

UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS
DEPARTMENT OF PHYSICS

Ihor Prudiiev

Search for $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ decays
at Belle and Belle II

DOCTORAL THESIS

ADVISER: Prof. Dr. Peter Križan

Ljubljana, 2026

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA FIZIKO

Ihor Prudiiev

Iskanje razpadov $B^0 \rightarrow K_s^0 \tau^+ \tau^-$
pri Belle in Belle II

DOKTORSKA DISERTACIJA

MENTOR: prof. dr. Peter Križan

Ljubljana, 2026

Acknowledgements

I am very grateful to all the people who helped and supported me during this journey.

First of all, I would like to thank Prof. Peter Križan for his constant support, not only in academic matters but also in many other aspects of life. Thank you for your mentorship and care; many things would not have been possible without you. You have helped me find direction and look toward the future full of opportunities with greater confidence.

The Belle II group at Jožef Stefan Institute consists of some of the kindest people I have had the opportunity to work with, and I am grateful for the genuine friendships and warm connections that have grown from our time together. I am especially grateful to Gaetano de Marino for invaluable support for this work, as well as contribution to my scientific and personal development. What I have learned from you will not be forgotten.

As part of an international collaboration, I have had the opportunity to interact with many exceptionally talented and inspiring people. I would like to thank all my collaborators for the opportunity to work with them and to carry out this work in such an interesting environment.

During these wonderful years in Slovenia, I have met very special people. My heartfelt thanks go to my fiancée Kseniia and her younger sister Milana for their support, love, and for making my life feel brighter.

I cannot express enough gratitude to my parents, who have given me everything in this world. I am grateful beyond words to all members of my family for their unconditional love, support, and belief in me.

In loving memory of my grandparents

Search for $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ decays at Belle and Belle II

ABSTRACT

We present the first search for the rare decay $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ using approximately 1 ab^{-1} of e^+e^- collision data collected with the Belle and Belle II experiments at the KEKB and SuperKEKB colliders between 1999–2009 and 2019–2022, respectively. The decay proceeds via a $b \rightarrow s \tau^+ \tau^-$ flavor-changing neutral current transition, which is highly suppressed in the Standard Model and sensitive to possible contributions from new heavy mediators with preferential couplings to third-generation leptons. Since τ lepton decays involve neutrinos which escape detection, part of the decay information is inevitably lost, making the experimental reconstruction of the $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ particularly challenging.

We fully reconstruct the companion non-signal $\bar{B}^0$ meson in hadronic modes, enabling an exclusive study of the signal-side decay. The signal B^0 candidates are reconstructed in τ decay channels with a single charged particle, including leptonic and hadronic modes. A multivariate classifier is trained to enhance the discrimination between signal and background, and a simultaneous binned maximum-likelihood fit to the classifier output is performed across five orthogonal categories based on $\tau^+ \tau^-$ decay modes for both Belle and Belle II datasets.

No significant excess above the expected background is observed. We set the only upper limit to date on the branching fraction, $\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-) < 0.83 \times 10^{-3}$ at 90% confidence level.

Keywords: Belle, Belle II, Electroweak Penguins, Missing energy, B -anomalies, $b \rightarrow s \ell \ell$

Iskanje razpadov $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ pri Belle in Belle II

IZVLEČEK

Pričujoče doktorsko delo predstavlja prvo iskanje redkega razpada $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ z uporabo približno 1 ab^{-1} podatkov trkov e^+e^- , zbranih z eksperimentoma Belle in Belle II na trkalnikih KEKB in SuperKEKB v obdobjih 1999–2009 oziroma 2019–2022. Razpad poteka prek prehoda $b \rightarrow s \tau^+ \tau^-$, ki je v Standardnem modelu močno potlačen in zato občutljiv na morebitne prispevke novih težkih posrednikov z močnejšimi sklopitvami z leptoni tretje generacije. Prisotnost dveh leptonov τ v končnem stanju vodi do več nevtrinov in s tem do znatne manjkajoče gibalne količine, zaradi česar ta razpadni kanal spada med najzahtevnejše meritve prehodov $b \rightarrow s \ell \ell$.

V vsakem dogodku rekonstruiramo asociirani mezon $\bar{B}^0$ v hadronskih razpadnih načinih, kar omogoča ekskluzivno preučevanje razpada na preostali signalni strani dogodka. Kandidate za signalni B^0 rekonstruiramo v več razpadnih načinih, kjer lepton τ razpadeta v en sam nabit delec, ki je lahko lepton ali hadron. Za izboljšanje ločljivosti med signalom in ozadjem uporabimo multivariatni klasifikator ter izvedemo metodo maksimalne zanesljivosti za prilaganje diskretnih porazdelitev dogodkov njegovega izhoda v petih ortogonalnih kategorijah, določenih glede na razpadne načine $\tau^+ \tau^-$.

V podatkih ne opazimo statistično pomembnega presežka nad pričakovanim ozadjem. Postavimo trenutno edino zgornjo mejo na razpadno razmerje: $\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-) < 0.83 \times 10^{-3}$ pri 90% intervalu zaupanja.

Ključne besede: Belle, Belle II, elektrošibki pingvini, manjkajoča energija, anomalije v razpadih mezonov B , $b \rightarrow s \ell \ell$

Contents

1	Introduction	15
1.1	The Standard Model of particle physics	16
1.2	Flavor and quark mixing	17
1.3	Flavor-changing neutral currents	18
1.4	Status of B anomalies and $b \rightarrow s\tau^+\tau^-$ searches	19
1.5	New-physics interpretations of $b \rightarrow s\tau^+\tau^-$	21
1.6	Motivation for the search for $B^0 \rightarrow K_S^0\tau^+\tau^-$	22
2	The Belle and the Belle II experiments	25
2.1	The KEKB and SuperKEKB accelerators	25
2.1.1	Physics motivation and the $\Upsilon(4S)$ environment	25
2.1.2	The KEKB collider	26
2.1.3	The SuperKEKB collider	28
2.1.4	The nano-beam collision scheme	29
2.1.5	Luminosity performance and beam-induced backgrounds	31
2.1.6	Impact on Belle and Belle II physics	31
2.2	The Belle and Belle II detectors	32
2.2.1	Coordinate system and global detector layout	32
2.2.2	Tracking and vertex reconstruction	32
2.2.3	Charged-particle identification	37
2.2.4	Electromagnetic calorimetry and neutral particles	40
2.2.5	K_L^0 and muon identification	41
2.2.6	Trigger and data acquisition	41
2.2.7	Beam-induced backgrounds and Belle II upgrades	42
2.2.8	Performance summary	42
2.3	Summary	43
3	Analysis strategy	45
3.1	Dataset, Software and Monte Carlo simulation	45
3.1.1	Frameworks and event generators	46
3.1.2	Monte Carlo samples	47
3.2	Multivariate analysis	47
3.3	Hadronic B -tag reconstruction	47
3.4	Template-based maximum-likelihood fit	49
3.5	Statistical inference	49
4	Event selection	51
4.1	Partner B reconstruction	51
4.2	Signal B reconstruction	52

4.3	Rest of Event	57
4.4	Event categorization	58
5	Calibrations	65
5.1	Corrections to the simulation	65
5.1.1	B -tagging efficiency corrections	65
5.1.2	Tracking efficiency calibration	66
5.1.3	K_S^0 efficiency calibration	67
5.1.4	π^0 efficiency calibration	67
5.1.5	Lepton identification calibration	68
5.1.6	Correction of branching fractions of signal-side τ leptons . . .	68
5.2	Data-MC comparison in background enriched samples	68
5.2.1	Off-resonance	69
5.2.2	K_S^0 sideband	69
5.2.3	M_{bc} sideband	72
5.3	Summary	73
6	Background suppression and classification	75
6.1	Background characterization and vetoes	75
6.2	BDT training	76
6.2.1	BDT features	79
6.3	Background characterization	83
7	Validation	89
7.1	K_S^0 and M_{bc} sidebands	89
7.2	D -vetoed samples	90
8	Signal extraction	97
8.1	Fitting strategy	97
8.2	Systematic uncertainties	102
8.3	Results on simulation	109
8.4	Results on data	114
8.4.1	Measurement in control sample	114
8.4.2	Belle II partial result	114
8.4.3	Belle partial result	115
8.4.4	Final result	117
8.5	Isospin average	122
9	Conclusion and outlook	125
10	Bibliography	127
Appendix A Data management plan		133
Appendix B BDT training variables		135
B.1	BDT training - additional plots	140
Appendix C BDT training variables check in K_S^0 sideband		145
Appendix D BDT training variables check in M_{bc} sideband		151

Appendix E	Signal composition in Belle II MC	157
Appendix F	Signal composition in Belle MC	161
Appendix G	K_s^0 SR and SB comparison for Belle II MC	165
Appendix H	K_s^0 SR and SB comparison for Belle MC	167
Appendix I	Disjoint fits to Belle and Belle II MC samples	169
I.1	Fit to Belle MC	169
I.2	Fit to Belle II MC	174
Appendix J	“Post-fit” distributions of the BDT input features in signal enhanced region	179
Razširjeni povzetek v slovenskem jeziku		191
Uvod in fizikalna motivacija		191
Eksperimentalno okolje Belle in Belle II		192
Rekonstrukcija dogodkov in začetni izbor		195
Kategorije signalnih kandidatov		198
Popravki simulacije in validacija ozadij		199
Odstranjevanje ozadja z BDT		201
Statistični model in sistematične negotovosti		202
Rezultati		204
Izospinsko povprečenje		205
Sklep		205
List of publications related to this doctoral thesis		207

1. Introduction

The Standard Model (SM) of particle physics provides a remarkably successful description of the known elementary particles and their interactions over a wide range of energies. It has been tested in thousands of measurements with impressive precision, and the discovery of the Higgs boson at the Large Hadron Collider (LHC) completed its particle content [1, 2, 3]. Nonetheless, the SM is widely regarded as an effective theory, which does not account for dark matter nor includes gravity. Furthermore, it cannot explain the observed matter–antimatter asymmetry of the Universe, and it does not provide a dynamical origin for the hierarchy of fermion masses and mixing angles [3]. These open questions strongly suggest the existence of physics beyond the Standard Model (BSM).

Indirect searches—precision measurements of processes in which virtual, heavy BSM particles can contribute—offer a powerful and complementary approach to direct searches at high-energy colliders [4]. In particular, the study of flavor-changing processes involving heavy quarks and leptons can be sensitive to energy scales far beyond those directly accessible in current experiments. This thesis focuses on such an indirect probe: a search for the rare decay $B^0 \rightarrow K_S^0 \tau^+ \tau^-$, which proceeds via a flavor-changing neutral current (FCNC) transition $b \rightarrow s \tau^+ \tau^-$ and is highly suppressed in the SM [5]. Charge-conjugate processes are implied throughout this work.

Over the past decade, measurements of semileptonic B decays have revealed intriguing deviations from SM expectations, often referred to as “ B anomalies”. These include tension in observables testing lepton-flavor universality (LFU) in charged-current $b \rightarrow c \tau \nu$ and neutral-current $b \rightarrow s \mu^+ \mu^-$ transitions [4, 6]. Many BSM interpretations of these anomalies predict an enhanced coupling to third-generation leptons, and consequently sizable enhancements of $b \rightarrow s \tau^+ \tau^-$ decay rates, while remaining consistent with existing constraints [5, 7, 8]. However, $b \rightarrow s \tau^+ \tau^-$ processes remain heavily unexplored experimentally [9, 10], due to the challenge of reconstructing final states with two τ leptons and multiple neutrinos.

In this chapter, we introduce the theoretical and experimental context of this work. We begin with a brief review of the SM and the flavor sector, then describe FCNC processes and their role as probes of new physics. We summarize the present status of B anomalies and discuss theoretical motivations for enhanced $b \rightarrow s \tau^+ \tau^-$ transitions. Finally, we motivate the search for the specific decay $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ at Belle and Belle II.

The remainder of this thesis is organized as follows. Chapter 2 describes the KEKB and SuperKEKB colliders and the Belle and Belle II detectors, with emphasis on the experimental features relevant for rare decays with missing energy. Chapter 3 introduces the general analysis strategy, including the datasets, simulated samples, hadronic B tagging, multivariate classification, and statistical meth-

ods used for signal extraction. Chapter 4 describes the reconstruction and selection of $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ candidates, including the definition of the signal categories. Chapter 5 discusses the corrections applied to the simulation and the calibration of the dominant data-MC differences. Chapter 6 presents the background composition and the multivariate background-suppression procedure. Chapter 7 documents the validation of the analysis in data control regions and sidebands before the signal-enhanced region is inspected. Chapter 8 describes the signal-extraction fit, the treatment of systematic uncertainties, the results obtained with Belle and Belle II data, and the isospin average with the charged pseudoscalar channel. Finally, Chapter 9 summarizes the results and discusses their implications for future searches for $b \rightarrow s \tau^+ \tau^-$ transitions.

1.1 The Standard Model of particle physics

The SM is a relativistic quantum field theory based on the gauge symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y, \quad (1.1)$$

which describes the strong, weak, and electromagnetic interactions, respectively [3]. Here, C denotes color charge, associated with the strong interaction; L denotes the left-handed chiral component of the weak interaction; and Y denotes weak hypercharge, the quantum number associated with the $U(1)$ gauge symmetry. The fundamental degrees of freedom are quantum fields; their excitations correspond to the particles observed in experiments.

Matter is composed of spin- $\frac{1}{2}$ fermions, which are arranged in three generations. Each generation contains two quarks and two leptons, grouped into left-handed weak-isospin doublets and right-handed singlets. In the quark sector, the left-handed fields form the doublets

$$Q_L^{(1)} = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad Q_L^{(2)} = \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \quad Q_L^{(3)} = \begin{pmatrix} t_L \\ b_L \end{pmatrix}. \quad (1.2)$$

Here u, c, t are the up-type quarks and d, s, b are the down-type quarks. In the lepton sector, the corresponding left-handed doublets are

$$L_L^{(1)} = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \quad L_L^{(2)} = \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \quad L_L^{(3)} = \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}. \quad (1.3)$$

The right-handed components of quarks and charged leptons are electroweak singlets. All quarks carry color charge and interact via the strong interaction, while leptons are colorless and do not participate in QCD interactions [3].

The interactions between fermions are mediated by spin-1 gauge bosons, which are the excitations of the gauge fields associated with the SM symmetry group. The eight gluons g mediate the strong interaction described by quantum chromodynamics (QCD); the $W^\pm$ and Z^0 bosons mediate the weak interaction; and the photon γ mediates the electromagnetic interaction. The weak and electromagnetic interactions are unified into the electroweak theory, in which the physical gauge bosons arise as linear combinations of the $SU(2)_L$ and $U(1)_Y$ gauge fields [3].

Gauge invariance forbids explicit mass terms for the fermions and gauge bosons. Mass generation is achieved via the Brout–Englert–Higgs mechanism: a complex

scalar $SU(2)_L$ doublet field Φ acquires a nonzero vacuum expectation value (vev), spontaneously breaking the electroweak symmetry down to $U(1)_{\text{em}}$ [3]. Three of the four real degrees of freedom in Φ provide the longitudinal polarizations of the $W^\pm$ and Z^0 bosons, which thereby acquire masses, while the remaining degree of freedom corresponds to a physical scalar, the Higgs boson. Yukawa interactions between Φ and the fermion fields generate fermion masses proportional to their Yukawa couplings once the Higgs field develops its vev .

Despite its success, the SM leaves many fundamental questions open. It does not explain the pattern of fermion masses and mixing angles, the tiny but nonzero neutrino masses, the observed baryon-antibaryon asymmetry of the Universe, or the nature of dark matter [3]. These shortcomings motivate searches for BSM physics, among which precision studies of flavor-changing processes play a central role.

1.2 Flavor and quark mixing

The term *flavor* denotes the replication of fermion generations with identical gauge quantum numbers but different masses. In the quark sector, quarks of a given electric charge mix under the weak interaction: the weak interaction eigenstates are not aligned with the mass eigenstates. This misalignment is encoded in the Cabibbo–Kobayashi–Maskawa (CKM) matrix [11, 12].

In the SM, charged-current weak interactions couple left-handed up-type and down-type quarks via the $W^\pm$ boson. The corresponding piece of the Lagrangian reads

$$\mathcal{L}_W = -\frac{g}{\sqrt{2}} (\bar{u}'_L, \bar{c}'_L, \bar{t}'_L) \gamma^\mu V_{\text{CKM}} \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix} W_\mu^+ + \text{h.c.}, \quad (1.4)$$

where primed fields denote weak eigenstates and V_{CKM} is the 3×3 unitary CKM matrix,

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (1.5)$$

Unitarity of V_{CKM} ensures the absence of flavor-changing neutral currents at tree level and reduces the number of independent parameters. For three generations, the CKM matrix can be parametrised in terms of three real mixing angles and one irreducible complex phase [13]. This phase allows weak-interaction amplitudes to differ from those of the corresponding CP -conjugate processes, giving rise to observable asymmetries between certain decays of particles and antiparticles.

It is often convenient to express V_{CKM} using the Wolfenstein parametrisation, an expansion in powers of the small parameter $\lambda \approx 0.23$, related to the Cabibbo angle [14]. In this parametrisation,

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \quad (1.6)$$

where A , ρ , and η are additional real parameters. The nonzero value of η implies CP -violating couplings.

Global fits to flavor data show that the CKM mechanism provides a consistent description of quark mixing and CP violation, but they still allow for sizable BSM contributions, especially in loop-induced processes [4].

1.3 Flavor-changing neutral currents

Flavor-changing neutral currents are processes in which the flavor of a quark changes while its electric charge remains the same, such as $s \rightarrow d$, $b \rightarrow s$, or $b \rightarrow d$ transitions. In the SM, FCNCs are absent at tree level because neutral gauge bosons, such as the photon and the Z boson, do not change quark flavor. Flavor changes occur through the charged weak interaction mediated by the $W^\pm$ bosons. FCNCs can therefore only occur via higher-order loop diagrams, such as penguin and box diagrams with internal W bosons and up-type quarks, as shown in Fig. 1.1.

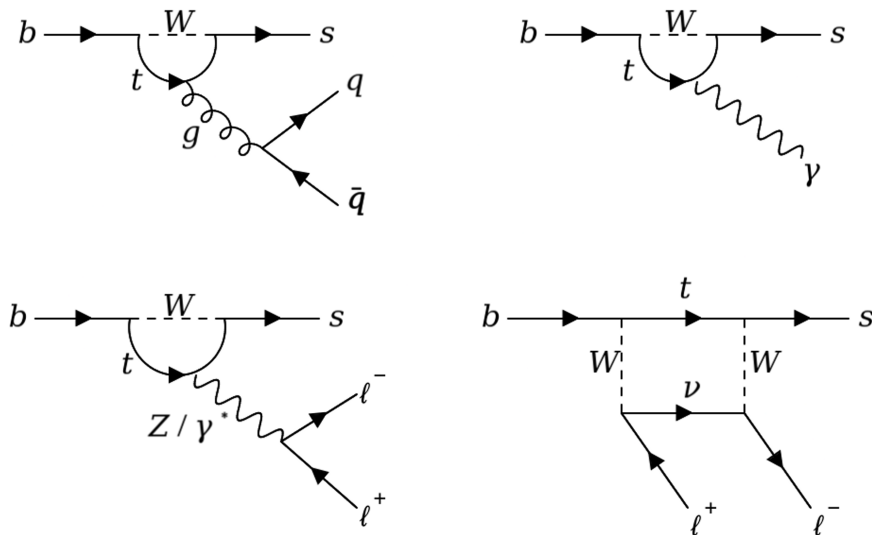


Figure 1.1: Examples of Feynman diagrams for FCNC $b \rightarrow s$ processes in the SM. Different name conventions are adopted depending on the final states: gluonic penguin (top left), radiative penguin (top right), electroweak penguin (bottom left) and box diagram (bottom right).

For b -quark decays, FCNC transitions like $b \rightarrow s\ell^+\ell^-$ (with $\ell = e, \mu, \tau$) and $b \rightarrow s\nu\bar{\nu}$ are strongly suppressed compared to tree-level charged-current transitions $b \rightarrow c\ell\nu$. The corresponding branching fractions are typically at the level of 10^{-7} – 10^{-6} in the SM, depending on the specific exclusive mode [4, 5]. Theoretical predictions for these processes involve short-distance Wilson coefficients, calculable in perturbation theory, and hadronic form factors, which encode the nonperturbative QCD dynamics and are obtained from lattice QCD or other nonperturbative methods [3].

The loop-induced nature of FCNCs makes them sensitive probes of BSM physics: new heavy particles carrying appropriate quantum numbers can enter the loops and modify the effective couplings. This can affect decay rates, angular observables, and CP asymmetries [4]. The effect is particularly clean in observables where hadronic uncertainties partially or completely cancel, such as certain ratios testing LFU.

So far, most experimental attention has focused on $b \rightarrow s\ell^+\ell^-$ transitions with

$\ell = e, \mu$, for which many exclusive and inclusive channels have been measured [3]. In contrast, decays involving $b \rightarrow s\tau^+\tau^-$ are experimentally much more challenging, because the τ leptons decay promptly into final states with additional neutrinos, leading to significant missing energy. As a result, the SM predictions for exclusive $b \rightarrow s\tau^+\tau^-$ modes, such as $B \rightarrow K^{(*)}\tau^+\tau^-$, are of the order of 10^{-7} [5].

1.4 Status of B anomalies and $b \rightarrow s\tau^+\tau^-$ searches

In the SM, LFU implies that the electroweak couplings to the three lepton families are identical up to corrections from the lepton masses [3]. LFU can be tested experimentally by forming ratios of branching fractions that differ only by the lepton flavor in the final state. Examples include

$$R_{D^{(*)}} = \frac{\mathcal{B}(\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau)}{\mathcal{B}(\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell)}, \quad R_{K^{(*)}} = \frac{\mathcal{B}(B \rightarrow K^{(*)}\mu^+\mu^-)}{\mathcal{B}(B \rightarrow K^{(*)}e^+e^-)}, \quad (1.7)$$

where $\ell = e, \mu$. In these ratios, many hadronic uncertainties cancel, making them theoretically clean observables [4].

Over the past years, measurements of $R_{D^{(*)}}$ by BaBar, Belle, and LHCb (see Fig. 1.2), as well as related observables such as $R_{J/\psi}$, showed deviations from the SM predictions at the level of a few standard deviations when combined [6].

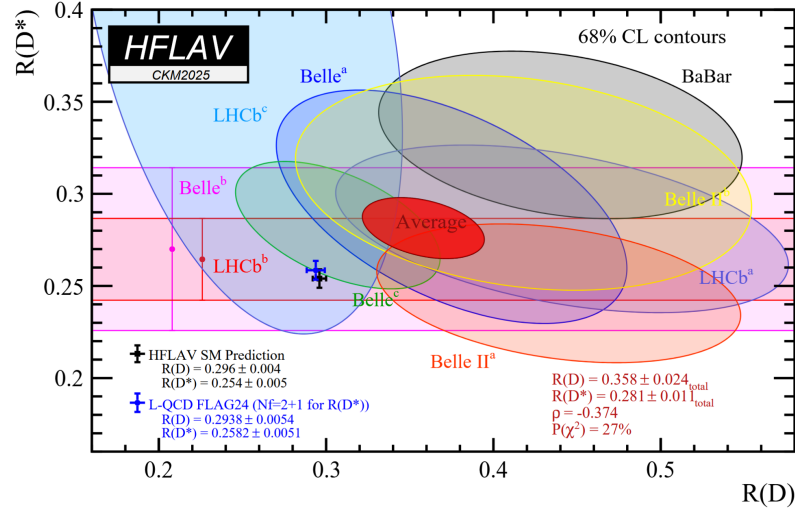


Figure 1.2: Status of R_D and R_{D^*} measurements. The red ellipse represents the world average, including both the separate measurements of $R_{D^{(*)}}$ and results from simultaneous measurements of $R_{(D)}$ and $R_{(D^*)}$. The black point with error bars indicates the average of SM calculations. Plot from HFLAV averages [6].

In the neutral-current sector, earlier measurements of $R_{K^{(*)}}$ exhibited tensions with the SM. However, recent results from LHCb are consistent with LFU within uncertainties, significantly reducing the evidence for a violation in these observables. Despite this, the global picture of $b \rightarrow s\ell^+\ell^-$ transitions still shows persistent tensions with the SM. In particular, deviations are observed in angular observables such as P'_5 in $B \rightarrow K^*\mu^+\mu^-$ decays and in the differential branching fractions of several $b \rightarrow s\mu^+\mu^-$ modes [15, 16] (see Fig. 1.3).

In addition, recent measurements of $B \rightarrow K\nu\bar{\nu}$ decays indicate a branching fraction above the SM expectation [18]. When interpreted in a model-independent

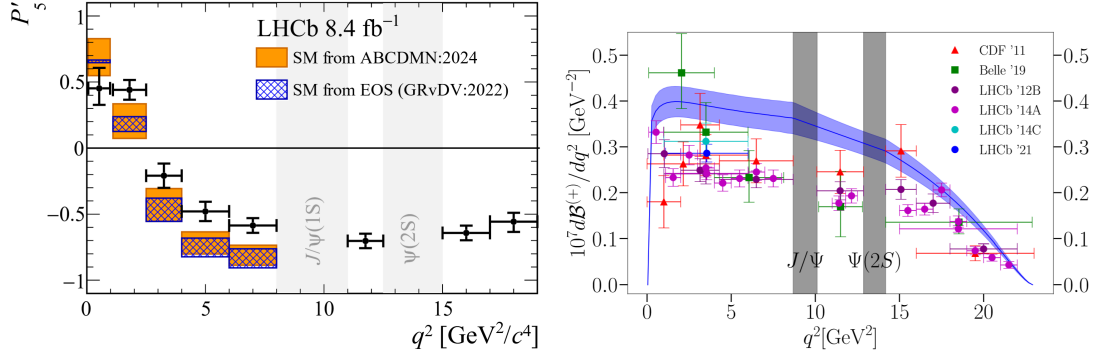


Figure 1.3: Left: Measurement of the angular observable P'_5 in $B \rightarrow K^* \mu^+ \mu^-$ decays compared to SM predictions. Right: Differential branching fractions for $b \rightarrow s \mu^+ \mu^-$ transitions. Figures adapted from Refs. [15, 17].

framework together with constraints from $b \rightarrow s \ell^+ \ell^-$ processes, these results favor scenarios with non-universal couplings that can be enhanced for third-generation leptons [5].

Taken together, while no single observable currently provides a conclusive indication of physics beyond the SM, the pattern of deviations continues to motivate models with lepton-flavor non-universal interactions, particularly involving τ leptons.

From the theory side, several global fits to flavor data have considered effective-operator modifications of the $b \rightarrow s \ell^+ \ell^-$ Hamiltonian [4, 5]. Scenarios that simultaneously address the charged-current and neutral-current anomalies often involve new heavy particles, such as leptoquarks or Z' bosons, with couplings that preferentially involve τ leptons [7, 8]. In these frameworks, the effective couplings for $b \rightarrow s \tau^+ \tau^-$ transitions can be significantly enhanced compared to the SM, leading to potentially large increases in the branching fractions of modes such as $B \rightarrow K^{(*)} \tau^+ \tau^-$, $B_s \rightarrow \tau^+ \tau^-$, or $\Lambda_b \rightarrow \Lambda \tau^+ \tau^-$ [5].

Experimentally, $b \rightarrow s \tau^+ \tau^-$ processes remain very weakly constrained. The presence of at least two τ leptons and multiple neutrinos leads to severe missing energy in the final state and makes full kinematic reconstruction impossible. As a consequence, only a few searches have been performed so far. The BaBar experiment has set an upper limit on the branching fraction $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)$ at the level of a few 10^{-3} using hadronic B tagging and leptonic τ decays [19]. The Belle experiment has searched for $B^0 \rightarrow K^{*0} \tau^+ \tau^-$ and set a similar upper limit [20], while Belle II improved it by about a factor of two [21], and LHCb has searched for $B_s^0 \rightarrow \tau^+ \tau^-$ and $B^0 \rightarrow \tau^+ \tau^-$, also obtaining upper limits at $\mathcal{O}(10^{-3})$ [22]. Recent searches [9, 10] have set upper limits on the branching fractions of $B^+ \rightarrow K^+ \tau^+ \tau^-$, $B^0 \rightarrow K^{*0} \tau^+ \tau^-$ decays at the level of a few 10^{-4} , significantly improving upon earlier results. These limits still remain about three orders of magnitude above the SM expectations [5] and summarized in Fig. 1.4, leaving ample room for BSM contributions.

The large gap between the SM predictions and current experimental constraints implies that sizeable BSM effects in $b \rightarrow s \tau^+ \tau^-$ transitions are still allowed. Improving the experimental sensitivity to these modes is therefore a high priority in the flavor physics programme of B factories and LHC experiments.

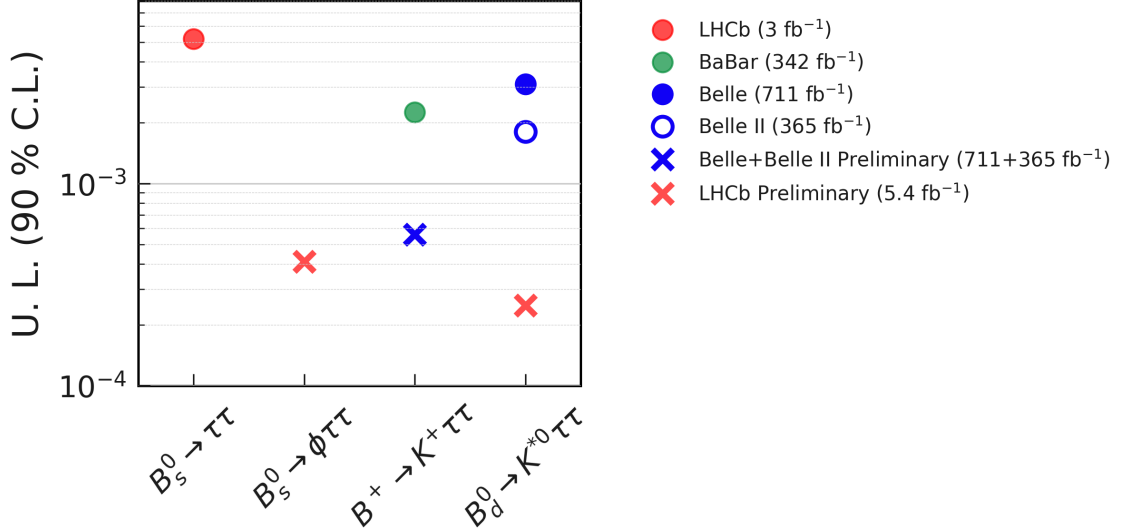


Figure 1.4: Upper limits on $b \rightarrow s\tau\tau$ transitions. Results from Refs. [9, 10, 19, 20, 21, 22]

1.5 New-physics interpretations of $b \rightarrow s\tau^+\tau^-$

Rare $b \rightarrow s\ell^+\ell^-$ (with $\ell = e, \mu, \tau$) transitions are conveniently described in the framework of an effective field theory, where heavy degrees of freedom are integrated out and their effects are encoded in short-distance Wilson coefficients multiplying local operators. The effective Hamiltonian governing these processes can be written as

$$\mathcal{H}_{\text{eff}}(b \rightarrow s\ell\ell) = -\frac{4G_F}{\sqrt{2}}V_{tb}V_{ts}^* \sum_i C_i^\ell \mathcal{O}_i^\ell + \text{h.c.}, \quad (1.8)$$

where G_F is the Fermi constant, $V_{tb}V_{ts}^*$ are elements of the Cabibbo–Kobayashi–Maskawa matrix, and C_i^ℓ are the Wilson coefficients encoding short-distance physics.

The dominant semileptonic operators relevant for $b \rightarrow s\ell^+\ell^-$ transitions are

$$\mathcal{O}_9^\ell = \frac{\alpha}{4\pi}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \ell), \quad \mathcal{O}_{10}^\ell = \frac{\alpha}{4\pi}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \gamma_5 \ell). \quad (1.9)$$

In the SM, these operators are generated by loop-level electroweak penguin and box diagrams. Physics beyond the SM can modify their Wilson coefficients and generate additional interactions via chirality-flipped operators $\mathcal{O}'_{9,10}^\ell$, obtained by replacing P_L with P_R in the quark current, as well as scalar and pseudoscalar operators

$$\mathcal{O}_S^\ell = \frac{\alpha}{4\pi}(\bar{s}P_R b)(\bar{\ell}\ell), \quad \mathcal{O}_P^\ell = \frac{\alpha}{4\pi}(\bar{s}P_R b)(\bar{\ell}\gamma_5 \ell). \quad (1.10)$$

Here α is fine-structure constant and $P_{L,R}$ are the left- and right-handed projection operators,

$$P_L = \frac{1 - \gamma_5}{2}, \quad P_R = \frac{1 + \gamma_5}{2}. \quad (1.11)$$

In the Standard Model, the Wilson coefficients C_i^ℓ are lepton-flavor universal up to small corrections and are dominated by loop-level contributions. Physics beyond the Standard Model can modify their magnitude, leading to shifts

$$C_i^\ell = C_i^{\text{SM}} + C_i^{\text{NP},\ell}. \quad (1.12)$$

In particular, non-universal contributions, where the shifts for $\ell = \tau$ differ from those for $\ell = e, \mu$, naturally lead to enhanced $b \rightarrow s\tau^+\tau^-$ rates [4, 5].

Prominent examples include:

- **Leptoquark models:** Leptoquarks are hypothetical bosons that couple directly to a quark and a lepton. Depending on their spin and gauge quantum numbers, they can contribute at tree level to both $b \rightarrow c\tau\nu$ and $b \rightarrow s\ell^+\ell^-$ transitions (see Fig. 1.5, left). In particular, vector leptoquark models, such as those involving a U_1 leptoquark, have been shown to accommodate the pattern of anomalies while predicting enhancements of up to three orders of magnitude in $b \rightarrow s\tau^+\tau^-$ branching fractions compared to the SM [5, 7].
- **Z' models:** Additional neutral gauge bosons Z' with flavor-changing couplings to quarks and non-universal couplings to leptons can modify the effective $b \rightarrow s\ell^+\ell^-$ couplings (see Fig. 1.5, right). If the couplings to τ leptons are larger than those to electrons and muons, sizable enhancements of $b \rightarrow s\tau^+\tau^-$ rates are expected [8].

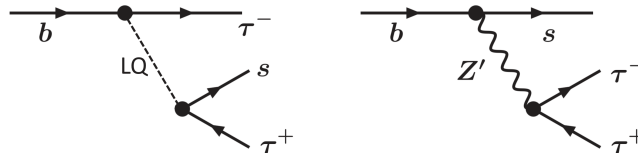


Figure 1.5: Examples of Feynman diagrams for $b \rightarrow s\tau^+\tau^-$ tree-level transitions in leptoquark (left) and Z' (right) BSM scenarios.

A common feature of many of these scenarios is that they predict strong correlations between the observables in the $b \rightarrow c\tau\nu$ and $b \rightarrow s\ell^+\ell^-$ sectors and the rates of rare $b \rightarrow s\tau^+\tau^-$ decays [5, 7]. In New Physics fits that accommodate the current data, the branching fractions for $B \rightarrow K^{(*)}\tau^+\tau^-$ and $B \rightarrow K\tau^+\tau^-$ can be enhanced by up to two or three orders of magnitude over their SM values (see Fig. 1.6), bringing them potentially within reach of current or near-future experiments at B factories [5].

Therefore, any measurement or improved upper limit on $b \rightarrow s\tau^+\tau^-$ decays places strong constraints on such BSM explanations and helps to clarify whether the observed B anomalies originate from New Physics or are due to statistical fluctuations and/or underestimated SM effects.

1.6 Motivation for the search for $B^0 \rightarrow K_S^0\tau^+\tau^-$

Within the family of exclusive $b \rightarrow s\tau^+\tau^-$ transitions, the decay $B^0 \rightarrow K^0\tau^+\tau^-$ plays a particularly interesting role. $B \rightarrow K\tau^+\tau^-$ and $B \rightarrow K^*\tau^+\tau^-$ form a set of modes that probe the same underlying quark-level process but with different hadronic final states and isospin combinations[5]. Measurements of these channels are complementary: they test the same short-distance physics while being subject to different experimental systematics, hadronic form factor uncertainties and possible couplings to new mediators.

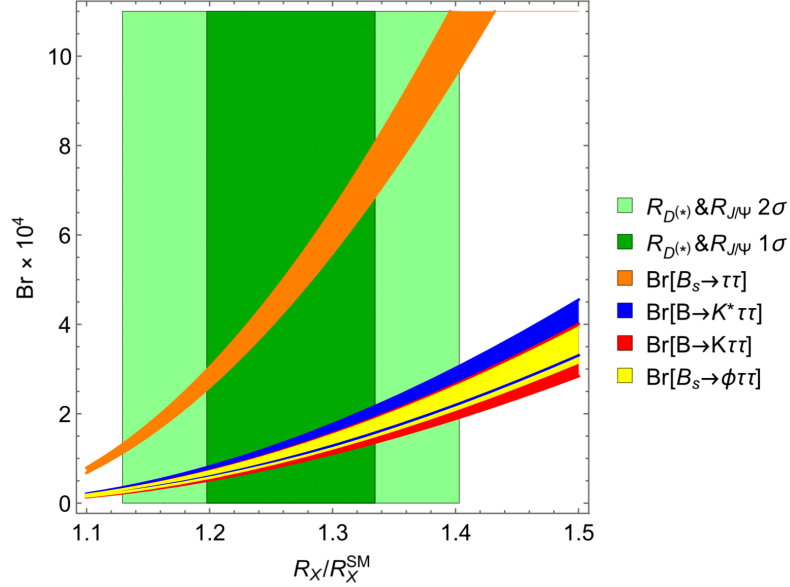


Figure 1.6: Predicted branching ratios for $b \rightarrow s \tau^+ \tau^-$ processes as a function of the anomaly, with respect to the SM, of the R_X ($X \in \{D^{(*)}, J/\psi\}$) observables [5]. The widths of the green vertical bands are the current experimental uncertainties at one and two standard deviations, respectively. The widths of the parabolic bands represent only the theoretical uncertainties.

From an experimental point of view, this work focuses on the $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ mode, where the K_S^0 is reconstructed in the $\pi^+ \pi^-$ decay mode [3]. The displaced K_S^0 vertex provides a clean signature that helps suppress background from continuum $e^+ e^- \rightarrow q \bar{q}$ processes and $B \bar{B}$ decays. However, the final state also contains two τ leptons, each decaying further into lighter leptons or hadrons plus at least one neutrino. As a consequence, the signal topology features multiple neutrinos and substantial missing energy, making full kinematic reconstruction impossible and rendering the analysis particularly challenging.

These difficulties can be partly overcome at B factories such as Belle and Belle II, where $B \bar{B}$ pairs are produced in $\Upsilon(4S)$ decays at a well-defined center-of-mass energy [3]. By fully reconstructing one of the B mesons (the *tag* side) in hadronic modes using multivariate algorithms, such as the Full Event Interpretation, one can infer the four-momentum of the companion B meson (the *signal* side) from energy-momentum conservation. This technique greatly suppresses backgrounds and provides powerful constraints on the signal kinematics, at the price of a low overall efficiency due to the difficulty of fully reconstructing hadronic B decays.

The decay $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ thus lies at the intersection of several key themes in theoretical and experimental flavor physics:

- It is a loop-suppressed FCNC process with a tiny SM branching fraction, making it an excellent probe of virtual BSM effects [5].
- It involves τ leptons in the final state, directly testing scenarios in which New Physics couples preferentially to third-generation leptons and providing an important counterpart to the measurements of $R_{D^{(*)}}$ and $R_{K^{(*)}}$ [4, 6].
- It is experimentally accessible at B factories using hadronic B tagging and the

clean $K_s^0 \rightarrow \pi^+\pi^-$ signature, especially when combining the large data sets collected by Belle and Belle II [3].

To date, no experiment has reported a measurement or limit specifically for $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ decay rate. The present work therefore represents the first dedicated search for $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ using the combined Belle and Belle II data sets. By exploiting hadronic B tagging, carefully optimised event selections, and a detailed treatment of backgrounds and systematic uncertainties, this search aims to significantly improve the experimental sensitivity to $b \rightarrow s \tau^+ \tau^-$ transitions.

Any observation of $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ at a branching fraction much larger than the SM expectation would be an unambiguous sign of New Physics and would strongly constrain the space of viable models addressing the B anomalies [5, 7]. Conversely, even in the absence of a signal, improved upper limits will provide valuable input to global fits and sharpen our understanding of the flavor structure of possible BSM interactions.

2. The Belle and the Belle II experiments

The search presented in this thesis uses data collected by the Belle and Belle II experiments at the KEKB and SuperKEKB asymmetric-energy e^+e^- colliders. Both experiments operate at or near the $\Upsilon(4S)$ resonance, where the initial state is precisely known and the final state is dominated by pairs of B mesons.

This chapter describes the experimental conditions relevant for the work. First, the KEKB and SuperKEKB colliders are introduced, with emphasis on the $\Upsilon(4S)$ environment, asymmetric beam energies, and luminosity performance. The Belle and Belle II detectors are then described by reconstruction function: tracking and vertexing, particle identification, calorimetry, K_L^0 and muon detection, and trigger and data acquisition. The chapter closes with a short discussion on accelerator and detector performance, and differences between the experiments.

2.1 The KEKB and SuperKEKB accelerators

The Belle and Belle II experiments are located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan, at the interaction point of two successive e^+e^- colliders: KEKB and its upgraded successor SuperKEKB. Both machines are designed to operate primarily at the center-of-mass (CM) energy corresponding to the $\Upsilon(4S)$ resonance, $E_{\text{CM}} \simeq 10.58$ GeV, which lies just above the threshold for $B\bar{B}$ production as shown in Fig. 2.1. In this energy regime the hadronic final state is dominated by B -meson pairs with relatively low additional multiplicities, providing a clean environment for precision flavor-physics measurements.

In this section we first introduce the general requirements for a high-luminosity B factory and the role of the $\Upsilon(4S)$ resonance. We then describe the KEKB and SuperKEKB accelerators in more detail, including the asymmetric beam-energy configuration, the nano-beam collision scheme, and the resulting luminosity performance and beam-induced backgrounds.

2.1.1 Physics motivation and the $\Upsilon(4S)$ environment

In an e^+e^- collider the center-of-mass energy $\sqrt{s}$ is determined by the beam energies via

$$s = (p_{e^-} + p_{e^+})^2 \approx 4E_{e^-}E_{e^+}, \quad (2.1)$$

where the approximation neglects the electron mass. The $\Upsilon(4S)$ resonance lies just above the threshold for $B\bar{B}$ production and decays almost exclusively into $B^0\bar{B}^0$ and B^+B^- pairs. Operating the collider at $\sqrt{s} \simeq m_{\Upsilon(4S)}$ therefore yields a very clean sample of B -meson events with well-defined initial-state kinematics and a modest multiplicity of additional hadrons.

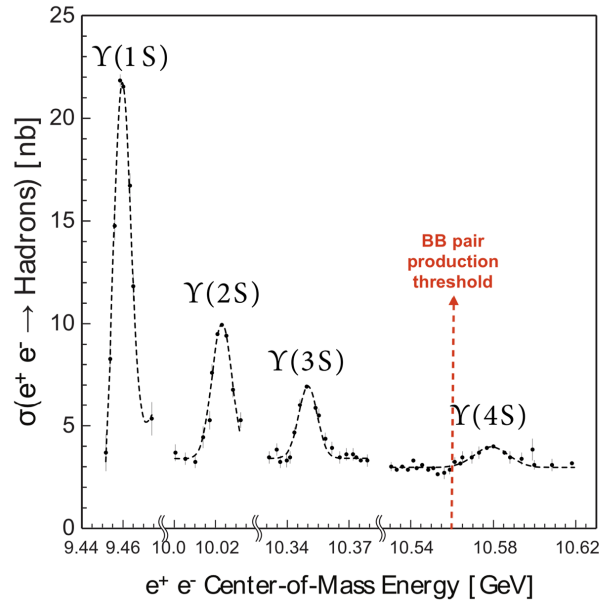


Figure 2.1: Hadronic cross-section as a function of the center-of-mass energy in the region of the first four S -wave Υ resonances.

If the beams had equal energies, the $\Upsilon(4S)$ would be produced nearly at rest in the laboratory frame and the B mesons would have momenta of only $\mathcal{O}(10 \text{ MeV}/c)$. With a typical B -meson lifetime of about 1.5 ps, the resulting decay length would be of order $50 \mu\text{m}$, comparable to or smaller than the intrinsic spatial resolution of modern vertex detectors. In such a configuration the decay vertices of the two B mesons would be difficult to separate, limiting the sensitivity to time-dependent CP -violation observables.

To overcome this limitation, the B factories employ *asymmetric* beam energies so that the $\Upsilon(4S)$ system is boosted along the beam (longitudinal) direction in the laboratory frame. The boost factor $\beta\gamma$ can be expressed as

$$\beta\gamma \simeq \frac{E_{e^-} - E_{e^+}}{m_{\Upsilon(4S)}}. \quad (2.2)$$

At KEKB, with beam energies of 8 GeV (electrons) and 3.5 GeV (positrons), $\beta\gamma \approx 0.43$, corresponding to an average B -meson decay length of $\mathcal{O}(200 \mu\text{m})$ in the laboratory frame. At SuperKEKB the beam energies are 7 GeV and 4 GeV, giving a slightly smaller boost, $\beta\gamma \approx 0.28$, and a decay length of about $130 \mu\text{m}$. In both cases the boosted topology permits precise measurements of decay-time differences between the two B mesons using high-resolution silicon vertex detectors. The smaller boost in SuperKEKB implies less missing momentum, which is beneficial for searches for decays with missing energy.

2.1.2 The KEKB collider

KEKB was a high-luminosity asymmetric-energy e^+e^- collider optimized for B -physics studies at the $\Upsilon(4S)$ resonance [23, 24]. It consisted of two storage rings,

- a high-energy ring (HER) for 8 GeV electrons, and
- a low-energy ring (LER) for 3.5 GeV positrons,

sharing a single interaction region where the Belle detector was installed. Electrons and positrons were accelerated by a linear accelerator (LINAC) and injected into the rings in a continuous top-up mode to maintain nearly constant beam currents. The schematic of the accelerator complex is shown in Fig. 2.2.

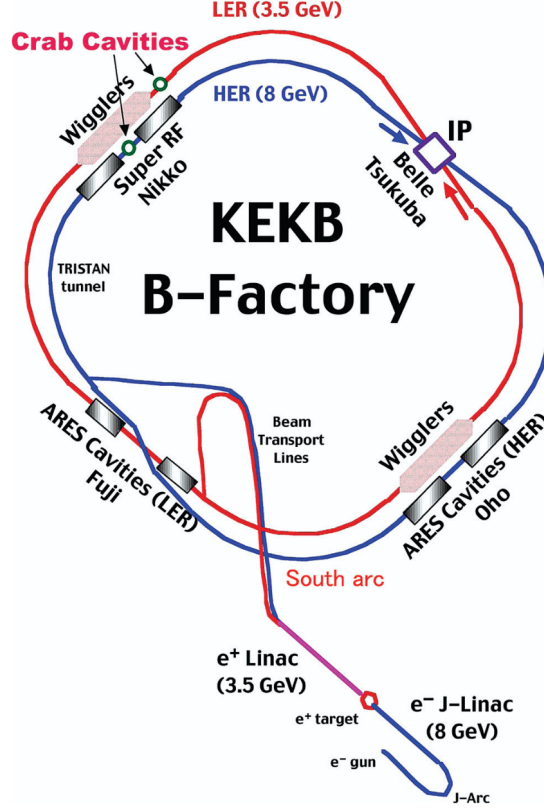


Figure 2.2: Schematic overview of the KEKB accelerator complex, showing the LINAC, the high- and low-energy storage rings, and the interaction point surrounded by the Belle detector.

The machine employed a double-ring design with separate vacuum chambers and optics for electrons and positrons. The two beams were brought into collision at a small horizontal crossing angle at the interaction point (IP). The main design parameters and the achieved operating values during Belle data taking are summarized schematically in Table 2.1. The peak instantaneous luminosity reached $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, which exceeded the original design goal and established a world record that was held for nearly a decade.

Several innovations contributed to the high performance of KEKB:

- **Continuous injection**, allowing the beams to be replenished without interrupting physics data taking, thereby maximizing the delivered integrated luminosity.
- **Crab cavities**, which were introduced in the final years of operation to compensate the crossing angle and restore head-on collisions in the bunch rest frame, improving the effective overlap of the bunches and thus the luminosity.
- **Sophisticated optics and feedback systems**, which stabilized the beam orbits and mitigated various collective instabilities.

Table 2.1: Illustrative comparison of design and achieved KEKB parameters relevant for luminosity performance. The values are indicative; precise numbers can be found in Refs. [23, 24].

Quantity	Design value	Achieved value
Beam energies (E_{e^+}, E_{e^-})	(3.5, 8.0) GeV	(3.5, 8.0) GeV
Beam currents (LER, HER)	$\sim (2.6, 1.1)$ A	similar
Vertical β at IP, β_y^*	10 mm	5–6 mm
Peak luminosity	$1.0 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$	$2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$

During its operation from 1999 to 2010, KEKB delivered a total integrated luminosity of about 1 ab^{-1} at and around the $\Upsilon(4S)$, corresponding to 711 fb^{-1} recorded by Belle at the $\Upsilon(4S)$ resonance and additional data collected at other bottomonium states and in the continuum. This dataset is the foundation of the legacy B -physics measurements from Belle.

2.1.3 The SuperKEKB collider

SuperKEKB is a major upgrade of KEKB designed to reach a target luminosity of

$$\mathcal{L}_{\text{design}} = 6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}, \quad (2.3)$$

approximately a factor forty higher than the peak luminosity of KEKB. The accelerator retains the general double-ring layout but incorporates significant upgrades to the injector, the damping ring, the main-ring optics, and the interaction region [25, 26]. The schematic of the accelerator complex is shown in Fig. 2.3

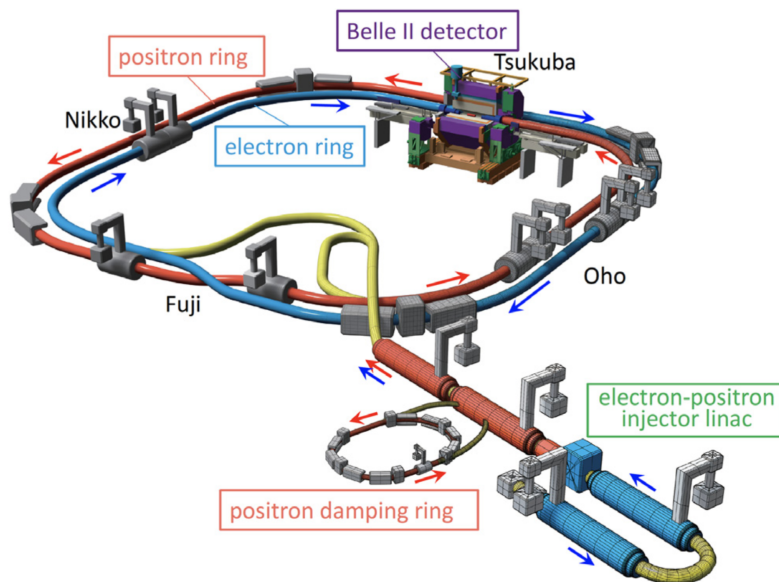


Figure 2.3: Schematic view of the SuperKEKB accelerator complex, including the LINAC, the positron damping ring, and the two main storage rings (LER and HER) that intersect at the interaction point.

The key design changes relative to KEKB include:

- **Revised beam energies:** the beam energies are set to $E_{e^-} = 7$ GeV (HER) and $E_{e^+} = 4$ GeV (LER), yielding a center-of-mass energy at the $\Upsilon(4S)$ and a boost $\beta\gamma \approx 0.28$. The reduced boost compared to KEKB is a compromise between physics needs and the requirement to minimise emittance and maximise luminosity.
- **Upgraded LINAC and positron source:** the injector is upgraded to deliver higher-current, lower-emittance beams. A dedicated damping ring at 1.1 GeV reduces the positron emittance before injection into the LER.
- **New final-focus superconducting quadrupoles:** the interaction region is redesigned with compact superconducting quadrupoles close to the IP to achieve extremely small β_y^* values in the nano-beam scheme.
- **Improved vacuum system:** new beam pipes and antechamber structures, together with extensive vacuum pumping, mitigate synchrotron-radiation-induced gas desorption and associated beam-gas backgrounds.

Table 2.2 summarizes representative design and currently achieved values for key machine parameters.

Table 2.2: Indicative design and achieved values for selected SuperKEKB parameters in the LER/HER rings, relevant for luminosity. Numerical values are illustrative and based on Refs. [25] and current machine operation status.

Quantity	Design value	Achieved (Run 1–2)
Beam energies (E_{e^+}, E_{e^-})	(4.0, 7.0) GeV	(4.0, 7.0) GeV
Beam currents ($I_{\text{LER}}, I_{\text{HER}}$)	(3.6, 2.6) A	$\mathcal{O}(1.3, 1.1)$ A
Vertical β at IP, β_y^*	(0.27, 0.41) mm	(0.9, 0.9) mm
Peak luminosity	$6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$	$5.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$

2.1.4 The nano-beam collision scheme

The significant luminosity increase at SuperKEKB is achieved primarily through the *nano-beam* scheme. In an e^+e^- collider, the instantaneous luminosity can be expressed approximately as

$$\mathcal{L} = \frac{\gamma_{\pm}}{2er_e} \left(1 + \frac{\sigma_y^*}{\sigma_x^*} \right) \frac{I_{\pm} \xi_{y,\pm}}{\beta_y^*} \frac{R_L}{R_{\xi_y}}, \quad (2.4)$$

where e is the electron charge, r_e is the classical electron radius, $\gamma_{\pm}$ are the relativistic Lorentz factors for electrons and positrons, $\sigma_{x,y}^*$ are the horizontal and vertical beam sizes at the IP, $I_{\pm}$ are the beam currents, $\xi_{y,\pm}$ are the vertical beam-beam tune-shift parameters that describe the focusing force exerted on a bunch by the electromagnetic field of the opposite bunch, and β_y^* is the vertical β function at the IP. The factors R_L and R_{ξ_y} account for geometric reductions due to the finite crossing angle of the beams.

From Eq. (2.4) it is clear that the luminosity can be increased by:

- increasing the beam currents $I_{\pm}$,
- increasing the beam–beam parameter $\xi_{y,\pm}$, and
- decreasing the vertical β function β_y^* (and therefore the vertical beam size σ_y^*).

However, simply increasing the beam current and reducing the beam size leads to instabilities and enhanced non-linear beam–beam interactions. Moreover, the so-called *hourglass effect* limits the benefit of very small β_y^* if the bunch length σ_z is comparable to or larger than β_y^* . In that case, only a fraction of the bunch is focused to the minimum beam size at the IP, and the effective luminosity is reduced.

The nano-beam scheme (Fig. 2.4) addresses these issues by arranging for the two beams to collide with a larger horizontal crossing angle φ (about 41.5 mrad at SuperKEKB, compared to 11 mrad of KEKB) and strongly squeezing the vertical beam size while also reducing the horizontal size σ_x . The effective overlap region of the two bunches has a length of order

$$d \sim \frac{\sigma_x}{\sin \varphi}, \quad (2.5)$$

which can be made much shorter than the physical bunch length σ_z . In this way the condition $d \lesssim \beta_y^*$ can be satisfied even with relatively long bunches, mitigating the hourglass effect.

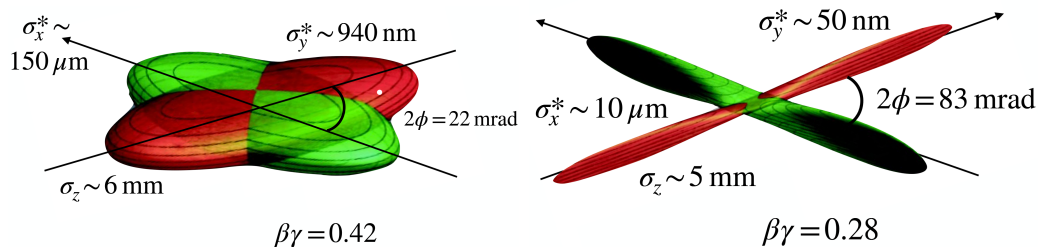


Figure 2.4: Illustration of the nano-beam collision scheme at SuperKEKB (right), compared to a conventional collision scheme with a small crossing angle (left). In the nano-beam configuration the beams are strongly focused and collide with a large crossing angle, reducing the effective overlap length and mitigating the hourglass effect.

In practice the nano-beam scheme requires:

- low-emittance beams from the injector and damping ring,
- strong final-focus optics with compact superconducting quadrupoles near the IP,
- precise control of the beam orbits and optics to maintain the small beam sizes.

These requirements pose stringent constraints on the interaction-region design and increase the sensitivity to machine imperfections and vibrations.

2.1.5 Luminosity performance and beam-induced backgrounds

The performance of a collider can be characterized by the integrated luminosity

$$\mathcal{L}_{\text{int}}^{\text{delivered}}(T) = \int_0^T \mathcal{L}(t) dt, \quad (2.6)$$

which directly determines the number of events N of a process with cross section σ via $N = \mathcal{L}_{\text{int}}^{\text{delivered}} \sigma$. KEKB delivered an integrated luminosity of about 1 ab^{-1} , while the design goal for SuperKEKB is $30 - 50 \text{ ab}^{-1}$ over its operational lifetime.

The push to much higher instantaneous luminosities at SuperKEKB leads to significantly increased beam-induced backgrounds compared to KEKB. Several mechanisms contribute:

- **Touschek scattering**, in which intra-bunch Coulomb scattering transfers particles from the core of the bunch into the high-amplitude tails, where they can be lost on the aperture.
- **Beam–gas interactions**, arising from collisions of beam particles with residual gas molecules in the vacuum chamber, producing hadronic showers and scattered leptons.
- **Synchrotron-radiation and radiative Bhabha processes**, which generate energetic photons that may interact with material around the beam pipe.
- **Injection-related backgrounds**, as fresh particles are injected into the rings at high rates to maintain the beam currents. Before these particles are fully damped onto the nominal orbit, they can undergo large betatron oscillations and be lost on apertures or collimators, producing short background bursts in the detector.

These backgrounds manifest themselves in the Belle and Belle II detectors as:

- increased hit occupancies, particularly in the innermost vertex detectors,
- additional clusters and noise in the electromagnetic calorimeter,
- spurious hits and noise in the muon system and tracking detectors,
- higher radiation doses on sensitive electronics and sensors.

The detector design, shielding, and data-acquisition systems must therefore accommodate these conditions. In particular, Belle II employs faster and more granular detectors, enhanced radiation hardness, and sophisticated online and offline background-rejection strategies.

2.1.6 Impact on Belle and Belle II physics

The accelerator parameters discussed above directly influence the physics reach of the Belle and Belle II experiments. The asymmetric beam energies and the resulting boost of the $\Upsilon(4S)$ system in the laboratory frame enable precise time-dependent measurements of CP violation in the B -meson system. The high integrated luminosities achieved and targeted by KEKB and SuperKEKB translate into very large samples of B mesons, τ leptons and charmed hadrons.

At the same time, the nano-beam scheme and high beam currents at SuperKEKB introduce substantial beam-related backgrounds and operational challenges. The Belle II detector is therefore designed to provide improved vertex resolution, tracking, and particle identification performance in a significantly harsher environment than that experienced by Belle. These aspects are discussed in detail in the following sections, where we describe the Belle and Belle II detectors and their subsystems.

2.2 The Belle and Belle II detectors

The Belle and Belle II detectors are large-solid-angle magnetic spectrometers located at the interaction points of the KEKB and SuperKEKB asymmetric-energy e^+e^- colliders, respectively. Both experiments operate at or near the $\Upsilon(4S)$ resonance, where the production of coherent $B\bar{B}$ pairs provides a clean environment for precision flavor-physics measurements. The Belle detector was designed to reconstruct and identify charged and neutral particles produced in e^+e^- collisions, with particular emphasis on time-dependent CP violation and rare B -meson decays [27, 28]. Belle II is its successor and largely retains the same global detector concept, while incorporating substantial upgrades needed for the much higher luminosity and harsher beam-background conditions of SuperKEKB [28, 29].

The two detectors share a similar cylindrical layout. From the interaction point outwards, particles traverse a low-mass beam pipe, vertex and tracking detectors, charged-particle identification (PID) systems, an electromagnetic calorimeter, a superconducting solenoid, and an instrumented iron flux-return used for muon and K_L^0 identification. A schematic cut-away view of Belle and Belle II detectors is shown in Fig. 2.5. This section follows the detector functions relevant for the analysis: tracking, particle identification, reconstruction of neutral particles, K_L^0 and muon identification, and online event selection.

2.2.1 Coordinate system and global detector layout

Belle and Belle II use a right-handed coordinate system with the origin at the nominal interaction point. The z axis is defined approximately along the electron-beam direction, the x axis points horizontally outward from the centre of the accelerator rings, and the y axis points vertically upward. The polar angle θ is measured from the positive z axis, and the azimuthal angle ϕ from the positive x axis in the xy plane. In cylindrical coordinates, the radial coordinate is defined as $r = \sqrt{x^2 + y^2}$.

The term “barrel region” refers to polar angles roughly transverse to the beam direction, while the forward and backward endcaps cover the regions closer to the beam directions. In both Belle and Belle II, the barrel region corresponds approximately to $32^\circ \lesssim \theta \lesssim 130^\circ$, with the endcaps extending the angular coverage to about 12° in the forward direction and 155° in the backward direction.

The Belle detector was approximately 8.4 m long and 7.9 m in diameter, with a total weight of about 1 400 t [27]. Belle II is about 7 m long and 7 m in diameter and is embedded, like Belle, in a 1.5 T solenoidal magnetic field. The approximate angular coverage of the principal Belle II subdetectors is summarized in Table 2.3.

2.2.2 Tracking and vertex reconstruction

Charged-particle tracking is performed using the silicon detectors close to the interaction point and a large central drift chamber at larger radii. These systems reconstruct the trajectories of charged particles (tracks) in the 1.5 T magnetic field,

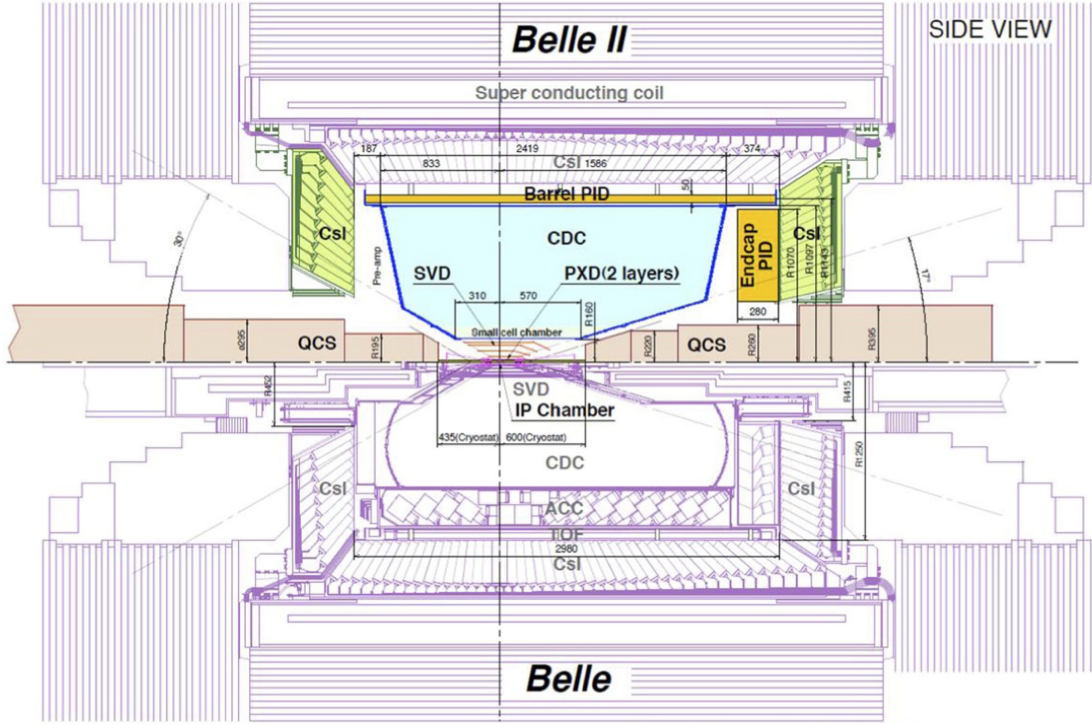


Figure 2.5: A schematic of the Belle (lower half) and Belle II (upper half) detectors, showing the arrangement of the main subdetectors around the interaction point. [28]

Table 2.3: Indicative overview of the Belle II subdetectors and their approximate polar-angle coverage. Detailed specifications are given in Ref. [28].

Subdetector	Function	Typical polar-angle coverage
PXD + SVD (VXD)	Precise vertexing, tracking	$17^\circ\text{--}150^\circ$
CDC	Tracking, vertexing, dE/dx	$17^\circ\text{--}150^\circ$
TOP (barrel)	Charged-particle identification	$32^\circ\text{--}129^\circ$
ARICH (forward)	Charged-particle identification	$12^\circ\text{--}32^\circ$
ECL	γ reconstruction, $e^\pm$ identification	$12^\circ\text{--}155^\circ$
KLM	$\mu^\pm$ identification and K_L^0 reconstruction	$20^\circ\text{--}155^\circ$

provide momentum measurements, and determine decay vertices. They also measure track impact parameters, which describe the distance of closest approach of a track to the interaction point, separated into transverse (dr) and longitudinal (dz) components with respect to the beam axis. This information is essential for reconstructing short-lived particles such as B mesons, charm hadrons, τ leptons, and displaced $K_S^0 \rightarrow \pi^+\pi^-$ decays.

Beam pipe and interaction region

At Belle, the beam pipe was made of beryllium to minimize the amount of material close to the interaction point, thereby reducing multiple scattering and photon conversions. Its inner radius was approximately 20 mm, with a wall thickness of order 1 mm. This provided a compromise between mechanical stability, vacuum requirements, and a low material budget.

At Belle II, the smaller beam size provided by the nano-beam collision scheme

allowed the beam-pipe radius to be reduced to about 10 mm at the interaction region. This improves vertex resolution by placing the innermost detector layers closer to the interaction point, but also increases the need for radiation-hard sensors, fast readout, and careful background control.

Silicon vertex detectors

The Belle silicon vertex detector (SVD) was the innermost tracking subsystem and played a central role in reconstructing decay vertices and track impact parameters. It consisted of double-sided silicon strip detectors (DSSDs) arranged in concentric cylindrical layers around the beam pipe. In the initial Belle configuration, the SVD comprised three layers; in 2003 it was upgraded to a four-layer system, extending the radial coverage and improving the vertex resolution [27]. The SVD covered the polar-angle range $17^\circ < \theta < 150^\circ$, corresponding to nearly the full detector acceptance. A schematic view of the Belle SVD is shown in Fig. 2.6.

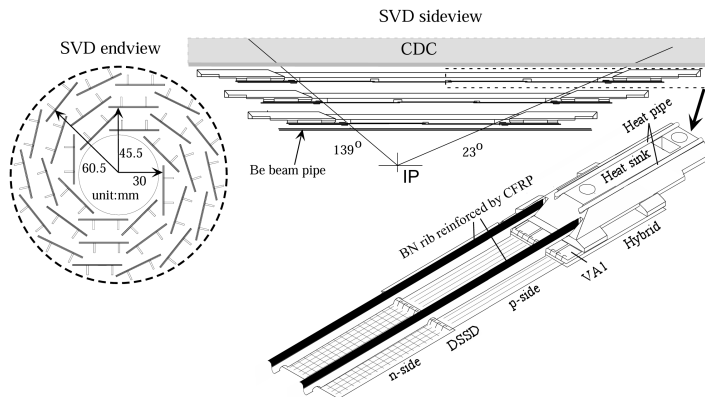


Figure 2.6: Schematic view of the Belle silicon vertex detector, showing the arrangement of the DSSD ladders and the coverage in polar angle. [27]

Each Belle DSSD module featured orthogonal strip implants on the two sides. Strips on one side were oriented approximately along the beam direction and provided precise measurements of the azimuthal coordinate ($r-\phi$), while strips on the opposite side measured the longitudinal coordinate (z). Typical strip pitches were of order $25\text{--}50\text{ }\mu\text{m}$ on the $r-\phi$ side and $50\text{--}80\text{ }\mu\text{m}$ on the z side. To limit the number of readout channels, every second z strip was read out, leading to an effective pitch of about $80\text{ }\mu\text{m}$.

The impact-parameter resolutions in the transverse (x, y) and longitudinal (z) projections can be parameterized as

$$\sigma_{xy} \simeq a_{xy} \oplus \frac{b_{xy}}{p\beta \sin^{3/2} \theta}, \quad (2.7)$$

$$\sigma_z \simeq a_z \oplus \frac{b_z}{p\beta \sin^{5/2} \theta}, \quad (2.8)$$

where p is the particle momentum in GeV/c , β is its velocity divided by the speed of light, and θ is the polar angle. The constants a and b encode the intrinsic resolution and the contribution from multiple scattering, respectively. Typical values at Belle were of order $a_{xy} \sim 20\text{ }\mu\text{m}$, $a_z \sim 30\text{ }\mu\text{m}$, and $b_{xy}, b_z \sim 30\text{--}35\text{ }\mu\text{m GeV}/c$ [27]. This performance was essential for time-dependent CP violation measurements in B^0 decays.

In Belle II, the vertex detector (VXD) is a major upgrade of the Belle SVD. It consists of a two-layer DEPFET pixel detector (PXD) surrounded by a four-layer double-sided silicon-strip detector (SVD) [28, 30]. The VXD is closer to the interaction point and extends to a larger outer radius, improving both the intrinsic vertex resolution and the lever arm for impact-parameter measurements. The PXD consists of two cylindrical layers at radii of approximately 14 mm and 22 mm, using thin DEPFET sensors in which the sensor and first amplification stage are integrated in the same device. A schematic of the Belle II PXD is shown in Fig. 2.7.

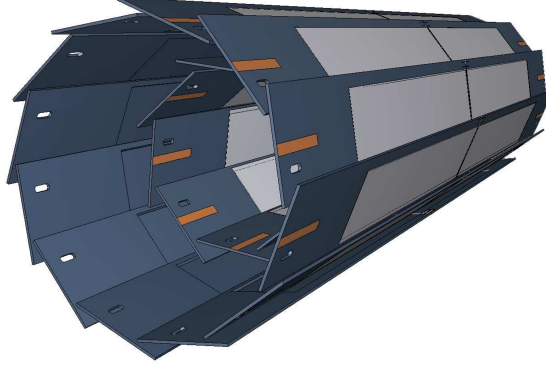


Figure 2.7: Schematic view of the two-layer DEPFET pixel detector (PXD) of Belle II, illustrating the ladder arrangement and support structures. [28]

The hit data from PXD are transferred to the readout system, where they are associated to regions of interest defined by the high-level trigger in order to reduce the data volume. During the first Belle II running period (2019–2022), a partial pixel detector configuration was installed, with a full inner layer and only a fraction of the second layer instrumented. During the long shutdown LS1 (2022–2023), it was replaced by a full-coverage detector with two fully instrumented layers, improved cooling, and updated readout electronics [28].

The Belle II SVD surrounds the PXD and consists of four layers of DSSDs at radii from about 38 mm to 135 mm. It provides additional space points for track reconstruction, extends the lever arm for impact-parameter measurements, and improves pattern recognition together with the rest of the tracking system. A picture of the Belle II VXD after installation is shown in Fig. 2.8.

Each Belle II SVD layer consists of ladders built from DSSDs with orthogonal strip implants on the two sides. The strip pitches are of order 50–75 μm in $r-\phi$ and 75–100 μm in z , depending on the layer. The impact-parameter resolutions achieved by the full VXD+CDC tracking system in Belle II can be parameterized as

$$\sigma_{xy} \simeq 10 \mu\text{m} \oplus \frac{25 \mu\text{m}}{p\beta \sin^{3/2} \theta}, \quad (2.9)$$

$$\sigma_z \simeq 15 \mu\text{m} \oplus \frac{27 \mu\text{m}}{p\beta \sin^{5/2} \theta}, \quad (2.10)$$

where p is the track momentum in GeV/ c [28, 30]. This represents a substantial improvement over Belle and is particularly important for reconstructing displaced vertices from K_s^0 mesons in the analysis presented in this thesis.

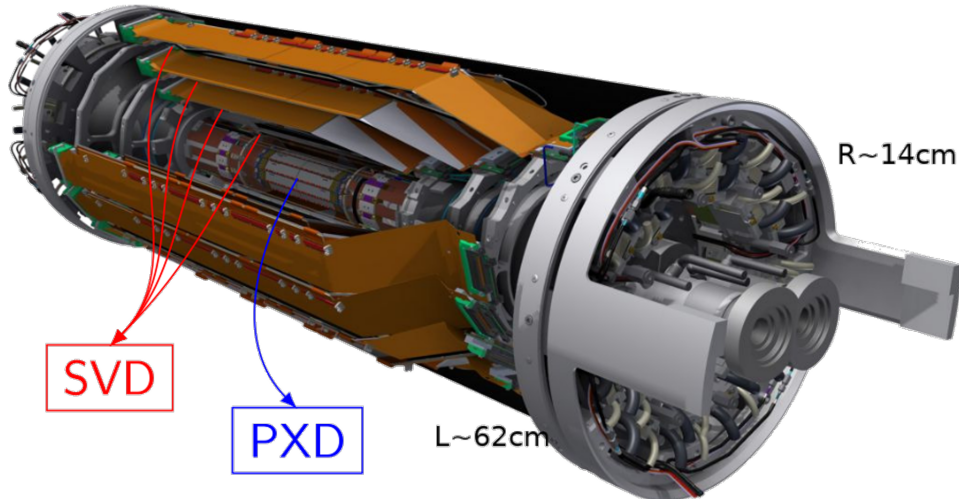


Figure 2.8: Picture of the Belle II vertex detector (VXD) after installation, showing the PXD and SVD layers. [28]

Central drift chamber

The central drift chamber (CDC) is the main tracking detector at larger radii. It provides charged-particle tracking over a large solid angle, contributes to momentum determination, and supplies specific ionization (dE/dx) information for particle identification.

The Belle CDC was located between the SVD and the particle-identification detectors. It consisted of 50 cylindrical layers of sense wires arranged in a jet-type cell structure [27]. The innermost layers used small cells optimized for high hit density near the beam pipe, while the outer layers used larger cells. The sensitive gas was a helium-based mixture chosen to minimize multiple scattering and improve momentum resolution. The transverse cross section of the Belle CDC is shown in Fig. 2.9.

The Belle CDC reconstructed tracks as helices in the uniform magnetic field. For tracks reconstructed with the CDC and SVD, the transverse-momentum resolution could be parameterized as

$$\frac{\sigma_{p_T}}{p_T} \simeq a_{p_T} p_T \oplus \frac{b_{p_T}}{\beta}, \quad (2.11)$$

where p_T is in units of GeV/c . The coefficients were of order $a_{p_T} \sim 0.003$ and $b_{p_T} \sim 0.005$ [27], yielding a relative resolution of about 0.5% at $p_T = 1 \text{ GeV}/c$.

The Belle II CDC is an enlarged and finer-grained successor of the Belle CDC [28]. It extends radially from about $r = 160 \text{ mm}$ to $r = 1130 \text{ mm}$ and covers the polar-angle range $17^\circ < \theta < 150^\circ$. It contains 56 layers of sense wires grouped into superlayers with alternating axial and stereo orientations. Axial layers provide measurements in the r - ϕ plane, while stereo layers are tilted with respect to the beam axis to provide z information. The total number of sense wires is 14 336, surrounded by field wires that shape the electric field in each drift cell. The cell sizes are smaller and the outer radius is larger than in Belle, improving the momentum resolution and helping to cope with higher occupancies.

The Belle II CDC uses a helium-isobutane gas mixture optimized to reduce multiple scattering while maintaining good drift properties.

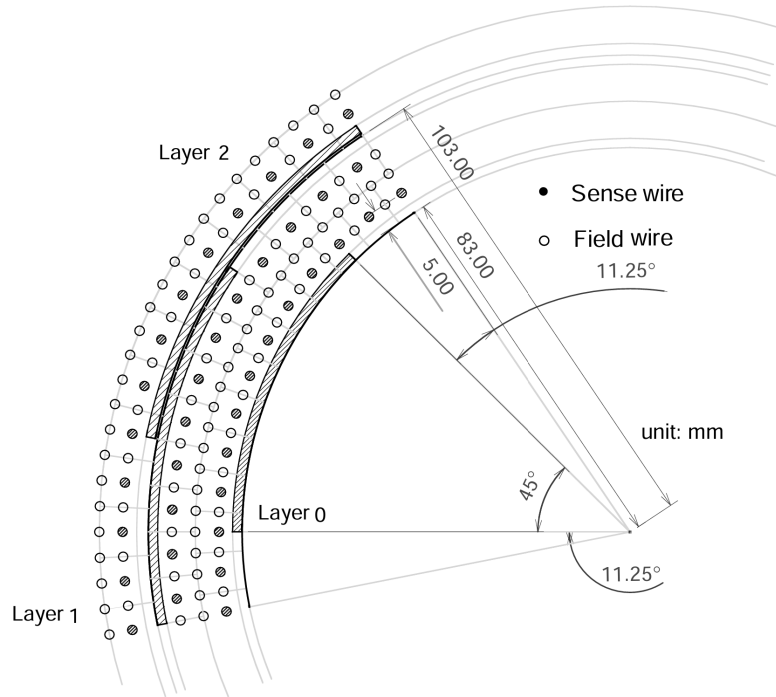


Figure 2.9: Transverse cross section of the Belle central drift chamber, illustrating the arrangement of sense and field wires and the cell structure. [27]

Superconducting solenoid and magnetic field

Both Belle and Belle II use a superconducting solenoid that provides a magnetic field of 1.5 T along the $+z$ direction [27, 28]. The field volume includes the beam pipe, vertex detector, CDC, particle-identification detectors, and ECL, ensuring that charged particles traverse the tracking and PID systems in a well-defined magnetic environment. The iron yoke outside the solenoid acts as a magnetic flux return and also serves as absorber material for the KLM. Precise knowledge of the magnetic field is essential for track momentum reconstruction, and the field was mapped and monitored during detector operation.

