Faculté de physique et ingénierie
Université de Strasbourg

Belle II

IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

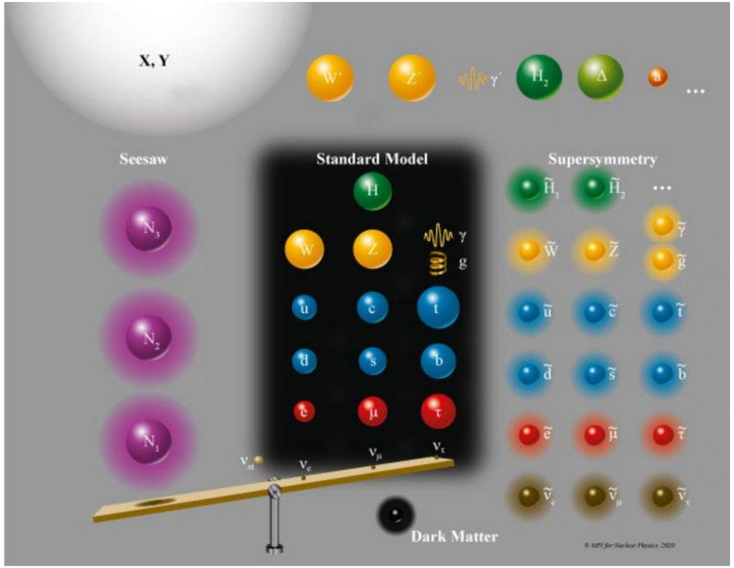
Search for $B^+ \to K^+\nu\bar{\nu}$ at Belle II using Graph Neural Networks

Merna Abumusabh

Supervised by: Giulio Dujany and Jacopo Cerasoli

From March 4th to July 26th

The Standard Model and Beyond ..



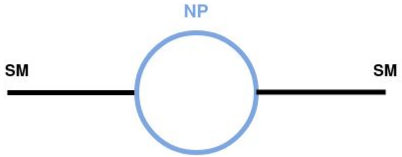
The standard model (SM) stands as the cornerstone of modern particle physics, but:

Dark matter, Dark energy,....

Energy frontier
(direct approach)

- Directly produce new particles using high energy collision.
- Sensitive to the energy scale of NP.

Intensity frontier
(indirect approach)



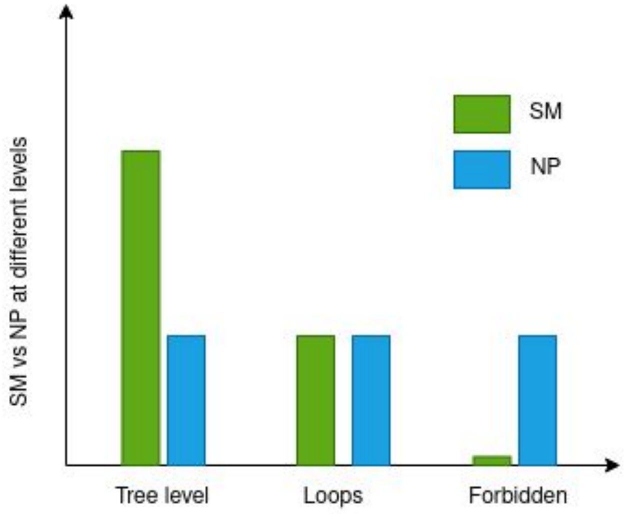
- Find signatures of new particles in the intermediate state.
- Sensitive to the flavor structure of NP.

Introduction: Indirect searches of New Physics

Search for new physics can be probed by indirect research with the aid of Effective Field Theories.

Probes for New Physics at much higher scales than direct searches can be performed on:
Flavor Changing Neutral Current processes.

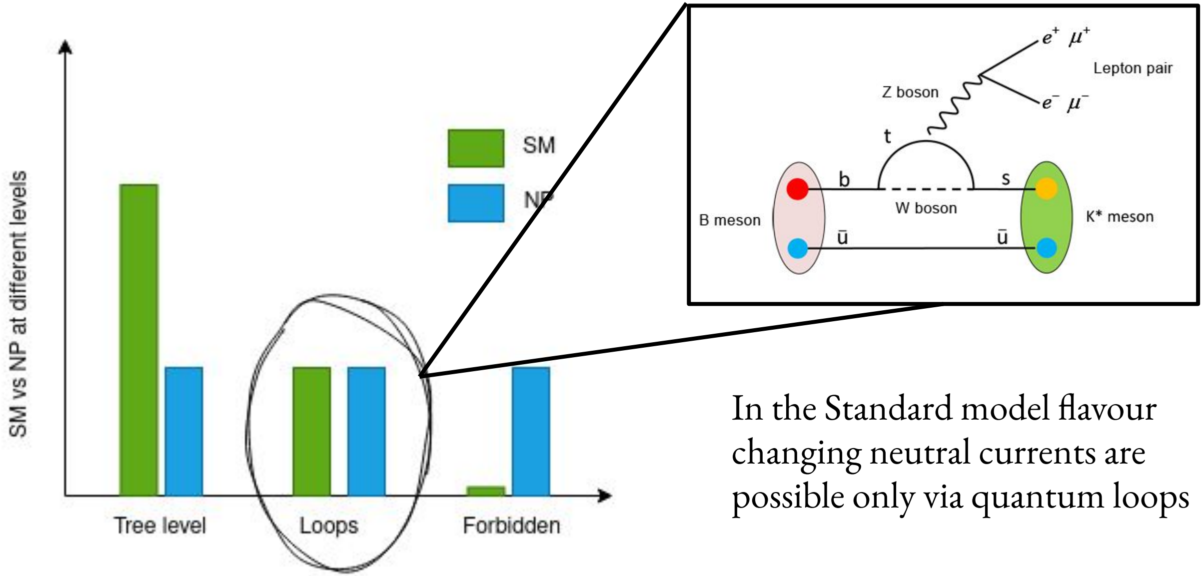
Such as rare B-mesons decays



Introduction: Indirect searches of New Physics

Search for new physics can be probed by indirect research with the aid of Effective Field Theories.

Probes for New Physics at much higher scales than direct searches can be performed on:
Flavor Changing Neutral Current processes. → Such as rare B-mesons decays



In the Standard model flavour changing neutral currents are possible only via quantum loops

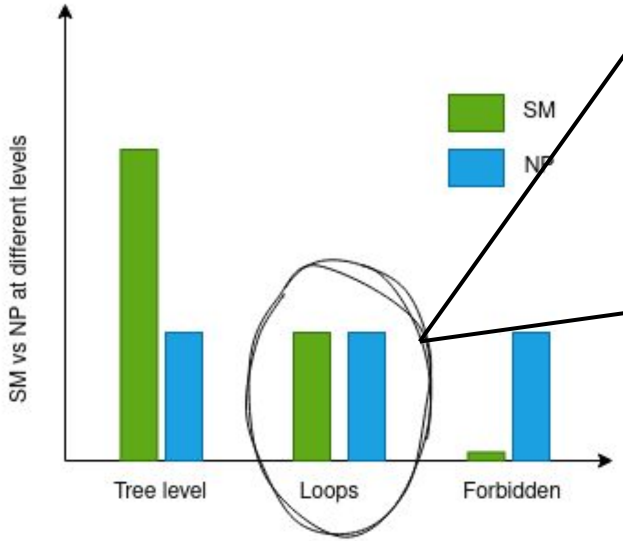
Rare process, but we have the perfect place to search for it!!

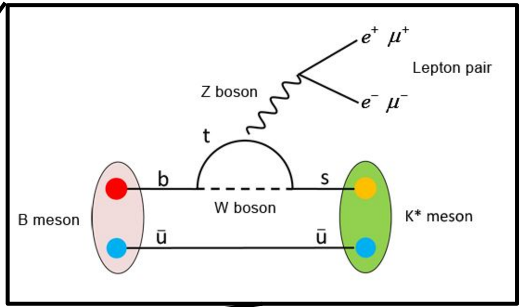
Introduction: Indirect searches of New Physics

Search for new physics can be probed by indirect research with the aid of Effective Field Theories.

Probes for New Physics at much higher scales than direct searches can be performed on:
Flavor Changing Neutral Current processes.

Such as rare B-mesons decays





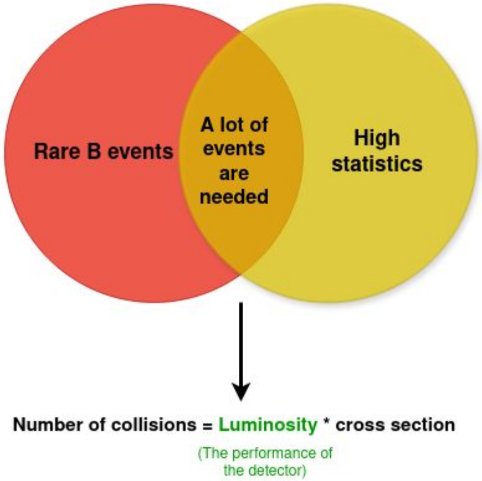
In the Standard model flavour changing neutral currents are possible only via quantum loops

Rare process, but we have the perfect place to search for it!! (A search for a "rare beauty")

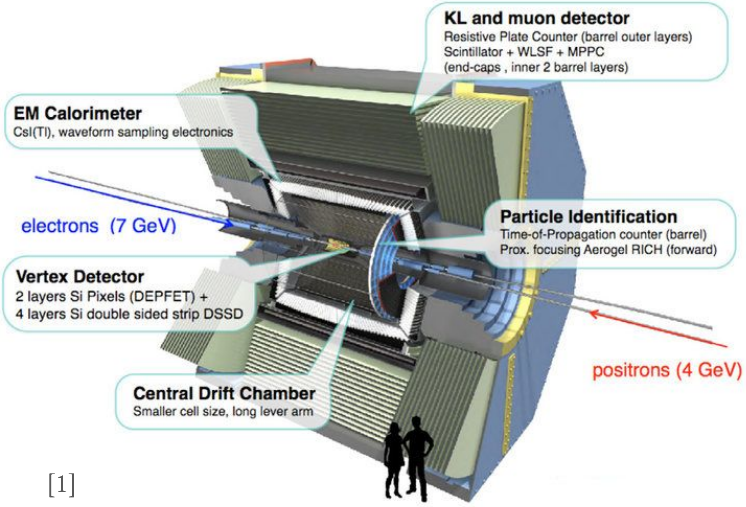
The BELLE II Experiment

SuperKEK, operates at the Upsilon(4S) resonance with 7GeV - 4 GeV beams.

