Faculté de physique et ingénierie
Université de Strasbourg

master PSA Physique Subatomique Astroparticules

Master Science, Mention Physique
Spécialité Physique Subatomique et Astroparticules

Année universitaire **2023-2024**

Merna ABUMUSABH

Search for $B^+ \to K^+ \nu\bar{\nu}$ at Belle II using Graph Neural Networks

Rapport de stage de Master
sous la direction de Giulio DUJANY et Jacopo CERASOLI

04 March 2024 au 26 July 2024

IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

Belle II

QMat QUANTUM SCIENCE AND NANOMATERIALS

Abstract

This work presents the search for the decay of a charged B-meson into a charged kaon and a pair of neutrinos via graph-tagged analysis, which is an approach based on graph neural network, in Belle II, which is operating at the asymmetric electron-positron collider SuperKEKB located at KEK, Tsukuba, Japan. $B^+ \to K^+\nu\bar{\nu}$ is a flavour changing neutral-current process that is highly suppressed in the Standard Model, and imposes many experimental challenge posed by the undetected pair of neutrinos in its final state. However, its branching fraction is accurately predicted in the Standard Model of particle physics, thus, a precise measurement of this branching fraction offers a unique opportunity to probe beyond Standard Model contributions. The expected upper limit on the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ is determined to be 0.987×10^{-5} at 90% confidence level, using simulated events.

Abstract

Ce travail présente la recherche de la désintégration d'un méson B chargé en un kaon chargé et une paire de neutrinos via une analyse basée sur un étiquetage par graphe, qui est une approche fondée sur les réseaux de neurones graphiques, dans l'expérience Belle II, qui fonctionne au collisionneur électron-positron asymétrique SuperKEKB situé à KEK, Tsukuba, Japon. Le processus $B^+ \to K^+\nu\bar{\nu}$ est une transition de courant neutre à changement de saveur, fortement supprimée dans le Modèle Standard, et pose de nombreux défis expérimentaux en raison de la paire de neutrinos non détectée dans l'état final. Cependant, sa fraction d'embranchement est précisément prédite dans le Modèle Standard de la physique des particules, et une mesure précise de cette fraction d'embranchement offre une opportunité unique de sonder des contributions au-delà du Modèle Standard. La limite supérieure attendue sur la fraction d'embranchement de $B^+ \to K^+\nu\bar{\nu}$ est déterminée à 0.987×10^{-5} avec un niveau de confiance de 90%, à partir d'événements simulés.

Acknowledgments

I am deeply grateful to my esteemed supervisor, Dr. Giulio Dujany, whose unwavering guidance and support have been invaluable pillars throughout my internship. His encouragement have not only shaped my professional growth but also enriched my personal development in profound ways.

I extend my heartfelt thanks to Dr. Jacopo Cerasoli, whose patience, tireless assistance, and willingness to address my queries with utmost care have been a source of immense support. His light-hearted spirit, even when explaining the minutest details repeatedly, has made daunting tasks seem manageable and fostered a sense of ease in my learning journey.

Furthermore, my gratitude extends to the entire BELLE II team, whose inclusive and encouraging environment has made each day of my internship a joyous and fulfilling experience.

The biggest thank is for my family, my mom, my dad, and my siblings who always surround me with their love, support and prayers, and my grandma's soul in heavens, who is watching me now.

To my fellow Gazans, from the moment we draw our first breath behind the bars, until our last breath beneath the rubble, your resilience and courage inspire me every day. May we metamorphose into blossoms, meandering rivers, polite turtles, or perhaps even into a gentle breeze - anything that melds seamlessly with the rhythm of a nature that refused to rhyme with us. May we find harmony in our divergent perspectives, may we agree on the core of life, and disagree on the things that only keep us from sleeping at night out of pondering or exhilaration, not out of pain or sorrow.

Contents

Introduction

The Standard Model (SM) stands as the cornerstone of modern particle physics, offering a remarkably successful framework for understanding the fundamental building blocks of the universe and their interactions. Over decades of experimental scrutiny, the SM has consistently proven its worth by accurately predicting and explaining numerous phenomena observed in particle collisions. However, beneath its impressive veneer lie intriguing puzzles and unresolved mysteries that continue to challenge our understanding.

One such enigma is the matter-antimatter asymmetry, a profound imbalance in the universe's composition that remains unexplained within the confines of the SM. This asymmetry, intricately linked to the magnitude of CP violation in the quark sector, underscores the need for a deeper understanding of fundamental particle interactions [1].

Therefore, motivations to extend the SM in order to address those questions are plentiful. In addition to the direct methods of searching for new physics, indirect searches can be performed, which consist in probing beyond-the-standard-model (BSM) physics by parameterizing the effect of new, unknown phenomena using effective theories.

The most important indirect constraints come from flavor physics observables and electroweak precision tests [2]. Flavor-changing-neutral-currents (FCNC) are excellent probes for BSM physics. FCNC processes are a class of particle interactions where a quark changes flavor without altering its electric charge. In the SM, FCNC processes occur only at higher orders in perturbation theory, making them highly suppressed. This suppression arises from the structure of the CKM (Cabibbo-Kobayashi-Maskawa) matrix, which governs the transitions between different quark flavors. The CKM matrix encodes the strength of flavor-changing transitions and introduces a hierarchy of suppression factors, leading to rare FCNC interactions within the SM [3].

Due to their rarity within the SM, FCNC processes are sensitive probes for physics beyond the Standard Model. The presence of new particles, such as those predicted by theories like supersymmetry or models addressing the hierarchy problem, can enhance FCNC's branching fractions. Therefore, studying FCNC processes provides valuable insights into the underlying structure of particle interactions and can help identify signatures of new physics.

Experimental investigations of FCNC processes, particularly in rare B meson decays, offer a rich testing ground for exploring deviations from SM predictions. These rare decays involve the transformation of B mesons, which contain bottom quarks, into final states involving different quark flavors. By precisely measuring the rates and kinematic distributions of these decays, physicists can test the predictions of the SM and search for deviations that may signal the presence of new particles or interactions beyond the SM [4].

The particular channel of interest in this work is the $B^+ \to K^+\nu\bar{\nu}$ decay. This mode is theoretically very clean due to the absence of virtual photons, resulting in a well-predicted branching fraction, so any deviation from that value would be a clear sign of BSM physics. However, studying this decay mode presents two main challenges. Firstly, it is extremely rare, meaning that the sensitivity and uncertainty of measurements are heavily influenced by the size of the data sample. Secondly, the final state involves two neutrinos, making the process almost invisible.

The BELLE II experiment [5] provides an ideal environment for studying rare decays like $B^+ \to K^+\nu\bar{\nu}$. With its world-record luminosity and highly sensitive detectors, BELLE II offers the opportunity to explore this rare process with unprecedented precision. BELLE II's capabilities synergize with the Hadronic Tag Analysis (HTA) and Inclusive Tag Analysis (ITA) approaches, which rigorously reconstruct signal B mesons and validate each other's results. This work introduces a new approch to study the decay using the Graph-based Full Event Interpretation (graFEI) framework. This method uses information from the final state particles to understand the underlying decay structure.

1 Context

1.1 Belle II Experiment

Belle II is a particle physics experiment collecting data at the SuperKEKB collider. This collider accelerates positrons to 4 GeV and electrons to 7 GeV. As a lepton-lepton collider, it avoids the pile-up and event-induced background issues seen in hadron colliders. The collider achieves a center-of-mass energy of 10.58 GeV, corresponding to the $\Upsilon(4S)$ meson, a bound state of $b\bar{b}$ quarks. This meson typically decays into B^+B^- or $B^0\bar{B}^0$. Belle II provides a clean environment ideal for studying partially invisible B decays. Due to the asymmetric acceleration of the two leptons, the produced B mesons are boosted forward in the direction of the higher energy beam.

One key feature of Belle II is its hermeticity, which is crucial for invisible particles like neutrinos, because it allows to reconstruct the Rest-of-Event (ROE) and hence, constrain their kinematics. This is managed by a series of sub-detectors, arranged from the nearest to the collision point to the furthest: the Vertex Detector (VXD), composed of the Silicon Vertex Detector (SVD) and Silicon PiXel Detector (PXD); the Central Drift Chamber (CDC); the Time Of Propagation counter (TOP) and Aerogel Ring Imaging CHerenkov (ARICH) for particle identification; the Electromagnetic CaLorimeter (ECL); and finally the K_L^0 and Muon detector (KLM). The arrangement of these detectors is illustrated in Figure 1.

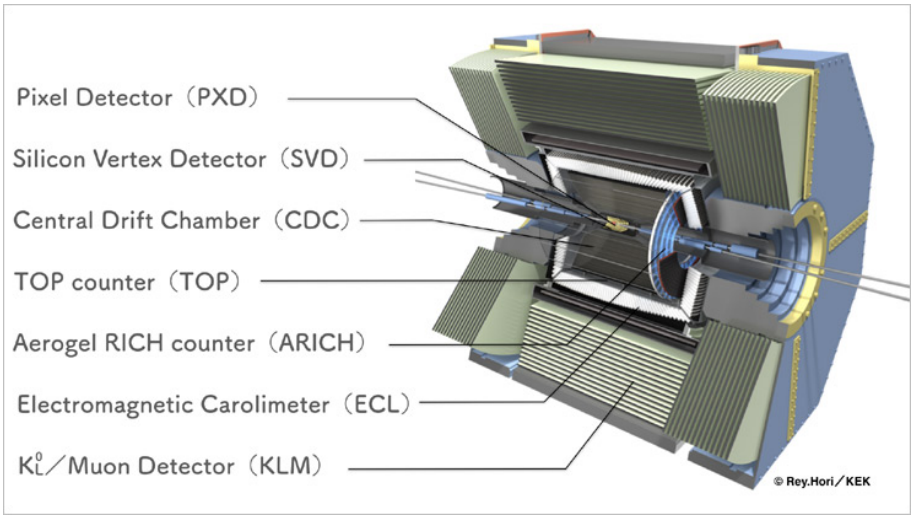


Figure 1: The Belle II detector[6].

1.2 $B^+ \to K^+\nu\bar{\nu}$

As outlined earlier, the detection challenge of this decay channel stems primarily from the presence of two neutrinos, which are undetected by Belle II. This decay does not occur at tree level in the SM and only proceeds via loop processes, as shown in Fig 2. As such, it is stringently suppressed by the Glashow-Illiopoulos-Maiani (GIM) mechanism, a fundamental aspect of the SM that prohibits certain flavor-changing processes at leading order.

Furthermore, the absence of charge in neutrinos and their inability to couple to photons makes this decay cleaner compared to similar ones involving charged leptons, such as $B \to Kl^+l^-$. Indeed, the inclusion of photon interactions in the latter process increases the theoretical uncertainties. This distinction underscores the unique characteristics of $B^+ \to K^+\nu\bar{\nu}$ and its potential significance in BSM physics research [7].

