Penguin-Hunting with Neural Nets

Time-Dependent $\mathcal{CP}$ Violation Measurement of $B^0 \to K^0_S \pi^0$

Pak Hin (Brian) Chan

Under the Supervision of
A/Prof. Martin Sevior

The School of Physics
The University of Melbourne

Submitted On: October 25, 2019

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

Abstract

As it stands, the $\mathcal{CP}$ violating parameters of $B^0 \to K_S^0 \pi^0$ show tension between the Standard Model and experimental results at the 2σ level, making it a prime candidate to investigate New Physics particles such leptoquarks and Z'. This discrepancy can be resolved at the Belle II experiment as we expected to achieve a forty-fold increase of data when compared to its predecessor used for the current results. Assuming these parameters do not deviate significantly from their current values, we show in this thesis the tensions in $B^0 \to K_S^0 \pi^0$ may be resolved as early as 2023.

Before a measurement can be taken, one must establish an unbiased framework tested against Monte-Carlo simulated data. In this thesis we fully establish and validate the analysis methodology in preparation for future datataking at Belle II. Using this workflow, alternative vertexing algorithms are investigated, and the uncertainty in the resolution is accounted for in a per-event basis. We assess the statistical uncertainty and bias associated with these methods. $B^0 \to J/\psi K_S^0$ is used as a control mode to estimate systematic uncertainties in $B^0 \to K_S^0 \pi^0$, and correction factors relating the two are determined.

Statement of Contribution

Chapter 1 was an original review of the literature. Chapter 2 is an accumulation of knowledge pertaining to the experiment acquired through cited references or internal documents. Chapter 3 is an outline of the analysis methodology adapted from previous work by Martin Sevior, and implemented by the author. In particular, the grid search method outlined in Section 3.1.2 was a novel feature appended to the analysis. Chapters 4 & 5 contains original work unless otherwise stated.

Statement of Acknowledgement

Contents

4 Validation and Results

Chapter 1

Introduction & Motivation

Since the dawn of humanity, the objective of natural philosophy has been an inquiry into the inner workings of the universe – a goal which has been elusive to the brightest of minds. Particle Physics is the investigation of fundamental particles which constitute our reality as well as the laws which govern them. The pinnacle of our cumulative efforts in the past century is referred to by the unassuming title of: "The Standard Model".

This chapter discusses the features and limitations of this model. In particular, we investigate how the difference in the decay of B^0 and $\overline{B}^0$ to $K_S^0 \, \pi^0$ is the measure of the violation of a fundamental discrete symmetry in the universe known as $\mathcal{CP}$ symmetry. $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$, the parameters which represent $\mathcal{CP}$ violation are empirically shown to deviate from their theoretical Standard Model (SM) values by approximately 2σ, which can be a probe into theories beyond the SM. In Chapter 2, we discuss how the Belle II experiment provides a clean[1] environment making this this the best way to observe this anomaly. In Chapter 3, we explore using Monte-Carlo simulated data to build and validate physics analyses – a process which allows for the extraction physical observables from raw detector data. Finally, in Chapter 4, we fully establish this physics analysis framework under different experimental approaches, to minimise the error in $\mathcal{S}_{CP}$ and model bias. A relationship between the well-established $B^0 \rightarrow J/\psi \, K_S^0$ as a control mode and $B^0 \rightarrow K_S^0 \pi^0$ is found, which will allow us to estimate of the systematic uncertainty of $\mathcal{CP}$ violating parameters. Assuming, $\mathcal{S}_{CP}$ and $\mathcal{A}_{CP}$ hold their central value, we estimate a resolution of this anomaly in 2023, and a detection of the $B^0 \rightarrow K_S^0 \pi^0$ decay channel with a 5σ statistical significance by 2020.

1.1   The Standard Model of Particle Physics

The SM is a relativistic Lorentz-invariant Quantum Field Theory (QFT) which describes fundamental spin-$\frac{1}{2}$ fermions (particles), and how they interact via the strong, weak, and electromagnetic forces. These fermions are divided into two groups: quarks and leptons, and they exist in three flavours (or generations), which share the same physical properties except mass. Each fermion has an antimatter counterpart. Quarks interact with all three forces, whereas leptons only interact with the weak force and electromagnetism (if charged). Quarks exists only in bound states, including:

- Mesons – comprised of quark and anti-quark pairs ($q\bar{q}$)

- Baryons – comprised of three quarks (or anti-quarks) (qqq or $\overline{qqq}$)

Interactions in the SM are mediated by spin-1 gauge boson particles. The strong force is mediated by a set of 8 coloured gluons, whereas the weak force and electromagnetism are mediated by $W^{\pm}$ and Z^0, and

---

[1]At Belle II measurements are taken from collisions of fundamental particles rather than composite particles (at experiments such as LHCb).

photons γ respectively (see Figure 1.1). The mass of fundamental particles depends on the strength of their coupling to the Higgs field, which necessitate the existence of the Higgs Boson.
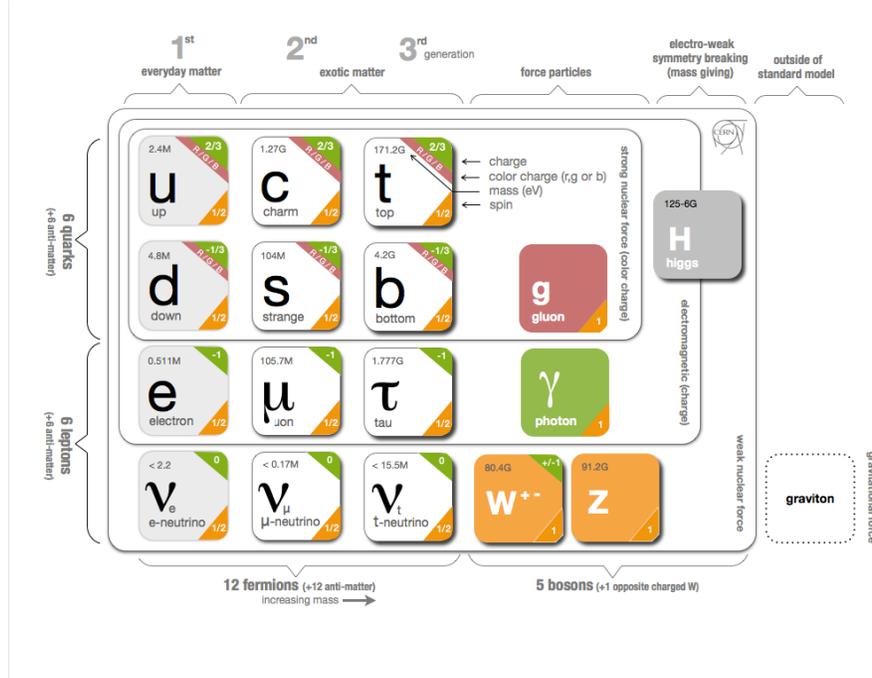
Figure 1.1: This is an overview of the particles (fermions, gauge bosons and the Higgs boson) which constitute the Standard Model. Note the presence of the "graviton" which is a particle theorised to exist beyond the Standard Model, which allows gravity to be unified with the three other fundamental forces. Credits: CERN

1.1.1   Quark Mixing and the CKM Matrix

Within the SM, quarks are only allowed to change flavour in the weak interaction through coupling with the $W^{\pm}$ bosons. These flavour transitions must be from an up-type quark (with a charge of $+\frac{2}{3}$ such as u, c, t) quark to a down-type quark (with a charge of $-\frac{1}{3}$ such as d, s, b), or vice versa. The strength of these interactions is given by the unitary 3×3 Cabibbo-Kobayashi-Maskawa (CKM) matrix:

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

where V_{ij} represents the quark transition $i \to j$ and must be determined experimentally. Three mixing angles $\theta_{12}, \theta_{23}, \theta_{13}$ and a single irreducible complex phase δ is sufficient to uniquely define this matrix. This complex phase is the source of $\mathcal{CP}$ violation in the SM (see Section 1.3). The standard parameterisation of the CKM matrix [1] is given by four free parameters:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}s_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13} & c_{23}c_{13} \end{pmatrix} \tag{1.2}$$

where $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$ involves one of the three mixing angles. The diagonal elements approaches unity. Elements far off from the diagonal ($|V_{td}|, |V_{ub}|$) are exeedingly small, resulting in a heavy suppression in the (liklihood of) their respective quark transitions.

1.1.2 The Unitary Triangle

Because the matrix is unitary, $(V_{CKM}V_{CKM}^\dagger = I)^2$, a set of conditions[3] such as:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.3}$$

can be imposed. This is often represented by a triangle in the complex plane. The angles of this Unitary Triangle (UT) shown in Figure 1.2 are obtained by re-scaling Equation 1.3 giving three angles to represent the CKM matrix:

$$\phi_1(=\alpha) = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right) \tag{1.4}$$

$$\phi_2(=\beta) = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right) \tag{1.5}$$

$$\phi_3(=\gamma) = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right) \tag{1.6}$$
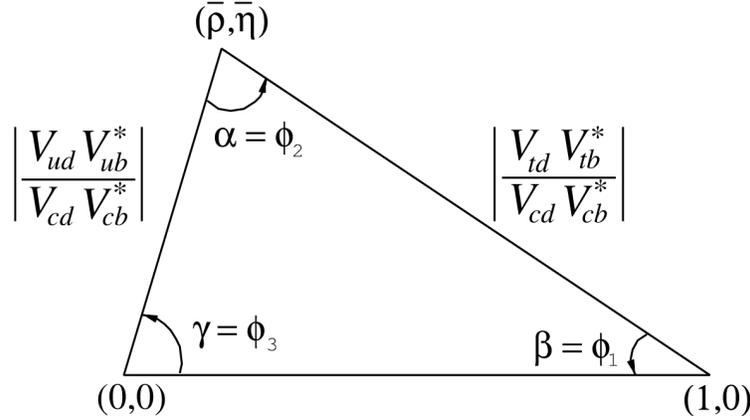
Figure 1.2: Sketch of the Unitary Triangle in the complex plane. Note the parameters $\bar{\rho}, \bar{\eta}$ relate to a different parameterisation of the CKM matrix not discussed in this thesis. Credits: Particle Data Group [2].

1.1.3 Limitations of the Standard Model

Despite being the successful and robust theory that it is, the SM falls short of the ultimate goal of physics since time immemorial – a theoretical description of all natural phenomena. Gravity, the weakest of the four fundamental forces and described by Einstein's General Relativity, is unable to be reconciled with the SM. Dark matter, whose existence is evidenced by numerous cosmological observations, is omitted from the SM. Neutrinos, considered to have small but non-zero mass due to neutrino oscillation, are required to be massless in the SM. Finally, the $\mathcal{CP}$ violation generated in the SM is insufficient to account for the observed baryogenesis (see Section 1.2.1). This set of discrepencies provide us with strong motivation to explore New Physics (NP) beyond the SM. Within the scope of this thesis, however, we will focus on the final tension in $\mathcal{CP}$ violation.

---

[2]Note that the † represents a Hermitian conjugate which takes the transpose of a matrix and then conjugates element-wise

[3]The condition which arises from unitarity is not unique. The choice of the constraint cited was made since each matrix element has the same order of magnitude.

1.2 The Baryon Asymmetry of the Universe

In the beginning of the universe, the abundance of matter and antimatter was thought to have been equal[4] [3]. In present day, however, we observe a universe heavily dominated by matter. The asymmetry in (baryonic) matter is parameterised by

$$\eta = \frac{N_B - N_{\overline{B}}}{N_\gamma} \tag{1.7}$$

where N_B ($N_{\overline{B}}$) gives the number of baryons (antibaryons). The normalisation N_γ is the number of photons in the universe. The transition of η from zero to a non-zero state is known as the Baryon Asymmetry of the Universe (BAU), and the process by which this happens is baryogenesis. In order for baryogenesis to take place, Andrei Sakharov [4] proposed three necessary conditions:

1. **Baryon Number Violation** – processes which violate baryon number[5] must be allowed to occur
2. $\mathcal{C}$**-symmetry and** $\mathcal{CP}$**-symmetry Violation** – baryogensis processes must favour baryons over antibaryons (see Section 1.3 for a detailed discussion of $\mathcal{CP}$ violation)
3. **Interactions Outside Thermal Equilibrium** – to ensure that the process which creates baryons is not reversed[6]

The second condition will be the focus of this thesis.

1.2.1 Observations vs. Standard Model Predictions of BAU

The SM fulfills all three of Sakharov's conditions, and thus allows for baryogensis. The degree of baryogensis depends on the level of $\mathcal{CP}$ violation, generated by the complex phase of the CKM matrix. Huet [5] placed an upper limit on the asymmetry within the SM at $|\eta| \sim \mathcal{O}(10^{-26})$. However, from multiple astrophysical experiments [6], we find that $\eta \sim \mathcal{O}(10^{-10})$ which is seventeen orders of magnitude above what was expected by our model. As such, this BAU problem is an area which may hint at NP beyond the SM.

1.3 Discrete Symmetries and $\mathcal{CP}$ Violation

Since the amount of baryogensis and $\mathcal{CP}$ violation is intrinsically linked, a discussion of $\mathcal{CP}$ violation must not be omitted when considering this anomaly. This section begins the discussion in general discrete symmetries, and then establishes the formalism for $\mathcal{CP}$ violation in neutral B-mesons.

A discrete symmetry is preserved when the laws of physics (weak, strong, electromagnetism), remain invariant under a discrete transformation. Otherwise the symmetry is said to be violated. Some discrete transformations include:

- Charge Conjugation $\mathcal{C}$: particle and antiparticle are interchanged $[q \rightarrow -q]$
- Parity $\mathcal{P}$: spatial axes are reversed $[(x, y, z) \rightarrow (-x, -y, -z)]$[7]
- Time Reversal $\mathcal{T}$: time-direction of processes are reversed $[t \rightarrow -t]$

---

[4]Even if asymmetry existed very early on, under the theory of cosmic inflation, it would have diluted, and caused the difference between the number of baryons and antibaryons to approach zero.

[5]Each baryon (antibaryon) in a system is assigned a baryon number of $+1$ (-1). This parameter essentially 'counts' the difference in numbers of baryons and antibaryons.

[6]When a system is in thermal equilibrium, baryon violating processes are allowed to proceed both forwards and backwards. If the $\mathcal{CPT}$ symmetry (see Section 1.3) is indeed preserved, the probability of observing a state of $B = +1$ is equally likely as $B = -1$. The average B of the universe is required then to be zero.

[7]This transformation can be thought of as taking a 'mirror' image of an object.

It follows that the $\mathcal{CP}$ transformation is the combination of the $\mathcal{C}$ and $\mathcal{P}$ symmetries.

Historically, the laws of nature were thought to be the same for matter and antimatter and invariant under $\mathcal{C}$, $\mathcal{P}$ and $\mathcal{T}$ transformations. However, in 1956, a gap in the literature was found in the $\mathcal{P}$ symmetry for the weak interaction by Lee & Yang [7]. Wu [8] subsequently provided empirical evidence for $\mathcal{P}$ symmetry being maximally violated in the (weak) nuclear decay of ^{60}Co. Later in 1964, the $\mathcal{CP}$ symmetry was found to be violated in the neutral kaon systems by the Cronin & Fitch experiment [9]. In 2001, $\mathcal{CP}$ violation was discovered in the B-meson systems by the Belle [10] and BaBar [11] collaborations.

Although our focus is $\mathcal{CP}$ violation in the weak interaction, it is interesting to note that experimental evidence thus far points to a stronger symmetry, $\mathcal{CPT}$ remaining unbroken in all the fundamental interactions. If this discrete symmetry were shown to be broken, then Lorentz -invariance would not be guaranteed and our current understanding of QFT will have to be reformulated. It is also interesting to note that although not theoretically forbidden, the strong interaction shows no sign of $\mathcal{CP}$ violation in nature.

1.3.1 Formalism for $\mathcal{CP}$ Violation

In order to measure $\mathcal{CP}$ violation, one must first establish the undergirding quantum mechanics framework. It turns out that B-mesons are particularly sensitive to NP phenomena as it shows many areas of tension with the SM (see Section 1.4). Within this sector, we will explore the $B^0 \to K_s^0 \pi^0$ decay, and as such this section will be devoted to establishing the formalism of $\mathcal{CP}$ violation in neutral B-mesons (although it may be generalised to any neutral meson system – see Ref. [2, 12]).

When a quantum mechanical object exists in a definitive state, it is commonly referred to as an eigenstate[8]. Due to its probabilistic nature, the physical observables of mesons are not always necessarily eigenstates (when in a superposition for instance). It follows that the flavour eigenstate of the B-meson is when it is either B^0 or $\overline{B}^0$. Over time, the particular B-meson will oscillate (mix) between these two flavour eigenstates. Suppose we would like to define an arbitrary neutral B-meson state $|\phi(0)\rangle$, we can do so by expressing it as a superposition of the flavour eigenstates:

$$|\phi(0)\rangle = a(0) \left| B^0 \right\rangle + b(0) \left| \overline{B}^0 \right\rangle \tag{1.8}$$

Over time, these mesons may decay into a set of final states f_i, or oscillate. The wavefunction becomes:

$$|\phi(t)\rangle = a(t) \left| B^0 \right\rangle + b(t) \left| \overline{B}^0 \right\rangle + \sum_i c_i(t) \left| f_i \right\rangle \tag{1.9}$$

The time evolution of this state may be described by Schrödinger's equation:

$$i\frac{\partial}{\partial t} |\phi(t)\rangle = \mathcal{H}_{eff} |\phi(t)\rangle \tag{1.10}$$

where $\mathcal{H}_{eff}$ is the effective Hamiltonian comprised of Hermitian matrices $\mathcal{M}$ and $\mathcal{G}$:

$$\mathcal{H}_{eff} = \mathcal{M} - \frac{i}{2}\mathcal{G} = \begin{pmatrix} M_{11} - \frac{i}{2}\Gamma_{11} & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{21} - \frac{i}{2}\Gamma_{21} & M_{22} - \frac{i}{2}\Gamma_{22} \end{pmatrix} \tag{1.11}$$

Assuming the $\mathcal{CPT}$ symmetry is unbroken, the following constraints can be applied to the matrix elements of $\mathcal{M}$ and $\mathcal{G}$

$$M_{11} = M_{22} \qquad \Gamma_{11} = \Gamma_{22} \tag{1.12}$$

$$M_{21} = M_{12}^* \qquad \Gamma_{21} = \Gamma_{12}^* \tag{1.13}$$

---

[8]$|x\rangle$ is an eigenstate of the observable A, if $A|x\rangle = \alpha |x\rangle$, where $\alpha \in \mathbb{C}$ is its corresponding eigenvalue.

