THE UNIVERSITY OF
MELBOURNE

MSc Thesis

Time-Dependent Analysis of $B^0 \to K_S\pi^0$

Author:
Timothy Green

Supervisor:
Martin Sevior

Department of Physics
The University of Melbourne

28 May 2020

Abstract

Current results from previous B-factory experiments such as BaBar and Belle for the $B^0 \to K_s\pi^0$ mode indicate a 2σ deviation between the Standard Model and experimental data fo CP parameters (A_{CP}) and S_{CP}. As such, this is one of the leading places to look for New Physics particles for example the Z' boson, leptoquarks, and diquarks. This is hoped to be resolved by the current Belle-II experiment which will have a significantly higher integrated luminosity than either Belle or BaBar by the time data collection is completed.

The aim of this work is to establish an analysis framework to study this decay mode using Monte-Carlo generated data. This analysis framework will then be tested by the use of Monte-Carlo toy fits and applied to the $B^0 \to J/\Psi K_s$ control mode. Various forms of continuum suppression are tested, and the expected results for various levels of integrated luminosity are examined. An in-depth study of the DeltaT error distribution function will be completed including a range of detector based results.

Contributions and Acknowledgements

The first two chapters include a review of the current literature on the topic, and an overview of the Belle-II experiment featuring an in-depth discussion of the workings behind the detector and each of the sub-detectors, and a comparison between this experiment and the previous Belle experiment. The third chapter outlines the analysis framework previously studied by Martin Sevior and Brian Chan and the developments by the author. The fourth chapter is an original study of the distribution of DeltaT errors and how this applies to the analysis framework. The fifth and final chapter consists of an original work detailing the methods behind validating the model.

I would like to acknowledge my supervisor, Martin Sevior, who kept me in the right direction and provided valuable feedback and insights into my results and techniques. I would also like to thank the Belle-II time-dependent CP violation group lead by Yusa-san and Stefano, and the hadronic and charmless B-decay working group lead by Diego, Martin, and Pablo. These groups have provided answers to many of the problems I have had throughout my studies.

Contents

1 Literature Review **4**
- 1.1 The Standard Model . . . 4
  - 1.1.1 The Standard Model of Particle Physics . . . 4
  - 1.1.2 The CKM Matrix and Quark Mixing . . . 5
  - 1.1.3 Unitary Relations and the Unitary Triangle . . . 6
- 1.2 CP Violation . . . 7
  - 1.2.1 Discrete Symmetries . . . 7
  - 1.2.2 Discovery of CP Violation . . . 7
  - 1.2.3 Mechanism for CP Violation in the Standard Model . . . 7
  - 1.2.4 Types of CP Violation . . . 8
  - 1.2.5 The Sakharov Conditions for Baryogenesis . . . 9
- 1.3 The $B \to K\pi$ Puzzle . . . 10
  - 1.3.1 The $B \to K\pi$ System . . . 10
  - 1.3.2 The Puzzle for Direct CP Violation . . . 10
  - 1.3.3 Interference CP Violation in $B \to K^0\pi^0$ . . . 11
  - 1.3.4 Possible Resolutions . . . 12

2 The Belle II Experiment **13**
- 2.1 Introduction to Belle II . . . 13
  - 2.1.1 The SuperKEKB Accelerator . . . 13
  - 2.1.2 Comparison to Belle . . . 13
- 2.2 The Belle II Detector . . . 14
- 2.3 Event Variables . . . 16
- 2.4 Event Classification . . . 17

3 Analysis Strategy **18**
- 3.1 Analysis Software . . . 18
- 3.2 Vertexing . . . 18
  - 3.2.1 BTube . . . 18
- 3.3 Flavour Tagging . . . 19
- 3.4 Continuum Suppression . . . 20
  - 3.4.1 Training Variables . . . 21
  - 3.4.2 Method of Continuum Suppression . . . 23
- 3.5 Fitting . . . 25
  - 3.5.1 DeltaT Probability Density Function . . . 25
  - 3.5.2 All Fits . . . 26
  - 3.5.3 Signal DeltaT Fits with Flavour Tagging . . . 29

4 Studies in DeltaT and DeltaT Errors **30**
- 4.1 The DeltaT Error distribution . . . 30
- 4.2 The K_s vertex location . . . 31

4.2.1 The PXD . . . 31
4.2.2 Number of SVD Hits . . . 32
4.3 DeltaT Residues . . . 34

5 Validation **35**
5.1 Expected Number of Events . . . 35
5.1.1 Signal . . . 35
5.1.2 Background . . . 36
5.2 Toy Single Fits . . . 36
5.3 MC Toy Results . . . 37
5.3.1 $1ab^{-1}$ Toy MC Results . . . 38
5.3.2 $62.8fb^{-1}$ Toy MC Results . . . 38
5.4 Linearity Fits . . . 39
5.4.1 The Initial Value of Scp . . . 39
5.5 Control Modes . . . 40
5.5.1 Control Mode Fits . . . 41
5.5.2 Control Mode Toy MC Results . . . 42
5.6 B-Meson Lifetime Fits . . . 44

Conclusion and Future Work **45**

References **45**

Chapter 1

Literature Review

1.1 The Standard Model

1.1.1 The Standard Model of Particle Physics

During the 1960's and 70's, the Standard Model of Particle Physics was developed as a way to classify the fundamental particles as well as describe their behaviours and interactions with each other. It is widely believed to be one of the great successes in modern physics describing three of the four fundamental forces in nature - Electromagnetism, the weak interaction, and the strong interaction. It is able to describe these with very high precision and for the most part provides exceptional agreement with experimental data. Currently, the Standard Model is not able to explain the fourth fundamental force - gravity, however attempts have been made by incorporating a spin-2 'graviton'.

The Standard Model classifies particles as quarks, leptons or bosons depending on their properties and how they interact. Quarks interact through all three of the forces however the leptons can only interact through electroweak interactions (Electromagnetism or the weak interaction). The bosons are the carriers of each fundamental force. The photon (γ) mediates electromagnetism, the $W^{\pm}$ and Z bosons mediate the weak interaction and the eight coloured gluons (g) mediated the strong interaction. If the hypothetical graviton was included it would mediate gravitational forces. The famous Higgs boson is able to interact with particles and provides a mechanism for obtaining mass. Leptons and quarks are known as fermions and come in three generations differing only by mass. All fermions are spin-1/2 and each has its own associated anti-particle.

Because of a principle called colour confinement, all matter made of quarks must exist in a bound state usually consisting of two or three quarks. Mesons are the states with two quarks and consist of a quark-antiquark pair ($q\bar{q}$). Baryons are the states with three quarks, either all regular particles (qqq) or all anti-particles ($\bar{q}\bar{q}\bar{q}$). Bound states with more than three quarks (e.g. tetraquarks) have been experimentally shown to exist however are very rare.

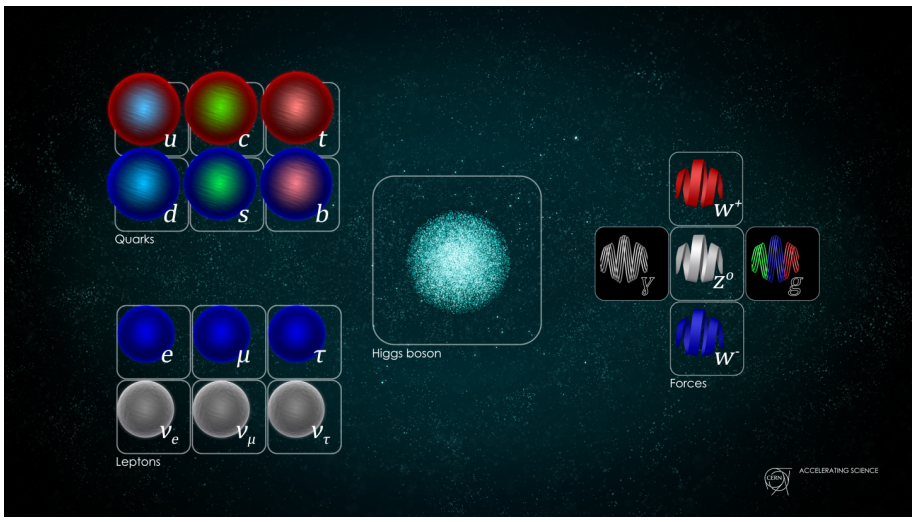

Figure 1.1: The fundamental particles in the Standard Model. (Image: CERN)

1.1.2 The CKM Matrix and Quark Mixing

In the Standard Model it is possible for a quark to spontaneously change flavour through interactions with the $W^{\pm}$ bosons. Because of the charge of the mediating boson, this must occur only between down type quarks to up type quarks, or up type quarks to down type quarks*. These flavour changing currents arise from Yukawa interaction terms in the SM Lagrangian[8]:

$$\frac{-g}{\sqrt{2}}(\bar{u_L}, \bar{c_L}, \bar{t_L})\gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + h.c. \tag{1.1}$$

Here, the elements of the vectors describe the three different generations of quarks, γ^μ is the set of Dirac matrices, and V_{CKM} is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix[7, 18]. The non-diagonal elements of this matrix are responsible for the flavour changing process. This matrix is given as follows:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.2}$$

These matrix elements are free parameters in the Standard Model and must be determined experimentally. If the absolute value of one of these elements is very low, any interaction involving this element will be suppressed. This is known as Cabibbo suppression and is notable in $|V_{ub}| = (3.38 \pm 0.36) \times 10^{-3}$.[20]

In its most general form, this matrix can be written in terms of three mixing angles $(\theta_{12}, \theta_{13}, \theta_{23})$ and one complex phase (δ) using the standard parametrisation†[27]:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{1.3}$$

*Up type particles have charge 2/3 while down type particles have charge -1/3. The difference in charges therefore matches that of the W bosons.

†The notation $c_{ij} \equiv \cos(\theta_{ij})$ and $s_{ij} \equiv \sin(\theta_{ij})$ is used.

δ is a CP-violating phase and is the cause of CP Violation in quark mixing.[9]

It is worth noting that because there is no equivalent term in the SM Lagrangian for the Z boson, thus quark mixing can only occur using charged currents. As a result, flavour-changing neutral currents are not allowed in the Standard Model (at least to tree level diagrams).

1.1.3 Unitary Relations and the Unitary Triangle

Since flavour-changing processes must conserve probability (either it changes or it does not), the CKM Matrix must be unitary. By definition this gives the relation:

$$V_{CKM}^{\dagger} V_{CKM} = V_{CKM} V_{CKM}^{\dagger} = I_3 \tag{1.4}$$

Where I_3 represents the 3x3 identity matrix. This gives a set of nine equations relating products of CKM matrix elements. The equation of interest is given below:

$$V_{ud} V_{ub}^{*} + V_{cd} V_{cb}^{*} + V_{td} V_{tb}^{*} = 0 \tag{1.5}$$

This is then divided by the term known to the highest accuracy ($V_{cd} V_{cb}^{*}$) and shown pictorially using a triangle on the complex plane with vertices at $(0,0)$, $(0,1)$ and $(\bar{\rho}, \bar{\eta})$. Here, $\bar{\rho}$ and $\bar{\eta}$ are parameters used in an alternate parametrisation of the CKM matrix known as the Wolfenstein parametrisation[28]. This triangle is known as the Unitary Triangle[8].

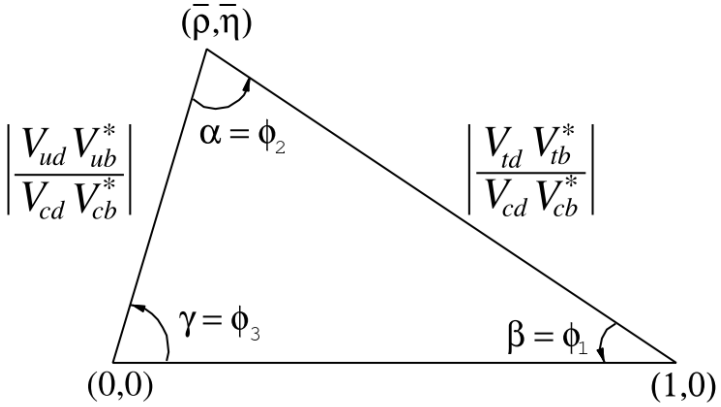

Figure 1.2: The Unitary Triangle.

The angles here are given by:

$$\phi_1 = \beta = \arg\left(-\frac{V_{cd} V_{cb}^{*}}{V_{td} V_{tb}^{*}}\right) \tag{1.6}$$

$$\phi_2 = \alpha = \arg\left(-\frac{V_{td} V_{tb}^{*}}{V_{ud} V_{ub}^{*}}\right) \tag{1.7}$$

$$\phi_3 = \gamma = \arg\left(-\frac{V_{ud} V_{ub}^{*}}{V_{cd} V_{cb}^{*}}\right) \tag{1.8}$$

The use of ϕ_i or Greek characters varies throughout the literature.

1.2 CP Violation

1.2.1 Discrete Symmetries

There are three important discrete transformations in the Standard Model. These are given as follows:

- $\mathcal{C}$ (charge) Transformation: Negation of the charge of a particle (e.g. $e^- \to e^+$).
- $\mathcal{P}$ (parity) Transformation: Reversal of the spatial coordinates (e.g. $\psi(t, \vec{x}) \to \psi(t, -\vec{x})$).
- $\mathcal{T}$ (time) Transformation: Reversal of the temporal coordinate (e.g. $\psi(t, \vec{x}) \to \psi(-t, \vec{x})$).

If a theory remains unchanged under any of these transformations it is symmetric under that transformation. An example of this is that electromagnetism is symmetric under $\mathcal{C}$ transformations. If the theory is not symmetric, the symmetry is said to be violated.

The $\mathcal{CPT}$ theorem states that any Lorentz invariant Quantum Field Theory must be symmetric under the combination of all three of these transformations[26].

1.2.2 Discovery of CP Violation

Up until the late 1950's it was believed that each symmetry ($\mathcal{C}$, $\mathcal{P}$ and $\mathcal{T}$) was conserved under all interactions. In 1957 C. S. Wu et al.[29] showed experimentally that the $\mathcal{P}$ symmetry was violated in the weak interactions by looking at beta decay in ^{60}Co. After this it was believed that $\mathcal{T}$ and the combination $\mathcal{CP}$ were symmetries.

Cronin and Fitch's famous experiment[10] put this theory to rest in 1964 when they discovered that the K_2^0 meson was found to decay into two pions in contradiction to the $\mathcal{CP}$ symmetric theory which held that only three pion decays were possible. This meant the weak eigenstates of the K^0 meson could not have been the same as the $\mathcal{CP}$ eigenstates which directly implies $\mathcal{CP}$ violation.

This phenomenon of CP violation in the Standard Model has been confirmed many times since its discovery, notably in the Belle and BaBar experiments in the early 2000's. It is a major focus of the current Belle II experiment.

1.2.3 Mechanism for CP Violation in the Standard Model

$\mathcal{CP}$ violation can be formulated in the Standard Model[‡] by writing a general state as a superposition of both flavours of the B^0 meson[2]. Over time this state is allowed to decay into final state particles.

$$|\psi(t)\rangle = a(t)|B^0\rangle + b(t)|\bar{B}^0\rangle + c_1(t)|f_1\rangle + c_2(t)|f_2\rangle + \dots \tag{1.9}$$

In Eq. 1.9, $c_n(t)$ must be zero for all n. The time dependence of $a(t)$ and $b(t)$ is given by the usual Schrödinger equation using a non-Hermitian[§] Hamiltonian represented by a 2x2 matrix which can be deconstructed into two Hermitian matrices:

$$\mathbf{H}_{eff} = \mathbf{M} - \frac{i}{2}\mathbf{G} \tag{1.10}$$

[‡]Specifically in the B^0 sector here.

[§]Non-Hermitian because the particle must be allowed to decay.

These two matrices **M** and **G** are associated with on-shell and off-shell transitions not involving the final states. The eigenstates of this Hamiltonian must have a definite mass and decay width. They can be written as a linear combination of B^0 flavours[¶].

$$|B_L\rangle = p|B^0\rangle + q|\bar{B}^0\rangle \tag{1.11}$$
$$|B_H\rangle = p|B^0\rangle - q|\bar{B}^0\rangle \tag{1.12}$$

From these definitions, two important quantities are defined - the mass and width splittings:

$$\Delta m = m_H - m_L \tag{1.13}$$
$$\Delta\Gamma = \Gamma_H - \Gamma_L \tag{1.14}$$

Solving the Schrödinger equation using Eq. 1.10 gives the following relation between p and q.

$$\left(\frac{p}{q}\right)^2 = \frac{\mathbf{M}_{12}^* - (i/2)\mathbf{G}_{12}^*}{\mathbf{M}_{12} - (i/2)\mathbf{G}_{12}} \tag{1.15}$$

1.2.4 Types of CP Violation

There are three types of CP violation in the Standard Model[23]. Throughout this section it will be notationally convenient to define the following decay amplitudes to any final state f:

$$A_f = \langle f|\mathbf{H}|B^0\rangle \tag{1.16}$$
$$\bar{A}_f = \langle f|\mathbf{H}|\bar{B}^0\rangle \tag{1.17}$$
$$A_{\bar{f}} = \langle \bar{f}|\mathbf{H}|B^0\rangle \tag{1.18}$$
$$\bar{A}_{\bar{f}} = \langle \bar{f}|\mathbf{H}|\bar{B}^0\rangle \tag{1.19}$$

Direct CP Violation

Direct CP violation occurs when a decay has a different decay rate than its CP conjugate. In terms of decay amplitudes:

$$|A_f| \neq |\bar{A}_{\bar{f}}| \tag{1.20}$$

This is parametrised by A_{cp} defined as:

$$A_{cp} = \frac{\Gamma(B^0 \to f) - \Gamma(\bar{B}^0 \to \bar{f})}{\Gamma(B^0 \to f) + \Gamma(\bar{B}^0 \to \bar{f})} \tag{1.21}$$

This can occur when there are both two different weak phases, and two different strong phases involved in the decay amplitudes.

