



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**

UNIVERSITÀ DEGLI STUDI DI TRIESTE

DIPARTIMENTO DI FISICA

XXXVII CICLO DEL DOTTORATO DI RICERCA IN FISICA

**SEARCH FOR $B^+ \rightarrow K^+ \tau^+ \tau^-$ DECAYS
AT BELLE II**

Settore scientifico-disciplinare: FIS/04

Ph.D. student:
Debjit Ghosh

Ph.D. program coordinator:
Prof. Francesco Longo

Thesis supervisor:
Dr. Diego Tonelli

ANNO ACCADEMICO 2023/2024

Abstract

This experimental particle-physics thesis reports on a search for the $B^+ \rightarrow K^+ \tau^+ \tau^-$ rare decay. This decay, unobserved to date, is of particular interest for its sensitivity to potential contributions from non-Standard-Model amplitudes. The analysis uses the full data set of 387 million bottom-antibottom meson pairs from electron-positron collisions at the $\Upsilon(4S)$ collected by the Belle II experiment as of 2023. Belle II is an hermetic solenoidal magnetic spectrometer surrounded by particle-identification detectors, a calorimeter, and muon detectors installed at the KEK SuperKEKB collider. The search uses leptonic decays of the τ leptons, which offer the most sensitivity. I develop the analysis using simulated and control data and inspect the signal-search region only once all procedures are finalized. A restriction to the events in which the pair-produced nonsignal B meson decays into fully reconstructed hadronic final-states suppresses significant backgrounds and enables inference of signal properties. In addition, an innovative restriction on the mass of the opposite-charge kaon-lepton pair suppresses dominant, and poorly modeled, residual background. An optimization of the signal selection further enhances sensitivity. I validate thoroughly the calorimeter energy unassociated with $\Upsilon(4S)$ reconstruction and determine the signal yield by counting the excess events over the expected background in its distribution. The analysis is currently under internal collaboration review and the signal-search region has not yet been undisclosed. The sensitivity of this first Belle II search is two times better than that of the world-leading result, in spite of relying on a smaller sample.

Contents

Introduction	1
1 Flavor physics to overcome the Standard Model	3
1.1 The Standard Model of particle physics	3
1.2 Where do we stand?	5
1.3 Flavor physics in the Standard Model	6
1.4 Flavor physics to overcome the Standard Model	7
1.5 Current flavor status	8
1.6 Recent B anomalies	9
1.7 Motivation for searching for $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays	11
1.7.1 Theoretical predictions	11
2 The Belle II experiment at the SuperKEKB collider	15
2.1 The SuperKEKB collider	15
2.1.1 SuperKEKB performance	19
2.2 The Belle II detector	19
2.2.1 Tracking system	22
2.2.1.1 Silicon-pixel vertexing detector	22
2.2.1.2 Silicon-strip vertexing detector	22
2.2.1.3 Central drift chamber	24
2.2.2 Electromagnetic calorimeter	25
2.2.3 Particle identification	27
2.2.3.1 Time-of-propagation detector	27
2.2.3.2 Aerogel ring-imaging Cherenkov counter	27
2.2.3.3 K_L^0 and muon detection system	28
2.2.4 Trigger and data acquisition system	28
2.3 Reconstruction of stable particles	29
2.3.1 Charged-particle reconstruction	30
2.3.2 Charged-particle identification	31
2.3.3 Photon reconstruction	34
3 Experimental considerations	38
3.1 A $B^+ \rightarrow K^+ \tau^+ \tau^-$ decay in Belle II	38
3.2 Analysis strategy and overview	38
3.3 Current experimental status	40
3.4 Relevant analysis tools	41
3.4.1 B tagging	41
3.4.2 Lower-level observables	41
3.4.3 B meson observables	43

3.4.4	Global event observables	45
4	Signal reconstruction and baseline selection	49
4.1	Online event selection	49
4.2	Samples	50
4.2.1	Experimental data	50
4.2.2	Simulated data	50
4.3	Centralized data reduction	51
4.4	Offline selection and reconstruction	52
4.4.1	Introduction to full event interpretation	52
4.4.2	FEI preselection	53
4.4.3	FEI training	56
4.4.4	B_{tag} selection	57
4.4.5	Multiple B_{tag} candidates	58
4.4.6	Choice of τ final states and treatment of lepton final states	58
4.4.7	Signal B meson selection	60
4.5	Baseline selection	61
4.6	Opposite-charge kaon-lepton mass	64
4.7	Further discriminating observables	65
5	Residual calorimeter energy	67
5.1	Importance and strategy	67
5.2	Control samples	69
5.3	E_{ECL} sources	70
5.3.1	Misreconstructed photons	70
5.3.2	Beam-induced background photons	70
5.3.3	Collision photons	71
5.3.4	Source identification	71
5.3.5	Baseline photon selection and isolation of source-specific samples	72
5.4	Validation	74
5.4.1	Photon energy	75
5.4.2	Photon multiplicity	76
5.4.3	Intermediate summary	78
5.4.4	Photon-multiplicity correction in the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample	82
5.4.5	Sample-dependence checks	83
5.5	Final photon selection	84
6	Selection optimization and sensitivity	89
6.1	Introduction	89
6.2	Realistic simulated sample	89
6.2.1	Simulated signal-efficiency calibration	89
6.2.2	Simplified estimation of expected background	92
6.3	Sensitivity estimation	93
6.4	Signal selection optimization	95
6.4.1	Final background composition	95
6.5	Impact of baseline selection on sensitivity	97
6.5.1	PID selection	97
6.5.2	Photon selections	99
6.5.3	π^0 veto	99

7	Final validation and background estimate	101
7.1	E _{ECL} validation for the final sample composition	101
7.2	Final background estimation in the signal region	101
8	Systematic uncertainties	107
8.1	Uncertainty in background estimate	107
8.2	Uncertainty in signal efficiency	108
8.3	Uncertainties in external inputs	109
9	Results and summary	111
A	An alternative, BDT-based, selection	116
A.1	Introduction	116
A.2	Boosted decision trees	116
A.3	Choice of discriminating observables	117
A.4	Classifier training and testing	118
A.5	Selection optimization	121
A.6	Impact of BDT input choices	121
B	Sensitivity from exclusive analysis of lepton channels	124
C	Sensitivity from pion channels	126
C.1	Introduction	126
C.2	$K\ell\pi$ final state	126
C.2.1	Reconstruction and baseline selection	126
C.2.2	Study of charge correlations	127
C.2.3	BDT-based selection	128
C.2.4	Selection optimization and sensitivity	129
C.2.5	Auxiliary checks on expected sensitivity	132
C.2.5.1	Default selection	132
C.2.5.2	Impact of lepton particle-identification selection	132
C.2.6	Choice of discriminating observables	133
C.2.7	Conclusion on $K\ell\pi$ study	134
C.3	$K\pi\pi$ final state	134
C.3.1	Reconstruction and baseline selection	134
C.3.2	Background suppression	135
C.3.3	BDT-based selection	136
C.3.4	Selection optimization and sensitivity	136
C.4	Sensitivity comparison	138
	Bibliography	140

Introduction

The Standard Model of elementary particles and their interactions (SM) is the currently accepted theory of particle physics. It is widely recognized as the ultimate success of the reductionist paradigm for describing microphysics at its most fundamental level. By means of about twenty parameters, the Standard Model accurately describes thousands of measurements involving processes mediated by the electromagnetic, weak, and strong interactions that span more than ten orders of magnitude in energy. However, theoretical considerations and, possibly, experimental inconsistencies support the general belief that the Standard Model might still be an *effective theory* — a theory valid at the energies probed so far, but incorporated in a yet-unknown and more general theory extending to higher energies. Completing the Standard Model is the principal goal of today’s particle physics.

Direct approaches, which broadly consist in searching for decay products of non-SM particles produced on mass-shell in high-energy collisions, have been traditionally fruitful. However, their reach is limited by the collision energy of today’s accelerators and by the large investments needed to further it in future. Complementary approaches consist in comparing with predictions precise measurements in lower-energy processes in which virtual non-SM particles could contribute. The reach of such *indirect* approaches is not constrained by collision energy, but rather by the precision attainable, both in measurements and predictions.

The Belle II experiment is an international collaboration of about 700 physicists that *indirectly* tests the Standard Model by studying hundreds of millions of decays of mesons containing the quarks b and c , which are heavier and longer-lived partners of the fundamental constituents of nuclear matter, and τ leptons which are heaviest partners of the electron. These particles are pair-produced in electron-positron collisions at energies near 10.58 GeV. This is the threshold energy for the production of the $\Upsilon(4S)$ meson, a bound state of a b quark and a $\bar{b}$ antiquark that predominantly decays in a bottom-antibottom meson (B , bound states of a b quark and a lighter antiquark) pair yielding large, and low-background samples for physics analysis. Since the start of physics data-taking in 2019, Belle II collected data corresponding to 428 fb⁻¹ of integrated luminosity. The 365 fb⁻¹ portion collected until 2023 contains 387 million $B\bar{B}$ pairs, which are used in this first Belle II search for $B^+ \rightarrow K^+\tau^+\tau^-$ decays.

These decays are governed by $b \rightarrow s\ell^+\ell^-$ quark transitions, which imply a change in quark flavor without a change in electric charge ($\ell^\pm$ indicates a charged lepton). These transitions are both sensitive to contributions from hypothetical non-SM particles and suppressed in the Standard Model, yielding approximately one SM decay in ten millions. Comparing the observed branching fractions with SM predictions offers sensitive probes to investigate a broad range of potential plausible non-SM contributions. In addition, the motivation for studying $B^+ \rightarrow K^+\tau^+\tau^-$ decays has been reinforced in recent years by the observed pattern of flavor anomalies in $B^+ \rightarrow \bar{D}^{(*)0}\tau^+\nu$ decays, which suggests investi-

gating non-SM particles that have significant couplings with τ leptons. The only available experimental information on $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays is an upper limit at a 10^4 -times-greater rate than the SM prediction. This was reported in 2017 by the BaBar experiment with its full data set of 471 million $B\bar{B}$ pairs. With data from a more advanced detector, analyzed with an innovative and thoroughly optimized analysis, we plan on improving on that result — event though our sample is smaller.

The $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays are rare thus requiring large B -decay samples and effective background suppression for a sensitive search. Indeed, the primary challenge arises from the contamination of backgrounds 10^9 larger than signal at production. An equally challenging difficulty is the lack of straightforward discriminating signal features. The pair of τ leptons yield multiple neutrinos, which are undetected in Belle II. This leaves incomplete the signal kinematic information, leading to lack of narrow, or otherwise distinctive, signal distributions to discriminate against background. The analysis must therefore rely significantly on inclusive event properties to recover discriminating power. This is uniquely possible at Belle II due to its particular experimental environment, but it implies additional experimental challenges. Near-threshold $B\bar{B}$ pair-production from collisions of point-like particles reconstructed with an hermetic detector provides stringent constraints on the kinematic properties of the initial state. The analysis exploits these constraints with a two-pronged approach. I restrict the sample to collisions in which the pair-produced nonsignal B meson is fully reconstructed in a hadronic decay. This allows to infer statistically the signal properties and suppress background by relating the nonsignal B kinematic information with known production kinematics. The disadvantage is that fully reconstructable hadronic decays are relatively rare. This results in low efficiency, making the search for a rare decay even more difficult. Inclusive event properties are essential for signal extraction too. A decisive observable is the residual energy detected in the calorimeter after reconstruction of the $B\bar{B}$ pair. Due to on-threshold production, properly reconstructed signal events exhibit a distinctively peaking residual-energy shape that offers a convenient signal-extraction observable. However, modeling the residual energy is nontrivial as its properties depend on sample composition and on the properties of all particles in the event, which need to be carefully described. Hence, a thorough validation of the data-simulation consistency of residual energy is done as an original and important part of this work. Finally, this analysis introduces a simple, but innovative, background-suppression choice by restricting the search to events in which the opposite-charge kaon-lepton mass exceeds the D meson mass. This rejects kinematically most semileptonic $B^+ \rightarrow \bar{D}^{(*)0} \ell \bar{\nu}$ decays, which are abundant and prone to mismodeling. The resulting search region, with low and controlled background, enables a simpler analysis with no loss of sensitivity.

This thesis is structured as follows. Chapter 1 introduces the flavor sector of the Standard Model and discusses the relevance of $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays; Chapter 2 provides an overview of the SuperKEKB accelerator and the Belle II experiment; Chapter 3 outlines the experimental features relevant for the measurement; Chapter 4, in which the description of my direct, original contributions begins, reports on the reconstruction and selection of the analysis sample; Chapter 5 documents a thorough study of the principal signal-extraction observable and its validation; Chapter 6 describes the selection optimization; Chapter 7 discusses the final validation and the estimate of the background in the signal-search region; Chapter 8 discusses sources of systematic effects and associated uncertainties; Chapter 9 reports the results of the search and a summary.

Charge- and flavor-conjugate processes are implied throughout the document unless specified otherwise. Generic particle symbols (B , K , ...) indicate indistinctly charged or neutral particles.

Chapter 1

Flavor physics to overcome the Standard Model

This is a concise introduction to the weak interactions of quarks and how they are incorporated in the Standard Model of particle physics. Emphasis is on their role in searches for as-yet unknown particles that may complete the Standard Model at high energies. The final part of the chapter discusses $B^+ \rightarrow K^+ \tau^+ \tau^-$ dynamics and motivations for the search.

1.1 The Standard Model of particle physics

The Standard Model (SM) is an effective quantum field theory that describes all fundamental interactions in nature without gravity [1–6]. The quantum-field-theory framework results from the unification of quantum mechanics with special relativity and offers the most fundamental description of nature known to date. A field is a set of values, associated to certain physical properties, assigned to every point in space and time. Quantum fields are fields that pervade the whole spacetime and obey the rules of quantum mechanics. If a quantum field is modified by an appropriate perturbation, the resulting oscillatory states, called field excitations, carry more energy than the resting state and are called ‘particles’. For instance, the electron is the massive excitation of the electron field. The quantized nature of the description implies that only certain perturbations that satisfy precise energetic conditions are capable of generating field excitations. It is not possible, for example, to generate a wave in the electron field that corresponds to half an electron.

Quantum fields interact with each other. The Standard Model is the theory that describes their dynamics at energy scales relevant for the subnuclear world. Particles and their interactions are described in a Lagrangian formalism, in which every combination of fields and interaction operators that is not forbidden by the symmetries of the dynamics is, in principle, included. Local gauge symmetry, i.e., the invariance of the Lagrangian under space-time-dependent transformations applied to the phases of fields, is the key overarching concept. Interaction terms appear in the free-field Lagrangian after requiring it to be invariant under local gauge symmetries. The Standard Model is based on the symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y,$$

where $SU(3)_C$ is the standard unitary group that describes the strong interactions (quantum chromodynamics, QCD), and C stands for the color charge; $SU(2)_L \otimes U(1)_Y$ is the product of groups that describe the combination of the weak and electromagnetic interactions, where $SU(2)_L$ is the standard unitary group of weak isospin doublets (L standing

for *left*¹) and $U(1)_Y$ stands for the unitary group of *hypercharge* Y .

Spin-1 particles called *gauge bosons* mediate the interactions. Strong interactions are mediated by eight massless particles corresponding to the $SU(3)_C$ generators, called *gluons*: they carry a charge that is of three kinds, called *color*. Weak interactions are mediated by two charged massive bosons, $W^\pm$, and a neutral massive boson, Z^0 . Electromagnetic interactions occur between particles carrying electric charge and are mediated by a neutral massless boson, the photon γ . The physical electroweak bosons ($W^\pm$, Z^0 , γ) arise from the following linear combinations of $SU(2)_L \otimes U(1)_Y$ generators:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \quad \text{and} \quad \begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix},$$

where θ_W is a free parameter, called *Weinberg angle*. The $W^\pm$ mass depends on the Z mass via θ_W . Particles acquire mass via the interaction with the Higgs field, which is mediated by a spin-0 particle, the Higgs boson.

Matter particles correspond to excitations of spin-1/2 fields and are called *fermions*. Their masses are free parameters of the theory. Each fermion is also associated with an antiparticle that has the same mass and opposite internal quantum numbers. Fermions are further classified into two classes, quarks, which are the fundamental constituents of nuclear matter, and leptons, each organized in three weak-isospin doublets.

- Quark doublets are composed of an up-type quark, with charge $(2/3)e$, and a down-type quark, with charge $(1/3)e$ each,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}.$$

They couple with both the strong and electroweak interactions. Each quark has color and a ‘flavor’ quantum number, which comes in six varieties and is conserved in the electromagnetic and strong interactions, but not in the weak interactions. Due to color confinement free quarks are not observable [7]. They are only observed in their colorless bound states, which include mesons, typically composed of a quark and an anti-quark, and baryons, composed of three quarks. Baryons are assigned a quantum number, called baryon number, found to be conserved even if no symmetry of the Lagrangian implies that.

- Lepton doublets are composed each by an almost massless neutral neutrino and a massive particle with electric charge $-e$;

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}.$$

They couple only with the electroweak interaction. Each lepton has a lepton-family quantum number; their sum in a process, called global lepton number, is found to be conserved in all interactions, although no symmetry of the dynamics prescribes that; individual lepton numbers are not conserved in neutrino oscillations.

Figure 1.1 shows a scheme of the Standard Model particles and their interactions.

¹Only particles with *left* chirality are influenced by the weak interaction.

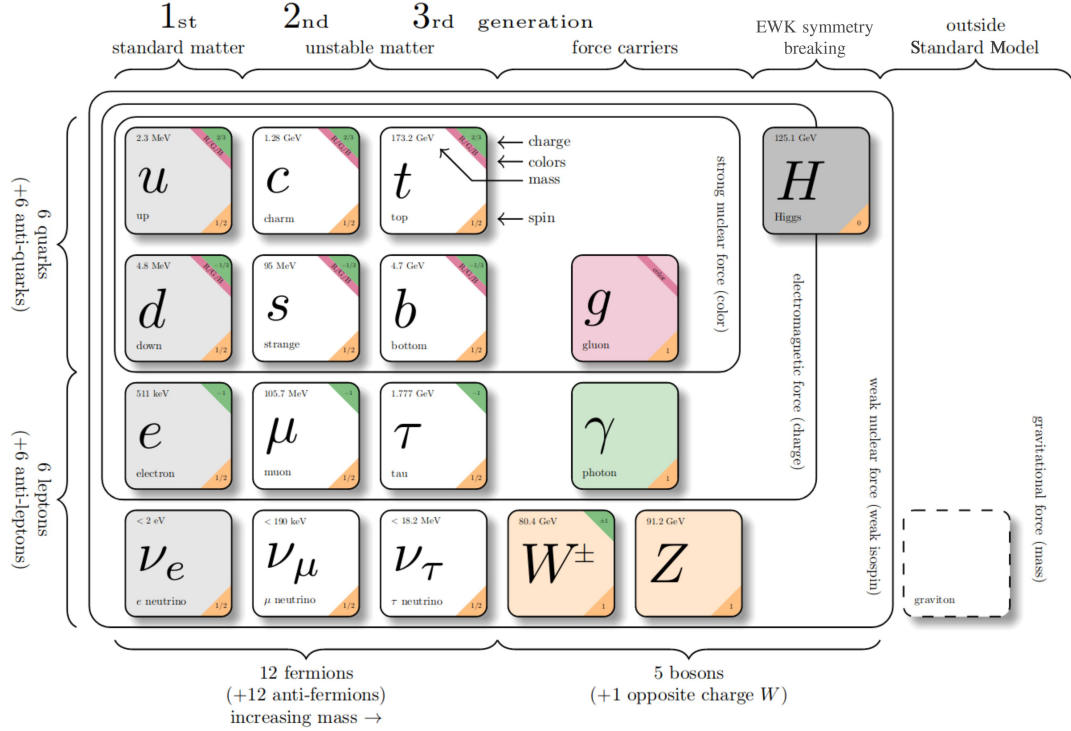


Figure 1.1: Scheme of particles and interactions in the Standard Model.

In addition to gauge symmetry, discrete symmetries are important in constraining the dynamics. Parity (P) is a transformation that inverts all spatial coordinates; charge conjugation (C) is the exchange of every particle with its own antiparticle; and time reversal (T) inverts the time axis. The product of these three discrete symmetry transformations is found to be conserved in all interactions, as prescribed by foundational axioms of field theory [8,9], but the symmetries are not conserved individually. Parity symmetry is maximally violated in the weak interactions, while the combined CP symmetry is violated in the weak interactions at the 0.1% level. In principle, the strong interaction too could violate CP symmetry, but no experimental evidence of that has ever been observed. The existence of as-yet unobserved particles, called axions, has been postulated to account for that [10].

1.2 Where do we stand?

The Standard Model was completed in the 1970's and has been successfully tested since, in thousands of measurements whose fractional precisions reach one part per trillion [11]. However, observations and theoretical considerations suggest that the Standard Model is likely to be an effective theory, valid at the eV–TeV energies probed so far, that should be completed by a more general full theory valid over a broader range of high energies. Open questions that support this interpretation include the lack of an explanation for a dynamical origin for the observed asymmetry between matter and antimatter in the universe, the strikingly large differences observed between fermion masses, the possible instability of the Higgs vacuum, the conceptual and technical difficulties in achieving a

description of gravity consistent with quantum mechanics, or the postulated large amounts of non-interacting matter (dark matter), introduced to justify cosmological observations.

Extending the Standard Model to higher energy-scales is the main goal of today's particle physics, in an attempt at addressing these and other open issues. Current strategies to extend the Standard Model can broadly be classified into two synergic approaches.

The energy-frontier, *direct* approach aims at using high-energy collisions to produce on-shell particles (that is, particles satisfying the energy-momentum conservation at production) not included in the Standard Model, and detect directly their decay products, thus gaining direct evidence of their existence.² Historically this offered striking experimental evidence of new phenomena, when energetically accessible, but its reach is limited by the maximum energy available at colliders.

The intensity-frontier, *indirect* approach broadly consists in searching for significant differences between precise measurements and equally precise SM predictions in lower-energy processes sensitive to non-SM contributions. A semi-intuitive, although simplified conceptual representation of the subtending idea is that exchanges of virtual (off-mass-shell) particles of arbitrary high mass, including those not described in the Standard Model, occur in the transition, thus altering the amplitudes in an observable manner. The presence of virtual particles, which may imply a temporary non-conservation of energy if interpreted classically, is allowed by Heisenberg's uncertainty principle $\Delta E \Delta t > \hbar/2$. Experimental evidence is typically harder to establish, but the reach is not bounded by the maximum collision energy reachable by experiments. A large portion of the effort in this approach is centered on the weak-interactions of quarks (so called 'flavor physics').

1.3 Flavor physics in the Standard Model

Although technically flavor physics includes also lepton interactions, I restrict the scope by referring solely to the quark interactions here.

The role of flavor in shaping the Standard Model has been central since the early days of particle physics. However, its prominence in determining the theory can perhaps be tracked down to the early 1960's with the apparent inconsistency between weak coupling constants measured in muon decay, neutron decay, and strange-particle decays. Such inconsistency was first addressed by Gell-Mann and Levy [12] and then Cabibbo [13], who postulated differing mass (d) and weak (d') eigenstates for down-type quarks. This was achieved by introducing a mixing angle (θ_C) between the s quark and d quark, the only two down-type quarks known at the time. While Cabibbo's theory addressed efficiently the difference of weak coupling constants, it also predicted a rate for the $K_L^0 \rightarrow \mu^+ \mu^-$ and other kaon process inconsistent with the experimental exclusion limits at the time. Glashow, Iliopoulos, and Maiani addressed the conundrum by postulating the existence of a fourth quark (c) of $2 \text{ GeV}/c^2$ mass, whose contribution in the $K_L^0 \rightarrow \mu^+ \mu^-$ decay amplitude would cancel the u quark contribution, suppressing the branching fraction down to values consistent with experimental limits [14]. The charm quark was then discovered four years after the prediction, showing the compelling power of the indirect approach. In addition, in 1973, when only three quarks were known, Kobayashi and Maskawa generalized Cabibbo's theory from a four-quark model to a six-quark model to accommodate the phenomenon of CP violation observed in 1964 [15]. They introduced a complex unitary matrix to describe the relations between mass (unprimed) and weak (primed) interaction eigenstates of quarks as

²*Mass shell* is jargon for mass hyperboloid, which identifies the hyperboloid in energy-momentum space describing the solutions to the mass-energy equivalence equation $E^2 = (pc)^2 + m^2 c^4$. A particle *on-mass-shell* satisfies this relation.

seen by $W^\pm$ bosons. This is known as the Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix or V_{CKM} ,

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

Each V_{ij} matrix element encapsulates the weak-interaction coupling between an up-type i and down-type j quarks. The CKM matrix is a $N \times N$ CKM matrix with $(N-1)^2$ free parameters, in which N is the number of quarks families [16]. If $N = 2$, the only free parameter is the Cabibbo angle $\theta_C \approx 13^\circ$ whereas if $N = 3$, the free parameters are three Euler angles (θ_{12} , θ_{13} , and θ_{23}) and a complex phase (δ), which allows for CP -violating couplings [11]. The matrix is most conveniently written in the so-called *Wolfenstein parametrization*, an expansion in the small parameter $\lambda = \sin \theta_C \approx 0.23$ that exposes explicitly the observed hierarchy between its elements [17],

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) ,$$

where

$$\lambda = \frac{V_{us}}{\sqrt{V_{ud}^2 + V_{us}^2}}, \quad A\lambda^2 = \lambda \frac{V_{cb}}{V_{us}}, \quad \text{and} \quad A\lambda^3(\rho + i\eta) = V_{ub}^* .$$

The real parameter λ expresses the mixing between the first and second quark generations, A and ρ are real parameters, and η is a complex phase that introduces CP violation. The unitarity condition $V_{\text{CKM}} V_{\text{CKM}}^\dagger = \mathbb{1}$ yields nine relations,

$$\begin{aligned} |V_{ud}|^2 + |V_{cd}|^2 + |V_{td}|^2 &= 1 & V_{us}^* V_{ud} + V_{cs}^* V_{cd} + V_{ts}^* V_{td} &= 0 & V_{ud} V_{cd}^* + V_{us} V_{cs}^* + V_{ub} V_{cb}^* &= 0 , \\ |V_{us}|^2 + |V_{cs}|^2 + |V_{ts}|^2 &= 1 & V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} &= 0 & V_{ud} V_{td}^* + V_{us} V_{ts}^* + V_{ub} V_{tb}^* &= 0 , \\ |V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 &= 1 & V_{ub}^* V_{us} + V_{cb}^* V_{cs} + V_{tb}^* V_{ts} &= 0 & V_{cd} V_{td}^* + V_{cs} V_{ts}^* + V_{cb} V_{tb}^* &= 0 , \end{aligned}$$

which are sums of three complex numbers each. The six equations summing to zero prompt a convenient geometric representation in the complex plane in terms of so-called *unitarity triangles*. A CP conserving theory would yield null-area triangles or, equivalently, a vanishing Jarlskog invariant $J = \Im(V_{us} V_{cb} V_{ub}^* V_{cs}^*)$ [18–20]. All elements of the second equation in the second row have similar magnitudes, yielding a notable triangle referred to as ‘the Unitarity Triangle’, shown in fig. 1.2. Conventionally, side sizes are normalized to the length of the base, and the three angles are labelled α or ϕ_2 , β or ϕ_1 , and γ or ϕ_3 .

The phenomenon of flavor mixing, due to the noncoincidence of strong-interaction eigenstates, relevant in quark production, and physical eigenstates of the full Hamiltonian, which have defined mass and lifetime, further enriches the phenomenology.

1.4 Flavor physics to overcome the Standard Model

Many physicists find the current understanding of flavor dynamics unsatisfactory. The observed hierarchies between quark masses and couplings seem too regular to be accidental and the abundance of free parameters (six quark masses and four couplings) suggests the possibility of a deeper, more fundamental theory possibly based on a reduced set of

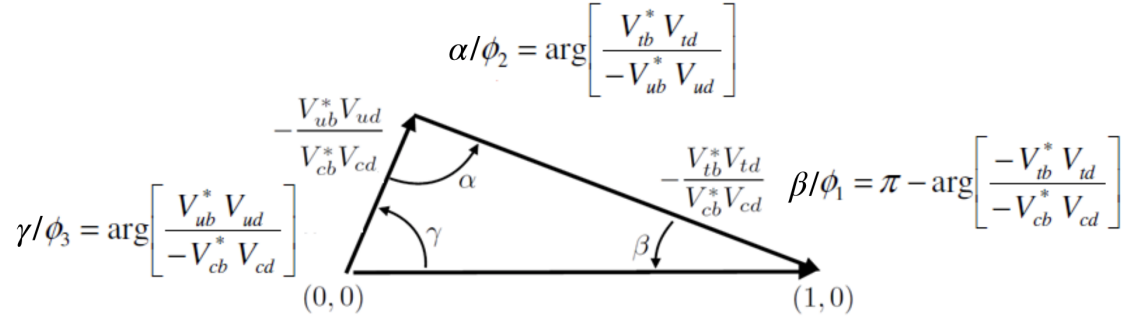


Figure 1.2: Graphical representation of the Unitarity Triangle.

parameters. In addition, while the CKM mechanism offers a framework to include CP violation in the Standard Model, it does not really enlightens the origin for such a singular phenomenon. But even in the absence of a deeper understanding of the origin of CP violation, naturalness arguments indicate that most generic extensions of the Standard Model would involve additional sources of CP violation. These and other considerations support the notion that a more detailed and complete study of the phenomenology of quarks dynamics may reveal useful information to guide searches for SM extensions. This is further supported by experimental advantages. The abundance and diversity of experimentally accessible processes to measure redundantly a reduced set of parameters makes indirect searches in the flavor sector a promising option for exploring non-SM dynamics. In fact, even if no deviations from the Standard Model are found, the resulting stringent constraints on SM extensions are expected to remain useful in informing future searches.

The two classes of flavor-physics processes most promising for probing contributions of non-SM particles are CP -violating and *flavor-changing-neutral-current* processes. The phenomenon of CP violation offers numerous avenues to uncover or characterize possible non-SM contributions. Alterations of the CP -violating phases with respect to those predicted by the SM are generically expected in a broad class of SM-extensions. Observing experimental evidence of those phases offers further opportunities to explore the dynamics, even if total rates are unaffected. Flavor-changing neutral currents (FCNC) are processes in which quark flavor changes, but quark electric charge does not. These processes are suppressed in the Standard Model because they occur only through higher-order amplitudes involving the internal exchange of $W^\pm$ bosons ('loop amplitudes'), as shown in fig. 1.3. Such amplitudes are naturally sensitive to non-SM contributions, since any particle with proper quantum numbers and nearly arbitrary mass can replace the SM-quark closed-lines in these diagrams thus altering the rate. Hence, FCNCs are powerful in signaling contributions from non-SM particles if rate enhancements, or suppressions, with respect to Standard Model expectations are observed. The search for a FCNC decay is the objective of this thesis work.

1.5 Current flavor status

Measurements of parameters associated with quark-flavor physics have been performed in many dedicated, or general-purpose, experiments in the last three decades, including CLEO, CPLEAR, NA32, NA48, NA62, KTeV, SLD, OPAL, L3, ALEPH, DELPHI, BaBar, Belle, Belle II, CDF, CDFII, LHCb, BESIII, ATLAS, and CMS [21].

The current status of constraints on sides and angles of the Unitarity Triangle is shown

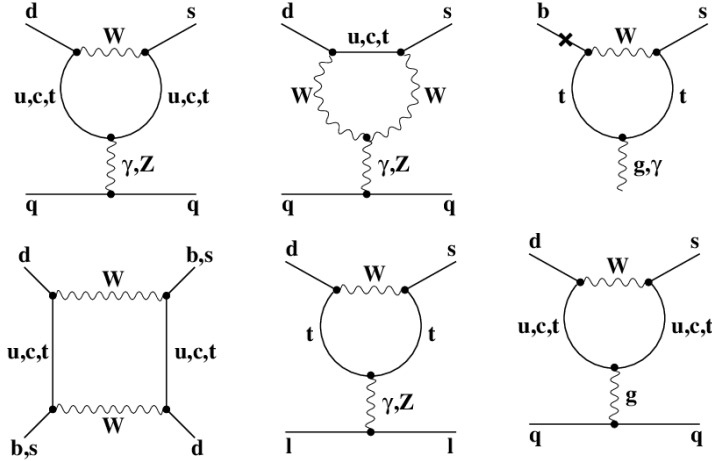


Figure 1.3: Examples of FCNC diagrams.

in fig. 1.4 [22]. Measurements of $\sin 2\beta$ reached a precision of 1%, mainly due to the availability of large samples of $B^0 \rightarrow J/\psi K^0$ decays in e^+e^- and pp collisions, while the angle α is known down to a 4% precision from $B \rightarrow hh$ decays, where h represents a charged or neutral π or ρ) in e^+e^- collisions. The angle γ is measured with 4% precision using combinations of several measurements involving $B \rightarrow DK$ decays reconstructed in e^+e^- and pp collisions. Discrepancies in the determinations of $|V_{cb}|$ and $|V_{ub}|$ are found between values measured using different analyses of semileptonic decays, mainly performed in e^+e^- collisions. The decay-width difference of the $B_s^0 - \bar{B}_s^0$ system is determined with 5% precision in pp collisions, while measurements are not yet precise enough to discern the expected nonzero value for the $B^0 - \bar{B}^0$ system. Mass differences in both systems are known with better than 1% precision from pp and $p\bar{p}$ collisions. In addition, many other measurements in charm and kaon physics contribute that are not represented in the Unitarity Triangle.

The resulting global picture is that the CKM interpretation of CP violation is the dominant mechanism at play in the dynamics. However, despite the first-order consistency, possible deviations of up to 10%–15% are still unconstrained, especially those associated with loop-mediated processes, leaving sufficient room for non-SM physics. It is especially promising that most of the relevant measurements are currently dominated by statistical uncertainties, offering therefore fruitful opportunities for the two experiments that will contribute the most in the next decade, LHCb and Belle II. This is all the more attractive because direct searches for non-SM physics, mainly in pp collisions at the Large Hadron Collider (LHC) at CERN, excluded large portions of the parameter space for several proposed SM extensions, but showed no conclusive evidence of non-SM physics to date. Since plans for a higher-energy collider in the near future are still fluid, the systematic study of flavor physics emerges as a promising program to search for non-SM in the next decade.

1.6 Recent B anomalies

In the SM, the coupling strength to the $W^\pm$ bosons is identical for all lepton doublets, a property called lepton-flavor universality (LFU), which is an accidental symmetry of the theory. Tests of LFU during the last decade, by the BaBar, LHCb, and Belle (II) experiments, have shown indications of deviations from the SM predictions in the $b \rightarrow c\ell\bar{\nu}_\ell$

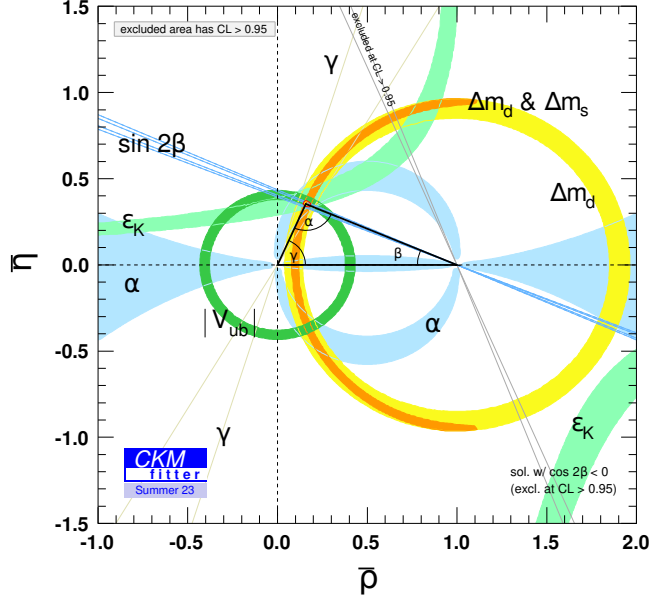


Figure 1.4: Current constraints on sides and angles of the Unitarity Triangle. Reproduced from Ref. [22]

transitions, commonly referred to as ‘ B anomalies’. These tests usually consist in measuring ratios of branching fractions of decays that only differ by the final-state lepton flavor, in search for significant deviations from the SM expectations. Departures from LFU are found in the ratios

$$R_{D^{(*)}}^{\tau/\ell} = \frac{\Gamma(\bar{B} \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau)}{\Gamma(\bar{B} \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell)}, \quad (1.1)$$

in which ℓ indicates an electron or a muon, yielding global averages (fig. 1.5) [23]

- $R_D^{\tau/\ell} = 0.342 \pm 0.026$ and
- $R_{D^*}^{\tau/\ell} = 0.287 \pm 0.012$,

which exceed the SM predictions by 2.5σ and 1.6σ , respectively. Considering the -0.39 linear correlation between them, the combined result differs from the SM predictions by about 3.3σ . A similar, though milder (2σ), deviation is observed in [24]

$$R_{J/\psi}^{\tau/\mu} = \frac{\Gamma(\bar{B}_c \rightarrow J/\psi \tau^- \bar{\nu}_\tau)}{\Gamma(\bar{B}_c \rightarrow J/\psi \ell^- \bar{\nu}_\ell)} = 0.71 \pm 0.25$$

If conclusively confirmed not to be due to experimental mistakes or underestimated systematic uncertainties, these discrepancies may point towards a $b \rightarrow c \tau^- \bar{\nu}_\tau$ enhancement via a non-SM mediator. This could occur as a tree-level exchange of a W' boson (the Z' partner in the vector-boson model) [25], a leptoquark [26, 27], or a charged Higgs [28]. A large tree-level contribution, $\mathcal{O}(10\%)$, would be needed to make up for the enhancement, and would therefore be likely to affect sensibly various other processes. Hence, more data are crucial to test the experimental reliability of the anomaly and explore possible implications in other processes. The close association of these enhancements with third-generation leptons prompts a dedicated exploration of $b \rightarrow s \tau \tau$ transitions.

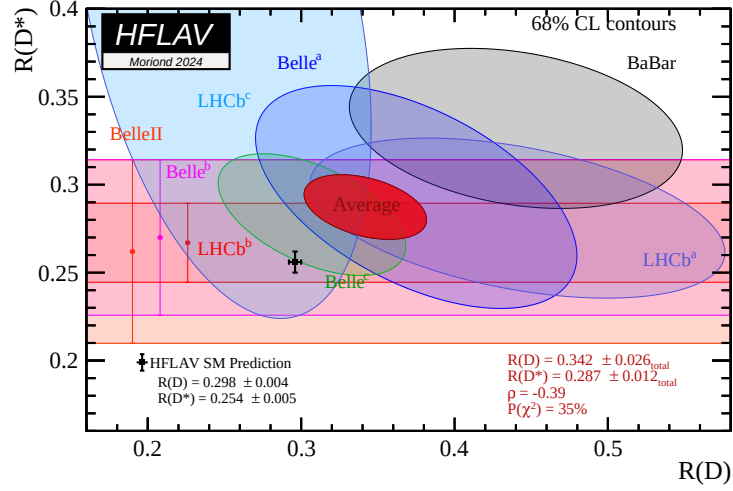


Figure 1.5: Results of $R_{D^*}^{\tau/\ell}$ as a function of $R_D^{\tau/\ell}$ results compared with the SM predictions. Reproduced from Ref. [23].

1.7 Motivation for searching for $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays

Decays $B^+ \rightarrow K^{(*)} \tau^+ \tau^-$ have an important role in this exploration. These are FCNC decays that occur through penguin or box amplitudes in weak perturbation theory (fig. 1.6), which make them strongly suppressed. However, the decay amplitude would be enhanced in SM extensions that imply large coupling with particles of the third generation, the two τ leptons in this case. This makes searches and study of $B^+ \rightarrow K^{(*)} \tau^+ \tau^-$ decays particularly compelling. However, the presence of up to four neutrinos in the final state renders the measurements challenging, and explains why only a few measurements are available to date for such decays.

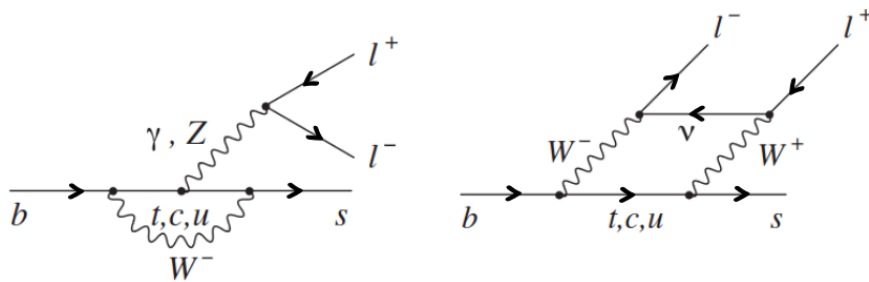


Figure 1.6: Lowest-order Feynman (left) ‘penguin’ and (right) ‘box’ diagrams of the $b \rightarrow s \ell^+ \ell^-$ transition.

1.7.1 Theoretical predictions

The one-loop $b \rightarrow s$ transition can be described by an effective Hamiltonian using the operator product expansion, which is a phenomenology technique that allows to factorize amplitudes into short-distance (high-energy) terms and long-distance (low-energy) terms, thus facilitating calculations [29]. The effective Hamiltonian is

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_{i=1} \mathcal{C}_i(\mu) \mathcal{O}_i(\mu) + \text{hermitian conjugate}, \quad (1.2)$$

where G_F is the Fermi constant; $V_{qq'}$ are relevant CKM matrix-elements; local operators $\mathcal{O}_i$ represent effective point-like vertices, such as current-current $\mathcal{O}_1$ and $\mathcal{O}_2$, QCD penguin $\mathcal{O}_{3-6}$, electromagnetic and chromomagnetic $\mathcal{O}_7$ and $\mathcal{O}_8$, (axial)vector components of the electroweak penguin $\mathcal{O}_{(10)9}$, and (pseudo)scalar operators $\mathcal{O}_{(P)S}$; and the Wilson coefficients $\mathcal{C}_i(\mu)$ are renormalization-scale-dependent couplings to the corresponding operators describing short-distance physics. For the $b \rightarrow s\ell^+\ell^-$ transition, only the operators

$$\begin{aligned} \mathcal{O}_7^{(\prime)} &= \frac{\alpha}{4\pi} m_b (\bar{s} \sigma_{\mu\nu} P_{R/L} b) F^{\mu\nu}, \\ \mathcal{O}_9^{(\prime)} &= \frac{\alpha}{4\pi} (\bar{s} \gamma_\mu P_{L/R} b) (\bar{\ell} \gamma^\mu \ell), & \mathcal{O}_{10}^{(\prime)} &= \frac{\alpha}{4\pi} (\bar{s} \gamma_\mu P_{L/R} b) (\bar{\ell} \gamma^\mu \gamma^5 \ell), \\ \mathcal{O}_S^{(\prime)} &= \frac{\alpha}{4\pi} (\bar{s} P_{R/L} b) (\bar{\ell} \ell), & \mathcal{O}_P^{(\prime)} &= \frac{\alpha}{4\pi} (\bar{s} P_{R/L} b) (\bar{\ell} \gamma^5 \ell) \end{aligned}$$

contribute to leading order (fig. 1.7). Here, $\mathcal{O}_i^{(\prime)}$ are chirally flipped operators; $\alpha = e^2/4\pi$ is the electromagnetic coupling constant; m_b is the bottom-quark mass; s , b , and ℓ are the strange quark, bottom quark, and lepton fields respectively; γ_μ is the 4×4 matrix used to describe spin-half particles; $\sigma_{\mu\nu}$ is the commutator of γ matrices; $P_{L/R}$ are the left/right-handed projection operators, and $F_{\mu\nu}$ is the electromagnetic field tensor.

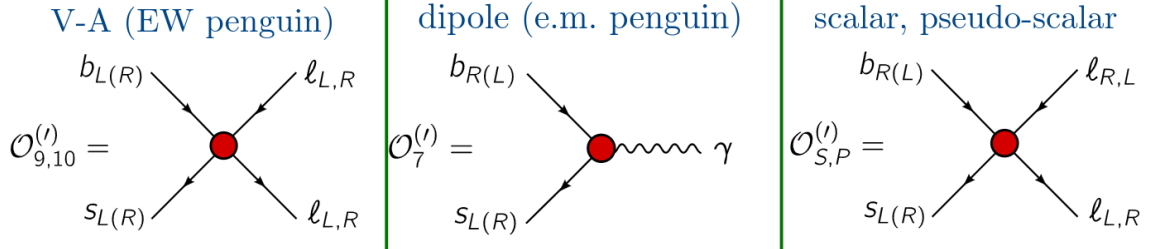


Figure 1.7: Diagrams representing the effective operators that contribute to the $b \rightarrow s\ell^+\ell^-$ transition at the leading order in electroweak theory. Reproduced from Ref. [30].

The SM branching fraction for the decay of interest, $B^+ \rightarrow K^+ \tau^+ \tau^-$, is calculated by replacing ℓ by τ . Currently, the most precise SM prediction uses recent lattice-QCD form factors [31], which are effective parameters that encapsulate the nonperturbative effects of the strong interaction in the hadronic-matrix elements of vector and tensor currents between b and s quarks, and is [32]

$$\mathcal{B}_{\text{SM}}(B^+ \rightarrow K^+ \tau^+ \tau^-) = (1.49 \pm 0.10) \times 10^{-7}. \quad (1.3)$$

Here, the branching fraction is calculated by restricting the phase space to the region of squared ditau mass $q^2 > 14.18 \text{ GeV}^2/c^4$ to exclude contributions from $B \rightarrow K\psi(2S)[\rightarrow \tau^+ \tau^-]$ decays, which have a much larger effective branching fraction of 1.8×10^{-6} , and does not include QED uncertainties associated with electromagnetic interactions between final-state charged-particles and between electrically-charged quarks within the B and K mesons.

Usually, the effect of non-SM physics is generically parametrized as an additive modification of Wilson coefficients, $\mathcal{C}_i \rightarrow \mathcal{C}_i^{\text{SM}} + \Delta_i^{\text{NP}}$. Hence, fits to experimental results sensitive

to the same sets of operators allow, under mild assumptions, to infer generic, model-independent constraints on non-SM models by determining bounds to the Δ_i^{NP} quantities. An example is the leptoquark U_1 hypothesis (represented by $(3, 1)_{2/3}$ in Weyl notation) [33], which has attracted interest because it could explain the B anomalies and remain consistent with LHC high-energy bounds. Leptoquarks (LQs) are color-triplet bosons that carry both lepton and baryon numbers thus mediating interactions between quarks and leptons [34]. In general, a LQ can be a scalar or a vector field, denoted by its quantum number with respect to the SM gauge group $(SU(3)_c, SU(2)_L)_Y$, whose electric charge can take values $Q = \pm 5/3, \pm 4/3, \pm 2/3, \pm 1/3$. Leptoquarks offer a plausible interpretation of B anomalies because they would enter at the loop level in interactions that showed no deviation from the SM (into four quarks or four leptons), while they could contribute to the semileptonic B transitions at the tree level, where anomalies are seen. The fact that leptoquarks have not yet been directly observed would also be compatible with the predicted mass of $\mathcal{O}(1)$ TeV.

To assess the compatibility of the U_1 LQ model with recent B anomalies, the effective Hamiltonian is first transformed into the form introduced in Ref. [33] and the two Wilson coefficients ($\mathcal{C}_{LL}^c$ and $\mathcal{C}_{LR}^c$) are fitted using data from three sets of related measurements,

- LFU ratios R_D , R_{D^*} , and R_{A_c} from $b \rightarrow c\tau\bar{\nu}_\tau$ [35].
- $\mathcal{B}(B^- \rightarrow \tau^- \bar{\nu}_\tau)$ [11, 36].
- Drell-Yan production $pp \rightarrow \tau^+ \tau^-$, dominated by $b\bar{b} \rightarrow \tau^+ \tau^-$ [37, 38].

The first two results are collectively referred to as low-energy observables and the latter as high-energy observables. A recent fit to low-energy observables is shown in fig. 1.8 [39].

Two scenarios of U_1 LQs are considered, one with a purely left-handed interaction ($\mathcal{C}_{LR}^c = 0$), and another with right-handed currents of equal magnitude ($\mathcal{C}_{LL}^c = -\mathcal{C}_{LR}^c$). Both scenarios are found to be equally compatible with the $b \rightarrow c$ observables, as seen in the left panel of fig. 1.8. Comparison of the model with $pp \rightarrow \tau^+ \tau^-$ constraints shows that low-and high-energy data are compatible too. Using the parameter space preferred by the low-energy fit to the U_1 model, one can infer predictions for $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)$ [39]. The right panel of fig. 1.8 shows these predictions in both U_1 LQ scenarios as functions of the relative deviation in R_{D^*} from the SM prediction. The resulting non-SM predictions for the $B^+ \rightarrow K^+ \tau^+ \tau^-$ branching fraction are about two orders of magnitude larger than the expected SM rate, indicating the attractive possibility to probe conclusively the LQs hypothesis with sensitivity that may be already available in present data.

In summary, extensions to the SM, such as U_1 LQs, that could explain current B anomalies may point to the presence of non-SM couplings that favor τ leptons and increase the $b \rightarrow s\tau^+ \tau^-$ transition rates by large factors. This constitutes a strong motivation to search for these modes given the paucity of available results. The work exposed in this thesis aims at partially addressing that.

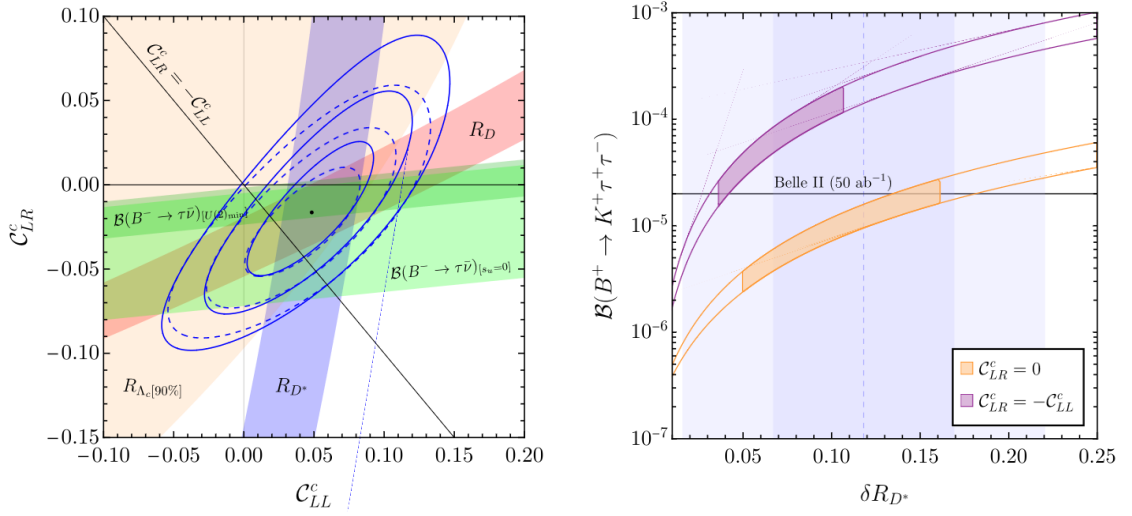


Figure 1.8: (Left) An example two-dimensional χ^2 fit of Wilson coefficients to the results from low-energy observables available in 2023. The black dot indicates the best-fit result, $(0.05, -0.02)$. The full, blue elliptical lines denote one, two, and three standard-deviation contours based only on $b \rightarrow c$ observables. The dashed ellipses are obtained by additionally including $\mathcal{B}(B^- \rightarrow \tau^- \bar{\nu}_\tau)$ measurements. Preferred regions corresponding to $\Delta\chi^2 = 1$ for each observable are indicated as bands, except for R_{A_c} , in which the orange-shaded surface indicates the 90% C.L. region. The two bands associated with $\mathcal{B}(B^- \rightarrow \tau^- \bar{\nu}_\tau)$, represent two different symmetry-breaking scenarios of left-handed quark doublets. (Right) Prediction for $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)$ according to the U_1 model, as a function of $\delta R_{D^*} = (R_{D^*}/R_{D^*}^{\text{SM}}) - 1$. The filled orange and purple regions are the 90% CL regions preferred by the low-energy fit. The blue vertical bands denote current experimental one or two standard-deviation regions for δR_{D^*} . The black line shows the estimated Belle II sensitivity with a 50 ab^{-1} sample [40]. Reproduced from Ref. [39].

Chapter 2

The Belle II experiment at the SuperKEKB collider

The data used in this work are collected by the Belle II experiment. This chapter outlines the Belle II detector at the SuperKEKB accelerator, with emphasis on the subdetectors more relevant for the reconstruction of $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays.

2.1 The SuperKEKB collider

SuperKEKB is an electron-positron (e^+e^-) energy-asymmetric collider, designed to produce more than 600 $B\bar{B}$ pairs per second ($B^0\bar{B}^0$ and B^+B^- in approximately equal proportions) via decays of $\Upsilon(4S)$ mesons produced at threshold [41]. Such colliders are called ‘ B -factories’, and were proposed in the 1990’s for the dedicated exploration of CP violation in B mesons. The main goal of B -factories is to produce low-background quantum-correlated $B\bar{B}$ pairs at high rates and with sufficient boost to study their time evolution.

Intense beams of electrons and positrons are brought to collision at the energy corresponding to the $\Upsilon(4S)$ meson mass, 10.58 GeV, which is just above the $B\bar{B}$ production kinematic threshold. The great majority of collisions yield electromagnetic processes $e^+e^- \rightarrow e^+e^-$, $e^+e^- \rightarrow \gamma\gamma$, etc., (fig. 2.1) that are scarcely interesting and straightforwardly discarded using global event quantities such as charged-particle multiplicity or total energy detected in the event. More interesting for flavor physics are the collisions that produce hadrons, henceforth called hadronic events. Figure 2.2 shows the hadron-production cross-section in e^+e^- collisions as a function of the final-state mass. The various peaks are radial excitations of the Υ meson overlapping the nearly uniform background at about 4 nb from so-called continuum of lighter-quark pair-production from the process $e^+e^- \rightarrow q\bar{q}$, where q identifies a u , d , c , or s quark. These are useful for charm physics, some selected topics in hadron physics, and as control channels. At the $\Upsilon(4S)$ collision energy, approximately 70% of the hadronic events involves the production of continuum. The rest are $\Upsilon(4S)$ events, which decay to $B\bar{B}$ pairs more than 96% of the times. At-threshold production implies little available energy to produce additional particles in the $B\bar{B}$ events, resulting in low-background conditions. These are the collisions relevant for the analysis described in this work as they produce $B^+ \rightarrow K^+ \tau^+ \tau^-$ signal, among many other processes. Colliding beams of point-like particles imply precisely known collision energy, which sets stringent constraints on the collision’s kinematic properties, thus offering means of further background suppression. Since bottom mesons are produced in a strong-interaction decay, flavor is conserved, and the null net bottom content of the initial state implies production

of a flavorless $B\bar{B}$ pair. Even though B^0 and $\bar{B}^0$ undergo flavor oscillations before decaying, their time-evolution is quantum-correlated in such a way that no same-flavor B^0B^0 or $\bar{B}^0\bar{B}^0$ pairs are present at any time. Angular-momentum conservation implies that the decay of the spin-1 $\Upsilon(4S)$ in the two spin-0 bottom mesons yields total angular momentum $J = 1$. Because the simultaneous presence of two identical bosons in an antisymmetric state would violate Bose statistics, the system evolves coherently as an oscillating $B^0\bar{B}^0$ particle-antiparticle pair until either one decays. This allows efficient identification of the bottom (or antibottom) content of one meson at the time of decay of the other, if the latter decays in a final state accessible only by either bottom or antibottom states. This important capability is called ‘flavor tagging’ and allows measurements of flavor-dependent decay rates of neutral bottom mesons, as needed in many determinations of CP -violating quantities.

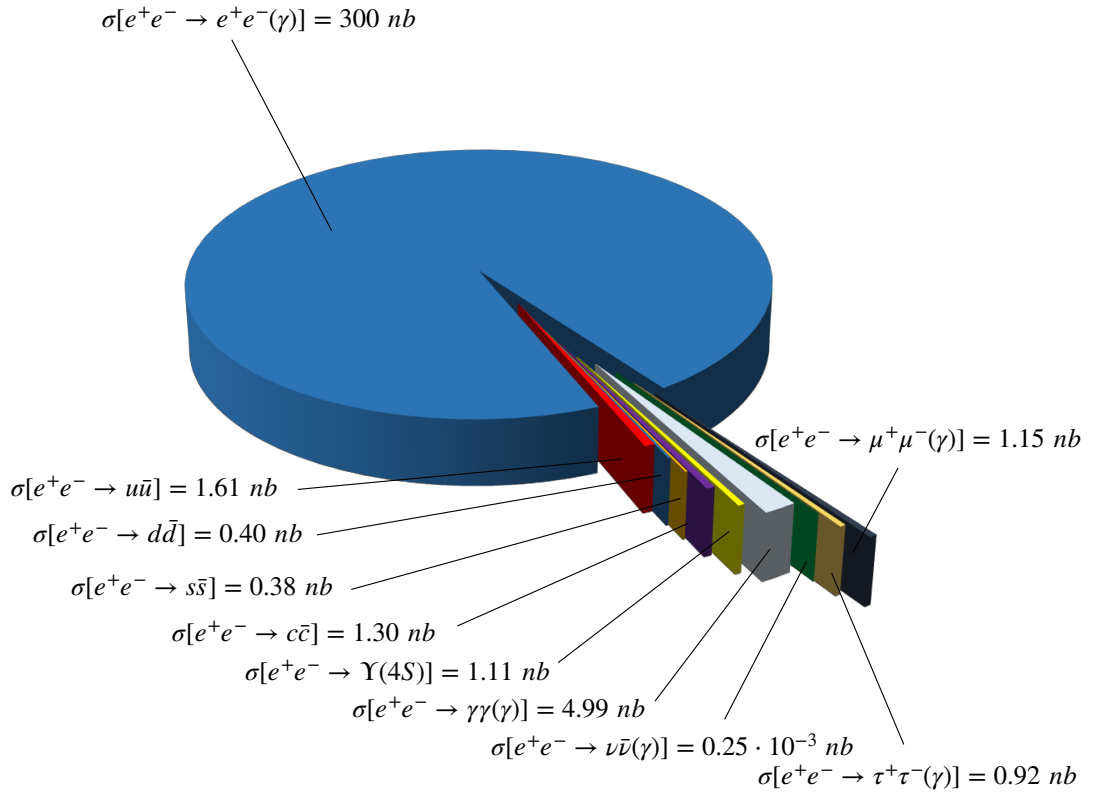


Figure 2.1: Cross sections of the main final states produced in e^+e^- collision at the $\Upsilon(4S)$ center-of-mass energy.