- **Belle II is an intensity frontier experiment.**
- **Asymmetric collider producing B mesons**

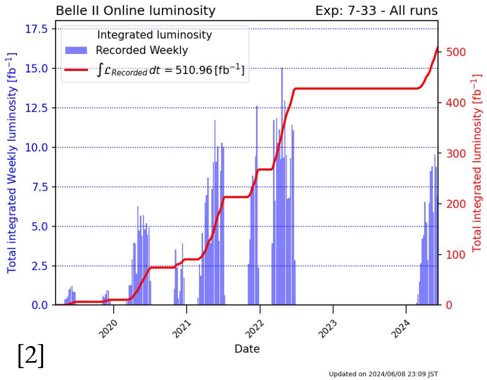


$$\int L_{\text{Recorded}} dt = 511.5 \text{fb}^{-1}$$



[1]

Belle II Technical Design Report

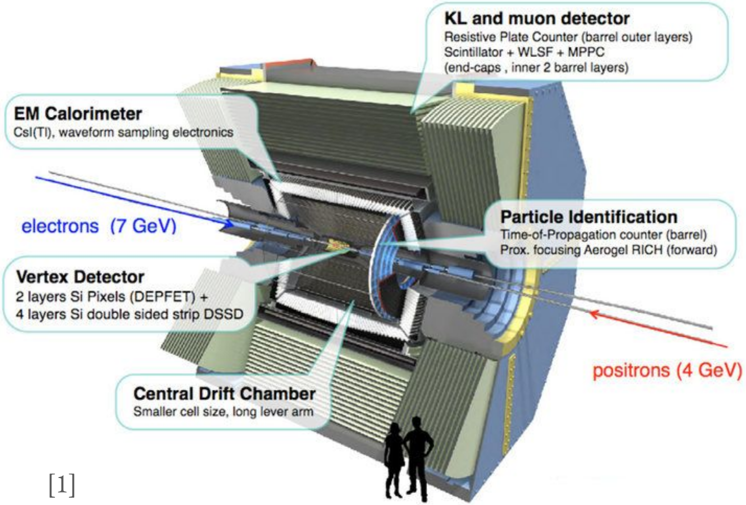


[2]

The BELLE II Experiment

SuperKEK, Operates on the Upsilon(4S) resonance with 7 GeV(e-) on 4 GeV(e+) beams.

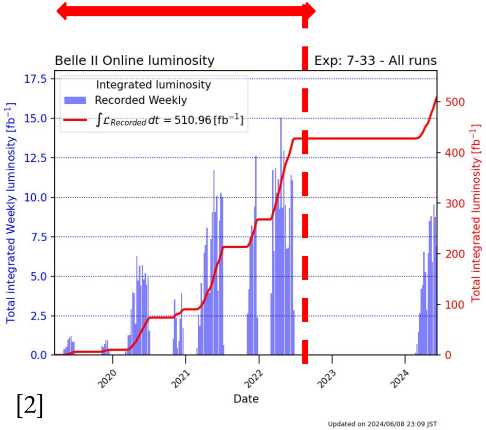
- **Belle II is an intensity frontier experiment.**
- **Asymmetric collider producing B mesons**



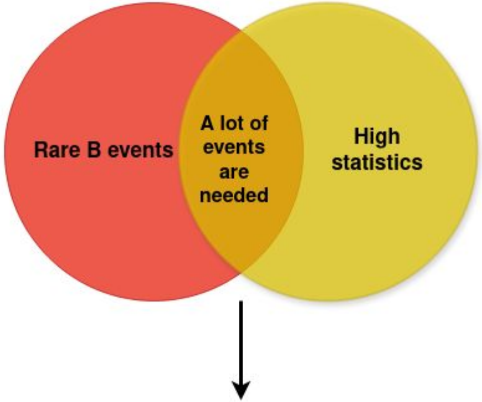
[1]

Belle II Technical Design Report

The L range in this work



[2]



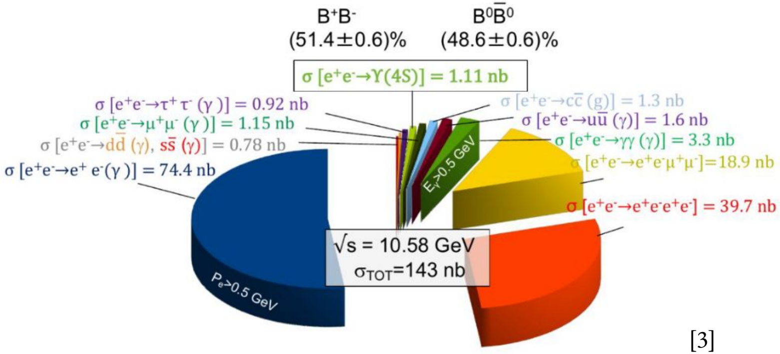
Number of collisions = Luminosity * cross section

(The performance of the detector)

$$\int L_{\text{Recorded}}\, dt = 511.5 \text{fb}^{-1}$$

B production at Belle II

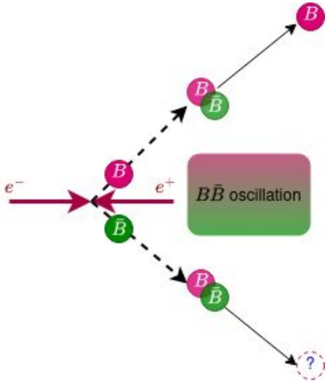
B factories: collide e^+e^- asymmetrically at $\Upsilon(4S)$ resonance.



[3]

$\approx 0.78\%$ of the e^+e^- collisions produces $\Upsilon(4S)$

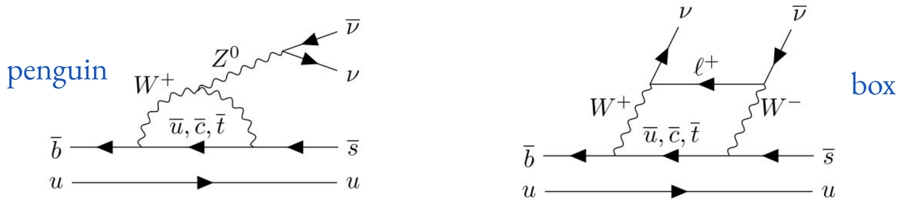
Many B mesons are produced, but also so much more τ leptons and quarks.



At the KEK B-factory, a B-meson and anti-B-meson are generated as a pair.

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

The particular channel of interest in this work is the $B^+ \to K^+\nu\bar{\nu}$ decay.



Lowest-order diagrams for the decay in the SM

Good place to look at!

- Theoretically very clean due to the absence of virtual photons:
  The branching fraction is well predicted by the SM, so any deviation from that could reveal potential contribution from New Physics (NP)

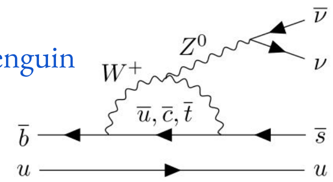
But challenging!

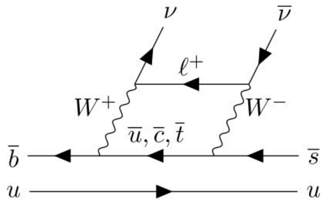
- Pretty rare
  The sensitivity and the uncertainty are highly affected by the size of the sample.
- Two neutrinos at the final state
  The process is almost invisible

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

The particular channel of interest in this work is the $B^+ \to K^+\nu\bar{\nu}$ decay.

penguin





box

Lowest-order diagrams for the decay in the SM

B-tagging

Good place to look at!

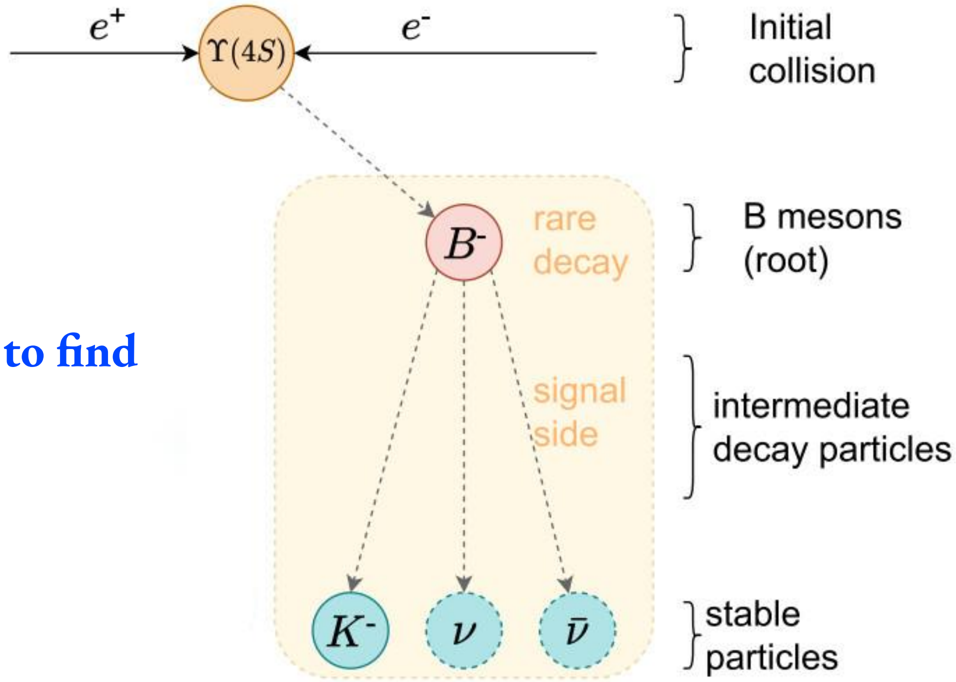
- Theoretically very clean due to the absence of virtual photons: The branching fraction is well predicted by the SM, so any deviation from that could reveal potential contribution from New Physics (NP)

But challenging!