Indirect CP Violation

Indirect CP Violation occurs when the mixing rates for each flavour eigenstate differs from each other. This requires the condition

$$|q| \neq |p| \tag{1.22}$$

This is usually small in B decays and is not significant here.

[¶]This can be generalized if $\mathcal{CPT}$ violation is allowed. For simplicity it will be assumed here

Interference CP Violation

CP Violation through interference occurs when a final state is obtainable not only from a direct decay, but also from a decay via a mixed state[6]. For the final state being $K_s\pi^0$, it is from the decay rates as follows:

$$B^0 \to K_s\pi^0 \tag{1.23}$$
$$B^0 \to \bar{B}^0 \to K_s\pi^0 \tag{1.24}$$

This type of CP Violation is interesting because it results in a time dependence in the asymmetry due to the mixing. The asymmetry is now parametrised as:

$$A_{cp}(t) = \frac{\Gamma(B^0(t) \to f) - \Gamma(\bar{B}^0(t) \to \bar{f})}{\Gamma(B^0(t) \to f) + \Gamma(\bar{B}^0(t) \to \bar{f})} \tag{1.25}$$

This can be further simplified so that it is in terms of three parameters, A_{cp}, S_{cp} and Δm.

$$A_{cp}(t) = A_{cp}\cos(\Delta m t) + S_{cp}\sin(\Delta m t) \tag{1.26}$$
$$A_{cp} = \frac{1 - |\xi_f|^2}{1 + |\xi_f|^2} \tag{1.27}$$
$$S_{cp} = \frac{2\Im(\xi_f)}{1 + |\xi_f|^2} \tag{1.28}$$

This uses the parameter ξ_f defined as:

$$\xi_f = \frac{q}{p}\frac{\bar{A}_f}{A_f} \tag{1.29}$$

For the decay $B^0 \to K_s\pi^0$, Eq. 1.28 reduces to an expression related to the angle ϕ_1 in the Unitary Triangle:

$$S_{cp} = \sin(2\phi_1) \tag{1.30}$$

1.2.5 The Sakharov Conditions for Baryogenesis

One of the many reasons CP violation is important to study in regards to the current state of physics is its applicability to the astrophysical phenomenon of baryogenesis; Why is the universe dominated by matter and not antimatter? How this occurs is known as the Baryon Asymmetry of the Universe (BAU).

In 1967, Andrei Sakharov gave a set of three conditions required for baryogenesis to occur[25]:

- Baryon number violation - An obvious necessity allowing the difference between matter and antimatter to occur.
- C and CP violation - There must be some processes which favour decays to regular matter.
- Non-thermal Equilibrium Interactions - Required to decrease the rate of matter-antimatter annihilation processes which reduce the asymmetry.

Defining the asymmetry as the number of baryons per photon N_B/s, the total amount of asymmetry allowed by the CKM mechanism was calculated by Huet and Sather to be[14]:

$$\left|\frac{n_B}{s}\right| < 6 \times 10^{-27} \tag{1.31}$$

This is much too small to account for the quoted asymmetry observed through astrophysical observations:

$$\left|\frac{n_B}{s}\right| \approx 6 \times 10^{-11} \tag{1.32}$$

A better understanding of CP Violation could lead to a solution to this problem, possibly through the introduction of new physics.

1.3 The $B \to K\pi$ Puzzle

1.3.1 The $B \to K\pi$ System

The $B \to K\pi$ system refers to all possible decays from a B meson (possibly charged) to all charge conserving variations of a K meson and a pion. This consists of four decays: $B^+ \to K^+\pi^0$, $B^+ \to K^0\pi^+$, $B^0 \to K^+\pi^-$ and $B^0 \to K^0\pi^0$. Each has an associated branching ratio and A_{cp}, and the $B^0 \to K^0\pi^0$ mode has an associated S_{cp}.

To simplify notation, the decay amplitudes for each mode are written as:

$$\begin{aligned} A^{+-} &= A(B^+ \to \pi^+K^-), \quad & A^{+0} &= A(B^+ \to \pi^+K^0) \\ A^{0+} &= A(B^+ \to \pi^0K^+), \quad & A^{00} &= A(B^0 \to \pi^0K^0) \end{aligned} \tag{1.33}$$

These amplitudes are related through a quadratic isospin relation[3]:

$$\sqrt{2}A^{00} + A^{-+} = \sqrt{2}A^{0+} + A^{+0} \tag{1.34}$$

There are six Feynman diagrams contributing to these decays representing colour favoured and colour suppressed gluonic penguins (P'_{tc} and P'_{uc}), colour favoured and colour suppressed electroweak penguins (P'_{EW} and P'^C_{EW}) and colour favoured and colour suppressed tree diagrams (T' and C'). The SU(3) flavour symmetry allows the following relations between electroweak penguin amplitudes and tree amplitudes to good approximation:

$$P'_{EW} = \frac{3c_9}{2c_1}RT', \qquad P'^C_{EW} = \frac{3c_9}{2c_1}RC' \tag{1.35}$$

Where c_i are known as Wilson coefficients and R is composed of CKM matrix elements.

1.3.2 The Puzzle for Direct CP Violation

Using Eq. 1.35 and calculations for gluonic penguins, it can be shown that to first order, only P'_{tc}, T' and P'_{EW} should contribute. The decay amplitudes can then be written as[3]:

$$A^{+0} = -P'_{tc} \tag{1.36}$$

$$\sqrt{2}A^{0+} = -T'e^{i\gamma} + P'_{tc} - P'_{EW} \tag{1.37}$$

$$A^{-+} = -T'e^{i\gamma} + P'_{tc} \tag{1.38}$$

$$\sqrt{2}A^{00} = -P'_{tc} - P'_{EW} \tag{1.39}$$

The CP conjugated decay rates must also include a negation in the CP violating phase γ. Because P'_{EW} has no weak phase, the A_{cp} for A^{0+} and A^{-+} should be the same. Experimentally, it is shown that this is far from reality. The difference in A_{cp} ($\Delta A_{cp} = A^{0+} - A^{-+}$) is given by:

$$(\Delta A_{cp})_{exp} = (12.2 \pm 2.2)\% \tag{1.40}$$

The puzzle for direct CP Violation in this system is that this is a 5.5σ difference from zero.

1.3.3 Interference CP Violation in $B \to K^0\pi^0$

Because the neutral K_L has a large lifetime, it decays far from where it is produced and hence is very hard to detect. For this reason it is more experimentally far easier to look at the neutrally charged K_S meson and for which it is easier to measure its associated S_{cp} parameter. R. Fleischer goes into detail for these modes in references [12] and [6]. A relation bounding S_{cp} is shown to exist[12]:

$$S_{cp} = \sin(\phi_d - \phi_{00})\sqrt{1 - A_{cp}^2} \approx \sin(\phi_d - \phi_{00}) \tag{1.41}$$

This uses the mixing phase ϕ_d and a hadronic parameter ϕ_{00} given by $\phi_{00} = \arg \bar{A}_{00} A_{00}^*$. This is related to the angle ϕ_3. Given the small uncertainty in this angle, it is now possible to predict this with good accuracy and provide a precision prediction for S_{cp}. The approximation can be made since A_{cp} should be small.

Eq. 1.34 is also used to place limits on A_{cp} using the observed asymmetries in the other modes in the $B \to K\pi$ system. BaBar and Belle have made the following observations for the $B^0 \to K_s\pi^0$ mode. The results are around a 2σ discrepancy from the SM predictions.

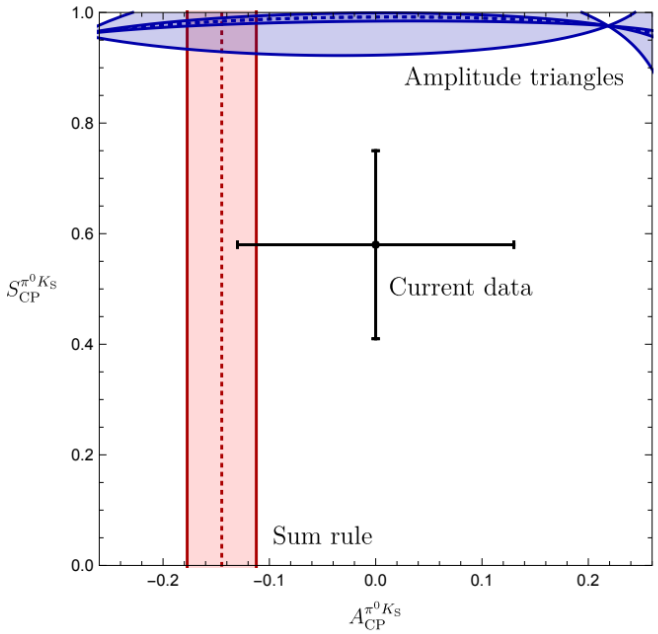

Figure 1.3: Current data and expected values for A_{cp} and S_{cp} in the $B^0 \to K_s\pi^0$ mode.

1.3.4 Possible Resolutions

The possible resolutions to this puzzle may be found in contributions from a new physics electroweak penguin with a different CP violating phase would significantly decrease the discrepancy[15].

There are currently three main ways to do this, corresponding to adding three beyond Standard Model particles - The leptoquark, the diquark, or the Z' boson[11]. The diquark solution would also solve another issue with the Standard Model by adding a mechanism for neutrinos to gain mass.

The other possibility is that this is a statistical anomaly. More data for this mode is needed to reduce the errors to show if this is the case or not.

Chapter 2

The Belle II Experiment

2.1 Introduction to Belle II

2.1.1 The SuperKEKB Accelerator

The Belle II detector is designed to look at the decays of B mesons. In order to do this it must be located at a "B-factory". This B-factory is the SuperKEKB accelerator located in Tsukuba, Japan. The SuperKEK accelerator is an electron-positron collider designed to accelerate electrons up to an energy of 7 GeV in its High Energy Ring (HER) and 4 GeV positrons in its Low Energy Ring (LER). These electrons and positrons are then collided at an Interaction Point (IP). The energies are chosen in order to produce an $\Upsilon(4S)$ resonance consisting of a b and a $\bar{b}$ quark. This will then either decay into $q\bar{q}$ or neutral and charged B-mesons ($\Upsilon(4S) \rightarrow B^0\bar{B}^0$ or $\Upsilon(4S) \rightarrow B^+B^-$) at approximately an equal rate. The asymmetry between the energies of the electron and the positron is designed to give a boost in momentum to the B mesons which enables the location of the B-meson decay point to infer the time of B-meson decay and hence makes it possible to make time-dependent CP-violation measurements.[19]

2.1.2 Comparison to Belle

The Belle II detector is designed to be an upgrade to the original Belle detector which collected data between 1999 to 2010 at the KEKB accelerator (the predecessor to SuperKEKB). The more significant upgrades from the original include:

- Improved spatial resolution due to the implementation of a silicon pixel detector.
- A larger volume of silicon strip detectors results in a higher efficiency of reconstruction for K_s to two charged pions.
- The implementation of new Particle Identification Devices (PID) in the barrel and endcap regions.
- Better electronics in the electromagnetic calorimeter which significantly reduce background noise.

Along with these upgrades to the detector, the change to SuperKEKB from KEKB will result in a 50 times increase in the integrated luminosity (from around 1ab^{-1} to around 50ab^{-1}). This will result in much lower statistical errors compared to the original Belle experiment. [1]

2.2 The Belle II Detector

The Belle II detector consists of several sub-detectors, each serving different purposes and having their own strengths and weaknesses. From closest to the IP to furthest, this includes the Pixel Vertex Detector (PXD), the Silicon Vertex Detector (SVD), the Central Drift Chamber (CDC), the Particle IDentification (PID) device, the Electromagnetic Calorimeter (ECL), and the K_L/Muon Detector (KLM).[1]

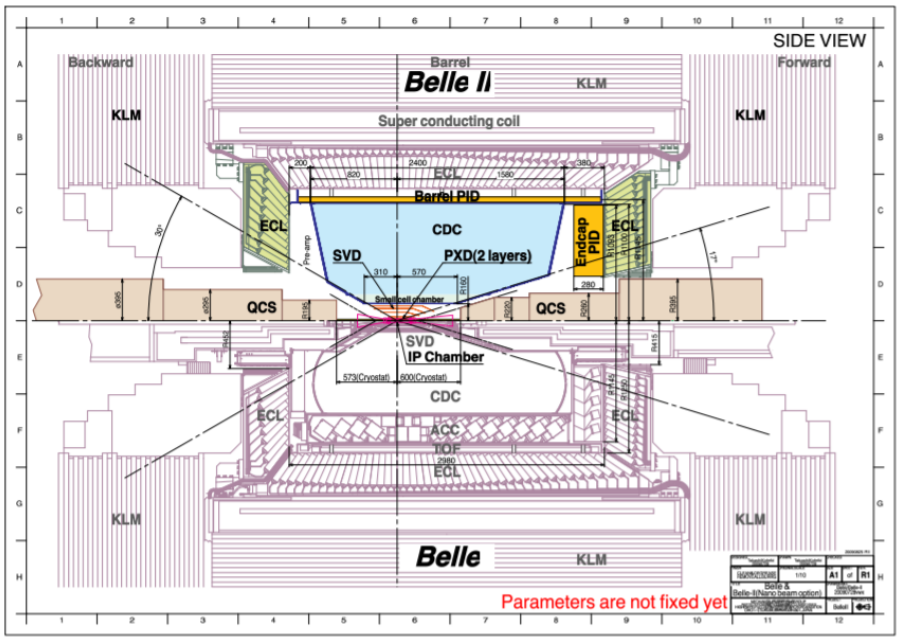

Figure 2.1: Side view of the Belle II detector (top) and the original Belle detector (bottom) showing the location of each sub-detector.

Pixel Vertex Detector (PXD)

The Pixel Vertex Detector (PXD) is a new addition to the Belle II detector from the original, and consists of two layers of very thin (50 micron) silicon sensors which utilize Depleted Field Effect Transistor (DEPFET) technology. These layers are at a radius of 14mm and 22mm from the interaction point. The PXD is required because the high levels of beam-related background causes a high hit rate which strip detectors (such as the SVD) could not handle. This detector allows for very precise vertexing of decays inside it.

Silicon Vertex Detector (SVD)

Similar to the PXD, the Silicon Vertex Detector (SVD) uses DEPFET technology, however the SVD uses sensors in strips instead of pixels. This detector consists of four layers and is used for vertexing at a radii of between 38mm and 140mm, and can extrapolate data obtained in the PXD and the Central Drift Chamber. The SVD is also particularly useful in vertexing decay channels involving D-mesons and τ-leptons

Central Drift Chamber (CDC)

The Central Drift Chamber (CDC) is located between a radii of 160mm and 1130mm from the IP. It is used for precisely measuring the tracks from charged particles along with their momentum. It is filled with Helium Methane gas mixture and a complex wire structure in order to perform a three-dimensional reconstruction and provide some particle identification information via energy loss within the gas volume.

Particle IDentification (PID)

The Particle IDentification (PID) device is primarily designed to determine whether a particle is a kaon or a pion. It consists of two devices located in the Barrel and EndCap regions of the detector. The Time of Propagation (TOP) detector is employed in the Barrel region. It consists of fused silica (quartz) to produce Cherenkov photons when a charged particle passes through. The photons are captured via total internal reflection and are directed to micro channel plate photon detectors which record the location and time of arrival of the photons. This information is employed to determine the cherenkov radiation cone and hence the velocity of the particle. This enables the determination of the pions or kaons. There is also a PID located in the forward endcaps which use Aerogel Rich Imaging Cherenkov (ARICH) detectors. Here, transparent Aerogel of refractive index near 1.05 is used to generate Cherenkov photons. These are used to discriminate between muons, pions, and electrons below 1 GeV/c momentum.