Therefore, the deviations from the Standard Model (SM) predictions can be quantified by introducing the *signal strength*, which is defined as the ratio of the measured branching fraction

of $B^+ \to K^+\nu\bar{\nu}$ and the value predicted in the SM:

$$\mu = \frac{\mathcal{B}(B^+ \to K^+\nu\bar{\nu})}{\mathcal{B}(B^+ \to K^+\nu\bar{\nu})_{\rm SM}} \tag{1}$$

where $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})_{\rm SM} = 5.56 \times 10^{-6}$ from Ref [8] is known with a relative uncertainty of 8%, which includes a contribution from loop diagrams of about 4.97×10^{-6}, and a 0.61×10^{-6} contribution from long-distance double-charged-current $B^+ \to \tau^+(\to K^+\bar{\nu})\nu$ decay. The SM expectation for the $B^+ \to K^+\nu\bar{\nu}$) branching fraction used as a reference is the one from the loop contributions, and the contribution from τ decays is excluded.

Figure 2: Leading orders of the $B^+ \to K^+\nu\bar{\nu}$ decay in the SM.

1.3 Recoil B-meson reconstruction

B-meson decays involving one or more neutrinos typically provide limited or no kinematic constraints, which are crucial for distinguishing these events from the main backgrounds [9]:

- Continum backgrounds: $e^+e^- \to q\bar{q}$, where $q \in u, d, s, c$
- On-resonanace backgrounds[1]: B^+B^- and $B^0\bar{B}^0$, referred to as *charged* and *mixed* backgrounds.

The measurement of rare signal decays such as $B^- \to K^-\nu\nu$, which include invisible particles like neutrinos, presents significant challenges due to their lack of detectable signatures in the detector. As a result, information about the full B meson decays is incomplete. To address this, the decay of the counterpart B meson (the tag-side) in the $\Upsilon(4S) \to B^+B^-$ event is reconstructed, as depicted in Fig 3. This approach, known as *recoil B-meson reconstruction*, leverages the full event information by combining observations from both the signal and tag sides of the $\Upsilon(4S)$ decay to enhance the robustness of the analyses.

[1]The $\Upsilon(4S)$ resonance

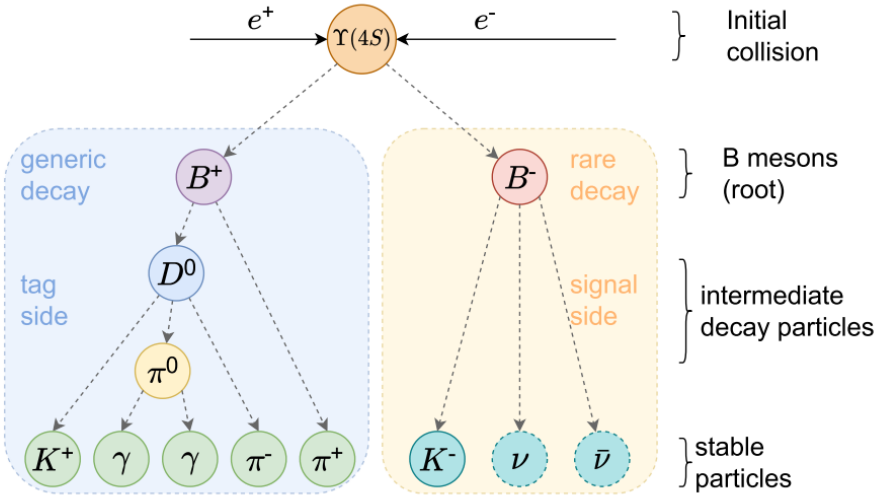


Figure 3: An electron-positron (e^--e^+) collision initiates a $\Upsilon(4S)$-event, which subsequently decays into two B-mesons, delineated as the signal-side (right) and tag-side (left) B-decays. The process is illustrated through the decay chains, capturing both intermediate and stable particles involved in the transitions.[10].

The rapid decay of B mesons into a multitude of channels, coupled with the fact that only stable final state particles (FSPs) such as electrons, muons, kaons, pions, and photons can be detected, further complicates the reconstruction process. These stable particles are crucial for reconstructing the decay sequence and inferring the properties of the invisible components of the decay. By meticulously analyzing the detected FSPs and reconstructing the decay chain, the missing information can be peiced together and the understanding of the underlying physical processes can be significantly improved.

1.4 Belle II measurements of $B^+ \to K^+\nu\bar{\nu}$

At the Belle II experiment two complementary methods were employed to analyze the decay $B^+ \to K^+\nu\bar{\nu}$ using 362fb^{-1} data collected in the period 2019-2022. The first approach, called Inclusive Tag Analysis (ITA) [11], reconstructs only the signal B meson and uses the rest of the event (ROE) to constrain the kinematics, and imposes a tagging efficiency of almost 100%, but with a low signal purity. The second method, known as Hadronic Tag Analysis (HTA) [12], follows a more conventional path to validate the ITA method, but imposes a lower tagging efficiency.

The branching fractions of the $B^+ \to K^+\nu\bar{\nu}$ found by the HTA and ITA are, respectively, $[1.1^{+0.9}_{-0.8}(\text{stat})^{+0.8}_{-0.5}\ (\text{syst})]\ \times 10^{-5}$ and $[2.7 \pm 0.5((\text{stat}) \pm 0.5(\text{syst})] \times 10^{-5}$. The HTA approach allows an the high purity but low tagging efficiency while the ITA has an high tagging efficiency but low purity. A new tagging strategy with at the same time a high efficiency and a high purity is sought for. This objective will be addressed through the implementation of the Graph-tagged Analysis (GTA), which is performed via the graFEI.

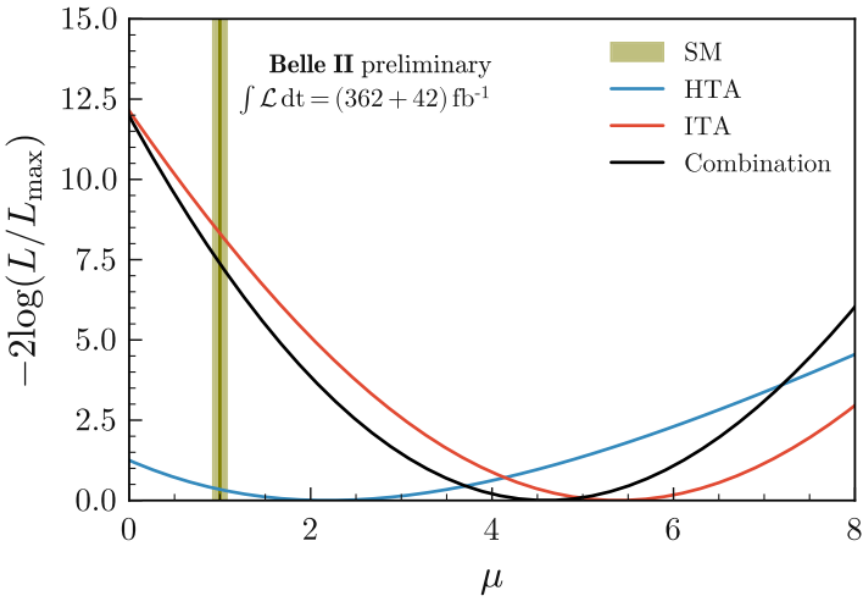


Figure 4: Twice the negative profile log-likelihood ratio as a function of the signal strength μ for the ITA, HTA, and the combined result [13].

2 Methods

2.1 The Graph-based Full Event Interpretation Model (graFEI)

The *Graph-based Full Event Interpretation* (graFEI) is a machine learning model devised to inclusively reconstruct events at Belle II by solely utilizing information on the final state particles, without presupposing any underlying decay chain structure [14].
This is efficiently accomplished using graph neural networks (GNNs), a specialized class of neural networks designed to operate on graph data structures. Graphs, denoted as G, consist of nodes $V = \{v_i\}_{i=1}^{N}$ and edges $E = \{e_{ij} | i \neq j\}$, which connect these nodes.
In graFEI, each event is conceptualized as a graph where nodes represent detected particles and edges denote potential interactions or relationships between these particles.

Decay trees are represented as rooted directed acyclic graph structures. The adjacency matrix A is one method to algebraically represent graphs, where elements of A indicate the presence of an edge between nodes, reflecting the direct connections within the decay tree. However, since only final state particles are detectable in particle physics experiments, the number of intermediate particles is unknown, making the direct reconstruction of A impossible. This challenge is addressed by employing the Lowest Common Ancestor (LCA) matrix, adapted for particle physics as the LCAS matrix, where 'S' stands for "stage" [15] [10]. The LCAS matrix elements identify the nearest common ancestor shared by any two particles, facilitating the reconstruction of the decay sequence from only the final state particles.

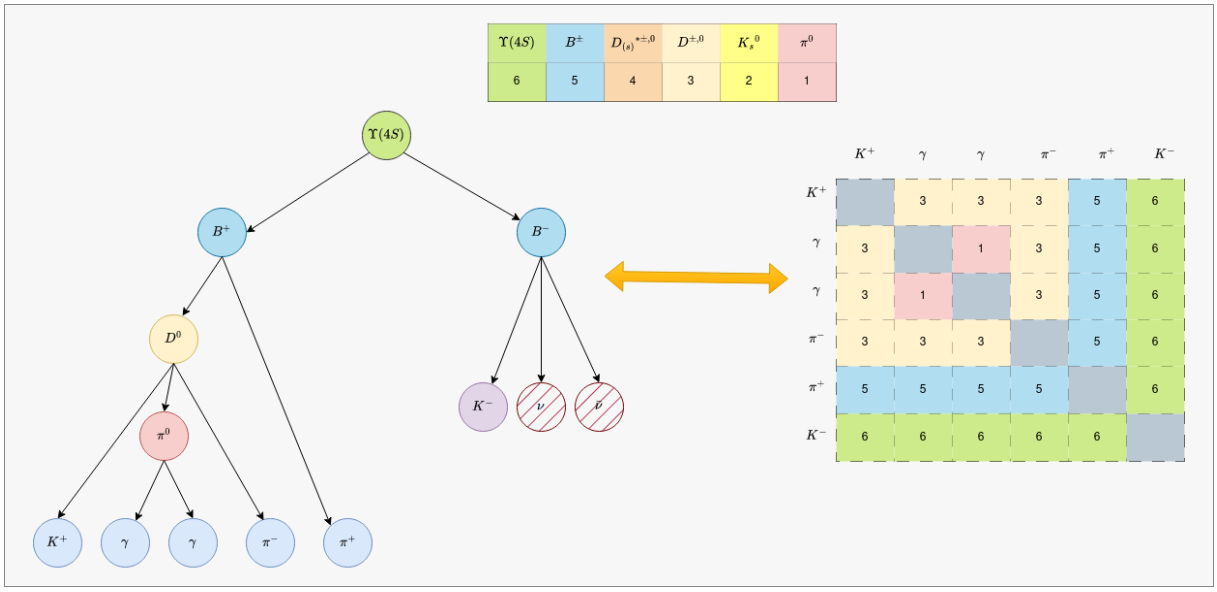


Figure 5: Example of a reconstructed decay tree with the corresponding LCAS matrix, demonstrating how the matrix encodes relationships between final state particles.