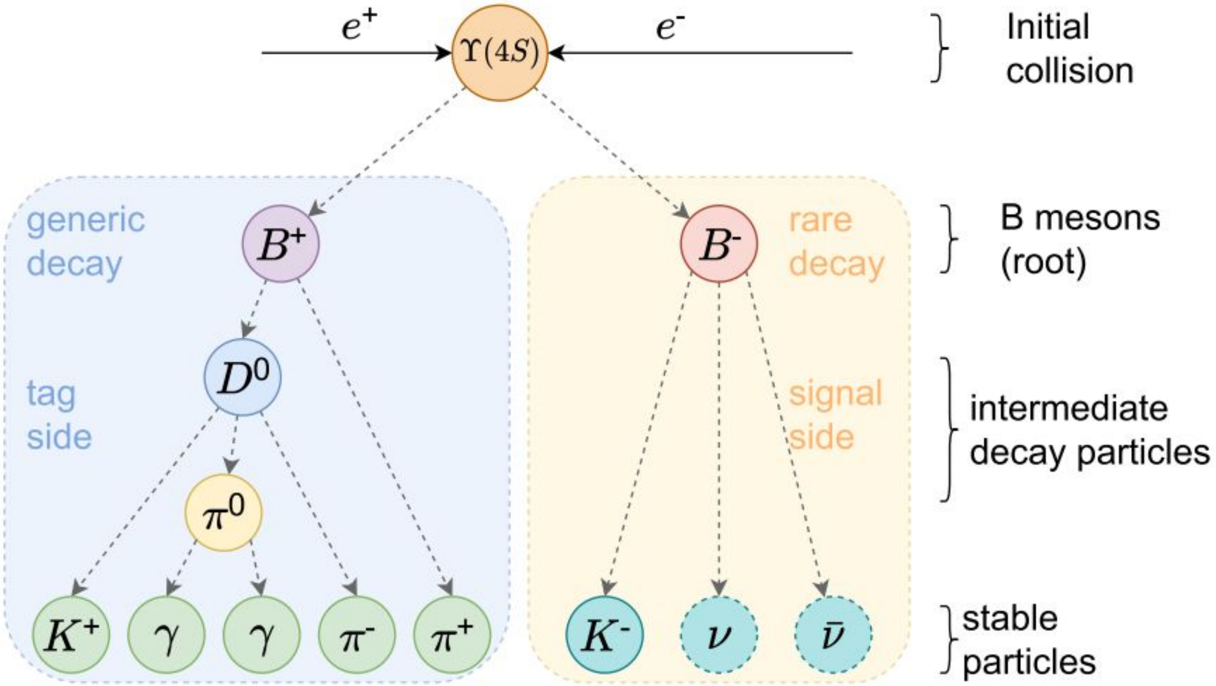
- Pretty rare
  The sensitivity and the uncertainty are highly affected by the size of the sample.
- Two neutrinos at the final state
  The process is almost invisible

B-tagging at Belle II

The decay we want to find

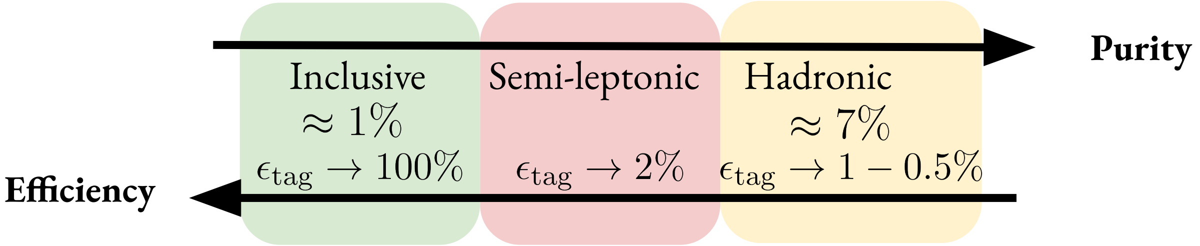


B-tagging at Belle II



[4]

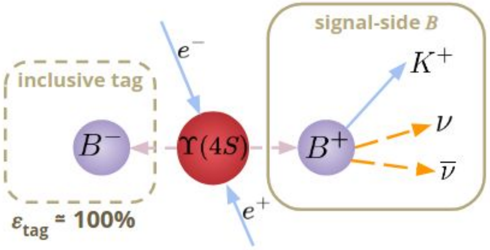
B-tagging at Belle II



$B^+ \to K^+ \nu\bar{\nu}$ at Belle II

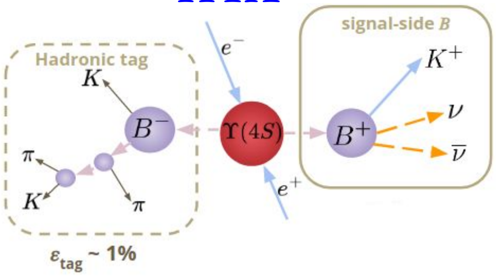
Two complementary approaches are used at Belle II with 362fb^{-1}

ITA



Reconstruct signal B only, exploit the rest of the event (ROE) to suppress backgrounds.

HTA



Fully reconstruct the B-tag kinematics
Lower background contamination

B-tagging at Belle II

Purity →

Inclusive	Semi-leptonic	Hadronic
$\approx 1\%$		$\approx 7\%$
$\epsilon_{\text{tag}} \to 100\%$	$\epsilon_{\text{tag}} \to 2\%$	$\epsilon_{\text{tag}} \to 1 - 0.5\%$

← **Efficiency**

But can we get good purity, and good tagging efficiency?

$B^+ \to K^+ \nu\bar{\nu}$ at Belle II

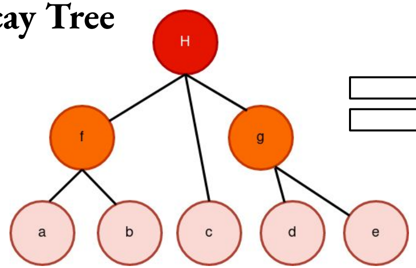
Two complementary approaches are used at Belle II with 362 fb

Reconstruct signal B only, exploit the rest of the event (ROE) to suppress backgrounds.

Fully reconstruct the B-tag kinematics
Lower background contamination

Graph-based Full Event Interpretation Model (graFEI)

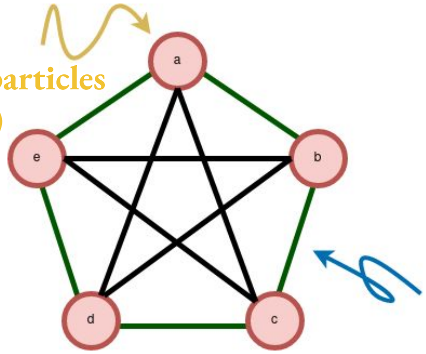
Decay Tree



The Graph-based Full Event Interpretation (graFEI) is a machine learning model to inclusively reconstruct events at Belle II using only the information on the final state particles, without presupposing any underlying decay chain structure.

Represented by a graph

Detected particles (node)

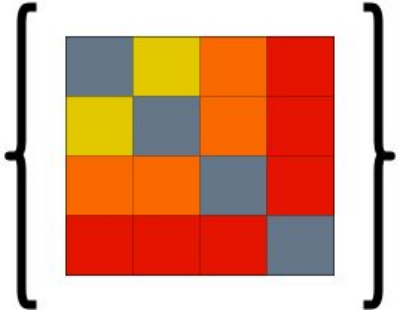


Relations (edge)

Transformed to

LCAS matrix

Lowest Common Ancestor Matrix

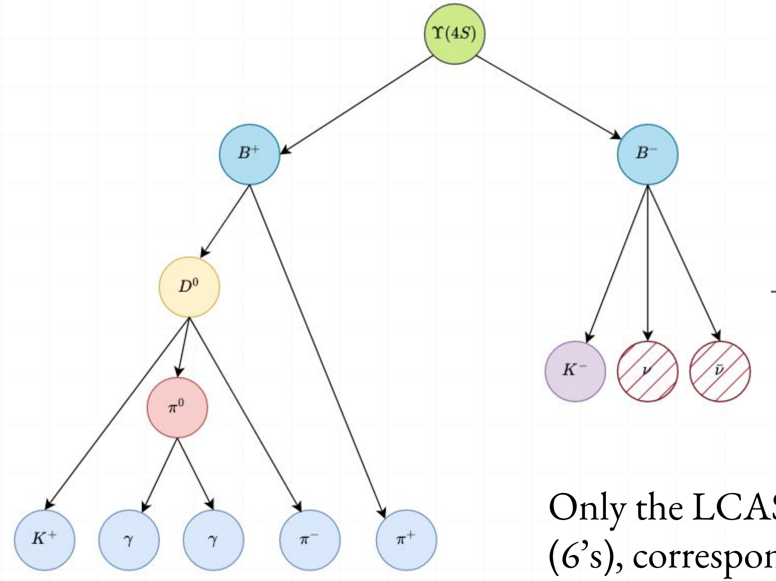


graFEI: Decay trees as graphs

The LCAS matrix elements identify the nearest common ancestor shared by any two particles.

$\Upsilon(4S)$	$B^{\pm}$	$D_{(s)}^{*\pm,0}$	$D^{\pm,0}$	K_s^0	π^0
6	5	4	3	2	1

Each particle is labeled by an identifying number



LCAS representation

	K^+	γ	γ	π^-	π^+	K^-
K^+		3	3	3	5	6
γ	3		1	3	5	6
γ	3	1		3	5	6
π^-	3	3	3		5	6
π^+	5	5	5	5		6
K^-	6	6	6	6	6	

Only the LCAS matrix with full (6's), corresponding to the signal side, are kept

Analysis strategy

- **Reconstructing the decay with the graFEI**

- **Candidate Selection:**
  Identify the K^+ candidates.

