Sensitivity studies of rare $B^+ \to K^+\tau^+\tau^-$ decays at Belle II

Rosalía Soto-Rey

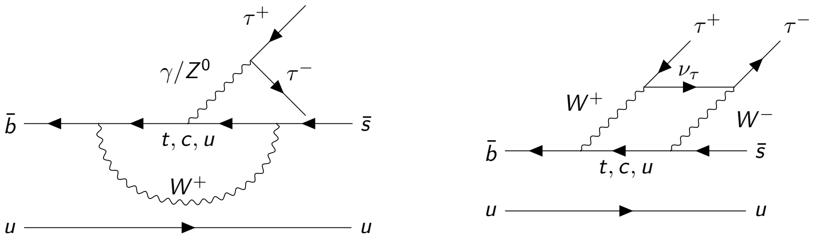
Supervisors: Niharika Rout and Debjit Ghosh

DESY Summer Student Programme 2025

September 11, 2025

$B^+ \to K^+\tau^+\tau^-$ decays

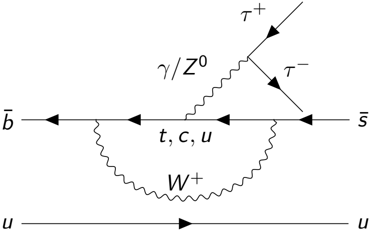
$$B^+ \to K^+\tau^+\tau^-$$

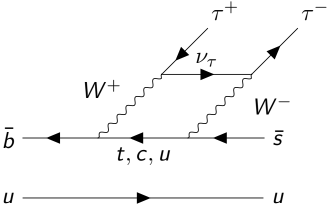


Standard Model prediction: $\mathcal{B}_{SM} = (1.49 \pm 0.10) \cdot 10^{-7}$

Motivation

$$B^+ \to K^+\tau^+\tau^-$$





$$\mathcal{B} > \mathcal{B}_{SM}?$$

- Potentially sensitive to non-SM particles

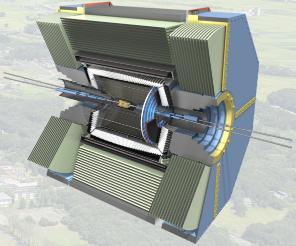
- Recent B anomalies $\rightarrow$ ▶ $R_{D^{(*)}}^{\tau/\ell}$ ▶ $B^+ \to K^+\nu\bar{\nu}$

Experimental Aspects

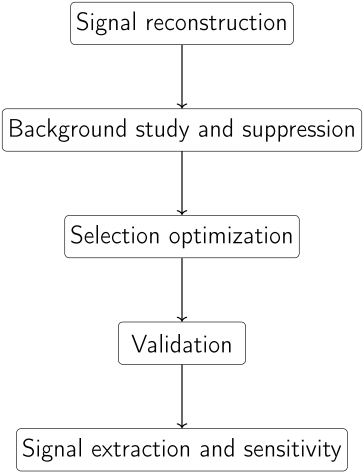
- Only published search available from BaBar experiment in 2017
  - No evidence for a signal
  - $\mathcal{B} < 2.25 \cdot 10^{-3}$ (90 % CL)
- Experimentally challenging:
  - Rare decay
  - No signal-peaking structure due to undetected multiple ν

The Belle II Experiment at SuperKEKB

- e^+e^- energy-asymmetric collider
- The Belle II detector is installed at SuperKEKB interaction point
- Belle II has been designed to:
  - Measure EW parameters precisely
  - Study exotic hadrons
  - Search for BSM phenomena
- Belle II's advantage: on-threshold $B\bar{B}$ production, precise initial known kinematics and full event reconstruction



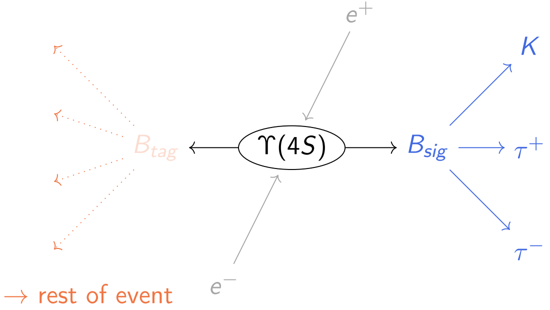
Analysis Flow



Reconstruction

Untagged Analysis

Conventional exclusive hadronic tagging: significant background reduction, but prone to low signal efficiency

