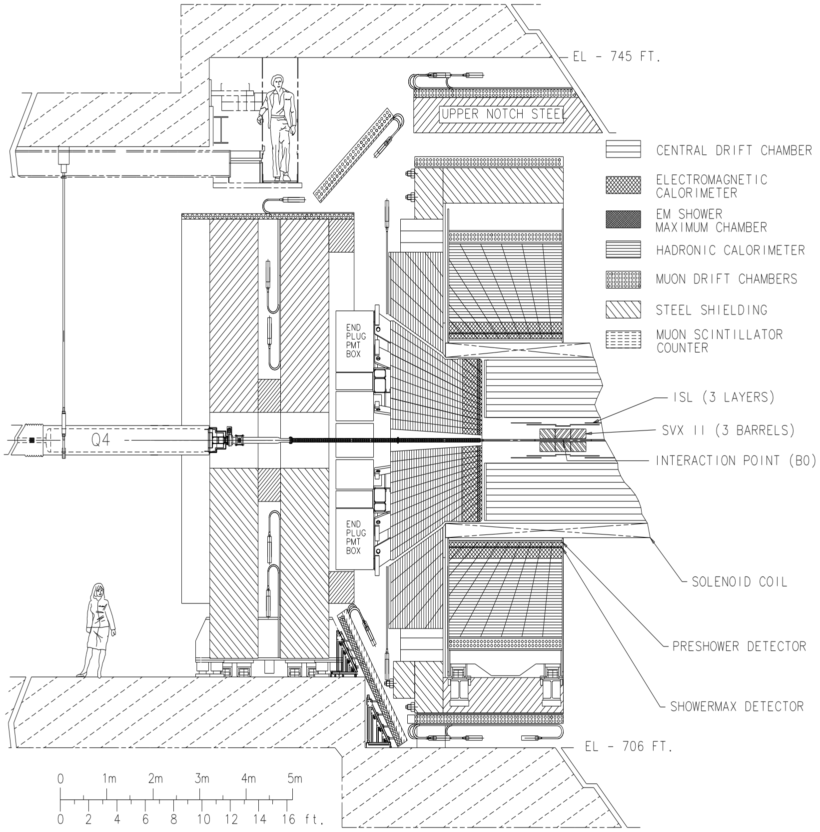


Figure 2.4: An elevation of the CDF Run II detector. Only one half is shown [32].

close to the center of the CDF detector. In the Cartesian coordinate system the z axis is parallel to the beam line with the increasing positive axis in the direction of proton circulation. The y axis is defined as vertically upwards and the x axis is radially outwards from the Tevatron center. This is shown in Figure 2.5.

The x-y plane is referred to as the transverse plane. It can also be useful to define coordinates using cylindrical polar or spherical polar coordinates. In both cases the $r-\phi$ plane is transverse. The angle θ is the polar angle relative to the z axis. Particle production is constant over the variable defined as rapidity, Y, in Equation 2.2. In the case where the mass of the particle can be ignored in relation to its momentum the expression simplifies to η, the pseudo rapidity which is a function of the polar angle

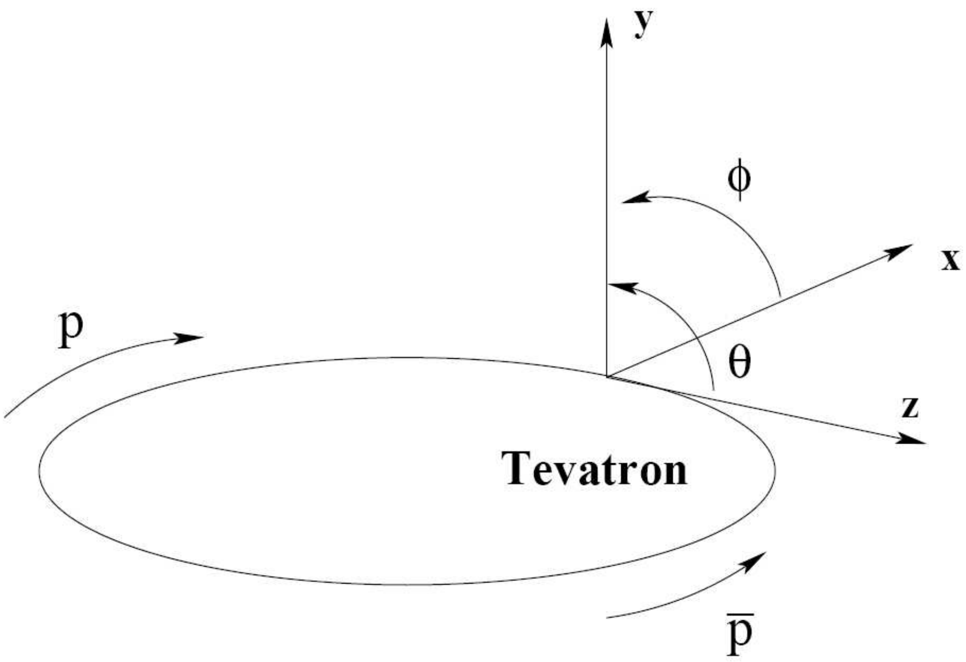


Figure 2.5: The Coordinate system

alone as defined in Equation 2.3.

$$Y = \frac{1}{2} ln \frac{(E + P_z)}{(E - P_z)} \tag{2.2}$$

$$\eta = -ln(tan(\frac{\theta}{2})) \tag{2.3}$$

2.2.3 The Silicon Tracker

The innermost component of the CDF-II detector is the silicon microstrip vertex detector (SVX) [33]. This component allows for accurate measurement of tracks close to the beam line which is necessary to get the best possible impact parameter and secondary vertex resolution. This information is crucial not only to B meson lifetime analyses but also used for other work such as b-tagging jets used for Higgs and Top

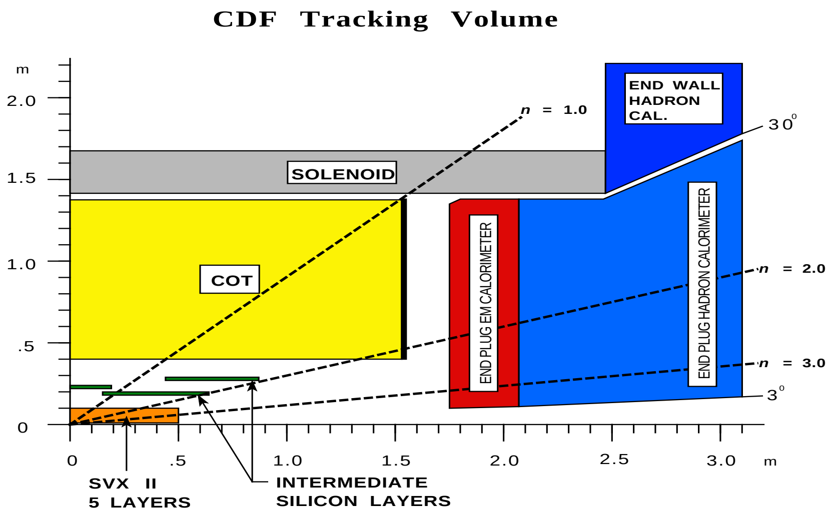


Figure 2.6: The CDF detector showing the tracking volume in η. The muon chambers are not shown in this diagram. This is a side on schematic in the y-z plane. Only one quadrant is shown, the detector is rotationally symmetric about the z axis and symmetric with a reflection in the y axis.

studies. Measurements in the silicon detector also improve the momentum resolution.

Silicon micro-strip detectors are an ideal choice to perform the task required. Silicon is able to withstand the higher radiation dose found closer to the beam pipe. The silicon is a small band gap semiconductor which is ideal for providing fast electronic readout and can be finely segmented to give high precision position measurements. The silicon strip is a reverse biased p-n junction. When a charged particle passes through it, it leaves an electron-hole pair. Electrons drift towards the n side while holes drift to the p side. The charge deposition will be read out on one or more strips and the localized signal gives $r - \phi$ information.

There are 3 different silicon detectors and they are shown in Figure 2.7.

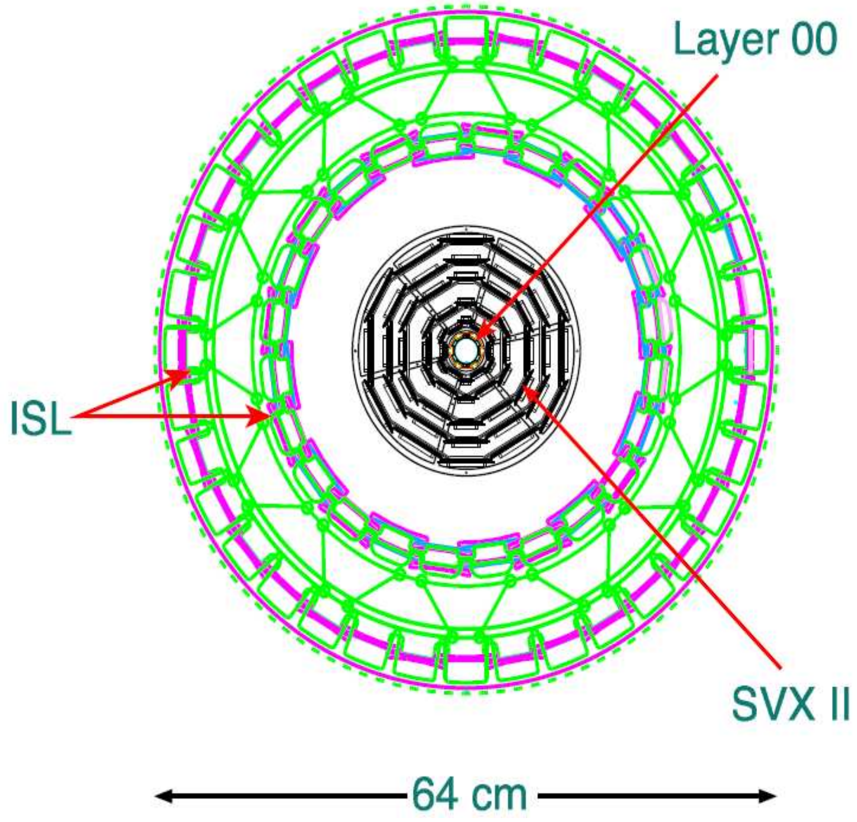


Figure 2.7: An end view of the CDF II silicon system. The sensors are depicted in black for the SVX II, pink for the ISL and red for L00. The SVX II cooling bulkheads and the ISL support structure are also shown. Full details can be found [33].

The Silicon Vertex Detector II (SVX II)

SVX II is the most important system and consists of 3 cylindrical barrels 29 cm long [34]. The total subsystem therefore extends to a z of ± 45 cm from the interaction point. Each barrel has 5 concentric layers of double sided silicon divided into 12 wedges called ladders. There is a combination of 90° and small angle stereo layers. In the detector there is overlap between the ladders for alignment purposes and to improve the impact parameter resolution. The distance between silicon strips, the strip pitch, is 50 μm. The $r\phi$ resolution of a hit in the SVX is $9\mu m$, which is smaller than the strip pitch. This is achieved by analysing the electron-hole pair creation in neighbouring strips.

Layer 00(L00)

Layer00 is single sided silicon layer directly on the beam pipe at a radius of 2 cm [35]. It has full azimuthal coverage and $|z| <$47 cm. As radiation damage occurs, it will no longer be possible to fully deplete the inner most layers of the SVX at the maximum appliable voltage. L00 is single sided and a higher voltage can be applied and so will continue to work after the innermost layers of the SVX II can no longer provide information. This ensures that high resolution tracking can continue for the lifespan of the CDF detector.

During the initial period of this analysis, the data from the L00 detector was not correctly aligned or calibrated. For this reason, the information from L00 has not been included.

The Intermediate Silicon Layers (ISL)

The ISL is made from double sided silicon. The silicon layers have a 1.2° stereo angle. They are arranged in 12 wedges similar to the SVX-II. There is one central layer at a radius of 23 cm to provide an intermediate measurement between the SVX-II and the outer drift chamber (COT). There are a further 2 layers at a radius of 20 and 29 cm that are positioned at $1.0 < |\eta| < 2.0$ which give forward tracking information. At high η a track will not traverse all the layers of the COT and so the hits from the ISL become more important. The forward tracking information from the ISL can be used in conjunction with energy deposits in the forward calorimeters to form tracks without the need of the COT.

2.2.4 The Central Outer Tracker

The second part of the tracking system is the Central Outer Tracker [36]. The COT is a cylindrical wire drift chamber that extends from a radius of 43.4 cm to 132.3 cm from the beam pipe and has a length of 310 cm. Radially the COT is divided into 8 super layers, where super layer 1 is closest to the interaction point. Every super layer is segmented into supercells in ϕ as shown in Figure 2.8. The number of super cells scales appropriately with the radius of the super layer. Each cell has 12 sense wires and 12 potential wires made out of 2mm gold plated copper beryllium wire. The cell layout is shown in Figure 2.9. The cathode is a made of gold on 0.25mm Mylar and the sides of the cell are mylar strips with field shaping wires. In super layers 2,4,6,8 the sense wires, those that are used to detected the ionisation, are axial (parallel to the beam line). Therefore, these super layers can give information on the $r-\phi$ plane only. Super layers 1,3,5,7 have their wires at a $\pm 2^{\circ}$ angle to the beam line so that some z information can also be extracted.

The COT is filled with a 50/50 mix of argon and ethane gas. When a charged particle travels though the COT the gas is ionised. The wires in the COT produce an electric gradient and this causes the electrons to drift towards the sense wires. They are detected when they reach the wire. In this gas the electrons have a sufficient drift velocity so that the maximum drift time is 100 ns which is smaller than the bunch crossing rate. This means that the tracks from successive bunch crossing interactions are separated.

Tracks with $|\eta| < 1$ will traverse all the layers of the COT as shown in Figure 2.6. In this ideal case, the charged track will have 48 axial and 48 stereo measurements and these hit measurements are used to reconstruct tracks as will be described in section 2.3. The $r-\phi$ resolution of a hit is approximately 140 μm. The resolution in

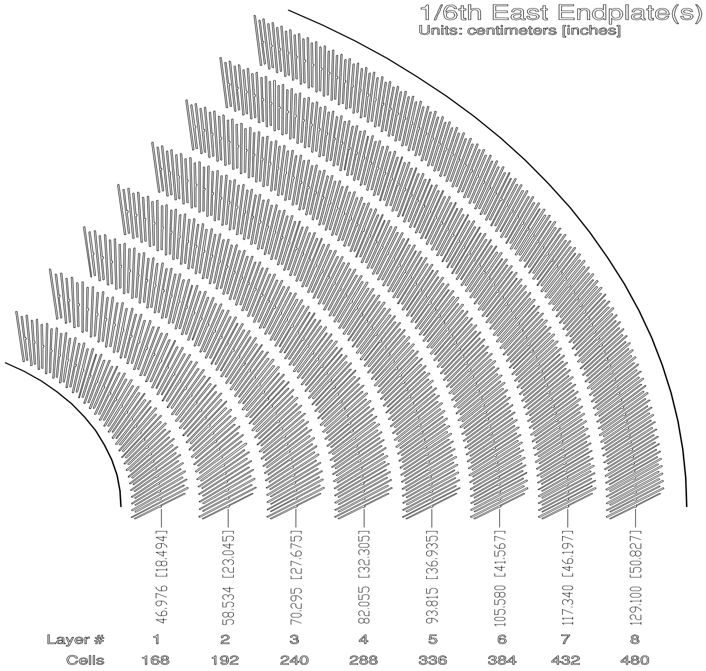


Figure 2.8: The cell layout in a section of the COT. The even super layers are axial and the odd super layers are stereo. The diagram gives the number of cells per superlayer. It also gives the radius of the midpoint of each superlayer from the beam line in cm[inches]. [32]

the z direction is better than 5 cm.

2.2.5 The Solenoid

A charged particle in a homogeneous magnetic field will travel in a helix around that magnetic field. The magnetic field is provided by a solenoid magnet situated outside the silicon and COT. It provides a 1.4 T field in the direction of the beam line. The solenoid is superconducting and operates at liquid helium temperatures and is made from Aluminium stablilised NbTi. The solenoid current, approximately 4665 A, is regulated by feedback driven by an NMR probe. The time interval for feedback is approximately 1 minute and this ensures a stable magnetic field.

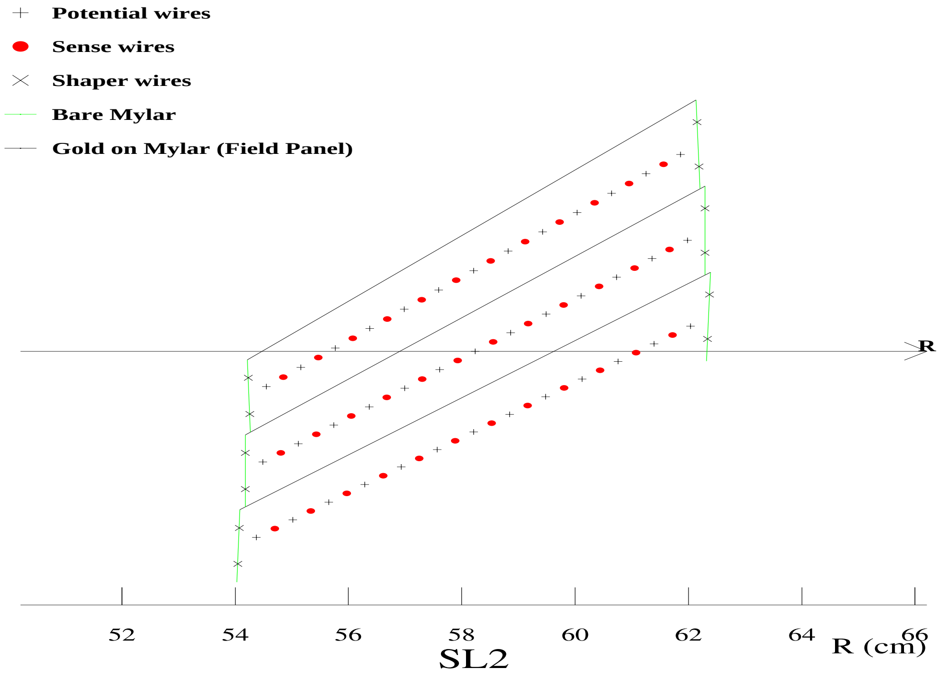


Figure 2.9: The wire layout within a super cell

2.2.6 Time of Flight

The Time of Flight detector [37](TOF) is used to help separate pions, kaons and protons. This particle identification (PID) information can be used to enhance B physics studies. The detector is used to give the time in the laboratory frame for a particle to reach the TOF. With knowledge of the momentum, this can be used to calculate the invariant mass which is used to identify particle type. The timing resolution is 100 ps. For tracks with $P_T < 1.5$ GeV the TOF gives 2σ separation between pions and kaons.

The detector itself is located between the COT and the solenoid and consists of 216 plastic scintillator strips of dimension 2.8m x 4cm x4cm. Each has a photo multiplier tube(PMT) at each end to measure pulse height and width. The difference in times of hits between a pair of PMTs is used to give the z position. All the PMT hits are fitted offline to calculate the difference between the time of the bunch crossing and the time the particle reaches the TOF. Using the difference in times it is possible to calculate the speed of the particle and determine its mass.

The TOF detector is best used for separating low momenta tracks, whereas the data in this thesis has selection requirements that force at least 2 out of 3 tracks to have momentum greater than 2 GeV. Studies have not been done to show what effect the information from the TOF detector may have on this lifetime analysis and therefore the information is not used.

2.2.7 Calorimetry

The calorimetry is located outside the tracking systems and the solenoid. Energy deposition in the hadronic and electro-magnetic calorimeters can help in the detection of photons, hadronic jets and leptons. Moreover, information from the distribution of energy and tracks can be used to detect missing transverse energy and hence infer the presence of neutrinos. Missing transverse energy is the energy that you would expect to measure due to the conservation of momentum and energy. The energy however is not detected (and hence missing) as it is the energy of a particle that does not interact with the calorimeter such as a neutrino.

The calorimetry is divided into 2 sections, the central region with $|\eta| < 1.1$ and the plug region that covers $1.1 < |\eta| < 3.64$. Each region consists of an electromagnetic calorimeter(CEM, PEM) [38] and a hadronic one(CHA,PHA) [39] [40]. There is also a further hadronic calorimeter on the end wall to cover the gap between the central and plug region (WHA).

The calorimeters use a scintillator sampling system with tower segmentation. There

	CEM	**CHA**	**WHA**	**PEM**	**PHA**
Coverage	$\lvert\eta\rvert < 1.1$	$\lvert\eta\rvert < 0.9$	$0.7 < \lvert\eta\rvert < 1.3$	$1.1 < \lvert\eta\rvert < 3.6$	$1.1 < \lvert\eta\rvert < 3.6$
Layers	31	32	15	23	23
Material	Lead	Steel	Steel	Lead	Iron
Depth	$18\ X_0$	$4.7\Lambda_0$	$4.5\Lambda_0$	$21\ X_0$	$7\Lambda_0$
$\frac{\sigma_E}{E}$	$\frac{13.5\%}{\sqrt{Esin\theta}} \oplus 2\%$	$\frac{80\%}{\sqrt{E}}$	$\frac{80\%}{\sqrt{E}}$	$\frac{14.4\%}{\sqrt{E}} \oplus 0.7\%$	$\frac{80\%}{\sqrt{E}} \oplus 5\%$

Table 2.1: A description of the energy resolution of the calorimeters

are alternate layers of absorber and plastic scintillator. The particles lose energy in the absorber and create a shower of particles which are detected in the scintillator. The energy of the shower is roughly proportional to the number of particles in the shower. Wavelength shifters in the scintillator carry the readout to the photomultiplier tube. The electromagnetic calorimeters use a lead absorber, while the hadronic ones use steel or iron absorber. The properties of the calorimeters are given in Table 2.2.7. The amount of material is given in terms of radiation interaction lengths X_0, which is the mean length of material traversed for an electron to lose all but 1/e of its energy. The radiation length for lead using a 7.6 GeV electron is approximately 0.56 cm. The hadronic calorimeters are described in terms of Λ_0, which is the interaction length for nuclear interactions. The interaction length for iron is about 16 cm. The electrons and photons will be stopped in the electromagnetic calorimeter, but the other particles will deposit most of their energy in the hadronic calorimeter.

In the electromagnetic calorimeters, there are the proportional chambers (CES, PES) at the depth that represents the approximate shower maximum. The proportional chambers have wires in $r - \phi$ that allow measurement of the the 3D position and allow matching of the shower to a track. In front of the Electromagnetic calorimeter are the shower presamplers (CPR,PPR). Information from this subsystem is useful for $\pi^0 - \gamma$ separation.

2.2.8 Muon Detectors

Muons are 200 times heavier than electrons, so they have less bremsstrahlung radiation. Additionally, muons do not have strong interactions with nuclei. For these two reasons muons are virtually undetected by the calorimeters and so the muon detectors are placed outside the calorimeter systems as this is an effective way to identify muons. Between the interaction point and the muon systems there is heavy material shielding from the calorimetry and in some cases additional steel absorbers due to the magnet return yoke. This heavy shielding reduces the number of hadrons reaching the muon detectors and creating a fake muon signal. However, it increases multiple Coulomb scattering, which means a muon may have had a number of small angle scatterings before it reaches the muon detector. There are 4 muon systems [41].

- CMU Central MUon detector
- CMP Central Muon uPgrade
- CMX Central Muon eXtension
- IMU Intermediate MUon detector

All the detectors have 4 layers of rectangular cells and each cell is a drift tube with a single steel wire that runs parallel or almost parallel to the beam position. This allows for measurement of ϕ. The gas is a 50:50 mix of argon and ethane and the principle of operation is similar to that of the COT; charged particles ionize the gas in the chamber and the ionization electrons drift towards the sense wire. Hits from the detector are used to form short tracks (stubs) which are then matched to tracks from the main tracking chambers. The multiple scattering does make this matching harder but the mismatch is Gaussian in position and can be accounted for.

There are also layers of scintillator that provide timing information. This allows the association of a muon with a particular interaction, as the drift time in the muon chambers exceeds the time between bunch crossings. It also allows for rejection of cosmic rays. The muon detectors all have differing coverage in η and ϕ as shown in 2.10.

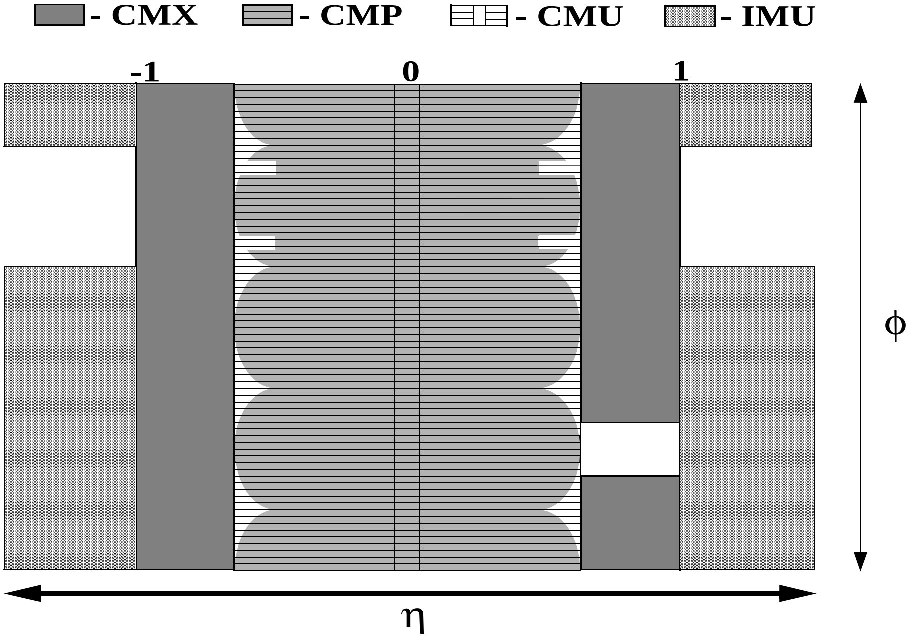


Figure 2.10: The $\eta - \phi$ coverage of the Muon Chambers

The CMU is situated outside the hadronic calorimeter at a distance of r=347cm from the beam line and has a coverage of $|\eta| < 0.6$. The hit resolution is 1mm in z and 250 μm in $r - \phi$. Before a particle reaches the CMU it will only pass through 5.5 radiation lengths of material, the smallest of any of the muon detectors. Therefore a fake or "punch through" is most likely for the CMU.

The CMP is located behind the CMU and the magnet return yoke which also provides additional shieleding, increasing the number of radiation lengths the particle traverses before it gets to the CMP. The CMP is a rectangle and is the only major system not to be azimuthally symmetric, hence the η coverage changes as a function of ϕ. The CMX increases the η coverage to 1.0 and the IMU is used with tracking to identify muons in the forward region. The final decay state considered in this thesis does not contain a muon and therefore information from these detectors is not used in the analysis.

2.2.9 Luminosity Measurement

The Cherenkov Luminosity Counters (CLC) are used to measure the luminosity. They are situated close to the beam line in the forward and backward regions with $3.7 < |\eta| < 4.7$. Each has 48 isobutane filled Cherenkov counters with timing resolution that allows the CLC to differentiate between beam loses, which are typically out of time, and particles from $p\overline{p}$ interactions. The luminosity is calculated from equation 2.4,

$$\mu \times f_{BC} = \sigma_{p\overline{p}} \times L \tag{2.4}$$

where f_{BC}, the bunch crossing frequency is known from the Tevatron RF system and $\sigma_{p\overline{p}}$, the inelastic cross-section, is known to within 4% uncertainty. The average number of interactions per bunch crossing, μ, is measured by the CLC, and this gives a total uncertainty on the luminosity less than 6%. The integrated luminosity measurement is important for measuring cross-sections but is not necessary for a lifetime analysis.

2.3 Charged Particle Tracking

The analysis presented in this thesis uses only the tracking information. This section describes how the information from the COT and silicon detectors is used to reconstruct the tracks and their parameters.

The helical path of a charged particle can be described by 5 parameters. The first three below describe the circular motion in the $r-\phi$ plane, and the latter two describe the straight line motion in the $r-z$ plane. The five parameters used in CDF tracking are defined as follows:

- **Impact parameter** d_0: This is defined as $[\hat{\boldsymbol{p}} \times \boldsymbol{d}] \cdot \hat{\boldsymbol{z}}$ where $\hat{\boldsymbol{p}}$ is the unit vector in the direction of the particle momentum, $\boldsymbol{d}$ is the vector from the primary interaction to the point of closest approach and $\hat{\boldsymbol{z}}$ is the unit vector in the z direction. $|d_0|$ is the distance of closest approach to the primary vertex.
- **Azimuthal angle** ϕ_0: This is the azimuthal angle of the particle trajectory at the point of closest approach to the interaction point.
- **Curvature C**: If the track helix is projected onto the $r-\phi$ plane it will form a circle radius R_C. The track curvature is defined as $C = q/(2R_C)$ where q is the charge of the track.
- **cot**θ: This is the cotangent of the angle between the momentum of the particle and the z axis
- z_0: This is the position along the z axis at the point of closest approach to the interaction point.

The transverse momentum of the particle is related to the radius of curvature through the relationship below:

$$P_T = 0.3BR_C \tag{2.5}$$

where B is the magnetic field and R_C the radius of curvature. As the momentum of a particle increases so does its radius of curvature. A larger radius of curvature is harder to measure and therefore the resolution of the momentum of a track decreases as the momentum increases.

Offline track reconstruction at CDF first finds tracks in the COT as described in detail in [42]. The electronic readout of the COT provides the drift time, which is the difference between the wire signal and the beam crossing times, for every hit. The drift velocity and drift time are used to find the position where the ionizing track must have crossed. There is ambiguity in these hit positions as it is not known which side of the wire the particle traversed but this ambiguity is removed when all the hits are considered and a track is reconstructed.

The tracking algorithm starts by finding track segments within an axial super layer. These segments are formed by combining hits within a super layer to form a straight line segment. To extend the track across the COT volume, two different algorithms are used. The first is called “segment linking”. Segment linking matches segments in different super layers to form a track. The other strategy is “Histogram linking”. Histogram linking takes a segment in one super layer and extrapolates the segment to the beam position. For each of the other layers that could contain hits belonging to the track a histogram is filled with the log likelihood (as a function of track curvature) that the initial segment and hit in the current layer belong to the same track. The histograms for each layer are summed, and in ideal measurement circumstances all contributions due to hits belonging to the track that includes the initial segment will have the same curvature. This means that hits from the same track will populate the same bin. If the histogram has sufficient entries in one bin then this combination of hits is accepted as a track. The hits from either track finding strategy are then fit to get the parameters of a track in 2-D. Hits are then added from the stereo super layers and the track is fit again to make the 3-D track with all 5 track parameters measured. At this point any hits that are more than 600 μm in the $r - \phi$ plane from the fit track are dropped and the track is fit once more. These hits are expected to be poorly measured or coming from another track so removing them improves the quality of the fit.

There is significant overlap between the tracks that are found by the two strategies of segment linking and histogram linking, however using both maximizes the track finding efficiency. Duplicate tracks from the two strategies, which are those that share more than 15% of hits in common, are removed from the tracklist. Tracking studies using D mesons (D^0, D^{+*}, D^+, D_s^+) that satisfy the fiducial and kinematic criteria of the hadronic trigger are found to have a track finding efficiency of between 94-98% depending on D meson type and P_T(D) [43].

The strategy for incorporating the information from the silicon detector into the track reconstruction is described in detail in [44]. The algorithm is seeded by the COT track. As the COT resolution in the z direction is an order of magnitude worse than in the $r - \phi$ plane careful linking in stereo hits is required. The first algorithm considers only axial information. The track reconstruction is progressively "Outside-In". This means that the COT track is extrapolated to the edge of the silicon. At the silicon layer a window around the COT track is defined. This window is defined using the 1 standard deviation uncertainty of the point where the extrapolated track intersects the silicon. The window is of radius 4 standard deviations. For each hit within this window a new track is created. The two track candidates with the greatest number of silicon hits are extrapolated into the next layer where the procedure is repeated. At each silicon layer the track error matrix is adjusted to reflect the amount of scattering material encountered. The procedure is then repeated but this time hits from the stereo silicon layers are also added making 3-D tracks. For each seeding COT track the silicon track that has the hits in the largest number of silicon layers is the best candidate and this track is put into the event's track list, where each track is defined by the 5 parameters described above.

The momentum resolution of a track is dominated by the information from the COT. For tracks reconstructed in the COT only the resolution is $\frac{\sigma_{P_T}}{P_T} = 0.15\% P_T$ $[GeV/c]^{-1}$ and for tracks that use COT and silicon this is reduced to $\frac{\sigma_{P_T}}{P_T} = 0.07\% P_T$ $[GeV/c]^{-1}$. The silicon improves the impact parameter resolution. The impact parameter resolution for tracks without L00 information is $9 \oplus 50/P_T$ μm and with the L00 information the impact parameter resolution is $6 \oplus 25/P_T$ μm.

2.3.1 The Beam Profile and Interaction Vertex

The beam profile and interaction vertex are determined by the track information from the COT and silicon detectors. The transverse profile of the beam can be described by a Gaussian in x and y ,i.e., circular. Beam width studies at CDF [45] have found that the width of the beam averaged over z is $26\mu m$ for both the x and y directions. The beam profile in the z direction is described by a Gaussian with mean 0 and width of approximately 25 cm. The mean of the Gaussian determining the beam profile in the x and y direction give the interaction point for the event. The interaction point is found by taking a vertex fit of all tracks in the event as described in [46]. This vertex fit finds two vertices, the first using information from COT tracks only, and the second using tracks with COT and silicon information. They are both calculated to allow comparison and to check global alignments. The vertex fit is an iterative process in which tracks that do not appear to originate from the primary vertex are successively removed from the fit. The interaction point calculated from the vertex fit is averaged over a number of events, and the beam profile and interaction point are stored for each run. The interaction point found by using tracks with silicon information is the default used. Subsequently, algorithms have been introduced that allow for calculation of an interaction point for each event [47] and not a run averaged vertex position as described above. However these algorithms were not utilized for the analysis presented in this thesis.

2.4 The CDF Trigger System

The rate of collisions at the CDF interaction point is determined by the beam structure and is 1.7MHz. Each event would contain 200 Kb of information if it were read out and written to tape. It is not possible to record such a volume of data. Currently approximately 120 events can be written to tape per second, and therefore the trigger must be able to provide a rejection factor of 14000.

Most interactions do not contain physics processes of interest. For example the inelastic $p\overline{p}$ crosssection is 50 mb, however the b quark production cross section within the CDF detector acceptance $|\eta| < 1$ and $P_T > 6$ GeV is only 10 μb. The top production cross section is even smaller at 5 pb [48]. Therefore it is the role of the trigger not only to simply provide a rejection factor of 14000 but also to be able to preferentially select the interesting processes and provide a higher signal to background ratio in the collected events.

2.4.1 Trigger Overview

A flow chart of the Trigger System is given in Figure 2.11. The CDF trigger system consists of 3 levels. At each level more processing time is available before a decision must be made and at each level the information available is more accurate as better reconstruction algorithms can be used. There are different trigger paths that look for different signatures in the data. For example the calorimeter information is used to create simple objects such as electrons and photons, and the trigger criteria will be passed if the object satisfies a minimum P_T threshold. The accuracy of the object's energy or momentum increases for each level due to the processing time available to reconstruct the object. There are also triggers that look for missing transverse energy or muons or simply tracks that do not come from the primary vertex. The data used in this thesis is collected through the track triggers and they are described in more detail after the overview.

The Level 1 trigger is a synchronous deadtimeless hardware trigger with a pipeline of that can store events for just over 5.5 μs and so this is the L1 processing time. A decision to discard or keep the event must be made in that time. The output from L1 is approximately 25,000 events per second. If an event passes a L1 trigger path criteria, the detector front ends send more information to Level 2(L2). L2 has 4 buffers to store L1 accepted events. If all the buffers are full, then the CDF detector incurs deadtime and cannot process more events. The goal during data taking is to keep dead time below 5%. The L2 trigger is an asynchronous software trigger that has a processing time of approximately 50 μs. The output from L2 is approximately 550 Hz. If an event is accepted by a L2 trigger then the whole detector is read out. The Level 3 trigger [49] is purely software and performs full reconstruction of the event using a modified version of the offline production executable on a PC farm. The output from Level 3 is approximately 120 Hz. This rate is limited by how fast events can be recorded.

The trigger rates will grow with instantaneous luminosity. To deal with this there are a number of mechanisms in place. These include switching on the high bandwidth trigger paths only after the instantaneous luminosity has fallen beyond a defined threshold. Other mechanisms include prescales. A prescale is the fraction of events that are rejected at random by a trigger path. At CDF there is dynamic prescaling which means that the prescale fraction changes based on the available trigger bandwidth.

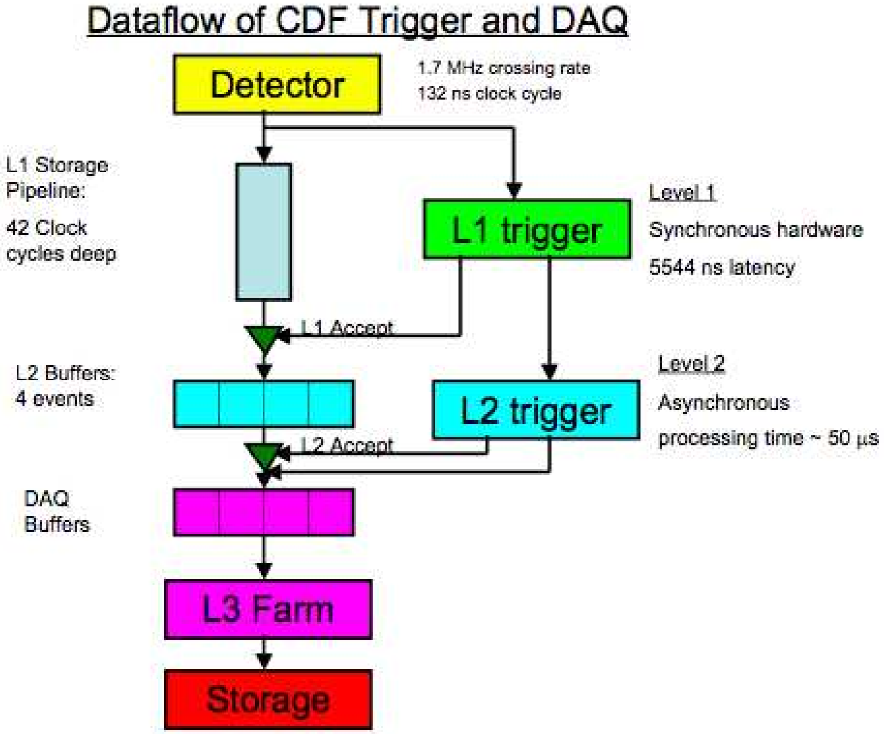


Figure 2.11: A flow diagram representing the CDF trigger system

2.4.2 The B Physics Track Triggers

The B physics track triggers were designed to collect events containing B hadrons or mesons in hadronic final states. The CDF Run I b physics trigger strategy was to select events containing leptons, however this strategy suffers due to the low branching fraction of B mesons to final states that include leptons. The track triggers select tracks with large impact parameter and events with displaced vertices. This takes advantage of the long B lifetime which will produce this topology. As this strategy does not require the presence of leptons, the trigger can collect decays in a wide variety of hadronic final states.

The selection criteria for each level is given in Table 2.4.2. The L_{xy} refers to the transverse displacement between the primary vertex and the track pair intersection point. The information flow from the detector subcomponents is shown in Figure 2.12. There are three different trigger paths in this family of triggers. The main difference between them is the requirement of opposite charge of the two tracks and the P_T requirements made on the tracks. Each trigger path is made available at a different luminosity and has different prescales. For example the `BCHARMHIGHPT` has the highest background rejection rate due to tighter cuts and so uses less bandwidth and is therefore available at a higher instantaneous luminosity where bandwidth is limited. On the other hand, the `BCHARMLOWPT` path has the loosest selection criteria and will collect a larger fraction of signal events, but due to the high bandwidth required, it is only switched on once the instantaneous luminosity has fallen and more bandwidth is available.

At L1 there is not sufficient time to read out the silicon or incorporate that information into a fast track reconstruction algorithm and therefore the L1 decision is made using information from the COT hits only. These L1 tracks are known as XFT tracks, reconstructed using the eXtremely Fast Tracker. The axial XFT Finder searches for hit segments using 4 super layers [50], and then the Linker searches for segments from tracks with $P_T >$1.5 GeV. The search is performed in slices of ϕ with one track allowed per 1.25°. At the same time, the Stereo Finder searches for hit segments in the 3 outer stereo layers. The axial and stereo track are combined using the stereo linker association modules (SLAM) which are custom built chips that are programmed to combine

Trigger Criteria L1 (XFT)	**Units**	BCHARMLOWPT	BCHARM	BCHARMHIGHPT
Minimum Track P_T	GeV	2.0	2.0	2.5
Two Track Charge Product	-	na	-1	-1
Two Track Max $\Delta\phi$	degrees	90°	135°	135°
Two Track P_T scalar sum	GeV	0	5.5	6.5
Trigger Criteria L2 (SVT)				
Minimum d_0	μm	120	120	120
Maximum d_0	μm	1000	1000	1000
Minimum Track P_T	GeV	2.0	2.0	2.5
Maximum Track χ^2	-	15	15	15
Track Pair Charge Product	-	na	-1	-1
Maximum Pair $\Delta\phi$	degrees	90°	90°	90°
Minimum Pair $\Delta\phi$	degrees	2°	2°	2°
Minimum Pair P_T scalar sum	GeV	0	5.5	6.5
Minimum Pair L_{xy}	μm	200	200	200
Trigger Criteria L3				
Minimum d_0	μm	80	80	80
Maximum d_0	μm	1000	1000	1000
Minimum Track P_T	GeV	2.0	2.0	2.5
Maximum Track η	-	1.2	1.2	1.2
Track Pair Charge Product	-	na	-1	-1
Maximum Pair $\Delta\phi$	rad	1.571	1.571	1.571
Maximum Pair $\Delta\phi$	rad	0.035	0.035	0.035
Maximum Pair Δz_0	cm	5.0	5.0	5.0
Minimum Pair L_{xy}	μm	200	200	200
Minimum Sum P_T	GeV	4.0	5.5	6.5

Table 2.2: This table gives the description of the trigger selection criteria at each level

the axial and stero tracks. The stereo confirmed tracks are available $1.8\mu s$ after the bunch crossing. At L1 there is no information about displaced vertices available and so the criteria simply select central high P_T tracks.

If the event passes at L1, then the XFT tracks are sent to the Silcon Vertex Tracker (SVT) [48] and global L2 for more sophisticated algorithms. The tracks used to make the Level 2 decision are called SVT tracks. The SVT finds SVT tracks using the stereo-confirmed XFT tracks and the digitized SVX pulse heights. The hit finder finds clusters of hits in each layer of silicon. These clusters of hits are compared to “roads” which are pre defined patterns of tracks stored in memory. Coarse pattern recognition to these roads is performed and then the roads and associated hits are fitted to give track parameters including the impact parameter. The Level 2 algorithm steps are shown in Figure 2.13.

If the event passes the L2 decision, it is passed to the L3 processors. The L3 decision is made based on L3 tracks. L3 tracks are found using an algorithm very similar to offline reconstruction as described in Section 2.3. The differences between the offline reconstruction algorithm and the L3 algorithm are that at L3 the segment linking algorithm is not used, and hits are not dropped after a track is fitted. The L3 algorithm is different as it needs to be faster, and these two changes speed the reconstruction by a factor of 2. An event must pass the selection criteria of all levels in one of the trigger paths to be recorded.

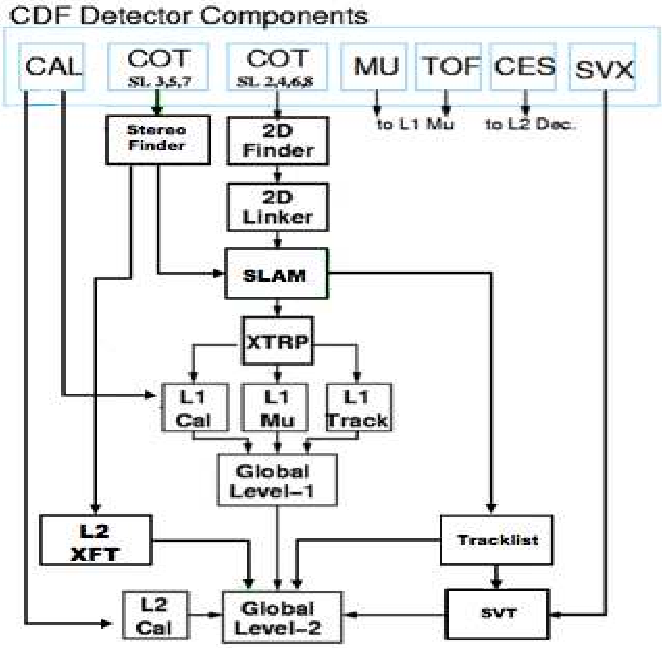


Figure 2.12: This diagram shows which information from the detector sub components is used to make the trigger decision

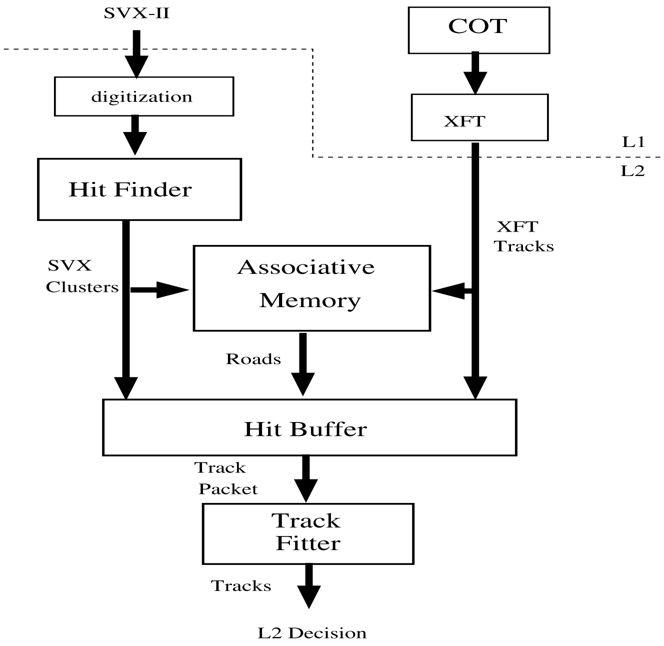


Figure 2.13: Flow diagram of the decision making process of the L2 track trigger.

Chapter 3

Event Reconstruction and Selection

> This chapter contains a description of the data sample and procedure for the search and reconstruction of B-mesons from charged daughter tracks within it. A description of all selection criteria to optimize the signal to background ratio and the description of the Monte-Carlo sample used to validate the selection criteria is also given. Finally a description of criteria applied by the trigger is given, as this is crucial in determining the lifetime.

3.1 Data Sample

The measurement in this thesis uses data collected between February 2002 and March 2006 and corresponds to an integrated luminosity of 1.0 fb^{-1}. Events that were collected through the `BCHARM` family of triggers are analyzed to look for the signature of a B meson decay to $D^0\pi$. A diagram of the expected topology of a decay in this mode is shown in figure 3.1. In this chapter the reconstruction of $B^- \rightarrow D^0\pi$ is described. The charge conjugate reconstruction for the B^+ is also carried out. After decay reconstruction, further selection cuts are applied to remove fake events and improve the signal to background ratio. A selection cut defines the minimum or maximum value a given reconstructed quantity must have in an event for the event to be part of the final sample on which the measurement is made. Standard CDF reconstruction tools are used to select events, and the description of the algorithms are given in this chapter. Section 2.3 described how the hit information in the COT and silicon detectors was used to reconstruct tracks and find the primary vertex position. These tracks and the information about the primary vertex are now used to reconstruct the B meson decay.