Because the $\Upsilon(4S)$ mesons are produced at threshold, they would be nearly at rest in the laboratory in an energy-symmetric collider. The resulting B mesons too would be produced with low momentum (about $10 \text{ MeV}/c$) in the laboratory, because of the $21 \text{ MeV}/c^2$ difference between the $\Upsilon(4S)$ mass and the $B\bar{B}$ pair mass. With such low momenta they would only travel approximately $1 \mu\text{m}$ before decaying, rendering the $10 \mu\text{m}$ typical spatial resolution of vertex detectors insufficient to separate B -decay vertices and study the decay-time evolution. Asymmetric beam energies are used to circumvent this limitation. By boosting the collision center-of-mass along the beam in the laboratory frame, B -decay vertex separations are achieved that are resolvable with current vertex detectors [42]. SuperKEKB

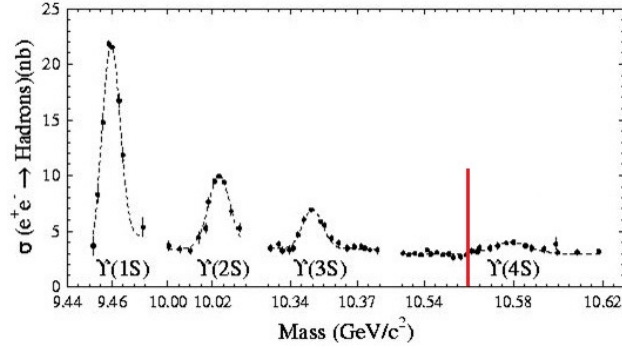


Figure 2.2: Hadron production cross section from e^+e^- collisions as a function of the final-state mass. The vertical red line indicates the $B\bar{B}$ production threshold.

(Fig. 2.3) implements a 7-on-4 GeV energy-asymmetric double-ring design, which achieves a vertex displacement of about $130 \mu\text{m}$.

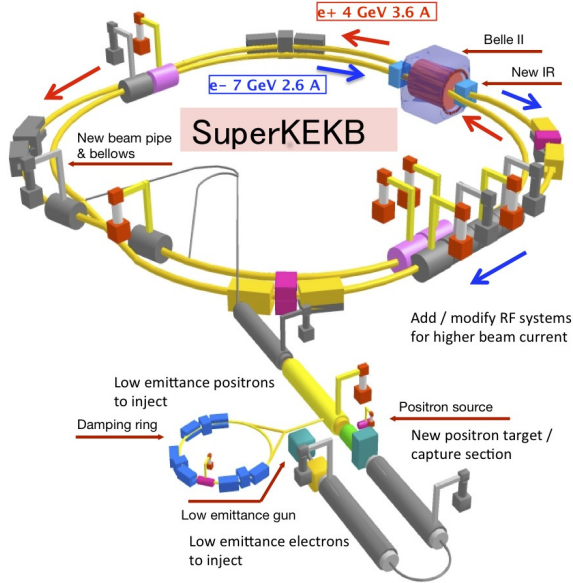


Figure 2.3: Illustration of the SuperKEKB collider.

Electrons are produced in a thermionic gun with a barium-impregnated tungsten cathode, then accelerated to 7 GeV with a linear accelerator (linac) and injected in the high-energy ring (HER). Positrons are produced by colliding electrons on a tungsten target, then isolated by a magnetic field, accelerated to 4 GeV with the linac and injected in the low-energy ring (LER).

The electrons and positrons continuously collide at a single interaction point, surrounded by the Belle II detector. To achieve high luminosities, a nano-beam, large crossing-angle collision scheme is implemented [43]. This is an innovative configuration based on keeping small horizontal and vertical emittance, which is a measure of the spread and size of the particle beam in the phase space of position and momentum, and large crossing angle, as shown in Fig. 2.4. Such configuration is obtained with the production of low emittance beams, in addition to a sophisticated final-focus superconducting-quadrupole-magnet system, made of magnets, corrector coils, and compensation solenoids installed

at each longitudinal end of the interaction region. Conceptually the nano-beam scheme mimics a collision with many short micro-bunches, allowing significant advantages in luminosity with respect to previous conventional schemes. The reduction of the luminous volume size to about 5% with respect to the predecessor KEKB, combined with doubled beam currents, is expected to yield a factor 40 gain in intensity. The penalty for such high intensities are significant challenges in achieving the design performance and operating steadily, and higher beam-induced backgrounds.

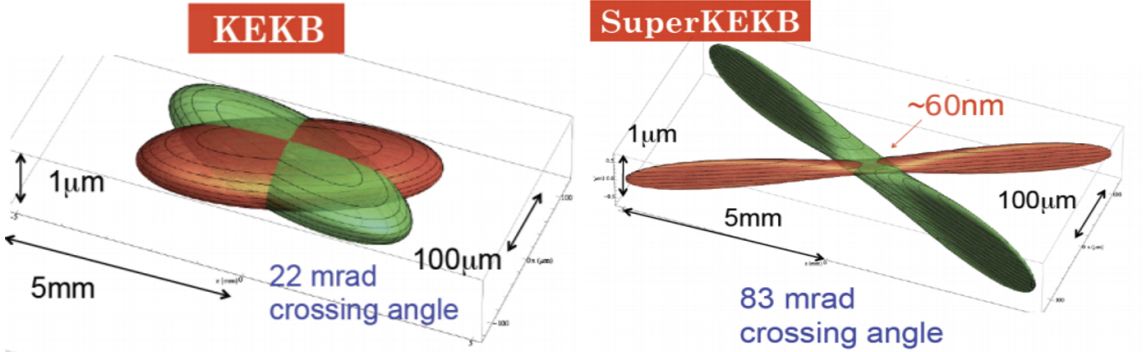


Figure 2.4: Two-dimensional sketch of the nano-beam mechanism implemented in SuperKEKB (right) compared with the previous KEKB collision scheme (left).

The performance of the SuperKEKB collider is mainly characterized in terms of the instantaneous luminosity $\mathcal{L}$, which is a measure of collision intensity,

$$\mathcal{L} = \frac{\gamma_{\pm}}{2e r_e} \left(1 + \frac{\sigma_y^*}{\sigma_x^*} \right) \frac{I_{\pm} \xi_{y\pm}}{\beta_{y\pm}^*} \cdot \frac{R_{\mathcal{L}}}{R_{\xi_y}}$$

where γ is the relativistic Lorentz factor, e is the absolute value of the electron charge, r_e is the classical radius of the electron, σ_x^* and σ_y^* are the bunch widths at the interaction point (IP) in the plane orthogonal to the beam direction (transverse plane), I is the current of the beam, β_y^* is the vertical betatron function at the IP [44], ξ_y is the vertical beam-beam parameter, $R_{\mathcal{L}}$ and R_{ξ_y} are the luminosity reduction factors and the vertical beam-beam parameter, respectively, due to non-vanishing crossing angle [45]. The ratio of these reduction factors is close to unity, while the design values for the other parameters are reported in Table 2.1. The rate of any given process

$$\text{rate} [\text{events s}^{-1}] = \mathcal{L} [\text{cm}^{-2} \text{s}^{-1}] \times \sigma [\text{cm}^2],$$

is the product of its cross-section and $\mathcal{L}$.

	Design	Achieved (as of 2023)
Energy [GeV]	4.0/7.0	4.0/7.0
ξ_y	0.090/0.088	0.0407/0.0279
β_y^* [mm]	0.27/0.41	1.0/1.0
I [A]	3.6/2.62	1.321/1.099

Table 2.1: Design and achieved values for SuperKEKB fundamental parameters for the LER/HER rings.

The integral of instantaneous luminosity over time T , called integrated luminosity,

$$\mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t') dt'$$

is a direct measure of the number of produced events of interest $N = \mathcal{L}_{\text{int}}\sigma$.

2.1.1 SuperKEKB performance

Physics data-taking started in March 2019, and Belle II has integrated $427.9 \pm 2.0 \text{ fb}^{-1}$ of luminosity at the time of this writing. In 2022, SuperKEKB achieved the instantaneous-luminosity world record, $4.7 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. In spite of these achievements, technological and scientific challenges have significantly reduced SuperKEKB performance with respect to design goals. A number of issues associated with beam injection, collimation, and short beam lifetime due to the reduction of their dynamic aperture, which also causes high uncontrollable beam backgrounds, has been limiting the capability to deliver the expected samples of data in its first five years. Consolidation, improvement and development work has been made to overcome these difficulties during the first long interruption of collider operations from mid-2022 to late 2023. However, operations in 2024 do not show yet evidence of major improvement.

2.2 The Belle II detector

Belle II (fig. 2.5) is a large-solid-angle, multipurpose magnetic spectrometer surrounded by a calorimeter and particle-identification systems, installed around the SuperKEKB interaction point. It is designed to determine energy, momentum, and identity of a broad range of particles produced in 10.58 GeV e^+e^- collisions. Belle II is approximately a cylinder of about 7 m in length and 7 m in diameter. It employs a right-handed Cartesian coordinate system with origin in the interaction point. The z axis corresponds to the principal axis of the solenoid, which is approximately parallel to the electron beam direction at the interaction point; the y axis points vertically upward, and the x axis is horizontal and pointing outward of the accelerator tunnel. The polar angle, θ , is referred to the positive z axis. The azimuthal angle, ϕ , is referred to the positive x axis in the xy plane. The radius, $r = \sqrt{x^2 + y^2}$, is defined in cylindrical coordinates and measured from the origin in the xy plane. Throughout this thesis, *longitudinal* means parallel to the electron beam direction (to the z axis), and *transverse* means perpendicular to the electron beam direction, i.e., in the xy plane.

Belle II comprises several subsystems, each dedicated to a specific aspect of event reconstruction. From the interaction point outward, a particle would traverse the beam pipe, a two-layer silicon-pixel vertex-detector (PXD), a four-layer silicon-strip vertex-detector (SVD), a central wire drift-chamber (CDC), a time-of-propagation central Cherenkov counter (TOP) or an aerogel threshold forward Cherenkov counter (ARICH), an array of CsI(Tl) crystals (ECL), a superconducting solenoidal magnet, and multiple layers of resistive plate counters (KLM).

The principal experimental strengths are hermetic coverage, which allows for reconstruction of final states involving neutrinos; efficient and precise reconstruction of charged-particle trajectories (tracks), which provide accurately reconstructed decay-vertices and good momentum resolution; high-purity charged-particle identification and neutral-particle reconstruction. A summary of the technological specifications of the Belle II subsystems is in Table 2.2. A detailed description of Belle II and its performance is in Ref. [46].

Purpose	Acronym	Technology	Configuration	Channels	Polar coverage (θ)
Beam pipe		Beryllium	Cylindrical, inner radius 10 mm, 10 μm Au, 0.6 mm Be, 1 mm paraffin, 0.4 mm Be		
Tracking	PXD	Silicon pixel	Sensor size: $15 \times (L1 \text{ } 136, L2 \text{ } 170) \text{ mm}^2$, Pixel size: $50 \times (L1a \text{ } 50, L1b \text{ } 60, L2a \text{ } 75, L2b \text{ } 85) \mu\text{m}^2$; two layers at radii: 14, 22 mm	10^6	$[17^\circ; 150^\circ]$
	SVD	Silicon strip	Rectangular and trapezoidal, strip pitch: $50(p)/160(n) - 75(p)/240(n) \mu\text{m}$, with one floating intermediate strip; four layers at radii: 38, 80, 115, 140 mm	2.45×10^5	$[17^\circ; 150^\circ]$
	CDC	Drift chamber with $\text{He-C}_2\text{H}_6$ gas	14336 wires in 56 layers, inner radius of 160mm outer radius of 1130 mm	1.4×10^5	$[17^\circ; 150^\circ]$
	Solenoid	Solenoid coil	Length: 4.41 m. Radius: 1.80 m		
Particle ID	TOP	RICH with quartz radiator	16 segments in ϕ at $r \approx 120$ cm, 275 cm long, 2cm thick quartz bars with 4×4 channel MCP PMTs	8×10^3	$[31^\circ; 128^\circ]$
	ARICH	RICH with aerogel radiator	2×2 cm thick focusing radiators with different n , HAPD photodetectors	7.8×10^4	$[14^\circ; 30^\circ]$
Calorimetry	ECL	CsI(Tl) crystals	Barrel: $r = 125 - 162\text{cm}$, end-cap: $z = -102 - +196\text{cm}$	6624 (Barrel), 1152 (FWD), 960 (BWD)	$[12.4^\circ; 31.4^\circ]$, $[32.2^\circ; 128.7^\circ]$, $[130.7^\circ; 155.1^\circ]$
Muon ID	KLM	Barrel: RPCs and scintillator strips	2 layers with scintillator strips and 12 layers with 2 RPCs	$\theta \text{ } 1.6 \times 10^4$, $\phi \text{ } 1.6 \times 10^4$	$[40^\circ; 129^\circ]$
	KLM	End-cap: scintillator strips	12 layers of $(7-10) \times 40 \text{ mm}^2$ strips	1.7×10^4	$[25^\circ; 40^\circ]$, $[129^\circ; 155^\circ]$

Table 2.2: Summary of the Belle II subdetectors and their specifications.

The beam pipe is a passive element that is strictly not part of the detector. However, the impact of its geometry and technological choices is so intertwined with key detector performances that it warrants a description here. The beam pipe is a 3 km-long vacuum enclosure to allow beams circulating inside the detector. In the following, I refer only to the straight section of the beam pipe surrounding the interaction point. Multiple Coulomb scattering in the beam-pipe wall of the final-state charged particles would spoil the vertex-position resolution; this dictates a thin beam-pipe wall made of a low- Z material. Moreover, since the vertex resolution is inversely proportional to the distance between the interaction point and the first track sampling, the beam pipe has to be narrow. The possibility for beam-halo to interact with the beam pipe, thus inducing beam backgrounds, and heating of the pipe wall due to charge induction complicates the design. Hence, the beam pipe is constantly cooled and shielded from the vertex detector. The Belle II beam pipe is made of two beryllium cylinders, 0.6 mm thick at radius of 10 mm, and 0.4 mm thick at radius of 12 mm, respectively. A 1.0 mm gap between the inner and outer walls of the pipe is filled with paraffin for cooling. The beam pipe is coated with a 10 μm gold sheet that absorbs low-energy photons, which could damage the silicon detector.

2.2.1 Tracking system

At Belle II, reconstruction of charged particles and ensuing measurement of their momenta and charges is achieved through an integrated system consisting of six layers of silicon and a drift chamber, surrounding the beam pipe and immersed in a 1.5 T axial magnetic field maintained in a cylindrical volume 3.4 m in diameter and 4.4 m in length. The field is oriented along the z direction and provided by an aluminum-stabilized superconducting solenoid made of NbTi/Cu alloy. The solenoid surrounds all the subdetectors up to the KLM. The iron yoke of the detector serves as the return path of the magnetic flux.

2.2.1.1 Silicon-pixel vertexing detector

The innermost detector is a pixel vertex detector (PXD) [47]. Its goal is to sample the trajectories of final-state charged particles in the vicinity of the decay position (vertex) of their long lived ancestors, so that the decay point can be inferred by extrapolation inward.

PXD sensors are based on the technology of depleted field-effect transistors [47]. They are made of p-channel MOSFET integrated on a silicon substrate, which is fully depleted by applying an appropriate voltage. Incident particles generate electron-hole pairs in the depleted region. The charge carriers drift towards the minimum of potential placed under the transistor channel, and thus modulate a current passing through the MOSFET. Sensors are 75 μm thick.

The PXD has two layers at 14 mm and 22 mm radii, respectively, and a full length of 174 mm at the radius of the outer layer. It comprises around 8 million pixels, $50 \times (50 - 55)\mu\text{m}^2$ (inner layer) and $50 \times (70 - 85)\mu\text{m}^2$ (outer layer) each. The polar acceptance ranges from 17° to 150° . The design impact-parameter resolution is 12 μm , achieved by weighting the charge deposited in neighboring pixels. For the data used in this thesis, the full first pixel layer is used, along with a 1/6 azimuthal sector of the second layer, as completion of the pixel detector only happened in 2023.

2.2.1.2 Silicon-strip vertexing detector

Around the PXD is SVD, a silicon detector aimed at reconstructing decay vertices and low-momentum charged-particles at high resolution [48].

Items	Parameters
Cryostat	
Radius: outer/inner	2.00 m/1.70 m
Central field	1.5 T
Total weight	23t
Effective cold mass	≈ 6 t
Length	4.41 m
Coil	
Effective radius	1.8 m
Length	3.92 m
Conductor dimensions	$3 \times 33 \text{ mm}^2$
Superconductor	NbTi/Cu
Stabilizer	99.99% aluminium
Nominal current	4400 A
Inductance	3.6 H
Stored energy	35 MJ
Typical charging time	0.5 h
Liquid helium cryogenics	Forced flow two phase
Cool down time	< 6 days
Quench recovery time	< 1 day

Table 2.3: Main design parameters of the solenoid coil.

It uses double-sided silicon strip sensors. Each sensor is made of a silicon n-doped bulk on one side, and a perpendicular highly p-doped implant on the other side. This means that for each sensor, one side has strips parallel to the beams direction, and the other perpendicular. A voltage is applied to enhance the depletion region at the p-n junction, and removes intrinsic charge-carriers from the region. Traversing charged particles ionize the silicon, freeing electron-hole pairs that drift due to the electric field thus inducing a signal in highly granular strip electrodes implanted at both ends of the depletion region. The fine segmentation and fast charge collection of SVD sensors make possible to deal with large track density environments.

The SVD structure consists of four concentric layers at radii of 39, 80, 104 and 135 mm, composed by, respectively, 7, 10, 12, and 16 independently-readout longitudinal modules called ladders, arranged in a cylindrical geometry. As shown in fig. 2.6, SVD has a polar-asymmetric geometry that mirrors the asymmetry in particle density resulting from the center-of-mass boost. The polar acceptance ranges from 17° to 150° .

Sensors are $300 \mu\text{m}$ thick, and the separation between adjacent strips (d_{pitch}) ranges from $50 \mu\text{m}$ to $240 \mu\text{m}$. Hence, the nominal spatial resolution $d_{\text{pitch}}/\sqrt{12}$ varies with the polar angle. Since the charge associated with an incident particle is usually distributed among several strips, position resolution is improved by interpolation.

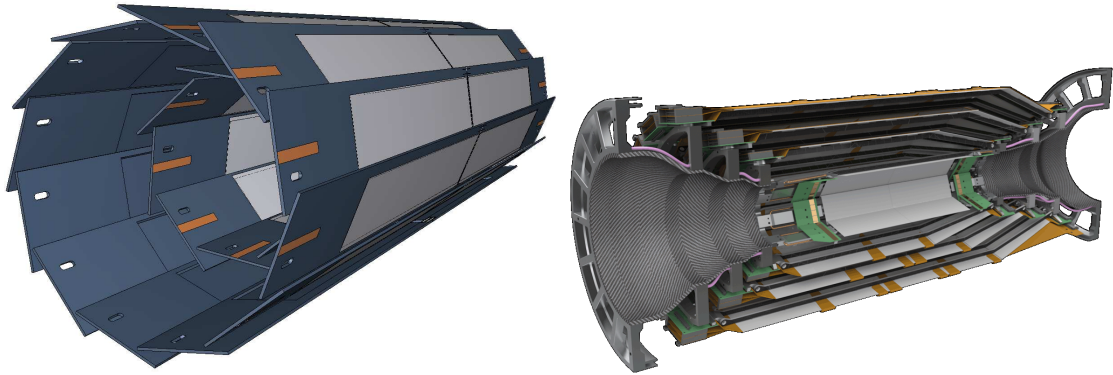


Figure 2.6: (Left) sketch of the PXD detector and (right) exploded view of a SVD detector half.

2.2.1.3 Central drift chamber

The CDC is a drift chamber [49]. It samples charged-particle trajectories at radii between 16 cm and 113 cm, thus providing accurate measurements of momentum and electric charge, trigger information for events containing charged particles, and information on identification of charged-particle species by measuring their specific-ionization energy-loss (dE/dx). When a charged particle traverses the CDC volume, it ionizes the gas, freeing electrons and positive ions from gas atoms. A stationary electric field then accelerates these charges until they approach the sense wires. In their vicinity high field gradients cause an abrupt acceleration that causes secondary ionizations, which induce an electric signal whose time is digitized. The particle trajectory is inferred from the time between the collision and the signal.

The chamber volume contains 14336 30- μm -diameter sense wires, divided in 56 layers, immersed in a gaseous mixture of 50% He and 50% C_2H_6 , while 42240 126- μm -diameter aluminum wires shape the electric field. Layers of wires are installed with either "axial" orientation, i.e., aligned with the solenoidal magnetic field, or skewed with respect to the axial wires with a "stereo" orientation. The azimuthal acceptance ranges from 17° to 150° .

The spatial resolution is about 100 μm and the dE/dx resolution is 11.9% for an incident angle of 90° . Figure 2.7 shows a sliced view of the CDC and the wire configurations.

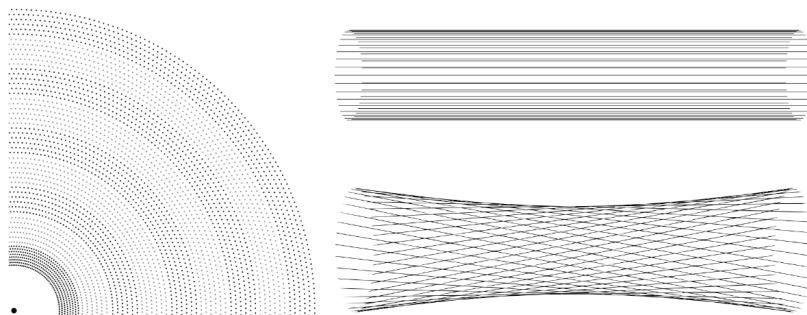


Figure 2.7: (Left) A transverse quadrant of the drift chamber, organized in layers, called *superlayers*; wire orientation for (top-right) axial and (bottom-right) stereo layers. Skew is exaggerated for visualization purposes.

2.2.2 Electromagnetic calorimeter

The ECL measures the energy of photons and electrons [50]. High energy photons and electrons entering the calorimeter initiate an electromagnetic shower through bremsstrahlung and electron-positron pair production. The energy is mostly converted to photons, which are collected by the photodiodes. In contrast to hadrons, which pass through the calorimeter with minimal energy loss, most photons and electrons dissipate their entire energy.

The configuration, mechanical structure, and crystals of Belle II ECL are those of the Belle's calorimeter. The readout electronic boards have been upgraded to cope with SuperKEKB's higher luminosity. The layout is shown in fig. 2.8. The ECL consists of three polar compartments: the barrel, the forward endcap, and the backward endcap section. The barrel section is 3.0 m long with 1.25 m of inner radius; the endcaps are located at $z = +2.0$ m (forward) and -1.0 m (backward) from the interaction point. Table 2.4 summarizes the geometrical parameters of each section.

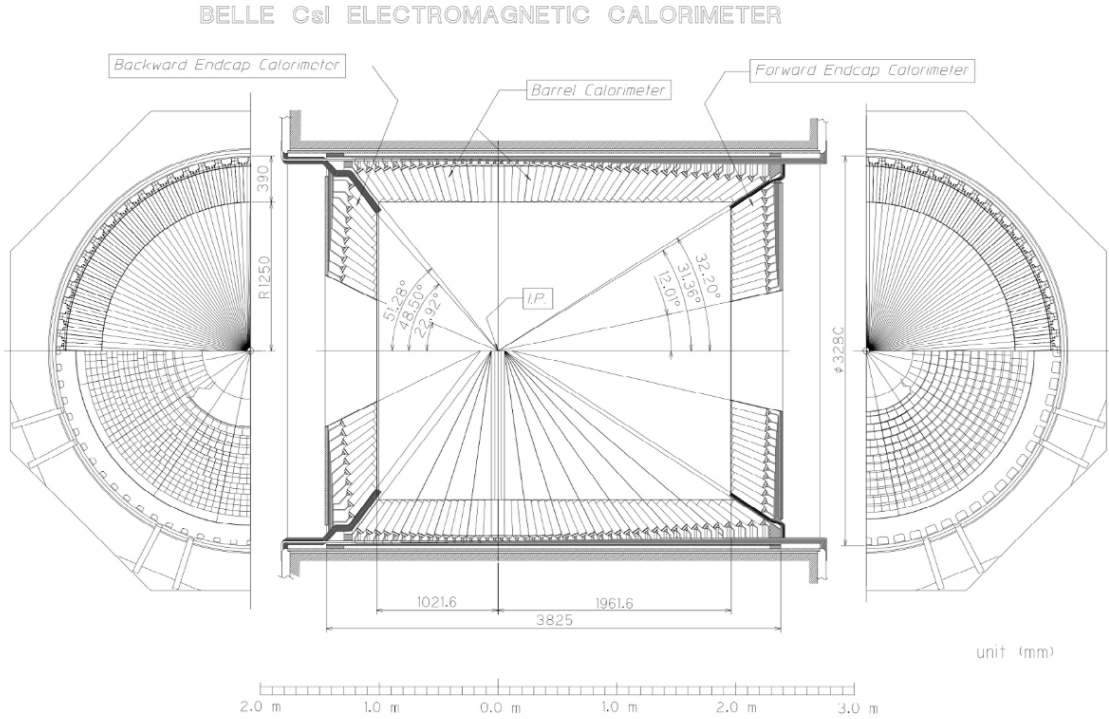


Figure 2.8: ECL layout.

Item	θ coverage	θ segmentation	ϕ segmentation	Number of crystals
Forward endcap	12.4°–31.4°	13	48–144	1152
Barrel	32.2°–128.7°	46	144	6624
Backward endcap	130.7°–155.1°	10	64–144	960

Table 2.4: Summary of ECL parameters.

High momentum π^0 detection is essential in this work and requires good separation of two nearby photons and a precise determination of the opening angle. This requires a segmented calorimeter. The ECL is a highly segmented array of 8736 cesium iodide crystals

doped with thallium (CsI(Tl)). Thallium shifts the energy of the excitation light into the visible spectrum. The light is detected by a independent pair of silicon PIN photodiodes [50] and charge-sensitive preamplifiers installed at the outer end of each crystal.

A typical crystal in the barrel section has a $55 \times 55 \text{ mm}^2$ active surface on the front face and $65 \times 65 \text{ mm}^2$ on the rear face; the dimensions of the crystals in the endcap sections vary from 44.5 to 70.8 mm and from 54 to 82 mm for front and rear faces, respectively. A diagram of an ECL crystal is shown in fig. 2.9. The 30-cm crystal length, corresponding to 16.1 radiation lengths reduces the fluctuations of shower leakages out of the outermost end of the crystals, which spoils energy resolution. The crystals are designed in such a way that a photon injected at the center of the crystal would deposit 80% of its energy in the crystal on average. The crystals principal axes do not point exactly to the nominal interaction point, but they are inclined to prevent photons from escaping through gaps between crystals by about 1.3° in the θ and ϕ directions in the barrel section, and by about 1.5° and about 4° in the θ direction in the forward and backward sections.

Considering the ECL structure – gaps, crystal wrapping, mechanical structure – the fraction of photons that do not leave a detectable signal in the calorimeter is only 0.2%.

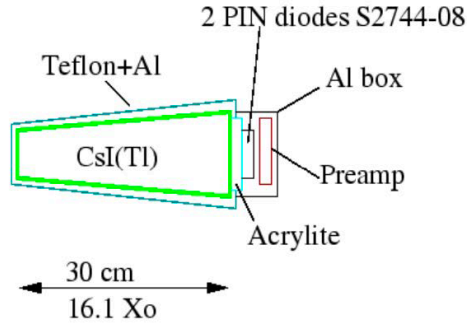


Figure 2.9: Schematic design of a CsI(Tl) crystal with attached readout electronic circuits.

Signals from the photodiodes are sent to two preamplifiers mounted on the rear of the crystal for charge integration. The two resulting signals are sent to a readout board [51] located outside the Belle II detector and containing 16 analog shaper circuits. These channels receive signals from up to 16 CsI(Tl) crystals. A total of 576 shaper modules are needed to process signals from all 8736 crystals. Shaper circuits amplify and shape the analog signal to provide a clean and well-defined pulse while removing noise, pile-up due to overlapping pulses from neighbouring events, and other unwanted features from the signal. The signal is sampled by a digitizer at 1.76 MHz, which corresponds to an interval between measurements of 567 ns. After collecting 31 samples, the signal waveform is processed using a photon template fit to compute the signal amplitude of the signal, the time relative to the trigger signal, and the χ^2 fit quality.

The photon emission spectrum peaks at around 550 nm, which is convenient for photodiode readout. However, the time for the light in the crystals to decay is relatively long, increasing considerably the overlap of pulses from neighboring (background) events. This means that scintillation light may be present when a particle from a later event arrives, generating pile-up background.

The ECL also uses Bhabha scattering to measure luminosity. Because the Bhabha cross section is predicted with high accuracy in QED, a precise inference of luminosity is

achieved from the measured rate of Bhabha events in a volume of known acceptance.

2.2.3 Particle identification

Belle II combines measurements of time-of-propagation, Cherenkov radiation, and ionization energy loss in the tracker and drift chamber to identify charged particles.

2.2.3.1 Time-of-propagation detector

The TOP detector measures the time of propagation of the Cherenkov photons emitted from charged particles passing through its quartz bars and internally reflected within a radiator [52]. It is made of 16 quartz bars mounted at 1.2 m from the IP. Each bar has three main components (fig. 2.10): a long bar acts as Cherenkov radiator, where photons are generated and propagated; a focusing mirror is mounted at the forward end; and a prism mounted at the backward end collects photons and guides them to a photomultiplier. The polar coverage ranges from 31° to 128° . On average, photons originated from slower particles take more time to reach the photomultipliers, because of the inverse proportionality between β and the cosine of the Cherenkov photon-emission angle.

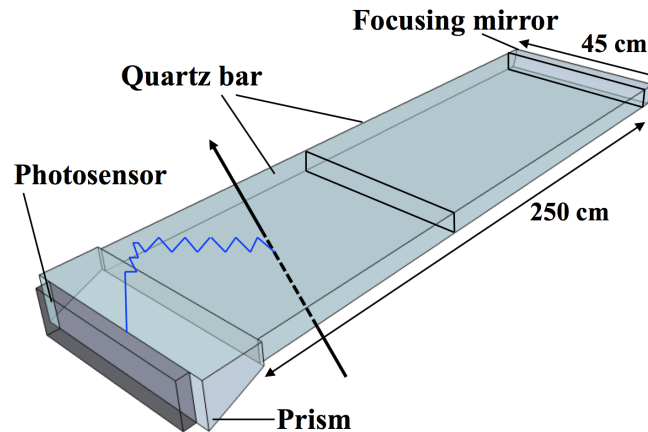


Figure 2.10: Scheme of a TOP bar. A charged particle crossing the radiator and emitting Cherenkov photons, which are collected at the photomultipliers, is also represented.

2.2.3.2 Aerogel ring-imaging Cherenkov counter

The ARICH detector identifies charged particles by measuring the Cherenkov ring produced when passing through a radiator [53]. It consists of 420 modules for photon detection in seven layers extending from 0.56 to 1.14 m radius, and 248 aerogel tiles installed on the detector endcaps. The aerogel radiator produces Cherenkov photons when traversed by charged particles of a certain momentum range. Next to the radiator is an expansion volume where photons are propagated, to form rings on position-sensitive photodiodes. Photocathodes then convert photons into photoelectrons and generate electric signals. Two adjacent radiators with different refraction indexes generate enough photons for achieving sufficient resolution, as shown in fig. 2.11.

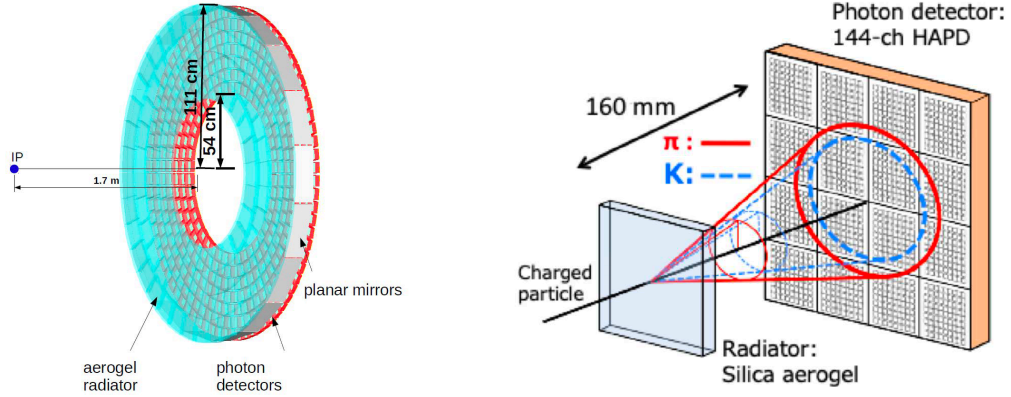


Figure 2.11: Sketch of (left) the ARICH with its main components and (right) diagram of the difference in the photon path for Cherenkov photons from kaons and pions.

2.2.3.3 K_L^0 and muon detection system

The KLM detects muons and neutral particles that do not get absorbed in the inner detectors, such as K_L^0 mesons [54]. It is made of alternating 4.7-cm-thick iron plates and active detector elements. Iron elements act also as magnetic flux returns for the tracking solenoid. In the inner layers, the active material is scintillator, in the outer layers are glass-electrode resistive-plates chambers, with a gas mixture filling the space between electrodes. When particles traverse the KLM, they produce charges that are collected by applying an appropriate voltage. The barrel section of the detector covers 45° to 125° in polar angle. The endcaps cover 20° to 45° and 125° to 155° .

2.2.4 Trigger and data acquisition system

The e^+e^- collisions at the $\Upsilon(4S)$ resonance occur with an approximate 200 MHz rate and produce a variety of processes. As the events of interest are only a fraction of the total cross section and it would be impossible to record all collisions on permanent memory, an online event-selection system (trigger) is used to distinguish them from background in real time, and to feed only the interesting events to the data acquisition system (DAQ), compatibly with data processing resources. The physics processes of interest include hadronic, μ/τ -pair, Bhabha, and two photon events. Accept rates of Bhabha and $\gamma\gamma$ events, which have high cross section and can be identified by their distinct signature, are artificially reduced by a factor of 100 to comply with the data acquisition limitations. Preferably discarded events include beam-related background resulting from synchrotron radiation, scattering of the beams on the residual gas, interactions in the beam pipe, and cosmic-ray events.

The Belle II trigger is organized according to a two-level logic, with a level 1 (L1) hardware trigger followed by a software-based, high-level trigger (HLT).

The L1 trigger, designed for a maximum rate of 30 kHz, uses input from four subdetectors: (i) the CDC, that provides three-dimensional track information to suppress tracks not originating from the interaction point; (ii) the ECL, that gives information on total energy deposit and cluster multiplicity; (iii) the TOP, that provides timing and hit topology information; and (iv) the KLM, that gives high-efficiency trigger for muons. These are used to achieve a low-level reconstruction that is fed to the global decision logic, which sends the proper trigger signal if the event passes the selection requirements. The L1 logic is implemented using field-programmable gate arrays that have a fixed latency of 5 μ s, with

an uncertainty on the trigger timing (jitter) of approximately 10 ns.

Cross sections and expected L1 trigger accept rates for physics processes of interest at the design instantaneous luminosity of $8 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$ are given in Table 2.5.

Process	σ [nb]	Trigger rate [Hz]
$e^+e^- \rightarrow \Upsilon(4S)$	1.2	960
$e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$)	2.8	2200
$e^+e^- \rightarrow \mu^+\mu^-$	0.8	640
$e^+e^- \rightarrow \tau^+\tau^-$	0.8	640
$e^+e^- \rightarrow e^+e^-$ (Bhabha scattering, $\theta_{\text{lab}} > 17^\circ$)	44	350*
$e^+e^- \rightarrow \gamma\gamma$ ($\theta_{\text{lab}} > 17^\circ$)	2.4	19*
Two-photon events ($\theta_{\text{lab}} > 17^\circ$ and $p_T > 0.1 \text{ GeV}/c$)	≈ 80	≈ 1500

Table 2.5: Expected cross sections and trigger rates ($\sigma \times \mathcal{L}$) of various physics processes at $8 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$ luminosity [46]. Bhabha and $\gamma\gamma$ accept rates (*) are artificially reduced by a factor of 10^2 to cope with data-acquisition limitations.

Events selected by the L1 trigger are input to the HLT, which makes a decision using information from all the subdetectors except for PXD. The online software reconstruction is similar to that used offline. A first selection, performed after the first step of the reconstruction and aimed at discarding about half of the events, is based on requirements on track multiplicity, vertex position, and total ECL energy deposit. After the remaining steps of the standard reconstruction are completed, further physics-level selection are performed. After this stage, the number of events is reduced to about 1/5 of those passing the L1 trigger. The efficiency of the HLT for $\Upsilon(4S) \rightarrow B\bar{B}$ events is higher than 99%.

Data from the PXD for events that pass the L1 selection are stored in a dedicated online data reduction system. Once an event passes the selection, HLT extrapolates the tracks found by CDC and SVD to the PXD layers, defining regions of interest (ROIs). These are passed to the data reduction system, and only hits matching with a ROI are transmitted to the DAQ system. This keeps the PXD data size to about 100 kB/event.

Fully reconstructed events are stored in DST files. The size of a DST of a typical hadronic event is 100 kB. The large amount of information stored in DST files is reduced into mini-DST to isolate subsets of events of physics processes of interest like hadronic events. The size of a mini-DST of an hadronic event is around 40 kB.

2.3 Reconstruction of stable particles

Reconstruction is the process through which raw data collected by the detectors are transformed into manageable physics information, in terms of quantity, quality, and meaningfulness. Several algorithms use low-level objects (detector signals, alignment, and calibration information) combined with our knowledge of relativistic kinematics to produce higher-level objects (tracks, energy deposits, etc). In the work described in this thesis I use information associated to two types of stable particles, that are particles that do not decay prior to the tracking system, and that are reconstructed by the detector: charged particles, related for example to the second B meson produced in the collision or reconstructed in the final states of various control channels, and photons. An outline of the essential aspects of the reconstruction of these, along with the associated performance quantities follows.

2.3.1 Charged-particle reconstruction

The ideal trajectories of charged particles in a solenoidal magnetic field are helical, with radius proportional to their transverse momentum. This ideal configuration can be altered by effects such as Coulomb scattering or other energy losses. When reconstructing a track, that is, measuring its momentum and position of closest approach to the interaction point, we need to take into account for these possible effects.

Track reconstruction, or “tracking”, in Belle II [55] consists in the combination of sequences of hits (measurement space-points) into tracks (full trajectories) after a charged particle crosses multiple active layers. The first step is called track finding; the second, track fitting. Tracking relies on PXD, SVD, and CDC information. Due to the different properties of these detectors, specific algorithms are used for each.

As a first step of track finding, hits in the outer tracking volume (CDC), where lower occupancy aids track finding, are filtered and reconstructed by two independent algorithms. One is a global track finding based on the Legendre algorithm [56], that transforms the position of each hit into a (θ, ρ) pair, which represents all the circles traversing both the IP and the considered hit. Another is a local algorithm that takes into account possible non-circular trajectories. The global track finding searches for patterns of hits consistent with helical trajectories, accounting for layer inefficiencies, while local track finding detects extended patterns of nearby hits, to complement the global search and detect short tracks and tracks displaced from the IP. The results of both algorithms are merged and the resulting CDC-only tracks are fitted by an iterative fitter based on the Kalman filter technique, that accounts also for possible random perturbations on the trajectory due for example to multiple scattering or energy losses [57].

Then, tracks are extrapolated inward making sure to avoid duplications, and SVD information is added. They are fitted again, before being extrapolated further inward to the PXD to define regions of interest around their expected intersection points. If an excited pixel is found inside this region, it is included in the pattern recognition algorithm, otherwise it is discarded.

Finally, the parameters of the track are determined thanks to a fitting algorithm and by assuming a mass hypothesis (fig. 2.12):

- d_0 , the distance of the point of the closest approach to the z axis;
- ϕ_0 , the angle between the transverse momentum and the x axis at the point of the closest approach;
- ω , the track curvature signed according to the particle charge;
- z_0 , the z coordinate at d_0 ;
- $\tan \lambda$, the tangent of the angle between track momentum and transverse momentum.

Track reconstruction is subjected to uncertainties and errors. A track might sometimes be a fake track, if it includes hits from beam-induced background or combines hits from two different particles, or a clone track, if other tracks are reconstructed from the same particle.

Tracking efficiency, that is the efficiency in reconstructing the track of a particle produced after a collision in the detector acceptance, varies from 75% at low transverse momenta ($\mathcal{O}(10)$ MeV) to 95% around 4 GeV/ c . It degrades the closer the track is to the beam axis (small or large polar angles), while it is mostly constant around 90% regardless of the azimuthal angle.

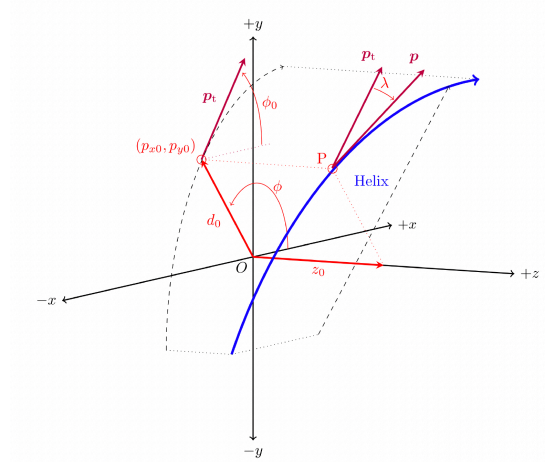


Figure 2.12: Three-dimensional representation of the helical trajectory of a track at the point of closest approach (P) to the IP (the origin O). Symbol p is the momentum of the charged particle at the point P , p_t its transverse momentum and λ is the angle between the two vectors.

The observed transverse momentum resolution is $\sigma(p_T)/p_T = 0.0011p_T[\text{GeV}/c] \oplus 0.0025/\beta$ as shown in fig. 2.13. The momentum- and angle-dependent impact parameter resolutions are $\sigma_{xy} = 10 \oplus 25/(p\beta\sin^{3/2}\theta)\mu\text{m}$ and $\sigma_z = 15 \oplus 27/(p\beta\sin^{5/2}\theta)\mu\text{m}$ for the transverse and longitudinal projections, respectively.

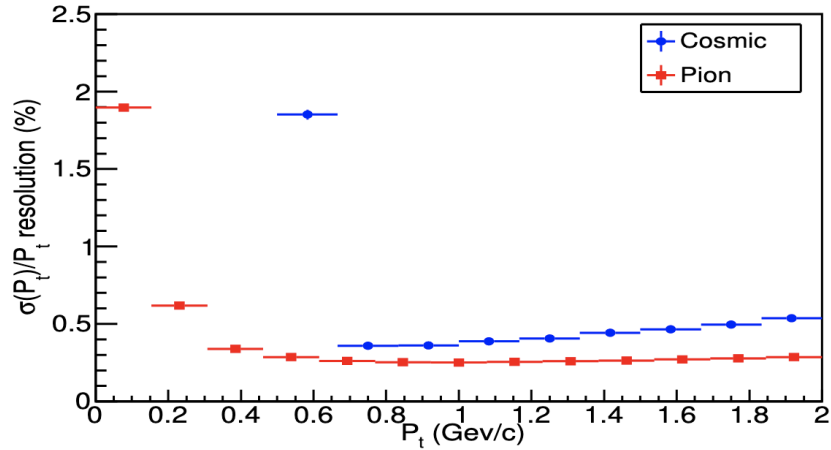


Figure 2.13: Transverse momentum resolution for collision and cosmic ray data.

2.3.2 Charged-particle identification

Particle identification is essential in flavor physics as most of the interesting channels are suppressed and therefore affected by signal-like backgrounds that only differ by the identity of some final-state hadrons. Particle identification (PID) at Belle II is achieved by combining information from several subdetectors. The trajectories of charged particles reconstructed by the tracking detectors, PXD, SVD, and CDC, are extrapolated outward to the TOP, ARICH, ECL and KLM detectors, where geometric matching between the tracks and observed signals is attempted. Offline reconstruction associates PID-detector

information sensitive to its identity to each matching track. For example, the drift chamber output encodes information on the specific ionization energy loss associated with each track. The raw information is further processed to provide higher-level quantities that are more convenient for usage in analysis. These are typically ‘likelihood’ values associated to the track. For each of six possible mass hypotheses, kaon, pion, electron, muon, proton and deuteron, the likelihood expresses the probability to observe the detected PID if the mass hypothesis was true.

For each detector and particle-hypothesis, the likelihood is usually obtained by comparing the expected and the observed value of the raw information, taking into account the uncertainties. For instance, in the CDC such information is $dE/dx_{\text{obs}}(h)$, the specific-ionization energy-loss observed for a charged particle h , averaged across the CDC wires. The resulting (natural logarithm of) the likelihood is

$$\ln \mathcal{L}_{\text{hyp}}^{\text{CDC}}(h) = -\frac{1}{2} \left[\frac{\frac{dE}{dx}_{\text{obs}}(h) - \frac{dE}{dx}_{\text{exp-hyp}}(h)}{\sigma_{\text{obs}}(h)} \right]^2, \quad (2.1)$$

where ‘hyp’ represents the particle hypothesis and $\sigma_{\text{obs}}(h)$ is the observed uncertainty on $dE/dx_{\text{obs}}(h)$, which mainly depends on the number of CDC hits associated to h . The expected value $dE/dx_{\text{exp-hyp}}(h)$ is the average ionization-energy loss from a charged particle h that has the observed momentum, assuming the hypothesis ‘hyp’, calculated using the Bethe-Bloch equation [58, 59] modified according to minor empirical adjustments to adapt to the details of the CDC response. Figure 2.14 shows the $dE/dx_{\text{obs}}(h)$ distribution for various particle species in Belle II data and the expected energy loss for each of the six mass hypotheses considered.

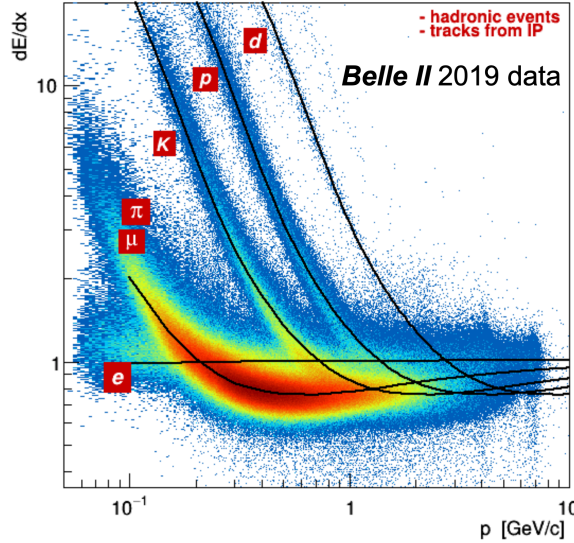


Figure 2.14: Distribution of (points) observed ionization-energy loss as a function of momentum for charged particles from hadronic events reconstructed in Belle II data, along with (solid lines) average expected values. Reproduced from Ref. [60].

In the TOP, the likelihood is calculated by comparing the observed number of detected photons associated to the charged particle with the photon yield expected from simulation [61, 62],

$$\ln \mathcal{L}_{\text{hyp}}^{\text{TOP}}(h) = \sum_{i=1}^N \ln \left[\frac{S_{\text{hyp}}(x_i, t_i, h) + B(x_i, t_i)}{N_e(h)} \right] + \ln P_N(N_e(h)), \quad (2.2)$$

where x_i and t_i are, respectively, the positions and times of arrival of the N Cherenkov photons excited by the charged hadron h . The term $S_{\text{hyp}}(x, t, h)$ is the signal distribution for the hypothesis ‘hyp’; $B(x, t)$ is the distribution of background; and $N_e(h) = N_{\text{hyp}}(h) + N_B$ is the expected number of detected photons, which is the sum of the expected number of signal photons $N_{\text{hyp}}(h)$ for hypothesis ‘hyp’ and background photons N_B . The second term in Eq. (2.2) is a probability for a Poisson with mean N_e to generate N photons [61]. Figure 2.15 shows an example of the identification of a kaon in the TOP detector: the positions and arrival times of Cherenkov photons are compared with the values expected for a pion or a kaon.

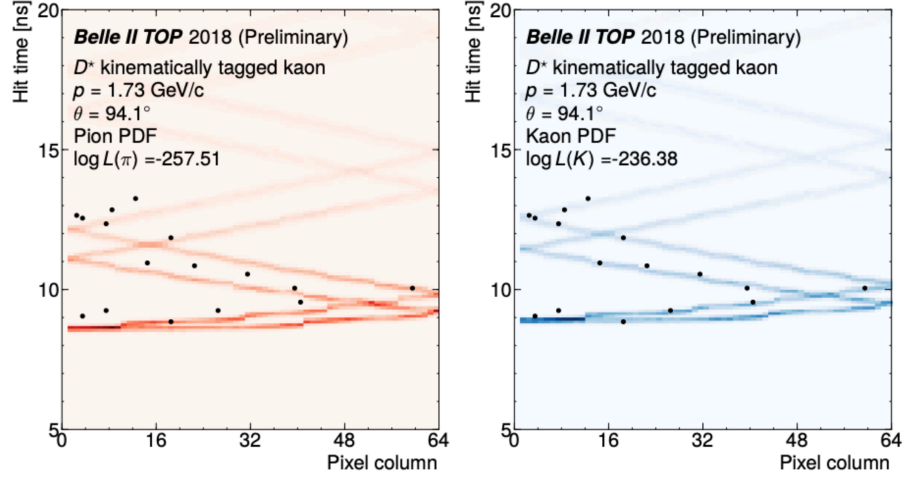


Figure 2.15: Example of kaon identification in the TOP detector. Arrival time of the Cherenkov photons as a function of position is compared with the expectations for (left) a pion and (right) a kaon passing in the TOP. Reproduced from Ref. [60].

Using the likelihoods for the various mass hypotheses, Belle II algorithms construct a particle identification variable $\text{PID}_{\text{hyp}}^{\text{det}}$ for every detector,

$$\text{PID}_{\pi}^{\text{det}}(h) = \frac{\mathcal{L}_{\pi}^{\text{det}}(h)}{\mathcal{L}_{\pi}^{\text{det}}(h) + \mathcal{L}_K^{\text{det}}(h) + \mathcal{L}_e^{\text{det}}(h) + \mathcal{L}_{\mu}(h) + \mathcal{L}_p^{\text{det}}(h) + \mathcal{L}_d^{\text{det}}(h)}, \quad (2.3)$$

which is directly used in physics analyses. As an example, this was the $\text{PID}_{\text{hyp}}^{\text{det}}$ expression associated with the pion mass hypothesis¹, but the $\text{PID}_{\text{hyp}}^{\text{det}}$ values for other mass hypotheses are obtained by replacing the likelihood at the numerator $\mathcal{L}_{\pi}^{\text{det}}(h)$ with the corresponding hypothesis-specific value $\mathcal{L}_{\text{hyp}}^{\text{det}}(h)$. The quantity $\text{PID}_{\text{hyp}}^{\text{det}}(h)$ is defined similarly to a likelihood ratio $\mathcal{L}_0/\mathcal{L}_1$, which is the best-performing quantity to test two alternative simple statistical hypotheses [63]. The $\text{PID}_{\text{hyp}}^{\text{det}}$ variable assumes values from 0 to 1. The larger the $\text{PID}_{\text{hyp}}^{\text{det}}$, the higher the probability of observing the reconstructed track assuming true the chosen mass hypothesis.

Information from individual detectors is combined to improve the identification performance. The detector-specific likelihoods are combined together as a product,

$$\mathcal{L}_{\text{hyp}}(h) = \mathcal{L}_{\text{hyp}}^{\text{TOP}}(h) \mathcal{L}_{\text{hyp}}^{\text{CDC}}(h) \mathcal{L}_{\text{hyp}}^{\text{SVD}}(h) \mathcal{L}_{\text{hyp}}^{\text{ARICH}}(h) \mathcal{L}_{\text{hyp}}^{\text{ECL}}(h) \mathcal{L}_{\text{hyp}}^{\text{KLM}}(h), \quad (2.5)$$

¹In practice, the Belle II software expresses PID_{hyp} using only the natural logarithm of the likelihood values,

$$\text{PID}_{\text{hyp}} = \frac{e^{\ln \mathcal{L}_{\text{hyp}} - \ln \mathcal{L}_{\text{MAX}}}}{\sum_i (e^{\ln \mathcal{L}_i - \ln \mathcal{L}_{\text{MAX}}})} = \frac{e^{\Delta \ln \mathcal{L}_{\text{hyp}}}}{\sum_i (e^{\Delta \ln \mathcal{L}_i})} \quad (2.4)$$

where $\mathcal{L}_{\text{MAX}}$ is the largest of the likelihood values over the six hypotheses.

and the result is used in Eq.(2.6) to obtain the detector-combined PID. If a particle does not get reconstructed in a detector because, for instance, it escapes its acceptance, no PID information from that detector is available and the corresponding individual likelihood is set to one.

Of the two main PID detectors, TOP allows separating pions from kaons at $0.4 - 4 \text{ GeV}/c$ momenta with kaon identification efficiency of 85% and pion misidentification rate of about 10%, while the ARICH separates pions from kaons across all their momentum spectrum and discriminates also pions, electrons, and muons below $1 \text{ GeV}/c$ with 4σ separation or more.

Combining information from all detectors, the electron and muon identification efficiencies are respectively 86% and 88.5% after requiring the binary PID to be larger than 0.9, with pion misidentification rates of 0.4% and 7.3%, respectively. Binary PID is an additional PID variable that compares only two mass hypotheses, for example

$$\text{PID}_{\mu,\pi}^{\text{det}}(h) = \frac{\mathcal{L}_{\mu}^{\text{det}}(h)}{\mathcal{L}_{\mu}^{\text{det}}(h) + \mathcal{L}_{\pi}^{\text{det}}(h)}. \quad (2.6)$$

Data and simulation agree, except at low momenta where discrepancies within 20% are observed. Performance of kaon identification for a threshold of 0.8 on the kaon-pion binary PID is summarized in fig. 2.16. Efficiency varies from 95% to around 60%, depending on kaon momentum and polar angle. The pion misidentification varies from about 20% to less than 5%.

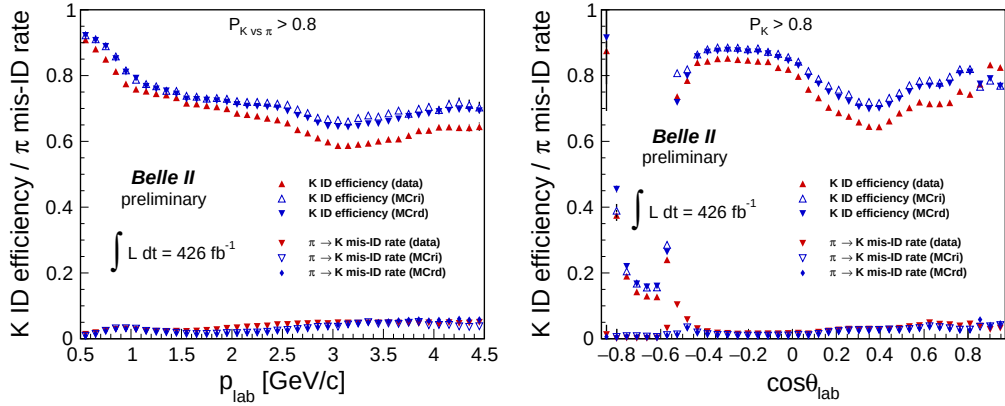


Figure 2.16: Kaon identification efficiencies and pion misidentification rates for events having a binary PID larger than 0.8, in data and simulation (MC) as functions of (left) kaon momentum, and (right) cosine of the polar angle in the laboratory frame. The “ri” and “rd” suffixes in the simulation labels indicate simulation that does not reproduce, or does reproduce, time-dependent variations in beam conditions.

2.3.3 Photon reconstruction

Photon reconstruction is particularly important in this search as the signal-extraction observable is constructed from all neutral electromagnetic-calorimeter energy deposits remaining after $B\bar{B}$ reconstruction. Photons are reconstructed from energy deposits in the ECL not geometrically associated with any outward CDC track extrapolation. When photons impinge on the ECL, they undergo electromagnetic interactions with the crystals thus depositing energy, which is commonly spread across multiple crystals and might overlap

with energy deposits from other nearby photons. In each collision that passes the trigger, the information of all 8736 crystals in the calorimeter is recorded for offline analysis. Raw crystal-level information is translated into photon candidates used in Belle II analyses through the photon reconstruction algorithms.

Adjacent sets of calorimeter crystals showing energy deposits are called clusters, and the process by which the energies and times observed in each of the crystals are converted into a set of clusters is called cluster reconstruction. The cluster-reconstruction algorithm, illustrated in fig. 2.17, starts by dividing the calorimeter active surface into connected regions, which are contiguous sets of crystals containing significant energy, isolated from all other connected regions. A connected region is assembled starting from a seed crystal,

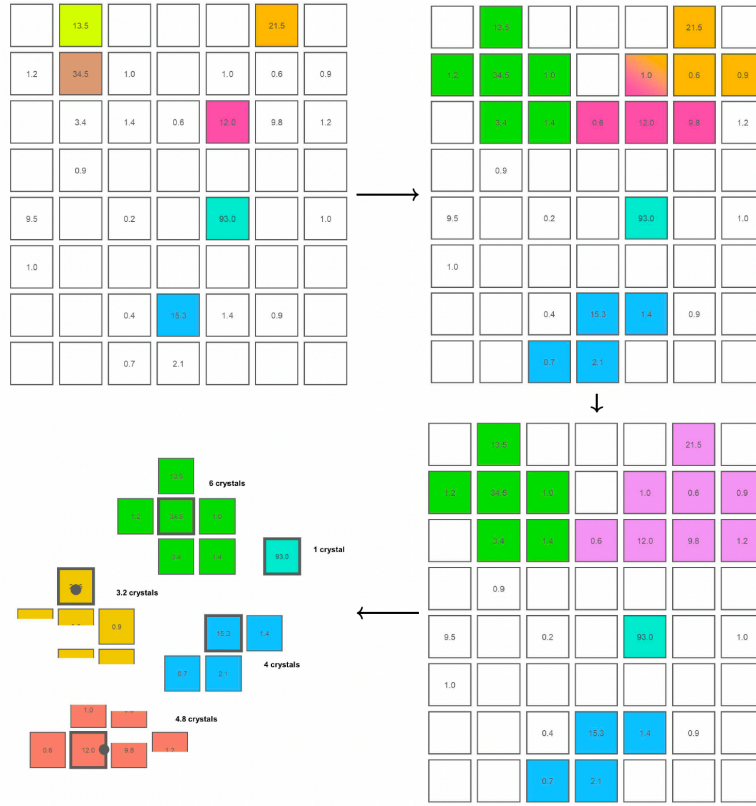


Figure 2.17: Scheme of the steps of cluster reconstruction. (Top left) Seed crystals are identified, (top right) neighboring crystals are attached if they have a large enough energy, (bottom right) connected regions are created, and (bottom left) finally if a connected region has more than one local maxima, it is split in more clusters.

whose energy must exceed 10 MeV. The eight immediate neighbors to the seed crystal are examined and included in the connected region if their energy exceeds 0.5 MeV. If any of these have energy greater than 10 MeV, their neighbors are also examined, and the process repeats. Each connected region is then divided into clusters, one per local maximum. A local maximum is a crystal whose energy is greater than the seed energy and is greater than that of its eight immediate neighbors. If more than one local maximum is found, the energy in each crystal is shared among the resulting clusters. Each cluster in the connected region is assigned an energy weight that depends on the energy of the cluster and the distance between the crystal and the local maximum crystal (centroid position). The properties

of the cluster, including the centroid position, are derived from the corresponding set of associated crystals. The procedure is iterative. The centroid position is initially taken to be the center of the crystal exhibiting a local maximum, and the cluster energy to be 50% more than the energy in the crystal showing the local maximum. The position and energy are updated after each iteration, and the process repeats until the centroid positions are stable. Cluster-shape quantities are calculated using all crystals in the cluster. The energy is the sum of energies in the N most energetic crystals, where N depends on the location in the detector, the photon energy, and the level of beam-induced backgrounds in the event, estimated using out-of-time events, that are events with times far from the trigger time.