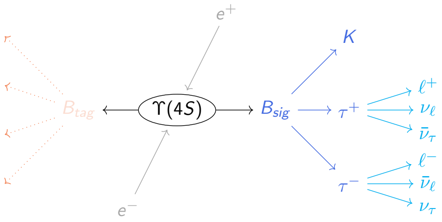


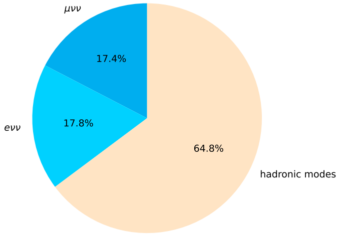
No restriction on the other B decay modes → higher ϵ_s

→ but this also increases background level

Untagged Analysis

Focus on leptonic decays $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$



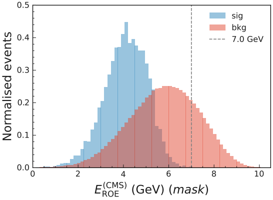


These represent $\sim 35\ \%$ of all τ decays

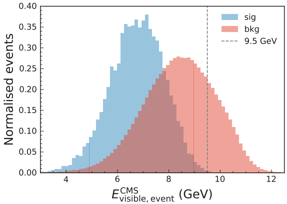
Background study and suppression

Preliminary Backgound Rejection

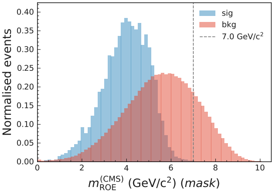
- $E_{\text{ROE}}^{(\text{CMS})}$: rest–of–event energy in CMS frame



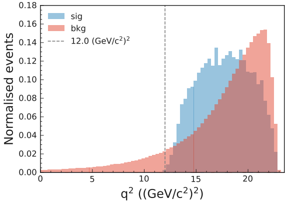
- $E_{\text{visible}}^{(\text{CMS})}$: visible energy of event (CMS)



- $m_{\text{ROE}}^{(\text{CMS})}$: rest–of–event mass in CMS frame



- q^2: invariant mass squared of the τ pair

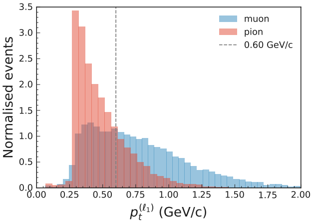


33 % background rejection with no signal loss (ϵ_s = 3.69 %)

Suppress Fake Leptons from Pions

- $e - \pi$ misidentification for $p_t \sim 0.3$ GeV/c
- $\mu - \pi$ misidentification for $p_t \lesssim 0.6$ GeV/c

  $\rightarrow$ tracks entering muon chamber



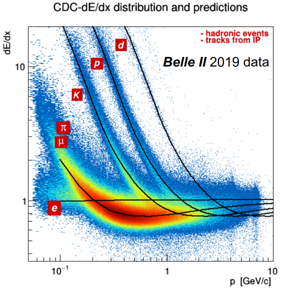
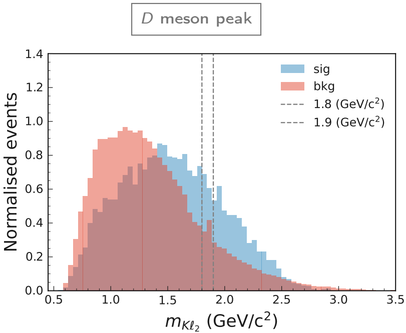
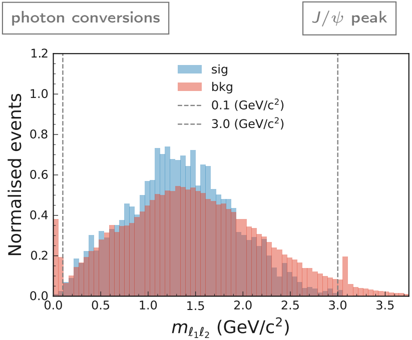


Figure: M. Milesi, Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers, CHEP 2019