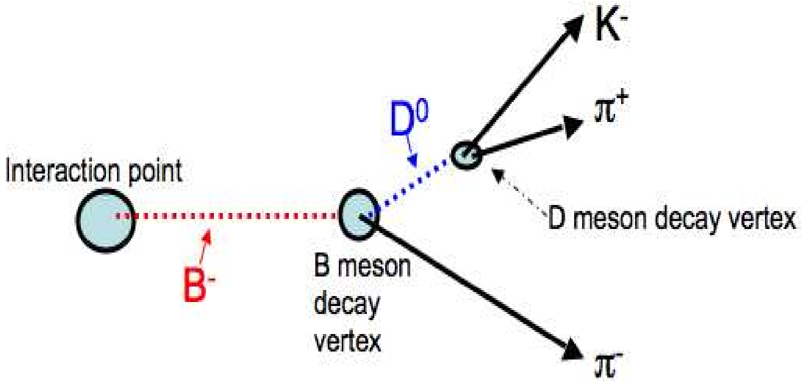


Figure 3.1: An example of the decay topology

It is useful to introduce some nomenclature to refer to the three daughter tracks individually. The pion that originates from the B vertex is called π_B and the pion and kaon from the D meson are called π_D and k_D respectively.

3.2 Track Selection

Before combining tracks to form a B meson, standard CDF track selection cuts are applied. This removes from the sample tracks that are poorly measured due to a low number of hits. Tracks with low transverse momentum are also removed as they are

are more likely to be fake tracks or poorly measured tracks due to a larger probability of multiple scattering interactions with the detector material.

- Each track has transverse momentum P_T greater than 0.4 GeV.
- Each track is required to have at least 5 hits in at least 2 axial COT super layers
- Each track is required to have at least 5 hits in at least 2 stereo COT super layers
- Each track is required to have hits in a minimum of 3 silicon r-ϕ layers.
- Each track is required to have an $|\eta| < 2$

The decay channel analyzed has both kaons and pions in the final state. At CDF, particle identification cannot definitively separate tracks that are pions from those that are kaons, therefore all possible tracks must be considered as though they were pions and considered again as though they were kaons.

3.3 $B \rightarrow D^0\pi$ Reconstruction

Decay reconstruction takes place by combining tracks, finding the common vertex and calculating the properties of a parent particle. Events that have candidate particles with the desired quantities such as mass in the desired range are selected for further analysis. At the reconstruction stage only loose requirements are made on the parent and daughter particles. Once these events are selected from the full data sample, further selection criteria can be applied to optimize the chosen subsample.

3.3.1 D^0 Candidates

The reconstruction of the decay $B^- \rightarrow D^0\pi^-$ begins with the reconstruction of the D^0 in the decay $D^0 \rightarrow \pi^+K^-$. Charged tracks are selected based on the track quality criteria already described. Oppositely charged pairs of tracks are chosen with their energy assigned based on the hypothesis that they are pions or kaons with $E = \sqrt{m_{K,\pi}^2 + \boldsymbol{p} \cdot \boldsymbol{p}}$ where $m_{K,pi}$ is the PDG [11] average pion or kaon mass of 0.140 and 0.494 GeV respectively. The momentum $\boldsymbol{p}$ determined from the tracking geometry and magnetic field has no direct assumptions about the mass of the particle. The invariant mass of the pair is then calculated, and this is the candidate D^0 mass $m_D = \sqrt{(E_\pi + E_k)^2 - (\boldsymbol{p_K} + \boldsymbol{p_\pi})^2}$. If this mass lies within ±0.055 GeV of the mass of the D^0 PDG mass, 1.864 GeV, and the z_0 parameters for both the pion and kaon are within 5cm of each other, the candidate is considered further.

For all D candidates that pass the basic selection a vertex fit is performed using CTVMFT [51]. This algorithm uses an iterative process, and works by taking a set of tracks and constraining them to the best common vertex point. The starting point of the iteration is important in finding the best vertex position. The iteration starts by finding the two intersection points of the helical path of the first two tracks in the $r - \phi$ plane. The one with the least separation in z is used as the algorithm starting point. To find the vertex the track parameters and the vertex point are treated as variables and the χ^2 is minimized.

$$\chi^2 = \sum_{i,j=1}^{N} \boldsymbol{\eta_i^t} G_{ij}^{-1} \boldsymbol{\eta_j} \tag{3.1}$$

where $\boldsymbol{\eta}$ is the 5 component vector of differences between the unconstrained (measured) and constrained track parameters, N is the number of tracks used in the vertex fit and G is the re-calculated error matrix of the track parameters.

The CTVMFT algorithm described above returns a vertex position for the D candidate. It also returns the error on the length between the primary and secondary vertices in the transverse plane. This error includes the propagation of errors from the fitted track parameters and the errors on the primary vertex position. The uncertainty on the transverse length ($\approx 30\ \mu m$) is dominated by the uncertainty of the primary vertex position as this has the poorest resolution ($\approx 26\ \mu m$). There is now a candidate D meson with a calculated transverse momentum and decay point.

3.3.2 Finding B Meson Candidates

The next stage of reconstruction is to find candidate B mesons in the events already selected. As shown in the diagram 3.1 the B meson daughter particles are the D meson and a pion. For each D^0 candidate, all remaining negative tracks with momentum greater than 0.8 GeV are assumed to have the mass of a pion and are in turn combined with the D^0 to form a candidate B. If the resulting B^- meson has a mass between 4.5 and 6.5 GeV and a transverse momentum $P_T > 4GeV$ it is considered further. The daughter tracks of the D candidate and the B pion are refitted taking into account the particle mass assignment of the track and the expected energy loss corrections due to material traversed by the particle. The B decay vertex is then found using the vertex finding algorithm using the pion and "D^0 track". To improve the mass resolution of the B meson, the additional constraint that the D meson is constrained to its PDG mass of 1.8645 GeV is included in the χ^2 minimization [51].

It is possible that a single event will yield more than one B meson candidate. These are all kept in the sample so that the true B meson is not discarded. The selection criteria at this stage are relatively loose. To determine the value of a particular cut it is beneficial to study the expected distributions of particle properties in signal and background. For signal events, realistic MC is used to give a good representation of the B decays. The MC is described in section 3.4 before the selection criteria are determined in section 3.6.

3.4 The Monte Carlo Sample

There are two types of simulation samples used to validate the methods and evaluate the selection criteria in this analysis. The most important is the Fast Simulation, which will be described in chapter 6. It is used for method validation and systematic uncertainty determination. Realistic Monte Carlo(MC) simulation is also used primarily to determine selection cuts and to do some basic testing. The realistic Monte Carlo is not used for validation as it is not possible to generate the sample sizes required due to computer processing time limitations arising from full detector simulation. To test the method and determine systematic errors over four hundred million events are required. The full detector simulation can generate approximately 1000 events using one CPU for one day. Therefore it is not possible to use the realistic Monte Carlo generation to validate this method. The realistic Monte Carlo generation is described in this section.

For the realistic Monte Carlo generation, only $B^- \rightarrow D^0\pi^-$ events and their charge conjugate are generated. The b quarks are generated using the one particle inclusive differential cross section for heavy quark production in hadronic collisions including $\mathcal{O}(\alpha_s^3)$ radiative corrections [52] [53], and then decayed using the EvtGen package [54]. The events are then put through a full detector simulation using GEANT 3 [55]. This simulates all the hits in the detector sub components. The events are passed to the TrigSim program, which is an offline implementation of the L1 and L2 trigger algorithms. After that, the offline track reconstruction algorithm is run. This algorithm leaves the data in the same format as CDF data. Decay reconstruction programs are run on the generated data to give a sample of B meson events. As the realistic Monte Carlo is not used to provide any input into the final result, there is no re-weighting of the sample in momentum or trigger category to better match the data.

3.5 Variables on which Selection is Based

In this section, variables that are used for selection are defined. As the vertex measurement in the z direction is not as accurate as the measurement in the xy plane, proper lifetimes and proper decay lengths are calculated in terms of the transverse quantities so as to keep uncertainties on the measurement small.

- L_{xy}(X): For an ideal measurement, the L_{xy} would be the distance between the interaction primary vertex and the decay point of a secondary particle, (X). However, due to measurement uncertainty, the vector pointing to the secondary from the primary vertex and the particle momentum vector do not lie on the same path. Therefore a signed quantity is defined as

  $$L_{xy}(X) = \boldsymbol{r} \cdot \boldsymbol{\hat{P}_T} \tag{3.2}$$

  where $\boldsymbol{r}$ is the vector from the primary to secondary vertex in the transverse plane, and $\boldsymbol{\hat{P}_T}$ is the unit vector along the transverse momentum of the particle (X).

- ct: This is the proper decay length, i.e, the lifetime of the particle in its own rest frame multiplied by the speed of light. It is given by

$$ct = \frac{L_{xy}(X) M_X}{|\boldsymbol{P_T}|} \tag{3.3}$$

where M_X is the reconstructed mass of the particle X and $\boldsymbol{P_T}$ is its transverse momentum. The factor $\frac{M_X}{\boldsymbol{P_T}}$ is the Lorentz boost in the transverse plane.

- σ_t: This is the error on the measured proper lifetime. The error is dominated by the measurement of the primary vertex but also correctly incorporates errors on the secondary vertex measurement and track parameters. The quantity σ_{ct} is more often used, and $\sigma_{ct} = \sigma_t \times c$ is the error on the measured proper decay length.

- ΔR(A,B): For two particles A and B with η and ϕ measured as η_A, η_B, ϕ_A, ϕ_B, respectively, this is defined as

$$\Delta R(A, B) = \sqrt{(\eta_A - \eta_B)^2 + (\phi_A - \phi_B)^2} \tag{3.4}$$

- $\Delta\phi$(A,B): This is the angular separation in ϕ between two tracks or particles A and B.

- χ^2: The vertex fit returns a χ^2 which relates to the quality of the fit.

3.6 Selection Criteria on Candidate B Mesons

The full selection criteria are based heavily on the optimization carried out for a previous lifetime CDF analysis on the same decay mode [56]. The selection criteria have diverged from the previous analysis where lifetime based cuts are concerned. The technique used in this measurement uses analytical calculation to correct the bias due to lifetime based selection criteria of the trigger. It is easier to incorporate cuts on the L_{xy} of the B and D meson into the calculation rather than on the significance, $\frac{L_{xy}}{\sigma_{L_{xy}}}$, which is a quantity used in the previous analysis [56].

The selection criteria taken from [56] are given below:

- $\Delta R(k_D, \pi_D) < 2$
- $\Delta R(D, \pi_B) < 2$
- $P_T(D^0) \geq$2.4 GeV
- $P_T(B^{\pm}) \geq$5.5 GeV
- P_T $(\pi^{\pm}_B) \geq 1$ GeV
- The candidate B vertex $\chi^2 < 15$
- The impact parameter of the B with respect to the transverse primary interaction point is $< 80\mu$m.
- All the B daughter tracks have a z_0 within 5 cm of each other.

Cuts on L_{xy} for the B and D particle are chosen to remove as much background as possible with minimal impact on the signal yield. To understand the impact on signal yield, the realistic MC can be used to represent the distributions expected for real signal events. To understand the fraction of background removed by any cut the data events with reconstructed mass above 5.38 GeV are used. The reconstructed MC shows that events with mass above 5.38 GeV are unlikely to be true B meson decays in this decay mode.

The $L_{xy}(B)$ is considered first, and the distribution of this quantity is shown in figure 3.2 after all other selection cuts described above have been applied. In figures 3.2-3.4 the areas of the the MC and background distributions have each been normalized to 1 event. The cut is placed at $350\mu m$ as shown by the dotted black line. Placing the cut here is therefore expected to remove 94% of the background events but only 12% of the signal. This loss in signal is acceptable due to the large fraction of background events that are removed.

After the cut on the $L_{xy}(B)$ has been made, the $L_{xy}(D)$ (from the primary vertex) distribution of remaining events is shown in figure 3.3. (These distributions have also been normalized to 1 event) This figure shows that if a cut is placed requiring events to have $L_{xy}(D)$ >-100μm then only 0.1% of signal events will be removed whereas 37% of the remaining background will be removed from the sample. Therefore this is a very efficient cut as the loss in signal is minimal.

After the cut on $L_{xy}(D)$, one further cut is made on σ_{ct}. Events with large σ_{ct} are poorly measured and undesirable in a lifetime analysis. Events with σ_{ct} >100 μm are removed. The distribution of σ_{ct} is shown in figure 3.4 after all other cuts. The figure demonstrates that this is also an efficient cut as very little signal (0.3%), is lost but 15% of the remaining background is removed.

A further lifetime cut is applied and requires that the B meson candidate has a proper decay length less than 3000 μm. As the mean lifetime is of order 500 μm this selection cut removes few signal events but aids in the numerical stability of the analytical calculation that corrects for selection bias.

The final selection cut made is that the reconstructed B mass is required to lie between 5.23 and 5.5 GeV and the reasons for this choice are further explained with the description of the mass model used to parameterize the distribution in section 5.4.

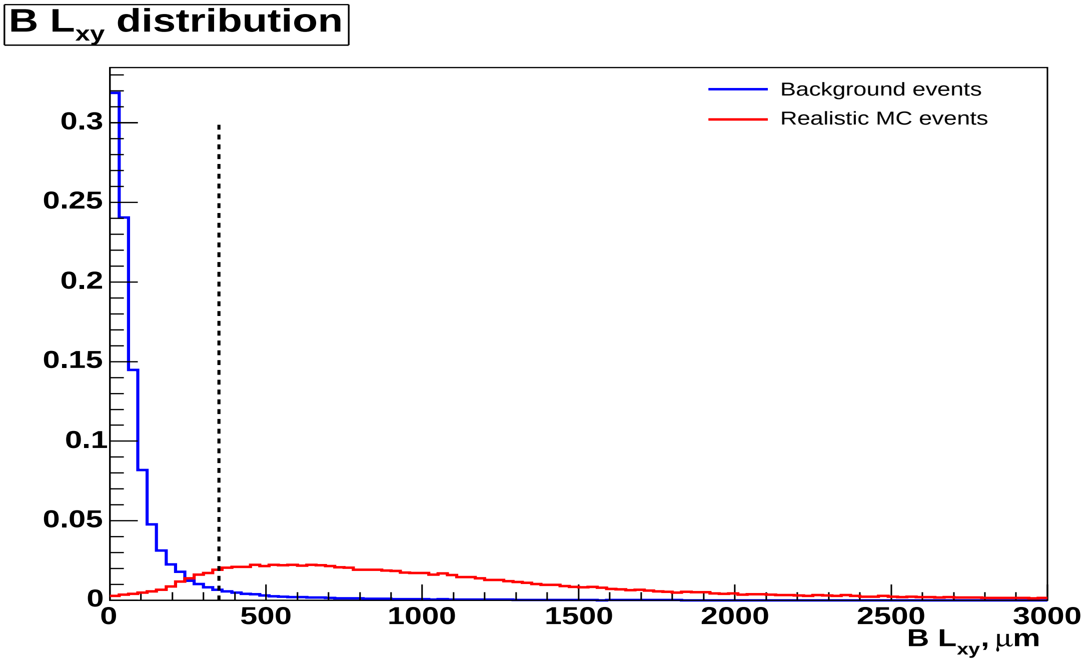


Figure 3.2: The distribution of the $L_{xy}(B)$ for background events (blue) from data and MC signal events (red). The dotted black line depicts the applied cut. Events to left of this line are removed from the sample. The areas of each distribution have been normalized to one event.

3.6.1 Trigger Reconfirmation

One final step in selecting the sample is to make sure the daughter tracks of the B meson can pass the trigger and that the event did not pass the trigger based on other tracks in the interaction. This analysis analytically corrects for the biases induced by the trigger selection criteria. It is therefore important to ensure that events in the final sample can pass the trigger so that the bias can be corrected for. In the `BCHARM` family of triggers there are three separate trigger paths. At high collision luminosity, only the `BCHARMHIGHPT` path is available to accept events. As the luminosity falls over the course of a store, the other two trigger paths also become available. During reconstruction, the information on which trigger paths were passed during data taking are recorded and and the corresponding trigger emulation is applied to the event. If more than one

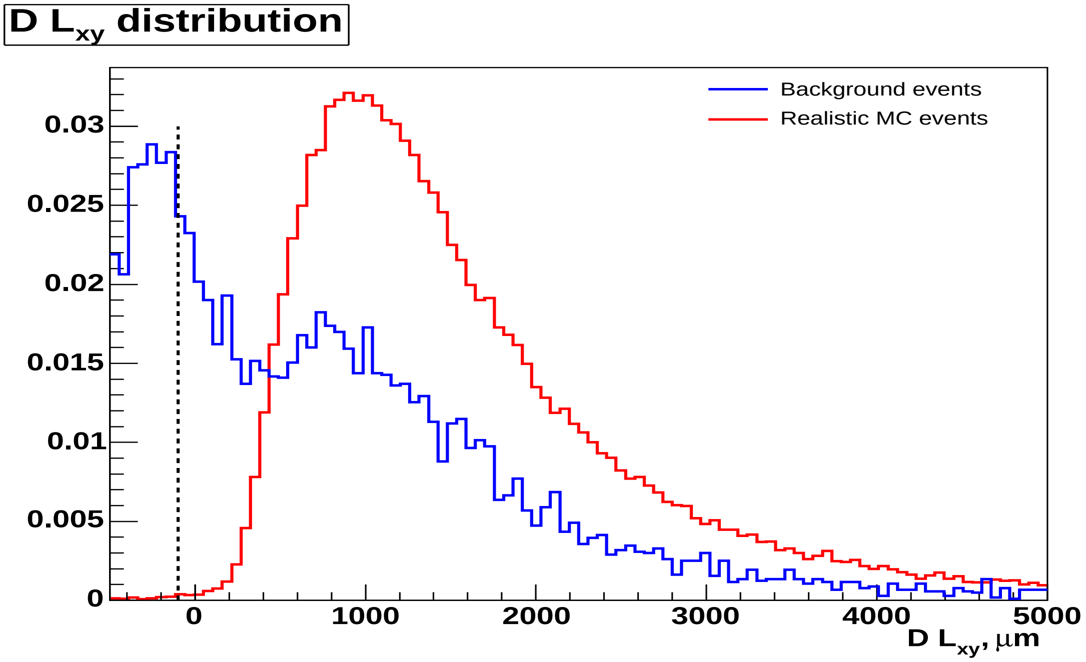


Figure 3.3: The distribution of the $L_{xy}(D)$ for background events (blue) from data and MC signal events (red). The dotted black line depicts the applied cut. Events to left of this line are removed from the sample. The areas of each distribution have been normalized to one event.

trigger path has been satisfied then the trigger with the loosest selection criteria is emulated.

The trigger criteria for each path was given in Table 2.4.2 and is repeated below in Table 3.6.1 for clarity. The main difference between the three trigger paths is that the `BCHARMLOWPT` path does not require the pair of tracks to be oppositely charged and that the `BCHARM` and `BCHARMHIGHPT` paths make demands on the scalar sum of the P_T of the two tracks. The scalar sum is for two tracks is simply $|\boldsymbol{P_{T1}}| + |\boldsymbol{P_{T2}}|$

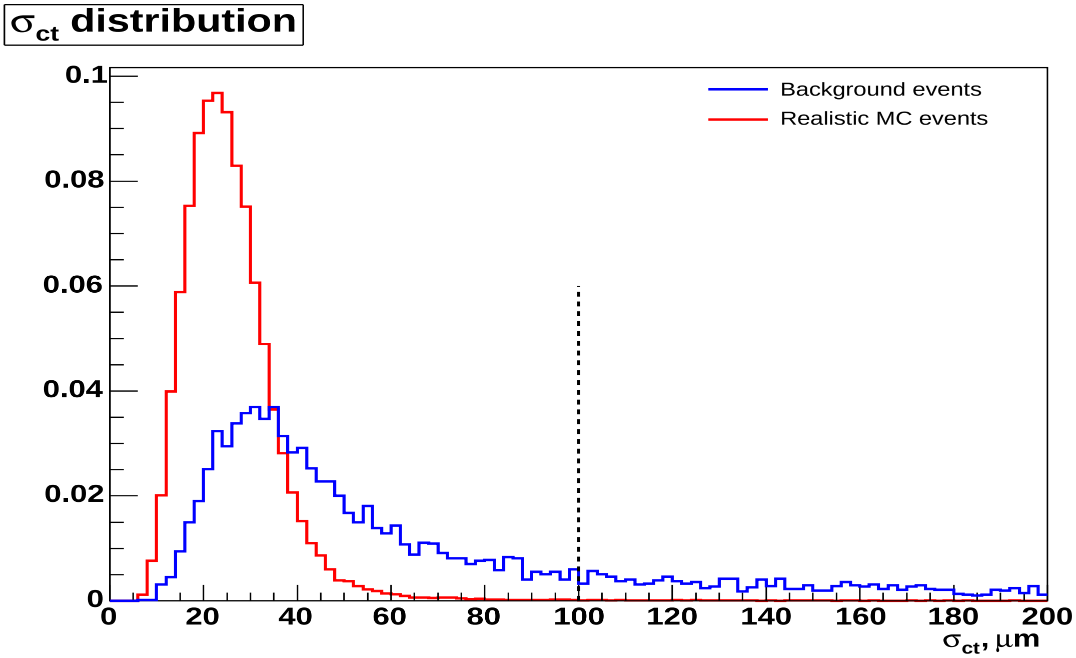


Figure 3.4: The distribution of the σ_{ct} for background events (blue) from data and MC signal events (red). The dotted black line depicts the applied cut. Events to right of this line are removed from the sample. The areas of each distribution have been normalized to one event.

Trigger Emulation and L3 Matching

To emulate the trigger, the information used to confirm the trigger must be the information that was available to the trigger at the time of the event, not the more accurate information available after offline reconstruction. For example, to know if a track has passed the requirement at Level 2 that its d_0 is greater than $120\mu m$, then it must be the SVT online impact parameter that satisfies this condition, not the offline impact parameter. During offline reconstruction, the tracks are matched to SVT tracks as part of standard reconstruction. The information from the SVT tracks is used to confirm the L2 trigger path. If the offline track has not been matched to an SVT track, then this track cannot be used to confirm the L2 trigger as it was not found by the SVT algorithms during data taking.

Trigger Criteria L1	**Units**	`BCHARMLOWPT`	`BCHARM`	`BCHARMHIGHPT`
Minimum Track P_T	GeV	2.0	2.0	2.5
Two Track Charge Product	-	na	-1	-1
Two Track Max $\Delta\phi$	degrees	90°	135°	135°
Two Track P_T scalar sum	GeV	0	5.5	6.5
Trigger Criteria L2				
Minimum d_0	μm	120	120	120
Maximum d_0	μm	1000	1000	1000
Minimum Track P_T	GeV	2.0	2.0	2.5
Maximum Track χ^2	-	15	15	15
Track Pair Charge Product	-	na	-1	-1
Maximum Pair $\Delta\phi$	degrees	90°	90°	90°
Minimum Pair $\Delta\phi$	degrees	2°	2°	2°
Minimum Pair P_T scalar sum	GeV	0	5.5	6.5
Minimum Pair L_{xy}	μm	200	200	200
Trigger Criteria L3				
Minimum d_0	μm	80	80	80
Maximum d_0	μm	1000	1000	1000
Minimum Track P_T	GeV	2.0	2.0	2.5
Maximum Track η	-	1.2	1.2	1.2
Track Pair Charge Product	-	na	-1	-1
Maximum Pair $\Delta\phi$	rad	1.571	1.571	1.571
Maximum Pair $\Delta\phi$	rad	0.035	0.035	0.035
Maximum Pair Δz_0	cm	5.0	5.0	5.0
Minimum Pair L_{xy}	μm	200	200	200
Minimum Sum P_T	GeV	4.0	5.5	6.5

Table 3.1: This table gives the description of the trigger cuts made at each level. 'na' stands for not applied.

The L3 trigger uses tracks reconstructed by algorithms specifically used by L3. The difference between L3 tracks and offline tracks is that the L3 tracks were found using only the Histogram linking algorithm, and no hits are dropped for a subsequent refit. The reason for this difference is that the processing time is faster by a factor of 2. There are no standard algorithms to match the L3 track to the offline track. As it is important to get the trigger response correct, in order to properly understand the bias, L3 matching is introduced as follows:

For each event the L3 track list can be accessed. This contains tracks only with $P_T \geq 1.2GeV$, however as the trigger only uses tracks with $P_T \geq 2GeV$, this does not present a problem. For each offline track, a match χ^2 is calculated between the offline and each L3 track. The match χ^2 depends on the P_T and ϕ of the track. It is defined as:

$$\chi^2 = (\frac{\Delta P_T}{\sigma_{P_T}^{L3}})^2 + (\frac{\Delta \phi}{\sigma_{\phi}^{L3}})^2 \tag{3.5}$$

where $\Delta P_T,\ \Delta\phi$ are the differences between the offline and L3 measured track P_T and ϕ respectively, and $\sigma_{P_T}^{L3}$, σ_{ϕ}^{L3} are the errors on the L3 track P_T and ϕ measurements.

The L3 track that has the smallest χ^2 is defined as the matched track. In the data analyzed there was no instance where the same L3 track was matched to more than one offline track.

Using the matched SVT and L3 tracks, the trigger can be reconfirmed on the event. If all the selection criteria and the trigger confirmation is passed, then the event is part of the sample on which the lifetime measurement is carried out.

Chapter 4

Lifetime Probability Distribution Function for Signal Events

In this chapter, the probability distribution function (PDF) is derived for the lifetime of signal events. The cuts applied by the trigger cause significant biases which can be accounted for solely by knowledge of decay kinematics. This chapter shows how this is indeed possible and considers all the detector effects that must be incorporated before the the calculations are complete enough to fit the lifetime of events detected at CDF.

4.1 Basic Lifetime Probabilities

B mesons are unstable particles and have mean lifetimes of order 1 picosecond. The lifetime distribution of a pure sample of $B^{\pm}$ mesons with ideal measurement accuracy is an exponential. This leads to a simple PDF and likelihood. For a series of measurements, the probabilities for each measured time t_i can be multiplied to give the likelihood given the mean decay time parameter τ. All lifetimes are the proper decay time in the rest frame of the B meson.

Before considering any trigger or detector effects, 3 hypothetical samples of B mesons are introduced, and the differences in their likelihoods are considered. The 3 samples are as follows:

- Sample 1: This is a pure sample where no selection criteria has been applied. The measured lifetime of the ith event is t_i
- Sample 2: In this sample, all events with lifetime lower than t_{low} and higher than t_{high} have been removed
- Sample 3: In this sample a different t_{low} and t_{high} are associated to each event. To avoid confusion, they are referred to as $t_{\min i}$ and $t_{\max i}$. $t_{\min i}$ and $t_{\max i}$ are different for every event (hence the label i). If an event's lifetime lies between its $t_{\min i}$ and $t_{\max i}$, the event is retained in the sample, otherwise it is discarded. The $t_{\min i}$ and $t_{\max i}$ are not dependent on the mean lifetime of the unbiased sample τ.

The lifetime distributions of these samples are shown in figure 4.1.

For sample 1, the probability density function, P, and likelihood for N events, $\mathcal{L}$, are very simple and given in equation 4.1 and 4.2 respectively. In the probability density function the observables of data and the parameters are separated by a semi-colon.

$$\mathrm{P}(t_i;\tau) = \frac{1}{\tau}e^{\frac{-t_i}{\tau}} \tag{4.1}$$

$$\begin{aligned} \mathcal{L} &= \prod_i^N \mathrm{P}(t_i;\tau) \\ \log\mathcal{L} &= -N\log\tau - \sum_{i=1}^{N}\frac{t_i}{\tau} \end{aligned} \tag{4.2}$$

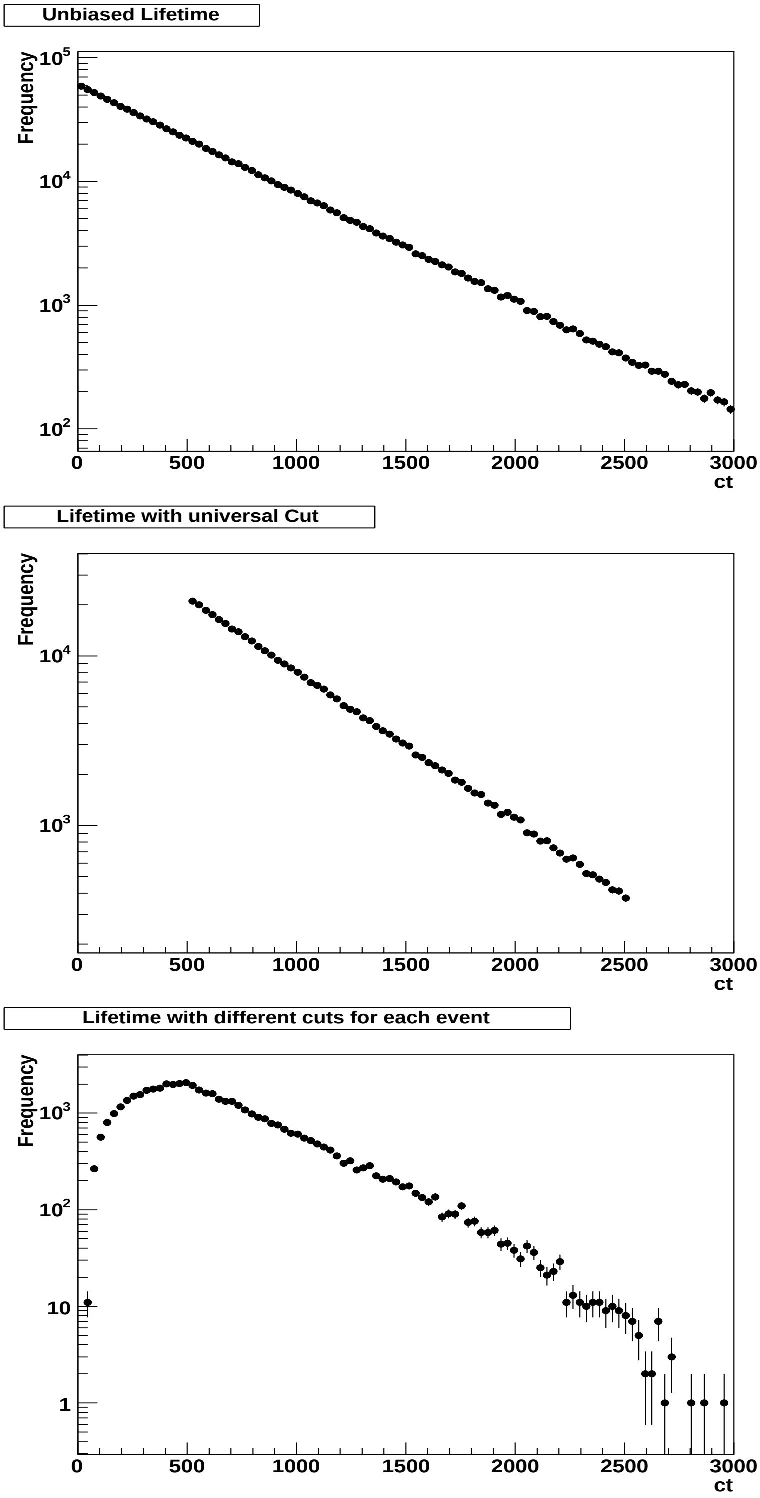


Figure 4.1: The lifetime distribution for the three samples described in section 4.1. The top plot is sample 1, the middle is sample 2 and the bottom plot is sample 3.

Consider now the effect of applying a single upper and lower lifetime cut as in sample 2. In this case, the probability to observe a lifetime at a given time must be normalized over the region accepted by the lifetime cuts. The changes to the PDF and likelihood are given in equation 4.3 and 4.4.

$$\begin{aligned}
\mathrm{P}(t_i;\tau) &= \mathrm{P}(t_i;\tau|t_i \in [t_{low}, t_{high}]) \\
\mathrm{P}(t_i;\tau) &= \frac{\frac{1}{\tau}e^{\frac{-t_i}{\tau}}}{\int\limits_{t_{low}}^{t_{high}} \frac{1}{\tau}e^{\frac{-t'}{\tau}}dt'} \\
\mathrm{P}(t_i;\tau) &= \frac{\frac{1}{\tau}e^{\frac{-t_i}{\tau}}}{e^{\frac{-t_{low}}{\tau}} - e^{\frac{-t_{high}}{\tau}}} \qquad (4.3)
\end{aligned}$$

$$\log\mathcal{L} = -N\log\tau - N\log(e^{\frac{-t_{low}}{\tau}} - e^{\frac{-t_{high}}{\tau}}) - \sum_{i=1}^{N}\frac{t_i}{\tau} \qquad (4.4)$$

Hence the difference between equations 4.2 and 4.4 is the extra middle term that depends on t_{low} and t_{high}.

Consider now sample 3, where there are differing lower and upper lifetime cuts for each event. The effects of the $t_{\min}$ and $t_{\max}$ proper lifetime selection criteria are shown in figure 4.1. It is apparent that the resulting distribution is significantly biased. The probability to find an event with decay time t_i is the product of the probability to find t_i given that t_i must lie between $t_{\min i}$ and $t_{\max i}$ and the probability that this particular $t_{\min i}$ and $t_{\max i}$ define the selection criteria for that event. For sample 3 the probability can be written as

$$\begin{aligned}
\mathrm{P}(t_i, t_{\min i}, t_{\max i}; \tau) &= \mathrm{P}(t_i; \tau | t_i \in [t_{\min i}, t_{\max i}]) \cdot \mathrm{P}(t_{\min i}, t_{\max i} | \, t_i) \\
&= \frac{\frac{1}{\tau} e^{\frac{-t_i}{\tau}}}{\int\limits_{t_{\min i}}^{t_{\max i}} \frac{1}{\tau} e^{\frac{-t'}{\tau}} dt'} \cdot \mathrm{P}(t_{\min i}, t_{\max i} | \, t_i) \\
&= \frac{\frac{1}{\tau} e^{\frac{-t_i}{\tau}}}{e^{\frac{-t_{\min i}}{\tau}} - e^{\frac{-t_{\max i}}{\tau}}} \cdot \mathrm{P}(t_{\min i}, t_{\max i} | \, t_i) \qquad (4.5)
\end{aligned}$$

where the index i labels the event, each of which has its measured decay time t_i and its upper and lower lifetime selection criteria $t_{\min i}$ and $t_{\max i}$. The probability includes dependence on the term $\mathrm{P}(t_{\min i}, t_{\max i} | t_i)$. If the assumption is made that the value of $t_{\min i}$ or $t_{\max i}$ do not depend on the measured lifetime, t_i, then $\mathrm{P}(t_{\min i}, t_{\max i} | t_i) \rightarrow \mathrm{P}(t_{\min i}, t_{\max i})$. In the log-likelihood, the sum over $\log(\mathrm{P}(t_{\min i}, t_{\max i}))$ is simply a constant that can be ignored. The log-likelihood function used to determine τ for a set of N ideal decays that have varying lifetime cuts is given by

$$\begin{aligned}
\log \mathcal{L} &= -N \log(\tau) \\
&\quad - \sum_{i=1}^{N} \left(\frac{t_i}{\tau} + \log\left(e^{-t_{\min i}/\tau} - e^{-t_{\max i}/\tau} \right) \right) \qquad (4.6)
\end{aligned}$$

Note that the difference between the likelihood function for events that have this kind of selection and those that have no selection at all is only

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

The upper lifetime cut has a dramatic effect on the precision with which a lifetime can be measured; finding an event with lifetime t_i contains less information if the range of possible t has been restricted by lifetime cuts. For example, an upper lifetime cut at twice the B mean lifetime loses only 14% of events. However, the statistical error of the measurement is increased by a factor of 2, equivalent to a signal loss of 75% [57].

4.1.1 The Effect of an Impact Parameter Based Trigger

The effect of an impact parameter based trigger is akin to the situation in Sample 3. Upper and lower impact parameter cuts directly translate into cuts on the proper decay length and hence on the lifetime of the decaying particle. The CDF triggers used in this analysis have as their primary requirement that there are two tracks with impact parameters in the range $120 - 1000$ μm.

The consequence of the trigger is best illustrated in the decay shown in figure 4.2. It depicts a B meson decay to two daughter particles, for example a kaon and a pion. In each of the diagrams the decay vertex has been translated to a different decay point, along the path of the B momentum. The kinematics of each track (i.e their momentum) have been held constant. If the B meson decays at a point corresponding to a short proper decay time, both the kaon and pion do not have impact parameters in the correct range, and so the event can not pass the trigger. If the decay occurs in the region where both tracks have impact parameters in the trigger range, then the event will pass the trigger. Finally, if the decay occurs at a longer time, one or both tracks will have impact parameters too large and will fail trigger selection.

Therefore, the impact parameter cuts have effectively defined a $t_{\min}$ and $t_{\max}$ cut for that event. For the event shown, $t_{\min}$ and $t_{\max}$ are not dependent on the actual decay time, t_i, as the decay kinematics do not depend on the decay time. This is exactly the situation described for sample 3 at the start of this section. Hence, the likelihood used to extract the lifetime for a sample of events collected by the impact parameter based trigger will be similar to equation 4.6.

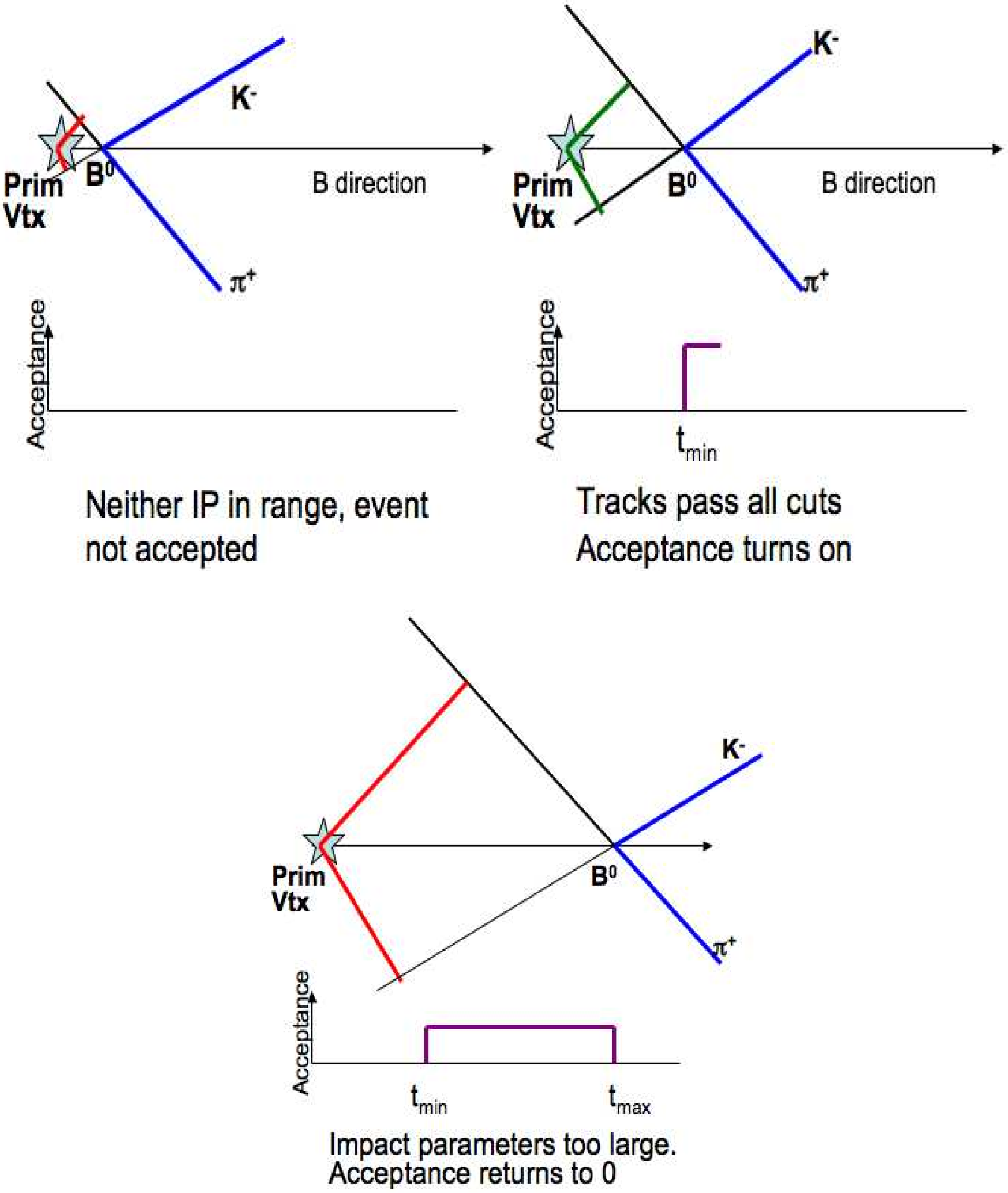


Figure 4.2: The diagrams here show how the impact parameter cuts directly translate to upper and lower ct cuts for the B meson

4.2 The Acceptance Function

The acceptance function, $A_i(t)$, unique to each event, is the probability that an event could be accepted by the trigger and selection cuts if it decays at proper time t. The acceptance function for the event in figure 4.2 is drawn, and is simply defined by t_{min} and t_{max} . If the PDF for each event is normalised over $A_i(t)$, then this is all that is required to correct for the bias induced by the trigger. The time between t_{min} and t_{max} is referred to as an interval.

To correctly calculate $A_i(t)$ all the selection requirements of the trigger and the analysis selection must be considered. To calculate $A_i(t)$ the decay vertex is translated along the B momentum direction. The kinematics of the decay are held fixed. At each decay point, which corresponds to a proper decay time, the properties of the decay are calculated to see if they meet all the selection requirements. If they do not, the value of $A_i(t)$ is zero.

For more complicated decay geometries, for example 3 final state daughters, the impact parameter cuts on the decay products can translate into a series of disjoint time-intervals (k)

$$\log \mathcal{L}_{\text{cut}} = -\sum_{i=1}^{N} \log \left(\sum_{k=1}^{n_i} e^{-\frac{t_{\text{min ik}}}{\tau}} - e^{-\frac{t_{\text{max ik}}}{\tau}} \right) \tag{4.7}$$

which changes the final piece of the likelihood as detailed in equation 4.7 and depicted in figure 4.3. These disjoint intervals can occur as the acceptance function can be viewed as the sum of the acceptance functions due to each track pair combination separately, and these will not necessarily overlap.

The correct calculation of $A_i(t)$ and the PDF is the subject of the remainder of this

Figure 4.3: Given the 3-momenta of all particles in the decay, and the decay lengths 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 function. For the depicted decay there are 2 intervals.

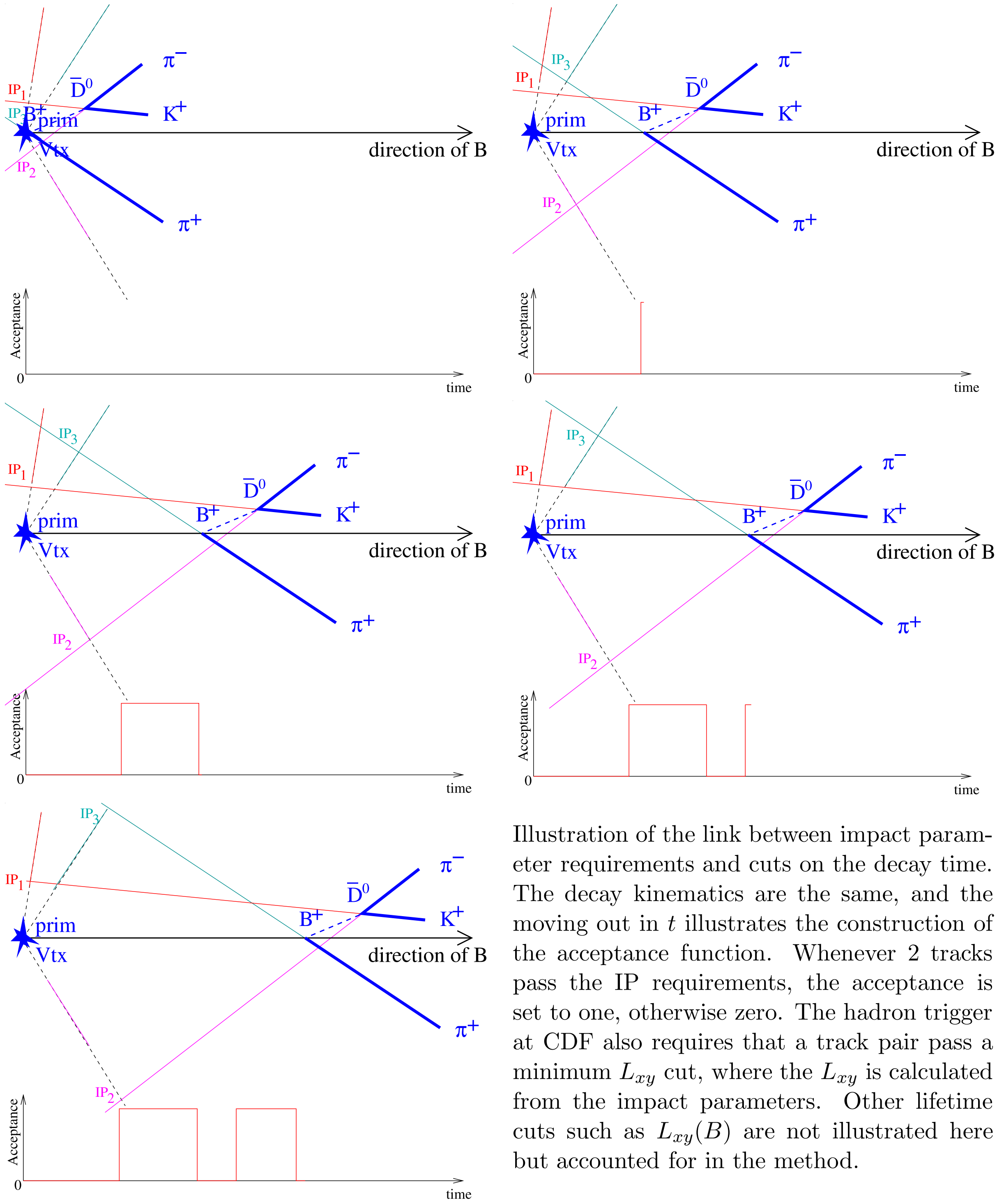


Illustration of the link between impact parameter requirements and cuts on the decay time. The decay kinematics are the same, and the moving out in t illustrates the construction of the acceptance function. Whenever 2 tracks pass the IP requirements, the acceptance is set to one, otherwise zero. The 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. Other lifetime cuts such as $L_{xy}(B)$ are not illustrated here but accounted for in the method.

chapter. To reach the final PDF, measurement errors and the track finding effieciencies must also be considered.

4.3 The PDF Including Measurement Errors

So far for the purpose of simplicity no measurement errors have been considered. In practice there are measurement uncertainties on all reconstructed and measured quantities. In addition, trigger selection is based upon the SVT measured quantities while the decay time is calculated after full reconstruction and vertexing and uses the offline measured impact parameters. These differ from the SVT measured impact parameters used by the trigger. The next few subsections detail how these measurement errors are included in the likelihood.

4.3.1 Incorporating Measurement Errors on the Decay Point

Imagine a trigger that makes its selection based on offline impact parameters and assume that there are only measurement errors on the proper lifetime.

The acceptance function is still a top-hat function or perhaps a series of them as shown in figure 4.3. It is as a function of the measured decay time, t, as opposed to the true decay time, t', which is unknown. Assuming Gaussian measurement errors, the probability to measure a decay time t_i can be written 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'_i, given its mean lifetime τ,

$$\mathrm{P}(t_i;\tau) = \frac{1}{\tau} e^{\frac{-t'_i}{\tau}}.$$

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

$$\mathrm{P}(t_i; \tau, \sigma_t) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t_i'-t_i)^2}{2{\sigma_t}^2}}.$$

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

$$A_i\left(t\right) = \sum_{\substack{k=\text{all}\\ \text{intervals}}} \left(\theta(t_i - t_{\min k}) - \theta(t_i - t_{\max k})\right)$$

where θ is the Heaviside function.

