SEARCH FOR $B^+ \to K^+\tau^+\tau^-$ AT THE BELLE EXPERIMENT AND PROSPECTS AT BELLE II

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHYSICS

DECEMBER 2024

By

Chris Ketter

Thesis Committee:

Peter Lewis, Chairperson
Thomas Browder
Jason Kumar
Sven Vahsen
Rufus Willett

Keywords: Belle, Belle II, KEK, EWP, KLM, firmware, TARGETX

To Gary S. Varner,

Instrumentation in high-energy particle physics has forever changed thanks to your brilliant idea for an oscilloscope on a chip, just as I have forever changed under your tutelage.

Rest in peace.

ACKNOWLEDGMENTS

I want to thank my original advisor, Gary S. Varner, for his persistance, patience, and optimism, without which this would not have been possible. To Sven Vahsen for agreeing to advise me after Gary's pasisng and for his encouragement to finish my dissertation. To Peter Lewis advising me in my final year and enthusiastically reviewing my work. To Thomas Browder for giving me a million chances and for always laughing at his own jokes. I am also thankful to the many others that have helped along the way, including but not limited to Kurtis Nishimura, Vasily Shebalin, Isar Mostafanezhad, Brandon Kunkler, Mikihiko Nakao, Matt Andrew, and Leo Pillonen. This work was supported by U.S. Department of Energy and Research Award No. DE-SC0010504.

ABSTRACT

The branching fraction for the electroweak penguin decay $B \to K\tau\tau$ has not been measured. Motivated by speculative physics scenarios that predict a large enhancement in the branching fraction of $B \to K\tau\tau$ compared to the Standard Model, and motivated further by recent success of the inclusive tagging technique, we develop an inclusive-hadronic-tagging technique for measuring the decay $B^+ \to K^+\tau^+\tau^-$. We use simulated Belle data equivalent in size to the $711\,\mathrm{fb}^{-1}$ recorded by the Belle experiment. The goal is to show that we can improve sensitivity over existing hadronic-tagging measurements of this decay. Based on simulation, we project an upper limit for $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 1.14 \times 10^{-4}$ at 90 % confidence level, a factor of about 20 better than the world's best constraint. The second part of this dissertation addresses a critical weakness in τ-lepton reconstruction, which is muon identification for τ decays to muon final states. As Belle II aims to record $50\,\mathrm{ab}^{-1}$ at the $\Upsilon(4\mathrm{S})$ resonance, it is of utmost importance that such calibration work is done now in order to enable future searches for this and other rare B-meson decays. Some of this poor muon identification can be attributed to lack of calibration and to unused features of the readout electronics for the Belle II K-Long and Muon (KLM) detector subsystem. We develop firmware that enables waveform readout of the silicon photomultipliers installed in the KLM detector. We employ a novel method for efficiently calibrating many channels at once, and we improve the hit-time granularity by a factor of four compared to the former firmware that did not use the waveform readout.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AC	Analog current
ACC	Aerogel Cherenkov counter
ADC	Analog-to-digital converter
ASIC	Application-specific integrated circuit
basf2	Belle-II analysis software framework
BDT	Boosted-decision tree
CDC	Central drift chamber
CKM	Cabibbo, Kobayashi, and Maskawa
CL	Confidence level
CMSSM	Constrained minimal supersymmetric model
DAC	Digital-to-analog converter
DC	Direct current
DSSD	Double-sided silicon strip detector
ECL	Electromagnetic calorimeter
EWP	Electroweak penguin
FCNC	Flavor-changing neutral current
FIFO	First-in first-out
FPGA	Field-programable gate array
FR	Full reconstruction
GDL	Global decision logic
GTLP	Gunning Transceiver Logic Plus
HER	High-energy ring
IP	Interaction point
JTAG	Joint test action group
KEK	High-energy accelerator research organization (Japanese)
KEKB	B factory at KEK
KLM	K-long and muon
LED	Light-emitting diode
LER	Low-energy ring
LFV	Lepton flavor violation
LU	Lepton universality
LUV	Lepton universality violation
MC	Monte Carlo

ML	Machine learning
mSUGRA	Minimal super gravity
PID	Particle identification
PIN	P-doped, intrinsic (undoped), N-doped
QCD	Quantum chromodynamics
QED	Quantum electrodynamics
RAM	Random-access memory
ROC	Receiver operating characteristic
RF	Radio frequency
RHIC	Ribbon-header interface card
RJ45	Registered jack 45
ROI	Region of interest
RPC	Resistive-plate counter
rSUGRA	Relaxed minimal super gravity
SCROD	Standard control and readout device
SiPM	Silicon photomultiplier
SM	Standard model
SPS	Single-photon spectra
SRAM	Static random-access memory
SSB	Spontaneous symmetry breaking
SVD	Silicon vertex detector
TARGET	TeV array readout with GSa/s sampling and event trigger
TDC	Time-to-digital converter
TOF	Time of flight
TSC	Trigger scintillation counter
VME	Versa Module Eurocard

PREFACE

Let us begin in a lighthearted way by means of a nutty and *quarky* story. Physicists are known for their whimsical naming of particles. We called them strange because we did not understand the strange nature of cosmic rays. We called them charm because the newly found strangeness was unsettling, and the restored symmetry that came with the charm felt somehow charming. Not only do we call the three generations of leptons *flavors*, but we have a conserved quantum number called flavor. Now we can talk about flavor conservation and not be referring to food preservation.

Going back to the ancient Greeks, Democritus and Leucippus thought that matter was grainy. But like peanut butter to peanuts, Aristotle thought it was infinitely divisible, infinitely smooth. Now fast forward many centuries. John Dalton noticed a pattern of discrete units at work and called them atoms after the Greek word $\alpha\tau o\mu o\sigma$, meaning "uncuttable." They did not remain uncuttable for long. Soon, J. J. Thompson's discovery of the electron cast some doubt. Ernest Rutherford's scattering experiments revealed the nucleus, this time named from Latin, *nucleus*, meaning "kernel." So matter is made out of nuts, with kernels inside and electron shells. Okay.

And Rutherford got to name yet another. Further scattering experiments led him to believe that the hydrogen atom has a nucleus with a singular positively charged particle. He called it the proton, from Greek $\pi\rho\omega\tau o\nu$ meaning "first," as in nothing before it. This came with a bonus new particle that must have a mass similar to the proton but without an electric charge. Sticking with the Greek -$o\nu$ suffices, W. D. Harkins gave it the name neutron. So matter is made of protons, neutrons, and electrons. Surely that's it. Well, naming something as first is just as bad of an idea as naming it uncuttable.

In 1964 when Murray Gell-Mann discovered a substructure to the proton and neutron that, oddly, comes in threes, he knew better than to apply a misnomer this time. He had the sound of the name he wanted in his head, "kwork," but not the spelling. Upon perusing the book Finnegans Wake by James Joyce, he came across the line

> Three quarks for Muster Mark,
> Sure he hasn't got much bark,
> And sure any he has it's all beside the mark.

Rather than quark with an "arrr" sound, he reimagined the meaning by Joyce as "Three *quarts* for Muster Mark" as though ordering ales for a friend. The spelling stuck, and to this day we do not know of any substructure to the quark, but we know better than to name it something that means indivisible. Nevertheless, we did name it after a measure of volume that will forever cement a connection to the imperial unit system.

Let us now talk about penguins.

This dissertation is about a rare transition between quarks (electroweak penguins) and about the development of modern electronics that can aid in detecting the decay products of the former.

In Part I we present a feasibility study on the use of a recently developed analysis technique to search for the rare B meson decay $B^+ \to K^+\tau^+\tau^-$. We begin with an overview of particle physics. The overview assumes the reader has a working knowledge of quantum mechanics and has a decent mathematics background, including linear algebra, group theory, and relativistic calculations using 4-vectors. However, a reader without the mathematics background may still glean some intuition about the standard model of particle physics. In Chapter 2 we introduce $B^+ \to K^+\tau^+\tau^-$. We cover existing measurements in this field, tests for lepton universality, and two pieces of motivation for this analysis, including a new physics model and a recent analysis on the search for $\mathrm{B}^+ \to \mathrm{K}^+\nu\overline{\nu}$. In Chapter 3 we cover the theoretical prediction for the branching fraction of $B^+ \to K^+\tau^+\tau^-$. We continue in Chapter 4 with a description of the Belle detector. In the rest of Part I, we go over the analysis itself, beginning with our reconstruction technique, then the application of machine learning for background rejection, a maximum likelihood estimate to fit our data, some discussion on systematic uncertainty, and finally a presentation of our findings on the feasibility of using this analysis technique for this particular decay.

In Part II, we switch gears and look to the future. Namely, we recognize the importance of muon detection to the success of this analysis, and we acknowledge that Belle II, already taking data, needed considerable work in the implementation of the readout electronics for its muon detection system. In Chapters 10 and 11 we describe the upgraded detector and readout electronics of the KLM subsystem between Belle and Belle II. In Chapter 12 and 15 we cover our work on the implementation of the readout electronics. Chapters 13 and 14 are dedicated to the upgrade of the other readout electronics for the portion of the KLM detector that was not replaced for the Belle II experiment. Finally, we provide some of our performance measurements on data collected from Belle II.

Part I

Part 1: $B^+ \to K^+\tau^+\tau^-$ Inclusive-Tagging Feasibility Study

CHAPTER 1
PARTICLE PHYSICS

The standard model of particle physics, also called the Weinberg-Salam model [1], is a model that explains the fundamental constituents of matter and three out of the known four forces of nature. It consists of matter particles called fermions and interaction particles called bosons (Fig. 1.1). The fermions are spin-1/2 particles that are categorized into quarks and leptons. The bosons are integer spin, including the spin-1 gauge bosons which are the carriers of the strong, electromagnetic, and weak forces, and the spin-0 Higgs boson which is an excitation of the Higgs field (a mechanism to generate particle masses in a gauge invariant way) [2].

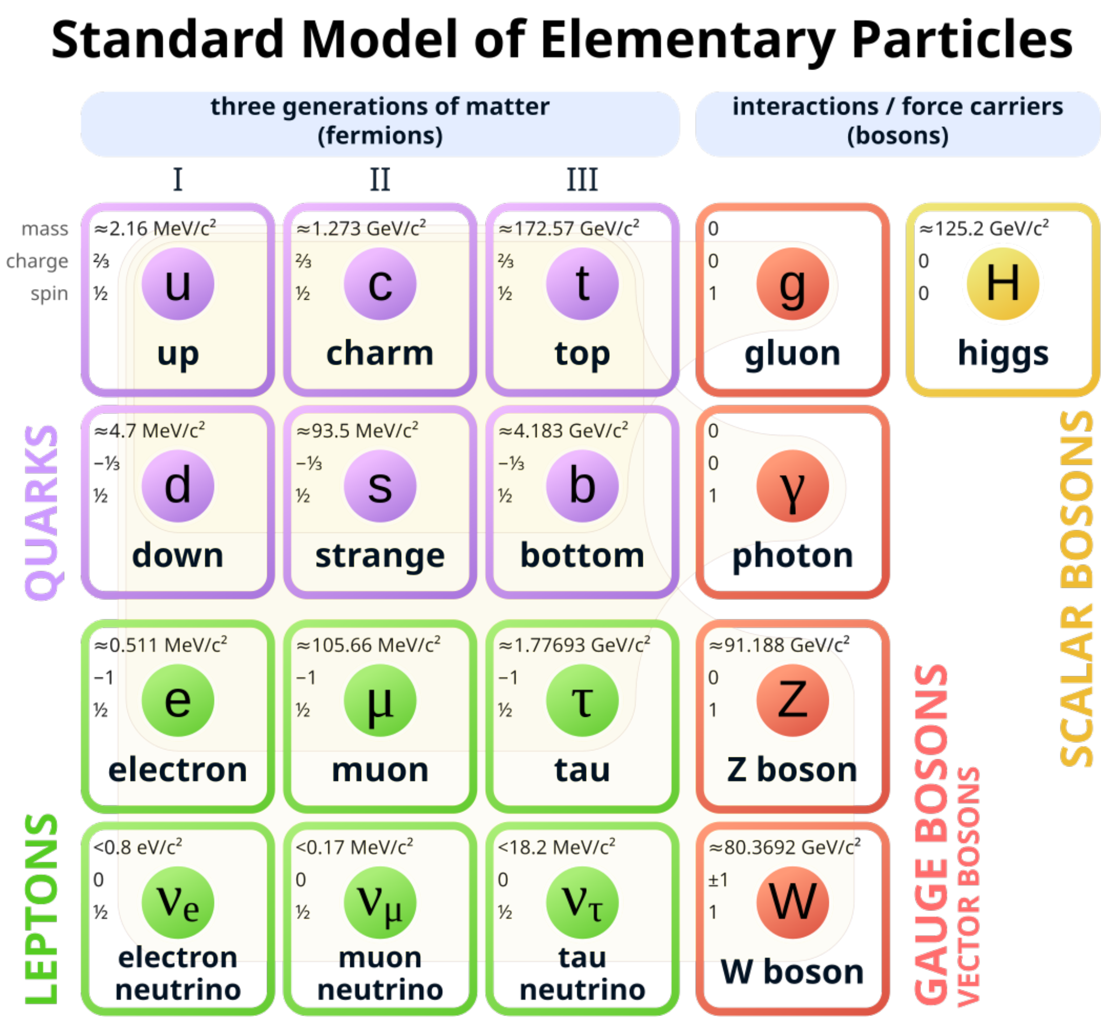

Figure 1.1: Elementary particles of the standard model.

There are six quarks and six antiquarks, all with fractional electric charge relative to that of a proton (which has an electric charge of +1). The *up*-type quarks (up, charm, top) all have $\pm 2/3$ electric charge and are denoted u, c, and t, while the *down*-type quarks (down, strange, bottom) all have $\mp 1/3$ electric charge and are denoted d, s, and b. Further, quarks carry a "color" charge (red, green, or blue) and antiquarks carry anti color charge ($\overline{\text{red}}$, $\overline{\text{green}}$, $\overline{\text{blue}}$). This is not to be confused with the human experience of seeing color in the visible part of the electromagnetic spectrum. The name color is an analogy to our experience with combining the three primary colors to make white, which we associate with colorless or color neutral. Color confinement and electric-charge quantization dictate that quarks can only exist in combinations, called hadrons (from $\dot{\alpha}\delta\rho\acute{o}\sigma$ meaning *thick*), wherein the total color charge is neutral and total electric charge is an integer value. Neutral color can be achieved by combining three quarks of color charges red, green, and blue (called baryons), and by combining two quarks with the same color/anticolor, e.g., red $-$ $\overline{\text{red}}$ (called mesons). A few examples are the proton, which consists of the quarks uud, the B^+ meson consisting of the quarks $\bar{b}u$, and the K^- meson consisting of the quarks $s\bar{u}$.

The gluon is the gauge boson that interacts with color charge, the so-called strong nuclear force. Unlike the electromagnetic or gravitational force, the strength of the strong force does not diminish with distance. Hence any attempt to split apart quarks results in energies exceeding the rest mass of a new quark/antiquark pair. In simpler terms, one could say the gluon breaks and the two new endpoints at the break are replaced with a quark and antiquark.

Besides the quarks, there are six leptons and six antileptons (from $\lambda\epsilon\pi\tau\acute{o}\sigma$ meaning *fine*). The electron and its antiparticle the positron ($e^{\mp}$), the muon and antimuon ($\mu^{\mp}$), and the tauon and antitauon ($\tau^{\mp}$) are all massive and have an integer electric charge. They can interact via the weak force, whose carriers are the massive W^+, W^-, and Z^0 bosons, and since they carry non-zero electric charge, they can interact via the electromagnetic force whose carrier is the photon. For each of these, there is a corresponding neutrino (ν_e, $\overline{\nu_e}$, ν_μ, $\overline{\nu_\mu}$, ν_τ, $\overline{\nu_\tau}$), all of which have zero charge, are nearly massless, and only interact via the weak force.

Everyday matter consists of up and down quarks and their antiparticles along with electrons, all of which are part of the first generation of matter (i.e., the first column in Fig. 1.1). Fermion masses (except perhaps for the neutrinos) increase substantially in the second generation and even more so in the third generation. To study the exotic particles containing these heavier fermions we generate them in the lab with particle colliders or wait for cosmic rays (mostly protons) to hit our atmosphere. The Belle experiment in particular collides electrons and positrons at a center-of-mass energy $\sqrt{s} = 10.58$ GeV to generate a bottomonium ($b\bar{b}$) resonance called Upsilon-4S, denoted $\Upsilon(4S)$.

It is prudent to mention a few things about bottomonium particles. Just as with the hydrogen atom, bottomonium has quantized energy levels (n), different spatial probability distributions (S-wave, P-wave, D-wave) based on angular momentum (l), and singlet (spin 0) or vector (spin 0, 1, or 2) states due to hyperfine splitting (e.g., ${}^1\mathrm{S}_0$, ${}^3\mathrm{S}_1$, ${}^1\mathrm{P}_1$, …). Unlike hydrogen, the order of 10 to 100 MeV energy differences between states inclines us to label each state as a different particle. The $l = 0$, $s = 0$ bottomonium are called $\eta_\mathrm{b}(\mathrm{nS})$, the $l = 0$, $s = 1$ bottomonium are called $\Upsilon(\mathrm{nS})$, the $l = 1$, $s = 0$ bottomonium are called $\mathrm{h_b}(\mathrm{nP})$, and the $l = 1$, $j = 0, 1, 2$ bottomonium (where $j = l + s$) are called $\chi_{\mathrm{bj}}(\mathrm{nS})$. B mesons contain a $\bar{b}$ quark paired with one quark from generation I. The $\Upsilon(4\mathrm{S})$ decays to $B\overline{B}$ pairs more than 96 % of the time. It is considered a quasi-bound state because its mass is larger than the mass of two B mesons. The next most massive in the group, the $\Upsilon(3\mathrm{S})$, is less massive than two B mesons. Therefore, to make a B factory, the center-of-mass energy of the collider should be large enough that $\Upsilon(4\mathrm{S})$ production is kinematically possible.

A second point worth mentioning is that when an electron and positron collide at a center-of-mass energy $\sqrt{s} = 10.58$ GeV, an $\Upsilon(4\mathrm{S})$ is not guaranteed. There is a whole host of lighter resonances that can be produced instead. We call this continuum background. However, continuum processes, like $e^+e^- \to c\bar{c}$, for example, tend to hadronize into two jets. This is because $\sqrt{s}$ is substantially larger than the $c\bar{c}$ rest mass. The excess energy, therefore, must go into momentum. Conservation of momentum tells us that it must be split equally between the quark and anti-quark, and in opposite directions.

1.1 Quantum Field Theory

The following sections up to Section 1.6 relied heavily on Ref. [3] with help from Refs. [4, 5]. The Schrödinger equation for a free particle's wave function $\phi(\vec{r}, t)$ (in units $\hbar = c = 1$) is

$$-\frac{1}{2m}\nabla^2\phi = i\frac{\partial}{\partial t}\phi\,, \tag{1.1}$$

where m is the particle's mass and ∇^2 is the divergence $\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 + \left(\frac{\partial}{\partial z}\right)^2$. Although very successful in some of its predictions, it has some major drawbacks. Namely, it is not a relativistic theory, and it cannot account for the annihilation or creation of matter. When making the Schrödinger equation fully relativistic via the relation $m^2 = E^2 - p^2$ one gets the Klein-Gordon equation

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2\right)\phi = m^2\phi, \tag{1.2}$$

which is 2nd order in the time derivative. Using a mostly-minus Minkowski metric and a 4-gradient $\partial^\mu = \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)$, 1.2 can be written

$$(\partial_\mu\partial^\mu + m^2)\phi = 0. \tag{1.3}$$

Paul Dirac sought a way to make the equation first order in time. He simply wrote down a version as a product of two factors and demanded that the cross terms cancel out. The result is called the

Dirac equation,

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \tag{1.4}$$

where the γ^μ are the 4×4 gamma matrices satisfying the needed anticommutation relation that cancels out the unwanted cross terms in Dirac's derivation. In reconciling this equation with reality by asking what are the real observables (in quantum mechanics real observables are the expectation values of Hermitian operators), a major problem arose wherein the expectation value of the Hamiltonian, i.e., energy, was allowed to take on negative values. It was known that the negative energy problem could be resolved with a positively charged particle, but the proton was too massive to fit the bill and no other such particle was known at the time. This problem was ultimately resolved with the discovery of the positron.

Solutions to these equations can be thought of as excitations of quantum fields, a spin-0 scalar field in the case of the Klein-Gordon equation and a spin-1/2 field in the case of the Dirac equation. Today, we use the terms from both of these equations in parts of the standard model Lagrangian to represent free particles.

The standard model is a quantum field theory with a local gauge symmetry made up of the groups

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y, \tag{1.5}$$

where the $SU(3)$ symmetry describes interactions between quarks and gluons while $SU(2) \otimes U(1)$ pertains to the electroweak sector. The subscript C refers to the color charge, L to the left chirality of the weak interaction, and Y to the weak hypercharge. Starting with a free-particle Lagrangian, one can demand that the Lagrangian be invariant under local transformations of U(1), SU(2), and SU(3). Global invariance is already satisfied, for example, a global U(1) transformation amounts to a phase shift everywhere and the Lagrangian is unchanged, but a phase shift that depends on a space-time coordinate is not satisfied without adding an extra term to the Lagrangian. Such new terms then describe interactions between particles, including mass terms for the bosons. This exercise predicts a non-force-carrying boson called the Higgs after one applies spontaneous symmetry breaking to account for non-zero vacuum energy. A particular choice of the direction in which to break the symmetry rotates the masses of the other bosons in such a way that the Z^0 and $W^\pm$ become very massive. It was this feature that made the Higgs boson worth searching for for several decades.

With a Lagrangian in hand, calculating real-world phenomena like scattering cross sections and decay rates can be just a matter of writing the corresponding Hamiltonian and performing perturbation theory. However, not all processes are perturbative. The strong coupling constant, α_s, increases at long range (low energy), which makes higher-order terms in perturbation theory contribute more than lower-order terms.

1.2 Quantum Electrodynamics

If we begin with a free-particle Lagrangian for a fermion

$$\mathcal{L}_0 \;=\; i\,\overline{\psi}(x)\gamma^\mu\partial_\mu\psi(x) \;-\; m\,\overline{\psi}(x)\psi(x), \tag{1.6}$$

and require that it be locally invariant under U(1) transformations, we pick up a term involving a spin-1 gauge field in the Lagrangian that describes the interaction between fermions and the electromagnetic field:

$$\mathcal{L} = \mathcal{L}_0 \;+\; eQA_\mu(x)\,\overline{\psi}(x)\gamma^\mu\psi(x). \tag{1.7}$$

For this new gauge field to have a free-particle term associated with it, we add

$$\mathcal{L}_{\rm kin} \;\equiv\; -\frac{1}{4}\,F_{\mu\nu}(x)\,F^{\mu\nu}(x)\;, \tag{1.8}$$

where $F_{\mu\nu} \;\equiv\; \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field-strength tensor. Adding a mass term would render our Lagrangian no longer invariant under U(1) transformations, so the gauge boson associated with this new gauge field must be massless, i.e., the photon. When 1.8 is combined with 1.7 one can derive the Maxwell equations

$$\partial_\mu F^{\mu\nu} = J^\nu\,, \qquad\qquad J^\nu = -eQ\,\overline{\psi}\gamma^\nu\psi\;, \tag{1.9}$$

where $\partial_\mu F^{\mu\nu} = J^\nu$ and $J^\nu = -eQ\,\overline{\psi}\gamma^\nu\psi$, where J^ν is the electromagnetic 4-current. This is the quantum electrodynamics (QED) Lagrangian.

1.3 Quantum Chromodynamics

Let q_f represent a quark field of flavor f. The free particle Lagrangian is then

$$\mathcal{L}_0 \;=\; \sum_f\; \bar{q}_f\;(i\gamma^\mu\partial_\mu \;-m_f)\,q_f \tag{1.10}$$

Noting that quarks carry a color charge of one of three colors (or an anticolor charge for antiquarks), we require that our Lagrangian be locally invariant under SU(3) transformations. The result, just as above for QED, is eight massless gauge bosons, $G^\mu_a(x)$, which are the gluons. Defining $G^{\mu\nu}_a$ as the quantum chromodynamics (QCD) field-strength tensor and $D^\mu \equiv [\partial^\mu - ig_s\,G^\mu(x)]$ as our covariant derivative, we can now write down the QCD Lagrangian:

$$\mathcal{L}_{\rm QCD} \;\equiv\; -\frac{1}{4}\,G^{\mu\nu}_a G^a_{\mu\nu} + \sum_f\; \bar{q}_f\;(i\gamma^\mu D_\mu - m_f)\;q_f\,. \tag{1.11}$$

it is hidden in the notation, but this Lagrangian contains generators for SU(3) matrices and structure constants for SU(3). The interaction terms include quark-gluon interactions and gluon-gluon interactions, unlike the case of QED which does not have photon-photon interactions.

1.4 The Electroweak Interaction

To understand the symmetry of the electroweak interaction, first note some observations which that symmetry must describe.

1. Neutrinos have a left-handed chirality and antineutrinos have a right-handed chirality.

2. Only left-handed fermions and right-handed antifermions participate in charged-current weak interactions (those involving a $W^{\pm}$ boson).

3. The mass eigenstates of the quarks (states before a decay) are different from the weak eigenstates (states after a decay). They are related by a 3×3 rotation matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{weak} = \begin{pmatrix} \mathrm{V}_{ud} & \mathrm{V}_{us} & \mathrm{V}_{ub} \\ \mathrm{V}_{cd} & \mathrm{V}_{cs} & \mathrm{V}_{cb} \\ \mathrm{V}_{td} & \mathrm{V}_{ts} & \mathrm{V}_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{mass} , \tag{1.12}$$

   which is known as the CKM matrix.

4. Left-handed (right-handed) fermions (antifermions) appear in flavor doublets, e.g.,

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L , \quad \begin{pmatrix} \overline{\nu}_\mu \\ \mu^+ \end{pmatrix}_R , \quad \begin{pmatrix} u \\ d \end{pmatrix}_L , \quad etc.$$

5. The photon interacts with fermions of left- and right-handed chiralities equally whereas the Z boson acts differently between the two. Further, the Z-boson-to-neutron interaction only involves left-handed chiralities.

The symmetry describing rotations between flavor eigenstate doublets is SU(2). To include electromagnetic interactions we also require U(1) symmetry and arrive at

$$SU(2)_L \otimes U(1)_Y \ . \tag{1.13}$$

Requiring local invariance under transformations of this symmetry, as with QED and QCD, generates the electroweak interaction terms and four massless gauge bosons.

The charged-current Lagrangian is

$$\mathcal{L}_{\rm CC} \;=\; \frac{g}{\sqrt{2}} \left\{ W_\mu^\dagger \left[\sum_{ij} \bar{u}_i \, \gamma^\mu P_L \, \mathbf{V}_{ij} \, d_j \;+\; \sum_l \bar{\nu}_l \, \gamma^\mu P_L \, l \right] \;+\; \text{h.c.} \right\}, \tag{1.14}$$

where W represents four fields for massless gauge bosons (called the Goldstone bosons) and P_L is the left projection operator defined as $\frac{1-\gamma_5}{2}$. We want to identify W_1 and W_2 with the $W^\pm$ and W_3 and W_4 with neutral bosons. To get the desired difference in behavior between photons and Z bosons, the electroweak mixing angle θ_W is introduced

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \equiv \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} . \tag{1.15}$$

Now we can write the neutral current Lagrangian as

$$\mathcal{L}_{\rm NC} = \sum_i \overline{\psi}_j \, \gamma^\mu \left\{ A_\mu \left[g \, \frac{\sigma_3}{2} \, \sin\theta_W + g' \, y_j \, \cos\theta_W \right] \;+\; Z_\mu \left[g \, \frac{\sigma_3}{2} \, \cos\theta_W - g' \, y_j \, \sin\theta_W \right] \right\} \, \psi_j \, . \tag{1.16}$$

Finally, to recover QED, we choose θ_W such that

$$g \sin(\theta_W) = g' \cos(\theta_W) = e \tag{1.17}$$

and the electroweak hypercharge via the Gell-Mann–Nishijima formula

$$Y = 2(Q - T_3), \tag{1.18}$$

where Q is the electric charge and T_3 is the third component of the weak isospin. A summary of the standard model fermions arranged in left-handed doublets and right-handed singlets, along with their weak isospin (magnitude and 3rd component), their electroweak hypercharge, and their electric charge is listed in Table 1.1.

Table 1.1: Standard model fermions with representation (doublet or singlet), chirality, weak isospin (T, T_3), and charge.

I	II	III	T	T_3	Y	Q
$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	$\frac{1}{2}$	$+1/2$ $-1/2$	-1	0 -1
e_R	μ_R	τ_R	0	0	-2	-1
$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	$\frac{1}{2}$	$+1/2$ $-1/2$	$+1/3$	$+2/3$ $-1/3$
u_R	c_R	t_R	0	0	$+4/3$	$+2/3$
d_R	s_R	b_R	0	0	$-2/3$	$-1/3$

1.5 Spontaneous Symmetry Breaking

The choice of θ_W almost gave us the $W^\pm$, Z, and γ bosons, except they are all four massless. If one writes a Lagrangian for two $SU(2)_L$ complex scalar fields, they will identify two terms corresponding to the potential energy. The potential well for this two-dimensional potential energy function takes the shape of a Mexican hat, depending on the sign of an arbitrary constant. Rather than choosing the zero-state energy to lie at the point of the hat, one can choose their zero state at a lower potential in the rim of the hat. This is called the spontaneous breaking of a continuous global symmetry (SSB). Doing so, however, always introduces one or more massless spin-0 particles. These are the Goldstone bosons. An astute choice of phase, i.e., precisely where in the rim you call zero, has the effect of transforming away the Goldstone bosons leaving behind a massive scalar particle — the Higgs boson — and a massive gauge field. Further, when the gauge field picked up mass, it acquired a 3$^{\text{rd}}$ polarization state. This is the Higgs mechanism, and it is responsible for giving mass to the $W^\pm$ and Z. The Goldstone bosons may be thought of as components of the spin-0 Higgs field. A common way that this is described is that the $W^\pm$ and Z *ate* the Goldstone bosons and acquired mass in the process. The Higgs mechanism also gives masses to the remaining massive particles in the standard model. That the Higgs mechanism gives mass to the $W^\pm$ and Z made the theory compelling enough to build the Large Hadron Collider to look for the Higgs boson. The connection between electroweak theory and QED can now be summarized as

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \quad \stackrel{SSB}{\longrightarrow} \quad SU(3)_C \otimes U(1)_{\text{QED}} \,. \tag{1.19}$$

1.6 Performing Calculations

The connection between the theory described above and the measurement of a phenomenon, such as a branching fraction, comes from perturbation theory. To first order in perturbation theory, the amplitude for decay from an initial state (I) to a final state (F) is given by

$$\mathcal{A}(I \to F) = \langle F|\mathcal{H}_{int}|I\rangle, \tag{1.20}$$

also called the matrix element, $\mathcal{M}$, and $\mathcal{H}_{int}$ is the interaction Hamiltonian, $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{int}$. When computing a matrix element, the field operators are written in terms of field equations with annihilation and creation operators that create or destroy quanta of their respective fields. The factoring of the matrix element involves a lot of commutation relations of the Dirac algebra. The factored form has interaction terms and propagator terms which can be represented graphically as the vertices and connectors of a Feynman diagram. At higher orders in perturbation theory, powers of $\mathcal{H}_{int}$ result in internal loops in the diagrams. A single matrix element is usually represented by multiple diagrams.

To get a width, Fermi's golden rule is used:

$$d\Gamma = |\mathcal{M}|^2 d\rho(\mathbf{p}_f) \tag{1.21}$$

where $\rho(\mathbf{p}_f)$ is the density of final states which includes normalization factors, delta functions to conserve energy and momentum, and the continuous phase space of kinematically allowed final-state momenta. The natural unit of Γ is [Energy]. Since the natural unit of time is 1/[Energy], the proper decay time is the inverse of Γ, and Γ itself also describes the decay rate. The branching fraction, $\mathcal{B}(I \to F)$, is $\Gamma_{I\to F}/\Gamma_{I\to X}$, where $\Gamma_{I\to X}$ is the inclusive decay width of $I \to$ anything.

CHAPTER 2
INTRODUCTION TO $B^+ \to K^+\tau^+\tau^-$

The decay $B^+ \to K^+\tau^+\tau^-$ is part of the $b \to s\ell\ell$ family of decays. As it involves a transition from a b quark to an s quark, which both have the same electric charge, it is an example of a flavor-changing neutral current (FCNC). In the standard model of particle physics, this decay does not happen at tree level but rather requires additional loops in the diagram. It can proceed via the electroweak penguin (EWP) diagram or the W^+W^- box diagram (Fig. 2.1). Rare decays like these, because

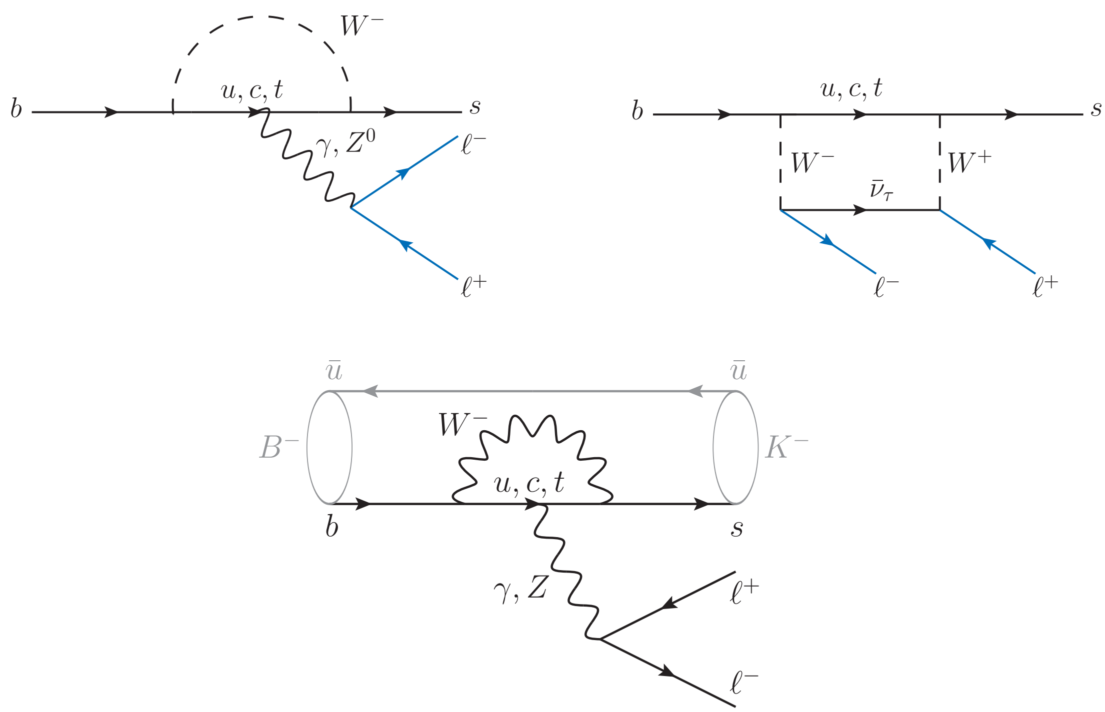

Figure 2.1: Diagrams for $b \to s\ell\ell$. Left: electroweak penguin. Right: W^+W^- box. Bottom: $\overline{u}$ quark included to elucidate $B^+ \to K^+\tau^+\tau^-$.

they contain loops, are a good probe for beyond-standard-model physics. New physics processes can enter via virtual particles at a level comparable to the standard model processes alone [6].

2.1 Existing Measurements

The decay $B^+ \to K^+\tau^+\tau^-$, along with its neutral $B^0 \to K^0\tau^+\tau^-$ counterpart and the $K^*(892)$ counterparts, have not been observed. The current world-best upper limit for the $B^+ \to K^+\tau^+\tau^-$ branching fraction is 2.25×10^{-3} at a 90 % confidence level (CL), which was published by BaBar in 2017 [7]. The Belle experiment has not published any journal results for this decay mode, but a recent PhD dissertation from Universite Paris-Saclay which used hadronic tagging to search for $B^+ \to K^+\tau^+\tau^-$ in the combined Belle and Belle II dataset of $1073\,\mathrm{fb}^{-1}$ established an upper limit of 6×10^{-4} at a 90 % CL [8].

As for the neutral counterparts, a 2023 Belle analysis established an upper limit on $B^0 \to K^{*0}\tau^+\tau^-$ of 3.1×10^{-3} at 90 % CL, also using hadronic tagging.

Existing measurements of $B^+ \to K^+\ell^+\ell^-$ include $\mathcal{B}(B^+ \to K^+e^+e^-)$ at $(5.6 \pm 0.6) \times 10^{-7}$ and $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$ at $(4.53 \pm 0.35) \times 10^{-7}$. The modes with $K^*(892)^+$ are about twice as prevalent with $\mathcal{B}(B^+ \to K^*(892)^+e^+e^-)$ at $(1.55 + {}^{+\,0.40}_{-\,0.31}) \times 10^{-6}$ and $\mathcal{B}(B^+ \to K^*(892)^+\mu^+\mu^-)$ at $(9.6 \pm 1.0) \times 10^{-7}$ [9]. We summarize these and the corresponding B^0 decay modes in Table 2.1.

Table 2.1: Summary of $b \to s\ell\ell$ measurements from Ref. [9].

Decay	$\mathcal{B}$		Significance
$B^+ \to K^+e^+e^-$	(5.6 ± 0.6)	$\times 10^{-7}$	—
$B^+ \to K^+\mu^+\mu^-$	(4.53 ± 0.35)	$\times 10^{-7}$	1.8
$B^+ \to K^*(892)^+e^+e^-$	$(1.55 {}^{+\,0.40}_{-\,0.31})$	$\times 10^{-6}$	—
$B^+ \to K^*(892)^+\mu^+\mu^-$	(9.6 ± 1.0)	$\times 10^{-7}$	—
$B^0 \to K^0e^+e^-$	$(2.5 {}^{+\,1.1}_{-\,0.9})$	$\times 10^{-7}$	1.3
$B^0 \to K^0\mu^+\mu^-$	(3.39 ± 0.35)	$\times 10^{-7}$	1.1
$B^0 \to K^*(892)^0e^+e^-$	$(1.03 {}^{+\,0.19}_{-\,0.17})$	$\times 10^{-6}$	—
$B^0 \to K^*(892)^0\mu^+\mu^-$	(9.4 ± 0.5)	$\times 10^{-7}$	—

2.1.1 Standard Model Tests

In the standard model, the electroweak interaction strength between the three generations of leptons is equal. For example, one would expect the decays $b \to ce\overline{\nu}_e$, $b \to c\mu\overline{\nu}_\mu$, and $b \to c\tau\overline{\nu}_\tau$ to be equally likely aside from differences in the amount of momentum phase space available to each. We call this lepton universality (LU). It is an accidental symmetry, that is to say, the standard model does not have a mechanism to break LU, but it is not a fundamental symmetry of the theory. We test it by looking for examples of lepton-universality violation (LUV). Note that lepton flavor is conserved in each of the above examples due to the antineutrinos of the same generations. If lepton flavor were not conserved, it would be an example of lepton-flavor violation (LFV), which is distinctly different from LUV.

Some tests for LUV include

$$\mathcal{R}(h) = \frac{\mathcal{B}(B \to hL...)}{\mathcal{B}(B \to h\ell...)}, \tag{2.1}$$

where h is a hadron, L is a heavier lepton (τ or μ depending on the test), and ℓ is a lighter lepton

($\ell = e, \mu$ or $\ell = e$). Measuring ratios of branching fractions has the advantage that a lot of the systematic uncertainties, both in theory and experiment, cancel out in the ratio. Another test of LUV is $\Delta\mathcal{A}_{FB}$, where the forward-backward asymmetry

$$\mathcal{A}_{FB} = \frac{1}{\Gamma}\left(\int_{-1}^{0} - \int_{0}^{1}\right) d\cos\theta \frac{d\Gamma}{d\cos\theta} \tag{2.2}$$

is compared between decay modes that only differ by the flavor of the final-state lepton(s). We summarize some measurements of these quantities in Fig. 2.2.

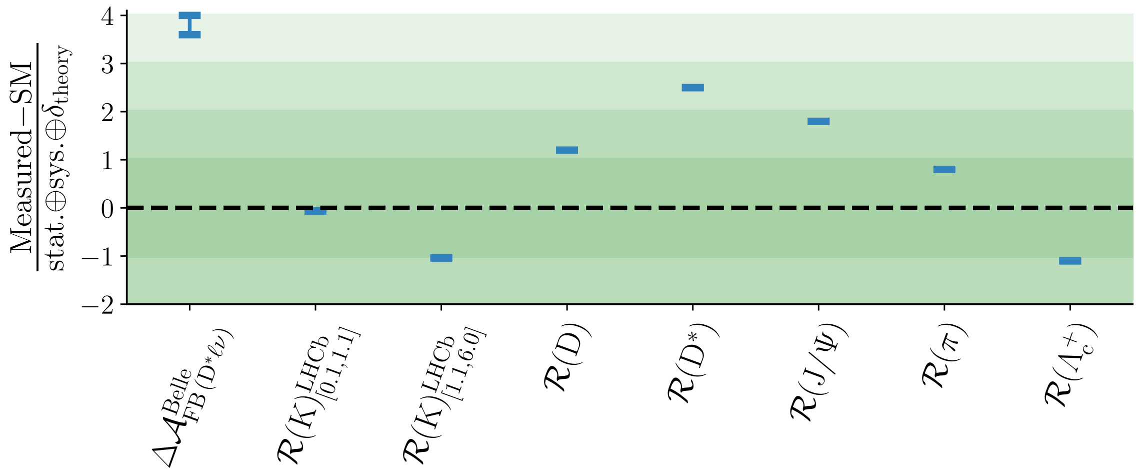

Figure 2.2: Lepton universality measurements. The vertical axis indicates deviation from the standard model prediction normalized by the uncertainties provided in each analysis. The error bars for $\Delta A_{FB}(D^*\ell\nu)$ indicate a range provided by the authors in Ref. [10] that depends on the extent of correlation between the electron and muon modes. The previous LHCb result for $\mathcal{R}(K)$ of $3.1\,\sigma$ disappeared in late 2022 [11].

2.2 Motivation

The motivation for this analysis began with LHCb run1 results showing a $2.5\,\sigma$ deviation from SM predictions in $\mathcal{R}(K)$, and later a $3.1\,\sigma$ deviation with run1 and run2 combined. This deviation, however, disappeared by late 2022 when LHCb announced they had incorrectly modeled the pion background in their earlier $B^+ \to K^+e^+e^-$ analyses. Nevertheless, if there is a deviation from SM predictions in $\mathcal{R}(K)^{\mu}_{e}$, that is $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)/\mathcal{B}(B^+ \to K^+e^+e^-)$, then it would be interesting to measure $\mathcal{R}(K)^{\tau}_{\ell}$, where $\ell = \mu \; or \; e$. Aside from $\mathcal{R}(K)$, we had additional motivation for measuring $B^+ \to K^+\tau^+\tau^-$, not only because it had not been measured in Belle data, but for two more reasons — a theory paper predicting enhancements in $b \to s\tau^+\tau^-$ processes, and a Belle II analysis of $B^+ \to K^+\nu\bar{\nu}$.

2.2.1 New Physics Predictions

S. Descotes-Genon looked at measurements of lepton-universality violation in the charm sector and considered how they might imprint on $b \to s\tau^+\tau^-$ [12]. Specifically, they find that by assuming a common new physics explanation for R_D, $R_D^{(*)}$, and $R_{J/\psi}$, then new physics contributions to $b \to c\tau^-\overline{\nu}_\tau$ and $b \to s\tau^+\tau^-$ are generated together. Such a model could be realized by the presence of a leptoquark. Parameterizing their predictions by the ratio $\frac{R_X}{R_X^{SM}}$, they plot the expected branching fractions for various $b \to s\tau^+\tau^-$ processes (Fig. 2.3). For example, taking $\frac{R_X}{R_X^{SM}} = 1.27$, which is

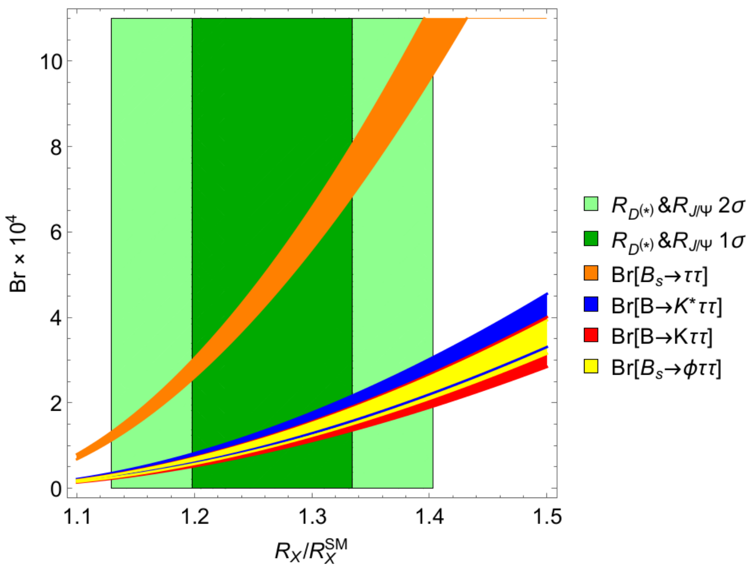

Figure 2.3: Various branching fraction predictions for $b \to s\tau^+\tau^-$ processes shown in [12]. For reference, SM predictions for these branching fractions are on the order of 10^{-7}. At $R_X/R_X^{SM} = 1.27$, our mode, $B \to K\tau^+\tau^-$, has a branching fraction prediction around 1×10^{-4}

the central value for measurements of $\frac{R_{D^{(*)}}}{R_{D^{(*)}}^{SM}}$ and $\frac{R_{J/\psi}}{R_{J/\psi}^{SM}}$, they predict $\mathcal{B}(B \to K\tau^+\tau^-) \approx 1 \times 10^{-4}$, nearly three orders of magnitude larger than the standard model prediction.

2.2.2 $B \to K\nu\overline{\nu}$

By 2021, Belle II had an exciting result on the search for $B \to K\nu\overline{\nu}$ using a novel inclusive tagging method. The analysis [13] on only $63\,\text{fb}^{-1}$ of data from Belle II achieved results competitive with earlier analyses, on about 10 times the data, from Belle [14, 15] and BaBar [16]. We share their comparison in Fig. 2.4.

The success of the inclusive tagging method when used for a decay mode with a lot of missing energy was a primary motivation for this analysis.

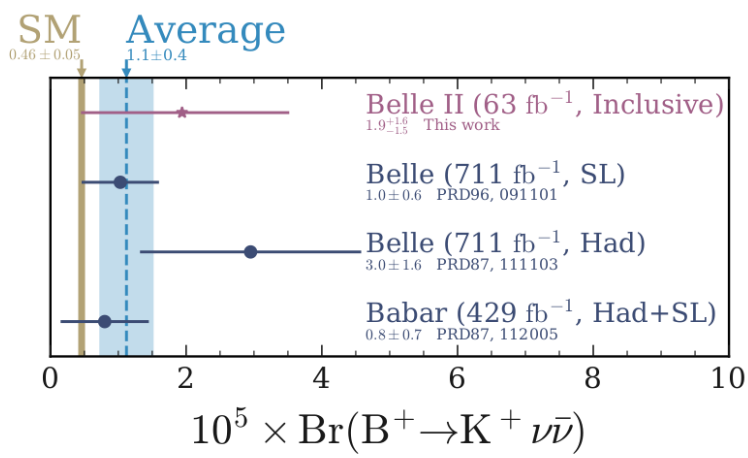

Figure 2.4: $\mathcal{B}(B^+ \to K^+\nu\overline{\nu})$ comparisons between different data sets and different tagging methods. This figure is shown in the appendix (preprint only) of [13].

CHAPTER 3
THEORY

3.1 Effective Field Theory

The standard-model branching fraction for $B \to K\tau^+\tau^-$ is calculated using an effective field theory. The Hamiltonian is projected as an effective Hamiltonian [17] using an operator product expansion [18, 19]. In this process, the short-range effects that can be calculated perturbatively are separated from the long-distance processes by integrating out the heavy degrees of freedom. In the present case, we use an effective Hamiltonian at the energy scale of the b mass giving us an effective five-quark theory. For one-loop processes like $b \to s\ell\ell$ transitions, the effective Hamiltonian used is

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_{i=1}^{10} \mathcal{C}_i(\mu)\mathcal{O}_i(\mu) + h.c., \tag{3.1}$$

where G_F is the Fermi constant, $V_{qq'}$ are CKM matrix elements, the $\mathcal{C}_i$ are the Wilson coefficients which encode the short-range effects, and the $\mathcal{O}_i$ are the local operators

$$\mathcal{O}_1 = (\bar{s}_i\gamma_\mu P_L c_j)\,(\bar{c}_j\gamma^\mu P_L b_i)\,, \tag{3.2}$$

$$\mathcal{O}_2 = (\bar{s}_i\gamma_\mu P_L c_i)\,(\bar{c}_j\gamma^\mu P_L b_j)\,, \tag{3.3}$$

$$\mathcal{O}_3 = (\bar{s}_i\gamma_\mu P_L b_i) \sum_{q=u,\, d,\, s,\, c,b} \left(\bar{q}_j\gamma^\mu P_L q_j\right), \tag{3.4}$$

$$\mathcal{O}_4 = (\bar{s}_i\gamma_\mu P_L b_j) \sum_{q=u,\, d,\, s,\, c,b} \left(\bar{q}_j\gamma^\mu P_L q_i\right), \tag{3.5}$$

$$\mathcal{O}_5 = (\bar{s}_i\gamma_\mu P_L b_i) \sum_{q=u,\, d,\, s,\, c,b} (\bar{q}_j\gamma^\mu P_R q_j), \tag{3.6}$$

$$\mathcal{O}_6 = (\bar{s}_i\gamma_\mu P_L b_j) \sum_{q=u,\, d,\, s,\, c,b} (\bar{q}_j\gamma^\mu P_R q_i), \tag{3.7}$$

$$\mathcal{O}_7 = \frac{e}{16\pi^2}\bar{s}_i\sigma_{\mu\nu}\,(m_s P_L + m_b P_R)\, b_i F^{\mu\nu}, \tag{3.8}$$

$$\mathcal{O}_8 = \frac{g}{16\pi^2}\bar{s}_i\sigma_{\mu\nu}(m_s P_L + m_b P_R)T^a_{ij} b_j G^{a\mu\nu}, \tag{3.9}$$

$$\mathcal{O}_9 = \frac{e^2}{16\pi^2}(\bar{s}_i\gamma^\mu P_L b_i)(\bar{\ell}\gamma_\mu \ell), \tag{3.10}$$

$$\mathcal{O}_{10} = \frac{e^2}{16\pi^2}(\bar{s}_i\gamma^\mu P_L b_i)(\bar{\ell}\gamma_\mu\gamma_5 \ell). \tag{3.11}$$

Here, u, d, s, c, and b are quark fields, where the subscripts i and j are color-charge indices. The ℓ is a lepton field. The overlines represent antiparticles. The P_L and P_R are shorthand for the left and right projections $\frac{1-\gamma_5}{2}$ and $\frac{1+\gamma_5}{2}$, respectively. The e and g represent the electron charge and strong coupling constant. The commutator $\frac{i}{2}[\gamma^\mu, \gamma^\nu]$ is denoted $\sigma^{\mu\nu}$. The masses of the strange and bottom quarks are denoted m_s and m_b. Finally, $F^{\mu\nu}$ is the electromagnetic field-strength tensor while $G^{a\mu\nu}$ is the strong field-strength tensor, the latter being accompanied by T^a_{ij} which is the generator for $SU(3)$ color symmetry.