2.2.3 Charged-particle identification

Efficient separation of charged hadrons and identification of electrons and muons are crucial for many B -physics analyses. Both Belle and Belle II combine information from several subdetectors into likelihoods or multivariate selectors for different particle hypotheses. The CDC measures the specific ionization energy loss dE/dx by averaging charge information over multiple sense wires along a track. This provides useful particle-identification information, especially for low-momentum particles and for tracks outside the most sensitive regions of the dedicated particle-identification PID detectors. The electromagnetic calorimeter contributes to electron identification, and the KLM provides muon identification. The main difference between Belle and Belle II lies in the dedicated hadron-identification systems: Belle used threshold Cherenkov and time-of-flight detectors, whereas Belle II uses imaging Cherenkov techniques with precise timing and spatial information.

Particle identification at Belle: ACC and TOF

At Belle, charged-hadron identification was based primarily on aerogel Cherenkov counters (ACC), time-of-flight (TOF) scintillation counters, and CDC dE/dx measurements. The ACC and TOF detectors were located outside the CDC and provided complementary information in different momentum regions.

The ACC were threshold Cherenkov detectors using silica aerogel as the radiator material [27]. The refractive index of the aerogel tiles was chosen so that pions and kaons had different Cherenkov thresholds in the momentum region relevant for B decays. Photons produced in the aerogel were reflected by wavelength-shifting plates and detected by photomultiplier tubes. The response was parameterized in terms of the number of photoelectrons, and likelihoods for the π , K , and p hypotheses were formed by comparing the measured response with the expectation for a given track momentum.

The ACC provided K/π separation up to momenta of about $3.5 \text{ GeV}/c$ in the forward region. The TOF system consisted of scintillator counters arranged in a barrel just outside the ACC. Each counter was read out at both ends by photomultiplier tubes, providing timing information with a typical resolution of order 100 ps [27]. For a track of path length L and reconstructed momentum p , the expected time of flight for a particle of mass m is

$$t = \frac{L}{c} \sqrt{1 + \left(\frac{mc}{p}\right)^2}. \quad (2.12)$$

The TOF system was particularly effective for K/π separation below about $1.5 \text{ GeV}/c$. In combination with CDC dE/dx and ACC information, it provided robust hadron identification over the momentum range relevant for Belle B -decay analyses.

Particle identification at Belle II: TOP and ARICH

At Belle II, the higher luminosity and increased background rates motivated the replacement of the Belle ACC and TOF systems with more advanced Cherenkov detectors. In the barrel region, charged-hadron identification is provided by the time-of-propagation counter (TOP), while in the forward endcap it is provided by the aerogel ring-imaging Cherenkov detector (ARICH) [28, 31].

The TOP is a Cherenkov detector based on the precise measurement of photon propagation times in quartz radiator bars [32]. It consists of 16 modules arranged azimuthally around the CDC. Each module contains a synthetic fused-silica radiator bar, about 2.7 m long, coupled to micro-channel plate photomultiplier tubes (MCP-PMTs) at one end. Cherenkov photons emitted by a charged particle undergo total internal reflection inside the bar and propagate to the photon sensors. A schematic of a TOP module is shown in Fig. 2.10.

The TOP exploits both the time of propagation and the spatial distribution of detected photons. For a given particle hypothesis and measured track parameters, the expected photon hit times and positions can be predicted. The observed pattern is compared with expectations to form likelihoods, for example

$$\ln \mathcal{L}_{\text{TOP}}^{\text{hyp}}(h) = \sum_{i=1}^{N_\gamma} \ln P(t_i, \mathbf{x}_i | \text{hyp}, h), \quad (2.13)$$

where N_γ is the number of detected photons and P is the probability density for observing a photon at time t_i and position $\mathbf{x}_i$ given the track hypothesis h and

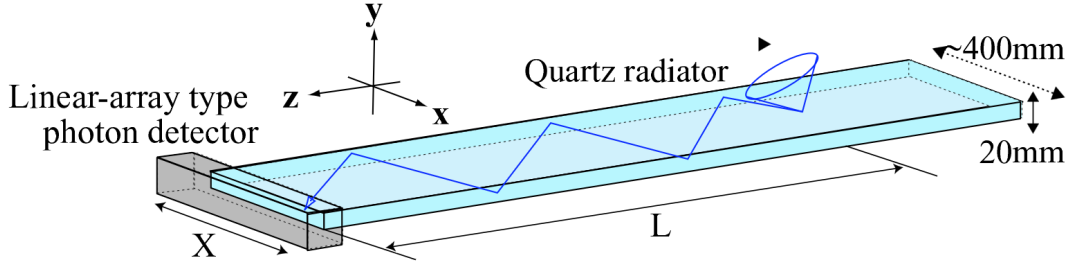


Figure 2.10: Schematic diagram of a Belle II TOP module, showing the quartz radiator bar and the MCP-PMT array at the readout end. [28]

particle type “hyp”. The TOP provides excellent K/π separation, typically exceeding $3\text{--}4\sigma$ up to momenta of about $3\text{ GeV}/c$ in the barrel region.

The ARICH detector is installed in the forward endcap region and uses silica aerogel as the radiator material. It is optimized for particle identification in the forward direction, where the TOP does not provide coverage. The ARICH uses aerogel tiles with different refractive indices in a dual-layer configuration to improve the Cherenkov-angle resolution. Photons emitted in the aerogel are detected by an array of photon sensors downstream of the radiator. The operating principle of the ARICH detector is shown in Fig. 2.11.

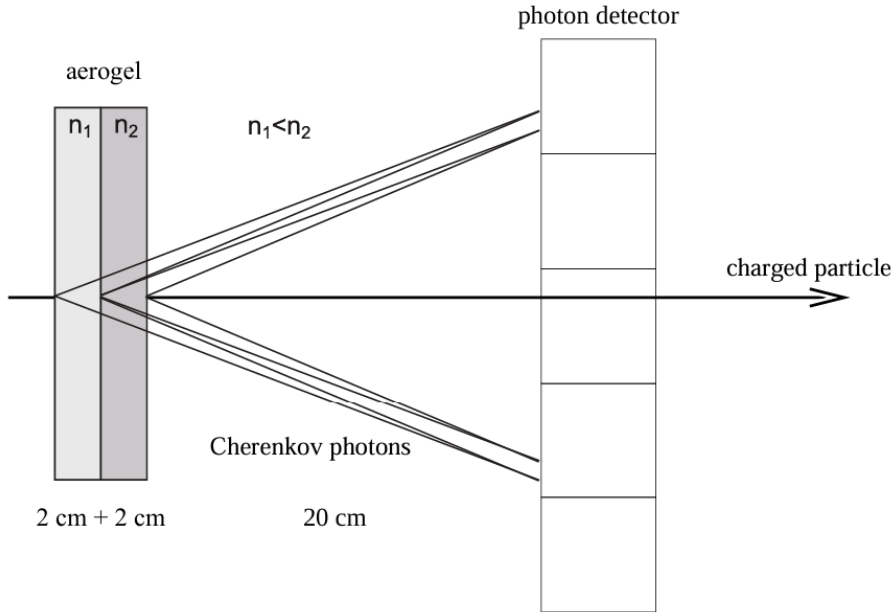


Figure 2.11: Principle of the ARICH detector operation. The charged particle crosses layers of aerogel with different refractive indexes and the emitted Cherenkov photons are detected by the HAPDs that reconstruct the Cherenkov ring.

For a given track of momentum p , the expected Cherenkov angle for a particle of mass m in a medium of refractive index n is

$$\cos \theta_C = \frac{1}{n\beta} = \frac{1}{n} \sqrt{1 + \left(\frac{mc}{p} \right)^2}, \quad (2.14)$$

where β is the particle velocity in units of c . The measured ring pattern is compared with expectations for the π , K , and p hypotheses to form likelihoods. The ARICH provides effective K/π separation in the forward region up to momenta of several GeV/ c , complementing the TOP coverage and enabling robust PID across the Belle II acceptance.

2.2.4 Electromagnetic calorimetry and neutral particles

The electromagnetic calorimeter (ECL) reconstructs photons and electrons, provides the main input for $\pi^0 \rightarrow \gamma\gamma$ reconstruction, and contributes to global event quantities such as the visible and missing energy. These capabilities are particularly important in analyses with neutrinos or other invisible particles in the final state, including the search for $B^0 \rightarrow K_s^0 \tau^+ \tau^-$.

The Belle ECL was composed of 8736 thallium-doped cesium iodide crystals, CsI(Tl), arranged in barrel and endcap modules [27]. The crystals were typically 16–17 radiation lengths deep and read out by photodiodes. The calorimeter covered the polar-angle range $12^\circ \lesssim \theta \lesssim 155^\circ$. Its layout is shown in Fig. 2.12.

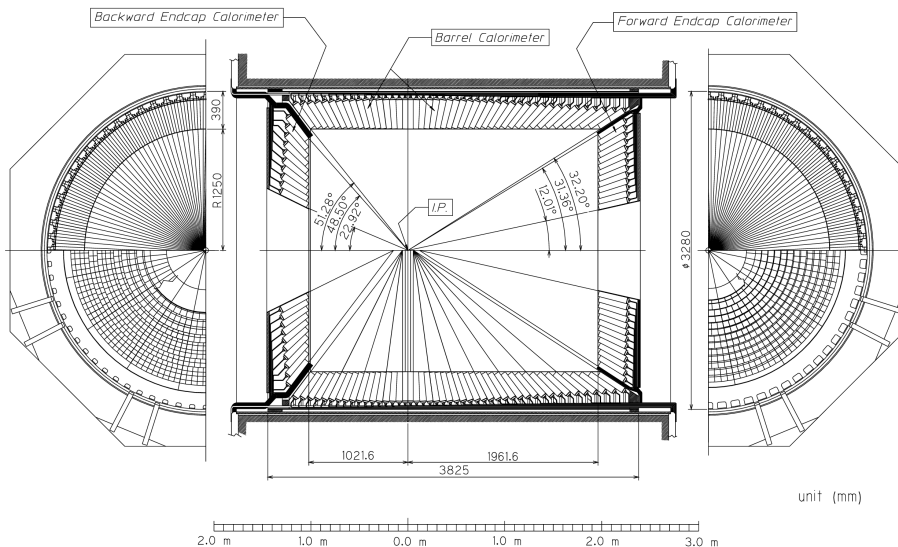


Figure 2.12: Layout of the Belle electromagnetic calorimeter, showing the barrel and endcap crystal arrays. [27]

The energy resolution for electromagnetic showers in Belle could be parameterized as

$$\frac{\sigma_E}{E} \simeq \frac{a}{\sqrt{E [\text{GeV}]}} \oplus b, \quad (2.15)$$

with $a \sim 1.5\%$ – 2% and $b \sim 1\%$. The position resolution for photon clusters was of order a few mm in the barrel. Electron identification used the cluster energy, shower shape, and E/p ratio, combined with tracking and CDC dE/dx information. The ECL also provided timing information used to suppress beam backgrounds and assist in trigger decisions.

Belle II reuses the CsI(Tl) crystal calorimeter from Belle but with substantially upgraded waveform-sampling front-end electronics [28]. This upgrade provides more precise timing information and allows operation under the higher beam-background conditions at SuperKEKB. The energy resolution remains similar to Belle, while

the improved timing, of order a few ns for electromagnetic showers, helps reject out-of-time background clusters.

Neutral-particle reconstruction in both experiments relies primarily on photon candidates measured in the ECL. Pairs of photons are combined to reconstruct neutral pions, while individual photons are also used in the reconstruction of radiative decays and excited hadron states. This information is important for reconstructing both the companion B meson in hadronic modes and decays on the signal-side such as $\tau \rightarrow \rho\nu$ with $\rho^\pm \rightarrow \pi^\pm\pi^0$. In addition, the calorimeter hermeticity helps identify missing-energy signatures and suppress backgrounds with undetected neutral particles by constraining the additional neutral energy in the rest of the event.

2.2.5 K_L^0 and muon identification

Muon identification and detection of long-lived neutral kaons are provided by the K_L^0 and muon detector (KLM), located in the outermost region of the detector. In both Belle and Belle II, the KLM is integrated into the iron flux-return yoke of the solenoid. The iron acts as a hadron absorber and magnetic flux return, while active detector layers provide hit information from penetrating particles.

In Belle, the KLM consisted of alternating layers of iron and resistive plate chambers (RPCs) with orthogonal strip readout [27]. Tracks reconstructed in the inner detectors were extrapolated into the KLM, and the patterns of hits and their penetration depths were used to identify muons and K_L^0 candidates. Muon identification efficiency exceeded 90% in the relevant momentum range, with a pion fake rate of a few percent. The KLM also provided directional information for K_L^0 mesons and contributed to the trigger.

Belle II retains the same general KLM concept, but parts of the system were upgraded for the higher-background SuperKEKB environment [28]. In the barrel region, RPCs are employed, while in the endcaps the original RPCs were replaced by plastic scintillator strips with wavelength-shifting fibres and silicon photomultipliers. The scintillator-based detectors provide better performance in the high-background endcap regions and improve timing capabilities. The KLM remains important for muon identification, K_L^0 detection, event hermeticity, and trigger decisions.

2.2.6 Trigger and data acquisition

The trigger and data-acquisition (DAQ) systems select events of interest in real time and record them for offline reconstruction. Their design reflects the different running conditions of KEKB and SuperKEKB. Belle operated with high bunch-crossing rates but significantly lower backgrounds than Belle II, while Belle II must handle much larger event rates and detector occupancies.

Belle employed a multi-level trigger architecture [27]. The hardware Level-1 trigger used information from the CDC, ECL, TOF, and KLM to select events consistent with hadronic interactions, leptonic final states, and calibration or control samples. It applied topological criteria such as track and calorimeter-cluster multiplicities, total visible energy thresholds, and identification of Bhabha events and two-photon processes for luminosity monitoring and calibration. The typical Level-1 trigger rate was a few kHz. Events accepted by the Level-1 trigger were transferred to the DAQ system, where software selections using partial event reconstruction further reduced the sample before data were written to mass storage.

Belle II uses a two-level trigger system designed for the higher rates at Su-

perKEKB [28]. The Level-1 hardware trigger operates with a fixed latency of about $5\ \mu\text{s}$ and a maximum (current) accept rate of 30(10) kHz. It receives input from the CDC, ECL, TOP, and KLM, and field-programmable gate arrays implement the trigger logic and distribute global decisions to the front-end electronics. The high-level trigger (HLT) is a software trigger running on an array of processors. It receives events from the readout electronics via a multi-stage event builder and performs more complex reconstruction and selections, including track and vertex reconstruction in selected cases. The HLT reduces the rate of accepted events by about a factor of 5–10 for permanent storage.

The Belle II DAQ system uses a modular architecture with common readout boards, optical links, and high-throughput network switches for event building. A particular challenge is the integration of PXD data, which are read out separately and merged with regions of interest defined by the HLT. This is necessary to keep the PXD data volume manageable while retaining the precision vertex information provided by the pixel detector.

2.2.7 Beam-induced backgrounds and Belle II upgrades

The increase in luminosity at SuperKEKB compared to KEKB is accompanied by substantially higher beam-induced backgrounds from Touschek scattering, beam–gas interactions, synchrotron radiation, and radiative Bhabha processes. These backgrounds lead to elevated occupancies and radiation doses in the inner detectors and increased noise and fake hits in the outer subsystems.

The Belle II detector upgrades were designed to cope with this environment. The vertex detector uses finely segmented, radiation-tolerant sensors and fast readout electronics. The CDC has smaller cells and faster electronics than the Belle CDC. The TOP and ARICH replace the Belle ACC and TOF systems with detectors capable of high-rate operation and improved particle-identification performance. The ECL readout electronics were upgraded to use waveform sampling and improved timing, while the KLM endcaps were upgraded to scintillator-based detectors. Additional shielding, collimation, and online and offline background filters are used to reduce the impact of beam backgrounds.

The first long shutdown of Belle II in 2022–2023 was used for several major detector and accelerator upgrades. On the detector side, the completed PXD2 pixel detector was installed, conventional TOP MCP-PMTs were replaced with longer-lifetime ALD MCP-PMTs, and the back-end readout was upgraded from the COPPER-based system to PCIe40 for several subdetectors. The CDC gas system and monitoring were also improved. On the accelerator side, work included improvements to the injection system, shielding near the interaction region, and collimation, including the installation of a non-linear collimator. These upgrades are critical for maintaining detector performance as the SuperKEKB approaches higher luminosities.

2.2.8 Performance summary

The combination of precise vertexing, high-efficiency tracking, robust particle identification, good calorimetry, and large hermetic coverage made Belle a highly successful detector for B physics. Key Belle performance figures included impact-parameter resolutions of order 20–30 μm , relative transverse-momentum resolution around 0.5% at $p_T \sim 1\ \text{GeV}/c$, K/π separation over a broad momentum range using CDC, ACC, and TOF information, and percent-level photon and electron energy resolution in

the ECL. During its operation, Belle recorded approximately 711 fb^{-1} at the $\Upsilon(4S)$ and additional data at other energies.

Belle II extends this programme with improved vertex resolution from the upgraded vertex detector, excellent momentum resolution and tracking efficiency from the larger CDC, improved charged-particle identification from the TOP and ARICH, upgraded calorimetry and KLM systems, and a modern trigger and DAQ system designed for high-background conditions. These capabilities, together with the large integrated luminosity provided by SuperKEKB, enable precision measurements and rare-decay searches that extend and surpass the achievements of Belle. The dataset recorded by Belle II used in this work is smaller than the one recorded by Belle and equals approximately 365 fb^{-1} at the $\Upsilon(4S)$ resonance, which motivates the usage of both datasets for increased signal yields.

The analysis presented in this work exploits many of the detector features. Precise vertexing is needed for reconstructing displaced K_S^0 candidates, tracking and particle identification are essential for selecting τ decays with one charged daughter particle, calorimeter information is used for neutral-pion reconstruction and rest-of-event observables, and the overall hermeticity of Belle and Belle II is crucial in a final state with multiple neutrinos.

2.3 Summary

In this chapter we introduced the Belle and Belle II detectors at the KEKB and SuperKEKB e^+e^- colliders. KEKB operated with asymmetric beam energies of 8 GeV electrons and 3.5 GeV positrons at the $\Upsilon(4S)$ resonance, delivering a total integrated luminosity of about 1 ab^{-1} , of which 711 fb^{-1} were collected at the $\Upsilon(4S)$ by Belle. SuperKEKB is a major upgrade based on the nano-beam collision scheme and revised beam optics, with a design luminosity of $6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$ and a target total (used for this work at $\Upsilon(4S)$ resonance) integrated luminosity of 50 ab^{-1} (365 fb^{-1}) for Belle II.

Belle was a highly successful general-purpose spectrometer featuring a silicon vertex detector, a central drift chamber, aerogel Cherenkov counters, a time-of-flight system, a CsI(Tl) electromagnetic calorimeter, and a KLM system for muon and K_L^0 identification, all immersed in a 1.5 T solenoidal magnetic field. Its performance enabled precise measurements of CP violation, rare decays, and spectroscopy.

Belle II is a comprehensive upgrade of Belle. Its new PXD+SVD vertex detector, larger and finer-grained CDC, TOP and ARICH particle-identification systems, upgraded ECL readout, improved KLM, and modern trigger and DAQ system are designed to cope with the higher luminosity and beam-induced backgrounds of SuperKEKB [28]. These improvements provide better vertex and momentum resolution, enhanced particle identification, and robust operation at much higher luminosities.

The clean e^+e^- environment, excellent detector performance, and sophisticated reconstruction tools available at Belle and Belle II provide a powerful setting for precision studies of rare B decays and searches for physics beyond the Standard Model. In the following chapters, these capabilities are used to reconstruct and select $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ candidates, characterize backgrounds, and extract results from the combined Belle and Belle II datasets.

3. Analysis strategy

The search for the decay $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ is experimentally challenging due to the presence of multiple neutrinos in the final state, which prevents a full kinematic reconstruction of the signal decay. The expected branching fraction is extremely small, while backgrounds from $B\bar{B}$ decays and continuum processes exceed the signal by several orders of magnitude.

To mitigate experimenter bias, the signal-enhanced data region is kept hidden while the analysis procedure is developed. All selection criteria, background-suppression steps, and validation procedures are defined using simulated samples and data control regions only. The data in the signal-enhanced region are inspected only after the analysis strategy has been finalized and approved within the Belle II collaboration. The software used for simulation and reconstruction is discussed in Section 3.1.

The analysis exploits the well-defined initial state of e^+e^- collisions at the $\Upsilon(4S)$ resonance. In the center-of-mass (CM) frame, the $\Upsilon(4S)$ decays into a pair of B mesons produced nearly at rest. By fully reconstructing one of the two B mesons in hadronic decay modes (tag-side), the four-momentum of the signal-side B meson can be inferred. This approach provides strong kinematic constraints and enables powerful background suppression in the presence of missing energy. The algorithm used for this work is based on multivariate classification discussed in Section 3.2 and the exclusive reconstruction of non-signal B -meson is introduced in Section 3.3.

The background suppression is achieved by the usage of multivariate binary classification into signal-like and background-like. The output of the classifier serves as the discriminating variable in the simultaneous binned maximum likelihood fit to the Belle and Belle II data. Such fitting procedure is introduced in Section 3.4.

In case of statistically insignificant signal yields, the measurement can be reinterpreted through setting an upper limit on the rate of the searched for process. The details of such statistical inference are introduced in Section 3.5.

3.1 Dataset, Software and Monte Carlo simulation

As mentioned in the previous chapters, the Belle dataset consists of about 1 ab^{-1} , of which 711 fb^{-1} were collected at the $\Upsilon(4S)$ resonance. The Belle II dataset used for this work (data taking period 2019-2022) consists of 424 fb^{-1} , of which 365 fb^{-1} were collected at the $\Upsilon(4S)$ resonance.

Monte Carlo (MC) simulation plays a central role in the Belle and Belle II analyses. It is used to optimize event-selection criteria, to determine signal efficiencies and acceptances, to estimate backgrounds from various physics processes, and to evaluate systematic uncertainties associated with detector response and modelling, while minimizing potential experimenter bias arising from prior expectations. In both experiments, simulated events are processed through a full detector simulation

and the same reconstruction and analysis chains as the data.

3.1.1 Frameworks and event generators

The Belle and Belle II experiments use similar simulation and reconstruction chains based on well-established software frameworks and detector-simulation tools.

For Belle, the generation and simulation of events are handled by the **basf** (Belle analysis framework) and its associated libraries. Belle II uses the **basf2** software framework [33] (Belle II analysis framework), which integrates event generation, full detector simulation, reconstruction, and analysis in a modular environment.

The main generators employed for the simulation are:

- EVTGEN for the simulation of b and c -hadron decays, including mixing and CP -violating effects [34],
- KKMC for $e^+e^- \rightarrow q\bar{q}$ and lepton-pair production with initial-state radiation [35],
- PYTHIA for hadronisation and fragmentation (PYTHIA8 for Belle II and PYTHIA6 for Belle) in both $q\bar{q}$ and $B\bar{B}$ events [36, 37],
- PHOTOS for the simulation of final-state radiation in decays [38, 39].

In the Belle simulation, beam-induced backgrounds from KEKB (such as beam–gas interactions, Touschek scattering, and synchrotron radiation) are accounted for by overlaying randomly triggered background events onto the simulated signal and background events. This overlay procedure reproduces the occupancies and noise conditions observed in the data and ensures that the MC adequately reflects the detector conditions.

For Belle II, Monte Carlo samples are produced in campaigns corresponding to different data-taking periods and detector configurations. Each campaign is characterized by specific machine and detector conditions, including SuperKEKB optics, beam currents, background levels, and detector status.

Beam-background conditions at SuperKEKB are modelled by dedicated background simulations for each data-taking period. Simulated hits from these background samples (arising from Touschek scattering, beam–gas interactions, radiative Bhabha processes, etc.) are overlaid at the hit level on top of the physics events in the detector simulation. The overlay rates are tuned to match the occupancies observed in the data for each run period [28].

In both experiments, the generated events are passed through the detector simulation (GEANT3 in Belle, GEANT4 in Belle II) to model interactions with the detector material, including energy loss, multiple scattering, electromagnetic and hadronic showers, and secondary interactions.

For the combined Belle+Belle II analysis, Belle events are converted into a Belle II-like data format using the **b2bii** conversion framework [40]. This enables a uniform reconstruction and analysis chain in **basf2**, with common object definitions and calibration strategies. The conversion preserves the kinematics and event structure of the original Belle data and MC while mapping detector objects to the Belle II reconstruction framework.

The reconstructed MC events are then processed through the same calibration and skimming procedures as the real data, ensuring that selection efficiencies and background estimates derived from MC are directly applicable to the analysis.

3.1.2 Monte Carlo samples

The Belle MC samples used in this analysis include:

- Realistic $B\bar{B}$ samples, where $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ is generated and both B mesons decay inclusively into all known channels with branching fractions taken from world averages [3, 6]. The size of the simulated sample used for this work corresponds to 1422 fb^{-1} (1440 fb^{-1}) in Belle (II).
- Continuum $q\bar{q}$ samples, where $q = u, d, s, c$ quarks are produced and hadronize into light- and charm-hadron final states. The size of the simulated sample used for this work corresponds to 1422 fb^{-1} (1440 fb^{-1}) in Belle (II).
- Dedicated background samples of $b \rightarrow u\ell^-\bar{\nu}_\ell$ ($\ell = e, \mu$) and other, rarer $b \rightarrow (s, d, u)$ processes (only in Belle, sample corresponds to $20 - 50 \times 711 \text{ fb}^{-1}$),
- Signal-enriched samples in which one B meson from $\Upsilon(4S)$ decay is forced to decay into the signal mode of interest, while the other decays according to the SM prediction, existing measurements or empirical modeling. In the scope of this work, the decay mode of interest is $B^0 \rightarrow K_s^0\tau^+\tau^-$, for which 100 (60) millions of events are generated for the Belle (II) dataset.

The simulations of other processes are also available, such as $e^+e^- \rightarrow \tau^+\tau^-$ and $e^+e^- \rightarrow \mu^+\mu^-$ and other low-multiplicity processes. They are not utilized in the scope of this work due to negligible contributions to expected backgrounds.

3.2 Multivariate analysis

If we consider a dataset of N elements: $\mathbf{x}_1, \dots, \mathbf{x}_N$, each exhibiting M features $\mathbf{x}_i = (x_i^1, \dots, x_i^M)$, which can be seen as the values of M observed variables for each event. The multivariate classification consists in assigning to each element $\mathbf{x}_i$ a label $\mathcal{O} \in [0, 1]$, which is a probability of how signal-like the element is.

Decision trees are a supervised learning algorithm used for multivariate classification. They consist of chains of nodes with binary decisions, determined by a selection on the features of events, illustrated in Fig 3.1. However, the decision trees are often inaccurate (prone to overfitting statistical fluctuations), therefore boosted decision trees (BDTs) are preferred in high energy physics. Boosting is an ensemble learning method that combines a set of less accurate models (called “weak learners”) to create a single, highly accurate model (a “strong learner”).

In the scope of this work, a **FastBDT** package is used [41], which implements a speed-optimized and cache-friendly gradient-boosted decision tree multivariate classification algorithm. Such classification is used for electron identification in Belle II, the reconstruction of partner (non-signal) B meson, and background suppression for the $B^0 \rightarrow K_s^0\tau^+\tau^-$ search.

3.3 Hadronic B -tag reconstruction

The reconstruction of the tag-side (non-signal) B meson is performed using the Full Event Interpretation (FEI) algorithm. The FEI performs a hierarchical reconstruction of B meson decay chains (only hadronic decay chains are used in the scope of this work) using **FastBDT** multivariate classifiers at each stage, assigning a signal

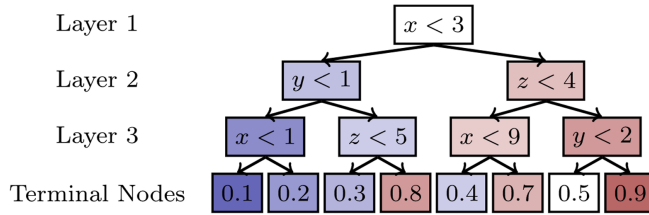


Figure 3.1: Simple example of a decision tree for a binary classifier. Here the features of the dataset are called x, y, z . Decision trees are structured with nodes followed by binary splits, arranged in multiple layers. The terminal node gives the output of the tree. [41]

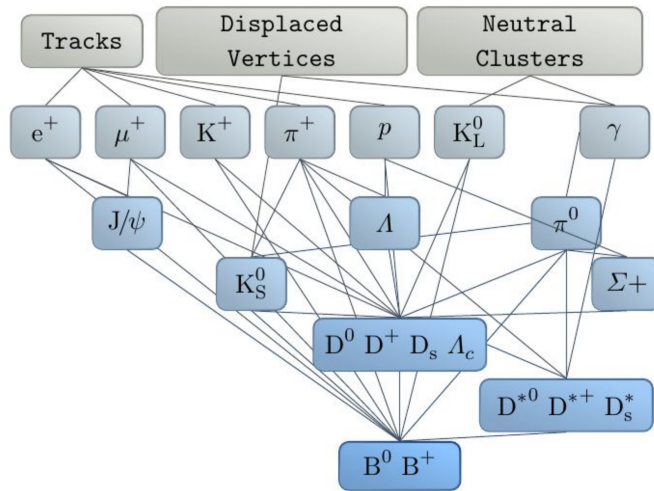


Figure 3.2: Schematic representation of the FEI algorithm. Starting from basic detector-level objects, final states particle are reconstructed and then combined to form intermediate resonances and eventually the final B -tag candidate.

probability $\mathcal{P}_{\text{FEI}}$ to each reconstructed B_{tag} candidate [42]. The hierarchy of the FEI algorithm is depicted in Fig. 3.2.

Kinematic variables can be used to discriminate correctly and wrongly reconstructed B_{tag} candidates. If the B_{tag} is reconstructed in exclusive hadronic decay modes, the beam-constrained mass

$$M_{\text{bc}} = \sqrt{(E_{\text{beam}}^*/c^2)^2 - |\mathbf{p}_B^*/c|^2} \quad (3.1)$$

and the energy difference

$$\Delta E = E_B - E_{\text{beam}}^* \quad (3.2)$$

are used at B -factories. Quantities denoted with asterisk are computed in the CM frame, where the e^+e^- collision is symmetric, therefore $E_{\text{beam}}^* = \sqrt{s}/2$ and the expected energy of both B mesons is $E_B^* = E_{\text{beam}}^*$.

Once the tag-side B meson is reconstructed, the four-momentum of the signal-side B meson can be inferred as

$$p_{B_{\text{sig}}} = p_{\Upsilon(4S)} - p_{B_{\text{tag}}}. \quad (3.3)$$

This information enables the construction of kinematic observables sensitive to missing energy and missing mass, which are essential for suppressing background processes.

3.4 Template-based maximum-likelihood fit

The signal yield is extracted using a binned maximum-likelihood fit to the distribution of the multivariate classifier output, determined from simulation. The fit compares the observed number of events in each bin with the expected signal-plus-background predictions. The likelihood function is constructed as

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_i \text{Poisson}(n_i | \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})) \prod_k C(\theta_k), \quad (3.4)$$

where n_i is the observed number of events in bin i , s_i and b_i are the expected signal and background yields, μ is the signal strength parameter proportional to the branching fraction, and $\boldsymbol{\theta}$ denotes nuisance parameters associated with systematic uncertainties. The functions $C(\theta_k)$ represent constraint terms encoding prior knowledge of the uncertainties.

The information on the parameters of interest μ and the nuisance parameters $\boldsymbol{\theta}$ is encoded in the likelihood function $\mathcal{L}(\mu, \boldsymbol{\theta})$. Point estimates of these parameters are obtained using the maximum-likelihood (ML) principle. The corresponding estimators, denoted by $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$, are defined as the values that maximize the likelihood function.

Equivalently, the ML estimators satisfy the conditions

$$\left. \frac{\partial \mathcal{L}(\mu, \boldsymbol{\theta})}{\partial \mu} \right|_{\mu=\hat{\mu}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \quad \left. \frac{\partial \mathcal{L}(\mu, \boldsymbol{\theta})}{\partial \theta_i} \right|_{\mu=\hat{\mu}, \theta_i=\hat{\theta}_i} = 0, \quad (3.5)$$

for all components θ_i of the nuisance-parameter vector $\boldsymbol{\theta}$.

3.5 Statistical inference

In the absence of a significant signal excess, upper limits on the $\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-)$ are set using the modified frequentist CL_s method. The test statistic is based on the profile likelihood ratio

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \quad (3.6)$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ denote the unconditional maximum-likelihood estimators, and $\hat{\boldsymbol{\theta}}_\mu$ is the conditional estimator for a fixed signal strength μ .

The compatibility of the observed data with a given signal-strength hypothesis μ is quantified using a p -value constructed from the distribution of the test statistic q_μ . For a fixed value of μ , the p -value for the signal-plus-background hypothesis is defined as

$$CL_{s+b}(\mu) \equiv p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu | \mu) dq_\mu, \quad (3.7)$$

where $f(q_\mu | \mu)$ denotes the probability density function of the test statistic under the signal-plus-background hypothesis, and q_μ^{obs} is the value observed in data.

Similarly, the compatibility of the data with the background-only hypothesis ($\mu = 0$) is quantified by the corresponding p -value

$$CL_b \equiv p_0 = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu | 0) dq_\mu. \quad (3.8)$$

The modified frequentist confidence level CL_s is defined as the ratio of these two quantities,

$$CL_s(\mu) \equiv \frac{CL_{s+b}(\mu)}{CL_b}. \quad (3.9)$$

This construction reduces the probability of excluding signal hypotheses to which the analysis has little sensitivity.

An upper limit on the signal strength μ at confidence level $1 - \alpha$ is obtained by determining the largest value of μ for which

$$CL_s(\mu) \geq 1 - \alpha. \quad (3.10)$$

4. Event selection

This chapter describes the reconstruction and selection of $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ candidates. The analysis targets events in which the companion B meson is fully reconstructed in a hadronic decay mode, while the signal side contains a K_s^0 candidate and two τ -daughter candidates, where only τ -decays with a single charged particle (one-prong) are considered. This topology provides sensitivity to the signal while retaining enough kinematic information to suppress backgrounds with missing energy.

The event reconstruction is organized around three parts of the event:

1. the fully reconstructed tag-side B meson, denoted B_{tag}^0 ;
2. the signal-side $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ candidate, reconstructed from particles not used in the B_{tag}^0 candidate;
3. the rest of event (ROE), defined as all remaining reconstructed objects not associated with either the tag-side or signal-side candidate.

The chapter first describes the reconstruction of these objects and the baseline selection criteria. It then introduces the categorization of signal candidates according to the reconstructed τ -daughter composition, followed by the best-candidate selection and the resulting signal efficiencies and background yields.

4.1 Partner B reconstruction

The reconstruction of the tag-side B meson is performed using the hadronic Full Event Interpretation algorithm [42] for B^0 mesons. This algorithm exclusively reconstructs B^0 candidates in the 32 decay modes listed in Tab. 4.1. By fully reconstructing the B_{tag}^0 , we can infer the kinematic properties of the $B_{\text{sig}}^0 \rightarrow K_s^0 \tau \tau$, which would be impossible to get directly as it has from 2 to 4 neutrinos in the final state. At the cost of the low tag-side efficiency, we obtain a pure sample: in signal MC, a purity of $\sim 70\%$ can be achieved for an efficiency of $0.4 \div 0.5\%$.

For the tag side selection, we use the signal probability requirement ($\mathcal{P}_{\text{FEI}}$) to reduce the combinatorial background coming from continuum events and the mis-reconstructed $B\bar{B}$ events. As can be seen from Fig. 4.1 (left) the cut $\log_{10}(\mathcal{P}_{\text{FEI}}) > -2$ greatly helps to purify our B_{tag} sample: by reducing the truth-matched B_{tag} efficiency of $\sim 7\%$, we can reject $\sim 64\%$ of background. This cut can hence be seen as an efficient online pre-filter, which speeds-up the processing and reduces the output size.

In addition to the $\mathcal{P}_{\text{FEI}}$, the beam-constrained mass M_{bc} is also used: with the rather tight requirement $M_{bc} > 5.27 \text{ GeV}/c^2$ we can remove the combinatorial background and reduce the mis-reconstructed backgrounds where $\mathbf{p}_{B_{\text{tag}}}$ is not properly reconstructed. The $-0.15 \text{ GeV} < \Delta E < 0.1 \text{ GeV}$ requirement on the energy difference is also applied to further suppress misidentified B_{tag} 's with wrong $E_{B_{\text{tag}}}$. This

Table 4.1: FEI B^0 hadronic tag modes.

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

cut is not as effective as that on M_{bc} , given that the FEI classifier uses ΔE in the training, resulting in a peaking distribution even for combinatorial backgrounds (Fig. 4.2). Finally, for each event the B_{tag} candidate with the highest $\mathcal{P}_{\text{FEI}}$ is retained to remove the candidate multiplicity coming from the B_{tag} reconstruction.

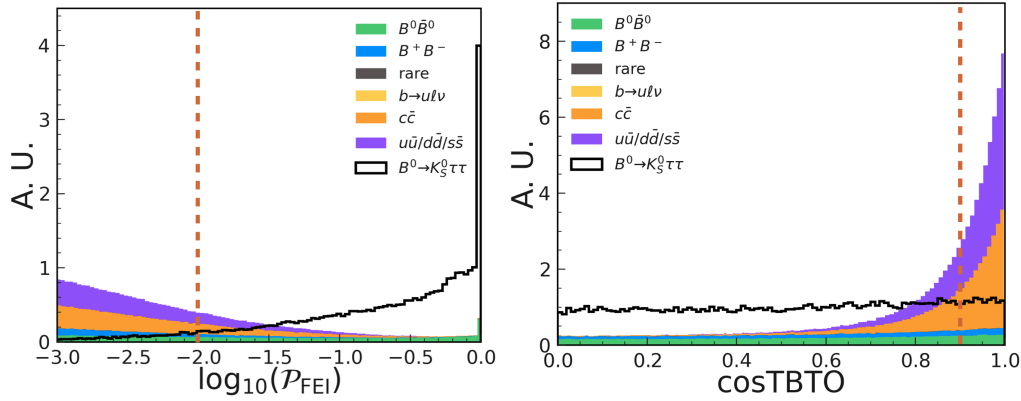


Figure 4.1: Selection cuts on the FEI signal probability (called $\mathcal{P}_{\text{FEI}}$) and on the $\cos T_{\text{BT0}}$ (Belle MC). B_{tag} candidates of signal events (in black) are truth-matched.

Another selection requirement related to the B_{tag} is the cut $\cos T_{\text{BT0}} < 0.9$ (Fig. 4.1, right). In fact, this variable is defined as the cosine of the angle between the thrust axis of the B_{tag} and thrust axis of everything else in the event (B_{sig} and rest of event). For a collection of N momenta, the thrust axis is the unit vector along which their total projection is maximal.

4.2 Signal B reconstruction

The exclusive reconstruction of the B_{tag} also allows to know its flavor and hence that of the companion B (B_{sig}). As will be explained later in this section, the

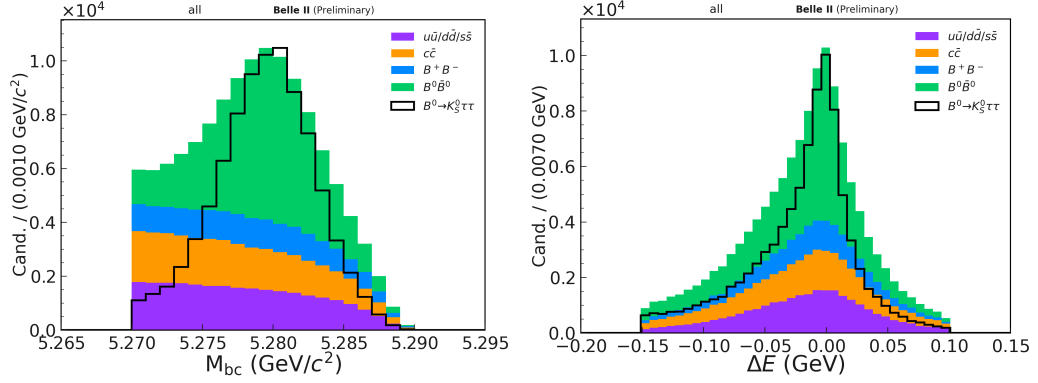


Figure 4.2: M_{bc} and ΔE distributions of signal and background candidates with $\mathcal{P}_{\text{FEI}} > 0.01$ (Belle II MC).

flavor information is used for the background suppression. The two τ -daughters candidates, completely symmetric in signal events, have instead specific signatures depending on their charge (as a function of the B flavor) in background events. Due to $B^0 - \bar{B}^0$ oscillations, the flavor information inferred from the B_{tag} is not always correct. The fraction of events with a pair of same-flavor B -mesons is equal to $\chi_d = 0.1874 \pm 0.0018$ [3], the time-integrated mixing probability for the B^0 meson, assuming no CP violation. On the contrary, for unmixed events, the flavor of the B_{sig} candidate can be correctly inferred, provided that the B_{tag} meson is properly reconstructed.

The B_{sig} candidates consists of three particles: one K_s^0 and two oppositely charged particles, each corresponding to a τ^- and a τ^+ lepton. The K_s^0 candidate is selected in Belle using from $\pi^+\pi^-$ pairs. We apply a loose mass requirement after performing a vertex fit using (though no cut on the vertex quality is applied, except for the existence of the vertex itself). We also require the candidate to satisfy the cut `goodBelleKshort==1` (gBKS). It consists of a binned selection based on the following variables:

- **p** The K_s^0 momentum;
- **$d\mathbf{r}$** $\min(d_1, d_2)$ where d_i is the d_0 computed with respect to the IP of the i -th K_s^0 daughter;
- **$d\phi$** The azimuthal angle between the momentum vector and the decay vertex vector of the K_s^0 candidate;
- **z_dist** The distance in z between the two daughter tracks at their measured vertex;
- **fl** The flight length of K_s^0 candidate in the x - y plane.

The exact combination of cuts can be checked in Tab. 4.2.

This criterion helps to significantly reduce the combinatorial background in signal events (98%) while retaining around 80% of truth-matched (TM) K_s^0 's, as can be seen in the plot of Fig. 4.3. For background events, the rejection of wrongly reconstructed K_s^0 is even more effective; see Fig. 4.4, which shows the large combinatorial suppression, at the cost of little signal inefficiency.

Table 4.2: `goodBelleKshort` cuts.

$p(\text{GeV}/c)$	$dr(\text{cm})$	$d\phi(\text{rad})$	$z_dist(\text{cm})$	$fl(\text{cm})$
<0.5	>0.05	<0.3	<0.8	-
$0.5-1.5$	>0.03	<0.1	<1.8	>0.08
>1.5	>0.02	<0.03	<2.4	>0.22

 Table 4.3: Efficiency of the `goodBelleKshort` cuts..

	Belle	Belle II	Belle/Belle II
dr	0.924	0.911	1.013
$d\phi$	0.861	0.884	0.974
z_dist	0.824	0.871	0.946
fl	0.823	0.869	0.947

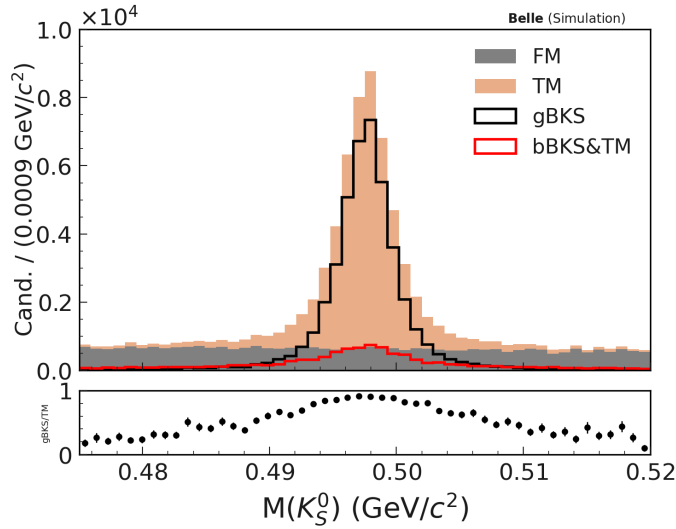


Figure 4.3: Invariant mass of K_S^0 candidates divided into truth-matched (TM) and false-matched (FM) categories. The candidates passing the `goodBelleKshort==1` candidates (gBKS) are overlaid (black). Those that do not pass the `goodBelleKshort` cut (bBKS) but are true K_S^0 are represented in red. The lower pad contains the ratio between the gBKS and TM candidates (Belle MC).

In Belle II, we select K_S^0 from the from $\pi^+\pi^-$ pairs in such a way that the displaced-vertex candidates are prioritised with respect to the others. The vertex fitting is performed and the K_S^0 daughters' momenta are updated after the fit. A similar background rejection coming from random combinations of (π^+, π^-) pairs is observed in Belle compared to Belle II. However, a higher signal efficiency is observed when applying the cut `goodBelleKshort==1`. This effect can be seen in Fig. 4.5: for K_S^0 with decay vertex distances above $\sim 1\text{cm}$ the efficiency is higher for the Belle II sample.

The other two particle objects forming the B_{sig} candidates are the two τ daughters, called $t_{\text{correlated}}$ and $t_{\text{uncorrelated}}$ based on their charge sign compared to that of the B_{sig} 's b -quark ¹. The distinction between *correlated* and *uncorrelated* particles

¹For instance, the possible combinations are either $\bar{B}_{\text{tag}}^0 B_{\text{sig}}^0 (\rightarrow K_S^0 t_c^- t_u^+)$ or $B_{\text{tag}}^0 \bar{B}_{\text{sig}}^0 (\rightarrow K_S^0 t_c^+ t_u^-)$

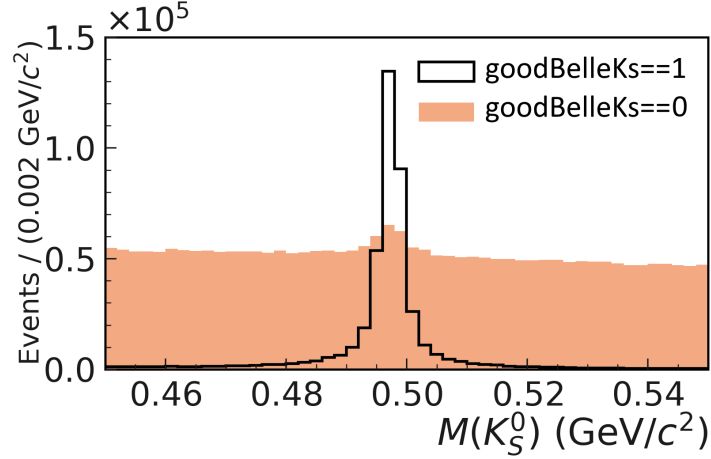


Figure 4.4: Invariant mass of the K_S^0 candidates in $B^0 \bar{B}^0$ background events (Belle II MC).

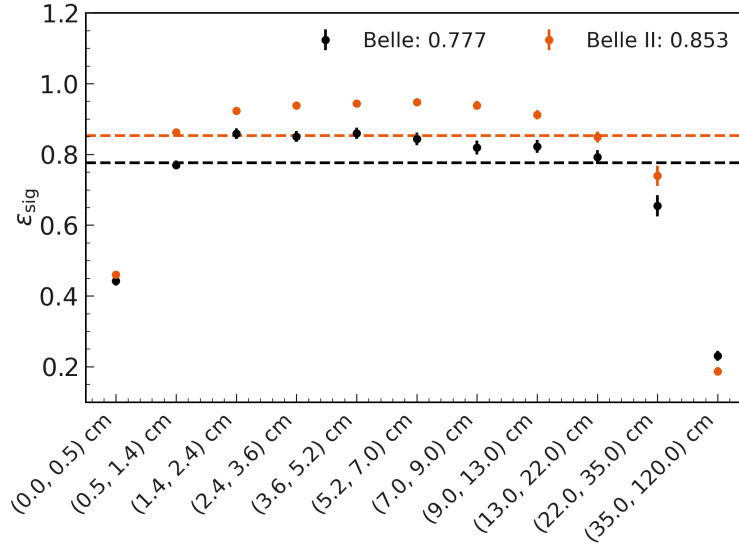


Figure 4.5: `goodBelleKshort==1` cut efficiency as a function of the 3D-vertex distance. The orange (black) points correspond to Belle (II) signal samples.

helps in background suppression, as will be explained in Section 6. For example, in $B \rightarrow DX$ transitions, the $t_{\text{correlated}}$ would correspond to the charged particle coming from the D along with the kaon, while $t_{\text{uncorrelated}}$ would rather come directly from the B . For simplicity, $t_{\text{correlated}}$ and $t_{\text{uncorrelated}}$ will be from now on referred to as t_c and t_u , respectively.

As we target the one-prong τ decays, we select two oppositely charged tracks, to which only impact parameter cuts ($dr < 2\text{cm}$ and $|dz| < 4\text{cm}$) and a cut on the minimal momentum $p > 50\text{MeV}/c$ are applied. The visible decay composition of individual τ leptons and of $\tau^+ \tau^-$ pairs is shown in Fig. 4.6; the large fraction of modes with exactly one charged particle motivates this signal-side topology. We explicitly store the information about particle identity type for each track. The cri-

teria used to categorize our tracks as leptons or hadrons are described in Section 4. We also reconstruct $\tau \rightarrow \rho\nu$ decays by combining the selected tracks to a π^0 candidate. Neutral pions are reconstructed using photons with energies of more than 120 MeV, 30 MeV, and 80 MeV in the forward, barrel, and backward regions of the ECL, respectively. We also require the polar angle of the photons to lie in the range $(17^\circ, 150^\circ)$ and apply quality cuts on the number of crystals in the cluster associated to the photon, and on the energy distribution among the crystals within the cluster. Photon pairs are combined to form a π^0 candidate if their invariant mass is in the range $(121, 142)$ MeV/ c^2 , and their maximum angular separation is 1.0 rad in the transverse plane and 0.9 rad in the three-dimensional space. We also require the π^0 to have momenta in the range $(0.2, 3.0)$ GeV/ c . We save the ρ candidate if the $\pi^\pm\pi^0$ invariant mass belongs to the range $(0.48, 1.07)$ GeV/ c^2 . In order to get a reasonable candidate multiplicity, we consider only two random π^0 candidates per event. This criterion does not bias the final $M(\pi^0)$ distribution, which is restricted to the range $(0.120, 0.150)$ GeV/ c^2 , nor sculpts the $M(\pi^\pm\pi^0)$ – that will be used for the background suppression. The π^0 multiplicity is small in the signal sample, justifying the choice of random π^0 candidate selection.

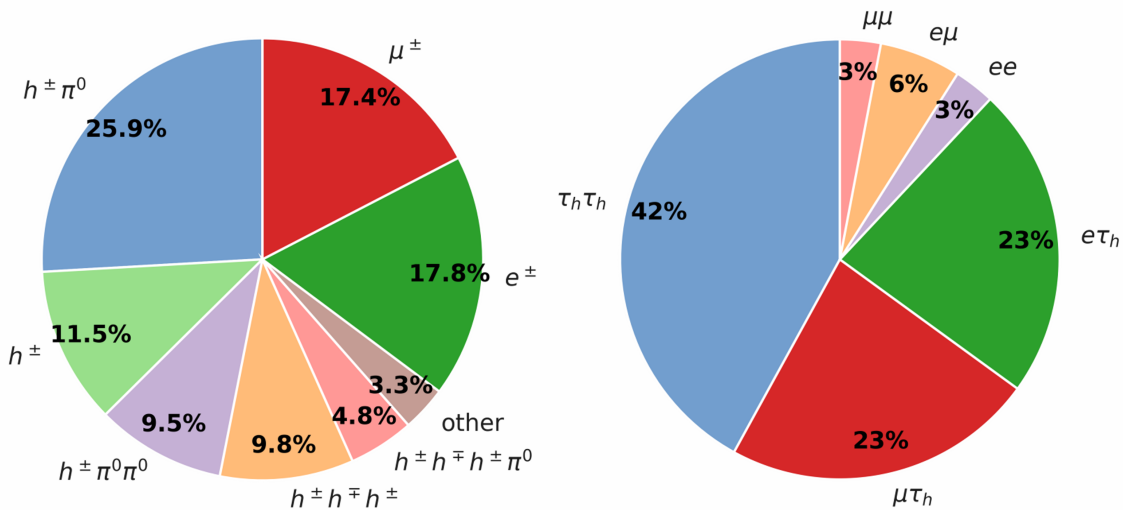


Figure 4.6: Composition of visible τ -lepton decay modes. The left panel shows the main decay modes of a single τ lepton, grouped according to the visible charged and neutral particles in the final state. The right panel shows the corresponding visible final-state categories for a $\tau^+\tau^-$ pair. Neutrinos, which are present in all τ decays, are not shown explicitly. One-prong τ decays make up about 85% of all τ decays and define the signal-side topology used in this analysis.

A further cut is applied on the q^2 variable, computed using the reconstructed B_{tag} and K_s^0 4-momenta following the formula:

$$q_{\text{rec}}^2 = m_B^2 + m_{K_s^0}^2 - 2 \left(E_{\text{beam}}^* E_{K_s^0}^* / c^4 + \mathbf{p}_{B_{\text{tag}}}^* \cdot \mathbf{p}_{K_s^0}^* / c^2 \right) \quad (4.1)$$

where m_B is the PDG B^0 mass [3], E_{beam}^* is the beam energy in the center-of-mass frame. All other values correspond to the reconstructed ones. The q^2 resolution can be seen in Fig. 4.7 (left) for different event categories, based on whether the B_{tag} and/or the signal K_s^0 are correctly reconstructed. We use true-level information

from simulation distinguish the truth-matched (TM) objects and the false-matched (FM) ones. For TM K_S^0 's and TM (FM) B_{tag} 's, the resolution is $\sim 0.07\text{GeV}^2$ ($\sim 0.2\text{GeV}^2$), with slightly better resolutions for Belle II.

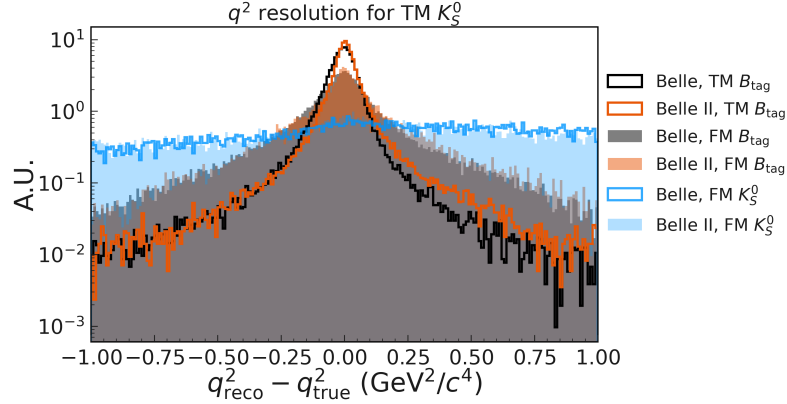


Figure 4.7: On the left, q^2 resolution for truth-matched (TM) and false-matched (FM) B_{tag} 's or K_S^0 's. On the right, Belle and Belle II signal events are compared.

We apply the cut $q_{\text{rec}}^2 > 12\text{GeV}^2/c^4$ as this allows to reject $\sim 17\%$ of background events with a minimal signal loss ($\sim 0.5\%$). In fact, the minimum allowed q^2 for $B \rightarrow K\tau\tau$ decays is $q_{\text{min}}^2 = 4m_\tau^2 \sim 12.63\text{GeV}^2/c^4$. Therefore our cut should retain all signal events and the little signal inefficiency is due to resolution effects (Fig. 4.8).

We stress that the SM predictions for $B^0 \rightarrow K^0\tau\tau$ decays are computed in the full range $q^2 \in (4m_\tau^2, q_{\text{max}}^2) \text{ GeV}^2/c^4$ ($\mathcal{B} = 1.55(11) \times 10^{-7}$) or $q^2 \in (14.18, q_{\text{max}}^2) \text{ GeV}^2/c^4$ ($\mathcal{B} = 1.375(95) \times 10^{-7}$) to exclude the long-distance contribution from $B^0 \rightarrow \psi(2S)(\tau\tau)K^0$ decays [16]¹. We do not exclude the resonant contribution because it is very small $\mathcal{B}(B^0 \rightarrow K_S^0\psi(2S)(\rightarrow \tau^+\tau^-)) \sim 9 \times 10^{-7}$.

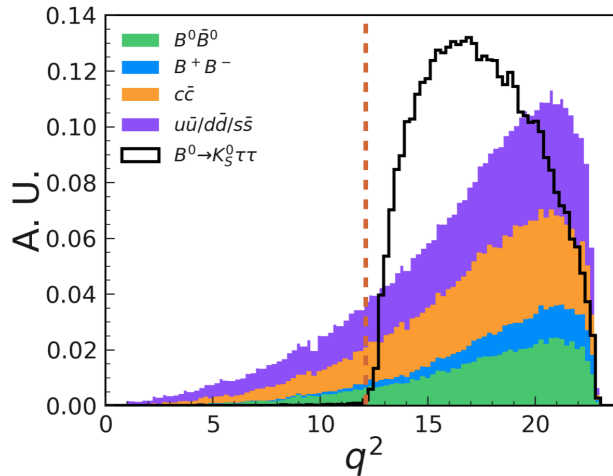


Figure 4.8: Selection cut on the reconstructed q^2 (Belle MC).

4.3 Rest of Event

Once the B_{tag} and B_{sig} are selected, the remaining tracks and ECL clusters in the event (rest of event, or ROE) should be treated. We require no tracks in ROE

¹In the above sentence and throughout this note, q_{max}^2 equals $(m_B - m_K)^2 \sim 23\text{GeV}^2/c^4$.

(tracks with the requirement $dr < 2\text{cm}$ and $|dz| < 4\text{cm}$), as it should be the case for well-reconstructed $\Upsilon(4S)$ with two 1-prong τ decays. Concerning the residual deposits in the ECL, we apply some selection to suppress those coming from beam backgrounds and those related to clustering issues or hadronic split-offs (the so-called *fake photons*).

The variables we use to separate the true photons from the collision events from the rest are:

- **minC2TDist** defined as the distance between the ECL cluster and the closest track in the event. The distance is computed considering the extrapolated hits in the ECL volume. The **minC2TDist** variable can help to suppress the fake photon component: ‘real’ photons coming from collision or from beam backgrounds) tend to have higher isolation, hence larger distance with the closest track.
- cluster energy. Photons from beam background tend to have lower energies and hit the endcap regions more than the barrel. For this reason, usually a region-dependent cut on the cluster energy with higher threshold for the backward/forward photons is enough to largely suppress background. At Belle, the thresholds were defined as:

1. $\theta \in (18^\circ, 37^\circ)$ (forward region): $E_\gamma > 0.10\text{GeV}$
2. $\theta \in (47^\circ, 122^\circ)$ (barrel region): $E_\gamma > 0.05\text{GeV}$
3. $\theta \in (130^\circ, 155^\circ)$ (backward region): $E_\gamma > 0.15\text{GeV}$.

A qualitative representation of how *true*, *fake* and *beam-induced* ECL clusters appear in the energy, **minC2TDist** and polar angle dimensions is shown in Fig. 4.9.

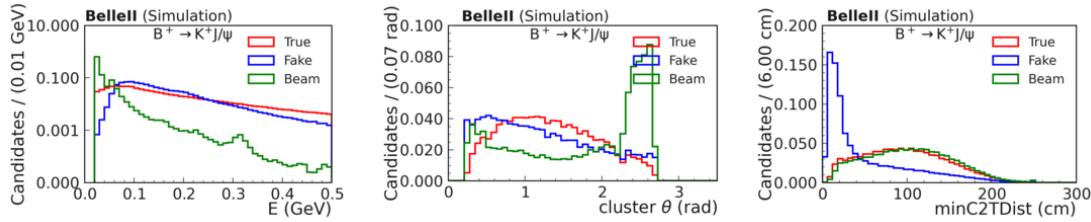


Figure 4.9: Energy, polar angle and **minC2TDist** of ECL clusters for simulated Belle II events (Obtained from $B^+ \rightarrow J/\psi K^+$ validation samples, plots are meant as illustrative).

4.4 Event categorization

Once the main particle objects are reconstructed, we combine them to form $\Upsilon(4S)$ candidates into the following modes:

- $\Upsilon(4S) \rightarrow K_S^0 B_{\text{tag}}^0 \text{ track}^+ \text{ track}^-$
- $\Upsilon(4S) \rightarrow K_S^0 B_{\text{tag}}^0 \rho^+ \text{ track}^-$
- $\Upsilon(4S) \rightarrow K_S^0 B_{\text{tag}}^0 \text{ track}^+ \rho^-$

- $\Upsilon(4S) \rightarrow K_S^0 B_{\text{tag}}^0 \rho^+ \rho^-$

To each track a probability of the corresponding particle belonging to a specific particle type is assigned. In fact, as mentioned earlier, we target the $\tau^+ \rightarrow \{e^+ \nu_e, \mu^+ \nu_\mu, h^+, \rho^+\} \bar{\nu}_\tau$ decays (with $\rho^- \rightarrow \pi^+ \pi^0$), which are selected with multiple requirements on the tracks. Leptons are identified using selectors $\mathcal{P}^\ell$ that combine information from multiple subdetectors. In Belle, electron identification relies on the ECL, CDC, and ACC [43], while muons are identified with the KLM [44]; in Belle II, all subdetectors except the PXD and SVD contribute [45]. The selectors are likelihood-ratio-type observables, except for electrons in Belle II where a BDT classifier is used. We require $\mathcal{P}^{\mu(e)} > 0.9$, $p > 200$ MeV/c and polar-angle ranges $17^\circ\text{--}150^\circ$ ($18^\circ\text{--}151^\circ$) for muons (electrons) in Belle, and $23^\circ\text{--}150^\circ$ ($13^\circ\text{--}155^\circ$) in Belle II. The resulting efficiencies are 77% (81%) for electrons and 44% (67%) for muons in Belle (Belle II), while the misidentification rates are below 1% (1%) and 6% (19%), respectively, as estimated from signal MC. Other tracks are identified as hadrons (h/ρ).

Given that two τ 's are present in our B_{sig} , we end up with ten possible final states. We rank them based on the fact that leptonic final states should be preferred over hadronic and that μ should be preferred over e . Furthermore, whenever there is a ρ candidate, there must also be a π one; in that case we prefer the former given the further mass constraint (despite not being particularly loose). The variable

$$\text{ID} = (\text{number of muons}) \cdot 3.1 + (\text{number of electrons}) \cdot 3.0 + (\text{number of } \rho \text{ mesons}) \cdot 0.5$$

is designed to assign an importance score to each of the ten $\tau\bar{\tau}$ final states.

Table 4.4: Possible final states of $\tau\bar{\tau}$ decays with the corresponding ID and category assignment. Importance of category assignment is based on the descending order of the ID variable.

$\tau\bar{\tau}$ final state	ID	Category
$\mu \mu$	6.2	$\ell\ell$
$e \mu$	6.1	
$e e$	6.0	
$\mu \rho$	3.6	$\rho\ell$
$e \rho$	3.5	
μh	3.1	$\ell_c h_u, \ell_u h_c$
$e h$	3.0	
$\rho \rho$	1.0	$\ell 0$
ρh	0.5	
$h h$	0.0	

Table 4.4 shows the ID number for each final state. With these criteria, for each event we pick the candidate(s) belonging to the final state with highest ID. The candidates are then divided into the four orthogonal categories defined as:

- $\ell\ell$ (lepton-lepton category): $\text{ID} \geq 6.0$
- $\rho\ell$ (ρ -lepton category): $\text{ID} \geq 3.5$ and $\text{ID} < 4$
- ℓh (lepton-hadron category): $3.0 \leq \text{ID} < 3.5$

- $\ell 0$ (lepton-less inclusive category): ID < 2.

The ℓh category is further split into $\ell_c h_u$ and $\ell_u h_c$ categories for a more effective background rejection. The two subsets of ℓh have the same selection, and are split via an orthogonal cut on whether we reconstruct the t_c as a lepton or not. Doing so, we raise the amount of categories to 5. At this stage, there can be multiple $\Upsilon(4S)$ candidates belonging to the same category (i.e. having the same ID). The residual candidate multiplicity, shown in Fig. 4.10 (Belle II) and Fig. 4.11 (Belle), can be due to many reasons. First, we do not truth-match here our signal events, allowing for more swaps between signal and tag side side and for candidate multiplicity coming, for example, from the π^0 multiplicity (we allow for two π^0 's in the event because of the $B^0 \rightarrow K_s^0 \rho^+ \rho^-$ final state). Another source of multiplicity is the K_s^0 : there are cases where the K_s^0 is truth-matched but one (or both) of its daughters generates clone tracks. This is the main cause of multiplicity in, for example, the $\ell \ell$ category.

Because of the several reasons behind the residual multiplicity, we decide to remove it by randomly choosing a single candidate. The fractions of preserved fully truth-matched signal events after performing a random candidate selection are 99.93% for the $\ell \ell$ category, 99.84% for the $\ell_c h_u$ category, 99.81% for the $\ell_u h_c$ category, 92.43% for the $\rho \ell$ category, 80.67% for the $\ell 0$ category. Similar numbers are obtained for the Belle MC: $\ell \ell$ - 98.89%, $\ell_c h_u$ - 97.11%, $\ell_u h_c$ - 96.93%, $\rho \ell$ - 94.40%, $\ell 0$ - 83.19%. The higher preserved amount of fully truth-matched signal events for the $\rho \ell$ and $\ell 0$ in the Belle case is related to slightly lower candidate multiplicity, therefore increased efficiency of random candidate selection. The explanation of the truth-matching logic is given later in the text and is consistent with the one used in Tab. 4.7.

In Tab. 4.5 the true τ decay mode is shown as a function of the reconstructed one (for Belle II MC). Throughout this note, we obtain the *true* categories using **topoAna** [46], a C++-based tool providing full access to the whole generated chain in MC from the generation-level information of the samples. One can see that $\sim 25\%$ of the $\tau \rightarrow e$ decays are mis-categorized as $\tau \rightarrow h$ due to the low momentum and that the fraction is even higher for $\tau \rightarrow \mu$ decays taken as $\tau \rightarrow h$ (the minimum p_T to reach KLM, $\sim 500\text{MeV}$, already removes 40% of the $\tau \rightarrow \mu$ events, as demonstrated in Fig. 4.12). Among the reconstructed $\tau \rightarrow h$ we have a large fraction of $\tau \rightarrow \pi\pi^0$ decays, for which the π^0 was lost due its low momentum and finally, in the reconstructed $\tau \rightarrow \rho$ category, a significant fraction comes from $\tau \rightarrow \pi\pi^0\pi^0$ events or other decays.

We also show in Tab. 4.6 the composition of signal events in terms of reconstructed $\tau^+\tau^-$ decay combinations. In Belle II, most of the events belong to the $\ell + \text{hadron}$ category, given the fact that $\tau \rightarrow \ell$ and $\tau \rightarrow \rho$ have the largest branching fractions (34% and 25%, respectively), and that, as seen before, many ρ 's are reconstructed as pions. For Belle, the largest fraction of events falls into the hh category, due to the lower lepton-ID efficiency.

While Table 4.6 gives averaged results over all the categories, Tables E.1 to E.5 in Appendix E show the true composition in terms of t_c, t_u pairs (for Belle results see Appendix F). One can see that for the $\ell \ell$ (Tab. E.1) category $\sim 80\%$ of the events are correctly selected and that for the $\ell_{c,u} h$ (Tab. E.2, Tab. E.3) categories, $\sim 90\%$ of the events have $\ell \equiv t_{c,u}$. Among the $\rho \ell$ (Tab. E.4) events, only $\sim 50\%$ of them are real $\tau \rightarrow \pi\pi^0\nu \times \tau \rightarrow \ell\nu\nu$, and finally, for the $\ell 0$ (Tab. E.5) category, $\sim 80\%$ of the events have at least one $\tau \rightarrow \pi\pi^0\nu$ decay.

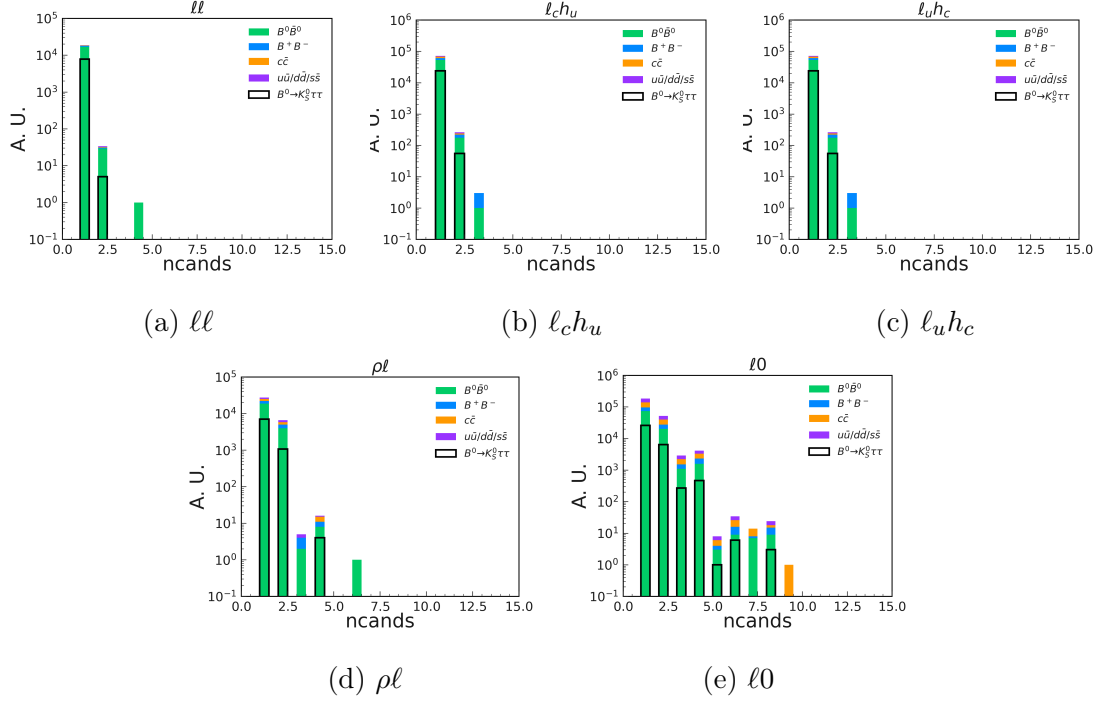


Figure 4.10: $\Upsilon(4S)$ multiplicity per category before the random candidate selection within the same category. The mean candidate multiplicities for $\ell\ell$, $\ell_c h_u$, $\ell_u h_c$, $\rho\ell$ and $\ell 0$ are 1.0006, 1.0023, 1.0023, 1.1306, 1.2526 respectively. Belle II MC signal and backgrounds are shown.

Table 4.5: Belle II MC true composition of the reconstructed decay mode of the τ^+ (only one track taken). Integrated among all reconstructed categories. The table is normalized to unity. Last row (column) shows the integrated values for a given column (row).

reco true \	e	μ	ρ	h	
e	0.16		0.01	0.04	0.21
μ		0.14	0.01	0.06	0.21
π		0.01	0.02	0.11	0.13
$\pi \pi^0$		0.02	0.10	0.17	0.29
$\pi \pi^0 \pi^0$		0.01	0.03	0.05	0.09
K				0.01	0.01
rest			0.02	0.04	0.06
	0.16	0.17	0.19	0.48	

Finally, we check the goodness of our reconstruction by counting the number of events with good B_{tag} 's and good B_{sig} 's. While for the B_{tag} we use the `basf2` variable `isSignal` to distinguish between good (TM) and bad (FM) tags, for signal we check one by one the possible causes of mis-reconstruction – whether something is wrong in the reconstruction of K_S^0 , t_c or t_u or all of them. The numbers are summarized in Tab. 4.7 for the five categories. It can be seen that most of the B_{tag} 's are correct and that $\sim 60\%$ of events are totally correct for the $\ell\ell, \ell h$ categories. A large fraction

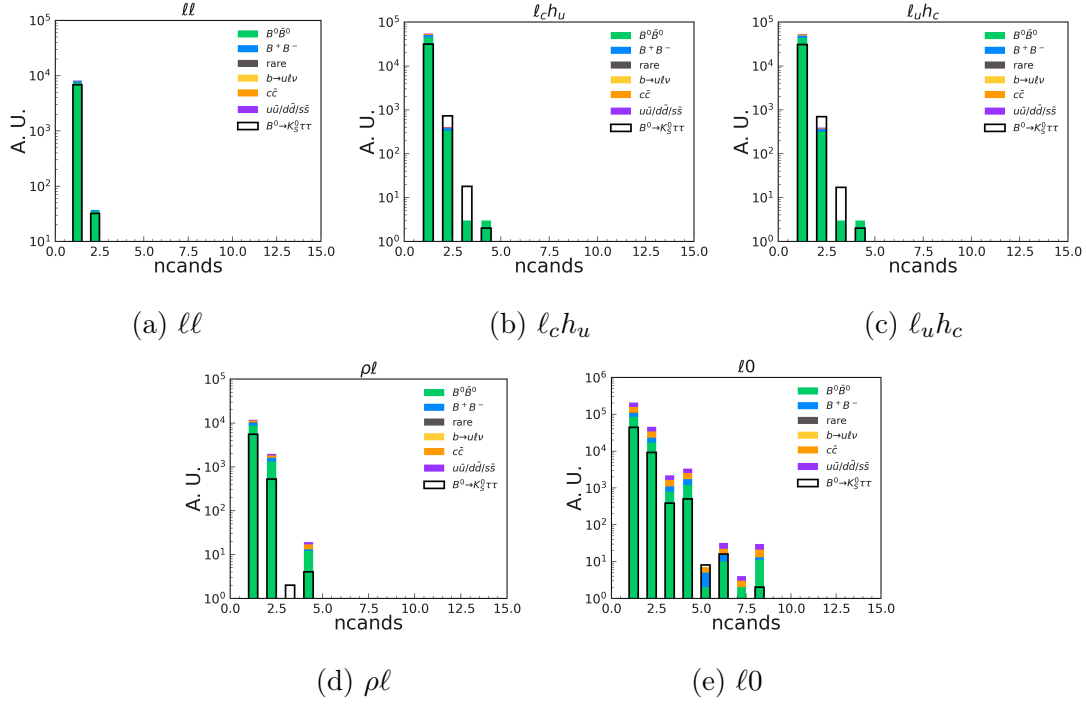


Figure 4.11: $\Upsilon(4S)$ multiplicity per category before the random candidate selection within the same category. The mean candidate multiplicities for $\ell\ell$, $\ell_c h_u$, $\ell_u h_c$, $\rho\ell$ and $\ell 0$ are 1.0046, 1.0238, 1.0234, 1.0887, 1.2151 respectively. Belle MC signal and backgrounds are shown.

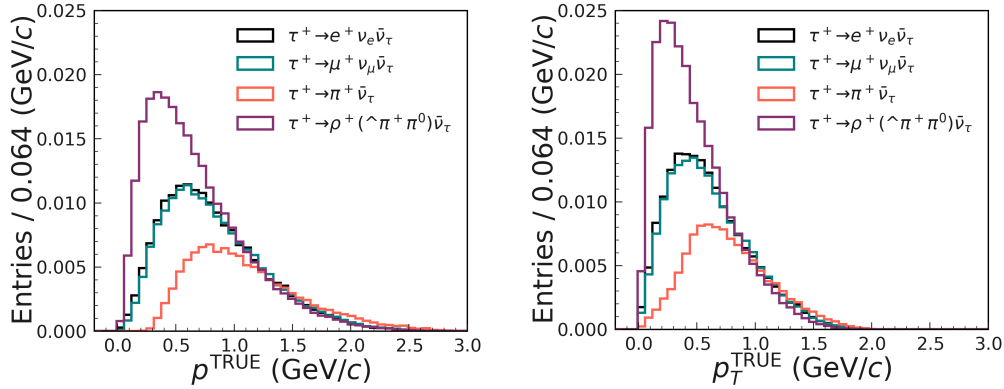


Figure 4.12: True p and p_T of signal τ charged daughters ($e^\pm$, $\mu^\pm$, $\pi^\pm$ prompt and $\pi^\pm$ from $\rho^\pm$) before any selection (Belle MC).

Table 4.6: Reconstructed composition of $\tau\bar{\tau}$ pair in the event for the Belle (left) and Belle II (right) dataset.

$t_u t_c$	ℓ	ρ	h	ℓ	ρ	h
ℓ	0.0707	0.0314	0.1636	0.0970	0.0559	0.1504
ρ	0.0307	0.0131	0.0753	0.0543	0.0249	0.1190
h	0.1659	0.0818	0.3675	0.1458	0.1229	0.2297

Table 4.7: Fraction of signal events in different logical categories, according to the number of wrong particles in the event (Belle II signal MC). Each group corresponds to a reconstructed category; respectively $\ell\ell$, ℓ_ch_u , ℓ_uh_c , $\rho\ell$, $\ell 0$ categories. See text for more details.

$B_{\text{tag}} \backslash B_{\text{sig}}$	TM	Wrong K_S^0	Wrong $t_{c/u}$	Wrong $K_S^0, t_{c/u}$	Wrong t_u, t_c	FM	Total	
TM	0.59	0.02	0.13	0.01	0.01	0.00	0.75	$\ell\ell$
FM	0.18	0.01	0.05	0.00	0.01	0.00	0.25	
Total	0.76	0.03	0.18	0.01	0.01	0.00	1.00	
TM	0.61	0.02	0.06	0.00	0.00	0.00	0.70	ℓ_ch_u
FM	0.21	0.01	0.06	0.01	0.01	0.00	0.30	
Total	0.83	0.03	0.12	0.01	0.01	0.00	1.00	
TM	0.60	0.02	0.06	0.00	0.00	0.00	0.69	ℓ_uh_c
FM	0.22	0.01	0.06	0.00	0.01	0.00	0.31	
Total	0.82	0.04	0.13	0.01	0.01	0.00	1.00	
TM	0.39	0.02	0.24	0.02	0.03	0.00	0.70	$\rho\ell$
FM	0.14	0.01	0.12	0.01	0.02	0.00	0.30	
Total	0.52	0.02	0.36	0.03	0.05	0.01	1.00	
TM	0.53	0.03	0.00	0.00	0.06	0.01	0.63	$\ell 0$
FM	0.25	0.02	0.00	0.00	0.09	0.01	0.37	
Total	0.79	0.05	0.00	0.00	0.15	0.02	1.00	

of mis-reconstruction comes from the B_{tag} but there is also a non-negligible amount of candidates where the $t_{c,u}$ are not correctly assigned. On the other hand, the fraction of candidates having a false-matched K_S^0 is almost negligible, thanks to the `goodBelleKshort` cut. The reconstructed lepton object is considered correct, if it is a truth-matched lepton that comes from τ . The reconstructed ρ object is truth-matched by requiring that its π^0 is correctly reconstructed and that it comes from τ ; additional true π^0 's are allowed (therefore, if the ρ is matched, but the τ has decayed via the $\pi^+\pi^0\pi^0$ channel, we still consider it as correct even though the second π^0 is lost). The h object is truth-matched as a true track coming from the τ , so it can also be a partially reconstructed ρ . Given a much looser requirement on h , compared to ρ , the fractions of wrong track logical conditions differ a lot between lepton-hadron and lepton- ρ categories. It can be seen that the higher the number of leptons there is in the final state, the purer are both the tag-side and the signal-side reconstructions. The exception is the $\rho\ell$ category, which has a significant fraction of wrong tracks due to a strict condition on what is a true ρ , but it is still expected to be a “purer” object than the h . The tables with internal “true” composition for each category, together with the table that summarizes the goodness of reconstruction are shown in Appendix F for the Belle sample.