In terms of these parameters, the total probability is:

$$P\left(t_i; \tau, \sigma_t\right) \quad = \quad \frac{\int\limits_0^\infty \frac{1}{\tau} e^{\frac{-t_i}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t_i'-t_i)^2}{2{\sigma_t}^2}} A_i\left(t\right)|_{t=t_i} dt'}{\int\limits_{-\infty}^\infty dt \int\limits_0^\infty \frac{1}{\tau} e^{\frac{-t}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t'-t)^2}{2{\sigma_t}^2}} A_i\left(t\right) dt'} \tag{4.8}$$

4.3.2 Including the Difference Between Online and Offline Impact Parameters

The trigger uses fast, SVT-measured quantities for selection. So far, the acceptance function calculation uses offline quantities. A relation between the two must be established so that as the event is slid along its momentum direction the recalculated offline impact parameter at the new decay point are related to what the SVT would measure. For the purposes of this section, it is assumed that all tracks are found by the SVT.

To correctly calculate the acceptance function, d_0^{SVT} , which is the impact parameter calculated by the SVT must be estimated for each track at each possible decay point. The difference in impact parameter as measured by the SVT and the subsequent offline reconstruction, $(\Delta d_0)_{L2}$ where $(\Delta d_0)_{L2} = d_0^{\mathrm{SVT}} - d_0^{off}$, is shown in figure 4.4.

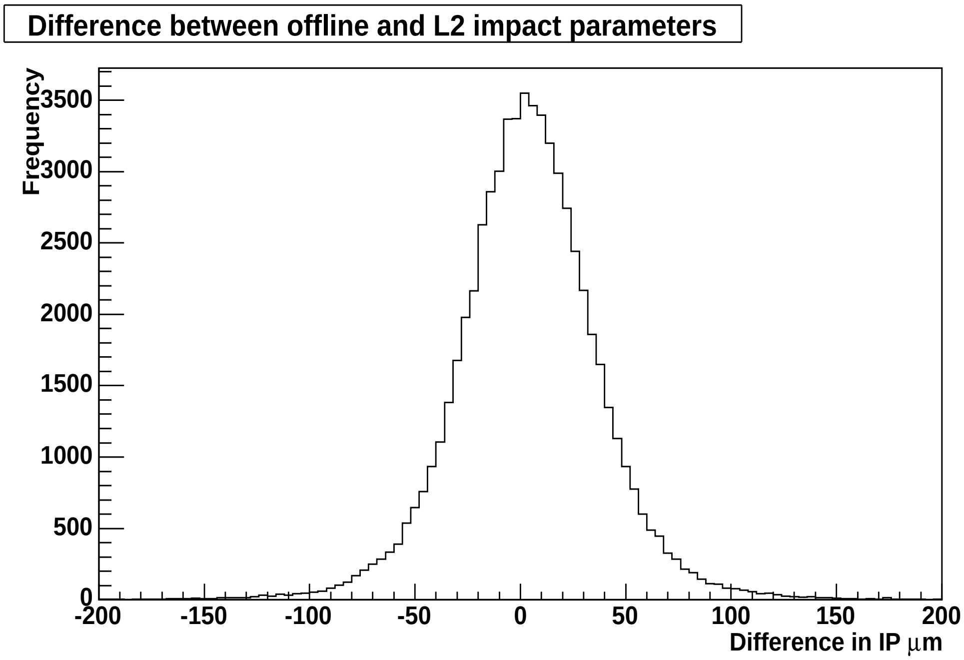


Figure 4.4: The difference between the offline and SVT measured impact parameter in data. All tracks where SVT information is available are used.

Using $(\Delta d_0)_{L2}$

The difference between the offline impact parameter, d_0^{off}, and the SVT impact parameter, d_0^{SVT}, is assumed to be independent of the decay lifetime. The distribution of $(\Delta d_0)_{L2}$ is shown for events with $ct > 550\ \mu m$ and $ct < 550\ \mu m$ in figure 4.5. The figure demonstrates that this is a valid assumption. The relative offset for one track, at the real point of decay, $(\Delta d_0)_{L2}$, is assumed to be the offset for that track at all possible decay times. The SVT impact parameter for a track is estimated as

$d_0^{SVT} = d_0^{off} + (\Delta d_0)_{L2}$. The acceptance function is now calculated by using the estimated value of d_0^{SVT} to see if the trigger conditions have been met. This is shown in figure 4.6. The acceptance function retains is basic top hat function style shape, but the values for $t_{\min}$ and $t_{\max}$ change.

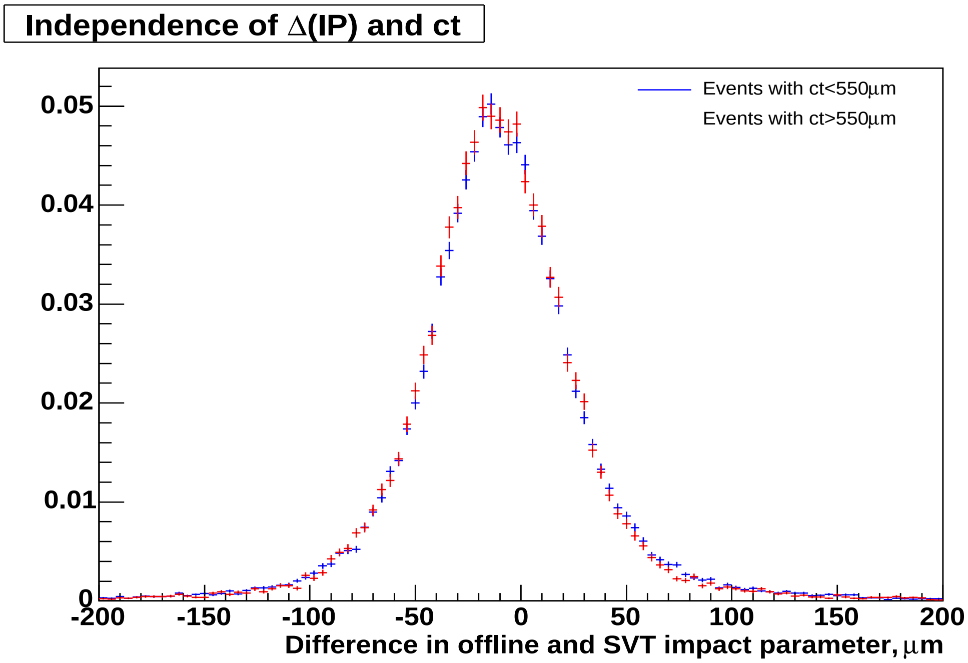


Figure 4.5: The difference between the offline and SVT measured impact parameter where the real data has been split into two regions of *ct*. No significant difference is observed between the two distributions and hence Δd_0 is independent of the B lifetime

One of the advantages of the simulation free lifetime measurement method is that it is not necessary to know the actual impact parameter error distribution. The method is also capable of handling systematic shifts in the impact parameter measurement provided that those shifts are uniform within the allowed impact parameter range. This means that this method of accounting for the trigger bias is more robust and requires a less detailed consideration of the SVT performance than any simulation based method.

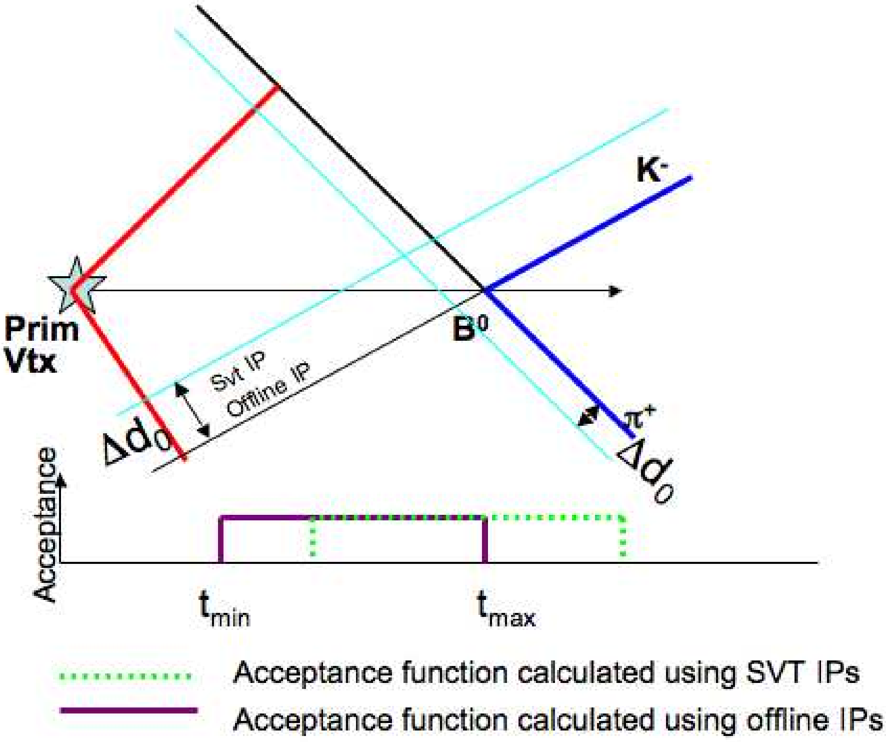


Figure 4.6: The diagram here shows how using the difference between the SVT measured and offline impact parameters can be used to calculate the acceptance function taking into account what the SVT would have measured

The Discretised SVT d_0

One further step is to incorporate the fact that the SVT algorithm used to fit track parameters returns impact parameters in multiples of 10 μm. A typical SVT-measured impact parameter distribution in Monte Carlo for values between 200 μm and 800 μm is shown in figure 4.7. Although steps of 10 μm are small compared to the typical resolution of about 35 μm [48], it turns out that ignoring this effect has significant effects on the fitted lifetime. This effect is demonstrated using realistic Monte Carlo events: Using 15000 events, a decay length of $ct = 448 \pm 6$ μm is measured for a sample where the input value is 496 μm. This is a 10 σ discrepancy. The effect is easy to take into account by modifying the calculation of d_0^{SVT}, to

$$d_0^{SVT} = [\text{nearest multiple of 10 of}](d_0^{off} + (\Delta d_0)_{L2}) \tag{4.9}$$

With this modification and using the same events as before, the fitted lifetime is measured to be 494±7 μm which is in good agreement with the input value of 496 μm.

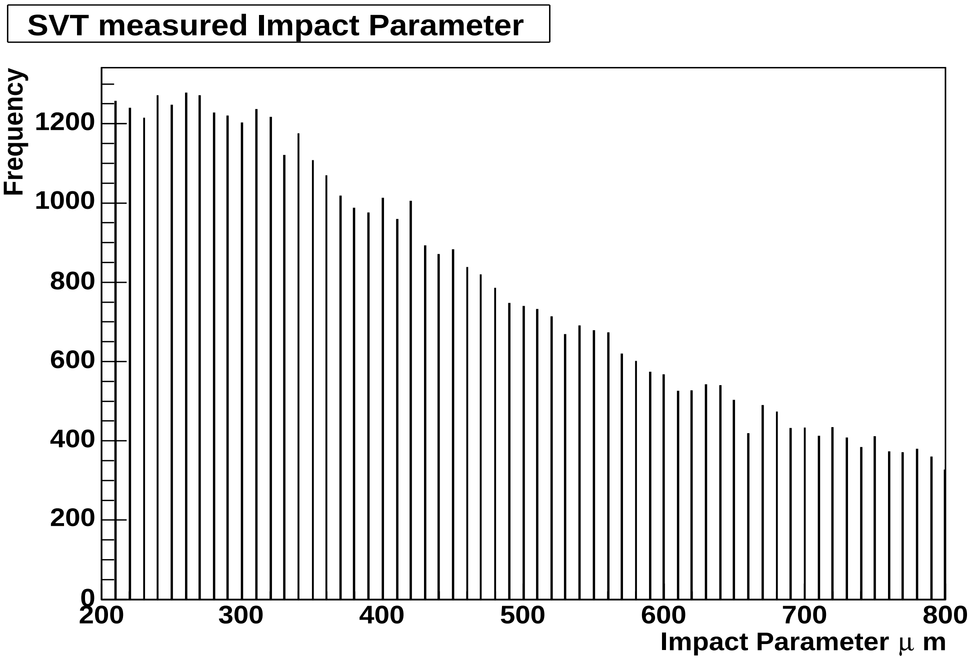


Figure 4.7: The plot shows that impact parameters are only measured to the closest 10 by the SVT using realistic Monte Carlo. For clarity only the region between 200 and 800 microns is shown.

Level 3 Impact Parameters

As the Level 3 trigger uses a reduced version of offline reconstruction, there is a difference between the L3 impact parameters, d_0^{L3}, and d_0^{off}. The same technique used in section 4.3.2 is applicable here where $(\Delta d_0)_{L3}$ measured at the point of decay is kept constant. Therefore $d_0^{L3} = d_0^{off} + (\Delta d_0)_{L3}$.

4.3.3 The PDF Incorporating SVT Measurement Errors

The acceptance function is now calculated using the estimated values of d_0^{SVT} and d_0^{L3} for each track at each decay point. The acceptance function is non zero where the L2 trigger requirements are satisfied by the track d_0^{SVT} and the L3 trigger requirements are satisfied by the track d_0^{L3} and all other selection criteria are also satisfied.

The acceptance function still has the top hat shape but the values of each t_{min} and t_{max} are different to what would have been calculated using the offline values to satisfy the trigger conditions.

Therefore 4.8 becomes:

$$P\left(t_i;\tau,\sigma_t\right) \quad = \quad \frac{\int\limits_0^\infty \frac{1}{\tau} e^{\frac{-t_i}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t'_i-t_i)^2}{2\sigma_t{}^2}}\, dt' \times A_i(t)|_{t=t_i}}{\sum\limits_{\substack{k=\text{all}\\ \text{intervals}}} \int\limits_{t_{\text{min k}}}^{t_{\text{max k}}} dt \int\limits_0^\infty \frac{1}{\tau} e^{\frac{-t}{\tau}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t'-t)^2}{2\sigma_t{}^2}} \times A_i(t)\, dt'} \tag{4.10}$$

Using the frequency function

$$\mathrm{F}(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{\frac{-y^2}{2}}\, dy \tag{4.11}$$

this can be written as

$$\begin{aligned} P\left(t_i;\tau,\sigma_t\right) &= \frac{\frac{1}{\tau}e^{\frac{-t_i}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t_i}{\sigma}-\frac{\sigma}{\tau}\right)A_i(t)|_{t=t_i}}{\sum\limits_{\substack{k=\text{all}\\ \text{intervals}}}\int\limits_{t_{\text{min k}}}^{t_{\text{max k}}}\frac{1}{\tau}e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t}{\sigma}-\frac{\sigma}{\tau}\right)dt} \\ &= \frac{\frac{1}{\tau}e^{\frac{-t_i}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t_i}{\sigma}-\frac{\sigma}{\tau}\right)\times A_i(t)|_{t=t_i}}{\sum\limits_{\substack{k=\text{all}\\ \text{intervals}}}\left[-e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t}{\sigma}-\frac{\sigma}{\tau}\right)+\mathrm{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\text{min k}}}^{t=t_{\text{max k}}}} \end{aligned} \tag{4.12}$$

Dividing both numerator and denominator by $e^{\frac{1}{2}\frac{\sigma^2}{\tau^2}}$ makes the formula numerically robust. The PDF becomes:

$$P\left(t_i;\tau,\sigma_t\right) = \frac{\frac{1}{\tau}e^{\frac{-t_i}{\tau}}\mathrm{F}\left(\frac{t_i}{\sigma}-\frac{\sigma}{\tau}\right)\times A_i(t)|_{t=t_i}}{\sum\limits_{\substack{k=\text{all}\\ \text{intervals}}}\left[-e^{\frac{-t}{\tau}}\mathrm{F}\left(\frac{t}{\sigma}-\frac{\sigma}{\tau}\right)+e^{-\frac{1}{2}\frac{\sigma^2}{\tau^2}}\mathrm{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\text{min k}}}^{t=t_{\text{max k}}}} \tag{4.13}$$

4.4 Considerations Regarding the SVT Track Finding Efficiency

In the calculation of the PDF in equation 4.13, it is assumed that all tracks are found by the SVT. As the SVT tracks are found during data taking and with limited time, it is not possible to find all tracks. Variations in the single track finding efficiency may cause a lifetime bias in the sample if, for example, tracks with certain impact parameters are more likely to be found, as there is a correlation between impact parameter and lifetime. Any large variations in the single track finding efficiency that will cause large biases in the measured result must be incorporated into the PDF to keep the uncertainty on the lifetime measurement small.

4.4.1 Determining the Single Track Finding Efficiency as a Function of the Impact Parameter

The most important variation to consider is that of the the single track finding efficiency as a function of d_0, as d_0 is a quantity that changes as the event is slid along its ct. The realistic MC is not used to find the single track finding efficiency, as the sample does not contain a simulation of run conditions over the whole period of data taking. There are known differences between the response of the real and simulated detector.

In data, the single track efficiency cannot be found by considering all tracks, as this is a biased sample. As all the tracks in the data sample belong to events that have passed the trigger, at least two of them must have been found by the SVT. To obtain a sample of unbiased tracks, the following procedure is carried out using the track nomenclature defined in section 3.3.2: A sub sample of events is found that can pass the trigger using the daughter tracks of the D meson alone (π_D, k_D). For those events, the pion from the B meson (π_B) can be considered an unbiased set of tracks. For these tracks, the efficiency, which is the fraction of offline tracks that are found by the SVT, can be plotted. Further samples of unbiased tracks can be found by considering the π_D track in events where the trigger can be passed by the π_B and k_D tracks, and considering the k_D track in events where the trigger can be passed by the π_B and π_D tracks alone.

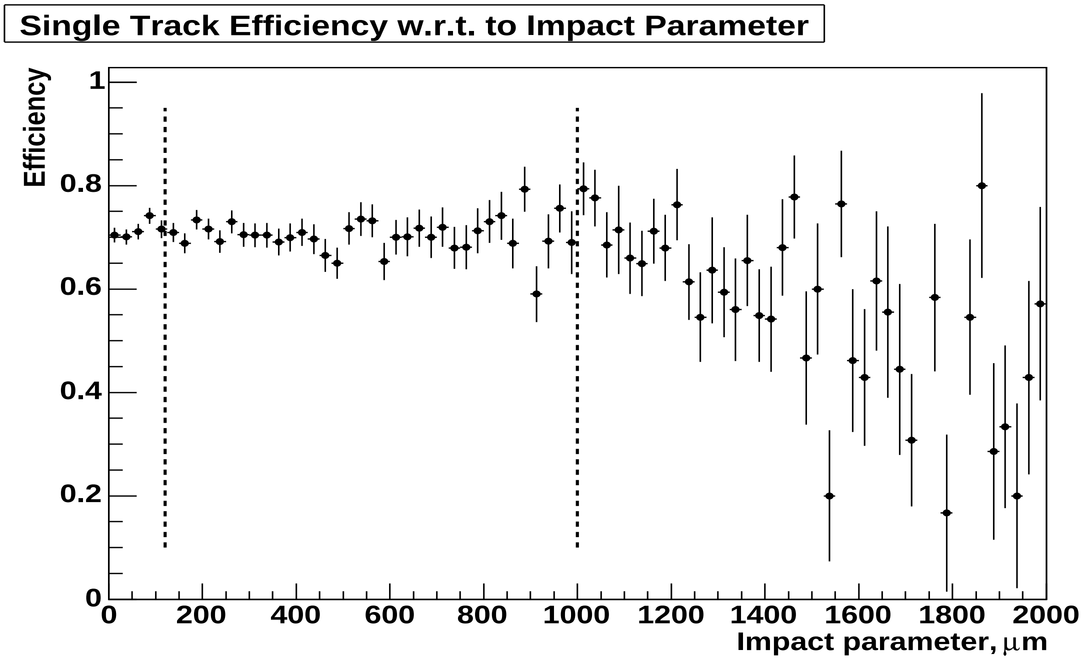


Figure 4.8: The single track finding efficiency as a function of the impact parameter, determined using data and the procedure described in section 4.4.1.

The single track efficiency is plotted in figure 4.8. As expected, the efficiency is not uniform and there is also a sharp fall in the efficiency after an impact parameter of 1000 μm. The track finding efficiency is approximately 70% for tracks with impact parameter less than 1000 μm.

4.4.2 Complications for more than Two Tracks

A complication that arises is that with three or more tracks there is the possibility that more than one track pair pass the trigger selection criteria. As the decay is slid along, there will be regions in *ct* where one track pair is available for triggering and others where there are two. In general, the trigger efficiency should be higher if two track pairs satisfy the trigger requirement rather than only one. This is because for a given single track finding efficiency 70%, the probability of finding two tracks out of three is higher than finding two tracks out of two. For the previously assumed case of track finding efficiency at 100% in section 4.3.2, there was no need to consider this scenario, as the event either fulfilled the selection criteria or it did not. Now, even if the track pair does fulfill the trigger selection criteria, the probability of whether or not those tracks required to pass the trigger were actually found by the SVT must be considered. This results in the calculation of the trigger efficiency, which is the probability that a sufficienct number of tracks, which when combined can pass the trigger, are found. The trigger efficiency is the value of the acceptance function at any time t. For the previous calculation of the PDF, and in the diagrams, this has been 1.

As a digression, it is noted that, if the SVT track finding efficiency is independent of impact parameter, the acceptance function can be calculated using only the tracks that were actually found. Given the found tracks, the trigger efficiency is either 1 or 0 no matter how many tracks are available for triggering.

However, the clear dependence between impact parameter and finding efficiency prevents this simple modification. After an impact parameter of 1000 μm, there is a rapid drop in efficiency. In a three body decay, for example, the situation can arise where two tracks with impact parameter 120 $\mu m < |d_0| <$ 1000 μm pass the trigger while the third track has an impact parameter beyond 1000 μm. At this point, the SVT track finding efficiency will be close to 0 and so the third track is not found. As the event is slid along t, at some point the impact parameter of the third track will be less that 1000 μm. Now, it could potentially play a role in the trigger decision. It cannot be assumed that as the track was not found by the SVT at its original position, it will not be found at the new position.

There are two pieces of information that would need to be available before knowing the trigger efficiency at this ct:

- What is the probability that the track that is originally at high impact parameter is found by the SVT once the decay point has been moved so that the impact parameter is now in the trigger range?
- What would the SVT-measured impact parameter be?

4.4.3 Possible Ways of Avoiding SVT Efficiency Complications when Final States have more than 2 Tracks

It should be noted that neither of the complications discussed in section 4.4.2 arise when there are only two tracks in a decay. This is because the trigger requires the IP to be below 1000 μm where the efficiency is flat. Since no track with $|d_0| > 1000 \mu m$ appears in the sample it does not matter that the efficiency is not flat beyond this point.

There are two ways that the complications in section 4.4.2 can be avoided. These are to reduce every decay to a two track decay by calculating the acceptance function using one track pair only; or by applying a cut on all tracks to have $d_0 < 1000\ \mu m$. The advantages and disadvantages of each solution are discussed below:

- "Two tracks": Treat a multi-body decay like a two body decay by declaring two tracks as trigger tracks and re-applying the trigger condition using these two tracks only. This reduces the size of the sample. If this particular track pair does not fulfill the trigger criteria then the event is removed from the sample. The PDF used to extract the lifetime is equation 4.13, where there will only be one interval to consider as there is only one track pair. However, the loss in statistics greatly reduces the precision of the measurement, and so this approach is not advantageous. It should be noted that the sample size cannot be increased by changing which decay products will be the trigger tracks based on which pair will pass the trigger for a particular event. The choice of which tracks are the trigger tracks must be the same for all events in the sample.

- "Fiducial cut": Use all tracks in the trigger decision, but impose a cut requiring *all* of them to have an impact parameter $|d_0| < 1000\mu m$. The effect of this cut on the acceptance as a function of *ct* can be calculated in the same way as that of the other impact parameter cuts: No track with $|d_0| > 1000\mu m$ affects the calculation of the trigger acceptance, and it does not matter how the SVT single-track finding efficiency looks like beyond $1000\mu m$ so equation 4.13 can be used to calculate the probability. This solution is not very costly in the number of events, but since it reduces the width of the acceptance function it significantly reduces the statistical power per event.

Therefore both these simple solutions significantly reduce the precision of the measurement, and another approach must be considered. These simple solutions have been detailed here to illustrate the the complications more fully.

4.5 Solving the >2 Track Problem

The two proposed solutions in section 4.4.3 for dealing with decays with more than 2 tracks in the final state are unsuitable for maximizing precision. Another solution that would not suffer from loss in precision is described in this section. This approach involves finding an answer to the following questions:

- For a track that is not found by the SVT what would the SVT-measured d_0 have been, had the track been found?

- As a function of impact parameter how likely is it that a track is found by the SVT ?

The information can then be used to calculate the absolute value of the trigger efficiency, which will vary depending on how many tracks pass the trigger requirements at a given ct.

4.5.1 Assigning a Value for the SVT d_0 for those Tracks with no SVT Match

For tracks found by the SVT, the SVT d_0 at a given value of ct is estimated from the offline d_0 and the difference between the measured SVT and offline impact parameter of the track at its original position. For tracks with no match, $(\Delta d_0)_{L2}$ must be assigned. The first step is to plot the $(\Delta d_0)_{L2} \equiv d_0^{\mathrm{SVT}} - d_0^{\mathrm{off}}$ distribution for those tracks where this information is available. This histogram is shown in figure 4.4. Secondly, for all tracks without d_0^{SVT} , a random value of $(\Delta d_0)_{L2}$is drawn from this plot. This is used to calculate d_0^{SVT} for that track. This process answers the first question: what would d_0^{SVT} be for tracks that were not found by the SVT.

4.5.2 Absolute Trigger Efficiency from the SVT Single Track Finding Efficiency

As tracks are included from the region where the SVT efficiency is not flat, the simple calculation of trigger efficiency given that the SVT found exactly those tracks cannot be used as this condition is no longer independent of ct. The trigger efficiency can no longer be 0 or 1. Instead it will depend on the number of tracks available for the

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	$3\varepsilon_s^2 - 2\varepsilon_s^3$

Table 4.1: Trigger efficiency in terms of the SVT-single track finding efficiency, for a three particle final state.

trigger decision, and the probability that the SVT finds those tracks given their impact parameters. In order to decide which track combinations could have fired the trigger, the d_0^{SVT} is needed for all tracks involved, including those which were not actually found by the SVT. For the tracks not found in the SVT, the d_0^{SVT} values generated from the measured Δd_0 distribution, as described in section 4.5.1 above, are used.

In order not to have to model the complicated turn-off curve of the SVT efficiency near $|d_0^{\mathrm{off}}| = 1000\ \mu m$, the SVT single track finding efficiency is approximated as follows: it is described as flat for $|d_0^{\mathrm{off}}| < 1000\ \mu m$ and zero elsewhere. For this to be accurate, the tracks with $|d_0^{\mathrm{off}}| > 1000\ \mu m$ are treated as having not been found by the SVT. The effect of the small drop in efficiency between $0 < |d_0^{\mathrm{off}}| < 1000\ \mu m$ is considered in the systematic error studies. With this simple form, the SVT single-track finding efficiency as a function of impact parameter is described by a single parameter, the SVT single track finding efficiency for tracks with $|d_0^{\mathrm{off}}| < 1000\ \mu m$, ε_s. At each given ct, for each given track, the SVT single track finding efficiency is either 0 or ε_s. This is the information required to answer the second question: what is the probability that a track is found by the SVT. The total trigger 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 trigger efficiency for the three track case are given in

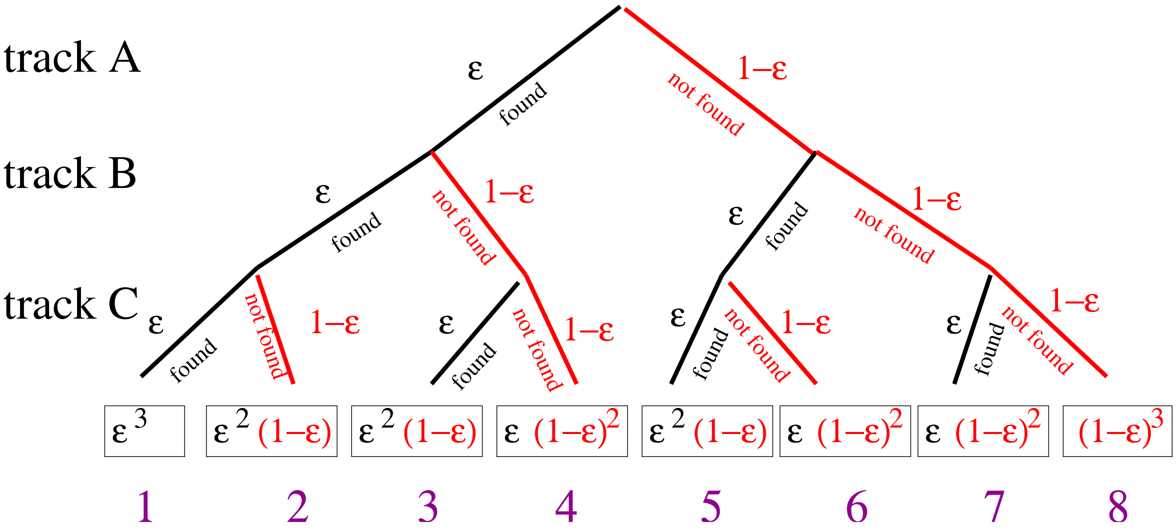


Figure 4.9: Decision tree for the example of 3 tracks in the final states. Each track can either be found (probability ε_s) or not be found (probability $1-\varepsilon_s$) by the SVT, giving $2^3 = 8$ possible combinations. The total trigger efficiency is calculated by adding up the individual probabilities of those combinations that pass the trigger cuts.

Table 4.1. For 4 or more tracks in the final states, this is a bit more complicated, as there is the need to distinguish two possible ways in which 2 track pairs could pass the trigger: the pairs may or may not have a track in common. In the computer program calculating those efficiencies, this is handled in the most general way, allowing it 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 are 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. This process is illustrated for the three-track case in figure 4.9, where the tracks are labeled A, B and C, from which the results listed in table Table 4.1 can be read off in the following way:

- 1 pair: If for example only the track pair (A,B) passes the trigger cuts. The probability required is the probability that both tracks A and B are found. To calculate this combinations 1 and 2 are added together, 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 probability required is that at least one of the two track pairs is found. The possible combinations are $1, 2$ and 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). In this case the probability required is that at least one of the 3 track pairs is found. The combinations $1, 2, 3$ and 5, are added, giving $3\varepsilon_s^2 - 2\varepsilon_s^3$.

With the trigger efficiency calculated the final acceptance function can be calculated that takes into account the SVT track finding efficiency. Whereas before the acceptance was either 0 or 1 (as the efficiency was assumed to be 100%), now it can take any of the values from Table 4.1. The value it has at any given *ct* depends on the number of track pairs that satisfy the trigger criteria at that time. A sample acceptance function is shown in figure 4.10.

4.5.3 Extracting the Value of ε_s

The method described above requires the value of the SVT single track finding efficiency so that the trigger efficiency can be calculated. $\varepsilon_s$can be fitted simultaneously with the lifetime. The extra information required to fit the single track finding efficiency is the number of tracks found in each event, relative to the minimum of 2 required to pass the trigger. For the three track case, it is the frequency of finding two tracks in the SVT versus three, for those events where all three tracks are within the SVT's reach,

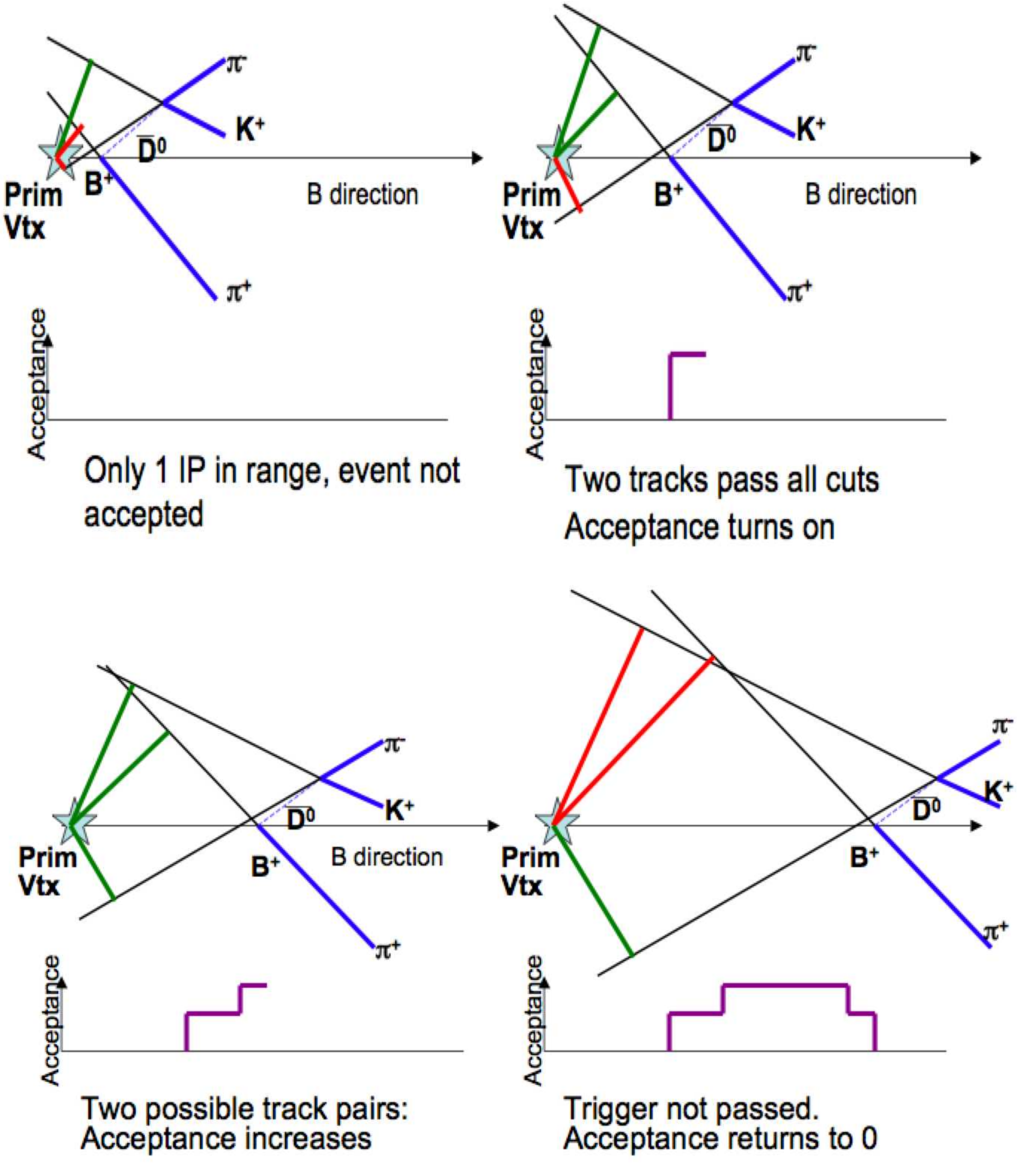


Figure 4.10: The diagrams here show how the acceptance function calculation changes when the SVT single track finding efficiency is included.

i.e, have $|d_0^{\text{off}}| < 1000\ \mu m$ and a minimum P_T greater than 2 GeV. The probabilities associated with those track configurations are exactly those at the end of the decision tree in figure 4.9. The probability is given by

$$P(trk_i;\varepsilon_s|t_i) = \frac{\varepsilon_s^r(1-\varepsilon_s)^{n-r}}{A_i(t)|_{t=t_i}} \tag{4.14}$$

where r is the number of tracks found, n is the number of tracks within the reach of the SVT and the denominator provides the correct normalisation. Note that the probability that any r out of n tracks are found is not considered, but rather that those specific r tracks that have SVT matches are found, and that the others are not.

4.6 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 4.13 is sufficient to fit a lifetime to an SVT-biased signal sample. For multi-body decays, the PDF needs to modified to take into account the above considerations. The following definitions are used:

- $P(\text{trk};\varepsilon_s|t_i)$: The probability to find the actual track combination. It is given in equation 4.14.
- $P(\text{trigger}|trk_i,t_i)$ The probability that the given track configuration fires the trigger, given the impact parameters, etc, calculated for the measured decay time t_i , using the sliding method. This is either 1 or 0.
- $P(\text{trigger}|\varepsilon_s,t_i)$: The probability that the trigger fires, given ε_s, but summed over

all possible track combinations that could have fired the trigger, $P(\text{trigger}; \varepsilon_s | t_i) = \sum_{\text{trk}} P(trk_i; \varepsilon_s | t_i) P(\text{trigger} | trk_i, t_i)$. This corresponds to the entries in Table 4.1. It is the normalization factor to go with $P(\text{trk}; \varepsilon_s | t_i) P(\text{trigger} | trk_i, t_i)$

- $\text{poly}_k(\varepsilon_s)$: Since $P(\text{trigger} | \varepsilon_s)$ is constant for t_i within one time interval with constant track configuration, it can be replaced by $\text{poly}_k(\varepsilon_s)$, where the index k labels the time interval, and $\text{poly}_k(\varepsilon_s)$ is one of the polynomials in table 4.1.

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

$$\begin{aligned} P\left(t_i; \tau, \sigma_t, \varepsilon_s\right) &= \frac{P(\text{trk}|\varepsilon_s) P(\text{trigger}|\text{trk}, t_i) \frac{1}{\tau} e^{\frac{-t_i}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t_i}{\sigma} - \frac{\sigma}{\tau}\right) A_i(t)|_{t=t_i}}{\sum\limits_{\text{all trk}} \sum\limits_{\substack{k=\text{all} \\ \text{intervals}}} \int\limits_{t_{\text{min k}}}^{t_{\text{max k}}} P(\text{trk}|\varepsilon_s) P(\text{trigger}|\text{trk}, t_i) \frac{1}{\tau} e^{\frac{-t}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) dt} \\ &= \frac{\varepsilon_s^r (1 - \varepsilon_s)^{n-r} \frac{1}{\tau} e^{\frac{-t_i}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t_i}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum\limits_{\substack{k=\text{all} \\ \text{intervals}}} \text{poly}_k(\varepsilon_s) \left[-e^{\frac{-t}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right) + \mathrm{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\text{min k}}}^{t=t_{\text{max k}}}} \qquad (4.15) \end{aligned}$$

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

Chapter 5

The PDF for a mixture of signal and background events

This chapter introduces the full PDF when background is included. The implications of certain terms and assumptions that have to be made are introduced. The chapter finishes by describing the particular parameterizations that are chosen to describe the data.

5.1 Addition of Background

In the previous chapter the PDF for signal events alone was derived and is given in equation 4.15. To extract a lifetime from data, the PDF must also have a component that describes the background. In the signal PDF, the probability to find a lifetime given the acceptance function A_i is calculated, where each A_i is calculated from the decay kinematics that translate the trigger cuts into different lifetime cuts for each event. The PDF is written as

$$\mathrm{P}(t_i, A_\mathrm{i}; \tau) = \mathrm{P}(t_i; \tau | A_\mathrm{i})\mathrm{P}(A_\mathrm{i}) \tag{5.1}$$

P($A_{\rm i}$) will be a function of the decay kinematics, i.e, track P_T, opening angles, relative position of the B and D vertex etc. When considering signal alone it is possible to ignore the term P($A_{\rm i}$) as it is simply a factor and has no dependence on t_i. If background is added, then the full expression changes to

$$P(t_i, A_{\rm i}; \tau) = P(s)P(t_i; \tau|A_{\rm i}, s)P(A_{\rm i}|s) + P(b)P(t_i|A_{\rm i}, b)P(A_{\rm i}|b) \tag{5.2}$$

where $P(s)$ is the signal probability (the signal fraction) and $P(b) = 1 - P(s)$ is the background probability. In this case the terms $P(A_{\rm i}|s)$ and $P(A_{\rm i}|b)$ only factor out if they are the same for signal and background. If they are different for signal and background, ignoring these factors will effectively give a wrong signal fraction when used in the fit. This statement is made more obvious by rewriting equation 5.2 as

$$\begin{aligned} P(t_i, A_{\rm i}; \tau) &= P(s)P(A_{\rm i}|s)P(t_i; \tau|A_{\rm i}, s) + P(b)P(A_{\rm i}|b)P(t_i|A_{\rm i}, b) && (5.3) \\ &= \{P(s|A_{\rm i})P(t_i|A_{\rm i}, s) + P(b|A_{\rm i})P(t_i|A_{\rm i}, b)\} \, P(A_{\rm i}) && (5.4) \end{aligned}$$

The two above expressions are equivalent and set out two ways of approaching the full PDF. Either the probability to find a particular acceptance function given that the event is signal must be calculated as in equation 5.3, or, alternatively, the individual event signal probability given the acceptance function must be calculated as in equation 5.4. The final factor in equation 5.4 factors out.

If the probability for observing t_i, depends on other observables of the event such as $A_{\rm i}$ or σ_{ti} then the probability for observing that particular $A_{\rm i}$ or σ_t must also be considered. Ignoring such terms can lead to a bias and has been discussed in [58]. For this analysis, it is possible to write a PDF that includes using the observed σ_{ti} as well as the $A_{\rm i}$ however correlations between these two variables introduce significant complexity and therefore this method uses an average σ_t to calculate the probability for observing t_i.

5.2 The Full Likelihood

The following notation is used:

- $P(A)$ "probability of A"
- $P(A,B)$ "probability of A and B"
- $P(A|B)$ "probability of A given B"
- $P(AorB)$ "probability of A or B"

The probability of A and B is given by

$$P(A,B) = P(A)P(B|A) \tag{5.5}$$

and this leads to Bayes' Theorem:

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} \tag{5.6}$$

The probability of A or B is

$$P(A \text{ or } B) = P(A) + P(B) - P(A,B) \tag{5.7}$$

The following measured observable quantities are in the fit:

- The measured lifetime, t_i .
- The acceptance function A_{i} , calculated from the decay kinematics and the trigger cuts.
- The mass, m_i .
- The track-configuration observed, trk_i . This configuration is dependent on which tracks are found by the SVT. It is necessary to retain this information so that the single track finding efficiency can be extracted from the fit.

The data sample contains only those events that have passed the trigger. The probability that an event has a particular t_i , m_i , A_{i} and trk_i , given that it has passed the trigger, $P\left(t_i, m_i, trk_i, A_{\mathrm{i}}; \tau|\text{trigger}\right)$, is the probability that needs to be considered. The probability will depend on other parameters in addition to τ, such as $\varepsilon_s$and σ_t. In addition there will be further parameters that describe the mass and background lifetime distributions. Given that τ is the parameter of interest, the dependence on τ is explicitly written. Dependence on other parameters is not explicitly written.

It is necessary to remember the distinctions between $P(A_{\mathrm{i}})$ and $P(\text{trigger})$. $P(\text{trigger})$ is the probability that the trigger accepts the event. $P(A_{\mathrm{i}})$ depends only on decay kinematics and is the probability to find an event where the decay kinematics translate the trigger cuts to the given acceptance function. The term $P(\text{trigger}|A_{\mathrm{i}})$ is the probability that a decay with these kinematics passes the trigger. It is important to distinguish these as $P(A_{\mathrm{i}})$ does not depend on the mean lifetime whereas $P(\text{trigger})$ and $P(\text{trigger}|A_{\mathrm{i}})$ do depend on the mean lifetime.

To incorporate background, $P\left(t_i, m_i, trk_i, A_{\mathrm{i}}|\text{trigger}\right)$ is split into a signal and a background part. The letters s and b are used for signal and background. Using equation 5.7, $P(s,b)=0$, and $P(s \text{ or } b)=1$:

$$\begin{aligned} P\left(t_i, m_i, trk_i, A_{\mathrm{i}}; \tau|\text{trigger}\right) &= P\left(s, t_i, m_i, trk_i, A_{\mathrm{i}}; \tau|\text{trigger}\right) \\ &+ P\left(b, t_i, m_i, trk_i, A_{\mathrm{i}}|\text{trigger}\right) \end{aligned} \tag{5.8}$$

Considering the first term on the right hand side in equation 5.8, only:

$$P\left(s, t_i, m_i, trk_i, A_{\mathrm{i}}; \tau|\text{trigger}\right) \tag{5.9}$$

Using equation 5.6 on equation 5.9 gives:

$$\begin{aligned} &P\left(s, t_i, m_i, trk_i, A_{\mathrm{i}}; \tau|\text{trigger}\right) \\ &= \frac{P\left(s, t_i, m_i, trk_i, A_{\mathrm{i}}; \tau\right) P\left(\text{trigger}|s, t_i, m_i, trk_i, A_{\mathrm{i}}\right)}{P(\text{trigger}; \tau)} \end{aligned} \tag{5.10}$$

The term $P\left(\text{trigger}|s, t_i, \sigma_t, m_i, trk_i, A_{\mathrm{i}}\right)$ is either 1 or 0, because the trigger decision is completely determined by the acceptance function, the decay time, and which tracks have actually been found by the SVT. The denominator in equation 5.10 is the probability that the trigger accepts the event. It is more useful to rewrite this in terms of the probability the trigger accepts the event given the acceptance function, $P(\text{trigger}; \tau|A_{\mathrm{i}})$. It is also easier to calculate the denominator for signal and background separately, so the aim is to find an expression in terms of $P(\text{trigger}; \tau|A_{\mathrm{i}}, s)$. Using Bayes' theorem (5.6), the denominator in equation 5.10 becomes:

$$P(\text{trigger}; \tau) = P(\text{trigger}; \tau|A_{\mathrm{i}}, s)\frac{P(A_{\mathrm{i}}, s)}{P(A_{\mathrm{i}}, s|\text{trigger})} \tag{5.11}$$

The left-hand term in the numerator of equation 5.10 can be written as

$$P\left(s, t_i, m_i, \mathit{trk}_i, A_{\mathrm{i}}; \tau\right) = P\left(s, A_{\mathrm{i}}\right) P\left(t_i, m_i, \mathit{trk}_i; \tau | s, A_{\mathrm{i}}\right) \tag{5.12}$$

Putting these together (note the cancellation of $P(A_{\mathrm{i}}, s)$), and abbreviating

$P\left(\mathrm{trigger}|s, t_i, m_i, \mathit{trk}_i, A_{\mathrm{i}}\right)$ as $P\left(\mathrm{trigger}|\mathrm{all}\right)$, the equation becomes:

$$\begin{aligned} & P\left(s, t_i, m_i, \mathit{trk}_i, A_{\mathrm{i}}; \tau | \mathrm{trigger}\right) \\ = & \frac{P\left(t_i, m_i, \mathit{trk}_i; \tau | s, A_{\mathrm{i}}\right) P(A_{\mathrm{i}}, s|\mathrm{trigger}) P\left(\mathrm{trigger}|\mathrm{all}\right)}{P(\mathrm{trigger}; \tau | A_{\mathrm{i}}, s)} \end{aligned} \tag{5.13}$$

The first factor in the numerator can be split further

$$\begin{aligned} & P\left(t_i, m_i, \mathit{trk}_i; \tau | s, A_{\mathrm{i}}\right) \\ = \ & P\left(t_i; \tau | s, A_{\mathrm{i}}\right) P\left(\mathit{trk}_i | t_i, s, A_{\mathrm{i}}\right) P\left(m_i | \mathit{trk}_i, t_i, s, A_{\mathrm{i}}\right) \end{aligned} \tag{5.14}$$

At this stage only the basic rules of manipulating probabilities have been used. To progress further, some sensible assumptions are made:

- t_i , the measured lifetime, is independent of A_{i} . This is sensible as A_{i} is related to decay kinematics only. $P(t_i; \tau | A_{\mathrm{i}})$ is not the probability of measuring t_i given the trigger, but it is the probability of finding t_i given the decay kinematics that translate trigger cuts into lifetime cuts, *before the trigger is applied.*

- trk_i , the number of tracks found by the SVT, is independent of the A_{i} . As above, it is important to distinguish that A_{i} is related to decay kinematics only, so $P(\mathit{trk}_i | A_{\mathrm{i}})$ is not the probability to find r out of n tracks in the SVT given the trigger, but it is the probability of finding r out of n tracks in the SVT given the decay kinematics that translate trigger cuts into lifetime cuts.

- m_i , the reconstructed mass is independent of trk_i , t_i and $A_{\rm i}$. Again this is a sensible assumption for signal events. For combinatoric background, that is, background events formed by random track combinations, this assumption is revisited in the discussion of sources of systematic uncertainty in chapter 8.