- **Separate signal and backgrounds:**
  - **Preselection cuts**
  - **Training the BDT**
  - **Defining a signal region (SR)**

- **Extract the signal strength:**
  Simultaneous binned profile scan to get the total number of signal events in the signal region and extract the signal strength.

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu})_{\mathrm{SM}} = 5.56 \times 10^{-6}$$

The signal strength is defined as:

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

Input samples

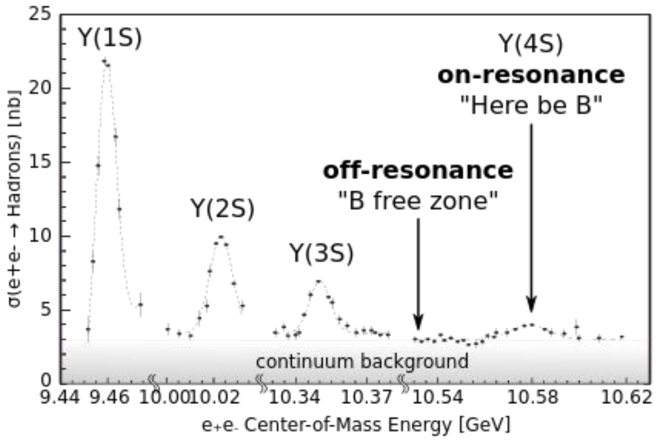
The following samples were generated to to mimic the events at the Belle II detector:

- 98×10^6 events of the signal $B^+ \to K^+\nu\bar{\nu}$.

Backgrounds, correspond to 4 times the luminosity of the Belle II data set:

- On-resonance: 28.2×10^6 of B^+B^- and $B^0\bar{B}^0$
- Off-resonance: 211.7×10^6 of $e^+e^- \to q^+q^-$

Signal events were normalized to the SM branching fraction, and backgrounds were scaled to match the integrated luminosity on of the Belle II data at 362fb^{-1}



[4]

Input samples

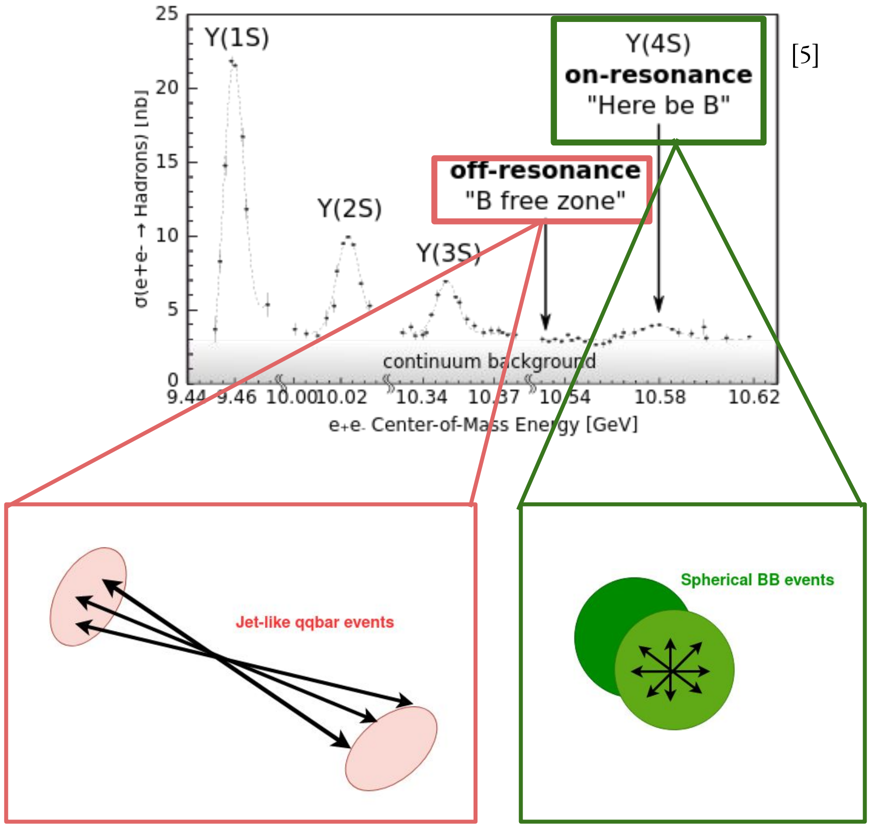
The following samples were generated to mimic the luminosity of the Belle II detector:

- 98×10^6 events of the signal $B^+ \to K^+ \nu \bar{\nu}$.

Backgrounds, correspond to 4 times the luminosity of the Belle II data set:

- On-resonance: $28.2 \times 10^6$of B^+B^- and $B^0\bar{B}^0$
- Off-resonance: 211.7×10^6 of $e^+e^- \to q^+q^-$

Signal events were normalized to the SM branching fraction, and backgrounds were scaled to match the integrated luminosity on of the Belle II data at 362fb^{-1}

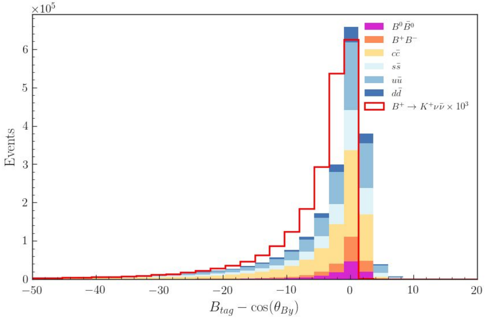


[5]

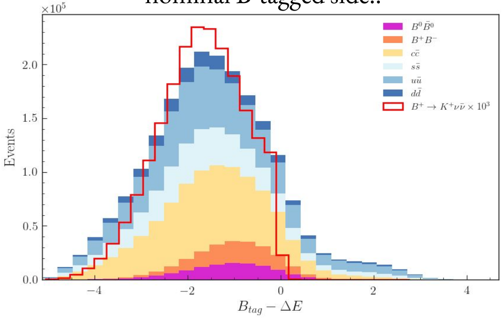
Discriminating variables

The input variables are divided into three main categories:

1- Kinematics



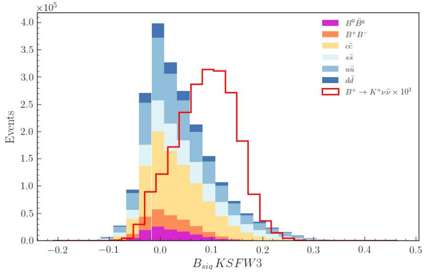
Cosine of the angle in CMS between momentum the particle and a nominal B-tagged side..

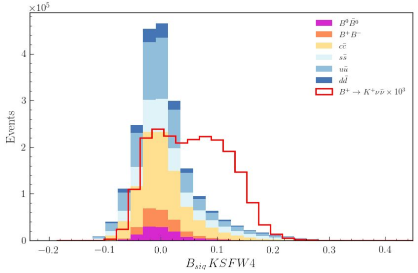


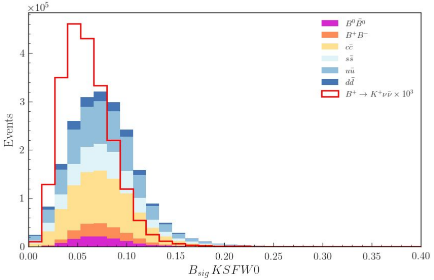
The difference between the energy of the tagged side and half the CMS energy

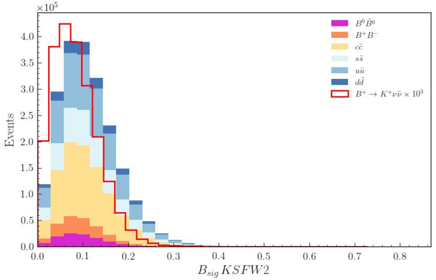
2- Event shape variables

Such as the modified Fox-Wolfram moments





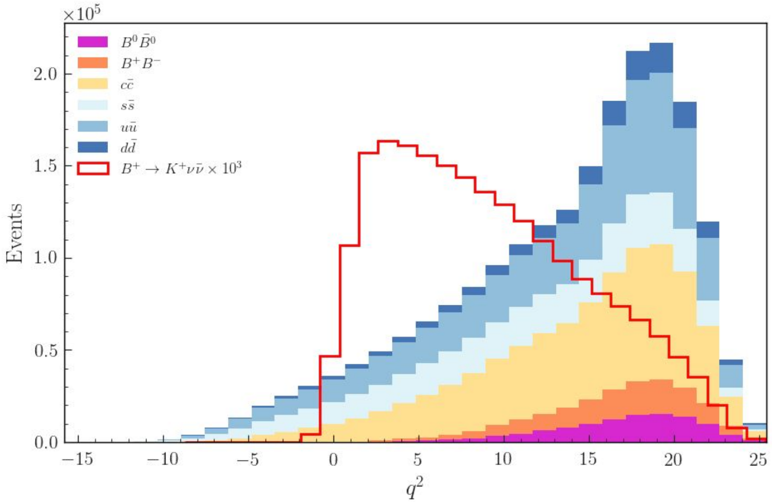




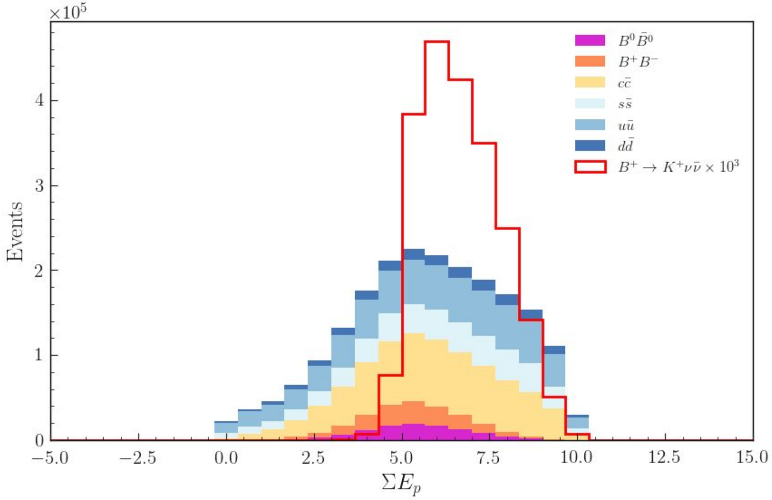
Discriminating variables

3- Other variables

The mass squared of the neutrino pair



Missing energy of event in the CMS +
missing momentum of event in CMS



Preselection cuts

1- GraFEI reconstruction requirement

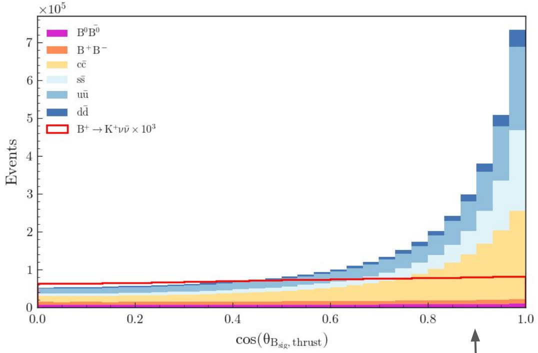
The LCAS matrix is required to contain the signal kaon and the tag side.
The signal reconstruction efficiency is found to be 27.29%