The mass eigenstates can be defined as a linear combination of the flavour eigenstates:

$$|B_L\rangle \equiv p\,|B^0\rangle + q\,|\overline{B}^0\rangle \tag{1.14}$$

$$|B_H\rangle \equiv p\,|B^0\rangle - q\,|\overline{B}^0\rangle \tag{1.15}$$

with the normalisation $|p|^2 + |q|^2 = 1$. Using this, we can obtain:

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}} \tag{1.16}$$

Suppose the $\omega_{L,H}$ are the eigenvalues to the mass eigenstate $|B_{L,H}\rangle$. The difference in mass of the B_L and B_H and the difference in their decay lengths are referred to as the mass Δm and decay length $\Delta\Gamma$ splitting respectively, which can be expressed in terms of these mass eigenstates:

$$\Delta m \equiv m_H - m_L = \mathrm{Re}(\omega_H - \omega_L) \tag{1.17}$$

$$\Delta\Gamma \equiv \Gamma_H - \Gamma_L = -2\,\mathrm{Im}(\omega_H - \omega_L) \tag{1.18}$$

It is also convenient to define the decay amplitudes (see Section 1.4.1) as:

$$A_f = \langle f|\mathcal{H}|B^0\rangle, \quad \overline{A}_f = \langle f|\mathcal{H}|\overline{B}^0\rangle \tag{1.19}$$

$$A_{\bar{f}} = \langle \overline{f}|\mathcal{H}|B^0\rangle, \quad \overline{A}_{\bar{f}} = \langle \overline{f}|\mathcal{H}|\overline{B}^0\rangle \tag{1.20}$$

1.3.2 Types of $\mathcal{CP}$ Violation

$\mathcal{CP}$ violation can appear in three different forms: direct, indirect, and interference.

I. $\mathcal{CP}$ Violation in Decays (Direct)

Direct $\mathcal{CP}$ violation occurs when the decay rates of B^0 and $\overline{B}^0$ differ,

$$\Gamma(B^0 \to f) \neq \Gamma(\overline{B}^0 \to \overline{f}) \tag{1.21}$$

which is satisfied when

$$\left|\frac{\overline{A}_{\bar{f}}}{A_f}\right| \neq 1 \tag{1.22}$$

For the $B^0 \to K_S^0\pi^0$ decay, the final states are $\mathcal{CP}$ eigenstates (ie. $f_{CP} = \overline{f}_{CP}$), and the asymmetry is defined as

$$A_{CP} = \frac{\Gamma(\overline{B^0} \to K_S^0\pi^0) - \Gamma(B^0 \to K_S^0\pi^0)}{\Gamma(\overline{B^0} \to K_S^0\pi^0) + \Gamma(B^0 \to K_S^0\pi^0)} \tag{1.23}$$

This occurs as an interference between one leading order strong decay and one leading order weak decay (a condition which is fulfilled in $B^0 \to K_S^0\pi^0$, see Figure 1.3).

II. $\mathcal{CP}$ Violation in Mixing (Indirect)

Indirect $\mathcal{CP}$ violation occurs when a particular final state can come only from either B^0 or $\overline{B}^0$, but not both. This is equivalent to:

$$\left|\frac{q}{p}\right| \neq 1 \tag{1.24}$$

This type of asymmetry is caused by the mixing probability between the two flavour eigenstates being unequal. This is most prominent when B decays into a final state containing both hadrons and leptons. This asymmetry does not apply to $B^0 \to K_S^0\pi^0$.

III. $\mathcal{CP}$ Violation via Interference

Considering only $\mathcal{CP}$ eigenstates as final states, $\mathcal{CP}$ violation via interference occurs when there exists two ways of reaching the final state, the rates of which are not the same.

- Direct decay: $B^0 \to f_{CP}$
- Mixing first, then decay: $B^0 \to \overline{B}^0 \to f_{CP}$

Note that this can occur even if $\mathcal{CP}$ violation for direct and indirect are zero. This occurs when:

$$\text{Im}(\lambda_f) \neq 1 \quad \text{where} \quad \lambda_f = \frac{q}{p}\frac{\overline{A}_f}{A_f} \tag{1.25}$$

The asymmetry is time-dependent as the rates of reaching the final state between the two methods are not the same:

$$A_{CP}(t) = \frac{\Gamma(\overline{B^0}(t) \to K_S^0\pi^0) - \Gamma(B^0(t) \to K_S^0\pi^0)}{\Gamma(\overline{B^0}(t) \to K_S^0\pi^0) + \Gamma(B^0(t) \to K_S^0\pi^0)} \tag{1.26}$$

Time Dependence of $\mathcal{CP}$ Violation

Equation 1.26 can further be expressed as:

$$A_{CP}(t) = \mathcal{A}_{CP}\cos(\Delta m t) + \mathcal{S}_{CP}\sin(\Delta m t) \tag{1.27}$$

where Δm is the mass splitting (see Equation 1.17), and $\mathcal{A}_{CP}$ gives the direct $\mathcal{CP}$ violation and the $\mathcal{S}_{CP}$ gives the $\mathcal{CP}$ violation via interference:

$$\mathcal{A}_{CP} = \frac{|\lambda_f|^2 - 1}{|\lambda_f|^2 + 1} \quad \mathcal{S}_{CP} = \frac{2\,\text{Im}(\lambda_f)}{1 + |\lambda_f|^2} \tag{1.28}$$

The decay rate of the neutral B is given by:

$$\mathcal{P}_{K_S^0\pi^0}(\Delta T, q) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}}\left[1 + q\left(\mathcal{A}_{CP}\cos(\Delta m \Delta T) + \mathcal{S}_{CP}\sin(\Delta m \Delta T)\right)\right] \tag{1.29}$$

Where the τ_{B^0} is the lifetime of the B^0 and Δm is the aforementioned mass splitting. These time-dependent measurements are usually taking at collider experiments where a pair of coherent B-mesons are produced. When one of them decays to $K_S^0\,\pi^0$, the other one decays according to their relevant amplitudes. The q in the above expression indicates the flavour of the latter B-meson[9], and differs from the q defining the superposition of the mass eigenstate.

1.4 Motivation: $B^0 \to K_S^0\pi^0$ puzzle

The $B \to K\pi$ system is a collection of four decays ($B^+ \to K^0\pi^+$, $B^+ \to K^+\pi^0$, $B^0 \to K^+\pi^-$, $B^0 \to K^0\pi^0$) (and their conjugates) which contains nine observables: 4 branching ratios[10], 4 $\mathcal{A}_{CP}$, and 1 $\mathcal{S}_{CP}$.

When a global fit of the theoretical parameters of this system is performed with the experimentally verified observables, the fit is found to not converge well [13]. This inconsistency[11] between the SM and

---

[9] q is $+1$ and -1 for B^0 and $\overline{B}^0$ respectively.

[10] This describes the fraction of B mesons which will have a particular decay chain.

[11] With the current data, it is more precise to say the $B \to K\pi$ system *allows for* rather than it being a *source of* NP.

observeables is known as the "$B \rightarrow K\pi$ puzzle". The most common method of reconciling this tension is by introducing NP particles which act as intermediary states in the decay. Some prime candidates are the modified boson Z', diquarks and leptoquarks, the latter of which could also contribute to neutrino masses (addressing another limitation of the SM) [14].

This section begins with a detailed description of the $B \rightarrow K\pi$ puzzle, and then two pieces of this puzzle will be explored. The first piece is the statistically significant (defined to be beyond 5σ in Particle Physics) deviation from the SM in direct $\mathcal{CP}$ violation. The second piece is the 2σ tension pertinent to the $B^0 \rightarrow K_S^0\pi^0$ decay mode.

1.4.1 Penguins: Details of the puzzle

A theoretical framework used to describe how particles will behave can be established with QFT. In particular, the decay amplitude stores information about the decay process (transition from initial to final state) in a complex valued matrix element, which can often by represented by Feynman diagrams. Tree level processes give first order amplitudes, making these processes dominant (with some exception). By introducing 'loops' (see for example Figure 1.3) in the diagram, second order effects can be accounted for with Feynman diagrams known in the literature as 'penguins'. Depending on whether a gluon or $W^\pm$ boson runs in the loop, these are known as gluonic or electroweak (EW) penguins.

In this system, six possible Feynman diagrams contribute to the overall amplitude: the top (up) mediated gluonic penguins P_{tc} (P_{uc}), the colour-allowed (-suppressed[12]) tree amplitudes T (C), and EW penguins P_{EW} (P_{EW}^C). An example of the corresponding Feynman diagrams for P_{EW}^C and P_{uc}/P_{tc} is shown in Figure 1.3. The T amplitude requires a direct transition of quarks $b \rightarrow u$ ($\bar{b} \rightarrow \bar{u}$), but the corresponding CKM matrix element $|V_{ub}|$ is extremely small, suppressing this amplitude. The penguin decay P_{tc}, on the other hand, becomes the leading order diagram. Furthermore, EW penguins play a vital role in the puzzle. The four decays can be grouped by whether or not colour-allowed EW penguins contribute to the overall amplitude [15, 16]:

- Only P_{EW}^C contribute – shows *consistency* with the SM ($B^0 \rightarrow K^+\pi^-$, $B^+ \rightarrow K^0\pi^+$)

- Both P_{EW} and P_{EW}^C contribute – shows *tension* with the SM ($B^0 \rightarrow K^0\pi^0$, $B^+ \rightarrow K^+\pi^0$)
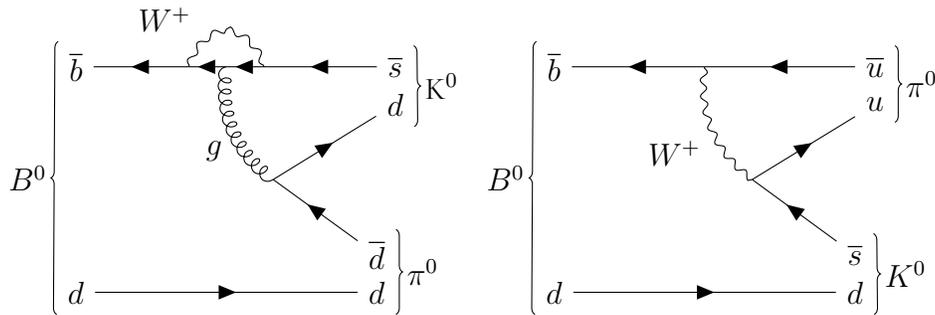
Figure 1.3: Feynman diagrams of $B^0 \rightarrow K^0\pi^0$ processes. Normally these loop processes are relatively minor compared to the tree level. However, due to the suppression of $|V_{ub}|$, the gluonic (left) and EW (right) penguins dominate the decay amplitude.

---

[12]Colour suppression occurs since each quark has an associated colour (red, green, blue) or anticolour charge. The pair of quarks which make up a meson must have colour-anticolour charge in order to make it overall colourless. During decays, if a particular meson final state comes from the *same* meson, any colour combination is allowed. However if they come from *different* mesons, the colours must match reducing the possible number of combinations allowed, thus restricting the amplitude.

1.4.2 Direct $\mathcal{CP}$ Violation of $B^0 \to K^+\pi^-$ and $B^+ \to K^+\pi^0$

Initially it was believed that the difference in direct $\mathcal{CP}$ violation between $B^0 \to K^+\pi^-$ and $B^+ \to K^+\pi^0$ was zero in the SM. However, a non-zero value was observed [13]:

$$\Delta\mathcal{A}_{CP} \equiv \mathcal{A}_{CP}(B^0 \to K^+\pi^-) - \mathcal{A}_{CP}(B^+ \to K^+\pi^0) = (12.2 \pm 2.2)\% \tag{1.30}$$

at a 5.5σ from the SM expectation of $\Delta\mathcal{A}_{CP} = 0.0$.

1.4.3 $\mathcal{CP}$ Violation via interference for $B^0 \to K^0\pi^0$

In the following section, all parameters and formulae are sourced from Ref. [17, 18] unless otherwise stated. For the $B^0 \to K^0\pi^0$ decay, the expected direct $\mathcal{CP}$ violation can be calculated via the sum rules. $\mathcal{S}_{CP}$ on the other hand, has a dependence on $\mathcal{A}_{CP}$:

$$\mathcal{S}_{CP} = \sin(\phi_d - \phi_{00})\sqrt{1 - \mathcal{A}_{CP}^2} \tag{1.31}$$

where the B^0-$\bar{B}^0$ mixing phase[13] is $\phi_d = (43.2 \pm 1.8)°$. ϕ_{00}, on the other hand, involves calculations of isospin relations with hadronic parameters. The expected $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ in the SM is plotted in Figure 1.4 along with their 1σ bands and quoted in Table 1.1. Here it is shown that the tension between the SM and observation stand at approximate 2σ. This should be further explored with data to reduce the statistical uncertainty at experiments such as Belle II.

Relation between K^0 and K^0_S

The focus of the discussion thus far has been on $B^0 \to K^0\pi^0$. However the K^0 meson exists in two flavour eigenstates: the (relatively) short-lived K^0_S and the long-lived K^0_L. They are both $\mathcal{CP}$ eigenstates with different eigenvalues (ξ):

$$\mathcal{CP}\left|K^0_S\right\rangle = -\left|K^0_S\right\rangle, \quad \mathcal{CP}\left|K^0_L\right\rangle = \left|K^0_L\right\rangle \tag{1.32}$$

It turns out that due to contributions from the colour-suppressed tree amplitude, $\mathcal{S}_{CP} = -\xi\sin 2\phi_{eff}$, where ϕ_{eff} is a shifted value of the unitary angle ϕ_1 [19]. Measurement of $\mathcal{S}_{CP}$ in $B^0 \to K^0_L\pi^0$ is beyond reach of current experimental setups as the K^0_L decays far outside the sensitive regions of any conceivable detector. However, since a measurement of the $\mathcal{S}_{CP}$ of $B^0 \to K^0\pi^0$ is equivalent to $B^0 \to K^0_S\pi^0$, this latter decay can be used to estimate the time-dependent $\mathcal{CP}$ violating parameter.

The $\mathcal{A}_{CP}$ of K^0_S and K^0_L mode is the same, so in order to get the direct $\mathcal{CP}$ violation of $B^0 \to K^0\pi^0$, the measurements of both modes are averaged. Since the $B^0 \to K^0_S\pi^0$ is sensitive to both $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ measurements, it is an extremely attractive probe into $\mathcal{CP}$ violation in $B^0 \to K^0\pi^0$.

1.4.4 Previous Results: Belle and BaBar

This decay channel has been thoroughly explored by the Belle and BaBar experiments. The results they obtained can be found in Table 1.1. The analysis was performed at Belle [19] with a 3D fit to M_{bc}, ΔE, and ΔT (see Section 3.1.10). In this analysis we introduce a fourth fitting variable which is a logarithmic transformation of the continuum suppression variable (see Section 3.1.9). Furthermore, modifications were made to the binning models (see Section 4.1). For these reasons, and the far larger dataset Belle-II is expected to collect, we hope to resolve an aspect of the puzzle by 2023 (see Section 4.3).

---

[13]This is the phase difference between the decay amplitude and its $\mathcal{CP}$ conjugate.

[14]The value of $\mathcal{S}_{CP}$ depends on $\mathcal{A}_{CP}$ which was set to $+0.14$. The theoretical uncertainty was estimated with the values provided in Ref. [21] and the updated global experimental uncertainty on $\mathcal{A}_{CP}$. No attempt to fully replicate the error analysis
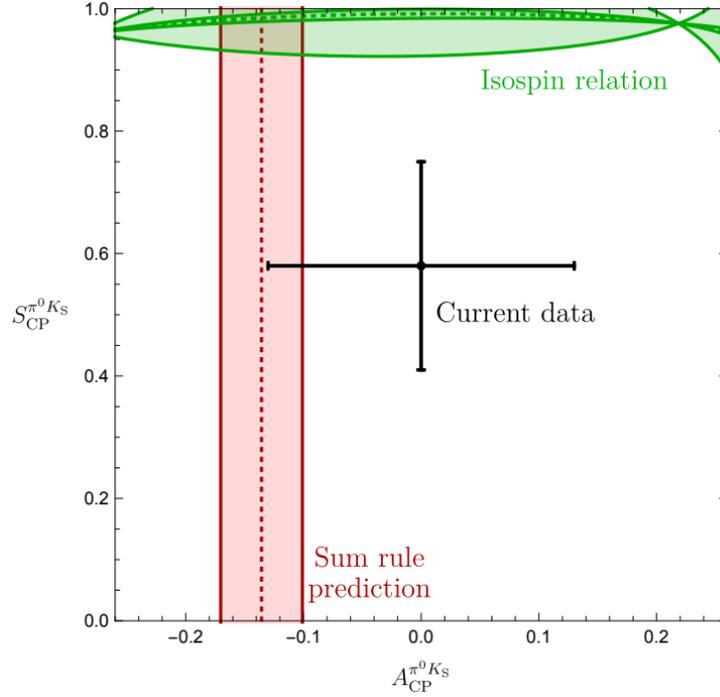
Figure 1.4: The dotted red line is the $\mathcal{A}_{CP}$ calculated from the sum rule, and the dotted green line is the $\mathcal{S}_{CP}$ ascertained from the isospin rule. Their 1σ uncertainty bands are shown. Both are calculated under SM assumptions. The results observed from current data is also shown with their 1σ error bars. Credit: [17]

Source	$\mathcal{A}_{CP}$	$\mathcal{S}_{CP}$
Belle [19]	$+0.14 \pm 0.13$ (stat) ± 0.06 (sys)	$+0.67 \pm 0.31$ (stat) ± 0.08 (sys)
BaBar [20]	-0.13 ± 0.13 (stat) ± 0.03 (sys)	$+0.55 \pm 0.20$ (stat) ± 0.03 (sys)
Global Average [2]	0.00 ± 0.13 (expt)	$+0.58 \pm 0.17$ (expt)
SM Expectation [17]	-0.14 ± 0.03 (theory)	$+0.99 \pm 0.13$ (theory)[14]

Table 1.1: Previous results from Belle and BaBar, where (stat) and (sys) are the statistical and systematic uncertainties respectively. The global average and theoretical SM values are shown with their uncertainties.

---

was made by the author, as such the theoretical uncertainty may differ from the bands shown in Figure 1.4. This calculation has been included for completeness.

Chapter 2

Belle II Experiment

Having motivated the measurement in the $\mathcal{CP}$ violating parameters of $B^0 \to K^0_S \pi^0$, this Chapter outlines the Belle II experiment exploited for this purpose.

