Searches for the $B \to K^{(*)}\nu\bar{\nu}$ decays using an inclusive tagging method at the Belle II experiment

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Yulan Fan
aus Shandong (China)

Hamburg
2026

Gutachter/innen der Dissertation:	Dr. Alexander Glazov Prof. Dr. Kerstin Tackmann
Zusammensetzung der Prüfungskommission:	Dr. Alexander Glazov Prof. Dr. Kerstin Tackmann Prof. Dr. Günter H. W. Sigl Prof. Dr. Elisabetta Gallo Dr. Katharina Behr
Vorsitzender der Prüfungskommission:	Prof. Dr. Günter H. W. Sigl
Datum der Disputation: Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Markus Drescher
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

The rare decays $B \to K\nu\bar{\nu}$ represent a class of flavor-changing neutral current processes that are highly suppressed in the Standard Model and thus serve as sensitive probes for new physics. These decays are theoretically clean, with uncertainties primarily stemming from hadronic form factors, and their branching fractions are predicted with high precision. However, their experimental study poses significant challenges due to the presence of two undetectable neutrinos in the final state, necessitating sophisticated reconstruction techniques and robust background suppression.

This thesis investigates the decays $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ using data corresponding to $365\ \mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance within the Belle II experiment. The analyses employ the inclusive tagging method, in which the companion B meson is reconstructed inclusively to infer the kinematics of the signal-side decay. This approach provides substantial gains in signal efficiency compared to exclusive tagging, albeit at the cost of increased background complexity.

The branching fraction of $B^+ \to K^+\nu\bar{\nu}$ is extracted from a maximum likelihood fit, yielding $[2.7 \pm 0.5(\mathrm{stat}) \pm 0.5(\mathrm{syst})] \times 10^{-5}$, which corresponds to a significance of 2.9 standard deviations from the Standard Model expectation. This measurement establishes the experimental viability of the inclusive tagging technique and has already prompted refinements of theoretical predictions, motivating renewed investigations into potential new physics effects.

Building upon this framework, the analysis of $B^0 \to K^{*0}\nu\bar{\nu}$ is currently in the pre-unblinding stage. This channel poses additional challenges due to the more complex form-factor structure of the K^{*0} meson and the higher track multiplicity of its final states. Nevertheless, the analysis follows a similar strategy to the $B^+ \to K^+\nu\bar{\nu}$ case, with dedicated improvements in background modeling and event selection tailored to the K^{*0} topology. While no results are presented at this stage, the $B^0 \to K^{*0}\nu\bar{\nu}$ study is expected to provide complementary sensitivity to new physics through its unique helicity structure and kinematic features.

In addition, this thesis includes a complementary study on a background filter in the central drift chamber, aimed at improving Belle II tracking performance under high-background conditions. While independent of the $B \to K^{(*)}\nu\bar{\nu}$ analyses, this work constitutes an independent contribution to optimization of the detector performance.

Zusammenfassung

Die seltenen Zerfälle $B \to K\nu\bar{\nu}$ gehören zur Klasse der flavor-wechselnden neutralen Ströme, die im Standardmodell stark unterdrückt sind und daher empfindliche Sonden für neue Physik darstellen. Diese Zerfälle sind theoretisch besonders sauber, da die Unsicherheiten hauptsächlich aus hadronischen Formfaktoren stammen, und ihre Verzweigungsverhältnisse mit hoher Präzision vorhergesagt werden können. Ihre experimentelle Untersuchung ist jedoch äußerst anspruchsvoll, da zwei nicht nachweisbare Neutrinos im Endzustand auftreten. Dies erfordert ausgefeilte Rekonstruktionstechniken und eine robuste Unterdrückung von Untergrundprozessen.

Diese Dissertation untersucht die Zerfälle $B^+ \to K^+\nu\bar{\nu}$ und $B^0 \to K^{*0}\nu\bar{\nu}$ anhand von Daten mit einer integrierten Luminosität von $365\ \mathrm{fb}^{-1}$, die bei der $\Upsilon(4S)$-Resonanz im Belle-II-Experiment aufgezeichnet wurden. Die Analysen verwenden die inklusive Tagging-Methode, bei der das Begleit-B-Meson inklusiv rekonstruiert wird, um die Kinematik des Signalzerfalls zu erschließen. Dieser Ansatz ermöglicht deutliche Gewinne in der Signaleffizienz im Vergleich zum exklusiven Tagging, geht jedoch mit einer erhöhten Untergrundkomplexität einher.

Das Verzweigungsverhältnis von $B^+ \to K^+\nu\bar{\nu}$ wird aus einem Maximum-Likelihood-Fit bestimmt und ergibt $[2.7 \pm 0.5(\mathrm{stat}) \pm 0.5(\mathrm{syst})] \times 10^{-5}$, was einer Signifikanz von 2.9 Standardabweichungen gegenüber der Standardmodell-Erwartung entspricht. Dieses Ergebnis belegt die experimentelle Realisierbarkeit der inklusiven Tagging-Technik und hat bereits zu einer Verfeinerung theoretischer Vorhersagen geführt, die neue Untersuchungen möglicher Effekte jenseits des Standardmodells motivieren.

Aufbauend auf diesem Rahmen befindet sich die Analyse des Zerfalls $B^0 \to K^{*0}\nu\bar{\nu}$ derzeit in der Vor-Unblinding-Phase. Dieser Kanal stellt zusätzliche Herausforderungen dar, da die Formfaktoren des K^{*0} komplexer sind und seine Endzustände eine höhere Spurenmultiplizität aufweisen. Dennoch folgt die Analyse einer ähnlichen Strategie wie im Fall von $B^+ \to K^+\nu\bar{\nu}$, ergänzt durch gezielte Verbesserungen in der Untergrundmodellierung und Ereignisselektion, die speziell an die Topologie des K^{*0} angepasst sind. Obwohl zum jetzigen Zeitpunkt keine Ergebnisse präsentiert werden, verspricht die Untersuchung von $B^0 \to K^{*0}\nu\bar{\nu}$ eine komplementäre Sensitivität gegenüber neuer Physik durch ihre einzigartige Helizitätsstruktur und kinematischen Eigenschaften.

Darüber hinaus beinhaltet diese Dissertation eine ergänzende Studie zu einem Untergrundfilter in der zentralen Driftkammer, der die Leistungsfähigkeit der Spurrekonstruktion von Belle II unter Hochuntergrund-Bedingungen verbessern soll. Unabhängig von den $B \to K^{(*)}\nu\bar{\nu}$-Analysen stellt diese Arbeit einen eigenständigen Beitrag zur Optimierung der Detektorleistung dar.

Disclaimer

The simulation and data analysis presented in this thesis are the results of a collaborative effort. The subject of this thesis was proposed by my PhD advisor Dr. Alexander Glazov (AG).
This work is based on data collected with the Belle II detector, which was designed, installed, and maintained by members of the Belle II Collaboration. The Belle II detector records data from electron–positron collisions provided by the SuperKEKB accelerator, which is operated by the SuperKEKB Collaboration. The work presented in this thesis partially relied on the Belle II software framework, which is maintained by members of the Belle II Collaboration. Whenever internal results from the collaboration are used, they are explicitly cited. Part of the analysis was performed on the National Analysis Facility hosted at DESY.
The $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ analyses are implemented by a team of researchers. I was directly involved in all stages of both analyses: setup, validation, background suppression, statistical interpretation of results, development and review of analysis code, and preparation of documentation. For the $B^+ \to K^+\nu\bar{\nu}$ channel, I also handled the responses to collaboration comments up to the final publication stage. I am a co-author of the publication (PHYS. REV. D 109, 112006 (2024)) [1], based on the search for the $B^+ \to K^+\nu\bar{\nu}$ decay, which is presented in Part II, Chapter 1 of this thesis. Apart from the specific points listed below, all contributions presented in this thesis are my own work:

- Figure style: Dr. Cyrille Praz (CP).
- Fake-photon study: Eldar Ganiev (EG).
- K_L^0 efficiency studies: AG.
- $B\bar{B}$ Background correction: mainly AG, EG, and Slavomira Stefkova.
- Systematics and signal extraction: introduced mainly by AG, CP, and Dr. Simon Kurz; subsequently carried out as a team effort.
- Signal-selection efficiency and validation in $B^+ \to K^+\nu\bar{\nu}$: Dr. Filippo Dattola, Yubo Han, and AG.
- $\mathrm{BDT}_{\mathrm{CA}}$ and signal-selection-efficiency method: introduced by AG.

The complementary study was carried out mainly by me, under the supervision of AG, with comments and feedback from friends and experts as listed below:

- **CDC background-suppression Filter**
  - Comments from the Belle II tracking group, Prof. Giulia Casarosa, Dr. Christian Wessel, Longke Li, and Boyang Zhang.
  - XGBoost optimization within basf2: introduced by AG.
  - Reconstruction of raw hits: CP, AG, and me.
- **Modeling Studies**
  - $\gamma\gamma \to K_S^0 K_S^0$ modeling: commented by AG, Xu Dong, Hailong Ma, Sen Jia, Xiang Pan, and Prof. Sadaharu Uehara.
  - $K^+\pi^+\pi^-$ modeling: substantially commented by Mengchuan Du.

Declaration on oath

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.
If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Name: Yulan Fan

Place: Hamburg, Germany

Date: 5 October 2025

To my Oma and Opa

Contents

Part I

Introduction

This chapter establishes the theoretical and experimental foundations for the study of rare $b \to s\nu\bar{\nu}$ transitions, which constitute some of the cleanest probes of flavor-changing neutral currents in the Standard Model. The first part provides the theoretical motivation, beginning with an overview of the Standard Model and its flavor sector, followed by the effective Hamiltonian formalism for $b \to s$ transitions. Special emphasis is placed on the Standard Model predictions for exclusive $B \to K^{(*)}\nu\bar{\nu}$ decays, together with possible contributions from physics beyond the Standard Model that could alter these channels.
The second part introduces the experimental framework, focusing on the Belle II experiment at the SuperKEKB collider. The accelerator design, detector subsystems, and trigger and data acquisition systems are described, along with the simulation and event reconstruction chain that enables precision studies of rare B-meson decays.
Finally, the chapter outlines the data analysis strategies and techniques employed in later sections of this thesis. These include inclusive tagging algorithms, signal-from-background discriminating topological observables used in multivariate classifiers for background suppression, and maximum likelihood estimation methods. Together, these discussions provide the theoretical motivation, experimental context, and methodological basis for the results presented in the next chapters.

Chapter 1

Theoretical Framework

1.1 Standard Model

The central aim of particle physics is to uncover the most elementary constituents of matter and the laws that govern their interactions. Over the past century, progress has been driven by repeatedly asking: What lies inside? Atoms were revealed to contain nuclei and electrons, nuclei were shown to consist of protons and neutrons; and protons and neutrons themselves turned out to be made of quarks, bound together by hadronic interaction. Along the way, leptons (the electron, muon, tau, and their neutrinos), photons, W and Z bosons, gluons, and most recently the Higgs boson were added to the catalogue of fundamental particles of nature. Together, these entities and their interactions are described by the Standard Model of particle physics (SM), the most successful and precise framework we currently possess [2–4].
The Standard Model encompasses three essential components: matter fields, gauge bosons, and the Higgs boson:

- **Matter fields:** These comprise twelve spin-$\frac{1}{2}$ fermions arranged into three families. Each family contains two quarks and two leptons, distinguished by their electric charge and masses. Quarks (u, d, c, s, t, b) carry fractional charges and participate in the strong interaction through their color charge, while leptons (e, μ, τ and their associated neutrinos) are colorless. The fermionic nature of these particles has been experimentally confirmed through angular distribution measurements in scattering processes [5, 6].

- **Gauge bosons:** These spin-1 particles mediate the fundamental forces of the SM, their existence and properties dictated by the underlying gauge symmetries. The photon mediates electromagnetism and is massless. Eight gluons mediate the strong interaction and, although massless, are confined within hadrons. The gluon was independently confirmed in three-jet events observed at PETRA in 1979 [7]. The $W^{\pm}$ and Z bosons mediate the weak interaction and are massive, with their discovery at CERN [8, 9] providing direct confirmation of the electroweak gauge structure.

- **Higgs boson:** The only fundamental scalar in the theory, the Higgs boson arises from the mechanism of electroweak symmetry breaking, which endows fermions and weak gauge bosons with mass. Its discovery at the LHC in 2012 completed the SM particle spectrum [10, 11].

In addition to its particle content, the SM is characterized by a set of free parameters, comprising the coupling couplings, fermion masses, Cabibbo–Kobayashi–Maskawa (CKM) mixing angles, and CP-violating phases. These parameters are determined experimentally. The predictive power of the SM is extraordinary. Precision tests performed at LEP and SLC confirm its validity at the per-mille level [12], which global electroweak fits successfully anticipated the mass of the top quark and constrained the Higgs boson mass prior to their experimental discoveries. The anomalous magnetic moment of the electron, measured to one part in 10^{12}, provides a further striking demonstration of its accuracy.
The guiding principle of the SM is gauge symmetry, the requirement that the laws of physics remain invariant under certain local transformations. In quantum mechanics, multiplying the wavefunction by a constant

phase leaves physical predictions unchanged, corresponding to a global U(1) symmetry. By Noether's theorem, such symmetries imply conserved quantities, in this case, electric charge [13]. The principle gains greater power when the symmetry is required to hold locally, for instance, when the transformation may vary from point to point in spacetime. To maintain invariance in this case, compensating fields - gauge bosons - must be introduced, with dynamics dictated by the symmetry itself. Thus, interactions among particles arise directly from the demand of local gauge invariance [14, 15].

The Standard Model is constructed on the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ sector describes the strong interaction. Quarks are observed to carry color charge, and the existence of their bound states (for instance, mesons and baryons) confirms the presence of an underlying $SU(3)$ symmetry. Experiments at SLAC and DESY first revealed point-like quark constituents in deep inelastic scattering (DIS), providing direct evidence for quarks and gluons [6]. The electroweak interaction is described by the $SU(2)_L \times U(1)_Y$ sector. Left-handed fermions transform as $SU(2)$ doublets, while right-handed fermions are singlets. The discovery of the W and Z bosons at CERN in 1983 [8, 9] confirmed this structure, with their measured masses $M_W \approx 80.379 \pm 0.012$ GeV and $M_Z \approx 91.1876 \pm 0.0021$ GeV [12]. The gauge structure is highly constrained: It is the minimal anomaly-free group consistent with observed particle properties. Once specified, the symmetry fixes the form of the interactions, leaving only a finite set of measurable parameters.

Gauge symmetry forbids explicit mass terms for fermions and gauge bosons, yet all observed particles except the photon and gluons are massive. The Higgs mechanism resolves this by introducing a complex scalar doublet with a nonzero vacuum expectation value (VEV), the electroweak symmetry $SU(2)_L \times U(1)_Y \rightarrow U(1)_em$ is spontaneously broken. In this process, three of the scalar degrees of freedom are absorbed as the longitudinal components of the $W^{\pm}$ and Z bosons, thereby giving them mass while leaving the photon massless. The remaining scalar degree of freedom manifests as the physical Higgs boson. Fermion masses arise through Yukawa interactions, in which left- and right-handed fermion fields couple to the Higgs doublet. Once the Higgs acquires its VEV, these interactions generate fermion mass terms proportional to their Yukawa couplings. This explains why the SM accommodates massive quarks and leptons while retaining gauge invariance. However, the Yukawa couplings span over twelve orders of magnitude-from the electron to the top quark-without explanation, reflecting an unresolved flavor puzzle of the SM.

The Higgs boson itself was discovered at the LHC in 2012 [16,17], confirming this mechanism experimentally. Its mass is now precisely measure, $M \approx$125 GeV/c^2 [10, 11], and its couplings to other particles has been tested to be within 5-15% [18,19], all consistent with SM predictions. The Higgs VEV is determined from the Fermi constant measured in muon decay, giving $v = 246.22 \pm 0.06$ GeV [4], which set the electroweak scale. Yet, when compared to the Planck scale, $M_{\text{Pl}} = \sqrt{\frac{\hbar c}{G_N}} \approx 1.22 \times 10^{19}$ GeV [20], a striking tension appears.

In quantum field theory, loop corrections to the Higgs boson mass are quadratically sensitive to the highest energy scale up to which the theory is valid. If the SM is treated as an effective field theory with a cutoff at M_{Pl}, quantum corrections would naturally push the Higgs mass toward that enormous scale. The fact that the observed Higgs mass remains near the electroweak scale, therefore, requires an extreme fine-tuning between the bare mass and the quantum corrections. Explaining this stability of the weak scale against Planck-scale corrections is known as the hierarchy problem, and it represents one of the most profound open questions in particle physics.

Despite these achievements, the SM is manifestly incomplete. It does not explain dark matter or dark energy. Neutrino oscillations have conclusively demonstrated that neutrinos possess nonzero masses, which requires an extension of the minimal model [21]. Furthermore, the observed baryon–antibaryon asymmetry of the universe cannot be explained by the level of CP violation encoded in the CKM matrix [22, 23]. Finally, the values of the fundamental parameters such as fermion masses, mixing angles, and gauge couplings remain unexplained: they enter the theory only as empirical inputs.

Thus, the SM stands as both a triumph of modern physics and a signpost pointing beyond itself. It provides a framework that explains a vast array of observations with exquisite precision, yet leaves many of the most profound questions in particle physics unresolved.

One of the most intriguing aspects of the SM is its treatment of flavor. The theory accommodates three generations of quarks and leptons, but does not explain the replication of families, the large hierarchies in fermion masses, or the observed pattern of mixing between quark flavors [24, 25]. These open questions point to flavor physics as a promising sector in which to search for phenomena beyond the SM. In particular,

rare processes in which quark flavors change without altering electric charge, known as flavor-changing neutral currents (FCNCs), are especially valuable. Within the SM such transitions are forbidden at tree level by the Glashow–Iliopoulos–Maiani (GIM) mechanism [26] and occur only through higher-order loops, making them both highly suppressed and theoretically well controlled [27, 27]. This dual property — rarity and precision — renders FCNCs an excellent probe of potential new dynamics, and places $b \to s$ transitions at the forefront of flavor physics studies [28, 29].

1.2 $b \to s$ Transition

The transition $b \to s$ is a classic example of FCNCs. The underlying amplitudes involve all up-type quarks circulating in loops with W bosons, with the top quark giving the numerically dominant contribution [30]. To describe such processes at scales relevant for B decays, one integrates out the heavy fields W, Z, t, H at $\mu \simeq m_W$ using the operator product expansion. This procedure yields a local effective Hamiltonian [27, 31]

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} \left[\sum_{q=u,c,t} V_{qb}V_{qs}^* \big(C_1\,\mathcal{O}_1^q + C_2\,\mathcal{O}_2^q\big) + V_{tb}V_{ts}^* \sum_{i=3}^{10} C_i\,\mathcal{O}_i \right] + \text{h.c.}\,, \tag{1.1}$$

The operators $\mathcal{O}_i$ are local combinations of quark, lepton, and gluon fields that encode the possible low-energy interaction structures. The Wilson coefficients $\mathcal{C}_i(\mu)$ are short-distance quantities which incorporate the effects of heavy particles and higher-order corrections. In practice, they are obtained by matching the full SM onto the effective theory at the electroweak scale and subsequently evolved down to the b-quark scale using renormalization-group techniques [27]. In this way, the Wilson coefficients act as numerical weights for the various operators, quantifying the relative importance of different interaction channels.

The operator basis includes several distinct classes. The current-current operator $\mathcal{O}_{1,2}^c$ contains four-quark structures with internal charm fields, and although they originate at tree level, they influence rare processes indirectly via operator mixing. The quantum chromodynamics (QCD) penguin operators $\mathcal{O}_3$–$\mathcal{O}_6$ describe gluonic four-quark interactions relevant to hadronic decays of the type $b \to sq\bar{q}$. Electromagnetic and chromomagnetic dipole operators $\mathcal{O}_{7\gamma}$ and $\mathcal{O}_{8g}$, mediate $b \to s\gamma$ and gluonic penguin effects, respectively. In the semileptonic sector, the operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$ describe vector and axial-vector interactions with charged leptons, governing processes such as $b \to s\ell^+\ell^-$.

Finally, the neutrino operator $\mathcal{O}_L^\nu$ mediates the $b \to s\nu\bar{\nu}$ transition, which, though experimentally challenging, is theoretically very clean. After renormalization group running to the scale of the b-quark mass, typical SM values for the Wilson coefficients are approximately $C_7^{\text{eff}} \sim -0.3, C_9^{\text{eff}} \sim +4.1, C_{10}^{\text{eff}} \sim -4.3$ [32]. These numbers encapsulate the relative strength of dipole, vector, and axial-vector contributions in $b \to s$ transitions.

The phenomenology of these operators is well illustrated by the inclusive decay channels. The radiative decay $b \to s\gamma$ is dominated by the electromagnetic dipole operator $O_{7\gamma}$ and has a branching fraction $\mathcal{B}_{\text{SM}}(B \to X_s\gamma) = (3.40 \pm 0.17) \times 10^{-4}$ [32], in excellent agreement with the world average $\mathcal{B}_{\text{exp}} = (3.49 \pm 0.19) \times 10^{-4}$ [33]. The semileptonic process $b \to s\ell^+\ell^-$ ($\ell = e, \mu$) involves interference among $\mathcal{O}_7$, $\mathcal{O}_9$, and $\mathcal{O}_{10}$, yielding a branching fraction at the level of $\mathcal{B}_{\text{SM}}(B \to X_s\ell^+\ell^-) \simeq (4-6) \times 10^{-6}$ [34]. Finally, the invisible decay $b \to s\nu\bar{\nu}$ is the cleanest theoretically, since it is described by a single left-handed operator with a coefficient $C_L^\nu = -X_t/\sin^2\theta_W$. The resulting spectrum has the simple form $d\Gamma/d\hat{s} \propto |C_L^\nu|^2(1-\hat{s})^2(1+2\hat{s})$ with three neutrino flavours, yielding an integrated branching fraction $\mathcal{B}_{\text{SM}}(B \to X_s\nu\bar{\nu}) \simeq (3-4) \times 10^{-5}$ [35].

Unlike the radiative or charged-lepton modes, $b \to s\nu\bar{\nu}$ is free from long-distance electromagnetic effects, making it exceptionally clean from a theoretical perspective. Although experimentally challenging due to the invisibility of neutrinos, the decay offers a unique and precise probe of flavor-changing neutral currents. In the following, we discuss the $b \to s\nu\bar{\nu}$ transition within the SM framework and explore its sensitivity to potential new physics contributions.

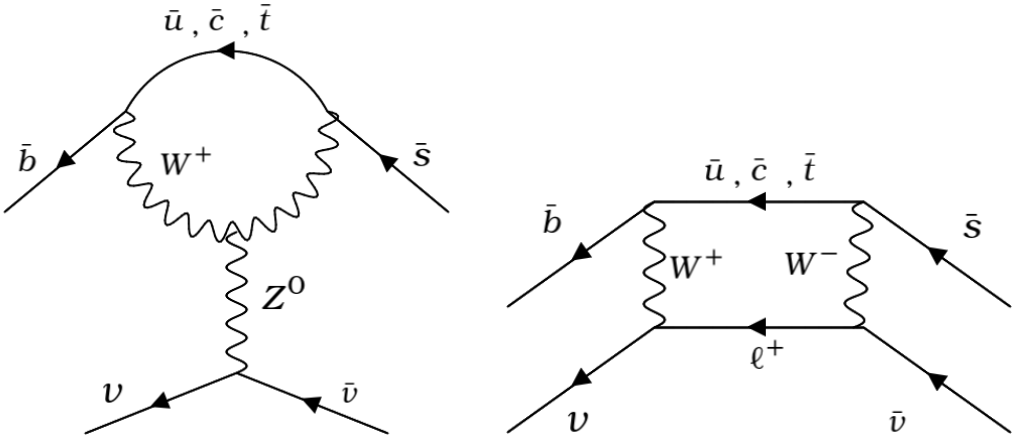

Figure 1.1: Lowest-order diagrams for the $b \to s\nu\bar{\nu}$ decay in the SM are of the penguin or box type.

1.3 $b \to s\nu\bar{\nu}$ Transitions in the Standard Model

The rare decay $b \to s\nu\bar{\nu}$ is among the cleanest FCNCs in the SM. At the weak scale they arise from the electroweak box and Z-penguin diagrams, as shown in Fig. 1.1. The corresponding effective Hamiltonian is particularly clean, as only a single left-handed operator contributes:

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}}\, V_{tb}V_{ts}^* \,\frac{\alpha_e}{4\pi s_W^2}\, C_L\, \mathcal{O}_L, \qquad \mathcal{O}_L^\mu = (\bar{s}\gamma^\mu P_L b)(\bar{\nu}\gamma_\mu P_L \nu), \tag{1.2}$$

The Wilson coefficient is given in the SM by $C_L^{\text{SM}} = -X_t$, where the short-distance function X_t includes next-to-leading QCD and full two-loop electroweak corrections, yielding percent-level theoretical precision [36]. The momentum transfer to the neutrino pair is defined as

$$q^\mu = (p_\nu + p_{\bar{\nu}})^\mu, \tag{1.3}$$

which represents the invariant mass squared of the di-neutrino system. Since neutrinos are massless in SM, the leptonic current is conserved, $q_\mu\, \bar{\nu}\gamma^\mu P_L \nu = 0$. This property eliminates hadronic contributions proportional to q_μ, leaving only the transverse components of the hadronic current. As a result, the pseudoscalar and vector final states, K^+ and K^{*0}, exhibit distinct but constrained structures.

1.3.1 Decay Rate of $B^+ \to K^+\nu\bar{\nu}$

Exclusive $b \to s\nu\bar{\nu}$ decays are governed by hadronic matrix elements of the quark current $\bar{s}\gamma_\mu b$, which encode the non-perturbative QCD dynamics of the B meson transitioning to a light pseudoscalar or vector meson. These matrix elements are parameterized by Lorentz-invariant form factors that depend only on the momentum transfer 1.3.
For the pseudoscalar final state, the most general decomposition of the vector current is

$$\langle K(p_K)|\,\bar{s}\gamma_\mu b\,|B(p_B)\rangle = f_+(q^2)\Big[(p_B + p_K)_\mu - \frac{m_B^2 - m_K^2}{q^2}\, q_\mu\Big] + f_0(q^2)\,\frac{m_B^2 - m_K^2}{q^2}\, q_\mu. \tag{1.4}$$

As a result, the contributions proportional to q_μ vanish, and only the vector form factor $f_+(q^2)$ survives, while the scalar form factor $f_0(q^2)$ enters only through the constraint

$$f_+(0) = f_0(0). \tag{1.5}$$

The differential decay rate is then given by

$$\frac{d\Gamma(B \to K\nu\bar{\nu})}{dq^2} = \frac{G_F^2 \alpha_e^2}{256\pi^5} \frac{|V_{tb}V_{ts}^*|^2}{s_W^4} |X_t|^2 \frac{\lambda^{3/2}(m_B^2, m_K^2, q^2)}{m_B^3} |f_+(q^2)|^2 \times 3, \tag{1.6}$$

where $\lambda(a,b,c) = (a+b-c)^2 - 4ab$, and the factor of 3 accounts for the neutrino flavors. The form factor $f_+(q^2)$ describes the overlap between the B meson and the recoiling K meson, as probed by the conserved vector current. It determines both the normalization and shape of the differential spectrum.
The High Precision Lattice QCD collaboration (HPQCD) provides high-precision lattice QCD results for $f_{+,0}(q^2)$ using the Bourrely-Caprini-Lellouch (BCL) z-expansion [37]. The conformal variable is defined as

$$z(q^2, t_0) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}}, \qquad t_\pm = (m_B \pm m_K)^2, \tag{1.7}$$

with t_0 chosen near $t_- = (m_B - m_K)^2$ to minimize $|z|$ in the physical region. The form factors are then expanded as

$$f_+(q^2) = \frac{1}{1 - q^2/m_{B_s^*}^2} \sum_{k=0}^{K} b_k^{(+)} z(q^2, t_0)^k, \tag{1.8}$$

$$f_0(q^2) = \sum_{k=0}^{K} b_k^{(0)} z(q^2, t_0)^k, \tag{1.9}$$

with the kinematic constraint $f_+(0) = f_0(0)$. The coefficients $b_k^{(+,0)}$ and their covariance matrices are provided by HPQCD.
The most recent SM prediction from HPQCD gives [37]

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu})_{\rm SM}^{\rm total} = (5.58 \pm 0.37) \times 10^{-6}, \tag{1.10}$$

where $(0.61 \pm 0.06) \times 10^{-6}$ arises from the long-distance cascade $B^+ \to \tau^+\nu$, $\tau^+ \to K^+\bar{\nu}$. This tree-level cascade contribution, as shown in Fig. 1.2, though not part of the genuine FCNCs amplitude, mimics the same final state and therefore can be separated from the short-distance signal. $b \to s\nu\bar{\nu}$ contribution. A complementary analysis [38] finds

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu})_{\rm SM} = (5.22 \pm 0.15 \pm 0.28) \times 10^{-6}, \tag{1.11}$$

based on updated lattice and light-cone sum-rule (LCSR) form factors. The two determinations are fully consistent, with the small difference traceable to the explicit treatment of the τ-mediated contribution and the choice of form-factor inputs.

1.3.2 Decay Rate of $B^0 \to K^{*0}\nu\bar{\nu}$

The case of the vector meson K^* is more intricate, since both vector and axial-vector form factors contribute. The hadronic current is parameterized as [38]

$$\begin{aligned}\langle K^*(p,\epsilon)|\, \bar{s}\gamma_\mu(1-\gamma_5)b\, |B(p_B)\rangle = {} & \epsilon_{\mu\nu\rho\sigma}\, \epsilon^{*\nu} p_B^\rho p^\sigma \frac{2V(q^2)}{m_B + m_{K^*}} - i\epsilon_\mu^* (m_B + m_{K^*}) A_1(q^2) \\ & + i(\epsilon^* \cdot q) \frac{(p_B + p)_\mu}{m_B + m_{K^*}} A_2(q^2) \\ & + i(\epsilon^* \cdot q) \frac{2m_{K^*}}{q^2} q_\mu \big[A_3(q^2) - A_0(q^2)\big],\end{aligned} \tag{1.12}$$

with the auxiliary form factor defined by

$$A_3(q^2) = \frac{m_B + m_{K^*}}{2m_{K^*}} A_1(q^2) - \frac{m_B - m_{K^*}}{2m_{K^*}} A_2(q^2). \tag{1.13}$$

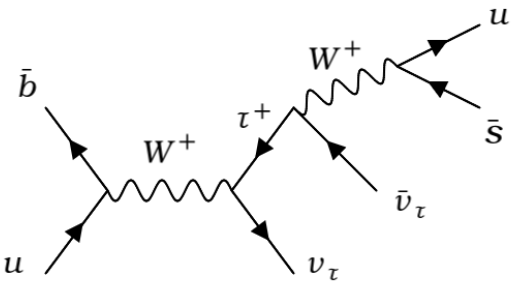

Figure 1.2: The long-distance double-charged-current diagram arising at tree level in the SM also contributes to the $B^+ \to K^+ \nu\bar{\nu}$ decay.

In the limit of massless neutrinos, the helicity amplitudes depend only on the form factors $V(q^2)$, $A_1(q^2)$, and $A_2(q^2)$, while the pseudoscalar form factor $A_0(q^2)$ does not contribute.
The relevant helicity amplitudes of $B \to K^* \nu\bar{\nu}$ are

$$H_\pm(q^2) = (m_B + m_{K^*})\, A_1(q^2) \;\mp\; \frac{\sqrt{\lambda_*}}{m_B + m_{K^*}}\, V(q^2), \tag{1.14}$$

$$H_0(q^2) = \frac{1}{2m_{K^*}\sqrt{q^2}}\Big[(m_B^2 - m_{K^*}^2 - q^2)(m_B + m_{K^*})A_1(q^2) - \frac{\lambda_*}{m_B + m_{K^*}}\, A_2(q^2)\Big], \tag{1.15}$$

with $\lambda_* = \lambda(m_B^2, m_{K^*}^2, q^2)$. At first glance H_0 where $V(q^2)$ governs transverse singular as $q^2 \to 0$. This is avoided by the condition $A_3(0) = A_0(0)$, which enforces a cancellation in the numerator of H_0. Thus, the longitudinal helicity amplitude remains finite. In practice, it is convenient to introduce the combination $A_{12}(q^2)$ [38], which is constructed precisely to guarantee the proper behavior of $H_0(q^2)$ near $q^2 = 0$.
The helicity amplitudes $H_\pm(q^2)$ and $H_0(q^2)$ capture the hadronic dynamics. The full angular dependence of the decay enters when considering the angle θ between the K^* momentum in the B rest frame and the direction of one neutrino in the $\nu\bar{\nu}$ rest frame. The double-differential rate is

$$\frac{d\Gamma(B \to K^* \nu\bar{\nu})}{dq^2 d\cos\theta} = \mathcal{N}(q^2)[\frac{3}{4}(1 + \cos^2\theta)(|\mathrm{H}_+|^2 + |\mathrm{H}_-|^2) + \frac{3}{2}\sin^2\theta|\mathrm{H}_0|^2] \tag{1.16}$$

with

$$\mathcal{N}(q^2) = \frac{G_F^2 \alpha_e^2}{256\pi^5}\, \frac{|V_{tb}V_{ts}^*|^2}{s_W^4}\, |X_t|^2\, \frac{\sqrt{\lambda_*}\, q^2}{m_B^3} \times 3. \tag{1.17}$$

Integration over $\cos\theta$ recovers the single-differential rate. The differential decay rate takes the form

$$\frac{d\Gamma(B \to K^* \nu\bar{\nu})}{dq^2} = \mathcal{N}(q^2)(|H_0|^2 + |H_+|^2 + |H_-|^2). \tag{1.18}$$

A particularly clean observable is the longitudinal polarization fraction

$$F_L(q^2) = \frac{|H_0|^2}{|H_0|^2 + |H_+|^2 + |H_-|^2}, \tag{1.19}$$

which depends almost entirely on the hadronic form factors and is largely insensitive to short-distance physics [39].
Each form factor is expressed in a pole-dominated z-expanded form:

$$F(q^2) = \frac{1}{1 - q^2/m_{\mathrm{pole}}^2} \sum_{k=0}^{K} a_k^F\, z(q^2, t_0)^k, \qquad F \in \{V, A_1, A_2\}. \tag{1.20}$$

The coefficients a_k^F are fitted to LCSR results at low q^2 and constrained by lattice data at higher q^2.
Using this form-factor treatment and two-loop–electroweak improved short distance, the SM expectation is [38]

$$\mathcal{B}(B^0 \to K^{*0}\nu\bar{\nu})_{\rm SM} = \left(9.47 \pm 1.28_{\rm FF} \pm 0.57_{\rm param}\right) \times 10^{-6}. \tag{1.21}$$

where FF denotes the uncertainty from the form factors, and param denotes parametric uncertainties such as CKM elements and quark masses.
The $B \to K$ channel depends on the single form factor $f_+(q^2)$, known to high precision from lattice QCD with a BCL parametrization. The $B \to K^*\nu\bar{\nu}$ channel requires three independent form factors V, A_1, A_2, obtained from LCSR and lattice-informed fits, with A_{12} ensuring proper kinematics. Together, these parametrizations provide the essential non-perturbative input for reliable SM predictions of branching ratios and polarization observables in $b \to s\nu\bar{\nu}$ transitions.

1.4 New Physics Scenarios in $b \to s\nu\bar{\nu}$

The SM predicts that the decays $B \to K^{(*)}\nu\bar{\nu}$ proceed via electroweak Z-penguin and W-box diagrams 1.1. The amplitudes are doubly suppressed by the loop factor and by the small CKM combination $V_{tb}V_{ts}^*$, leading to branching fractions of order $\mathcal{O}(10^{-6})$ [37, 38]. Moreover, the effective Hamiltonian is dominated by a single left-handed operator

$$\mathcal{O}_L^\mu = (\bar{s}\gamma^\mu P_L b)(\bar{\nu}\gamma_\mu P_L \nu), \;\; \mathcal{O}_R^\mu \simeq 0 \tag{1.22}$$

Any significant deviation from this structure would be a sensitive probe of new physics beyond the SM. Broadly, new physics scenarios that can enhance the rates fall into three categories, depending on the nature of the mediator and final state: tree-level FCNCs mediators, invisible mediator scenarios, and sterile-neutrino extensions. The central mechanism is to remove the loop and CKM suppression that makes the SM process so rare, allowing sizeable enhancements of the branching fractions or the appearance of right-handed operators.

Tree-level FCNCs mediators. The first class of new-physics scenarios involves mediators that couple only to SM fields and generate tree-level FCNCs without requiring a dark sector. A canonical example is a heavy neutral Z' boson with flavour-violating couplings to quarks and couplings to neutrinos. Such a state induces direct contributions to the effective operators $\mathcal{O}_L^\nu$ or $\mathcal{O}_R^\nu$ [40–42], as shown in Fig. 1.3 (left). The relevant interaction Lagrangian can be written as

$$\mathcal{L} \supset \left[g_{bs}^L\, \bar{s}\gamma_\mu P_L b + g_{bs}^R\, \bar{s}\gamma_\mu P_R b\right] Z'^\mu + g_\nu\, \bar{\nu}\gamma_\mu P_L \nu\, Z'^\mu, \tag{1.23}$$

where the new couplings $g_{bs}^{L,R}$ act as direct flavour-changing parameters, in contrast to the CKM-suppressed loop structure of the SM.
Leptoquark models achieve a similar effect by exchanging scalar or vector particles that link quarks and leptons, producing $b \to s\nu\bar{\nu}$ at tree level through Yukawa-type couplings [43–46], as shown in Fig. 1.3 (right). For instance, a scalar triplet S_3 and a vector singlet U_1 couple via

$$\mathcal{L} \supset y^{ij}\, \bar{Q}_i^c\, i\sigma_2(\vec{\tau}\cdot\vec{S}_3)\, L_j \;+\; x_L^{ij}\, \bar{Q}_i\gamma_\mu L_j\, U_1^\mu, \tag{1.24}$$

so that off-diagonal entries in y^{ij} or x_L^{ij} directly induce FCNCs with neutrinos in the final state.
A further example is provided by R-parity-violating supersymmetry. Squark exchange induced by the $\lambda' LQD^c$ superpotential operator generates a tree-level four-fermion interaction between quarks and neutrinos [47]. The effective interaction arises from

$$\mathcal{L} \supset \lambda'_{i2k}\, \tilde{d}^*_{Rk}\, \bar{\nu}_i P_L s + \lambda'_{i3k}\, \tilde{d}^*_{Rk}\, \bar{\nu}_i P_L b, \tag{1.25}$$

which, after integrating out the squark, leads to effective operators of the form $(\bar{s}\gamma_\mu P_R b)(\bar{\nu}_i\gamma^\mu P_L \nu_i)$.
In all of these cases, the SM suppression is lifted because the new mediator provides a direct, tree-level FCNCs involving only SM neutrinos. As a result, the Wilson coefficients $C_{L,R}^\nu$ can receive large additional contributions, and the branching fractions of $B \to K^{(*)}\nu\bar{\nu}$ may be substantially enhanced.

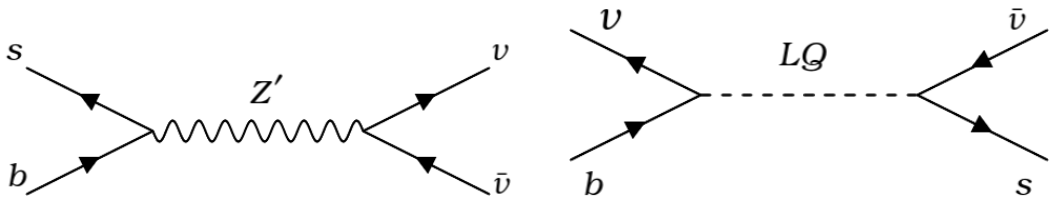

Figure 1.3: The beyond SM diagrams for the $b \to s\nu\bar{\nu}$ decay via new mediators, Z' (left) and leptoquark (right), in the tree-level.

Invisible-mediator scenarios. The second class of models introduces new mediators that couple to the b–s current and decay invisibly, either into SM neutrinos or into dark-sector states χ. In these models, the severe loop and CKM suppression of the SM is replaced by tree-level couplings, potentially leading to order-one enhancements of the rate. A light mediator typically produces a distortion or or excess in the low-q^2 region of the missing-mass spectrum, while heavy mediators lead to an overall rescaling of the rate.
A prototypical example is the Higgs-portal scalar S (HS), which mixes with the SM Higgs and couples directly to $\bar{s}b$ [48–50]. The relevant interactions are

$$\mathcal{L} \supset y_{bs}\, \bar{s} P_R b\, S + y_\nu\, S\, \bar{\nu}\nu + \lambda_{HS}\, S^2 |H|^2 + \kappa\, S |H|^2, \tag{1.26}$$

Here $b \to sS$ occurs at tree level, and the invisible decay $S \to \nu\nu$ or $\chi\chi$ mimics the SM signature.
Another realization is provided by dark photons or new $U(1)_D$ gauge bosons [51, 52], with interactions

$$\mathcal{L} \supset g_{bs}^{A'}\, \bar{s}\gamma^\mu P_{L,R} b\, A'_\mu + g_D\, \bar{\chi}\gamma^\mu \chi\, A'_\mu - \frac{\epsilon}{2}\, F'_{\mu\nu} B^{\mu\nu}, \tag{1.27}$$

where $g_{bs}^{A'}$ is an effective flavour-changing coupling, g_D is the dark gauge coupling, and ϵ parameterizes kinetic mixing with hypercharge. In this case, a vector mediator replaces the SM penguin, inducing the $(\bar{s}\gamma_\mu b)(\bar{\chi}\gamma^\mu\chi)$ at tree level.
Axion-like particles or Majorons a can also mediate invisible channels through derivatives interactions [53, 54]

$$\mathcal{L} \supset \frac{\partial_\mu a}{f_a}\, \bar{s}\gamma^\mu \left(c_V^{sb} + c_A^{sb}\gamma_5\right) b, \tag{1.28}$$

if a decays invisibly, the final state is again indistinguishable from $\nu\bar{\nu}$.
Extended Higgs sectors such as two-Higgs-doublet models can permit tree-level flavour-changing couplings of a neutral Higgs boson [51, 55]. A generic structure is

$$\mathcal{L} \supset Y_{bs}\, \bar{s} P_R b\, H^0 + y_\chi\, H^0\, \bar{\chi}\chi, \tag{1.29}$$

so that $b \to sH^0$ occurs directly, followed by the invisible decay $H^0 \to \chi\chi$.
Finally, a more model-independent SM effective field theory approach extended with dark-sector currents (dark-standard-model-effective-field-theory). [41, 42]. The leading operator takes the form

$$\mathcal{L} \supset \frac{c_L}{\Lambda^2}\, (\bar{s}\gamma_\mu P_L b)(\bar{\chi}\gamma^\mu\chi), \tag{1.30}$$

capture the low-energy effects of integrating out a heavy invisible mediator.
In all of these invisible mediator scenarios, the essential feature is that a new scalar or vector boson mediates $b \to s$ transitions at tree level and subsequently decays invisibly. As a result, the suppression characteristic of the SM loop is removed, and the branching fractions can receive order-one enhancements, often accompanied by distinctive low-q^2 distortions in the missing-mass spectrum.

Sterile-neutrino extensions. The third class enlarges the neutrino sector by introducing sterile neutrinos, often referred to as heavy neutral leptons (HNLs). These are gauge-singlet fermions that mix with the active neutrinos and thus can appear as invisible final states in B decays. The minimal interaction takes the form

$$\mathcal{L} \supset -y_N \, \bar{L}\tilde{H}N - \tfrac{1}{2} M_N \bar{N}^c N, \tag{1.31}$$

where y_N is a Yukawa coupling, M_N is the Majorana mass of the sterile state, and $\tilde{H} = i\sigma_2 H^*$ is the conjugate Higgs doublet. After electroweak symmetry breaking, this generates an active-sterile mixing angle $U_{\ell N} \sim y_N v/M_N$, which allows the sterile neutrino N to participate in weak processes.
In this framework, the decays such as $b \to s\nu N$ or $b \to sNN$ occur at tree level, opening new invisible channels that add to the SM $b \to s\nu\nu$ rate [56]. Consequently, $\mathcal{B}(B \to K^{(*)}$ + invisible) need not be saturated by SM neutrinos alone, but could also include contributions from sterile states. In contrast to the invisible mediator scenarios, the enhancement here is directly tied to the mechanism of neutrino mass generation through Yukawa couplings and Majorana terms. Characteristic signatures include modifications of the missing-mass spectrum and, if the sterile neutrino is sufficiently long-lived, displaced vertices that could be observable in dedicated searches.
It should be emphasized that, although some of these scenarios have recently attracted renewed attention in light of the published $B^+ \to K^+\nu\bar{\nu}$ result [1] (to be discussed in Part 2), the purpose of the present discussion is broader: to outline the possible theoretical scenarios that can contribute to $B \to K^{(*)}\nu\bar{\nu}$ decays, thereby providing a comprehensive picture of the landscape beyond the SM.

1.5 Overview of the Previous Searches

The search for $B \to K^{(*)}\nu\bar{\nu}$ decays is strongly motivated theoretically, but presents significant experimental challenges because the two neutrinos escape detection. Consequently, the signal must be reconstructed indirectly from the missing energy and momentum in the event. The previous generation of B factories enabled the Belle and BaBar experiments to perform dedicated searches for these decays. No signal has been observed so far, and only upper limits on the branching fractions have been reported.
All previous searches employed either the hadronic tagging analysis (HTA), the semileptonic tagging analysis (STA), or the inclusive tagging analysis (ITA) (see Sec. 3.1). Table 1.1 summarizes the experimental results of previous searches for $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ decays, including the collaboration, integrated luminosity, analysis technique, and the resulting upper limits on the branching fractions at 90% confidence level. For $B^0 \to K^{*0}\nu\bar{\nu}$ decays, the reconstructed decay modes of K^{*0} are listed explicitly.

Experiment	Luminosity	Technique	Decay Channel(s)	Limits(90% CL)$\times 10^{-5}$
Belle II	63 fb^{-1}	ITA	$B^+ \to K^+\nu\bar{\nu}$	4.1 [57]
Belle	711 fb^{-1}	HTA	$B^+ \to K^+\nu\bar{\nu}$	5.5 [58]
			$B^0 \to K^{*0}\nu\bar{\nu}$, $K^{*0} \to K^+\pi^-$	5.5 [58]
Belle	711 fb^{-1}	STA	$B^+ \to K^+\nu\bar{\nu}$	1.9 [59]
Belle	711 fb^{-1}	STA	$B^0 \to K^{*0}\nu\bar{\nu}$, $K^{*0} \to K^+\pi^-$	1.8 [59]
BABAR	429 fb^{-1}	HTA	$B^+ \to K^+\nu\bar{\nu}$	3.7 [60]
			$B^0 \to K^{*0}\nu\bar{\nu}$, $K^{*0} \to K^+\pi^-, K^0_S\pi^0$	9.3 [60]
BABAR	418 fb^{-1}	STA	$B^+ \to K^+\nu\bar{\nu}$	1.3 [61]

Table 1.1: Experimental results of previous searches for $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ decays with collaboration name, luminosity of data, measured technique and measured upper limit on the branching fraction at 90% confidence level.

Chapter 2

The Belle II experiment

The Belle II experiment is a next-generation facility dedicated to precision studies of the flavor sector of the SM and to the search for physics beyond it through indirect signatures [62]. Situated at the interaction point of the SuperKEKB collider in Tsukuba, Japan, Belle II builds upon the legacy of its predecessor, Belle, and forms the core of what is known as a B factory—a facility engineered for the prolific production and precise study of B mesons, as well as other heavy flavor states. The primary physics motivation for Belle II lies in the rich phenomenology of the heavy quark flavor sector, particularly processes involving bottom and charm quarks, and in complementary precision studies of the charged lepton sector. These areas provide sensitive probes of the SM through over-constrained tests of the CKM matrix, searches for CP violation (CPV) beyond the Standard Model paradigm, and measurements of rare or forbidden decays that could reveal virtual effects of new particles or interactions. Belle II operates at the $\Upsilon(4S)$ resonance, a bottomonium state just above the kinematic threshold for $B\bar{B}$ pair production. The electron–positron (e^+e^-) collision environment provides a uniquely clean setting for studying such decays, free from the complexities of hadronic initial states. An essential feature of the B factory design is the use of asymmetric beam energies—the electron and positron beams have unequal momenta—resulting in a net boost to the center-of-mass frame. At the $\Upsilon(4S)$, this corresponds to $\beta\gamma \approx 0.28$, allowing reconstruction of proper-time differences between B meson decays [62]. This asymmetry is essential for time-dependent analyses of neutral B meson decays, where precise measurements of decay time differences enable detailed studies of CP-violating phenomena arising from the interference between mixing and decay. Beyond CPV, Belle II addresses a broad program at the intensity frontier, where the unprecedented luminosity of SuperKEKB allow for precision tests of the SM and enhances sensitivity to rare processes. These include searches for lepton flavor violation, measurements of CKM matrix elements such as $|V_{ub}|$ and $|V_{cb}$, tests of lepton universality, and the exploration of potential contributions from new physics scenarios such as supersymmetry, leptoquarks, or dark sector mediators. In this context, Belle II represents a major advancement in flavor physics, leveraging both technological upgrades in detection and dramatic increases in luminosity provided by the SuperKEKB accelerator. Together, they enable Belle II to perform high-precision measurements and sensitive searches that complement the energy frontier probed by hadron colliders, offering a unique window into phenomena that may lie beyond the reach of direct production. Table 2.1 summarizes representative cross sections for various physics processes in e^+e^- collisions at $\sqrt{s} = 10.58$ GeV, near the $\Upsilon(4S)$ resonance. The highest cross section arises from QED processes such as Bhabha scattering and multi-lepton final states, while hadronic processes like $q\bar{q}(\gamma)$ and $\Upsilon(4S) \to B\bar{B}$ occur at the nanobarn level. These values guide event generation and background estimation in Belle II analyses. The following sections describe the detector subsystems in detail, followed by the trigger, data acquisition, and reconstruction frameworks that enable the Belle II physics program.

2.1 SuperKEKB

SuperKEKB is a next-generation electron–positron collider developed at KEK in Tsukuba, Japan, as a high-luminosity upgrade of the earlier KEKB facility. Engineered to deliver a 40-fold increase in instantaneous lumi-

Table 2.1: The presentative cross sections for physics processes in e^+e^- collisions at $\sqrt{s} = 10.58$ GeV, near the $\Upsilon(4S)$ resonance [62].

Physics Process	**Cross Section[nb]**	**Generator**
$e^+e^- \to \Upsilon(4S)$	1.11	Pythia8 [63], EvtGen [64]
$e^+e^- \to u\bar{u}(\gamma)$	1.61	KKMC [65]
$e^+e^- \to d\bar{d}(\gamma)$	0.40	KKMC
$e^+e^- \to s\bar{s}(\gamma)$	0.38	KKMC
$e^+e^- \to c\bar{c}(\gamma)$	1.30	KKMC
$e^+e^- \to e^+e^-(\gamma)$	300	BABAYAGA.NLO [66]
$e^+e^- \to \mu^+\mu^-(\gamma)$	1.15	KKMC
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.92	KKMC
$e^+e^- \to e^+e^-e^+e^-$	39.7	AAFH [67]
$e^+e^- \to e^+e^-\mu^+\mu^-$	18.9	AAFH
$e^+e^- \to \gamma\gamma(\gamma)$	3.30	BABAYAGA.NLO
$e^+e^- \to \nu\bar{\nu}(\gamma)$	0.25×10^{-3}	KKMC

nosity over its predecessor, SuperKEKB provides the intense beam environment necessary for the Belle II experiment's precision measurements in flavor physics. The collider with a design goal of 8×10^{35} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [68], and as of December 2024 has achieved a peak luminosity 5.1×10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ corresponding to an integrated dataset of 575 fb^{-1}.
The structural layout of SuperKEKB is shown in Figure 2.1 of the design reference [68], which provides a schematic overview of the entire accelerator complex. SuperKEKB is composed of two independent storage rings: the High Energy Ring (HER) for 7 GeV electrons and the Low Energy Ring (LER) for 4 GeV positrons. These rings share a 3-km tunnel and intersect at a single interaction point (IP) within the Belle II detector. The schematic highlights not only the storage rings but also essential infrastructure such as the injector linear accelerator, the positron damping ring, and beam transport lines. A notable feature of the interaction region is the integration of advanced superconducting quadrupole magnets, final-focus systems, and specialized beam optics designed to handle high-current, low-emittance beams.

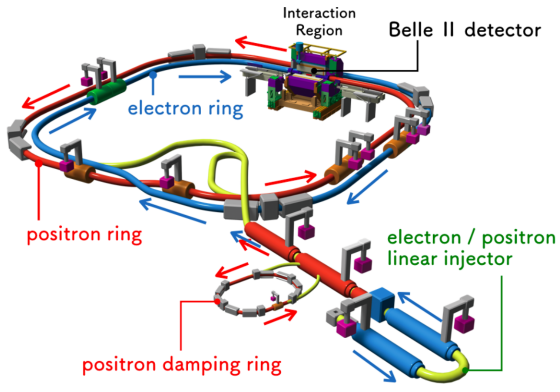

Figure 2.1: Schematic layout of the SuperKEKB accelerator complex [68], showing the HER (7 GeV e^-) and LER (4 GeV e^+) storage rings, the interaction point (IP), and supporting infrastructure such as the injector linac, damping ring, and beam transport lines.

A central innovation of SuperKEKB is the adoption of the nanobeam collision scheme [68, 69], depicted in Figure 2.2. This scheme employs a large horizontal crossing angle—83 mrad in total—so that the beams intersect at a shallow angle rather than head-on. The geometry increases the Piwinski angle, which in turn allows a dramatic reduction in the vertical beta function at the interaction point (β_y^*) to approximately 1 mm.

As a result, the vertical beam size is compressed to the sub-micrometer scale (σ_y^* ≈250 nm), significantly enhancing the probability of particle collisions per bunch crossing. Importantly, the crossing angle also suppresses the hourglass effect by reducing the longitudinal overlap of the bunches, enabling strong focusing without requiring ultrashort bunch lengths. SuperKEKB employs asymmetric beam energies (7 GeV e^-, 4 GeV e^+), inherited from KEKB and maintained to preserve compatibility with the existing tunnel and detector geometry. This asymmetric beam energies yield a Lorentz boost of $\beta\gamma \approx 0.28$, which is essential for time-dependent CPV studies. Accelerator optimization instead concentrates on sustaining ultra-low emittance, minimizing beam–beam effects, and preserving collision precision under high-current operation.

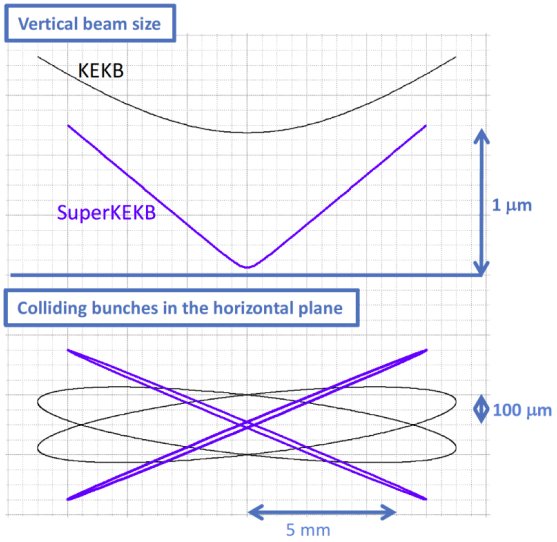

Figure 2.2: Illustration of the nanobeam collision scheme implemented at SuperKEKB. The upper panel shows the reduction of the vertical beam size compared to KEKB, achieved by squeezing the vertical beta function at the interaction point. The lower panel depicts the bunches colliding in the horizontal plane with a total crossing angle of 83 mrad, which enhances the Piwinski angle and suppresses the hourglass effect [68].

SuperKEKB achieves its performance through a combination of high stored beam currents, tight focusing, and precise orbit control. The table 2.2 below summarizes its key numerical features. These design innovations enable SuperKEKB to sustain ultra-low emittance beams under high current operation, realizing the luminosity required by Belle II.

Parameter	**HER (e^-)**	**LER (e^+)**
Beam energies	7(8) GeV	4(3.5) GeV
Beam current	2.6(1.19)A	3.6(1.64)A
Horizontal crossing angle	41.5(0, 11) mrad	41.5(0, 11) mrad
Vertical beta function at IP β_y^*	0.30(5.9) mm	0.27(5.9) mm
Vertical beam size at IP σ_y^*	62(940) nm	48(940) nm
Piwinski angle	19.3(0) rad	24.6(0) rad

Table 2.2: Key design parameters of the SuperKEKB collider for the High Energy Ring (HER, e^-) and Low Energy Ring (LER, e^+). Values in parentheses indicate the corresponding parameters of the predecessor KEKB machine [68].

2.2 Detector

The Belle II detector [70], as shown in Fig 2.3, is a general-purpose spectrometer optimized for high-precision studies of heavy flavor physics at the SuperKEKB collider. It is designed to operate in a high-luminosity e^+e^- environment, offering excellent vertex resolution, tracking, particle identification, and muon detection. At its core, a 0.6 mm thick beryllium beam pipe with a 10 mm inner radius minimizes multiple scattering and enables precise vertex measurements near the interaction point. Surrounding this is the Vertex Detector (VXD), composed of the Pixel Detector (PXD) and Silicon Vertex Detector (SVD), providing high-resolution vertexing for time-dependent CPV studies. Surrounding the VXD is the Central Drift Chamber (CDC), responsible for tracking and measuring the momentum of charged particles via curvature in the magnetic field. The Electromagnetic Calorimeter (ECL) detects and measures the energy of photons and electrons using $CsI(Tl)$ crystals. Particle identification is handled by two systems: the Time-of-Propagation (TOP) counter in the barrel and the Aerogel Ring Imaging Cherenkov (ARICH) detector in the forward endcap. These distinguish between charged hadrons over a wide momentum range using Cherenkov light detection. The tracking and particle identification (PID) systems are enclosed within a 1.5 T superconducting solenoid magnet, which provides the magnetic field for momentum analysis. Outside the magnet is the KLM system (K_L^0 and μ detector), consisting of resistive plate chambers interleaved with iron plates, used to detect muons and long-lived neutral hadrons. A multi-level trigger system, combining hardware and software, ensures efficient event selection under high background and collision rates. Together, these subsystems provide Belle II with comprehensive capabilities for tracking, vertexing, particle identification, and calorimetry, enabling precision studies of rare processes and stringent tests of the Standard Model.

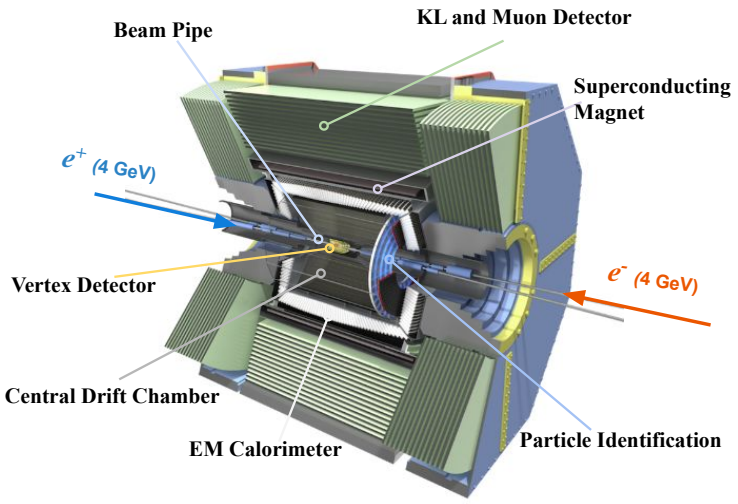

Figure 2.3: Schematic view of the Belle II detector, showing its major subsystems: beam pipe, vertex detector (PXD and SVD), central drift chamber, particle identification devices (TOP and ARICH), electromagnetic calorimeter, superconducting solenoid, and the KLM system [70].

2.2.1 Tracking: VXD

The VXD comprises two subsystems: the PXD and the SVD. Together, they provide six layers of precisely position-sensitive detectors closely surrounding the interaction point. The PXD consists of two cylindrical layers positioned at radii of 14 mm and 22 mm from the beam axis. Each layer comprises 8 and 12 ladders,

respectively (or 16 and 24 modules when accounting for both sides of the interaction region) [71]. The sensitive thickness of the Depleted P-channel Field Effect Transistor (DEPFET) sensors is 75 μm, offering an excellent compromise between mechanical robustness and material budget. The DEPFET matrices are operated at 50 kHz frame rate with a 10 MHz row readout speed, matching the high background conditions expected at SuperKEKB.
Surrounding the PXD is the SVD [72], composed of four layers of double-sided silicon strip detectors (DSSDs), placed at radii from 38 mm to 135 mm. Strip pitch ranges from 50−75 μm (p-side) and 160 μm (n-side), balancing resolution and occupancy. The SVD provides additional tracking redundancy and improves pattern recognition and momentum resolution, particularly for low-momentum tracks. The combined VXD achieves a typical impact parameter resolution of approximately

- $\sim$ 15 μm in the transverse plane
- $\sim$ 20-30 μm along the beam (z) direction for charged tracks with momenta above a few hundred MeV/c

Fig 2.4 shows the xy-plane views of the Pixel Detector (left) and the Silicon Vertex Detector (right), illustrating their geometry around the beam pipe. This high-resolution vertexing capability is essential for precise reconstruction of decay time differences in neutral B^0-$\bar{B}^0$ mixing and for identifying displaced vertices from charm or tau decays in a dense event environment.

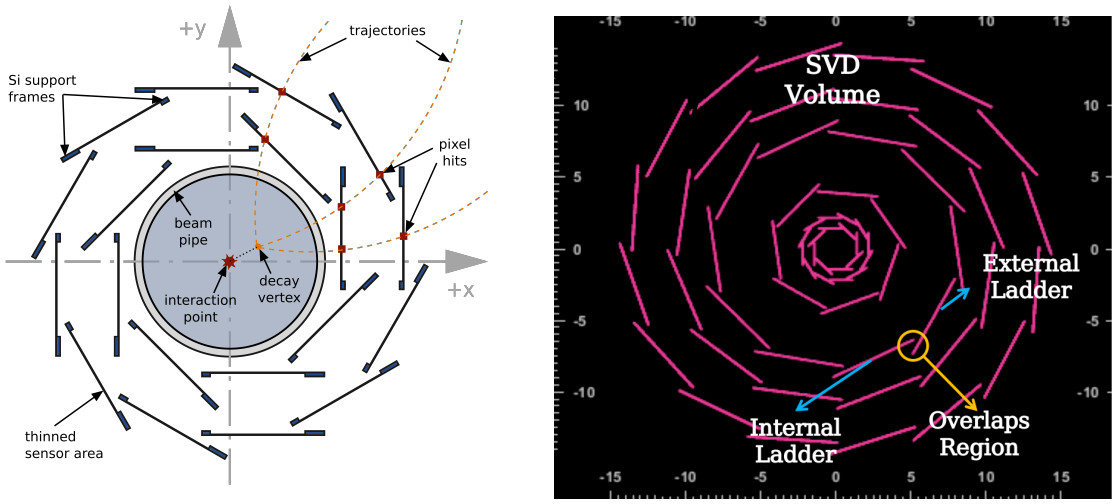

Figure 2.4: xy-plane layout of the two subsystems of the Belle II Vertex Detector. Left (Anselm Baur): PXD, consisting of two cylindrical DEPFET layers at radii of 14 mm and 22 mm. Right [72]: the four-layer SVD composed of double-sided strip sensors, extending the radial coverage to 135 mm. Combined, the PXD and SVD deliver high-resolution vertexing close to the interaction point,

2.2.2 Tracking: CDC

The CDC is the principal tracking device outside the Vertex Detector in Belle II. It serves to measure the trajectories and momenta of charged particles, determine their ionization energy loss (dE/dx), and contribute to particle identification.
The CDC covers a polar angle range of $17° < \theta < 150°$, and consists of 56 concentric wire layers, arranged into alternating axial and stereo layers, further grouped into 9 superlayers. The axial layers have sense wires aligned parallel to the beam axis, providing precise measurements in the transverse plane (r-ϕ). The stereo layers are tilted at small angles, allowing for the determination of the longitudinal (z) coordinate and enabling full 3D track reconstruction. To accommodate the high-luminosity environment of SuperKEKB, the CDC adopts a small-cell structure in its innermost layers, reducing the drift distance and improving

occupancy handling. The drift cells are filled with a gas mixture of 50% helium and 50% ethane (C_2H_6, chosen to minimize multiple scattering while maintaining good spatial resolution.

The CDC achieves a typical spatial resolution of approximately 100 μm per hit, and a momentum resolution of $\delta p_T/p_T \approx 0.3\%$ for tracks with high transverse momentum. Its large outer radius (1.13 m) and length (2.3 m) provide excellent coverage over a wide solid angle, and its integration with the VXD ensures efficient track matching and pattern recognition. Fig 2.5 shows an overview of the CDC in x-y plane geometry and arrangement of axial and stereo wire layers.

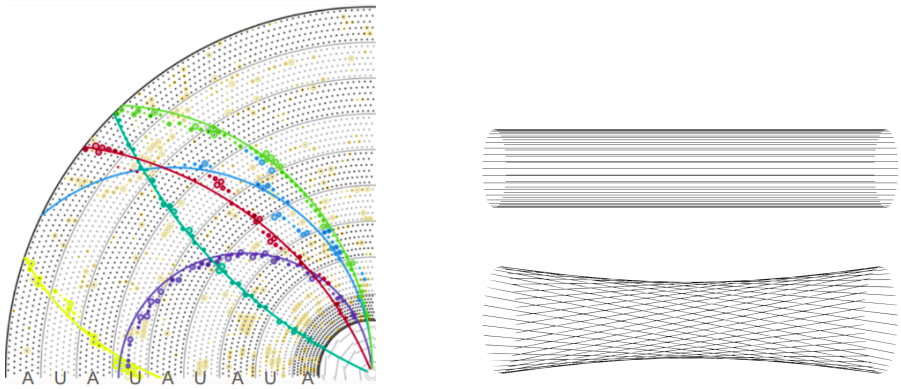

Figure 2.5: Views of the Belle II CDC. Left: xy-plane geometry of the cylindrical detector [73]. Right: arrangement of axial and stereo wire layers, enabling 3D tracking [74].

2.2.3 Particle Identification

Belle II employs two advanced Cherenkov detectors for charged hadron identification, crucial for discriminating kaons and pions with high accuracy over a wide momentum range.

The ARICH, positioned in the forward endcap, uses multilayer silica aerogel radiators to generate Cherenkov light rings [70]. These are imaged by Hybrid Avalanche Photo-Detectors (HAPDs), which is radiation-tolerant, high-gain photon detectors developed specifically for Belle II, to reconstruct the Cherenkov angle, allowing efficient kaon/pion separation up to momenta around 4 GeV/c. Figure 2.6 (left) illustrates the working principle of the proximity-focusing ARICH detector. Charged particles passing through the aerogel radiator emit Cherenkov photons at an angle θ_c relative to their trajectory. These photons propagate through a small gap (the proximity gap) without focusing optics and are detected on a photodetector plane, forming a ring image.

The radius of the Cherenkov ring on the detector plane directly relates to the Cherenkov angle θ_c, which depends on the particle velocity and the aerogel's refractive index. By measuring this ring radius and combining it with momentum information from tracking, the ARICH identifies particle species, effectively distinguishing kaons from pions in the forward region of Belle II.

In the barrel region, the TOP utilizes internally reflected Cherenkov photons within quartz bars, combining precise timing ($\sim$50 ps) and photon hit pattern measurements to identify particle species [75]. Figure 2.6 (right) depicts the TOP counter's side view, where charged particles emit Cherenkov photons inside a quartz bar at the Cherenkov angle θ_c. These photons internally reflect along the bar and reach photodetectors at the end. By measuring their arrival times and positions, the detector reconstructs θ_c, allowing particle identification through velocity-dependent Cherenkov light patterns. The pattern recognition relies on likelihood fits comparing observed photon hit distributions to templates for different mass hypotheses. Together, the ARICH and TOP achieve kaon identification efficiencies of about 90% with pion misidentification probabilities of 5-10%, depending on momentum and polar angle.

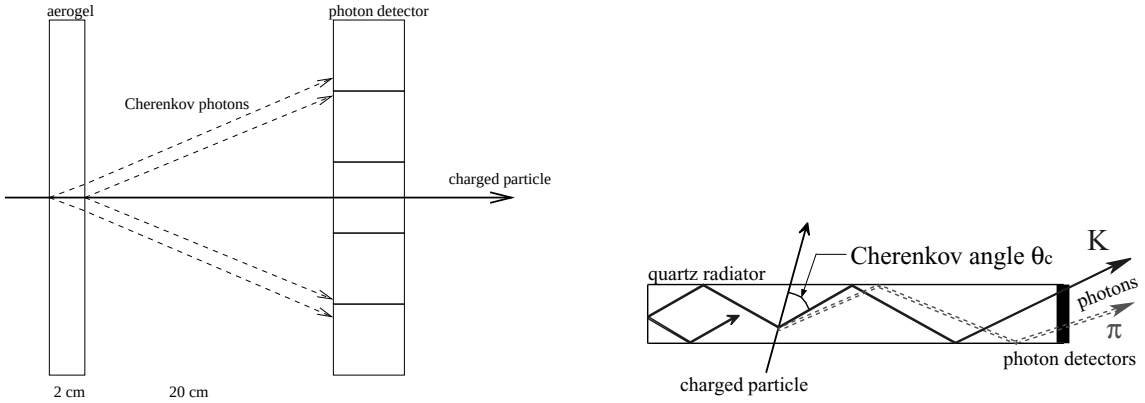

Figure 2.6: Illustration of the Belle II Cherenkov-based PID detectors. Left: proximity-focusing ARICH, where charged particles produce Cherenkov rings detected by HAPDs [70]. Right: TOP counter, in which Cherenkov photons propagate by internal reflection inside quartz bars, with their arrival times and hit positions measured at the bar end to reconstruct the Cherenkov angle [70].

2.2.4 ECL

The Electromagnetic Calorimeter of the Belle II detector is a key subdetector responsible for measuring the energy and position of electrons, positrons, and photons. It plays a central role in electromagnetic particle identification and in reconstructing neutral particles such as π^0 and η mesons [70]. The ECL is composed of 8736 thallium-doped cesium iodide crystals arranged in a barrel (6624 crystals) and two endcap sections (1152 crystals each), providing hermetic coverage over a polar angular range of approximately 12° to 155°. Each $CsI(Tl)$ crystal has dimensions of approximately 6×6 ×30 cm^3, corresponding to about 16 radiation lengths, which is sufficient to contain the full electromagnetic shower. The scintillation light generated by particle interactions is read out via silicon PIN photodiodes coupled to waveform-sampling front-end electronics. These electronics provide both fast and slow signal shaping (typically around 0.2 μs, and 1 μs, respectively), and are capable of time-resolved signal extraction for pile-up rejection and pulse-shape discrimination. The energy measurement in the ECL is based on the total scintillation light yield produced by the electromagnetic shower. The light output is proportional to the energy deposited, with a typical light yield of 50,000 photoelectrons per MeV, but only a fraction is collected. The signals are digitized and calibrated to reconstruct the deposited energy. The upgraded Belle II readout system enables high precision under high-luminosity conditions and improved timing resolution, essential for background suppression.
The performance of the ECL has been validated both in test beam campaigns and using Belle II collision data. The achieved energy resolution is approximately 1.8% at 1 GeV and 4% at 100 MeV. The spatial resolution for photon and electron showers is of the order of 5 mm/$\sqrt{E(\mathrm{GeV})}$. The timing resolution is approximate 2−3 ns for 1 GeV showers, enabling effective discrimination against beam-related backgrounds.
In summary, the Belle II ECL achieves high-resolution, high-efficiency detection of electromagnetic particles, with excellent energy and spatial resolution. Its upgraded design ensures robust operation in the high-luminosity environment of SuperKEKB.

2.2.5 KLM

The K_L^0 and Muon Detector constitutes the outermost subsystem of the Belle II spectrometer and plays a critical role in the identification of muons and long-lived neutral kaons (K_L^0). It is embedded in the iron flux return yoke of the superconducting solenoid and covers the polar angle range from approximately 20° to 155°, providing nearly hermetic coverage for penetrating and weakly interacting particles [70].
The primary detection principle relies on the different interactions of muons and K_L^0 mesons with the detector material. Muons, being minimum ionizing particles, penetrate deeply through multiple layers of

iron absorber and leave aligned hits in the active detection modules. In contrast, K_L^0 mesons are detected via hadronic interactions within the iron, resulting in secondary particle showers that produce spatially localized hit clusters.
The KLM is segmented into alternating layers of iron plates ($\sim$4.7 cm thick) and active detector modules. In the Belle II upgrade, the active layers in the two innermost barrel regions and in the forward and backward endcaps were replaced with plastic scintillator strips coupled to wavelength-shifting fibers and silicon photomultipliers (SiPMs). This upgrade was required to cope with the high background conditions at SuperKEKB, which exceeded the rate capability of the resistive plate chambers (RPCs) used in Belle. The remaining outer barrel layers continue to utilize RPCs due to their robustness and cost-effectiveness.
Each scintillator layer provides both spatial and timing information. The strip geometry yields a spatial resolution of a few centimeters, while the timing information enhances the ability to suppress background and correctly associate hits with the interaction vertex.
The KLM exhibits excellent performance with muon identification efficiencies above 90% for momenta above a few GeV/c, and pion misidentification rates at the level of 2–3%. The K_L^0 detection efficiency is approximately 60–70%, as expected from the stochastic nature of hadronic interactions in dense absorbers. This capability is essential for reconstructing missing energy signatures and neutral particle decays in final states involving neutrinos or invisible particles.
In summary, the KLM provides essential information for muon identification and K_L^0 detection. Its hybrid design, combining scintillator modules and RPCs, ensures robust performance in the challenging high-luminosity environment of SuperKEKB, making it a vital component of the Belle II physics program.

2.3 Trigger and Data Acquisition System of Belle II

The Belle II experiment at SuperKEKB employs a sophisticated Trigger and Data Acquisition (DAQ) system to efficiently manage the high event rates generated by an instantaneous luminosity exceeding $5\times10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. This system is designed to selectively record events of interest while controlling data volume and maintaining high operational efficiency.

2.3.1 Trigger System

The trigger system is structured as a two-level hierarchy:

- **Level-1 (L1) Trigger:** A hardware-based system with a latency of approximately 5 μs, processing inputs primarily from the TOP, CDC, ECL, and KLM. It reduces the raw event rate from the 250 MHz bunch crossing frequency to about 30 kHz. The L1 trigger achieves an efficiency exceeding 90$\%$ across a broad range of physics channels, including low-momentum tracks and missing energy events critical for rare decay analyses [70].

- **High-Level Trigger (HLT):** A software-based trigger running on a large computing farm that performs full event reconstruction with detailed data from all sub-detectors. The HLT further reduces the input rate to a sustainable $\sim$ 5 kHz for permanent storage, in line with bandwidth and storage constraints.

2.3.2 Data Acquisition System

Post L1 acceptance, the DAQ system handles the digitized data streams from over a million detector channels, achieving an aggregate throughput of several gigabytes per second [76]. The modular front-end electronics utilize high-speed optical links to transmit data to event builders that synchronize detector fragments and assemble complete event records with sub-microsecond precision.
The system maintains a minimal dead time below 1%, enabling near-continuous data taking with nearly 100% live-time. Integration with online monitoring and calibration tools ensures data quality and detector stability throughout operations. This process is shown in diagram 2.7

Together, the Belle II trigger and DAQ systems efficiently select and record physics events in the challenging high-luminosity environment of SuperKEKB, maximizing sensitivity to rare processes and precision measurements in flavor physics.

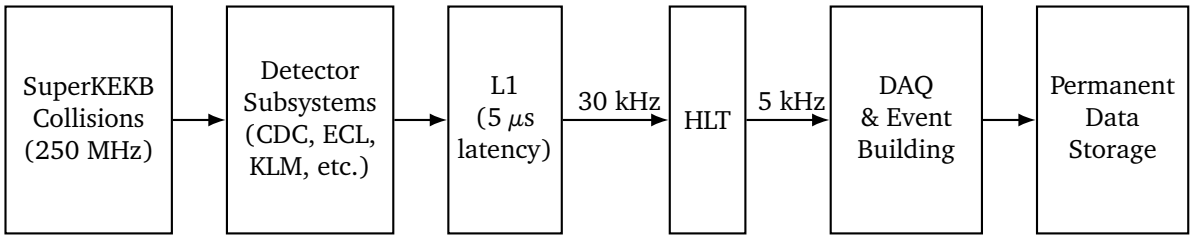

Figure 2.7: Schematic of the Belle II trigger and DAQ chain. Collision events at 250 MHz are processed by the detector subsystems and passed through the two-level trigger hierarchy (L1 hardware and HLT software) before final assembly in the DAQ system and transfer to permanent storage.

2.4 Simulation and Beam-Induced Background in the Belle II Experiment

2.4.1 Analysis Software Framework

The Belle II Analysis Software Framework 2 (`basf2`) is a modular, object-oriented software platform developed for the full chain of data processing in the Belle II experiment [77]. Written primarily in $C++$ with Python bindings, `basf2` supports event generation, detector simulation, reconstruction, calibration, and user-level physics analysis. Its modular design enables flexible pipeline configuration, allowing researchers to tailor processing sequences for different stages of data analysis. Integration with conditions databases and support for parallel processing make `basf2` both scalable and adaptable to evolving computing needs. The framework is also essential for large-scale Monte Carlo campaigns and real-time processing during data-taking periods.
Beam-induced backgrounds represent a significant challenge for the Belle II experiment operating at the high luminosity and beam currents of the SuperKEKB collider. These backgrounds originate from various sources, each contributing to increased detector occupancy and potential radiation damage [78].

2.4.2 Main Background Types

Beam-induced backgrounds [70, 78] at Belle II arise from several key processes intrinsic to high-luminosity e^+e^- collisions at SuperKEKB. The principal sources include:

- **Single-Beam Backgrounds:** These arise from individual beams circulating in the collider and include:
  - Touschek scattering: Intra-beam scattering causing particle losses and secondary radiation.
  - Beam-gas interactions: Collisions between beam particles and residual gas molecules in the beam pipe, producing secondary particles.
  - Synchrotron radiation: Emission of photons by charged particles traveling in curved trajectories.
- **Collision-Induced Backgrounds:** Resulting from interactions between the two beams at the interaction point, leading to:
  - Pile-up: Overlapping events from multiple bunch crossings.
  - Low-energy secondary particles: Produced in hard scatterings, including radiative Bhabha scattering ($e^+e^- \to e^+e^-\gamma$), which generates abundant low-energy particles that can mimic signal events.

Given the varied origins and complex nature of these backgrounds, a multifaceted mitigation strategy is essential to preserve detector performance and longevity. These include the installation of beam collimators to intercept scattered particles before they reach the detector, implementation of radiation shielding around sensitive components, and the use of fast front-end electronics capable of handling high hit rates. The design also anticipates higher-than-nominal background rates by incorporating safety margins into subsystem tolerances. Key mitigation techniques include:

- Implementation of collimation systems to absorb off-axis particles.
- Optimized beam pipe design to reduce residual gas and synchrotron radiation effects.
- Use of detector shielding to protect sensitive components.
- Optimization of beam optics and currents to adjust optics and currents, minimizing background production.

To ensure safe and efficient operation of the detector, extensive efforts have been made to quantify and mitigate beam-induced backgrounds through a combination of simulation and empirical studies. Measurements indicate:

- For single-beam backgrounds, detectors such as the TOP counter and the CDC experience hit rates approximately half of their maximum tolerable levels.
- Collision-induced backgrounds are expected to remain manageable up to a luminosity of $2.8 \times 10^{35}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ with a vertical beta function $\beta_y^* = 0.6$ mm.

2.4.3 Simulation in Belle II

Monte Carlo (MC) simulation plays a central role in the Belle II experiment, enabling realistic modeling of both physics processes and detector responses. The MC framework integrates several specialized event generators to simulate different types of events, providing high-fidelity input for performance studies and physics analyses.
The simulation of $B\bar{B}$ events, which are produced at the $\Upsilon(4S)$ resonance, is primarily handled by `EvtGen` [64], a dedicated generator for heavy flavor physics that includes detailed models of B meson decays and CP violation effects. For continuum $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) events, the `Pythia` [63] generator is employed, offering a comprehensive treatment of parton showering and hadronization. `EvtGen` is interfaced with `Pythia` for fragmentation and hadronization.
The main physics processes and their primary event generators are summarized in Table 2.1. In addition, dedicated generators such as `TAUOLA` [79] (for τ decays), `PHOTOS` [80, 81] (for quantum electrodynamics (QED) radiative corrections), and `BGGEN` [70] (for two-photon and radiative Bhabha processes) are employed in the simulation framework. Leptonic processes such as $e^+e^- \to \mu^+\mu^-$ or $e^+e^- \to \tau^+\tau^-$ are simulated using `KKMC`, which includes full $\mathcal{O}(\alpha^2)$ radiative corrections and spin correlations. Two-photon processes and radiative Bhabha scattering, which contribute to backgrounds and luminosity measurements, are modeled with `BabayagaNLO`, `TREPS` [82] and `BGGEN`.
The detector response is simulated with `Geant4`, which propagates generated particles through a detailed geometrical and material description of the Belle II detector. This includes interactions with detector components, energy deposition, scattering, and secondary particle production. The output is digitized to match the real data acquisition format, enabling seamless integration with the `basf2` reconstruction framework. This layered simulation approach ensures high precision in reproducing both physics signatures and background conditions under realistic operational settings at SuperKEKB.

2.5 Event Reconstruction in Belle II

Event reconstruction in the Belle II experiment involves the transformation of raw detector data into high-level physics objects, such as particle tracks, vertices, and decay candidates. This process is essential for

interpreting collision events and extracting meaningful physics results. The reconstruction is implemented within the basf2, a modular system designed for scalability, flexibility, and integration with calibration and simulation workflows.
The process begins with converting raw digitized signals from sub-detectors into reconstructed hits. For example, signals from the PXD, SVD, and CDC are converted into space points representing the passage of charged particles. Similarly, energy deposits in the ECL and hits in the KLM system are digitized and recorded as preliminary inputs for clustering and PID.

2.5.1 Tracking

Charged-particle tracking in Belle II is a multi-stage process designed to robustly reconstruct particle trajectories within the complex geometry and high-background environment of the SuperKEKB collider. The tracking system integrates measurements from the CDC and the VXD. An overview of the complete tracking chain is shown in Figure 2.8 (left).

2.5.1.1 Track Finding and Fitting

The reconstruction begins in the CDC, where two complementary algorithms—the global and local finders—identify track candidates. The global finder [83] employs a Legendre transformation in conformal space, mapping curved trajectories into straight lines to simplify pattern recognition and detect tracks with small curvature. In contrast, the local finder uses a cellular automaton to link segments, enabling the recovery of displaced or non-prompt tracks. Candidate tracks from both algorithms are merged and initially fitted using a Deterministic Annealing Filter (DAF), producing preliminary estimates of the track parameters.
From the earliest stages of reconstruction, each track candidate is represented using the perigee parametrization, a five-parameter helical model defined at the point of closest approach (POCA) to the beamline. As illustrated in Figure 2.8 (right), the parameters include the signed transverse impact parameter d_0, the azimuthal angle ϕ_0 of the momentum at POCA, the signed curvature ω (inverse transverse momentum), the longitudinal impact parameter z_0, and the dip angle $tan\lambda$. This representation not only provides a compact and analytical model of the track geometry but is also central to all fitting and extrapolation procedures throughout the tracking chain.
After CDC track seeds are formed, they are propagated inward into the VXD using a Combinatorial Kalman Filter (CKF), which predicts hit positions using the perigee parameters and updates them iteratively with silicon measurements. A multivariate classifier is applied to suppress accidental hit combinations arising from the background. Simultaneously, a standalone VXD tracking algorithm reconstructs low-momentum or tightly curved tracks that do not reach the CDC. Finally, all track candidates are merged and refitted under various mass hypotheses to account for multiple scattering and energy loss in the detector material. The resulting tracks, described entirely by their perigee parameters and associated covariances, are ready for use in vertex fitting, particle identification, and high-level physics analysis.

2.5.1.2 Vertex Reconstruction

Tracks are extrapolated to identify interaction and decay vertices, a critical step for reconstructing secondary vertices from B and D meson decays. Vertex fits minimize spatial distances between extrapolated track helices and help discriminate signal from background [83, 84].
A typical example is the decay of a neutral B^0 meson into a J/ψ and a K^0_S, where the J/ψ subsequently decays into a pair of oppositely charged leptons ($J/\psi \to \mu^+\mu^-, e^+e^-$), and the K^0_S decays into two charged pions ($K^0_S \to \pi^+\pi^-$). Each charged track is described by perigee parameters and associated covariance matrices. The reconstruction proceeds in two stages:

- The J/ψ vertex is determined by fitting the two muon tracks to a common origin, with an additional mass constraint applied to improve precision.
- The K^0_S vertex is obtained by fitting the two pion tracks, typically displaced from the interaction point by several centimeters due to the K^0_S lifetime.

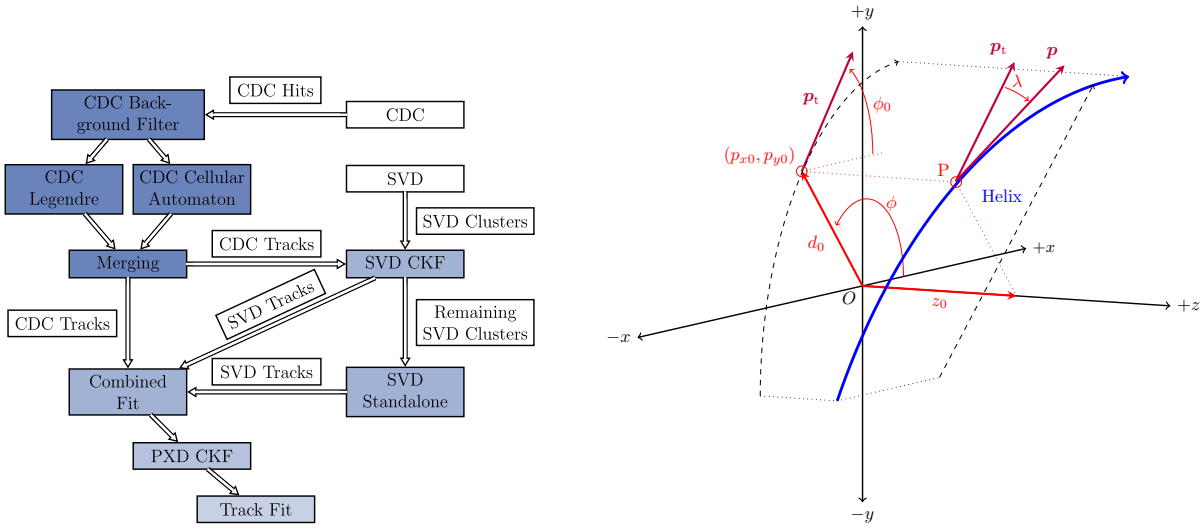

Figure 2.8: Left: Overview of the steps performed for track reconstruction at Belle II [83]. Right: The perigee parametrisation of the track helix [84].

To accurately reconstruct the full decay chain, Belle II employs a global decay chain vertex fitting algorithm, which simultaneously fits all vertices and intermediate resonances under both kinematic and geometric constraints. The fit incorporates the known invariant masses of intermediate states (for instance, J/ψ, K_S^0), the spatial positions of calorimeter clusters, and the beam interaction point profile. The J/ψ vertex is determined from the intersection of the lepton tracks, while the K_S^0 decay vertex—composed exclusively of neutral particles—is inferred using the reconstructed charged pions. This illustrates how Belle II combines prompt vertexing with displaced-vertex reconstruction, a capability central to time-dependent CPV measurements.

2.5.2 Particle Identification and Clustering

Charged PID in Belle II is a crucial step in event reconstruction. It utilizes information from various subdetectors—such as the TOP counter, ARICH, CDC, ECL, and the KLM system—to distinguish among particle hypotheses.

2.5.2.1 Charged Likelihood-Based Approach

The traditional PID method employs a likelihood function for each particle hypothesis h, conditioned on observables x from the detectors:

$$\mathcal{L}_h = P(\mathbf{x} \mid h) \tag{2.1}$$

The observables x may include:

- Cherenkov angle or hit time from TOP and ARICH
- Specific ionization dE/dx in the CDC
- Energy deposit E and ratio E/p in the ECL
- Penetration depth in the KLM

Instead of a binary likelihood ratio, Belle II uses a global likelihood-based probability defined as:

$$\mathcal{P}(h \mid \mathbf{x}) = \frac{\mathcal{L}_h}{\sum_j \mathcal{L}_j} \tag{2.2}$$

where the sum in the denominator runs over all considered hypotheses j. For example, the kaon identification score is:

$$\mathcal{P}(K \mid \mathbf{x}) = \frac{\mathcal{L}_K}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d} \tag{2.3}$$

where d denotes a deuteron. This formulation allows for probabilistic classification, enabling more flexible and accurate decision-making during analysis. Selection cuts on these probabilities can be optimized for different working points, balancing efficiency and misidentification rates. Typically, kaon (pion) identification efficiencies are above 90% with misidentification rates around 5–10%, depending on the momentum and detector region.

2.5.2.2 Neural Network-Based Approach

To account for nonlinear correlations among detector observables and achieve higher classification performance, Belle II employs multivariate classifiers, particularly neural networks (NNs). These are trained to approximate the posterior probability: $P(h \mid \mathbf{x}) = \mathrm{NN}_h(\mathbf{x})$ where $\mathbf{x}$ is the feature vector consisting of:

- Cherenkov angle and hit timing (TOP, ARICH)
- dE/dx from the CDC
- E/p ratio from the ECL
- Track kinematics and angles
- Penetration depth and hit pattern in the KLM

The neural networks are implemented using `TMVA`, `TensorFlow`, or `PyTorch` within the `basf2` framework. These classifiers offer improved performance, particularly in complex or high-background environments, and can be used standalone or in combination with likelihood ratios. While charged PID relies on sub-detector responses, neutral reconstruction begins with calorimetric clustering.
In calorimetric detectors such as the ECL, adjacent energy deposits are grouped into clusters, which are used to reconstruct the position and energy of neutral particles like photons and neutral pions. Clustering contributes to PID by providing shower-shape variables that help discriminate photons and electrons from hadrons.
Neutral particles are reconstructed by analyzing clusters in the ECL (for instance, photons) and hits in the KLM (for instance, K_L^0). Timing, spatial distributions, and energy deposition patterns are used to distinguish signal clusters from beam backgrounds and pile-up effects.
The combination of likelihood and neural network approaches enables Belle II to achieve high-fidelity PID, essential for flavor tagging, rare decay searches, and precision Standard Model tests. Together, these reconstruction procedures transform raw detector signals into high-level objects suitable for physics analysis, ensuring optimal efficiency and precision in the challenging environment of SuperKEKB.

Chapter 3

Data Analysis Strategy and Technique

In the study of rare decay processes, precise measurement of branching fractions requires careful control of statistical and systematic uncertainties, particularly when signal yields are expected to be small. To mitigate potential analysis bias, a blinding strategy is employed during the development phase of the analysis. This approach prevents researchers from unintentionally adjusting selection criteria or modeling choices in response to features in the data that may mimic a signal. Blinding typically involves concealing the signal region of a key observable while optimizing event selection and validating background models using control regions and simulated samples. By excluding the signal region from view until all analysis procedures are finalized, the blinding strategy ensures that decisions are made independently of any possible signal fluctuations, thereby preserving the objectivity of the result. This practice is especially important in rare decay searches, where the signal-to-background ratio is often low and small adjustments to the analysis can significantly affect the outcome.
After the selection strategy is established, including the use of multivariate techniques to suppress background, the signal region is unblinded. Statistical fitting methods are then applied to extract the signal yield. In the absence of a significant excess over the expected background, an upper limit on the branching fraction is determined. This overall strategy reflects a standard and robust framework employed in high-energy physics to ensure unbiased and statistically sound results in searches for rare processes.

3.1 Inclusive Tagging at Belle II

Belle II employs several tagging strategies to reconstruct one of the two $B-$mesons in each $\Upsilon(4S) \rightarrow B_{\mathrm{tag}}B_{\mathrm{sig}}$ event. The way the tag-side (B_{tag}) is reconstructed defines the tagging method, and three principal approaches are used:

- Hadronic Tagging [85]. The tagging side is fully reconstructed in a set of exclusive hadronic final states. The hadronic tagging method offers a very pure sample with excellent knowledge of B_{sig} kinematics but suffers from low efficiency ($\sim$0.1%-0.3%) since only a small fraction of all B decays can be fully reconstructed.

- Semileptonic Tagging [85]. The tagging side is partially reconstructed through semileptonic decays. It offers higher efficiency than hadronic tagging but loses some kinematic precision due to the missing neutrino.

- Inclusive Tagging [62]. Reconstructs only the signal-side decay explicitly. All remaining detected particles are assigned collectively to B_{tag}. This strategy benefits from high efficiency and minimal requirements on tag-side decay details, making it well-suited for rare or invisible-mode decays. The drawback, however, is reduced purity—i.e., a higher rate of misassigned tag-side components.

Figure 3.1 shows an illustration of three tagging methods with signal $B^{\pm} \rightarrow K^{\pm}\nu\bar{\nu}$.

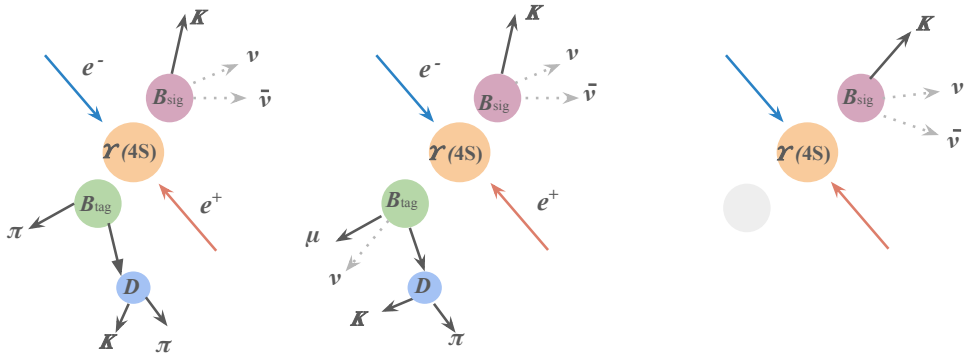

Figure 3.1: Illustration of the three tagging methods using $B^+ \to K^+\nu\bar{\nu}$ as the signal: (left) hadronic tagging, where the tagging B meson is fully reconstructed from hadronic tracks; (middle) semileptonic tagging, where the tagging B is reconstructed in semileptonic decays with one neutrino missing; and (right) inclusive tagging, where the remaining particles in the event are used to infer the tagging B

3.2 Classifiers

In modern high-energy physics analyses, the separation of signal from background processes often requires sophisticated multivariate methods due to the high dimensionality and complex correlations among input features. Multivariate classifiers are algorithms that learn to map a set of input variables, such as kinematic or topological observables, to a discriminant output, which represents the likelihood that an event belongs to a signal or background class. One of the most widely used multivariate methods in particle physics is the tree-based methods, particularly boosted decision trees (BDT), which have been the standard approach for signal–background classification tasks for over a decade [86]. They are favored for their robustness, speed, and interpretability, and are especially effective on tabular, low- to moderate-dimensional datasets, which are typical in collider analyses.