$\mathcal{O}_{1-2}$ are the current-current operators, $\mathcal{O}_{3-6}$ are the quantum chromodynamic penguin operators, $\mathcal{O}_7$ is the electromagnetic operator, $\mathcal{O}_8$ is the chromomagnetic operator, $\mathcal{O}_9$ is vector component of the electroweak penguin operator, and $\mathcal{O}_{10}$ is the axial-vector component of the electroweak penguin operator.

To leading order, only the operators $\mathcal{O}_{7,9,10}$ contribute to $b \to s\ell\ell$.

3.2 Branching Fraction Calculation

Starting with the effective Hamiltonian above, the theoretical branching fraction for $B \to K\ell\ell$ has been calculate by Bobeth et al. in Ref. [20]. With only $\mathcal{O}_{7,9,10}$, we have

$$\begin{aligned} H_{\text{eff}} = &-\frac{G_\text{F}}{\sqrt{2}}\frac{\alpha_e}{\pi}V_{tb}V_{ts}^* \\ &\times \left(\mathcal{C}_7\frac{m_b}{e}\left[\bar{s}\sigma_{\mu\nu}P_R b\right]F^{\mu\nu} + \mathcal{C}_9\left[\bar{s}\gamma_\mu P_L b\right]\left[\bar{\ell}\gamma^\mu\ell\right] + \mathcal{C}_{10}\left[\bar{s}\gamma_\mu P_L b\right]\left[\bar{\ell}\gamma^\mu\gamma_5\ell\right]\right) + \text{h.c.} \end{aligned} \tag{3.12}$$

To go from $b \to s\ell\ell$ to $B \to K\ell\ell$, they note that there are also contributions from the current-current and quantum chromodynamic operators. They account for these contributions using effective Wilson coefficients

$$\mathcal{C}_7^{\text{eff}} = \mathcal{C}_7 - \frac{1}{3}\left[\mathcal{C}_3 + \frac{4}{3}\mathcal{C}_4 + 20\,\mathcal{C}_5 + \frac{80}{3}\mathcal{C}_6\right] + \frac{\alpha_s}{4\pi}\left[(\mathcal{C}_1 - 6\,\mathcal{C}_2)\,A(q^2) - \mathcal{C}_8 F_8^{(7)}(q^2)\right], \tag{3.13}$$

$$\begin{aligned} \mathcal{C}_9^{\text{eff}} =&\mathcal{C}_9 + h(0,q^2)\left[\frac{4}{3}\mathcal{C}_1 + \mathcal{C}_2 + \frac{11}{2}\mathcal{C}_3 - \frac{2}{3}\mathcal{C}_4 + 52\,\mathcal{C}_5 - \frac{32}{3}\mathcal{C}_6\right] \\ &- \frac{1}{2}\,h(m_b,q^2)\left[7\,\mathcal{C}_3 + \frac{4}{3}\mathcal{C}_4 + 76\,\mathcal{C}_5 + \frac{64}{3}\mathcal{C}_6\right] + \frac{4}{3}\left[\mathcal{C}_3 + \frac{16}{3}\mathcal{C}_5 + \frac{16}{9}\mathcal{C}_6\right] \\ &+ \frac{\alpha_6}{q^2}\left[\left(\frac{4}{9}\mathcal{C}_1 + \frac{1}{3}\mathcal{C}_4 + 76\,\mathcal{C}_6\right) + \frac{4}{3}\left[\mathcal{C}_3 + \frac{16}{3}\mathcal{C}_2\right) - \mathcal{C}_8\,F_8^{(0)}(q^2)\right] \\ &+ 8\,\frac{m_c^2}{q^2}\left[\left(\frac{4}{9}\mathcal{C}_1 + \frac{1}{3}\mathcal{C}_2\right)\left(1 + \hat{\lambda}_u\right) + 2\,\mathcal{C}_3 + 20\,\mathcal{C}_5\right], \end{aligned} \tag{3.14}$$

where q^2 is the invariant mass squared of the $\ell^+\ell^-$ (the magnitude of the 4-momentum squared recoiling off the kaon). After defining the amplitudes

$$\begin{aligned} F_A &= C_{10}\,, \qquad\qquad F_V = C_9^{\text{eff}} + \kappa\,\frac{2\,m_b\,m_B}{q^2}\,C_7^{\text{eff}}\,, \text{ and} \\ F_P &= -m_\ell\left[1 + \frac{m_B^2 - m_K^2}{q^2}\left(1 - \frac{f_0}{f_+}\right)\right] C_{10}\,, \end{aligned} \tag{3.15}$$

they calculate the decay amplitude (Eq. 1.20) with equation 3.12 and arrive at

$$\mathcal{A}(\bar{B} \to \bar{K}\ell^+\ell^-) = i\,\frac{G_{\text{F}}e^2}{\sqrt{32}\pi^2}\,V_{tb}V_{ts}^*\,f_+(q^2)\left[F_V\,p^\mu(\bar{\ell}\gamma_\mu\ell) + F_A\,p^\mu(\bar{\ell}\gamma_\mu\gamma_5\ell) + F_P\,(\bar{\ell}\gamma_5\ell)\right] \tag{3.16}$$

where p^μ is the 4-momentum of the B meson. Explicit form factor dependence has been added in terms of f_+ and f_0.

Over the range $14.18 < \mathrm{q}^2 < (\mathrm{m}_B - \mathrm{m}_K)^2$ GeV^2 their result for $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ is 1.26×10^{-7}, with the largest uncertainty coming from form factors ($^{+\,0.40}_{-\,0.21}\times10^{-7}$). A summary of central values and uncertainties are provided in Table 3.1.

Table 3.1: Summary of $\mathrm{B} \to \mathrm{K}\ell\ell$ predictions from Ref. [20] over $14.18 < \mathrm{q}^2 < (\mathrm{m}_B - \mathrm{m}_K)^2$ GeV^2.

Observable	Central Value	Form Factor	Short Distance	Sub Leading	CKM
$10^7 \times \mathcal{B}(\mathrm{B}^+ \to \mathrm{K}^+\ell^+\ell^-)$	1.04	$^{+\,0.60}_{-\,0.27}$	$^{+\,0.04}_{-\,0.02}$	$^{+\,0.03}_{-\,0.03}$	$^{+\,0.04}_{-\,0.07}$
$10^7 \times \mathcal{B}(B^+ \to K^+\tau^+\tau^-)$	1.126	$^{+\,0.40}_{-\,0.21}$	$^{+\,0.04}_{-\,0.30}$	$^{+\,0.02}_{-\,0.02}$	$^{+\,0.05}_{-\,0.09}$

3.3 Susceptibility to Theoretical Uncertainty

Modern particle physics analyses rely heavily on simulation, and this analysis relies on simulation of a never-before-seen decay. As simulation is based on theory, it is important to understand the theoretical uncertainty, both for standard-model predictions and beyond-standard-model predictions. In this section we briefly look at predictions for the branching fraction of $B^+ \to K^+\tau^+\tau^-$ as a function q^2 , first for the standard model calculation and second for a beyond-standard-model scenario, to gain a qualitative understanding of how these uncertainties could impact our measurement.

3.3.1 Form Factor Uncertainty

Figure 3.1 shows the theoretical differential branching fraction as a function of 4-momentum transfer to the dilepton system ($d\mathcal{B}/d\mathrm{q}^2$) for the decay $B^+ \to K^+\tau^+\tau^-$. The shape difference between the $\pm 1\,\sigma$ bands is considerable. Any search for this decay will necessarily have shape-related systematic uncertainty due to this. For completeness, we've included their Fig. 3.2 showing $d\mathcal{B}/d\mathrm{q}^2$ for $\mathrm{B}^+ \to \mathrm{K}^+\mu^+\mu^-$ to elucidate that, one, $B^+ \to K^+\tau^+\tau^-$ is kinematically allowed only for high q^2 (soft kaon recoil) and two, that the ditauon mode is favored slightly over the lighter lepton modes in this region.

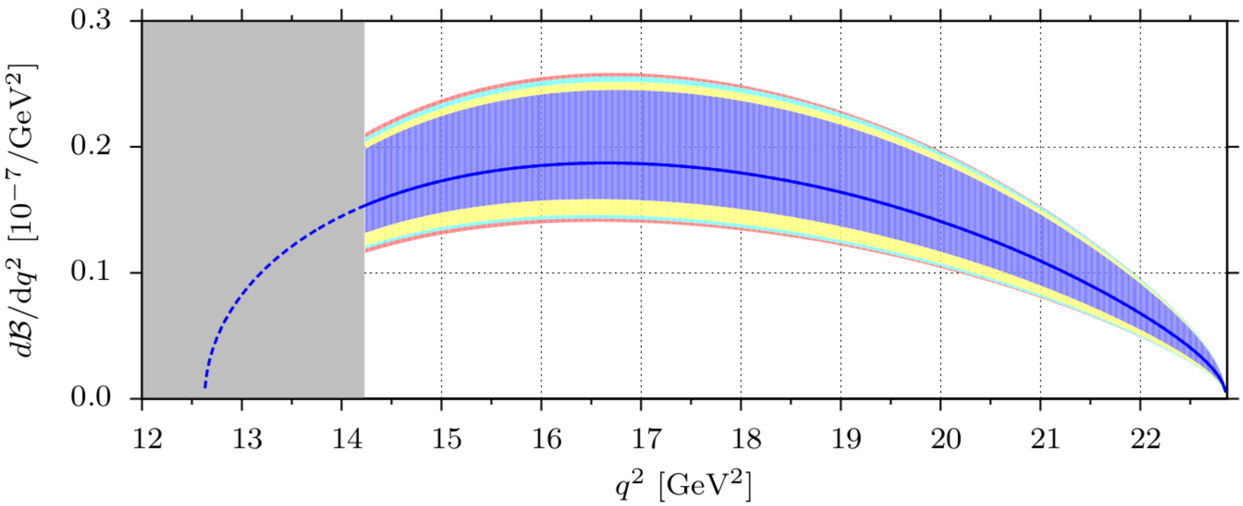

Figure 3.1: Theoretical $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ as a function of q^2 from Ref. [20]. The bands show uncertainties from form factors (blue), CKM matrix elements (yellow), short-distance effects (cyan), and sub-leading corrections (red). The gray band shows the q^2 cut used by BaBar which was placed to avoid the $\psi(2\mathrm{S})$ resonance.

Further, compared to other form-factor extrapolations commonly used in this type of calculation, form-factor estimations vary by up to roughly 25 % over the whole q^2 range, and at high q^2 (soft kaon recoil) they differ by as much as 75 %. If nature has a different model in mind than the one we used to generate our Monte Carlo data, it would cause systematic uncertainties due to modeling.

3.3.2 New Physics

Some new physics models have directly calculated their effects on $B^+ \to K^+\tau^+\tau^-$.

S. Choudhury et al. looked for supersymmetric effects on $B^+ \to K^+\tau^+\tau^-$ in Ref. [21]. They present a different shape for $d\mathcal{B}/d\mathrm{q}^2$ when using the minimal supergravity (mSUGRA) model, also known as constrained minimal supersymmetric model (CMSSM), or when using the relaxed supergravity (rSUGRA) model (Fig. 3.3). They also find a strong dependence on $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ from their mSUGRA model parameters m_0 and $\tan(\beta)$.

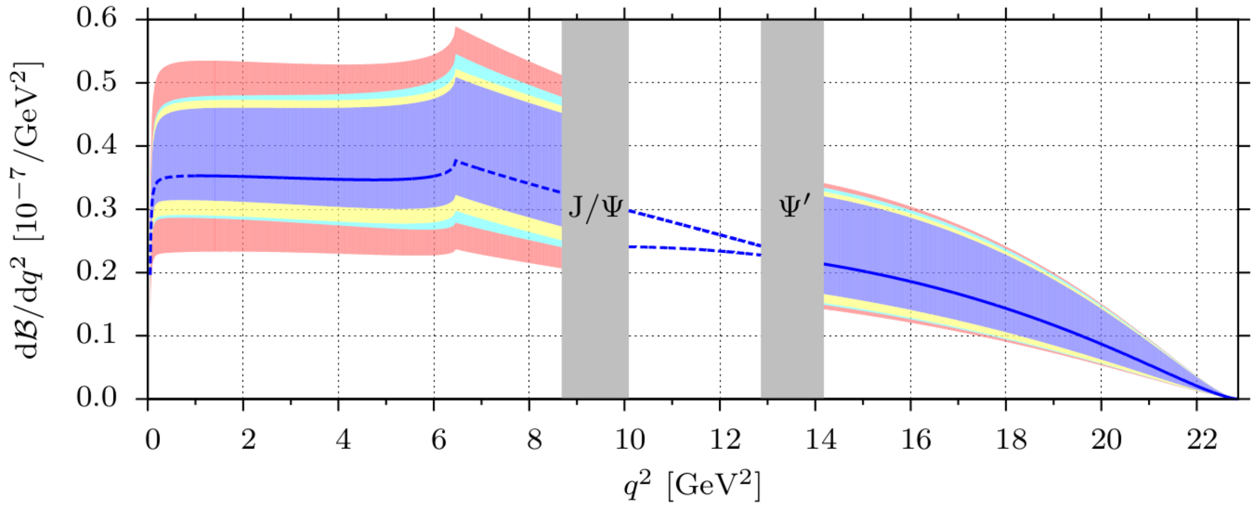

Figure 3.2: Theoretical $\mathcal{B}(\mathrm{B}^+ \to \mathrm{K}^+\mu^+\mu^-)$ as a function of q^2 from Ref. [20]. The blue bands show total theoretical uncertainty.

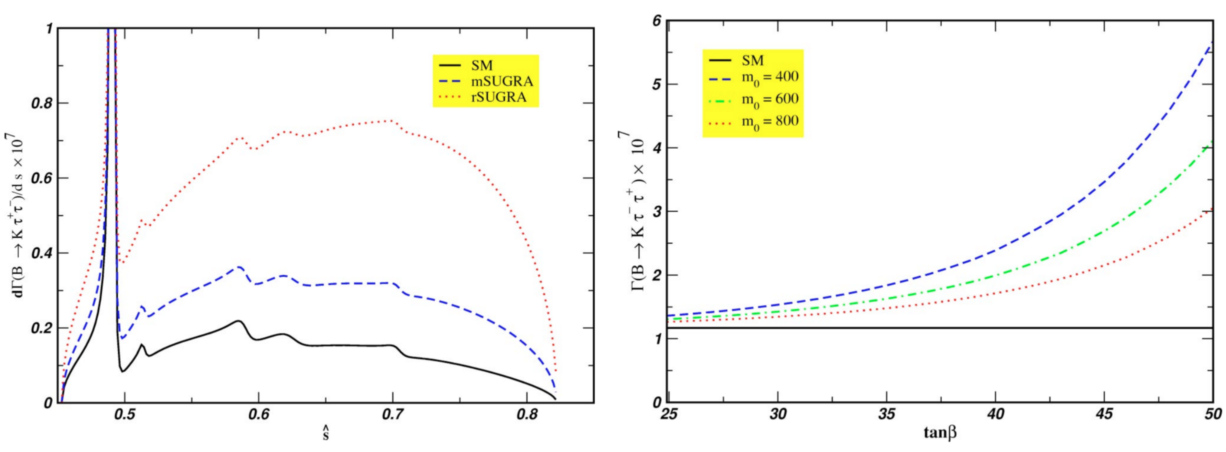

Figure 3.3: $d\mathcal{B}/d\mathrm{q}^2$ and $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ predictions from Ref. [21]. Note $\hat{s} = \mathrm{q}^2/\mathrm{m}_B^2$, so the range is comparable to 3.1.

CHAPTER 4
BELLE EXPERIMENT

The Belle experiment was designed to make copious amounts of B mesons and accurately measure charge-parity (CP) violation. The KEKB [22] asymmetric e^+e^- accelerator provides beams of 8 GeV electrons and 3.5 GeV positrons. At the interaction point (IP) is a 3-story tall multipurpose particle spectrometer, the Belle [23] detector. It is designed to detect particles with energies between about 50 and 7000 MeV. The experiment ran from 1999–2010 and recorded 980 fb^{-1} of data, primarily at the $\Upsilon(4S)$ resonance [24]. Belle's measurement of CP violation was recognized in the 2008 Physics Nobel Prize.

4.1 The KEKB Accelerator

Construction of the KEKB accelerator began in 1994, using the tunnels from the former TRISTAN collider, and was completed by the end of 1998. It is a two-ring e^+e^- asymmetric-energy collider with an 8 GeV electron beam, the high-energy ring (HER), and a 3.5 GeV positron beam, the low-energy ring (LER). The ring energies are chosen to achieve a center-of-mass energy near the $\Upsilon(4S)$ resonance. The rings are 3.016 km in length. The linear accelerator used for TRISTAN was upgraded directly to 8 GeV with new accelerating sections, eliminating the need for the TRISTAN accumulation ring. The design luminosity for KEKB is $10 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which is made possible by high design beam current (1.1 A for the HER and 2.6 A for the LER) and small beam profile (90 µm $\times$ 1.9 µm) at the interaction point. The KEKB accelerator is equipped with new kinds of accelerating radio-frequency (RF) cavities that strategically remove higher-order resonance modes from the cavities to reduce coupled-bunch instabilities. The crossing angle of the beams is 22 mrad which eliminates the need for dipole steering magnets in the interaction region (IR) and improves bunch recovery after collision. The Belle detector is aligned with the direction of the LER as the positrons would suffer more from bending in the detector's 1.5 T solenoidal magnetic field than the electrons in the HER would suffer. To squeeze the beams at the IP, a pair of 4.8 T superconducting final focus magnets are installed.

The largest beam instability for KEKB is due to electron clouds in the LER beam pipe. These clouds form when synchrotron radiation hits the beam pipe wall and frees electrons via the photoelectric effect. The positively-charged positron beam bunches attract these electrons toward the center of the beam pipe. To remedy this effect, a total of 800 m of the LER beam pipe was wound with copper coils to create a solenoidal magnetic field aligned with the direction of the beam. Emmitted photoelectrons curve in this field and remain close to the wall of the beam pipe. Specific luminosity nearly doubled after the installation of these solenoids. [22, 23].

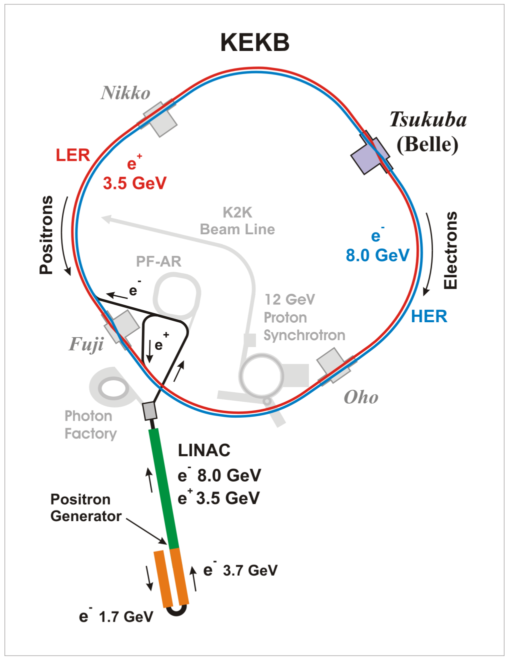

Figure 4.1: Schematic of the KEKB accelerator.

4.2 Belle Detector

At the interaction point of the KEKB accelerator, the two beam pipes are joined with an X-shaped double-walled beryllium structure with an inner (outer) radius of 2.0 cm (2.35 cm) and 0.5 mm wall thickness. The 2.5 mm gap between is flowed with helium gas for cooling. Just outside of this structure begins the Belle detector subsystems. [23, 24, 25]

4.2.1 Silicon Vertex Detector

The innermost detector subsystem of Belle is the Silicon Vertex Detector (SVD). It measures asymmetries in the proper time distribution of B-meson decays. The SVD uses double-sided silicon strip detectors (DSSDs) for particle detection. It comprises three layers, where the first, second, and third layers use two, three, and four DSSDs respectively. The strip pitch of each DSSD is 25 µm for the p-side and 42 µm for the n-side. Altogether, the SVD has 81,920 readout channels. The three SVD layers are arranged in a barrel shape around the beam pipe at radii of 3 cm, 4.5 cm, and 6 cm. Performance studies show the SVD can resolve the impact parameter down to ∼25 µm in the $r - \phi$ direction and ∼40 µm in the z direction for most particle momenta [25].

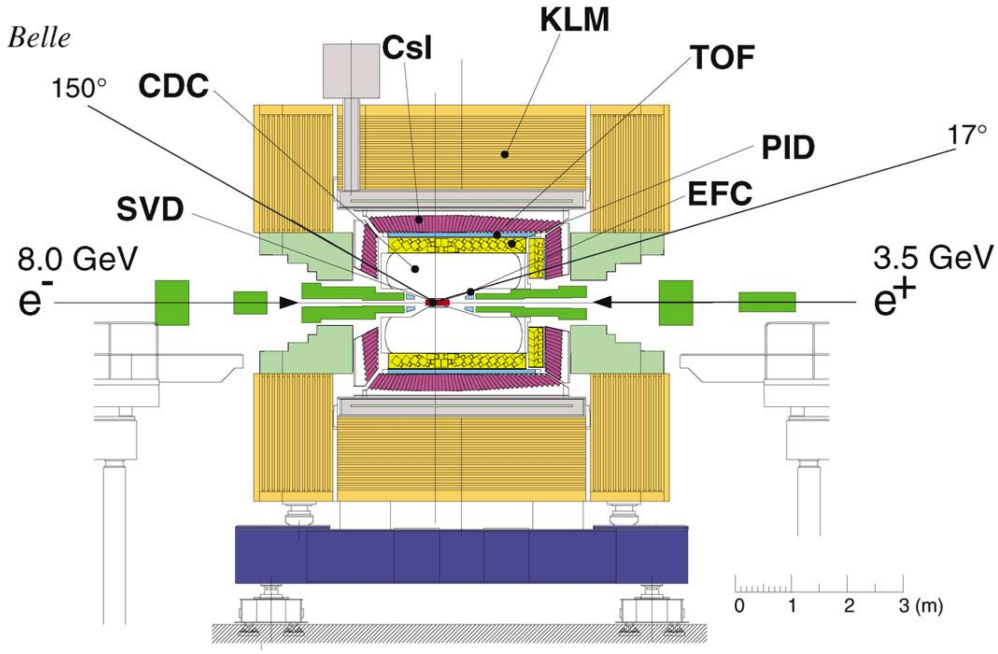

Figure 4.2: Schematic of the Belle detector from Ref. [23].

4.2.2 Central Drift Chamber

To provide efficient reconstruction of charged-particle tracks and precise measurement of particle momentum, a gaseous tracking chamber called the Central Drift Chamber (CDC) occupies the space outside of the SVD. To minimize multiple Coulomb scattering, a low-Z gas mixture of 50 % helium and 50 % ethane is used. The CDC comprises 50 layers, each layer being a group of field wires and sense wires that are strung in an axial or stereo (slightly off-axis) direction. The longest wire is 2.4 m The radius of the innermost (outermost) layer is 10.35 cm (87.4 cm). In total, there are 8400 drift cells — a nearly square region between adjacent field wires and surrounding a single sense wire. When a charged particle traverses the CDC, it leaves a trail of gas ions that drift through the strong electric field to the sense wires. Drift time and wire position are used to reconstruct the track accurately. Momentum and dE/dx are used for particle species identification (PID) [23].

Additionally, the first, third, and fourth layers are composed of cathode strips rather than field wires. The cathode strips form segmented rings with eight segments in the ϕ direction and an 8.2 mm pitch in the z-direction. These strips provide z information near the IP for charged tracks, the purpose of which is to reject background from cosmic rays, beam gas events, and interactions between spent electrons and the beam pipe.

4.2.3 Aerogel Cherenkov Counter

Using dE/dx alone does not provide good PID at all particle momenta. Of particular interest to the goals of the Belle experiment is the ability to distinguish charged pions ($\pi^{\pm}$) from charged kaons ($K^{\pm}$). Cherenkov radiation happens when a charged particle exceeds the speed of light in a material, $v_0 = c/n$, where n is the index of refraction. If $\beta_n = v/v_0$, then the intensity of Cherenkov radiation is proportional to $1 - 1/\beta_n^2$ [26, 27]. For particle velocities just slightly larger than v_0 the intensity increases rapidly with increased velocity. As the charged kaon mass is approximately 3.5 times larger than the charged pion mass, this effect can be beneficial for kaon–pion separation if a material with an index of refraction suitable to the momentum range of interest is used.

For this reason, Belle uses an aerogel Cherenkov counter (ACC) in the barrel region just outside the CDC and in the forward end-cap direction. The index of refraction of the aerogel used ranges from 1.01 to 1.03 for different polar angles in the detector. Each ACC module detects Cherenkov photons using one or two fine-mesh photomultiplier tubes. There are 960 ACC modules in the barrel region and 228 in the forward endcap region. Measurements of charm decays yield kaon efficiencies above 80 % and pion fake rates below 10 % in the momentum range of 0.5 GeV/c to 4.0 GeV/c [23].

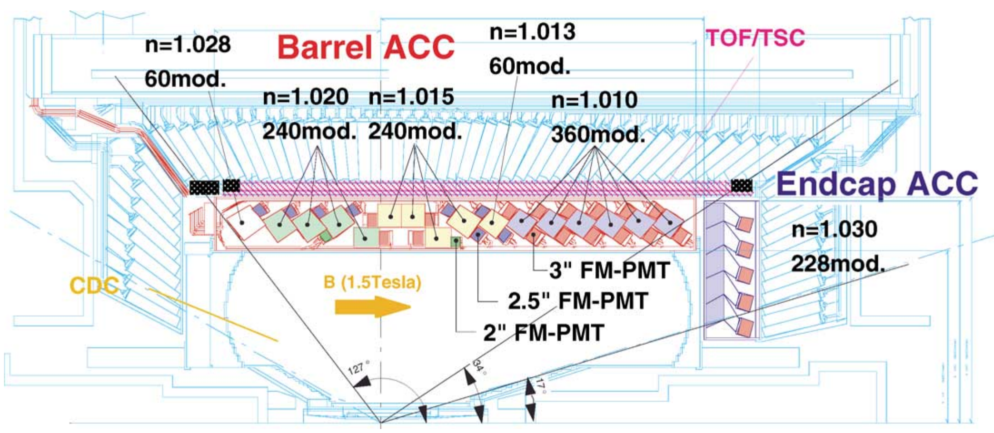

Figure 4.3: Schematic of the ACC from Ref. [23].

4.2.4 Time of Flight Counter

The ACC cannot measure kaons with momenta less than 0.5 GeV/c, and electron efficiency from the ACC is less than 90 % for tracks less than ∼1.0 GeV/c. In this lower-momentum regime, the time-of-flight (TOF) counters use precise timing (100 ps) for additional particle identification. If velocity and momentum are known, particle mass can be computed directly. The TOF system comprises 128 scintillating bars with a trapezoidal cross-section and a length of 2.55 m. The bars are arranged as a cylinder around the inner detectors with a radius of ∼1.2 m. To achieve a timing resolution of 100 ps, the TOF counters are instrumented with fine-mesh photomultiplier tubes on either end. As the TOF system also provides a fast trigger signal to the trigger system (described below), it must be resistant to the beam background. For this purpose, one 0.5 cm thick trigger scintillation counter (TSC) is situated radially inward of every two adjacent TOF counters and separated from the TOF counters with a 1.5 cm air gap. Only those hits that are coincident between a TOF and a TSC are propagated to the trigger system.

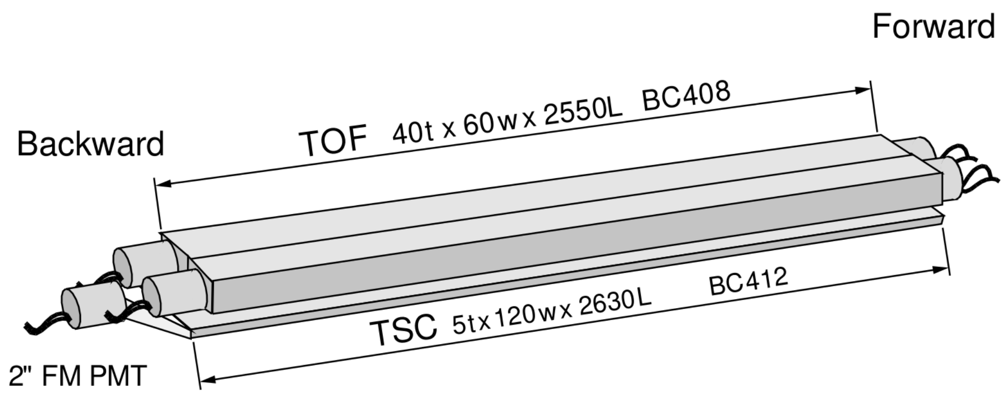

Figure 4.4: TOF/TSC module from Ref. [28].

Studies of muon pairs show that the weighted average of hit times from photomultiplier tubes on either end of a TOF counter achieves a timing resolution between 70 ps to 100 ps. The measured separation power between pions and kaons ranges from $6\,\sigma$ to $2\,\sigma$ in the momentum range of 0.6 GeV/c to 1.4 GeV/c.

4.2.5 Electromagnetic Calorimeter

The detectors described thus far provided track reconstruction and particle identification of charged tracks using several techniques. The purpose of the electromagnetic calorimeter (ECL) is to measure the energy deposition of photons (which are charge neutral and thus unseen by the inner detectors) and of charged tracks. A good resolution on energy and position is needed for the reconstruction of $\pi^0 \to \gamma\gamma$ decays among others. This is achieved using cesium iodide crystals doped with thallium, CsI(Tl). Charged tracks and photons entering a crystal undergo electromagnetic showering creating many lower-energy photons that are sensed by a pair of silicon PIN photodiodes (Hamamatsu S2744-08).

The ECL consists of 6624 crystals in the barrel region, with an inner radius of 1.25 m and length of 3 m, and an additional 2112 crystals in the forward and backward endcap regions. Crystal dimensions vary, but a typical crystal is 30 cm in length with a front face of 5.5 cm $\times$ 5.5 cm and rear face of 6.5 cm $\times$ 6.5 cm. The width is chosen to contain 80 % of the electromagnetic shower for a photon incident on the center of the crystal while maintaining good positional resolution, and

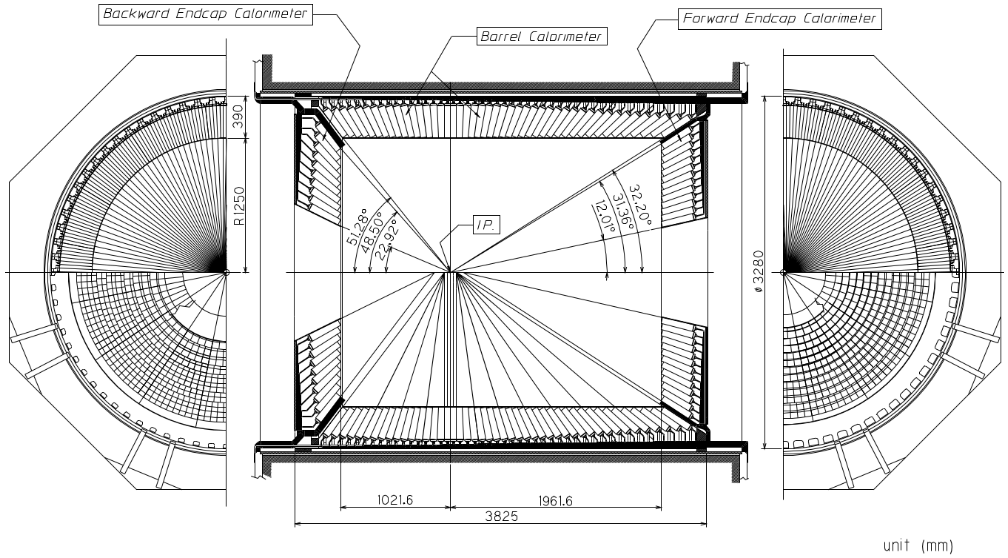

Figure 4.5: CAD drawing of the ECL [23].

the length is chosen to provide 16.2 radiation lengths of stopping power.

The measured energy resolution of the ECL is roughly 2.85 % in the backward endcap calorimeter and 1.7 % to 1.74 % in the barrel and forward endcap calorimeters. The spatial resolution determined from photon beams at the ROKK-1M facility is between 0.4 cm and 1 cm depending on the transverse position of the beam relative to the center of the crystal. The measured mass resolution for π^0 mesons is 4.9 MeV/c^2.

4.2.6 Superconducting Solenoid

As a general-purpose particle spectrometer, the Belle detector requires a strong magnetic field to curve tracks of charged particles allowing for measurement of charge sign and rigidity, p/q, or momentum p per charge q. The solenoid uses niobium-titanium superconductors clad with copper, NbTi/Cu, to generate a magnetic field of 1.5 T parallel to the direction of the positron beam. The inner radius of the superconducting solenoid is 1.7 m and its length is 4.4 m. It is cooled using liquid helium and is energized with 4.4 kA of current. Outside of the solenoid, the structure of the Belle detector takes the shape of 14 concentric octagonal shells of 4.7 cm thick iron plates which serve as the flux-return path for the magnetic field. Similarly, the endcap regions are composed of 14 parallel octagonal plates of 4.7 cm thick iron. The gaps between these iron plates are instrumented with particle detectors which will be the subject of the next section.

4.2.7 K-Long and Muon Detector

Long-lived neutral particles and charged particles that lose minimal energy to ionization, i.e., long-lived neutral kaons (K_L^0) and minimum ionizing charged particles — muons ($\mu^{\pm}$), may transit the ECL and solenoid before decaying. For this reason, the gaps within the iron yoke, that is to say, the flux return for the magnetic field, are instrumented with resistive-plate counters (RPCs). This detector subsystem is called the K-long and muon (KLM) detector.

An RPC is a detector consisting of two plate electrodes, at least one with high bulk resistivity, with a thin gas gap between the plates. High voltage is applied to the plates, and the resulting strong electric field between them allows traversing particles to ionize the gas and promptly create a localized discharge in the electrodes. This discharge induces a voltage pulse on readout strips that are not connected to the plate electrodes [29].

Each RPC module in the KLM detector is a sandwich of two single RPC layers and readout strips above and below the pair of RPC layers. This two-RPC-layer design results in larger than 98 % efficiencies and further provides redundancy if a problem arises with one of the RPC layers. The readout strips above are orthogonal to those below to provide two-dimensional hit information. The barrel modules vary in size, depending on location within the detector, and have (ϕ, θ) readout strips that are 5 cm wide and vary in length from 1.5 m to 2.7 m. The endcap modules are quarter-washer shaped (Fig. 4.6) and have (ϕ, θ) readout strips that vary in length from 1.83 m to 5 m. Here, the θ strips are 3.6 cm wide and the ϕ strip widths are tapered from 1.9 cm to 4.7 cm.

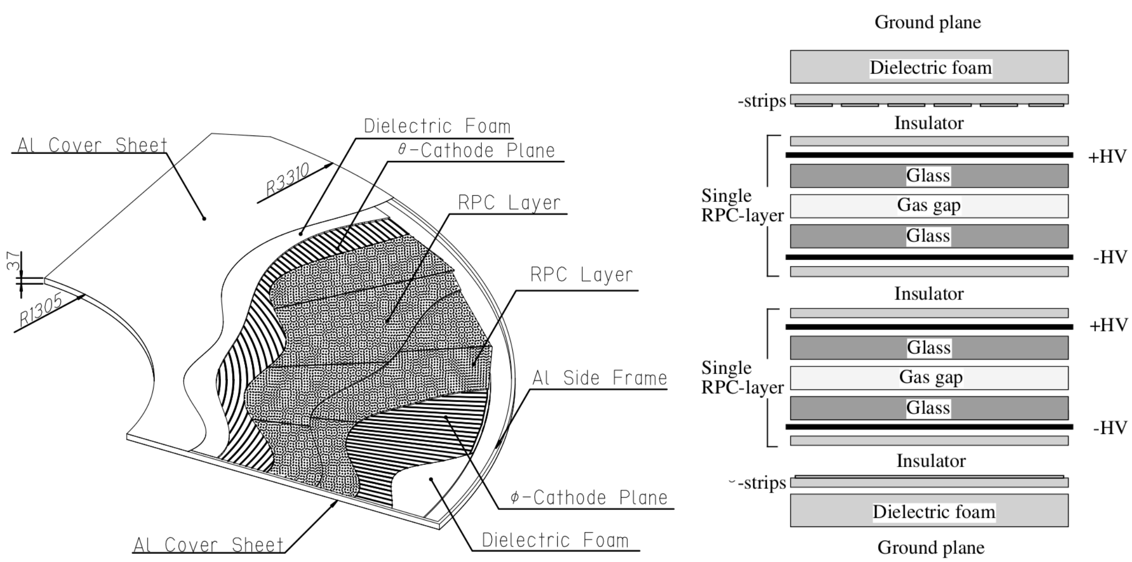

Figure 4.6: Left: Cross-sectional drawing of an endcap RPC module. Right: Layer stackup of an RPC module. Source [23].

A K_L^0 meson originating near the IP transits the detector unseen until it is involved in a hadronic interaction with the material of the detector. The resulting hadronic shower creates a cluster of

2D hits that is either contained in the ECL or contained within a few layers of the KLM detector depending on where the hadronic interaction happened. The ECL provides 0.8 interaction lengths of material and the KLM detector provides an additional 3.9 interaction lengths. Muons on the other hand, being minimum-ionizing particles, leave tracks in the SVD and CDC, small energy depositions in the ECL, and tracks of 2D hits through the layers of the KLM. Muons require momenta of at least 600 MeV/c to reach the KLM detector.

The measured positional resolution of hits in the KLM detector has a standard deviation of 1.2 cm and an angular resolution of less than 10 mrad. Measured muon detection efficiency is roughly 90 % for momenta above 1.5 GeV and ranges from 40 % to 90 % at lower momenta [23]. The RPCs do have some drawbacks, however. They require a steady flow of expensive gas to operate, including refrigerants (R134a). They are susceptible to damage by water contamination in the gas and to overpressurization, the latter causing irreversible damage. Perhaps the largest drawback is the dead time. When a streamer forms and discharges a localized region of the resistive plate, that region is dead until that area of the plate charges back up. At very high luminosity and or high beam background rates, this dead time becomes significant in some parts of the detector. At background rates 20 to 40 times higher than the design for the Belle experiment, the RPC efficiency may degrade to 50 % [30].

4.3 Trigger System

The cross-section for $e^+e^- \to \Upsilon(4S)$ is 1.2 nb, whereas for processes like Bhabha scattering ($e^+e^- \to e^+e^-$) the cross section is 44 nb. At the design luminosity of $10^{34}\,\mathrm{cm^{-2}s^{-1}}$ this amounts to 12 Hz and 4400 Hz for $\Upsilon(4S)$ meson production and Bhabha scattering respectively. With an average event size of 30 kB, Bhabha events alone would account for 11.4 TB of data recorded daily.

To decrease the required storage space a triggering system is used. The global decision logic (GDL) collects primitive trigger information from the detector subsystems within 1.85 µs of an event and decides to provide a trigger to the data acquisition system in the next 350 ns. This is the level-1 trigger. The inputs to the GDL include r–ϕ and r–z track triggers from the CDC, energy-sum and cluster-count triggers from the ECL, timing and topology triggers from the TOF subsystem, and muon triggers from the KLM subsystem. For Bhabha and $e^+e^- \to \gamma\gamma$ events, the level-1 trigger is reduced by a factor of 100 to avoid unnecessary data storage while still using these processes to measure luminosity.

The KEKB accelerator operates at a frequency of 509 MHz with bunch crossings happening every 2 ns in the full bunch-fill pattern. A copy of the 509 MHz accelerator clock is divided and distributed to the Belle trigger system and most of the detector-subsystem trigger systems. By using an accelerator clock divided by eight, the GDL can maintain a fixed level-1 trigger latency of 2200(8) ns with respect to the bunch crossing in which the e^+e^- collision happened. This is known as the fast-trigger-and-gate scheme [23].

4.4 Data Acquisition System

Most of the detector subsystems are read out using a charge-to-time (Q-to-T) converter and a time-to-digital converter (TDC). The Q-to-T technique stores the charge in a capacitor that is discharged at a constant rate. By measuring discharge-start time and discharge-stop time with respect to a common gate time, both amplitude and hit time are measured at once. The KLM subsystem only uses a TDC. The SVD uses its own data acquisition and sparsification system which relies on a special level-0 trigger provided by the TOF system with a fixed latency of $0.85\,\mu s$. The designed maximum trigger rate of Belle is 500 Hz with a maximum dead time of 10 %.

After a level-1 trigger is issued, event data from the SVD and from the TDCs of the remaining detector subsystems are merged into the event builder. The combined event data then goes to an online computer farm where it is formatted for offline storage and a level-3 trigger (software trigger) is applied. The software trigger further reduces the background data. Finally, the event data is transferred to a tape library 2 km away for permanent storage [23].

CHAPTER 5
ANALYSIS STRATEGIES

After describing the detector and how the data is acquired, we must now consider how we want to analyze the data. First of all, most modern particle-physics analyses are blind analyses. Blind means that, to avoid biasing ourselves, we do not look at the data until our analysis technique is fully developed. To achieve this, we build our analysis technique by looking at simulated data. The point at which one examines the actual data is called the unblinding or unboxing. There is also partial unblinding, which can mean many different things, but generally has the aim of quantifying the level of agreement between simulated data and actual data. One can look at off-resonance data to check the modeling of continuum background. Portions of on-resonance data in which one does not expect to see the decay mode they are studying can be unblinded. Partial unblinding can also refer to control modes — modes that resemble to the decay mode being studied and may serve to check that a particular analysis technique is producing the correct results or to check the detection efficiency of a certain particle species. Fully unblinding is the final step and usually requires permission from collaborators. The analysis should not be modified after the unblinding, but additional work may follow to better understand the results.

Aside from the fact that this is a blind analysis, we also want to discuss how we will identify our decay mode during event reconstruction and how we will extract our result.

5.0.1 Tagging Methods

As previously mentioned, $\Upsilon(4S)$ decays to $B\overline{B}$ more than 96 % of the time. Of these, 51.6 ± 0.6 % are B^+B^- and 48.4 ± 0.6 % are $B^0\,\overline{B}^0$. If reconstructing a particular B decay, whether a charged B or a neutral B, there is always a second B decay in the event. A tagged analysis means that both B decays are reconstructed to some extent with one being labeled as the tag-B meson and the other the signal-B meson. An untagged analysis means that one only attempts to reconstruct their signal-B meson and does nothing with the other B meson. The most common tagging methods are hadronic tagging and semileptonic tagging. Another approach, that will be used in this study, is inclusive tagging.

Hadronic Tagging

Hadronic tagging is the process of singling out events in which one of the B mesons decays fully hadronically, that is to say, there are no leptons in any of the decay products, including secondary decays. When a fully hadronic B decay is identified, it is labeled as the tag-B meson, and the remaining tracks and clusters in the event are used to search for the signal-B meson. While hadronic B decays are fairly common, fully hadronic decays, where the daughters and granddaughters and so

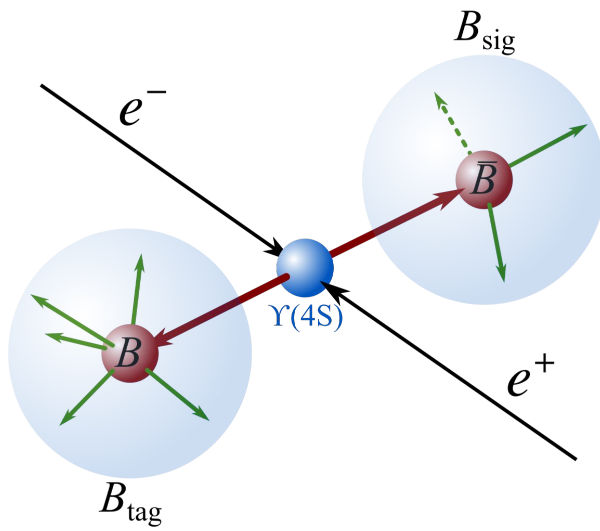

Figure 5.1: Schematic of B meson generation in e^+e^- collisions. In the center-of-mass frame, $\Upsilon(4S)$ is produced at rest and decays into two B mesons ($B^0\,\overline{B}^0$ or B^+B^-) moving in opposite directions. In this example, one B decays hadronically and is labeled B_{tag}. Precise knowledge of the collision spot and of the tag-B meson's direction can be used to determine the direction of the other B meson thus allowing for the determination of missing energy as shown by the dashed line in the B_{sig} decay.

on all decay hadronically, are much less common. In Belle, the Full Reconstruction (FR) algorithm was developed [31]. Using 1104 exclusive decay channels, the FR algorithm correctly reconstructs a $B^{\pm}$ (B^0) in only 0.28 % (0.18 %) of $B\overline{B}$ events.

Although the efficiency of hadronic tagging is very low, the precision is very high. With one B meson fully reconstructed without any missing energy, the direction of the other B meson can be determined precisely (the magnitude of the two B momenta are already known from the e^+e^- center-of-mass energy and the B rest mass, but the directions are not known). This approach has advantages when looking for B decays with missing energy or when making measurements of a particular decay topology.

Semileptonic Tagging

Like hadronic tagging, semileptonic tagging singles out events in which one B meson decays to an exclusive semileptonic decay mode. This B meson is labeled as the tag-B meson and the remaining tracks and clusters in the event are used to search for the signal-B meson. Using the same variables and hierarchical approach as the hadronic FR algorithm, a semileptonic FR algorithm was developed for use at Belle [32]. The algorithm only uses 60 exclusive decay channels, each with a single electron or muon, but manages to correctly reconstruct a $B^{\pm}$ (B^0) in 0.31 % (0.34 %) of $B\overline{B}$ events.

While the efficiency for semileptonic tagging is slightly higher than that of hadronic tagging, the precision is less. With only one lepton in the tag-B decay, and since the magnitude of the B momentum is known, the directions of the B and $\overline{B}$ can be constrained to a conical surface. Still they cannot be completely determined as in the case of hadronic tagging.

Inclusive Tagging

Quite the opposite to the previously described FR algorithms, inclusive tagging involves first reconstructing the B decay of interest and labeling it as the signal B. The remaining tracks and clusters in the event are then classified as the tag-B meson. As no exclusive B decays are used to form the tag this method is called inclusive tagging. To increase the tagging purity, selections are applied to the tag-B meson. The selections can differ from one analysis to the next but may include requiring that the summed momenta and energy are consistent with the B mass and its momentum. Other selections can be applied to the tracks and clusters that form the inclusive tag-B meson to reduce background sources not associated with the $\Upsilon(4S)$ decay. A tree fit can be used to constrain the decay vertex of the tag-B meson and help eliminate false candidates. An inclusive-tagging analysis differs from an untagged analysis in that rather than ignoring the unused tracks and clusters, they are used to place additional constraints on the overall $\Upsilon(4S)$ decay.

The benefit of inclusive tagging is high efficiency. Depending on the applied selections, one can search for their signal-B meson in up to 100 % of $B\overline{B}$ events as opposed to a fraction of a percent seen in the FR algorithms. This can be advantageous when searching for very rare decays. The downside to inclusive tagging is the large backgrounds.

In our analysis, we use a modified version of inclusive tagging which we refer to as inclusive hadronic tagging. The idea is that since we have a known amount of electrons and muons from our tau decays (depending on each specific tau decay mode), we can require that the entire event has exactly that many electrons and/or muons. The tag-B meson decay, therefore, would most likely be fully hadronic. We choose this approach because our decay, $B^+ \to K^+\tau^+\tau^-$, is expected to be very rare, and missing energy is a distinguishing feature. The strategy is one, to require that the summed momenta and energy are consistent with the B mass and its momentum, and two, to require that the signal-B meson, aside from having the requisite decay products, also have considerable missing energy.

5.0.2 Result Extraction Strategies

There are a handful of strategies for measuring a branching fraction for a particular decay mode.

The cut-and-count approach is the simplest. In this approach, the number of signal events and the number of background events that pass the selection criteria of the analysis are counted. The ratio of signal-to-background and the knowledge of the total number of events are used to calculate the branching fraction.