The LCAS matrix offers a streamlined and interpretable framework to represent complex particle interactions and decay processes. It enables the GNN to learn and predict the masses of final state particles and their relationships, structured within the LCAS matrix. As shown in Figure 5, each ancestor is identified in the LCAS matrix: 6 for $\Upsilon(4S)$, 5 for $B^{\pm}$, 4 for D^{*}, 3 for $D^{\pm,0}$, 2 for ${K^0}_s$ and 1 for π^0. The processing steps include:

- The input graph is a fully-connected one, meaning each final state particle share an edge with all the others. nodes contain particle information (particle charge, `pT, dr, dz, electronID, protonID, muonID, kaonID`, while edges containg information related to pairs of particles (the angle between particles' momenta ($\cos\theta$), the distance of closest approach between two tracks (`DOCA`)).
- The graph is passed through the network, where dedicated neural networks update the node and edge features
- The output graph is a fully connected graph with same structure as the input one but updated features. The output features are used to predict the LCAS elements and the mass hypotheses of final state particles.

2.2 Binary classification

2.2.1 Gradient-boosted decision tree

Single decision trees [16] are widely used in machine learning for their simplicity and interpretability. They work by recursively splitting the dataset based on feature values to create a tree-like model of decisions, which are in this case, the event being a signal or a background. However, single decision trees often suffer from high variance, leading to overfitting, or high bias, leading to underfitting. These limitations make them less robust and less accurate on complex datasets.

Boosted Decision Trees (BDTs) address these issues by forming an ensemble of multiple decision trees, each contributing to the final prediction. Unlike single decision trees, BDTs iteratively correct the errors of the previous trees, thereby improving overall classification and

regression performance. This ensemble method reduces both variance and bias, resulting in better generalization on unseen data.

For a given observation $x \in \mathbb{R}^{N_v}$ and a set of N_t decision trees, each tree assigns a weight $w_i(x) \in \mathbb{R}$ to x. The weights from all trees in the ensemble are summed to define a global weight $W(x)$:

$$W(x) = \sum_{i=1}^{N_t} w_i(x), \tag{2}$$

with an associated global prediction probability $\hat{y}_g$ given by:

$$\hat{y}_g = P(W(x)) = \frac{1}{1 + e^{-W(x)}} \tag{3}$$

2.2.2 Feature Importance

Feature importance is a crucial aspect of model interpretability in gradient-boosted decision trees. It quantifies the contribution of each feature to the predictive performance of the model. One common method to measure feature importance is by calculating the gain, which is the improvement in accuracy brought by a feature to the branches it is involved in.

The importance of a given variable v can then be defined as the sum of the gains across all nodes featuring v, normalized by the total gain:

$$I(v) = \frac{\sum_{i \in S_0} \text{Gain}(i)}{\sum_{j \in S} \text{Gain}(j)}, \tag{4}$$

where $I(v)$ is the relative importance of v, S_0 is the set of nodes featuring v, and S is the set of all nodes present in the tree.

This measure of feature importance allows practitioners to identify and focus on the most significant features in their data, leading to more efficient and interpretable predictions.

2.3 Statistical Methods

To estimate the signal sensitivity μ, defined in equation 2.1, a binned maximum-likelihood fit is implemented. For a set of N_b bins, where events are counted after a given selection, the expected number of events $f_b(\mu, \theta)$ in each bin is estimated from simulations for each type of contribution, including one signal sample and $n \geq 1$ background sources:

$$f_b(\mu, \theta) = \sum_{t \in \{\text{event types}\}} f_{b,t}(\mu, \theta) \tag{5}$$

where $f_{b,t}$ is the expected number of events in bin b for the event category sample t, and θ is a vector of N nuisance parameters that impact the base expectations.

It is assumed that, with $n \geq 1$ background sources, θ contains n nuisance parameters $\mu_1, \ldots, \mu_n$, and $N - n$ additional nuisance parameters such that:

$$\theta = (\mu_1, \ldots, \mu_n, \theta_{N-n}, \ldots, \theta_N)^T \tag{6}$$

The normalization parameters μ_i, $i \in \{1, \ldots, n\}$, are voluntarily named similarly to the signal strength μ, as each μ_i corresponds to a given background strength.

The likelihood of the observation of $n_1, \ldots, n_{N_b}$ data events in these bins is modeled as:

$$L(\mu, \theta | n_1, \ldots, n_{N_b}) = \frac{1}{Z} \prod_{b \in \{\text{bins}\}} \frac{e^{-f_b(\mu,\theta)} (f_b(\mu, \theta))^{n_b}}{n_b!} p(\theta) \tag{7}$$

where Z is a normalization factor, $\frac{e^{-f_b(\mu,\theta)}(f_b(\mu,\theta))^{n_b}}{n_b!}$ represents the Poisson probability of observing n_b events given the expected $f_b(\mu,\theta)$, and $p(\theta)$ is the prior probability given to the different nuisance parameters.

The prior probability $p(\theta)$ is composed of several Gaussian densities, centered at unity for normalization variations and at zero for additive variations:

$$p(\theta) = \prod_{i=1}^{n} \text{Gauss}(\theta_i|1, \sigma^2_{\text{norm},i}) \prod_{j=N-n+1}^{N} \text{Gauss}(\theta_j|0,1) \tag{8}$$

The background normalization uncertainties $\sigma_{\text{norm},i}$ and the factors $\delta_i^{b,t}$ are inputs to the model, describing the systematic uncertainties.

Finally, μ is extracted from the MC events by the maximization of the likelihood function $L(\mu,\theta|n_1,\ldots,n_{N_b})$.

2.3.1 Upper Limit Determination

Assuming a specified value of μ, the likelihood ratio, $\lambda(\mu)$, is defined by the following equation:

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}}|n_1,\ldots,n_{N_b})}{L(\hat{\mu},\hat{\theta}|n_1,\ldots,n_{N_b})} \tag{9}$$

where μ represents the signal strength and θ are the nuisance parameters. The parameters $(\hat{\mu},\hat{\theta})$ are those that maximize the likelihood across the observed data set, and $\hat{\hat{\theta}}$ are the parameters that maximize the likelihood for a fixed μ.

The likelihood ratio test statistic, Λ_μ, is then given by:

$$\Lambda_\mu = -2\ln\lambda(\mu), \tag{10}$$

where Λ_μ corresponds to the square of the sensitivity. The factor of -2 is introduced to scale the logarithmic likelihood ratio, ensuring that under the null hypothesis, the test statistic Λ_μ follows a chi-squared distribution, as per Wilks' Theorem [17]. This transformation allows for using the chi-squared distribution to determine statistical significance, facilitating comparisons of the observed data against expected distributions under the null hypothesis.

An upper limit on μ at a specified confidence level (CL) is determined by finding the μ value that satisfies the following condition:

$$\Lambda_\mu = \chi^2_{\text{CDF}^{-1}}(C), \tag{11}$$

where C represents the required confidence level, and $\chi^2_{\text{CDF}^{-1}}$ is the inverse cumulative distribution function of the chi-squared distribution.

Assuming that the value of μ that satisfies the equation above is μ^*, the upper limit is expressed as:

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) < 4.49 \times 10^{-6} \times \mu^* \tag{12}$$

The binned maximum likelihood fitting and the determination of the upper limit are conducted using the `pyhf` package [18].

3 Results

3.1 Input Simulated Samples

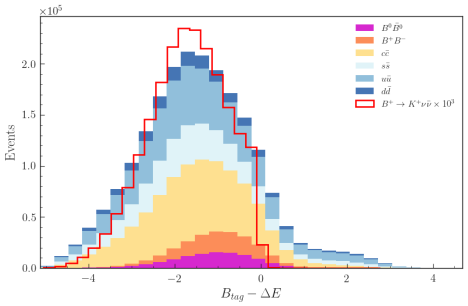
Simulated $B^+ \to K^+\nu\bar{\nu}$ signal and background samples are generated in Belle II reflecting data-taking:

- 98.7×10^6 events of the signal $B^+ \to K^+ \nu\bar{\nu}$.
- 211.7×10^6 continuum background corresponding to 1444 fb^{-1} of luminosity.
- 28.2×10^6 $B\bar{B}$ background of mixed and charged events, corresponding to 1444 fb^{-1} of luminosity.

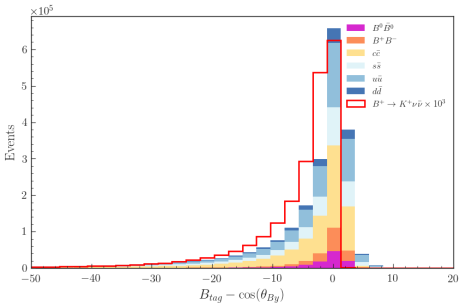
3.2 Events variables

The signal is normalized to the SM branching fraction, while the backgrounds are normalized to the BELLE II luminosity of 362 fb^{-1}. Some of the event variables used in this analysis are as follows:

- kinematics:
  - The difference between the energy of the tagged side and half the center of mass energy
  - Cosine the angle in Center-of Mass (CMS) frame between momentum the particle and the tag-side B .



(a)

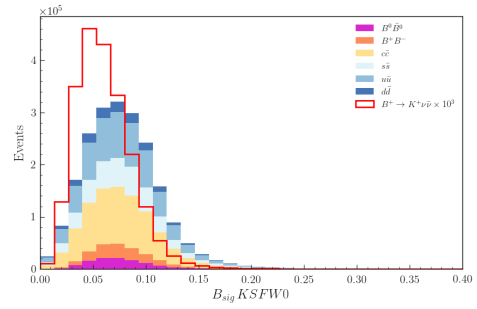


(b)

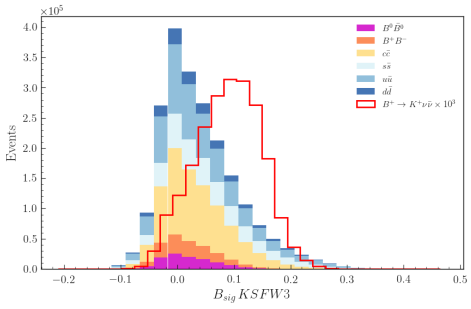
- Event shape variables: Such as the modified Fox-Wolfram moments, which describe how compatible the event topology is with each of the spherical harmonics [19].
  The signal-ROE (so) modified Fox-Wolfram moment of degree l for the a particle is

$$H_{xl}^{s0} = \frac{1}{Z} \sum_{i=1}^{N_s} \sum_{j_x=1}^{N_x} C_{ij_x}^{l} p_{j_x} P_l(\cos \alpha_{ij_x}) \tag{13}$$

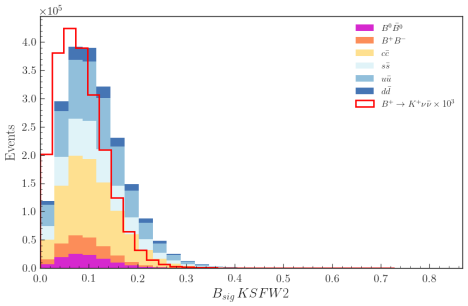
  - Z, a normalization factor, verifying $Z = 2(\sqrt{s} - E_B^*)$, where $\sqrt{s}$ is the available energy in the center-of-mass frame and E_B^* is the energy of the signal B-meson candidate in the center-of-mass frame.
  - $C_l^{ijx} \in \{-1, 0, 1\}$, the product of the charges for the candidates i and jx if l is odd; $C_l^{ijx} = 1$ if l is even.
  - P_l, the Legendre polynomial of l-th order.
  - α_{ijx}, the angle between the momenta p_i and p_{jx}.