3.2.1 Decision Tree

Decision Trees [87] are supervised learning algorithms that partition the input space into regions associated with target predictions. Each internal node applies a binary decision rule based on a feature threshold, and each terminal node (leaf) assigns an output value, as illustrated in Fig. 3.2.
Given a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ with input features $\mathbf{x}_i \in \mathcal{R}^d$ and targets y_i, the tree recursively splits the data at each node using decision rules of the form:

$$x_j \leq \tau, \tag{3.1}$$

where x_j is the j-th feature and τ is the threshold chosen to optimize a splitting criterion.
The optimal split at each node is determined by maximizing the reduction in impurity. For classification, common impurity measures include:

- **Gini impurity:**

$$G(S) = 1 - \sum_{k=1}^{K} p_k^2, \quad p_k = \frac{1}{|S|} \sum_{i \in S} \infty \, (y_i = k),$$

  where S is the set of samples at a node and K is the number of classes.

- **Entropy (Information Gain):**

$$H(S) = -\sum_{k=1}^{K} p_k \log_2 p_k.$$

The impurity reduction (gain) from a split into subsets S_L and S_R is:

$$\Delta I = I(S) - \left(\frac{|S_L|}{|S|} I(S_L) + \frac{|S_R|}{|S|} I(S_R) \right),$$

where $I(\cdot)$ is either G or H.
Decision Trees are intuitive and interpretable, and can naturally handle both categorical and numerical data. However, they are prone to overfitting and high variance. Ensemble methods like BDT address these issues by aggregating multiple trees.

3.2.2 Gradient Boosted Decision Trees: Foundations and Advances in FastBDT and XGBoost

BDTs are ensemble models that combine multiple weak learners—typically shallow decision trees—to form a strong predictor. The key idea is to sequentially add trees that correct the errors of the existing ensemble. Given training data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, the model prediction after m trees is:

$$F_m(\mathbf{x}) = \sum_{t=1}^{m} f_t(\mathbf{x}), \tag{3.2}$$

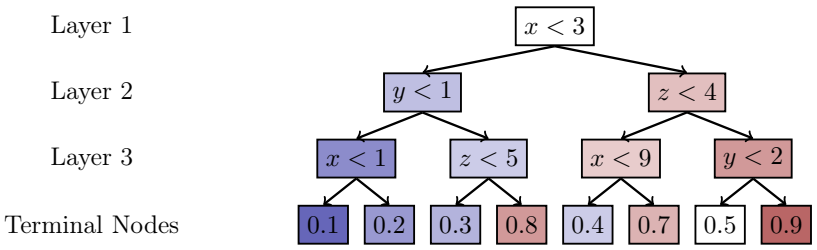

Figure 3.2: Example of a three-layer decision tree. At the root node (Layer 1), the sample is split according to the feature x<3. In Layer 2, further branching occurs using the feature y<1 (z<4). In Layer 3, the decision rules involve conditions on x and z (y). Each terminal node represents a region of feature space, where events are classified according to the distribution of training labels that fall into that node [88].

where each f_t is a regression tree. Trees are added to minimize a differentiable loss function $\mathcal{L}(y, F(\mathbf{x}))$ using gradient descent in function space.
At each iteration, a new tree f_m is trained to predict the negative gradient:

$$g_i = - \left. \frac{\partial \mathcal{L}(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right|_{F=F_{m-1}} . \tag{3.3}$$

Each tree is trained on pseudo-residuals (gradients), and a learning rate η controls update size:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \zeta f_m(\mathbf{x}). \tag{3.4}$$

Regularization techniques such as limiting tree depth, shrinkage, and sub-sampling are used to prevent overfitting. BDTs are highly flexible and perform well on structured data. They can approximate complex functions, handle feature interactions, and are robust to different data distributions. Their effectiveness makes them widely used in scientific applications, including high-energy physics.
Despite performance enhancements, FastBDT [88] adheres to the core gradient boosting framework. The differences lie primarily in data preprocessing and split evaluation. The optimization target remains the same: minimizing the overall loss via sequential, additive tree fitting. FastBDT introduces several innovations to accelerate training:

- Equal-frequency binning: Each feature is transformed into integer-valued bins based on quantiles. This discretization enables fast, cache-friendly histogram construction.
- Integer-based histogramming: Tree splits are computed using cumulative histograms over binned data rather than scanning raw float values, dramatically reducing computation time.
- Branchless training: Avoids conditional branching (if statements) inside hot loops to maintain CPU pipeline efficiency.
- Structure of Arrays (SoA) memory layout: Enhances cache locality and vectorization by storing features column-wise.

XGBoost (Extreme Gradient Boosting) is a general-purpose gradient boosting library that builds upon the standard framework by introducing second-order optimization, regularization, and distributed training capabilities [89].
XGBoost optimizes an objective function combining a differentiable loss and a regularization term to control model complexity. Formally, the objective at iteration t is:

$$\mathcal{J}^{(t)} = \sum_{i=1}^{N} \mathcal{J}\big(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\big) + \Omega(f_t), \tag{3.5}$$

where

- $\hat{y}_i^{(t-1)}$ is the prediction of the ensemble up to iteration $t-1$,
- $f_t \in \mathcal{F}$ is the newly added regression tree,
- $\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$ is the regularization term,
- T is the number of leaves in tree f,
- w_j is the weight of leaf j,
- γ and λ are regularization parameters.

To efficiently optimize this objective, XGBoost applies a second-order Taylor expansion of the loss around current predictions:

$$\mathcal{J}^{(t)} \approx \sum_{i=1}^{N} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t), \tag{3.6}$$

where g_i and h_i are the first and second derivatives of the loss; the algorithm derives optimal leaf weights:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \tag{3.7}$$

and the corresponding loss reduction to choose the best splits.
XGBoost incorporates second-order gradient optimization, flexible regularization terms (L1 and L2), and supports distributed computing, making it both powerful and robust for a wide range of problems.
The contribution of each input variable to the classification is quantified with the normalized feature importance . Let f_j denote the j-th feature, and consider an ensemble of T decision trees. The un-normalized importance of f_j is computed as the sum of the improvements in the splitting criterion (FastBDT) or the gain (XGBoost) over all nodes where f_j is used:

$$\tilde{I}(f_j) = \sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t(f_j)} \Delta C_n \quad \text{(FastBDT)}, \quad \tilde{I}(f_j) = \sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t(f_j)} \text{Gain}_n \quad \text{(XGBoost)}, \tag{3.8}$$

where $\mathcal{N}_t(f_j)$ is the set of nodes in tree t where feature f_j is used, and ΔC_n or Gain_n is the improvement at node n.
The normalized feature importance is then defined as

$$I(f_j) = \frac{\tilde{I}(f_j)}{\sum_k \tilde{I}(f_k)}, \tag{3.9}$$

so that the total importance of all input features sums to 1. This normalization facilitates comparison between features and clearly identifies the most influential variables in the classifier.

3.3 Maximum Likelihood Estimation

In particle physics, when analyzing experimental data, we often want to estimate parameters by comparing observed event counts in different bins (ranges of an observable) to predicted counts. A binned maximum likelihood fit is a statistical method to extract these parameters by maximizing the probability (likelihood) of observing the data given a model [90]. The signal strength μ is defined as

$$\mu = \frac{\mathcal{B}(B \to K^{(*)}\nu\bar{\nu})}{\mathcal{B}(B \to K^{(*)}\nu\bar{\nu})_{\text{SM}}}, \tag{3.10}$$

which compares the branching ratio of a rare decay observed in data to the SM prediction. The binned maximum likelihood fit is a powerful method to extract this parameter from experimental data. Suppose our data is divided into N bins indexed by $i = 1, \ldots, N$.

- n_i: observed number of events in bin i.
- $s_i(\boldsymbol{\theta})$: expected signal events in bin i according to the SM prediction, potentially dependent on nuisance parameters $\boldsymbol{\theta}$.
- $b_i(\boldsymbol{\theta})$: expected background events in bin i, also potentially dependent on nuisance parameters.
- μ: signal strength parameter scaling the expected signal.

In this case, μ rescales the signal part of the expected events, representing how the branching ratio of a rare decay compares to the SM prediction. The expected number of events in bin i is modeled as

$$\nu_i(\mu, \boldsymbol{\theta}) = \mu\, s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}). \tag{3.11}$$

Assuming each bin count n_i follows a Poisson distribution, the likelihood function for the entire dataset is

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{[\nu_i(\mu, \boldsymbol{\theta})]^{n_i} e^{-\nu_i(\mu,\boldsymbol{\theta})}}{n_i!}. \tag{3.12}$$

Systematic uncertainties are encoded by nuisance parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots\}$, which affect both signal and background predictions: $s_i(\boldsymbol{\theta}), \quad b_i(\boldsymbol{\theta})$. Prior knowledge about these uncertainties is introduced through constraint terms, often Gaussian:

$$\pi(\theta_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(\theta_j - \theta_{j,0})^2}{2\sigma_j^2}\right], \tag{3.13}$$

and $\pi(\boldsymbol{\theta}) = \prod_j \pi(\theta_j)$, where $\theta_{j,0}$ and σ_j are the nominal values and uncertainties of the nuisance parameters. The full likelihood becomes

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = L(\mu, \boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}). \tag{3.14}$$

To estimate μ, we maximize the likelihood over nuisance parameters for fixed μ:

$$\hat{\boldsymbol{\theta}}(\mu) = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mu, \boldsymbol{\theta}). \tag{3.15}$$

The profile likelihood is

$$\mathcal{L}_p(\mu) = \mathcal{L}(\mu, \hat{\boldsymbol{\theta}}(\mu)). \tag{3.16}$$

The maximum likelihood estimate $\hat{\mu}$ is the value of μ maximizing $\mathcal{L}_p(\mu)$.

3.4 Upper Limits

When searching for rare processes or new particles, the observed data may not show a statistically significant deviation from SM expectations. In such cases, the result is quoted as an upper limit on the relevant physical quantity, e.g., a branching fraction $\mathcal{B}$, at a chosen confidence level (CL) [91]. This section outlines the general frequentist framework. The statistical framework uses the same likelihood construction as in the signal extraction fits:

$$L(\mu, \theta) = \prod_i f(x_i \,|\, \mu\, s_i(\theta) + b_i(\theta)) \times \prod_j \pi_j(\theta_j). \tag{3.17}$$

where $s_i(\theta)$, $b_i(\theta)$ and $\pi_j(\theta_j)$ are the expected signal yields, expected background yields and constraint term in bin i, separately, θ are nuisance parameters for systematic uncertainties. The test statistic is the profile likelihood ratio, which compares the likelihood for a fixed μ to the globally best-fit value. To set limits, the $\mathrm{CL_s}$ method is used. It is defined as:

$$\mathrm{CL_s}(\mu) = \frac{\mathrm{CL_{s+b}}(\mu)}{\mathrm{CL_b}} \tag{3.18}$$

where $\mathrm{CL_{s+b}}$ and $\mathrm{CL_b}$ denote the probabilities, under the signal-plus-background and background-only hypotheses, of obtaining a test statistic at least as extreme as observed. The 95% confidence level upper limit μ_{up} is the smallest value of μ for which $\mathrm{CL_s}(\mu_{\mathrm{up}})$ =0.05. Finally, this limit is expressed as a branching fraction: $\mathcal{B}_{\mathrm{up}} = \mu_{\mathrm{up}} \times \mathcal{B}_{\mathrm{SM}}$.

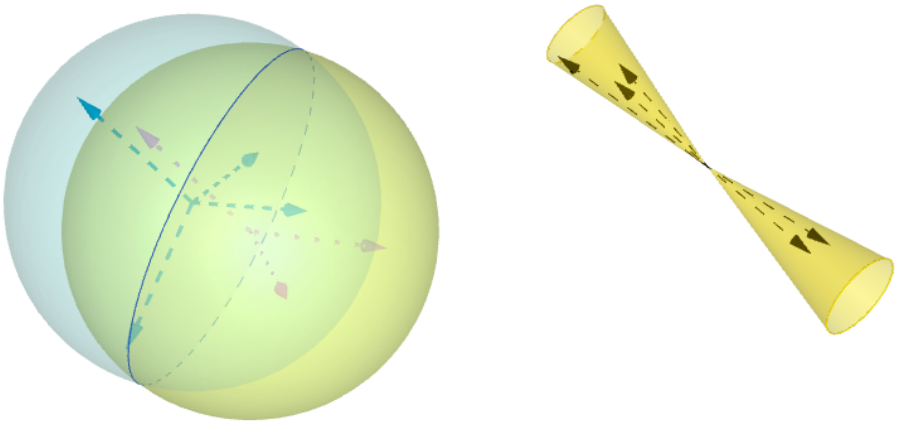

Figure 3.3: Comparison of event topologies in e^+e^- collisions: (left) isotropic distribution characteristic of $B\bar{B}$ decays; (right) back-to-back jet-like structure typical of continuum $q\bar{q}$ events. The dashed arrows represent particle trajectories.

3.5 Topological Variables in the Analysis

In high-energy e^+e^- collisions at B factories, understanding the global structure of particle final states is essential for distinguishing signal processes, such as $B\bar{B}$ decays, from dominant background sources like continuum $q\bar{q}$ production. Event shape variables quantify the geometrical and dynamical topology of events based on the momenta of final-state particles. These variables capture differences in spatial distributions, jet-like behavior, and isotropy, offering valuable tools for signal-background separation. Such observables are widely used in event classification and selection, forming a foundational component of precision B physics measurements. Fig 3.3 illustrate event shape of a $B\bar{B}$ event and a $q\bar{q}$ event.

3.5.1 Sphericity and Thrust

Sphericity and thrust are classical event-shape variables sensitive to isotropy and jet-like topology. Sphericity S is defined in terms of the momentum tensor $S^{\alpha\beta}$, constructed from the momenta of the final-state particles as:

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2} \tag{3.19}$$

where $\alpha, \beta \in \{x, y, z\}$, and $\vec{p}_i$ are the three-momenta of the particles in the event. The sphericity S and aplanarity are given in terms of the eigenvalues λ_1, λ_2, λ_3 of the sphericity tensor [92]:

$$\begin{aligned} S &= \frac{3}{2}(\lambda_2 + \lambda_3) \\ A &= \frac{3}{2}(\lambda_3) \end{aligned} \tag{3.20}$$

with the constraint:

$$\lambda_1 + \lambda_2 + \lambda_3 = 1, \quad \text{and} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \tag{3.21}$$

λ_1 corresponds to the direction along which the majority of the momentum is concentrated, while λ_2 is orthogonal to λ_1, captures the spread of momentum within the event plane. λ_3 quantifies the component of momentum perpendicular to this plane, representing the out-of-plane momentum. $S \in [0, 1]$, where $S = 0$ for perfectly jet-like events and $S = 1$ for sotropic events. $B\bar{B}$ events tend to have higher sphericity and

aplanarity, due to the more random distribution of decay products from near-rest B mesons, while $q\bar{q}$ events tend to have lower sphericity and near-zero aplanarity, reflecting their jet-like topology.
The thrust T is an event shape variable that quantifies how collimated the momenta of final-state particles are along a single axis. The thrust is defined as:

$$T = \max_{\vec{n}} \left(\frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|} \right) \tag{3.22}$$

where $\vec{p}_i$ is the momentum vector of the i-th final-state particle, $\vec{n}$ is a unit vector (the thrust axis) along which the scalar sum of projected momenta is maximized. Thrust emphasizes the dominant direction of momentum flow, making it complementary to sphericity, which captures the overall angular isotropy of the event. $T \in [0.5, 1]$, where $T = 1$ indicates perfectly collimated two-jet events. Particles are collimated along one or two directions in jet-like $e^+e^- \to q\bar{q}$ events, whereas they are emitted isotropically in all directions in more spherical $B\bar{B}$ events.

3.5.2 Harmonic Moments and Fox-Wolfram Moments

Harmonic moments probe single-particle angular distributions relative to a fixed axis, while Fox-Wolfram moments extend this to two-particle correlations, providing rotationally invariant measures of event shape. Harmonic moments are global event-shape observables that characterize the angular distribution of final-state particles with respect to a chosen reference axis, typically the thrust axis or the beam (collision) axis. In Belle II, they are denoted as B_l, and defined as:

$$B_l = \sum_i \frac{|\vec{p}_i|}{\sqrt{s}} P_l(\cos\,\theta_i) \tag{3.23}$$

where P_l is the Legendre polynomial of order l, $\vec{p}_i$ is the momentum of the i-th particle, $\sqrt{s}$ is the center-of-mass energy, and θ_i is the angle between $\vec{p}_i$ and the chosen axis. harmonic moments are not rotationally invariant, making them sensitive to the event geometry relative to the selected axis.
While harmonic moments probe the angular distribution of individual particles with respect to a fixed axis, a more comprehensive approach involves examining correlations between all particle pairs, leading to the definition of the Fox-Wolfram moments. The l-th Fox–Wolfram moment H_l is defined as [93, 94]:

$$\begin{aligned} H_l &= \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{s} P_l(\cos\theta_{ij}) \\ P_l(\cos\ \theta_{ij}) &= \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{lm}^*(\hat{p}_i) Y_{lm}(\hat{p}_j) \end{aligned} \tag{3.24}$$

where $\cos\,\theta_{ij} = \hat{p}_i \cdot \hat{p}_j$ is the angle between the momentum directions $\vec{p_i}$ and $\vec{p_j}$. Low l values (like l=1, 2) are sensitive to broad patterns, like whether most particles are back-to-back (for instance, $e^+e^- \to q\bar{q}$), while higher l values capture fine angular details, like subtle clustering. Fox–Wolfram moments are fully rotationally invariant because they involve pairwise correlations projected over the entire solid angle. Normalizing the H_l by H_0, $R_l = \frac{H_l}{H_0}$, ensures that R_l is dimensionless and typically constrained within a finite range, making it easier to compare across events with different energies or multiplicities.
The Kakuno-Super Fox-Wolfram (KSFW) moments are a generalized form of the traditional Fox-Wolfram moments by incorporating particle types, charge information, and event topology, offering enhanced discrimination power in B-factory analyses [95]. The KSFW moments introduce more refined correlations by:

- Separating particles into signal-side and rest-of-event (ROE) categories;
- Treating ROE particles by type (charged, neutral, missing).

Take an example of inclusive tagging $B \to K^* \nu\bar{\nu}$, the signal-ROE(so) KSFW moments of degree $l \in \mathbb{N}$ and $\xi \in$ charged(c), neutral(n), missing(m) is defined as

$$H^{so}_{\xi,l} = \frac{1}{Z} \sum_{i=K,\pi} \sum_{j_\xi=1}^{N_\xi} C^l_{ij\xi} p_{j_\xi} P_l(\cos \alpha_{\mathrm{ij}_\xi}) \tag{3.25}$$

where i denotes signal-side paritcles(K and π), and j_ξ refers to ROE particles of type ξ (c, n, or m). The angle α_{i,j_ξ} is the angle between $\vec{p}_i$ and $\vec{p}_{j_\xi}$. The weighting factor $C_{ij_\xi} \in$-1, 0, 1 is given by the product of the charges of candidate i and candidate j_ξ if l is odd and C_{ij_ξ}=1 if l is even (The missing momentum is treated as a neutral charge). The normalization factor Z ensures consistent scaling across events. The ROE−ROE(oo) moment of order l is defined as

$$R^{oo}_l = \frac{1}{Z} \sum_{j_\xi=1}^{N_\xi} \sum_{j'_\xi=1}^{N_\xi} C^l_{j\xi j'_\xi} p_{j_\xi} p_{j'_\xi} P_l(\cos \alpha_{\mathrm{j}_\xi \mathrm{j}'_\xi}) \tag{3.26}$$

R^{oo}_l captures the structure or collimates the ROE by evaluating angular correlations among ROE particles only.

Part II

Searches for the $B^+ \to K^+ \nu\bar{\nu}$ and $B^0 \to K^{*0} \nu\bar{\nu}$ decays using an inclusive tagging method

In this chapter, I present the two analyses in which I was directly involved, following their chronological order. The first concerns the study of the decay $B^+ \to K^+\nu\bar{\nu}$, on which I worked for the summer of 2022 until its publication in November 2023. Following that, I contributed to the inclusive tagging analysis of $B^0 \to K^*(892)^0\nu\bar{\nu}$, which I begin in the early Summer 2024 and continue to pursue. For convenience, in the remainder of this chapter, I will use the notation K^{*0} to denote $K^*(892)^0$.

Chapter 4

Search for $B^+ \to K^+\nu\bar{\nu}$

This chapter reports searches for the rare decays $B^+ \to K^+\nu\bar{\nu}$ [1] using the inclusive-tagging technique in Belle II experiment. The $B^0 \to K^{*0}\nu\bar{\nu}$ is treated as the background.
It begins by describing the data sample and corresponding simulated Monte Carlo datasets, followed by the application of data–simulation corrections to improve agreement between simulation and experiment. The main background suppression is achieved using two multivariate classifiers, and the signal region is defined based on the classifier output. The signal selection efficiency is validated with dedicated control samples. After BDT selection, the remaining background is studied in detail, separating continuum ($e^+e^- \to q\bar{q}$) and B-decay contributions, and the background estimation is validated against sideband data. Consistency checks are performed before opening the signal region. Systematic uncertainties affecting both signal efficiency and background estimation are then evaluated. Finally, the chapter presents the unblinded results, including the measured branching fraction (or upper limit) and the associated statistical significance.

4.1 Samples and $B^+ \to K^+\nu\bar{\nu}$ Reconstruction

4.1.1 Data sample

We use e^+e^- collision data collected with the SuperKEKB collider between 2019 and 2022. The primary dataset contains $362\ \mathrm{fb}^{-1}$ recorded at a center-of-mass energy of $\sqrt{s} = 10.58\ \mathrm{GeV}$ [96], corresponding to the $\Upsilon(4S)$ resonance, and includes $(387 \pm 6) \times 10^6\ B\bar{B}$ pairs. To study backgrounds from continuum processes, including $e^+e^- \to \tau^+\tau^-$ and light-quark $e^+e^- \to q\bar{q}$ (u, d, s, c) production, we also analyze a $42\ \mathrm{fb}^{-1}$ off-resonance sample taken $60\ \mathrm{MeV}$ below the $\Upsilon(4S)$ mass.

4.1.2 Simulation Sample

This analysis makes use of the Belle II Monte Carlo framework described in Sec. 2.4.3, with full detector simulation and reconstruction. In this section we summarize the specific simulated samples employed, together with their effective luminosities.

Signal samples. The $B^+ \to K^+\nu\bar{\nu}$ signal is generated using $1 \times 10^6\ B\bar{B}$ pairs, where at least one B meson decays to the $K\nu\bar{\nu}$ final state according to a phase-space model, which differs from the SM expectation. To correct for this, the events are reweighted to match the SM form factor calculations [37]. A consistent weighting scheme is applied to $B \to K^*\nu\bar{\nu}$ background decays. The long-distance contributions from $B^+ \to \tau^+\nu$ with $\tau^+ \to K^+\bar{\nu}$ is modeled separately, normalized to its branching fraction, and incorporated into the B^+B^- background sample. Fig. 4.1 shows the SM expectation of $B^+ \to K^+\nu\bar{\nu}$ for the squared di-neutrino mass distribution obtained in the simulated signal sample.

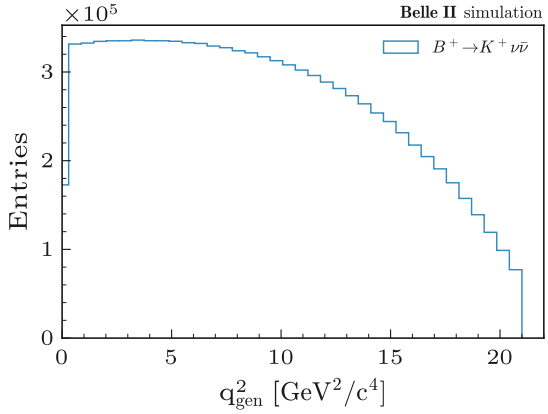

Figure 4.1: Distribution of the q^2 obtained in the simulated $B^+ \to K^+\nu\bar{\nu}$ sample generated according to the SM prediction in [37].

Generic samples. The run-independent production corresponds to an effective luminosity of 2.8 ab^{-1} of integrated luminosity for $B\bar{B}$ events at the $\Upsilon(4S)$ resonance, 1 ab^{-1} for continuum $q\bar{q}$ events ($q = u, d, s, c$), and 1 ab^{-1} for $\tau^+\tau^-$ events. These samples serve multiple purposes, including the evaluation of signal-selection efficiencies, the characterization of background processes, and the modeling of fit components in the data analysis. Furthermore, an off-resonance simulation corresponding to 50 fb^{-1} is used as an auxiliary check.

Dedicated background samples. Certain background processes are insufficiently modeled in generic MC due to limited experimental knowledge. Since their kinematic properties can mimic signal events, dedicated samples are used to replace these components, with the corresponding generic contributions removed to avoid double-counting. The second row in Tab. 4.1 shows the dedicated samples that we updated based on the relevant studies, which will be discussed in 4.2.

dedicated sample	model
$B \to K^*\nu\bar{\nu}$	SM formfactor [37]
$B^+ \to K^+p\bar{p}$	To model isospin counterpart of $B^+ \to K^+n\bar{n}$
$B^+ \to K^+K_L^0K_L^0$	Dalitz $B^+ \to K^+K_S^0K_S^0$
$B^+ \to K^+K_S^0K_L^0$	$B^+ \to K^+\phi(\to K_S^0K_L^0)$ and $B^0 \to K_S^0K^+K^-$

Table 4.1: Dedicated simulation samples used to model specific background processes not well described by generic simulation

4.1.3 $B^+ \to K^+\nu\bar{\nu}$ Reconstruction

This analysis employs an inclusive tagging method 3.1. In this approach, a charged kaon is reconstructed and the remaining tracks and energy deposits are classified as the rest-of-event(ROE). The online event selection (trigger) systems used in this analysis rely on either the number of charged-particle trajectories reconstructed in the CDC or the total energy deposits measured in the ECL, achieving an efficiency close to 100% for signal decays.

4.1.3.1 Charged Tracks and Neutral Clusters Selection

The reconstruction begins with the selection of charged tracks and neutral clusters, which are then identified and combined to reconstruct signal candidates. To suppress nonphysical candidates arising from detector mis-reconstruction or cosmic-ray muons, each charged particle and photon is required to have an energy less than 5.5 GeV. Both charged tracks and photon candidates are required to lie within the polar-angle acceptance of the CDC, $17^\circ \leq \theta_{\mathrm{CDC}} \leq 150^\circ$, ensuring they fall within regions of reliable detector coverage and high reconstruction fidelity.
Charged tracks at Belle II predominantly originate from pions, kaons, electrons, muons, and protons. For each track, likelihoods for the different particle hypotheses are calculated by combining particle identification (PID) information from all relevant Belle II sub-detectors. The most probable hypothesis is then assigned, using prior probabilities obtained from simulated $e^+e^- \to \Upsilon(4S), \Upsilon(4S) \to B\bar{B}$ events. Charged tracks are required to satisfy a transverse momentum constraint of $p_T > 0.1\ \mathrm{GeV}/c$, ensuring that only well-measured tracks are selected and suppressing low-momentum background from beam interactions or soft particles.
To differentiate tracks originating from the primary vertex (prompt tracks) from those stemming from displaced decays such as K^0_S, the following criteria are applied:

- **Primary-vertex tracks:** Prompt tracks must satisfy impact-parameter requirements $|d_z| < 3.0\ \mathrm{cm}$ and $d_r < 0.5\ \mathrm{cm}$. Here, d_r denotes the transverse impact parameter, defined as the distance of closest approach of the reconstructed track to the average interaction point (IP) in the plane perpendicular to the beam axis, while d_z denotes the longitudinal impact parameter, defined as the separation between the track and the IP along the beam axis. The requirements $d_r < 0.5\ \mathrm{cm}$ and $|d_z| < 3.0\ \mathrm{cm}$ suppress charged particles originating from regions displaced from the IP, thereby preferentially selecting prompt tracks from B-meson decays.

- **Secondary tracks from K^0_S decays:** are identified using the following criteria: Secondary tracks consistent with $K^0_S \to \pi^+\pi^-$ are identified using a mass window of $0.495 < M_{\pi\pi} < 0.500$ GeV/c^2, a vertex fit p-value > 0.001, a minimum decay length corresponding to $t > 0.007$ ns, and directional alignment $\cos\theta_{\mathrm{flight}} > 0.98$. These criteria ensure that the two tracks form a displaced vertex consistent with the known properties K^0_S.

These selection criteria reliably tag K^0_S decay products, excluding them from prompt track reconstruction and ensuring that only genuine signal-candidate tracks are retained.
Photon (Neutral Cluster) Selection: Photon candidates are reconstructed from clusters in the ECL. The selection is designed to retain genuine photons while suppressing residual electronic noise and misassociated energy deposits:

- **Energy threshold:** $E > 0.1\ \mathrm{GeV}$, The minimum energy threshold suppresses clusters from beam-related background and low-energy deposits from charged hadrons that fail track matching.

- **Angular acceptance:** restricted to the fiducial volume of the ECL, corresponding to the polar-angle coverage of the CDC, ensuring reliable reconstruction.

- **Track-matching veto:** photon candidates must not be associated with any reconstructed charged track, thereby avoiding double counting and ensuring the neutral origin of the cluster.

4.1.3.2 Charged Kaon Track Selection

Signal Kaon candidates are selected from charged tracks that are not associated with reconstructed $K_S^0 \to \pi^+\pi^-$ decays. Tracks are required to have at least 20 associated CDC hits to ensure good tracking quality and momentum resolution. Additionally, tracks must have at least one hit in the PXD, providing precise vertex resolution and improved impact-parameter determination. Kaon candidates are then selected by applying a tight particle-identification requirement with a KaonID likelihood ratio greater than 0.9 based on CDC, ECL, TOP, and ARICH information. These criteria collectively ensure that the selected signal Kaon candidates are well-reconstructed, clean, and highly reliable for use in subsequent B or D meson reconstruction.

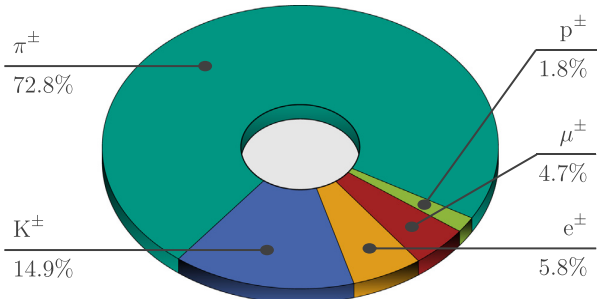

Figure 4.2: Composition of reconstructed charged particles at Belle II, adapted from [83]

4.1.3.3 Signal Event Reconstruction

After selecting high-quality signal tracks, events are further required to contain between four and ten reconstructed charged tracks. This suppresses backgrounds from both very low-multiplicity processes and high-multiplicity continuum processes, while maintaining more than 90% signal efficiencies. Low-multiplicity backgrounds, such as those from two-photon collisions, are further suppressed by requirements on the event's total visible energy and the polar angle of the missing momentum. The total energy from all reconstructed particles in the event is required to be at least 4 GeV, ensuring that the event is sufficiently energetic and well-measured. The polar angle of the missing momentum, calculated in the center-of-mass frame from the total reconstructed momentum, is restricted to $(17°, 160°)$, ensuring that the missing momentum points into the instrumented detector volume.
To account for the presence of the two undetected neutrinos, the momentum transfer squared to the neutrino pair (q^2) is computed as follows:

$$\begin{aligned} q^2 &= (p_B^* - p_K^*)^2 \\ &= (E_B^* - E_K^*)^2/c^2 - |\vec{p}_B^* - \vec{p}_K^*|^2 \\ &= M_B^2c^4 + M_K^2c^4 - 2\left(E_B^*E_K^* - c^2\vec{p}_B^* \cdot \vec{p}_K^*\right) . \end{aligned} \tag{4.1}$$

where $p_B^* \equiv (E_B^*, \vec{p}_B^*)$ and $p_K^* \equiv (E_K^*, \vec{p}_K^*)$ is the four-momentum of B meson and the kaon in the center-of-mass (c.m.) frame. The on-shell conditions $E_B^{*2} - |\vec{p}_B^* c|^2 = M_B^2c^4$ and $E_K^{*2} - |\vec{p}_K^* c|^2 = M_K^2c^4$ are applied.
Since the four-momentum of the B meson is not directly measurable in the experiment, the momentum transfer to the neutrino pair must be reconstructed using the observable kaon kinematics. In the e^+e^- c.m. frame, the B meson is nearly at rest, the four-momentum is approximated as $(\sqrt{s}/2, 0)$ and $\vec{p}_B^* \cdot \vec{p}_K^* 0 \approx 0$. The reconstructed momentum transfer squared to the neutrino pair q^2_{rec} is calculated as

$$q^2_{\text{rec}} = \frac{s}{4c^4} + M_K^2 - \frac{\sqrt{s}E_K^*}{c^4} \tag{4.2}$$

When multiple signal-kaon candidates are present in an event, the candidate with the lowest reconstructed q^2_{rec} is chosen, since overestimates in q^2_{rec} come from mis-reconstructed candidates, so the lowest is more

physical. After the best signal-kaon candidate is selected, all remaining reconstructed particles in the event are assigned to the ROE. The ROE is formed by fitting the remaining charged tracks to a common vertex and combining them with all reconstructed photons and K_S^0 candidates not used in the signal candidate. For signal events, the ROE corresponds to the decay products of the accompanying B meson. Table 4.2 summarizes the step-by-step selections.

charged tracks	$p_t >= 0.1$GeV/c, E<5.5 GeV, $17^\circ < \theta_{\rm CDC} < 150^\circ$ tracks $\in K_S^0$: $0.495 < M_{\pi\pi} < 0.500$ GeV/c^2, p-value > 0.001, proper decay time > 0.007 ns, $\cos\theta_{\rm flight} > 0.98$. tracks $\notin K_S^0$: $\lvert d_z\rvert < 3.0$ cm, $d_r < 0.5$ cm
neutral clusters	E>0.1 GeV, E<5.5 GeV, $17^\circ < \theta_{CDC} < 150^\circ$, not matched to tracks
K^+	$\notin K_S^0$, KaonID>0.9, nCDCHits>=20, nPXDHits>=1
signal event	Ntracks$\in$[4, 10], visible energy larger than 4 GeV, $\theta_{\rm missing} \in (17^\circ, 160^\circ)$
best signal candidate	lowest-$q^2_{\rm rec}$

Table 4.2: Summary of event reconstruction and selection criteria for $B^+ \to K^+\nu\bar{\nu}$ analysis.

4.1.4 Corrections to Simulated Sample

Detector simulation is validated against dedicated data control samples. Residual mismodeling is corrected with scale factors or event weights, and the associated systematic uncertainties are propagated to the signal efficiency and background estimates.
Charged particle reconstruction is studied using $e^+e^- \to \tau^+\tau^-$ events. Data and simulation agree well, with a residual 0.9% per-track systematic uncertainty is assigned. This efficiency, determined before any analysis-specific selection, accounts for detector acceptance and reconstruction performance.
Kaon identification efficiency and pion-to-kaon misidentification rates are measured using $D^{*+} \to \pi^+ D^0, D^0 \to K^-\pi^+$. These correction factors are applied per candidate as a function of momentum and polar angle, ensuring consistency between data and simulation.
Neutral particles in Belle II, primarily photons and neutral hadrons (K_L^0), are reconstructed in the ECL. Their detection and energy reconstruction are treated with dedicated studies:

- **Real photons**: The photon detection efficiency and energy calibration are determined using $e^+e^- \to \mu^+\mu^-\gamma$ events. Efficiency is defined as the fraction of expected photons that are reconstructed with the correct kinematics. The associated uncertainty on the photon energy is 0.5%, with the effect on signal yields evaluated by varying photon energies in simulated events.

- **Fake photons**: Neutral hadron deposits are studied using the control channel $B^+ \to K^+ J/\psi$. A scale factor $f_h \approx 1.05$ (with 0.9-1.1 at 68% CL) is applied, and the hadronic-energy correction is varied by $\pm$10% to estimate a conservative 100% relative systematics.

- K_L^0 **efficiency**: Detection of high-energy K_L^0(larger than 1.6 GeV) in the ECL is studied using radiative-return $e^+e^- \to \gamma\phi(\to K_S^0 K_L^0)$. The K_L^0 momentum is inferred from the reconstructed high-energy photons and K_s^0, and the trajectory of K_L^0 is extrapolated to the ECL to match clusters. Simulation is found to overestimate the efficiency by 17%, and a correction with an associated $\pm$8.5% systematic is applied. For lower-energy K_L^0, where cluster identification is less reliable, the effect is absorbed within the hadronic-energy systematic derived from unmatched photon studies.

Corrections for continuum mismodeling and signal-like $B\bar{B}$ backgrounds are derived from regions adjacent to the signal region and will be discussed in Sec. 4.4. Together, these corrections ensure that the simulated samples reproduce the detector performance observed in data. The associated systematic uncertainties are included in the likelihood fit and contribute to the total uncertainty on the branching fraction measurement.

4.1.5 Definition of Discriminating Variables

After event selection, we construct a set of kinematic and topological variables designed to discriminate between signal and background. These variables form the input feature set for the boosted decision trees (BDTs) used in the analysis.
For a given track, the POCA is defined as the point minimizing the distance to a line passing through the average IP and parallel to the solenoid symmetry axis (z-axis). The transverse impact parameter d_r is this minimal distance, while the longitudinal impact parameter d_z is the z-coordinate of the POCA relative to the IP.

4.1.5.1 Signal-Kaon-related Observables

- $d_r(K^+)$ Radial distance between the POCA of the K^+ candidate track and the IP. (Fig. 4.3 (left))
- $\cos(thrust_B, z)$ Cosine of the angle between the thrust axis of the signal B candidate and the z axis.(Fig. 4.3 (right))

Signal kaons originate near the IP, giving $d_r(K^+) \approx 0$, while background tracks from secondary decays yield broader distributions. Fig. 4.3 shows the distributions of $dr(K^+)$ and $\cos(\text{thrust}_{B,z})$ in data and simulation.

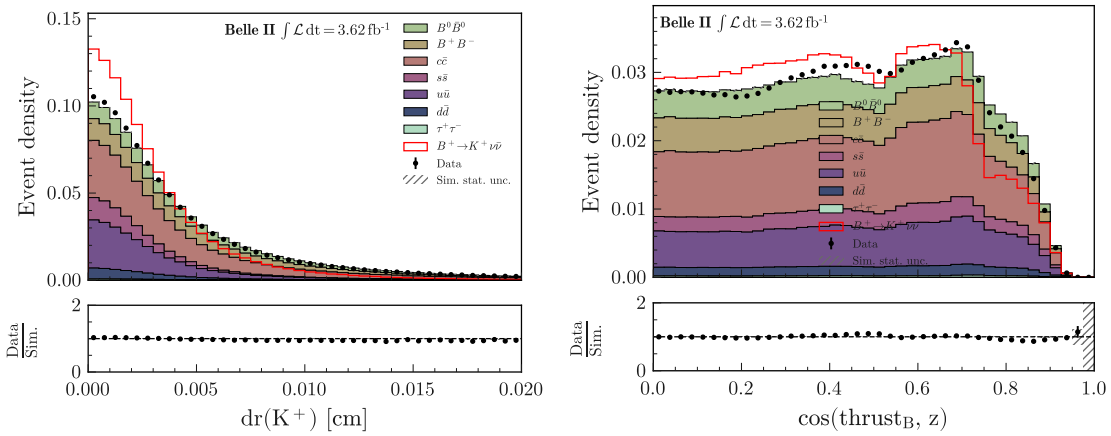

Figure 4.3: Distribution of signal-Kaon-related observables, $dr(K^+)$ and $\cos(\text{thrust}_{B,z})$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

4.1.5.2 ROE-related Observables

- ΔE_{ROE} Difference between the ROE energy in the c.m. and the energy of one beam of c.m. ($\sqrt{s}/2$)
- p_{ROE} Magnitude of the ROE momentum
- $\theta(p_{\text{ROE}})$ Polar angle of the ROE momentum
- Modified Fox-Wolfram R_0^{oo}, R_2^{oo} Moment of the oo type (i.e. ROE-ROE), calculated in the c.m. frame
- $\text{Var}(p_T^i | i \in \text{ROE})$ Variance of the transverse momentum of the ROE tracks
- $d_x(\text{TagVertex})$, $d_z(\text{TagVertex})$ Two variables corresponding to the x, z components of the vector from the average interaction point to the ROE vertex.

- p-value(ROE) p-value of the ROE vertex fit

The variables derived from the ROE provide strong discrimination between signal events and background events, especially continuum $q\bar{q}$ background. Kinematic properties of the ROE, such as the energy difference ΔE_{ROE}, the momentum magnitude p_{ROE}, the angle θ_p(ROE), and the variance of the transverse momentum of ROE tracks, are largely determined by the momentum distribution of ROE particles. In signal events, these quantities reflect the well-defined kinematics of the tag-side B decay, leading to characteristic peaks or narrow distributions. In contrast, for the continuum background, the ROE often collects particles from jet-like fragmentation, resulting in broader or shifted distributions. In signal events, ΔE_{ROE} peaks slightly below zero due to the missing neutrinos, whereas in continuum $q\bar{q}$ events, jet fragmentation leads to an excess of reconstructed energy, making it one of the most discriminative variables.
In addition to kinematics, the global event shape of the ROE provides further discrimination. Modified Fox–Wolfram moments of the oo type, R_0^{oo} and R_2^{oo}, capture the isotropy of the ROE in the center-of-mass frame. Signal $B\bar{B}$ events are nearly spherical, yielding smaller R_2^{oo} and more uniform R_0^{oo}, whereas continuum $q\bar{q}$ events are jet-like, producing larger and more anisotropic moments. Together, the ROE kinematics and event-shape variables form a complementary set of inputs that effectively separate signal from background.
Fig. 4.4 and 4.5 show the distributions of nine ROE-related variables in data and simulation.

4.1.5.3 Missing-energy-related Observables

- M^2_{missing} Square of the missing invariant mass
- $\theta(p_{\mathrm{missing}})$ Polar angle of the missing three-momentum in the c.m. frame

The missing mass squared, M^2_{missing}, and the polar angle of the missing momentum, $\theta_{\mathrm{missing}}$, provide complementary information about the unobserved neutrinos in the signal. For signal events, M^2_{missing} is broadly distributed, reflecting the two missing neutrinos and detector resolution effects. In contrast, continuum $q\bar{q}$ background peaks, as most particles are detected and only a small apparent missing momentum arises, with a long tail extending to higher values due to occasional mis-reconstruction. For the continuum $q\bar{q}$ background, the polar angle of the missing momentum, θ_{miss}, shows no preferred direction. This is because the missing momentum in such events typically arises from mis-measured or undetected particles, which are distributed randomly in space. Fig. 4.6 shows the distributions of M^2_{missing} and $\theta(p_{\mathrm{missing}})$ in data and simulation.

4.1.5.4 Entire-event-related Observables

- Fox-Wolfram R_1, R_2, R_3 First, second, and third normalized Fox-Wolfram moment in the c.m. frame
- Sphericity Event sphericity in the c.m. frame
- Harmonic Moment B_0, B_2 Zeroth-order and second-order harmonic moment with respect to the thrust axis in the c.m. frame
- Modified Fox-Wolfram $H^{\mathrm{so}}_{c,2}$ Second-order modified Fox-Wolfram moment related to charged particles in the ROE calculated in the c.m. frame
- Modified Fox-Wolfram $H^{\mathrm{so}}_{n,2}$ Second-order modified Fox-Wolfram moment related to neutral particles calculated in the c.m. frame
- Modified Fox-Wolfram $H^{\mathrm{so}}_{m,2}$, $H^{\mathrm{so}}_{m,4}$ Second-order and fourth-order modified Fox-Wolfram moment related to missing momentum calculated in the c.m. frame
- $d_r(K^+, \mathrm{Tag\,Vertex})$, $d_z(K^+, \mathrm{Tag\,Vertex})$ Radial and longitudinal distance between the POCA of the K^+ candidate track and the tag vertex

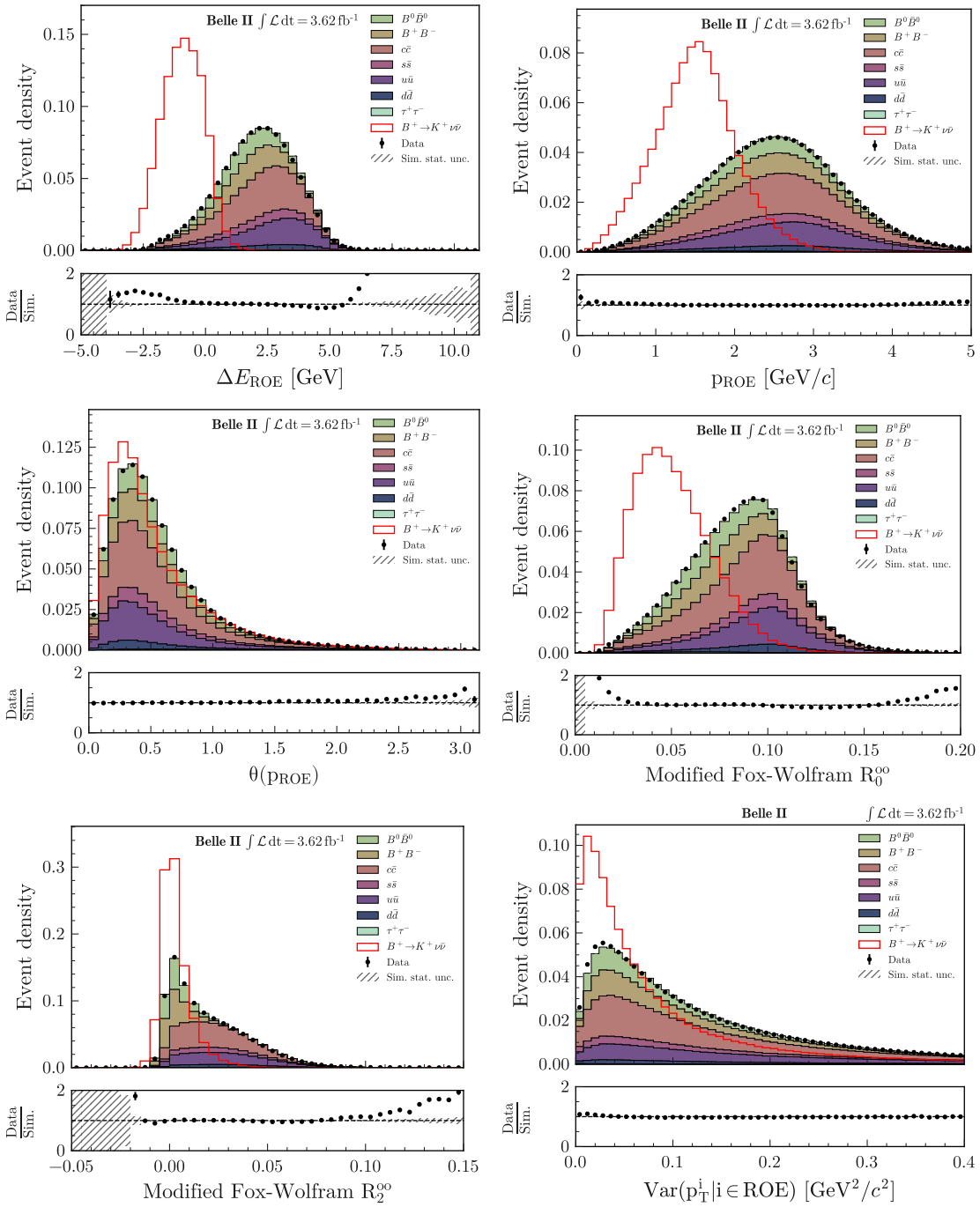

Figure 4.4: Distributions of ROE-related observables, ΔE_{ROE}, p_{ROE}, $\theta(p_{\mathrm{ROE}})$, Modified Fox-Wolfram R_0^{oo}, Modified Fox-Wolfram R_2^{oo} and $\mathrm{Var}(p_T^i | i \in \mathrm{ROE})$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

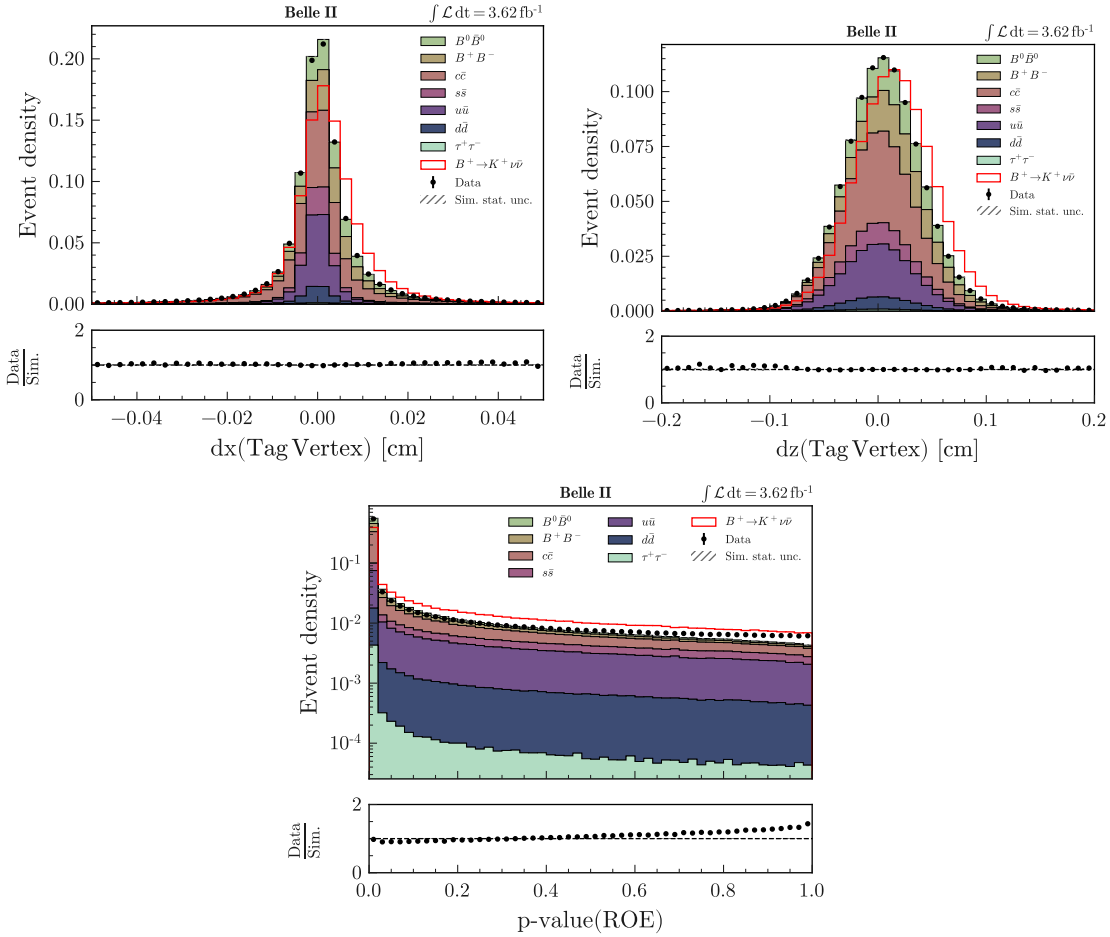

Figure 4.5: Distributions of $d_x(\text{TagVertex})$, $d_z(\text{TagVertex})$ and $p-\text{value(ROE)}$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

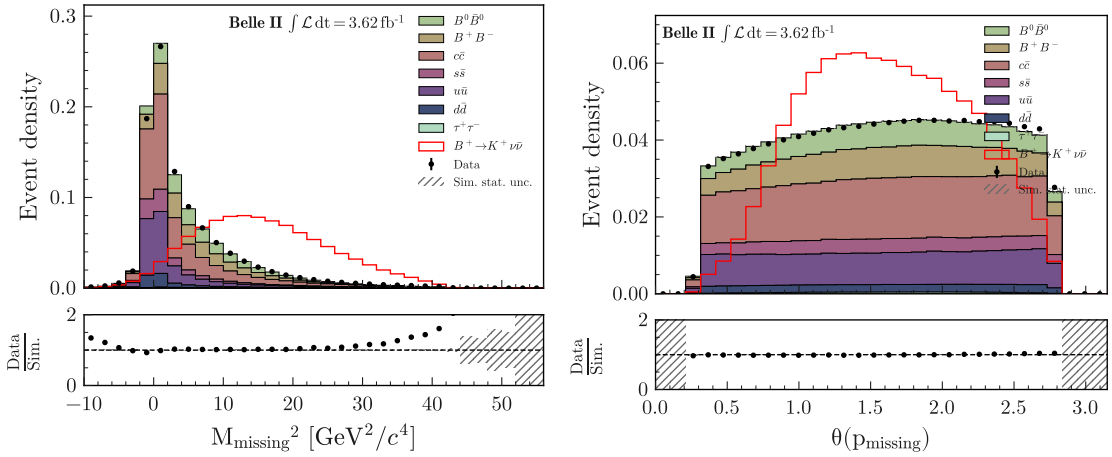

Figure 4.6: Distributions of missing-energy-related observables, M^2_{missing} and $\theta(p_{\text{missing}})$, in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

- N_{lepton} Number of $e^{\pm}$ and $\mu^{\pm}$ candidates
- N_{γ} Number of photon candidates
- N_{tracks} Number of charged particle candidates
- e^2 Square of the total charge of tracks in the event
- $\cos(\theta(\text{thrust}))$ Cosine of the polar angle of the thrust axis in the c.m. frame
- $\cos(\text{thrust}_K, \text{thrust}_{\text{ROE}})$ Cosine of the angle between the thrust axis of the signal B candidate and the thrust axis of the ROE computed in the c.m. frame

Event-shape-related variables are powerful in distinguishing the continuum, $B\bar{B}$ background events and signal $B^+ \to K^+\nu\bar{\nu}$ signal events, with continuum being jet-like, $B\bar{B}$ events more spherical, and the signal events exhibiting an intermediate topology due to a spherically distributed inclusive B decay combined with a single high-p_T track in the signal $\bar{B}$ side. The normalized Fox–Wolfram moments R_1, R_2, and R_3 are sensitive to the angular correlations of all particles in the event, capturing dipole, quadrupole, and octupole structures that reflect the transition from elongated continuum events to spherical $B\bar{B}$ events. Harmonic moments B_0 and B_2 characterize the angular distribution of particles relative to the event's thrust axis, with B_0 measuring the average alignment along the axis and B_2 capturing quadrupole deviations relative to the axis. For the Modified Fox-Wolfram $H^{\text{so}}_{m,2}$, $H^{\text{so}}_{m,4}$ signal events are broader since missing neutrinos carry unmeasured momentum, whereas background peaks narrowly around zero (no genuine missing momentum). Fig. 4.7 and 4.8 show distribution of Fox-Wolfram $R_{1,2,3}$, sphericity, Harmonic momentum $B_{0,2}$, Modified Fox-Wolfram and $d_{r,z}(K^+, \text{Tag Vertex})$ in data and simulation. Apart from event-shape variables, global event quantities such as the number of leptons, number of photons, total number of tracks, and the squared total charge are used to capture additional discriminative features between signal and background events.
Fig. 4.9 shows the distributions of number of $e^{\pm}$ and $\mu^{\pm}$ candidates, photon candidates, charged particle candidates, square of the total charge of tracks, sphericity, $\cos(\theta(\text{thrust}))$ and $\cos(\text{thrust}_K, \text{thrust}_{\text{ROE}})$ in the data and simulation.

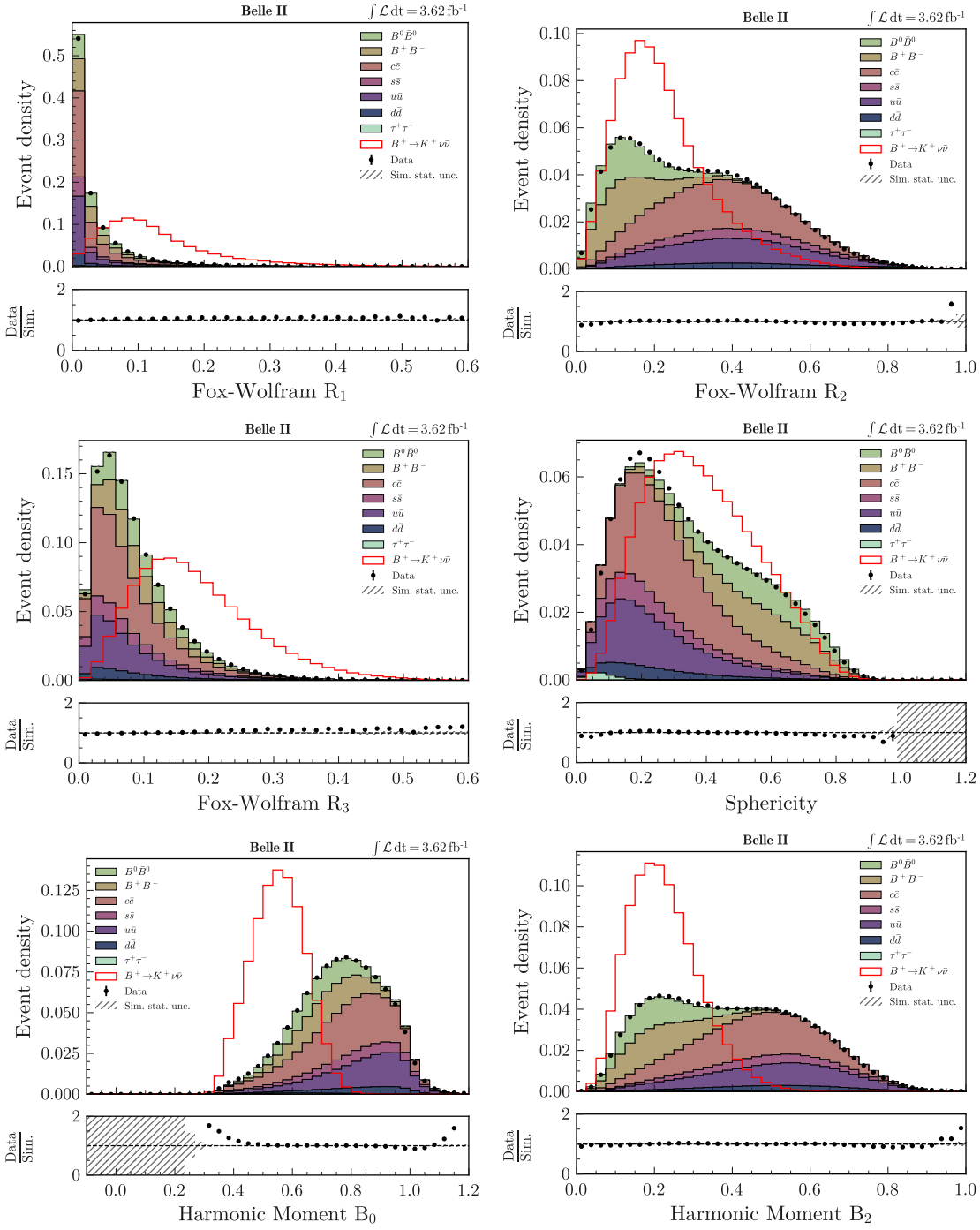

Figure 4.7: Distributions of entire-event-related observables, Fox-Wolfram R_1, R_2, R_3, sphericity and Harmonic Moment B_0, B_2 in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

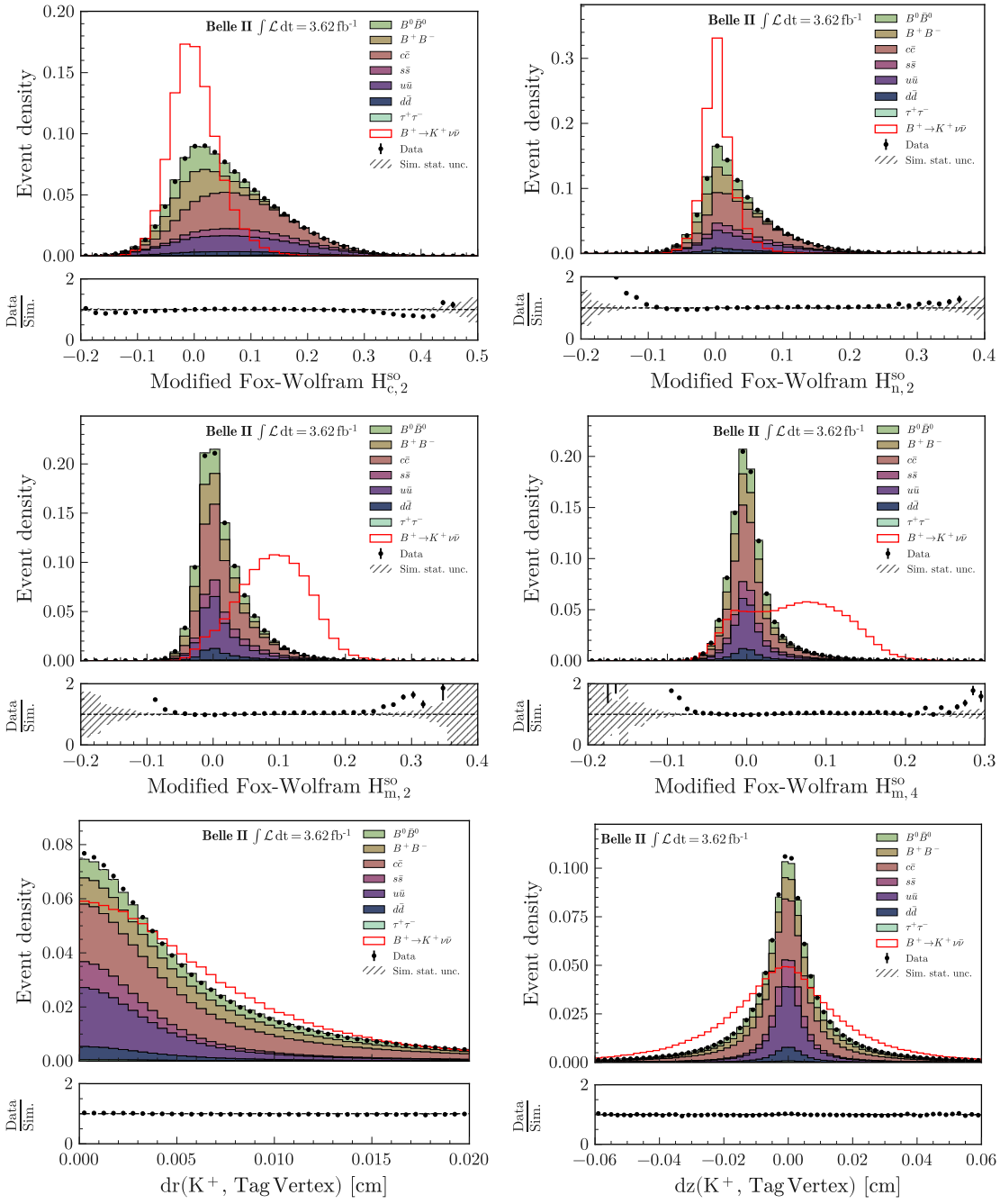

Figure 4.8: Distributions of entire-event-related observables Modified Fox-Wolfram $H^{\mathrm{so}}_{c,2}$, $H^{\mathrm{so}}_{n,2}$, $H^{\mathrm{so}}_{m,2}$, $H^{\mathrm{so}}_{m,4}$, $d_r(K^+, \text{Tag Vertex})$, and $d_z(K^+, \text{Tag Vertex})$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity ($\sim 3.62\ \mathrm{fb}^{-1}$) is used.

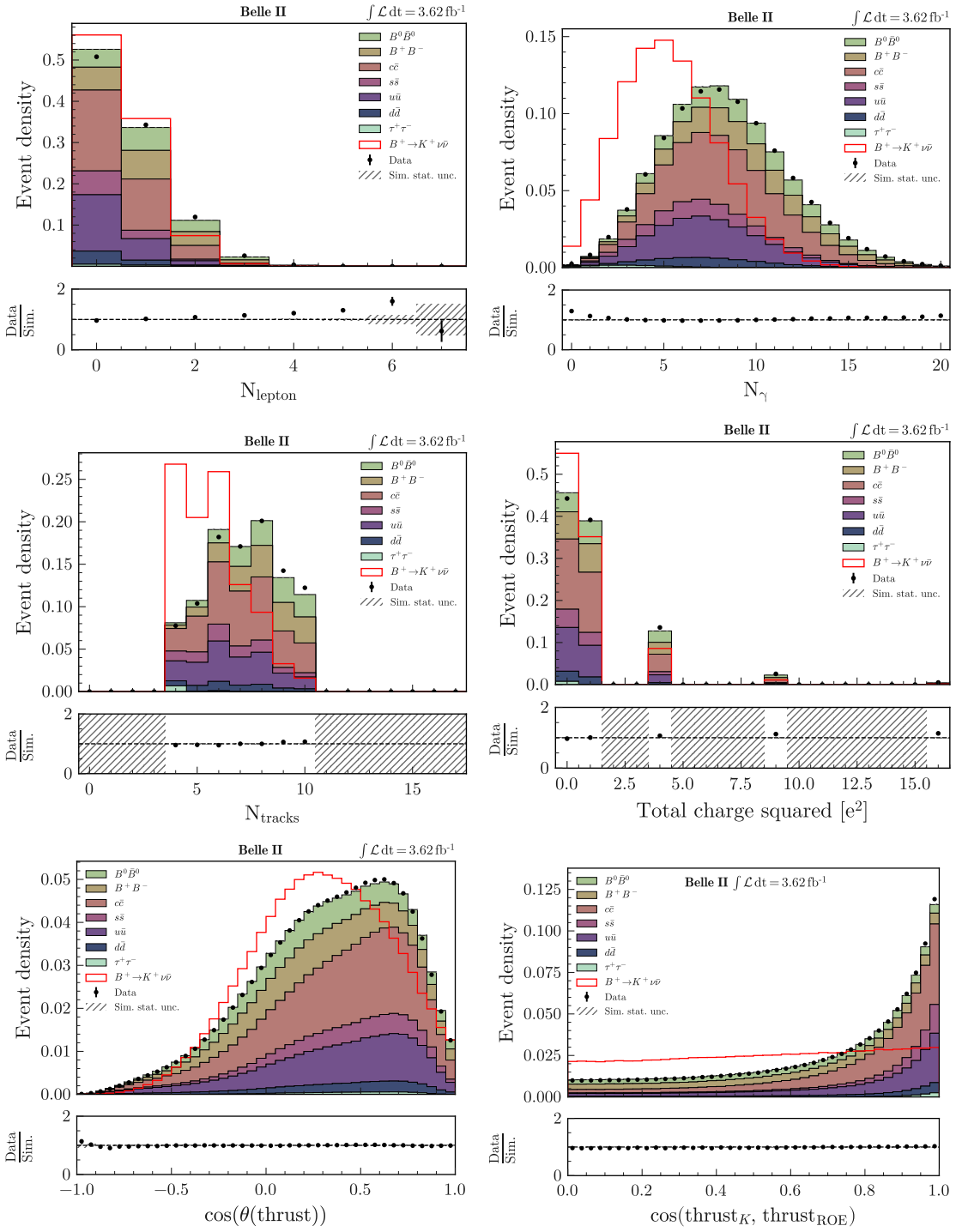

Figure 4.9: Distributions of entire-event-related observables, N_{lepton}, N_{γ}, N_{tracls}, e^2, $\cos(\theta(\text{thrust}))$ and $\cos(\text{thrust}_K, \text{thrust}_{ROE})$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

4.1.5.5 D-meson-suppression-related Observables

Charm decays, such as $D^0 \to K^-\pi^+$ and $D^+ \to K^-\pi^+\pi^+$ constitute an important source of background, since they mimic the signal topology. To suppress them, a dedicated D-meson veto is implemented. A kinematic fit is performed using the signal kaon and either one oppositely charged track (for D^0 candidates) or two additional tracks ($\pi^+\pi^-$ or $\pi^+\pi^+$ for D^+ candidates) to form a common vertex. Among possible combinations, the candidate with the best vertex fit quality is selected as the best D^0/D^+ candidate.

- $M(K^+X^-)$ Mass of the best D^0 candidate, where X^-. The charged tracks from the ROE
- $d_r(D^+)$ Radial distance between the best D^+ candidate vertex and the IP
- Median(p-value(D^0)) Median p-value of the vertex fits of the D^0 candidates
- p-value(D^0) p-value of the best D^0 candidate vertex fit
- p-value(D^+) p-value of the best D^+ candidate vertex fit

The very interesting features shown in $M(K^+X^-)$, clear and narrow peaks are observed around $K^*(892), \phi$ and $D^{0/+}$ of background events, while no particular peaks appear in signal events. Other discriminating variables include the p-value from the kinematic vertex fit of $D^{0/+}$ candidates, in signal events, the lack of true D-mesons yields poor vertex fits (low p-values), whereas background events often give higher p-values consistent with true vertices. Fig. 4.10 shows distributions of five D-veto related variables in data and simulation.

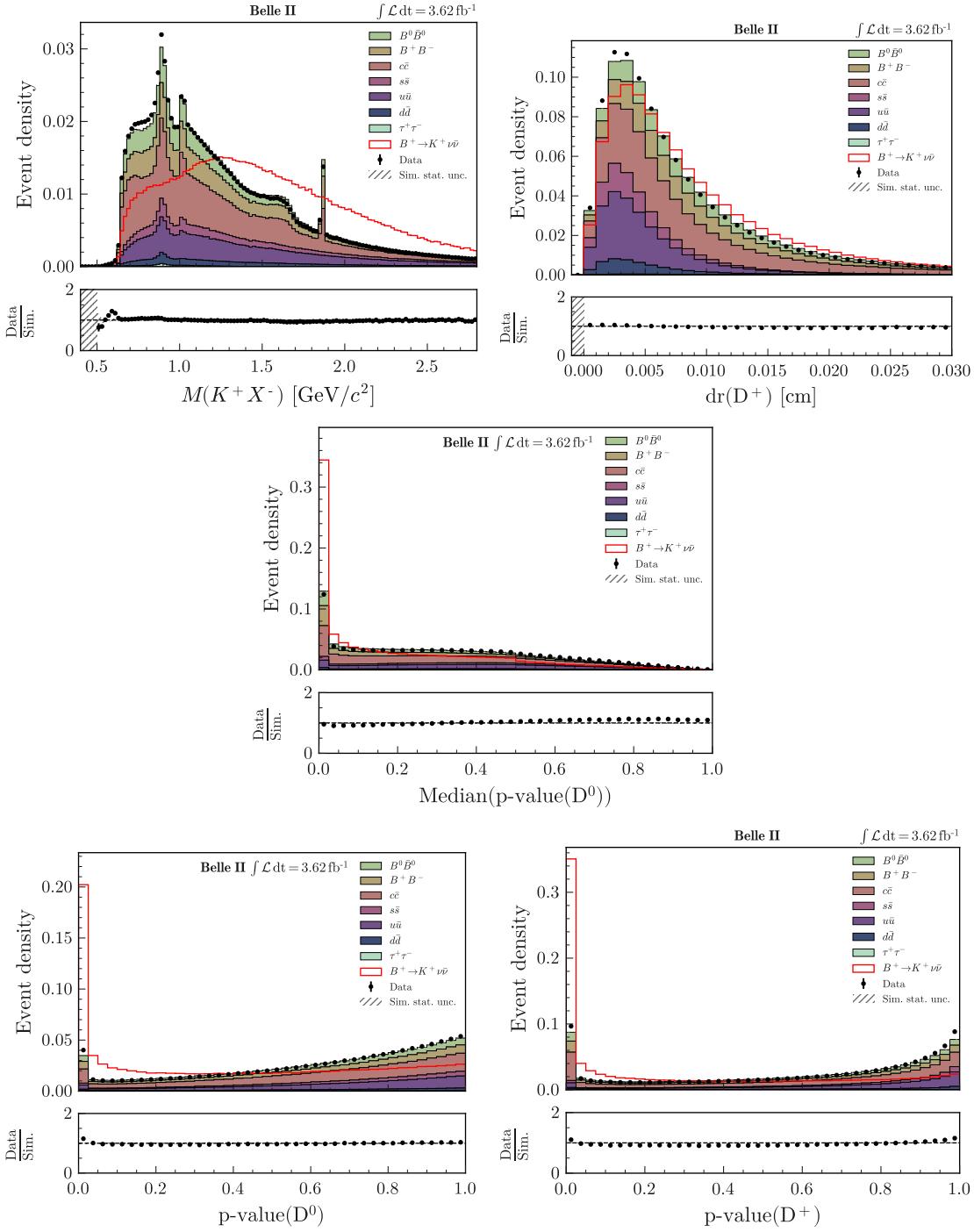

Figure 4.10: Distributions of D-suppression-related observables, $M(K^+X^-)$, $d_r(D^+)$, Median(p-value(D^0)), p-value(D^0) and p-value(D^+) in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram, and is not included in the stack. 1% of the total $\Upsilon(4S)$ data luminosity (~ 3.62 fb^{-1}) is used.

4.2 Background Suppression

4.2.1 The First Classifier

After applying the basic event selection criteria, a substantial amount of residual background remains. As shown in the variables distribution in Sec. 5.1.5, continuum $e^+e^- \to q\bar{q}$ events dominate after pre-selection and exhibit topologies distinct from signal. To further suppress this background, we employ a multivariate classifier based on the FastBDT algorithm. The input variables Tab. 4.3 are dominated by event-shape observables, which are highly effective in distinguishing the jet-like structure of continuum events from the more spherical topology of signal events. The classifier is trained using simulated samples of both signal and background, and its output is subsequently used to reject events with continuum-like characteristics, thereby enhancing the overall purity of the selected signal sample. For the training and validation of the FastBDT classifiers, simulated samples corresponding to the major background categories were employed. Specifically, 10^6 events were used for each background class: $e^+e^- \to B^+B^-, B^0\bar{B}^0, u\bar{u}, d\bar{d}, c\bar{c}, s\bar{s}$ and $e^+e^- \to \tau^+\tau^-$ production. In order to account for the relative production cross sections and to ensure that the training is representative of the expected composition in data, an event weight was applied to each category such that the effective yields are scaled to the same integrated luminosity. In addition, a sample of 10^6 simulated signal events with SM-q^2-dependence was used. The simulated sample is split equally into statistically independent training and validation sub-samples, enabling both optimization of the classifier and verification of its performance while avoiding bias due to overtraining.

Variable	$B^+ \to K^+\nu\bar{\nu}$
ΔE_{ROE}	0.592
Modified Fox-Wolfram $H^{so}_{m,2}$	0.135
p_{ROE}	0.045
Modified Fox-Wolfram R^{oo}_2	0.037
Fox-Wolfram Moment R_1	0.036
Harmonic Moment B_0	0.036
Modified Fox-Wolfram $H^{so}_{m,4}$	0.035
Modified Fox-Wolfram R^{oo}_0	0.034
$\cos(\mathrm{thrust}_B,\ \mathrm{thrust}_{\mathrm{ROE}})$	0.017
$\theta(p_{\mathrm{ROE}})$	0.013
$\cos(\theta(\mathrm{thrust}))$	0.012
Harmonic Moment B_2	0.007

Table 4.3: Features and relative importance of BDT_1, where the importance is normalized to unity.

The hyperparameters of the BDT_1 used in the analyses are listed in Tab. 4.4. These values were optimized to achieve a balance between model performance and robustness against overfitting. The number of trees is set to 2000, with tree depths of 2, a shrinkage (learning rate) of 0.2, a sampling rate of 0.5, and 256 equal-frequency bins for discretizing continuous variables. Increasing the value of any of these parameters generally increases the complexity of the model, which can improve fit to training data but also raises the risk of overfitting. The chosen values reflect an optimized trade-off between expressiveness and generalization.
We first ensure that the classifier is not overfitting. To check for potential overfitting in BDT_1, we performed a standard train-test comparison. The BDT prediction, which takes values between 0 and 1 after applying the event weights, is compared for signal and background between the training and testing samples. A score close to 0 indicates a background-like event, while a score close to 1 indicates a signal-like event. Good agreement between the training and testing distributions, Fig. 4.11 (left), indicates that the classifier generalizes well and is not overfitting the training samples.
Next, we interpret the feature importances. The feature importance of BDT_1 for the $B^+ \to K^+\nu\bar{\nu}$ analysis are shown in Tab. 4.3. It is evident that ΔE_{ROE} dominates the classifier, contributing approximately 59% of the total importance, followed by the Modified Fox-Wolfram $H^{so}_{m,2}$ with 13.5%. These two variables clearly

Parameter	Value
Number of trees	2000
Tree depth	2 (BDT_1) / 3 (BDT_2)
Shrinkage	0.2
Sampling rate	0.5
Number of equal-frequency bins	2^8

Table 4.4: Optimised hyperparameters of the classification models of first and second classifiers.

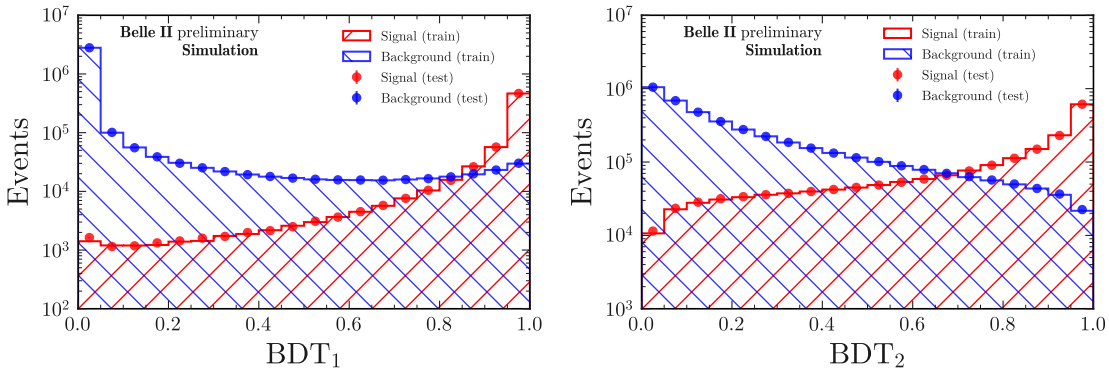

Figure 4.11: Simulated signal and background events taken from the training and the test sample in bins of the BDT_1 (left) and BDT_2 (right) output for $B^+ \to K^+\nu\bar{\nu}$. For the BDT_2 output, we applied $\mathrm{DBT}_1 > 0.9$. Good agreement indicates no evidence of overfitting.

carry the most discriminating power between signal and background events, capturing the overall $B\bar{B}$ versus $q\bar{q}$ event shape as well as the missing momentum in the event. The remaining features individually contribute much smaller fractions (ranging from 0.7% to 4.5%), but collectively they provide complementary information that enhances the overall performance of the BDT_1. While BDT_1 is highly effective at rejecting continuum $q\bar{q}$ background, it lacks sensitivity to specific $B\bar{B}$ backgrounds containing charm and resonance structures. BDT_2 addresses this by incorporating missing-energy and D-veto observables, significantly enhancing separation power in the signal region.

4.2.2 The second Classifier

Distributions are compared only in the sideband region ($\mathrm{BDT}_1 > 0.9$) to preserve blindness in the signal-enhanced region. With this cut, approximately 34% of the signal events are retained while 98.5% of the background is rejected, significantly improving the signal-to-background ratio.
In order to further investigate the performance of the classifier inputs, we compare the absolute event yields of various variables with the background scaled to the luminosity of the whole dataset while the signal contribution is scaled by a factor of 10^3 to enable a fair comparison with the background. In order to examine the quality of the data-simulation correction and the discrimination between signal and background without biasing the analysis, the signal region is blinded by requiring $\mathrm{BDT}_1 < 0.99$. The ratio panels are restricted to the range $[0.75, 1.25]$ to emphasize possible discrepancies while suppressing the impact of statistical fluctuations. The three representative variables, q^2_{rec}, $M(K^+X^-)$, and $\cos(\mathrm{thrust}_B, \mathrm{thrust}_{\mathrm{ROE}})$ are shown in Fig. 4.12. The three displayed variables have good agreement between data and simulation, but excluding D-veto inputs in BDT1 leaves peaks near D, K^*, and ϕ in $M(D^0)$ ($M(K^+X^-)$). Even scaling the signal by a thousand, it remains negligible relative to the background. A single BDT using only a few event-shape inputs is insufficient for optimal separation. To improve the discrimination power, we train a second classifier (BDT_2) on events with $\mathrm{BDT}_1 > 0.9$, making use of an extended feature set. In addition to

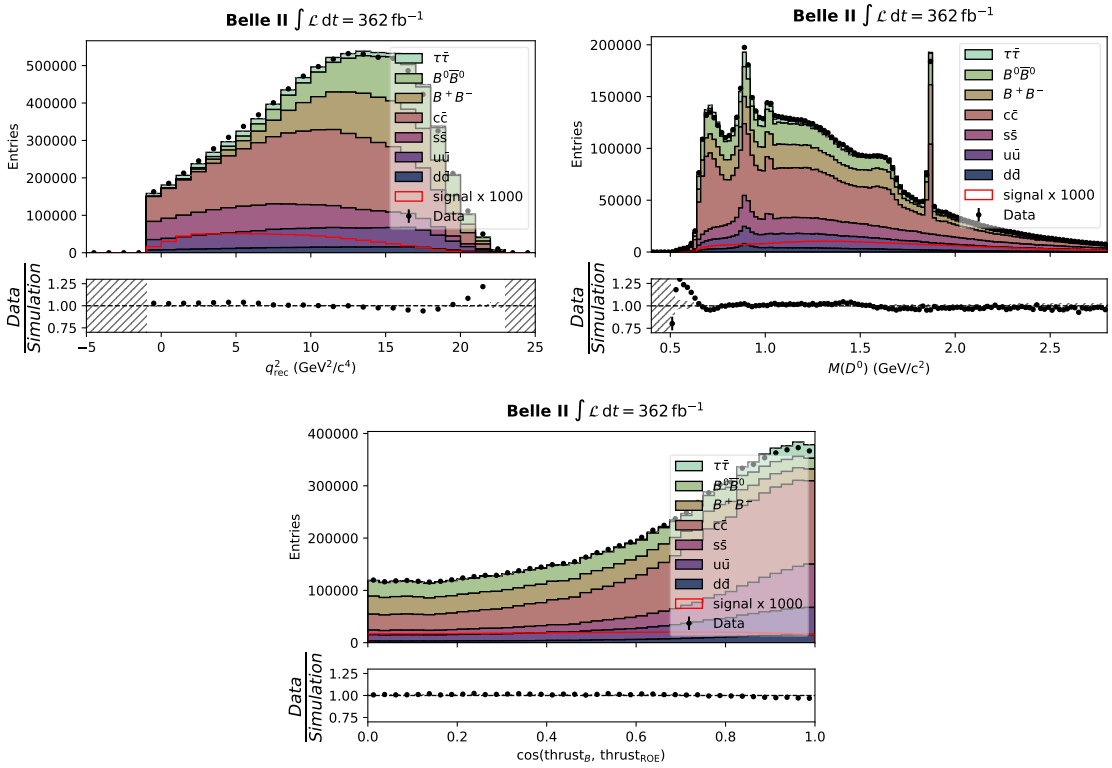

Figure 4.12: Distributions of representative variables, q^2_{rec}, $M(K^+X^-)$ and $\cos(\mathrm{thrust}_B, \mathrm{thrust}_{\mathrm{ROE}})$, in the BDT_1 sideband ($\mathrm{BDT}_1 < 0.99$). Data (points with error bars) are compared to simulation (stacked histograms) scaled to the full dataset luminosity, with the $B^+ \to K^+\nu\bar{\nu}$ signal overlaid as a line histogram scaled by 10^3 for visibility. The total $\Upsilon(4S)$ data luminosity is used.

event-shape observables that separate $q\bar{q}$ from $B\bar{B}$ events, the input space is enriched with variables related to missing energy, charm-meson vetoes, and other kinematic quantities. In total, 35 variables, with good consistency of data and simulation, are used as inputs (listed in the Appendix 9.1), where their distributions are shown in density-normalized form for clarity and with $\mathrm{BDT}_1 < 0.99$ to avoid unblinding.

For BDT_2, we adopt the XGBoost framework, which provides fast training, robust handling of correlations, and efficient regularization. The background training sample corresponds to 200 fb^{-1} of simulated data (about 4.2 M events), with another independent 200 fb^{-1} sample used for testing. On the signal side, 3.4 M simulated events are split into statistically independent training and test samples.

It should be noted that XGBoost does not automatically rebalance the signal-to-background statistics. In this analysis, the class imbalance is handled explicitly by balancing the ratio of background to signal events, in combination with the event weights provided from the simulation.

As summarized in Table 4.4, the XGBoost classifier (BDT_2) explores one additional tree depth compared to the FastBDT classifier (BDT_1), allowing it to capture more complex correlations among the input variables while still being protected against overfitting by the chosen shrinkage and sampling parameters. The comparison of the training and testing responses in Fig. 4.11 (right) shows excellent consistency, demonstrating that BDT_2 achieves stable performance without signs of overtraining.

The feature importance of the second classifier is listed in Tab. 4.5. The 35 inputs span event-shape, missing-energy, and charm-veto categories. Their relative importance demonstrate that BDT_2 leverages complementary information from all domains, ensuring robust separation between signal and background.

Once the BDT classifiers (BDT_1 and BDT_2) are trained, they are applied to events to evaluate the expected signal efficiency and signal–background significance. The signal efficiency, ϵ_{sig}, is defined as the fraction of generated signal events passing all selections, including both the initial selection and the BDT response. The

Variable	$B^+ \to K^+\nu\bar{\nu}$
$\cos(\text{thrust}_B,\ \text{thrust}_{\text{ROE}})$	0.099
Median(p-value(D^0))	0.090
Modified Fox-Wolfram $H^{so}_{m,2}$	0.079
Modified Fox-Wolfram R^{oo}_2	0.068
p-value(Tag Vertex)	0.050
p-value(D^+)	0.050
$M(D^0)$	0.048
$\theta(p_{\text{missing}})$	0.042
Modified Fox-Wolfram $H^{so}_{n,2}$	0.037
Total charge squared	0.036
Modified Fox-Wolfram $H^{so}_{m,4}$	0.036
Modified Fox-Wolfram $H^{so}_{c,2}$	0.034
N_{lepton}	0.030
$dr(K^+)$	0.029
$dz(K^+,\ \text{Tag Vertex})$	0.021
ΔE_{ROE}	0.019
$dz(\text{Tag Vertex})$	0.019
$\theta(p_{\text{ROE}})$	0.019

Variable	$B^+ \to K^+\nu\bar{\nu}$
p_{ROE}	0.017
$\cos(\text{thrust}_B,\ z)$	0.017
p-value(D^0)	0.016
N_{tracks}	0.015
N_γ	0.014
$dx(\text{Tag Vertex})$	0.013
$dr(D^+)$	0.013
Fox-Wolfram Moment R_3	0.012
Fox-Wolfram Moment R_2	0.012
${M_{\text{missing}}}^2$	0.011
Fox-Wolfram Moment R_1	0.009
Sphericity	0.009
$\cos(\theta(\text{thrust}))$	0.009
$\text{Variance}_{\text{ROE}}(p_T)$	0.009
Modified Fox-Wolfram R^{oo}_0	0.008
$dr(K^+,\ \text{Tag Vertex})$	0.007
Harmonic Moment B_0	0.004

Table 4.5: Features and relative importance of BDT_2, normalized to unity. The table is split into two columns for readability.

expected number of signal events after the classifier, s, is calculated as

$$s = \epsilon_{\text{sig}} \times N^{\text{SM}}_{\text{sig}} \tag{4.3}$$

$N^{\text{SM}}_{\text{sig}}$is the expected number of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ events produced in $\mathcal{L} = 362\ \text{fb}^{-1}$ of data according to the SM prediction $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = 4.97 \times 10^{-6}$ [37].

$$N^{\text{SM}}_{\text{sig}} = \mathcal{L} \times \sigma(e^+e^- \to \Upsilon(4S)) \times 2 \times \mathcal{B}(\Upsilon(4S) \to B^+B^-) \times \mathcal{B}(B^+ \to K^+\nu\bar{\nu}) \tag{4.4}$$

The factor of 2 accounts for the fact that each $\Upsilon(4S) \to B^+B^-$ decay produces one B^+ meson that may undergo the signal decay. The cross-section of $\sigma(e^+e^- \to \Upsilon(4S)) = 1.077 \times 10^6$ fb, and branching fraction of $\mathcal{B}(\Upsilon(4S) \to B^+B^-) = 0.516$ are applied. The background yield, b, is determined from the number of background events that survive the classifier predictions. The statistical significance is then evaluated as

$$\frac{s}{\sqrt{s+b}}, \tag{4.5}$$

This definition is appropriate in the regime of moderate to large statistics. For small yields, alternative treatments (for instance, profile likelihoods) would be required, but here $s/\sqrt{s+b}$ provides a robust figure of merit for optimization. The significance versus signal efficiency plot, as shown in Fig. 4.13(left), illustrates the trade-off between retaining signal events and rejecting background. The maximum achievable signal efficiency is 0.36. For BDT_1, the significance peaks at approximately 0.45, occurring at a low signal efficiency range of 0.04-0.08, indicating that stringent cuts yield high purity but low signal retention. In contrast, BDT_2 achieves a substantially higher peak significance of about 1.4 at a slightly higher efficiency of 0.04–0.05, demonstrating that the additional variables and improved training allow for stronger signal–background discrimination. The improvement of BDT_2 over BDT_1 indicates that including charm-veto, missing-energy, and additional kinematic observables provides genuine additional discrimination power beyond event-shape information.

We map the raw BDT_2 output score onto a uniform scale in signal efficiency. The inefficiency metric $\eta(\text{BDT}_2)$ is defined such that it increases linearly with the fraction of signal events rejected when tightening the BDT_2

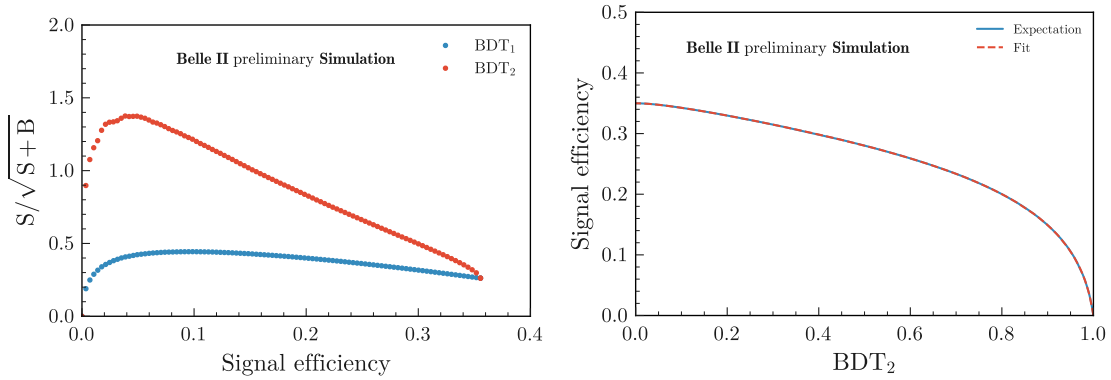

Figure 4.13: Left: Expected significance as a function of signal efficiency for BDT_1 and BDT_2. Right: Signal efficiency as a function of the BDT_2 score, used to define the uniform inefficiency metric $\eta(\mathrm{BDT}_2)$.

score cut. By construction, simulated signal events are uniformly distributed in $\eta(\mathrm{BDT}_2)$.

$$\eta(\mathrm{BDT}_2) = 1 - \int_c^1 \xi(c')\, dc', \tag{4.6}$$

where $\xi(c)$ is the total signal-selection efficiency density for a given BDT_2 score c. This definition effectively maps the BDT_2 score onto a physically meaningful scale, reflecting the fraction of signal events yet to be selected as the score threshold varies. Thus, the $\eta(\mathrm{BDT}_2)$ is uniform for simulated signal events. We examine the distributions of $\eta(\mathrm{BDT}_2)$, q^2_{rec}, and $M(D^0)$ in the sideband region defined by $0.75 < \eta(\mathrm{BDT}_2)$<0.90, which is blinded from the signal extraction, shown in Fig. 4.14.

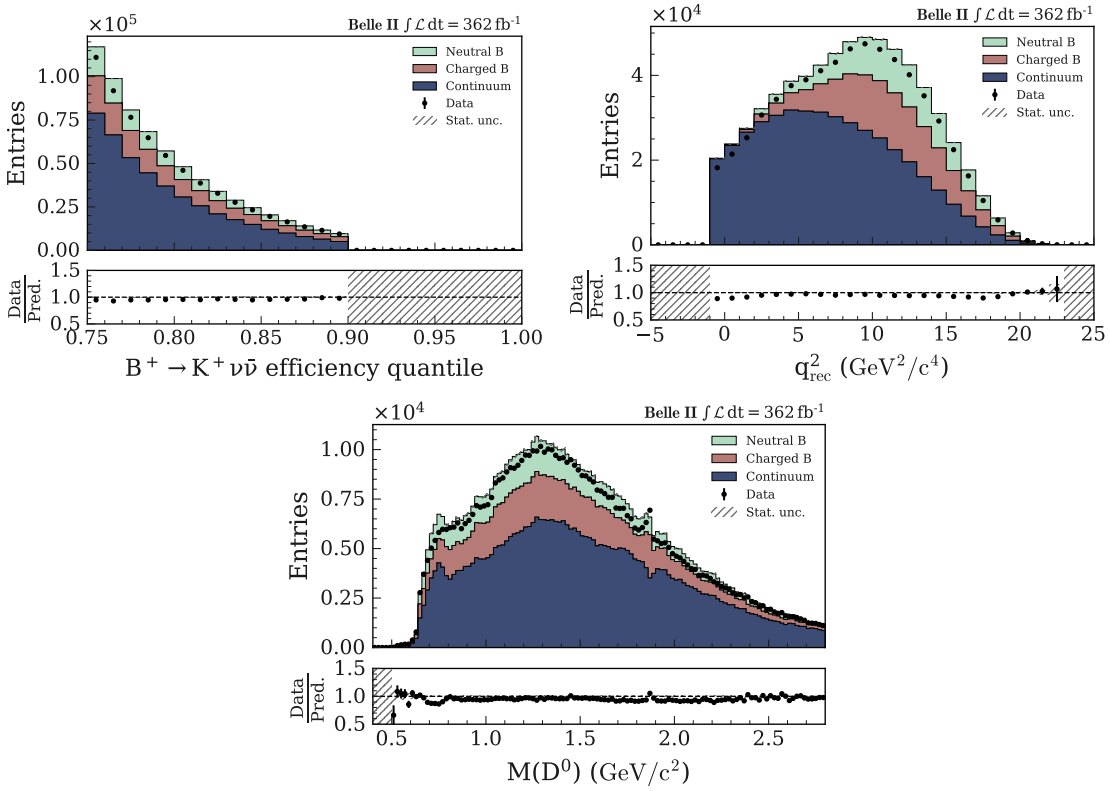

Figure 4.14: Distributions of $\eta(\mathrm{BDT}_2)$, q^2_{rec}, and $M(D^0)$ in the BDT_2 sideband ($0.75 < \eta(\mathrm{BDT}_2) < 0.90$). The sideband is blinded from the final signal extraction. The total $\Upsilon(4S)$ data luminosity is used.

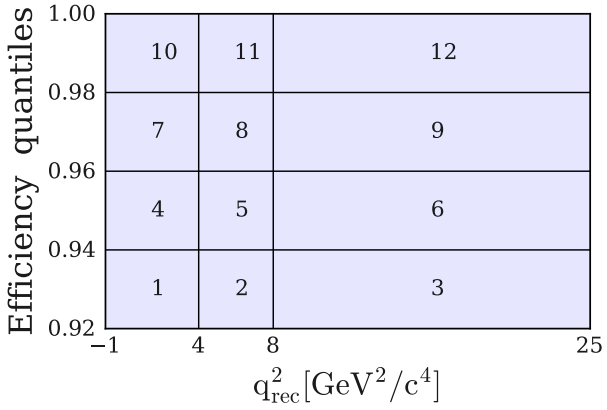

Figure 4.15: Two-dimensional signal region binning in $\eta(\mathrm{BDT}_2)$ and q^2_{rec} for the $B^+ \to K^+\nu\bar{\nu}$ analysis. Equal-width binning is used in $\eta(\mathrm{BDT}_2)$, while q^2_{rec} bins follow theoretical guidance with statistical adjustments at high q^2_{rec}.