After all selections, categorization and best candidate selection, the signal efficiencies and the expected background yields (normalized to 100/fb) are shown in the Table 4.8 (Belle II) and Table 4.9 (Belle) per each category. As expected, the $\ell\ell$ category is the purest, and ℓ_uh_c has more $B\bar{B}$ background compared to ℓ_ch_u due to $B \rightarrow X\ell\nu$ events. The $\rho\ell$ has more $q\bar{q}$ background (in Belle II) coming from random combinations and $\ell 0$ has the most $q\bar{q}$ background – at the same level of $B\bar{B}$. The

differences in the yields between the Belle and Belle II samples are coming from multiple effects: lower overall lepton reconstruction efficiency in Belle, lower K_s^0 reconstruction efficiency in Belle, lower π^0 reconstruction efficiency for the ρ^+ in Belle etc. In general, a lower reconstruction efficiency is observed, leading to decreased both signal and background yields. Only the lepton-hadron modes have comparable yields, induced by a higher $\rho \rightarrow h$ leak in Belle, while also suffering only from lower efficiency of a single lepton, as opposed to the $\ell\ell$ category. Comparison of yields within the symmetric $\ell 0$ and $\ell\ell$ categories gives the most fair representation of the reconstruction efficiency differences in Belle and Belle II, given the similar selection for the two datasets. However, the lower lepton-ID efficiency in Belle leads to larger $\ell \rightarrow h$ migration. Similarly, the difference in the π^0 reconstruction efficiency reduces the ρ^+ reconstruction efficiency in Belle, but the $\rho \rightarrow h$ leak makes all the ρ^+ mesons that would fall into the ρh subcategory be added into the hh category, which makes these events stay in the same “leptonless” $\ell 0$ category.

It is worth mentioning that the quoted numbers are not corrected for any data/MC discrepancies and do not perfectly reflect the final working point. Such numbers will be presented in Section 8.

Table 4.8: Signal efficiencies and background yields per 100fb^{-1} for the 5 categories after the initial selection and the best candidate selection. The efficiency is computed as the number of reconstructed events within a given category divided by the total number of generated signal MC events. (Belle II MC).

Category	Signal Efficiency (10^{-4})	Backgrounds/ 100 fb^{-1}		
		$B\bar{B}$	$q\bar{q}$	total
$\ell\ell$	1.30 ± 0.02	1251 ± 9	53 ± 2	1304 ± 10
$\ell_c h_u$	2.01 ± 0.02	1673 ± 11	364 ± 5	2037 ± 12
$\ell_u h_c$	1.95 ± 0.02	2484 ± 13	342 ± 5	2826 ± 14
$\rho\ell$	1.31 ± 0.01	1723 ± 11	458 ± 6	2181 ± 12
$\ell 0$	5.41 ± 0.03	8492 ± 24	7800 ± 23	16293 ± 34

Table 4.9: Signal efficiencies and background yields per 100fb^{-1} for the 5 categories after the initial selection and the best candidate selection. The efficiency is computed as the number of reconstructed events within a given category divided by the total number of generated signal MC events. (Belle MC).

Category	Signal Efficiency (10^{-4})	Backgrounds/ 100 fb^{-1}		
		$B\bar{B}$	$q\bar{q}$	total
$\ell\ell$	0.69 ± 0.01	559 ± 6	14 ± 1	573 ± 6
$\ell_c h_u$	1.58 ± 0.02	1178 ± 9	139 ± 3	1317 ± 10
$\ell_u h_c$	1.60 ± 0.01	2324 ± 13	144 ± 3	2468 ± 13
$\rho\ell$	0.60 ± 0.01	838 ± 8	134 ± 3	972 ± 8
$\ell 0$	5.28 ± 0.02	9258 ± 26	8379 ± 24	17628 ± 35

5. Calibrations

Before the background-suppression and signal-extraction steps, simulated MC events are corrected for known differences with respect to collision, *real* data. These corrections affect both the signal efficiency and the predicted background composition. The largest effects arise from the hadronic B -tagging efficiency, while smaller corrections are applied for charged-track reconstruction, K_S^0 reconstruction, lepton identification, and other effects.

In addition to efficiency corrections, the normalizations of the dominant background components are constrained using data control samples. Continuum $q\bar{q}$ backgrounds are calibrated with off-resonance data, while the normalization and uncertainty of the combinatorial $B\bar{B}$ component are studied in dedicated sidebands.

This chapter describes the origin and impact of these corrections. The associated systematic uncertainties are introduced here and propagated to the signal-extraction model described in Chapter 8.

5.1 Corrections to the simulation

5.1.1 B -tagging efficiency corrections

The tagging and tag-side efficiency of FEI can differ in data with respect to MC for different reasons; the dominant one being that the modeling of B -decays heavily depends on PYTHIA and is often based on old measurements. Moreover, as the Belle II decay file developed independently from Belle's, the data/MC differences measured for the efficiencies of FEI-reconstructed B_{tag} decay channels are different for the two experiments. The FEI calibration is performed using two control samples based on $B \rightarrow D\pi$ or $B \rightarrow X\ell\nu$ events. In the first case, the B_{tag} is reconstructed using FEI and the signal pion is selected; the B_{tag} mode-dependent yields are extracted as fits to the recoil mass of the $B_{\text{tag}}\pi$ system. For the $B \rightarrow X\ell\nu$ sample, the B_{tag} is analogously reconstructed using FEI and the signal lepton is selected; the B_{tag} mode-dependent yields are extracted as fits to the lepton momentum. The second approach has higher statistics due to the larger branching fractions but suffers from the systematic uncertainty related to the fitting procedure and poorly known $B \rightarrow D^{**}\ell\nu, Dn\pi\ell\nu$ components. In both approaches the weights are provided as functions of the B_{tag} final state for the main decay modes (mainly $D^{(*)}n\pi$) while the rest of the modes are combined due to the lower statistics. The treatment of FEI corrections differs between Belle and Belle II:

Belle Only the calibration based on $B \rightarrow D\pi$ decays is available for the considered selection (B^0 tags with $\mathcal{P}_{\text{FEI}} > 0.01$).

Belle II The correction factors are obtained from a χ^2 combination of three independent calibrations: one from $B \rightarrow D\pi$ and two from $B \rightarrow X\ell\nu$ (separately for $\ell = e$ and $\ell = \mu$).

The Uncertainties on the calibration factors are evaluated using the full covariance matrix provided by the calibration procedure. Variations are generated according to a multidimensional normal distribution, and the resulting spread is used to estimate the combined statistical and systematic uncertainty.

In simulation, the `isSignal` flag allows one to identify perfectly reconstructed B_{tag} candidates. However, candidates that do not satisfy this condition still contain a non-negligible fraction of correctly reconstructed decays, as observed in control samples such as $B \rightarrow D\pi$. To account for this, we introduce a refined classification of B_{tag} candidates based on their reconstruction quality.

Candidates are considered mis-reconstructed (combinatorial) if they include one or more particles that are inconsistent with originating from the same B meson. This is determined using information on the origin of the reconstructed particles in simulation. All remaining candidates, which are largely consistent with a correctly reconstructed decay, are classified as ‘peaking’.

The effect of this refined definition is illustrated in Fig. 5.1. The top row shows the classification obtained with the default `isSignal` flag, while the bottom row shows the custom classification introduced here. The comparison is shown both in the beam-constrained mass M_{bc} and in the recoil-mass distribution used in the $B \rightarrow D\pi$ control sample.

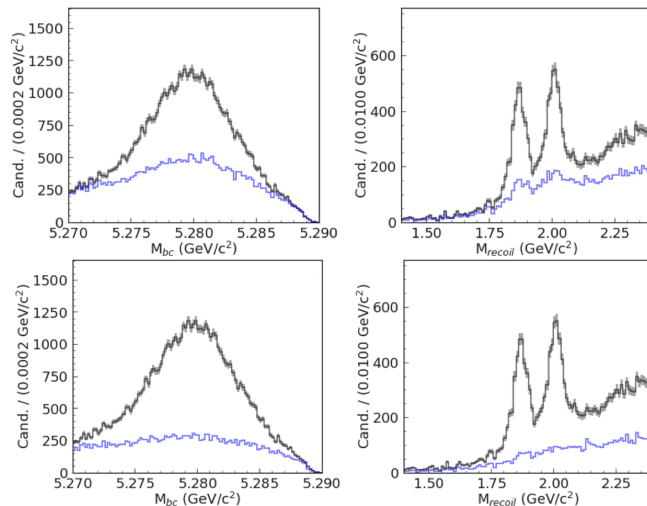


Figure 5.1: Comparison between the default `isSignal`-based classification and the custom classification of reconstructed B_{tag} candidates in simulation. The top row uses the default flag, while the bottom row uses the custom definition introduced in the text. The left column shows the beam-constrained mass M_{bc} , and the right column shows the recoil-mass distribution in the $B \rightarrow D\pi$ control sample. ‘Peaking’ candidates are shown in black and ‘combinatorial’ candidates in blue.

The mean correction factors applied to the peaking tag-side B^0 mesons in $B\bar{B}$ background events are 0.768 ± 0.052 for Belle and 0.786 ± 0.050 for Belle II, which are the largest in both datasets.

5.1.2 Tracking efficiency calibration

Corrections to the charged-track reconstruction efficiency are applied to all reconstructed tracks, including those originating from τ decays and prompt K_s^0 daughters.

For momenta above 200 MeV/c:

Belle: Corrections are evaluated in Belle using a high-purity sample of $D^{*+} \rightarrow D^0 \pi^+$, with $D^0 \rightarrow K_s^0 \pi^+ \pi^-$. The pion from the K_s^0 decay serves as the probe track, while the event is kinematically constrained using the remaining two pions from the D^0 and the pion from the D^{*+} decay.

Belle II: The efficiency is determined using $e^+e^- \rightarrow \tau^+\tau^-$ events, where $\tau^+ \rightarrow \ell^+ \nu_\ell \bar{\nu}_\tau$ and $\tau^- \rightarrow \pi^- \pi^- \pi^+ \nu_\tau$, with one track from the hadronic τ decay used as the probe.

In both Belle and Belle II, no efficiency corrections are needed and per-track systematic uncertainties of 0.35% and 0.27% are assigned, respectively.

For momenta below 200 MeV/c: Momentum-binned corrections from the $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0 \pi^-) \pi^+$ calibration samples are applied, as the pion π^- emitted by the D^{*-} allows to probe the low-momentum regime for both datasets.

The mean of the correction applied and the mean of the associated systematic uncertainty in simulated background is 0.999 ± 0.004 for Belle and 0.998 ± 0.003 for Belle II.

5.1.3 K_s^0 efficiency calibration

Due to the displaced decay topology of K_s^0 mesons, their reconstruction efficiency requires dedicated calibration. Data–MC efficiency differences in the K_s^0 vertex reconstruction are measured using the $D^{*+} \rightarrow D^0(\rightarrow K_s^0 \pi^+ \pi^-) \pi^+$ calibration channel.

Belle The efficiency corrections are determined as functions of the K_s^0 momentum and $\cos \theta$, averaged over different detector configurations.

Belle II The corrections are provided in bins of momentum, $\cos \theta$, and the three-dimensional decay vertex displacement.

For Belle II, it is verified that the corrections are stable under the application of additional quality requirements on the K_s^0 candidates. The same assumption is adopted for Belle. Due to negligible contribution of incorrectly identified K_s^0 candidates, every object reconstructed as K_s^0 is efficiency-corrected.

The average correction factors are 0.985 ± 0.006 for Belle and 0.989 ± 0.021 for Belle II.

5.1.4 π^0 efficiency calibration

Corrections to the π^0 reconstruction efficiency are applied to $\tau \rightarrow \rho \nu$ decays where the π^0 candidate is correctly reconstructed and matched to the true particle.

Belle The corrections are provided as functions of the π^0 momentum, derived from studies of $\tau \rightarrow \pi \pi^0 \nu$ decays.

Belle II The corrections are given as functions of $(p, \cos \theta)$ and are derived from control samples of D -meson decays.

The mean correction factors are 0.948 ± 0.022 for Belle and 1.010 ± 0.041 for Belle II.

5.1.5 Lepton identification calibration

For $\tau \rightarrow \ell$ decays, corrections are applied to account for differences between data and simulation in lepton identification efficiencies and hadron misidentification rates. The contribution from $\pi \rightarrow e$ misidentification is not considered due to its negligible rate in the selected sample. The corrections to efficiencies and fake rates are applied based on true particle identity.

The lepton identification efficiencies are determined with $e^+e^- \rightarrow \ell^+\ell^-\gamma$ (Belle II only), $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$, and $J/\psi \rightarrow \ell^+\ell^-$ events. The $\pi \rightarrow \mu$ misidentification rates are obtained from $K_s^0 \rightarrow \pi^+\pi^-$ decays and, for Belle II, supplemented by $e^+e^- \rightarrow \tau^+(\rightarrow \pi^+\pi^+\pi^-\bar{\nu}_\tau)\tau^-$ events. In Belle (Belle II), the mean systematic uncertainty related to electron identification efficiency is 2.6% (1.9%). The efficiency of muon identification has a systematic uncertainty of 2.1% (0.4%). The systematic uncertainty associated with incorrect identification of pions as muons is 4.2% (5.4%).

5.1.6 Correction of branching fractions of signal-side τ leptons

The branching fractions of τ decay modes in our simulations have small differences with respect to the latest PDG averages, especially for the Belle MC. We fix the branching fractions of the dominant τ decays in both Belle and Belle II signal MC samples to the current best experimental knowledge. The correction are overall small – one to few percent for the most abundant τ decay modes and are applied only to signal MC, as it directly impacts signal efficiency. For each τ decay mode the uncertainty from PDG is assigned as a systematic uncertainty.

5.2 Data-MC comparison in background enriched samples

Studies in background enriched samples (sidebands) are performed to check the goodness of the corrections applied to MC to better reflect data. Furthermore, these ad-hoc samples can also be used to measure possible additional data/MC mismatches due to simulation issues. For example, it is known that $q\bar{q}$ samples suffer from some normalization issues and that specific background components are mismodelled (inclusive productions, spectra etc.). We carry out the studies mentioned above mainly using three sideband samples:

1. E_{cms} sideband. With off-resonance data, we can check the data/MC agreement obtained with the full set of cuts in a $B\bar{B}$ -free sample.
2. K_s^0 sideband. With a selection on the K_s^0 candidate that excludes good, displaced V0's, one can check in on-resonance data not only the $B\bar{B}$ normalization but also the BDT training variables for each fitting category.
3. M_{bc} sideband. Despite the lower statistics compared to the K_s^0 sideband, this sample provides a direct access to the combinatorial B_{tag} component in events with true K_s^0 on the signal side.

Other sidebands were tried but they were considered as less appropriate. Below a short summary of the attempts made and the main reason(s) why each sample was discarded.

$\mathcal{P}_{\text{FEI}}$ sideband. The sample is obtained by picking the B_{tag} candidate with $0.001 < \mathcal{P}_{\text{FEI}} < 0.01$. With this selection the B_{tag} 's are mostly wrongly reconstructed. So it could have been used to derive combinatorial $B\bar{B}$ scaling factor. However, the final state composition of such B_{tag} 's is very different from our signal.

“Charge” sideband. $B^0 \rightarrow K_S^0 \tau^+ \tau^+$ decays are reconstructed, allowing for charge violation. This sample, despite being convenient as all the corrections can be applied, shows enhancements in fake tracks for the τ daughters.

q^2 sideband. This sample can in principle be used for the $B\bar{B}$ normalization but it has lower statistics compared to the K_S^0 sidebands (see Fig. 4.8). Moreover, with the same sample one cannot check the BDT training features, being the q^2 one of them.

5.2.1 Off-resonance

When selecting the B_{tag} 's, one must consider that the M_{bc} values get shifted towards lower values due to the shift in the CM energy, E_{cm} , of approximately 60 MeV. Hence, for the off-resonance samples, the selection reads $M_{\text{bc}}^{\text{corr}} > 5.27 \text{ GeV}/c^2$, where the ‘corrected’ M_{bc} is defined as in Eq. 5.1.

$$M_{\text{bc}}^{\text{corr}} = \sqrt{\left(\frac{E_{\text{cm}} + 0.06}{2}\right)^2 / c^4 - \left(\left(\frac{E_{\text{cm}} + 0.06}{E_{\text{cm}}}\right) p^* / c\right)^2} \quad (5.1)$$

The effect is less pronounced on the ΔE variable, which also gets corrected with the same E_{cm} shift. It is important to stress that the M_{bc} correction relies on the precise knowledge of the E_{cm} .

Due to less-known effects in the hadronization of produced $q\bar{q}$ pair, significant mis-modelings can be present in simulation. For this work we notice that simple scaling factors (SF) are sufficient to obtain a good agreement. They are computed as:

$$SF_{q\bar{q}} = \frac{N_{\text{data}}}{N_{\text{MC}}} \quad (5.2)$$

and are evaluated for the $\ell X \equiv \ell\ell \cup \ell_u h_c \cup \ell_c h_u \cup \rho\ell$ and $\ell 0$ categories separately. Here we merge all the categories with at least one lepton into ℓX because of the low statistics (especially in $\ell\ell$). The values of $SF_{q\bar{q}}$ for Belle and Belle II and their statistical errors are shown in Tab. 5.2.

The $M_{\text{bc}}^{\text{corr}}$ distributions for Belle and Belle II off-resonance samples are shown in Fig. 5.2; MC shows good agreement once simple scaling factors are applied. This holds true also for other dimensions like $M(K_S^0)$ or E_{ECL} . Fig. 5.3 shows that the data/MC yield discrepancies can be related to an excess of K_S^0 in off-resonance MC; as a proof of that, such discrepancy disappears for the K_S^0 -depleted and $B\bar{B}$ -enhanced samples.

5.2.2 K_S^0 sideband

While for well-reconstructed B^0 tags official data/MC efficiency corrections exist, for the cross-feed component, made of wrongly-reconstructed tags in B^0 or B^+ events,

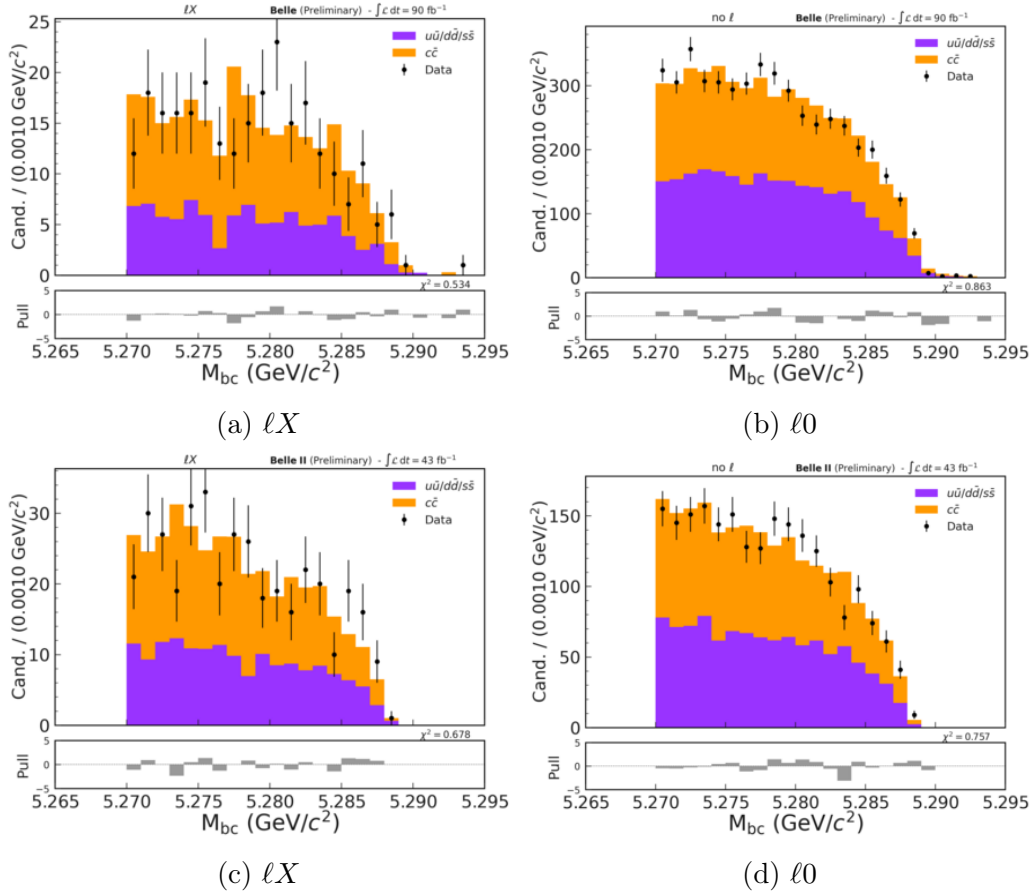


Figure 5.2: M_{bc}^{corr} for Belle (upper) and Belle II (lower) off-resonance samples – ℓX category (left) and $\ell 0$ category (right). The scaling factors defined in Eq. 5.2 are applied to MC events.

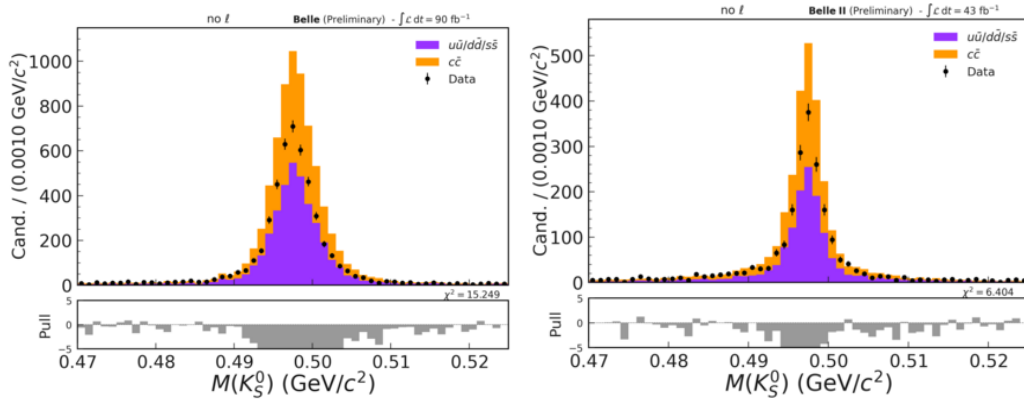


Figure 5.3: $M(K_s^0)$ for Belle (left) and Belle II (right) off-resonance samples, $\ell 0$ categories. No scaling factors are applied.

possible mis-modelings must be addressed using control samples. For this check, we choose the K_s^0 sideband, defined using the selection criteria in Table 5.1.

As shown in Fig. 5.4, true K_s^0 's that have the `goodBelleKshort==0` flag (called bBKS, for ‘bad’ Belle K_s^0) are more likely to be formed by combining two oppositely charged pions from the IP. For those passing the `goodBelleKshort==1` flag (called gBKS, for ‘good’ Belle K_s^0), the trend is opposite and they are mostly displaced

5.2. Data-MC comparison in background enriched samples

Table 5.1: Definition of K_s^0 sideband.

Object	Selection
	$M(\pi^+\pi^-) \in \{(0.480, 0.486) \cup (0.510, 0.516)\} \text{ GeV}/c^2$
	successful vertex fit
	goodBelleKshort==0 (inverted quality criteria)

vertices originating from particle with zero charge(V0). As we want our sideband to be approximately “signal-free” and the bBKS still contains true K_s^0 ’s, we add the invariant mass requirement to obtain two symmetric bands with same width. The width was chosen in such a way that the obtained sample had enough events, and that the fake K_s^0 ’s mass would not too far compared to the true value.

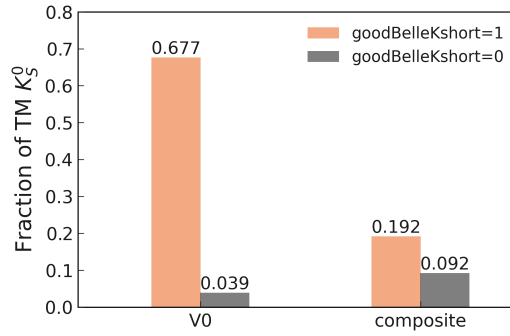


Figure 5.4: Composition of signal events in terms of K_s^0 reconstruction in Belle II. Most of gBKS are reconstructed by the **V0Finder** while for the bBKS the trend is opposite (more are reconstructed as combinations of prompt pions).

From Fig. 5.5 with the M_{bc} and decay modes for the mis-reconstructed B_{tag} ’s, one can see that the composition of “unmatched” B_{tag} ’s in the K_s^0 sideband is close enough to our background.

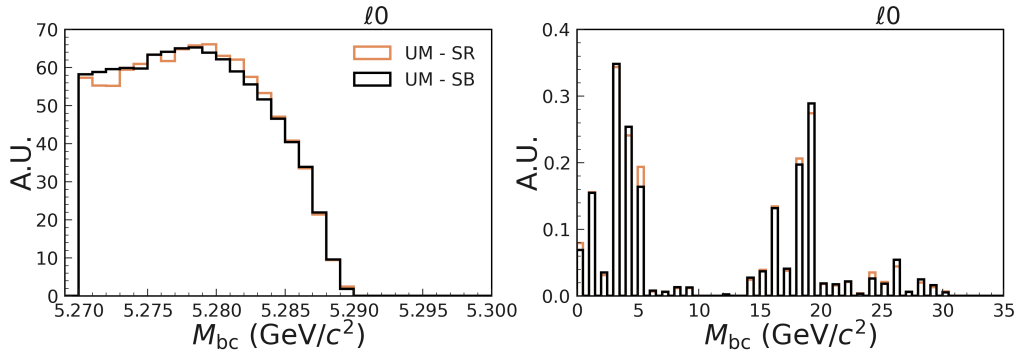


Figure 5.5: M_{bc} and FEI mode distributions for the *unmatched* (UM) B_{tag} ’s in $B\bar{B}$ events in the K_s^0 sideband (SB), compared to signal region (SR). Plots refer to Belle II MC. The distributions refer to the $\ell 0$ category, taken as example.

The scaling factor SF for the combinatorial $B_{tag} BB$ component is obtained with the formula in Eq. 5.3:

$$SF_{B\bar{B}}^{comb} = \frac{N_{\text{data}} - N_{q\bar{q}} - N_{B\bar{B}}^{\text{TM}}}{N_{B\bar{B}}^{comb}}, \quad (5.3)$$

where N_{data} is the number of events in on-resonance data, $N_{q\bar{q}}$ is the MC yield of $q\bar{q}$ events corrected using off-resonance data (in the K_s^0 sideband), $N_{B\bar{B}}^{\text{TM}}$ is the corrected yield of truth-matched $B\bar{B}$ events and finally $N_{B\bar{B}}^{comb}$ is the MC yield of unmatched $B\bar{B}$ events. A visual representation of how the combinatorial B component is calibrated can be seen in Fig. 5.6, where the M_{bc} for false-matched events is compared to the data subtracted.

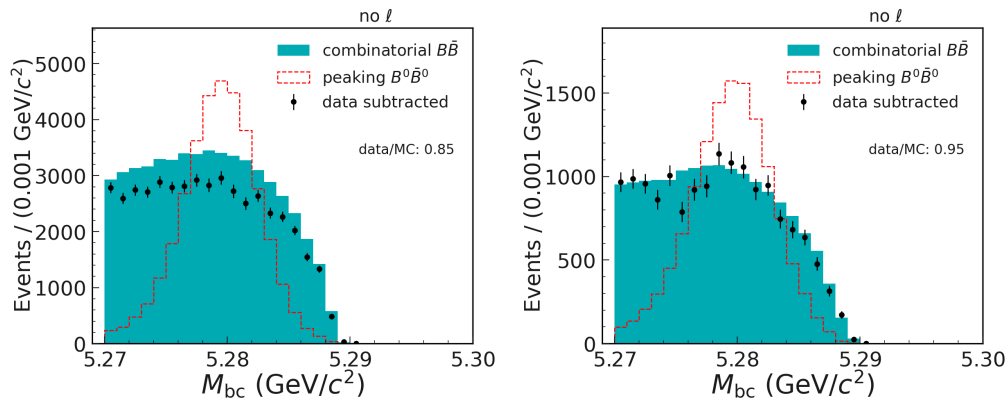


Figure 5.6: M_{bc} for Belle (left) and Belle II (right) in K_s^0 sideband (on-resonance) samples, $\ell 0$ categories. The ratios between the data subtracted yields and the MC ‘combinatorial’ yields provide the $SF_{B\bar{B}}^{comb}$.

As for the $q\bar{q}$ scaling factors, those for the combinatorial $B\bar{B}$ events are also listed in Tab. 5.2 for Belle and Belle II MC.

In off-resonance K_s^0 SB, we notice a better data/MC agreement even without any correction (i.e. scaling factors are closer to one). The computed factors for off- and on-resonance K_s^0 SB samples are listed in Tab. 5.2.

5.2.3 M_{bc} sideband

As an additional handle for background validation, we explore the M_{bc} sideband, defined by the cut $M_{bc} \in [5.24, 5.27) \text{ GeV}/c^2$. We obtain this sample by just shifting the M_{bc} requirement to $M_{bc} > 5.24 \text{ GeV}/c^2$, making sure, when we check the outputs, that the signal region is kept veiled. As done for the signal region, we apply a best-candidate selection based on the highest $\mathcal{P}_{\text{FEI}}$. As can be seen from the M_{bc} distributions, the BCS has some sculpting effect (positive slope) at lower M_{bc} values in Belle.

To derive the $SF_{B\bar{B}}^{comb}$ for this sideband, we proceed analogously to the K_s^0 sideband: first we derive normalization factors from the off-resonance data (Fig. 5.7), which we then apply to the $q\bar{q}$ of the on-resonance data. Finally, with the computation of Eq. 5.3, we obtain the scaling factors for the combinatorial B_{tag} component. It should be stressed that no $B\bar{B}$ background is taken as ‘peaking’ in this sideband. The obtained correction factors are listed in Tab. 5.2. From Fig. 5.8 one can see that simple normalization factor gives already a good shape agreement, although some slight shape mismatch can be seen in Belle, possibly due to the sculpting mentioned earlier.

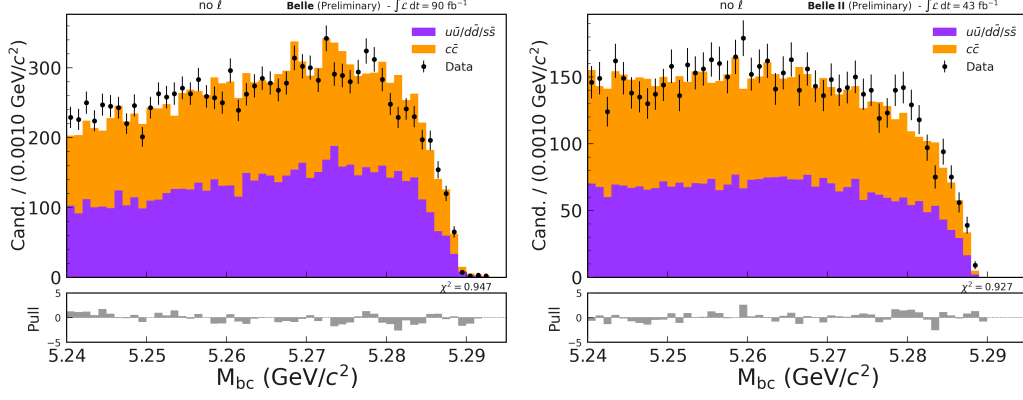


Figure 5.7: M_{bc} for Belle (left) and Belle II (right) off-resonance samples, $\ell 0$ categories. Scaling factors are applied.

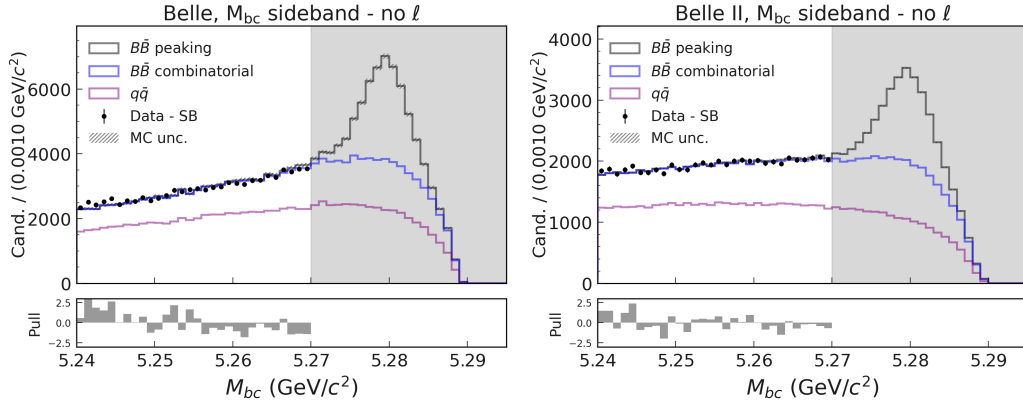


Figure 5.8: M_{bc} for Belle (left) and Belle II (right) on-resonance samples, $\ell 0$ categories. The signal region is veiled and both $q\bar{q}$ and $B\bar{B}$ scaling factors are applied.

5.3 Summary

Tab. 5.3 (Belle II) and Tab. 5.4 (Belle) show the final efficiencies and background yields with all corrections applied, to be compared to the ‘uncorrected’ ones. As will be presented in Section 6, the corrected samples are then used for the background suppression, as they provide a more realistic description of data.

Table 5.2: Data/MC scaling factors for Belle and Belle II.

SF	$\ell\ell$	ℓ_ch_u	ℓ_uh_c	$\rho\ell$	$\ell 0$	
$q\bar{q}$ (OFF res)	0.74 ± 0.05				0.659 ± 0.011	Belle
$q\bar{q}$ (OFF res, K_s^0 SB)	0.98 ± 0.05				0.747 ± 0.009	
$B\bar{B}$ (ON res, K_s^0 SB)	1.16 ± 0.13	0.83 ± 0.05	0.76 ± 0.03	0.75 ± 0.05	0.851 ± 0.018	
$q\bar{q}$ (OFF res, M_{bc} SB)	0.87 ± 0.06				0.703 ± 0.010	
$B\bar{B}$ (ON res, M_{bc} SB)	0.82 ± 0.09	0.84 ± 0.05	0.76 ± 0.03	0.84 ± 0.06	0.82 ± 0.03	
$q\bar{q}$ (OFF res)	0.79 ± 0.04				0.718 ± 0.017	Belle II
$q\bar{q}$ (OFF res, K_s^0 SB)	0.96 ± 0.05				0.989 ± 0.020	
$B\bar{B}$ (ON res, K_s^0 SB)	0.82 ± 0.07	0.85 ± 0.05	0.95 ± 0.04	0.93 ± 0.06	0.93 ± 0.03	
$q\bar{q}$ (OFF res, M_{bc} SB)	0.701 ± 0.027				0.692 ± 0.011	
$B\bar{B}$ (ON res, M_{bc} SB)	0.78 ± 0.06	0.86 ± 0.05	0.93 ± 0.04	0.94 ± 0.05	1.08 ± 0.04	

Table 5.3: Signal efficiencies and expected background yields for Belle II MC. In the corrected columns the reconstruction efficiencies, fake rates and normalization are corrected for data/MC discrepancies. Additionally, the approximate overall weight is shown. Uncertainties shown are coming from MC statistics only.

	Signal eff. (10^{-4})			$q\bar{q}$ yields (100 fb^{-1})			$B\bar{B}$ yields (100 fb^{-1})		
	uncorr	corr	weight	uncorr	corr	weight	uncorr	corr	weight
$\ell\ell$	1.30 ± 0.02	0.97 ± 0.01	0.74	53 ± 2	44 ± 2	0.83	1251 ± 9	925 ± 7	0.74
ℓ_ch_u	2.01 ± 0.02	1.54 ± 0.01	0.77	364 ± 5	293 ± 4	0.80	1673 ± 11	1308 ± 9	0.78
ℓ_uh_c	1.95 ± 0.02	1.48 ± 0.01	0.76	342 ± 5	273 ± 4	0.80	2484 ± 13	2044 ± 11	0.82
$\rho\ell$	1.31 ± 0.01	1.05 ± 0.01	0.80	458 ± 6	376 ± 5	0.82	1723 ± 11	1443 ± 9	0.84
$\ell 0$	5.41 ± 0.03	4.32 ± 0.02	0.80	7800 ± 23	5542 ± 16	0.71	8492 ± 24	7324 ± 21	0.86

Table 5.4: Signal efficiencies and expected background yields for Belle MC. In the corrected columns the reconstruction efficiencies, fake rates and normalization are corrected for data/MC discrepancies. Additionally, the approximate overall weight is shown. Uncertainties shown are coming from MC statistics only.

	Signal eff. (10^{-4})			$q\bar{q}$ yields (100 fb^{-1})			$B\bar{B}$ yields (100 fb^{-1})		
	uncorr	corr	weight	uncorr	corr	weight	uncorr	corr	weight
$\ell\ell$	0.69 ± 0.01	0.48 ± 0.01	0.70	14 ± 1	8 ± 1	0.61	559 ± 6	433 ± 5	0.77
ℓ_ch_u	1.58 ± 0.02	1.14 ± 0.01	0.72	139 ± 3	90 ± 2	0.65	1178 ± 9	878 ± 7	0.75
ℓ_uh_c	1.60 ± 0.01	1.17 ± 0.01	0.73	144 ± 3	93 ± 2	0.65	2324 ± 13	1681 ± 9	0.72
$\rho\ell$	0.60 ± 0.01	0.44 ± 0.01	0.73	134 ± 3	84 ± 2	0.63	838 ± 8	578 ± 6	0.69
$\ell 0$	5.28 ± 0.02	4.03 ± 0.01	0.76	8379 ± 24	5344 ± 15	0.64	9258 ± 26	7333 ± 21	0.79

6. Background suppression and classification

After the baseline event selection, the selected sample is still dominated by background. The main contributions arise from $B\bar{B}$ decays, especially semileptonic and hadronic B decays involving charm, and from continuum $e^+e^- \rightarrow q\bar{q}$ production. Their relative importance depends strongly on the reconstructed $\tau\tau$ category, motivating a category-dependent background-suppression strategy.

This chapter first characterizes the dominant backgrounds after all selection and calibration steps. Several notable background contributions are identified and, where convenient, removed with dedicated vetoes. The remaining separation between signal and background is performed with multivariate classifiers trained separately for each signal category and dataset. The classifier output is transformed into the variable $\mathcal{O}'$, which is used both to define the signal-enhanced region and as the fit observable in the signal-extraction model.

The final part of the chapter describes the choice of input variables, the classifier training, the transformation of the classifier output, and the resulting expected signal and background distributions.

6.1 Background characterization and vetoes

After the basic selection, the best candidate selection and the corrections to the simulation derived from the off-resonance sample and K_S^0 sideband, the overall background composition in terms of $B\bar{B}$ and $q\bar{q}$ events is given in Tab. 5.4 (5.3) for Belle (II). Generally speaking, the $B\bar{B}$ events tend to dominate: they are $>90\%$ for the $\ell\ell$ category, $\sim 80 - 90\%$ for the 1-lepton categories, and $\sim 50\%$ for the $\ell 0$. On a deeper level, we know that most of the $B\bar{B}$ events are actually $B^0\bar{B}^0$ events, given the quite tight cut on $\mathcal{P}_{\text{FEI}}$ —the fraction of neutral B 's being (75-95%).

Within the $B^0\bar{B}^0$ events, a very large fraction ($\sim 97\%$) have a generated hadronic B_{tag} , while the nature of the companion B_{sig} depends on the category. In particular, the most frequent background is:

- $B \rightarrow D_{\text{sl}}\ell\nu$ decays for the $\ell\ell$ category ($\sim 60\%$);
- $B \rightarrow D_{\text{sl}}X_{\text{had}}$ decays for the $\ell_c h_u$ category ($\sim 40\%$);
- $B \rightarrow D_{\text{had}}\ell\nu$ decays for the $\ell_u h_c$ category ($\sim 60\%$);
- $B \rightarrow D_{\text{had}}\ell\nu$ decays for the $\rho\ell$ category ($\sim 35\%$);
- $B \rightarrow D_{\text{had}}X_{\text{had}}$ decays for the $\ell 0$ category ($\sim 55\%$).

The numbers above refer to Belle II samples but in Belle a similar composition is found.

Given this composition, we expect that our background suppression based on MVA will easily reduce backgrounds with large missing energy but little missing mass ($B \rightarrow D\ell\nu$ decays). Other components with unreconstructed particles and less clear signatures ($B \rightarrow D^{**}\ell\nu$, $B \rightarrow D^{(*)}\tau\nu$, charmless B decays) will populate our final signal region.

Furthermore, there are some backgrounds that can be visualized as peaks in specific invariant mass distributions. For example:

1. $D^- \rightarrow K_S^0\{\pi^-, \rho^-\}$ in $M(K_S^0 t_c)$
2. $\gamma \rightarrow e^+e^-$ in $M(t_c t_u)$
3. $J/\psi \rightarrow \ell^+\ell^-$ in $M(t_c t_u)$
4. $D^0 \rightarrow K_S^0\pi^0$ in $M(K_S^0\pi^0)$

At this stage, we do not explicitly veto those resonances except for the 1. as it is the most abundant. In particular, we apply the cut only for the $\ell_u h_c$, $\ell 0$ and the $\rho\ell$ categories using the cuts $M(K_S^0, t_c) \notin [1.8, 1.9] \text{ GeV}/c^2$. Before the veto is applied, the distributions are shown in Fig 6.1. The vetoed events provide a useful validation sample enriched in $B^0 \rightarrow D^{(*)-}\ell^+\nu_\ell$ events, as will be described in Section 7.2.

To suppress the backgrounds, we rely on a Boosted Decision Tree classifier (BDT). Events like those in 2. and 3. are easily rejected with the multivariate approach.

6.2 BDT training

The MVA (MultiVariate Analysis) package called FastBDT [41] is employed as a tool for classifying signal and background events. Since five different reconstructed categories are created, each with specific background sources, the set of variables used for training are not the same. Training is performed with unbalanced samples of signal and background, where the sample with more entries gets weighted, therefore receiving a higher penalty for the training. The BDT output score $\mathcal{O}$ is transformed into $\mathcal{O}'$ via a Probability Integral Transform [47]. After the transformation, the signal $\mathcal{O}'$ shape is uniform in the $[0.0, 1.0]$ range (see Fig. 6.2 for an example); therefore a cut $\mathcal{O}' > Y$ corresponds to a final efficiency of Y (provided that all correction weights are already applied).

In order to use all the available MC events (2 streams for Belle, 4 streams for Belle II) we split the events into two sets based on their event number – *even* and *odd*. In this way, the weights obtained by training with odd/even subsets can be applied to the full even/odd testing subsets, after having checked that the two classifiers have same performance (as demonstrated in Fig. 6.3 and Fig. 6.4 for the $\ell_c h_u$ category – the others are similarly good). This allows us to exploit the full available MC dataset for the fit template creation to have the smallest possible statistical uncertainty on the fit template, which would come from MC statistics.

As mentioned before, all the data/MC corrections described in Section 5 are applied to the MC distributions prior to the training, in order to provide more realistic distributions, both in terms of shapes and relative composition of the background events.

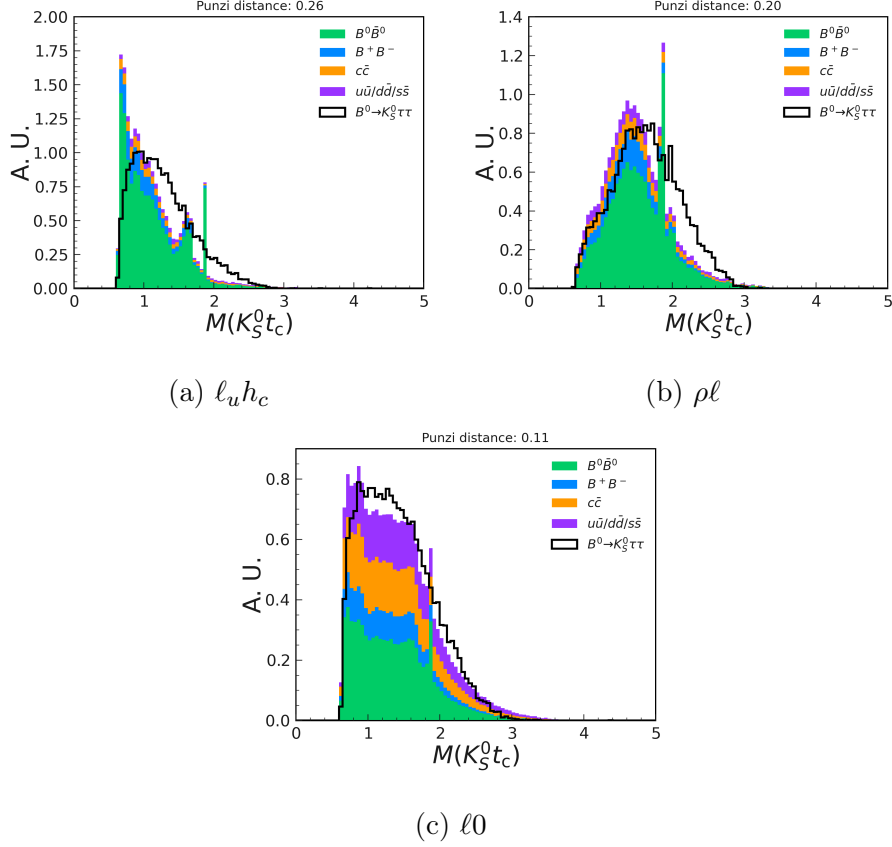


Figure 6.1: $M(K_S^0, t_c)$ distribution for the 3 categories, in which the D veto is applied. For the $\rho \ell$ category, the distribution is not shown specifically for the case when ρ is a t_c . (Belle II MC)

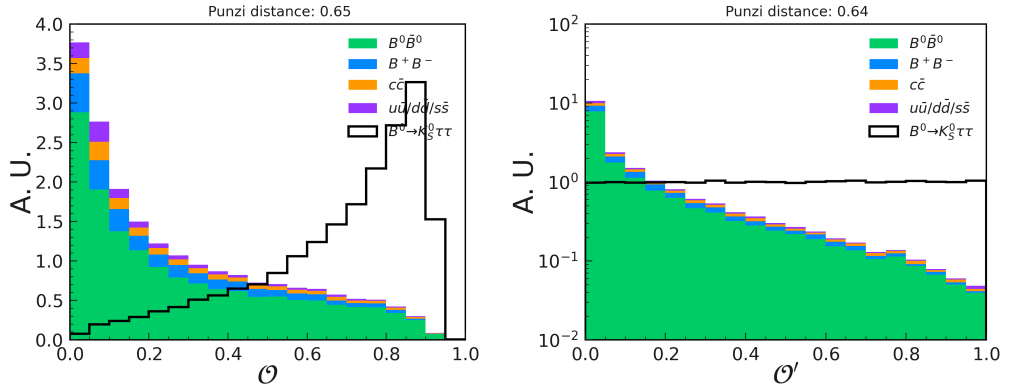


Figure 6.2: Example of the BDT outputs before (left) and after (right) the Probability Integral Transformation for the $\ell_u h_c$ category (Belle II MC)

The BDT classifiers are trained separately for each category and sample (Belle and Belle II). The hyper-parameters used for the training are optimised in such a way that, for a fixed ratio of train-test dataset sizes (equal to 80%) for *odd* and *even* trainings independently, the Area Under Curve (AUC) of the ROC for the training set is maximised, while not letting the AUC of the ROC for the test datasets differ from the training ones by more than 2%. We do not tune all the hyper-parameters

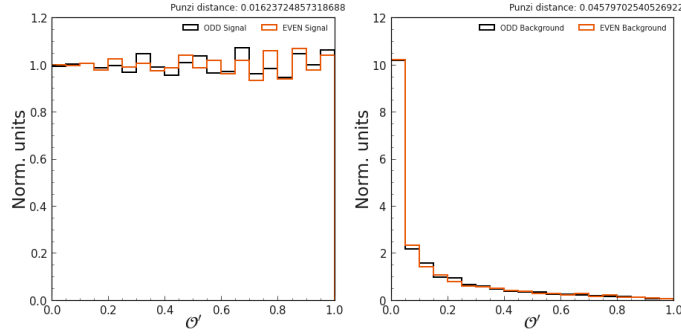


Figure 6.3: Comparison of the BDT response for *even* and *odd* events (Belle II MC, $\ell_c h_u$ category). Signal MC is shown on the left and background events on the right.

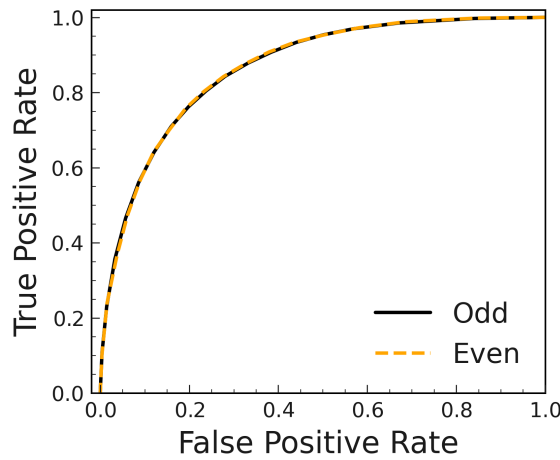


Figure 6.4: Comparison of the ROC (Receiver Operating Characteristic) curves for the training sets for *even* and *odd* samples (Belle II MC, $\rho\ell$ category).

but rather focus on three of them: **depth** (maximum number of consecutive cuts for each decision tree), **nTrees** (the number of decision trees used for the training), **shrink** (the level of learning rate for the method of gradient boosting). The optimization results are shown in Table 6.1. The consistency of the *odd* and *even* trainings is also checked as shown above.

Table 6.1: BDT training hyper-parameters. The values were optimised for each fit category and for Belle and Belle II samples separately.

	Belle			Belle II		
	depth	nTrees	shrink	depth	nTrees	shrink
$\ell\ell$	2	150	0.05	3	160	0.04
$\ell_c h_u$	3	160	0.05	3	160	0.04
$\ell_u h_c$	3	180	0.04	3	180	0.04
$\rho\ell$	3	160	0.035	3	200	0.04
$\ell 0$	4	200	0.04	4	250	0.04

The ROC curves for each category are shown in Fig. 6.5 for the Belle II dataset, *odd* split. Similar performance and ROC curves are obtained for Belle. ROC curves

for the *odd* and *even* splits for Belle and *even* for Belle II are shown in Appendix B.1, Fig. B.14, Fig. B.15, Fig. B.13.

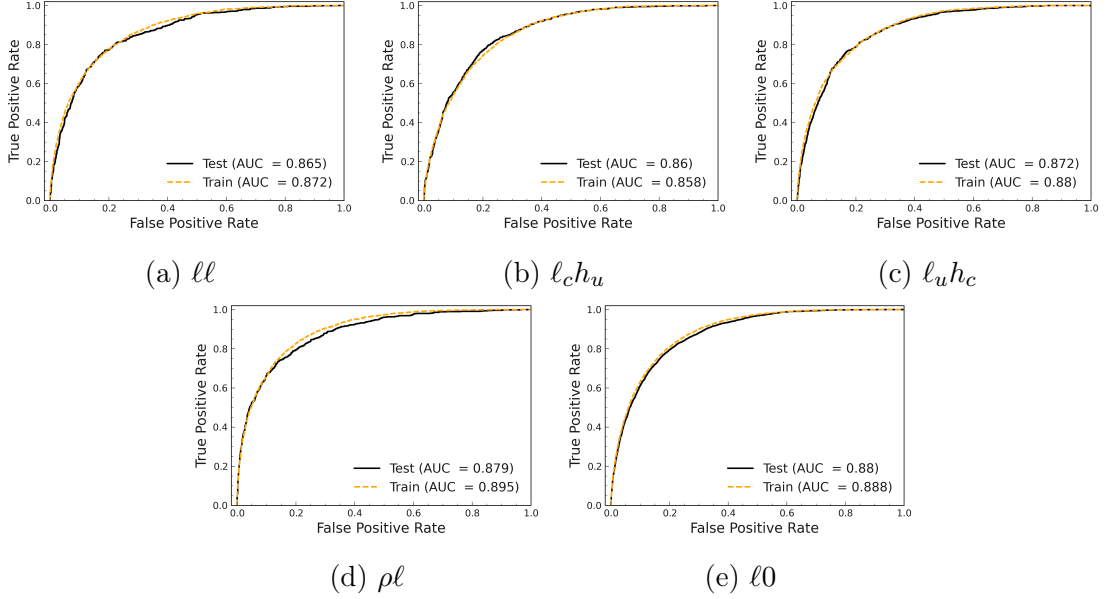


Figure 6.5: ROC curve for test and train sets (Belle II MC), *odd* split. ROC curves for Belle and *even* split for Belle II are in Appendix B.1.

The $\mathcal{O}'$ (transformed BDT score output) distributions are shown in Fig. 6.6 (Belle) and Fig. 6.7 (Belle II). The correlations of the input features to the $\mathcal{O}'$ are shown in Fig. B.16 in Appendix B.1.

The working point for the signal region is chosen at cut of $\mathcal{O}' > 0.3$ for all the categories, which retains exactly 70% signal and rejects $\gtrsim 80\%$ of background events. The choice of the cut to define signal region is not optimized but rather picked to cut out the backgrounds easily rejected by the BDT and retain reasonably high signal efficiency.

6.2.1 BDT features

To quantify the separation power between signal and the backgrounds, we use the distance metric defined in Ref. [48]: $P = \sqrt{f(1-f) \int \frac{(p_{\text{sig}}(x) - p_{\text{bkg}}(x))^2}{p_{\text{tot}}(x)} dx}$. The integral is replaced by a sum over all the bins in which the 1D distributions for signal p_{sig} and background events p_{bkg} are divided. f is the fractional area covered by signal. Given the way P is constructed, perfect overlap between the two distributions would lead to 0 while 100% separation translates to $P = 1$. We pick the features for the training based on their separation power. The features used to train the BDT classifier are introduced and described below.

- $M(K_s^0 t_c)$ The invariant mass of the system composed of the K_s^0 candidate and the *correlated* τ daughter candidate. This variable is used to suppress $B \rightarrow DX$ background. Since the K_s^0 does not carry flavor information, the flavor of the B_{sig} candidate is inferred from the reconstructed flavor of the B_{tag} . This determines which τ daughter is assigned as the correlated one. The assignment is reliable for unmixed $B^0 \bar{B}^0$ events, but is incorrect for mixed events, as illustrated in Fig. 6.8.

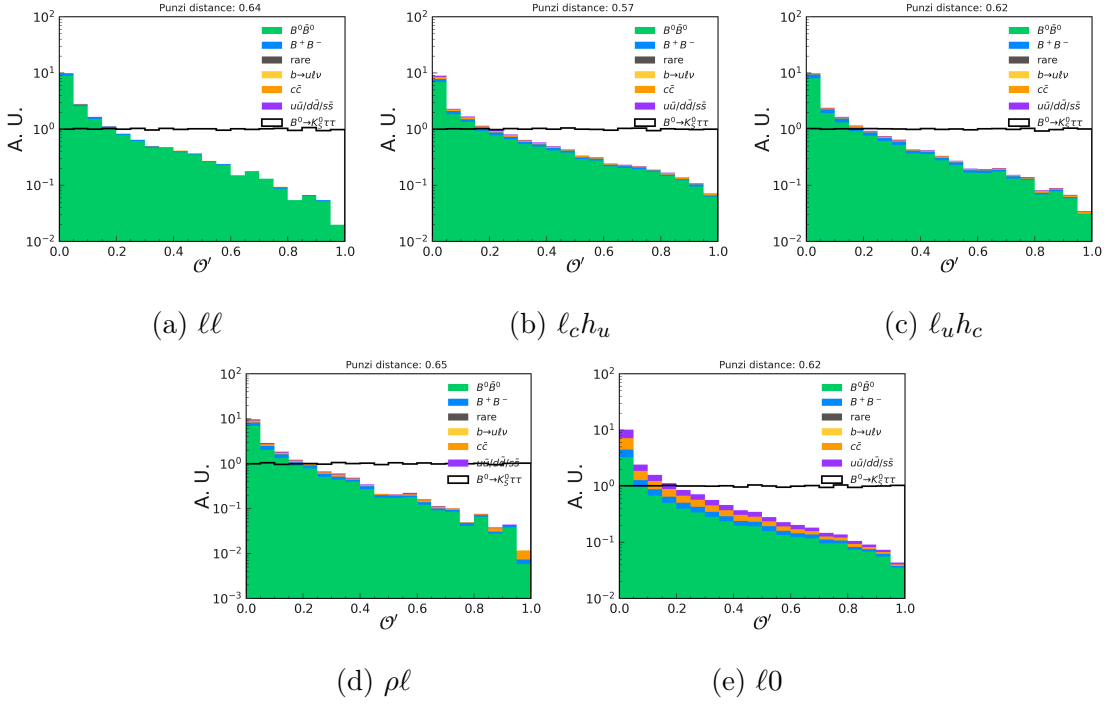


Figure 6.6: BDT outputs after Probability Integral Transformation (Belle MC).

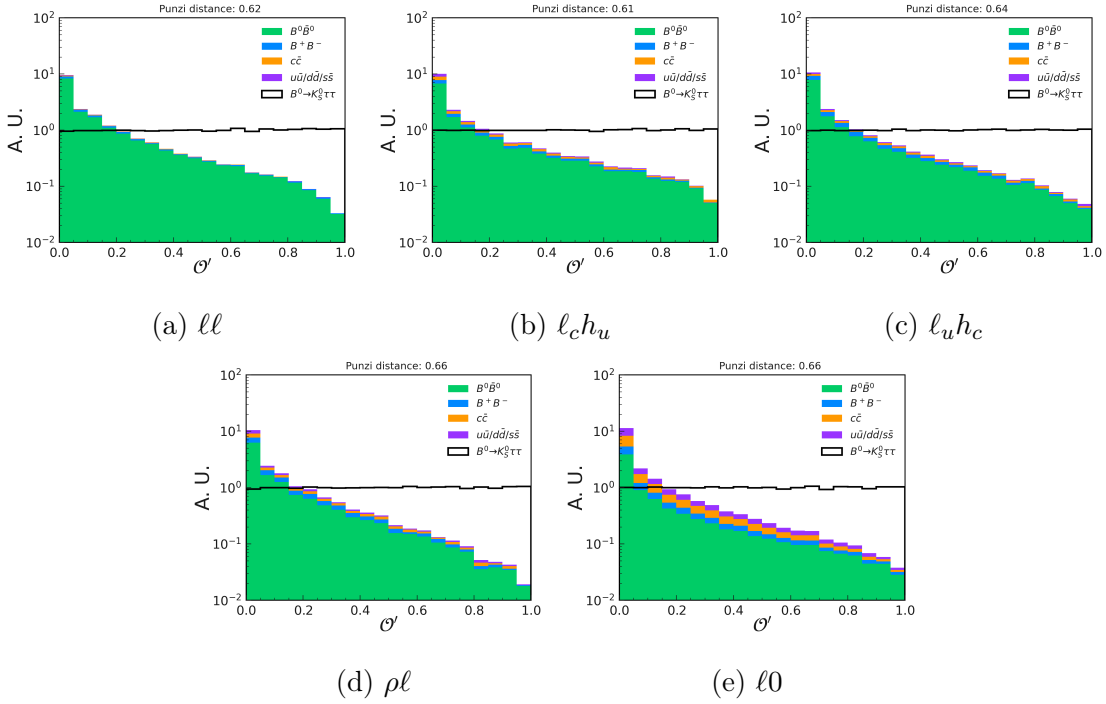


Figure 6.7: BDT outputs after Probability Integral Transformation (Belle II MC).

For the $\ell_u h_c$ category, the spectrum presents several resonant structures besides the combinatorial background where the K_S^0 and the t_c do not originate from the same particle:

1. $D^+ \rightarrow K_S^0 \pi^+$ events (vetoed);
2. Partial $D^+ \rightarrow K_S^0 \pi^+ X$ events;

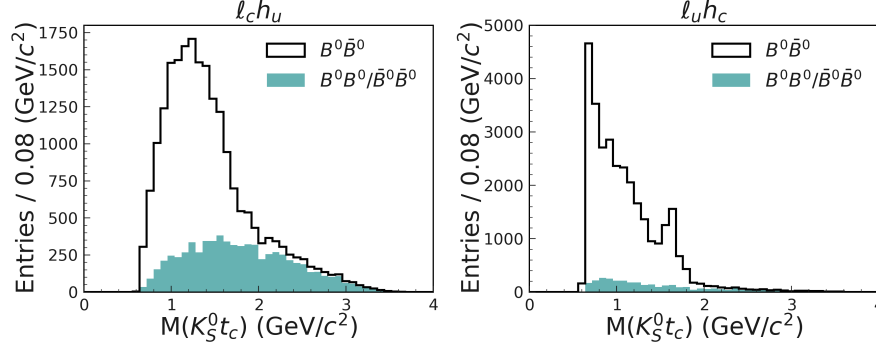


Figure 6.8: $M(K_S^0 t_c)$ for mixed and unmixed neutral B events in $\ell_c h_u$ (left) and $\ell_u h_c$ (right) categories (Belle II MC).

3. $D^{*+} \rightarrow D^0 \pi^+$ events where the t_c is the pion from the D^{*+} and the K_S^0 is from the D^0 (typically the D^0 decays into a K_S^0 and a number of π^0 's or $\eta^{(\prime)}$'s);
4. $K^{*+} \rightarrow K_S^0 \pi^+$ events.

This can be seen qualitatively in Fig. 6.9.

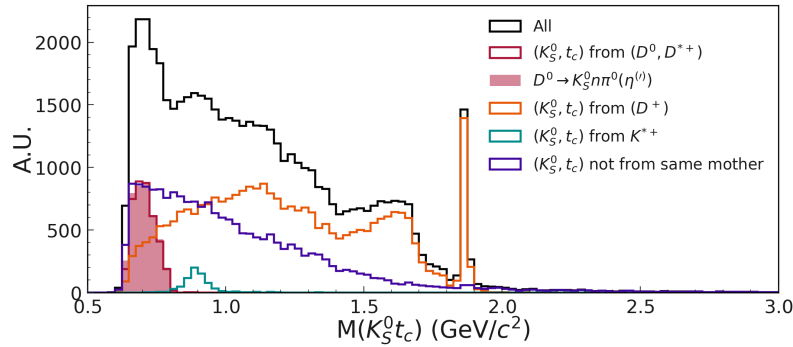


Figure 6.9: $M(K_S^0 t_c)$ for unmixed neutral B events in $\ell_u h_c$ category (Belle II MC). The main resonant structures are highlighted.

- m_{miss}^2 , defined as

$$m_{\text{miss}}^2 = p_{\text{miss}}^2 = \left(\frac{E_{\text{miss}}^*}{\mathbf{p}_{\text{miss}}^*} \right)^2; \quad p_{\text{miss}} = p_{e^+e^-} - p_{B_{\text{sig}}} - p_{B_{\text{tag}}},$$

is mostly used to separate between signal and $B \rightarrow D_{\text{had}}^{(*)} \ell \nu$ decays, for which the m_{miss}^2 tends to be closer to zero. This background is predominant in $\ell\ell$ and ℓh categories.

- E_{ECL} The energy of the deposits in ECL that are not associated to the object used for B_{tag} nor B_{sig} reconstruction. For well-reconstructed signal events, E_{ECL} should be zero or small, in case of beam-background induced or residual neutral hadrons. This variable is particularly important for the $\ell 0$ and $\rho\ell$ categories, in fact for the ‘cleaner’ categories, the E_{ECL} tends to peak at zero in a similar way to the signal.

- Another variable related to the amount of missing 4-momentum of the event is U_{MISS} , defined as:

$$U_{\text{MISS}} = E_{\text{MISS}} + c|\mathbf{p}|_{\text{MISS}}; \quad p_{\text{MISS}} = p_{e^+e^-} - p_{B_{\text{sig}}} - p_{B_{\text{tag}}} - p_{\text{ROE}}.$$

We keep this variable as it proves to improve the classifier performance despite its large correlation with E_{ECL} .

- $\mathbf{p}^*(\mathbf{t}_u)$ The momentum in the CM frame of the *uncorrelated* τ daughter candidate. The momentum tends to be higher for $B \rightarrow X\ell\nu$ background events with $t_u \equiv \ell$.
- q_{rec}^2 The reconstructed di- τ squared mass, obtained as a recoil of the $B_{\text{tag}}^0 K_S^0$ system. The background events more frequently have q^2 closer to $q^2 = (m_B - m_K)^2$ while signal events have a more phase-space like distribution (for SM-only contributions). We have also tried using the related variable $M(t_c, t_u)$, which provides information on background events with J/ψ and photon conversions, but given the correlation with q_{rec}^2 and the mild importance of the latter, we decide not to include it in our final feature list.
- ΔZ The difference in the z -coordinate between the B_{tag} and B_{sig} . The B_{sig} vertex is obtained from the $(K_S^0 t_c)$ system. The ΔZ distribution is larger for $B\bar{B}$ events compared to $q\bar{q}$ events, where the vertices are formed out of random combination.
- $M(K_S^0 \pi_{\text{sig}}^0)$ The invariant mass of the system composed of the K_S^0 candidate and the π^0 used for the daughter of the ρ from τ . The variable is used to check backgrounds having K^{*0} or D^0 resonances.
- $\cos TBTO$ See definition in 4.1. Although a pre-selection cut is applied at the reconstruction level to easily discard a large fraction of $q\bar{q}$ background, the rest of the distribution still allows to get some separation between $B\bar{B}$ and $q\bar{q}$ events. Other event-shape variables but were discarded due to high correlations with other observables used for the training.
- $M(\rho)$ When the ρ from τ is properly reconstructed, it peaks at the true $m_\rho \sim 770 \text{ MeV}/c^2$. Due to the large width of the ρ resonance, this variable is not highly discriminating but it still holds some separation power and is used for the $\rho\ell$ and the $\ell 0$ categories.

Figures 6.10 to 6.14 summarize the signal-to-background separation and the validation in the sidebands for the most-important variable (Belle II training) for each category. In Appendix B we show the complete set of input features used for building the ten classifiers (one for each fit category and experiment – Belle and Belle II): see Figures B.1 to B.10. The correlation matrices with the input features are collected in the Appendix B.1, Fig. B.11 (for Belle II), Fig. B.12 (for Belle).

The input features, as well as their importance computed by FastBDT, are also summarized in Fig. 6.15 for the $10 \times \{\text{odd}, \text{even}\}$ trainings. The importance is computed by summing up all separation gains along the path of the event through the trees.

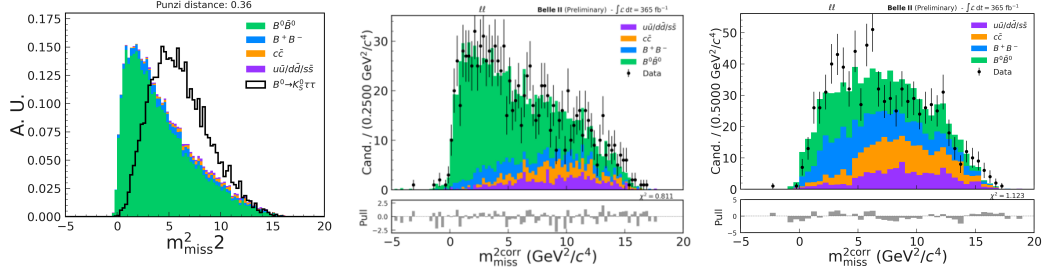


Figure 6.10: Most important training feature for the $\ell\ell$ category (Belle II MC). On the left, the MC backgrounds are compared to the signal shape. In the middle plot, the data/MC agreement is checked in the K_S^0 sideband. On the right, the agreement is checked in the M_{bc} sideband.

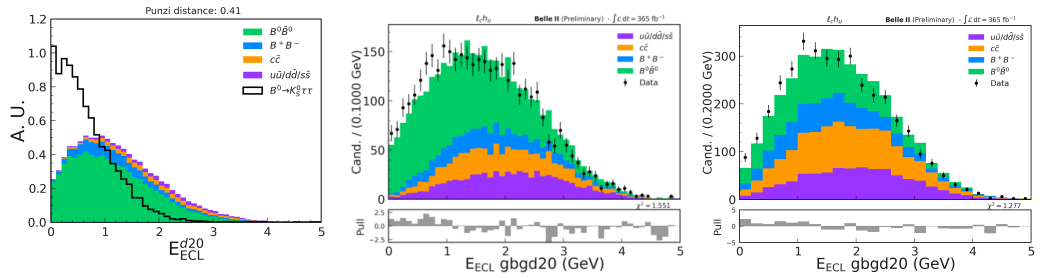


Figure 6.11: Most important training feature for the $\ell_c h_u$ category (Belle II MC). On the left, the MC backgrounds are compared to the signal shape. In the middle plot, the data/MC agreement is checked in the K_S^0 sideband. On the right, the agreement is checked in the M_{bc} sideband.

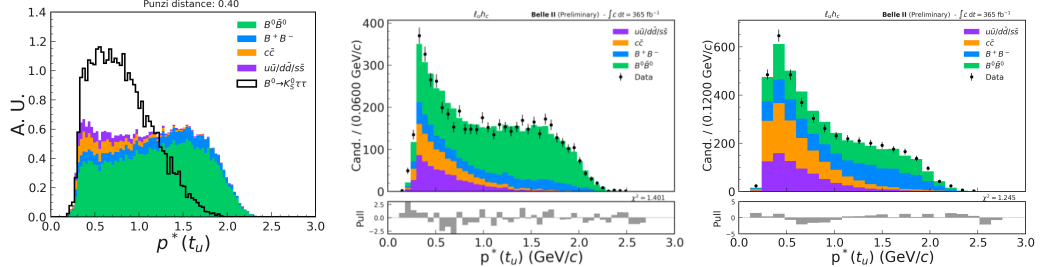


Figure 6.12: Most important training feature for the $\ell_u h_c$ category (Belle II MC). On the left, the MC backgrounds are compared to the signal shape. In the middle plot, the data/MC agreement is checked in the K_S^0 sideband. On the right, the agreement is checked in the M_{bc} sideband.

6.3 Background characterization

We study the background composition as a function of the the signal region definition. First, we check the relative fractions of $B^0\bar{B}^0$, B^+B^- , $c\bar{c}$, etc. at different $\mathcal{O}'$ thresholds: see Figures 6.16 for Belle MC and Fig. 6.17 for Belle II MC. For all fit categories, the fraction of $B^0\bar{B}^0$ backgrounds is dominant and its relative abundance raises at higher signal purities, showing that BDT is more effective at rejecting continuum and charged B backgrounds, while neutral B decays have more irreducible-background signatures.

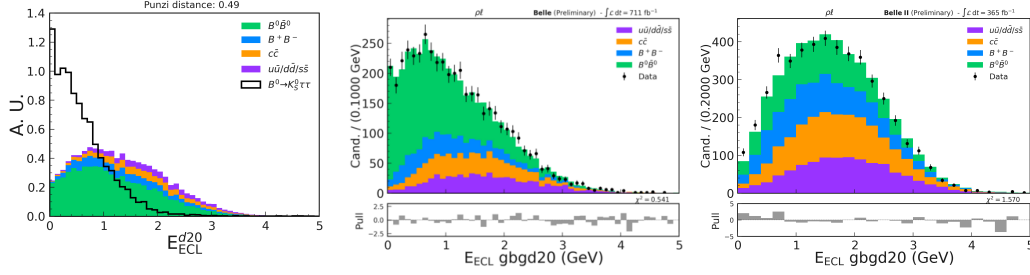


Figure 6.13: Most important training feature for the $\rho\ell$ category (Belle II MC). On the left, the MC backgrounds are compared to the signal shape. In the middle plot, the data/MC agreement is checked in the K_S^0 sideband. On the right, the agreement is checked in the M_{bc} sideband.

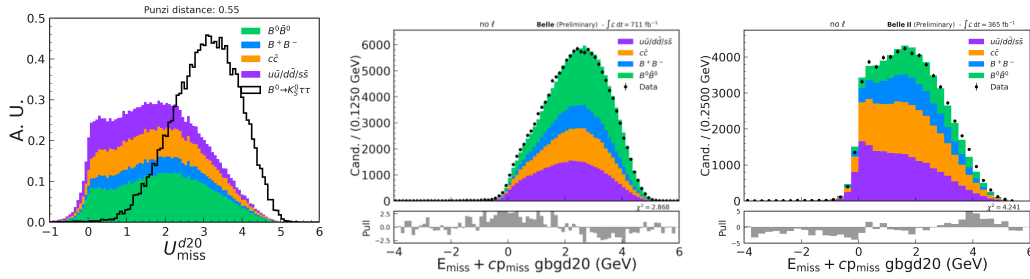


Figure 6.14: Most important training feature for the $\ell 0$ category (Belle II MC). On the left, the MC backgrounds are compared to the signal shape. In the middle plot, the data/MC agreement is checked in the K_S^0 sideband. On the right, the agreement is checked in the M_{bc} sideband.

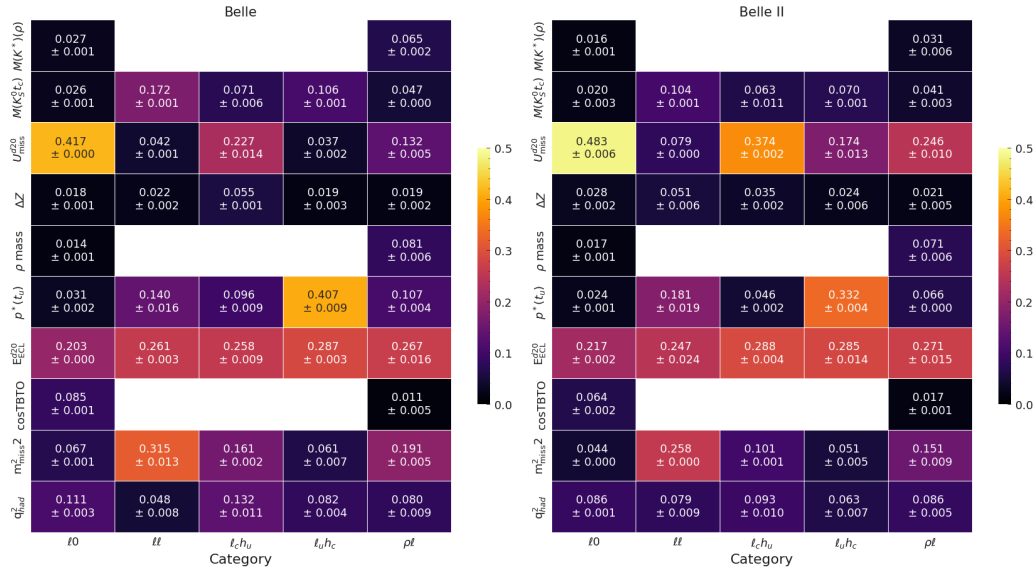


Figure 6.15: Summary of BDT training features and their importance for the Belle (left) and Belle II (right) datasets. The average of the odd/even training ranking value is shown, $\pm$ the standard deviation of the two.

Secondly, we check how the background composition, described at the beginning of Section 6, changes while approaching the most sensitive region; this is shown in Fig. 6.18 (Belle), Fig. 6.19 (Belle II) for the $B^0 \bar{B}^0$ simulation.

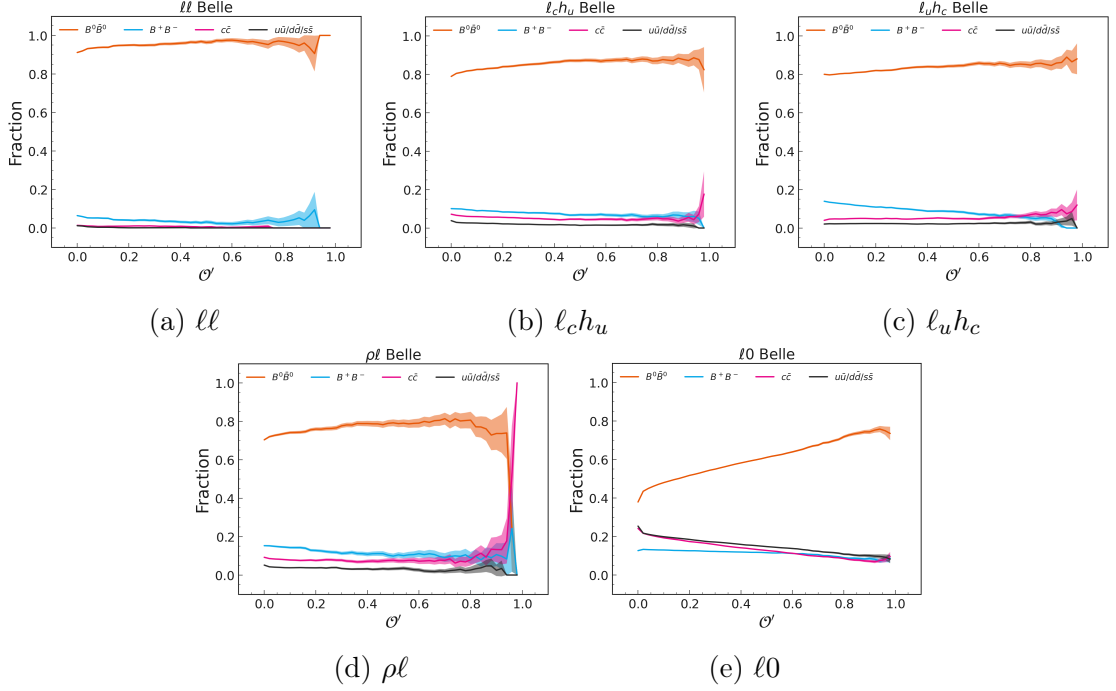


Figure 6.16: Background composition in MC background events as a function of the BDT cut (Belle MC).