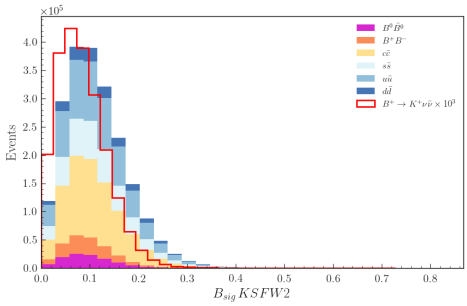
(a)



(b)



(c)

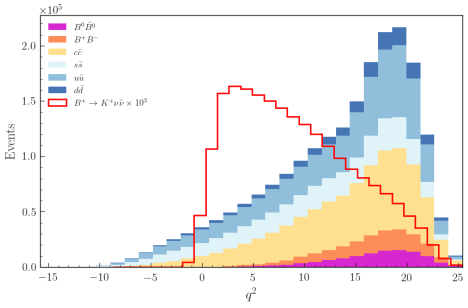


(d)

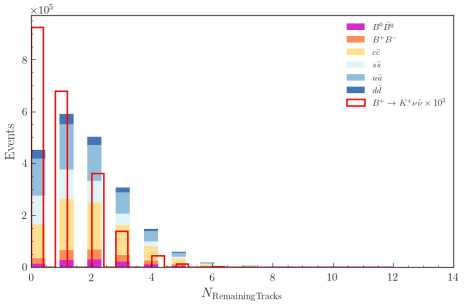
Figure 7

others:

- The mass squared of the neutrino pair.
- The number of extra tracks in the event.



(a)



(b)

The full set of the variables used in the analysis, their descriptions and distributions are listed in Appendix A.

3.3 Event classification

3.3.1 GraFEI reconstruction requirement

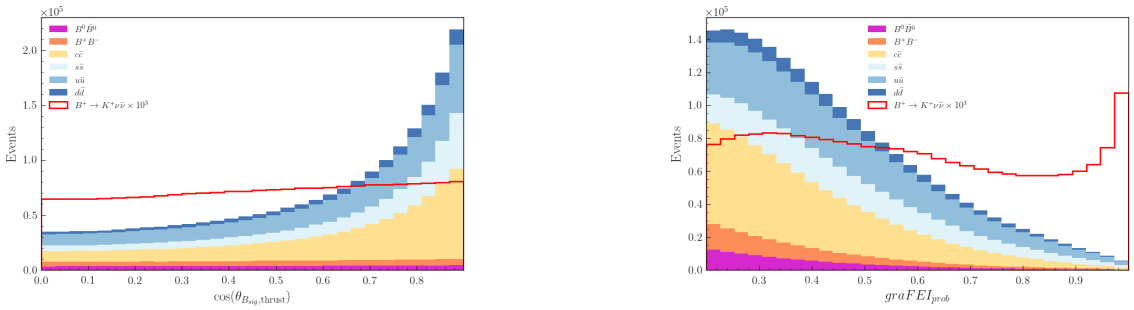
The LCAS matrix is required to contain only the signal kaon and the tag B meson. The signal reconstruction efficiency ϵ^{Reco} is found to be 27.29% [2].

[2]Performed by the IPHC Belle II team

3.3.2 Pre-training cuts

The pre-training cuts are applied on the most glaring backgrounds, in order to enable the subsequent training to concentrate on rejecting more difficult backgrounds. These cuts are applied on the two following variables:

- The likelihood of the signal-side final state particle being a kaon is > 0.9.
- The cosine of the angle between thrust axis (axis along which the projection of a collection of momenta is maximized) if the signal B and the thrust axis of the Rest of Event (ROE) < 0.9.
- The graFEI assigns a probability to each LCAS element that it predicts. The geometric mean of all those probabilities is required to be > 0.2.



(a) (b)

Figure 9

The pre-training efficiency $\epsilon^{\text{Pre-train}}$ is found to be 53.61%.

3.3.3 Classifier training

Events are classified according to their similarity to the expected signal using a BDT implemented with XGBoost [20]. In order to achive a good performance, the variables with strong correlations in both the signal and the backgrounds were removed from the input list, and the BDT was trained with 16 variables. To further enhance the model performance, the hyperparameters were optimized using `RandomizedSearchCV`, which is a method for hyperparameter tuning that searches across a specified hyperparameter space by sampling a fixed number of parameter settings from the specified distributions rather than trying all combinations. The optimiziation was tested over the following ranges:

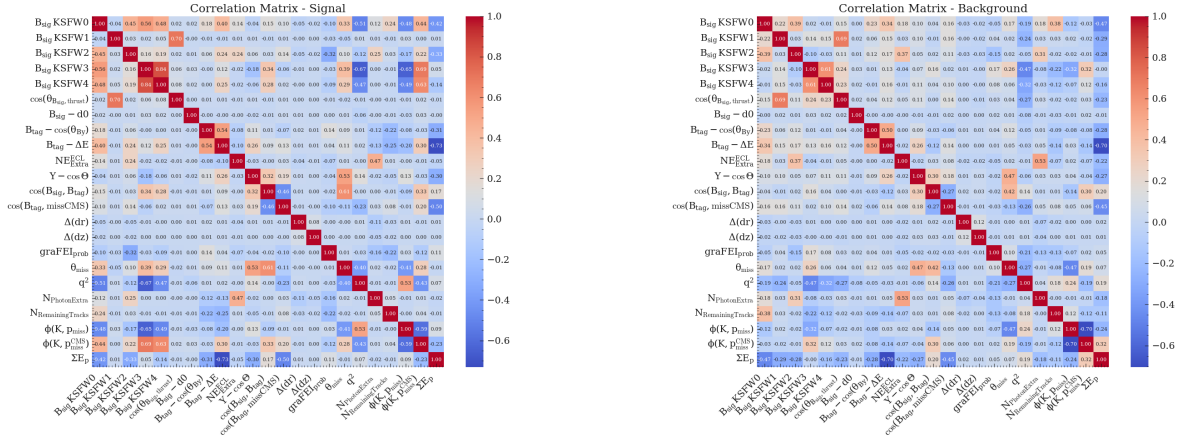
- `max_depth`: [3, 5]
- `learning_rate`:[0.01, 0.2]
- `subsample`: [0.6, 1.0]
- `n_estimators`: [200, 600]

The optimized values are listed in Table 1.

BDT parameter	Value
Tree depth	5
Learning rate	0.1547
Number of estimators	600
Subsample Ratio	0.8912

Table 1: The optimized BDT hyperparameters

The BDT is trained using 10% of the input samples, further divided into 80% for training and 20% for testing, then applied on the remaining 90% of the samples.
The tree reconstruction algorithm used in training is (`tree_method = hist`), which speeds up gradient boosting by using histogram-based algorithms to compute the best split points. Instead of scanning all possible split points, it discretizes feature values into bins and constructs histograms to efficiently find the optimal splits, reducing computational overhead and enhancing training speed.

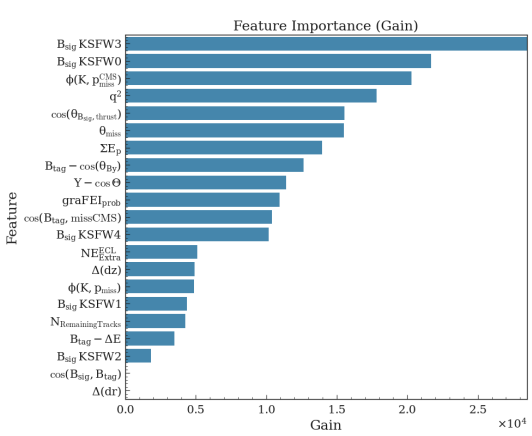


(a) (b)

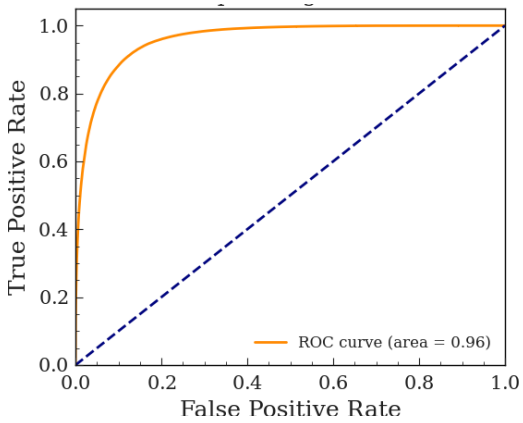
Figure 10: The matrix of linear correlation coefficents of the input variables to the BDT of the signal (left) and the backgrounds (right)

After training, the importance of each feature in classifying events can be assessed. Feature importance indicates the contribution of each feature to the model's predictive performance. Figure 11(a) illustrates the relative importance of features in the classification process.

Figure 12 showcases the BDT output distribution for both signal (S) and background (B) events. The BDT classifier effectively discriminates between signal and background, as evidenced by the separation between the two distributions across the BDT output range.



(a)



(b) Area Under the Curve (AUC) for the classifier parameters.

Figure 11

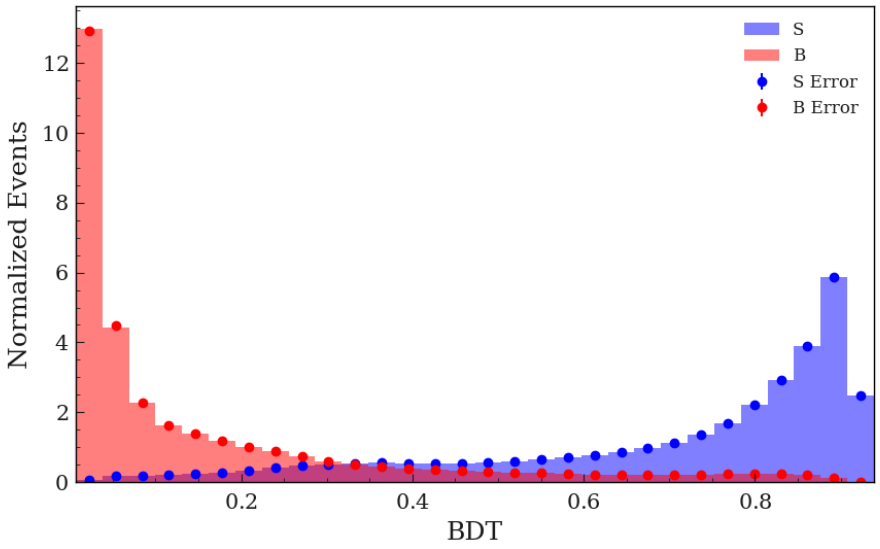


Figure 12: The classifier output of the agreement between the testing and training of the 90% sample