4.3 Signal Region Definition and Signal Efficiency Validation

After suppressing the dominant backgrounds with the two classifiers, the next step is to define a signal region (SR) where the final search is performed. Since the statistical analysis is based on a maximum binned likelihood fit, the SR must be chosen to maximize sensitivity while allowing for an appropriate partitioning into bins. The strategy balances three considerations: (i) maximizing expected significance, (ii) preserving sufficient signal statistics, and (iii) exploiting kinematic features such as the q^2 spectrum. With these goals in mind, the SR is defined using the inefficiency metric $\eta(\mathrm{BDT}_2)$ in combination with the BDT_1 selection.

4.3.1 Signal Region Definition

For this analysis, the optimal selection corresponds to $\eta(\mathrm{BDT}_2) > 0.92$, in conjunction with the requirement of $\mathrm{BDT}_1 > 0.9$. This choice retains approximately 8% of the total signal, reflecting a balance between signal purity and statistical yield. To further exploit kinematic information and the distribution of signal events, the selected signal region is subdivided into bins of reconstructed di-neutrino mass q^2_{rec} and $\eta(\mathrm{BDT}_2)$. This two-dimensional binning captures variations in both the di-neutrino mass spectrum and the signal significance across the event sample, thereby allowing a more granular optimization of sensitivity. Such a binning strategy ensures that regions with high signal-to-background ratios contribute maximally to the final statistical evaluation, while maintaining control over background fluctuations.
Given that the inefficiency variable $\eta(\mathrm{BDT}_2)$ is uniformly distributed for signal events, equal-width binning is adopted in the selected signal region. Specifically, we divide the range [0.92, 1.00] into four bins, [0.92,0.94, 0.96, 0.98, 1.00]. For the reconstructed di-neutrino mass variable, q^2_{rec}, the kinematic range is [-1, 25]. While theory guidance in [36] suggests binning of q^2 as [0, 4, 8, 12, 16, 25], the limited statistics in the higher q^2_{rec} region necessitates combining the last three intervals. Accordingly, we adopt three bins [-1, 4, 8, 25].
This binning of signal region in a two-dimensional 4×3 scheme of $\eta(\mathrm{BDT}_2) \times \mathrm{q}^2_{\mathrm{rec}}$ shown in Fig. 4.15. Such binning allows us to simultaneously exploit both the discriminating power of the BDT-based inefficiency metric and the theoretically motivated kinematic structure of the signal. In particular, the highest $\eta(\mathrm{BDT}_2)$ bin, $\eta(\mathrm{BDT}_2) > 0.98$ provides the dominant sensitivity to the signal, is called the most sensitive region (denoted as HSR), while the lowest bin, $\eta(\mathrm{BDT}_2) < 0.94$ serves primarily to constrain background contributions. This design balances statistical power, theoretical motivation, and background control, thereby optimizing the expected performance of the binned likelihood fit.
Within the defined signal region, the expected yields of the Standard Model signal and background are 160 and 16,793 events, respectively. Restricting to $\eta(\mathrm{BDT}_2)$>0.98, the expected SM signal yield decreases to 40 events, accompanied by 977 background events. This illustrates the trade-off between maximizing signal

Selection stage	signal selection efficiency ϵ
Object selection (acceptance)	0.89
Signal candidate selection	0.55
First signal candidate selection	0.53
Basic event selection	0.41
BDT_1 filter	0.34
Signal search region	0.08
The most sensitive signal search region	0.02

Table 4.6: Cumulative signal-selection efficiency at each stages of the analysis for the $B^+ \to K^+\nu\bar{\nu}$. Uncertainties are statistical only. PID weights are applied starting from the basic event selection step.

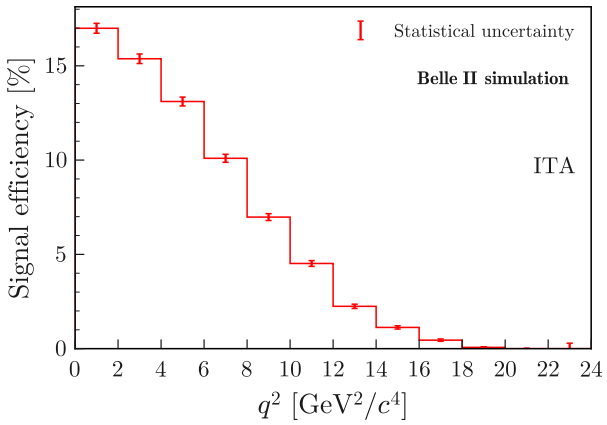

Figure 4.16: Signal-selection efficiency as a function of the true di-neutrino invariant mass squared (q^2) for simulated signal events passing all selections in the signal region.

purity and retaining statistical power: while the tighter selection enhances the signal-to-background ratio, it significantly reduces the available signal statistics. The two-dimensional binning in $\eta(\mathrm{BDT}_2) \times \mathrm{q}^2_{\mathrm{rec}}$ mitigates this tension by enabling the fit to simultaneously exploit the high-purity region for sensitivity to the signal while using lower-purity bins to constrain the background contribution.

Table 4.6 summarizes the cumulative signal-selection efficiency at each stage of the analysis. Starting from the object-level acceptance, where most of the generated signal is retained ($\epsilon = 0.89$), the successive selections from signal candidate to the whole event keep two-fifth of signals. The application of the BDT_1 classifier further improves background rejection, with an efficiency of $\epsilon = 0.34$. The signal region and high sensitivity region are kept at 8% and 2% signal events in enhancing the overall sensitivity of the analysis.

To study potential kinematic biases, the signal efficiency is evaluated as a function of the generated di-neutrino mass squared q^2, rather than the reconstructed variable, shown in Fig. 4.16. Within the defined signal region, the efficiency exhibits a dependence on q^2. At the low-q^2 region, corresponding to high-momentum kaons in the final state, the efficiency reaches values as high as 17%. As q^2 increases, the kaon momentum decreases, making the signal topology less distinctive and more easily mimicked by background processes. Consequently, the efficiency falls at higher q^2.

4.3.2 Signal Selection Efficiency Validation

Because the $B^+ \to K^+\nu\bar{\nu}$ signal has neutrinos in the final state, direct validation of efficiency in data is impossible. Instead, we employ an embedding method using the control channel $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$, which provides a large branching fraction and a clean experimental signature. By replacing the K^+ candidate

in this channel with the kaon from simulated $B^+ \to K^+\nu\bar{\nu}$ decays, while ignoring the two muons, we can mimic the signal topology in a realistic event environment. The embedding procedure proceeds as follows:

a. **Selection of the control sample:** $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ candidates are reconstructed in both data and simulation. The two oppositely charged muons are required to have an invariant mass within ± 50 MeV/c^2 of the nominal J/ψ mass. To suppress combinatorial background, we impose $|\Delta E| < 100$ MeV and $M_{\rm bc} > 5.27$ GeV/c^2. These requirements yield a high-purity sample with minimal background contamination (below 2%).

b. **Retention of the rest-of-event:** All reconstructed objects associated with the $B^+ \to K^+J/\psi$ candidate are removed, while the rest-of-event, which contains the decay products of the accompanying B^- meson, is retained. This ensures that the event maintains realistic tagging and detector conditions.

c. **Preparation of signal decays:** Simulated $B^+ \to K^+\nu\bar{\nu}$ events are processed, retaining only the objects associated with the signal decay. These serve as the replacement signal candidates.

d. **Event embedding:** The isolated $B^+ \to K^+\nu\bar{\nu}$ decay is merged with the ROE from the control sample to construct an embedded event. In this way, the event background and environment come from real data, while the kaon originates from the simulated signal process. This produces signal-like events under realistic experimental conditions.

e. **Kinematic alignment:** The kinematics of the embedded signal decay are adjusted to match those of the original $B^+ \to K^+J/\psi$ candidate. Specifically, the reconstructed kaon momentum and decay vertex are shifted and rotated so that the B^+ direction and decay vertex coincide with those measured in the control channel. This guarantees that the embedded signal inherits the same production and event geometry as the well-measured $B^+ \to K^+J/\psi$ decay.

Distribution of the classifier outputs for BDT_1 (main panel) and BDT_2 restricted to $\text{BDT}_1 > 0.9$ (inset), as shown in Fig. 4.17. Results are shown for $B^+ \to K^+J/\psi$ candidates in both data and simulation, before embedding and after the embedding procedure in which the muons are removed and the kaon momentum is replaced. For comparison, the corresponding distributions obtained directly from simulated $B^+ \to K^+\nu\bar{\nu}$ signal events are also overlaid. The simulated histograms are normalized to the number of $B^+ \to K^+J/\psi$ events observed in data.

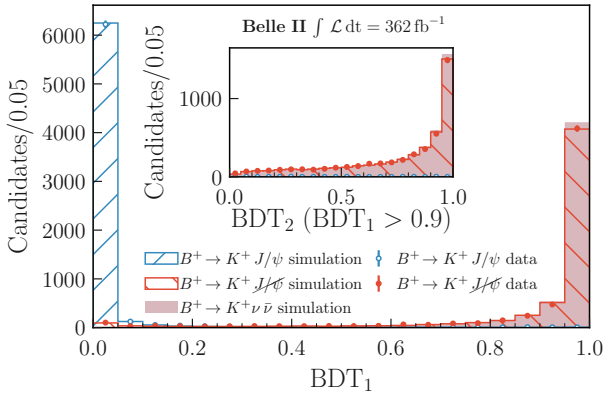

Figure 4.17: Distributions of classifier outputs for BDT_1 (main) and BDT_2 restricted to $\text{BDT}_1 > 0.9$ (inset), comparing $B^+ \to K^+J/\psi$ candidates in data and simulation before and after embedding. The overlaid histogram corresponds to simulated $B^+ \to K^+\nu\bar{\nu}$ events, normalized to the $B^+ \to K^+J/\psi$ yield in data.

The validation of the signal selection efficiency using the embedded $B^+ \to K^+J/\psi$ control sample is summarized in Table 4.7. The efficiency in the signal region is measured to be (18.2±0.5)% in data and (18.1±0.3)% in the corresponding simulation, yielding a ratio of (100 ± 3)%. This excellent agreement

Sample	signal selection efficiency(%)
Data	18.2±0.5
MC	18.1±0.3
Ratio	100±3

Table 4.7: Signal-region selection efficiency for the embedded $B^+ \to K^+ J/\psi$ control sample, comparing data and simulation. The ratio demonstrates excellent agreement, validating the efficiency modeling in simulation.

demonstrates that the reconstruction, selection, and classifier performance are well modeled in the Monte Carlo samples, and provides confidence in the reliability of the signal efficiency estimation used in the analysis.
This validation confirms that the signal efficiency estimated from the simulation is trustworthy, providing a solid foundation for the subsequent signal extraction.

4.4 Background Modeling and Validation

After the application of the second classifier, a comprehensive background study is performed. This includes the evaluation of $q\bar{q}$ modeling uncertainties, the treatment of specific $B\bar{B}$ background contributions at different levels, and validation with dedicated control samples. These studies provide the basis for a consistent definition of sideband regions and their use in the overall background validation strategy. To this end, a sideband region is defined as $0.75 < \eta(\mathrm{BDT}_2) < 0.9$. The choice of this sideband region is motivated by the following considerations:

- Correction for $q\bar{q}$ modeling is derived from a dedicated classifier, BDT_C, which has been trained using events with $\eta(\mathrm{BDT}_2) > 0.75$.
- Corrections for certain $B\bar{B}$ background components have been studied in the regions $\eta(\mathrm{BDT}_2) > 0.75$ and $\eta(\mathrm{BDT}_2) > 0.9$, which are expected to yield mutually consistent results.
- The signal region of the analysis is defined by $\eta(\mathrm{BDT}_2) > 0.92$.

Consequently, the region $0.75 < \eta(\mathrm{BDT}_2) < 0.9$ serves as a well-motivated sideband: it is adjacent to but distinct from the signal region, ensuring negligible signal contamination while remaining consistent with the regions used for training and background corrections. This construction allows for a robust validation of the background modeling before extracting results in the signal region.

4.4.1 Continuum Background Correction

After suppressing the dominant backgrounds with two multivariate classifiers, the $q\bar{q}$ continuum background remains significantly about 40% in the signal region and 17% in the most sensitive region. To validate this component, we compared off-resonance data with both off-resonance simulation and with the $e^+e^- \to q\bar{q}$ component in $\Upsilon(4S)$ simulations. These studies revealed a moderate disagreement between data and simulation, affecting both the normalization and the shape. Such a disagreement is most likely due to deficiencies in the fragmentation model of `Ptythia`, which is known to be a major source of uncertainty for continuum hadronization.

To mitigate the shape discrepancy, we applied a data-driven reweighting procedure following the method in Ref. [97]. A dedicated classifier, denoted BDT_C, was trained using the XGBoost algorithm. The training sample consisted of 43 fb^{-1} of off-resonance data, corresponding to 5.26×10^4 events, treated as signal, while 50 fb^{-1} of off-resonance simulation sample, corresponding to 5.33×10^4 events, treated as background. The input variables comprised the 35 BDT_2 training variables, supplemented by q^2_rec and BDT_2 output score. The inclusion of q^2_rec and BDT_2 score is particularly motivated, since these variables directly define the signal region and the binning of the maximum likelihood fit. Ensuring good data–simulation agreement in these observables is therefore essential. Because the $q\bar{q}$ control sample contains no signal contribution, including these variables as inputs to BDT_C does not introduce bias.

To ensure that the correction captures only broad, first-order differences between data and simulation, the maximum tree depth was restricted to one. This choice prevents the classifier from learning high-order correlations that could distort physics-sensitive distributions, while enabling effective reweighting based on the most relevant features. The hyperparameters of BDT_C are summarized in Tab. 4.8.

Parameter	Value
Number of trees	2000
Tree depth	1
Shrinkage	0.01
Sampling rate	0.01
Number of equal-frequency bins	2^8

Table 4.8: Optimised hyperparameters of the BDT_{rmC} classifier.

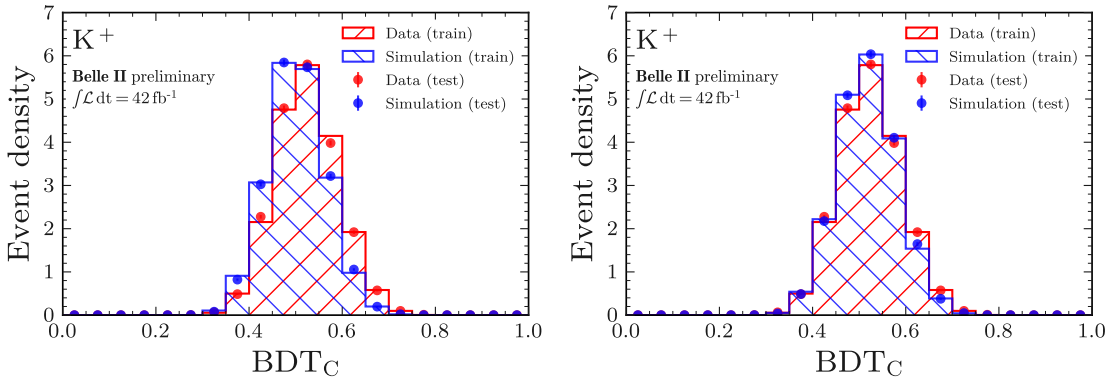

Figure 4.18: Validation of $\mathrm{BDT_C}$: output distribution (left) demonstrating no overfitting, and effect of applying $\mathrm{BDT_C}$ weights to continuum simulation (right)

source of BDT_c weight	$\eta(\mathrm{BDT_2}) > 0.75$	$\eta(\mathrm{BDT_2}) > 0.92$	$\eta(\mathrm{BDT_2}) > 0.92$ with BDT_c weight
off-resonance MC	1.27	1.50	1.40±0.05
continuum MC	1.27	1.49	1.36±0.05

Table 4.9: Ratio of off-resonance data and continuum MC with different BDT C weights: training with off-resonance MC (2nd row) and continuum MC (3rd row). The ratios are consistent within uncertainty

Fig. 4.18 (left) shows the output distribution of the $\mathrm{BDT_C}$. For this validation, half of the data and simulated samples were used for training, while the remaining half were reserved as independent test samples. This separation allows us to verify that the classifier is not overfitting and that its discriminating power is genuine. For the final $\mathrm{BDT_C}$ training used in the analysis, the full data and simulation samples were employed to maximize statistical precision.
The classifier output, p_c, is defined such that $p_\mathrm{c} = 0.5$ indicates equal consistency with data and simulation. Interpreting p_c as a likelihood ratio, the event weight is defined as

$$w = \frac{p_\mathrm{c}}{1 - p_\mathrm{c}} \tag{4.7}$$

which corresponds to likelihood ratio of data and simulation, $\frac{L(\mathrm{data})}{L(\mathrm{simulation})}$, This $\mathrm{BDT_C}$ weight is applied to simulated $q\bar{q}$ events to correct the fragmentation-driven shape discrepancies. The effect of applying the $\mathrm{BDT_C}$ weights is illustrated in Fig. 4.18 (right). As expected, this reweighting procedure improves the agreement between data and simulation.
Tab. 4.9 summarizes the normalization ratios of off-resonance data to continuum MC under different $\mathrm{BDT_C}$ weighting schemes. The first $\mathrm{BDT_C}$ training uses off-resonance MC as background (nominal), while the second uses the continuum component in $\Upsilon(4S)$ MC.
Since the $\mathrm{BDT_C}$ is trained in the region $\eta(\mathrm{BDT_2}) > 0.75$ and the weights are normalized, the overall normalizations remain consistent across the different weighting schemes. Small differences are observed in the signal region ($\eta(\mathrm{BDT_2}) > 0.92$), reflecting the sensitivity of the reweighting to the choice of training sample and the fine details of the continuum modeling. Nevertheless, the ratios are consistent within the statistical uncertainties, demonstrating that the $\mathrm{BDT_C}$ reweighting is robust with respect to the choice of training background sample.
We compare the off-resonance data and $e^+e^- \to q\bar{q}, \tau^+\tau^-$ components in $\Upsilon(4S)$ simulation for key kinematic variables. Figure 4.19 shows the distributions of the Fox-Wolfram moment R_2 and the q^2_rec. On the left panels, the simulated events are shown without applying the $\mathrm{BDT_C}$ weights, while on the right panels, the $\mathrm{BDT_C}$ weights are applied to the simulation. This demonstrates that the $\mathrm{BDT_C}$ procedure effectively corrects for

the shape discrepancies, resulting in a continuum $q\bar{q}$ simulation that is consistent with the observed data across these key variables.

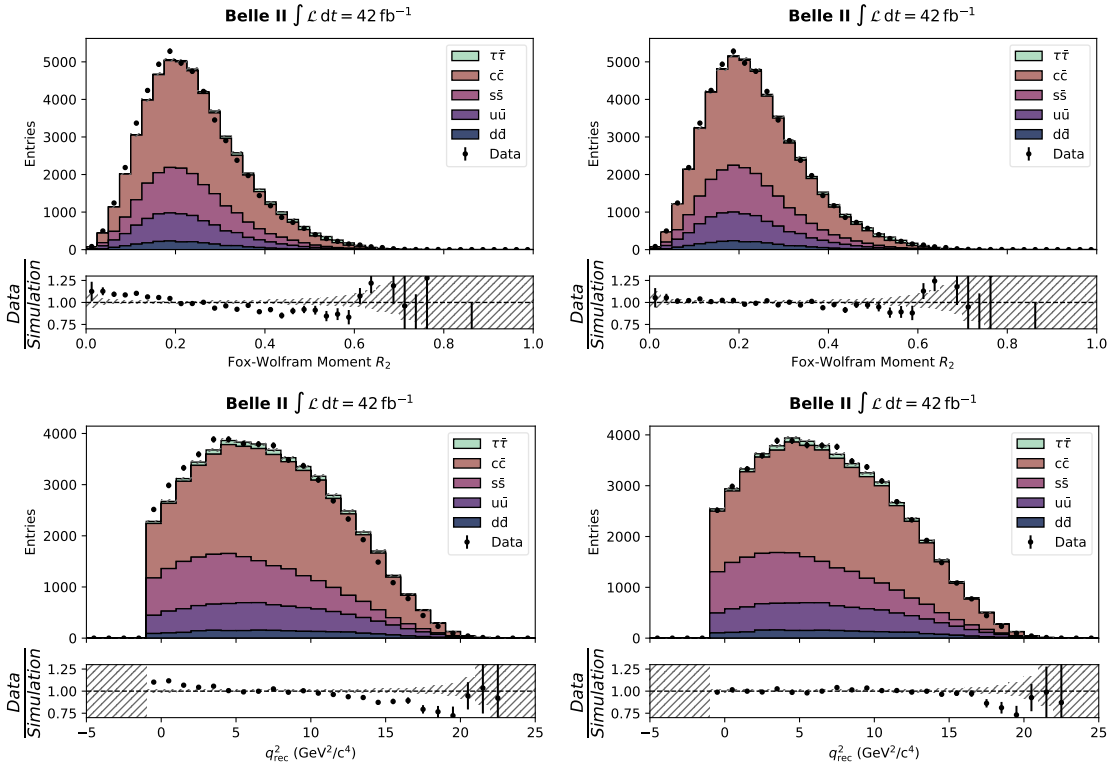

Figure 4.19: Comparison of off-resonance data and simulation for the Fox–Wolfram moment R_2 and reconstructed q^2_{rec}. Left: simulation without $\mathrm{BDT_C}$ reweighting. Right: simulation with $\mathrm{BDT_C}$ reweighting applied.

Within the defined sideband region, the background is decomposed into three dominant components: $q\bar{q}$, B^+B^-, and $B^0\bar{B}^0$. The relative contributions of these categories are illustrated in Fig. 4.20, where the $q\bar{q}$ component is reweighted with $\mathrm{BDT_C}$ and normalized by a residual factor of 1.3.

After applying the $\mathrm{BDT_C}$ weights, a residual normalization discrepancy remains in the off-resonance sample. In the signal region, the simulation is scaled by a factor of 1.40 ± 0.05 to match the data. Figure 4.21 shows the comparison in the two-dimensional signal-region search bins, after reweighting and scaling.

4.4.2 $B\bar{B}$ Background Corrections

The $B\bar{B}$ background is decomposed into several categories, illustrated in Fig. 4.22. The dominant contribution arises from semileptonic B decays to charm mesons, in which the kaon candidate originates from a charm-meson decay. This process accounts for approximately 47% of the total B background in the signal region. The next largest components are hadronic B decays involving charmed mesons (about 38%) and other hadronic B decays (about 14%). Smaller but non-negligible contributions arise from $B^+ \to \tau^+\nu_\tau$ decays and rare $B \to K^*\nu\bar{\nu}$ processes. Within the charm background, direct $B \to D$ transitions dominate over decays involving D^* resonances, due to their lower particle multiplicity. Contributions from higher excitations of charm mesons (D^{**} modes), which are less precisely measured experimentally, amount to roughly 4% of the total B background. These states are modeled according to the `Pythia 8` simulation [63]. In the following sections, the modeling of the major background categories is discussed in detail, together with the treatment of specific decay modes that require special handling.

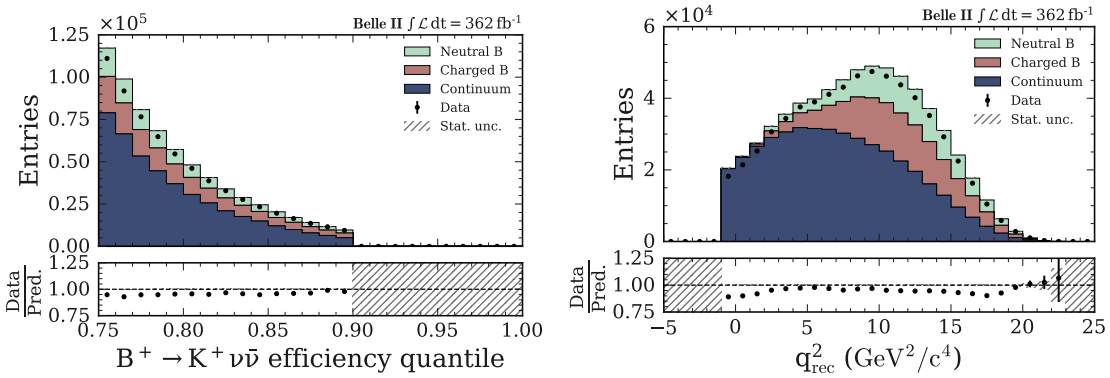

Figure 4.20: Background composition in the BDT_2 sideband region. Distributions are shown after applying BDT_C reweighting to the $q\bar{q}$ component and scaling by the residual normalization factor.

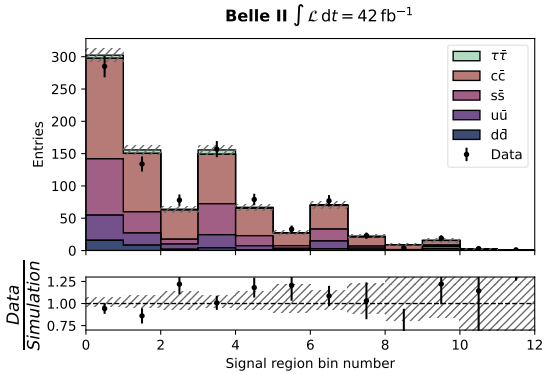

Figure 4.21: Comparison of off-resonance data and simulated continuum background in the two-dimensional signal-region search bins of the K^+ channel. The simulation includes BDT_C reweighting and is further scaled by a factor of 1.40 ± 0.05 to match the total data yield. The total $\Upsilon(4S)$ data luminosity is used.

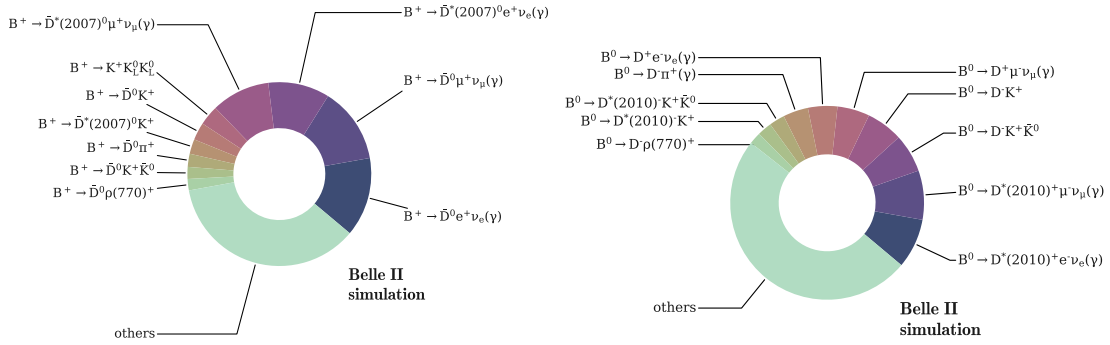

Figure 4.22: Composition of $B\bar{B}$ background in the signal region: charged B decays (left) and neutral B decays (right)

	π enrich sample	μ enriched sample	e enriched sample
$K^{\pm}$	K fake π	-	-
$\pi^{\pm}$	$\pi ID > 0.9$	π fake μ	-
$\mu^{\pm}$	μ fake π	$\mu ID > 0.9$	$eID < 0.9$
$e^{\pm}$	e fake π	$\mu ID < 0.9$	eID>0.9
normalization factor	1.30±0.02	1.35±0.01	1.38±0.01

Table 4.10: Definition of PID sidebands and considered momentum- and polar-angle-dependent PID weights. The corrections are obtained from the systematics framework. Fake rate of leptons to pions are also considered, with the caveat that it is a flat factor 2.4 (2.35) for muons (anti-muons) and 2.35 (2.48) for electrons (positrons).

4.4.2.1 Modeling of D-meson Decays

The accurate modeling of B-meson decays producing kaons is particularly critical in the high-sensitivity region of this analysis. In particular, decays such as $B^0 \to K^+ D^{(*)-}$ and $B^+ \to K^+ D^{(*)0}$ are especially importance. A sizable fraction of these decays proceeds through channels where the D-meson decays into final states containing a long-lived neutral kaon (K_L^0). Since the inclusive branching fractions of $D \to K_L^0 X$ transitions are not precisely measured, they introduce sizable uncertainties into the background description. Moreover, such decays closely mimic the signal topology: the prompt K^+ from the B-meson decay resembles the signal kaon candidate, while the undetected K_L^0 contributes to the missing energy in the final state. This produces background events with kinematic features that overlap strongly with the signal region, especially in the reconstructed q^2_{rec} distribution. To constrain this contribution, dedicated control samples are constructed by relaxing the standard kaon identification requirements, thereby providing an independent handle on the fraction of $B \to D(\to K_L^0 X)$ decays.
Three control samples with modified particle identification criteria are exploited: π-enriched, μ-enriched, and e-enriched samples. These samples are constructed by applying the same selection criteria as the nominal analysis, except that the kaon-identification requirement for the signal candidate is replaced by a tight $\pi/\mu/e$-identification criterion, as summarized in Tab. 4.10. For each sample, the corresponding momentum- and polar-angle-dependent PID efficiencies and fake rates are applied, derived from the systematic framework. In addition to the standard PID corrections, flat misidentification factors are applied for lepton-to-pion fake rates: 2.4 (2.35) for muons (anti-muons) and 2.35 (2.48) for electrons (positrons). These corrections ensure that the simulation reproduces the observed lepton misidentification probabilities in data, which are particularly important in the lepton-enriched sidebands. By construction, the three sideband samples are orthogonal to the nominal signal selection (Kaon ID>0.9) and thus provide an independent and unbiased handle on the modeling of $B \to D(\to K_L^0 X)$ decays. To improve the data−simulation agreement and maintain consistency with the nominal analysis, the same BDT_C is applied to the continuum component in each control sample, while its normalization with the modeling correction in the signal region, 1.30 ± 0.05, is taken from the corresponding off-resonance sample rather than the nominal one to reflect the actual conditions. To study the $B \to D(\to K_L^0 X)$ contribution, the sample is divided into three categories: continuum ($q\bar{q}$) events, B-meson decays containing a D meson whose decay products include a K_L^0, and B-meson decays without such a D meson in the decay chain. A structured discrepancy is observed in the data-simulation ratio, shown in Fig 4.23(upper). The q^2_{rec} quantifies the squared invariant mass of the system recoiling against the π emitted in the B decay, reflecting the kinematics of the unseen part of the decay.

- In the continuum-dominated region with $q^2_{\text{rec}} < 0$, the data lie below the simulation, suggesting a mild overestimation from the off-resonance normalization
- Below the D-meson mass threshold, the simulation overestimates the data, dominated by continuum events.
- Above the D-meson mass threshold, simulation lies below the data, reflecting imperfections in model-

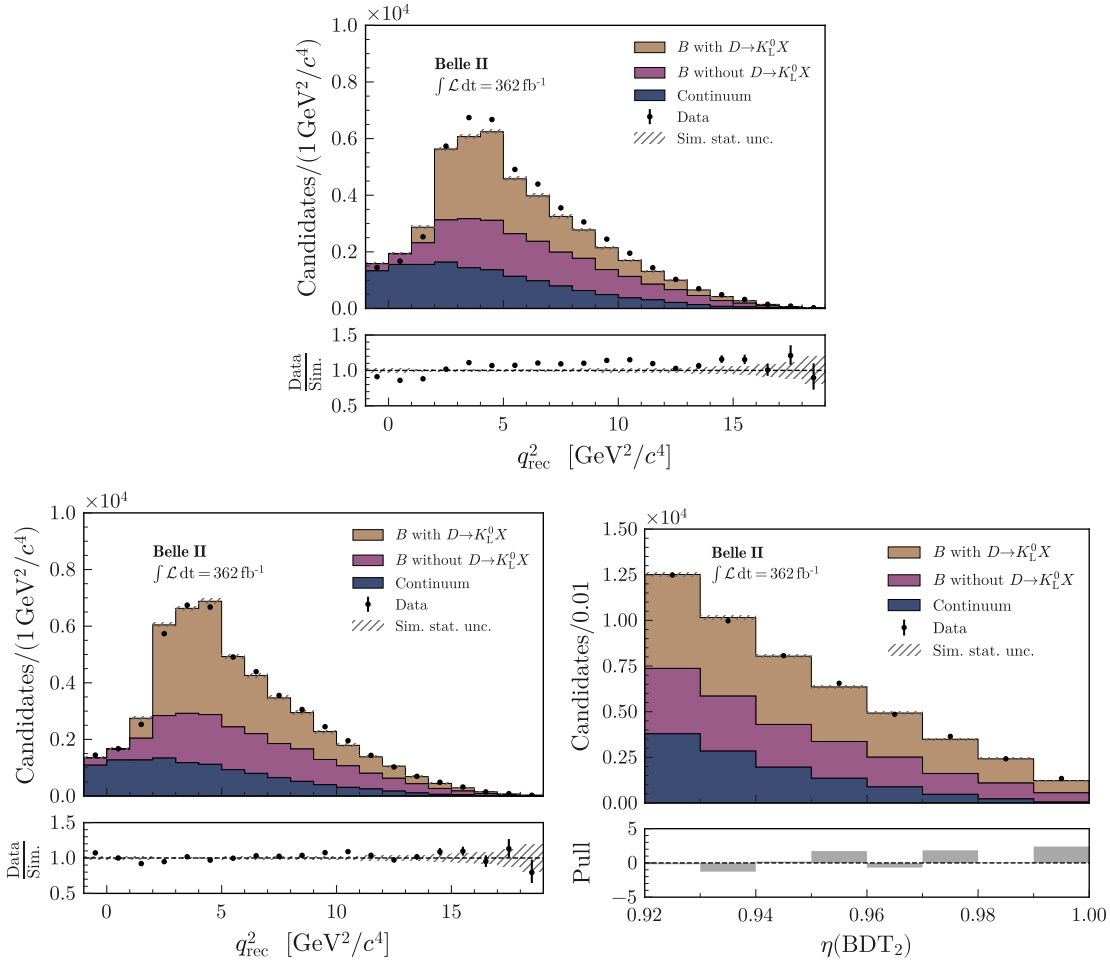

Figure 4.23: Distributions in the π-enriched control sample: $q^2_{\rm rec}$ before correction (upper), $q^2_{\rm rec}$ after correction (bottom left), and $\eta(\mathrm{BDT}_2)$ after correction (bottom right).

ing $B \to D(\to K^0_L X)$ decays and their kinematic distributions.

A binned fit to the $q^2_{\rm rec}$ distribution is used to determine the normalizations of the three components in the signal region. The B-meson decays containing $D \to K^0_L$ are left free in the fit, while the complementary B decays are constrained by the uncertainty in the total number of $B\bar{B}$ pairs. The continuum component is assigned a 50% uncertainty, consistent with the nominal analysis. The normalization factors for the π, μ and e-enriched samples, including $D \to K^0_L X$ decays, are determined to be 1.30±0.02, 1.35±0.01, and 1.38±0.01, respectively. Due to the observed discrepancies in the correction factors among the samples, a systematic uncertainty of 10% is assigned, yielding a correction of 1.30±0.10 to the nominal sample. The middle and right panels of the figure present the fit results of $q^2_{\rm rec}$ and $\eta(\mathrm{BDT}_2)$ in π-enriched sample, respectively, illustrating that the data-to-simulation ratio is much closer to unity after applying the fit. The $q^2_{\rm rec}$ distribution with this 1.3 correction applied to $B \to D(\to K^0_L X)$ in μ/e-enriched sample is shown as validation in Appendix 9.2 . The normalization of the B-meson sample without D$\to K^0_L$ decays agrees with expectations within 1.2 σ. The continuum background normalization is reduced by (19±2)%, aligning it more closely with the simulation. This difference between off- and on-resonance normalizations is used to assign the systematic uncertainty on the off-resonance scale in the nominal analysis.
To further assess the robustness of this study, an investigation in the nominal sample within the sideband region $0.75 < \eta(\mathrm{BDT}_2) < 0.90$ is performed, shown in Fig. 4.24. To illustrate the impact of the $D \to K^0_L$

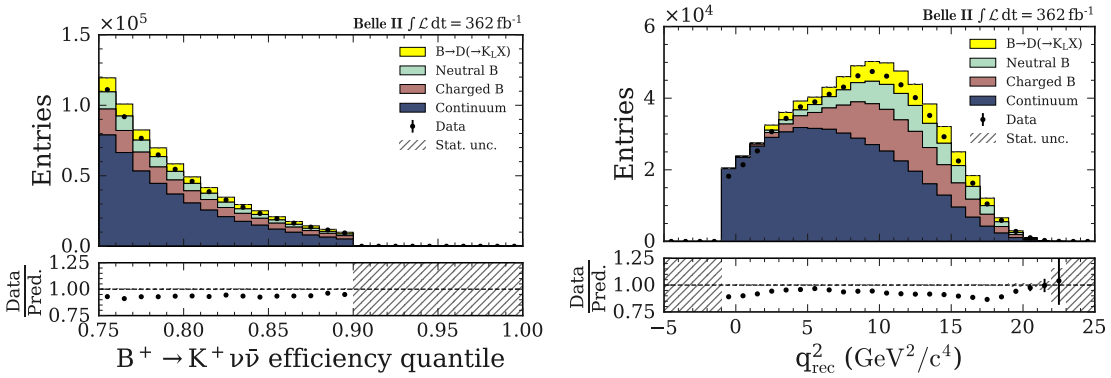

Figure 4.24: Background composition in the BDT_2 sideband region, of $\eta(\mathrm{BDT}_2)$ and q^2_{rec} with $B \to D(\to K^0_L X)$ events separated (yellow) and scaled by a factor of 1.3 for visibility. The total $\Upsilon(4S)$ data luminosity is used.

contribution, the B^+B^- and $B^0\bar{B}^0$ samples are further subdivided into components with and without this decay. The $D \to K^0_L$ component is scaled by a factor of 1.3 to enhance its visibility, and is represented by the yellow histogram in Fig. 4.24. The remaining B^+B^- and $B^0\bar{B}^0$ backgrounds are shown after explicitly removing decays involving $D \to K^0_L$.

4.4.2.2 Modeling of the $B^+ \to K^+K^0\bar{K}^0$ Background

The decay $B^+ \to K^+K^0\bar{K}^0$ constitutes a non-negligible background due to one or both neutral kaons may escape detection, thereby mimicking the signal topology. In Belle II simulations, these decays are modeled either according to phase space or as independent resonant contributions in Belle II simulations, which does not fully reproduce the true dynamics.
The $B^+ \to K^+K^0\bar{K}^0$ decay proceeds three isospin-related modes: $K^0_SK^0_S$, $K^0_LK^0_L$, and $K^0_SK^0_L$. Although amplitude analyses from Babar [98] provide detailed Dalitz structures, but cannot be used directly, because they incorporate reconstruction efficiencies and detector effects specific to Babar, which are not directly transferable to Belle II simulation.
Instead, the $B^+ \to K^+K^0_SK^0_S$ mode is reconstructed directly in data. The corresponding phase-space (PHSP) simulation of $B^+ \to K^+K^0_SK^0_S$ is generated for both signal and background, reflecting the uniform phase-space distribution. To obtain a realistic event distribution, an `sPlot` fit is performed on the reconstructed data to separate signal from background, producing the signal-only $M(K^0_SK^0_S)$ distribution. This distribution is then used as an event weight to correct the PHSP simulation of $B^+ \to K^+K^0_SK^0_S$ and its isospin partner $B^+ \to K^+K^0_LK^0_L$, thereby incorporating the resonant and non-resonant structures observed in data. Fig 4.25(left) shows the invariant mass distribution of $K^0_SK^0_S$ of simulated $B^+ \to K^+K^0_SK^0_S$ and background-subtracted data.
The $B^+ \to K^+K^0_SK^0_L$ has not been directly measured. Its modeling combines resonant and non-resonant contributions to reproduce the expected invariant-mass and angular distributions.

- The resonant p-wave contribution arises from $B^+ \to K^+\phi$, with $\phi \to K^0_SK^0_L$. This component produces the expected narrow peak in the $K^0_SK^0_L$ invariant mass spectrum.
- The non-resonant p-wave contribution is modeled using the isospin-related decay $B^0 \to K^0_SK^+K^-$. Assuming isospin symmetry, this channel provides the shape of the orbital angular momentum equal to 1 component between the K^+ and the $K^0_SK^0_L$ system, with a correction applied for the ratio of B^0 and B^+ lifetimes.

These components are combined in simulation, applied as a weight to correct the phase-space generated events. The modeling is validated using the isospin-related decay $B^0 \to K^0_SK^+K^-$, comparing the observed

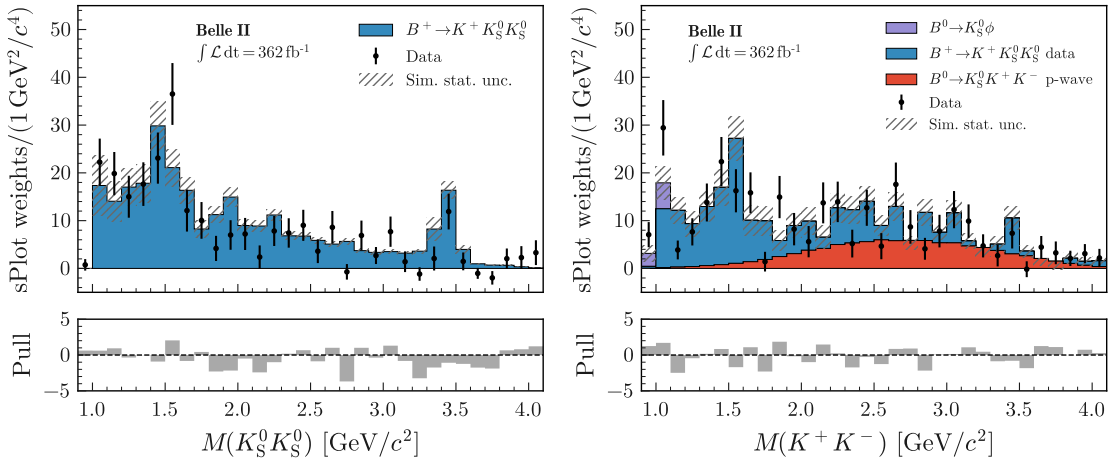

Figure 4.25: Validation of $B^+ \to K^+ K^0 \bar{K}^0$ modeling. Left: $M(K_S^0 K_S^0)$ distribution for $B^+ \to K^+ K_S^0 K_S^0$ in data (background-subtracted) and phase-space simulation after reweighting. Right: $M(K^+ K^-)$ distribution for the isospin-related decay $B^0 \to K_S^0 K^+ K^-$ compared with the modeled contributions (s-wave, resonant p-wave, and non-resonant p-wave). The total $\Upsilon(4S)$ data luminosity is used.

$K^+ K^-$ invariant mass spectrum to the sum of modeled s-wave, resonant p-wave, and non-resonant p-wave contributions. This comparison demonstrates satisfactory agreement in both shape and normalization, supporting the conclusion that the simulation adequately describes the $B^+ \to K^+ K_S^0 K_L^0$ background. Fig. 4.25 (right) shows the invariant mass distribution of the $K^+ K^-$.

4.4.2.3 Modeling of $B^+ \to K^+ n\bar{n}$

The $B^+ \to K^+ n\bar{n}$ background, although rare, is non-negligible because of an observed threshold enhancement in $B^{+/0} (\to K^{(*)+/0} p\bar{p}$ decays [99, 100]. If modeled according to pure phase space, this contribution would amount to only about 0.2% of the total B background in the signal region. However, when the threshold enhancement is included, the fraction increases to 0.4% in the signal region and 1.0% in the most sensitive region, making it relevant for the analysis. To account for this, the $B^+ \to K^+ n\bar{n}$ component is excluded from the generic background simulation and instead generated separately using phase-space decays. These events are then reweighted using the measured $B^0 \to K^0 p\bar{p}$ spectrum [100], which exhibits the threshold enhancement, under the assumption that a similar threshold effect applies to the $n\bar{n}$ final state. This procedure ensures that the simulated $B^+ \to K^+ n\bar{n}$ background reproduces both the normalization and the shape expected in data.

4.4.2.4 Modeling of $B^0 \to K^{*0} \nu\bar{\nu}$ and $B^+ \to \tau^+ \nu_\tau, \tau^+ \to K^+ \bar{\nu}_\tau$ Backgrounds

The $B \to K^* \nu\bar{\nu}$ background arises when the pion from $K^* \to K\pi$ is not reconstructed, causing the event to mimic the signal. At the time of this analysis, no dedicated $B \to K^* \nu\bar{\nu}$ form-factor predictions were available. Therefore, we adopted the same form-factor model as for $B^+ \to K^+ \nu\bar{\nu}$ [37], and explicitly includes this contribution in the background composition. It represents a small but structurally important background, since its kinematics overlap strongly with the signal. The yield is suppressed by a factor of five relative to the signal, and by about a factor of ten in the most sensitive region.
The long-distance contribution from $B^+ \to \tau^+ \nu_\tau, \tau^+ \to K^+ \bar{\nu}_\tau$ produces the same visible final state as the signal: a charged kaon accompanied by missing energy from neutrinos. This tree-level decay is treated as a background. These events are simulated separately and normalized to branching fraction $\mathcal{B}(B^+ \to \tau^+ \nu_\tau, \tau^+ \to K^+ \bar{\nu}_\tau) = (0.61 \pm 0.06) \times 10^{-6}$ [37]. The selection efficiency for this background (∼9.7%) is slightly higher than that for the signal (8.0%). Despite this higher efficiency, the overall contribution is small

due to the suppressed branching fractions. The expected yield is about six times smaller than the signal in the most sensitive region.

4.5 Signal Extraction Setup

Signal yields are extracted with a binned maximum-likelihood fit to the observed event counts in a two-dimensional space defined by the variables $\eta(\mathrm{BDT}_2)$ and q^2_{rec}. The data are partitioned into four bins in $\eta(\mathrm{BDT}_2)$ and three bins in q^2_{rec}, yielding 12 signal-region search bins. Both $\Upsilon(4S)$ and off-resonance samples are employed, resulting in 2×12 bins included in the simultaneous fit. For the $\Upsilon(4S)$ simulation, the sample corresponds to 0.8 ab^{-1} of continuum events and 2.8 ab^{-1} of $B\bar{B}$ events, taken from datasets orthogonal to those used in the classifier training. These events are weighted to match an integrated luminosity of data, 362 fb^{-1}. For the off-resonance background, an independent simulated continuum sample, corresponding to an equivalent integrated luminosity of 200 fb^{-1}, is employed. This sample is weighted to match the integrated luminosity of the off-resonance data, 42.3 fb^{-1}.

A common set of nuisance parameters $\mu_{q\bar{q}}$ with $q = u, d, s, c$ and $\mu_{\tau^+\tau^-}$ governs the continuum background normalization across all 24 bins, such that the off-resonance data provide direct constraints on the continuum component in the on-resonance signal region. Templates are constructed to approximate the distributions of the relevant observables for each class of event, and the likelihood function is defined as the product of Poisson probability densities corresponding to the event counts in the signal-region bins.

The degrees of freedom of this model include the signal strength μ, defined as the signal branching fraction relative to its SM expectation, $\mu = 1$. The SM reference value for the branching fraction is taken as 4.97×10^{-6} [37]. Systematic effects and finite simulation statistics are incorporated as nuisance parameters, implemented as either multiplicative or additive modifiers of the expected bin yields. These are constrained by auxiliary information via Gaussian priors.

The statistical analysis is primarily performed using the pyhf framework [101], which constructs the likelihood as a product of Poisson densities. As a cross-check, fits are also carried out with the sghf program, which approximates the Poisson densities by Gaussian functions and is additionally employed in control-sample studies. While the pyhf implementation is used to obtain the final results, the sghf model, being computationally less demanding, is adopted for large-scale validation tests.

The expected signal and background contributions in the signal search region are evaluated using the simulated samples described above, which have been properly weighted to correspond to an integrated luminosity of 362 fb^{-1}. The two-dimensional signal region plot, shown in Fig. 4.26 (top), illustrates the expected simulation of signal and background in binning of $\eta(\mathrm{BDT}_2) \times \mathrm{q}^2_{\mathrm{rec}}$, in projections of the signal region onto $\eta(\mathrm{BDT}_2)$(bottom left) and q^2_{rec} (bottom right). Along the $\eta(\mathrm{BDT}_2)$ axis (upper axis), the events are divided into four $\eta(\mathrm{BDT}_2)$ ranges. The first range, defined by 0.92<$\eta(\mathrm{BDT}_2)$<0.94, serves as a control region and is dominated by background, in particular B^+B^- events. The highest $\eta(\mathrm{BDT}_2)$ range corresponds to the most sensitive signal region, where the background contribution is minimal. Each $\eta(\mathrm{BDT}_2)$ range is further subdivided into three bins in q^2_{rec} (bottom axis). Within each $\eta(\mathrm{BDT}_2)$ range, the first q^2_{rec} bins are predominantly populated by continuum events, whereas the last q^2_{rec} bins are dominated by B^+B^- background. Overall, the background decreases progressively from the control region to the most sensitive signal region, while both signal and background distributions remain relatively smooth across q^2_{rec}.

The comparison between off-resonance data and reweighted continuum simulation, shown in the two-dimensional signal-region binning of $\eta(\mathrm{BDT}_2) \times \mathrm{q}^2_{\mathrm{rec}}$, shown in Fig. 4.26 (upper right), provides a direct validation of the $q\bar{q}$ background modeling where simulation is scaled up by a factor of 1.40 ± 0.05 to correct for the observed normalization discrepancy.

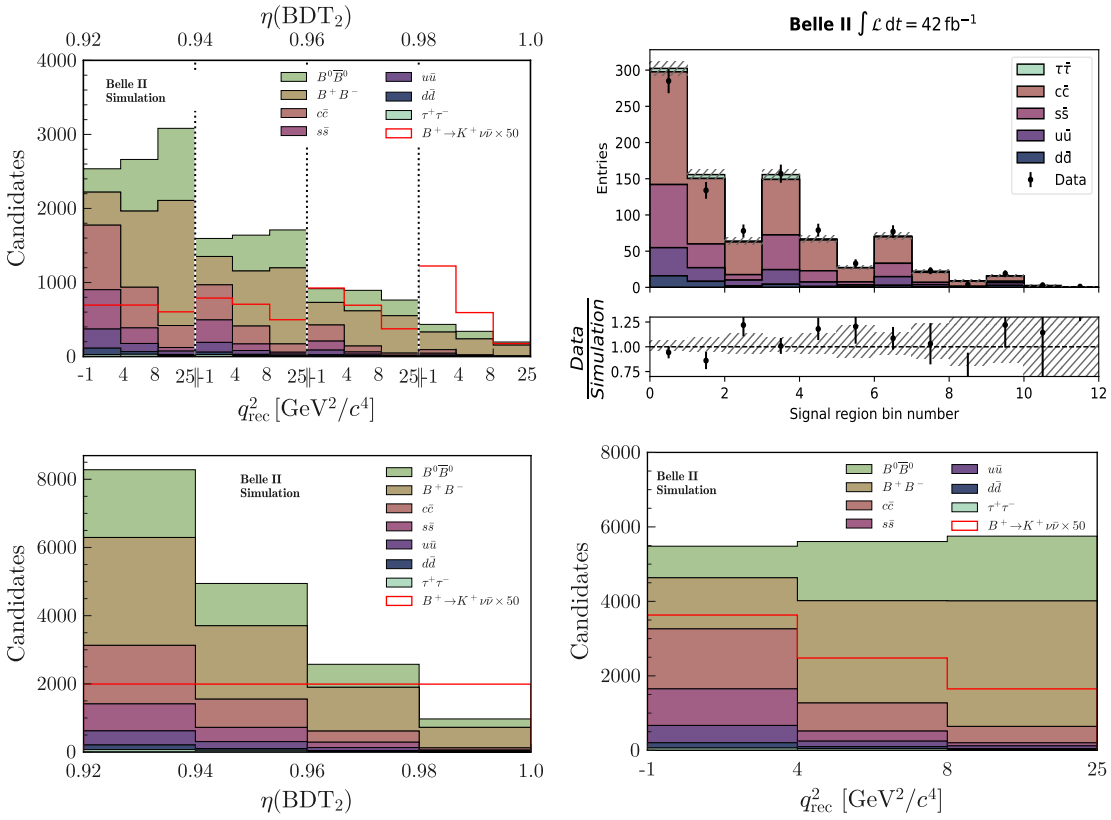

Figure 4.26: Expected background and signal in signal-region search bins of the $B^+ \to K^+\nu\bar{\nu}$. Top left/right: two-dimensional binning in $\eta(\mathrm{BDT}_2)\times q^2_{\mathrm{rec}}$ of $\Upsilon(4S)$/off-resonance simulation. Bottom left/right: projections onto $\eta(\mathrm{BDT}_2)$ and q^2_{rec}. Off-resonance data (black points) are compared to reweighted continuum simulation (colored histograms), scaled by 1.40 ± 0.05 to match the data yield.

4.6 Systematics

The validation studies performed in both the control sample and the sideband region demonstrate a satisfactory level of consistency between the observed data and the simulated expectations, thereby supporting the reliability of the background modeling. On this basis, the signal region has been precisely defined, together with the parameter of interest, μ, and the statistical framework of the fitting strategy has been established. The subsequent step of the analysis is devoted to a systematic evaluation of the uncertainties that may influence the determination of the signal branching fraction.
Systematic uncertainties enter the statistical model through dedicated nuisance parameters, which modify the nominal expectations in the bins of the signal search region. Two main categories of nuisance parameters are employed. The first corresponds to normalization variations. These parameters are constrained by Gaussian priors centered at unity, with widths determined from auxiliary inputs, thereby allowing controlled deviations from the nominal yields. The second category accounts for shape-related variations, and each variation enters the likelihood together with a variation vector, which encodes correlations among bins and across samples. Their priors are standard Gaussian distributions centered at zero, such that a unit variation corresponds to the application of a single variation vector.

4.6.1 External Input Uncertainties

The systematic uncertainties considered in this analysis are grouped into three broad categories: those related to physics modeling, those arising from detector response, and those associated with external inputs and global normalization factors. This classification provides a structured framework for evaluating their impact on the signal extraction. In the following, we begin with the discussion of the third category, which includes uncertainties from luminosity, the number of $B\bar{B}$ pairs, finite signal simulation statistics, and external branching-fraction inputs.

- Integrated luminosity. The on- and off-resonance integrated luminosities are taken from Belle II measurements [102], with an uncertainty of approximately 1%. This directly scales the expected number of events and impacts both signal and background yields proportionally.
- Number of $B\bar{B}$ pairs. Determined by the Belle II B-counting group [103], with a 1.5% uncertainty. This uncertainty affects the calculation of the expected signal events.
- Simulated sample size. The finite size of the simulated samples introduces bin-by-bin statistical uncertainties on the expected yields. In the pyhf likelihood framework, each bin is treated as a Poisson-distributed observation, and the uncertainties $\mathcal{O}(1\%)$ from limited simulation statistics are implemented as nuisance parameters.
- Normalization of continuum background of $\Upsilon(4S)$ events. The reweighting of continuum simulation to off-resonance data is assigned a conservative 50% uncertainty per component, based on observed discrepancies in the signal-region bins shown in Fig. 4.26 (upper right).
- Normalization of off-resonance background. An additional 5% uncertainty is applied to account for potential differences in continuum composition between on- and off-resonance conditions, motivated by Tab. 4.9. This is implemented in the likelihood model through five independent normalization nuisance parameters, $\mu_{u\bar{u}}$, $\mu_{d\bar{d}}$, $\mu_{c\bar{c}}$, $\mu_{s\bar{s}}$ and $\mu_{\tau^+\tau^-}$, each constrained by a Gaussian prior centered at 1.0 with a width of 0.5.
- Normalization of $B\bar{B}$ background. A similarly conservative 50% uncertainty is applied to the $B\bar{B}$ events to cover the normalization observed in Fig 4.20, Fig. 4.24 and Fig. **??** in sideband. Two nuisance parameters, $\mu_{B^+B^-}$ and $\mu_{B^0\bar{B}^0}$ are introduced.
- Branching fractions of leading B^+ and B^0 in signal region. The branching fractions of the decay modes that account for approximately 80% of B^+B^- decays and 70% of $B^0\bar{B}^0$ decays in the signal region, as shown in pie chart 4.22, are varied within their reported uncertainties [104] by $\mathcal{O}(1\%)$. This

treatment reflects the fact that the branching fractions used in the simulation correspond to central values, and the associated uncertainties must be propagated to assess their impact on the background expectations.

- Branching fraction for $B \to D^{**}$. The fraction of signal candidates containing excited charm mesons (D^{**}) is found to be about 3% for charged-B backgrounds and 5% for neutral-B backgrounds. To account for potential mis-modeling, a conservative 50% systematic uncertainty is assigned to their branching fractions, resulting in relative uncertainties below 5% in all bins of the signal region.

The finite simulated sample size does not correspond to the statistical uncertainty of the measurement itself (which comes from the finite number of observed data events). Instead, it is a technical/statistical limitation of the simulation that introduces extra uncertainty into the expected yields used in the likelihood. Let n_i denote the observed data count in signal-region bin i and $N_{\mathrm{MC},i}$ the number of simulated events in the same bin, with expected yield $\lambda_i = w\, N_{\mathrm{MC},i}$, where w is the luminosity weight. The data counts follow a Poisson distribution

$$n_i \sim \mathrm{Pois}(\lambda_i). \tag{4.8}$$

Since $N_{\mathrm{MC},i}$ is finite, λ_i itself carries a statistical uncertainty, which is implemented in the likelihood via the Barlow–Beeston method by introducing an auxiliary Poisson term for the simulation. In the large-sample limit this reduces to a Gaussian uncertainty of relative size $1/\sqrt{N_{\mathrm{MC},i}}$. Frameworks such as `pyhf` include these terms so that finite-simulation statistical uncertainties are consistently propagated to the signal strength μ.
The uncertainties on the branching fractions of the dominant B-meson decay modes are propagated into the likelihood model through a replica-based reweighting procedure. A collection C of branching fractions and their uncertainties is constructed from [104]. For each event e in the signal search region, 1000 weights $\{w_1^e, \ldots, w_{1000}^e\}$ are generated according to

$$w_i^e = \max\left(1 + \frac{s_i^e}{\mathcal{B}(e)}, 0\right), \tag{4.9}$$

where $\mathcal{B}(e)$ is the branching fraction of the decay mode associated with event e, and s_i^e is sampled from a Gaussian distribution centered at zero with a width equal to the corresponding uncertainty. This procedure produces 1000 replicas of each simulated sample, from which pseudo-observations are constructed by summing weights in each signal-region bin.
A covariance matrix Σ_{ij} is then computed across bins and samples, and its five leading eigenvectors are retained as variation vectors. Each variation vector enters the likelihood model through a dedicated nuisance parameter $\theta_k^{(\mathcal{B})}$ ($k = 1, \ldots, 5$). Residual terms from the eigen-decomposition are added in quadrature to uncorrelated uncertainties. The resulting relative uncertainties are typically of order $\mathcal{O}(1\%)$.
Tab. 4.11 summarizes the sources and size of uncertainty of the above external-input uncertainties.

4.6.2 Detector Response Uncertainties

Detector response uncertainties account for the possible mis-modeling of the reconstruction and identification performance of the Belle II detector. These include the efficiency of track finding for all charged particles, the particle identification of signal K^+, the modeling of the detector's energy response to photons and hadrons, and the efficiency for reconstructing K_L^0 mesons in ECL. Each of these sources, as discussed in Sec.4.1.4 is evaluated to ensure that differences between data and simulation are properly taken into account.

- Track finding efficiency. A systematic uncertainty of 0.3% per track is recommended by the Belle II tracking group, and we conservatively assign 0.9% per track to propagate potential differences between data and simulation, as discussed in Sec. 4.1.4. Simulated signal and background samples are reconstructed both normally and with tracks randomly removed with a per-track probability of 0.9%, and the event yield in the signal region are compared between the two cases.

Source	Size of uncertainty	Impact on σ_μ
Integrated luminosity	1%	<0.01
Number of $B\bar{B}$	1.5%	0.02
Simulated-sample size	$\mathcal{O}(1\%)$	0.52
Normalization of continuum background	50%	0.10
Off-resonance sample normalization	5%	0.05
Normalization of $B\bar{B}$ background	50%	0.90
Branching fraction for $B \to D^{**}$	50%	0.42
Leading B-decays branching fractions	$\mathcal{O}(1\%)$	0.22

Table 4.11: Summary of external-input systematic uncertainties for the $B^+ \to K^+\nu\bar{\nu}$ analysis. The table lists the relative size of each uncertainty, expressed as a percentage where applicable, together with its impact on the uncertainty of the fitted signal strength parameter μ (denoted σ_μ).

- Signal kaon PID. A requirement of kaonID > 0.9 selects the signal kaon candidate. Simulation is weighted to correct for PID efficiency differences between data and MC, using calibration weights in $(p_T, \cos\theta)$. Associated uncertainties (≈1% for efficiencies, ≈10% for mis-ID rates) are propagated as systematics. Lepton-to-kaon mis-ID corrections are included but found to be negligible.

- Photon energy. For the calorimeter photon detection efficiency, studies indicate an uncertainty of order 1%, but following the Belle II neutral group's recommendation, a 0.5% systematic uncertainty is assigned by randomly modifying 0.5% of photon clusters in the ECL.

- Hadronic energy. A study of the control channel $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ indicates a possible 5% upward bias in the neutral hadron energy scale, while the nominal simulation applies a 10% downward correction to unmatched photons to improve data–simulation consistency. To conservatively cover this effect, a 10% systematic uncertainty (corresponding to 100% relative to the nominal correction) is assigned.

- K_L^0 efficiency in ECL. The K_L^0 reconstruction efficiency in simulation is corrected by 17% based on dedicated studies. A relative uncertainty of half this correction, 8.5%, is assigned as a systematic.

The methodology for evaluating the tracking and PID efficiency uncertainties and their propagation to the binned maximum likelihood fit is documented in Ref. [105]. Apart from PID, which is described by seven nuisance parameters, each remaining source of systematic uncertainty is represented by a single nuisance parameter in the likelihood fit. Tab. 4.12 summarizes the sources and size of uncertainty of the above detector-response uncertainties.

Source	Size of uncertainty	Impact on σ_μ
Track finding efficiency	0.9%	0.20
Signal kaon PID	$\mathcal{O}(1\%)$	0.07
Photon energy	0.5%	0.08
Hadronic energy	10%	0.37
K_L^0 efficiency in ECL	8.5%	0.22

Table 4.12: Summary of detector-response systematic uncertainties for the $B^+ \to K^+\nu\bar{\nu}$ analysis. The relative size of each uncertainty and its impact on the uncertainty of the fitted signal strength σ_μ are shown.

4.6.3 Physics Modeling Uncertainties

Physics-modeling uncertainties arise from imperfect knowledge of signal and background processes, including the SM signal form factors, branching fractions for decays such as $B^+ \to K^+ K_L^0 K_L^0$, $B^+ \to K^+ n\bar{n}$, and $D \to K_L X$, as well as the p-wave component in $B^+ \to K^+ K_S^0 K_L^0$. Additional uncertainties stem from continuum background modeling, the $\mathrm{BDT_C}$ classifier, and the global signal efficiency. These sources reflect either improvements in decay modeling specific to this analysis or uncertainties in the modeling of signal kinematics.

- Signal SM form-factors. The signal form-factor uncertainty is evaluated by reweighting simulated events using eigenvector variations of the form-factor covariance matrix, propagated via SVD decomposition. The induced variations in the q^2 spectrum are at the level of about 2% relative to the nominal form-factor model, and this residual is taken as the systematic uncertainty.

- Global signal efficiency. The signal efficiency validation (Sec. 4.1.4) indicates a 3% difference between data and simulation. To cover this, a nuisance parameter allowing $\pm 3\%$ variations is included, applied coherently across all bins.

- Branching fraction for $B^+ \to K^+ K_L^0 K_L^0$. The branching fraction of $B^+ \to K^+ K_L^0 K_L^0$ is estimated using the same weight as $B^+ \to K^+ K_S^0 K_S^0$, whose measured branching ratio is about 6%. Data-driven cross-checks constrain this procedure to about 10% precision. The dominant additional uncertainty arises from the assumption of isospin symmetry between the $K_S K_S$ and $K_L K_L$ final states, with studies indicating that any isospin-breaking effects are small. Since the uncertainties on $B^+ \to K^+ K_S^0 K_S^0$ and $B^+ \to K^+ K_L^0 K_L^0$ are correlated, a single 20% systematic uncertainty is conservatively assigned to cover both modes, which is implemented as a single nuisance parameter in the fit.

- Branching fraction for $B^+ \to K^+ K_S^0 K_L^0$. The branching fraction of $B^+ \to K^+ K_S^0 K_L^0$ is estimated from $B^0 \to K_S^0 K^+ K^-$ decays using isospin arguments. The $B^0 \to K_S^0 K^+ K^-$ channel includes three main components: (i) a non-resonant p-wave contribution, and (ii) the resonant p-wave contribution from $B \to K\phi$. The branching fraction of $B \to K\phi$ is known with a precision of about 10% ([4]). To account for possible isospin-breaking effects (20%) and the uncertainties in modeling the non-resonant p-wave component (20%), an overall 30% systematic uncertainty is assigned to the normalization of the $B^+ \to K^+ K_S^0 K_L^0$ branching fraction.

- Branching fraction of $B^+ \to K^+ n\bar{n}$. A systematic uncertainty is included in the analysis for the potential $B^+ \to K^+ n\bar{n}$ background, based on 100% of the exclusion of this mode. This covers several sources: (i) imperfections in modeling the threshold enhancement, (ii) the assumption of isospin symmetry relating $B^+ \to K^+ n\bar{n}$ to $B^+ \to K^+ p\bar{p}$, (iii) modeling of additional baryonic channels in the final state, and (iv) uncertainties in the anti-neutron response in the ECL.

- Branching fraction for $D \to K_L^0 X$. The fraction of D-meson decays involving K_L^0 mesons is corrected by 30%, based on scaling factors determined from several samples (1.3, 1.35, 1.38). To account for the observed variations between samples and limited statistics, a conservative 10% absolute systematic uncertainty is assigned to this correction.

- Continuum background modeling, $\mathrm{BDT_C}$. The effect of the $\mathrm{BDT_C}$ reweighting is used as a 100% shape systematic, meaning that the full difference between the distributions before and after $\mathrm{BDT_C}$ reweighting is assigned as the shape uncertainty.

Tab 4.13 summarizes the sources and size of uncertainty of the above physics modeling uncertainties.
To estimate the impact of a given systematic source, we modify the priors of the corresponding nuisance parameters (setting them very large, for instance, 10000). The change in the uncertainty on μ relative to the nominal fit gives the effective contribution of that systematic, disentangling correlated and uncorrelated effects. By repeating this procedure for groups of nuisance parameters, we can build a ranking of systematics by their influence on the signal strength.

Source	Size of uncertainty	Impact on σ_μ
Signal SM form-factors	$\mathcal{O}(1\%)$	0.02
Global signal efficiency	3%	0.03
Branching fraction for $B^+ \to K^+ K_L^0 K_L^0$	20%	0.49
Branching fraction for $B^+ \to K^+ K_S^0 K_L^0$	30%	0.02
Branching fraction for $B^+ \to n\bar{n}K^+$	100%	0.20
Branching fraction for $D \to K_L^0 X$	10%	0.14
Continuum background modeling, $\mathrm{BDT_C}$	100% of corrections	0.01

Table 4.13: Summary of physics-modeling systematic uncertainties for the $B^+ \to K^+\nu\bar{\nu}$ analysis. The relative size of each assigned uncertainty and its impact on the fitted signal strength uncertainty σ_μ are shown.

The dominant contributions are the $B\bar{B}$ background normalization (0.90), MC statistical precision (0.52), the branching fraction uncertainty for $B^+ \to K^+ K_L^0 K_L^0$ (0.49) and the $B \to D^{**}$ branching fractions (0.42). Several calibrations (hadronic energy, K_L efficiency) and tracking also gives sizable contributions. Some very large assigned uncertainties (e.g. 100% continuum shape correction) have little impact on σ_μ because their effect is orthogonal to the signal-sensitive bins and is constrained by the fit.

4.7 Results

4.7.1 Pre-unblinding Checks

Before unblinding the signal region, a series of validation studies are performed to test the stability of the fit and the robustness of the background modeling.

4.7.1.1 Fit Stability Study

To evaluate the robustness of the binned likelihood fit before unblinding, toy datasets were generated for each bin with Poisson fluctuations and systematic shifts from the covariance matrix. Each toy was fitted independently to extract the signal strength μ. The distribution of fitted μ quantifies the expected spread due to statistical and systematic uncertainties, while the linearized uncertainty provides a cross-check. The distribution of $\mu_{\text{toy}} - \mu_{\text{central}}$ is used to test fit robustness. The observed data are used as central values without evaluating the overall fit quality, thereby preserving blindness while testing robustness. A mean near zero indicates an unbiased fit, and the width reflects the expected pre-unblinding uncertainty, shown in Fig. 4.27.

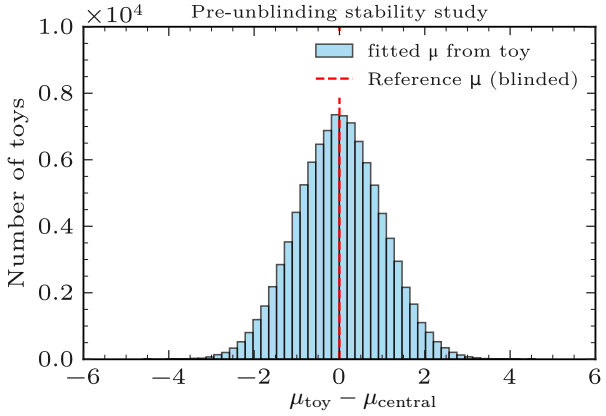

Figure 4.27: Distribution of fitted signal strengths μ from the toy stability study. The width reflects the expected spread from statistical and systematic fluctuations, and the mean is consistent with zero bias.

4.7.1.2 Signal Injection Study

A signal injection study was performed to validate the maximum likelihood fit procedure and the estimation of the signal strength μ. Known signal strengths $\mu_{\text{injected}} = 1, 5, 20$ were injected, and the fit response was evaluated under statistical and systematic fluctuations , as shown in Fig.4.28.
For each injected μ:

a. **Toy generation:** 10,000 pseudo-datasets (toys) are generated by fluctuating the expected signal plus background yields. Poisson fluctuations account for statistical uncertainties, while Gaussian fluctuations (if enabled) simulate correlated and uncorrelated systematic effects.

b. **Fit to toys:** Each toy dataset is refitted using the same maximum likelihood fitter as in the nominal analysis. The fitted signal strength μ_{fit} and its uncertainty σ_{fit} are recorded.

c. **Analysis of fitted results:**

- The mean of μ_{fit} across toys is compared with μ_{injected} to check for bias.
- The spread of μ_{fit} is compared to the fitted uncertainties σ_{fit} to validate their accuracy.

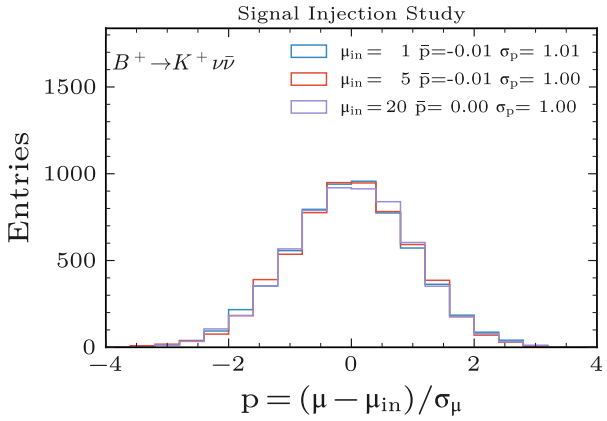

Figure 4.28: Validation of the fitter using injected signals $\mu = 1, 5, 20$. The fitted values are consistent with the injections within uncertainties, and the pull distributions follow the expected standard normal behavior.

- Pulls are computed as

$$\text{pull} = \frac{\mu_{\text{fit}} - \mu_{\text{injected}}}{\sigma_{\text{fit}}},$$

  which should follow a standard normal distribution if the fit is unbiased and uncertainties are correctly estimated.

This study tests the accuracy and reliability of the fit, ensuring that the measured μ is unbiased and that the uncertainties are properly estimated. It also checks the stability of the fitter across a range of signal strengths from near the expected sensitivity ($\mu = 1$) to high signal scenarios ($\mu = 20$).

4.7.1.3 Background Normalization Test

In the likelihood fit, each background process carries a normalization uncertainty — here set to 50% pre-fit. These nuisance parameters are among the dominant sources of systematic uncertainty. To test their behavior before unblinding, we generated 100,000 toy datasets with Poisson fluctuations per bin. Each toy was fitted with the full likelihood, and the pulls of the background normalizations were recorded with respect to their priors, which are Gaussian-constrained with 50% width. This study verifies whether the fit tends to drive the normalizations systematically away from their nominal values or whether they remain well-constrained within their priors, ensuring that the statistical procedure does not artificially sculpt the background contributions.

4.7.2 Unblinded Fit Results

With the various pre-unblinding checks in place, the analysis strategy and background modeling were verified to be robust. Having established confidence in the control samples and the behavior of the multivariate discriminants, the signal region was then unblinded, and the full statistical fit was performed. In the following, I present the outcome of this fit, including the observed event yields, the extracted signal strength, and the assessment of compatibility between the fitted model and the data.
Fig. 4.29 shows the comparison between data and the fitted model in both the off-resonance control sample and the signal region. In each case, the fit describes the data well, indicating that the modeling of both background and potential signal contributions is reliable. In the signal region, the data exhibit an excess relative to the background expectation, consistent with the presence of the rare decay $B^+ \to K^+ \nu\bar{\nu}$.
The proportion of events identified as signal is relatively small in the full signal region, about 5%. However, in the three bins with the most signal-like classifier output, $\eta(\text{BDT}_2) > 0.98$, the signal purity rises significantly

to 19%. This demonstrates that the multivariate selection is effective at isolating signal-rich regions. The signal strength is extracted as

$$\mu = 5.4 \pm 1.0(\text{stat}) \pm 1.1(\text{syst}) = 5.4 \pm 1.5, \tag{4.10}$$

where the first uncertainty is statistical and the second reflects systematic contributions. The statistical uncertainty is obtained by generating ensembles of pseudo-experiments, where the number of events in each bin is fluctuated according to Poisson statistics. The total uncertainty on μ is determined using a profile likelihood scan. This evaluates the negative log-likelihood by fitting the model with μ fixed at values around its minimum. From this procedure, the systematic component of the uncertainty is then determined by subtracting the statistical part in quadrature. An additional 8% theoretical uncertainty, related to the external knowledge of the branching fraction normalization, is quoted separately and not included in the result above. Finally, the overall compatibility of the fitted model with the data is tested using pseudo-experiments. The corresponding p-value is found to be 47%, indicating no tension between the observations and the fitted expectations.

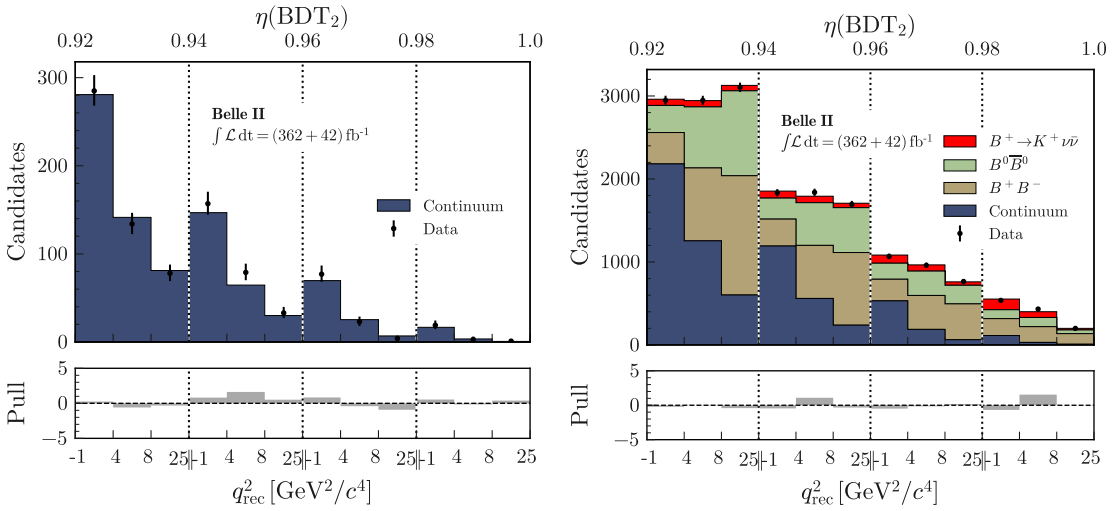

Figure 4.29: Observed yields and fit results in bins of the $\eta(\text{BDT}_2) \times \text{q}^2_{\text{rec}}$ space obtained by the simultaneous fit to the off- and on-resonance data, corresponding to an integrated luminosity of 42 fb^{-1} and 362 fb^{-1}, respectively. The yields are shown individually for the $B^+ \to K^+\nu\bar{\nu}$ signal, neutral and charged B-meson decays, and the sum of the five continuum categories. The yields are obtained in bins of the $\eta(\text{BDT}_2) \times \text{q}^2_{\text{rec}}$ space. The pull distributions are shown in the bottom panel.

To further validate the fit, it is instructive to compare data and simulation in the signal region and in the most sensitive region after applying the post-fit corrections. Fig. 4.31 and Fig. 4.32 displays several representative distributions for events in these regions. For simulated events, bin-by-bin weights are applied according to the ratio of post-fit to pre-fit yields, separated by signal region bin and event category, in order to reflect the outcome of the fit.
In the signal region, the reconstructed q^2_{rec} distributions are described reasonably well overall, but some local discrepancies remain. At very low q^2_{rec} (lower than 3 GeV^2/c^4), the simulation tends to predict more events than observed, while in the intermediate region ($3 < q^2_{\text{rec}} < 7$) the data yield is higher than the MC expectation. Smaller excesses of data are also visible around $9 < q^2_{\text{rec}} < 10$ and $12 < q^2_{\text{rec}} < 13$, whereas the simulation overshoots again in the adjacent windows $7 < q^2_{\text{rec}} < 9$ and $10 < q^2_{\text{rec}} < 12$. These patterns reflect fluctuations around the transition regions of the available phase space, where the efficiency and background composition vary most rapidly. In the most sensitive region, the agreement is even closer: only a modest deficit of data with respect to simulation is visible at very low q^2_{rec} (lower than 3 GeV^2/c^4) and in the range $3 < q^2_{\text{rec}} < 5$, while the remainder of the spectrum is well described.

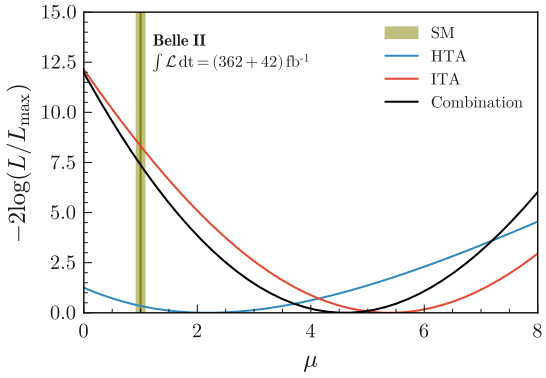

Figure 4.30: Likelihood scan of the signal μ. ITA denotes the inclusive tagging analysis (this analysis), HTA denotes the hadronic tagging analysis, and Combination represents the combined fit of ITA and HTA. The green band indicates the SM prediction.

Interpreting the fitted signal strength μ in terms of a branching fraction leads to

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (2.7 \pm 0.5_{\text{stat}} \pm 0.5_{\text{syst}}) \times 10^{-5}. \tag{4.11}$$

This measurement does not include the potential contribution from the long-distance double-charged-current process $B^+ \to \tau^+\nu_\tau$ followed by $\tau^+ \to K^+\bar{\nu}_\tau$.
The statistical significance of the observed excess is evaluated using the profile likelihood method, as shown in Fig. 4.30. For several fixed values of μ, the profile likelihood L is computed while allowing all nuisance parameters to float. The difference between $-2\ln L$ at $\mu = 0$ (background-only hypothesis) and its minimum value provides a test statistic whose square root gives the significance of the observed signal with respect to the background-only expectation. This procedure yields a significance of $3.5\ \sigma$, corresponding to strong evidence for the decay.
A similar calculation is performed to compare the observed signal yield to the SM prediction. In this case, the difference between $-2\ln L$ at $\mu = 1$ (SM normalization) and at the best-fit point is used. The resulting significance is $2.9\ \sigma$, suggesting that the measured branching fraction lies above the SM expectation, though the deviation is not yet sufficient to claim a discovery or a definitive inconsistency with the SM.
Several consistency checks were performed to test the robustness of the result. In split-sample studies, the data were divided into statistically independent subsets according to criteria such as data-taking period, event topology, and kaon properties, and the fit was repeated for each subset. All results were mutually consistent, with the largest deviation being 2.4σ for the total-charge split; this was investigated in detail and found compatible with statistical fluctuations ($\chi^2/\text{ndf} = 12.5/9$). Variations of the fit strategy (for instance, background constraints, restricted data ranges, or q^2 subdivisions) led to shifts in the signal strength μ of less than 0.2 and only minor changes in its uncertainty.
As an additional cross-check, the analysis was repeated for the control channel $B^+ \to \pi^+ K^0$, which closely resembles the signal topology. The extracted branching fraction, $(2.5 \pm 0.5) \times 10^{-5}$, agrees well with the world average $(2.34 \pm 0.08) \times 10^{-5}$, thereby confirming the reliability of the procedure.

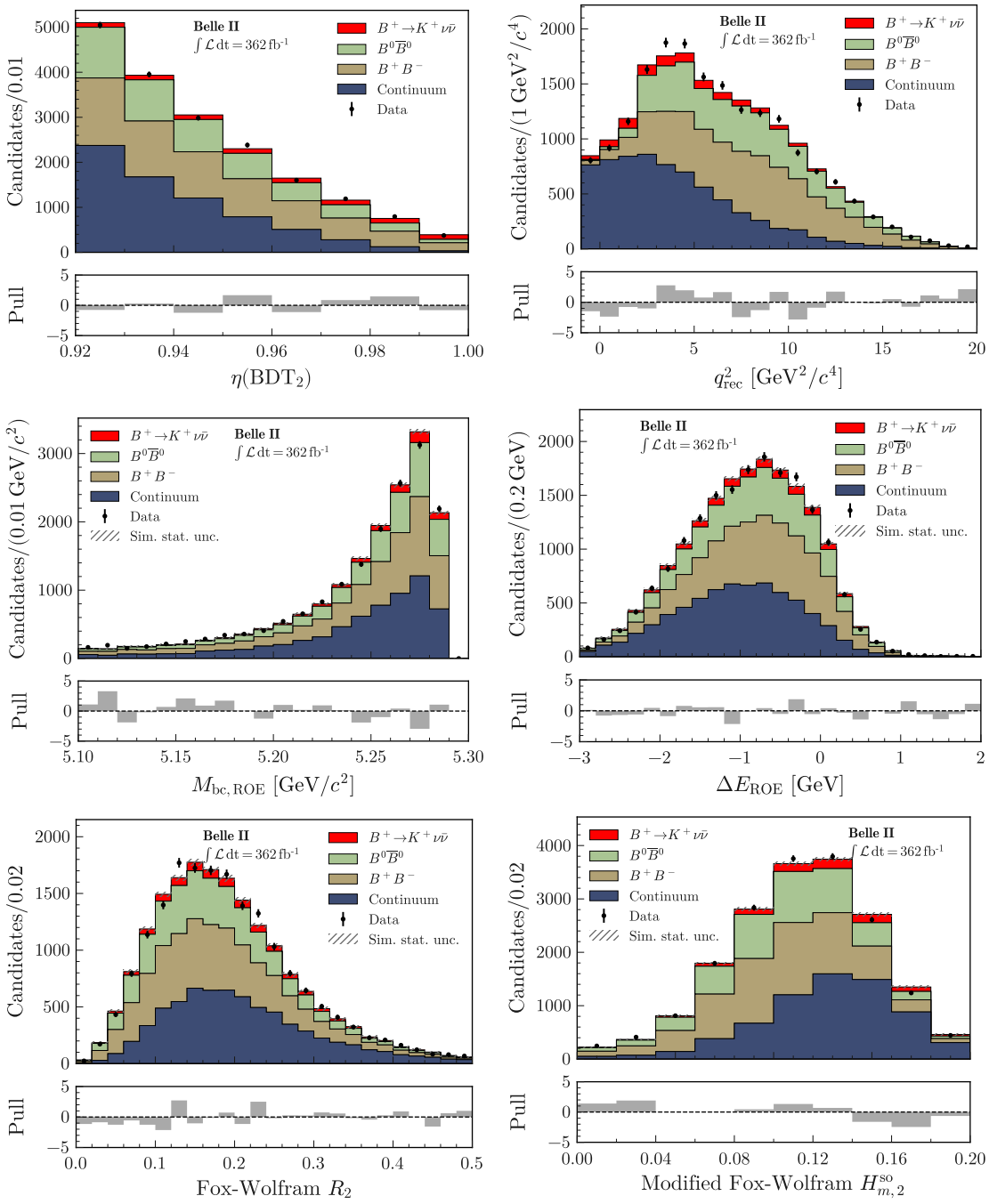

Figure 4.31: Distributions of $\eta(\text{BDT}_2)$, q^2_{rec}, beam-constrained mass of the ROE $M_{\text{bc,ROE}}$, ΔE_{ROE}, Fox-Wolfram R_2, and modified Fox-Wolfram $H^{so}_{m,2}$ in data (points with error bars) and simulation (filled histograms) shown individually for the $B^+ \to K^+\nu\bar{\nu}$ signal, neutral and charged B-meson decays, and the sum of the five continuum categories. Events in the full signal region, with $\eta(\text{BDT}_2) > 0.92$, are shown. Simulated samples are normalized according to the fit yields, and the pull distributions are shown in the bottom panels. The total $\Upsilon(4S)$ data luminosity is used.

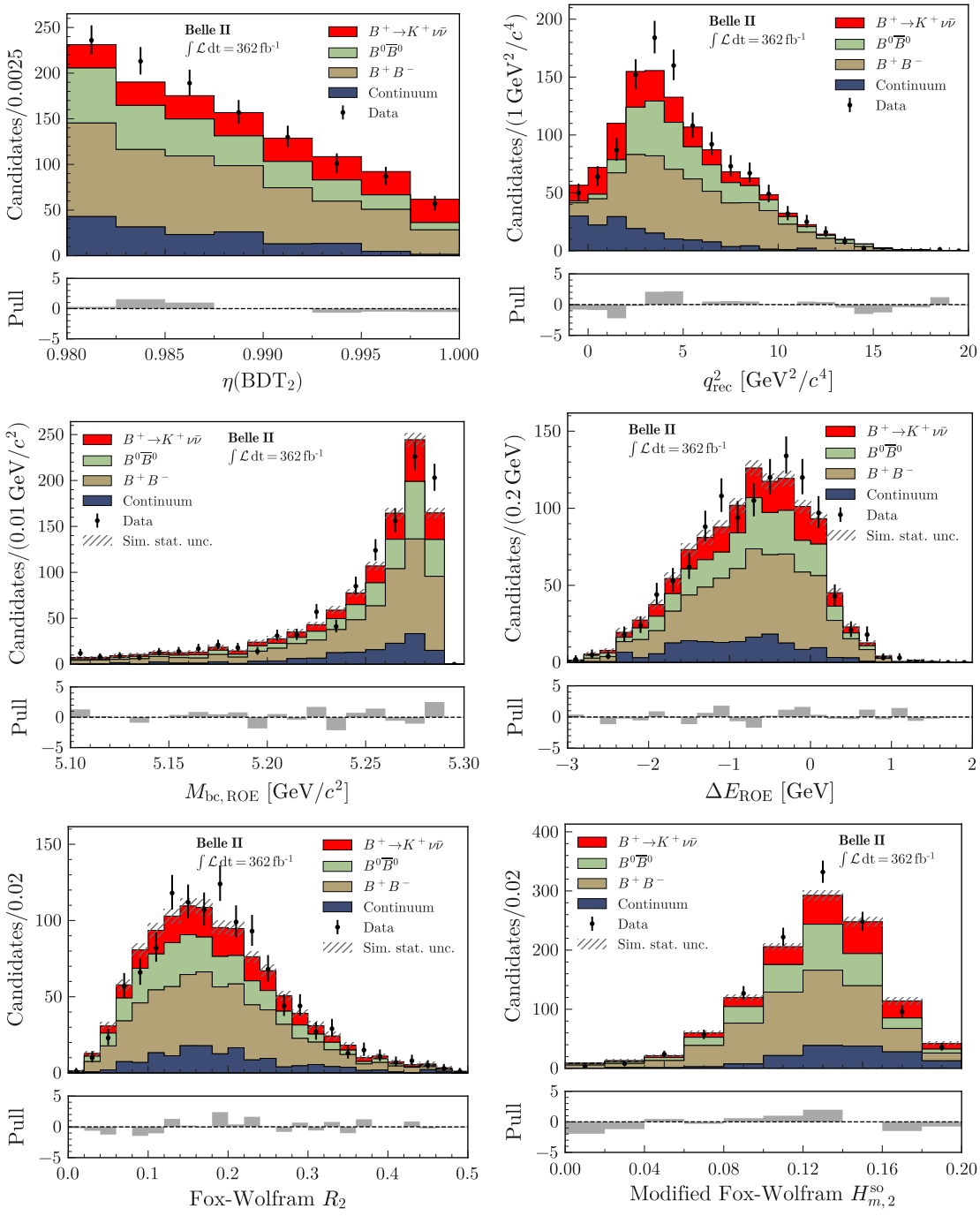

Figure 4.32: Distributions of $\eta(\text{BDT}_2)$, q^2_{rec}, beam-constrained mass of the ROE $M_{\text{bc,ROE}}$, ΔE_{ROE}, Fox-Wolfram R_2, and modified Fox-Wolfram $H^{so}_{m,2}$ in data (points with error bars) and simulation (filled histograms) shown individually for the $B^+ \to K^+\nu\bar{\nu}$ signal, neutral and charged B-meson decays, and the sum of the five continuum categories. Events in the full signal region, with $\eta(\text{BDT}_2) > 0.98$, are shown. Simulated samples are normalized according to the fit yields and the pull distributions are shown in the bottom panels. The total $\Upsilon(4S)$ data luminosity is used.

4.8 Discussion

Search for $B^+ \to K^+\nu\bar{\nu}$ presents a significant advancement in the study of $B^+ \to K^+\nu\bar{\nu}$, achieving a level of precision comparable to the current best results despite utilizing a smaller data sample. This analysis is further complemented by its combination with the HTA, which contributes additional independent branching fraction measurements to the comparison and enhances the overall statistical robustness of the study. To ensure a rigorous comparison, previous measurements were recalculated to account for variations in $f^{+/-}$ assumptions and reporting formats as following:

- $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (1.5^{+1.7+0.4}_{-0.8-0.2}) \times 10^{-5}$ [60].
- $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (0.2^{+0.8}_{-0.7}) \times 10^{-5}$ [61].
- $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = \frac{N_{\text{sig}}}{\epsilon N_{B^+B^-}}$ where $N_{B^+B^-}$ is calculated based on number of $B\bar{B}$ in the paper [58] with the charged production fraction $f^{+-} = 0.516$.
- $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = \frac{N_{\text{sig}}}{\epsilon N_{B^+B^-}}$ where $N_{B^+B^-}$ is calculated based on number of $B\bar{B}$ in the paper [59] with the charged production fraction $f^{+-} = 0.516$.
- $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (2.7 \pm 0.7) \times 10^{-5}$ [57].

While the results align well with previous hadronic and inclusive tagging measurements, the observed tensions with semileptonic tagging results from BABAR ($2.3\ \sigma$) and Belle ($1.8\ \sigma$) warrant further investigation into potential systematic effects inherent to tagging methodologies, as shown in Fig 4.33. The resulting simplified weighted average of $(1.3 \pm 0.4) \times 10^{-5}$, supported by a χ^2 p-value of 35%, indicates that the global ensemble of measurements remains statistically consistent. This suggests that the ITA method in this analysis is a highly efficient tool for probing rare decays in the flavor physics sector.

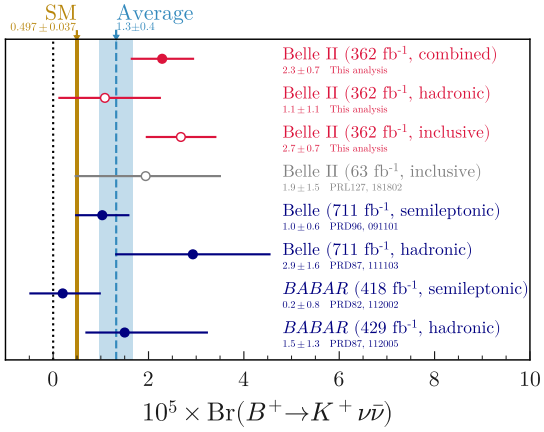

Figure 4.33: Branching fraction value of $B^+ \to K^+\nu\bar{\nu}$ from SM prediction [37] (golden-color area), previous measurements [57–61] (blue and grey dots with error bars), this analysis (ITA) (red circle with error bar) and combined measurements (red dot with error bar) with HTA (red circle with error bar). The average value (right blue area) is weighted without the superseded measurement of Belle II [57].

Chapter 5

Search for $B^{*0} \to K^{*0}\nu\bar{\nu}$

This chapter presents the search for the rare decay $B^0 \to K^{*0}\nu\bar{\nu}$ at Belle II, performed with the inclusive-tagging technique. The analysis strategy closely follows that of the $B^+ \to K^+\nu\bar{\nu}$ study, beginning with a description of the data and simulated samples, followed by data-simulation corrections and background suppression using multivariate classifiers. The signal region is defined from the classifier output, and control samples are employed to validate the signal selection efficiency. After selection, the residual backgrounds are studied and validated against sideband data. Systematic uncertainties on both efficiency and background estimates are then evaluated. In this channel, $B^+ \to K^+\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$, and $B^+ \to K^{*+}\nu\bar{\nu}$ are treated as cross-feed signals and are floated in the template fits to allow for a combined interpretation.

5.1 Samples and $B^0 \to K^{*0}\nu\bar{\nu}$ Reconstruction

5.1.1 Data sample

We use the same e^+e^- collision data sample described in Sec. 4.1.1, collected with the SuperKEKB collider between 2019 and 2022. The dataset corresponds to $365\ \mathrm{fb}^{-1}$ [96] at $\sqrt{s} = 10.58\ \mathrm{GeV}$ (the $\Upsilon(4S)$ resonance), containing $(387 \pm 6) \times 10^6\ B\bar{B}$ pairs. An additional $42\ \mathrm{fb}^{-1}$ off-resonance sample, taken $60\ \mathrm{MeV}$ below the $\Upsilon(4S)$, is used to study continuum backgrounds from $e^+e^- \to q\bar{q}$ (u, d, s, c) and $\tau^+\tau^-$. The primary dataset corresponds to an integrated luminosity of 362 fb^{-1} at the $\Upsilon(4S)$ resonance. A recent dedicated measurement of the integrated luminosity using Bhabha, dimuon, and digamma-dominated channels gives $365.37 \pm 1.70\ \mathrm{fb}^{-1}$ [96], which is now adopted as the official Belle II value (rounded to $365\ \mathrm{fb}^{-1}$).