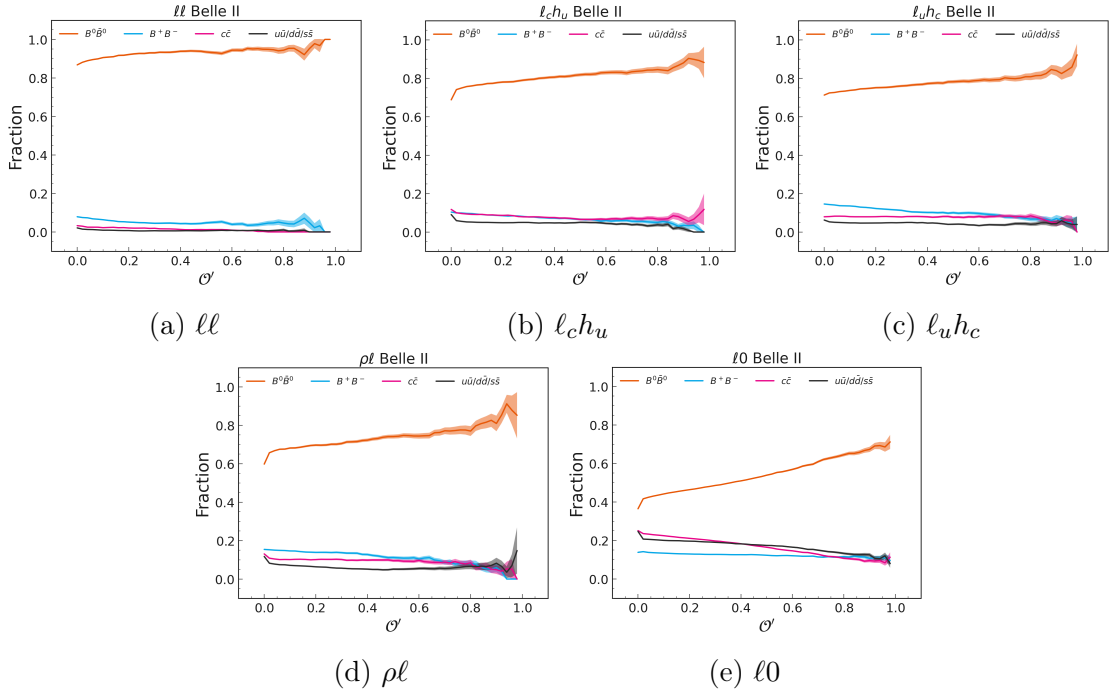


Figure 6.17: Background composition in MC background events as a function of the BDT cut (Belle II MC).

Additionally, we extract information about which B decays are present on the signal side of the reconstructed events surviving the $\mathcal{O}' > 0.3$ cut is applied. The fractions of the signal-side modes that contribute to our backgrounds, summarized in Tables 6.2 and 6.3, are obtained without applying any weights and without a requirement on the tag-side to be truth-matched. This is done to quantify the

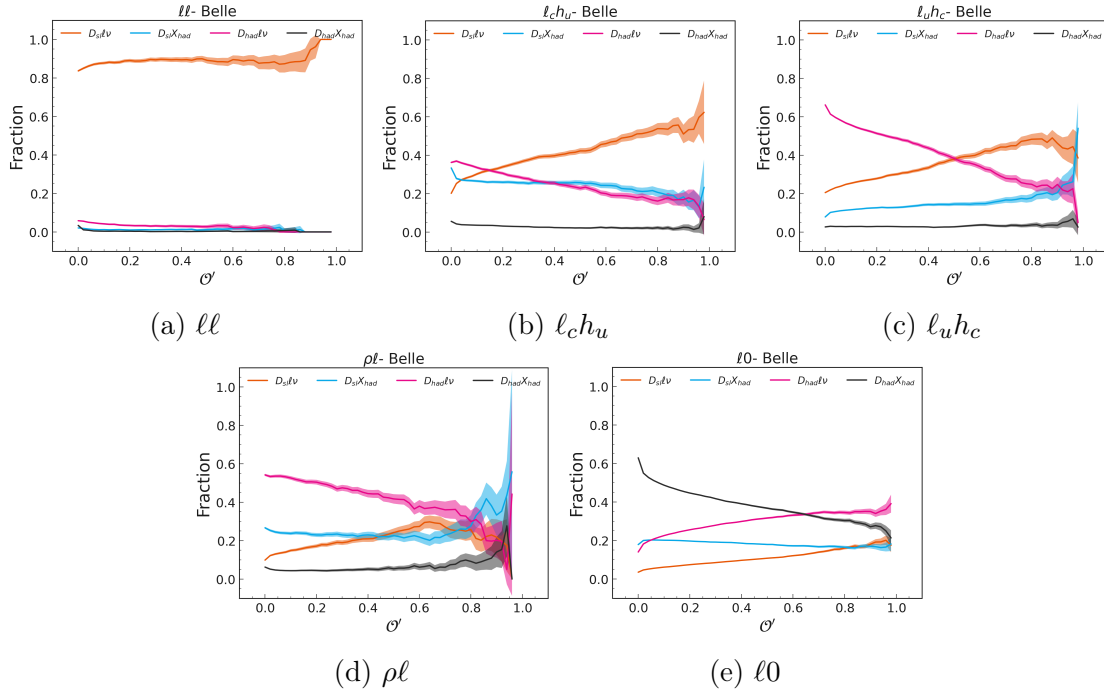


Figure 6.18: Background composition in $B^0 \bar{B}^0$ events as a function of the BDT cut (Belle MC). Here the fraction is computed using the total number of events (unmixed+mixed) for the normalization. The shaded bands correspond to the statistical uncertainties. In the plot legend, the labels sl and had indicate whether the decay is semileptonic or hadronic.

contribution of the modes that are less experimentally known, like the gap modes or modes with the D^{**} . Overall, we observe a high contribution of semileptonic B decays in both Belle and Belle II samples.

Table 6.2: Fractional composition of the signal-side B -meson in the $B \bar{B}$ background within the reconstructed categories in the fit region ($\mathcal{O}' > 0.3$) in Belle MC.

Background modes	ll	$\ell_c h_u$	$\ell_u h_c$	$\rho \ell$	$\ell 0$
$B \rightarrow \text{SL}$	0.939	0.644	0.808	0.665	0.361
$B \rightarrow D^{(*)} \ell \nu$	0.761	0.444	0.565	0.347	0.206
$B \rightarrow D^{(*)} \tau \nu$	0.104	0.103	0.098	0.089	0.086
$B \rightarrow D^{**} \ell \nu$	0.066	0.083	0.127	0.202	0.055
$B \rightarrow D^{**} \tau \nu$	0.007	0.008	0.006	0.007	0.010
$B \rightarrow D^{(*)} n(\eta/\pi)(\tau/\ell)\nu$	0.002	0.006	0.013	0.020	0.005
$B \rightarrow \text{HAD (without } D^{**})$	0.055	0.327	0.173	0.297	0.590
$B \rightarrow \text{HAD (with } D^{**})$	0.006	0.028	0.018	0.038	0.049

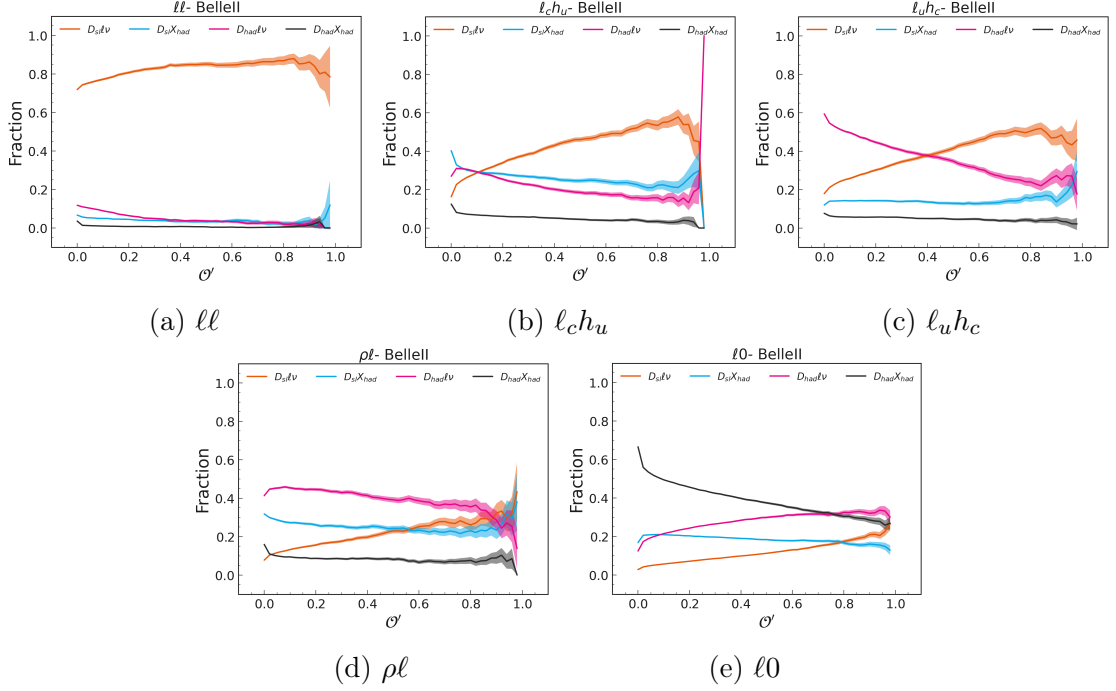


Figure 6.19: Background composition in $\bar{B}^0 \bar{B}^0 / B^0 B^0$ events as a function of the BDT cut (Belle II MC). Here the fraction is computed using the total number of events (unmixed+mixed) for the normalization. The shaded bands correspond to the statistical uncertainties. In the plot legend, the labels sl and had indicate whether the decay is semileptonic or hadronic.

Table 6.3: Fractional composition of the signal-side B -meson in the $B\bar{B}$ background within the reconstructed categories in the fit region ($\mathcal{O}' > 0.3$) in Belle II MC.

Background modes	$\ell\ell$	$\ell_c h_u$	$\ell_u h_c$	$\rho\ell$	$\ell 0$
$B \rightarrow \text{SL}$	0.884	0.617	0.773	0.619	0.349
$B \rightarrow D^{(*)}\ell\nu$	0.665	0.420	0.539	0.314	0.188
$B \rightarrow D^{(*)}\tau\nu$	0.113	0.120	0.112	0.103	0.096
$B \rightarrow D^{**}\ell\nu$	0.071	0.044	0.076	0.127	0.035
$B \rightarrow D^{**}\tau\nu$	0.010	0.012	0.013	0.023	0.015
$B \rightarrow D^{(*)}n(\eta/\pi)(\tau/\ell)\nu$	0.024	0.020	0.032	0.053	0.016
$B \rightarrow \text{HAD (without } D^{**})$	0.110	0.371	0.219	0.364	0.630
$B \rightarrow \text{HAD (with } D^{**})$	0.006	0.013	0.008	0.016	0.021

7. Validation

This chapter presents the validation of the analysis strategy in data control regions. The goal is to test the modeling of the variables used for background suppression and signal extraction in samples that are kinematically close to the signal region but depleted in signal. These studies are performed before interpreting the data in the signal-enhanced region.

Two main classes of validation samples are used. The first consists of the K_s^0 and M_{bc} sidebands, which probe different mixtures of continuum and combinatorial $B\bar{B}$ backgrounds. The second is a sample enriched in $B^0 \rightarrow D^{(*)-} \ell^+ \nu_\ell$ decays, obtained from events removed by the charm veto applied in the signal selection. Together, these regions test the modeling of the BDT input variables, the transformed classifier output $\mathcal{O}'$, and the normalization corrections applied to the signal and dominant background components.

7.1 K_s^0 and M_{bc} sidebands

We compare the distributions obtained in the K_s^0 signal region (SR) with those in the K_s^0 sideband (SB). If a good agreement is obtained, then this indicates that the K_s^0 sideband can be used to check the good modeling of the BDT training variables. The BDT score and the most separating variable are shown for each background source ($u, d, s, c\bar{c}, B^+B^-$ and $B^0\bar{B}^0$ events). This comparison between SR and SB sounds more reasonable compared to that of the total backgrounds because we know there can be differences in the relative contributions of the background sources. The plots for the Belle II categories are shown in Figures G.1 to G.5 in Appendix G (the distributions for Belle are collected in Appendix H, where we observe a slightly worse agreement). Overall, we see that the BDT scores and the input features agree well for all background sources and categories except for expected differences. This also means that the K_s^0 SB can be used for data/MC comparisons in the final fitting region – $\mathcal{O}' > 0.3$, corresponding to $\mathcal{O} > (0.7 \div 0.8)$ depending on the mode.

The plots with the data/MC comparison of the training features in the K_s^0 sideband can be checked in Figures C.1 to C.10 (Appendix C). Overall, the simulated distributions agree well with data, although for some categories we do observe a slope in the pulls distributions in variables involving ROE information (U_{miss}, E_{ECL}) but we believe dedicated corrections are not needed for this level of discrepancy. Because we observe a good agreement in the high $\mathcal{O}'$ region for the K_s^0 sideband, the BDT absorbs such discrepancies. While we also observe good agreement between the K_s^0 sideband and signal region MC, the BDT is also expected to absorb such discrepancies in the signal window.

In Figures 7.2 and 7.3 we compare the $\mathcal{O}'$ distributions in data and MC for the K_s^0 sidebands in the signal region (defined as $\mathcal{O}' > 0.3$). Given the fact that the scaling factors obtained on all events (Tab. 5.2) are not necessarily constant over

the $\mathcal{O}'$ threshold, we repeat the procedure described in Sections 5.2.1 and 5.2.2.

The same comparisons are done for the M_{bc} sideband. Compared to the K_s^0 sideband, and even more, to signal region, this sample is much more $q\bar{q}$ enhanced. Nevertheless, the agreement of the training features is checked in Figures D.1 to D.10 (Appendix D), probing different backgrounds with respect to K_s^0 sideband. The distributions agree well with data, however some residual mismatches between data and simulation can be seen. Since the sideband is heavily $q\bar{q}$ contaminated, we assume that such mis-modelings would be much less pronounced in the signal region. In Figures 7.4 and 7.5 we compare the $\mathcal{O}'$ distributions in data and MC for the M_{bc} sidebands in the signal region ($\mathcal{O}' > 0.3$) with the $SF_{B\bar{B}}^{comb}$ factors applied.

The normalization factors computed in the fit window are summarized for both sidebands in Tab. 7.1. Overall we observe that a larger correction is required in Belle compared to Belle II, and that the factors obtained in the $\mathcal{O}' > 0.3$ region are more consistent with the ‘global’ ones in Belle II. For each of the two sidebands, we cannot be sure whether the data/MC discrepancy really comes from the combinatorial B_{tag} component and not from other sources (for example, in the K_s^0 SB from the two prompt pions combined into a K_s^0 candidate, or, in the M_{bc} SB, from a different composition in the B_{tag} modes). Of course, the fact that the two sets of factors are compatible in most cases (see Fig. 7.1) reassures us on the goodness of the calibration procedure. Nevertheless, due to our lack of certainty we decide to not correct the combinatorial B_{tag} component and keep the conservative approach where we assign an uncertainty of 50% to it.

Coming to the $q\bar{q}$ simulations, the scaling factors from off-resonance signal region are used to correct their normalizations, and the statistical uncertainties on the corrections are assigned as systematic uncertainties.

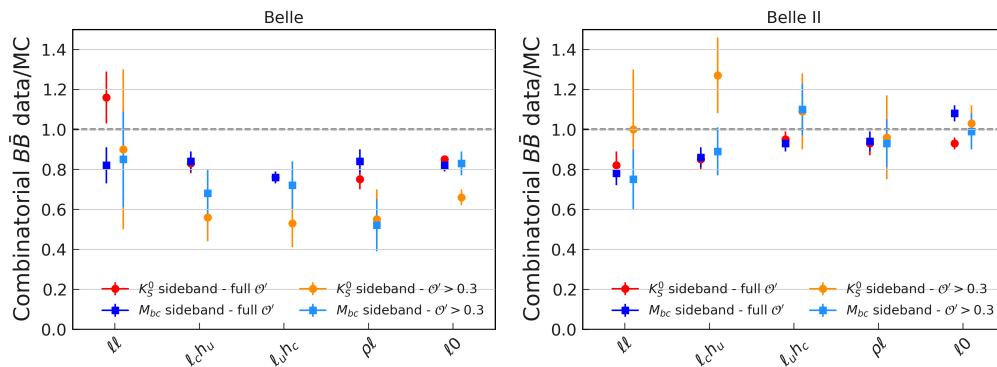


Figure 7.1: Scaling factors for B_{tag} combinatorics for Belle (left) and Belle II (right) in K_s^0 sideband and M_{bc} sideband (on-resonance) samples.

7.2 D -vetoed samples

As mentioned in other sections, an obvious background for the analysis consists of semileptonic B^0 decays followed by the $D^- \rightarrow \{K_s^0 \pi^-, K_s^0 \rho^-\}$ decays. In Section 6 we explain that this background is vetoed by applying the selection $M(K_s^0, t_c) \notin [1.8, 1.9] \text{ GeV}/c^2$ for the $\ell_u h_c$, $\rho\ell$ and $\ell 0$ categories. Similarly to what was done in [19, 20, 49], we can check the analysis procedure before the BDT cut using the vetoed regions. Figures 7.6 and 7.7 show the data/MC comparison in the $M(K_s^0, t_c)$ vetoed region for the $\ell_u h_c$ and $\ell_u \rho_c$ categories, respectively. For both Belle and

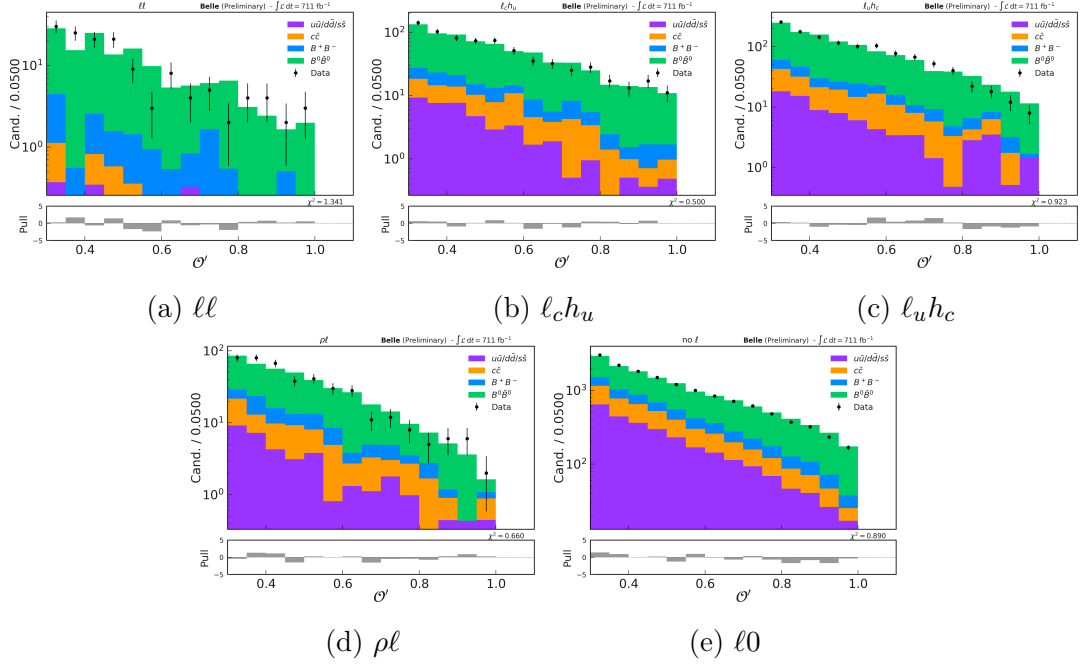


Figure 7.2: Transformed BDT score in signal region, Belle K_S^0 sideband, on-resonance data (Belle). All signal-region scaling factors are applied.

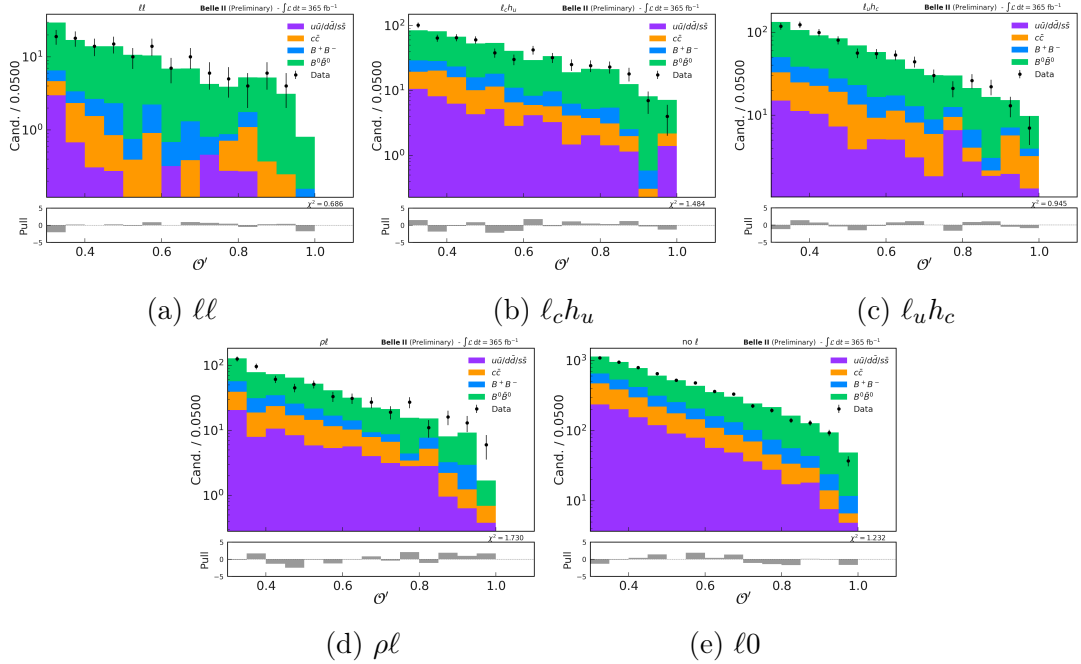


Figure 7.3: Transformed BDT score in signal region, Belle II K_S^0 sideband, on-resonance data. All signal-region scaling factors are applied.

Belle II data we check the distributions of $M(K_S^0, t_c)$, m_{miss}^2 and $\mathcal{O}'$. Despite the low statistics, especially for Belle II, this indicates that the main efficiency corrections (lepton ID, B_{tag} , K_S^0 and π^0 reconstruction) applied to the MC samples yield to a satisfactory agreement with data.

The different background components could be selected according to their m_{mis}^2 values: $D^-\ell^+\nu_\ell$ in $m_{\text{mis}}^2 < 0.5\text{GeV}^2/c^4$, $D^{*-}\ell^+\nu_\ell$ in $m_{\text{mis}}^2 \in (0.5, 1.0)\text{GeV}^2/c^4$ or

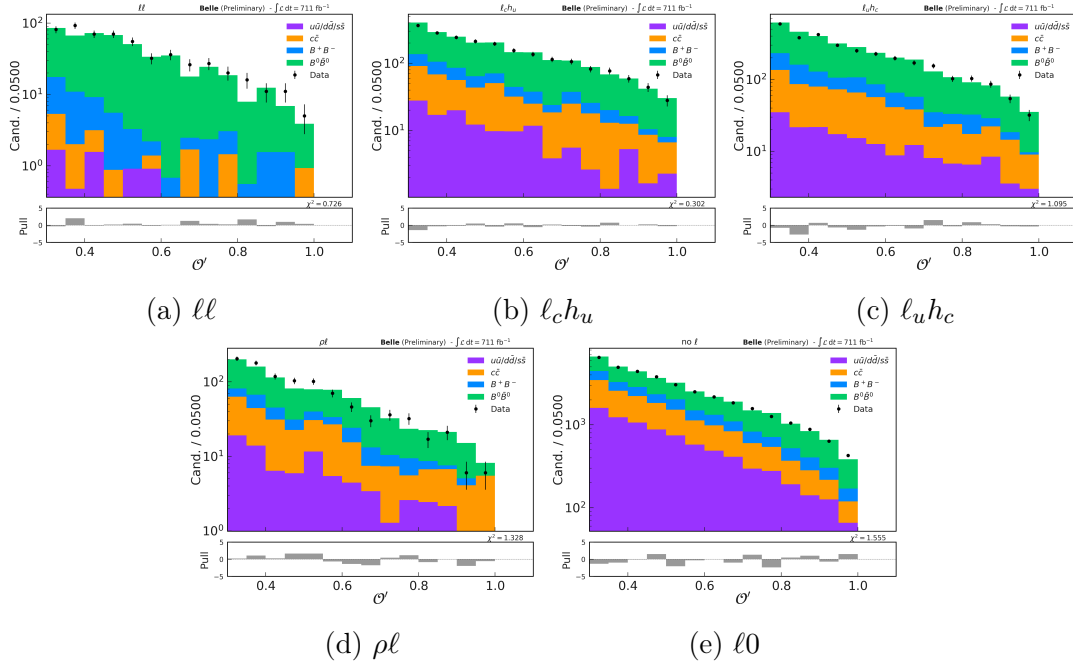


Figure 7.4: Transformed BDT score in signal region, Belle M_{bc} sideband, on-resonance data. All scaling factors are applied.

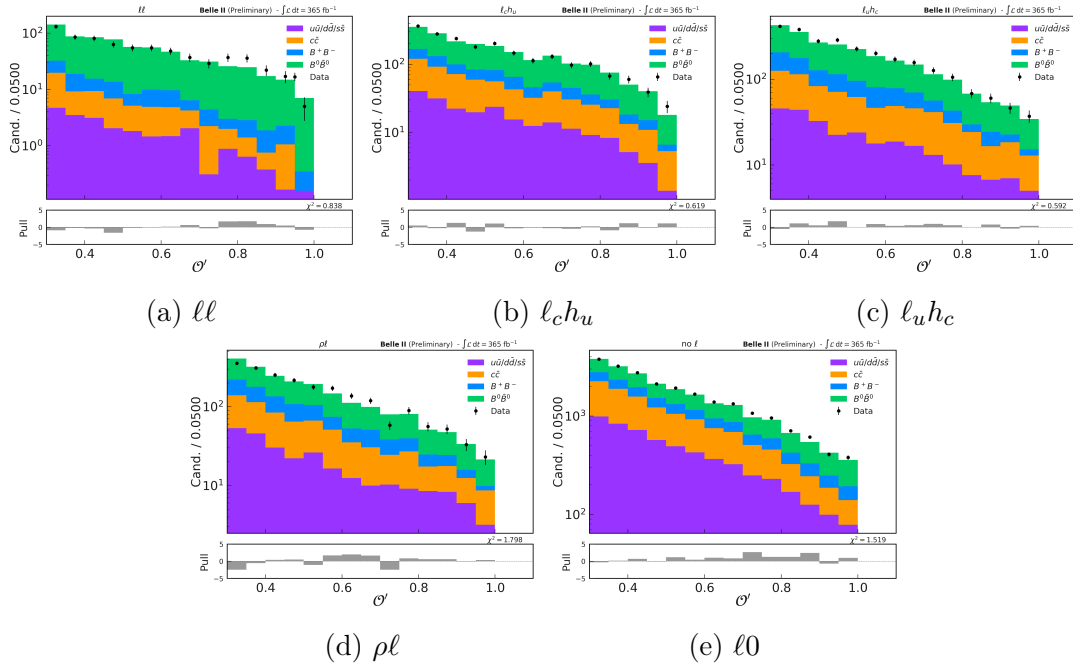


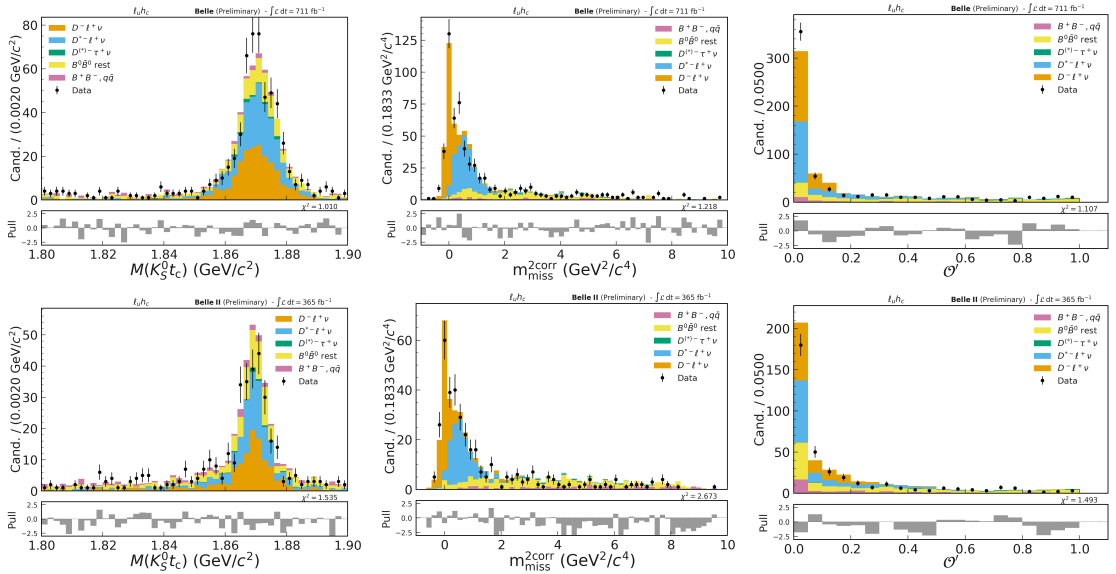
Figure 7.5: Transformed BDT score in signal region, Belle II M_{bc} sideband, on-resonance data. All signal-region scaling factors are applied.

$D^{(*)-}\tau^+\nu_\tau$ in $m_{\text{mis}}^2 > 1.0\text{GeV}^2/c^4$. Similarly, low $\mathcal{O}'$ values are enriched in $B^0 \rightarrow D^{(*)-}\ell^+\nu_\ell$ events while the right-end tails include $B^0 \rightarrow D^{(*)-}\tau^+\nu_\tau$ and other $B\bar{B}$ events which are more signal-like.

In Fig. 7.8 we show the M_{bc} distributions for the three categories ℓ_hh_c , $\ell_u\rho_c$ and $\ell 0$ to study the data/MC agreement for peaking/combinatorial B_{tag} 's in both full and signal BDT regions. We can see that for the ℓ_hh_c and $\ell_u\rho_c$ (selected subset of $\rho\ell$)

Table 7.1: Data/MC scaling factors for Belle and Belle II derived in the $\mathcal{O}' > 0.3$ region.

SF	$\ell\ell$	$\ell_c h_u$	$\ell_u h_c$	$\rho\ell$	$\ell 0$	
$q\bar{q}$ (OFF res)	0.57 ± 0.12				0.67 ± 0.03	Belle
$q\bar{q}$ (OFF res, K_S^0 SB)	0.99 ± 0.18				0.75 ± 0.04	
$B\bar{B}$ (ON res, K_S^0 SB)	0.9 ± 0.4	0.56 ± 0.12	0.53 ± 0.12	0.55 ± 0.15	0.66 ± 0.04	
$q\bar{q}$ (OFF res, M_{bc} SB)	0.99 ± 0.17				0.69 ± 0.03	
$B\bar{B}$ (ON res, M_{bc} SB)	0.85 ± 0.24	0.68 ± 0.12	0.72 ± 0.12	0.52 ± 0.13	0.83 ± 0.06	
$q\bar{q}$ (OFF res)	1.05 ± 0.16				0.68 ± 0.05	Belle II
$q\bar{q}$ (OFF res, K_S^0 SB)	1.17 ± 0.20				1.03 ± 0.07	
$B\bar{B}$ (ON res, K_S^0 SB)	1.0 ± 0.3	1.27 ± 0.19	1.09 ± 0.19	0.96 ± 0.21	1.03 ± 0.09	
$q\bar{q}$ (OFF res, M_{bc} SB)	0.84 ± 0.09				0.77 ± 0.04	
$B\bar{B}$ (ON res, M_{bc} SB)	0.75 ± 0.15	0.89 ± 0.12	1.10 ± 0.13	0.93 ± 0.12	0.99 ± 0.09	

Figure 7.6: $M(K_S^0 t_c)$ for $\ell_u h_c$ events in the vetoed $M(K_S^0 t_c)$ region for Belle (upper row) and Belle II (lower row) datasets. MC events are scaled to 711/fb and 365/fb respectively. All efficiency corrections are applied.

categories, where most of the B_{tag} 's are peaking, the agreement is good, confirming the goodness of the FEI corrections. The only category having a significant amount of combinatorial B_{tag} 's is $\ell 0$ (plots in the rightmost column); here we observe an undershoot of data compared to MC, somewhat confirming what was obtained in the K_S^0 sideband. This further check validates that the combinatorial B_{tag} component

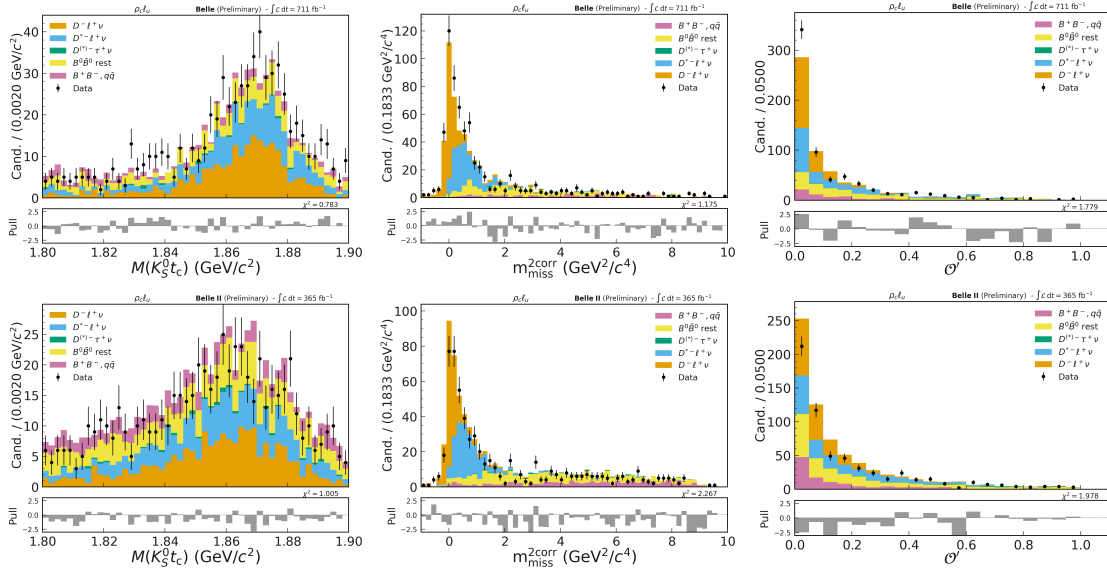


Figure 7.7: $M(K_S^0 t_c)$ for $\ell_u \rho_c$ events in the vetoed $M(K_S^0 t_c)$ region for Belle (upper row) and Belle II (lower row) datasets. MC events are scaled to 711/fb and 365/fb respectively. All efficiency corrections are applied.

does need an ad-hoc scaling, which we cover with the 50% uncertainty assigned to it.

To validate signal efficiency, or the BDT cut efficiency, we again utilize the $\ell_u h_c$ and $\rho_c \ell_u$ categories in the D -vetoed region, as this sample is very pure in $B^0 \bar{B}^0$ with well-reconstructed tag side. Additionally, the sample includes all signal final state objects: e, μ, ρ, h, K_S^0 . However, since both $M(K_S^0 t_c)$ and m_{mis}^2 are used in the BDT training as input variables, the $B \rightarrow D^{(*)} \ell \nu$ events are very much background-like and so is the $\mathcal{O}'$ distribution, leading to very low statistics in the signal region. To mitigate this difference between the D -vetoed sample and the signal, and enhance the available statistics, we inject signal-like $M(K_S^0 t_c)$ and m_{mis}^2 distributions into both data and MC samples. This procedure flattens the $\mathcal{O}'$ distribution, making it more signal-like. Figure 7.9 shows the $\mathcal{O}'$ distributions in the signal region before and after the injection. The data/MC ratios obtained after the injection are listed in Table 7.2: as they are all compatible with 1 within the uncertainties (at $\sim 1 \sigma$ -level), we do not apply efficiency corrections but assign 10% systematic uncertainty to the signal efficiency.

Table 7.2: Data/MC ratios after the injection in the D -vetoed sample.

	Belle	Belle II
$\ell_u h_c$	0.99 ± 0.07	0.98 ± 0.09
$\ell_u \rho_c$	1.11 ± 0.07	0.95 ± 0.06

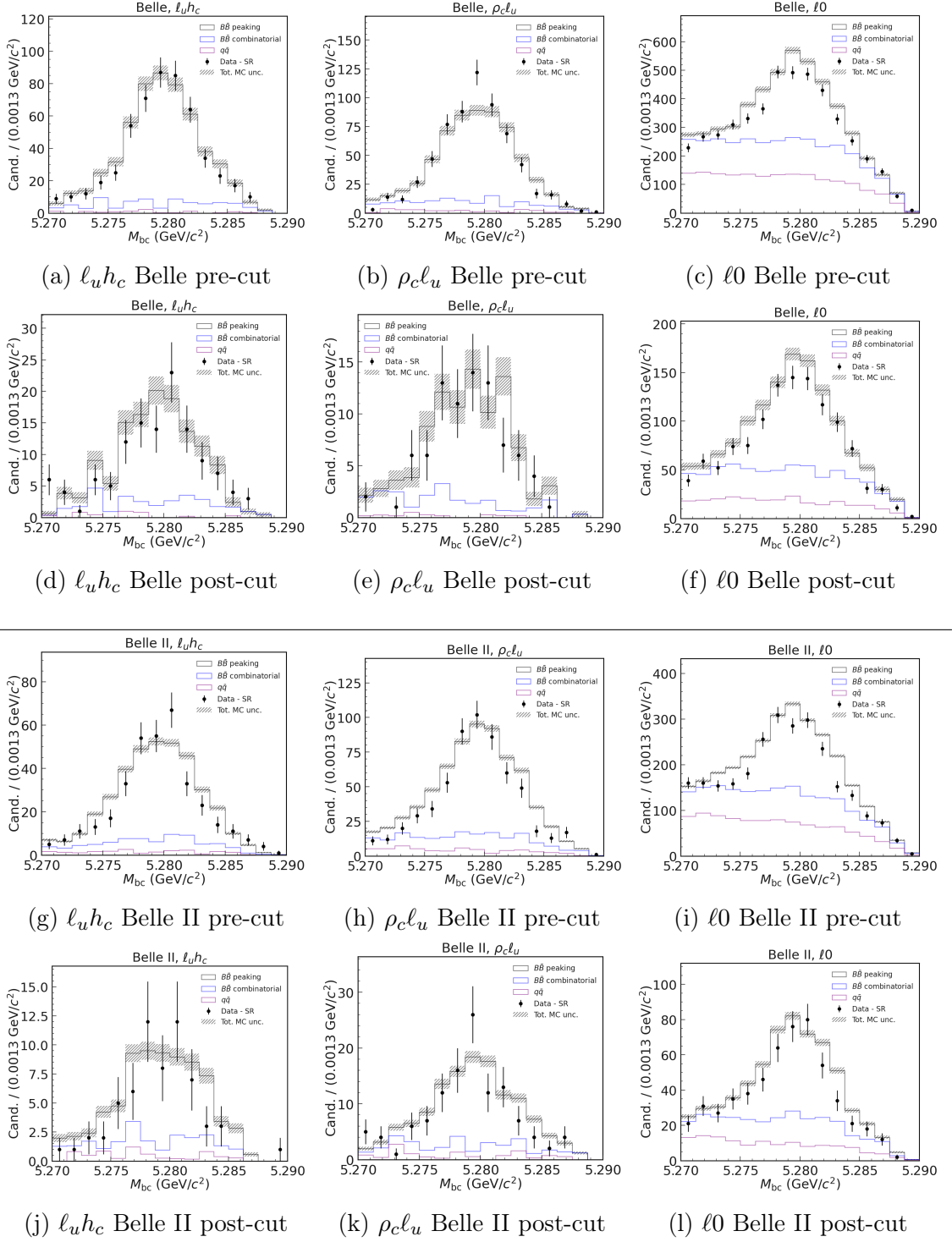


Figure 7.8: M_{bc} distributions in the D -vetoed region in data and MC. $q\bar{q}$ and both peaking and combinatorial $B\bar{B}$ components are shown in MC. Belle (II) MC is shown above (below) the horizontal divider.

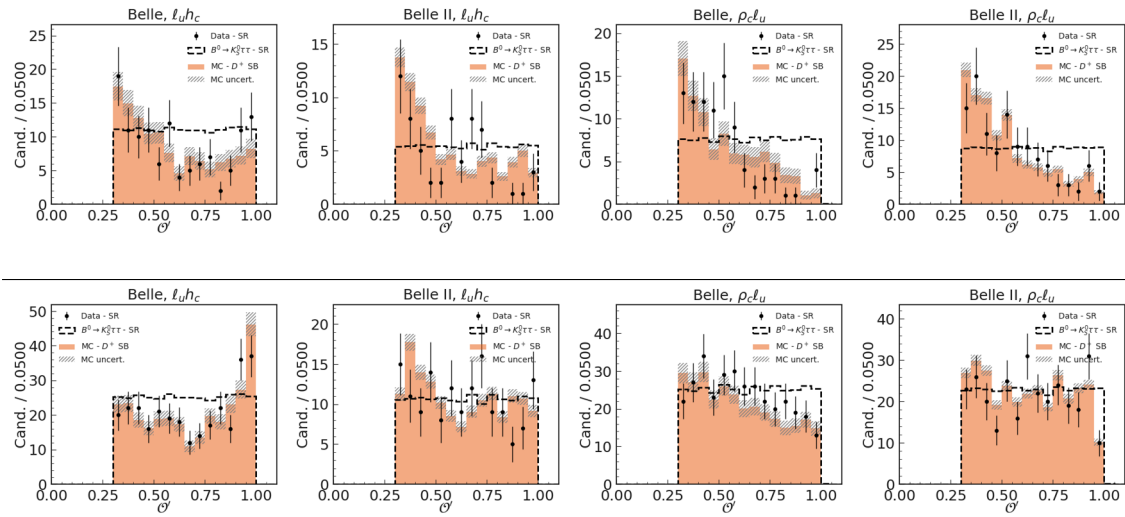


Figure 7.9: $\mathcal{O}'$ distributions in the D -vetoed sample before (upper row) and after (lower row) the injection of signal MC distributions into $M(K_S^0 t_c)$ and m_{mis}^2 .

8. Signal extraction

This chapter describes the extraction of the $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ branching fraction. The signal is extracted with a simultaneous binned maximum-likelihood fit to the transformed BDT output variable $\mathcal{O}'$ in the Belle and Belle II datasets and in the five reconstructed $\tau\tau$ categories. The chapter first introduces the fit observable, the template construction, the binning choice, and the statistical model. It then describes the implementation of systematic uncertainties and the validation of the fit on simulated samples. The results obtained from Belle, Belle II, and the combined dataset are then presented. Finally, the neutral-mode result is combined with the charged pseudoscalar channel to obtain an isospin-averaged constraint on $B \rightarrow K \tau^+ \tau^-$ decays.

8.1 Fitting strategy

The signal extraction is performed with a binned maximum-likelihood fit to the transformed BDT output variable $\mathcal{O}'$ in the range $[0.3, 1.0]$. The fit is performed simultaneously across the five signal categories and the Belle and Belle II datasets. The statistical model is implemented with `cabinetry`, which uses `pyhf` [50, 51], a pure-Python implementation of the ROOT `HistFactory` [52] framework for histogram-based analyses and interval estimation.

The measurement relies on comparing the observed event yields with the predictions from simulation. This comparison is described by a statistical model, which gives the probability of observing the data for a given set of model parameters. In a `HistFactory` or `pyhf` model, the data are represented by several binned distributions, referred to as channels or categories. In each bin of each category, the expected yield is obtained by summing the contributions from different physics processes, or samples. These expected yields can depend on free parameters, such as the signal strength, and on constrained nuisance parameters, which encode statistical and systematic uncertainties.

As described in Section 4, the fit categories are constructed using orthogonal selections based on the reconstructed $\tau\tau$ -pair daughter composition. The number of observed signal events, N_{sig} , is related to the branching fraction by

$$N_{\text{sig}} = 2 \times N_{\Upsilon(4S)} \times f^{00} \times \mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) \times \varepsilon, \quad (8.1)$$

where $f^{00} = \mathcal{B}(\Upsilon(4S) \rightarrow B^0 \bar{B}^0)$ is the fraction of neutral B -meson pairs produced in $\Upsilon(4S)$ decays, $N_{\Upsilon(4S)}$ is the number of produced $\Upsilon(4S)$ mesons, and ε is the signal reconstruction efficiency. The value of $N_{\Upsilon(4S)}$ differs between the Belle and Belle II datasets, while ε depends on the fit category. The central values and uncertainties of $N_{\Upsilon(4S)}$ and f^{00} are discussed in Section 8.2.

For each category, three binned histograms, referred to as *templates*, are built from simulation. They describe the expected distributions of the signal, $B\bar{B}$ back-

ground, and $q\bar{q}$ background in the fit variable. The three templates are combined to form the nominal model for the expected event yields. The templates are not fixed: their normalizations and, where relevant, their shapes are allowed to vary according to statistical and systematic uncertainties, whose implementation is described below and in Section 8.2.

The $B\bar{B}$ template contains the peaking $B^0\bar{B}^0$ component, in which the B_{tag}^0 is correctly reconstructed, and the combinatorial $B^0\bar{B}^0$ and B^+B^- components. Since the flavor of the signal-side B^0 meson is not known reliably on an event-by-event basis, the $B\bar{B}$ template also includes peaking and combinatorial contributions from same-flavor B^0B^0 and $\bar{B}^0\bar{B}^0$ events, where the B_{tag} candidate is combinatorial. The $q\bar{q}$ template describes combinatorial backgrounds from continuum events.

All templates are weighted to account for known data/MC differences. The background templates are additionally scaled to the integrated luminosity of the data. The signal template is normalized to correspond to a branching fraction

$$\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = 10^{-3},$$

which defines the reference signal yield before the fit. As shown in Section 6, the signal distribution is uniform by construction in the full range of $\mathcal{O}'$, up to small variations due to finite sampling.

In addition to the templates, a pseudodata sample is constructed from the same simulated background samples, assuming the absence of signal. This sample is also referred to as Asimov data and is used to validate the fit and estimate the expected sensitivity on simulation. The signal contribution is set to zero because the Standard Model prediction, $\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) \approx 10^{-7}$, is approximately four orders of magnitude below the current expected experimental sensitivity. The pseudodata therefore correspond to the no-signal hypothesis.

Since signal events are uniformly distributed in $\mathcal{O}'$, the binning is naturally chosen according to signal-efficiency quantiles. The initial choice was to use bins corresponding to 5% signal-efficiency quantiles in all categories. An additional requirement was then imposed that each bin should contain at least 10 expected background events. For this reason, the bin size was increased to 10% signal-efficiency quantiles in the $\ell\ell$ and $\rho\ell$ categories. In contrast, the $\ell 0$ category has much larger bin-wise background yields, so a finer binning corresponding to 2.5% signal-efficiency quantiles is used. Even with this finer binning, the expected number of entries per bin remains well above 10, so the asymptotic formulae used for the test statistic are valid. Their applicability has also been checked with toy experiments.

In summary, all bins within a category have the same width in signal efficiency, and the number of bins is chosen to be the following for both Belle and Belle II:

$\ell\ell, \rho\ell$: 7 bins, corresponding to 10% signal-efficiency quantiles;

$\ell_c h_u, \ell_u h_c$: 14 bins, corresponding to 5% signal-efficiency quantiles;

$\ell 0$: 28 bins, corresponding to 2.5% signal-efficiency quantiles.

The pre-fit distributions of the constructed templates and pseudodata for the combined fit are shown in Fig. 8.1, with the yields scaled to the data integrated luminosity. The uncertainties shown on the templates are statistical only and are computed as

$$\sigma_{\text{stat}}(i\text{-th bin}) = \sqrt{N^i/N_s},$$

where N^i is the number of entries in the i -th bin and N_s is the ratio between the effective integrated luminosities of simulation and data. The uncertainty on the pseudodata is taken as the square root of the number of entries in each bin, assuming the integrated luminosity of data. As discussed in Section 6, two effective luminosities of Belle MC and four effective luminosities of Belle II MC are used to reduce the statistical uncertainty on the background templates. For Belle, the effective integrated luminosity corresponds approximately to the same number of events as in the Belle II MC sample.

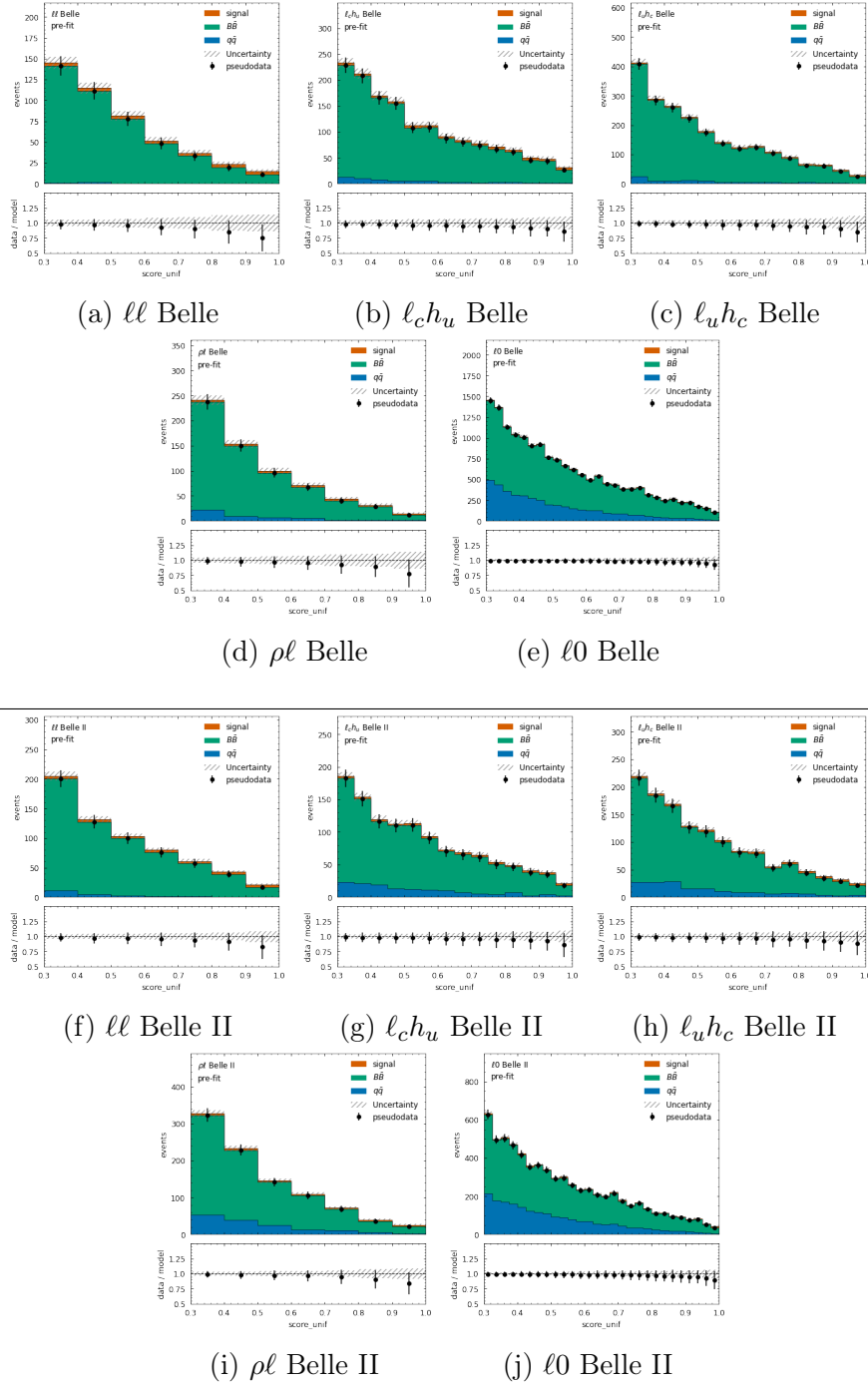


Figure 8.1: Pre-fit distributions of the BDT output $\mathcal{O}'$. The signal template corresponds to $\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = 10^{-3}$. Data points labeled as pseudodata are constructed assuming the absence of signal and correspond to the Asimov data. The shown yields correspond to the integrated luminosity in data. The five upper plots correspond to Belle MC, while the five lower plots correspond to Belle II MC. Uncertainties on MC correspond to statistical uncertainty only.

The setup described above is implemented in the `cabinetry` framework, which allows one to:

- create binned templates from the signal $B^0 \rightarrow K_S^0 \tau \tau$ and background samples, namely $B\bar{B}$ and $q\bar{q}$;
- include an unconstrained normalization scale factor for the signal sample, denoted by μ ;
- parametrize variations in the normalization and/or shape of any sample due to systematic effects;
- include bin-by-bin statistical uncertainties from the limited MC sample sizes;
- combine multiple categories and correlate nuisance parameters across them.

The templates in each category are characterized by statistical and systematic uncertainties, which can be either unique to a category or shared among several categories. The separation into categories is useful not only for background suppression, but also for the treatment of systematic effects related to the background composition. For example, normalization uncertainties on the $q\bar{q}$ or $B\bar{B}$ yields can be derived in a category-specific or multi-category way and then treated as correlated or uncorrelated among categories, depending on their origin.

The systematic uncertainties are included in the fit through nuisance parameters with Gaussian constraints. The modifiers used in the model are:

- **normsys**: a constrained scaling of a template, representing a normalization uncertainty. It is applied coherently to all bins of a given sample and is controlled by a single nuisance parameter. This modifier is used for uncertainties related to

the sample luminosity,

the number of $\Upsilon(4S)$ mesons, $N_{\Upsilon(4S)}$,

the $q\bar{q}$ normalization.

- **histosys**: a correlated per-bin shape uncertainty applied to a sample. Two variations, corresponding to upward and downward shifts of the expected event rate, are provided for each bin of the sample. In this analysis, some uncertainties are implemented with a combined **normsys+histosys** treatment, using a single Gaussian constraint term for both the normalization and shape components. This category includes uncertainties related to

efficiency corrections, including FEI, lepton ID, K_S^0 and π^0 reconstruction, and tracking;

normalization corrections, such as the correction to the combinatorial $B\bar{B}$ component, which affects only part of the $B\bar{B}$ template and therefore changes its shape;

MC generation inputs, such as χ_d , f^{00} , the signal form factor, and B -meson branching fractions.

- **normfactor**: an unconstrained scaling of a sample. It is used for the signal strength parameter μ .
- **statererror**: a set of bin-wise scale factors applied to each sample, used to model the uncertainty due to finite MC statistics. There is one nuisance parameter per bin.

In summary, the likelihood function f is a product of Poisson probability density functions, which combines the information from all bins and categories, multiplied by Gaussian constraint terms for the nuisance parameters:

$$f(\mu | \chi) = \prod_{\substack{c \in \\ \text{categories}}} \prod_{\substack{b \in \\ \text{bins}_c}} \text{Pois}(n_{c,b} | \nu_{c,b}(\mu, \chi)) \times \prod_{\chi \in \chi} \text{Gaus}_\chi(a_\chi | \chi). \quad (8.2)$$

Here, $n_{c,b}$ is the number of observed events in bin b of category c , while the corresponding expected value is denoted by $\nu_{c,b}$. The systematic uncertainties discussed in the next section are embedded into the likelihood through the nuisance parameters χ , which modify the expected event counts and are constrained by Gaussian probability density functions Gaus_χ linked to auxiliary data a_χ .

The expected number of events in each bin can be written as

$$\nu_{c,b}(\mu, \chi) = \mu \cdot 2 \cdot N_{\Upsilon(4S)} \cdot f_{00} \cdot \mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) \cdot \varepsilon_{c,b}(\chi) + \nu_{c,b}^{B\bar{B}}(\chi) + \nu_{c,b}^{q\bar{q}}(\chi). \quad (8.3)$$

Here, $\nu_{c,b}^{B\bar{B}}$ and $\nu_{c,b}^{q\bar{q}}$ are the expected yields of the $B\bar{B}$ and $q\bar{q}$ templates, respectively. Although the nominal signal efficiency is uniform in the fit variable, it becomes bin-dependent through the inclusion of shape systematic uncertainties and is therefore denoted by $\varepsilon_{c,b}$. The multiplicative factor μ , which is free in the fit, scales the overall normalization of all signal templates relative to the assumed reference branching fraction.

8.2 Systematic uncertainties

Other than those related to the efficiency corrections described in Section 5, our fitting procedure suffers from other systematic uncertainties that we should estimate and take into account. Below we list the sources of uncertainty and how we incorporate them into our fit model. The summary of the systematic uncertainties, the corresponding categories and the samples that are affected by them is shown in the modifier grid in Fig. 8.6.

- **Efficiency corrections** The efficiency corrections (ℓ ID, fake rates, K_S^0 , FEI, π^0 , tracking) come with systematic uncertainties. These corrections are usually binned, and the usual way is to treat the systematic uncertainties across different bins as fully correlated. We treat such systematic sources as follows: the systematic uncertainty is assigned to every event belonging to the bin for which the correction is provided. Then, as these events populate the bins of our $\mathcal{O}'$ template, they still remain fully correlated. For the “up” or “down” variation of the template, the central value of the event weight is raised up or down by a corresponding uncertainty, unique to each event (as displayed in Fig. 8.2). In such a way, we treat the reconstruction efficiency uncertainties and fake rates.

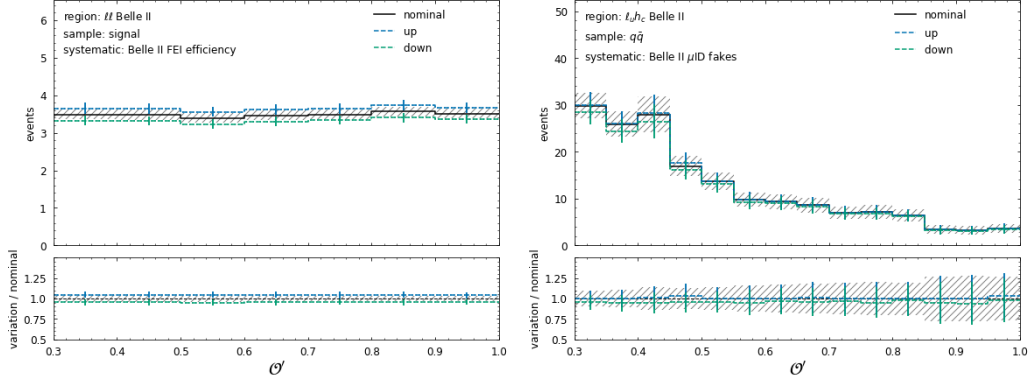


Figure 8.2: Left: Signal template for FEI efficiency modifier in $\ell\ell$ category. Right: $q\bar{q}$ template for μ ID fake rate modifier in $\ell_u h_c$ category (Belle II MC).

If there are both statistical and systematic uncertainty for the correction to the simulation, we treat the statistical part separately and consider it as fully uncorrelated. In order not to create a large amount of nuisance parameters, these uncertainties are combined with the MC statistical uncertainty of the template, thus slightly enlarging it. In fact, the fully uncorrelated shape treatment of such statistical uncertainties would result in $correction \times category \times sample \times bin$ more nuisance parameters (where the *category*, *bin*, *sample* are affected by a given correction). Although, due to a simplicity in implementation, the statistical uncertainties of a specific correction within a specific sample are treated as fully correlated within a given bin, therefore our treatment is slightly more conservative compared to the fully uncorrelated treatment. The inclusion of the statistical uncertainties into the **statererror** modifier is performed by following the steps below:

- A fully-correlated among bins (b) normalization+shape systematic is created for each correction co , same as done for the systematic uncertainties, with the $\sigma_{stat,co,b}$ uncertainties used.
- The $\pm\sigma_{stat,co,b}$ shifts are symmetrized.
- This symmetrical uncertainty is extracted for every bin of every affected sample.
- For each bin, we sum in quadrature the $\sigma_{stat,co,b}$ of all the affected corrections and add the resulting $\sigma_{stat,b}$ to the initial statistical uncertainty $\sigma_{MCstat,b}$ of the sample by a sum in quadrature.
- Then, as the **statererror** modifier is created per category, the per-sample per-bin modified statistical uncertainties are summed in quadrature into the total per-bin statistical uncertainty $\sigma_{Modified,b}$.
- The initial fully correlated normalization+shape systematic modifiers are removed from the model.

This way we keep the same number of nuisance parameters as before the implementation of the statistical uncertainties of the corrections. The impact on the result is validated to be same as if the statistical uncertainties of the corrections would be implemented as fully correlated, while significantly im-

proving the stability of the fit. The effect of the increase of the total statistical uncertainty per bin is illustrated in Fig. 8.3 for the $\ell\ell$ category in Belle II.

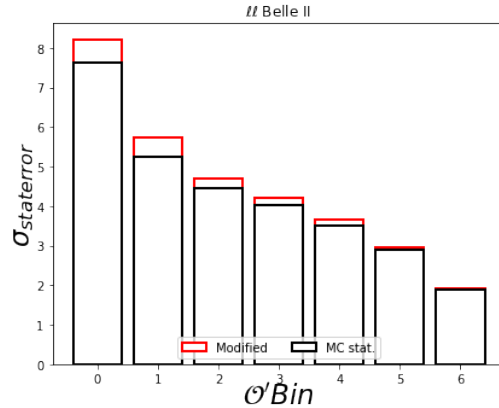


Figure 8.3: In red, the per-bin statistical error in $\ell\ell$ category after the statistical (uncorrelated) part of the efficiency corrections was merged into the total statistical uncertainty per bin. In black, the initial statistical uncertainty per bin, known from MC statistics (Belle II MC).

- **Luminosity** The luminosity collected at the $\Upsilon(4S)$ resonance is $365.4 \pm 1.7\text{fb}^{-1}$ at Belle II [53] and $711 \pm 10\text{fb}^{-1}$ at Belle [54]. The uncertainties on the luminosities are incorporated to the $q\bar{q}$ and $B\bar{B}$ templates as normalization nuisance parameters (one fully correlated parameter for each experiment).
- **$N_{\Upsilon(4S)}$** The number of $B\bar{B}$ pairs produced at the $\Upsilon(4S)$ resonance is $(387 \pm 6) \times 10^6$ at Belle II [55] and $(772 \pm 11) \times 10^6$ at Belle [56]. These uncertainties are taken as normalization uncertainties and treated as fully correlated for the signal templates among all five categories within the Belle or Belle II samples. We do not consider correlations between the $N_{\Upsilon(4S)}$ and the luminosity normalization.
- **$N_{B^0\bar{B}^0}$** The number of neutral B -meson pairs is $N_{B^0\bar{B}^0} = N_{\Upsilon(4S)} \times f^{00}$. The numerical value we consider for f^{00} from the latest HFLAV average [6] is

$$f^{00} = 0.4861 \pm 0.0080$$

As mentioned in Eq. 8.1, the central values of $N_{\Upsilon(4S)}$ and f^{00} are used to deduce the expected number of signal events for a given branching fraction, here set to 10^{-3} . The uncertainties (we sum in quadrature the statistical and the systematic terms) are propagated to the signal and $B\bar{B}$ templates. While for signal the uncertainty on the number of B -pairs can be treated as simple normalization term, for the $B\bar{B}$ templates the underlying B^+B^- and $B^0\bar{B}^0$ events undergo the same fractional variation, but opposite in sign, resulting in an overall change of the template shape. Nonetheless, to fully correlate both signal and the $B^0\bar{B}^0$ part of the $B\bar{B}$ template, the modifier for signal is also a normalization+shape, but with no shape variation.

- **χ_d** The amount of mixed events changes as a function of the BDT cut and is different for each fit category. The events are generated with a mixing rate χ_d

that is compatible with the PDG value for Belle II samples and *almost* compatible for Belle ($\sim 2\sigma$ discrepancy). Therefore, we decide to use the uncertainty on the current PDG best-fit value of χ_d as a systematic uncertainty to be applied to all B^0 events ($B\bar{B}$ and signal templates) as a shape+normalization uncertainty (e.g. if the fraction of mixed events goes up, then fraction of unmixed events goes down).

- **$q\bar{q}$ normalization** For each fit category, the data/MC yield ratio of the off-resonance events belonging to the region $\mathcal{O}' > 0.3$ is computed; the central value is used to re-scale the $q\bar{q}$ overall yield and the (statistical) uncertainty is used as normalization systematic uncertainty for the $q\bar{q}$ fit template.
- **Combinatorial B_{tag} normalization** For each fit category, a 50% “normalization” systematic uncertainty is assigned to the $B\bar{B}$ template’s events with non-peaking B_{tag} ’s. Because this nuisance affects only part of the template, it is implemented as normalization+shape (and not pure normalization, despite being called such). The justification for this systematic is discussed in Section 7.1. The fractions of events with the non-peaking B_{tag} ’s within the $B\bar{B}$ template, are:
 - **Belle:** 14% ($\ell\ell$), 28% ($\ell_c h_u$), 29% ($\ell_u h_c$), 35% ($\rho\ell$), 46% ($\ell 0$)
 - **Belle II:** 19% ($\ell\ell$), 30% ($\ell_c h_u$), 30% ($\ell_u h_c$), 37% ($\rho\ell$), 45% ($\ell 0$)
- **Signal modelling** The BTOSLLBALL model in `evtGen`, used to generate the signal events, is based on the $B \rightarrow K$ form factors (FF) computed in [57] using the light cone QCD sum rule approach. In the default configuration, the set of values are those providing the loosest bounds on the $\mathcal{C}_9$ and $\mathcal{C}_{10}$ Wilson coefficients. As an alternative signal model, we use the SM predictions from the `flavio` package [58], which relies on different form-factor parameterizations and Wilson-coefficient inputs [59, 60]. We have checked that the total integrated BF for $B^0 \rightarrow K_S^0 \tau \tau$ obtained with `flavio` is consistent with the SM prediction computed in [61]. The binned differences in the q_{true}^2 distribution between the *nominal* (BTOSLLBALL) and the *alternative* model provided by `flavio` (see left plot of Fig. 8.4) are used to establish event-based weights w :

$$w_i = \frac{d\Gamma_{\text{alt}}/dq_i^2 - d\Gamma_{\text{nom}}/dq_i^2}{d\Gamma_{\text{nom}}/dq_i^2}.$$

The right plot of Fig. 8.4 shows an example of how the signal template varies if one considers the alternative model. The modifier type used for such template is a one-sided normalization+shape. The fact that at high $\mathcal{O}'$ values we get slightly higher signal yield, when using the alternative model of the left plot of Fig. 8.4, can be easily explained: the BDT favors lower q^2 values (because backgrounds tend to have larger values).

- **Uncertainties on the $B\bar{B}$ background decays** While for the B_{tag} side the data/MC efficiency differences are corrected and the systematics uncertainties computed, for the B_{sig} side there can be additional mismatches we want to address. For example, the B modes selected in our $B\bar{B}$ background and belonging to the B_{sig} side can have generated branching fractions (BF) that significantly differ from the most recent PDG/HFLAV averages. Even when

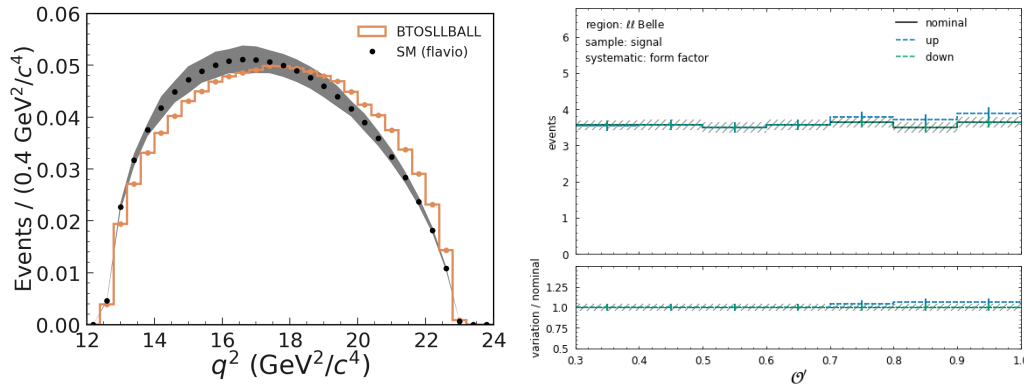


Figure 8.4: Systematic uncertainty related to the signal model used for the selection optimization. On the left, the default model (BTOSLLBALL) is compared to the alternative one (SM `flavio`). The q^2 -binned event-by-event weight resulting by comparing the two distribution is propagated into the template modifiers whose envelope is represented in the right plot for the Belle $\ell\ell$ category.

the generated BFs are up to date, the uncertainties on the averages and their impact on the $B\bar{B}$ templates shapes should be taken into account. As described in Section 6.3 (and summarized in Tables 6.2 and 6.3), a large fraction of $B\bar{B}$ background events has a semileptonic decay on the signal side. This fraction ranges from $\sim 90\%$ for the $\ell\ell$ category to $\sim 40\%$ for the $\ell 0$ category. Besides the well known $B \rightarrow D^{(*)}\ell\nu$ and $B \rightarrow D^{(*)}\tau\nu$ decays, we observe a non-negligible component of $B \rightarrow D^{**}\ell\nu$, which are less precisely measured. Similarly, among the hadronic B_{sig} decay modes, many of them have not been measured or have poor precision. The general strategy is the following: we correct the number of events entering the signal region by applying weights w_{BF} equal to

$$w_{\text{BF}} = BF_{\text{meas}}/BF_{\text{MC}}.$$

where BF_{meas} uses either the PDG [3] or the HFLAV [62] averages depending on the mode. Once the known branching fractions are corrected, we must also propagate their uncertainties $\sigma(BF_{\text{meas}})$; for the unmeasured modes, we assume a 100% uncertainty. This approach is rather conservative and results in large uncertainty on the final result. To get the variations, an ensemble of 5000 toy experiments following a lognormal distribution ($\ln N(w_{\text{BF}}, \sigma(BF_{\text{meas}}))$) is generated to get the correlation and the covariance matrices. As we treat the PDG/HFLAV averages as a single auxiliary measurement, the toy weight applied to the BF is fully correlated between the Belle and Belle II datasets for each mode within a toy replica. Five toy replicas are shown in Fig. 8.5, where the spread in the bin yields roughly represents the uncertainty on the $B\bar{B}$ template due to this systematic source. The total dimensionality of such matrices are 140×140 , dictated by the 70(Belle)+70(Belle II) bins in $\mathcal{O}'$.

To capture the uncertainty obtained via toy variation, we take all eigenvectors from the obtained covariance matrix and incorporate them as yield and shape modifiers via 140 (70 for Belle or Belle II disjoint results) nuisance parameters into the fit model.

- **Uncertainty on K_L production modeling** Following Ref. [18], we consider

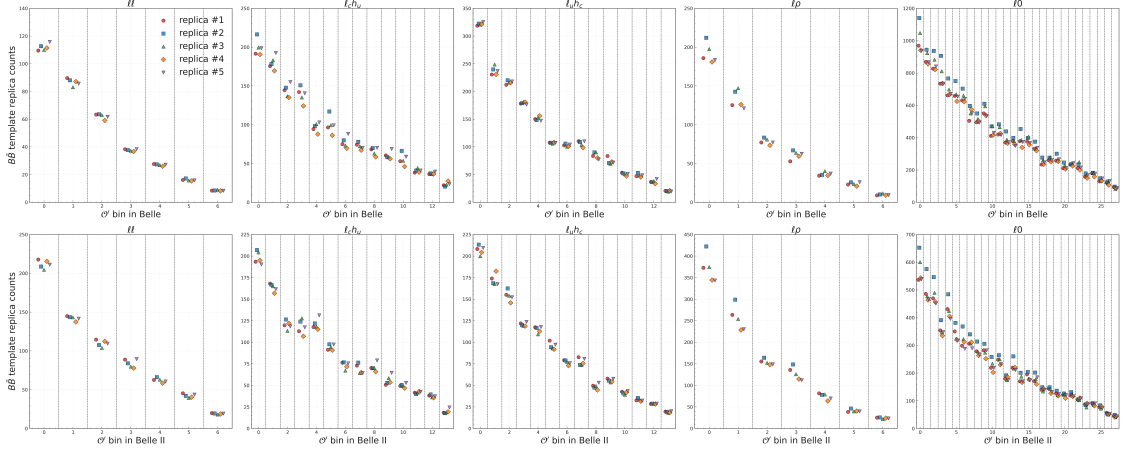


Figure 8.5: Five toy replicas of the $B\bar{B}$ template in bins of $\mathcal{O}'$ in the fit region.

a large uncertainty on the K_L inclusive production. As we do not perform a dedicated measurement ourselves to investigate this component directly, we do not upscale the $D \rightarrow K_L^0$ component, instead, we assign a $\pm 40\%$ normalization uncertainty on this component in B backgrounds (separately in Belle and Belle II) to cover both the quoted central value and its uncertainty. We do not allow this component to additionally float in $q\bar{q}$, as our scaling factor obtained from the off-resonance sample corrects also the data/MC differences induced by $D \rightarrow K_L^0$ modeling. Therefore, the systematic uncertainty is assigned to $B \rightarrow D \rightarrow K_L^0$ inclusive events. This uncertainty is treated as uncorrelated for the Belle and Belle II, while within the experiment, this systematic is implemented as fully correlated normalization+shape.

- BDT selection efficiency uncertainty** An uncertainty associated with the signal efficiency after the BDT selection is derived using the modified $D(K_S^0 t_c)$ -vetoed region, described in Sec. 7.2. In this region, the data/MC scale factors can be measured with a statistical precision of about 10%, which is taken as the corresponding systematic uncertainty. The calibration is performed only in two of the five fit categories, $\ell_u h_c$ and $\ell_u \rho_c$, where the $B^0 \bar{B}^0$ purity is high, approximately 98% for $t_c = \pi$ and 90% for $t_c = \rho$. For this reason, the uncertainty is treated as fully correlated among all five fit categories within the same experiment. The uncertainties assigned to Belle and Belle II are treated as uncorrelated and affect signal normalization.
- τ Branching Fraction uncertainties** As the τ decay mode composition directly impacts the signal shape and efficiency, we assign the uncertainty from PDG as a systematic uncertainty for each considered mode. Such systematic is implemented as normsys+histosys for each weighted τ decay mode (10 nuisance parameters). As a reminder, the most abundant τ branching fractions are fixed to PDG values within the signal template.

The comparison of pre- and post-cut signal efficiencies and (normalized) background yields is shown in Tab. 8.1. The normalization factors for the $q\bar{q}$ and combinatorial B_{tag} components for the pre-cut case are “global” as described in Paragraphs 5.2.1 and 5.2.2. For the post-cut case the $q\bar{q}$ weights are derived using the

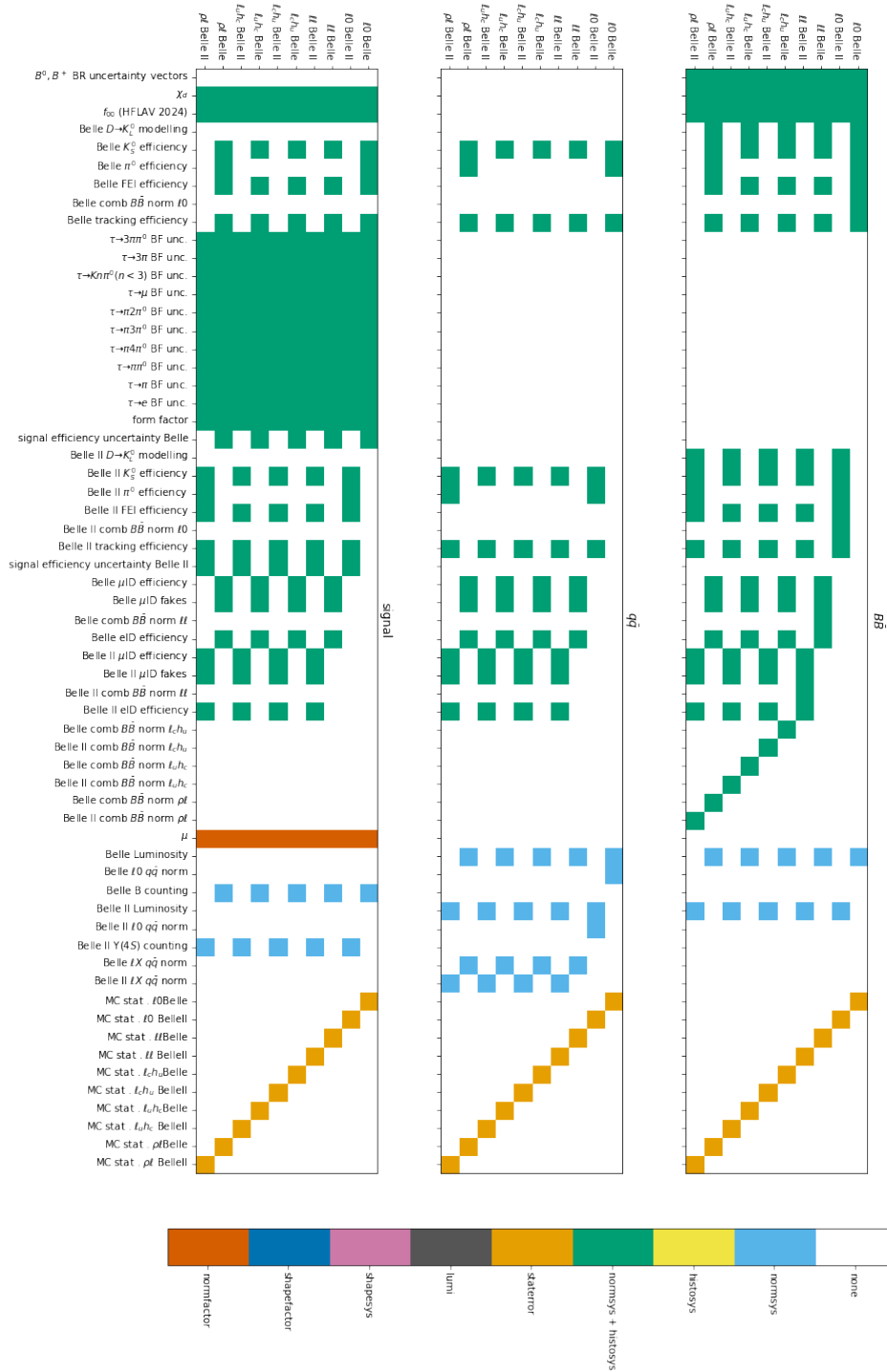


Figure 8.6: Modifier grid showing the systematic uncertainties, their type and the affected samples and categories.

the same procedure, but after the BDT cut is applied and explained in Section 7.1. For the combinatorial B_{tag} no scaling is applied for the post-cut case to be representative of the template yields. Additionally, for the post-cut columns the branching fractions of the most abundant known background decays are weighted in MC to match the latest PDG/HFLAV averages as described in this section.