Electromagnetic Calorimeter (ECL)

The Electromagnetic Calorimeter uses Cesium Iodide (Thallium) (CsI(Tl)) as a scintillating material to create an 'electromagnetic shower'. This occurs via electromagnetic interactions of high energy photons and electrons or positions through pair production or bremsstrahlung. As this occurs it initiates a shower of electrons and positions which generate scintillation light in the CsI crystals which in turn is collected by photodiodes placed on the back of the crystals. The net result is a pulse from the electronics which is proportional to the energy depositied by ionising particles in the crystals. In the case of photons or electrons(positrons) this is close to their entire energy. The ECL is used for a variety of purposes including detecting photons and determining their energy, electron identification, and some K_L detection.

K_L & μ Detector (KLM)

The K-long and Muon Detector is the outermost detector and is used for detecting K_L and μ particles, both of which have a relatively long lifetime. It consists of alternating iron plates and detector elements. These iron plates also form the magnetic flux return for the superconducting solenoid. K_L mesons also are able to interact hadronically with these iron plates and hence leave a signal. This iron provides an extra 3.9 interaction lengths for the K_L mesons over the 0.8 interaction lengths provided by the ECL. They also act as a material which causes muons to lose energy through ionisation. This is required because muons are minimum ionising particles which have a very low loss of energy through matter. Because of this, it is unlikely that other types of particles (other than K_L and muons) will be detected in the KLM beyond the first few layers. The muons can be distinguished from the K_L mesons since muons are charged and the K_L meson is neutral. This means muons will create tracks in other parts of the detector while K_L mesons will remain invisible until interacting with the iron.

2.3 Event Variables

When reconstructing an event, the raw detector data can not be used directly. Instead, the raw data is used to determine the kinematics of possible decay trees back to the original B-meson (or $q\bar{q}$) and identifies the most likely scenario. Information about the reconstructed B-meson is then used along with information about the other reconstructed B-meson from the $\Upsilon(4S)$ to form useful variables - ΔT, ΔE and M_{bc}.

Time Difference (ΔT)

In order to look at the time-dependent CP parameters, it is necessary to look at the difference in decay time between both B-mesons. This will either be positive or negative depending on whether the B^0 or $\bar{B}^0$ decays first.

This can be determined because of the asymmetry in beam energies provides a momentum-boost to the B-meson in the z-direction. The change in the vertexed z-position ($z_{sig} - z_{tag}$) found from the vertex detector can therefore be converted into the difference in time between the decay of the B-mesons:

$$\Delta z = \beta\gamma c \Delta T \tag{2.1}$$

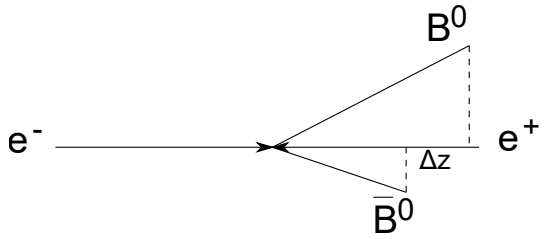

Figure 2.2: Δz is used to determine ΔT.

Energy Difference ΔE

In the Centre-of-Mass (COM) frame of the $\Upsilon(4S)$, both B-mesons should have the same energy (5.26 GeV) equal to half the energy of the $\Upsilon(4S)$ (called E_{beam}). This is then compared to the sum of the energies of each of the final state particles to give ΔE.

$$\Delta E = E_{beam} - \sum_i E_i \tag{2.2}$$

Here, the sum is over all final state particles. This should peak around zero for all decays involving a B-meson and events far from zero can be ignored.

Beam-Constrained Mass M_{bc}

The beam-constrained mass M_{bc} uses the momentum of the final state to compare to the $\Upsilon(4S)$ Centre-of-Mass B-meson energy. It is defined as:

$$M_{bc} = \sqrt{E_{beam}^2 - \left(\sum_i \vec{p}_i\right)^2} \tag{2.3}$$

This quantity should peak around the mass of the B-meson, $m_B = 5.279$ GeV.

2.4 Event Classification

For each electron-positron collision, there are a wide range of decay channels possible. Of these channels, there are only three which are of interest - one signal and two background channels.

- Signal - Any event where a B-meson decays into a K-short and a neutral pion. In this case, the B-meson of interest is labelled B_{sig} and the other labelled B_{tag}. The other B-meson is used for flavor tagging.
- Continuum - Any event where the electron and positron decay into quark-antiquark pairs ($e^+e^- \rightarrow q\bar{q}$). The only possible quarks which can be formed are u, d, c, and s, with c being the most common. These quarks will create 'hadronic showers' and various mesons will be formed.
- $B\bar{B}$ background - Any event where a neutral B-meson is formed but does not result in a signal event.

Of these, the continuum is expected to dominate.

Chapter 3

Analysis Strategy

3.1 Analysis Software

All simulation, reconstruction, and analysis tasks done throughout this chapter were completed using the Belle-II Analysis Framework (`basf2`)[21]. This framework is developed by the Belle-II organisation and uses steering scripts made using python. `release-05-01-13` is used throughout. The `ROOT`[5] toolkit developed for C++ by CERN was also used for fitting curves to data (using `MINUIT` [16]) and handling toy models. It was also used for continuum suppression when using TMVA (see section 3.4.2). `EvtGen` [24] was used to generate the MC signal. The majority of tasks were run on Spartan[22] (the University of Melbourne HPC) however the DESY grid system[13] was also used extensively.

3.2 Vertexing

Vertexing is the process of determining the location of the decay of each particle in an event and finding various kinematic quantities (such as the invariant mass, or momentum). For the $B^0 \to K_s\pi^0$ mode, the K_s will decay 67% of the time into a set of two charged pions (each of opposite charge). This is primary decay channel which can be used for fitting since the charged tracks of the pions can be used. The π^0 almost always decays into two photons which makes it almost useless for vertexing. The signal B-meson vertex is therefore found by extrapolating the K_s back towards the IP. This is known as imposing an IP constraint.

There are two techniques available for vertexing - `RAVE`, and `TreeFit`. `RAVE` is currently deprecated and is known to give slightly inferior results to `TreeFit`. Unfortunately, for technical reasons it is impossible using `TreeFit` to use only a subset of final state particles for vertexing. This functionality is necessary when applying the same technique to a control mode so `RAVE` must be used. There is currently work being done to make this feature available.

3.2.1 BTube

Because the beam spot in Belle-II is much smaller than the beam spot for Belle, it is possible for low life-time mesons (such as B-mesons or D-mesons) to decay outside of it. The technique used for finding the z positions for the tag-side B-meson must therefore be modified to account for this. This is done by introducing the BTube constraint. This constraint looks at the signal-side B-meson and uses energy-momentum conservation to form a 'tube' from the IP where the tag-side B-meson vertex should be (Figure. 3.1). The tag-side B-meson is constrained to this tube instead of to the IP. Using this method should significantly reduce errors in tag-side vertexing which should result in a better resolution function (DeltaT distribution).

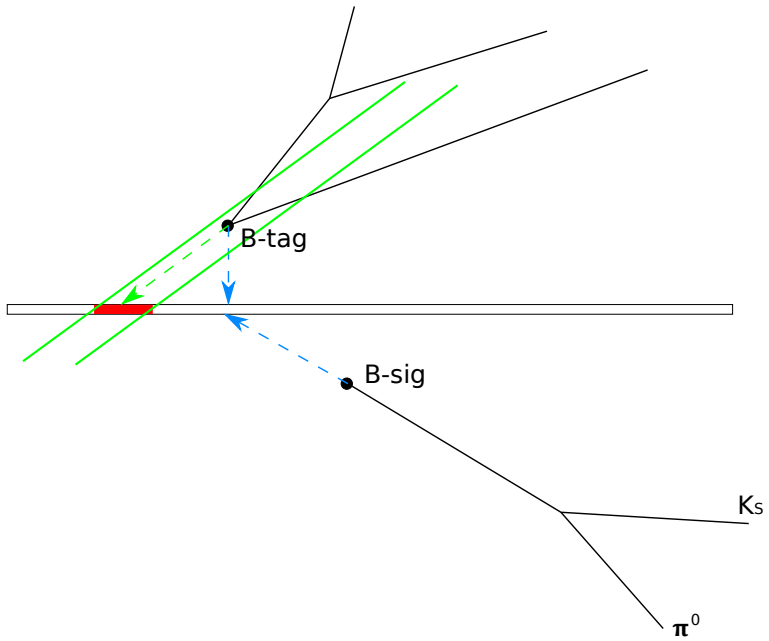

Figure 3.1: Mechanism of BTube: The signal-side B-meson is extrapolated back to the beam, and instead of placing the B-meson there (shown with blue arrows), a tube is formed (shown in green) tracing back to the actual IP.

3.3 Flavour Tagging

In order to obtain CP-violating parameters for this decay mode, it is necessary to determine the flavour of the B-meson before it decays into the K_S and π^0. It is impossible to directly determine the flavour of the signal-side B-meson because both the B^0 and $\bar{B}^0$ decay into the same final state. Because the B^0 and $\bar{B}^0$ are produced as a quantum entangled state, finding the flavour of the tag-side B-meson allows determination of the flavour of the signal side B-meson. This can be done when the tag-side B-meson decays semileptonically - For example $\bar{B}^0 \to D^+(\to K^-\pi^+\pi^+)\mu^-\nu$. In this case the charge of the lepton and the kaon are used to determine the flavour of the B-meson (Figure. 3.2). `basf2` provides a mechanism for this flavour-tagging which gives the expected flavour q, and a factor r, which is the probability of this flavour being correct. These are then multiplied together to give a new variable qr.

This variable is required to model the signal DeltaT however it must first be converted into the 'mistag fraction' (w). The mistag fraction, as the name suggests, is just the fraction of events that are tagged incorrectly. This can be straightforwardly obtained from qr using Eq. 3.1.

$$w = \frac{1}{2}\left(1 - |q.r|\right) \tag{3.1}$$

In line with the Belle-II flavour tagging guidelines, this variable is broken up into seven bins - the first of which has almost no information about the flavour and the last having a high confidence in it. The bin edges are outlined in Table. 3.1.

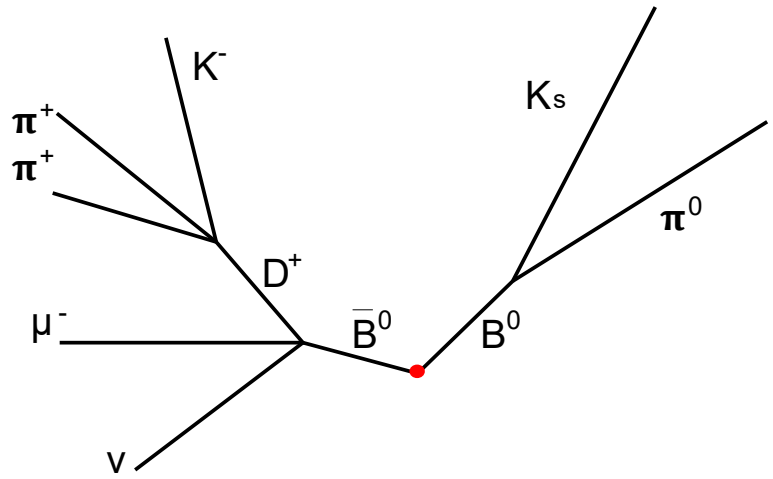

Figure 3.2: Example decay of $\bar{B}^0 \to D^+(\to K^-\pi^+\pi^+)\mu^-\nu$. The charge of the lepton and kaon are used to determine the flavour of the tag-side B-meson. The flavour of the signal-side B-meson can then be inferred.

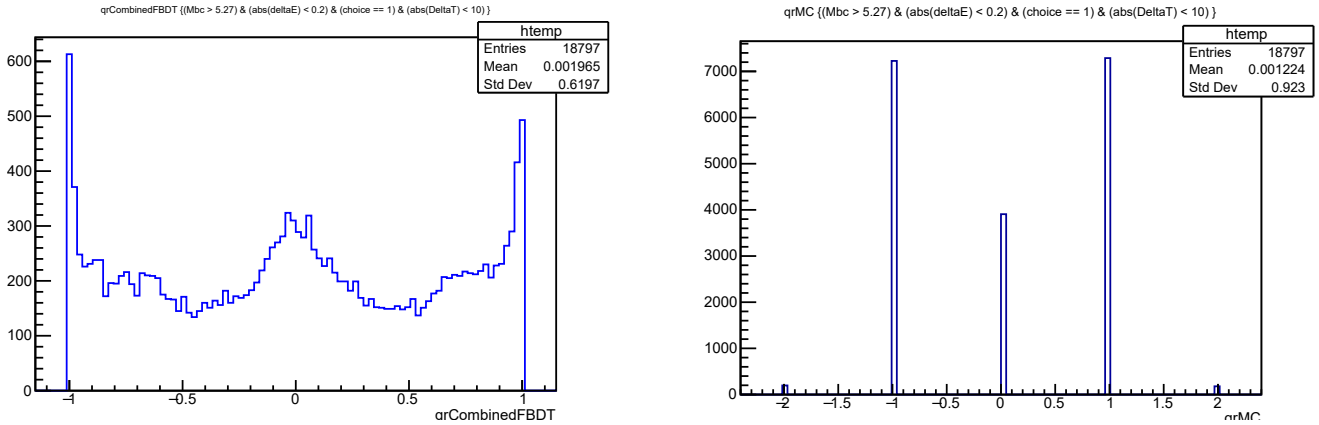

Figure 3.3: The left image shows a histogram of the output of the flavour-tagging on the MC signal, and the right image shows the true value of qr.

	Edge 1	Edge 2	Edge 3	Edge 4	Edge 5	Edge 6	Edge 7	Edge 8
qr	0.0	0.1	0.25	0.5	0.625	0.75	0.875	1.0
w	0.5	0.45	0.375	0.25	0.1875	0.125	0.0624	0.0

Table 3.1: The eight bin edges describing the seven bins for both qr and w.

The difference in bin sizes is because at high values of qr (or low values of w), there are many more events than in the low (high) valued bins. Another useful variable is the difference in the mistag rates between B^0 and $\bar{B}^0$ for the i'th bin. This is called Δw_i.

$$\Delta w_i = w_i(B^0) - w_i(\bar{B}^0) \tag{3.2}$$

This variable is also used to model DeltaT.

3.4 Continuum Suppression

To get fits to data to be as accurate as possible, there must be as many continuum events removed while still maintaining enough signal to produce good results. This process is called

continuum suppression. Continuum suppression is done using multivariate analysis techniques trained on variables describing the geometry of an event.

The main distinguishing feature between signal and continuum events is that signal events produce a much more spherical shape when the B-mesons decay into its daughter particles. Continuum events usually produce 'jet-like' shapes where the daughter particles are grouped together much more tightly (Figure. 3.12).

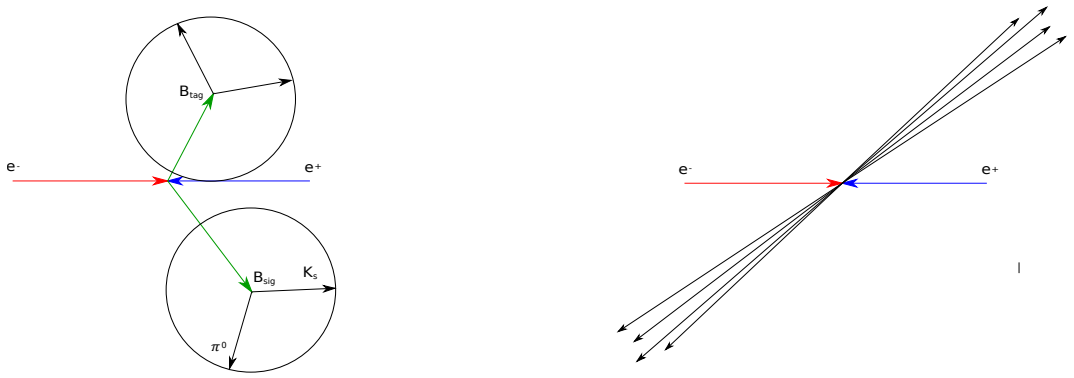

Figure 3.4: Comparison between the shape of the signal events (left) and the continuum events (right).

3.4.1 Training Variables

The set of training variables is given here. There are many that have also been excluded due to having a high correlation with other variables.

- `cosTBz`
- `thrustOm`
- `KSFW_[mm2, et]`
- `KSFW_hso[00, 04, 10, 14, 20, 22, 24]`
- `KSFW_hoo[0, 1, 2, 3, 4]`
- `CleoCone_[4, 5, 6, 7, 8, 9]`
- `cosMisMom`
- `cmsCosTheta`

Thrust Variables

The thrust T is obtained by summing the projection of the momentum of all the daughter particles from either B_{sig} or B_{tag} and normalizing it to the sum of all the daughter particle's momentum (Eq. 3.3).

$$T = \frac{\sum_i |\hat{n} \cdot \vec{p_i}|}{\sum_i |\vec{p_i}|} \tag{3.3}$$

This value is then maximized by finding the vector $\hat{n}$ that gives the highest T. The vector $\hat{n}$ is called the thrust axis.

The `thrustOm` variable is the thrust obtained using the daughter particles from the tag B-meson. Another variable (called `thrustBm`) describes the thrust for the signal B-meson however it is highly correlated with other variables.