However, suppose for a moment that the decay mode being studied, after looking at many thousand simulated events, has a mass distribution with a shape that is distinctly different from all of the simulated background events. The use of the cut-and-count approach here would throw away that shape information. In this case, it would be beneficial to extract the branching fraction from a fit to the mass distribution rather than just counting events.

Sometimes the fit variable (mass in the former example) can be modeled well with analytic functions like a Gaussian distribution for the signal peak and polynomials for the background. In such a case one can use a least-squares fitter. Often the shape cannot be well described with analytic functions. In that case, the data can be fit to a combination of model histograms using a binned maximum-likelihood estimate (MLE).

There is no reason to limit the fit to a single variable. It might be the case that shape differences between signal and background are more pronounced in two-dimensional histograms. Here, one might pursue a new variable that is a function of two other variables. Alternatively, a binned MLE can be performed using two-dimensional model histograms.

As mentioned above, missing energy is a distinguishing feature of our decay. We will use machine-learning (ML) techniques to reduce background events further, and the output classifier of our ML algorithm will also have distinctly different shapes for signal and background respectively. Therefore, we will extract our branching fraction by performing a binned MLE using two-dimensional histograms of the ML output classifier versus missing-mass squared of the event.

CHAPTER 6
RECONSTRUCTION

We now describe the method used to search for $B^+ \to K^+\tau^+\tau^-$ in Belle data. This chapter will begin with a description of the $\tau^\pm$ decay modes that are considered. Next, we cover the Monte Carlo data used, including the generation of our own signal-enriched Monte Carlo data. We continue with the selection of our final-state particles, combining those particles to form signal candidates, best candidate selection, and selection criteria for our inclusive tag-B meson.

6.1 Decay Modes

The $\tau^\pm$, the heaviest of the leptons at 1776.93 MeV, has a short lifetime of 290.3×10^{-15} s [9]. Even for the most energetic taus in Belle from $e^+e^- \to \tau^+\tau^-$ events, such a tau would have a Lorentz factor $\gamma \approx 3$ and would have a mean decay length of 0.24 mm. As this is much less than the diameter of the beam pipe at the interaction point, it is safe to assume we do not expect to see $\tau^\pm$ leptons directly in the tracking detectors of Belle. Rather, we reconstruct some of the $\tau^\pm$ decays and anticipate missing energy due to neutrinos.

Of the known τ decay modes, $\tau^- \to \pi^-\pi^0\nu_\tau$ has the largest branching fraction at 25.49 %. Note that we will omit the charge-conjugated modes for ease of reading, but you may regard them as equal. The next three largest branching fractions are $\tau^- \to e^-\overline{\nu}_e\nu_\tau$, $\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$, and $\tau^- \to \pi^-\nu_\tau$ at 17.82 %, 17.39 %, and 10.82 % respectively. These are examples of "1-prong" τ decays as they all have one charged particle in the final state. To round off our list of the largest branching fractions we have the 1-prong decay $\tau^- \to \pi^-\pi^0\pi^0\nu_\tau$ and the 3-prong decay $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ coming in at 9.26 % and 9.31 % respectively.

Despite the best efforts to improve pion identification, $\pi - K$ separation and $\pi - \mu$ separation are among the least-performing in efficiency and fake rate. To eliminate backgrounds due to fake pions, this analysis only considers 1-prong τ decays to leptons. This gives us the four decay modes:

- $B^+ \to K^+\ \tau^+[e^+\overline{\nu}_\tau\nu_e]\ \tau^-[e^-\nu_\tau\overline{\nu}_e]$
- $B^+ \to K^+\ \tau^+[\mu^+\overline{\nu}_\tau\nu_\mu]\ \tau^-[\mu^-\nu_\tau\overline{\nu}_\mu]$
- $B^+ \to K^+\ \tau^+[e^+\overline{\nu}_\tau\nu_e]\ \tau^-[\mu^-\nu_\tau\overline{\nu}_\mu]$
- $B^+ \to K^+\ \tau^+[\mu^+\overline{\nu}_\tau\nu_\mu]\ \tau^-[e^-\nu_\tau\overline{\nu}_e]$

Throughout this document, we will stick to the convention that the charged kaon is the first daughter of the B meson, and the second daughter is the lepton with the same charge as the B and K mesons. Our search includes the charge conjugated $B^- \to K^-\tau^-\tau^+$ decay, but for brevity we will frequently only refer to $B^+ \to K^+\tau^+\tau^-$ where charge conjugation is assumed, just as in the title of this paper.

We will use the following short-hand notation to refer to these modes

- the "ee" mode $\Leftrightarrow K[{}^\tau e][{}^\tau e]$
- the "$\mu\mu$" mode $\Leftrightarrow K[{}^\tau \mu][{}^\tau \mu]$
- the "$e\mu$" mode $\Leftrightarrow K[{}^\tau e][{}^\tau \mu]$
- the "μe" mode $\Leftrightarrow K[{}^\tau \mu][{}^\tau e]$

where the τ superscript preceding the lighter lepton is there to remind us that these leptons came specifically from τ decays and hence missing neutrinos are involved. The charge is implied by the ordering $(+,+,-)$ for B^+ decays and $(-,-,+)$ for B^- decays.

Combined, these four decay modes account for 12.40 % of the possible $B^+ \to K^+\tau^+\tau^-$ final states.

6.2 Monte Carlo Data

The characterization of our B decay, of backgrounds from other B decays, of backgrounds stemming from e^+e^- annihilating to create other resonances lighter than the $\Upsilon(4S)$, and of contamination from beam backgrounds are all done using simulated data. We refer to this as Monte Carlo (MC) data.

We generated 10^6 events each for our four decay modes using the `EvtGen` [33] and `PYTHIA` [34] packages. Final-state radiation is simulated using `PHOTOS` [35]. The decay description begins with $\Upsilon(4S) \to B_{\mathrm{sig}}\overline{B}_{\mathrm{gen}}$, in other words, one B meson decays to our signal mode and the other decays generically according to the branching fractions specified in the Belle master decay file. In each case, 50 % are $B^+ \to K^+\tau^+\tau^-$ and 50 % $B^- \to K^-\tau^-\tau^+$.f We will frequently refer to this as signal Monte Carlo or signal MC.

Detector response to our decays is simulated using `GEANT3` [36]. The output files are in the same format as recorded Belle data which allows simulated data to be reconstructed as if it were real data.

6.3 Particle Selection

The reconstruction of either Monte Carlo data or real data is carried out using the Belle II Analysis Software Framework (`basf2`) [37, 38]. The `b2bii` [39] module (pronounced B2B2) allows users to reconstruct Belle data using the `basf2`.

To elucidate the steps taken in this analysis, we will describe the reconstruction of the $K[^{\tau}e][^{\tau}\mu]$ mode as it contains each of the lighter lepton species. The other three decay modes will be summarized in the appendix.

Each event is stored as a table of track objects and cluster objects. After converting a Belle event using `b2bii`, the first step is to fill particle lists (e.g., a list of kaon candidates, a list of moun candidates . . .) with the final-state particles in the decay chain. The purity of a particle list depends on selections applied when filling the list. There are many predefined variables available in `basf2` to aid in this process.

6.3.1 Skim

We begin by applying our skim criteria to each event. A skim is an event-based selection applied prior to reconstruction for the purpose of decreasing the total number of events that need to be reconstructed. First, it is necessary to describe a few event-shape variables. Sphericity,

$$S = (3/2)(\lambda_2 + \lambda_3), \tag{6.1}$$

is a linear combination of the 2nd and 3rd eigenvalues of the sphericity tensor. Fox-Wolfram R2, or just R2, is the ratio of the 2nd to the 0th Fox-Wolfram moments [40]:

$$H_l = \sum_{i,j} \frac{|p_i||p_j|}{E_{\text{tot}}^2} P_l(\cos\phi_{ij}), \tag{6.2}$$

where P_l is a Legendre polynomial. In our case, we require sphericity > 0.05 and R2 < 0.75. These selections throw away very jet-like events (particle fluxes are highly collimated). We require fewer than 30 tracks in the event, each with at least one hit in the CDC and momentum < 5.3 GeV/c. This eliminates events with extensive contamination from beam background and events with more than one e^+e^- collision. Finally, and very specific to this analysis, we reconstruct electron and muon candidates using the criteria described below, and we require that the event has exactly the number of muons and electrons needed to reconstruct our decay mode. We will discuss tagging below, but remember that this step is responsible for our approach of inclusive hadronic tagging. Essentially, we do not want tag-B mesons with lepton daughter particles.

6.3.2 Particle Identification

To reconstruct our decay we need to identify kaons, electrons (or positrons), and muons. This is done by identifying charged tracks that pass selections on things like origin, momentum, and particle-species identification.

Common variables used for charged-track selections include θ, the polar angle of the reconstructed track with respect to the positron beam axis, and impact parameters dr and dz, which are the radial and longitudinal distance of the closest approach between a reconstructed track and the estimated e^+e^- interaction point. Particle momentum (determined by track curvature, magnetic field strength, an electric-charge hypothesis, and the polar angle) is also used for all of our charged-track selections. The momentum can be boosted to any reference frame, but the lab frame and $\Upsilon(4\mathrm{S})$ rest frame are the most common. Components of momentum may also be used. For example, transverse momentum, $\mathrm{p_T}$, is the component of momentum perpendicular to the magnetic field and therefore the only component acted on by the Lorentz force. In addition, transverse momentum does not change appreciably when boosting from the lab frame to the $\Upsilon(4\mathrm{S})$ rest frame (only the small x component due to the electron-positron crossing angle affects the transformation).

The particle species is determined using particle identification (PID) variables. These variables differ from one particle species to another but generally are built using information like energy loss in the CDC (dE/dx), Cherenkov photon counts in the ACC, flight time from the TOF counters, and energy deposition and shower shape in the ECL. In the case of muons with transverse momentum larger than 600 $\frac{\mathrm{MeV}}{\mathrm{c}}$, the matching of CDC tracks and ECL clusters with tracks in the KLM detector is also used.

6.3.3 Kaons

For candidate kaons, the polar angle must be within the acceptance of the CDC, and the radial and longitudinal impact parameters are required to be within an ellipsoid around the IP. The size of the ellipsoid scales inversely with the transverse momentum of the kaon so that high p_T tracks have tighter impact-parameter selections.

$$\frac{\mathrm{dr}^2}{\left(\frac{0.004537}{\mathrm{p_T}-0.058975}+0.002159\right)^2}+\frac{\mathrm{dz}^2}{\left(\frac{0.004499}{\mathrm{p_T}-0.059205}+0.339197\right)^2}<4^2\,. \tag{6.3}$$

This equation needs some explanation. The impact-parameter resolution of a kaon improves with increasing transverse momentum of the kaon. Basically, as transverse momentum increases, more layers of the CDC are involved in the track reconstruction. The usual approach is to ignore transverse momentum and make loose selections on the radial and longitudinal impact parameters, effectively drawing a cylinder around the interaction point. This helps reduce beam-induced background that did not originate near the interaction point but does not remove tracks with poor track-fit results. To try to improve this, we studied the impact-parameter resolution in different transverse-momentum ranges for kaons and muons using our $K[{}^{\tau}\mu][{}^{\tau}\mu]$ signal MC.

Using 16 transverse-momentum bins, we plot the radial (longitudinal) impact parameter distributions in each and fit them to one-sided Gaussian (full Gaussian) distributions to extract the average width (Figures 6.1 and 6.3). We plot the width-fit results from these 16 fits in Figures 6.2 and 6.4. In these plots, "daughter0" is the kaon while "daughter1" and "daughter2" are the muons. The trends look like rectangular hyperbolas, so we fit them to $1/\mathrm{p_T}$ with free parameters for horizontal and vertical offset.

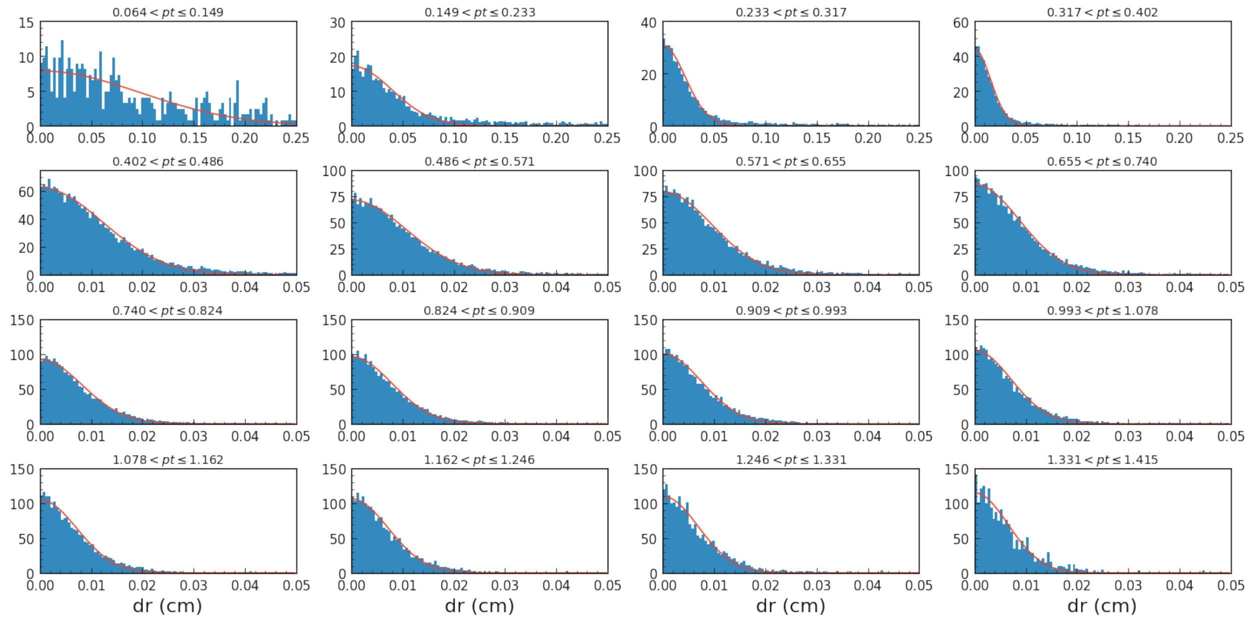

Figure 6.1: Charged-kaon radial-impact parameter distributions in 16 bins of kaon transverse momentum for the $K[^\tau\mu][^\tau\mu]$ mode. Each panel is fitted to a Gaussian distribution with the mean constrained at zero to extract the width. The fit results and fit errors are saved for the next step. The same procedure is repeated for the signal muons (not shown).

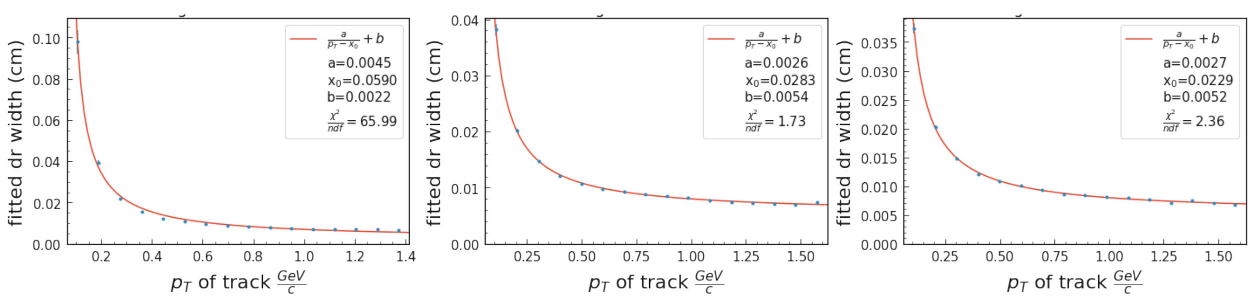

Figure 6.2: Width-fit results and errors from Fig. 6.1. The first panel is for our signal kaon while the second and third are for signal muons, all from the $K[^\tau\mu][^\tau\mu]$ mode. As we later select muons with momenta larger than 600 $\frac{\mathrm{MeV}}{\mathrm{c}}$ we find that this procedure is not beneficial for muons. The red line is a fit to a rectangular hyperbola with floating vertical and horizontal offsets.

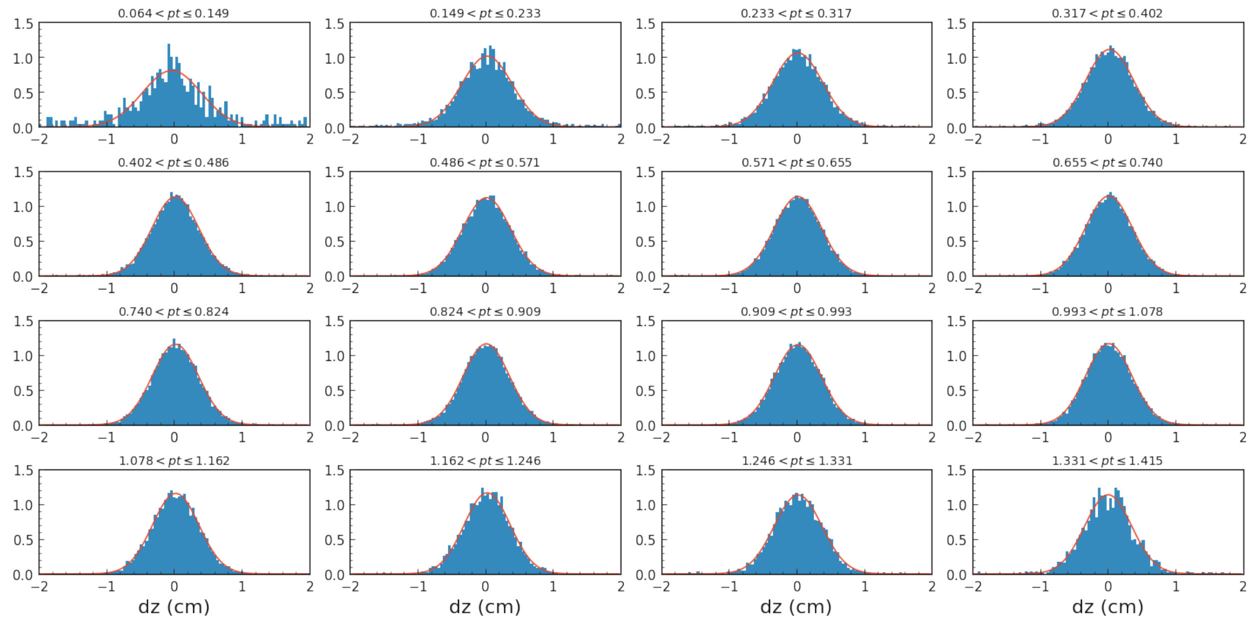

Figure 6.3: Charged-kaon longitudinal-impact parameter distributions in 16 bins of kaon transverse momentum for the $K[^{\tau}\mu][^{\tau}\mu]$ mode. Each panel is fitted to an unconstrained Gaussian distribution. The fit results and fit errors are saved for the next step. The same procedure is repeated for the signal muons (not shown).

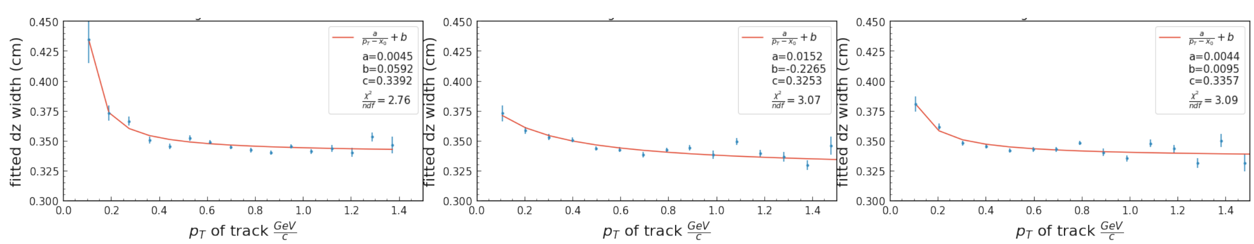

Figure 6.4: Width fit results and errors from Fig. 6.3. The first panel is for our signal kaon while the second and third are for signal muons, all from the $K[^{\tau}\mu][^{\tau}\mu]$ mode. Again, we see the muon results are rather flat above 600 $\frac{\mathrm{MeV}}{\mathrm{c}}$.

We see that muon dependence on transverse momentum is not very strong, and since we'll later select only muons with momentum larger than 600 $\frac{\mathrm{MeV}}{\mathrm{c}}$ to avoid complications from misidentified pions, we choose not to apply this procedure to muon selection.

With functions for the radial and longitudinal impact parameters in terms of transverse momentum, we can build an ellipsoid with our radial impact parameter function as the semiminor axis and our longitudinal impact parameter function as the semimajor axis. By construction, $\mathrm{dr}(\mathrm{p_T})$ and $\mathrm{dz}(\mathrm{p_T})$ specify the expected $1\,\sigma$ Gaussian width at different transverse momenta. To get $\mathrm{N}\,\sigma$, these functions should be multiplied by N. To turn this into a selection criterion, we make the equation of the ellipsoid an inequality and move the N to the right side, thus arriving at Eq. 6.3. For our final selection, we choose the very loose value of $\mathrm{N} = 4$. To visualize how this selection

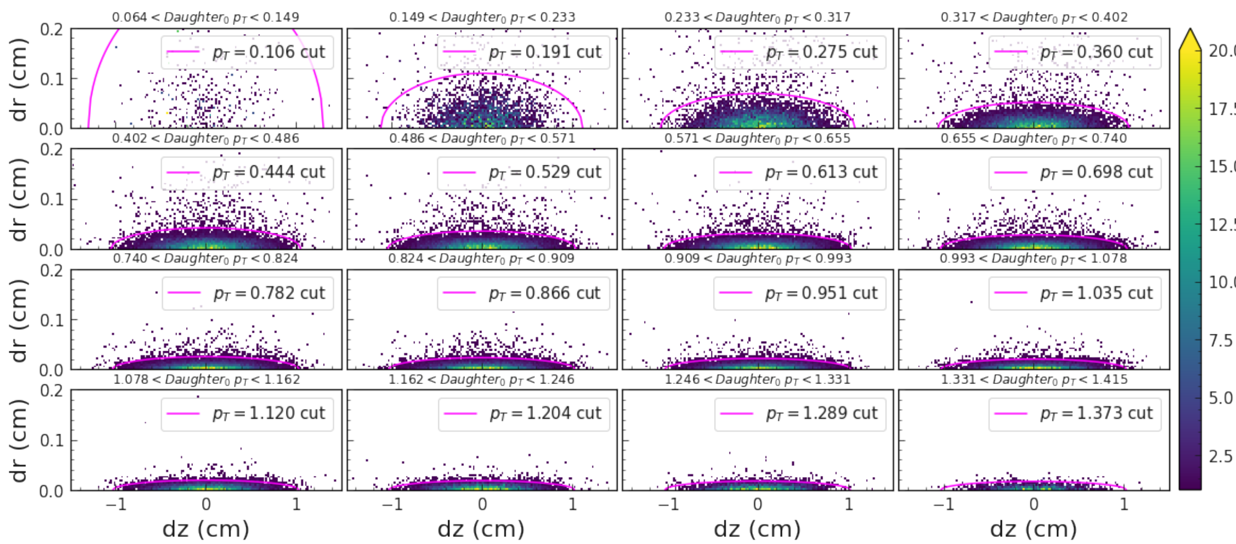

Figure 6.5: Visualization for elliptical selection of kaon impact parameters. The 16 panels are 16 bins of transverse momentum. Each panel is a 2D histogram in dr vs. dz. Overlaid on each panel is Eq. 6.3 for the $3\,\sigma$ case evaluated at the average transverse momentum for that bin.

behaves, we've plotted the case for $\mathrm{N} = 3$ in Fig. 6.5.

Aside from our complicated impact parameter selection, there are three selections placed on binary PID variables,

$$\frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_e} > 0.3, \quad \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\mu} > 0.4, \quad \text{and } \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi} > 0.4, \tag{6.4}$$

which are combined likelihoods based on information from the ACC, TOF counter, and CDC. These PID selections are plotted in 6.6.

To study kaon identification, we take a subset of the generic B^+B^- MC, reconstruct all charged tracks using loose selection criteria, and create an output file with the above binary PID variables as well as a code for the true simulated particle species. To minimize kaon misidentification, we

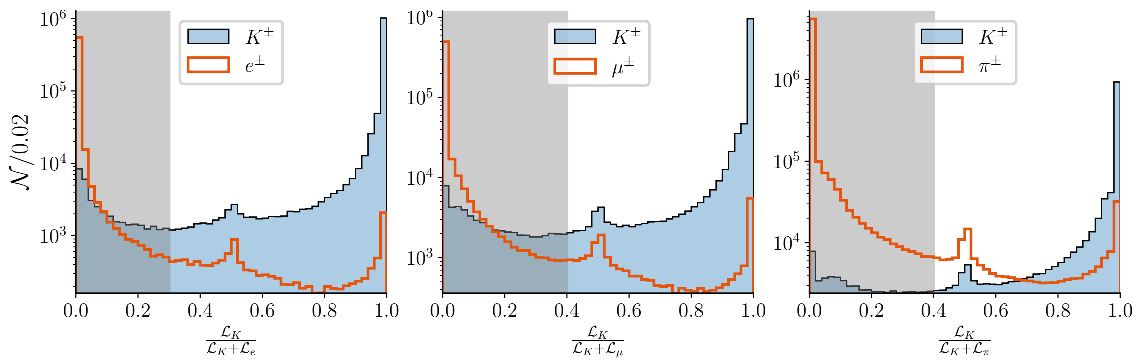

Figure 6.6: Likelihood variables used for $K^\pm$ selection. The data is a subset of the generic B^+B^- MC. The gray band indicates values that are excluded by our selection.

construct the simple figure of merit

$$\frac{\epsilon(K)}{\sqrt{\epsilon(e)+\epsilon(\mu)+\epsilon(\pi)+\epsilon(K)}}\,, \tag{6.5}$$

where ϵ_i is the fraction of particle species i passing the binary PID selection of interest. Taking the fraction of each particle species rather than the total is equivalent to normalizing the data so that there are equal numbers of each species. We then scan over this figure of merit by varying values of the binary PID selection of interest. In each scan, the other two binary PIDs are also applied using their respective selections shown in Eq. 6.4. The results of the three scans (one for each binary PID) are shown in Fig. 6.7.

We choose a selection of 0.3 for kaon-electron separation which is very near the peak figure of merit. For kaon-muon and kaon-pion separation, we choose a selection 0.4 because the figures of merit are quite flat for these variables and we wish to maximize event retention when reconstructing $B^+ \to K^+\tau^+\tau^-$. For this subset of the generic B^+B^- MC, we achieve a true-kaon efficiency of 91.8 % and fake-electron, -muon, and -pion efficiencies of 2.2 %, 4.2 %, and 2.7 %, respectively.

To examine the impact of each selection, we prepared a reconstruction script that applies the selections one at a time and tabulates the average candidate multiplicity (average number of candidate kaons per event) and retention (fraction of events for which there was at least one kaon). This reconstruction was run on a subset of the signal MC data, generic $B\overline{B}$ MC data, and $q\overline{q}$ MC data. The results are shown in Table 6.1. We found that for signal MC we retain 83.5 % of events while for all the background MC types we retain between 55.2 % and 73.2 % of events. Again, we are focusing on the $K[^\tau e][^\tau\mu]$ mode, but the results for the other three modes are similar. Equivalent tables for the other three modes may be found in the appendix.

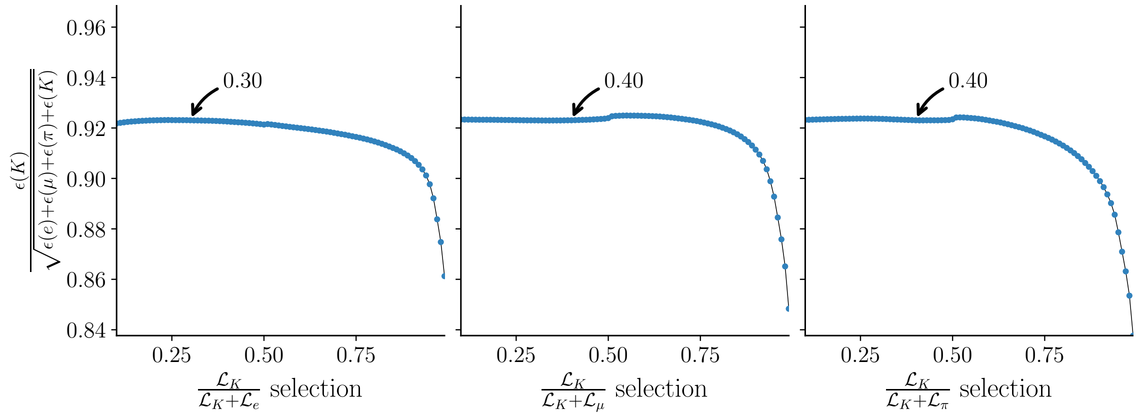

Figure 6.7: Figure of merit scans for charged kaon identification. The variables kaon vs. electron, muon, or pion are the binary PID variables from Eq. 6.4. In the vertical axis, ϵ_i is the fraction of particle species i passing the binary PID selection of interest. This amounts to a normalization of each particle species. In the first panel, we choose a selection of 0.3 for kaon-electron separation very near the peak figure of merit. In the last two panels, we choose a selection of 0.4 for kaon-muon and kaon-pion separation because the figures of merit are quite flat for these variables and we wish to maximize event retention when reconstructing $B^+ \to K^+\tau^+\tau^-$.

Table 6.1: Average candidate multiplicity per event/retention for selections made to the signal kaon particle list in the $K[{}^{\tau}e][{}^{\tau}\mu]$ mode. The function $f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T})$ is a dynamic ellipsoid around the interaction point with size inversely proportional to transverse momentum.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.992	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in CDC Acceptance	-	9.7 / 0.992	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_e} > 0.3$	-	3.2 / 0.979	4.1 / 0.977	4.0 / 0.979	3.5 / 0.928	3.6 / 0.944
3	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\mu} > 0.4$	-	2.1 / 0.926	2.5 / 0.859	2.4 / 0.879	2.1 / 0.737	2.2 / 0.822
4	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\pi} > 0.4$	-	2.0 / 0.916	2.3 / 0.838	2.3 / 0.861	2.0 / 0.703	2.1 / 0.799
5	$f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T}) < 4^2$	cm	1.6 / 0.835	1.6 / 0.684	1.7 / 0.732	1.5 / 0.552	1.6 / 0.653

6.3.4 Electrons

For candidate $e^{\pm}$, we also require the polar angle to be within the acceptance of the CDC, we use the standardized impact-parameter selections of dr < 0.5 cm and $|\mathrm{dz}| < 1.5$ cm. We require lab-frame momentum to be larger than $0.25\,\frac{\mathrm{GeV}}{\mathrm{c}}$ and we require the center-of-mass frame ($\Upsilon(4\mathrm{S})$ rest frame) momentum to be smaller than $1.9\,\frac{\mathrm{GeV}}{\mathrm{c}}$. The reason for using a minimum momentum in the lab frame but maximum momentum in the center-of-mass frame is simply that in signal-MC data we see a more defined minimum boundary in the lab frame whereas we see more background rejection by placing the maximum boundary in the center-of-mass frame. Further, we use two electron identification variables. One, $e^{\pm}$ ID (BDT), which is the output classifier of a boosted-decision tree (BDT) that was recently trained on Belle data, and the second, $e^{\pm}$ ID, which is the legacy electron identification variable used for Belle analyses. Specifically, $e^{\pm}$ ID is only applied to tracks for which $e^{\pm}$ ID (BDT) does not work, namely charged tracks without associated ECL energy deposits. On the tracks for which $e^{\pm}$ ID (BDT) does work, we require $e^{\pm}$ ID (BDT) > 0.3, and for those in which it does not work we require $e^{\pm}$ ID > 0.6.

We take candidates passing the above selection criteria and pass them through a bremsstrahlung correction module. Bremsstrahlung correction recalculates the four momenta of the electron candidates after including ECL clusters within 0.1 rad of tangent lines to the electron track as seen in the detector reference frame. The list of bremsstrahlung corrected electron candidates is used for our skim criteria as described in 6.3.1. Our final selection on electrons before forming candidate B mesons is that the center-of-mass-frame (Υ(4S) rest frame) momentum, $p^{\star}_{e^{\pm}}$, be less than 1.9 $\frac{\mathrm{GeV}}{\mathrm{c}}$.

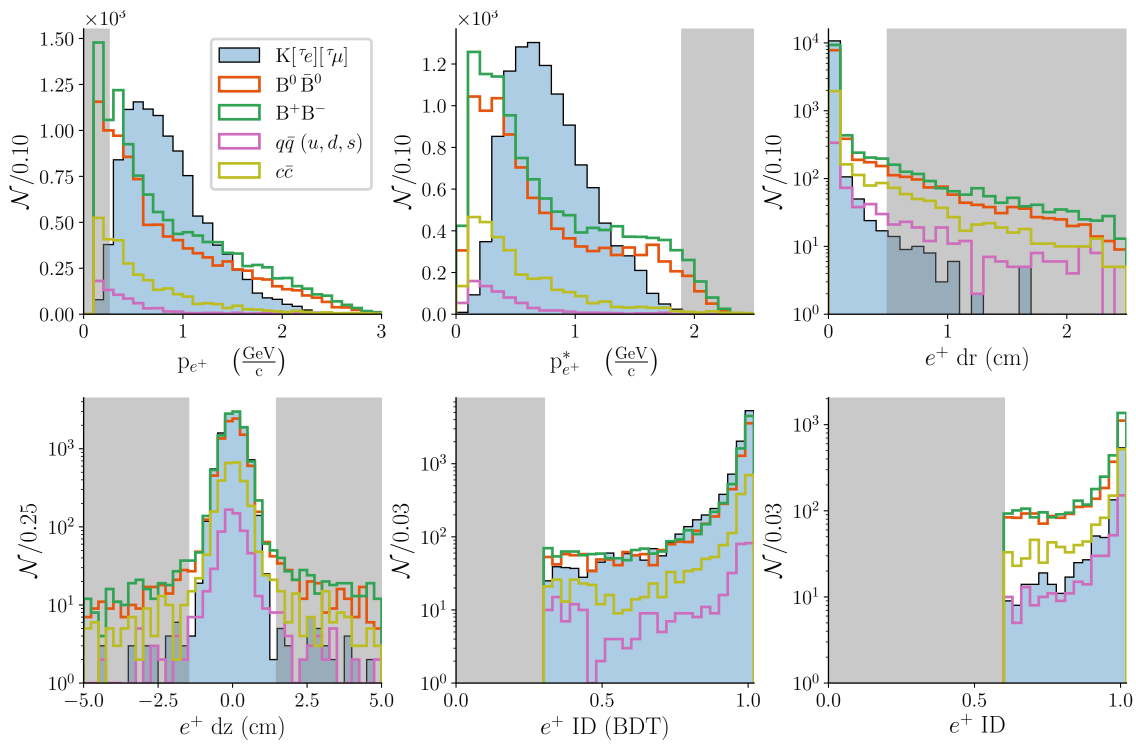

Figure 6.8: Variables used for $e^{\pm}$ selection. The $K[^{\tau}e][^{\tau}\mu]$ decay mode is shown. The shaded regions are excluded by the selection. The filled blue histograms are from signal MC, and the unfilled histograms are from the various background components. For minimum electron momentum, we see a steeper rise in signal MC when using the lab frame than when using the center-of-mass frame, so we choose $p_{e^+} > 0.25\,\frac{\mathrm{GeV}}{\mathrm{c}}$. For maximum electron momentum, however, we see we can reject more $B\overline{B}$ background in the center-of-mass frame, so we choose $p^{\star}_{e^+} < 1.9\,\frac{\mathrm{GeV}}{\mathrm{c}}$. For the radial and longitudinal impact parameters we use the standard cuts of 0.5 cm and 1.5 cm, respectively. Last, we use two electron ID variables. We choose the BDT-based electron ID larger than 0.3, unless it returns an undefined number, in which case we choose the non-BDT-based electron ID larger than 0.6 .

As with the kaon candidates, we present a cut-flow table for the $K[^{\tau}e][^{\tau}\mu]$ mode in Table 6.2 and you may find the other three decay modes in the appendix. We find an event retention of 81.4 % for the signal MC and between 8.3 % and 34.0 % for the various background MC datasets. Note that the average candidate multiplicity per event for signal electrons in this mode is 1.2, which indicates that up to 20 % of these events may not have a fully-hadronic inclusive tag-B meson. For the decay mode with two final-state electrons this value is 1.84, which indicates that at most 92 % of them can form $K[{}^{\tau}e][{}^{\tau}e]$ candidates.

Table 6.2: Average candidate multiplicity per event/retention for selections made to the signal electron particle list in the $K[{}^{\tau}e][{}^{\tau}\mu]$ mode.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.992	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.7 / 0.992	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	p > 0.25	$\frac{\mathrm{GeV}}{\mathrm{c}}$	7.1 / 0.992	8.9 / 0.994	8.8 / 0.994	7.3 / 0.951	7.6 / 0.962
3	dr < 0.5	cm	6.0 / 0.992	6.9 / 0.994	6.9 / 0.994	5.9 / 0.951	6.0 / 0.962
4	|dz| < 1.5	cm	5.7 / 0.992	6.5 / 0.994	6.5 / 0.994	5.6 / 0.951	5.7 / 0.962
5	$e^{\pm}$ ID (BDT) > 0.3 or $e^{\pm}$ ID > 0.6	-	1.2 / 0.817	1.2 / 0.354	1.2 / 0.342	1.2 / 0.087	1.1 / 0.183
6	$\|\vec{\mathrm{p}}\star\| < 1.9$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	1.2 / 0.814	1.2 / 0.340	1.2 / 0.326	1.2 / 0.083	1.1 / 0.176

6.3.5 Muons

The $\mu^{\pm}$ candidates are selected in very much the same way as the $e^{\pm}$ candidates: standardized impact-parameter selections; two momenta selections; and two particle ID selections. The selections are summarized in Table 6.3 and plotted in Fig. 6.9.

Table 6.3: Average candidate multiplicity per event/retention for selections made to the signal muon particle list in the $K[{}^{\tau}e][{}^{\tau}\mu]$ mode.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.992	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.8 / 0.992	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	p > 0.6	$\frac{\mathrm{GeV}}{\mathrm{c}}$	3.9 / 0.987	3.9 / 0.986	3.9 / 0.988	3.9 / 0.942	3.9 / 0.951
3	dr < 2.0	cm	3.7 / 0.987	3.6 / 0.982	3.7 / 0.985	3.7 / 0.939	3.6 / 0.948
4	|dz| < 4.0	cm	3.6 / 0.985	3.4 / 0.979	3.5 / 0.982	3.6 / 0.936	3.5 / 0.944
5	$\mu^{\pm}$ ID (BDT) > 0.4 or $\mu^{\pm}$ ID > 0.95	-	1.1 / 0.438	1.1 / 0.203	1.1 / 0.209	1.0 / 0.039	1.0 / 0.086
6	$\|\vec{\mathrm{p}}\star\| < 1.9$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	1.1 / 0.430	1.1 / 0.187	1.1 / 0.190	1.0 / 0.036	1.0 / 0.079

The results of our muon selection for the $K[{}^{\tau}e][{}^{\tau}\mu]$ decay mode is that we retain 43 % of signal MC events with an average muon multiplicity of 1.1. For the $K[{}^{\tau}\mu][{}^{\tau}\mu]$ decay mode, we retain 63 % of signal MC events, but the average muon multiplicity is only 1.3, which is problematic because we need two muons for this decay mode. For background MC, we retain as little as 3.6 % of events

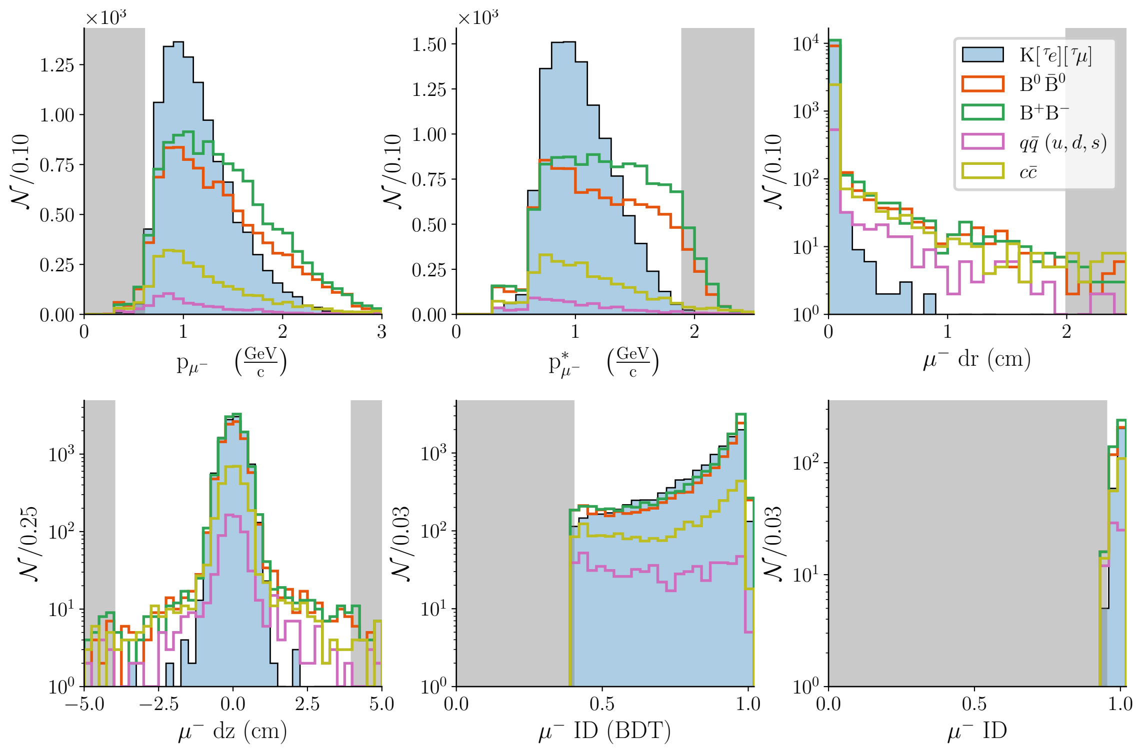

Figure 6.9: Variables used for $\mu^{\pm}$ selection. The $K[\,^{\tau}e][\,^{\tau}\mu]$ decay mode is shown. The shaded regions are excluded by the selection. The filled blue histograms are from signal MC, and the unfilled histograms are from the various background components. For minimum muon momentum, we see a steeper rise in signal MC when using the lab frame than when using the center-of-mass frame, so we choose $\mathrm{p}_{\mu^-} > 0.6$. For maximum muon momentum, however, we can reject more $B\overline{B}$ background in the center-of-mass frame, so we select $\mathrm{p}^{\star}_{\mu^-} < 1.9$. For impact parameters dr and dz, we use the standard cuts of 2 cm and 4 cm, respectively. Last, we use two muon ID variables. We choose the BDT-based muon ID larger than 0.4 unless it returns an undefined number, in which case we choose the non-BDT-based muon ID larger than 0.95.

(u, d, s continuum MC), and as much as 19 % (B^+B^- MC). The background retentions due to muon selection alone are nearly identical across all four decay modes.

6.4 B-Meson Candidates Selection

To form a B^+-meson candidate from an event, we combine the 4-momenta from one K^+ candidate, one e^+ candidate, and one μ^- candidate. This is repeated for all combinations of final state particles and again for the charge-conjugated decay. As with the particle selection, we now make selections on our B candidates to decrease the candidate multiplicity and otherwise throw out candidates that are not kinematically consistent with our decay mode. These include selections on the three invariant-mass-squared combinations of our final state particles, beam-constrained mass,

$$M_{bc} \equiv \sqrt{E^{\star 2}_{\text{beam}} - \left|\vec{p}^{\star}_{B}\right|^2}, \tag{6.6}$$

and energy difference,

$$\Delta E \equiv E^{\star}_{B} - E^{\star}_{\text{beam}}, \tag{6.7}$$

where $\star$ denotes the center-of-mass reference frame. These selections are tabulated in Table 6.4 and plotted in Fig. 6.10.

Table 6.4: Average candidate multiplicity per event/retention for selections made during signal $B^{\pm}$ reconstruction in the $K[^{\tau}e][^{\tau}\mu]$ mode. Note that event retention starts smaller than in the previous three cut-flow tables. This is the effect of combining candidates from the three particle lists to form B candidates. The B selections reduce backgrounds by more than 20 % in all except the $q\bar{q}$ background.

	Selection	Unit	$B^{\pm}_{\text{sig}}B^{\mp}_{\text{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			1.1 / 0.088	$1.1 / 4.64 \times 10^{-3}$	$1.1 / 5.64 \times 10^{-3}$	$1.2 / 1.92 \times 10^{-4}$	$1.2 / 9.84 \times 10^{-4}$
1	$\lvert\ell^+\ell^-\rvert^2 < 9$	$\frac{\text{GeV}^2}{c^4}$	1.1 / 0.088	$1.1 / 4.56 \times 10^{-3}$	$1.1 / 5.51 \times 10^{-3}$	$1.2 / 1.90 \times 10^{-4}$	$1.2 / 9.70 \times 10^{-4}$
2	$0.5 < \lvert K^+\ell^-\rvert^2 < 8$	$\frac{\text{GeV}^2}{c^4}$	1.1 / 0.088	$1.1 / 4.47 \times 10^{-3}$	$1.1 / 5.39 \times 10^{-3}$	$1.1 / 1.74 \times 10^{-4}$	$1.2 / 9.42 \times 10^{-4}$
3	$0.3 < \lvert K^+\ell^+\rvert^2 < 8$	$\frac{\text{GeV}^2}{c^4}$	1.1 / 0.088	$1.1 / 4.42 \times 10^{-3}$	$1.1 / 5.27 \times 10^{-3}$	$1.1 / 1.72 \times 10^{-4}$	$1.2 / 9.28 \times 10^{-4}$
4	$-3.7 < \Delta E < -1.3$	GeV	1.1 / 0.087	$1.1 / 4.13 \times 10^{-3}$	$1.1 / 4.84 \times 10^{-3}$	$1.1 / 1.72 \times 10^{-4}$	$1.1 / 8.90 \times 10^{-4}$
5	$M_{bc} > 4.8$	$\frac{\text{GeV}}{c^2}$	1.1 / 0.084	$1.1 / 3.73 \times 10^{-3}$	$1.1 / 4.29 \times 10^{-3}$	$1.1 / 1.54 \times 10^{-4}$	$1.1 / 8.14 \times 10^{-4}$

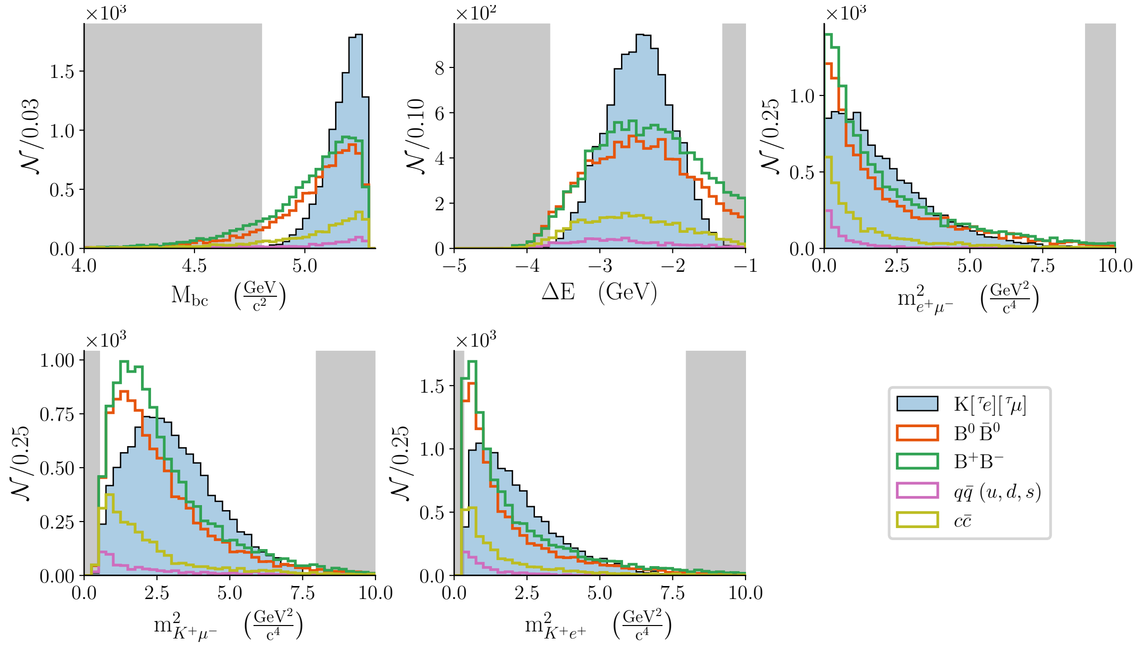

Figure 6.10: Variables used for selections of B-meson candidates. To produce these plots, our reconstruction was run on a subset of the MC data for our signal and background components without applying any cuts. The shaded regions are excluded by the selections.