2.1 B-Factory: SuperKEKB Accelerator

In order to make such a measurement, a source of numerous B-mesons is required. The SuperKEKB accelerator in Tsukuba, Japan hosts a "B-factory" built for this purpose. It accelerates electrons to an energy of 7 GeV in its High Energy Ring (HER) and 4 GeV positrons in its Low Energy Ring (LER), and they collide at an Interaction Point (IP) to produce $\Upsilon(4S)$ resonance (made of $b\bar{b}$) which decays into a pair of either charged or neutral coherent B-mesons. These leptons are accelerated around a beampipe, where the forward direction is defined as the direction momentum of the e^-. The asymmetric collision energy gives a boost to the lifetime of these mesons in the lab-frame so that time-dependent measurements can be conducted with precision.

At the IP, the Belle-II detector is built so that the decay products of these B-mesons can be observed. In its entire lifetime, Belle, the predecessor of Belle-II, obtained an integrated luminosity[1] of up to $1\,\mathrm{ab}^{-1}$. Belle-II is expected to reach at least forty times this amount of data at the conclusion of the experiment [22].

2.2 Sub-Detectors of Belle-II

Beginning with the innermost (closest to the beampipe) components, this section will describe the functionality and purpose of each sub-detector [23].

2.2.1 Pixel Vertex Detector (PXD) and Silicon Vertex Detector (SVD)

The innermost component of the Belle-II detector is comprised of two layers PiXel Vertex Detectors (PXD) made from silicon using DEPleted Field Effect Transistor (DEPFET) technology. Its highly granular nature allows for precise vertexing of the B-mesons. Just outside of this lies four layers of silicon Strip Vertex Detectors (SVD) which provides less precise[2] information. As charged particles pass through these detectors, the pixels/strips will be triggered to a readout placed outside the sensitive region. This provides vertex information which is vital for a time-dependent analysis.

---

[1]The integrated luminosity $\mathcal{L}$ represents the amount of data taken. Combined with the cross section of producing the $\Upsilon(4S)$ resononce σ at an experiment, one can determine the number of $B\overline{B}$ produced: $N_{B\overline{B}} = \mathcal{L}\sigma$.

[2]Multiple considerations including cost factored into the 2 PXD and 4 SVD design in Belle II.
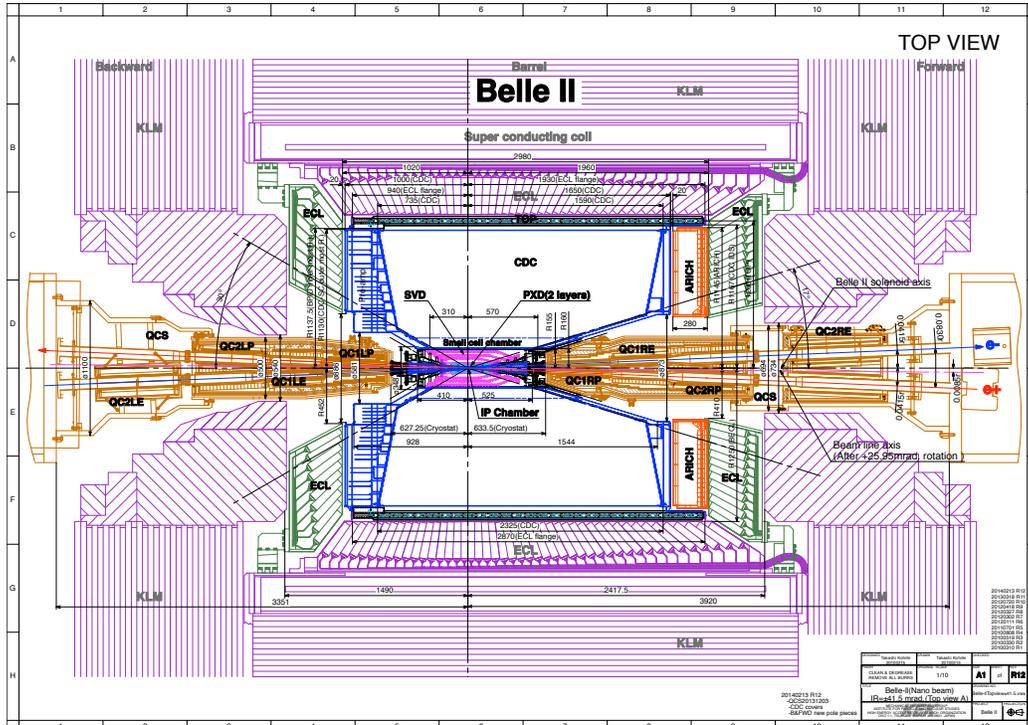
Figure 2.1: The Belle-II detector with the central horizontal orange cylinder containing the beampipe, and the leptons are made to collide at the centre of the diagram in the IP chamber. The inner detector with the PXD and SVD is shown in pink. The CDC is shown in blue, with PID components TOP and ARICH surrounding it. The ECL is highlighted in green. Finally the outer purple detector is the KLM. The central area is known as the barrel region, the right and left sides are the forward and backward endcap regions respectively (labelled on the diagram).

2.2.2 Central Drift Chamber (CDC)

The next sub-detector is the Central Drift Chamber (CDC), which serves the two-fold purpose of reconstructing charged tracks (and their momenta), and identifying particles (according to energy deposits). The latter is particularly helpful for particles which decay before reaching outer particle identification parts of Belle-II. The CDC is filled with Helium-Methane gas and constructed with wires in such a way that 3D particle information can be reconstructed.

2.2.3 Particle Identification (PID)

The purpose of the Particle Identification (PID) detectors is to primarily distinguish between kaons and pions. In the barrel region, the particles passing through the Time-of-Propagation (TOP) counters emit Cherenkov photons which are internally reflected at different Cherenkov angles for pions and kaons resulting unique propagation times. From these arrival times of these photons at the end of the quartz bar, particle type can be distinguished. In the end-cap region, as these particles pass through Aerogel Ring-Imaging CHerenkov (ARICH) detectors, the Cherenkov photons cause rings of different radii to be emitted which is used as a discriminating feature. Note the ARICH can discriminate between Pions, Muons or low energy electrons.

2.2.4 Electromagnetic Calorimeter (ECL)

The Electromagnetic Calorimeter (ECL) is the next layer, where charged particles and photons would interact with the scintillating material causing additional photons to be emitted. This is known as an electromagnetic 'shower' as a signal particle can cause a cascade of photons which are detected and recorded. The ECL serves many purposes, which includes energy deposit measurements, angular information detection, and particle identification. The ECL is the primary means of luminosity measurements. The ECL covers both the barrel and the endcap areas.

2.2.5 K_L^0 and μ detector (KLM)

The outermost component of Belle-II is the K-long and Muon (KLM) detector used to identify long lived kaons K_L^0 and μ. This iron component also provides additional decay length for these longer living particles. The KLM component will not be particularly important for this analysis.

2.3 Event Type

We define each detected e^+e^- collision as an event. Although the beam energies are set to maximise the production of $\Upsilon(4S)$, not every event will be successful. $\Upsilon(4S)$ almost always[3] decays into a pair of coherent B-mesons. The different features of the decay products of these mesons are recorded in each sub-detector as the particles pass through it. Each event can be described by the set of tracks (reconstructed objects) which represent the decay chain B-mesons followed. However not all events are deemed interesting, so on both a hardware and software level, a majority of events are veto-ed by the trigger system. Yet, there remains very significant background. It is beneficial to classify the types of events detected by Belle-II:

1. **Signal** – the decay pertinent to this analysis ($e^+e^- \to \Upsilon(4S) \to B^0/\overline{B}^0 \to K_S^0\pi^0$)
2. **Continuum Background** – $e^+e^- \to q\overline{q}$ (where $q = u, d, s, c$) which hadronises into streams of mesons
3. $B\overline{B}$ **Background** – coherent B-mesons that decay via some *other* decay channel. Particular combinations of tracks could lead one to misidentify this as a signal event. (Note that this can also be known in literature as combinatorial background, but will hitherto be referred to as "BB background").

The discussion of backgrounds has been at an event-level so far. However, each event may contain tracks whose source are not e^+e^- collisions, but derive from the general radioactive environment of the active accelerator, or cosmic radiation referred to as **Beam Background**

2.4 Belle-II Analysis Software Framework

All analysis tasks for this analysis were done within Belle-II Software Analysis Framework (`basf2`) [24] which is built off `C++` modules and `python` steering scripts. Unless otherwise stated, `release-03-02-04` of this software was used. In addition CERN's analysis package `ROOT` [25] was exploited for the analysis.

2.5 Reconstructed Variables

Reconstruction begins with the final state particles. The reconstruction process takes raw detector-level information (eg. hits in the CDC), and determines the precise decay chain a particular signal event followed, as well as a particle's kinematic information. By taking all viable combinations of their four-momenta, intermediate particles are reconstructed. This process is repeated until information about the B candidate is

---

[3]More than 96% of the time [2].

found. The B-meson which is fully reconstructed[4] is the signal $B_{sig}(\to K_S^0 \pi^0)$. The other coherent B_{tag} is useful for flavour tagging (see Section 3.1.5).

A multiplicity of candidates for B_{sig} is found by combining four momenta, where only one of them can represent the true B_{sig}. To reduce the multiplicity of candidate for each event, it is beneficial to introduce two variables for the reconstructed B-meson: ΔE and M_{bc}.

2.5.1 The Energy Difference ΔE

Each B-mesons should contain half the total energy of $\Upsilon(4S)$ in its Centre-of-Mass (COM) frame. Given a set of B_{sig} their mean energy should be equal to $E_{beam} = \frac{1}{2}E_{\Upsilon(4S)} = 5.29\,\text{GeV}$. Any B candidate which has strong deviation from this is likely to be incorrectly reconstructed. We define the difference in energy as:

$$\Delta E = E_{beam} - \sum_i E_i \tag{2.1}$$

where E_i is the energy of the i-th daughter particle of B_{sig}. A collection of true B_{sig} should have a ΔE distribution centred at zero.

2.5.2 The Beam-Constrained Mass M_{bc}

The beam-constrained mass is another variable to assess the performance of the B_{sig} reconstruction. Rather than using the energies of the daughters, their momenta along with E_{beam} is used to calculate the following:

$$M_{bc} = \sqrt{E_{beam}^2 - \left(\sum_i \vec{p}_i\right)^2} \tag{2.2}$$

where $\vec{p}_i$ is the three-momenta of the i-th daughter of B_{sig}. A set of B_{sig} events should have M_{bc} peaking at $m_B = 5.2785\,\text{GeV}\,c^{-2}$.

A correction to M_{bc}

In the signal decay, high energy $\pi^0 \to \gamma\gamma$ creates electromagnetic showers which are not fully contained in the clusters of the ECL. This causes the reconstructed pion energy, and thus the momentum of the π^0 to be underestimated. A low energy tail in the M_{bc} can appear which does not reflect reality. Furthermore, for the $B^0 \to K_S^0 \pi^0$ analysis, although M_{bc} and ΔE are required to be independent[5], a Pearson correlation[6] [26] of 0.11 was found between them. These problems can be mitigated by making a correction to the M_{bc}, using the known mass of π^0 and energy of K_S^0 as constraints:

$$\vec{p}_{corr} = \vec{p}_{Ks} - \frac{\vec{p}_{\pi_0}}{|\vec{p}_{\pi_0}|}\sqrt{(E_{beam} - E_{Ks})^2 - m_{\pi_0}^2} \tag{2.3}$$

$$M_{bc}^{corr} = \sqrt{E_{beam}^2 - \vec{p}_{corr}^2} \tag{2.4}$$

Note that a similar correction could have been made in ΔE to the same effect. The reduction in correlation and removal of the low energy tail in M_{bc} is shown in Figure 2.2.

---

[4]From the conditions we set in analysis, we ensure the signal-B is the one which is fully reconstructed

[5]Independence is an important assumption we require in the fitting stage (see Section 3.1.10) of the analysis

[6]A Pearson correlation of 0 (± 1) indicates no (maximum) correlation between the two input variables.
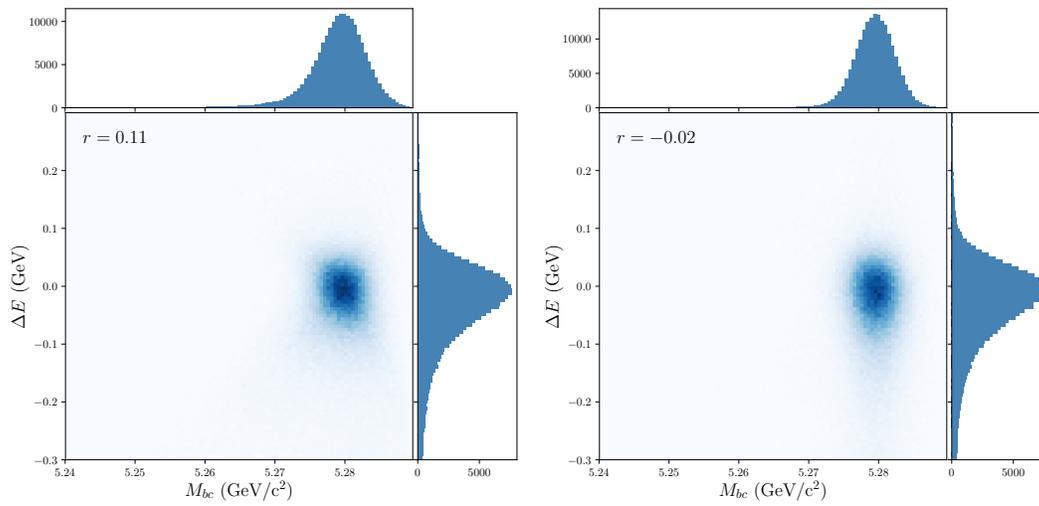
Figure 2.2: A comparison of the correlation of M_{bc} and ΔE before (Left) and after (Right) making a correction on the M_{bc}. The correlation was reduced and the low energy tail is less apparent after making this correction.

Chapter 3

Physics Analysis

The process of making inferences about physical observable from raw Belle-II data is highly non-trivial. This task of a physics analysis extracts five *parameters*, two $\mathcal{CP}$ violating parameters $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$, and yields of each event type N_{sig}, N_{cont} and N_{BB}. From these we can ascertain three *observables* of interest: $\mathcal{A}_{CP}$, $\mathcal{S}_{CP}$ and the branching ratio[1]. Of these, we focus on $\mathcal{S}_{CP}$ as $B^0 \to K_s^0 \pi^0$ is the only decay that can make this measurement. Physics analysis can be broken down into three distinct stages (see Figure 3.1):

1. **Build** analysis workflow
2. **Validate** workflow for different $\mathcal{CP}$ violating parameters
3. **Measure** parameters of interest using the workflow
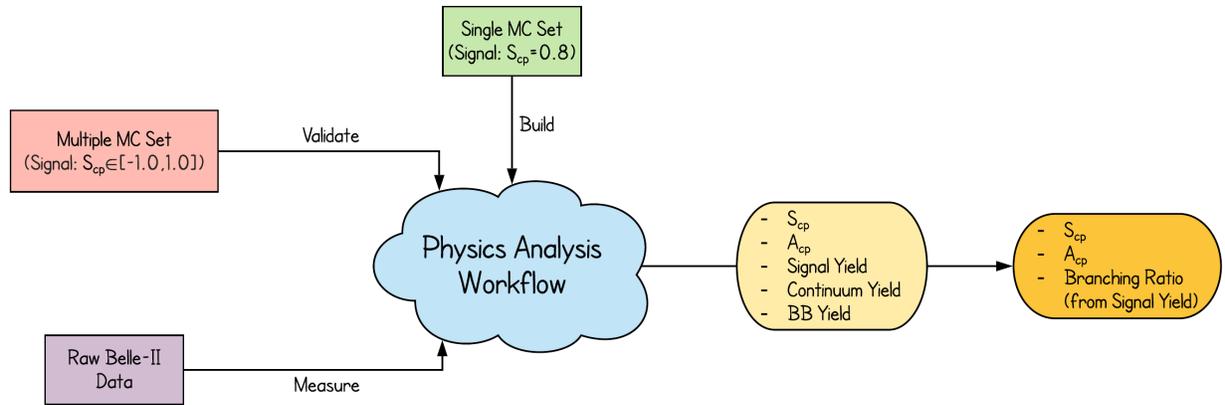
Figure 3.1: The methodology of physics analysis which involves building, and validating the physics analysis workflow or model. In the latter two, parameters of interest can be extracted either for verification or measurement.

Steps 1 & 2 utilise Monte-Carlo (MC) simulations where events are generated and simulated in the Belle-II experiment with software packages rather than actual data. MC simulations require $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ as input parameters. By validating with MC data before viewing real data, this "blind analysis" technique reduces experimenter's bias and maintains scientific integrity [27]. The implementation for measurement and validation can be expressed as a series of sequential steps:

1. **Skim Data** – reconstructing B_{sig} and keeping events which pass a set of broad selection criteria

---

[1]Obtained from signal yield (Equation 3.16)

2. **Reconstruct Rest of Event** – tracks not used in B_{sig} reconstruction are used to reconstruct the B_{tag} (other coherent B)

3. **Vertexing** – determining the decay vertices of the coherent B-mesons, calculating their positional displacement Δz (where the positive direction of the z-axis is defined as the forward direction of the beampipe) and difference in decay time (resolution) ΔT

4. **Flavour Tagging** – ascertaining the flavour (with associated confidence) of the B_{tag}, then parameterising this information in a single variable qr

5. **Continuum Suppression** – using event-shape variables and multivariant techniques, determine the probability a particular event is signal/continuum. The result of this undergoes a logarithmic transformation to define a fitting variable: `logSupp`

6. **4D Simultaneous Fit** – a fit is conducted simultaneously on the distribution of each of the four fitting variables: M_{bc}, ΔE, ΔT, and `logSupp` to extract the parameters of interest

Data is progressively removed via a set of selection criteria given in Table 3.1, which has the intended effect of maximising signal acceptance and minimising background rejection.

Section 3.1 discusses how the above model is built. In particular, it outlines how the weights for continuum suppression and PDF parameters for the simultaneous fit are determined. For the remainder of this thesis, a physics model is defined by a set of user-specified choices when building the analysis (eg. which vertexing algorithm is used). Section 3.2 is outlines the validation process for each physics model that is tested. Section 3.3 explores a complementary analysis, the $B^0 \rightarrow J/\psi K_S^0$ decay, which is used as a control mode to estimate systematic uncertainties in the signal mode.