`cosTBz` is the cosine of the angle between the thrust axis for B_{tag} and the forward direction of the beampipe. There is similarly another correlated variable `cosTBTO` which describes the same thing for B_{sig}.

Kakuno-Super-Fox-Wolfram Moments

The Kakuno-Super-Fox-Wolfram (KSFW) moments were designed by Belle to discriminate between the shapes of signal and continuum events. There are three main types of KSFW moments - signal, tag, and missing momentum moments.[4]

The signal moments are given by:

$$H_{xl}^{so} = \sum_{i} \sum_{jx} |\vec{p}_{jx}| \, P_l(\cos\theta_{i,jx}) \tag{3.4}$$

In this equation, i and j sum over the signal-side and tag-side tracks respectively. x describes the charge of the tag-side particle (0 for charged, 1 for neutral, and 2 for missing). For odd l these terms vanish. $P_l(\cos(\theta))$ are the Legendre polynomials.

For the tag-side KSFW moments, l is split into both even and odd moments.
The even l terms are given by:

$$H_{l}^{oo} = \sum_{i} \sum_{j} |\vec{p}_i| \, |\vec{p}_j| \, P_l(\cos(\theta_{ij})) \tag{3.5}$$

And the odd l terms are given by:

$$H_{l}^{oo} = \sum_{i} \sum_{j} Q_i Q_j \, |\vec{p}_i| \, |\vec{p}_j| \, P_l(\cos(\theta_{ij})) \tag{3.6}$$

The last type of KSFW moment is the missing momentum moments. This is simply the difference between the expected mass and the reconstructed mass.

$$M_{miss}^2 = \left(E_{\Upsilon(4S)} - \sum_{n=1}^{N_t} E_n \right)^2 - \sum_{n=1}^{N_t} |p_n|^2 \tag{3.7}$$

N_t is the total number of daughter particles.
The last variable is the transverse energy. This is given by:

$$E_t = \sum_{n=1}^{N_t} (|p|_t)_n \tag{3.8}$$

Cleo Cones

Cleo cones use the jet-like shape of the continuum events and the spherical shape of the signal events to form nine discriminating variables. It does this by forming cones around the B-mesons thrust axis at intervals of 10°. It then transforms these cones and looks at the momentum flow.

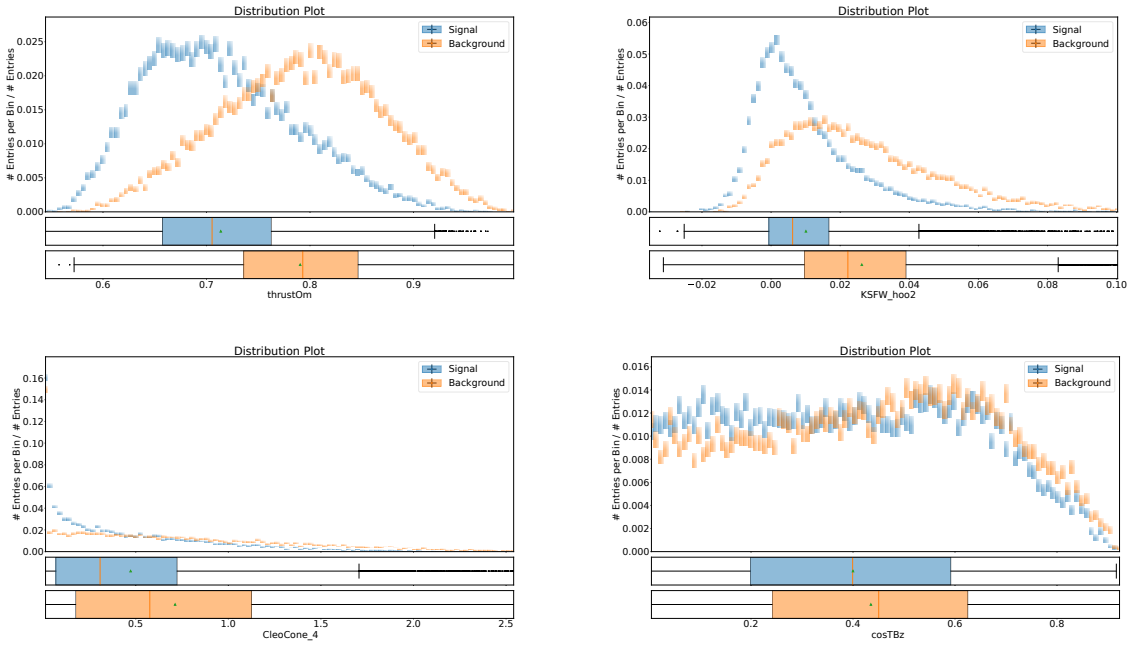

Figure 3.5: Some of the variables used for training in continuum suppression.

3.4.2 Method of Continuum Suppression

Continuum Suppression is achieved using a Multi-Variate Analysis (MVA). This consists of looking at all of the training variables for each event and training an algorithm to decide whether another arbitrary event is signal (or BB background) or continuum. When this algorithm is applied to independent events it creates another variable based on how confident it is. To train this algorithm, a dataset was divided in half, with one half being used for training and the other used for testing.

Figure of Merits

When the new continuum suppression variable is created, a cut must be applied to remove as much continuum as possible while maintaining a good amount of signal. This is done by maximising the 'Figure of Merit' (FOM). Doing this provides a good payoff between signal removed and continuum kept. The FOM is given by Eq. 3.9.

$$FOM = \frac{S}{\sqrt{S+B}} \tag{3.9}$$

TMVA & FastBDT (FBDT)

In this analysis, there were two toolkits used for training and applying continuum suppression - The Toolkit for Multivariate Analysis (TMVA) and FastBDT (FBDT) [17]. Both of these use a Boosted Decision Tree (BDT) method. This consists of training a set of decisions - each with certain weights, and layering more weighted decisions for each outcome. The outcome of this boosted decision tree is the probability of the event being signal or continuum. Both TMVA and FBDT were compared against each other, and for FBDT, two different sets of hyper-parameters were also tested (see below). The technique that maximized the FOM is the one that was used.

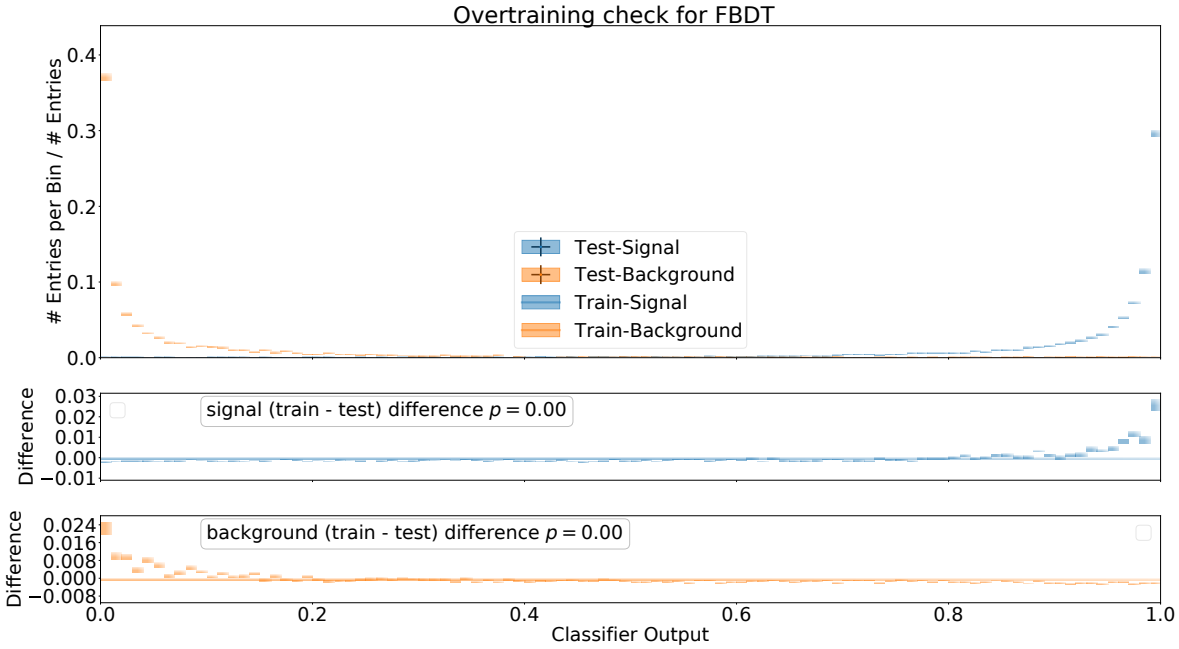

Figure 3.6: Output of the continuum suppression on the train and test data from FBDT.

The two sets of hyper-parameters tested were the default set, and an optimized set found by David Porter (Undergraduate Research Project, University Melbourne via Private Communication). This looks at the difference between the values of `nTrees`, `nLevels`, and `nCuts`. The values used are given in table 3.3.

Parameter Set	nTrees	nCuts	nLevels
Default	100	8	3
Optimized	400	8	6

Table 3.2: The two sets of hyper-parameters used for training with FBDT.

For each training method, the FOM was found using the expected number of events in $100 fb^{-1}$ of MC data - 117 signal events and 94,198 continuum events (see section 5.1). The result is that the use of FBDT using the optimized set of hyper-parameters results in a significantly higher FOM than the other two.

Method	Max. FOM	MVA Cut	Signal Events	Continuum Events
TMVA	2.58	0.14	36	163
FBDT (default)	2.86	0.92	49	245
FBDT(optimized)	3.54	0.93	61	236

Table 3.3: Results for continuum suppression using different methods. Note that the MVA cut for TMVA is much lower than the other two due to this method producing an output between negative one and positive one, whereas FBDT produces an output between zero and one.

Logarithmic Transform

In order to use the continuum suppression output as a fitting variable, it is convenient to define a transformed variable to create a Gaussian shape. This transform uses the maximum value of the MVA output (M), the intended cut value (R), and output of the continuum suppression (y). The logarithmic transform (y') is defined by Eq. 3.10.

$$y' = \log\left(\frac{y - R}{M - y}\right) \tag{3.10}$$

Both the original and transformed continuum suppression variables for the MC signal are shown below (after the cut).

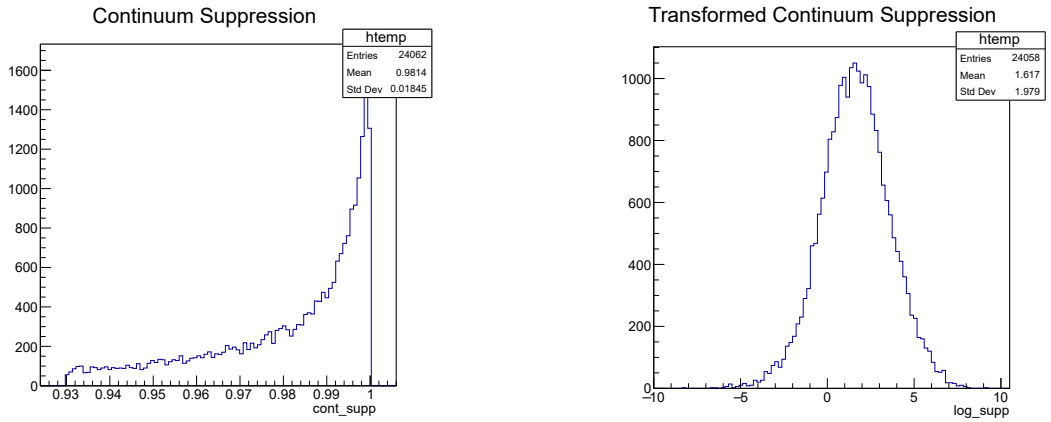

Figure 3.7: MVA output and the logarithmic transform.

3.5 Fitting

To fit to the yields, Acp, and Scp of a set of data containing signal, continuum, and BB background events, a set of Probability Density Functions (PDF's) is required for each of the four variables (DeltaE, Mbc, logSupp, DeltaT) for each type of event. For the signal case, fitting to each of these in the different bins of qr was tested against using the same PDF for every bin. The initial value of Scp used in the signal DeltaT PDF was also optimized.

3.5.1 DeltaT Probability Density Function

The most complicated PDF used was for the signal DeltaT. This PDF consists of a physical model convoluted with the sum of two Gaussians representing the spread in DeltaT from its errors. The physical model accounts for the different CP violation parameters and the mistag values. Because this contains information about the mistag information, it is fit simultaneously to each bin of qr resulting in seven different PDF's using the same parameters (unless testing using different ones) except for the mistag data. This results in the low bins of qr showing significantly less effects of CP violation compared to the higher bins.

The physical PDF is as follows:

$$\begin{aligned} f_i(\Delta T, q) = & \frac{e^{-|\Delta T|/\tau}}{4\tau}([1 - q\Delta w_i + q\mu_i(1 - 2w_i)] + [q(1 - 2w_i) + \mu_i(1 - q\Delta w_i)]\times \\ & [S_{cp}\sin(\Delta m\Delta T) + A_{cp}\cos(\Delta m\Delta T)]) \end{aligned} \tag{3.11}$$

Here, w_i and Δw_i are the mistag data introduced in Section 3.3, μ_i is the number of events in each bin, q is the charge, and Δm is described by Eq. 1.14. τ is the known lifetime of the neutral B-meson (1.52 ps).

The full PDF is given by the convolution of this with the sum of two Gaussians labelled R:

$$F(\Delta T, q) = f_i(\Delta T, q) \otimes R \tag{3.12}$$

It was found (Section 5.4.1) that the optimal value of Scp to initially fit to was zero. Neither Scp nor Acp are floated while doing the initial fits.

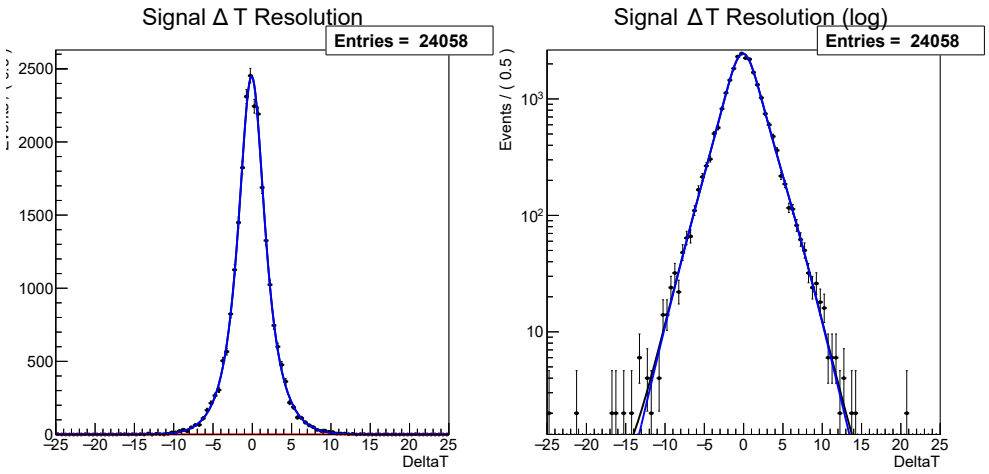

Figure 3.8: Fits to DeltaT on MC signal data with an Scp of zero shown on linear (left) and log (right) scales.

3.5.2 All Fits

Table 3.4 outlines the shapes used for fitting each variable for each event type.

Event Type	Mbc	DeltaE	logSupp	DeltaT
Signal	2 Gaussians	Crystal Ball Gaussian	3 Gaussians	Physical PDF 2 Gaussians
BB Background	Argus Gaussian	Chebychev Gaussian	2 Gaussians	Physical PDF (unbinned) 1 Gaussians
Continuum	Argus Gaussian	Chebychev	3 Gaussians	Physical PDF (unbinned) 1 Gaussians

Table 3.4: The composition of the fits for each variable for each event type.

The physical PDF's for the BB background and continuum DeltaT are much simpler than for the signal and do not account for flavour tagging, Scp, or Acp. Both types of continuum

were fit to $100 fb^{-1}$ of data after skimming and continuum suppression.