Finally, clusters associated with a CDC track, or with energy less than 50 MeV and detection times farther away from the collision time than (nominally) 99% of collision photons of that energy, are discarded.

The energy of the surviving clusters is calibrated using simulated single-photon events due to various effects that can modify their true energy: energy leakage out of the outermost boundary of the calorimeter, energy deposited in inactive material between crystals or in the interface preceding the innermost boundary of the calorimeter, or energy deposited in crystals not included in the cluster-energy calculation. Finally, we associate to the cluster a reconstructed photon (or neutral hadron) candidate.

A particle reconstructed in simulation is considered “correctly reconstructed” if it is “matched” with the underlying generator-level particle. In the case of a photon, a cluster is correctly matched if a certain fraction of its reconstructed energy is attributable to the generated particle. Each cluster can have weighted relations with up to 21 crystals. Each excited crystal, in turn, can have a weighted relation with none, one, or multiple generated particles. The weight between a cluster and a generated particle is given by the product (weight between the corresponding cluster and crystal) $\times$ (weight between the crystal and the generated particle). The weight between the cluster and the crystal, as described above, is just the fraction of energy the crystal contributes to the cluster. The weight between the crystal and the generated particle is calculated using the total energy deposited by the generated particle in each crystal. If multiple relations exist between a given cluster and generated particle, only the relation with the largest weight is used for photon matching. Truth matching is set between the reconstructed cluster and the generated particle if the following conditions are met: (i) the generated particle is actually a photon, (ii) $\text{weight}/E_{\text{rec}} > 0.2 \text{ GeV}$, and (iii) $\text{weight}/E_{\text{true}} > 0.3 \text{ GeV}$, where E_{rec} is the reconstructed energy and E_{true} is the true energy in simulation. If the generated particle is not a photon, such as an electron, no match occurs. This applies even if one of the other lower-weighted relations for the particle is correct.

Once clusters are reconstructed, the distribution of the detected energy within the crystals in each provides information about the spatial distribution of the released energy, so-called “shower-shape”. This in turn offers information useful to statistically identify the phenomenon that generated the interaction among the various sources.

Shower shape provides discrimination among the energy deposition sources, as illustrated in fig. 2.18. The simplest shower shape originates from a single photon where the maximal energy is deposited in the center crystal of a shower symmetric around the crystal main axis. The whole shower is typically contained in an array of about 5×5 crystals, even for high photon energies.

While electron-induced clusters are intrinsically similar to photon clusters, their shower shape is often different due to additional bremsstrahlung photons emitted in the interaction with material surrounded by the ECL. These photons lead to less well-defined cluster shapes for electrons when compared to photons. In addition, electrons are bent in the magnetic

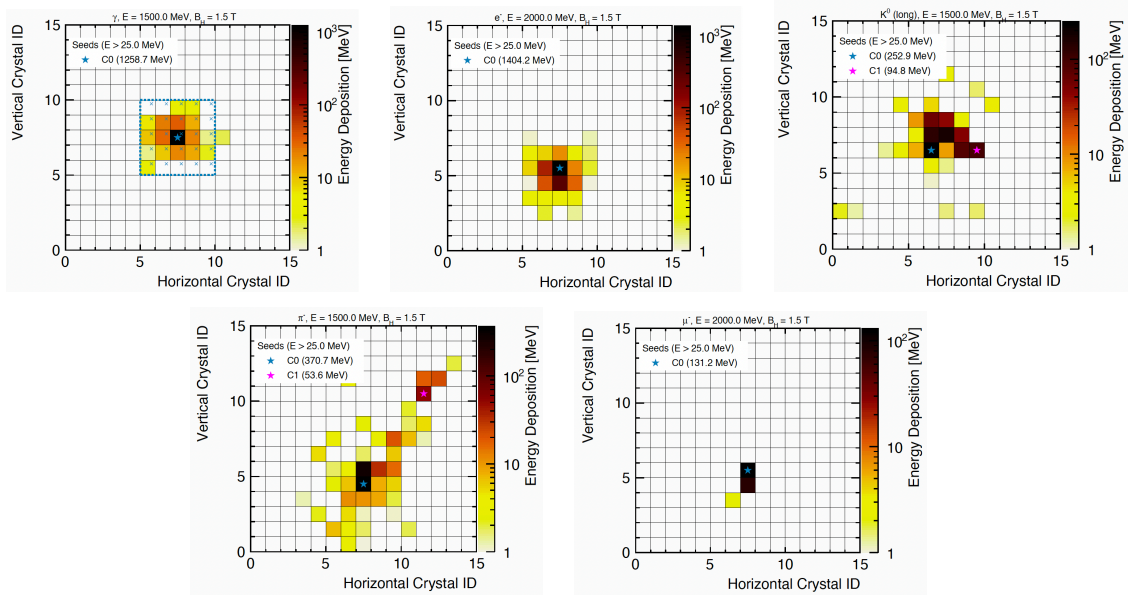


Figure 2.18: Simplified event displays of the energy deposition of simulated (top-left) photon, (top-middle) electron, (top-right) K_L^0 , (bottom-left) charged pion, and (bottom-right) muon.

field, which result in a different entry angle into the ECL and a deviation of the shower shape from radial symmetry shape even in the absence of additional radiation.

A fraction of neutral hadrons, such as neutrons and K_L^0 , undergo strong interactions in the ECL. The resulting shower shape is irregular and radially asymmetric.

If charged hadrons interact strongly within the ECL, they produce irregularly shaped showers in addition to a tilted entry angle due to the magnetic field.

Particles like muons and hadrons that do not interact strongly within the crystal are minimum-ionizing, depositing around 200 MeV almost solely in the crystals directly transversed by the particle. Additional radiated photons may lead to an overlapping minimal ionizing signal with electromagnetic showers.

The energy resolution for photons ranges from $\sigma_E/E = 7.7\%$ at 100 MeV to 2.2% at 1 GeV. The resolution on the reconstructed π^0 mass is $5.4 \text{ MeV}/c^2$.

Chapter 3

Experimental considerations

This chapter discusses experimental aspects relevant to searches for $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays and Belle II-specific analysis methods relevant for this analysis to facilitate the comprehension of subsequent technical descriptions. An overview of the analysis strategy is also given.

3.1 A $B^+ \rightarrow K^+ \tau^+ \tau^-$ decay in Belle II

Confined bunches of electrons and positrons are brought to collision in the interaction point (IP) at the rate of 200 million per second. Most of these collisions yield uninteresting electroweak events such as Bhabha scattering, photon-pair production, and others. In a small fraction of collisions though, a $\Upsilon(4S)$ meson is produced at threshold, and half of those immediately decay into a $B^+ B^-$ pair. The 1.6 ps B^+ meson lifetime, combined with a typical B momentum of 1.5 GeV/ c in the laboratory, results in a B flight length of about 130 μm . This implies that most B mesons decay before reaching the innermost tracking layer, but their decay products are sufficiently displaced from the IP that the silicon vertex measures nonzero impact parameters. One every approximately 10 million of them is expected to decay into a $K^+ \tau^+ \tau^-$ final state (fig. 3.1). The K^+ typically has 1.0 GeV/ c momentum and with its 12 ns lifetime it is considered stable in Belle II as it traverses the tracker and PID detectors that allow to identify it before decaying. The τ leptons have 1.3 GeV/ c momentum, typically, and with their 0.3 ps lifetime fly 65 μm before decaying into a muon or electron and a neutrino pair, 35% of times; or a charged pion, neutral pions, and a neutrino, 50% of times. Leptons and pions have typical momenta of 0.6 GeV/ c and with their large lifetime are also considered stable and traverse the tracker and the PID detectors. Electrons are absorbed in the calorimeter. Neutrinos are not reconstructed directly in Belle II and the vast majority of them traverse the detector undisturbed. This prevents from accessing the full kinematic information about the signal, which have significant consequences in the analysis strategy.

3.2 Analysis strategy and overview

This is a “blind” analysis; I develop the analysis using only simulated- and control-data samples without inspecting the signal-enriched region in data until all procedures are established and final. Such an approach reduces the chances of biasing the analysis choices towards the expected or desired result, reducing bias.

The $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays are rare, with a SM branching fraction of order 10^{-7} [64]. Hence, a large B -decay sample is a prerequisite to search for this decay. I use 387 million

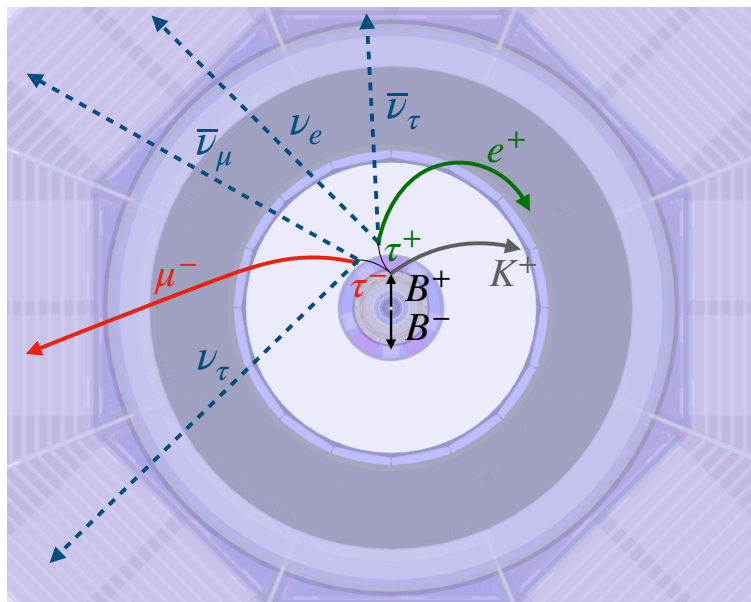


Figure 3.1: Sketch of a $B^+ \rightarrow K^+ \tau^+ \tau^-$ decay in the transverse view. Not to scale.

$B\bar{B}$ pairs. This is the third largest sample of this kind and per-se would contain 56 $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays, assuming the SM branching fraction and 100% signal efficiency.

A significant challenge is the contamination from background 10^9 times larger than the signal at production, which requires to devise highly discriminating rejection strategies. An equally challenging difficulty arises from the nature of $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays. The final state includes two τ leptons, each decaying into at least one neutrino, which are not detected at Belle II. This leads to significant unreconstructed energy and momentum that prevents the full reconstruction of the signal kinematic properties, thus depriving the analysis of any narrow, or otherwise distinctive, signal structure to suppress background. However, the Belle II experimental environment, which is uniquely apt for this search, partially compensates for this serious limitation. On-threshold $B\bar{B}$ pair production from collisions of point-like particles reconstructed with a hermetic detector provides stringent constraints on the kinematic properties of collision products, which help mitigating the lack of kinematic constraints associated with the signal decay itself. To this end, I restrict the sample to collisions in which the nonsignal B meson is fully reconstructed in a hadronic decay. By combining the resulting kinematic properties of the nonsignal B with those of the initial state, I infer statistically the τ -pair kinematic features and, consequently, the signal features while suppressing background too. The disadvantage is that fully reconstructable hadronic decays are relatively rare, with branching fractions of order 10^{-3} . This results in less than 1% efficiency, making the search for a rare decay even more difficult, as the expected SM signal yield gets reduced to about 0.6 events.

Absence of signal peaking-structures is also overcome by relying on inclusive event properties to extract signal. A key observable in this and similar analyses is the residual energy E_{ECL} detected in the calorimeter after reconstruction of the $B\bar{B}$ pair, which peaks at zero for signal events, allowing suppression of background. However, E_{ECL} usage is nontrivial. Susceptibility to the properties of all particles in the event makes it prone to data-simulation mismodeling. I conceive, develop, and execute an original and thorough study to validate the E_{ECL} data-simulation consistency, which has proven essential for

this and several other Belle II analyses. This consists in inspecting the basic, lower-level observables that contribute to E_{ECL} that is, photon energy and photon multiplicity and study their data-simulation consistency for various photon-candidate sources and selections in control samples. Based on the findings, I suppress misreconstructed photons as they show irrecoverable data-simulation inconsistencies, obtaining properly modeled E_{ECL} properties.

The sensitivity of the search also depends on the choice of τ decays. The τ lepton decays into electrons and muons 35% of the time, into a single charged pion with multiple neutral pions 50% of the time, and into multiple charged pions or other hadrons for the remaining cases [11]. Hence, before delving into analysis of collision data, I perform a thorough study to compare sensitivities between various combinations of τ final states using realistic replicas of the analysis in simulation. I classify the samples according to τ -pair final states $K^+\ell^+\ell^-$, $K^+\ell^+\pi^-$, and $K^+\pi^+\pi^-$. Each has a different background composition, requiring a dedicated analysis. The study, based on expected exclusion limits in realistic background-only simulated samples corrected to mirror data, shows that the $K^+\ell^+\ell^-$ sensitivity is 1.6 to 2.5 times better than the others. Hence I restrict the search to the $K^+\ell^+\ell^-$ final-state only. This further suppresses the expected SM signal yield in our sample down to 0.07 events. A similar study shows that treating separately the $K^+e^+e^-$, $K^+e^+\mu^-$, and $K^+\mu^+\mu^-$ final states does not provide any significant gain in sensitivity; hence I focus on lepton final states analyzed inclusively.

I introduce an innovative background-suppression criterion with respect to the only previous search of this channel. Selecting signal events in which the opposite-charge kaon-lepton mass exceeds the D -meson mass isolates a region with extremely low-background due to absence of prevailing and poorly modeled semileptonic $B^+ \rightarrow \bar{D}^{(*)0}\ell\bar{\nu}$ decays. This approach suppresses efficiently one of the most abundant backgrounds. In addition, it enables a simpler analysis based on sequential one-dimensional selections that achieve the same sensitivity to that based on a multivariate classifier with much less need of validations. I further improve sensitivity by optimizing the selections to provide the best average expected limit.

Signal-extraction involves counting the events observed in the E_{ECL} signal-region and subtracting the expected background yield as estimated from the extrapolation of yields in sidebands. I assess all significant sources of systematic uncertainties including the dominant contribution from potential discrepancies between data and simulation in the properties of the background populating the signal region, which is studied and validated in detail.

3.3 Current experimental status

In 2017, the BaBar experiment reported the first ever search for $B^+ \rightarrow K^+\tau^+\tau^-$ decays using its full dataset of 471 million $B\bar{B}$ pairs [65]. This is the only published experimental result available to date. One of the two B mesons was fully reconstructed in a hadronic final state, and the remaining tracks, calorimeter energy deposits, and missing energy were assigned to the candidate signal B meson. Only leptonic τ decays were considered and analyzed separately into ee , $e\mu$, and $\mu\mu$ final states. Multivariate classifiers were developed for each final state to suppress background from semileptonic decays and to extract the signal yield. A similar observable to the residual calorimeter energy was a key performer in the classifier. No signal was observed and an upper limit of $\mathcal{B}(B^+ \rightarrow K^+\tau^+\tau^-) < 2.25 \times 10^{-3}$ was set at the 90% confidence level. While the observed yields in the e^+e^- and $\mu^+\mu^-$ channels were consistent with expected backgrounds, an excess of 3.7σ over the background expectation was reported in the $e^\pm\mu^\mp$ channel.

3.4 Relevant analysis tools

In order to facilitate the understanding of the subsequent, more technical descriptions of the analysis, I introduce here conceptual discussions about some relevant analysis quantities.

3.4.1 B tagging

Near-threshold production of $B\bar{B}$ pairs, available in B -factory experiments, allows for using reconstruction of the nonsignal B -meson properties to suppress background and infer the kinematic and flavor properties of the signal B meson. This technique is known as “ B tagging” and offers a significant experimental advantage when reconstructing decays with final-state neutrinos. Approaches are broadly classified into two classes.

Exclusive B -tagging The nonsignal B_{tag} candidate is reconstructed in a set of several specific decays. Exclusive B -tagging is further divided into “semileptonic” and “hadronic” algorithms, depending on the choice of B_{tag} decays.

In semileptonic B tagging, the nonsignal B is reconstructed using final states comprising an electron or muon and one or multiple hadrons. This algorithm targets semileptonic decays, which account for about a quarter of the B meson width and are dominated by a few modes. The large branching fractions of the $B \rightarrow D^{(*)}\ell\nu$ decays typically used leads to a relatively high efficiency of up to approximately 2% [66]. However, the presence of neutrinos does not allow the full reconstruction of the B_{tag} four-momentum, resulting in weaker kinematic constraints on the signal.

In hadronic tagging, the nonsignal B is fully reconstructed using only hadrons. Although hadronic B decays account for three-quarters of the B meson width, the myriad of possible final states cause even the highest-rate decays to have only $\mathcal{O}(10^{-3})$ branching fraction. This leads to a hadronic B -tagging efficiency of less than 1%. Despite the low efficiency, hadronic B tagging provides effective background suppression and precise kinematic constraints, giving it a purity advantage over semileptonic B tagging.

Inclusive B -tagging Inclusive B tagging does not explicitly select and reconstruct any specific nonsignal B decay. It first isolates particles potentially coming from the signal candidate and assigns all remaining particles of the event to the B_{tag} candidate. Then statistical-learning classifiers are typically used to infer the presence of a B_{tag} , by exploiting distinctive topological and kinematic properties. The inclusive approach yields high signal efficiency of up to 4%, but it suffers from low purity. In addition, inclusive-tagging analyses are more prone to mismodelings associated with the inclusive character of the algorithm, which requires a more detailed control of sample composition and the description of all the particles in an event.

3.4.2 Lower-level observables

Track displacement Interactions between beam particles within the same bunch, or with residual gas, may result in interactions of the beam halo with the SuperKEKB or Belle II infrastructure, yielding intense showers of secondary particles that illuminate the detector. These are referred to as beam-induced backgrounds (beam-background in short). Track displacement from the interaction point is effective to suppress such backgrounds because beam-background tracks do not usually point back to the interaction point. The quantities typically used are the transverse (dr) and

longitudinal (dz) distances of the track from the interaction point, whose position is known within $350\ \mu\text{m}$ in the longitudinal direction, $15\ \mu\text{m}$ in the horizontal direction, and $0.2\ \mu\text{m}$ in the vertical one. In this analysis, I use the displacement of the tracks originating from both B mesons. Figure 3.2 shows an example of dr and dz distributions of a signal muon in simulation.

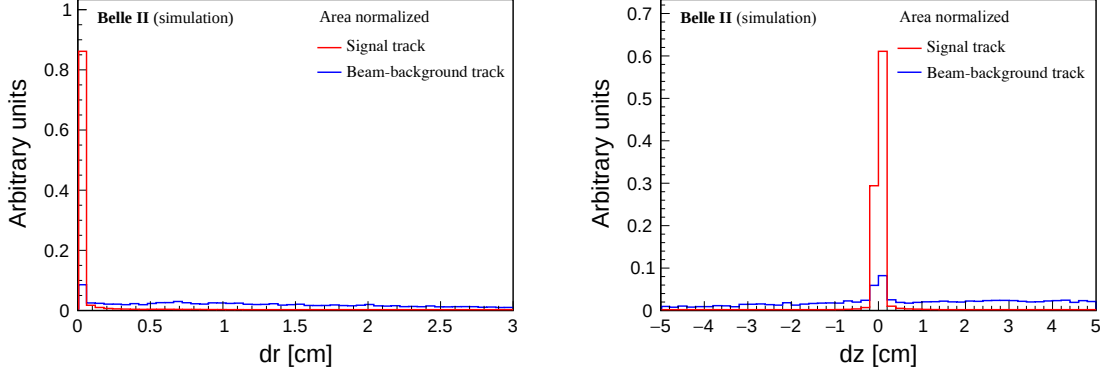


Figure 3.2: Distributions of (left) dr and (right) dz for (red) muons and (blue) beam-background tracks in simulated $B^+ \rightarrow K^+ \tau^+ \tau^-$ sample. Distributions are normalized to unity.

Hit multiplicity The number of samplings a charged particle undergoes in the tracking detectors (hits) is used to select on the quality of the tracks. Spurious tracks from accidental combinations of unrelated hits are typically suppressed by hit-multiplicity requirements.

Particle identification Several detectors provide track-specific statistical information on the identity of the corresponding charged particle (PID). This information expresses the probability of observing the detected PID signal assuming true a mass hypothesis among various possibilities and enhances discrimination against background, especially from prevailing pion components in continuum and misreconstructed B decays. If the comparison between observed and expected PID for a certain hypothesis is restricted between two particle types only, typically kaon and pion, it is referred to as binary particle identification. If the comparison is extended to multiple possibilities (kaon, pion, muon, electron, proton, and deuteron), it is referred to as global particle identification. The optimal usage choice depends on the specific composition of the targeted background. As an example, fig. 3.3 shows the distribution of the global electron PID for electrons in simulated $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays.

Lepton momentum in CM frame (p_{t+}^*) The signal lepton with same electric charge as the signal kaon has a distinctive momentum spectrum in the center-of-mass frame (fig. 3.3) because primary leptons from B -meson decays have higher momentum than leptons from the signal τ decay, due to the high B -meson mass. This makes p_{t+}^* an important discriminating observable to reduce semileptonic B to D decay backgrounds, such as $B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^- \pi^0) \ell^+ \nu$, which preferentially yield high p_{t+}^* .

Photon energy Photons originating from collision events typically have higher energies than beam-background photons, which are originated from the interaction of the

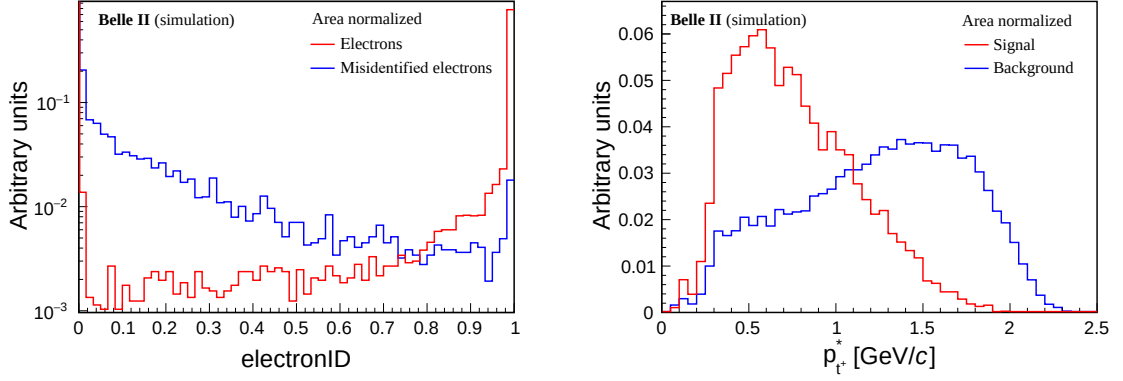


Figure 3.3: Distributions of (left) global electron PID and (right) p_{t+}^* for electrons in a simulated $B^+ \rightarrow K^+ \tau^+ \tau^-$ sample. Distributions are normalized to unity.

beam with the detector material (fig. 3.4). Hence photon-energy restrictions reduce beam background photons.

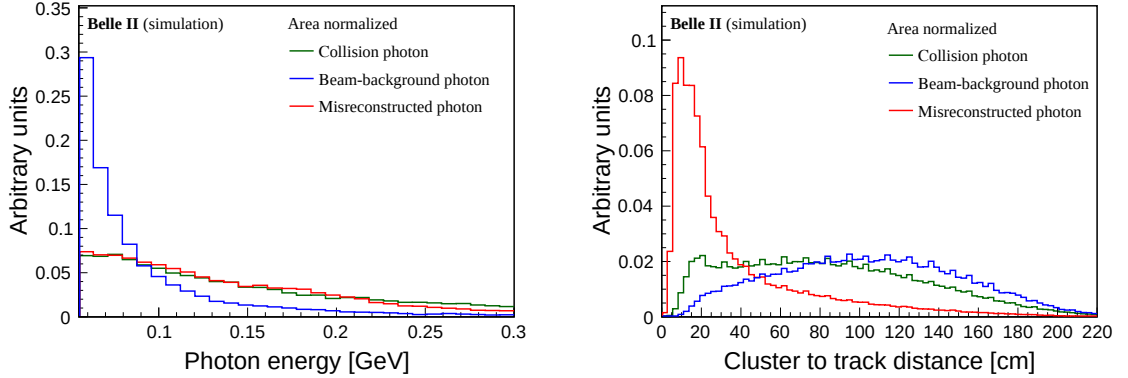


Figure 3.4: Distributions of (left) photon energy and (right) cluster-to-track distance in a simulated $B^+ \rightarrow K^+ \tau^+ \tau^-$ sample. Distributions are normalized to unity.

Cluster-to-track distance Charged hadrons may undergo hadronic interactions in the calorimeter and deposit energy. To distinguish them from photons, the distance along the calorimeter inner surface between the center of the highest-energy crystal in the cluster and the intersection of any track extrapolated outward from the drift chamber provides discrimination between genuine and misreconstructed photons (fig. 3.4).

Cluster-event time difference The observed energy-deposition time in the calorimeter relative to the collision time informs whether the source of the deposition is likely to be originated from the collision or from the interaction of the beam with the detector material (fig. 3.5).

3.4.3 B meson observables

Using distinctive kinematic information of signal candidates is a common approach to suppress background in many experimental environments. A widely used and effective

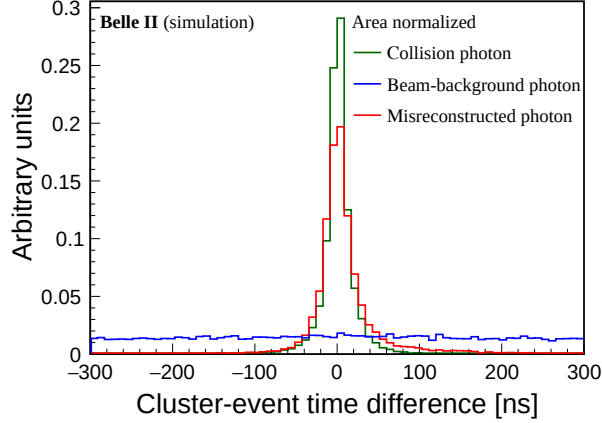


Figure 3.5: Distributions of cluster-event time difference in a simulated $B^+ \rightarrow K^+ \tau^+ \tau^-$ sample. Distributions are normalized to unity.

discriminator is the invariant mass, since fully reconstructed signal events cluster at a specific mass value and background shows typically broader distributions.

The peculiar kinematic environment of e^+e^- colliders at the $\Upsilon(4S)$ provides additional constraints that further background separation. The $\Upsilon(4S)$ is produced almost at threshold and decays in two same-mass particles, B and $\bar{B}$. If the B meson is correctly reconstructed, the energy of each of its decay products equals half of the collision energy in the center-of-mass frame. This is optimally exploited by two observables (all quantities in the $\Upsilon(4S)$ frame).

Beam-constrained mass

$$M_{bc} = \sqrt{s/4 - |\vec{p}_B^*|^2}, \quad (3.1)$$

where $\vec{p}_B^*$ is the momentum of the B meson reconstructed from the momenta of its decay products, and s is the squared collision energy. The observable is a distinctive reparametrization of the B -meson momentum that has a narrower signal distribution than the invariant B mass since the beam-energy spread is smaller than the uncertainty on the reconstructed B -meson energy. All B decays fully reconstructed using their decay products peak at the B -meson mass regardless if the products are correctly or incorrectly identified. Non- B events and partially reconstructed B decays have smooth distributions. Distributions of all components vanish at the kinematic limit of half the collision energy. The observable M_{bc} is effective in separating B events from continuum due to light-quark production.

Energy difference

$$\Delta E = E_B^* - \sqrt{s}/2, \quad (3.2)$$

is the difference between the reconstructed B -candidate energy and half of the collision energy, which is known with high precision. If the B meson is correctly reconstructed, the energy of the decay products equals approximately half of the collision energy. Therefore, B signals peak at zero, while continuum background follows a smooth distribution. In addition to discriminating against continuum, ΔE allows distinguishing background from misidentified B decays. If a B final-state particle is misidentified as another, its reconstructed energy, and consequently that of the B candidate, departs from its true energy because of the mismatch in mass, resulting in a ΔE shift.

Figure 3.6 shows an illustrative sketch of the ΔE and M_{bc} distributions for a generic correctly reconstructed exclusive B decay, continuum background events, and candidates reconstructed from other $B\bar{B}$ events. These observables are not relevant for the $B^+ \rightarrow K^+ \tau^+ \tau^-$ signal, but they are used in the reconstruction and selection of the nonsignal B in the analysis.

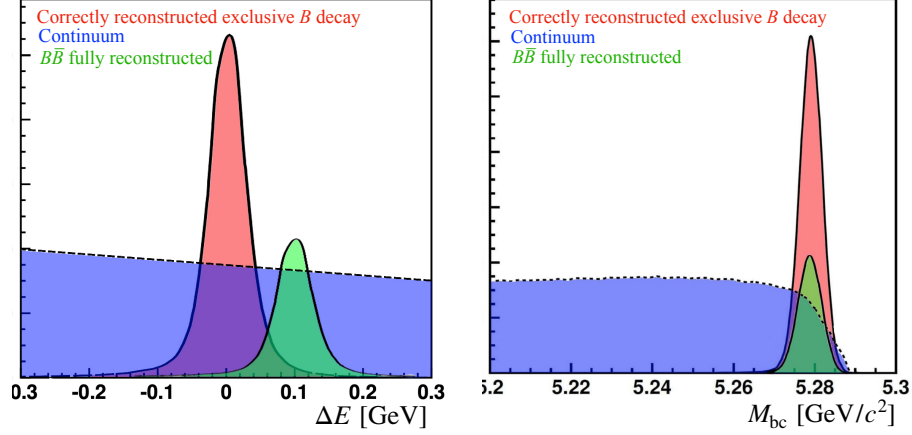


Figure 3.6: Sketched distributions of (left) ΔE and (right) M_{bc} for a correctly reconstructed exclusive B decay, continuum, and fully reconstructed candidates in $B\bar{B}$ events.

In addition to these standard observables, other analysis-specific B observables are employed in this analysis.

Lepton pair mass ($m(\ell^+ \ell^-)$) The invariant mass of the oppositely charged lepton pair in signal events suppresses backgrounds from photon conversions ($\gamma \rightarrow e^+ e^-$) and charmonium ($c\bar{c}$) resonances such as $J/\psi \rightarrow \ell^+ \ell^-$ decays.

Beam-constrained τ -pair mass squared (q^2) This is the square of the $\tau^+ \tau^-$ pair mass estimated from beam, B_{tag} , and signal kaon kinematic properties as $q^2 \equiv (p_{\text{beam}} - p_{B_{\text{tag}}} - p_K)^2$ where p are four-momenta. This observable has mostly phenomenological relevance. The theoretical model for the decay predicts the branching fraction within specific boundaries of q^2 . To ensure a consistent comparison between experimental results and theoretical predictions, it is important to align them within consistent q^2 boundaries.

3.4.4 Global event observables

Hadronic $e^+ e^-$ cross-sections are dominated by continuum background, consisting in the production of light $q\bar{q}$ pairs that mostly yield pions and kaons. The kinematic features associated with at-threshold $B\bar{B}$ production render observables capable to capture the “shape” of the event, that is, the spatial and phase-space distributions of final-state particles, powerful discriminators of $B\bar{B}$ events from continuum.

Figure 3.7 shows an illustrative sketch of the event shapes of a $B\bar{B}$ and a continuum event. In a $B\bar{B}$ event, both B mesons are nearly at rest in the $\Upsilon(4S)$ frame. The B decay products are therefore emitted isotropically in that frame, unlike light quarks, which are produced with a comparatively large initial momentum due to their small mass compared to the collision energy. This results in their fragmentation to develop along two collimated

back-to-back jets of light hadrons. Hence, the spatial and energy-momentum distributions of $B\bar{B}$ decay products are approximately spherical, compared to pencil-like shapes typical of continuum.

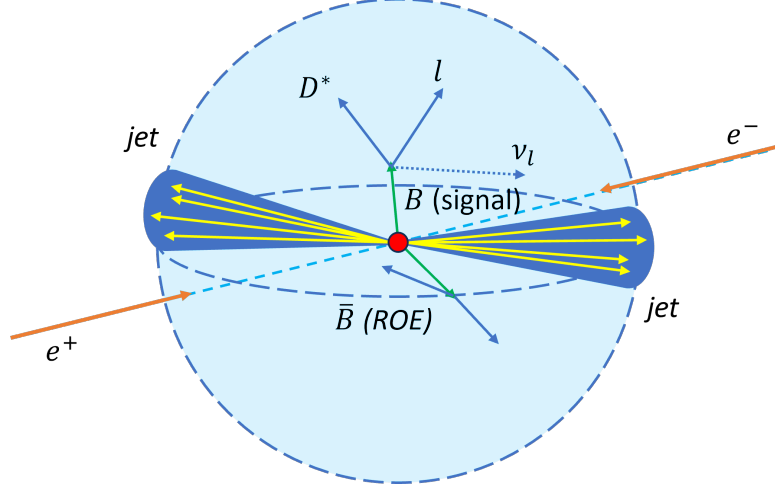


Figure 3.7: Event-shape sketch for continuum and $B\bar{B}$ events in the center-of-mass frame.

Information based on these distributions is exploited through several observables. These observables are used to suppress continuum background in nonsignal B selections of this $B^+ \rightarrow K^+ \tau^+ \tau^-$ search.

Sphericity. Sphericity quantifies the ordinary spatial distribution of all final-state particles. For a collection of three momenta $\mathbf{p}_i$, the sphericity tensor S is

$$S^{\alpha,\beta} = \frac{\sum_{i=1}^N p_i^\alpha p_i^\beta}{\sum_{i=1}^N |\mathbf{p}_i|^2}, \quad (3.3)$$

(with $\alpha, \beta = x, y, z$) and provides a three-dimensional representation of the spatial distribution of the $\mathbf{p}_i$ collection. For an isotropic distribution, the three eigenvalues λ_k have similar magnitude; for a directional distribution, the eigenvector oriented in the preferred direction has an eigenvalue considerably larger than the two others. An useful derived quantity is the sphericity (or sphericity scalar) $S = \frac{3}{2}(\lambda_2 + \lambda_3)$, where λ_2 and λ_3 are the two lowest eigenvalues. Figure 3.8 illustrates a comparison of sphericity between $B\bar{B}$ and continuum events. In $B\bar{B}$ events, sphericity is close to 1.0, corresponding to isotropically distributed momenta, while collimated distributions, as in continuum events, yield sphericity close to 0.0.

Thrust angle. Thrust captures the extent to which particle momenta in an event align along a principal axis. For N three-momenta in an event $\mathbf{p}_i$ ($i = 1, \dots, N$), thrust T is defined as

$$T = \frac{\sum_{i=1}^N |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_{i=1}^N |\mathbf{p}_i|}, \quad (3.4)$$

where $\mathbf{T}$ is the unit vector that maximizes the total momentum projection, and therefore, the longitudinal projection of the particle momenta [67]. The magnitude of the cosine of the angle between the thrust axis calculated using only particles from the nonsignal B ($\mathbf{T}_{\text{tag}}$) and the thrust axis based on all particles remaining

from the nonsignal B reconstruction ($\mathbf{T}_{\text{ROE}}$), $\cos\text{TBTO}$, offers the most powerful discrimination between $B\bar{B}$ and continuum events (fig. 3.8). Since the momenta of the B and $\bar{B}$ decay products are isotropically distributed, $\mathbf{T}_{\text{tag}}$ and $\mathbf{T}_{\text{ROE}}$ are randomly distributed, leading to uniform $\cos\text{TBTO}$ distributions in $B\bar{B}$. For $q\bar{q}$ events, particle momenta are collimated, resulting in strongly directional $\mathbf{T}_{\text{tag}}$ and $\mathbf{T}_{\text{ROE}}$, which yield peaking $\cos\text{TBTO}$ distributions.

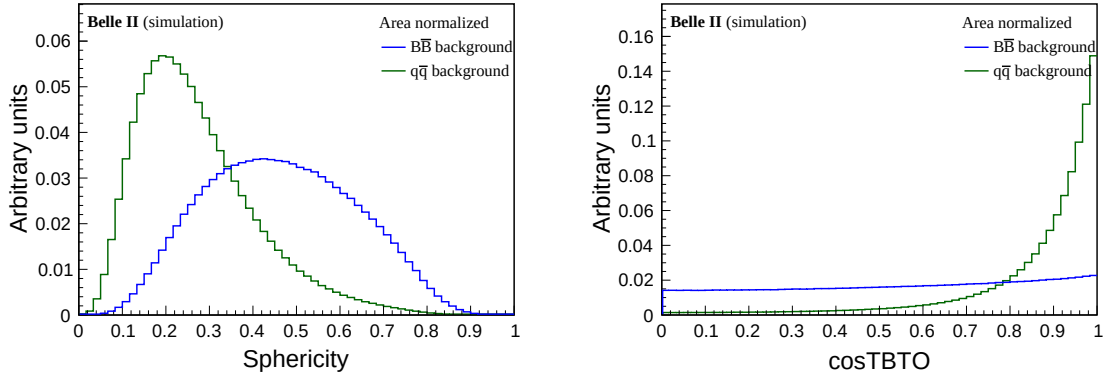


Figure 3.8: Distributions of (left) sphericity and (right) $\cos\text{TBTO}$ for simulated $B\bar{B}$ and continuum events. Distributions are normalized to unity.

In this analysis, sphericity and thrust are used to suppress the light-quark background in the selection of the nonsignal B_{tag} meson.

Residual calorimeter energy The residual energy E_{ECL} detected in the calorimeter after reconstruction of tag and signal B mesons is the principal signal-extraction observable in this work. In near-threshold production, a proper reconstruction of the $B\bar{B}$ pair would leave no residual energy in a perfectly hermetic calorimeter, implying that E_{ECL} peaks at zero for signal events. The background has instead a broader, smooth distribution due to misreconstruction or misassociation of particles that contribute extra energy. Figure 3.9 shows the E_{ECL} distribution in simulated signal and background events. A prominent signal peak is observed at zero smeared by resolution into a significant high- E_{ECL} tail. The background has a broad distribution peaking at about 0.3 GeV and extended to even larger values. The small secondary peak at $E_{\text{ECL}} \approx 0.0$ in the background distribution is due to backgrounds with same final states as signal, which are not discriminable with this observable. One significant drawback of inclusive event observables such as E_{ECL} is their susceptibility to mismatches between data and simulation. Proper modeling requires an accurate description of all phenomena intervening in a collision, including collision products, beam backgrounds, etc., and correlations between them. Therefore, thorough validation of E_{ECL} is essential and detailed in Chapter 5.

Missing mass squared (M_{miss}^2) This is the invariant mass of all undetected particles in the event. It is calculated by subtracting the sum of the four-momenta of all reconstructed charged particles and photons from the four-momentum of the beam. This observable serves as a powerful signal discriminant, as the signal and background exhibit different distributions due to differing final-state neutrino multiplicities (fig. 3.9).

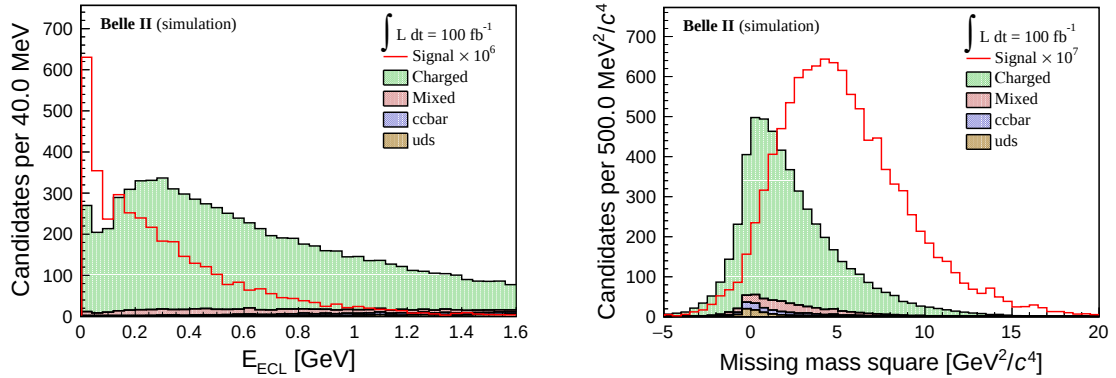


Figure 3.9: Distributions of (left) E_{ECL} and (right) M_{miss}^2 for simulated signal and various classes of background events. Labels ‘charged’, ‘mixed’, ‘ccbar’, and ‘uds’ indicate simulated B^+B^- ; $B^0\bar{B}^0$; $c\bar{c}$; and $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds, respectively.

Residual π^0 mass This is the invariant mass of photon pairs reconstructed from energy deposits that are not associated with the $B\bar{B}$ pair reconstruction. This observable helps to suppress background from misreconstructed π^0 , which is particularly insidious for its capability to spoil the E_{ECL} distributions.

Chapter 4

Signal reconstruction and baseline selection

This chapter describes the samples used, along with signal reconstruction and baseline selection.

4.1 Online event selection

Several processes in e^+e^- collisions at $\sqrt{s} = 10.58$ GeV occur with much higher rates than $B\bar{B}$ pair production (table 2.5). Events from QED such as Bhabha scattering, lepton or photon pair production, and beam-gas interactions in which a single electron or positron interacts with residual gas molecules, need to be suppressed to restrict the sample to hadronic events, which are those of interest in this work. As discussed in section 2.2.4, the trigger operates the first such filtering in real-time, using promptly available track information from the drift chamber and cluster data from the calorimeter. The details of trigger-selection criteria evolve throughout the data-taking period to cope with the changing environmental conditions, especially in terms of beam-induced backgrounds. However, the most typical selections implemented at the first trigger level for identifying B physics events are as follows

- Presence of three or more two-dimensional (transverse plane) tracks that have four hits in each of five axial super-layers in the drift chamber. Beam-injection veto active.
- Presence of three or more two-dimensional tracks that have four hits in each of five axial super-layers and one or more three-dimensional tracks with $|dz| < 20$ cm in the drift chamber. Beam-injection veto active.
- Presence of two or more two-dimensional tracks that have four hits in each of five axial super-layers and one or more three-dimensional tracks with $|dz| < 20$ cm in the drift chamber. Track opening angle greater than 90° . Beam-injection and Bhabha vetos active.
- Total cluster energy greater than 1 GeV. Beam-injection and Bhabha vetos active.
- Presence of three or more isolated clusters. Beam-injection and Bhabha vetos active.

The beam injection-veto rejects events during particle injection into the accelerator, as the beam is unstable at this stage, leading to high beam-induced backgrounds. The

Bhabha veto is activated when two clusters in the calorimeter have energies greater than 3 GeV, with at least one exceeding 4.5 GeV; the sum of the polar angles is between 165° and 190° , and the difference in azimuthal angles is within 160° to 200° in the center-of-mass frame.

The efficiency of this trigger selection and the close variants used to collect our sample is known to be in excess of 99% for our signal decay as well as for most B decays.

4.2 Samples

I use simulated and experimental data. Simulated data are used to design the analysis, optimize the event selection, estimate the efficiency for reconstructing and selecting signal, study the analysis sensitivity, and aid in the estimation of background in the signal-search region. Experimental data are those from which final results are extracted and validations performed.

4.2.1 Experimental data

The Belle II experiment started physics operations in March 2019, accumulating electron-positron collision data corresponding to an integrated luminosity of $427.9 \pm 2.0 \text{ fb}^{-1}$. I use the full data set collected at the energy of the $\Upsilon(4S)$ resonance up to late 2023, corresponding to $365.3 \pm 1.7 \text{ fb}^{-1}$, to reconstruct the signal and control modes. The analysis does not use data collected at energies other than the $\Upsilon(4S)$ resonance.

4.2.2 Simulated data

Simulated samples are based on the Monte Carlo approach. Monte Carlo samples are produced using event generators, which are computer programs that use pseudorandom number generators to produce sets of four-vectors reproducing final states of e^+e^- collisions according to theoretical models of particle kinematic properties and interactions. Generated data are then subjected to detector simulation, where models of the detector geometry and material are interfaced with models of interactions of particles with matter and signal formation to reproduce the expected values of the raw quantities observed in the detector. These are then subjected to processing and event reconstruction as if they were collision data. The resulting simulated data contain information about reconstructed particles and about the generated *true* particles. By matching these sets of information, we understand whether particles are reconstructed properly, or what are the most frequent misreconstruction occurrences, and what are principal backgrounds. This “truth-matching” procedure is useful to optimize selection requirements, classify sample components, and for many consistency checks.

Figure 4.1 shows a sketch of the generation sequence for a hadronic event in Belle II. The properties of virtual Z and photons created in the electron-positron annihilation, and their subsequent splitting into a quark-antiquark pair, which in turn produces the observed hadrons, are simulated by `Pythia8` [68]. The decay of the heavy hadron (top right corner in fig. 4.1) is simulated according to a `EvtGen` model for known decays [69] and using `Pythia8` for unmeasured decays. The photon emission by final-state charged particles is simulated by `PHOTOS` [70].

For signal studies, I simulate inclusive samples of $e^+e^- \rightarrow B^+B^-$ events, in which one of the B mesons is forced into decay to the $K\tau^+\tau^-$ final state, followed by all allowed τ decays, and the other B meson can decay in any allowed final state. The `EvtGen` program

$p_T > 200 \text{ MeV}/c$ and impact parameters $|dz| < 4 \text{ cm}$ and $|dr| < 2 \text{ cm}$ to suppress beam-background-related tracks. To be accepted into the `hlt_hadron` skim, an event should meet the following conditions:

1. Presence of three or more good tracks originated from a primary vertex, to suppress Bhabha, $e^+e^- \rightarrow \mu^+\mu^-$, two-photon, and low multiplicity beam-gas events (all processes that produce two or fewer tracks).
2. Inconsistency with the HLT selection for Bhabha events, which consists of requirements on the number of tracks identified as electrons or positrons, their momenta, the angle between them, and the total energy deposit in the ECL.

After applying these criteria, more than 99% (85%) of all $B\bar{B}(q\bar{q})$ events are retained, with a rejection of more than 99% of non-hadronic events as shown in Table 4.1.

	$B\bar{B}$	$c\bar{c}$	$q\bar{q}$	$\tau^+\tau^-$	$\mu^+\mu^-$	e^+e^-
ϵ	0.995	0.963	0.798	0.273	0.0096	0.0014

Table 4.1: Efficiencies for various processes selected through the hadronic stream selection.

4.4 Offline selection and reconstruction

In a search for a rare decay, event selection is a chief performance driver. In this analysis I first reconstruct the tag B meson (B_{tag}) through the standard Belle II hadronic tagging algorithm. Subsequently, I reconstruct the signal B meson (B_{sig}) from residual tracks not associated with the B_{tag} reconstruction. Then I sort the B_{tag} candidate chosen for further analysis among multiple candidates. Finally, the $\Upsilon(4S)$ is reconstructed by combining B_{tag} and B_{sig} mesons and the chosen B_{sig} candidate is sorted among multiple options, if any. As we reconstruct two B mesons, the selection of both is relevant. I adopt the B_{tag} selection optimized for B_{tag} efficiency calibration. While further margins of optimization could be available by specializing the B_{tag} selection to the specific signal I am seeking, I opt not to explore them to keep the B_{tag} treatment in this analysis consistent with the large body of validations and calibrations centrally available in Belle II. On the other hand, I perform a thorough and systematic study of the selections of the signal B to enhance the search sensitivity.

4.4.1 Introduction to full event interpretation

Typically hadronic tag algorithms reconstruct final-state particles from tracks and clusters, combine them to form the intermediate decay particles, and combine these in turn to form the B meson. Signal-enhancing information is used to increase the purity of each tagging B candidate. For this analysis I use an automated-boosted-decision tree algorithm known as “full event interpretation”, FEI. FEI reconstructs B_{tag} candidates by testing several of the most frequent hadronic B decay modes (36 decays in the case of B^+). Each mode is then further classified based on the relevant intermediate decays that are successfully reconstructed (tables 4.2 and 4.3). Their combinations correspond to over $\mathcal{O}(10^3)$ exclusive decay channels. The list of tested decays is the result of a compromise. On the one hand, it is important that the included decay modes have either branching fractions large enough to cover a large fraction of the width or that they yield sufficiently pure signal. On the other

hand, combinatorial complexities in reconstruction and processing time impose constraints on the number of tested possibilities.

Many decays in table 4.3 share the same final state so that their reconstructions are not mutually exclusive. For example, the $B^+ \rightarrow \bar{D}^{*0} (\rightarrow \bar{D}^0 \pi^0) \pi^+$ decay can form correctly reconstructed $B^+ \rightarrow \bar{D}^{*0} (\rightarrow \bar{D}^0 \pi^0) \pi^+$ decays, or $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0$ decays without the intermediate $\bar{D}^{*0}$. FEI is set to favor the modes with more intermediate resonances because they carry richer tagging information.

The FEI algorithm builds B_{tag} candidates from final states backward starting from tracks and clusters as shown in fig. 4.2 and applies kinematic fits at each successive reconstruction stage. At each stage, a probability ($\mathcal{P}_{\text{FEI}}$) is associated to each reconstructed particle. The probability acts as a quality metric of the reconstruction and of its purity. This probability, along with other information on tracks, clusters, intermediate particles is used to form heavier particles and ultimately the B_{tag} candidate. The B_{tag} signal probability is a powerful observable to suppress misreconstruction. Figure 4.3 illustrates a typical $\mathcal{P}_{\text{FEI}}$ distribution. Signal peaks at $\mathcal{P}_{\text{FEI}} \approx 1.0$ while backgrounds tend to reach lower values down to $\mathcal{P}_{\text{FEI}} = 0.0$.

As FEI produces multiple candidates of various modes in the same event and associates higher $\mathcal{P}_{\text{FEI}}$ to purer decays, choosing the candidate with the highest $\mathcal{P}_{\text{FEI}}$ in the event is equivalent to choosing the purest candidate in that event.

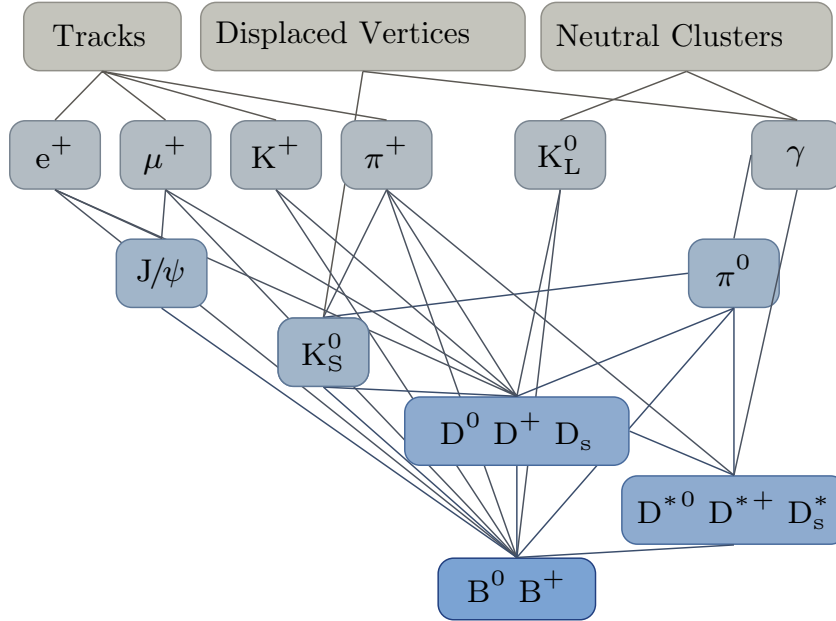


Figure 4.2: Hierarchical reconstruction of B candidates in the FEI algorithm, reproduced from Ref. [66].

4.4.2 FEI preselection

FEI forms final state particles ($e^-, \mu^-, K^-, \pi^-, \pi^0, p^-, K_L^0, K_S^0, \gamma$) from the tracks, ECL, and KLM clusters before feeding information into dedicated multivariate classifiers. The background is slightly reduced through fiducial preselections on the transverse and longitudinal distances from the interaction point to the track, which reject beam-induced

Decay mode		
0.	$D^0 \rightarrow K^- \pi^+$	$D^{*0} \rightarrow D^0 \pi^0$
1.	$D^0 \rightarrow K^- \pi^+ \pi^0$	$D^{*0} \rightarrow D^0 \gamma$
2.	$D^0 \rightarrow K^- \pi^+ \pi^0 \pi^0$	
3.	$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$	
4.	$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^- \pi^0$	
5.	$D^0 \rightarrow \pi^- \pi^+$	
6.	$D^0 \rightarrow \pi^- \pi^+ \pi^+ \pi^-$	
7.	$D^0 \rightarrow \pi^- \pi^+ \pi^0$	
8.	$D^0 \rightarrow \pi^- \pi^+ \pi^0 \pi^0$	
9.	$D^0 \rightarrow K_S^0 \pi^0$	
10.	$D^0 \rightarrow K_S^0 \pi^+ \pi^-$	
11.	$D^0 \rightarrow K_S^0 \pi^+ \pi^- \pi^0$	
12.	$D^0 \rightarrow K^- K^+$	
13.	$D^0 \rightarrow K^- K^+ \pi^0$	
14.	$D^0 \rightarrow K^- K^+ K_S^0$	
15.	$D^+ \rightarrow K^- \pi^+ \pi^+$	
16.	$D^+ \rightarrow K^- \pi^+ \pi^+ \pi^0$	
17.	$D^+ \rightarrow K^- K^+ \pi^+$	
18.	$D^+ \rightarrow K^- K^+ \pi^+ \pi^0$	
19.	$D^+ \rightarrow \pi^+ \pi^0$	
20.	$D^+ \rightarrow \pi^+ \pi^+ \pi^-$	
21.	$D^+ \rightarrow \pi^+ \pi^+ \pi^- \pi^0$	
22.	$D^+ \rightarrow K_S^0 \pi^+$	
23.	$D^+ \rightarrow K_S^0 \pi^+ \pi^0$	
24.	$D^+ \rightarrow K_S^0 \pi^+ \pi^+ \pi^-$	
25.	$D^+ \rightarrow K^+ K_S^0 K_S^0$	

 Table 4.2: D meson decay modes used in hadronic FEL.

Decay mode	
0.	$B^+ \rightarrow \bar{D}^0 \pi^+$
1.	$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0$
2.	$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0 \pi^0$
3.	$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^-$
4.	$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^- \pi^0$
5.	$B^+ \rightarrow \bar{D}^0 D^+$
6.	$B^+ \rightarrow \bar{D}^0 D^+ K_S^0$
7.	$B^+ \rightarrow \bar{D}^{*0} D^+ K_S^0$
8.	$B^+ \rightarrow \bar{D}^0 D^{*+} K_S^0$
9.	$B^+ \rightarrow \bar{D}^{*0} D^{*+} K_S^0$
10.	$B^+ \rightarrow \bar{D}^0 D^0 K^+$
11.	$B^+ \rightarrow \bar{D}^{*0} D^0 K^+$
12.	$B^+ \rightarrow \bar{D}^0 D^{*0} K^+$
13.	$B^+ \rightarrow \bar{D}^{*0} D^{*0} K^+$
14.	$B^+ \rightarrow D_s^+ \bar{D}^0$
15.	$B^+ \rightarrow \bar{D}^{*0} \pi^+$
16.	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0$
17.	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0 \pi^0$
18.	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^-$
19.	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^- \pi^0$
20.	$B^+ \rightarrow D_s^+ \bar{D}^0$
21.	$B^+ \rightarrow D_s^+ \bar{D}^{*0}$
22.	$B^+ \rightarrow \bar{D}^0 K^+$
23.	$B^+ \rightarrow D^- \pi^+ \pi^+$
24.	$B^+ \rightarrow D^- \pi^+ \pi^+ \pi^0$
25.	$B^+ \rightarrow J/\psi K^+$
26.	$B^+ \rightarrow J/\psi K^+ \pi^+ \pi^-$
27.	$B^+ \rightarrow J/\psi K^+ \pi^0$
28.	$B^+ \rightarrow J/\psi K_S^0 \pi^+$
29.	$B^+ \rightarrow \bar{\Lambda}_c p \pi^+ \pi^0$
30.	$B^+ \rightarrow \bar{\Lambda}_c p \pi^+ \pi^- \pi^+$
31.	$B^+ \rightarrow \bar{D}^0 p \bar{p} \pi^+$
32.	$B^+ \rightarrow \bar{D}^{*0} p \bar{p} \pi^+$
33.	$B^+ \rightarrow D^+ p \bar{p} \pi^+ \pi^-$
34.	$B^+ \rightarrow D^{*+} p \bar{p} \pi^+ \pi^-$
35.	$B^+ \rightarrow \bar{\Lambda}_c p \pi^+$

 Table 4.3: B^+ meson decay modes used in hadronic FEL.

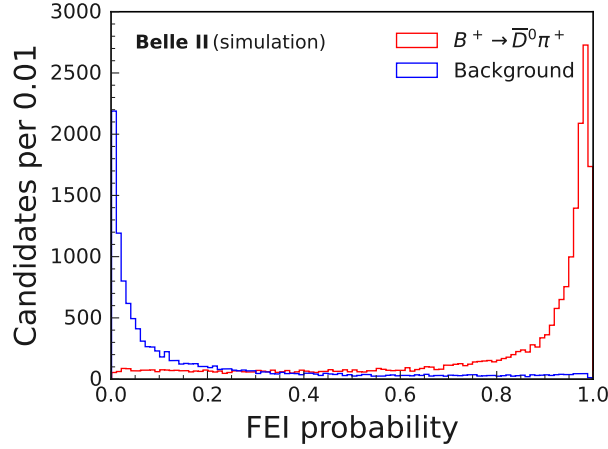


Figure 4.3: Distributions of FEI probability for (red) properly reconstructed and (blue) misreconstructed tag $B^+ \rightarrow \bar{D}^0 \pi^+$ candidates in simulated $B\bar{B}$ events, where signal side is reconstructed in $B^+ \rightarrow \bar{D}^0 \pi^+$ decays. Distributions are normalized to unity.

backgrounds and do not affect signal efficiency. Different photon energy thresholds are used in different polar regions of the calorimeter, as mentioned in table 4.4, to adjust them to polar-angle dependent backgrounds. More stringent requirements are used in the endcap regions compared to the barrel due to the forward boost of the collision. The resulting charged particles and photon candidates are ranked by particle-identification probability and energy, respectively. The number of candidates per event is restricted to 10 (e^- or μ^-), 20 (K^- , π^- , or p), or 40 (photons) to reduce the processing complexity associated with the training of the classifier.