5.1.2 Simulation Sample

Signal samples. For the $B^0 \to K^{*0}\nu\bar{\nu}$ signal, two types of simulated samples are produced. A run-independent sample of 10 million events is used for training, validating the classifiers, and calculating detector-related systematic uncertainties, while a run-dependent sample corresponding to 4 million $\Upsilon(4S)$ events is used to construct the signal template in the analysis. The generation is based on SM form factor calculations [38], with the K^{*0} mesons decayed inclusively into allowed hadronic final states. Figure 5.1 shows the simulated distributions of the di-neutrino invariant mass squared and the cosine of the helicity angle for $B^0 \to K^{*0}\nu\bar{\nu}$ decays, where the helicity angle is defined as the angle of the $K^\pm$ (or $\pi^\pm$) in the K^{*0} rest frame with respect to the B momentum direction.

Generic Samples. A run-independent generic sample corresponding to $400\ \mathrm{fb}^{-1}$ is used for the training and validation of multivariate classifiers, following the same generation strategy described in Sec. 4.1.2. In addition, a run-dependent generic sample corresponding to four times the integrated luminosity of the data is employed to study and validate the background modeling and compositions, and further model the background contributions in the signal region.

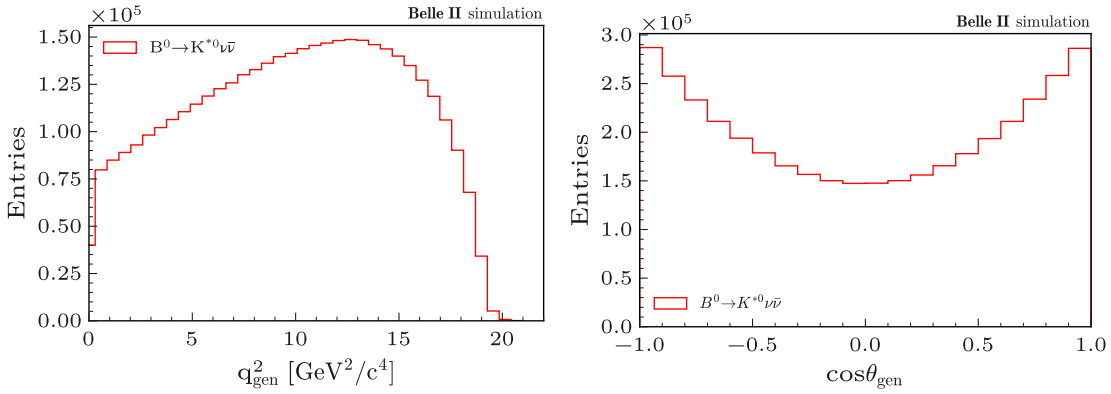

Figure 5.1: SM prediction for the simulated $B^0 \to K^{*0}\nu\bar{\nu}$ decays: (left) di-neutrino invariant mass squared,q^2, and (right) helicity angle cosine, $\cos\theta$ distributions [38].

Dedicated background samples. Certain background processes in the $B^0 \to K^{*0}\nu\bar{\nu}$ analysis are insufficiently modeled in the generic MC, either due to limited experimental input or to the complexity of their decay dynamics. Since these channels can mimic the kinematic properties of the signal, dedicated run-independent samples with improved modeling are introduced to replace the corresponding components in the generic MC, which are excluded to prevent double-counting. The dedicated samples employed for this purpose are summarized in Table 5.1.

dedicated sample	model
$B^+ \to K^+\nu\bar{\nu}$ $B^0 \to K^0_S\nu\bar{\nu}$ $B^+ \to K^{*+}\nu\bar{\nu}$	SM form factor [38]
$B^+ \to K^+p\bar{p}$	To model isospin counterpart $B^+ \to K^+n\bar{n}$
$B^+ \to K^{*+}p\bar{p}$	To model isospin counterpart $B^+ \to K^{*+}n\bar{n}$
$B^0 \to K^0_S\, p\bar{p}$	To model isospin counterpart $B^0 \to K^0_S\, n\bar{n}$
$B^0 \to K^{*0}p\bar{p}$	To model isospin counterpart $B^0 \to K^0_S n\bar{n}$
$B^0 \to K^{*0}K^0_SK^0_S$ $B^0 \to K^{*0}K^0_LK^0_L$ $B^0 \to K^{*0}K^0_SK^0_L$ $B^+ \to K^{*+}K^0_SK^0_S$ $B^+ \to K^{*+}K^0_LK^0_L$ $B^+ \to K^{*+}K^0_SK^0_L$	These are modeled with dedicated data-driven techniques to better reproduce kaon correlations.
$B^+ \to K_1(1270)^+\nu\bar{\nu}$ $B^+ \to K_1(1400)^+\nu\bar{\nu}$ $B^0 \to K_1(1270)^0\nu\bar{\nu}$ $B^0 \to K_1(1400)^0\nu\bar{\nu}$	To model $B \to X_s\nu\bar{\nu}$

Table 5.1: Dedicated simulation samples used to model specific background processes that are either poorly constrained experimentally or insufficiently described by generic MC. These replace the corresponding contributions in the generic samples to avoid double counting.

In addition, a run-dependent off-resonance MC sample corresponding to four times the integrated luminosity of the off-resonance data is used to model continuum backgrounds and to provide an auxiliary validation of the analysis.

5.1.3 $B^0 \to K^{*0}\nu\bar{\nu}$ Reconstruction

This analysis employs the inclusive-tagging method described in Sec. 3.1. In this case, a K^{*0} meson is reconstructed from its decay into $K^+\pi^-$, while the remaining charged tracks and neutral clusters are assigned to the ROE. The trigger relies on either the number of charged-particle trajectories reconstructed in the CDC or the total energy deposits measured in the ECL, and achieves an efficiency close to 100% for signal decays.

5.1.3.1 Charged Tracks and Neutral Clusters Selection

The selection of charged tracks and neutral clusters in the $B^0 \to K^{*0}\nu\bar{\nu}$ analysis follows the same general strategy as in the $B^+ \to K^+\nu\bar{\nu}$ channel, with modifications to accommodate the presence of K^0_S candidates and to optimize photon reconstruction.
Charged Tracks: Generic track-quality requirements, including transverse momentum and impact-parameter constraints, are applied as in the K^+ analysis (Sec. 4.1.3.1). For $K^0_S \to \pi^+\pi^-$ candidates, which contribute as background, a looser selection is adopted to maximize reconstruction efficiency of K^{*0}. Specifically, the requirements on the two-track invariant mass are loosened to (0.485, 0.510) GeV/c^2, the vertex-fit quality and minimum proper decay time requirement used in the $B^+ \to K^+\nu\bar{\nu}$ analysis are removed.
Photon (Neutral Cluster) Selection: Photon candidates are reconstructed from ECL clusters with a reduced minimum energy threshold of $E > 60$ MeV, which enhances sensitivity to low-energy photons. A new isolation criterion is imposed: the distance between the ECL cluster and the nearest charged track must exceed 20 cm, suppressing hadronic split-offs and improving photon purity. The details of the updated selections are provided in the Appendix 9.3. These modifications ensure higher reconstruction efficiency for displaced K^{*0} decays and better control of neutral-cluster backgrounds.

5.1.3.2 $K^{*0} \to K^+\pi^-$ Reconstruction

The K^{*0} meson is reconstructed in the charged mode $K^{*0} \to K^+\pi^-$. The neutral decay $K^{*0} \to K^0\pi^0$ is excluded due to its low reconstruction efficiency and poor purity. Candidate tracks are required to have at least one PXD hit and satisfy the general track-quality criteria discussed above. Particle identification is applied to enhance the selection of true kaons and pions, with the following requirements:

- **Kaon ($K^\pm$):** global likelihood probability $\mathrm{KaonID} > 0.75$.
- **Pion ($\pi^\pm$):** global pion likelihood probability $\mathrm{PionID} > 0.05$.

These criteria are optimized to maximize signal significance while controlling misidentification. The selected kaon and pion tracks are subsequently combined in a vertex fit to determine the decay vertex and momentum of the K^{*0} candidate. The reconstructed invariant mass of the $K\pi$ system is constrained to

$$0.8\ \mathrm{GeV}/c^2 < M(K\pi) < 1.0\ \mathrm{GeV}/c^2,$$

ensuring a well-defined K^{*0} signal.
The K^{*0} mass window retains 39.8% of the signal but produces high candidate multiplicities (5.2 signal and 10.5 background candidates per event). Adding kaon PID reduces the efficiency to 30.2% while cutting multiplicities to 1.8 and 2.4, respectively. Including pion PID further reduces them to 1.5 for signal and 1.9 for background, while the efficiency reaches 26.2%. The full evolution of signal K^{*0} selection is summarized in Tab. 5.2.
The nested pie chart, shown in Fig. 5.2, illustrates how candidate composition evolves through successive selection stages for neutral B decays in a run-dependent generic sample. From the inner to the outer layer, corresponding to the $M(K\pi)$ mass constraint, the addition of kaon PID, and finally full selection with pion PID, the total number of candidates decreases dramatically.
The inner layer is dominated by $\Upsilon(4S)$ (The two signal tracks are reconstructed from two different Bs) and B^0 (The two signal tracks are reconstructed from the same B) backgrounds, with $\Upsilon(4S)$ contributing roughly half of all candidates. Applying kaon PID in the middle layer reduces candidate counts by approximately an

cuts	signal efficiency [%]	avg. signal multiplicity	avg. background multiplicity
K^{*0} Mass	39.8	5.2	10.5
K^+ PXD hits and PID	30.2	1.8	2.4
π^- PXD hits and PID	26.2	1.5	1.9
MVA-two-candidate selection	26.1	1.3	1.5

Table 5.2: Signal efficiency and average candidate multiplicities after successive selection steps for $B^0 \to K^{*0}\nu\bar{\nu}$ reconstruction.

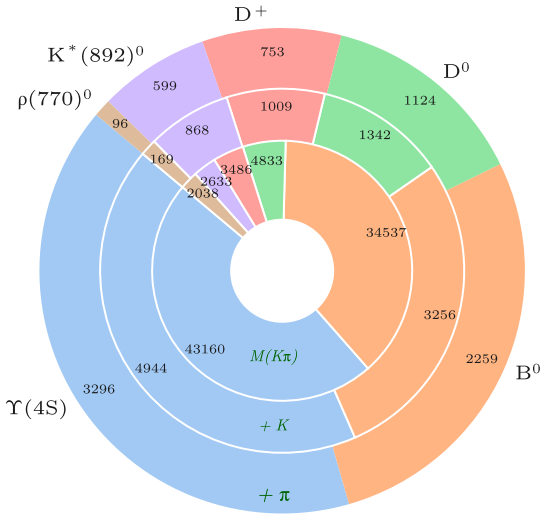

Figure 5.2: Evolution of candidate composition in neutral-B decays through successive selection steps: $M(K\pi)$ mass window (inner layer), addition of kaon PID (middle), and full selection with pion PID (outer)

order of magnitude, most strongly suppressing $\Upsilon(4S)$ and B^0, while the relative fraction of D-meson and resonance candidates (D^0, D^+, K^{*0}, ρ^0) increases. The outer layer, with pion PID included, further reduces the overall multiplicity, particularly for B^0 and ρ^0, leaving a smaller but cleaner sample dominated by $\Upsilon(4S)$ and the main D-meson channels.
Thus, the strategy effectively balances efficiency with manageable candidate multiplicity, in contrast to the K^+ channel, where unique candidate selection is feasible at an earlier stage.
To further reduce combinatorial background, a boosted decision tree classifier ($\mathrm{BDT_{CA}}$) trained using four input variables: the invariant mass of the $K^+\pi^-$ pair, the reconstructed dineutrino invariant mass squared (q^2_{rec}), the decay vertex displacement in the X-Y plane, and the vertex-fit χ^2-probability of the K^{*0} candidate—is applied to distinguish correctly reconstructed $K^{*0} \to K^+\pi^-$ candidates from random combinatorial combinations. The reconstructed momentum transfer squared, q^2_{rec}, defined in 4.2, is also employed. It is calculated under the assumption that the signal B meson is at rest in the e^+e^- c.m. frame, where $M_{K^{*0}}$ is the nominal K^{*0} mass and $E^*_{K^{*0}}$ is the reconstructed c.m. energy of the candidates.

$$q^2_{\mathrm{rec}} = \frac{s}{4c^4} + M^2_{K^{*0}} - \frac{\sqrt{s}E^*_{K^{*0}}}{c^4} \tag{5.1}$$

At this stage, two candidates with the highest $\mathrm{BDT_{CA}}$ scores are retained per event. This procedure achieves a signal efficiency of 26.1%, with an average signal multiplicity of 1.3 and an average background multiplicity of 1.5. Thus, the method reduces the background multiplicity while retaining high signal efficiency and defers the final best-candidate choice to later stages of the analysis to avoid introducing early selection bias.

The $\mathrm{BDT_{CA}}$ classifier is trained using 500 trees with a maximum depth of two. The relative importance of the input variables is summarized in Table 5.3. The reconstructed di-neutrino invariant mass squared, q^2_{rec}, is the most powerful discriminator, contributing about 56% of the separation power. The invariant mass of the K^{*0} candidate, $M(K\pi)$, provides an additional 29% importance by constraining candidates to be consistent with the expected resonance. The displacement of the K^{*0} decay vertex in the X–Y plane, $d_r(K^{*0})$, contributes nearly 10%, while the χ^2-probability of the vertex fit adds about 5%. Together, these features provide complementary information that allows the classifier to effectively distinguish true $K^{*0} \to K^+\pi^-$ candidates from random combinatorial combinations.

$\mathrm{BDT_{CA}}$ Parameters	Values
Number of trees	500
Tree depth	2
Features	Importances (%)
q^2_{rec}	56.1
$M(K\pi)$	28.9
$d_r(K^{*0})$	9.80
χ^2 of K^{*0} vertex fit	5.20

Table 5.3: The hyperparameters, relative features and importances of $\mathrm{BDT_{CA}}$ for the $B^0 \to K^{*0}\nu\bar{\nu}$ channel, where importances are normalized to unity.

5.1.3.3 Signal Event Reconstruction

After selecting the K^{*0} candidate, all remaining charged and neutral particles in the event are assigned to the ROE. Charged particles, excluding pions from K^0_S decays, are assigned mass hypotheses according to their PID likelihoods. The choice of hypothesis is weighted by the expected particle abundances in $\Upsilon(4S)$ decays. To ensure a well-reconstructed ROE, the number of charged tracks in the ROE is required to be between two and nine. The total visible energy must exceed $4\ \mathrm{GeV}$ and the polar angle of the missing momentum is restricted to $17^\circ < \theta < 160^\circ$ to suppress the low-multiplicity events and ensure the missing momentum points into the instrumented detector volume. Furthermore, the squared total event charge is required to be less than five, thereby suppressing events with non-physical charge imbalance. The details of low-multiplicity suppression are provided in the Appendix 9.4.

Tab 5.4 summarizes the event reconstruction and the selection criteria from charged and neutral objects to the best signal $B^0 \to K^{*0}\nu\bar{\nu}$ candidates.

Charged tracks	$p_t >= 0.1\ \mathrm{GeV}/c$, E<5.5 GeV, $17^\circ < \theta_{CDC} < 150^\circ$ tracks $\in K^0_S$: $0.485 < M(\pi^+\pi^-) < 0.510\ \mathrm{GeV}/c^2$, $\cos\theta_{\mathrm{flight}} > 0.98$. tracks $\notin K^0_S$: $\lvert d_z\rvert < 3.0$ cm, $d_r < 0.5$ cm
Neutral clusters	E>0.06 GeV, E<5.5 GeV, $17^\circ < \theta_{\mathrm{CDC}} < 150^\circ$, not matched to tracks, the distance between the ECL cluster and its nearest track is larger than 20 cm
Signal K^+	$\notin K^0_S$, KaonID>0.75, nPXDHits>=1
Signal π^-	$\notin K^0_S$, pionID>0.05, nPXDHits>=1
Signal K^{*0}	0.8 <$M(K\pi)$ < 1.0 GeV/c^2
Signal event	Ntracks∈[4, 11], visible energy>4 GeV, $\theta_{\mathrm{missing}} \in (17^\circ, 160^\circ)$, squared total charge < 5.
Best signal candidate	Two candidates with the highest $\mathrm{BDT_{CA}}$ scores are retained per event

Table 5.4: Summary of event reconstruction and selection criteria for the $B^0 \to K^{*0}\nu\bar{\nu}$ analysis.

5.1.4 Corrections to Simulated Sample

The charged-particle reconstruction is treated in the same way as in the K^+ channel, with a 0.9% per-track systematic uncertainty applied, as described in Sec. 4.1.4.

Particle identification corrections are applied for both kaons and pions. Kaon identification efficiencies and pion-to-kaon misidentification rates are measured using $D^{*+} \to \pi^+ D^0$, $D^0 \to K^-\pi^+$ decays, and applied per candidate as a function of momentum and polar-angle acceptance. Similarly, pion identification efficiencies and kaon-to-pion misidentification rates are incorporated to ensure that both kaon and pion PID performance is consistent between data and simulation.

For photon energy bias, we correct the data using standard `basf2` photon energy bias correction as recommended by the Neutrals Performance group [106], and evaluate the systematic by varying the photon energy in MC by $\pm$0.5%. For neutral clusters not associated with photons, we reduce their energy by 10% (validated with run-dependent simulation) and assign a 100% systematic to this correction. Corrections for the K_L^0 efficiency follows the same strategy as that in $B^+ \to K^+\nu\bar{\nu}$ analysis (Sec. 4.1.4).

Several of the branching fractions of leading B-meson decays are corrected in the simulation to match the most recent measurements, and the branching fractions of the main $B \to D^{**}X$ decays updated to reflect the most recent experimental measurements. The details of the correction procedure, including the definition of the weights, decay-mode classification, and associated systematic uncertainties, will be discussed in the systematic section 5.6.1, since they are studied in the signal region. Here, we only note that these corrections are applied consistently to the simulated samples.

In addition, continuum mis-modeling and signal-like $B\bar{B}$ backgrounds are corrected in the vicinity of the signal region (Sec. 5.4).

Together, these corrections ensure that the simulated samples reproduce the detector performance and physics branching fractions observed in data, and their associated uncertainties are propagated through to the signal efficiency and background estimates.

5.1.5 Definition of Discriminating Variables

In the $K^{*0} \to K^+\pi^-$ channel, the signal side consists of two charged tracks (K^+ and π^-). This enables the use of two additional two-body kinematic and vertex-topology observables that are not available in the single-track K^+ channel. Examples include the relative impact parameters of the kaon and pion, the opening angle between the kaon and pion momenta, Armenteros–Podolanski variables [107], and pointing variables that quantify the alignment of the reconstructed K^{*0} momentum with its decay vertex. These observables provide enhanced separation power against both $B\bar{B}$ and continuum backgrounds.
The discriminative kinematic and topological variables are constructed as input to the BDTs. As in the K^+ analysis, the transverse and longitudinal impact parameters of the signal tracks are included, and are discussed in Sec. 4.1.5.

5.1.5.1 Signal-K^{*0}-related Observables

- $M(K^{*0})$ Invariant mass of signal K and π.
- $d_r(K^{*0})$, $d_z(K^{*0})$ Radial and longitudinal distance between the POCA of the K^{*0} candidate and the IP.
- $d_r(\pi^{\pm})$ Radial distance between the POCA of the signal $\pi^{\pm}$ track and the IP.
- $q_T(K)$, $q_T(\pi)$ Transverse momentum of the K/π relative to the K^{*0} momentum direction, $q_T(K/\pi) = \frac{|\vec{p}_{K/\pi} \times \vec{p}_{K^{*0}}|}{|\vec{p}_{K^{*0}}|}$
- $\Delta\rho(K\pi)$, $\Delta z(K\pi)$ Distance between POCA of K^+ and POCA of π^- in the transverse plane and beam axis. $\Delta\rho(K\pi) = \sqrt{(d_{K,x} - d_{\pi,x})^2 + (d_{K,y} - d_{\pi,y})^2}$, $\Delta z(K\pi) = |d_{K,z} - d_{\pi,z}|$
- $\cos(\mathrm{thrust}_B, z)$ Cosine of the angle between the thrust axis of the signal B candidate and the z axis.
- $\cos\theta(p_K, p_\pi)$ Opening angle between signal K and signal π, $\cos\theta_{K\pi} = \frac{\vec{p}_K \cdot \vec{p}_\pi}{\|\vec{p}_K\| \, \|\vec{p}_\pi\|}$

Figures. 5.3 and 5.4 show the distributions of the signal-K^{*0}-related variables in data and simulation, where 1% of Experiment 18 data is used. The distributions of these discriminating variables are compared between data and simulation, where the data-to-simulation ratios are required to lie within the range (0, 2) to ensure good agreement. To make the signal visible in comparison with the dominant background, the signal contribution is scaled by a factor of 5000. As expected, the invariant mass $M(K^{*0})$ of the signal candidates peak within the defined mass window 0.8–1.0 GeV/c^2, consistent with the K^{*0} resonance, while the background exhibits a much broader distribution. In true two-body decays such as $K^{*0} \to K^+\pi^-$, the transverse momenta of the daughters relative to the K^{*0} momentum, q_T, are strongly constrained, whereas in combinatorial backgrounds they tend to be broader. The other kinematic and topological observables, such as the impact parameters d_r and d_z, vertex separation variables $\Delta\rho$ and Δz, as well as angular observables (thrust correlation and opening angle), are exploited in a multivariate approach to enhance the separation between $B \to K^{*0}\nu\bar{\nu}$ signal and generic $B\bar{B}$ or continuum $q\bar{q}$ background.

5.1.5.2 ROE-related Observables

- ΔE_{ROE} Difference between the ROE energy in the c.m. and the energy of one beam of c.m. ($\sqrt{s}/2$), $\Delta E_{\mathrm{ROE}} = E_{\mathrm{ROE}} - \sqrt{s}/2$.
- p_{ROE} Magnitude of the ROE momentum
- $\theta(p_{\mathrm{ROE}})$ Polar angle of the ROE momentum
- $N_{\gamma,\mathrm{ROE}}$ Number of photons in the ROE
- $M_{\mathrm{bc,ROE}}$ Beam-constraint mass of ROE, $M_{\mathrm{bc,ROE}} = \sqrt{(\sqrt{s}/2)^2 - |\vec{p}\mathrm{ROE}|^2}$

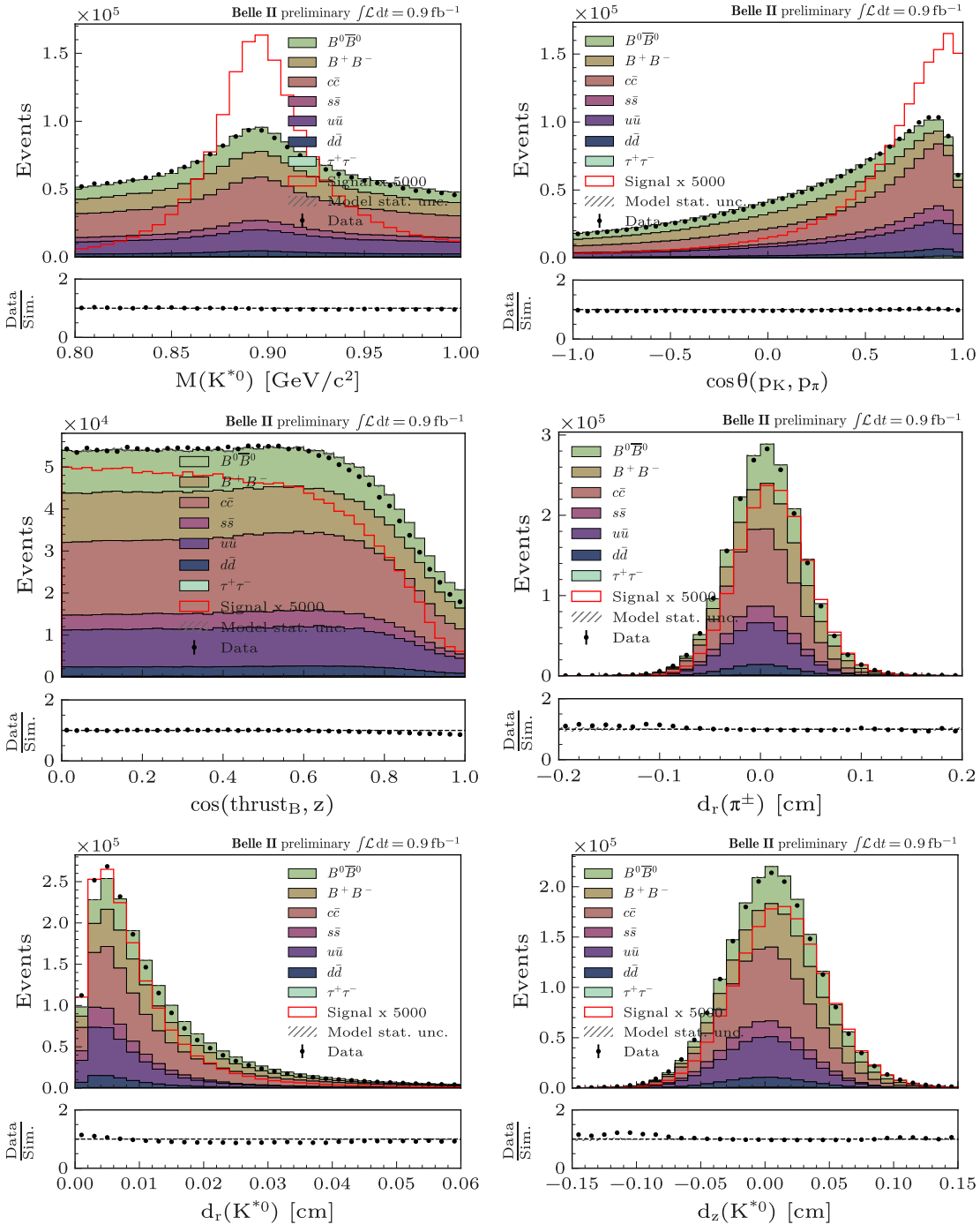

Figure 5.3: Distribution of signal-K^{*0}-related observables, $M(K^{*0})$, $\cos\theta(p_K, p_\pi)$, $\cos(\text{thrust}_B, z)$, $d_r(\pi^\pm)$, $d_r(K^{*0})$, and $d_z(K^{*0})$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \text{fb}^{-1}$) is used.

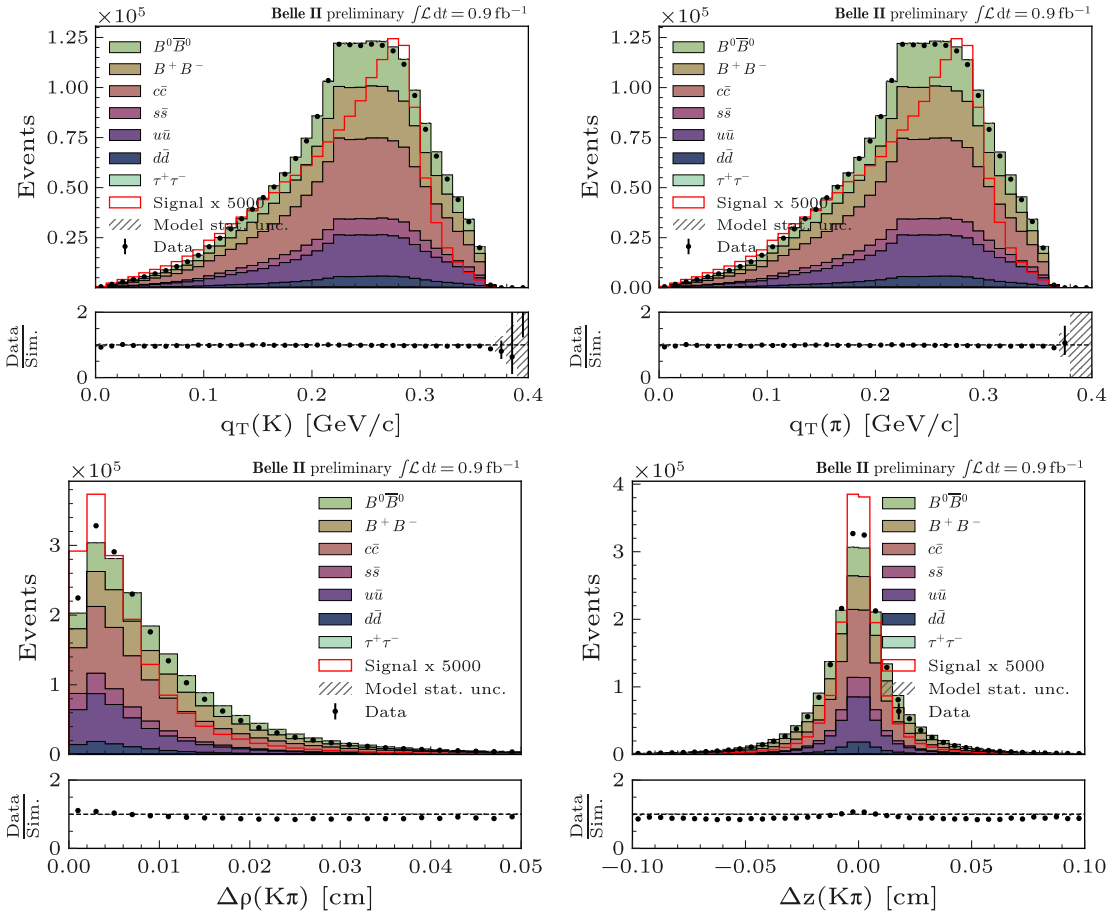

Figure 5.4: Distribution of signal-K^{*0}-related observables, $q_T(K)$, $q_T(\pi)$, $\Delta\rho(K\pi)$, and $\Delta z(K\pi)$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

- Modified Fox-Wolfram R_0^{oo}, R_2^{oo} Moment of the oo type (i.e. ROE-ROE), calculated in the c.m. frame

The observables constructed from the ROE provide complementary information on the global event topology and kinematics. They are powerful for distinguishing signal from continuum $q\bar{q}$ background. For example, the p_{ROE} and its polar angle $\theta(p_{\mathrm{ROE}})$ capture differences in momentum flow between signal and continuum. In ΔE_{ROE} distribution, signal events peak slightly below zero due to missing neutrinos, while continuum $q\bar{q}$ events, with jet-like fragmentation collecting more particles, show broader, shifted shapes above zero. The complementary variable $M_{\mathrm{bc,ROE}}$ uses the beam energy and reconstructed momentum, giving a sharp peak at $5.28\ \mathrm{GeV}/c^2$ for signal and lower values for continuum. And the modified Fox-Wolfram moments R_0^{oo} and R_2^{oo} quantify the overall isotropy, with continuum events being jet-like events and signal more spherical. The comparison between data and simulation is shown in Fig. 5.5 and 5.6.

5.1.5.3 Missing-energy-related Observables

- $\theta(p_{\mathrm{missing}})$ Polar angle of the missing three-momentum in the c.m. frame
- E_{missing} Missing energy
- $M^2_{\mathrm{missing}}/E_{\mathrm{missing}}$ Squared missing invariant mass over missing energy.
- $(p_l + p_\nu)^2$ Invariant mass squared of the lepton−neutrino system

We also construct variables related to the missing four-momentum in the event, which are particularly effective for semileptonic and invisible final states. These include the polar angle of the missing momentum vector $\theta(p_{\mathrm{missing}})$, the missing energy E_{missing}, and the ratio of missing mass squared to missing energy, $M^2_{\mathrm{missing}}/E_{\mathrm{missing}}$, which helps distinguish multi-neutrino events from mis-reconstruction. Additionally, the variable $(p_\ell + p_\nu)^2$ is introduced to suppress the semileptonic B decays, where the signal pion from $K^{*0} \to K^+\pi^-$ is treated as the lepton. This shifts the distribution for signal events to values around 25-75 $\mathrm{GeV}^2/\mathrm{c}^4$, whereas background events peak at lower values, typically 0-20 $\mathrm{GeV}^2/\mathrm{c}^4$. As before, the data-to-simulation ratio is constrained to (0, 2), and the signal is scaled by a factor of 5000 for visibility, shown in Fig. 5.7.

5.1.5.4 Entire-event-related Observables

- Fox-Wolfram R_1, R_2 First and second normalized Fox-Wolfram moment in the c.m. frame
- Harmonic Moment B_0, B_2 Zeroth-order and second-order harmonic moment with respect to the thrust axis in the c.m. frame
- Modified Fox-Wolfram $H^{\mathrm{so}}_{m,0}$, $H^{\mathrm{so}}_{m,2}$ and $H^{\mathrm{so}}_{m,4}$ Zeroth-order, second-order and fourth-order modified Fox-Wolfram moment related to missing momentum calculated in the c.m. frame
- Modified Fox-Wolfram $H^{\mathrm{so}}_{c,0}$, $H^{\mathrm{so}}_{c,2}$ Zeroth-order and second-order modified Fox-Wolfram moment related to charged particles in the ROE calculated in the c.m. frame
- Modified Fox-Wolfram $H^{\mathrm{so}}_{n,2}$ Second-order modified Fox-Wolfram moment related to neutral particles calculated in the c.m. frame
- $dz_{\mathrm{momentum\ line}}(K^{*0},\ \mathrm{Tag\ Vertex})$ Longitudinal distance between the tag vertex and the extrapolated momentum line of the K^{*0} candidate
- $d_r(K^{*0},\ \mathrm{Tag\ Vertex})$, $d_z(K^{*0},\ \mathrm{Tag\ Vertex})$ Radial and longitudinal distance between the POCA of the K^{*0} candidate and the tag vertex
- $N_{\ell^\pm}$ Number of $e^\pm$ and $\mu^\pm$ candidates
- $N_{e^\pm}$ Number of $e^\pm$ candidates
- N_{tracks} Number of charged particle candidates

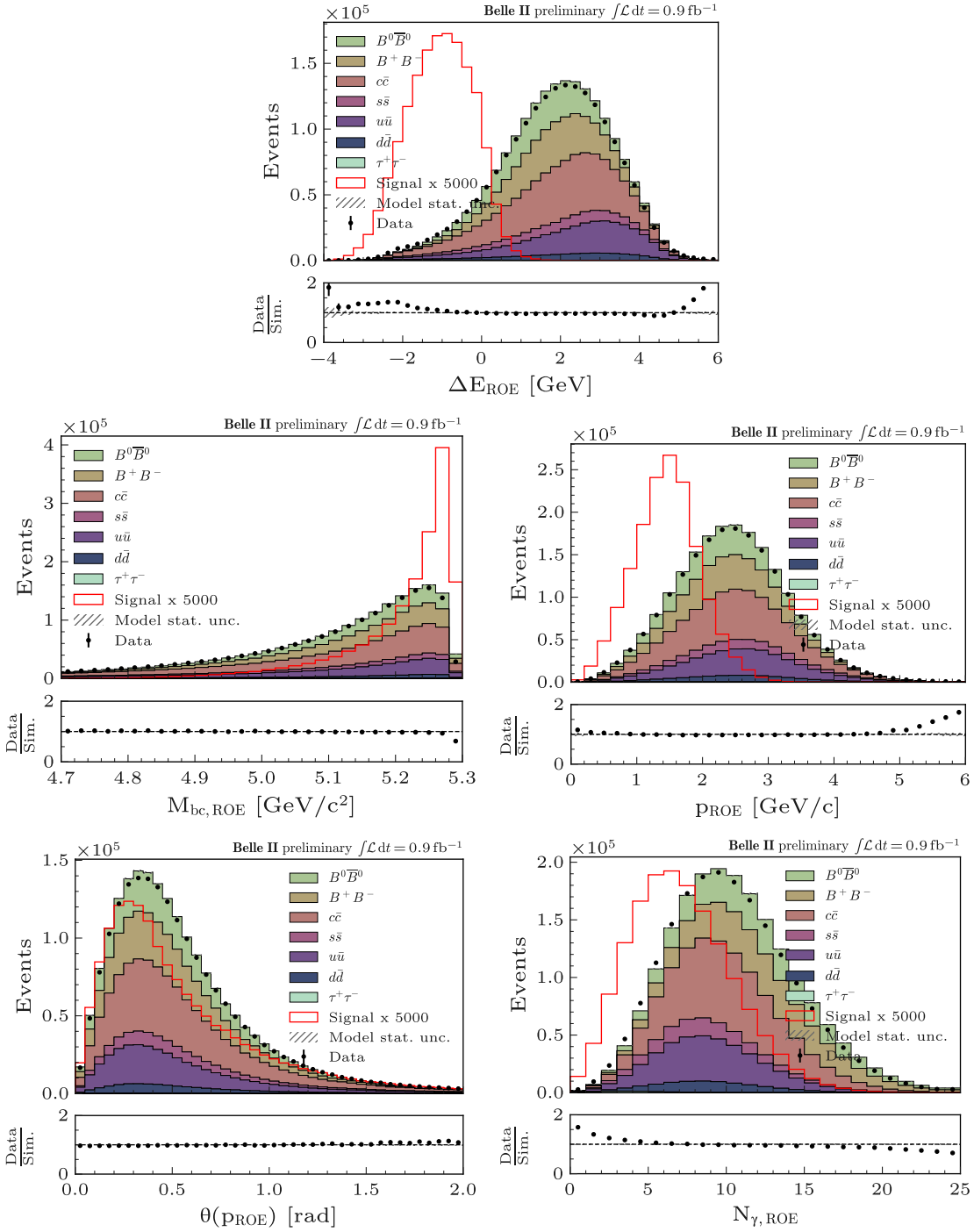

Figure 5.5: Distribution of ROE-related observables, ΔE_{ROE}, $M_{\mathrm{bc,ROE}}$, p_{ROE}, $\theta(p_{\mathrm{ROE}})$, and $N_{\gamma,\mathrm{ROE}}$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.

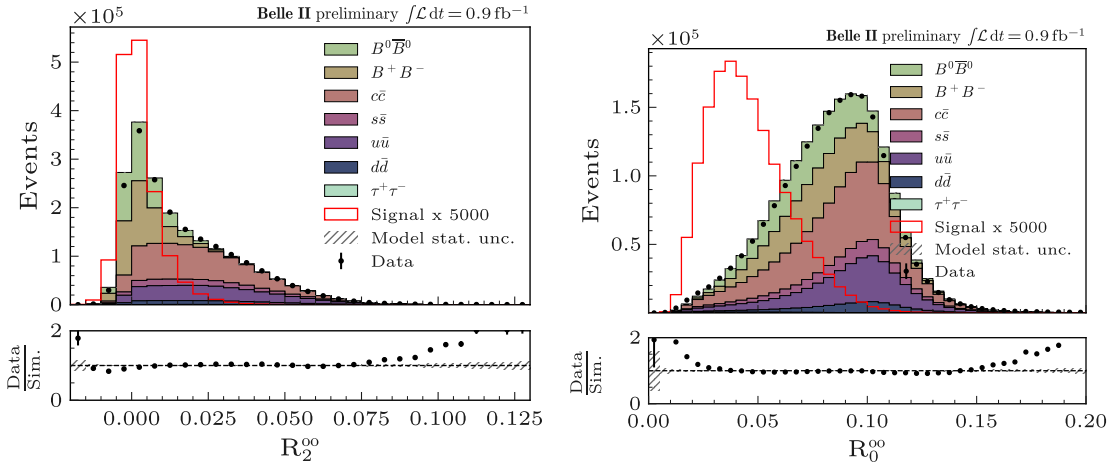

Figure 5.6: Distribution of Modified Fox-Wolfram R_0^{oo} and R_2^{oo} in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

- e^2 square of the total charge of tracks in the event
- $\cos\theta(\text{thrust})$ Cosine of the polar angle of the thrust axis in the c.m. frame
- $\cos(\text{thrust}_B, \text{thrust}_{\text{ROE}})$ Cosine of the angle between the thrust axis of the K^{*0} candidate and the thrust axis of the ROE computed in the c.m. frame
- $\sum p(z, \text{forward})$ Forward hemisphere momentum along the beam axis
- ξZ Event-level longitudinal momentum fraction.

Most of the event-shape and event-level variables, such as the normalized Fox–Wolfram moments(as shown in Fig. 5.8 and 5.9), harmonic moments, modified Fox–Wolfram moments, multiplicities of charged tracks and leptons, total charge, and thrust-related observables, follow the same definitions as in the K^+ channel. We additionally consider some new observables that exploit the global momentum flow of the event, such as the forward hemisphere momentum along the beam axis $\sum p(z, \text{forward})$, and the event-level longitudinal momentum fraction ξZ(as shown in Fig. 5.10 and 5.11). These variables provide complementary information about the overall topology of the event and help discriminate signal from continuum $q\bar{q}$ background.

5.1.5.5 D-meson-suppression-related Observables

For the D-veto related variables, we explore different combinations aimed to suppress peaking backgrounds from real charm meson decays (D^0 and D^+), which can mimic the signal topology. The variables used in the classifier training are reconstructed from the decay $B^+ \to K^+\pi^-\pi^+$. Further discussion is provided in Appendix 9.5.
In the $B^+ \to K^+\pi^-\pi^+$ reconstruction (a proxy reconstruction to form possible fake D candidates, not a literal signal mode):

- The K^+ is taken as the signal kaon track,
- One of the pions originates from the rest of the event (ROE), excluding K_S^0 candidates,
- The other pion is either from the ROE (not from K_S^0) or corresponds to the signal pion.

Additional requirements are imposed:

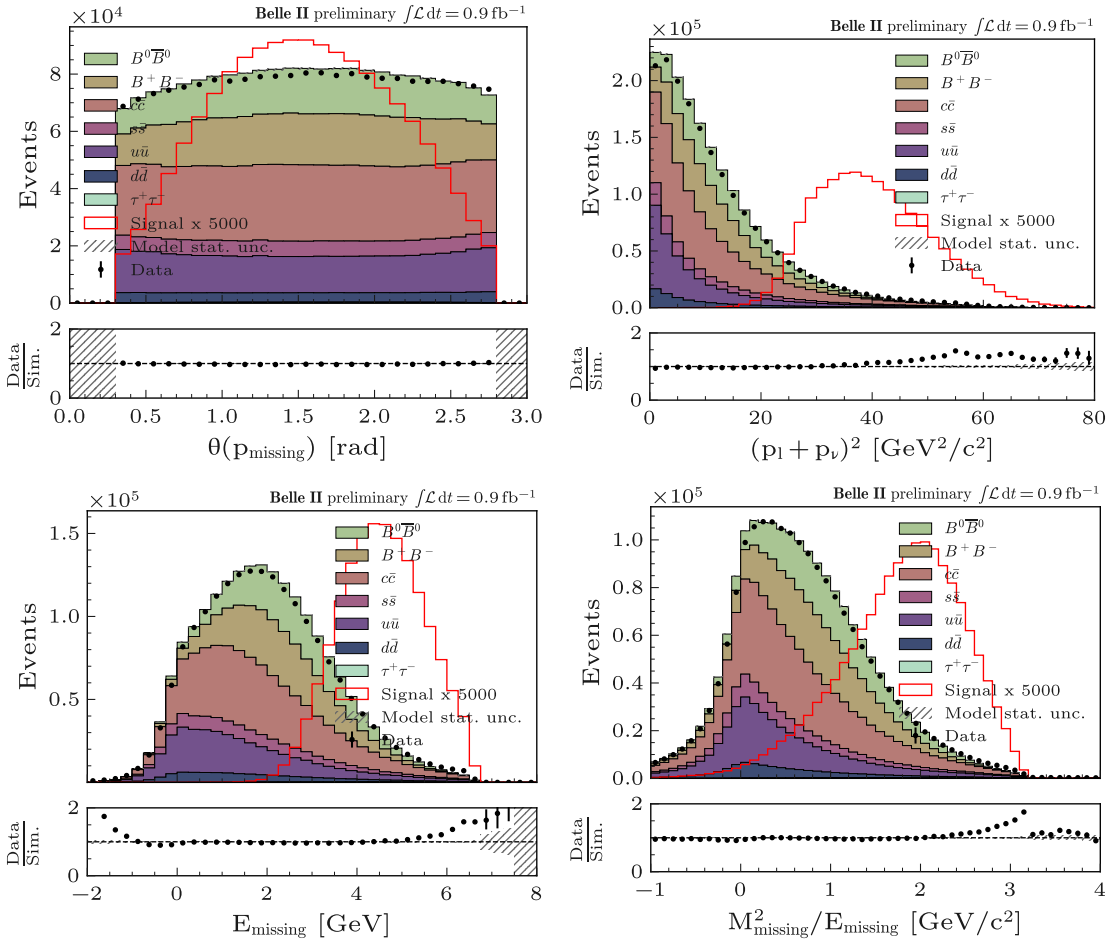

Figure 5.7: Distribution of missing-energy-related observables, $\theta(p_{\text{missing}})$, $(p_l + p_\nu)^2$, E_{missing} and $M^2_{\text{missing}}/E_{\text{missing}}$, in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

- The selected pion tracks must have at least one PXD hit,
- The invariant mass of the combination is required to be below 2.1 GeV/c^2,
- A kinematic fit is performed to reconstruct the vertex of the three tracks, and the best candidate is chosen based on the fit p-value.

The related variables are defined as:

- Vertex fit p-value of the best D^+ candidate
- $\cos\theta_{p_{D^+},r}$ Cosine of the D^+ meson momentum vector and the $K^+\pi^-\pi^+$ vertex in the xy-plane. $\cos\theta = \frac{\vec{p}_D \cdot \vec{r}_T}{|\vec{p}_D||\vec{r}_T|} = \frac{p_{D,x}(x_V - x_{IP}) + p_{D,y}(y_V - y_{IP})}{\sqrt{p^2_{D,x}+p^2_{D,y}}\sqrt{(x_V-x_{IP})^2+(y_V-y_{IP})^2}}$ where: $\vec{p}_D = (p_{D,x}, p_{D,y})$ is the transverse momentum of the D meson, (x_V, y_V) is the reconstructed D decay vertex position, (x_{IP}, y_{IP}) is the beam spot (interaction point), $\vec{r}_T = (x_V - x_{IP}, y_V - y_{IP})$ is the vector from IP to the D vertex.

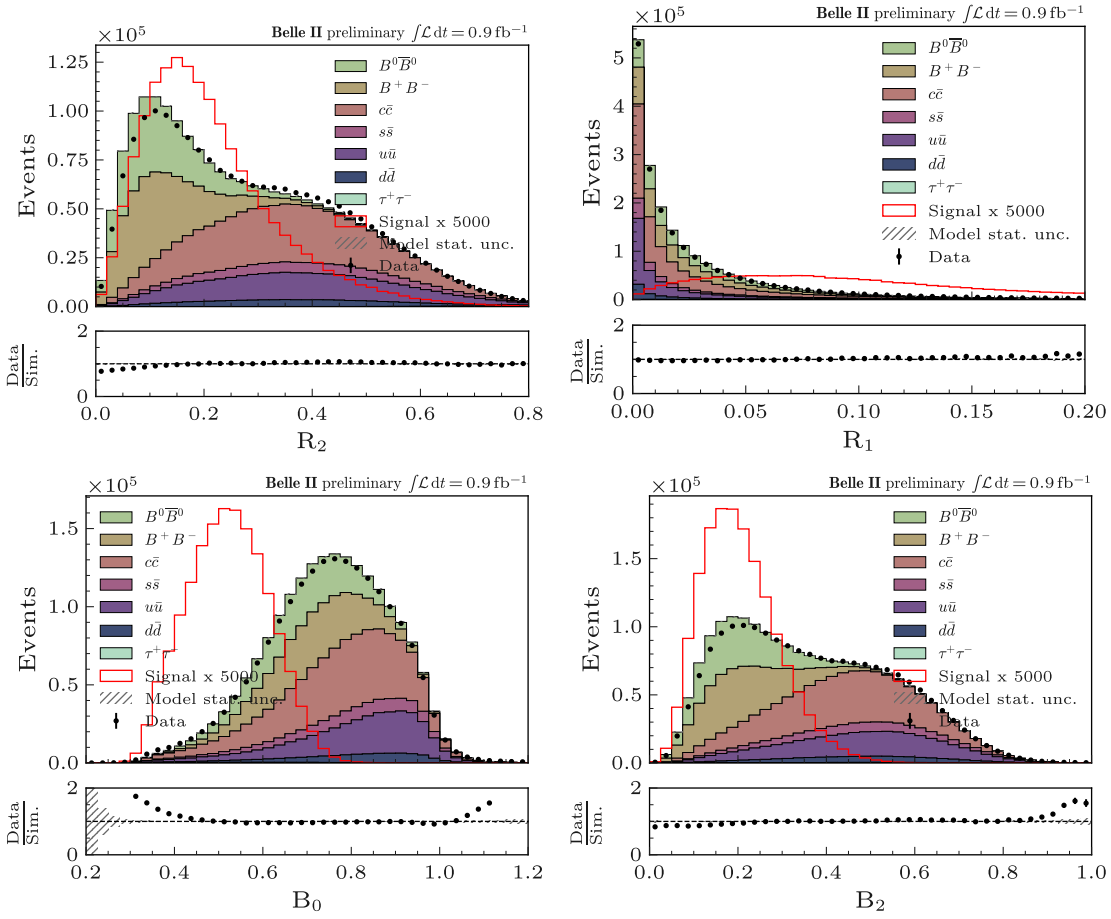

Figure 5.8: Distribution of entire-event-related observables, Fox-Wolfram R_1, R_2 and Harmonic Moment B_0,B_2 in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.

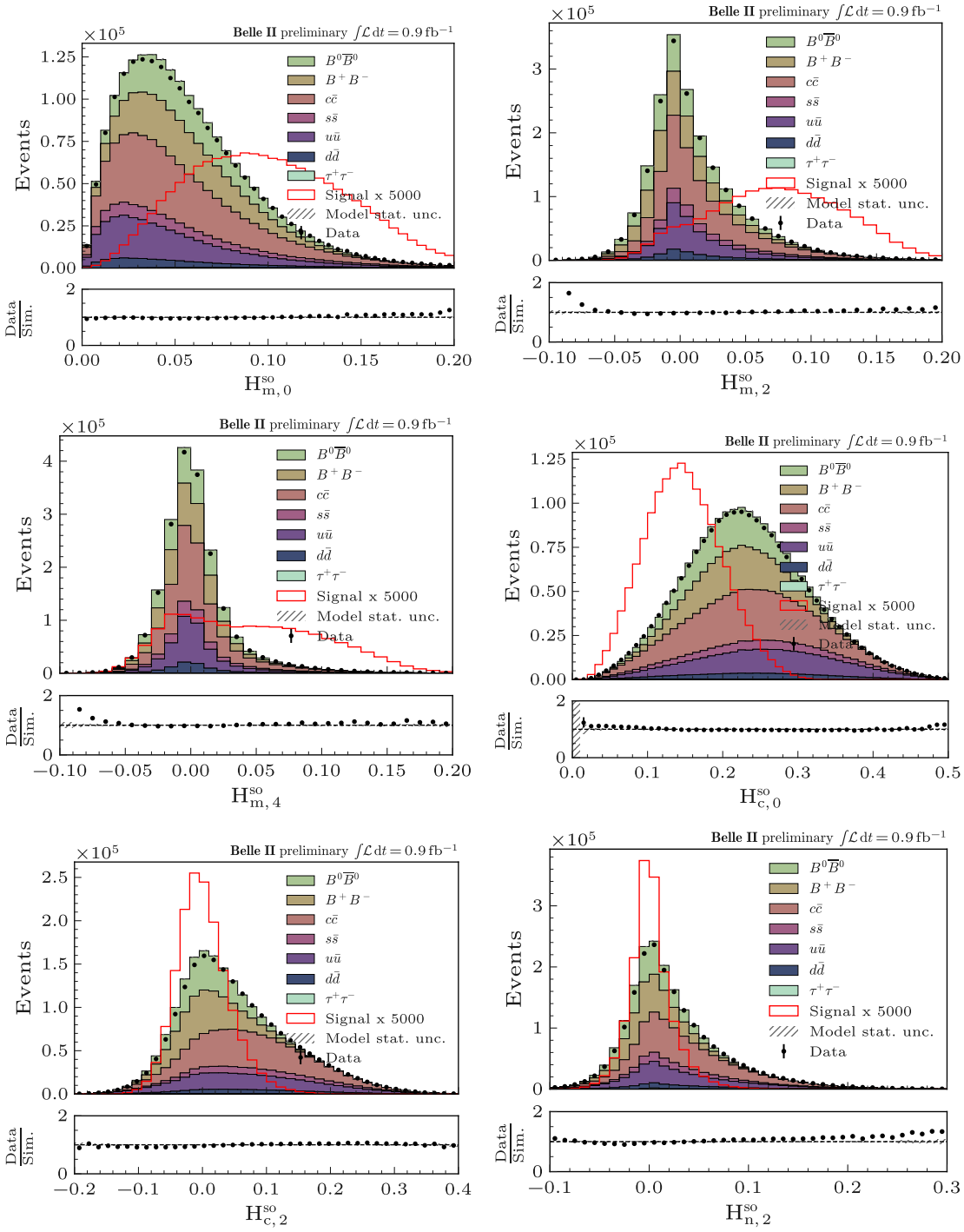

Figure 5.9: Distribution of entire-event-related observables, Modified Fox-Wolfram $H^{so}_{m,2}$, $H^{so}_{m,2}$ and $H^{so}_{m,4}$, $H^{so}_{c,0}$, $H^{so}_{c,2}$ and $H^{so}_{n,2}$ in data (points with error bars) and simulation(stacked filled histograms) for $B^0\overline{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.

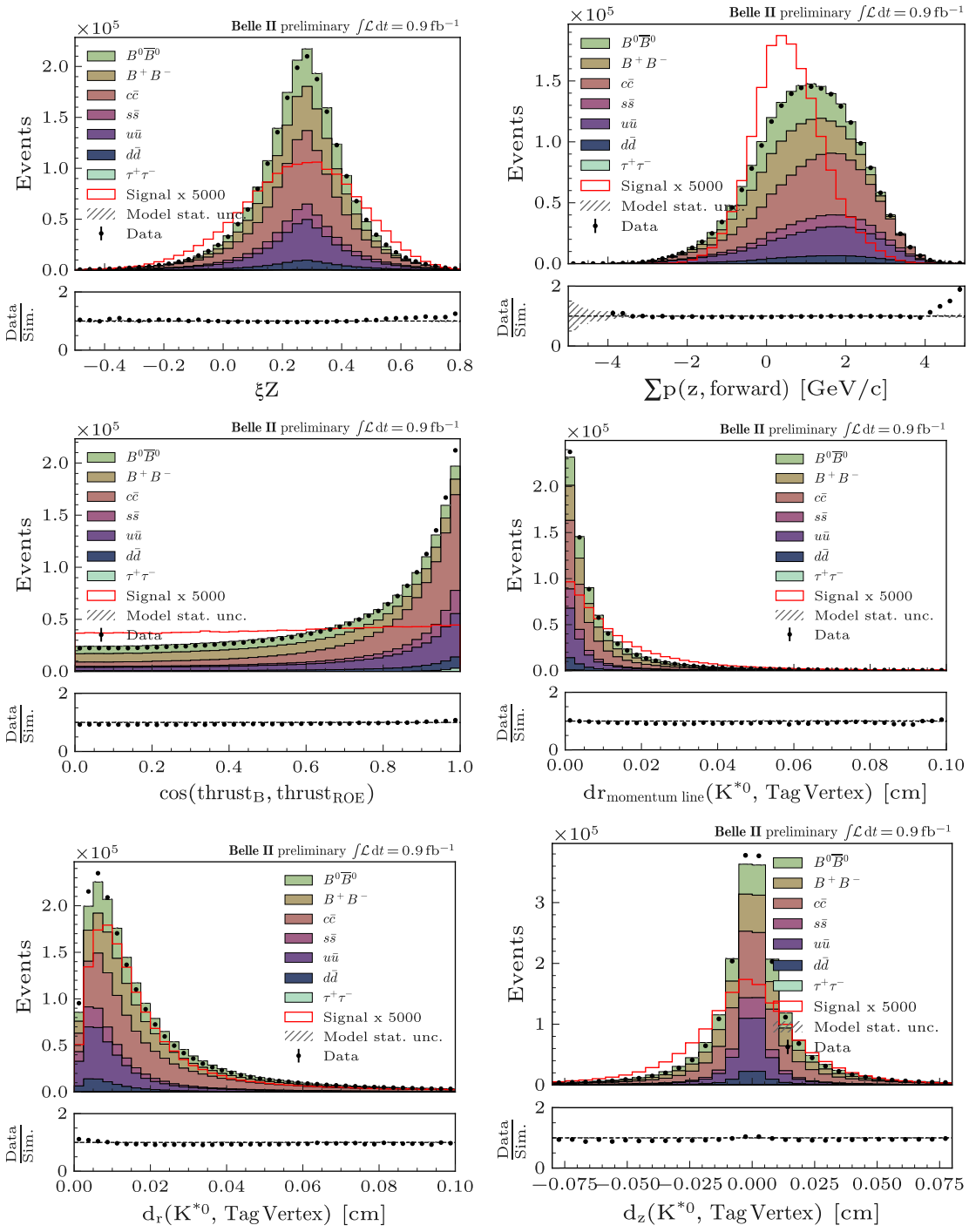

Figure 5.10: Distribution of entire-event-related observables, ξZ, $\sum p(z, \mathrm{forward})$, $\cos(\mathrm{thrust}_B, \mathrm{thrust_{ROE}})$, $dz_{\mathrm{momentum\ line}}(K^{*0},\ \mathrm{Tag, Vertex})$, $d_r(K^{*0},\ \mathrm{Tag, Vertex})$, and $d_z(K^{*0},\ \mathrm{Tag, Vertex})$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.

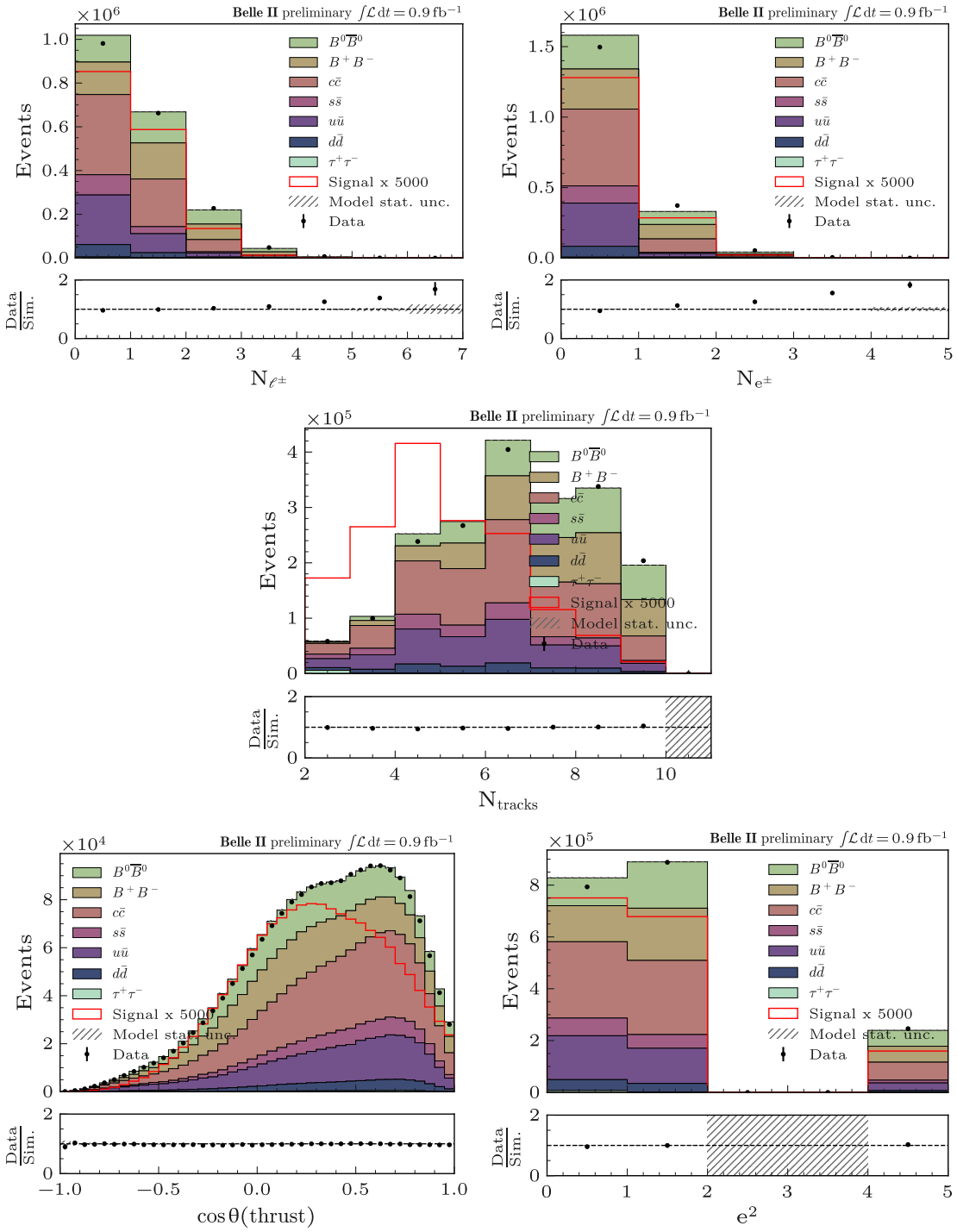

Figure 5.11: Distribution of entire-event-related observables, $N_{\ell^\pm}$, $N_{e^\pm}$, N_{tracks}, $\cos\theta(\text{thrust})$ and e^2 in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

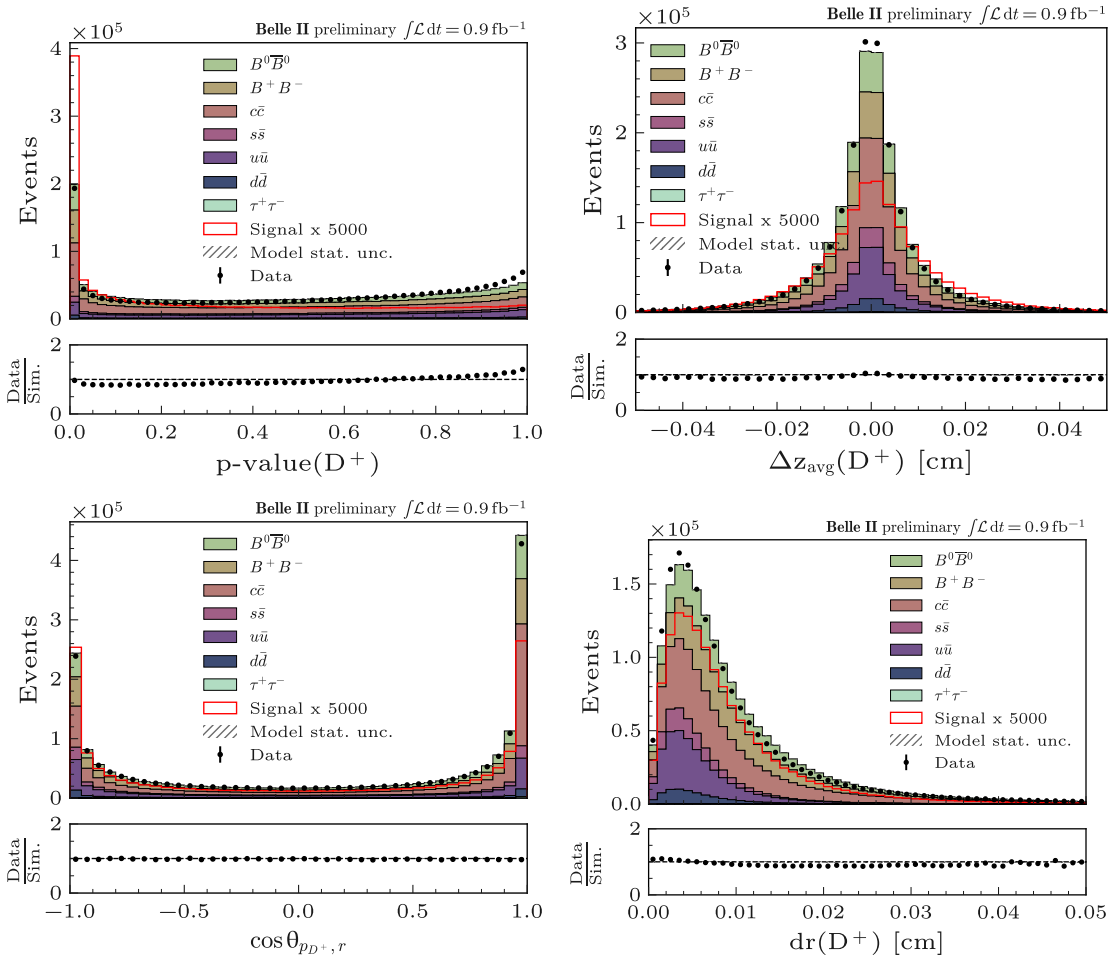

Figure 5.12: Distribution of D-meson-suppression-related observables, vertex fit p-value of the best D^+ candidate, $\Delta z_{\mathrm{avg}}(D^+)$, $\cos\theta_{p_{D^+},r}$ and $dr(D^+)$, in data (points with error bars) and simulation(stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram, scaled by a factor of 5000 for visibility, and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.

- $\Delta z_{\mathrm{avg}}(D^+)$ The average z-position of the two pions along the beam axis.
  $\Delta z_{\mathrm{avg}}(D^+) = \frac{(z_{\pi^-} - z_{\mathrm{vertex}}) + (z_{\pi^+} - z_{\mathrm{vertex}})}{2} = \frac{z_{\pi^-} + z_{\pi^+}}{2} - z_{\mathrm{vertex}}$
- $dr(D^+)$ Radial distance between the best D^+ candidate vertex and the IP

Figure. 5.12 shows the distribution of these D^+ veto-related variables.

5.2 Background Suppression

After the basic selection, the K^{*0} channel still exhibits substantial background. The first classifier using the same hyperparameters in Tab. 4.4 is applied, and input variables—primarily event-shape and ROE-related quantities—to discriminate $B\bar{B}$ from $q\bar{q}$ events is used in Tab. 5.5. The second classifier is then optimized using various combinations of features for training and validation.

5.2.1 The first classifier

Run-independent simulated samples with 12 discriminating variables are used to train a FastBDT classifier [88]. The training sample comprises approximately 3×10^5 reconstructed signal events and background contributions from $e^+e^- \to B\bar{B}$, $e^+e^- \to q\bar{q}$, and $e^+e^- \to \tau^+\tau^-$ processes, with 3-6$\times 10^5$ events per category across seven categories in total. To achieve a physically meaningful mixture, event weights are applied such that each category is scaled to the same effective luminosity, corresponding to about 1.18×10^8 events overall. The training uses candidate weights, which ensure that the reconstructed signal remains properly balanced with the weighted background composition. An independent sample of comparable size is reserved for validation.

Possible overtraining is investigated by comparing the weighted BDT response distributions of signal and background between training and validation samples. No significant discrepancy is observed, as illustrated in Fig. 5.13 (left).

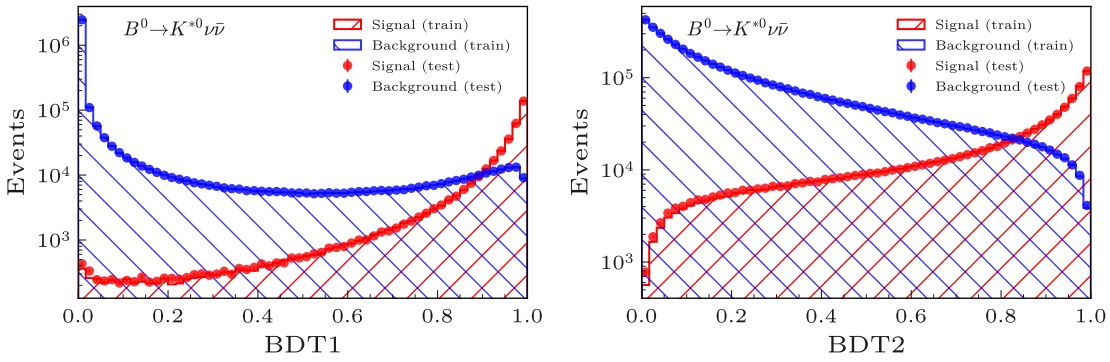

Figure 5.13: Simulated signal and background events taken from the training and the test sample in bins of the BDT_1 (left) and BDT_2 (right) output for $B^0 \to K^{*0}\nu\bar{\nu}$. For the BDT_2 output, we applied $\mathrm{DBT}_1 > 0.9$. Good agreement indicates no evidence of overfitting.

The relative importance of the input variables is summarized in Tab. 5.5. At this stage, ΔE_{ROE} provides the dominant discrimination power, consistent with observations in the K^+ channel. This behavior reflects the prevalence of $q\bar{q}$-dominated background in this selection stage. To enhance signal purity, we select candidates with a BDT_1 output exceeding 0.9. This requirement retains approximately 18% of the signal, compared to the K^+ channel; the lower signal retention reflects the additional combinatorial background in the two-track final state. The consistency between data and simulation is examined first, followed by the discriminating power of the input variables for separating signal from background. To blind the most sensitive region of the signal, only events with BDT_1 scores below 0.99 are considered. Representative variables are shown in Fig. 5.14, with the signal distribution scaled by a factor of 5000 for clarity.

These variables illustrate how different aspects of the decay are exploited: the resonance structure through $M(K^{*0})$, two-body decay topology via $q_T(K)$ and $\cos\theta(p_K, p_\pi)$, and charm suppression through the cosine of the D-meson momentum with respect to the IP−vertex vector. Together, they complement the global event-shape inputs and enhance the separation between signal and background.

Variable	$B^0 \to K^{*0}\nu\bar{\nu}$
ΔE_{ROE}	0.595
Modified Fox-Wolfram $H^{so}_{m,2}$	0.078
Modified Fox-Wolfram R^{oo}_0	0.069
p_{ROE}	0.066
Modified Fox-Wolfram $H^{so}_{m,4}$	0.040
Harmonic Moment B_0	0.036
Modified Fox-Wolfram R^{oo}_2	0.027
$\cos(\theta(\mathrm{thrust}))$	0.025
$\theta(p_{\mathrm{ROE}})$	0.025
$\cos(\mathrm{thrust}_B,\ \mathrm{thrust}_{\mathrm{ROE}})$	0.015
Fox-Wolfram Moment R_1	0.014
Harmonic Moment B_2	0.010

Table 5.5: Features and relative importance of BDT_1 for the $B^0 \to K^{*0}\nu\bar{\nu}$ channel, where importance are normalized to unity.

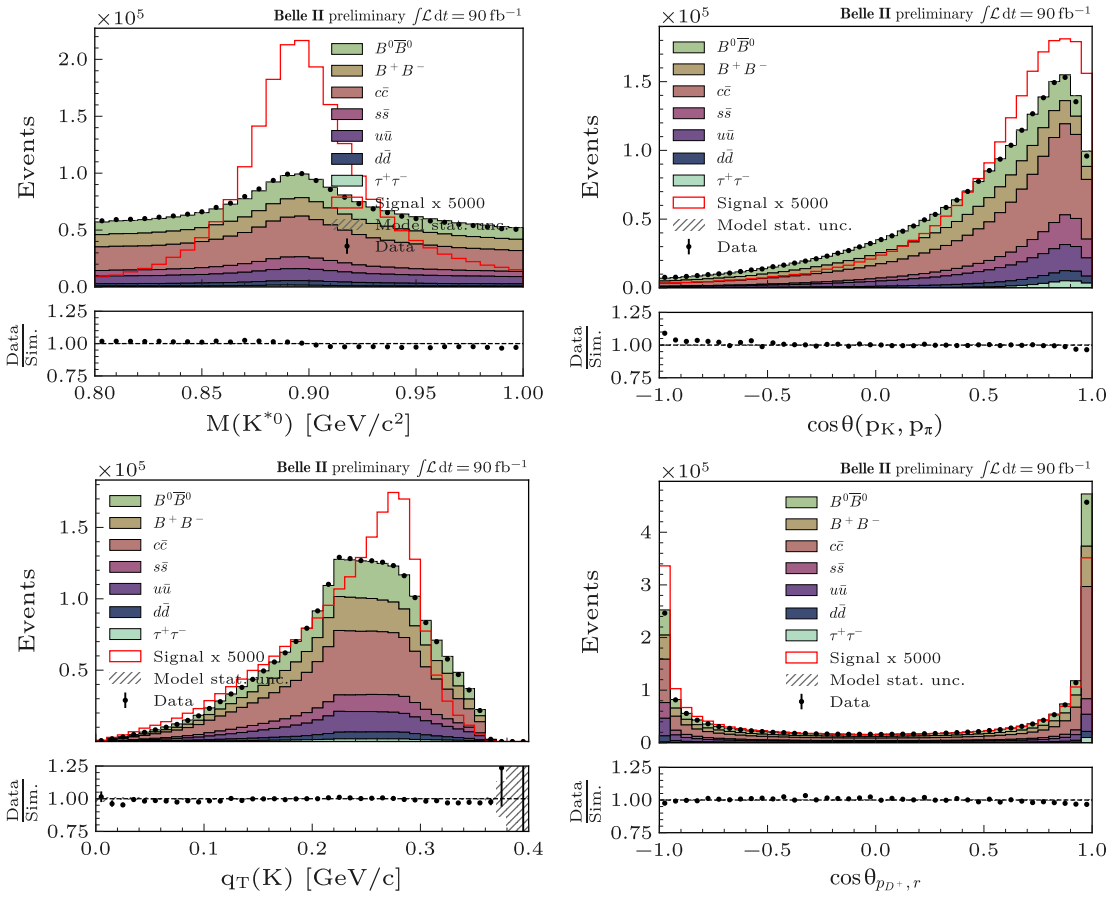

Figure 5.14: Distributions of representative input variables in the K^{*0} channel for events with $0.9 < \mathrm{BDT}_1 < 0.99$. Data (points with error bars) are compared with simulation (stacked histograms), with the $B^0 \to K^{*0}\nu\bar{\nu}$ signal overlaid as a line histogram scaled by a factor of 5000 for visibility. The $\Upsilon(4S)$ Experiment 18 data luminosity is used.

5.2.2 The second classifier

After the first stage, the background is dominated by $e^+e^- \to c\bar{c}$ and $e^+e^- \to B\bar{B}$ processes, which motivates the development of a second classifier to further improve signal discrimination. This classifier is implemented with the `XGBoost` algorithm [89], using the same configuration and hyperparameters as in the K^+ channel to ensure consistency across analyses. The training is performed with run-independent simulated samples, comprising about 1.0 $\times 10^6$ signal events and 4.8$\times 10^6$ background events, where the background events correspond to an integrated luminosity of 200 fb^{-1}. An independent sample of the same size is used for validation.
The choice of input variables is optimized by maximizing the figure of merit $S/\sqrt{S+B}$ and the signal efficiency while monitoring training quality metrics such as the receiver operating characteristic (ROC) curve [108] and the area under the ROC (AUC) score [109]. The final set includes 39 variables, covering kinematic information from the K^{*0} resonance, event-shape observables, D^+-veto quantities, ROE properties, and missing energy, as summarized in Tab. 5.6. Compared to the K^+ channel, the K^{*0} classifier draws more discriminating power from resonance-specific and charm-suppression variables, such as $\Delta z(K\pi)$, $q_T(K)$, and D^+ veto observables, while ROE and event-shape inputs retain complementary separation power. Because some variables, such as $M_{\mathrm{bc,ROE}}$, are potentially sensitive to run-dependent effects, the comparison between run-independent and run-dependent samples is carried out, confirming good consistency. Possible overtraining is evaluated by comparing classifier outputs for training and validation samples, with results shown in Fig. 5.13. The feature importance of the training variables are listed in 5.6 (left panel).

Variable	$B^0 \to K^{*0}\nu\bar{\nu}$
$M(K^{*0})$	0.181
Modified Fox-Wolfram $H^{so}_{m,2}$	0.069
Modified Fox-Wolfram R^{oo}_2	0.044
$N(\gamma_{\mathrm{ROE}})$	0.043
$\chi^2(D^- \to K^+\pi^-\pi^+)$	0.042
$\Delta\rho(K\pi)$	0.040
Modified Fox-Wolfram $H^{so}_{n,2}$	0.037
Modified Fox-Wolfram $H^{so}_{c,2}$	0.034
$dr(K^{*0})$	0.032
$\cos(\mathrm{thrust}_B,\ z)$	0.031
N_{lepton}	0.030
$M_{\mathrm{bc,ROE}}$	0.030
$\cos\theta(p_K, p_\pi)$	0.029
$\theta(p_{\mathrm{missing}})$	0.027
$dz_{\mathrm{momentum\ line}}(K^{*0},\ \mathrm{Tag\ Vertex})$	0.027
$(p_l + p_\nu)^2$	0.026
e^2	0.026
$N_{e^\pm}$	0.021
$p_T(\pi)$ with respect to the K^{*0}	0.019

Variable	$B^0 \to K^{*0}\nu\bar{\nu}$
$\Delta z(K\pi)$	0.018
$p_T(K)$ with respect to the K^{*0}	0.018
$dz(K^{*0},\ \mathrm{Tag\ Vertex})$	0.018
Fox-Wolfram Moment R_2	0.014
$\cos\theta_{p(D^+\to K^+\pi^-\pi^+),r}$	0.013
$dr(K^{*0},\ \mathrm{Tag\ Vertex})$	0.013
N_{tracks}	0.012
$\Delta z_{\mathrm{avg}}(D^+ \to K^+\pi^-\pi^+)$	0.012
$dr(D^+(\to K^-\pi^+\pi^-)$	0.010
$dr(\pi)$	0.010
$M^2_{\mathrm{missing}}/E_{\mathrm{missing}}$	0.010
Modified Fox-Wolfram $H^{so}_{m,0}$	0.009
Modified Fox-Wolfram $H^{so}_{c,0}$	0.009
ΔE_{ROE}	0.008
$\sum p(z, \mathrm{forward})$	0.008
$dz(K^{*0})$	0.006
$dr_{\mathrm{momentum\ line}}(K^{*0},\ \mathrm{Tag\ Vertex})$	0.006
ξZ	0.005
E_{missing}	0.004

Table 5.6: Features and relaive importance of BDT$_2$ for the $B^0 \to K^0\nu\bar{\nu}$ channel, where importance are normalized to unity. The left and right panels correspond to the first and second halves of the ranked feature list.

Following the strategy outlined in Sec. 4.2.2 for the $B^+ \to K^+\nu\bar{\nu}$ decay, we evaluate the signal efficiency, background suppression, and statistical significance for the $B^0 \to K^{*0}\nu\bar{\nu}$ channel. The SM expectation of $B^0 \to K^{*0}\nu\bar{\nu}$ is defined as

$$N^{\mathrm{SM}}_{\mathrm{sig}} = \mathcal{L} \times \sigma(e^+e^- \to \Upsilon(4S)) \times 2 \times \mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0) \times \mathcal{B}(B^0 \to K^{*0}\nu\bar{\nu}). \tag{5.2}$$

For this analysis, we use $\mathcal{L} = 365\ \mathrm{fb}^{-1}$, and the SM branching fraction $\mathcal{B}(B^0 \to K^{*0}\nu\bar{\nu}) = 9.05 \times 10^{-6}$ [38].

Analogous to the K^+ study (Sec. 4.2.2), we investigate the trade-off between signal efficiency and background rejection, shown in Fig. 5.15 (left). The maximum achievable efficiency in the K^{*0} case is found to be 0.187. For BDT_1, the significance curve again peaks at low efficiency, reaching a maximum around 0.28, while BDT_2 provides a more pronounced improvement. Specifically, BDT_2 attains a peak significance of about 0.94, achieved at efficiencies in the range (0.01, 0.04). This confirms that the enriched feature set of BDT_2 offers superior discrimination power also in the K^{*0} channel.
Since the BDT_2 score itself has no physical interpretation, we map it to the signal efficiency, ϵ_{sig}, Fig. 5.15 (right), and introduce the inefficiency quantile $\eta(\mathrm{BDT}_2)$ to quantify the fraction of discarded signal events. In this case, the most significant contributions are concentrated in the low-efficiency region,

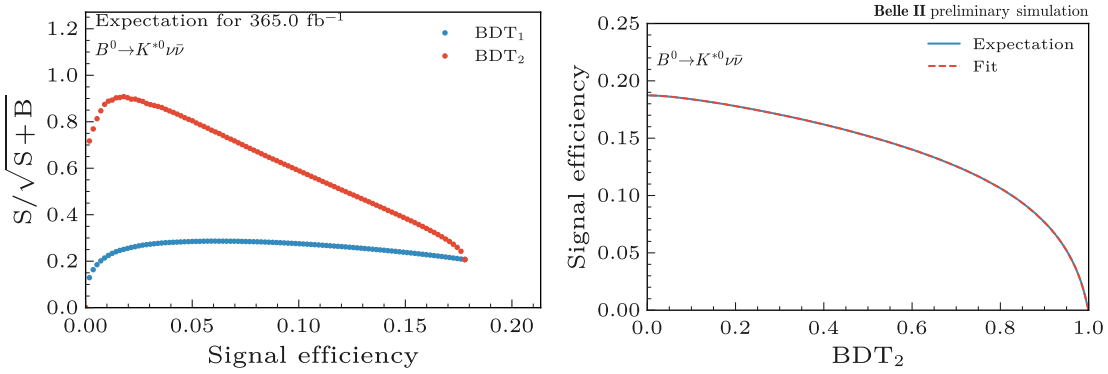

Figure 5.15: Signal significance versus efficiency (left) and efficiency as a function of the BDT_2 score (right) for the $B^0 \to K^{*0}\nu\bar{\nu}$ analysis. The results are shown for BDT_1 and BDT_2 selections.

To investigate the post-BDT_2 distributions in a controlled manner, we restrict the study to a safe, blinded region of phase space, as the signal region has not yet been explicitly defined. Figure 5.16 displays the distributions of $\eta(\mathrm{BDT}_2)$ and q^2_{rec} with $\eta(\mathrm{BDT}_2) < 0.9$, where the expected signal contribution is relatively small. Within this range, the data and simulated samples exhibit overall consistency, although additional validation studies are required and are presented in subsequent sections. Since $\eta(\mathrm{BDT}_2)$ is constructed to yield a uniform signal distribution, and the background yield is observed to decrease with increasing $\eta(\mathrm{BDT}_2)$, a dedicated optimization study is performed to determine the optimal $\eta(\mathrm{BDT}_2)$ threshold, ensuring a robust and unbiased definition of the signal region.

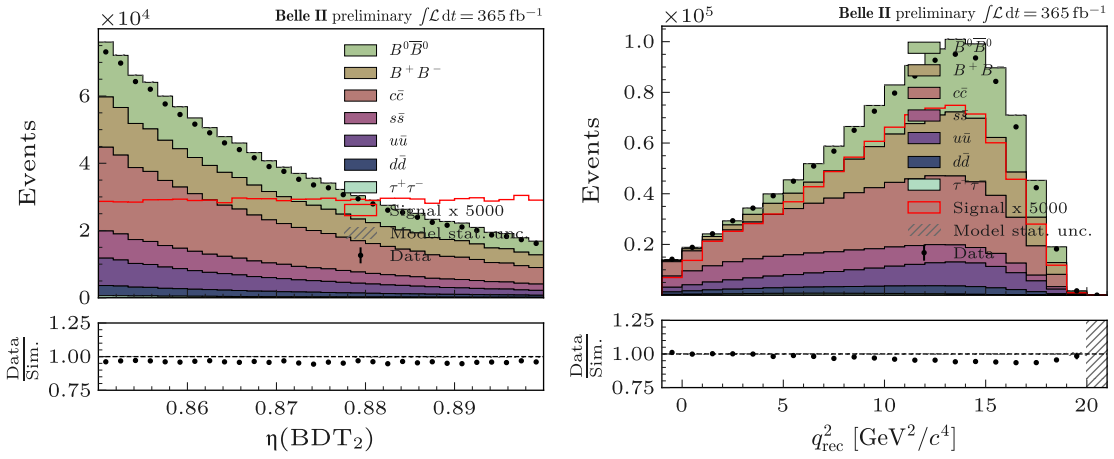

Figure 5.16: Distributions of $\eta(\mathrm{BDT}_2)$ and q^2_{rec} in the sideband region ($\eta(\mathrm{BDT}_2) < 0.9$), comparing data and run-dependent MC. The expected $B^0 \to K^{*0}\nu\bar{\nu}$ signal contribution is negligible in this region. The total $\Upsilon(4S)$ data luminosity is used.

5.3 Signal Region Definition and Signal Efficiency Validation

Following the background suppression achieved with the two BDT classifiers, we define the final signal region and optimize its binning for the statistical fit. The goal is to maximize sensitivity to the $B^0 \to K^{*0}\nu\bar{\nu}$ signal while maintaining robustness against statistical fluctuations. This is achieved by jointly binning in the classifier-derived inefficiency variable, $\eta(\mathrm{BDT}_2)$, and the reconstructed di-neutrino mass, q^2_{rec}. Each dimension requires a tailored binning strategy due to its distinct physical interpretations and event distributions.

5.3.1 Signal Region Definition

The optimization of the two-dimensional binning scheme is a critical step in balancing statistical precision with sensitivity to potential signal contributions. Since the two observables, $\eta(\mathrm{BDT}_2)$ and q^2_{rec}, exhibit different distributions and physical interpretations, their binning strategies are defined independently.

Binning in $\eta(\mathrm{BDT}_2)$. The inefficiency variable $\eta(\mathrm{BDT}_2)$, derived from the classifier response, is divided into equal-width bins in the range $[x, 1]$ with $x = 0.90, 0.91, \ldots, 0.98$. This choice is motivated by two considerations. First, it provides a uniform granularity across the classifier output, allowing a transparent interpretation of how signal efficiency and background rejection evolve with $\eta(\mathrm{BDT}_2)$. Second, it avoids introducing artificial structures into the signal efficiency distribution, which is observed to be approximately flat in the blinded control region. The uniform binning, therefore, ensures consistent coverage of the classifier space while maintaining statistical stability in each bin.

Binning in q^2_{rec}. In contrast, the reconstructed variable q^2_{rec} exhibits a highly non-uniform distribution, dominated by background contributions in the sideband region. A uniform binning in q^2_{rec} would lead to bins with insufficient statistics near the edges of the kinematic range. Instead, an adaptive binning strategy is employed over the interval $[-1, 25]$, with the number of bins varying between three and seven. To ensure sufficient statistics per bin, the following constraints are imposed:

$$\Delta q_i^2 \geq 3 \quad \text{for all bins}, \qquad \Delta q^2_{\text{last}} \geq 10, \tag{5.3}$$

where Δq_i^2 denotes the width of the i-th bin. These requirements prevent overly fine subdivisions in sparsely populated regions while retaining flexibility to resolve non-trivial structures in the distribution. As an illustration, a subset of possible three-bin q^2_{rec} configurations satisfying the statistical constraints is shown below. Each configuration is defined by the bin edges:

$$[-1, 2, 5, 25], \quad [-1, 4, 7, 25], \qquad\qquad [-1, 9, 12, 25], \quad [-1, 12, 15, 25]$$

These illustrate how the first and last bins are constrained to have sufficient width, while the intermediate bins are varied to maximize coverage of the sideband region without creating bins with too few events. The full set of valid binning configurations, combined with the uniform $\eta(\mathrm{BDT}_2)$ bins, results in the 11,790 unique two-dimensional binning schemes evaluated in this study.

Evaluation of binning combinations. By combining the nine possible $\eta(\mathrm{BDT}_2)$ configurations with all valid q^2_{rec} configurations, a total of 11,790 unique two-dimensional binning schemes are generated. Each configuration is subsequently evaluated using sghf, providing an estimate of the statistical uncertainty on the signal strength parameter, μ. This systematic procedure enables a quantitative comparison across binning strategies, ensuring that the selected configuration minimizes statistical uncertainty while preserving robustness.

The final strategy is thus asymmetric: equal-width bins in $\eta(\mathrm{BDT}_2)$, where the signal is approximately flat, combined with variable-width bins in q^2_{rec}. The comprehensive scan of 11,790 possible binning configurations, each assessed through its statistical precision, provides a principled basis for defining the signal region, ensuring optimized statistical sensitivity to a possible $B^0 \to K^{*0}\nu\bar{\nu}$ signal.

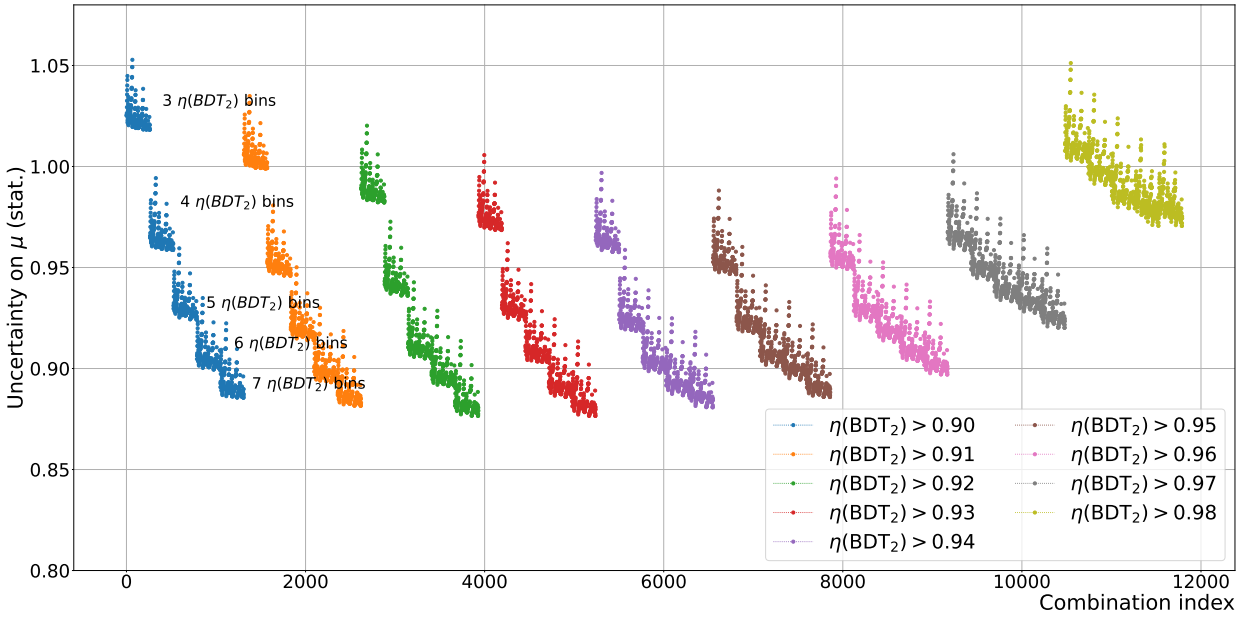

Figure 5.17: Statistical uncertainty on the signal strength parameter μ as a function of the 11,790 evaluated two-dimensional binning configurations. Each point corresponds to one binning scheme, ordered along the x-axis by configuration index.

The statistical uncertainties were evaluated using a representative dataset, with the x-axis corresponding to the combination index of the 2D binnings and the y-axis showing the statistical uncertainty on the signal strength parameter, μ, shown in Fig. 5.17.
For a fixed $\eta(\mathrm{BDT}_2)$ region, increasing the number of $\eta(\mathrm{BDT}_2)$ bins generally leads to a reduction in the statistical uncertainty, reflecting the improved resolution of the classifier space. In contrast, the impact of the q^2_{rec} binning on the overall uncertainty is comparatively small, indicating that the finer subdivision in q^2 contributes less to the statistical precision than the classifier binning.
In principle, seven bins of $\eta(\mathrm{BDT}_2)$ yield the lowest uncertainty; however, excessively fine binning can introduce statistical fluctuations due to limited events per bin, compromising the robustness of the measurement. Examining the effect of varying the lower boundary of the $\eta(\mathrm{BDT}_2)$ region while keeping the number of bins fixed, the uncertainty decreases when moving from the $(0.9, 1)$ range to $(0.95, 1)$, reflecting higher signal purity in the upper range of $\eta(\mathrm{BDT}_2)$. A slight increase in uncertainty is observed when extending the lower boundary further to $(0.97, 1)$, which is attributed to the reduced number of events surviving the tighter selection.
Overall, this analysis demonstrates that both the number of $\eta(\mathrm{BDT}_2)$ bins and the choice of the lower boundary significantly influence the statistical uncertainty, whereas variations in q^2_{rec} binning play a secondary role. Optimal sensitivity is achieved by balancing bin granularity with sufficient event statistics in each bin.

5.3.2 Final Signal Region and Binning Choice

Based on the statistical uncertainty study discussed above, the signal region is defined by selecting events with $\eta(\mathrm{BDT}_2) > 0.95$. For the 2D binning within this signal region, the $\eta(\mathrm{BDT}_2)$ axis is divided into 5 equal-width bins: (0.95, 0.96, 0.97, 0.98, 0.99, 1.00), allowing a fine-grained study of the classifier response while avoiding excessive statistical fluctuations in any single bin. The q^2_{rec} axis is divided into three bins: (-1, 4, 8, 21). This binning scheme provides a balance between maximizing sensitivity to potential signal contributions and maintaining robust statistical precision, serving as the basis for the subsequent signal extraction and fit procedures. This binning of signal region in a two-dimensional 5×3 scheme of $\eta(\mathrm{BDT}_2) \times \mathrm{q}^2_{\mathrm{rec}}$, shown in Fig. 5.18.

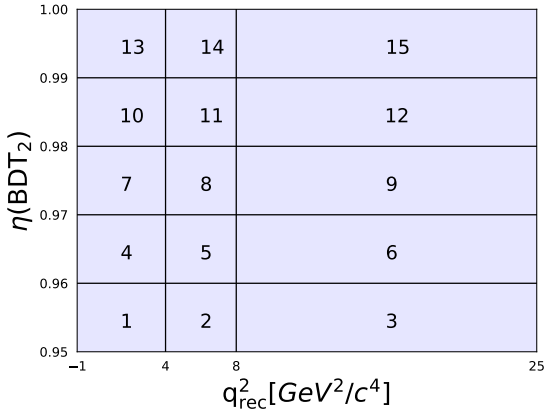

Figure 5.18: Final two-dimensional signal region binning scheme in $\eta(\mathrm{BDT}_2)$ and q^2_{rec} for the $B^0 \to K^{*0}\nu\bar{\nu}$ analysis

Selection stage	signal selection efficiency, $\epsilon(\mathrm{selection})$
Basic event selection	0.195
BDT_1 filter	0.18
Signal search region	0.05
Highest purity signal search region	0.01

Table 5.7: Cumulative signal-selection efficiency at each stages of the analysis for the $B^0 \to K^{*0}\nu\bar{\nu}$. Uncertainties are statistical only. PID weights are applied starting from the basic event selection step.

Following the definition of the signal region and the optimized two-dimensional binning, the signal efficiency is evaluated as a function of the reconstructed kinematic variables q^2_{rec} and the helicity angle θ, as shown in Fig. 5.19. The efficiency as a function of q^2_{rec} shows a clear dependence on the K^{*0} momentum because q^2_{rec} is calculated from the K^{*0} energy . In particular, small q^2_{rec} corresponds to high K^{*0} momentum (see Eq. 5.1). The highest efficiency, approximately 12%, is observed in the second q^2 bin $(1, 3)$, after which it gradually decreases with increasing q^2. This behavior indicates that higher-momentum K^{*0} candidates are more likely to pass the selection criteria and the BDT classifier selection threshold. The signal efficiency as a function of the helicity angle θ exhibits a non-uniform structure. Within the defined signal region ($\eta(\mathrm{BDT}_2) > 0.95$), the efficiency reaches a maximum of approximately 9% at $\cos\theta$ = -1, gradually decreases to a plateau of around 4% in the intermediate region $\cos\theta \in (-0.9, 0.8)$, and drops sharply to 2.5% in the last bin $(0.8, 1.0)$. Detailed studies indicate that, apart from the last bin, the decrease in efficiency in the intermediate region is primarily induced by the $\eta(\mathrm{BDT}_2) > 0.95$ requirement. To verify this, we examined the efficiency with looser selections, specifically $\eta(\mathrm{BDT}_2) > 0.85$, as shown in Fig. 5.19. In these looser selections, the plateau is significantly reduced, and the efficiency drops from $\cos\theta \sim -0.8$ to $\cos\theta \sim -0.2$ can be clearly attributed to the BDT_2 classifier response rather than detector acceptance or intrinsic kinematic effects.

