With funding from the:
Federal Ministry of Research, Technology and Space

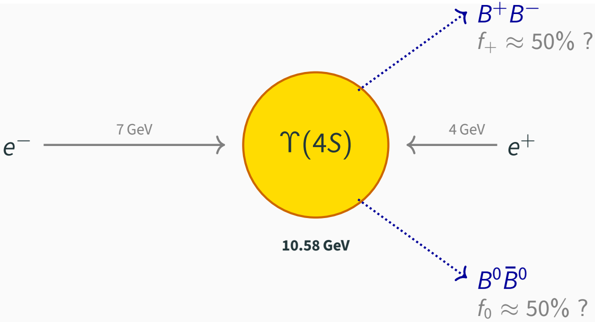
Measurement of the energy dependence of strong isospin violation in $\Upsilon(4S)$ decays using a neural network event classifier

Lena Nowatzki, Thomas Kuhr, Thomas Lück

ML4Jets Vienna 2026

September 14, 2026

B Factories and $\Upsilon(4S)$



- e^+e^- collisions at $\sqrt{s} \approx 10.58$ GeV produce the $\Upsilon(4S)$ resonance
- $m_{B^\pm} \approx 5.27941$ GeV,
  $m_{B^0} \approx 5.27972$ GeV
  $2m_B \approx 10.56$ GeV — just below $\sqrt{s}$
  $\Rightarrow$ decays almost exclusively to $B\bar{B}$

Key assumption

Many B branching fraction measurements assume equal production: $R^{\pm 0} = f_+/f_0 = 1$
If wrong $\Rightarrow$ systematic bias on all these measurements

Isospin Violation in $\Upsilon(4S)$ Decays

Isospin symmetry predicts $R^{\pm 0} = \frac{\Gamma(\Upsilon(4S)\to B^+B^-)}{\Gamma(\Upsilon(4S)\to B^0\bar{B}^0)} = 1$

$$R^{\pm 0} = R^{\pm 0}_{PS} \cdot R^{\pm 0}_{CE}$$

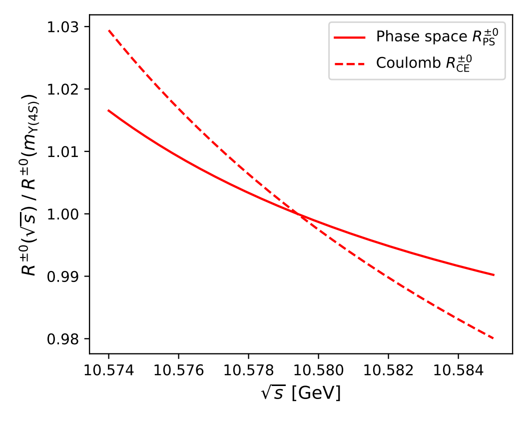
Mass difference ($m_{B^+} \neq m_{B^0}$):

$$R^{\pm 0}_{PS} = \frac{p_{\pm}^3}{p_0^3}$$

Coulomb interaction (B^+B^- charged):

$$R^{\pm 0}_{CE} = \frac{2\pi\lambda(1+\lambda^2)}{1-\exp(-2\pi\lambda)}$$

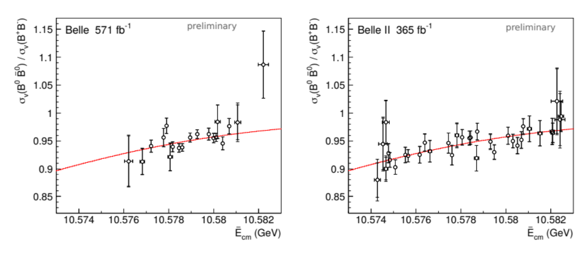
where $\lambda = \alpha/(2v_{\pm})$ and $v_{\pm} = (1 - 4m^2_{B^{(*)\pm}}/m^2_{\Upsilon})^{1/2}$



Prior Measurement

Belle II Analysis:

- Exclusive reconstruction in $\sim$ 1000 hadronic final states
- $R^{\pm 0}(E_{\mathrm{cm}})$ measured in 0.2 MeV bins



Belle II (2025), arXiv:2511.15926

Note: plot shows the **inverted** ratio $\sigma_V(B^0\bar{B}^0)/\sigma_V(B^+B^-)$

This work

Can we use **all** $B\bar{B}$ events?