3.1 Building the Analysis Workflow

3.1.1 Generating MC Signal Events

The model building stage begins by generating[2] $\Upsilon(4S) \rightarrow B^0 \overline{B}^0$ events and with decay amplitude calculations, we determine the kinematic information about the intermediate and final states. This is done in the `EvtGen` package [28]. Here one of the neutral B-mesons are forced to decay as $B_{sig} \rightarrow K_S^0 \pi^0$ and the other B_{tag} decays generically (according to their appropriate decay amplitudes). These events are then simulated in the Belle-II detector environment via the `GEANT4` package [29], which gives a set of detector-level information (eg. hits in the CDC) for each event. One million of these events with $\mathcal{S}_{CP} = 0.8$, $\mathcal{A}_{CP} = 0.0$[3] was generated and simulated. This dataset will be referred to as signal MC.

Continuum and BB background events are common across all analyses so samples are available collaboration wide. This analysis used campaign `MC11x2` MC data which simulates the background expected in 1 ab^{-1} of data-taking.

3.1.2 Reconstruction and Skimming

In Section 2.5 recall that reconstructed variables can be used to remove unlikely B candidates. For this analysis we set the following requirements $M_{bc} > 5.2\,\mathrm{GeV}\,c^{-2}$ and $|\Delta E| < 0.3\,\mathrm{GeV}$, which has the effect of

---

[2]More specifically, given known energies and momenta of e^+e^-, one can, with a level of statistical/random (MC) fluctuation, calculate the kinematics of the $\Upsilon(4S)$, and with known decay amplitudes, determine the kinematics of its daughters also.

[3]The input $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ used to **build** the model should not affect performance. However for particular values (ie. $\mathcal{A}_{CP} = 0.0, \mathcal{S}_{CP} = 0.67$), significant bias was observed. That is, for a given input $\mathcal{S}_{CP}$, the extracted $\mathcal{S}_{CP}$ was vastly different (outside of statistical uncertainty). The model building input $\mathcal{CP}$ violating parameters were selected to ensure little bias is observed in the validation stage.

eliminating a large amount of unnecessary information. The distributions of these variables are shown in Figure 2.2. The process of applying this broad criteria is known as skimming.

Skim Efficiency

A time-dependent measurement can only be made if the signal decay chain is as follows:

$$B^0 \to K_S^0(\to \pi^+\pi^-) \, \pi^0(\to \gamma\gamma) \tag{3.1}$$

This is because charged final states are required for vertexing and $K_S^0 \to \pi^+\pi^-$ is one of the few decay chains of K_S^0 that fulfills this condition. Unfortunately, this has a branching fraction of only 0.69. On the other hand, $\pi^0 \to \gamma\gamma$ almost always occurs. Each charged pion and photon has an 80% chance of being properly reconstructed, and the K_S^0 has a 70% probability of being reconstructed from properly reconstructed pions. This places an upper limit on the number of signal events that can survive reconstruction:

$$0.69 \times 0.8^2 \times 0.8^2 \times 0.7 = 0.24 \tag{3.2}$$

It turns out the actually efficiency is far lower than this (see Table 3.3) due to further selection criteria outlined in Table 3.1.

Determining Selection Criteria

Further constraints can be applied to the selection criteria for improved background removal. This method provides rather weak discrimination, however it can be a good baseline. A more restrictive set of selection criteria gives a good signal purity[4] where as looser conditions will result in a good efficiency[5]. This tradeoff is often parameterised by the Figure of Merit (FOM) calculated from S and B, the number of signal and background events, respectively:

$$FOM = \frac{S}{\sqrt{S + B}} \tag{3.3}$$

which satisfy a set a particular set selection criteria. To find this set, a grid search method divides parameter space into discrete blocks. The FOM is calculated for each block, and the one which corresponds to the highest value is selected. This optimisation was already done in previous analyses resulting the requirements outlined in Table 3.1 for the signal decay. Here we consider $B^0 \to J/\psi K_S^0$ which is used as a control mode (see Section 3.3). The parameters M_{bc}, ΔE and the reconstructed mass of the $M_{J/\psi}$ were each divided into discrete sections[6] and after applying a grid search, the best cut is outlined in Table 3.1.

Since more sophisticated continuum suppression techniques are used at a later stage, it is important to keep as many signal events as possible. To that end, a signal acceptance $> 90\%$ is required. A visualisation (projection) of the FOM for different cut values is shown in Figure 3.2.

### 3.1.3 Rest of Event and Clean Mask

Once appropriate tracks have been used to reconstruct the B_{sig} the tracks in the Rest of Event (ROE) are often associated with the coherent B_{tag}. However, many of these tracks can also be assigned to beam background. As such, a set of requirements[7] known as a "clean mask" is applied on the track kinematics. The remaining

---

[4]Out of the events which survive the cuts, the proportion which are signal events.

[5]Out of the signal events, the proportion which survive the cuts.

[6]The set of cut values are: $M_{bc} = \{5.23, 5.24, 5.25, 5.26, 5.27\}$, $\Delta E^{lower} = \{-0.3, -0.275, -0.25, -0.225, -0.2\}$, $\Delta E^{upper} = \{0.2, 0.225, 0.25, 0.275, 0.3\}$, $M_{J/\psi}^{lower} = \{2.6, 2.7, 2.8, 2.9\}$, $M_{J/\psi}^{upper} = \{3.2, 3.3, 3.4, 3.5\}$. Energy units are GeV and mass units are GeV/c^2

[7]In this analysis the clean mask applied the following conditions: At least one hit in the CDC, track momentum at least $0.05\,\text{GeV}\,c^{-1}$, and the centre of mass track momentum $<3.2\,\text{GeV}\,c^{-1}$.
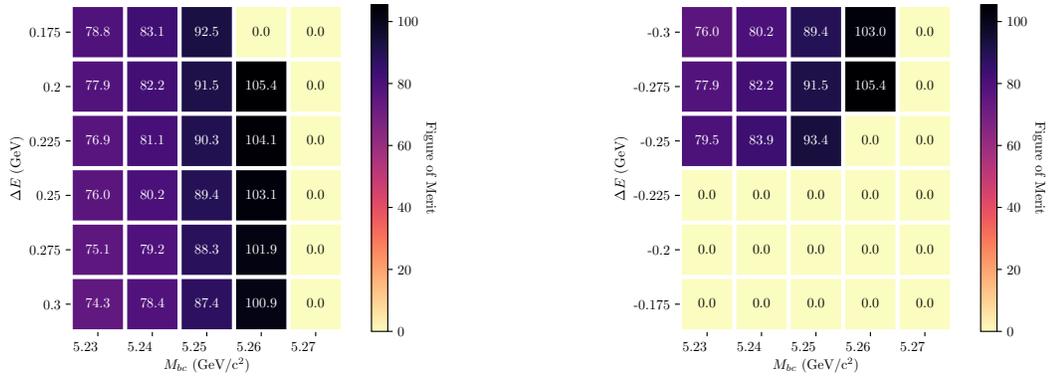
Figure 3.2: The grid search method gives a value of FOM for each discrete section of a 5-dimensional parameter space. A projection of this onto $M_{J/psi} \in [2.8, 3.3]$ GeV c^{-2}, $\Delta E > -0.275$ GeV (left) or $\Delta E < 0.3$ GeV (right) is shown. Zero FOM regions are associated with signal acceptance falling below the required 90%.

tracks are used to reconstruct the tag-side B_{tag}.

3.1.4 Vertexing

The $B_{sig} \to K_S^0 \pi^0$ decay products are neutral, so determining its decay vertex directly is impossible. $\pi^0 \to \gamma\gamma$ gives neutral final states, so this cannot contribute to vertexing either. K_S^0, however, decays to charged pions a majority of the time. As such its decay vertex can be reconstructed. If this process is successful, we can then use the momentum and position of the K_S^0 to extrapolate the decay position (z_{sig}) of B_{sig}. This process is known as applying the IP Tube constraint. Since the mass of intermediate particles K_S^0 and π^0 are known, an additional mass constraint may be applied to improve the position of vertexing. Two vertexing algorithms are used in this analysis:

1. **VertexRAVE** – a geometric fitter which does not include the mass constraint. It has been used in previous analysis, but the software is no longer maintained, so this serves as a good baseline.
2. **TreeFitter** [30] – a kinematic fitter which provides a global fit allowing both constraints to be used.

The generic decay of the tag-side often leaves charged tracks so its vertexing is not normally problematic. A constraint similar to IP Tube is applied to find the decay vertex (z_{tag}) of B_{tag}. The difference in decay length is then $\Delta z = z_{sig} - z_{tag}$ which can be transformed[8] into a decay time (resolution) via $\Delta z = \beta\gamma c\Delta T$.

3.1.5 Flavour Tagging

Since B_{sig} decays into $\mathcal{CP}$-eigenstates, its daughters give no indication of the B_{sig} flavour. B_{tag}, on the other hand, decays generically, so it is possible for this flavour to be determined in a process known as 'flavour tagging'. By accounting for meson mixing, and due to the fact that B_{sig} and B_{tag} are coherent, the flavour of B_{sig} can also be ascertained. The result of flavour tagging gives a variable q which is $+1$ (-1) if B_{tag} is B^0 ($\bar{B}^0$). The confidence of the flavour tagger is parameterised by the dilution factor $r \in [0, 1]$ which gives 1 for an unambiguous flavour assignment.

A correct assignment of flavour is vital for the time-dependent measurement as outlined in Equation 1.29. The quality of the tagger is accounted for by placing data in bins of qr according to their dilution factor which is detailed in Section 3.1.10.

---

[8] The boosting in Belle-II is known to be $\beta\gamma = 0.28$ [23]

3.1.6   Continuum Suppression

Continuum suppression is the process of distinguishing between signal and continuum events[9]. To this end, multivariant techniques trained on event shape or topology was used. The shape of an event (Figure 3.3) can be a powerful discriminator, and this is parameterised by a set of 30 topological variables (see Section 3.1.7 for more details):

- `cosTBTO` and `cosTBz` – angle between the thrust axis of B_{sig} and the thrust axis of the ROE, and beampipe respectively

- `thrustBm` and `thrustOm` – magnitude of the B_{sig} and ROE thrust respectively

- `R2` – reduced Fox-Wolfram moment

- `KSWF_hso{00,02,04,10,12,14,20,22,24}` – Signal KSFW

- `KSWF_hoo{0,1,2,3,4}` – ROE KSFW

- `KSWF_{mm2,et}` – missing mass and transverse energy

- `Cleocone{1-9}` – momentum flow along thrust axis in discrete cones of $10°$
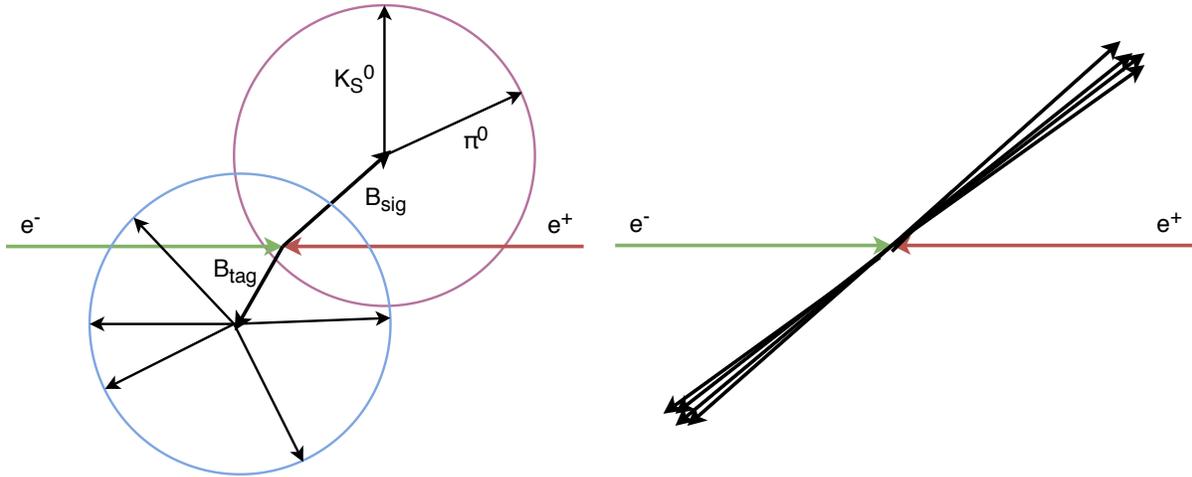
Figure 3.3: A signal event (left) has a more 'spherical' topology, whilst the continuum event (right) has more of a 'jet-like' shape of collimated streams of bounded quarks. Both are in the COM frame.

The following two sections will continue with a detailed discussion of continuum suppression. Section 3.1.7 describes the above features in detail and Section 3.1.8 outlines two multivariant techniques used.

3.1.7   Event Shape Variables

Angular and Thrust Variables

The thrust of a particle is defined as

$$T = \frac{\sum_{i=1}^{N} |\hat{n} \cdot \vec{p}_i|}{\sum_{i=1}^{N} |\vec{p}_i|} \tag{3.4}$$

---

[9]BB background is notably absent from this process as its signature is too close to signal events. Since they make up a small fraction of the total background, it is far easier to simply model them using PDFs later than to attempt to completely remove them.
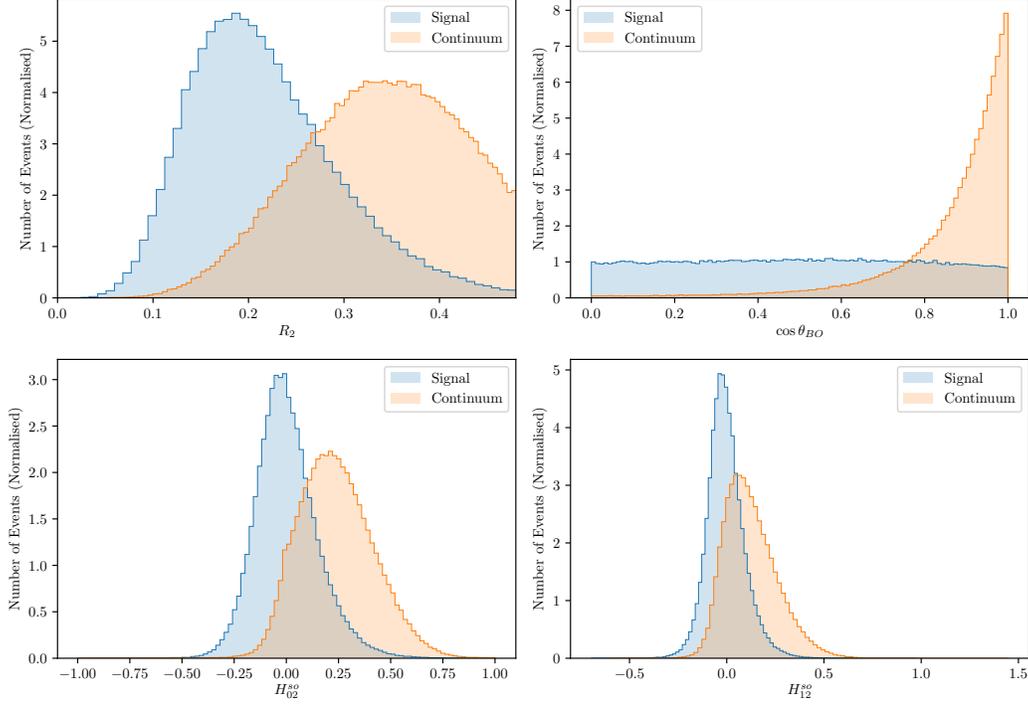
Figure 3.4: Comparing the distribution of signal and continuum in the top 4 most discriminating variables, in the order of: `cosTBTO`, R2, `KSFW_hso02`, `KSFW_hso12`.

where the thrust axis $\hat{n}$ is the unit vector which maximised T, and i runs over all (N) daughters of B_{sig} and B_{tag} for `thrustBm` and `thrustOm` respectively. `cosTBTO` gives the cosine of the angle between the thrust axes of B_{sig} and the B_{tag} whereas `cosTBz` does the same, but replaces B_{tag} with the forward direction of the beam pipe.

Reduced Wolfram Moment (R2)

Fox and Wolfram [31] formulated a rotationally invariant observable which can be calculated from QCD perturbation theory. The Fox-Wolfram moment is:

$$H_l = \sum_{ij} \frac{|\vec{p}_i||\vec{p}_j|}{s} P_l(\cos \varphi_{ij}) \tag{3.5}$$

where i, j runs over all reconstructed charged tracks, and $\vec{p}_{\{i,j\}}$ are their associated momenta. The s is the squared centre-of-mass energy[10], and P_l is an l-th order Legendre polynomial. φ_{ij} gives the angle between the i-th and j-th reconstructed tracks. The ratio of the Fox-Wolfram moments is:

$$R_k = \frac{H_k}{H_0} \tag{3.6}$$

It turns out that R_2 is a very useful event shape variable for continuum suppression. Note that this is referred to as the reduced Fox-Wolfram moment in Belle-II.

---

[10]In Belle-II this is simply $2E_{beam}$

Kakuno-Super-Wolfram-Fox Moment: KSWF

The reduced Fox-Wolfram moment may be a great discriminating variable, however it does not contain all possible information about the event shape. Belle[11] developed a set of variables which accounts for the signal-side and tag-side information in calculating the Fox-Wolfram moments (up to 4-th order). These Kakuno-Super-Fox-Wolfram (KSWF) moments [22] are separated into three categories: signal H_l^{so} and tag H_l^{oo} moments, and missing momentum.

Signal Moments
The signal moments H_l^{so} considers both signal- and tag-side tracks. The tag side is subdivided into charged $x = 0$, neutral $x = 1$ or missing $x = 2$. The moments vanish for odd values of l[12]. For even l it is given by:

$$H_{xl}^{so} = \sum_a \sum_b |\vec{p_b}| P(\cos \varphi_{ab}) \tag{3.7}$$

where a runs over the signal-side tracks and b runs over the tag-side. This is normalised to $H_0^{max} \equiv 2(E_{beam} - \Delta E)$. The normalisation removes the dependence on ΔE. Giving a set of 9 moments (3 for each tag-side type): $H_{\{00,02,04,10,12,14,20,22,24\}}^{so}$

Tag Moments
The tag moments considers only tag-side tracks. Here the ratio of moments is taken. For even l:

$$H_l^{oo} = \sum_a \sum_b |\vec{p_a}||\vec{p_b}| P_l(\cos \varphi_{ab}) \tag{3.8}$$

For odd l:

$$H_l^{oo} = \sum_a \sum_b q_a q_b |\vec{p_a}||\vec{p_b}| P_l(\cos \varphi_{ab}) \tag{3.9}$$

Analogous to the signal moments, both are normalised to $(H_0^{max})^2$. This gives a set of 5 moments: $H_{\{0,1,2,3,4\}}^{so}$

Missing Momentum
The squared-missing-mass can be found by combining the missing energy and momentum:

$$M_{miss}^2 = \left(2E_{beam} - \sum_{n=1}^{N} E_n\right)^2 - \sum_{n=1}^{N} |p_n|^2 \tag{3.10}$$

where N is the number of daughter particles in the event. The transverse energy is the sum of the transverse momentum of each particle $\sum_{n=1}^{N} |(P_t)_n|$. This gives 2 variables for continuum suppression.