→ Performed by the IPHC Belle II team

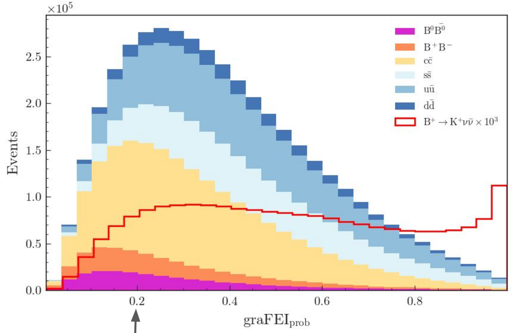
Ntuples

2- Pre-training cuts

1- A cut on the particle identification for Kaons > 0.9



2- The cosine of the angle between thrust axis of the signal B and the thrust axis of the Rest of Event (ROE) < 0.9



3- Geometric mean of probabilities associated to each predicted LCAS element > 0.2

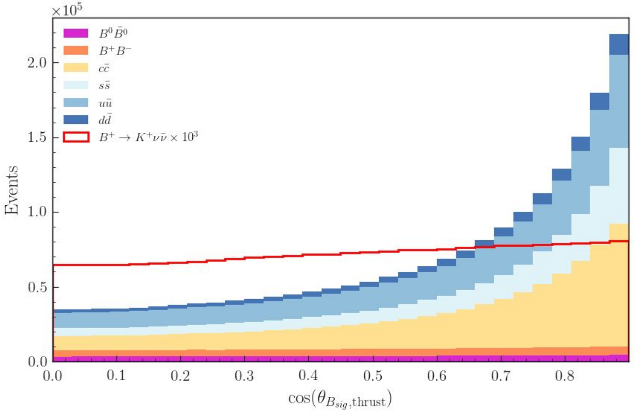
Preselection cuts

GraFEI reconstruction requirement

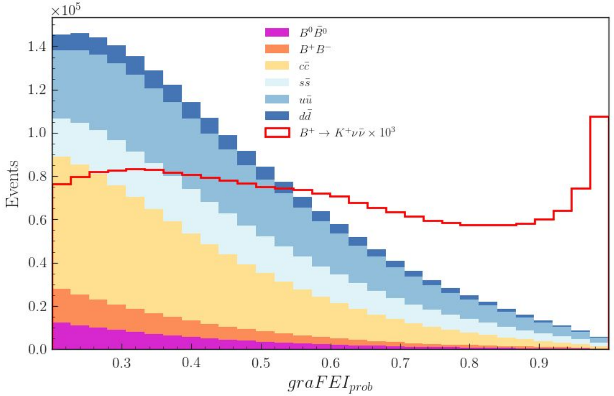
The LCAS matrix is required to contain the signal kaon and the tag side. → Performed by the IPHC Belle II team
The signal reconstruction efficiency is found to be 27.29%

Pre-training cuts

`Bsgn_cosTBTO` < 0.9

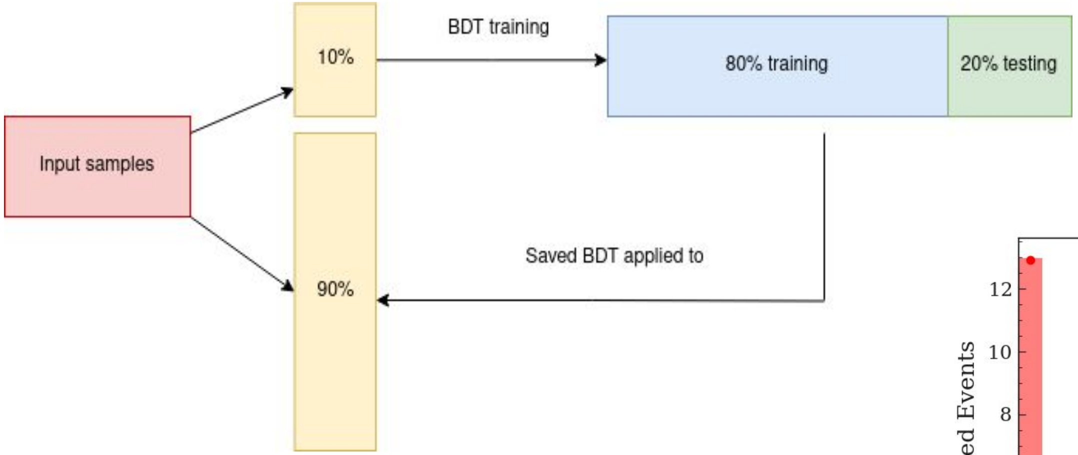


`graFEI_probegdeGeom` > 0.2

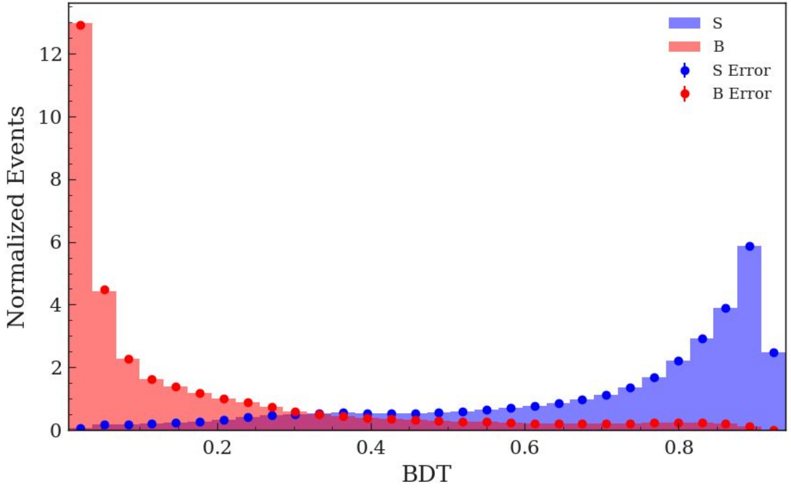


$\epsilon_{\text{pre-training}} = 53.61\ \%$

BDT Training



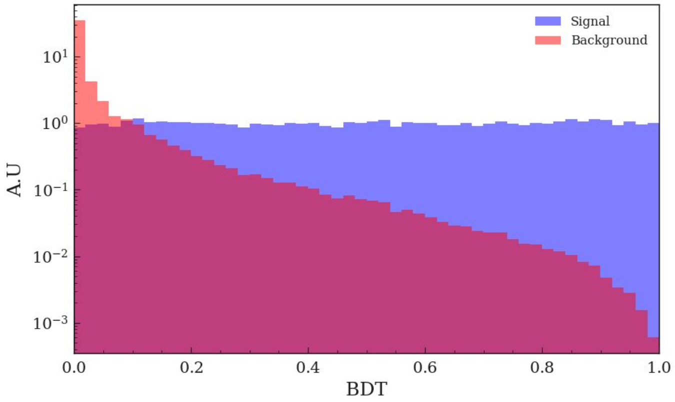
In order to achieve a good performance, the variables with strong correlations in both the signal and the backgrounds were removed from the input list, and the BDT was trained with 23 variables.



Binned likelihood fitting

1- Flatten the BDT output

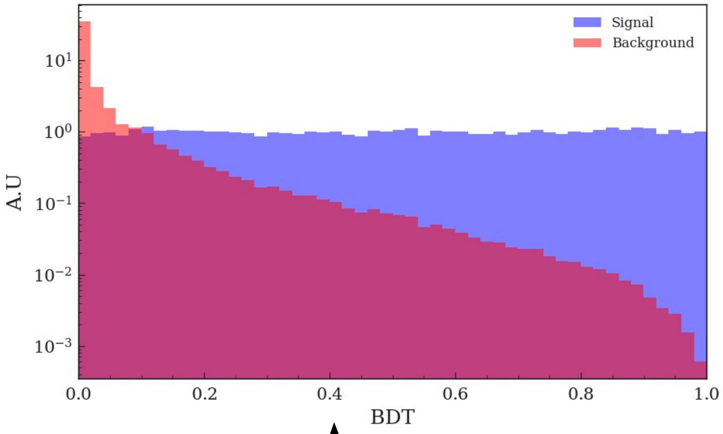
$$a \to \frac{\int_{-1}^{a} BDT(y)\, dy}{\int_{-1}^{1} BDT(y)\, dy}$$



Binned likelihood fitting

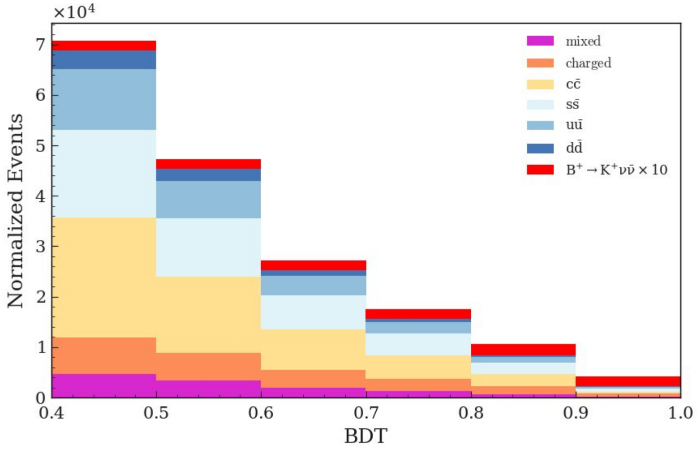
1- Flatten the BDT output

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



Cut the BDT here

2- Defining the signal search region (SR) by applying a cut of BDT > 0.4 to reject the backgrounds.
The SR is divided 6 equal bins of the classifier probability output.