For each simulated final-state particle type, individual BDT classifiers are trained targeting whether the reconstructed particle matches the generated one or not. BDTs targeting charged particles are trained using as discriminating observables impact parameters, PID likelihoods, and other observables that affect the performance of the PID such as momentum and χ^2 of the track fit. Photon classifiers use polar information to suppress efficiently beam background, as well as cluster-event time difference, energy, direction, and spatial shape of cluster information. For all particles, the rank computed in the process is used as a discriminating feature for the classifiers. After training, a restriction is applied to the classifier output $\mathcal{P}_{\text{FEI}}$, which encapsulates information from all discriminating observables and their correlations. Depending on particle type, only a subset of top candidates ranked by $\mathcal{P}_{\text{FEI}}$ is saved to reduce combinatorial complexity in forming combined particles from the final state particles. The maximum number of candidates saved per event is shown in the FEI post-selections column in table 4.4.

After the classifier is trained for final state particles, combined particles of the subsequent stage are reconstructed as shown in fig. 4.2. Candidates π^0 mesons are formed from photon pairs. After a loose restriction on the diphoton mass, classifiers are trained using diphoton mass, angle between photon momenta, energy, direction, and classifier outputs of the photons. After restricting on $\mathcal{P}_{\text{FEI}}$ and selecting the highest-ranked ten candidates, π^0 candidates are combined to form heavier particles.

For other combined particles, individual classifiers are trained for each of their decay modes after preselections. The classifiers for $\bar{D}^0$ and other intermediate particles use additional features over those used for π^0 . These include the invariant masses and angles between each combination of decay-product momenta, vertex-fit results, distances to the

Particle	FEI preselection	FEI selection
$e^\pm, \mu^\pm$	$dr < 2 \text{ cm}, dz < 4 \text{ cm}, 10 \text{ hi } \ell\text{ID}$	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 5 \text{ hi } \mathcal{P}_{\text{FEI}}$
$K^\pm, \pi^\pm, p$	$dr < 2 \text{ cm}, dz < 4 \text{ cm}, 20 \text{ hi } p\text{ID}$	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
γ	$E > \{0.10, 0.09, 0.16\} \text{ GeV}$ in {forward, barrel, backward} regions, 40 hi E	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 20 \text{ hi } \mathcal{P}_{\text{FEI}}$
π^0	$0.08 < M < 0.18 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
J/ψ	$2.6 < M < 3.7 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-3}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
Λ	$0.9 < M < 1.3 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
K_S^0	$0.4 < M < 0.6 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
Σ^+	$1.0 < M < 1.4 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-2}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
$D^0/\bar{D}^0, D^\pm$	$1.7 < M < 1.95 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-3}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
$D_s^\pm$	$1.68 < M < 2.1 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-3}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
Λ_c	$2.2 < M < 2.4 \text{ GeV}/c^2, 20 \text{ lo } dM $	$\mathcal{P}_{\text{FEI}} > 10^{-3}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
$D^{*0}/\bar{D}^{*0}, D_{(s)}^{*\pm}$	$Q = M - \sum_i m_i < 0.3 \text{ GeV}/c, 20 \text{ lo } dQ $	$\mathcal{P}_{\text{FEI}} > 10^{-3}, 10 \text{ hi } \mathcal{P}_{\text{FEI}}$
$B^\pm$	$ \Delta E < 0.5 \text{ GeV}, M_{\text{bc}} > 5.2 \text{ GeV}/c^2$ 20 hi $\Pi_i \mathcal{P}_{i, \text{FEI}}$	20 hi $\mathcal{P}_{\text{FEI}}$

Table 4.4: Summary of FEI selections applied to each particle type before and after dedicated BDT training. Reproduced from Ref. [66]. Shorthand ℓID stands for electron-identification or muon-identification for electrons and muons, respectively; $p\text{ID}$ stands for binary particle-identification ratio of kaon vs pion for $K^\pm$, pion vs kaon for π , and proton vs kaon for $p^\pm$; ‘hi’ stands for ‘highest’ and ‘lo’ stands for ‘lowest’. The subscript i runs over all the decay-products.

vertices, and momenta and angles of all decay products in the candidate rest frame. For D^* particles, a requirement is applied to the difference, Q , between the mass of the candidate and the masses of all decay products, instead on the invariant mass, to exploit its greater separation power.

In the final FEI preselection stage, in which B_{tag} candidates are formed, loose preselections on the beam-constrained mass M_{bc} and energy difference ΔE are applied. The beam-constrained mass serves as a control observable to estimate the fraction of correctly reconstructed B_{tag} candidates in data during FEI validation. To reduce correlation between the signal probability and M_{bc} , highly correlated observables, such as invariant masses, are excluded from the classifier training inputs. Instead, ΔE is used as the most discriminating observable in training.

4.4.3 FEI training

The FEI is centrally trained and tested on realistic $B\bar{B}$ simulation. Continuum events are excluded from the training to reduce bias, as they can exhibit properties similar to those of intermediate particles in B decays. For instance, the invariant mass of intermediate particles, an important input observable in FEI training, is the same regardless of the background type. This results in a suboptimal FEI performance in presence of continuum, which calls for additional continuum suppression after FEI reconstruction. Once training is complete, the results are made available to the collaboration. During validation of FEI reconstruction, all backgrounds including continuum are included in the simulation.

4.4.4 B_{tag} selection

The FEI retains the 20 highest-ranked tag B candidates per event for further processing. Among them one may be properly reconstructed and others result from misreconstruction. The tagging purity improves by further restricting the relevant kinematic selections. A proper B_{tag} reconstruction yields M_{bc} close to the B -meson mass of $5.28 \text{ GeV}/c^2$ and ΔE close to 0.0 GeV , as shown in fig. 4.4, as well as high $\mathcal{P}_{\text{FEI}}$ probability (fig. 4.6). As FEI is trained on simulation, its reconstruction efficiency is sensitive to data-simulation disagreement. Indeed, in typical Belle II analyses that use FEI, a 20% – 30% mismatch in FEI efficiency between data and simulation is observed. Chapter 6 discusses a correction to this effect and its validation. I impose the same B_{tag} selections as used in the validation study, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$, $-0.15 < \Delta E < 0.1 \text{ GeV}$, and $\mathcal{P}_{\text{FEI}} > 0.01$ to be able to apply consistently this correction in this analysis. When using B^+B^- -tagging, chances of producing a misreconstructed B_{tag} candidate are higher in a B^+B^- event than in a $B^0\bar{B}^0$ event. Therefore, I expect a larger contribution from charged B background in my sample.

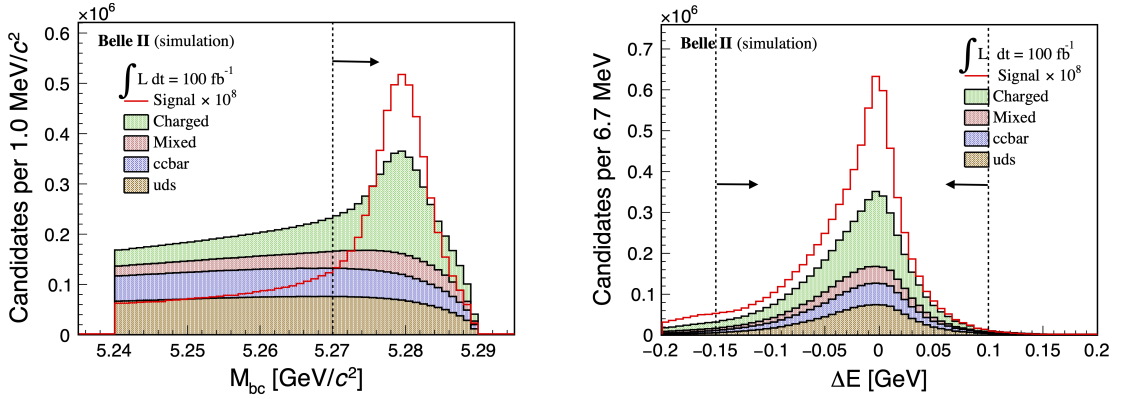


Figure 4.4: Distributions of (left) M_{bc} and (right) ΔE for realistically simulated B_{tag} candidates. ‘Charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; and ‘uds’ represents $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds in simulation. The vertical black dashed lines indicate the selection restrictions with arrows pointing toward the retained events.

The B_{tag} selection reduces 92% of background in simulation. The rest has a 63.2% proportion of continuum, which I further reduce using event-shape information. I use sphericity and $\cos\text{TBTO}$ since particle momenta in $B\bar{B}$ events are isotropically distributed tend to have higher sphericity whereas collimated “jet-like” $q\bar{q}$ event tend to have lower sphericity. Similarly, the cosine of the thrust-angle difference, $\cos\text{TBTO}$, in collimated $q\bar{q}$ jets peaks at 1.0 whereas it has a uniform distribution in $B\bar{B}$ events (fig. 4.5). Candidates B_{tag} are required to meet loose selections, sphericity greater than 0.2 and $\cos\text{TBTO}$ smaller than 0.9, which are consistent with those used for tagging-efficiency correction. These requirements reduce $q\bar{q}$ down to 42.2% with a 13.3% penalty in B_{tag} efficiency according to the simulation.

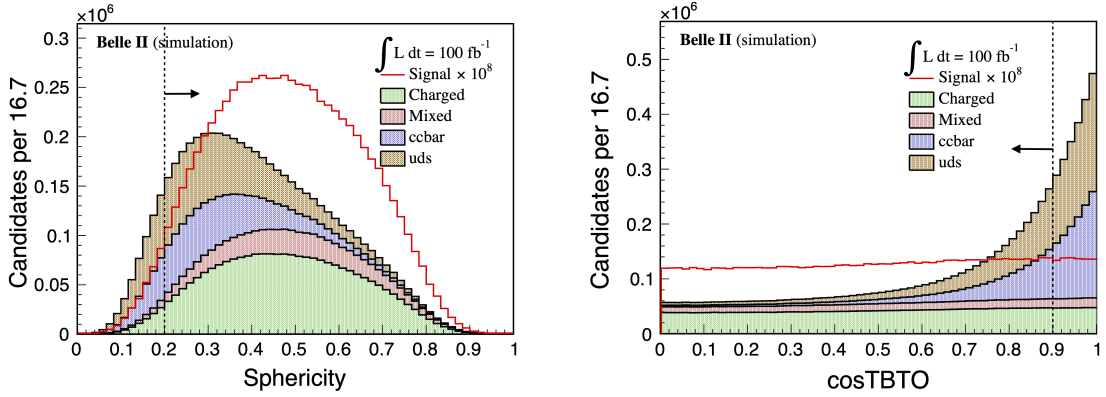


Figure 4.5: Distribution of (left) event sphericity and (right) $\cos\text{TBTO}$ for realistically simulated B_{tag} candidates. The vertical black dashed lines indicate the selection restrictions with arrows pointing toward the retained events. See previous plots for legend conventions.

4.4.5 Multiple B_{tag} candidates

Further background suppression is possible by sorting the B_{tag} meson to use for further analysis among the various B_{tag} candidates that are typically available after reconstruction. This is done by using also information from the signal side. In general some particles remain unassociated with B_{tag} reconstruction. These include signal decay products, misreconstructed particles, and particles not resulting from the e^+e^- collision. These last do not emerge from the beam interaction point preferentially. Hence, I restrict the tracks to originate in close proximity of the interaction point by imposing $dr < 0.5$ cm and $|dz| < 2$ cm, and polar angle within drift-chamber acceptance ($17^\circ < \theta < 150^\circ$). Tracks are also required to have at least 20 hits in the drift chamber. Figure 4.7 shows the resulting track multiplicity in simulation. The signal yields a charged kaon and two τ leptons, each decaying into one lepton and undetected neutrinos, for a total of three signal tracks. Hence, a restriction on the multiplicity of signal-side tracks helps to reduce high multiplicity backgrounds and, consequently, the multiplicity of B_{tag} candidates. Restricting the analysis to events having exactly three quality tracks after B_{tag} reconstruction reduces by 80% background with 62% signal efficiency loss in simulation.

Table 4.5 shows the simulated B_{tag} efficiency and purity as a function of the progressive imposition of selection criteria. After all selections, about 0.5% of all generated generic B events are retained. The B_{tag} purity is 44%. The corresponding average B_{tag} multiplicity is 1.8. I restrict the B_{tag} candidates to one per event by choosing the B_{tag} candidate with the highest $\mathcal{P}_{\text{FEI}}$.

4.4.6 Choice of τ final states and treatment of lepton final states

Before detailing the signal selection I motivate the choice of signal final states targeted in this search.

The search sensitivity for $B^+ \rightarrow K^+\tau^+\tau^-$ decays depends on the τ final states. I compare the averaged expected upper limits in realistic background-only samples among various τ -decay final states that contain one charged particle. These are classified into three categories, $K\ell\ell$, $K\ell\pi$, and $K\pi\pi$. In the pion modes, ρ mesons contribute significantly as

B_{tag} selection	B_{tag} efficiency (%)	Purity (%)
Loose selection	6.1	14.9
$M_{\text{bc}} > 5.27 \text{ GeV}/c^2$	3.5	25.6
$-0.15 < \Delta E < 0.1 \text{ GeV}$	3.3	26.5
$\mathcal{P}_{\text{FEI}} > 0.01$	1.5	50.5
Sphericity > 0.2	1.5	50.3
$\text{CosTBTO} < 0.9$	1.3	50.5
Signal-side track multiplicity of 3	0.5	43.9

Table 4.5: Evolution of efficiency and purity in simulated signal events after each B_{tag} selection criterion. ‘Loose selection’ criteria includes $M_{\text{bc}} > 5.24 \text{ GeV}/c^2$, $-0.2 < \Delta E < 0.2 \text{ GeV}$, and $\mathcal{P}_{\text{FEI}} > 0.001$.

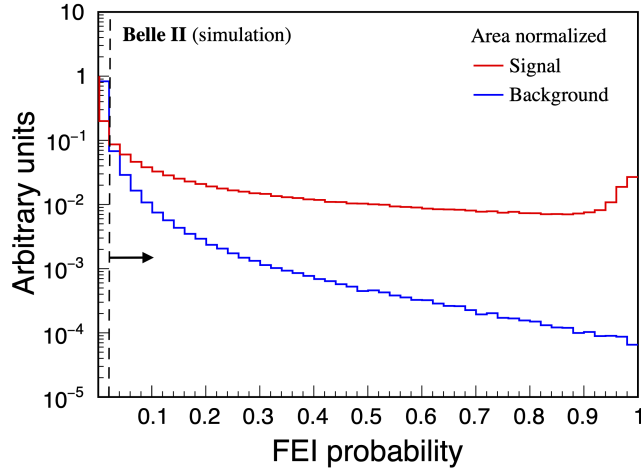


Figure 4.6: Distributions of FEI probability for realistically simulated B_{tag} candidates. Distributions are normalized to unity.

they decay into a charged pion and a neutral pion. This increases signal efficiency; however, unlike in the $K\ell\ell$ case, the light-quark background dominates in these modes.

Due to their differing kinematic properties, each category requires dedicated selection criteria to achieve optimal sensitivity. I optimize the selection criteria separately using realistic simulated samples and determine the final selection based on the best-expected limit. The study (detailed in appendix C) amounts to repeating the analysis on many simulated experiments and shows that the sensitivity for the $K\ell\ell$ final state is 1.6 to 2.5 times better than that of the pion final states. Hence, I focus exclusively on the $K\ell\ell$ final state for this first Belle II analysis. I therefore consider $B^+ \rightarrow K^+\tau^+\tau^-$ decays where τ^- decays into $e^-\nu_\tau\bar{\nu}_e$ or $\mu^-\nu_\tau\bar{\nu}_\mu$. This contributes to a 12.4% fraction of the $B^+ \rightarrow K^+\tau^+\tau^-$ width.

In leptonic τ decays, three final states are possible, Kee , $Ke\mu$, and $K\mu\mu$. Treating each separately may enhance the overall sensitivity as in principle they are affected by backgrounds of different nature and size. I therefore compare the expected sensitivity among them by optimizing selection criteria separately using realistic simulated samples as done for the choice of τ final states. The combined sensitivity of the three categories treated

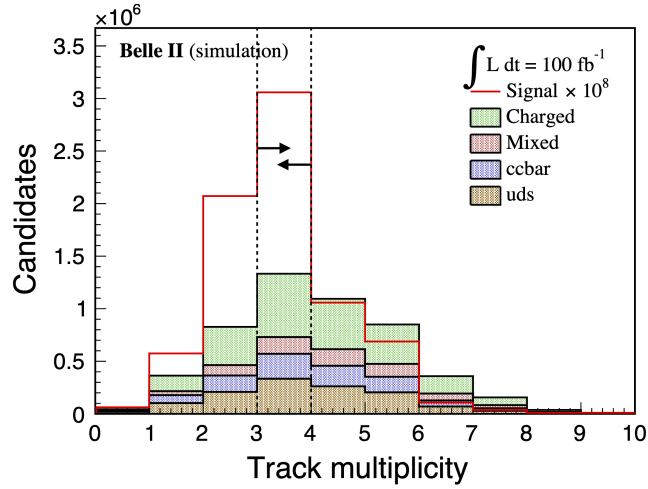


Figure 4.7: Multiplicity of simulated tracks not included in B_{tag} reconstruction. See previous plots for legend conventions.

individually shows only a 8% improvement over treating the leptonic final states inclusively (details in appendix B). I therefore choose to proceed with the inclusive treatment of the lepton final states for simplicity in this first Belle II analysis.

4.4.7 Signal B meson selection

The signal B meson (B_{sig}) is reconstructed from the particles not used in B_{tag} reconstruction. Signal yields three tracks and no clusters. Quality track requirements are described in section 4.4.5. Tracks are further associated to mass hypotheses depending on particle identification information (fig. 4.8). I identify a track as kaon if its kaon-over-pion binary likelihood under kaon hypothesis ($\mathcal{L}(K/\pi)$) exceeds 0.75. The criterion retains 99% of true kaons in simulated signal decays. I identify a track as an electron if its global likelihood under the electron hypothesis exceeds 0.9. This requirement retains 94% of true electrons in simulated signal decays. I identify a track as a muon if its global likelihood under muon hypothesis exceeds 0.9. This requirement retains 79% of true muons in simulated signal decays. The lower purity in the muon channel is due to the inability to distinguish from pions low-momentum muons, which may fail to reach the muon detector. This is shown in fig. 4.9, which displays the momentum distributions of truth-matched muons and pions misidentified as muons. In the subsequent steps of the analysis, a restriction on the minimum lepton momentum improves muon purity. All the above particle identification criteria are optimal, as described in chapter 6.

The signal is reconstructed using all final-state particle combinations that meet the selections. Hence the $B^+ \rightarrow K^+ \tau^+ (\rightarrow \ell^+ \nu \bar{\nu}) \tau^- (\rightarrow \ell^- \nu \bar{\nu})$ signal results in the following combinations: $K^\pm e^\pm e^\mp$, $K^\pm e^\pm \mu^\mp$, $K^\pm \mu^\pm e^\mp$, and $K^\pm \mu^\pm \mu^\mp$. Due to the open kinematics from undetected neutrinos, neither M_{bc} peaks at the B meson mass nor ΔE peaks at zero, as shown in fig. 4.10. Then, B_{sig} and B_{tag} of opposite charge are combined together to form the $\Upsilon(4S)$ candidate.

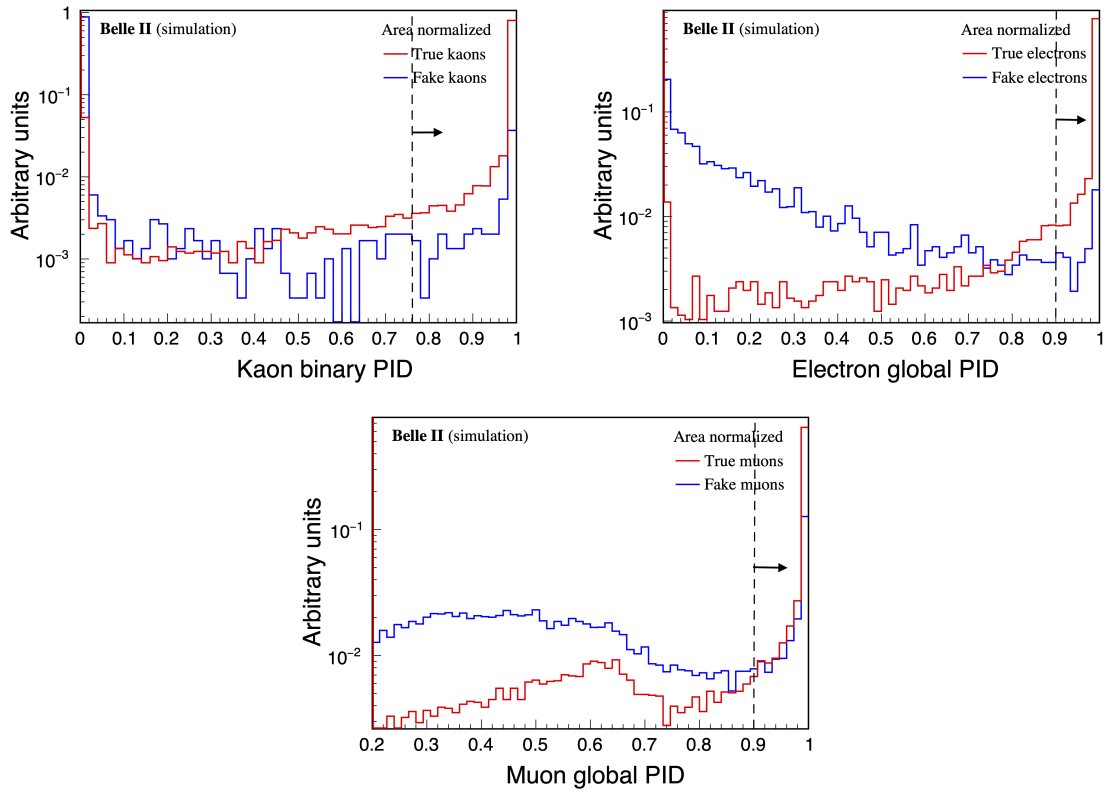


Figure 4.8: Distributions of (top-left) binary particle-identification for kaons, (top-right) global particle-identification for electrons, and (bottom) global particle-identification for muons in simulation.

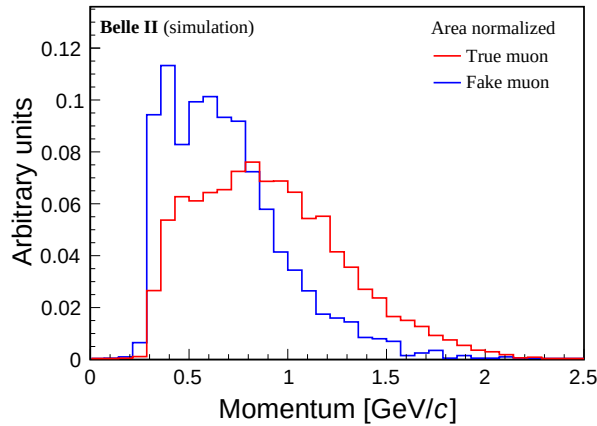


Figure 4.9: Distributions of momentum for genuine muons and for pions misidentified as muons in simulated signal events.

4.5 Baseline selection

Tag B meson reconstruction and selection suppresses large background from $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^0\bar{B}^0$, $q\bar{q}$, $\tau^+\tau^-$, and $\mu^+\mu^-$ processes. The remaining background is domi-

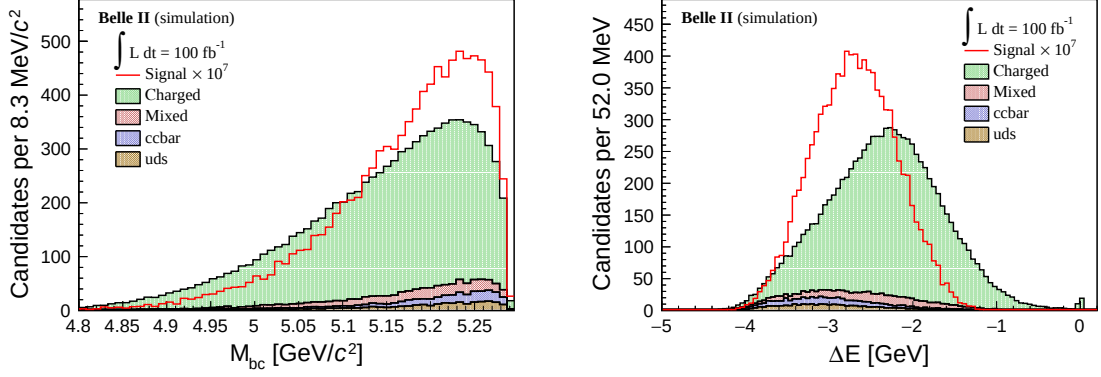


Figure 4.10: Distributions of (left) M_{bc} and (right) ΔE for simulated B_{sig} candidates in a realistic sample. See previous plots for legend conventions.

nated (88%) by B^+B^- decays. At this stage, signal efficiency in simulation is 4×10^{-4} , which reduces to 3×10^{-4} if signal candidates are truth-matched. Suppressing further background with acceptable loss in efficiency would enhance search sensitivity. I therefore apply restrictions on observables that have clearly distinct distributions between signal and background. These restrictions are not systematically optimized at this stage as visual inspection of distributions is sufficient for our purpose.

In the oppositely charged kaon-lepton pair mass ($m(K^+t^-)$), a prominent D meson peak is visible due to $\overline{D}^0 \rightarrow K^+\pi^-$ background where the pion is misidentified as lepton (mostly muon), as shown in fig. 4.11. I reject signal candidates with $1.8 < m(K^+t^-) < 1.9$ GeV/c^2 . Similarly, a prominent J/ψ peak is visible in lepton-pair mass ($m(\ell^+\ell^-)$), as shown in fig. 4.11. I suppress charmonium contributions by rejecting signal candidates with $m(\ell^+\ell^-) > 3$ GeV/c^2 . Dilepton mass also captures photon conversions and electron-positron pairs with $m(\ell^+\ell^-) < 100$ MeV/c^2 are removed. Those requirements reduce background by 4% with negligible loss of signal efficiency. Signal exhibits a symmetric q^2 distribution, reflecting the unconstrained nature of the $\tau^+\tau^-$ system, as shown in fig. 4.11. Background shows an asymmetric structure, thereby offering some discrimination. To enhance signal purity, a $q^2 > 14.18$ GeV^2/c^4 selection is applied. This removes the low- q^2 background tail and excludes contributions from $B^+ \rightarrow K^+\psi(2S)[\rightarrow \tau^+\tau^-]$ decays, which have an effective branching fraction of 1.8×10^{-6} , much higher than signal. In addition, to suppress this background, the requirement aligns consistently this search with the phase-space boundaries chosen for SM predictions [32]. This requirement reduces background by 17% with 8% loss in signal efficiency.

I also reconstruct π^0 mesons not included in the $\Upsilon(4S)$ reconstruction by combining pairs of photons not associated to the signal or tag B candidate with energies greater than 55 MeV in the calorimeter barrel and cluster-to-track distance greater than 30 cm.¹ All candidate π^0 are subjected to a mass-constrained fit, which imposes the diphoton mass to equal the known π^0 mass, and corrects the photon energies accordingly. Figure 4.11 shows the resulting diphoton mass distribution. I remove signal candidates if $0.131 < m(\gamma\gamma) < 0.138$ GeV/c^2 . This requirement reduces background by 27% with 10% loss in signal efficiency.

With all the above selections, 0.02% and 0.01% fractions of events have multiple sig-

¹These requirements increase photon purity (as described in chapter 5).

nal candidates with average multiplicities of 2.0 in realistic simulated events and data, respectively. I randomly choose one $\Upsilon(4S)$ candidate per event. This restriction has a negligible impact on background yield and signal efficiency. All the selection criteria introduced in this chapter up to this point collectively configure the **baseline selection**, which is referred to multiple times in subsequent chapters. Table 4.6 outlines the evolution of signal efficiency, truth-matched signal efficiency, and background rejection as a function of each selection step. After the above baseline selections, the background contains 90.0% of B^+B^- , 5.6% of $B^0\bar{B}^0$, and 4.3% of $q\bar{q}$ events. Table 4.7 provides a summary of background composition. The dominant processes are $B^+ \rightarrow \bar{D}^{(*)0}\ell^+\nu_\ell$ decays, where the D meson undergoes semileptonic decays, accounting for 71% of the total background.

Selection	Efficiency ($\times 10^{-4}$)	Truth- matched efficiency ($\times 10^{-4}$)	Fraction of background events (%)
B_{tag} reconstruction and continuum suppression	130.0	52.9	100
Multiplicity of tracks not associated to B_{tag}	53.7	23.6	20
B_{sig} reconstruction (τ BF, IP, and PID)	4.00	2.99	0.185
$m(K^+t^-) < 1.8$ or > 1.9 GeV/ c^2	3.80	2.83	0.180
$m(\ell^+\ell^-) < 3.0$ GeV/ c^2	3.80	2.83	0.176
$m(\ell^+\ell^-) > 0.1$ GeV/ c^2	3.79	2.82	0.173
$q^2 > 14.18$ GeV $^2/c^4$	3.47	2.59	0.143
π^0 veto	3.12	2.45	0.105
Multiple candidate selection	3.12	2.44	0.105

Table 4.6: Reconstruction efficiency, truth-matched efficiency, and background rejection at each step of the baseline selection, as expected from simulation.

Background	Fraction (%)
$B^+ \rightarrow \bar{D}^{0*}\mu^+\nu$	25.2
$B^+ \rightarrow \bar{D}^{0*}e^+\nu$	24.5
$B^+ \rightarrow \bar{D}^0e^+\nu$	9.9
$B^+ \rightarrow \bar{D}^0\mu^+\nu$	9.5
$B^+ \rightarrow \bar{D}^{0*}\tau^+\nu$	2.2
Other $B\bar{B}$ background	24.4
Continuum background	4.3

Table 4.7: Background composition in a simulated sample after baseline selection. All uncertainties are smaller than 0.1%.

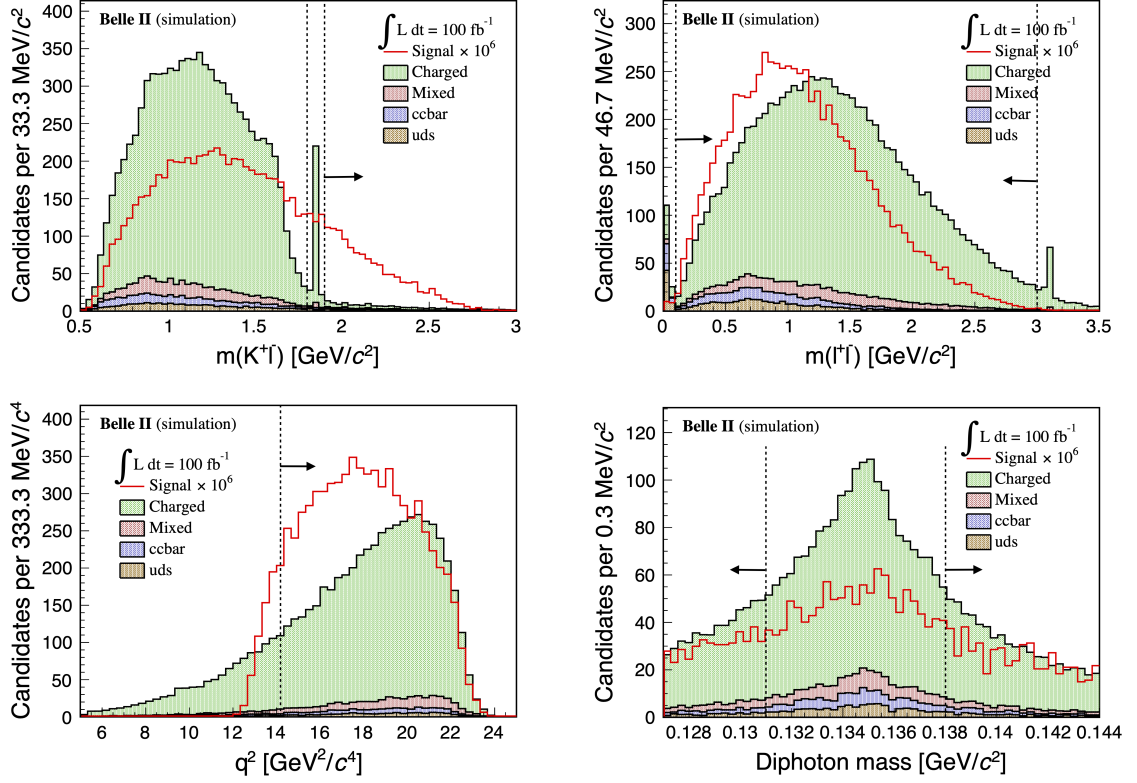


Figure 4.11: Distributions of (top-left) opposite-charge kaon-lepton pair mass, (top-right) lepton pair mass, (bottom-left) beam-constrained squared $\tau^+\tau^-$ mass (bottom-left), and (bottom-right) mass of photon pairs not included in $\Upsilon(4S)$ reconstruction in simulation at each step of their baseline selections. Vertical dashed lines show the selections with arrows pointing toward the accepted regions. See previous plots for legend conventions.

4.6 Opposite-charge kaon-lepton mass

Inspection of kinematic observables reveals an interesting opportunity. A major transition in background composition is observed in the $m(K^+t^-)$ spectrum at the D meson mass (fig. 4.11, top-left). The large component of semileptonic D meson decays or hadronic D meson decays with misreconstructed π^0 's (e.g., $D^0 \rightarrow K^- \pi^+ \pi^0$) clusters at $m(K^+t^-)$ mass below the charm mass. This allows to identify a signal-rich region with little background in the higher $m(K^+t^-)$ spectrum exploiting the signal tail that extends up to $m(K^+t^-) \approx 2.7$ GeV/c^2 . This is a very powerful finding that allows an innovation in this analysis in that it conveniently restricts the search to a background-depleted sample. The $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ requirement reduces background by 99.2% with 83.7% loss in efficiency. The background contains 54% B charge events, 14% B mixed events, and 32% continuum events. Table 4.8 shows the detailed background composition in the region $m(K^+t^-) > 1.9 \text{ GeV}/c^2$. In an attempt at further suppressing continuum, I study selections on various event-shape observables. This reduces continuum but also reduces the signal efficiency, resulting in worse sensitivity. I therefore do not suppress further continuum. A study discussed in appendix A) shows no improvement in sensitivity by using the full, unrestricted $m(K^+t^-)$ range as the increase in signal efficiency is spoiled by much higher backgrounds. I therefore opt for an analysis restricted in $m(K^+t^-)$. This also allows for basing the

selection on sequential one-dimensional selections rather than statistical-learning classifiers keeping the analysis simpler and more reliable. A multivariate analysis in the restricted $m(K^+t^-)$ region is not possible due to the small sample size, which is not sufficient to train the classifier.

4.7 Further discriminating observables

The M_{miss}^2 distribution, as shown in fig. 4.12, has higher values for signal events compared to background events, primarily due to the higher neutrino multiplicity in $B^+ \rightarrow K^+\tau^+\tau^-$ processes. This is not the case for background, which mostly has one or two neutrinos in the final state. Restricting events to larger values of M_{miss}^2 suppresses these backgrounds. Similarly, the lepton with the same charge as the signal kaon has a distinctive momentum spectrum in the center-of-mass frame, as shown in fig. 4.12. In chapter 6 I discuss an optimization of the selection of these observables and define the E_{ECL} signal region based on the best search sensitivity.

Background	Fraction (%)
$B^+ \rightarrow \overline{D}^{*0}\ell^+\nu$	11.1 ± 1.4
$B^0 \rightarrow D^{*-}\ell^+\nu$	5.7 ± 0.1
$B^+ \rightarrow \overline{D}^0\ell^+\nu$	5.6 ± 0.1
$B^+ \rightarrow J/\psi K^{*+}$	2.5 ± 0.1
$B^+ \rightarrow D_s^{*-}K^+\ell^+\nu$	2.5 ± 0.1
Other $B\overline{B}$ background	40.3 ± 3.0
Continuum background	31.9 ± 2.6

Table 4.8: Background composition in a simulated sample after $m(K^+t^-) > 1.9 \text{ GeV}/c^2$.

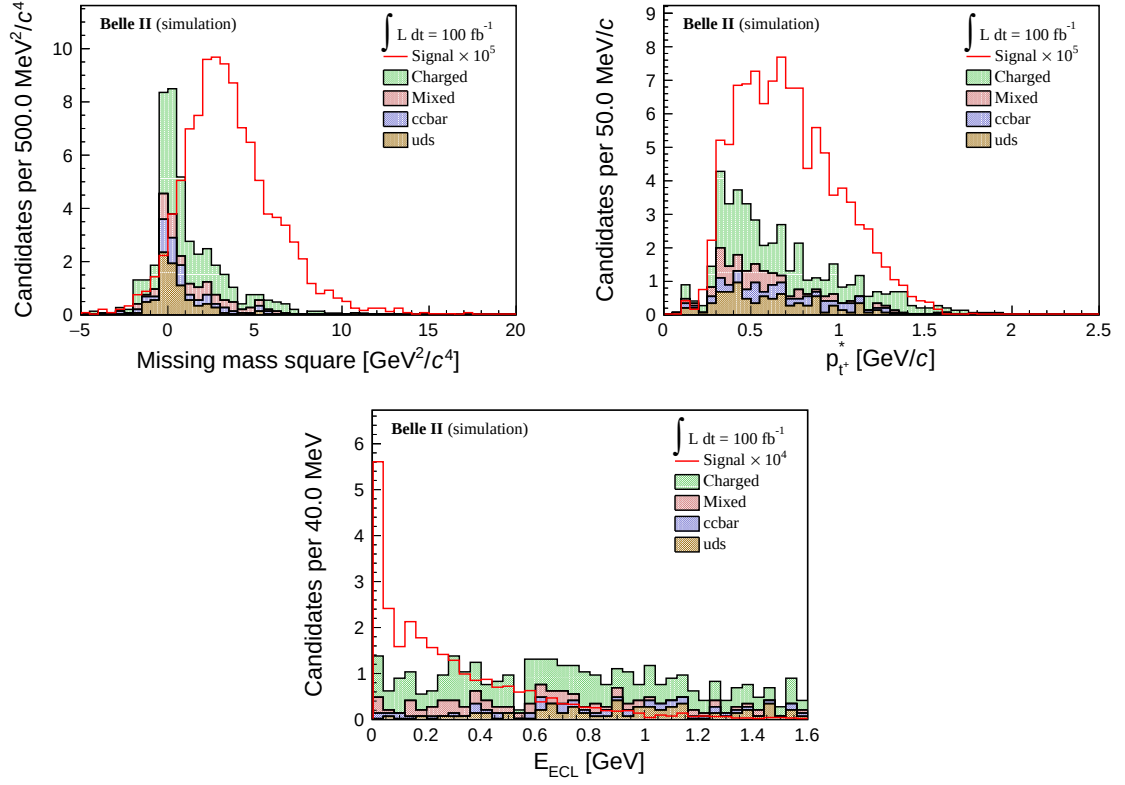


Figure 4.12: Distributions of (top-left) M_{miss}^2 , (top-right) p_{t+}^* , and (bottom) E_{ECL} in $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ region of a simulated sample. See previous plots for legend conventions.

Chapter 5

Residual calorimeter energy

This chapter describes the validation of the signal-extraction observable and its dependence on the underlying photon selection and on sample composition.

5.1 Importance and strategy

As mentioned in previous chapters, the residual energy detected in the calorimeter (E_{ECL}) after $B\bar{B}$ reconstruction, that is, the sum of the energies of all neutral clusters, is the signal-extraction observable. Figure 5.1 shows the E_{ECL} distribution in realistic $B^+ \rightarrow K^+ \tau^+ \tau^-$ simulation. Properly reconstructed signal events have a narrow E_{ECL} distribution peaking at zero with a moderate tail extending to up to 1 GeV. This shape is due to on-threshold production and an hermetic detector. These two aspects prevent large unaccounted for energy from being excluded from the signal reconstruction. Only small contributions from beam-background photons, tag-side photons not associated with the B_{tag} , and clusters not generated by photons may contribute. In contrast, background events have a smooth distribution extending to fairly large values as, in addition to the above sources, signal-side photons that are not properly associated contribute large, additional energy deposition. The small secondary peak at $E_{\text{ECL}} \approx 0.0$ in the background distribution originates from any background with same final states as signal, such as $B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \ell^- \bar{\nu}_\ell) \ell^+ \nu_\ell$ decays. Figure 5.2 compares E_{ECL} in $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ and $B^+ \rightarrow \bar{D}^0 \pi^+$ data and simulation. Significant mismodeling is observed, which exemplifies the susceptibility to mismodeling of this observable. This is expected by virtue of its inclusive nature. In order for simulation to describe properly E_{ECL} , it is necessary that not just all $\mathcal{T}(4S)$ final states are accurately simulated, but also beam background and other instrumental effects are so. Because mismodeling is likely to bias the search results, thorough E_{ECL} validation is essential in this work and for similar Belle II analyses.

The E_{ECL} properties primarily depend on how photons are reconstructed and selected and on the composition of the specific sample in which E_{ECL} is used. Photon selection and reconstruction are key as they determine the basic elements used to construct E_{ECL} . Restrictions on cluster properties may be used to selectively suppress specific photon-candidate sources; however, care must be taken for their effect on the signal E_{ECL} properties. Restrictions often limit contributions of certain cluster-energies thus sculpting the E_{ECL} distribution. As an example, fig. 5.3 illustrates the impact of the cluster selection on E_{ECL} for three simulated scenarios, loose photon candidates, standard candidates, and tighter candidates. Loose constraints increase the contamination from various sources, reducing signal resolution and discriminating power. Stringent constraints accept a lower

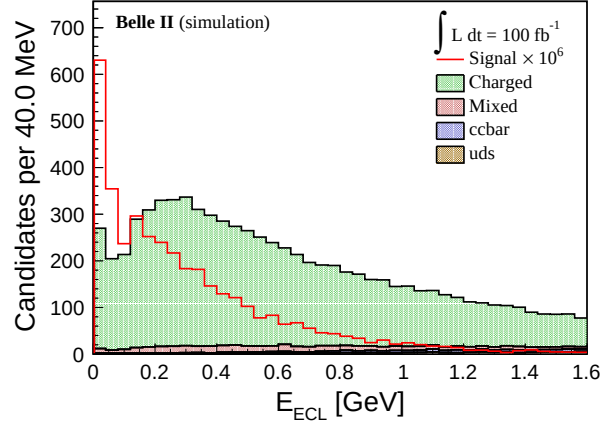


Figure 5.1: Distribution of E_{ECL} for simulated signal and background events in the signal region. ‘Charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; and ‘uds’ indicates $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds in simulation.

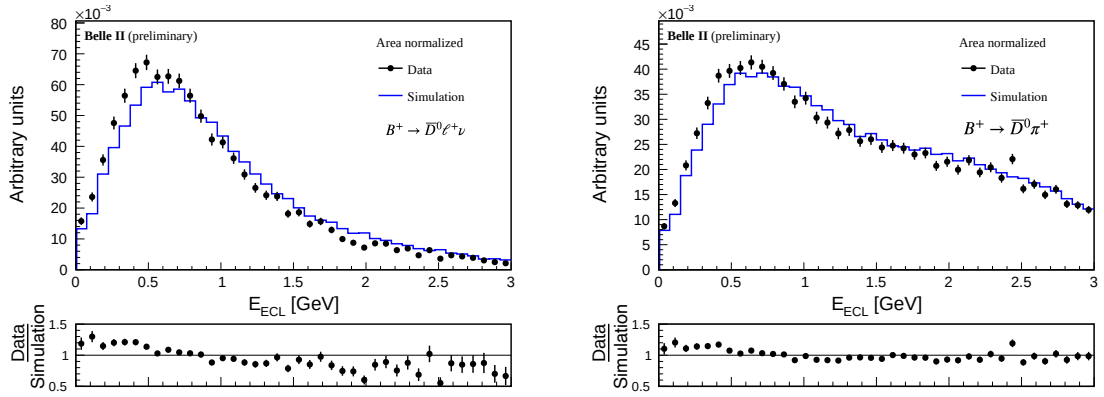


Figure 5.2: Distributions of E_{ECL} in (points) data and (histogram) simulation for (left) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (right) $B^+ \rightarrow \bar{D}^0 \pi^+$ events. In this and subsequent figures, the “simulation” label means simulated samples containing B^+B^- , $B^0\bar{B}^0$, and $q\bar{q}$ events. Distributions are normalized to unity.

number of photon candidates contributing to E_{ECL} , which results in both signal and background tending to peak at zero, thus reducing signal-to-background separation. Sample composition is equally relevant to E_{ECL} modeling as the relative abundance of sources of mismodeled photons may determine the overall quality of the E_{ECL} description.

At given photon selection and sample composition, the E_{ECL} depends on individual cluster energies and on the number of clusters not associated to the $\Upsilon(4S)$ (also called “multiplicity”). Any mismodeling of E_{ECL} should therefore originate from an underlying inconsistency in either or both these distributions. I therefore study the E_{ECL} description in several control samples and using photon candidates selected with various criteria. The goal is to achieve a photon selection capable of both yielding a properly modeled E_{ECL} spectrum while keeping the sensitivity of the search high.

In the first part of this study I examine and validate the E_{ECL} properties using several

control samples selected with baseline criteria looser than the final, optimized analysis selection. In addition to ensuring large sample sizes and thus providing statistically clearer information, this choice makes my findings sufficiently general to be applicable to similar analyses in Belle II. Once E_{ECL} validation is established in this “universal” configuration, I restrict the validation to the final selection for proper application in this analysis.

I begin the E_{ECL} investigation based on baseline photon selections across several control samples. Next, I examine the consistency of E_{ECL} descriptions in samples enriched in specific sources of photon candidates in an attempt at isolating the sources of discrepancies. Then, I develop dedicated photon selections that mitigate the observed inconsistencies. Finally, I validate E_{ECL} with the final photon selection.

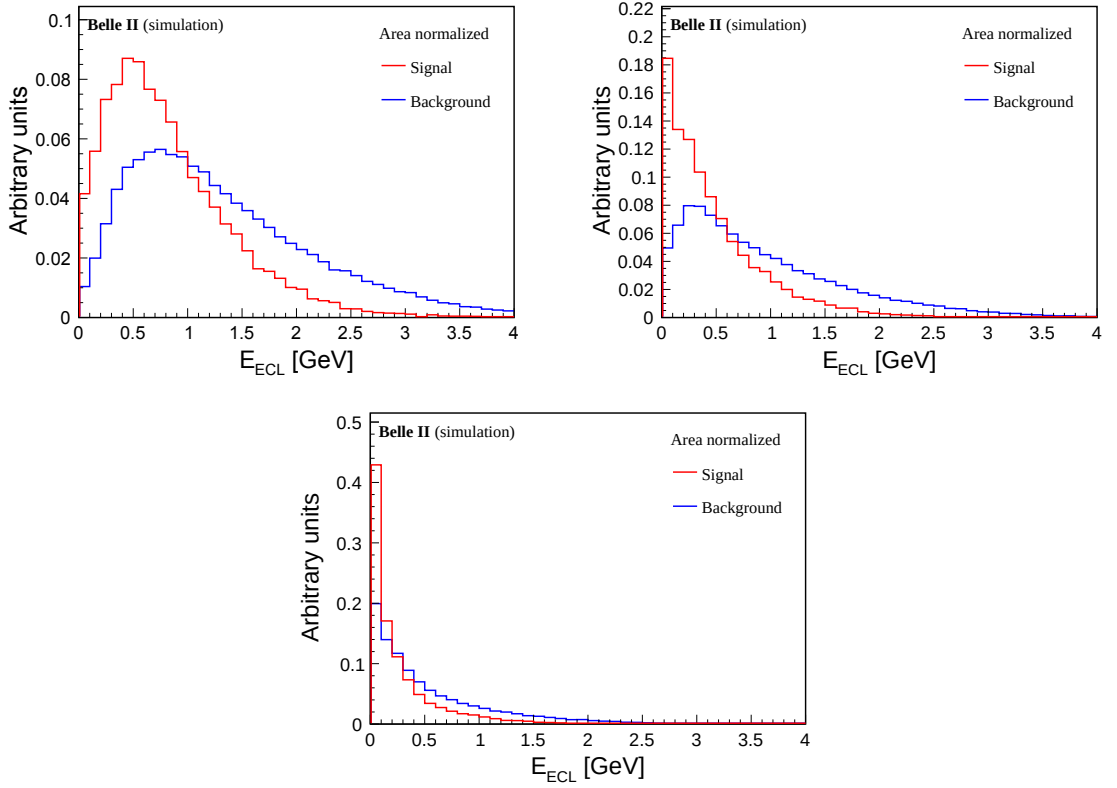


Figure 5.3: Distributions of E_{ECL} in simulated signal and background events containing (top-left) photons unrestricted in cluster-to-track distance, (top-right) photons with cluster-to-track distance greater than 30 cm, and (bottom) photons with cluster-to-track distance greater than 80 cm. Distributions are normalized to unity.

5.2 Control samples

To ensure reproducibility of the findings against variations of sample composition, I identify multiple data control-samples that have three tracks as the signal sample and are reconstructed with the signal baseline selection (chapter 4) except for individual criteria that may be changed to keep blind the signal-search region,

1. q^2 sideband: events with $q^2 < 12 \text{ GeV}^2/c^4$.

2. Wrong τ -pair charge sideband: $\tau^+\tau^+$ and $\tau^-\tau^-$ events.
3. Wrong B charge sideband: events with same-charge K and B_{tag} .
4. Particle-identification sideband: events with binary kaon-over-pion identification of the kaon candidate smaller than 0.6.
5. $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\ell^+\nu_\ell$ sideband: events in which kaons and pions are combined to reconstruct D^0 mesons with mass $1.84 < m(K^+\pi^-) < 1.89 \text{ GeV}/c^2$ and kaons (pions) are selected with binary kaon (pion) particle identification over pion (kaon) greater than 0.6.
6. $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\pi^+$ sideband: events with the same D^0 meson reconstruction and particle-identification criteria as for $B^+ \rightarrow \bar{D}^0\ell^+\nu_\ell$.

To keep findings of this study applicable to other Belle II analyses, I do not apply any π^0 veto (section 4.5) at this stage. This is added in a later validation stage once the rest of the photon selection is established.

5.3 E_{ECL} sources

Various sources of photon candidates deposit energy in the calorimeter, potentially contributing to E_{ECL} .

5.3.1 Misreconstructed photons

Misreconstructed photons are calorimeter-energy deposits identified as photons that are not generated from primary photon interactions in the calorimeter. Typically these photon candidates are partially or mistakenly reconstructed from scattered multiple clusters, whose main sources are as follows.

A charged hadron may undergo hadronic interactions in the calorimeter and generate an irregular shower shape, different from that of photons or electrons (section 2.3.3). Its energy deposition is usually spread into multiple clusters. Those detected sufficiently far from the related drift-chamber track may be mistakenly reconstructed as photons. Figure 5.4 (left) shows an illustrative example. In addition, charged particles with low transverse momentum and therefore high curvature may cross multiple calorimeter crystals at large angles with respect to the crystal main axes generating energy deposits spread into multiple clusters (fig. 5.4, right). Finally, another relevant source are neutral hadrons, such as neutrons and K_L^0 , which may undergo strong interactions in the calorimeter producing irregular, spread shower shapes not associated with tracks. Due to their expected small contribution to E_{ECL} , I do not study this last source separately. All these sources potentially contribute to E_{ECL} .

5.3.2 Beam-induced background photons

Beam-induced background photons are photons not originating from the hard scattering, but from the interaction of the beam with the accelerator or associated infrastructure material. They primarily result from beam particles being deviated due to bremsstrahlung or Coulomb scattering with residual gas molecules in the beam pipe or with other beam particles in the same bunch. Heavy-metal shields in the beam enclosure, superconducting

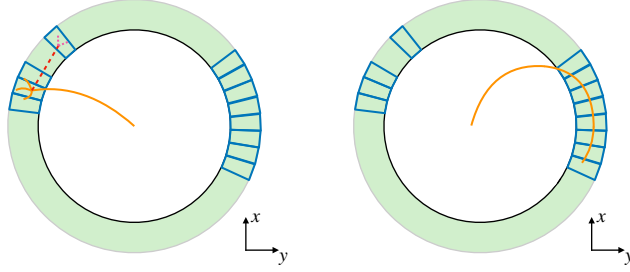


Figure 5.4: A simplified sketch of an axial section of the calorimeter showing (left) a hadronic interaction by a charged hadron and (right) a low transverse-momentum charged particle impinging in the calorimeter.

final-focus cryostats, and horizontal and vertical movable collimators suppress shower particles entering the Belle II acceptance, but they are not fully efficient, leading to sizable contaminations, which contribute to E_{ECL} .

5.3.3 Collision photons

A collision photon is a properly reconstructed photon that emerges from the cascade of interactions following e^+e^- annihilation. Physics backgrounds yielding final-state photons such as radiative or hadronic decays (e.g., $B^+ \rightarrow \bar{D}^{*0}(\rightarrow \bar{D}^0\gamma)\ell^+\nu_\ell$ or $B^+ \rightarrow \bar{D}^{*0}(\rightarrow \bar{D}^0\pi^0(\rightarrow \gamma\gamma))\ell^+\nu_\ell$) that are not properly reconstructed into the $\Upsilon(4S)$ candidate contribute to E_{ECL} .

5.3.4 Source identification

Understanding the contributions to E_{ECL} from each source using simulation may help to single out, and possibly suppress, the main effects. However, photon classification in simulation is challenging because photon truth matching (section 2.3.3) has shortcomings for a significant fraction of misreconstructed photons and for all beam-background photons in current Belle II simulation. Since misreconstructed photons spread their deposit across multiple clusters, truth matching often fails in identifying them, as it is technically nontrivial to devise algorithms that associate all the proper crystals to the particles that caused the signals. In addition, beam backgrounds are not simulated. They are sampled at random in data and overlaid to simulated events. As a result, beam backgrounds intrinsically lack generation information, which prevents any identification through truth matching. Hence, one cannot exploit solely “Monte Carlo truth” to study these sources and must also rely empirically on distinctive features at the reconstruction level. I achieve this by combining information of photon-matching weights and cluster-event time difference.

Photon-matching weights quantify the strength of the causal relation between reconstructed clusters and generated particles in simulation. A small weight means likely failure in associating the simulated particle that generated the cluster; a large weight indicates a reliable association. Figure 5.5 shows the weight distribution for simulated background in the $B^+ \rightarrow K^+\tau^+\tau^-$ signal-search region after the baseline selection. Since beam-background photons are not generated, their photon matching weights peak at zero. Misreconstructed photons have a long lower tail that extends to zero as well. To separate these two categories, cluster-event time difference helps. This is the difference between the time at which energy deposits are detected in the calorimeter and the collision time, as recorded

by various detectors.

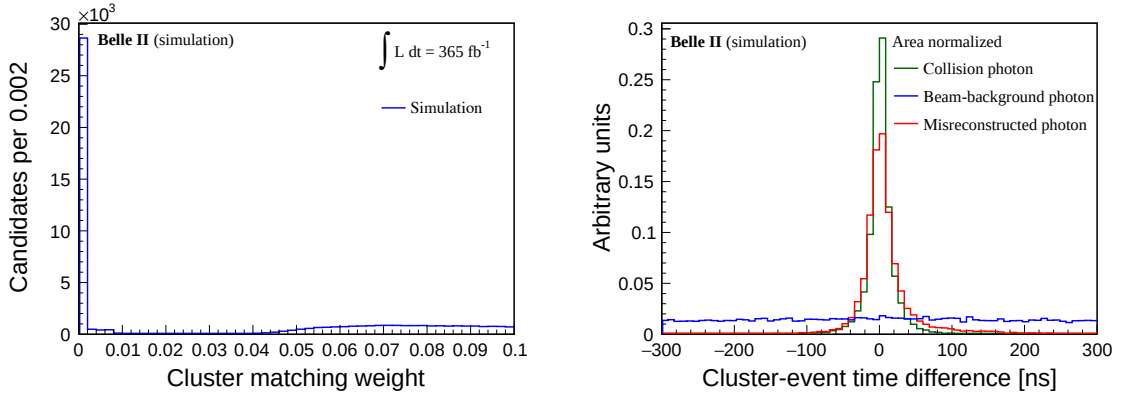


Figure 5.5: Distributions of (left) photon-matching weight for clusters that fail truth-matching and (right) cluster-event time difference for all clusters in a simulated sample of background populating the signal-region. (Right) Distributions are normalized to unity.

The shapes of the cluster-event time difference for misreconstructed and beam-background photons differ. The energy-deposition time of misreconstructed photons peaks at 0.0 ns because they deposit energy shortly after the e^+e^- collision. In contrast, beam-background photons do not originate from the collision and can deposit energy in the calorimeter at any time, in principle. As a consequence, beam-background photons have a smoother, nearly uniform cluster-event time difference distribution, as shown in fig. 5.5.

I inspect the matching weights of simulated clusters that fail truth matching and check the corresponding cluster-event time difference distribution. Figure 5.6 shows a comparison of time-difference for events with weights in the $[0.002, 0.020]$ and $[0.02, 0.03]$ ranges. The unmatched photon candidates with low weights have a uniform distribution, reflecting beam-background properties, whereas the photons with high weights peak around zero, reflecting collision properties. Based on these findings I identify as a beam-background photon a simulated photon-candidate that fails matching with weight less than 0.02. I identify as a misreconstructed photon that failed truth matching a simulated-photon candidate with weight greater than 0.02. This identification is only relevant in simulation studies because no photon matching is possible for experimental data obviously.

5.3.5 Baseline photon selection and isolation of source-specific samples

Photon selections alter the E_{ECL} distribution and impact its discriminating power, but they do not exclude any signal candidates. To investigate their role I start by comparing E_{ECL} in data and simulation for photon candidates selected with minimal criteria that are sufficiently generic to be useful for other Belle II analyses.

Since beam-background preferentially contributes low-energy photons, photon candidates with energies inferior to 20 MeV are not included in standard Belle II reconstruction. In addition, an online selection imposes that all photon candidates with energies between 20 to 50 MeV detected earlier or later from the event time than 99% of collision photons of that energy, are discarded to further reduce beam-background. This additional criterion introduces a significant data-simulation discrepancy near 50 MeV, as shown in fig. 5.7, which impacts all analyses. To circumvent this mismodeling, I select photons with 55 MeV minimum energy for this study and further analysis.

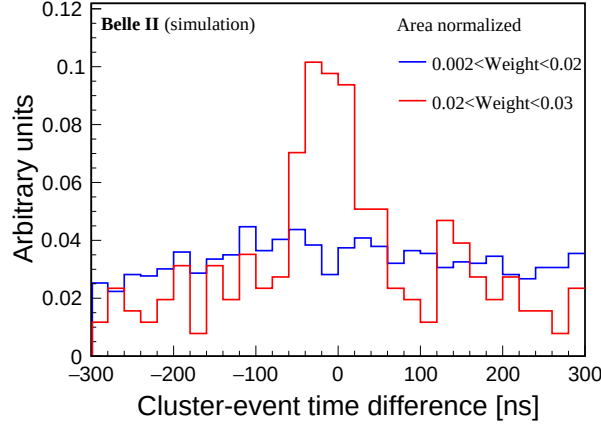


Figure 5.6: Distributions of cluster-event time difference for photon candidates with matching weight in the range (red) $[0.002, 0.020]$ and (blue) $[0.02, 0.03]$ from a simulated sample of background populating the signal region. Distributions are normalized to unity.