6.5 Inclusive Tag and Best Candidate Selection

For each remaining B-meson candidate in an event, we combine the remaining tracks and clusters to form the inclusive tag B meson associated with that candidate. When doing this, we apply additional selections to these tracks and clusters to eliminate beam background and noise that did not originate from either B decay. These are summarized in tables 6.5 and 6.6. These selections are important to our analysis since they can directly affect the quality of our inclusive tag-B meson. Because it is an inclusive tag reconstruction, we use the width of the tag B ΔE and $\mathrm{M_{bc}}$ distributions as a metric for evaluating our selections. A special reconstruction of our signal MC was prepared that saved ΔE and $\mathrm{M_{bc}}$ for many variations of our selections. For each variation, we numerically measured the width of these two distributions. For $\mathrm{M_{bc}}$, we measured the point at which the reversed cumulative distribution reached 68.27 % of the total number of entries, effectively summing the histogram bins from right to left until we reached the $1\,\sigma$ threshold. For ΔE, we split the distribution at the peak and measured the cumulative distribution in both the left and right directions away from the peak. This gave us two $1\,\sigma$ values for the left and right sides respectively. The sum of these two was used as a measure of the ΔE width. A good performing selection should have a high $1\,\sigma$ threshold for the $\mathrm{M_{bc}}$ distribution (indicating a narrower distribution) and a small $1\,\sigma$ value for the ΔE distribution. An example of this optimization performed on the selection for the minimum number of CDC hits is shown in Fig. 6.11. All of the track selection criteria were

Table 6.5: Mask for tracks that can be used to form the inclusive tag-B meson. Row three is a piecewise selection for impact parameters: for tracks with center-of-mass momentum above 500 MeV/c, it is a dynamic ellipsoid shrinking at a hyperbolic rate as track momentum increases; for tracks with momentum below 500 MeV/c, it is a fixed ellipsoid with a semiminor axis of 15 cm and a semimajor axis of 34 cm. As impact parameters are determined from track fits, their uncertainties can help eliminate tracks with poor track fit results, which may include fake tracks built out of noise hits.

	cut	unit
1	Number of CDC hits > 9	-
2	$.06 < \lvert \mathrm{p}^\star \rvert \leq 3.2$	$\frac{\mathrm{GeV}}{c}$
3a ($\mathrm{p}^\star > .5$)	$\frac{\mathrm{dr}^2}{\left(\frac{4}{\mathrm{p}^\star - .261002} - 1.736513\right)^2} + \frac{\mathrm{dz}^2}{\left(\frac{7}{\mathrm{p}^\star - .241694} - 2.099669\right)^2} < 1$	cm
3b ($\mathrm{p}^\star \leq .5$)	$\frac{\mathrm{dr}^2}{15^2} + \frac{\mathrm{dz}^2}{34^2} < 1$	cm
4	Radial impact parameter error < .125	cm
5	Longitudinal impact parameter error < .8	cm

optimized in the same way. For cluster selection, we used standard selections for θ, for the number of crystals involved in a cluster, and for the cluster time. For the cluster minimum energy in the backward, barrel, and forward regions of the ECL, we used the same optimization procedure as above.

Table 6.6: Mask for ECL clusters that can be used to form the inclusive tag-B meson. The number of crystals in a cluster is a weighted sum in the case of overlapping clusters. Cluster time is calibrated so that a photon originating from the IP at the same time as event t_0 should have a cluster time consistent with zero.

	cut	unit
1	$0.2967 < \theta < 2.618$	rad.
2	Number of crystals > 1.5	-
3	Cluster time: $\lvert \Delta t \rvert < 200$	ns
4a	Backward Endcap: E > 0.06	GeV
4b	Barrel: E > 0.06	GeV
4c	Forward Endcap: E > 0.08	GeV

For the impact parameter selections, we wanted something similar to the dynamic ellipsoid used for the kaon selection, but we went about it differently. We began with studying the 2D impact parameter distribution for good tracks (good meaning the fit minus the truth pulls were less than three for both the radial and longitudinal impact parameters). It can be seen that below a momentum of $0.5\,\frac{\mathrm{GeV}}{\mathrm{c}}$, there is no dependence on track momentum. We began by constructing a constant symmetric ellipsoid for this momentum range. The semiminor and semimajor axes of this ellipsoid were optimized using the same procedure as above (minimizing the width of ΔE and M_{bc} for the tag B meson). Above $0.5\,\frac{\mathrm{GeV}}{\mathrm{c}}$, the size of the ellipsoid should decay hyperbolically

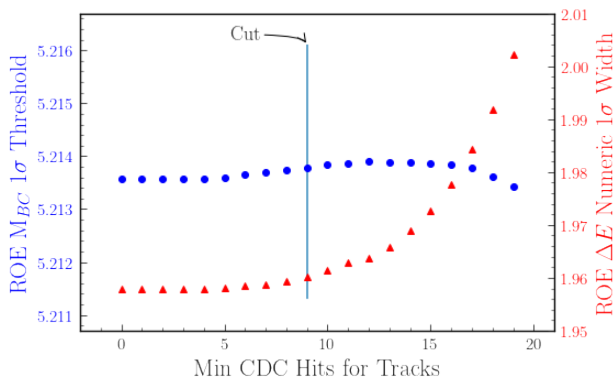

Figure 6.11: Example of our scan to optimize selection criteria for tracks that are used in the inclusive tag-B meson. Here, ROE means "rest of event" and is a synonym for our inclusive tag B. Each test point is one selection for minimum number of CDC hits that a track must have. The blue is the $1\,\sigma$ threshold when summing $\mathrm{M_{bc}}$ from right to left, so a higher value means a narrower $\mathrm{M_{bc}}$ distribution. The red is the sum of the $1\,\sigma$ widths on the left side and on the right side of the ΔE peak, so smaller value means a narrower ΔE distribution. Our final selection is annotated "cut" and was chosen where separation between the red and blue curves is close to the maximum separation but also high enough that we can avoid poor track fit results when there are too few hits for a good fit.

as a function of momentum. We chose the plateau of our hyperbolas for both the radial and longitudinal directions, so basically we would never draw an ellipsoid smaller than that for even the highest momentum track. We optimized the shape parameter of each hyperbolic function, again, using the same procedure as above. The terms with six decimal places were then calculated to make sure our ellipsoid equals the maximum size at $0.5\,\frac{\mathrm{GeV}}{\mathrm{c}}$ and approaches the minimum size at higher momenta.

This procedure involved a lot of iteration, and unfortunately, at the time of writing this, we discovered that the constant ellipsoid for track momenta below $0.5\,\frac{\mathrm{GeV}}{\mathrm{c}}$ does not exactly equal the dynamic ellipsoid at precisely $0.5\,\frac{\mathrm{GeV}}{\mathrm{c}}$. The radial sizes agree but the longitudinal size becomes abruptly smaller, from 34 cm to 25 cm, when momentum increases past $0.5\,\frac{\mathrm{GeV}}{\mathrm{c}}$. We do not anticipate this should have a large impact on our results but it should be corrected going forward.

An example of the tag B ΔE and $\mathrm{M_{bc}}$ distributions, before any further selections are applied, is presented in Fig. 6.12.

Now, after having constructed our signal B candidates and a corresponding tag B candidate, we have one or more $\Upsilon(4\mathrm{S}) \to B_{\mathrm{sig}}B_{\mathrm{tag}}$ candidates for our event. We want tag-B mesons that are kinematically consistent with a hadronic decay. To achieve this, we apply selections on $\mathrm{M_{bc}}$ and ΔE. Based on a 2D histogram of tag B ΔE vs. tag B $\mathrm{M_{bc}}$ from our signal Monte Carlo data, we chose an ellipse-shaped selection that would preserve the majority of correctly reconstructed signal events without allowing the extra background that would have been present if a box selection was

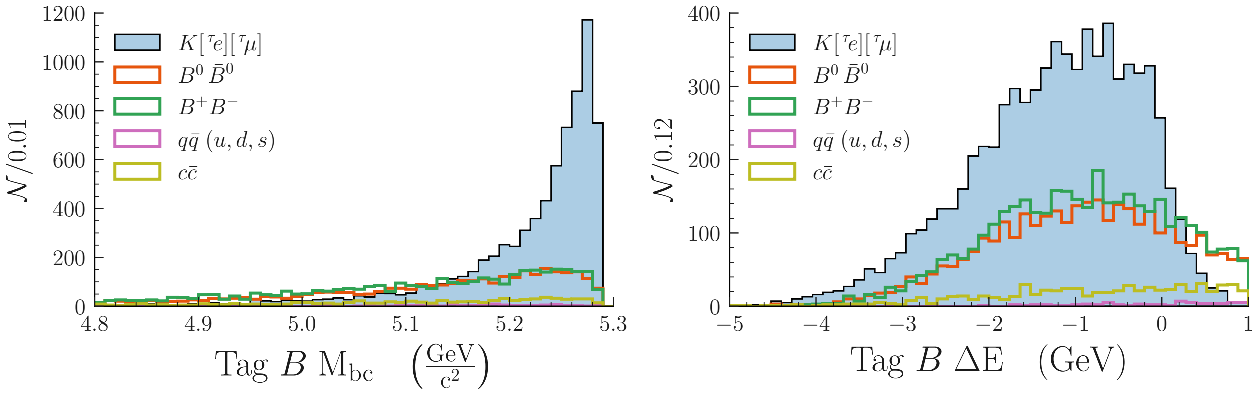

Figure 6.12: Tag B ΔE and $\mathrm{M_{bc}}$ distributions after finalizing both our signal particle selections and our track and cluster selections. The data shown is from a subset of all our MC data.

used. This is the third and final time you'll hear about an ellipse in this analysis. In Fig. 6.13 the chosen ellipse is drawn on top of the distribution. It has a semi-minor (major) axis of $0.13\,\mathrm{GeV/c^2}$ ($1.7\,\mathrm{GeV}$) and is rotated by 35 mrad. All $\Upsilon(4S)$ candidates with a tag-B meson not passing this selection are removed from the event.

Finally, the list of $\Upsilon(4S)$ candidates is as small as it is going to get. A single candidate must be chosen. We perform a vertex fit of each candidate signal-B meson using a global decay chain vertex fitter [41]. The fitter simultaneously fits the entire decay chain by finding the best decay vertex positions and particle momenta. The inclusive tag-B meson decay vertex is fit using a geometric fit that is designed for cases when the tag is not explicitly reconstructed. We take the combined p-value using the methods of Ref. [42] and we choose the candidate with the best combined p-value.

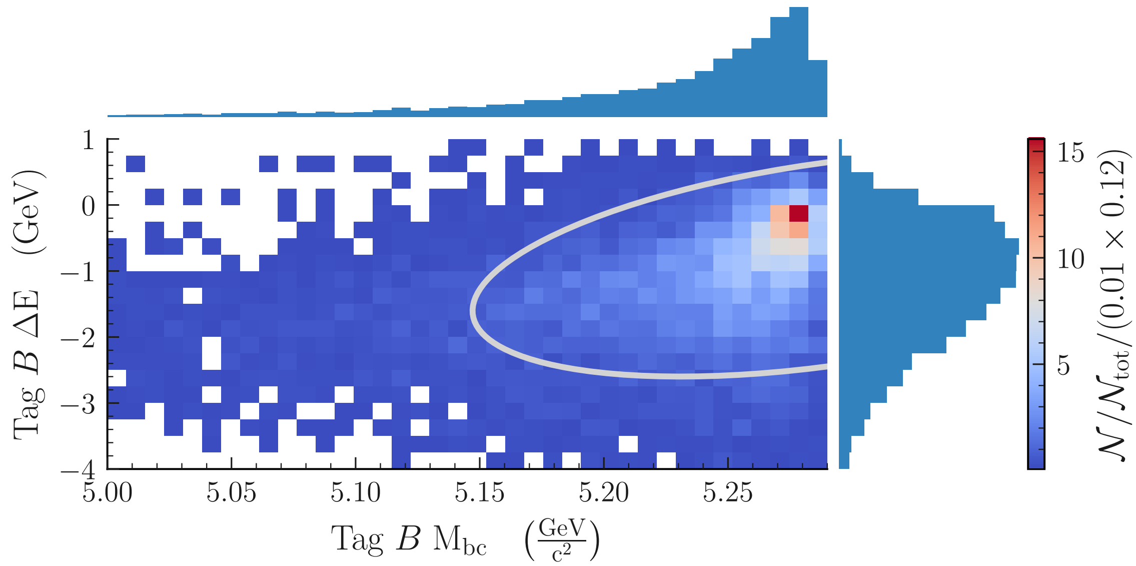

Figure 6.13: A 2D histogram of the tag B ΔE vs. the tag B $\mathrm{M_{bc}}$ for 10^6 MC events of the $K[^{\tau}e][^{\tau}\mu]$ mode. In order to preserve as much signal as possible and to select an inclusive tag-B meson that is consistent with a hadronic B decay, we require that our tag B is in the region indicated by the ellipse. It has a semi-minor (major) axis of $0.13\,\mathrm{GeV/c^2}$ ($1.7\,\mathrm{GeV}$) and is rotated by 35 mrad. The color scale times the bin size integrates to one.

6.6 Reconstruction Summary

The result of running the reconstruction is a file containing all the variables we wish to keep for those events that pass the selection criteria. One such variable is a flag that specifies whether or not our reconstruction assembled the particles of an event correctly, including correct particle species identification and correct order of the entire decay chain. This is called Monte Carlo truth matching. This enables us to calculate our efficiency for reconstructing a particular decay from simulated data. In Table 6.7 we summarize our efficiencies. Reconstruction efficiency, ϵ_{reco} is the fraction of events that passed the selection criteria. Of those events, the fraction that matched Monte Carlo truth is listed in the second column as $\epsilon_{\mathrm{matched}}/\epsilon_{\mathrm{reco}}$. In the last column, we list the overall Monte Carlo matching efficiency, which is just the product of the first two columns.

Table 6.7: Reconstruction efficiencies (ϵ_{reco}) and Monte Carlo-matching efficiencies (ϵ_{match}) for 1-prong leptonic modes of $B^+ \to K^+\tau\tau$.

Decay Mode	ϵ_{reco}	$\frac{\epsilon_{matched}}{\epsilon_{reco}}$	$\epsilon_{matched}$
$K[^{\tau}e][^{\tau}e]$	0.257	0.708	0.182
$K[^{\tau}\mu][^{\tau}\mu]$	0.057	0.722	0.041
$K[^{\tau}e][^{\tau}\mu]$	0.103	0.870	0.090
$K[^{\tau}\mu][^{\tau}e]$	0.103	0.882	0.091

We find that our decay mode with two final-state $e^{\pm}$ has the highest Monte Carlo-matched efficiency at 18 %. Modes with one final state muon come in at 9 %. The decay mode with two final-state muons comes in last at a mere 4 %.

CHAPTER 7
SIGNAL–BACKGROUND SEPARATION

To further reduce backgrounds from other B-decay channels and continuum backgrounds like $e^+e^- \to q\overline{q}$, where $q = (u,\ d,\ s,\ c)$, we employ multivariate analysis (MVA). We train a boosted decision tree (BDT) using signal Monte Carlo and background Monte Carlo that passed our cuts during reconstruction. The software package used for BDT training is `fastBDT` [43].

A decision tree is a multivariate technique with advantages over a simple cut-based analysis. Instead of discarding events that do not satisfy a selection applied to one particular variable, other variables are checked first, further improving event classification (Fig. 7.1). Once trained, a decision

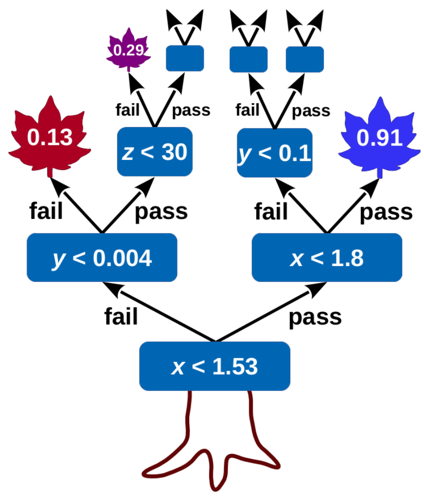

Figure 7.1: Schematic of a decision tree from Ref. [44]. Internal nodes are split by a decision on the value of some variable. Terminal nodes are called leaves and have a purity associated with them.

tree is a cut boundary that can be described by an $(\mathcal{N}-1)$-dimensional surface in the phase space of $\mathcal{N}$ variables, and this surface does not have to be connected. To boost a decision tree means to iteratively add on more trees, each additional one being enriched with training samples that were misclassified by the previous one. In this way, each additional tree is encouraged to explore neighborhoods of the phase space that are poorly classified by the previous trees. Each additional tree is assigned a weight based on its performance, and the final classifier is a weighted average of all the trees. While adding an excessive number of trees can lead to overtraining, it is mitigated by weighted averaging. Even if overtrained, it is not the error rate on the training sample that is important but rather the error rate on the independent testing sample [44].

For training data, the reconstruction is run on a third of all 10^6 signal MC events of each decay mode, and it is run on one stream each of charged B MC (B^+B^-), neutral B MC ($B^0\,\overline{B}^0$), charm continuum MC ($c\overline{c}$), and $(u,\ d,\ s)$ continuum MC ($u\overline{u}, d\overline{d}, s\overline{s}$). A stream refers to a dataset equal in size to the full Belle dataset.

Two BDTs are trained for each decay mode. The first BDT is trained with all the variables listed in section 7.0.1 below. These include seven variables pertaining to the event shape, the 4-momentum transfer to the $\tau^+\tau^-$ pair, the beam-constrained mass and energy difference of the signal-B meson, the center-of-mass momenta and radial impact parameters for all of the final-state leptons, the fit quality of the signal-B vertex fit, the error estimation of the transverse momentum of the signal-B meson, the angle between the two final-state leptons, the polar angle of the signal-B meson, the angle between the $K^\pm$ and signal-B mesons, the angle between the signal-B and tag-B mesons, the number of tracks and photons in the event, and the invariant-mass-squared combinations of the kaon and either final-state lepton.

The second BDT is trained on the same data with the same variables, except we include the output classifier of BDT1 in the second training vector. This second training improves the separation between signal and background by a non-negligible amount. In each training, 10 % of the data is set aside prior to training for use as independent test data.

7.0.1 Variables

When running our reconstruction, we specify a list of variables that we wish to keep for offline analysis. They include event-shape variables which describe the topology of the decay, kinematic variables for any of the particles in our decay chain, kinematic variables for the tag-B meson, and more. The following sections will briefly describe the variables we use for BDT training. Plots of each of these variables are shown in Fig. 7.2.

Event Topology

In an e^+e^- event (at $\sqrt{s} = 10.58\,\mathrm{GeV}$ in this case), whether scattering, producing heavier lepton pairs, producing light $q\bar{q}$ pairs, or producing an $\Upsilon(4\mathrm{S})$, a lot can be learned by the event shape. At the $\Upsilon(4\mathrm{S})$ resonance, there is little leftover momentum when $\Upsilon(4\mathrm{S})$ decays to $B\overline{B}$. In fact, $\sqrt{\mathrm{E}_B^2 - \mathrm{m}_B^2} = 0.325\,\frac{\mathrm{GeV}}{\mathrm{c}}$. We say these B mesons are nearly at rest. When they decay, due to the conservation of momentum, the decay products tend to go in all directions. Such an event is spherical. For resonances below the $\Upsilon(4\mathrm{S})$ resonance, there is considerable leftover momentum. The result is an event shape that is not very spherical but rather tends to make two back-to-back *jets* of particles.

A popular collection of variables that describe event shape are the Fox-Wolfram moments [40], which characterize event shape by expanding all track combinations in a basis of Legendre polynomials. The Fox-Wolfram moments are defined as

$$\mathrm{H}_\ell = \sum_{i,j} \frac{|\vec{\mathrm{p}_i}||\vec{\mathrm{p}_j}|}{s} \mathrm{P}_\ell(\cos\,\phi_{i,j})\,, \tag{7.1}$$

for all track pairs (i, j) with the angle between them $\phi_{i,j}$ and center-of-mass energy for the e^+e^- pair equal to $\sqrt{s}$. Taking it a step further, the Kakuno-Super-Fox-Wolfram (KSFW) moments [45] are a version of this in which the tracks are treated separately according to their association with either the signal B or the tag-B meson. These associations are denoted as s (signal) or o (other) respectively. Further, the KSFW moments use the event-missing-momentum vector as an additional category in the calculation. We denote the KSFW moments as $\mathrm{H}^{ab}_{\ell_a \ell_b}$, where $ab = oo$ or so and ℓ is the order of the Legendre polynomial for each. For training our BDT to classify based on overall event shape, we use seven of the possible 16 KSFW moments which we chose according to their relative significance as determined by the `fastBDT` software.

Because the B mesons are not exactly at rest, inferring the direction of their momentum can be useful. The thrust axis of a group of particles with a common origin is the direction in which the particles collectively have the largest momentum [46]. If thrust is given by

$$\vec{\mathrm{T}} = \sum_i |\vec{\mathrm{p}_i} \cdot \hat{\mathrm{n}}| \hat{\mathrm{n}} \,, \tag{7.2}$$

then the thrust axis is the $\hat{\mathrm{n}}$ for which $|\vec{\mathrm{T}}|$ is maximized. As our decay mode involves potentially a lot of missing momentum from the two tau decays, relations involving the thrust axis of the signal B decay can help differentiate our signal vs. other typical B decays. For this reason, we include in our training the cosine of the angle between the signal-B thrust axis and the beam axis, $\cos(\angle \vec{\mathrm{T}}_B \hat{\mathrm{z}})$, and the cosine of the angle between the signal-B thrust axis and the tag-B thrust axis, $\cos(\angle \vec{\mathrm{T}}_B \vec{\mathrm{T}}_O)$.

To help our BDT classify our signal against events with back-to-back leptons, like $J/\Psi \to \ell^+\ell^-$, we include the angle between the two final-state leptons in the final-state dilepton rest frame, $(\angle \ell\ell)_{\ell\ell-\mathrm{frame}}$. Last, because of kinematic relations between our signal kaon and signal-B meson, we include the cosine of the angle between the signal kaon and the recoil direction of the reconstructed tag-B meson, $\cos\left(\angle \vec{\mathrm{R}}_{\mathrm{tag}} \vec{K}\right)$, where we've chosen the tag-B recoil direction as an approximation of the true direction of the signal-B meson.

Leptons

For the final-state leptons in our decay ($e^\pm$, $\mu^\mp$) we use a few variables that we found to be useful when evaluating their significance using the `fastBDT` software package. The radial distances of the lepton tracks from the IP, dr_{e^+}, dr_{μ^-}, may help distinguish a displaced vertex from a tau decay as opposed to a mode where a lepton is a direct descendant of the B meson. The magnitudes of the lepton momenta in the center-of-mass frame, $\mathrm{p}^\star_{e^+}$ and $\mathrm{p}^\star_{\mu^-}$, may be on average smaller than those originating directly from a B decay. Finally, the binary muon–pion PID variable using information from the ACC, the TOF counters, and the CDC, $\mathcal{L}^\mu_\pi = \frac{\mathcal{L}(\mu)}{\mathcal{L}(\mu)+\mathcal{L}(\pi)}$, could help distinguish our decay mode from modes involving a $D \to K\pi$ decay where the pion is misidentified as a muon.

Signal B Meson

From the signal-B meson, we use the uncertainty of the transverse momentum, δp_T, and the χ^2-probability of the vertex fit, χ_{prob}. The transverse-momentum uncertainty, presumably the quadrature sum of the δp_T errors of all three track-fit results from our signal-B reconstruction, may be of value to our BDT classifier as it collectively quantifies the quality of all our tracks simultaneously, and may guard against background candidates made up of one or more ghost tracks that are made up of hits from noise and multiple other tracks. Such a candidate could erroneously have a decay topology consistent with our decay mode. The χ^2-probability of the vertex fit can help train the BDT classifier against backgrounds from wrong-mother reconstructions where tracks from both B mesons are combined to create a signal candidate.

Tag B Meson

From the tag-B meson, we use beam-constrained mass, M_{bc}^{tag}, energy difference, ΔE^{tag}, and the number of ECL clusters associated with the tag B decay, $\mathcal{N}_{\gamma}^{tag}$. As this is an inclusive-hadronic-tagging analysis, ensuring the inclusively reconstructed tag-B meson is consistent with a true B meson also guards against wrong-mother reconstructions.

Others

Finally, we use the following four variables. The number of tracks in the whole event, $\mathcal{N}_{trk}^{evt}$, is on average smaller for our signal than for typical background events. The 4-momentum transfer to the dilepton system, q^2_{simp}, which is calculated as the 4-momentum recoiling off of the kaon in the B rest frame, has a distinctly different shape for our signal than for typical backgrounds. The subscript "simple" means the calculation is performed assuming the B is at rest in the $\Upsilon(4S)$ rest frame. The magnitude squared of summed 4-momenta for the signal kaon and oppositely-charged (equally-charged) final-state lepton, $p^2_{K^+\ell^-}$ ($p^2_{K^+\ell^+}$), has a distinctly different (somewhat different) shape between our signal and typical backgrounds.

In Fig. 7.2 we present histograms of all our training variables for our signal $K[^{\tau}e][^{\tau}\mu]$ MC, generic $B\overline{B}$ MC, and $q\bar{q}$ MC where $q = (u,\, d,\, s,\, c)$. Some of the distributions look quite similar for signal and background which indicates that their utility comes from neighborhoods of 25-dimensional phase space, something not readily apparent in one dimension over all events.

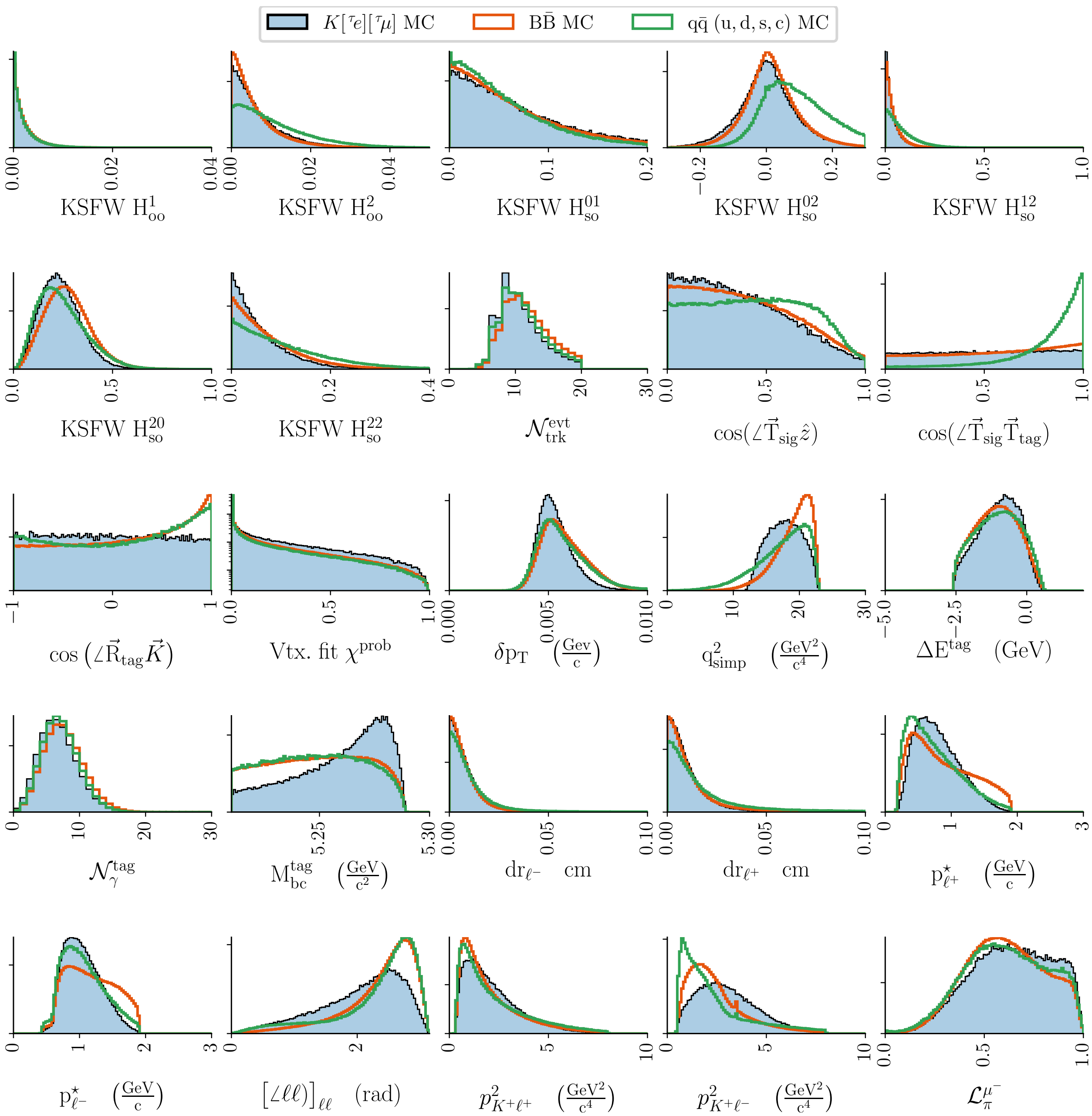

Figure 7.2: Histograms of variables used for training the BDT to identify signal vs. background in the $K[{}^{\tau}e][{}^{\tau}\mu]$ mode. Bin widths vary, but counts are normalized so that each integrates to unity. Some of the distributions look quite similar for signal and background which indicates that their utility comes from neighborhoods of 25-dimensional phase space, something not readily apparent in one dimension over all events.

7.0.2 Boosted Decision Tree Performance

We train one classifier for each of the four decay modes in this analysis. As with the reconstruction Chapter (6), we will limit our scope to the $K[\,^{\tau}e][\,^{\tau}\mu]$ mode, but plots for the other three decay modes may be found in the appendix.

To quantify BDT classifier performance, we plot a receiver operating characteristic (ROC) curve and calculate the area under the curve. A ROC curve is the true-positive rate vs. the false-positive rate. It is a standard tool for evaluating the performance of a binary classifier. A ROC curve with a slope of one is equivalent to a coin toss and has an area under the curve of 0.5. As the classifier becomes better than a coin toss, the ROC curve bends upward toward the upper-left corner. A perfect classifier has a capital Γ shape and has an area under the curve of 1.0.

The first classifier, BDT1, is trained against one stream of generic B and continuum background using the 25 variables described above. We show the train–test spectrum and the ROC curve in Fig. 7.3. The comparison between training data and independent test data does not show any signs of overtraining. We measure an area under the ROC curve of 0.777 for the independent test data.

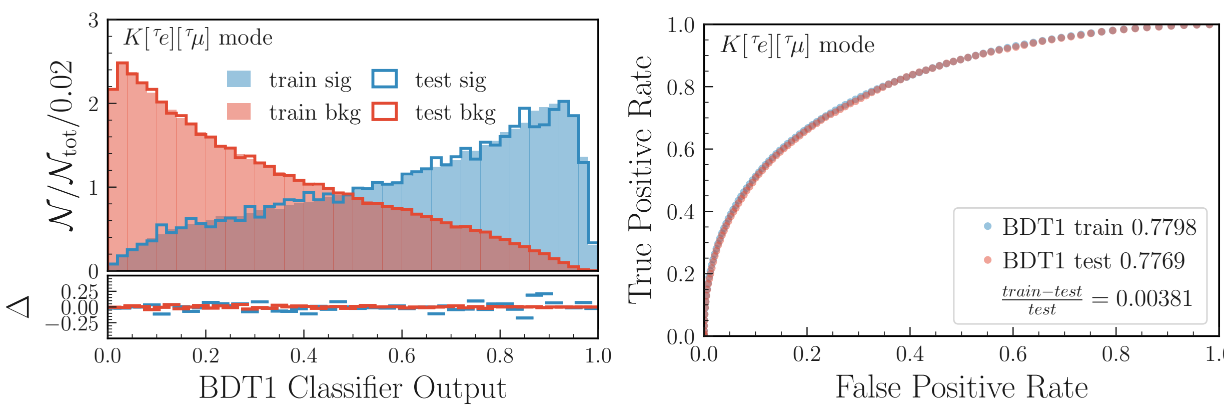

Figure 7.3: Train–test spectrum and ROC curve for BDT1 on the $K[\,^{\tau}e][\,^{\tau}\mu]$ mode. The train–test spectrum (left) shows moderate separation between signal and background. The ROC curve (right) shows an overtraining factor of 0.38 %.

The second classifier, BDT2, is trained on the same data using the 25 variables described above plus the output classifier of BDT1. We choose this approach because the presence of BDT1 in the BDT2 training vector heavily influences the first decision tree, thus encouraging the subsequent trees to explore neighborhoods of phase space that are even more signal-like. Again, we do not see any signs of overtraining (Fig. 7.4). Including the first BDT in the training vector improves the area under the receiver-operator curve by 5.8 %. When comparing the train–test spectra of the two BDTs, the improvement in signal–background separation is evident.

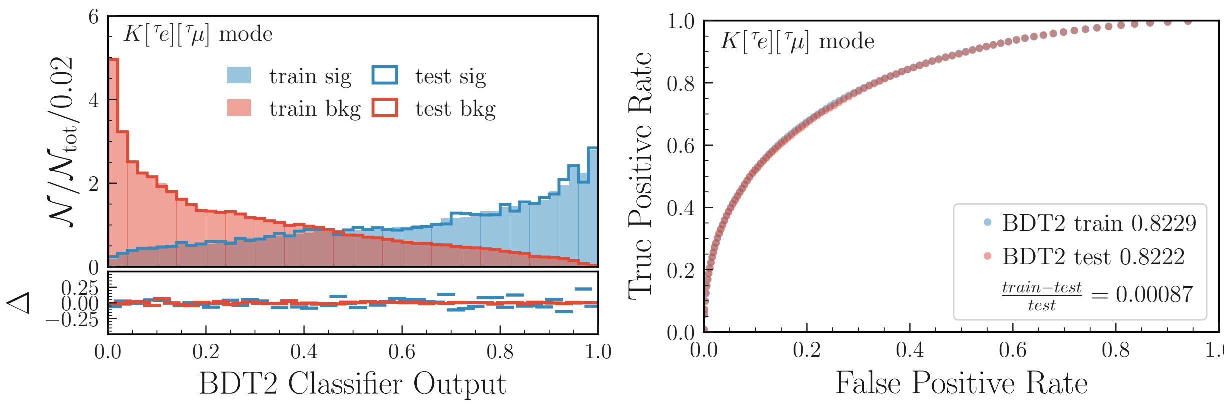

Figure 7.4: Train–test spectrum and ROC curve for BDT2 on the $K[\,^\tau e][\,^\tau\mu]$ mode, trained on the same variables and same data but with BDT1 included in the training variables. The train–test spectrum has visibly improved compared to BDT1 (Fig. 7.3). The overtraining factor is still negligible, and the area under the ROC curve has improved by 5.8 %

CHAPTER 8
FITTING

8.1 Constructing Model Histograms

One of the most salient features of $B^+ \to K^+\tau^+\tau^-$ is the missing mass in the final state, especially with four missing neutrinos in our chosen τ decay modes. To that end, we quantify our signal component by performing a 2D binned maximum-likelihood estimate using the variables missing-mass squared ($\mathrm{m}^2_{\mathrm{miss}}$) and BDT2. Using the notation where B, K, e, and μ represent 4-momenta,

$$\mathrm{m_{miss}}^2 = |B - K - e - \mu|^2, \tag{8.1}$$

where the B is calculated from the beam energy and nominal B mass ($\mathrm{E}_B = \sqrt{s}/2$, $|\mathrm{p_B}| = \sqrt{s}/2 - \mathrm{m}_B$), while the direction of the B momentum is fixed to be opposite that of the tag-B meson in the center-of-mass frame. The distributions of these two variables by themselves are shown in Fig. 8.1 . The 2D distribution for signal and the most prominent background types are shown in Fig. 8.2 . We start with fine binning in the BDT2 (vertical) direction and coarse binning in the $\mathrm{m}^2_{\mathrm{miss}}$ (horizontal) direction to best exploit the shape differences between components without creating such fine bins that the statistics become poor.

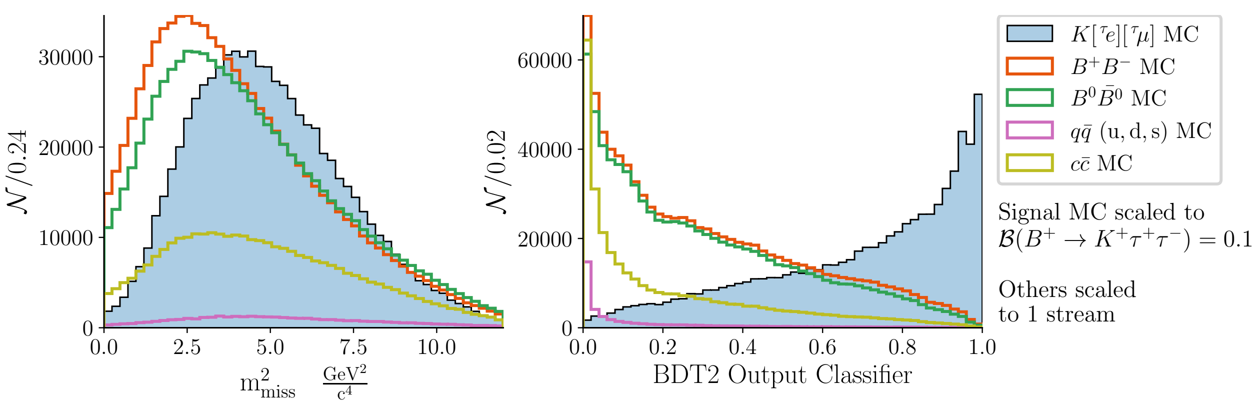

Figure 8.1: Missing-mass squared and BDT2 distributions for the $K[^\tau e][^\tau\mu]$ mode. On the left, the signal missing-mass squared (solid blue) has a different shape and is shifted right as compared to the backgrounds. On the right, we now see the BDT2 distributions broken up according to each background component. The largest background components are from the neutral and charged $B\overline{B}$ components.

We also reconstruct all the available streams of rare Monte Carlo data. These are then normalized to one stream and plotted in Fig. 8.3 . We find that the contribution from rare MC is negligible.

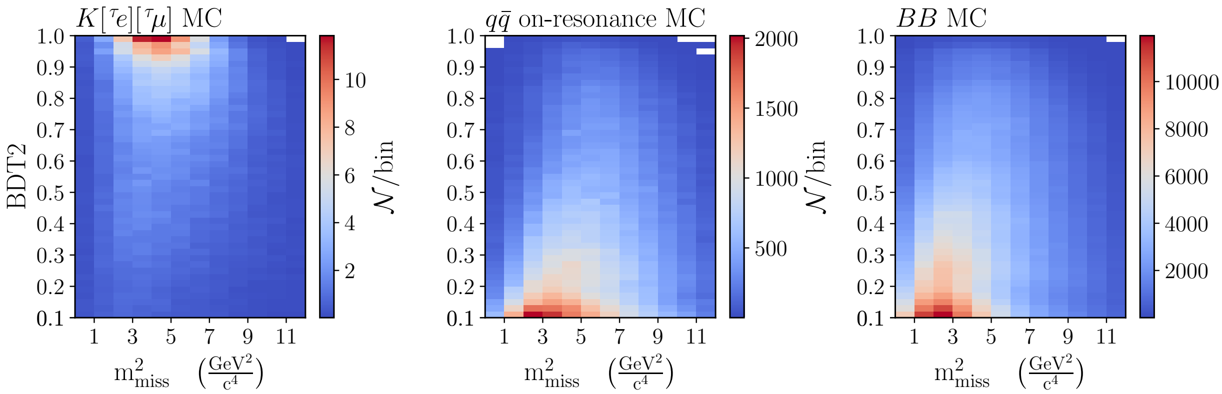

Figure 8.2: BDT2 vs. missing-mass squared distributions for signal and prominent background components in the $K[^{\tau}e][^{\tau}\mu]$ reconstruction.

Before proceeding to the fit, we check for peaking backgrounds and consider an alternate binning strategy. To check for peaking backgrounds, we plot the three possible invariant mass squared distributions for pairs of our final state particles (Fig. 8.4). We do see a resonance at the D mass in the $K^{\pm}\mu^{\mp}$ invariant mass squared distribution, primarily in the B^+B^- and $B^0\,\overline{B}^0$ backgrounds, and to a lesser extent in the $c\bar{c}$ background. This is due to pions from $D^0 \to K^-\pi^+$ decays that are misidentified as muons.

To check the effect that this peaking background has on our fit, we isolate the entries under the peak and those entries just left and right of the peak. We histogram both of these samples using the same 2D binning as above then subtract the latter histogram from the former. What remains is a representation of just the contribution from $D^0 \to K^-\pi^+$ decays (Fig. 8.5). We find that this distribution is concentrated at the lower end of the missing mass squared range and does not contaminate our signal region.

As it is, our 2D histogram has 540 bins. When combining all four decay modes into a simultaneous fit, we have 2160 bins. While this many bins can be fit using a maximum-likelihood estimate, the processing time can be problematic, especially if bin-wise systematics are included. The purpose of using such wide ranges over our fitting variables is to provide good control over the normalization of the background components in the fit. To solve this problem, we rebin the data. We want fine bins over the region where the signal is most significant. Coarse bins in the remaining region will suffice so long they provide good separation power between the two background components. To achieve this, the bins are first sorted by signal significance, $\mathcal{N}_{\mathrm{sig}}/\sqrt{\mathcal{N}_{\mathrm{tot}}}$, and the top 40 are used as fine bins. Next, the remaining ~ 500 bins are sorted by $q\bar{q}$ significance, $\mathcal{N}_{q\bar{q}}/\sqrt{\mathcal{N}_{\mathrm{tot}}}$, and are combined to create 30 roughly equal-sized coarse bins. This binning strategy reduces the total number of bins by a factor of 31. This does, however, result in some of the coarse bins being built up from disjointed regions of the BDT2 $\times$ $\mathrm{m}^2_{\mathrm{miss}}$ phase space (Fig. 8.6). These are the model histograms for the three components (Signal, $q\bar{q}$, $B\overline{B}$) of our 2D binned maximum-likelihood estimate. The other

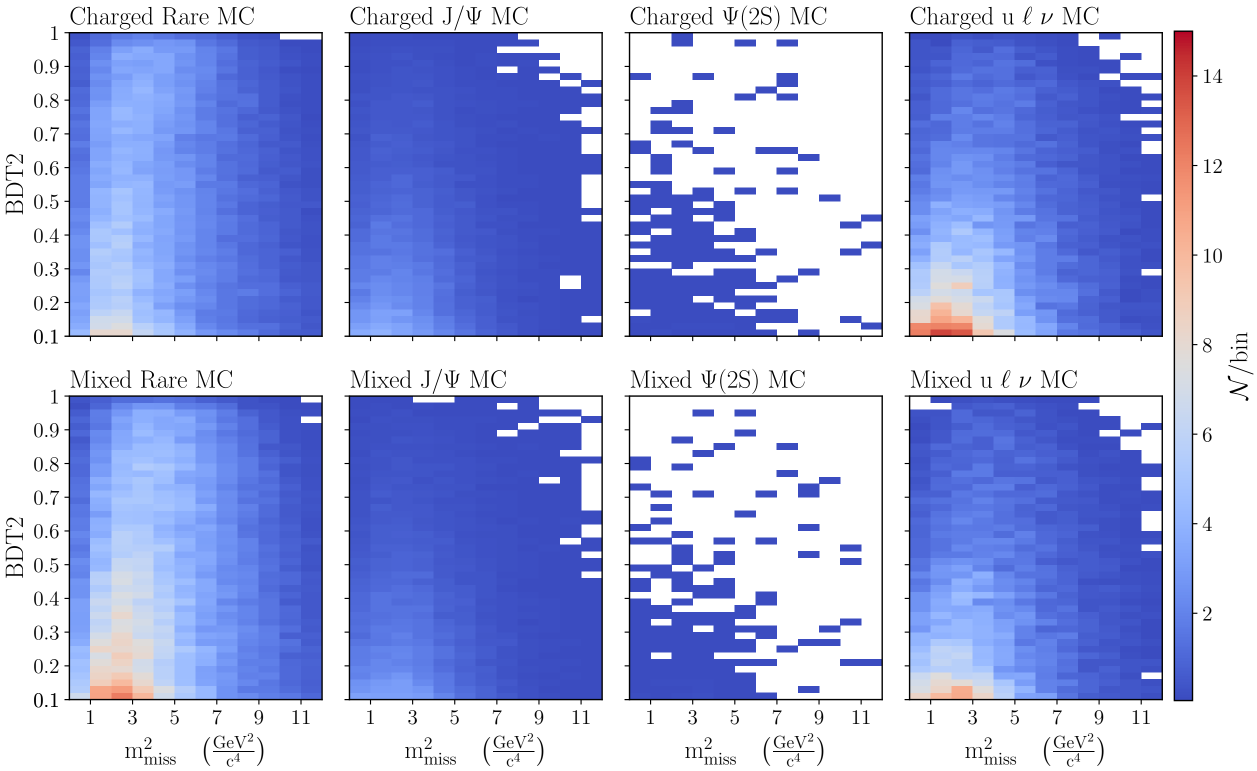

Figure 8.3: BDT2 vs. missing-mass squared distributions for rare components in the $K[{}^{\tau}e][{}^{\tau}\mu]$ reconstruction.

three decay modes are treated identically.

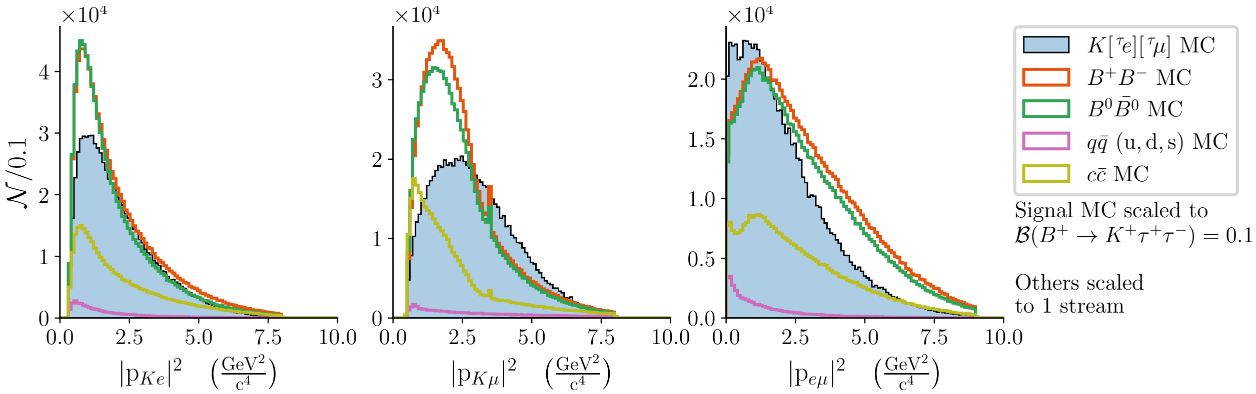

Figure 8.4: Invariant mass squared combinations for the $K[^\tau e][^\tau\mu]$ mode. These have been plotted to check for peaking backgrounds. There is a resonance at the D mass in the $K^\pm\mu^\mp$ distribution, primarily in the B^+B^- and $B^0\overline{B}^0$ backgrounds (middle panel). The same resonance exists for the $K[^\tau\mu][^\tau\mu]$ mode. The other two modes do not have this resonance.

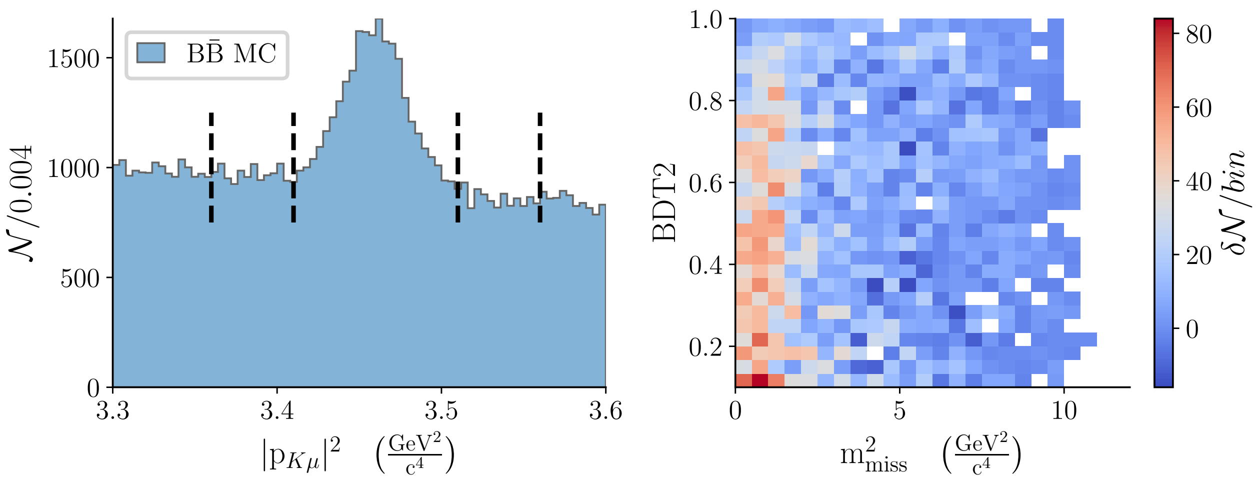

Figure 8.5: Contribution from the $D^0 \to K^-\pi^+$ peak in the $K[^\tau e][^\tau\mu]$ reconstruction. The histogram on the left is zoomed in on the D^0 mass peak seen in Fig. 8.4. The dashed lines indicate the peaking region and two sidebands with half the width of the peaking region. The 2D histogram on the right is made from entries from the peaking region of the histogram on the left minus the entries from the sideband regions. It is a representation of the contribution due to the D^0 mass peak. Because this 2D histogram contributes mostly to lower BDT2 values and low missing-mass squared values we do not consider it a problem for our fit.