With these assumptions, the probability simplifies to

$$P\left(t_i, m_i, trk_i; \tau | s, A_{\rm i}\right) \quad = \quad P\left(t_i; \tau | s, \right) P\left(trk_i | t_i, s\right) P\left(m_i | s\right) \tag{5.15}$$

so Equation 5.9 becomes:

$$\begin{aligned} & P\left(s, t_i, m_i, trk_i, A_{\rm i}; \tau | \text{trigger}\right) \\ & \quad = \quad \frac{P\left(t_i; \tau | s\right) P\left(trk_i | t_i, s\right) P\left(m_i | s\right) P(A_{\rm i}, s | \text{trigger}) P\left(\text{trigger} | \text{all}\right)}{P(\text{trigger}; \tau | A_{\rm i}, s)} \end{aligned} \tag{5.16}$$

The factor P($A_{\rm i}$,s| trigger) can be rewritten as $P(A_{\rm i}|\text{trigger})P(s|A_{\rm i}, \text{trigger})$. The first factor is the same for signal and background and does not depend on the parameters of interest. Therefore, this factor can be ignored (it would become a constant term in the likelihood). The second factor, $P(s|A_{\rm i}, \text{trigger})$ is the signal fraction given an $A_{\rm i}$. There exist other ways to rewrite P($A_{\rm i}$, s| trigger), but the form chosen is easiest to parameterize. This is demonstrated in section 5.6.

The final version of the signal part of the total PDF is now

$$\begin{aligned} & P\left(s, t_i, m_i, trk_i, A_{\mathrm{i}}; \tau | \mathrm{trigger}\right) \\ = \;& P\left(s | A_{\mathrm{i}}, \mathrm{trigger}\right) P\left(A_{\mathrm{i}} | \mathrm{trigger}\right) \times \\ & \frac{P\left(t_i; \tau | s,\right) P\left(m_i | s\right) P\left(trk_i | t_i, s\right) P\left(\mathrm{trigger} | \mathrm{all}\right)}{P(\mathrm{trigger}; \tau | A_{\mathrm{i}}, s)} \end{aligned} \tag{5.17}$$

The PDF for background will be analogous. The different factors are listed below, however the detailed parameterizations where necessary are explained in other sections.

An explanation of each factor in equation 5.17 and the background counterpart follows:

$P\left(t_i; \tau | s\right)$: This is the probability of measuring a lifetime t_i , for signal events with a mean lifetime τ and a resolution σ_t. It is given by

$$P\left(t_i; \tau | s\right) = \frac{1}{\tau} e^{\frac{-t_0}{\tau} + \frac{1}{2}\frac{\sigma^2}{\tau^2}} \mathrm{F}\left(\frac{t_0}{\sigma} - \frac{\sigma}{\tau}\right) \tag{5.18}$$

The value of $c \times \sigma$ is 26 μm. This choice is discussed and explained in section 5.3.

$P\left(t_i | b\right)$: This is the probability of measuring a lifetime t_i for background events. Details of the parameterization are given in section 5.5.

$P\left(m_i | s\right)$: The probability of observing a mass m given that the event is signal. The parameterization is described in section 5.4

$P\left(m_i | b\right)$: The probability that a mass m_i is observed given that the event is background. The parameterization is also described in section 5.4.

$P\left(trk_i|t_i,s\right)$: 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

$$P\left(trk_i|t_i,s\right) = P\left(\text{SVT found } r \text{ tracks out of } n|t_i,s\right) = \varepsilon_s^r\left(1-\varepsilon_s\right)^{(n-r)} \tag{5.19}$$

where r is the number of tracks found in the SVT, and n is the number of tracks available; n is smaller or equal to the total number of tracks in the final state. A track is "available" if it has $|d_0^{\text{off}}| < 1mm$ and $P_T > 2$ GeV. The analogous term for background has the same form.

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

$$P\left(\text{trigger}|\text{all}\right) = \left\{\begin{array}{ll} 1 & \text{if event passes trigger cuts} \\ 0 & \text{else} \end{array}\right\} \tag{5.20}$$

So its value is 1 for all events in the sample and could be omitted.

$P(\text{trigger};\tau|A_{\text{i}},s)$: The probability that the trigger accepts the signal event, given the acceptance function A_{i} . It is the normalization factor. It is given by:

$$\begin{aligned}
&\sum_{\text{all trk}}\ \sum_{\substack{k=\text{all}\\ \text{intervals}}}\ \int\limits_{t_{\text{min k}}}^{t_{\text{max k}}} P\left(trk_i|t_i,s\right) P\left(\text{trigger};\tau|\text{all}\right) \frac{1}{\tau} e^{\frac{-t_0}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}} \text{F}\left(\frac{t_0}{\sigma}-\frac{\sigma}{\tau}\right) dt_0 \\
=&\ \sum_{\substack{k=\text{all}\\ \text{intervals}}} \text{poly}_k(\varepsilon_s)\left[-e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma^2}{\tau^2}}\text{F}\left(\frac{t}{\sigma}-\frac{\sigma}{\tau}\right)+\text{F}\left(\frac{t}{\sigma}\right)\right]_{t=t_{\text{min k}}}^{t=t_{\text{max k}}} \qquad (5.21)
\end{aligned}$$

$P\left(s|A_{\text{i}},\text{trigger}\right)$: The signal fraction as a function of the acceptance function. The acceptance function depends only on the decay kinematics and all the event selection criteria (trigger and further selection after reconstruction). The parameterization for this factor is described in section 5.6.

The next few sections cover the models and parameterizations used for probabilities in mass and background lifetime, and also the value of σ_t used in $\mathrm{P}(t_i;\tau|s)$. Finally, the term $P\left(s|A_{\mathrm{i}}, \mathrm{trigger}\right)$ is considered, and it is shown how this term can be simplified using Fisher linear discriminant analysis and how, once this is done, the Fisher scalar variable that results from the analysis can be parameterized.

The measurement of the proper decay time for each B meson has been made by converting position measurement to time measurements. It is easier to consider the likelihood in terms of proper decay length, and fit for the mean decay length and so a change of variables is made where $t_i \to ct_i$, and $\tau \to c\tau$. This allows consideration of resolutions in terms of length which is more natural as the detector measures position.

5.3 Resolution Model

The likelihood assumes a detector resolution described by a single Gaussian of width 26 μm. In this section, it is shown that this description of the resolution is sensible and unlikely to affect the lifetime measurement. The sensitivity of the lifetime measurement to plausible variations from the single Gaussian of width 26 μm is studied alongside other sources of systematic uncertainty in chapter 8.

It is instructive to consider the lifetime PDF for signal events where the detector resolution is modeled by a single Gaussian as derived in section 4.3. equation 4.13 and repeated below in terms of mean decay length and proper decay length and σ is measured in microns:

$$P\left(ct_i; c\tau\right) = \frac{\frac{1}{c\tau} e^{\frac{-ct_i}{c\tau}} \mathrm{F}\left(\frac{ct_i}{\sigma} - \frac{\sigma}{c\tau}\right)}{\sum\limits_{\substack{i=\text{all}\\ \text{intervals}}} \left[-e^{\frac{-ct}{c\tau}} \mathrm{F}\left(\frac{ct}{\sigma} - \frac{\sigma}{c\tau}\right) + e^{-\frac{1}{2}\frac{\sigma^2}{(c\tau)^2}} \mathrm{F}\left(\frac{ct}{\sigma}\right) \right]_{t=t_{\text{min i}}}^{t=t_{\text{max i}}}} \tag{5.22}$$

It is seen from this equation that in the limit $1 >> \sigma^2/c^2\tau t_i$ and $ct_i/\sigma >> 1$ that this probability reduces to the probability obtained if measurement errors are ignored, and that the sensitivity to the resolution used is minimal. This is because the value of the computed Frequency function will be close to 1. It is also seen that if a prompt component of the sample is present (events that have small t_i), then the lifetime extracted will be dependent on the resolution. In this case, for the most accurate measurement, it is necessary to use a resolution based on a Gaussian with a width of the measurement error returned by the vertex fit rather than use the same average resolution for all events.

However, the requirement that $L_{xy}(B) > 350\mu m$ removes all events with small t_i, and the cut on σ_{ct} removes events that have a measurement error greater than 100 μm. $c\tau$ is approximately 490 μm and so it is evident that the majority of the events in the data sample will be insensitive to the resolution. This observation allows use of an average resolution function where each measurement of ct_0 is assumed to have the same resolution. It is advantageous to use an average resolution for all events for two reasons. Firstly, there is a correlation between an event's σ_{ct} and A_{i}, as the acceptance is dependent on the opening angle between tracks, as is σ_{ct}. This correlation would have to be properly modeled in the PDF and as the correlation is not simple this is not trivial. Secondly, the σ_{ct} distribution for signal and background is different. As discussed in equation 5.4, if a quantity is used dependent per event then the distribution of that quantity must be modeled for the PDF to be correct. This adds a further layer of complexity to the analysis. While modeling the distributions of σ_{ct} will increase the signal and background separation, and therefore slightly reduce the statistical uncertainty, the difficulties of correctly modeling the correlation leave using the average resolution as the most attractive option.

5.3.1 Studies on Lifetime Measurement Sensitivity to Resolution Model

The realistic MC can be used to understand the shape of the resolution function as the measured and generated proper decay time of all events are known. A plot showing the difference between the measured and generated individual proper decay time lifetime are shown in figure 5.1. It is roughly approximated by a Gaussian of width 27 μm. (A width of 26 μm is used for the data fit because this is the mean error in background subtracted data.)

The fitted Gaussian has a mean displaced from zero. In the analysis we use a resolution function centered on zero. Other resolution functions are considered in the discussion on systematic uncertainties.

To test the sensitivity to the chosen width of the resolution function, the lifetime of 75,000 realistic MC events are fitted using different values for σ_{ct}. The results are shown in Table 5.1. The results show that there is almost no difference in the lifetime measurement when the average σ_{ct} is between 2 and 32 μm, and this is the result expected from inspection of equation 5.22 above. It is also seen that if an unrealistic resolution of a Gaussian of width 100 μm is applied, then the lifetime measurement is different by 4 μm which represents 1% change. This study shows that if one single

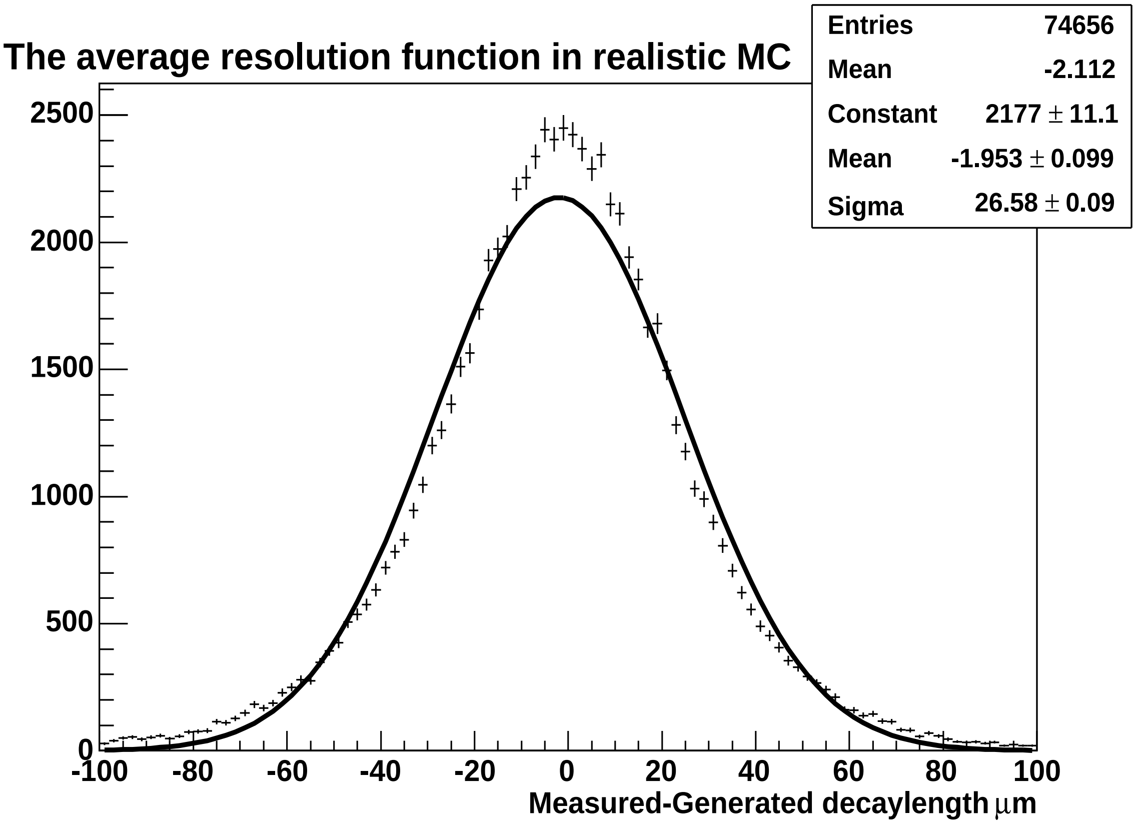


Figure 5.1: The difference between generated and measured lifetime in realistic MC

Gaussian is used as the resolution function, any sensible choice of width will do. The realistic MC shows that a better approximation could include a resolution function with longer tails. The effect of this is further investigated as a source of systematic uncertainty. The background subtracted histogram of the signal region has a mean σ_{ct} of 26 μm and this is used as the width of the Gaussian to model the detector resolution.

5.4 The Mass Model

The mass distribution, after the selection criteria described in chapter 3 have been applied, is shown in figure 5.2. This section details the motivations in the choice of

Average resolution	Best Fit ±Error
$\sigma_{c\tau}$=2 μm	494.158 ± 3.083 μm
$\sigma_{c\tau}$=21 μm	494.158 ± 3.083 μm
$\sigma_{c\tau}$=26 μm	494.156 ± 3.083 μm
$\sigma_{c\tau}$=32 μm	494.159 ± 3.083 μm
$\sigma_{c\tau}$=100 μm	490.694 ± 3.058 μm

Table 5.1: The fit results for 75000 realistic MC $B^{\pm} \rightarrow D^0\pi^{\pm}$. The truth lifetime is 496 μm.

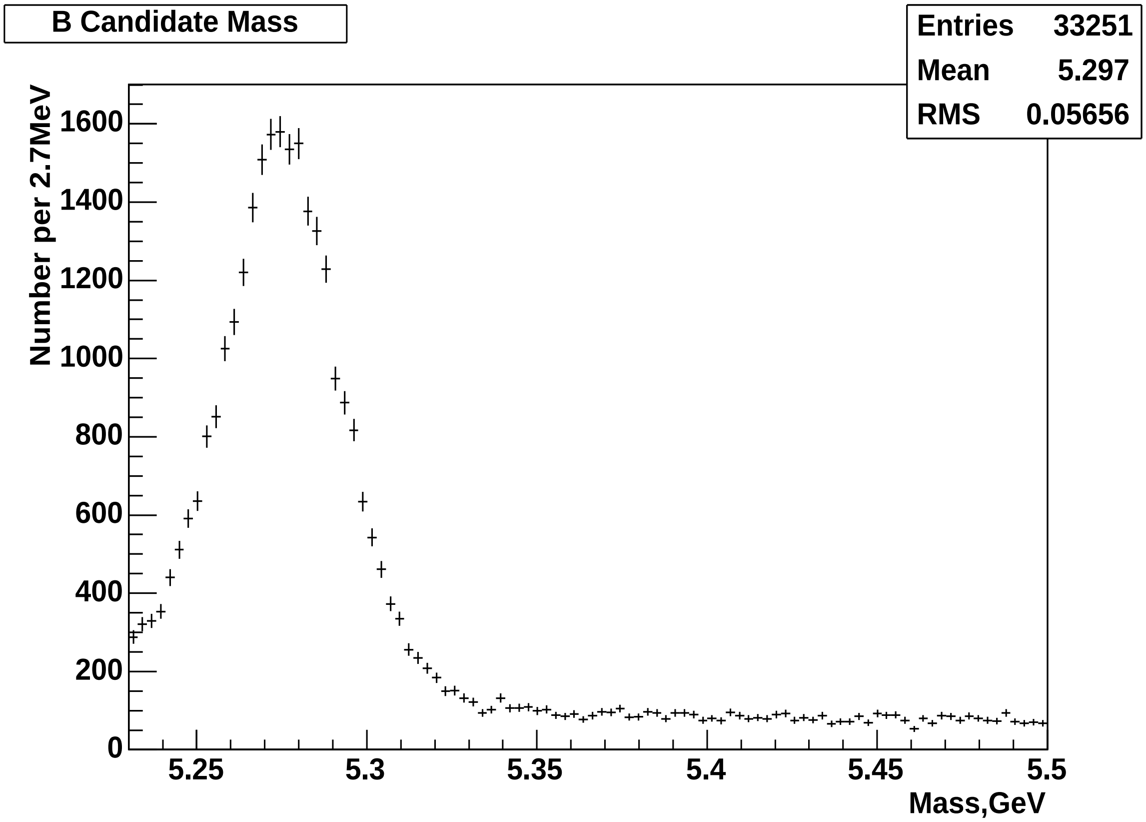


Figure 5.2: The distribution of mass after selection criteria has been applied.

mass window, and the model used to describe the mass distribution in the fit.

The upper mass cut is at 5.5GeV. This is to allow for a sufficient number of events in the upper sideband. Background events are required in the fit to give a good determination of what the background distributions of lifetime and acceptance function are underneath the peak. The lower sideband is removed entirely. This is because the lower sideband is dominated by partially reconstructed decays, and this type of background is not expected underneath the peak. Detailed studies of the extended mass spectrum of reconstructed $\bar{D}^0\pi$ events have been carried out for other CDF analyses, details of which can be found in [59]. A mass spectrum plot from such an analysis is included here in figure 5.3. The contribution from partially reconstructed modes (e.g $B \rightarrow D^*\pi$) is expected to be less than 1% above a mass of 5.2GeV, and even less over a mass of 5.23GeV. This cut also removes approximately two thirds of the Cabibbo suppressed decay mode $B \rightarrow DK$. The remainder of this Cabibbo suppressed decay mode is treated in the same way as signal, and systematic studies described in chapter 8 later show that treating it as signal adds negligible error to the measurement. The lower mass cut also removes signal. The reduction in yield is required to remove the partially reconstructed background as the analysis method is not yet equipped to model these events.

A first order polynomial is the model for background events, and the slope of the polynomial is determined by the likelihood fit. In this analysis, the Cabibbo suppressed decay mode ($B \rightarrow DK$ is treated as signal, so there is no need to model a separate component. To fit the signal mass distribution, two Gaussians, with different means and widths, are required. Hence the mass model is given by:

$$P(m_i|s) = \frac{f_1}{\sigma_1\sqrt{2\pi}} e^{-(m_i-m_1)^2/2\sigma_1^2} + \frac{(1-f_1)}{\sigma_2\sqrt{2\pi}} e^{-(m_i-m_2)^2/2\sigma_2^2} \tag{5.23}$$

and is additionally constrained by the normalization condition:

$$\int_a^b P(m_i|s) dm_i = 1 \tag{5.24}$$

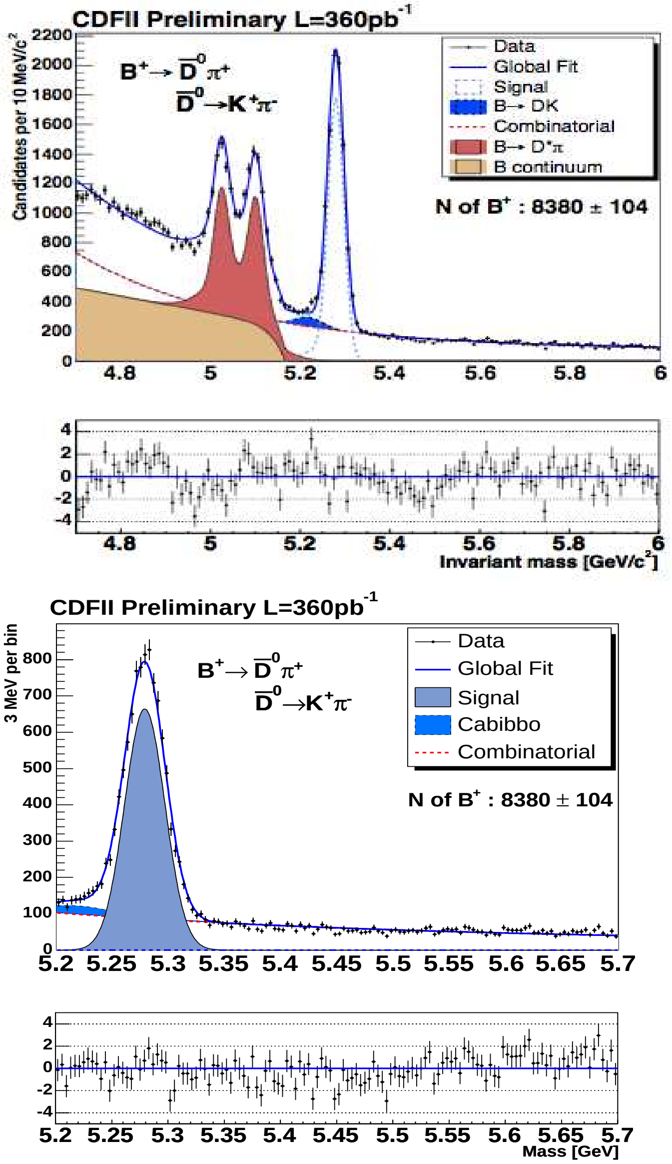


Figure 5.3: The mass spectrum determined by a previous CDF analysis [59] of $B \rightarrow D\pi$. This uses 360 pb^{-1} of data and a wider mass range than used in the analysis presented in this thesis. The top plot shows the fit over the full range, and the bottom plot shows a close up of the signal peak.

$$P(m_i|b) = (1 - \alpha m_i) \cdot (b - a - \frac{\alpha}{2}(b^2 - a^2)) \tag{5.25}$$

where a and b are the lower and upper mass limits, 5.23 and 5.5 GeV/c^2 respectively.

The free parameters in the mass fit are m_1, m_2, σ_1, σ_2, α, f_1 and f_s. The fit to the data with residuals is shown in figure 5.4. The best fit parameters are given in table 7.1. A further model involving 3 Gaussians was also fit to the data, but one of the Gaussian fraction best fit parameter was 0 which demonstrates that the extra degrees of freedom were not required. A second order polynomial was also tried for the background, but the coefficient for the 2nd order term was consistent with zero, which also shows that the first order polynomial is sufficient.

5.5 Parameterizing the Background Lifetime Distribution

To perform the fit, a parameterization for the background lifetime will also be required. The starting point to describe the lifetime for signal events is an exponential. The exponential is the distribution of the lifetime before detector and trigger effects are taken into account. The distribution of lifetime for background events before the trigger is applied is described by a function $y(t)$. As there is no meaning to the physical nature of the background lifetime distribution, the function $y(t)$ can be said to contain any detector resolution, effects and this removes the need for $y(t)$ to explicitly contain dependence on σ_t.

The probability of measuring a lifetime t_i given that an event is background and

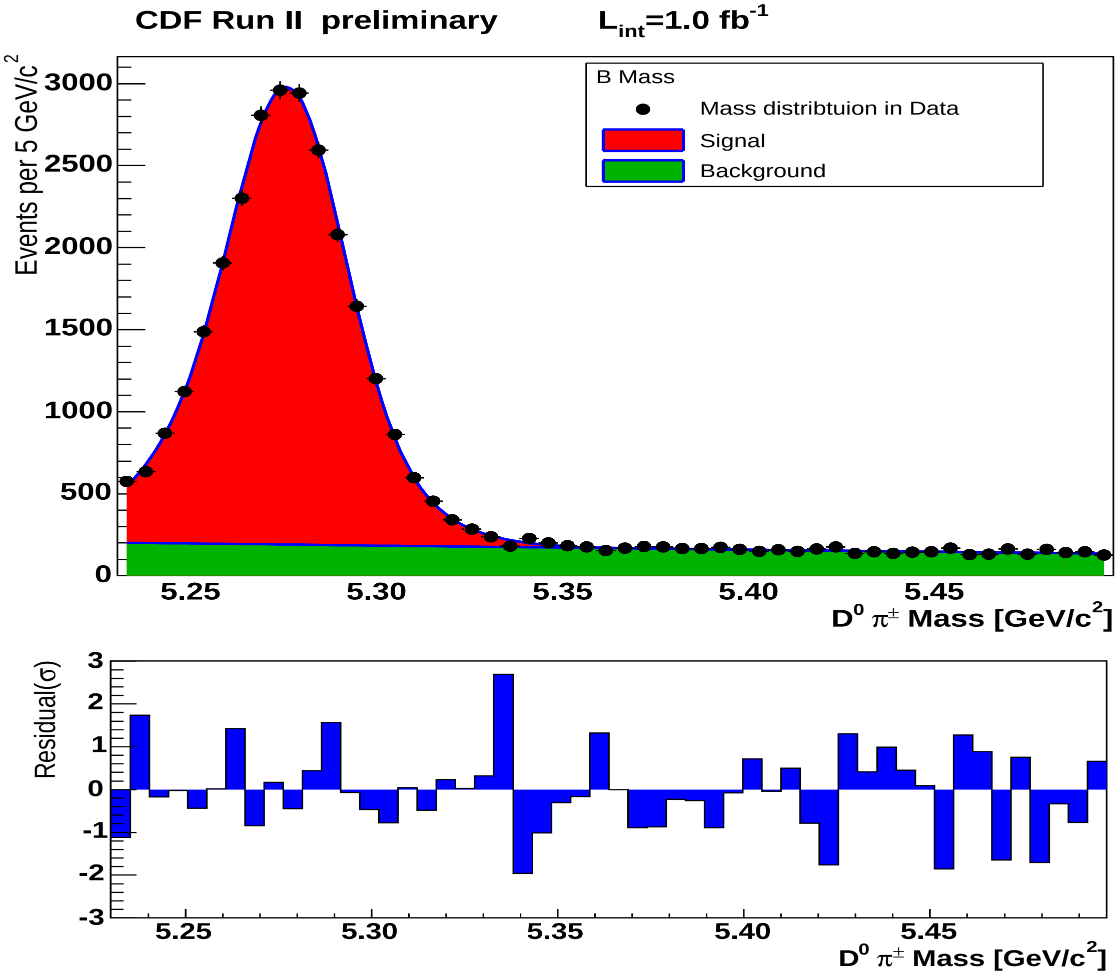


Figure 5.4: The mass fit projection and residuals.

given its acceptance function is

$$P(t_i|b, A_{\mathrm{i}}) = \begin{cases} \frac{y(t_i)\times A_{\mathrm{i}}|_{t=t_i}}{\int A_{\mathrm{i}}\times y(t)\ dt} & \text{for } t_{\min} \leq t_i \leq t_{\max} \\ 0 & \text{for all other } t. \end{cases} \tag{5.26}$$

where $t_{\min}$ and $t_{\max}$ represent the edges of the acceptance function and where $P(t|b, A_{\mathrm{i}})$ has been normalized so that the integral is 1 over all time.

Traditionally, in lifetime analyses, a sum of two or three exponentials convoluted with detector resolution has been used to model the background. These models give poor fits to the data in the upper sideband, and so a more complex $y(t)$ is required.

The background is parameterized by fitting the height of $y(t)$, at different times, t_j, and interpolating between these points using an exponential function, i.e, a linear interpolation on $logy$. The expression is given by:

$$y(t) = e^{a_j + \left(\frac{a_{j+1}-a_j}{t_{j+1}-t_j}\right)(t-t_j)} \qquad \text{for } \ t_j \leq t \leq t_{j+1} \tag{5.27}$$

where the constants a_j are returned by the fit. There is a freedom of choice in the number, n and spacing of the points t_j. It is desirable to add the least number of necessary points to the fit, and it is observed that the lifetime distribution after the trigger varies most rapidly at small lifetimes. Therefore it is prudent to have t_j closely spaced in region of low t, and let them be more widely spaced at large t. To avoid too much adjustment, the following algorithm is used to determine the spacing:

$$t_j = t_{\min} + \left(-c + e^{\left(ln(c) + \frac{ln(c+b)-ln(c)}{n} \cdot j\right)}\right) \cdot \frac{(t_{\max} - t_{\min})}{b}, \tag{5.28}$$

where c and b are constants which affect the scale of the logarithmic spacing. We choose $c = 1$. There are now two parameters, b and n, which determine the spacing of t_j. Decreasing b spaces t_j more equally.

The upper sideband is used to test this model and find judicious choices for b and n. The values $n = 10$ and $b = 4$ are chosen, as these give a good fit as shown in figure 5.5.

Other choices for b and n are available. Using a larger number for n does improve the residuals but also leads to fit instabilities, especially in the higher ct region above $1500\mu m$ where there are very few events. No significant difference is observed between using $b = 3$ or $b = 5$. The effect the choice of this parameterization has on the final

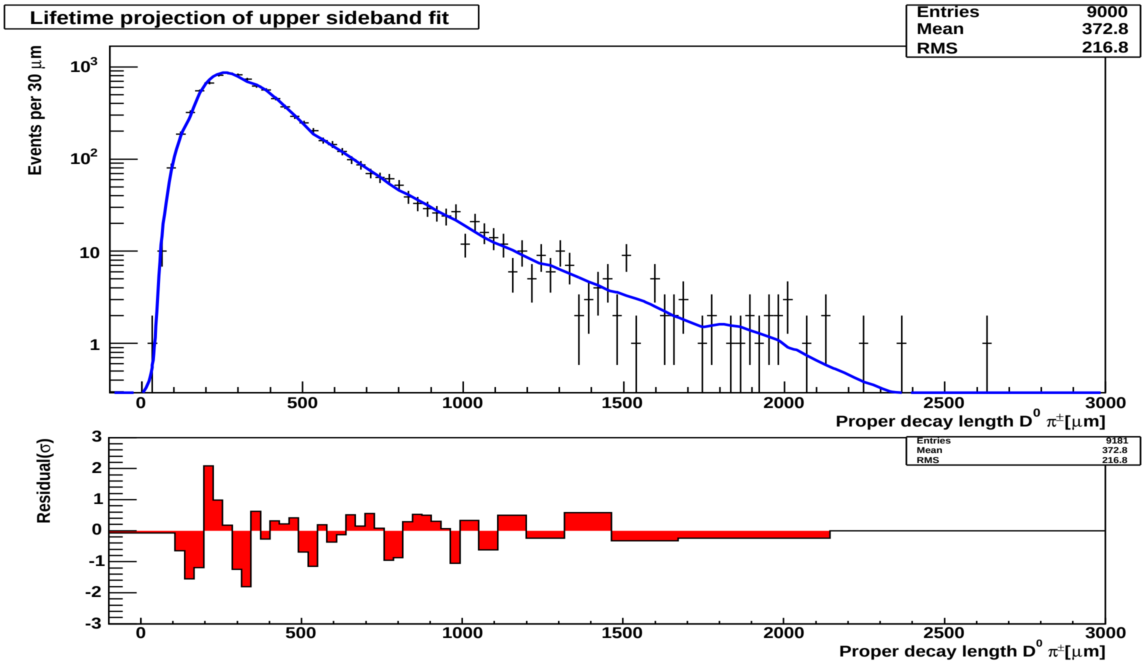


Figure 5.5: Lifetime fit projection for the upper sideband. $n = 10$ and $b = 4$

result is discussed along with other systematic uncertainties in chapter 8. The fit projections from performing fits on the upper sideband using alternative choices for b and n are shown in figures 5.6, 5.7 and 5.8.

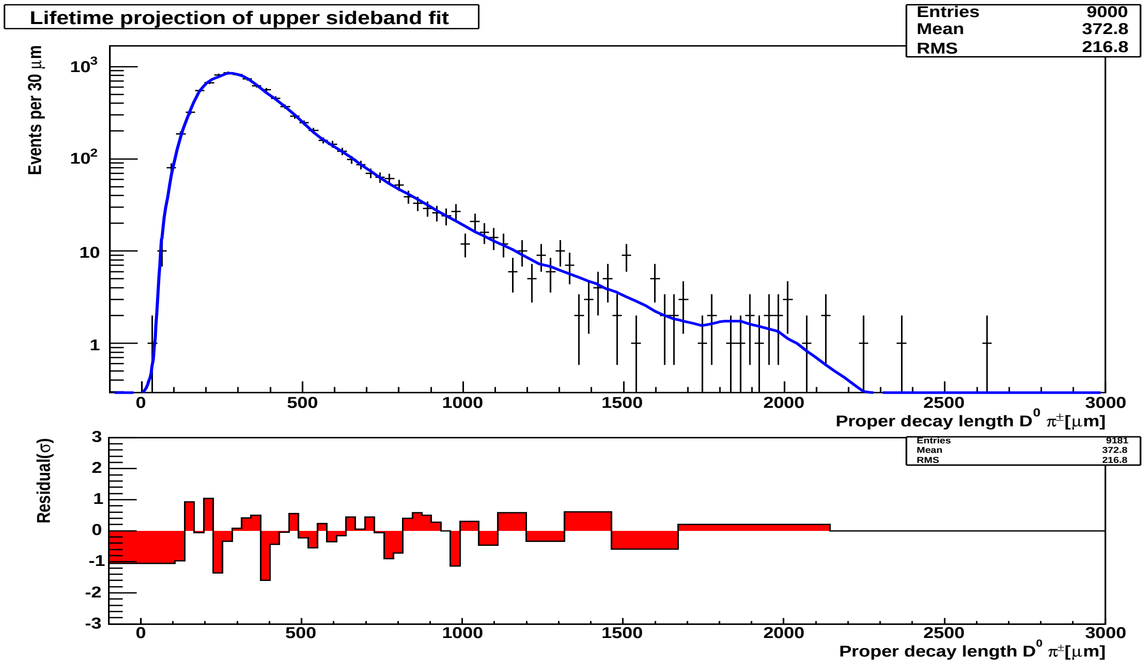


Figure 5.6: Lifetime fit projection for the upper sideband using 12 fit points. $n = 12$ and $b = 4$.

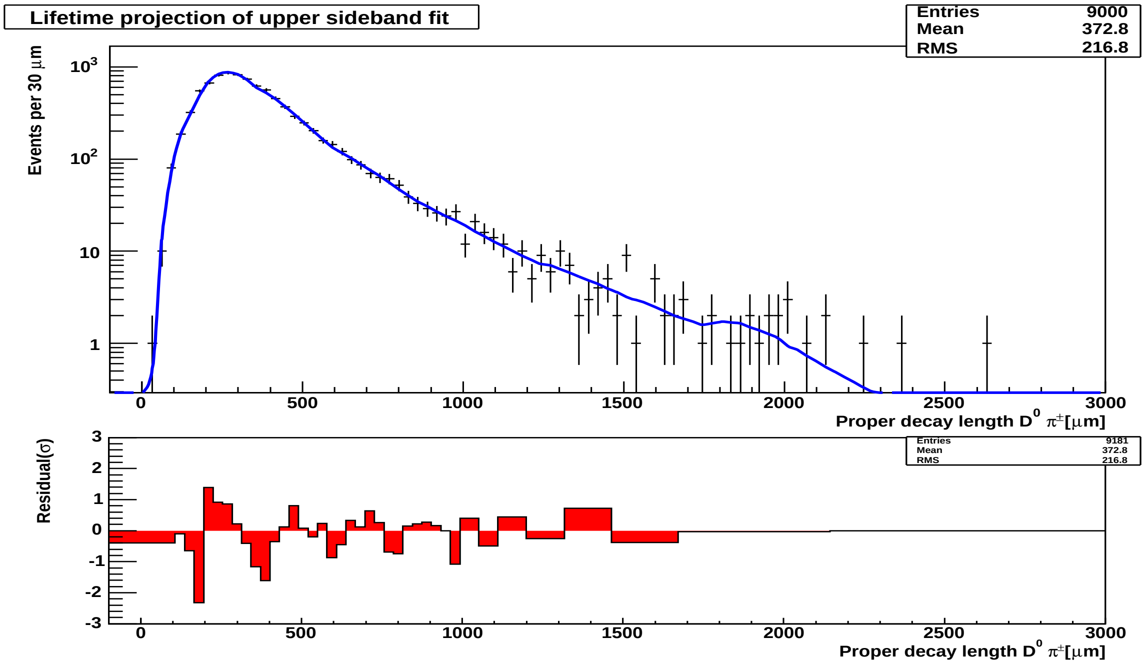


Figure 5.7: Lifetime fit projection for the upper sideband, with $n = 10$ and $b = 3$.

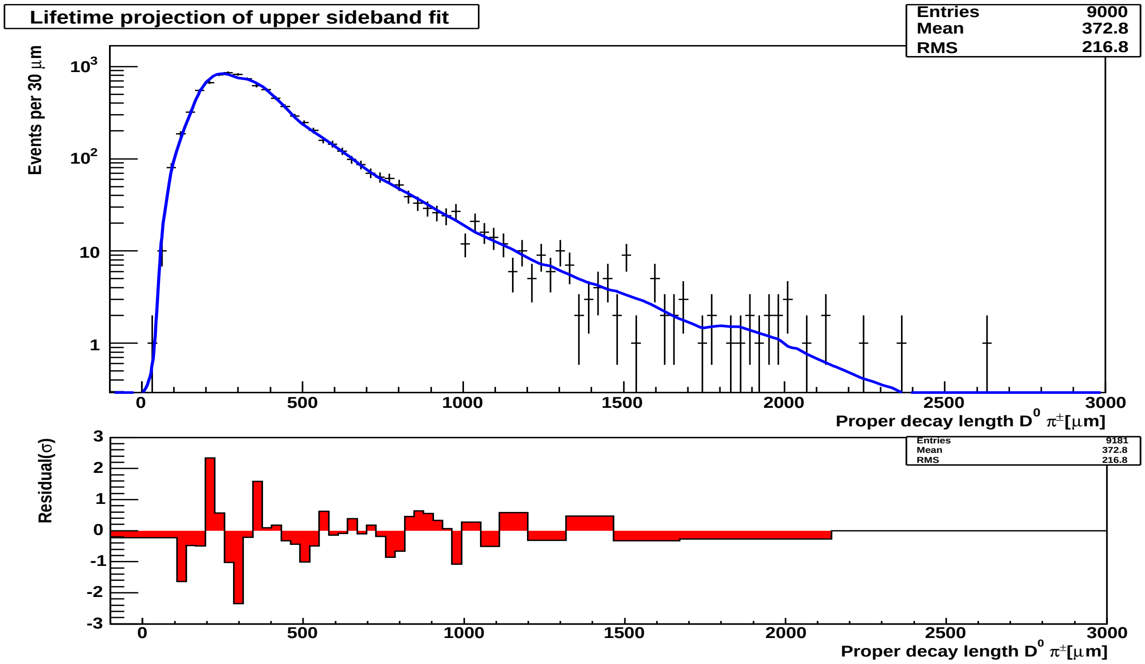


Figure 5.8: Lifetime fit projection for the upper sideband with $n = 10$ and $b = 5$.

5.6 The Signal Fraction Based on the Acceptance Function

The likelihood requires a signal probability for each event as a function of the acceptance function. The mean acceptance function for signal and background is shown in figure 5.9. The mean acceptance function is calculated by adding each event's acceptance function expressed as a histogram and normalizing this sum to one event. For the mean background acceptance, events in the upper sideband are used. For the mean signal acceptance, the events in the signal region are used and a background subtraction is performed.

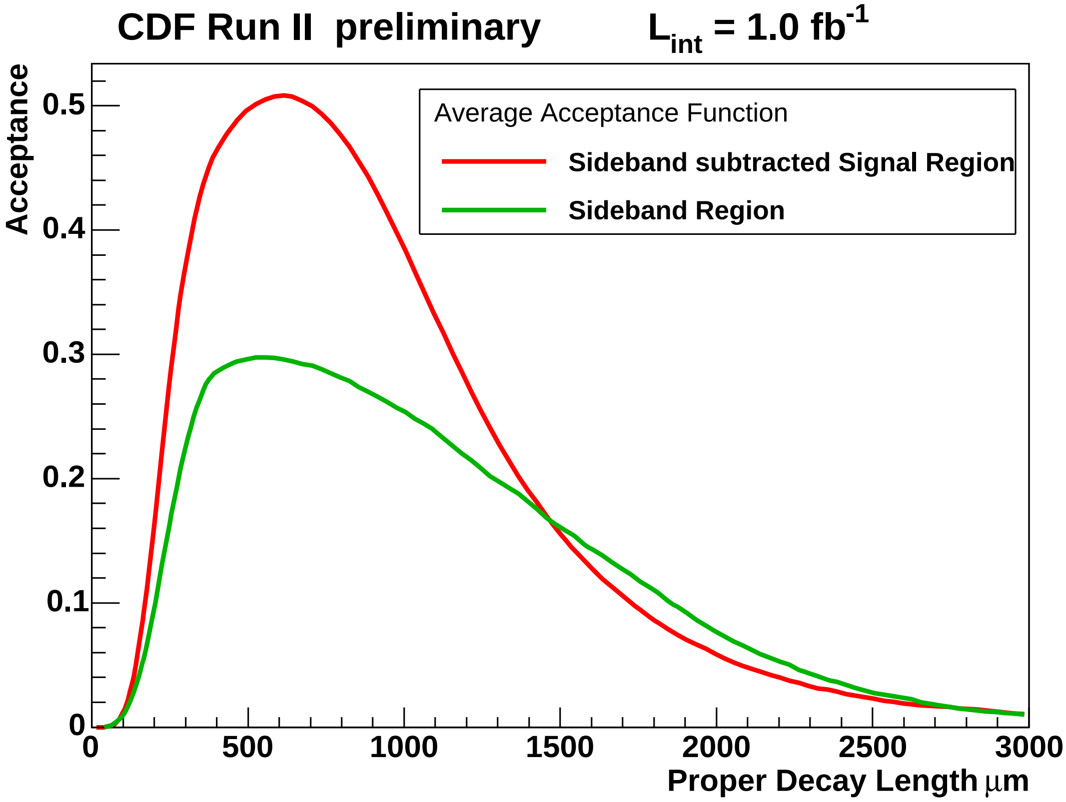


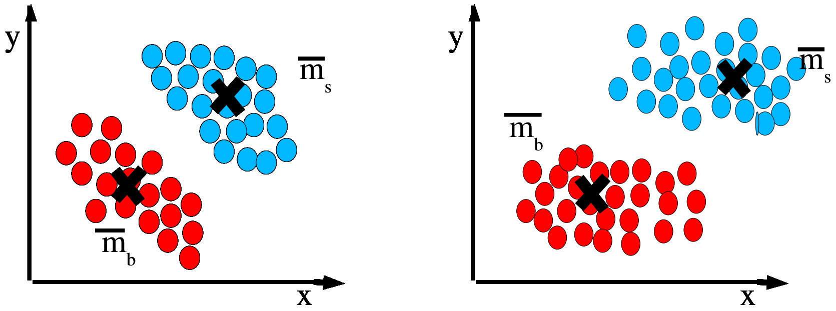
Figure 5.9: Mean acceptance function for signal (red, obtained by background subtraction) and background (green).

If the distribution of acceptance functions were the same for signal and background, the factors $P(s|A_{\rm i})$ and $P(b|A_{\rm i})$ would simplify to the overall signal fraction. However, given that the mean acceptance for signal and background is clearly different, this factor cannot be replaced by the overall signal fraction.

If the acceptance were a scalar variable, then all that is required is to find a suitable parameterization for the signal fraction as a function of that variable. However, the acceptance is not a scalar variable; it is a function. Parameterizing a distribution of functions is required but is not trivial. One way to overcome this problem is to characterize the acceptance function by a scalar and parameterize it to give the signal probability. It is best to preserve as much information about the acceptance function as possible, particularly the parts where signal and background differ most. Fisher discriminant analysis is used to perform the transformation from function to scalar. The next section describes the basics of Fisher linear Discriminant analysis, before describing how Fisher Discriminant analysis is applied to this problem. Fisher Discriminant analysis is described in [60], [61], [62].

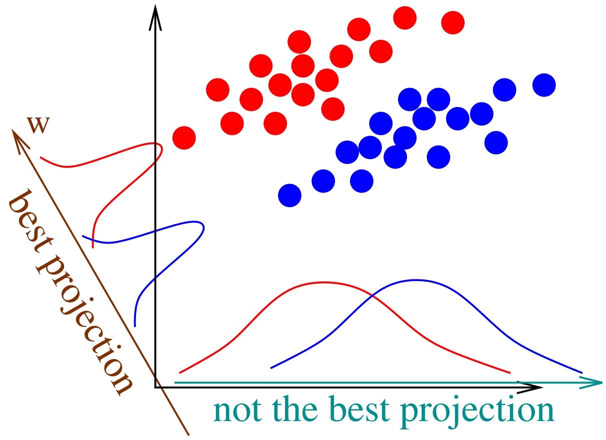
5.6.1 Basics of Fisher Linear Discriminant Analysis

Imagine two classes of events, eg, signal and background with their own distributions of variables x and y as shown in figures 5.10(a), 5.10(b) and 5.10(c). The means of variables x and y for each distribution are shown as $\overline{m_s} = (\overline{x_s}, \overline{y_s})$ and $\overline{m_b} = (\overline{x_b}, \overline{y_b})$. The vector from the origin to any point can be projected along any direction. If, for example, the vector to each point is projected along the hypothetical direction $\boldsymbol{z}$, then the result of the projection can be binned into a histogram of variable z. In order to achieve separation between signal and background, it is best to project along a direction which gives maximal separation in the resulting single variable distribution. For example, in figure 5.10(a), there will be good separation if the direction for projection is that along the two mean points, whereas in figure 5.10(b), there will be good separation if the data is projected onto the y axis. The best separation will be the one where the distance between the projected means of each class of event is large while the spread around each mean remains small. The direction which results in maximal separation is referred to as $\boldsymbol{w}$.



(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) Here the separation is best along the direction w.

Figure 5.10: These diagrams illustrate that it is necessary to take the means and spread of each variable in finding the direction of best separation

Fisher discriminant analysis finds this direction $\boldsymbol{w}$. A derivation follows and starts by defining the following matrices and vectors: $\mid \overline{\boldsymbol{m_s}}\rangle$ is the mean vector describing signal events, $\mid \overline{\boldsymbol{m_b}}\rangle$ is the mean vector describing background events, and $\mid \boldsymbol{p_i}\rangle$ is the vector representing each event.