Cleocones

The CLEO collaboration introduced a set of variables which describe the momentum flow around the thrust axis of the B_{sig} at $10°$ intervals [32]. The i-th Cleocone describes the momentum flow in the region between $10°(i-1)$ and $10°i$. This gives a set of 9 variables to be used for continuum suppression

---

[11]Note that for Belle, all these variables were combined using a Fisher linear discriminant, however for Belle-II we feed the entire set directly into our MVA

[12]There is a charge dependence on the signal side, but since K_S^0 and π^0 are both neutral, this term vanishes.

Pak Hin (Brian) Chan

MSc Thesis

3.1.8 Multivariant Methods

Multivariant Analysis (MVA) is a method (classifier) which combines all 30 parameters into a single test statistic $y \in [-1.0, 1.0]$ where an event with $y = 1$ ($y = -1$) is classified as signal (continuum) with perfect confidence. The underlying principle for any MVA technique is quite similar – complex patterns in the event shape variables are used to determine weights. These weights can then be later applied to each event to calculate y. Building the continuum suppression is divided into these stages:

1. **Training** – observing a subset of the total available data to determine the weights of the MVA
2. **Testing** – using the weights to determine the y of the remaining events and assessing performance
3. **Implementing** – on the full dataset to find y for each event using pretrained weights

At this point, there are almost $500,000$ events which are valid for training[13]. Due to the relatively high statistics[14] 50% of this set is selected for training and the remainder is used for testing. Once trained and tested, the weights are implemented on the entire MC dataset (signal, countinuum, and BB). The Boosted Decision Tree (BDT) and Multi-Layer Perceptron (MVA) methods are explored in this section. All MVA implementation was done with the `TMVA` library [33] in `ROOT`.

Hyperparameter

Each MVA has user-specified parameters unique to the method known as hyperparameters. Variables such as the maximum depth of each tree or number of trees in the forest of BDT, or number of perceptrons in MLP all need to be tuned for maximum performance. This is a time consuming process in `TMVA`. Since these methods already perform well on this data (with AUC > 90), hyperparameters were not optimised in this analysis.

I. Boosted Decision Tree

Decision trees divide the feature space into signal and background regions. In their implementation, an event begins at the root (top) node where a single 'decision' (usually threshold on a single variable) has to be made, the result of which will determine the branch this event follows. This process is repeated until the final node is reached. Depending on which node the event is in, it is assigned to be either signal or continuum.

Decision trees are trained by placing the most discriminating variable threshold at the root of the tree. The Gini index [33] measures how well split the data for a particular decision tree. The nodes are chosen such that this metric is maximised. Since the features are a set of 30 continuous variables, a perfect separation of signal/continuum can theoretically be reached in the training data. However this will not be representative of the actual discrimination in the test data as signal regions will be localised to the neighbourhood of each point of the signal event. This leads to a loss in test performance in a process known as **overfitting**.

To prevent overfitting, we set a limit on the depth of a decision tree (define as maximum number of 'decisions' a particular event has to make in a tree – this is 3 for Figure 3.5). This leads to the creation of a relatively weak classifer which does not overfit. To increase the predictivitiy of the MVA, we employ a 'forest' of up to $N = 850$ weak trees in a process known as AdaBoost. Boosting involves taking a randomly selected (with repetition) 50% subset of the training data for each tree. This is adaptive in the sense that events which tend to be misclassified are more likely to be selected for subsequent trees, making the entire process Adaptive Boosting (AdaBoost). Along with a learning rate $\beta = 0.5$ each tree is weighted according

---

[13]This value depends on the vertexing method. The condition of validity depends on whether best candidate selection has been made (see Section 3.1.8: Best Candidate Selection).

[14]The train:test split ratio has to be balanced such that there is sufficient data for *both* training and testing. However the relatively high statistics ensures a 50:50 split gives high performance. An *ansatz* confirmed in Figure 3.7.
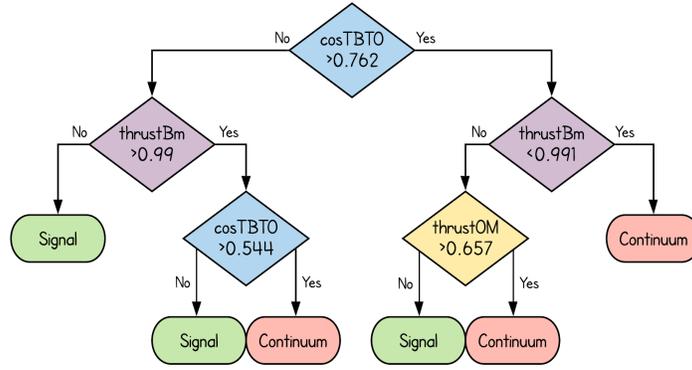
Figure 3.5: An example of a decision tree. Beginning at the top, whether an event is signal or background is determined by whether or not it passes the indicated selection criteria.

to how good their performance is (parameterised by the error rate[15] ϵ). The weight of each tree is given by α^β, where α is defined as:

$$\alpha = \frac{1-\epsilon}{\epsilon} \tag{3.11}$$

When implementing a BDT, events are passed through each tree giving a result[16] of h_i for the i-th tree. The output is then the weighted sum of these:

$$y = \frac{1}{N} \sum_{i=1}^{N} \ln(\alpha_i) \cdot h_i \tag{3.12}$$

II. Multi-Layer Perceptron

Another class of MVA uses perceptrons, where the set of 30 parameters $\vec{x} = (x_1, \ldots, x_n)$ is combined with a pre-determined set of weights[17] $\vec{w} = (w_1, \ldots w_n)$ via an inner product for each event $s = \vec{x} \cdot \vec{w}$. The result of this is passed through an activation function, which has the effect of scaling s to give the required range of the output y. Here we used $y = \tanh(s)$ as an activation function.

Perceptrons themselves are not very powerful classifiers. However, their outputs can be feed into other perceptrons creating a network known as a Multi-Layer Perceptron in an architecture known as artificial neural networks.

In order to train MLPs, a subset of the training data known as a 'batch' is passed through the neural net. The misidentified events are passed backwards (from output to input) adjusting the weights $\vec{w}$ along the way. An 'epoch' is reached when the entire dataset has passed through the neural network once. The MLP is trained over 5 epochs. Just like decision trees, MLPs are very prone to overfitting and methods such as early stopping when the performance of the network no longer improves are implemented to mitigate this.

Results of Continuum Suppression

The primary metric for the performance of an MVA classifier is the Area Under the Curve (AUC) of a Receiver Operator Curve (ROC). The output y is a continuous variable. As such one must set a threshold ζ such that

---

[15]Proportion of total event misidentified

[16]+1($-$1) for signal (continuum)

[17]An additional weight term called the 'bias' is present here, where the bias is independent of any of the input parameters. This allows us to shift the activation function (see text) to best optimised to our data.
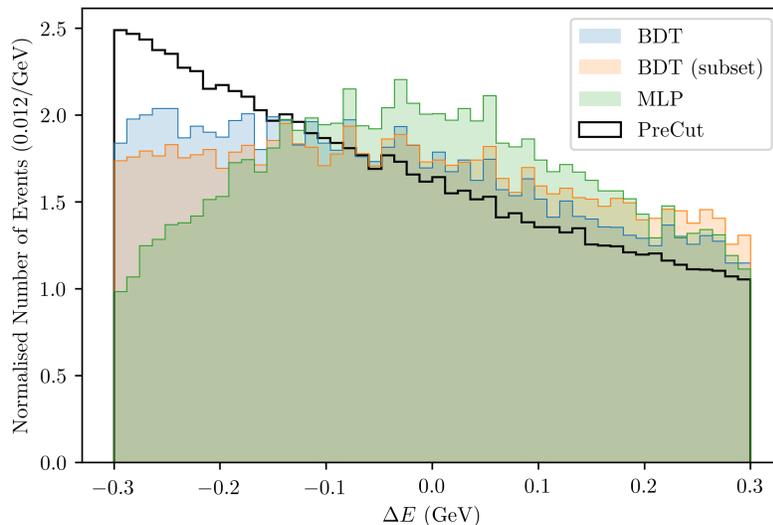
Figure 3.6: Highlight of the significant sculpting seen in the MLP, and the impact of removing the top 3 correlated variables in BDT (BDT subset). Note that PreCut is the shape of ΔE before continuum suppression.

$y > \zeta$ are assigned 'signal' and everywhere else is 'continuum'. A high value of ζ gives high signal purity and low ζ gives high signal efficiency. A ROC curve is a plot of background rejection against the signal acceptance for different values of ζ. An AUC of 0.5 means the classifier performs no better than random, and 1.0 indicates a perfect discrimination.

Performance is not the only consideration for continuum suppression. The fitting variables ΔE, ΔT, M_{bc} were spectators for training the classifier, and so their histogram shapes should remain invariant after implementing the MVA weights. This is important as the final stage of the workflow involves fitting to these variables. Changing the shape of these histogram reduces the ability of the simultaneous fit to distinguish between signal and background, deteriorating the quality of the parameters of interested extracted in the final stage. Thus the shape must be maintained as much as possible.

The ΔE variable for MLP show significant sculpting (shape change) after applying continuum suppression. The BDT also shows some sculpting. The cause of this was hypothesised to be correlation between ΔE and topological variables used to train the classifier. Indeed three variables, `thrustBm`, `Cleocone1` and `Cleocone2` show significant[18] correlation. However removing these variables had little effect on the sculpting as seen in Figure 3.6 with a deteriorated AUC (Figure 3.7). It should be noted that some sculpting should be expected as it is impossible to get zero correlation between the topological variables and ΔE. No significant sculpting is observed in the other fitting variables.

For these reasons, despite the slight performance boost of MLP, the BDT with the full set of 30 event shape variables were used for continuum suppression.

Best Candidate Selection

There are multiple candidates for B_{sig} in each event, where only one is the true signal B. However, when real Belle-II data is processed, only the best candidate (highest Pearson's χ^2 goodness of fit [26]) is selected. When training the data, whether or not this best candidate selection is made for the MC signal and continuum can be a user-specified parameter. The performance for either case was compared in Figure 3.7, and it was

---

[18]Pearson correlation > 0.1

shown that the AUC is virtually identical between the two. Note that when **testing** and **implementing**, best candidate selection is *always* made. For the sake of consistency, the classifier trained with best candidate selection was used.
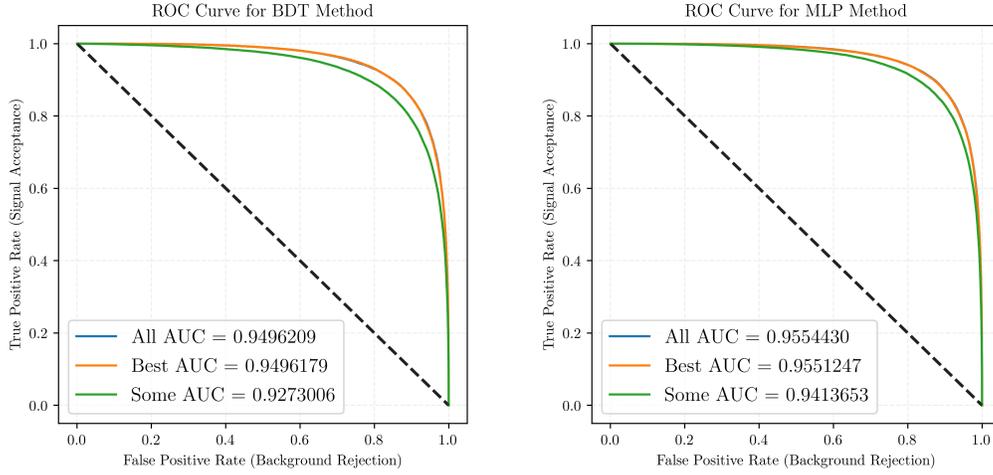
Figure 3.7: The ROC curve comparing BDT (left) and MLP (right). "Best" indicates best candidate selection being done and "All" indicates all candidates being used for the training. In addition, "Some" shows the performance of the classifier (with best candidate selection), when highly correlated variables are removed in an attempt to remove sculpting.

3.1.9 Logarithmic Transformation

A major difference between the Belle analysis and this one is the use of continuum suppression as a fitting variable. By including it in the (3+1)D fit, better discrimination between signal and background can be made during the simultaneous fit, leading to a subsequent improvement in measuring the parameters of interest. The shape of the classifier output y tends to be asymmetric and not conducive to fitting. A monotonic transformation is made to approximate a Gaussian shape whilst preserving the information. By determining the maximum value of y in the MC dataset M and a rough threshold of y which gives $> 90\%$ signal acceptance R, the following transformation is made:

$$\text{logSupp} = \log\left(\frac{y - R}{M - y}\right) \tag{3.13}$$

Note that to ensure the argument of the log is positive, a threshold $y > R$ is set.

3.1.10 Fitting PDF

In the final stage of the workflow, when validation and measurements are conducted, we are required to have a Probability Density Function (PDF) to represent each of the fitting variables $\{M_{bc}, \Delta E, \Delta T, \text{logSupp}\}$ and event type {signal, continuum, BB}, giving a total of 12 PDFs (see Table 3.2). The yields are represented by the normalisation of the PDF of each event type, and the $\mathcal{CP}$ violating parameters are built into the resolution function. Although the form of the PDF is user-specified and known, the exact parameters which best describe each of the 12 PDFs must be determined in a process known as 'fitting'. The determination of these parameters is the final stage of building the analysis. This section begins with the nuances which exist

Stage	Selection Criteria (Units: E – GeV, M – GeV c^{-2}, T – ps, r – cm)
Skim (Broad)	$M_{bc} > 5.2$ and $|\Delta E| < 0.3$
Skim (Tight – Signal Mode)	$E_\gamma > 0.05$, $-0.85 < \cos\theta_\gamma < 0.9$ and $dr_{K_S^0} > 0.03$
Skim (Tight – Control Mode)	$M_{bc} > 5.26$, $-0.275 < \Delta E < 0.2$
	$2.8 < M_{J/\psi} < 3.3$ and $dr_{K_S^0} > 0.03$
Pre-Continuum Suppression	$|H_{\{*\}}^{so}|, |H_{\{*\}}^{oo}|, |R_2| < 10$, $|\text{KSFW}_{\{mm2,et\}}| < 100$, Cleocone$_{\{*\}} < 10$
	$|\cos\theta_{BO}|, |\cos\theta_{Bz}| < 1.01$ and $|\texttt{thrustBm}|, |\texttt{thrustOm}| < 1.01$
Post-Continuum Suppression†	$M_{bc} > 5.24$, $|\Delta T| < 70$ and $\texttt{logSupp} > -9$
Pre-Fitting	$M_{bc} < 5.292$ and $\Delta T_{err} < 2.5$

Table 3.1: A list of selection criteria made at each stage of the analysis. The γ refers to the final state of $\pi^0 \to \gamma\gamma$. The cuts in violet is one of the major differences between $B^0 \to K_S^0 \pi^0$ and $B^0 \to J/\psi\, K_S^0$ (which done as a complementary analysis – see Section 3.3). $\{*\}$ indicated all relevant variables (eg. all Cleocones). Cleocone has units of momenta GeV c^{-1}, and the KSFW moments has units pertaining to their formulae in Section 3.1.7. Note that from the † stage onwards, the M_{bc} correction has been made.

for signal events and the resolution function. Then the Maximum Likelihood Estimation used to find PDF parameters is outlined.

Firstly, the probability of misidentifying (mistagging) the flavour of B_{tag} must be accounted for. Intuitively, that events whose flavour tagging confidence is high should be weighted more heavily, the signal[19] MC is first divided into 7-8 bins of qr (for the exact parameters of these bins, see Section 4.1). For each bin (and each fitting variable), unique PDF parameters are determined (thus in reality, more than 12 PDFs must be defined).

The PDFs for the resolution shown in Table 3.2 only fit to the shape of the distribution. In order to account for the $\mathcal{CP}$ violation, we introduce another PDF f^{phys} describing the physics effects. The overall resolution PDF is a convolution of the physics PDF and the shape of the resolution given by:

$$\mathcal{P}_{K_S^0 \pi^0}(\Delta T, q) = f^{phys} \circledast \mathcal{R} \tag{3.14}$$

Note that the mistag information in each bin is accounted for in signal events with a physics PDF given by[20]:

$$f_i^{phys}(\Delta t, q) = \frac{e^{|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \left([1 - q\Delta w_i + q\mu_i(1 - 2w_i)] + [q(1 - 2w_i) + \mu_i(1 - q\Delta w_i)] \times \right.$$
$$\left. (\mathcal{S}_{CP}\sin(\Delta m\Delta t) + \mathcal{A}_{CP}\cos(\Delta m\Delta t))) \right)$$

where w_i represents the average mistag probability, $\Delta w_i = w_i^{B^0} - w_i^{\bar{B}^0}$ give the difference in mistag rate between B^0 and $\bar{B}^0$ and μ_i gives the fraction of events in the i-th bin. Unlike the signal events, background events do not contribute to the $\mathcal{CP}$ violation, thus their flavour assignment is meaningless, so only one set of PDF parameters is assigned to each fitting variable. Furthermore, only the pre-factor exponential term is preserved for continuum and BB in f^{phys}.

---

[19] The $\mathcal{CP}$ violating effects is only exhibited in the signal channel.
[20] This function was adapted from the $\texttt{RooFit}$ library [34].

Fitting Variable	Event Type		
	Signal	Continuum	BB
ΔT (Resolution)*		Core Gaussian	
		Tail Gaussian	
		Outlier Gaussian	
M_{bc}	2 Gaussians	Gaussian	Argus [35]
		Argus	
ΔE	Gaussian	Chebychev Polynomial	Gaussian
	Bifurcated Gaussian	(2nd order)	
	Chebychev Polynomial (1st order)		
`logSupp`	3 Gaussians	3 Gaussians	Gaussian
ΔT_{err}	Gaussian	-	-
	Landau	-	-

Table 3.2: A summary of PDFs selected to fit to each of the fitting variables in each event type. * – the listed PDF for ΔT is the fit the physical shape of the data. The actual PDF used for the model accounts for $\mathcal{CP}$ violation and is more complicated (see text). Note the final row pertains to the per-event error analysis in Section 4.1.