The results of the fits are shown below:

Signal

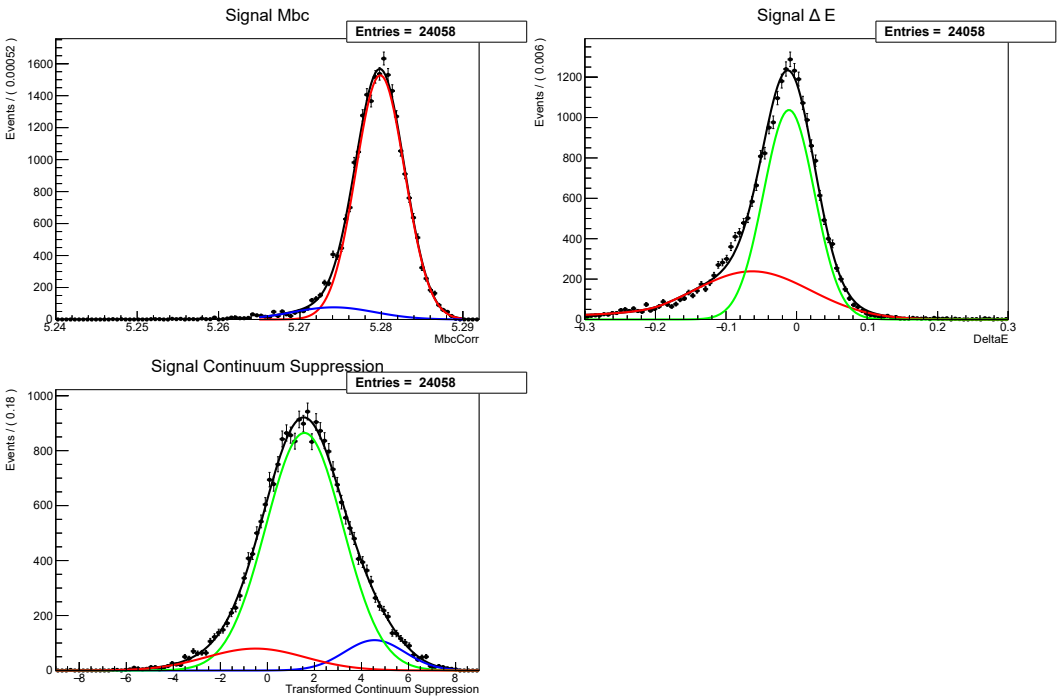

Figure 3.9: Fits to the signal PDF's except DeltaT.

Continuum

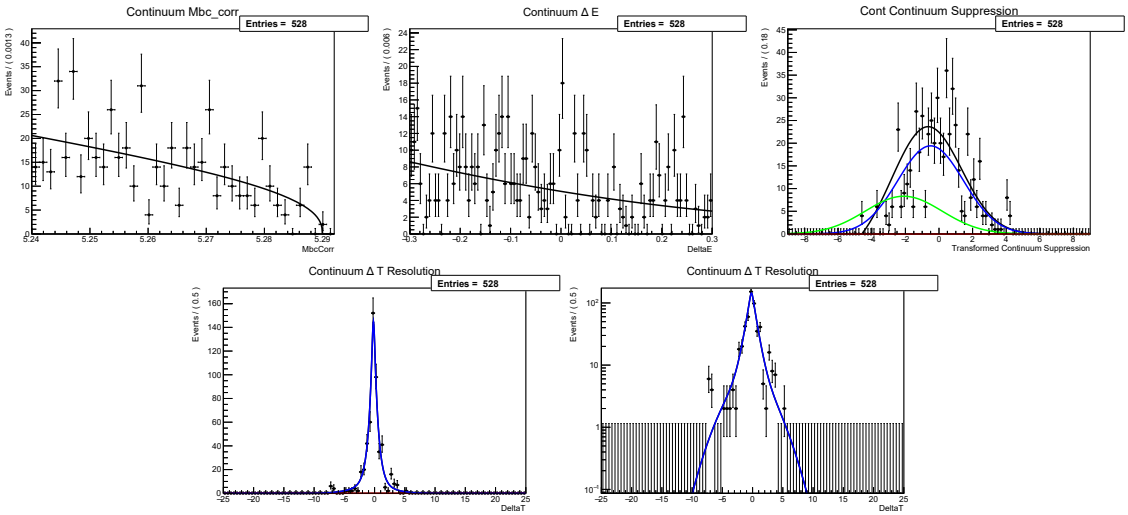

Figure 3.10: Fits to the continuum PDF's.

BB Background

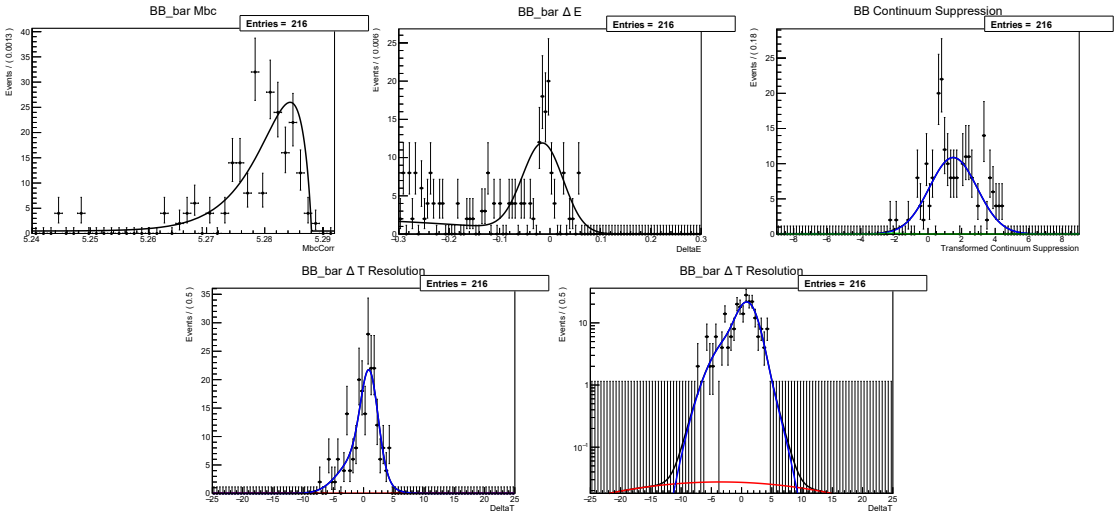

Figure 3.11: Fits to the BB background PDF's.

3.5.3 Signal DeltaT Fits with Flavour Tagging

As stated in Section 3.5.1, there is a set of PDF's for the signal DeltaT - One for each bin of qr. Since the bins of qr tell how good the information about flavour tagging is, the low bins would be almost randomly assigned a flavour and the high bins would be assigned a flavour with very good confidence. This results in each PDF for the signal DeltaT showing different levels of CP violation.

This can be seen by plotting DeltaT with cuts on the assigned flavour.

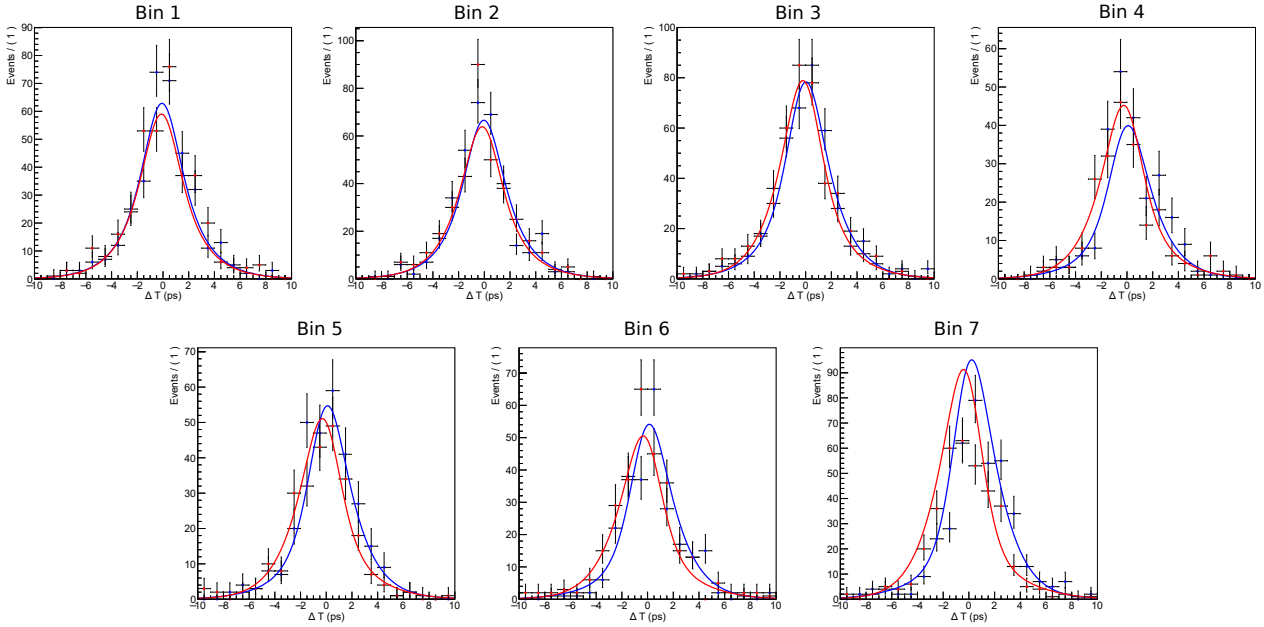

Figure 3.12: PDF's for each bin of DeltaT normalised to the number of events in each. it is apparent that the PDF's for low bins are almost identical for both flavours however the high bins show a large seperation between the two. These plots were generated using MC signal data with a value of Scp of 0.6.

Chapter 4

Studies in DeltaT and DeltaT Errors

4.1 The DeltaT Error distribution

To optimize the results for DeltaT and hence lower the errors in the time dependent CP parameters (Scp), it is important to understand how the distribution of DeltaT Errors (DeltaTErr) is formed. It is also essential in order to verify that a chosen control mode is acceptable.

The main features of the distribution are the peaks at around 0.5, 1, 1.2, 4.5 and 10. The main goal of this chapter is to identify the cause of each relevant peak and to find an appropriate cut.

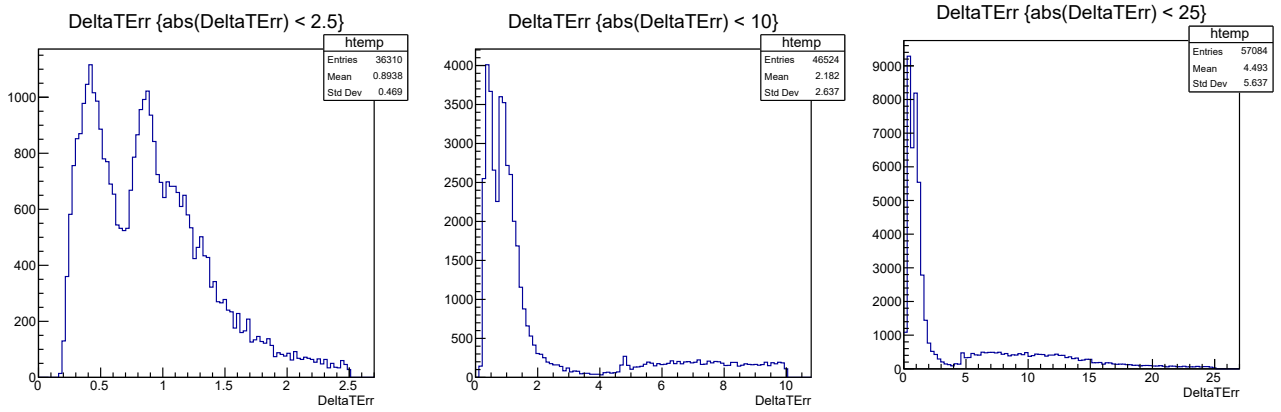

Figure 4.1: Distributions for Delta T Errors cut at 2.5 (left), 10 (centre), and 25 (right). There are five clear peaks below DeltaT errors of around 20.

The neutral B-meson only has a lifetime of around 1.52 ps. This means that all entries with errors above around four contain almost no information about DeltaT. This means that a cut can be established somewhere between two and four, removing the last two peaks entirely. Unfortunately, this also means cutting almost 40% of events. This cut in events will result in lower statistics for the yield and Acp so the errors for these values will be higher compared to a time independent analysis.

4.2 The K_s vertex location

4.2.1 The PXD

The first theory relating to the cause of the first peak is whether the K_s decays inside the PXD or not. This is relevant because the outer layer of the PXD is 22mm from the IP and the average distance it takes for the K_s meson to decay is very similar. This means that it should be a common occurrence for the K_s vertex to decay both inside or outside this layer. Because the K_s meson is not charged, it will not produce any tracks until it decays into a π^+ and a π^-. If the K_s decays outside the PXD, the PXD will not provide any vertexing information.

This theory was tested by taking cuts on the radius of the K_s to either be inside or outside the PXD. It was found that this cut was very effective at isolating the first peak and so this was determined to be the cause.

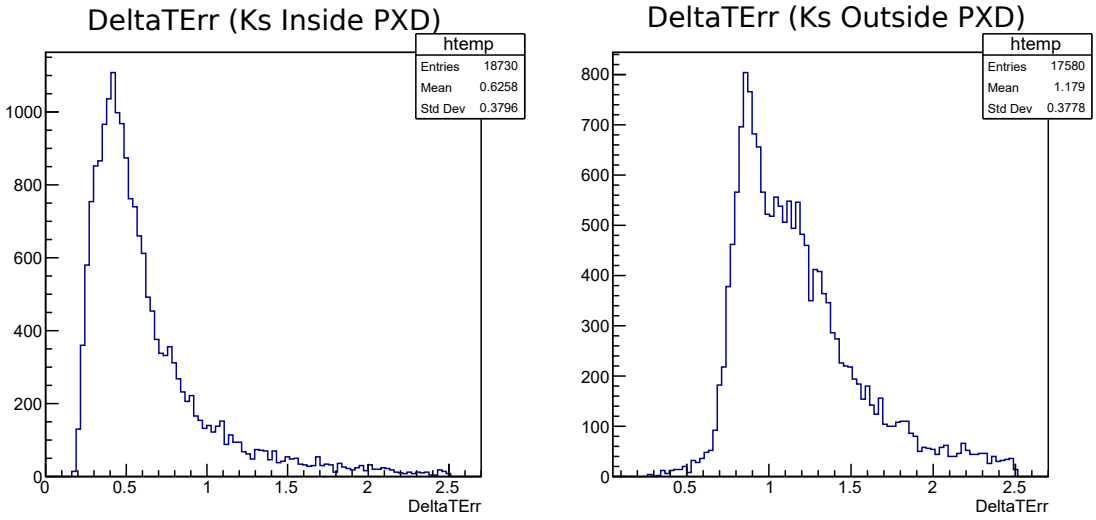

Figure 4.2: Distributions for DeltaT errors after cutting the Ks vertex to be inside (left) and outside (right) of the PXD. The peak at around 0.5 is entirely removed when the K_s vertex is outside and the peak at around 1.0 is similarly removed when the K_s decays within the PXD.

The effect can also be confirmed by looking at the location of the K_s vertex while taking cuts cuts on DeltaT errors (Figure. 4.4). This shows that the majority of events with a DeltaT error less than 2.5 have a K_s vertex within the PXD. There are still some events visible that decay outside the PXD visible due to the tail end of the second peak, however this number is very small compared to the number of events within the PXD.

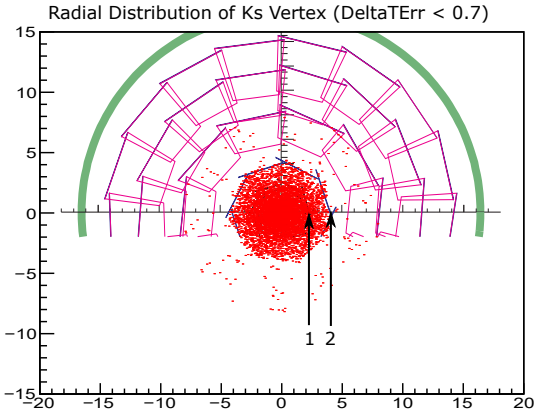

Figure 4.3: The distribution of the K_s vertex X-Y location overlayed on the layout of the Belle-II detector (end-on). While the vast majority of events occur within the PXD (labelled 1), due to the high number of events there are still many events outside the PXD but before the SVD (labelled 2). All events with DeltaT errors above 0.7 were cut to try and isolate the first peak.

4.2.2 Number of SVD Hits

As can be seen in Figure. 4.1, the distributions from the first three peaks and the fourth peak have very little overlap. This means that finding the cause of the last two peaks, and cutting a variable based on that should entirely remove the fourth and fifth peaks. This is equivalent to finding a value for DeltaT errors and cutting all events above that.

The variable found to be the cause of these peaks is the number of hits in the SVD. In order for the SVD to be able to reconstruct good tracks for the K_s vertex, there must be at least two hits for each charged pion in the silicon detector. This is because it is impossible to form a straight line from only one point. In order for there to be two hits in the SVD, the K_s must decay before the fifth layer of the VXD (both the PXD and the SVD). This is shown in Figure. 4.4.

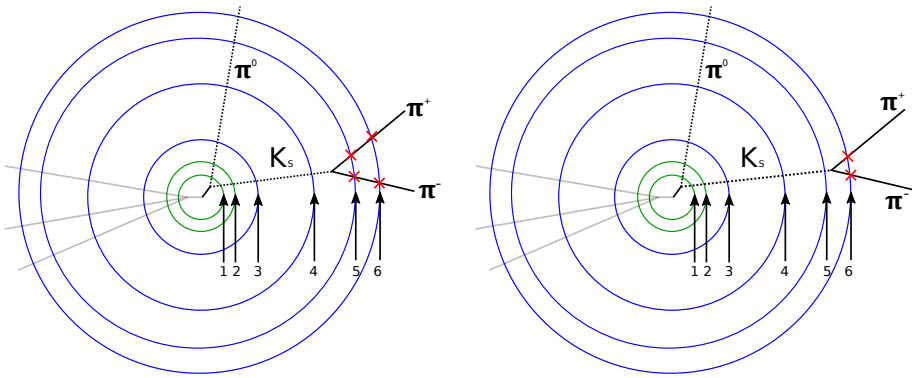

Figure 4.4: Example of how a decay might look on the X-Y plane (end-on) inside the VXD (each layer numbered 1-6). The gray tracks represent the tag-side B-meson decay, and the black represents the signal-side. The dashed lines show the invisible π^0 and K_s tracks. The solid black lines are the tracks from the visible charged pions. The hits on the VXD are shown with a red 'x', with the left image showing two hits per pion and the right showing only one hit per pion.