Results of GTA

Signal Efficiency

Reminder: # of generated signals 98710654

Selection	Number of signals after the cut	$\varepsilon_{sig} \times \%$
graFEI event reconstruction	26940105	27.295 ± 0.004
Preselection cuts	14429334	53.607 ± 0.003
Signal Region	7885101	54.646 ± 0.004

Total efficiency =

$$7.989 \pm 0.001\%$$

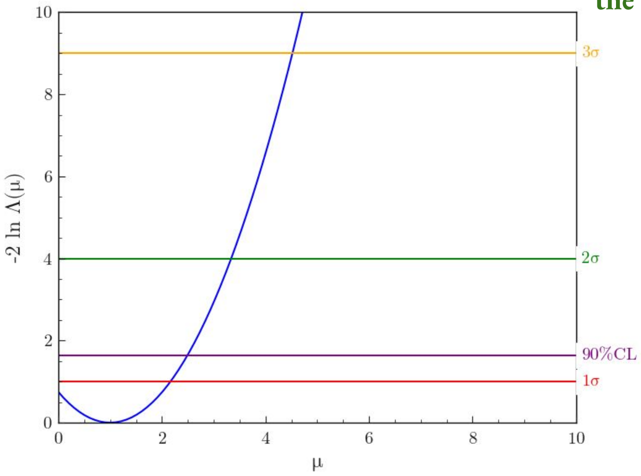
Correcting the uncertainty value in the report

- A binned profile scan was performed to extract the signal strength at 90% CL for the SM prediction.
- The fit parameter is signal-strength

$$\mu < 2.29$$

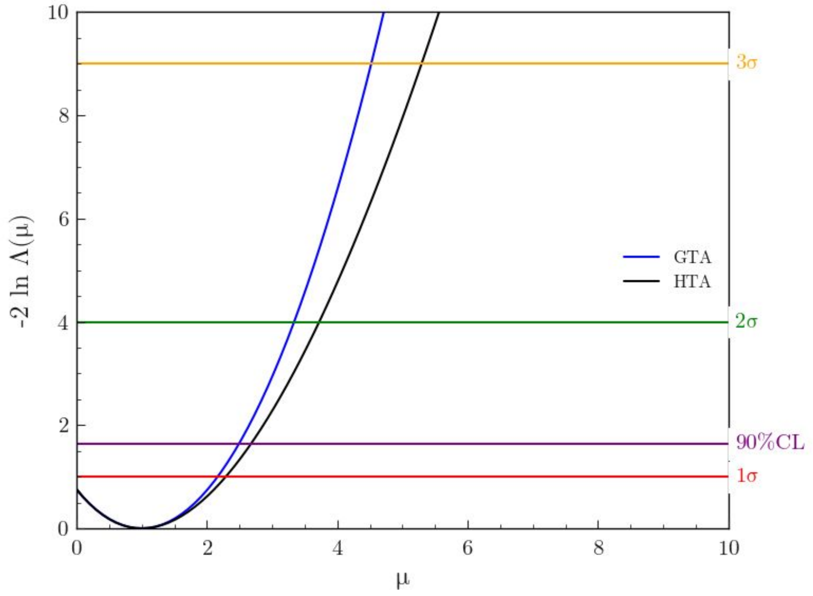
And so, the upper limit on the branching fraction is:

$$\mathcal{B}(B^+ \to K^+ \nu\bar{\nu}) < 1.38 \times 10^{-5}$$



Comparing with previous results

	μ at 90% CL	Upper limit on $\mathcal{B}(B^+ \to K^+ \nu\bar{\nu})$ at 90% CL	σ^+
GTA	2.49	1.38×10^{-5}	1.16
HTA	2.67	1.48×10^{-5}	1.28



GTA shows better sensitivity to the signal at the SM prediction than HTA at 90% CL.

Comparing with previous results

HTA central value

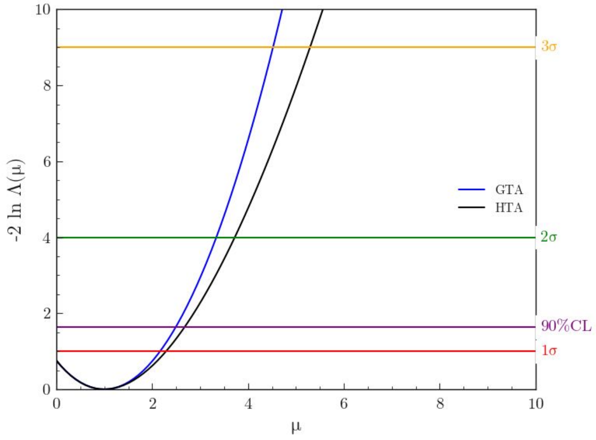
$\mu = 2.2$

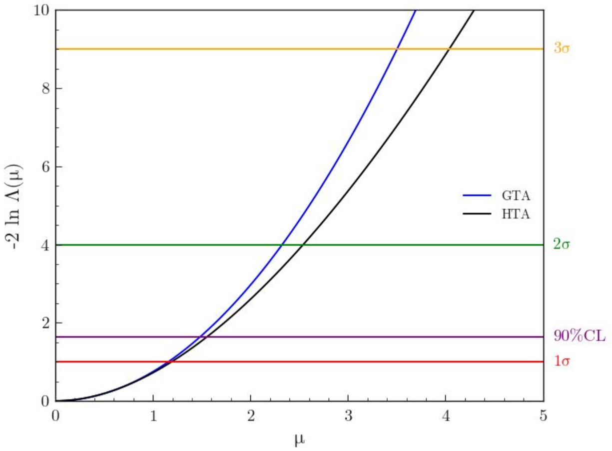
	μ at 90% CL	Upper limit on $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ at 90% CL	σ^+
GTA	3.70	2.05×10^{-5}	1.17
HTA	4.01	2.22×10^{-5}	1.38

BG-only hypothesis

$\mu = 0$

	μ at 90% CL	Upper limit on $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ at 90% CL	σ^+
GTA	1.48	0.82×10^{-5}	1.15
HTA	1.55	0.86×10^{-5}	1.19

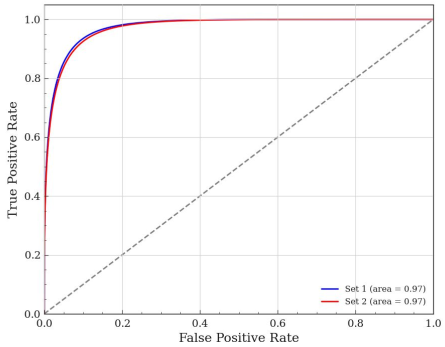




Some extra Results and prospects

Some extra Results and prospects

The BDT was trained with less variables (from 23 to 16) → Did not affect the results



Hyperparameters Optimization

RandomizedSearchCV : Integrates directly with scikit-learn's cross-validation

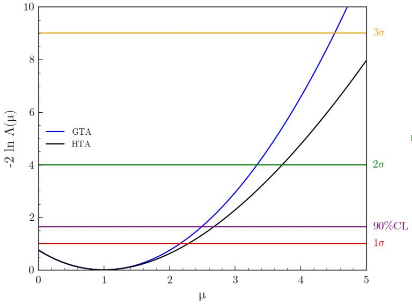
Best parameters found:
- learning_rate: 0.15470772691245635
- max_depth: 5
- n_estimators: 600
- subsample: 0.8912865394447438
- Best cross-validation ROC AUC score: 0.9721

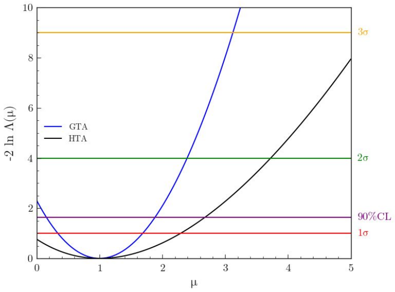
Also checked by Optuna: a Bayesian Optimization.

Optimizing the SR

BDT [0.8, 1]

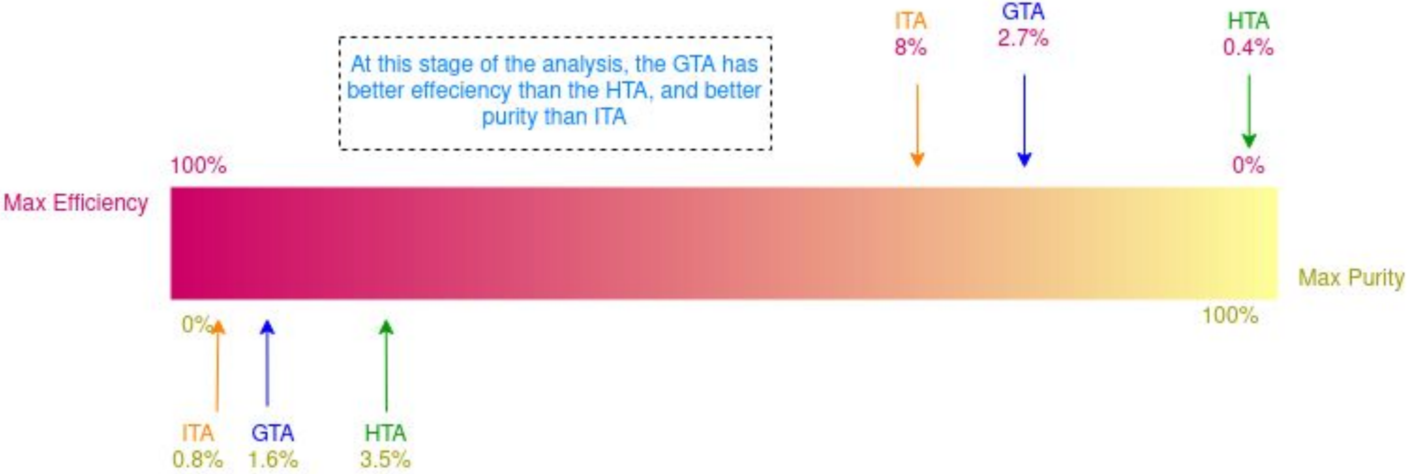
	# BG	U.L on BF at 90% CL
Before	167205	1.384 $(\times 10^{-5})$
After	4204	1.045 $(\times 10^{-5})$





Some extra Results and prospects

- The background events have been reduced by 97.4%, and the upper limit on the branching fraction at 90% CL by 24.5%. However, the efficiency has decreased to one-third of its original value.
- Before optimization, the purity was 0.11%. After the optimization, it increased to 1.56%.