This observation confirms that the helicity-angle dependence of the signal efficiency is a combined effect of the kinematics and the selection thresholds imposed by the multivariate classifiers. The sharp drop in the last bin $(0.8, 1.0)$ remains dominated by geometrical acceptance and lower event statistics in the forward region. These studies underscore the importance of carefully accounting for classifier-induced distortions when evaluating efficiencies across kinematic structures, particularly in the context of angular analyses.

The signal efficiency of each step is summarized in Tab. 5.7.

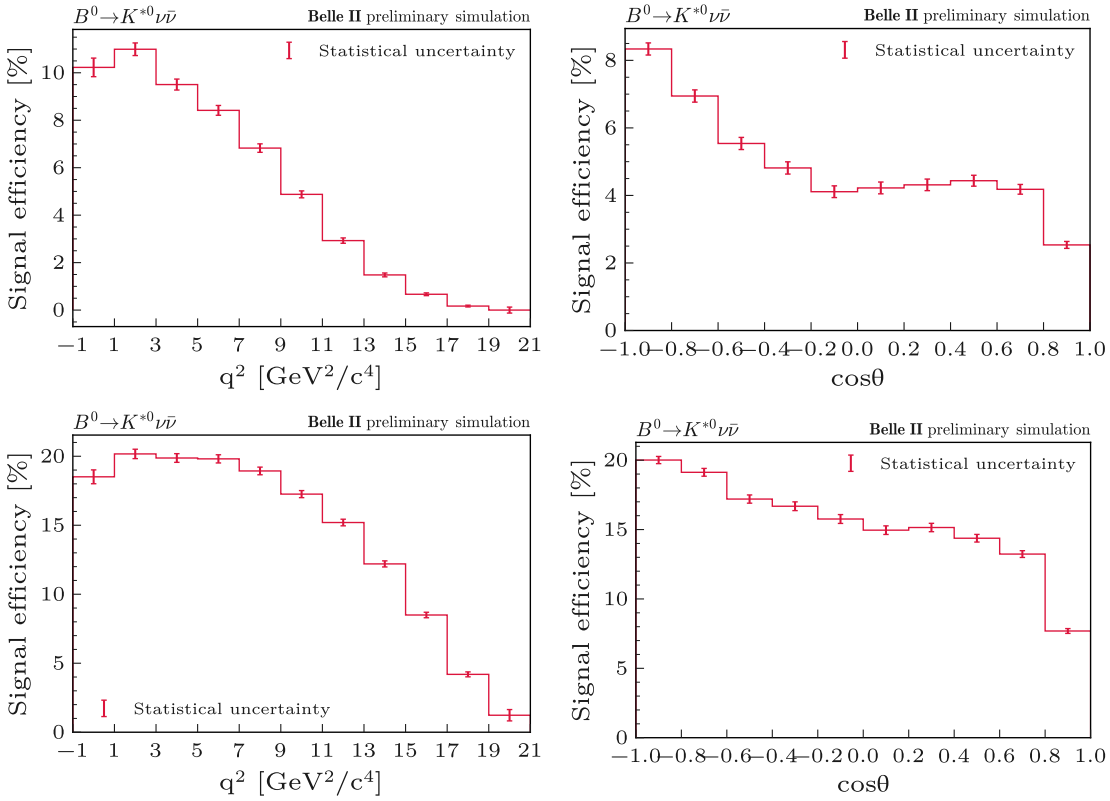

Figure 5.19: Signal efficiency as a function of generated q^2 (left) and helicity angle $\cos\theta$ (right). Top: nominal signal region ($\eta(\mathrm{BDT}_2) > 0.95$); bottom: looser requirement ($\eta(\mathrm{BDT}_2) > 0.85$)

.

5.3.3 Signal Efficiency Validation

To assess the signal reconstruction efficiency of $B^0 \to K^{*0}\nu\bar{\nu}$ decays, an embedding strategy is employed in which simulated signal decays are combined with the ROE from control samples. For the $B^+ \to K^+\nu\bar{\nu}$ channel, the control mode $B^+ \to K^+ J/\psi(\to \ell^+\ell^-)$ is utilized. The reconstructed kaon in this channel is replaced with the corresponding particle from simulated signal decays, retaining the ROE ensures realistic detector occupancy and background. For the $B^0 \to K^{*0}\nu\bar{\nu}$ channel, the procedure is modified to account for the explicit K^{*0} decay vertex. Here, $B^0 \to K^{*0} J/\psi(\to \ell + \ell^-)$ candidates reconstructed in reduced data format provide the kinematic and vertex information of the B^0 meson. The simulated decay is forced to occur at the reconstructed B^0 decay vertex, effectively neglecting the nominal lifetime in order to align with the measured geometry, forcing the decay at the measured vertex. The simulated $B^0 \to K^{*0}\nu\bar{\nu}$ decay is then merged with the ROE from the corresponding control sample, yielding embedded samples that accurately reproduce the signal topology in a realistic event environment.

a. **Control sample selection and candidate reconstruction:** Use $B^0 \to K^{*0}(\to K^+\pi^-)J/\psi(\to \ell^+\ell^-)$ candidates from reduced data format, where $\ell = \mu, e$. Apply basic track quality requirements: transverse momentum $p_T > 0.1$ GeV, track angles within CDC acceptance, longitudinal impact parameter $|dz| < 3.0$ cm, transverse impact parameter $dr < 0.5$ cm, more than 4 tracks passing standard quality criteria in the event, and at least one hit in the pixel detector. Apply particle identification:

   - Muon channel ($J/\psi \to \mu^+\mu^-$): muons with μID > 0.5; kaons with μID< 0.5 and kaonID > 0.9; pions with μID < 0.5 and kaonID < 0.9.
   - Electron channel ($J/\psi \to e^+e^-$): electrons with eID > 0.5; kaons with kaonID > 0.9 and eID < 0.5; pions with kaonID < 0.9 and eID < 0.5; include Bremsstrahlung photons with $E_\gamma < 1$ GeV in the electron reconstruction.

   Reconstruct $K^{*0} \to K^+\pi^-$ with a kinematic fit, require invariant mass in [0.8, 1.0] GeV/c^2, and χ^2 probability of fit is larger than 0.001. If multiple candidates exist, rank by reconstructed $q^2_{\rm rec}$ and select the best candidate. Reconstruct $J/\psi \to \ell^+\ell^-$ with $|\Delta M| < 0.05$ GeV/c^2 and apply a mass-constrained kinematic fit. Form $B^0 \to J/\psi K^{*0}$ candidates with beam-constrained mass $M_{\rm bc} > 5.25$ GeV/c^2 and energy difference $|\Delta E| < 0.10$ GeV.

b. **Extract kinematics and simulate the signal:** For the selected B^0 candidate and its daughters, retain momentum components (p_x, p_y, p_z), spatial coordinates (x, y, z), and impact parameters relative to the interaction point. Store the decay vertex of the B^0 meson to preserve the full decay geometry. This information captures the essential kinematics and spatial configuration for embedding while avoiding contamination from other tracks or calorimeter deposits. Generate $B^0 \to K^{*0}\nu\bar{\nu}$ decays using the stored kinematic and vertex information. Force the B^0 decay at the measured vertex to maintain correct spatial alignment with the rest of the event.

c. **Embed the signal:** Merge the simulated $B^0 \to K^{*0}\nu\bar{\nu}$ decay with the ROE from the control $B^0 \to J/\psi K^{*0}$ candidate. The resulting embedded samples allow a direct measurement of reconstruction efficiency in a realistic background, while controlling systematic uncertainties from modeling of the ROE.

Since the kinematics of the $K^\pm$ and $\pi^\pm$ mesons, as well as the K^{*0} decay vertex, are preserved in this embedding procedure, the study mainly validates the basic selection requirements and the performance of the multivariate classifiers (BDT$_1$ and BDT$_2$). Figure 5.20 shows the $\theta_{p_{\rm missing}}$ distributions at three stages: pre-BDT$_1$ (basic two candidates selection), the BDT$_1$ stage (BDT$_1$ > 0.9), and the BDT$_2$ stage (η(BDT$_2$) > 0.85 with BDT$_1$ > 0.9), for the signal $B^0 \to K^{*0}\nu\bar{\nu}$, the embedded simulation, and data. At the BDT$_1$ stage, the selection efficiency is 24.6% ± 0.5% for data and 26.8% ± 0.1% for simulation, giving a ratio of 0.919 ± 0.021. This indicates a modest data–simulation discrepancy. At the BDT$_2$ stage, the efficiency is 14.3% ± 0.4% for data and 14.7% ± 0.1% for simulation, yielding a ratio of 0.98 ± 0.03.
Figure 5.21 shows the K^{*0} invariant mass, the rest-of-event beam-constrained mass ($M_{\rm bc,ROE}$), and the reconstructed $q^2_{\rm rec}$ for events with BDT$_2$ > 0.85. The distributions demonstrate consistent behavior between

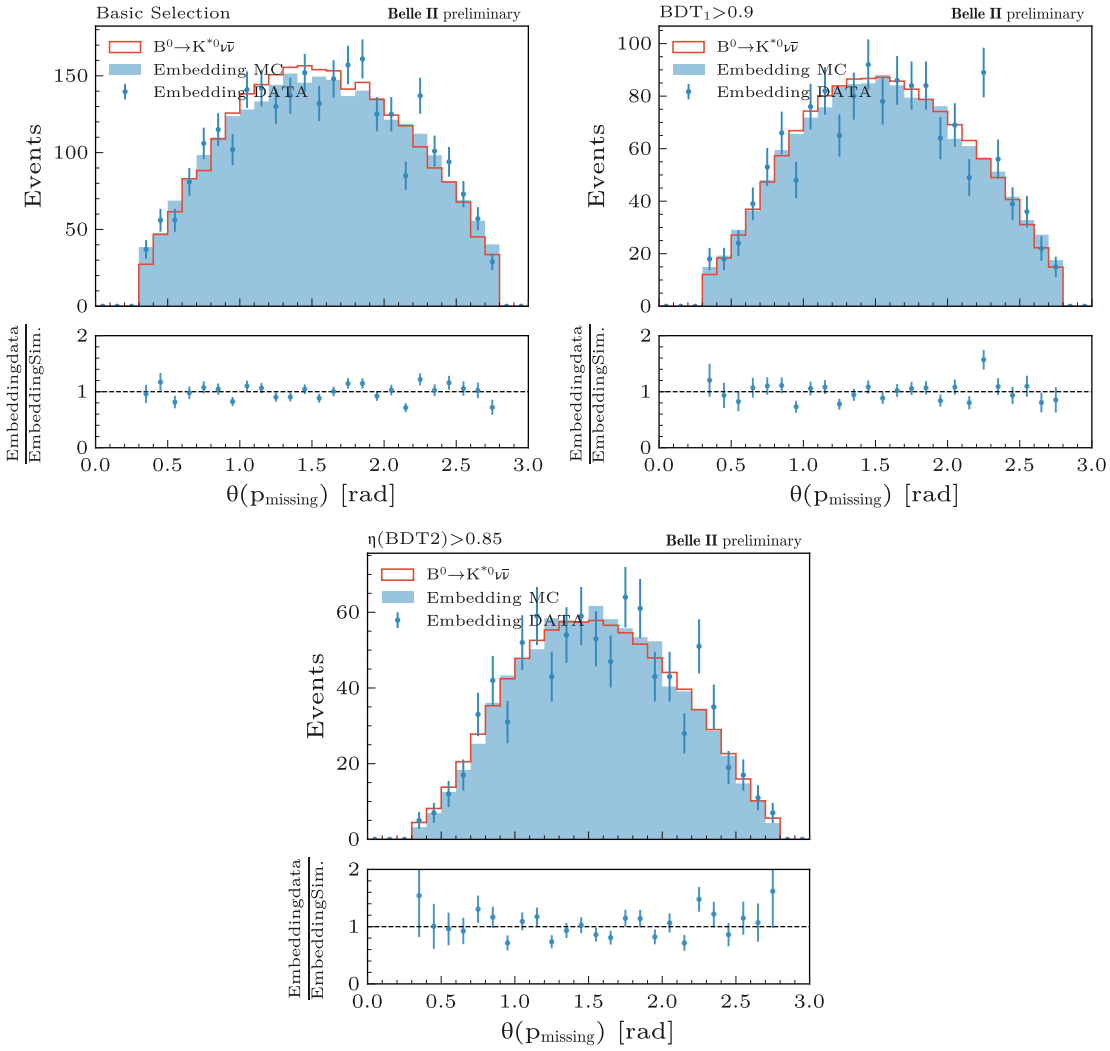

Figure 5.20: Validation of the $\theta_{p_{\text{missing}}}$ distribution at three selection stages: pre-BDT$_1$(upper left), BDT$_1$ (upper right), and $\eta(\text{BDT}_2) > 0.85$ (bottom) for $B^0 \to K^{*0}\nu\bar{\nu}$ comparison between embedded data, embedded simulation, and nominal signal simulation.

embedded simulation and data, confirming that the embedding procedure reproduces the kinematics and event environment reliably.

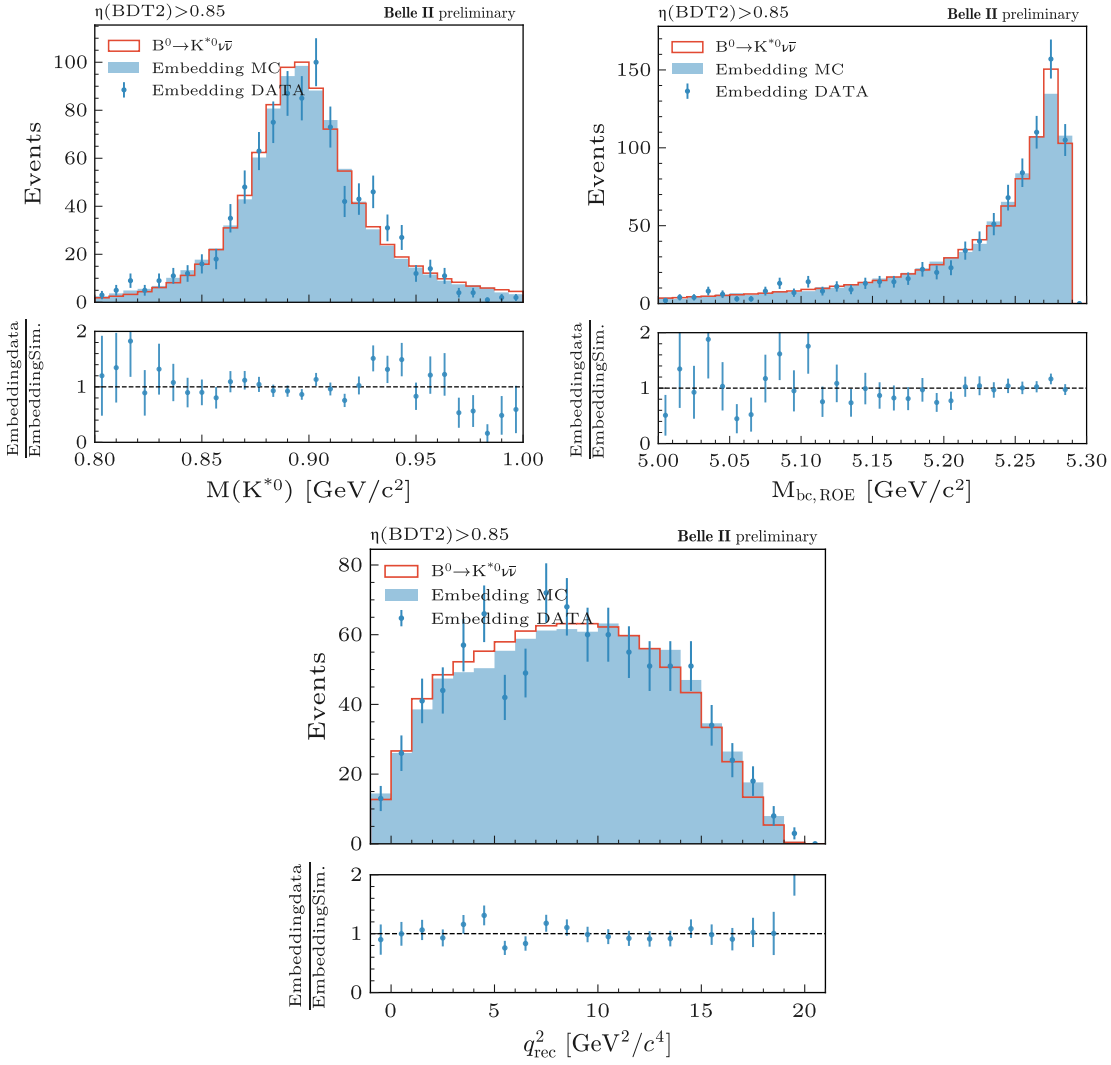

Figure 5.21: Distributions of the K^{*0} invariant mass, $M_{\mathrm{bc,ROE}}$, and q^2_{rec} for $B^0 \to K^{*0}\nu\bar{\nu}$ candidates with $\mathrm{BDT}_2 > 0.85$: comparison of embedded data, embedded simulation and the nominal signal simulation.

We further examine embedded data, embedded simulation, and nominal signal simulation in the signal search region (Fig. 5.22). The selection efficiency in this region is 4.03% $\pm$ 0.25% for embedded data and 4.05% $\pm$ 0.07% for embedded simulation, yielding a ratio of 0.994 $\pm$ 0.064. This indicates excellent consistency between embedded simulation and data.

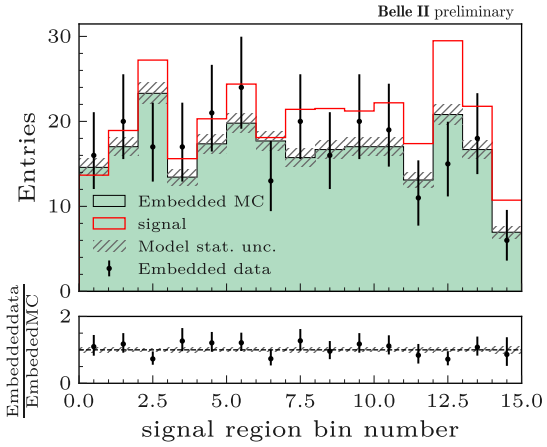

Figure 5.22: Distributions in the signal search region for $B^0 \to K^{*0}\nu\bar{\nu}$: comparison of embedded data (combined $\mu^+\mu^-$ and e^+e^- channels), embedded simulation, and nominal signal simulation.

5.4 Background Modeling and Validation

After the application of the second classifier, a comprehensive background study is performed for the $B^0 \to K^{*0}\nu\bar{\nu}$ channel. This includes the evaluation of continuum modeling, the treatment of specific $B\bar{B}$ background contributions at different levels, and validation with dedicated control samples. These studies provide the basis for a consistent definition of sideband regions and their use in the overall background validation strategy.
To this end, two complementary sidebands are defined: a BDT sideband and a mass sideband.

- The BDT sideband is chosen as $0.85 < \eta(\mathrm{BDT}_2) < 0.95$. The choice of this range is motivated by the following considerations:
  - Corrections for continuum modeling are derived from a dedicated classifier trained using events with $\eta(\mathrm{BDT}_2) > 0.85$.
  - Studies of $B\bar{B}$ background components in the regions $\eta(\mathrm{BDT}_2) > 0.85$ and $\eta(\mathrm{BDT}_2) > 0.95$ show consistent results.
- In addition, because the K^{*0} is reconstructed from its $K\pi$ decay products, a mass sideband is constructed to validate the background independently of the signal peak. This sample follows the same selection criteria as the nominal control sample, with the additional requirement that the reconstructed K^0 mass lies outside the signal window of [0.9, 1.0] GeV/c^2. The mass sideband thus provides an independent dataset for robust cross-checks of the background modeling.

5.4.1 Continuum Background Correction

Experiment 25 generation correction (simulation beam-energy mismatch) Because BDT_2 includes run-dependent variables that depend explicitly on the beam energy, we examined their distributions separately for each experiment. This study revealed systematic shifts in $M\mathrm{bc}, \mathrm{ROE}$
These variables are defined as follows:

$$M_{\mathrm{bc,ROE}} = \sqrt{E_{\mathrm{ROE}}^2 - |\vec{p}_{\mathrm{roe}}|^2}, \tag{5.4}$$

with $E_{\mathrm{ROE}} = E_{\mathrm{cms}}/2$.
Further investigation showed that the Experiment 25 off-resonance simulation was generated at a center-of-mass energy of $E_{\mathrm{cms}} \simeq 10.525$ GeV, approximately 8 MeV higher than the corresponding data $E_{\mathrm{cms}} \simeq 10.517$ GeV. This mismatch directly explains the observed shifts in the beam-energy dependent variables.
The top panels of Fig. 5.23 compare data and simulation for Experiments 12, 18, and 25. While Experiments 12 and 18 show good data–simulation agreement, Experiment 25 exhibits clear offsets in $M_{\mathrm{bc,ROE}}$, consistent with the beam-energy mismatch.
The corresponding corrections are applied to the experiment 25 off-resonance simulation only: for $M_{\mathrm{bc,ROE}}$, the correction is

$$\Delta M_{\mathrm{bc}} = \sqrt{(10.525/2)^2} - \sqrt{(10.517/2)^2} \simeq 4 \text{ MeV}/c^2, \tag{5.5}$$

The bottom panels of Fig. 5.23 show the same comparisons after applying the corrections. The Experiment 25 simulation distributions are shifted into alignment with the data, bringing them into consistency with Experiments 12 and 18.

Off-resonance application correction (BDT_2 training energy mismatch) The beam energies differ between the $\Upsilon(4S)$ and off-resonance samples by $\Delta E_{\mathrm{cms}} \simeq 60$ MeV/c^2 (see Fig. 5.24). Since BDT_2 was trained using $M_{\mathrm{bc,ROE}}$ values computed with the $\Upsilon(4S)$ beam energy, applying it directly to off-resonance data would introduce a mismatch. To correct for this, $M_{\mathrm{bc,ROE}}$ in the off-resonance sample is shifted by +30 MeV/c^2, aligning it with the $\Upsilon(4S)$ energy during BDT_2 evaluation. Afterward, $M_{\mathrm{bc,ROE}}$ is restored to its original value.
In practice, this procedure (such as $M_{\mathrm{bc,ROE}}$)is implemented as follows:

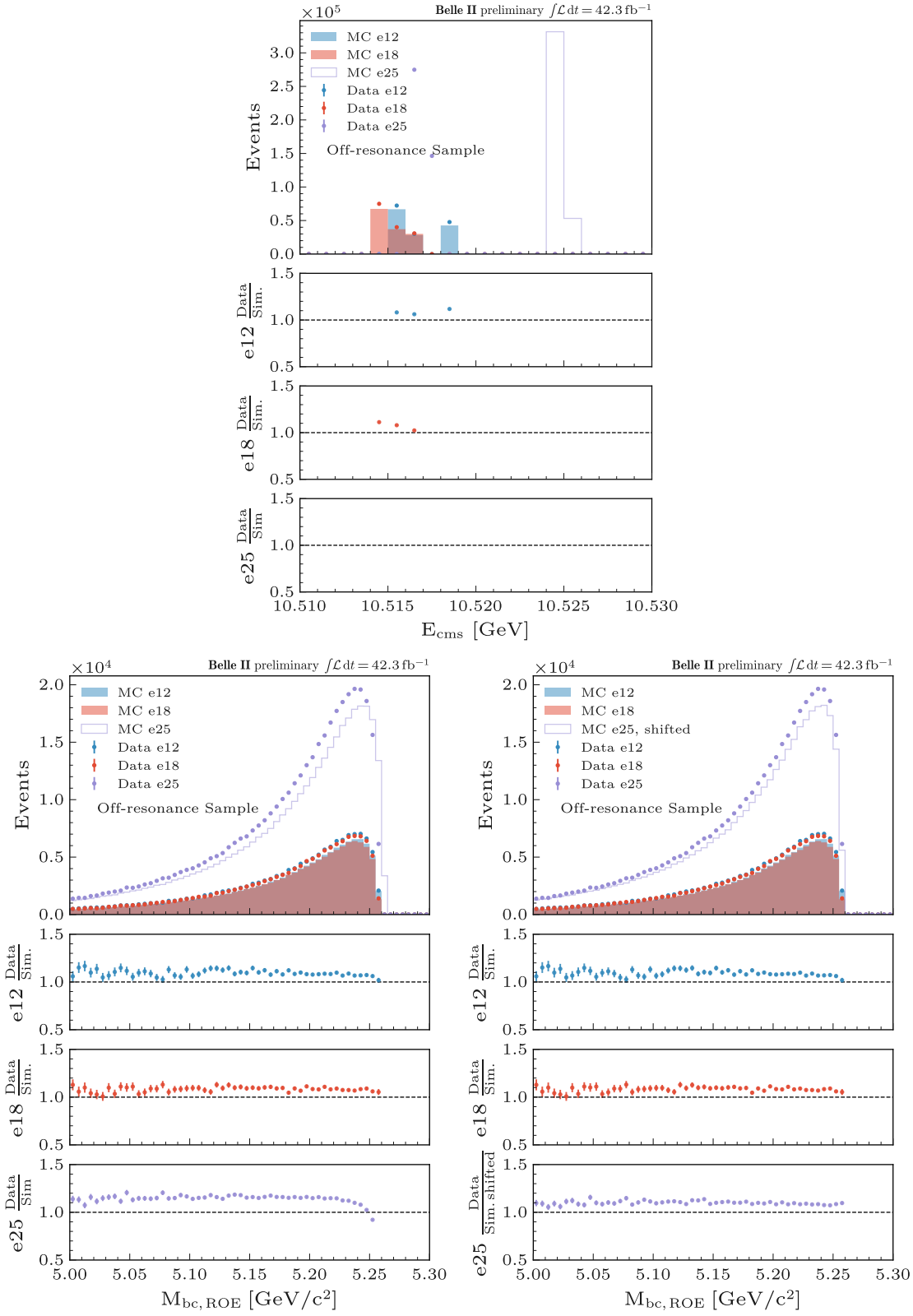

Figure 5.23: Comparison of off-resonance data and simulation for Experiments 12, 18, and 25, illustrating the effect of the beam-energy mismatch in Experiment 25. The variables shown are E_{cms} (upper) $M_{\text{bc,ROE}}$ (bottom left). Before correction, Experiment 25 simulation shows systematic shifts relative to data, while Experiments 12 and 18 remain consistent. After applying corrections $\Delta M_{\text{bc}} \simeq 4$ MeV/c^2 (bottom right), the Experiment 25 simulation aligns with the data, achieving consistency across experiments.

a. Save the original value of $M_{\mathrm{bc,ROE}}$.

b. Apply a +30 MeV/c^2 shift to $M_{\mathrm{bc,ROE}}$.

c. Evaluate the BDT$_2$ output.

d. Restore the original value of $M_{\mathrm{bc,ROE}}$.

This procedure ensures that the BDT$_2$ output remains consistent with the energy used during training, while preserving the original off-resonance kinematics for subsequent analysis.

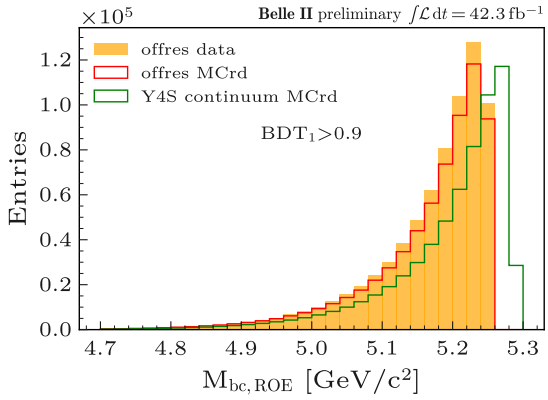

Figure 5.24: Comparison of off-resonance data (orange filled), off-resonance simulation (red), and continuum simulation at the $\Upsilon(4S)$ (green) for events with BDT$_1 > 0.9$. The off-resonance simulation follows the data shape but differs in normalization, whereas the $q\bar{q}$ in $\Upsilon(4S)$ simulation exhibits clear shifts in $M_{\mathrm{bc,ROE}}$ (+30 MeV/c^2), due to the beam-energy difference between the $\Upsilon(4S)$ and off-resonance samples.

Figure 5.25 demonstrates the effect of applying BDT$_2$ with and without the correction. If BDT$_2$ is applied directly to the off-resonance data and simulation, their distributions remain mutually consistent but are suppressed relative to expectations. Once the correction procedure is applied, the distributions recover the proper shapes, confirming the validity of the method.

BDT$_{\mathrm{C}}$ Training In the K^{*0} channel, the $q\bar{q}$ continuum background contributes about 37% in the signal region and 16% in the most sensitive region. The validation procedure and the treatment of data-simulation discrepancies follow the same strategy as in the K^+ channel. For the reweighting, a dedicated BDT$_C$ is trained with the XGBoost algorithm, using 42.3 fb^{-1} of off-resonance data ($\sim 1 \times 10^5$ events), treated as the classifier `signal`, and the corresponding fourfold-larger off-resonance simulation sample ($\sim 4 \times 10^5$ events), treated as `background`. Both data and simulation samples were randomly split, with three-quarters used for training and one-quarter reserved for validation. The input variables consisted of the 39 BDT$_2$ training variables, augmented by q^2_{rec} and the BDT$_2$ score. Events were selected with the requirement $\eta(\mathrm{BDT}_2) > 0.85$. The hyperparameters were kept the same as those in the K^+ channel (see Tab. 4.8).
Possible overfitting is checked by comparing the classifier output for the training and validation sub-samples. Fig. 5.26 (left) shows the output score of the BDT$_{\mathrm{C}}$. The practical effect of the reweighting procedure is also illustrated in Fig. 5.26 (right), where the BDT$_{\mathrm{C}}$ score is used as an event weight for the simulated $e^+e^- \to q\bar{q}$ and $\tau^+\tau^-$ components. Without reweighting, discrepancies are visible between off-resonance data and simulation, particularly in global event-shape observables such as the Fox–Wolfram moment R_2 (Fig. 5.27 (left panel). After applying the BDT$_{\mathrm{C}}$ weights (Fig. 5.27 right panels), the simulation agrees more closely with the data. A residual normalization discrepancy of about 3% remains in the $\eta(\mathrm{BDT}_2) > 0.85$ region of the off-resonance sample. The impact of the BDT$_{\mathrm{C}}$ procedure on the two-dimensional signal region search bins is illustrated in Fig. 5.28. The left panel compares off-resonance data to the uncorrected simulated continuum background. Shape mismatches are visible, reflecting the limited modeling accuracy

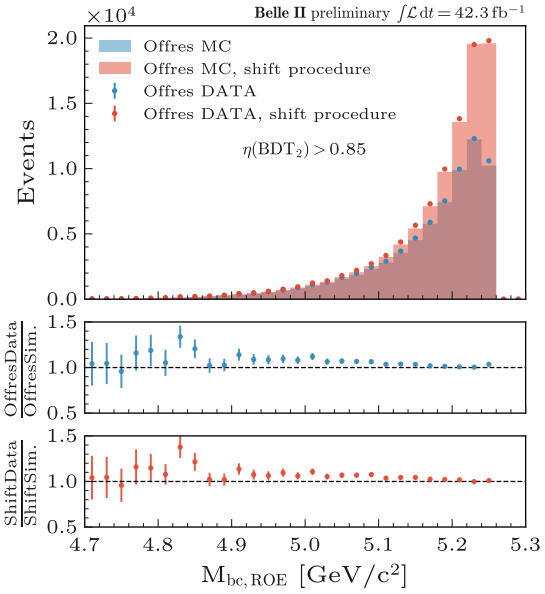

Figure 5.25: Effect of the beam-energy correction procedure when applying BDT2 to off-resonance data and simulation. Left panels: $M_{\mathrm{bc,ROE}}$. Without correction (blue), data and simulation are mutually consistent but exhibit suppressed shapes. After applying the correction (red), the shapes are restored, validating the procedure.

of the raw simulation. After applying the $\mathrm{BDT_C}$ weights (upper right panel in Fig. 5.28), the agreement between data and simulation is improved across the 2D plane. However, a residual normalization difference of about 10% remains across the 2D plane, with data systematically overshooting the weighted simulation. To account for this effect, a global normalization factor is applied in addition to the $\mathrm{BDT_C}$ weights. The result, shown in the bottom panel of Fig. 5.28, demonstrates that the combination of $\mathrm{BDT_C}$ shape reweighting and normalization correction brings the simulation into excellent agreement with the off-resonance data in both overall rate and bin-by-bin distributions.

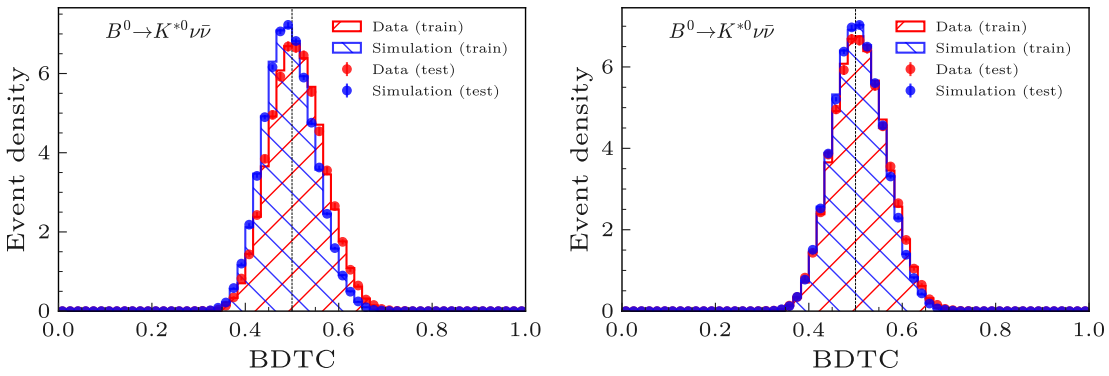

Figure 5.26: Check of $\mathrm{BDT_C}$ training stability. Left: classifier output for training and validation samples, showing no sign of overfitting. Right: effect of applying $\mathrm{BDT_C}$ weights to simulated $q\bar{q}$ and $\tau^+\tau^-$ components.

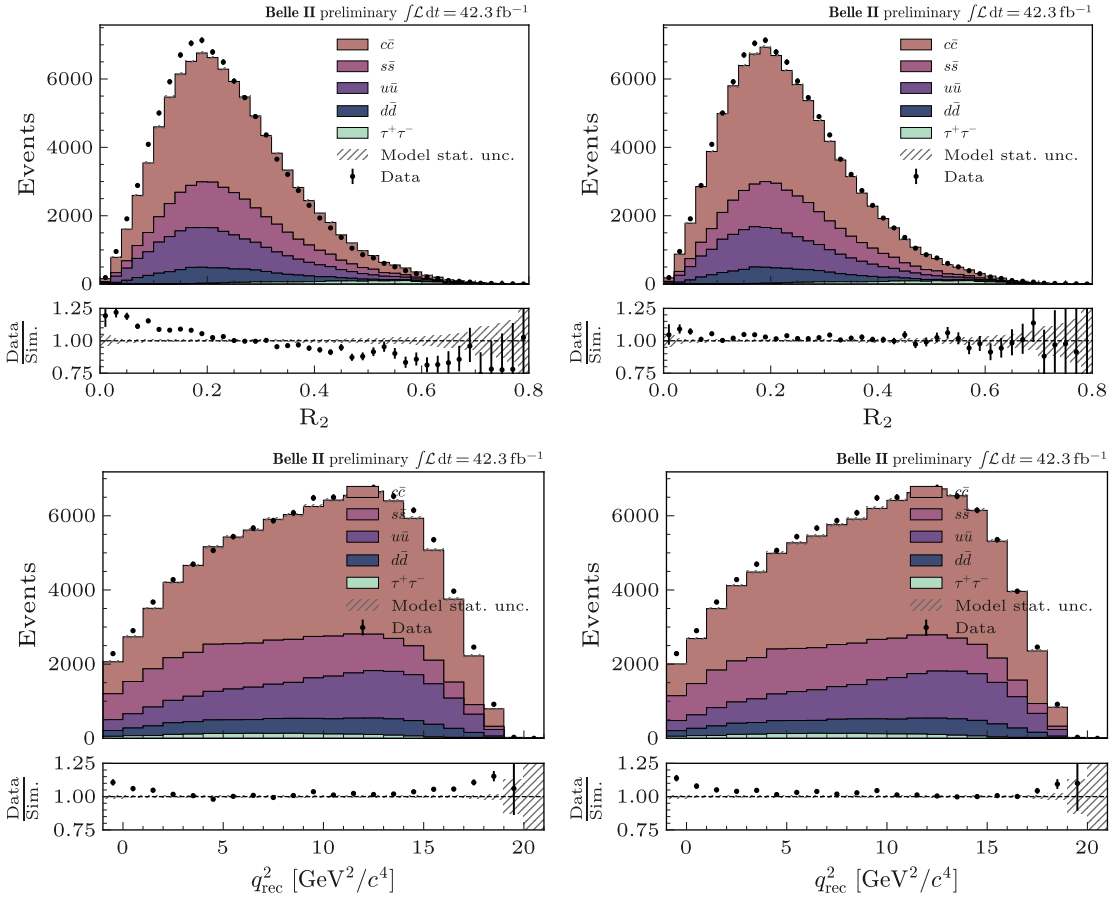

Figure 5.27: Impact of $\mathrm{BDT_C}$ reweighting on global observables. Left panels: comparison of off-resonance data and simulated background without reweighting, showing discrepancies in event-shape observables (R_2, reconstructed q^2_{rec}). Right panels: after applying $\mathrm{BDT_C}$ weights, the simulation is in better agreement with data.

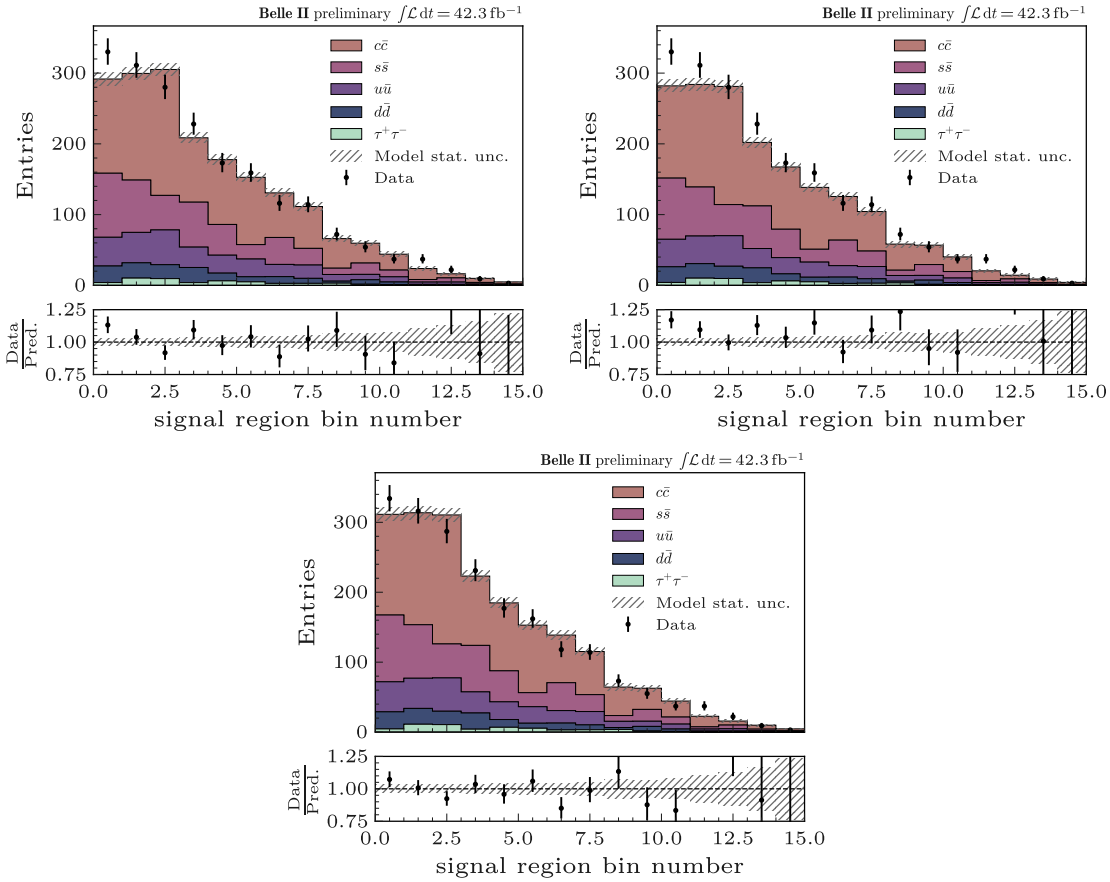

Figure 5.28: Effect of $\mathrm{BDT_C}$ reweighting in the two-dimensional signal region search bins. Upper left: unweighted simulation shows significant mis-modeling. Upper right: after applying $\mathrm{BDT_C}$ weights, shape agreement improves, but a $\sim$ 10% normalization difference remains. Bottom: applying an additional global normalization factor yields excellent agreement between data and simulation.

5.4.2 $B\bar{B}$ Background Corrections

To validate the background modeling, we first compare data and simulation for the q^2_{rec} and $\eta(\text{BDT}_2)$ in the BDT_2 sideband, where the signal region remains blinded. Figure 5.29 shows q^2_{rec} and $\eta(\text{BDT}_2)$ distributions in the sideband ($0.85 < \eta(\text{BDT}_2) < 0.95$). Data (points with error bars) are compared to stacked background components, including $B\bar{B}$ and corrected continuum, with ratios shown below each distribution. In this comparison, the $q\bar{q}$ background is corrected using the BDT_c weighting procedure, with an additional 3% normalization factor applied. After this correction, a residual normalization difference of about 5% remains. The sideband comparison confirms that the continuum background is well modeled, allowing us to turn to a more detailed study of the $B\bar{B}$ background.

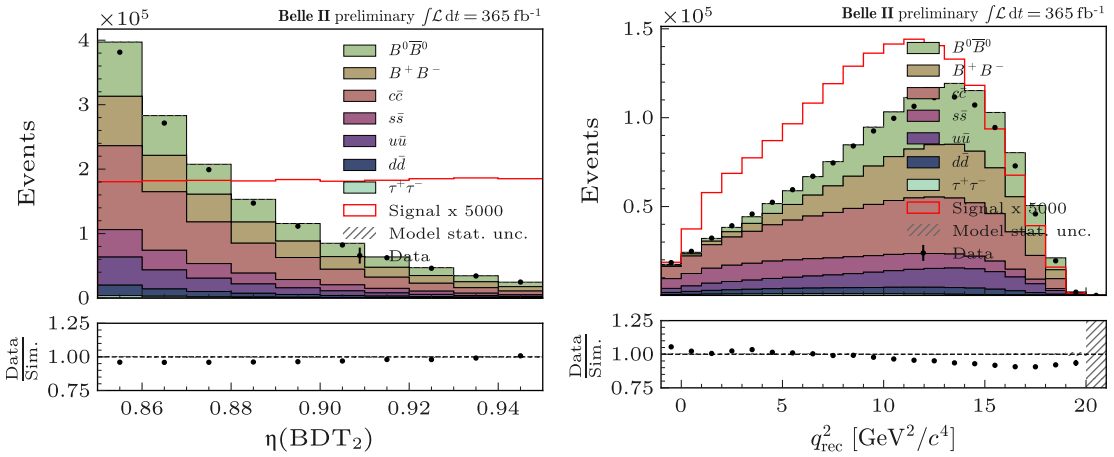

Figure 5.29: Comparison of data (points with error bars), stacked $B\bar{B}$ and continuum background components (filled histograms), and signal scaled by a factor of 5000 (red outline) in the BDT_2 sideband ($0.85 < \eta(\text{BDT}_2) < 0.95$). Left: $\eta(\text{BDT}_2)$ distribution. Right: q^2_{rec} distribution. Ratios of data to the summed background prediction are shown in the lower panels. Continuum background is corrected using the BDT_C weights with an additional 3% normalization, leaving a residual ∼5% normalization difference. The total $\Upsilon(4S)$ data luminosity is used.

In the signal region, the $B\bar{B}$ background constitutes the dominant contribution. In the signal region, the $B\bar{B}$ background constitutes about 63% of the total background, rising to 84% in the most sensitive region. The composition of the $B\bar{B}$ background shows approximately equal fractions from B^+B^- and $B^0\bar{B}^0$ events, suggesting no strong channel dependence in the overall contamination.
A more detailed investigation is performed to identify the sources of reconstructed candidates that mimic the signal K^{*0} in the signal region (Fig. 5.30). A significant component arises from a combinatorial background, where the two tracks forming the K^{*0} candidate originate from different B mesons in the $\Upsilon(4S) \to B\bar{B}$ decay. Since the fake candidate is effectively built across the $B\bar{B}$ pair, this background is referred to as $\Upsilon(4S)$ fakes. This source contributes about 19% (27%) of the total $B^0\bar{B}^0$ background in the signal region (most sensitive region). The composition of reconstructed K^{*0} candidates in simulation is shown in Fig. 5.30 with pie charts illustrating the particle sources in the signal region (top) and most sensitive region (bottom), separately for B^0 and B^+.
To investigate whether different categories exhibit distinctive shapes in the variables used for fitting, the distributions of q^2_{rec} and the $\eta(\text{BDT}_2)$ with each particle category are examined in the signal region. Figure 5.31 compares the $\eta(\text{BDT}_2)$ and q^2_{rec} distributions for the main categories identified in Fig. 5.30. The comparison reveals no significant differences among the categories, indicating that the shapes of q^2_{rec} and $\eta(\text{BDT}_2)$ are largely insensitive to the specific particle type responsible for the fake K^{*0}.
For the $\Upsilon(4S)$ fake background, it is not possible to unambiguously determine which of the two mesons should be identified as the signal B. For the $B\bar{B}$ background, we examine the composition in the signal region. We find that a large fraction originates from semileptonic decays. Specifically, about 45% of the

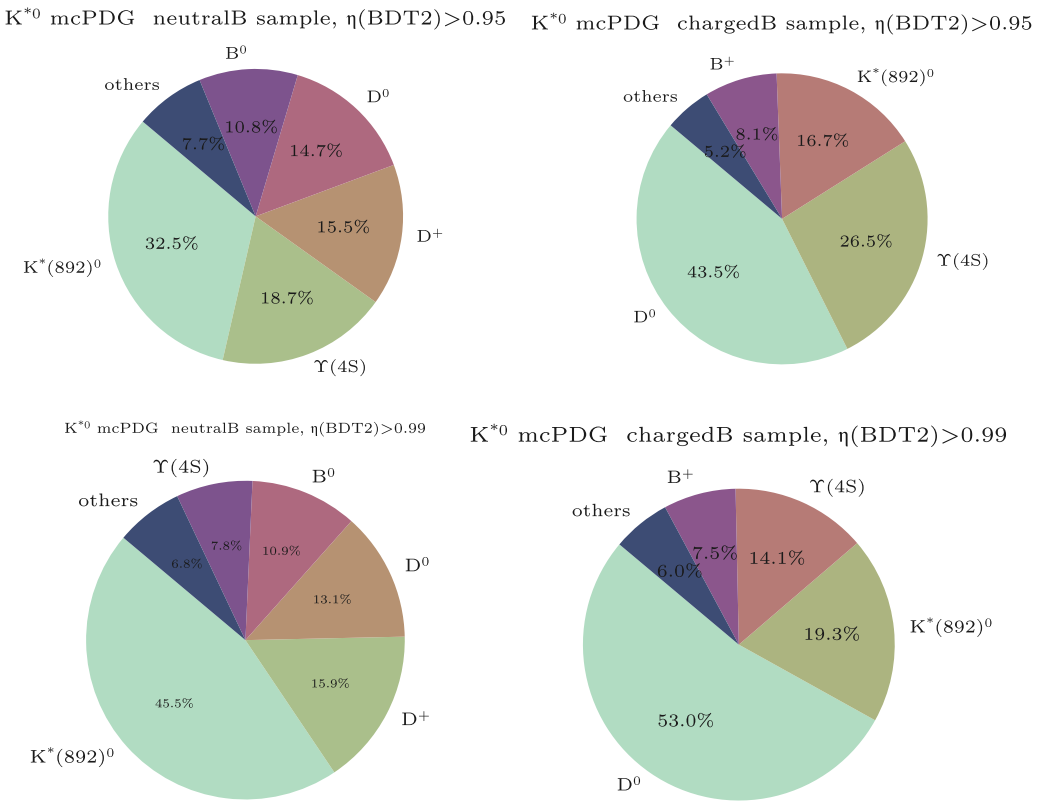

Figure 5.30: Composition of reconstructed K^{*0} candidates in simulated $B^0\bar{B}^0$ (left) and B^+B^- (right) samples. Upper row: signal region ($\eta(\mathrm{BDT}_2) > 0.95$). Lower row: most sensitive region ($\eta(\mathrm{BDT}_2) > 0.99$). Each pie chart shows the particle sources contributing to the K^{*0} candidate, highlighting the relative contributions of combinatorial ($\Upsilon(4S)$ fakes), true K^{*0}, and D-meson–induced fakes.

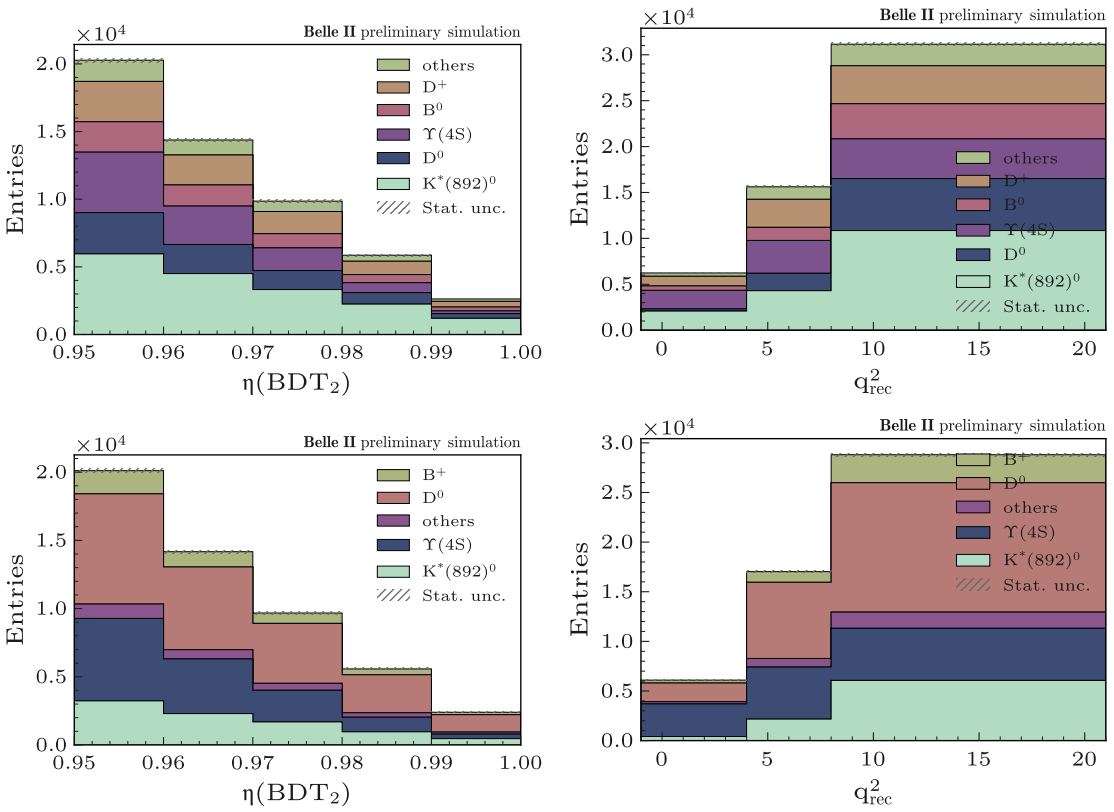

Figure 5.31: Distributions of $\eta(\mathrm{BDT}_2)$ (left) and q^2_{rec} (right) for the categories contributing to the reconstructed K^{*0} candidates. Upper panels: $B^0\bar{B}^0$ background in the signal region ($\eta(\mathrm{BDT}_2) > 0.95$). Lower panels: B^+B^- background in the signal region. The distributions are compared using the binning of the signal search. No distinctive features are observed across categories, indicating that the shapes of q^2_{rec} and $\eta(\mathrm{BDT}_2)$ are largely insensitive to the fake K^{*0} source.

$B^0\bar{B}^0$ background and 61% of the B^+B^- background can be attributed to the following modes:

- $B^0 \to D^*(2010)^+ \ell^- \nu$ and $B^0 \to D^+ \ell^- \nu$,
- $B^+ \to D^*(2007)^0 \ell^+ \nu$ and $B^+ \to D^0 \ell^+ \nu$.

The remaining part of the $B\bar{B}$ background consists mainly of decays involving hadrons accompanied by charmoninum or other composite particles. The backgrounds that mimic the signal K^{*0} can be classified into three main categories, each requiring a dedicated treatment strategy.

a. **Combinatorial fakes (including $\Upsilon(4S)$ contributions):** In this category, the two tracks forming the K^{*0} candidate are associated with the B mesons produced at the $\Upsilon(4S)$. They may originate from the same B meson (directly from the B or via its intermediate composite particles) or from two different mesons in the $B\bar{B}$ pair. To study and validate these contributions, we employ the K^{*0} mass sideband sample, which provides a control region enriched in combinatorial candidates and allows us to check consistency with the observed background in the signal region.

b. **True K^{*0} mesons from B decays:** These backgrounds correspond to genuine K^{*0} mesons produced in B decays, where other particles in the final state are not reconstructed. The missing particles are not limited to the two neutrinos expected in the signal but may also include neutral hadrons, leading to an apparent signal-like topology. Decay modes of particular relevance are

   - $B^0 \to K^{*0} K^0 \bar{K}^0$,
   - $B^0 \to K^{*0} n\bar{n}$.

c. **Charm-induced fakes (D-meson decays):** In these cases, the B meson decays into a charm meson, and the subsequent decay of the D produces two tracks that mimic a K^{*0} candidate. To investigate this contribution, we construct a dedicated control sample based on decays of the type

   - $D \to K^{(*)} X$,

   which allows us to test whether such charm-induced fakes reproduce the observed background composition in the signal region.

5.4.2.1 Background Validation with K^{*0} Mass Sideband

As an additional cross-check of the $B\bar{B}$ background modeling, we define a K^{*0} mass sideband sample that allows validation outside of the signal peak. This sample uses the same selection criteria as the nominal sample, but requires the reconstructed K^{*0} mass to lie outside the nominal signal window, which is [0.65, 0.8] GeV/c^2 and [1.0, 1.4] GeV/c^2. This mass sideband allows for orthogonal cross-checks of the background modeling. A more detailed treatment is needed because the reconstructed K^{*0} mass enters into several steps of the candidate selection and multivariate discriminants. In particular, the K^{*0} mass is used as an input feature in the BDT used for candidate arbitration ($\mathrm{BDT_{CA}}$) to select the best two candidates. To preserve consistency with the nominal analysis while minimizing modifications, we replace the reconstructed K^{*0} mass with its nominal value [4] when evaluating sideband candidates. The first-stage discriminator, $\mathrm{BDT_1}$, which relies mainly on event-shape variables and does not use the K^{*0} mass explicitly, is applied without changes, and the same selection requirement $\mathrm{BDT_1} > 0.9$ is imposed.
For the second-stage discriminator, $\mathrm{BDT_2}$, the K^{*0} mass is one of the most discriminating variables in the training. Therefore, the nominal $\mathrm{BDT_2}$ model cannot be applied directly to the sideband sample. Instead, we retain the nominal $\mathrm{BDT_2}$ training, but when evaluating predictions on sideband candidates, we replace the reconstructed $M(K^{*0})$ with the nominal constant value. This procedure ensures that the discriminating power and efficiency remain consistent with the nominal selection. Efficiency maps are derived using these modified $\mathrm{BDT_2}$ predictions. The same substitution strategy is applied to the candidate-level discriminator $\mathrm{BDT_C}$, ensuring consistency between the nominal and sideband analyses. Fig. 5.32 shows the distributions of the reconstructed K^{*0} mass, q^2_{rec}, and $\eta(\mathrm{BDT_2})$ in the region $\eta(\mathrm{BDT2})$ >0.85, with the BDT_C weight

applied to continuum events in the full data set and in the corresponding run-dependent simulation sample. Since the sideband excludes the K^{*0} signal peak, the nominal signal region is not blinded. Overall, the data and simulation are generally consistent, with a slight undershoot of the data compared to the simulation in the high-mass region.

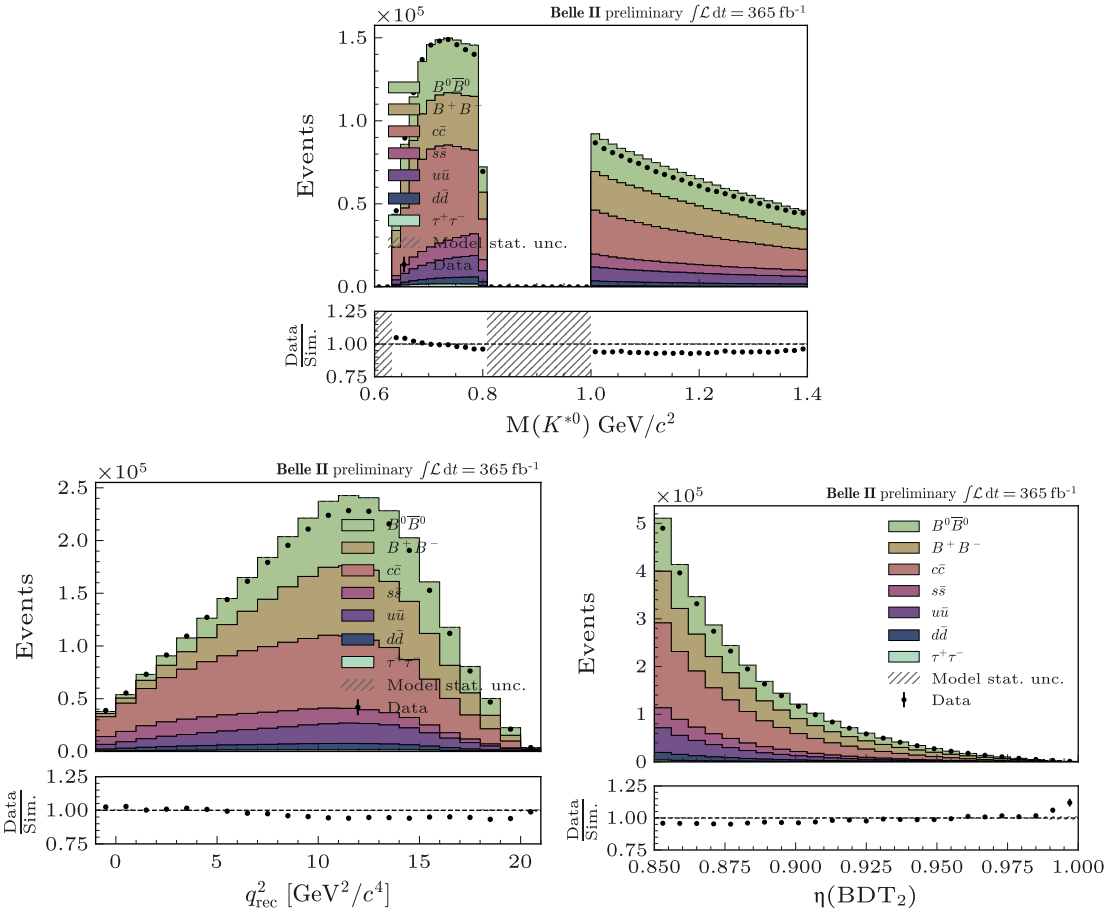

Figure 5.32: Distributions in the K^{*0} mass sideband sample with $\eta(\mathrm{BDT}_2) > 0.85$. Upper: reconstructed K^{*0} mass. Bottom left: q^2_{rec}. Bottom right: $\eta(\mathrm{BDT}_2)$. Data (points with error bars) are compared to the run-dependent simulation (stacked histograms), with continuum background corrected using $\mathrm{BDT_C}$ weights. Ratios of data to simulation are shown in the lower panels. The comparison shows good overall agreement, with a slight undershoot of data relative to simulation in the high-mass region. The total $\Upsilon(4S)$ data luminosity is used.

The composition of the $B\bar{B}$ background in the mass sideband sample is examined to assess its consistency with the nominal background model. In the $B^0\bar{B}^0$ sample, approximately 28% of the candidates arise from mis-reconstructed $\Upsilon(4S)$ decays, while the remaining 72% originate from single-B decays. Within the latter category, semileptonic processes provide the dominant contributions, with about 20% from $B^0 \to D^*(2010)^+\ell^-\nu$ and about 17% from $B^0 \to D^+\ell^-\nu$.
A similar pattern is observed in the B^+B^- sample, where 32% of the background is attributed to $\Upsilon(4S)$ fakes and 67% to single-B decays. The leading channels in this case are $B^+ \to D^*(2007)^0\ell^+\nu$ (30%) and $B^+ \to D^0\ell^+\nu$ (35%).
To identify which reconstructed particles mimic the signal K^{*0}, we study the composition in both the nominal signal region and the most sensitive region. Fig. 5.33 shows pie charts of the reconstructed K^{*0} composition for $B^0\bar{B}^0$ (left) and B^+B^- (right), comparing the signal region (top row, $\eta(\mathrm{BDT}_2) > 0.95$) and the most

sensitive region (bottom row, $\eta(\mathrm{BDT}_2) > 0.99$). After removing the genuine K^{*0} contribution (for instance, restricting to sideband-like candidates), the composition of the sideband sample is consistent with that of the nominal sample (Fig. 5.30). This agreement indicates that the sideband provides a reliable and orthogonal validation of the background model.

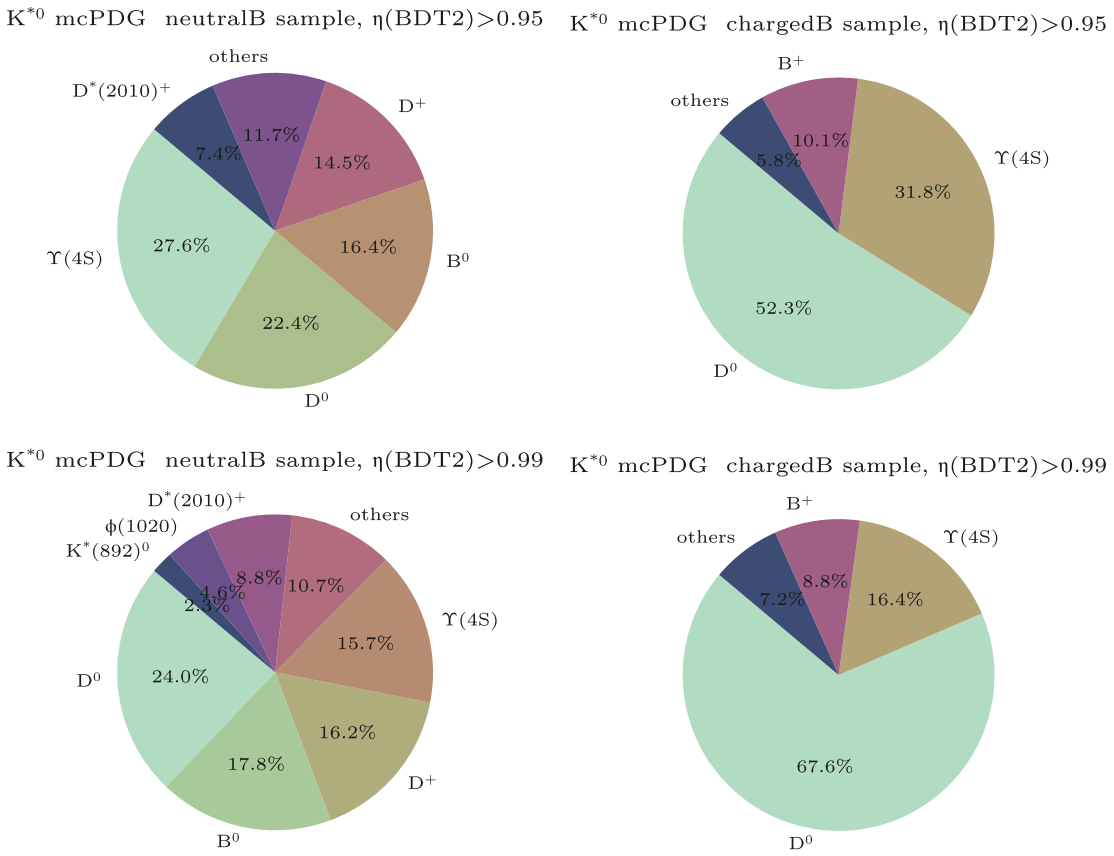

Figure 5.33: Composition of reconstructed K^{*0} candidates in the mass sideband. Left: $B^0\bar{B}^0$. Right: B^+B^-. Top row: signal region ($\eta(\mathrm{BDT}_2) > 0.95$). Bottom row: high-sensitivity region ($\eta(\mathrm{BDT}_2) > 0.99$).

Within the single-B component, semileptonic channels dominate, with about 30% from $B^+ \to D^*(2007)^0\ell^+\nu$ and about 35% from $B^+ \to D^0\ell^+\nu$. A detailed comparison with the nominal sample, after removing the true K^{*0} contribution (mostly located within $0.8 < M(K\pi) < 1.0$ GeV), shows that the mass sideband composition closely matches the nominal sample. This agreement is observed both in the overall category fractions and in the distributions of these categories across the signal and high-sensitivity regions, as shown in Fig. 5.34 (right panel), indicating a robust consistency between the two samples.
For the $B^0\bar{B}^0$ background, the mass sideband is generally similar to the nominal sample, but notable differences exist. For the $B^0\bar{B}^0$ background, the mass sideband composition is broadly similar to the nominal sample, but with fewer D^+ fakes (14.5% vs. 23%) and more D^{*+} decays (7.4% vs. 2%), primarily appearing in the lower mass region ([0.65, 0.8] GeV/c^2), as shown in Fig. 5.34 (left panel). These differences manifest as smooth variations in $\eta(\mathrm{BDT}_2)$ and q^2_{rec} distributions. A closer examination of the D^{*+} decays shows that the dominant channel is $D^{*+} \to D^0\pi^+$, where the kaon from the D^0 and the pion from the D^{*+} can combine to form a fake K^{*0}. Dedicated BDT_2 variables, as discussed in Section 5.1.5.5, are designed to suppress backgrounds from such three-track final states. While this suppression is effective in the nominal sample, it is less pronounced in the sideband. Importantly, this difference is physically motivated, and the sideband composition is not expected to perfectly match the nominal sample.

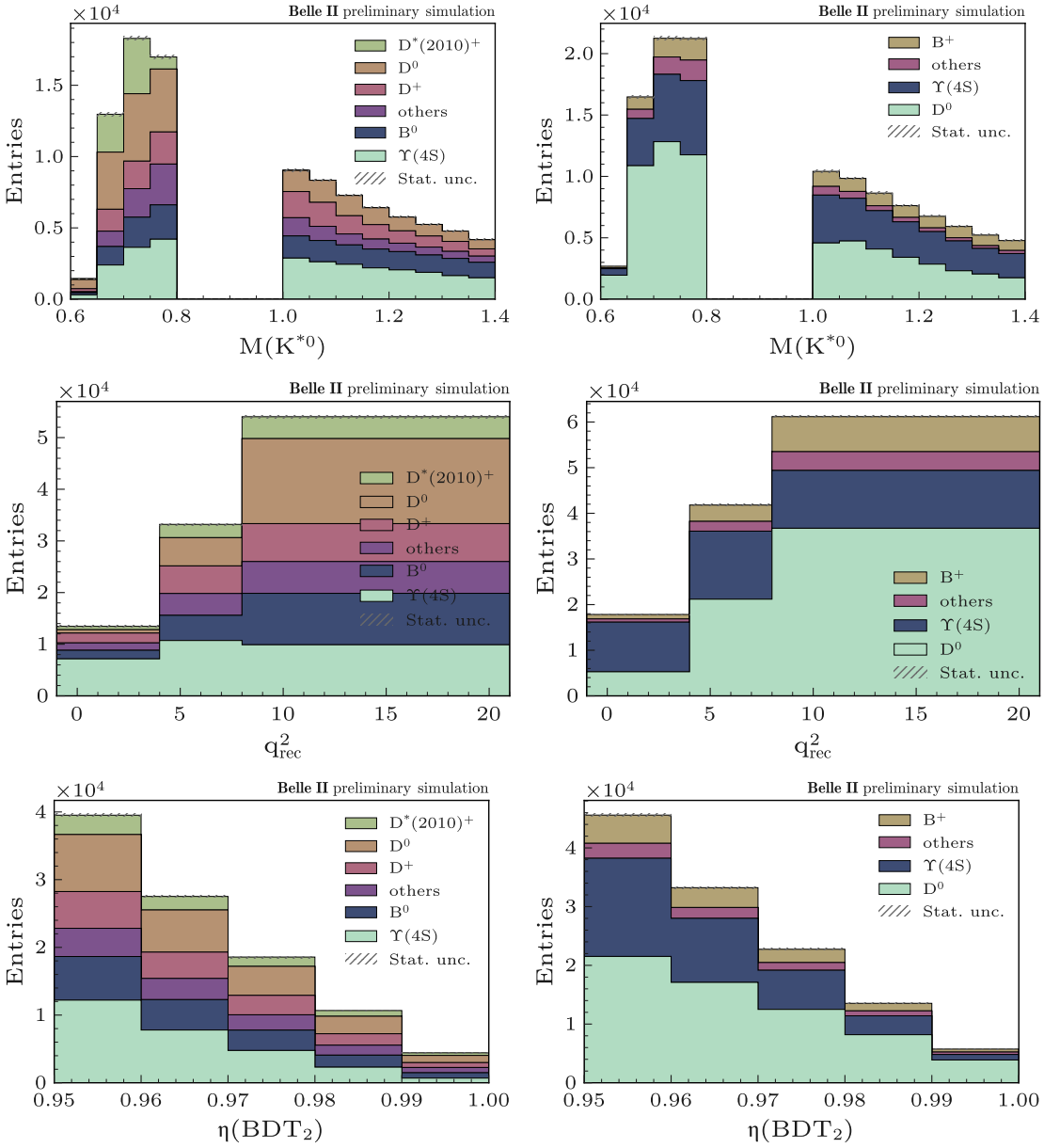

Figure 5.34: Distributions of reconstructed K^{*0} mass (top), q^2_{rec} (middle), and $\eta(\text{BDT}_2)$ (bottom) for the categories contributing to fake K^{*0} candidates in the mass sideband. Left: $B^0\bar{B}^0$. Right: B^+B^-.

5.4.2.2 Modeling of $B^0 \to K^{*0}K^0\bar{K}^0$

The decay $B^0 \to K^{*0}K^0\bar{K}^0$ has not yet been measured experimentally, so its modeling in simulation relies on theoretical assumptions. In the Belle II simulation, the decay is modeled as a phase-space process including several resonant components with assumed branching fractions. To obtain a more realistic description, we study the related decay $B^0 \to K^{*0}K^0_S K^0_S$.
The reconstruction follows the general selection strategy used in rare B-decay analyses. Charged tracks are required to satisfy impact parameter and momentum requirements. K^0_S candidates are reconstructed from two oppositely charged tracks with mass and vertex constraints, while K^{*0} candidates are formed from a kaon–pion pair with (0.8 1.0) GeV/c^2. The B^0 candidates are built from one K^{*0} and two K^0_S mesons. To suppress continuum background, standard event-shape variables and kinematic selections are applied, including beam-constraint mass larger than 5.27 GeV/c^2, $|\Delta E| < 0.15$ GeV, $R_2 < 0.4$, and $\cos(\mathrm{thrust}_K, \mathrm{thrust}_{\mathrm{ROE}})$ <0.95.
A fit to the ΔE distribution yields 265 ± 49 signal events, with the signal modeled by a Gaussian function and the background by a second-order Chebyshev polynomial. Using the sPlot technique, the $M(K^0_S K^0_S)$ distribution is examined. In simulation, a clear peak is seen near 1.5 GeV/c^2, corresponding to f'_2 and f_2 resonances, No such structure is observed in the data, indicating that these resonant contributions are overestimated in the simulation. Consequently, the simulation model is adjusted by removing the explicit $B^0 \to K^{*0}f'_2$ and $B^0 \to K^{*0}f_2$ components. The final configuration retains:

- $B^0 \to K^{*0}K^0_L K^0_L$ and $B^0 \to K^{*0}K^0_S K^0_S$ as phase-space decays,
- $B^0 \to K^{*0}K^0_S K^0_L$ with only non-f_2 resonances contribution,
- no explicit f'_2 or f_2 components.

This adjustment yields a more consistent and data-driven description of the $B^0 \to K^{*0}K^0\bar{K}^0$ background, ensuring that the simulation does not introduce spurious resonant features absent in data.

5.4.2.3 Modeling of $B^0 \to K^{*0}n\bar{n}$

The decay $B^0 \to K^{*0}n\bar{n}$ contributes about 0.3% of the total background in the signal region. Since this mode has not been directly measured, its branching fraction is estimated by applying isospin symmetry to the measured $\mathcal{B}(B^+ \to K^{*+}p\bar{p}) = 3.6 \times 10^{-6}$, corrected for the lifetime difference between B^+ and B^0 mesons. This gives $\mathcal{B}(B^0 \to K^{*0}n\bar{n}) \approx 3.35 \times 10^{-6}$.
In the standard Belle II simulation, these decays are generated according to a uniform three-body phase-space model. However, baryonic B decays are known to exhibit a strong enhancement near the baryon–antibaryon threshold, a feature not reproduced by the phase-space model [110]. To incorporate this effect, the external data were fitted with a threshold-plus-exponential function, and dedicated samples of 10^4 events were generated. These samples were reweighted from phase space to match the fitted distribution, and subsequently used to replace the generic $B \to Kn\bar{n}$ contribution in simulation.
Given the absence of direct measurements and the strong modeling assumptions, a conservative 100% systematic uncertainty is assigned to the branching fraction. This covers uncertainties from the threshold-enhancement modeling, the assumption of isospin symmetry, the simulation of (anti)neutrons in the calorimeter, and possible contributions from other baryonic modes.

5.4.2.4 Modeling of Charm-meson Backgrounds ($D \to K^{(*)}X$)

In both the nominal and sideband background samples, D^0 and D^+ decays account for 30-50% of the fake K^{*0} candidates in the signal region. The dominant modes are $D^0 \to K^-\pi^+$ and $D^+ \to K^-\pi^+\pi^+$, which directly mimic the K^{*0} topology. Since these modes are expected to play a central role in the background modeling, a dedicated validation is performed.
D candidates are reconstructed by combining the kaon used in the K^{*0} candidate with additional pions from the rest of the event, as described in appendix. 9.5.2 for $D^0 \to K^-\pi^+$ and in Sec. 5.1.5.5 for $D^+ \to K^-\pi^+\pi^+$.
To probe the modeling, maximum-likelihood fits are applied to the invariant-mass spectra of $M(D^0)$ and

$M(D^+)$ in both data and simulation. The fits are performed in the sideband region adjacent to the signal, $0.93 < \eta(\mathrm{BDT}_2) < 0.95$, using the full dataset and simulated samples. Gaussian functions describe the signal peaks, while linear functions model the combinatorial background, with all parameters floated independently for data and simulation.
For $D^0 \to K^-\pi^+$, the extracted signal yields are 608 ± 38 in data and 516 ± 17 in simulation, corresponding to a data-to-simulation ratio of 1.18 ± 0.08. For $D^+ \to K^-\pi^+\pi^+$, the yields are 732 ± 67 in data and 810 ± 30 in simulation, with a ratio of 0.90 ± 0.09. These results correspond to deviations of 2.3 and 1.1 standard deviations, respectively.
Overall, the comparison indicates that the simulation provides a statistically consistent description of the charm-meson contributions to the $B\bar{B}$ background within current uncertainties, thereby validating their use in the background modeling.

5.4.2.5 Treatment of Crossfeed Signals

The $K^{*0}\nu\bar{\nu}$ decay is analyzed together with the other three: $B^0 \to K^{(*)}\nu\bar{\nu}$ modes: $B^+ \to K^+\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$, and $B^+ \to K^{*+}\nu\bar{\nu}$. Because of their similar topologies, non-negligible crossfeed between channels is expected. The crossfeed is evaluated using simulated signal samples with all relevant corrections applied. Events are processed through the nominal BDT_1 and BDT_2 selections and compared to the corresponding target signal mode (Figure 5.35). In the $B^0 \to K^{*0}\nu\bar{\nu}$ channel, crossfeed from the other modes amounts to about 8.6% of the total signal yield. A separate study assesses crossfeed from background events, where a single event can be reconstructed in multiple signal channels, potentially leading to double counting. For example, a kaon from $B^+ \to K^{*+}(\to K^+\pi^0)\nu\bar{\nu}$ may combine with a pion to form a K^{*0} candidate and satisfy the $K^{*0}\nu\bar{\nu}$ selection. Using all simulated background samples within the signal regions, such overlapping events contribute approximately 2% of the background.
Background overlaps are negligible at the present level of precision. Crossfeed from other signal channels, though non-negligible, is explicitly included in the signal model, ensuring an unbiased measurement and a correct interpretation of the $B^0 \to K^{*0}\nu\bar{\nu}$ yield.

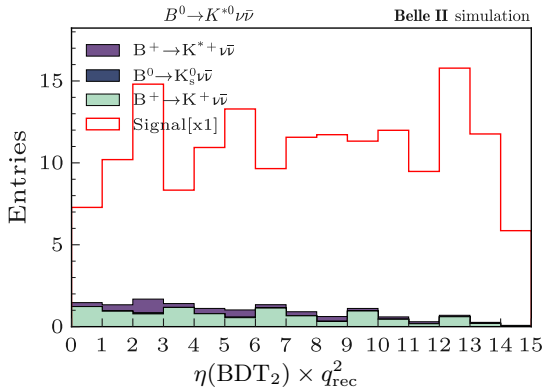

Figure 5.35: Crossfeed contributions in the $B^0 \to K^{*0}\nu\bar{\nu}$ statistical model. The distributions are shown in the signal region search bins, $\eta(\mathrm{BDT}_2) \times q^2_{\mathrm{rec}}$, with stacked filled histograms representing the three other signal modes ($B^+ \to K^+\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$, and $B^+ \to K^{*+}\nu\bar{\nu}$) reconstructed as $B^0 \to K^{*0}\nu\bar{\nu}$. The crossfeed fraction amounts to $\sim$ 8.6% of the total expected signal yield in this channel.

5.4.2.6 Modeling of Inclusive $B \to X_s\nu\bar{\nu}$ Background

In the Belle II simulation, $B \to X_s\nu\bar{\nu}$ denotes decays in which the s-quark hadronic system X_s is produced via $b \to s\nu\bar{\nu}$ transitions. The hadronization of X_s is modeled with `Pythia` fragmentation. In this context, $B \to X_s\nu\bar{\nu}$ includes all final states except the four exclusive modes $B \to K/K^*(892)\nu\bar{\nu}$. Consequently, the

X_s system includes higher kaon resonances such as $K_1(1270)$, $K_1(1400)$, $K(1430)$, and others with invariant mass of X_s larger than 1.1 GeV/c^2.
We find that the overall contribution of $B \to X_{s(u/d)}\nu\bar{\nu}$ is less than 1% compared with the total background in the signal region, which is relatively small. Nevertheless, there is concern that if the signal branching fraction is enhanced, the $B \to X_s\nu\bar{\nu}$ background could also be enhanced, since both processes originate from the same $b \to s\nu\bar{\nu}$ transition. In particular, fragmentation of X_s can produce final states such as $K^{*0}\eta/\omega/\pi$, or multi-hadron states, which can mimic the signal either when a neutral particle is not reconstructed ($\eta/\omega/\pi$) or when a track from a K^{*0} decay is missed, leading to a fake K^{*0} candidate when combined with another pion.
A detailed study of the composition in the signal region shows that the dominant background is associated with $K^*\pi$ final states, mainly from $B^+ \to K^{*0}\pi^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\pi^0\nu\bar{\nu}$, where a pion is missed in the reconstruction. Other final states contribute either through combinatorial background from two B-mesons ($\Upsilon(4S)$ fakes) or through combinatorial background within a single B-meson (X_s fakes). In particular, we investigated which particles mimic the signal K^{*0}, as shown in Fig. 5.36. The dominant contribution originates from genuine K^{*0} mesons produced in $X_s \to K^{*0}\pi$ decays. The second-largest contribution comes from fake K^{*0} candidates, where the dominant source is again $X_s \to K^{*0}\pi$. In this case, one kaon from the true K^{*0} combines with an additional pion from the X_s system to mimic the signal candidate. Overall, $X_s \to K^{*0}\pi$ decays constitute the primary source of $B \to X_s\nu\bar{\nu}$ background in the signal region.

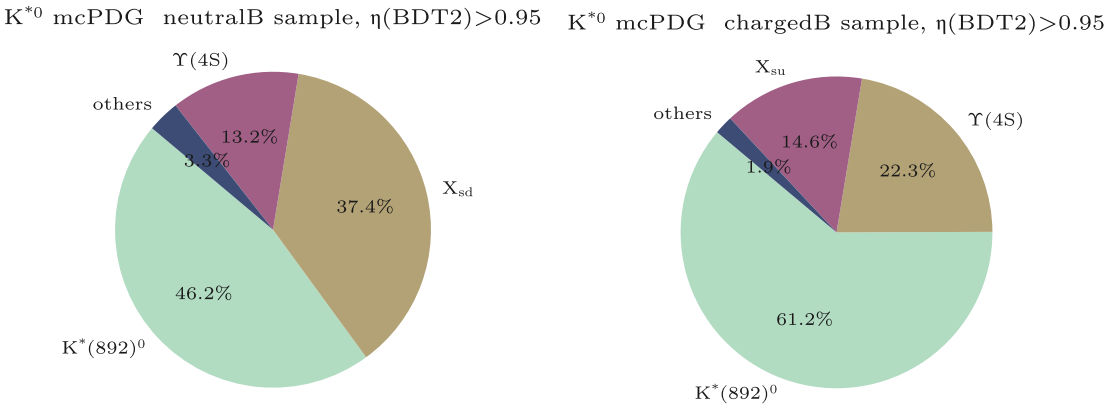

Figure 5.36: Composition of reconstructed K^{*0} candidates in the $B \to X_s\nu\bar{\nu}$ generic simulation sample, shown in the signal region ($\eta(\mathrm{BDT}_2) > 0.95$). Left: B^0 decays. Right: B^+ decays. The charts indicate whether the K^{*0} candidates originate from genuine K^{*0} mesons or from combinatorial fakes within the X_s system.

Currently, SM predictions provide branching fractions for both the inclusive $B \to X_s\nu\bar{\nu}$ decays and the exclusive modes $B \to K^*\nu\bar{\nu}$, no detailed form-factor calculations or event generators are available for higher-mass X_s states. By subtracting the exclusive contributions from the inclusive prediction, one can estimate the branching fraction for $B \to X_s\nu\bar{\nu}$ with $M(X_s) > 1$ GeV/c^2. However, no detailed theoretical model or form factor calculation exists for these higher-mass X_s states, which introduces significant modeling uncertainty. Experimentally, only upper limits on $B \to X_s\nu\bar{\nu}$ have been established. For instance, current measurements set [111]

$$\mathcal{B}(B \to X_s\nu\bar{\nu}) = [5.1^{+9.2}_{-8.8}(\mathrm{stat})^{+12.9}_{-11.0}(\mathrm{syst})] \times 10^{-6}, \tag{5.6}$$

with uncertainties dominated by both statistical and systematic contributions. This highlights that the experimental constraints are still rather loose, particularly for higher-mass X_s states. The $B \to X_s\nu\bar{\nu}$ decays are expected to exhibit a complex structure dominated by several K^* resonances, which can interfere with one another. For the current analysis, low-multiplicity decays are most relevant. Guidance for modeling

these decays can be drawn from measurements of analogous channels, such as $B^+ \to K^+\pi^-\pi^+\mu^+\mu^-$ observed by LHCb [112]. At low X_s invariant masses, the decays are dominated by the $K_1(1270)$ and $K_1(1400)$ resonances, while contributions from higher-mass states are relatively small, as shown in the partial wave analysis of $K\pi\pi$ system in $B \to J/\psi K\pi\pi$ [113].
For the present analysis, we adopt a simplified benchmark model in which equal mixtures of $B \to K_1(1270)\nu\bar{\nu}$ and $B \to K_1(1400)\nu\bar{\nu}$ decays are generated according to phase space. The total branching fraction is taken to be $(1.6 \pm 0.5) \times 10^{-5}$ for both B^+ and B^0 decays. Since the K_1 resonances have different branching fractions to $K^*\pi$ and $K\rho$ final states, they contribute differently to the $B \to K\nu\bar{\nu}$ and $B \to K^*\nu\bar{\nu}$ backgrounds. To account for this modeling uncertainty, the relative contributions of $K_1(1270)$ and $K_1(1400)$ are varied between 0.25/0.75 and 0.75/0.25. In this case, the interferences between Kaon states are ignored for simplicity.

5.5 Signal Extraction Setup

Signal yields are extracted for $B^0 \to K^{*0}\nu\bar{\nu}$ using a binned maximum-likelihood fit in the two-dimensional space defined by $\eta(\mathrm{BDT}_2)$ and q^2_{rec}. The data are divided into five bins in $\eta(\mathrm{BDT}_2)$ and three bins in q^2_{rec}, giving a total of 15 signal-region search bins. Both $\Upsilon(4S)$ and off-resonance datasets (the latter used to constrain the continuum background) are included, yielding 2×15 bins for the simultaneous fit. For the $\Upsilon(4S)$ simulation, four times the equivalent of 365 fb^{-1} of data are generated, which are then reweighted to match the integrated luminosity of the on-resonance data, 365 fb^{-1}. For the off-resonance background, an independent continuum simulation corresponding to 4×42.3 fb^{-1} is employed and reweighted to the off-resonance data luminosity of 42.3 fb^{-1}. This setup mirrors the methodology used for the $B^+ \to K^+\nu\bar{\nu}$ analysis, with adjustments in the binning scheme to accommodate the kinematic and classifier distributions relevant for the K^{*0} final state. The use of both on-resonance and off-resonance simulations ensures a reliable modeling of signal and continuum backgrounds, while the simultaneous fit across all bins allows for optimal extraction of the signal yields.

A common set of nuisance parameters $\mu_{q\bar{q}}, q = u, d, s, c$ and $\mu_{\tau^+\tau^-}$ controls the continuum normalization across all bins, with off-resonance data constraining the continuum in the on-resonance signal region. Templates model the observable distributions for each background class, and the likelihood is the product of Poisson probabilities over all signal bins. The fit covers 5×3 bins in ($\eta(\mathrm{BDT}_2)$, q^2_{rec}) and is performed simultaneously for on-resonance and off-resonance samples.

The signal strength μ is defined as the branching fraction relative to the Standard Model expectation, $\mathcal{B}_{\mathrm{SM}}(B^0 \to K^{*0}\nu\bar{\nu}) = 9.05 \times 10^{-6}$ [38]. The measurement is performed using a binned maximum likelihood fit, as discussed in Sec. 3.3. Systematic uncertainties and finite simulation statistics are included as nuisance parameters, implemented as multiplicative or additive modifiers of the expected bin yields and constrained via Gaussian priors.

The primary analysis is performed using the `pyhf` framework [101], with `sghf` used for cross-checks and validation. The expected signal and background contributions in the $B^0 \to K^{*0}\nu\bar{\nu}$ signal region is evaluated using simulated samples weighted to correspond to the on-resonance integrated luminosity of 365 fb^{-1}. The expected background and signal in the signal region search bins, $\eta(\mathrm{BDT}_2) \times \mathrm{q}^2_{\mathrm{rec}}$ are shown in Fig. 5.37 (upper). The region of $0.95 < \eta(\mathrm{BDT}_2) < 0.96$ is predominantly populated by background events. Overall, the background contribution decreases with increasing $\eta(\mathrm{BDT}_2)$ (Fig. 5.37 (bottom left)), while within each $\eta(\mathrm{BDT}_2)$ bin, it exhibits a modest increase from low to high q^2_{rec} values. In contrast, the signal distribution is relatively uniform across $\eta(\mathrm{BDT}_2)$ and increases along the q^2_{rec} bins, as shown in Fig. 5.37 (bottom right). The signal achieves its highest significance in the most sensitive region ($\eta(\mathrm{BDT}_2) > 0.99$), which is primarily dominated by $B\bar{B}$ events.

This binning scheme and the observed distribution patterns ensure that the most sensitive regions are those in which the signal contribution is maximized relative to the background, optimizing the extraction of the $B^0 \to K^{*0}\nu\bar{\nu}$ yield.

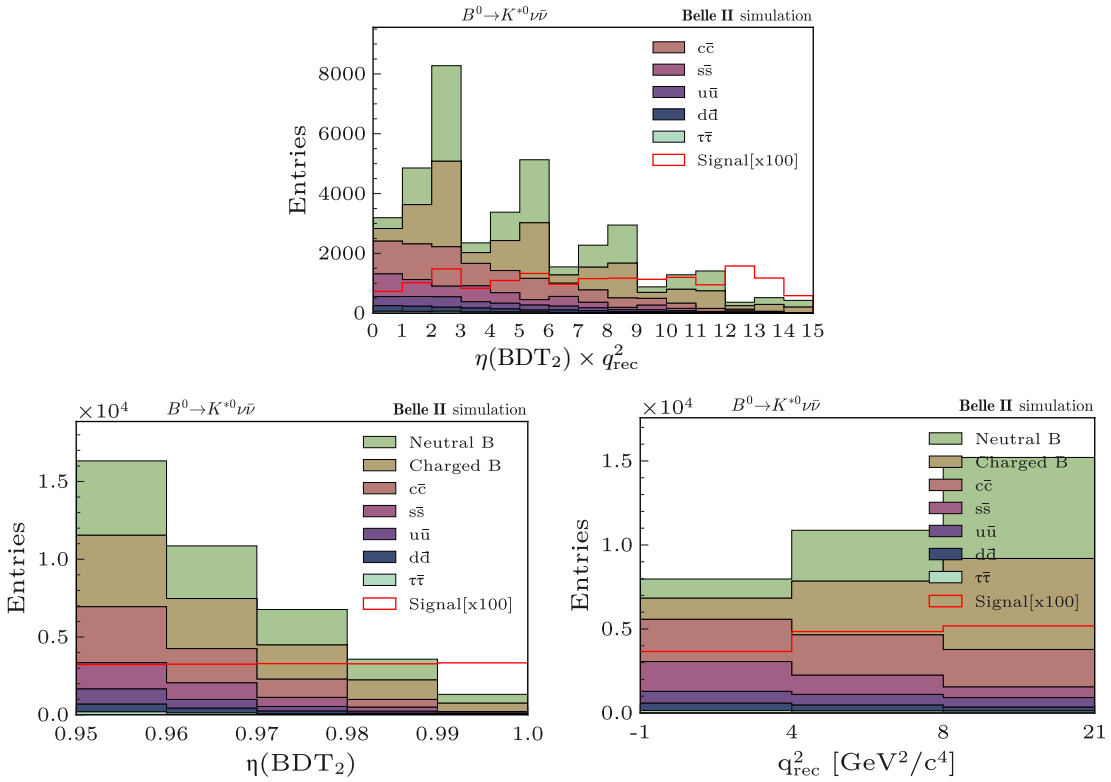

Figure 5.37: Expected background and signal distributions in the $B^0 \to K^{*0}\nu\bar{\nu}$ signal region. Top: two-dimensional binning in $\eta(\mathrm{BDT_2}) \times \mathrm{q}^2_{\mathrm{rec}}$. Bottom left/right: projections onto $\eta(\mathrm{BDT_2})$ and q^2_{rec}.

5.6 Systematics

In the case of the $B^0 \to K^{*0}\nu\bar{\nu}$ analysis, the incorporation of systematic uncertainties adheres to the same methodology as outlined for the $B^+ \to K^+\nu\bar{\nu}$ analysis in Sec. 4.6. The systematic uncertainties are grouped into three categories(physics modeling, detector response, and external inputs with global normalization factors) as discussed for the $B^+ \to K^+\nu\bar{\nu}$ channel in Sec. 4.6. For the $B^0 \to K^{*0}\nu\bar{\nu}$ channel, the same classification is employed, with minor updates to improve accuracy and to incorporate new or revised modeling inputs.

5.6.1 External Input Uncertainties

The following discussion covers uncertainties from luminosity, the number of $B\bar{B}$ pairs, finite signal simulation statistics, and external branching fractions.

- Integrated luminosity. The on- and off-resonance integrated luminosities are taken from updated luminosity measurements [96], with an uncertainty of approximately 1%. This directly scales the expected number of events and impacts both signal and background yields proportionally.

- Number of $B\bar{B}$ pairs. Determined by the Belle II B-counting group [103], with an updated uncertainty of 2%. This affects the calculation of the expected signal events.

- Number of B mesons: f^{+-} and f^{00}. To obtain the number of B^+ and B^0 mesons in the sample, two nuisance parameters are introduced: one for the ratio f^{+-}/f^{00}, constrained to 1.052 ± 0.031 [114], and another for the fraction of non-$B\bar{B}$ decays of the $\Upsilon(4S)$, $f_{\not B} = 0.0027^{+0.0138}_{-0.0002}$ [114], where $f^{+-} + f^{00} + f_{\not B} = 1$.

- Simulated sample size. The finite size of the simulated samples introduces bin-by-bin statistical uncertainties on the expected yields. Each bin is treated as a Poisson-distributed observation, with uncertainties implemented as nuisance parameters.

- Normalization of continuum background of $\Upsilon(4S)$ events. The reweighting of continuum simulation to off-resonance data is assigned a conservative 50% uncertainty per component, to account for potential mis-modeling and to remain consistent when combining the four signal channels.

- Normalization of off-resonance background. A 5% uncertainty is applied, the same as in the $B^+ \to K^+\nu\bar{\nu}$ analysis (Sec. 4.6.1), to account for potential differences in continuum composition between on- and off-resonance conditions.

- Normalization of $B\bar{B}$ background. A conservative 50% uncertainty is applied to cover normalization variations in the signal region, even though the observed normalization in the K^{*0} signal region is roughly 10%. This choice ensures consistency when combining all signal channels.

- Branching fractions of leading B^+ and B^0 in the signal region. The leading B^+ and B^0 decay backgrounds in the signal region are corrected using weights that adjust the branching fractions from the Belle II simulation to the latest nominal values [4], with the associated uncertainties propagated as a dedicated systematic, as discussed in Sec. 4.6. The decays are grouped into four categories, each treated differently: semileptonic decays involving D mesons use direct nominal value [4]; two-body hadronic decays are corrected similarly; hadronic decays involving D mesons use the updated version of `basf2` branching fractions validated by full event interpretation studies [85]; and charmless hadronic decays employ modified non-resonant branching fractions. Additional treatments include assigning 100% relative uncertainty to Pythia-generated decays and a 5% extra uncertainty for $B \to J/\psi K$ decays to account for possible $J/\psi \to n\bar{n}$ contamination. The semileptonic decays dominate the background, while the largest fractional uncertainties occur for the less common hadronic and charmless decays.