One can determine the PDF parameters $\vec{p}$ which best represent each dataset x by defining the likelihood[21]:

$$\mathcal{L}_{k,j} = \text{Poisson}(N_{obs}|N_{exp}(\vec{p})) \prod_{i=0}^{N_{obs}} f(x;\vec{p}) \tag{3.15}$$

where N_{exp} is the number of events in a dataset, and N_{obs} is the result of a Poisson fluctuation centred around N_{exp}. We would like to find the set of $\vec{p}$ which maximises $\mathcal{L}$. This process is known as Maximum Likelihood Estimation. Since the data we have at this stage is MC, the $\mathcal{A}_{CP}$, $\mathcal{S}_{CP}$ and yields are known. First these parameters are fixed so only the PDF parameters are allowed to change. Then a multi-dimensional version of gradient descent[22] on the negative log-transformed $\mathcal{L}$ via the `MINUIT` library [36] in `ROOT`. Examples of individual fits to PDFs are shown in Figure 3.8-3.9.

---

[21]Theoretically this formalism is for an *extended* maximum likelihood which is used in the simultaneous fit. In this case we can either use the number of events remaining in each MC (signal, continuum, BB) set to normalise this, or set the prefactor to be unity.

[22]Beginning at an initial point, the gradient of the target function (negative log-likelihood) is calculated towards the direction of steepest descent. The next point is iteratively calculated along this direction with a change in distance defined by the Hessian matrix which parameterised the curvature of this space. This is done with the `HESSE` package implemented by `MINUIT`.
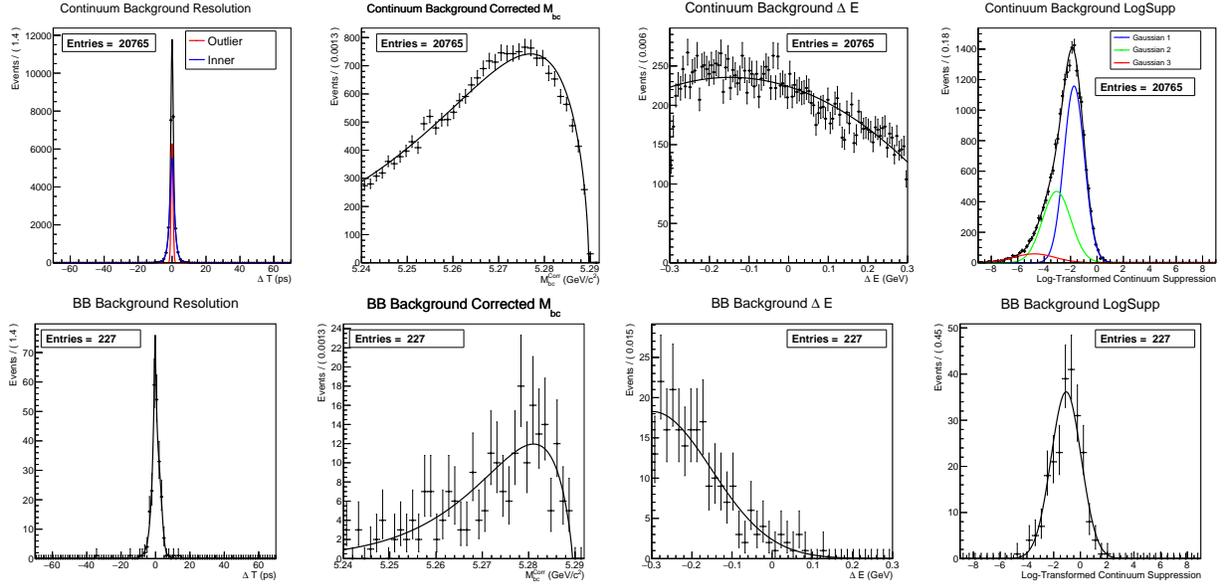
Figure 3.8: Individual fits find PDF parameters for continuum (top) and BB (bottom) to build the model. This was for the experimental set up described in Section 4.1 (treeFitter with per-event error).
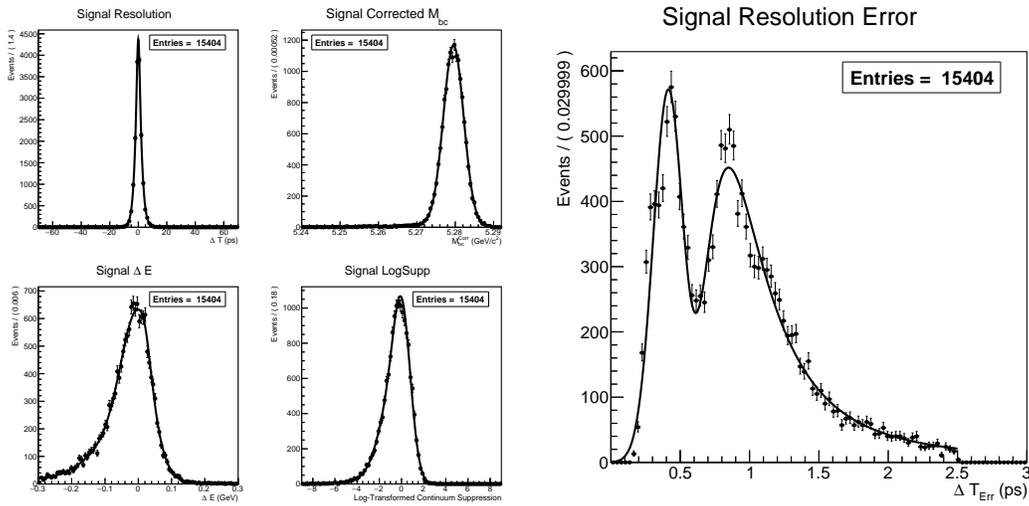

Figure 3.9: Individual fits find PDF parameters for signal to build the model. This was for the experimental set up described in Section 4.1 (treeFitter with per-event error). As such the fit to the uncertainty distribution of the resolution ΔT_{err} is also included.

3.2 Validating the Model

Having successfully built a model, it is important to assess it for bias, in a process known as validation. We require the model to accurately extract out parameters of interest to reflect their true values in nature when a measurement is performed. Since MC data requires an input $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$, MC data can be used to create a dataset with known $\mathcal{A}_{CP}$, $\mathcal{S}_{CP}$ and yields. By mixing MC data in the correct proportions, we can mimic real Belle II data, and apply the physics analysis to this mock dataset. A comparison of input and extracted values validates the physic model. Within the scope of this thesis, we will only validate for $\mathcal{S}_{CP}$.

One million signal events were generated with $\mathcal{A}_{CP} = 0.0$ and $\mathcal{S}_{CP} \in [-1.0, 1.0]$ in steps of 0.1 for a total of 21 test cases. Skimming, ROE, vertexing, and flavour tagging are identical to what was done in Section 3.1. Since continuum suppression has already been trained, the resulting weights are used to evaluate `logSupp`. This section discusses how the 'cocktail' dataset is created and how the parameters of interest can be extracted.

3.2.1 The Mock Data

The mock data is a collection of signal, continuum and BB MC events which are combined in a single set to imitate Belle II data. The proportions of this is set to the expected number of events, and depends on the integrated luminosity we wish to imitate. This section simulates 5 ab^{-1} of Belle II data, however it can easily be extended to different amounts of data.

Expected Number of Signal Events

The reconstruction efficiency ϵ_{reco} is the proportion of signal events which pass the selection criteria given in Table 3.1 and is approximately 13% for $B^0 \to K_S^0 \pi^0$. The number of neutral B-mesons expected[23] in 5 ab^{-1} of data is the product of the cross section of $B\overline{B}$ events at Belle-II, σ =1.1 nb, and the integrated luminosity, giving $N_{B^0\overline{B}^0}$ =5.5 × 10^9. The number of signal events expected is:

$$N_{sig} = \epsilon_{reco} \times \mathcal{B}(B^0 \to K_S^0 \pi^0) \times N_{B^0\overline{B}^0} \tag{3.16}$$

where $\mathcal{B}(B^0 \to K_S^0 \pi^0)$ =4.45 × 10^{-6} [2] is the globally averaged branching ratio[24] of the signal decay.

Expected Number of Background Events

Since 1 ab^{-1} of background events were obtained for this analysis, the number of expected events in 5 ab^{-1} is simply a five-fold increase to the events which remain the signal and background MC.

Creating the Mock Data

The expected events for signal, continuum, and BB are first Poisson fluctuated. The number of signal MC events available to create the mock data far exeeds the number of expected events. As such these events are randomly selected from the pool of events. The background MC, however, have far fewer events than required for the mock data. As such to reach the required amount of events in continuum and BB, these events are 'generated' according to their qr distributions, as the proportion of events in each qr is required to be preserved for a time-dependent measurement. By mixing the randomly selected signal events, and generated background events, a mock dataset is created. This is also known as a toy MC sample.

---

[23]A factor of $1/2$ for charged B and of 2 for the coherent pair is accounted for.

[24]Although it seems to be a tautology to use the branching ratio to determine the branching ratio. Only an estimate here is required as N_{sig} is Poisson fluctuated.

3.2.2 Toy MC Experiments

Using the toy MC sample, we are now in a position to extract the parameters of interest. First the sample is divided into the bins of qr. Analogous to Section 3.1.10, these parameters are found using the method of maximum likelihood fit. However, in this case, the PDF parameters are fixed, and the five parameters of interest, $\mathcal{A}_{CP}$, $\mathcal{S}_{CP}$ and yields of each event type are determined. This fit is *simultaneous* in the sense that we are required to find the parameters which maximise the a global likelihood function across *all* of the binned dataset and fitting variables. A visual inspection of the fit quality is conducted to the fitted PDF matches with the distribution of the data (see Figure 3.10). Note that for a measurement the process is identical, but Belle II data will replace the toy MC sample.

A single fit will provide limited information about the quality of the physics model. As such we create 500 toy MC samples for each test set of input $\mathcal{S}_{CP}$, and extract parameters of interest. This gives a distribution of $\mathcal{A}_{CP}$, $\mathcal{S}_{CP}$ and yields. Increasing the number of fits will cause the mean of these parameters to converge to their true values. This process is known as a toy MC experiment. The spread of each parameter in the toy MC experiment represent the statistical uncertainty. We can estimate the bias of each toy MC experiment by calculating the statistical pull, which is a distribution defined by the difference between the expected (input) value, and the extracted value, normalised by the spread. An unbiased model should give a standard normal distribution as the pull. Notice in Figure 3.11, the small pull in the $\mathcal{S}_{CP}$ indicates little bias for that toy MC. The statistical uncertainty is a measure of the performance of the *physics model*. One can assess the quality of the toy MC experiment by calculating the uncertainty *on the mean* which pertains to the *entire toy MC experiment*. This vanish as the number of fits increases.

3.3 Control Mode: $B^0 \rightarrow J/\psi\, K_S^0$

The $B^0 \rightarrow J/\psi\, K_S^0$ is considered a golden mode since it has a relatively high branching ratio at 8.73×10^{-4} and a set of well-defined[25] $\mathcal{CP}$ violating parameters [2]. This makes it ideal as a "control" mode for validating the resolution PDF of $B^0 \rightarrow K_S^0 \pi^0$ (see Section 4.2).

A set of one million signal MC events was generated for $\mathcal{S}_{CP} \in [-1.0, 1.0]$ in steps of 0.2 analogous to $B^0 \rightarrow K_S^0 \pi^0$. Notice from Table 3.3 we are required to select approximately half of the signal MC pool for toy experiments. Sampling from such low statistics means the results may not be representative of the population. As such after generating one set of signal MC to calculate ϵ_{reco}, we force the J/ψ to decay to either $e^+ e^-$ or $\mu^+ \mu^-$ in `EvtGen`. This channel occurs generically around 10% of the time, however it is the only channel of J/ψ which can be properly reconstructed. By restricting the generation, we increase the number of signal MC events which pass the selection criteria. There is a 10-fold increase in the number of signal MC events which makes the random sub-sampling more representative. All other steps of building and validating the analysis were identical to $B^0 \rightarrow K_S^0 \pi^0$. Only $1\,\mathrm{ab}^{-1}$ of data was simulate in the toy MC experiments.

---

[25]Uncertainty on $\mathcal{CP}$ violating parameters $\sim \mathcal{O}(10^{-2})$
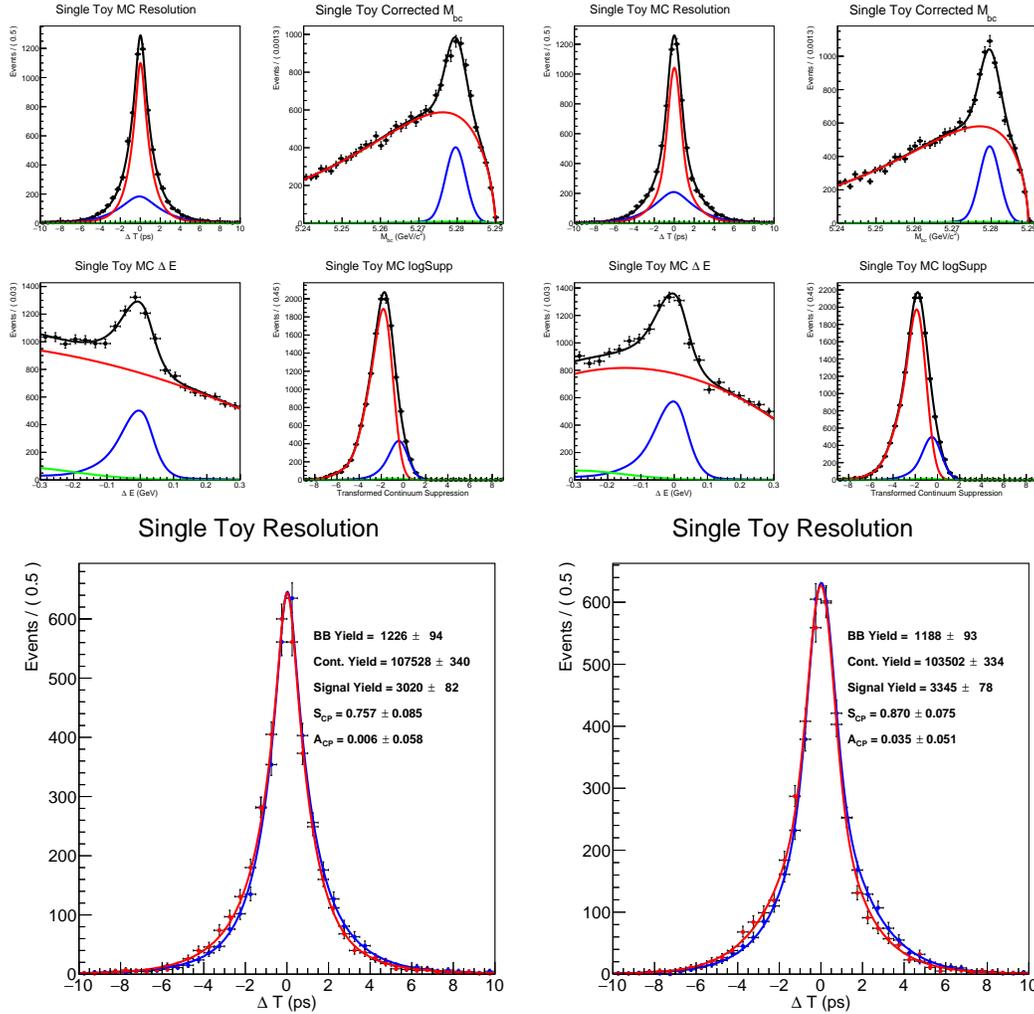
Figure 3.10: (Top) One of the 500 toy MC fits to the data. Note that each plot is a projection of a 4D PDF onto one of four fitting variables. On the left is the baseline model described in Section 4.1, and on the right is the best performing model outlined in Section 4.1 (treeFitter with per-event error). The red, blue and green represent signal, continuum and BB respectively. (Bottom) The resolution is plotted for B^0 (blue) and $\overline{B}^0$ (red) to compare the difference in resolution.
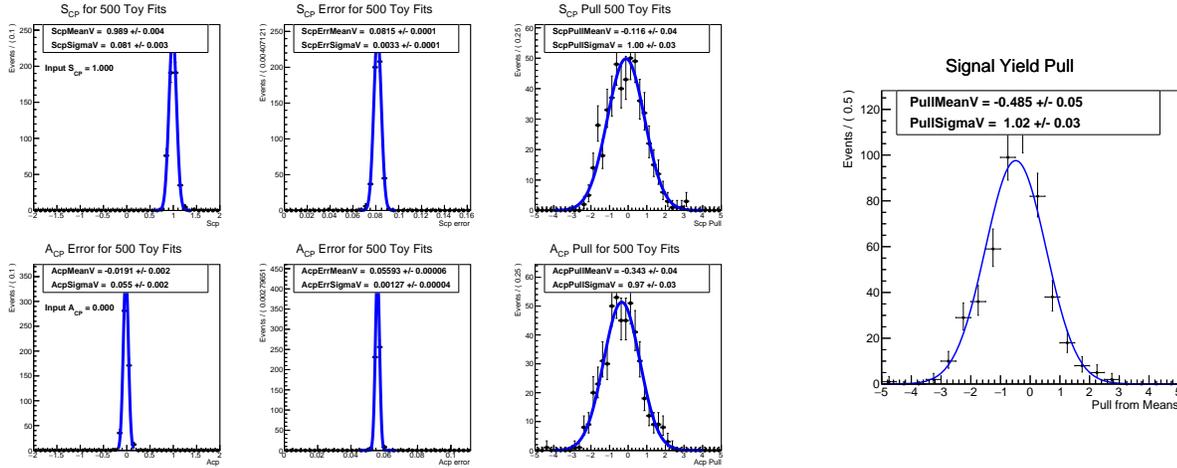
Figure 3.11: (Left) The distribution of $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ as well as their statistical uncertainties. Their pull distributions are also displayed. (Right) The pull distribution of the signal yield is shown. All pulls here show relatively small bias as the pull distribution is close to the standard normal.

Toy MC Experiment	Integrated Luminosity (ab^{-1})	ϵ_{reco} (%)	Number of Expected (Survived) Events		
			Signal	Continuum	BB
RAVE (Baseline)	5	12.72	3113 (127185)	107295 (21459)	1165 (233)
Tree		13.54	3315 (135428)	103825 (20765)	1135 (227)
K_S^0 Vertex	1	1.78	8524 (17752)	5925 (5925)	453 (453)
J/ψ Vertex		2.65	12702 (26455)	9935 (9935)	723 (723)

Table 3.3: A summary of the number of events which survive all selection criteria is displayed in the parenthesis. The reconstruction efficiency ϵ_{reco} is the fraction of signal MC events which survived the selection criteria. The expected number of events is the result of calculations from Sections 3.2.1-3.2.1, giving an estimate of each event type is available for fitting for the given integrated luminosity.