$\Rightarrow$ Complementary inclusive approach using a neural network

Inclusive Approach

Idea:

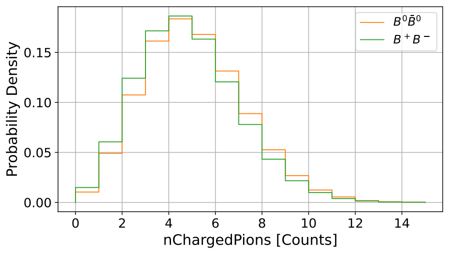
- Global event-level variables separate B^+B^- from $B^0\bar{B}^0$ — no decay is reconstructed
- Neural network classifier trained on MC
- Extract $R^{\pm 0}$ from a template fit to the classifier output

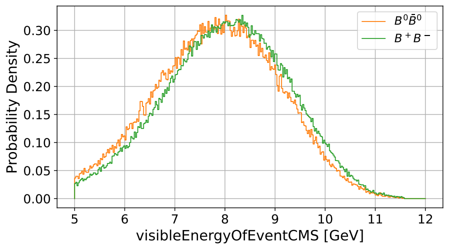
Event variables → Neural network → Template fit → $R^{\pm 0}$

Input Features

26 event-level variables in three categories:

- Particle multiplicities (nTracks, nChargedPions, ...)
  charge conservation: $B^{\pm} \to$ odd, $B^0 \to$ even
- Event kinematics (p_T, E_{vis}, ...)
- Event shape (thrust, sphericity, Fox-Wolfram moments, ...)
  $B\bar{B}$ isotropic (B nearly at rest), background jet-like, aligned along the thrust axis

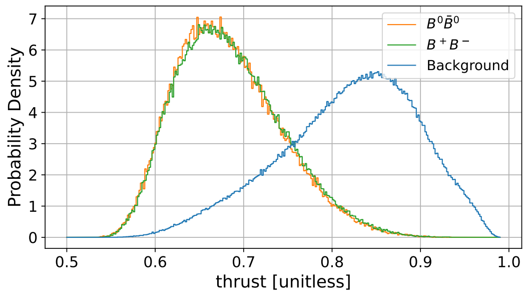




Input Features

26 event-level variables in three categories:

- Particle multiplicities (nTracks, nChargedPions, ...)
  charge conservation: $B^{\pm} \rightarrow$ odd, $B^{0} \rightarrow$ even
- Event kinematics (p_T, E_{vis}, ...)
- Event shape (thrust, sphericity, Fox-Wolfram moments, ...)
  $B\bar{B}$ isotropic (B nearly at rest), background jet-like, aligned along the thrust axis



Background: Non-$B\bar{B}$ events

$$e^+e^- \rightarrow q\bar{q}\ (q = u, d, s, c), \tau^+\tau^-$$

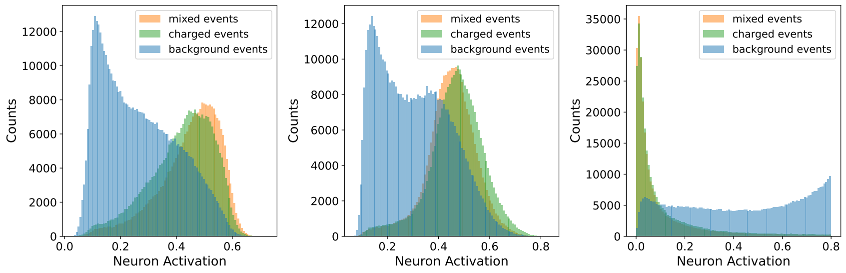
Neural Network Classifier — First Approach

Architecture

- $26 \to 126 \to 65 \to 100 \to 3$
- dropout $\sim 10\%$, batch norm, ReLU
- trained with cross-entropy loss
- hyperparameters via Optuna
  Score $= \frac{1}{2}(F_1 + F_1^{\text{mod}})$

Training

- MC16, statistically independent subsets for training / templates / fit ($\sim 0.5\,\text{fb}^{-1}$ each)
- background $\sim 80\%$ of events $\Rightarrow$ training set reweighted



Template Fit

Method:

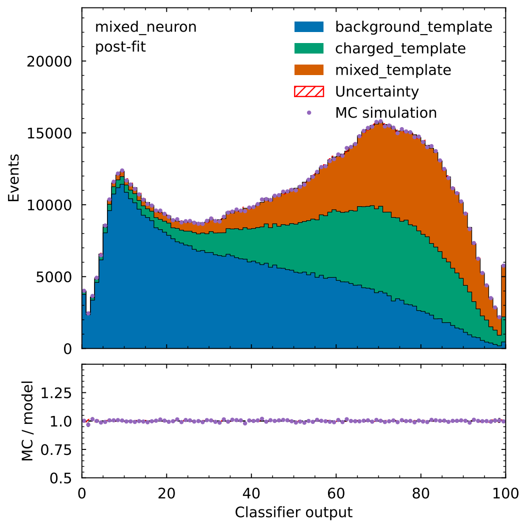
- Fit NN output as weighted sum of the three MC templates
- `pyhf`: `normfactor` per template, `shapesys` per bin (MC statistics)

$$R^{\pm 0} = \frac{\mu_{B+} \cdot N_{B+}^{\text{template}} / \varepsilon_{B+}}{\mu_{B0} \cdot N_{B0}^{\text{template}} / \varepsilon_{B0}}$$

μ_s: normalization, ε_s: efficiency, N_s^{template}: template yield

Result on MC: $R^{\pm 0} = 1.06 \pm 0.04$

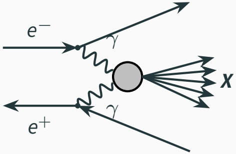
True value: $R^{\pm 0} = 1.0626$



MC/Data Disagreements

Applied corrections:

- Differential track momentum scale and energy loss corrections
- Photon-energy bias correction
- $E_{\mathrm{vis}} > 5\,\mathrm{GeV} \rightarrow$ removes a two-photon background not included in MC

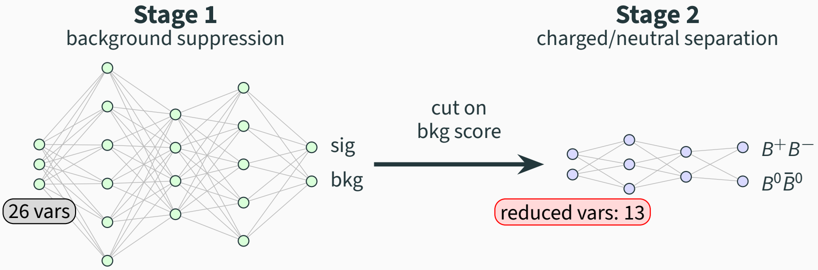


Two-photon process $e^+e^- \rightarrow e^+e^-X$:
$e^{\pm}$ escapes undetected, X produced at the vertex

Remaining challenge:

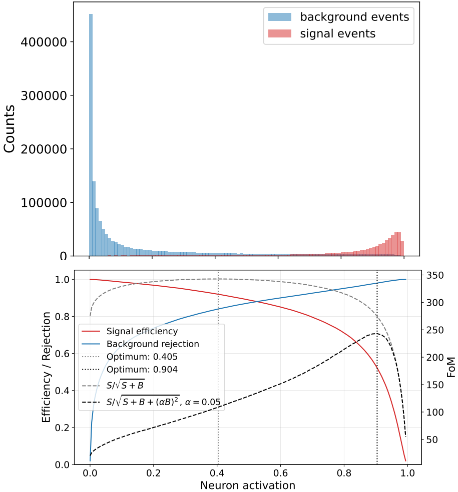
- Residual disagreement concentrated in the **event-shape variables**
- Event-shape variables are essential for **background suppression** but carry little charged/neutral information $\Rightarrow$ split the two tasks into two separate steps

Two-Stage Classifier



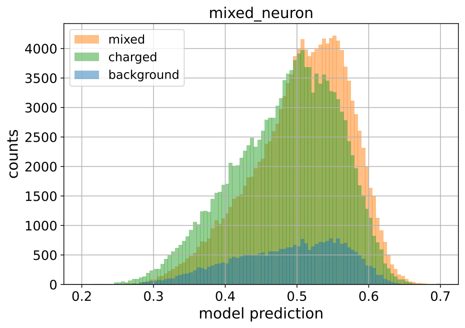
- **Stage 1**: three-class architecture, all 26 variables, AUC $= 0.95$; cut at 0.9 raises background rejection 85% $\rightarrow$ 98% (signal efficiency $\sim$50%)
- **Stage 2**: architecture from a KL-divergence-ranked scan, event-shape variables removed

Stage 1: Cut Selection



- S_1 peaks at 0.405, S_2 ($\alpha = 0.05$) at 0.904
- S_2 penalizes residual background, which is the source of the data/MC disagreement
- chosen 0.9: rejection 85% $\rightarrow$ 98%, signal efficiency $\sim$50%