3.4 Signal region search

The signal region (SR) is the region on which the binned-likelihood fit will be performed, in order to get the value of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$. The region is achieved by applying a cut of BDT $>$ 0.8. This region is then divided into 6 equal bins of the classifier probability output. In order to easily treat the classifier output, it is modified so that the distribution of the $B^+ \to K^+\nu\bar{\nu}$ signal is uniform along the whole region (Fig 13), that is, for each event y, its BDT output a is transformed as follows:

$$a \to \frac{\int_{-1}^{a} BDT(y)\,dy}{\int_{-1}^{1} BDT(y)\,dy} \tag{14}$$

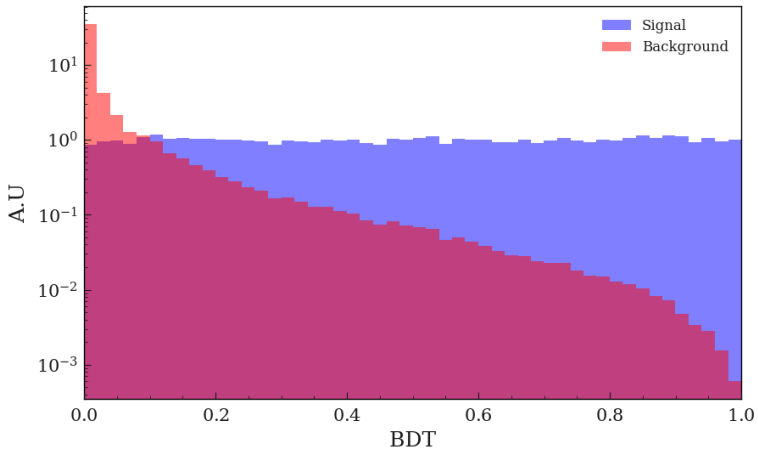


Figure 13: The flattened BDT output distributions for signal and background samples

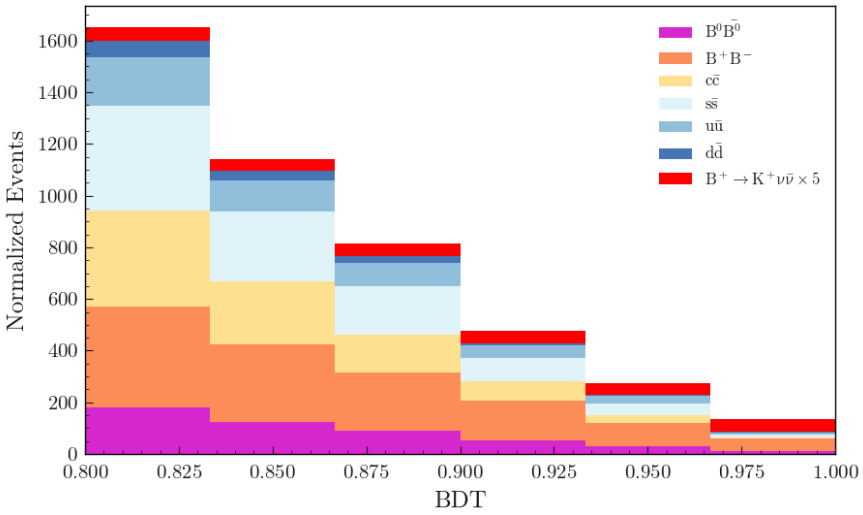


Figure 14: Distribution of events in the signal search region for the GTA

3.5 Results and Discussion

3.5.1 Efficiencies

Selection	Number of signals after the cut	$\varepsilon_{sig} \times \%$
graFEI event reconstruction	26940105	27.295 ± 0.005
Preselection cuts	14429334	53.607 ± 0.010
Signal Region	2671942	18.517 ± 0.010

Table 2: Signal selection efficiency at various steps of the selection

The total efficiency of the selection process is, therefore, the product of the following terms:

$$\epsilon^{\text{Total}} = \epsilon^{\text{Reco}} \times \epsilon^{\text{Pre-train}} \times \epsilon^{\text{Selec}} = (2.709 \pm 0.002)\% \tag{15}$$

3.5.2 Signal upper limit

Based on the likelihood $L(\mu, \theta)$, the expected upper limit on the value of $B(B^+ \to K^+\nu\bar{\nu})$ is determined. Assuming the Standard Model (SM) value for $B(B^+ \to K^+\nu\bar{\nu})$, Figure 15 presents the distribution of the likelihood-ratio test Λ_μ as defined in Equation 10. From this distribution, the expected upper limit for the GTA is:

$$B(B^+ \to K^+\nu\bar{\nu}) < 0.987 \times 10^{-5} \tag{16}$$

at a 90% confidence level. The asymmetrical error σ_{As} associated with the signal strength errors is:

$$\sigma_{\text{As}} = \sigma^+ - \sigma^- = 0.025 \tag{17}$$

where σ^+ and σ^- are the positive and negative errors, respectively:

$$\sigma^+ = \mu(2\ln\Lambda = 1) - \mu(2\ln\Lambda = \text{min})$$

$$\sigma^- = \mu(2\ln\Lambda = \text{min}) - \mu(2\ln\Lambda = 1)$$

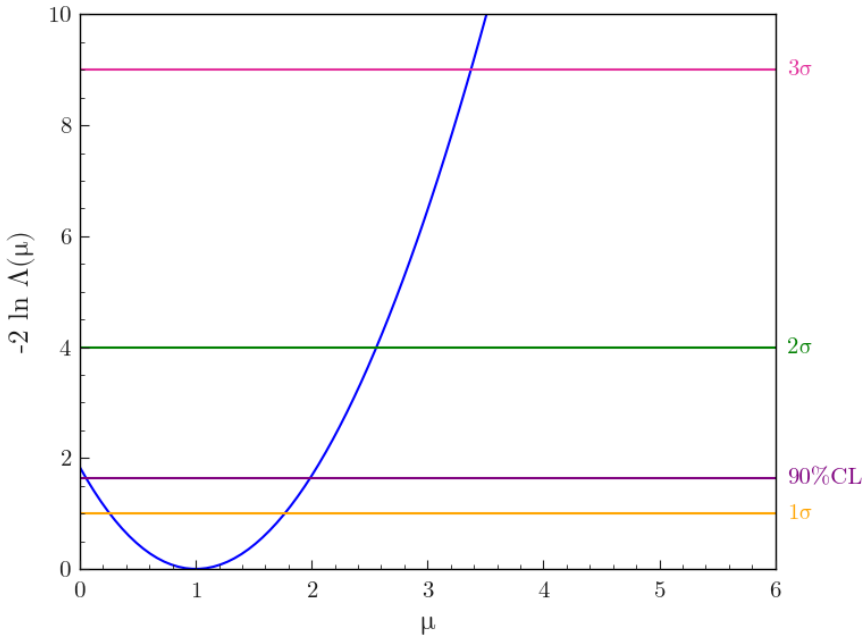


Figure 15: Twice the negative likelihood-ratio λ_μ, expecting a SM value for the $B^+ \to K^+\nu\bar{\nu}$ branching fraction.

3.5.3 Comparing with previous results

The expected upper limit on branching fraction is compared with previous measurements in Figure 16 in both the SM prediction and the HTA and ITA combined central value central value.

Table 3: SM predictions, with $\mu = 1$

	μ at 90% CL	Upper limit at 90% CL	σ^+	σ^-	$\sigma^+ - \sigma^-$
GTA	1.987	0.987×10^{-5}	0.768	0.743	0.025
HTA	2.676	1.329×10^{-5}	1.289	–	–
ITA	1.766	0.878×10^{-5}	0.597	0.589	0.008

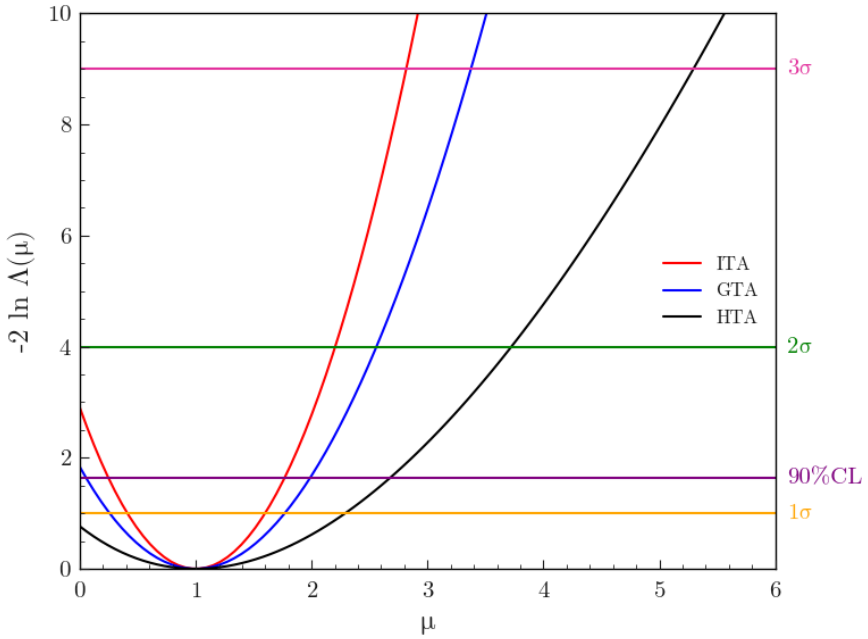


Figure 16

The branching fraction of the signal obtained with the (HTA + ITA) combined results is 2.3×10^{-5}, which corresponds to a signal strength of 4.6 [3][13].

[3]Systematic uncertainties are not being considered at this stage of the analysis

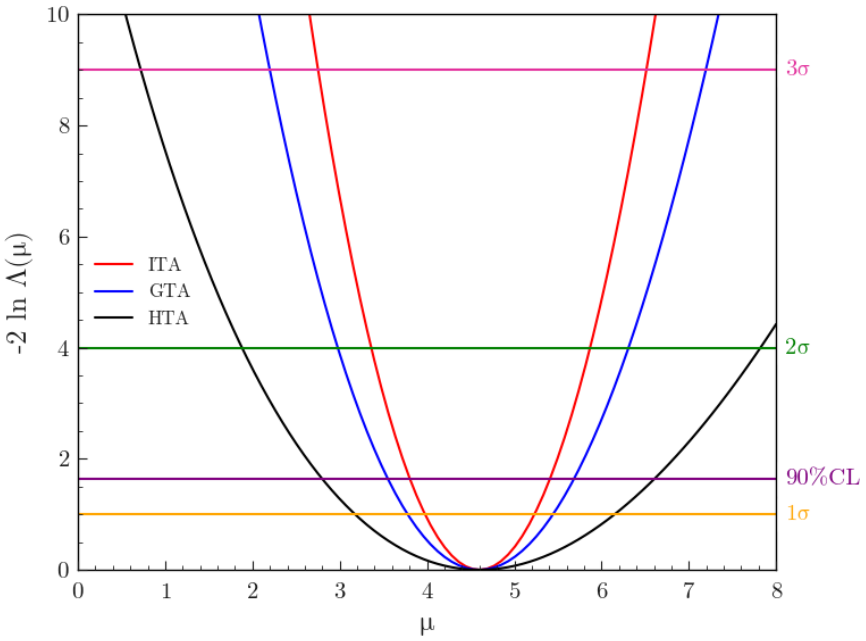


Figure 17

Table 4: HTA and ITA combined central value, $\mu = 4.6$

	μ at 90% CL	Upper limit at 90% CL	σ^+	σ^-	$\sigma^+ - \sigma^-$
GTA	5.583	2.824×10^{-5}	0.843	0.822	0.021
HTA	6.605	3.282×10^{-5}	1.549	1.424	0.125
ITA	5.407	2.687×10^{-5}	0.629	0.623	0.006

The signal purity is defined as:

$$\mathcal{P}^{\text{sig}} = \frac{\text{The number of the final signal events}}{\text{The number of the final background events}}$$

The signal purity and total efficiencies of the GTA, HTA and ITA at the SM prediction are shown in Table 5:

Table 5

	$\epsilon^{\text{Total}}(\%)$	$\mathcal{P}^{\text{sig}}(\%)$
GTA	2.7	1.4
HTA	0.4	3.5
ITA	8	0.8

The GTA demonstrates superior performance in both total efficiency and signal purity compared to both the HTA and the ITA, respectively. While the ITA shows a higher total efficiency compared to the GTA, it does so at the expense of lower signal purity. These findings suggest that the GTA offers a balanced approach in achieving robust testing outcomes, particularly when considering both efficiency and signal integrity as critical metrics in the evaluation process.