Chapter 4

Validation and Results

So far, the discussion has revolved around a single toy MC experiment for a particular input $\mathcal{S}_{CP}$. In order to fully validate the analysis we employ an ensemble of toy MC experiments for each test case. There are two overall metrics we use to assess the performance of each physics model:

- **Error in $\mathcal{S}_{CP}$** – the statistical uncertainty on $\mathcal{S}_{CP}$ should be minimised. The average uncertainty is taken across all toy MC experiments in the ensemble so that each physics model has a single statistic representing the error in $\mathcal{S}_{CP}$.

- **Bias** – a linear fit is performed on the extracted $\mathcal{A}_{CP}$, $\mathcal{S}_{CP}$ and signal yield as a function of the input $\mathcal{S}_{CP}$. If the entire ensemble is unbiased, the linear fit in $\mathcal{S}_{CP}$ should have a gradient of 1.0 with an offset of 0.0. Deviation in the gradient is the measure of bias in $\mathcal{S}_{CP}$. For $\mathcal{A}_{CP}$, the gradient and offset should be 0.0. The offset deviation is the measure of bias in $\mathcal{A}_{CP}$. The signal yield should have a zero gradient and an offset to match the expected yield. The percentage deviation from this value is quoted.

This Chapter will first begin with exploring different methods of reducing the error on $\mathcal{S}_{CP}$ without compromising the bias of the analysis. Adjustments of the qr binning and vertexing methods are assessed, and a novel method of accounting for error in the resolution on a per-event basis is investigated. Then, we discuss the results of the validation which was done for the $B^0 \rightarrow J/\psi K_S^0$ control mode and determine correction factors relating this back to the signal mode. Finally we estimate when Belle II will reach two milestones in datataking: obtaining a statistically significant yield (rediscovery) and resolving this section of the $B \rightarrow K\pi$ puzzle.

4.1   Comparing Models

First we would like to investigate the effect the qr binning has on the performance of the model. Recall that a simultaneous fit is done in bins of qr to account for the mistag information. Increasing the number of bins will increase the granularity of this mistag, if this number is too high, there is insufficient events for the fits to individual PDFS to converge. A baseline is established by using 8-bins of qr and vertexRAVE following previous analyses[1] We also consider a 7-bin (where the highest two qr bins were combined), and a modified 7-bin (where the 2nd and 3rd highest bins were combined) model, with the following edges on the dilution factor r:

- **8-bin** – 0.0,0.1,0.25,0.5,0.625,0.75,0.875,0.9375,1.0

---

[1]It should be noted that prior to this work, a similar analysis was validated in release 2 of `basf2`. However in the new release 3, several functions have been depreciated, and the flavour tagger has been updated.

- **7-bin** – $0.0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.875, 1.0$

- **7-binMOD** – $0.0, 0.1, 0.25, 0.5, 0.625, 0.75, 0.9375, 1.0$

It was shown in Table 4.1 that the 8-bin model was the (marginally) best performing with an error of 0.0856. Both $\mathcal{S}_{CP}$ and $\mathcal{A}_{CP}$ show a 2% bias in this case. The modified 7-bin model showed similar performance for error in $\mathcal{S}_{CP}$, but has more significant bias in $\mathcal{A}_{CP}$. The bias in the yield was approximately 1% across all three. As such the 8-bin

The performance of these binning methods are shown in Table 4.1 with the 8-bin method slightly out-performing others with an error of 0.0856. Notice all $\mathcal{S}_{CP}$ has an approximately 2% bias. The $\mathcal{A}_{CP}$ also has a 2% bias except for the modified 7-bin which as a two-fold increase in that. The signal yield is consistently underestimated by approximately 1%. The 8-bin vertexRAVE method serves as a baseline for the following comparisons.

Since vertexRAVE software is no longer maintained, alternative vertexing techniques is preferred for the analysis. Thus its performance is compared to treeFitter[2]. The treeFitter shows a slight improvement with an error in $\mathcal{S}_{CP}$ at 0.0845, but there is a significant bias in the $\mathcal{A}_{CP}$ at 8%.

Finally, the consideration of vertex quality has been neglected so far in our discussion. Depending on the final state particle trajectories, the experimental uncertainties associated with the resolution differ event-by-event. This is the principle of per-event error. Rather than the resolution of each event being characterised by ΔT, a pair of values $(\Delta T, \Delta T_{err})$ which are correlated with each other take on this role. Suppose we can model the conditional[3] using a PDF ($\mathcal{E}(\Delta T | \Delta T_{err})$) just like the fitting variables, then the fully resolution PDF $\mathcal{P}'(\Delta T, \Delta T_{err})$ is a convolution of the original PDF $\mathcal{P}(\Delta T)$ [34]:

$$\mathcal{P}'_{K_S^0 \pi^0}(\Delta T, \Delta T_{err}, q) = \mathcal{P}_{K_S^0 \pi^0}(\Delta T, q) \otimes \mathcal{E}(\Delta T | \Delta T_{err}) \tag{4.1}$$

The physical PDFs fitted to the error in resolution is in Table 3.2. Implementing Per-Event error also improves the performance of the model as the error in $\mathcal{S}_{CP}$ drops to 0.0825, with only a slight increase in the $\mathcal{A}_{CP}$ bias of 1%, and almost no bias in $\mathcal{S}_{CP}$.

A combination of treeFitter and per-event error show the most significant improvement, reducing the error in $\mathcal{S}_{CP}$ by 6.8%, giving 0.0798. Although there is virtually no bias in the signal yield, there is a 5% bias in $\mathcal{S}_{CP}$ and 8% in the $\mathcal{A}_{CP}$. From Figure 4.6, it is clear the bias in the $\mathcal{S}_{CP}$ is not too significant. However we find the bias in the $\mathcal{A}_{CP}$ needs to be addressed. Although for 5 ab^{-1}, we find that this bias lies well within the statistical uncertainty of a single fit, but when we reach 50 ab^{-1}, we expect this to reduce to 0.0159, making this bias problematic. Luckily, this bias can be addressed. A correction factor yet to be accounted for is the fact that the detector interacts differently between matter and antimatter. It the correction on detector bias can address this issue, then the treeFitter per-event error is the best physics model.

4.2 Control Mode

Since the $B^0 \rightarrow J/\psi \, K_S^0$ has such well-defined parameters, it is easier to estimate the systematic uncertainties for this channel. This section establish as relationship between the signal and control modes so by analysing the latter, we can estmiate the systematic uncertainties in the PDF resolution for the former.

Unlike our signal decay however, both K_S^0 and J/ψ can be vertexed[4]. We consider the case where B_{sig} is vertex by K_S^0 only and J/ψ only. This will be known as K_S^0 vertex and J/ψ vertex. This is because the

---

[2]TreeFitter is also shown to reduce the number of BB background slightly in this analysis (Table 3.3), which was mentioned in Ref. [30].

[3]This is conditional in the sense that the shape of the PDF depends on the value of ΔT

[4]Here we opted to use vertexRAVE as it has single-vertex functionalities which is not expected of a global fitter like treeFitter.
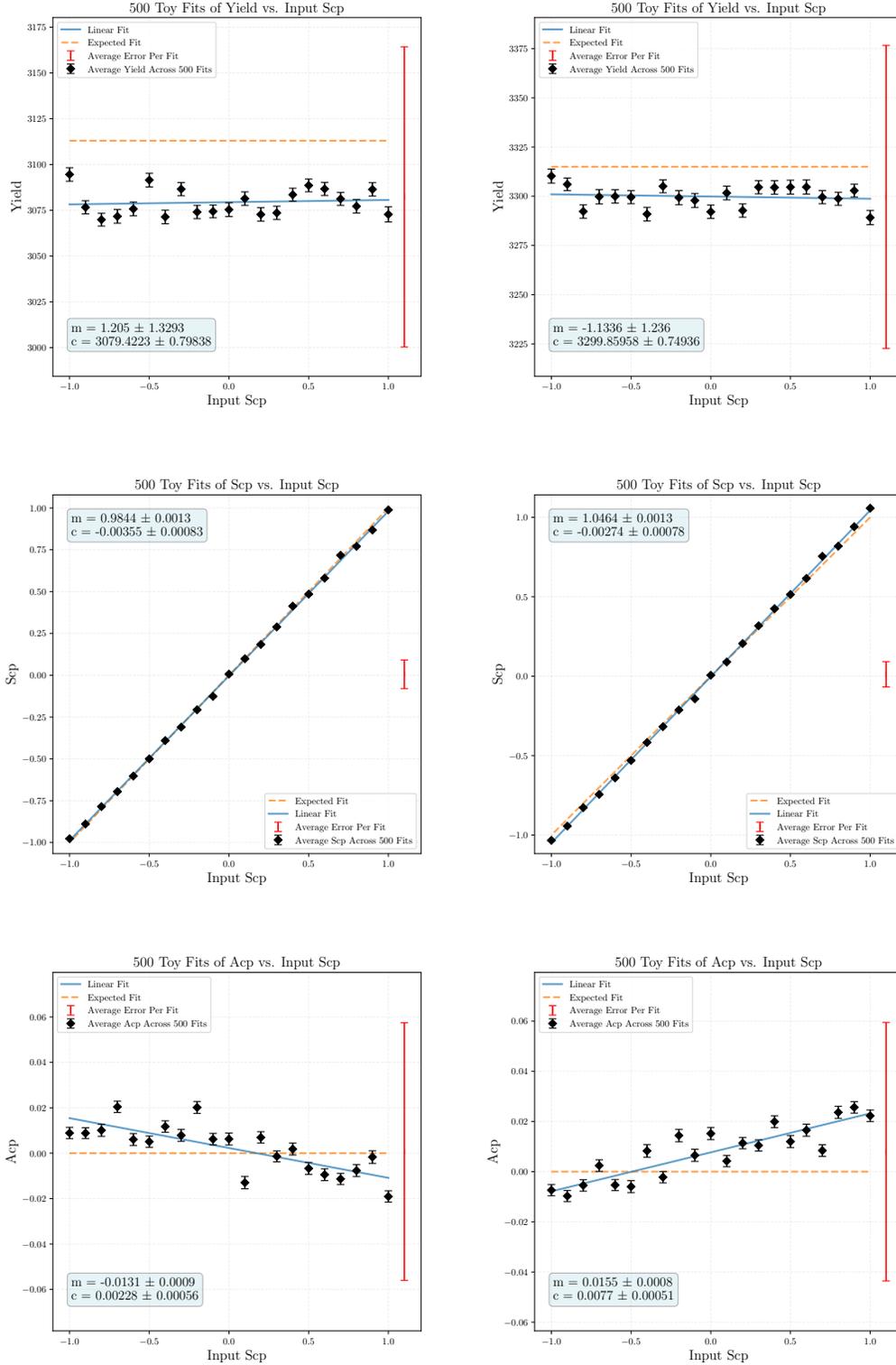
Figure 4.1: Plots showing trends in signal yield (top), $\mathcal{S}_{CP}$ (middle), and $\mathcal{A}_{CP}$ (bottom) for the baseline (left) and tree fitter with per-event error (right). The error bars on each point is the uncertainty on the mean of the output parameter which is inversely proportional to the number of fits. The large red error bar is the averaged uncertainty expected in a single fit – this is an approximation of the statistical error of the analysis.

analysis for $B^0 \rightarrow J/\psi K_S^0$ is well investigated, thus we can employ this mode to investigate the vertexing on the K_S^0. First we consider the fit to the fitting to extract PDF parameters when we build the models. As shown in Figure 4.2, the K_S^0 vertex fit is quite reasonable, but the J/ψ vertex does not converge, due to the significant asymmetry in the resolution. This is actually a known issue. In previous experiments, the beamspot[5] was broad enough that the IP tube constraint gives an unbiased vertex for B_{sig}, However, in Belle-II, with the advent of nano-beam technology, the B vertices lies just outside the beamspot. Thus using the IP tube constraint introduces bias to resolution, and fitting the usual PDF to it becomes challenging. This can be addressed by an additional constraint in release 4[6] of `basf2`. Nonetheless the K_S^0 being successful means the resolution can still be validated.
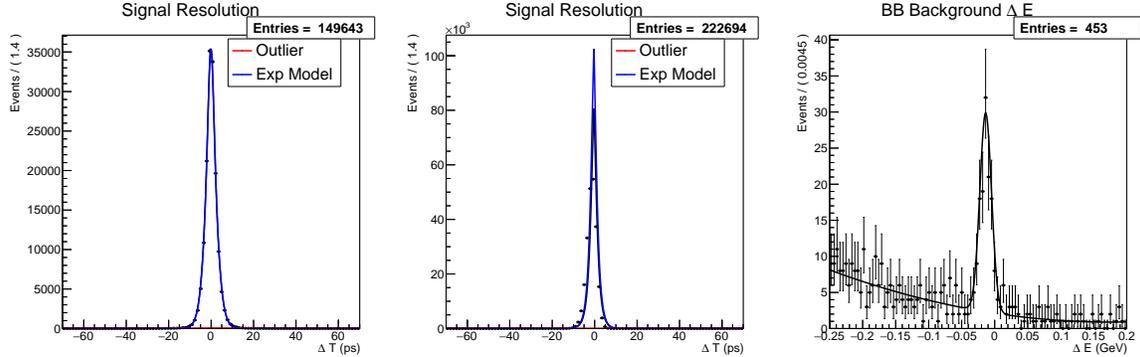
Figure 4.2: The fits to resolution in the control mode for a K_S^0 vertex (left), and J/ψ vertex (middle). The peaking BB background exists in both the K_S^0 and J/ψ vertices, but is shown for the former (right).

Analogous to the previous section an ensemble of 500 toy fits was performed to assess the bias of the model. A single toy fit is shown in Figure 4.3, and one can see that the fit to the resolution is not great in either case. This can be caused by a peak close to $\Delta E = -0.02$ GeV in the BB background. The ideal shape is shown in Figure 3.8. The source of this peaking background is likely due to a charm-meson misidentified as J/ψ. The investigation will need to be further investigated in future work.

Nonetheless, the framework to establish the relationship between the signal and control modes can still be created. In order to do this, we take the baseline of 8-bin vertexRAVE, and replace the ΔT parameters with those from the K_S^0 vertex control mode. An ensemble fit is done to this. Doing the signal mode with control mode parameters was seen to be successful (see Figure 4.5) with a higher error in $\mathcal{S}_{CP}$ at 0.0934.

For each input $\mathcal{S}_{CP}$, the difference in the control- and signal-mode extracted $\mathcal{S}_{CP}$ is plotted in Figure 4.4. From this linear relationship we formulate a correction factor:

$$\mathcal{S}_{CP}^{signal} = -0.0789 * \mathcal{S}_{CP}^{input} + (-0.00186 + \mathcal{S}_{CP}^{control}) \qquad (4.2)$$

4.3   Belle II Milestones

Thus far we have considered only the scenario of 5 ab^{-1} of Belle II data. At the time of this thesis, however, 10 fb^{-1} of integrated luminosity has been reached, which is much less than the amount we have considered. At the end of data taking, we expected approximately 50 ab^{-1}. In the low statistics regime, it is useful to consider

---

[5]At Belle II, rather than collide individual leptons, collections of them known as 'beams' functionally collided. The beamspot is a localised area with e^+e^-

[6]Which has just been release at the time of writing this thesis.
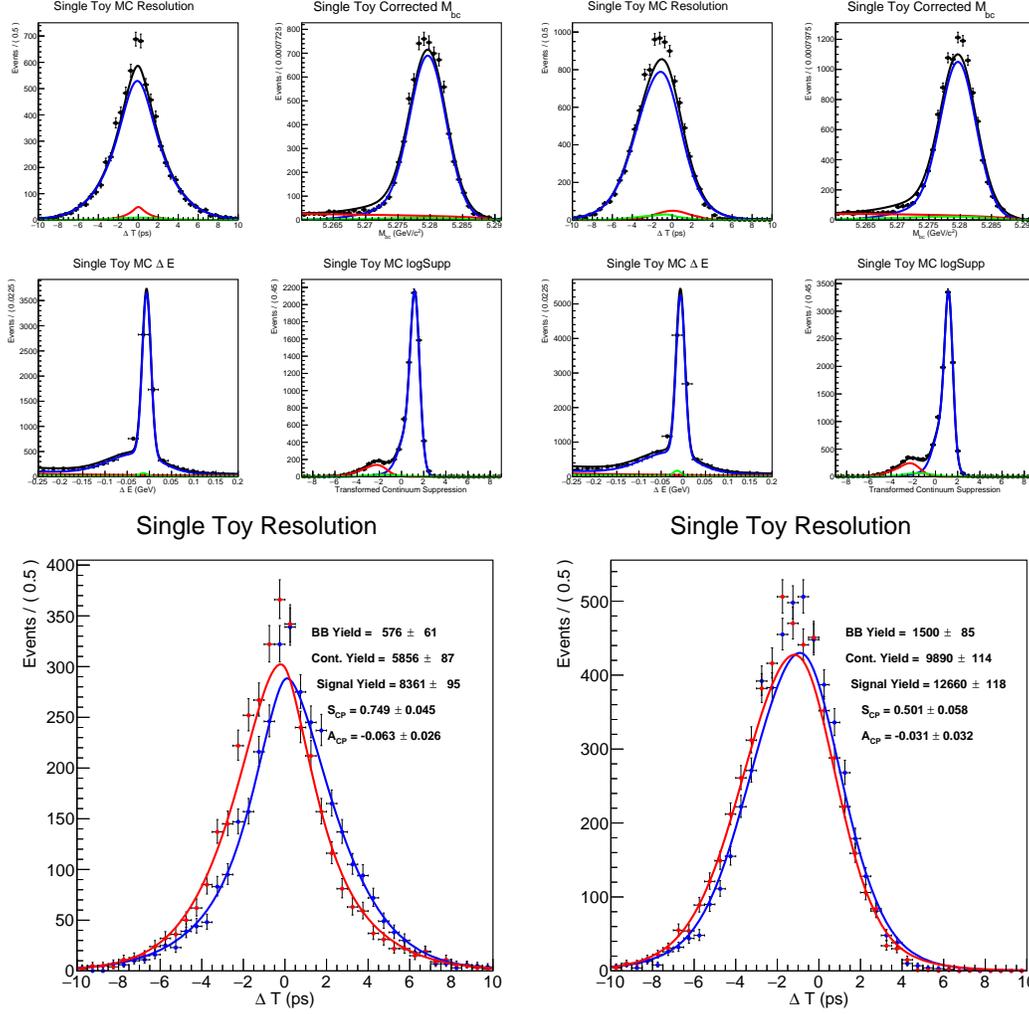
Figure 4.3: A single toy fit in the control mode for the K_S^0 (left) and J/ψ (right) vertices. Notices the PDF does not fit well to the resolution data, which may be due to the peaking BB background.

when we are able to obtain a statistically significant (non-zero) yield of $B^0 \to K_S^0 \pi^0$ at a milestone known as rediscovery. In the high statistics regime, assuming that $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ remain at their experimental and theoretical central values, we would like to estimate when a resolution to this sector of $B \to K\pi$-puzzle can be achieved.