The difference between no cut, and taking a cut on the K_s vertex being within the fifth layer of the VXD is shown in Figure. 4.5. While there is still a very small peak at around 4, the majority of events with DeltaT errors above around 2.5 are cut. This means that taking the cut on the K_s vertex location to be before the fifth VXD layer is almost equivalent to cutting all events with DeltaT errors above 2.5. The fifth peak in the DeltaT error distribution must therefore be due to vertexing using the CDC.

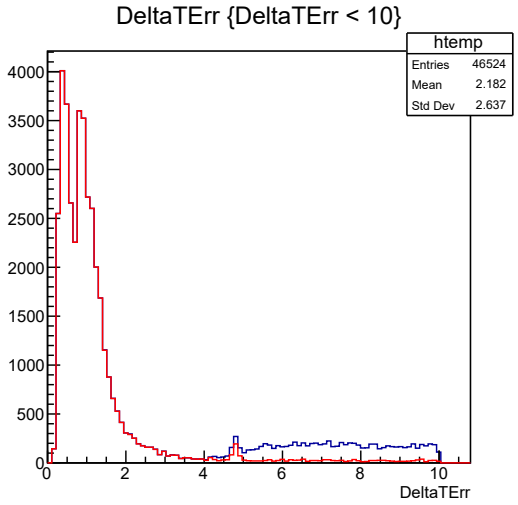

Figure 4.5: Distribution of DeltaT errors with no cut (blue), compared to with a cut on the K_s vertex location (red).

Again, this can be further illustrated by looking at the distribution of the K_s vertex while placing a cut on DeltaT errors at 2.5 (Figure 4.6).

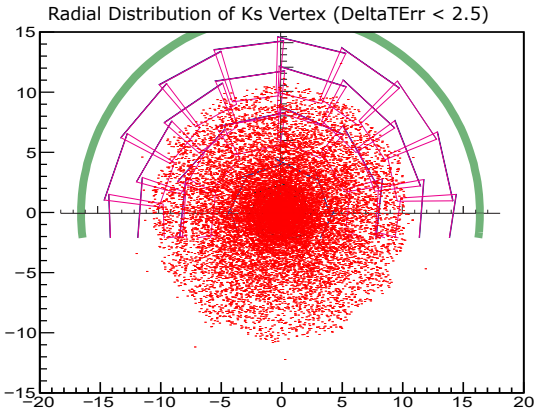

Figure 4.6: Distribution of K_s vertex X-Y positions with a cut on DeltaT errors at 2.5. Almost all events are within the fifth layer of the VXD.

Figure 4.6 shows the distribution stops around 1cm before the fifth layer of the VXD. This could be because the K_s's that decay close to it could potentially decay into two charged pions which then go through the same SVD strip. This would result in less information from the SVD and hence a higher error in vertexing. This could also be the cause of the small third peak seen in the DeltaT distribution although this has not been tested.

4.3 DeltaT Residues

When using Monte-Carlo generated signal data, it is important to make sure there are no correlations between variables and their corresponding MC truth value, as this would result in a bias in the shape of its distribution. In this section, the correlation between DeltaT and its MC truth values is addressed.

To look at the correlations, the DeltaT residue is defined ($\delta\Delta T$). This is equal to the difference between DeltaT and its true value:

$$\delta\Delta T = \Delta T - \Delta T_{MC} \tag{4.1}$$

To check for correlations, the distribution of this variable is plotted taking cuts between different values of DeltaT. Ideally, each distribution would be identical.

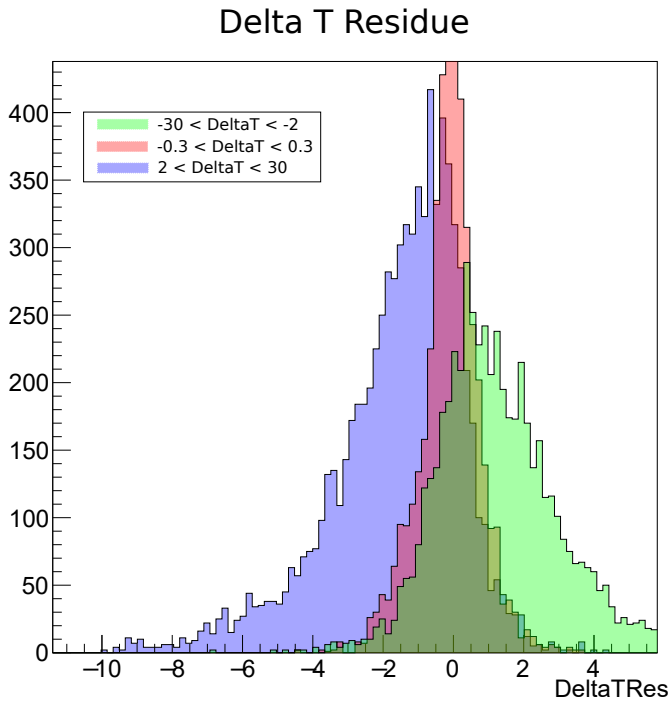

Figure 4.7: Distribution of the DeltaT Residue taking cuts of different values of DeltaT.

As can be seen from Figure. 4.7, the distributions of this variable are not the same when different cuts of DeltaT is taken.

It turns out that the difference in distribution shapes is not actually an issue, since the MC DeltaT is formed from the true DeltaT distribution. This forms a PDF without the convolution with the two Gaussians. The convolution with these Gaussians spreads the distribution out slightly pushing the negative values to be more negative and the positive values to be more positive. This accounts for the difference in shapes. What is not expected is the difference in the number of events in each side region (the blue and green areas). The number of events in the blue region is 7828 and the number of events in the green area is 4389. The explanation for this is currently unknown.

Chapter 5

Validation

Now that the model has been created, it must be tested. To do this, a data set will be created that simulates what should be found in the real data. This involves finding the number of each type of event that should be expected and adding it all into a single data set. Since there are many more signal events generated than could be found in the real data (at this point), a random selection of these can be added to the mix. For the background however, only 100fb^{-1} was used. This means that a random sample can not have enough statistical variance to be useful. The solution to this is to use the PDF models for them to create new events with correct distributions. It is important that not only are the PDF's are correct, but also the the fraction of events in each bin of qr is maintained. Doing this will guarantee a statistically good sample of background.

The goal is to fit to the signal and background yields, and ultimately the values of Acp and Scp. These can be specified when the signal is generated. Describe here is how data sets of both 62.8fb$^-$1 and 1ab^{-1} are simulated.

A control mode will then be used with the same analysis strategy in order to confirm it works properly before unblinding the signal mode.

5.1 Expected Number of Events

5.1.1 Signal

The number of signal events expected is given by the reconstruction efficiency (ϵ) multiplied by the branching ratio and the number of $B^0\bar{B}^0$ events. This includes cuts of $Mbc > 0.52$ and $|DeltaE| < 0.2$. The number of $B^0\bar{B}^0$ events is equal to the cross-section of $B^0\bar{B}^0$ events times the integrated luminosity. Belle-II expects that this cross-section is equal to 1.1nb. The expected number of signal events is then given by Eq. 5.1.

$$N_{sig} = \epsilon \times Br(B^0 \to K_s\pi^0) \times 1.1nb \times \mathcal{L} \tag{5.1}$$

The efficiency can be obtained by looking at how many of the events generated survived after all the cuts and reconstruction. Here, 100,000 events were generated and 24,062 survived. This gives $\epsilon = 0.24$. The branching ratio is half that of $B^0 \to K^0\pi^0$ equal to 4.45×10^{-6} resulting in the expected number of signal events to be 1175 events per ab^{-1} before continuum suppression (see Section. 3.4.2).

$$N_{sig}(1\text{ab}^{-1}) = 0.24 \times 4.45 \times 10^{-6} \times 1.1 \times 10^9 = 1175 \tag{5.2}$$

5.1.2 Background

100fb^{-1} of data was used for both the continuum and BB background. From this data set, 94,198 continuum events and 498 BB background events survived. After continuum suppression this is down to 236 continuum events and 23 BB background events. This is scaled up to 1ab^{-1} to be 2360 continuum and 230 BB background events per ab^{-1}. The expected number of events for each relevant data size is shown in table 5.1.

	Signal	Continuum	BB Background
5ab^{-1}	3050	11800	1150
1ab^{-1}	610	2360	230
100fb^{-1}	61	236	23
62.8fb^{-1}	38	148	14

Table 5.1: Numbers of events in various data sizes.

5.2 Toy Single Fits

In order to verify the analysis technique, a dataset has been generated using the signal events and sampled background events. This was then fit to the yields, Scp, and Acp. This section shows the results from a single fit to these values for both 1ab^{-1} and 62.8fb^{-1}. The input values of Scp and Acp are 0.6, and 0.0 respectively. There values were chosen since they are close to the known values.

1ab^{-1} Results

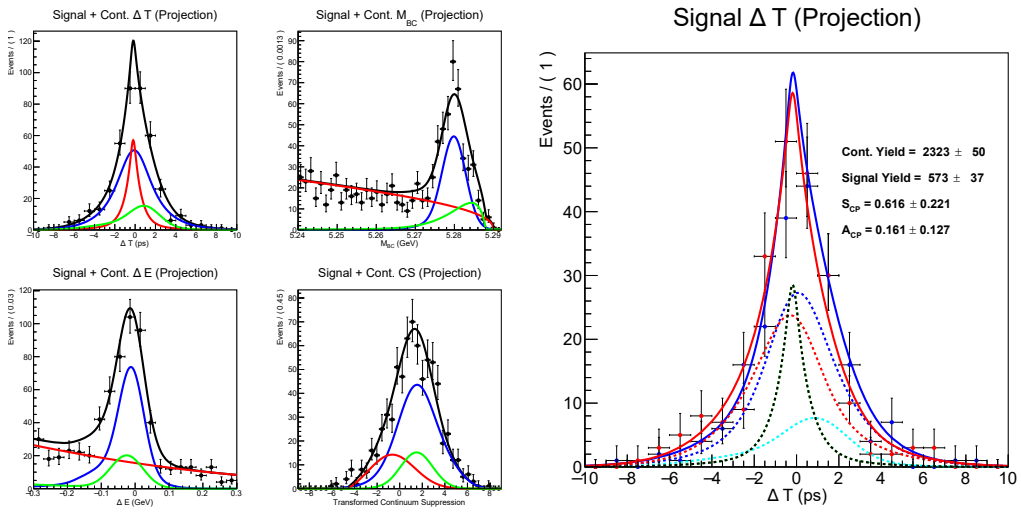

Figure 5.1: Fits to a synthesized data set expected from 1ab^{-1} of integrated luminosity.

62.8fb^{-1} Results

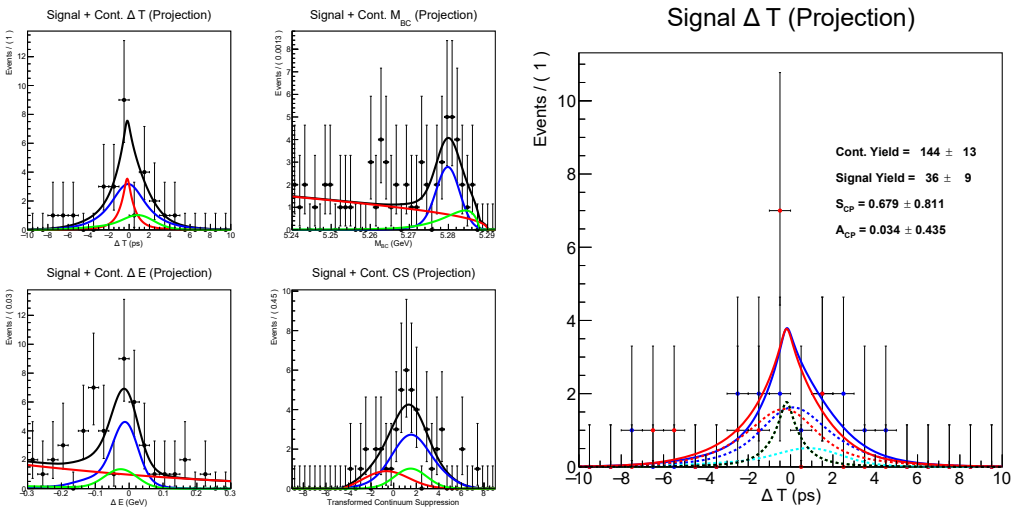

Figure 5.2: Fits to a synthesized data set expected from 62.8fb^{-1} of integrated luminosity.

Figures. 5.1 and 5.2 show both the fits to the yield (left) and the flavour integrated DeltaT fit (right). On the left, the signal is blue, the continuum is red, and the BB background is green. On the right, the B^0's are blue and the $\bar{B}^0$'s are red. The continuum is green and the BB background is cyan.

5.3 MC Toy Results

Because a single fit is not necessarily indicative of how good the model is, these single toys are repeated 400 times and the results for all the yields, Acp, and Scp are kept along with their respective errors. These are then displayed in a histogram and fit to a Gaussian. The mean of these Gaussians are then used to compare the success of different models - Ideally, each of these means will match the input data. The pull is also shown which shows how well the model compares to the input, with the mean of the pull showing on average how many standard deviations away the acquired mean is from the expected. A perfect model would have a pull with mean of zero, and a standard deviation of one.

As with the single fits, the input values are an Acp and Scp of 0.0 and 0.6 respectively, and yields as outlined in table 5.1.

5.3.1 1ab^{-1} Toy MC Results

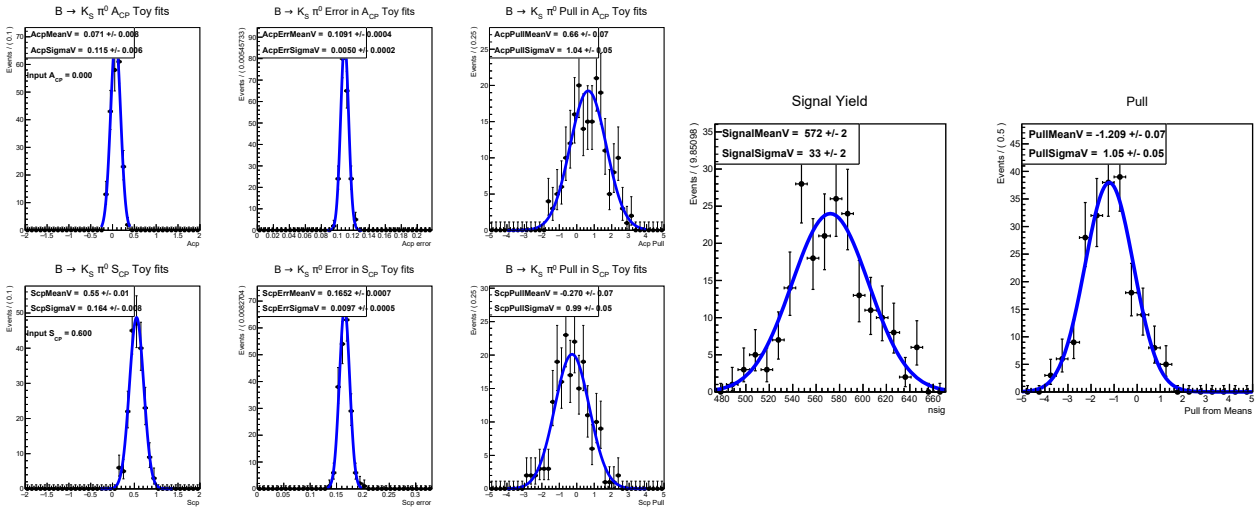

Figure 5.3: Toy MC fits of 1ab^{-1} of integrated luminosity. The left shows the value, error, and pull of Acp (top) and Scp (bottom), and the right shows the yield and pull of the signal yield.