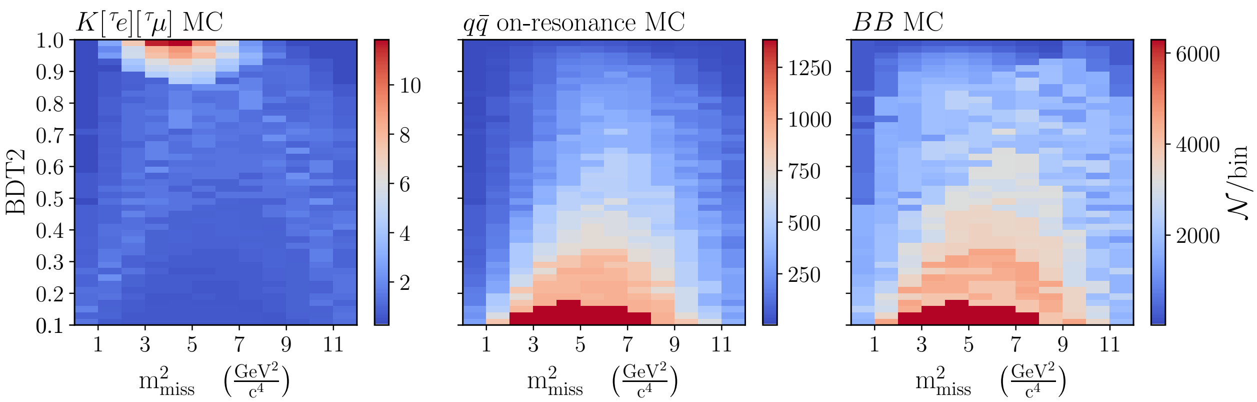

Figure 8.6: Model histograms for the $K[^\tau e][^\tau\mu]$ mode after rebinning. The rebinning procedure ranked the fine bins from Fig. 8.2 according to signal significance. The top 40 fine bins are kept as is, and the remaining were combined into 30 regions, forming coarse bins. The coarse-bin regions are not necessarily connected. There are 70 bins total here, down from 540 before rebinning. These model histograms as well as those from the other three decay modes make up the overall model that is provided to the fitter.

8.2 Maximum-Likelihood Estimate

The Poisson probability of observing some number of events (n) in a single bin is given by

$$P(n, s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)}\,, \tag{8.2}$$

where the expectation value $\lambda = (s + b)$ is characterized by the expected number of signal events, s, and expected number of background events, b. The likelihood function for many (N) bins is a product of Poisson probabilities with the free parameter, μ, for the signal strength:

$$L(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^n_j}{n_j!} e^{-(\mu s_j + b_j)}\,. \tag{8.3}$$

Here, θ is a vector of nuisance parameters for the expected number of s and b in each of the bins. The signal strength is a multiplicative parameter for the amount of fitted signal in the data. It is reported as a multiple of the signal model histograms, $\mathbf{s}$. If auxiliary measurements are made to constrain, say, the background normalization or perhaps the shape of the signal and background, then additional Poisson products may be tacked on to $L(\mu, \theta)$. Because we do not perform any such auxiliary measurements, we can exclude this term from our description without any loss of generality. The best fit is given by the values of $\hat{\mu}$, $\hat{\theta}$, that maximize $L(\hat{\mu}, \hat{\theta})$. It can, of course, be obtained by setting the first derivative to zero:

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{N} \left(\frac{n_i}{(\hat{\mu} s_i + b_i)} - 1 \right) \frac{\partial(\hat{\mu} s_i + b_i)}{\partial \theta_j} = 0\,. \tag{8.4}$$

In the limit of large sample sizes, the inverse covariance matrix can be calculated as

$$V_{ij}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right]\,, \tag{8.5}$$

which can then be inverted to find standard deviations on the best-fit values [47].

Our goal is to construct a maximum likelihood estimate for fitting data to our model histograms. The parameter of interest in this fit is the signal strength, μ. We aim to demonstrate how small of a value of the signal strength we can reliably detect using our model. We use the `cabinetry` application to construct our likelihood which is then fit using a maximum-likelihood estimate by the `pyhf` [48, 49] software package. The package `pyhf` is a `python` implementation of `root`'s Histogram Factory. We construct a fit model with four channels (i.e., our four decay modes) and three components for each channel. The three components are signal, continuum background ($q\overline{q}$), and generic B background ($B\overline{B}$). Even though we simulated 4×10^6 signal events, we construct our signal model with 1/3 of these and normalize it to $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) = 1 \times 10^{-4}$. The background models each correspond to one stream of MC data, i.e., $1\times$ the full Belle dataset. The fitter has nine free parameters: signal normalization as the parameter of interest, and 8 nuisance parameters for the normalization modifiers relating to the two background components on each of the four channels. The normalization modifiers used for the background components are Gaussian constraints in the likelihood specification that use a nominal histogram as well as $\pm 1\,\sigma$ histograms as inputs. We choose $\pm\sqrt{\mathcal{N}_{bin}}$ for our $\pm 1\,\sigma$ histograms to account for statistical variations.

8.3 Asimov Fit and Toy Studies

We begin with a fit of the Asimov dataset. The Asimov dataset is a representative dataset defined such that when using it to estimate parameters using a maximum-likelihood estimate, the estimated parameters are equal to the true parameters [47]. If the Asimov data set is used, Eq. 8.4 is still true, which means that the inverse covariance matrix (Eq. 8.5) can be calculated from a likelihood constructed with the Asimov dataset. Therefore, we perform a fit on the Asimov dataset to measure the standard deviations of our fit parameters.

In our case, the Asimov dataset is the sum of the model histograms, and the Asimov fit is a fit of our model to the Asimov dataset. We summarize the fit results to two decimal places in Table 8.1 . The measured signal strength in the Asimov fit is 1.00 ± 0.51, which tells us that our

Table 8.1: Fit results for Asimov fit. Signal is a normalization factor while the others are reporting the number of standard deviations away from nominal given a nominal plus high/low variation in the model.

Mode	Component	Best Fit	Uncertainty
ee	$q\overline{q}$	0.00	0.76
ee	$B\overline{B}$	0.00	0.32
$\mu\mu$	$q\overline{q}$	0.00	0.82
$\mu\mu$	$B\overline{B}$	0.00	0.31
$e\mu$	$q\overline{q}$	0.00	0.79
$e\mu$	$B\overline{B}$	0.00	0.31
μe	$q\overline{q}$	0.00	0.82
μe	$B\overline{B}$	0.00	0.29
Combined	Signal	1.00	0.51

model has a resolving power of about 5×10^{-5}.

The correlation matrix is shown in Fig. 8.7 . We see a large anticorrelation between our two background components in each channel. This anticorrelation is not concerning as long as the total background is estimated properly. That they are not fully correlated or anticorrelated tells us there is still merit to fitting them as separate components. We also see moderate anticorrelation between the $B\overline{B}$ background and signal in each channel. This anticorrelation is somewhat concerning because an underestimation of the four $B\overline{B}$ components would lead to an overestimation

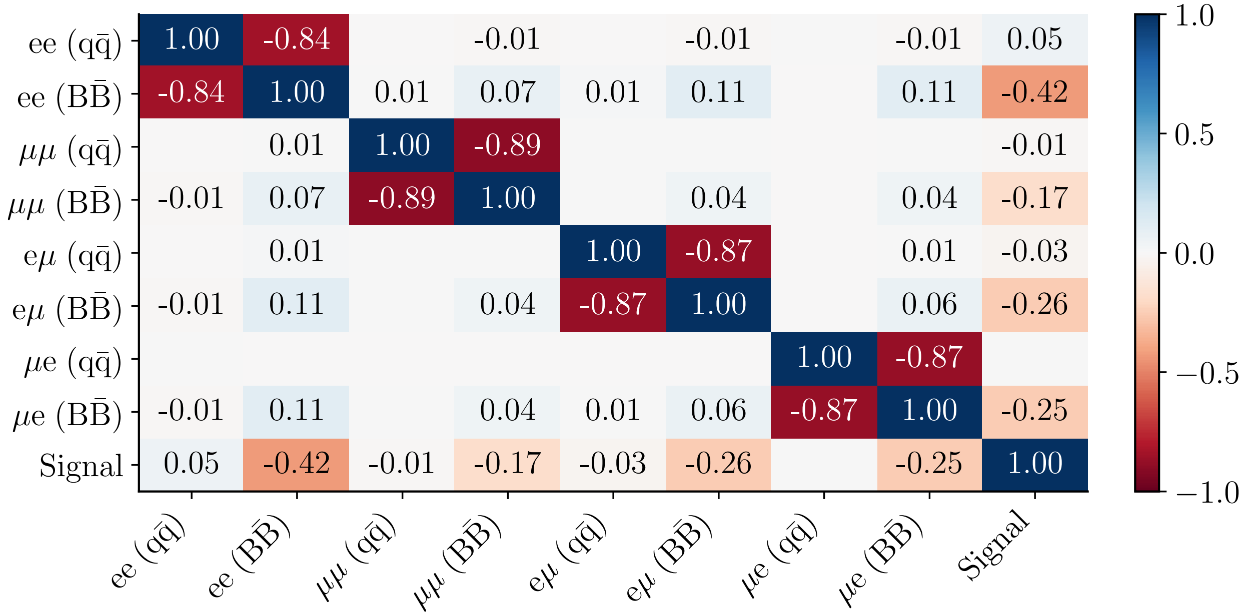

Figure 8.7: Correlation matrix for our maximum likelihood estimate. Empty entries are consistent with zero within two decimals. The two background components are very anticorrelated, but that they are not fully anticorrelated tells us that there is still merit to fitting them as separate components. The moderate anticorrelation between the $B\overline{B}$ background and the signal in each channel will contribute to the standard deviation of the signal strength estimate.

of our signal component, and vice versa. However, because the correlation matrix is related to the covariance matrix, this possible misestimation of the signal component is accounted for in the standard deviation.

To ensure our fitter finds a maximum-likelihood estimate that is a global maximum and not just a local maximum, we test different hypotheses for the signal strength, μ. To test a signal-strength hypothesis, a likelihood profile is used:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \,. \tag{8.6}$$

The denominator is the maximum likelihood, and the $\hat{\hat{\theta}}$ in the numerator represents the values of the nuisance parameters that maximize $L(\mu, \hat{\hat{\theta}})$. Standard practice is to consider $-2\log(\lambda(\mu))$ because it becomes a χ^2 distribution in the limit of large sample sizes. This equation has computational benefits as well because the ratio cancels out the factorials in $L(\mu, \theta)$, thus also allowing for the use of fractional bin counts, and the logarithm turns the products into summations. Note we can avoid writing $\lambda(\mu)$ altogether and write $-2\Delta\log(L)$. In Figure 8.8 we present a scan of $-2\Delta\log(L)$ over different hypotheses of μ both near the best-fit value and far from it. We see that there is

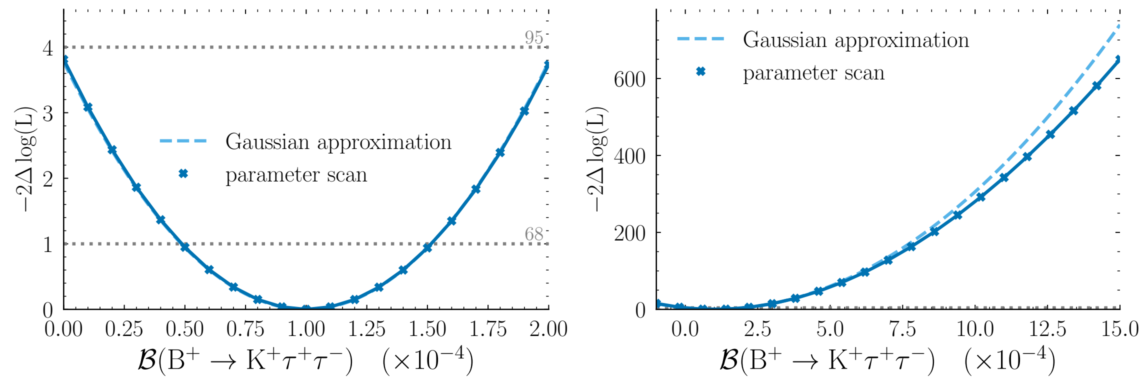

Figure 8.8: Likelihood profile scans over the signal-strength parameter. Near the best-fit value (left) the profile crosses the $1\,\sigma$ and $2\,\sigma$ lines at values consistent with the standard deviation reported by our fitter. Over the whole allowed range that we specified to our fitter (right), we see no signs of other minima which indicates that our fitter converged on the correct value.

no flat region or non-global minima over the region of allowed signal strengths in our model (-1 to 15), and when zoomed in around the best-fit value of the Asimov fit we see that the Gaussian approximation is in good agreement with the actual scan.

We test our fitter using toy Monte Carlo datasets. A toy is generated by drawing random numbers for each bin in the model histograms from Poisson distributions with expectation values for each bin equal to the bin count in the model. This creates fluctuations bin by bin based on Poisson statistics. We check the behavior of our fitter by generating 10,000 toys, each with the contribution from the signal model varying between branching fractions of $B^+ \to K^+\tau^+\tau^-$ from zero up to 10^{-3}. We plot the fit results of each along with error bars and pull distributions in Fig. 8.9. The pull distribution is fitted to a Gaussian using a least-squares fit. The mean pull being consistent with zero and width being consistent with one indicate that our fitter is providing accurate error estimates. We perform this test to make sure our model provides consistent results regardless of how much signal is present in the dataset. Our fitter must not be biased toward overestimating or underestimating signal content at any level. Our pull distribution tells us that the output vs. input has a slope of one and that the standard deviations estimated by the fitter correctly model the expected statistical fluctuations.

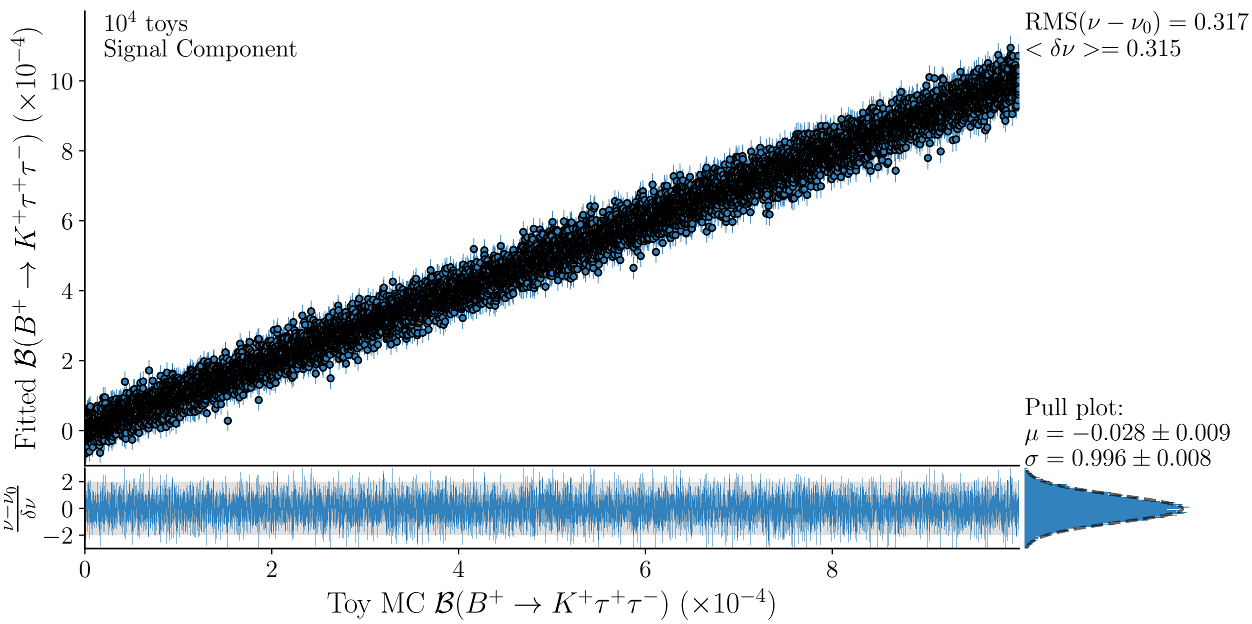

Figure 8.9: Linearity test on 10,000 toy fits with varying amounts of signal in the test data. To check linearity the deviation of each fit from a line with a slope of one, normalized using the uncertainty provided by the fitter is plotted on the bottom. These values populate the histogram on the right, which is the pull distribution. With a mean consistent with zero and a width consistent with one, the pull distribution indicates the fitter is providing accurate error estimates.

CHAPTER 9
UNCERTAINTIES

Any particle physics analysis must take a careful accounting of the uncertainties involved for the results to be valid. Systematic uncertainties arise from inaccuracies in measurement devices, which can include things like resolution on time and position of a hit, but also include things like particle identification and Monte Carlo modeling (which always has some level of disagreement with data). Statistical uncertainties arise from the quantity of collected data, which includes the quantity of simulated data. As this analysis is a proof of concept using MC data and not a measurement of collected data, a full accounting of the involved uncertainties is not required. However, we will look at a few of them to get an idea of their impact.

9.1 Monte Carlo Sample Size

The 4×10^6 signal MC events we generated amount to over 400 times our sensitivity target of 1×10^{-4}, so when creating our signal model histograms, we scale down all the bin contents by the appropriate amount. To estimate the contribution to the total statistical uncertainty due to the finite size of the generated signal MC data, we generate 10,000 toys with random Poisson fluctuations on the full signal MC prior to scaling down. We keep our model histograms fixed and create toy fit datasets that are the sum of these scaled-down toys as well as our Asimov $B\overline{B}$ and $q\bar{q}$ datasets.

The result, shown in Fig. 9.1, is a signal strength variance of 6.1×10^{-7}. Assuming $\Upsilon(4S)$ decays to charged B mesons 50 % of the time, this corresponds to an uncertainty of 470 signal events for the full Belle dataset. Compared to the theoretical branching fraction, $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) = 1.26 \times 10^{-7}$, which predicts about 97 events, this is a lot. However, compared to our model's signal strength uncertainty of 5.1×10^{-5}, it is only 1.2%.

Similarly, we check the dependence on the amount of $B\overline{B}$ and $q\bar{q}$ Monte Carlo by varying each independently and combining them with Asimov datasets of the unvaried components (Fig. 9.2). The result is a 4.72×10^{-5} $B\overline{B}$ MC statistical uncertainty and 1.65×10^{-5} $q\bar{q}$ MC statistical uncertainty. The former is nearly equivalent to the signal strength uncertainty itself. This indicates that the analysis is statistically limited by the size of the $B\overline{B}$ MC dataset. This can be reduced by combining all 10 streams of the available $B\overline{B}$ MC rather than using just one.

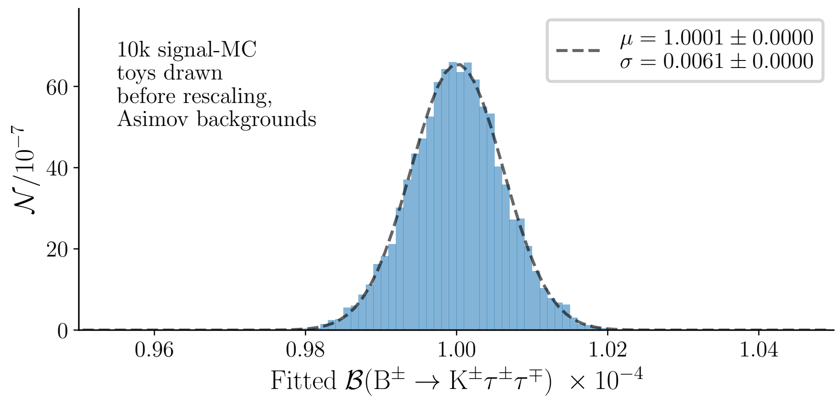

Figure 9.1: Measurement of signal MC statistical uncertainty. As signal MC histograms are scaled down by more than 100 to make the model histograms, this measurement was performed by statistically fluctuating the full signal MC histograms prior to scaling down. In each of the 10k toys, the two background components are represented by their Asimov datasets. The uncertainty of 6.1×10^{-7} equates to 1.2% of our signal-strength uncertainty.

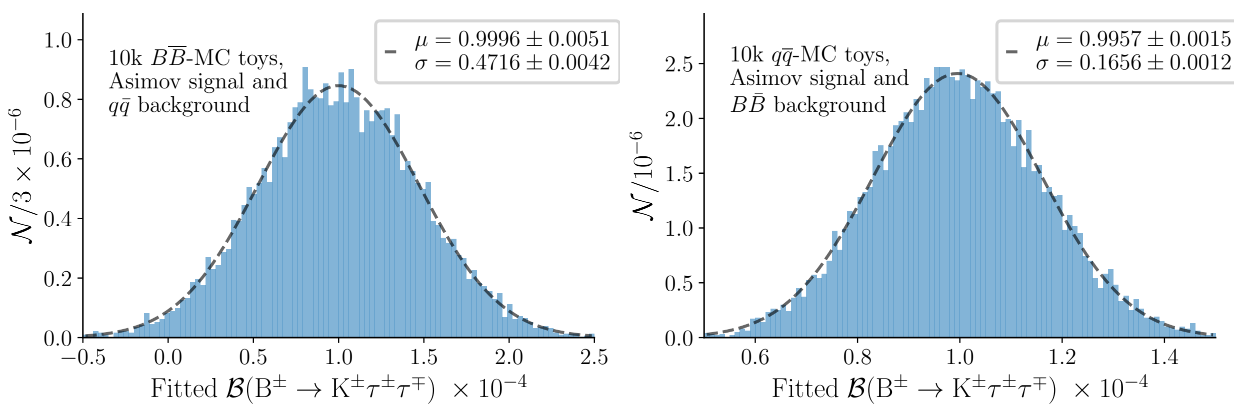

Figure 9.2: Measurement of background MC statistical uncertainty. Left, toys are drawn from $B\overline{B}$ MC 10k times while signal and $q\bar{q}$ MC are represented by their Asimov datasets. The uncertainty of 4.72×10^{-5} is of the same order as our signal-strength uncertainty. Right, toys are drawn from $q\bar{q}$ MC 10k times while signal and $B\overline{B}$ MC are represented by their Asimov datasets. The uncertainty of 1.65×10^{-5} is also of the same order as our signal-strength uncertainty, but only 1/3 as much.

9.2 Reconstruction Efficiency

The Belle detector does not and can not cover the full solid angle due to the need for beam pipes, final focus magnets, the steel mechanical structure of the detector, and electronics to some extent. As such, some events will not be contained in the active solid angle of the detector. Further, reconstruction efficiency of our signal MC is not perfect even when all the tracks are contained in the active solid angle of the detector. Together, these two contribute to the total systematic error. Fortunately, reconstruction efficiency automatically includes contributions solely from detector acceptance, so all we have to do is calculate the binomial error for each of our four decay modes:

$$\delta\epsilon = \sqrt{\frac{\epsilon_{\text{matched}}(1-\epsilon_{\text{matched}})}{\mathcal{N}_{\text{generated}}}}\,. \tag{9.1}$$

The results are presented in Table 9.1. Adding these four binomial errors in quadrature we find a systematic uncertainty due to reconstruction efficiency $\delta\epsilon_s = 0.06\,\%$. Compared to our statistical uncertainties discussed above, this is a rather small effect.

Table 9.1: Binomial error estimation for reconstruction efficiency.

Decay Mode	$\epsilon_{\text{matched}}$ %	$\delta\epsilon$ %
$K[^{\tau}e][^{\tau}e]$	18.24	0.039
$K[^{\tau}\mu][^{\tau}\mu]$	4.10	0.020
$K[^{\tau}e][^{\tau}\mu]$	8.98	0.029
$K[^{\tau}\mu][^{\tau}e]$	9.06	0.029
Total	10.10	0.059

9.3 Number of B^+B^-

The Belle experiment recorded $(771.581\pm10.566)\times10^6$ $B\overline{B}$ pairs [50]. Also, the branching fraction of $\Upsilon(4S)$ to charged B pairs is 0.514 ± 0.006 [9]. Multiplying the two and propagating the uncertainties gives

$$\delta\mathcal{N}_{B^+B^-} = (\mathcal{N}_{B\overline{B}}\cdot\delta\mathcal{B}_{B^+B^-})^2 + (\delta\mathcal{N}_{B\overline{B}}\cdot\mathcal{B}_{B^+B^-})^2 = 7.14\times10^6\,. \tag{9.2}$$

In terms of a fraction, we have the expected number of charged B pairs, $\mathcal{N}_{B^+B^-} = \mathcal{B}_{B^+B^-}\cdot\mathcal{N}_{B\overline{B}} = 396.593\times10^6$, and $\delta\mathcal{N}_{B^+B^-}/\mathcal{N}_{B^+B^-}$ comes out to $1.80\,\%$. Even though we reconstruct modes of both $B^{\pm}$ charges, there is no need to multiply this figure by two because it would enter in both the numerator and denominator of $\delta\mathcal{N}_{B^+B^-}/\mathcal{N}_{B^+B^-}$. This constitutes another systematic uncertainty as it pertains to the accuracy of our apparatus, not to the size of the dataset.

9.4 Uncertainties Summary

We have computed MC statistical errors for each of our fit components in terms of a branching-fraction uncertainty and systematic errors as a percent. At this point we cannot state statistical plus systematic errors in the same units because we have not stated a branching fraction or upper limit. Nevertheless, we can tabulate the uncertaintes all as percents using our target sensitivity we've been considering so far of 1×10^{-4}. We present a summary of our brief error estimates in Table. 9.2.

Table 9.2: Error summary. Our brief look at uncertainties includes three sources of MC statistical uncertainty, systematic uncertainty due to the recorded number of $B\overline{B}$ pairs by the Belle experiment, and systematic uncertainty due to reconstruction efficiency, which includes acceptance due to the solid-angle coverage of the detector.

Source	Error
$\mathcal{N}_{\text{Signal}}$ MC	0.61 %
$\mathcal{N}_{B\overline{B}}$ MC	47.0 %
$\mathcal{N}_{q\bar{q}}$ MC	17.0 %
$\mathcal{N}_{B^+B^-}$ Belle	1.8 %
Reconstruction Efficiency	0.06 %

The largest contribution is 47% error due to $\mathcal{N}_{B\overline{B}}$ MC statistics followed by 17% due to $\mathcal{N}_{q\bar{q}}$ MC statistics. This is a clear indication that any future work on this analysis would benefit from building the model histograms from all available MC data, which includes 10 full streams of $B\overline{B}$ MC and six full streams of $q\bar{q}$ MC. If MC statistical uncertainty is improved, we anticipate that systematic uncertainty due to lepton identification would be the largest contribution. This is by no means a comprehensive list of uncertainties. Nevertheless, what we have presented here amounts to an error of $\pm$ 50 % (stat) $\pm$ 1.8 % (sys).

CHAPTER 10
PROSPECTS AND DISCUSSION FROM PART I

We have established a method of inclusive hadronic tagging for the measurement of $B^+ \to K^+\tau^+\tau^-$. We selected our inclusive tag-B mesons to retain as much of our signal B decays as possible and to require that the tag B have values of ΔE and $\mathrm{M_{bc}}$ that are mostly consistent with fully-hadronic B decays. The choice of inclusive tagging was motivated by the $B^+ \to K^+\nu\bar{\nu}$ analysis, and the inclusive hadronic tagging was motivated by the idea of signal quantification by fitting missing mass squared, wherein we use the direction opposite to the tag-B meson to assign the direction of the signal-B meson in the missing mass squared calculation. The latter assignment of signal-B direction would not be a logical choice if the tag-B were to frequently decay semileptonically.

Our goal is to show that this approach can improve the sensitivity for measurement of $B^+ \to K^+\tau^+\tau^-$ over previous searches. Bear in mind that, while current sensitivity is orders of magnitude larger than the standard-model prediction, some theoretical physics scenarios predict a large enhancement in the branching fraction of $B^+ \to K^+\tau^+\tau^-$, so improving experimental sensitivity is a worthwhile effort. Indeed, we hope to achieve a sensitivity that can rule out some of the theoretical physics scenarios. We will conclude Part I by calculating our expected sensitivity on the upper limit of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ and comparing it to previous searches, then we will highlight a weakness of this analysis that motivates Part II.

We will calculate the expected upper limit at a confidence level of 90 % to compare directly to other searches. We determine the upper limit by performing a toy MC study on the null hypothesis, i.e., the case where zero signal is present in the data, but we fit each toy dataset to our model, which does have a signal component. The expectation for these fits is that the signal strength is zero. However, we expect a normal distribution of signal strengths around zero with some characteristic width. To say that a branching fraction has an upper limit at a 90 % confidence level means that repeated measurements of the null hypothesis will not measure a branching fraction above that upper limit 90 % of the time. In other words, our model would *falsely* predict a branching fraction higher than the upper limit only 10 % of the time — falsely in the sense that the hypothesis is null so one might think of it as a false-positive outcome.

Let us begin with a naive estimation of the upper limit based on our fit model alone. We drew 25,000 toy datasets from our full background MC model. Because there is zero signal in the toy datasets, they represent null hypothesis datasets. Each toy is fit to our model. For our upper limit estimation, we only consider entries with a fitted signal strength larger than zero — in the parlance used above, only these entries constitute false positives. We take the normalized cumulative distribution of entries with a branching fraction larger than zero and determine where the distribution crosses the 90 % line (Fig. 10.1). We find an expected upper limit of $\mathcal{B}(B^+ \to$

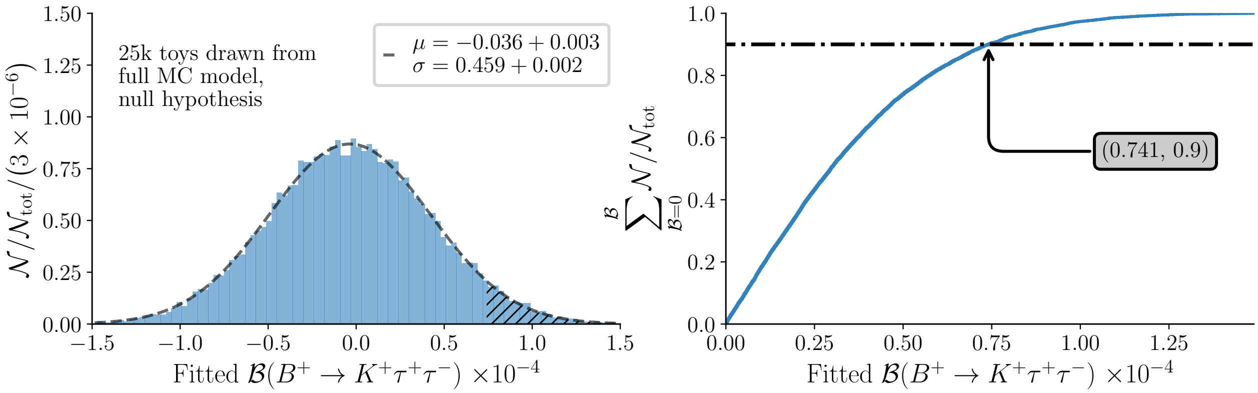

Figure 10.1: Naive upper limit estimation using 25k toys drawn from the full MC model with zero signal (null hypothesis). Left: histogram of fit results for signal strength. The fit to a Gaussian distribution shows a bias toward negative values in terms of least-squares Gaussian fit uncertainty, but the bias is small compared to the width. The width of the Gaussian is consistent with the signal-strength uncertainty of the Asimov fit. The hatched region is the 10 % tail probability for positive values of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$. Right: the cumulative distribution of the plot on the left starting from a branching fraction of zero. The intersection with 0.9 is indicated as this point corresponds to the upper limit of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ at a 90 % confidence level.

$K^+\tau^+\tau^-) < 7.41 \times 10^{-5}$ at 90 %C L .

To determine our projected upper limit based on this Monte Carlo study, we combine the sensitivity 4.59×10^{-5} from the left panel of Fig. 10.1 with our MC statistical uncertainty from Section 9.4. Adding them in quadrature, we find a sensitivity of 6.79×10^{-5}. This is one standard deviation (1σ). To go from 1σ to a 90 % confidence level, we use the inverse Gauss error function, $\sqrt{2}\,\mathrm{erf}^{-1}(0.90) \approx 1.64\sigma$. Finally we apply the systematic error from Section 9.4 as a multiplicative error and get our result. Based on this inclusive-hadronic-tagging feasibility study we project an upper limit of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 1.14 \times 10^{-4}$ at 90 % C.L.

Our Monte Carlo dataset is equivalent in size to the full Belle dataset of 711 fb^{-1}. This projected result improves upon the upper limit found in a recent PhD dissertation[8], which used hadronic tagging and the combined Belle and Belle II dataset of 1073 fb^{-1}, by over five times. Compared to the BaBar analysis on 424 fb^{-1}[7], our prospective upper limit is about 20 times lower (Fig. 10.2). Nonetheless, 1.14×10^{-4} is still nearly three orders of magnitude larger than the standard model theoretical prediction of 1.26×10^{-7}. The method of inclusive hadronic tagging shows promise, and in the new physics scenarios where a large enhancement in this mode is predicted, it may already be sensitive enough. However, to push our sensitivity close to the SM prediction, more data is needed.

One aspect of this analysis that makes $B^+ \to K^+\tau^+\tau^-$ challenging is the efficiency of muon reconstruction. Our efficiency for reconstructing signal MC in the mode with two final-state muons was only 4.1 %, and it was only 9 % for the modes with a single final-state muon. As the full

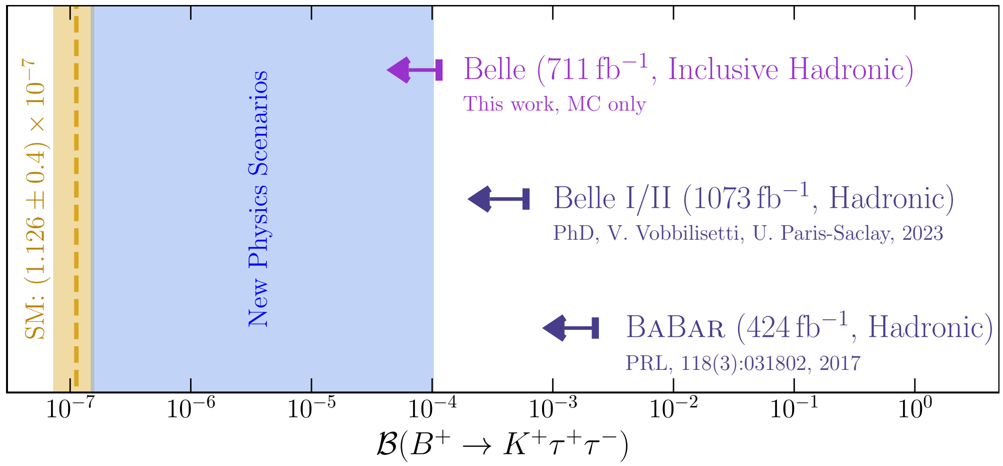

Figure 10.2: Comparison of upper limits for the branching fraction of $B^+ \to K^+\tau^+\tau^-$ at a 90 % confidence level. This work, which is a Monte Carlo study, establishes an upper limit one to two orders of magnitude lower than hadronic tagging analyses on the same mode.

Belle II dataset is expected to be far larger than the Belle dataset, it is important that Belle II is able to achieve the best muon identification efficiency possible. Investigating the muon identification system of Belle II in Part II, we identify and address some improvements that need to be made to the electronic readout system in order that Belle II has the best chance for discovering the decay $B^+ \to K^+\tau^+\tau^-$.

Part II

Part 2: Enabling Next-Generation Readout

CHAPTER 11
MUON SYSTEM UPGRADE

Following the end of Belle data taking in 2010, work began on upgrades for the next run period. The KEKB accelerator was upgraded to SuperKEKB. The positron beamline was completely replaced and a separate damping ring was added to cool down the positrons prior to injection into the main ring. The Belle detector was upgraded to the Belle II detector. With a new pixel detector, a new silicon vertex detector, a new drift chamber, the switch from aerogel Cherenkov counters to Cherenkov ring-imaging detectors, and a new readout system, Belle II would be able to have improved vertex resolution, improved particle identification, and be capable of event rates up to 30 kHz.

The K-long (K_L^0) and muon (μ) detector, or KLM, which makes up the outermost active volume of both the Belle and the Belle II detector, was also upgraded. The flux return of the detector's 1.5 T solenoidal magnetic field consists of 14 layers of 4.7 cm-thick steel plates. These plates form octagons in the Belle II barrel region and flat disks in the endcap regions. The gaps between the plates are interleaved with active particle detection modules: 15 layers in the barrel, 14 in the forward endcap, and 12 in the backward endcap. Besides serving as the flux return, the steel plates also offer more stopping power for hadrons, contributing an additional 3.9 interaction lengths in addition to the 0.8 interaction lengths of the electromagnetic calorimeter [30].

During the first-generation Belle experiment [23] (1999-2010), the KLM detector was instrumented exclusively with resistive-plate counters (RPCs) [51]. Because of the expected high neutron background when Belle II is operating at its design luminosity, the inner two barrel layers and all of the endcap layers were replaced with plastic scintillators. The scintillators used are long strips with a 1 cm $\times$ 4 cm cross section in the barrel and 0.7 cm $\times$ 4 cm in the endcaps. Strip lengths vary due to the geometry of the detector. Each scintillating strip contains a multi-clad 1.2 mm diameter wavelength shifting fiber running down its central axis, and a silicon photo-multiplier (SiPM) at one end of the fiber. Details of the scintillator and wavelength-shifting fiber selection and construction can be found in Ref. [52].

When Belle II began data taking, the SiPMs were not properly calibrated, and the electronic readout for the scintillator readout system was operating in a latched mode because firmware did not exist to read out its waveform-sampling ASICs (application-specific integrated circuit). Further complications arose in that there were no calibration sources built into the detector. Our specific contribution to these issues is the development of waveform-readout firmware to utilize the full capability of the installed electronics, including a scheme for calibrating SiPM gain without a calibration source.

The following chapters describe the electronic readout system of the KLM detector. We discuss the RPC and the scintillator readout systems separately up to the point where the two data streams are merged. We organize the remaining chapters according to the flow of information, starting with a charged particle passing through a single RPC module or scintillator bar and following the signal through the data acquisition system. Details about the calibration of the SiPMs and the waveform-sampling ASIC, as well as performance measurements form the last two chapters.

CHAPTER 12
SCINTILLATOR READOUT ELECTRONICS

Charged particles passing through the KLM generate scintillation light in the scintillator strips. Some of the light in each strip is absorbed by its central wavelength-shifting fiber. The wavelength shifter is a material that absorbs light at a higher frequency and re-emits it at a lower frequency, and it has a longer attenuation length than the plastic scintillator material. As the wavelength shifter re-emits this light isotropically, photons emitted at angles less than the critical angle of the fiber propagate to the ends of the fiber. When a photon hits one of the 667 pixels of the SiPM[1] (each pixel is an avalanche photodiode), there is a $\sim$20 % possibility that an avalanche forms. In an avalanche, one photoelectron is rapidly amplified to $\sim$750 k electrons. Multiple photons can fire multiple pixels and their outputs are combined in parallel. Current across quenching resistors in the SiPM decreases the bias voltage across the photodiode and the avalanche(s) cease. The result is a signal with a steep leading edge, a long tail, and an amplitude proportional to the number of pixels that fired.

The following sections describe the electronics encountered by a signal on its way to the Belle II data acquisition (DAQ) system [53]. Figure 12.1 depicts the multiplicity of each component in the KLM readout and how each is connected.

[1] Hamamatsu s10362-13-050c

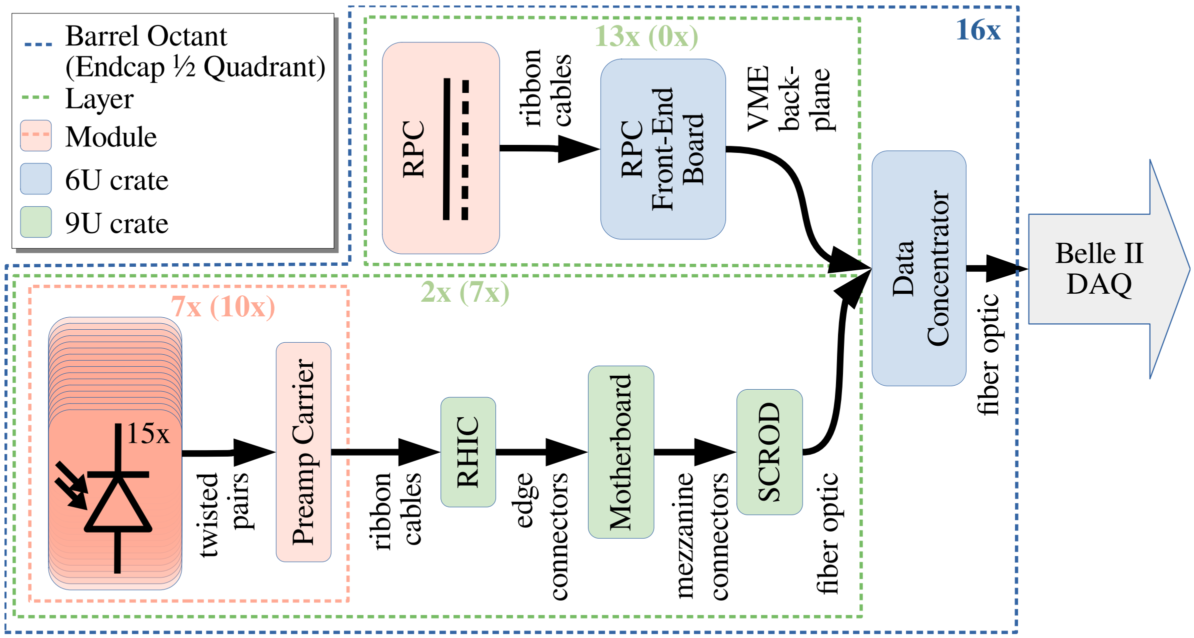

Figure 12.1: Flow diagram of KLM readout. For each layer of each octant (8 forward and 8 backward), there is one RPC or scintillator module. In each layer of each endcap quadrant, there are only scintillator modules. Each RPC module connects via ribbon cables to an RPC front-end board in a 6U VME crate, while each scintillator module connects via ribbon cables to one RHIC-and-motherboard combination in a 9U crate. In a barrel (endcap) scintillator module, there are 7 (10) preamplifier carriers and about 78 (exactly 150) SiPMs in each. In the barrel (endcap), 15 (7) front-end boards connect to one Data Concentrator.

12.1 Preamplifiers

In each scintillator module, groups of 15 SiPMs are connected to a preamplifier carrier card via twisted-pair cables, and, in turn, each carrier card is connected to the readout and control electronics located at the top of the Belle II magnet yoke via two ribbon cables several meters in length. One ribbon cable supplies a unique bias voltage to each SiPM, while the other delivers each preamplified signal to the readout system.

We use a custom-designed preamplifier that is a fully-differential operational amplifier. Each is assembled on a 2 cm by 2 cm printed circuit board that plugs edgewise into the preamplifier carrier card. The gain of the preamplifier is large enough to resolve single-photoelectron pulses from a SiPM. This was an oversight, perhaps, as even minimum-ionizing particles traversing the KLM detector tend to saturate the preamplifier. While this large preamplifier gain enables measurement of SiPM gain by using single-photon spectra, which we describe later, the saturation prevents us from resolving the full pulse height of some hits during Belle II operation. Further, we cannot measure the leading-edge time of a pulse by using a dynamic threshold that is set relative to the pulse height. Rather, we opt to measure leading-edge time using a fixed threshold. The preamplifier input from a SiPM is indicated in Fig. 12.2 in the following section.

12.2 Ribbon Header Interface Card

All of the ribbon cables from a single scintillator module are connected to a single set of readout and control electronics via the Ribbon Header Interface Card (RHIC). This circuit board connects all of the preamplified signals to the scintillator motherboard, and it also has two octal 8-bit 5 V digital-to-analog converters[2] (DACs) for each group of 15 SiPMs to fine-tune the SiPM bias voltage with a precision of 20 mV. We refer to these as the HV-trim DACs.

Each DAC output is buffered through a charge-sensitive amplifier[3] — the DAC drives the (−) terminal, the (+) terminal terminates the SiPM, and the amplifier output serves as a current monitor (Fig. 12.2). Each RHIC also contains a 2-pin high-voltage (HV) connector and distributes the HV to all of the preamplifier carriers over the ribbon cables.

[2]Texas Instruments DAC088S085
[3]Texas Instruments LMV324

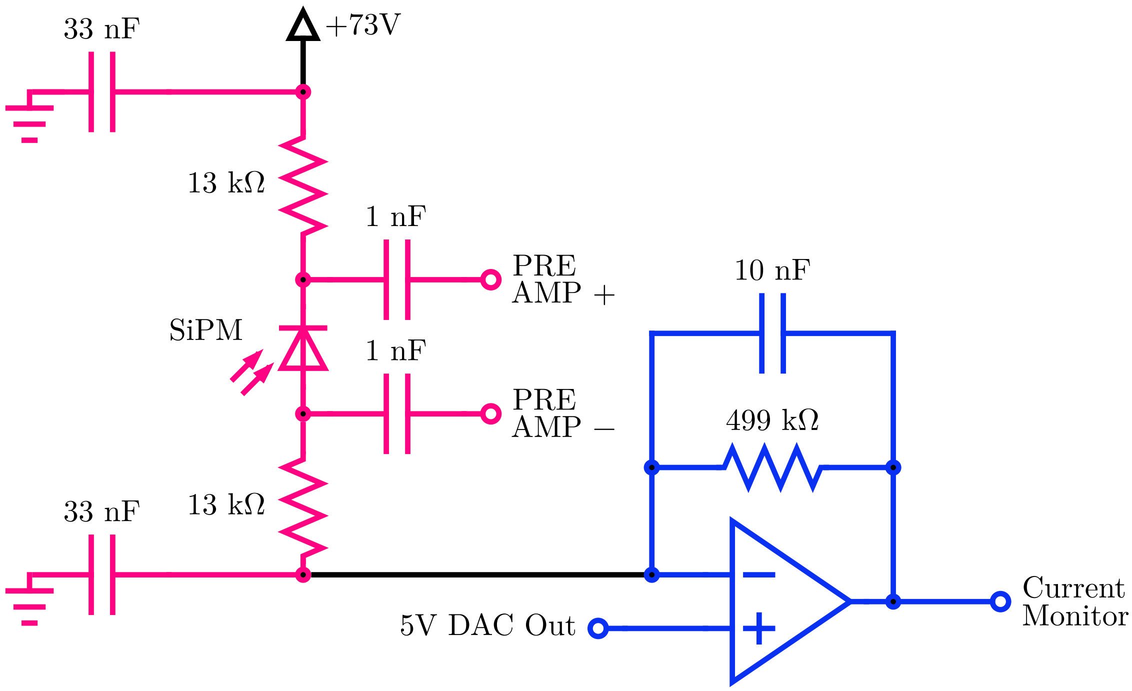

Figure 12.2: Simplified schematic of SiPM bias and current monitor. Pink components, including the preamplifier itself, are located inside a scintillator module and blue components are part of the RHIC. The 5 V DAC output is used to create a virtual termination for the SiPM via the inputs of a charge-sensitive amplifier. This allows fine tuning of the gain in 20 mV steps. The current monitors of all channels are connected to the Standard Control and Read-Out Device (SCROD) via multiplexers and an analog-to-digital converter. Their purpose is to monitor dark current of each channel over the life of the experiment. SiPM output is captured by the inputs of the preamplifier prior to being transmitted over the ribbon cables.

12.3 Scintillator Motherboard

Each RHIC is connected edge-to-edge with a scintillator motherboard that is installed in a 9U Versa Module Eurocard (VME) crate (Fig. 12.3). The VME crate is only used for low-voltage supply to the scintillator readout electronics. The scintillator motherboard has 10 TARGETX daughter cards (waveform-sampling ASICs described in the next section) connected to its top. Each group of 15 SiPMs is routed to one daughter card. The motherboard is arranged in a two-bus configuration, each with independent control circuits and each with 15 data lines. As the largest scintillator modules (located in the endcaps) contain 75 vertical (x) strips and 75 horizontal (y) strips, this two-bus configuration allows for the readout of x hits and y hits simultaneously. These buses are routed to a single Standard Control and Read-Out Device (SCROD) board.

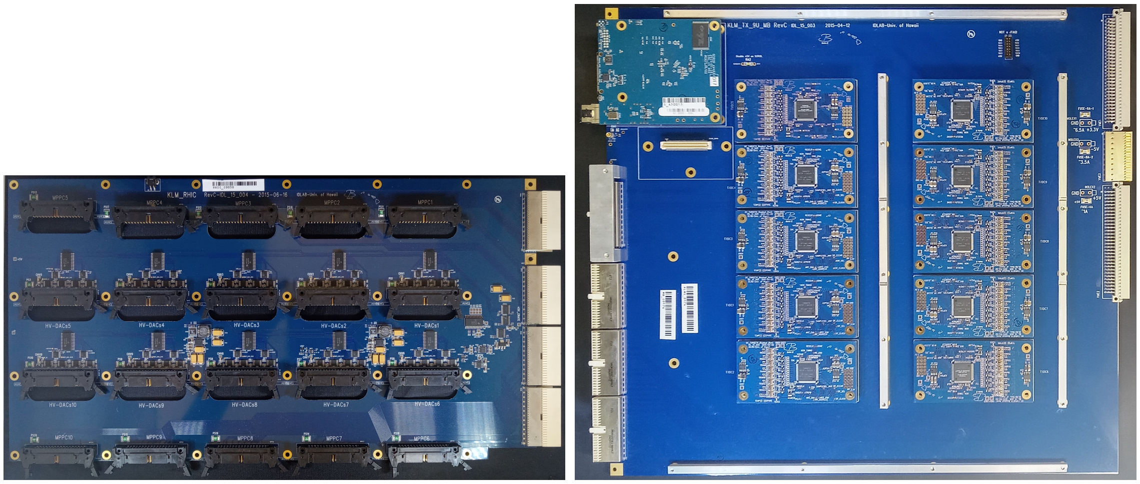

Figure 12.3: Ribbon header interface card (RHIC) and scintillator motherboard. The RHIC Provides voltage to SiPMs and collects signals. It contains 20 octal DACs for setting the voltage. They are 8-bit 5 V DACs and are used to fine tune the ground side of the SiPM bias. The scintillator motherboard has 10 TARGETX daughter cards and a SCROD mounted to it.

12.4 TARGETX Daughter Card and the TARGETX

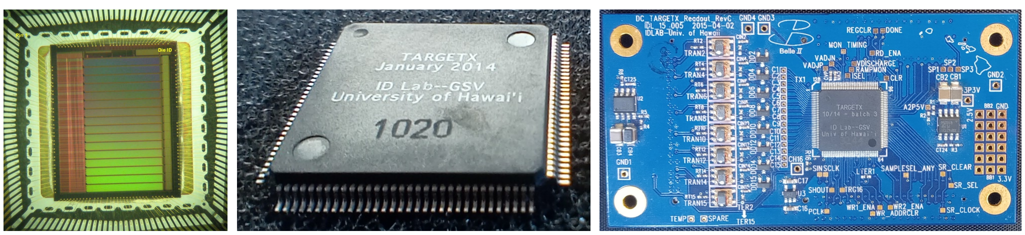

Figure 12.4: TARGETX die photograph, TARGETX in its package, and TARGETX Daughter Card.