Table 8.1: Signal efficiencies and expected background yields for Belle II MC both before and after the cut $\mathcal{O}' > 0.3$ is applied. In both cases the reconstruction efficiencies, fake rates and normalization are corrected for data/MC discrepancies. The pre-cut values are given with the normalization corrections obtained in the full $\mathcal{O}'$ region (global) as described in Sections 5.2.1 and 5.2.2, while the post-cut values are shown with separate combinatorial $q\bar{q}$ normalizations, obtained in the fit region. There is no combinatorial B_{tag} normalization factor for the post-cut case to be representative of the yields of the templates. The uncertainties are not shown as they are entangled through complicated systematic and statistical effects discussed in this section.

	Signal eff. (10^{-4})		$q\bar{q}$ yields (100 fb^{-1})		$B\bar{B}$ yields (100 fb^{-1})	
	pre-cut	post-cut	pre-cut	post-cut	pre-cut	post-cut
$\ell\ell$	0.97	0.68	44	6	925	165
ℓ_ch_u	1.54	1.08	293	40	1308	278
ℓ_uh_c	1.48	1.04	273	48	2044	318
$\rho\ell$	1.05	0.74	376	41	1443	216
$\ell 0$	4.32	3.03	5542	567	7324	1295

Table 8.2: Signal efficiencies and expected background yields for Belle MC both before and after the cut $\mathcal{O}' > 0.3$ is applied. In both cases the reconstruction efficiencies, fake rates and normalization are corrected for data/MC discrepancies. The pre-cut values are given with the normalization corrections obtained in the full $\mathcal{O}'$ region (global) as described in Sections 5.2.1 and 5.2.2, while the post-cut values are shown with separate combinatorial $q\bar{q}$ normalizations, obtained in the fit region. There is no combinatorial B_{tag} normalization factor for the post-cut case to be representative of the yields of the templates. The uncertainties are not shown as they are entangled through complicated systematic and statistical effects discussed in this section.

	Signal eff. (10^{-4})		$q\bar{q}$ yields (100 fb^{-1})		$B\bar{B}$ yields (100 fb^{-1})	
	pre-cut	post-cut	pre-cut	post-cut	pre-cut	post-cut
$\ell\ell$	0.48	0.34	8	0.5	433	61
ℓ_ch_u	1.14	0.80	90	13	878	195
ℓ_uh_c	1.17	0.82	93	14	1681	281
$\rho\ell$	0.44	0.31	84	7	578	81
$\ell 0$	4.03	2.82	5344	593	7333	1808

8.3 Results on simulation

To get the result on MC and validate our fitting procedure we perform the fit of our model to the Asimov data.

One of the common health checks for the fit performance is the pull plot of the fit to Asimov data. A pull is computed by calculating the difference between the fitted parameter value ($\hat{\theta}$), the initial value (θ_0), and dividing it by the width of that constrained parameter ($\Delta\theta$) as shown in the equation:

$$pull = \frac{\hat{\theta} - \theta_0}{\Delta\theta} . \quad (8.4)$$

The pull plot for the combined fit to the Asimov data can be seen in Fig. 8.7. As expected, all the nuisance parameters in the fit give zero pulls after the fit to the Asimov data. The errors on the pulls do not deviate much from the nominal 1σ constraints, although the errors on the uncertainties derived with the most conservative approaches are slightly shrank after the fit. The post-fit correlations of the nuisance parameters, obtained from the fit to the Asimov dataset are shown in Fig. 8.8.

It can be noticed that some systematics have rather high (anti-)correlations between each other. In particular, a high anticorrelation is seen between the $D \rightarrow K_L^0$ modeling uncertainty and the normalization uncertainties assigned to the combinatorial B_{tag} components in the $B\bar{B}$ templates. This arises from the relatively similar shape of these components and the comparable size of their associated uncertainties. Moderate anticorrelations are also observed between the first two leading eigenvectors from the branching fraction (BF) uncertainty decomposition. This behavior indicates some level of mutual compensation between these parameters, which is expected due to the construction of eigenvectors. In addition, some anticorrelations are found between various systematic sources and the leading BF uncertainty vector. This is attributed to their entanglement through the dominant $B \rightarrow D$ background contributions, which result in similar shape variations. On the other hand, high correlations are observed in combinatorial B_{tag} $B\bar{B}$ normalization systematics within a given experiment, but show no correlation to B_{tag} $B\bar{B}$ normalization systematics from the other experiment as expected. This is an expected behavior, as each normalization uncertainty covers the same part of the template and is allowed to fluctuate by the same amount. MC statistical uncertainties, as anticipated, generally show weak correlations. Rare exceptions occur when normalization systematics or certain efficiency corrections mimic similar behavior in specific bins. These are uncorrelated statistical uncertainties, hence their general independence.

The impact of the single systematic uncertainties on the POI is computed by repeating the fit to the Asimov while varying the nominal value of the nuisance parameter by its systematic uncertainty as explained in [63, 64]. The nuisance parameter is not held fixed, as would be done by using the method implemented in `cabinetry`, as fixing the parameter would modify the shape of the likelihood. The deviation of μ from 0 for a given fit with a shifted systematic represents the impact of that systematic source on the fit observable. The values of the “up” and “down” impacts on the signal strength are averaged. For the MC statistical uncertainty, the values of impact on POI are summed in quadrature across all bins. The total uncertainty is again represented by a sum in quadrature of the impacts on POI from every systematic. The numeric values for the combined fit are shown in the Table 8.3.

We check for possible fit biases by fitting our model to an ensemble of toy Asimov datasets, sampled with a Poisson distribution. A 10,000 toy samples are used for each signal strength hypothesis. If the central value and the uncertainty of the parameter of interest μ is correctly estimated by the fitter, the distribution of the fitted μ is expected to be centered at the injected μ hypothesis with a width corresponding to the statistical uncertainty. The distribution of the fitted μ for four signal strength hypotheses is shown in Fig. 8.9, where the median is utilized to estimate the fit bias and 68% quantiles around the median are used to estimate the statistical uncertainty ($\sigma_{\pm}$). The fit shows almost no bias ($\mathcal{O} \lesssim 1\%$), which is also influenced by the number of toys being used. Due to the toy precision, to estimate the statistical

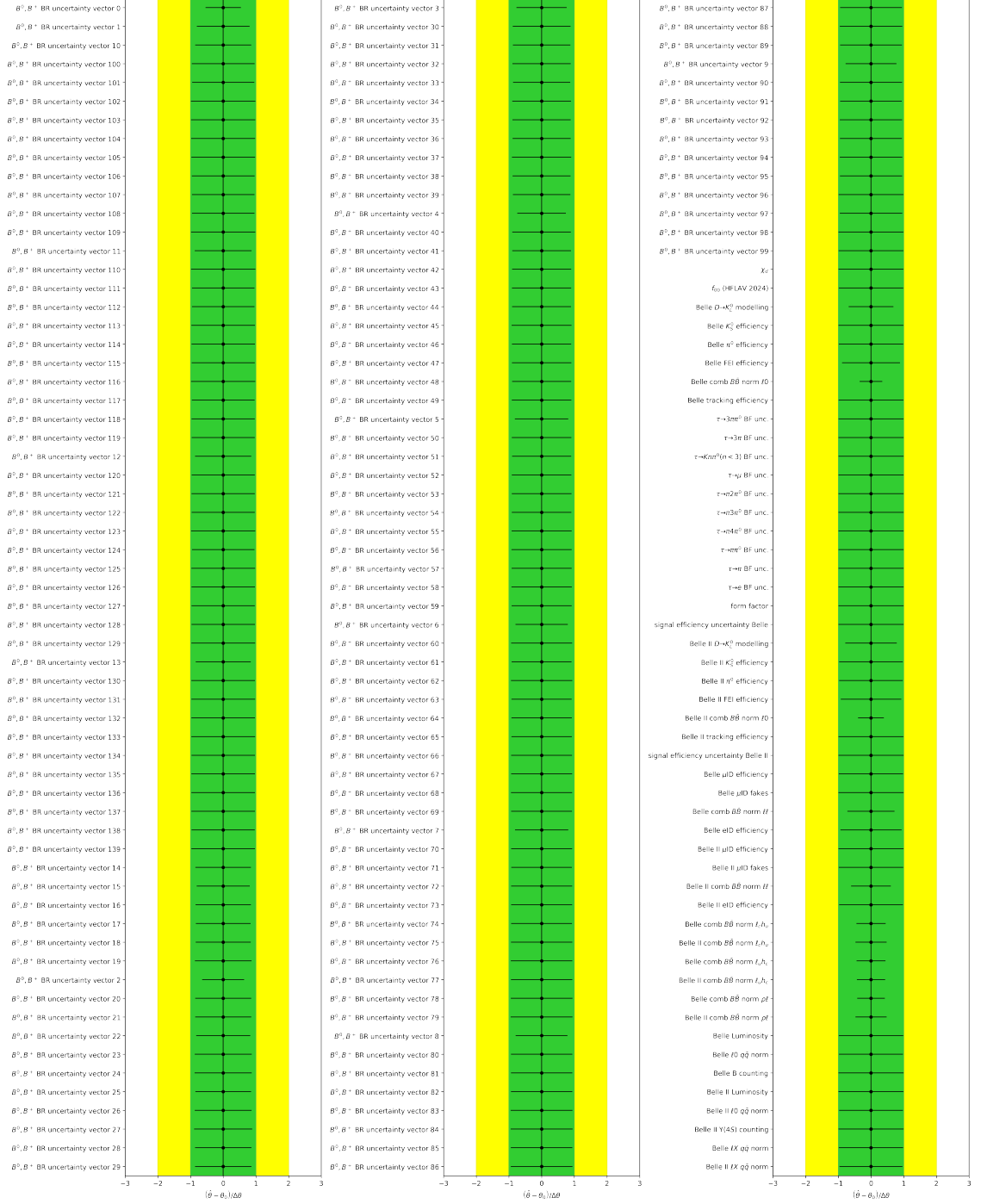


Figure 8.7: Pull plot from the fit to the Asimov dataset.

uncertainty we prefer to subtract the systematic uncertainty in quadrature from the best-fit value of μ , while the result of such operation is almost compatible with the σ_+ and the σ_- , obtained from the toy Asimov data. The fact that the values are not entirely compatible is a consequence of: 1) number of toys precision 2) symmetrized uncertainty of the systematics estimate 3) the distribution of the fitted μ is not entirely Gaussian itself and is slightly asymmetric due to the Poissonian sampling of the bins with low yields. Therefore some information is lost due to the symmetrization of statistical and systematical uncertainties on μ . Same tests as mentioned here are performed for the disjoint fits of Belle and Belle II; they can be

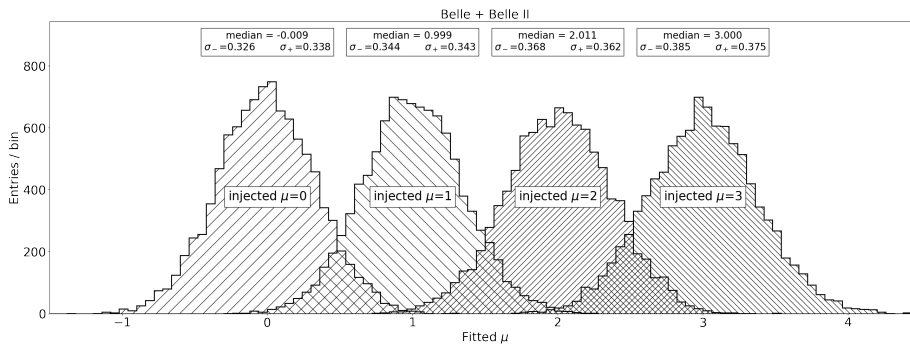
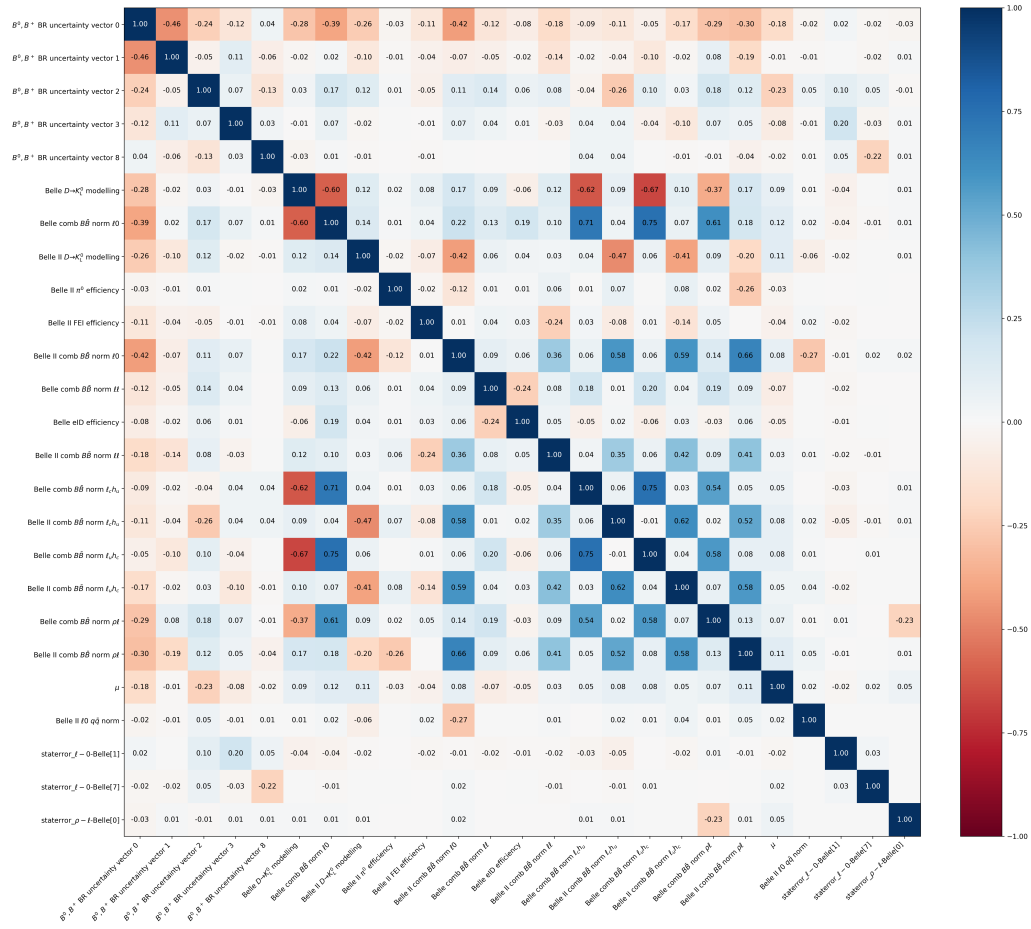


Table 8.3: Uncertainty on fitted signal strength μ for the combined fit to the Asimov data, ranked from highest to lowest.

Source	$\Delta\mu$	Source	σ_μ
B branching fraction uncertainties	0.20	MC stat	0.17
Belle FEI efficiency	0.05	Belle II $D \rightarrow K_L^0$ modelling	0.04
f_{00} (HFLAV 2024)	0.03	Belle $D \rightarrow K_L^0$ modelling	0.02
Belle II FEI efficiency	0.02	Belle II comb. $B\bar{B}$ norm. $\rho\ell$	0.02
Belle eID efficiency	0.02	Belle comb. $B\bar{B}$ norm. $\ell 0$	0.02
Belle II comb. $B\bar{B}$ norm. $\ell_c h_u$	0.02	Belle II comb. $B\bar{B}$ norm. $\ell 0$	0.02
Belle comb. $B\bar{B}$ norm. $\ell\ell$	0.02	Belle comb. $B\bar{B}$ norm. $\ell_u h_c$	0.02
Belle comb. $B\bar{B}$ norm. $\rho\ell$	0.01	Belle II π^0 efficiency	0.01
Belle II eID efficiency	0.01	Belle comb. $B\bar{B}$ norm. $\ell_c h_u$	0.01
Belle II comb. $B\bar{B}$ norm. $\ell_u h_c$	0.01	Belle II $\ell 0$ $q\bar{q}$ norm.	0.01
Belle luminosity	0.01	Belle II comb. $B\bar{B}$ norm. $\ell\ell$	0.01
Belle μ ID efficiency	0.01	Belle II μ ID fakes	0.01
Belle II ℓX $q\bar{q}$ norm.	< 0.01	Belle ℓX $q\bar{q}$ norm.	< 0.01
Belle II K_S^0 efficiency	< 0.01	Belle K_S^0 efficiency	< 0.01
Belle tracking efficiency	< 0.01	Belle π^0 efficiency	< 0.01
Belle $\ell 0$ $q\bar{q}$ norm.	< 0.01	χ_d	< 0.01
Belle II μ ID efficiency	< 0.01	Belle II luminosity	< 0.01
Belle μ ID fakes	< 0.01	Belle II tracking efficiency	< 0.01
$\tau \rightarrow \pi 2\pi^0$ BF unc.	< 0.01	Signal efficiency uncertainty Belle II	< 0.01
Belle II $\Upsilon(4S)$ counting	< 0.01	$\tau \rightarrow \pi 3\pi^0$ BF unc.	< 0.01
Signal efficiency uncertainty Belle	< 0.01	$\tau \rightarrow 3\pi\pi^0$ BF unc.	< 0.01
Belle B counting	< 0.01	$\tau \rightarrow \pi 4\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \mu$ BF unc.	< 0.01	$\tau \rightarrow e$ BF unc.	< 0.01
$\tau \rightarrow K n \pi^0$ ($n < 3$) BF unc.	< 0.01	$\tau \rightarrow 3\pi$ BF unc.	< 0.01
$\tau \rightarrow \pi\pi^0$ BF unc.	< 0.01	$\tau \rightarrow \pi$ BF unc.	< 0.01
Form factor	< 0.01		
Total		0.29	

is shown in Fig. 8.10

The simultaneous fit to the 10 (5 in Belle and 5 in Belle II) fit categories leads to the results on the fitted parameter of interest μ shown in Table 8.4, together with the Belle and Belle II disjoint fit results. The result of the branching fraction measurement with the separate impacts of the statistical and systematic uncertainties are shown in Table 8.5. The quoted systematic uncertainty is taken from the Table 8.3 (or a corresponding table for a disjoint fit from Appendix I), while the contribution of the statistical uncertainty is computed as a difference in quadrature between the total uncertainty (obtained from the fit) and the systematic uncertainty.

 Table 8.4: Results based on Belle ($\int \mathcal{L} dt = 711\text{fb}^{-1}$) and Belle II ($\int \mathcal{L} dt = 365\text{fb}^{-1}$) MC samples with the background-only hypothesis.

Sample	μ	UL @ 90% CL
Belle	0.00 ± 0.55	0.95×10^{-3}
Belle II	0.00 ± 0.66	1.12×10^{-3}
Combined	0.00 ± 0.44	0.74×10^{-3}

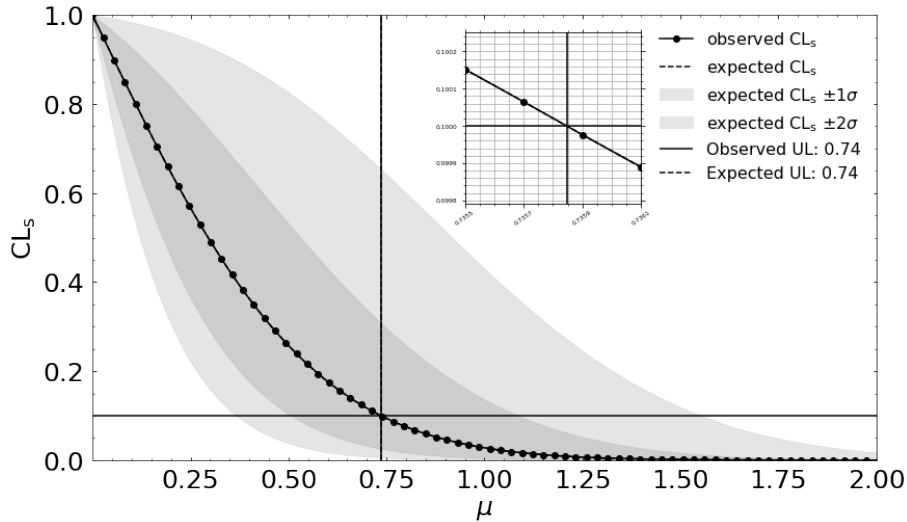

 Figure 8.10: CL_s scan from the fit to the Asimov dataset.

 Table 8.5: Branching fraction measurement results based on Belle ($\int \mathcal{L} dt = 711 \text{fb}^{-1}$) and Belle II ($\int \mathcal{L} dt = 365 \text{fb}^{-1}$) MC samples with the background-only hypothesis.

Sample	$\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) (10^{-3})$
Belle	$0.00 \pm 0.41(\text{stat.}) \pm 0.36(\text{syst.})$
Belle II	$0.00 \pm 0.49(\text{stat.}) \pm 0.43(\text{syst.})$
Combined	$0.00 \pm 0.33(\text{stat.}) \pm 0.29(\text{syst.})$

One can notice that the statistical uncertainty for Belle and Belle II scale almost correctly with the square root of the luminosity $(0.49/0.41) \times \sqrt{365/711} \approx 0.86 \sim 1$. The mismatch is driven by the differences in reconstruction efficiencies between Belle and Belle II and the separate BDT trainings.

8.4 Results on data

8.4.1 Measurement in control sample

We validate the efficiency in simulation and the fit procedure using semileptonic B^0 decays with $D^- \rightarrow K_S^0 \pi^-$ or $K_S^0 \rho^-$, selected within the vetoed region of $M(K_S^0 t_c)$. First, we fit $N_{\text{data}}(D^{(*)-})/N_{\text{MC}}(D^{(*)-})$ in $B^0 \rightarrow D^{(*)-} \ell^+ \nu$ events and measure 1.09 ± 0.11 (1.04 ± 0.14), thereby confirming good agreement between data and simulation.

8.4.2 Belle II partial result

The result of the fit to Belle II data shows satisfactory agreement between the data and the model. No significant signal is observed. The measured branching fraction is found to be $\mathcal{B}_{\text{Belle II data}}(B^0 \rightarrow K_S^0 \tau \tau) = [0.20 \pm 0.50(\text{stat.}) \pm 0.44(\text{syst.})] \times 10^{-3}$. The result is compatible with the MC expectation. The upper limit on the $\mathcal{B}_{\text{Belle II data}}(B^0 \rightarrow K_S^0 \tau \tau)$ is found to be 1.31×10^{-3} at 90% confidence level. The pull plot of the Belle II only fit is shown in Fig. 8.11. The contributions of all systematic sources are summarized in Tab. 8.6.

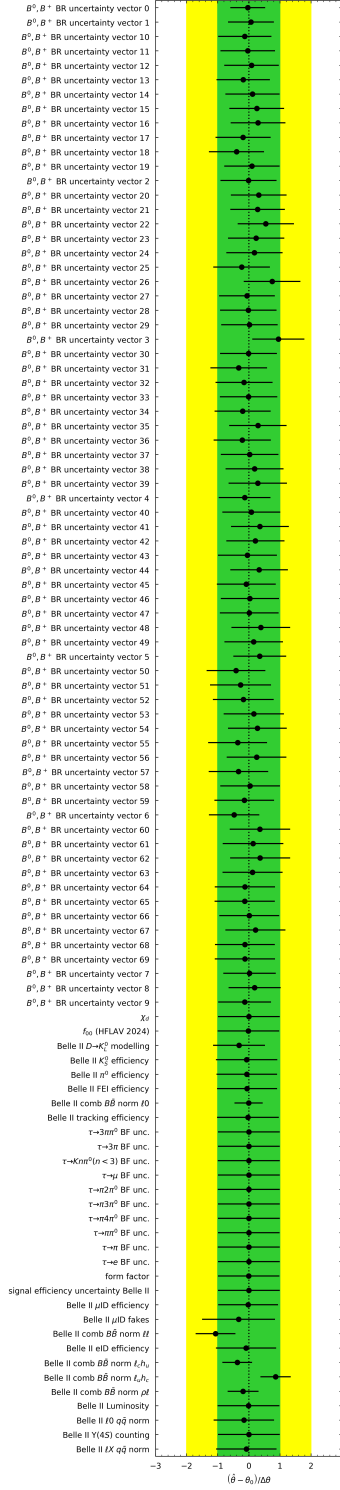


Figure 8.11: Pull plot from the fit to Belle II data only.

Table 8.6: Uncertainty on fitted signal strength μ for the Belle II-only fit.

Source	σ_μ
B BF uncertainties	0.35
MC stat	0.23
Belle II FEI efficiency	0.09
Belle II $D \rightarrow K_L^0$ modelling	0.07
Belle II eID efficiency	0.04
f_{00} (HFLAV 2024)	0.04
Belle II π^0 efficiency	0.04
Belle II K_S^0 efficiency	0.03
Belle II comb $B\bar{B}$ norm $\rho\ell$	0.03
Belle II $\ell 0$ $q\bar{q}$ norm	0.03
signal efficiency uncertainty Belle II	0.02
Belle II comb $B\bar{B}$ norm $\ell_u h_c$	0.02
Belle II comb $B\bar{B}$ norm $\ell 0$	0.02
Belle II comb $B\bar{B}$ norm $\ell\ell$	0.02
Belle II μ ID fakes	0.02
form factor	0.01
Belle II ℓX $q\bar{q}$ norm	0.01
Belle II Luminosity	0.01
Belle II tracking efficiency	0.01
Belle II comb $B\bar{B}$ norm $\ell_c h_u$	< 0.01
Belle II μ ID efficiency	< 0.01
Belle II $\Upsilon(4S)$ counting	< 0.01
$\tau \rightarrow Kn\pi^0 (n < 3)$ BF unc.	< 0.01
$\tau \rightarrow \pi\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi$ BF unc.	< 0.01
$\tau \rightarrow \pi 4\pi^0$ BF unc.	< 0.01
$\tau \rightarrow 3\pi\pi^0$ BF unc.	< 0.01
$\tau \rightarrow 3\pi$ BF unc.	< 0.01
$\tau \rightarrow K\pi\pi^0 (n < 3)$ BF unc.	< 0.01
$\tau \rightarrow \eta$ BF unc.	< 0.01
$\tau \rightarrow \pi 2\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi 3\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi 4\pi^0$ BF unc.	< 0.01
$\tau \rightarrow m^0$ BF unc.	< 0.01
$\tau \rightarrow \pi$ BF unc.	< 0.01
$\tau \rightarrow e$ BF unc.	< 0.01
form factor	< 0.01
signal efficiency uncertainty Belle II	< 0.01
Belle II μ ID efficiency	< 0.01
Belle II μ ID fakes	< 0.01
Belle II comb $B\bar{B}$ norm $\ell\ell$	< 0.01
Belle II eID efficiency	< 0.01
Belle II comb $B\bar{B}$ norm $\ell_u h_c$	< 0.01
Belle II comb $B\bar{B}$ norm $\rho\ell$	< 0.01
Belle II Luminosity	< 0.01
Belle II f_0 $q\bar{q}$ norm	< 0.01
Belle II ℓX $q\bar{q}$ norm	< 0.01
χ_d	< 0.01
Total	0.44

8.4.3 Belle partial result

The result of the fit to Belle data shows satisfactory agreement between the data and the model. No significant signal is observed. The measured branching fraction is found to be $\mathcal{B}_{\text{Belle data}}(B^0 \rightarrow K_S^0 \tau \tau) = [0.07 \pm 0.42(\text{stat.}) \pm 0.36(\text{syst.})] \times 10^{-3}$. The re-

sult is compatible with the MC expectation. The upper limit on the $\mathcal{B}_{\text{Belle data}}(B^0 \rightarrow K_S^0 \tau \tau)$ is found to be 1.01×10^{-3} at 90% confidence level. The pull plot of the Belle only fit is shown in Fig. 8.12. The contributions of all systematic sources are summarized in Tab. 8.7.

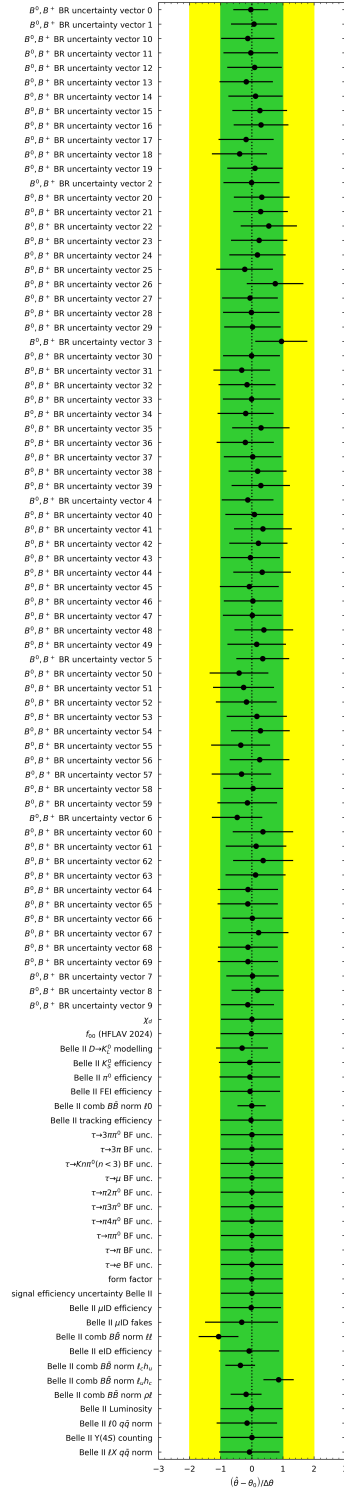


Figure 8.12: Pull plot from the fit to Belle data only.

Table 8.7: Uncertainty on fitted signal strength μ for the Belle-only fit.

Source	σ_μ
MC stat	0.24
B BF uncertainties	0.23
Belle FEI efficiency	0.09
Belle comb $B\bar{B}$ norm $\ell\ell$	0.05
Belle $D \rightarrow K_L^0$ modelling	0.04
Belle comb $B\bar{B}$ norm $\ell 0$	0.03
Belle comb $B\bar{B}$ norm $\rho\ell$	0.03
Belle eID efficiency	0.03
f_{00} (HFLAV 2024)	0.03
Belle comb $B\bar{B}$ norm $\ell_u h_c$	0.02
Belle comb $B\bar{B}$ norm $\ell_c h_u$	0.02
Belle Luminosity	0.02
Belle π^0 efficiency	0.02
Belle μ ID efficiency	0.02
form factor	0.01
Belle ℓX $q\bar{q}$ norm	0.01
Belle tracking efficiency	0.01
Belle K_S^0 efficiency	0.01
signal efficiency uncertainty Belle	0.01
$\tau \rightarrow \pi 2\pi^0$ BF unc.	0.01
χ_d	0.01
Belle B counting	< 0.01
Belle $\ell 0$ $q\bar{q}$ norm	< 0.01
$\tau \rightarrow 3\pi$ BF unc.	< 0.01
$\tau \rightarrow e$ BF unc.	< 0.01
$\tau \rightarrow 3\pi\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi 3\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi 4\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi$ BF unc.	< 0.01
$\tau \rightarrow \mu$ BF unc.	< 0.01
$\tau \rightarrow \pi\pi^0$ BF unc.	< 0.01
$\tau \rightarrow K n \pi^0 (n < 3)$ BF unc.	< 0.01
Belle μ ID fakes	< 0.01
Total	0.36

8.4.4 Final result

The result of the simultaneous fit to the full available data yields no observation of significant signal. Already at the pre-fit stage the agreement between data and model is good as shown in Fig. 8.13. The distributions of the fit variable after the fit is performed are shown in Fig. 8.14. A selection of approximations of the post-fit distributions of the BDT input features is summarized in Appendix J. The measured branching fraction is found to be

$$\mathcal{B}(B^0 \rightarrow K_S^0 \tau \tau) = [0.12 \pm 0.33(\text{stat.}) \pm 0.29(\text{syst.})] \times 10^{-3}.$$

The result is compatible with the MC expectation. It is worth noting that the central value of the $\mathcal{B}(B^0 \rightarrow K_S^0 \tau \tau)$ from the simultaneous fit does not lay in between the disjoint Belle and Belle II results, with the reason for this being different fit models used for the three fits and inclusion of correlations of the nuisance parameters for the models of the two datasets. The upper limit is set to be

$$\mathcal{B}_{UL}(B^0 \rightarrow K_S^0 \tau \tau) < 0.83 \times 10^{-3}$$

at 90% confidence level. The CL_s scan plot is shown in Fig. 8.15. The pull plot of the full simultaneous fit is shown in Fig. 8.16. The contributions of all systematic sources are summarized in Tab. 8.8.

Table 8.8: Sources of systematic uncertainties, their treatment in the fit and their contribution to the uncertainty on the signal strength μ .

Source	Uncertainty type, parameters	σ_μ
Leading B -decay branching fractions	Correlated shape, 140	0.20
MC statistics	Uncorrelated shape, 140	0.18
B -tagging efficiency	Correlated shape, 2	0.06
Combinatorial B_{tag}^0 normalization	Correlated shape, 10	0.05
$D \rightarrow K_L^0$ modeling	Correlated shape, 2	0.05
Lepton ID	Correlated shape, 6	0.03
f_{00}	Correlated shape, 1	0.02
Signal efficiency uncertainty	Normalization, 2	0.02
π^0 efficiency	Correlated shape, 2	0.02
$q\bar{q}$ normalization	Normalization, 4	0.02
Signal model	Correlated shape, 1	0.01
K_S^0 efficiency	Correlated shape, 2	< 0.01
Number of $\Upsilon(4S)$	Normalization, 2	< 0.01
Luminosity uncertainty	Normalization, 2	< 0.01
Tracking efficiency	Correlated shape, 2	< 0.01
Leading τ -decay branching fractions	Correlated shape, 10	< 0.01
χ_d	Correlated shape, 1	< 0.01
Total		0.29

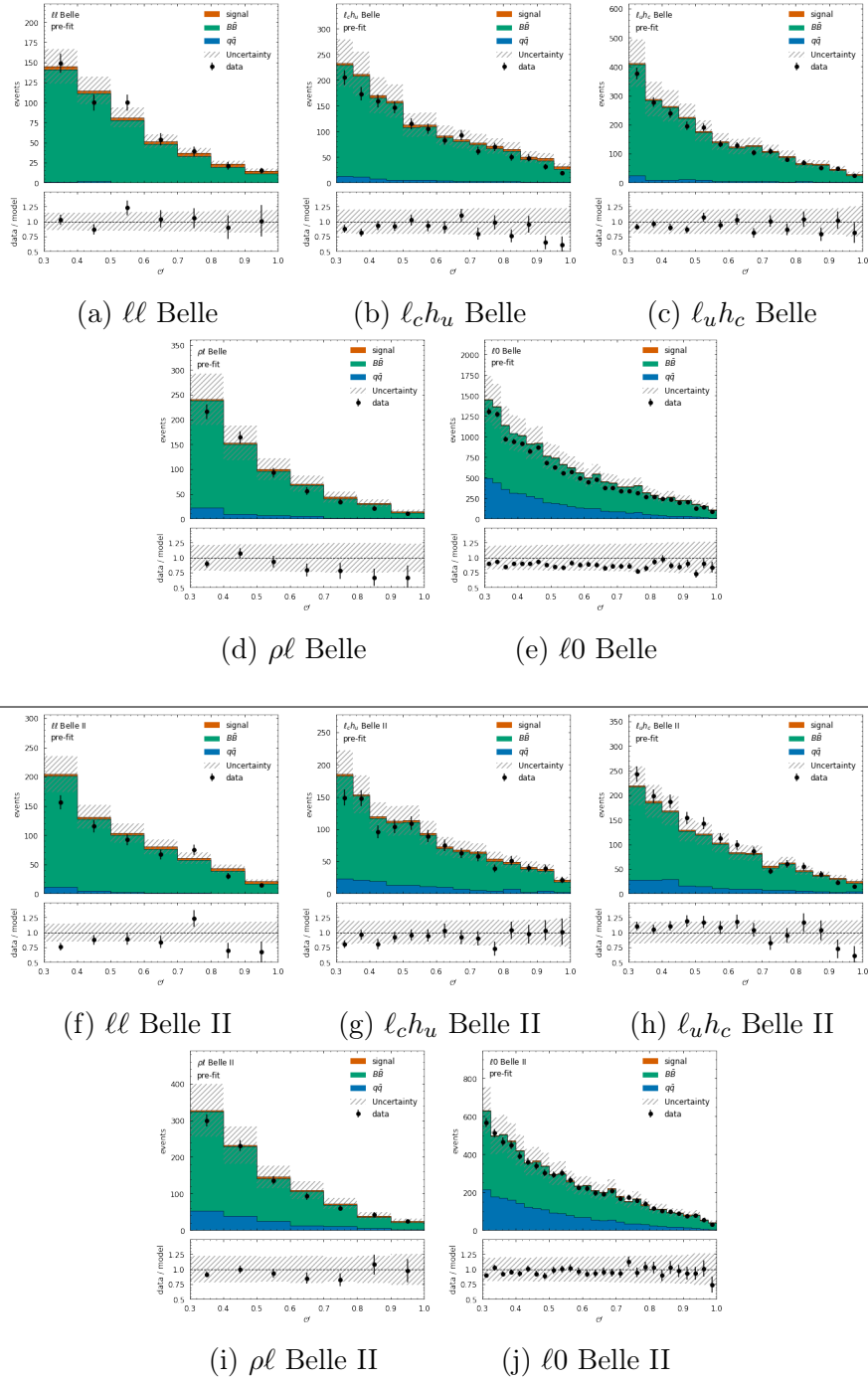


Figure 8.13: Pre-fit distributions of the BDT output $\mathcal{O}'$ in Belle (upper 5 plots) and Belle II (lower 5 plots) from a simultaneous fit. The signal template corresponds to $\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = 10^{-3}$. Uncertainty bands include the systematic uncertainties embedded into the fit model.

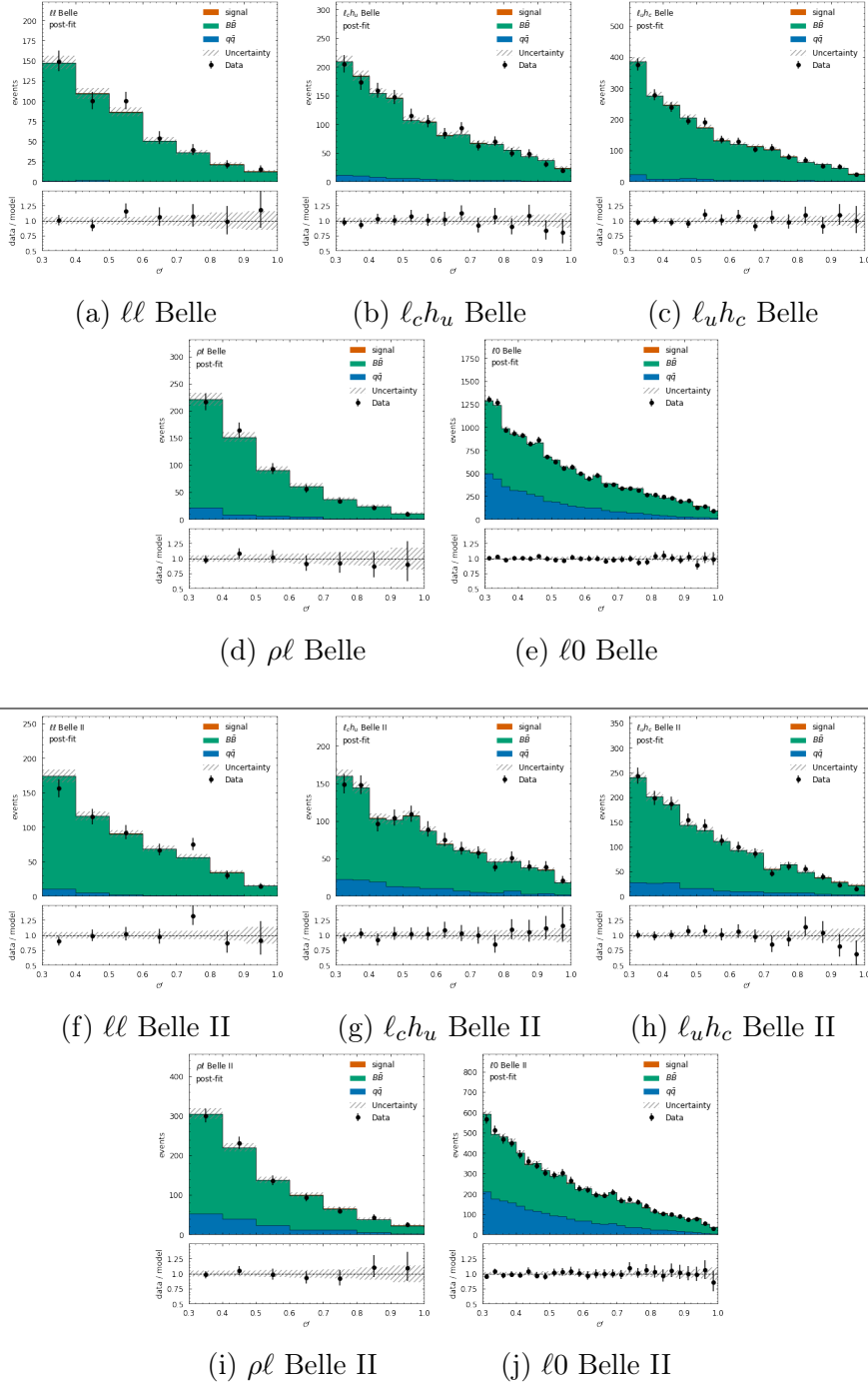


Figure 8.14: Post-fit distributions of the BDT output $\mathcal{O}'$ in Belle (upper 5 plots) and Belle II (lower 5 plots) from a simultaneous fit. The signal shown corresponds to measured $\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = 0.12 \times 10^{-3}$. Uncertainty bands shows the total post-fit uncertainty in each bin.

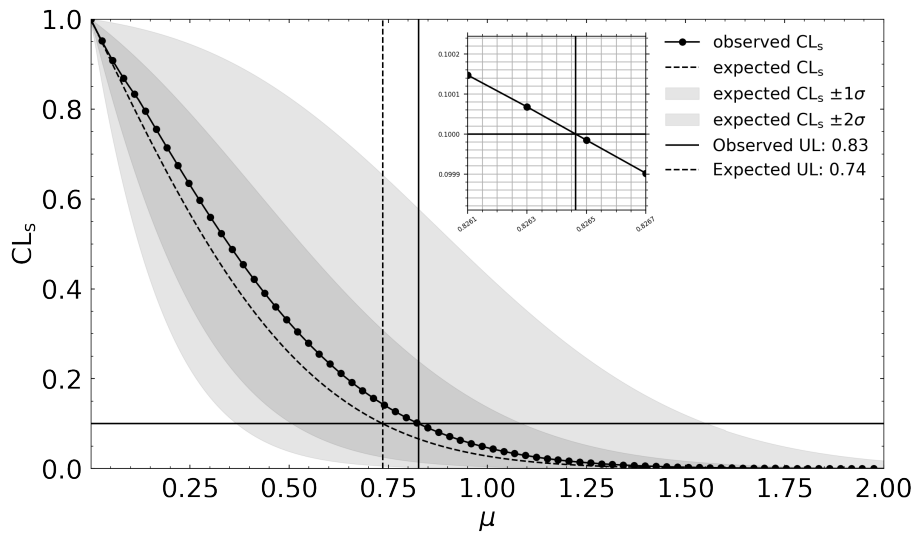


Figure 8.15: CL_s scan from the fit to Belle+Belle II datasets.

8.4. Results on data

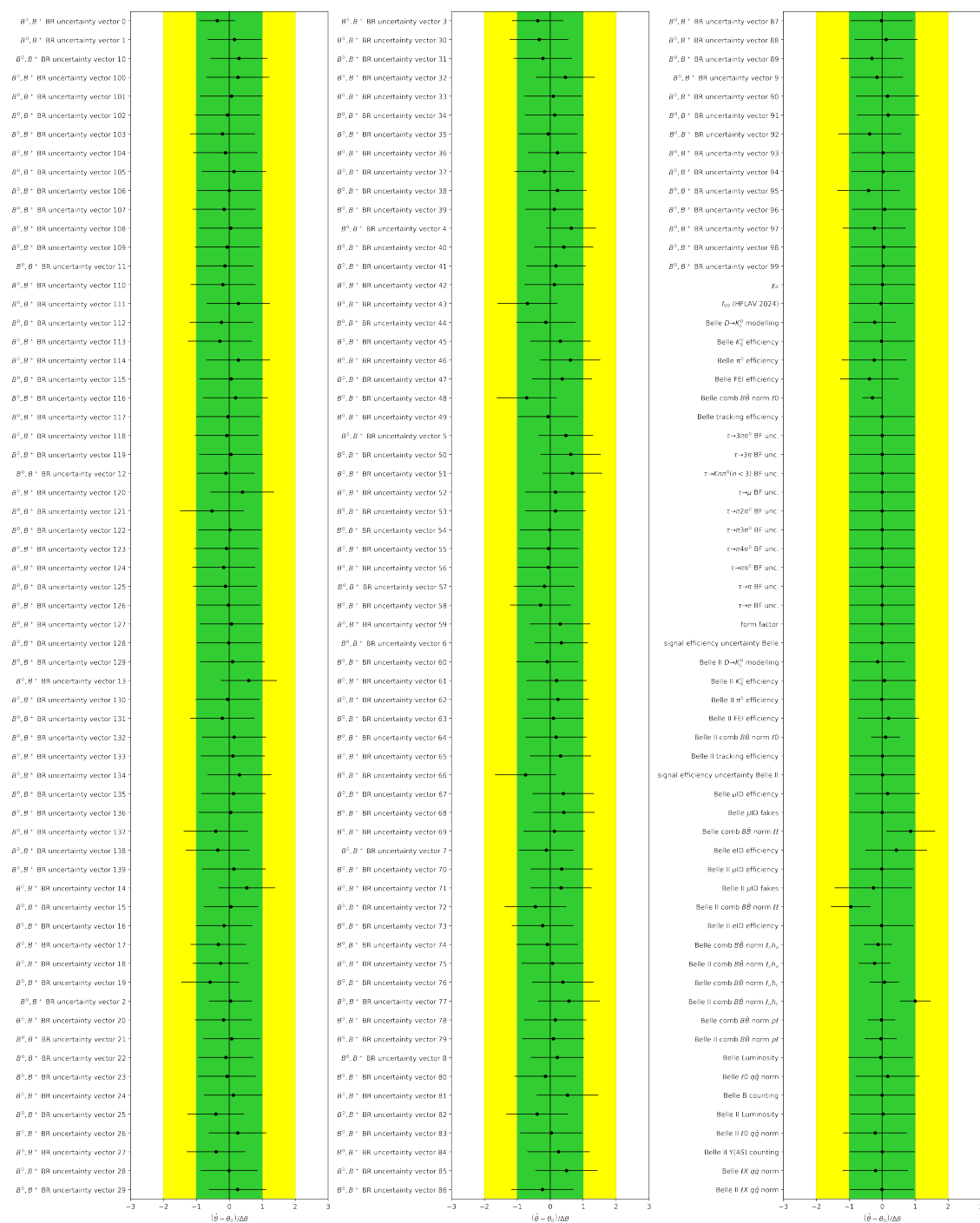


Figure 8.16: Pull plot from the fit to Belle+Belle II datasets.

8.5 Isospin average

The recent search for $B^+ \rightarrow K^+\tau^+\tau^-$ using Belle and Belle II data provides the most stringent constraint in the pseudoscalar-kaon sector, setting an upper limit of

$$\mathcal{B}_{\text{UL}}(B^+ \rightarrow K^+\tau^+\tau^-) < 0.56 \times 10^{-3}$$

at 90% confidence level [9]. Together with the search for $B^0 \rightarrow K_s^0\tau^+\tau^-$ presented in this work, it probes the same underlying $b \rightarrow s\tau^+\tau^-$ transition with a pseudoscalar kaon in the final state. In many beyond-SM scenarios, the NP contribution is expected to affect the charged and neutral modes in a closely related way, up to known lifetime factors. The neutral mode therefore provides an independent measurement with different signal selection, background composition, and systematic uncertainties, allowing the two results to be combined into an isospin-averaged constraint on $B \rightarrow K\tau^+\tau^-$.

The two measurements are not identical in their experimental strategy. The $B^0 \rightarrow K_s^0\tau^+\tau^-$ analysis described in this work uses several $\tau\tau$ final-state categories and extracts the signal from a simultaneous binned maximum-likelihood fit to the transformed BDT output distribution. In contrast, the $B^+ \rightarrow K^+\tau^+\tau^-$ analysis uses only leptonic τ decays and determines the signal yield from the event counts in the most sensitive region of the residual calorimeter energy. Although these approaches appear different, the counting measurement can be written as a binned likelihood with one signal-region bin for Belle and one for Belle II. This makes it possible to reinterpret the charged-mode result in the same `pyhf` framework used for the neutral-mode analysis.

For this purpose, the observed yields, expected background yields, signal efficiencies, and systematic uncertainties of the $B^+ \rightarrow K^+\tau^+\tau^-$ measurement are used to construct a two-bin `pyhf` model. The two bins correspond to the Belle and Belle II signal regions. Signal-efficiency uncertainties are implemented as normalization uncertainties, while the uncertainty on the expected background yield is implemented as a bin-wise statistical uncertainty. The resulting model reproduces the nominal charged-mode central value and uncertainty,

$$\mathcal{B}(B^+ \rightarrow K^+\tau^+\tau^-) = (0.04 \pm 0.30) \times 10^{-3},$$

as well as the published upper limit. This validates the use of the `pyhf` implementation for the isospin average.

The neutral-mode likelihood is then modified to express the result in terms of the underlying $B^0 \rightarrow K^0\tau^+\tau^-$ decay rather than the experimentally reconstructed $B^0 \rightarrow K_s^0\tau^+\tau^-$ final state, assuming $\mathcal{B}(B^0 \rightarrow K^0\tau^+\tau^-) = 2 \cdot \mathcal{B}(B^0 \rightarrow K_s^0\tau^+\tau^-)$. The lifetime difference between charged and neutral B mesons is also included, since the isospin relation applies to the decay widths rather than directly to the branching fractions. The common branching fraction parameter is therefore defined through

$$\mathcal{B}(B \rightarrow K\tau^+\tau^-) \equiv \mathcal{B}(B^+ \rightarrow K^+\tau^+\tau^-) = 2 \frac{\tau_{B^+}}{\tau_{B^0}} \mathcal{B}(B^0 \rightarrow K_s^0\tau^+\tau^-),$$

where $\tau_{B^+}/\tau_{B^0} = 1.076 \pm 0.004$ [3]. The uncertainty on the lifetime ratio is included as an additional normalization systematic uncertainty.

The charged- and neutral-mode models are combined into a single likelihood with one shared parameter of interest, $\mathcal{B}(B \rightarrow K\tau^+\tau^-)$. Correlations between systematic

uncertainties are treated according to the origin of each effect. The uncertainties related to the number of produced $B\bar{B}$ pairs and to tracking efficiencies are correlated between the charged and neutral modes within the same experiment. The uncertainty associated with the $\Upsilon(4S)$ branching fractions into charged and neutral B -meson pairs is treated as anticorrelated between the two modes. The uncertainties related to the hadronic B -tagging efficiency are kept uncorrelated between the charged and neutral channels, due to different performance and calibration for neutral and charged channels. The particle-identification uncertainties are also not correlated between the two measurements, as lepton-identification uncertainty for the charged channel is combined with the hadron-identification uncertainty, absent in $B^0 \rightarrow K^0\tau^+\tau^-$ measurement.

The fit to the combined likelihood gives

$$\mathcal{B}(B \rightarrow K\tau^+\tau^-) = (0.06 \pm 0.28) \times 10^{-4}.$$

The upper limit is derived using the same CLs prescription as for the neutral-mode analysis. The expected upper limit improves by nearly 10% relative to the charged-mode result alone, while the observed limit improves by about 5%. The comparison of the observed and expected CLs scans for the charged-mode-only and combined fits is shown in Fig. 8.17.

We obtain

$$\mathcal{B}_{\text{UL}}(B \rightarrow K\tau^+\tau^-) < 0.54 \times 10^{-3}$$

at 90% confidence level. This result provides the most stringent constraint to date on $b \rightarrow s\tau^+\tau^-$ transitions involving a pseudoscalar kaon in the final state.

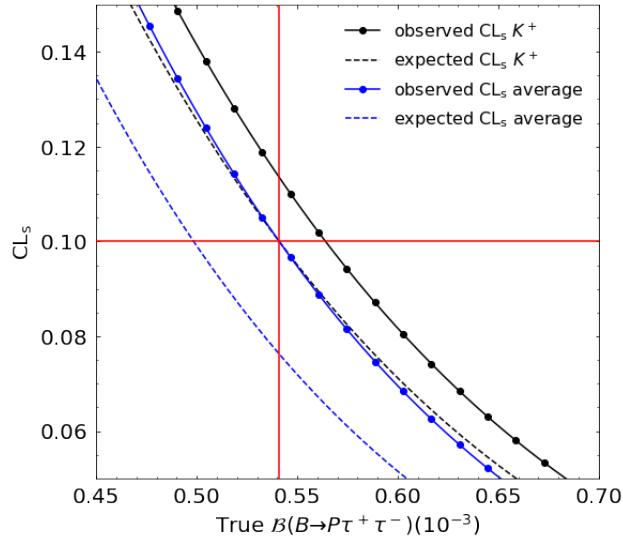


Figure 8.17: Comparison of the observed and expected CLs scans for the $B^+ \rightarrow K^+\tau^+\tau^-$ fit and for the isospin-averaged $B \rightarrow K\tau^+\tau^-$ combination.

As an illustrative reinterpretation, the isospin-averaged upper limit can be translated into a constraint on a benchmark new-physics contribution to the left-handed $b \rightarrow s\tau^+\tau^-$ operators that can explain R_X excess (see Fig. 1.6 and Sec. 1.5). Following Ref. [5], we consider the scenario

$$C_9^{\text{NP},\tau} = -\Delta, \quad C_{10}^{\text{NP},\tau} = +\Delta, \quad (8.5)$$

so that, using the notation of Eq. (1.12),

$$C_9^\tau = C_9^{\text{SM}} - \Delta, \quad C_{10}^\tau = C_{10}^{\text{SM}} + \Delta. \quad (8.6)$$

In the large- Δ limit, related to the R_X/R_X^{SM} as

$$\Delta = \frac{2\pi}{\alpha} \frac{V_{cb}}{V_{tb}V_{ts}^*} \left(\sqrt{\frac{R_X}{R_X^{\text{SM}}}} - 1 \right), \quad (8.7)$$

the branching fraction of the pseudoscalar mode is

$$\mathcal{B}(B \rightarrow K\tau^+\tau^-) = (8.8 \pm 0.8) \times 10^{-9} \Delta^2. \quad (8.8)$$

Using the observed isospin-averaged limit, we obtain (see Fig. 8.18)

$$|\Delta| < 2.6 \times 10^2. \quad (8.9)$$

This bound should be interpreted as a benchmark recast rather than a model-independent constraint, since it assumes that the dominant new-physics effect enters through the left-handed vector and axial-vector operators, with opposite shifts in C_9^τ and C_{10}^τ .

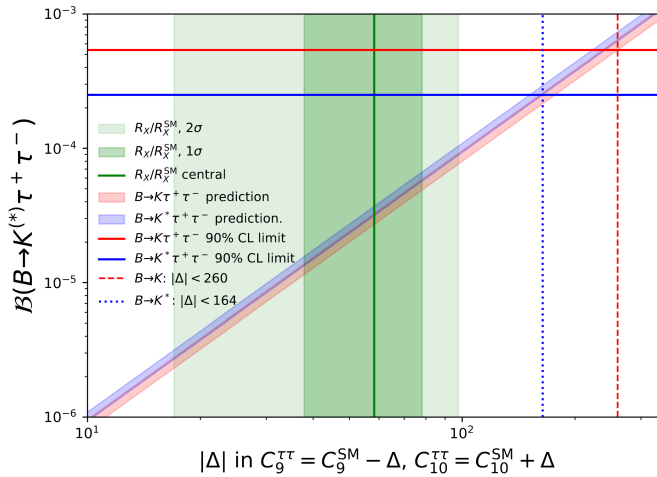


Figure 8.18: Benchmark reinterpretation of the upper limits on $\mathcal{B}(B \rightarrow K\tau^+\tau^-)$ and $\mathcal{B}(B \rightarrow K^*\tau^+\tau^-)$ in terms of the parameter $|\Delta|$, defined through $C_9^{\tau\tau} = C_9^{\text{SM}} - \Delta$ and $C_{10}^{\tau\tau} = C_{10}^{\text{SM}} + \Delta$. The red and blue shaded regions show the large-new-physics predictions for the K and K^* modes, including the theory uncertainties. The horizontal lines show the observed 90% confidence-level upper limits, while the vertical lines indicate the corresponding conservative constraints on $|\Delta|$. The green bands show the $|\Delta|$ region associated with the R_X/R_X^{SM} excess and its $\pm 1\sigma$ and $\pm 2\sigma$ ranges.

9. Conclusion and outlook

This thesis presents the first search for the rare decay $B^0 \rightarrow K_s^0 \tau^+ \tau^-$. The analysis uses the full Belle dataset and the Belle II Run 1 dataset collected at the $\Upsilon(4S)$ resonance, corresponding to integrated luminosities of 711 fb^{-1} and 365 fb^{-1} , respectively. The decay proceeds through the flavor-changing neutral-current transition $b \rightarrow s \tau^+ \tau^-$, which is highly suppressed in the Standard Model but can be significantly enhanced in presence of possible New Physics mediators that couple preferentially to third-generation leptons.

The experimental signature is challenging because the two τ leptons produce multiple neutrinos, preventing a full reconstruction of the signal decay. The analysis therefore exploits the clean environment of $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ events. One of the two B mesons is fully reconstructed in hadronic decay modes, providing constraints on the signal-side kinematics, otherwise inaccessible in final states with missing energy.

Signal candidates are formed by combining a $K_s^0 \rightarrow \pi^+ \pi^-$ candidate with two oppositely charged tracks. Only the candidates consistent with topologies where τ decays to a final state with a single charged particle are considered. The selected events are divided into five mutually exclusive categories according to the reconstructed $\tau\tau$ final-state composition: $\ell\ell$, $\ell_c h_u$, $\ell_u h_c$, $\rho\ell$, and $\ell 0$. This categorization exploits differences in signal efficiency, background composition, and charge-correlation information between the reconstructed τ daughters and the inferred B_{sig} flavor.

Several corrections are applied to the simulated samples to account for known differences with data. The dominant correction is associated with the hadronic B -tagging efficiency, while additional corrections account for charged-track reconstruction, K_s^0 reconstruction, lepton identification, π^0 reconstruction, τ branching fractions, and background-normalizations. Continuum backgrounds are constrained using off-resonance data, and the combinatorial $B\bar{B}$ component is studied with data sidebands. The uncertainties associated with these corrections are propagated to the final statistical model.

Background suppression is performed with multivariate classifiers trained separately for the different categories and datasets. The classifier output is transformed into the variable $\mathcal{O}'$, for which the signal distribution is uniform by construction. The final signal extraction is performed in the range $\mathcal{O}' > 0.3$, using a simultaneous binned maximum-likelihood fit to the Belle and Belle II samples and to all signal categories. The fit model includes templates for the signal, $B\bar{B}$ background, and continuum background, with statistical and systematic uncertainties implemented as nuisance parameters. Main systematic sources arise from poor knowledge of the signal-side B -meson decay branching fractions and from limited simulated sample statistics.

The analysis strategy is validated in several control regions before inspecting the signal-enhanced region. The K_s^0 and M_{bc} sidebands are used to test the modeling of the BDT input variables, the transformed classifier output, and the background-normalization corrections. The validation of the signal efficiency is performed in a sample enriched in $B^0 \rightarrow D^{(*)-} \ell^+ \nu_\ell$ decays, obtained from the charm-vetoed region. The agreement observed in these regions supports the use of the simulated templates and justifies the uncertainties assigned for the signal extraction.

No significant signal is observed in the Belle, Belle II, or combined datasets. The simultaneous fit gives

$$\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-) = [0.12 \pm 0.33 \text{ (stat.)} \pm 0.29 \text{ (syst.)}] \times 10^{-3},$$

which is compatible with the background-only expectation. The corresponding upper limit is

$$\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-) < 0.83 \times 10^{-3}$$

at 90% confidence level. **This result represents the first experimental constraint on this decay mode.**

The neutral-mode result is also combined with the recent measurement of $B^+ \rightarrow K^+ \tau^+ \tau^-$ to obtain a constraint on the $B \rightarrow K \tau^+ \tau^-$ transition. This isospin average is performed with a common likelihood in which the charged and neutral modes share a single branching-fraction parameter, and correlations between systematic uncertainties are treated according to their physical origin. The combined fit gives

$$\mathcal{B}(B \rightarrow K \tau^+ \tau^-) = (0.06 \pm 0.28) \times 10^{-3},$$

and the corresponding upper limit is

$$\mathcal{B}(B \rightarrow K \tau^+ \tau^-) < 0.54 \times 10^{-3}$$

at 90% confidence level. **This provides the most stringent constraint to date on $b \rightarrow s \tau^+ \tau^-$ transitions involving a pseudoscalar kaon in the final state.**

The results presented in this thesis were also reported as preliminary results at the Electroweak and QCD sessions of the 60th Rencontres de Moriond, and are documented in the corresponding conference proceedings [65, 66]. At the time of writing, the paper describing the full analysis is in the final stages of the Belle II collaboration review, with submission for publication expected to follow after the completion of this process.

The methodology developed in the scope of this work, in particular the simultaneous treatment of Belle and Belle II data and simulation, the improvements in category-dependent multivariate suppression, and the likelihood-based combination with related channels, provides a foundation for future searches for rare $b \rightarrow s \tau^+ \tau^-$ processes at Belle II. The achieved sensitivity is still several orders of magnitude above the Standard Model prediction, but close to the range of branching fractions relevant for some new-physics scenarios with enhanced couplings to τ leptons. In order to confirm or exclude such models, a significant advancement in the experimental techniques is required. Possible improvements involve the B -tagging performance (either hadronic or with higher efficiency approaches), the particle identification, and more precise calibration of tag-side and signal-side efficiencies. A further boost to the sensitivity to $b \rightarrow s \tau^+ \tau^-$ decays will be driven by the significantly larger dataset expected to be collected in the near future by the Belle II experiment.

10. Bibliography

- [1] G. Aad *et al.* (ATLAS Collaboration), *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, Phys. Lett. B **716**, 1 (2012).
- [2] S. Chatrchyan *et al.* (CMS Collaboration), *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716**, 30 (2012).
- [3] S. Navas *et al.* (Particle Data Group), *Review of Particle Physics*, Phys. Rev. D **110**, 030001 (2024).
- [4] W. Altmannshofer and D. M. Straub, *New Physics in $b \rightarrow s$ Transitions after Moriond 2021*, Eur. Phys. J. C **81**, 952 (2021).
- [5] B. Capdevila, A. Crivellin, S. Descotes-Genon, L. Hofer and J. Matias, *Searching for New Physics with $b \rightarrow s\tau^+\tau^-$ Processes*, Phys. Rev. Lett. **120**, 181802 (2018).
- [6] S. Banerjee *et al.* (Heavy Flavor Averaging Group), *Averages of b -Hadron, c -Hadron, and τ -Lepton Properties as of 2023*, Phys. Rev. D **113**, 012008 (2026), with online updates at <https://hflav.web.cern.ch/>.
- [7] A. Angelescu, D. Bećirević, D. A. Faroughy, F. Jaffredo and O. Sumensari, *Single Leptoquark Solutions to the B -Physics Anomalies*, Phys. Rev. D **104**, 055017 (2021).
- [8] 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**, 054013 (2015).
- [9] M. Abumusabh *et al.* (Belle and Belle II Collaborations), *Search for the Decay $B^+ \rightarrow K^+\tau^+\tau^-$ Using Data from the Belle and Belle II Experiments*, arXiv 10.48550/arXiv.2603.24437 (2026).
- [10] R. Aaij *et al.* (LHCb Collaboration), *Searches for $B^0 \rightarrow K^+\pi^-\tau^+\tau^-$ and $B_s^0 \rightarrow K^+K^-\tau^+\tau^-$ Decays* (2025), arXiv:2510.13716 [hep-ex] .
- [11] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531 (1963).
- [12] M. Kobayashi and T. Maskawa, *CP -Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49**, 652 (1973).

- [13] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation*, Phys. Rev. Lett. **55**, 1039 (1985).
- [14] L. Wolfenstein, *Parametrization of the Kobayashi–Maskawa Matrix*, Phys. Rev. Lett. **51**, 1945 (1983).
- [15] R. Aaij *et al.* (LHCb Collaboration), *A Comprehensive Analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ Decay* (2025), arXiv:2512.18053 [hep-ex] .
- [16] 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**, 014511 (2023), erratum: Phys. Rev. D **107**, 119903 (2023).
- [17] 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**, 014511 (2023).
- [18] I. Adachi *et al.* (Belle II Collaboration), *Evidence for $B^+ \rightarrow K^+ \nu \bar{\nu}$ Decays*, Phys. Rev. D **109**, 112006 (2024).
- [19] J. P. Lees *et al.* (BaBar Collaboration), *Search for $B^+ \rightarrow K^+ \tau^+ \tau^-$ at the BaBar Experiment*, Phys. Rev. Lett. **118**, 031802 (2017).
- [20] T. V. Dong *et al.* (Belle Collaboration), *Search for the Decay $B^0 \rightarrow K^{*0} \tau^+ \tau^-$ at the Belle Experiment*, Phys. Rev. D **108**, L011102 (2023).
- [21] I. Adachi *et al.* (Belle II Collaboration), *Search for $B^0 \rightarrow K^{*0} \tau^+ \tau^-$ Decays at the Belle II Experiment*, Phys. Rev. Lett. **135**, 151801 (2025).
- [22] R. Aaij *et al.* (LHCb Collaboration), *Search for the Decays $B_s^0 \rightarrow \tau^+ \tau^-$ and $B^0 \rightarrow \tau^+ \tau^-$* , Phys. Rev. Lett. **118**, 251802 (2017).
- [23] S. Kurokawa and E. Kikutani, *Overview of the KEKB Accelerators*, Nucl. Instrum. Meth. A **499**, 1 (2003).
- [24] T. Abe *et al.*, *Achievements of KEKB*, Prog. Theor. Exp. Phys. **2013**, 03A001 (2013).
- [25] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB Collider*, Nucl. Instrum. Meth. A **907**, 188 (2018).
- [26] Y. Ohnishi *et al.*, *Accelerator Design at SuperKEKB*, Prog. Theor. Exp. Phys. **2013**, 03A011 (2013).
- [27] A. Abashian *et al.*, *The Belle Detector*, Nucl. Instrum. Meth. A **479**, 117 (2002).
- [28] T. Abe *et al.* (Belle II Collaboration), *Belle II Technical Design Report* (2010), arXiv:1011.0352 [physics.ins-det] .
- [29] E. Kou *et al.* (Belle II Collaboration), *The Belle II Physics Book*, Prog. Theor. Exp. Phys. , 123C01 (2019).

-
- [30] K. Adamczyk *et al.* (Belle II SVD Collaboration), *The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector*, JINST **17** (11), P11042.
 - [31] I. Adachi *et al.*, *Aerogel RICH Counter at the Belle II Detector* (2025), arXiv:2512.19146 [hep-ex] .
 - [32] D. Kotchetkov *et al.*, *Front-End Electronic Readout System for the Belle II Imaging Time-of-Propagation Detector*, Nucl. Instrum. Meth. A **941**, 162342 (2019).
 - [33] T. Kuhr *et al.*, *The Belle II Core Software*, Comput. Softw. Big Sci. **3**, 1 (2019).
 - [34] D. J. Lange, *The EvtGen Particle Decay Simulation Package, Proceedings of the 7th International Conference on B Physics at Hadron Machines*, Nucl. Instrum. Meth. A **462**, 152 (2001).
 - [35] S. Jadach, B. F. L. Ward and Z. Wąs, *The Precision Monte Carlo Event Generator KK for Two-Fermion Final States in e^+e^- Collisions*, Comput. Phys. Commun. **130**, 260 (2000).
 - [36] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (05), 026.
 - [37] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015).
 - [38] E. Barberio, B. van Eijk and Z. Wąs, *PHOTOS: A Universal Monte Carlo for QED Radiative Corrections in Decays*, Comput. Phys. Commun. **66**, 115 (1991).
 - [39] E. Barberio and Z. Wąs, *PHOTOS: A Universal Monte Carlo for QED Radiative Corrections. Version 2.0*, Comput. Phys. Commun. **79**, 291 (1994).
 - [40] M. Gelb *et al.*, *B2BII: Data Conversion from Belle to Belle II*, Comput. Softw. Big Sci. **2**, 10.1007/s41781-018-0016-x (2018).
 - [41] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, Comput. Softw. Big Sci. **1**, 2 (2017).
 - [42] T. Keck *et al.*, *The Full Event Interpretation*, Comput. Softw. Big Sci. **3**, 6 (2019).
 - [43] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima and T. Tsukamoto, *Electron Identification in Belle*, Nucl. Instrum. Meth. A **485**, 490 (2002).
 - [44] A. Abashian *et al.*, *Muon Identification in the Belle Experiment at KEKB*, Nucl. Instrum. Meth. A **491**, 69 (2002).
 - [45] M. Milesi, J. Tan and P. Urquijo, *Lepton Identification in Belle II Using Observables from the Electromagnetic Calorimeter and Precision Trackers*, EPJ Web Conf. **245**, 06023 (2020).

- [46] X. Zhou, S. Du, G. Li and C. Shen, *TopoAna: A Generic Tool for the Event Type Analysis of Inclusive Monte Carlo Samples in High Energy Physics Experiments*, Comput. Phys. Commun. **258**, 107540 (2021).
- [47] J. E. Angus, *The Probability Integral Transform and Related Results*, SIAM Rev. **36**, 652 (1994).
- [48] G. Punzi, *Notes on Statistical Separation of Classes of Events* (2006), arXiv:physics/0611219 [physics.data-an] .
- [49] R. Cheaib, *A Search for the Decay of a B Meson into a Kaon and a Tau Lepton Pair at the BaBar Experiment*, Ph.D. thesis, McGill University (2016), accessed: 2025-01-11.
- [50] L. Heinrich, M. Feickert and G. Stark, *pyhf: v0.7.6* (2024), <https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6>.
- [51] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: Pure-Python Implementation of HistFactory Statistical Models*, J. Open Source Softw. **6**, 2823 (2021).
- [52] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke (ROOT Collaboration), *HistFactory: A Tool for Creating Statistical Models for Use with RooFit and RooStats*, CERN-OPEN-2012-016 10.17181/CERN-OPEN-2012-016 (2012).
- [53] I. Adachi *et al.* (Belle II Collaboration), *Measurement of the Integrated Luminosity of Data Samples Collected During 2019–2022 by the Belle II Experiment* (2024), arXiv:2407.00965 [hep-ex] .
- [54] A. J. Bevan *et al.*, *The Physics of the B Factories*, Eur. Phys. J. C **74**, 3026 (2014).
- [55] C. Cecchi, G. De Nardo, E. Manoni and M. Merola, *B Counting Measurement of the Run 1 Belle II Data Sample*, https://docs.belle2.org/record/3468/files/BELLE2-NOTE-PH-2023-011_v2.pdf (2023).
- [56] Belle Collaboration, *Number of B Events in Hadron B(J)*, <https://belle.kek.jp/secured/nbb/nbb.html> (2008).
- [57] A. Ali, P. Ball, L. T. Handoko and G. Hiller, *Comparative Study of the Decays $B \rightarrow (K, K^*)\ell^+\ell^-$ in the Standard Model and Supersymmetric Theories*, Phys. Rev. D **61**, 074024 (2000).
- [58] D. M. Straub, *flavio: A Python Package for Flavour and Precision Phenomenology in the Standard Model and Beyond* (2018), arXiv:1810.08132 [hep-ph] .
- [59] N. Gubernari, M. Reboud, D. van Dyk and J. Virto, *Dispersive Analysis of $B \rightarrow K^{(*)}$ and $B_s \rightarrow \phi$ Form Factors*, JHEP **12** (12), 153, erratum: JHEP **01**, 125 (2025).
- [60] J. Aebischer, J. Kumar, P. Stangl and D. M. Straub, *A Global Likelihood for Precision Constraints and Flavour Anomalies*, Eur. Phys. J. C **79**, 509 (2019).

-
- [61] W. G. Parrott, C. Bouchard and C. T. H. Davies (HPQCD Collaboration), *Erratum: 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**, 119903 (2023).
 - [62] Y. Amhis *et al.* (Heavy Flavor Averaging Group), *Averages of b -Hadron, c -Hadron, and τ -Lepton Properties as of 2021*, Phys. Rev. D **107**, 052008 (2023), with specific result from HFLAV [webpage](#).
 - [63] D. van Dyk and L. Lyons, *How to Incorporate Systematic Effects into Parameter Determination* (2023), [arXiv:2306.05271 \[hep-ex\]](#) .
 - [64] A. Pinto, Z. Wu, F. Balli, N. Berger, M. Boonekamp, É. Chapon, T. Kawamoto and B. Malaescu, *Uncertainty Components in Profile Likelihood Fits*, Eur. Phys. J. C **84**, 593 (2024).
 - [65] M. Liu, *$b \rightarrow s\ell^+\ell^-$ ($\ell = e, \mu, \tau$) and $b \rightarrow s\nu\bar{\nu}$ at Belle and Belle II* (2026), [arXiv:2604.22079 \[hep-ex\]](#) .
 - [66] V. Bertacchi, *Measurements of electroweak penguins and B decays to final states with missing energy at Belle and Belle II* (2026), [arXiv:2604.17644 \[hep-ex\]](#) .

Appendix A

Data management plan

0. General information

0.1 Name of the doctoral student Ihor Prudiiev

0.2 Doctoral programme and scientific field Doctoral programme in Physics, scientific field: experimental particle physics

0.3 Title of the doctoral dissertation Search for $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ decays at Belle and Belle II

1. Type of data and methods used if new data was collected or produced

1.1 What data have you collected and/or produced? The research presented in this dissertation is based on experimental data collected by the Belle and Belle II detectors in electron–positron collisions at the $\Upsilon(4S)$ resonance, together with simulated Monte Carlo samples produced within the Belle and Belle II collaborations. These data were not collected individually by the doctoral student, but as part of the long-term data-taking programmes of the Belle and Belle II collaborations.

1.2 How did you collect and/or produce new data and how did you use existing data for your thesis? The primary collision data were collected by the Belle experiment at the KEKB collider and by the Belle II experiment at the SuperKEKB collider. These data were not collected individually by the doctoral student, but as part of the long-term data-taking programmes of the Belle and Belle II collaborations. The corresponding simulated samples were produced using the standard collaboration production chains, including event generation, detector simulation, beam-background overlay where applicable, and event reconstruction.

For this dissertation, the existing centrally reconstructed datasets were processed to produce analysis-specific ntuples. The selected samples were then used to optimize the event selection, train and validate multivariate classifiers, evaluate signal efficiencies, estimate backgrounds, derive systematic uncertainties, construct binned templates, and perform the statistical interpretation of the result.

The main derived datasets produced in the course of the analysis are:

- reduced analysis ntuples containing reconstructed event-level and candidate-level quantities;
- histograms and templates of the final discriminating variable used in the signal extraction fit;

- observed data histograms in the fit region;
- signal-efficiency estimates, including their dependence on the generated q^2 distribution;
- generated q^2 distributions used to evaluate the signal-model dependence;
- systematic-uncertainty inputs, including normalization and shape variations;

1.3 If you have been dealing with sensitive data, please state how you handled the data in accordance with ethical standards and legislation. The research does not involve sensitive data.

2. Data storage and protection during research

2.1 How did you store the data? The primary Belle and Belle II collision data and centrally produced Monte Carlo samples are stored and managed according to the computing and data-management policies of the Belle and Belle II collaborations. The storage is distributed across collaboration computing resources, including Grid storage and computing sites, as well as the KEK Computing Center (KEKCC) in Japan. Access to these datasets is restricted to authorized collaboration members and is controlled through the standard collaboration authentication and authorization mechanisms.

2.2 If you have worked with sensitive data, how have you ensured that it was kept safe and secure? The research does not involve sensitive data.

3. Long-term data availability and storage

3.1 In which (data) repository will you publish and ensure long-term preservation of the data after the completion of the research work in accordance with Article 50 of the Rules and Regulations for Doctoral Studies at UL? The primary collision and simulated datasets are governed by the data-management and publication policies of the Belle and Belle II collaborations. These datasets are preserved within the collaboration computing infrastructure and are not deposited independently by individual analysts.

After publication of the analysis, the numerical material needed to interpret and reuse the published result will be made publicly available in a suitable high-energy-physics data repository, such as HEPData, subject to the approval procedures of the Belle and Belle II collaborations. The material planned for public release includes the binned fit templates and observed data entering the final signal extraction, the signal efficiencies, the signal-efficiency dependence on q^2 , the generated q^2 distribution used in the analysis, and other numerical inputs needed to reproduce or reinterpret the published result.

3.2 If you plan to restrict access to the data for a certain period, please state the reasons for this. Access to the raw and centrally reconstructed Belle and Belle II data is restricted according to the policies of the Belle and Belle II collaborations.