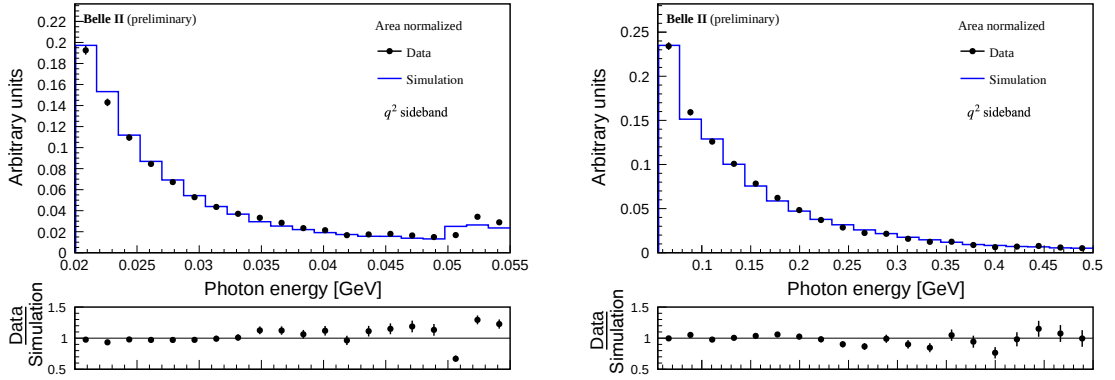


Figure 5.7: Distributions of energy for baseline-selected photon candidates reconstructed (points) data and (histogram) simulation with energies (left) $E < 55$ MeV and (right) $E > 55$ MeV and restricted to the q^2 sideband. Distributions are normalized to unity.

To understand the discrepancies associated with a specific source, I isolate various samples, each enriched in a specific photon class.

Misreconstructed photon sample As misreconstructed photons mostly originate from energy depositions by charged particles, the associated cluster is usually spatially close to the extrapolation onto the calorimeter of a drift-chamber track. Hence, photon candidates with small cluster-to-track distance have a higher probability of being misreconstructed photons. Figure 5.8 shows the cluster-to-track distance distribution of simulated background in the signal region. Misreconstructed photons peak at lower values, which allows to isolate a misreconstructed-photon sample through a cluster-to-track distance restriction to below 20 cm.

Beam-background photon sample Beam-background photons tend to have a uniform event-time difference distribution (fig. 5.5). Hence, photons at higher cluster-event time differences are most likely beam-background photons. Moreover, the chance of being a misreconstructed photon decreases if the associated cluster is distant from

the extrapolated track. I therefore isolate a beam-background-enriched sample by requiring cluster-event time difference greater than 150 ns and cluster-to-track distance greater than 20 cm.

Collision photon sample A straightforward way to obtain a collision photon sample is suppressing both misreconstructed and beam-background photons by inverting the above criteria. A photon with a large cluster-to-track distance and a small cluster-event time difference is likely to be a collision photon. I therefore construct a sample enriched in collision photons by requiring cluster-to-track distance greater than 50 cm, cluster-event time difference less than 150 ns, and cluster restricted to the barrel region ($32.2^\circ < \theta < 128.7^\circ$) of the calorimeter as beam-background photons originate from interactions of a single beam, with which they tend to be collinear, and are therefore more concentrated in the endcap compartments of the calorimeter.

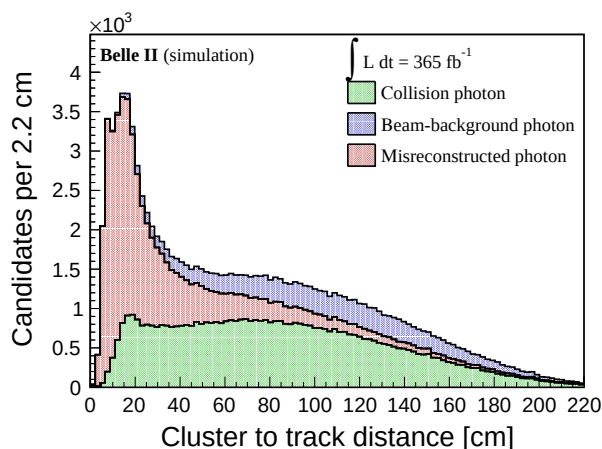


Figure 5.8: Distribution of cluster-to-track distance in a simulated sample of background populating the signal region.

Figure 5.9 shows the results of this classification in the control samples introduced in section 5.2. Distributions of cluster polar-angle in each sample are shown, each enriched in photons from one category to give a visual representation of sample composition. Although modest contaminations remain, my classification criteria achieve fairly pure samples for each component, with compositions that are qualitatively consistent across the various studied. I expect that a data-simulation consistency study of each photon-candidate source would be sufficient to reliably identify the most serious sources of E_{ECL} mismodeling.

5.4 Validation

Now that I established the factors impacting E_{ECL} , isolated quasi-pure samples of the relevant sources, and identified multiple control samples to test the universality of the effects, I proceed to the proper E_{ECL} study and validation. Figure 5.10 compares the E_{ECL} shape between data and simulation in the control samples introduced in section 5.2. All samples show data-simulation consistency, except for the $B \rightarrow D$ channels, which show large discrepancies.

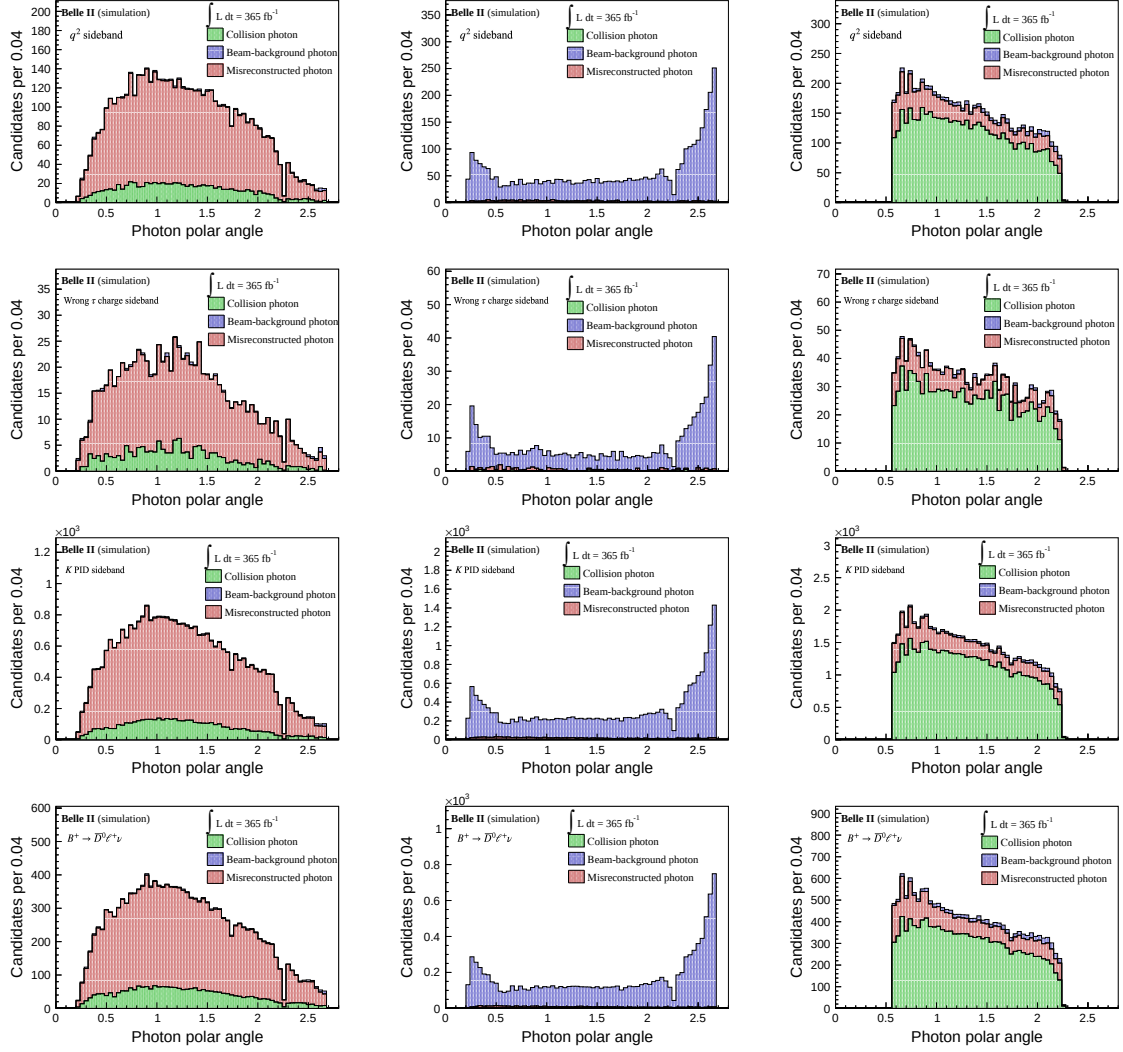


Figure 5.9: Distributions of polar angle for photon candidates in simulated samples enriched in (left) misreconstructed photons, (middle) beam-background photons, and (right) collision photons, from events from the (first row) q^2 sideband, (second row) wrong τ charge sideband, (third row) kaon PID sideband, (fourth row) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$. Simulated distributions are normalized to data luminosity.

This motivates checking consistency separately for photon energy and photon multiplicity for those samples, in an attempt at identifying the cause of the mismodeling and possibly improve it (figs. 5.11 and 5.12). As expected, $B \rightarrow D$ control channels have simulated photon-energy and multiplicity distributions inconsistent with data.

5.4.1 Photon energy

To further specialize the investigation, I compare photon-energy distributions between data and simulation for each of the photon categories and for each control sample (fig. 5.13). The most significant discrepancies are observed for samples enriched in misreconstructed photons. This calls for suppressing misreconstructed photons by requiring cluster-to-track

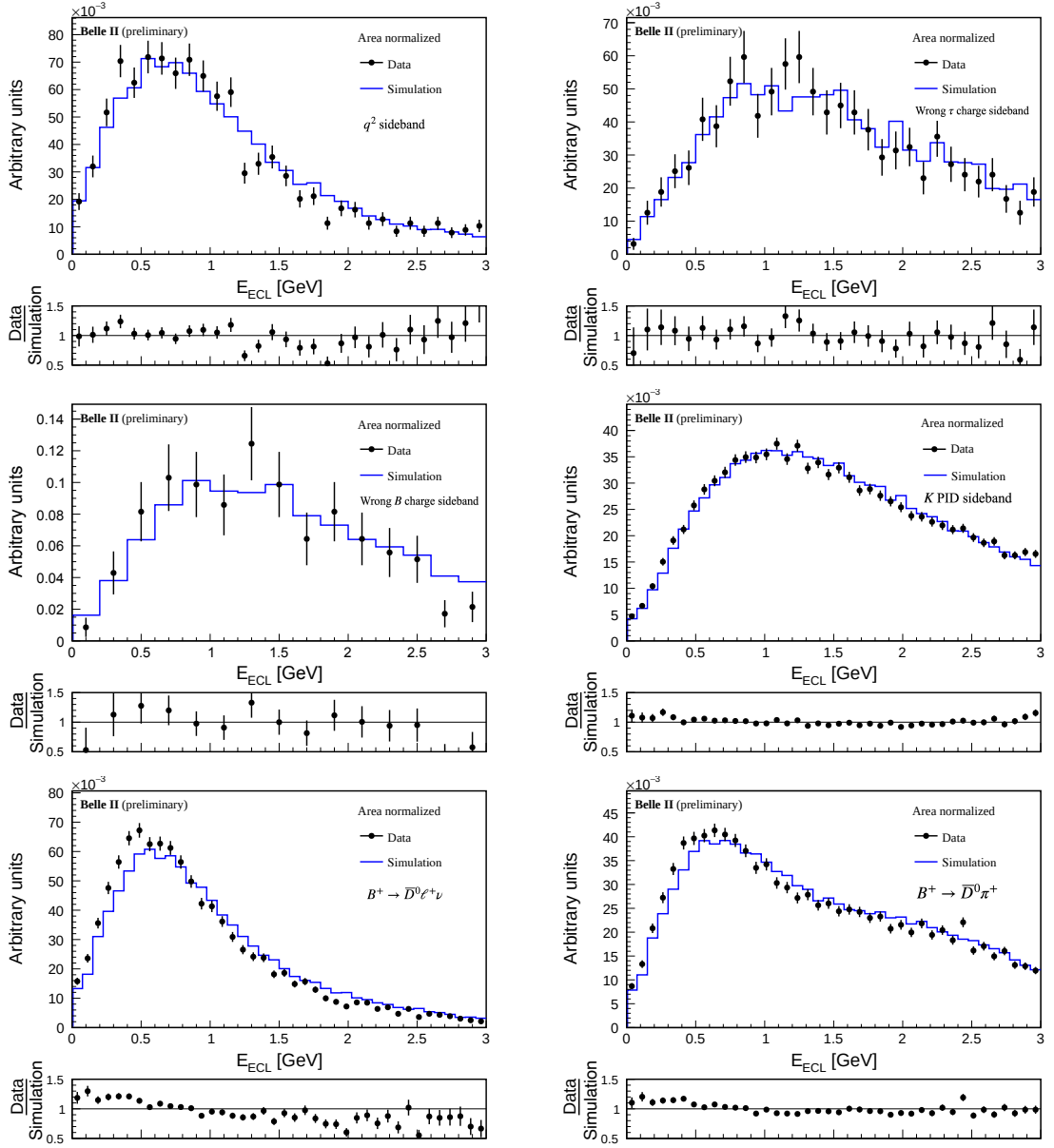


Figure 5.10: Distributions of E_{ECL} for (points) data and (histogram) simulated events in the (top-left) q^2 sideband, (top-right) wrong τ charge sideband, (middle-left) wrong B charge sideband, (middle-right) kaon PID sideband, (bottom-left) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (bottom-right) $B^+ \rightarrow \bar{D}^0 \pi^+$. Distributions are normalized to unity.

distance greater than 20 cm. Figure 5.14 shows the impact on E_{ECL} . The general picture is improved, but distributions show a residual mismodeling in $B \rightarrow D$ samples, prompting a further check of photon multiplicity of beam-background and collision photons.

5.4.2 Photon multiplicity

In analogy with what is done with photon energy, I compare photon-multiplicity distributions (fig. 5.15). All show consistent descriptions except for the $B \rightarrow D$ channels, which are

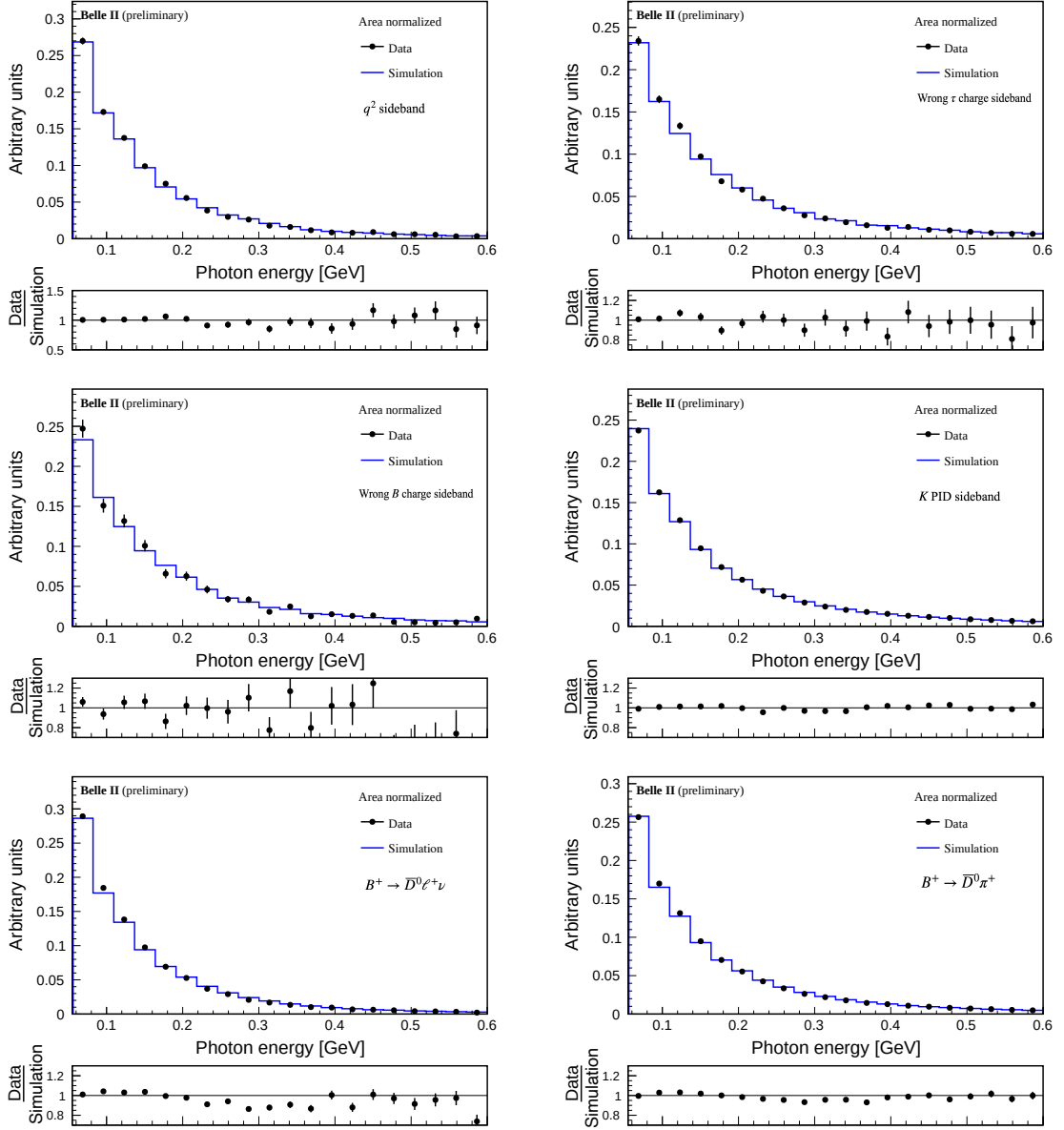


Figure 5.11: Distributions of photon energy for (points) data and (histogram) simulated events in the (top-left) q^2 sideband, (top-right) wrong τ charge sideband, (middle-left) wrong B charge sideband, (middle-right) kaon PID sideband, (bottom-left) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$, and (bottom-right) $B^+ \rightarrow \bar{D}^0 \pi^+$. Distributions are normalized to unity.

affected by large discrepancies for collision photons. The mismodeling in the region of very low multiplicity is particularly concerning as that is the region most relevant for the E_{ECL} signal-search region. The observed discrepancy suggests that the residual E_{ECL} mismodeling in $B \rightarrow D$ channels is primarily due to the mismodeling of collision-photon multiplicity. A dedicated investigation of the fundamental reason for this mismodeling would be lengthy and complicated and it is outside the scope of this work. However, a plausible hypothesis is that the mismodeling may origin in the poor experimental knowledge of contributions from decays involving charm resonances, such as D^{**} . Here an empirical solution such as

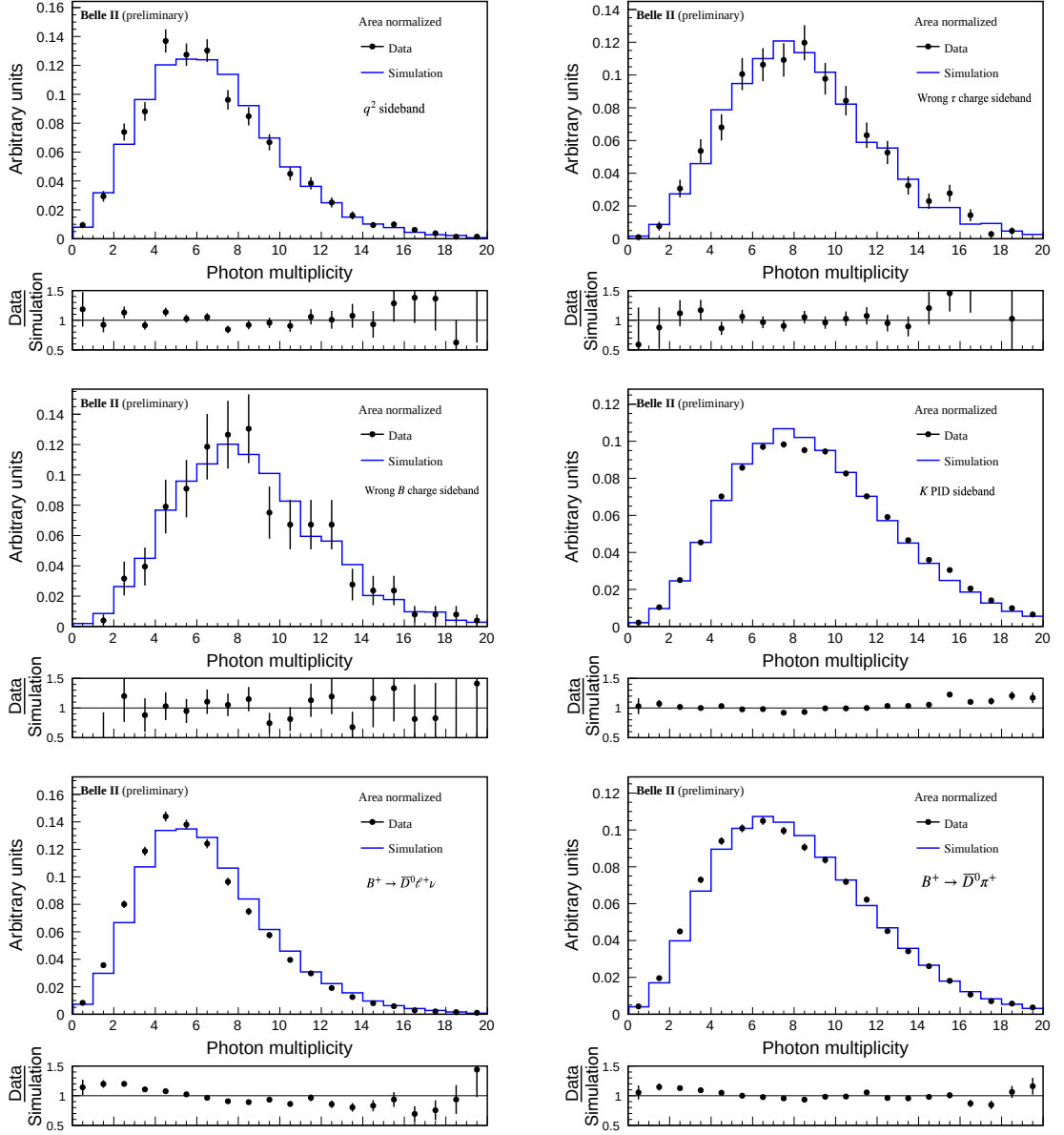


Figure 5.12: Distributions of photon multiplicity for (points) data and (histogram) simulated events in the (top-left) q^2 sideband, (top-right) wrong τ charge sideband, (middle-left) wrong B charge sideband, (middle-right) kaon PID sideband, (bottom-left) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (bottom-right) $B^+ \rightarrow \bar{D}^0 \pi^+$. Distributions are normalized to unity.

collision-photon suppression would not work because that would bias the E_{ECL} distributions of signal and background to become similar, thus degrading discriminating power.

5.4.3 Intermediate summary

In summary, misreconstructed photons exhibit photon-energy inconsistencies across all samples (table 5.1). However, these can be reduced by more restrictive cluster quality selections, specifically on the distance from extrapolated tracks. Mismodeling in E_{ECL} is

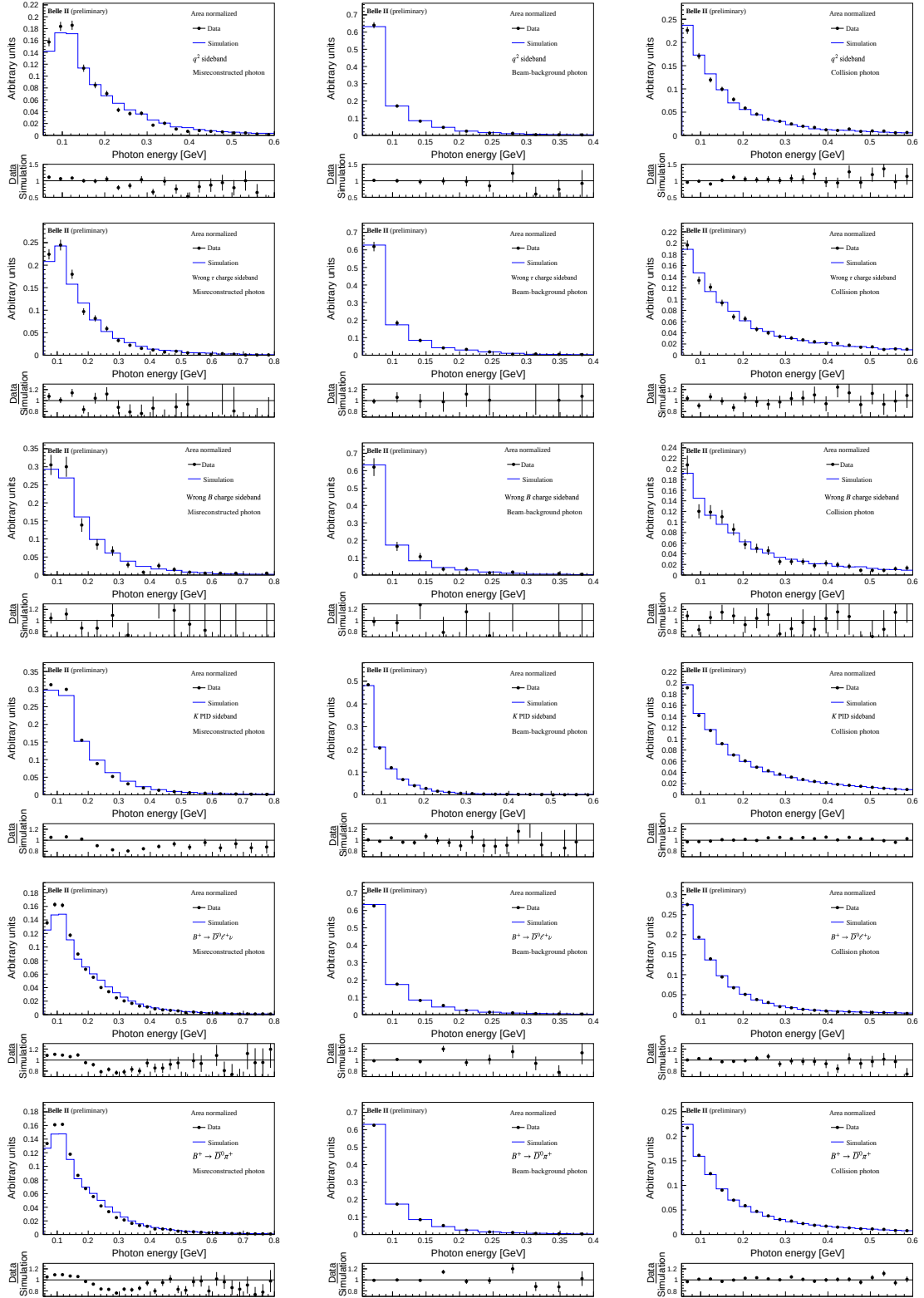


Figure 5.13: Distributions of photon energy for (point) data and (histogram) simulated samples enriched in (left) misreconstructed photons, (middle) beam-background photons, and (right) collision photons, from events in the (first row) q^2 sideband, (second row) wrong τ charge sideband, (third row) wrong B charge sideband, (fourth row) kaon PID sideband, (fifth row) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (sixth row) $B^+ \rightarrow \bar{D}^0 \pi^+$ samples. Distributions are normalized to unity.

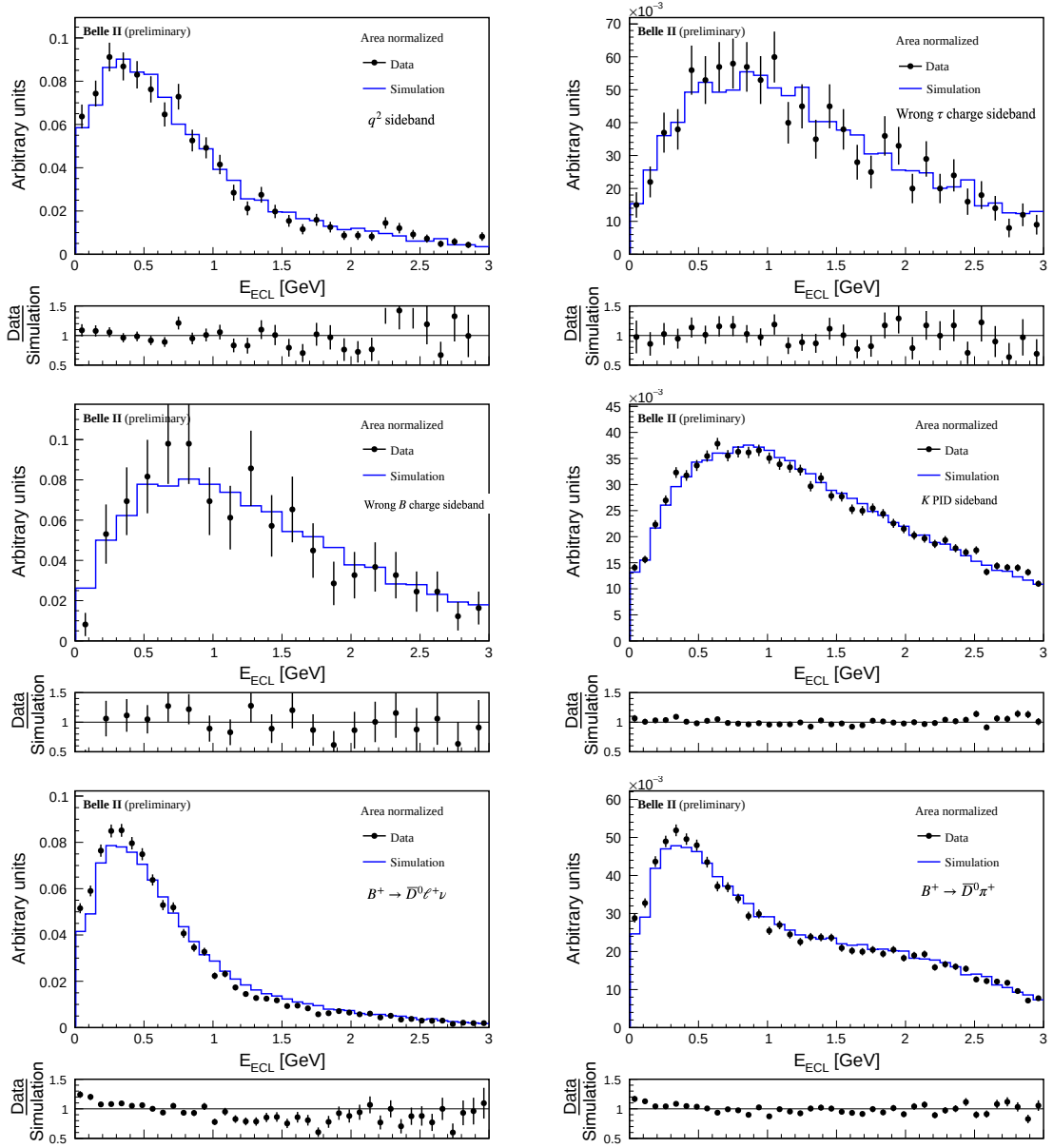


Figure 5.14: Distribution of E_{ECL} for (points) data and (histogram) simulated photon candidates with cluster-to-track distance greater than 20 cm in the (top-left) q^2 sideband, (top-right) wrong τ charge sideband, (middle-left) wrong B charge sideband, (middle-right) kaon PID sideband, (bottom-left) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (bottom-right) $B^+ \rightarrow \bar{D}^0 \pi^+$. Distributions are normalized to unity.

more pronounced and harder to reduce when due to photon-multiplicity discrepancies. This happens most prominently in $B \rightarrow D$ channels, where the collision-photon multiplicity is poorly modeled resulting in an overall poor E_{ECL} description. In the following, I discuss an attempt at a correction of the collision-photon multiplicity mismodeling in $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$.

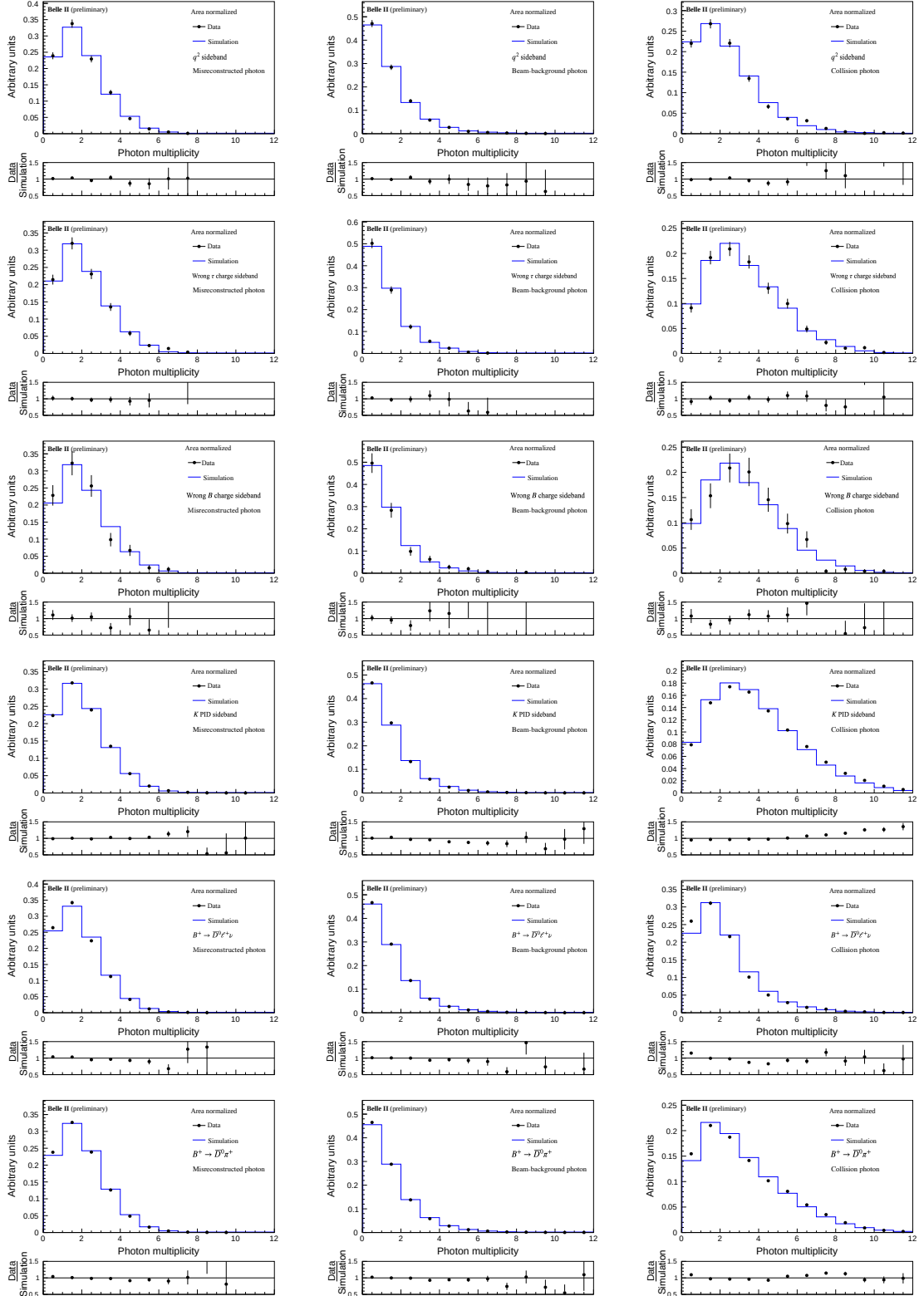


Figure 5.15: Distributions of photon multiplicity for (point) data and (histogram) simulated samples enriched in (left) misreconstructed photons, (middle) beam-background photons, and (right) collision photons, from events in the (first row) q^2 sideband, (second row) wrong τ charge sideband, (third row) wrong B charge sideband, (fourth row) kaon PID sideband, (fifth row) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (sixth row) $B^+ \rightarrow \bar{D}^0 \pi^+$ samples. Distributions are normalized to unity.

Sample	E_{ECL}	Misreconstructed photon energy	Collision photon multiplicity	All other sources
q^2 sideband	Consistent	Inconsistent	Consistent	Consistent
Wrong τ charge sideband	Consistent	Inconsistent	Consistent	Consistent
Wrong B charge sideband	Consistent	Inconsistent	Consistent	Consistent
Kaon PID sideband	Consistent	Inconsistent	Consistent	Consistent
$B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$	Inconsistent	Inconsistent	Inconsistent	Consistent
$B^+ \rightarrow \bar{D}^0 \pi^+$	Inconsistent	Inconsistent	Inconsistent	Consistent

 Table 5.1: Summary of the E_{ECL} consistency study.

5.4.4 Photon-multiplicity correction in the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample

Reducing the collision-photon multiplicity mismodeling may improve E_{ECL} mismodeling in $B \rightarrow D$ samples. Multiplicity mismodeling means that rates of collision photons at certain multiplicity values differ between data and simulation. I explore an empirical correction based on weighting the collision-photon multiplicity of a simulated $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample, enriched in collision photons, to mirror the data. Here the selection aimed at enhancing the collision-photon contents is used and weights are being bin-by-bin data-to-simulation ratios taken from fig. 5.15, right panel in the fifth row. Figure 5.16 shows the impact by comparing collision-photon multiplicities with and without corrections in the same, collision-photon-enriched $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample. As expected, weighting significantly reduces the data-simulation inconsistency. The remaining small discrepancies are due to 25% of contamination from non-collision photons, as shown in fig. 5.9. After this correction, I check the consistency of photon multiplicity in a more realistic sample that has larger contributions from non-collision photons.

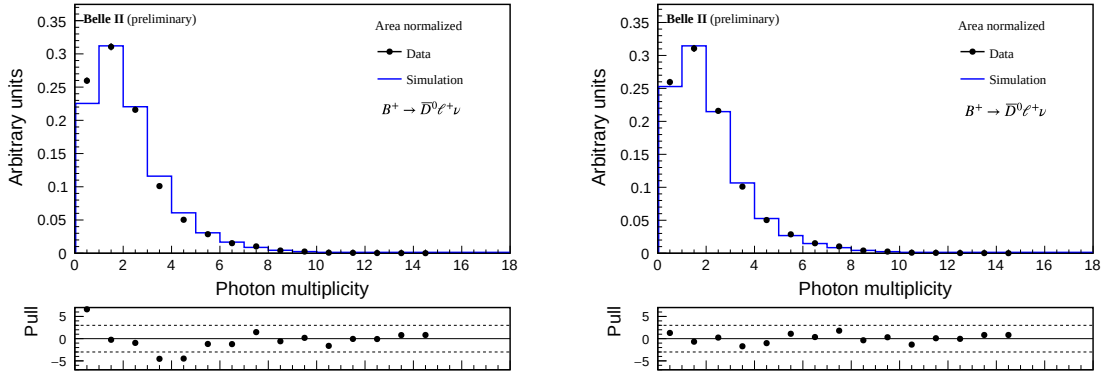


Figure 5.16: Distribution of collision-photon multiplicity for $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ (point) data and (histogram) simulation (left) without and (right) with collision-photon multiplicity weights. Distributions are normalized to unity.

Figure 5.17 shows the impact of the weighting on photon multiplicity in a $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample that is more loosely selected (same selection as in Fig. 5.14) and therefore includes larger contamination from non-collision photons. A residual mismodeling is visible, but

weighting improves data-simulation consistency. I therefore check how E_{ECL} changes after the correction is applied (fig. 5.18). Corrected simulation shows consistency with experimental data. Hence, the misreconstructed-photon suppression and collision-photon multiplicity correction fix the E_{ECL} mismodeling in $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ events.

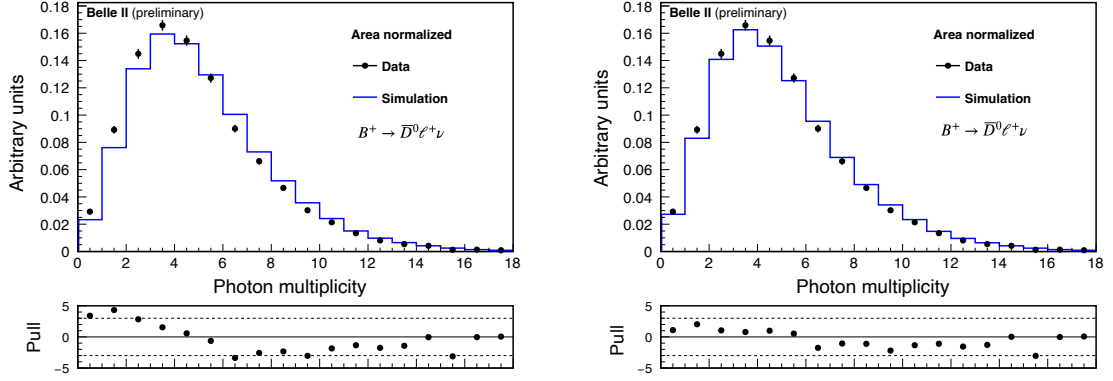


Figure 5.17: Distribution of photon multiplicity for $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ (point) data and (histogram) simulation (left) without and (right) with collision-photon multiplicity weights. Distributions are normalized to unity.

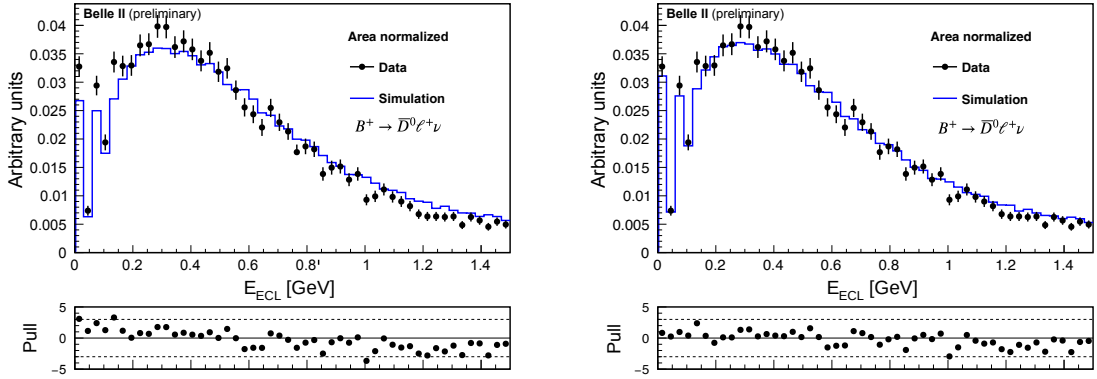


Figure 5.18: Distribution of E_{ECL} for $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ (point) data and (histogram) simulation (left) without and (right) with collision-photon multiplicity weights. Distributions are normalized to unity.

5.4.5 Sample-dependence checks

As an empirical correction is involved, it is important to check its applicability to samples other than that from which it is derived, otherwise, it cannot be trusted to properly correct the signal sample. I test the connections in the channel $B^+ \rightarrow \bar{D}^0 \pi^+$, an independent sample showing large E_{ECL} mismodeling (fig. 5.14). As in $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ decays, the inconsistency originates from collision-photon multiplicity mismodeling (fig. 5.15). I therefore correct the $B^+ \rightarrow \bar{D}^0 \pi^+$ collision-photon multiplicity with the weights derived from the

$B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ samples. The correction improves the $B^+ \rightarrow \bar{D}^0 \pi^+$ data-simulation consistency showing a satisfactory degree of “universality” of weights, at least across $B \rightarrow D$ decays (fig. 5.19).

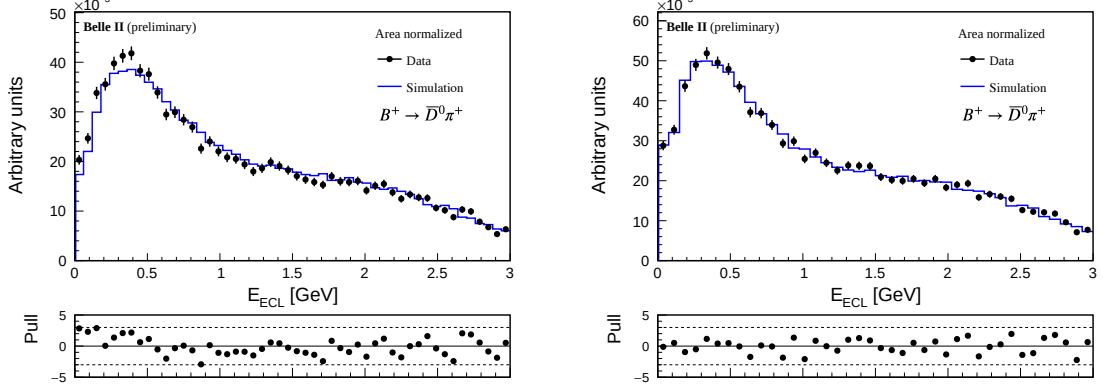


Figure 5.19: Distribution of E_{ECL} in $B^+ \rightarrow \bar{D}^0 \pi^+$ samples in (point) data and (histogram) simulation (left) without and (right) with collision-photon multiplicity weights derived from the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample. Distributions are normalized to unity.

However, it is essential to verify that the weights are applicable to other control samples, and that they do not introduce large data-simulation discrepancies in samples that already show consistent E_{ECL} descriptions. I therefore apply the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ collision-photon multiplicity weights in all other control samples. Figure 5.20 illustrates the impact. The $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ weights improve data-simulation E_{ECL} consistency in the q^2 sideband; however, they introduce new data-simulation E_{ECL} differences in all other samples.

The nonuniversality of the empirical collision-photon multiplicity weighting indicates that data-simulation differences in the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ and $B^+ \rightarrow \bar{D}^0 \pi^+$ samples have a different origin from those in other samples. This is plausible, as sample composition and kinematic properties, which are important drivers of mismodelings, differ across the control samples. In the absence of further studies that investigate and possibly pinpoint the fundamental reasons of such mismodelings, it is unrealistic to find an empirical weighting that simultaneously corrects the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ and $B^+ \rightarrow \bar{D}^0 \pi^+$ photon multiplicities without spoiling consistency in other samples.

I therefore decide to cover the possibility of additional E_{ECL} mismodeling in the signal-search region with an additional systematic uncertainty, discussed in detail in the next chapters. This is based on the difference in the amount of expected background observed between applying, or not, the collision-photon-multiplicity weighting that fixes the discrepancies in the $B \rightarrow D$ samples. While the composition and kinematic properties of our signal sample and the $B \rightarrow D$ samples differ, we assume that the difference induced in the results by a correction that fixes the largest discrepancy observed in any of the control samples is a sufficiently generous range to bracket any possible residual and unaccounted for mismodeling.

5.5 Final photon selection

These studies show the achievement of a satisfactory E_{ECL} model for events not containing $B \rightarrow D$ decays owing to a dedicated photon selection including requirements of energy

greater than 55 MeV and cluster-to-track distance greater than 20 cm. This is a promising outcome for my analysis and useful information for all other Belle II analyses using E_{ECL} , which can apply these findings as they are, or adapt them to their photon selections using the same conceptual logic. However, additional photon-selection criteria are required to improve E_{ECL} resolution and discriminating power in this search (discussed in section 5.1). Photon reconstruction requires tracking information to discard clusters associated with charged hadrons. But the criterion is only effective for clusters within drift-chamber acceptance as outside drift-chamber coverage tracks are not reconstructed. I hence choose photon candidates within drift-chamber polar angle coverage, 17° to 150° .

The cluster-to-track distance distribution shows that misreconstructed photons contribute within 30 cm (fig. 5.8). To suppress them further, I therefore change the requirement to that threshold.

Cluster-event time difference effectively suppresses beam-background photons. However, its distribution is imperfectly modeled near collision time due to poor event-time modeling (fig. 5.21), leading to possible additional mismodeling. I therefore search for an alternative way to reject beam-background photons. Due to their origin from the interactions of individual beams, beam-background photons deposit energy predominantly in the endcap calorimeter sectors. I compare the simulated photon-energy distribution in each calorimeter sector in fig. 5.22 for the various sources. As expected, beam-background photons have large contributions backward, with energy deposits extending up to 200 MeV. I therefore suppress those with polar-dependent energy criteria. For all photon candidates, I require energy greater than 100 MeV in the forward endcap, greater than 55 MeV in the barrel sector, and greater than 150 MeV in the backward endcap.

These refined photon selections improve the E_{ECL} resolution, as shown in fig. 5.23. With the final photon selection, E_{ECL} is sufficiently well modeled in all control samples, except for the $B \rightarrow D$ sidebands, for which we add a dedicated systematic uncertainty (fig. 5.24). I therefore apply these photon selections in my analysis.

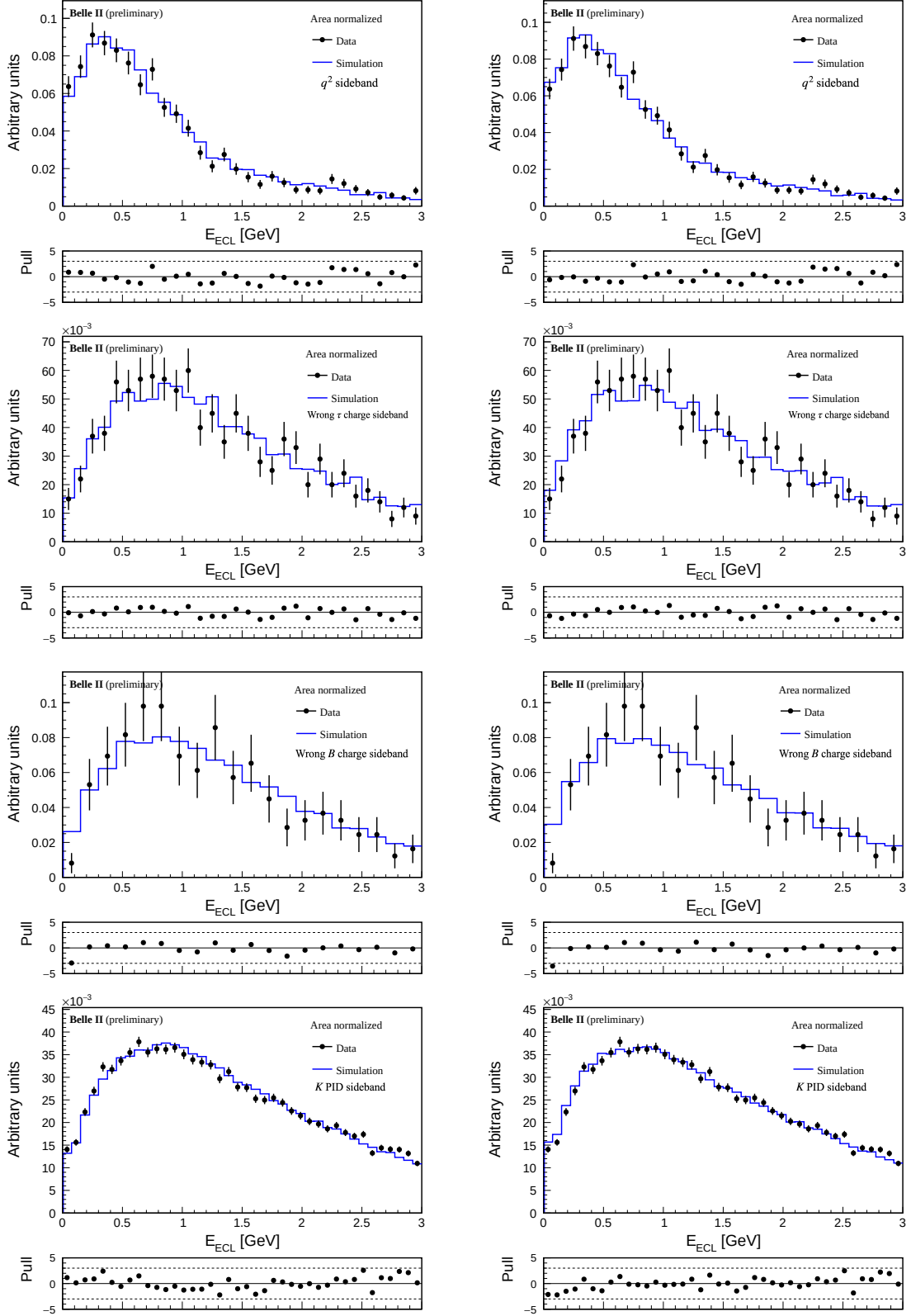


Figure 5.20: Distributions of E_{ECL} (left) without and (right) with collision-photon weights derived from the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ sample for (point) data and (histogram) simulation in the (first row) q^2 sideband, (second row) wrong τ charge sideband, (third row) wrong B charge sideband, and (fourth row) kaon PID sideband. Distributions are normalized to unity.

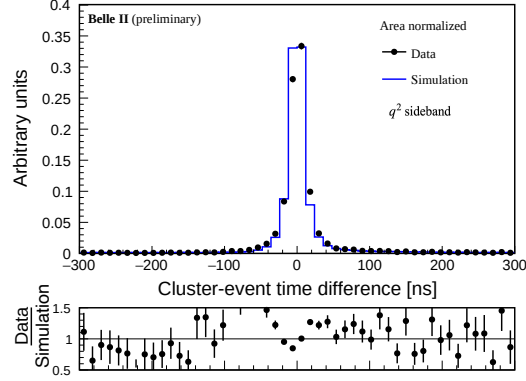


Figure 5.21: Distribution of cluster-event-time difference for photon candidates in (point) data and (histogram) simulated q^2 -sideband events. Distributions are normalized to unity.

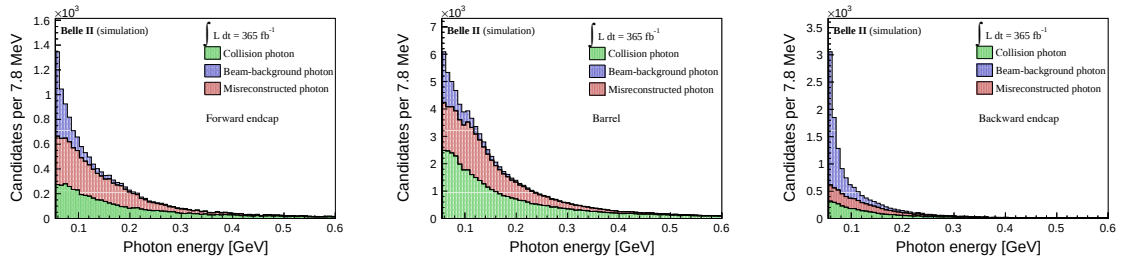


Figure 5.22: Energy distribution for simulated background photons in the (left) forward endcap, (middle) barrel, and (right) backward endcap regions of the calorimeter, divided by source.

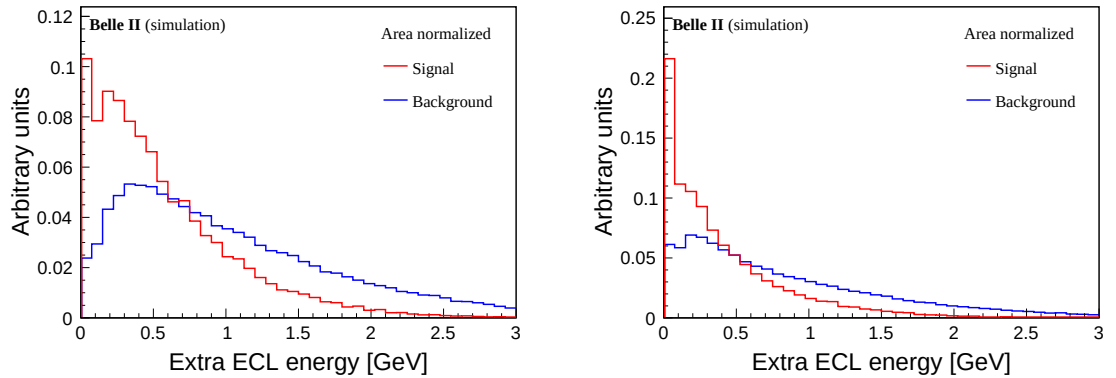


Figure 5.23: Distribution of E_{ECL} for simulated photons with (left) cluster-to-track distance greater than 20 cm and (right) final selection. Distributions are normalized to unity.

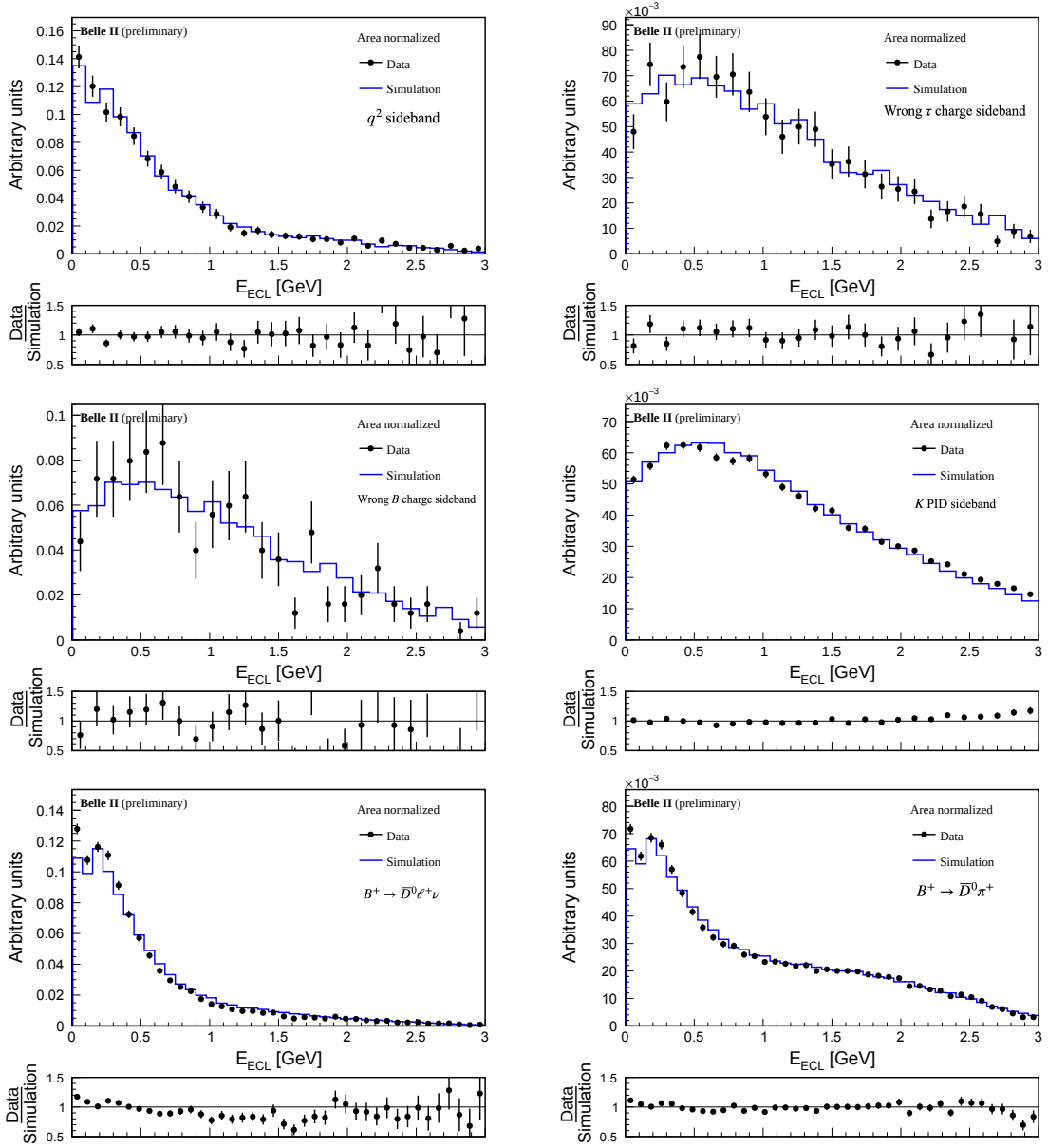


Figure 5.24: Distribution of E_{ECL} with final photon selection for (point) data and (histogram) simulation in the (top-left) q^2 sideband, (top-right) wrong τ charge sideband, (middle-left) wrong B charge sideband, (middle-right) kaon PID sideband, (bottom-left) $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$, and (bottom-right) $B^+ \rightarrow \bar{D}^0 \pi^+$. Distributions are normalized to unity.