The matrix $\boldsymbol{S_M}$ is defined by considering the square of the separation of projected means of signal and background events along the projection direction $\mid \boldsymbol{w}\rangle$. The definition is given in equation 5.29.

$$\begin{aligned}(\langle \boldsymbol{w} \mid \overline{\boldsymbol{m_s}}\rangle - \langle \boldsymbol{w} \mid \overline{\boldsymbol{m_b}}\rangle)^2 &= \langle \boldsymbol{w} \mid (\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}})\rangle\langle(\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}}) \mid \boldsymbol{w}\rangle \\ &= \langle \boldsymbol{w} \mid \boldsymbol{S_M} \mid \boldsymbol{w}\rangle\end{aligned} \tag{5.29}$$

The matrix $\boldsymbol{S_{sig}}$ is defined by considering the square of the spread of the projected signal events around the projected mean. The definition is given in equation 5.30.

$$\begin{aligned}&\sum_{All\ signal\ events} (\langle \boldsymbol{w} \mid \boldsymbol{p_i}\rangle - \langle \boldsymbol{w} \mid \overline{\boldsymbol{m_s}}\rangle)^2 \\ &\qquad = \sum_{All\ signal\ events} \langle \boldsymbol{w} \mid (\boldsymbol{p_i} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{p_i} - \overline{\boldsymbol{m_s}}) \mid \boldsymbol{w}\rangle \\ &\qquad = \sum_{All\ signal\ events} \langle \boldsymbol{w} \mid \boldsymbol{S_{sig}} \mid \boldsymbol{w}\rangle\end{aligned} \tag{5.30}$$

The matrix $\boldsymbol{S_{bg}}$ is defined in a similar way in equation 5.31.

$$\begin{aligned}&\sum_{All\ bkg\ events} (\langle \boldsymbol{w} \mid \boldsymbol{p_i}\rangle - \langle \boldsymbol{w} \mid \overline{\boldsymbol{m_b}}\rangle)^2 \\ &\qquad = \sum_{All\ bkg\ events} \langle \boldsymbol{w} \mid (\boldsymbol{p_i} - \overline{\boldsymbol{m_b}})\rangle\langle(\boldsymbol{p_i} - \overline{\boldsymbol{m_b}}) \mid \boldsymbol{w}\rangle \\ &\qquad = \sum_{All\ bkg\ events} \langle \boldsymbol{w} \mid \boldsymbol{S_{bg}} \mid \boldsymbol{w}\rangle\end{aligned} \tag{5.31}$$

It is clear that the best projection direction is one in which the means of the two types of events fall far apart with smallest spread. This is Fisher's criterion and is expressed mathematically as finding the $\boldsymbol{w}$ for which $\boldsymbol{J(w)}$ is maximized, where $\boldsymbol{J(w)}$ is given in equation 5.32.

$$\boldsymbol{J(w)} = \frac{\langle \boldsymbol{w} \mid \boldsymbol{S_M} \mid \boldsymbol{w} \rangle}{\langle \boldsymbol{w} \mid (\boldsymbol{S_{sig}} + \boldsymbol{S_{bg}}) \mid \boldsymbol{w} \rangle} = \frac{\langle \boldsymbol{w} \mid \boldsymbol{S_M} \mid \boldsymbol{w} \rangle}{\langle \boldsymbol{w} \mid \boldsymbol{S_W} \mid \boldsymbol{w} \rangle} \tag{5.32}$$

Equations 5.33 and 5.34 show that an eigenvalue equation is formed as a result from the maximization of $\boldsymbol{J(w)}$. Furthermore, using the definition of the matrix $\boldsymbol{S_M}$, the eigenvalue equation can be simplified to remove the need to find the actual eigenvalues. Only ${\boldsymbol{S_W}}^{-1}$ and the vector $\mid (\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}})\rangle$ are required to find the projection direction $\mid \boldsymbol{w}\rangle$, where the matrix $\boldsymbol{S_W}$ is defined in equation 5.32. Hence the condition

$$0 = \nabla_w(\boldsymbol{J(w)}) = \frac{2\boldsymbol{S_M} \mid \boldsymbol{w}\rangle}{\langle \boldsymbol{w} \mid \boldsymbol{S_W} \mid \boldsymbol{w} \rangle} - \frac{\langle \boldsymbol{w} \mid \boldsymbol{S_M} \mid \boldsymbol{w} \rangle}{\langle \boldsymbol{w} \mid \boldsymbol{S_W} \mid \boldsymbol{w} \rangle} . \frac{2\boldsymbol{S_W} \mid \boldsymbol{w}\rangle}{\langle \boldsymbol{w} \mid \boldsymbol{S_W} \mid \boldsymbol{w} \rangle} \tag{5.33}$$

simplifies to

$$\begin{aligned} \boldsymbol{S_M} \mid \boldsymbol{w}\rangle - \lambda \boldsymbol{S_W} \mid \boldsymbol{w}\rangle &= 0 \\ \boldsymbol{S_M} \mid \boldsymbol{w}\rangle &= \lambda \boldsymbol{S_W} \mid \boldsymbol{w}\rangle \\ {\boldsymbol{S_W}}^{-1}\boldsymbol{S_M} \mid \boldsymbol{w}\rangle &= \lambda \mid \boldsymbol{w}\rangle \end{aligned} \tag{5.34}$$

where λ is a constant given by

$$\lambda = \frac{\langle \boldsymbol{w} \mid \boldsymbol{S_M} \mid \boldsymbol{w} \rangle}{\langle \boldsymbol{w} \mid \boldsymbol{S_W} \mid \boldsymbol{w} \rangle} \tag{5.35}$$

Using the definition of $\boldsymbol{S_M}$ from equation 5.29, $\boldsymbol{S_M} \mid \boldsymbol{w}\rangle$ can be rewritten in the following way.

$$\begin{aligned} \boldsymbol{S_M} \mid \boldsymbol{w}\rangle &= \mid (\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}})\rangle\langle(\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}}) \mid \boldsymbol{w}\rangle \\ \boldsymbol{S_M} \mid \boldsymbol{w}\rangle &\propto \mid (\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}})\rangle \end{aligned} \tag{5.36}$$

When this is inserted into equation 5.33, it can be seen that it is not necessary to solve for the eigenvalues and all that is needed to find $\mid \boldsymbol{w}\rangle$ is the inverse of $\boldsymbol{S_W}$ and $\mid (\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}}\rangle$.

$$\boldsymbol{S_W}^{-1} \mid (\overline{\boldsymbol{m_s}} - \overline{\boldsymbol{m_b}})\rangle \propto \mid \boldsymbol{w}\rangle \tag{5.37}$$

The value of the discriminating variable is given by the scalar product of each event vector $\mid \boldsymbol{p_i}\rangle$ with $\mid \boldsymbol{w}\rangle$, and this Fisher scalar variable is the best one for distinguishing between the two classes of events.

While the derivation discussed above was for 2 variables describing the events, this technique can be generalised to any number of variables. The matrices $\boldsymbol{S_M}$ and $\boldsymbol{S_W}$ are $n \times n$ square matrices where n is the number of variables describing each event.

5.6.2 Using the Fisher Discriminant Analysis to Characterize the Acceptance Function by a Scalar Variable

To use the Fisher Discriminant Analysis to transform the acceptance functions to a scalar variable, two things are required:

- A description of each acceptance function in terms of a set of common variables.
- A knowledge of the mean value of each of these variables for signal and background, i.e, the vectors $\mid \overline{\boldsymbol{m_s}}\rangle$ and $\mid \overline{\boldsymbol{m_b}}\rangle$.

The acceptance function is a series of top hat functions, each of which is characterized by $t_{\min}$, $t_{\max}$ and ε_s. The series of $t_{\min}$ and $t_{\max}$ cannot be used as variables as there is no way to connect the $t_{\min}$,$t_{\max}$ of one event with another as they could, for example, come from different track combinations. Another issue is that the height of the top hat function is characterized by a polynomial in ε_s, but the value of ε_s itself is not known until after the likelihood has been maximized.

To resolve these issues, the acceptance function is split into different parts that have a different polynomial in ε_s for their height. This resolves the issue of not knowing the value of the efficiency before the fit as regions of differing polynomial can be compared directly. For example, the region where the acceptance function has height ε_s^2 can be compared directly, with another acceptance function without knowledge of the value of ε_s. Once the acceptance is split into different parts it is then binned accordingly. The height of each bin is now the variable that describes the acceptance function. Nominally the height in each bin is either 0 or 1; however, where the acceptance turns on or off within the bin, an intermediate value is taken to represent the acceptance in that bin. Each part of the acceptance is binned from 0 to 3000 microns into 20 bins and, as there are three types of polynomial, this means that each acceptance function is represented by 60 variables. Tests in chapter 6 have shown that this is sufficient, and to bin more finely can cause instabilities in the inversion of the matrix $\boldsymbol{S_W}$. Figure 5.11 illustrates how the acceptance function is described by a vector of variables.

Now that each event is characterized by a vector of variables the matrix $\boldsymbol{S_W}$ and the vector $| \overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}} \rangle$ need to be extracted.

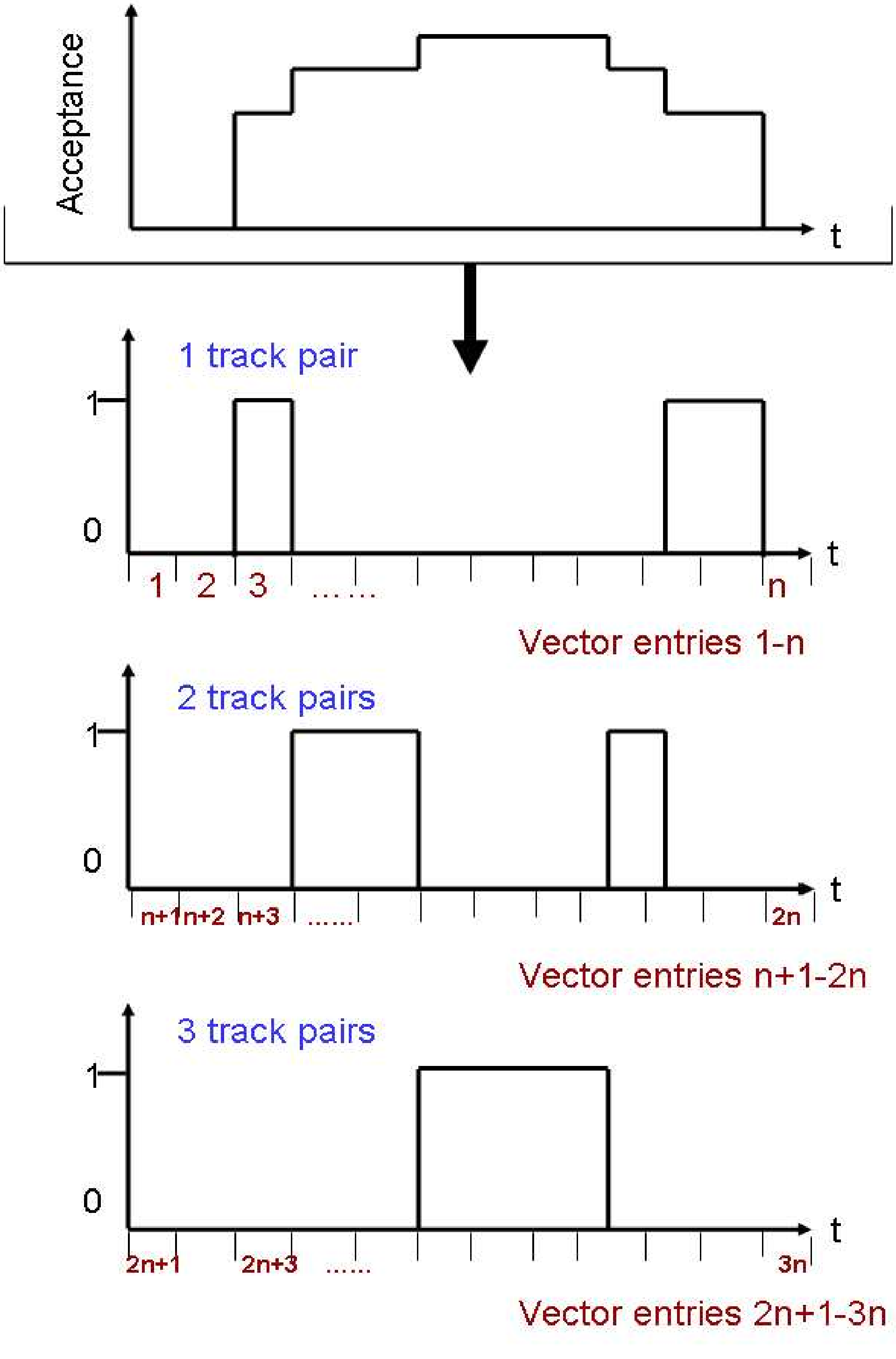


Figure 5.11: To characterize the acceptance function by a vector, the acceptance function is split into the sum of its parts. From this acceptance function, a vector of length 3n is constructed where the nth entry is the height shown in the lower three diagrams.

5.6.3 Extracting $|\,\overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}}\rangle$ from the Dataset

The required vector is simply the difference between the mean vectors describing signal and background events. In data, it is not possible to distinguish between true signal and true background events. The mass fit and background subtraction are used to calculate the mean vectors. The mass fit can be used to define two regions. The first is the upper sideband which contains events with reconstructed mass greater than 5.37 GeV. It is assumed that all the events in the sideband region are background, and that this background is typical of all the background in the sample. The vector $|\,\overline{\boldsymbol{m_b}}\rangle$ is simply the sum of acceptance vectors in the sideband region divided by the number of events in this region. The validity of the assumption that the background in the upper sideband is representative of all the background events is discussed and justified in chapter 8.

The signal region is defined between 5.25 and 5.31 GeV, as this gives a larger signal to background ratio than using an extended region between 5.23 and 5.37 GeV. The vector $|\,\overline{\boldsymbol{m_r}}\rangle$ is defined as the mean vector for events in the signal region. The signal fraction in the signal region is f_s. The vector $|\,\overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}}\rangle$ can be extracted as shown in equation 5.38 below.

$$\begin{aligned}
\overline{\boldsymbol{m_r}}\rangle &= \frac{\sum_{\substack{events\ in\\ signal\ region}} |\,\boldsymbol{v_i}\rangle}{\sum_{\substack{events\ in\\ signal\ region}} 1} \\
|\,\overline{\boldsymbol{m_r}}\rangle &= f_s\,|\,\overline{\boldsymbol{m_s}}\rangle + (1-f_s)\,|\,\overline{\boldsymbol{m_b}}\rangle \\
|\,\overline{\boldsymbol{m_r}} - \overline{\boldsymbol{m_b}}\rangle &= f_s\,|\,(\overline{\boldsymbol{m_s}}\rangle - |\,\overline{\boldsymbol{m_b}}\rangle)
\end{aligned} \tag{5.38}$$

The vector ($|\,\overline{\boldsymbol{m_r}}$ - $\overline{\boldsymbol{m_b}}\rangle$) is proportional to ($|\,\overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}}\rangle$). Although f_s can be calculated from the mass fit, given that the constant of proportionality will not change the final direction $\boldsymbol{w}$, ($|\,\overline{\boldsymbol{m_r}}$ - $\overline{\boldsymbol{m_b}}\rangle$) is used to represent $|\,\overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}}\rangle$.

A further note is made here regarding the vector of variables that describes the acceptance function. It is possible that some of the variables in the vector $| \overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}}\rangle$ have value 0 , which means that the variable corresponding to this entry provides little discriminating power. If these variables are kept, problems may be encountered when inverting $\boldsymbol{S_W}$. This issue is documented in other uses of Fisher discriminants. Therefore rows in which $| \overline{\boldsymbol{m_s}}$ - $\overline{\boldsymbol{m_b}}\rangle$ =0 are removed and corresponding rows in the event vectors are also removed. This effect was occasionally encountered with fast simulation studies and data studies in other decay channels, but was not observed in the data and decay mode used for the lifetime measurement in this thesis.

5.6.4 Finding $\boldsymbol{S_W}$

The matrix $\boldsymbol{S_W}$ can be written as equation 5.39, where $\boldsymbol{v_s}$ and $\boldsymbol{v_b}$ are the vectors describing the acceptance of pure signal and pure background events respectively. The definition is taken from equations 5.30, 5.31 and 5.32.

$$\boldsymbol{S_W} = \sum_{\substack{Signal\\events}} | (\boldsymbol{v_s} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{v_s} - \overline{\boldsymbol{m_s}}) | + \sum_{\substack{Background\\events}} | (\boldsymbol{v_b} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{v_b} - \overline{\boldsymbol{m_s}}) | \tag{5.39}$$

Consider first the sideband region. The matrix $\boldsymbol{S_{bk}}$ is defined in equation 5.40 once $\overline{\boldsymbol{m_b}}$ is calculated. The $\boldsymbol{v_i}$ are the event vectors for events in the region under consideration.

$$\boldsymbol{S_{bk}} = \sum_{\substack{Sideband\\events}} | (\boldsymbol{v_i} - \overline{\boldsymbol{m_b}})\rangle\langle(\boldsymbol{v_i} - \overline{\boldsymbol{m_b}}) | \tag{5.40}$$

Using the value of f_s from the mass fit, the vector $\mid \overline{\boldsymbol{m_s}}\rangle$ can also be calculated using equation 5.39. Another matrix $\boldsymbol{S_{bk-as-sg}}$ is defined in equation 5.41.

$$\boldsymbol{S_{bk-as-sig}} = \sum_{\substack{Sideband\\events}} \mid (\boldsymbol{v_i} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{v_i} - \overline{\boldsymbol{m_s}}) \mid \tag{5.41}$$

Finally the matrix $\boldsymbol{S_{sg+b}}$ is defined in equation 5.42.

$$\begin{aligned}\boldsymbol{S_{sg+b}} &= \sum_{\substack{Signal\\region\\events}} \mid (\boldsymbol{v_i} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{v_i} - \overline{\boldsymbol{m_s}}) \mid \\ &= \sum_{\substack{Signal\\events\ insignal\\region}} \mid (\boldsymbol{v_s} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{v_s} - \overline{\boldsymbol{m_s}}) \mid \\ &+ \sum_{\substack{background\\events\ in\ signal\\region}} \mid (\boldsymbol{v_b} - \overline{\boldsymbol{m_s}})\rangle\langle(\boldsymbol{v_b} - \overline{\boldsymbol{m_s}}) \mid \end{aligned} \tag{5.42}$$

The 3 matrices defined in equations 5.40, 5.41 and 5.42 can all be calculated and combined to give the desired matrix $\boldsymbol{S_W}$ as shown in equation 5.43.

$$\boldsymbol{S_W} = \boldsymbol{S_{sg+b}} - \alpha \boldsymbol{S_{bk-as-sig}} + \beta \boldsymbol{S_{bk}} \tag{5.43}$$

where

$$\begin{aligned}\alpha &= \frac{\mathrm{N}^0 \text{ Background in signal region}}{\mathrm{N}^0 \text{ Background in sideband}} \\ \beta &= \frac{\mathrm{N}^0 \text{ Background in signal and sideband region}}{\mathrm{N}^0 \text{ Background in sideband region}}\end{aligned} \tag{5.44}$$

To do the above calculations a single signal fraction, f_s, is not used. Instead, the calculations are optimized by using the signal probability for each event derived from the mass fit, i.e, the information that events near the center of the B mass peak have a larger signal probability than those at the edges of the signal window is used.

5.6.5 Fisher Scalar

To get the direction $\mid \boldsymbol{w}\rangle$, the formula in equation 5.37 is used with $\boldsymbol{S_W}$ and $\mid \overline{\boldsymbol{m_s}}$ -$\overline{\boldsymbol{m_b}}\rangle$ as calculated above. To get the Fisher scalar for every event the scalar product $\langle \boldsymbol{w}.\boldsymbol{v_i}\rangle$ is taken for every event. This variable, referred to as the Fisher scalar, characterizes the acceptance function for that event, and therefore the factor in the likelihood $P(s \mid A_{\rm i})$ is now replaced by $P(s \mid fish_i)$. It is not necessary to model how the Fisher scalar is itself distributed but only how the signal fraction varies as function of the Fisher scalar.

At this point it becomes instructive to look at simulated data and see what distribution the Fisher scalar may have, how the signal fraction varies as a function of the Fisher scalar, and what possible parameterizations could be used. What cannot be done is to use MC, realistic or otherwise, to make a template, as this template then simply incurs all systematic errors associated with a Monte Carlo dependent method to correct for the bias due to the impact parameter based trigger.

As an example, the Fisher scalar distribution of signal and background events from a fast MC simulation is used. The fast simulation itself is described in the next chapter. The distribution of the Fisher scalar variable is shown in figure 5.12 with the MC signal events in red and the MC background events in blue. The signal fraction as a function of Fisher scalar can be obtained from this MC by simply dividing the signal events by the total events for each bin. This distribution is shown in figure 5.13.

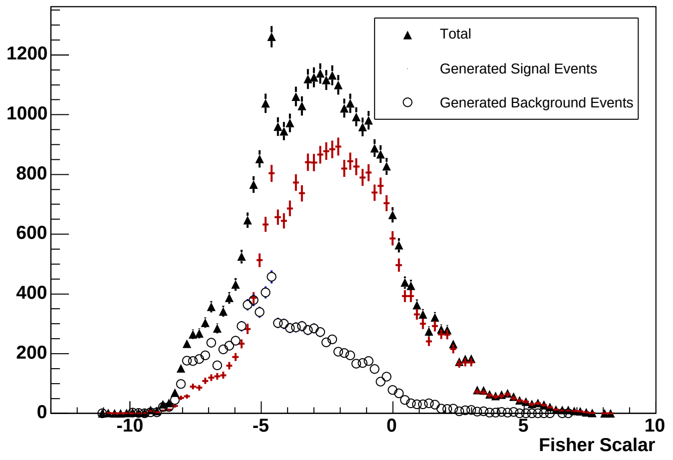


Figure 5.12: The distribution of Fisher scalar using fast simulated data. Signal Monte Carlo is red and background Monte Carlo is blue. It is clear that their distributions are different.

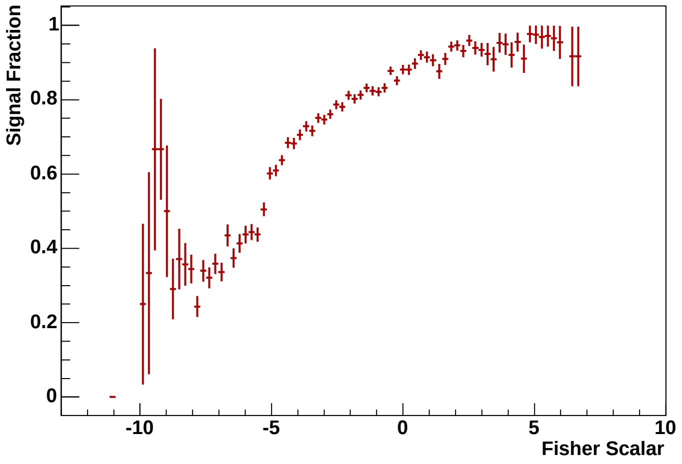


Figure 5.13: True variation in signal fraction as a function of fisher scalar using a sample of simulated data.

To parameterize this distribution, a first attempt might simply be to bin the distribution into 10 bins and fit the signal fraction in each bin, adding 10 fit parameters to the likelihood fit. The advantage of this method is that it is entirely general and the method could be used to test a variety of scenarios without the need for detailed re-parameterization. However the disadvantage of simple binning is that the signal fraction is observed to vary rapidly in some regions, and therefore a more robust solution is needed. A rapidly varying function has more changes over a short range. An interpolating function provides a means to smoothly vary the function over a short range in variable. The Lagrange interpolating polynomials [63] [64] take a series of values for a function and interpolate a smooth curve through the points. Lagrange interpolating polynomials are used to parameterize the signal fraction as a function of the Fisher scalar, as they retain the advantageous features of binning the distribution, but also provide the means to give a smooth function.

To use the Lagrange interpolating polynomials, the order of the polynomial must be chosen, and this gives the number of fit parameters. The distribution is binned into n bins where n is the order of the polynomials. The fit parameter fits the signal fraction at the center of a particular bin and the interpolating polynomial fits a smooth function through those points.

In figure 5.13, there are low statistics in the tails of the fisher scalar distribution. These regions are difficult to fit using the Lagrange interpolating polynomials, as with low statistics there is nothing to stop the fitted function rapidly oscillating between 0 and 1 for these bins. The final fit function is chosen such that the tails of the distribution, the first and last bin, are fit to a single signal fraction, and that in between the first and last bin the Lagrange interpolating polynomial of order 11 is used.

Therefore 13 parameters are added to the fit. The fit itself does not have access to a histogram such as the one shown in figure 5.13. It has access only to the value of Fisher scalar for each event. There is no knowledge over which events are signal and which are background. Using the information from the mass and lifetime fits that are simultaneously fitted with the signal fraction as a function of Fisher scalar the fit can converge upon a Lagrange interpolating polynomial. For the MC sample shown above, the fit function is overlaid, and there is good agreement between the truth signal fraction as a function of Fisher scalar, and that given by the fitted Lagrange interpolating polynomial.

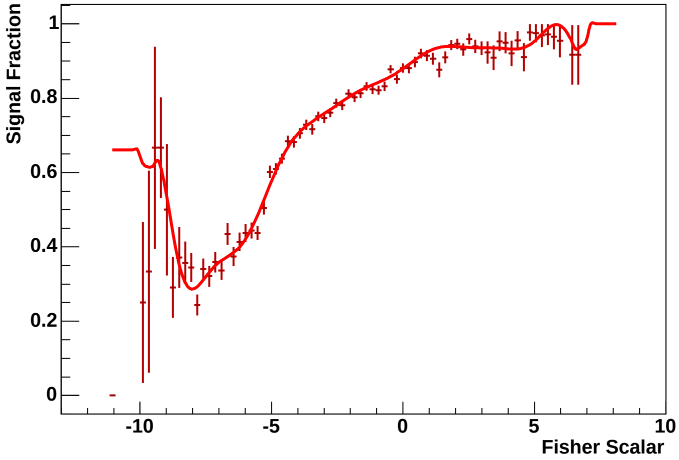


Figure 5.14: The points show the truth signal fraction in the fit from a sample of fast MC. The red line shows the fitted function. The points are shown to demonstrate that the correct signal fraction has been found. The fit itself does not know which events are signal or background and so does not know the truth, yet manages to match it well.

Chapter 6

Fast Monte-Carlo Simulation

This chapter describes the features of a fast Monte Carlo simulation used to test the full PDF derived in chapters 4 and 5. The fast Monte-Carlo simulation generates long-lived decays with kinematics modeled on the data and emulates the acceptance of the detector and trigger realistically. Due to its speed, the fast simulation can generate the several thousand samples required to fully test this method and assess the systematic errors. This chapter also contains the studies that use the fast simulation to demonstrate that the PDF derived in chapters 4 and 5 can remove the biases due to the impact parameter based trigger. The effect of the assumptions made in deriving the PDF, and other detector effects on the final results, is discussed in chapter 8

6.1 The Fast Simulation

The method for using the calculated acceptance to correct for biases introduced by the impact parameter based triggers needs to be verified. To be able to make quantitative statements on the performance of this method, and in particular the size of any remaining residual bias, the PDF derived in chapters 4 and 5 needs to be tested on several thousand samples of simulated data. While a full, detailed simulation of the CDF detector and trigger does exist, it is far too CPU intensive due to the modeling of all the detector components such as the response of all the wires in the COT. It will not be possible to use this to generate the required size of samples. However it is possible to generate B decays and model the response of the detector. This way, the millions of events required for rigorous testing of the method can be easily generated.

The purpose of the fast simulation is to create large samples of events that can be treated like data and be used to validate the method and assess systematics. It is not possible to create a fast simulation where all the features of data are reproduced in exact detail, but the aim is to broadly reproduce the kinematic variables seen in data. As information from the simulation is not used as an input into the result measured on data, minor discrepancies between the simulation and data are not important.

The measurement of the B meson lifetime using the simulation-free technique depends on all the kinematics of the B meson and all daughter tracks. All the kinematic information is required to calculate $A_{\rm i}$. Therefore, a fast simulation must also produce all these quantities so that acceptance functions can be calculated. The fast simulation has three main steps, each of which is elaborated upon further in this chapter. The first step is to create the B meson and decay it to all the final state daughter tracks with their associated momenta and positions. The second step is to apply a very simple detector simulation. The third step is to apply the trigger and then the selection criteria. If the selection criteria are satisfied then the event is in the sample. At this point, the event can be treated as if it was from data; acceptance functions can be calculated and the lifetime of the sample can be extracted using the likelihood described in chapter 4.

Kinematic and lifetime variables are drawn randomly from parameterized functions or histograms. The software required to do this resides in the CERN ROOT package.

The uniform deviate is drawn using the Mersenne twistor generator [65]. As the periodicity of this algorithm is approximately 10^{6000} this generator is more than sufficient for the generation of the fast simulation. This is the recommended generator to avoid problems with correlations. To compare the fast simulation with the data, background subtracted histograms of the signal region are created. In this chapter, where comparison plots are shown, the distribution in data and in simulation have been normalized to one event. The following describes the generation of the default simulation. For various studies, deviations from the standard are introduced and they are explained in the relevant sections.

There are 3 different trigger paths in data, with Low, Medium and High P_t requirements as described in section 3.6. To create the fast simulation, 3 sets of events are created, one for each trigger path. After this has been done, the events for the sample are chosen to match the relative number in each trigger path seen in data.

6.2 Signal Event Generation

6.2.1 The B Meson

In simulating the B meson the aim is to generate a lifetime and momentum in such a way that the distribution of remaining events after the trigger emulation matches that of the events observed in data. To do this, the lifetime and momentum of the B meson must be generated from a distribution that represents that variable before the trigger is applied. The distributions observed in data cannot be used as the generating function as they are the distribution after the trigger, not before.

A B meson is created at the primary vertex. The direction of travel is chosen so that the overall distribution before the trigger will be uniform in ϕ and η. These two choices are made because there is no preferred direction in the transverse plane and production cross sections are flat in the η variable. Once the ϕ and η have been randomly assigned the momentum vector of the B meson, $\boldsymbol{P(B)}$, is given as:

$$\boldsymbol{P(B)} = (cos(\phi)sin(\theta), sin(\phi)sin(\theta), cos(\theta)) \times P_B \tag{6.1}$$

where P_B is the magnitude of the B momentum.

To create simulated events so that the final transverse momentum distribution matches the distribution seen in data, the parent distribution is created in the following way and the value of P_B is drawn at random from this distribution. First, for data events passing each trigger path, a background subtracted histogram of the B momentum in the signal region is created. The background template that is subtracted comes from the upper sideband. Every entry in the histogram of P_B is scaled down by a factor of 0.72. This distribution, now with a lower mean, yields a momentum distribution of signal closest to what is observed in data. It is more important to match the data in transverse momentum than in the 3-dimensional momentum, as it is the transverse variables such as the B P_T and track P_T that are used to calculate the acceptances that are at the center of this technique. A plot that compares the transverse momentum distribution of simulated events that pass the trigger and the background subtracted signal region is shown in figure 6.1.

After momentum, the lifetime t_B is assigned at random from an exponential decay with decay constant $\tau_B = 1.638$ ps , the mean B^+ lifetime taken from the PDG [11]. The lifetime of each event is then smeared by adding a value generated randomly from a single Gaussian centered on zero with width 0.009 ps (or σ_{ct}=26 μm). This is the resolution model chosen for the data fit as described in section 5.3. Alternatives to

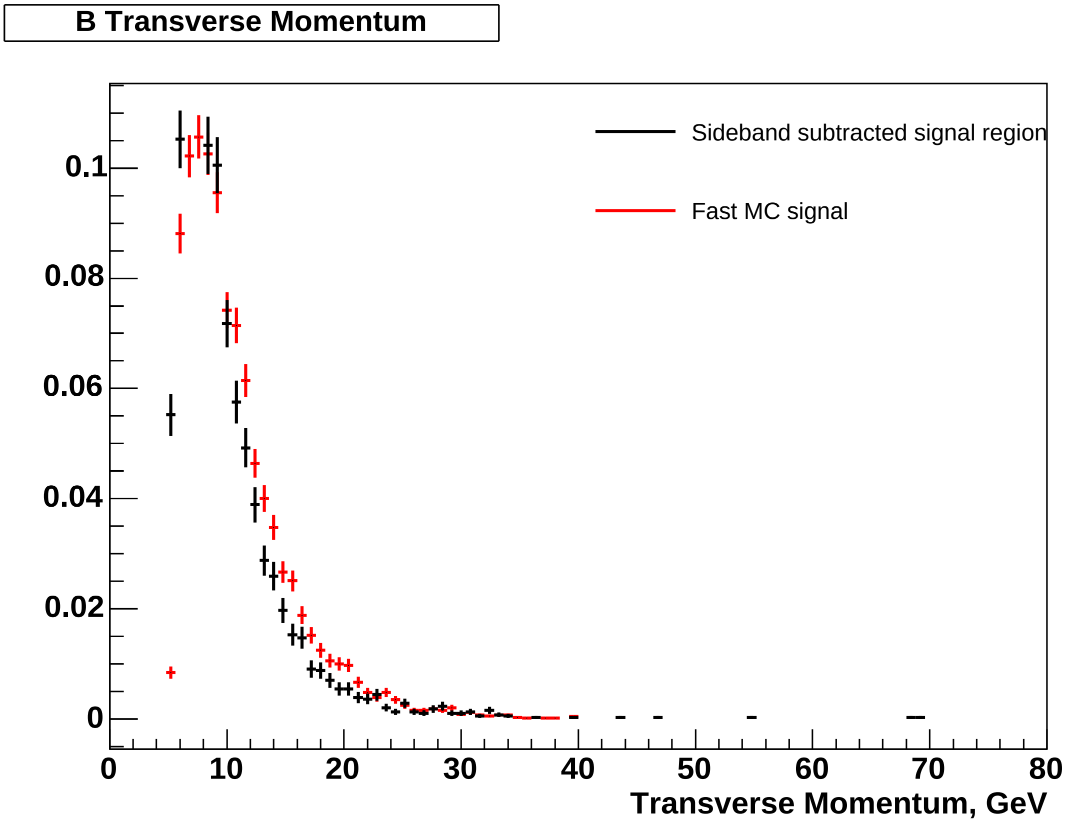


Figure 6.1: A comparison of transverse momentum distributions between the signal in data and simulated signal events

this resolution model and the effect on the fitted lifetime are discussed in chapter 8.

The mass of the B meson m_B is assigned at random from a Gaussian with mean M_B, the PDG mass, and width 18 MeV.

The coordinates of the decay point of the B meson, $Vtx(B)$ can then be calculated:

$$\boldsymbol{Vtx(B)} = ct_B \frac{\boldsymbol{P(B)}}{M_B} \tag{6.2}$$

where M_B, the B meson mass, is the PDG value of 5.279GeV and the factor $\frac{\boldsymbol{P(B)}}{M_B}$ is required to boost to the laboratory frame.

6.2.2 Calculation of the kinematics of the daughter particles

In the rest frame of the B meson, the decay daughters D and π_B will have equal $|\boldsymbol{P}|$ and travel in opposite directions, and their distribution will be isotropic. In the rest frame the magnitude of the momentum of the decay daughters is given by

$$|\boldsymbol{P}|^2 = \frac{(m_B^2 + m_D^2 - m_\pi^2)^2 - 4m_D^2 m_B^2}{4m_B^2} \tag{6.3}$$

where m_B, m_D, m_π are the masses of the B meson, and its decay daughters D and π mesons respectively.

In the rest frame of the B meson, the direction of the daughter pion is distributed isotropically, as this is the decay of a scalar particle to 2 other scalar particles. Hence the ϕ direction of the pion is chosen randomly so that the overall distribution lies uniformly between 0 and 2π radians. The direction of flight of the pion has its θ coordinate assigned by choosing a uniform deviate and assigning this value to $\cos(\theta)$. With this information, the momentum of the pion in the B rest frame is defined and the momentum of the D is opposite in direction and equal in magnitude. These vectors are then transformed to the laboratory frame, thereby giving the momentum of the pion and the D in the laboratory frame.

The D meson also travels a distance before decaying into a kaon and pion. The lifetime of the D meson, t_D, is generated randomly from an exponential with mean 0.41 ps (the mean measured lifetime of the D^0 meson from the PDG [11]). The secondary vertex of the D meson is given by

$$\boldsymbol{Vtx(D)} = \boldsymbol{Vtx(B)} + ct_D \frac{\boldsymbol{P(D)}}{M_D} \tag{6.4}$$

The laboratory momenta for the kaon and pion daughters from the D are calculated in the same way as for daughters of the B meson. In the D rest frame, the pion daughter direction is isotropic, and a direction is chosen at random. Momentum vectors are then calculated and then transformed to the laboratory frame.

The simulated event now has 3 tracks, each with a momentum vector, and is associated with either the B or D vertex. For each track, the impact parameter ϕ_0 and η can be calculated. These quantities are defined as the offline quantities. No smearing of the track parameters is applied.

6.2.3 Detector Simulation and Selection

Having generated the decays, they are now required to be in the fiducial volume of the detector. The efficiencies of certain detector components are also taken into consideration. The tracking volume of the COT does not extend beyond η=2 so any events with tracks $\eta \geq 2$ are removed, as it can be assumed that the offline tracks would not have been found. The SVT single track finding efficiency is approximately 70% as shown in figure 4.8. To emulate this, a random 30% of all tracks are given no simulated SVT information. For tracks that are given SVT information, the only difference introduced between the offline and SVT quantities is in the impact parameter. As introduced in section 4.3.2, in data there is a difference between the SVT and offline impact parameters and this must be accounted for in data, and therefore this difference is included in the simulation. For the default simulation, the offline quantity is rounded to the nearest multiple of 10 microns to emulate the discretisation of the SVT.

As all the track and vertex information is known, all other variables that are used in selection criteria as listed in section 3.6 are calculated, and the selection criteria are applied to the simulated events. If all these criteria are satisfied, the event is then passed onto the trigger emulation, which applies all the trigger cuts to the tracks and rejects any events that do not pass the trigger criteria. The data passing each of the three trigger paths- low, medium and high P_T- are simulated separately. For each final sample, the number of events from each trigger path is adjusted to match that observed in data. Table 6.1 gives the abundances observed in each path for the data in the fit in two regions of mass. To find the ratio that should be applied to signal events a background subtraction is carried out. The observed ratio of events in different trigger paths is 57.5 : 35.1 : 7.4 for the Low : Medium : High P_T paths respectively.

Trigger Path	$5.25 < M_B < 5.31$ GeV	$M_B > 5.37$
Low	13304	2075
Medium	8246	1463
High	1744	321

Table 6.1: Trigger abundances observed in data.

6.3 Simulated Signal and Data Comparison

The method of correction for the impact parameter bias is via the calculation of the acceptance function. Furthermore, as shown in the derivation of the full likelihood in section 5.1, the probability that a particular event is signal given its acceptance cannot be ignored. Therefore the acceptance is a crucial observable of the event and it is important that the acceptance functions of signal and background match those of data to a reasonable approximation. To compare the acceptance functions from simulated data and the sideband subtracted signal region, the acceptance functions are categorized in the same way as done for the Fisher discriminant part of the analysis, i.e, in terms of track pairs (1, 2 or 3) that are able to pass the trigger as demonstrated in figure 5.11 Each category is compared separately. To compare the distribution of acceptances the mean acceptance is constructed and compared, as it is difficult to visualise two distribution of functions. The mean gives information about the sample, and the means of each sample can be compared. A histogram for the mean acceptance for each category is drawn by summing the histograms representing the acceptance for each event and dividing by the number of events in the sample. For the data sample, the background subtracted plots of the mean acceptances can be made for events lying in the signal region. Figures 6.2, 6.3 and 6.4 show the comparison between data and simulated events.

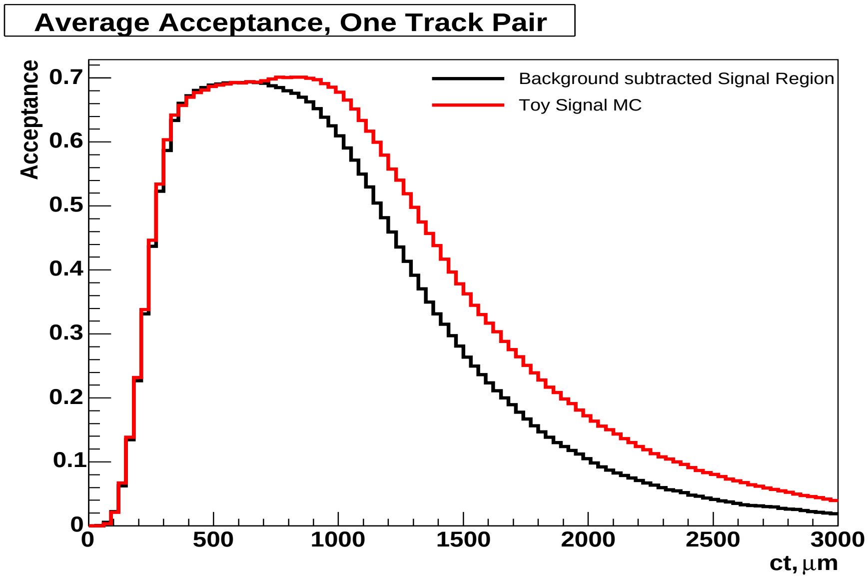


Figure 6.2: A comparison of the mean acceptance function for one track pair in simulation and signal region in data. The mean acceptance in red is that of the simulated data, and the black is that of the background subtracted signal region.

From figures 6.2 to 6.4 it can be seen that the mean acceptances of simulation and data have a similar breadth and height. These are similar enough for the purposes of testing the method, recalling that no information from the simulation is used in the fits. In general the acceptance functions of the simulation are wider than those observed in

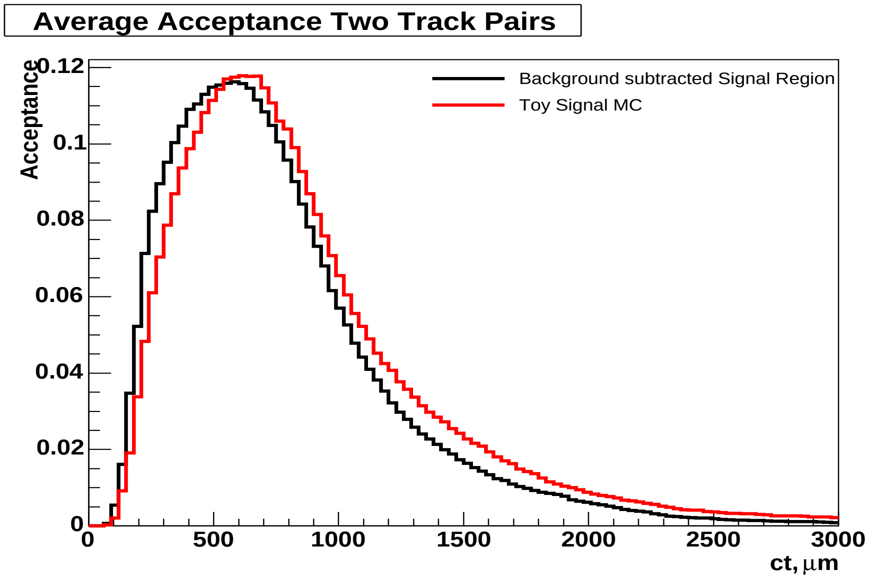


Figure 6.3: A comparison of the mean acceptance function for two track pairs in simulation and signal region in data. The mean acceptance in red is that of the simulated data, and the black is that of the background subtracted signal region.

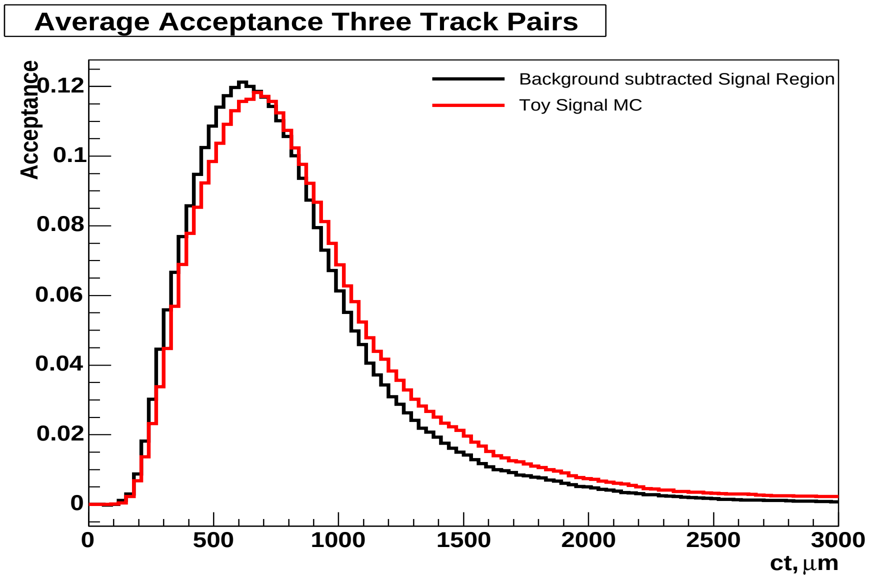
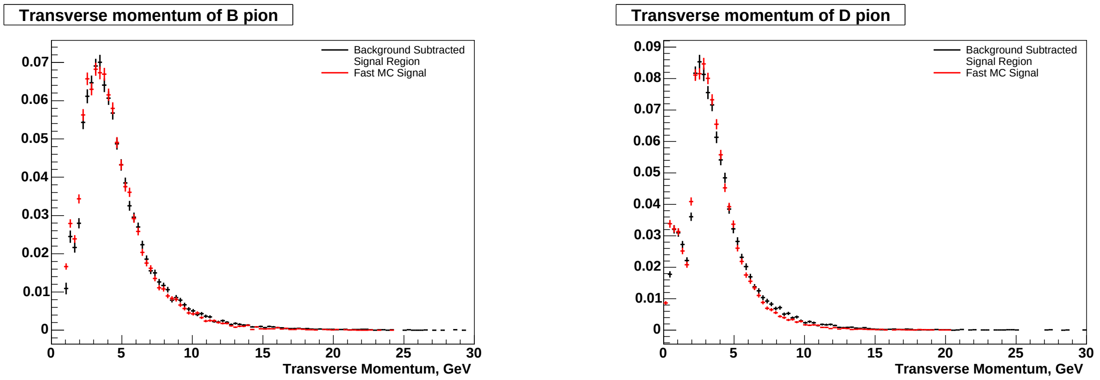


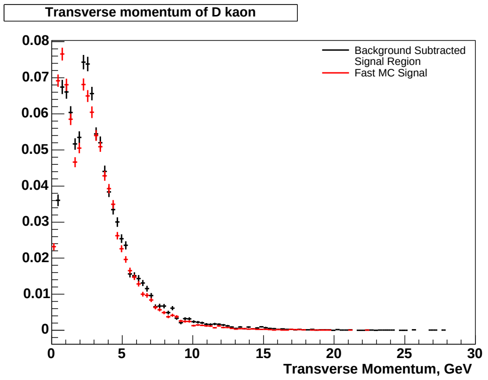
Figure 6.4: A comparison of the mean acceptance function for three track pairs in simulation and signal region in data. The mean acceptance in red is that of the simulated data and the black is that of the background subtracted signal region.

data, and this is most apparent in the one track pair case. It is also instructive to look at a comparison of the track impact parameter, transverse momentum, and η from simulation and the background subtracted signal region. In addition, other quantities to compare are the opening angle between any particular pair of tracks. Comparison plots of the normalized quantities follow in figures 6.5 to 6.9.



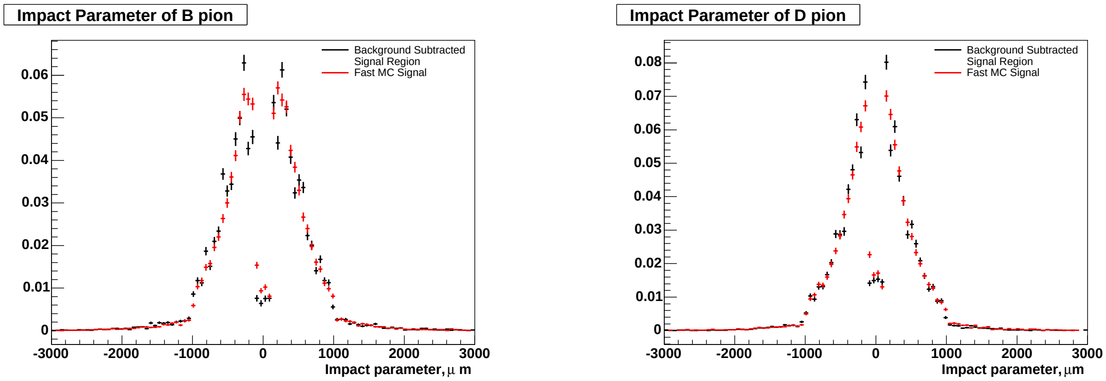
(a) The transverse momentum spectrum of the pion from the B

(b) The transverse momentum spectrum of the pion from the D



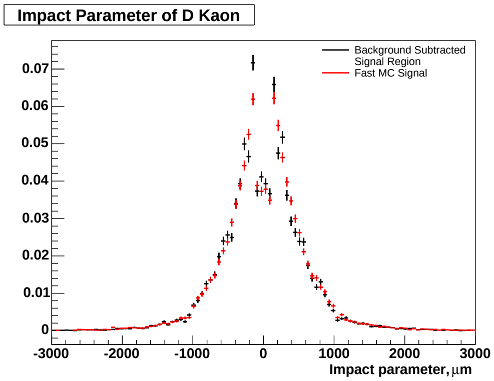
(c) The transverse momentum spectrum of the kaon from the D

Figure 6.5: A comparison of the sideband subtracted signal region in data and the transverse momentum of tracks derived from the fast simulation. The sculpting effect of the trigger accepting events with two tracks that have $P_T \geq 2GeV$ is visible. The distribution of simulated events is drawn in red, while that of background subtracted data is shown in black. The area of each distribution is normalized to one.