4 Conclusion

This study focuses on the search of the $B^+ \to K^+\nu\bar{\nu}$ decay using simulations corresponding to a luminosity of 1443 fb^{-1}, employing the graph-neural network approach, graFEI.

The analysis involves meticulous selection steps, including the application of Boosted Decision Trees (BDTs) to mitigate backgrounds, and rigorous evaluation of selection efficiency at various stages. The culmination of the study features a binned maximum likelihood fit on the BDT classifier output, establishing an upper limit on the branching fraction of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) < 0.987 \times 10^{-5}$ at a 90% confidence level (CL).

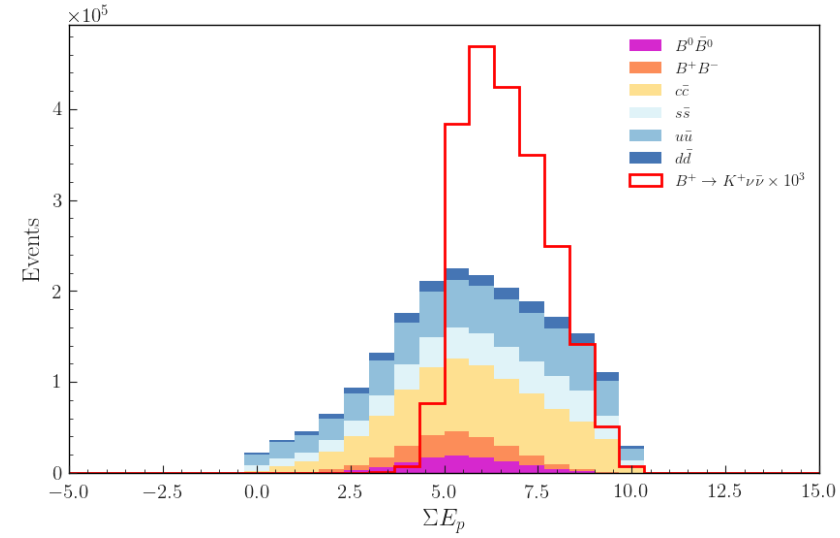
Importantly, this investigation highlights the comparative efficacy of different methodologies: the graph-neural network approach (GTA), the Hadronic Tag Analysis (HTA), and Inclusive Tag Approach (ITA). The GTA demonstrates superior performance in both total efficiency and signal purity compared to HTA and ITA. This underscores the GTA's capability to effectively discern rare signal events amidst background noise, crucial for advancing our understanding of rare decays in particle physics.

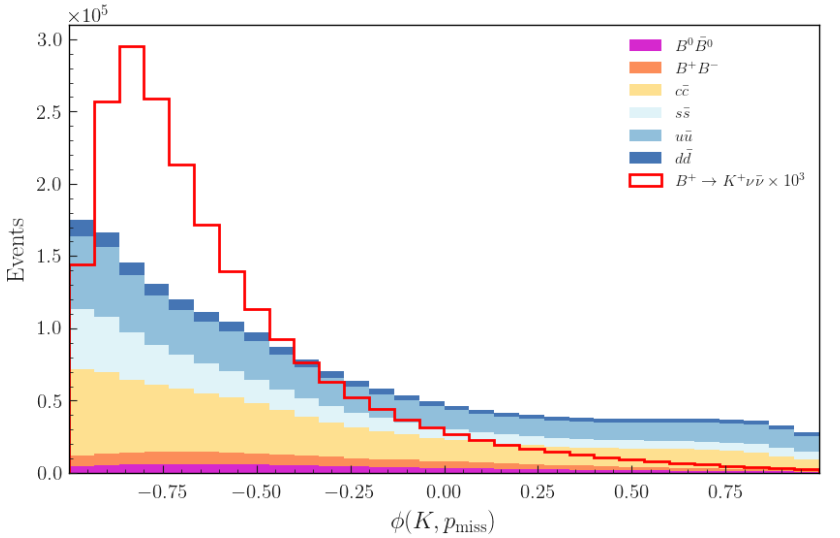
Looking ahead, ongoing refinement of selection techniques and meticulous consideration of statistical and systematic uncertainties promise to enhance sensitivity and improve upper limits in future analyses. These efforts are essential for pushing the boundaries of particle physics research, aiming to uncover new physics phenomena and contribute to the ongoing evolution of the Standard Model.

A List of Variables

Variable	Description
sumEp	Missing energy of event in the CMS + missing momentum of event in CMS
miss_theta	Angle between missing momentum and z axis
Ups_cosTheta	Angle between Upsilon total momentum and z axis
NEExtra_ROEClusters_v702	Total energy of neutral clusters in calorimeter in the rest-of-event
cos_tag_miss_CMS	Angle between Btag and missing momentum in CMS
delta_dz	
phi_K_pmiss_CMS	Angle between signal kaon and missing momentum in CM frame

Table 6: Variable descriptions


×10^5
Events
ΣE_p
B^0B̄^0
B^+B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+νν̄ × 10^3


×10^5
Events
φ(K, p_miss)
B^0B̄^0
B^+B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+νν̄ × 10^3

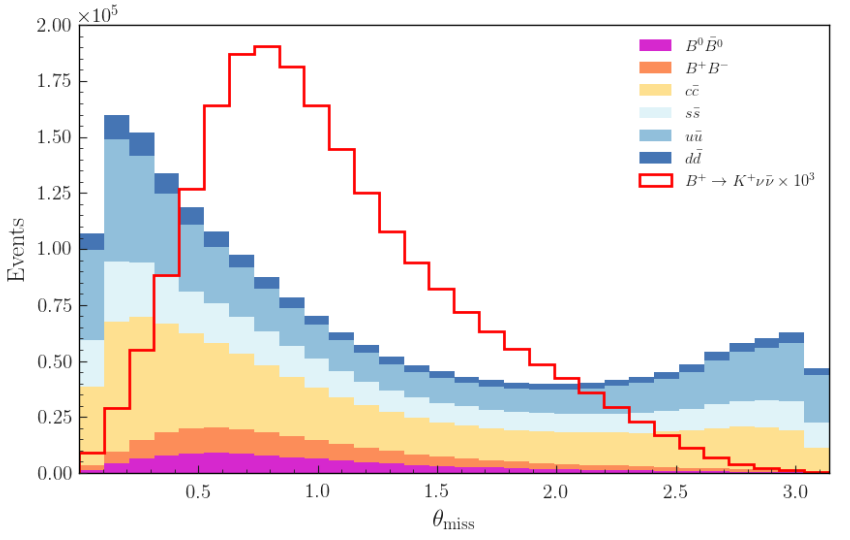
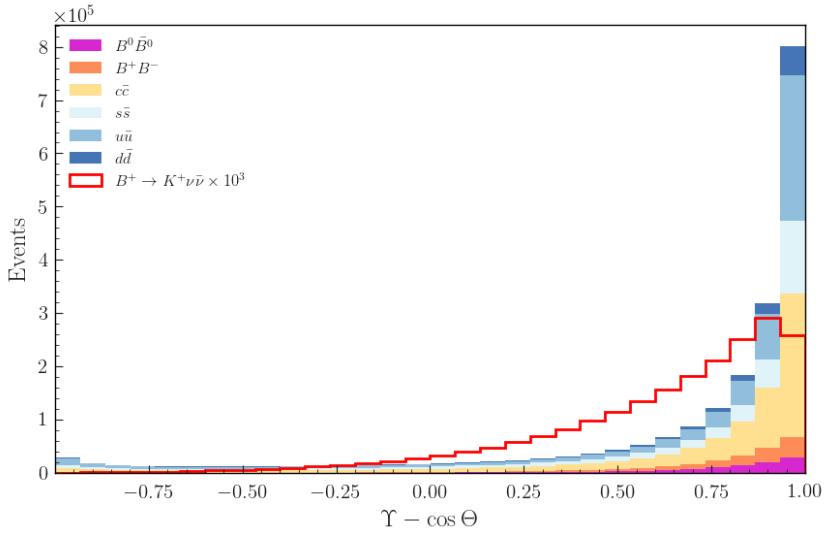
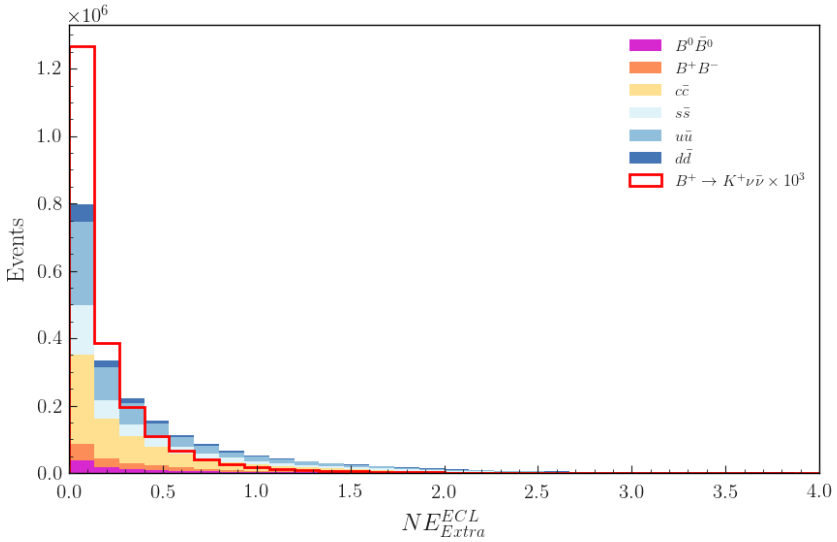

×10^5
Events
θ_miss
B^0B̄^0
B^+B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+νν̄ × 10^3