Summary and Next Steps

- A search for the $B^+ \to K^+\nu\bar{\nu}$ signal was performed using a graph-based algorithm, the graFEI.
- Consistency in the signal strength between the GTA and the HTA, with the GTA having better sensitivity.

Next steps:

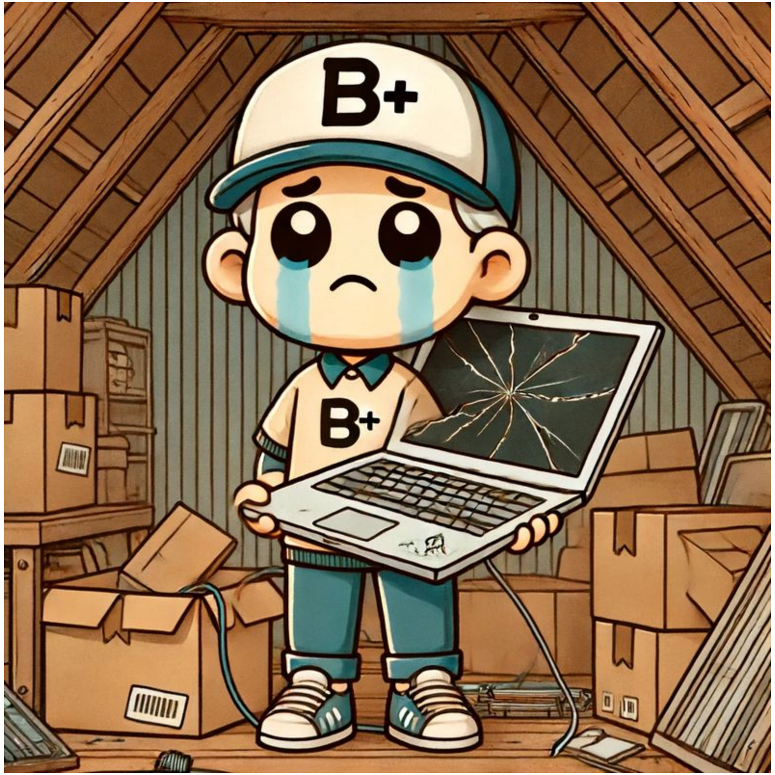
- Optimize the Selection efficiency and the signal purity.
- Compare the results to the inclusive tagged analysis.
- Investigate systematic uncertainties.

Thanks!

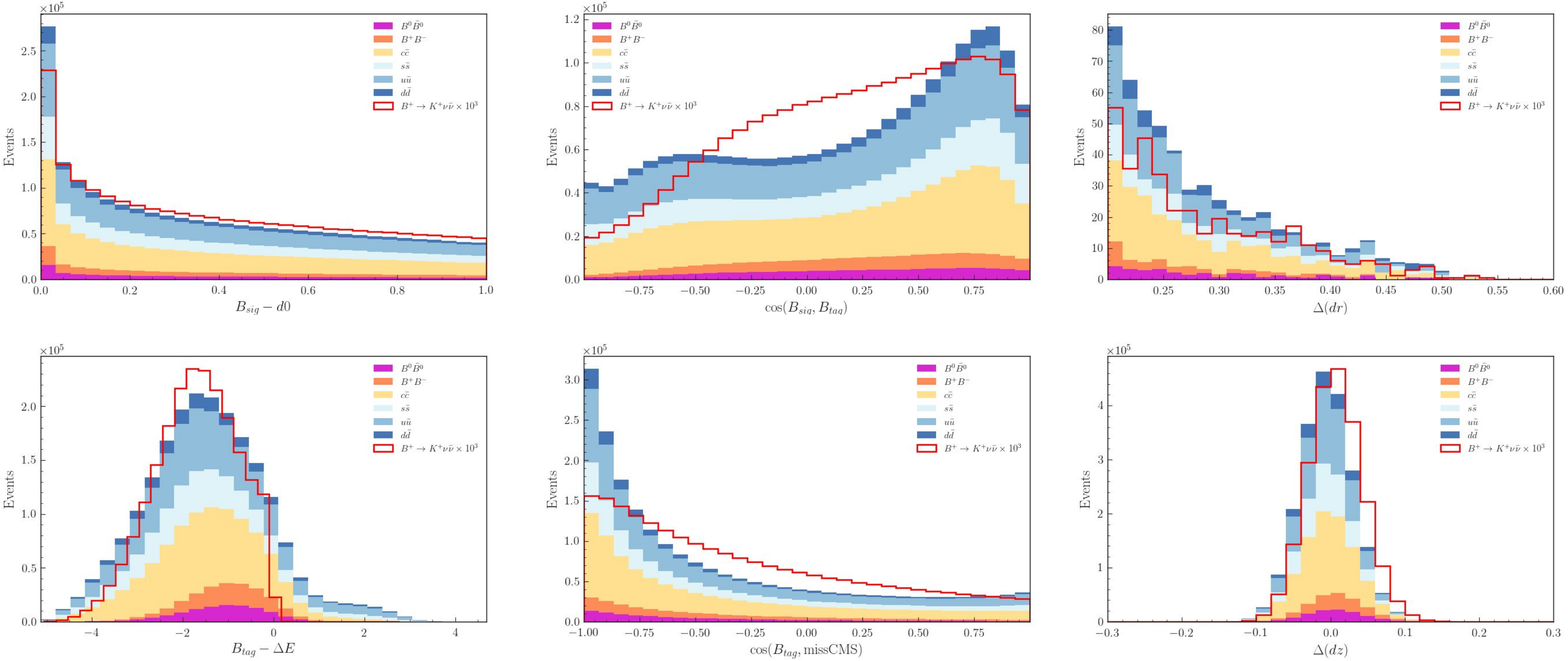
Neutrinos alone, among all the known particles, have ethereal properties that are striking and romantic enough both to have inspired a poem by John Updike and to have sent teams of scientists deep underground for 50 years to build huge science-fiction-like contraptions to unravel their mysteries.