Appendix B

BDT training variables

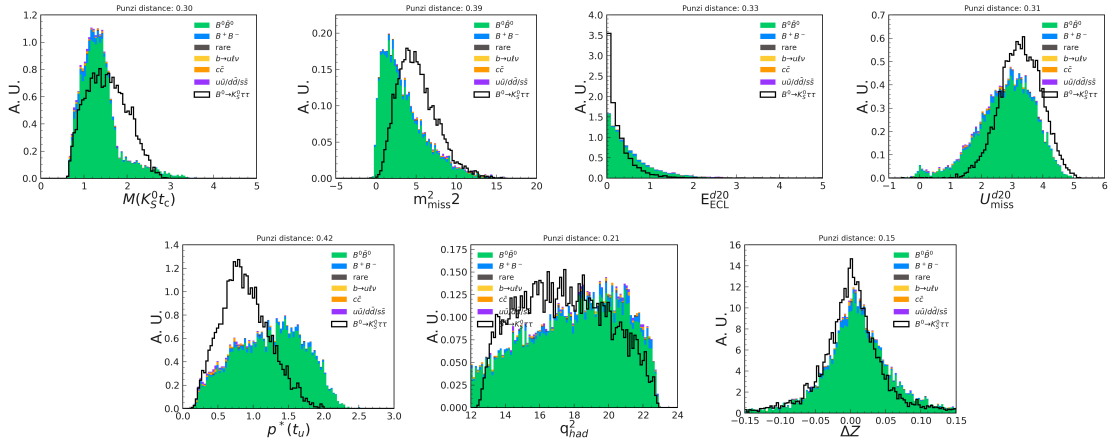


Figure B.1: Training features for the $\ell\ell$ category (Belle MC).

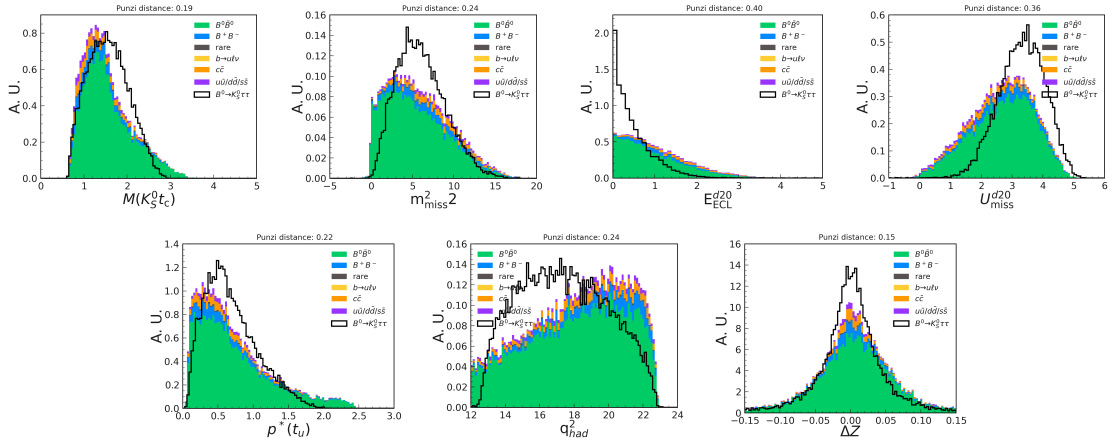


Figure B.2: Training features for the $\ell_c h_u$ category (Belle MC).

Appendix B. BDT training variables

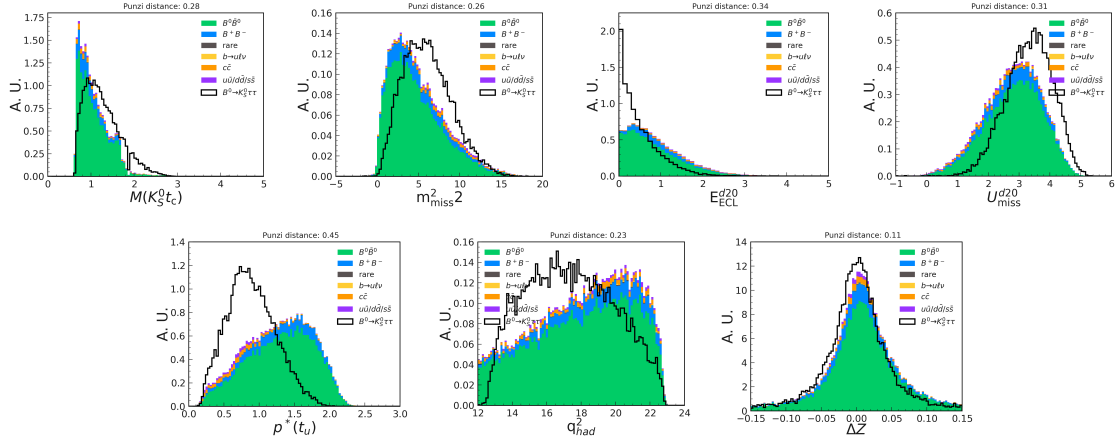


Figure B.3: Training features for the $\ell_u h_c$ category (Belle MC).

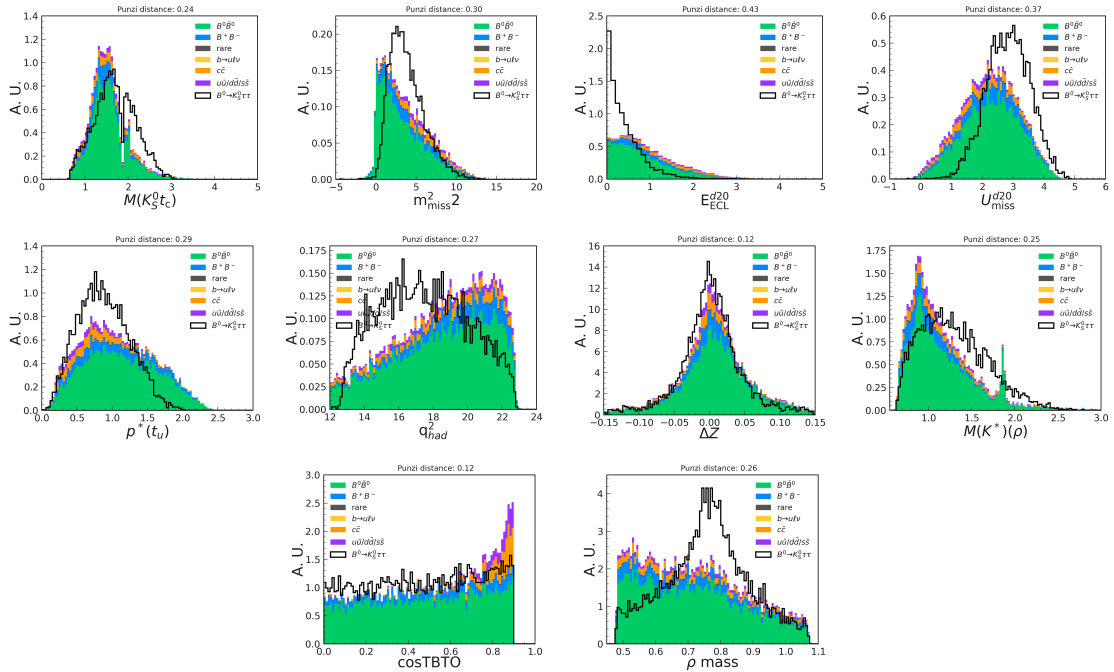


Figure B.4: Training features for the $\rho \ell$ category (Belle MC).

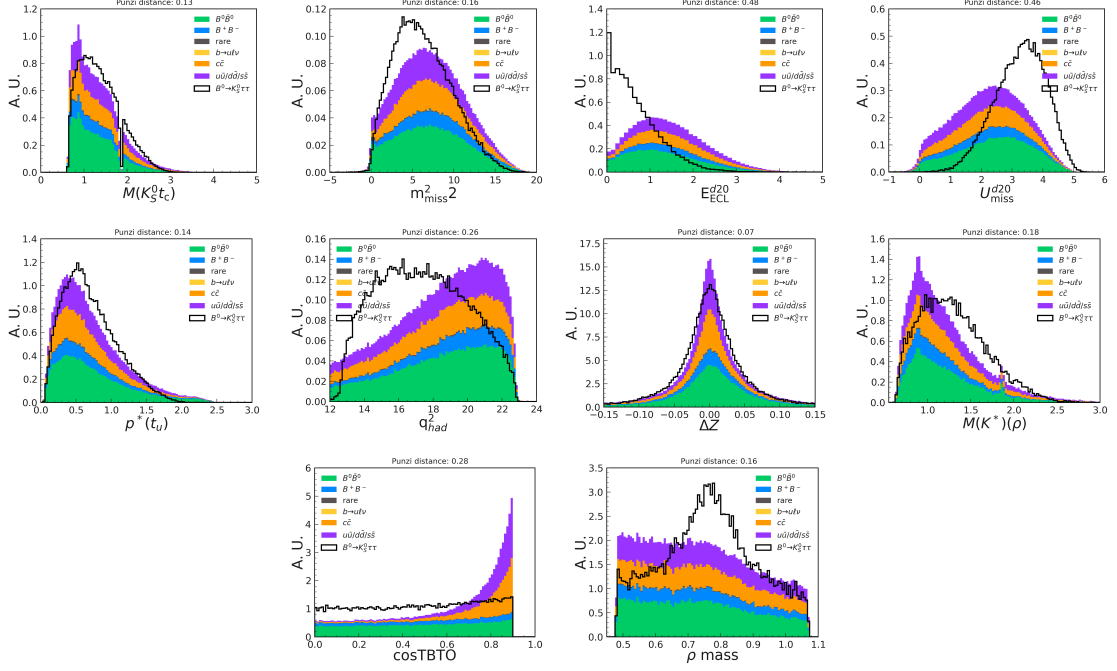


Figure B.5: Training features for the ℓ_0 category (Belle MC).

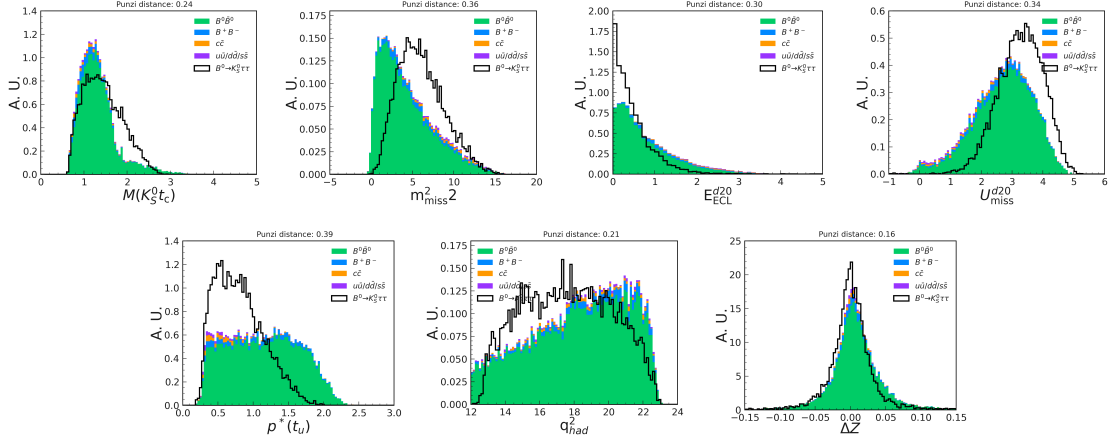


Figure B.6: Training features for the $\ell\ell$ category (Belle II MC).

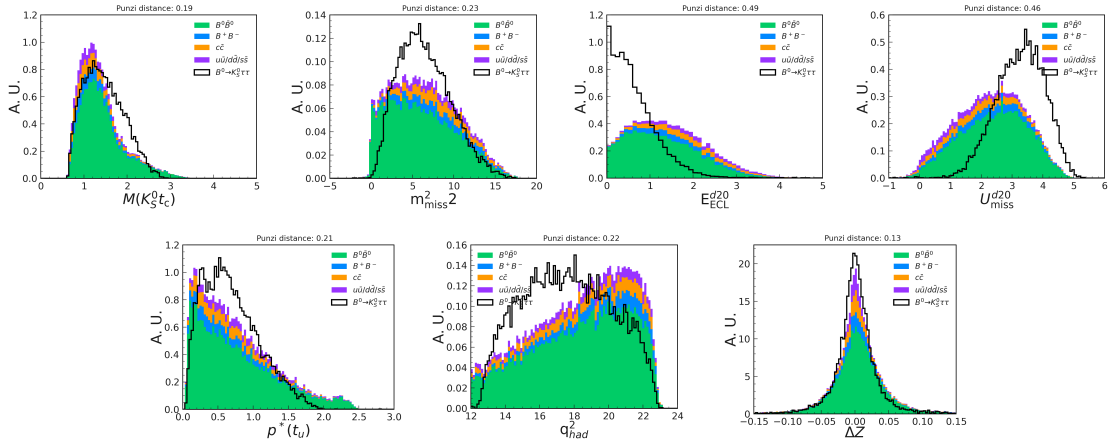


Figure B.7: Training features for the $\ell_c h_u$ category (Belle II MC).

Appendix B. BDT training variables

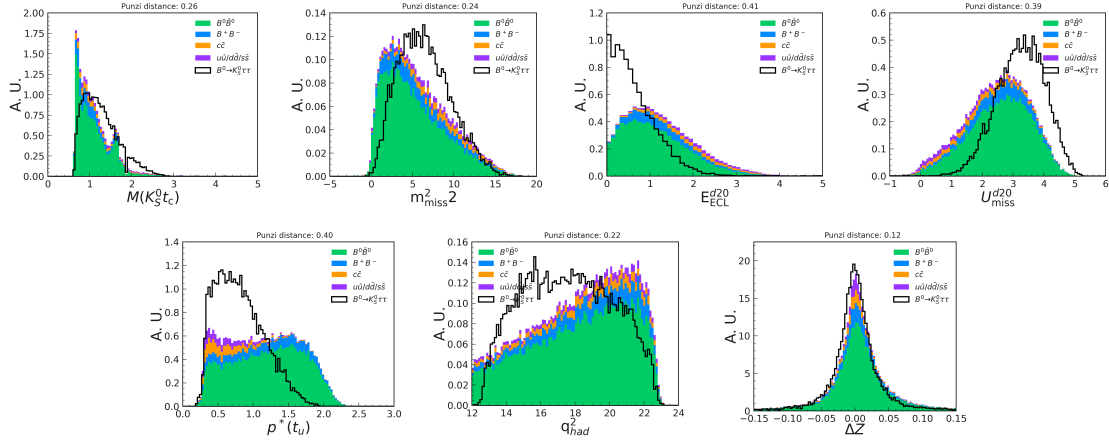


Figure B.8: Training features for the $\ell_u h_c$ category (Belle II MC).

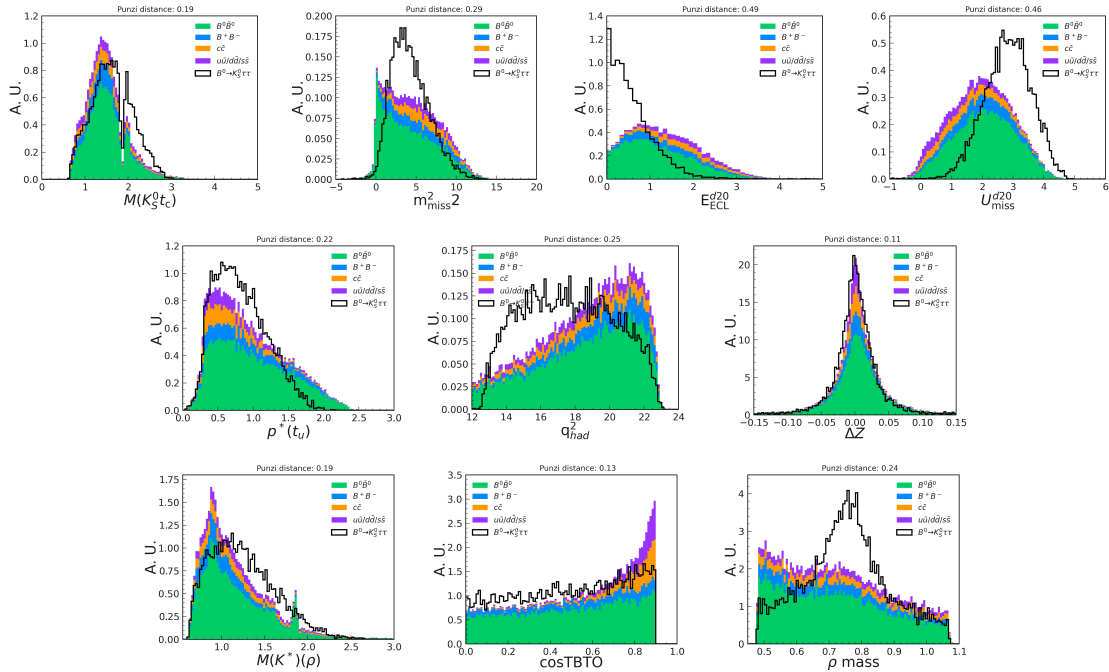


Figure B.9: Training features for the $\rho \ell$ category (Belle II MC).

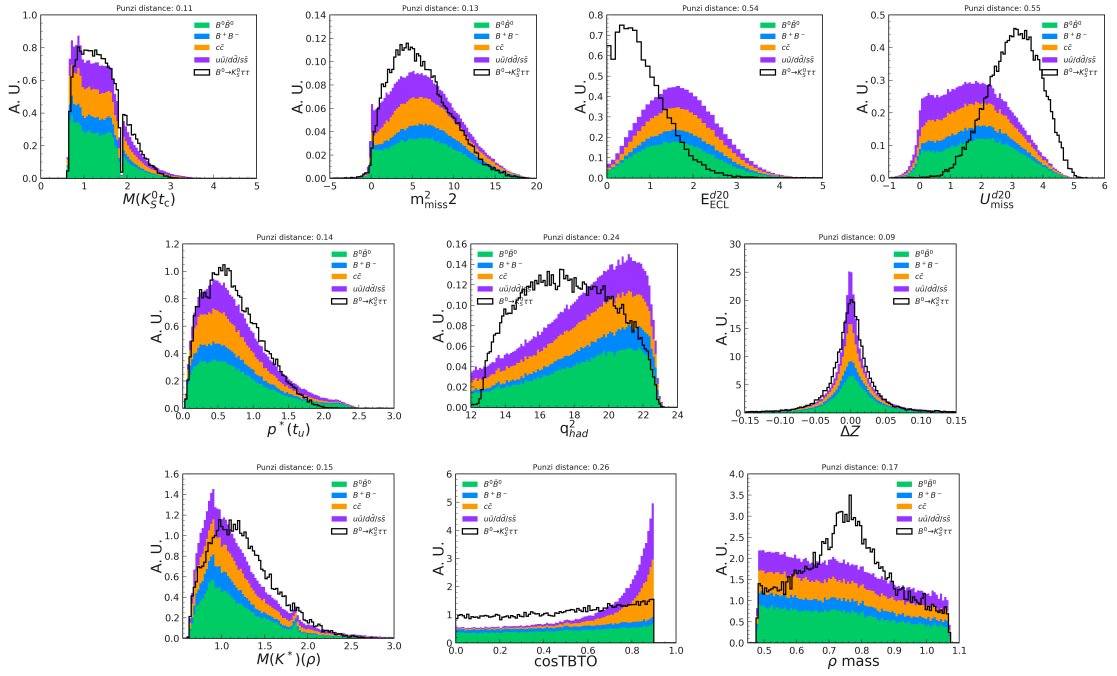


Figure B.10: Training features for the ℓ_0 category (Belle II MC).

B.1 BDT training - additional plots

This Appendix contains additional plots related to the background suppression, in particular, the BDT training. Fig. B.11 (Belle II) shows the correlations among the features used for the training.

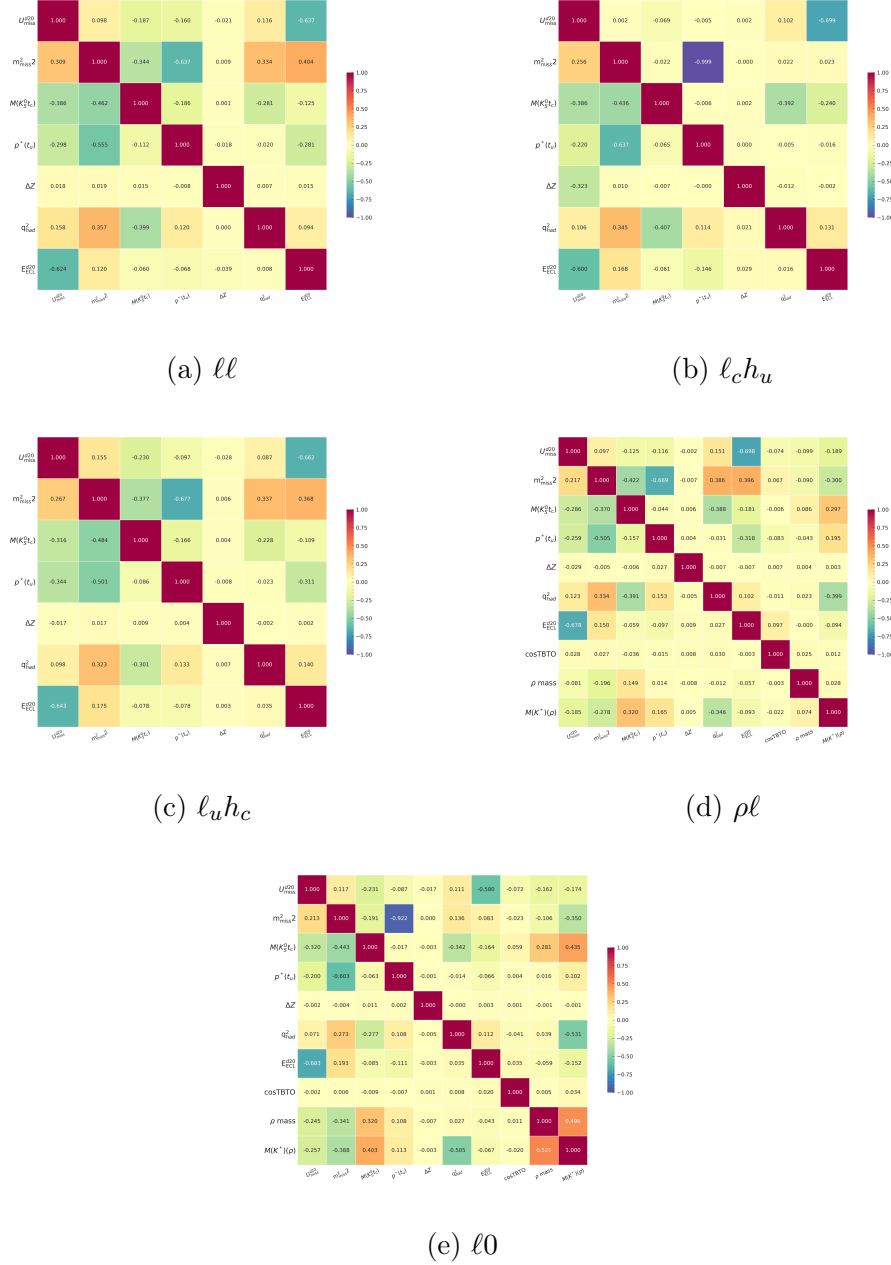
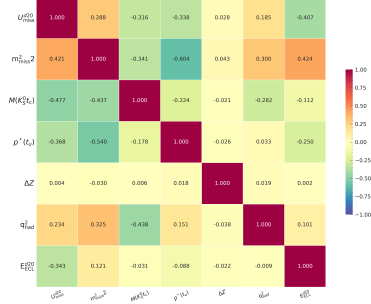
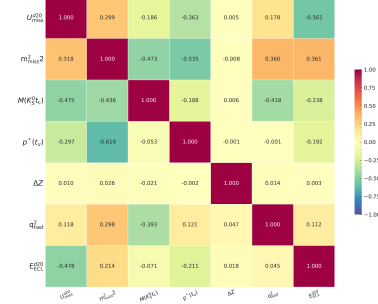


Figure B.11: Correlation matrix for the input variables used for the BDT training. The lower-left (upper-right) diagonal part represents the signal (background) events (Belle II MC).

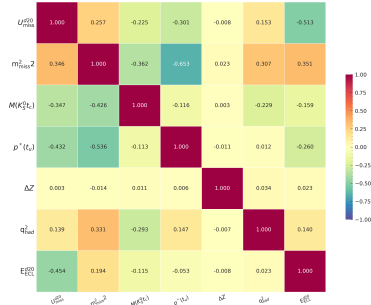
B.1. BDT training - additional plots



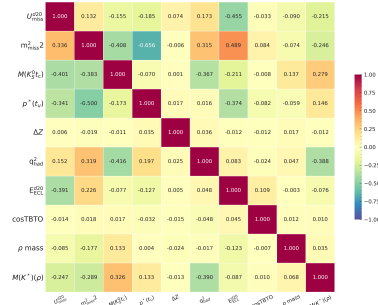
(a) $\ell\ell$



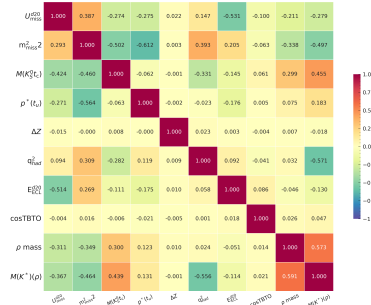
(b) ℓchu



(c) $\ell_u h_c$



(d) $\rho\ell$



(e) $\ell 0$

Figure B.12: Correlation matrix for the input variables used for the BDT training. The lower-left (upper-right) diagonal part represents the signal (background) events (Belle MC).

Appendix B. BDT training variables

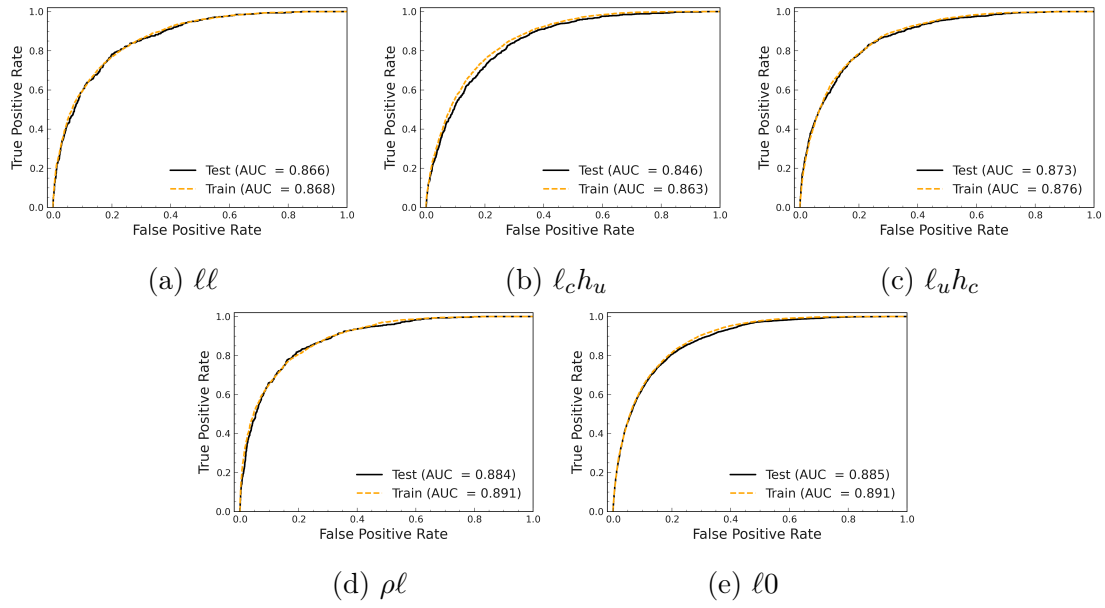


Figure B.13: ROC curve for test and train sets (Belle II MC), *even* split.

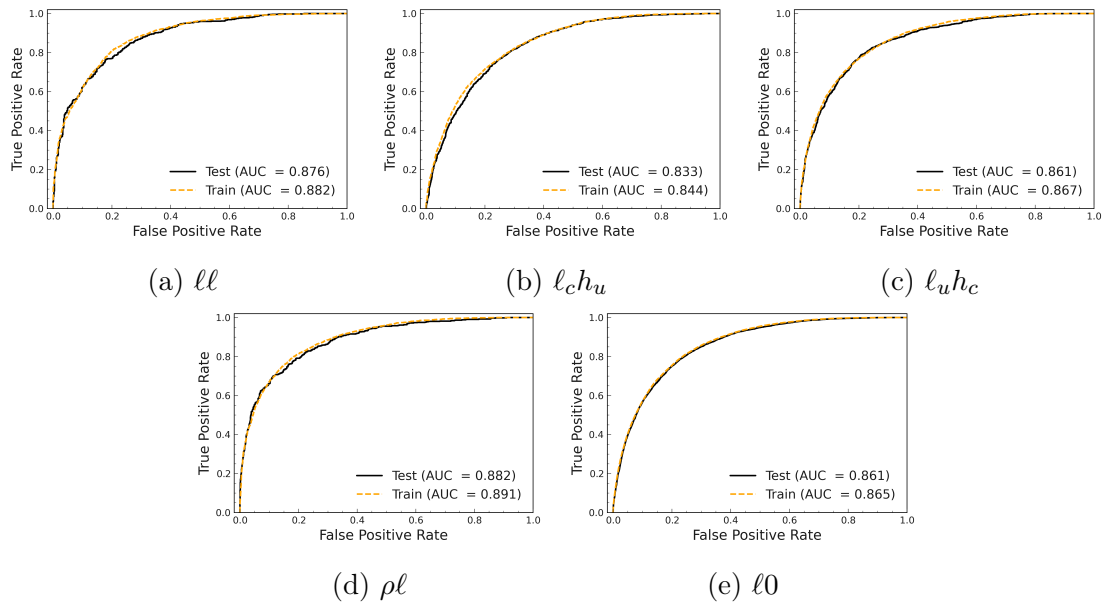


Figure B.14: ROC curve for test and train sets (Belle MC), *odd* split.

B.1. BDT training - additional plots

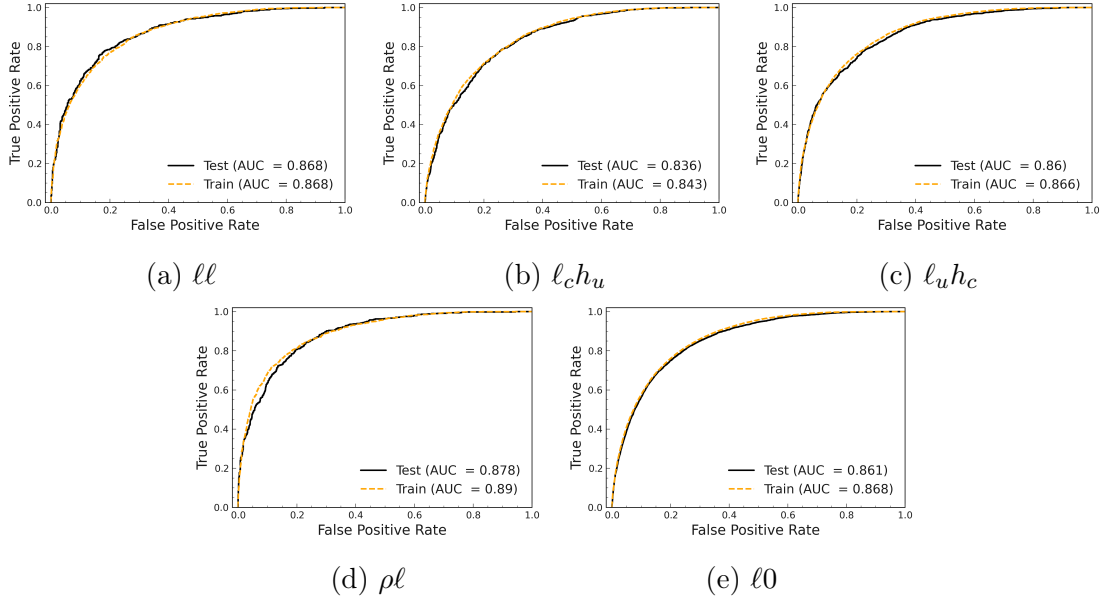


Figure B.15: ROC curve for test and train sets (Belle MC), *even* split.

	$\ell\ell$ Belle		$\ell_c h_u$ Belle		$\ell_u h_c$ Belle		$\rho\ell$ Belle		$\ell\ell$ Belle		
U_{miss}	0.148	0.390	0.355	0.512	0.328	0.424	0.241	0.438	0.421	0.523	U_{miss}
m_{miss}^2	0.152	0.296	-0.105	0.058	0.042	0.236	-0.113	0.060	-0.421	-0.105	m_{miss}^2
$M(K_S^0 t_c)$	0.402	0.166	0.125	-0.015	0.466	0.252	0.301	0.172	0.205	0.061	$M(K_S^0 t_c)$
$\rho^*(t_u)$	-0.419	-0.461	0.034	-0.065	-0.512	-0.540	-0.173	-0.266	0.114	-0.028	$\rho^*(t_u)$
ΔZ	0.005	-0.009	-0.026	-0.012	-0.001	-0.009	0.008	0.004	-0.017	-0.009	ΔZ
q_{had}^2	-0.297	-0.028	-0.370	-0.159	-0.279	-0.109	-0.378	-0.142	-0.361	-0.190	q_{had}^2
E_{ECL}	-0.516	-0.313	-0.679	-0.521	-0.587	-0.341	-0.654	-0.425	-0.720	-0.605	E_{ECL}
	SIG	BG	SIG	BG	SIG	BG	-0.032	-0.051	-0.288	-0.287	$\cos\text{TBTO}$
							0.246	0.159	0.071	-0.042	ρ mass
							0.393	0.073	0.108	-0.010	$M(K^*)(\rho)$
							SIG	BG	SIG	BG	
	$\ell\ell$ Belle II		$\ell_c h_u$ Belle II		$\ell_u h_c$ Belle II		$\rho\ell$ Belle II		$\ell\ell$ Belle II		
U_{miss}	0.318	0.411	0.455	0.571	0.434	0.449	0.366	0.449	0.441	0.496	U_{miss}
m_{miss}^2	0.199	0.301	-0.091	0.004	0.021	0.126	-0.102	-0.043	-0.326	-0.017	m_{miss}^2
$M(K_S^0 t_c)$	0.304	0.120	0.094	0.051	0.311	0.175	0.235	0.122	0.140	0.013	$M(K_S^0 t_c)$
$\rho^*(t_u)$	-0.473	-0.446	-0.014	-0.011	-0.445	-0.378	-0.145	-0.150	0.067	-0.034	$\rho^*(t_u)$
ΔZ	-0.008	-0.011	-0.028	-0.001	-0.003	-0.008	0.008	0.005	0.006	-0.006	ΔZ
q_{had}^2	-0.316	-0.068	-0.278	-0.139	-0.245	-0.102	-0.292	-0.159	-0.308	-0.153	q_{had}^2
E_{ECL}	-0.417	-0.268	-0.612	-0.518	-0.533	-0.377	-0.542	-0.419	-0.588	-0.499	E_{ECL}
	SIG	BG	SIG	BG	SIG	BG	-0.085	-0.088	-0.257	-0.224	$\cos\text{TBTO}$
							0.185	0.088	0.021	-0.060	ρ mass
							0.277	0.041	0.067	-0.029	$M(K^*)(\rho)$
							SIG	BG	SIG	BG	

Figure B.16: Correlations of input features to the transformed BDT output $\mathcal{O}'$ for signal and background for all reconstructed categories for Belle (upper plot) and Belle II (lower plot).

Appendix C

BDT training variables check in K_S^0 sideband

In all the plots shown here, the MC distributions are corrected with the “global scaling factors from Tab. 5.2, K_S^0 SB.

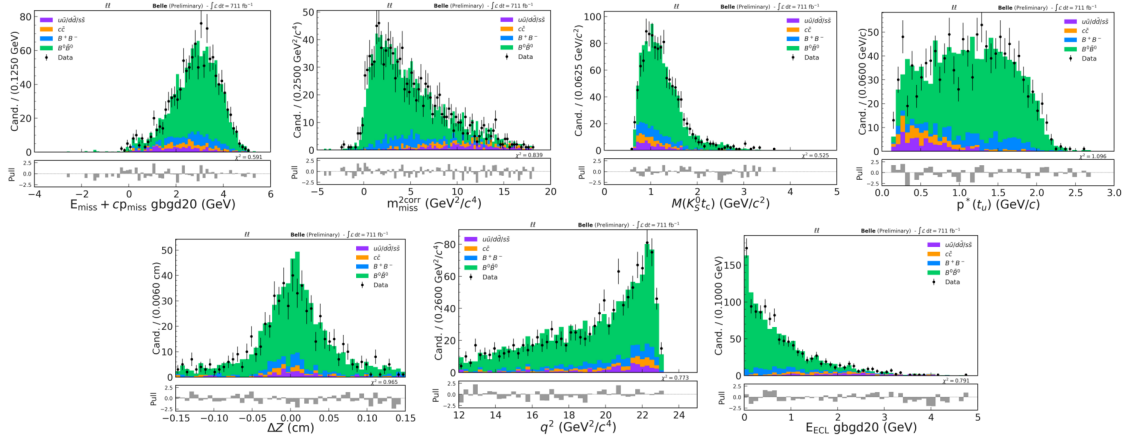


Figure C.1: Data/MC comparison of the training features for the $\ell\ell$ category (Belle K_S^0 sideband).

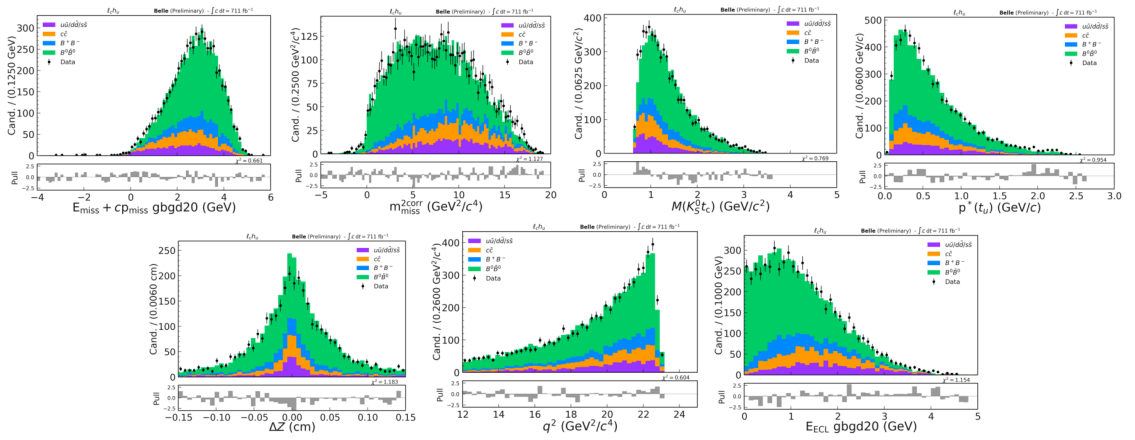


Figure C.2: Data/MC comparison of the training features for the $\ell_c h_u$ category (Belle K_S^0 sideband).

Appendix C. BDT training variables check in K_s^0 sideband

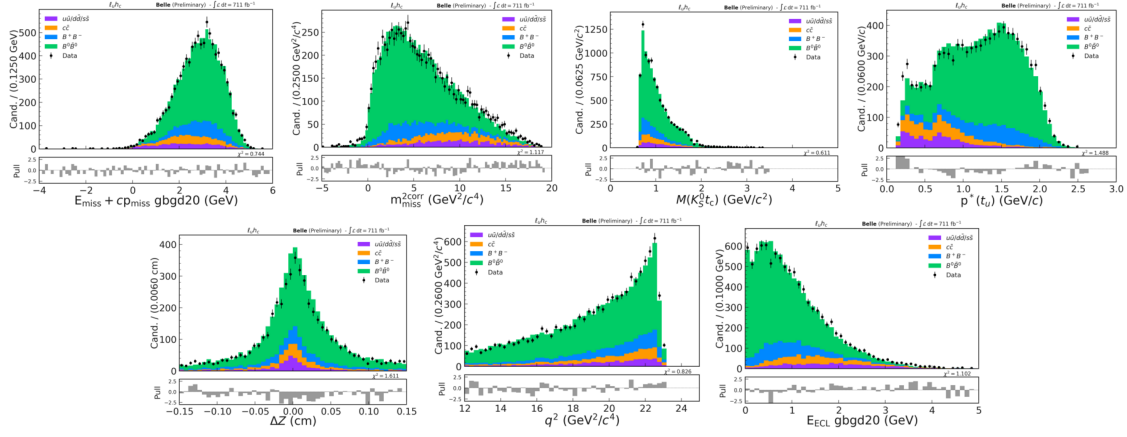


Figure C.3: Data/MC comparison of the training features for the $\ell_u h_c$ category (Belle K_s^0 sideband).

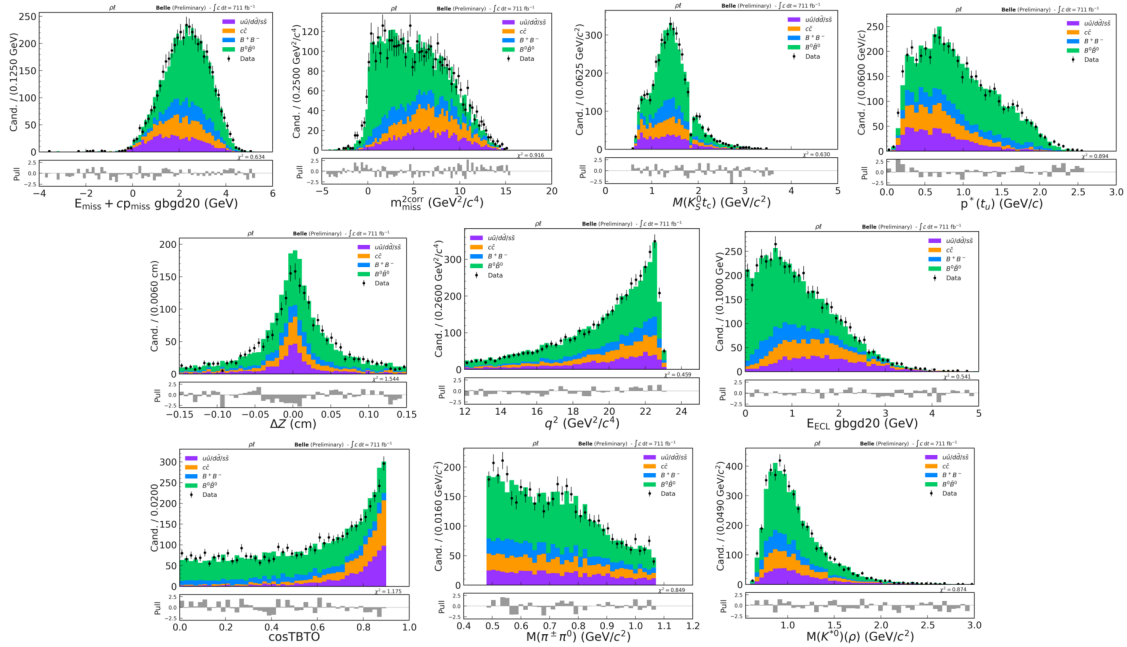


Figure C.4: Data/MC comparison of the training features for the $\rho \ell$ category (Belle K_s^0 sideband).

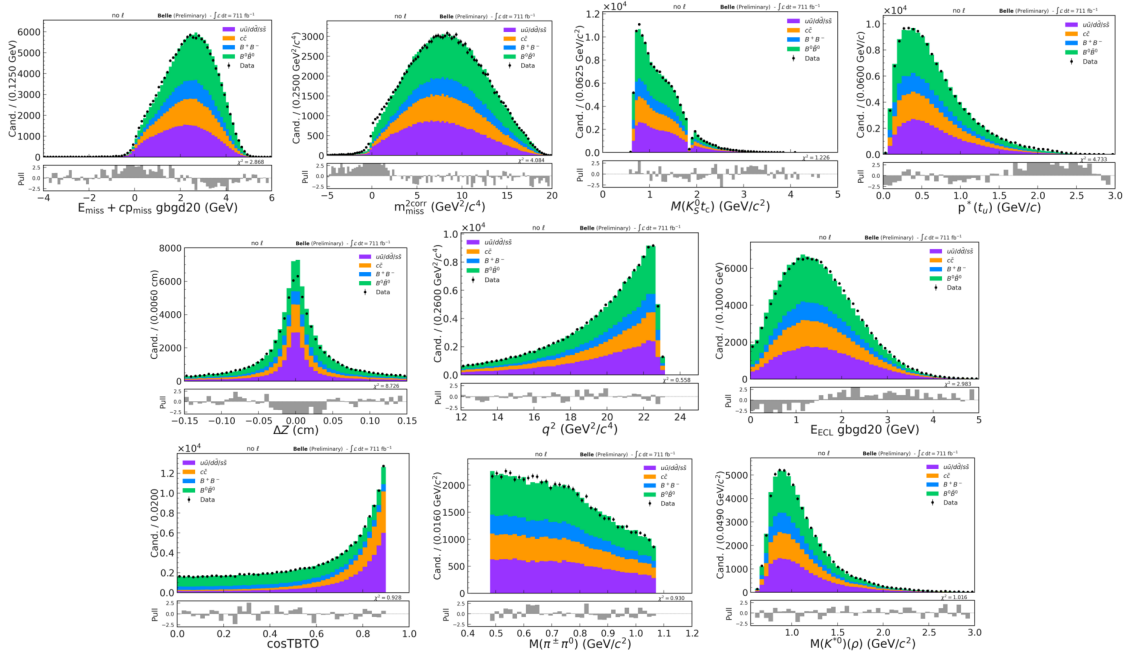


Figure C.5: Data/MC comparison of the training features for the $\ell 0$ category (Belle K_S^0 sideband).

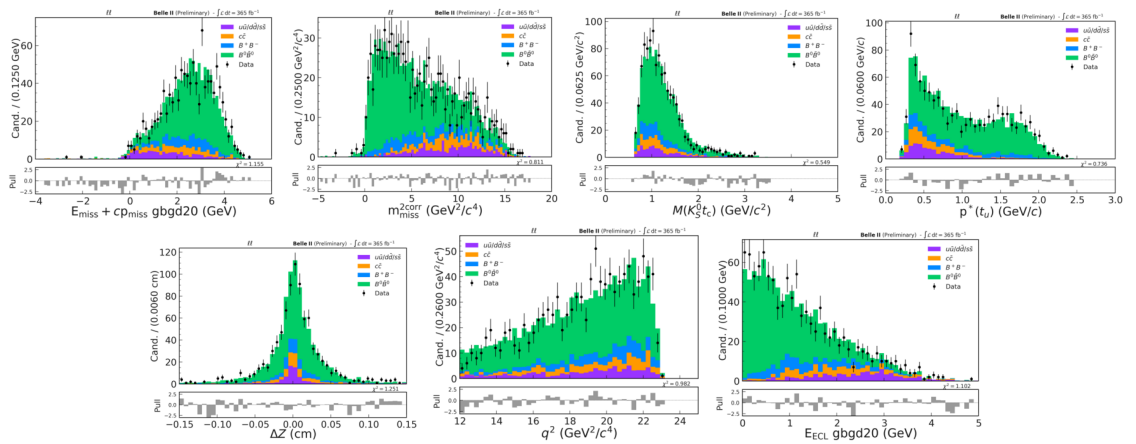


Figure C.6: Data/MC comparison of the training features for the $\ell\ell$ category (Belle II K_S^0 sideband).

Appendix C. BDT training variables check in K_S^0 sideband

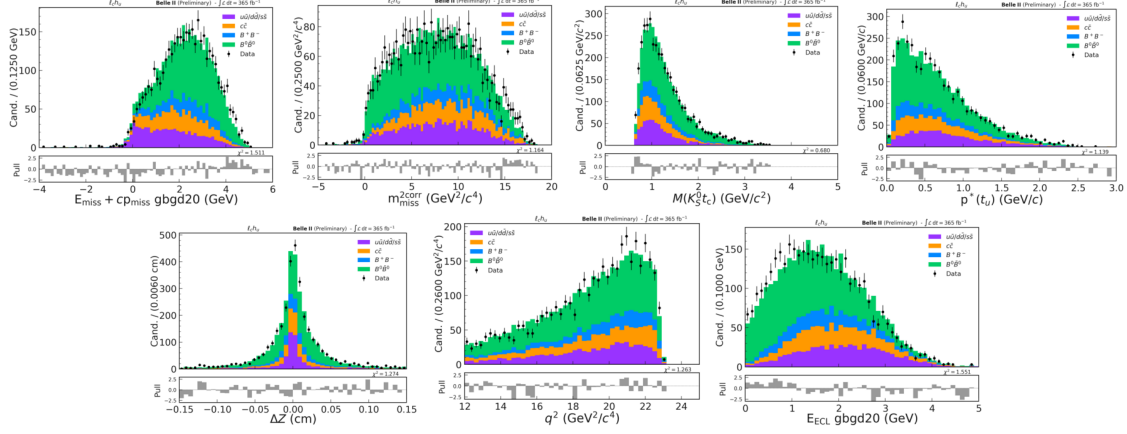


Figure C.7: Data/MC comparison of the training features for the $\ell_c h_u$ category (Belle II K_S^0 sideband).

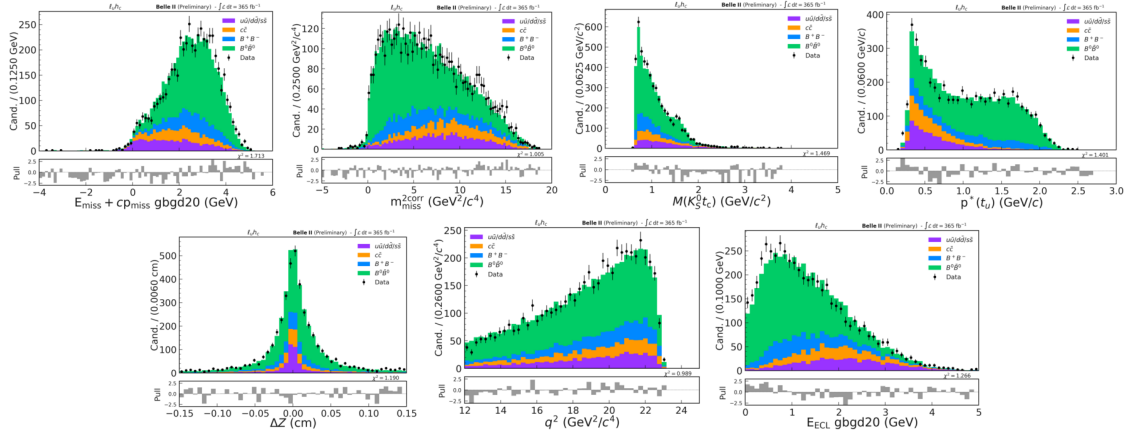


Figure C.8: Data/MC comparison of the training features for the $\ell_u h_c$ category (Belle II K_S^0 sideband).

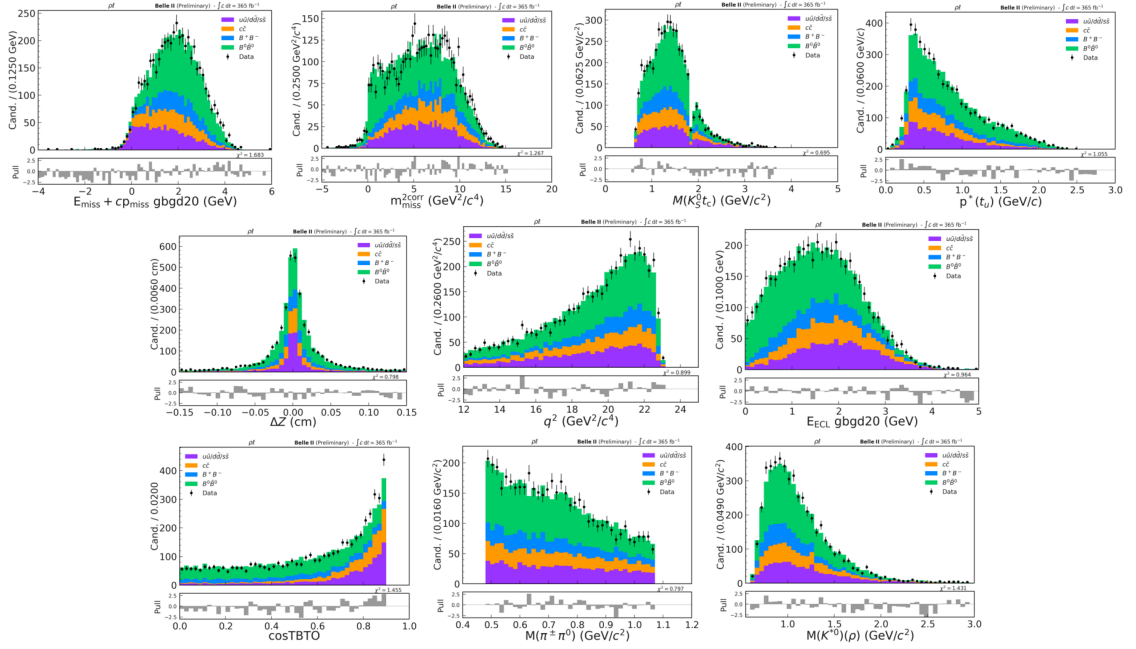


Figure C.9: Data/MC comparison of the training features for the $\rho\ell$ category (Belle II K_S^0 sideband).

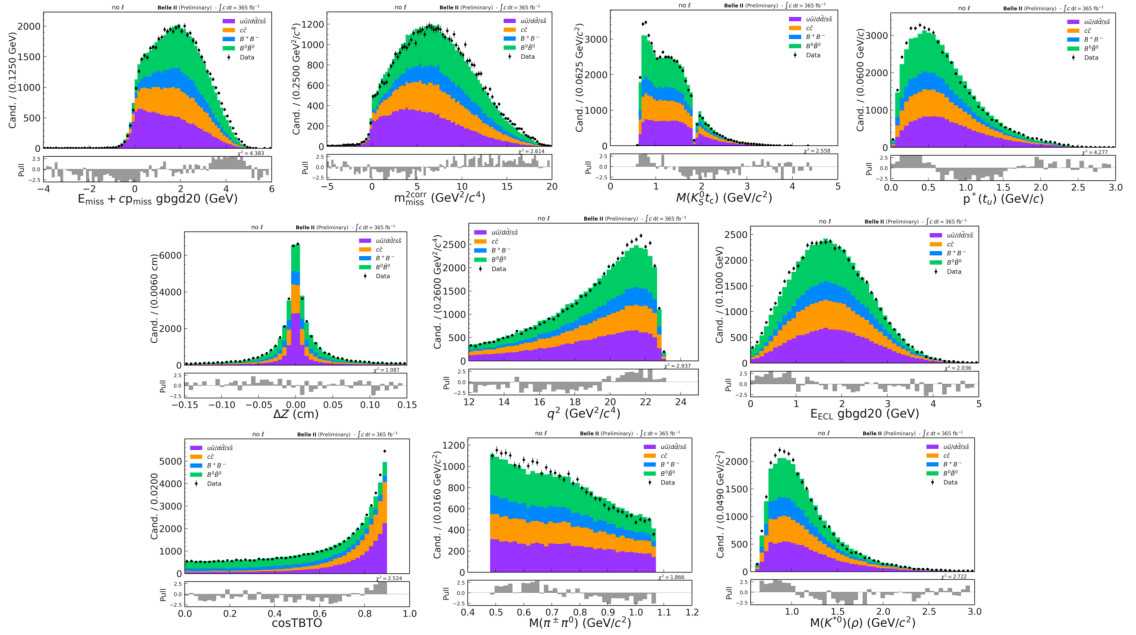


Figure C.10: Data/MC comparison of the training features for the $\ell 0$ category (Belle II K_S^0 sideband).

Appendix D

BDT training variables check in M_{bc} sideband

In all the plots shown here, the MC distributions are corrected with the “global scaling factors from Tab. 5.2, M_{bc} SB.

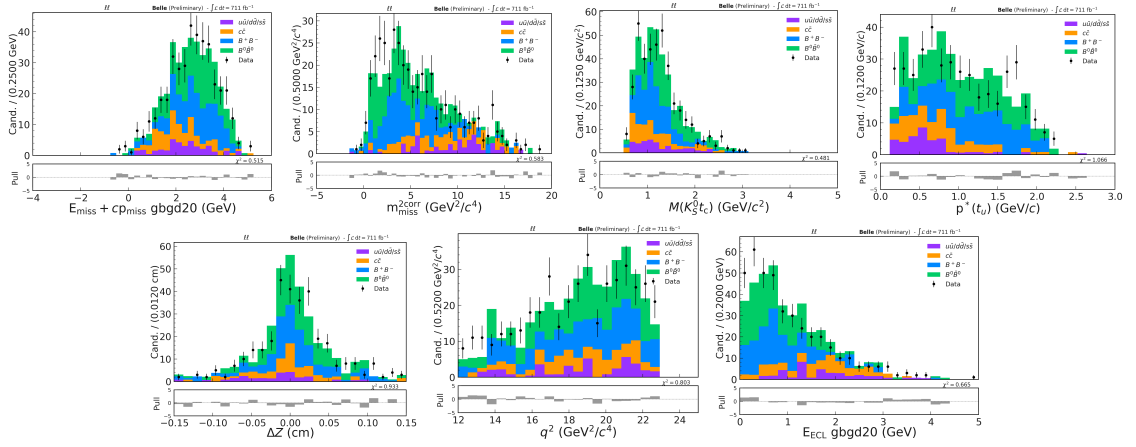


Figure D.1: Data/MC comparison of the training features for the $\ell\ell$ category (Belle M_{bc} sideband).

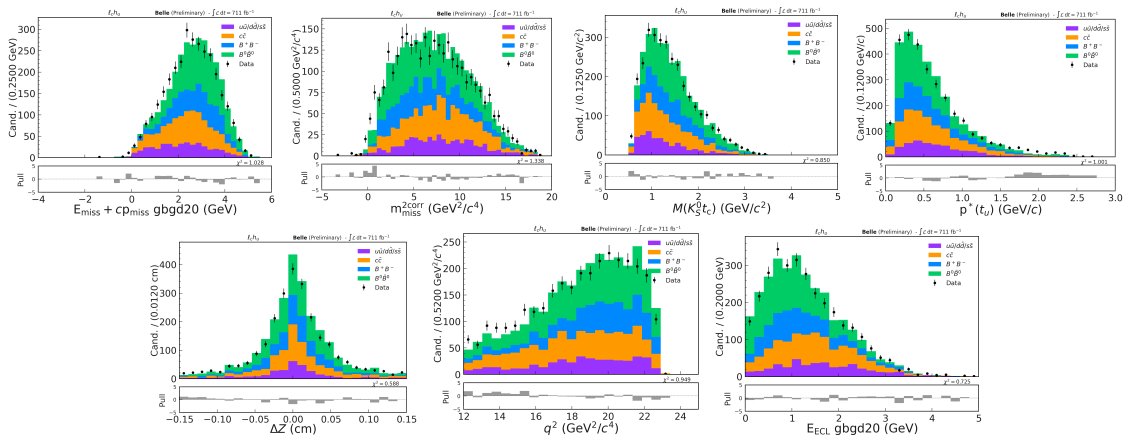


Figure D.2: Data/MC comparison of the training features for the $\ell_c h_u$ category (Belle M_{bc} sideband).

Appendix D. BDT training variables check in M_{bc} sideband

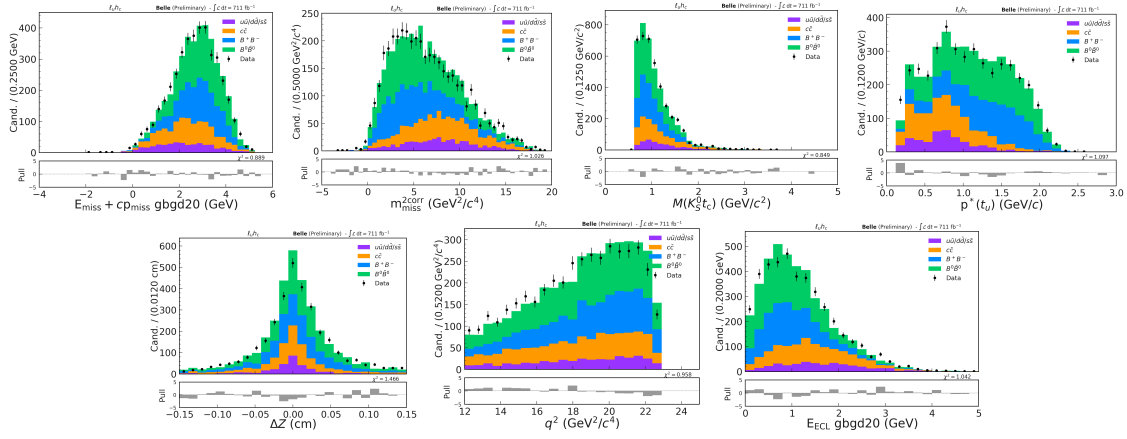


Figure D.3: Data/MC comparison of the training features for the $\ell_u h_c$ category (Belle M_{bc} sideband).

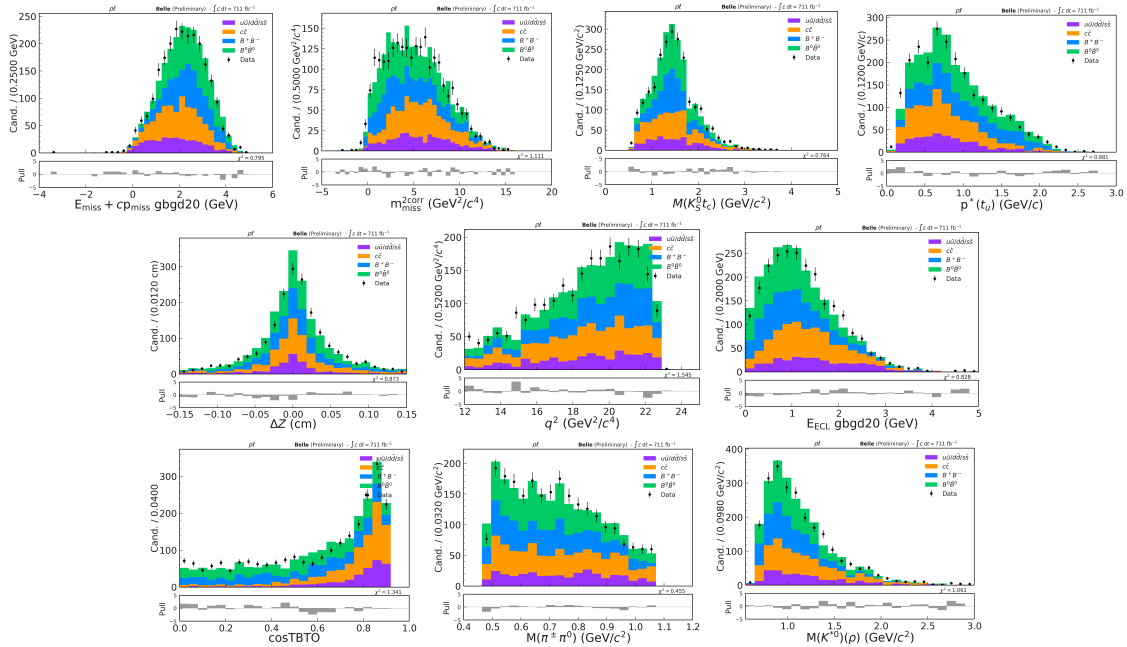


Figure D.4: Data/MC comparison of the training features for the $\rho \ell$ category (Belle M_{bc} sideband).

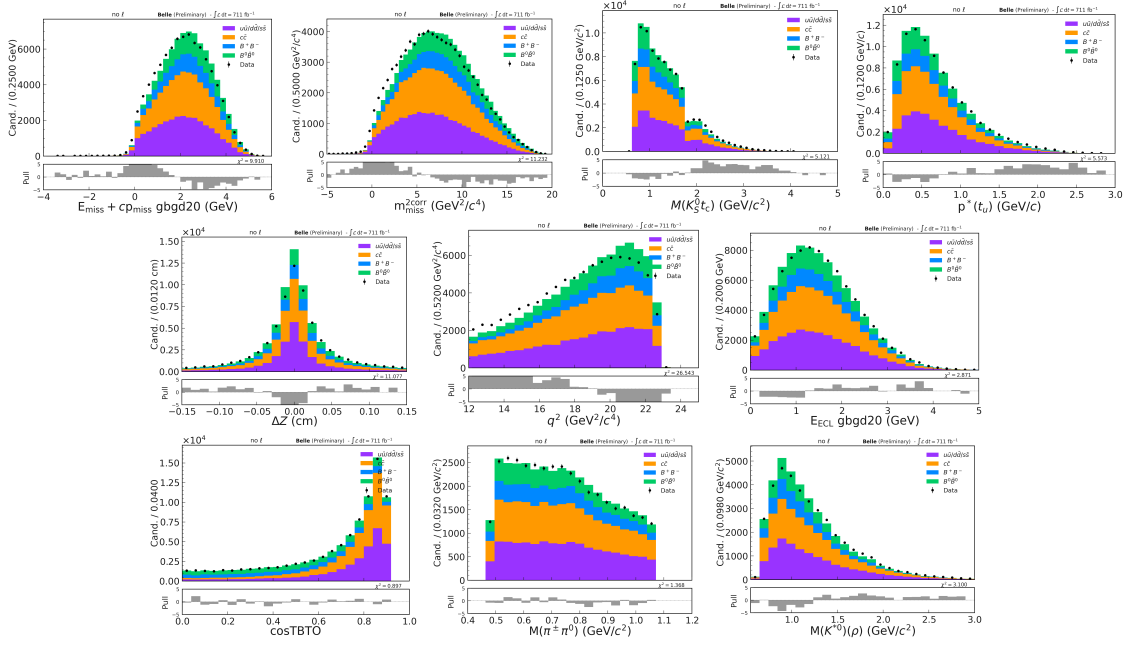


Figure D.5: Data/MC comparison of the training features for the $\ell 0$ category (Belle M_{bc} sideband).

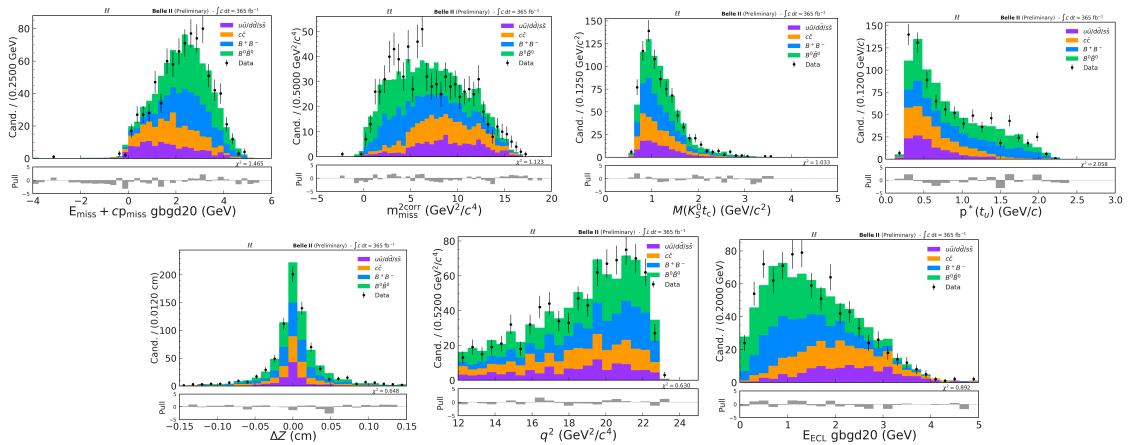
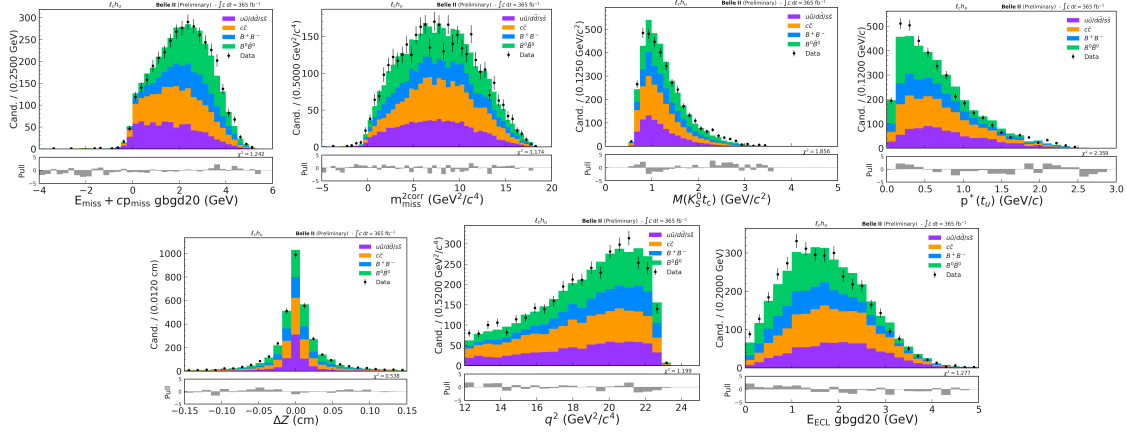
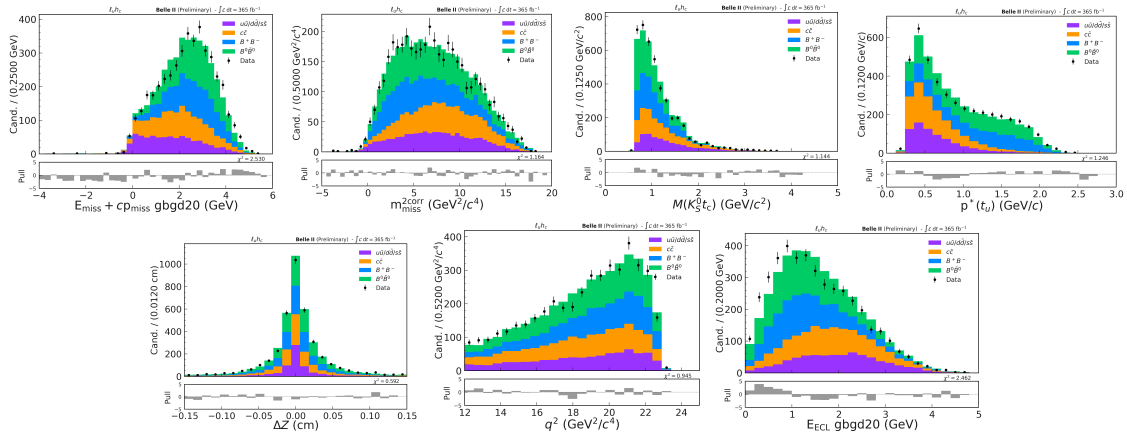


Figure D.6: Data/MC comparison of the training features for the $\ell\ell$ category (Belle II M_{bc} sideband).


 Figure D.7: Data/MC comparison of the training features for the $\ell_c h_u$ category (Belle II M_{bc} sideband).

 Figure D.8: Data/MC comparison of the training features for the $\ell_u h_c$ category (Belle II M_{bc} sideband).

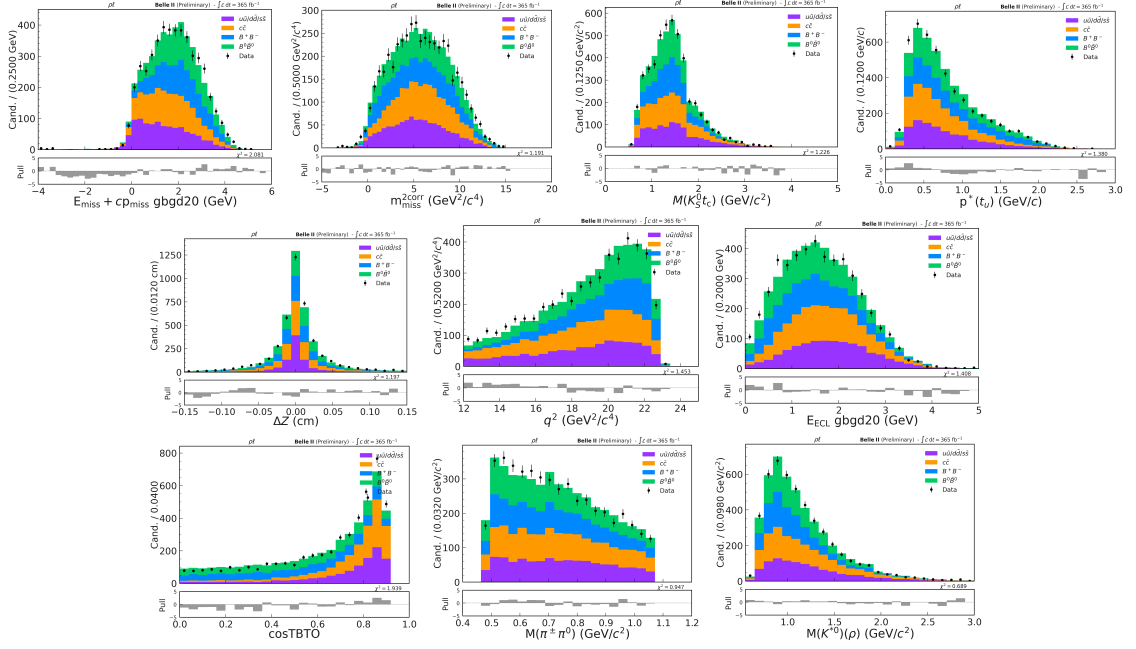


Figure D.9: Data/MC comparison of the training features for the $\rho\ell$ category (Belle II M_{bc} sideband).

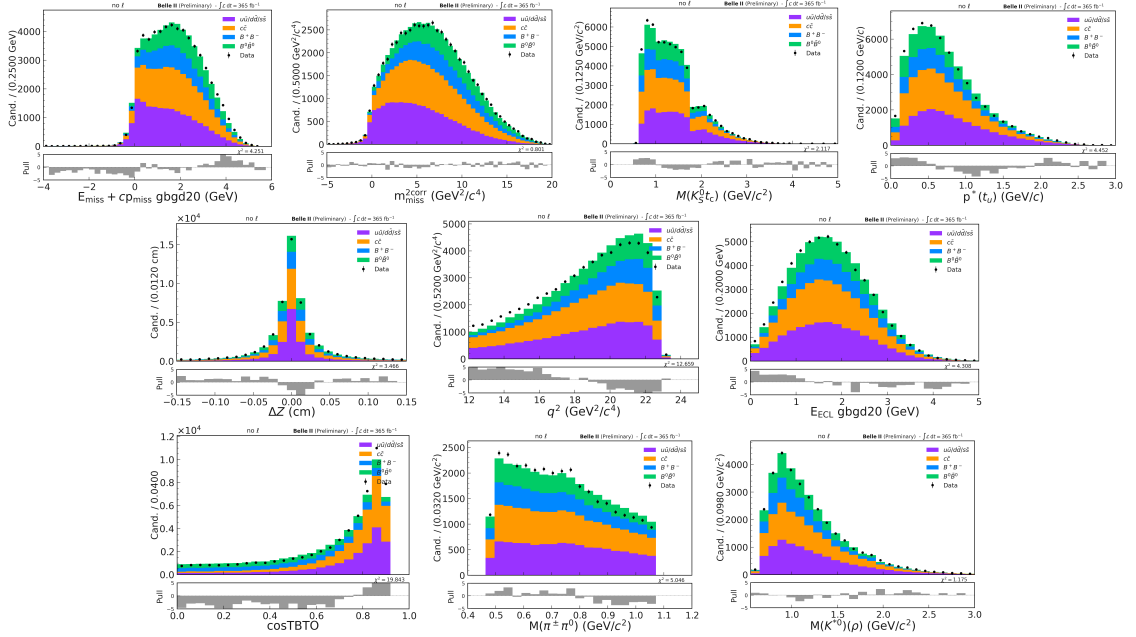


Figure D.10: Data/MC comparison of the training features for the $\ell 0$ category (Belle II M_{bc} sideband).

Appendix E

Signal composition in Belle II MC

Table E.1: Belle II MC true composition of the reconstructed $\ell\ell$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell\ell$ category.

$t_u \backslash t_c$	e	μ	π	$\pi \pi^0$	$\pi \pi^0 \pi^0$	K	rest	
e	0.23	0.20	0.01	0.03	0.01			0.48
μ	0.19	0.17	0.01	0.02	0.01			0.41
π	0.01	0.01						0.02
$\pi \pi^0$	0.03	0.02						0.06
$\pi \pi^0 \pi^0$	0.01	0.01						0.02
K								0.00
rest								0.01
	0.47	0.41	0.02	0.06	0.02	0.00	0.01	

Table E.2: Belle II MC true composition of the reconstructed $\ell_c h_u$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell_c h_u$ category.

$t_u \backslash t_c$	e	μ	π	$\pi \pi^0$	$\pi \pi^0 \pi^0$	K	rest	
e	0.05	0.04						0.10
μ	0.07	0.06		0.01				0.14
π	0.11	0.10	0.01	0.01				0.23
$\pi \pi^0$	0.17	0.15	0.01	0.02				0.36
$\pi \pi^0 \pi^0$	0.04	0.04						0.08
K	0.01	0.01						0.02
rest	0.03	0.03						0.07
	0.49	0.42	0.03	0.05	0.01	0.00	0.01	

Table E.3: Belle II MC true composition of the reconstructed $\ell_u h_c$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell_u h_c$ category.

$t_u \backslash t_c$	e	μ	π	$\pi \pi^0$	$\pi \pi^0 \pi^0$	K	rest	
e	0.04	0.07	0.11	0.17	0.04	0.01	0.04	0.48
μ	0.04	0.06	0.10	0.15	0.04	0.01	0.03	0.42
π			0.01	0.01				0.03
$\pi \pi^0$		0.01	0.01	0.02				0.05
$\pi \pi^0 \pi^0$				0.01				0.01
K								0.00
rest								0.01
	0.09	0.14	0.23	0.36	0.08	0.02	0.08	

Table E.4: Belle II MC true composition of the reconstructed $\rho \ell$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\rho \ell$ category.

$t_u \backslash t_c$	e	μ	π	$\pi \pi^0$	$\pi \pi^0 \pi^0$	K	rest	
e				0.13	0.06		0.02	0.23
μ	0.01	0.01	0.01	0.11	0.06		0.02	0.21
π	0.01	0.01		0.01				0.03
$\pi \pi^0$	0.13	0.11	0.01	0.04	0.02		0.01	0.33
$\pi \pi^0 \pi^0$	0.06	0.05	0.01	0.02	0.01			0.14
K								0.00
rest	0.02	0.02		0.01				0.06
	0.23	0.21	0.03	0.32	0.15	0.00	0.06	

Table E.5: Belle II MC true composition of the reconstructed $\ell 0$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell 0$ category.