4.3.1 Low Statistics

A measurement of statistically significant yield for each toy fit[7] can be made by first extracting the maximum log-likelihood[8] $\mathcal{L}_{max}$ which correspond to the extracted values. Then we fix the yield at zero and fit to the same data, also extracting the maximum log-likelihood associated with a zero yield $\mathcal{L}_0$. The statistical significance is given by the number of standard deviations s from a null hypothesis of zero yield, where s is:

$$s = \sqrt{2\mathcal{L}_{max} - 2\mathcal{L}_0} \tag{4.3}$$

---

[7]At these low statistics it is impossible to extract $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$. We conduct a fit where we fix these to zero in order so that the fitting can more easily converge.

[8]This is simply the logarithm of the likelihood.
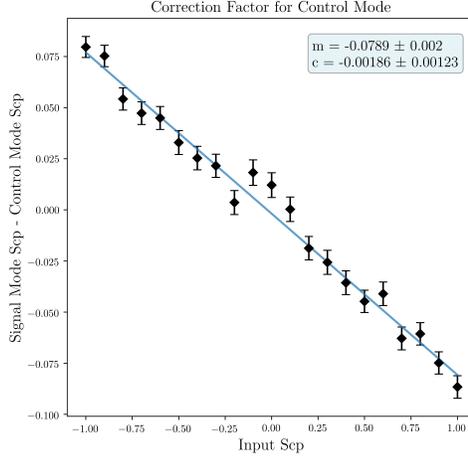
Figure 4.4: The linear trend in the difference in extracted $\mathcal{S}_{CP}$ against the input $\mathcal{S}_{CP}$ is plotted. Using the gradient and offset found from this, we can formulate a correction factor.
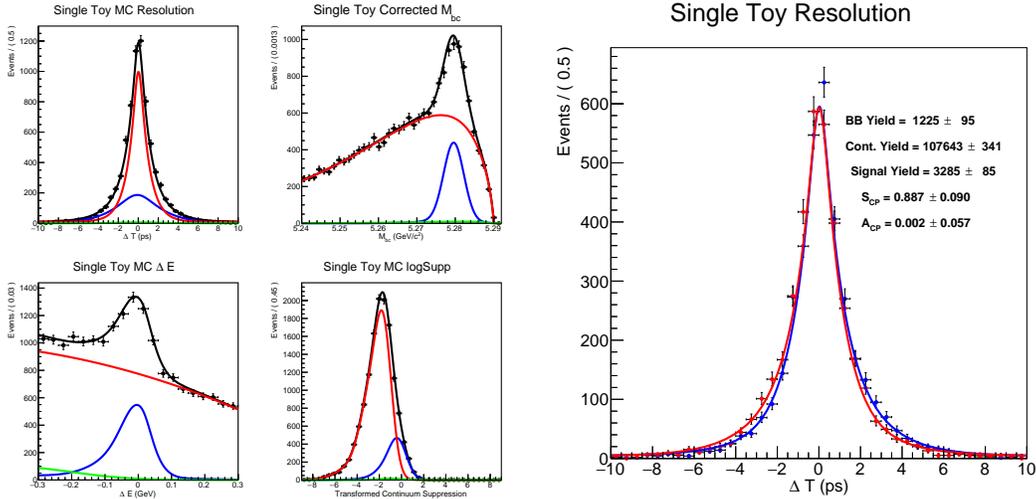

Figure 4.5: A single toy fit to the signal mode using resolution parameters obtained from the control mode.

We investigate the statistical significance of 5 fb⁻¹, 10 fb⁻¹, 20 fb⁻¹ and 50 fb⁻¹. Only 50 fb⁻¹ give a statistically significant yield. The upper limit on the branching ratio can also be found by taking the 90-percentile of the extracted yield and replacing it with N_{sig} in Equation 3.16. Both of these are shown in Table 4.2.

4.3.2 High Statistics

To estimate when the statistical uncertainty will allow the $B \to K\pi$-puzzle to be resolved, we estimate the error in $\mathcal{S}_{CP}$ and $\mathcal{A}_{CP}$ which are their estimated statistical uncertainties σ. To calculate when these reach significance, we must find the integrated luminosity when the difference of the $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ is greater than 5σ. This was found to be 20 ab⁻¹ for $\mathcal{A}_{CP}$ and 10 ab⁻¹ for $\mathcal{S}_{CP}$. Their combined statistical uncertainty is defined by the p-value, of the $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ individually. We then take a product of that and calculate the significance. As seen from Table 4.3, a resolution to the tensions in $B^0 \to K_S^0\pi^0$ is found at 5 ab⁻¹. Note these calculations involve only experimental uncertainties, future calculations involving theoretical uncertainties
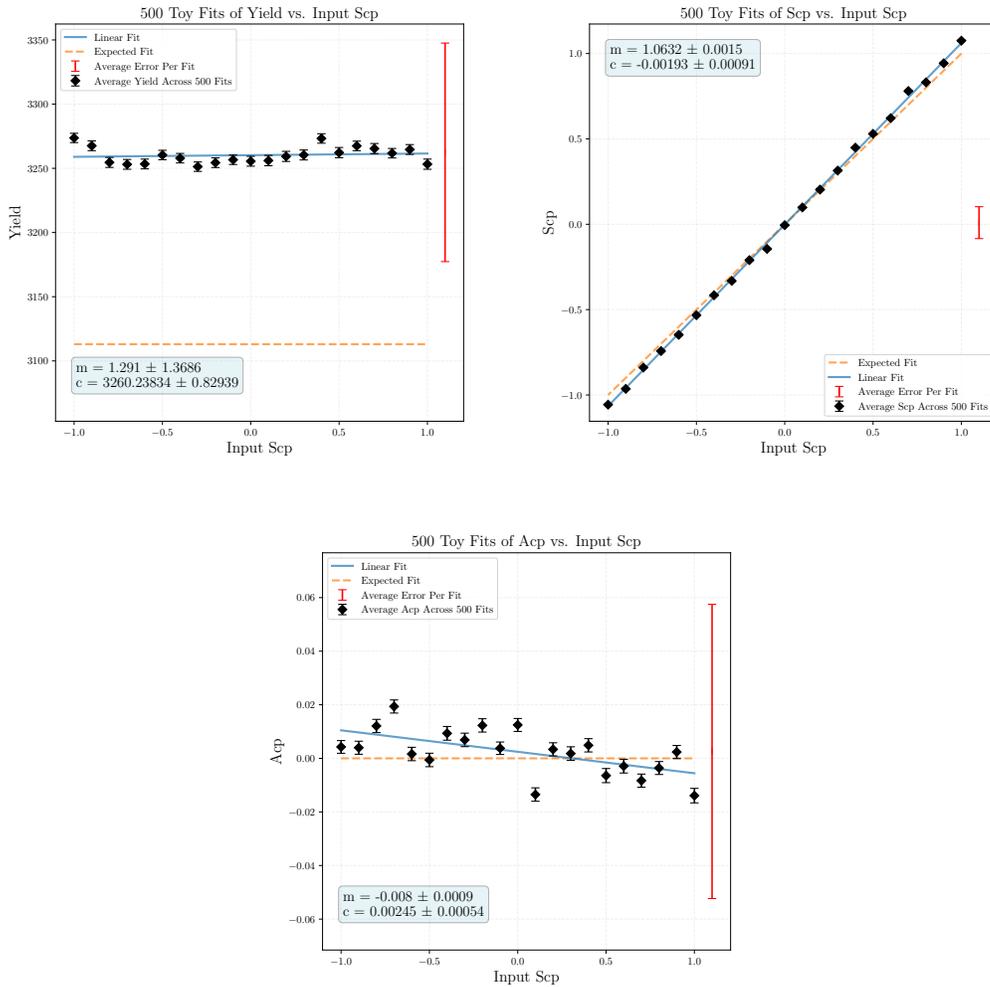
**Figure 4.6:** Assessing the bias of the signal mode with control mode resolution parameters. An increased bias is seen in the $\mathcal{S}_{CP}$ when compared with the baseline. The signal yield is significantly overestimated by the fitting program.

Model	Actual (Expected) Signal Yield	Gradient (Offset) $\mathcal{S}_{CP}$	Gradient (Offset) $\mathcal{A}_{CP}$	Mean Error in $\mathcal{S}_{CP}$
8-bin	3079 (3113)	0.984 (-0.0036)	-0.013 (0.0023)	0.0856
7-bin	3082 (3113)	0.988 (-0.0018)	-0.015 (0.0025)	0.0880
7-binMOD	3082 (3113)	0.984 (-0.0017)	-0.014 (0.0049)	0.0858
Per-Event	3136 (3113)	1.004 (0.0008)	-0.011 (0.0032)	0.0825
Tree Fitter	3249 (3315)	1.021 (-0.0021)	0.016 (0.0077)	0.0845
Per-Event + Tree Fitter	3300 (3315)	1.046 (-0.0027)	0.016 (0.0077)	0.0798
Control Mode	3260 (3113)	1.063 (0.0019)	-0.008 (0.0025)	0.0934

Table 4.1: A summary of the bias and performance from each model. The bias is measured by extracting the gradient and offset from a linear fit to different input $\mathcal{S}_{CP}$.

Integrated Luminosity (fb^{-1})	Statistical Significance (σ)	90% Upper Limit on Branching Ratio ($\times 10^{-6}$)	Mean Signal Yield Obtained (Expected)
5	1.83	8.28	2(3)
10	2.52	7.12	5(6)
20	3.56	6.49	11(12)
50	5.91	5.68	30(31)

Table 4.2: Outline of when a statistically significant signal yield can be extracted for $B^0 \to K_S^0 \pi^0$, as well as upper limits on the branching ratio set by these results.

should improve the accuracies of these estimations.

Integrated Luminosity (ab^{-1})	$\mathcal{A}_{CP}$ Significance	$\mathcal{S}_{CP}$ Significance	Combined Significance
5	2.47	4.79	5.71
10	3.59	6.79	7.95
20	5.13	9.95	> 8.00
30	6.39	> 8.00	> 8.00
40	7.69	> 8.00	> 8.00
50	> 8.00	> 8.00	> 8.00

Table 4.3: Assuming the theoretical and experimental values of $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ hold their central values, an estimate of the data required to obtain a resolution of the $B \rightarrow K\pi$ puzzle after. This is calculated by obtaining the statistical uncertainty from MC simulations.

Chapter 5

Conclusion and Further Work

In this thesis, the framework for the analysis of $B^0 \rightarrow K_S^0 \pi^0$ has been fully validated for 8-bins of qr with the vertexRAVE algorithm for determining decay position of the B_{sig}. By accounting for the experimental uncertainty of the resolution at a per-event basis, and using treeFitter for vertexing, we observe a 6.8% improvement in the statistical uncertainty of the measurement in $\mathcal{S}_{CP}$ for an ensemble of toy MC experiments. This comes at a cost of increasing the bias of $\mathcal{A}_{CP}$ to 8%. However if detector bias is accounted for, there should be an improvement in this result. With the analysis framework completedly established, other novel techniques (such as training continuum suppression on raw detector information) can be implemented in order to minimise error in $\mathcal{S}_{CP}$ without compromising the bias.

In this work, $B^0 \rightarrow J/\psi\, K_S^0$ as a control mode was was investigated, and a relationship in the resolution function is established with the signal mode. Although significant problems such as the asymmetric J/ψ vertex resolution function, and peaking BB background persist. The these issues, can be addressed by a modified vertexing constraint, and investigation into charm-backgrounds. Once these issues have been resolved, a natural continuation of this work invovles using real Belle II data to estimate the systematic uncertainties in the PDF parameters of the resolution function.

Although data-taking has already begun at Belle II, due to the low branching ratio of the $B^0 \rightarrow K_S^0 \pi^0$, we are unable to obtain enough statistics to detect this mode with the required 5σ statistical significance. However hopefully by 2020, an integrated luminosity of at least $50\,\text{fb}^{-1}$ will allows us rediscover this decay channel. By 2023, if $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$ maintains their central value (barring any issues with data-taking at the experiment), the $5\,\text{ab}^{-1}$ necessary to resolve this sector of the $B \rightarrow K\pi$ puzzle should be attainable.

References

[1] Ling-Lie Chau and Wai-Yee Keung. "Comments on the Parametrization of the Kobayashi-Maskawa Matrix". In: *Phys. Rev. Lett.* 53 (1984), p. 1802.

[2] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev.* D98.3 (2018), p. 030001.

[3] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. "Matter and Antimatter in the Universe". In: *New J. Phys.* 14 (2012), p. 095012. arXiv: `1204.4186 [hep-ph]`.

[4] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35.

[5] Patrick Huet and Eric Sather. "Electroweak baryogenesis and standard model CP violation". In: *Phys. Rev.* D51 (1995), pp. 379–394. arXiv: `hep-ph/9404302 [hep-ph]`.

[6] Vimal Simha and Gary Steigman. "Constraining The Early-Universe Baryon Density And Expansion Rate". In: *JCAP* 0806 (2008), p. 016. arXiv: `0803.3465 [astro-ph]`.

[7] T. D. Lee and Chen-Ning Yang. "Question of Parity Conservation in Weak Interactions". In: *Phys. Rev.* 104 (1956), pp. 254–258.

[8] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (1957), pp. 1413–1414.

[9] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Phys. Rev. Lett.* 13 (1964), pp. 138–140.

[10] Kazuo Abe et al. "Observation of large CP violation in the neutral B meson system". In: *Phys. Rev. Lett.* 87 (2001), p. 091802. arXiv: `hep-ex/0107061 [hep-ex]`.

[11] Bernard Aubert et al. "Observation of CP violation in the B^0 meson system". In: *Phys. Rev. Lett.* 87 (2001), p. 091801. arXiv: `hep-ex/0107013 [hep-ex]`.

[12] P. Kooijman and N. Tuning. *Lectures on CP violation (or: The Physics of Anti-matter) Particle Physics II.* 2015.

[13] Nicolas Boisvert Beaudry et al. "The $B \to \pi K$ puzzle revisited". In: *JHEP* 01 (2018), p. 074. arXiv: `1709.07142 [hep-ph]`.

[14] Alakabha Datta, Divya Sachdeva, and John Waite. "Unified explanation of $b \to s\mu^+\mu^-$ anomalies, neutrino masses, and $B \to \pi K$ puzzle". In: *Phys. Rev.* D100.5 (2019), p. 055015. arXiv: `1905.04046 [hep-ph]`.

[15] Robert Fleischer, Stefan Recksiegel, and Felix Schwab. "On Puzzles and Non-Puzzles in $B \to \pi\pi, \pi K$ Decays". In: *Eur. Phys. J.* C51 (2007), pp. 55–61. arXiv: `hep-ph/0702275 [HEP-PH]`.

[16] Yuuji Unno. "Experimental status and prospects of the 'Kpi puzzle'". In: *CKM unitarity triangle. Proceedings, 6th International Workshop, CKM 2010, Warwick, UK, September 6-10, 2010.* 2011. arXiv: `1101.0675 [hep-ex]`.

[17] Robert Fleischer et al. "Exploring $B \to \pi\pi, \pi K$ decays at the high-precision frontier". In: *Eur. Phys. J.* C78.11 (2018), p. 943. arXiv: `1806.08783 [hep-ph]`.

[18] Robert Fleischer et al. "Utilising $B \to \pi K$ Decays at the High-Precision Frontier". In: *53rd Rencontres de Moriond on QCD and High Energy Interactions (Moriond QCD 2018) La Thuile, Italy, March 17-24, 2018.* 2018. arXiv: `1805.06705 [hep-ph]`.

[19] M. Fujikawa et al. "Measurement of CP asymmetries in B0 —> K0 pi0 decays". In: *Phys. Rev.* D81 (2010), p. 011101. arXiv: `0809.4366 [hep-ex]`.

[20] Bernard Aubert et al. "Measurement of time dependent CP asymmetry parameters in B0 meson decays to omega K0(S), eta-prime K0, and pi0 K0(S)". In: *Phys. Rev.* D79 (2009), p. 052003. arXiv: `0809.1174 [hep-ex]`.

[21] Robert Fleischer et al. "Benchmarks for the New-Physics Search through CP Violation in B0 —> pi0 K(S)". In: *Phys. Rev.* D78 (2008), p. 111501. arXiv: `0806.2900 [hep-ph]`.

[22] W. Altmannshofer et al. "The Belle II Physics Book". In: (2018). Ed. by E. Kou and P. Urquijo. arXiv: `1808.10567 [hep-ex]`.

[23] T. Abe et al. "Belle II Technical Design Report". In: (2010). arXiv: `1011.0352 [physics.ins-det]`.

[24] T. Kuhr et al. "The Belle II Core Software". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 1. arXiv: `1809.04299 [physics.comp-ph]`.

[25] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86.

[26] Olaf Behnke et al., eds. *Data analysis in high energy physics.* Weinheim, Germany: Wiley-VCH, 2013.

[27] Aaron Roodman. "Blind analysis in particle physics". In: *eConf* C030908 (2003). [,166(2003)], TUIT001. arXiv: `physics/0312102 [physics.data-an]`.

[28] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth.* A462 (2001), pp. 152–155.

[29] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303.

[30] J. F. Krohn et al. "Global Decay Chain Vertex Fitting at B-Factories". In: (2019). arXiv: `1901.11198 [hep-ex]`.

[31] Geoffrey C. Fox and Stephen Wolfram. "Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes". In: *Phys. Rev. Lett.* 41 (1978), p. 1581.

[32] D. M. Asner et al. "Search for exclusive charmless hadronic B decays". In: *Phys. Rev.* D53 (1996), pp. 1039–1050. arXiv: `hep-ex/9508004 [hep-ex]`.

[33] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: (2007). arXiv: `physics/0703039 [physics.data-an]`.

[34] W Verkerke and D Kirkby. *RooFit users manual v2. 91.* 2008.

[35] H. Albrecht et al. "Search for Hadronic $b \to u$ Decays". In: *Phys. Lett.* B241 (1990), pp. 278–282.

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