The TARGETX (Fig. 12.4) is a waveform-sampling ASIC. The TARGETX daughter card (Fig. 12.4) AC-couples the SiPM signals via transformers and provides over- and under-voltage protection for the TARGETX inputs. Aside from containing various probe points for testing, its sole purpose is to house a single TARGETX ASIC.

The TARGET (TeV Array Readout with GSa/s sampling and Event Trigger) series of ASICs was originally designed for the readout of Cherenkov cameras [54]. The TARGETX is the latest version of TARGET and was designed specifically for KLM scintillator-based readout to include triggering capabilities. It is a 16-channel[4] analog storage device, capable of sampling at 1 GHz using 2^{14} sample-storage cells per channel, allowing it to store 16.384 µs of analog data. Every channel has two switched-capacitor sampling arrays of 32 sampling capacitors each; while one array is sampling

[4] KLM only uses 15 TARGETX channels. The 16th channel of each ASIC is used for bench testing.

the other is being transferred to a capacitor-based analog-storage array. Each channel's storage array is composed of 512 *windows* of 32 storage capacitors each. Storage takes place in a round-robin fashion, always moving the write-address pointer sequentially across the storage windows. The write pointer can be synchronized with the user's application by asserting a clear signal thus resetting the write address to zero. To prevent overwriting, the user can deassert the write-enable signal as the write pointer moves across some region of interest, but this will cause some dead time as incoming samples will have nowhere to be written. Each channel also has a fast trigger output with a programmable 12-bit trigger threshold. Finally, for digitization, the TARGETX uses a Wilkinson analog-to-digital converter (ADC). There is one Wilkinson ramp generator and 32 fast Gray-code counters per channel allowing all 32 samples of a selected storage window to be digitized simultaneously on all 16 channels. The device has one data-out pin per channel and a 14-bit address select bus; hence, data from all of its channels can be shifted out in parallel, starting and stopping on any of the 2^{14} storage cells desired. Refs. [54, 55, 56] provide more information about the TARGET series of ASICs.

12.5 Standard Control and Read Out Device (SCROD) Board

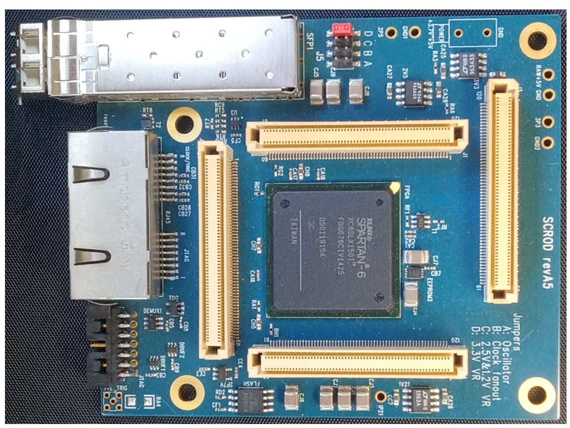

Figure 12.5: A SCROD board. This board is basically the command center for the scintillator readout electronics. It controls up to 10 TARGETX ASICs all the while listening for triggers and command words, processing waveform data, and delivering data packets.

Also connected to the top of the motherboard is the final piece of dedicated readout electronics for the scintillator system, the SCROD (Fig. 12.5). This board contains one field-programmable gate array (FPGA)[5] and serves as the interface to the rest of the Belle II readout system. Global (detector-wide) synchronous clock and global trigger arrive via low-voltage differential signals over network cables to one registered-jack 45 (RJ45) connector on the SCROD. A second RJ45 brings

[5] Xilinx Spartan-6 XLS-150T

in the JTAG (joint-test action group) signals needed for reprogramming the FPGA. Lastly, a serial fiber transceiver provides two-way communication with the rest of the detector's readout system. It can operate using either an external clock source or its onboard oscillator by the addition of a jumper resistor. A static random-access memory chip (SRAM)[6], which is used in its 4 M by 8-bit configuration, provides additional storage for the FPGA.

[6] Infineon Technologies CY62177EV30

CHAPTER 13
SCINTILLATOR READOUT FIRMWARE

Control of 10 TARGETX ASICs, waveform readout and processing, L0 trigger buffering, and L1 trigger processing are all managed by a single SCROD (L0 refers to channel self triggers from the TARGETX while L1 refers to triggers from the Belle II global decision logic, basically a request to the front-end electronics to check their buffers and report any hits within the time window of interest).

The firmware can be described by breaking it into two principal paths through which data flows. These two main paths are the L0 trigger path and the digital waveform path. In Fig. 13.1, these two paths are colored orange and yellow, respectively. For brevity, peripheral processes like the configuration and status register interface, phase-locked global clock input, fiber transceiver interface, and SRAM access are not discussed.

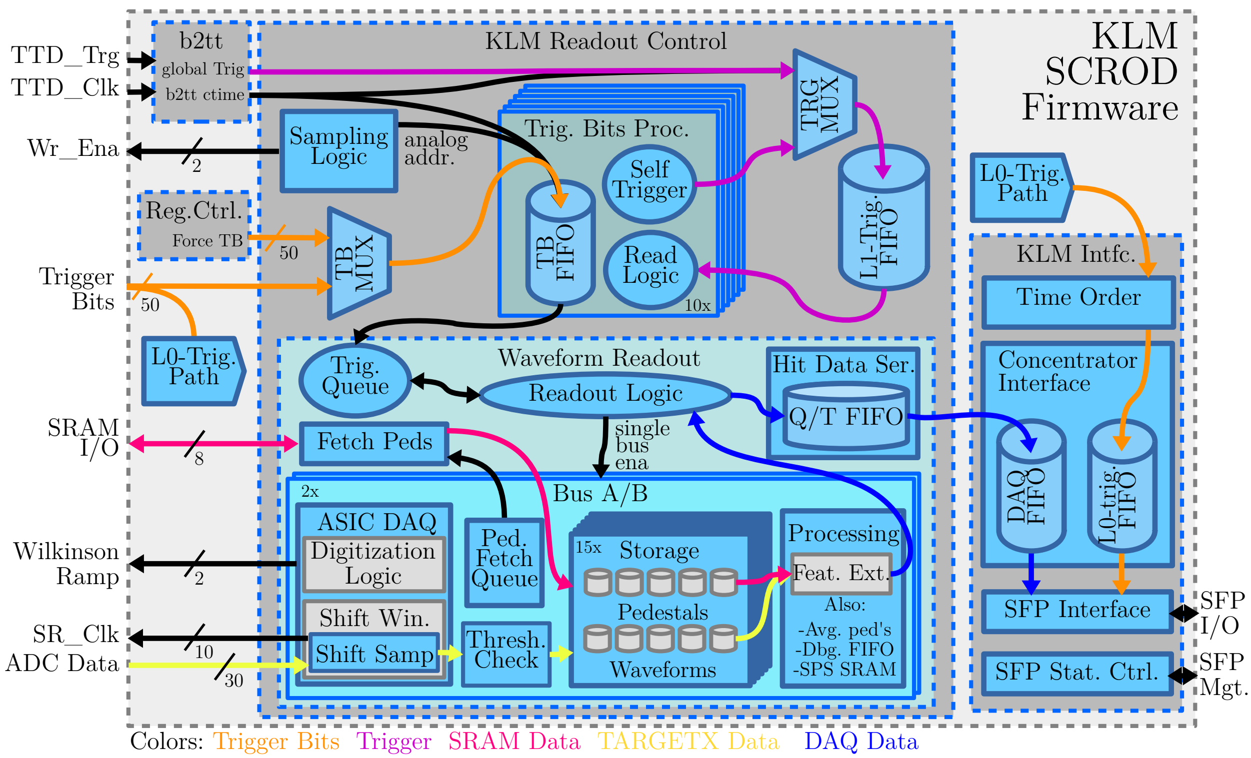

Figure 13.1: Block diagram depicting the layout of the SCROD firmware. The core entity of the waveform readout logic is duplicated twice, once for each bus of five TARGETX ASICs. Similarly, the storage entity is duplicated 15 times on each bus since all 15 channels of a TARGETX can be read out simultaneously

13.1 L0 Trigger Path

In the TARGETX, the analog input of each channel is tied to one input of a comparator. The comparator reference input is the programmable trigger threshold. When the analog input crosses the threshold, the comparator triggers a fixed-length digital pulse generated by a one-shot timer. One-shots from all 15 channels are fed into an address encoder, and finally delivered to the FPGA using five *trigger* bits. Four bits are sufficient to encode 15 channels by reporting the channel number in binary. Sixteen-channel encoding is not possible because zero is reserved for the case when there are no triggers. The fifth (most significant) trigger bit handles the case when more than one channel fires simultaneously on a single TARGETX. In this case, trigger bit 5 is held high and the first four (least significant) bits encode which group of 4(3) channels may have been hit (the last group contains 3 channels). For example, the five-bit pattern 0b10001 means multiple channels in the first group (channels 1, 2, 3, and 4) have been hit, while the five-bit pattern 0b11100 means multiple channels in the last two groups (channels 9 through 15) have been hit. We refer to these as multi-channel hits. For such hits, waveform digitization is required to disambiguate which channels in the channel groups actually had hits. If waveform digitization is not used, the detector resolution is degraded from 4 cm to 16 cm, and the number of strips hit in each group is unknown.

When L0 triggers from any of the TARGETX ASICs on the motherboard arrive at the FPGA, they are copied, timestamped, and split into two paths. One path performs time ordering of the hits and sends the time-ordered hit information via fiber optics to the Data Concentrator. These hits are transmitted to the Belle II trigger decision logic. The other path timestamps the bits again, this time using the TARGETX write address, and writes them to a FIFO (first-in first-out), where they wait to either be matched to an L1 trigger and go on to digitization, or to exceed the maximum look-back time and be cleared from the FIFO. There is one such FIFO allocated for each of the 10 TARGETX ASICs.

13.2 L1 Trigger Handling

After an L0 trigger is sent, the Belle II global trigger logic calculates (based on L0 triggers from all of the subdetectors) whether or not a global (L1) trigger should be issued. The latency of this decision is fixed, so when a trigger is issued, every subdetector only need look back a finite amount of time (less than 5 μs) to see if there were any corresponding L0 triggers within that time frame.

When an L1 trigger is received, the SCROD firmware quickly checks its hit buffer and earmarks any hits for digitization. The first step is to set up a mask over the region of interest to prevent the TARGETX from overwriting the analog storage cells before digitization can finish. If more L1 triggers are received while a previous one is being digitized, they are also masked immediately and earmarked for later digitization. To prevent event pileup, the digitization queue has a programmable threshold at which it will forego digitization in order to catch up. When this threshold is reached and a previous L1 trigger has just finished processing, the remaining jobs in the queue are processed in *simple* mode, wherein hit time is just the timestamp of the L0 trigger, pulse height is reported as zero, and one extra bit within the data packet is toggled to indicate that waveform digitization was not carried out for that hit.

Simulation tests verify that this scheme allows the SCROD firmware to keep pace with a 30 kHz L1 trigger rate (Poisson-distributed with a minimum of 200 ns between consecutive L1 triggers), the requirement for Belle II unified readout-system design [57]. Analysis of physics data taken at a luminosity of $1.9 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ and with an L1 trigger rate of 2.8 kHz shows that digitization is skipped for only 0.4 % of hits.

13.3 Digitization

To fetch waveforms from the TARGETX analog memory, the SCROD firmware asserts a 9-bit window address to each and enables a Wilkinson ramp on one or more of them. The 32 samples in a given storage window are digitized simultaneously. The time required to digitize depends on the slope of the Wilkinson ramp, which is tunable and can range from less than one microsecond to many tens of microseconds. Tuning of the Wilkinson ramp is described in Section 16.1.

When ADC conversion is finished, the SCROD firmware supplies a shift-register clock to one TARGETX and reads back all 15 channels in parallel, one bit at a time. While all 10 TARGETX ASICs can be digitized simultaneously, data can only be shifted out from two of them at a time (one on each bus). Of the 32 samples digitized for a given window, a 5-bit sample select signal is asserted to choose which sample to shift out. Once a sample is shifted from an ASIC to the SCROD, it is written to a FIFO (*Waveform FIFOs*). A total of 150 FIFOs (each 12-bit wide and 512-bit deep) are allocated for this purpose — one for each channel on the motherboard.

The region of interest (ROI) is calculated by the SCROD firmware in advance based on the timestamp of the L0 trigger. It may be contained in one storage window or span several storage windows. In the latter case, as in Fig. 13.2, each window must be digitized and shifted out in sequence. The use of one FIFO per channel avoids complications that would arise from multiplexing data across multiple channels and ASICs.

A register in the SCROD firmware sets the number of samples to read out. The minimum number of samples is four, limited by the ROI calculation logic, and the maximum is 512, limited by the depth of the waveform FIFOs. Another register sets a sample-number offset relative to the L0 timestamp. This allows fine-tuning of the *lookback* time — just as one would turn the horizontal-translation knob on an oscilloscope to center their signal in the screen.

Before analyzing the waveforms, they must be cleaned up. Every storage cell in the TARGETX has a slightly different DC offset, known as the pedestal voltage. The pedestal voltage for each storage cell must be measured beforehand in a calibration sequence so that it can later be subtracted from the waveform. Another 150 FIFOs (also 12-bit by 512-bit) are allocated for the waveform

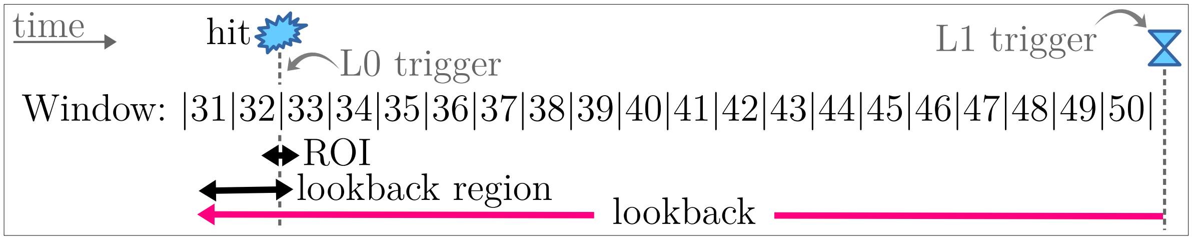

Figure 13.2: Depiction of lookback time, region of interest, and TARGETX storage windows. In this example, the initial hit is timestamped (L0 trigger) in the firmware at the end of storage window number 32. After an L1 trigger is issued, the lookback region is calculated and the hit buffer is checked for hits. Since there is a hit contained in the lookback region, the region of interest is calculated, and in this case, it spans the boundary of windows 32 and 33, so both windows must be digitized.

pedestal subtractions.

13.4 Pedestal Management

Pedestals are stored on the SCROD's 4 M×8-bit static RAM (SRAM) chip. Whether reading or writing, SRAM access takes 55 ns per address. Since TARGETX samples have a 12-bit resolution, pedestal values for two sample cells are stored over three SRAM addresses. On the FPGA, the logic for reading and writing the waveform and pedestal FIFOs contains a normal (data acquisition) mode and a measurement mode.

In the normal mode, while the region of interest is being digitized and shifted out by the digitization logic, another firmware entity begins fetching pedestal values from the SRAM and writing them to the pedestal FIFOs. As there is only one SRAM to store pedestals for all ten TARGETX ASICs, pedestal reading necessarily happens sequentially. If a TARGETX ASIC experiences a multi-channel hit, pedestals must be fetched for every channel that may have been hit (between 3 and 15 channels, depending on the trigger-bit pattern on the multi-channel hit). Pedestal reading generally takes less time than shifting out waveforms from the TARGETX ASICs, but in the case of a multi-channel hit with a trigger bit pattern requiring several channels to be checked, e.g., 0b11111, SRAM access becomes the bottleneck.

The firmware entity responsible for pedestal reading has a fair arbiter that prioritizes SRAM scheduling for one channel on one bus, and then one channel on the opposite bus. It remembers which bus was serviced last and services the opposite bus next if arbitration between the two buses is required. This feature was added to prevent biasing one bus over the other in case a make-haste signal is applied and some hits in an event have to forego both digitization and/or feature extraction. Such a signal is not implemented at this time but may be required in the future when SuperKEKB luminosity, detector background rates, and L1 trigger rates all increase.

Figure 13.3: Resource-saving scheme to use existing FIFOs for measuring pedestals. Concatenating inputs and outputs of two 12-bit FIFOs allows pedestals to be measured up to 2^{12} times. A shift register performs the averaging.

In measurement mode, the firmware hijacks the digitization machinery described above, the voltage of each storage cell is measured many times (up to 2^{12}), and the average pedestal values are written to the SRAM. This must be done in advance, with the HV turned off, so that the values are available during data taking. In this configuration (Fig. 13.3), inputs and outputs of one waveform FIFO and one pedestal FIFO are concatenated to make a single 24-bit wide FIFO. Prior to pedestal measurement, all the FIFOs are primed with 32 writes of the value zero (the number of samples in a TARGETX storage window). During pedestal measurement, a command to write to the FIFOs causes them to first be read once, their 24-bit output is added to the 12-bit sample, and the sum is written back into the FIFOs. It is repeated $2^{\mathcal{N}}$ times for each of the 512 storage windows, where $\mathcal{N}$ is set by a SCROD register. Averaging is achieved by shifting the 24-bit result to the right $\mathcal{N}$ times and then keeping the 12 least-significant bits.

13.5 Feature Extraction

While processing an L1 trigger, the feature extraction entity is enabled as soon as any of the channels have both their associated waveform and pedestal FIFOs ready. The two FIFOs are read in tandem and the pedestals are subtracted from the waveform samples. To avoid the use of negative numbers, a baseline value of 3072 (3/4 full scale) is added to each sample during pedestal subtraction.

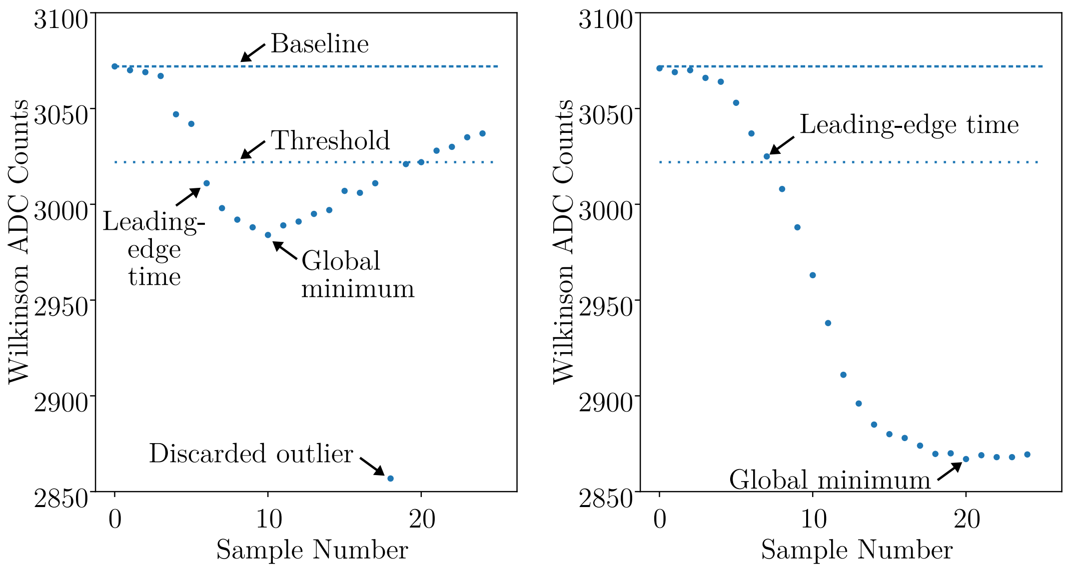

Figure 13.4: Example waveforms showing the SCROD firmware's feature-extraction algorithm. Left: Normal pulse. The sample closest to the discrimination threshold is recorded as leading-edge time. The baseline minus the global minimum is recorded as pulse height. The outlier is rejected. Right: Pulse that saturated its preamplifier as can be seen by the plateau shape.

Leading-edge time is measured using constant-value discrimination with linear interpolation to find the sample nearest the threshold. Constant fraction discrimination is not used due to the saturation of the preamplifiers for very large pulses (Fig. 13.4). Pulse height is measured by finding the global minimum (pulses are negative going), requiring equal or larger samples on either side, and recording its magnitude with respect to the baseline. An outlier monitor discards any potential minimum sample that is further than 128 ADC counts from the average of its two neighbor samples. Outliers may occur if there is a bit error when shifting a sample out of the TARGETX. If a small pulse does not cross the discrimination threshold, the time of the minimum is substituted for the leading-edge time. If a global minimum is not found (waveform is constantly increasing or decreasing), then the last sample is used for the pulse height. Time and pulse height are written to a data packet, and sent over a Xilinx Aurora data link to the Data Concentrator.

At the same time (for calibration purposes) the waveform is written to a debugging FIFO, and the measured pulse height is written to a histogram, both of which can be read back through the register interface. Three debugging FIFO modes are available: write the waveform with pedestal subtraction, write the waveform without pedestal subtraction, and write the pedestals only.

Once all valid channels in the event have been processed, the remaining waveform FIFOs corresponding to channels that did not have valid hits but were shifted out of the TARGETX anyway are all read until they are empty, so that the firmware is ready to process the next event in the queue.

CHAPTER 14
RPC FRONT-END BOARD

RPC signals are generated when a charged particle leaves an ionization trail in the gas gap of an RPC module. The strong electric field creates an avalanche, and electric current is induced on the two orthogonal readout-strip planes in the vicinity of the avalanche. Like the scintillator readout, the RPC readout relies on ribbon cables to transport the signal to the readout electronics. Unlike the scintillator readout, the RPC readout is not instrumented with waveform digitizers. It discriminates the signal and timestamps it before transmitting the timestamp to the Data Concentrator.

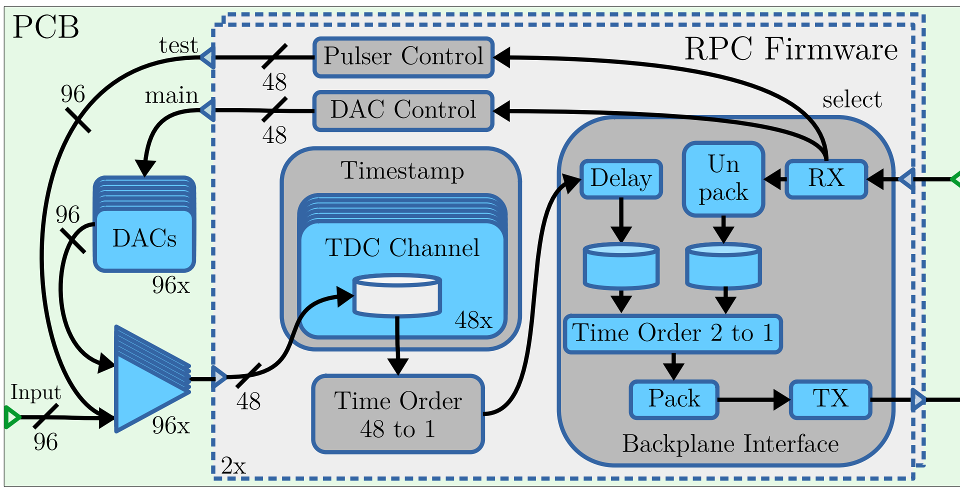

Figure 14.1: Schematic of an RPC front-end board and its firmware. It has two FPGAs, each one handling 48 RPC channels. Signals from RPC channels that exceed the programmable digitization threshold are timestamped and transmitted over the VME backplane to the Data Concentrator. There are 13 RPC front-end boards per Data Concentrator in the barrel region of the KLM detector.

A hit is formed when an electric pulse from an RPC readout strip exceeds a discrimination threshold set by a DAC[1] on the RPC front-end board. The basic layout of the RPC front-end board and its firmware is depicted in Fig. 14.1. In the RPC front-end firmware, hits are timestamped in a time-to-digital conversion (TDC) module with a resolution of 3.94 ns (2×system clock). Each RPC front-end board contains 96 line receivers and discriminator channels, 48 per front-panel (ribbon cable) connector. Channels 1-48 connect to negative RPC pulses while channels 49-96 connect to positive RPC pulses. An analog test pulser provides an independent built-in test of each channel. Two FPGAs[2] are used for discriminator control and timestamp generation. These FPGAs are configured over the backplane using SERA/A08/A09. The threshold and pulser are operated with run control commands over the Belle2Link. The discriminators only generate a rising edge for the

[1] Analog Devices LTC2636
[2] Xilinx Spartan-6 XC6SLX25

FPGA timestamp generator. Hits are also time ordered, using their timestamps, to simplify event building in the Data Concentrator.

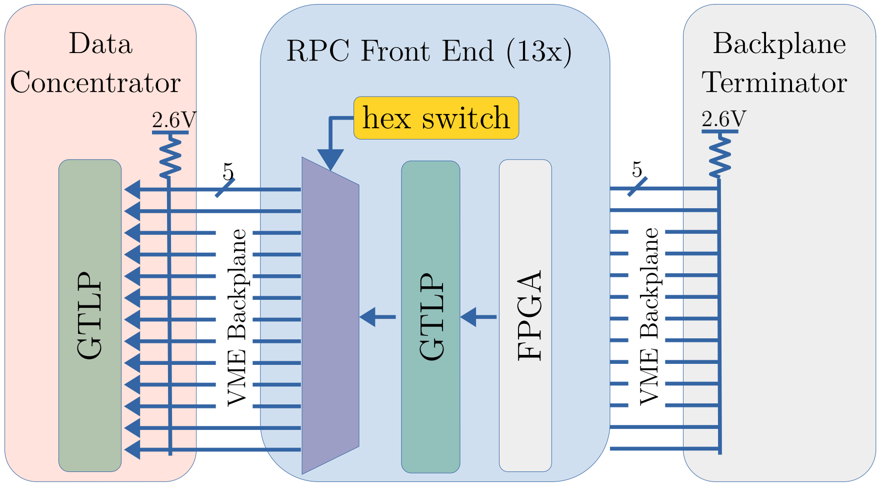

Figure 14.2: Custom backplane protocol using rotary switches to address a demultiplexer and communication over 5-bit buses using GTLP.

Hits are transmitted to the Data Concentrator over the VME backplane using a custom protocol. The backplane termination voltage is lowered to 2.6 V for GTLP (Gunning Transceiver Logic Plus). TDC data is transmitted over dedicated 5-bit buses. Each 5-bit bus is 1-13 demultiplexed to the desired slot/position which corresponds to its layer. The position is selected by a hex rotary switch on the RPC front-end board which must be set during installation (Fig. 14.2).

CHAPTER 15
THE DATA CONCENTRATOR

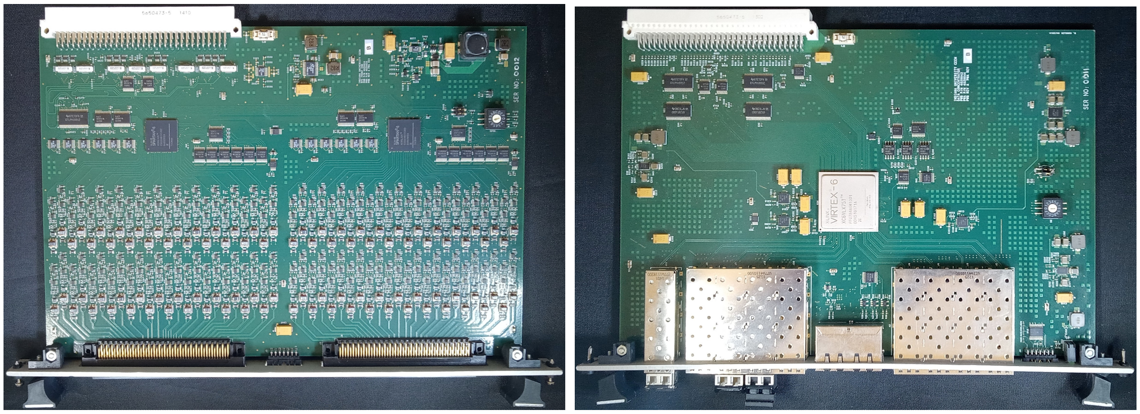

Figure 15.1: Left: RPC front-end board. Right: Data Concentrator.

Each Data Concentrator (Fig. 15.1) contains nine serial fiber transceivers, two RJ45 connectors, and a single FPGA.[1] It collects data packets from either two scintillator modules over fiber and thirteen RPC modules over a VME backplane in the barrel region, or from six to seven scintillator modules (all over fiber) in the endcap regions. Two additional fiber transceivers are used to transmit hit packets to the Belle II data acquisition system and to transmit trigger packets to the Belle II trigger decision system, respectively. One RJ45 provides clock and trigger inputs while the other is used for FPGA programming. A block diagram of the Data Concentrator firmware is depicted in Fig. 15.2

For any particular L1 trigger, the Data Concentrator gathers any RPC packets it has and then waits for each connected SCROD to respond with either a valid or null data packet. The combined data packets are then sent to the Belle II data acquisition (DAQ) system via fiber using a custom protocol called Belle2Link.

For sending configuration data packets to the front-end electronics, the Belle2Link protocol is also used. Further, to send data to the SCRODs, the Data Concentrator translates everything from Belle2Link and passes it on to the SCROD via the Xilinx Aurora protocol.

[1] Xilinx Virtex-6 XC6VLX75T

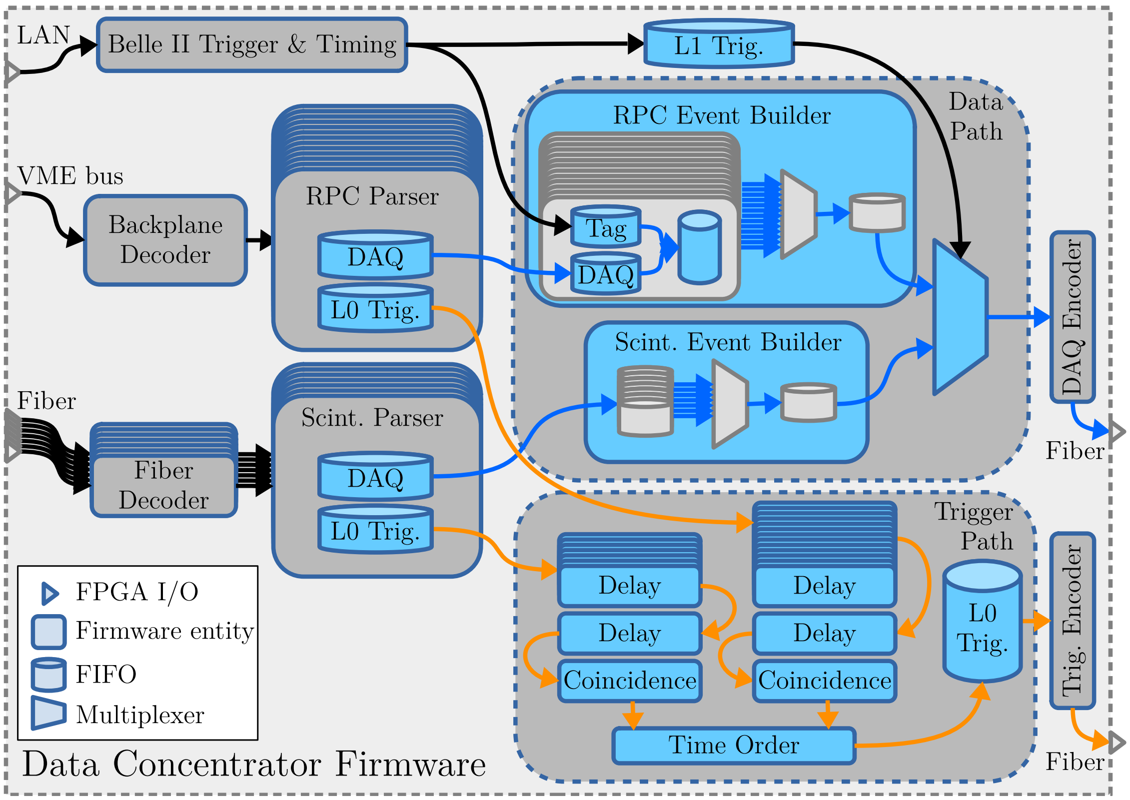

Figure 15.2: Flow chart depicting the top level of the Data Concentrator firmware. The blue (orange) arrows represent the data path (trigger path). Clock distribution, register interface, and fiber transceiver interface omitted.

CHAPTER 16
CALIBRATION OF THE TARGETX ASICS AND SIPMS

The TARGETX ASIC has 61 registers for device calibration. Fifteen registers set the width of the trigger bit pulses from the one-shot circuits, another 15 set the trigger threshold comparators for each channel, and one register enables a test pattern. The remaining 30 registers are used to optimize the performance of the TARGETX. Of these 30 remaining registers, some are for tuning the shape of the Wilkinson ramp, some for timing of sample storage and addressing, and some for time-base corrections. The initial tuning of these registers is described in Ref. [58]. Further, on the RHIC, there is one 5 V 8-bit DAC for fine-tuning the SiPM gain on each channel.

16.1 Wilkinson Ramp Tuning

The TARGETX ASIC uses a Wilkinson ADC to digitize its analog samples. During digitization, one input of a comparator is driven by the sample cell that is being digitized and the other by a linearly increasing voltage source (the Wilkinson ramp). The time it takes for the ramp voltage to reach the sample voltage is proportional to the sample voltage. When the ramp begins, a 12-bit Gray-code counter is activated. When the ramp exceeds the sample voltage, the comparator output latches the Gray-code counter to its present value. We refer to this value as the number of Wilkinson ADC counts — a digital value that corresponds to the voltage of the sample. The Wilkinson clock that increments the Gray-code counter is provided over low-voltage differential signals generated by the SCROD FPGA.

In the TARGETX ASIC, the I-select register controls the slope of the Wilkinson ramp, and the V-discharge register controls the starting voltage of the ramp (Fig. 16.1:left). Increasing I-select decreases the slope. We measure the behavior of I-select by turning off the HV and digitizing samples corresponding to the input offset voltage of the TARGETX channels, which is set at about 3/4 of the dynamic range for TARGETX inputs. We measure the mean number of Wilkinson ADC counts for groups of 32 samples at every I-select setting. At higher values of I-select, and with a Wilkinson clock period of 7.87 ns, the input offset voltage is high enough that the Wilkinson counter exceeds its maximum and continues counting from zero. In the measurement, we correct this overflow in software (Fig. 16.1:right). The quadratic shape below $V_D/2$ is characteristic of the P-channel MOSFET that is driving the ramp generator. We select a value for I-select within this quadratic band that maximizes the dynamic range of the Wilkinson ADC without overrunning it.

One unfortunate design flaw of the preamplifiers is that they were optimized for single-pixel detection, yet they saturate quite easily when a large number of SiPM pixels fire in tandem, causing the negative-going SiPM pulse to have a flat bottom rather than a well-defined extremum. This hardware choice could not be changed, so with this in mind, we set V-discharge so that saturated

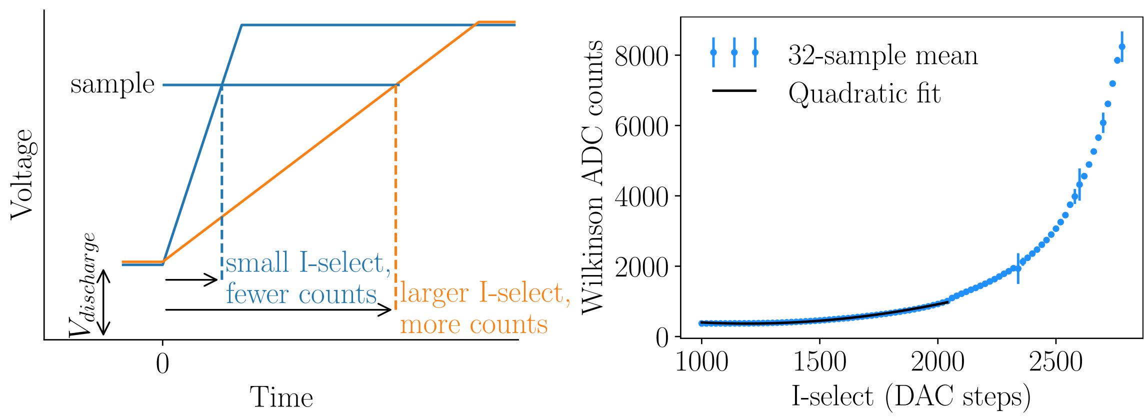

Figure 16.1: Left: Wilkinson ADC principle. A counter is enabled when the ramp begins and continues counting until the ramp exceeds the sample voltage. The discharge voltage of the ramp generator sets the starting voltage. Lower values of I-select make steeper slopes, which leads to shorter digitization times and fewer Wilkinson ADC counts. Right: Measured baseline as a function of ramp slope. Baseline refers to a constant-voltage input without signal present. Quadratic behavior reflects saturation mode of a p-type FET at drive strengths below $V_D/2$. The 12-bit Gray-code counter overflows twice near the end of the scan. The data was overflow corrected at intervals of 2^{12}, hence the larger error bars near these boundaries. Also, outlier samples are more common near 2^{11} ADC counts hence the remaining larger error bars.

pulses will have a minimum near zero Wilkinson ADC counts. This choice allows for quicker digitization times.

16.2 Aligning the Trigger Threshold Baseline

The analog input of each TARGETX channel is not only sampled by the sampling array but also provides one input of the trigger-threshold comparator. The other input is provided by a 12-bit DAC. These trigger-threshold DAC settings are each tuned independently. Nominally, all of the 15 analog inputs would have the same DC offset, and the trigger-threshold DACs would be identical. In practice, however, there is variance.

The first step to aligning all channels is to turn the HV off and measure the frequency of trigger bits versus the trigger-threshold DAC setting for each channel (Fig. 16.2). This is done in firmware by counting the number of trigger bits within a programmable time interval, then reading out the count via the status register interface. The result is a normal distribution centered on the value of interest. The average width of all the distributions is 1.495(22) trigger DAC steps.

This measurement provides the trigger threshold that corresponds to zero pixels hit on the SiPM. We call this the trigger-threshold baseline value. This measurement is performed on all 18,560 installed channels and saved to a database. A histogram of the measured baseline values for all channels is shown on the left of Fig. 16.3. Ultimately, we want to tune the trigger threshold to

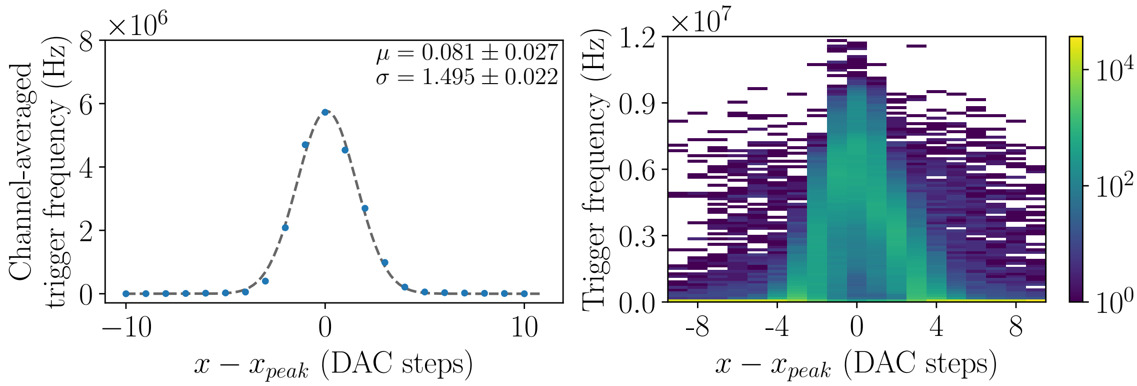

Figure 16.2: The variable x refers to the trigger-threshold DAC setting of any particular channel. Left: Sum of all baseline scans divided by number of installed channels. Right: Histogram of all the scans in order to demonstrate the variance between all installed TARGETX channels.

a value that corresponds to a predetermined number of fired pixels from the SiPM. However, the voltage seen by a single pixel depends on the gain of the SiPM.

16.3 Coarse Gain Adjustment

Before finely tuning the gain on more than 18,000 channels, it is prudent to first perform a coarse measurement to get close to the desired value. According to the SiPM vendor, at a 70 V bias, the frequency of SiPM pulses larger than 1.2 pixels is 75 kHz, and 750 kHz for 0.5 pixels, respectively. Extensive testing on a few channels pinned down a trigger-threshold DAC setting of 35 below the baseline value as corresponding to a trigger threshold of 1.2 pixels.

Because the low side of the SiPM bias voltage is set with an 8-bit 5 V DAC, we set the high side of all SiPMs to 73 V — about 1 V to 5 V above the breakdown. Now, with knowledge of the trigger-threshold baseline value from the previous section, we set each channel's trigger threshold DAC to 35 below its baseline value, and we measure the trigger bit frequency versus the HV-trim DAC setting. We select the DAC setting that yields a trigger-bit frequency closest to 75 kHz as the coarse setting. A histogram of the best HV-trim DAC values and corresponding frequencies for all channels is shown on the right of Fig. 16.3.

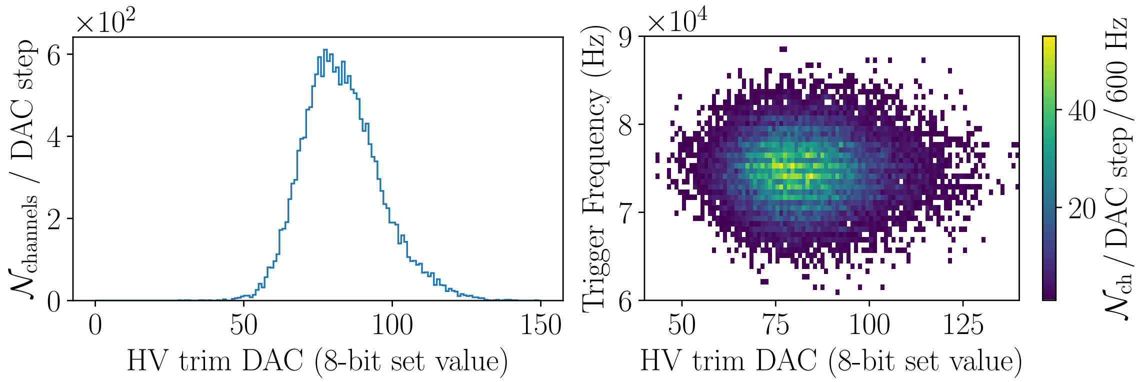

Figure 16.3: Left: Histogram of HV trim DAC values for all installed channels. Right: Histogram of coarse HV-trim DAC adjustment and corresponding trigger scalar frequency for each (when the trigger thresholds are set to 35 trigger DAC steps below the baseline value).

16.4 Normalizing Gain on All SiPMs

Single-photon spectra (SPS) are one way to measure gain (Fig. 16.4). We use a histogram of pulse height measurements for this. Each peak in the spectrum corresponds to a number of pixels fired. The separation between adjacent peaks is the gain of the SiPM plus preamplifier. To set the gain uniformly on all channels, we need to know the gain as a function of the HV-trim DAC setting. This requires many SPS and many fits. Ordinarily, this is achieved by using a calibration source such as an LED, and a climate chamber can allow for testing at different temperatures.

The KLM detector was designed without calibration sources in mind. This leaves the inherent dark rate of the SiPM as the only means of measuring gain. A further complication is encountered in the data acquisition system: the Data Concentrator firmware and the readout PC were designed to accept 8-byte packets from the scintillator front end, not waveforms. Finally, whatever procedure is used, it must be repeated on all 18,560 channels. To overcome these issues, we record SPS in firmware by keeping a histogram of waveform peak measurements in the FPGA's block RAM. Each address in the allocated RAM corresponds to one bin of the histogram, and filling the histogram just requires reading the RAM, adding 1, and writing again.

To make the measurement, a special reset signal clears the SPS RAM contents, and trigger thresholds are turned off for all but one channel on the motherboard, which is set at a threshold of about one photoelectron. The firmware is configured in self-triggering mode, and pulse-height data is collected for 90 seconds. In this way, all 150 channels on a motherboard can be measured in 225 minutes. Offline software executes the measurement on all installed motherboards in parallel.

While the experiment hall is climate-controlled, some daily and seasonal variation in temperature is expected. Seasonal variation is mitigated by performing the calibration all at once. The absolute gain will fluctuate due to seasonal variation, but normalization across all channels should not be affected. Daily variation in temperature cannot be mitigated.

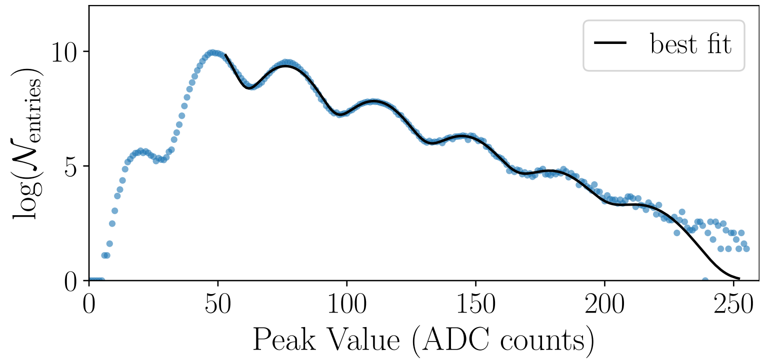

Figure 16.4: Example of darkrate-based single-photon spectrum. The trigger threshold is set to about one PE, which has the effect of diminishing the first two photo peaks. The best fit to a function with six peaks is drawn on top of the data.

After each 90 s measurement, the RAM contents are read out using the register interface for offline analysis. The procedure is repeated at different bias voltages over the range of 2 V to 3 V past breakdown to establish gain as a function of bias voltage for each channel.

We fit each SPS using a function that is a sum of Gaussian distributions which are regularly spaced by the parameter a_0 and whose amplitudes decrease by a power law χ^k, $0 < \chi < 1$:

$$f(x) = A \sum_{k=1}^{N_{PE}} \chi^k e^{\frac{-(x-(x_0+ka_0))^2}{2(\sigma_0+k\sigma_1)^2}} \,, \tag{16.1}$$

where χ is the optical crosstalk probability, x_0 is a pedestal offset, σ_0 describes electronic noise and ADC resolution, and σ_1 scales with the number of pixels fired. The only purpose of the fit is to extract the parameter a_0, i.e., the gain in units of ADC counts per photoelectron. The power law stems from crosstalk between pixels. This is due to infrared photons created in the avalanche. If these photons are emitted isotropically, then the probability that they will hit any other pixels is given by a power law. The crosstalk probability also depends on how many infrared photons are created in a typical avalanche, and on the quantum efficiency for converting infrared photons to photoelectrons. These can all be lumped into the parameter χ in the fit.

To measure gain as a function of bias voltage, we repeat the above procedure for different bias voltages (Fig. 16.5). For 10 data points per channel, we have to perform 185,600 fits. A non-linear least-squares minimizer is used. The fitting procedure is optimized by studying many examples. The best results are achieved by fitting the logarithm of the number of entries to $\log(f(x))$, and by providing the minimizer with the loss function $\rho(\delta^2) = 2(\sqrt{1+\delta^2} - 1)$, where ρ is the loss and δ is a residual. These choices improve the sensitivity of the fit to a_0, the only parameter of interest.

The result of this procedure on all 18,560 channels is a converging linear fit on all but approximately 1,000 channels (Fig. 16.6). The mean gain slope is 15 ADC counts / PE / V, and the mean

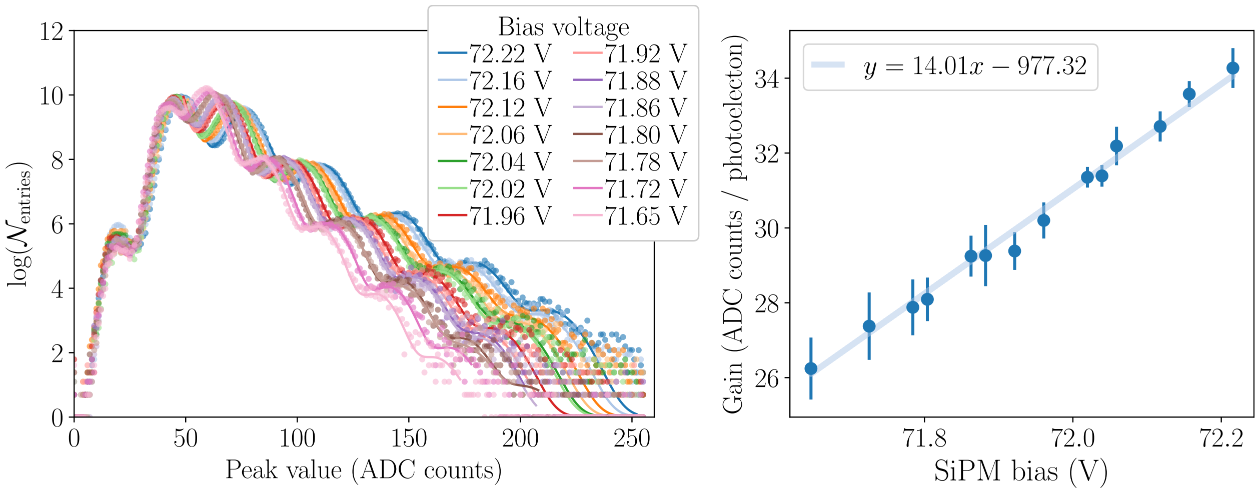

Figure 16.5: Left: Single-channel example of multiple gain fits at different SiPM bias voltages. Right: Mean separation between peaks, a_0, as a function of bias voltage.

breakdown voltage (x-intercept) is around 70 V. After the fit procedure, for all channels with a converged linear fit, the required HV-trim DAC value needed to achieve 30 ADC counts / PE is calculated using the linear fit function, and these values are written to the KLM detector's configuration database. For the channels without a converged fit, the HV-trim DAC value from the coarse gain adjustment is retained.