5.3.2 62.8fb^{-1} Toy MC Results

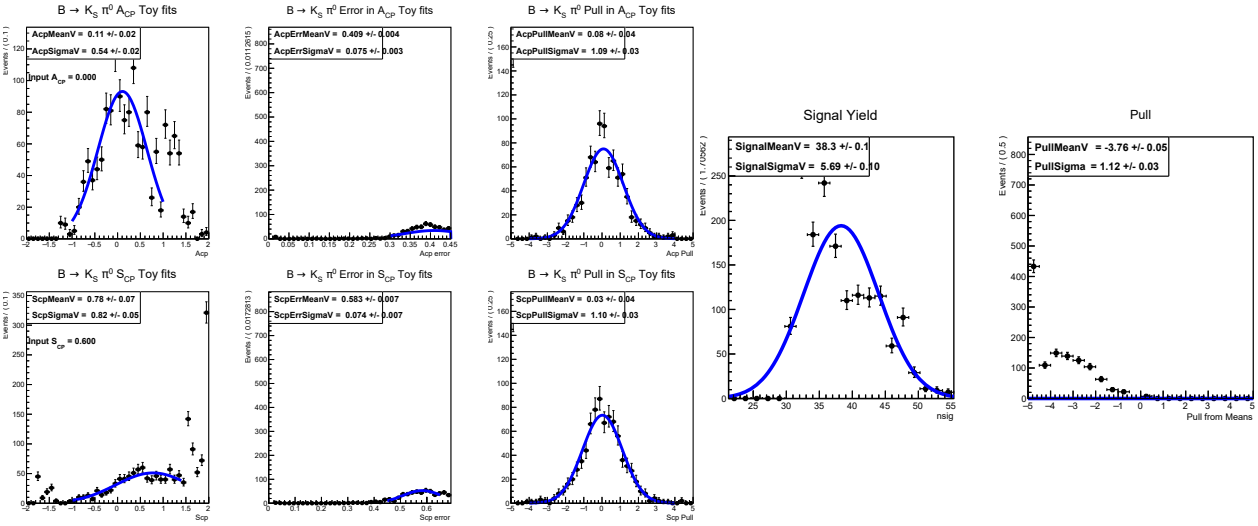

Figure 5.4: Toy MC fits of 62.8fb^{-1} of integrated luminosity.

Based on these results, it is likely that 62.8fb^{-1} of integrated luminosity will be insufficient to make proper measurements of Scp. The results for the yield and Acp will also be very limited. It is currently estimated that approximately 100fb^{-1} will be required for Acp and yield results and probably more for Scp. The results are shown in table. 5.2.

	Yield	Yield Error	Acp	Acp Error	Scp	Scp Error
$1ab^{-1}$	572	31	0.07	0.1	0.55	0.17
$62.8fb^{-1}$	38	8	0.11	0.409	0.78	0.583

Table 5.2: Results for Toy MC fits.

5.4 Linearity Fits

To check for a bias in the model, an MC toy model is done for a variety of values for the yield, Acp, and Scp. A plot of obtained values against expected values is then plotted and a line of best fit is taken. In a perfect model this will form a straight line with a gradient of one. This would mean the obtained values perfectly match what is expected. The result of this linearity fit is shown in Figure. 5.6

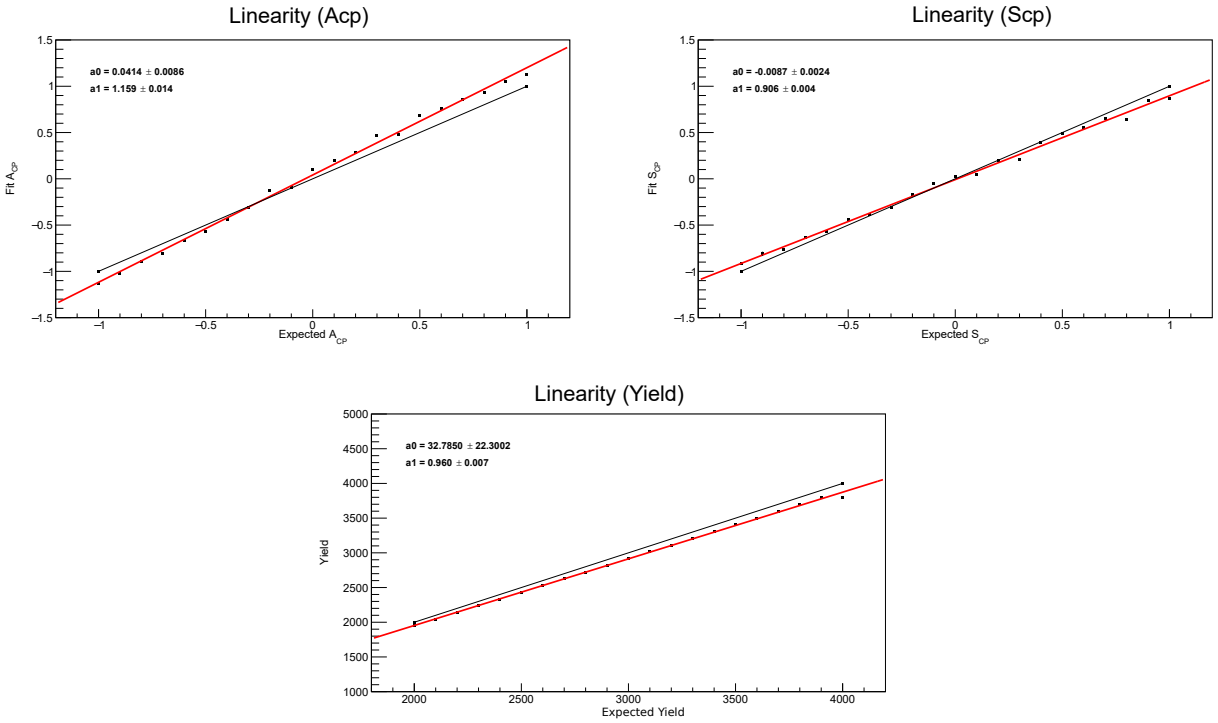

Figure 5.5: Linearity fits for the means of the Yield, Scp, and Acp.

This shows that there is a small bias. These linearity fits used $5ab^{-1}$ of integrated luminosity for each trial and 2000 fits were used for each data point. This bias can be accounted for by introducing a correction factor however this introduces systematic errors so the bias will need to be further studied. This bias was not present when TMVA was used (see section 5.4.1) for continuum suppression however FBDT is still preferred as it is more efficient (section 3.4.2).

5.4.1 The Initial Value of Scp

When initially fitting to DeltaT, an initial value of Scp must be chosen. This value remains unchanged up until doing the toy fits. Because of this, the initial value can have an impact on the bias in Scp. This section also looks at effect of fitting each signal variable (Mbc, DeltaE,

logSupp, DeltaT) using the same parameters or different parameters for each bin of qr. The model used in this section uses an older model which used eight bins of qr instead of seven. It also uses TMVA for continuum suppression. This should not affect the results however the bias here was found to be much less.

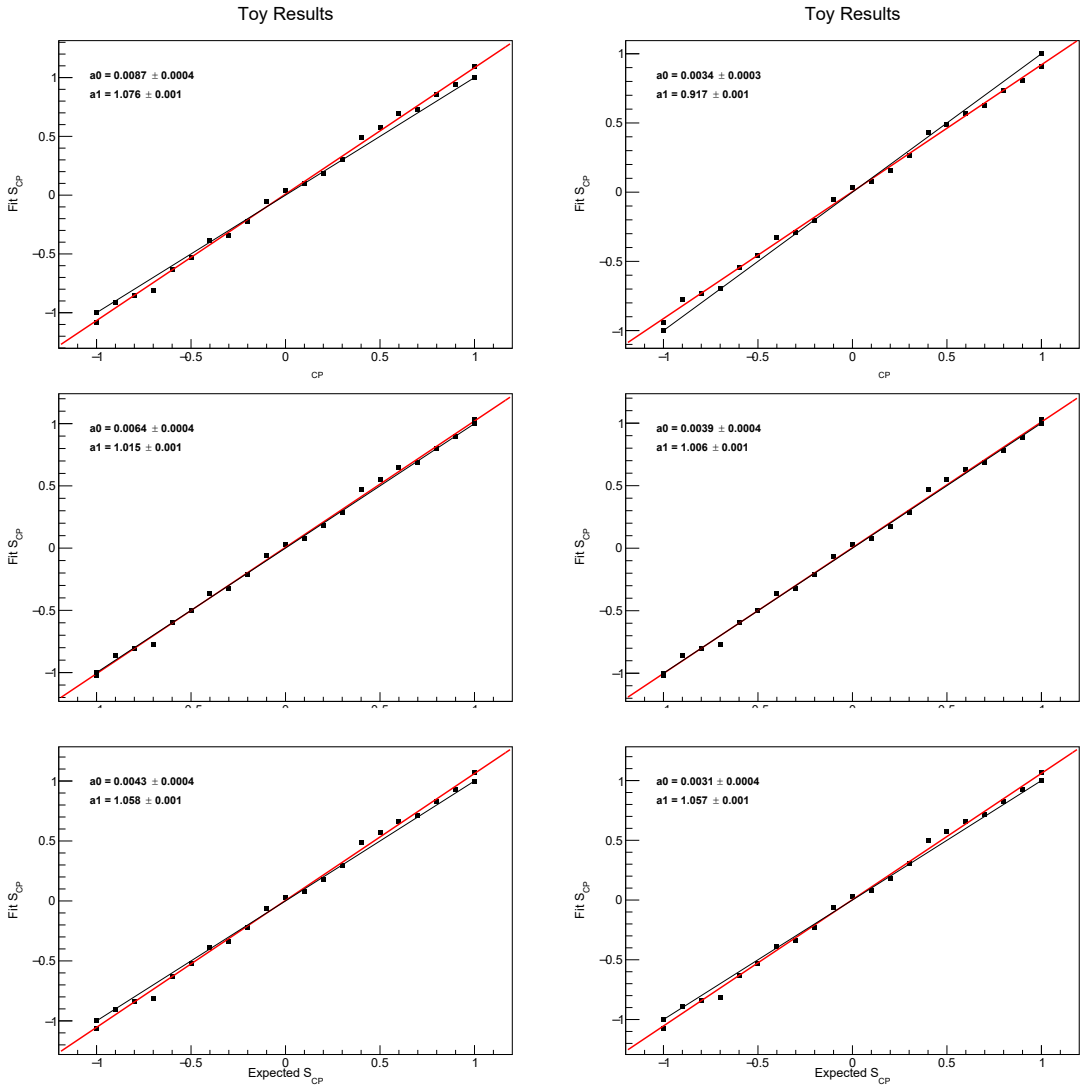

Figure 5.6: Linearity using the same parameters for each bin of signal PDF (left) and different parameters for each bin of signal PDF (right). The top row was initially fit to an Scp of negative one, the middle row used zero, and the bottom row used positive one.

These results show that initially fitting to a Scp value of zero is the best choice. While the bias was slightly lower when different parameters for each bin were used (0.6% compared to 1.5% for an initial Scp of zero), the difference is not enough to justify the increased systematic errors due to the number of parameters needed to fit multiple bins of qr. Therefore the same parameters for each bin are used.

5.5 Control Modes

In order to validate the analysis technique on real data without unblinding the $B^0 \to K_s\pi^0$ mode, the analysis is first used on another mode with similar properties but is already unblinded. This is called a control mode. The control modes deemed appropriate for this study

were the $B^0 \to J/\Psi K_s$ mode and the $B^+ \to K_s\pi^+$ mode. Since both of these modes can be vertexed using the J/Ψ (which decays via the $J/\Psi \to e^+e^-$ and $J/\Psi \to \mu^+\mu^-$ channels around 5% of the time each) and the π^+, it is important that the vertexing is done only using the K_s to imitate the $B^0 \to K_s\pi^0$ mode as closely as possible. This is the reason `RAVE` is used for vertexing and not `TreeFit` (section 3.2).

The difference can be seen by looking at the DeltaT error distributions (Figure. 5.7). When only the K_s is used for vertexing, the DeltaT error distribution has significantly more events at higher values. It also shows the peaking structure displayed in section 4.1.

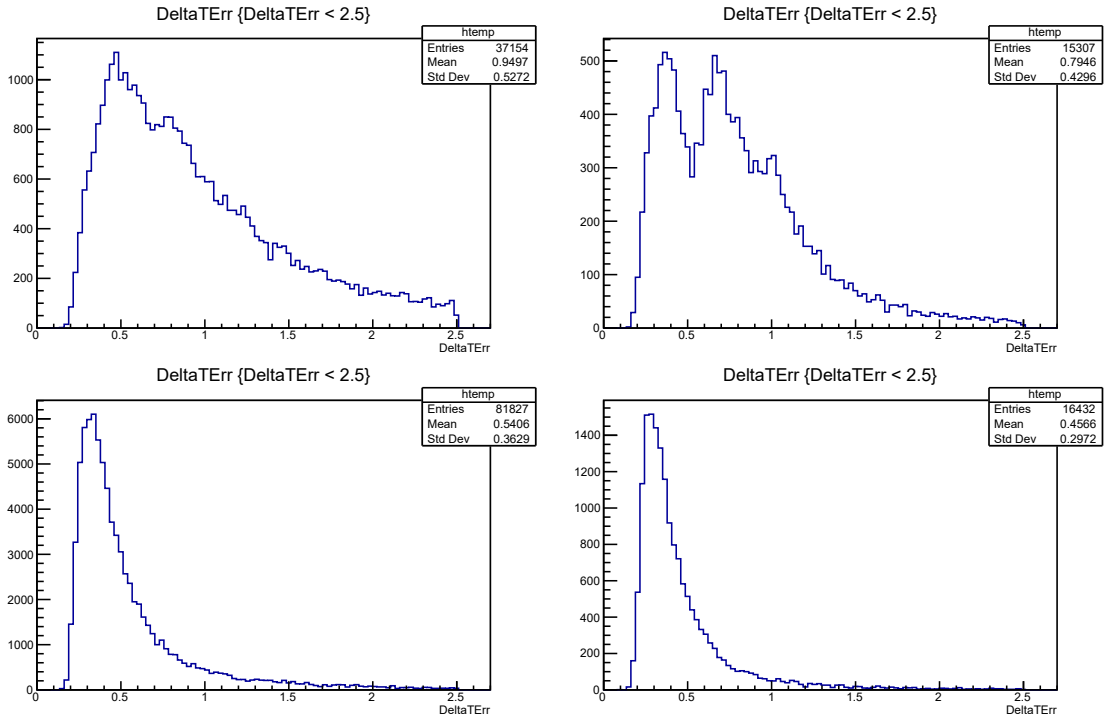

Figure 5.7: DeltaT error distribtions for $B^0 \to J/\Psi K_s$ (left) and $B^0 \to K_s\pi^+$ (right). The top plots show the distributions when only the K_s is used for vertexing, and the bottom plots show the distributions when all particles are used.

Since both of these modes were viable, the $B^0 \to J/\Psi K_s$ mode was ultimately chosen because it has a much lower expected number of background events. Here, the $J/\Psi \to e^+e^-$ channel was used however the $J/\Psi \to \mu^+\mu^-$ will be incorporated in future to increase the number of signal events. A cut on the reconstructed invariant mass of the J/Ψ was used to drastically reduce the number of background events. The cut was $dM < 0.11$.

5.5.1 Control Mode Fits

As with the $B^0 \to K_s\pi^0$ mode, the distribution of the variables must be fit. These are fit with the same shapes as was done in section 3.5 however the parameters are not expected to be the same. The fits for this mode are shown in Figure. 5.8.

The main difference between these fits and the fits for $B^0 \to K_s\pi^0$ is the long tail in the negative side for the DeltaE distribution. The cause of this is due to how the J/Ψ decays. Since the $J\Psi$ decays into e^+ and e^- in this case, the electrons lose energy due to bremsstrahlung radiation. This means that when the energy of the signal-side B-meson is reconstructed, it will be slightly lower than expected.

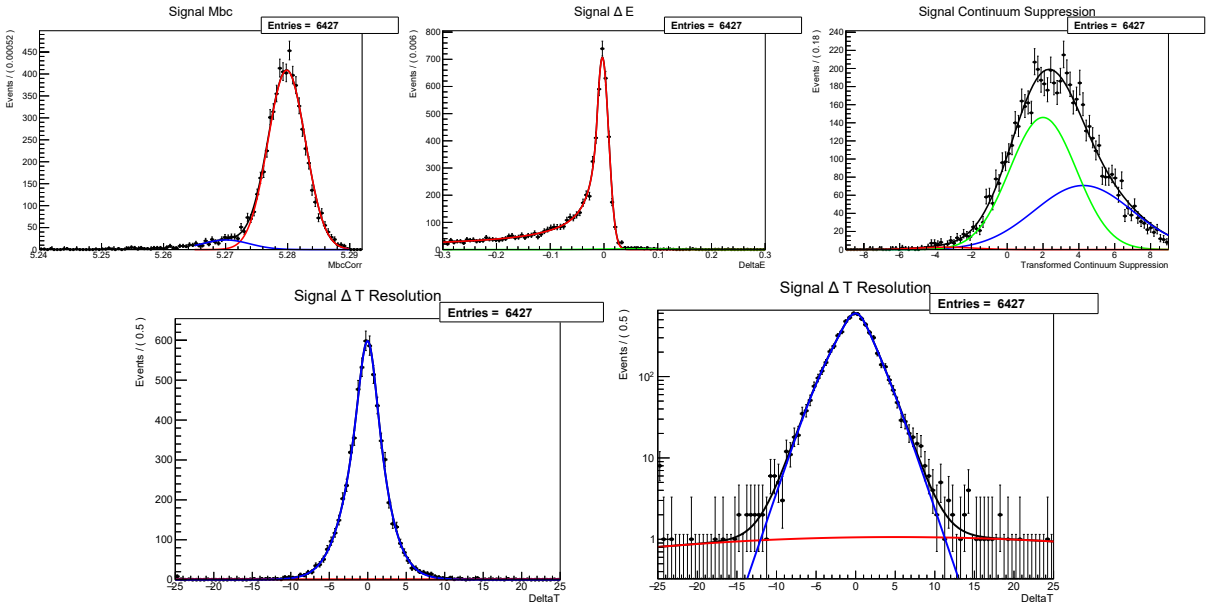

Figure 5.8: Fits to the signal for the control mode.