Lawrence M. Krauss

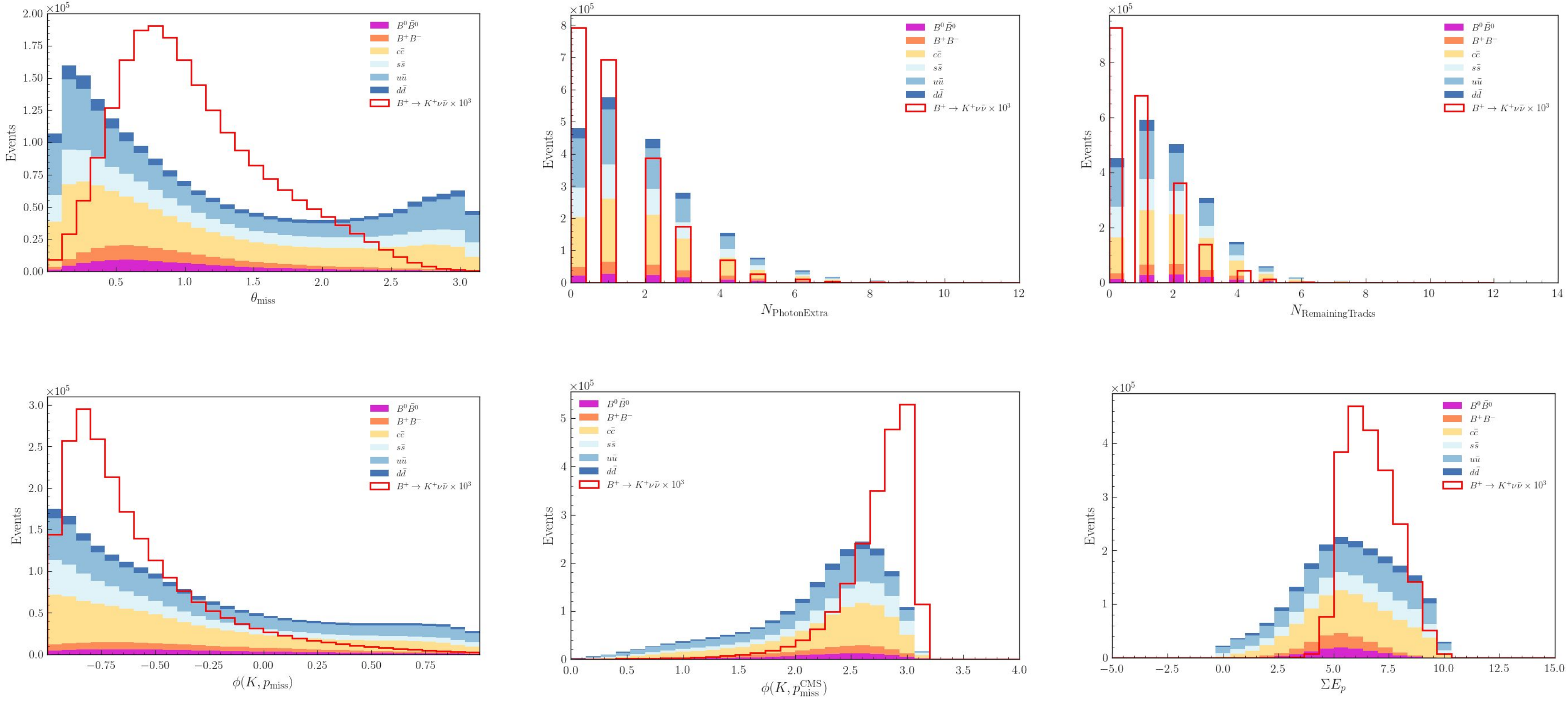
Backups



Input Variables Distribution



Input Variables Distribution



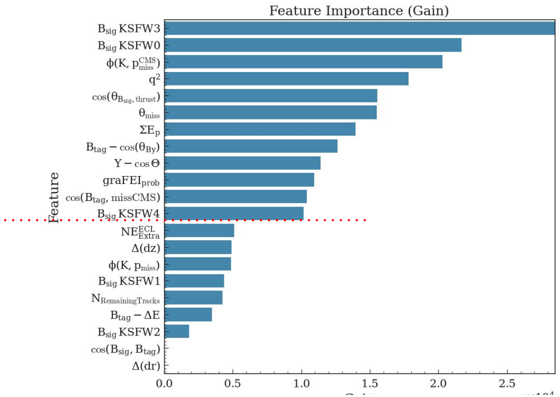
Input Variables Distribution

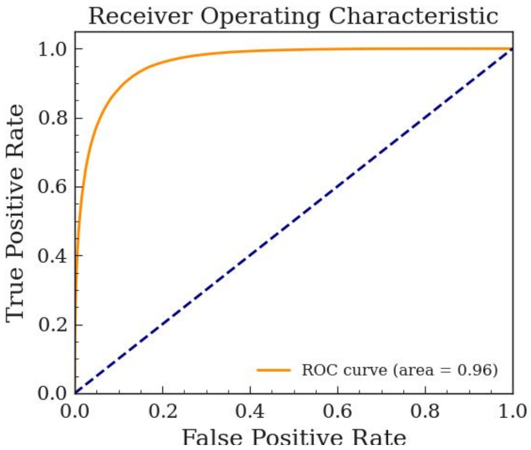
Variable	Description
sumEp	Missing energy of event in the CMS + missing momentum of event in CMS
phi_K_pmiss	Angle between signal kaon and missing momentum in lab frame
miss_theta	Angle between missing momentum and z axis
Ups_cosTheta	Angle between Upsilon total momentum and z axis
NEExtra_ROEClusters_v702	Total energy of neutral clusters in calorimeter in the rest-of-event
Btag_deltaE	ΔE of Btag
cos_tag_miss_CMS	Angle between Btag and missing momentum in CMS
delta_dr	Distance between Bsig impact parameter and Btag impact parameter (impact parameter is the distance of a track from the interaction point)
delta_dz	
nPhotonExtra_ROEClusters_v702	number of photons in the calorimeter used to compute NE-Extra_ROECluster_v702
phi_K_pmiss_CMS	Angle between signal kaon and missing momentum in CM frame
cos_sgn_tag	Angle between Bsig and Btag

BDT Training

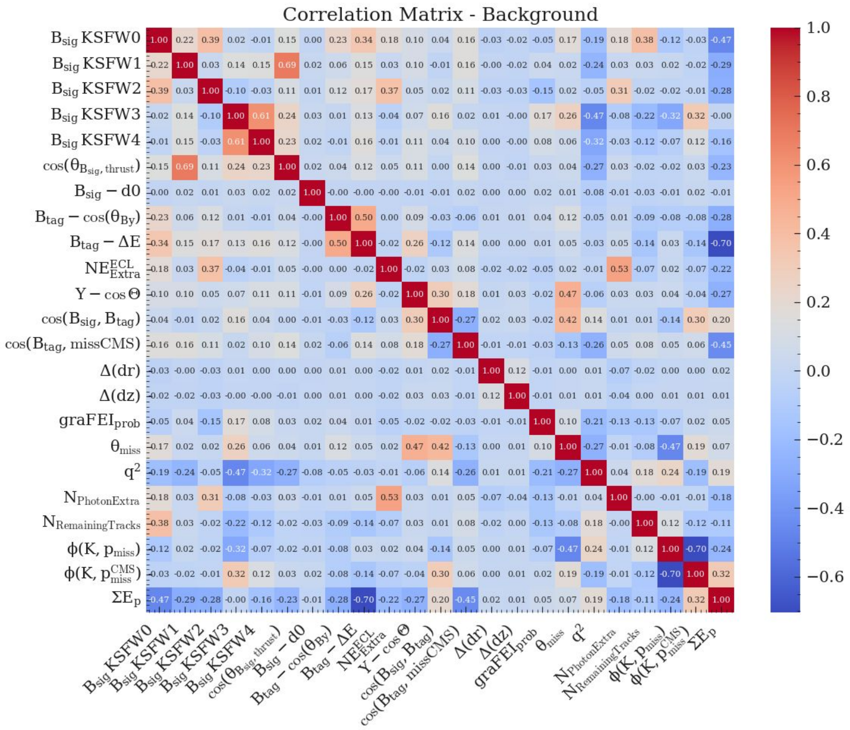
BDT parameter	**Value**
Tree depth	4
Learning rate	0.01
Number of estimators	400
Subsample Ratio	0.6
Evaluation metric	Negative log likelihood
Tree construction algorithm	"hist"

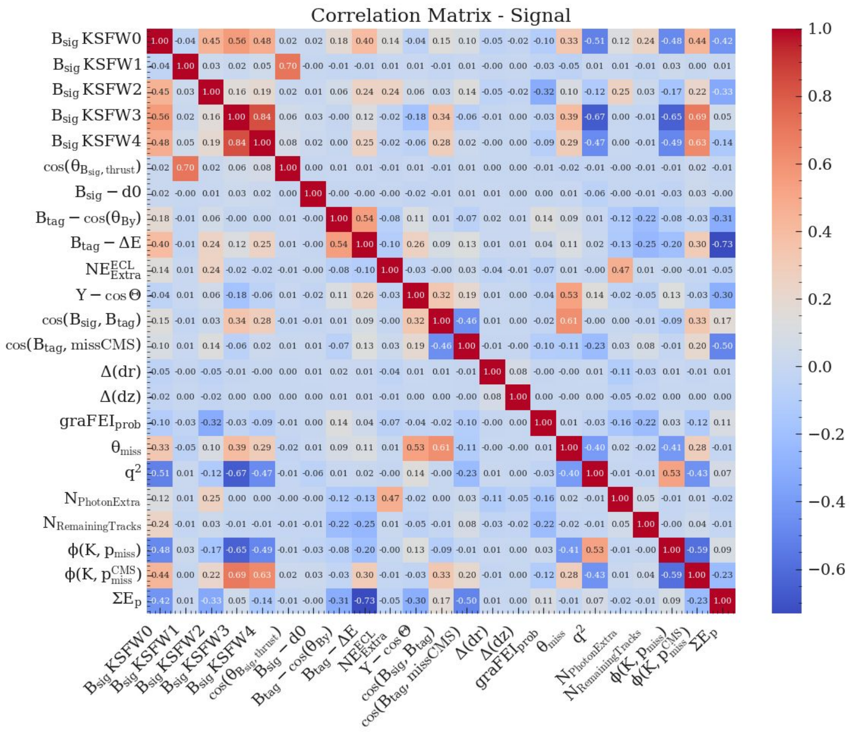
A histogram-based algorithm to find the optimal splits, reducing computational overhead and enhancing training speed.





BDT Training





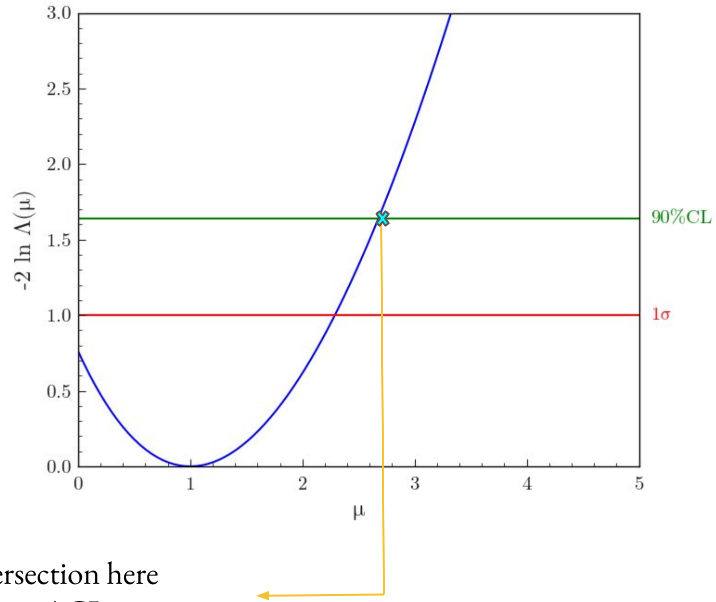
Confidence Intervals

The likelihood function represents probabilities for a particular set of parameters, so integrating the likelihood function in a region directly gives the probability that "the answer" lies in that region.

$$\Lambda = -2\ln\left(\frac{\text{Null Hypothesis}}{\text{Alternative Hypothesis}}\right)$$

$$\Lambda = -2\ln\left(\frac{\mathcal{B}(B^+ \to K^+\nu\bar{\nu})}{\mathcal{B}(B^+ \to K^+\nu\bar{\nu})_{\text{SM}}}\right)$$

As is shown by Wilks' theorem, times the logarithm of a likelihood ratio (where one likelihood represents a null hypothesis (the best-fit) and another likelihood is for an alternative set of parameters) is approximately Chi-squared distributed for sufficiently large datasets (where statistical error is small and the likelihood is well described by a Gaussian distribution).



The intersection here with the 90% CL gives us the upper limit on the branching fraction at 90% CL.

Figures References

[1] https://www.interleptons.eu/Belle-II/
[2] https://www.belle2.org/research/luminosity/
[3] https://indico.in2p3.fr/event/32285/
[4]Full Event Interpretation using Graph Neural Networks-Lea Reuter.
[5] https://software.belle2.org/sphinx/release-05-02-18/online_book/fundamentals/02-datataking.html