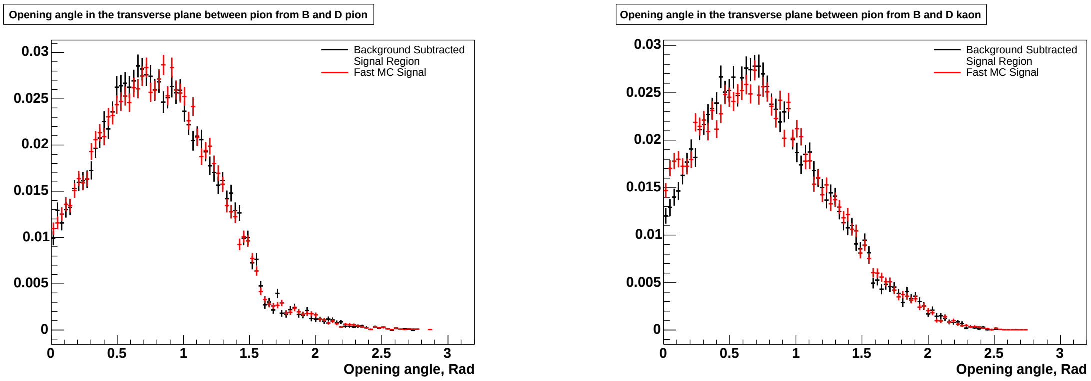
(a) The impact parameter distribution of the pion from the B

(b) The impact parameter distribution of the pion from the D



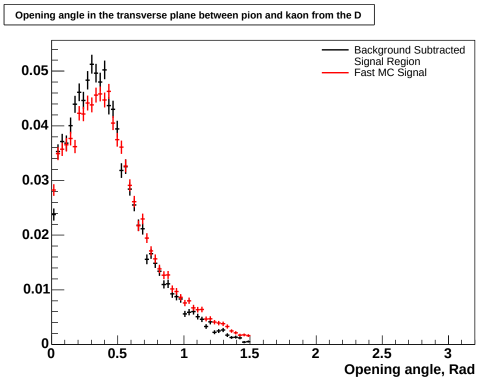
(c) The impact parameter distribution of the kaon from the D

Figure 6.6: A comparison of the sideband subtracted signal region in data and the fast simulation. These plots show the impact parameter distribution. The sculpting effect of the trigger is visible here too as only tracks with $d_0 \geq 120\mu m$ are used in the trigger decision. The distribution of simulated events is drawn in red, while that of background subtracted data is shown in black. The area of each distribution is normalized to one.



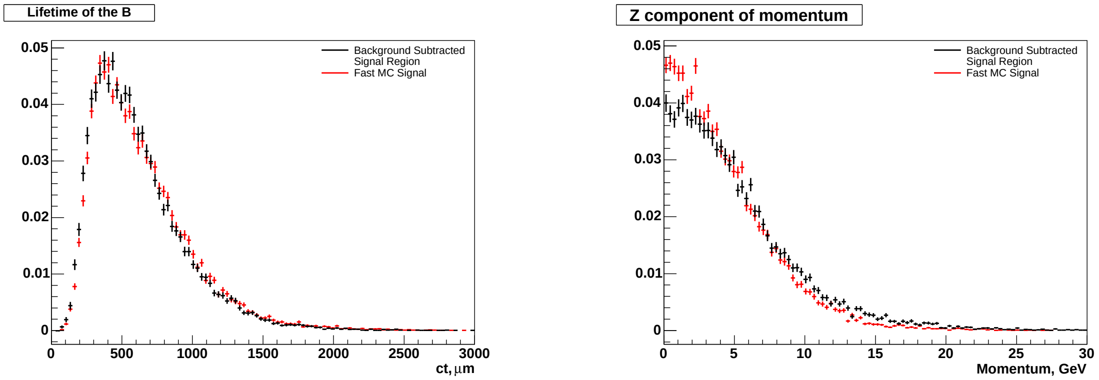
(a) The opening angle distribution between the pion from the B and the pion from the D

(b) The opening angle distribution between the pion from the B and the kaon from the D



(c) The opening angle distribution between the pion and kaon from the D

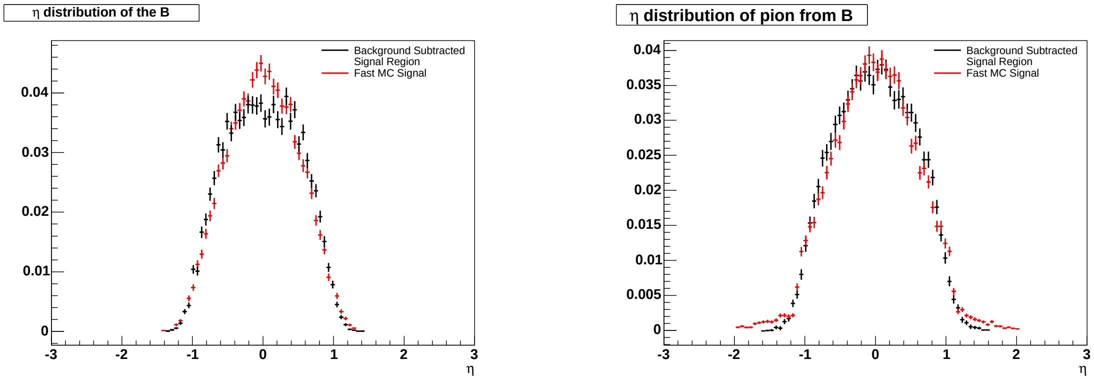
Figure 6.7: A comparison of the sideband subtracted signal region in data and the fast simulation. These plots show the opening angle in the transverse plane distribution. The distribution of simulated events is drawn in red, while that of background subtracted data is shown in black. The areas of each distribution are normalized to one.



(a) The *ct* distribution of the B meson

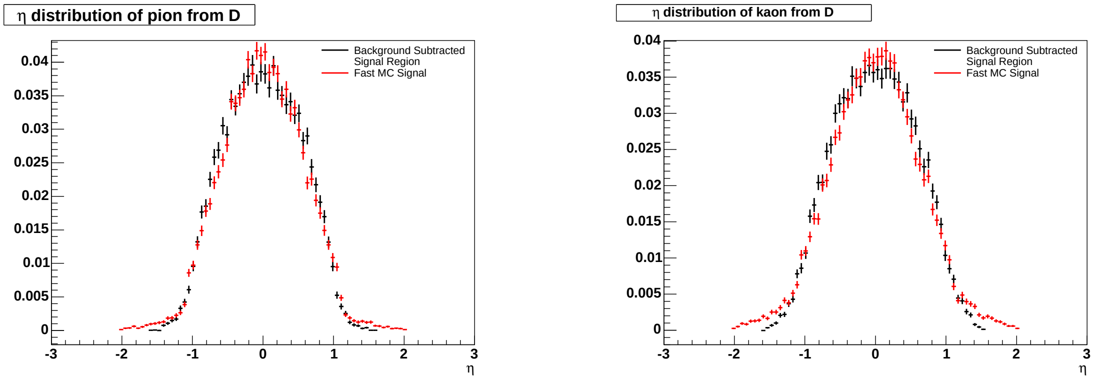
(b) The P_z momentum component distribution of the B meson

Figure 6.8: A comparison of the sideband subtracted signal region in data and the fast simulation. These plots show the distribution of the z component of momentum and the B lifetime. The distribution of simulated events is drawn in red, while that of background subtracted data is shown in black. The area of each distribution is normalized to one.



(a) The η distribution of the B meson

(b) The η distribution of the pion from the B



(c) The η distribution of the pion from the D

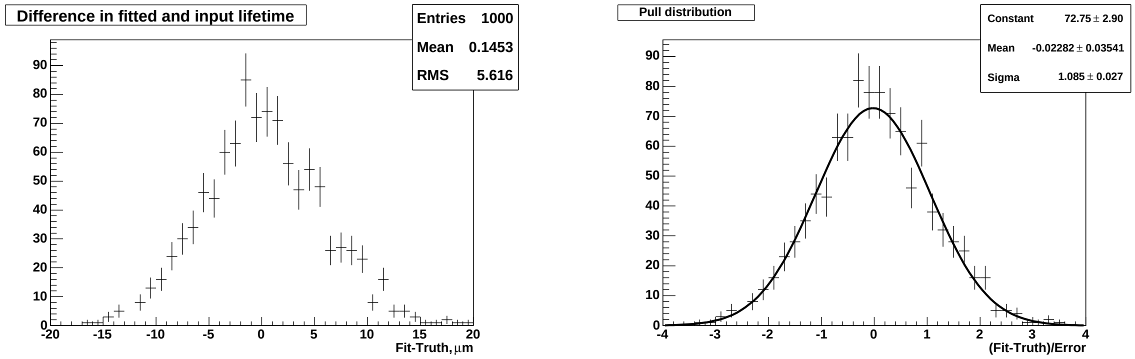
(d) The η distribution of the kaon from the D

Figure 6.9: A comparison of the sideband subtracted signal region in data and the fast simulation. These plots show the η distributions of the tracks and that of the B meson itself. The distribution of simulated events is drawn in red, while that of background subtracted data is shown in black. The area of each distribution is normalized to one.

Figures 6.5- 6.9 show that there is reasonable agreement between the quantities in the fast simulation after the trigger has been applied and the quantities in the background subtracted data. This fast simulation has been created using minimal input. The result of this minimal input is a fast simulation that is, in general terms, very like the data. As all detector efficiencies, tracking algorithms, magnetic field and the underlying event itself have not been simulated it is unsurprising that the fast simulation does not match the data perfectly. One improvement that could be made is to reject events so that there is an improved match to the B meson η and P_z distributions. However as the fast simulation is required to generate millions of events adding time to generation is undesirable. The fast simulation is sufficiently like data to assess the effect of assumptions in the PDF and to test whether such a method for correcting the bias in selecting events with high impact parameter is capable of extracting the true lifetime correctly. The added advantage is that the fast simulation is very flexible. The response of detector components can be changed or incorporated, which is useful to test sources of systematic uncertainty, and test the robustness of the method.

6.4 Using the Simulation to Test the Signal Only PDF

To test whether the constructed PDF can correct for the selection bias, 1000 independent samples of 24,000 events each are generated. For each sample, the likelihood is calculated and maximized to give the most probable lifetime. In practice, the lifetime is considered in terms of decay length, and the input mean proper decay length is $491.1\mu m$. Figures 6.10(a) and 6.10(b) show residual and pull distributions. The residual is the difference between the input lifetime and fitted lifetime for each sample. The pull is the residual divided by the error on the fitted lifetime. Figure 6.10(b) is fitted to a Gaussian. This Gaussian is consistent with a Gaussian of mean 0 and width 1. This demonstrates that the method does correct for the bias induced by the trigger.



(a) The residual distribution for the extracted lifetime in 1000 samples of Fast MC

(b) The residual distribution for the extracted lifetime in 1000 samples of Fast MC

Figure 6.10: Fast MC is used to demonstrate how calculating the acceptance function and applying the PDF described in equation 4.15 can remove the bias introduced by the impact parameter based trigger

In the standard fast simulation generation, the only difference between the SVT-measured impact parameter and the offline impact parameter is that the SVT impact parameter is rounded to the nearest multiple of 10 microns to emulate the discretization of the SVT. To further test how the method performs with larger differences between the SVT and offline impact parameters, some changes are introduced. The acceptance function is calculated on the SVT-measured impact parameters, where the SVT-measured impact parameter at any point is different from the offline impact parameter by the same amount as at the point of decay. Two tests are carried out on pseudo experiments. For each test 500 independent samples of 24,000 signal events are generated. The tests are

- Experiment 1: The SVT-measured impact parameter is distributed normally around the offline impact parameter with a width of 35 μm. The value is then rounded to the nearest multiple of 10 to emulate the discritization of the SVT measured impact parameters. The motivation for this test comes from figure 4.4, which shows the difference between offline and SVT measured impact parameters. This histogram is reasonably modeled by a Gaussian. This is also the quoted resolution of the SVT [48]

- Experiment 2: The SVT-measured impact parameter is normally distributed around the offline impact parameter with a width of 35 μm and in addition an offset of a further +35 μm is added. This means that the mean SVT measured impact parameter is on average larger than the offline impact parameter. This tests an extreme situation that is unlikely to be realized in the data. This alteration tests the assumption that the method is robust against such differences, and that the true nature of the distribution between the SVT and offline impact parameter does not need detailed understanding.

500 samples of each type were generated. The pull and residual means are listed in Table 6.2. The means of the pull distribution are statistically consistent with zero, indicating no bias. The mean residuals are also small$\approx$0.1 μm for the realistic scenario. This is negligible when compared to the smallest systematic error of 0.3 μm, as will be shown in chapter 8. These tests demonstrate that the method is robust against differences between the SVT and offline impact parameters, and that precise understanding of the relation between the two is not required.

Input $c\tau$ (μm)	Residual mean	Pull mean
No SVT error distribution	0.1	-0.02±0.04
Gaussian SVT error distribution	-0.1	-0.02±0.05
Gaussian + offset SVT error distribution	-0.2	-0.07±0.05

Table 6.2: Summary of validation tests using different functional forms of the difference between the offline and SVT measured impact parameter.

6.4.1 Changes in Efficiency Due to Changes in Hardware and Software During the Period of Data Taking

The single track finding efficiency is an important part of the acceptance calculation, and the efficiency is dependent on detector performance. There have been changes to the hardware and software in the track finding algorithms during the data taking period to improve performance. Over the period of data collection there have been four changes to the SVT [66]. The SVT tracks are found by pattern matching of silicon hits to predefined "roads". The changes are:

- (Period 1) The initial running from start to 12/26/02
- (Period 2) On 12/26/02 begin using roads that only require hits in 4 silicon layers as opposed to 5.
- (Period 3) On 07/24/04 use new "roads" and new geometry file
- (Period 4) On 07/22/05 installation of new memory (AM) board used in storing "roads". Increases the computational power of the SVT.
- (Period 5) On 08/31/05 move to using 128K "road" patterns from 64K.

The impact on the single track finding efficiency due to these changes is determined first. If the efficiency is plotted in these different periods, it is seen that the last two

Data Period	Mean efficiency
Period 1	0.606 ± 0.025
Period 2	0.662 ± 0.009
Period 3	0.741 ± 0.007
Period 4	0.751 ± 0.017
Period 5	0.757 ± 0.011

Table 6.3: The mean efficiencies in different data taking periods defined in section 6.4.1 by changes in the SVT.

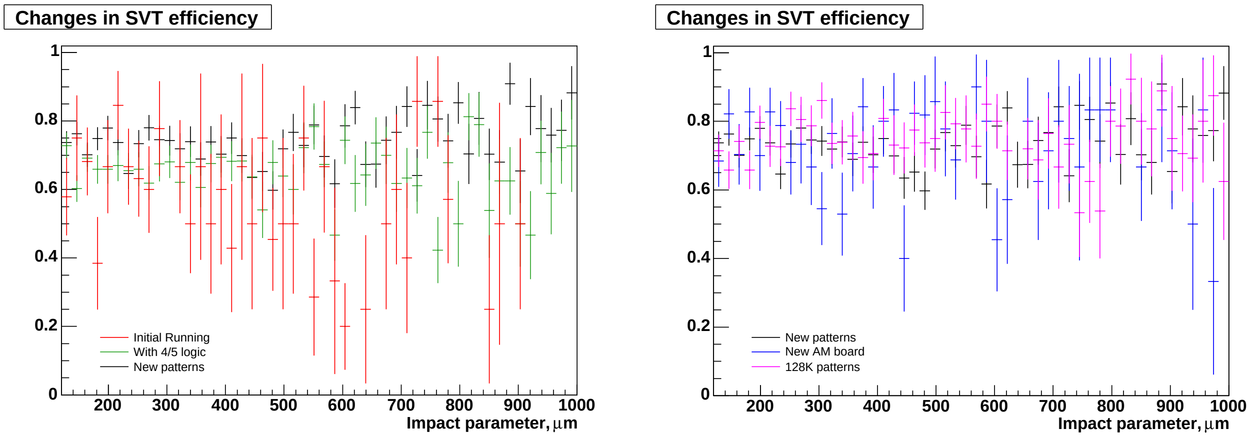


Figure 6.11: SVT track finding efficiency as a function of impact parameter and split into different time periods. The left plot shows the first two changes that did affect the efficiency in the range of interest. The right plot shows the final two changes which did not affect significantly the data in the range of interest.

had little effect on the efficiency as a function of impact parameter in the range 0-1000 microns, the only region of interest. This is expected, as these changes were made primarily to improve the efficiency at higher impact parameters. The single track finding efficiency beyond 1000 microns is irrelevant to this analysis as all tracks with matches above 1000 microns are deemed to have been not found. The first two changes, however, made significant increases to the efficiency. These results are shown in two plots in figure 6.11. The mean efficiency for each period is listed in table 6.3. This is determined by a fit to a zeroth order polynomial to the efficiency in each data sample.

As there is an observed change in the single track finding efficiency the impact this may have on the lifetime measurement must be considered. The effect that these changes may have on the lifetime measurement is studied using the fast simulation. To do this the toy generation is modified so that the signal is generated with three different efficiencies 0.49, 0.66, 0.73 in periods 1, 2 and 3-5 respectively. The values of the efficiency parameters used come from the full fit to data, shown in table 7.2. The percentage of data generated with each efficiency is 3%: 32 % : 65%, as this is the proportion of data collected in each period. For each sample, 24,000 events are generated, and there are 1000 samples which are fitted with one floating efficiency parameter for all the data in one sample. The pull distribution resulting from these fits is shown in figure 6.12. It is observed that the distribution is wide, 1.19±0.03, and not consistent with 1. This implies that the statistical error returned by each fit has been underestimated by approximately 20%. The statistical power of each event is related not only to the width of the acceptance function but also the height which is determined by the efficiency. The fit incorrectly assumes that all tracks with impact parameter between 120 and 1000 μm have the same efficiency. This is likely to be the cause of the statistical error underestimation.

To try and overcome this problem, the samples are fitted again but this time there are three floating efficiency parameters. Each event will only be sensitive to one parameter. For example, all events that were generated with the efficiency 0.73 will be used to fit one efficiency parameter. The pull distribution of these fits where the three efficiencies are fitted separately is shown in figure 6.13. The result of these fits is a pull distribution where the width is consistent with 1.

Therefore, in the fit on data, three separate $\varepsilon_s$will be fitted where each one is dependent only on events collected during the relevant data taking period.

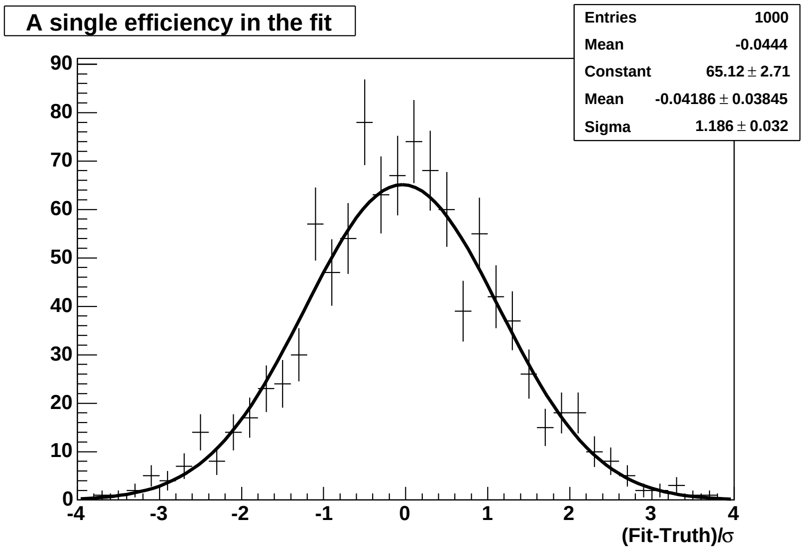


Figure 6.12: The pull distribution for samples that have been generated using 3 efficiencies but are fitted using only a single efficiency parameter.

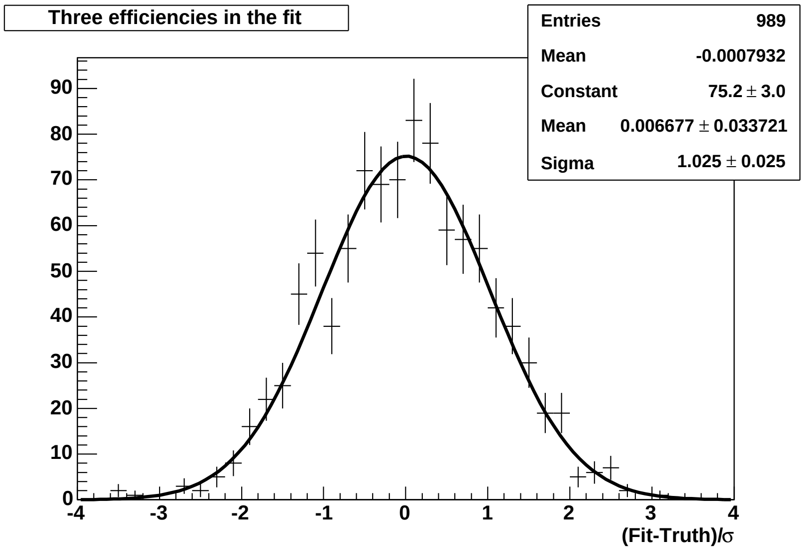


Figure 6.13: The pull distribution for samples that have been generated using 3 efficiencies and where each of the three subsamples is used to fit one of three efficiency parameters

6.5 Creating Simulated Background Data

The same techniques are used for the fast simulation background data as for signal. The upper sideband of the reconstructed data contains “B meson” candidates that decay to “D meson candidates”. The analysis assumes that the kinematic distributions observed in the B meson candidates in the sideband represent the kinematic distributions of the events in the signal region that pass the selection criteria but do not really represent a true signal event (i.e., background events under the signal peak).

For the fast simulation, the lifetime and momentum of the candidate B meson and the lifetime of the candidate D meson need to be generated. The pre-trigger distribution of the candidate B mesons found in the upper sideband is given by the function $y(t)$ as described in section 5.5. Lifetimes are assigned at random according to the distribution $y(t)$. It is not known what the background lifetime distribution for the D meson candidate looks like before the trigger. Therefore the D meson candidate lifetime distribution found in the upper sideband is used. Simulated events are assigned candidate D meson lifetimes at random according to this distribution. While this is an approximation, it turns out that the comparison of the kinematics of the sideband and simulated data is sufficiently good that there is no need to improve the generating distribution for candidate D meson lifetimes.

The momentum distribution used to generate the magnitude of the candidate B momentum P_B is given by the following function:

$$P_B = \frac{e^{-x/5.5} + 5.5}{2} GeV \tag{6.5}$$

where x is any number greater than zero. This distribution was used as it produces a good match between the simulated and sideband data distributions of transverse momentum after the trigger, as seen in figure 6.14

Background events are mixed in the trigger ratio of Low:Med:High as 53.8 : 37.9 : 8.3 as these are the fractions observed in the sideband. The absolute abundances are given in table 6.1. The SVT single track finding efficiency is set to 60%. It is observed to be lower for the background, as in general background events in data have lower momentum tracks, tracks with higher η and tracks with fewer SVT hits. Tracks with these qualities are known to have a lower probability of being found by the SVT. The value of 60% is the approximate value of the efficiency when plotted for events in the sideband. One further change is required to make the kinematic distribution of the simulated background match the kinematic distribution in the sideband. Using the decay technique of calculating the momenta of the daughter particles in the relevant rest frame, the transverse momentum distribution of the pion from the B (π_B) has a similar shape to that of simulated signal. However, in data this distribution is quite distinct between signal and background. Figure 6.15 shows the comparison between the B pion transverse distribution in sideband and signal. There are clear differences, mainly that the sideband distribution in data has a lower mean momentum. To get good agreement between the acceptance functions in sideband and simulation, it is necessary to adjust the simulation to reproduce this distribution of transverse momentum. To do this, a fraction of events with large B pion transverse momentum are removed from the sample so that the generated distribution agrees better with the distribution observed in the sideband. Rejecting these events sculpts the distributions of other parameters as well, such as the impact parameters of the other tracks, and brings them closer to the distribution observed in the sideband.

The comparisons for the acceptance functions of sideband and simulated background are shown in figures 6.16- 6.18. There is broad agreement in shape, breadth

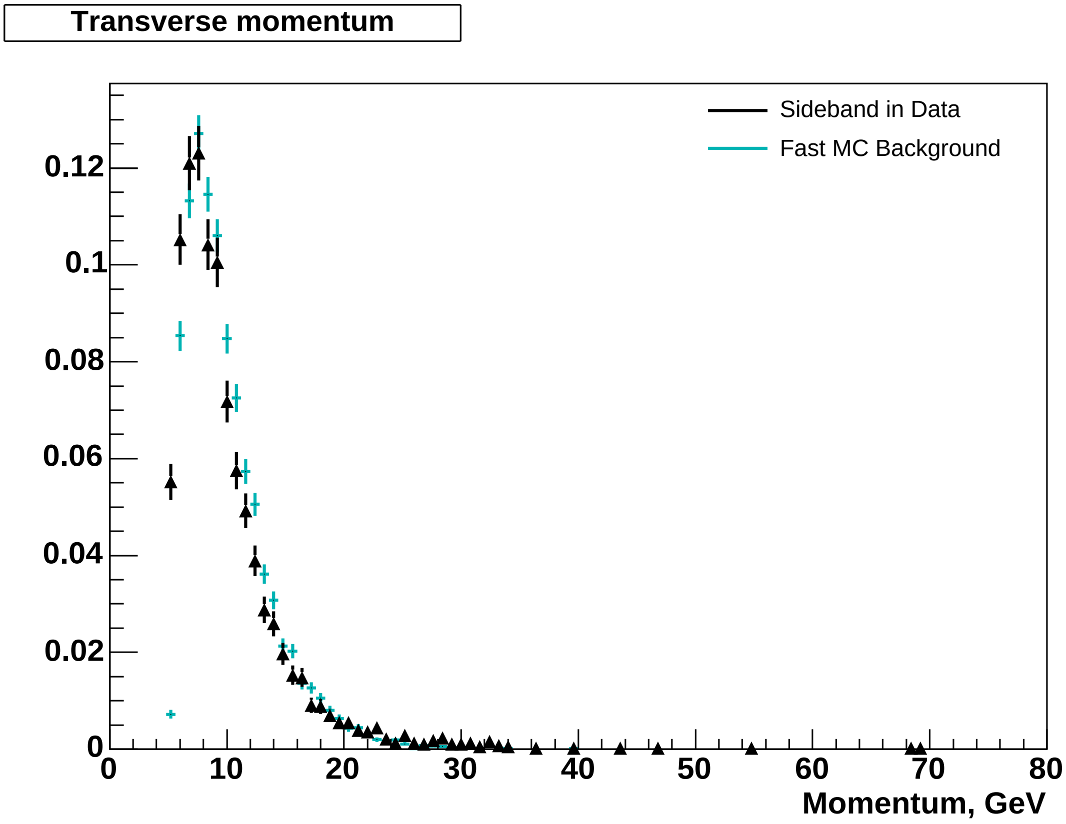


Figure 6.14: A comparison of transverse momentum distributions between the sideband in data and simulated background events

and height, particularly in the case where only one track pair satisfies the trigger conditions. This case is the most dominant contribution to the total acceptance. Also shown are the plots in figures 6.19- 6.23 comparing the sideband data and simulated background in various kinematic quantities. The simulation is very close to the background in the sideband, and so a mix of the simulated signal and background can be used to validate the method.

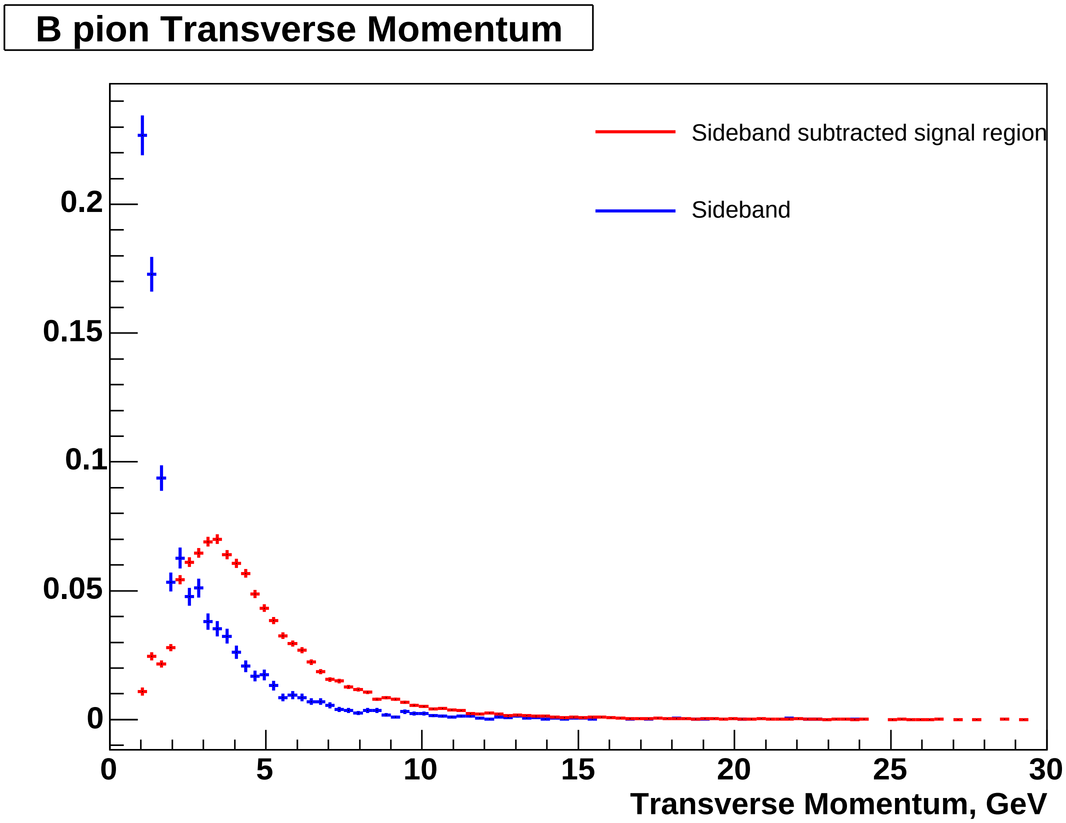


Figure 6.15: A comparison of the B pion transverse momentum distribution between the sideband and sideband subtracted signal region in data

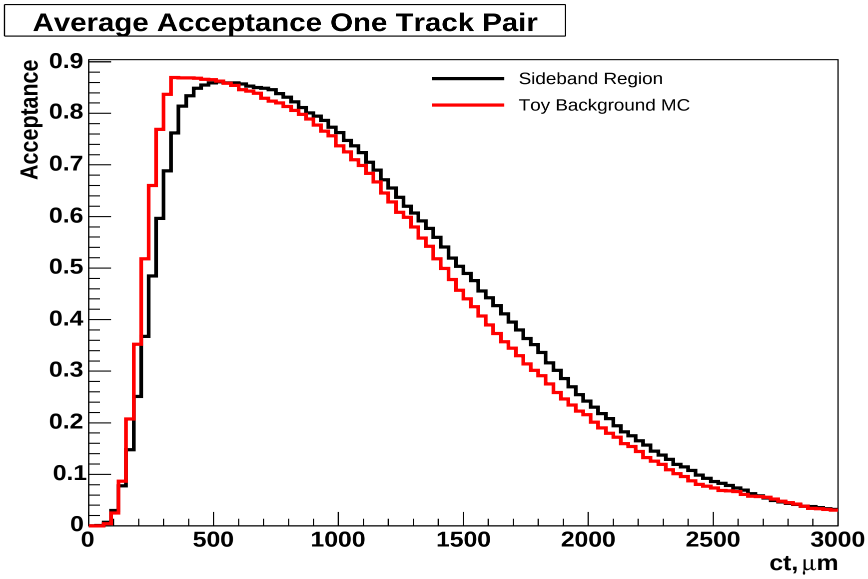


Figure 6.16: A comparison of the acceptance function for one track pair in simulated background and background in data. The red curve is from simulated data, the black one from data in upper sideband.

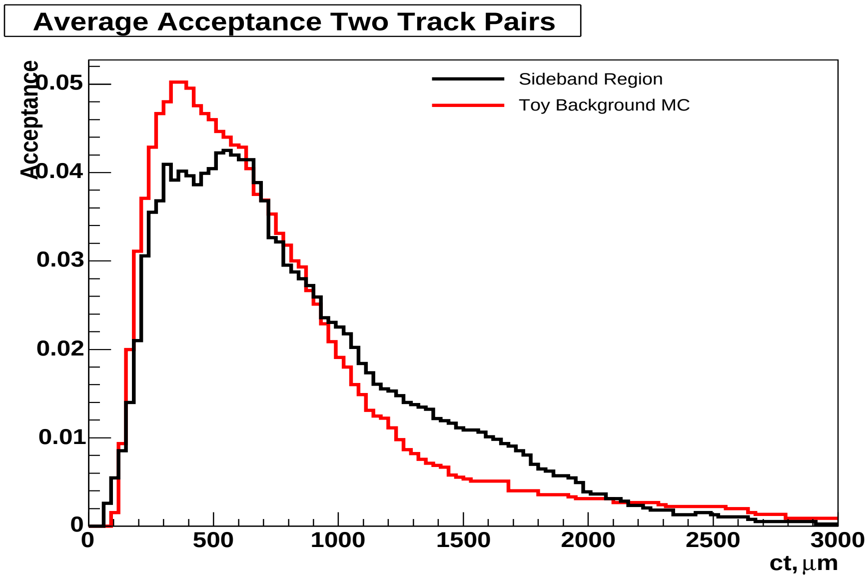


Figure 6.17: A comparison of the acceptance function for two track pairs in simulated background and data. The red curve is from simulated data, the black one from data in upper sideband.

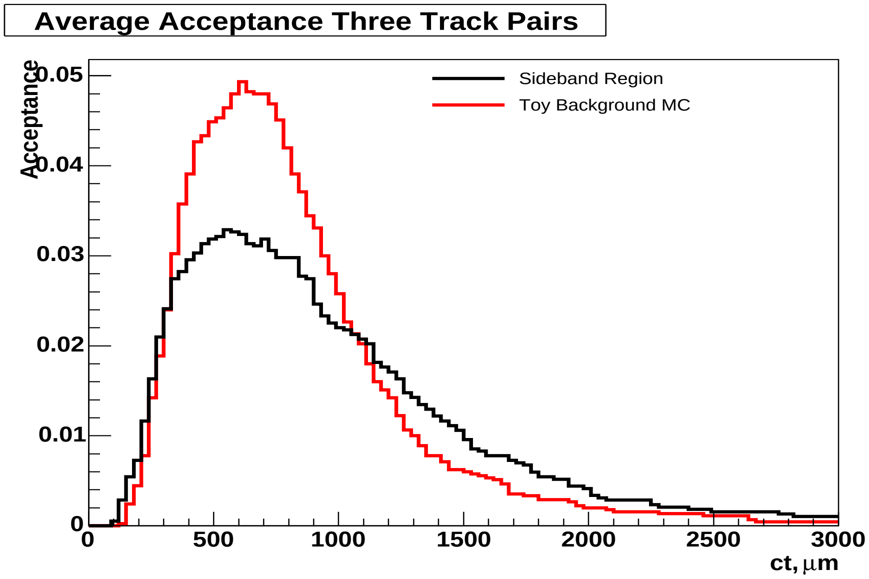
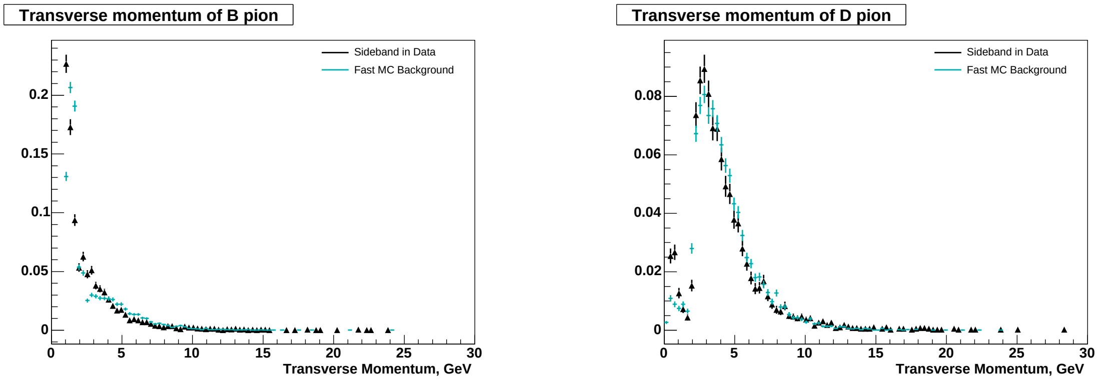
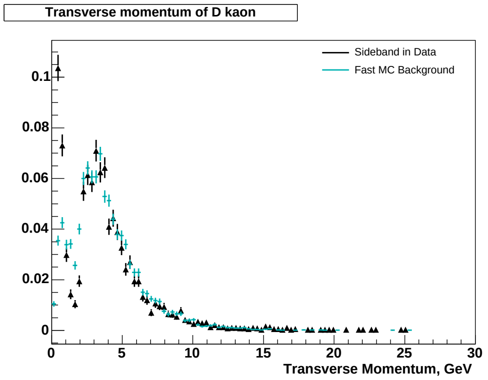


Figure 6.18: A comparison of the acceptance function for three track pairs in simulated background and data. The red curve is from simulated data, the black one from data in upper sideband.



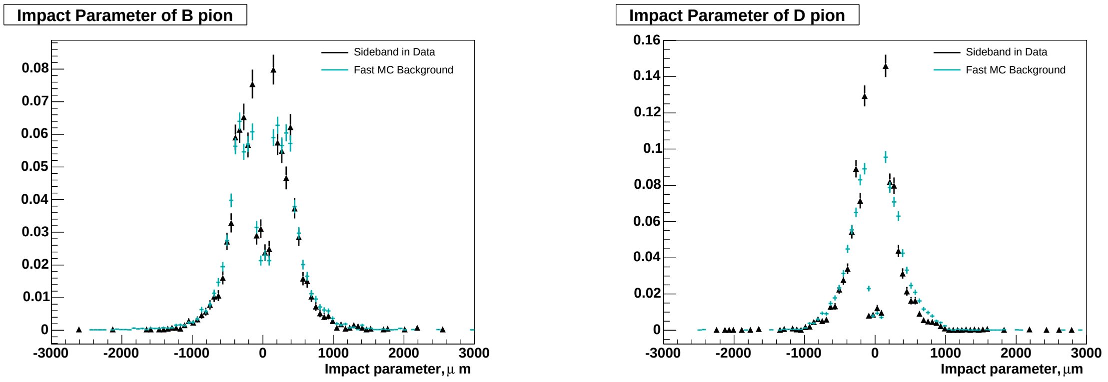
(a) The transverse momentum spectrum of the pion from the B

(b) The transverse momentum spectrum of the pion from the D



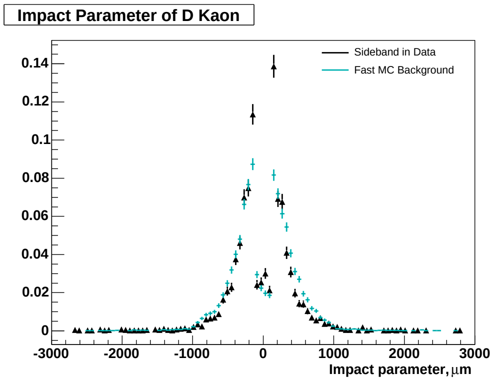
(c) The transverse momentum spectrum of the kaon from the D

Figure 6.19: A comparison of the sideband region in data and the transverse momentum of tracks derived from the fast simulation. The sculpting effect of the trigger accepting events with two tracks that have $P_T \geq 2GeV$ is visible. The blue distribution is that of simulated data, and the black distribution is from the upper sideband. The area of each distribution have been normalized to one.



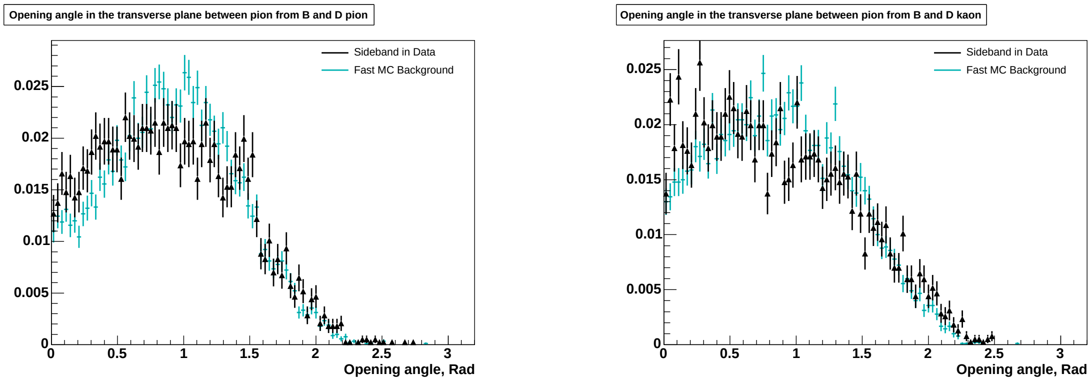
(a) The impact parameter distribution of the pion from the B

(b) The impact parameter distribution of the pion from the D



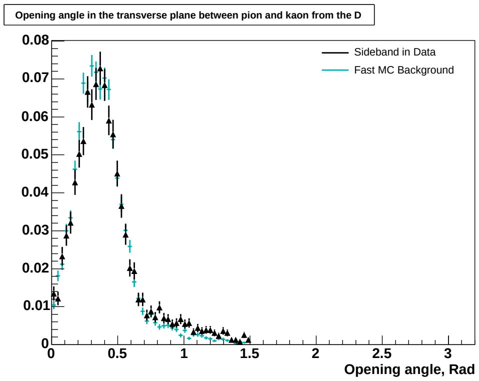
(c) The impact parameter distribution of the kaon from the D

Figure 6.20: A comparison of the sideband region in data and the fast simulation. These plots show the impact parameter distribution. The sculpting effect of the trigger is visible here too as only tracks with $d_0 \geq 120\mu m$ are used in the trigger decision.The blue distribution is that of simulated data and the black distribution is from the upper sideband. The area of each distribution has been normalized to one.



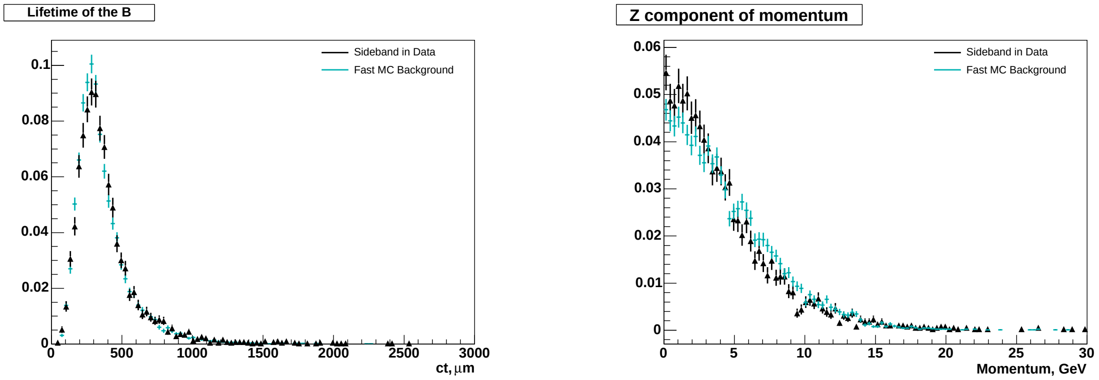
(a) The opening angle distribution between the pion from the B and the pion from the D

(b) The opening angle distribution between the pion from the B and the kaon from the D



(c) The opening angle distribution between the pion and kaon from the D

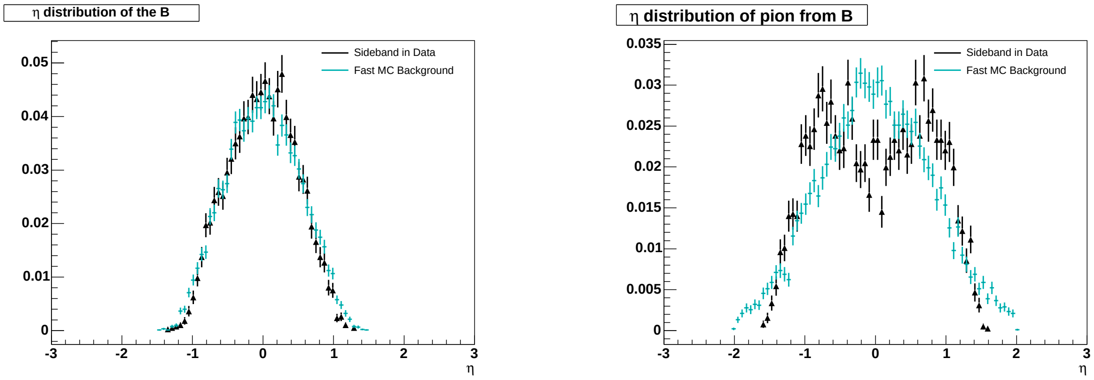
Figure 6.21: A comparison of the sideband region in data and the fast simulation. These plots show the opening angle in the transverse plane distribution.The blue distribution is that of simulated data and the black distribution is from the upper sideband. The area of each distribution has been normalized to one.



(a) The ct distribution of the B meson

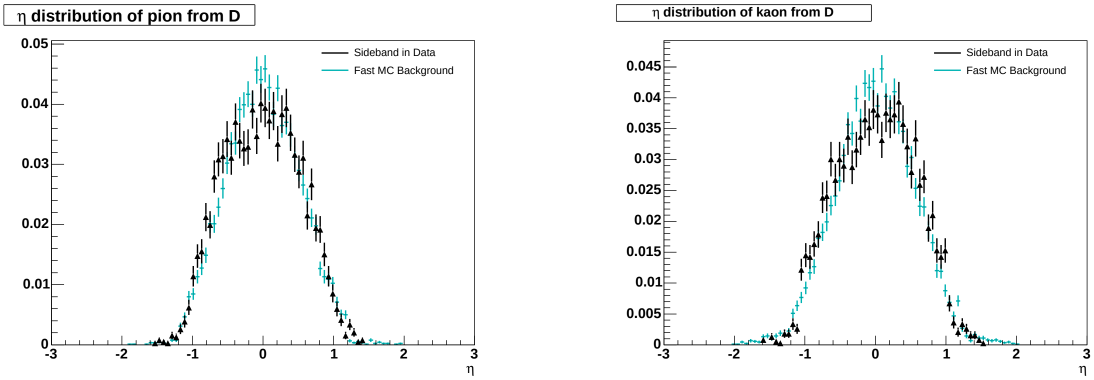
(b) The P_z momentum component distribution of the B meson

Figure 6.22: A comparison of the B meson z momentum and ct in the sideband region in data and the fast simulation. The blue distribution is that of simulated data, and the black distribution is from the upper sideband. The area of each distribution has been normalized to one.



(a) The η distribution of the B meson

(b) The η distribution of the pion from the B



(c) The η distribution of the pion from the D

(d) The η distribution of the kaon from the D

Figure 6.23: A comparison of the sideband region in data and the fast simulation. These plots show the η distributions of the tracks and that of the B meson itself. The analysis requirements cause a sculpting in the η of the π_B, this is not reproduced in the simulation. The blue distribution is that of simulated data and the black distribution is from the upper sideband. The area of each distribution has been normalized to one.

6.6 Validating the Method for a Mixture of Signal and Background

The fast simulation has been used to show, in section 6.4, that the biases induced by the impact parameter-based trigger can be corrected for by using information of the kinematics of the event only. In this section, the simulation free method is tested on samples containing a mixture of signal and background events. Thousands of simulated data samples are generated in a variety of configurations to provide a thorough and robust test of the PDF derived in chapter 5.

In chapter 5, the complications of adding background were discussed. In particular there is the difference in distributions of acceptance functions in signal and background which, if ignored, lead to additional biases. To characterize the acceptance function distributions, a Fisher discriminant analysis is applied that transforms the acceptance function to a scalar variable. The validation here includes testing whether this transformation can work, or whether too much information is lost in the reduction from the function to the scalar variable.