- Branching fraction for $B \to D^{**}$. Backgrounds from B decays to excited D mesons (D^{**}, D_{sJ}) enter the signal region when a kaon from the D decay is reconstructed, while the rest of the event is not. The fraction of signal candidates from these decays is about 3% for B^+ and 5% for B^0. The roughly 80 individual modes are grouped into four categories: (1) semileptonic ($B \to D^{**}\ell\nu$), (2) hadronic ($B \to D^{**}h$), (3) decays with $D_s^{(*)}$, and (4) other D_{sJ} decays. Each group has a dedicated nuisance parameter: known branching fractions are corrected in simulation and assigned measured uncertainties, while unknown `Pythia`-generated modes are assigned 100% uncertainty. This simplified treatment allows easy future updates. The largest impact comes from semileptonic decays (0.15%), while the other groups contribute below 10^{-3} relative uncertainty on the signal branching fraction.

The treatment of finite simulation sample size and the propagation of branching-fraction uncertainties follow the same strategy as outlined for the K^+ channel in Sec. 4.6, and the corresponding systematic variations are implemented identically in the K^{*0} analysis. Tab. 5.8 summarizes the sources and size of uncertainty of the above external-input uncertainties.

Source	Size of uncertainty	Impact on σ_μ
Integrated luminosity	1%	$\mathcal{O}(10^{-3})$
Number of $B\bar{B}$	2%	0.04
Simulated-sample size	$\mathcal{O}(1)$%	0.59
Normalization of continuum background	50%	0.53
Off-resonance sample normalization	5%	0.11
Normalization of $B\bar{B}$ background	50%	0.60
Branching fraction for $B \to D^{**}$	see text	0.31
Leading B-decays branching fractions	see text	0.20

Table 5.8: Summary of external-input systematic uncertainties for the $B^0 \to K^{*0}\nu\bar{\nu}$ channel. The second column lists the size of the applied uncertainty, while the third column shows the resulting impact on the signal strength σ_μ.

5.6.2 Detector Response Uncertainties

For the K^{*0} channel, the same sources of detector-response systematic uncertainty are considered as in the K^+ channel, since most corrections are based on reconstruction- or sample-independent studies. The only difference is in the particle identification, where both a kaon and a pion must be reconstructed. The applied corrections are treated as correlated across channels, while the associated uncertainties are conservatively treated as uncorrelated. The considered sources are:

- **Track finding efficiency.** As in the K^+ channel, we assign a systematic uncertainty of 0.9% per track, following the Belle II tracking group's recommendation. Signal and background samples are reconstructed both normally and with tracks randomly removed according to this probability, and the resulting differences in signal-region yields are propagated as systematic uncertainties.

- **Kaon and pion PID.** The K^{*0} candidate is reconstructed from two charged tracks, requiring kaonID$>$0.75 for the kaon and pionID$>$0.05 for the pion. Simulated events are weighted with corrections derived from Belle II PID performance studies in bins of (p_T, $\cos\theta$), and their associated uncertainties are propagated. For kaons, the uncertainties are at the level of $\sim$1% for efficiencies and $\sim$10% for fake rates. Pion-ID uncertainties are smaller but propagated in the same way. In the combined fit across the four channels, the PID corrections are correlated while the uncertainties are uncorrelated.

- **Photon energy response.** A systematic uncertainty of 0.5% is applied to cover possible mis-modeling of the ECL photon detection efficiency, following Belle II neutral reconstruction recommendations. This is implemented by randomly modifying 0.5% of photon clusters in the simulation.

- **Hadronic energy response.** Control studies of $B^+ \to K^+ J/\psi \to \mu^+\mu^-$ suggest a potential 5% upward bias in neutral-hadron energy scale, while the nominal simulation applies a 10% downward correction. To conservatively cover this discrepancy, a 10% uncertainty (for instance, 100% of the nominal correction) is assigned.
- K_L^0 **efficiency in ECL.** The K_L^0 reconstruction efficiency in simulation is corrected by 17% based on dedicated studies, with a systematic uncertainty of half this correction, i.e. 8.5%. This uncertainty is propagated bin-by-bin into the likelihood.

The evaluation and propagation of tracking and PID efficiency uncertainties follow the methodology of Ref. [105], with PID described by seven nuisance parameters and all other detector-response sources represented by a single nuisance parameter each. Tab. 5.9 summarizes the sources and size of uncertainty of the above detector-response uncertainties.

Source	Size of uncertainty	Impact on σ_μ
Track finding efficiency	0.9%	$\mathcal{O}(10^{-3})$
Signal kaon and pion PID	$\mathcal{O}(1\%)$	0.07
Photon energy	0.5%	0.08
Hadronic energy	10%	0.06
K_L^0 efficiency in ECL	8.5%	0.34

Table 5.9: Summary of detector-response systematic uncertainties for the $B^0 \to K^{*0}\nu\bar{\nu}$. The second column lists the assigned size of the uncertainty, and the third column shows the resulting impact on the signal strength σ_μ.

5.6.3 Physics Modeling Uncertainties

For the K^{*0} channel, the treatment of physics modeling uncertainties follows the same strategy as in the K^+ channel, with additional sources relevant to vector final states. Since all four signal channels are combined in the final fit, the K^+-related systematics are shared, while K^{*0}-specific effects are added separately. The main sources of uncertainty are summarized below:

- SM-Form-factors of signal. For the K^0 channel, q^2-dependent weights derived from updated longitudinal and transverse form-factor parameterizations are applied to simulated events. The induced variations in the q^2 spectrum are at the level of about 2% relative to the nominal form-factor model, and this residual is taken as the systematic uncertainty.
- Global signal efficiency. Embedded-sample studies show very good agreement between data and simulation; therefore, no correction is applied to the signal efficiency. The statistical uncertainty on the embedded data-to-simulation yield ratio is taken as a global normalization uncertainty on the signal.
- Branching fractions of $B \to K^{(*)}K^0K^0$. The dominant contribution comes from $B^0 \to K^{*0}K^0K^0$, with a 15% uncertainty assigned for the $K^{*0}K_L^0K_L^0$ mode based on the precision of our $B^0 \to K^{*0}K_S^0K_S^0$ measurement. For $B^+ \to K^{*+}K^0K^0$, where no significant signal is observed, the branching fraction is estimated via isospin arguments and a conservative 30% uncertainty is assigned. Backgrounds from $B^0 \to K^0K^0K^0$ and related decays ($B^0 \to K_S^0K_S^0K_S^0$, $K_S^0K_S^0K_L^0$, $K_S^0K_L^0K_L^0$) are corrected in simulation to account for known mismodeling and an obsolete decay definition, with branching fractions updated using isospin relations and the latest nominal values [4].
- Branching fractions of $B \to K^{(*)}n\bar{n}$. The main effect arises from the possible $B^0 \to K^{*0}n\bar{n}$ contribution, which produces signal-like kinematics. Simulation is corrected using measured $B \to K^{(*)}p\bar{p}$ decays, and a 100% uncertainty is assigned to cover isospin breaking and modeling of the invariant mass of $M(n\bar{n})$ spectrum. Other related channels, such as $B^+ \to K^+n\bar{n}$ (already included in the K^+ analysis), are also assigned the same conservative treatment.

- Branching fraction of $D \to K_L^0 X$. The same correction and uncertainty as in the K^+ channel is applied: 30% correction with a $\pm$10% absolute uncertainty.
- Continuum background modeling (BDT_C). As in the K^+ channel, the full difference between distributions before and after BDT_C reweighting is taken as a 100% shape systematic.
- $B \to X_s \nu\bar{\nu}$ Signal-like contributions from $B \to X_s \nu\bar{\nu}$, where X_s denotes final states with strangeness excluding K and K^*, are modeled as a sum of $B \to K_1(1270)\nu\bar{\nu}$ and $B \to K_1(1400)\nu\bar{\nu}$. A 50% uncertainty is assigned on the total contribution and an additional 50% uncertainty on the relative fraction of the two components.
- Fake K^{*0} candidates. A significant fraction of reconstructed K^{*0} mesons in the signal region originates from wrong combinations of kaon and pion tracks. These are grouped into categories: fakes from different B mesons at the $\Upsilon(4S)$ (10%), from the same B meson (20%), and from charm decays (D^0, D^+) (30%). To avoid double-counting, these uncertainties are not applied to decays already covered by branching-fraction uncertainties.

Tab. 5.10 summarizes the sources and size of uncertainty for the K^{*0} physics modeling systematics.

Source	Size of uncertainty	Impact on σ_μ
$B \to X_s \nu\bar{\nu}$	50% 10% for normalization / composition	0.14 / 0.06
Fake K^{*0} candidates	10%/20%/30% for $\Upsilon(4S)/D^0/D^+$	0.14 / 0.08 / 0.01
Signal SM form-factors	$\mathcal{O}$(1%)	0.13
Global signal efficiency	6%	0.06
Branching fraction for $B \to K^{(*)} K_L^0 K_L^0$	15(30)% $B \to K^{*0(+)} K_L^0 K_L^0$	0.22
	20% $B^+ \to K^+ K_L^0 K_L^0$	0.01
	30% $B^+ \to K^+ K_L^0 K_L^0$	$\mathcal{O}(10^{-3})$
	20% $B^0 \to K^0 K^0 K^0$	$\mathcal{O}(10^{-3})$
Branching fraction for $B \to K^* n\bar{n}$	100%	0.08
Branching fraction for $D \to K_L X$	10%	0.02
Continuum background modeling, BDT_C	100% of corrections	0.41

Table 5.10: Summary of physics-modelling systematic uncertainties for the $B^0 \to K^{*0}\nu\bar{\nu}$ channel. The second column lists the assigned size of the uncertainty, and the third column shows the corresponding impact on the signal strength σ_μ

For the K^{*0} channel, the leading sources of uncertainty arise from the normalization of the $B\bar{B}$ background (0.60), the continuum normalization (0.53), and the limited size of the simulated samples (0.59). At a comparable though somewhat lower level, the uncertainties associated with the leading B-meson branching fractions (0.20) and with the branching fraction of $B \to K^* K_L^0 K_L^0$ (0.22) contribute, together with those from the K_L^0 reconstruction efficiency (0.34) and the $B \to D^{**}$ branching fractions (0.31). Additional non-negligible contributions are observed from the BDT_C reweighting (0.41) and from the $B \to X_s \nu\bar{\nu}$ component (0.14). Despite the large size of several assigned systematic variations, their overall impact on σ_μ remains limited, since their effects are largely orthogonal to the bins most sensitive to the signal and are further constrained by the fit. The relative ranking of systematics is similar across all four channels, ensuring a consistent combination.
For the K^{*0} channel, the pre-unblinding validation studies (fit stability, signal injection, and background normalization checks) are still ongoing. Accordingly, the signal region remains blinded, and no unblinded results are reported in this work.

Part III

Complementary Studies

In addition to the main analysis of $B \to K^{(*)}\nu\bar{\nu}$ decays, several complementary studies were carried out. These include the development of a CDC background filter to improve tracking performance, as well as exploratory modeling efforts of the two-photon process $\gamma\gamma \to K^0_S K^0_S$ and the $K\pi\pi$ system in $B \to K\pi\pi\nu\bar{\nu}$ decays. While the latter two studies were not ultimately used in the main measurement, they provided valuable cross-checks and methodological experience.

Chapter 6

CDC Background-Suppression Filter

In the context of the Belle II experiment, the suppression of background in the CDC is crucial for maintaining high-quality tracking performance under the challenging conditions of high-luminosity operation. Within the Belle II tracking performance group, I have developed and implemented a dedicated CDC background filter aimed at mitigating the impact of non-physical signals in the chamber. This filter has already been applied during cosmic data taking 2025, and it is scheduled to be used in the official physics data-taking period beginning in 2025.
To study the nature of different types of CDC hits, a visualization tool was prepared in the form of a heatmap of a representative event (Experiment 26, Run 1939, Event 26840) Fig. 6.1. This visualization distinguishes three categories of hits within the CDC:

- Hits attached to a reconstructed track (physical signal candidates).
- Hits not attached to a track but flagged with the ASIC background marker (attributed to electronic noise).
- Hits not attached to a track and without the ASIC background flag (interpreted as crosstalk effects).

These two types of background hits, ASIC noise and crosstalk, originate from fundamentally different mechanisms and present distinct challenges:

- ASIC Noise (Electronic Noise). ASIC noise is generated within the readout electronics (ASICs) due to thermal fluctuations, baseline drifts, or digital switching noise. It's random in both time and spatial distribution; uncorrelated with physical activity in the detector and often flagged by firmware.
- Crosstalk (Signal Coupling). Crosstalk is caused by capacitive or inductive coupling between adjacent wires or channels when a true charged particle generates an avalanche on one sense wire. It's correlated with genuine physical hits in both time and position; typically lower in amplitude than the true hit and not flagged by ASIC markers. Crosstalk is particularly challenging to filter, since it closely resembles real physics hits. If not suppressed, it can produce fake tracks, distort track fitting, and increase the combinatorial load of pattern recognition.

To address the crosstalk contamination, the exploitation of a machine learning–based classifier capable of distinguishing true track-related hits (signal) from crosstalk hits (background) is proposed. This approach enables a more data-driven and adaptive background filtering strategy that goes beyond static flagging.
The remainder of this discussion will be organized as follows:

- **Motivation:** Crosstalk suppression is essential for reliable CDC performance in Belle II. Machine learning classifiers can exploit correlations between observables for better background rejection than simple threshold-based filters.
- **Training and implementation:** Classifiers were trained on labeled samples derived from track attachment and background flags, using key hit-level variables (ADC, TOT, TDC, and superlayer). Integration

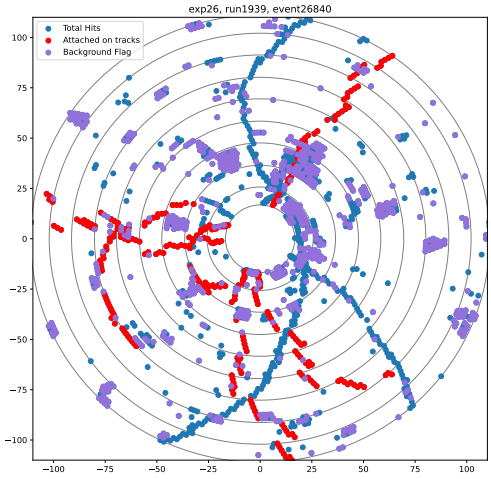

Figure 6.1: Event display of CDC hits for a representative event. Hits attached to reconstructed tracks are shown in red (without background flag). Hits flagged as ASIC background are shown in purple. Unattached hits without a background flag, interpreted as crosstalk, are also shown in blue. The visualization illustrates the coexistence of physical signals, electronic noise, and crosstalk within the same event.

into the reconstruction framework is achieved via the wire-hit preparation stage and database-managed weights.

- **Initial checks:** After implementation, the classifier was evaluated against the default database filter by comparing hit selection efficiency and GPU (runtime) performance with the same experiment as training samples.
- **Validation:** Evaluated the classifier on various datasets.
- **Performance:** Official checks after merging confirmed improved background suppression and preserved signal efficiency, with minimal computational overhead, demonstrating suitability for large-scale reconstruction.

For clarity, in the following, we adopt simplified notation: the detector hits associated with reconstructed tracks are referred to as hits-on-tracks, while hits not associated with any track are referred to as hits-off-tracks.

6.1 CDC Hit Characterization

The Belle II experiment is designed to operate at unprecedented luminosities, where the CDC must cope with high hit rates while preserving excellent tracking performance. Under such conditions, even modest levels of background can substantially degrade the quality of track reconstruction. Background hits that are uncorrelated with real charged particles add to the combinatorial complexity of pattern recognition, while correlated but spurious hits can bias track fits or produce fake tracks. To address this, one must look closely at the observables recorded for each CDC hit, namely:

- ADC (Analog-to-Digital Converter counts): Represents the integrated charge of the signal pulse, proportional to the ionization energy deposition.
- TOT (Time-over-Threshold): The time duration for which the signal remains above a discriminator threshold, correlated with signal amplitude.
- TDC (Time-to-Digital Converter counts): Encodes the drift time, i.e., the arrival time of the signal relative to the trigger, reflecting the spatial position of the ionization along the drift cell.

These observables exhibit distinct distributions for hits attached to reconstructed tracks (genuine signals) and for crosstalk hits. Their behavior is also constrained by the CDC geometry, which is organized into superlayers: The inner superlayers are made of small drift cells, located closest to the beam pipe. These provide high granularity to cope with dense hit environments and large background fluxes. And the outer superlayers use larger drift cells, where hit densities are lower, but precise momentum resolution is required. For quality assurance, database-level requirements are applied differently depending on superlayer type, denoted as DB in the thesis:

- Inner superlayers (small cells): ADC $>$ 15, TOT $>=$ 2, and ADC/TOT $>=$ 3.
- Outer superlayers (large cells): ADC $>$ 18, TOT $>=$ 2, and ADC/TOT $>=$ 3.

When examining the hit distributions under these requirements, several challenges become evident, as shown in Fig. 6.2 where the data is from Experiment 26: Crosstalk contributes strongly at very low ADC values but also shows a long tail extending to high ADC (value larger than 200), overlapping with genuine track hits. And it dominates the short TOT region (value smaller than 5), yet overlaps significantly with track-associated hits at moderate values, making a strict cut inefficient. In addition to overlaps, crosstalk shows strong contributions in both low TDC (value among 4200−4600) and very high TDC (value larger than 4900) regions Attempts to apply single-variable cuts (e.g. high ADC or restricted TDC ranges) are inefficient and unstable. Because these distributions vary across runs, fixed thresholds that work for one dataset often fail to generalize. Even extending to two-dimensional correlations such as ADC vs. TOT or ADC vs. TDC, the observed patterns are complex and vary across runs.
These considerations motivate the development of a classifier-based approach. A classifier can simultaneously combine information from ADC, TOT, and TDC. It can learn non-linear decision boundaries that adapt to overlapping distributions. It offers robustness against run-dependent variations that defeat static cuts. By training a classifier to distinguish track-associated hits (signal) from crosstalk (background) using these observables, we aim to achieve effective suppression of spurious hits while preserving efficiency for true tracks. This is crucial for ensuring the CDC delivers reliable tracking performance in the official Belle II data-taking period starting in 2025.

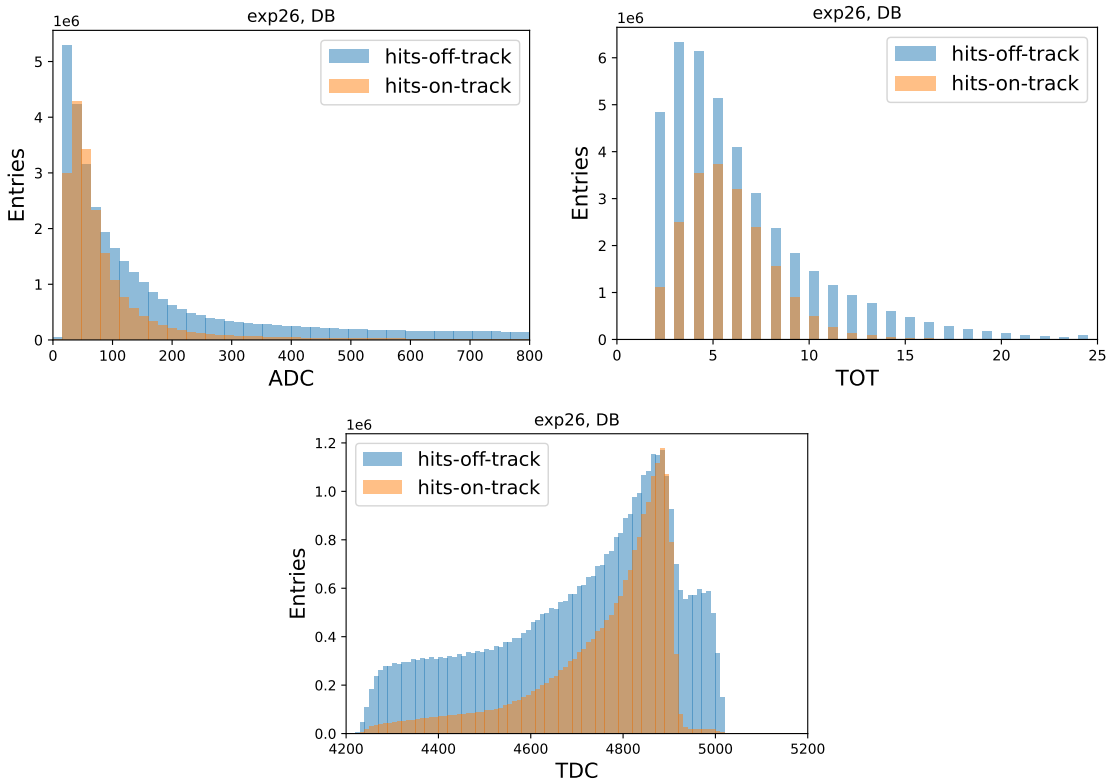

Figure 6.2: Distributions of CDC hit observables after database-level quality cuts using data from Experiment 26 with database filter. One-dimensional distributions of ADC, TOT, and TDC. Hits attached to reconstructed tracks (genuine signals) are shown in orange, while hits not attached to tracks (dominated by crosstalk) are shown in blue. Crosstalk contributes strongly at low values but overlaps significantly with genuine hits across the full observable range, highlighting the need for a multivariate classification approach.

6.2 Training and Implementing

The main objective is to introduce a MVA classifier to suppress CDC background hits, particularly those at the extremes of the ADC distribution, where single-variable cuts fail to provide efficient separation. Instead of relying solely on database-level requirements, the classifier exploits the full information from raw hit observables to provide a more flexible and data-driven filtering strategy.
The first step is to extract hit-level information directly from the raw data. Hits attached to reconstructed tracks are labeled as signal (positive class), where hits not attached to reconstructed tracks are labeled as background (negative class). ASIC-flagged noise is excluded, since it is already removed at the hardware level. To ensure broad coverage and allow the classifier to replace the strict database-level requirements, a minimal pre-selection (denoted as All) is applied: ADC $>= 1$, TOT $>= 1$, and ADC/TOT $>= 0$. This retains the full dynamic range of observables, ensuring the classifier is not biased by strict cuts.
The training features are ADC, TOT, TDC, and superlayer index (inner vs. outer). Two classifiers, FastBDT and XGBoost, were trained and compared. Both models achieved comparable discrimination power at this stage. Hyperparameters were optimized using `Optuna` [115], yielding the following best configuration, as shown in Tab. 6.1 Feature importance studies indicate that ADC provides the strongest discrimination ($\sim$46% contribution), followed by TOT ($\sim$25%), with the superlayer index providing the remaining discrimination power, as shown in Fig. 6.3 (upper left panel).

Parameter	Value
Number of trees	110
Tree depth	3
Shrinkage	0.53
Sampling rate	0.54
Number of equal-frequency bins	2^3

Table 6.1: Optimised hyperparameters of the classifier used in the CDC background filter.

Validation studies demonstrated that the models do not suffer from overfitting, with training and test performance in close agreement. A scan of the MVA output showed that the optimal operating point is around 0.3-0.5, as shown in Fig. 6.3 (upper right panel). If a working point is chosen to preserve more than 90% efficiency for hits attached to tracks, the classifier can reject up to 65% of unattached hits, as shown in Fig. 6.3 (bottom left). The efficiency curve versus MVA output shows that the 90−95% efficiency plateau corresponds to an MVA threshold of $\sim$0.4. Thus, the decision was made to adopt a cut at MVA score > 0.4 as the working point for deployment.
The integration of MVA techniques, such as FastBDT and XGBoost, into the `basf2` framework for CDC wire hit filtering represents an important step toward improving both the efficiency and robustness of track reconstruction in Belle II. By embedding the models directly into the reconstruction chain, it becomes possible to exploit correlations between detector observables that are inaccessible to conventional requirement-based filters, thereby providing enhanced background suppression while maintaining high signal efficiency.
A key aspect of the design is the use of the conditions database for managing the trained models. By storing weights and cut values as database payloads, the integration preserves the modularity and reproducibility of the reconstruction framework.
The MVA filter was integrated into the existing `CDCWireHitPreparer` module as an additional option. When the filter is set to `mva`, the module redirects wire hits to an `MVAFilter` that operates on the variable set `CDCWireHitVarSet`. For each wire hit, the variable extraction function retrieves key observables, including the ADC count, TOT, TDC count, and the superlayer identifier. These variables were chosen because they are both readily available in the hit information and directly relevant to distinguishing between signal and background hits. Consistency is ensured by using the same set of variables during both training and inference. The deployment of trained models is managed through the conditions database. The function `write_tracking_mva_filter_payloads_to_db` is used to store MVA weights along with associated cut values as database payloads. During reconstruction, the `TFCDC_WireHitPreparer` module retrieves the

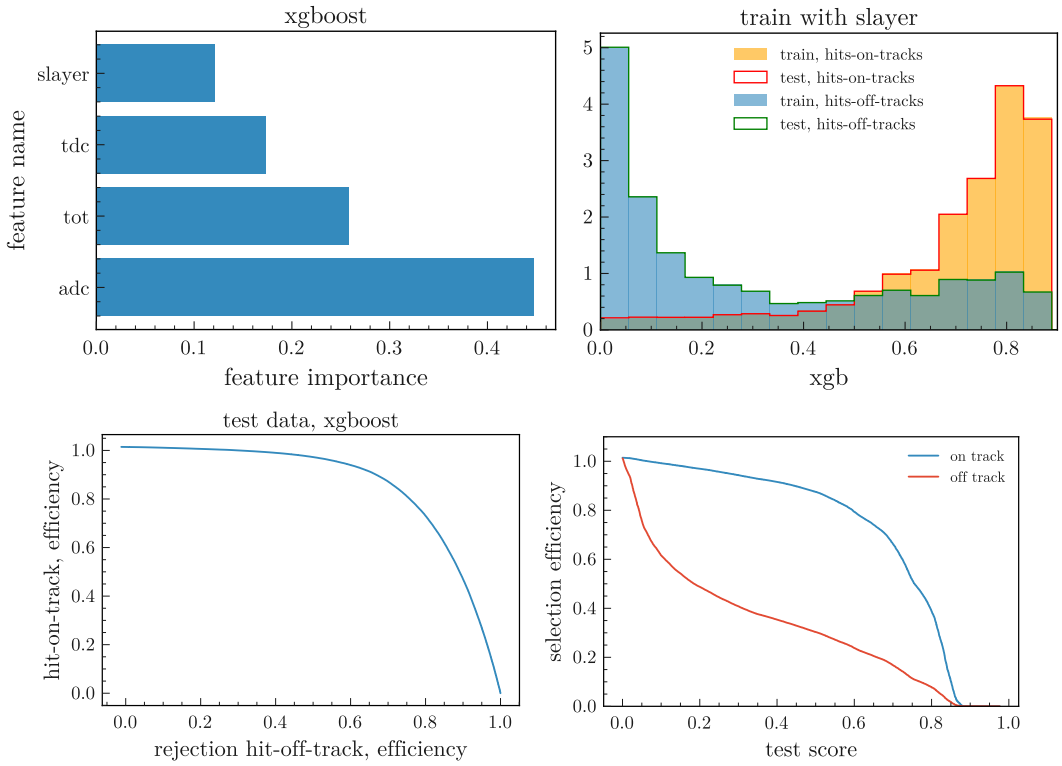

Figure 6.3: Performance of the CDC background classifier. Top left: feature importance of the input variables. Top right: comparison of training and test scores, showing no evidence of overfitting. Bottom left: distribution of efficiency of attached-on-track hits versus efficiency of unattached-on-track hits. Bottom right: selection efficiency versus MVA threshold for hits-on-track (signal) and hits-off-track (background), indicating a good separation and a stable working point around 0.4.

corresponding payload by its identifier, applies the MVA model to each hit, and rejects hits that fall below the specified cut. This design integrates seamlessly with the existing `basf2` conditions system, ensuring modularity, reproducibility, and flexibility. Moreover, it allows systematic performance studies across different MVA models or threshold settings simply by updating the database configuration, without modifying the reconstruction code itself.

6.3 Initial Checks with Experiment 26 Hadron Data

To evaluate the effect of the MVA-based filter, comparisons were made between the MVA and the DB requirement filter across different ADC ranges for hits both attached and not attached to a track (excluding hits flagged as ASIC background), as shown in Fig. 6.4. In the low ADC region (0-20), both signal hits (attached to track) and background hits (off track) exhibit similar trends under the MVA and DB filters, gradually increasing from 10 to 19 hits. In this range, the MVA retains slightly more signal while suppressing background marginally more effectively than the DB filter. In the dominant middle ADC region (20-400), the MVA preserves a comparable number of hits attached to reconstructed track as the DB filter while reducing approximately 20% of hits not attached to track, indicating enhanced background rejection without significant signal loss. In the high ADC region, dominated by crosstalk, both signal and background counts decrease under the MVA, with a substantial reduction of background hits compared to the DB filter, demonstrating its improved ability to suppress crosstalk.

A broader examination of the distributions of ADC, TOT, and TDC (Fig. 6.5) confirms that the MVA maintains nearly all signal hits while effectively removing extreme background contributions: hits not attached to track with very large TDC values (lager than 4950) are eliminated, and 30-40% of hits not attached to track across the TOT range are suppressed. Performance comparisons between XGBoost and FastBDT show virtually identical behavior, consistent with their use of the same hyperparameters. These results collectively indicate that the MVA-based filter can selectively reduce background while preserving signal, providing a promising improvement over the default database requirements.

In addition to wire-hit level performance, the impact of the MVA-based filter was assessed at the event levels. Two key variables were examined: number of CDC Hits (`nCDCHits`), defined as the number of CDC hits associated with a reconstructed track, and `nExtraCDCHitsPostCleaning`, which counts the CDC hits in the event not assigned to any track and not flagged as likely beam background in a event, for instance, hits surviving a post-cleanup selection. The variable `nExtraCDCHits` represents all CDC hits not associated with any track, including possible background contributions, while `nExtraCDCHitsPostCleaning` isolates hits that are not likely background or crosstalk.

Analysis of these variables shows that the MVA filter preserves nearly all signal hits, as shown in Fig. 6.6 (left panel), with a slight increase of approximately 40 hits in the lower range and a small decrease in tracks with more than 50 hits. For the post-cleaning variables, as shown in Fig. 6.6 (right panel), the distribution of hits shifted from the range (200-1000) in the filter to (50-600) under the MVA, indicating that while the number of hits contributing to a track remains similar, there is a substantial reduction in extraneous hits considered during track reconstruction. This demonstrates that the MVA effectively suppresses unattached-track hits, including crosstalk, without compromising the integrity of genuine track signals.

Overall, these results suggest that the MVA filter not only improves hit-level background suppression but also leads to a cleaner track-level reconstruction, potentially enhancing downstream event-level analyses by reducing ambiguities from spurious hits.

Track-finding efficiency was quantified as the fraction of reconstructed tracks satisfying nCDCHits$> i$ relative to all reconstructed tracks,

$$\varepsilon(\theta; i) = \frac{\text{tracks with nCDCHits} > i}{\text{all reconstructed tracks}}. \tag{6.1}$$

and was evaluated for thresholds $i \in [5, 25]$ as a function of the reconstructed polar angle θ, as shown in Fig. 6.7. Across this range of hit-count thresholds, the MVA-based filter and the DB requirement yield very similar efficiency curves, indicating that the MVA implementation is robust against variations in the hit-count requirement.

The observed patterns can be summarized as follows. Near very small θ, 17°, where tracks are nearly collinear with the forward beam direction, the efficiency is dominated by statistical fluctuations; here, the DB filter sometimes shows a marginally higher efficiency than FastBDT, while XGBoost tracks the DB more closely. For intermediate θ, the efficiency for both filters rises and reaches a plateau around $\theta \approx 0.85$, after which a local minimum appears near $\theta \approx$1.5-1.6 (roughly transverse to the beam, near 90°). Beyond this valley, the efficiency increases again, with the MVA showing a small but consistent advantage over the DB filter across a wide θ range, until both efficiencies decline toward the largest polar angles ($\sim$ 2.2-2.5), where behaviour becomes more variable.

These features admit plausible detector-level interpretations that are consistent with the geometry and operational context. The low-θ region is subject to low statistics and to enhanced beam−line effects (boosted kinematics and backgrounds close to the beam pipe), so small differences among algorithms there are expected and should be interpreted with caution. The valley near $\theta \sim$ 1.5-1.6 likely reflects geometric transitions in the tracking system (barrel/endcap boundaries and changes in wire geometry and stereo angles) where pattern recognition is intrinsically more difficult and where hit assignment ambiguities increase. The small advantage of MVA in much of the mid-to-large θ region suggests that systematic removal of spurious hits (crosstalk and other unattached-track contributions) improves the purity of the input hit sets used by the track finder, reducing confusion during pattern formation without substantially sacrificing genuine hits.

To quantify the global impact of hit removal on track-finding, the efficiency $\varepsilon(i)$ was evaluated as a function of the minimum hit requirement nCDCHits$> i$, with i ranging from 5 to 25, as shown in Fig. 6.8. The resulting values were displayed as a heatmap, which clearly illustrates the monotonic decrease in efficiency as the hit threshold increases. For the DB and both MVA classifiers, the efficiency decreases smoothly from about 95% at nCDCHits$>$5 to approximately 71-72% at nCDCHits> 5. The overall behavior of FastBDT and XGBoost closely follows that of the DB requirement, demonstrating that the MVA does not adversely bias the selection even under progressively stricter hit requirements.

To highlight relative differences, a normalized efficiency was defined as

$$\varepsilon_{\mathrm{rel}}(i) = \frac{\varepsilon(i)}{\varepsilon_{\mathrm{DB}}(i)} \tag{6.2}$$

such that the DB filter corresponds to unity at all thresholds. Both FastBDT and XGBoost remain essentially identical to the DB across the entire range, with relative efficiencies clustering around unity. At the most stringent hit requirement, nCDCHits$>$23, XGBoost even exceeds the DB by approximately 1%, indicating that in this regime, it may preserve a slightly higher fraction of genuine tracks without compromising background suppression. This stability across hit thresholds provides further evidence that the MVA-based filter is a reliable replacement for the DB requirement, with negligible risk of systematic efficiency loss.

In Belle II, the CDC standalone tracking reconstructs tracks from filtered CDC hits using parallel track-finding algorithms: the Legendre transformation for global pattern recognition and the cellular automaton (FacetAutomation) for local segment building. These two methods are merged into the final CDC standalone tracks, which are subsequently combined with SVD tracks in the global fit. While the Legendre and cellular automaton modules dominate the total runtime of CDC tracking, it is important to verify that the introduction of an MVA-based background filter in `TFCDC_WireHitPreparer` does not introduce significant additional time consumption.

When comparing classifiers, both FastBDT and XGBoost provide nearly identical track-finding efficiencies. However, without optimization, the XGBoost evaluation in `TFCDC_WireHitPreparer` was extremely slow, requiring approximately 4543.61 $\pm$ 1288.58 ms per event, compared to 3.77 $\pm$ 1.00 ms for the DB filter. This overhead arises from repeated Python wrapper calls for each hit within `MVAExpert`, which accumulate substantially even though `TFCDC_WireHitPreparer` is not the dominant module. To address this, the classifier interface was redesigned to support batch evaluation of hits. A dedicated prediction function accepts a vector of input features, and the ordering of training variables is obtained directly from the payload. The filter interface was restructured from object-wise to vectorized evaluation, with corresponding updates to `MVAFilter`. After this optimization, the per-event runtime of XGBoost in `TFCDC_WireHitPreparer` was reduced to roughly 10-15 ms, negligible relative to the total CDC standalone runtime dominated by Legendre and cellular automaton modules.

To quantify the impact of MVA-based background filtering on computational performance, the per-event runtimes of three key CDC standalone modules, `TFCDC_WireHitPreparer`, `TFCDC_SegmentFinderFacetAutomation`, and `TFCDC_AxialTrackFinderLegendre`, were extracted from the log files and averaged over several runs. In `TFCDC_WireHitPreparer`, the DB filter requires approximately 2-6 ms per event, whereas FastBDT consumes 5-11 ms per event and XGBoost consumes 9-16 ms per event. These differences are run-dependent, reflecting varying event complexity and detector occupancy. On average, FastBDT adds 5-7 ms and XGBoost adds 7-10 ms relative to the DB, indicating a modest overhead from MVA evaluation that remains negligible compared to the overall CDC tracking time, as shown in Fig. 6.9.

For the TFCDC_SegmentFinderFacetAutomation module, the DB runtime is run-dependent, ranging from 30-50 ms per event, while both FastBDT and XGBoost are nearly run-independent at approximately 24 ms per event. This results in a 10-20 ms per-event speedup relative to the DB, as shown in Fig. 6.9. Similarly, in TFCDC_AxialTrackFinderLegendre, the DB runtime varies between 25-85 ms per event depending on the run, whereas the MVA-based classifiers are nearly constant around 30 ms per event, providing 5-70 ms per-event improvement, as shown in Fig. 6.9.
Overall, these measurements show that implementing the MVA background filter in TFCDC_WireHitPreparer does not compromise computational efficiency. In fact, due to slight optimizations in subsequent CDC standalone algorithms, the three main modules combined become almost twice as fast compared to the run-dependent DB baseline, while preserving track-finding performance.
Here is the same event display, as Fig. 6.1, if we loose the selection (left) and apply the classifier Fig. 6.10.

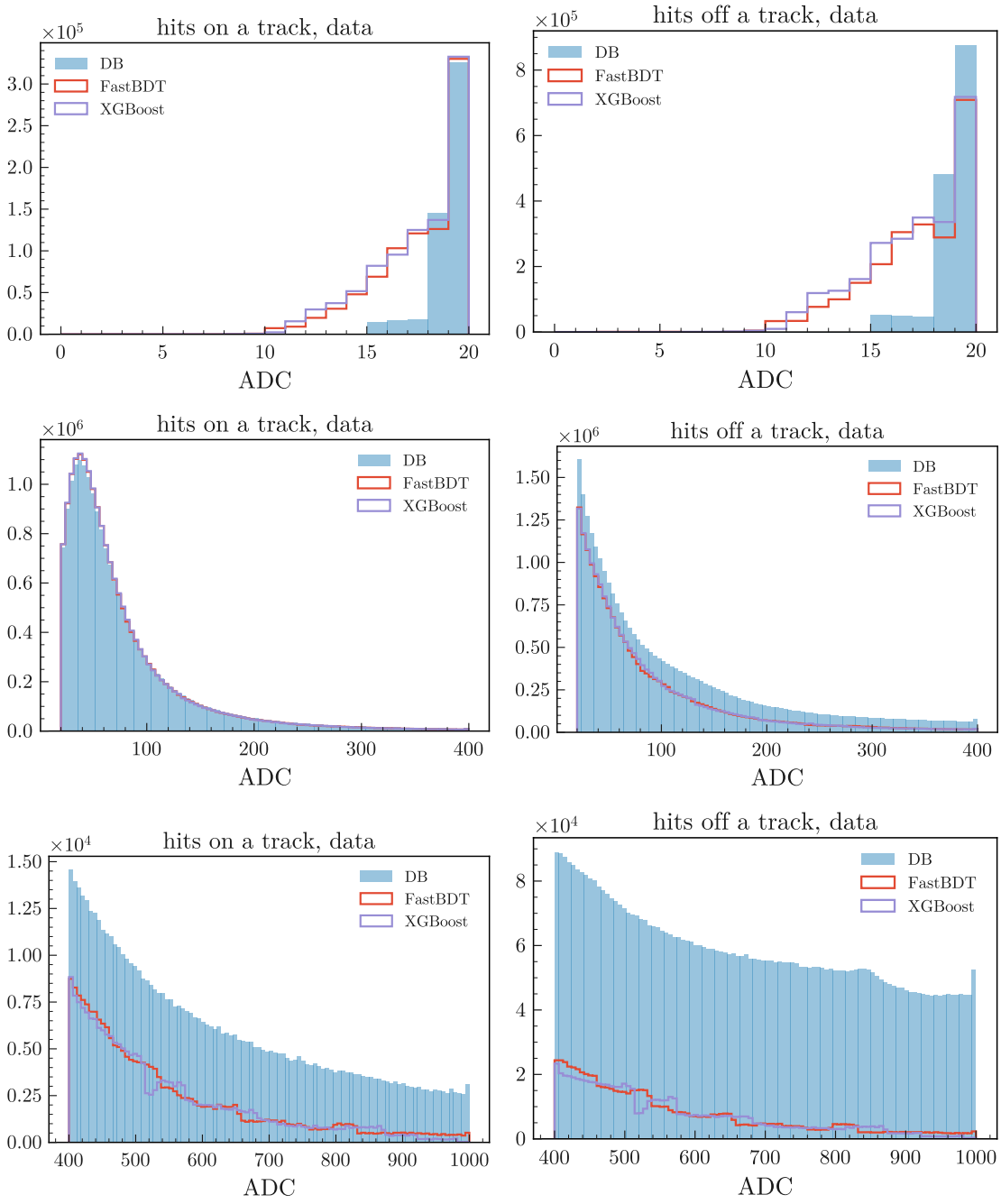

Figure 6.4: Comparison of the number of CDC hits selected by the database (DB) filter (blue), FastBDT (red), and XGBoost (purple) for experiment 26 hadron data. Left panels: hits attached to reconstructed tracks. Right panels: hits not attached to tracks (dominated by crosstalk). From upper to bottom: low-ADC region (ADC $<$ 20), mid-ADC region (20 $<$ ADC $<$ 400), and high-ADC region (ADC $>$ 400). The MVA filters preserve track-associated hits while suppressing hits not attached to track contributions more effectively than the DB filter, especially in the high-ADC region.

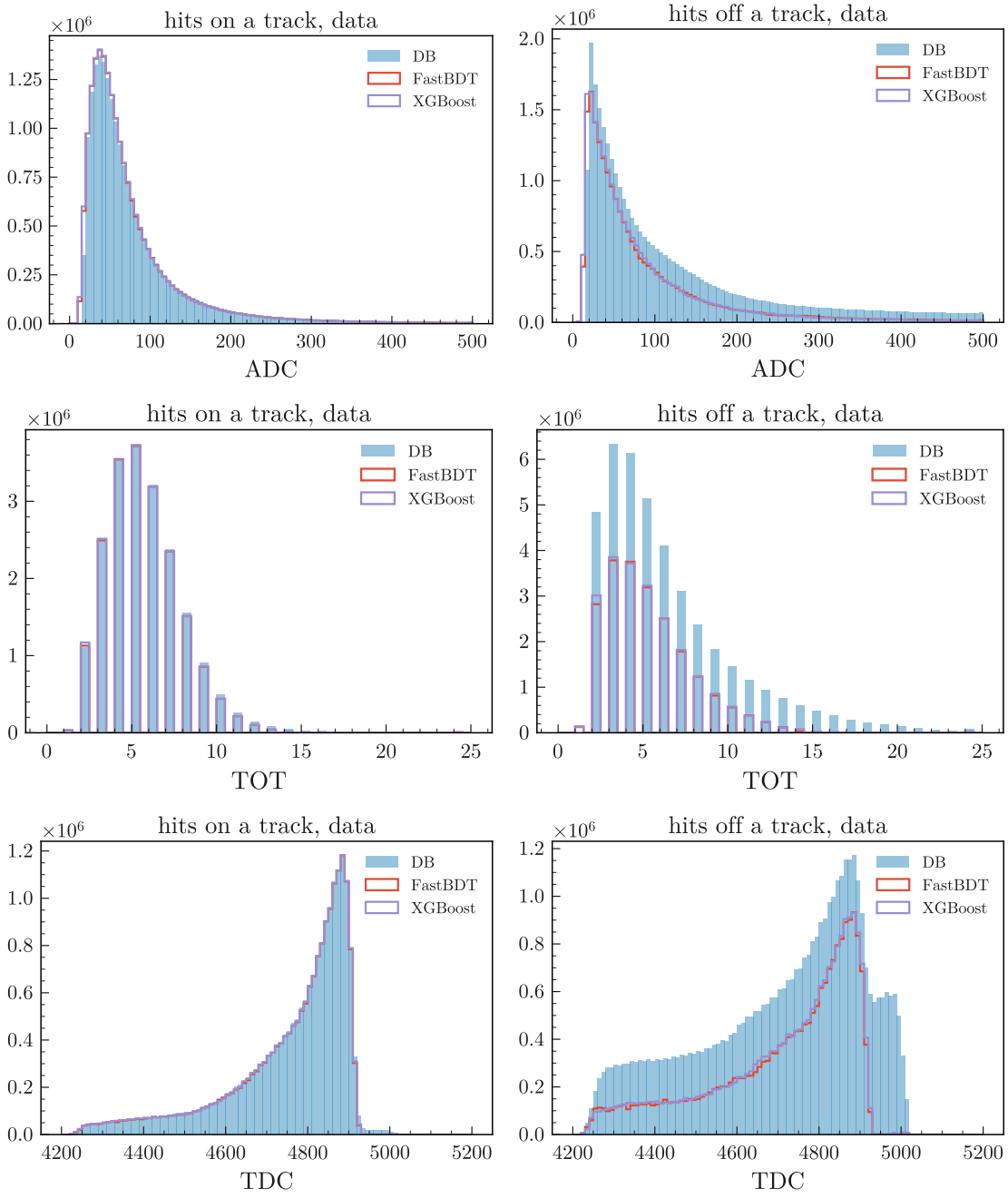

Figure 6.5: Distributions of CDC hit observables for experiment 26 hadron data after filtering. Left panels: ADC, TOT, and TDC distributions. Right panels: unattached-track hits in the same observables. Blue: DB filter, red: FastBDT, purple: XGBoost. The MVA filters preserve nearly all track-associated hits while significantly suppressing extreme unattached-to-track contributions, such as very large TDC values and 30–40% of hits across the TOT range.

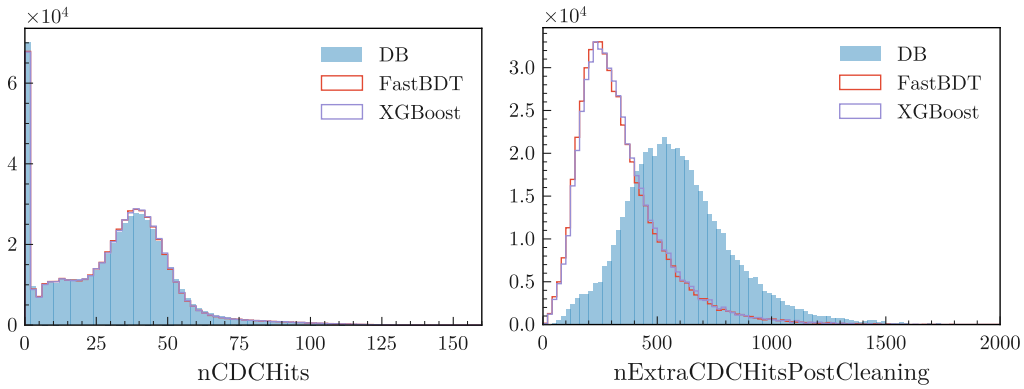

Figure 6.6: Event-level tracking variables comparing the DB filter (blue), FastBDT (red), and XGBoost (purple) for experiment 26 hadron data. Left: distribution of the number of CDC hits attached to reconstructed tracks (nCDCHits). Right: number of extra CDC hits after post-cleaning. The MVA filters preserve nearly all track-associated hits while substantially reducing extraneous hits, leading to cleaner inputs for track reconstruction.

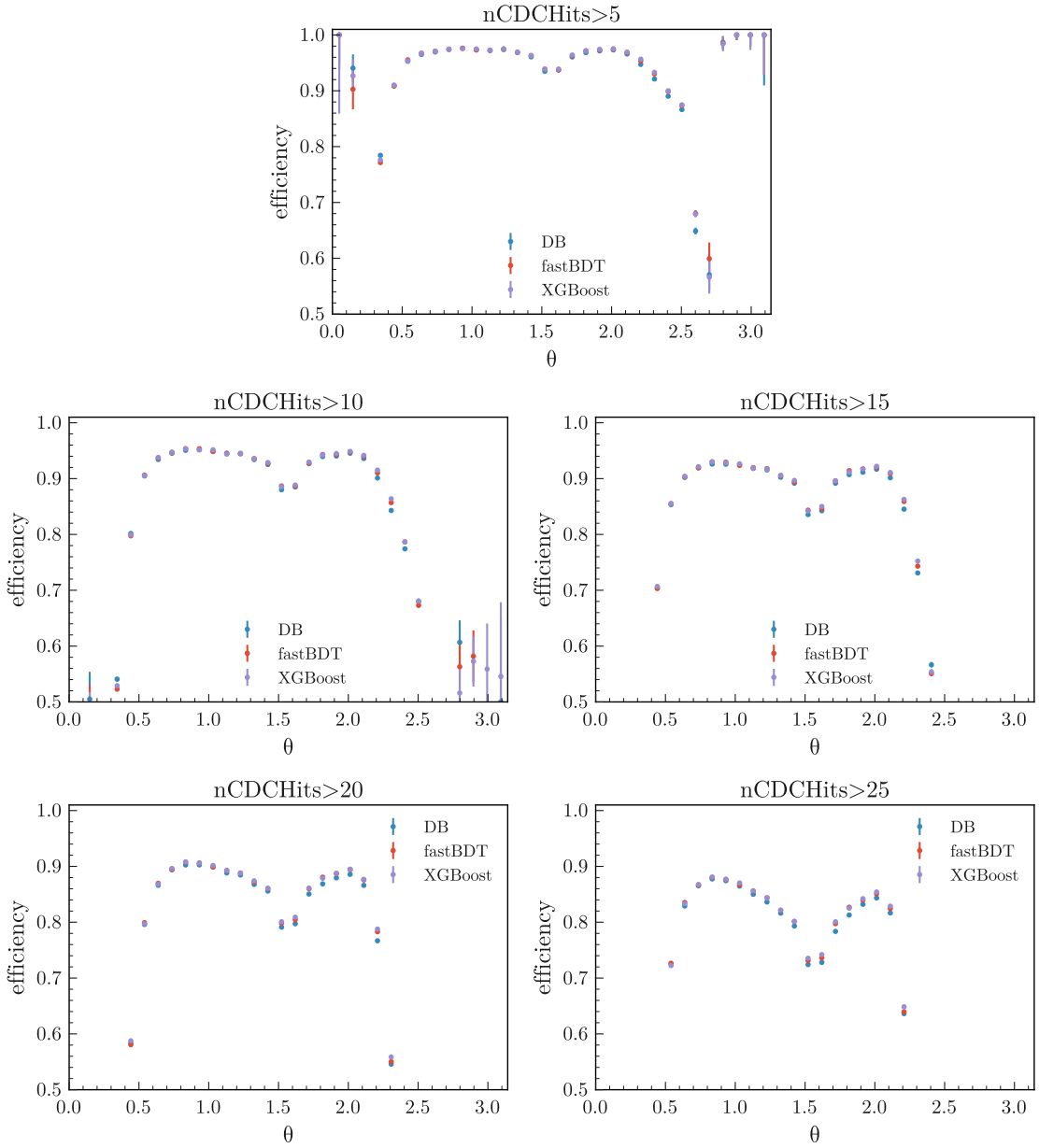

Figure 6.7: Track-finding efficiency $\varepsilon(\theta; i)$ as a function of the polar angle θ for minimum hit requirements nCDCHits >5, 10, 15 (top and middle rows) and 20, 25 (bottom row). Comparison of DB filter (blue), FastBDT (red), and XGBoost (purple). The MVA-based filters closely follow the DB baseline across all thresholds, with small gains in efficiency over a broad θ range, while preserving stability in challenging regions (e.g. forward angles and the barrel–endcap transition).

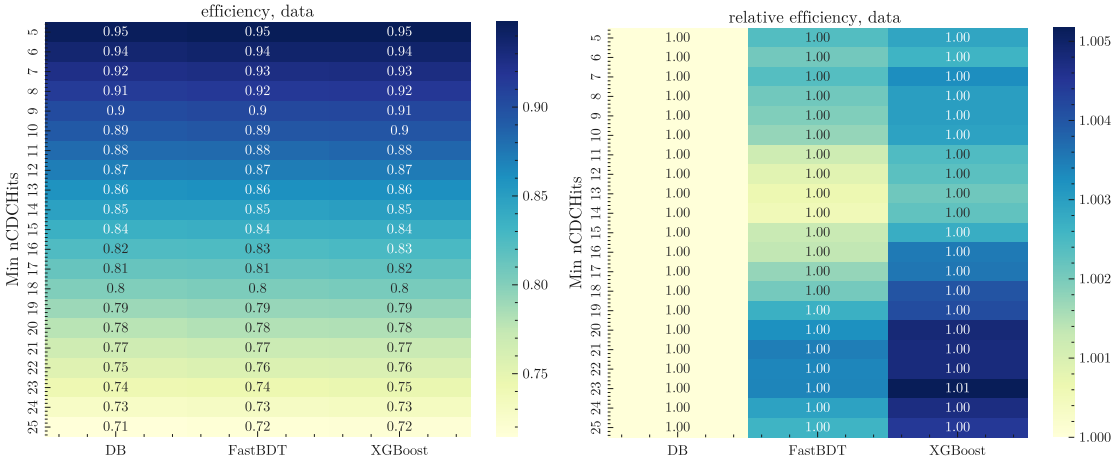

Figure 6.8: Global track-finding efficiency as a function of the minimum hit requirement (left) and the relative efficiency ε_{rel} normalized to the DB filter (right). For the DB filter, FastBDT and XGBoost.

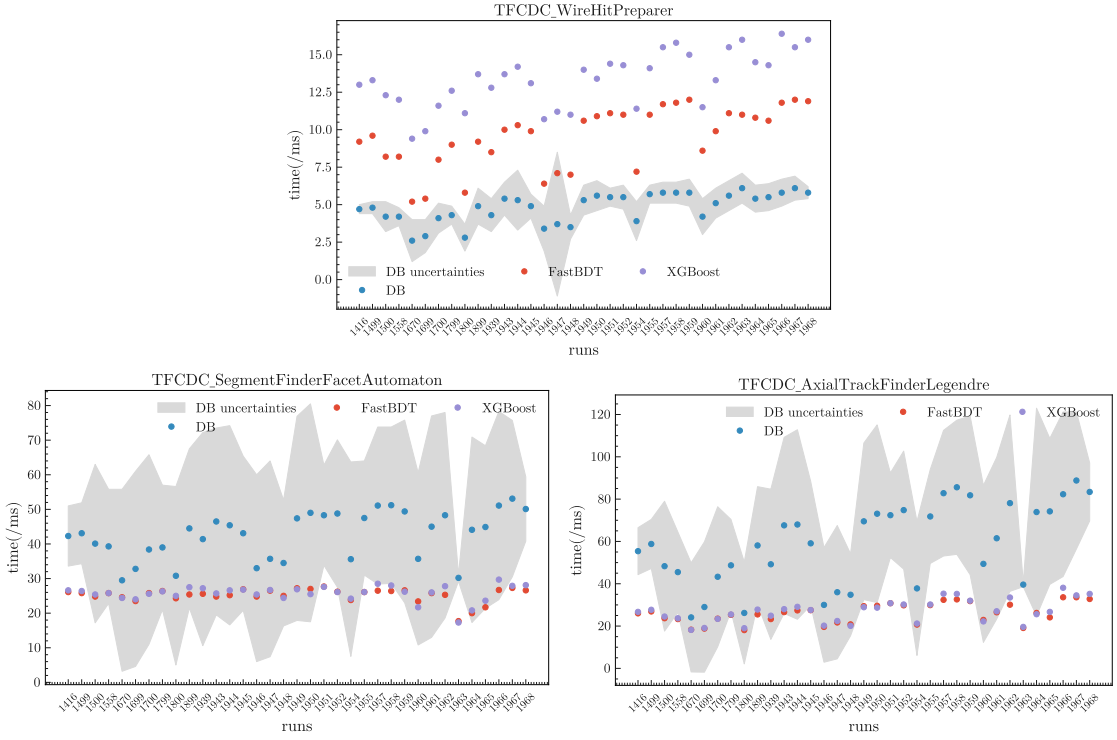

Figure 6.9: Average per-event runtime of key CDC standalone modules in experiment 26 hadron data, comparing DB filter (blue), FastBDT (red), and XGBoost (purple) for `TFCDC_WireHitPreparer`, `TFCDC_SegmentFinderFacetAutomation`, and `TFCDC_AxialTrackFinderLegendre`.

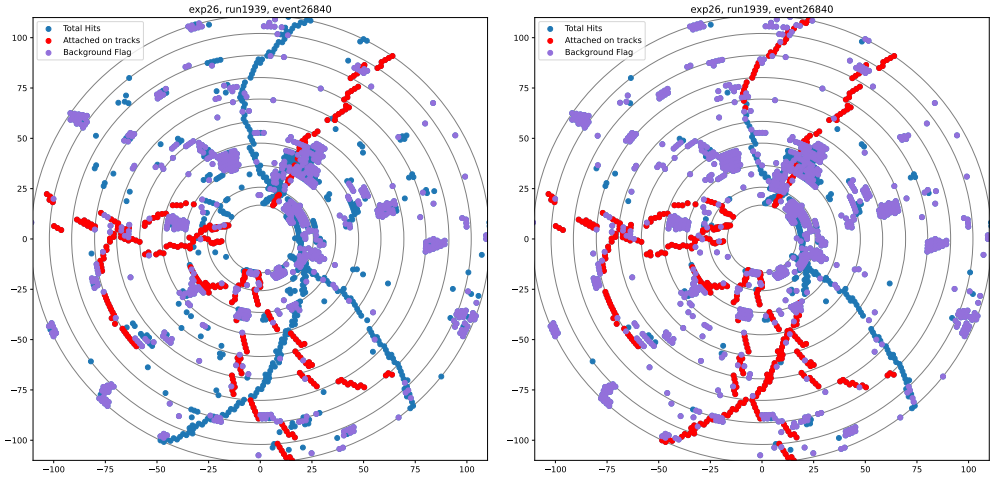

Figure 6.10: CDC event display for Experiment 26, Run 1939, Event 26840. Left: hits with loosened database-level requirements (All), including significant background contamination. Right: the same event after applying the FastBDT filter, showing suppression of unattached-track hits and preservation of track-associated signals.

6.4 Validations

While FastBDT and XGBoost demonstrate similar performance in both efficiency and runtime, FastBDT is simpler to integrate within the `basf2` framework and is therefore chosen as the preferred background filter. The above checks were carried out using data from the same experiment as the training sample (Experiment 26). To confirm that the classifier generalizes well and is not biased toward this particular dataset, further validation was conducted on a broad range of independent samples:

- **Experiment 30 data:** independent data to test robustness across different detector conditions.
- **Di-muon samples:** both data and simulation, providing a clean and well-understood control channel for validation.
- **Run-independent simulations:** MC16 samples from Experiments 0, 1003, and 1004, ensuring coverage of multiple run periods without experiment-specific biases.
- **Run-dependent simulations:** samples matched to specific experimental conditions, verifying stability of the classifier under realistic time-dependent variations.
- **Proton-dominated sample:** a dedicated dataset enriched in protons, which are characterized by higher ionization losses and reduced curvature in the CDC, providing a stringent test of classifier behavior for particle-species dependence.

Since the main classifier is trained on hadron tracks, an alternative MVA trained on muons is also prepared. Applying this muon-specific classifier to the same datasets provides an input-output consistency check for the framework, ensuring reliable performance across particle types and experimental conditions.

6.4.1 Validation with Experiment 30 Data

As a first cross-check, the FastBDT filter was validated on Experiment 30 data, which corresponds to a different run period and therefore provides an independent test of stability under varying detector conditions. The event-display heatmaps reveal that Experiment 30 data are comparatively cleaner, as shown in Fig. 6.11, exhibiting reduced levels of crosstalk background relative to Experiment 26. As a result, the background suppression task is less demanding, and the track-finding efficiency is correspondingly higher for both the DB filter and the FastBDT-based filter.
This is also reflected in the hit-count distributions: the raw number of CDC hits shows a slight shift toward lower values compared to Experiment 26, as shown in Fig. 6.12, whereas the distribution of post-cleaning hits, the number of CDC hits in the event not assigned to any track, and not flagged, is also shifted but to a lesser extent, compared it in Experiment 26, as shown in Fig. 6.12. The reduced impact of post-cleaning illustrates the intrinsically cleaner conditions of Experiment 30, in which fewer spurious hits are present to begin with.
Track-finding efficiency, defined in Eq. 6.1, was then evaluated for thresholds $i \in [5, 25]$, with results shown in Fig. 6.13. Across this range, the DB filter achieves efficiencies decreasing from approximately 97% to 79%, while the MVA-based filter yields efficiencies of 94% to 77%. The latter remains consistently within 0-2% of the DB values, demonstrating that the classifier maintains robust performance on an independent dataset.
Runtime studies further highlight differences between modules. For the `TFCDC_WireHitPreparer` stage, the DB filter exhibits stable, run-independent timing of $\sim$ 2.8 ms per event, whereas the MVA shows more variation, ranging from 5.0 to 6.2 ms per event. By contrast, in the `TFCDC_SegmentFinderFacetAutomation` and `TFCDC_AxialTrackFinderLegendre` modules, the MVA achieves more stable performance, with average run-times of 28 ms (compared to 30-38 ms for DB) and $\sim$20 ms (compared to 22-30 ms for DB), respectively. Taken together, these results correspond to an overall gain of approximately 6-12 ms per event when using the MVA filter, as shown in Fig. 6.14, while preserving efficiency curves comparable to the DB filter.
These results confirm that the FastBDT filter generalizes well to cleaner experimental conditions, preserving tracking efficiency while reducing runtime.

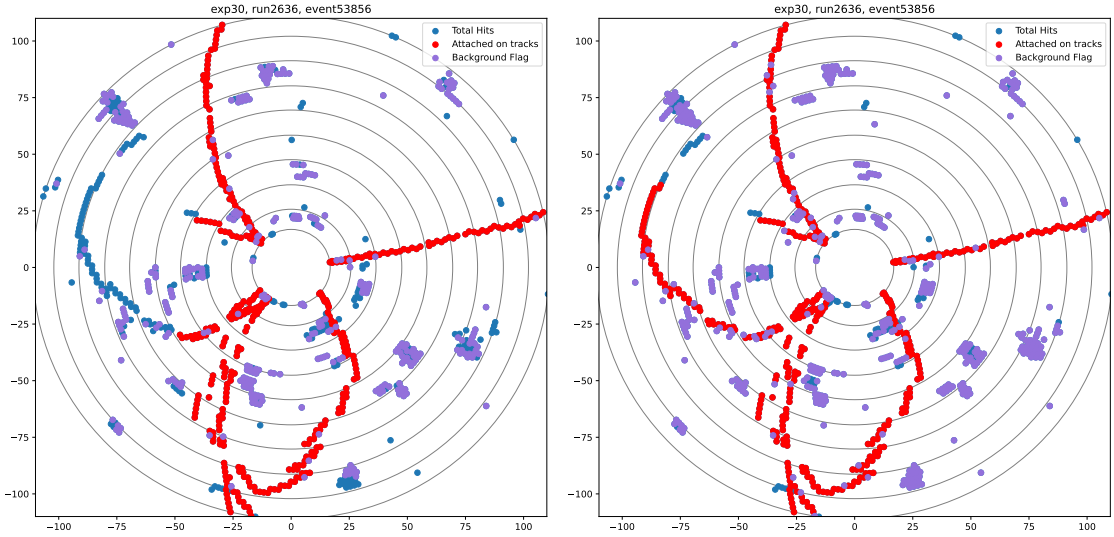

Figure 6.11: Event-display heatmaps for Experiment 30 data. Left: default DB filter; Right: FastBDT filter. Compared to Experiment 26, these data are intrinsically cleaner, with fewer crosstalk hits.

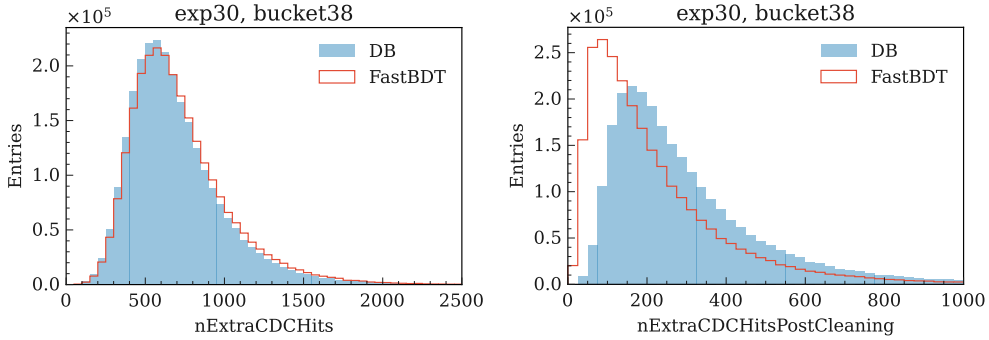

Figure 6.12: Distributions of extra CDC hits in Experiment 30. Left: total number of hits not assigned to tracks; Right: post-cleaning hits, with DB filter (blue) and nominal FastBDT classifier (red).

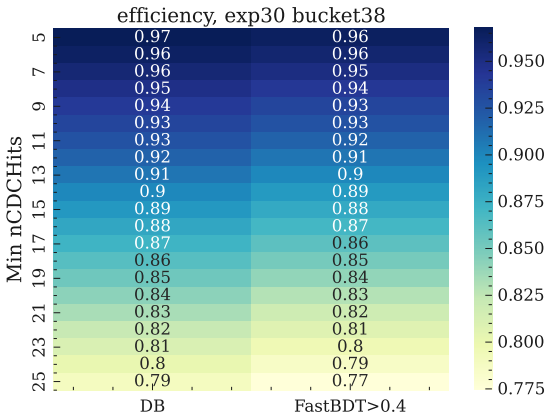

Figure 6.13: Tracking efficiency in Experiment 30 as a function of the minimum hit requirement i of DB filter and nominal FastBDT classifier. The two methods agree within 0–2% across the full range $i \in [5, 25]$.

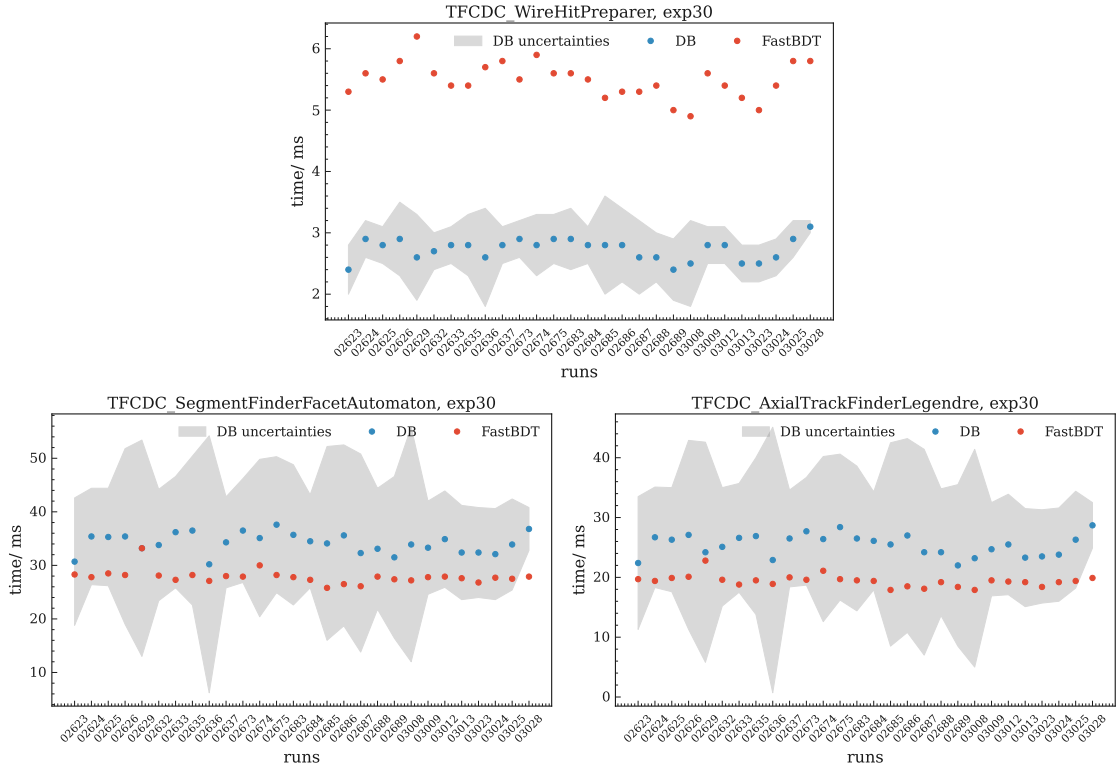

Figure 6.14: Average per-event runtime of key CDC standalone modules in Experiment 30 hadron data, comparing DB filter (blue) and FastBDT (red) for `TFCDC_WireHitPreparer`, `TFCDC_SegmentFinderFacetAutomation`, and `TFCDC_AxialTrackFinderLegendre`.

6.4.2 Validation with Simulation Samples

Since the characteristics of crosstalk background and other hit-level observables in the CDC are strongly event-dependent, and in some cases also run-dependent, it is essential to assess the robustness of the MVA classifier on simulated samples. This is particularly important because the classifier was trained directly on data, and its ability to generalize to different experimental conditions must be established.
Two categories of simulated samples provided by the data production group were employed:

- **Run-dependent MC15 sample:** 100,000 $B^0\bar{B}^0$ pairs matched to specific detector conditions of the Experiment 26 data.
- **Run-independent MC16 samples:** 200,000 $B^0\bar{B}^0$ pairs generated for each of three phases, Experiment 1003 (early Phase 1), Experiment 1004 (Run 2), and Experiment 0 (nominal Phase 3 conditions) yielding a total of 600,000 events.

The nominal FastBDT classifier was applied to these samples and compared against the default DB filter.

Run-dependent validation (MC15rd) At the hit level, FastBDT slightly redistributes the number of hits per track: tracks with 30-45 CDC hits become more frequent, while those with around 50 hits are somewhat reduced compared to DB. In the post-cleaning variable, FastBDT significantly suppresses excess hits, shifting the distribution from about 300-1000 hits (DB) to roughly 100-700 hits per event, as shown in Fig. 6.15.

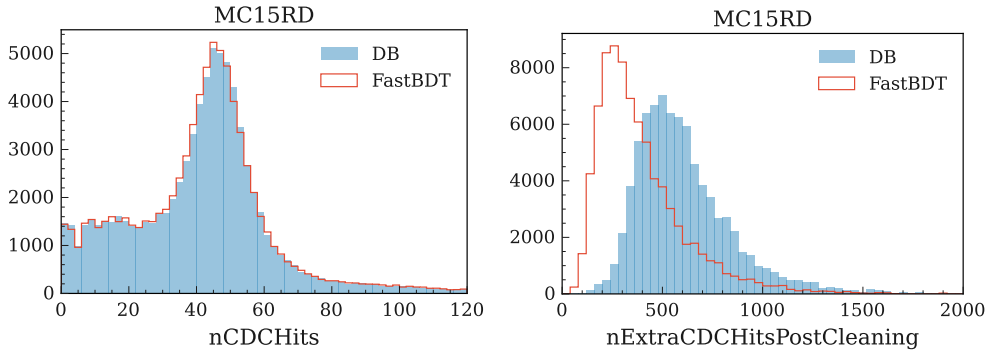

Figure 6.15: Distributions of extra CDC hits in MC15rd. Left: total number of hits not assigned to tracks; Right: post-cleaning hits, with DB filter (blue) and nominal FastBDT classifier (red).

At the track-finding level, the efficiency as a function of polar angle remains stable. A small overall gain is observed, except for a local dip around θ ∼1.5 rad (perpendicular to the beam axis), consistent with the transition region already seen in real data. At fixed working points, the global efficiency of FastBDT is essentially identical to DB, confirming robustness, as shown in Fig. 6.16.
For the computational performance, the hit-preparation stage roughly doubles in runtime under FastBDT (from 4.6 ms to 9 ms), reflecting the cost of MVA evaluation. However, this overhead is offset by faster downstream processing: the axial track finder and segment finder are accelerated by about a factor of two (for instance, from 70 ms to 35 ms and from 55 ms to 40 ms, respectively). Thus, the overall tracking chain benefits from a net speed-up, as shown in Fig. 6.17.

Run-independent validation (MC16ri) The run-independent MC16 samples correspond to different data-taking periods and background conditions, providing a complementary test of the FastBDT classifier beyond the run-dependent samples. These include Experiment 0, representing nominal phase 3 conditions with a high background; Experiment 1004, with a moderate background, and Experiment 1003, with a low background. The nominal classifier was applied to all samples and compared with the default database requirements.

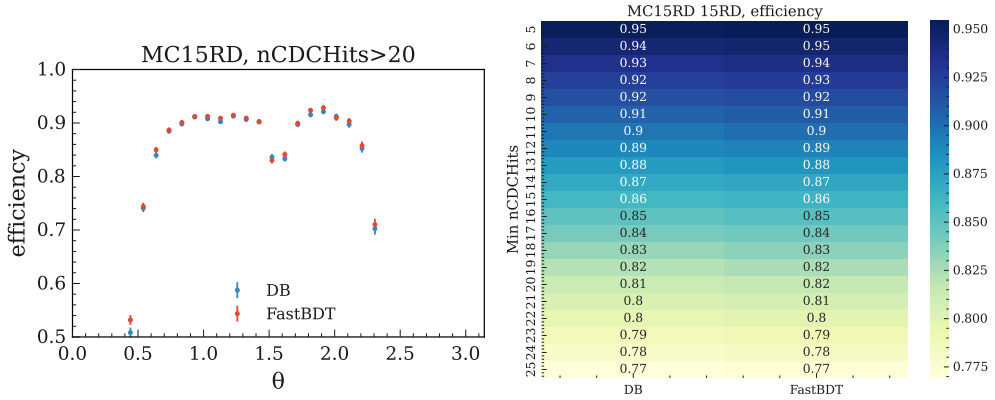

Figure 6.16: Left: Track-finding efficiency $\varepsilon(\theta; i)$ as a function of the polar angle θ for minimum hit requirements nCDCHits > 20. Right: Tracking efficiency in MC15rd as a function of the minimum hit attached on tracks requirement. Compared between the DB filter and the nominal FastBDT classifier.

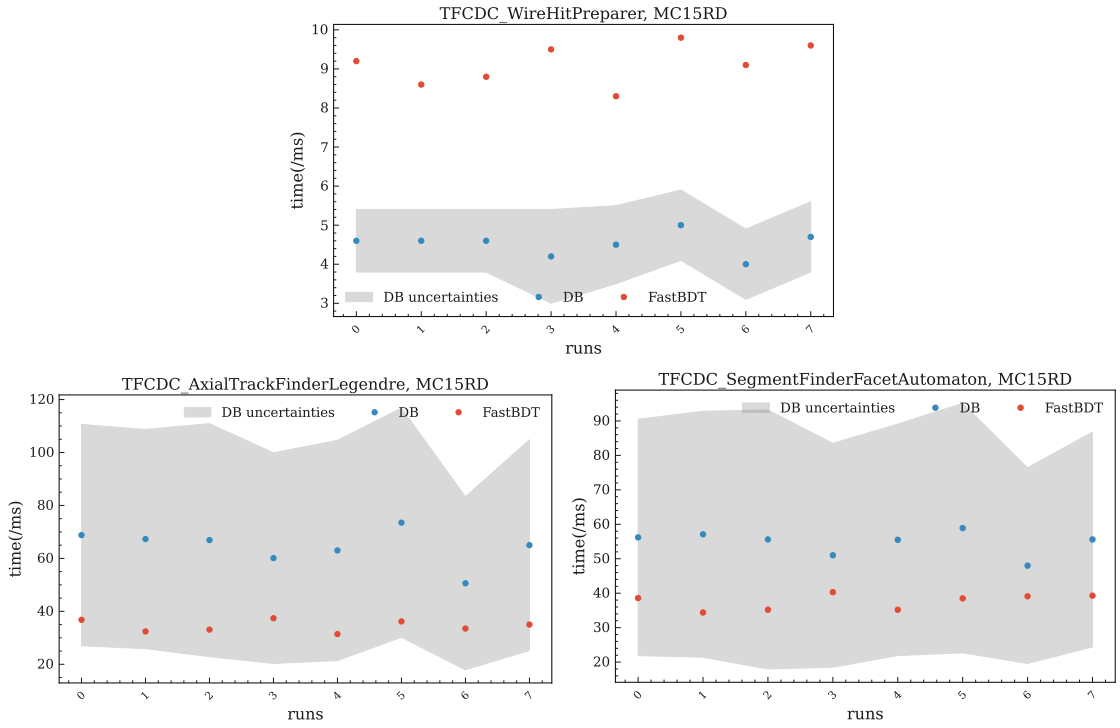

Figure 6.17: Average per-event runtime of key CDC standalone modules in MC15rd, comparing DB filter (blue) and FastBDT (red) for `TFCDC_WireHitPreparer`, `TFCDC_SegmentFinderFacetAutomaton`, and `TFCDC_AxialTrackFinderLegendre`.

At the hit level, FastBDT consistently reduces the number of post-cleaning hits across all experiments. In the high-background sample (Experiment 0), post-cleaning hits per event decrease from roughly 1600-2500 under the database filter to about 550-1300 with FastBDT, while hits per track show fewer events with very low counts (0-20) and more tracks containing 40-50 hits. In the moderate-background sample (Experiment 1004), post-cleaning hits shift from 600-1300 to 200-800, with a modest increase in tracks containing 40-50 hits and a slight decrease in tracks with 50-60 hits. In the low-background sample (Experiment 1003), post-cleaning hits decrease from 250-1000 to 100-750, while the distribution of hits per track remains largely unchanged, as shown in Fig. 6.18.
At the track level, the FastBDT filter generally preserves or slightly improves track-finding efficiency. In the high-background sample, efficiency rises across the full polar-angle range, including near perpendicular tracks, with working-point efficiencies increasing by approximately 1-3%, from 92-70% under the database filter to 93-73% with FastBDT. In the moderate-background sample, overall working-point efficiencies increase by around 1% (95-79%), while in the low-background sample, efficiency is largely unchanged (96-81%), with only a small improvement in the forward region of the detector, as shown in Fig. 6.19.
Regarding computational performance, the hit-preparation stage shows increased runtime due to the MVA evaluation: approximately 9 ms for Experiment 0, 4 ms for Experiment 1004, and 3.5 ms for Experiment 1003. Despite this overhead, downstream modules benefit from reduced processing time. The axial track finder runtime decreases from roughly 460 ms to 110 ms in Experiment 0, from 120 ms to 45 ms in Experiment 1004, and from 60 ms to 35 ms in Experiment 1003. Similarly, the segment finder runtime is reduced from 200 ms to 60 ms in Experiment 0, from 75 ms to 50 ms in Experiment 1004, and from 60 ms to 40 ms in Experiment 1003, as shown in Fig. 6.20, 6.21 and 6.22. These results indicate that the computational advantage of the classifier is greatest in high-background conditions, while still providing moderate gains in cleaner samples.
Validation on both run-dependent and run-independent simulations confirms that the FastBDT-based filter provides consistent hit-level and tracking-level performance compared to the DB filter. It achieves substantial background suppression, reproduces features observed in real data, and improves the overall runtime of the CDC standalone tracking chain. Importantly, the MC16ri studies demonstrate that the classifier is particularly powerful under high-background conditions, where it simultaneously enhances tracking efficiency and reduces runtime, while maintaining stable performance in low-background environments. Together, these simulation studies confirm that the FastBDT filter is robust across both matched and mismatched conditions, with stable efficiency and runtime improvements.

6.4.3 Validation with $e^+e^- \to \mu^+\mu^-(\gamma)$ Data

To verify the generality of the FastBDT classifier trained on hadron samples, its performance was evaluated on a lepton-dominated channel, specifically $e^+e^- \to \mu^+\mu^-(\gamma)$. This sample contains only two tracks per event, providing a clean and well-controlled test of the classifier in a low-background environment. In addition to applying the nominal classifier, a dedicated classifier was trained on the di-muon sample itself using the same hyperparameters and hit-level observables, serving as an input-output consistency check for the original classifier.
Since the reconstruction differs from the B meson sample, primary selections were applied to suppress beam-induced backgrounds and spurious tracks. These selections required hits to satisfy $|d_0| < 2.0$ cm, $|z_0| < 4.0$ cm, $p_T > 2.0$ GeV/c, and momentum in c.m. frame $p^* > 0.5$ GeV/c. Furthermore, the $e^+e^- \to e^+e^-$ (`Bhabha`) events were identified as contaminating sharp peaks in both the momentum and polar-angle distributions. These events were removed by applying the veto

$$\sim ((\omega < 0 \;\&\; \theta < 40^\circ \;\&\; \theta > 17^\circ) \mid (\omega > 0 \;\&\; \theta > 120^\circ \;\&\; \theta < 135^\circ)),$$

where ω denotes the track charge.
Examination of the training variables revealed slight differences in 2D correlations, particularly ADC versus TDC and ADC versus TOT, reflecting the cleaner nature of the dimuon sample. The dedicated di-muon classifier was trained with the same variables and hyperparameters as the hadron classifier, showing no signs of overfitting. TOT emerged as the most important feature ($\sim 45\%$), followed by ADC ($\sim 28\%$),

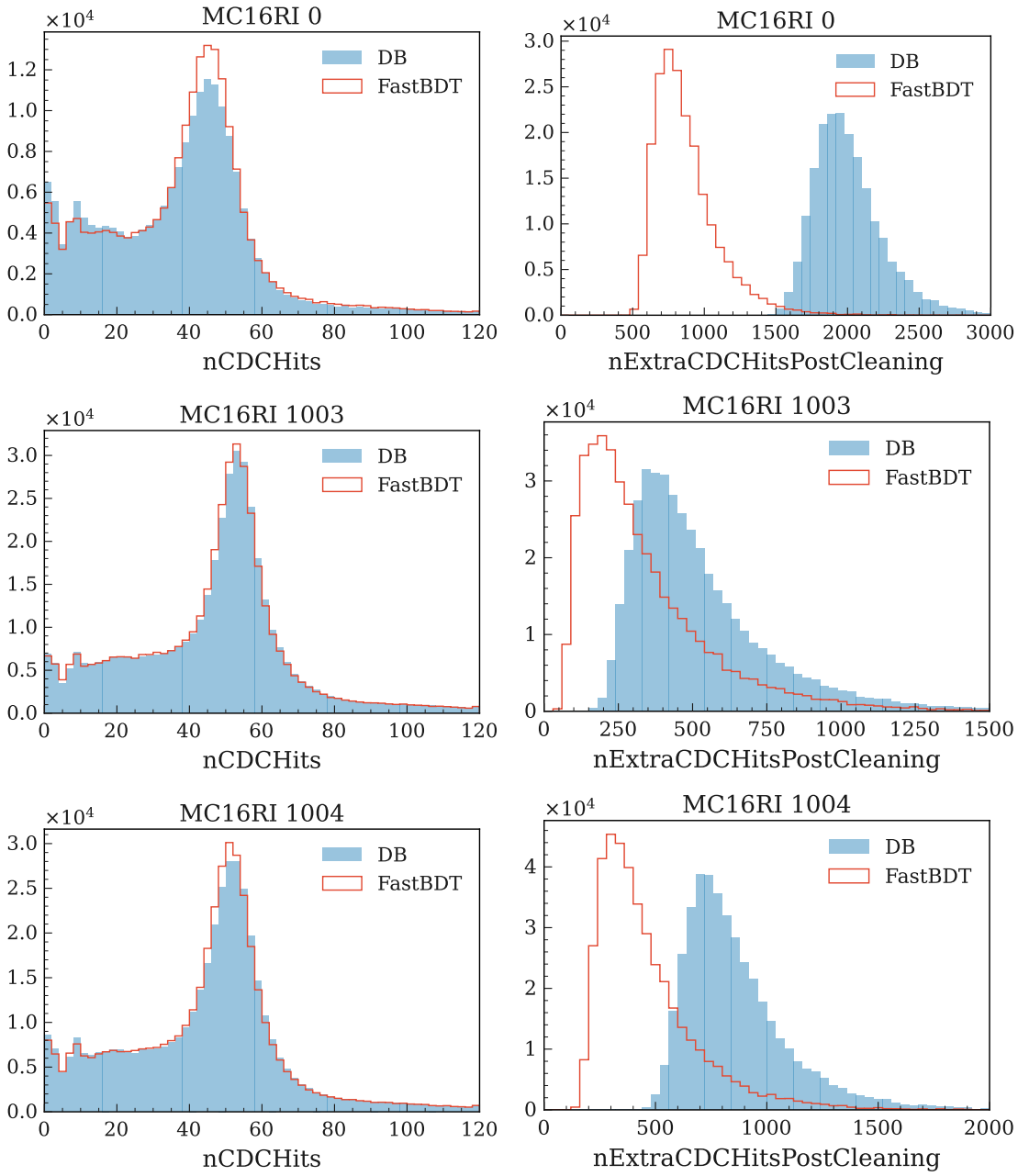

Figure 6.18: Distributions of extra CDC hits in MC16ri. Left: total number of hits not assigned to tracks; Right: post-cleaning hits, with DB filter (blue) and nominal FastBDT classifier (red) with Experiment 0, 1003, and 1004.

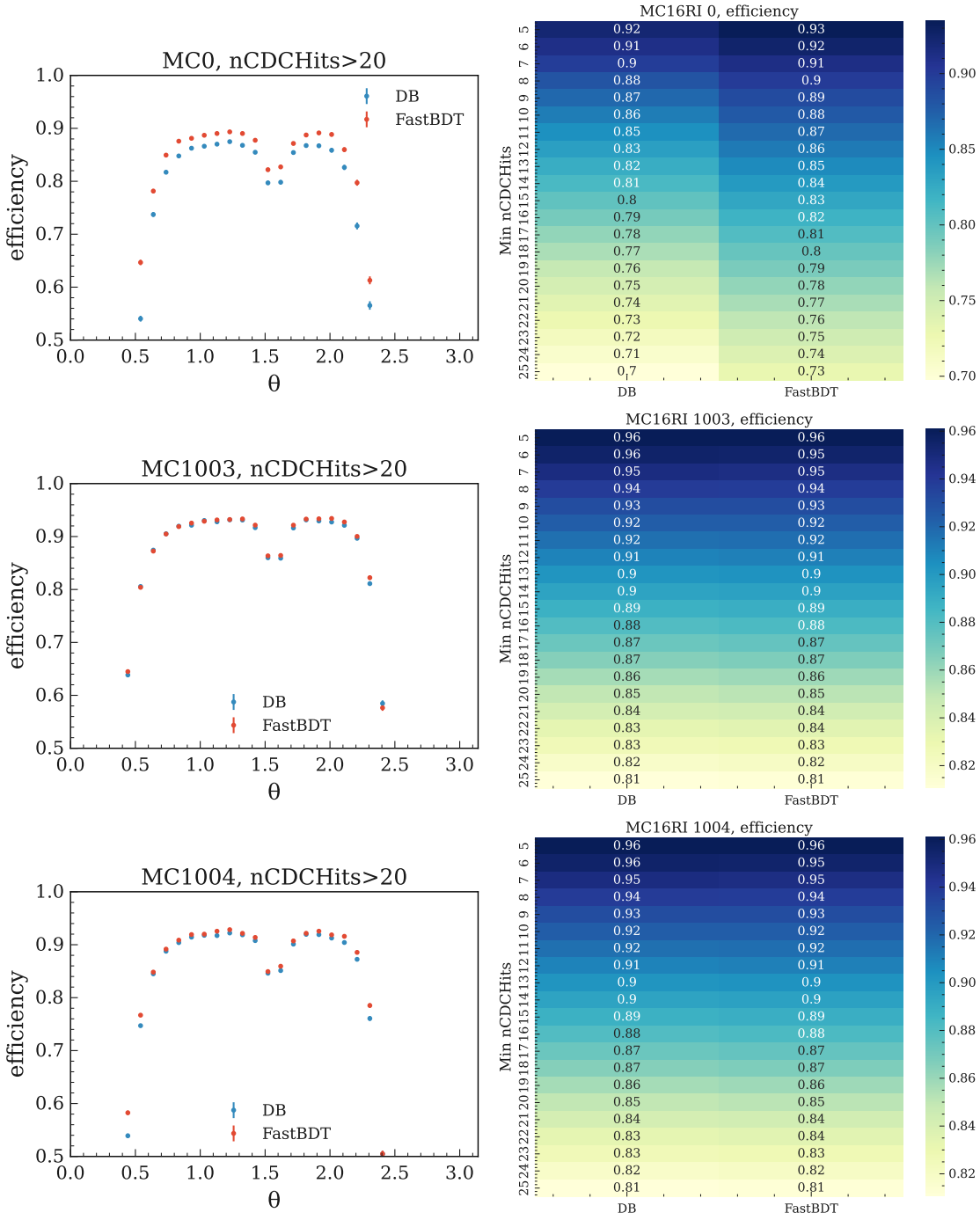

Figure 6.19: Left: Track-finding efficiency $\varepsilon(\theta; i)$ as a function of the polar angle θ for minimum hit requirements nCDCHits > 20. Right: Tracking efficiency in MC16ri as a function of the minimum hit requirement i. Compared between the DB filter and the nominal FastBDT classifier on the MC16ri simulation.

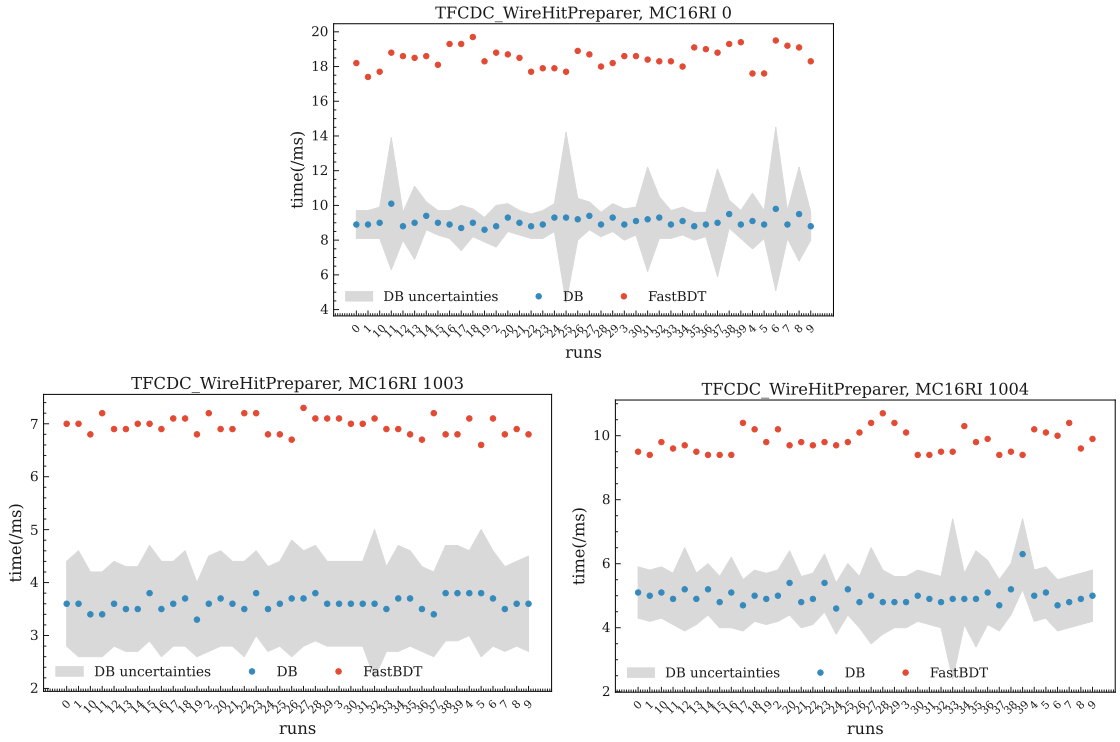

Figure 6.20: Average per-event runtime of key CDC standalone modules in MC16ri, comparing DB filter (blue) and FastBDT (red) for `TFCDC_WireHitPreparer`,

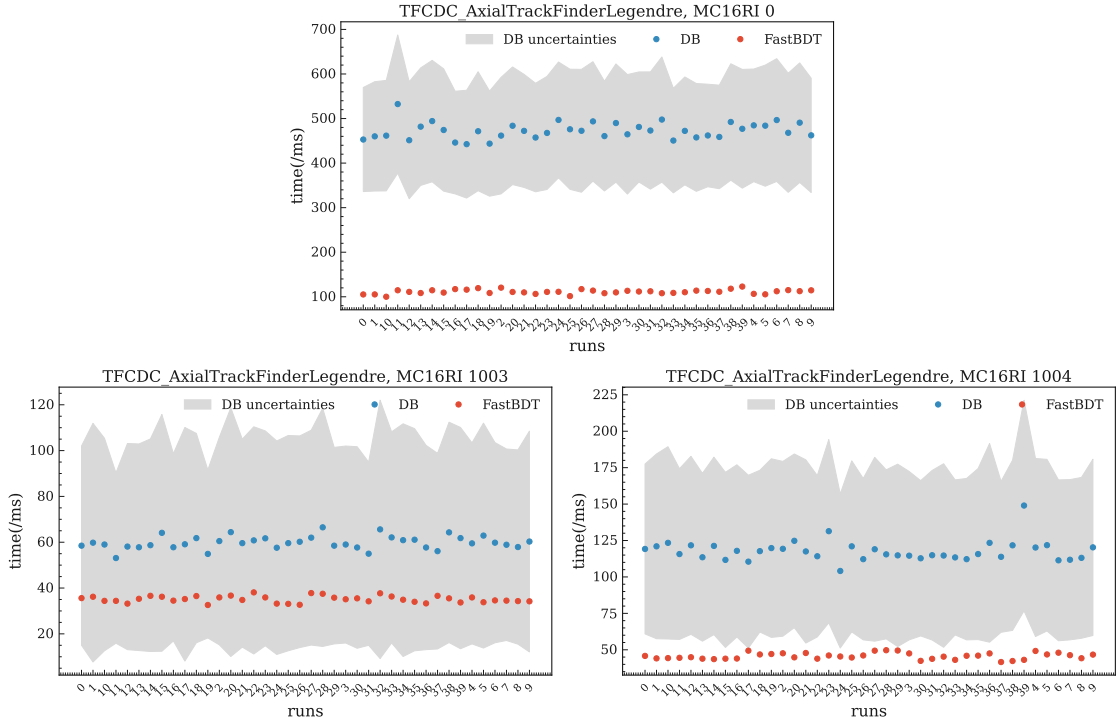

Figure 6.21: Average per-event runtime of key CDC standalone modules in MC16ri, comparing DB filter (blue) and FastBDT (red) for `TFCDC_SegmentFinderFacetAutomation`

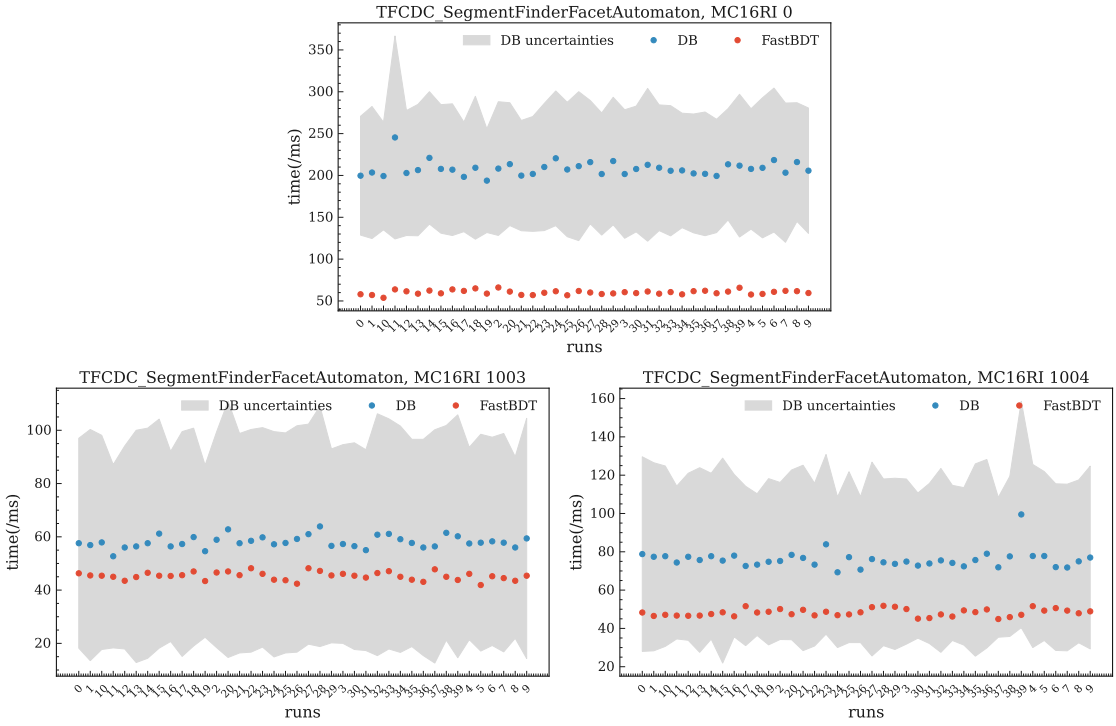

Figure 6.22: Average per-event runtime of key CDC standalone modules in MC16ri, comparing DB filter (blue) and FastBDT (red) for `TFCDC_AxialTrackFinderLegendre`

with the remaining discrimination provided by the superlayer index and TDC. Efficiency studies indicated that retaining more than 90% of attached-track hits allows rejection of up to 80% of background hits, corresponding to a working-point cut of dimuon-FastBDT score larger than 0.5, shown in Fig. 6.23
Comparisons of the dimuon-trained classifier (denoted as `FastBDT, dimuon`), The nominal hadron classifier and the default database filter, as shown in Fig. 6.24, reveal the following trends:

- In the low- and mid-ADC regions, the dimuon FastBDT more effectively suppresses background while maintaining similar attached-track hit efficiency in the mid-ADC range.
- A minor reduction in signal is observed in the lowest ADC region, consistent with TOT being the most important variable.
- For TOT distributions, the dimuon classifier prioritizes background suppression, whereas the nominal hadron classifier retains slightly more signal at low ADC.
- For TDC, the dimuon classifier removes more background compared to the nominal classifier, but both BDTs outperform the database filter by reducing background while preserving signal hits.