$t_u \backslash t_c$	e	μ	π	$\pi \pi^0$	$\pi \pi^0 \pi^0$	K	rest	
e		0.01	0.01	0.03	0.01		0.01	0.07
μ	0.01	0.01	0.02	0.05	0.01		0.01	0.11
π	0.01	0.02	0.03	0.07	0.02		0.01	0.16
$\pi \pi^0$	0.03	0.05	0.07	0.17	0.05		0.04	0.41
$\pi \pi^0 \pi^0$	0.01	0.02	0.02	0.06	0.02		0.01	0.14
K								0.01
rest	0.01	0.01	0.02	0.04	0.01		0.01	0.10
	0.07	0.11	0.17	0.42	0.13	0.01	0.10	

Appendix F

Signal composition in Belle MC

Table F.1: Belle MC true composition of the reconstructed decay mode of the τ^+ (only one track taken). Integrated among all reconstructed categories. The table is normalized to unity. Last row (column), separated by a thicker line shows the integrated values for a given column (row).

reco true	e	μ	ρ	h	
e	0.16			0.05	0.22
μ		0.09	0.01	0.12	0.22
π				0.12	0.13
ρ			0.06	0.21	0.27
a_1			0.02	0.08	0.10
K				0.01	0.01
rest			0.01	0.05	0.06
	0.17	0.10	0.11	0.63	

Appendix F. Signal composition in Belle MC

Table F.2: Fraction of signal events in different logical categories, according to the number of wrong particles in the event (Belle signal MC). Each group corresponds to a reconstructed category; respectively $\ell\ell$, ℓ_ch_u , ℓ_uh_c , $\rho\ell$, $\ell 0$ categories.

$B_{\text{tag}} \backslash B_{\text{sig}}$	TM	Wrong K_S^0	Wrong $t_{c/u}$	Wrong $K_S^0, t_{c/u}$	Wrong t_u, t_c	FM	Total	
TM	0.57	0.01	0.04	0.00	0.00	0.00	0.62	$\ell\ell$
FM	0.34	0.01	0.03	0.00	0.00	0.00	0.38	
Total	0.91	0.02	0.07	0.00	0.00	0.00	1.00	
TM	0.55	0.01	0.02	0.00	0.00	0.00	0.58	ℓ_ch_u
FM	0.35	0.01	0.05	0.00	0.01	0.00	0.42	
Total	0.90	0.02	0.07	0.00	0.01	0.00	1.00	
TM	0.55	0.01	0.02	0.00	0.00	0.00	0.59	ℓ_uh_c
FM	0.34	0.01	0.06	0.00	0.00	0.00	0.41	
Total	0.89	0.02	0.08	0.01	0.00	0.00	1.00	
TM	0.39	0.01	0.20	0.01	0.02	0.00	0.62	$\rho\ell$
FM	0.18	0.00	0.16	0.01	0.02	0.00	0.38	
Total	0.57	0.01	0.36	0.02	0.04	0.00	1.00	
TM	0.47	0.01	0.00	0.00	0.05	0.00	0.54	$\ell 0$
FM	0.34	0.01	0.00	0.00	0.10	0.01	0.46	
Total	0.81	0.03	0.00	0.00	0.15	0.02	1.00	

Table F.3: Belle MC true composition of the reconstructed $\ell\ell$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell\ell$ category.

$t_u \backslash t_c$	e	μ	π	ρ	a_1	K	rest	
e	0.38	0.20	0.01	0.01				0.61
μ	0.22	0.12		0.01				0.36
π	0.01							0.01
ρ	0.01							0.01
a_1								0.01
K								0.00
rest								0.00
	0.62	0.34	0.01	0.02	0.01	0.00	0.00	

Table F.4: Belle MC true composition of the reconstructed $\ell_c h_u$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell_c h_u$ category.

$t_u \backslash t_c$	e	μ	π	ρ	a_1	K	rest	
e	0.05	0.03						0.09
μ	0.12	0.07						0.20
π	0.13	0.07						0.21
ρ	0.20	0.11		0.01				0.32
a_1	0.06	0.04						0.11
K	0.01							0.01
rest	0.04	0.02						0.06
	0.61	0.35	0.01	0.02	0.01	0.00	0.00	

Table F.5: Belle MC true composition of the reconstructed $\ell_u h_c$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell_u h_c$ category.

$t_u \backslash t_c$	e	μ	π	ρ	a_1	K	rest	
e	0.06	0.12	0.13	0.19	0.06	0.01	0.04	0.61
μ	0.03	0.07	0.07	0.11	0.04		0.02	0.35
π								0.01
ρ								0.02
a_1								0.01
K								0.00
rest								0.00
	0.09	0.20	0.20	0.31	0.11	0.01	0.07	

Table F.6: Belle MC true composition of the reconstructed $\rho\ell$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\rho\ell$ category.

$t_u \backslash t_c$	e	μ	π	ρ	a_1	K	rest	
e				0.19	0.08		0.03	0.31
μ				0.09	0.05		0.02	0.17
π								0.01
ρ	0.19	0.11		0.01	0.01			0.33
a_1	0.08	0.05						0.14
K								0.00
rest	0.03	0.02						0.05
	0.31	0.18	0.01	0.31	0.14	0.00	0.05	

Table F.7: Belle MC true composition of the reconstructed $\ell 0$ category (normalized to unity). Last row (column), separated by a thicker line shows the integrated values for a given column (row), or the total contribution of a given τ decay mode to t_c (t_u) within the $\ell 0$ category.

$t_u \backslash t_c$	e	μ	π	ρ	a_1	K	rest	
e	0.01	0.01	0.01	0.03	0.01		0.01	0.08
μ	0.01	0.03	0.03	0.06	0.02		0.01	0.17
π	0.01	0.03	0.03	0.06	0.02		0.01	0.17
ρ	0.03	0.06	0.06	0.13	0.05		0.03	0.36
a_1	0.01	0.02	0.02	0.05	0.02		0.01	0.15
K								0.01
rest	0.01	0.01	0.01	0.03	0.01		0.01	0.07
	0.07	0.17	0.17	0.36	0.14	0.01	0.08	

Appendix G

K_S^0 SR and SB comparison for Belle II MC

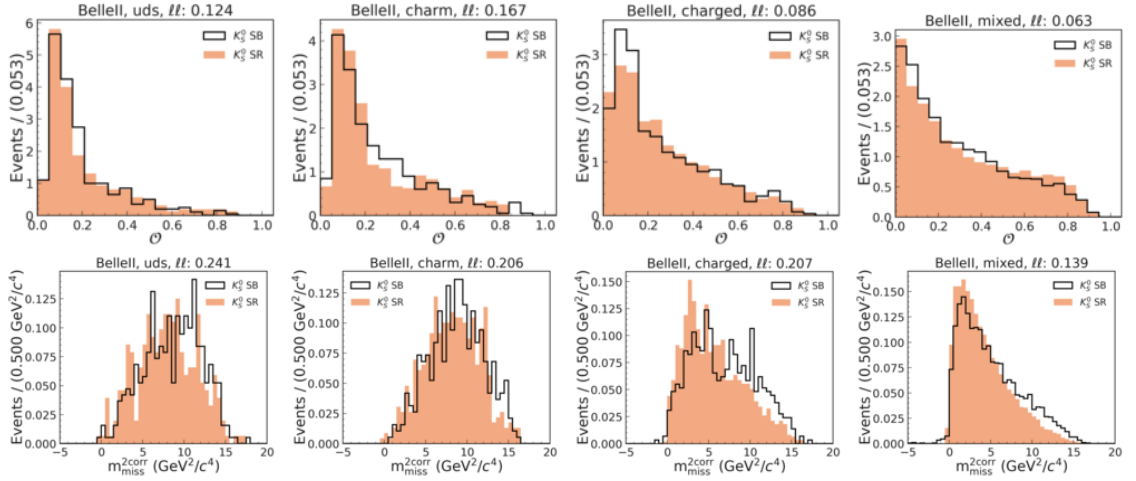


Figure G.1: BDT output and most important training feature for each background component ($\ell\ell$ category, Belle II MC).

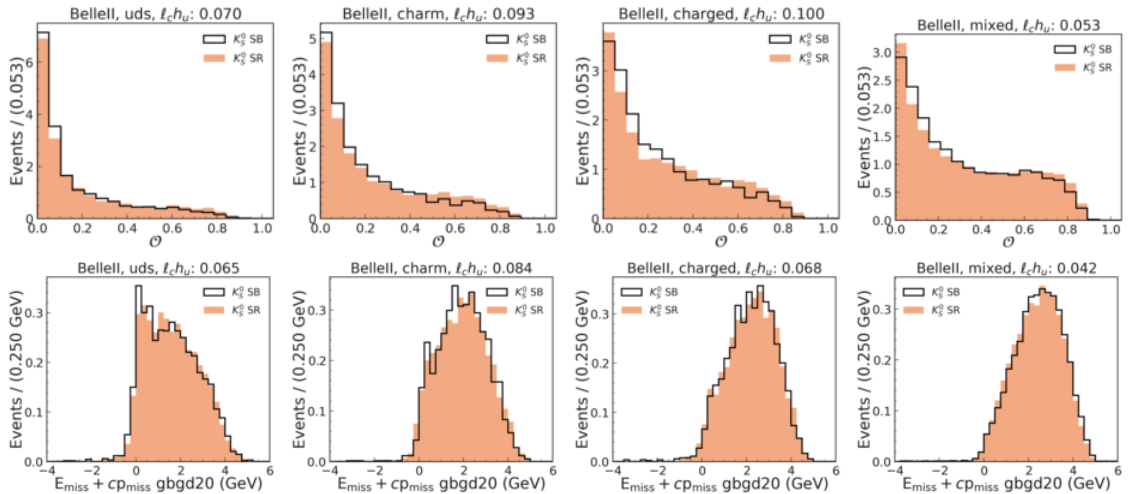


Figure G.2: BDT output and most important training feature for each background component ($\ell_c h_u$ category, Belle II MC).

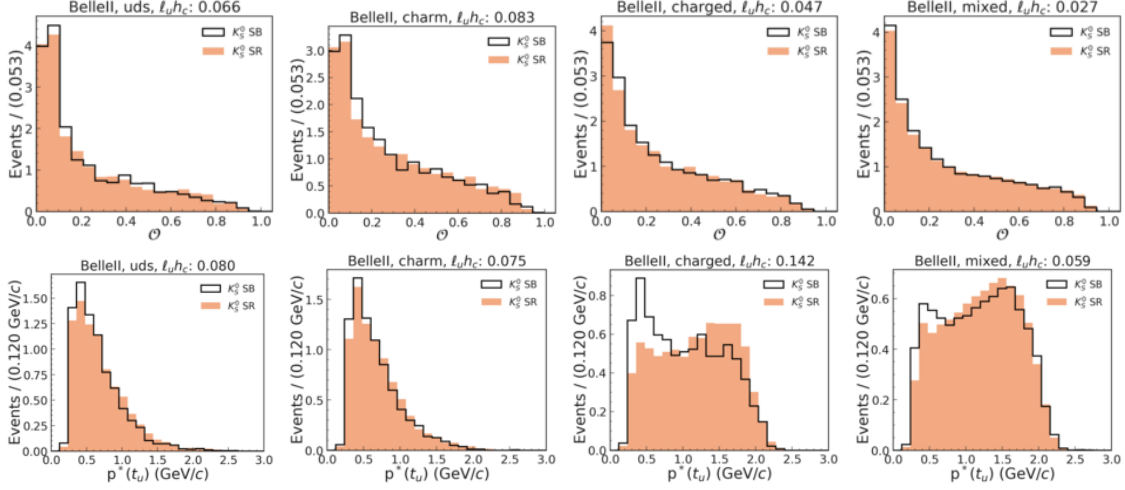


Figure G.3: BDT output and most important training feature for each background component ($\ell_u h_c$ category, Belle II MC).

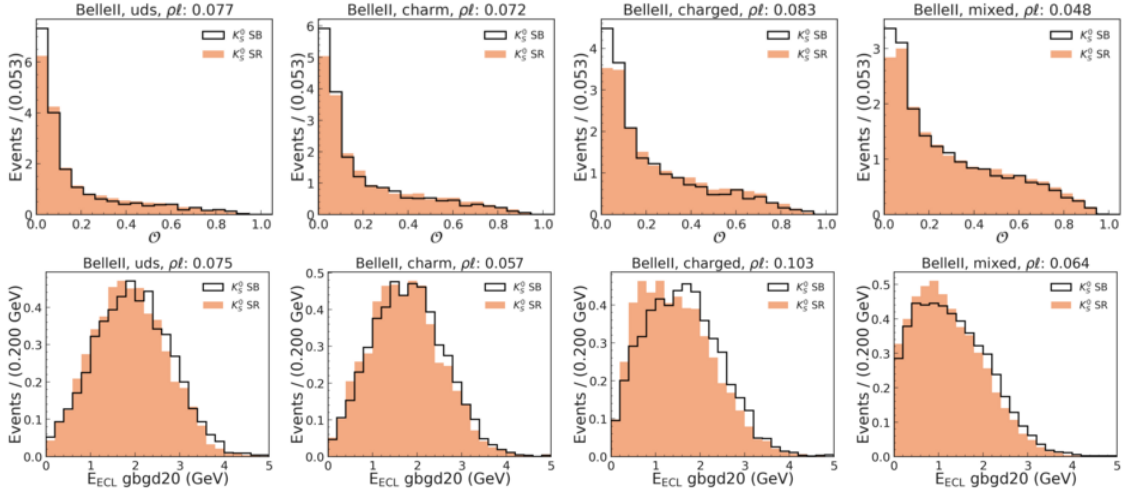


Figure G.4: BDT output and most important training feature for each background component ($\rho \ell$ category, Belle II MC).

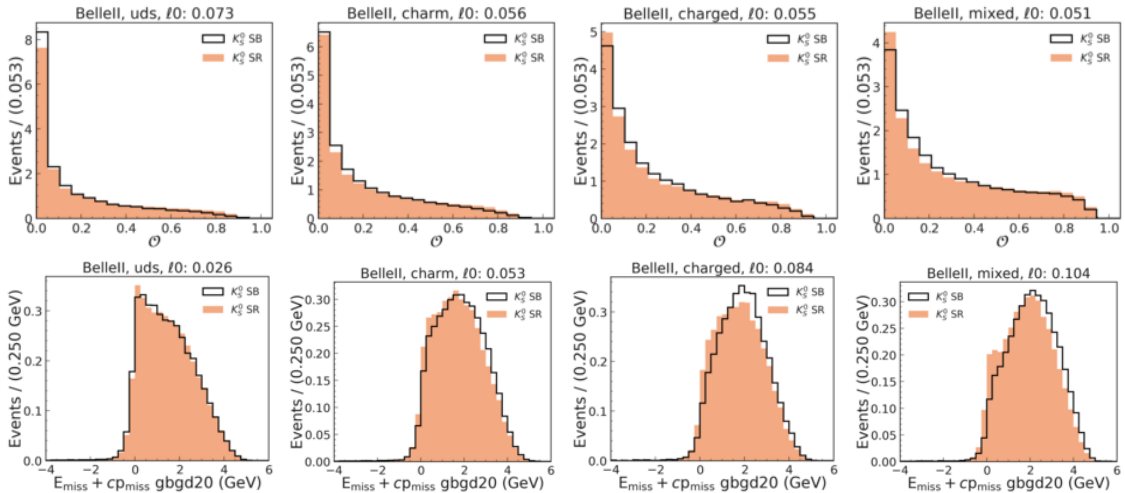


Figure G.5: BDT output and most important training feature for each background component ($\ell 0$ category, Belle II MC).

Appendix H

K_S^0 SR and SB comparison for Belle MC

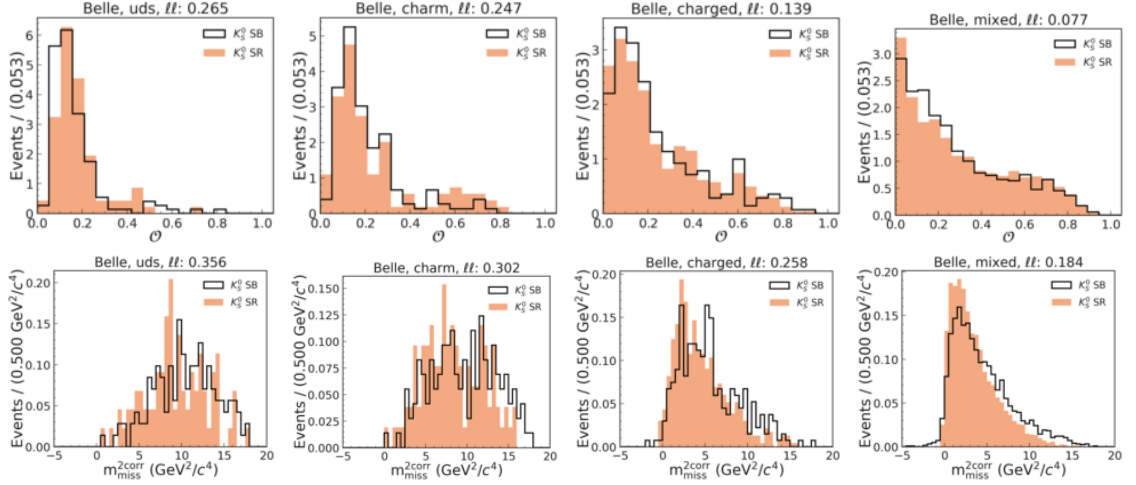


Figure H.1: BDT output and most important training feature for each background component ($\ell\ell$ category, Belle MC).

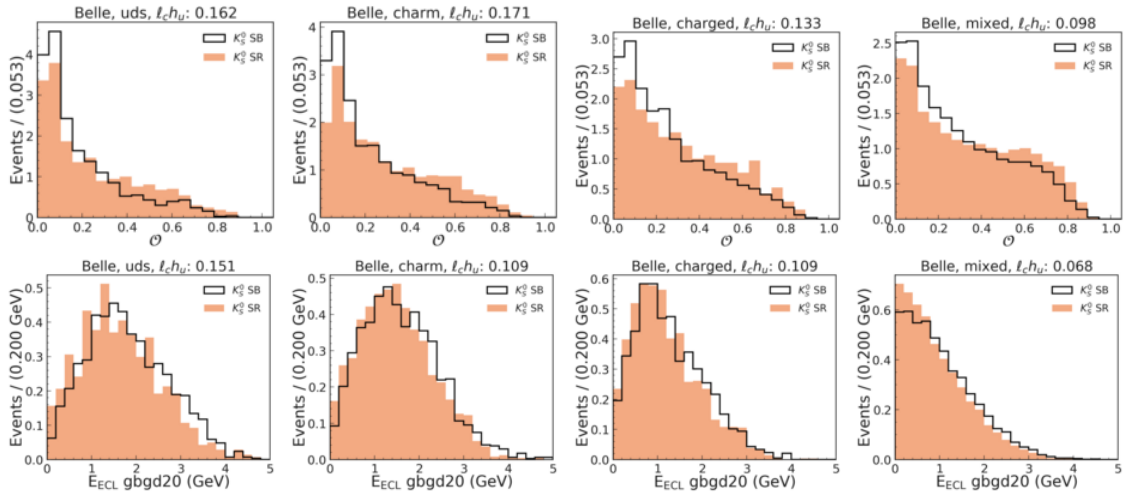


Figure H.2: BDT output and most important training feature for each background component ($\ell_c h_u$ category, Belle MC).

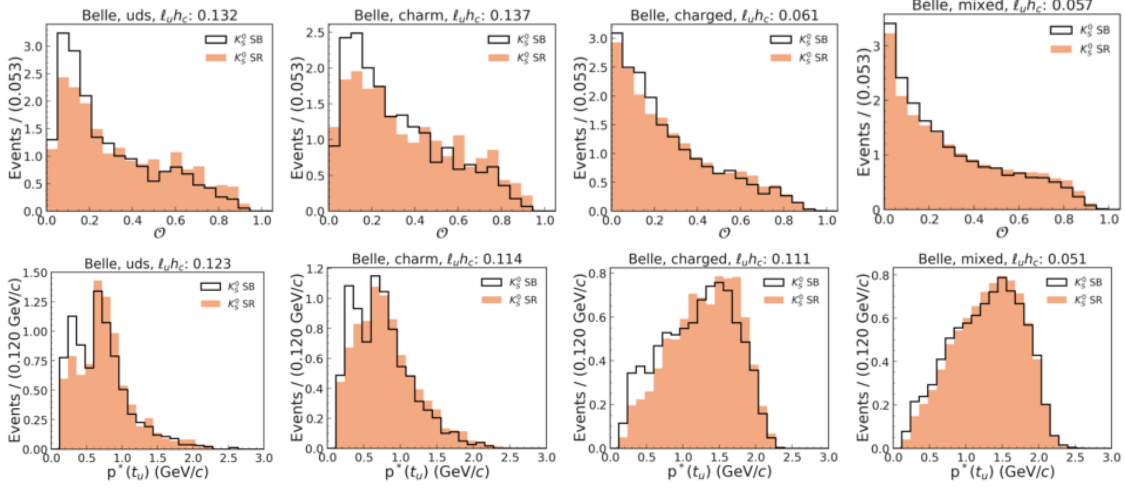


Figure H.3: BDT output and most important training feature for each background component ($\ell_u h_c$ category, Belle MC).

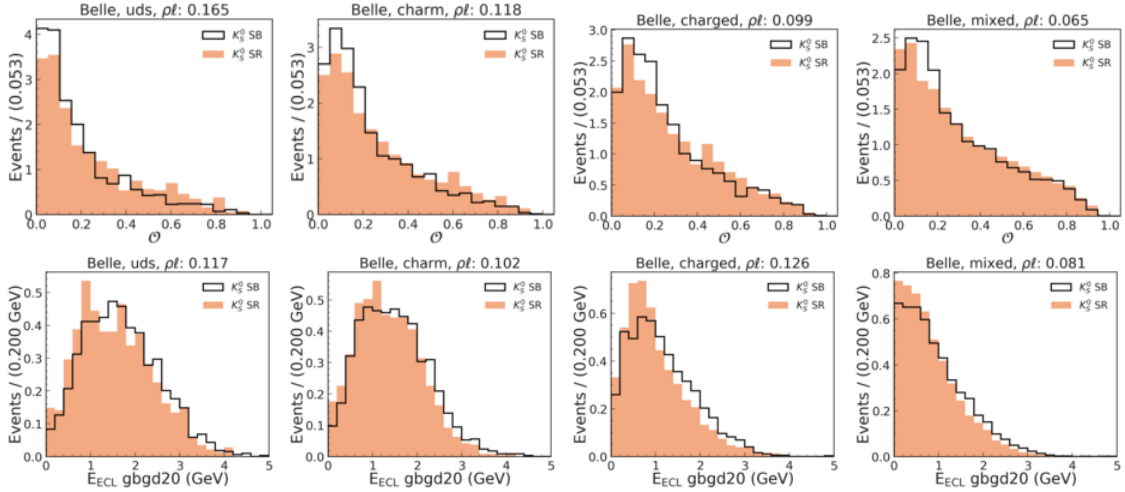


Figure H.4: BDT output and most important training feature for each background component ($\rho \ell$ category, Belle MC).

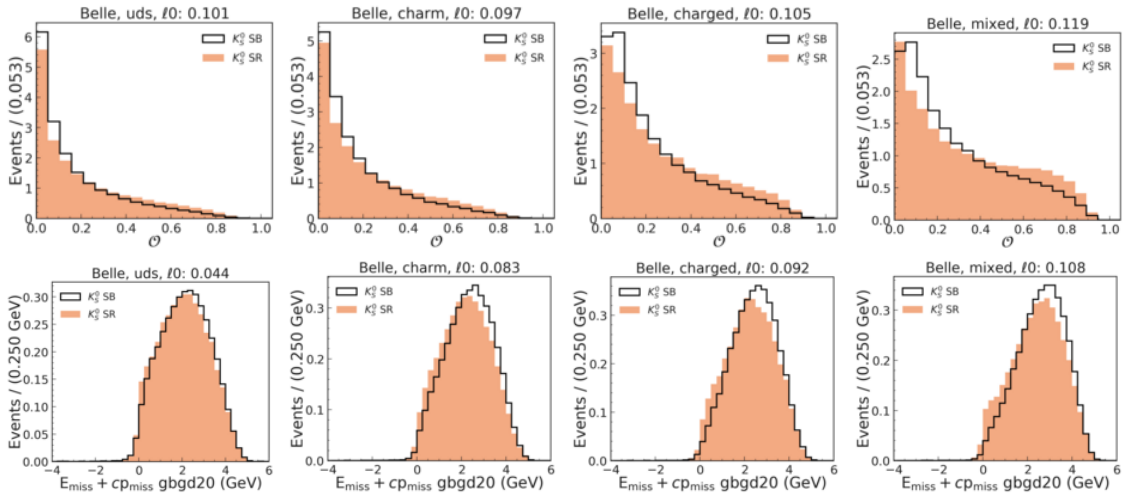


Figure H.5: BDT output and most important training feature for each background component ($\ell 0$ category, Belle MC).

Appendix I

Disjoint fits to Belle and Belle II MC samples

I.1 Fit to Belle MC

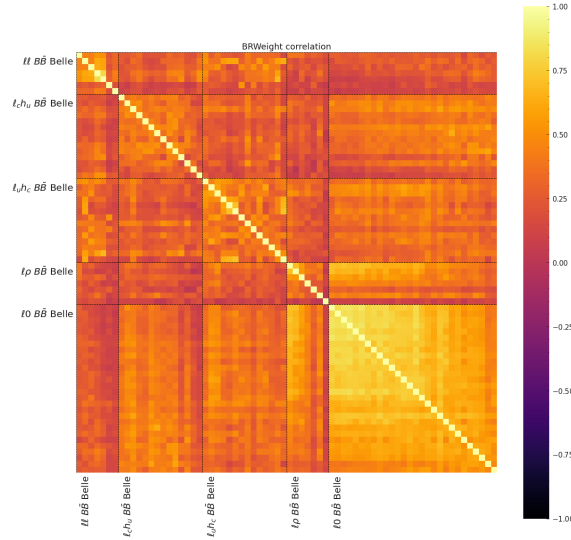


Figure I.1: Correlation matrix for the B branching fractions corrections computed with 1000 toys(Belle only).

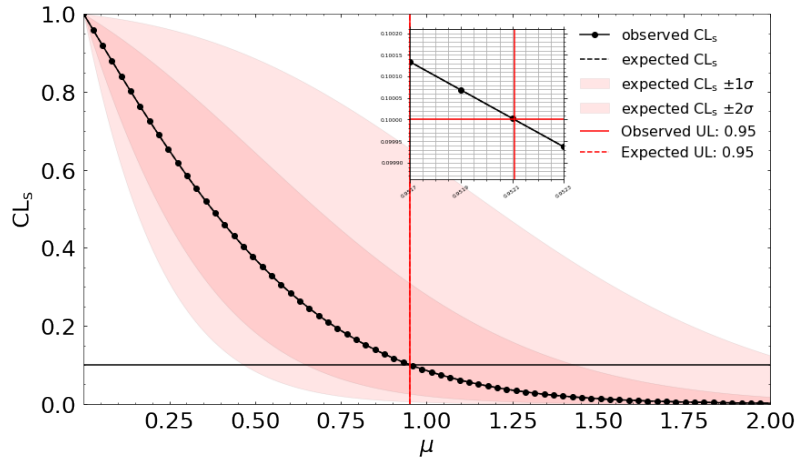


Figure I.2: CL_s scan from the fit to the Asimov dataset (Belle only).

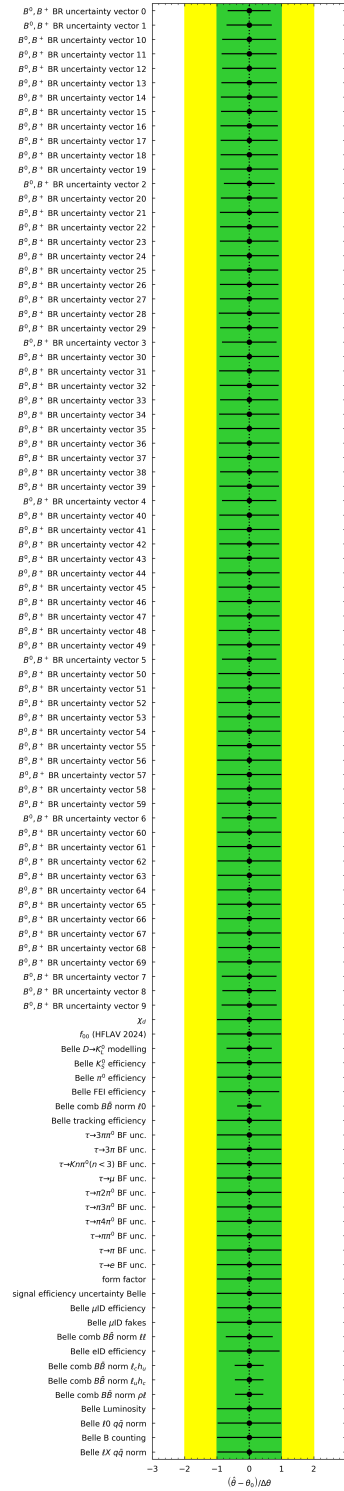


Figure I.3: Pull plot from the fit to the Asimov dataset (Belle only).

Appendix I. Disjoint fits to Belle and Belle II MC samples

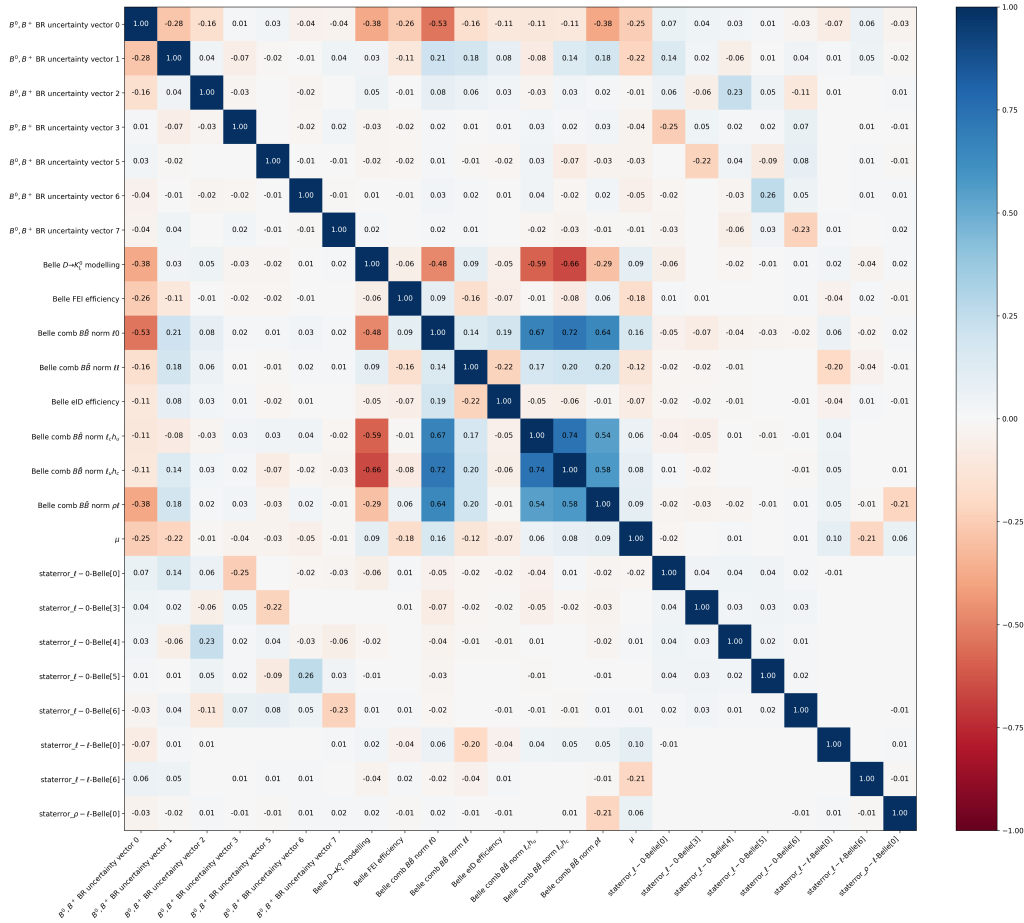


Figure I.4: Correlation matrix for the nuisance parameters in the fit model, obtained from the fit to the Asimov dataset (Belle MC). Parameters that have correlation less than 0.2 are not shown.

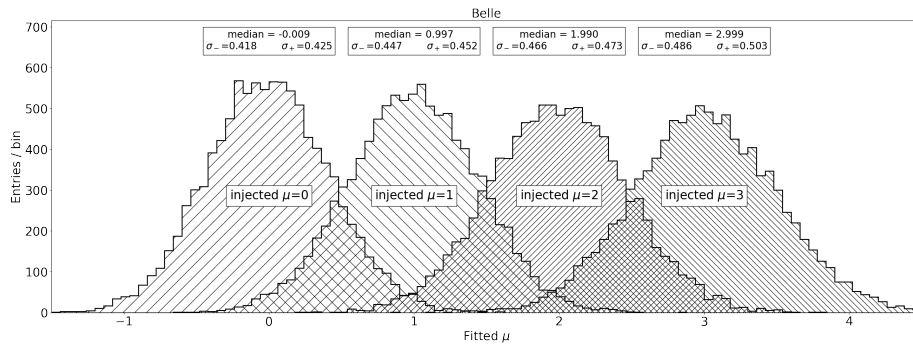


Figure I.5: The distribution of the fitted μ to the ensemble of the toy Asimov datasets for four injected signal strength hypotheses (10000 toys each). The median and the σ_+, σ_- values are shown, estimated from the 68% quantiles of the distribution. Belle-only fits.

Table I.1: Uncertainty on fitted signal strength μ for the Belle-only fit to the Asimov data.

Source	σ_μ
B Branching fractions uncertainties	0.23
MC stat	0.23
Belle FEI efficiency	0.09
Belle comb $B\bar{B}$ norm $\ell\ell$	0.05
Belle eID efficiency	0.04
Belle $D \rightarrow K_L^0$ modelling	0.04
Belle comb $B\bar{B}$ norm $\ell 0$	0.03
f_{00} (HFLAV 2024)	0.03
Belle comb $B\bar{B}$ norm $\rho\ell$	0.02
Belle comb $B\bar{B}$ norm $\ell_u h_c$	0.02
Belle Luminosity	0.02
Belle comb $B\bar{B}$ norm $\ell_c h_u$	0.01
Belle π^0 efficiency	0.01
Belle μ ID efficiency	0.01
Belle $\ell X q\bar{q}$ norm	0.01
Belle tracking efficiency	0.01
Belle K_S^0 efficiency	0.01
Belle $\ell 0 q\bar{q}$ norm	< 0.01
χ_d	< 0.01
Belle μ ID fakes	< 0.01
$\tau \rightarrow \pi 2\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi 3\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi \pi^0$ BF unc.	< 0.01
form factor	< 0.01
$\tau \rightarrow 3\pi \pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi$ BF unc.	< 0.01
$\tau \rightarrow e$ BF unc.	< 0.01
$\tau \rightarrow K n \pi^0 (n < 3)$ BF unc.	< 0.01
$\tau \rightarrow \pi 4\pi^0$ BF unc.	< 0.01
signal efficiency uncertainty Belle	< 0.01
$\tau \rightarrow \mu$ BF unc.	< 0.01
$\tau \rightarrow 3\pi$ BF unc.	< 0.01
Belle B counting	< 0.01
Total	0.36

I.2 Fit to Belle II MC

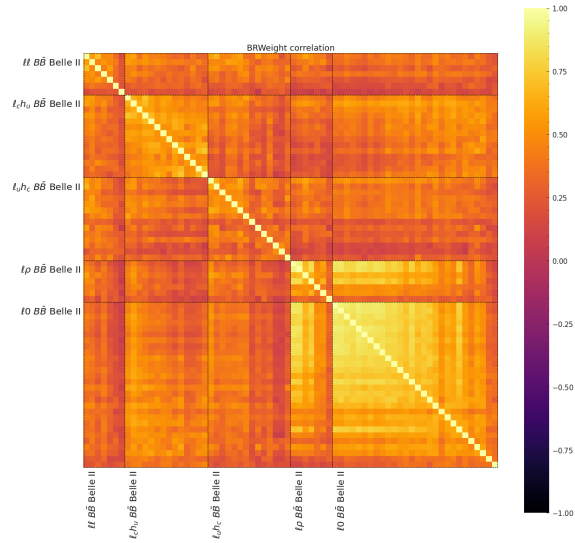


Figure I.6: Correlation matrix for the B branching fractions corrections computed with 1000 toys (Belle II only).

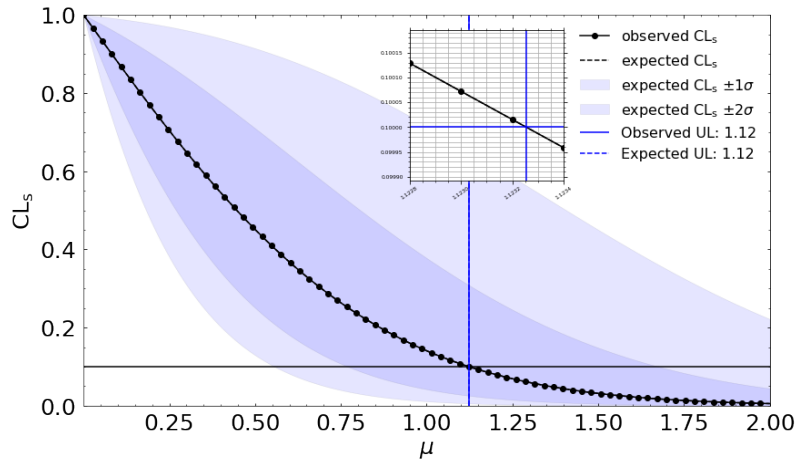


Figure I.7: CL_s scan from the fit to the Asimov dataset (Belle II only).

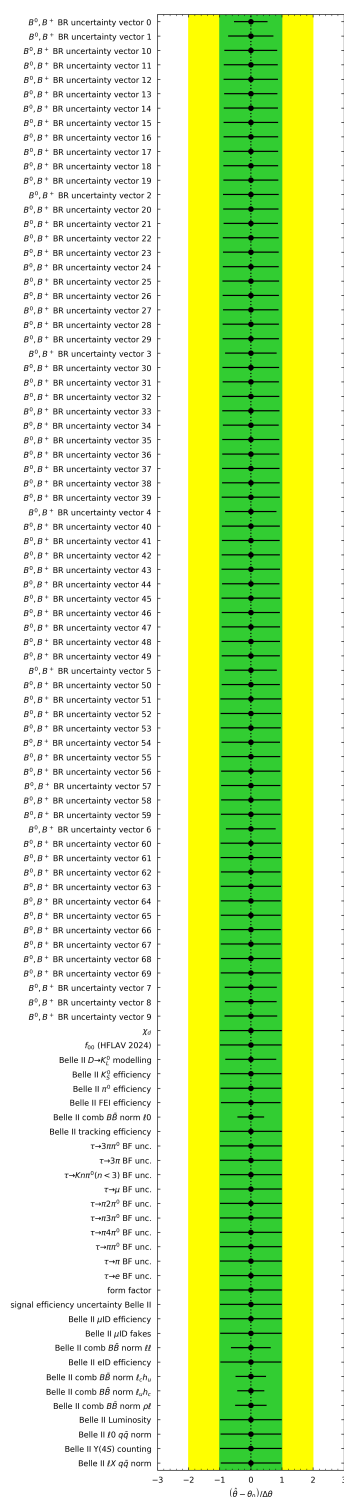


Figure I.8: Pull plot from the fit to the Asimov dataset (Belle II only).

Appendix I. Disjoint fits to Belle and Belle II MC samples

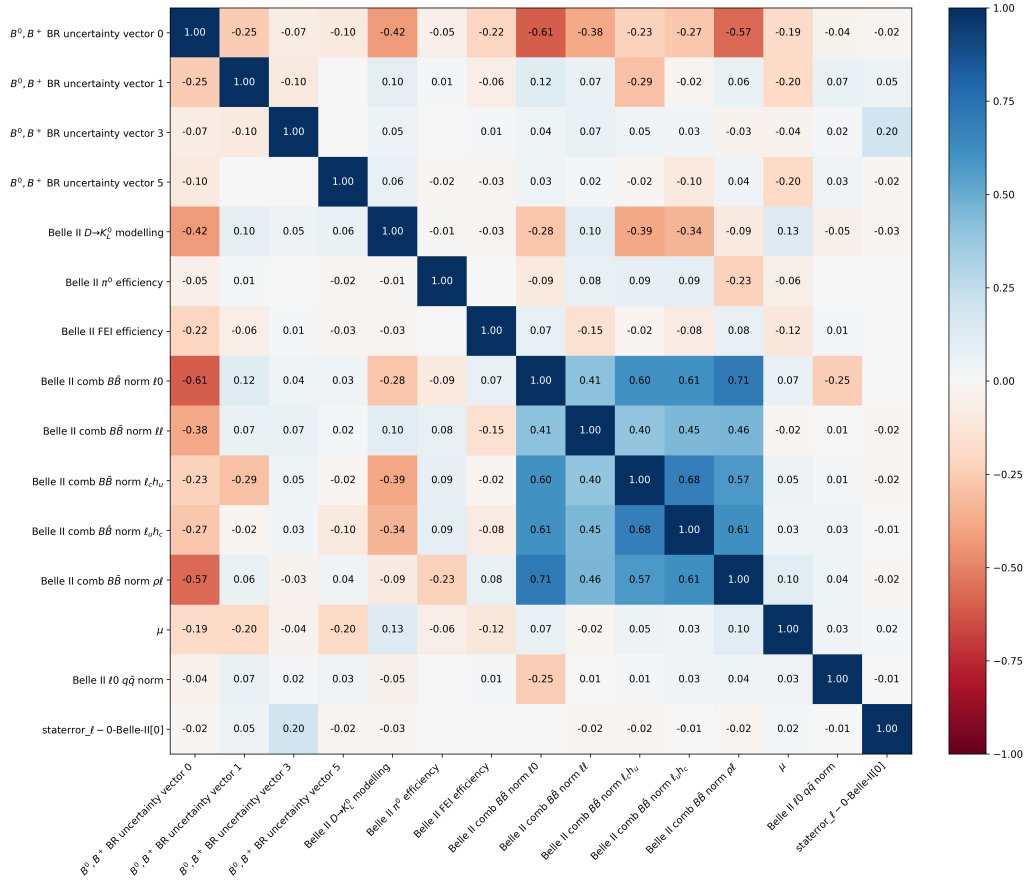


Figure I.9: Correlation matrix for the nuisance parameters in the fit model, obtained from the fit to the Asimov dataset (Belle II MC). Parameters that have correlation less then 0.2 are not shown.

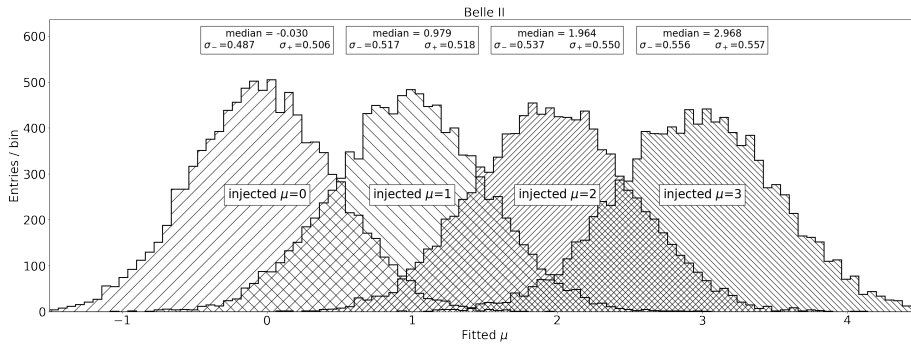


Figure I.10: The distribution of the fitted μ to the ensemble of the toy Asimov datasets for four injected signal strength hypotheses (10000 toys each). The median and the σ_+, σ_- values are shown, estimated from the 68% quantiles of the distribution. Belle II-only fits.

Table I.2: Uncertainty on fitted signal strength μ for the Belle II-only fit to the Asimov data.

Source	σ_μ
B Branching fractions uncertainties	0.34
MC stat	0.22
Belle II FEI efficiency	0.07
Belle II $D \rightarrow K_L^0$ modelling	0.07
f_{00} (HFLAV 2024)	0.04
Belle II π^0 efficiency	0.04
Belle II eID efficiency	0.04
Belle II comb $B\bar{B}$ norm $\rho\ell$	0.03
Belle II K_S^0 efficiency	0.03
Belle II $\ell 0$ $q\bar{q}$ norm	0.02
Belle II comb $B\bar{B}$ norm $\ell 0$	0.02
Belle II comb $B\bar{B}$ norm $\ell_c h_u$	0.02
Belle II μ ID fakes	0.01
Belle II comb $B\bar{B}$ norm $\ell\ell$	0.01
Belle II comb $B\bar{B}$ norm $\ell_u h_c$	0.01
Belle II Luminosity	0.01
Belle II tracking efficiency	0.01
Belle II μ ID efficiency	< 0.01
Belle II ℓX $q\bar{q}$ norm	< 0.01
form factor	< 0.01
$\tau \rightarrow 3\pi$ BF unc.	< 0.01
$\tau \rightarrow \pi 2\pi^0$ BF unc.	< 0.01
$\tau \rightarrow 3\pi\pi^0$ BF unc.	< 0.01
$\tau \rightarrow Kn\pi^0 (n < 3)$ BF unc.	< 0.01
signal efficiency uncertainty Belle II	< 0.01
$\tau \rightarrow \pi 3\pi^0$ BF unc.	< 0.01
$\tau \rightarrow \pi\pi^0$ BF unc.	< 0.01
χ_d	< 0.01
$\tau \rightarrow \pi$ BF unc.	< 0.01
$\tau \rightarrow e$ BF unc.	< 0.01
$\tau \rightarrow \mu$ BF unc.	< 0.01
Belle II $\Upsilon(4S)$ counting	< 0.01
$\tau \rightarrow \pi 4\pi^0$ BF unc.	< 0.01
Total	0.43

Appendix J

“Post-fit” distributions of the BDT input features in signal enhanced region

The binned weight is extracted in post-fit $\mathcal{O}'$ and applied to the pre-fit distributions of BDT input features in order to approximate post-fit distributions of them. A signal-enhanced region $\mathcal{O}' > 0.7$ is shown in this Appendix. As the post-fit uncertainty of the fit variable can not be translated into the uncertainties of the histograms in other dimensions, only the data uncertainty is used to calculate the residuals. Therefore, the significance of the residuals is overestimated. Nevertheless, a good overall data-simulation agreement is observed. Signal in the plots corresponds to the measured branching fraction $\mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = 0.21 \times 10^{-3}$. The excluded signal corresponds to the $\mathcal{B}_{UL}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = 0.84 \times 10^{-3}$ and is multiplied by ten for better visibility.

Appendix J. “Post-fit” distributions of the BDT input features in signal enhanced region

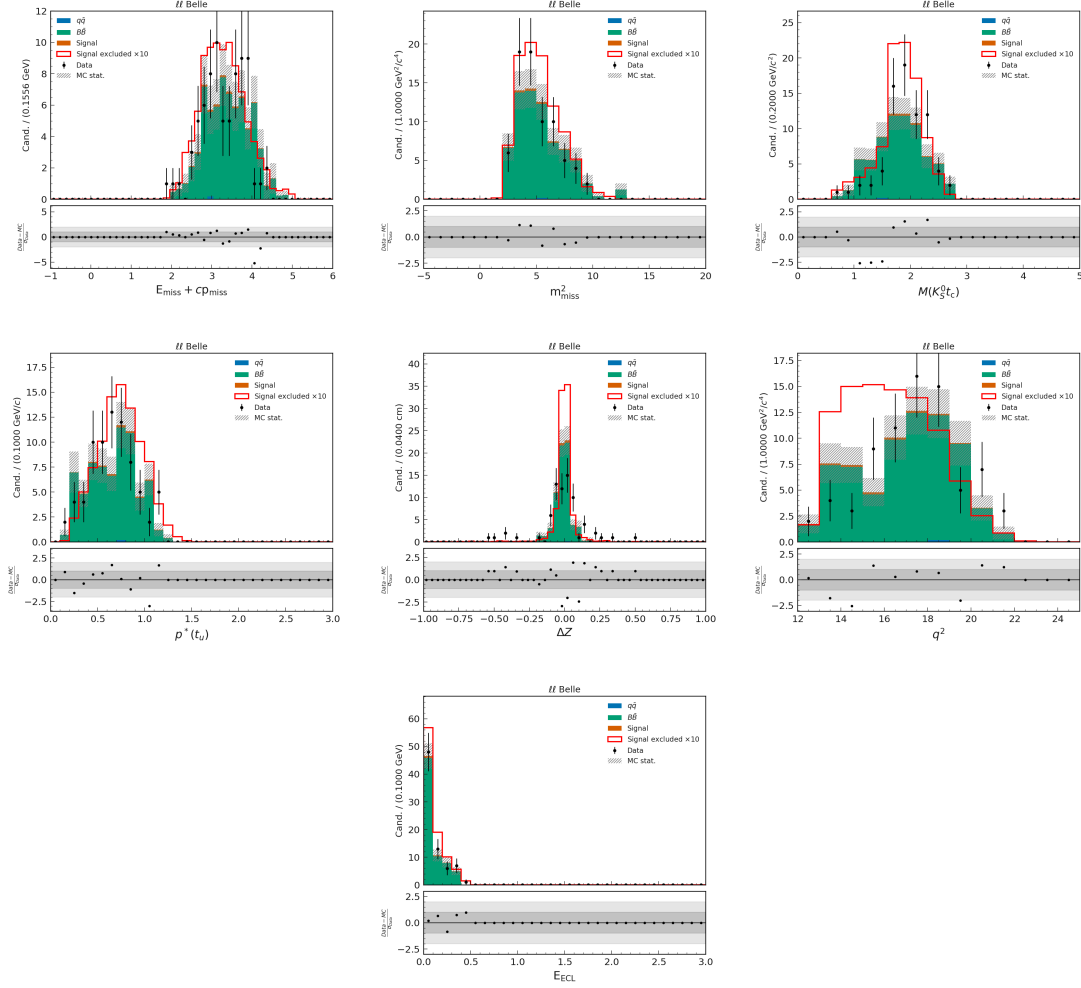


Figure J.1: Data/MC comparison of the training features for the $\ell\ell$ category (Belle signal enhanced region).

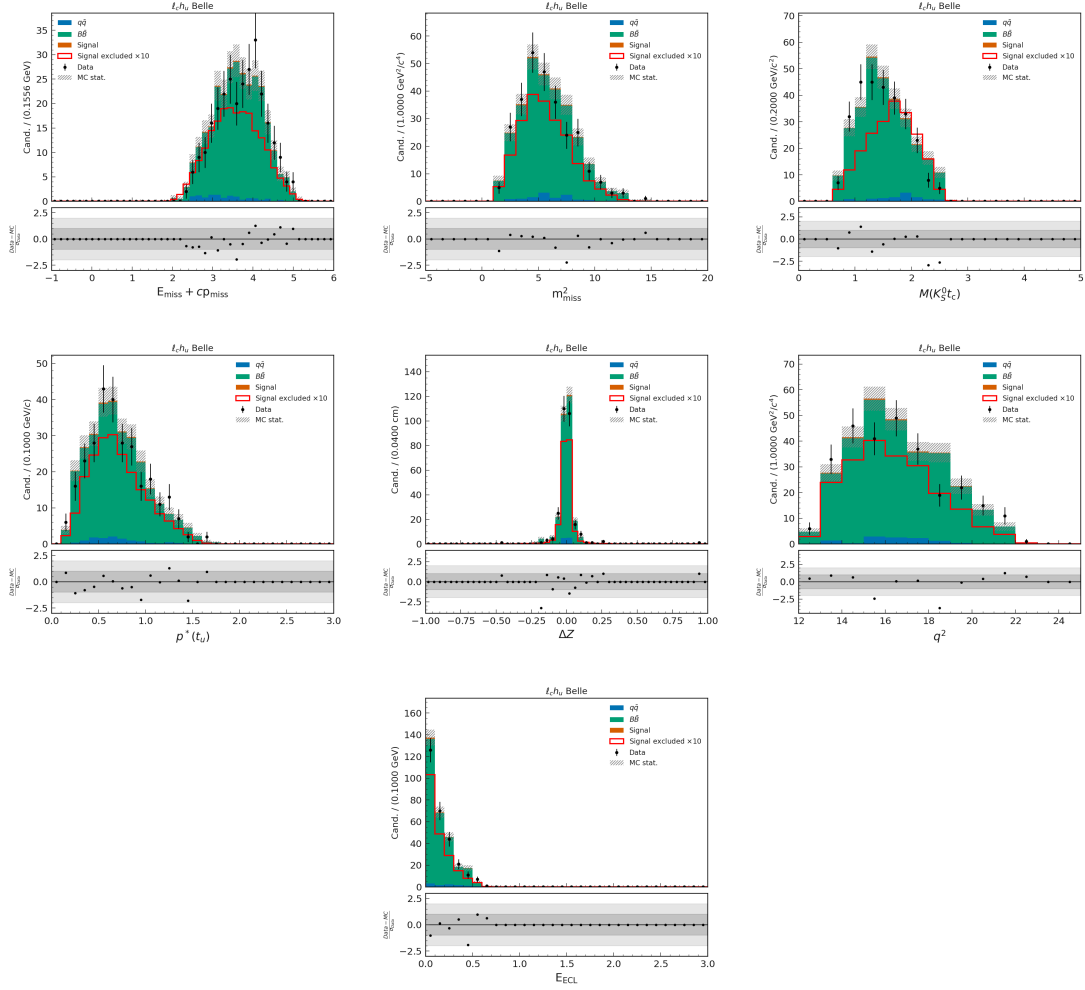


Figure J.2: Data/MC comparison of the training features for the $\ell_c h_u$ category (Belle signal enhanced region).

Appendix J. “Post-fit” distributions of the BDT input features in signal enhanced region

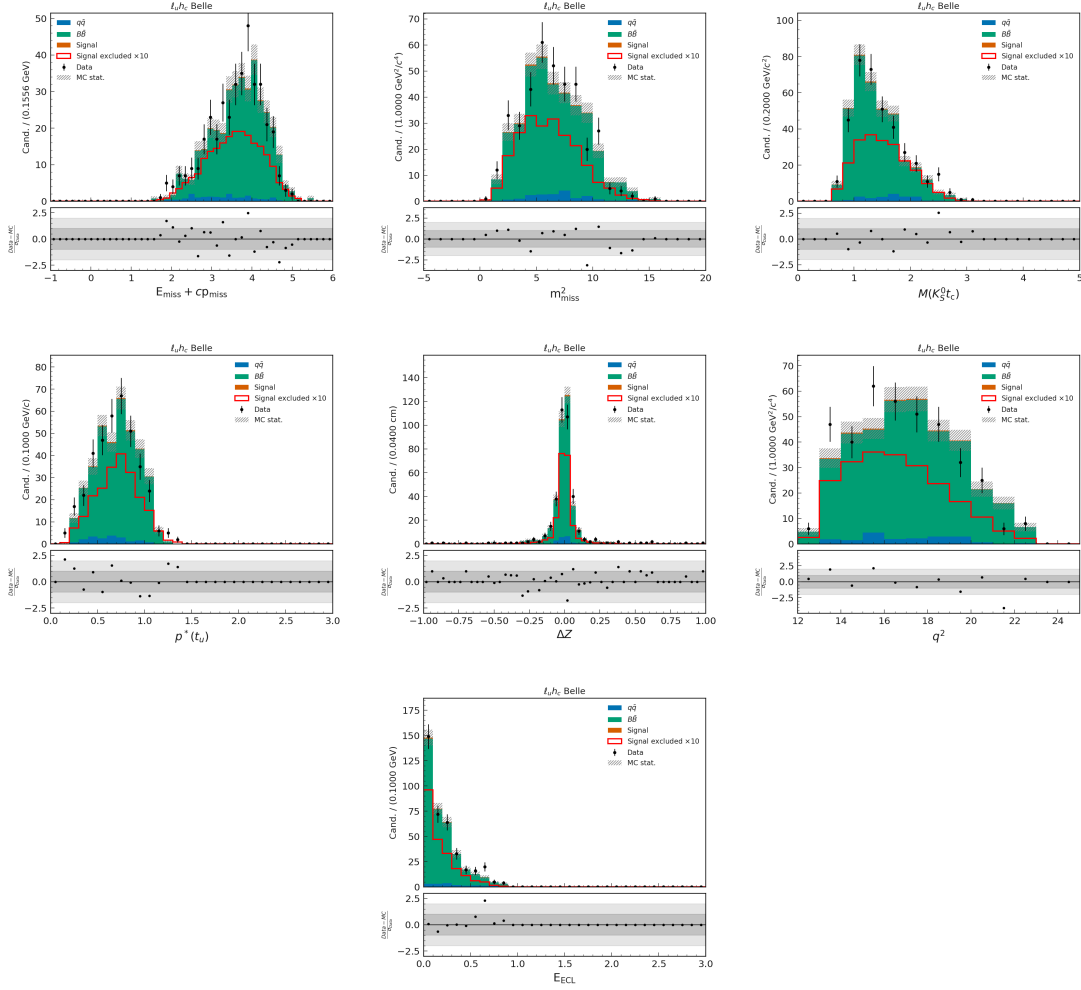


Figure J.3: Data/MC comparison of the training features for the $\ell_u h_c$ category (Belle signal enhanced region).

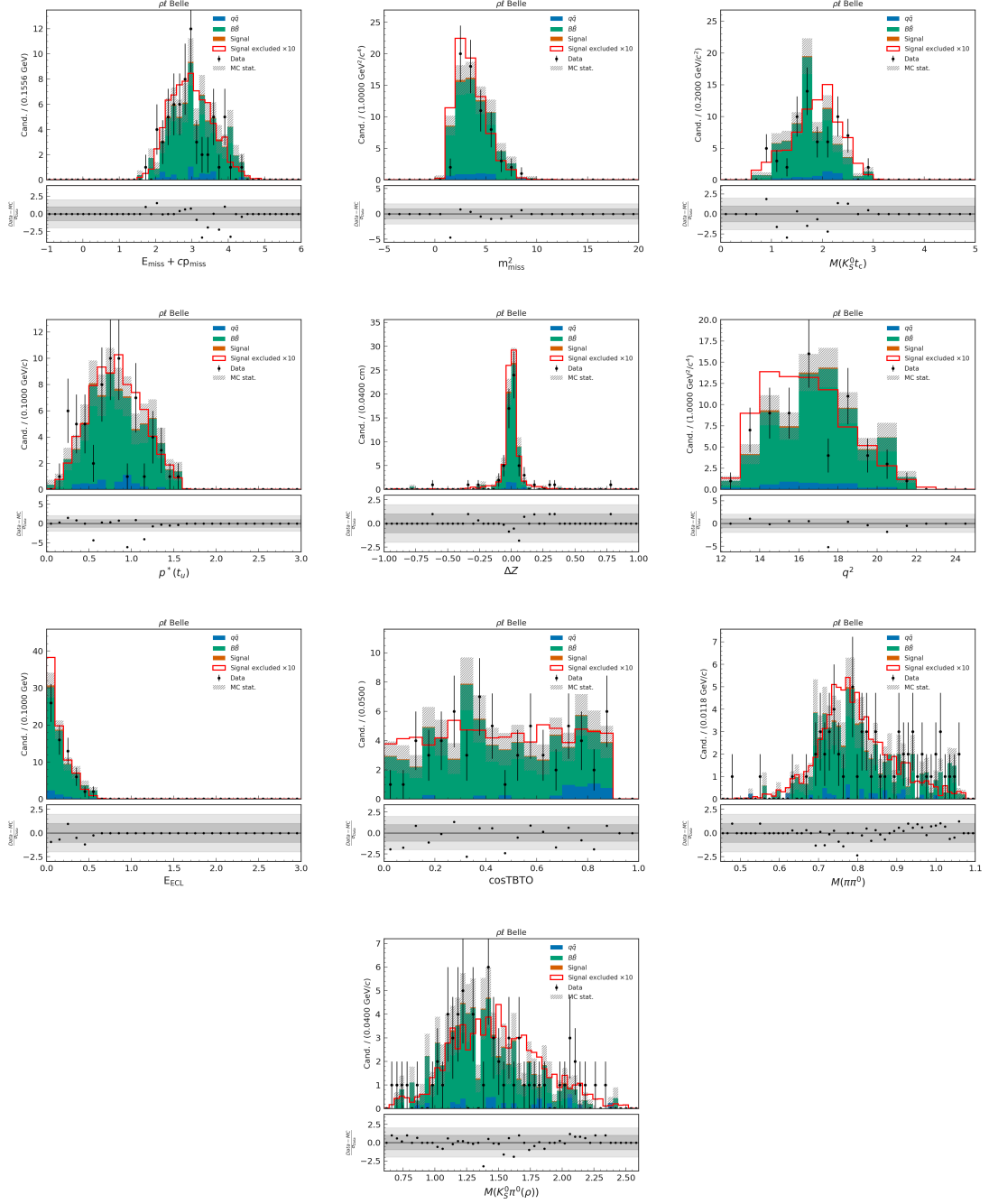


Figure J.4: Data/MC comparison of the training features for the $\rho\ell$ category (Belle signal enhanced region).

Appendix J. “Post-fit” distributions of the BDT input features in signal enhanced region

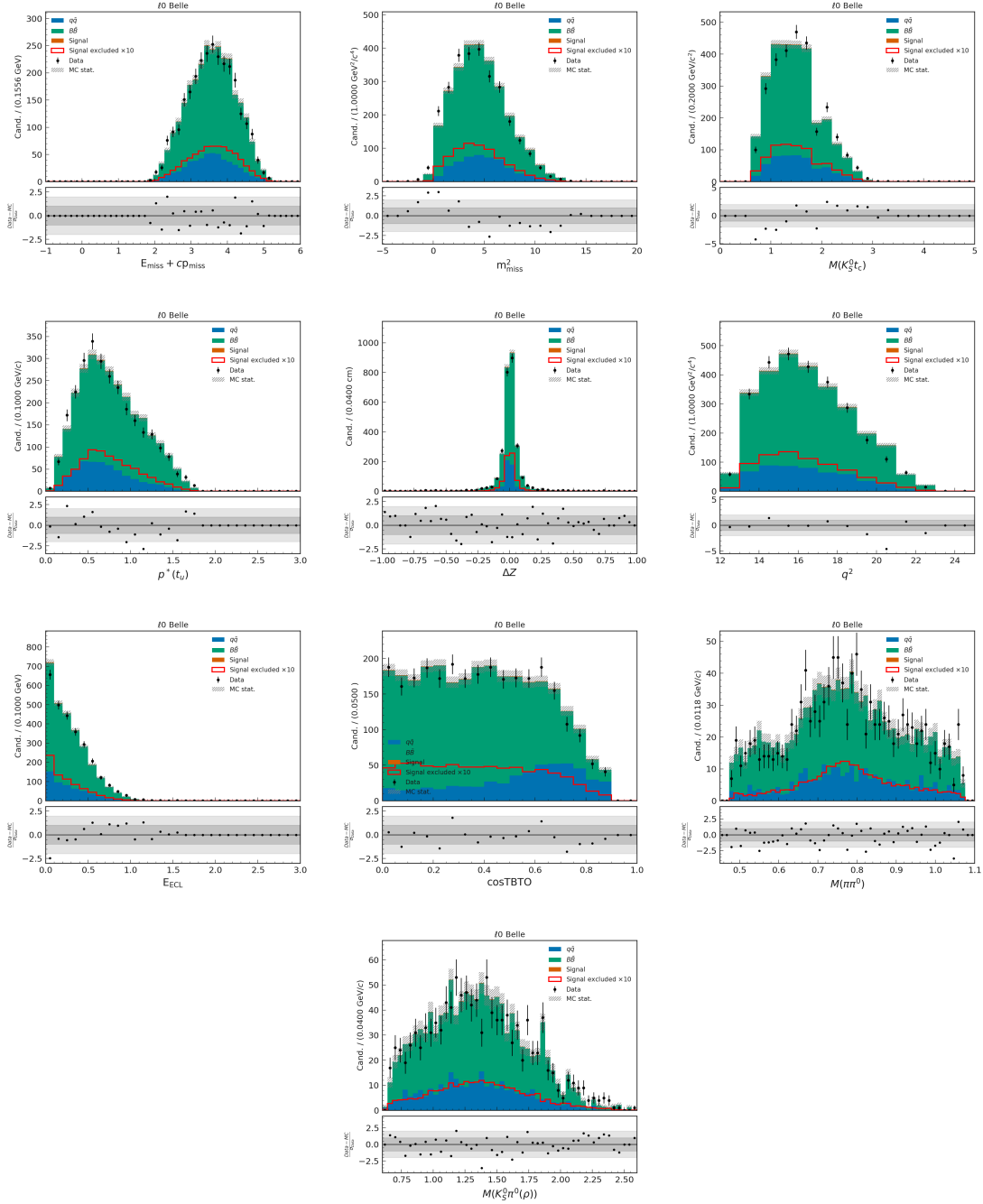


Figure J.5: Data/MC comparison of the training features for the ℓ_0 category (Belle signal enhanced region).

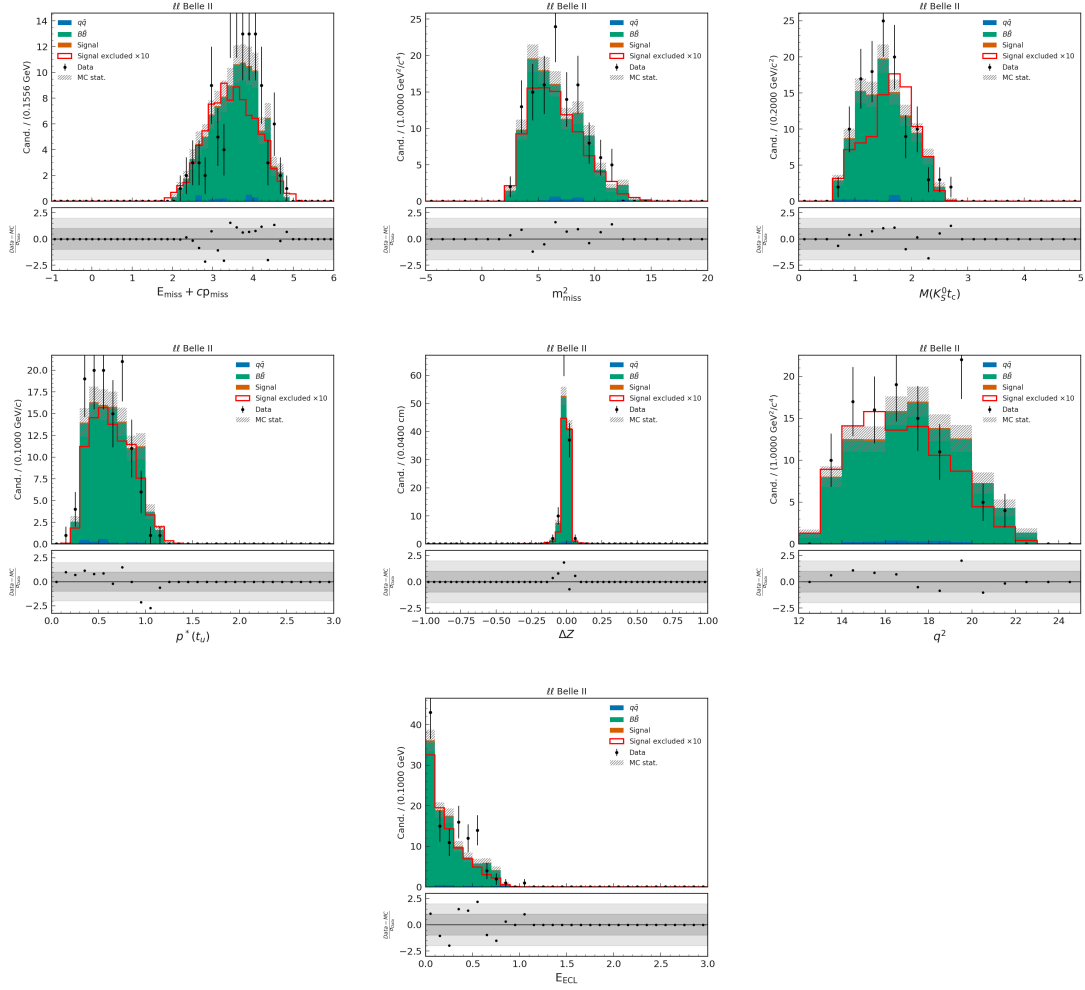


Figure J.6: Data/MC comparison of the training features for the $\ell\ell$ category (Belle II signal enhanced region).

Appendix J. “Post-fit” distributions of the BDT input features in signal enhanced region

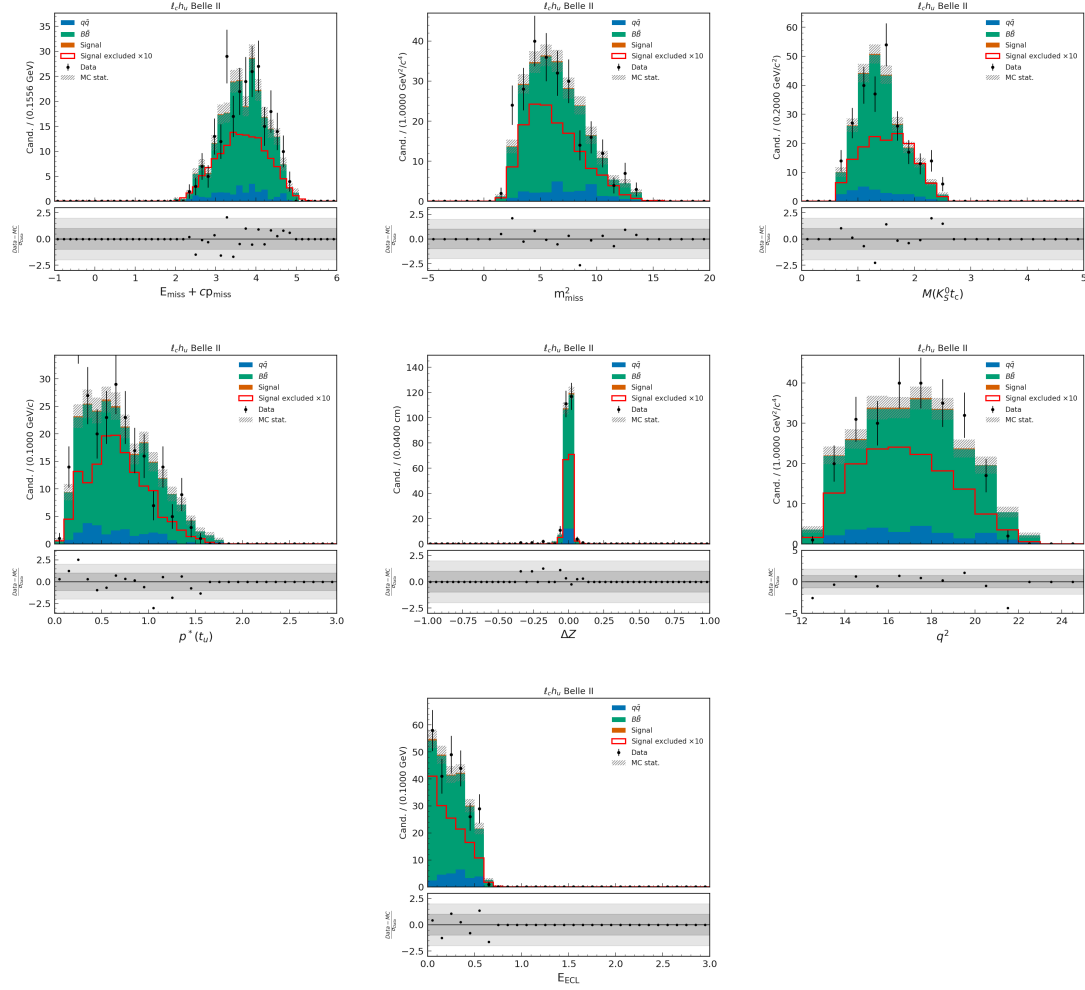


Figure J.7: Data/MC comparison of the training features for the $\ell_c h_u$ category (Belle II signal enhanced region).

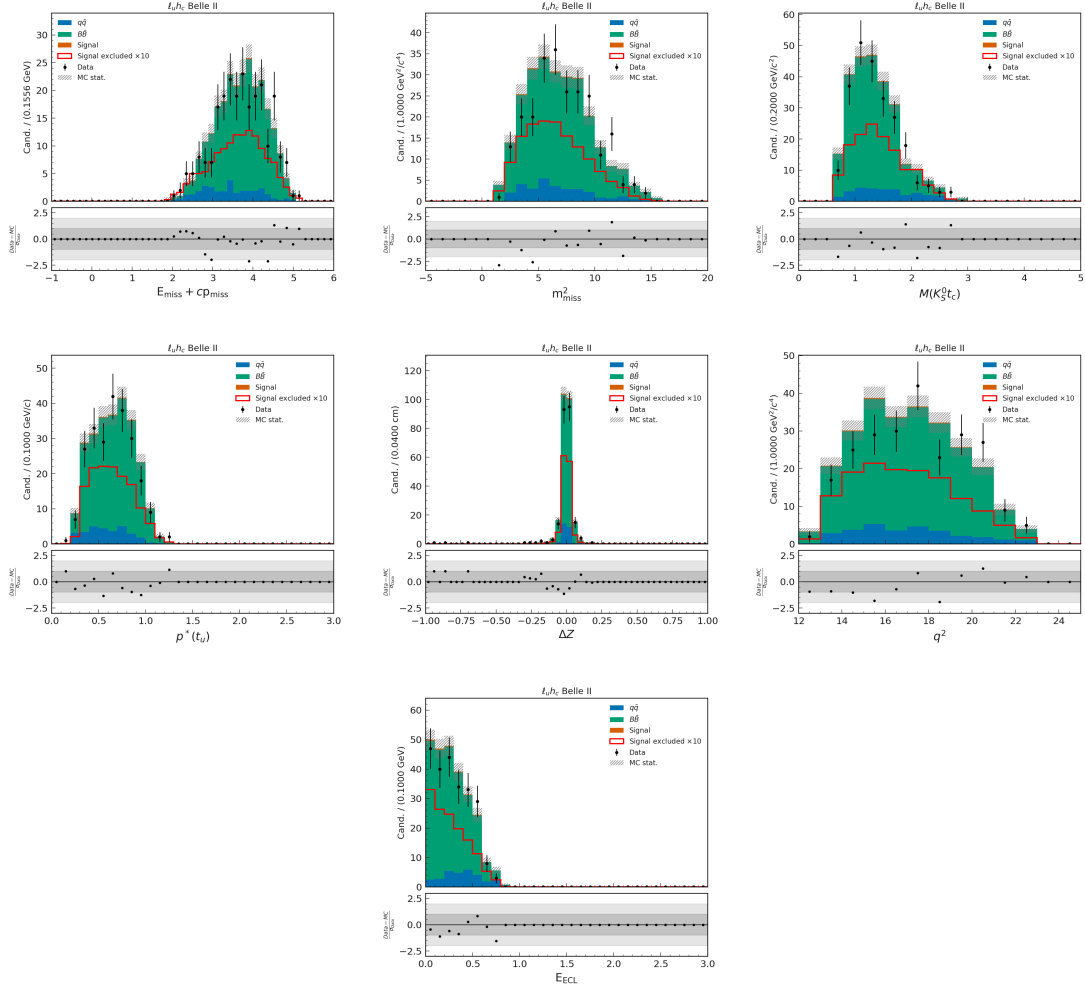


Figure J.8: Data/MC comparison of the training features for the $\ell_u h_c$ category (Belle II signal enhanced region).

Appendix J. “Post-fit” distributions of the BDT input features in signal enhanced region

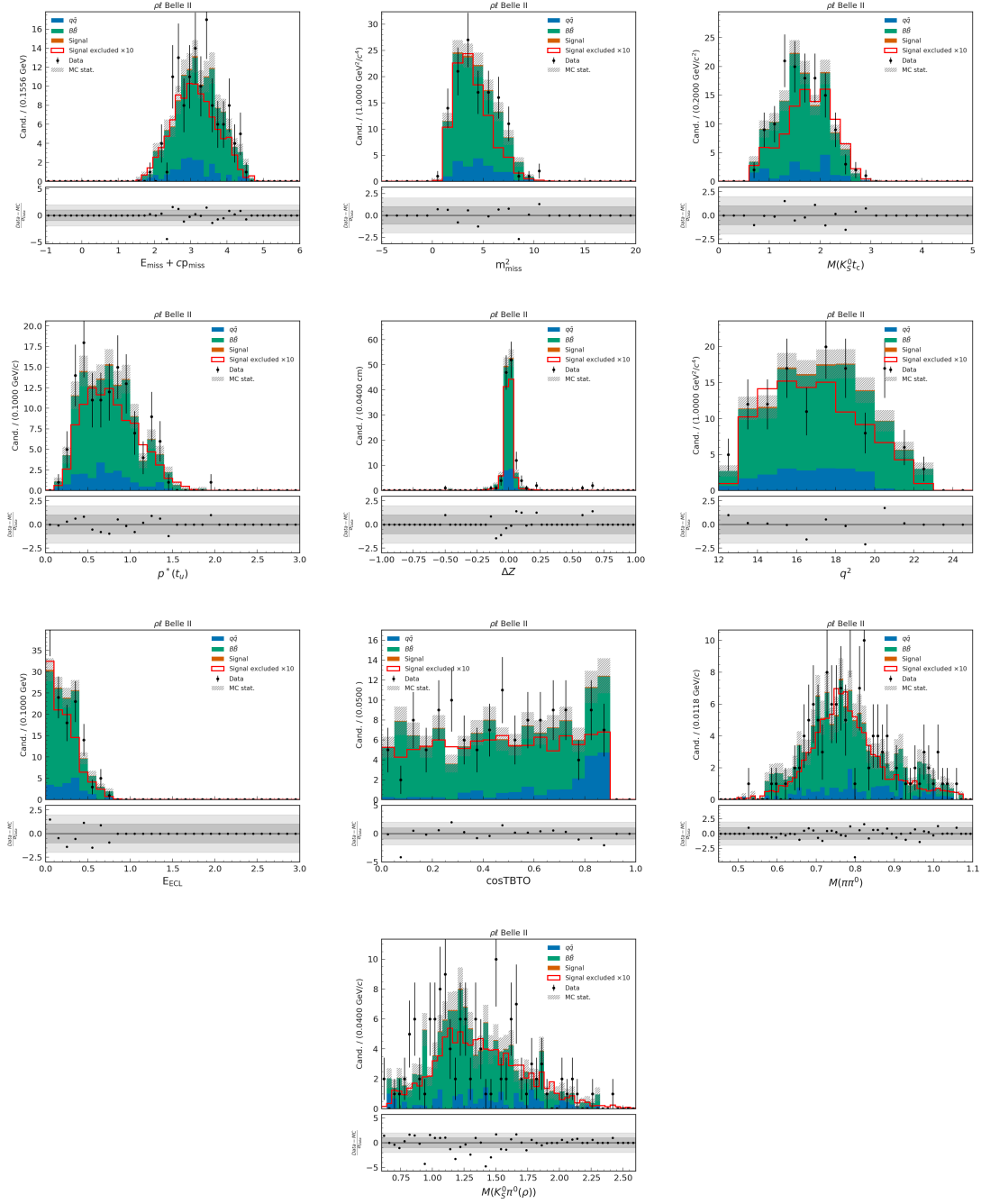


Figure J.9: Data/MC comparison of the training features for the $\rho\ell$ category (Belle II signal enhanced region).