75 % background rejection with $\epsilon_s = 2.03$ %

Selections on Invariant Mass Distributions





77% background rejection with $\epsilon_s = 1.89$ %

Outcome

- After all these selections, we obtain:
  - Reasonable signal efficiency: $\epsilon_s = (1.8870 \pm 0.0019)\%$
  - High background rejection: $(77.4404 \pm 0.0068)\%$

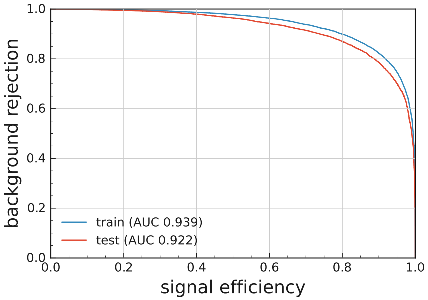
    - 26 % continuum and $\tau^+\tau^-$ background events
    - 74 % $B\bar{B}$ background events

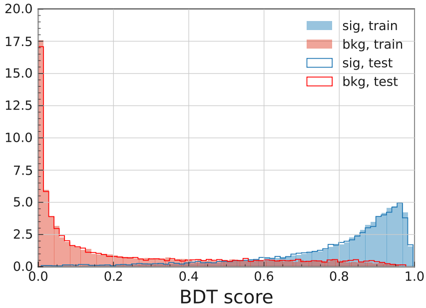
- Apply multidimensional selections using boosted-decision tree (BDT) statistical technique instead of one-dimensional selection to suppress remaining background

BDT training and performance

BDT Performance

We train a BDT to improve the separation signal/background





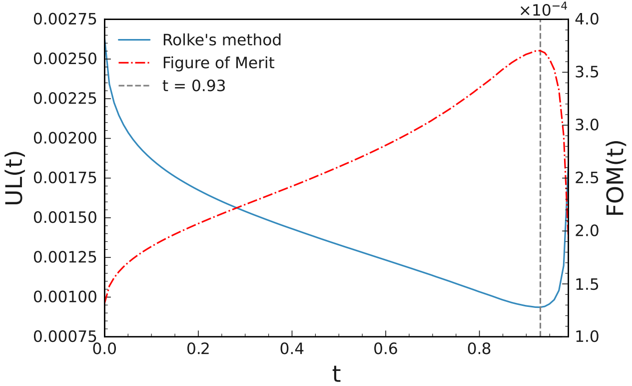
✓ Relatively high AUC value with less overtraining

✓ Good separation between signal and background

Selection on BDT Output

[Evaluation on 200 fb^{-1}]

$FOM = \frac{S}{\sqrt{S+B}}$



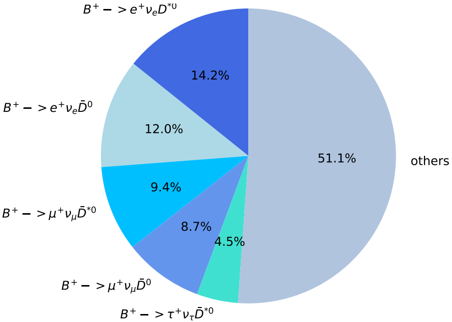
Signal region: t > 0.93

99.83% background rejection with $\epsilon_s = (0.3607 \pm 0.008)$ %

Signal region in simulated sample

Background Composition in the Signal Region

Signal region: 101 903 background events (85 % $B\bar{B}$, 15 % $q\bar{q}$ and τ pair)



Cascade decay channels mainly with

$K^+\ell^-\bar{\nu}_\ell$

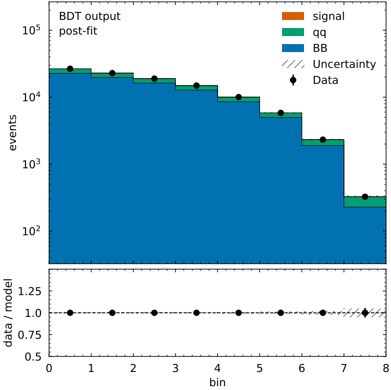
$+$

neutrals $(\pi^0, \gamma \dots)$

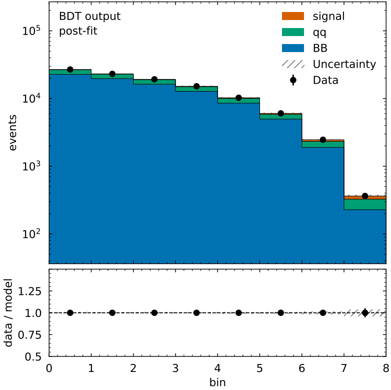
in the final state

Fit and Signal Injection

Template fit in 8 bins in 3 categories (signal + $B\bar{B}$ + $q\bar{q}$)