6.6.1 Generating the Mass Variable

Before the simulation can be used to validate the method, the mass resolution must be taken into account. The primary intention at this stage is to verify whether the full PDF discussed in chapter 5 will correctly extract the lifetime, or whether there remains a significant bias. As the Fisher scalar variable is calculated using information from the mass fit, for validation purposes, the simulation uses a very simple mass distribution from which to generate the event masses. This avoids the need to consider whether the mass fit could be the source of extra biases in the result. The mass distribution that is used is as follows. Background events are distributed uniformly between 5.23 and 5.5 GeV. Signal events are given a Gaussian mass distribution with mean 5.276 GeV and width 18 MeV. For the final tests, a different mass distribution is used that is closer to the distribution observed in data; this is discussed later.

6.6.2 Validation Method

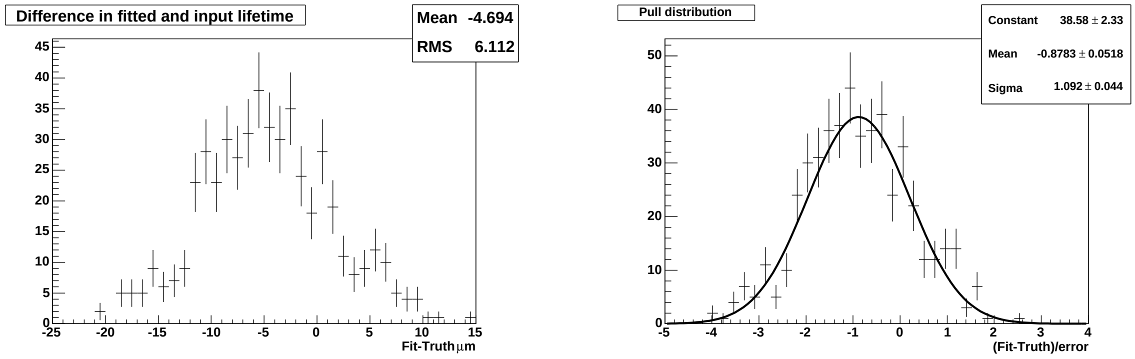
To validate the method, 500 independent samples of simulated data are created. Each sample contains 24000 signal events, similar to the yield observed in data. The simulated data also contains 9000 background events. This results in S/B = 5.0 between the mass range 5.23 and 5.37 GeV, which is comparable to 4.8 which is observed in data in figure 5.4.

The simulated data is fit as though it were data in all except one respect. In deriving the full PDF and the functions to describe it, two assumptions are made: firstly, that the factor $P(s|fish_i)$ is a good approximation for $P(s|A_\mathrm{i})$, and, secondly, that the Lagrange interpolating polynomials can fit the distribution of the Fisher scalar variable well. In validating this method, it is useful to consider each assumption separately so that any biases can be attributed to the correct cause.

For the simulated data, it is known exactly which events are signal and which are background, and therefore a histogram for each sample describing the change in signal fraction as a function of Fisher scalar can be constructed as in figure 5.13. For validation, this histogram is used to calculate $P(s|fish_i)$. Proceeding this way tests only the first assumption, which is the validity of approximating $P(s|A_\mathrm{i})$ by $P(s|fish_i)$. (The second assumption is tested later.)

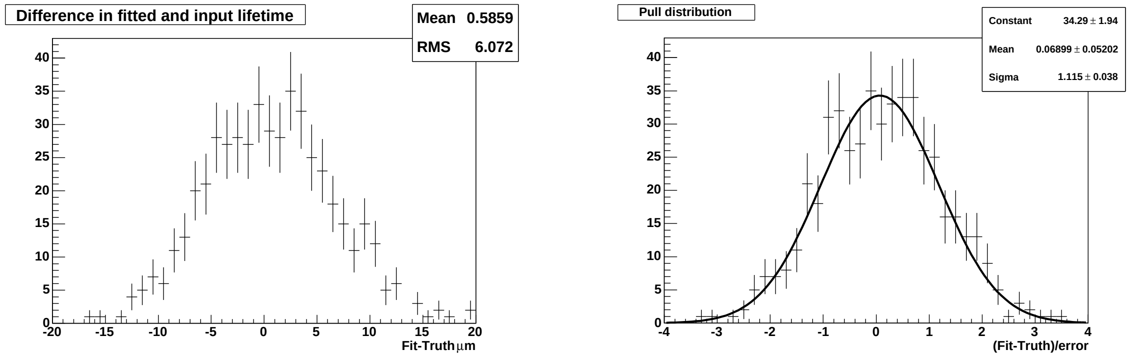
To test the method, two PDFs are used. The first ignores the difference between the signal and background acceptance function distributions. To do this, the factor $P(s|fish_i)$ is replaced by an overall signal fraction. The second PDF that is used is the full PDF described in chapter 5. For each sample a plot of the residual distribution is shown in figure 6.24. The residual for each sample is given by $c\tau_i - c\tau$ where $c\tau_i$ is the fitted lifetime and $c\tau$ is the input lifetime. If the difference between acceptance function distributions in signal and background are ignored, it is found that the lifetime is biased by approximately -4.7 microns. This is the size of the bias corrected for by including the distribution of acceptance functions in the PDF. The pull distribution, which is given by $\frac{c\tau_i - c\tau}{\sigma_{c\tau i}}$ where $\sigma_{c\tau i}$ is the statistical error on the lifetime for each fit, is fit to a Gaussian. The mean of this Gaussian is -0.88 ± 0.05 which is a deviation of 17σ. This demonstrates that a -4.7μm shift is not a mere statistical fluctuation. Figure 6.24 shows also the results when the same samples are fitted with the PDF that takes into account the acceptance function distributions. Here we find the residual mean is only 0.6 μm and that from the pull distribution is 0.06 $\pm$0.05, which is statistically consistent with 0. Therefore maximizing the PDF described in chapter 5 does extract the correct lifetime. This shows that the assumption that the Fisher scalar variable contains enough information to be able to correct any bias due to the difference in distribution of acceptance functions is valid.

It is noted here that the fit to the actual data can be performed neglecting the terms that model the signal fraction as a function of the Fisher scalar. A shift of approximately -7 μm is observed between the fit that neglects the Fisher scalar term and the term that includes it. The shift of -4.7 μm observed in simulated data is comparable to the -7 μm observed in data, and this shows that the simulated data is behaving in a similar way to the real data.



(a) The residual distribution when the acceptance function distribution is not taken into account

(b) The pull distribution when the acceptance function distribution is not taken into account



(c) The residual distribution using the full PDF

(d) The pull distribution using the full PDF

Figure 6.24: Toy studies that show the effect of ignoring the acceptance function distribution, and how by using Fisher discriminants to characterize the acceptance function, and taking into account the distribution of the Fisher scalar the biases are corrected for.

6.6.3 Further Cross Checks

As the distributions in the simulation do vary slightly from the data, further tests are carried out to ensure that the PDF to extract the lifetime is truly correct and that the results shown in figure 6.24 are not a fluke. To test the method further, the simulated data is generated in non standard configurations.

Altering the Signal Input Lifetime

The first test is to alter the input lifetime. As before, 500 samples of simulated data are generated with B $c\tau = 450\mu m$ and $c\tau = 540\mu m$. The mean of the residual distribution and mean of the fitted pull distribution are shown in Table 6.4. The method extracts the lifetime from these samples as well as for the standard simulation.

Input $c\tau$ (μm)	Residual mean	Pull mean
491.1	0.6	0.07±0.05
450	0.1	0.03±0.05
540	0.2	0.06±0.05

Table 6.4: Summary of validation tests using differing input lifetimes

Altering the Background Momentum Spectrum

Another test carried out is to alter the momentum spectrum of the background events. The generating function for background B momenta is altered so that the mean of the distribution varies by ± 1 GeV in comparison to the default simulation. The distributions of the momentum after the trigger in the 3 samples is shown in figure 6.25. The results of the 500 tests are given in Table 6.5. These show that the method continues to work with these different background kinematics. Although the mean of the fitted pull is 3σ from 0, the residual mean is similar to that of other tests. This is not investigated further in the work presented in this thesis.

These studies show that the method can handle the presence of background and any additional biases it produces. In particular the characterization of the acceptance function by the Fisher scalar results in the removal of the bias due to the difference in acceptance function distributions between signal and background.

Input $c\tau$ (μm)	Residual mean	Pull mean
Standard B momentum spectrum	0.6	0.07±0.05
Softer B momentum spectrum	0.7	0.18±0.06
Harder B momentum spectrum	0.0	-0.07±0.06

Table 6.5: Summary of validation tests using background generated with different momentum spectra

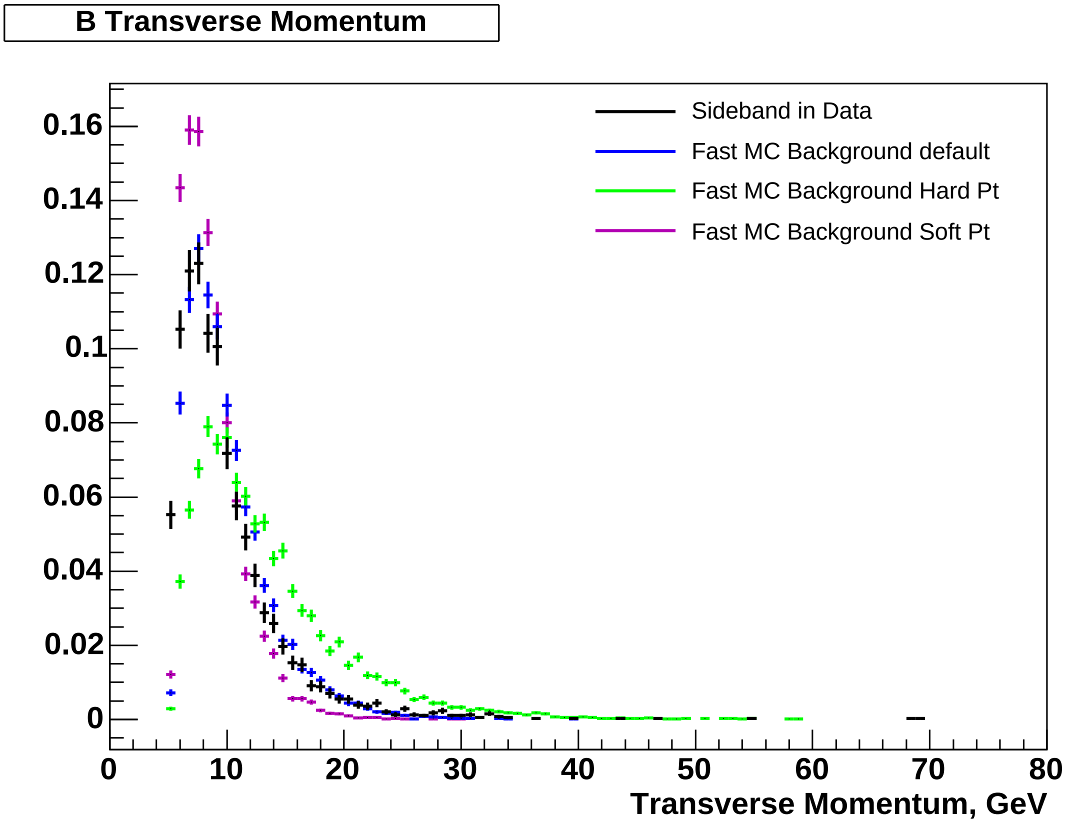
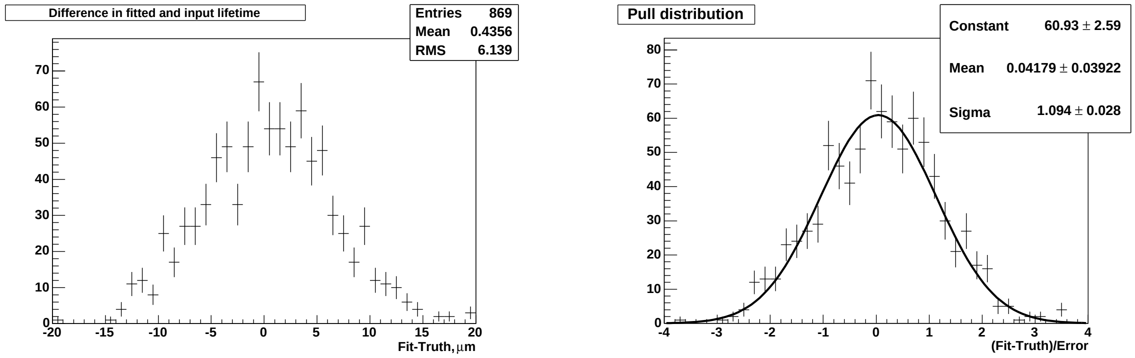


Figure 6.25: The difference in background momentum spectra used for validation tests.

6.6.4 Validating the Use of the Lagrange Interpolating Polynomials to Parameterize the Fisher Scalar Distribution

Sections 6.6.2 and 6.6.3 demonstrate that the method is robust and can model the difference between the distributions of acceptance function in signal and background sufficiently well using the Fisher scalar variable. This section tests the second assumption, that the Lagrange interpolating polynomials can provide a good model for the signal fraction as a function of Fisher scalar.

To test this assumption, 870 samples are generated and fit. This time, at no point is information regarding the true signal or background nature of any event used. The Lagrange interpolating polynomials are used to fit the signal fraction as a function of Fisher scalar. The residual and pull distributions of the tests are shown in figure 6.26. The mean of the pull distribution is 0.04±0.04, which is statistically consistent with all the previous tests where truth information was used to determine the signal fraction as a function of Fisher scalar. This shows that the Lagrange interpolating polynomials are able to parameterize the signal fraction as a function of Fisher scalar well enough to not affect the lifetime.



(a) The residual distribution for fits where the Lagrange interpolating polynomials have been used to fit the signal fraction as a function of the Fisher scalar

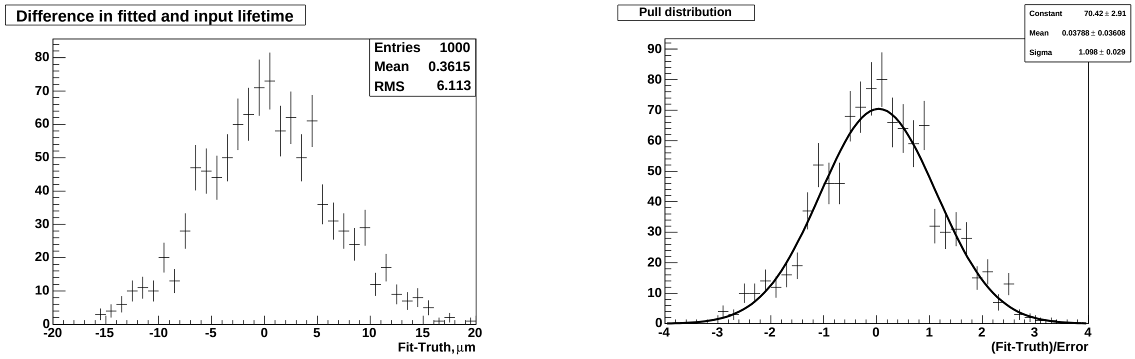
(b) The pull distribution for fits where the Lagrange interpolating polynomials have been used to fit the signal fraction as a function of the Fisher scalar.

Figure 6.26: Simulation studies show that the Lagrange interpolating polynomials give a good fit to the signal fraction as a function of Fisher scalar. This is demonstrated by the pull distributions for best fit lifetime, as there is no significant bias.

6.6.5 Generating Fast Simulation with a more Realistic Mass Distribution

A further set of 1000 samples are generated. For this set of samples, the masses of each event are sampled according to the mass model used to fit data, where the parameters of the model come from the best fit parameters from the mass fit on data. The mass model is given in equations 5.23 and 5.25.

One further change is also introduced into the fit procedure at this point and this is the procedure that is also used in the data fit. An initial fit to the mass related variables alone has to be performed to allow calculation of the Fisher scalar for each event. For the subsequent fit using the full PDF, the mass parameters relating to the signal gaussians and background slope are fixed to their values from the previous mass-only fit. This is found to improve the stability of the fit. The Lagrange interpolating polynomials are used to parameterize the signal fraction as a function of Fisher scalar distribution. The residual and pull distributions of the tests are shown in figure 6.27



(a) The residual distribution for fits where a more realistic mass model has been used in the generation

(b) The lifetime pull distribution for fits where a more realistic mass model has been used in the generation

Figure 6.27: Simulation studies show that the presence of a more complicated mass model does not result in any additional biases

The pair of plots in figure 6.27 demonstrate that no additional bias comes from using a more complex mass model as the mean is 0.04±0.04. There is a residual bias of 0.4 microns as shown in the residual histogram in figure 6.27. To be conservative, this is taken as a systematic error of the fitting procedure. The pull distribution has a width of 1.1±0.3 which is 3σ discrepant from 1. There are indications from the fits to signal events that this wide pull is present before background is added as seen in figures 6.10 6.13.

This suggests that the statistical error returned by the fit could be underestimated by 10%. It is assumed that the result in data will be similarly affected, and as this effect is small, the statistical error returned by the fit to data is scaled by a factor of 1.1 to compensate for this effect.

In conclusion, this chapter has detailed a fast simulation that matches the kinematic distributions of signal and background. This fast simulation has been used to generate multiple samples. These samples were used to test the likelihood function, and have demonstrated that the method described in chapters 4 and 5 is robust and valid. Variations of the standard generation were also applied, and the fitting procedures continued to extract the correct lifetime. It is concluded therefore that a simulation-free technique can be used to correct for biases induced by impact parameter-based triggers.

Chapter 7

The $B^{\pm}$ Lifetime

This chapter describes the measurement of the $B^{\pm}$ lifetime using the data.

7.1 The Lifetime Fit on Data

The data has been reconstructed and selected according to the procedures and selection criteria described in chapter 3. In chapters 4 and 5, the PDF for extracting the best fit lifetime and correcting for the selection biases has been derived. The simulation studies in chapter 6 have shown that this PDF is valid and that the models used to fit the variable distributions discussed in chapter 5 do not cause additional biases. The lifetime fit can now be performed on the data.

7.1.1 The Mass Distribution Fit

The first stage is to fit the the mass distribution on its own. The model used to fit the mass distribution has been described in section 5.4. There are 7 free parameters in the mass fit: the mean and width of both signal Gaussians, the relative fraction of each, the background slope and the overall signal fraction. The full fit results for each parameter are given in table 7.1. The fit projection is given in figure 7.1. There are correlations between the best fit parameter results as expected due to the nature of the fitting function. The yield of signal events is 24,200±200.

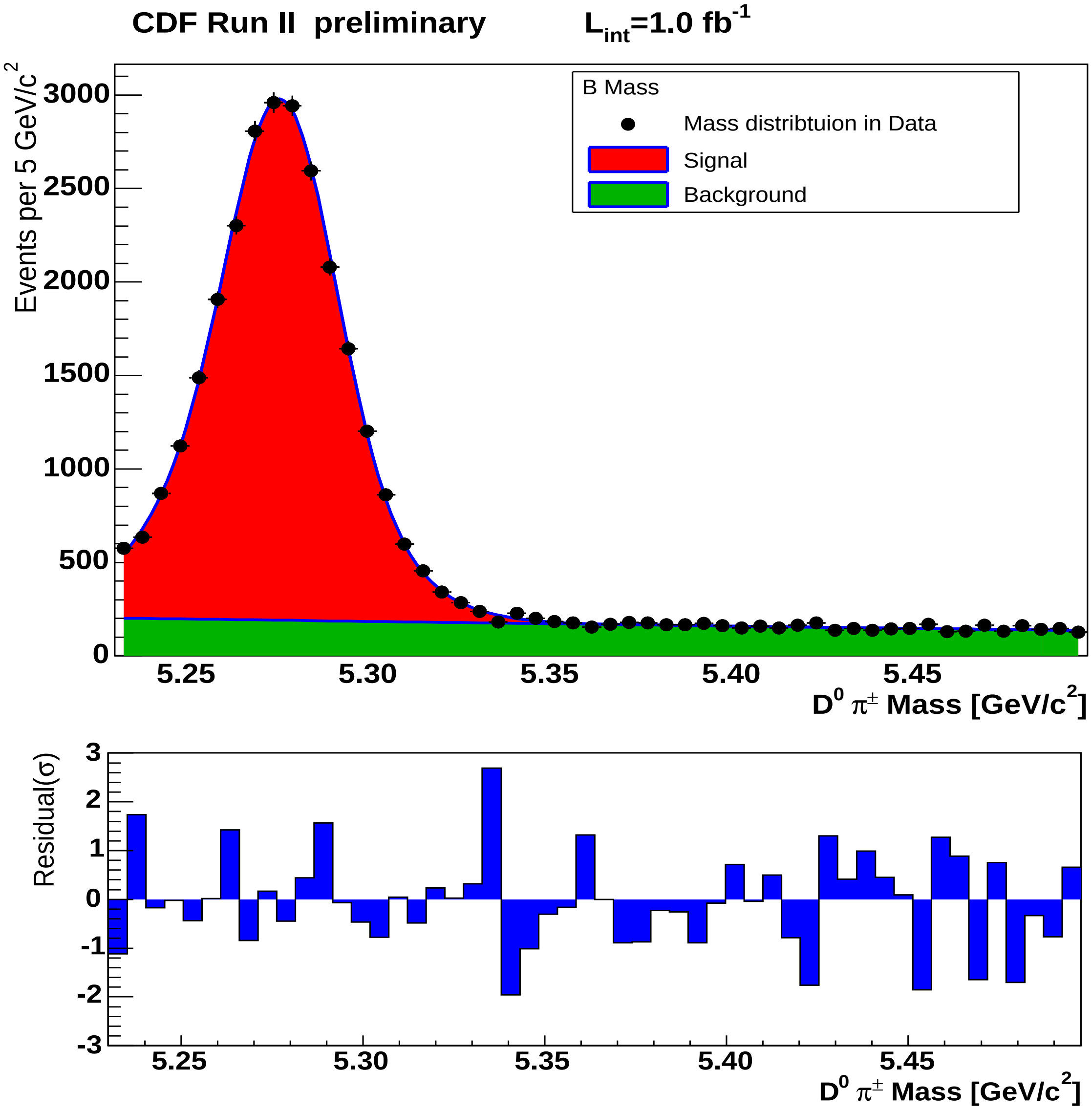


Figure 7.1: The mass fit projection on the data

Parameter	Best Fit	Error
m_1	5.2761	±0.0004
m_2	5.2699	±0.0025
σ_1	0.0264	±0.0033
σ_2	0.0144	±0.0010
f_1	0.439	±0.13
f_s	0.730	±0.0050
α	-0.1657	±0.0036

Table 7.1: Summary of best fit mass parameters

7.1.2 Calculation of the Fisher Scalar Variable

The Fisher scalar distribution is then calculated using the information from the mass fit as discussed in section 5.6. The resulting Fisher scalar distribution is shown in figure 7.2.

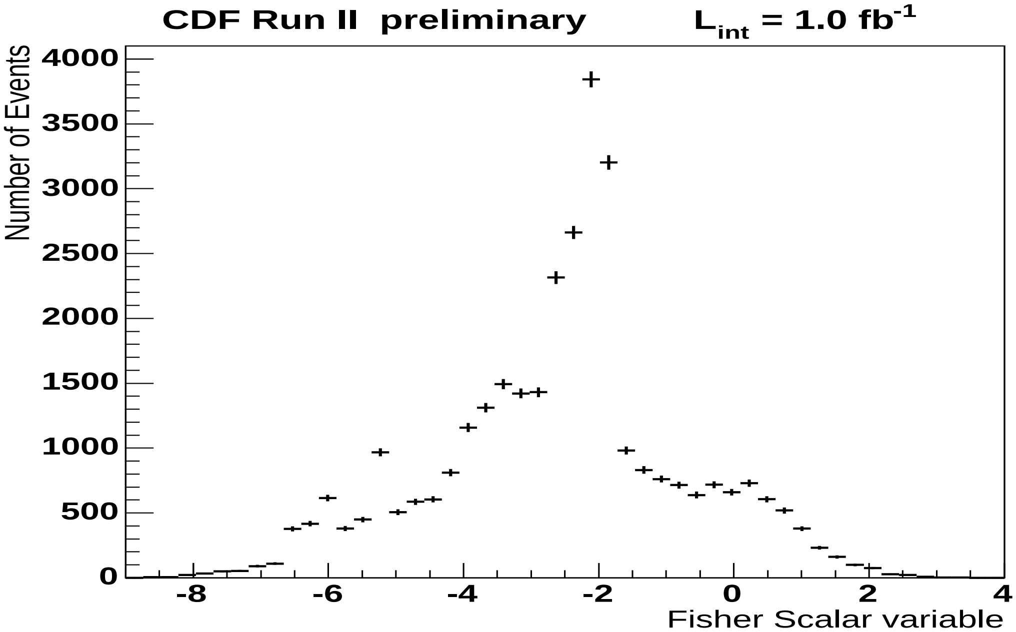


Figure 7.2: The Fisher scalar distribution

7.1.3 Lifetime Fit

Finally, the full fit is performed and the likelihood derived in section 5.2 is maximized to find the best fit lifetime. The parameters of the mass fit are fixed to the values determined in the previous mass only fit. The signal fraction is re-parameterized by Lagrange interpolating polynomial. This is determined by the fit to the Fisher scalar variable and gives a signal fraction per event based on the event's Fisher scalar. In total there are 30 free parameters in the fit. These are the best fit lifetime itself and 3 parameters to determine the signal single track finding efficiency in 3 different data taking periods, 3 parameters to determine the background single track finding efficiency in each of the 3 data taking periods. There are also 10 parameters to describe the background lifetime distribution and 13 parameters to describe the signal fraction as a function of the Fisher scalar variable.

The lifetime fit is run simultaneously with the fit to the Fisher scalar variable. The best measured lifetime is

$$c\tau = 498.2 \pm 6.8\mu m \tag{7.1}$$

$$\tau = 1.662 \pm 0.023ps \tag{7.2}$$

where the quoted error is statistical only and has been scaled by the factor 1.1 as determined in section 6.6.5. The result is consistent with the current world average of 491.1 ± 3.0 μm [11]. The projection of the fit on the lifetime variable is shown in figure 7.3. The Lagrange interpolating polynomial for the fit is shown in figure 7.4. To assess the validity of this function, events are binned in regions of the Fisher scalar and then a mass fit is performed on each bin. The signal fraction resulting from these fits is overlaid on the plot of the Lagrange interpolating polynomial. As there is good agreement between the signal fraction determined from a series of mass fits and the

Lagrange interpolating polynomial, there is confidence that the fit has found a good function to represent the signal fraction as a function of the Fisher scalar.

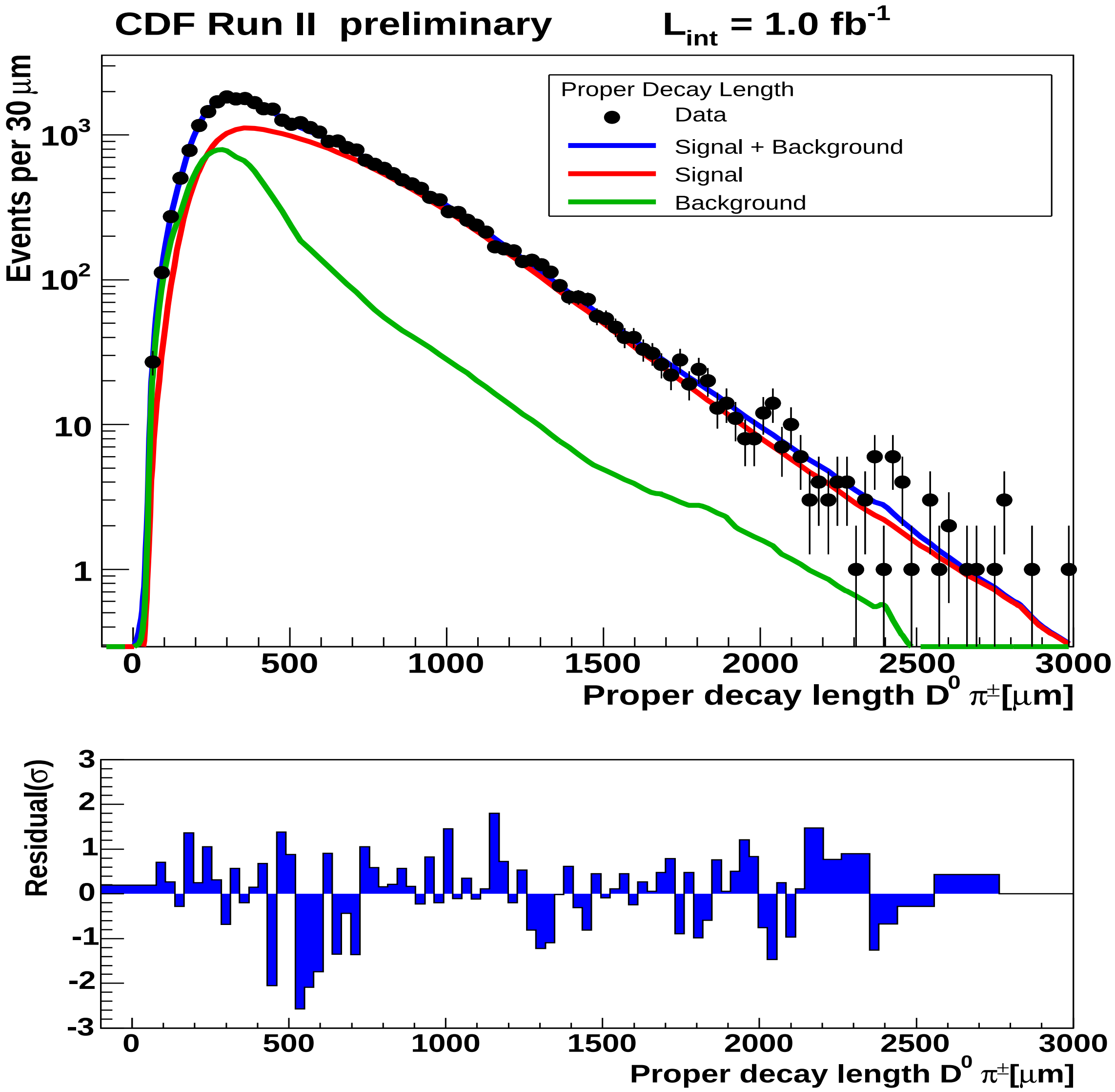


Figure 7.3: The projection of the fit to the lifetime variable

The fit results for the other parameters are given in the tables 7.2 - 7.4, where the best fit results have been split into those for efficiency, those describing the background lifetime and those describing the signal fraction as a function of Fisher scalar for clarity.

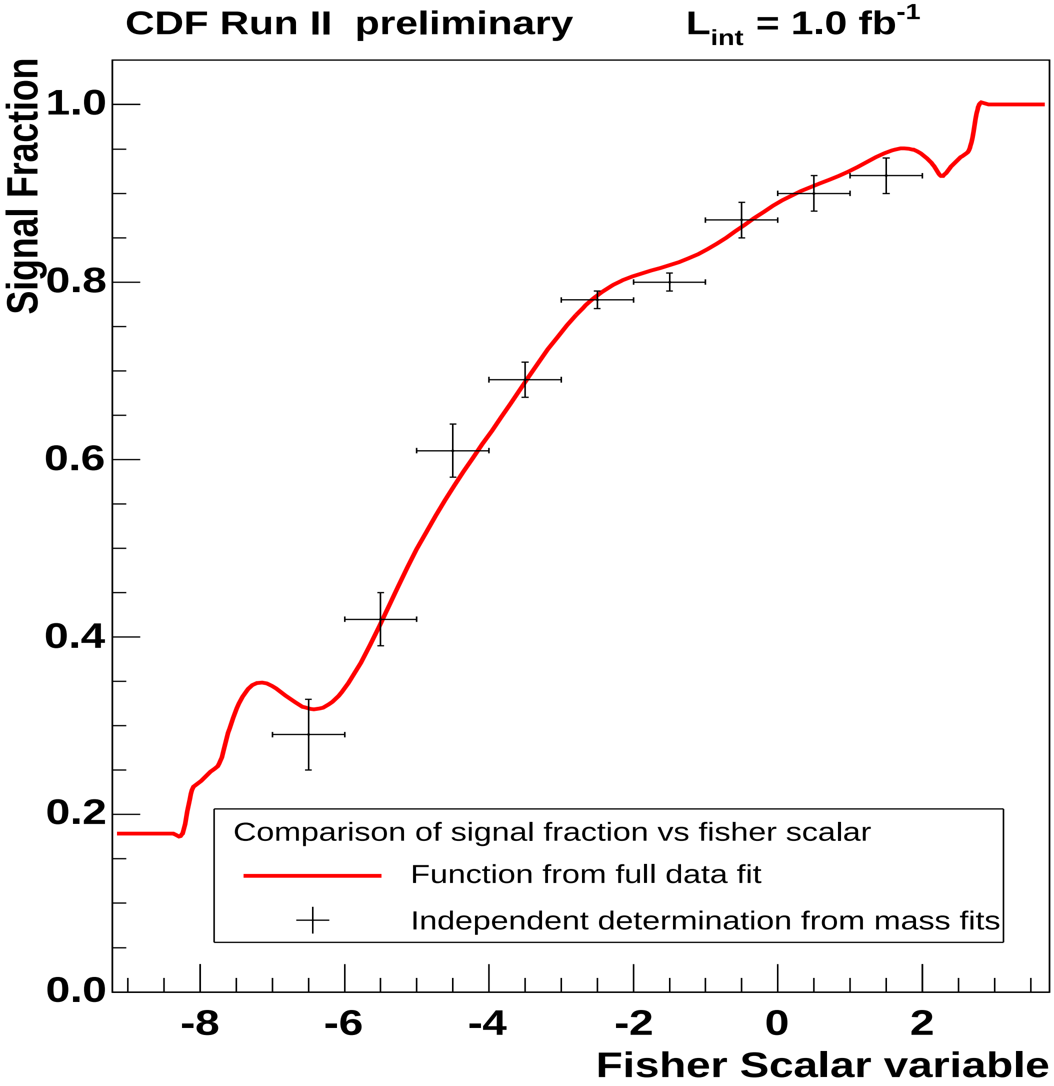


Figure 7.4: The Lagrange interpolating polynomial found by the fit to best describe the signal fraction as a function of Fisher scalar. Overlaid is the signal fraction determined from mass fits to events within regions of Fisher scalar.

Efficiency Parameter	Best Fit	Error
Signal Period 1	0.488	±0.033
Signal Period 2	0.656	±0.008
Signal Period 3-5	0.725	±0.006
Background Period 1	0.508	±0.066
Background Period 2	0.522	±0.017
Background Period 3-5	0.560	±0.017

Table 7.2: Summary of best fit efficiency parameters. The periods are defined in section 6.4.1

Background ct Parameter	ct_j (μm)	Best Fit	Error
a_1	0	10.50	±0.35
a_2	146.9	7.08	±0.06
a_3	322.6	4.77	±0.04
a_4	532.7	2.72	±0.04
a_5	783.9	1.28	±0.07
a_6	1084.3	1.28	±0.10
a_7	1443.5	-1.19	±0.17
a_8	1873.1	-1.94	±0.29
a_9	2386.7	-2.78	±0.47
a_{10}	3000	-7.16	±2.63

Table 7.3: Summary of best fit background lifetime parameters. The ct_j values represent the points where the background lifetime distribution is sampled as defined by t_j in equation 5.27

To fully determine the lifetime, the error due to systematic uncertainties must also be considered. These are discussed in the next chapter.

Fisher Parameter	Fisher Scalar	Best Fit Signal fraction	Error
Parameter 1	-8.35	0.179	±0.076
Parameter 2	-7.05	0.259	±0.063
Parameter 3	-6.19	0.328	±0.023
Parameter 4	-5.32	0.379	±0.013
Parameter 5	-4.45	0.539	±0.011
Parameter 6	-3.59	0.660	±0.008
Parameter 7	-2.73	0.768	±0.006
Parameter 8	-1.86	0.813	±0.005
Parameter 9	-1.00	0.849	±0.008
Parameter 10	-0.13	0.900	±0.007
Parameter 11	0.73	0.937	±0.010
Parameter 12	1.60	0.919	±0.036
Parameter 13	2.89	1.00	±0.048

Table 7.4: Summary of best fit signal fraction as a function of Fisher scalar. The values of Fisher scalar give the mid point of each bin used by the Lagrange interpolating polynomial function as described in section 5.6.5

Chapter 8

Systematic Uncertainties

This chapter provides a description of the sources and sizes of the systematic uncertainty of this measurement. The sources in order of size are as follows: dependence of the SVT single track finding efficiency on impact parameter and transverse momentum, a correlation between invariant mass and lifetime in background, the parameterization of the background lifetime, a possible misalignment of the silicon detector, a possible bias in the fitter, the resolution of the lifetime measurement and the dependence of the SVT single track finding efficiency of η.

8.1 Assessing the Size of Systematic Errors

The derivation of the likelihood in chapters 4 and 5 uses certain assumptions about the performance of various detector components. While these assumptions are realistic, plausible variations that may exist in data, and their effect on the measured lifetime, are considered in this chapter. In addition to the effect of assumptions on detector performance, other sources of systematic error such as the choice of a particular parameterization are also studied.

The fast simulation described in chapter 6 is the main tool that is used to understand the size of systematic uncertainty. The fast simulation can be used to generate large numbers of samples which can then be fit using the derived likelihood. Small variations from the assumptions on features in the data or detector performance can be incorporated into the fast simulation generation. The samples can then be fit using the derived likelihood. The size of the difference between the best fit lifetime and the input lifetime provides an estimate for the size of the systematic uncertainty. For each variation, 1000 samples, each containing a number of events representative of data, are fit. The estimate of the uncertainty, $\delta_{c\tau}$, due to ignoring a feature in data is then the mean of the residual distribution: $\delta_{c\tau} = \overline{c\tau_i} - c\tau_m$, where $c\tau_i$ is the best fit lifetime in the ith generated sample and τ_m is the input mean lifetime.

To understand when the value of $\delta_{c\tau}$ is significant the pull distribution is used. It is defined as $\frac{\overline{c\tau_i - c\tau_m}}{\sigma_{c\tau_i}}$ where $c\tau_i$ is the best fit lifetime in the ith generated sample (1000 samples are generated), $\sigma_{c\tau_i}$ is the uncertainty of the ith fit, and $c\tau_m$ is the input (generated) lifetime. The pull distribution is fit with a Gaussian. This fit yields the best fit mean and uncertainty of the pull: $\overline{a} \pm \sigma_a$. The estimate of the uncertainty of the residual is then $\sigma_a \times \overline{\sigma_{c\tau}}$,where $\overline{\sigma_{c\tau}}$ is the average of the best fit errors of the lifetimes over all samples. The value of σ_a for 1000 samples is approximately 0.035 and the value of $\overline{\sigma_{c\tau}}$ is approximately 6 μm, where there are 24,000 signal events per sample. This results in an error on the mean residual, $\delta_{c\tau}$, of approximately 0.2 μm. For sources of systematic uncertainty that result in a mean residual smaller than 0.2 μm, the systematic error is deemed as negligible. This is also a sensible cut off in the context of the size of the statistical and leading systematic errors, which are approximately 6 and 3 μm, respectively.

The systematic errors are approached in turn in this chapter. They are divided into two groups; the first are those which affect only the signal, i.e, those that would be present even if there was a pure sample of signal events collected by the CDF detector. The second group are systematic errors that arise due to the background events that are present and how well the likelihood models them.

8.1.1 Systematic Error Due to Approximations made on the Single Track Finding Efficiency

One of the most important assumptions of the simulation free method is that the single track finding efficiency is constant for tracks with impact parameters between 120 and 1000 μm, which are the upper and lower trigger cuts. This is a good assumption as can be seen in figure 8.1; the efficiency only has a small deviation from flat in the region of interest (to the left of the dotted line).

If the SVT single track finding efficiency is not constant over the relevant range of impact parameter, then it can preferentially accept events with high or low impact parameter. As tracks with high impact parameter tend to be correlated with longer lifetimes (and vice-versa), a non-constant SVT single track finding efficiency can bias the lifetime.

The method also assumes that the single track finding efficiency is constant as a function of P_T and η. This is also an important assumption to consider as the height of the acceptance function is dependent on the value of the efficiency. If there are variations in the efficiency as a function of P_T or η, then the height calculated will be different from the true value and this could result in a further source of systematic uncertainty. The following sections consider how neglecting a small drop in efficiency for larger impact parameters, and the form of the efficiency with respect to P_T or η

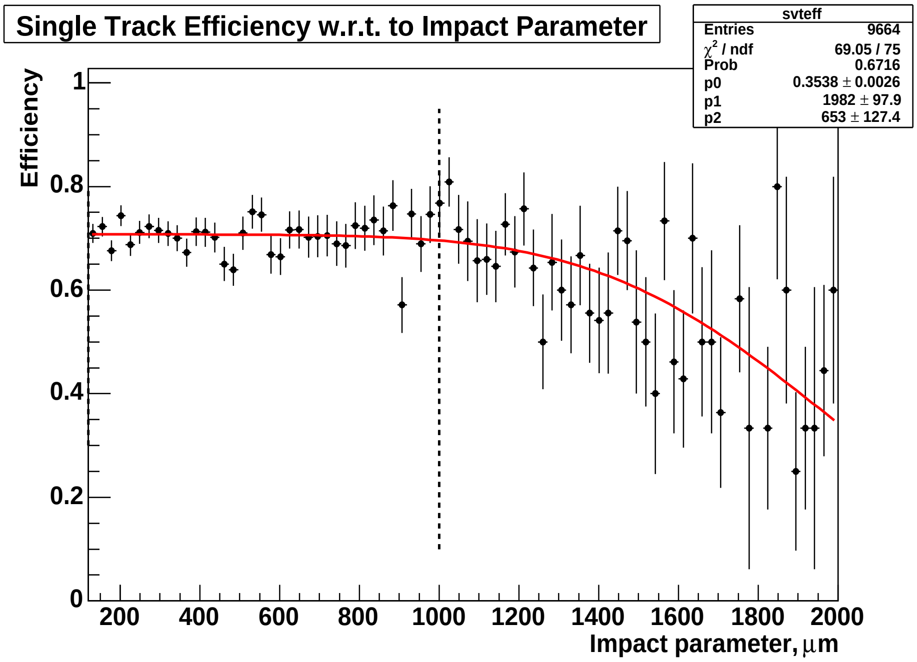


Figure 8.1: The single track finding efficiency as a function of the track impact parameter. The points come from data and the fitted curve is the result of the fit to the function described in equation 8.2.

observed in data, will affect the measured lifetime.

Variations in the Single Track Finding Efficiency as a Function of Impact Parameter

The realistic Monte Carlo can not be used to determine the single track finding efficiency, as the samples available do not cover the data taking period and there is uncertainty on how well the realistic Monte Carlo emulates the detector. Changes in the SVT finding efficiency during the course of the data taking period must also be considered, and so the realistic Monte Carlo will not be a good representation of the data. To determine the single track finding efficiency from the data sample, it is necessary to obtain an unbiased sample. The method by which this is obtained is explained in section 4.4.1 and results in 3 unbiased samples.

Using these unbiased samples of tracks from the data, the single track finding efficiency can be found by dividing the number tracks found by the SVT, N^{SVT}, by the number of tracks found by the offline algorithm , N^{OFF}, as a function of the offline impact parameter d_0.

Offline tracks for this study of the SVT single track finding efficiency are selected according to the following criteria:

i . The number of Silicon hits(found offline) in $r-\phi$, $N_{r-\phi}$, is ≥ 3

ii . The track transverse momentum has $P_T \geq 2$ GeV

iii . The track $\mid \eta \mid$ is < 1.1

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

$$\epsilon_{SVT}(d_0) = \frac{N^{SVT}(N_{r-\phi} \geq 3, P_T \geq 2GeV, \mid \eta \mid < 1.1)}{N^{OFF}(N_{r-\phi} \geq 3, P_T \geq 2GeV, \mid \eta \mid < 1.1)} \tag{8.1}$$

and is plotted in figure 8.1. The efficiency is fit to the following function:

$$\epsilon_{SVT} = p_0 \times erfc(\frac{d_0 - p_1}{p_2}) \tag{8.2}$$

where $erfc$ is the complementary error function, d_0 is the offline impact parameter, and p_0, p_1 and p_2 are free parameters. The complementary error function is used, as it has the correct functional form; it starts flat and then turns down towards zero, which is the pattern expected of the SVT performance.

Although the approximation of constant efficiency is good, figure 8.1 shows a small drop in efficiency as the impact parameter approaches the 1000 μm trigger cut off. To test the effect of this kind of variation in efficiency, simulated data is generated containing this effect. To do this, the parameters of the fit function are used to define the probability of finding a track based on its impact parameter. For a given impact parameter, the probability of finding the track is calculated from the best fit function and then the track is designated as found or not found accordingly.

1000 samples of 24,000 are generated, and the lifetime of each sample is fitted using the standard likelihood, i.e, ignoring any deviations from a flat single track finding efficiency. The difference between the fitted and input lifetimes is shown in figure 8.2. This plot shows a mean residual of -1.7 μm. Having demonstrated that using the real shape of the SVT single track finding efficiency dependence on impact parameter changes the best fit lifetime by 1.7 μm, it is prudent to further investigate this source of systematic uncertainty by varying $\varepsilon_s(d_0)$ further.

The fitted function describing the variation in single track finding efficiency is described by 3 fitted parameters, each of which has an uncertainty derived from the fit. The best fit parameters are varied within one σ of the fit-derived uncertainty, taking into account the correlation matrix. Varying the parameters like this results in six sets of values for p_0, p_1, p_2. Of the six possible input values for the function in equation 8.2, there are two that have a greater change in efficiency between 120 μm and 1000 μm than the function already used. These are shown as the blue and green curves in figure 8.3, where the red curve is the result of the fit with no further variations to the parameters. The values of the parameters are given in table 8.1.

1000 samples of events are generated assuming the efficiency is represented by the green curve, and 1000 further samples are generated assuming the efficiency is

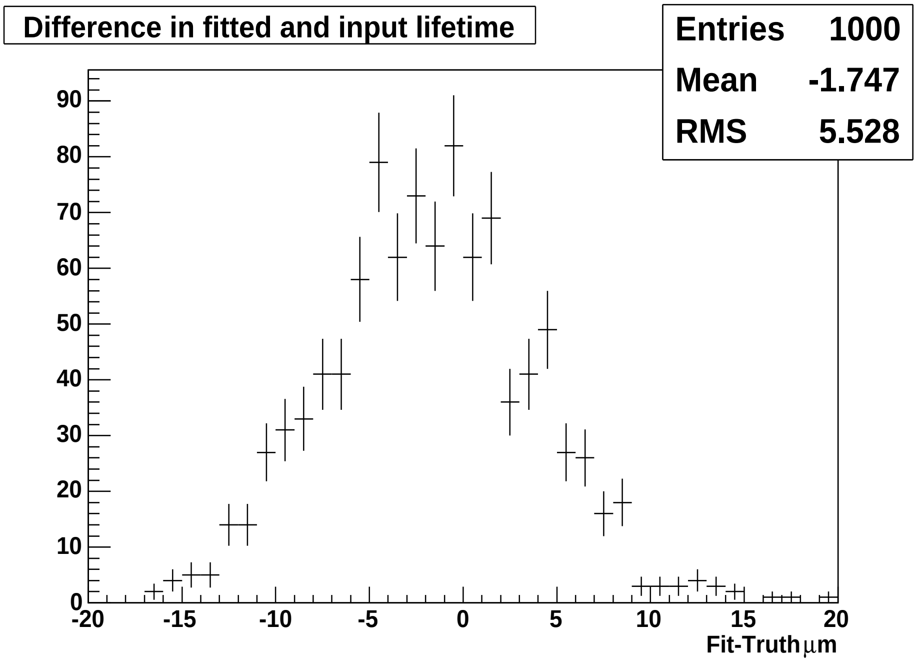


Figure 8.2: The difference between the fitted and input lifetimes for 1000 samples of fast simulation events.