5.5.2 Control Mode Toy MC Results

The number of expected signal events is calculated using the same method as was done in section 5.1. The result is that for an integrated luminosity of 62.8fb^{-1}, 800 signal events are expected. Due to issues with DESY's grid computing system, the number of background events could not be fully calculated however it is estimated to be around 1200. There are no BB background events expected for this mode.

All the PDF's used here are preliminary due to delays in DESY's grid computing system. The preliminary PDF's are likely the cause of the overestimation of the yield which in turn is the cause of the higher value of Scp than expected. As before, the expected values are an Acp and Scp of 0.0 and 0.6 respectively.

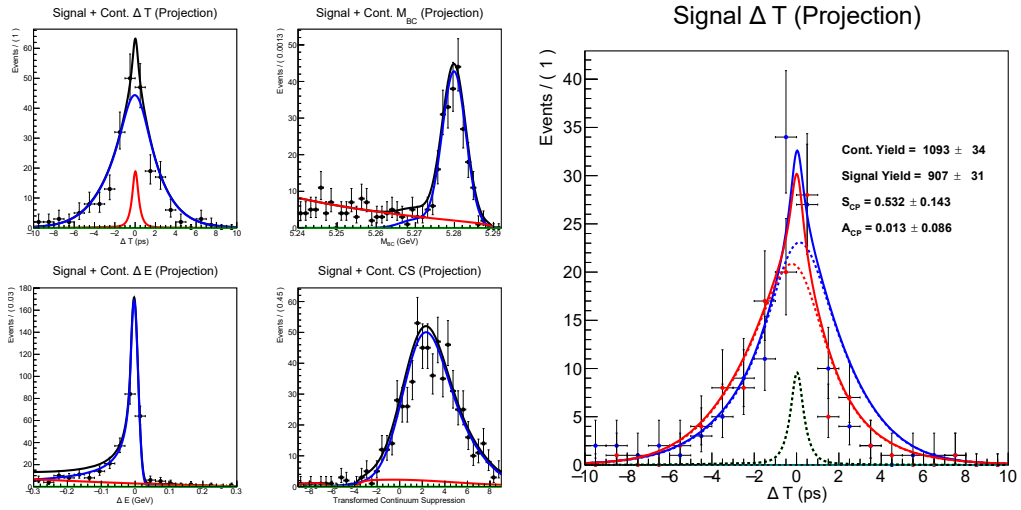

Figure 5.9: Fits to a synthesized data set expected from 62.8fb^{-1} of integrated luminosity for the control mode.

There were two values of Scp that were tested for the full MC toy which used values of Scp

of 0.0 and 0.6. 400 fits were used for each of these values. The results show that there is still a significant bias on Scp. This should be fixed when the PDF's are finalized.

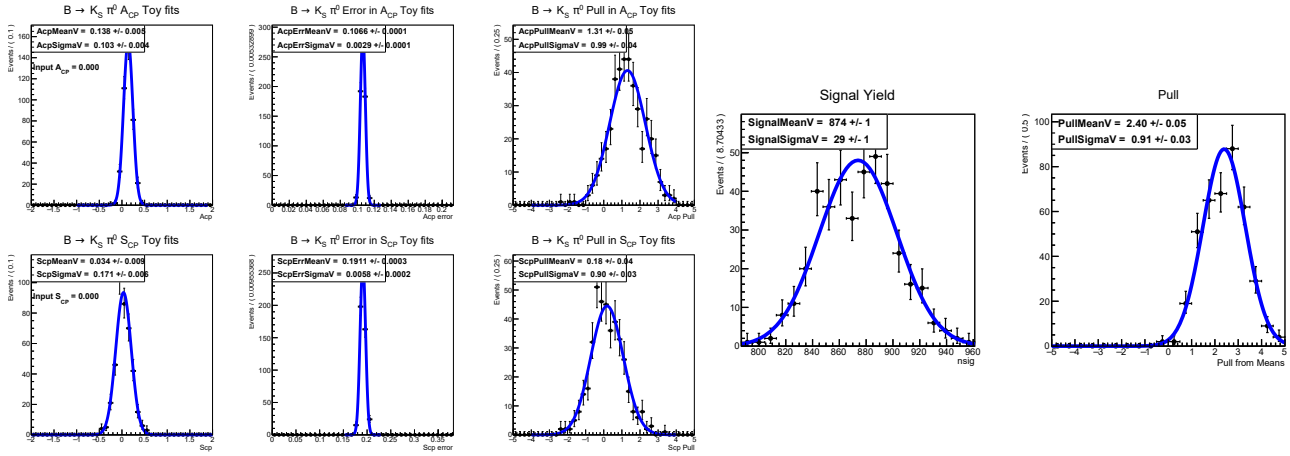

Figure 5.10: Toy MC fits of 62.8fb^{-1} of integrated luminosity for the control mode using an Scp value of zero.

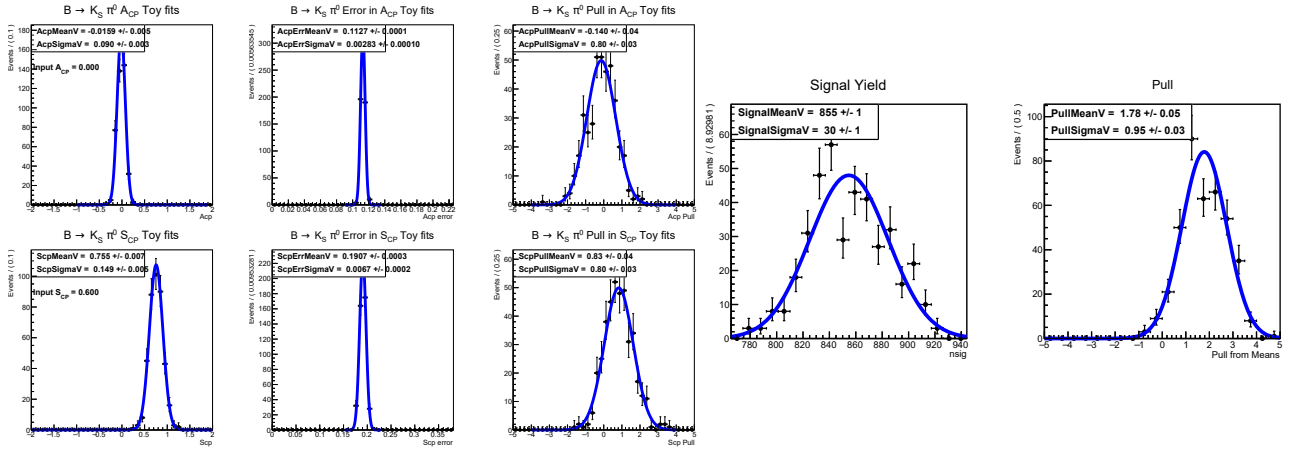

Figure 5.11: Toy MC fits of 62.8fb^{-1} of integrated luminosity for the control mode using an Scp value of 0.6.

When the control mode PDF's are finalised, a proper linearity fit will need to be done. The results are outlined in table 5.3.

	Yield	Yield Error	Acp	Acp Error	Scp	Scp Error
Scp: 0.0	874	31	0.138	0.107	0.034	0.19
Scp: 0.6	855	33	-0.016	0.113	0.755	0.149

Table 5.3: Results for Toy MC fits.

5.6 B-Meson Lifetime Fits

Another form of validation is to fit for the lifetime of the B-meson. This is done so that the analysis can be tested without unblinding the flavour-tagging data. Throughout this section only MC data is used. Real data for the control mode should be ready as soon as resources become available on the grid computing system.

The lifetime fits are achieved by fixing all parameters except for the yields and the B-lifetime. This was done using samples both with and without background. Doing this confirms that the fit works on MC signal data alone.

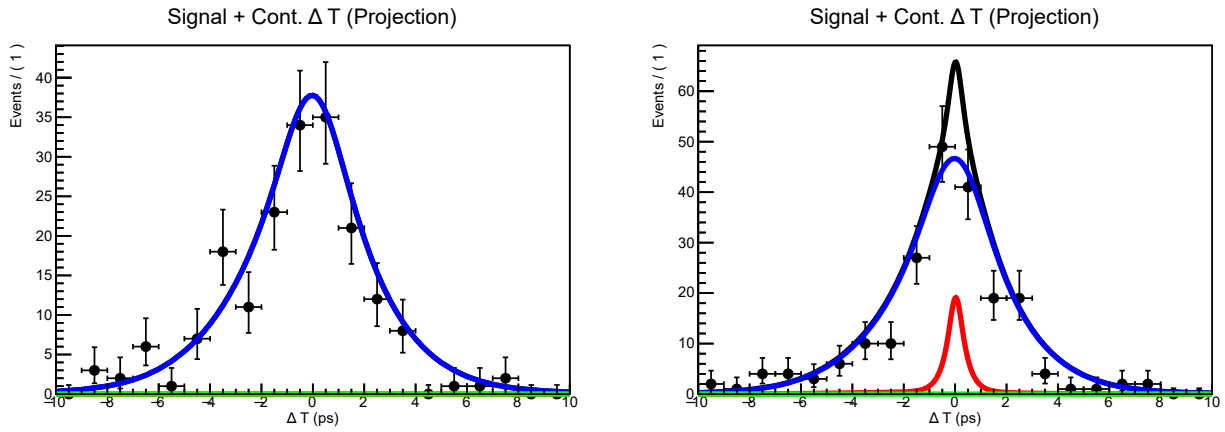

Figure 5.12: Lifetime fits for the B-meson using no background (left) and with background (right).

The results of this fit are shown in table. 5.4.

With Background:	(1.571 ± 0.08)ps
Without Background:	(1.354 ± 0.08)ps

Table 5.4: Results for the lifetime fits using the control mode. This uses 62.8fb^{-1} of integrated luminosity.

The expected value for the lifetime is 1.52ps. As with the other results in this section, the underestimation of the lifetime when the background is introduced is probably due to the overestimation of the yield.

Conclusion and Future Work

Throughout this work, a working analysis framework has been established for the $B^0 \to K_s\pi^0$ decay channel. This has been done using `RAVE` for vertexing, and a signal DeltaT PDF using seven bins of qr. The use of the BTube constraint for the tag-side vertexing has been implemented. This framework currently shows around a 9% bias in the value of Scp and Acp, and a yield which is consistently around one standard deviation from the expected value. The usage of TMVA and two sets of hyper-parameters for FBDT have been compared, and an optimized version of FBDT was found to be the best form of continuum suppression. A detailed analysis of the DeltaT error distribution has been studied and an optimal cut on DeltaT errors at 2.5ps has been found. A preliminary analysis of the $B^0 \to J/\Psi K_s$ control mode has been implemented including toy models and a framework for B-meson lifetime fits has been established.

Future work of this analysis will need to include finding the cause of the bias in Scp, Acp, and the yield, and finalization of the control mode PDF's. A full linearity fit will then need to be completed for the control mode and the $J/\Psi \to \mu^+\mu^-$ decay channel will be included. This can all happen when resources become available on the grid computing system. After this is completed, the B-lifetime fits can be completed on real data for the control mode and if there are no issues, Acp and Scp can be determined. If the real data control mode is successful, the $B^0 \to K_s\pi^0$ mode will then be looked at. The main limiting factor right now is the current integrated luminosity which is only good enough to give a very rough value of Scp. When more data is obtained a very high precision value of the branching fraction, Acp, and Scp can be obtained.

Bibliography

[1] Tetsuo Abe et al. "Belle II technical design report". In: *arXiv preprint arXiv:1011.0352* (2010).

[2] C Amsler et al. "12. CP VIOLATION IN MESON DECAYS". In: (2012).

[3] Nicolas Boisvert Beaudry et al. "The B→ πK puzzle revisited". In: *Journal of High Energy Physics* 2018.1 (2018), p. 74.

[4] Adrian Bevan et al. *The physics of the B factories*. Springer Nature, 2017.

[5] Rene Brun et al. *root-project/root: v6.18/02*. Version v6-18-02. Aug. 2019. DOI: 10.5281/zenodo.3895860. URL: https://doi.org/10.5281/zenodo.3895860.

[6] Andrzej J Buras and Robert Fleischer. "Quark mixing, CP violation and rare decays after the top quark discovery". In: *Advanced Series on Directions in High Energy Physics* 15 (1998), p. 65.

[7] Nicola Cabibbo. "Unitary symmetry and leptonic decays". In: *Physical Review Letters* 10.12 (1963), p. 531.

[8] A Ceccucci, Y Sakai, and Z Ligeti. "The CKM quark-mixing matrix". In: *J. Phys. G* 37 (2010), p. 075021.

[9] Ling-Lie Chau and Wai-Yee Keung. "Comments on the parametrization of the Kobayashi-Maskawa matrix". In: *Physical Review Letters* 53.19 (1984), p. 1802.

[10] James H Christenson et al. "Evidence for the 2 π Decay of the K 2 0 Meson". In: *Physical Review Letters* 13.4 (1964), p. 138.

[11] Alakabha Datta, John Waite, and Divya Sachdeva. "Unified explanation of b→ s μ+ μ- anomalies, neutrino masses, and B→ π K puzzle". In: *Physical Review D* 100.5 (2019), p. 055015.

[12] Robert Fleischer et al. "Utilising *B to* π K Decays at the High-Precision Frontier". In: *arXiv preprint arXiv:1805.06705* (2018).

[13] A Haupt et al. "The DESY Grid Centre". In: *Journal of Physics: Conference Series* 396.4 (Dec. 2012), p. 042026. DOI: 10.1088/1742-6596/396/4/042026. URL: https://doi.org/10.1088/1742-6596/396/4/042026.

[14] Patrick Huet and Eric Sather. "Electroweak baryogenesis and standard model CP violation". In: *Physical Review D* 51.2 (1995), p. 379.

[15] R Jaarsma. "Utilising B→ πK Decays at the High-Precision Frontier". In: *The International Conference on B-Physics at Frontier Machines-BEAUTY2018*. Vol. 6. 2018, p. 11.

[16] F. James. "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1". In: (1994).

[17] Thomas Keck. "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification". In: *CoRR* abs/1609.06119 (2016). arXiv: 1609.06119. URL: http://arxiv.org/abs/1609.06119.

[18] Makoto Kobayashi and Toshihide Maskawa. “CP-violation in the renormalizable theory of weak interaction”. In: *Progress of theoretical physics* 49.2 (1973), pp. 652–657.

[19] E Kou et al. “The Belle II physics book (2018)”. In: *arXiv preprint arXiv:1808.10567* ().

[20] Robert Kowalewski and Thomas Mannel. “Determination of vcb and vub”. In: *J. Phys. G* 33 (2006), p. 1.

[21] T. Kuhr et al. “The Belle II Core Software”. In: *Computing and Software for Big Science* 3.1 (Nov. 2018). ISSN: 2510-2044. DOI: 10.1007/s41781-018-0017-9. URL: http://dx.doi.org/10.1007/s41781-018-0017-9.

[22] BERNARD MEADE et al. “Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility”. In: (Apr. 2017). DOI: 10.4225/49/58ead90dceaaa. URL: https://melbourne.figshare.com/articles/online_resource/Spartan_HPC-Cloud_Hybrid_Delivering_Performance_and_Flexibility/4768291.

[23] Antonio Pich. “CP violation”. In: *arXiv preprint hep-ph/9312297* (1993).

[24] Anders Ryd et al. “EvtGen: A Monte Carlo Generator for B-Physics”. In: (May 2005).

[25] Andrei D Sakharov. “Violation of CP-invariance, C-asymmetry, and baryon asymmetry of the Universe”. In: *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16*. World Scientific, 1998, pp. 84–87.

[26] Henry P Stapp. “Derivation of the CPT Theorem and the Connection between Spin and Statistics from Postulates of the S-Matrix Theory”. In: *Physical Review* 125.6 (1962), p. 2139.

[27] Masaharu Tanabashi et al. “Review of particle physics”. In: *Physical Review D* 98.3 (2018), p. 030001.

[28] Lincoln Wolfenstein. “Parametrization of the Kobayashi-Maskawa matrix”. In: *Physical Review Letters* 51.21 (1983), p. 1945.

[29] Chien-Shiung Wu et al. “Experimental test of parity conservation in beta decay”. In: *Physical review* 105.4 (1957), p. 1413.