Overall, these studies demonstrate that the nominal FastBDT generalizes well to lepton-dominated events. The dedicated dimuon classifier can provide slightly improved background rejection in low- and mid-ADC regions. The `Bhabha` veto ensures that contamination from high-occupancy QED events is removed, confirming that the performance of the classifier is robust across different event topologies and particle types.
For `nCDCHits`, both BDTs exhibit a shift in the distribution toward tracks containing fewer hits compared to the default DB filter. More tracks fall into the 10-45 hit range, while the population of tracks with more than 45 hits decreases under the BDT-based selection. The dimuon classifier shows a slightly stronger shift than the nominal hadron classifier, indicating a somewhat more aggressive suppression of high-hit-count tracks. The post-cleaning variable also shows clear improvements under the BDTs. The DB filter displays a

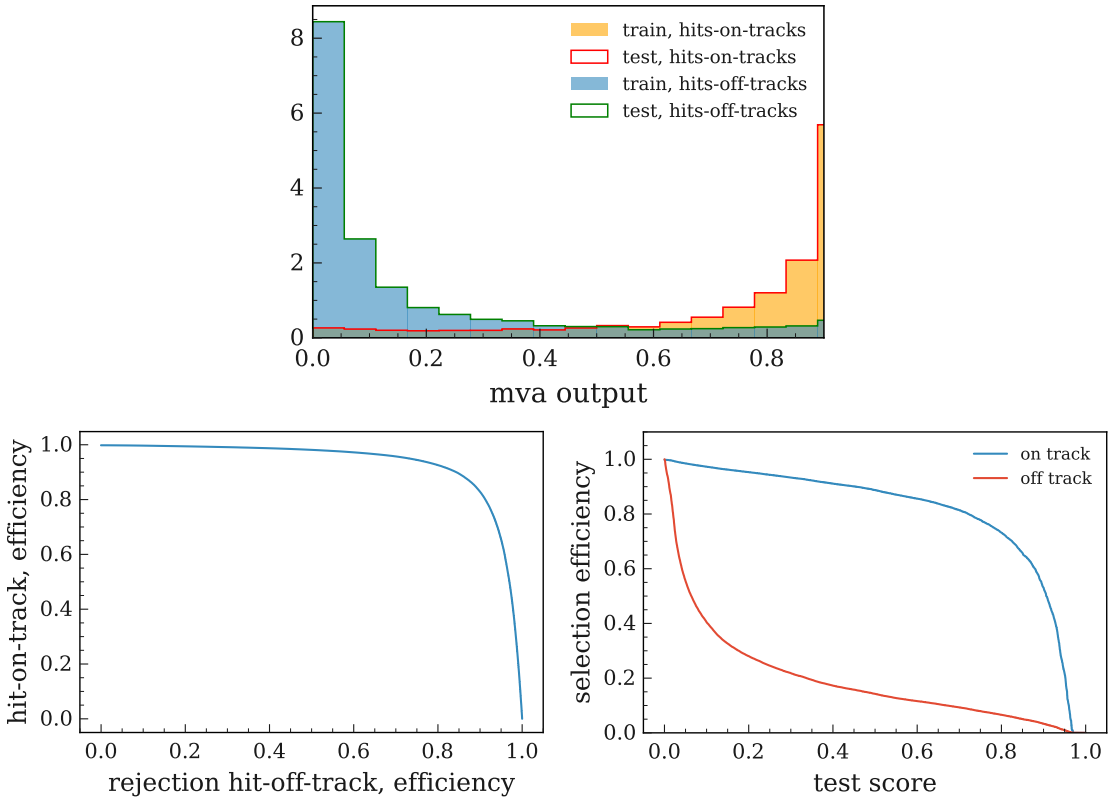

Figure 6.23: Performance of the classifier trained on $e^+e^- \to \mu^+\mu^-(\gamma)$ data. Upper: comparison of training and test scores, showing no evidence of overfitting. Bottom left: distribution of hits on-track efficiency versus off-track efficiency. Bottom right: selection efficiency versus MVA threshold for track-associated (signal) and off-track (background) hits, indicating a good separation and a stable working point around 0.5.

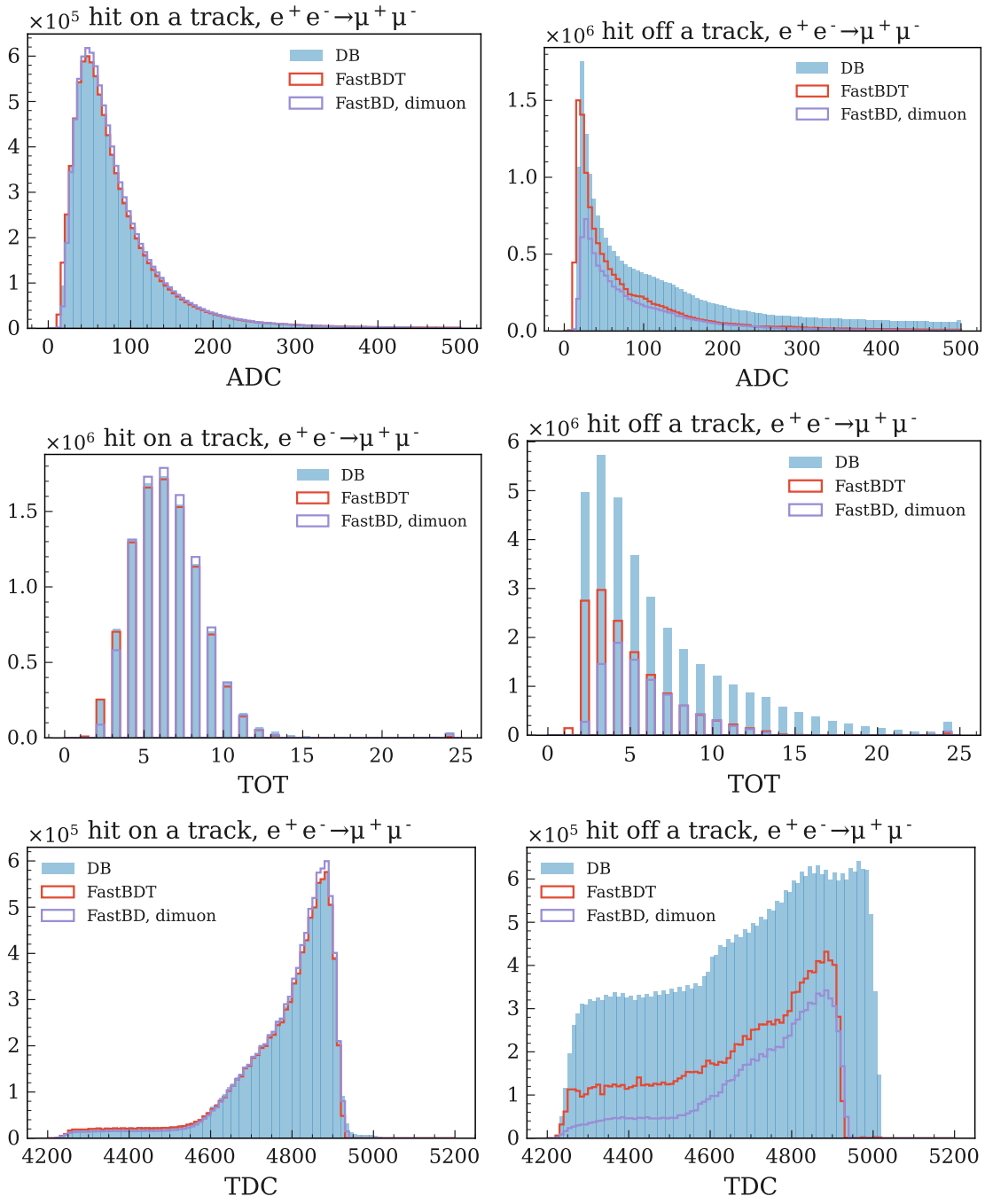

Figure 6.24: Distributions of CDC hit observables for $e^+e^- \to \mu^+\mu^-(\gamma)$ data after filtering. Left panels: ADC, TOT, and TDC distributions. Right panels: unattached-track hits in the same observables. Blue: DB filter, red: nominal FastBDT, purple: dimuon-FastBDT. The MVA filters preserve nearly all track-associated hits while significantly suppressing extreme unattached hits contributions, such as very large TDC values and 30–40% of hits across the TOT range.

relatively broad distribution, ranging from approximately 25 to 250 hits per event, with a slowly decreasing tail extending beyond 600 hits. Both BDTs significantly reduce this background: The nominal classifier concentrates most events in the 0-300 hit range with a peak near 90 hits, while the dimuon classifier achieves even stronger suppression, reducing the distribution to 0-200 hits with a peak around 60, as shown in Fig. 6.25. This demonstrates the effectiveness of the classifiers in cleaning residual crosstalk and spurious hits beyond what is achievable with a static DB filter.

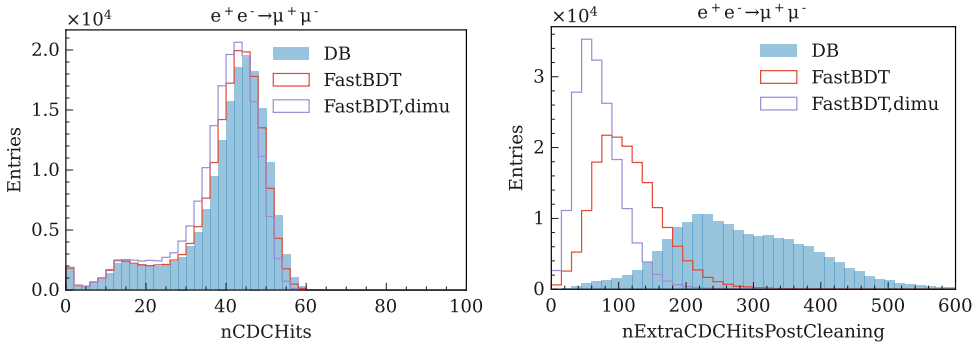

Figure 6.25: Distributions of extra CDC hits in $e^+e^- \to \mu^+\mu^-(\gamma)$ data. Left: total number of hits not assigned to tracks; Right: post-cleaning hits, with DB filter (blue), nominal FastBDT classifier (red), and FastBDT trained on $e^+e^- \to \mu^+\mu^-(\gamma)$ data (purple).

The efficiency as a function of polar-angle acceptance was also examined for the dimuon sample. Overall, the distributions are smooth and uniform, with only two small deviations: a drop in efficiency at very forward angles close to the beam axis, and a minor reduction at exactly perpendicular directions. Both effects are small, and in particular, the drop at 90° is negligible. As expected for an extremely clean two-track final state, the performance of the BDT classifiers is very similar to that of the database filter. Across working points from i >5 to i >25, the database filter yields efficiencies decreasing from about 99% to 89%, while the BDT classifiers give values ranging from 99% to 88%. This corresponds to only a 0-1% difference, as shown in Fig. 6.26 indicating that the classifiers do not introduce any significant inefficiency in the acceptance, apart from the extreme beam-axis regions, where reconstruction is intrinsically challenging.

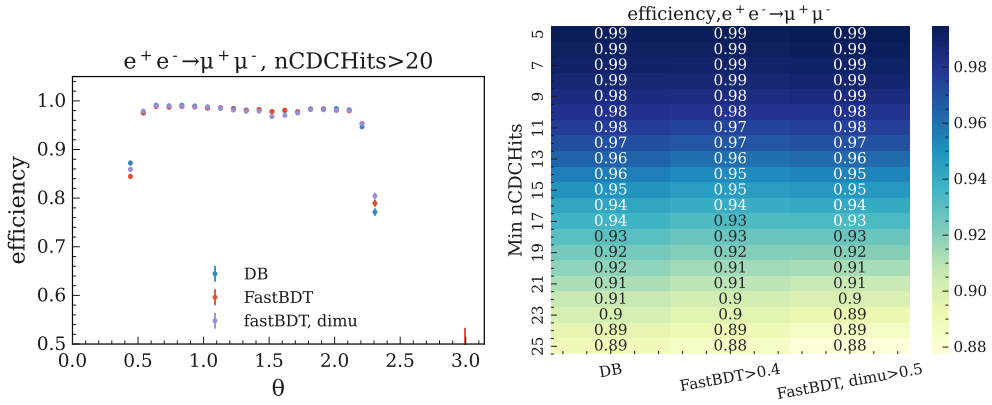

Figure 6.26: Left: Track-finding efficiency $\varepsilon(\theta; i)$ as a function of the polar angle θ for minimum hit requirements nCDCHits > 20. Right: Tracking efficiency in $e^+e^- \to \mu^+\mu^-(\gamma)$ data as a function of the minimum hit requirement i. Compared between the DB filter, the classifier trained on Experiment 26 hadron data and on $e^+e^- \to \mu^+\mu^-(\gamma)$ data.

The computational performance of the classifiers was further evaluated through module-level timing studies

as a function of the run number.
For the `TFCDC_WireHitPreparer` module, the DB filter shows a smooth run-dependent increase in processing time, rising gradually from approximately 2 to 5 ms per event as the run number increases. The BDT classifiers follow the same overall trend, but at higher values (4-10 ms per event) and with larger fluctuations between runs. For the `AxialTrackFinderLegendre` module, the DB filter again exhibits strong run dependence, with the timing rising from about 10 ms to 30 ms per event and reaching fluctuations up to 50 ms for large run numbers. By contrast, the BDT classifiers behave more mildly, with times increasing only from 1 ms to 12 ms over the same run range. In this context, the BDT classifiers appear to be more run-independent than the DB filter.
A similar pattern is observed for the `TFCDC_SegmentFinderFacetAutomation` module. The DB filter shows times increasing from about 10 ms to 18 ms per event, with some runs peaking near 24 ms. The BDT classifiers again behave more stably, with times gradually rising from 2 ms to 9 ms per event, Fig. 6.27. Overall, these comparisons suggest that, despite higher average processing times in some modules, the BDT classifiers provide a more stable and less run-dependent timing behavior than the database filter.

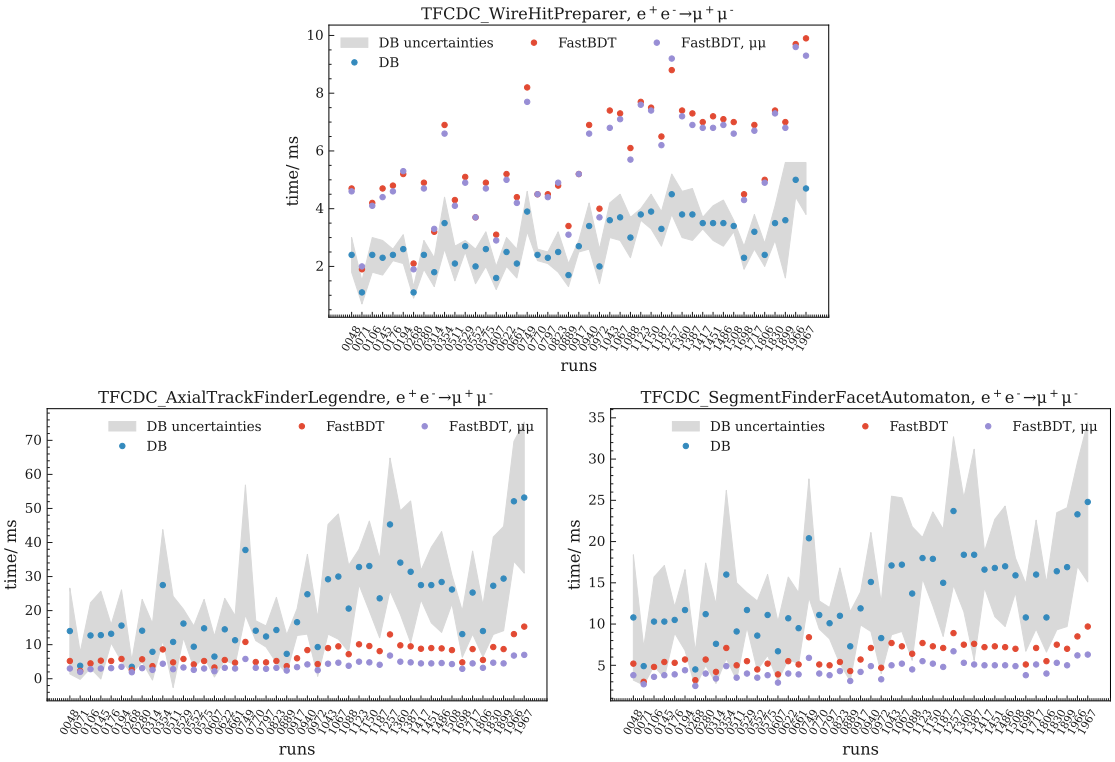

Figure 6.27: Average per-event runtime of key CDC standalone modules in $e^+e^- \to \mu^+\mu^-(\gamma)$ data, comparing DB filter (blue), nominal FastBDT (red), and dimuon-FastBDT (purple) for `TFCDC_WireHitPreparer`, `TFCDC_SegmentFinderFacetAutomation`, and `TFCDC_AxialTrackFinderLegendre`.

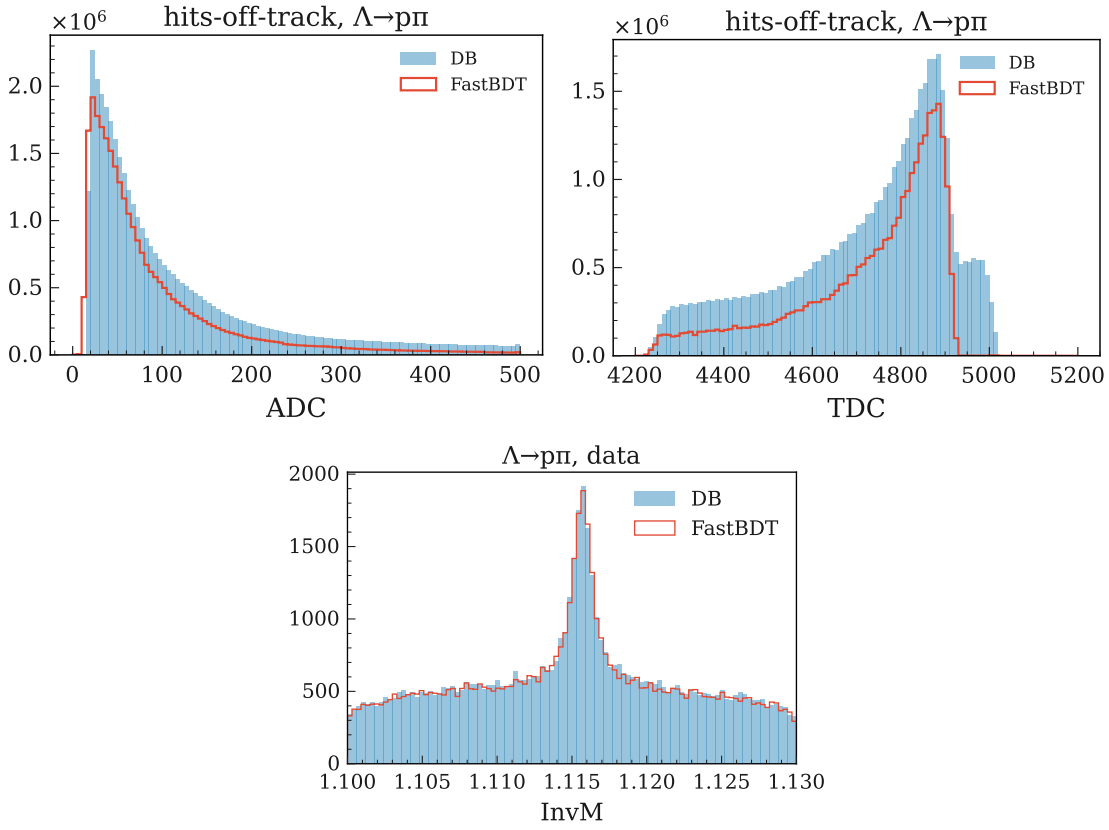

Figure 6.28: Distributions of ADC, TDC, and invariant mass in $\Lambda \to p\pi$ data. Blue: DB filter; Red: FastBDT. The FastBDT preserves the expected distributions while reducing spurious contributions.

6.4.4 Validation on Low-momentum Proton Sample

Protons at low momentum represent a particularly challenging case for CDC reconstruction and background filtering. Due to their higher ionization energy loss dE/dx and stronger multiple scattering compared to lighter particles such as pions or muons, they typically leave denser clusters of hits and suffer larger track curvature. In the CDC, these characteristics can mimic background-like features and therefore require careful validation to ensure that the MVA filter does not misclassify genuine proton hits as crosstalk or noise. Since the background filter operates at the raw hit and wire-hit level, no dedicated low-momentum proton control sample is directly available.

As no official skim dedicated to proton exists, a $\Lambda \to p\pi$ sample is produced from hadron data by applying a custom skim and reconstructing from hadron data. The skimming procedure selects Λ candidates from hadron data using the standard Belle II vertex fit and mass window requirements, followed by proton-pion separation through the momentum asymmetry and a minimum flight significance cut. This strategy ensures a reliable sample of protons with momenta down to 0.1 GeV/c, suitable for validating the robustness of the CDC background filter in the low-momentum regime. Initial studies using hadron data showed that the hit- and track-level distributions under both the database filter and the MVA-based filter behaved as expected. In particular, FastBDT preserves the expected distributions even after post-cleaning, as shown in Fig. 6.28.

However, the Λ invariant mass spectrum exhibited a sharp peak accompanied by substantial combinatorial background. Such contamination would bias any attempt to quantify the filter response to genuine protons. To address this, the `sPlot` [116] technique was adopted, which statistically separates signal and background contributions, provided that a reliable simulation-based template for the signal shape is available.

A dedicated simulated Λ sample was therefore generated to provide a clean signal shape for the `sPlot`

procedure. The generation process used inclusive continuum $c\bar{c}$ events, allowing hadronization into $\Lambda_c^{\pm}$ baryons, which were forced to decay via

$$\Lambda_c^{\pm} \to \Lambda^0 \pi^{\pm}, \quad \Lambda^0 \to p\pi^{-}. \tag{6.3}$$

Background levels were controlled to ensure high purity. Reconstruction of simulated Λ candidates was performed with a mass-constrained vertex fit, followed by quality selections including flight significance

$$\text{Flight Significance} = \frac{\text{Flight Distance}}{\text{Uncertainty of Flight Distance}} > 3, \tag{6.4}$$

momentum asymmetry between proton and pion

$$\text{Asymmetry of Daughters} = \frac{p_{\text{proton}} - p_{\text{pion}}}{p_{\text{proton}} + p_{\text{pion}}} > 0.4, \tag{6.5}$$

proton identification likelihood $\text{proton}_{\text{ID}}$>0.1, at least one CDC hit for both daughters, and geometrical consistency $\cos\theta(\Lambda, \text{vertex})$>0.9. Under these selections, the reconstructed Λ invariant mass distribution exhibits a clean, narrow peak with negligible combinatorial background, which is ensured by the generated decay chains, as shown in Fig. 6.29 (upper panel).

Several interesting features are observed in the distributions of reconstructed $\Lambda \to p\pi$ candidates. The polar angle distribution exhibits two clear peaks, one in the range 0.5-1.5, which corresponds predominantly to forward-going lambda, and another in 1.5-2.5, corresponding predominantly to backward-going pions originating from the $\Lambda_c^{\pm}$, as shown in Fig. 6.29 (bottom left panel). This separation reflects the kinematics of the $\Lambda_c^{\pm}$ decay in the laboratory frame and provides a natural distinction between the daughter particles in the detector acceptance. Examination of the post-cleaning variable reveals a corresponding two-peak structure, shifted relative to the single-peak distribution observed with the database filter, as shown in Fig. 6.29 (bottom right panel). This indicates that the MVA-based FastBDT filter applies different cleaning strength to protons and pions, likely reflecting the differential hit patterns and ionization characteristics of the two particle types. Such observations support the conclusion that the filter effectively suppresses extraneous hits while retaining genuine signals, with nuanced behavior depending on particle species and kinematics.

To isolate genuine $\Lambda \to p\pi$ decays from the combinatorial background, the `sPlot` technique is employed. The signal shape is modeled using the simulated $\Lambda \to p\pi$ sample, with a double-Gaussian convolution to account for the detector resolution. The combinatorial background is described by a Chebyshev polynomial function. Separate fits are performed for events passing the DB filter and the FastBDT filter. The fit results, shown in Fig. 6.30, yield 8,466 signal and 50,319 background events for the DB filter, and 8,536 signals with 48,942 background events for the FastBDT filter.

The `sPlot` method assigns to each event a weight, called the `sWeights`, which represents the probability that the event originates from the signal rather than a background component. Physically, this allows the extraction of pure signal distributions from the data, even in the presence of significant background, without introducing bias from the fitting procedure. By applying the `sWeights` to the hadron data, a clean $\Lambda \to p\pi$ sample is obtained, providing an unbiased set of protons suitable for validating the CDC background filter at the hit and track levels.

Using the `sPlot`-extracted clean $\Lambda \to p\pi$ sample, we first examine the kinematic distributions under the DB and FastBDT filters, as shown in Fig. 6.31. The invariant mass (InvM) of Λ distribution exhibits a sharp peak around $1.115\ \text{GeV}/c^2$, consistent with the nominal Λ mass. Negative bins observed in the invariant mass residuals in the ranges 1.100-1.113 GeV/c^2 and 1.118-1.130 GeV/c^2 reflect the background-subtracted `sWeights`. Physically, these negative weights arise because, in regions dominated by combinatorial background, the `sPlot` assigns negative weights to over-subtracted events, ensuring that the sum of weighted events reproduces the total signal yield without bias. This is a well-understood feature of the `sPlot` formalism [116]. Examining the daughter particle momenta, the pions (p_{pion}) are slow, with momenta primarily in the $0 - 0.6\ \text{GeV}/c$ range, while protons cover a broader range, from 0.1 up to $3\ \text{GeV}/c$. This reflects the Λ decay kinematics, where the heavier proton typically carries more momentum than the lighter pion.

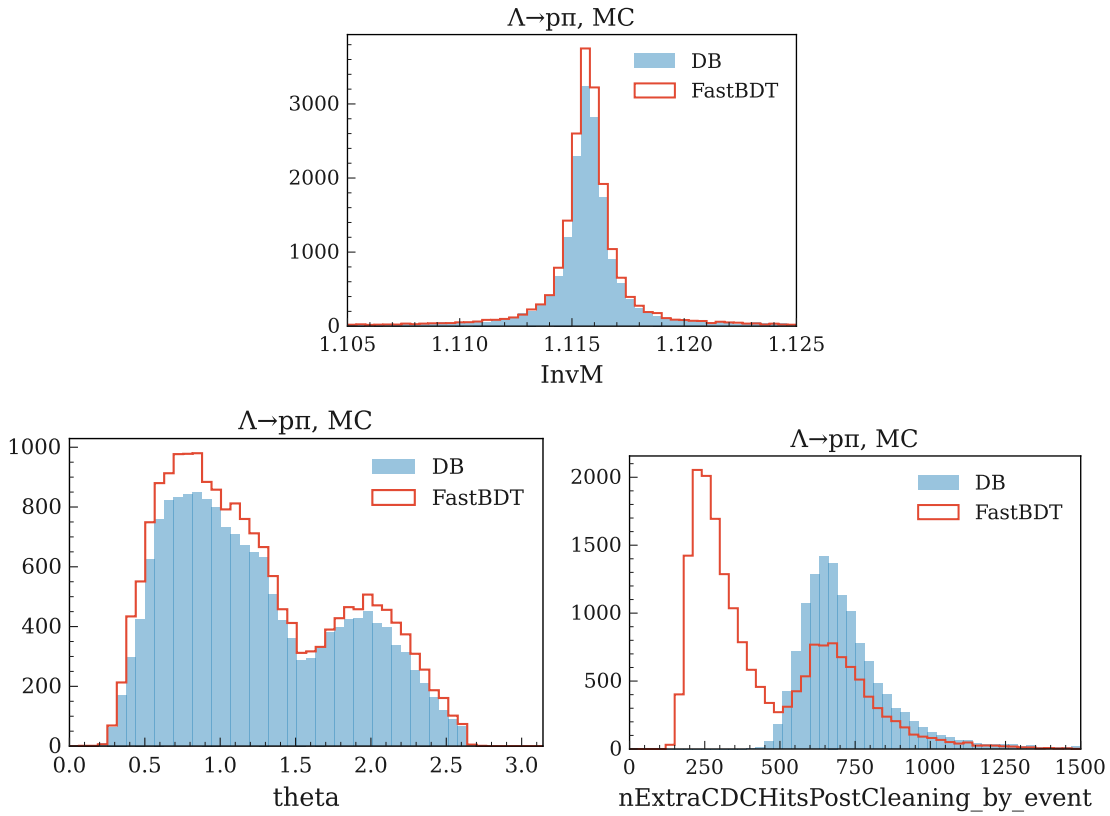

Figure 6.29: Simulated $\Lambda \to p\pi$ events. Upper: invariant mass; Bottom left: polar angle distribution showing forward/backward peaks; Bottom right: post-cleaning hit distribution, where FastBDT reduces excess hits compared to the DB filter.

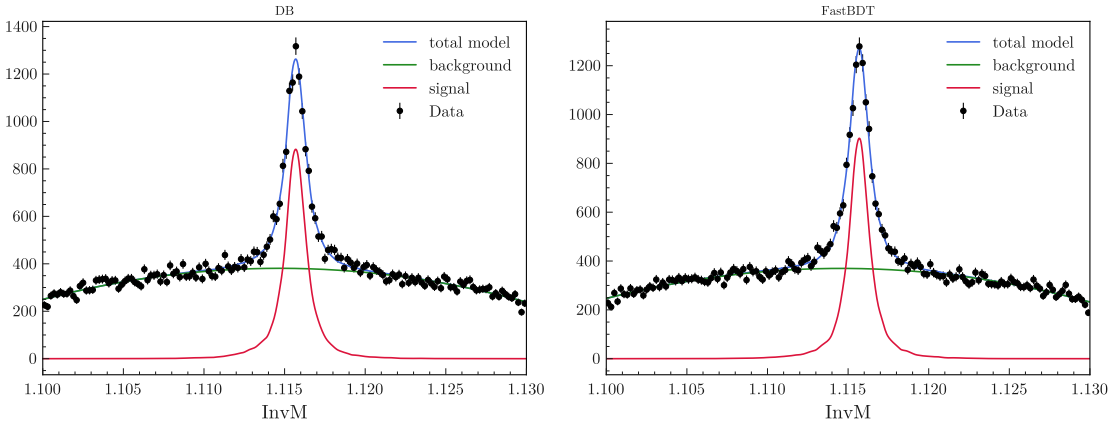

Figure 6.30: Fits to the invariant mass distribution in $\Lambda \to p\pi$ data, using signal templates from simulation and a polynomial background. Left: DB filter; Right: FastBDT filter. The FastBDT retains the same signal yield while reducing background.

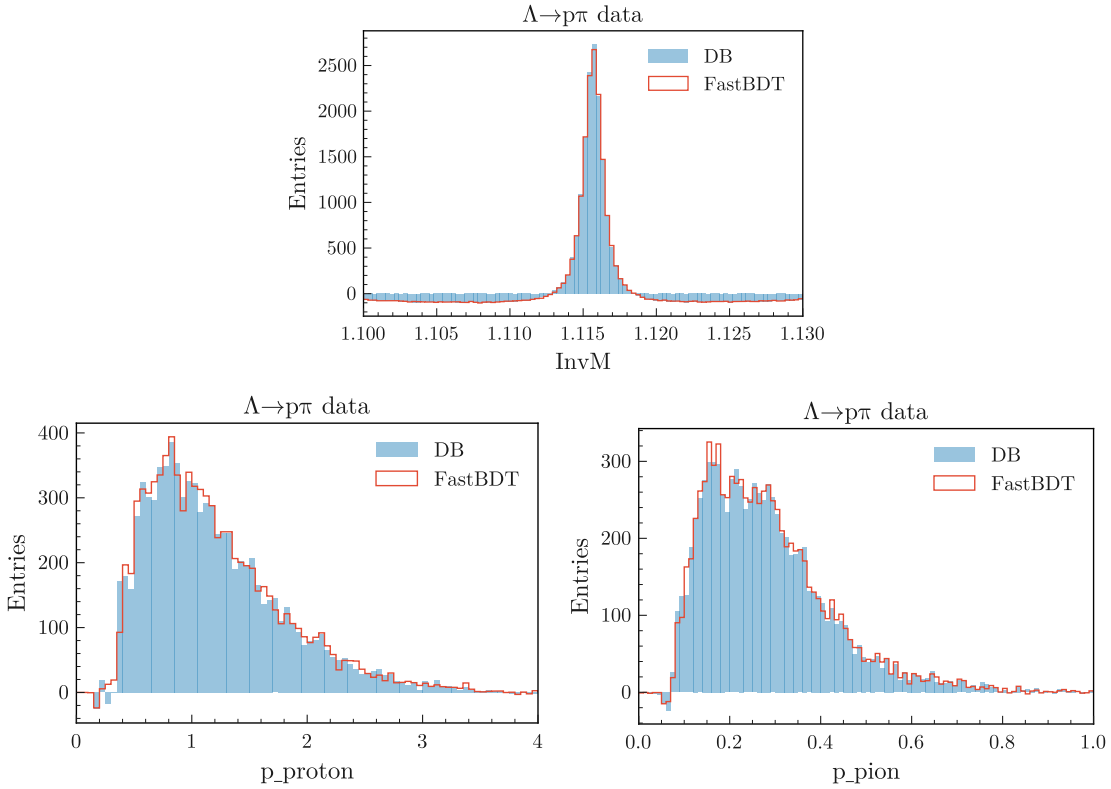

Figure 6.31: sPlot-extracted $\Lambda \to p\pi$ sample. Left: invariant mass after background subtraction; Middle: proton momentum spectrum; Right: pion momentum spectrum. Negative bins arise from s-weights in background-dominated regions.

The tracking post-cleaning variable shows a clear shift between DB and FastBDT filters. For DB-filtered events, the number of CDC hits per Λ event spans roughly $200 - 1000$, whereas FastBDT reduces this range to $50 - 800$, indicating a substantial reduction in extraneous hits while preserving the genuine signal.
Using the `sPlot`-extracted clean $\Lambda \to p\pi$ sample, the track-finding efficiency was evaluated as a function of proton momentum rather than detector acceptance. The efficiency is defined as in Eq. 6.1, based on the number of CDC hits associated with a reconstructed track ($n_{\text{CDCHits}} \in [5, 25]$).
For protons with momentum larger than 0.4 GeV/c, the overall efficiency under the DB filter ranges from 99% to 91%, while FastBDT shows a slight decrease from 99% to 89%, corresponding to a maximal efficiency loss of ∼2% (see the first two columns in Fig. 6.32). To investigate the contributions from high- and low-momentum protons, the sample is split into $p < 1$ GeV/c and $p > 1$ GeV/c. High-momentum protons exhibit nearly identical efficiencies for both filters (last two columns in Fig. 6.32), whereas low-momentum protons display a small efficiency reduction under FastBDT (middle two columns).
A more detailed evaluation was performed by calculating the efficiency as a continuous function of proton momentum in the range $0.4 - 2.3$ GeV/c, i.e., $\varepsilon(p < i/p_{\text{all}})$. The efficiency for very low-momentum protons ($p < 0.4$ GeV/c) is only 1.8% under DB and further reduced to 1.1% under FastBDT. A slight increase is observed for $p < 0.5$ GeV/c, with efficiencies of 5.7% and 5.6% for DB and FastBDT, respectively. As the momentum increases to 1 GeV/c, the efficiency rises to ∼44%, and for $p < 2.3$ GeV/c, it reaches 94%, with FastBDT closely following the same behavior, as shown in Fig. 6.33.
Time consumption was not evaluated for this study, since the application of `sWeights` [116] to the hadron data sample prevents a fair comparison of computational performance between the DB and FastBDT filters.
These results demonstrate that FastBDT preserves genuine proton tracks with only a marginal efficiency loss (∼2%) at the lowest momenta, confirming its reliability across particle species.

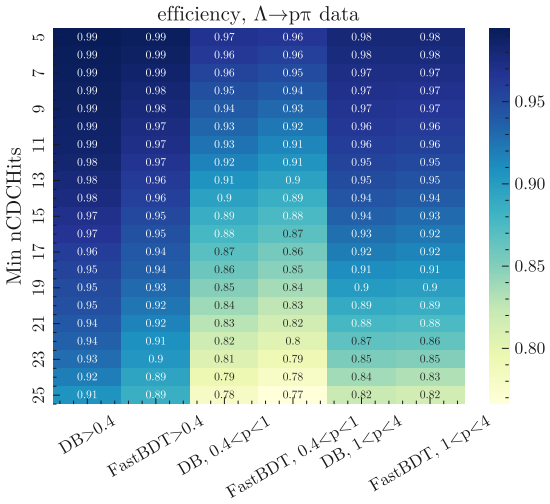

Figure 6.32: Track-finding efficiency for protons in the sPlot-cleaned $\Lambda \to p\pi$ sample. Efficiencies for different proton momentum ranges and filters. The first two columns show DB and FastBDT for $p_{\text{proton}} > 0.4\,\text{GeV}/c$. The next two columns show DB and FastBDT for low-momentum protons ($0.4 < p < 1\,\text{GeV}/c$). The last two columns show DB and FastBDT for higher-momentum protons ($1 < p < 4\,\text{GeV}/c$).

6.4.5 Standard Validation

The `basf2` framework provides an automated, general tracking validation using experiment 0 simulation. This validation computes key performance metrics to evaluate the full tracking chain, including CDC standalone reconstruction as well as the global track fit incorporating SVD and PXD hits. The following quantities are calculated [77]:

- **Finding efficiency:** $\varepsilon_{\text{find}} = \frac{\text{matched tracks}}{\text{all primary tracks}}$
- **Charge efficiency:** $\varepsilon_{\text{charge}} = \frac{\text{matched tracks with correct charge}}{\text{matched primary tracks}}$
- **Finding + charge efficiency:** $\varepsilon_{\text{find+charge}} = \frac{\text{matched tracks with correct charge}}{\text{all primary tracks}}$
- **Fake rate:** fraction of pattern-recognition tracks not associated with any true particle
- **Clone rate:** ratio of duplicate reconstructed tracks relative to all tracks associated with a particle
- **Charge asymmetry:** $\frac{N^+_{\text{matched}} - N^-_{\text{matched}}}{N^+_{\text{matched}} + N^-_{\text{matched}}}$
- **Hit efficiency:** efficiency of hits associated with matched tracks

This standard validation serves as a baseline assessment of CDC standalone and full tracking performance under nominal conditions. It enables consistent monitoring of the tracking quality and provides a framework to evaluate the impact of any modifications, such as the implementation of a CDC background filter, on key performance indicators.
The results of the standard `basf2` tracking validation are summarized in Tab. 6.2 for the CDC standalone performance and in Tab. 6.3 for the full tracking chain. Four configurations are compared: (i) the default database requirements (DB filter), (ii) the nominal hadron-trained MVA (FastBDT), (iii) an alternative MVA trained on $e^+e^- \to \mu^+\mu^-$ events with `FastBDT` > 0.5 ($e^+e^- \to \mu^+\mu^-$ FastBDT), and (iv) an MVA trained on experiment 0 simulation (MC Experiment 0 FastBDT), which is expected to yield the best performance since the training sample has a background composition closely matching that of the validation dataset.

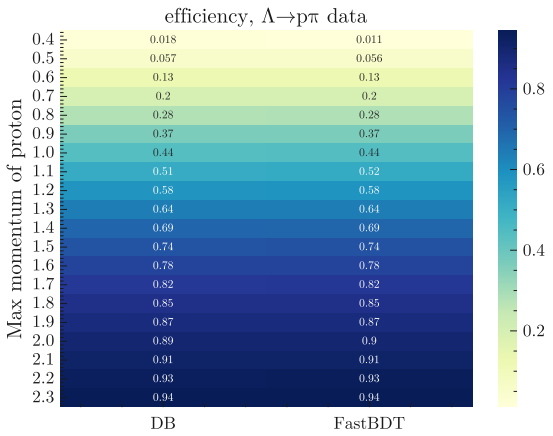

Figure 6.33: Track-finding efficiency for protons in the sPlot-cleaned $\Lambda \to p\pi$ sample. Efficiency as a continuous function of proton momentum. FastBDT closely follows the DB filter, with only a small (∼2%) reduction at very low momentum, while matching DB performance for higher-momentum protons.

At the CDC standalone level, both the hadron- and muon-trained MVAs show significant improvements over the default database filter. The finding efficiency increases from 71.3% (DB) to 80.9% (FastBDT) and 81.7% ($e^+e^- \to \mu^+\mu^-$ FastBDT), with a further rise to 83.2% for the MC Experiment 0 FastBDT. Charge efficiency remains consistently high (∼97%) across all methods, while charge asymmetry is reduced compared to the database filter, indicating better balance between positive and negative tracks. The fake and clone rates are stable across all configurations, with a slight reduction in the MVA cases. The hit efficiency shows a notable improvement, from 80.6% (DB) to 85.7% (MC Experiment 0 FastBDT).

CDC standalong (%)	DB filter	FastBDT	$e^+e^- \to \mu^+\mu^-$ FastBDT	MC Experiment 0 FastBDT
ϵ(finding + charge)	69.1	78.5	79.6	80.8
ϵ(finding)	71.3	80.9	81.7	83.2
ϵ(charge)	96.9	97.1	97.5	97.1
charge asymmetry	2.31	1.60	0.68	0.84
fake rate	3.39	3.52	2.88	2.93
clone rate	0.65	0.64	0.58	0.53
ϵ(hit)	80.6	82.4	81.0	85.7

Table 6.2: Standard CDC standalone tracking validation results from Experiment 0 simulation. Comparison of the DB filter, the nominal FastBDT, the dimuon-trained FastBDT, and the MVA trained directly on Experiment 0 simulation. Efficiencies, fake and clone rates, and charge asymmetry are expressed as percentages.

In the full tracking validation, which includes SVD and PXD information, the relative improvements are smaller but still visible. Finding efficiency increases from 91.4% (DB) to 93.1% (MC Experiment 0 FastBDT), while charge efficiency remains unchanged at ∼98.9%. The charge asymmetry, already small in the database filter, is further reduced and nearly vanishes in the muon-trained and MC Experiment 0 MVA cases. Both fake and clone rates show systematic decreases, with the clone rate reduced from 3.86% (DB) to 3.31% (MC Experiment 0 FastBDT). Finally, hit efficiency improves from 75.4% to 81.1%. The MC Experiment 0 trained MVA achieves the best performance in this validation, which is expected since the training and validation samples share the same background composition. This result should therefore be interpreted as an approximate upper bound on achievable performance under nominal conditions.
Overall, these results demonstrate that MVA-based CDC background filters enhance track-finding efficiency and hit efficiency without increasing fake or clone rates. The hadron-trained filter provides robust per-

formance in realistic conditions, while the MC Experiment 0 trained filter yields the best results in this validation, as expected, given the close match between training and validation samples. These findings confirm that MVA-based filtering constitutes a reliable and effective replacement for the static database requirements.

full tracking (%)	DB filter	FastBDT	$e^+e^- \to \mu^+\mu^-$ FastBDT	MC Experiment 0 FastBDT
ϵ(finding + charge)	90.3	91.5	91.9	92.0
ϵ(finding)	91.4	92.5	92.9	93.1
ϵ(charge)	98.8	98.9	98.9	98.8
charge asymmetry	0.37	0.34	0.07	-0.15
fake rate	7.05	6.44	6.09	6.38
clone rate	3.86	3.60	3.41	3.31
ϵ(hit)	75.4	78.5	77.8	81.1

Table 6.3: Standard full tracking validation results (including CDC, SVD, and PXD) from Experiment 0 simulation. Comparison of the DB filter, nominal FastBDT, the dimuon-trained FastBDT, and the MVA trained directly on Experiment 0 simulation. Efficiencies, fake and clone rates, and charge asymmetry are expressed as percentages.

6.5 Performance

6.5.1 Performance in Official Belle II Software

The MVA-based CDC background filter was officially merged into the Belle II software framework with `Release 9` (July 2024). The impact on tracking performance was evaluated by the Belle II tracking group through the standard full tracking validation. Fig. 6.34 compares the track-finding efficiency as a function of transverse momentum, p_T, between Release 8 (without this MVA background filter) and `Release 9` (with the filter included). A visible improvement in efficiency is observed, extending from very low p_T up to approximately $1.6\ \mathrm{GeV}/c$.

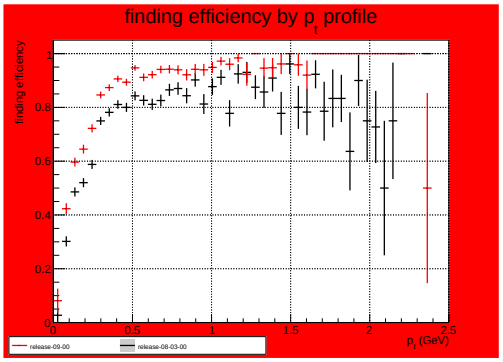

Figure 6.34: Track-finding efficiency as a function of transverse momentum p_T, comparing Belle II software Release 8 (without the MVA CDC background filter) and Release 9 (with the filter). A visible efficiency gain is seen from very low p_T up to about 1.6 GeV/c.

In addition to efficiency improvements, a substantial reduction in computational time was reported. The average per-event runtime of the `TFCDC_SegmentFinderFacetAutomation` module decreased from approximately $150\ \mathrm{ms}$ to $30\ \mathrm{ms}$, while the `TFCDC_AxialTrackFinderLegendre` module dropped from $\sim 400\ \mathrm{ms}$ to $\sim 90\ \mathrm{ms}$. Similarly, the `ToCDCCKF` module reduced from $35\ \mathrm{ms}$ to $20\ \mathrm{ms}$ per event, as shown in Fig. 6.35.
The results, summarized in Figs. 6.34 and 6.35, confirm that the official implementation of the CDC background filter in Release 9 achieves the dual goals of enhanced tracking efficiency, particularly at low transverse momenta and reduced computational cost, thereby improving both the physics performance and operational scalability of Belle II tracking.

6.5.2 Impact on Neutral Hadron Reconstruction

A dedicated study from the Belle II Tracking Group further investigated the impact of the CDC background filter on the reconstruction of K_S^0. In particular, the reconstruction efficiency of K_S^0 in function of decay radius was examined. These (late) K_S^0 mesons, with decay vertices beyond $8\ \mathrm{cm}$, rely heavily on the CDC for successful reconstruction.
The study demonstrated a significant improvement in efficiency with the new filter. In Run 1 (Experiment 26 data), the reconstruction efficiency increased by approximately 4%, while in Run 2 (Experiment 35 data), the gain reached nearly 10%. These results highlight the importance of improved CDC hit cleaning for neutral hadron reconstruction, especially in the outer tracking regions where combinatorial background previously dominated.
Fig. 6.36, taken from the Belle II Tracking and Vertexing Group meeting in July 2025 (presented by D. Pitzl), illustrates these efficiency improvements for late K_S^0 decays. This demonstrates that the new filter directly translates into improved sensitivity for physics channels with displaced vertices, where the CDC plays a decisive role.

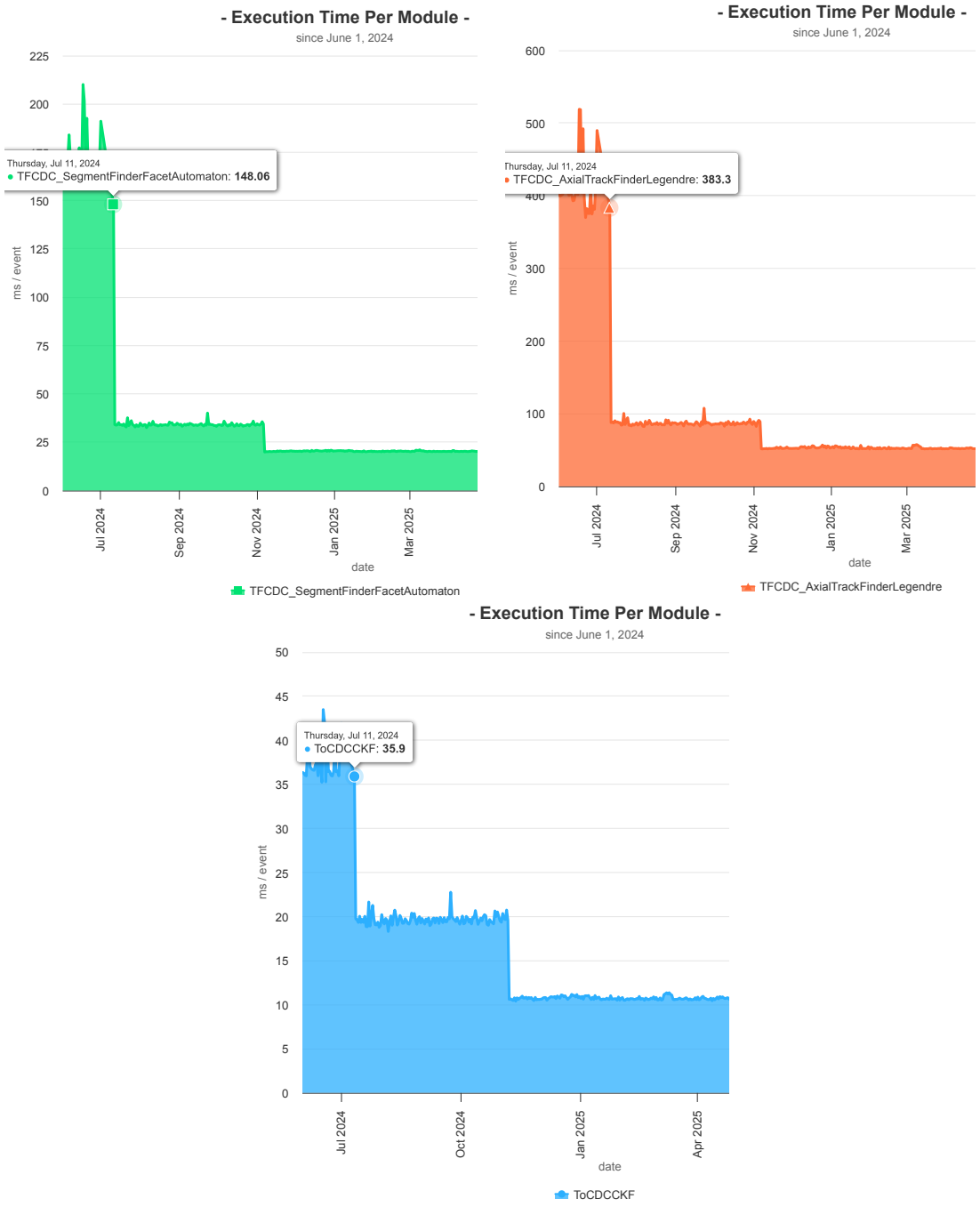

Figure 6.35: Average per-event runtime of key tracking modules before and after the MVA-based CDC background filter was introduced in Release 9. Upper left: `TFCDC_SegmentFinderFacetAutomation`; Upper right: `TFCDC_AxialTrackFinderLegendre`; Bottom: `ToCDCCKF`. All show substantial reductions in execution time.

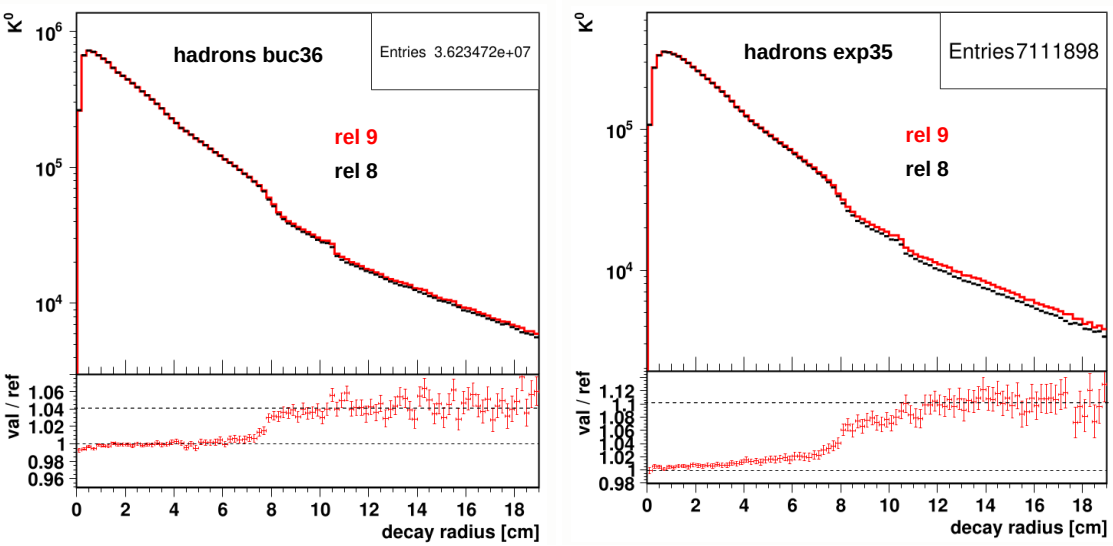

Figure 6.36: Reconstruction efficiency of K_S^0 mesons as a function of decay radius, from a Belle II tracking study (D. Pitzl, July 2025). Left: Run 1 (experiment 26 data), Right: Run 2 (experiment 35 data). Late decays beyond $8\ \mathrm{cm}$ show significant improvements (4% in Run 1, 10% in Run 2) when using the MVA-based CDC background filter.

Chapter 7

Modeling of $\gamma\gamma \to K^0_S K^0_S$ in Belle II Simulation

As discussed on Sec 9.4, a significant data-simulation discrepancy has been observed in regions dominated by low-multiplicity backgrounds, particularly in the forward and backward directions. This excess in data over simulation suggests that certain background processes are not adequately modeled. A likely missing contribution arises from two-photon processes in which one or both final-state electrons remain undetected.

In the Belle II simulation framework, several classes of two-photon processes are included, such as $\gamma\gamma \to \ell^+\ell^-(\ell = e, \mu, \tau)$, $\gamma\gamma \to K^+K^-$, $\pi^+\pi^-$, $p\bar{p}$. However, channels involving resonant hadronic systems, for instance $\gamma\gamma \to K^0_S K^0_S$, $\gamma\gamma \to K^{*0}\bar{K}^{*0}$, or $\gamma\gamma \to K^{*0}K\pi$, are not yet implemented. These omissions are supposed to account for part of the observed mismatch between data and simulation. In order to reduce the discrepancy, corresponding data regions have often been excluded from analyses. Nevertheless, a more systematic treatment of these backgrounds is essential for precision measurements at Belle II.

The study of two-photon processes has a long history in hadronic physics, with numerous measurements reported in the literature [117]. The underlying reaction,

$$e^+e^- \to e^+e^-\,\gamma\gamma \to e^+e^- + X, \tag{7.1}$$

is characterized by the emission of two photons from the incoming beams. Depending on the detection of the scattered electrons, events are classified into *zero-tag*, *single-tag*, and *double-tag* topologies. In the *zero-tag* configuration, both photons are quasi-real ($Q^2 \approx 0$), while *single-* and *double-tag* events provide access to processes with one or two virtual photons, respectively.

Because the production of vector meson pairs ($\gamma\gamma \to VV'$) proceeds through states with the same J^{PC} quantum numbers as the photon, such final states are particularly sensitive to hadronic dynamics in the non-perturbative regime of QCD. Previous studies, including the measurement of $\gamma\gamma \to K^{*0}\bar{K}^{*0}$ [118], have demonstrated the rich phenomenology of these processes. Belle has also reported a comprehensive set of two-photon measurements in both *zero-tag* and *single-tag* configurations [119–121]. In this study, motivated by personal interest and as a challenge task, we explore modeling these missing two-photon processes in the Belle II simulation framework, rather than performing a new measurement.

7.1 Overview of TREPS

The TREPS generator [82] provides a simulation framework for simulating two-photon processes $e^+e^- \to e^+e^-X$, where X represents a hadronic final state. Its central approach relies on the Equivalent Photon Approximation (EPA), which models the incoming electrons as sources of quasi-real photons. The two-photon cross section for the production of a final state X with invariant mass $W = \sqrt{(q_1 + q_2)^2}$ is then given by

$$\sigma(e^+e^- \to e^+e^-X) = \int dW\, d\Omega\, \frac{dL_{\gamma\gamma}}{dW}\, \frac{d\sigma(\gamma\gamma \to X)}{d\Omega}, \tag{7.2}$$

where $d\sigma(\gamma\gamma \to X)/d\Omega$ is the differential cross section for the two-photon subprocess. TREPS implements a partial-wave decomposition of the scattering amplitude:

$$\mathcal{M}(\gamma\gamma \to X) = \sum_{J,\lambda_1,\lambda_2} (2J+1)\, \mathcal{M}^J_{\lambda_1\lambda_2}\, D^J_{\lambda,\mu}(\theta,\phi), \tag{7.3}$$

with J the total angular momentum of the final state, λ_i the photon helicity, $D^J_{\lambda,\mu}(\theta,\phi)$ the Wigner D-functions describing the angular distribution, and (θ,ϕ) the scattering angles in the two-photon frame. The simulation procedure in TREPS consists of four principal steps:

a. **Photon Sampling:** Photon energies ω_1, ω_2 and virtualities Q_1^2, Q_2^2 are sampled according to the EPA.

b. **Phase-Space Generation:** Final-state momenta are generated uniformly in the allowed n-body phase space.

c. **Weight Evaluation:** Each event is assigned a weight

$$w = \frac{dL_{\gamma\gamma}}{dW}\, |\mathcal{M}(\gamma\gamma \to X)|^2 \,\text{Jacobian}, \tag{7.4}$$

which includes the two-photon luminosity, the partial-wave amplitude squared, and the phase-space Jacobian.

d. **Acceptance–Rejection:** Events are un-weighted according to the maximum weight to generate a sample suitable for detector simulation.

From the perspective of a user, the essential inputs for TREPS are the final-state particle masses, the physics model selection (including resonances, partial waves, and form factors), and event counts or cross sections for given energy bins. Differential cross-section tables can also be provided to guide event weighting, although explicit luminosity functions are often optional if the desired number of events per energy bin is specified.

7.2 Implementation for $\gamma\gamma \to K^0_S K^0_S$

Belle has measured the process $\gamma\gamma \to K^0_S K^0_S$ in both *single-tag* and *zero-tag* configurations [119, 120].
For the *single-tag*, since no HEPData entry exists, the digitized event yields from [119] are used, as shown in Fig. 7.1 (left panel). For the *zero-tag* measurement [120], the published differential cross sections, provided by [122] are applied, including statistical and systematic uncertainties, as shown in Fig. 7.1 (right panel) These values are interpolated across the invariant-mass range and supplied to TREPS for event weighting.
The $K^0_S K^0_S$ final state is then generated within TREPS using these measured cross sections as input, and the decay chain is verified at the generator level. A simplified reconstruction is applied uniformly to both the *single-tag* and *zero-tag* samples, without imposing explicit requirements on the final-state electrons or positrons. The K^0_S mesons are reconstructed from charged pion pairs only, requiring tracks within the CDC acceptance and a mass window of $0.45 < M(K^0_S) < 0.55$ GeV/c^2. The resulting distribution of the two-photon invariant mass W, of *single-tag* (left panel) and *zero-tag* (right panel), shown in Fig. 7.2, reproduces the shape of the input cross section, providing a shape-level validation of the event generation. Absolute normalization was not attempted in this study.
This study represents a personal attempt to explore the feasibility of implementing $\gamma\gamma \to K^0_S K^0_S$ within the existing TREPS framework. The exercise demonstrates that the generator can in principle accommodate this channel using published Belle cross sections as input, without requiring a full integration into the Belle II software framework. However, due to the limited understanding of two photon processes and the restricted time available for this work, no systematic validation of the modeling has been performed. The results should be considered as an initial proof of concept. Only shape-level validation was performed, and no systematic study of normalization or resonance modeling was attempted.

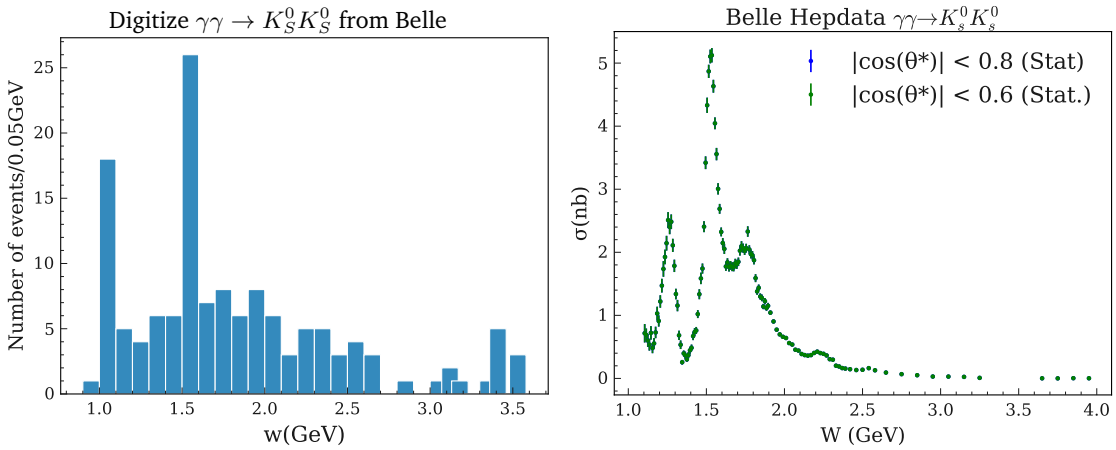

Figure 7.1: Differential cross sections for $\gamma\gamma \to K^0_S K^0_S$ as measured by Belle: *single-tag* (left) and *zero-tag* (right). Data are taken from [119, 122]

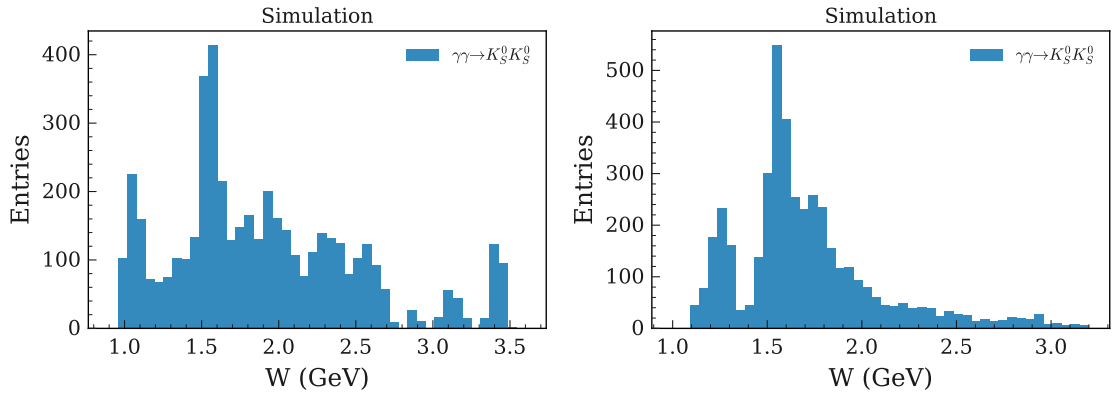

Figure 7.2: Generated and reconstructed W distributions $\gamma\gamma \to K^0_S K^0_S$ in TREPS of Belle II simulation: *single-tag* (left) and *zero-tag* (right). The shapes follow the input cross sections, validating the implementation.

Chapter 8

Modeling the $K\pi\pi$ System in $B \to K\pi\pi\nu\bar{\nu}$

In Belle II, $B \to X_s\nu\bar{\nu}$ signal MC is generated flat in PHSP, and the hadronic system X_s is subsequently fragmented with `Pythia`. Studies in the K^{*0} signal window, as mentioned in , indicate that the dominant fragmentation background arises from

$$X_s \to K^{*0}\pi \quad \text{with} \quad K^{*0} \to K\pi, \tag{8.1}$$

i.e. an effective $K\pi\pi$ final state. A realistic model of the $K\pi\pi$ system is therefore important both for background estimation and for reweighting of PHSP samples.
LHCb has shown that the $K\pi\pi$ distributions in non-resonant $B \to K\pi\pi\mu^+\mu^-$ (excluding the charmonium regions) are qualitatively similar to those in $B \to J/\psi K\pi\pi$ [112]. From this, we assume that the hadronic amplitude of the $K\pi\pi$ system is largely universal and can be taken from channels with high-statistics measurements such as $B \to J/\psi K\pi\pi$, while short-distance current effects ($\ell^+\ell^-$ or $\nu\bar{\nu}$) are negligible for the $K\pi\pi$ substructure.
Belle has performed a partial-wave analysis (PWA) of the $K\pi\pi$ system in $B \to J/\psi K\pi\pi$ [113]. The analysis shows that the dynamics are dominated by a rich set of kaonic resonances, including

$$K_1(1270), K_1(1400), K^*(1410), K_2^*(1430), K^*(1600), K^*(1680), K_2^*(1270), K_2^*(1710), \text{and } K_2^*(1980), \tag{8.2}$$

with significant interferences among them. These resonances contribute through different orbital angular momentum components, and their coherent superposition determines the observed $K\pi\pi$ distributions. The fitted complex amplitudes and line-shapes obtained by Belle therefore provide a realistic and data-driven input model for describing $X_s \to K\pi\pi$ in $B \to X_s\nu\bar{\nu}$.
In this framework the decay is described as a cascade $B \to J/\psi + R_1$, followed by $R_1 \to R_2 + (\text{bachelor})$, where R_1 denotes the quasi-two-body kaonic resonance that carries the entire $K\pi\pi$ system (e.g. $K_1(1270)$), and R_2 denotes the isobar inside this system (e.g. K^*, ρ, or ω). A key point in this construction is the consistent treatment of invariant masses: the propagator of R_1 always depends on the full hadronic mass $m_{abc} = m(K\pi\pi)$, while the propagator of R_2 depends on the channel-specific two-body mass, $m(K\pi)$ for $K^*\pi$ submodes and $m(\pi\pi)$ for ρK or ωK submodes. The remaining hadron, referred to as the bachelor, fixes which combination of particles is treated as ab and which as bc, and therefore determines the inputs to the kinematic functions $p, q, \cos\theta$. Correct bookkeeping of this particle order is essential for reproducing the Belle partial-wave amplitudes.
To construct the amplitude for a given PHSP event, one starts from the four-momenta of the three charged hadrons and computes the relevant kinematic variables: the three-body invariant mass m_{abc}, the appropriate two-body invariants m_{ab} and m_{bc}, the breakup momenta p and q, the variable $z = p/m_{abc}$, and the helicity angle $\cos\theta$. The resonance parameters (M_R, Γ_R) are fixed to the values given in Tab. [113], while the complex coefficients for each submode, $c_i = |c_i|e^{i\phi_i}$, are taken from the fitted moduli and phases reported in Table V.
The signal model itself is expressed as the coherent sum of all included submodes, plus a possible nonresonant component,

$$A(\vec{x}) = c_{\rm nr}A_{\rm nr} + \sum_i c_i\, A_i(\vec{x}), \qquad w(\vec{x}) = |A(\vec{x})|^2, \tag{8.3}$$

where each submode amplitude A_i follows the form given in Eq. (28)-(36) in [113]: a product of Breit-Wigner terms for R_1 and R_2, a mass-dependent width for R_2, and the spin-angular factor $\alpha_i(\cos\theta, z)$. The complex coefficients c_i control the relative strength and phase of the interfering amplitudes, ensuring that the reweighted distribution reproduces the measured fractions and interference patterns.
In practice, the re-weighting can be carried out within `EvtGen` in two equivalent ways. One approach is to compute $A(\vec{x})$ for each event and pass it directly through the vertex function, so that the generator samples with probability proportional to $|A|^2$. Alternatively, one may compute the weight $w(\vec{x})$ for each flat PHSP event and use it either for accept-reject filtering or as an analysis weight. Both methods allow large PHSP samples to be transformed into samples with realistic $K\pi\pi$ substructure consistent with the Belle partial-wave analysis.

8.1 Custom Implementation in `EvtGen`

The reweighting strategy implemented directly inside `EvtGen` is adopted. Large flat PHSP samples are generated and, for each event, `EvtGen` evaluates the chosen amplitude model and re-weights the event by $|A(\vec{x})|^2$. A single PHSP dataset can thus be transformed into a distribution with realistic $K\pi\pi$ substructure without regenerating separate samples, while retaining the flexibility to swap or tune models. Importantly, in the custom amplitude model(hereafter referred to as the DIY model), the amplitudes are not generated as a single coherent model; instead, they are produced separately for each physical J^P sector (1^+, 1^-, 2^+, 2^-), thereby preserving the internal interference structure within each sector but not between different spin-parity states.
Fig. 8.1 compares the PHSP sample (orange, 5000 events) with the resonance contributions (colored steps) projected onto $M(\pi^+\pi^-)$, $M(K^+\pi^-)$, and $M(K\pi\pi)$ and the corresponding squared invariant mass $M^2(\pi^+\pi^-)$, $M^2(K^+\pi^-)$, and $M^2(K\pi\pi)$. Here the gray curve is the sum of all resonant components, and it is dominated by the 1^+ sector ($K_1(1270)$ and $K_1(1400)$). The $K_1(1270)$ peak is evident in $M(K\pi\pi)$, and the corresponding $K^*(892)$ band shows up clearly in $M(K\pi)$ through $K_1 \to K^*\pi$. However, within this DIY model the two $1^\pm$ states are not well distinguished: their contributions largely overlap in the $K\pi\pi$ mass window, making it difficult to separate $K_1(1270)$ from $K_1(1400)$.
For reference, the resonance content used in the DIY model is grouped as:

- 1^+: $K_1(1270)$, $K_1(1400)$
- 1^-: $K^*(1410)$
- 2^+: $K_2^*(1430)$, $K_2^*(1980)$
- 2^-: $K(1600)$, $K_2(1770)$

In contrast to the Belle Dalitz projections (e.g. Fig. 18 of Ref. [113]), our $M(\pi\pi)$ projection does not reproduce the expected $\rho(770)$ behavior. This indicates potential issues still to be resolved. This discrepancy may arise from several factors: the use of fit fractions without full interference, mismodeling of resonance lineshapes or barrier factors, incomplete or mis-configured K_1 decay submodes (including the $K_1(1270)$–$K_1(1400)$ mixing), or technical differences between our `EvtGen` setup and the Belle model.
We applied the same DIY configuration to the invisible mode, replacing the charged leptons by a di-neutrino system: $B^\pm \to K^\pm\pi^\pm\pi^\mp\nu\bar{\nu}$, with the same resonance content and per-J^P sectors as in the visible channel (1^+, 1^-, 2^+, 2^-; the gray curve in Fig. 8.1 denotes the sum of all resonances and is dominated by the 1^+ sector), while the neutrino pair is treated as a massless two-body system without additional dynamical constraints. The resulting $M(K\pi\pi)$ spectrum (Fig. 8.1) shows a low-mass region consistent with the expected K_1 structures ($\sim$ 1.2–1.6 GeV/c^2), but then rises and remains large up to 3 GeV/c^2 before falling off toward the kinematic endpoint near m_B, exhibiting an extended PHSP-like shoulder for $M(K\pi\pi) \in$ [1.6, 5] GeV/c^2. This behavior is in fact generic if the $b \to s\nu\bar{\nu}$ dynamics are omitted: with two massless neutrinos, the available hadronic mass is limited only by energy conservation, $M_{K\pi\pi} \leq m_B$, and the flat sampling of the invisible system populates large $q^2 = M^2_{\nu\bar{\nu}}$ values, pushing $M_{K\pi\pi}$ upward. In other words, we are enforcing

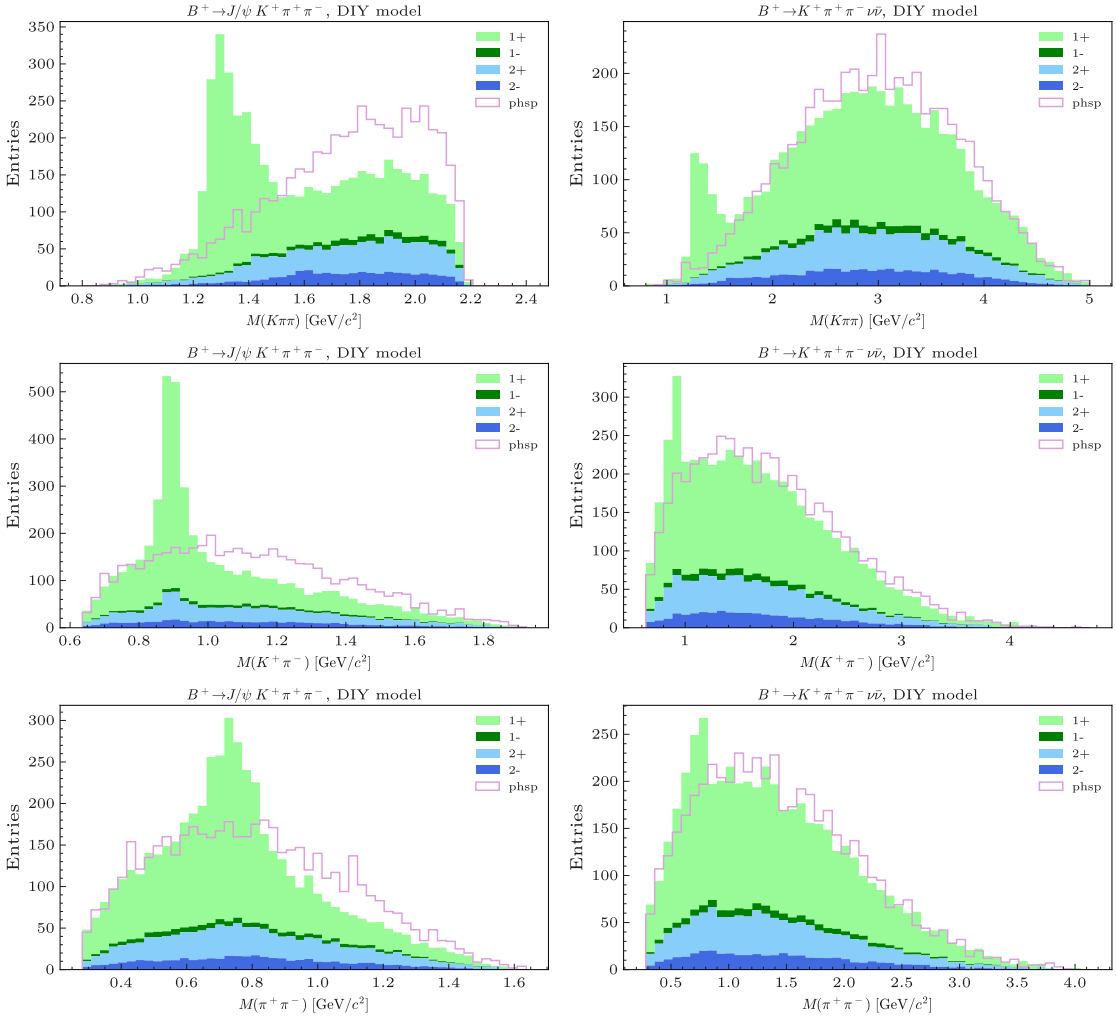

Figure 8.1: Dalitz projections of $B \to J/\psi K\pi\pi$ (left) and $B \to K\pi\pi\nu\bar{\nu}$ (right) using the DIY `EvtGen` model. Stacked histograms show the contributions of individual resonances by spin–parity sector, and the pink line shows the PHSP events.

realistic $K\pi\pi$ substructure but sampling the $(\nu, \bar{\nu})$ system as phase space; the joint density then factorizes approximately as

$$\mathrm{d}\Gamma \;\propto\; |A_{K\pi\pi}(\vec{x})|^2 \; \Phi_3(K\pi\pi) \; \Phi_2(\nu\bar{\nu}) \,, \tag{8.4}$$

which favors high $M(K\pi\pi)$ once integrated over the invisible degrees of freedom.
In the invisible channel, there is no established SM form-factor prediction for $B \to K\pi\pi\nu\bar{\nu}$ as a function of q^2 and the $K\pi\pi$ substructure. We therefore attempted a pragmatic fix by applying an additional suppression weight in the high hadronic-mass region ($M(K\pi\pi) > 2.5\,\mathrm{GeV}/c^2$) within the DIY `EvtGen` setup. This, however, did not alter the spectrum: the generated shape remained dominated by a broad PHSP-like shoulder extending up to the kinematic endpoint. Since this indicates that our extra weight was not effectively propagated through internal reweighting of `EvtGen` (or was absorbed by a subsequent normalization), we conclude that we do not fully control how `EvtGen` handles the invisible kinematics in this configuration. To overcome this limitation, we instead generate large flat PHSP samples and apply the desired amplitude weights externally on an event-by-event basis, thereby avoiding the internal ambiguities of `EvtGen`. This behavior is generic when the $b \to s\nu\bar{\nu}$ dynamics are omitted: the flat sampling of the invisible system populates large q^2 values, which in turn enhances high $M(K\pi\pi)$. Without incorporating SM formfactor distributions, the invisible system artificially enhances the spectrum in this region. Our pragmatic weighting fix did not resolve this within `EvtGen`, hence we moved to an external reweighting approach, described in the next subsection.

8.2 Amplitude Reweighting Method

To overcome the limitations of the DIY `EvtGen` model, we adopted an external reweighting strategy. For each generated PHSP event, we store the $(E, \vec{p})$ of all final-state particles and then evaluate dedicated amplitude functions corresponding to each decay mode. The sub-amplitudes are then combined into four spin-parity sectors (1^+, 1^-, 2^+, 2^-). For each sector, the squared amplitude $|A|^2$ is computed, normalized to the number of generated events, and finally mixed according to the Belle fit fractions. Squared invariant masses such as $M^2(K\pi\pi)$, $M^2(K\pi)$, and $M^2((\pi\pi)$ are also calculated event by event.
Applying the external reweighting to $B \to J/\psi K\pi\pi$, the $M(K\pi\pi)$ spectrum shows clear resonance structures. The dominant 1^+ contribution from $K_1(1270)$ and $K_1(1400)$ is visible, and the two states are now somewhat distinguishable. The 2^+ and 2^- components appear in their expected mass regions, while the 1^- contribution is also present. In the $\pi^+\pi^-$ spectrum, the low-mass contamination is suppressed, and although the ρ appears broader than in Belle, the peak is visible, with the $2^\pm$ sectors contributing in the proper mass region, as shown in Fig. 8.2.
For $B \to K\pi\pi\nu\bar{\nu}$, the reweighted spectrum behaves much better than in the internal `EvtGen` attempt. The $M(K\pi\pi)$ distribution is dominated by the 1^+ sector, with both $K_1(1270)$ and $K_1(1400)$ peaks visible, but the non-physical high-mass shoulder above $2.5\,\mathrm{GeV}$ is strongly suppressed. In the $M(K\pi)$ projection, the 1^+ resonance dominates, with additional visible contributions from 1^- and 2^+. In the $\pi^+\pi^-$ spectrum, a broader ρ peak is seen, while the 2^- sector contributes around $1.3\,\mathrm{GeV}/c^2$, consistent with $K_2(1770) \to K f_2(1270)$ contribution, as shown in Fig. 8.2.
This demonstrates that external reweighting restores the expected resonance patterns while avoiding non-physical artifacts, and provides a transparent, flexible framework for further studies.
Conceptually, both approaches start from the same PHSP distribution and apply amplitudes sector by sector. In the external reweighting method, PHSP events are generated, the (p, E) variables are extracted externally, amplitudes for each decay mode are then evaluated, and the results are normalized per J^P. In the DIY `EvtGen` model, PHSP events are generated, the (p, E) variables are accessed internally within `EvtGen`, amplitudes are applied via the `vertex()` interface, and the results are normalized per J^P. In principle the two strategies should yield the same PHSP basis. Any differences are likely due to how `EvtGen` implements the amplitude evaluation inside `vertex()`, including phase conventions, normalization factors, or treatment of overlapping resonances.
In this study, we developed and compared two approaches for modeling the $B \to K\pi\pi$ system: a DIY implementation inside `EvtGen` and an external reweighting method based on PHSP events. Both strategies

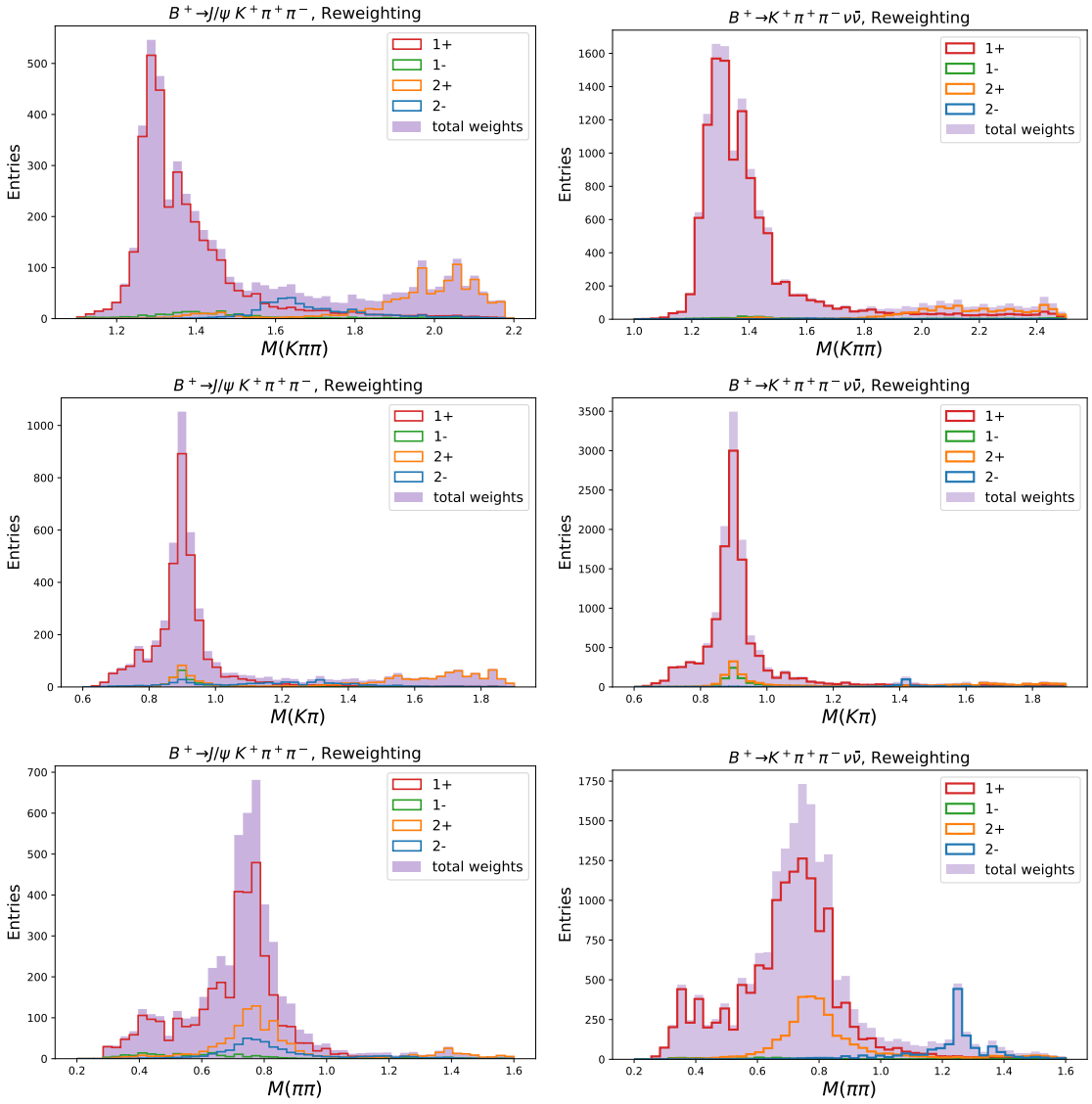

Figure 8.2: Dalitz projections of $B \to J/\psi K\pi\pi$ (left) and $B \to K\pi\pi\nu\bar{\nu}$ (right) after external reweighting. Step histograms show the separate resonance contributions; the purple-filled histogram shows their sum. Resonance structures from $K_1(1270)$ and $K_1(1400)$ are visible, and the non-physical high-mass shoulder above 2.5 GeV/c^2 is strongly suppressed.

reproduce the dominant resonance features, with the external reweighting providing greater transparency and control over the amplitude composition. Only shape validation was attempted; absolute normalization and full systematic uncertainties were not studied.

It is important to emphasize that the $B \to K^{*0}\nu\bar{\nu}$ analysis did not employ this reweighting procedure, largely due to its complexity and the absence of a direct Standard Model prediction for the full hadronic system. Instead, the present work should be regarded as an exploratory investigation into how generator configurations and amplitude modeling affect the simulated $K\pi\pi$ distributions. In principle, such an investigation inevitably involves assumptions about the hadronic dynamics, as well as our interpretation of amplitude analyses, particularly in the absence of detailed SM predictions or direct experimental measurements.

We stress that the results for $B \to K\pi\pi\nu\bar{\nu}$ reweighting must be treated with caution. The J/ψ system and the di−neutrino system are kinematically and dynamically distinct, and the lack of a rigorous theoretical framework introduces ambiguities in the modeling. As such, we do not place full confidence in the reweighted $K\pi\pi\nu\bar{\nu}$ spectra. Nevertheless, this exercise has been valuable as a learning process: it highlights the interplay between generator assumptions, amplitude modeling, and physics interpretation, and underscores the importance of carefully questioning the procedures we use when no detailed SM shape predictions or direct measurements are available.

Summary And Outlook

The thesis documents searches for $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^*(892)^0\nu\bar{\nu}$ with an inclusive tagging method in the Belle II experiment. The former has been published [1], the latter is under pre-unblinding checks.
Part I introduces the SM prediction for $B^0 \to K^{(*)}\nu\bar{\nu}$, at loop or box level of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (5.22 \pm 0.15 \pm 0.28) \times 10^{-6}$ and $\mathcal{B}(B^0 \to K^*(892)^0\nu\bar{\nu}) = (9.47 \pm 1.28 \pm 0.57) \times 10^{-6}$ [38]. The SM prediction for $B \to K^{(*)}\nu\bar{\nu}$ is relatively clean compared to the prediction for $B \to K^{(*)}\ell\ell$, which suffers from large theoretical uncertainties due to the breakdown of factorization due to photon exchanges [36]. The decay rate of $B \to K^{(*)}\nu\bar{\nu}$ can be enhanced through tree-level FCNCs processes involving new particles, such as $Z^{'}$ boson [40], or invisible particles, such as a dark photon [51], or sterile neutrinos [56]. The B factories, such as Babar and Belle, have performed searches for $B \to K^{(*)}\nu\bar{\nu}$ using hadronic tagging or semileptonic tagging methods with upper limits on the branching fractions at the 90% confidence level, where semileptonic tagging yields higher sensitivity than hadronic tagging [58–61]. Using the inclusive tagging method, Belle II [57] reaches the same precision for $B^+ \to K^+\nu\bar{\nu}$ as the hadronic tagging analysis performed by Belle [58], where Belle II uses the 63 fb^{-1} luminosity of data, which is less than 10% of the Belle luminosity. The inclusive method shows potential sensitivity to measure $B \to K^{(*)}\nu\bar{\nu}$ decays. It leads to the two searches, $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^*(892)^0\nu\bar{\nu}$, with an inclusive tagging method, documented in Part II.
The first chapter of Part II presents the search for $B^+ \to K^+\nu\bar{\nu}$ with inclusive tagging in a 362 fb^{-1} sample of electron-positron collisions at the $\Upsilon(4S)$ resonance collected with the Belle II detector at the SuperKEKB collider. The inclusive tagging analysis method exploits the global properties of the $B\bar{B}$ system produced together with the signal B meson. Relevant quantities are computed using the kaon candidate and the remaining particles in the event to discriminate between signal and background. Two successive classifiers are applied to distinguish the signal from the background. The first classifier retains 34% signal events while rejecting 98.5% background with the classifier score larger than 0.9. The two classifiers achieve a maximum expected significance of 1.4, with the first classifier alone reaching 0.5 (see Sec. 4.2.2). The second classifier is mapped to the efficiency quantile to quantify the fraction of retained signal events, $\eta(\mathrm{BDT}_2)$. To maximize the expected significance, the signal region is defined to retain 8% of the signal efficiency. The signal selection efficiency is validated with the kinematic-modified $B^+ \to K^+J/\psi(\to \ell\ell)$ sample. The simulation of $e^+e^- \to q\bar{q}$ and $\tau^+\tau^-$ processes, affected by generator-tuning differences, is corrected using offer-resonance data with a classifier, which is 60 MeV below the $\Upsilon(4S)$ resonance. The backgrounds from $B\bar{B}$ decays, which contribute 60% in the signal region and suffer modeling issues (such as D-meson decays involving a K^+/K^0_L meson, $B^+ \to K^+K^0\bar{K}^0$, $B^+ \to K^+n\bar{n}$, etc.), are carefully modeled. The background modeling is further validated using the events that are outside the signal region. Finally, after a series of unblinding checks, the signal yields are extracted via a binned maximum likelihood fit simultaneously to on- and off-resonance data in the two-dimensional signal region defined by $q^2_{\mathrm{rec}} \times \eta(\mathrm{BDT}_2)$. A product of the Poisson probability-density functions is used as the likelihood function, with the systematic uncertainties included as the nuisance parameters. The signal strength μ is determined to be $\mu = 5.4 \pm 1.0(\mathrm{stat}) \pm 1.1(\mathrm{syst})$, corresponding to a branching fraction to be $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (2.7 \pm 0.5(\mathrm{stat}) \pm 0.5(\mathrm{syst}))$. By evaluating the profile likelihood for μ, the significance is found to be 2.9 standard deviations from the SM expectation.
The second chapter of Part II presents a search for $B^0 \to K^*(892)^0\nu\bar{\nu}$ using the same dataset and the inclusive tagging method. The $K^*(892)^0$ is reconstructed with the charged decay mode, $K^*(892)^0 \to K^+\pi^-$. A classifier is applied to distinguish the signal from the combinatorial backgrounds arising from the high track multiplicity in the $K^*(892)^0$ vertex reconstruction. Two candidates with the highest classifier scores are retained per event, which are then passed through two successive classifiers. The first classifier is trained

with the same variables as those in the $B^+ \to K^+\nu\bar{\nu}$ channel. The second classifier exploits more K^*-vertex-related variables to distinguish the signal from various backgrounds. The first classifier alone yields a maximum expected significance of 0.3, while the combined two-classifier selection reaches 0.95. Once the classifier is mapped to the signal efficiency quantile, the signal region is then defined to retain 5% signal efficiency, chosen as a balance between statistical precision and background suppression. The signal efficiency is validated with a kinematic-modified $B^0 \to K^*(892)^0 J/\psi(\to \ell\ell)$ sample. The simulation of $e^+e^- \to q\bar{q}, \tau^+\tau^-$ is corrected using off-resonance data. Among the $B\bar{B}$ backgrounds, which is about 63% in the signal region, events where the two tracks originated from different B mesons contribute about 30%, and this is validated with a control sample with $K\pi$ invariant mass in the range of [0.6, 0.8] GeV/c^2 and [1.0, 1.4] GeV/c^2. The remaining 70% of the background originates from the same B-meson. It contains three main components: genuine $K^*(892)^0$ candidates with unreconstructed neutral particles, $K\pi$ pairs where both tracks originate from a charm meson, and combinations in which one track originates from a charm meson and the other from a lepton. These backgrounds are carefully controlled. The crossfeed background from other $B \to K^{(*)}\nu\bar{\nu}$ decays is either modeled based on SM predictions [38] (such as $B \to K^+/K_S^0/K^*(892)\nu\bar{\nu}$) or studied with a mixed PHSP model (such as $B \to X_s\nu\bar{\nu}$). The above background models are further validated using events outside the signal region, ($\eta(\mathrm{BDT}_2)$<0.95). The various checks and validations are ongoing; therefore, no unblinding results are presented in this thesis.

Statistically, increasing the dataset to 1 ab^{-1} and 5 ab^{-1} would improve the statistical precision of $B^+ \to K^+\nu\bar{\nu}$ to approximately 60% and 27% of the precision obtained in this thesis, respectively, assuming identical selection efficiency and background composition. In practice, increased luminosity will enable better detector calibration, larger simulation samples, and improved background modeling. Consequently, data-driven related systematics are expected to be reduced. For instance, improved modeling can increase the signal efficiency, while better separation of hits in the CDC can enhance the tracking performance, leading to higher signal efficiency at the analysis level as well.

The modeling $\gamma\gamma \to K_S^0 K_S^0$ and $B \to K\pi\pi\nu\bar{\nu}$, which serve as auxiliary studies, as well as the CDC background filter developed to improve tracking performance, are documented in Part III. The analogue simulation modeling serves as a complementary study in the absence of corresponding Belle II measurements; a one-time measurement effort would provide long-term benefits. For the CDC background study, the robust FastBDT shows improved tracking efficiency, while the graph neural network for multi-track reconstruction [123] is also appears promising. The current era of machine and deep learning may drive the upgrades to the underlying algorithms in basf2, making it more straightforward to integrate with external AI-based tools and data-driven workflows.

Table 8.2 is the main documentation of each chapter.