Parameter	Best Fit Parameters Red Curve	Varied Parameters Blue Curve	Varied Parameters Green Curve
p_0	0.3538 ± 0.0026	0.3520	0.3532
p_1	1982 ± 98	2071	1942
p_2	653 ± 127	777	682

Table 8.1: Summary of parameters used to generate non-flat single track finding efficiencies. The curves are defined by equation 8.2.

represented by the blue curve. The lifetime of these samples is fitted using the standard likelihood, which assumes that the efficiency is constant. If the efficiency function has the parameters in blue from the table 8.1, depicted by the blue curve in figure 8.3, then the average shift in the residual is largest for all configurations shown. The results of each fit are shown in figure 8.4. This shift is -3.1 μm and is the most conservative choice for the systematic uncertainty arising from the deviation of $\varepsilon_s(d_0)$ from a constant value.

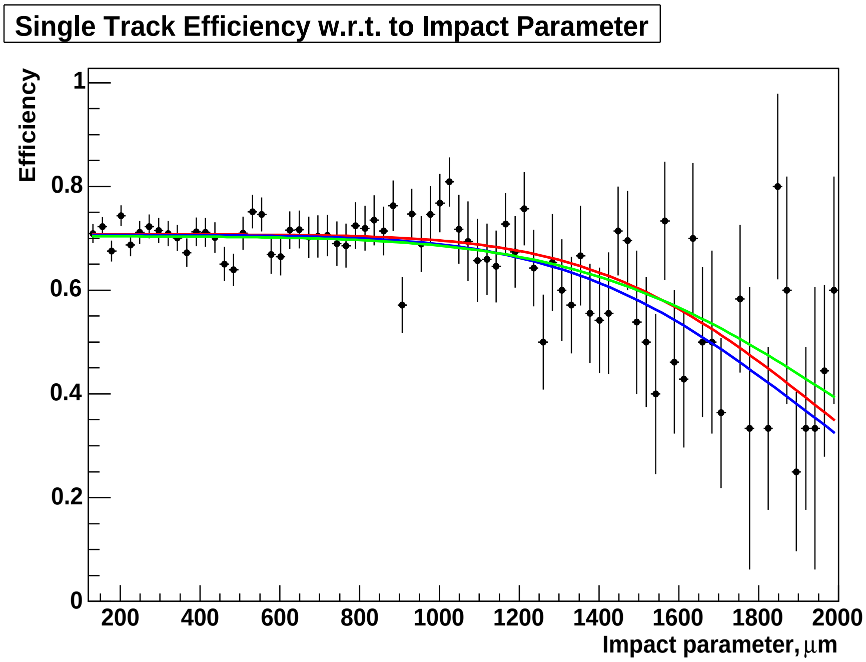


Figure 8.3: The single track finding efficiency as a function of the track impact parameter in data. 3 possible parameterizations of the efficiency are shown.

Systematic Error Due to the Variation in Single Track Finding Efficiency as a Function of Track P_T

In addition to assuming that $\epsilon_{SVT}(d_0)$ is constant it is also assumed that $\varepsilon_s(P_T, \eta)$ is constant, *i.e.*, the single track finding efficiency is independent of track P_T and η. The efficiency is drawn as a function of P_T in figure 8.5 using the unbiased samples found as described in section 4.4.1. It is fitted using a 3rd order polynomial. 1000 sets of 24000 events each are generated with this efficiency and are fitted using the standard likelihood. A mean shift of 1.8 μm is observed in the difference between fitted and input lifetime, and this value is taken as the systematic error due to ignoring variations in the single track finding efficiency as a function of transverse momentum.

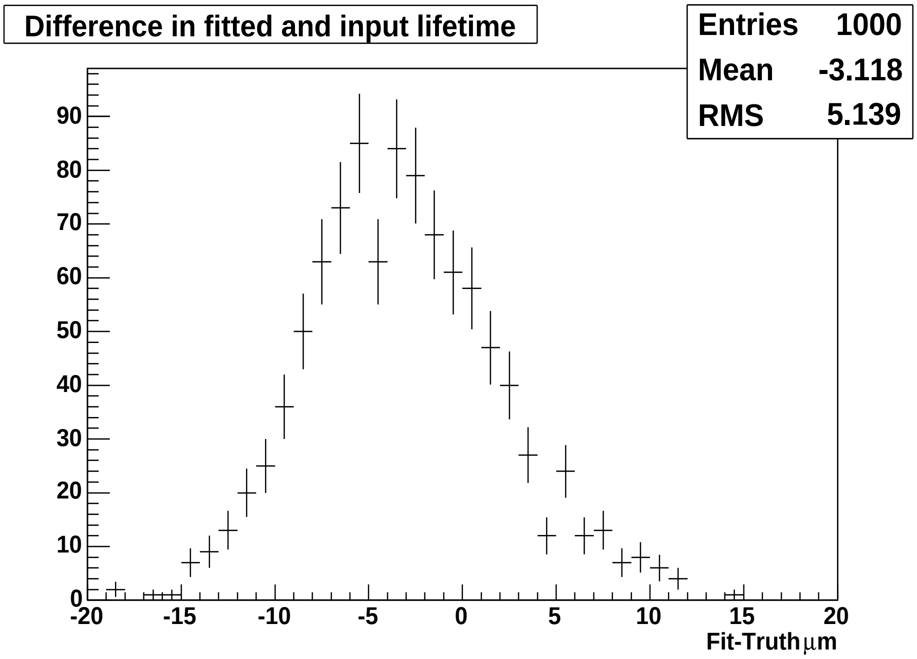


Figure 8.4: The difference between the fitted and input lifetimes for 1000 samples of fast simulation events where the single track finding efficiency is based on the blue curve in figure 8.3.

Systematic Error Due to the Variation in Single Track Finding Efficiency as a Function of Track η

Finally, the efficiency as a function of track η is also considered. The efficiency is plotted in figure 8.6 using the unbiased data samples described in section 4.4.1. As the η of a track bears no relation to the lifetime of the event, it is not expected that the efficiency as a function of η, despite its sharp drop, will cause a large systematic error. Also, as the majority of tracks have an $|\eta| < 0.8$ the drop observed after this will only affect a small subsample of events. Therefore, for simplicity the η dependent efficiency is modeled as constant between $0 < |\eta| < 0.8$ followed by a linear drop to 0 at $|\eta|$=1.1. Using this model, 1000 samples of simulated data are generated, and the lifetime of each is measured. The mean difference between the fitted and input lifetime

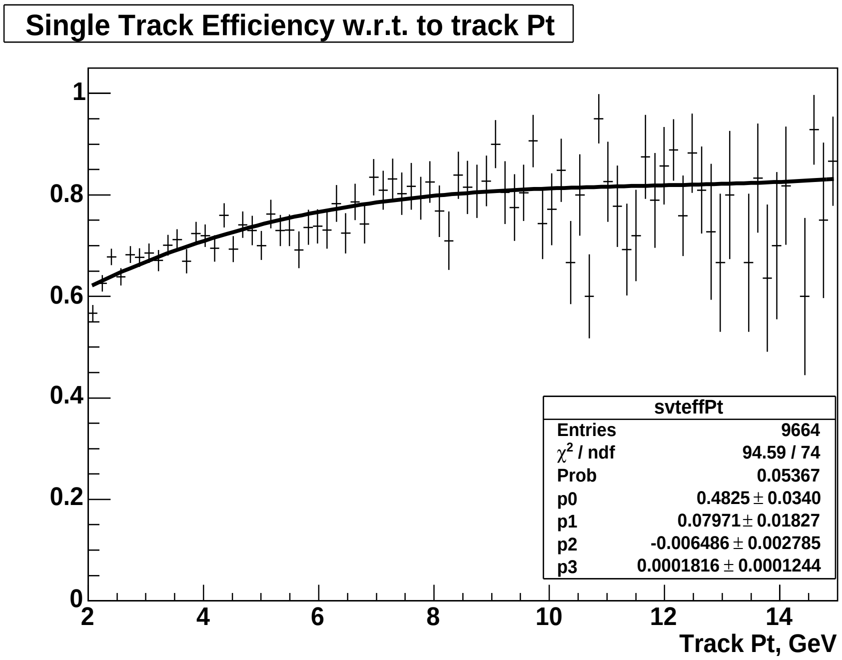


Figure 8.5: The single track finding efficiency as a function of the track transverse momentum. The points are from data. The fitted curve is a third order polynomial.

is 0.3 μm, and this is taken as the systematic error.

8.1.2 Systematic Uncertainties Due to a Possible Mis-alignment of the Silicon Detector

The track impact parameters and the event vertex position are calculated using information from the silicon detector, and this is dependent on knowing where there are hits along a silicon layer and also the position of the silicon layers relative to one another. Uncertainties in the exact position of the silicon can lead to mismeasurement of the impact parameters and vertex positions and therefore leads to an uncertainty in the measured lifetime. To quantify the size of such a systematic error, changes in the silicon alignment are considered and the effect on the measured lifetime calculated.

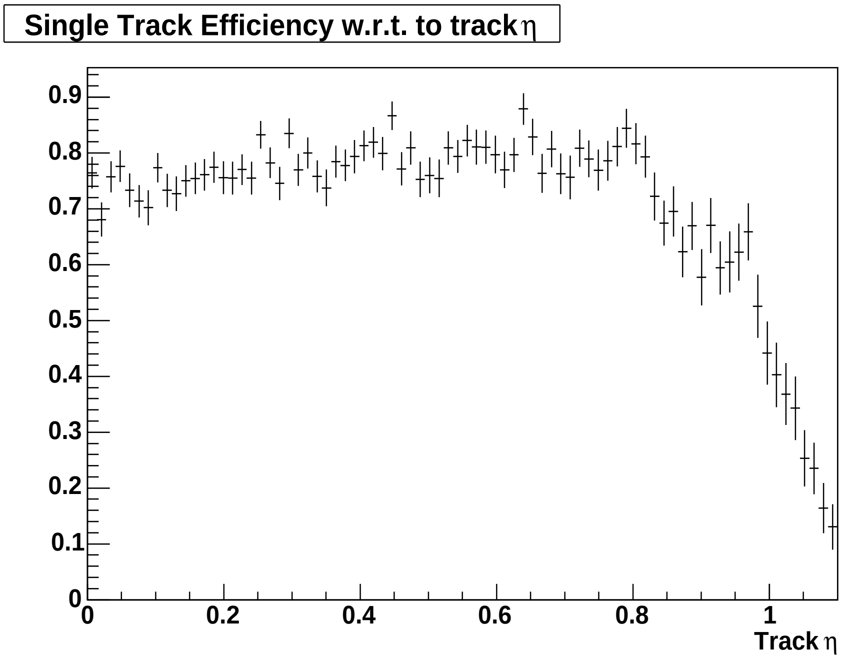


Figure 8.6: The single track finding efficiency as a function of the track η found in data.

The uncertainty in the radial position of the silicon layers at CDF is estimated as $\pm 50 \mu m$ [67]. The effect that such a shift can have on the track parameters is illustrated in the diagram in figure 8.7. The figure shows where a hit will be recorded given the assumed position of the silicon layer in consideration. The figure also shows where the hit will be recorded if in fact the silicon layer is not at its assumed position but instead 50 μm radially further away from the interaction point. The reconstruction algorithms assume knowledge of the silicon layer positions, and so if the layer is in a different position, it is clear that the r-ϕ position of the hit is miscalculated. Previous lifetime related studies at CDF have adopted 4 possible scenarios of misalignment in the silicon layers to study the effect the uncertainty in position of the detector layers will have on a lifetime measurement [67]. The following scenarios are considered:

1. The real positions of the Silicon layers are actually 50 μm further away from the

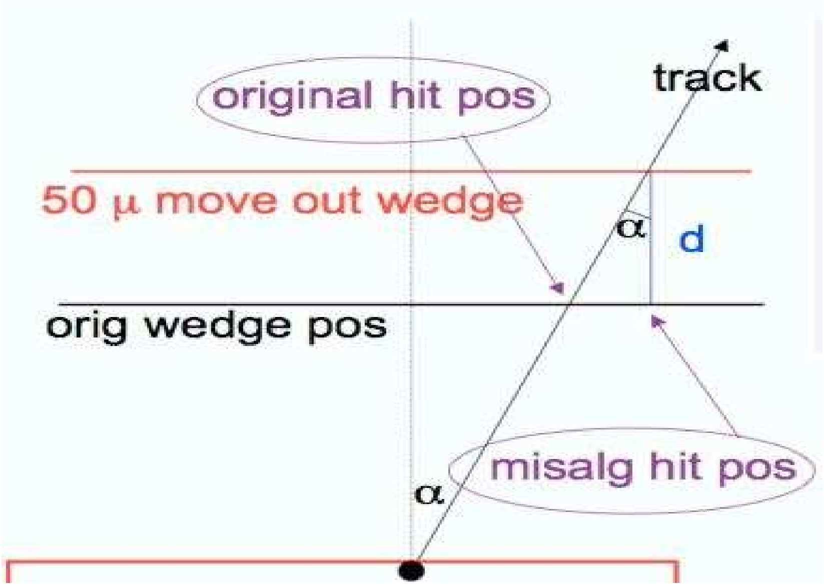


Figure 8.7: The diagram shows how the uncertainty in position of silicon wafer leads to the hit being reconstructed in the incorrect position

origin in the radial direction.

2. The real positions of the Silicon layers are actually 50 μm closer to the origin in the radial direction.
3. Silicon layers 2 and 3 as in item 1) above and the remainder in their assumed positions.
4. Silicon layers 2 and 3 as in item 2) above and the remainder in their assumed positions.

In the previous study the largest effect on measured lifetime was observed in scenario [1.] [67], and therefore this is the scenario that is considered for this analysis.

To simulate the effect of silicon misalignment in the fast simulation, a simple representation of the silicon detector is included. For layers that have been moved due to the alternate alignment scenario the the new hit position is calculated by:

$$hit_{shifted} = hit_{true} + R \cdot \sin(\alpha)$$

where in this expression hit is a measure of the hit position along a silicon layer, $hit_{shifted}$ is the miscalculated hit position due to the radial movement of the layers, and hit_{true} is the true hit position. The angle α is the angle of the track in the transverse plane measured from the bisector of a wedge as shown in figure 8.7. As there are 12 wedges each subtending an angle of 30° to the origin, the value of α must lie between $\pm 15°$. Finally, R is the radial shift, towards or away from the origin. The shifted hit positions are fitted to a straight line using the least squares technique, and new track impact parameters are calculated. Using the new impact parameters, the secondary vertices and new event lifetime are calculated. Using this generation method, 1000 samples of 24,000 events are generated according to scenario 1. described above. The mean difference between the fitted and input lifetimes for these samples is 0.5 μm.

To further test the effect that misalignment can have on the measured lifetime, scenario 3. described above is also considered. For the 1000 samples of events generated under misalignment scenario 3. the mean difference between the fitted and input lifetime is 0.4 μm.

Therefore a systematic error of 0.5 μm is assigned as the error due to alignment uncertainties of the silicon detector. The systematic error due to uncertainty in silicon alignment assigned by the previous studies was of order 1-2 μm; however, these were for samples using the J/Ψ trigger path, which does not select events based on the impact parameter of tracks found in the event. One of the effects of silicon misalignment is to change the detector resolution function, therefore the data samples that will be most affected by detector alignment will be those containing the prompt component that is most sensitive to the resolution function as discussed in section 5.3. Hence, it is expected that data samples containing a prompt component, such as those collected by the J/Ψ trigger, will be more sensitive to the silicon misalignment.

8.1.3 Systematic Error Due to Resolution Model

The measurement assumes that the detector resolution is perfectly Gaussian with a width of 26 μm. The basis for this model has been described in section 5.3. It is observed in the tests using realistic MC carried out in section 5.3 that even using the grossly unrealistic assumption on the resolution (a Gaussian width of 100 μm) the shift in lifetime is of order 4 μm and therefore only a small systematic uncertainty is expected due to the assumption of detector resolution. To determine the systematic error due to this choice of resolution function, other CDF analyses are used to derive an alternate resolution model. Fast simulated events are generated with the lifetimes smeared by this alternate resolution function, but the lifetime fit is performed assuming the resolution is a single Gaussian of width 26 μm. The difficulty in determining an accurate resolution model for data from the hadronic trigger is that the trigger removes most of the prompt component, and it is this component that is usually used to determine how the detector resolution smears lifetimes.

Determining an Alternate Resolution Model

To measure B_s mixing, an accurate resolution function was derived and described in [68]. This uses a sample of prompt D mesons which are then combined with a further track from the primary vertex. The lifetime distribution of these pseudo events is indicative of the detector resolution. The B_s mixing analysis used each event's individual σ_{ct} and did not use an average σ_{ct} as in this analysis. Therefore the resolution

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

Table 8.2: The alternate resolution function.

function quoted in [68] is written in terms of scale factors that multiply the event σ_{ct}. These results are re-interpreted to an average resolution using the distribution of σ_{ct} found in the data used for this analysis.

If σ_{res} is defined as the distribution of prompt lifetimes, then σ_{res} is the overall average resolution. The derived resolution function from the B_s mixing analysis [68] is used to weight each σ_{ct} in the data sample used in this thesis, and then they are smeared according to how the scale factors should affect each error on average. Using each σ_{ct} in the data sample the quantity $\sigma_{res}(t)$ is calculated:

$$\sigma_{res}(t) = \sum_i \frac{p_1}{\sigma_{cti}.s_1\sqrt{2\pi}} e^{-t^2/(2\sigma_{cti}.s_1)^2} + \frac{1-p_1}{\sigma_{cti}.s_2\sqrt{2\pi}} e^{-t^2/2(\sigma_{cti}.s_2)^2} \tag{8.3}$$

where p_1, s_1, s_2 are the fraction and scale factors for each resolution function component. The values of p_1, s_1, s_2 are taken from the resolution studies in [68] and are 0.17, 2.63, 1.28 respectively. The sum is calculated over events in the data sample used in this measurement. This quantity is calculated separately for the sideband and signal region, and then a background subtracted plot of σ_{res} is derived. The overall resolution function is then obtained by fitting σ_{res} to a sum of Gaussians. The resolution function is assumed to be of the form:

$$\sigma_{res}(t) = \frac{f_1}{\sigma_{cti}\sqrt{2\pi}} e^{-t^2/2\sigma_1^2} + \frac{(1-f_1)f_2}{\sigma_2\sqrt{2\pi}} e^{-t^2/2\sigma_2^2} + \frac{(1-f_1)(1-f_2)}{\sigma_3\sqrt{2\pi}} e^{-t^2/2\sigma_3^2} \tag{8.4}$$

The parameters of the resolution function obtained by the fit are given in table 8.2, and the plot of σ_{res} and the fit to σ_{res} shown in figure 8.8.

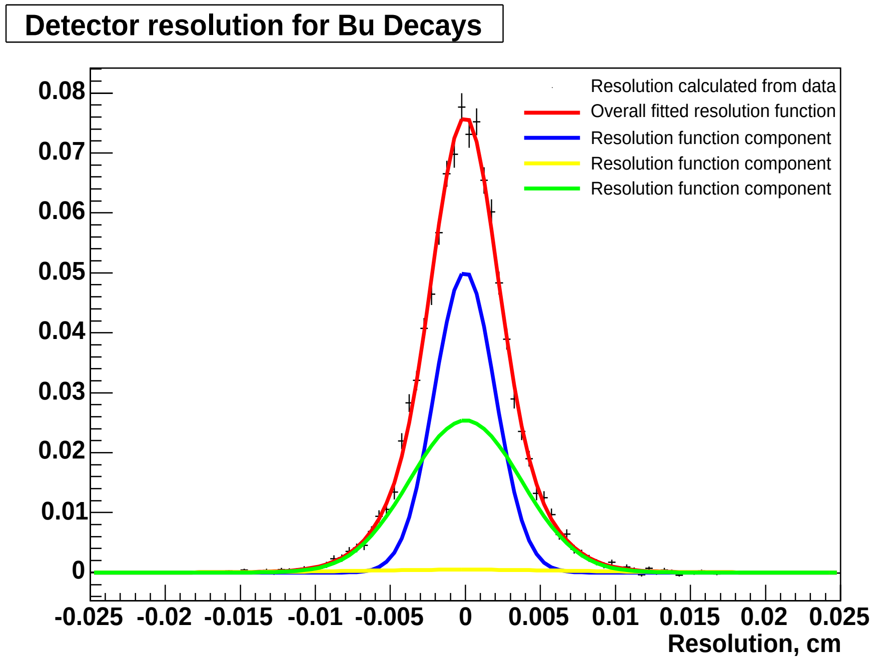


Figure 8.8: The resolution of the detector found by using information from other detailed studies. The fit is used to provide an alternate resolution model to the one that is used in the data fit.

Determining the Systematic Error Due to Resolution Function

Now that an alternative resolution function, σ_{res}, has been defined, events are generated using a modified fast simulation that smears the lifetime according to this alternate resolution. The samples are then fitted assuming that the resolution is a single Gaussian with width 26 μm. As usual, 1000 samples of 24000 events are generated and fit, and the mean difference between the fitted and input lifetime is found to be $0.3 \mu m$. This is taken to be systematic uncertainty due to the assumption that the detector resolution is a single Gaussian of width 26 μm. This small systematic uncertainty is expected, as from table 5.1 it is shown that the lifetime distribution of events in the sample leaves the likelihood only slightly sensitive to the value of σ_{ct}. As a cross check, a lifetime fit is performed on the data assuming this alternate resolution function. A shift of $0.2 \mu m$

is observed which is consistent with the systematic error assigned for this effect.

Only descriptions of detector resolution that are symmetric about 0 have been considered. Detailed studies of the detector resolution, carried out for a series of B hadron lifetime measurements including a J/Ψ in the final state, consider resolutions where the mean is shifted from zero. The difference in lifetimes between using models centered at zero and models shifted from zero was negligible [69].

8.1.4 Systematic Error Due to the Inclusion of the Cabibbo Suppressed Mode

The signal region contains events that are actually mis-reconstructed $B \rightarrow DK$ events, which means that the π_B track is really a kaon and not a pion. These events have the same true lifetime as our signal events but as the mass assignment for one track is incorrect (a kaon track has been given the mass of a pion), the mass and P_T of the B meson are shifted, and this causes the determined ct_i to be miscalculated. To estimate the effect of these events on the measured lifetime, it is first necessary to estimate what percentage of the signal is from the Cabibbo suppressed decay. The fast simulator is used to generate $B \rightarrow DK$ events with mean mass 5.276 GeV and width 20 MeV. These events are then reconstructed as though they were $B \rightarrow D\pi$. Of those events that pass the trigger, only 33% also pass the lower mass cut of 5.23 GeV. This information, in conjuction with the branching ratios from the PDG, is used to determine that only 3% of events in the sample are expected to be from the Cabibbo supressed mode.

To estimate the systematic uncertainty, 1000 samples of 24,000 events are generated such that 3% of them are $B \rightarrow DK$ events reconstructed as $B \rightarrow D\pi$. The lifetime is fitted and the observed mean difference betwen the fitted and input lifetime is less than 0.1 μm. This is condsidered to be neglgible.

8.2 Systematic Errors Associated with the Background

8.2.1 Systematic Errors Due to Assuming the Background in the Sideband is Representative of the Background Underneath the Peak

Another key assumption required to make the lifetime measurement is that the background in the upper sideband is representative of the the background under the peak. The selection criteria are chosen to remove any partially reconstructed B decays from the sample, and so it is a good assumption that the background underneath the peak is combinatoric like the upper sideband. However, changes in the distribution of lifetime and acceptance function as a function of background mass must be considered and the effect that these might have quantified.

8.2.2 Mass Lifetime Correlation in Background

It is assumed in chapter 5 that there is no correlation between the mass and lifetime of an event. This assumption is valid for signal and is verified using the realistic Monte Carlo. The mean lifetime for events with mass below and above 5.276 GeV are 618 μm and 615 μm respectively. The mass and lifetime for realistic Monte Carlo are shown in the profile plot in figure 8.9, and no significant correlation is observed. However, for background events, there is a correlation between the mass and lifetime of an event.

The correlation is shown in the profile plot in figure 8.10. The problem that arises from this is that the background lifetime distribution in the sideband is slightly different to that under the peak. To assess the effect of this, a correlation is introduced into the generated samples by rejecting events in such as way as to introduce a correlation similar to the one observed in data. The mean difference between the fitted and input lifetimes in 1000 samples containing 24000 signal events and 9000 background events each is 2.5 μm, which is taken to be the systematic error due to ignoring the correlation between mass and lifetime in the background.

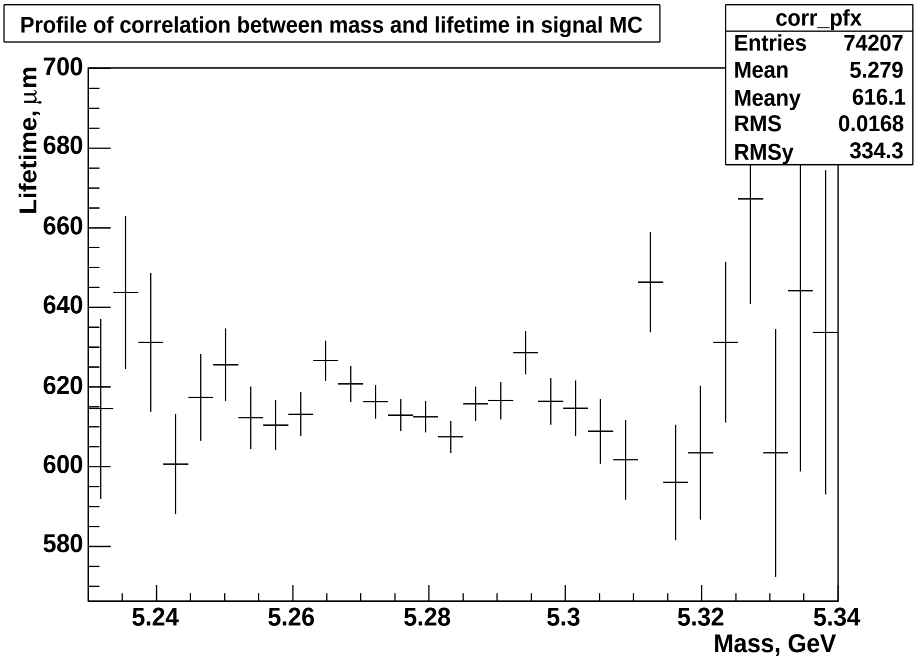


Figure 8.9: A profile plot to show how the lifetime varies with mass using realistic Monte Carlo events.

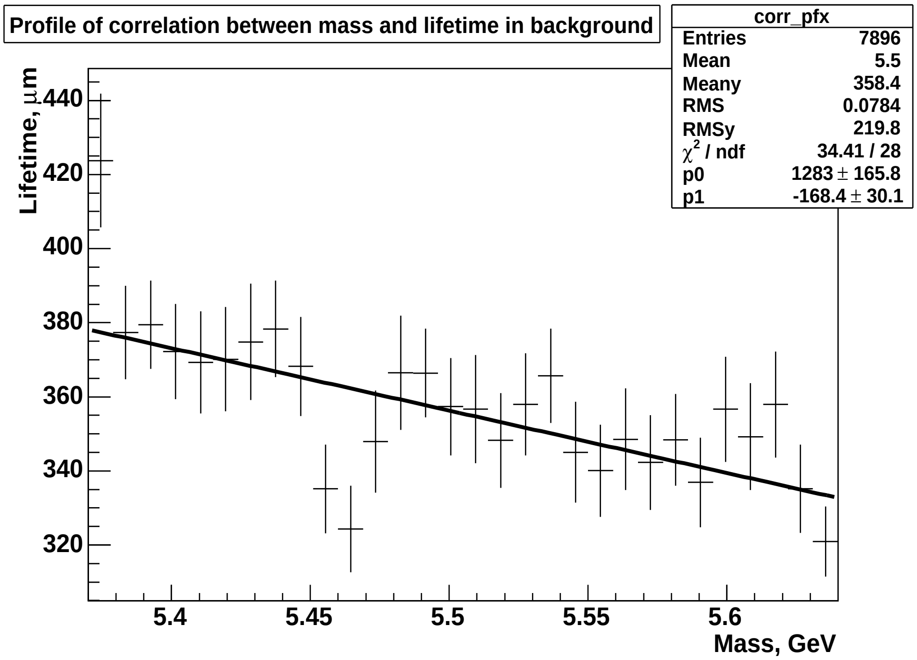


Figure 8.10: A profile plot to show how the lifetime varies with mass in background using the upper sideband.

Changes in the Distribution of Acceptance Function in the Background as a Function of Mass

The previous section discusses the change in lifetime as a function of mass in the upper sideband. A change in the distribution of acceptance functions should also be considered, as it is the acceptance functions of the upper sideband that are used to calculate the Fisher direction as described in section 5.6. For the nominal data fit, the events with mass between 5.37 and 5.5 GeV are used to represent the background acceptance distribution for calculation of the Fisher direction. To see how sensitive the result in data is to the distribution of acceptance functions in the sideband, a further fit to data is performed using the sideband events between a mass of 5.5 and 5.65 GeV to represent the background acceptance distribution. There is a shift in the data result of 0.1 μm, and so any systematic uncertainty due to a variation in the acceptance function across the background as a function of mass is negligible.

8.2.3 Background Lifetime Parameterization

The parameterization of the function $y(t)$ described in section 5.5 is defined by a choice of the number of parameters and spacing of the sampling points. To test the effect of different parameterizations on data, the data is fit using different configurations of $y(t)$. The two free variables in $y(t)$ are b and n and the default choice is b=4 and n=10. Fits to data are also performed using

- b=4 and n=11,
- b=4 and n=9,
- b=3 and n=10,
- b=5 and n=10.

In these 4 configurations the result in data did not shift by more than $0.4\mu m$.

However, as the parameterization is derived empirically, it is desirable to test the robustness of this parameterization more thoroughly. The fast simulation generates the background lifetime according to $y(t)$ and then uses the same functional form to fit the simulated data. To understand how well $y(t)$ can describe a distribution that is not generated with $y(t)$, an alternate function to generate background lifetime events must be defined. For a meaningful test, the alternate function must generate background that is similar to that observed in data. The upper sideband can be fitted by a sum of two exponentials as shown in figure 8.11. Although this is not a good quality fit the fitted function does broadly reproduce the features of the data. Therefore, 1000 samples of signal and background events are generated with the background lifetime generated using the exponential distribution. These samples are then fitted using the standard $y(t)$ to describe the background lifetime. The mean difference between the fitted and input lifetime is $0.8\mu m$, and this is taken as the systematic due to the background lifetime parameterization. It is consistent with the $0.4\mu m$ shift seen when using alternate parameterizations of $y(t)$ listed above on data.

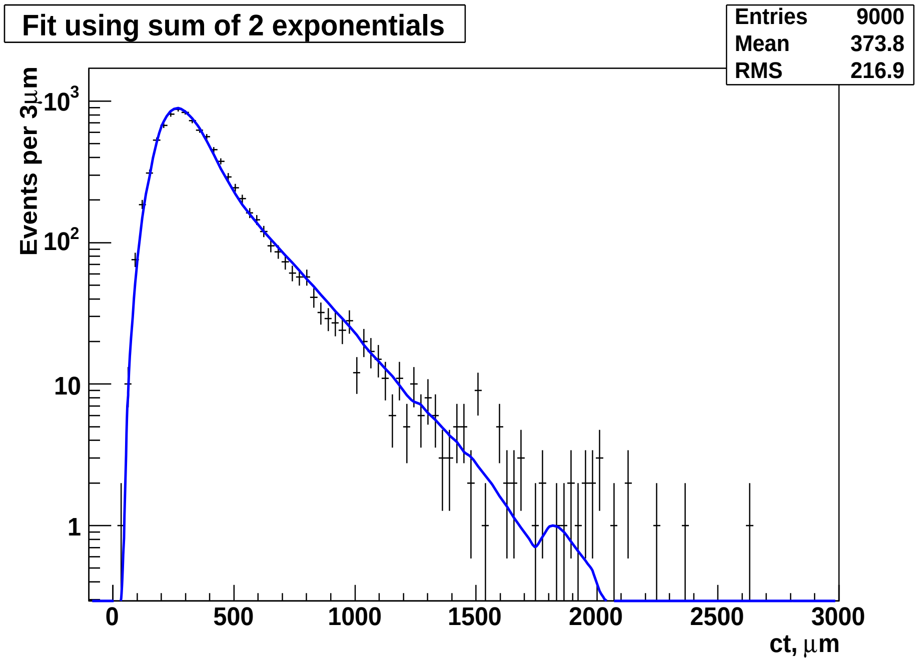


Figure 8.11: This plot shows a fit to the upper sideband using a sum of two exponentials. The shape resulting from this parameterization is close to that observed in data and so the parameters from this fit are used to generate an alternate set of background lifetimes.

8.2.4 Other Sources of Systematic Uncertainty

The final systematic error that is considered is the residual bias that remains from dealing with the difference between the signal and background acceptance functions. This is discussed in section 6.6.5 and is assigned to be $0.4\mu m$. This systematic also includes the general bias of the fitter and any systematic error due to using the Lagrange interpolating polynomials to describe the signal fraction as a function of Fisher scalar.

8.3 Summary of Systematic Errors

This chapter has investigated the assumptions made in formulating and parameterizing the likelihood and the effect that these assumptions will have on the final result. The assumptions made about detector performance are sensible and have been tested to the extent of variations observed in the data.

The systematic errors are summarized in table 8.3. The total systematic error has been calculated by adding the individual errors in quadrature. The leading systematic sources of uncertainty are the modeling of the single track finding efficiency with respect to impact parameter and transverse momentum, and the mass-lifetime correlation present in background. The total systematic uncertainty in the lifetime measurement is 4.5 μm, which is smaller than the statistical uncertainty in the measurement.

Source of Systematic Uncertainty	Error (μm)
Track finding efficiency dependence on d_0	3.1
Track finding efficiency dependence on P_T	1.8
Mass-Lifetime correlation in background	2.5
Background lifetime parameterization	0.8
Silicon Alignment	0.4
Fitter Bias	0.4
Lifetime resolution model	0.3
Track finding efficiency dependence on η	0.3
Total Systematic Uncertainty	4.5

Table 8.3: Summary of systematic uncertainties.

Chapter 9

Concluding Remarks

9.1 Results

Using 1.0 fb^{-1} of data collected by the B physics track triggers at CDF, the $B^{\pm}$ lifetime has been measured using the decay mode $B^{-} \rightarrow D^{0}\pi^{-}$. The measurement is

$$c\tau = 498.2 \pm 6.8 \pm 4.5\ \mu m \tag{9.1}$$

$$\tau = 1.662 \pm 0.023 \pm 0.015\ ps \tag{9.2}$$

where the first uncertainty is statistical and the second is systematic.

This result is consistent with the current world average 1.638±0.011 ps.

9.2 Concluding Remarks

This is the first measurement of a B hadron lifetime, using data collected by selecting displaced tracks, that has been performed without the use of simulation to correct for the selection biases.

One of the desired advantages of this new method was to remove the sources of systematic uncertainty due to differences between simulation and data that are faced by traditional methods. The aim of this new method was to devise a technique that resulted in smaller systematic sources of uncertainties than the traditional simulation dependent method.

As an example to see if this aim has been achieved, it is useful to consider a recent measurement of the Λ_b lifetime in the decay mode $\Lambda_b \rightarrow \Lambda_c^+\pi^-$ that has been performed using the traditional simulation dependent method. The measured lifetime is $c\tau = 422.8 \pm 13.8 \pm 8.8 \mu m$, where the first error is statistical and the second systematic. A summary of the systematic errors from this measurement is given in table 9.1 [25].

It is clear that the systematic error in the Λ_b lifetime is dominated by the modeling of the SVT in the CDF realistic Monte Carlo. In the simulation-free method, this does not have to be considered as no use of Monte-Carlo is made. Other sources of systematic uncertainty relate to Data-MC agreement of the Λ_b decay kinematics, as polarization or the Dalitz interference of the Λ_c decay are not included in the realistic Monte Carlo decay. Once again, using the simulation-free method, these sources of systematic uncertainty do not need to be considered. In fact, all the sources of systematic uncertainty marked with the ‡ symbol in table 9.1 would not need consideration if the simulation-free method is used, as these sources of systematic uncertainty relate to the differences between the data and simulation. Replacing all these sources of systematic

Source of Systematic Uncertainty	Error (μm)	
SVT Model	6.3	‡
Data-MC Agreement: Λ_c^+ Dalitz structure	3.7	‡
Combinatorial ct Template	2.9	
Data-MC Agreement: TrigCode re-weighting	2.0	‡
Alignment	2.0	
Data-MC Agreement: Λ_b^0 polarization	1.4	‡
Data-MC Agreement: Primary Vertex Position	1.2	‡
SVT-SVX d0 correlation	1.0	‡
Background Normalizations	1.0	
B^0 Efficiency	1.0	
B^0 Lifetime	1.0	
Other sources	negligible	
Total Systematic Uncertainty	8.8	

Table 9.1: Summary of systematic uncertainties for the Λ_b lifetime measurement using the traditional simulation dependent method. The items marked with ‡ are those that would not be present in the simulation free measurement.

uncertainly marked with ‡ would be the systematic uncertainty due to modeling the single track finding efficiency, which for the B^+ measurement is 3.6 μm (the sum in quadrature of systematic uncertainties due to single track finding efficiency modeling as a function of d_0, P_T and η).

This demonstrates the potential of the simulation free measurement as the systematic uncertainties due to Monte Carlo simulation and data agreement are significant compared to the systematic uncertainties due to assumptions on the single track finding efficiency.

Future improvements to the simulation-free method are also possible. The dependence of the single track finding efficiency on d_0 and P_T could, in the future, be incorporated into the PDF reducing the size of these systematic errors. The correlation of the mass and lifetime distribution in background can also be incorporated into the parameterization of the background. Such improvements have not been carried out for the $B^{\pm}$ measurement, as the statistical error is still larger than the systematic uncertainty and so the gain in precision is limited. Incorporating the dependence of the single track finding efficiencies on track parameters is also not trivial, however, these improvements will become necessary when there are larger samples of signal events available, and measurements become limited by the systematic uncertainties.

The LHC-b experiment is expected to collect large samples of B hadron decays whose lifetime distributions will also be biased due to impact parameter based triggers. The simulation-free method presented in this thesis could also be used at LHC-b, to measure the lifetimes of B mesons and B baryons. This technique will be particularly useful at LHC-b where the small systematic errors of this technique will complement the small statistical errors of measurement due to data that has high signal yields in a variety of decay modes.

In conclusion, this thesis has presented the precision measurement of the $B^{\pm}$ measurement using a new simulation-free technique to correct for the biases of the impact parameter based trigger. This new technique results in smaller systematic uncertainties than traditional methods that rely on simulation to correct for selection biases. Impact parameter based triggers can collect a wide range of decays that are interesting for lifetime measurement, as they are not limited to decay channels that include leptons. The simulation free technique is now ready to be used to measure the less precisely known lifetimes of the B_s, Λ_b, B_c and other B baryons as samples become available at CDF or LHC-b.

References

[1] E Fernandez et al. "Lifetime of particles containing a b quark.", *Phys Rev Lett*, **51** (1983) 1022–1025.

[2] N Lockyer et al. "Measurement of the lifetime of bottom hadrons", *Phys Rev Lett*, **51** (1983) 1316–1319.

[3] V Barger et al. "Lifetimes and branching fractions of mesons with heavy quark constituents", *Journal Phys G*, **5** (1979) L147.

[4] Paschos. "The charged current couplings and CP violation in the B meson system", *Physics Letters B*, **128** (1983) 240.

[5] Ginsparg. "Top quark mass and bottom quark decay", *Phys Rev Lett*, **50** (1983) 1415.

[6] Cleo. "Limit on the $b \to u$ coupling from semileptonic b decay", *Phys Rev Lett*, **52** (1984) 1084.

[7] Cusb. "Semileptonic decay of the B meson", *Physics Lett B*, **130** (1983) 444.

[8] I Bigi. "The lifetimes of heavy flavour hadrons - a case study in quark-hadron duality", *UND-HEP-99-BIG 07*.

[9] D Perkins. "Introduction to high energy physics". CUP (2000).

[10] W Burcham & M Jobes. "Nuclear and particle physics". Prentice Hall (1995).

[11] C Amsler et al. "Review of Particle Physics", *Physics Letters B*, **667** (2008) 1–1340.

[12] F Halzen & A Martin. "Quarks and Leptons: An introductory course in modern particle physics". John Wiley and Sons (1984).

[13] S Glashow. "Partial symmetries of weak interactions", *Nuclear Physics B*, **22**, 579.

[14] S Weinberg. "A model of leptons", *Phys Rev Lett*, **19**, 1264.

[15] N Cabibbo. "Unitary symmetry and leptonic decays", *Phys Rev Lett*, **10**, 531.

[16] K Maskawa M Kobayashi. "CP violation in the renormalizable theory of weak interaction", *Prog Theor Phys*, **49**, 652.

[17] S Stone (ed). "B decays". World Scientific Press (1994).

[18] N Uraltsev I Bigi, M Shifman. "Aspects of heavy quark theory", *Annu. Rev. Nucl. Part. Sci.*, **47** (1997) 591.

[19] C Tarantino. "Beauty Hadron Lifetimes and B-Meson CP violation parameters from Lattice QCD", *Eur Phys J*, **C33** (2004) S895.

[20] F Gabbiani et al. "Λ_b lifetime puzzle in heavy quark expansion", *Phys Rev D*, **68** (2003) 114006.

[21] F Gabbiani et al. "Spectator effects and lifetimes of heavy hadrons", *Phys Rev D*, **70** (2004) 094031.

[22] A Abulencia et al. "CDF Public webpage". `http://www-cdf.fnal.gov/physics/new/bottom/061130.blessed-bh-lifetime_v2/`

[23] A Abulencia et al. "CDF Public webpage". `http://www-cdf.fnal.gov/physics/new/bottom/080207.blessed-bs-lifetime/`

[24] A Abulencia et al. "Measurement of the Λ_b lifetime in $\Lambda_b \to J/\Psi\Lambda$ in $p\bar{p}$ collisions at $\sqrt(s) = 1.96 TeV$", *Phys Rev Lett*, **98** (2007) 122001.

[25] R Mumford et al. "Λ_b Lifetime using 2 displaced track triggers". CDF Note 8578 (2008).

[26] A Abulencia et al. "CDF Public webpage". `http://www-cdf.fnal.gov/physics/new/bottom/080703.blessed-lblcpi-ct/`

[27] Accelerator Division. "Accelerator Concepts", *Rookie Books*. `www-bdnew.fnal.gov/operations/rookie-books/rbooks.html`

[28] `http://www-cdf.fnal.gov/events/pic.html`

[29] J Marriner. "Stochastic Cooling Overview", *Nucl. Instrum. Methods. Phys. Res A*, **532** (2004) 11–18.

[30] A Sidorin I Meshkov. "Electron Cooling", *Nucl. Instrum. Methods. Phys. Res A*, **532** (2004) 19–25.

[31] `http://www.fnal.gov/pub/now/tevlum.html`

[32] The CDF collaboration. "Technical Design Report", *FERMILAB-PUB-96-390-E*.

[33] A Sill. "CDF Run II silicon tracking projects", *Nucl. Instrum. Methods. Phys. Res A*, **447** (2000) 1–8.

[34] J Antos et al. "The SVX II silicon vertex detector upgrade at CDF", *Nucl. Instrum. Methods. Phys. Res A*, **383** (1996) 13–20.

[35] C Hill. "Initial experience with the CDF layer 00 silicon detector", *Nucl. Instrum. Methods. Phys. Res A*, **511** (2003) 118–120.

[36] A Affolder et al. "CDF Central Outer Tracker", *Nucl. Instrum. Methods. Phys. Res A*, **526** (2004) 249–299.

[37] D Acosta et al. "A Time of Flight detector in CDF II", *Nucl. Instrum. Methods. Phys. Res A*, **518** (2004) 605–608.

[38] L Balka et al. "The CDF central electromagnetic calorimeter", *Nucl. Instrum. Methods. Phys. Res A*, **267** (1988) 272–279.

[39] S Bertolucci et al. "The CDF central and endwall hadron calorimeter", *Nucl. Instrum. Methods. Phys. Res A*, **267** (1988) 301–314.

[40] M Albrow et al. "The CDF Plug upgrade: electromagnetic calorimeter test beam results", *Nucl. Instrum. Methods. Phys. Res A*, **480** (2002) 524–546.

[41] A Artikov et al. "Design and construction of new central and forward muon counters at CDF II", *Nucl. Instrum. Methods. Phys. Res A*, **538** (2005) 385–371.

[42] C Hays et al. "The COT Pattern Algorithm and Offline Code". CDF Note 6992 (2004).

[43] C Chen et al. "Measurement of the track reconstruction efficiency in the COT using a MC track embedding technique". CDF Note 6394 (2003).

[44] The Tracking Group. ""Outside-In" Silicon Tracking at CDF". CDF Note 5991 (2002).

[45] S. Lai et al. "An updated measurement of beam width at CDF". CDF Note 6492 (2003).

[46] H. Stadie et al. "The Beam Position in Run II". CDF Note 6327 (2003).

[47] The Bs mixing group. "The Performance of the Event by Event Primary Vertex Finder". CDF Note 7512 (2005).

[48] B Ashmanskas et al. "The CDF Silicon Vertex Trigger", *Nucl. Instrum. Methods. Phys. Res A*, **518** (2004) 532–536.

[49] G Gomez-Ceballos. "Event Builder and L3 at the CDF exp", *Nucl. Instrum. Methods. Phys. Res A*, **518** (2004) 522–524.

[50] R Downing et al. "Track extrapolation and distribution for the CDF II trigger system", *Nucl. Instrum. Methods. Phys. Res A*, **570** (2007) 36–50.

[51] J Marriner. "Secondary vertex fit with mass constraints". CDF Note 1996 (1993).

[52] P Nason et al. "The total cross section for the production of heavy quarks in hadronic collisions", *Nuclear Physics B*, **303** (1988) 607–633.

[53] P Nason et al. "The one particle inclusive diffractive cross section for heavy quark production in hadronic collisions", *Nuclear Physics B*, **327** (1989) 49–92.

[54] David Lange. "The EvtGen particle decay simulation package", *Nucl. Instrum. Methods. Phys. Res A*, **462** (2001) 152–155.

[55] Application Software Group. "Detector description and simulation tool", *Cern programme library,Cern-DD-78-2-REV*.

[56] The B_s mixing group. "B meson lifetime determination in fully reconstructed hadronic decays". CDF Note 7386 (2005).

[57] J Rademacker. "Reduction of the statistical power per event due to upper lifetime cuts in lifetime measurements", *Nucl. Instrum. Methods. Phys. Res A*, **570** (2007) 525–528.

[58] G Punzi. "Comments on likelihood fits with variable resolutions", *arXiv:physics/0401045v1*. Published in conference proceedings Phystat 2003. SLAC.

[59] The CDF Bs mixing working group. "B meson lifetime determination in fully hadronic decays", *CDF note 7386*.

[60] R Fisher. "The use of multiple measurements in taxonomic problems", *Annals of Eugenics*, **7** (1936) 179–199.

[61] G Cowan. "Statistical data analysis", *Oxford Science Publications*, pages 51–54.

[62] W Krazanowski. “The algebraic basis of classical multivariate methods”, *The Statistician*, **20**, 4 (1971) 51–61.

[63] Brandon Archer & Weisstein. “Lagrange Interpolating polynomials”, *MathWorld A Wolfram Web Resource*. `http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html`

[64] E Waring. “Problems concerning interpolations”, *Philosophical transactions of the royal society of London*, **69** (1779) 59–67.

[65] M. Matsumoto & T. Nishimura. “Mersenne Twistor: a 623-dimensionally equidistributed uniform pseudorandom number generator”, *ACM Transactions on Modeling and Computer Simulation*, **8**, 1 (1998) 3.

[66] Stefano Torre. *Private communication.*

[67] C Paus K Anikeev. “New measurements of lifetimes of B mesons in exclusively reconstructed decays”. CDF Note 6551 (2004).

[68] The B_s mixing group. “Scale factors for proper time uncertainties”. CDF Note 7500 (2005).

[69] J. Boudreau et al. “Measurement of Exclusive B Lifetiems in the modes $B \to J/\Psi X$”. CDF Note 8524 (2008).