Figure 18


×10^5
Events
Υ − cos Θ
B^0 B̄^0
B^+ B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+ νν̄ × 10^3


×10^6
Events
NE^{ECL}_{Extra}
B^0 B̄^0
B^+ B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+ νν̄ × 10^3

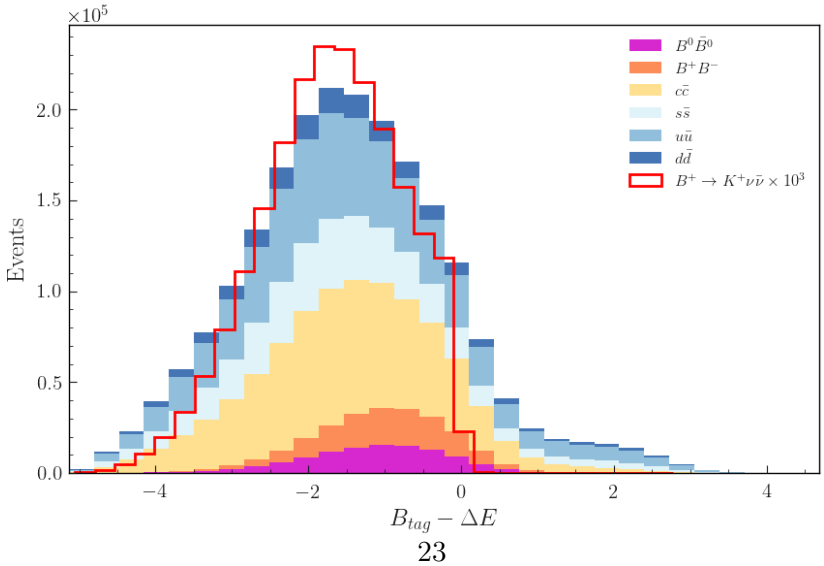
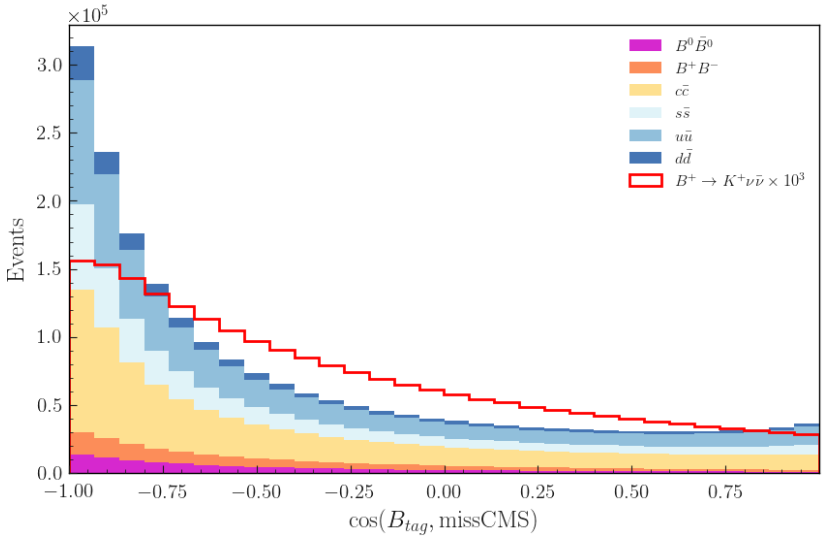
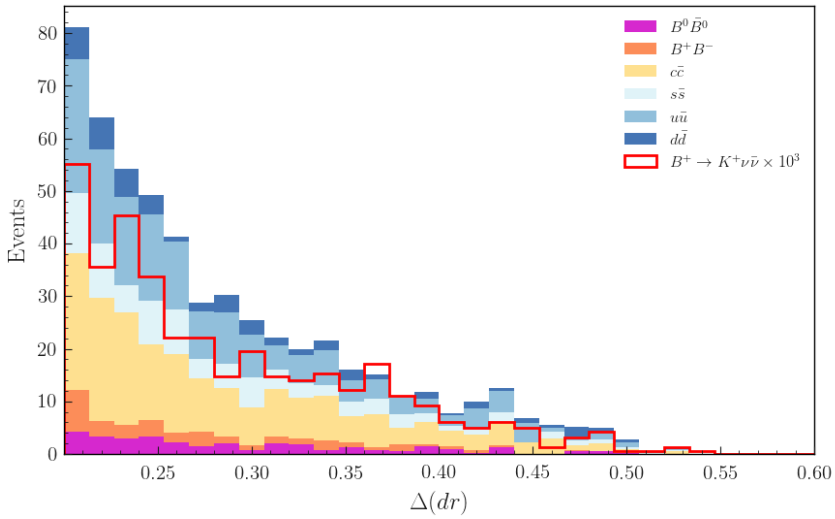

×10^5
Events
$B_{tag} - \Delta E$
B^0 B̄^0
B^+ B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+ νν̄ × 10^3

Figure 19


×10^5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Events
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
cos(B_tag, missCMS)
B^0 B̄^0
B^+ B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+ νν̄ × 10^3


80
70
60
50
40
30
20
10
0
Events
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
Δ(dr)
B^0 B̄^0
B^+ B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+ νν̄ × 10^3

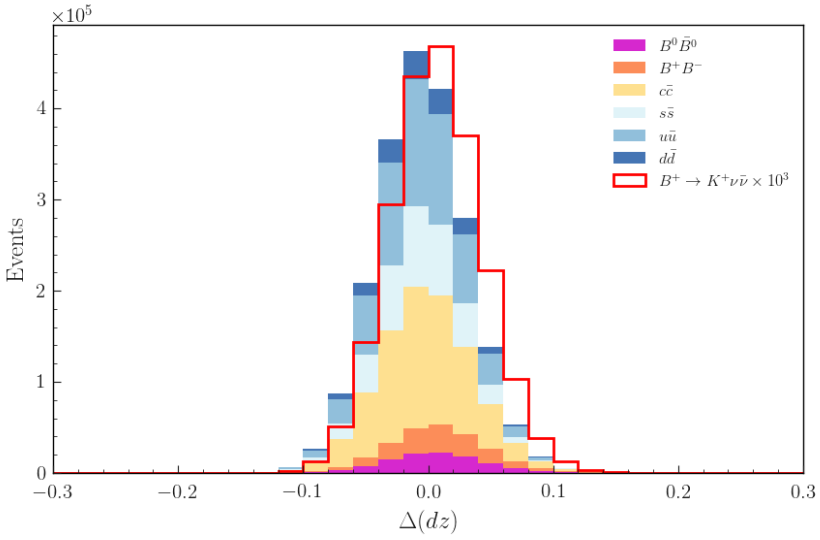
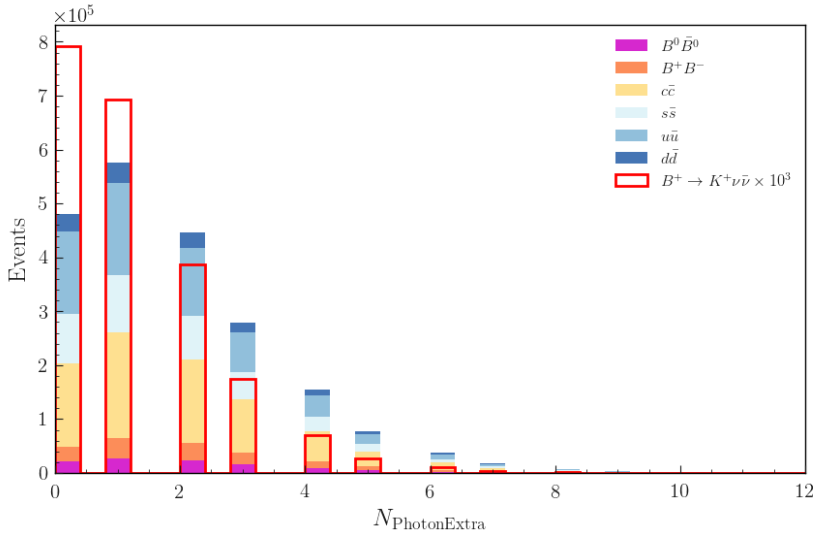
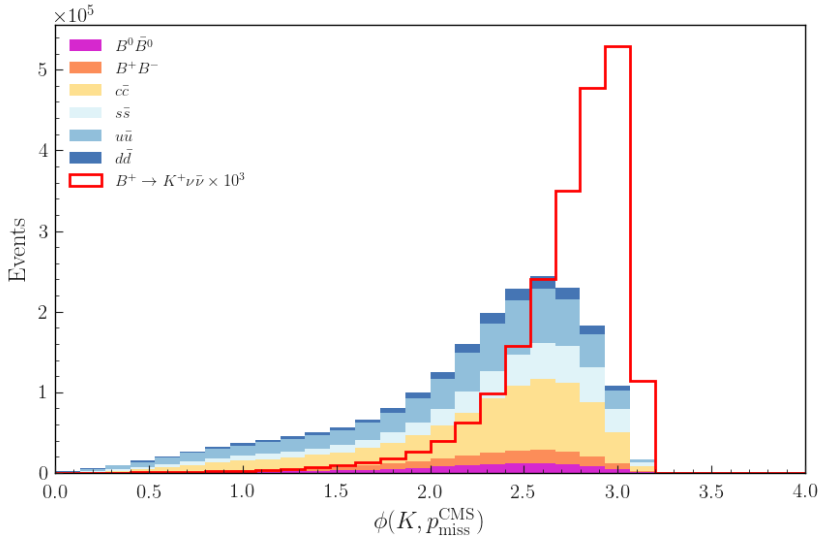

×10^5
4
3
2
1
0
Events
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
Δ(dz)
B^0 B̄^0
B^+ B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+ νν̄ × 10^3

Figure 20


×10^5
Events
N_PhotonExtra
B^0B̄^0
B^+B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+νν̄ × 10^3


×10^5
Events
φ(K, p_miss^CMS)
B^0B̄^0
B^+B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+νν̄ × 10^3

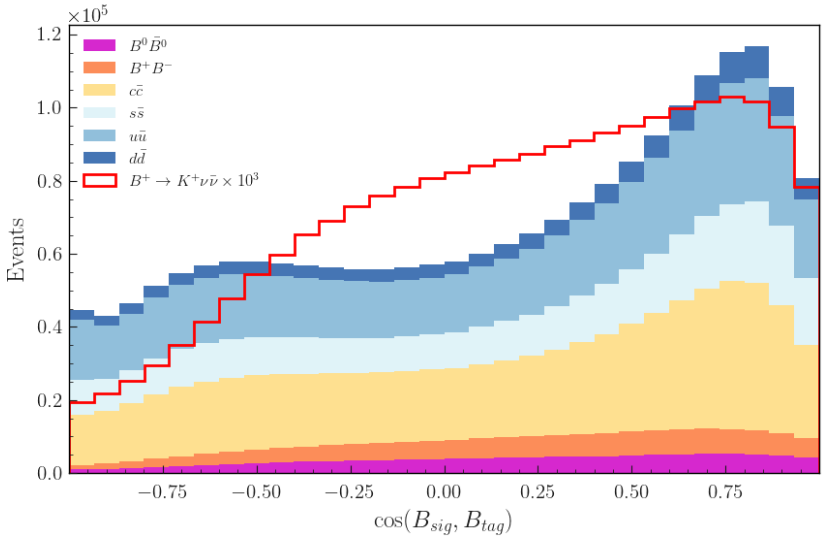

×10^5
Events
cos(B_sig, B_tag)
B^0B̄^0
B^+B^−
cc̄
ss̄
uū
dd̄
B^+ → K^+νν̄ × 10^3

Figure 21

B Selection Optimisation

The optimal BDT cut is evaluated by maximizing the Punzi Figure of Merit (FoM), defined as:

$$\text{FoM} = \frac{\varepsilon_{sig}}{\sqrt{N_{bkg}} + a/2} \tag{18}$$

Where ε_{sig} and N_{bkg} are the signal efficiency and the background yields for a given BDT cut. The parameter a is set to a = 5 for a statistical signal significance of 5σ. As shown in Fig 21, the maximum of the FoM is found at 0.9, so the cut on the events is placed at this BDT value. The BDT performance is quantified using the sensitivity ($\mathcal{S}$) at 90% CL, defined as:

$$\mathcal{S} = \frac{1.28 \times \sqrt{N_{bkg}}}{\epsilon_s \times N_{B^+}} \tag{19}$$

The factor of 1.28 arises from the upper limit of the interval $[-\infty, 1.28\sigma]$, which encompasses 90% of the area under a Gaussian distribution. The sensitivity pertains to estimating the limit on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction.