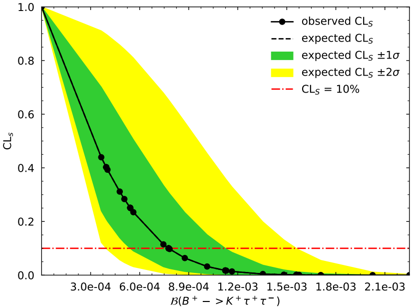


Assuming SM branching fraction



Assuming $\mathcal{B} = 2 \cdot 10^{-3}$

Upper limit (stat-only) on $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$



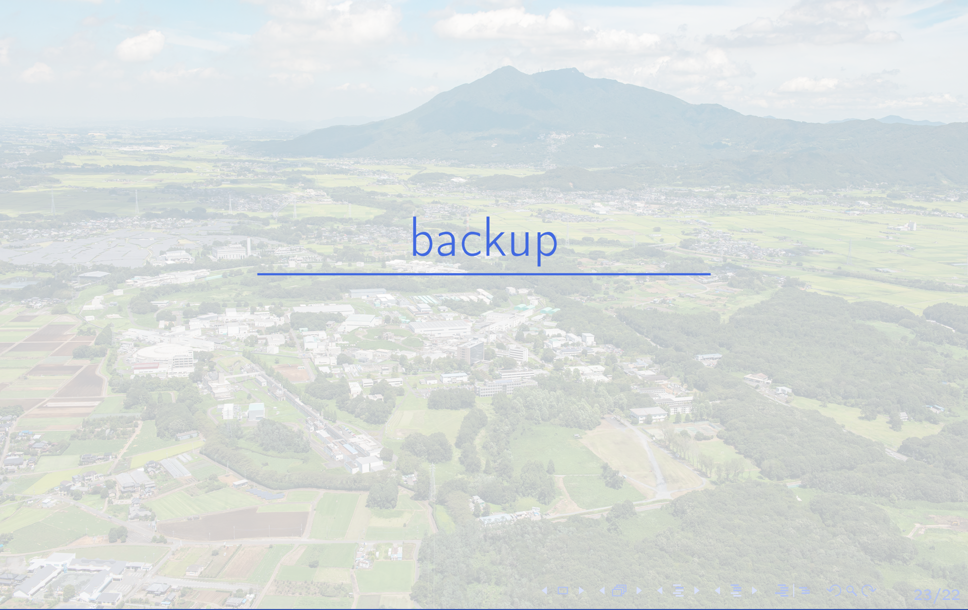
Projection assuming 424 fb^{-1}:

$$\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 5.3 \cdot 10^{-4} \ (90\ \%\ \mathrm{CL})$$

$\rightarrow$ improvement by about a factor of 4 relative to BaBar's analysis

Summary

- $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ is sensitive for search of New Physics
- Untagged analysis: no restriction on the non-signal B decay mode $\to$ higher signal efficiency, but also larger background
- Performance set by signal efficiency vs background rejection
  - Preselections already achieve strong background suppression
  - Further multidimensional selections applied via a BDT
  - BDT output $\to$ definition of signal region $\to$ fit
- UL (stat-only) on $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ approximately four times better compared to the world-leading result from BaBar experiment



backup

Branching Fraction ($\mathcal{B}$)

$$\mathcal{B} = \frac{N_{obs} - N_{exp}}{2\epsilon_s f^{+-} N_{B\bar{B}}}$$

Where ϵ_s is the signal efficiency:

$$\epsilon_s = \frac{\text{number of true events after the selections}}{\text{N}}$$

$$s(\epsilon_s) = \sqrt{\frac{\epsilon_s(1-\epsilon_s)}{N}}$$

$N \equiv$ number of generated events

Figure of Merit (FOM)

As a figure of merit, we are considering:

$$FOM = \frac{S}{\sqrt{S + B}}$$

Where $S(t)$ and $B(t)$ are the numbers of signal and background events surviving the selection (t).