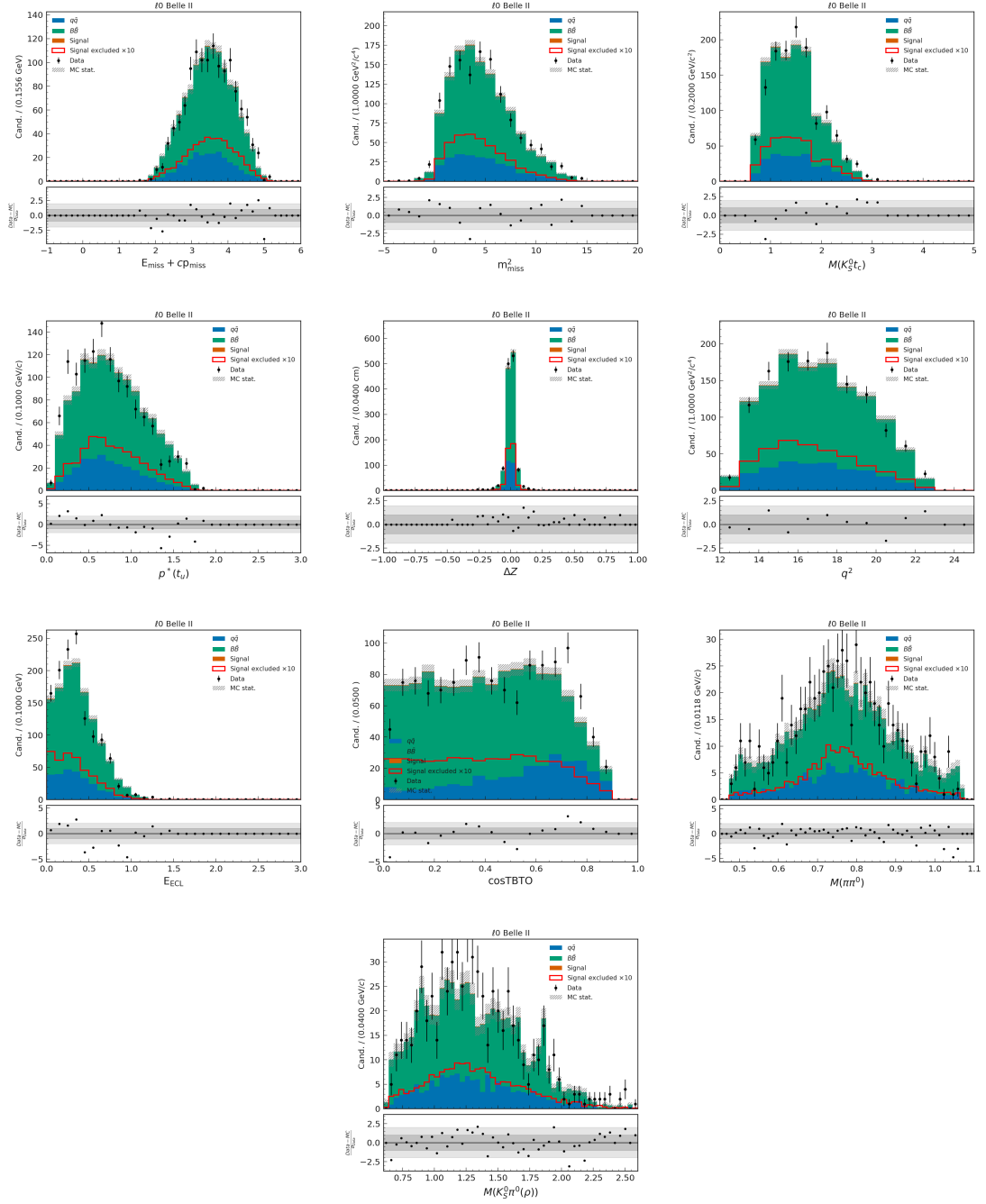


Figure J.10: Data/MC comparison of the training features for the ℓ_0 category (Belle II signal enhanced region).

Razširjeni povzetek v slovenskem jeziku

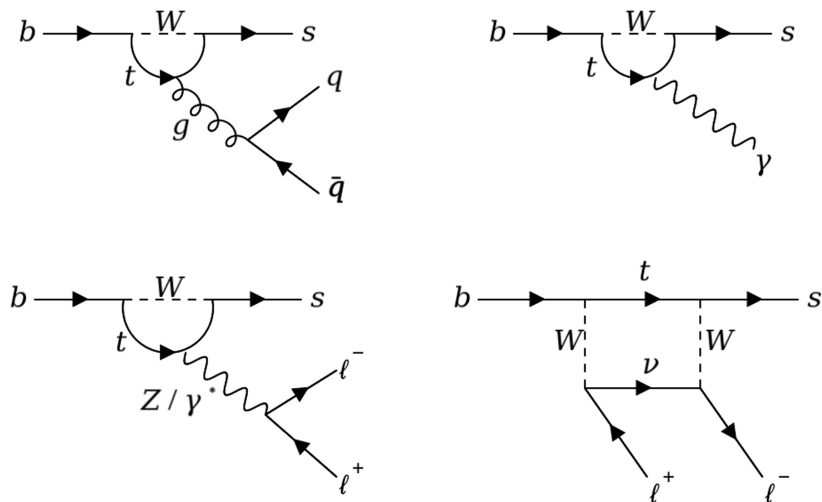
Uvod in fizikalna motivacija

Standardni model fizike osnovnih delcev izjemno uspešno opisuje elektromagnetno, šibko in močno interakcijo. Kljub temu ne predstavlja popolne teorije narave, saj ne vključuje gravitacije, ne pojasni izvora temne snovi, ne razloži hierarhije mas fermionov in ne daje zadostnega pojasnila za opaženo asimetrijo med snovjo in antisnovjo v vesolju. Zato se v fiziki osnovnih delcev izvajajo tako neposredna iskanja novih delcev pri visokih energijah, tipično pri čedalje višjih energijah, kot tudi posredna iskanja, pri katerih se natančne meritve redkih procesov primerjajo z napovedmi Standardnega modela. Drugi pristop je posebej pomemben v fiziki okusov, kjer bi lahko virtualni prispevki težkih delcev nove fizike (NF) znatno povečali verjetnost za redke razpade mezonov B .

Tema disertacije je prvo iskanje razpada $B^0 \rightarrow K_S^0 \tau^+ \tau^-$ z uporabo podatkov eksperimentov Belle in Belle II. Razpad pripada skupini procesov s preходом $b \rightarrow s \tau^+ \tau^-$. Takšni prehodi potekajo preko efektivnih nevtralnih tokov, ki spreminjajo okus; značilni prispevki diagramov v Standardnem modelu so prikazani na Sliki 1. Drevesni diagrami so prepovedani, zato potekajo le preko diagramov z zankami, kot so elektrošibki pingvinski in škatlasti diagrami. Zaradi tega potlačenja so njihova razvejitevna razmerja zelo majhna, tipično reda 10^{-7} za ekskluzivne kanale z dvema leptonoma τ . Hkrati so ravno zaradi tega občutljivi na prispevke težkih delcev NF, ki bi lahko vstopali v iste efektivne operatorje.

Motivacija za razpade z leptoni τ je povezana tudi z eksperimentalno opaženimi neskladji v semileptonskih razpadih mezonov B . V zadnjem desetletju so meritve razmerij, ki testirajo univerzalnost leptonskega okusa, pokazale več odstopanj od napovedi Standardnega modela. Nekatera neskladja v kanalih z elektroni in mioni so se z novejšimi meritvami zmanjšala, vendar globalna slika prehodov $b \rightarrow s \ell^+ \ell^-$ še vedno kaže odstopanja v kotnih opazljivkah in diferencialnih razvejitvenih razmerjih. Poleg tega so v prehodih $b \rightarrow c \tau \nu$ razmerja $R(D)$ in $R(D^*)$ še vedno nad napovedmi Standardnega modela. Veliko modelov NF, npr. modeli z lepto-kvarki ali dodatnimi bozoni Z' , predvideva okrepljene sklopitve s tretjo generacijo leptonov. V takih scenarijih lahko razvejitevna razmerja procesov $b \rightarrow s \tau^+ \tau^-$ narastejo za dva ali tri velikostne rede glede na napoved Standardnega modela.

V disertaciji je ta motivacija postavljena v širši okvir fizike okusov. Fermioni Standardnega modela so razporejeni v tri generacije, pri čemer imajo delci z enakimi kvantnimi števili, vendar iz različnih generacij, enake umeritvene sklopitve, razlikujejo pa se po njihovih masah. V kvarkovskem sektorju šibka interakcija ne deluje na lastna stanja mase, temveč na njihove linearne kombinacije. To neujemanje



Slika 1: Primeri diagramov okus spreminjajočij nevtralnih tokov za prehode $b \rightarrow s$ v Standardnem modelu.

opisuje matrika Cabibbo–Kobayashi–Maskawa (CKM),

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (1)$$

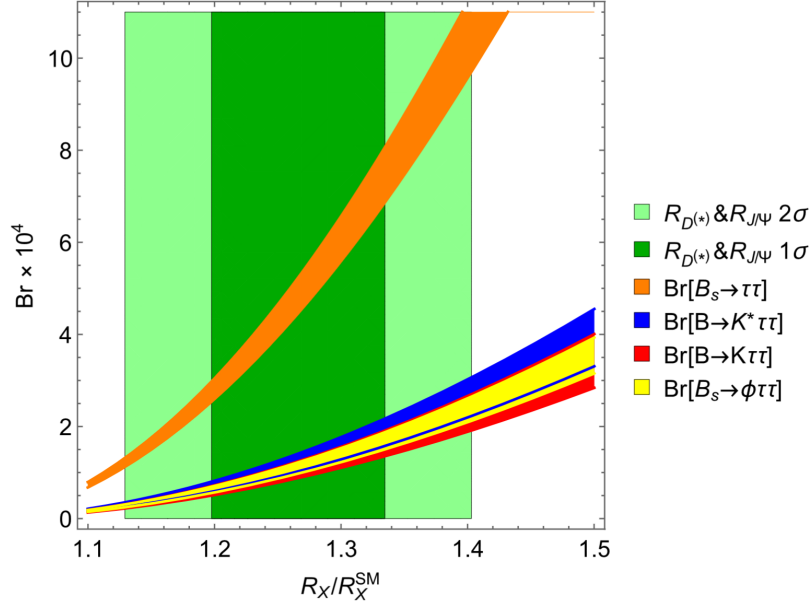
Nevtralni tokovi s spremembo okusa se v Standardnem modelu ne pojavijo na drevesni ravni, saj so na tej ravni prepovedani zaradi unitarnosti matrike CKM. Zato so procesi $b \rightarrow s\ell^+\ell^-$ dostopni le preko višjih redov perturbacijske teorije. Ta lastnost jim daje posebno vlogo: majhna amplituda Standardnega modela pomeni, da lahko tudi majhen prispevek Nove Fizike opazno spremeni merljivo razvejitevno razmerje ali obliko kinematičnih porazdelitev.

V primerjavi s kanali z elektroni in mioni so kanali z leptoni τ eksperimentalno veliko zahtevnejši. Lepton τ razpade še preden doseže aktivne dele detektorja, njegovi razpadi pa vsebujejo vsaj en nevtrino. Pri razpadu $B^0 \rightarrow K_S^0\tau^+\tau^-$ sta prisotna dva leptona τ , zato končno stanje vsebuje dva do štiri nevtrine. Zaradi tega ni mogoče uporabiti ozkih vrhov v invariantni masi ali popolne rekonstrukcije signalnega mezona B , kot je to pogosto pri drugih redkih razpadih. Analiza mora zato združiti informacije o celotnem dogodku, rekonstrukciji asociiranega ali označenega (ang. tag) mezona B , manjkajoči gibalni količini in dodatni aktivnosti v kalorimetru.

V efektivnem opisu lahko prehod $b \rightarrow s\ell^+\ell^-$ zapišemo s hamiltonianom

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_i C_i \mathcal{O}_i + \text{h.c.}, \quad (2)$$

kjer so C_i Wilsonovi koeficienti, $\mathcal{O}_i$ efektivni operatorji, oznaka h.c. predstavlja hermitsko konjugirani člen. Prispevki NF se kažejo kot premiki Wilsonovih koeficientov ali kot pojav novih operatorjev. Ker so procesi z leptoni τ eksperimentalno slabo omejeni, dopuščajo bistveno večji prostor za prispevke NF kot kanali z elektroni ali mioni.



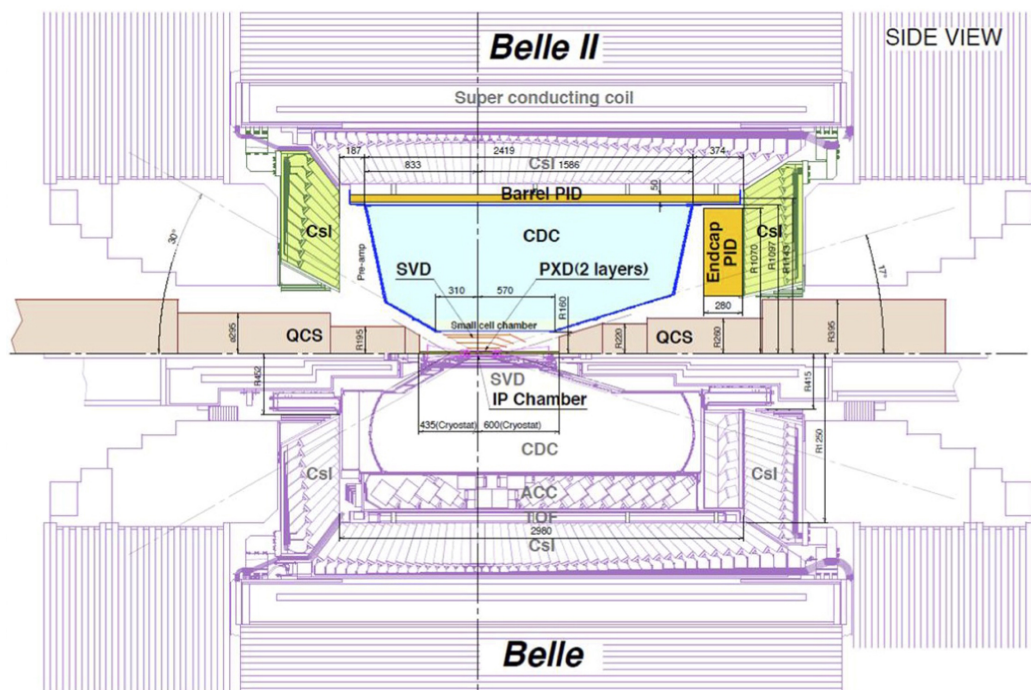
Slika 2: Napovedana razvejitevna razmerja za procese $b \rightarrow s\tau^+\tau^-$ kot funkcija anomalij v opazljivkah $R(X)$.

Eksperimentalno okolje Belle in Belle II

Analiza uporablja podatke, zbrane pri resonanci $\Upsilon(4S)$, ki skoraj vedno razpade v par mezonov $B\bar{B}$. Eksperiment Belle je deloval na trkalniku KEKB, eksperiment Belle II pa deluje na nadgrajenem trkalniku SuperKEKB. Obe napravi sta asimetrični tovarni B-mezonov, kjer trki elektronov in pozitronov potekajo pri energiji v težiščnem sistemu blizu mase resonance $\Upsilon(4S)$. Prednost takšnega okolja je dobro poznano začetno stanje. Če je eden od mezonov B v dogodku popolnoma rekonstruiran, lahko iz znane kinematike začetnega stanja sklepamo na lastnosti drugega mezona B , tudi kadar njegov razpad vsebuje nevtrine.

Uporabljena integrirana luminoznost znaša 711 fb^{-1} za Belle in 365 fb^{-1} za Belle II. To ustreza polnemu vzorcu Belle in podatkom prvega obdobja delovanja Belle II. Eksperimenta imata podobno geometrijo: okoli interakcijske točke so razporejeni sistemi za rekonstrukcijo sledi, identifikacijo delcev, merjenje energije fotonov in elektronov ter identifikacijo mionov in mezonov K_L^0 . Belle II vključuje pomembne nadgradnje glede na svojega predhodnika Belle, med drugim pikselni detektor PXD, izboljšan silicijev verteksni detektor SVD, centralno prehodno komoro CDC, detektorja TOP in ARICH za identifikacijo nabitih delcev ter elektromagnetni kalorimeter ECL.

KEKB in SuperKEKB sta zasnovana kot asimetrična trkalnika elektronov in pozitronov. Asimetrija energij žarkov povzroči, da je sistem $\Upsilon(4S)$ v laboratorijskem sistemu pospešen, kar je ključno za meritve časovno odvisne kršitve simetrije CP . SuperKEKB uporablja nadgrajeno optiko žarkov in tako imenovano shemo nano-žarkov, pri kateri se velikost trkalnega območja močno zmanjša. Namen nadgradnje je doseči bistveno večjo luminoznost kot pri KEKB, pri čemer večja intenziteta in manjše dimenzije žarkov hkrati povečajo ozadja, povezana z žarkom. Zaradi tega je pri Belle II še posebej pomembno razumevanje nevtralnih energijskih skupkov, za-



Slika 3: Shema detektorjev Belle in Belle II s prikazom glavnih poddetektorjev okoli interakcijske točke.

sedenosti poddetektorjev in vpliva ozadij na rekonstrukcijo dogodkov z manjkajočo energijo.

Detektorja Belle in Belle II sta shematsko prikazana na Sliki 3 in sta v disertaciji opisana po funkcionalnih sklopih. Sledilni sistem meri trajektorije nabitih delcev in omogoča rekonstrukcijo verteksov razpadov. Identifikacija nabitih delcev združuje informacije iz prehodne komore, detektorjev sevanja Čerenkova, kalorimetra in sistema za mione. Elektromagnetni kalorimeter meri energijo fotonov in elektronov ter je pomemben za rekonstrukcijo π^0 in za oceno dodatne energije v dogodku. Sistem KLM služi identifikaciji mionov in mezonov K_L^0 . Pri tej analizi so vsi ti sistemi pomembni: signal vsebuje $K_S^0 \rightarrow \pi^+\pi^-$, elektrone ali mione iz razpadov τ , potencialne nevtralne pione iz razpadov ρ , in manjkajočo energijo zaradi nevtrinov.

Simulirani vzorci Monte Carlo so uporabljeni za modeliranje signala in ozadja. Dogodki $e^+e^- \rightarrow q\bar{q}$ z $q = u, d, s, c$ opisujejo kontinuumsko ozadje, generični vzorci $B\bar{B}$ pa opisujejo razpade mezonov B . Signalni vzorci opisujejo razpad $B^0 \rightarrow K_S^0\tau^+\tau^-$ z ustreznimi oblikovnimi faktorji. Simulacija vključuje hadronizacijo, sevanje v končnem stanju, prekrivanje ozadja žarka in odziv detektorja. Podatki Belle so pretvorjeni v format programske opreme Belle II s pomočjo orodja `b2bi`, kar omogoča čim bolj enoten potek analize za oba eksperimenta.

Podatkovni in simulirani vzorci. Za Belle se uporablja vzorec pri resonanci $\Upsilon(4S)$ z integrirano luminoznostjo 711 fb^{-1} , za Belle II pa vzorec prvega obdobja delovanja z integrirano luminoznostjo 365 fb^{-1} . Simulacija ozadij je razdeljena na kontinuumске dogodke $e^+e^- \rightarrow q\bar{q}$, generične razpade $B\bar{B}$ in dodatne redke razpade, ki so pomembni za končno stanje z manjkajočo energijo. Ker je pričakovano razvejitveno razmerje signala zelo majhno, so signalni vzorci Monte Carlo mnogo večji od

podatkovnega vzorca. To omogoča dovolj natančno oceno signalne učinkovitosti po celotnem zaporedju izborov.

V primerjavi z meritvijo preprostega ekskluzivnega razpada je tukaj vloga simulacije ključnega pomena. Simulacija določa signalno učinkovitost po celotnem zaporedju izborov, prispevke različnih vrst ozadja, obliko porazdelitve $\mathcal{O}'$ in večino sistematičnih negotovosti. Hkrati se zaradi skrite analize optimizacija ne sme opirati na signalno območje v podatkih, zato se simulacija vedno preverja v kontrolnih območjih. Z izrazom stranski pas (ang. sideband) označujemo vzorec označujemo vzorec, osiromašen s signalom, ki leži ob signalnem območju v izbrani kontrolni spremenljivki. V disertaciji so opisani vzorci zunaj resonance, K_s^0 "sideband", M_{bc} "sideband" in območja z vetom na razpade preko mezonov D . Ti vzorci pokrivajo različne mehanizme ozadja in omogočajo ločene preglede kontinuumskih dogodkov, kombinatoričnih kandidatov K_s^0 , napačno rekonstruiranih asociiranih mezonov B in ozadij iz razpadov $B \rightarrow DX$.

Za Belle so dogodki najprej obdelani v izvornem okviru Belle, nato pa se s pretvorbo **b2bi** preslikajo v obliko, združljivo z analizo v **basf2**. To omogoča uporabo enotnih definicij objektov, podobnih popravkov učinkovitosti in enotne statistične obravnave. Vendar pa pretvorba ne pomeni, da sta Belle in Belle II eksperimentalno enaka. Razlike v detektorjih, rekonstrukciji, ozadijih in simulacijah ostanejo, zato se veliko popravkov in sistematičnih negotovosti obravnava ločeno za oba eksperimenta.

Vzorci zunaj resonance so posebej pomembni za preverjanje kontinuumskega ozadja, saj ne vsebujejo prispevka resonančnih razpadov $\Upsilon(4S) \rightarrow B\bar{B}$. S tem se loči modeliranje dogodkov $q\bar{q}$ od modeliranja razpadov mezonov B . Vsi vzorci so po rekonstrukciji obdelani v skupnem analiznem okviru, da so razlike med Belle in Belle II čim manjše.

Rekonstrukcija dogodkov in začetni izbor

Razpada $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ ni mogoče popolnoma rekonstruirati, ker leptona τ razpada v končna stanja z nevtrini. Zato je analiza zgrajena na hadronskem označevanju mezona B . V vsakem dogodku se najprej rekonstruira asociirani mezon B_{tag}^0 z algoritmom Full Event Interpretation (FEI), katerega shema je prikazana na Sliki 4. FEI hierarhično rekonstruira ekskluzivne hadronske razpade mezonov B s pomočjo zaporedja multivariatnih klasifikatorjev. S tem dobimo četverec gibalne količine, okus in kakovost rekonstrukcije asociiranega mezona B . Preostali del dogodka se nato uporabi za iskanje signalnega razpada.

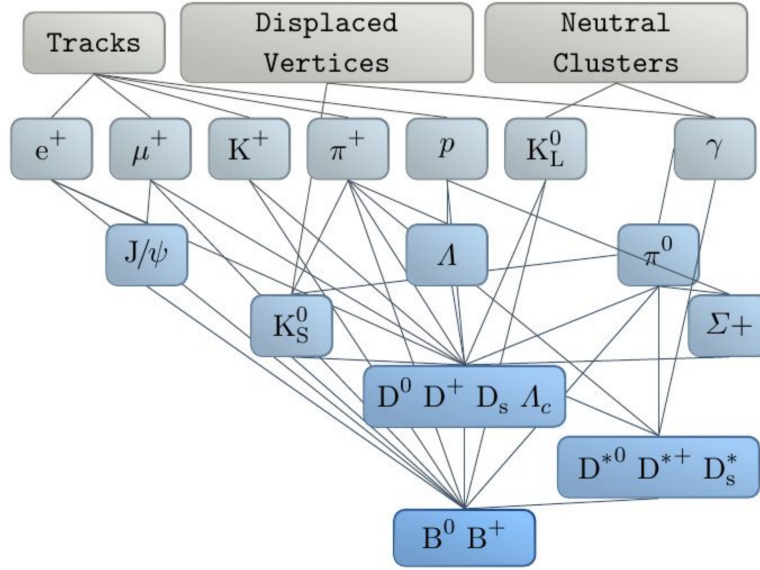
Za asociirani mezon B se zahteva dovolj dober rezultat iz algoritma, to je $\mathcal{P}_{\text{FEI}} > 0.01$. Hadronsko označevanje omogoči uporabo kinematičnih količin, katerih porazdelitvi sta prikazani na Sliki 5 in ki predstavljata močni zahtevi na pravilnost rekonstrukcije asociiranega mezona B ,

$$M_{bc} = \sqrt{E_{\text{beam}}^{*2}/c^4 - |\mathbf{p}_{B_{\text{tag}}}^*|^2/c^2} \quad (3)$$

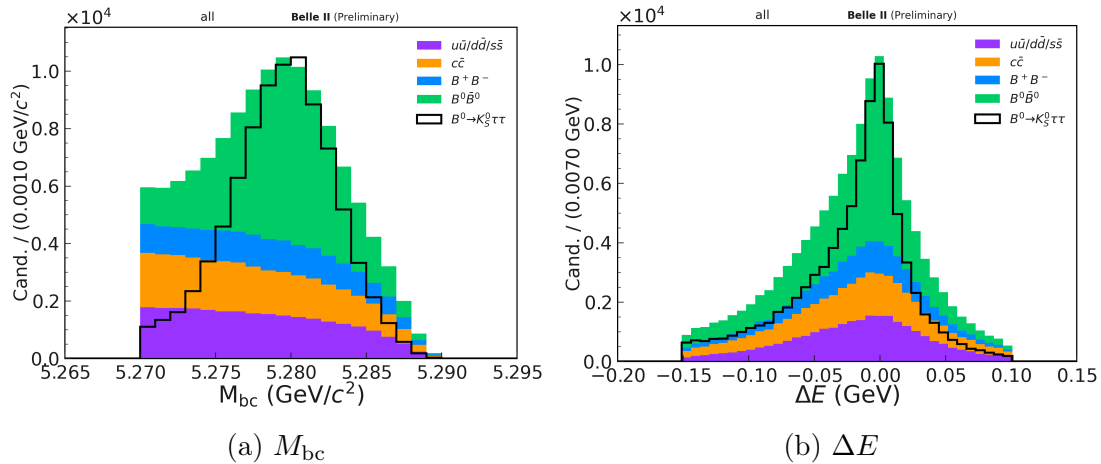
in

$$\Delta E = E_{B_{\text{tag}}}^* - E_{\text{beam}}^*, \quad (4)$$

kjer zvezdica označuje količine v težiščnem sistemu. Zahteve na M_{bc} in ΔE zmanjšajo kombinatorično ozadje in količino napačno rekonstruiranih asociiranih kandidatov. Če je v dogodku več asociiranih kandidatov, se izbere kandidat z največjo vrednostjo $\mathcal{P}_{\text{FEI}}$.



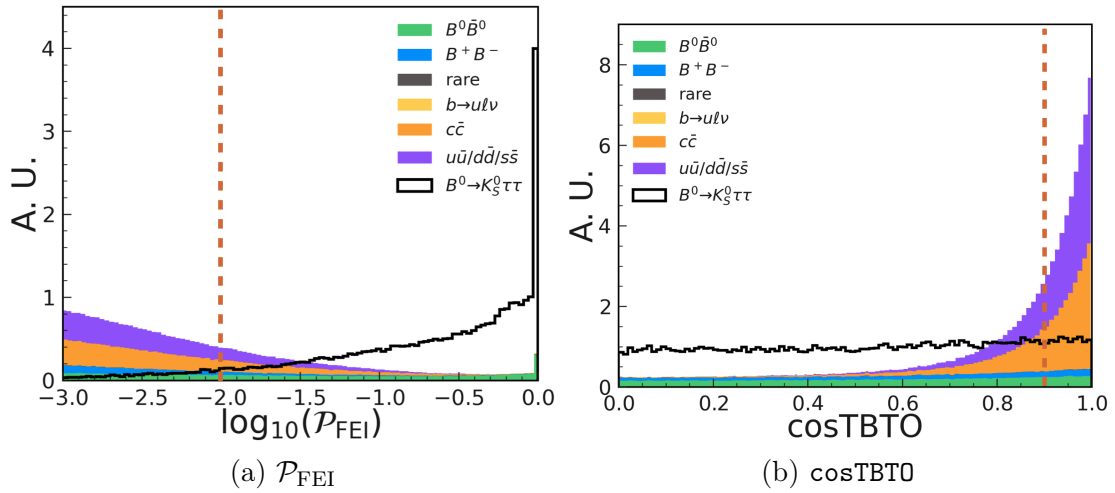
Slika 4: Shematski prikaz algoritma Full Event Interpretation.


 Slika 5: Porazdelitvi M_{bc} in ΔE za signalne in ozadne kandidate po zahtevi na $\mathcal{P}_{\text{FEI}}$.

Osnovna logika rekonstrukcije je naslednja. Najprej se iz vseh delcev v dogodku sestavi najbolj verjeten asociirani mezon B . Vsi delci, uporabljeni v tem kandidatu, se odstranijo iz nadaljnje obravnave signalne strani. Med preostalimi delci se poišče kandidat $K_S^0 \rightarrow \pi^+\pi^-$ in dve nasprotno nabiti sledi, ki sta interpretirani kot vidna razpadna produkta leptonov τ . Dodatni nabiti delci niso dovoljeni, ker bi kazali na napačno rekonstruiran dogodek ali na razpad τ z več nabitimi sledmi. Nevtralni energijski skupki pa se ne odstranijo popolnoma, saj so potrebni za rekonstrukcijo π^0 in za spremenljivke preostanka dogodka.

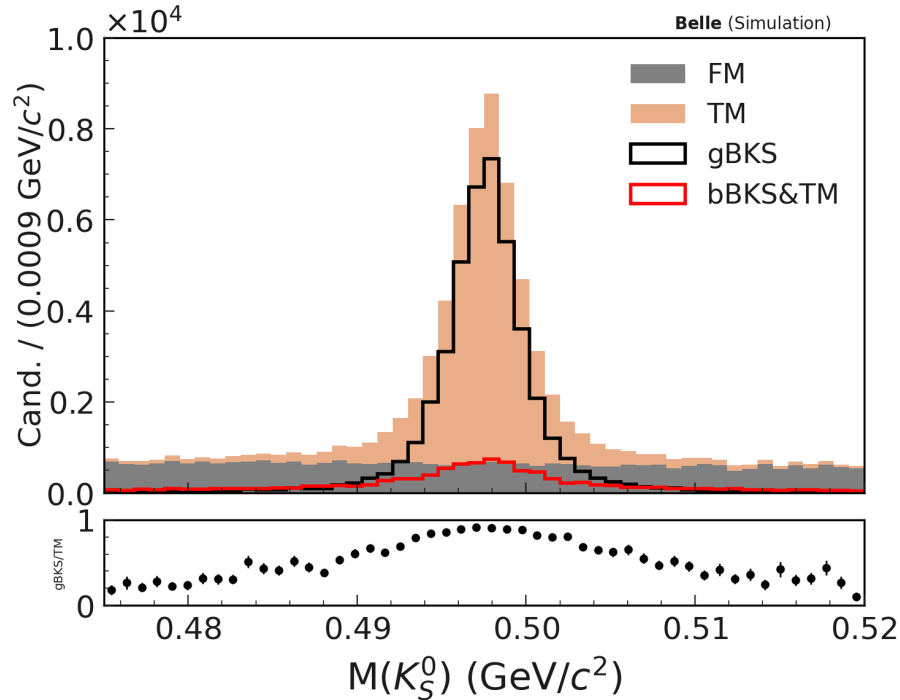
Poleg kinematičnih zahtev se uporabi tudi spremenljivka $\cos\text{TBT0}$; oba začetna izbora sta ilustrirana na Sliki 6. Spremenljivka $\cos\text{TBT0}$ opisuje kot med osjo potiska asociiranega mezona B in osjo potiska preostanka dogodka. Kontinuumski dogodki $e^+e^- \rightarrow q\bar{q}$ imajo bolj pljuskasto topologijo, medtem ko so dogodki $B\bar{B}$ pri resonanci $\Upsilon(4S)$ bolj sferični. Ta razlika je koristna že v začetnem izboru in se pozneje uporabi tudi v multivariatnem razločevanju ozadja.

Signalni mezon B je rekonstruiran iz kandidata K_S^0 in dveh nasprotno nabitih



Slika 6: Izbor na verjetnosti FEI in na kotni spremenljivki $\cos\text{TBTO}$.

sledi, ki predstavljata vidna razpadna produkta leptonov τ . Kandidat K_S^0 je rekonstruiran v razpadu $K_S^0 \rightarrow \pi^+\pi^-$. Zahteva se masa v širokem oknu okoli nominalne mase K_S^0 in uspešna prilagoditev verteksa. Za dodatno zmanjšanje kombinatoričnega ozadja se uporabi izbor `goodBelleKshort`, ki temelji na geometriji razpada K_S^0 , razdaljah sledi od interakcijske točke, smeri leta in položaju verteksa. Njegov učinek na porazdelitev mase je prikazan na Sliki 7.

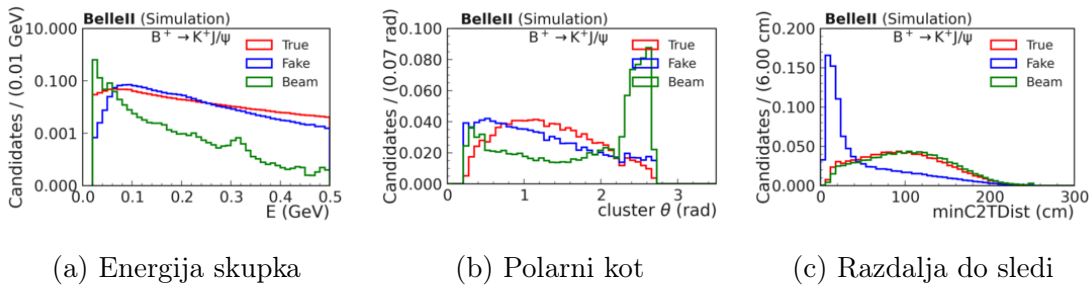


Slika 7: Invariantna masa kandidatov K_S^0 pred in po izboru `goodBelleKshort`.

Za signalno stran se rekonstruira razpad τ z eno nabito sledjo. Uporabljeni so leptonski načini $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$ in $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ ter hadronska načina $\tau^- \rightarrow \pi^- \nu_\tau$ in $\tau^- \rightarrow \rho^- \nu_\tau$. Nabite sledi morajo izvirati iz bližine interakcijske točke in ne smejo biti uporabljene pri rekonstrukciji asociiranega mezona B ali kandidata

K_S^0 . Identifikacija elektronov in mionov uporablja informacije iz več poddetektorjev. Sledi, ki niso identificirane kot leptoni, se obravnavajo kot hadroni; če skupaj s kandidatom π^0 tvorijo maso skladno z razpadom $\rho^\pm \rightarrow \pi^\pm \pi^0$, se razvrstijo kot kandidati ρ .

Razpadi $\tau \rightarrow \rho \nu$ zahtevajo rekonstrukcijo nevtralnega piona, kar je del istega opisa signalne strani. Kandidati π^0 so sestavljeni iz parov fotonov v elektromagnetnem kalorimetru. Pri tem se uporabijo zahteve na energijo fotonov, geometrijsko območje kalorimetra, obliko energijskega skupka in invariantno maso para fotonov. Ker lahko naključni fotoni povzročijo veliko število napačnih kombinacij, se število kandidatov π^0 na dogodek omeji. Dodatni nevtralni skupki, ki niso povezani z asociiranim mezonom B ali signalnim kandidatom, niso takoj zavrženi; namesto tega se uporabijo pri izračunu spremenljivk preostanka dogodka. Te spremenljivke vsebujejo pomembno informacijo o manjkajoči energiji in dodatni aktivnosti v kalorimetru; primeri lastnosti skupkov v ECL so prikazani na Sliki 8.



Slika 8: Lastnosti skupkov v ECL, uporabljene pri razumevanju preostale energije v kalorimetru.

Za signalne dogodke se pričakuje malo dodatne aktivnosti v detektorju, saj so vsi vidni delci pripisani asociirani ali signalni strani, nevtrini pa uidejo detekciji. Ozadja iz semileptonskih in hadronskih razpadov mezonov B pogosto vsebujejo dodatne fotone, nevtralne pione ali napačno pripisane delce. Zato so spremenljivke preostanka dogodka med najpomembnejšimi vhodnimi količinami za multivariatno klasifikacijo signala (Boosted Decision Tree – BDT).

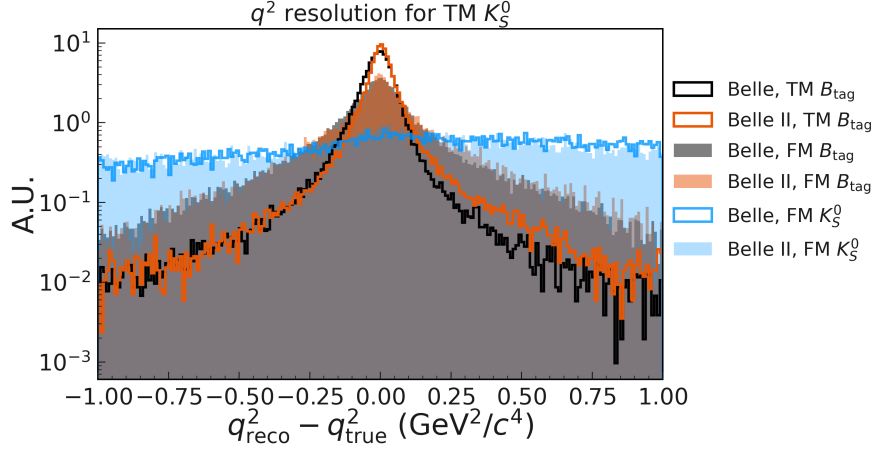
Pomembna kinematična količina je rekonstruirani kvadrat invariantne mase para $\tau^+ \tau^-$,

$$q_{\text{rec}}^2 = m_B^2 + m_{K_S^0}^2 - 2 \left(\frac{E_{\text{beam}}^* E_{K_S^0}^*}{c^4} + \frac{\mathbf{p}_{B_{\text{tag}}}^* \cdot \mathbf{p}_{K_S^0}^*}{c^2} \right). \quad (5)$$

Ker je minimalna dovoljena vrednost za signal približno $4m_\tau^2$, se zahteva $q_{\text{rec}}^2 > 12 \text{ GeV}^2/c^4$. Ta zahteva odstrani očitno ozadje z majhno izgubo signalne učinkovitosti, pri čimer je ločljivost q^2 povzeta na Sliki 9.

Kategorije signalnih kandidatov

Ekskluzivna rekonstrukcija asociiranega mezona B omogoča napoved okusa signalnega mezona B . Zaradi mešanja $B^0 - \bar{B}^0$ ta napoved ni vedno pravilna, vendar je za nemešane dogodke koristna za ločevanje signala od ozadja. Dve nabiti sledi na signalni strani se označita glede na korelacijo njunega naboja z okusom asociiranega mezona B : t_c je korelirana sled, t_u pa nekorelirana sled. Na podlagi identifikacije teh dveh sledi se dogodki razdelijo v kategorije, povzete v Tabeli 1. V končnem



Slika 9: Ločljivost rekonstruirane količine q^2 za različne kombinacije pravilno in napačno rekonstruiranih objektov.

prilagajanju se uporablja pet medsebojno izključujočih se kategorij: $\ell\ell$, ℓ_ch_u , ℓ_uh_c , $\rho\ell$ in $\ell 0$. Kategoriji ℓ_ch_u in ℓ_uh_c sta obravnavani ločeno, ker imajo vodilna ozadja drugačno odvisnost od naboja in okusa. S tem se izboljša izločevanje ozadja. Kategorija $\ell 0$ zajema dogodke brez identificiranega leptona med vidnimi razpadnimi produkti leptonov τ .

Tabela 1: Možna končna stanja razpadov $\tau\bar{\tau}$ z ustrezno vrednostjo ID in razvrstitvijo v kategorije. Pomembnost razvrstitve sledi padajočemu vrstnemu redu spremenljivke ID.

končno stanje $\tau\bar{\tau}$	ID	kategorija
$\mu\mu$	6.2	$\ell\ell$
$e\mu$	6.1	
ee	6.0	
$\mu\rho$	3.6	$\rho\ell$
$e\rho$	3.5	
μh	3.1	ℓ_ch_u, ℓ_uh_c
eh	3.0	
$\rho\rho$	1.0	$\ell 0$
ρh	0.5	
hh	0.0	

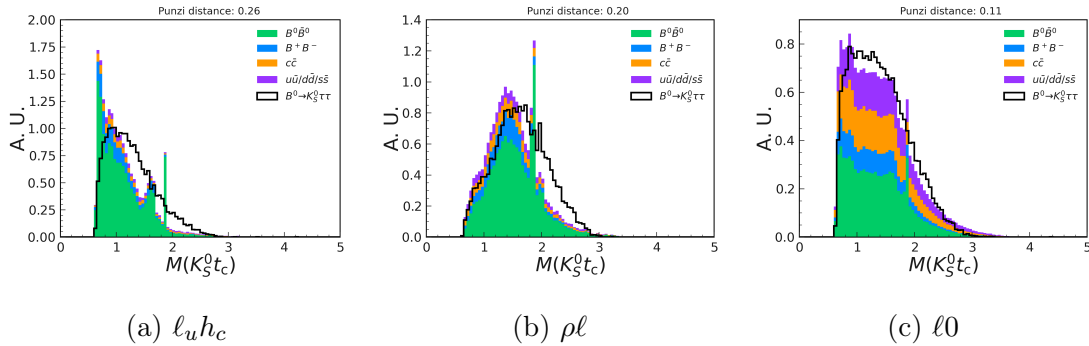
Popravki simulacije in validacija ozadij

Pred končnim prilagajanjem se na simuliranih dogodkih uporabijo korekcijski faktorji, ki popravijo znane razlike med podatki in simulacijo. Največji popravek je povezan z učinkovitostjo hadronskega označevanja asociiranega mezona B , saj je učinkovitost FEI v podatkih drugačna kot v simulaciji. Dodatni popravki upoštevajo učinkovitost rekonstrukcije sledi, rekonstrukcijo K_S^0 , identifikacijo leptonov, rekonstrukcijo π^0 , razvejitevna razmerja razpadov τ in normalizacijo ozadij. Kontinuumsko ozadje se omeji z uporabo podatkov izven območja resonance, kombinatorična komponenta ozadja $B\bar{B}$ pa se preveri v kontrolnih območjih.

Kalibracijski del analize je v disertaciji obravnavan ločeno, ker neposredno določa zanesljivost končnega modela. Učinkovitost rekonstrukcije nabitih sledi, identifikacije leptonov in rekonstrukcije nevtralnih pionov se popravi s faktorji, izmerjenimi v neodvisnih kontrolnih vzorcih. Pri K_S^0 je pomembna tako učinkovitost rekonstrukcije dveh pionov kot tudi pravilno modeliranje geometrijskih zahtev, ki izkoriščajo razmaknjen vertex razpada. Pri leptonih je potrebno upoštevati, da se učinkovitost in delež napačne identifikacije razlikujeta med Belle in Belle II ter med elektroni in mioni. Vsi te popravki vstopijo v predloge za signal in ozadje pred končnim prilaganjem.

Validacija v podatkovnih območjih z majhnim pričakovanim signalnim prispevkom je namenjena preverjanju, ali je mogoče simulaciji zaupati tam, kjer bo uporabljena za napoved ozadja v signalnem območju. K_S^0 "sideband" preverja kombinatorične kandidate K_S^0 , M_{bc} "sideband" preverja napačno rekonstruirane asociirane kandidate, podatki zunaj resonance pa izolirajo kontinuumsko ozadje. V vsakem od teh območij se primerjajo porazdelitve vhodnih spremenljivk BDT in tudi porazdelitve izhoda BDT. S tem se preveri, da klasifikator ne uporablja spremenljivke, ki bi bila v simulaciji slabo opisana.

Analiza uporablja več validacijskih vzorcev. Območje mase K_S^0 zunaj signalnega okna tvori K_S^0 "sideband". Dodatni "sideband" je definiran z uporabo spremenljivke M_{bc} . Ti vzorci preverjajo, ali simulacija pravilno opisuje porazdelitve spremenljivk, uporabljenih za treniranje BDT, ter izhod klasifikatorja. Posebej pomembna je tudi validacija v območjih, kjer se odstrani ozadje z razpadi preko mezonov D , saj so ta ozadja topološko podobna signalu.



Slika 10: Porazdelitev $M(K_S^0, t_c)$ v kategorijah, kjer se uporabi veto na ozadja z mezonu D .

Razkritje signalnega območja. Veto na območje mase mezona D je pomemben predvsem v kategorijah, kjer lahko korelirana hadronska sled skupaj s kandidatom K_S^0 tvori kinematično strukturo, značilno za razpade $B \rightarrow DX$. Slika 10 prikazuje porazdelitve, povezane z vetom na območje mase mezona D . Odstranitev tega območja zmanjša vodilna ozadja, hkrati pa zagotovi neodvisno kontrolno območje za preverjanje modeliranja teh komponent.

V disertaciji je poudarjeno, da validacijska območja niso samo tehnični dodatek, temveč osrednji del strategije meritve. Signalno območje je po definiciji majhno in statistično omejeno, zato morebitnega neskladja v njem ni mogoče enostavno razločiti od statistične fluktuacije. Nasprotno imajo "sideband" območja in območja z vetom na razpade preko mezonov D več dogodkov in so obogatena z znanimi komponentami ozadja. Če je v teh območjih vidno neskladje med podatki in simulacijo,

ga je mogoče obravnavati pred razkritjem signalnega območja. Takšna organizacija analize je posebej pomembna za razpade z manjkajočo energijo, kjer ne obstaja ena sama ozka kontrolna spremenljivka, temveč je zanesljivost rezultata odvisna od usklajenega opisa več medsebojno povezanih količin.

Pri validaciji se ne preverja samo skupna normalizacija, temveč tudi oblike porazdelitev. To je pomembno, ker končno prilagajanje uporablja obliko porazdelitve $\mathcal{O}'$, ne le števila dogodkov po izboru. Napačno modeliranje vhodne spremenljivke BDT se lahko po transformaciji preslika v napačno obliko prilagojene predloge. Zato so v disertaciji prikazane primerjave vhodnih spremenljivk v več kontrolnih območjih, nato pa še primerjave izhoda BDT. Šele ko so te primerjave zadovoljive, se isti model uporabi za napoved ozadja v signalnem območju.

Pri vseh validacijah se uporablja enak pristop: najprej se definira območje, ki ne vsebuje signalnega prispevka ali pa ga vsebuje zanemarljivo malo, nato se primerjajo podatki in simulacija po vseh relevantnih popravkih. Če se opazi sistematično neskladje, se uvede popravni faktor ali dodatna negotovost. S tem se zmanjša možnost, da bi končni rezultat izhajal iz napačnega modeliranja ozadja in ne iz pravega signalnega presežka.

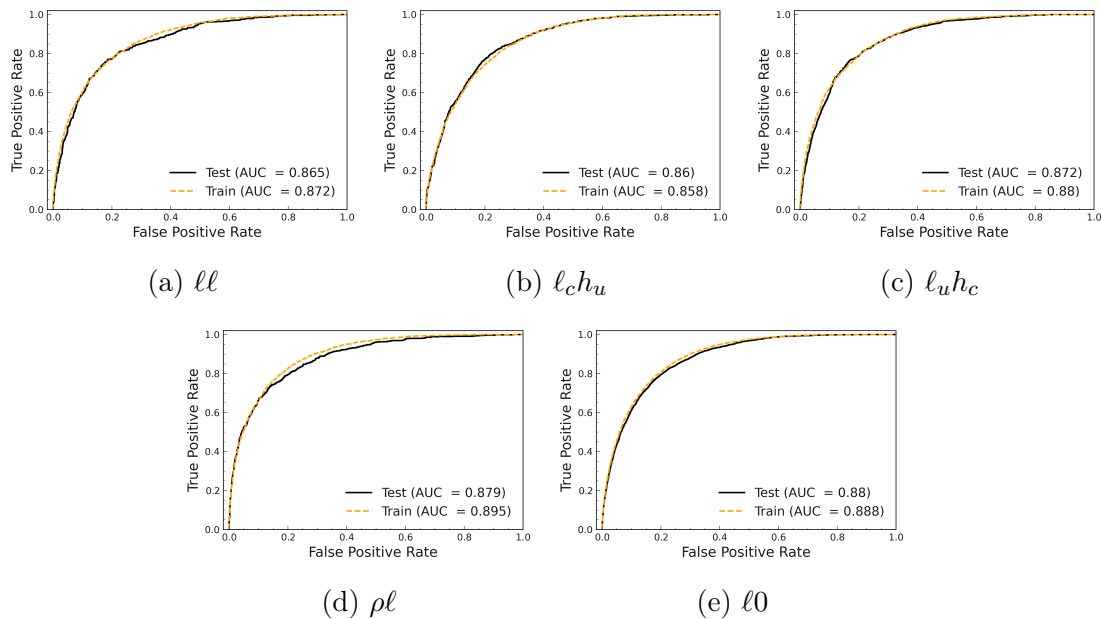
Odstranjevanje ozadja z BDT

BDT uporabimo kot fizikalno motivirano orodje za odstranjevanje ozadja, pri čemer so vhodne spremenljivke izbrane glede na njihovo povezavo s topologijo signalnih in ozadnih dogodkov. Nekatere spremenljivke opisujejo signalno stran, na primer invariantne mase kombinacij $K_s^0 t_c$, $K_s^0 \pi^0$ in $\pi^\pm \pi^0$. Druge opisujejo globalno kinematiko, kot so q_{rec}^2 , manjkajoča masa in kot med osmi potiska. Tretja skupina opisuje nesignalni preostanek dogodka, predvsem energijo dodatnih skupkov v ECL ter kombinacije manjkajoče energije in manjkajoče gibalne količine. Tak izbor vhodnih spremenljivk je pomemben, saj BDT ločuje signal od ozadja na podlagi robustnih fizikalnih razlik in ne na podlagi artefaktov simulacije.

Za vsako kategorijo in za vsak eksperiment je ozadje drugačno. V kategoriji $\ell\ell$ prevladujejo semileptonski razpadi z dvema leptonskima sledema. V kategorijah $\ell_c h_u$ in $\ell_u h_c$ je pomembno, ali je hadronska sled korelirana ali nekorelirana z okusom asociiranega mezona B , saj to loči različne prispevke iz razpadov preko mezonov D . Kategorija $\rho\ell$ vsebuje dodatno informacijo iz kandidata π^0 , kategorija $\ell 0$ pa ima večje ozadje, ker ne zahteva identificiranega leptona. Ločeno treniranje BDTjev omogoča, da izkoristimo informacijo teh razlik, namesto da bi bile povprečene v enem samem klasifikatorju.

Po začetnem izboru je razmerje med signalom in ozadjem še vedno zelo majhno. Učinkovitost ločevanja je za primer Belle II povzeta z ROC krivuljami na Sliki 11. Zato se za vsako kategorijo in za vsak eksperiment ločeno natrenira BDT. Izhodne spremenljivke BDT uporabljajo informacije o signalni strani, asociirani strani in preostanku dogodka. Med najpomembnejšimi so q_{rec}^2 , manjkajoča masa, energija dodatnih nevtralnih skupkov v ECL, invariantne mase kombinacij K_s^0 z vidnimi razpadnimi produkti τ , kot med osmi potiska obeh strani dogodka in količine, povezane z manjkajočo gibalno količino.

Uspešnost BDT se preveri z ločenimi učnimi in testnimi vzorci ter z razdelitvijo dogodkov na sode in lihe podvzorci. Takšna delitev preprečuje, da bi isti dogodek hkrati sodeloval pri učenju klasifikatorja in pri oceni njegove končne porazdelitve. Za vsako kategorijo se preveri stabilnost odziva med obema podvzorcema. Če bi bila



Slika 11: Primer ROC krivulj za testne in učne vzorce v Belle II.

razlika med sodim in lihim treningom velika, bi to kazalo na preučenje ali na premajhen simulirani vzorec. V končni analizi so odzivi stabilni, zato se transformirani izhod BDTja uporablja kot zanesljiva spremenljivka prilagajanja.

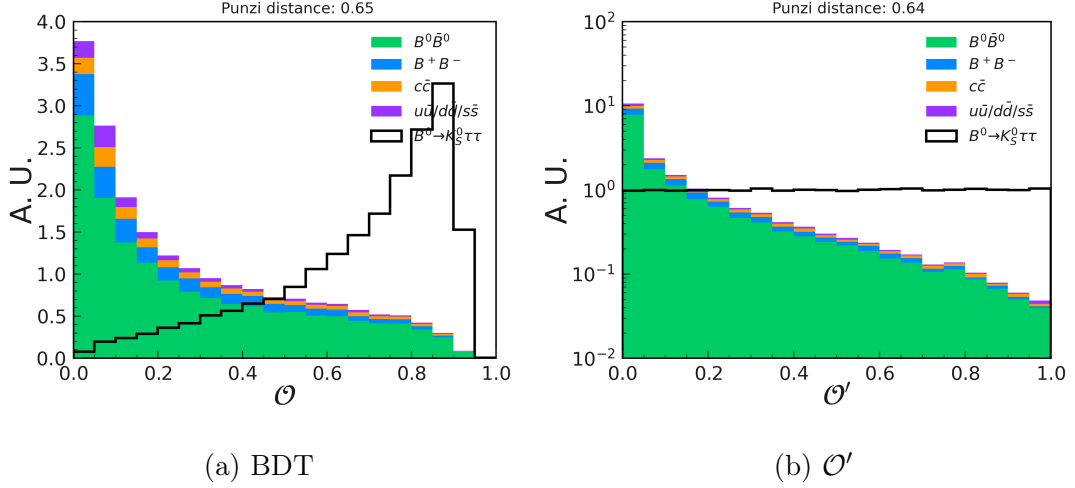
Primer izhoda pred in po transformaciji je prikazan na Sliki 12. Izhod BDTja se transformira s transformacijo verjetnostnega integrala, tako da je porazdelitev signala približno enakomerna. Transformirana spremenljivka se označi z $\mathcal{O}'$. Končno prilagajanje se izvede v območju $\mathcal{O}' > 0.3$, ki ohrani velik delež signala in hkrati močno zmanjša ozadje. Ker je $\mathcal{O}'$ končna spremenljivka prilagajanja, je njeno modeliranje preverjeno v več validacijskih območjih pred razkritjem signalnega območja. Analiza je zasnovana kot skrita analiza: signalno najbolj občutljivo območje se pregleda šele po določitvi izborov, popravkov, sistematičnih negotovosti in statističnega postopka.

Postopek razkritja signalnega območja zato nastopi šele po zaključku odstranjevanja ozadja in validaciji. Izbori, popravki simulacije, trening BDTjev, obravnava sistematičnih negotovosti in statistični model so določeni z uporabo simulacije in kontrolnih območij. Šele po zaključku teh korakov se pregleda območje z visoko vrednostjo $\mathcal{O}'$. Takšna skrita analiza zmanjšuje tveganje, da bi se izbori nenamerno prilagodili statističnim fluktuacijam v podatkih.

Statistični model in sistematične negotovosti

Statistični model mora hkrati izkoristiti informacijo iz desetih kanalov in pravilno prenesti sistematične negotovosti med njimi. Zato je prilagajanje zgrajeno kot skupen model, ne kot zaporedje neodvisnih prilagajanj. Parameter zanimanja je razvejeno razmerje oziroma ekvivalentna jakost signala. Vsi moteči parametri so profilirani, kar pomeni, da se pri vsaki testirani vrednosti signala izberejo tiste vrednosti motečih parametrov, ki najbolje opišejo podatke ob upoštevanju njihovih omejitev. Tak pristop omogoča, da podatki sami omejijo nekatere komponente ozadja, vendar le v meri, ki je skladna z zunanji negotovostmi.

Histogramske predloge so pripravljene za signal in več komponent ozadja. Nego-



Slika 12: Primer izhoda BDT pred in po transformaciji verjetnostnega integrala za kategorijo $\ell_u h_c$ v vzorcu Belle II.

tovosti oblike spremenijo relativno vsebino stolpcev v porazdelitvi $\mathcal{O}'$, medtem ko normalizacijske negotovosti spremenijo skupno pričakovano število dogodkov. Statistične negotovosti simuliranih vzorcev so pomembne, saj je po končnem izboru število simuliranih dogodkov v nekaterih stolpcih majhno. Njihova obravnava prepreči, da bi statistične fluktuacije simulacije bile interpretirane kot natančno poznana oblika ozadja.

Signal se izloči s hkratnim prilagajanjem histogramskih predlog v desetih kanalih: pet kategorij za Belle in pet kategorij za Belle II. V vsakem kanalu je spremenljivka prilagajanja $\mathcal{O}'$. Model je implementiran v okviru HistFactory s paketoma `pyhf` in `cabinetry`. Funkcija verjetja je produkt Poissonovih verjetnosti za vse stolpce in kategorije, pomnožen z Gaussovskimi omejitvenimi členi za moteče parametre:

$$f(\mu | \chi) = \prod_{\substack{c \in \\ \text{kategorije}}} \prod_{\substack{b \in \\ \text{bini}_c}} \text{Pois}(n_{c,b} | \nu_{c,b}(\mu, \chi)) \times \prod_{\chi \in \mathcal{X}} \text{Gaus}_{\chi}(a_{\chi} | \chi). \quad (6)$$

Tukaj je $n_{c,b}$ število opaženih dogodkov v stolpcu b kategorije c , $\nu_{c,b}$ pa ustrezno pričakovano število dogodkov. Sistematične negotovosti so vključene preko motečih parametrov χ , ki spreminjajo pričakovane vsebine stolpcev in so omejeni z Gaussovskimi členi Gaus_{χ} , povezanimi s pomožnimi podatki a_{χ} .

Pričakovano število dogodkov v posameznem stolpcu lahko zapišemo kot

$$\nu_{c,b}(\mu, \chi) = \mu \cdot 2 \cdot N_{\Upsilon(4S)} \cdot f_{00} \cdot \mathcal{B}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) \cdot \varepsilon_{c,b}(\chi) + \nu_{c,b}^{B\bar{B}}(\chi) + \nu_{c,b}^{q\bar{q}}(\chi). \quad (7)$$

Tukaj sta $\nu_{c,b}^{B\bar{B}}$ in $\nu_{c,b}^{q\bar{q}}$ pričakovani vsebini predlog za ozadji $B\bar{B}$ in $q\bar{q}$. Čeprav je nominalna signalna učinkovitost enakomerna v spremenljivki prilagajanja, postane zaradi negotovosti oblike odvisna od stolpca, zato je označena kot $\varepsilon_{c,b}$. Faktor μ , ki je prost parameter prilagajanja, množi celotno normalizacijo signalnih predlog glede na privzeto referenčno razvejitveno razmerje.

Sistematične negotovosti so razdeljene glede na njihov izvor. Največji prispevki izvirajo iz razvejitvenih razmerij vodilnih razpadov mezonov B , statistike simulacije, učinkovitosti hadronskega označevanja asociiranega mezona B , normalizacije kombinatorične komponente asociiranega mezona B in modeliranja razpadov $D \rightarrow K_L^0$.

Negotovosti oblike se vključijo kot spremembe histogramskih predlog, normalizacijske negotovosti pa kot pomnoževalni faktorji. Kadar je negotovost skupna Belle in Belle II ali več kategorijam, je v modelu opisana s koreliranim motečim parametrom. Glavni viri so v istem vrstnem redu kot v glavnem besedilu povzeti v Tabeli 2.

Tabela 2: Viri sistematičnih negotovosti, njihova obravnava v prilagajanju in njihov prispevek k negotovosti jakosti signala μ .

Vir	Vrsta negotovosti, parametri	σ_μ
Vodilna razvejitevna razmerja razpadov mezonov B	korelirana negotovost oblike, 140	0.20
Statistika MC	nekorelirana negotovost oblike, 140	0.18
Učinkovitost B -označevanja	korelirana negotovost oblike, 2	0.06
Normalizacija kombinatoričnih B_{tag}^0	korelirana negotovost oblike, 10	0.05
Modeliranje $D \rightarrow K_L^0$	korelirana negotovost oblike, 2	0.05
Identifikacija leptonov	korelirana negotovost oblike, 6	0.03
f_{00}	korelirana negotovost oblike, 1	0.02
Negotovost signalne učinkovitosti	normalizacija, 2	0.02
Učinkovitost π^0	korelirana negotovost oblike, 2	0.02
Normalizacija $q\bar{q}$	normalizacija, 4	0.02
Signalni model	korelirana negotovost oblike, 1	0.01
Učinkovitost K_S^0	korelirana negotovost oblike, 2	< 0.01
Število $\Upsilon(4S)$	normalizacija, 2	< 0.01
Negotovost luminoznosti	normalizacija, 2	< 0.01
Učinkovitost rekonstrukcije sledi	korelirana negotovost oblike, 2	< 0.01
Vodilna razvejitevna razmerja razpadov τ	korelirana negotovost oblike, 10	< 0.01
χ_d	korelirana negotovost oblike, 1	< 0.01
Skupaj		0.29

Rezultati

Rezultate na podatkih interpretiramo previdno. Ker je pričakovani signal v Standardnem modelu daleč pod trenutno eksperimentalno občutljivostjo, namen analize ni meritev Standardnega modela, temveč iskanje morebitnega velikega povečanja, ki bi lahko izviralo iz nove fizike. Zato je pomembno, da se poleg centralne vrednosti navede tudi pričakovana in opažena zgornja meja. Pričakovana meja opisuje občutljivost analize v primeru, ko je prisotno samo ozadje; opažena meja pa vključuje dejansko statistično fluktuacijo podatkov. Če sta si ti dve vrednosti blizu, to pomeni, da podatki ne kažejo nepričakovanega presežka ali primanjkljaja glede na model ozadja.

Ločeni rezultati Belle in Belle II so koristni za navzkrižno preverjanje. Belle ima večjo integrirano luminoznost, Belle II pa drugačen detektor in drugačne učinkovitosti rekonstrukcije. Če bi bil presežek prisoten samo v enem eksperimentu ali samo v eni kategoriji, bi bilo treba podrobno preveriti možnost napačnega modeliranja. V končnem rezultatu so centralne vrednosti ločenih prilagajanj skladne z ničlo znotraj negotovosti, združeni rezultat pa je prav tako skladen s hipotezo brez signala.

Končno prilagajanje je izvedeno hkrati za Belle in Belle II.

Na ravni posameznih eksperimentov sta dobljena rezultata

$$\mathcal{B}_{\text{Belle}}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = [0.07 \pm 0.42 (\text{stat.}) \pm 0.36 (\text{sist.})] \times 10^{-3}, \quad (8)$$

$$\mathcal{B}_{\text{Belle II}}(B^0 \rightarrow K_S^0 \tau^+ \tau^-) = [0.20 \pm 0.50 (\text{stat.}) \pm 0.44 (\text{sist.})] \times 10^{-3}. \quad (9)$$

Oba rezultata sta skladna s pričakovanjem v odsotnosti signala. Hkratno prilagajanje ni preprosto povprečje obeh ločenih rezultatov, ker so v združenem modelu nekatere sistematične negotovosti korelirane med Belle in Belle II, prilagajanje pa hkrati uporablja vseh deset kategorij. Zato lahko centralna vrednost skupnega rezultata nekoliko odstopa od intuitivnega aritmetičnega povprečja ločenih prilagajanj.

Pred prilagajanjem je ujemanje med podatki in modelom dobro. Po prilagajanju ni opaznega presežka v končnih porazdelitvah $\mathcal{O}'$, prikazanih na Sliki 13, ki bi kazal na signal $B^0 \rightarrow K_s^0 \tau^+ \tau^-$. Prilagojena vrednost razvejitvenega razmerja je

$$\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-) = [0.12 \pm 0.33 (\text{stat.}) \pm 0.29 (\text{sist.})] \times 10^{-3}. \quad (10)$$

Ker po prilagajanju ni opažen statistično značilen presežek, se rezultat izrazi tudi kot zgornja meja. Za določitev zgornje meje se uporablja metoda CL_s ,

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1 - p_b}. \quad (11)$$

Rezultat je združljiv s hipotezo ozadja. Sken CL_s je prikazan na Sliki 14, odmiki motečih parametrov pa na Sliki 15, prispevki sistematičnih negotovosti pa so povzeti v Tabeli 2. Ustrezna zgornja meja je

$$\mathcal{B}(B^0 \rightarrow K_s^0 \tau^+ \tau^-) < 0.83 \times 10^{-3} \quad (12)$$

pri 90-odstotni stopnji zaupanja. To je prva eksperimentalna omejitev za ta razpadni kanal.

Izospinsko povprečenje

Rezultat nevtralnega kanala združimo tudi z meritvijo razpada $B^+ \rightarrow K^+ \tau^+ \tau^-$, s čimer dobimo skupno mejo za psevdoskalarni prehod $B \rightarrow K \tau^+ \tau^-$. Pri tem je treba upoštevati, da opazovani končni stanji vsebujeta K^+ oziroma K_s^0 , medtem ko se izospinska povezava nanaša na razpade v psevdoskalarni kaon. V kombinaciji, povzeti z istimi enačbami kot v glavnem besedilu, se zato upoštevata projekcija $K^0 \rightarrow K_s^0$ in razmerje življenjskih dob nabitega in nevtralnega mezona B . Skupno prilagajanje uporablja en sam prost parameter, sistematične negotovosti pa so korelirane ali nekorelirane glede na njihov fizikalni izvor.

Kombinirani rezultat je

$$\mathcal{B}(B \rightarrow K \tau^+ \tau^-) = (0.06 \pm 0.28) \times 10^{-3}, \quad (13)$$

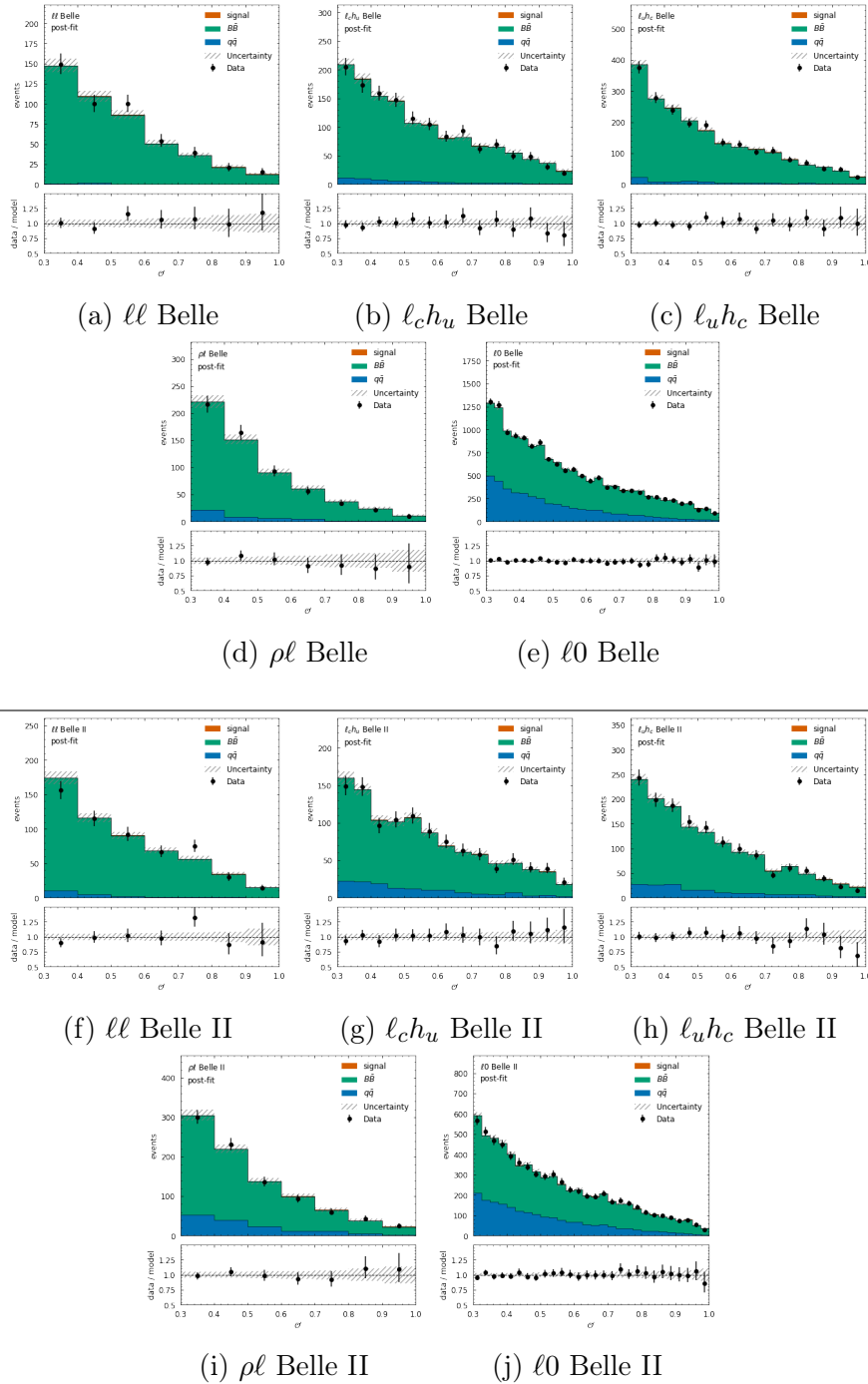
zgornja meja pa

$$\mathcal{B}(B \rightarrow K \tau^+ \tau^-) < 0.54 \times 10^{-3} \quad (14)$$

pri 90-odstotni stopnji zaupanja. Ta rezultat predstavlja najstrožjo trenutno omejitev za prehode $b \rightarrow s \tau^+ \tau^-$ s psevdoskalarnim kaonom v končnem stanju.

Sklep

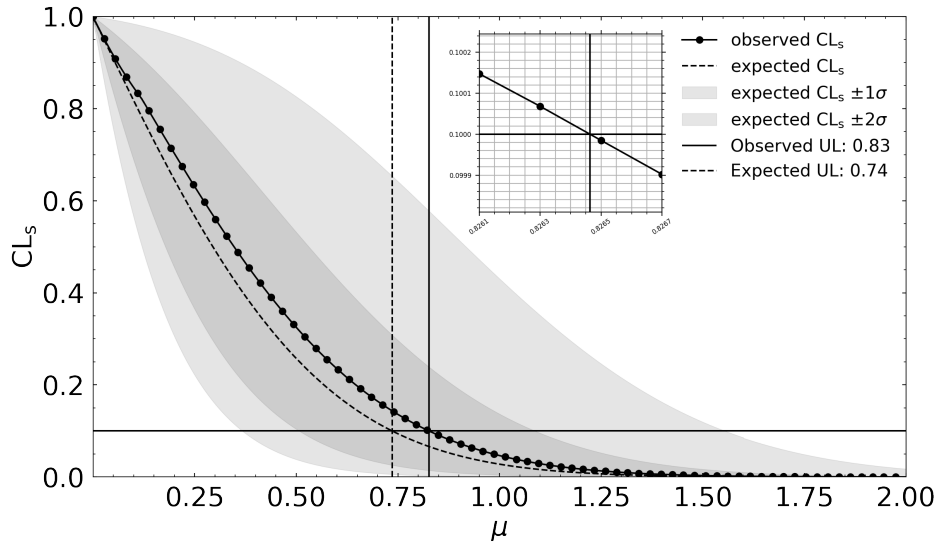
V okviru tega dela je bilo izvedeno prvo iskanje razpada $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ z uporabo podatkov Belle in Belle II. Analiza združi oba podatkovna vzorca v enotnem statističnem modelu, rekonstruira asociirani mezon B v hadronskih razpadih, razdeli signalne kandidate v več kategorij glede na vidne razpade leptonov τ , odstranjuje



Slika 13: Porazdelitve transformiranega izhoda BDT $\mathcal{O}'$ po prilagajanju za Belle in Belle II.

ozadje z ločeno natreniranimi klasifikatorji BDT in izvede hkratno prilagajanje porazdelitvam $\mathcal{O}'$. Sistematične negotovosti oblike in normalizacije so vključene kot moteči parametri, pri čemer so korelacije med eksperimenti in kategorijami obravnavane glede na njihov fizikalni izvor.

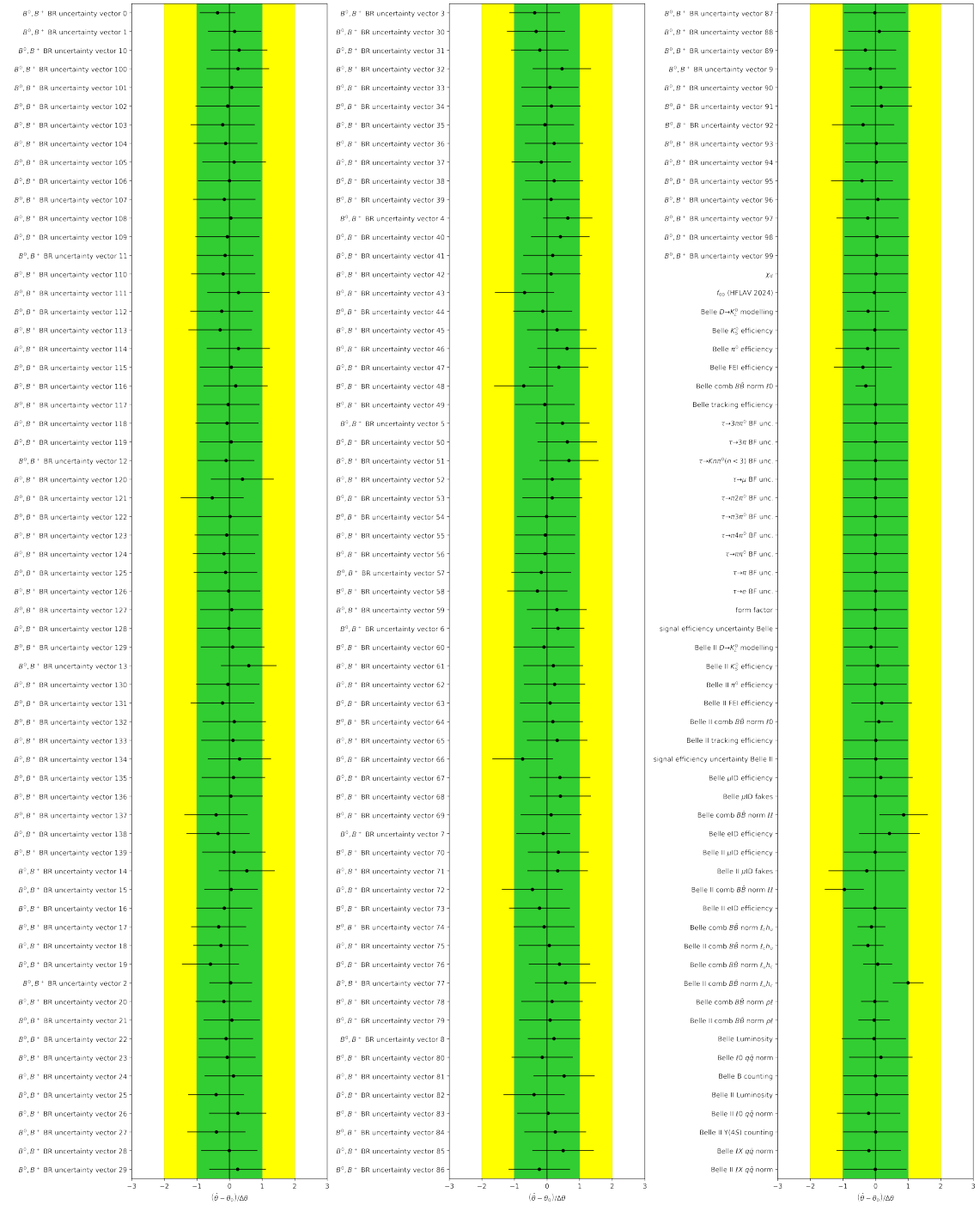
V končnem prilagajanju ni opazen statistično pomemben presežek nad pričakovanim ozadjem. Izmerjena vrednost razvejitenega razmerja je skladna z ničlo, zato je rezultat podan tudi kot zgornja meja. Delo prvič izmeri oziroma omeji razvejitevno razmerje za kanal $B^0 \rightarrow K_s^0 \tau^+ \tau^-$ in določi najstrožjo trenutno zgornjo mejo



Slika 14: Skeniranje CL_s za hkratno prilagajanje podatkom Belle in Belle II.

za prehode $b \rightarrow s\tau^+\tau^-$ s psevdoskalarnim kaonom v končnem stanju. Čeprav je občutljivost še vedno nad napovedjo Standardnega modela, rezultat omejuje določene scenarije NF z močno povečanimi sklopitvami na leptone τ .

Prihodnji podatki Belle II bodo neposredno izboljšali statistično komponento občutljivosti, vendar bo za dosego bistveno bolj natančne meritve pomembno tudi zmanjšanje sistematičnih napak. Med njimi so boljše modeliranje hadronskih razpadov mezonov B , natančnejša kalibracija FEI, večji simulirani vzorci, izboljšana obravnava ozadij žarka in morebitna uporaba dodatnih razpadov τ z eno nabito sledjo. Ker se pri večjih podatkovnih vzorcih statistične negotovosti zmanjšajo, bodo prav negotovosti oblike in zanesljivost modeliranja ozadja postale še bolj kritične.



Slika 15: Odmiki motečih parametrov v hkratnem prilagajanju podatkom Belle in Belle II.

List of publications related to this doctoral thesis

1. M. Liu, $b \rightarrow s\ell^+\ell^-$ ($\ell = e, \mu, \tau$) and $b \rightarrow s\nu\bar{\nu}$ at Belle and Belle II, arXiv:2604.22079 [hep-ex] (2026) [65].
2. V. Bertacchi, *Measurements of electroweak penguins and B decays to final states with missing energy at Belle and Belle II*, arXiv:2604.17644 [hep-ex] (2026) [66].