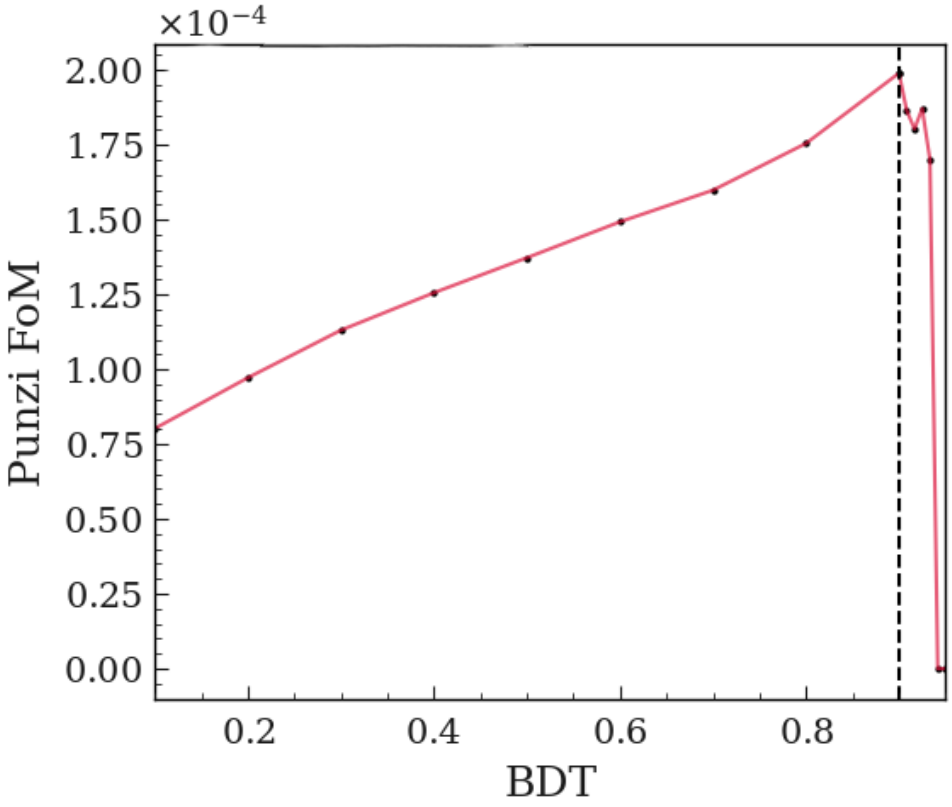


Figure 22: The dashed line corresponds to the optimal value

- Best BDT Cut: 0.90
- Estimation of the error: 27.92
- Number of Signal Events: 757886
- Signal Efficiency: 0.0023
- Upper Limit to 90% CL: 0.000085

BDT Cut	BG Events	Signals	Signal Efficiency
0.10	3198256	12886221	0.15
0.20	1924542	12605757	0.14
0.30	1227446	12102144	0.14
0.40	800815	11390872	0.13
0.50	512921	10504415	0.12
0.60	302979	9305582	0.10
0.70	148645	7587184	0.09
0.80	50657	5219835	0.06
0.90	779	757886	0.0023
0.91	314	482328	0.01
0.92	95	255088	0.00

Table 7: The number of signal events is not normalized

C Background-only hypothesis

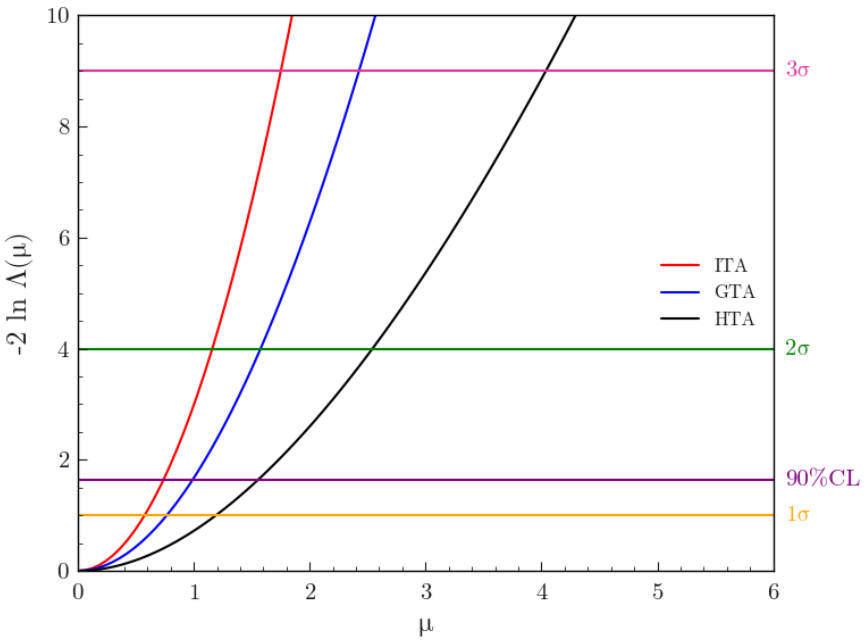


Figure 23: Twice the negative likelihood-ratio λ_μ at $\mu = 0$

Table 8: Background-only hypothesis, with $\mu = 0$

	μ at 90% CL	Upper limit on $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ at 90% CL	σ^+
GTA	0.95	0.47×10^{-5}	0.74
HTA	1.55	0.77×10^{-5}	1.19
ITA	0.75	0.37×10^{-5}	0.59

References

[1] Yann Gouttenoire. Beyond the standard model cocktail, 2022.

[2] Johnny Alejandro Mora Grimaldo and Hiroaki Aihara. *Measurement of the Flavor Changing Neutral Current Decays $B \to Kl+l-$ at the Belle II Experiment*. PhD thesis, Tokyo, University of Tokyo, Tokyo, 2020. Presented on 27 07 2020.

[3] Andrzej J. Buras. Flavor changing neutral current processes. In *28th International Conference on High-energy Physics*, pages 243–270, 10 1996.

[4] C. Greub. Theory of rare B meson decays, 1999.

[5] Belle II Technical Design Report, 2010.

[6] Joint Institute for Nuclear Research. Electrons and positrons collide for the first time in the superkekb accelerator, 2018. Accessed: 2024-06-07.

[7] Andrzej J. Buras, Jennifer Girrbach-Noe, Christoph Niehoff, and David M. Straub. $B \to K^{(*)}\nu\bar{\nu}$ decays in the Standard Model and beyond, 2014.

[8] W. G. Parrott, C. Bouchard, and C. T. H. Davies. Erratum: Standard Model predictions for $B \to K\ell^+\ell^-$, $B \to K\ell_1^-\ell_2^+$ and $B \to K\nu\overline{\nu}$ using form factors from $N_f = 2+1+1$ lattice QCD [Phys. Rev. D 107, 014511 (2023)]. *Phys. Rev. D*, 107:119903, Jun 2023.

[9] The physics of the b factories. *The European Physical Journal C*, 74(11), November 2014.

[10] Lea Reuter, Torben Ferber, Günter Quast, James Kahn, and Pablo Goldenzweig. *Full Event Interpretation using Graph Neural Networks*. PhD thesis, Karlsruhe, Karlsruhe Institute of Technology, Karlsruhe, 2022. Presented on 17 02 2022.

[11] Search for $B^+ \to K^+\nu\overline{\nu}$ Decays Using an Inclusive Tagging Method at Belle II. *Phys. Rev. Lett.*, 127:181802, Oct 2021.

[12] T. Keck, F. Abudinén, Florian U. Bernlochner, R. Cheaib, S. Cunliffe, M. Feindt, T. Ferber, M. Gelb, J. Gemmler, P. Goldenzweig, M. Heck, S. Hollitt, J. Kahn, J.-F. Krohn, T. Kuhr, I. Komarov, L. Ligioi, M. Lubej, F. Metzner, M. Prim, C. Pulvermacher, M. Ritter, J. Schwab, W. Sutcliffe, U. Tamponi, F. Tenchini, N. E. Toutounji, P. Urquijo, D. Weyland, and A. Zupanc. The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment. *Computing and Software for Big Science*, 3(1), February 2019.

[13] Evidence for $B^+ \to K^+\nu\bar{\nu}$ Decays, 2024.

[14] James Kahn, Ilias Tsaklidis, Oskar Taubert, Lea Reuter, Giulio Dujany, Tobias Boeckh, Arthur Thaller, Pablo Goldenzweig, Florian Bernlochner, Achim Streit, and Markus Götz. Learning tree structures from leaves for particle decay reconstruction. *Machine Learning: Science and Technology*, 3(3):035012, sep 2022.

[15] Ilias Tsaklidis, Pablo Goldenzweig, Isabelle Ripp-Baudot, James Kahn, and Giulio Dujany. *Demonstrating learned particle decay reconstruction using Graph Neural Networks at Belle II*. PhD thesis, Université de Strasbourg, Strasbourg and Karlsruhe, 2020. Presented on 19/06/2020.

[16] *Classification and Regression Trees, author=Breiman, Leo and Friedman, Jerome and Stone, Charles J and Olshen, Richard A*. Wadsworth International Group, 1984.

[17] Samuel Stanley Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals of Mathematical Statistics*, 9:60–62, 1938.

[18] Matthew Feickert, Lukas Heinrich, and Giordon Stark. pyhf: pure-python implementation of histfactory with tensors and automatic differentiation. 2022.

[19] L.A. Spiller. Modification of Fox-Wolfram moments for hadron colliders. *Journal of High Energy Physics*, 2016(3), March 2016.

[20] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, August 2016.