$$S(t) = \epsilon_s(t) \cdot \text{expected signal}$$

$$\text{expected signal} = 2f^{+-} N_{B\bar{B}}\ \mathcal{B}_{SM} \approx N_{B\bar{B}}\ \mathcal{B}_{SM}$$

Leptoquarks

Leptoquarks are color-triplet bosons
that carry both lepton and baryon numbers

$\rightarrow$ they can mediate interactions between quarks and leptons

✓ Leptoquarks offer a plausible interpretation of B anomalies

✓ Not yet directly observed $\rightarrow$ compatible with the predicted mass of $\mathcal{O}(1)$ TeV

Selections at Reconstruction Level

50M signal generated events & 200 fb^{-1} generic Monte Carlo

- Combine one kaon and two leptons to form signal B candidates
  - Cuts on impact parameter to suppress beam background

    `|dr| < 0.5 cm` `|dz| < 2 cm` `nCDCHits > 20` CDC polar angle acceptance
  - Better particle identification

    `binaryKpi > 0.6` `muonID > 0.9` `electronID > 0.9`
- All remaining tracks and clusters are assigned to the ROE
  - Tracks: `|dr| < 0.5 cm` `|dz| < 2 cm` CDC polar angle acceptance
  - Clusters: `energy` $>$ 100 `MeV (FWD)`, $>$ 55 `MeV (BRL)`, $>$ 150 `MeV (BWD)`, `cluster-to-track distance` $>$ 30 `cm`

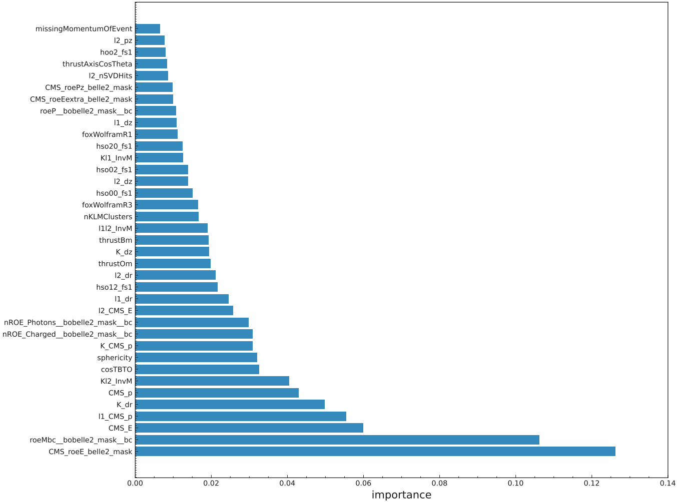
These selections reduce beam background for ROE

BDT Training

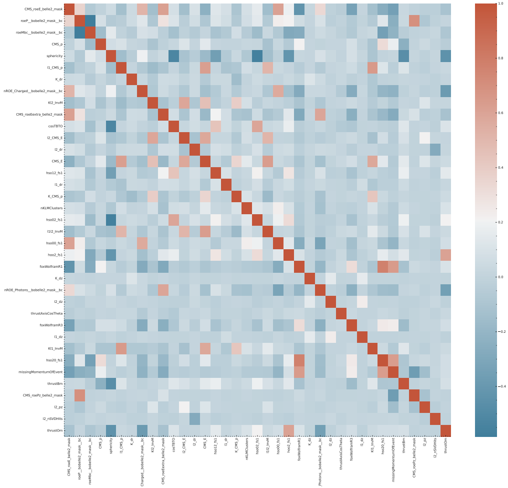
[50M signal generated events & 200 fb^{-1} generic MC]

- `XGBoost Classifier`
- Training sample: $\sim 2 \times 10^4$ signal and background events
- Evaluation: independent test sample
- Input features: 37 training variables
- Optimized hyperparameters

Feature Importances



Correlation Matrix



Highly-correlated variables are excluded

Hyperparameters Optimization

✓ Better BDT performance

✓ Reduce overtraining

Hyperparameter	Description	Value
Learning rate	Scaling factor of each tree's contribution	0.09772
Number of estimators	Number of gradient boosted trees	310
Maximum depth	Maximum depth of a tree	3
Subsample	Fraction of training events used per tree, can help prevent overtraining	0.58165

The Belle II Detector