Structure	Summary
Part I	Chapter 1: SM prediction and formfactor of $B \to K^{(*)}\nu\bar{\nu}$; Chapter 2: Belle II detector and software framework; Chapter 3: Inclusive tagging method, classifier framework, signal estimation techniques, and topology technique.
Part II	Chapter 4/5: Analysis of $B^+ \to K^+\nu\bar{\nu}/B^0 \to K^*(892)^0\nu\bar{\nu}$
	Section 1: Data and simulation sample, signal reconstruction; Section 2: Two classifiers to suppress background; Section 3: Signal region defined with the classifier output and signal efficiency validation with $B \to K^{(*)}J/\psi$; Section 4: $e^+e^- \to q\bar{q}, \tau^+\tau^-$ and $\Upsilon(4S) \to B\bar{B}$ background modeling and validation; Section 5: Signal extraction setup; Section 6: Systematic studied; Section 7: $B^+ \to K^+\nu\bar{\nu}$ results a 2.9 σ
Part III	Chapter 5: A hit-level MVA-based CDC background filter to improve the tracking performance. Chapter 6: An auxiliary modeling of $\gamma\gamma \to K^0_S K^0_S$ in Belle II simulation. Chapter 7: An auxiliary modeling of $B \to K\pi\pi\nu\bar{\nu}$ in Belle II simulation.

Table 8.1: A summary of each chapter in this thesis.

Zusammenfassung und Ausblick

Diese Dissertation dokumentiert die Suchen nach den Zerfällen $B^+ \to K^+\nu\bar{\nu}$ und $B^0 \to K^*(892)^0\nu\bar{\nu}$ unter Verwendung einer inklusiven Tagging-Methode im Belle-II-Experiment. Die erstgenannte Analyse wurde bereits veröffentlicht [1], während sich die letztere noch in der Phase der Vor-Entblindungs-Prüfungen befindet.

Teil I führt in die Standardmodell-(SM)-Vorhersage für $B^0 \to K^{(*)}\nu\bar{\nu}$ ein, die auf Schleifen- oder Boxdiagrammen basiert, mit den Zerfallswahrscheinlichkeiten $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (5.22 \pm 0.15 \pm 0.28) \times 10^{-6}$ und $\mathcal{B}(B^0 \to K^*(892)^0\nu\bar{\nu}) = (9.47 \pm 1.28 \pm 0.57) \times 10^{-6}$ [38]. Die SM-Vorhersage für $B \to K^{(*)}\nu\bar{\nu}$ ist vergleichsweise sauber im Vergleich zur Vorhersage für $B \to K^{(*)}\ell\ell$, die aufgrund des Zusammenbruchs der Faktorisierung infolge von Photonenaustauschprozessen großen theoretischen Unsicherheiten unterliegt [36]. Die Zerfallsrate von $B \to K^{(*)}\nu\bar{\nu}$ kann durch flavour-ändernde neutrale Ströme (FCNC) auf Baumebene verstärkt werden, die neue Teilchen wie ein Z'-Boson [40], unsichtbare Teilchen wie ein Dunkelphoton [51] oder sterile Neutrinos [56] einbeziehen. Die B-Fabriken BaBar und Belle haben Suchen nach $B \to K^{(*)}\nu\bar{\nu}$ unter Verwendung hadronischer oder semileptonischer Tagging-Methoden durchgeführt und obere Grenzen für die Verzweigungsverhältnisse auf dem 90%-Konfidenzniveau bestimmt, wobei das semileptonische Tagging eine höhere Empfindlichkeit zeigte als das hadronische [58–61]. Mit der inklusiven Tagging-Methode erreicht Belle II [57] die gleiche Präzision für $B^+ \to K^+\nu\bar{\nu}$ wie die hadronische Analyse von Belle [58], obwohl Belle II nur eine integrierte Luminosität von 63 fb^{-1} nutzt – weniger als 10 % derjenigen von Belle. Die inklusive Methode zeigt somit ein erhebliches Potenzial zur Messung der $B \to K^{(*)}\nu\bar{\nu}$-Zerfälle. Sie bildet die Grundlage für die beiden in Teil II dokumentierten Analysen zu $B^+ \to K^+\nu\bar{\nu}$ und $B^0 \to K^*(892)^0\nu\bar{\nu}$.

Das erste Kapitel von Teil II beschreibt die Suche nach $B^+ \to K^+\nu\bar{\nu}$ unter Verwendung der inklusiven Tagging-Methode in einer Stichprobe von 362 fb^{-1} Elektron-Positron-Kollisionen bei der $\Upsilon(4S)$-Resonanz, die mit dem Belle-II-Detektor am SuperKEKB-Collider aufgezeichnet wurden. Die inklusive Tagging-Analyse nutzt die globalen Eigenschaften des erzeugten $B\bar{B}$-Systems, das gemeinsam mit dem Signal-B-Meson produziert wird. Relevante Größen werden unter Verwendung des Kaon-Kandidaten und der verbleibenden Teilchen im Ereignis berechnet, um zwischen Signal und Untergrund zu unterscheiden. Zwei aufeinanderfolgende Klassifikatoren werden eingesetzt, um das Signal vom Untergrund zu trennen. Der erste Klassifikator behält 34 % der Signalereignisse bei, während 98,5 % des Untergrundes für Klassifikator-Scores größer als 0,9 unterdrückt werden. Beide Klassifikatoren zusammen erreichen eine maximale erwartete Signifikanz von 1,4, wobei der erste Klassifikator allein eine Signifikanz von 0,5 erzielt (siehe Abschnitt 4.2.2). Der zweite Klassifikator wird auf das Effizienz-Quantil $\eta(\mathrm{BDT}_2)$ abgebildet, um den Anteil der beibehaltenen Signalereignisse zu quantifizieren. Zur Maximierung der erwarteten Signifikanz wird die Signalregion so definiert, dass 8 % der Signaleffizienz beibehalten werden. Die Signalselektionseffizienz wird mit der kinematisch modifizierten $B^+ \to K^+J/\psi(\to \ell\ell)$-Stichprobe validiert. Die Simulation der Prozesse $e^+e^- \to q\bar{q}$ und $\tau^+\tau^-$, die durch Unterschiede in der Generatorabstimmung beeinflusst sind, wird mithilfe von Off-Resonanz-Daten korrigiert, die 60 MeV unterhalb der $\Upsilon(4S)$-Resonanz aufgenommen wurden. Die Untergründe aus $B\bar{B}$-Zerfällen, die etwa 60 % in der Signalregion ausmachen und Modellierungsunsicherheiten aufweisen (z. B. D-Meson-Zerfälle mit einem K^+/K^0_L-Meson, $B^+ \to K^+K^0\bar{K}^0$, $B^+ \to K^+n\bar{n}$ usw.), werden sorgfältig modelliert. Die Untergrundmodellierung wird zusätzlich anhand von Ereignissen außerhalb der Signalregion überprüft. Nach einer Reihe von Entblindungs-Vorprüfungen werden die Signalausbeuten mittels einer binned Maximum-Likelihood-Anpassung gleichzeitig an On- und Off-Resonanz-Daten in der zweidimensionalen Signalregion ($q^2_{\mathrm{rec}} \times \eta(\mathrm{BDT}_2)$) extrahiert. Als Likelihood-Funktion wird das Produkt von Poisson-Wahrscheinlichkeitsdichten verwendet, wobei systematische Unsicherheiten als Nuisance-Parameter

berücksichtigt werden. Die Signalstärke wird zu $\mu = 5.4 \pm 1.0\,(\mathrm{stat}) \pm 1.1\,(\mathrm{syst})$ bestimmt, was einer Verzweigungswahrscheinlichkeit von $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (2.7 \pm 0.5\,(\mathrm{stat}) \pm 0.5\,(\mathrm{syst}))$ entspricht. Durch Auswertung der Profil-Likelihood für μ ergibt sich eine Signifikanz von 2.9 Standardabweichungen gegenüber der SM-Erwartung.
Das zweite Kapitel von Teil II beschreibt die Suche nach $B^0 \to K^*(892)^0\nu\bar{\nu}$ unter Verwendung derselben Datenmenge und der inklusiven Tagging-Methode. Das $K^*(892)^0$ wird im geladenen Zerfallskanal $K^*(892)^0 \to K^+\pi^-$ rekonstruiert. Ein Klassifikator wird angewandt, um das Signal von kombinatorischen Untergründen zu unterscheiden, die aus der hohen Spurenmultiplizität bei der Rekonstruktion des $K^*(892)^0$-vertex stammen. Pro Ereignis werden zwei Kandidaten mit den höchsten Klassifikator-Scores beibehalten, die anschließend durch zwei aufeinanderfolgende Klassifikatoren weiterverarbeitet werden. Der erste Klassifikator wird mit denselben Variablen trainiert wie im $B^+ \to K^+\nu\bar{\nu}$-Kanal. Der zweite Klassifikator nutzt zusätzliche, K^*-Vertex-bezogene Variablen, um das Signal von verschiedenen Untergründen zu trennen.
Der erste Klassifikator allein liefert eine maximale erwartete Signifikanz von 0,3, während die kombinierte Zwei-Klassifikator-Selektion 0,95 erreicht. Nach der Abbildung des Klassifikatorausgangs auf das Signaleffizienzquantil wird die Signalregion so definiert, dass 5 % Signaleffizienz beibehalten werden – ein Kompromiss zwischen statistischer Präzision und Unterdrückung des Untergrundes. Die Signaleffizienz wird mithilfe einer kinematisch modifizierten $B^0 \to K^*(892)^0 J/\psi(\to \ell\ell)$-Stichprobe validiert. Die Simulation der Prozesse $e^+e^- \to q\bar{q}$ und $\tau^+\tau^-$ wird mit Off-Resonanz-Daten korrigiert. Unter den $B\bar{B}$-Untergründen, die etwa 63 % der Ereignisse in der Signalregion ausmachen, stammen rund 30 % aus Ereignissen, bei denen die beiden Spuren von unterschiedlichen B-Mesonen herrühren. Dies wird mit einer Kontrollstichprobe validiert, deren $K\pi$-invariante Masse im Bereich von [0.6, 0.8] GeV/c^2 und [1.0, 1.4] GeV/c^2 liegt. Die verbleibenden 70 % des Untergrundes stammen aus demselben B-Meson. Dieser Anteil besteht aus drei Hauptkomponenten: (i) echten $K^*(892)^0$-Kandidaten mit nicht rekonstruierten neutralen Teilchen, (ii) $K\pi$-Paaren, bei denen beide Spuren aus einem Charm-Meson stammen, und (iii) Kombinationen, in denen eine Spur von einem Charm-Meson und die andere von einem Lepton stammt. Diese Untergründe werden sorgfältig modelliert und kontrolliert. Der Crossfeed-Untergrund aus anderen $B \to K^{(*)}\nu\bar{\nu}$-Zerfällen wird entweder auf Grundlage der SM-Vorhersagen [38] modelliert (z. B. $B \to K^+/K^0_S/K^*(892)\nu\bar{\nu}$) oder mit einem gemischten PHSP-Modell untersucht (z. B. $B \to X_s\nu\bar{\nu}$). Die oben beschriebenen Untergrundmodelle werden anhand von Ereignissen außerhalb der Signalregion ($\eta(\mathrm{BDT}_2) < 0.95$) überprüft. Die verschiedenen Überprüfungen und Validierungen sind noch im Gange; daher werden in dieser Dissertation keine Entblindungsergebnisse präsentiert.
Statistisch gesehen würde eine Erhöhung des Datensatzes auf 1 ab^{-1} bzw. 5 ab^{-1} die statistische Präzision der Messung von $B^+ \to K^+\nu\bar{\nu}$ auf etwa 60 % bzw. 27 % der in dieser Arbeit erreichten Präzision verbessern, vorausgesetzt, die Selektionseffizienz und die Untergrundzusammensetzung bleiben unverändert. In der Praxis wird eine erhöhte Luminosität jedoch eine verbesserte Detektorkalibrierung, größere Simulationsstichproben und eine präzisere Untergrundmodellierung ermöglichen. Daraus folgt, dass datengetriebene systematische Unsicherheiten voraussichtlich reduziert werden. Beispielsweise kann eine verbesserte Modellierung die Signaleffizienz steigern, während eine bessere Trennung der Treffer in der CDC („Central Drift Chamber“) die Spurrekonstruktion verbessert und somit die Signaleffizienz auch auf Analyseebene erhöht. Die Modellierungen von $\gamma\gamma \to K^0_S K^0_S$ und $B \to K\pi\pi\nu\bar{\nu}$, die als Hilfsstudien dienen, sowie der zur Verbesserung der Spurrekonstruktion entwickelte CDC-Untergrundfilter werden in Teil III dokumentiert. Die analoge Simulationsmodellierung dient als ergänzende Studie in Ermangelung entsprechender Belle-II-Messungen; eine einmalige Messung in diesem Bereich würde langfristige Vorteile bieten.
Für die CDC-Untergrundstudie zeigt der robuste *FastBDT* eine verbesserte Spurrekonstruktionseffizienz, während das auf Graph-Neural-Networks basierende Verfahren zur Mehrspurrekonstruktion [123] ebenfalls vielversprechend ist. Die aktuelle Ära des maschinellen und tiefen Lernens könnte zukünftige Verbesserungen der zugrundeliegenden Algorithmen in `basf2` vorantreiben und die Integration externer KI-basierter Werkzeuge sowie datengetriebener Workflows erleichtern.
Tabelle 8.2 fasst die Hauptinhalte der einzelnen Kapitel zusammen.

Struktur	Zusammenfassung
Teil I	Kapitel 1: SM-Vorhersage und Formfaktoren von $B \to K^{(*)}\nu\bar{\nu}\nu$; Kapitel 2: Belle-II-Detektor und Software-Framework; Kapitel 3: Inklusive Tagging-Methode, Klassifikator-Framework, Signalbestimmung und Topologie-Techniken.
Teil II	Kapitel 4/5: Analysen von $B^+ \to K^+\nu\bar{\nu}$ / $B^0 \to K^*(892)^0\nu\bar{\nu}$
	Abschnitt 1: Daten- und Simulationsstichproben, Signalkonstruktion; Abschnitt 2: Zwei Klassifikatoren zur Untergrundunterdrückung; Abschnitt 3: Definition der Signalregion mithilfe des Klassifikator-Outputs und Validierung der Effizienz mit $B \to K^{(*)}J/\psi$; Abschnitt 4: Modellierung und Validierung der Untergründe $e^+e^- \to q\bar{q}, \tau^+\tau^-$ und $\Upsilon(4S) \to B\bar{B}$; Abschnitt 5: Signalextraktion; Abschnitt 6: Untersuchung der systematischen Unsicherheiten; Abschnitt 7: $B^+ \to K^+\nu\bar{\nu}$-Ergebnis mit einer Signifikanz von 2.9 σ.
Teil III	Kapitel 5: Hit-basiertes MVA-CDC-Untergrundfilter zur Verbesserung der Spurrekonstruktion. Kapitel 6: Hilfsmodellierung von $\gamma\gamma \to K^0_S K^0_S$ in der Belle-II-Simulation. Kapitel 7: Hilfsmodellierung von $B \to K\pi\pi\nu\bar{\nu}$ in der Belle-II-Simulation.

Table 8.2: Zusammenfassung der einzelnen Kapitel dieser Dissertation.

Chapter 9

Bibliography

[1] I. Adachi et al. Evidence for B+→K+$\nu\nu^{-}$ decays. *Phys. Rev. D*, 109(11):112006, 2024.

[2] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[3] C. Quigg. *Gauge Theories Of Strong, Weak, And Electromagnetic Interactions: First Edition*, volume 56 of *Advanced Book Program*. Benjamin-Cummings, USA, 1 edition, 9 1983.

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

[5] C. A. Heusch. Inelastic Charged-Lepton Scattering Off Nucleons. In *International Neutrino Conference 1976*, pages 426–448, 9 1976.

[6] Jerome I. Friedman. Deep inelastic scattering: Comparisons with the quark model. *Rev. Mod. Phys.*, 63:615–629, 1991.

[7] R. Brandelik et al. Evidence for a Spin One Gluon in Three Jet Events. *Phys. Lett. B*, 97:453–458, 1980.

[8] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[9] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[10] Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ decay channel with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector at the LHC. 4 2020.

[11] A measurement of the Higgs boson mass in the diphoton decay channel. 2019.

[12] S. Schael et al. Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427:257–454, 2006.

[13] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[14] Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, 8 2013.

[15] Ta-Pei [0000-0002-1137-0969] Cheng and Ling-Fong [0000-0002-8035-3329] Li. *Gauge Theory of Elementary Particle Physics*. Oxford University Press, Oxford, UK, 1984.

[16] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[17] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[18] Georges Aad et al. A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. *Nature*, 607(7917):52–59, 2022. [Erratum: Nature 612, E24 (2022)].

[19] Armen Tumasyan et al. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917):60–68, 2022. [Erratum: Nature 623, (2023)].

[20] Eite Tiesinga, Peter J. Mohr, David B. Newell, and Barry N. Taylor. CODATA recommended values of the fundamental physical constants: 2018*. *Rev. Mod. Phys.*, 93(2):025010, 2021.

[21] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[22] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[23] Andrei D Sakharov. Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe. *Soviet Physics Uspekhi*, 34(5):392, may 1991.

[24] Harald Fritzsch and Zhi-zhong Xing. Mass and flavor mixing schemes of quarks and leptons. *Prog. Part. Nucl. Phys.*, 45:1–81, 2000.

[25] Zhi-zhong Xing. Flavor structures of charged fermions and massive neutrinos. *Phys. Rept.*, 854:1–147, 2020.

[26] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, 2:1285–1292, 1970.

[27] Gerhard Buchalla, Andrzej J. Buras, and Markus E. Lautenbacher. Weak Decays beyond Leading Logarithms. *Rev. Mod. Phys.*, 68:1125–1144, 1996.

[28] Wolfgang Altmannshofer and Peter Stangl. New physics in rare B decays after Moriond 2021. *Eur. Phys. J. C*, 81(10):952, 2021.

[29] R. Aaij et al. Test of lepton universality in $b \to s\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 131(5):051803, 2023.

[30] T. Inami and C. S. Lim. Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes k(L) −−> mu anti-mu, K+ −−> pi+ Neutrino anti-neutrino and K0 <−−> anti-K0. *Prog. Theor. Phys.*, 65:297, 1981. [Erratum: Prog.Theor.Phys. 65, 1772 (1981)].

[31] Andrzej J. Buras. Weak Hamiltonian, CP violation and rare decays. In *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pages 281–539, 6 1998.

[32] M. Misiak, Abdur Rehman, and Matthias Steinhauser. Towards $\overline{B} \to X_s\gamma$ at the NNLO in QCD without interpolation in m_c. *JHEP*, 06:175, 2020.

[33] Swagato Banerjee et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023. 11 2024.

[34] Tobias Huber, Tobias Hurth, and Enrico Lunghi. Inclusive $\overline{B} \to X_s\ell^+\ell^-$: complete angular analysis and a thorough study of collinear photons. *JHEP*, 06:176, 2015.

[35] Tobias Felkl, Sze Lok Li, and Michael A. Schmidt. A tale of invisibility: constraints on new physics in $b \to s\nu\nu$. *JHEP*, 12:118, 2021.

[36] Andrzej J. Buras, Jennifer Girrbach-Noe, Christoph Niehoff, and David M. Straub. $B \to K^{(*)}\nu\overline{\nu}$ decays in the Standard Model and beyond. *JHEP*, 02:184, 2015.

[37] W. G. Parrott, C. Bouchard, and C. T. H. Davies. Standard Model predictions for B→Kℓ+ℓ-, B→Kℓ1-ℓ2+ and B→K$\nu\nu^-$ using form factors from Nf=2+1+1 lattice QCD. *Phys. Rev. D*, 107(1):014511, 2023. [Erratum: Phys.Rev.D 107, 119903 (2023)].

[38] Damir Bečirević, Gioacchino Piazza, and Olcyr Sumensari. Revisiting $B \to K^{(*)}\nu\bar{\nu}$ decays in the Standard Model and beyond. *Eur. Phys. J. C*, 83(3):252, 2023.

[39] Diganta Das, Gudrun Hiller, and Ivan Nisandzic. Revisiting $B \to K^*(\to K\pi)\nu\bar{\nu}$ decays. *Phys. Rev. D*, 95(7):073001, 2017.

[40] Julian Heeck. Unbroken B − L symmetry. *Phys. Lett. B*, 739:256–262, 2014.

[41] Feng-Zhi Chen, Qiaoyi Wen, and Fanrong Xu. Correlating $B \to K^{(*)}\nu\bar{\nu}$ and flavor anomalies in SMEFT. *Eur. Phys. J. C*, 84:1012, 2024.

[42] Biao-Feng Hou, Xin-Qiang Li, Meng Shen, Ya-Dong Yang, and Xing-Bo Yuan. Deciphering the Belle II data on $B \to K\nu\bar{\nu}$ decay in the (dark) SMEFT with minimal flavour violation. *JHEP*, 06:172, 2024.

[43] Javier Fuentes-Martín, Gino Isidori, Matthias König, and Nudžeim Selimović. Vector Leptoquarks Beyond Tree Level III: Vector-like Fermions and Flavor-Changing Transitions. *Phys. Rev. D*, 102:115015, 2020.

[44] Alessandra D'Alise, Giuseppe Fabiano, Domenico Frattulillo, Davide Iacobacci, Francesco Sannino, Pietro Santorelli, and Natascia Vignaroli. New physics pathways from B processes. *Nucl. Phys. B*, 1006:116631, 2024.

[45] Syuhei Iguro and Teppei Kitahara. Electric dipole moments as probes of the RD(*) anomaly. *Phys. Rev. D*, 110(7):075008, 2024.

[46] Chandan Hati, Julio Leite, Newton Nath, and José W. F. Valle. QCD axion, color-mediated neutrino masses, and B+→K++Emiss anomaly. *Phys. Rev. D*, 111(1):015038, 2025.

[47] Zeren Simon Wang, Herbert K. Dreiner, and Julian Y. Günther. The decay $B \to K + \nu + \bar{\nu}$ at Belle II and a massless bino in R-parity-violating supersymmetry. *Eur. Phys. J. C*, 85(1):66, 2025.

[48] Céline Boehm, Xiaoyong Chu, Jui-Lin Kuo, and Josef Pradler. Scalar dark matter candidates revisited. *Phys. Rev. D*, 103(7):075005, 2021.

[49] David McKeen, John N. Ng, and Douglas Tuckler. Higgs portal interpretation of the Belle II B+→K+$\nu\nu$ measurement. *Phys. Rev. D*, 109(7):075006, 2024.

[50] Kåre Fridell, Mitrajyoti Ghosh, Takemichi Okui, and Kohsaku Tobioka. Decoding the B→K$\nu\nu$ excess at Belle II: Kinematics, operators, and masses. *Phys. Rev. D*, 109(11):115006, 2024.

[51] T. M. Aliev, A. Elpe, L. Selbuz, and I. Turan. Explaining Belle data on B→K(*)$\nu\nu^-$ decays via dark Z resonances. *Phys. Rev. D*, 112(1):015025, 2025.

[52] Shu-Yu Ho, Jongkuk Kim, and Pyungwon Ko. Recent B+→K+$\nu\nu^-$ excess and muon g-2 illuminating light dark sector with Higgs portal. *Phys. Rev. D*, 111(5):055029, 2025.

[53] Xiyuan Gao and Ulrich Nierste. B→K+ axionlike particles: Effective versus UV-complete models and enhanced two-loop contributions. *Phys. Rev. D*, 112(5):055008, 2025.

[54] Luca Di Luzio, Alfredo Walter Mario Guerrera, Xavier Ponce Díaz, and Stefano Rigolin. Axion-like particles in radiative quarkonia decays. *JHEP*, 06:217, 2024.

[55] Xiao-Gang He, Xiao-Dong Ma, and German Valencia. Revisiting models that enhance B+→K+$\nu\nu^-$ in light of the new Belle II measurement. *Phys. Rev. D*, 109(7):075019, 2024.

[56] Tobias Felkl, Anjan Giri, Rukmani Mohanta, and Michael A. Schmidt. When energy goes missing: new physics in $b \to s\nu\nu$ with sterile neutrinos. *Eur. Phys. J. C*, 83(12):1135, 2023.

[57] F. Abudinén et al. Search for B+→K+$\nu\nu^{-}$ Decays Using an Inclusive Tagging Method at Belle II. *Phys. Rev. Lett.*, 127(18):181802, 2021.

[58] O. Lutz et al. Search for $B \to h^{(*)}\nu\bar{\nu}$ with the full Belle $\Upsilon(4S)$ data sample. *Phys. Rev. D*, 87(11):111103, 2013.

[59] J. Grygier et al. Search for $\boldsymbol{B \to h\nu\bar{\nu}}$ decays with semileptonic tagging at Belle. *Phys. Rev. D*, 96(9):091101, 2017. [Addendum: Phys.Rev.D 97, 099902 (2018)].

[60] J. P. Lees et al. Search for $B \to K^{(*)}\nu\bar{\nu}$ and invisible quarkonium decays. *Phys. Rev. D*, 87(11):112005, 2013.

[61] P. del Amo Sanchez et al. Search for the Rare Decay $B \to K\nu\bar{\nu}$. *Phys. Rev. D*, 82:112002, 2010.

[62] W. Altmannshofer et al. The Belle II Physics Book. *PTEP*, 2019(12):123C01, 2019. [Erratum: PTEP 2020, 029201 (2020)].

[63] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[64] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001.

[65] S. Jadach, B. F. L. Ward, and Z. Was. The Precision Monte Carlo event generator K K for two fermion final states in e+ e- collisions. *Comput. Phys. Commun.*, 130:260–325, 2000.

[66] G. Balossini, C. M. Carloni Calame, G. Montagna, O. Nicrosini, and F. Piccinini. Matrix elements and Parton Shower in the event generator BABAYAGA. *Nucl. Phys. B Proc. Suppl.*, 162:59–62, 2006.

[67] Frits A. Berends, P. H. Daverveldt, and R. Kleiss. Complete Lowest Order Calculations for Four Lepton Final States in electron-Positron Collisions. *Nucl. Phys. B*, 253:441–463, 1985.

[68] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. SuperKEKB Collider. *Nucl. Instrum. Meth. A*, 907:188–199, 2018.

[69] N. Ohuchi, Y. Arimoto, K. Akai, K. Aoki, N. Higashi, K. Kanazawa, M. Kawai, T. Kawamoto, H. Koiso, Y. Kondou, et al. Superkekb beam final focus superconducting magnet system. *Nuclear Instruments and Methods in Physics Research. Section A, Accelerators, Spectrometers, Detectors and Associated Equipment*, 1021, 10 2021.

[70] T. Abe et al. Belle II Technical Design Report. 11 2010.

[71] P. Ahlburg et al. The new and complete Belle II DEPFET pixel detector: Commissioning and previous operational experience. *Nucl. Instrum. Meth. A*, 1068:169763, 2024.

[72] R. Leboucher et al. Measurement of the cluster position resolution of the Belle II Silicon Vertex Detector. *Nucl. Instrum. Meth. A*, 1033:166746, 2022.

[73] Bianca Scavino. Tracking and vertexing in BELLE II. *PoS*, VERTEX2018:040, 2019.

[74] I. Adachi, T. E. Browder, P. Križan, S. Tanaka, and Y. Ushiroda. Detectors for extreme luminosity: Belle II. *Nucl. Instrum. Meth. A*, 907:46–59, 2018.

[75] Ezio Torassa. TOP detector for particle identification at Belle II. *Int. J. Mod. Phys. A*, 39(26n27):2442014, 2024.

[76] Satoru Yamada, Ryosuke Itoh, Katsuro Nakamura, Mikihiko Nakao, Soh Y. Suzuki, Tomoyuki Konno, Takeo Higuchi, Zhen'an Liu, and Jingzhou Zhao. Data Acquisition System for the Belle II Experiment. *IEEE Trans. Nucl. Sci.*, 62(3):1175–1180, 2015.

[77] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II Core Software. *Comput. Softw. Big Sci.*, 3(1):1, 2019.

[78] A. Natochii et al. Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider. *Nucl. Instrum. Meth. A*, 1055:168550, 2023.

[79] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was, and Z. Was. Universal Interface of TAUOLA Technical and Physics Documentation. *Comput. Phys. Commun.*, 183:821–843, 2012.

[80] Piotr Golonka and Zbigniew Was. PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays. *Eur. Phys. J. C*, 45:97–107, 2006.

[81] Elisabetta Barberio, Bob van Eijk, and Zbigniew Was. PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays. *Comput. Phys. Commun.*, 66:115–128, 1991.

[82] Sadaharu Uehara. TREPS: A Monte-Carlo Event Generator for Two-photon Processes at e^+e^- Colliders using an Equivalent Photon Approximation. 7 1996.

[83] Valerio Bertacchi et al. Track finding at Belle II. *Comput. Phys. Commun.*, 259:107610, 2021.

[84] J. F. Krohn et al. Global decay chain vertex fitting at Belle II. *Nucl. Instrum. Meth. A*, 976:164269, 2020.

[85] T. Keck et al. The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, 3(1):6, 2019.

[86] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009.

[87] H. Voss, Andreas Hocker, J. Stelzer, and F. Tegenfeldt. TMVA, the Toolkit for Multivariate Data Analysis with ROOT. *PoS*, ACAT:040, 2007.

[88] Thomas Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, 1(1):2, 2017.

[89] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. 3 2016.

[90] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[91] Alexander L. Read. Presentation of search results: The CL_s technique. *J. Phys. G*, 28:2693–2704, 2002.

[92] J. D. Bjorken and Stanley J. Brodsky. Statistical Model for electron-Positron Annihilation Into Hadrons. *Phys. Rev. D*, 1:1416–1420, 1970.

[93] Geoffrey C. Fox and Stephen Wolfram. Event Shapes in e+ e- Annihilation. *Nucl. Phys. B*, 149:413, 1979. [Erratum: Nucl.Phys.B 157, 543 (1979)].

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

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

[96] I. Adachi et al. Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*. *Chin. Phys. C*, 49(1):013001, 2025.

[97] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr. Advanced event reweighting using multivariate analysis. *J. Phys. Conf. Ser.*, 368:012028, 2012.

[98] J. P. Lees et al. Study of CP violation in Dalitz-plot analyses of B0 −-> K+K-K0(S), B+ −-> K+K-K+, and B+ −-> K0(S)K0(S)K+. *Phys. Rev. D*, 85:112010, 2012.

[99] Bernard Aubert et al. Measurement of the $B^+ \to p\bar{p}K^+$ branching fraction and study of the decay dynamics. *Phys. Rev. D*, 72:051101, 2005.

[100] Bernard Aubert et al. Evidence for the B0 −−> p anti-p K*0 and B+ −-> eta(c) K*+ decays and Study of the Decay Dynamics of B Meson Decays into p anti-p h final states. *Phys. Rev. D*, 76:092004, 2007.

[101] Matthew Feickert, Lukas Heinrich, and Giordon Stark. pyhf: a pure-Python statistical fitting library with tensors and automatic differentiation. *PoS*, ICHEP2022:245, 11 2022.

[102] F. Abudinén et al. Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment. *Chin. Phys. C*, 44(2):021001, 2020.

[103] F. Abudinén et al. B counting in moriond 2022/ichep 2022 dataset. *BELLE2-NOTE-PH-2022-007*.

[104] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[105] Cyrille Praz. *Search for B+→K+νν ¯ decays with a machine learning method at the Belle II experiment*. PhD thesis, Hamburg, University of Hamburg, Hamburg, 2022. See page 107-113 of https://ediss.sub.uni-hamburg.de/handle/ediss/9793.

[106] Priyanka Cheema. Suppressing Beam Background and Fake Photons at Belle II using Machine Learning. *EPJ Web Conf.*, 295:09035, 2024.

[107] J. Podolanski and R. Armenteros. Iii. analysis of v-events. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 45(360):13–30, 1954.

[108] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ROC Analysis in Pattern Recognition.

[109] Andrew P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[110] J. H. Chen et al. Observation of B0 −-> p anti-p K*0 with a large K*0 polarization. *Phys. Rev. Lett.*, 100:251801, 2008.

[111] J.Park. Measurements of electroweak penguin and lepton-flavour violating decays to final states with missing energy at belle and belle ii. Presented at 23rd Conference on Flavor Physics and CP Violation (FPCP2025),USA., 2025, 2025. `https://indico.cern.ch/event/1421485/contributions/6526355/`.

[112] Roel Aaij et al. First observations of the rare decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$. *JHEP*, 10:064, 2014.

[113] H. Guler et al. Study of the $K^+\pi^+\pi^-$ Final State in $B^+ \to J/\psi K^+\pi^+\pi^-$ and $B^+ \to \psi - prime K^+\pi^+\pi^-$. *Phys. Rev. D*, 83:032005, 2011.

[114] Swagato Banerjee et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023. 11 2024.

[115] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. 7 2019.

[116] Muriel Pivk and Francois R. Le Diberder. SPlot: A Statistical tool to unfold data distributions. *Nucl. Instrum. Meth. A*, 555:356–369, 2005.

[117] D Morgan, M R Pennington, and M R Whalley. A compilation of data on two-photon reactions leading to hadron final states. *Journal of Physics G: Nuclear and Particle Physics*, 20(8A):A1, aug 1994.

[118] H. Albrecht et al. Measurement of K* anti-K* production in two photon interactions. *Eur. Phys. J. C*, 16:435–444, 2000.

[119] M. Masuda et al. Study of K^0_S pair production in single-tag two-photon collisions. *Phys. Rev. D*, 97(5):052003, 2018.

[120] S. Uehara et al. Study of charmonia in four-meson final states produced in two-photon collisions. *Eur. Phys. J. C*, 53:1–14, 2008.

[121] S. Uehara et al. High-statistics study of eta pi0 production in two-photon collisions. *Phys. Rev. D*, 80:032001, 2009.

[122] Belle Collaboration. High-statistics study of K^0_S pair production in two-photon collisions. HEPData (collection), 2013. `https://doi.org/10.17182/hepdata.62221`.

[123] Lea Reuter et al. End-to-End Multi-track Reconstruction Using Graph Neural Networks at Belle II. *Comput. Softw. Big Sci.*, 9(1):6, 2025.

[124] H. Albrecht et al. First Observation of $\gamma\gamma \to K^*0\bar{K}^*0$. *Phys. Lett. B*, 198:255–260, 1987.

[125] H. Albrecht et al. First Observation of $\gamma\gamma \to K^* + K^*$-. *Phys. Lett. B*, 212:528–532, 1988.

[126] H. J. Behrend et al. The $K^0_S K^0_S$ Final State in $\gamma\gamma$ Interactions. *Z. Phys. C*, 43:91, 1989.

Appendix 9.1: Inputs Variables of the Second Classifier of $B^+ \to K + \nu\bar{\nu}$ Analysis

This appendix provides the set of 35 input variables employed for training the second multivariate classifier in the $B^+ \to K^+\nu\bar{\nu}$ analysis, as shown in Figs. 9.1, 9.2, 9.3, 9.4, 9.5 and 9.6,. In accordance with the adopted blinding strategy, the selection $\text{BDT}_1 < 0.99$ is imposed, even though the fact that the signal region has subsequently been unblinded, and the corresponding results have been published.

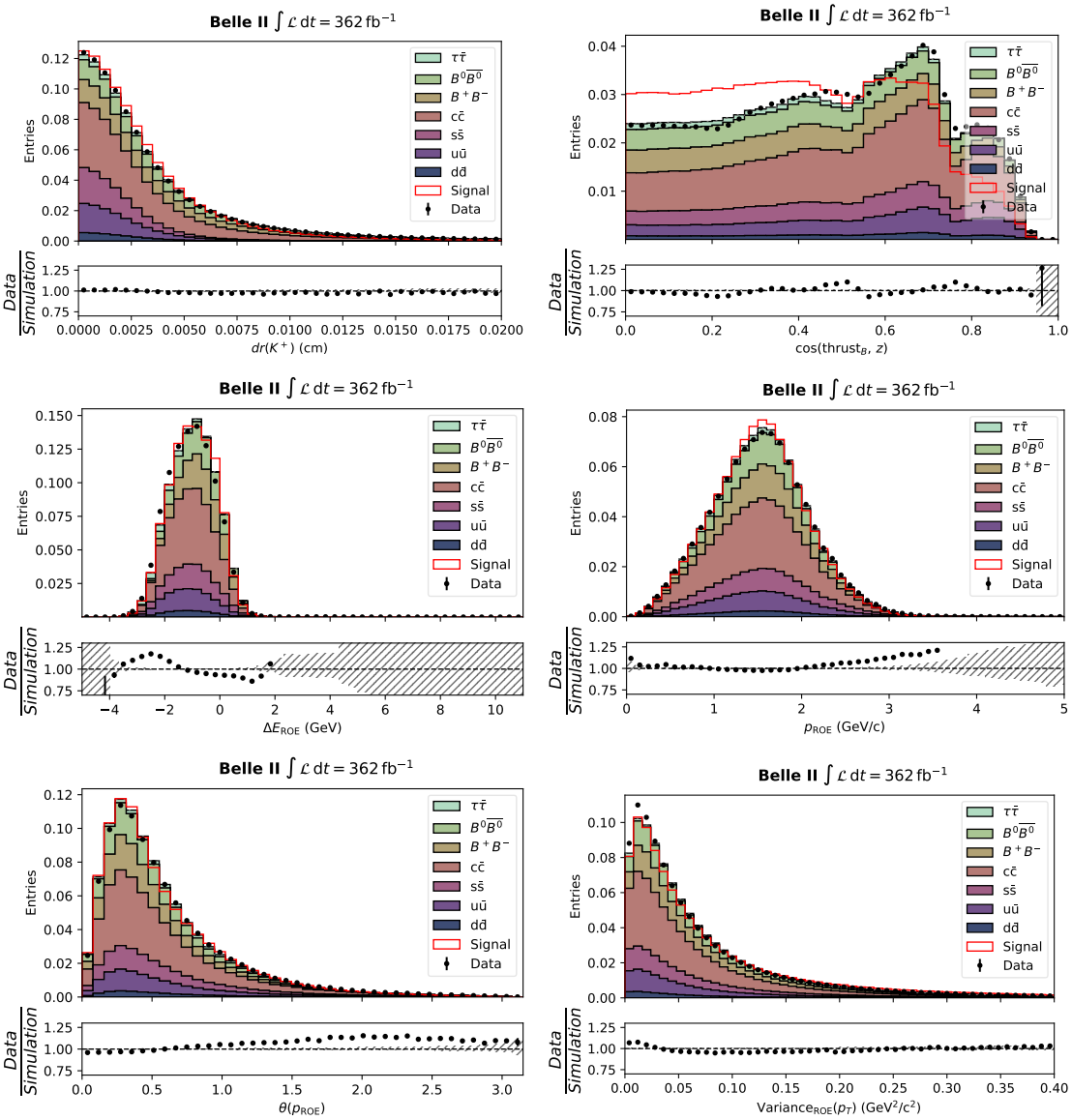

Figure 9.1: Distributions of $d_r(K^+)$, $\cos(\text{thrust}_B, z)$, ΔE_{ROE}, p_{ROE}, $\theta(p_{\text{ROE}})$ and $\text{Var}(p_T^i | i \in \text{ROE})$ of $0.9 < \text{BDT}_1 < 0.99$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. The total $\Upsilon(4S)$ data luminosity is used.

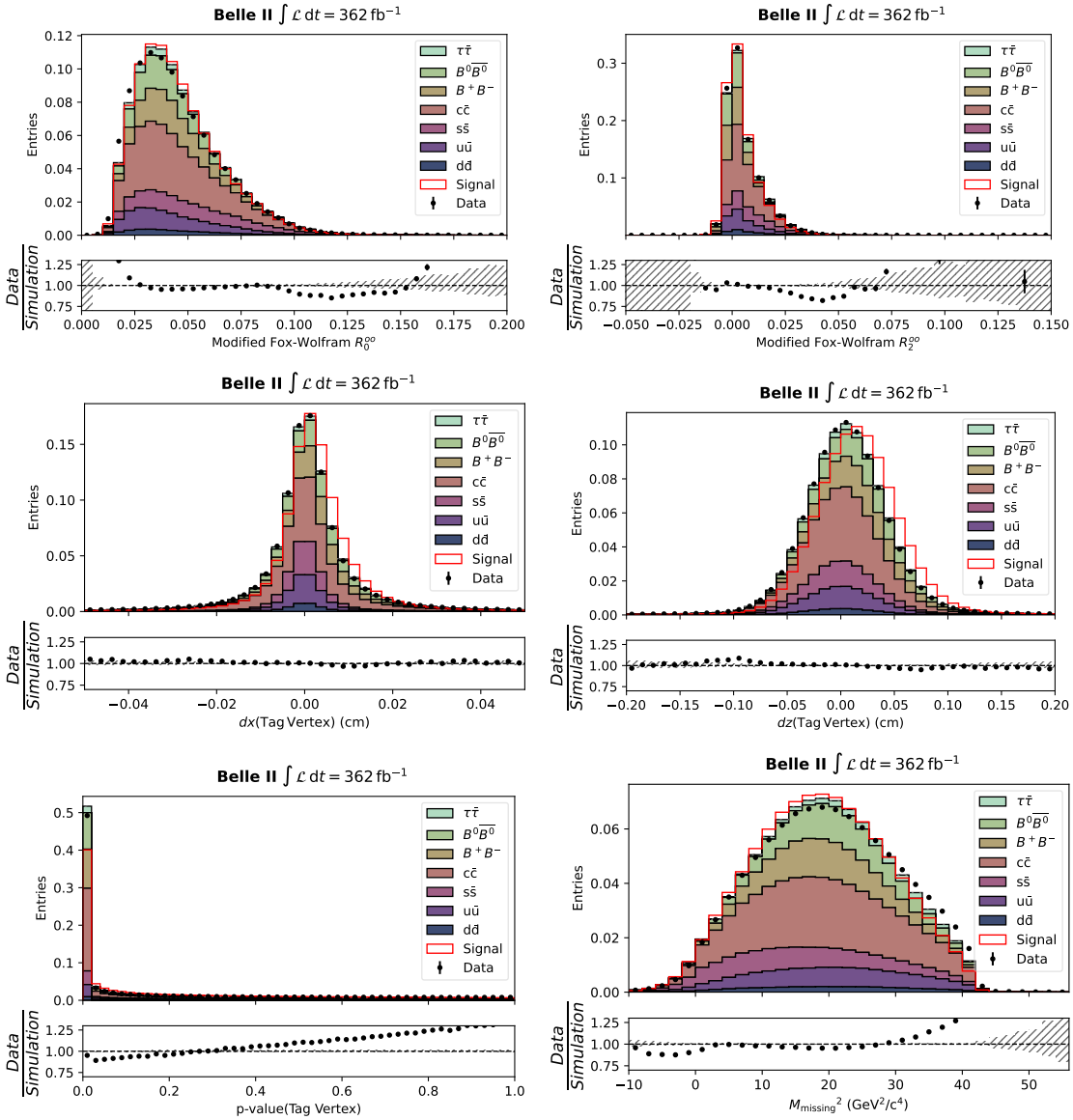

Figure 9.2: Distributions of R_0^{oo}, R_2^{oo}, $d_x(\mathrm{TagV})$, $d_z(\mathrm{TagV})$, p-value(ROE vertex) and M^2_{missing} of $0.9 < \mathrm{BDT}_1 < 0.99$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. The total $\Upsilon(4S)$ data luminosity is used.

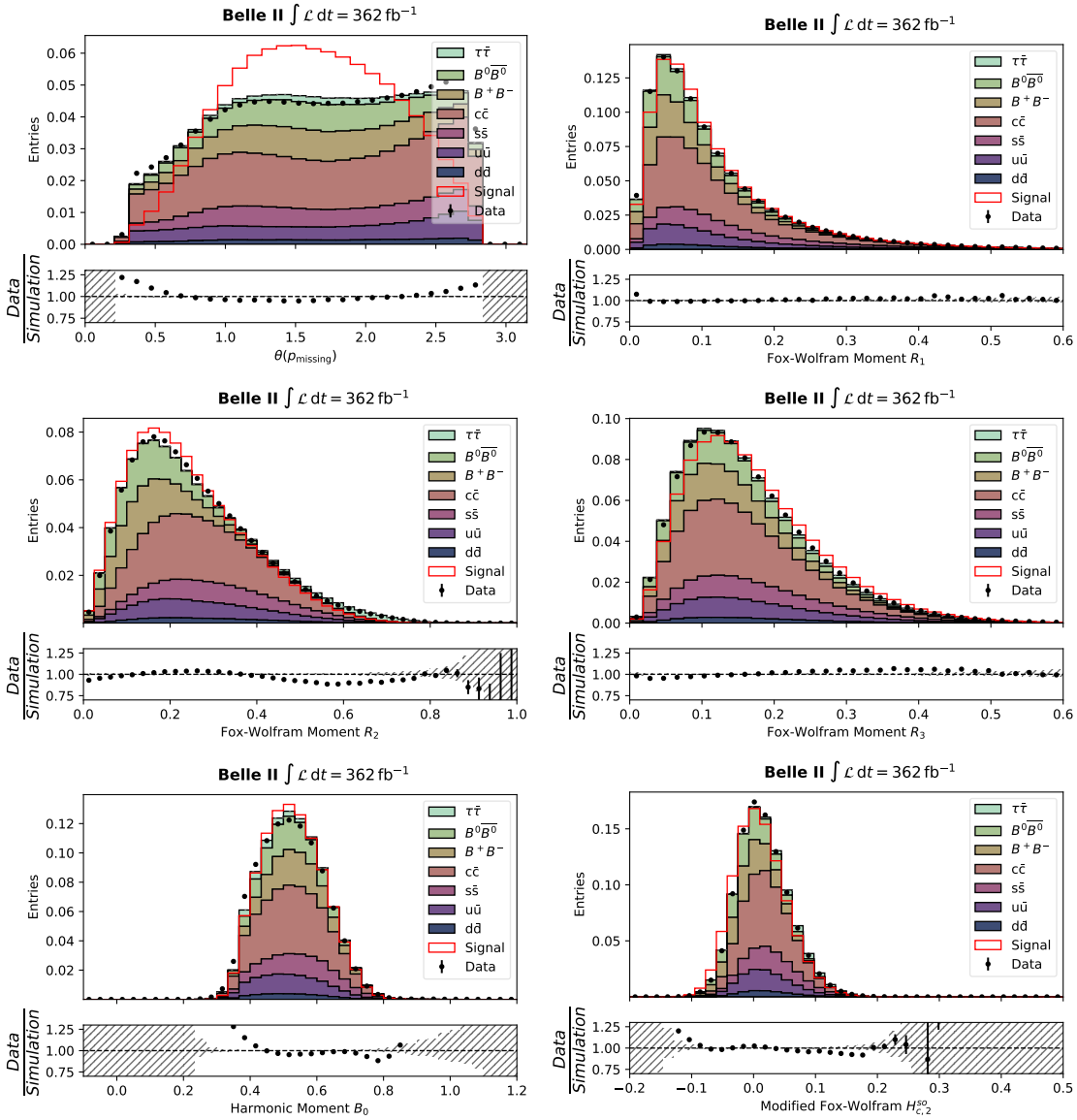

Figure 9.3: Distributions of the $\theta(p_{\text{missing}})$, Fox–Wolfram moment R_1, R_2, R_3, harmonic moment B_0, and modified Fox–Wolfram moments $H^{\text{so}}_{c,2}$ of $0.9 < \text{BDT}_1 < 0.99$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. The total $\Upsilon(4S)$ data luminosity is used.

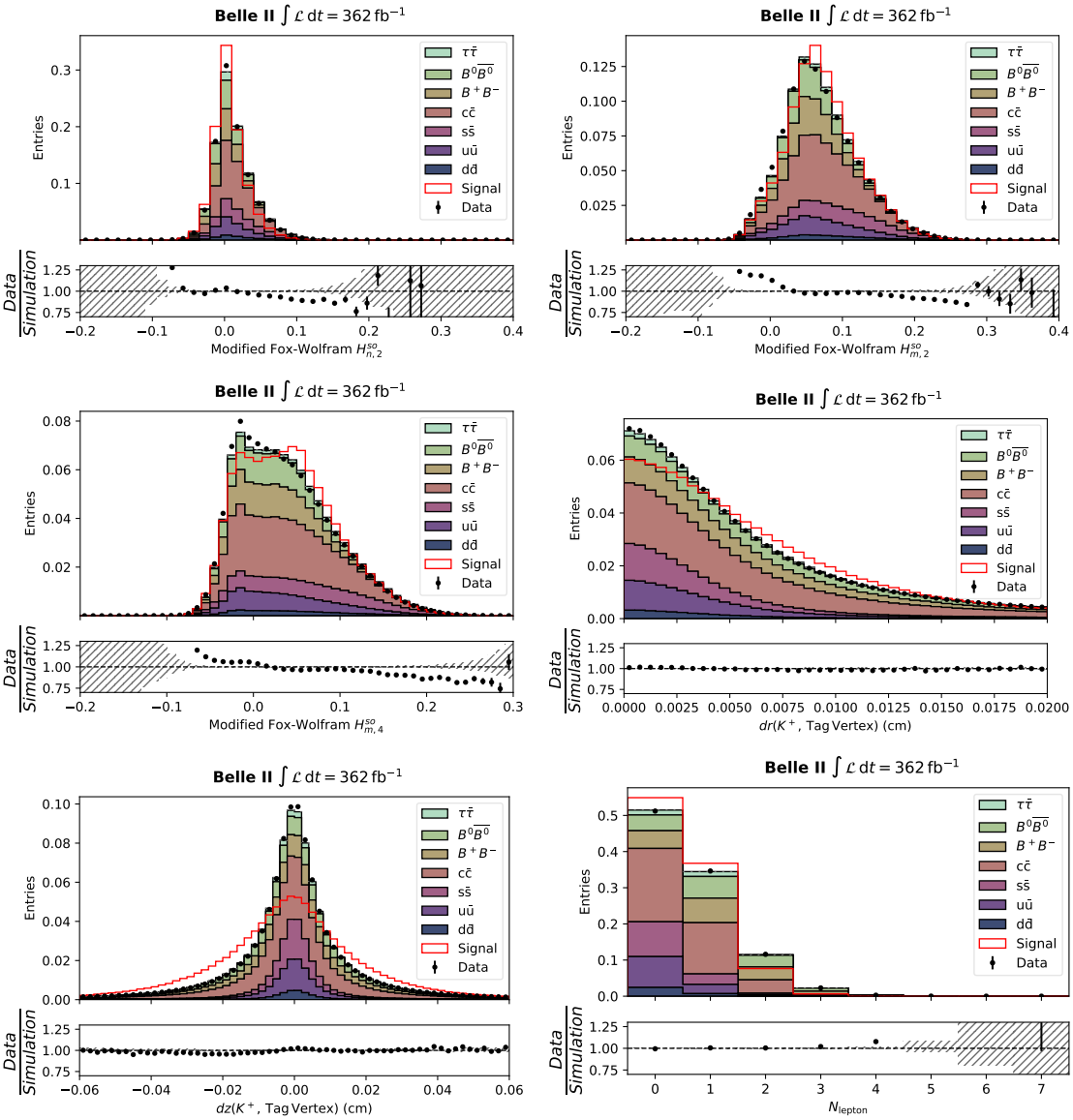

Figure 9.4: Distributions of the $H^{\text{so}}_{n,2}$, $H^{\text{so}}_{m,2}$, and $H^{\text{so}}_{m,4}$, the distances $d_r(K^+, \text{TagV})$ and $d_z(K^+, \text{TagV})$ and the multiplicities N_ℓ of $0.9 < \text{BDT}_1 < 0.99$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. The total $\Upsilon(4S)$ data luminosity is used.

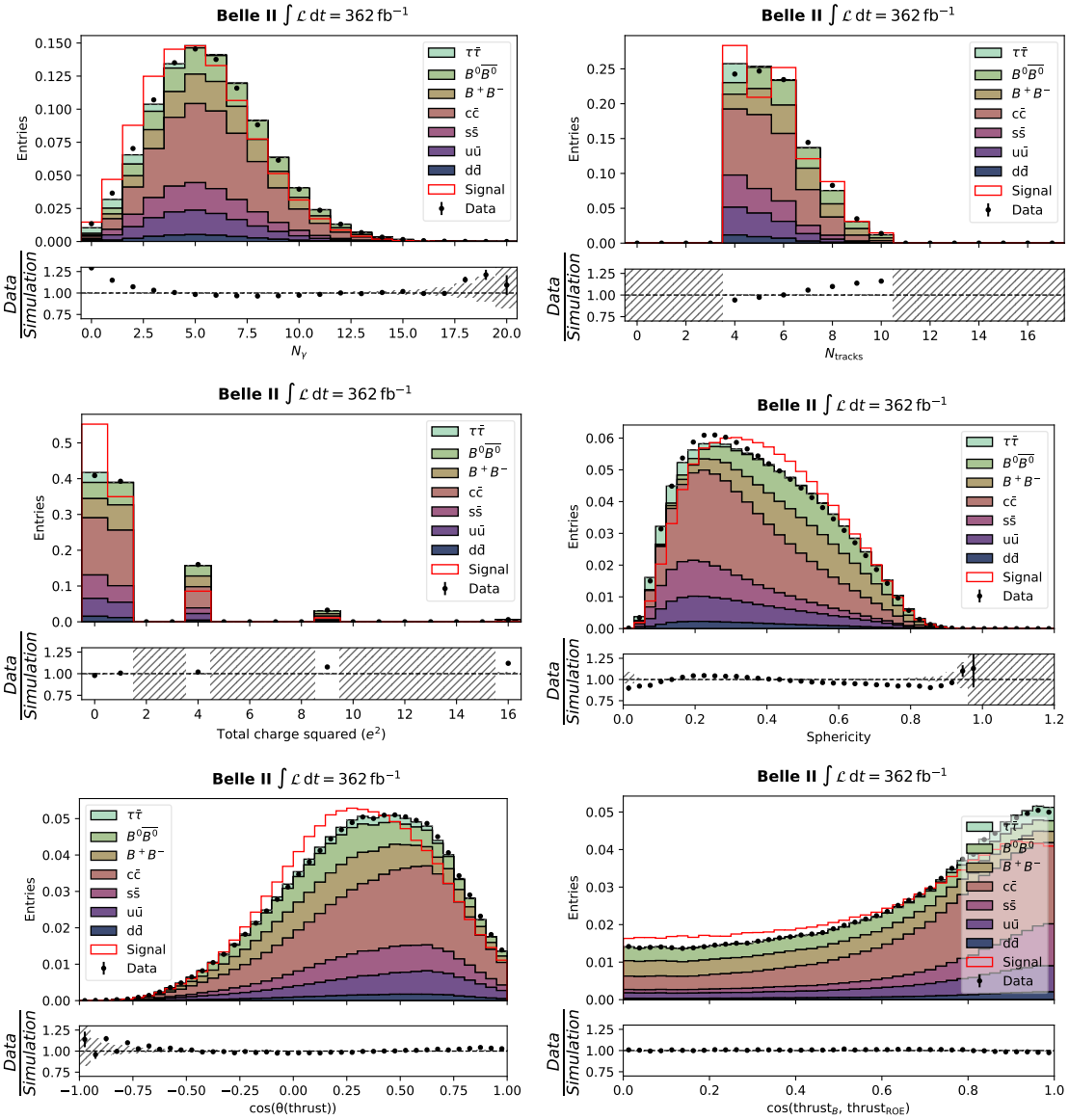

Figure 9.5: Distributions of the N_γ, and N_{tracks} and the squared event charge e^2, event sphericity, the angular variables $\cos\theta(\text{thrust})$, and the $\cos(\text{thrust}_K, \text{thrust}_{\text{ROE}})$ of $0.9 < \text{BDT}_1 < 0.99$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. The total $\Upsilon(4S)$ data luminosity is used.

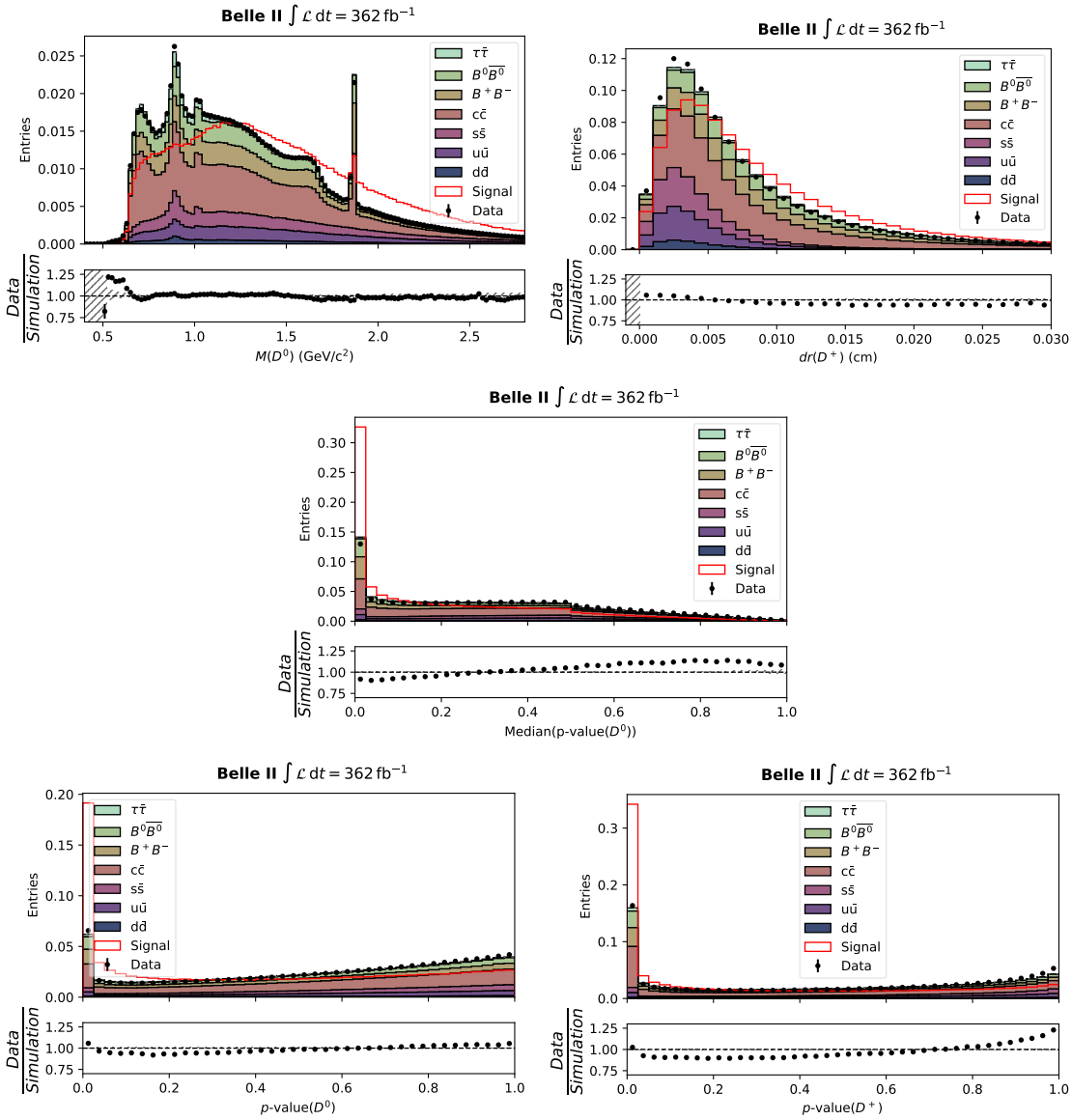

Figure 9.6: Distributions of the mass of the best D^0 candidate $M(K^+X^-)$, the radial distance $d_r(D^+)$ of the D^+ candidate vertex to the interaction point, the median p-value of the D^0 vertex fits, the p-value of the best D^0 candidate, and the p-value of the best D^+ candidate, for $0.9 < \mathrm{BDT}_1 < 0.99$ in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^+ \to K^+\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. The total $\Upsilon(4S)$ data luminosity is used.

Appendix 9.2: Lepton Sideband of $D \to K^0_L X$ study in $B^+ \to K^+ \nu\bar{\nu}$ Analysis

This appendix provides a validation of the correction derived for $D \to K^0_L X$ decays using lepton-enriched sideband samples. In particular, the $q^2_{\rm rec}$ distributions are examined in the muon- and electron-enriched control samples. These studies serve as an independent cross-check of the correction applied in the nominal analysis and demonstrate that the data-to-simulation agreement improves significantly after applying the normalization factor.

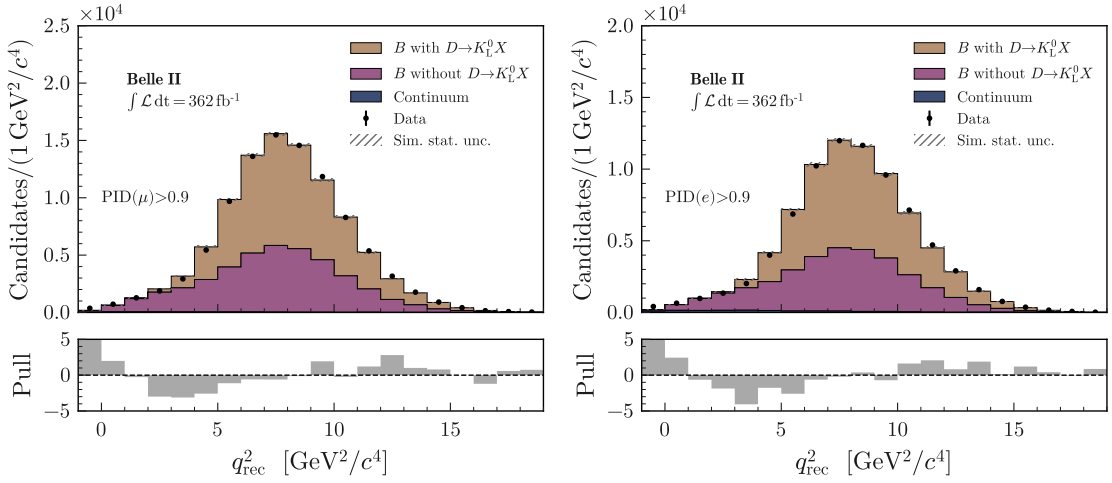

Figure 9.7: Distribution of $q^2_{\rm rec}$ in data (points with error bars) and simulation (filled histograms), separated into three categories: B-meson decays with and without subsequent $D \to K^0_L X$ decays, and the sum of the five continuum categories. The left (right) panel corresponds to the μ- (e-) enriched PID control sample, with $\eta(\text{BDT}_2) > 0.92$ in the ITA. The pull distributions are shown in the lower panels. The total $\Upsilon(4S)$ data luminosity is used.

Appendix 9.3: Optimization of Photon Selection for ROE in $B^0 \to K^{*0}\nu\bar{\nu}$ Analysis

In the published $B^+ \to K^+\nu\bar{\nu}$ analysis, photons in the rest of the event (ROE) were required to have an energy exceeding $100\ \mathrm{MeV}$. To assess the robustness of this criterion and explore possible improvements, we investigate an alternative photon selection in the $B^0 \to K^{*0}\nu\bar{\nu}$ channel. Specifically, the energy threshold is reduced to $60\ \mathrm{MeV}$, and potential mis-reconstructed photons originating from hadronic split-offs are suppressed by requiring the cluster–track distance ($\mathrm{minC2TDist}$) to exceed $20\ \mathrm{cm}$.
As illustrated in Fig. 9.8, this modified selection improves the agreement between data and simulation in the $\Delta E(\mathrm{ROE})$ distribution and yields enhanced consistency near the peak of H^{so}_{22}, albeit with some degradation at the distribution edges.

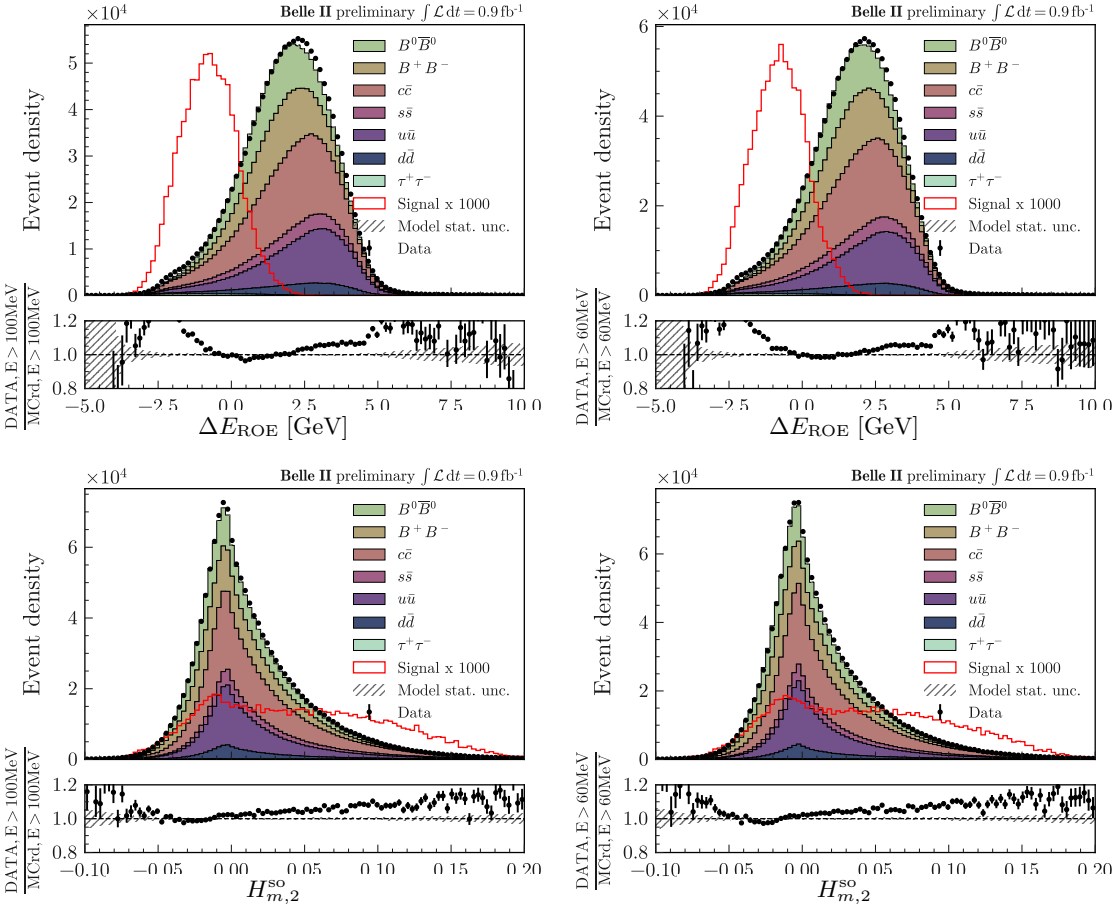

Figure 9.8: Data and simulation distributions of ΔE_{ROE}(top) and $H^{\mathrm{so}}_{m,2}$ (bottom) reconstructed candidates after applying (left) the selection $E > 100$ MeV and (right) the selection $E > 60$ MeV and $\mathrm{minC2TDist} > 20\mathrm{cm}$.

We quantify the impact of the two photon-selection strategies on both the agreement between data and simulation, and on the separation power between signal and background by means of the Jensen–Shannon (JS) distance, as illustrated in Fig. 9.9.
The change in data–simulation agreement for a given variable x is expressed as

$$\Delta d(x) = d_{\mathrm{JS}}^{\mathrm{new}}(x) - d_{\mathrm{JS}}^{\mathrm{nominal}}(x), \tag{9.1}$$

which represents the difference between the JS distance obtained with the modified selection and that of the nominal selection. The relative change in signal-background discrimination for a set of BDT_1 input variables X is defined as

$$\Delta d_X = \frac{d_{\mathrm{JS}}^{\mathrm{new}}(X) - d_{\mathrm{JS}}^{\mathrm{nominal}}(X)}{d_{\mathrm{JS}}^{\mathrm{nominal}}(X)}, \tag{9.2}$$

providing a normalized measure of the variation in separation performance.

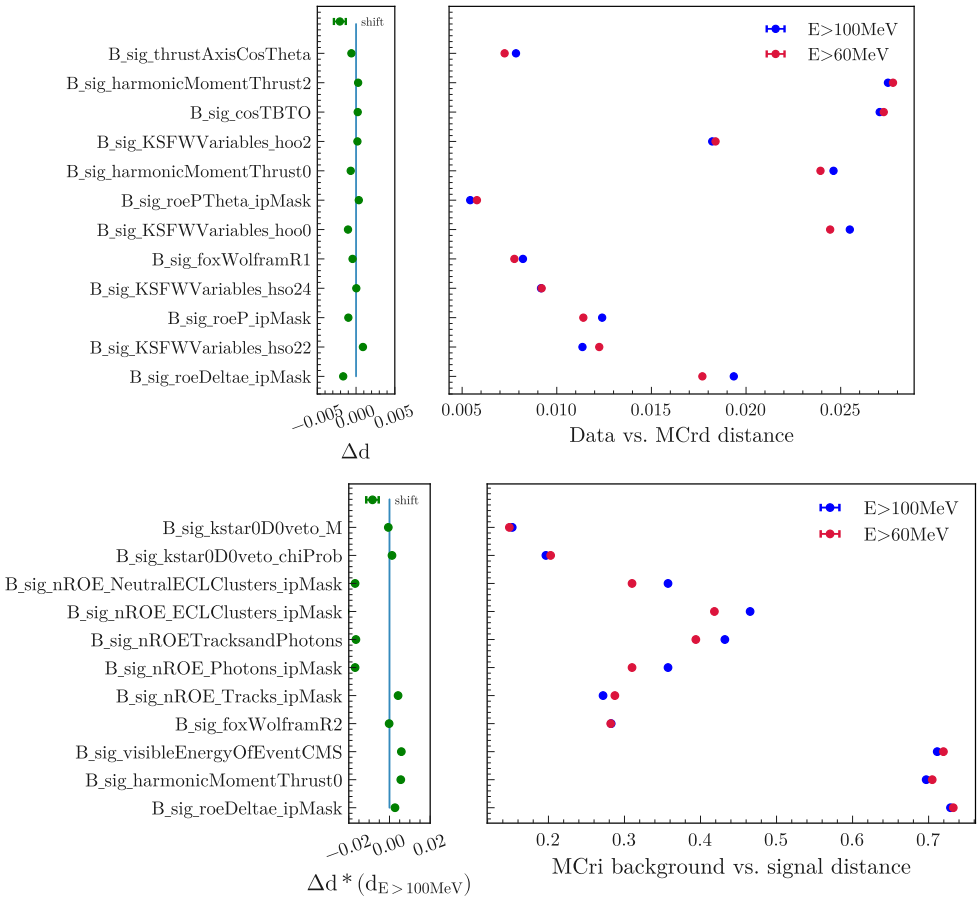

Figure 9.9: Jensen-Shannon distances between data and MC (upper) and signal and background (bottom) distributions for the BDT_1 training inputs. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.

The two photon-selection strategies yield comparable overall performance, and no clear preference can be established. Nevertheless, an improvement in the agreement between data and simulation is observed for events containing a large number of neutral clusters. Figure 9.10 presents the distributions of the number of ECL clusters in the ROE as well as the combined multiplicity of tracks and neutral clusters, comparing the two selection criteria. The reduced energy threshold provides superior data–simulation agreement, particularly in events with more than ten photons.

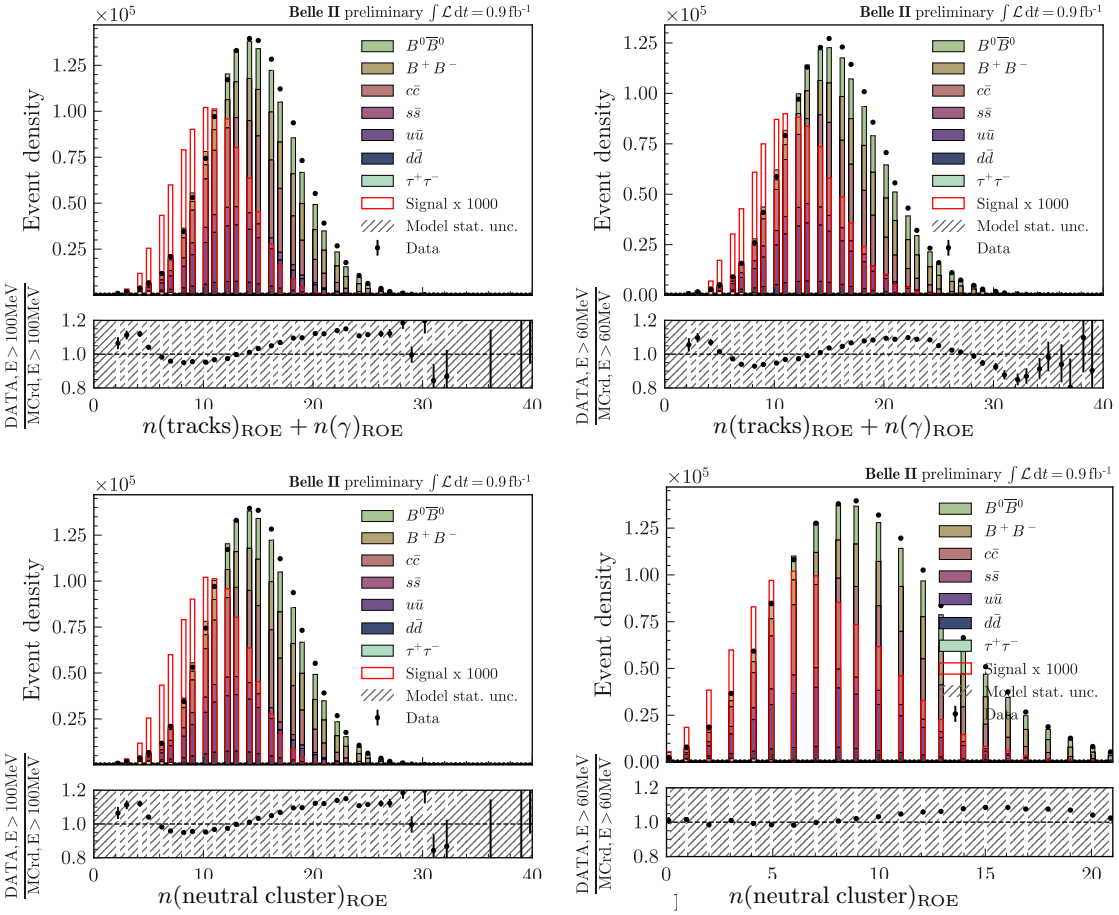

Figure 9.10: Data-MC distributions of number of tracks and photons in the ROE (top) and number of ECL neutral clusters in the ROE (bottom) reconstructed candidates after applying the selection $E > 100$ MeV (left) and the selection $E > 60$ MeV and $\text{minC2TDist} > 20$ cm (right). 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

Appendix 9.4: Low-multiplicity Events Suppression in $B^0 \to K^{*0}\nu\bar{\nu}$ Analysis

Discrepancies observed in the lower mass region of ROE and the angle of missing momentum distributions suggest that the current simulation does not fully account for specific low-multiplicity processes. In particular, the excess in the low-visible-energy region points to contributions from processes that produce only a few reconstructed particles, thereby mimicking the kinematic features of $B^0 \to K^{(*)}\nu\bar{\nu}$ decays. Such effects are especially relevant in channels with stringent event-level requirements, as they can introduce background components not modeled in the simulation and consequently bias efficiency estimates or background extrapolations.

Figure 9.11 illustrates the data−simulation discrepancies attributed to low-multiplicity processes. The effect is most pronounced in variables related to the missing momentum, indicating that such events preferentially populate regions of phase space where genuine $B^0 \to K^{*0}\nu\bar{\nu}$ decays are expected. Moreover, the discrepancy extends across the full invariant-mass range of the K^{*0} candidate, suggesting that it cannot be mitigated by a simple resonance selection and instead requires a more comprehensive treatment of the underlying low-multiplicity processes in the simulation.

To identify the source of the observed discrepancy, we examined possible missing or mis-modeled processes in the simulation. It was found that the Belle II simulation does not include low-multiplicity two-photon ($\gamma\gamma$) processes such as $\gamma\gamma \to K\pi K\pi$, $K^*K\pi$, $K^*\bar{K}^*$, $K^0_S K^0_S$, $\rho\phi$, and $\rho\omega$. The cross sections for these processes are comparatively large [124–126], and the presence of a $KK^{(*)}$-like system in the final state gives them a topology similar to that of $B^0 \to K^{*0}\nu\bar{\nu}$ signal events. To quantify their contribution, we focus on the $M_{\rm ROE} < 3.5$ GeV/c^2) and only two tracks in ROE, where such two-photon processes are expected to dominate, as shown in Fig. 9.12 (left).
These low-multiplicity contributions are effectively suppressed by applying the selection criteria $0.3 < \theta_{p_{\rm miss}} < 2.8$ and $E_{\rm vis} > 4$ GeV, which remove events with extreme missing-momentum directions and low region of $E_{\rm visible}$, as illustrated in Fig. 9.12 (right). Actually, various event-level variables and their combinations were studied to suppress low-multiplicity processes. Among them, the most effective selections were found to be ($Et > 2.8$ GeV and $0.3 < \theta_{p_{\rm miss}} < 2.8$) and ($E_{\rm visible} > 4$ GeV and $0.3 < \theta_{p_{\rm miss}} < 2.8$). For consistency with the published $B^+ \to K^+\nu\bar{\nu}$ analysis, we adopt the $E_{\rm visble} > 4$ GeV and $0.3 < \theta_{p_{\rm missing}} < 2.8$ requirement, which reduces the low-multiplicity contribution while retaining ∼96% of the signal efficiency.

Appendix 9.5: D-meson Suppression in $B^0 \to K^{*0}\nu\bar{\nu}$ Analysis

To further reduce backgrounds arising from mis-reconstructed D mesons, we consider the combination $K^{*0}\ell^+$, which can originate from $D \to K^{*0}e^+\nu_e$ or $D \to K^{*0}\mu^+\nu_\mu$.

9.5.1 $D^+ \to K^{*0}\ell^+\nu$ Suppression

Here, the lepton ($\ell = e$ or μ) comes from the rest of the event (ROE). The selection requirements are:

- Electron channel: eID > 0.9
- Muon channel: μID > 0.9
- Invariant mass of the $K^{*0}\ell^+$ combination: $M < 2.1$ GeV
- A three-track kinematic fit is performed, and the best candidate is chosen based on the fit p-value

Fig 9.13 shows the mass of $K^{*0}_{\rm sig}\ell^+_{\rm ROE}$ system together with its polar angle with respect to the beam axis, and the vertex fit p-value of best D^+ candidate distribution.

9.5.2 D^0 Suppression

In addition to the D^+ veto, we also investigate possible backgrounds from mis-reconstructed D^0 decays. For all modes considered, the kaon is taken from the signal K^{*0} candidate. Each reconstructed combination is subjected to a kinematic (or tree) vertex fit, and the best candidate is chosen according to the fit p-value. The following D^0 decay topologies are studied:

- $D^0 \to K^-\pi^+$: The pion is taken from the ROE (excluding tracks from K^0_S). Candidates are required to satisfy $M(K\pi) < 2.1$ GeV.
- $D^0 \to K^-\ell^+\nu$: The lepton (e or μ) is taken from the ROE, with electron ID > 0.9 or muon ID > 0.9. The reconstructed $K\ell$ mass is required to be smaller than $2.1\ {\rm GeV}/c^2$.
- $D^0 \to K^-\pi^+\pi^0$: The charged pion is selected from the signal side or the ROE (excluding tracks from K^0_S) and are required to have at least one PXD hit. A π^0 candidate is added, and a tree fit with a mass constraint on the π^0 is performed.
- $D^0 \to K^-\pi^+\pi^+\pi^-$: The pions are chosen from the signal or ROE, each with at least one PXD hit. The reconstructed mass must lie within $1.6\ {\rm GeV}/c^2 < M(K3\pi) < 2.1\ {\rm GeV}/c^2$.

Fig. 9.14 shows the mass of $K^-_{\rm sig}\pi^+\pi^0_{\rm ROE}$ and $K3\pi$ system together with its polar angle with respect to the beam axis, and vertex fit p-value of best D^0 candidate distribution of the vertex fit.
These vetoes substantially reduce backgrounds from semileptonic and hadronic D decays while preserving a high signal efficiency in the $B^0 \to K^0\nu\bar{\nu}$ channel.

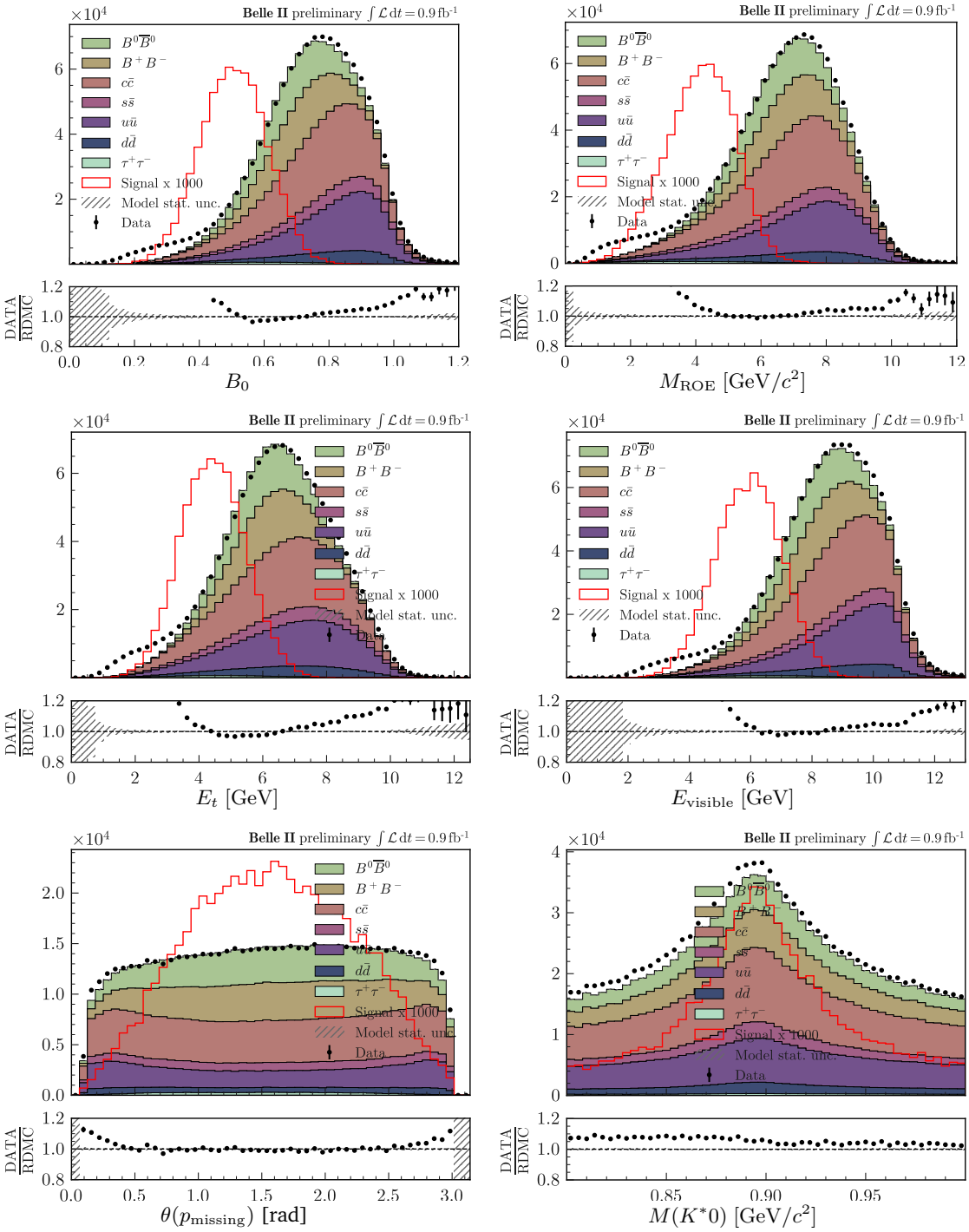

Figure 9.11: Distributions of the B_0, M_{ROE}, transverse energy E_t, visible energy in c.m frame E_{visible}, $\theta(p_{\text{missing}})$ and invariant mass of K^{*0} in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

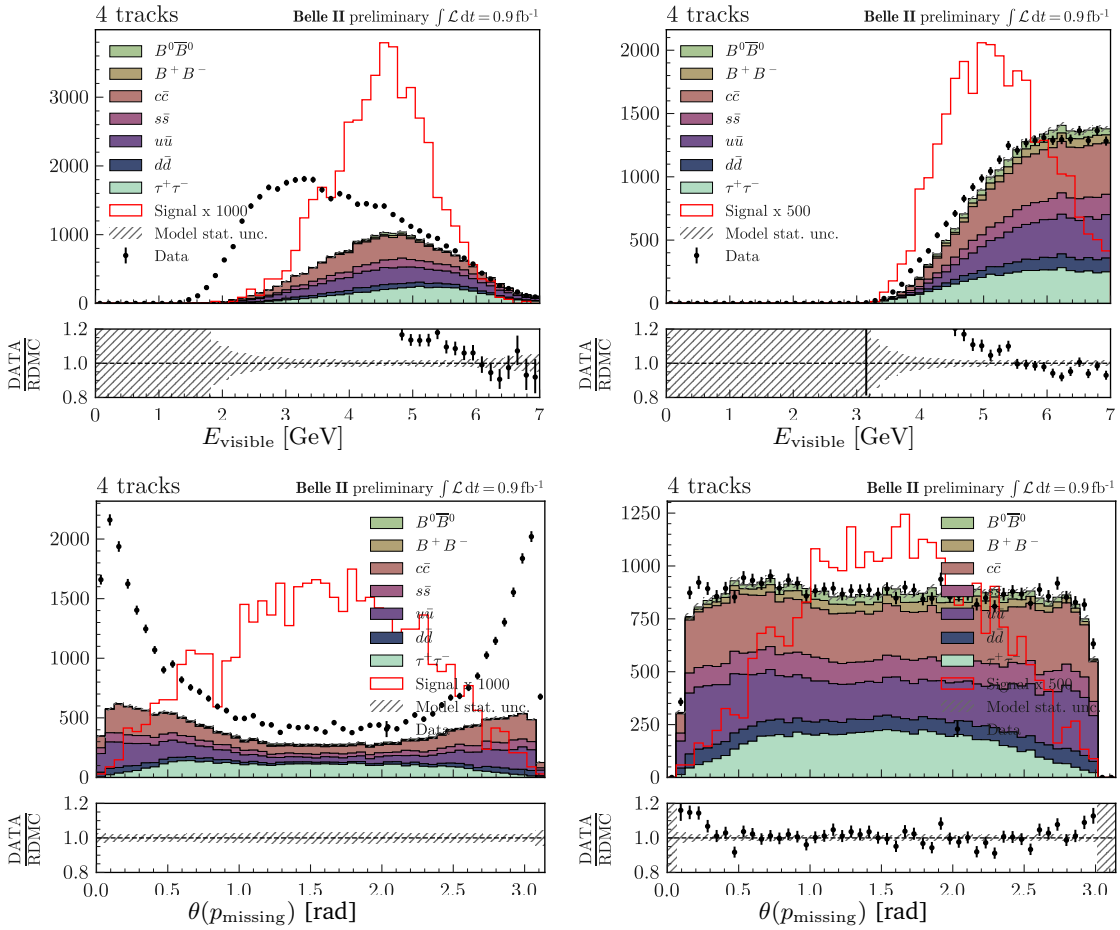

Figure 9.12: Distributions of the visible energy in c.m. frame E_{visible} and $\theta(p_{\text{missing}})$ of event with four tracks in an event and M_{ROE} <3.5 GeV/c^2 in data (points with error bars) and simulation (stacked filled histograms) for $B^0\bar{B}^0$, B^+B^-, and the five continuum categories. The $B^0 \to K^{*0}\nu\bar{\nu}$ signal is shown as an overlaid line histogram and is not included in the stack. With the $E_{\text{visible}} > 4$ GeV and 0.3<$\theta(p_{\text{missing}})$ <2.8, the low-multiplicity contribution in data is further suppressed(right). 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity (~ 0.9 fb^{-1}) is used.

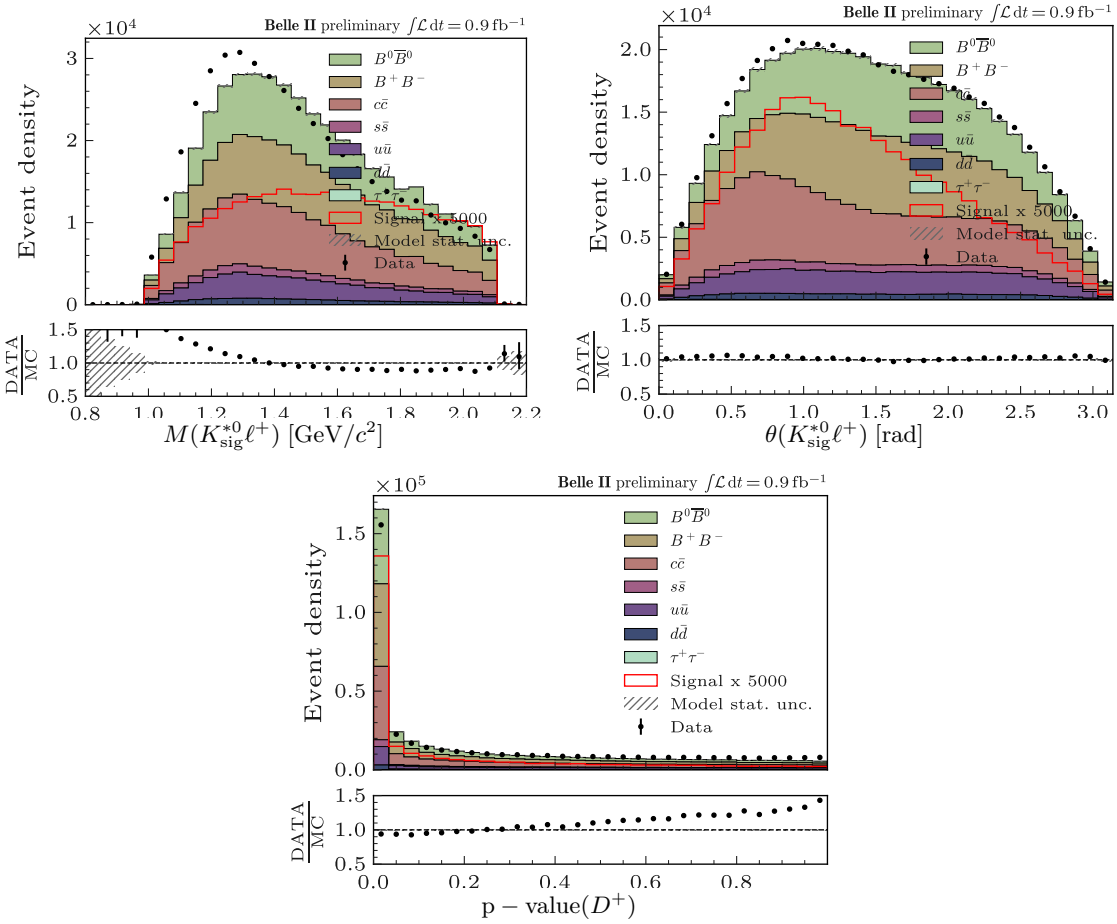

Figure 9.13: Data and simulation distribution of $M(K^{*0}_{\mathrm{sig}}\ell^+)$, θ and $\mathrm{p}-\mathrm{value}(D^+)$ in basic candidate selection level. These observables are used for the suppression of $D^+ \to K^0\ell^+\nu_\ell$ backgrounds. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used

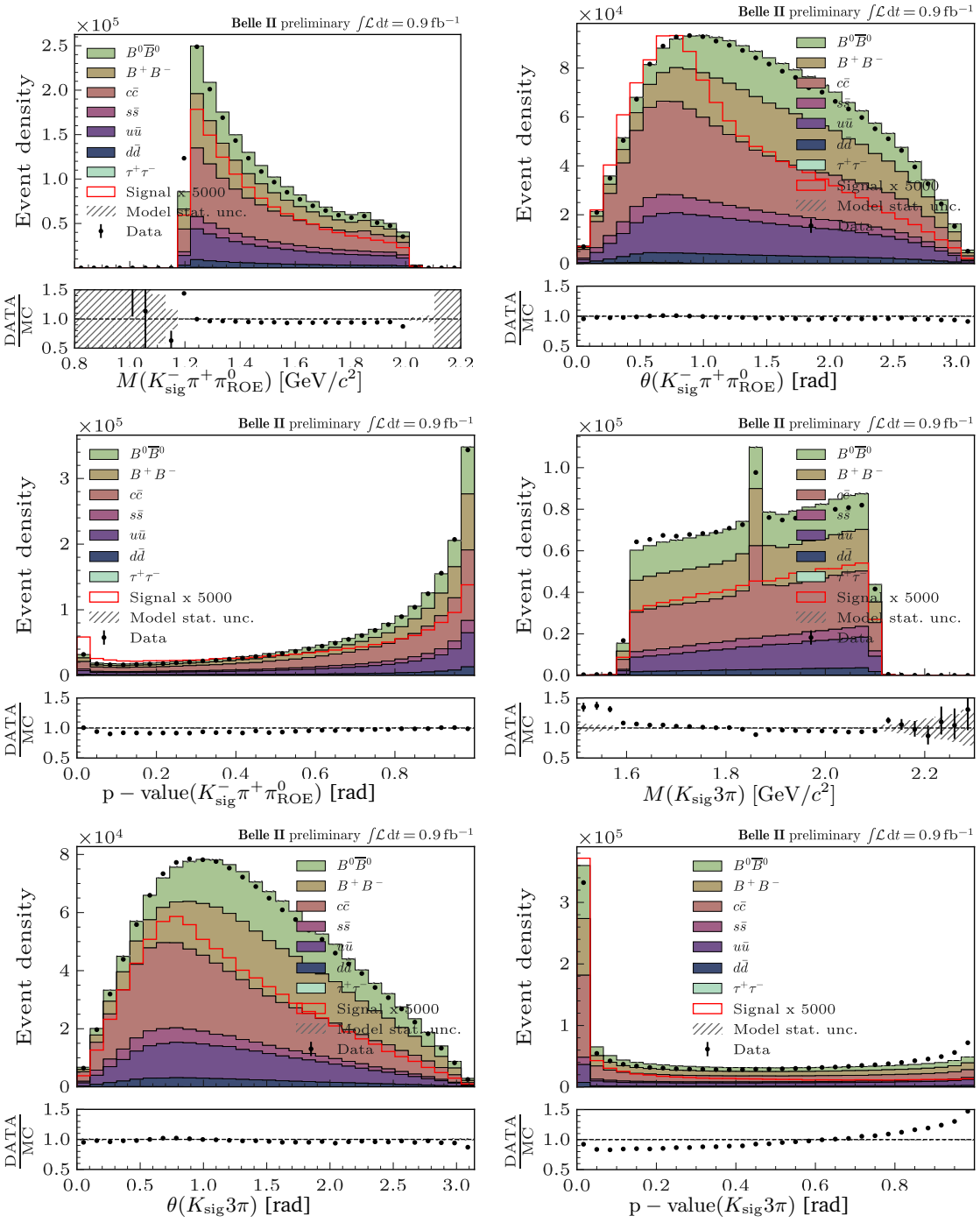

Figure 9.14: Data and simulation distribution of $M(K^{*0}_{\mathrm{sig}}\ell^+)$, θ and p − value(D^+) related to $D^0 \to K^-\pi^+\pi^0$ (upper panel and middle left panel) and $D^0 \to K^-\pi^+\pi^+\pi^-$ (middle right and bottom panel) in basic candidate selection level. 1% of the $\Upsilon(4S)$ Experiment 18 data luminosity ($\sim 0.9\ \mathrm{fb}^{-1}$) is used.