Chapter 6

Selection optimization and sensitivity

This chapter describes the optimization of the selection and the statistical sensitivity.

6.1 Introduction

In the previous chapter, I outlined the baseline selections and discussed key discriminating observables aimed at suppressing background. Optimizing the selection criteria for these observables may be effective to enhance the search sensitivity. This optimization is performed in simulated data, thus it is important to ensure that simulation describes experimental data accurately; otherwise, the optimization would lead to suboptimal choices. Several corrections are applied to make simulation mirror the data. Then, during optimization, I assess sensitivity by determining upper limits in realistically simulated background-only samples under various choices of selection criteria on the discriminating observables. The final selection is determined based on the criteria that yield the maximum statistical sensitivity.

6.2 Realistic simulated sample

Before any optimization, I ensure that the simulated samples are reliable descriptions of data. Otherwise, data-simulation mismodeling may lead to a suboptimal result. I am particularly concerned about two kinds of mismodeling that may severely impact sensitivity, mismodeling of signal efficiency and misestimation of backgrounds. To address these, I correct the simulated signal efficiency to mirror data and estimate the expected background in the signal region of E_{ECL} by studying several control samples in data.

6.2.1 Simulated signal-efficiency calibration

FEI calibration The efficiency for B_{tag} reconstruction with hadronic FEI differs between data and simulation. This discrepancy arises from mismodeling in the simulation, primarily due to limited knowledge of hadronic B decays, 40% of which are not measured. Control data samples are used to quantify this discrepancy, where one B meson is reconstructed into a hadronic final state via FEI, and only a charged pion is reconstructed among the other particles, assuming it originates from the partner B . Then, one looks at the invariant mass of all other particles produced except for the pion (recoil mass), and compares it between data and simulation. This approach uses the fact that the recoil mass shows prominent peaks of D^0 and D^{*0} mesons with low background, due to the large $b \rightarrow c$ width. The D^0 and D^{*0} yields offer

a precise data-driven metric to determine data-simulation mismatches in efficiency. The signal yields are extracted from a likelihood fit to the recoil mass distribution, as shown in fig. 6.1, and compared with simulation. Any discrepancy is attributed to a FEI efficiency mismatch, given that $B^+ \rightarrow \bar{D}^{(*)0} \pi^+$ branching fractions are well known. Table 6.1 lists data-simulation efficiency ratios for each B_{tag} decay mode. These are used to determine FEI efficiency calibration factors. These factors can be averaged across all available calibration channels for maximum statistical power, at the expense of systematic uncertainties associated with tag-composition differences with respect to the signal sample. Or they can be determined using only the relevant channels that contribute to the signal tag composition, in appropriate proportions. In this analysis we only use correction factors derived from $B^+ \rightarrow \bar{D}^0 \pi^+$ as D^0 has higher signal purity than D^{*0} . For simplicity, I use the averaged correction factor of 0.71 ± 0.03 in the optimization of the selection. However, I use the appropriately weighted sample-specific factor 0.75 ± 0.05 for the final signal efficiency.

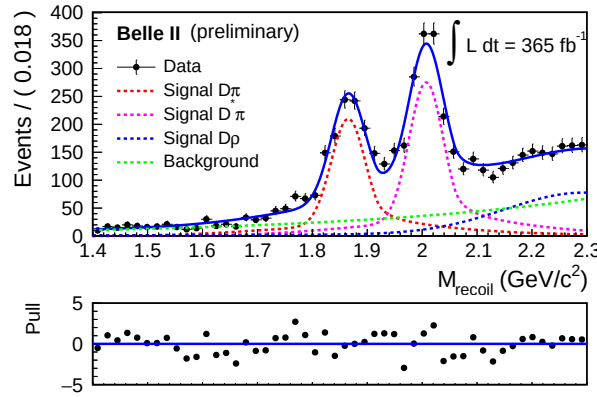


Figure 6.1: Distribution of recoil mass for $B^+ \rightarrow \bar{D}^{(*)0} \pi^+$ events reconstructed by FEI in Belle II data, with fit projections overlaid. The bottom panel shows the distribution of the difference between data and fit divided by the fit uncertainty.

π^0 veto efficiency A π^0 veto is applied to reject events containing a π^0 not associated with $\Upsilon(4S)$ reconstruction. This requirement may introduce differences in signal efficiency between data and simulation, which depend on the accuracy of simulation in properly describing π^0 production, reconstruction, and association. To address this, I correct the signal efficiency using the data sideband $1.8 < m(K^+ t^-) < 1.9$ GeV/c^2 , which contains predominantly $B^+ \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^-) \ell^+ \nu$ decays. The selection of this control sample is the signal baseline selection with an additional inverted criterion on $m(K^+ t^-)$. The efficiency is therefore determined by studying the variation in $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$ yield due to the π^0 veto and assuming that holds for the $B^+ \rightarrow K^+ \tau^+ \tau^-$ events as well. The correction is the ratio of the resulting efficiencies between data and simulation. This approach is approximately insensitive to data-simulation discrepancies, such as differences in selection criteria or branching fractions that could otherwise bias the desired efficiency.

I determine the signal yields by performing fits of the unbinned $m(K^+ t^-)$ distributions for two independent sets of events simultaneously, those that pass the π^0 veto selection, N_P , and those that fail it, N_F . The signal efficiency for the π^0 -veto is

B_{tag} decay mode	FEI efficiency correction	Fraction (%) in $B^+ \rightarrow \bar{D}^0 \pi^+$	Fraction (%) in $B^+ \rightarrow K^+ \tau^+ \tau^-$
$B^+ \rightarrow \bar{D}^0 \pi^+$	1.09 ± 0.05	8.8 ± 0.0	9.5 ± 1.1
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0$	0.73 ± 0.04	15.5 ± 0.1	19.4 ± 1.6
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^- \pi^+$	0.67 ± 0.03	17.0 ± 0.1	12.8 ± 1.2
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^- \pi^+ \pi^0$	0.45 ± 0.03	13.3 ± 0.1	11.9 ± 1.2
$B^+ \rightarrow \bar{D}^{*0} \pi^+$	1.05 ± 0.07	5.9 ± 0.0	9.6 ± 1.1
$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0$	1.09 ± 0.07	5.3 ± 0.0	5.4 ± 0.8
$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^- \pi^+$	0.76 ± 0.05	7.4 ± 0.0	7.0 ± 0.9
$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^- \pi^+ \pi^0$	0.62 ± 0.08	3.2 ± 0.0	2.6 ± 0.5
$B^+ \rightarrow D^- \pi^+ \pi^+$	0.71 ± 0.10	1.7 ± 0.0	2.1 ± 0.5
$B^+ \rightarrow D^- \pi^+ \pi^+ \pi^0$	0.49 ± 0.08	2.8 ± 0.0	2.7 ± 0.5
$B^+ \rightarrow \Lambda_c p^+ \pi^+ \pi^- \pi^+$	0.22 ± 0.04	3.3 ± 0.0	2.8 ± 0.6
Rest of B_{tag} decay modes	0.68 ± 0.03	15.8 ± 0.1	14.3 ± 1.3

Table 6.1: Data-simulation corrections to the FEI efficiency as functions of the decay occurring in the B_{tag} side using $B^+ \rightarrow \bar{D}^0 \pi^+$ decays as calibration signal. A comparison between B_{tag} sample-composition in this calibration signal and our $B^+ \rightarrow K^+ \tau^+ \tau^-$ signal is also shown.

calculated as

$$\varepsilon_{\text{veto}} = \frac{N_P}{N_P + N_F}, \quad (6.1)$$

The signal is modeled using a combination of a Johnson function [74] and a Crystal Ball function [75], while a straight line describes the background. This fit is done consistently in data and simulation. Figure 6.2 shows the distributions with fit projections overlaid. The resulting π^0 veto efficiency correction is

$$r_{\text{veto}} \equiv \frac{\varepsilon_{\text{veto}}^{\text{data}}}{\varepsilon_{\text{veto}}^{\text{MC}}} = \frac{0.886 \pm 0.016}{0.835 \pm 0.006} = 1.06 \pm 0.02, \quad (6.2)$$

where the uncertainty is used as a systematic uncertainty.

PID selection Charged-particle identification performance may differ between data and simulation, affecting both signal efficiency and background estimates. Signal efficiency is affected because (i) the efficiency for properly reconstructing $K\ell\ell$ decays in data differs from simulation and (ii) the efficiency has to account for misidentified $B^+ \rightarrow K^+ \tau^+ (\rightarrow \ell^+) \tau^- (\rightarrow \pi^-)$ or even $B^+ \rightarrow K^+ \tau^+ (\rightarrow \pi^+) \tau^- (\rightarrow \pi^-)$ that get reconstructed as signal and populate the search region. To account for these discrepancies, corrections for each charged-particle type are centrally derived in Belle II using control samples in the two-dimensional space of momentum and polar angle, as PID depends on both of these quantities. For kaon PID corrections, $D^{*+} \rightarrow D^0 (\rightarrow K^- \pi^+) \pi^+$ decays are used. For lepton PID corrections, various control samples are employed depending on momentum. For momenta smaller than 1.5 GeV/c, $e^+ e^- \rightarrow e^+ e^- e^+ e^-$ events are used for electron PID and $e^+ e^- \rightarrow e^+ e^- \mu^+ \mu^-$ for muon PID. For momenta larger than 1.5 GeV/c, decays $J/\psi \rightarrow e^+ e^-$ and $J/\psi \rightarrow \mu^+ \mu^-$ are used for electron and muon PID corrections, respectively. For

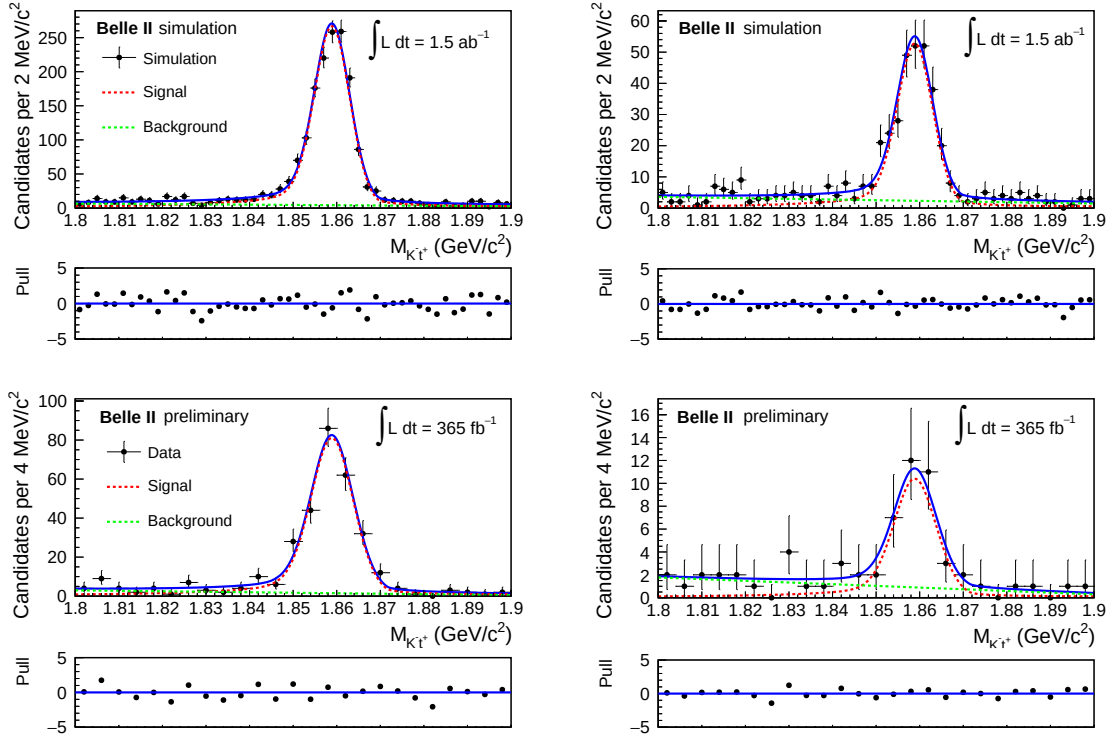


Figure 6.2: Distributions of $m(K^+t^-)$ for events (left) passing and (right) failing the π^0 veto in (top) realistic simulation and (bottom) data, with fit projections overlaid.

correction of the rate of pions misidentified as these charge particles are derived from $K_S^0 \rightarrow \pi^+\pi^-$ decays. Table 6.2 summarizes the resulting average corrections to the signal efficiency and background yield.

6.2.2 Simplified estimation of expected background

Accurate background estimation is critical for ensuring correct results in a counting experiment. For the purpose of selection optimization, I estimate the proper expected background under the signal by extrapolating event counts from the E_{ECL} and q^2 sidebands. Any data-simulation difference observed in these sidebands is extrapolated to the signal region to estimate the proper expected background. Since final selections on M_{miss}^2 and p_{t+}^* are not yet determined, I first examine the integrated rate difference between data and simulation without applying any restrictions on these observables. This approach enhances the statistical precision of the background estimate, but assumes that selections on M_{miss}^2 and p_{t+}^* would not introduce further data-simulation discrepancies in the final sample. This is reasonable for the purpose of selection optimization as any deviation from the assumption would lead, at most, to a suboptimal result, but not a biased one. Once the final selection criteria are set, I revisit this assumption and perform an updated estimation using the final selection.

I explore two reliable, independent data sidebands to estimate the rate difference between data and simulation, the E_{ECL} sideband ($E_{\text{ECL}} > 300 \text{ MeV}$) and the q^2 sideband ($q^2 < 14.18 \text{ GeV}^2/c^4$), for events without any restrictions on M_{miss}^2 and p_{t+}^* (fig. 6.3).

Charged particles	Efficiency corrections		Misidentification corrections	
	Signal	Background	Signal	Background
Kaon	0.976 ± 0.009	0.988 ± 0.016	0.606 ± 0.056	1.398 ± 0.044
Electron (same charge)	0.964 ± 0.002	0.942 ± 0.002	1.657 ± 0.058	1.194 ± 0.033
Electron (opposite charge)	0.986 ± 0.004	1.004 ± 0.005	0.983 ± 0.004	1.066 ± 0.027
Muon (same charge)	0.901 ± 0.008	0.891 ± 0.006	0.898 ± 0.028	0.981 ± 0.011
Muon (opposite charge)	0.947 ± 0.005	0.961 ± 0.007	0.782 ± 0.035	0.835 ± 0.015

Table 6.2: Corrections to the simulated particle-identification efficiency and misidentified pion rates for relevant charged particles. The adjective “same (opposite)” refers the concordance (discordance) of the charge with respect to the signal kaon charge.

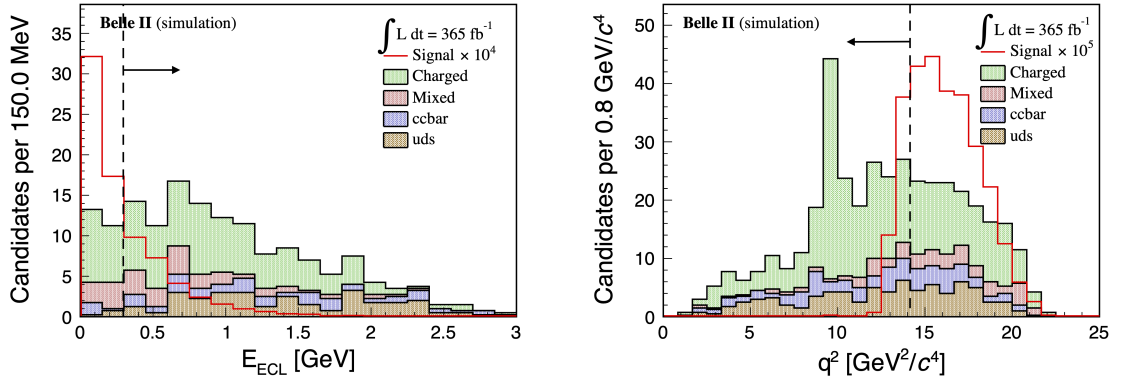


Figure 6.3: Illustration of sidebands used for estimating the background expectation in (left) E_{ECL} and (right) q^2 from a realistic simulated sample restricted to $m(K^+t^-) > 1.9 \text{ GeV}/c^2$. Distributions are normalized to data luminosity. Label ‘charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; ‘uds’ indicates $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds, all in simulation. The vertical black, dashed lines indicate the sideband boundaries with arrows pointing toward the sideband region.

- In the E_{ECL} sideband, the number of candidates in data is 98.0 ± 9.9 compared to 133.0 ± 5.8 in simulation. Assuming the same data-simulation yield ratio holds in the E_{ECL} signal region, the background observed in simulation is scaled by the ratio 0.74 ± 0.08 .
- In the q^2 sideband, the number of candidates in data is 166 ± 12.8 , while it is 247.3 ± 7.9 in simulation. The background scale-factor for this sideband is therefore 0.67 ± 0.06 .

The scale factors from both sidebands are consistent. I therefore use their weighted average 0.70 ± 0.05 in the optimization.

6.3 Sensitivity estimation

For the purpose of optimization, we assume that our analysis has insufficient sensitivity to see any signal and therefore optimize for the best expected limit using background-only

simulated events. Expected upper limits on the branching fraction of the $B^+ \rightarrow K^+ \tau^+ \tau^-$ decay are set by counting the number of observed events N_{obs} in the E_{ECL} signal-search window in each of a large set of simulated samples that are selected with a variety of different configuration of criteria. Observed yields are then compared to the expected numbers of background events, N_{exp} , and limits on the numbers of signal events in the sample, μ_s , are determined. The upper limit at the chosen $\beta\%$ confidence level is such that the unknown true value of μ_s would be $\beta\%$ of times smaller than the limit were the experiments be repeated multiple times. Limits on the signal yields are then converted into limits on the $B^+ \rightarrow K^+ \tau^+ \tau^-$ branching fraction, using

$$\mathcal{B} = \frac{N_{\text{obs}} - N_{\text{exp}}}{2\epsilon_s f^{+-} N(B\bar{B})}, \quad (6.3)$$

in which ϵ_s is the signal efficiency, $f^{+-} = 0.5113^{+0.0073}_{-0.0108}$ is the $\Upsilon(4S)$ branching fraction into $B^+ B^-$ pairs [76], and $N(B\bar{B}) = 387.1 \times 10^6$ is the B meson yield in our sample.

For the purpose of assessing swiftly numerous upper limits in an optimization that scans many selection configuration, I use the implementation of the frequentist approach proposed by Refs. [77, 78]. The likelihood is a combination of two components:

1. Poisson likelihood for the observed events,
 $\mathcal{L}(\mu_s, N_{\text{exp}} | N_{\text{obs}}) = (\mu_s + N_{\text{exp}})^{N_{\text{obs}}} e^{-(\mu_s + N_{\text{exp}})} / N_{\text{obs}}!$, which represents the probability of observing a given number of events, N_{obs} , assuming that the true number of events is the sum of signal, μ_s , and background, N_{exp} , contributions.
2. Gaussian likelihood for the background estimate N_{exp} ,
 $\mathcal{L}(\overline{N_{\text{exp}}}, \sigma | N_{\text{exp}}) = \exp\left(-\frac{(N_{\text{exp}} - \overline{N_{\text{exp}}})^2}{2\sigma^2}\right) / \sqrt{2\pi}\sigma$, which expresses the probability to this approximation is clearly unreliable for low numbers of expected events, but I consider it sufficient for the purpose of optimization. The final result on data is based on a Feldman-Cousins likelihood-ratio ordering, which properly takes into account the background yield distribution.

The full likelihood used in the optimization is

$$\mathcal{L}(\mu_s, \overline{N_{\text{exp}}}, \sigma | N_{\text{obs}}, N_{\text{exp}}) = \frac{(\mu_s + N_{\text{exp}})^{N_{\text{obs}}} e^{-(\mu_s + N_{\text{exp}})}}{N_{\text{obs}}!} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(N_{\text{exp}} - \overline{N_{\text{exp}}})^2}{2\sigma^2}\right) \quad (6.4)$$

which is first maximized with respect to the noninteresting (nuisance) parameters $\overline{N_{\text{exp}}}$ and σ at every point in the signal-yield space, effectively eliminating the dependence on unknown nuisance parameters. The resulting (profile) likelihood depends solely on the signal parameter, μ_s and is used to set limits.

These are based on background-only samples ($N_{\text{obs}} = N_{\text{exp}}$) and are assessed at 90% confidence level (CL). The optimization consists of the following steps,

- I start from simulated background events meeting the baseline selections described in section 4.5 with the additional restriction in opposite-charge kaon-lepton mass.
- I define a grid of selections in the three-dimensional space of the M_{miss}^2 , p_{t+}^* , and E_{ECL} observables.

- I apply each selection and correct the resulting simulated sample for efficiency and background estimate as discussed in section 6.2.
- Using the resulting number of expected background events and signal efficiency, I estimate the expected limit.
- As results of the optimization may be susceptible to fluctuations due to the small size of samples, I divide each simulated sample used in the procedure into five equipopulated subsamples and quote the average of the resulting limits as sensitivity.

6.4 Signal selection optimization

I restrict the optimization to data in the region $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ to suppress residual backgrounds with selections optimized simultaneously in M_{miss}^2 , p_{t+}^* , and E_{ECL} . A three-dimensional scan is performed across these observables, assessing sensitivity at each test point in the observable space. The scan ranges are $[-2.0, 6.0] \text{ GeV}^2/c^4$ in M_{miss}^2 with $0.2 \text{ GeV}^2/c^4$ step-width; $[0.2, 1.0] \text{ GeV}/c$ in p_{t+}^* with $0.05 \text{ GeV}/c$ step-width and $[0.1, 1.0] \text{ GeV}$ in E_{ECL} with 0.05 GeV step-width. The resulting expected upper limits are shown in fig. 6.4 as functions of the tested selection criteria, which are encoded in a scalar “selection configuration” number.

The optimized selection criteria are $M_{\text{miss}}^2 > 1.6 \text{ GeV}/c^2$, $p_{t+}^* > 0.5 \text{ GeV}/c$, and $E_{\text{ECL}} < 250 \text{ MeV}$ (fig. 6.5). The FEI-efficiency correction factor based on the B_{tag} composition of my signal sample after applying these selections is 0.75 ± 0.05 . The resulting signal efficiency is 1.3×10^{-5} with a background count of 4. The truth-matched signal efficiency is 1.2×10^{-5} . The five-folded averaged expected 90% CL upper limit, assuming a 70% background normalization scaling from section 6.2.2, is 1.03×10^{-3} . Table 6.3 reports a summary of these results.

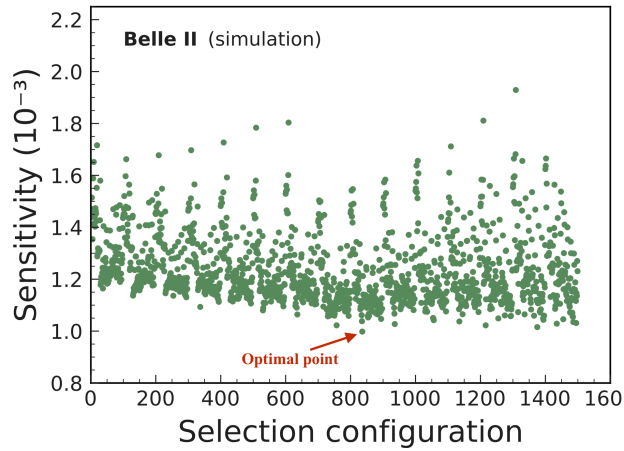


Figure 6.4: Expected upper limits from a realistically simulated background-only sample as functions of tested selection criteria, encoded one-dimensionally in the horizontal axis.

6.4.1 Final background composition

After applying all final selection criteria, including the results of the optimization, the background consists of 23 events in a simulated sample four times larger than the data.

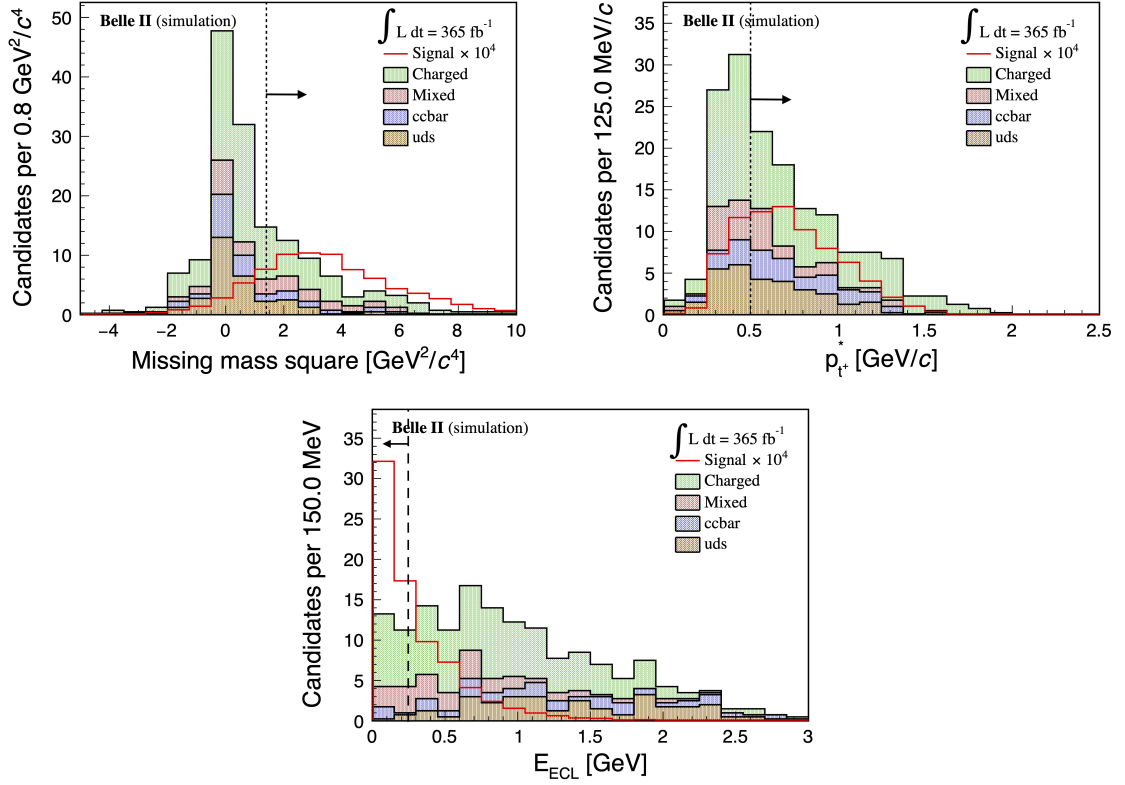


Figure 6.5: Distributions of (top-left) M_{miss}^2 , (top-right) p_{t+}^* and (bottom) E_{ECL} in a realistic simulated sample restricted to $m(K^+t^-) > 1.9 \text{ GeV}/c^2$. Distributions are normalized to data luminosity. The vertical black dashed lines indicate the optimized selections boundaries with arrows pointing toward the accepted signal regions. See previous plots for legend conventions.

Optimized selections	$M_{\text{miss}}^2 > 1.6 \text{ GeV}/c^2$ $p_{t+}^* > 0.5 \text{ GeV}/c$ $E_{\text{ECL}} < 250 \text{ MeV}$
Signal efficiency	1.3×10^{-5}
Background	4
Expected upper limit at 90% CL	1.03×10^{-3}

Table 6.3: Optimized selections and sensitivity in the realistic simulated sample.

Among them, 12 events (52%) are from B^+ decays, 7 events (30%) are B^0 mixed decays, and 4 events (17%) are from continuum. Table 6.4 provides further detail on the expected $B\bar{B}$ background composition in the signal-search region, showing main contributions from $B^0 \rightarrow D^{*-}\ell^+\nu$ decays (4 events) and $B^+ \rightarrow D_s^{(*)-}K^+\ell^+\nu$ decays (4 events) in which D_s^- predominantly decays into $\ell^-\bar{\nu}$ (3 events).

Background	Fraction (%)
$B^+ \rightarrow D_s^{(*)-} K^+ \ell^+ \nu$	17.4 ± 9.4
$B^0 \rightarrow D^{*-} \ell^+ \nu$	17.4 ± 9.4
$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	8.7 ± 6.4
$B^0 \rightarrow D^- \ell^+ \nu$	4.3 ± 4.4
Rest of $B\bar{B}$ background	34.8 ± 14.3
Continuum background	17.4 ± 9.4

Table 6.4: Composition of simulated background after final selection.

6.5 Impact of baseline selection on sensitivity

The baseline selection introduced in chapter 4 is chosen without a systematic optimization. I therefore revisit some of the choices to explore if further sensitivity could be available through those. The baseline selection includes particle-identification criteria for charged particles, photon criteria relevant for the E_{ECL} definition, and a π^0 veto. In these studies, all corrections to simulation discussed in previous sections are applied except for PID and π^0 -veto efficiency corrections, which have a marginal impact for the purpose of these studies. In addition, the signal efficiency is corrected using the averaged FEI efficiency calibration factor, 0.71. For a consistent comparison, the same simplifications are applied in the default analysis.

6.5.1 PID selection

Charged particles are identified through PID selection criteria. Modifying the PID selection alters the background composition and impacts sensitivity. The optimal E_{ECL} signal-search region may change as a result. I therefore perform a two-dimensional scan in PID of each charged particle and E_{ECL} signal window after applying all the selections.

Kaon binary PID The scan ranges are $[0.52, 1.00]$ in kaon binary PID with 0.06 step-width and $[0.1, 0.6]$ GeV in E_{ECL} with 0.05 GeV step-width. Figure 6.6 shows the sensitivity at each point. At any given choice of E_{ECL} signal window, kaon binary PID has a marginal impact on sensitivity. The optimal selection is found at a kaon binary PID greater than 0.75.

Electron global PID The scan ranges are $[0.7, 1.0]$ in electron global PID with 0.03 step-width and $[0.1, 0.6]$ GeV in E_{ECL} with 0.05 GeV step-width. Figure 6.7 shows the sensitivity at each point. At any given choice of E_{ECL} signal window, kaon binary PID has a marginal impact on sensitivity. Optimization of electron global PID has a negligible impact on the sensitivity, and I select it to be greater than 0.9, where the sensitivity is maximum.

Muon PID Similarly, I investigate the sensitivity dependence of muon global PID by varying it in the range $[0.7, 1.0]$ with 0.03 step-width along with the E_{ECL} window in the range $[0.1, 0.6]$ GeV with 0.05 GeV E_{ECL} step-width. Figure 6.8 shows the expected sensitivity at each point. The muon global PID has a negligible impact on the sensitivity, and I select it to be greater than 0.9.

The optimization above is performed by testing the PID selections of individual particles separately. Ideally, PID selections should be optimized simultaneously for all relevant

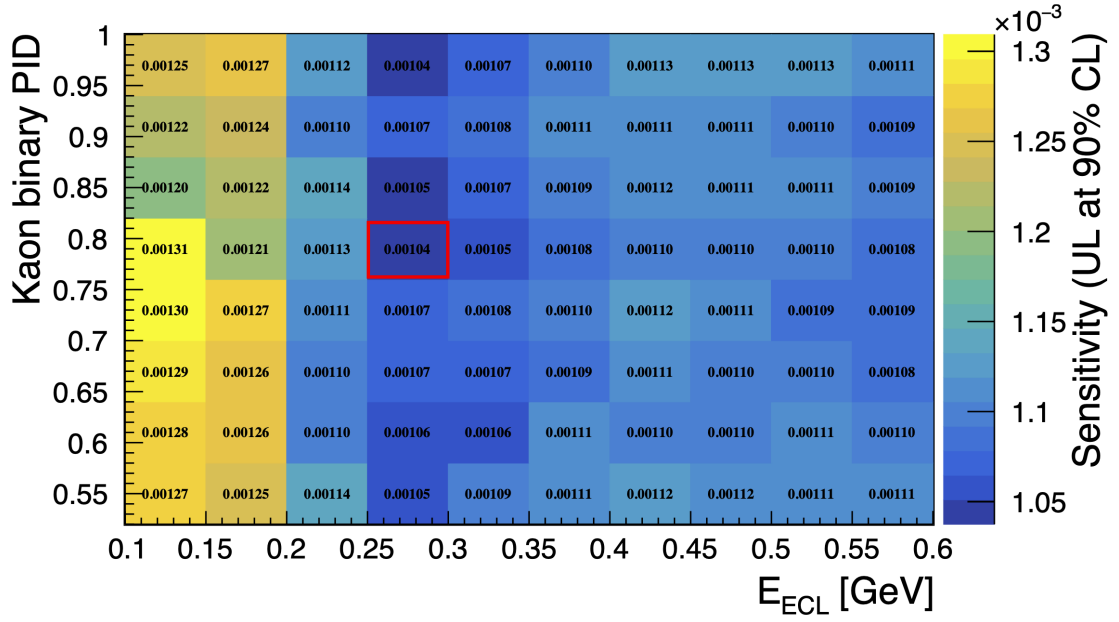


Figure 6.6: Sensitivity as a function of kaon binary PID and E_{ECL} signal-window selections in simulation. The red box shows the optimal selection chosen for the default analysis.

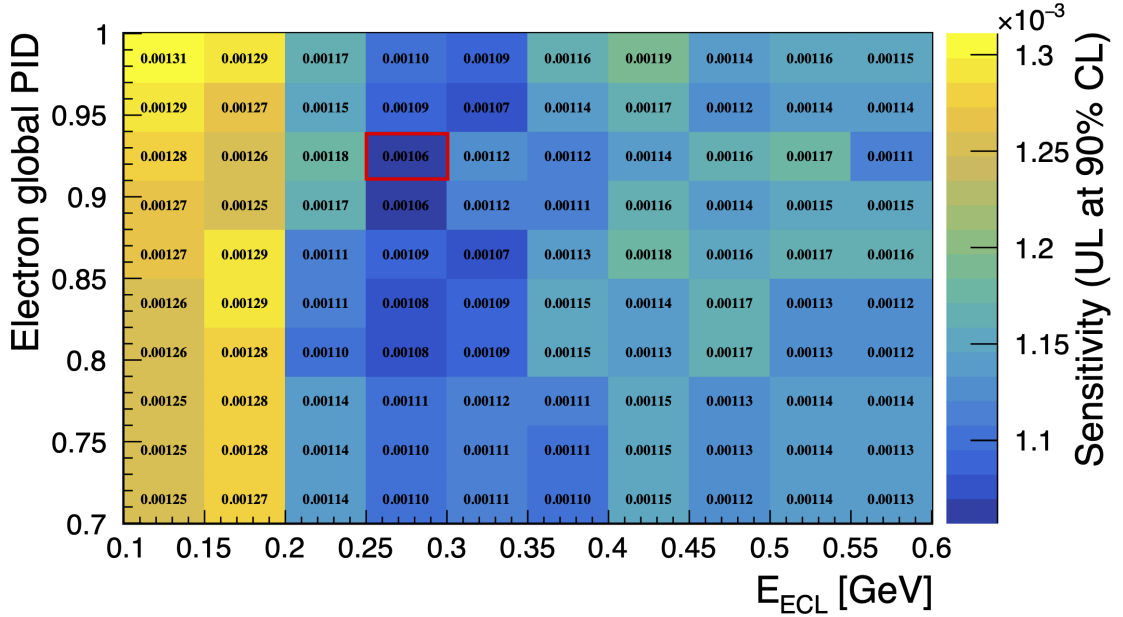


Figure 6.7: Sensitivity as a function of electron global PID and E_{ECL} signal-window in simulation. The red box shows the optimal selection chosen for the default analysis.

particles to account for potential correlations. However, I do not pursue this approach since the individual optimizations show that PID has a minimal impact on the expected sensitivity.

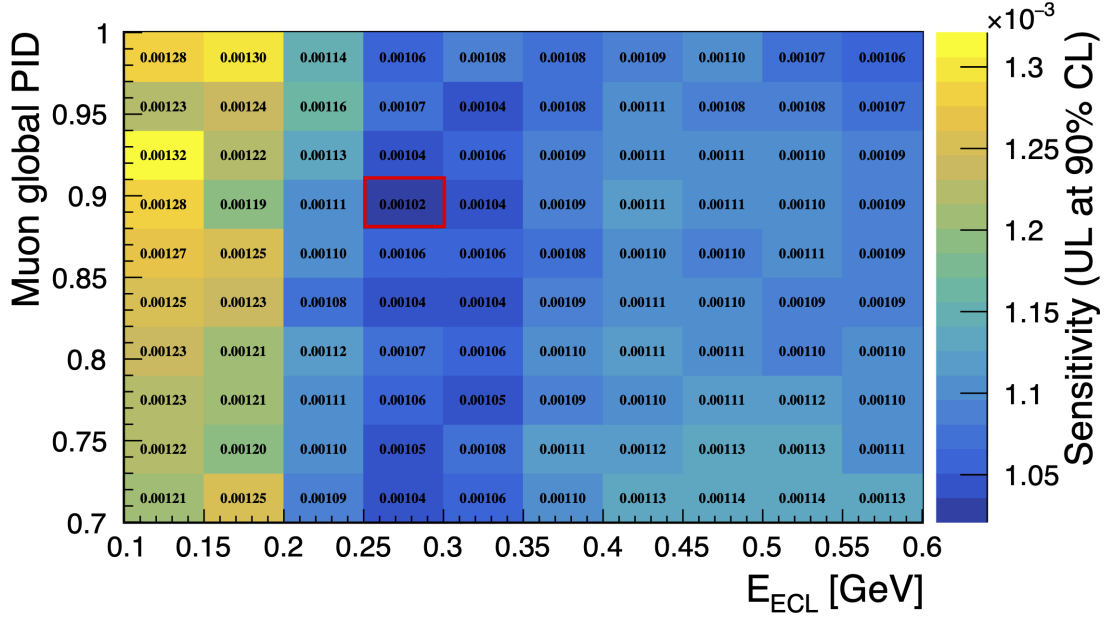


Figure 6.8: Sensitivity as a function of muon global PID and E_{ECL} signal-window selections in simulation. The red box shows the optimal selection chosen for the default analysis.

6.5.2 Photon selections

Photon selection is relevant in this analysis. Even though they do not directly impact signal efficiency, they bias the E_{ECL} shape. Loose constraints on photon energy degrade the resolution of the E_{ECL} signal, as they allow more beam-induced background and misreconstructed photons into the sample. Stringent constraints on the cluster-to-track distance alter the E_{ECL} distributions, reducing the distinction between signal and background and lowering the sensitivity. I therefore evaluate how my photon selections affect sensitivity using simulation adjusted to mirror data after reoptimizing the selection criteria.

I first study how the custom energy thresholds used in the default analysis — 55 MeV in the central calorimeter, 100 MeV in the forward calorimeter, and 150 MeV in the backward calorimeter — impact sensitivity. The results in table 6.5 show that sensitivity does not worsen with respect to the Belle II standard criteria of photon energy greater than 55 MeV for all polar angles. Hence, I retain these stringent selections as a further protection in the case that beam backgrounds in the signal region are larger than anticipated from the sidebands. Similarly, misreconstructed photons originating from tracks are suppressed restricting cluster-to-track distance. I investigate the impact on sensitivity of varying the minimum cluster-to-track distance in the range [20, 40] cm. Figure 6.9 provides a comparison of the resulting expected upper limits. Since no significant improvement is observed in any looser configuration, I retain the selection of distance greater than 30 cm.

6.5.3 π^0 veto

The π^0 veto selection is implemented to suppress background events containing misreconstructed π^0 s, which contribute noise to the residual energy in the calorimeter. After reoptimizing the selection criteria in each case, I evaluate the impact of the π^0 veto on sensitivity by comparing the expected sensitivity with and without the veto. As shown in table 6.6, rejecting events with a π^0 final state improves the sensitivity. Therefore, I decide

Optimized selection	This search photon selection	Photon baseline selection
M_{miss}^2	$> 1.6 \text{ GeV}^2/c^4$	$> 1.4 \text{ GeV}^2/c^4$
p_{t+}^*	$> 0.5 \text{ GeV}/c$	$> 0.35 \text{ GeV}/c$
E_{ECL}	$< 250 \text{ MeV}$	$< 450 \text{ MeV}$
Efficiency	1.4×10^{-5}	2.1×10^{-4}
Background	5	12
Expected upper limit	1.0×10^{-3}	1.1×10^{-3}

Table 6.5: Sensitivity for an alternative selection of photons that enter the E_{ECL} definition based on simulation adjusted to mirror data.

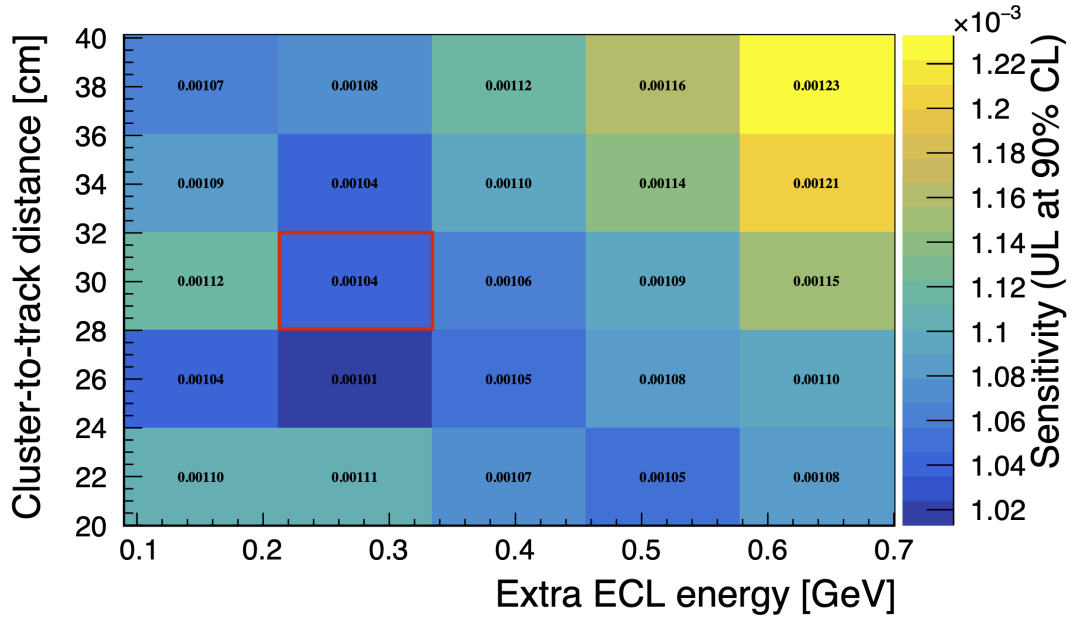


Figure 6.9: Sensitivity as a function of minimum cluster-to-track distance and E_{ECL} signal range in simulation adjusted to mirror data. The red box shows the optimal selection chosen for the default analysis.

to apply the π^0 veto in the analysis.

Optimized selection	With π^0 veto	Without π^0 veto
M_{miss}^2	$> 1.6 \text{ GeV}^2/c^4$	$> 1.4 \text{ GeV}^2/c^4$
p_{t+}^*	$> 0.5 \text{ GeV}/c$	$> 0.65 \text{ GeV}/c$
E_{ECL}	$< 250 \text{ MeV}$	$< 250 \text{ MeV}$
Efficiency	1.4×10^{-5}	1.02×10^{-5}
Background	5	3
Expected upper limit	1.00×10^{-3}	1.20×10^{-3}

Table 6.6: Sensitivity with an alternative veto selection in a realistic simulated sample.

Chapter 7

Final validation and background estimate

This brief chapter describes the validation of the signal-extraction observable specialized to the final selection used in the search and the full-fledged estimation of background expected in the signal-search region.

7.1 E_{ECL} validation for the final sample composition

At this stage, E_{ECL} is validated using the final photon selection, but for events selected according to loose selections on all other observables, as criteria $m(K^+t^-) > 1.9 \text{ GeV}/c^2$, $M_{\text{miss}}^2 > 1.6 \text{ GeV}^2/c^4$, and $p_{t+}^* > 0.5 \text{ GeV}/c$, and the π^0 veto are not applied. It is important to validate it for the final sample composition for this search and to account properly for both rate and shape mismodelings. After applying the final selection, the control sample size reduces considerably. In fact, most control samples introduced in section 5.2 lack sufficient event yields for proper validation, which can only rely on the $E_{\text{ECL}} > 300 \text{ MeV}$ sideband and the $q^2 < 14.18 \text{ GeV}^2/c^4$ sideband. With small sample sizes, detecting statistically significant discrepancies becomes difficult. I therefore perform an intermediate consistency check after lifting the M_{miss}^2 and p_{t+}^* selections to increase sample size and check for discrepancies before the final selection stage. Figure 7.1 shows a test of shape differences, in which data and simulation are normalized to each other both after the intermediate and the final selections. The E_{ECL} shape is sufficiently well modeled in both sidebands, at both the intermediate and final selections stages. Figure 7.2 shows the same test for rates. Here an indication of a rate difference is observed, even though the data set size is too limited for deriving conclusive statistical information. However, E_{ECL} can be safely used for signal extraction as the observed difference is corrected by the estimation of the background expectation described in the next section.

7.2 Final background estimation in the signal region

Estimating reliably the background expected in the signal-search region is a crucial element of this analysis. The strategy chosen consists of using the data observed in the q^2 and E_{ECL} sidebands to make an extrapolation into the signal-search region. The shape of simulated events scaled to mirror data in the sidebands is used in the extrapolation. Uncertainties in rates and shapes are accounted for in the systematic uncertainties.

As discussed, the E_{ECL} shape is validated across E_{ECL} and q^2 control samples selected

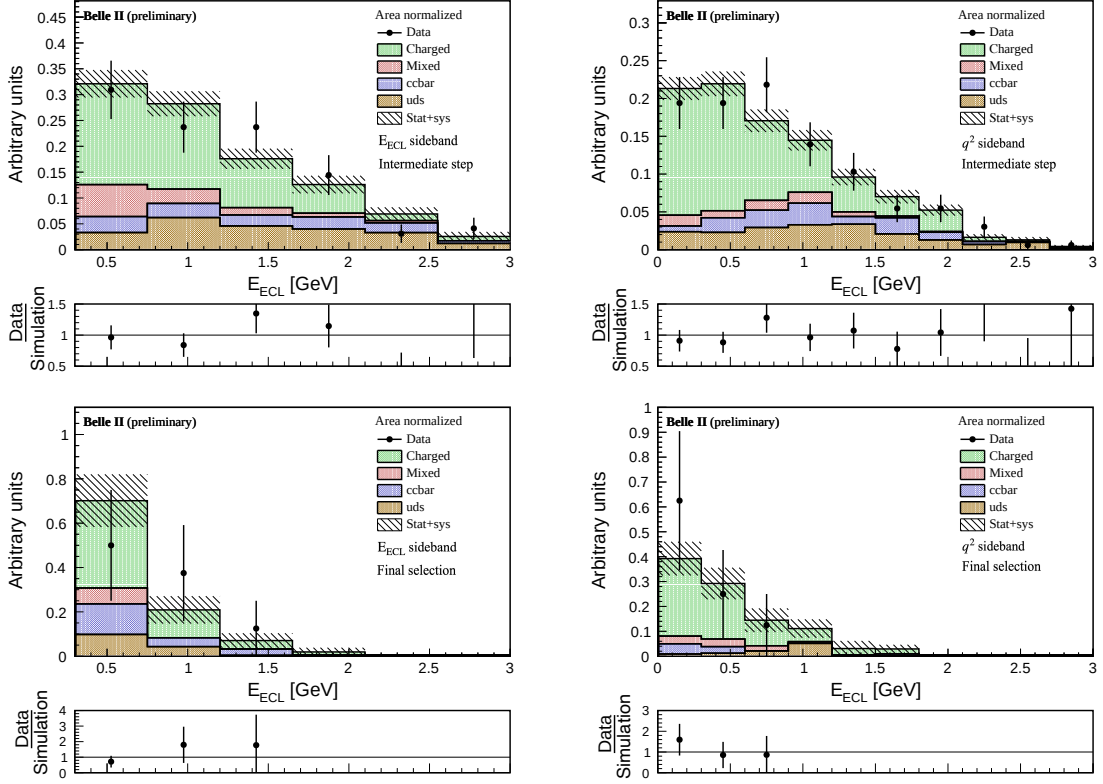


Figure 7.1: Distributions of E_{ECL} shape for (point) data and (histogram) simulation in the (top-left) E_{ECL} sideband without M_{miss}^2 and p_{t+}^* restrictions, (top-right) q^2 sideband without M_{miss}^2 and p_{t+}^* restrictions, (bottom-left) E_{ECL} sideband with final selection, and (bottom-right) q^2 sideband with final selection. Distributions are normalized to unity. Label ‘charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; ‘uds’ indicates $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds.

with all final selection criteria. However, this is not sufficient to demonstrate that E_{ECL} is under control in the signal region, if a comparison of sample composition between signal-search region and sidebands does not show general consistency. After the final selection, the E_{ECL} sideband contains 55 events simulated in a sample four-times larger than the collision data. Of these 32 (58%) are from charged B decays, 4 (7%) from neutral B decays, and 19 (35%) are from continuum. The q^2 sideband contains 85 events, of which 62 (73%) are from charged B decays, 7 (8%) are from neutral B decays, and 16 (19%) are continuum. Table 7.1 and 7.2 show more details and should be compared with table 7.3, which shows the expected background composition in the signal-search region. While differences in the detailed composition between signal-search region and sidebands are observed, the general consistency reassures that our background estimation procedure is sound. This is achieved as follows

Sample binning Events populating the E_{ECL} and q^2 sidebands after applying all final selection criteria are divided into bins.

Fit The data-to-simulation ratios from both E_{ECL} and q^2 sidebands are fit simultaneously assuming a multiplicative scaling.

Background estimation The expected background yield in the signal region is then

Background	Fraction (%)
$B^+ \rightarrow D_s^{(*)-} K^+ \ell^+ \nu$	16.5 ± 4.8
$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	11.8 ± 3.9
$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	8.2 ± 1.3
$B^+ \rightarrow \bar{D}_2^0 \ell^+ \nu$	4.7 ± 2.4
$B^+ \rightarrow \bar{D}_2^0 \rho^+$	3.5 ± 2.1
Other $B\bar{B}$	36.7 ± 7.7
Continuum	18.6 ± 5.1

 Table 7.1: Composition of simulated background in the q^2 sideband after the final selection.

Background	Fraction (%)
$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	12.7 ± 5.1
$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	7.3 ± 3.8
$B^+ \rightarrow D_s^{(*)-} K^+ \ell^+ \nu$	5.5 ± 0.3
$B^+ \rightarrow \psi(3770) K^+$	5.5 ± 0.3
$B^+ \rightarrow \bar{D}^0 D_s^{*+}$	3.6 ± 2.6
Other $B\bar{B}$	30.9 ± 8.6
Continuum	34.5 ± 9.1

 Table 7.2: Composition of simulated background in the E_{ECL} sideband after the final selection.

Background	Fraction (%)
$B^+ \rightarrow D_s^{(*)-} K^+ \ell^+ \nu$	17.4 ± 9.4
$B^0 \rightarrow D^{*-} \ell^+ \nu$	17.4 ± 9.4
$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	8.7 ± 6.4
$B^0 \rightarrow D^- \ell^+ \nu$	4.3 ± 4.4
Other $B\bar{B}$	34.8 ± 14.3
Continuum	17.4 ± 9.4

Table 7.3: Composition of a simulated background sample after the final selection.

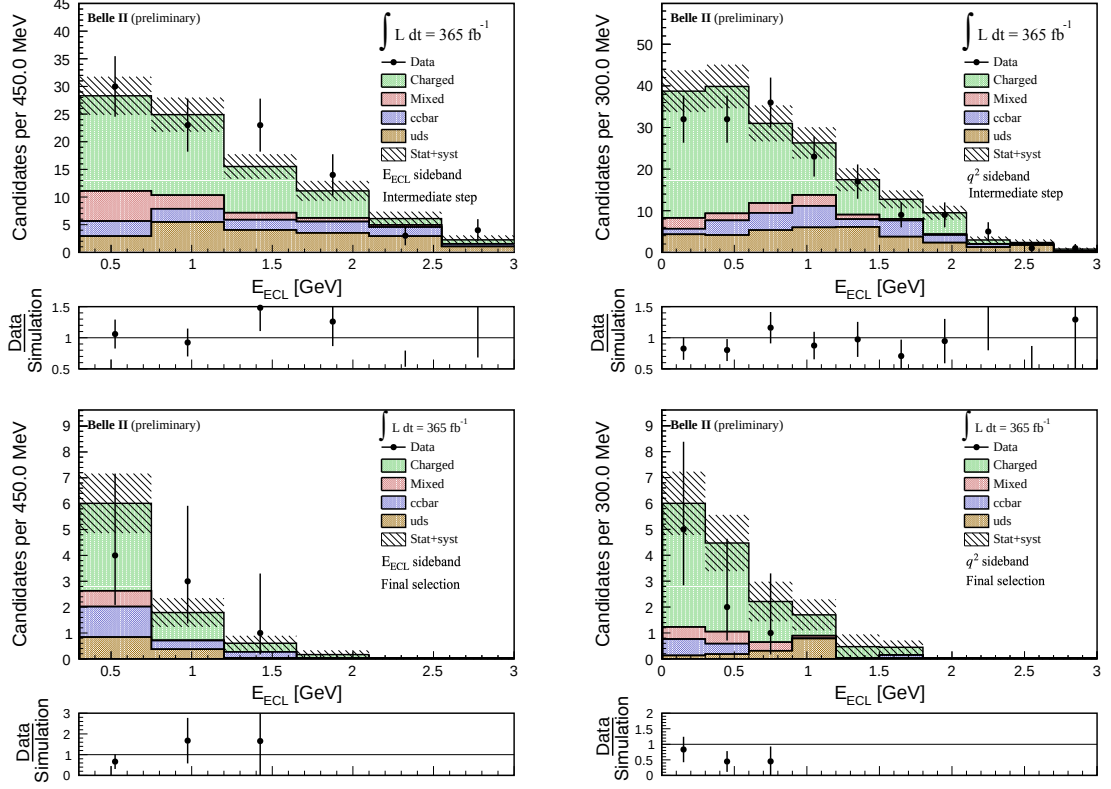


Figure 7.2: Distributions of E_{ECL} for (point) data and (histogram) simulation in the (top-left) E_{ECL} sideband without M_{miss}^2 and p_{t+}^* restrictions, (top-right) q^2 sideband without M_{miss}^2 and p_{t+}^* restrictions, (bottom-left) E_{ECL} sideband with final selection, and (bottom-right) q^2 sideband with final selection. Simulated distributions are normalized to data luminosity. See previous plots for legend conventions.

derived from the extrapolation of the resulting fit function from sidebands into the signal region. Alternative data-to-simulation scalings are used to assess systematic uncertainties.

After the final selection, the sidebands contain only eight data events each, divided into three bins. I perform a maximum likelihood fit to the binned data-to-simulation ratios simultaneously in both sidebands using a Poisson model where simulation and data are related by a constant scale $f(m_i, p) = m_i p$. The likelihood is

$$\mathcal{L}(p|d_i) = \prod_{i=1}^{\text{bin}} \mathcal{P}(d_i|f(m_i, p)) = f(m_i, p)^{d_i} e^{-f(m_i, p)} / d_i!,$$

where $\mathcal{P}$ indicates the single-bin probability-density function, p is the scale factor, m_i is the simulation yield in bin i , and d_i is the data yield in that bin. I repeat the fit 10^5 times by sampling simulated and experimental data in each bin according to a Poisson to account for statistical uncertainties. Figure 7.3 illustrates the fit projection of one such fit. I consider the best fit result in data as the central value of the scale factor and the mean of the distribution of the uncertainty as the uncertainty of the scale factor (fig. 7.4).

The fit result in our sample is $p = 0.55_{-0.13}^{+0.15}$. Extrapolating this to the signal region, I

obtain an expected background yield of $2.95 \pm 0.61^{+0.80}_{-0.70}$ events, where the first uncertainty is due to the simulated sample-size in the signal region and the second is due to the scale-factor uncertainty. Scale factors independently obtained from the individual sidebands are consistent with each other, at $0.66^{+0.26}_{-0.21}$ from the E_{ECL} sideband and $0.47^{+0.19}_{-0.15}$ from the q^2 sideband, supporting our choice of a fit simultaneous on both sidebands. An additional (systematic) uncertainty comes from the assumption of constant scaling. I assess this by assuming alternative models for the data-to-simulation ratio as a function of E_{ECL} . Table 7.4 provides a comparison of the resulting likelihood ratios, in terms of $-2 \ln \mathcal{L}$ showing that data favor our default model of constant scaling. The largest difference, resulting from many simplified simulated experiments, in best-fit results obtained with the default model, a linear model yielding 2.78 events, and a quadratic model yielding 3.22 events; provides an estimate of the model-dependence uncertainty, which is 0.27 events. This is summed in quadrature with the uncertainty on the scaling to yield the final expected background yield of $2.95 \pm 0.61^{+0.85}_{-0.75} = 2.95^{+1.05}_{-0.97}$. Further studies based on simplified simulated experiments show that these estimates might be affected by a positive bias amounting to 6% of the scale-factor (statistical) uncertainty, as shown in fig. 7.4. This is negligible with respect to the size of the other background-extrapolation systematic uncertainty and ignored.