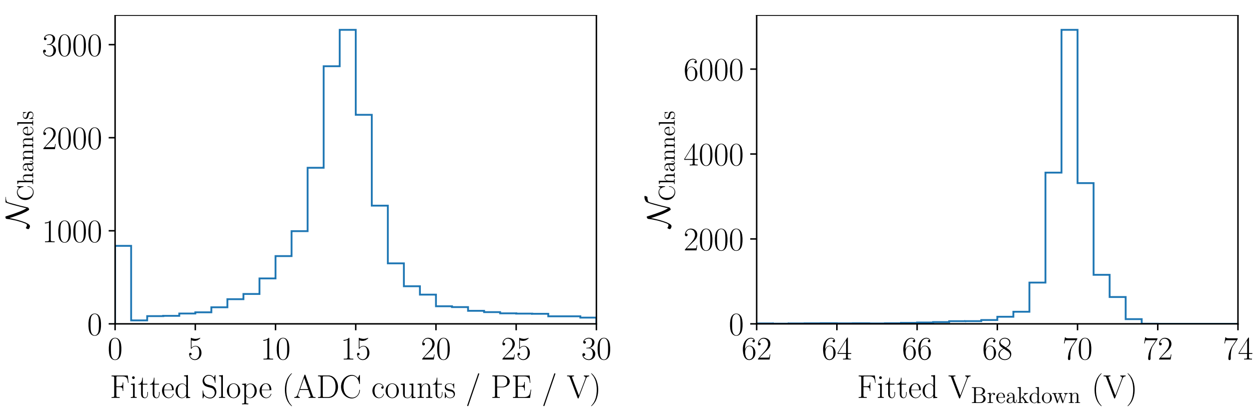

Figure 16.6: Results of SiPM gain calibration. Left: Slope of ADC counts / photoelectron / V for all installed SiPMs. The peak at zero corresponds to channels for which converging fits are not obtained using the automated measuring and fitting setup. For these channels, the coarse HV-trim DAC setting is retained. Right: x-intercept of each linear fit. This is the measured breakdown voltage of each SiPM

CHAPTER 17
PERFORMANCE

The hit time recorded by the firmware is delayed compared to the actual time that a particle interacts with a KLM module. This is due to the signal transmission from the hit position to the readout electronics. Because of the high background level expected at Belle II's design luminosity, precise measurement of hit time (t_0) resolution is necessary. Precise timing will lead to better reconstruction of tracks and K_L clusters. For a single-channel hit in the KLM detector, we define t_0 as,

$$t_0 = T_{\mathrm{rec}} - (T_0 + T_{\mathrm{flight}} + T_{\mathrm{propagation}} + T_{\mathrm{collect}} + T_{\mathrm{cable}}).$$

Here, T_{rec} is the time of the hit recorded by the firmware and is important for reconstructing 2D hits and subsequently tracks. The e^+e^- collision time for each event is T_0 (EventT0). We use EventT0 information from the Central Drift Chamber (CDC). The variable T_{flight} denotes the flight time of the particle from the interaction point to the hit position in the KLM module. For each recorded hit, T_{flight} is obtained by matching the KLM detector hit position to the extrapolated hit from CDC, which indicates the relative distance from the beam line. The variable $T_{\mathrm{propagation}}$ is the charge propagation (photon propagation) time for RPCs (scintillators) between the hit position and the end of the strip (fiber). In the scintillators, it is estimated using $L_{\mathrm{propagation}}/c_{\mathrm{eff}}$, where $L_{\mathrm{propagation}}$ is the propagation distance of the signal and c_{eff} is the effective speed of light in the fiber. The time between photoelectric conversion in a SiPM and leading-edge timestamping is T_{collect} and depends on the number of pixels fired in the SiPM. This contribution is small and currently treated as a constant. Finally, T_{cable} is the transmission time of the signal over the ribbon cables. The time delay of KLM detector hits is mainly from the ribbon cables, which is different for each strip due to the detector structure.

The distribution of $T_{\mathrm{cable}} = T_{\mathrm{rec}} - (T_0 + T_{\mathrm{flight}} + T_{\mathrm{propagation}})$ for each strip is fitted with a Gaussian function. The mean value of these Gaussian distributions is fitted with a constant function to obtain the weighted global mean. The difference between the mean value of one strip and the global mean value is used as the calibration constant for that strip and is stored in a database. In the Belle II software framework [37], T_{cable} values are subtracted when reconstructing the hits.

The results of the time calibration for data collected at the beginning of 2024 are shown in Fig. 17.1. For t_0 resolution, we state the standard deviation and full width at half maximum (FWHM). The t_0 resolutions (FWHMs) for RPCs, barrel scintillators, and endcap scintillators are 7.8 ns (14.0 ns), 5.4 ns (5.6 ns), and 4.7 ns (3.8 ns), respectively. The tails and asymmetric shapes of these distributions are likely due to a combination of the calibration algorithm itself and background hits. Currently, the algorithm does not account for the curvature of charged tracks within the Belle II magnetic field, and background hits that are not associated with reconstructed tracks in the KLM detector have not been removed.

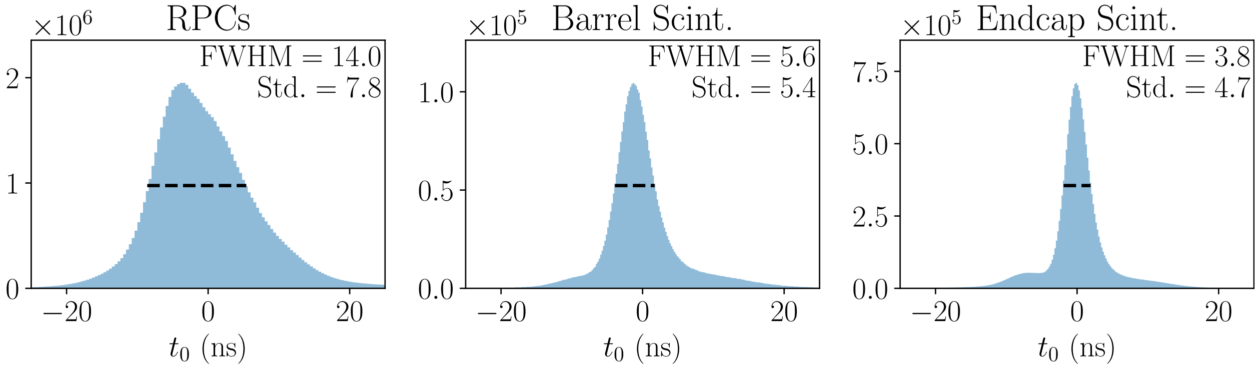

Figure 17.1: KLM hit time distributions for RPCs (left), barrel scintillators (middle), and endcap scintillators (right). The full width at half maximum (FWHM) and the standard deviation are indicated in each plot.

CHAPTER 18
CONCLUSION

In Part I a feasibility study on the use of inclusive hadronic tagging to improve sensitivity on $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ was presented. That the use of inclusive tagging to search for this rare decay improves sensitivity by 1–2 orders of magnitude over the world's leading measurement is very promising. Most notably, we are pleased to show that we can achieve sensitivities that make it possible to test some new physics scenarios where $b \to s\tau\tau$ processes are enhanced. With the anticipated full Belle 2 dataset of $50\,\mathrm{ab}^{-1}$ and continued development in software for particle identification, fake-rate reduction, and machine-learning techniques, it is hopeful that future analyses can reduce the sensitivity by another two orders of magnitude. Observation of this decay before we reach a sensitivity of 10^{-7} would be a smoking gun for new physics processes.

In our analysis, the largest source of background came from generic $B\overline{B}$ decays; future work should focus on reducing this background. One possible improvement, which was suggested at too late of a stage in this analysis for us to implement, would be to veto $K^{\pm}$ candidates that come from $D^0 \to K^+\pi-$ decays. Because this is a fully hadronic decay, a loose pion selection and tight cut around the D mass should be very effective in removing this background. To that end, such a veto should also include modes like $D^0 \to K^-\pi^+\pi^0$, $D^0 \to K^-\pi^+\pi^+\pi^-$, $D^0 \to K^-\pi^+\pi^+\pi^-\pi^0$, $D^+ \to K^-\pi^+\pi^+$, and $D^+ \to K^-\pi^+\pi^+\pi^0$.

In Part II, we described the implementation of a new electronic readout for the Belle II KLM subsystem. A challenging task was the creation of readout firmware for the scintillator readout that utilizes the waveform-digitization feature of the TARGETX ASIC at trigger rates up to 30 kHz. Having achieved this goal, we anticipate an improvement in particle identification for the KLM subsystem. With waveform readout now working, we can disambiguate multi-channel hits from a single ASIC, leading to improved track and cluster resolution. Waveform feature extraction has the potential to improve the hit-time resolution to about 1 ns, which will help with background rejection. Knowledge of SiPM pulse heights may also unlock new analysis techniques, perhaps incorporating pulse-height measurements to resolve low momentum muons from punch-through pions which made it into the KLM detector.

A second challenge—to calibrate gains on more than 18,000 SiPMs that were installed without any calibration sources—was resolved successfully. We deployed a procedure for efficiently recording single-photon spectra in firmware, and we developed an automated fitting procedure to extract the parameter of interest (the gain). We used the calibration results to homogenize the gain on more than 17,000 of the installed SiPMs.

Future work is required to improve the stability of the waveform digitization. Improvements in the fitting procedure or recollecting single-photon spectra on about 1,000 channels whose fits did not converge should be considered. Additionally, a more sophisticated feature-extraction algorithm, such as a finite impulse response filter, may lead to further improvements in time and peak resolution.

APPENDIX A
RECONSTRUCTION SELECTIONS

A.1 Kaons

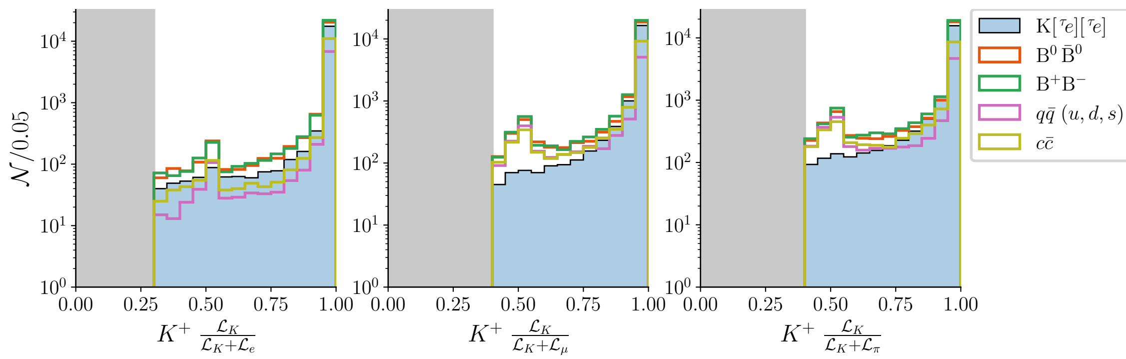

Figure A.1: Likelihood variables used for $K^{\pm}$ selection in the $K[^\tau e][^\tau e]$ mode.

Table A.1: Average candidate multiplicity per event/retention for selections made to the signal kaon particle list in the $K[^\tau e][^\tau e]$ mode. The function $f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T})$ is a dynamic ellipsoid around the interaction point with size inversely proportional to transverse momentum.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.6 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in CDC Acceptance	-	9.8 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_e} > 0.3$	-	3.0 / 0.975	4.1 / 0.977	4.0 / 0.979	3.5 / 0.928	3.6 / 0.944
3	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\mu} > 0.4$	-	2.1 / 0.928	2.5 / 0.859	2.4 / 0.879	2.1 / 0.737	2.2 / 0.822
4	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\pi} > 0.4$	-	2.0 / 0.918	2.3 / 0.838	2.3 / 0.861	2.0 / 0.703	2.1 / 0.800
5	$f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T}) < 4^2$	cm	1.6 / 0.836	1.6 / 0.684	1.7 / 0.732	1.5 / 0.553	1.6 / 0.653

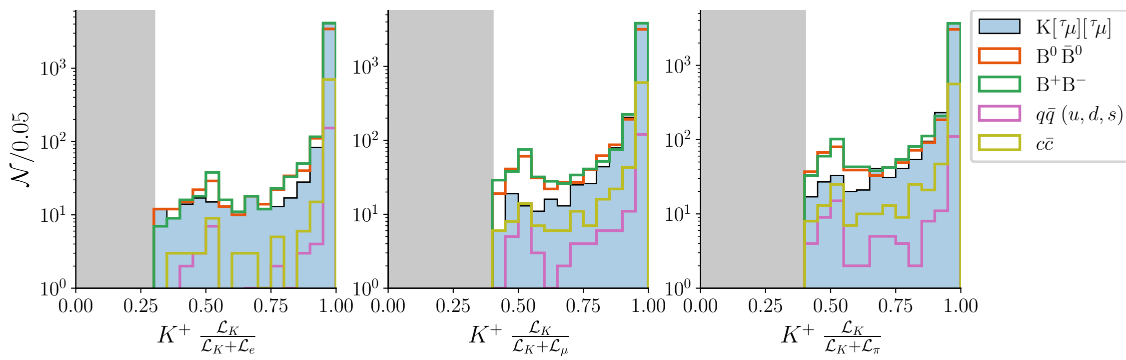

Figure A.2: Likelihood variables used for $K^{\pm}$ selection in the $K[{}^\tau\mu][{}^\tau\mu]$ mode.

Table A.2: Average candidate multiplicity per event/retention for selections made to the signal kaon particle list in the $K[{}^\tau\mu][{}^\tau\mu]$ mode. The function $f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T})$ is a dynamic ellipsoid around the interaction point with size inversely proportional to transverse momentum.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in CDC Acceptance	-	9.7 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_e} > 0.3$	-	3.4 / 0.982	4.1 / 0.977	4.0 / 0.979	3.5 / 0.928	3.6 / 0.944
3	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\mu} > 0.4$	-	2.1 / 0.926	2.5 / 0.859	2.4 / 0.879	2.1 / 0.737	2.2 / 0.822
4	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\pi} > 0.4$	-	2.0 / 0.915	2.3 / 0.838	2.3 / 0.861	2.0 / 0.703	2.1 / 0.800
5	$f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T}) < 4^2$	cm	1.6 / 0.836	1.6 / 0.684	1.7 / 0.732	1.5 / 0.552	1.6 / 0.654

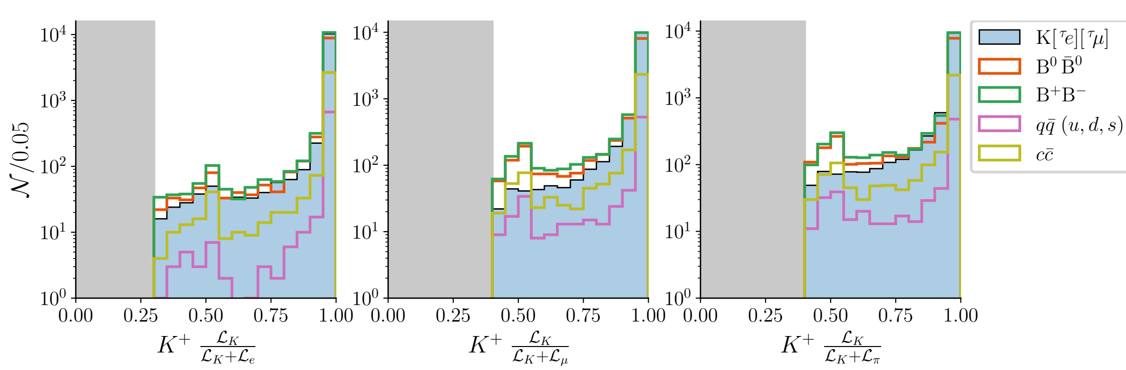

Figure A.3: Likelihood variables used for $K^{\pm}$ selection in the $K[{}^\tau e][{}^\tau\mu]$ mode.

Table A.3: Average candidate multiplicity per event/retention for selections made to the signal kaon particle list in the $K[^{\tau}e][^{\tau}\mu]$ mode. The function $f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T})$ is a dynamic ellipsoid around the interaction point with size inversely proportional to transverse momentum.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u},\,d\bar{d},\,s\bar{s}$	$c\bar{c}$
			10.5 / 0.992	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in CDC Acceptance	-	9.7 / 0.992	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_e} > 0.3$	-	3.2 / 0.979	4.1 / 0.977	4.0 / 0.979	3.5 / 0.928	3.6 / 0.944
3	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\mu} > 0.4$	-	2.1 / 0.926	2.5 / 0.859	2.4 / 0.879	2.1 / 0.737	2.2 / 0.822
4	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\pi} > 0.4$	-	2.0 / 0.916	2.3 / 0.838	2.3 / 0.861	2.0 / 0.703	2.1 / 0.799
5	$f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T}) < 4^2$	cm	1.6 / 0.835	1.6 / 0.684	1.7 / 0.732	1.5 / 0.552	1.6 / 0.653

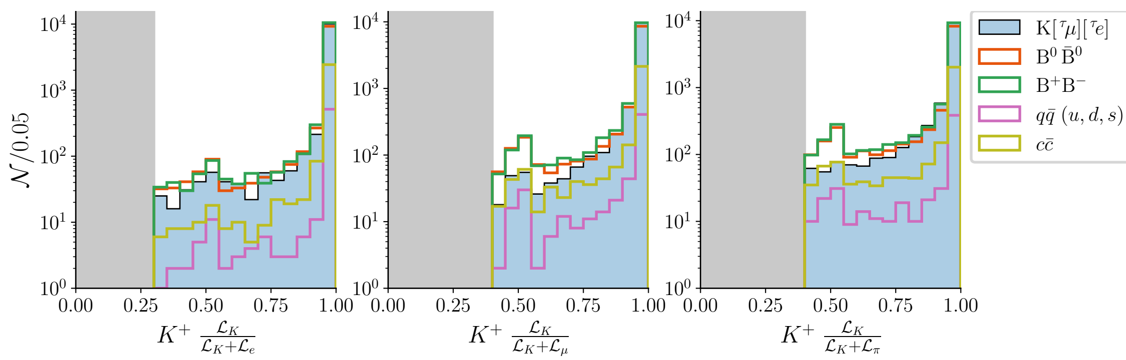

Figure A.4: Likelihood variables used for $K^{\pm}$ selection in the $K[^{\tau}\mu][^{\tau}e]$ mode.

Table A.4: Average candidate multiplicity per event/retention for selections made to the signal kaon particle list in the $K[^{\tau}\mu][^{\tau}e]$ mode. The function $f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T})$ is a dynamic ellipsoid around the interaction point with size inversely proportional to transverse momentum.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u},\,d\bar{d},\,s\bar{s}$	$c\bar{c}$
			10.5 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in CDC Acceptance	-	9.8 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_e} > 0.3$	-	3.2 / 0.979	4.1 / 0.977	4.0 / 0.979	3.5 / 0.928	3.6 / 0.944
3	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\mu} > 0.4$	-	2.1 / 0.927	2.5 / 0.859	2.4 / 0.879	2.1 / 0.737	2.2 / 0.822
4	$\frac{\mathcal{L}_K}{\mathcal{L}_K+\mathcal{L}_\pi} > 0.4$	-	2.0 / 0.915	2.3 / 0.838	2.3 / 0.861	2.0 / 0.703	2.1 / 0.799
5	$f(\mathrm{dr}, \mathrm{dz}, \mathrm{p_T}) < 4^2$	cm	1.6 / 0.833	1.6 / 0.684	1.7 / 0.732	1.5 / 0.552	1.6 / 0.653

A.2 Leptons

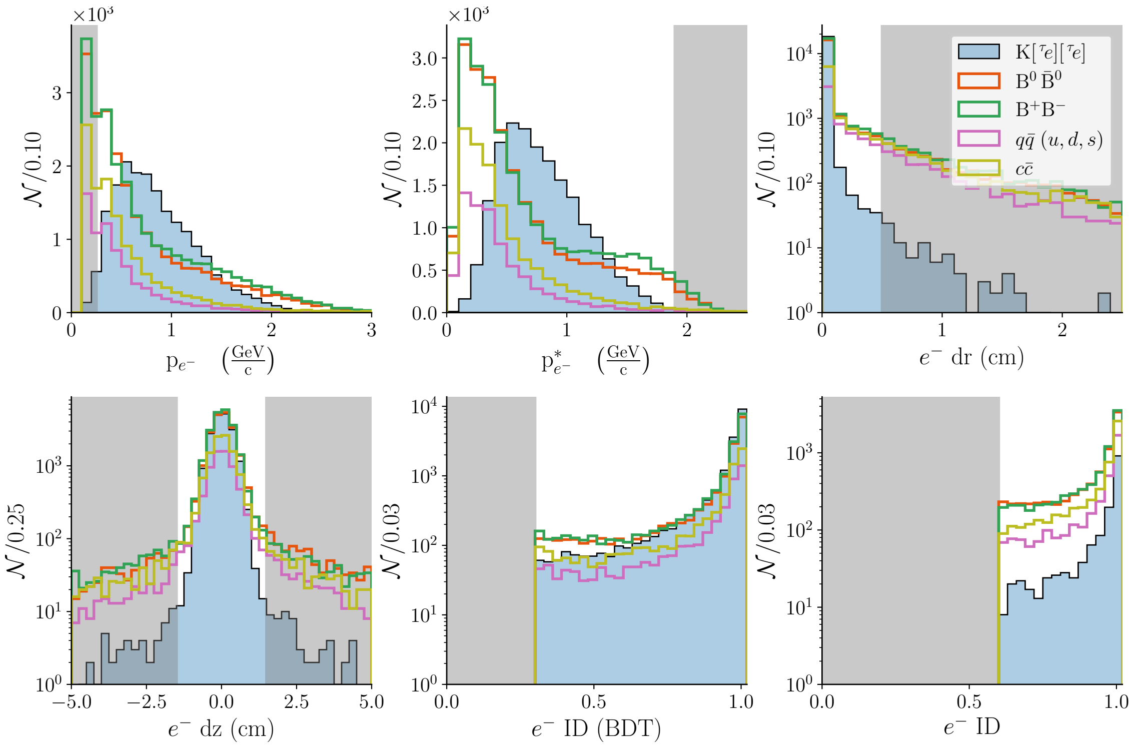

Figure A.5: Variables used for $e^{\pm}$ selection in the $K[{}^\tau e][{}^\tau e]$ mode.

Table A.5: Average candidate multiplicity per event/retention for selections made to the signal electron particle list in the $K[^{\tau}e][^{\tau}e]$ mode.

	Selection	Unit	$B^{\pm}_{\text{sig}}B^{\mp}_{\text{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.6 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.8 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	p > 0.25	$\frac{\text{GeV}}{\text{c}}$	7.1 / 0.991	8.9 / 0.994	8.8 / 0.994	7.3 / 0.951	7.6 / 0.962
3	dr < 0.5	cm	6.0 / 0.991	6.9 / 0.994	6.9 / 0.994	5.9 / 0.951	6.0 / 0.962
4	|dz| < 1.5	cm	5.7 / 0.991	6.5 / 0.994	6.5 / 0.994	5.6 / 0.951	5.7 / 0.962
5	$e^{\pm}$ ID (BDT) > 0.3 or $e^{\pm}$ ID > 0.6	-	1.9 / 0.956	1.2 / 0.353	1.2 / 0.342	1.2 / 0.087	1.1 / 0.183
6	$\|\vec{\text{p}}\star\| < 1.9$	$\frac{\text{GeV}}{\text{c}}$	1.8 / 0.955	1.2 / 0.340	1.2 / 0.327	1.2 / 0.083	1.1 / 0.176

Table A.6: Average candidate multiplicity per event/retention for selections made to the signal muon particle list in the $K[^{\tau}\mu][^{\tau}\mu]$ mode.

	Selection	Unit	$B^{\pm}_{\text{sig}}B^{\mp}_{\text{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.7 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	p > 0.6	$\frac{\text{GeV}}{\text{c}}$	3.9 / 0.988	3.9 / 0.986	3.9 / 0.988	3.9 / 0.942	3.9 / 0.951
3	dr < 2.0	cm	3.7 / 0.987	3.6 / 0.982	3.7 / 0.985	3.7 / 0.939	3.6 / 0.948
4	|dz| < 4.0	cm	3.6 / 0.986	3.4 / 0.979	3.5 / 0.982	3.6 / 0.936	3.5 / 0.944
5	$\mu^{\pm}$ ID (BDT) > 0.4 or $\mu^{\pm}$ ID > 0.95	-	1.3 / 0.638	1.1 / 0.203	1.1 / 0.209	1.0 / 0.039	1.0 / 0.086
6	$\|\vec{\text{p}}\star\| < 1.9$	$\frac{\text{GeV}}{\text{c}}$	1.3 / 0.633	1.1 / 0.187	1.1 / 0.190	1.0 / 0.036	1.0 / 0.079

Table A.7: Average candidate multiplicity per event/retention for selections made to the signal electron particle list in the $K[^{\tau}e][^{\tau}\mu]$ mode.

	Selection	Unit	$B^{\pm}_{\text{sig}}B^{\mp}_{\text{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.992	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.7 / 0.992	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	p > 0.25	$\frac{\text{GeV}}{\text{c}}$	7.1 / 0.992	8.9 / 0.994	8.8 / 0.994	7.3 / 0.951	7.6 / 0.962
3	dr < 0.5	cm	6.0 / 0.992	6.9 / 0.994	6.9 / 0.994	5.9 / 0.951	6.0 / 0.962
4	|dz| < 1.5	cm	5.7 / 0.992	6.5 / 0.994	6.5 / 0.994	5.6 / 0.951	5.7 / 0.962
5	$e^{\pm}$ ID (BDT) > 0.3 or $e^{\pm}$ ID > 0.6	-	1.2 / 0.817	1.2 / 0.354	1.2 / 0.342	1.2 / 0.087	1.1 / 0.183
6	$\|\vec{\text{p}}\star\| < 1.9$	$\frac{\text{GeV}}{\text{c}}$	1.2 / 0.814	1.2 / 0.340	1.2 / 0.326	1.2 / 0.083	1.1 / 0.176

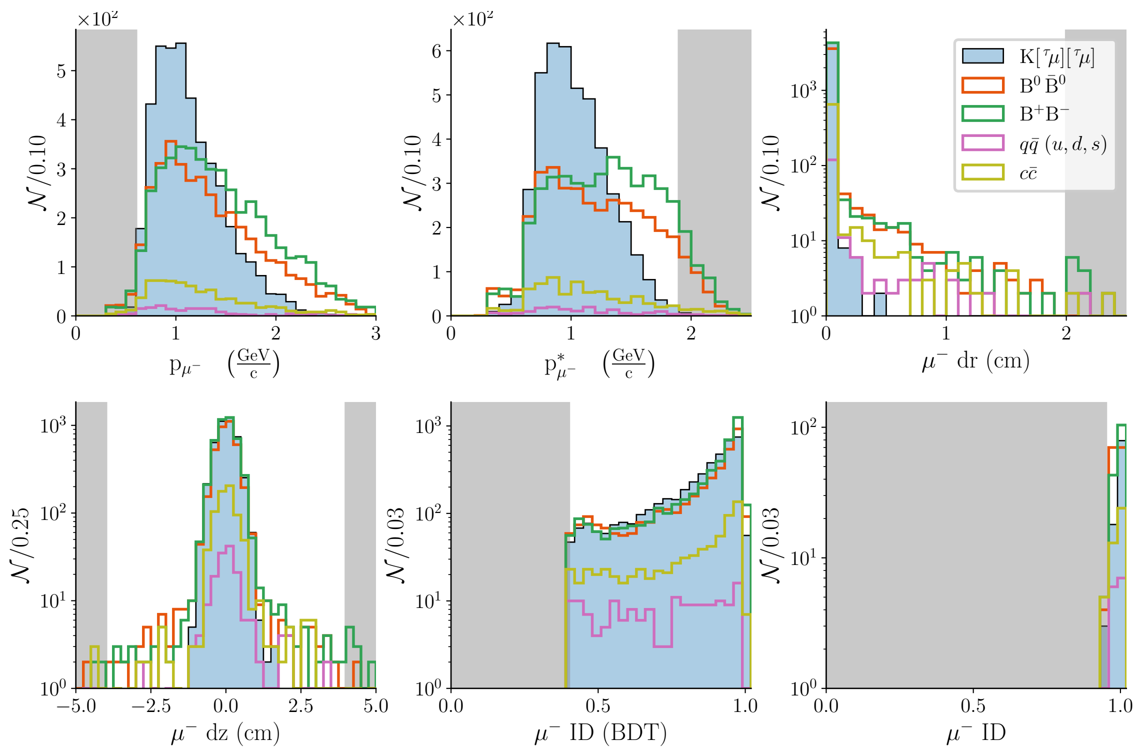

Figure A.6: Variables used for $\mu^{\pm}$ selection in the $K[^{\tau}\mu][^{\tau}\mu]$ mode.

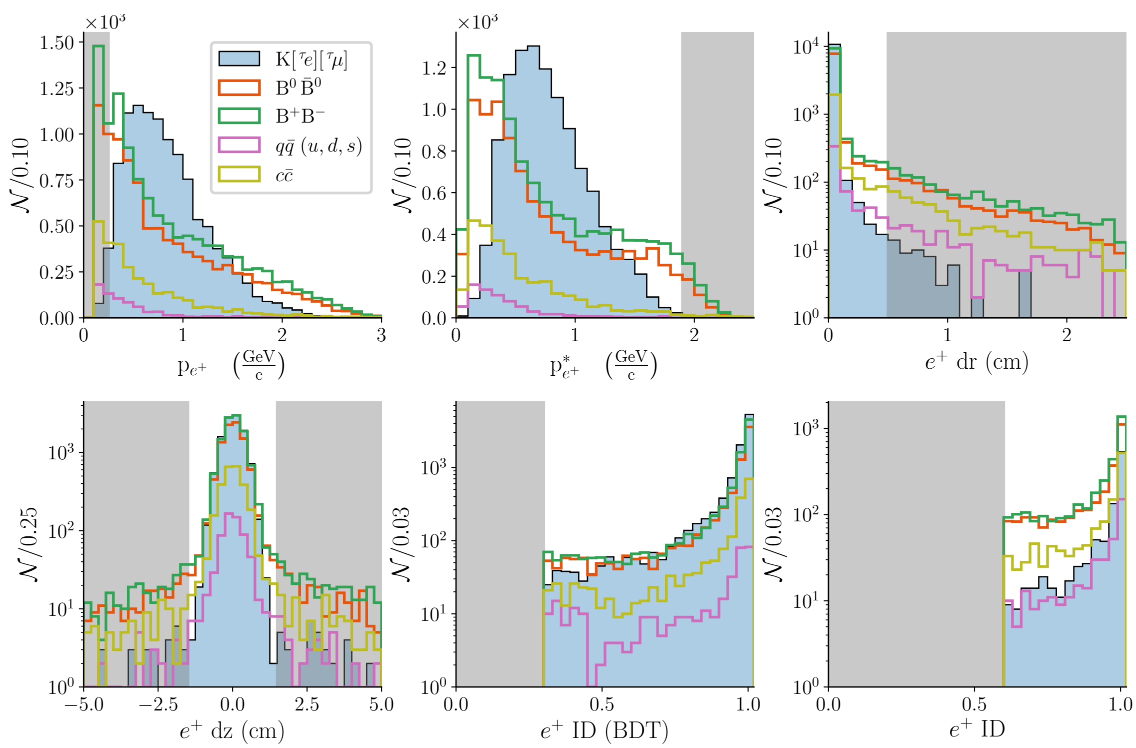

Figure A.7: Variables used for $e^{\pm}$ selection in the $K[^{\tau}e][^{\tau}\mu]$ mode.

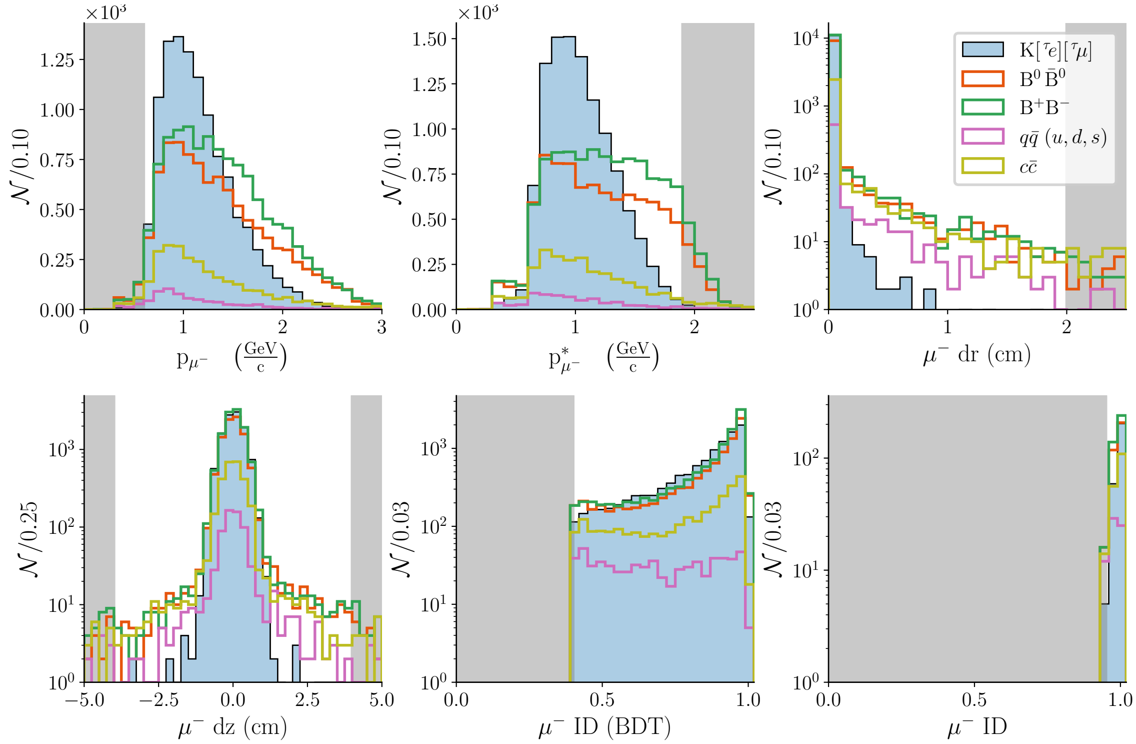

Figure A.8: Variables used for $\mu^{\mp}$ selection in the $K[{}^{\tau}e][{}^{\tau}\mu]$ mode.

Table A.8: Average candidate multiplicity per event/retention for selections made to the signal muon particle list in the $K[{}^{\tau}e][{}^{\tau}\mu]$ mode.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.992	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.8 / 0.992	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	p > 0.6	$\frac{\mathrm{GeV}}{\mathrm{c}}$	3.9 / 0.987	3.9 / 0.986	3.9 / 0.988	3.9 / 0.942	3.9 / 0.951
3	dr < 2.0	cm	3.7 / 0.987	3.6 / 0.982	3.7 / 0.985	3.7 / 0.939	3.6 / 0.948
4	|dz| < 4.0	cm	3.6 / 0.985	3.4 / 0.979	3.5 / 0.982	3.6 / 0.936	3.5 / 0.944
5	$\mu^{\pm}$ ID (BDT) > 0.4 or $\mu^{\pm}$ ID > 0.95	-	1.1 / 0.438	1.1 / 0.203	1.1 / 0.209	1.0 / 0.039	1.0 / 0.086
6	$\lvert\vec{\mathrm{p}}\star\rvert < 1.9$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	1.1 / 0.430	1.1 / 0.187	1.1 / 0.190	1.0 / 0.036	1.0 / 0.079

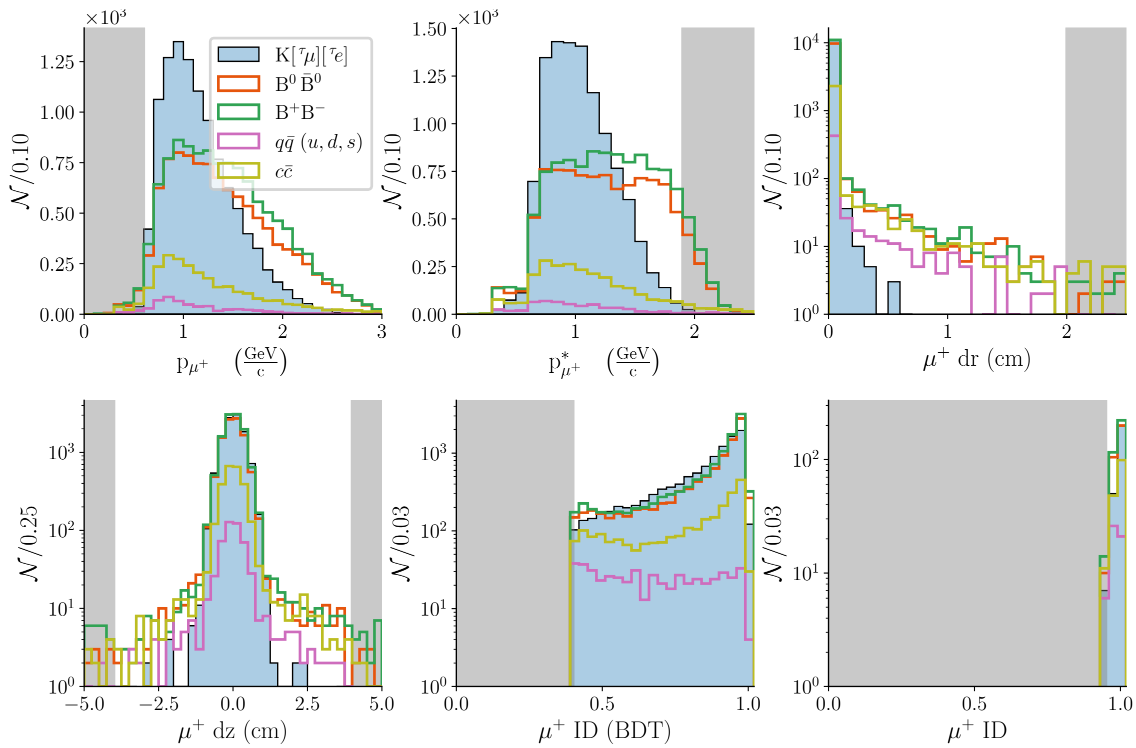

Figure A.9: Variables used for $\mu^\pm$ selection in the $K[^\tau\mu][^\tau e]$ mode.

Table A.9: Average candidate multiplicity per event/retention for selections made to the signal muon particle list in the $K[^\tau\mu][^\tau e]$ mode.

	Selection	Unit	$B^\pm_{\mathrm{sig}}B^\mp_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.8 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\mathrm{p} > 0.6$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	3.9 / 0.988	3.9 / 0.986	3.9 / 0.988	3.9 / 0.942	3.9 / 0.951
3	$\mathrm{dr} < 2.0$	cm	3.7 / 0.987	3.6 / 0.982	3.7 / 0.985	3.7 / 0.939	3.6 / 0.948
4	$\lvert\mathrm{dz}\rvert < 4.0$	cm	3.6 / 0.985	3.4 / 0.979	3.5 / 0.982	3.6 / 0.936	3.5 / 0.944
5	$\mu^\pm$ ID (BDT) > 0.4 or $\mu^\pm$ ID > 0.95	-	1.1 / 0.434	1.1 / 0.203	1.1 / 0.209	1.0 / 0.039	1.0 / 0.086
6	$\lvert\vec{\mathrm{p}}\star\rvert < 1.9$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	1.1 / 0.427	1.1 / 0.187	1.1 / 0.190	1.0 / 0.036	1.0 / 0.079

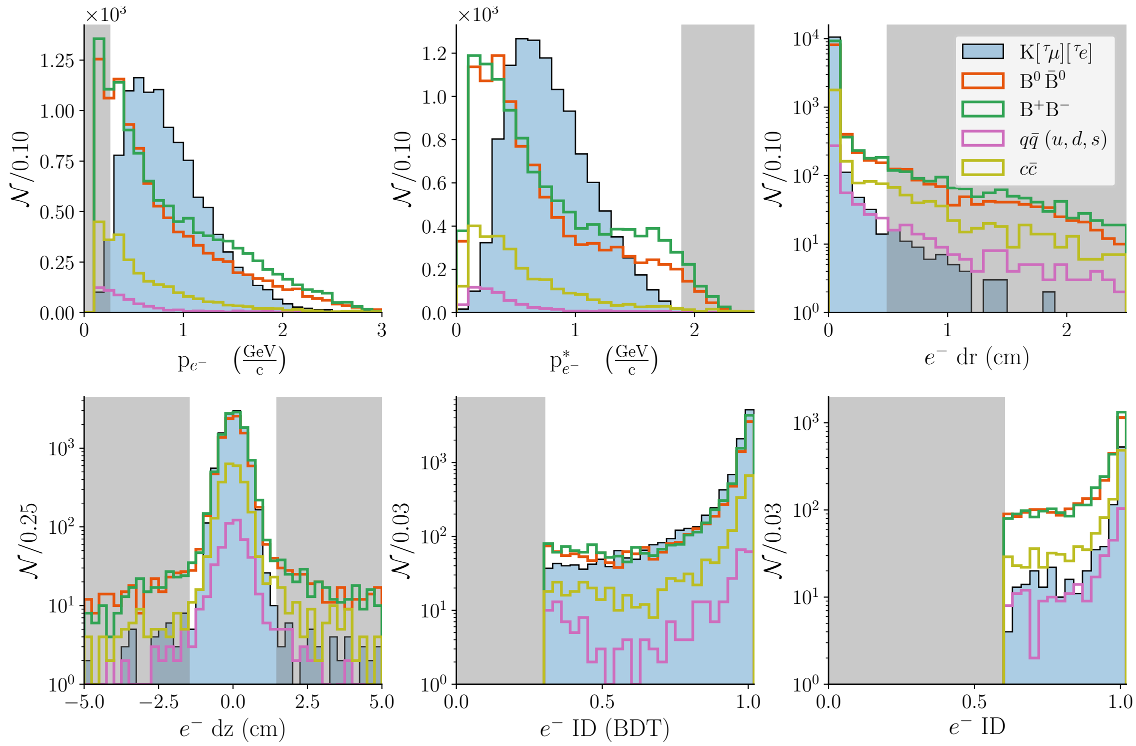

Figure A.10: Variables used for $e^\mp$ selection in the $K[^\tau\mu][^\tau e]$ mode.

Table A.10: Average candidate multiplicity per event/retention for selections made to the signal electron particle list in the $K[^\tau\mu][^\tau e]$ mode.

	Selection	Unit	$B^\pm_{\mathrm{sig}}B^\mp_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			10.5 / 0.991	15.7 / 0.994	14.6 / 0.994	10.7 / 0.951	11.8 / 0.962
1	θ in KLM Acceptance	-	9.8 / 0.991	14.1 / 0.994	13.2 / 0.994	10.0 / 0.951	10.8 / 0.962
2	$\mathrm{p} > 0.25$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	7.1 / 0.991	8.9 / 0.994	8.8 / 0.994	7.3 / 0.951	7.6 / 0.962
3	$\mathrm{dr} < 0.5$	cm	6.0 / 0.991	6.9 / 0.994	6.9 / 0.994	5.9 / 0.951	6.0 / 0.962
4	$\lvert\mathrm{dz}\rvert < 1.5$	cm	5.8 / 0.991	6.5 / 0.994	6.5 / 0.994	5.6 / 0.951	5.7 / 0.962
5	$e^\pm$ ID (BDT) > 0.3 or $e^\pm$ ID > 0.6	-	1.2 / 0.819	1.2 / 0.354	1.2 / 0.342	1.2 / 0.087	1.1 / 0.183
6	$\lvert\vec{\mathrm{p}}\star\rvert < 1.9$	$\frac{\mathrm{GeV}}{\mathrm{c}}$	1.2 / 0.815	1.2 / 0.340	1.2 / 0.327	1.2 / 0.083	1.1 / 0.176

A.3 Signal B Mesons

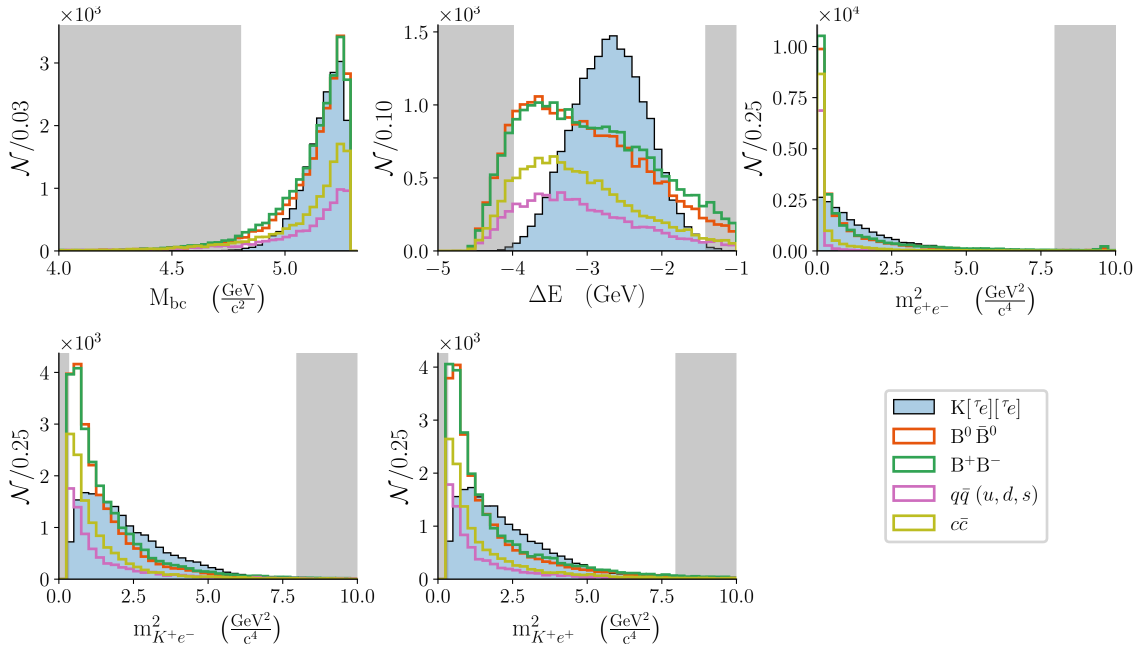

Figure A.11: Cuts placed on $B^{\pm}$ candidates in the $K[^{\tau}e][^{\tau}e]$ mode.

Table A.11: Average candidate multiplicity per event/retention for selections made during signal $B^{\pm}$ reconstruction in the $K[^{\tau}e][^{\tau}e]$ mode.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u},\ d\bar{d},\ s\bar{s}$	$c\bar{c}$
			1.1 / 0.280	1.2 / 0.011	1.2 / 0.013	$1.1 / 2.89 \times 10^{-3}$	$1.1 / 4.93 \times 10^{-3}$
1	$\lvert\ell^+\ell^-\rvert^2 < 8$	$\frac{\mathrm{GeV}^2}{c^4}$	1.1 / 0.279	1.2 / 0.011	1.2 / 0.012	$1.1 / 2.89 \times 10^{-3}$	$1.1 / 4.91 \times 10^{-3}$
2	$0.3 < \lvert K^+\ell^-\rvert^2 < 8$	$\frac{\mathrm{GeV}^2}{c^4}$	1.1 / 0.278	1.2 / 0.011	1.2 / 0.012	$1.1 / 2.80 \times 10^{-3}$	$1.1 / 4.84 \times 10^{-3}$
3	$0.3 < \lvert K^+\ell^+\rvert^2 < 8$	$\frac{\mathrm{GeV}^2}{c^4}$	1.1 / 0.277	1.2 / 0.011	1.2 / 0.012	$1.1 / 2.72 \times 10^{-3}$	$1.1 / 4.69 \times 10^{-3}$
4	$-4.0 < \Delta\mathrm{E} < -1.4$	GeV	1.1 / 0.275	1.2 / 0.010	1.2 / 0.011	$1.1 / 2.61 \times 10^{-3}$	$1.1 / 4.54 \times 10^{-3}$
5	$\mathrm{M_{bc}} > 4.8$	$\frac{\mathrm{GeV}}{c^2}$	1.1 / 0.268	$1.2 / 9.80 \times 10^{-3}$	1.1 / 0.010	$1.1 / 2.26 \times 10^{-3}$	$1.1 / 4.11 \times 10^{-3}$

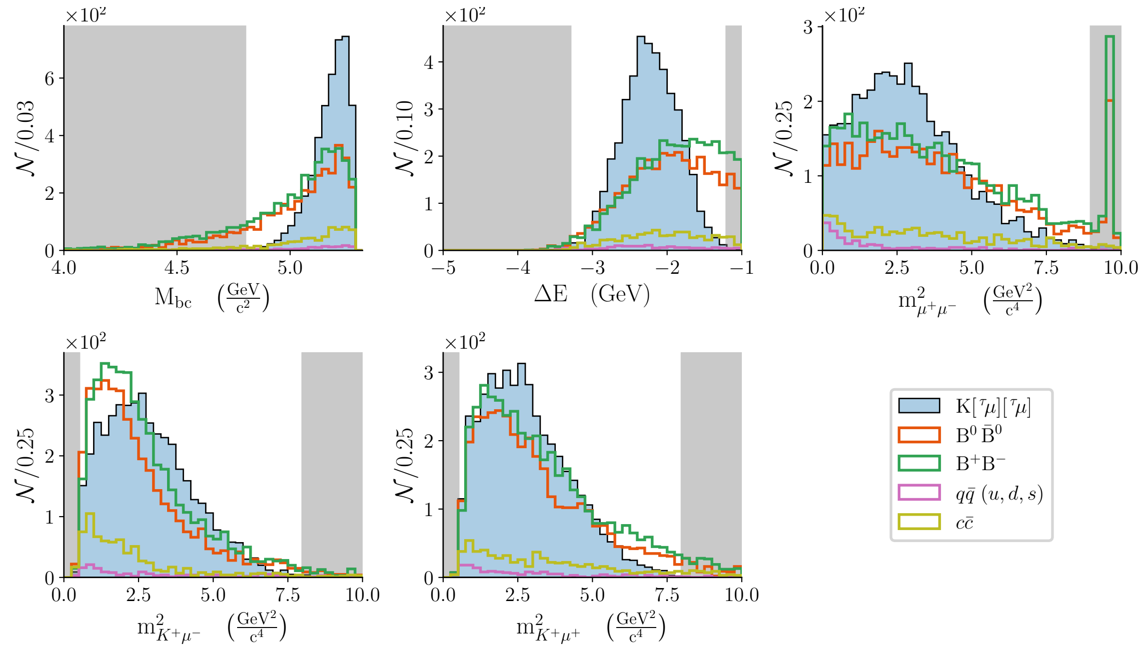

Figure A.12: Cuts placed on $B^\pm$ candidates in the $K[^\tau\mu][^\tau\mu]$ mode.