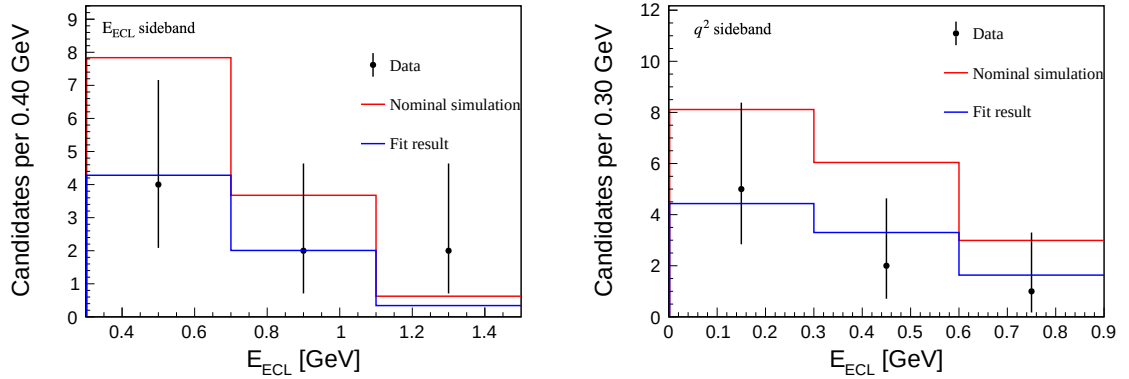


Figure 7.3: Distributions of E_{ECL} for (point) data, (histogram) simulation (blue) without and (red) with the scale factor applied in (left) E_{ECL} and (right) q^2 sidebands.

Function	$\mathcal{L}$	$-2 \ln(\mathcal{L})$
mp	2.35×10^{-5}	21.31
$m(p_0 + E_{\text{ECL}} \cdot p_1)$	2.42×10^{-5}	21.26
$m(p_0 + E_{\text{ECL}} \cdot p_1 + (E_{\text{ECL}})^2 \cdot p_2)$	3.56×10^{-5}	20.49

Table 7.4: Comparison of fit properties with various fit-functions in a single simplified experiment.

The large uncertainty in the background estimate renders this source a prominent contributor to the systematic uncertainty. Such background-estimate uncertainty is driven by the scarce population of the sidebands used for extrapolation, which in turn results from the optimization of the statistical precision of the analysis. By construction, an optimization targeted at providing the best expected limit in background-only samples, tends to suppress background as much as possible, thus depleting the sidebands as well. I therefore investigate if alternative selection configurations (see figure 6.4), although slightly

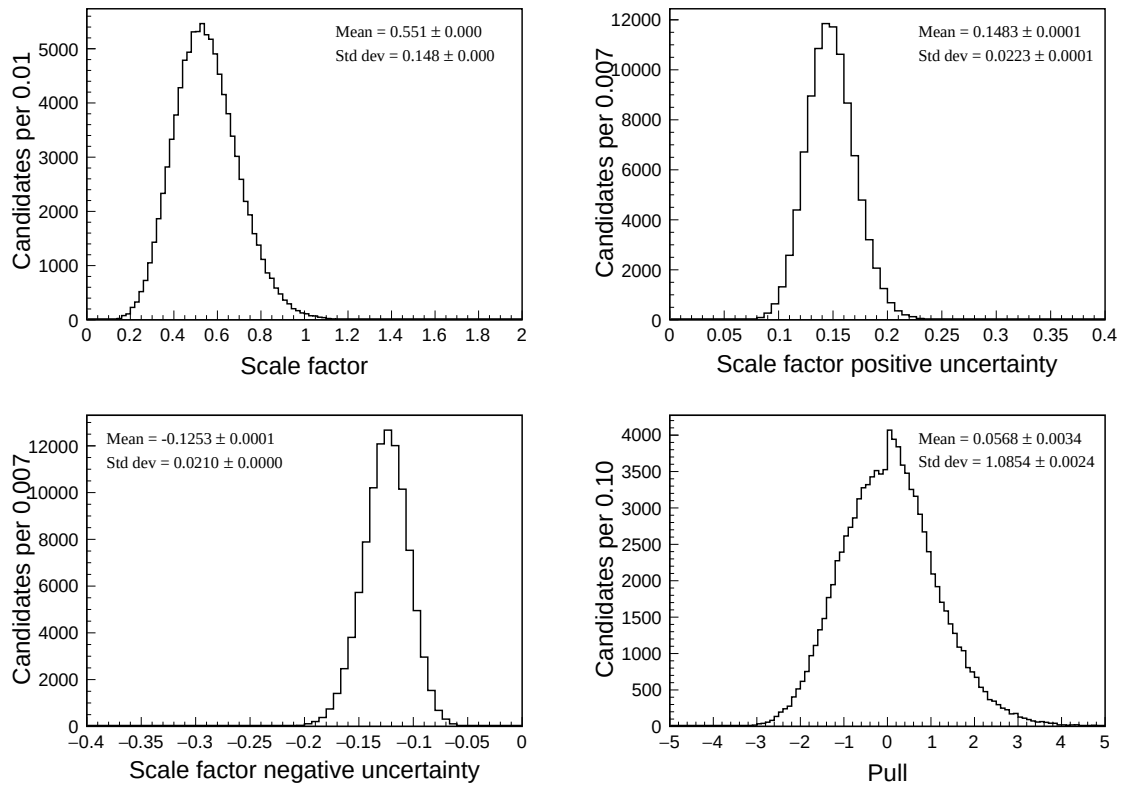


Figure 7.4: Distributions of fit results for (top-left) scale factor, (top-right) positive uncertainty, (bottom-left) negative uncertainty, and (bottom-right) pull in 10^5 simplified simulated experiments.

suboptimal from the purely statistical point of view, could achieve better total sensitivity by reducing background rejection and therefore reducing the systematic uncertainty due to the background-estimate. I repeat the background estimate for a number of alternative test selections, but fail to find any that provides an improved combined statistical and background-extrapolation uncertainty. The analysis therefore adopts the default optimized selection.

Chapter 8

Systematic uncertainties

This chapter describes the determination of the systematic uncertainties.

The branching fraction is obtained as

$$\mathcal{B} = \frac{N_{\text{obs}} - N_{\text{exp}}}{2\epsilon_s f^{+-} N(B\bar{B})}, \quad (8.1)$$

where N_{obs} and N_{exp} are numbers of observed and expected events in the signal region, ϵ_s is the signal efficiency, $f^{+-} = 0.5113^{+0.0073}_{-0.0108}$ is the $\Upsilon(4S)$ branching fraction into B^+B^- pairs; and $N(B\bar{B}) = 387.1 \times 10^6$ is the B -meson yield in the sample. I consider a number of sources of systematic uncertainty that have the potential of affecting the result. These are associated with (i) uncertainties in the models assumed for signal and background yields, i.e., ϵ_s and N_{exp} , which are typically addressed either using control samples in data or estimating the impact of alternative choices using simplified simulated experiments and (ii) uncertainties in the external inputs, i.e., f^{+-} and $N(B\bar{B})$, which are straightforwardly included in the signal-yield estimate using standard linear “error propagation”. As I aim at procedures ensuring frequentist coverage, results are assessed assuming the worst possible configuration of nuisance parameters, that is, the one making the results more background-like.

8.1 Uncertainty in background estimate

The estimate of the background yield in the E_{ECL} signal region is the most critical ingredient of the signal extraction, which contributes an important additive systematic uncertainty. In the default analysis, I estimate this contribution by extrapolating into the signal region the background events observed in adjacent control regions, according to the shape of the distribution observed in simulation, as described in section 7.2. This is associated with two sources of uncertainty — the statistical uncertainty due to the finite populations of the samples used in the extrapolation, i.e., the data events in the sideband, the simulated events in the sideband, and the simulated events in the signal-search region; and the systematic uncertainty associated with the shape assumed for the background distribution. These sources all reflect as systematic uncertainties in our inference. The contributions from the former source is straightforwardly assessed through ensembles of simplified experiments that simulate Poisson fluctuations. The contribution from the latter depends on the variations in expected background-yield observed using alternative shapes for the extrapolation. To this end, I test alternative assumptions for the chosen extrapolation models using the E_{ECL} distributions observed in the E_{ECL} sideband and in the q^2 sidebands and

use the deviation in results from the default extrapolation as uncertainty. In addition, I include a systematic uncertainty meant to cover possible mismodeling specific to low E_{ECL} values as possibly suggested by mismodeling in $B \rightarrow D$ control samples discussed in chapter 5. If present, these would not be captured by the simple extrapolation models used and would bias the results. This systematic uncertainty is based on the 0.27 difference in the amount of expected background observed between applying, or not, the collision-photon-multiplicity weighting that fixes the discrepancies in the $B \rightarrow D$ samples discussed in chapter 5. While the composition and kinematic properties of our signal sample and the $B \rightarrow D$ samples differ, we assume that the difference induced in the results by a correction that fixes the largest discrepancy observed in any of the control samples is a sufficiently generous range to bracket any possible residual and unaccounted for mismodeling. This yields an additive uncertainty of $^{+1.08}_{-1.00}$ events on the signal yield from which the branching fraction is determined, and it is the major source of systematic uncertainty in the analysis.

8.2 Uncertainty in signal efficiency

Any uncertainty in signal efficiency contributes a multiplicative uncertainty to the results. Even though multiplicative uncertainties are usually less impactful than additive uncertainties when dealing with small or not significant signals, it is important to account for all relevant sources properly.

- Uncertainty in the FEI-tagging efficiency.** This is the uncertainty associated with the efficiency for reconstructing and selecting the pair-produced nonsignal B meson. Given the complexity of FEI reconstruction and the uncertainties in $B\bar{B}$ sample composition, one cannot rely on native simulation. Dedicated procedures of so-called “FEI calibration” are implemented centrally in the collaboration and consist of applying FEI to control samples that offer independent and unbiased estimates of the efficiency, as discussed in Sec. 6.2. These estimates are then used to scale the simulation-based FEI efficiency to match data. This procedure involves various uncertainties associated with sample dependences and other effects that propagate into this analysis as systematic uncertainties. The FEI-calibration factor applied in this analysis, based on equalizing the channel-by-channel B_{tag} composition between signal and calibration sample, is 0.75 ± 0.05 . The statistical uncertainty on this correction reflects a systematic uncertainty in my search. In addition, a further contribution is included due to the possible dependence of the FEI calibration factors on the choice of calibration sample. This analysis uses FEI calibrations extracted from $B^+ \rightarrow \bar{D}^0 \pi^+$ samples. However, an alternative set of calibrations is available as obtained from $B^+ \rightarrow X \ell^+ \nu$ samples. I use the 0.08 difference in central values of the FEI calibration factors from these two samples as an additional multiplicative systematic uncertainty of 10.7%, which – combined in quadrature with 6.7% from the statistical uncertainty of the FEI calibration – totals a 12.6% relative systematic uncertainty, the second major source for this analysis.
- Tracking.** An additional systematic uncertainty of 0.27% per reconstructed charged particle is assigned to account for possible differences in tracking efficiency between data and simulation. This uncertainty is obtained centrally in the collaboration from the control sample $e^+e^- \rightarrow \tau^+\tau^-$, where one τ decays leptonically and the other decays to three charged pions. This uncertainty is fully correlated between the three signal tracks leading to a total systematic uncertainty on the branching fraction of 0.81%.

- **Uncertainty in the efficiency of signal PID requirements.** PID requirements final-state charged particles. The efficiency for these requirements is determined from simulation, and corrected with scale factors determined by data control samples as functions of relevant kinematic and time-dependent factors, as described in Sec. 6.2. The uncertainties on these corrections propagate into the analysis as a 1.6% relative systematic uncertainty on the branching fraction.
- **π^0 veto efficiency.** A π^0 veto rejects events containing a π^0 not associated to the B^+B^- reconstruction. A 1.06 ± 0.02 scale factor is applied to match the simulated signal-efficiency for this requirement to data, as described in Sec. 6.2. The uncertainty on this factor propagates as a 1.9% relative systematic uncertainty on the branching fraction.
- **Simulated sample size.** The signal efficiency ϵ_s is determined from a simulated signal sample with a finite number of events N_{tot} . The uncertainty on ϵ_s is approximated with $\sigma(\epsilon_s) = \sqrt{\epsilon_s(1 - \epsilon_s)/N_{\text{tot}}}$ and implies a 3.5% relative systematic uncertainty on the branching fraction.
- **Signal decay model.** The signal efficiency ϵ_s determined from simulation is the result of an integral over the assumed q^2 distribution subjected to the analysis selection. It therefore depends on the signal-decay model BTOSLLBALL used in generation [71]. If the true dynamics of the decay deviate from that model, the results may change. I account for this dependence by weighting the q^2 distribution assumed in the default analysis to mirror alternative phenomenological models [32] (fig. 8.1). I use the difference in results as systematic uncertainty, which amounts to a relative 4.3% contribution to the branching fraction.

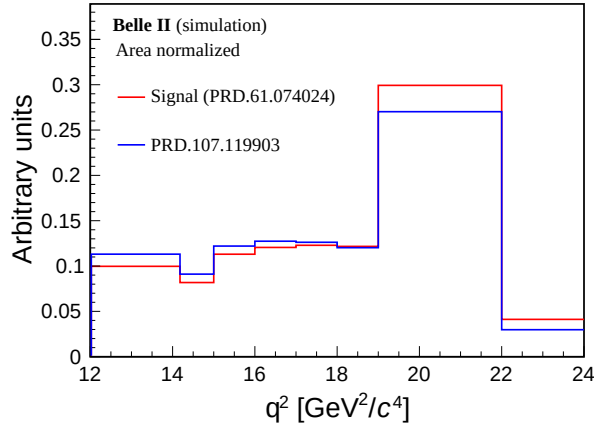


Figure 8.1: Generator-level (red) default [71] and (blue) alternative [32] signal q^2 distributions.

8.3 Uncertainties in external inputs

Conversion of the observed signal yield (or lack thereof) into a branching fraction implies division by the number of bottom-antibottom pairs contained in the Belle II sample $N(B\bar{B}) = (387 \pm 6) \times 10^6$ and by the branching fraction for a $\Upsilon(4S)$ to decay into a pair

Source	Size
Expected background yield	$^{+1.1}_{-1.0}$ (absolute)
FEI scale factor	12.6%
Tracking efficiency	0.8%
Particle identification corrections	1.6%
π^0 veto efficiency	1.9%
Simulated sample-size	3.5%
Decay model	4.3%
Total B -meson yield	1.6%
$\mathcal{R}(4S)$ branching fraction	$^{+1.4\%}_{-2.1\%}$
Total	$\pm[xx]$

Table 8.1: Sources of systematic uncertainty and their impact on the $B^+ \rightarrow K^+\tau^+\tau^-$ branching fraction. All values represent fractional uncertainties except the first, which is absolute.

of charged bottom-mesons $f^{+-} = 0.5113^{+0.0073}_{-0.0108}$. The uncertainties on the above factors propagate into multiplicative systematic uncertainties.

- **Total B meson yield.** The uncertainty on the number of $\mathcal{R}(4S) \rightarrow B\bar{B}$ events in the current dataset, $N_{B\bar{B}} = (387 \pm 6) \times 10^6$, constitutes a source of a 1.6% relative systematic uncertainty on the branching fraction.
- **$\mathcal{R}(4S)$ branching fraction.** The $f^{+-} = 0.5113^{+0.0073}_{-0.0108}$ value is taken from the recent HFLAV average [76], which newly includes an uncertainty due to possible isospin-symmetry violation between the decay modes used to determine it. The uncertainty on f^{+-} is propagated as $^{+1.4\%}_{-2.1\%}$ relative systematic uncertainty on the branching fraction.

All the systematic sources and their impact on the $B^+ \rightarrow K^+\tau^+\tau^-$ branching fraction are summarized in Table 8.1. A precise assessment of the total systematic uncertainty would need disclosing of the signal search-region to determine the fractional impact of the absolute uncertainty on the background estimate. However, by assuming the central branching-fraction value observed by Babar [65], $\mathcal{B} = 1.31 \times 10^{-3}$ this analysis would obtain a systematic uncertainty of 22%, which is significantly inferior to the 46% statistical uncertainty.

Chapter 9

Results and summary

I finally inspect the signal-search region $E_{\text{ECL}} < 250$ MeV in data. The difference between observed and expected background events determines the presence, or absence, of signal. I choose a counting approach as opposed to fitting the signal yield because the signal region is expected to be so sparsely populated that any gain in statistical information due to shape differences would be spoiled by the additional systematic uncertainties due to the poor knowledge of shapes themselves. Figure 9.1 shows the E_{ECL} distribution for data and simulation in the signal-search region after the final selections and including all the relevant corrections to simulation discussed in Chap. 6. I observe $N_{\text{obs}} = [xx]$ events. When subtracting the expected background yield $N_{\text{exp}} = [xx \pm xx]$, these correspond to a yield of $[xx \pm xx]$, showing evidence [no evidence] of signal. I derive the final results using a frequentist construction based on likelihood-ratio ordering, also known as Feldman-Cousins method [79]. This ordering is convenient as it prevents empty confidence intervals in the vast majority of physical inferences and provides a unified approach for setting limits, if no excess is observed, or constructing two-sided confidence intervals, if a significant signal is detected. The number of events observed in data after the final selection follows a Poisson distribution,

$$\mathcal{L}(\mu_s | N_{\text{obs}}) = \frac{(\mu_s + N_{\text{exp}})^{N_{\text{obs}}} e^{-(\mu_s + N_{\text{exp}})}}{N_{\text{obs}}!},$$

where μ_s indicates the signal yield. The goal is to determine a confidence interval for μ_s . The Feldman-Cousins method uses the likelihood-ratio ordering, comparing the likelihoods at different values of signal strength μ_s^i with its maximum value, $R = \mathcal{L}(\mu_s^i | N_{\text{obs}}) / \mathcal{L}(\hat{\mu}_s | N_{\text{obs}})$, where $\hat{\mu}_s$ is the signal yield that maximizes the likelihood for the given observed data. The possible outcomes of N_{obs} for each hypothesis μ_s^i are ranked in decreasing order of likelihood ratio and the confidence interval is constructed including progressively values of μ_s^i with high likelihood ratios until the desired probability content of the interval is reached.

Before proceeding with the derivation of the branching fraction, systematic uncertainties are included. The total additive systematic uncertainty, expressed in number of events, and the total multiplicative uncertainty, expressed as a fractional contribution, are obtained by summing in quadrature all systematic uncertainties of the same type as summarized in table 8.1. Statistically proper inclusion of systematic uncertainties in the construction of frequentist confidence regions is a notoriously difficult task that is regularly object of debate in statistics. While a variety of reasonable approaches tailored for specific use cases exist and are commonly used, none has been demonstrated to possess proper and optimal statistical properties independently of the problem at hand. In this search, I choose the

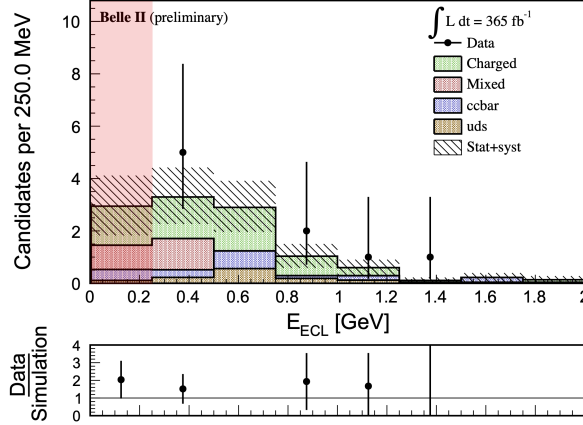


Figure 9.1: Distributions of E_{ECL} for (point) data and (histogram) corrected simulation in unblinded data. Label ‘charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; ‘uds’ indicates $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds, all in simulation. [The red-shaded signal-search region will be undisclosed when the collaboration approves].

so-called “supremum p -value” method, which prescribes that results should have at least nominal coverage whatever the true value of the nuisance parameters. In practice, this consists in repeating the confidence-region construction for various sets of extreme choices, among those plausible, of the true values of the nuisance parameters (i.e., values that make data to look more background-like) and pick the results that provide the lowest confidence level. While conceptually correct from a frequentist point of view, this method may lead to overcoverage for a significant fraction of the space of possible nuisance parameters. However, in a search for a process that has the potential to provide evidence for non-SM physics, this is preferable than risking to incur in an unsupported discovery claim generated by poor handling of some systematic effect.

Figure 9.2 reports Feldman-Cousins confidence belts. The left panel shows the expected acceptance region including only statistical uncertainties. The right panel shows the expected acceptance region including systematic uncertainties as well, under the assumption of the worst-case scenario of 1σ increase in expected background and 1σ decrease in signal efficiency. The resulting expected upper limit at 90% C.L. is 0.9×10^{-3} without systematic uncertainties and 1.0×10^{-3} with systematic uncertainties.

After undisclosed the signal-search region in data, the resulting $B^+ \rightarrow K^+\tau^+\tau^-$ branching fraction is

$$\mathcal{B}(B^+ \rightarrow K^+\tau^+\tau^-) = [xx \pm xx \pm xx]$$

[leading to a 90% confidence-level upper limit of $\mathcal{B}(B^+ \rightarrow K^+\tau^+\tau^-) < [xx]$]. This is the first evidence of this decay, which is determined with xx significance, as estimated using the observed ratio between the full likelihood and the background-only likelihood interpreted as a χ^2 variable with one degree of freedom [this limit is xx times better/worse than the only previous experimental result].

The largest contribution to the systematic uncertainty comes from the uncertainty on expected background in the signal region. This is because the available control data size is small after selection, leading to significant statistical fluctuations and shape uncertainty

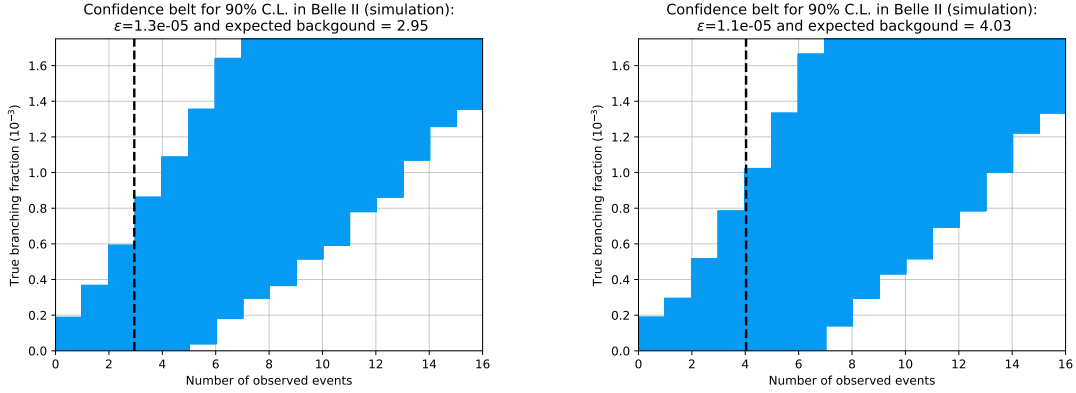


Figure 9.2: Feldman-Cousins confidence belts for (left) statistical-only and (right) complete results. The vertical dashed lines indicate an hypothetical observation in a signal-free simulated sample in which the observed events yield equals the expected background yield and allow to determine expected limits.

in the estimate. The second largest contribution is associated with the uncertainty on the B_{tag} efficiency correction, which is dominated by the systematic uncertainty on the FEI calibration factors.

A comparison of this analysis precision with existing best results is possible before undisclosed the signal-search region by assuming to observe in our data the signal rate observed by Babar. Our analysis achieves a statistical uncertainty of 46% and a systematic uncertainty of 22%. This is competitive with Babar, which reported a 48% statistical uncertainty and a 23% systematic uncertainty. However, the two analyses achieve similar performance through significantly different working points. The single-event sensitivity, which is the minimal branching fraction that would generate one event in the analysis sample is 19.3×10^{-5} in this search and 4.6×10^{-5} in the Babar search. This indicates that the Babar analysis favored a high signal efficiency, at the price of larger backgrounds, compared to our strategy of optimized suppression of backgrounds.

In summary, this thesis reports the first Belle II search for the $B^+ \rightarrow K^+ \tau^+ \tau^-$ decay, an important probe of non-SM physics whose impact is further enhanced by the persisting anomalies observed in rates of B decays to τ final states. The search is based on the full dataset collected prior to the 2023 shutdown, and it is the second of its kind globally. The sample contains 387 million pairs of bottom-antibottom mesons from decays of $\Upsilon(4S)$ mesons produced near-threshold in 10.58 GeV electron-positron collisions by the SuperKEKB collider.

I exploit the peculiar properties of the unique experimental environment by restricting the analysis to hadronic decays of the nonsignal B meson decays. This imposes kinematic constraints on the signal and substantially reduces the background. After a thorough sensitivity study I restrict the search to leptonic τ decays, which offer the most sensitivity. An innovative background-suppression choice based on opposite-charge kaon-lepton mass rejects the vast majority of prevailing and poorly modeled $B \rightarrow D$ backgrounds. Finally, I optimize the selections to improve the analysis sensitivity by testing the full analysis on realistically simulated samples selected through a large set of possible configurations of criteria. All of these choices allow for overcoming the challenge of the overwhelming backgrounds and lack of distinctive signal features. I extract signal using the residual energy detected in the calorimeter after having validated it in detail on control data. All

significant sources of systematic uncertainties are assessed and have a modest impact.

The analysis is currently under internal collaboration review and the signal-search region has not yet been undisclosed. The sensitivity is two times better than that of the current world-leading result in spite of our use of smaller sample.¹ In the broader picture, the impact of my work is multifaceted.

1. This work advances the global knowledge on $B^+ \rightarrow K^+ \tau^+ \tau^-$ decays with doubled sensitivity over the only previous result [65].
2. This work introduces significant innovation. Restricting the analysis to events where the opposite-charge kaon-lepton mass exceeds the D mass suppresses dominating and poorly modeled semileptonic $B \rightarrow D$ backgrounds simplifying the analysis without sacrificing sensitivity.
3. This work benefits or enables several ongoing and future analyses. In its capacity of first Belle II search for this channel and second of its kind worldwide, this analysis paves the way for future efforts in $B^+ \rightarrow K^+ \tau^+ \tau^-$ and similar channels. The study, characterization, and validation of E_{ECL} conducted for this work is broadly applicable to any analysis involving final-state neutrinos in the Belle II experiment and beyond. The restriction in kaon-lepton mass may benefit other $b \rightarrow s \ell \ell$ or $b \rightarrow s \tau \ell$ searches. The study of sensitivity contributed by $\tau \rightarrow \pi X$ final states reveals margins of gain, thus informing future efforts.

¹Since expected sensitivity is not provided in Ref. [65] this statement is based on the observed limit, which could be weaker than the actual analysis sensitivity as Babar observed a signal-like excess in one channel.

Appendices

Appendix A

An alternative, BDT-based, selection

A.1 Introduction

The default selection for the $B^+ \rightarrow K^+ \tau^+ \tau^-$ search is built upon sequential one-dimensional criteria because the opposite-charge kaon-lepton mass restriction isolates a signal-rich region with residual backgrounds that are already suppressed. An alternative approach is to lift the opposite-charge kaon-lepton mass reconstruction to increase signal efficiency and use nonlinear discriminators to control the increase in background by capturing and exploiting statistical differences present in the multidimensional correlations among discriminating observables. Usage of such discriminators on the sample restricted to $m(K^+ t^-) > 1.9 \text{ GeV}/c^2$ is not practical due to the limited size of the simulated sample. I combine non-linearly various discriminating observables in a boosted decision tree [80], a supervised multivariate classifier. I compare the expected sensitivity of the BDT approach to the nominal analysis after optimizing both the BDT score selection and the E_{ECL} signal window restriction. The results show that this alternative approach does not yield a significant gain in statistical sensitivity, which may induce larger systematic uncertainties associated with the increased complexity of the analysis.

A.2 Boosted decision trees

A multivariate classifier estimates the probability of a datum belonging to a given class; this probability is inferred from a set of observables $\mathbf{x} = (x_1, \dots, x_n)$. The algorithm operates in two phases. In the fitting phase, the classifier is "trained" using data with known classification (training sample). In this "supervised" phase the internal configuration of the classifying function that maps the inputs into a classification output is progressively adjusted to maximize the rate of successful classification. Successful classifications are known because the true classification is known for the training data. In the application phase, the resulting classifier is applied to new data with an unknown classification (testing sample). In this phase, the internal configuration of the classifying function is established from the training and used to classify the test data.

In this work I use a stochastic gradient-boosted decision tree. The decision tree is a specific type of supervised classifier that approximates the optimal classifying function by applying a set of consecutive binary requirements on each of the given discriminating observables. The maximum number of consecutive requirements is typically configurable and is called the depth of the tree. A schematic example of a decision tree is shown in fig. A.1 for a simple case of three discriminating observables and two classes. The setting of the selection requirements at each node is determined during training. By testing multiple

choices of the requirements at each node, the tree estimates the probability of training data to belong to a certain class of events for each. These probabilities are then compared with the known true classification. Only requirements that result in accurate predictions are implemented in the node. Hence, each requirement maximizes the separation between classes of events locally in the given training sample.

Predictions of trees with many consecutive requirements (deep trees) are often driven by the statistical fluctuations of the training data sample rather than by genuinely significant distinctive features. This "over-fitting" reduces the predictive power of the tree.

To reduce overfitting, ensembles of multiple shallow trees are combined into a "boosted" tree. While each shallow tree may give inaccurate predictions individually, combining them sequentially yields a model that is less likely to overfit, yielding good classification performance. Boosting proceeds by fitting an initial tree to the data; then a second tree is built targeted at classifying accurately only the subset of events where the first tree performs poorly; and then the sequence is repeated multiple times. Each successive tree attempts to correct the shortcomings of the combination of previous trees achieving an overall improved final performance. A commonly used boosting technique is gradient boosting [81]. The robustness of gradient-boosted decision trees against overfitting is improved by using random subsamples of the training data set instead of the full training sample in each boosting step. The strength of possible correlations between trees is reduced, thus achieving enhanced discriminating capabilities. This approach is called stochastic gradient boosted decision tree [82].

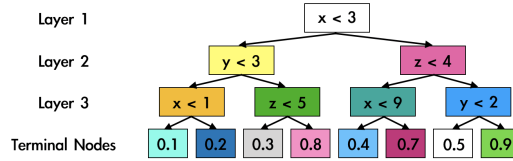


Figure A.1: Schematic example of a three-layer decision tree. At each node of the tree, a binary decision is made until a terminal node is reached. The numbers in the terminal node correspond to the probability of the test data point being signal candidates.

A.3 Choice of discriminating observables

I explore various sets of discriminating observables to choose the classifier inputs that offer optimal separation between signal and background, by comparing the discriminating power of each. Among the choices of invariant masses of any two final-states, angle between momenta of final-state particles, impact parameter differences between tracks, and event shape observables, I find that M_{miss}^2 , p_{t+}^* , $m(K^+t^-)$, and q^2 are sufficient for achieving nearly optimal performance, which is not surprising given that these observables are shown to be highly discriminating in the default analysis based on one-dimensional criteria. Figure A.2 shows their distributions in simulated signal and background events after baseline selections. The correlation matrix for these discriminating observables is shown in fig. A.3. Before feeding them into the classifier, it is important to validate them using a control data sample; otherwise, the classifier training may become suboptimal, reducing the sensitivity when applied to experimental data. Given the rarity and nature of our signal, a validation of signal distributions based on a suitable control decay is challenging and outside the scope of this test. However, I validate on data the background distributions using events in the q^2 sideband, $q^2 < 14.18 \text{ GeV}^2/c^4$, which has a similar composition as background in

the signal region except for larger $B^+ \rightarrow J/\psi K^+$ contribution (fig. A.4). I observe good simulation description for each.

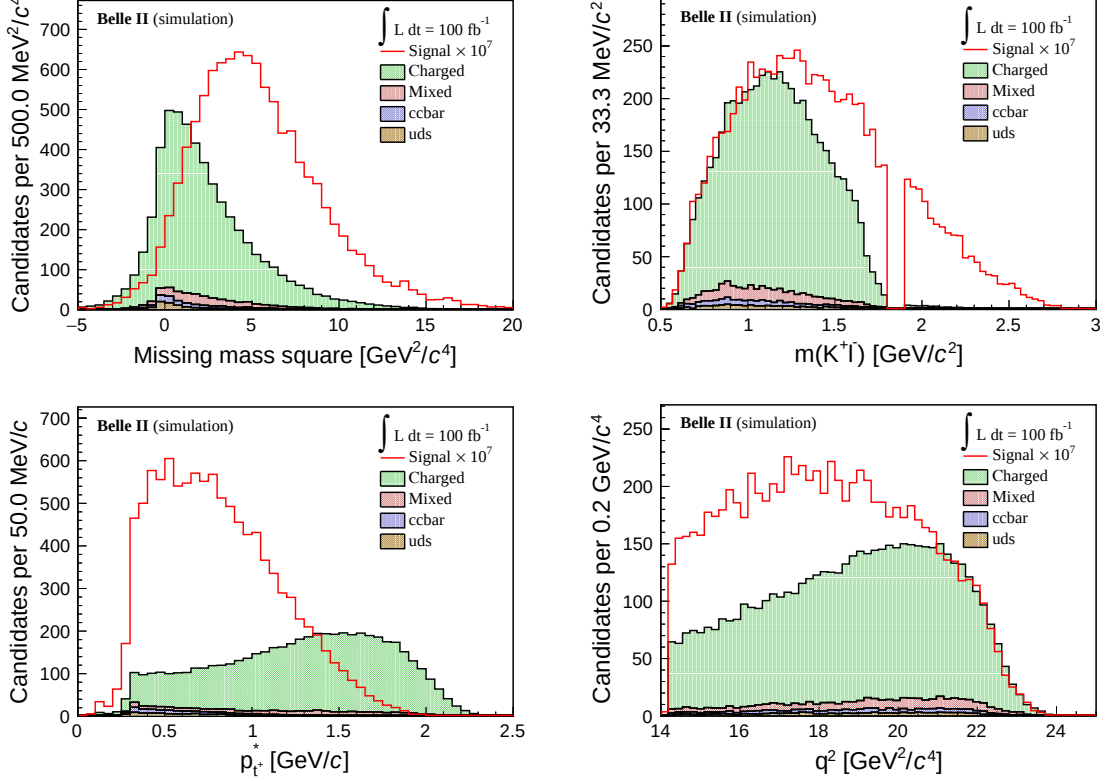


Figure A.2: Signal and background distributions of discriminating observables used as input in the BDT in simulation. Label ‘charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; ‘uds’ indicates $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds, all in simulation.

A.4 Classifier training and testing

I train the BDT using 10^4 simulated signal events and 10^4 simulated background events passing the baseline selection. A convenient way to estimate the classifier performance is through a receiver operating characteristic (ROC) curve, which represents signal efficiency as a function of background rejection (the complement to the background efficiency). The classifier performance improves as the ROC curve approaches the top-right corner of the two-dimensional space. This is expressed quantitatively by the area under the curve (AUC), which equals unity for the ideal case of total signal efficiency with full background rejection. The ROC curve resulting from my BDT is shown in fig. A.5.

I test the BDT on an independent simulated sample with the same composition as the training data to ensure that the model is not overfitted. Figure A.5 and A.6 compare the ROC curves and the distributions of classifier output for training and testing samples, respectively. The distributions agree for both signal and background components, showing negligible, if any, overfitting.

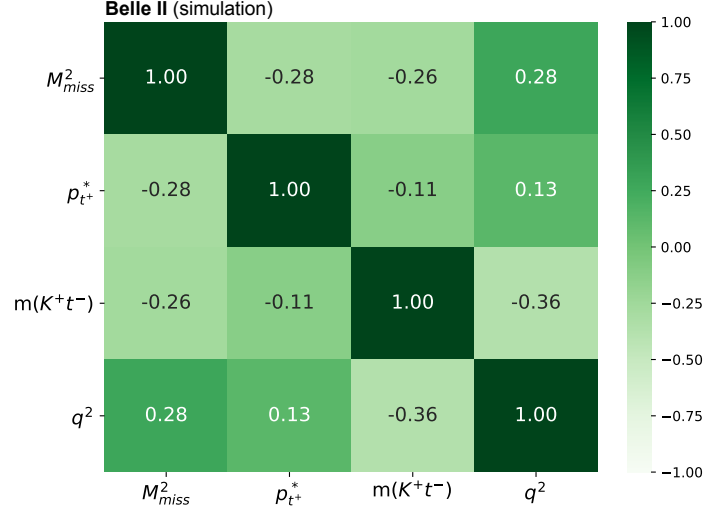
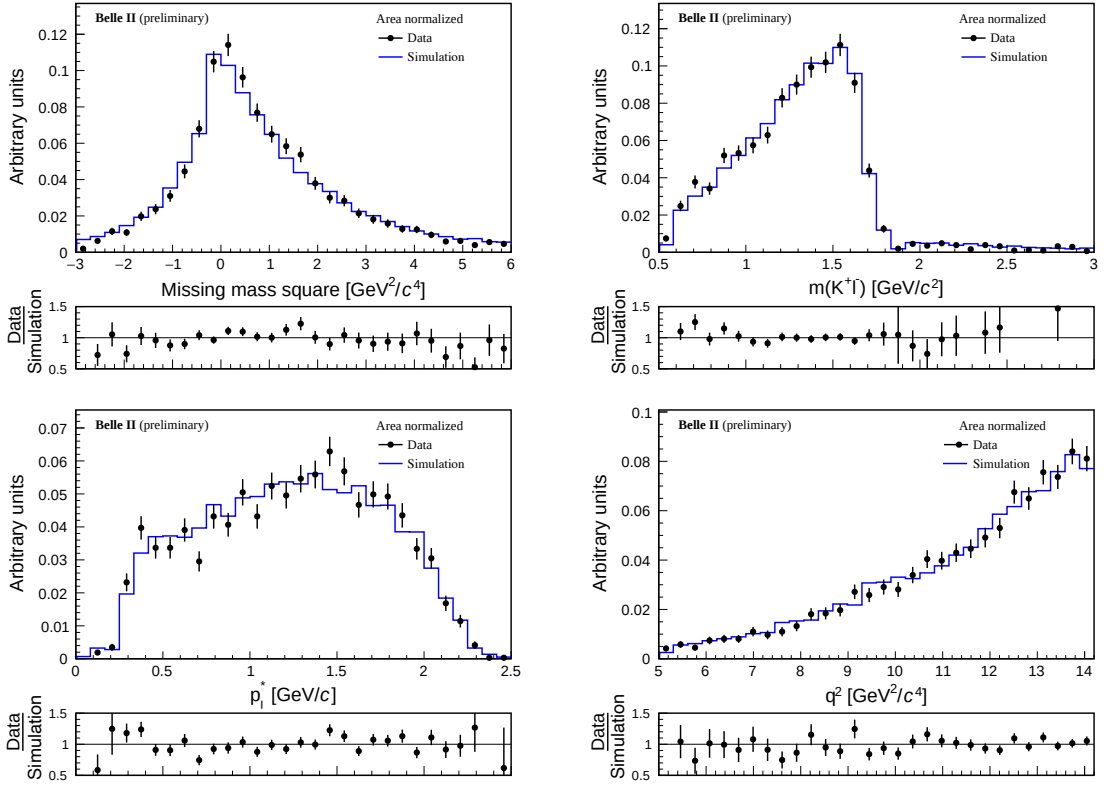


Figure A.3: Correlation matrix of BDT input observables in simulated signal sample


 Figure A.4: Distributions of (top-left) M_{miss}^2 , (top-right) $m(K^+t^-)$, (bottom-left) $p_{t^+}^*$, and (bottom-right) q^2 for (point) data and (histogram) simulation in q^2 sideband. Distributions are normalized to unity.

The BDT algorithm used allows for assessing the discriminating efficacy of each input

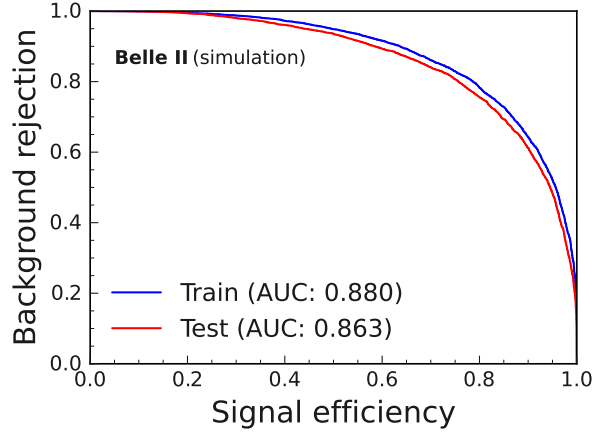


Figure A.5: Receiver operating characteristic for the BDT in (blue) training and (red) testing simulated samples.

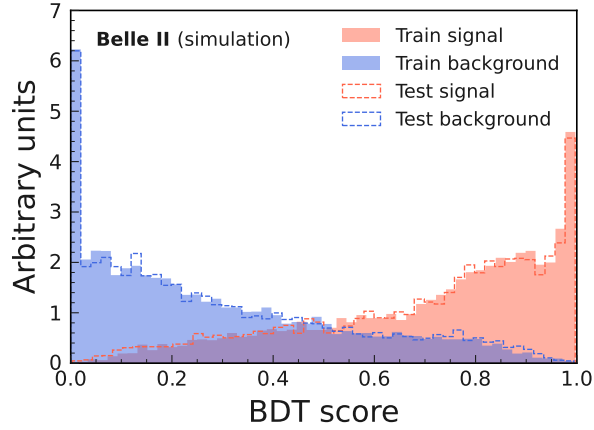


Figure A.6: BDT score on (solid) training and (dash) testing samples for simulated (red) signal and (blue) background events.

observable (table A.1). The lepton momentum exhibits the highest discriminating power.

The ROC curve qualifies the classifier performance, but it does not indicate which specific requirement on the classifier output provides the best selection, as this depends on the physics goal at hand. Here, I optimize the classifier output selection is targeting the best-expected limit, in analogy with the default analysis.

BDT input observable	Importance
p_{t+}^*	100
M_{miss}^2	71
$m(K^+t^-)$	63
q^2	0

Table A.1: Relative importance ranking of classifier input observables. The scale is relative and spans the 0 – 100 range.

A.5 Selection optimization

The BDT effectively separates signal from background as shown in fig. A.6, with signal candidates peaking closer to 1.0 and background candidates closer to 0.0. The optimization procedure is the same as discussed in chapter 6. This includes a two-dimensional scan in the space of the BDT score and E_{ECL} signal window. The scan ranges are $[0.70, 0.99]$ with 0.01 BDT score step-width and $[0.1, 1]$ GeV with 0.05 GeV E_{ECL} step-width. The optimal selection is a BDT score greater than 0.96 and E_{ECL} lower than 400 MeV. Figure A.7 illustrates the selected regions.

In these studies, all corrections to simulation are applied except PID and π^0 -veto efficiency corrections, which have a marginal impact for this purpose. The signal efficiency is corrected using the averaged FEI efficiency calibration factor. The same correction criteria are applied in the default analysis for consistency. The resulting signal efficiency is 1.9×10^{-5} , with 9 background events. The five-folded averaged expected limit, with FEI efficiency and background yield correction (70%), is 0.96×10^{-3} . A detailed comparison summary between my nominal analysis and this BDT alternative is provided in table A.2. The correction of PID and π^0 veto selection are not applied in this comparison as their effects are marginal.

The sensitivity of the BDT-based approach and the sequential one-dimensional selection approach are the same. I therefore favor the latter for its simplicity, transparency, and robustness.

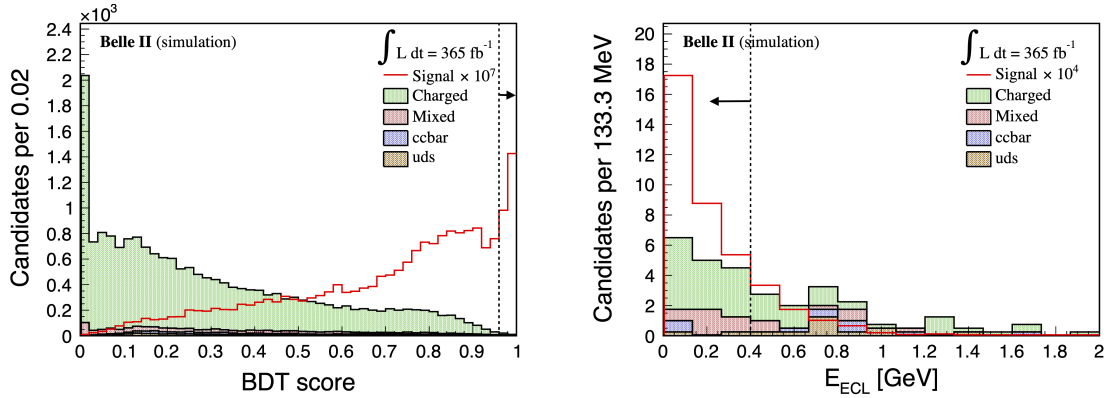


Figure A.7: Distributions of (left) BDT score and (right) E_{ECL} in realistic simulation. The vertical black dashed lines indicate the selection restrictions with arrows pointing toward the accepted events. See previous plots for legend conventions.

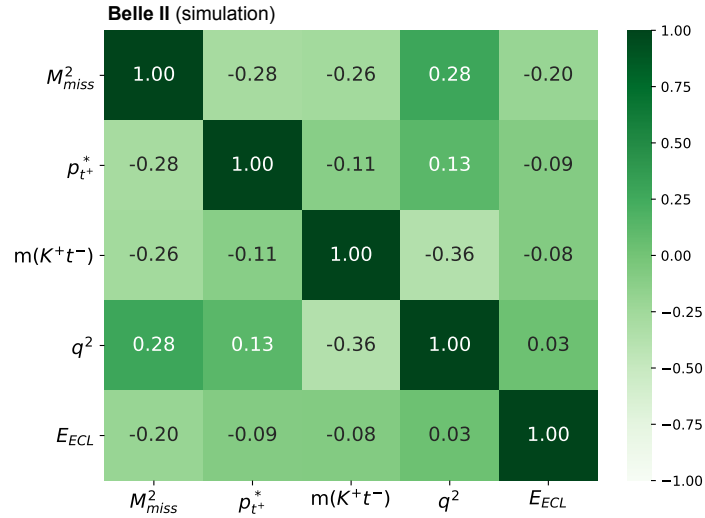
A.6 Impact of BDT input choices

In the BDT-based analysis, I exclude E_{ECL} as an input observable since it serves as the principal signal extraction observable. However, I explore an alternative approach by including E_{ECL} as a BDT input to assess whether it enhances analysis performance. In this case the signal extraction would be based on excess count of the BDT score. First, I examine the correlation between E_{ECL} and other input observables (fig. A.8). The E_{ECL} exhibits a low correlation with the other BDT input observables. Figure A.9 presents the BDT score and ROC curve. The inclusion of E_{ECL} slightly improves the area under the curve. I then optimize the BDT score to achieve the best sensitivity. The optimization

	Nominal analysis	BDT approach
Optimized selection	$\times$	BDT > 0.96
	$E_{\text{ECL}} < 250 \text{ MeV}$	$E_{\text{ECL}} < 400 \text{ MeV}$
	$p_{t^+}^* > 0.5 \text{ GeV}/c$	$\times$
	$M_{\text{miss}}^2 > 1.6 \text{ GeV}^2/c^4$	$\times$
Signal Efficiency	1.4×10^{-5}	1.9×10^{-5}
Background	5	9
UL at 90% C.L.	1.00×10^{-3}	0.96×10^{-3}

Table A.2: Optimized selections and expected sensitivity in the BDT-based approach.

scan range is $[0.4, 1.0]$ with 0.01 step-width in the BDT score. The optimal point is at a BDT score greater than 0.97. After final selection, the signal efficiency is 2.7×10^{-5} , the expected background is 18, and therefore the expected upper limit with 90% C.L. is 0.97×10^{-3} . Table A.3 compares the expected sensitivity between BDT models trained with and without E_{ECL} . No improvement is observed, likely due to the uncorrelated nature of E_{ECL} with other input observables, offering no clear advantage when used in the classifier alongside the separately optimized E_{ECL} signal window.


 Figure A.8: Correlation matrix of ‘BDT with E_{ECL} input’ input observables in simulated signal sample.

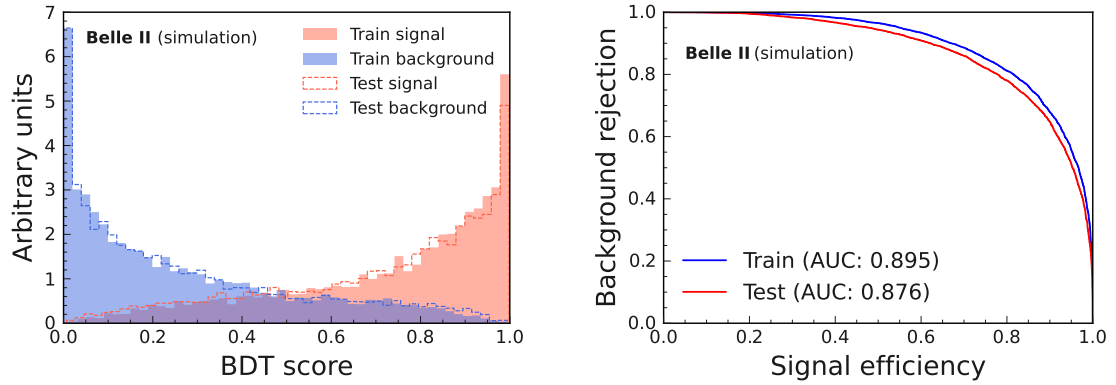


Figure A.9: (Left) ‘BDT with E_{ECL} inputs’ score on (solid) training and (dash) testing samples for (red) simulated signal and (blue) background events. (Right) ROC curve of this classifier.

	Without E_{ECL} input	With E_{ECL} input
BDT	> 0.96	> 0.97
E_{ECL}	$< 400 \text{ MeV}$	$\times$
UL at 90% C.L.	0.96×10^{-3}	0.97×10^{-3}

Table A.3: Expected sensitivity comparison between ‘BDT without E_{ECL} input’ and ‘BDT with E_{ECL} input’ in a realistic simulated sample.

Appendix B

Sensitivity from exclusive analysis of lepton channels

The default analysis treats inclusively the three leptonic τ final states by using the same selection criteria for Kee , $Ke\mu$, and $K\mu\mu$. However, signal shape and background composition may differ among them, opening up the possibility that treating them separately may enhance global sensitivity. I investigate channel-specific sensitivities by optimizing selection criteria separately on M_{miss}^2 , p_{t+}^* , and E_{ECL} signal range in realistic simulated samples restricted to $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ restriction. In these studies, all corrections to simulation are applied except for PID and π^0 -veto efficiency corrections, which have a marginal impact for this purpose. The signal efficiency is corrected using the averaged FEI efficiency calibration factor. The same correction criteria are applied in the inclusive analysis for consistency. The criteria test ranges are $[-2.0, 6.0] \text{ GeV}^2/c^4$ in M_{miss}^2 with $0.2 \text{ GeV}^2/c^4$ step-width; $[0.2, 1.0] \text{ GeV}/c$ in p_{t+}^* with $0.05 \text{ GeV}/c$ step-width and $[0.1, 1.2] \text{ GeV}$ in E_{ECL} with 0.05 GeV step-width.

Kee – Optimized selection criteria are $M_{\text{miss}}^2 > 2.0 \text{ GeV}/c^2$, $p_{t+}^* > 0.5 \text{ GeV}/c$, and $E_{\text{ECL}} < 950 \text{ MeV}$ (fig. B.1). The resulting signal efficiency is 3.9×10^{-6} , with 1 background event. The branching-fraction uncertainty assuming no signal is 9.2×10^{-4} . The five-folded average expected upper limit at 90% confidence level is 2.1×10^{-3} .

$Ke\mu$ – Optimized selection criteria are $M_{\text{miss}}^2 > 0.6 \text{ GeV}/c^2$, $p_{t+}^* > 0.4 \text{ GeV}/c$, and $E_{\text{ECL}} < 350 \text{ MeV}$ (fig. B.1). The resulting signal efficiency is 10.5×10^{-6} , with 7 background events. The branching-fraction uncertainty assuming no signal is 8.9×10^{-4} . The five-folded average expected upper limit at 90% confidence level is 1.5×10^{-3} .

$K\mu\mu$ – Optimized selection criteria are $M_{\text{miss}}^2 > 2.8 \text{ GeV}/c^2$, $p_{t+}^* > 0.45 \text{ GeV}/c$, and $E_{\text{ECL}} < 550 \text{ MeV}$ (fig. B.1). The resulting signal efficiency is 3.9×10^{-6} , with 1 background event. The branching-fraction uncertainty assuming no signal is 9.0×10^{-4} . The five-folded average expected upper limit at 90% confidence level is 1.9×10^{-3} .

Table B.1 summarizes these findings. For comparing the sensitivity with the default inclusive analysis, I combine the expected branching-fraction uncertainties assuming no signal by performing a weighted average of the exclusive results, which yields 5.2×10^{-4} . This shows only a 8% improvement over the branching-fraction uncertainty of the inclusive default analysis, 5.65×10^{-4} . I therefore choose to proceed with the inclusive treatment of the lepton final states for simplicity.

APPENDIX B. SENSITIVITY FROM EXCLUSIVE ANALYSIS OF LEPTON CHANNELS

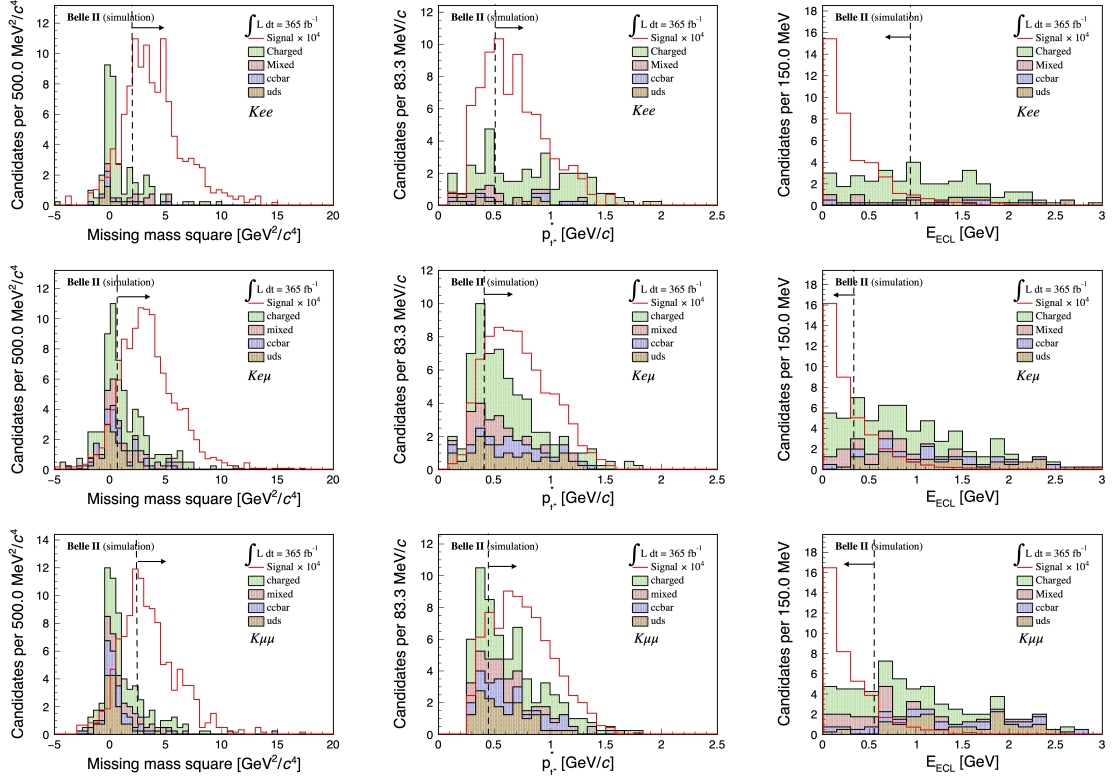


Figure B.1: Distributions of (left) M_{miss}^2 , (middle) p_{t+}^* , and (right) E_{ECL} in realistic simulated samples with (top row) Kee , (middle row) $Ke\mu$, and (bottom row) $K\mu\mu$ signal final states. Distributions are normalized to data luminosity. Label ‘charged’ indicates B^+B^- backgrounds; ‘mixed’ indicates $B^0\bar{B}^0$ backgrounds; ‘ccbar’ indicates $c\bar{c}$ backgrounds; ‘uds’ indicates $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ backgrounds, all in simulation.

	Kee	$Ke\mu$	$K\mu\mu$	Inclusive $K\ell\ell$
M_{miss}^2	$> 2.0 \text{ GeV}/c^2$	$> 0.6 \text{ GeV}/c^2$	$> 2.8 \text{ GeV}/c^2$	$> 1.6 \text{ GeV}/c^2$
p_{t+}^*	$> 0.5 \text{ GeV}/c$	$> 0.4 \text{ GeV}/c$	$> 0.45 \text{ GeV}/c$	$> 0.50 \text{ GeV}/c$
E_{ECL}	$< 950 \text{ MeV}$	$< 350 \text{ MeV}$	$< 550 \text{ MeV}$	$< 250 \text{ MeV}$
Efficiency	3.86×10^{-6}	10.5×10^{-6}	3.92×10^{-6}	14.0×10^{-6}
Background	1	7	1	5
UL at 90% C.L.	2.1×10^{-3}	1.5×10^{-3}	1.9×10^{-3}	1.0×10^{-3}

Table B.1: Expected upper limit from the default analysis of exclusive leptonic final states.

Appendix C

Sensitivity from pion channels

C.1 Introduction

The default analysis uses only $\tau^- \rightarrow \ell^- \nu \bar{\nu}$ final states. Incorporating pions into the final state increases signal efficiency. However, this inclusion introduces substantial background, and a dedicated study is needed to assess whether the increase in efficiency compensates for the additional background, yielding a gain in final sensitivity or not. To decide on which final states to include, I study the sensitivity brought by pion decay modes and compare it with the $K\ell\ell$ sensitivity.

I include all τ decay modes that yield one charged pion and any number of neutral particles ($\tau^- \rightarrow \pi^- \nu X$). This is because significant contributions come from decays involving one or more neutral pions, which arise from both nonresonant (e.g., $\tau^- \rightarrow \pi^- \pi^0 \pi^0$, 9.2% of τ decay width) and resonant processes (e.g., $\tau^- \rightarrow \rho^- (\rightarrow \pi^- \pi^0) \nu$ having 25.5% of τ decay width). Channels with high charged-pion multiplicity, such as $\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu$, are excluded for simplicity by restricting the sample to have exactly three charged particles. I broadly classify the final states into three independent categories:

1. $K\ell\ell$: both τ leptons decay into leptons yielding $K^\pm e^\pm e^\mp$, $K^\pm e^\pm \mu^\mp$, $K^\pm \mu^\pm e^\mp$, or $K^\pm \mu^\pm \mu^\mp$ final states. This is the state used in the default analysis.
2. $K\ell\pi$: one τ lepton decays into a lepton while the other decays into a pion, yielding $K^\pm e^\pm \pi^\mp$, $K^\pm \pi^\pm e^\mp$, $K^\pm \mu^\pm \pi^\mp$, or $K^\pm \pi^\pm \mu^\mp$ final states.
3. $K\pi\pi$: both τ leptons decay into pions, yielding the $K^\pm \pi^\pm \pi^\mp$ final state.

The most significant difference between the background composition of the leptonic and pion final states is the substantial contribution from light-quark backgrounds in pion final states. This occurs because pions are more susceptible to light-quark contamination compared to leptons.

C.2 $K\ell\pi$ final state

C.2.1 Reconstruction and baseline selection

The reconstruction and baseline selections are similar between pion and lepton decays (section 4.4). This includes B_{tag} selections, signal track selection, restrictions on signal track multiplicity, D meson veto, charmonium veto, photon conversion veto, and q^2 selections. The π^0 veto is lifted, as I observe that rejection of decays involving π^0 diminishes the $K\ell\pi$

signal efficiency because a large fraction of charged pions are accompanied by neutral pions. Additionally, particle identification for charged pions is required.

I ensure separation into independent $K\ell\ell$ and $K\ell\pi$ samples to prevent double counting. This is achieved by implementing mutually exclusive particle-identification requirements. Electron global particle-identification is imposed to be greater than 0.9 for electron candidates, muon global particle-identification greater than 0.9 for muon candidates, and both electron and muon global particle-identifications to be smaller than 0.9 for pion candidates. This selection achieves 80% purity in the signal pion. This is further improved by applying a pion global particle-identification requirement greater than 0.05. All these particle-identification selections are ensured to be optimal based on the best-expected limit in realistic simulated background -only samples. With these selections, according to simulation the pion candidate in the $K\ell\pi$ signal is truly a pion in 84% of events as described in table C.1.

Although $K\ell\ell$ and $K\ell\pi$ candidates are independent due to opposite particle-identification selections, multiple candidates with different final states may be reconstructed in the same event. In such cases, a candidate is selected based on the final state. I give the highest priority to $K\ell\ell$ candidates, if any, given lower background compared to the pion mode. This means that if multiple $K\ell\ell$ candidates are found, one is randomly selected; if else, a $K\ell\pi$ candidate is randomly chosen.

C.2.2 Study of charge correlations

The $K\ell\pi$ kinematic properties also depend on the charge correlation between kaon and pion. Figure C.1 shows $m(K^+t^-)$ distributions in inclusive $K\ell\pi$, $K^+\pi^+\ell^-$ and $K^+\ell^+\pi^-$ samples. In the opposite-charge $K\ell\pi$ sample, a large contribution from $\bar{D}^0 \rightarrow K^+\pi^-\pi^0$ is visible, which is shifted in $m(K^+t^-)$ due to the association of the π^0 . In contrast, the same-charge $K\ell\pi$ sample has a smooth $m(K^+t^-)$ distribution. This means that background composition is different in each (as shown in table C.2). The chief striking difference between these modes is the $q\bar{q}$ background. Simulation shows that opposite-charge $K\ell\pi$ has a 43% $q\bar{q}$ contribution, whereas same-charge $K\ell\pi$ has 17%.

I therefore study $K\ell\pi$ sensitivity separately for inclusive-charge, opposite-charge, same-charge for both the default analysis based on sequential one-dimensional restrictions and for the BDT-based analysis. In both approaches, the same-charge $K\ell\pi$ provides the same sensitivity, whereas the BDT selection in opposite-charge $K\ell\pi$ has a 1.4 times better sensitivity than in the default analysis. The result also shows that the same-charge $K\ell\pi$ sample has a superior expected sensitivity than the opposite-charge. However, the combined expected sensitivity of these two charge-correlated samples offers a marginal improvement over the sensitivity obtained from the inclusive charge $K\ell\pi$ sample. I therefore choose inclusive charge $K\ell\pi$ for the comparison with $K\ell\ell$.

Particle	Opposite charge	Same charge
Pion	84.3%	83.5%
Kaon	0.3%	1.4%
Electron	3.9%	3.7%
Muon	11.0%	10.9%

Table C.1: Particle-identification performance (in terms of possibility for a pion to be truly a pion, kaon, electron, or muon in the opposite- and same-charge samples.)

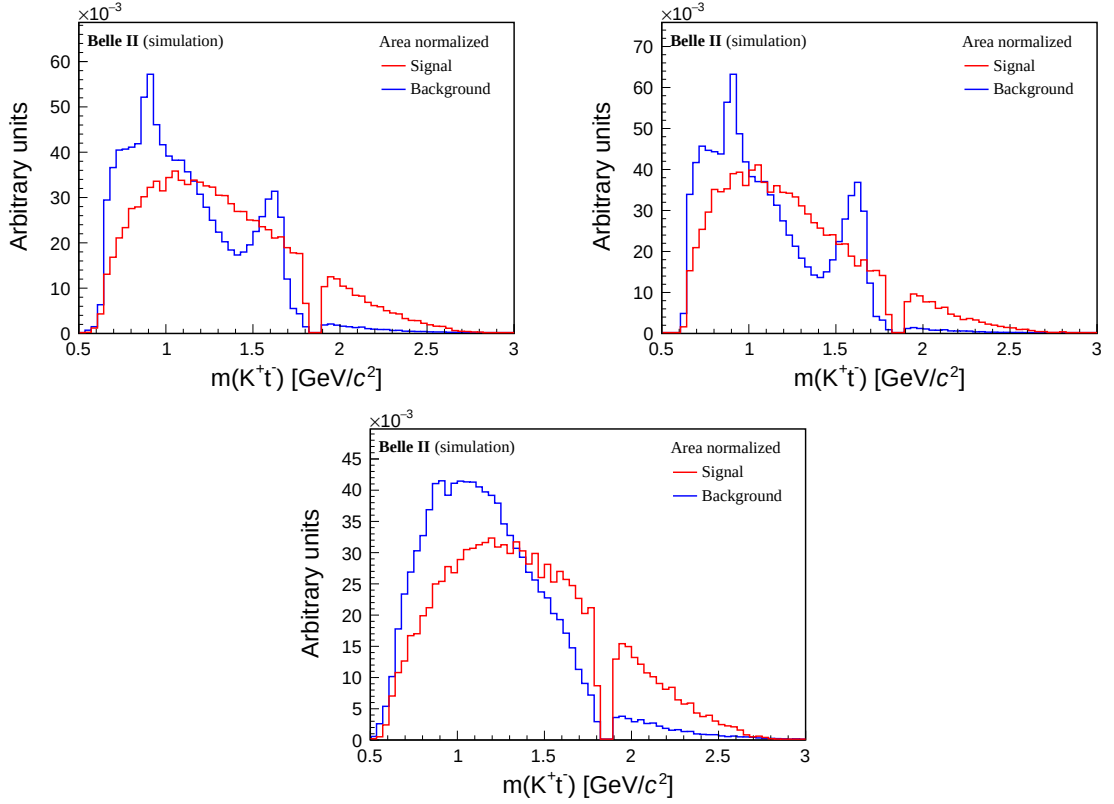


Figure C.1: Distributions of $m(K^+t^-)$ for (top-left) charge-inclusive $K\ell\pi$, (top-right) opposite-charge $K\ell\pi$, and (bottom) same-charge $K\ell\pi$ events in simulation. Distributions are normalized to unity.

Background type	Charge-inclusive $K\ell\pi$	Opposite-charge $K\ell\pi$	Same-charge $K\ell\pi$
Charged	62%	50%	69%
Mixed	11%	6%	14%
Continuum	27%	43%	17%

Table C.2: Background composition in $K\ell\pi$ analysis for simulated events with $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ and $E_{\text{ECL}} < 600 \text{ MeV}$.

C.2.3 BDT-based selection

As BDT-based selection in $K\ell\pi$ provides better sensitivity, I describe it first and then provide the expected sensitivity for the standard approach. I perform the BDT-based selection by exploring the entire $m(K^+t^-)$ spectrum except for the D meson veto region. The BDT algorithm is the same as that used for the $K\ell\ell$ case (appendix A). The discriminating observables are M_{miss}^2 , $p_{t^+}^*$, $m(K^+t^-)$, q^2 , and $\cos\text{TBTO}$ of B_{tag} . Here $\cos\text{TBTO}$ is introduced to suppress for the light-quark background. Figure C.2 illustrates their distribution in the charge-inclusive $K\ell\pi$ simulated sample. Figure C.3 shows their correlations in a simulated signal.

I train and test the BDT on independent samples with 10^4 candidates each. Figure C.4

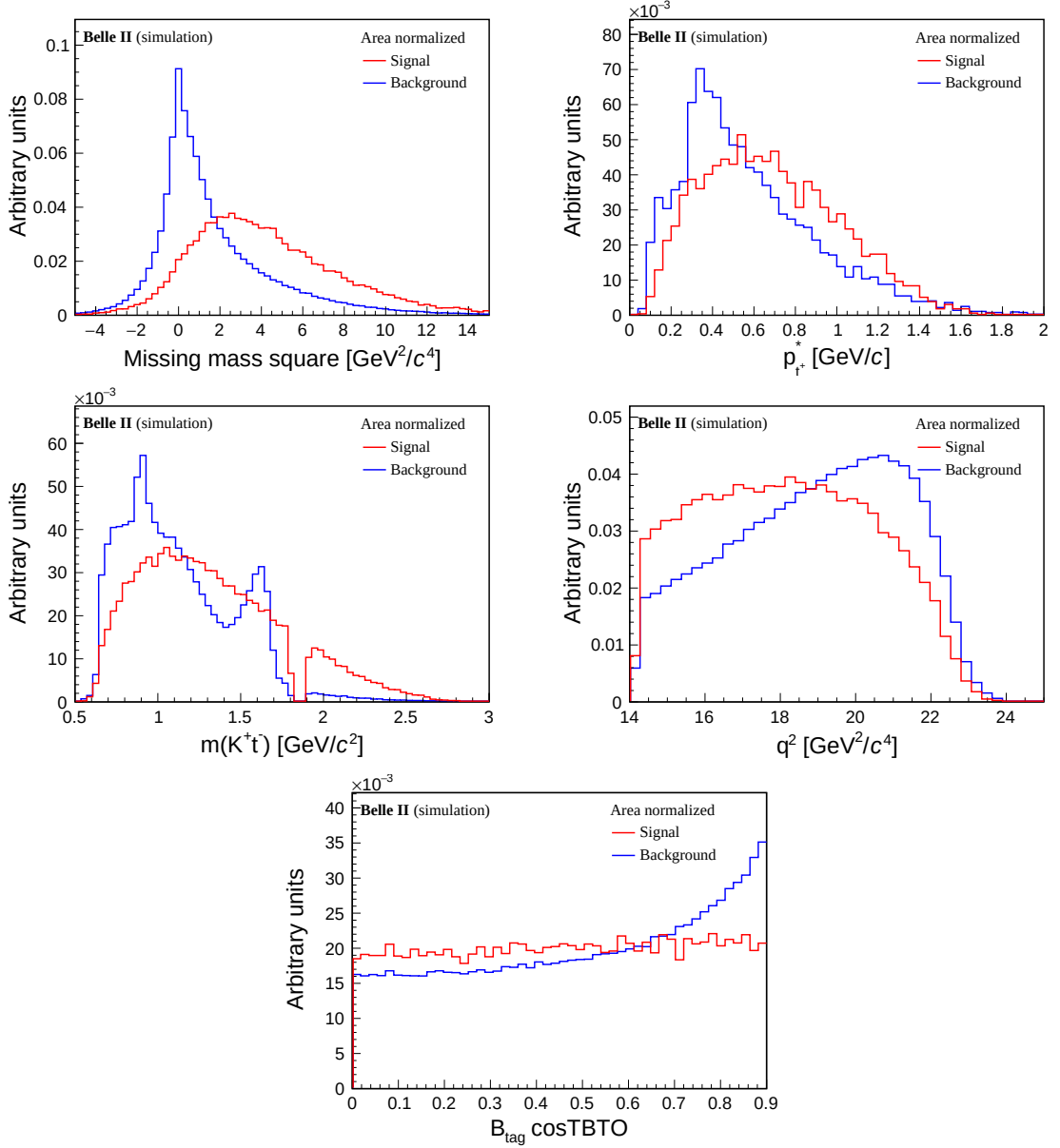


Figure C.2: Distributions of (top-left) M_{miss}^2 , (top-right) p_{t+}^* , (middle-left) $m(K^+t^-)$, (middle-right) q^2 , and (bottom) cosTBTO in charge-inclusive $K\ell\pi$ simulation. Distributions are normalized to unity.

shows the BDT output and the ROC curve. Table C.3 provides the importance metric of the input observables in the training. The BDT performance in the $K\ell\pi$ mode is inferior to that in the $K\ell\ell$ mode. This is reflected in the BDT score distribution in which $K\ell\ell$ has a distinct single peak with suppressed background, whereas the BDT score in $K\ell\pi$ has a broader signal peak with a higher background and flatter tail.

C.2.4 Selection optimization and sensitivity

As done in 6.3 for the $K\ell\ell$ analysis, I optimize selections on the BDT score selection and E_{ECL} signal window based on the expected upper limit in conditions of no signal.

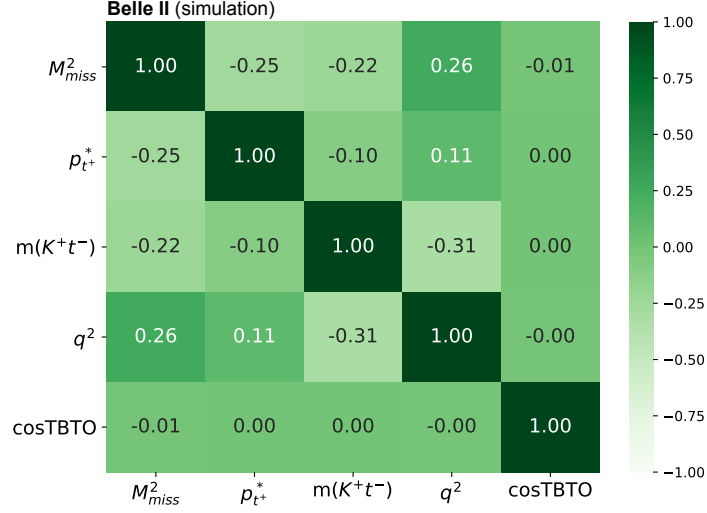


Figure C.3: Correlation matrix of BDT input observables in the simulated charge-inclusive $K\ell\pi$ signal sample.

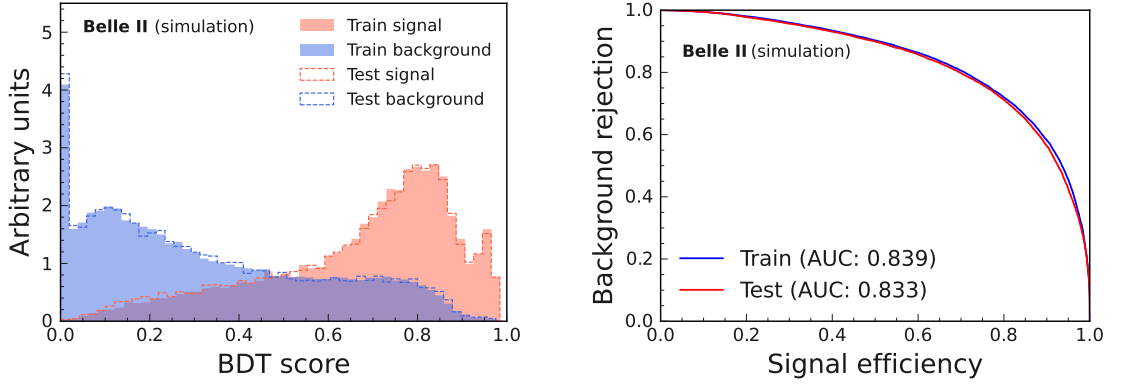


Figure C.4: (Left) BDT score on (solid) training and (dash) testing samples for simulated (red) signal and (blue) background events. (Right) Receiver operating characteristic of the BDT for (blue) training and (red) testing simulated samples.

The simulated sample is corrected with the averaged FEI-efficiency calibration factor and background is corrected with the 66% scaling obtained from E_{ECL} sideband ($E_{\text{ECL}} > 1.0$ GeV). Particle-identification and π^0 -veto efficiency corrections are not included in this study as their effects on sensitivity are marginal. The same correction criteria are applied in $K\ell\ell$ analysis for consistency. The scan ranges are $[0.4, 1.0]$ with 0.06 in BDT step-width and $[0.1, 1.5]$ GeV with 0.093 GeV E_{ECL} step-width. The optimal working points are BDT score greater than 0.7 and E_{ECL} smaller than 1.12 GeV (shown in fig. C.5).

The optimization tends to favor lower BDT values due to the broader $K\ell\pi$ signal shape, thereby accepting more background. Notably, the E_{ECL} signal resolution $K\ell\pi$ is poorer compared to that in $K\ell\ell$. This is expected as a significant portion of signal involves

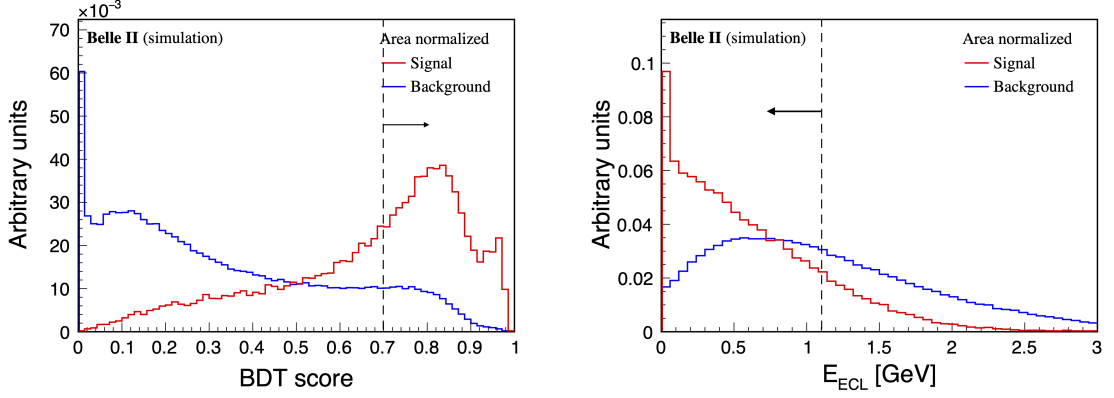


Figure C.5: Distributions of (left) BDT score and (right) E_{ECL} in charge-inclusive $K\ell\pi$ simulated samples. Distributions are normalized to unity.

π^0 s in the final state, for example $\tau^- \rightarrow \rho^-(\rightarrow \pi^-\pi^0)\nu$, $\tau^- \rightarrow \pi^-\pi^0\pi^0\nu$, and $\tau^- \rightarrow \rho(\rightarrow \pi^-\pi^0)\pi^0\nu$ decays. These π^0 are not involved in B_{sig} reconstruction and therefore contribute to E_{ECL} .

After applying the optimized selection, the signal efficiency is 2.6×10^{-4} , with 4522 events of background. The five-folded average expected upper limit at 90% confidence level is 1.6×10^{-3} . Table C.4 compares with this information with the $K\ell\ell$ sensitivity. Although the signal efficiency increases significantly in $K\ell\pi$, the background remains substantially large, resulting in an overall degradation of sensitivity compared to $K\ell\ell$.

BDT input observable	Importance
M_{miss}^2	100
p_{t+}^*	80
$m(K^+t^-)$	46
q^2	16
$\cos\text{TBTO}$	0

Table C.3: Relative importance ranking of classifier input observables in simulation. The scale is relative and spans the 0 – 100 range.

	$K\ell\ell$	Charge-inclusive $K\ell\pi$
BDT	> 0.96	> 0.7
E_{ECL}	$< 400 \text{ MeV}$	< 1120
Efficiency	1.9×10^{-5}	2.6×10^{-4}
Background	9	4522
UL at 90% C.L.	0.96×10^{-3}	1.6×10^{-3}

Table C.4: Expected upper limit of BDT-based analysis of $K\ell\pi$ final states.

C.2.5 Auxiliary checks on expected sensitivity

In this section, I study if alternative choices in the $K\ell\pi$ study would impact sensitivity.

C.2.5.1 Default selection

The $m(K^+t^-)$ mass restriction method followed by sequential one-dimensional selections provides a similar sensitivity as the mass-restricted BDT-selection for $K\ell\ell$ events. I therefore investigate if this is the case for the $K\ell\pi$ events as well.

The observables subject to optimization are M_{miss}^2 , p_{t+}^* , pion global PID, $\cos\text{TBTO}$, and E_{ECL} signal window. Figure C.6 shows distributions of these observables for events restricted in the $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ region.

The optimization scan ranges are $[0.0, 3.0] \text{ GeV}^2/c^4$ with $0.13 \text{ GeV}^2/c^4$ step-width in M_{miss}^2 , $[0.1, 2.5] \text{ GeV}/c$ with $0.07 \text{ GeV}/c$ step-width in p_{t+}^* , $[0.1, 1.1] \text{ GeV}$ with 0.05 GeV step-width in E_{ECL} , $[0.0, 0.95]$ with 0.05 step-width in pion global particle identification, and $[0.6, 0.9]$ with 0.05 step-width in $\cos\text{TBTO}$. The optimal working point is at

- M_{miss}^2 greater than $0.65 \text{ GeV}^2/c^4$,
- p_{t+}^* greater than $0.45 \text{ GeV}/c$,
- E_{ECL} smaller than 0.45 GeV ,
- pion global particle identification greater than 0.05 ,
- $\cos\text{TBTO}$ smaller than 0.8 .

After the final selection, the signal efficiency is 2.3×10^{-5} and the expected background in the signal region is 50 events. The expected upper limit at 90% confidence level is 1.9×10^{-3} (table C.5). The sensitivity of a default-like $K\ell\pi$ analysis is 19% worse than BDT-based selection, primarily due to poor discriminating performance in the opposite charge $K\ell\pi$ sample.

	$K\ell\ell$	Charge-inclusive $K\ell\pi$
M_{miss}^2	$> 1.6 \text{ GeV}/c^2$	$0.625 \text{ GeV}/c^2$
p_{t+}^*	$> 0.5 \text{ GeV}/c$	$> 0.45 \text{ GeV}/c$
E_{ECL}	$< 250 \text{ MeV}$	$< 450 \text{ MeV}$
Pion global ID	$\times$	> 0.05
$\cos\text{TBTO}$	$\times$	< 0.8
Efficiency	1.4×10^{-5}	2.3×10^{-5}
Background	4	50
UL at 90% C.L.	1.00×10^{-3}	1.9×10^{-3}

Table C.5: Expected upper limit from the default analysis of $K\ell\pi$ final states.

C.2.5.2 Impact of lepton particle-identification selection

The lepton particle-identification selections discussed in section C.2.1 are used to classify events into independent final states. There is no evidence that those choices are optimal.

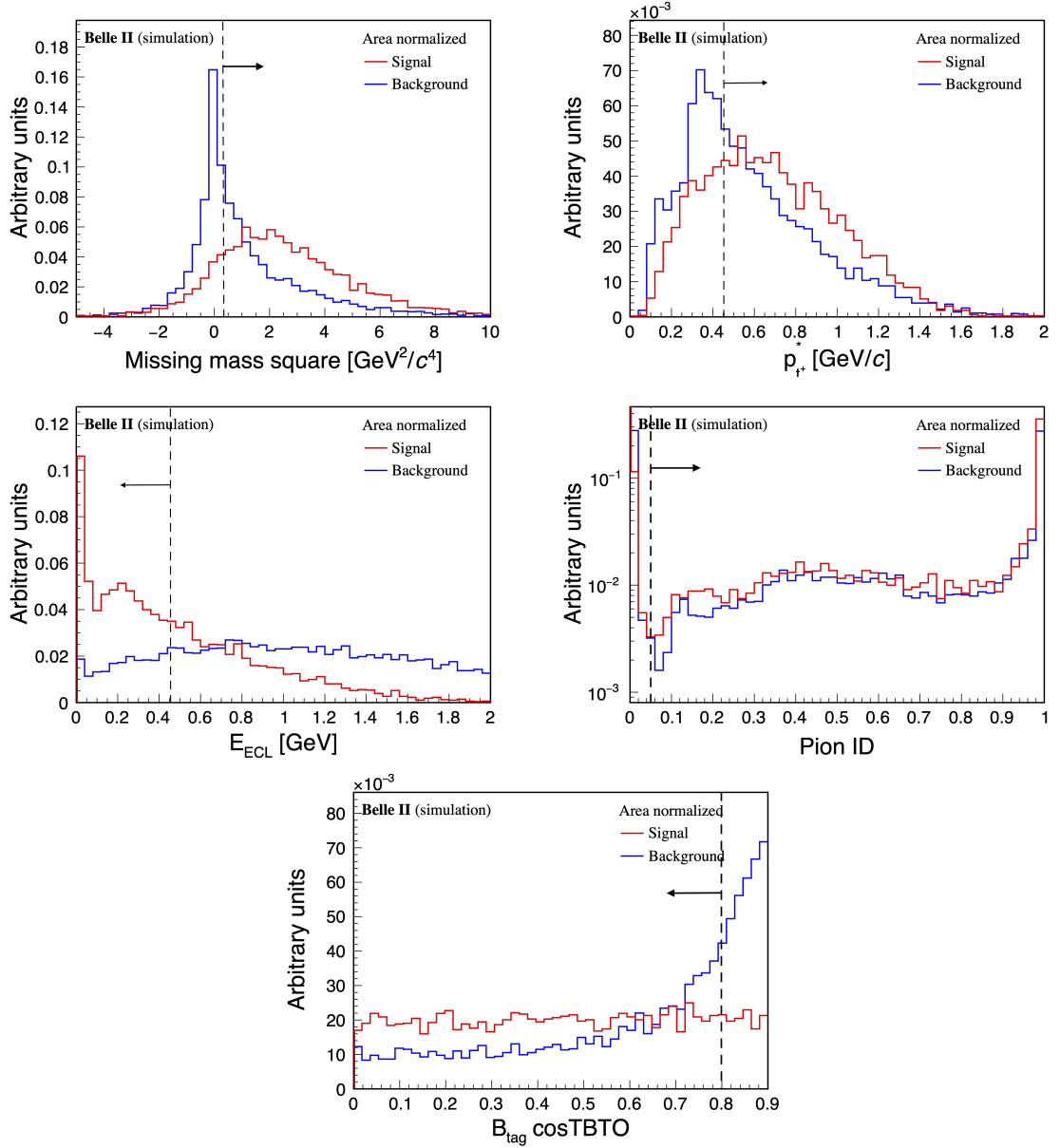


Figure C.6: Distributions of (top-left) M_{miss}^2 , (top-right) p_{t+}^* , (bottom-left) E_{ECL} , and (bottom-right) $\cos \text{TBTO}$ after applying $m(K^+t^-) > 1.9 \text{ GeV}/c^2$ selection in charge-inclusive $K\ell\pi$ simulation. Distributions are normalized to unity.

I therefore explore the possibility of sensitivity gains due to changes of pion particle-identification criteria. I set expected upper limits using the default analysis with lepton particle-identification thresholds at 0.7, 0.8, and 0.9. Table C.6 presents the expected upper limit for each scenario, with the average sensitivity derived from five-folded samples, similar to our prior estimations. The $K\ell\pi$ sensitivity is unaffected by lepton PID variations

C.2.6 Choice of discriminating observables

Another option for enhancing sensitivity could be to use an alternative signal extraction approach in which E_{ECL} is incorporated as an input observable in the BDT score and

Lepton particle identification	Expected upper limit at 90% C.L.
< 0.7	1.9×10^{-3}
< 0.8	1.9×10^{-3}
< 0.9	1.9×10^{-3}

Table C.6: Expected upper limit at 90% C.L. for various choices lepton particle-identification requirements on the signal pion in a realistic charge-inclusive $K\ell\pi$ simulated sample.

signal yield is determined by counting the excess events in the BDT score signal window. Table C.7 shows the result of the optimization of BDT score and the resulting expected five-folded averaged upper limit. Employing E_{ECL} as an input in the BDT provides a slight improvement over employing separate BDT score selection and E_{ECL} signal window.

	Without E_{ECL} input	With E_{ECL} input
UL at 90% C.L.	1.6×10^{-3}	1.5×10^{-3}

Table C.7: Expected 90% C.L. upper limit from an analysis based on ‘BDT with E_{ECL} input’ and ‘BDT without E_{ECL} input’ in a realistic charge-inclusive $K\ell\pi$ simulated sample.

C.2.7 Conclusion on $K\ell\pi$ study

My study of various choices to assess $K\ell\pi$ sensitivity and its comparison to $K\ell\ell$ sensitivity shows the equivalence between standard and BDT-based approaches across all modes, and a $K\ell\pi$ statistical sensitivity is 60% to 90% worse than in the $K\ell\ell$ analysis.

C.3 $K\pi\pi$ final state

C.3.1 Reconstruction and baseline selection

The strategy of a $K\pi\pi$ analysis is similar to that of $K\ell\ell$ and $K\ell\pi$ final states. Reconstruction and selection of B_{tag} and signal extraction are the same, except for different treatments of background suppression. After B_{tag} reconstruction, I reconstruct the signal B meson by combining a charged kaon with two oppositely charged pions. The kaon is identified with kaon-over-pion binary particle identification greater than 0.6. As for the $K\ell\pi$ analysis, inverse global electron and muon particle-identification requirements on the pion, required to be smaller than 0.9, are employed to make independent pion samples. This provides a pion purity of 76% in simulation. In addition, global pion particle-identification effectively suppresses further misreconstructed pions. After restricting the pion global particle-identification to be greater than 0.25, the pion purity increases to 90% (table C.8). This restriction is not optimized and is chosen by visual inspection (fig. C.7). The signal sample is restricted to $q^2 > 14.18 \text{ GeV}^2/c^4$ consistent with what was done for the other final states.

If multiple candidates with different final states are present. I assume leptons provide better sensitivity than pions and follow the $K\ell\ell$, $K\ell\pi$ $K\pi\pi$ ordering sorting. I randomly select a $K\ell\ell$ candidate; if it is absent, I randomly select a $K\ell\pi$ candidate. If both $K\ell\ell$ and $K\ell\pi$ are absent, I randomly choose a $K\pi\pi$ candidate.

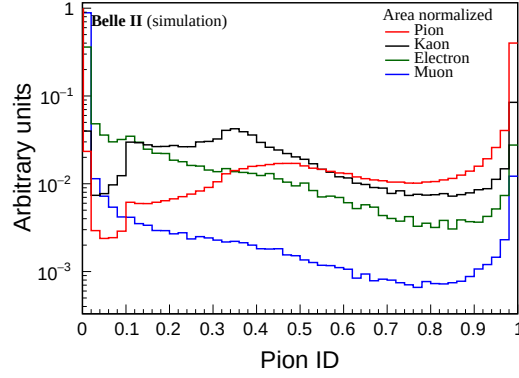


Figure C.7: Distributions of pion global particle-identification of simulated signal pions. Each color represents its true identity. Distributions are normalized to unity.

Particle type	Same charge pion as kaon (%)	Opposite charge pion as kaon (%)
Pion	89.4%	89.8%
Kaon	1.1%	0.3%
Electron	2.1%	2.2%
Muon	7.4%	7.7%

Table C.8: True particle identity of signal pion candidates in simulation.

C.3.2 Background suppression

The $K\pi\pi$ sample kinematic properties offer less discrimination than in the $K\ell\ell$ and $K\ell\pi$ cases. The only distinctive background structure is the from D^0 peak in the $m(K^+t^-)$ distribution (fig. C.8). I suppress this background by rejecting events with $1.85 < m(K^+t^-) < 1.88 \text{ GeV}/c^2$.

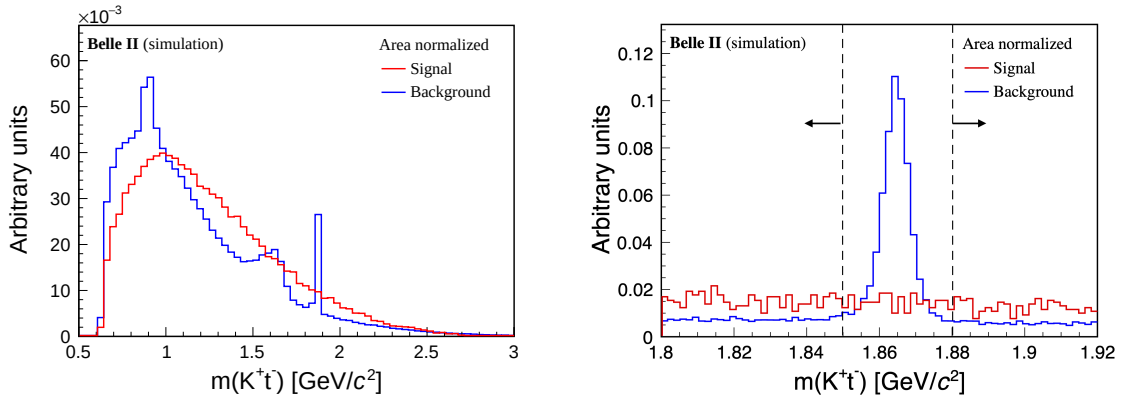


Figure C.8: Distributions of $m(K^+t^-)$ (left) in the whole spectrum and (right) near the D meson mass in simulated $K\pi\pi$ events. Distributions are normalized to unity.

The resulting background composition, provided in table C.9, shows a large contribution

of light-quark background. In addition, the B^+B^- background composition (table C.10) shows no dominating source. This makes suppression that targets a certain background inefficient.

An initial attempt at continuum suppression is made with a stringent B_{tag} cosTBTO selection. This provides a sensitivity improvement, which gets further enhanced by a multivariate classifier.

Background type	Fraction (%)
Charged	44.8%
Mixed	7.0%
Continuum	48.2%

Table C.9: Background composition after baseline selection in simulated $K\pi\pi$ sample.

Charged B backgrounds	Fraction (%)
$B^+ \rightarrow \bar{D}^{*0}\ell^+\nu$	9.9%
$B^+ \rightarrow \bar{D}^0\ell^+\nu$	4.9%
$B^+ \rightarrow \bar{D}^0 a 1^+$	0.8%
Rest	28.1%

Table C.10: Composition of B^+B^- background composition in the signal region of $K\pi\pi$ simulated sample.

C.3.3 BDT-based selection

The same BDT algorithm used in $K\ell\ell$ (section A.1), is adopted by training various event-shape observables to suppress both light quark and $B\bar{B}$ backgrounds. After searching for the best performance and small correlation among input observables, Fox-wolfram R1, cosTBTO, sphericity, M_{miss}^2 , $m(K^+t^-)$, and p_{t+}^* , provide the optimal performances. The Fox-wolfram R1 observable is the first-moment of a spacial function that captures the isotropical nature of $B\bar{B}$ events, as opposed to the collimated continuum background [83]. Figure C.9 and C.10 show relevant distributions and correlation in $K\pi\pi$ simulation, respectively. Before feeding these observables into the BDT, one must ensure that they are described well in the simulation. Signal validation would require identification of a suitable control sample in data, which is challenging and outside the scope of this exploratory study. I validate shape for backgrounds using the q^2 sideband, $q^2 < 14.18 \text{ GeV}^2/c^4$ because this has a similar background composition as the signal region. The result shows good consistency. I train and test the BDT in 10^4 simulated events. Figure C.11 illustrates the BDT output and ROC curve.

C.3.4 Selection optimization and sensitivity

The sensitivity is obtained after simultaneous optimization of the BDT score selection and the E_{ECL} signal region in a realistic simulated sample. I correct signal efficiency by applying the averaged FEI reconstruction efficiency correction (6.2). For a proper background estimation, I extrapolate the data-to-simulation integrated yield difference

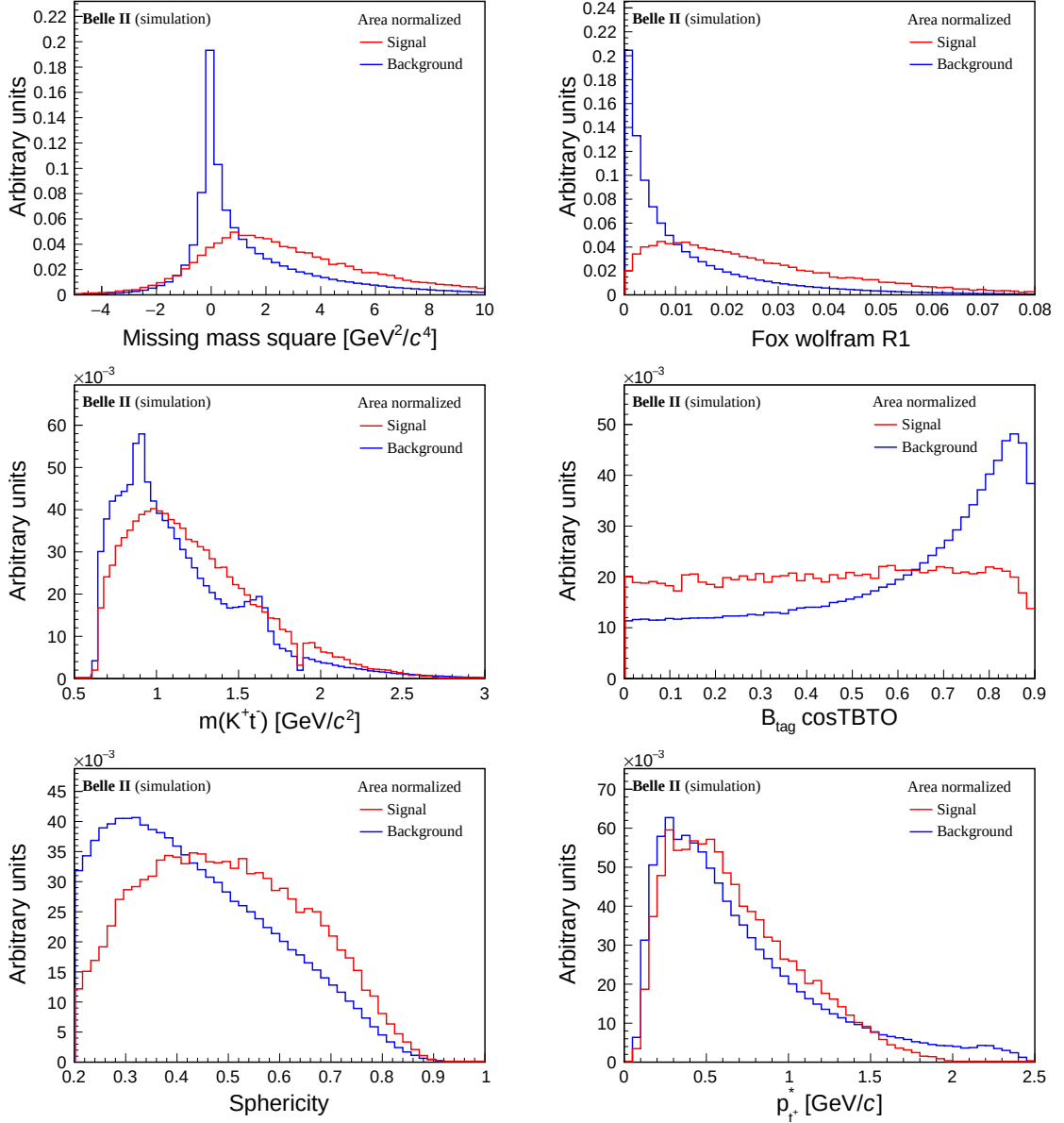


Figure C.9: Distributions of BDT input observables, (top-left) M_{miss}^2 , (top-right) Fox Wolfram R1, (middle-left) $m(K^+t^-)$, (middle-right) $\cos\text{TBTO}$, (bottom-left) sphericity, and (bottom-right) $p_{t^+}^*$, in simulated $K\pi\pi$ events. Distributions are normalized to unity.

observed in the q^2 sideband, $q^2 < 14.18 \text{ GeV}^2/c^4$, to the signal region. The derived ratio is 0.82 ± 0.01 . In this study, particle-identification and π^0 -veto efficiency corrections are not applied, which have a marginal impact for this purpose. The same correction criteria are applied in the $K\ell\ell$ analysis for consistency. The optimization scan ranges are $[0.4, 1.0]$ with 0.015 BDT score step-width and $[0.1, 2.0]$ GeV with 0.048 GeV E_{ECL} step-width.

The optimal point is at a BDT score greater than 0.63 and E_{ECL} smaller than 1.5 GeV, as shown in fig. C.12. The signal efficiency is 3.2×10^{-4} with 19426 expected background events in the signal region. The expected upper limit is 2.5×10^{-3} at the 90% confidence level, showing the worst sensitivity of $K\pi\pi$ compared to $K\ell\ell$ and $K\ell\pi$.

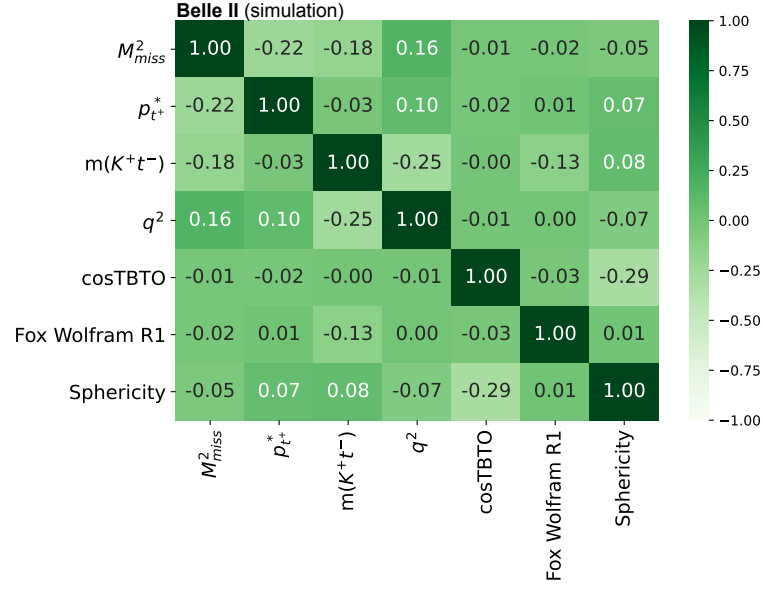


Figure C.10: Correlation matrix of BDT input observables in the simulated $K\pi\pi$ signal sample.

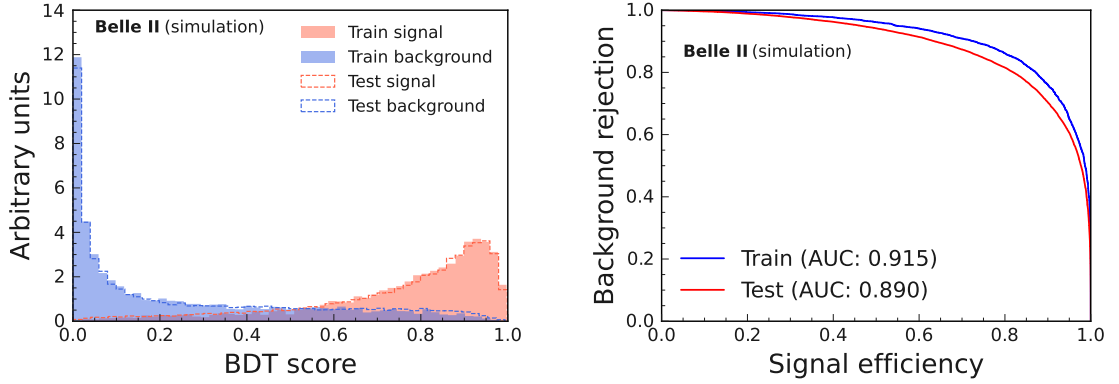


Figure C.11: (Left) BDT score on (solid) training and (dash) testing samples for simulated (red) signal and (blue) background events. (Right) Receiver operating characteristic for the BDT for (blue) training and (red) testing simulated samples.

C.4 Sensitivity comparison

The results of the investigation of sensitivity from three different $B^+ \rightarrow K^+\tau^+\tau^-$ final states are summarized in table C.11. The underlying selection and analysis choices are optimal in each category. The conclusion is that the $K\ell\ell$ sensitivity is 1.6 to 2.5 times better than the sensitivity in the other final states. I therefore choose to search for $B^+ \rightarrow K^+\tau^+\tau^-$ decays in leptonic τ final states for the first search in Belle II.

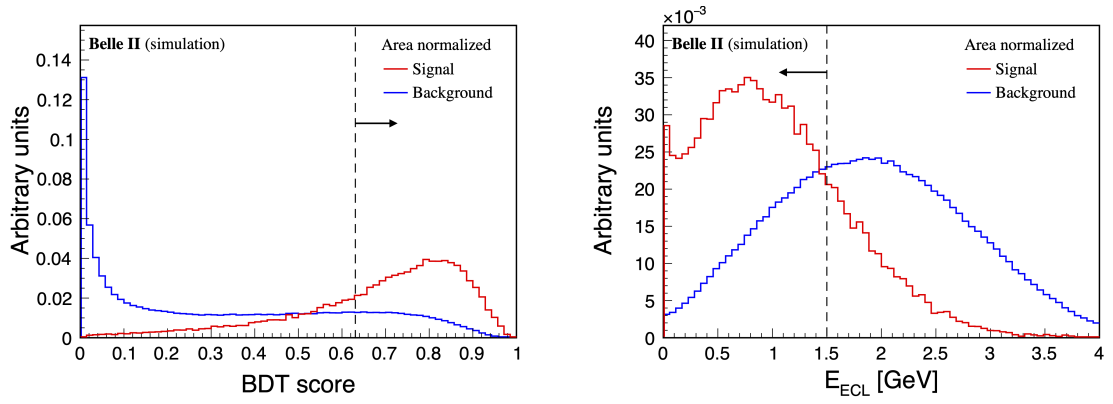


Figure C.12: Distributions of (left) BDT score and (right) E_{ECL} after baseline selection in simulated $K\pi\pi$ events. Distributions are normalized to unity.

	$K\ell\ell$	$K\ell\pi$	$K\pi\pi$
Signal efficiency	1.4×10^{-5}	2.6×10^{-4}	3.2×10^{-4}
Background	4	4522	19426
Sensitivity	1.00×10^{-3}	1.6×10^{-3}	2.5×10^{-3}

Table C.11: Sensitivity comparison between different final states based on repeating the analysis in realistic simulated samples.

Bibliography

- [1] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- [2] S. L. Glashow, *Partial-Symmetries of Weak Interactions*, Nuclear Physics **22** (1961) 579.
- [3] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367.
- [4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.
- [5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.
- [6] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.
- [7] To a good approximation, the attractive force between quarks grows linearly with distance. As two quarks are separated, putting energy into the bounded system, it becomes energetically favorable for a new quark–antiquark pair to appear, rather than extending the distance further. As a consequence new bounded states are generated instead of separated and free quarks.
- [8] G. Lüders, *Proof of the TCP Theorem*, Annals of Physics **2** (1957) 1.
- [9] W. Pauli, *Exclusion Principle, Lorentz Group, and Reversal of Space-time and Charge*, Niels Bohr and the Development of Physics (1955) 1.
- [10] C. B. Adams *et al.*, *Axion Dark Matter*, [arXiv:2203.14923 \[hep-ex\]](#).
- [11] R. L. Workman *et al.*, Particle Data Group, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.
- [12] M. Gell-Mann and M. Lévy, *The Axial Vector Current in Beta Decay*, Il Nuovo Cimento **16** (1960) 705.
- [13] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.
- [14] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.
- [15] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

- [16] A general $N \times N$ complex matrix U is defined with $2N^2$ free parameters. The unitary condition $UU^\dagger = \mathbb{1}$ eliminates N^2 degrees of freedom. We can then redefine $(N - 1)$ quark fields to absorb $(N - 1)^2$ free parameters. We end up with $2N^2 - N^2 - (2N - 1) = (N - 1)^2$ free parameters.
- [17] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.
- [18] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation*, Phys. Rev. Lett. **55** (1985) 1039.
- [19] C. Jarlskog, *A Basis Independent Formulation of the Connection Between Quark Mass Matrices, CP Violation and Experiment*, Z. Phys. C **29** (1985) 491.
- [20] I. Dunietz, O. W. Greenberg, and D. Wu, *A Priori Definition of Maximal CP Nonconservation*, Phys. Rev. Lett. **55** (1985) 2935.
- [21] Y. Amhis *et al.*, (HFLAV), *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2021*, Phys. Rev. D **107** (2023) 052008.
- [22] J. Charles *et al.*, (CKMfitter Group), *Unitarity Triangle Plot*, Eur. Phys. J. C **41** (2005) . <http://ckmfitter.in2p3.fr/>.
- [23] HFLAV Collaboration, *HFLAV: Averages of b -hadron, c -hadron, and τ -lepton properties as of 2024*, <https://hflav-eos.web.cern.ch/hflav-eos/semi/moriond24/html/RDsDsstar/RDRDs.html>, 2024. Accessed: 2024-10-19.
- [24] R. Aaij *et al.*, (LHCb collaboration), *Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \rightarrow J/\psi \tau^+ \nu_\tau)/\mathcal{B}(B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu)$* , Phys. Rev. Lett. **120** (2018) 121801.
- [25] A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski, and J. Rosiek, *Lepton-flavour violating B decays in generic Z' models*, Phys. Rev. D **92** (2015) 054013.
- [26] A. Angelescu, D. Becirevic, D. A. Faroughy, F. Jaffredo, and O. Sumensari, *Single leptoquark solutions to the B -physics anomalies*, Phys. Rev. D **104** (2021) 055017.
- [27] G. Hiller, D. Loose, and K. Schönwald, *Leptoquark Flavor Patterns and B Decay Anomalies*, JHEP **12** (2016) 027.
- [28] C. Niehoff, P. Stangl, and D. M. Straub, *Direct and indirect signals of natural composite Higgs models*, JHEP **01** (2016) 119.
- [29] K. Wilson and W. Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Commun.Math. Phys. **24** (1972) 87–106.
- [30] Quagliani, R, *LHCb presentation: measurements of R_K and R_{K^*} with the full LHCb Run 1 and 2 data*, https://indico.cern.ch/event/1187945/attachments/2569929/4431222/RX_LHC_Seminar_rquagliani.pdf, 2022.
- [31] W. G. Parrott, C. Bouchard, and C. T. H. Davies, HPQCD collaboration, *$B \rightarrow K$ and $D \rightarrow K$ form factors from fully relativistic lattice QCD*, Phys. Rev. D **107** (2023) 014510.

- [32] W. G. Parrott, C. Bouchard, and C. T. H. Davies, HPQCD collaboration, *Standard Model predictions for $B \rightarrow K\ell^+\ell^-$, $B \rightarrow K\ell_1^-\ell_2^+$ and $B \rightarrow K\nu\bar{\nu}$ using form factors from $N_f=2+1+1$ lattice QCD*, Phys. Rev. D **107** (2023) 014511. [Erratum: Phys.Rev.D 107, 119903 (2023)].
- [33] C. Cornella, D. A. Faroughy, J. Fuentes-Martín, G. Isidori, and M. Neubert, *Reading the footprints of the B-meson flavor anomalies*, JHEP **08** (2021) 050.
- [34] I. Dorsner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Kosnik, *Physics of leptoquarks in precision experiments and at particle colliders*, Phys. Rept. **641** (2016) 1–68.
- [35] R. Aaij *et al.*, (LHCb collaboration), *Observation of the decay $\Lambda_b^0 \rightarrow \Lambda_c^+ \tau^- \bar{\nu}_\tau$* , Phys. Rev. Lett. **128** (2022) 191803.
- [36] C. Alpigiani *et al.*, *Unitarity Triangle Analysis in the Standard Model and Beyond*, arXiv:1710.09644 [hep-ph].
- [37] G. Aad *et al.*, (ATLAS collaboration), *Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **125** (2020) .
- [38] A. Tumasyan *et al.*, (CMS collaboration), *Searches for additional Higgs bosons and for vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **07** (2023) 073.
- [39] J. Aebischer, G. Isidori, M. Pesut, B. A. Stefanek, and F. Wilsch, *Confronting the vector leptoquark hypothesis with new low- and high-energy data*, Eur. Phys. J. C **83** (2023) .
- [40] W. Altmannshofer *et al.*, (Belle II collaboration), *The Belle II physics book*, PTEP **2019** (2019) 123C01. [Erratum: PTEP 2020 (2020) 029201].
- [41] Y. Ohnishi *et al.*, *Accelerator Design at SuperKEKB*, Prog. Theo. Exp. Phys. (2013) 03A011.
- [42] P. Oddone, *Detector Considerations*, eConf **C870126** (1987) 423–446.
- [43] P. Raimondi, D. N. Shatilov, and M. Zobov, *Beam-Beam Issues for Colliding Schemes with Large Piwinski Angle and Crabbed Waist*, arXiv:physics/0702033.
- [44] D. R. Douglas, J. Kewisch, and R. C. York, *Betatron Function Parameterization of Beam Optics Including Acceleration*, Proceedings of the 1988 Linear Accelerator Conference (1988) .
- [45] S. Verdú-Andrés, *Luminosity geometric reduction factor from colliding bunches with different lengths*, .
- [46] T. Abe *et al.*, (Belle II Collaboration), *Belle II Technical Design Report*, arXiv:1011.0352.
- [47] C. Marinas, DEPFET, *The Belle II Pixel Detector: High precision with Low Material*, Nucl. Instrum. Meth. A **731** (2013) 31.
- [48] K. Adamczyk *et al.*, (Belle II Collaboration), *The Belle II Silicon Vertex Detector Assembly and Mechanics*, Nucl. Instrum. Meth. A **845** (2017) 38.

- [49] H. Hirano *et al.*, *A High-Resolution Cylindrical Drift Chamber for the KEK B-Factory*, Nucl. Instrum. Meth. A **455** (2000) 294.
- [50] V. Aulchenko *et al.*, *Electromagnetic Calorimeter for Belle II*, J. Phys. Conf. Ser. **587** (2015) 012045.
- [51] V. Aulchenko *et al.*, *Electromagnetic calorimeter for Belle II*, J. Phys. Conf. Ser. **587** (2015) no. 1, 012045.
- [52] K. Inami *et al.*, *TOP Counter for Particle Identification at the Belle II Experiment*, Nucl. Instrum. Meth. A **766** (2014) 5.
- [53] S. Iwata *et al.*, *Particle Identification Performance of the Prototype Aerogel RICH Counter for the Belle II Experiment*, Prog. Theo. Exp. Phys. **2016** (2016) 033H01.
- [54] A. Abashian *et al.*, Belle II Collaboration, *The K_L/μ detector subsystem for the BELLE experiment at the KEK B factory*, Nucl. Instrum. Meth. A **449** (2000) 112.
- [55] V. Bertacchi *et al.*, *Track finding at Belle II*, Computer Physics Communications **259** (2021) 107610.
- [56] T. Alexopoulos, M. Bachtis, E. Gazis, and G. Tsipolitis, *Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers*, Nucl. Instrum. Meth. A **592** (2008) 456.
- [57] R. Mankel, *Pattern recognition and event reconstruction in particle physics experiments*, Reports on Progress in Physics **67** (2004) 553.
- [58] H. Bethe, *Theory of the Passage of Fast Corpuscular Rays Through Matter*, Annalen Phys. **5** (1930) 325.
- [59] F. Bethe, *Stopping power of atoms with several electrons*, Z. Phys. **81** (1933) 363.
- [60] M. Milesi, *Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers*, Talk at CHEP 2019 .
https://indico.cern.ch/event/773049/contributions/3476156/attachments/1932835/3217425/mmilesi_mva_CHEP_07_11_19.pdf.
- [61] M. Staric, K. Inami, P. Krizan, and T. Iijima, *Likelihood analysis of patterns in a time-of-propagation (TOP) counter*, Nucl. Instrum. Meth. A **595** (2008) 252–255.
- [62] M. Staric, *Pattern recognition for the time-of-propagation counter*, Nucl. Instrum. Meth. A **639** (2011) 252–255.
- [63] J. Neyman and E. S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, Phil. Trans. R. Soc. Lond. A. **231** (1933) 289.
- [64] C. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, (HPQCD collaboration), *Standard Model predictions for $B \rightarrow K \ell^+ \ell^-$ with form factors from lattice QCD*, Phys. Rev. Lett. **111** (2013) . [Erratum: Phys. Rev. Lett. **112** (2014) 149902].
- [65] J. P. Lees *et al.*, (BaBar collaboration), *Search for $B^+ \rightarrow K^+ \tau^+ \tau^-$ at the BaBar experiment*, Phys. Rev. Lett. **118** (2017) 031802.

- [66] T. Keck *et al.*, *The Full Event Interpretation*, Computing and Software for Big Science **3** (2019) 6.
- [67] S. Brandt, C. Peyrou, R. Sosnowski, and A. Wroblewski, *The Principal Axis of Jets. An Attempt to Analyze High-Energy Collisions as Two-Body Processes*, Phys. Lett. **12** (1964) 57.
- [68] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026.
- [69] D. J. Lange, *The EvtGen Particle Decay Simulation Package*, Nucl. Instrum. Meth. A **462** (2001) 152.
- [70] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a Precision Tool for QED Corrections in Z and W Decays*, Eur. Phys. J. C **45** (2006) 97.
- [71] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A comparative study of the decays $B \rightarrow (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. D **61** (2000) 074024.
- [72] S. Agostinelli *et al.*, (GEANT4), *GEANT4 – a Simulation Toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.
- [73] B. Ward, S. Jadach, and Z. Was, *Precision Calculation for $e^+e^- \rightarrow 2f$: the KK MC Project*, Nuclear Physics B - Proceedings Supplements **116** (2003) 73.
- [74] N. L. Johnson, *Systems of frequency curves generated by methods of translation*, Biometrika **36** (1949) 149–176.
- [75] E. D. Bloom and C. Peck, *Physics with the Crystal Ball Detector*, Ann. Rev. Nucl. Part. Sci. **33** (1983) 143–197.
- [76] S. Banerjee *et al.*, Heavy Flavor Averaging Group (HFLAV), *Averages of b-hadron, c-hadron, and τ -lepton properties as of 2023*, [arXiv:2411.18639 \[hep-ex\]](https://arxiv.org/abs/2411.18639).
- [77] W. A. Rolke, A. M. Lopez, and J. Conrad, *Limits and confidence intervals in the presence of nuisance parameters*, Nucl. Instrum. Meth. A **551** (2005) 493–503.
- [78] R. Brun *et al.*, *root-project/root: v6.18/02*, <https://doi.org/10.5281/zenodo.3895860>, June, 2020.
- [79] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D **57** (1998) 3873–3889.
- [80] T. Keck, *FastBDT: A Speed-Optimized and Cache-Friendly Implementation of Stochastic Gradient-Boosted Decision Trees for Multivariate Classification*, [arXiv:1609.06119](https://arxiv.org/abs/1609.06119).
- [81] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, Ann. Statist. **29** (2001) 1189.
- [82] J. H. Friedman, *Stochastic Gradient Boosting*, Computational Statistics & Data Analysis **38** (2002) 367.
- [83] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978) 1581.