Table A.12: Average candidate multiplicity per event/retention for selections made during signal $B^\pm$ reconstruction in the $K[^\tau\mu][^\tau\mu]$ mode.

	Selection	Unit	$B^\pm_{\rm sig}B^\mp_{\rm tag}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u},\,d\bar{d},\,s\bar{s}$	$c\bar{c}$
			1.1 / 0.066	$1.1\,/\,2.88\times10^{-3}$	$1.1\,/\,3.50\times10^{-3}$	$1.1\,/\,1.36\times10^{-4}$	$1.1\,/\,6.18\times10^{-4}$
1	$\lvert\ell^+\ell^-\rvert^2<9$	$\frac{\rm GeV^2}{c^4}$	1.1 / 0.066	$1.1\,/\,2.71\times10^{-3}$	$1.1\,/\,3.23\times10^{-3}$	$1.1\,/\,1.36\times10^{-4}$	$1.1\,/\,5.90\times10^{-4}$
2	$0.5<\lvert K^+\ell^-\rvert^2<8$	$\frac{\rm GeV^2}{c^4}$	1.1 / 0.066	$1.1\,/\,2.66\times10^{-3}$	$1.1\,/\,3.17\times10^{-3}$	$1.1\,/\,1.24\times10^{-4}$	$1.1\,/\,5.78\times10^{-4}$
3	$0.5<\lvert K^+\ell^+\rvert^2<8$	$\frac{\rm GeV^2}{c^4}$	1.1 / 0.065	$1.1\,/\,2.59\times10^{-3}$	$1.1\,/\,3.09\times10^{-3}$	$1.1\,/\,1.12\times10^{-4}$	$1.1\,/\,5.16\times10^{-4}$
4	$-3.3<\Delta\mathrm{E}<-1.2$	GeV	1.1 / 0.064	$1.1\,/\,2.33\times10^{-3}$	$1.1\,/\,2.77\times10^{-3}$	$1.1\,/\,1.00\times10^{-4}$	$1.1\,/\,4.98\times10^{-4}$
5	$\mathrm{M_{bc}}>4.8$	$\frac{\rm GeV}{c^2}$	1.1 / 0.061	$1.1\,/\,2.08\times10^{-3}$	$1.1\,/\,2.37\times10^{-3}$	$1.1\,/\,8.00\times10^{-5}$	$1.1\,/\,4.52\times10^{-4}$

Table A.13: Average candidate multiplicity per event/retention for selections made during signal $B^\pm$ reconstruction in the $K[^\tau e][^\tau\mu]$ mode.

	Selection	Unit	$B^\pm_{\rm sig}B^\mp_{\rm tag}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u},\,d\bar{d},\,s\bar{s}$	$c\bar{c}$
			1.1 / 0.088	$1.1\,/\,4.64\times10^{-3}$	$1.1\,/\,5.64\times10^{-3}$	$1.2\,/\,1.92\times10^{-4}$	$1.2\,/\,9.84\times10^{-4}$
1	$\lvert\ell^+\ell^-\rvert^2<9$	$\frac{\rm GeV^2}{c^4}$	1.1 / 0.088	$1.1\,/\,4.56\times10^{-3}$	$1.1\,/\,5.51\times10^{-3}$	$1.2\,/\,1.90\times10^{-4}$	$1.2\,/\,9.70\times10^{-4}$
2	$0.5<\lvert K^+\ell^-\rvert^2<8$	$\frac{\rm GeV^2}{c^4}$	1.1 / 0.088	$1.1\,/\,4.47\times10^{-3}$	$1.1\,/\,5.39\times10^{-3}$	$1.1\,/\,1.74\times10^{-4}$	$1.2\,/\,9.42\times10^{-4}$
3	$0.3<\lvert K^+\ell^+\rvert^2<8$	$\frac{\rm GeV^2}{c^4}$	1.1 / 0.088	$1.1\,/\,4.42\times10^{-3}$	$1.1\,/\,5.27\times10^{-3}$	$1.1\,/\,1.72\times10^{-4}$	$1.2\,/\,9.28\times10^{-4}$
4	$-3.7<\Delta\mathrm{E}<-1.3$	GeV	1.1 / 0.087	$1.1\,/\,4.13\times10^{-3}$	$1.1\,/\,4.84\times10^{-3}$	$1.1\,/\,1.72\times10^{-4}$	$1.1\,/\,8.90\times10^{-4}$
5	$\mathrm{M_{bc}}>4.8$	$\frac{\rm GeV}{c^2}$	1.1 / 0.084	$1.1\,/\,3.73\times10^{-3}$	$1.1\,/\,4.29\times10^{-3}$	$1.1\,/\,1.54\times10^{-4}$	$1.1\,/\,8.14\times10^{-4}$

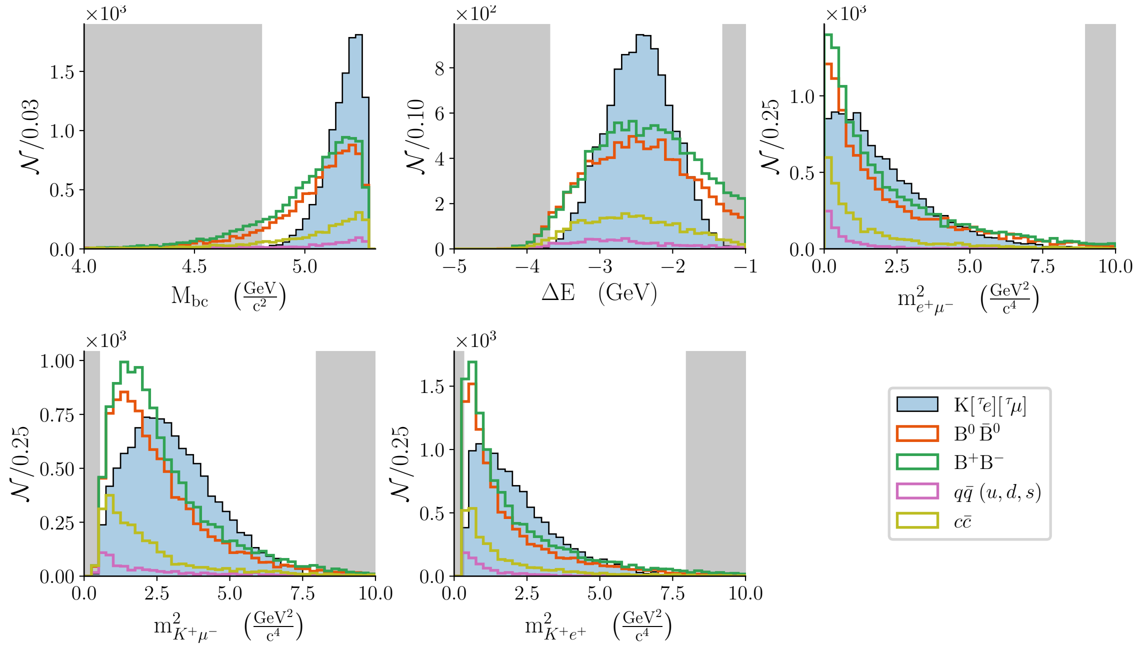

Figure A.13: Cuts placed on $B^{\pm}$ candidates in the $K[^{\tau}e][^{\tau}\mu]$ mode.

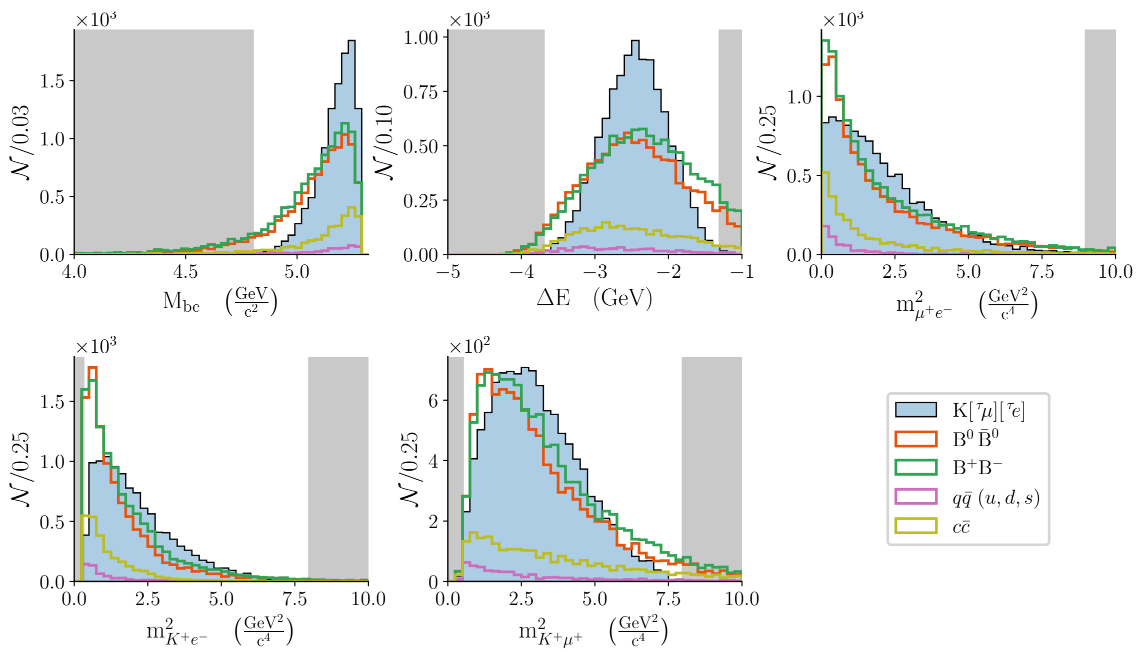

Figure A.14: Cuts placed on $B^{\pm}$ candidates in the $K[^{\tau}\mu][^{\tau}e]$ mode.

Table A.14: Average candidate multiplicity per event/retention for selections made during signal $B^{\pm}$ reconstruction in the $K[{}^{\tau}\mu][{}^{\tau}e]$ mode.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			1.1 / 0.076	$1.2\,/\,4.65\times10^{-3}$	$1.1\,/\,5.59\times10^{-3}$	$1.1\,/\,1.34\times10^{-4}$	$1.1\,/\,1.05\times10^{-3}$
1	$\lvert\ell^+\ell^-\rvert^2 < 9$	$\frac{\mathrm{GeV}^2}{c^4}$	1.1 / 0.076	$1.2\,/\,4.58\times10^{-3}$	$1.1\,/\,5.48\times10^{-3}$	$1.1\,/\,1.34\times10^{-4}$	$1.1\,/\,1.05\times10^{-3}$
2	$0.3 < \lvert K^+\ell^-\rvert^2 < 8$	$\frac{\mathrm{GeV}^2}{c^4}$	1.1 / 0.076	$1.2\,/\,4.55\times10^{-3}$	$1.1\,/\,5.42\times10^{-3}$	$1.1\,/\,1.28\times10^{-4}$	$1.1\,/\,1.03\times10^{-3}$
3	$0.5 < \lvert K^+\ell^+\rvert^2 < 8$	$\frac{\mathrm{GeV}^2}{c^4}$	1.1 / 0.076	$1.1\,/\,4.42\times10^{-3}$	$1.1\,/\,5.22\times10^{-3}$	$1.1\,/\,1.18\times10^{-4}$	$1.1\,/\,9.14\times10^{-4}$
4	$-3.7 < \Delta\mathrm{E} < -1.3$	GeV	1.1 / 0.075	$1.2\,/\,4.16\times10^{-3}$	$1.1\,/\,4.83\times10^{-3}$	$1.1\,/\,1.18\times10^{-4}$	$1.1\,/\,8.96\times10^{-4}$
5	$\mathrm{M_{bc}} > 4.8$	$\frac{\mathrm{GeV}}{c^2}$	1.1 / 0.075	$1.1\,/\,3.87\times10^{-3}$	$1.1\,/\,4.40\times10^{-3}$	$1.1\,/\,1.08\times10^{-4}$	$1.1\,/\,8.44\times10^{-4}$

A.4 Inclusive Hadronic Tag B Mesons

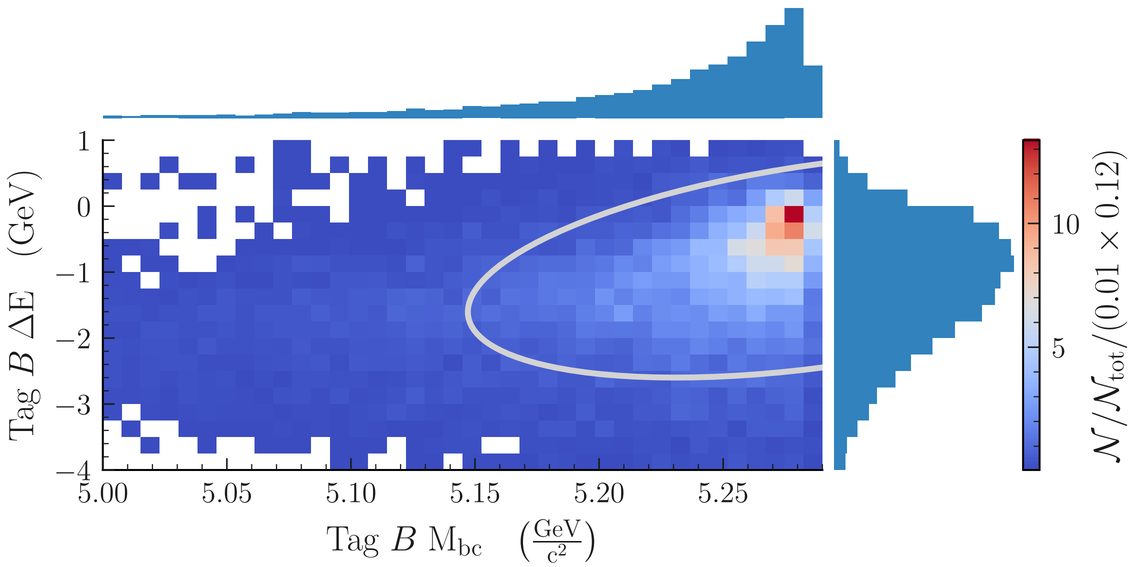

Figure A.15: 2D cuts placed on the inclusive tag $B^{\mp}$ candidates in the $K[^{\tau}e][^{\tau}e]$ mode.

Table A.15: Average candidate multiplicity per event/retention for event selections made to the inclusive tag $B^{\mp}$ in the $K[^{\tau}e][^{\tau}e]$ mode. The function $f(\mathrm{M_{bc}}, \Delta\mathrm{E})$ is an ellipse rotated by 35 mrad and can be seen in **??**.

	Selection	Unit	$B^{\pm}_{\rm sig}B^{\mp}_{\rm tag}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			1.5 / 0.387	1.4 / 0.016	1.5 / 0.016	$1.4 / 3.56 \times 10^{-3}$	$1.4 / 6.38 \times 10^{-3}$
1	$f(M_{bc}, \Delta E) < 1$	-	1.3 / 0.274	$1.2 / 4.89 \times 10^{-3}$	$1.2 / 4.82 \times 10^{-3}$	$1.1 / 3.82 \times 10^{-4}$	$1.1 / 1.22 \times 10^{-3}$

Table A.16: Average candidate multiplicity per event/retention for event selections made to the inclusive tag $B^{\mp}$ in the $K[^{\tau}\mu][^{\tau}\mu]$ mode. The function $f(\mathrm{M_{bc}}, \Delta\mathrm{E})$ is an ellipse rotated by 35 mrad and can be seen in **??**.

	Selection	Unit	$B^{\pm}_{\rm sig}B^{\mp}_{\rm tag}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			1.5 / 0.089	$1.4 / 3.41 \times 10^{-3}$	$1.4 / 3.73 \times 10^{-3}$	$1.3 / 1.40 \times 10^{-4}$	$1.4 / 7.26 \times 10^{-4}$
1	$f(M_{bc}, \Delta E) < 1$	-	1.3 / 0.058	$1.2 / 1.16 \times 10^{-3}$	$1.2 / 1.36 \times 10^{-3}$	$1.2 / 1.80 \times 10^{-5}$	$1.2 / 2.10 \times 10^{-4}$

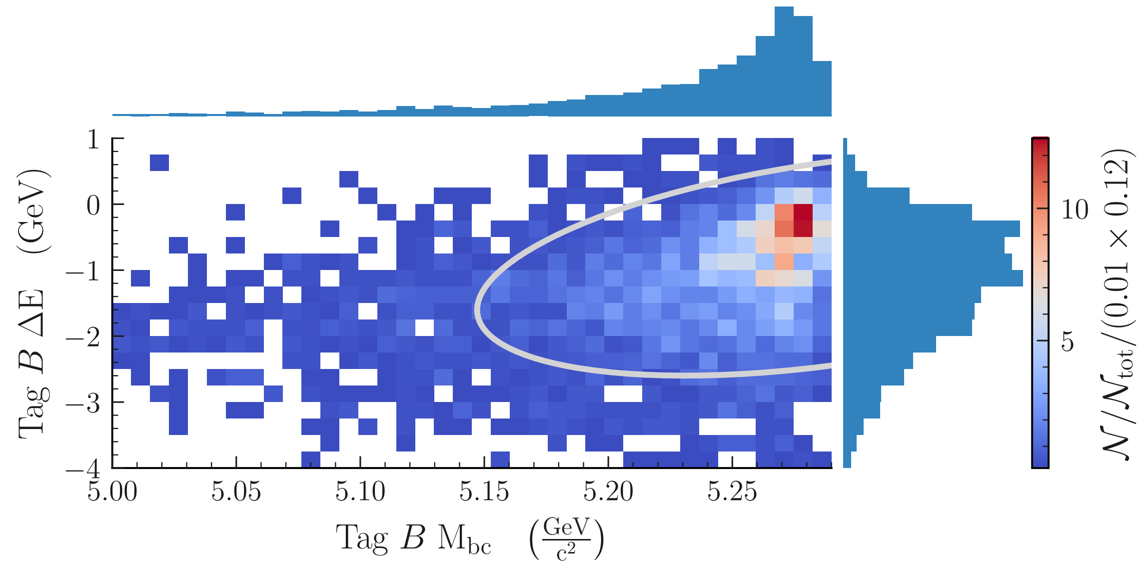

Figure A.16: 2D cuts placed on the inclusive tag $B^{\mp}$ candidates in the $K[^{\tau}\mu][^{\tau}\mu]$ mode.

Table A.17: Average candidate multiplicity per event/retention for event selections made to the inclusive tag $B^{\mp}$ in the $K[^{\tau}e][^{\tau}\mu]$ mode. The function $f(\mathrm{M_{bc}}, \Delta\mathrm{E})$ is an ellipse rotated by 35 mrad and can be seen in 6.13.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			$1.1\,/\,0.164$	$1.1\,/\,7.29\times10^{-3}$	$1.1\,/\,8.44\times10^{-3}$	$1.1\,/\,2.70\times10^{-4}$	$1.1\,/\,1.73\times10^{-3}$
1	$f(M_{bc}, \Delta E) < 1$	-	$1.1\,/\,0.110$	$1.1\,/\,2.28\times10^{-3}$	$1.1\,/\,2.50\times10^{-3}$	$1.0\,/\,2.00\times10^{-5}$	$1.0\,/\,3.74\times10^{-4}$

Table A.18: Average candidate multiplicity per event/retention for event selections made to the inclusive tag $B^{\mp}$ in the $K[^{\tau}\mu][^{\tau}e]$ mode. The function $f(\mathrm{M_{bc}}, \Delta\mathrm{E})$ is an ellipse rotated by 35 mrad and can be seen in **??**.

	Selection	Unit	$B^{\pm}_{\mathrm{sig}}B^{\mp}_{\mathrm{tag}}$	$B^0\,\overline{B}^0$	B^+B^-	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$	$c\bar{c}$
			$1.1\,/\,0.149$	$1.1\,/\,7.76\times10^{-3}$	$1.1\,/\,8.59\times10^{-3}$	$1.1\,/\,2.06\times10^{-4}$	$1.1\,/\,1.63\times10^{-3}$
1	$f(M_{bc}, \Delta E) < 1$	-	$1.1\,/\,0.106$	$1.1\,/\,2.59\times10^{-3}$	$1.1\,/\,2.88\times10^{-3}$	$1.0\,/\,2.00\times10^{-5}$	$1.0\,/\,3.88\times10^{-4}$

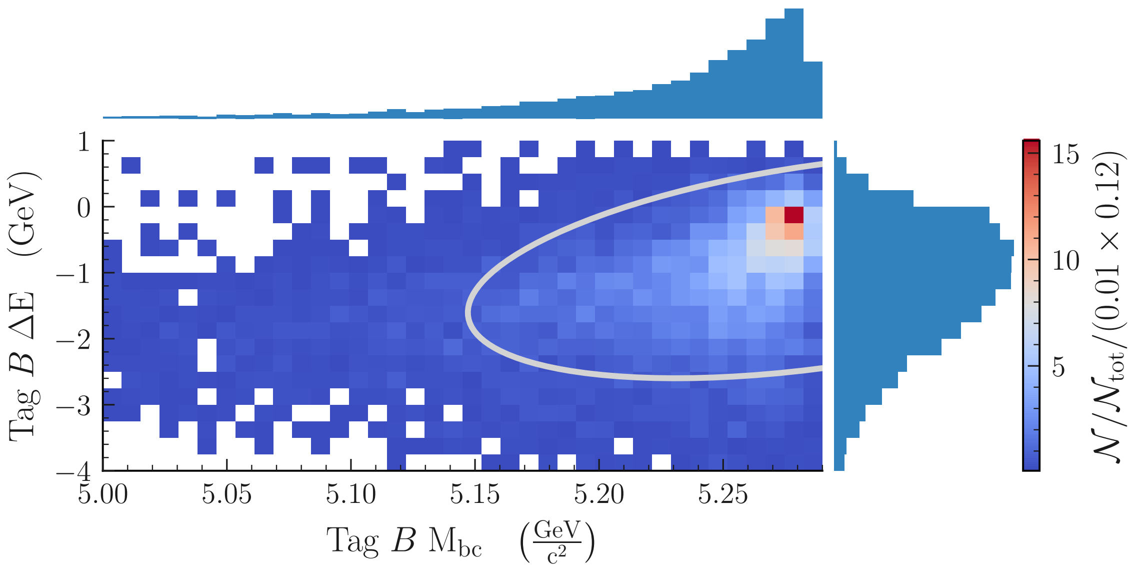

Figure A.17: 2D cuts placed on the inclusive tag $B^{\mp}$ candidates in the $K[^{\tau}e][^{\tau}\mu]$ mode.

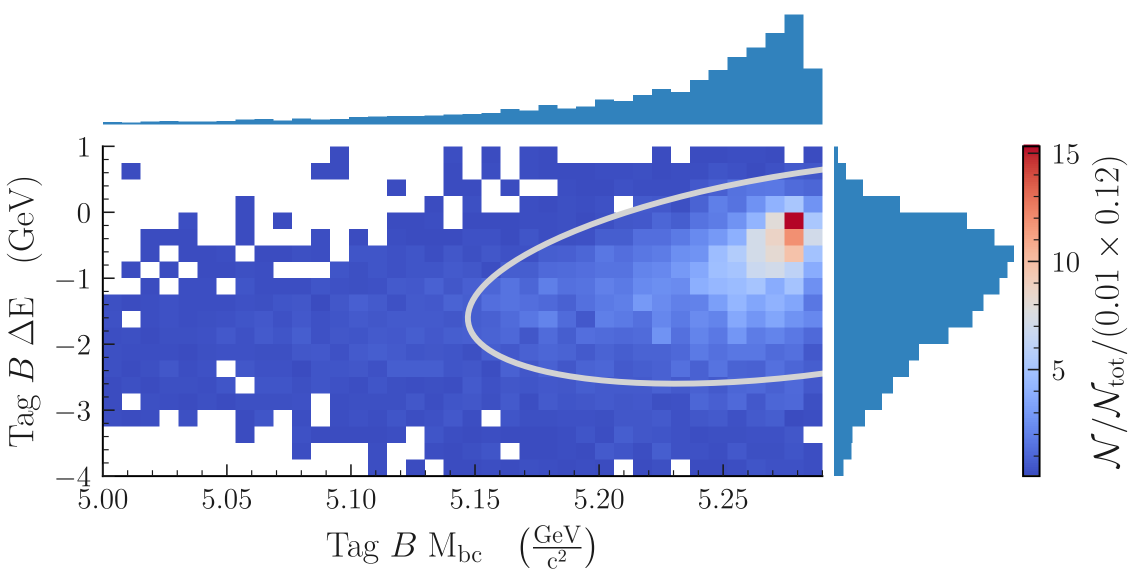

Figure A.18: 2D cuts placed on the inclusive tag $B^{\mp}$ candidates in the $K[^{\tau}\mu][^{\tau}e]$ mode.

APPENDIX B
TRAINING VARIABLES

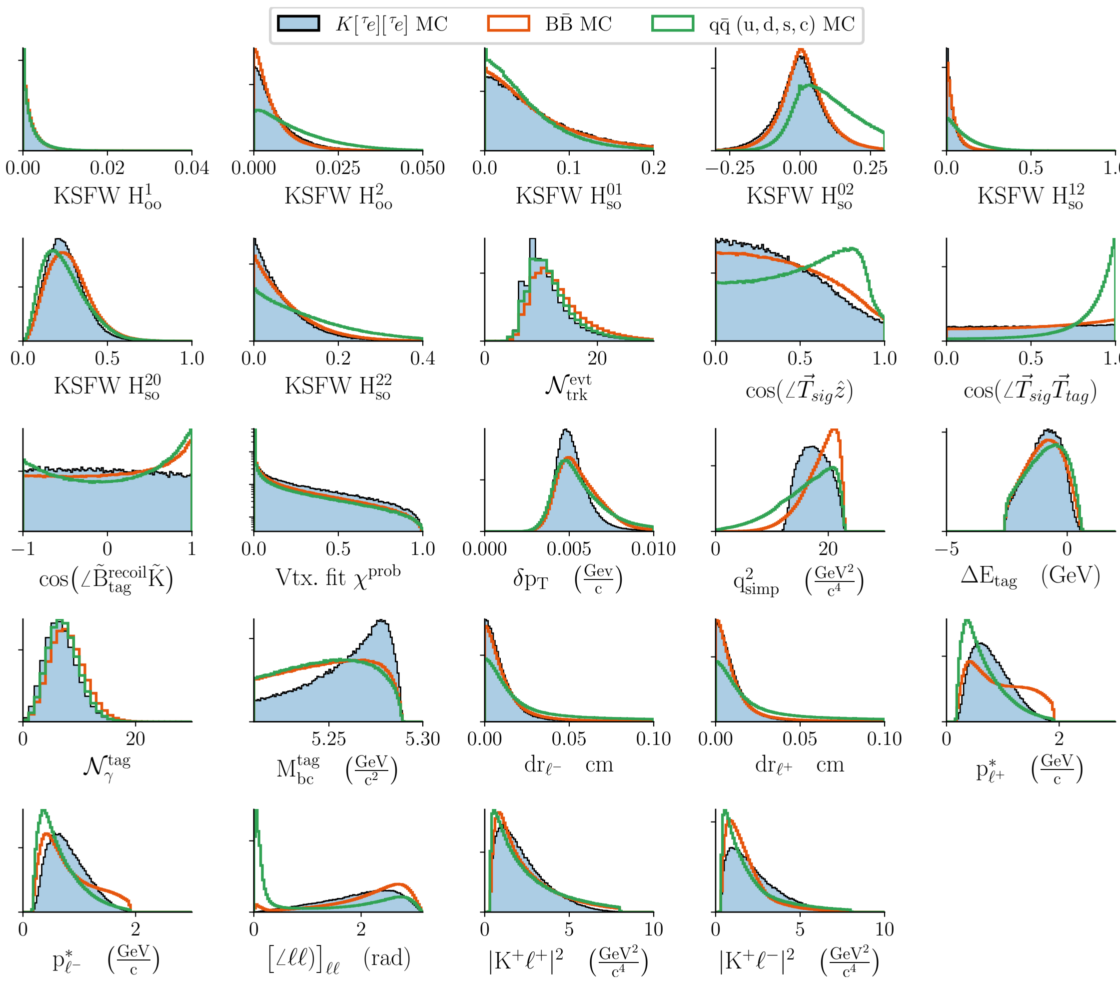

Figure B.1: Variables used for training the BDT to identify signal vs. background in the $K[^{\tau}e][^{\tau}e]$ mode.

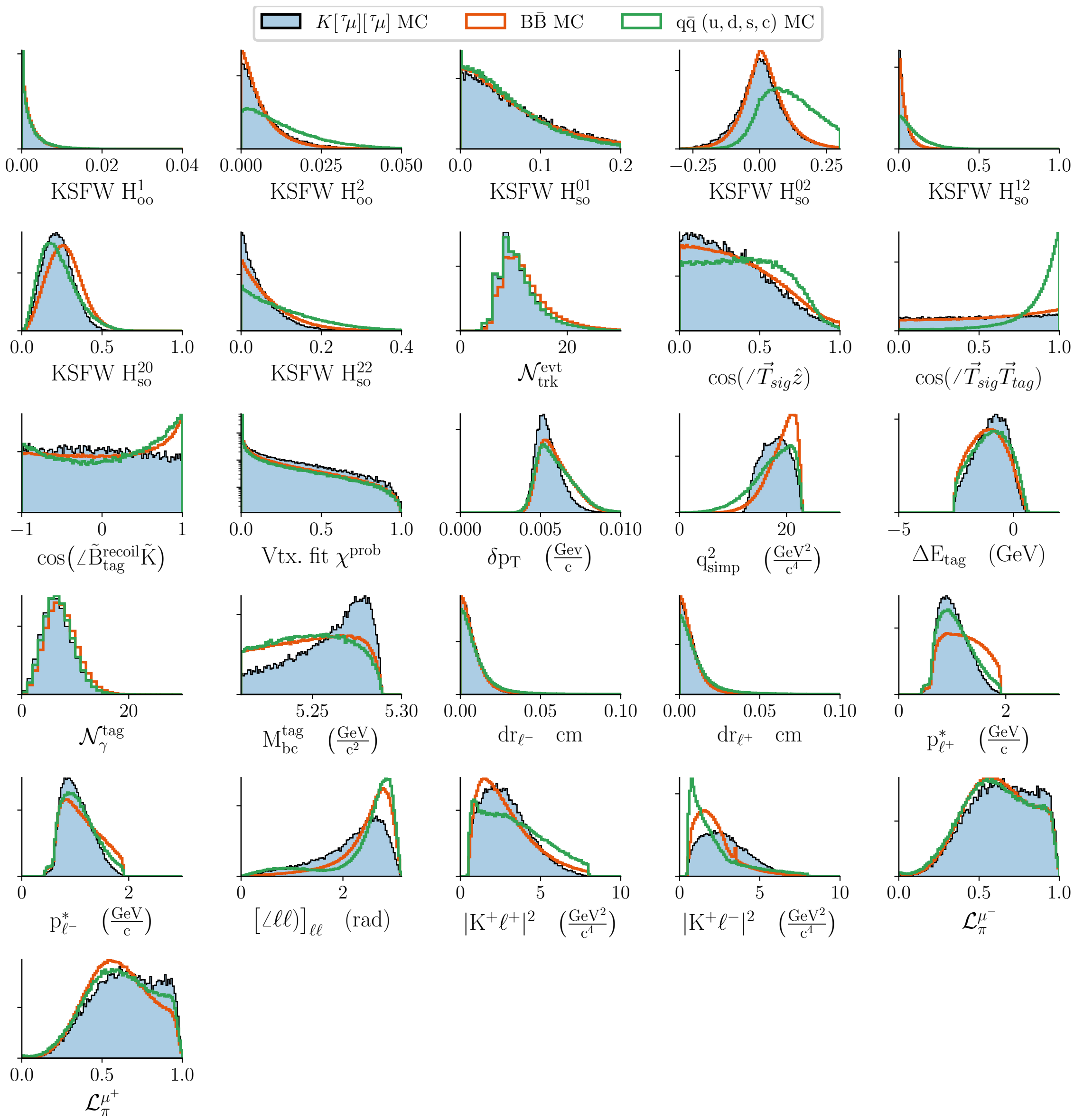

Figure B.2: Variables used for training the BDT to identify signal vs. background in the $K[^\tau\mu][^\tau\mu]$ mode.

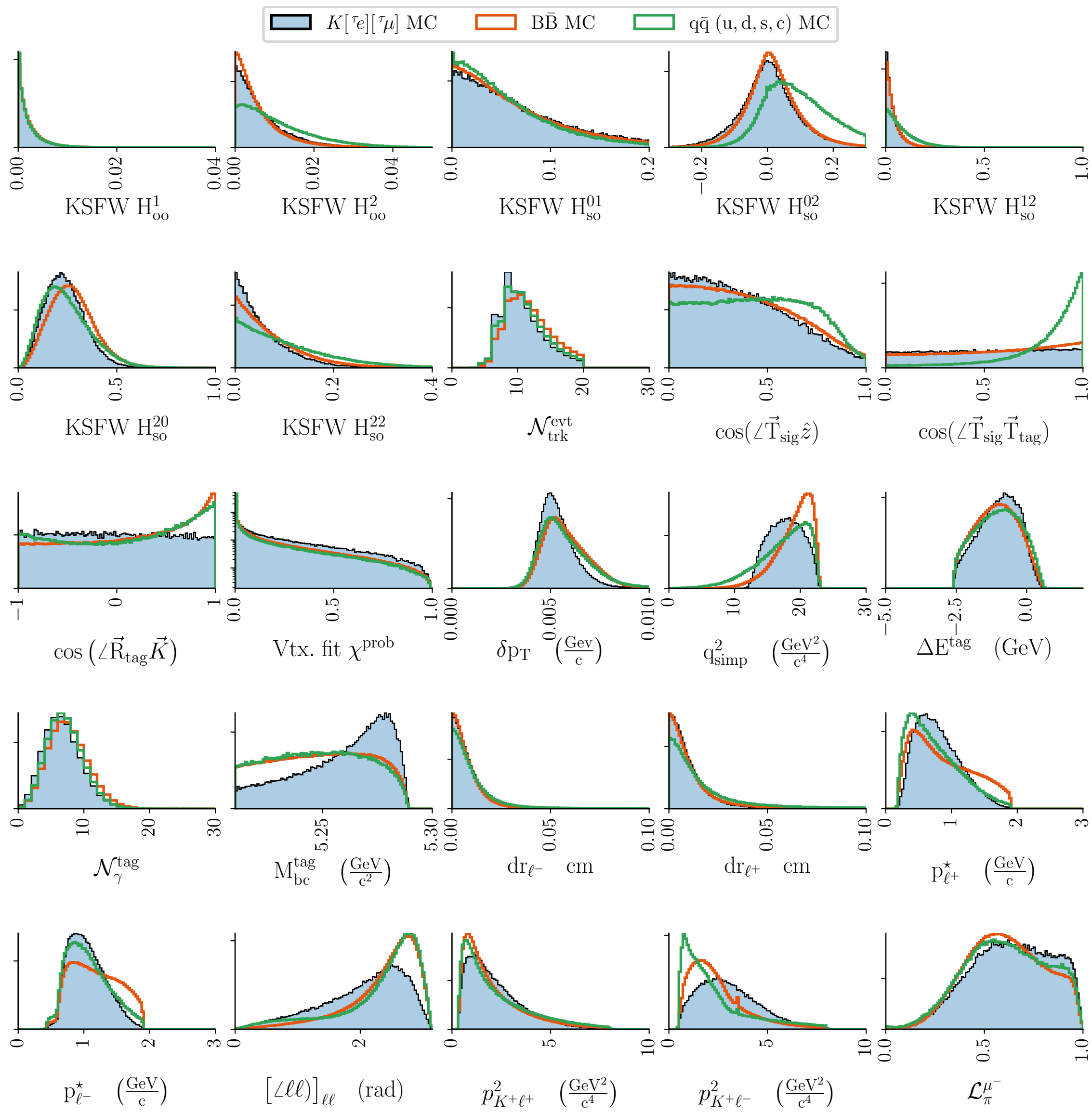

Figure B.3: Variables used for training the BDT to identify signal vs. background in the $K[^{\tau}e][^{\tau}\mu]$ mode.

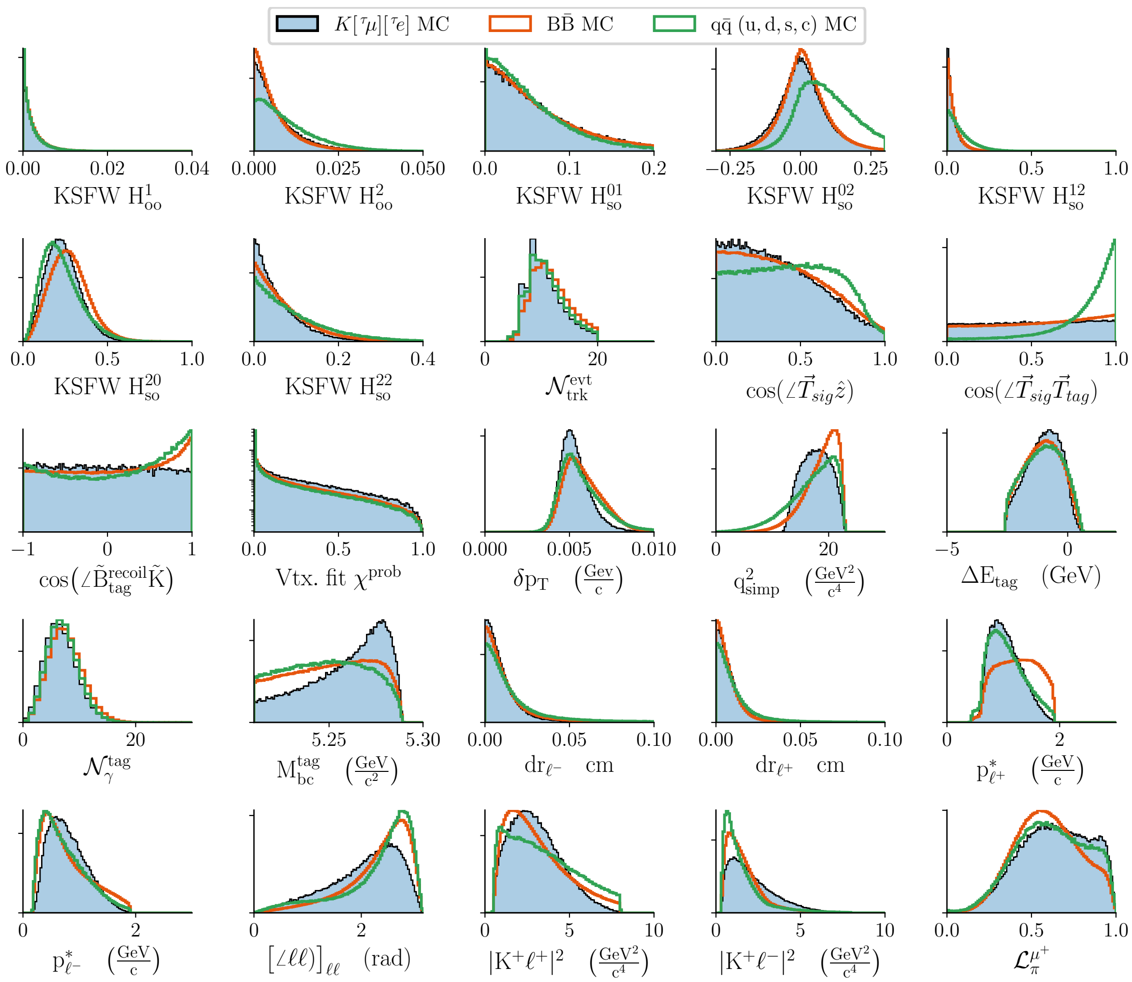

Figure B.4: Variables used for training the BDT to identify signal vs. background in the $K[^\tau\mu][^\tau e]$ mode.

APPENDIX C
ELECTRONIC RESOURCES

Dissertation	GitLab
Defense Presentation	GitLab
Analysis Scripts	GitLab
KLM SCROD Firmware	GitLab
KLM Readout Software	GitLab

BIBLIOGRAPHY

[1] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[2] F. Halzen and A. Martin. *Quarks & Leptons: An Introductory Course in Modern Particle Physics*. 1984.

[3] Antonio Pich. The Standard Model of Electroweak Interactions. In *2010 European School of High Energy Physics*, pages 1–50, 1 2012.

[4] D. Griffiths. *Introduction to Elementary Particles*. 2008.

[5] Duncan Carlsmith. *Particle physics*. Pearson, Boston, USA, 2013.

[6] Michel De Cian. Physics with electroweak penguins at lhcb, 2012.

[7] J. P. Lees et al. Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment. *Phys. Rev. Lett.*, 118(3):031802, 2017.

[8] Vidya Sagar Vobbilisetti. *Search for $B^+ \to K^+\tau\tau$ with Belle and Belle (II) data samples*. PhD thesis, Laboratoire de Physique des 2 Infinis Irène Joliot-Curie, France, U. Paris-Saclay, 2023.

[9] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024.

[10] Christoph Bobeth, Marzia Bordone, Nico Gubernari, Martin Jung, and Danny van Dyk. Lepton-flavour non-universality of $\bar{b} \to d^*\ell\bar{\nu}$ angular distributions in and beyond the standard model. *The European Physical Journal C*, 81(11), November 2021.

[11] R. Aaij et al. Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays. *Phys. Rev. D*, 108:032002, Aug 2023.

[12] Sébastien Descotes-Genon. Searching for New Physics with $b \to s\tau\tau$ Processes. *Acta Phys. Polon. B*, 49:1105–1119, 2018.

[13] F. Abudinén et al. (Belle II Collaboration). Search for $B^+ \to K^+\nu\overline{\nu}$ Decays Using an Inclusive Tagging Method at Belle II. *Phys. Rev. Lett.*, 127:181802, Oct 2021.

[14] J. Grygier, P. Goldenzweig, M. Heck, et al. Search for $b \to h\nu\overline{\nu}$ decays with semileptonic tagging at belle. *Phys. Rev. D*, 96:091101, Nov 2017.

[15] O. Lutz, S. Neubauer, M. Heck, T. Kuhr, A. Zupanc, et al. Search for $b \to h^{(*)}\nu\overline{\nu}$ with the full belle $\Upsilon(4s)$ data sample. *Phys. Rev. D*, 87:111103, Jun 2013.

[16] J. P. Lees et al. Search for $b \to K^{(*)}\nu\overline{\nu}$ and invisible quarkonium decays. *Phys. Rev. D*, 87:112005, Jun 2013.

[17] Herman Feshbach. A unified theory of nuclear reactions. ii. *Annals of Physics*, 19(2):287–313, 1962.

[18] Kenneth G. Wilson and Wolfhart Zimmermann. Operator product expansions and composite field operators in the general framework of quantum field theory. *Communications in Mathematical Physics*, 24, 1972.

[19] Stefan Hollands and Robert M. Wald. The operator product expansion in quantum field theory. 2023.

[20] C. Bobeth, G. Hiller, D. van Dyk, and C. Wacker. The decay $\bar{b} \to \overline{k}\ell^+\ell^-$ at low hadronic recoil and model-independent Δb = 1 constraints. *Journal of High Energy Physics*, 2012(107), 2012.

[21] S. Rai Choudhury, Naveen Gaur, A. S. Cornell, and G. C. Joshi. Supersymmetric effects on the forward-backward asymmetries of $\overrightarrow{B}\, k\tau^+\tau^-$. *Phys. Rev. D*, 69:054018, Mar 2004.

[22] S. Kurokawa and E. Kikutani. Overview of the kekb accelerators. *"Nucl. Instrum. Methods A*, 499(1):1–7, 2003. KEK-B: The KEK B-factory.

[23] A. Abashian *et al.* (The Belle Collaboration). The Belle detector. *Nucl. Instrum. Methods A*, 479(1):117–232, 2002.

[24] J. Brodzicka *et al.* Physics achievements from the Belle experiment. 2012.

[25] G Alimonti et al. The BELLE silicon vertex detector. *Nucl. Instrum. Methods A*, 453(1):71–77, 2000. Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics.

[26] P. A. Čerenkov. Visible radiation produced by electrons moving in a medium with velocities exceeding that of light. *Phys. Rev.*, 52:378–379, Aug 1937.

[27] I. M. Frank and Igor Y. Tamm. Coherent visible radiation of fast electrons passing through matter. *Physics-Uspekhi*, 93:388–393, 1937.

[28] J.W Nam et al. A detailed monte carlo simulation for the belle tof system. *Nucl. Instrum. Methods A*, 491(1):54–68, 2002.

[29] R. Santonico and R. Cardarelli. Development of resistive plate counters. *"Nucl. Instrum. Methods*, 187(2):377–380, 1981.

[30] T. Abe *et al.* Belle II technical design report, 2010.

[31] M. Feindt et al. A hierarchical neurobayes-based algorithm for full reconstruction of b mesons at b factories. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 654(1):432–440, October 2011.

[32] B. others Kronenbitter. Measurement of the branching fraction of $B^+ \to \tau^+\nu_\tau$ decays with the semileptonic tagging method. *Phys. Rev. D*, 92:051102, Sep 2015.

[33] David J. Lange. The evtgen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152–155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[34] Torbjörn Sjöstrand, Patrik Edén, Christer Friberg, Leif Lönnblad, Gabriela Miu, Stephen Mrenna, and Emanuel Norrbin. High-energy-physics event generation with pythia 6.1. *Computer Physics Communications*, 135(2):238–259, 2001.

[35] Elisabetta Barberio, Bob van Eijk, and Zbigniew Wąs. Photos — a universal monte carlo for qed radiative corrections in decays. *Computer Physics Communications*, 66(1):115–128, 1991.

[36] R Brun, F Bruyant, M Maire, A C McPherson, and P Zanarini. *GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version.* CERN, Geneva, 1987.

[37] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II core software: Belle II framework software group. *Computing and Software for Big Science*, 3(1), November 2018.

[38] The Belle II Collaboration. Belle II analysis software framework (basf2), aug 2022.

[39] Moritz Gelb, , Thomas Keck, Markus Prim, et al. B2BII: Data Conversion from Belle to Belle II. *Computing and Software for Big Science*, 2(9), 2018.

[40] Geoffrey C. Fox and Stephen Wolfram. Observables for the analysis of event shapes in e^+e^- annihilation and other processes. *Phys. Rev. Lett.*, 41:1581–1585, Dec 1978.

[41] J.-F. Krohn, F. Tenchini, P. Urquijo, et al. Global decay chain vertex fitting at belle ii. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 976:164269, 2020.

[42] P. Janot and F. Le Diberder. Optimally combined confidence limits. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 411(2):449–466, 1998.

[43] Thomas Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Computing and Software for Big Science*, 1(2), 2017.

[44] Yann Coadou. *Boosted Decision Trees*, page 9–58. WORLD SCIENTIFIC, February 2022.

[45] S. H. Lee, K. Suzuki, et al. Evidence for $B^0 \to \pi^0\pi^0$. *Phys. Rev. Lett.*, 91:261801, Dec 2003.

[46] S. Brandt, Ch. Peyrou, R. Sosnowski, and A. Wroblewski. The principal axis of jets — an attempt to analyse high-energy collisions as two-body processes. *Physics Letters*, 12(1):57–61, 1964.

[47] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), February 2011.

[48] Lukas Heinrich, Matthew Feickert, and Giordon Stark. pyhf: v0.7.6. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6.

[49] Lukas Heinrich, Matthew Feickert, Giordon Stark, and Kyle Cranmer. pyhf: pure-python implementation of histfactory statistical models. *Journal of Open Source Software*, 6(58):2823, 2021.

[50] Number of b events in hadron b(j). https://belle.kek.jp/secured/nbb/nbb.html. Internal documentation.

[51] J. G. Wang. RPC performance at KLM/BELLE. *Nucl. Instrum. Methods A*, 508:133–136, 2003.

[52] T. Aushev *et al.* A scintillator based endcap K_L and muon detector for the Belle II experiment. *Nucl. Instrum. Methods A*, 789:134 – 142, 2015.

[53] S. Yamada, R. Itoh, K. Nakamura, M. Nakao, S. Y. Suzuki, T. Konno, T. Higuchi, Z. Liu, and J. Zhao. Data Acquisition System for the Belle II Experiment. *IEEE Trans. Nucl. Sci.*, 62(3):1175–1180, 2015.

[54] K. Bechtol *et al.* TARGET: A multi-channel digitizer chip for very-high-energy gamma-ray telescopes. *Astroparticle Physics*, 36(1):156–165, Aug 2012.

[55] S. Funk *et al.* TARGET: A digitizing and trigger ASIC for the Cherenkov telescope array. In *AIP Conference Proceedings*, 2017.

[56] A. Albert *et al.* TARGET 5: A new multi-channel digitizer with triggering capabilities for gamma-ray atmospheric cherenkov telescopes. *Astroparticle Physics*, 92:49–61, Jun 2017.

[57] M. Nakao *et al.* Performance of the unified readout system of Belle II. *IEEE Transactions on Nuclear Science*, 68(8):1826–1832, 2021.

[58] B. Edralin. *Design and performance of an automated production test system for a 20,000 channel single-photon, sub-nanosecond electronic readout for a large area muon detector*. PhD thesis, University of Hawai'i, Honolulu, 11 2016.