Stage 2: Architecture Selection



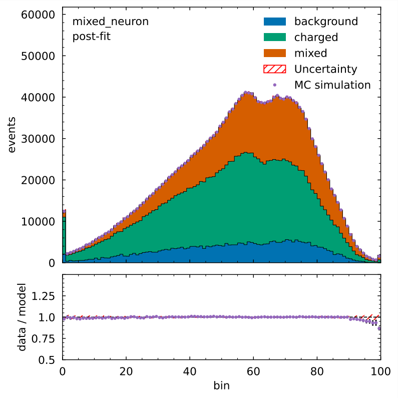
B^+B^- and $B^0\bar{B}^0$ differ only in charged-particle content — separation is limited by the physics, not the network (AUC $= 0.59$)

Architecture scan

- small grid search (120 configurations)
- Ranked by **KL divergence** between the charged and mixed output distributions $\rightarrow$ the fit uses template *shapes*, so rank on shape separation

chosen: rank-3 network — linear shape, 2 layers, max width 32

Result (MC Closure)



MC closure same fit setup as Stage 1

- $R^{\pm 0} = 1.11 \pm 0.05$, true value $1.0626 \rightarrow$ consistent
- background normalization is only weakly constrained

Pull study — 1000 toys

- pull widths consistent with unity $\rightarrow$ **fit is unbiased**

Summary

- Inclusive method: event-level classifier + template fit uses **all** $B\bar{B}$ events — no exclusive reconstruction, no channel-specific assumptions
- Two-stage design isolates the mismodelled event-shape variables from the stage that measures $R^{\pm 0}$
- MC closure: $R^{\pm 0} = 1.11 \pm 0.05$ (true 1.0626), pulls consistent with unity $\rightarrow$ fit unbiased
- Per-event separation is weak (AUC $= 0.59$) — limited by the physics, not the network
- Limitation is **systematic**, not statistical

Outlook

- Data-driven background templates (exclusive channel such as $B \to J/\psi K$, or Full Event Interpretation) — targets the dominant systematic at its origin

- Statistically $\sim$0.13% reachable on $\sim$700 fb^{-1}
  most precise existing measurement: 1.1% statistical
  Belle, 711 fb^{-1}, arXiv:2207.01194

- Achievable precision vs. number of energy bins still to be studied

Outlook

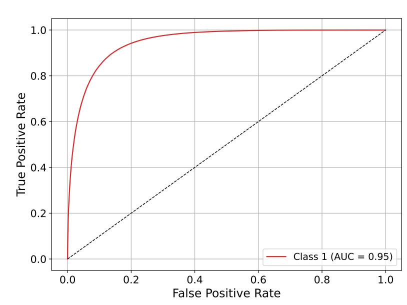
- Data-driven background templates (exclusive channel such as $B \to J/\psi K$, or Full Event Interpretation) — targets the dominant systematic at its origin
- Statistically $\sim 0.13\%$ reachable on $\sim 700\,\mathrm{fb}^{-1}$
  most precise existing measurement: 1.1% statistical
  Belle, $711\,\mathrm{fb}^{-1}$, arXiv:2207.01194
- Achievable precision vs. number of energy bins still to be studied

Thank you!

Backup

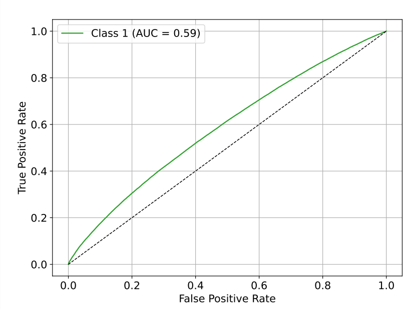
Classifier Performance

Stage 1 — background suppression



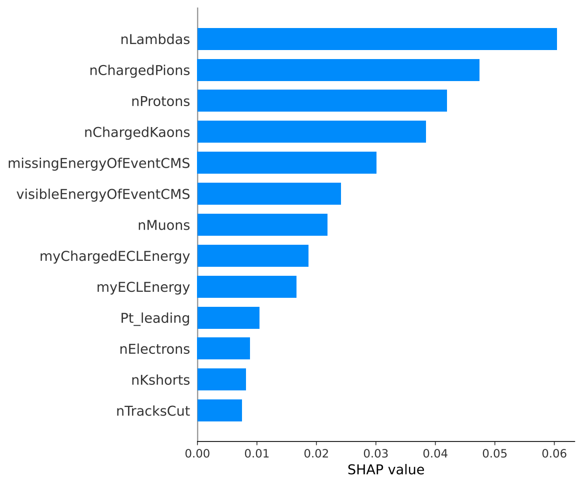
$\text{AUC} = 0.949$, $F_1 = 0.852$, $\kappa = 0.706$

Stage 2 — charged vs. mixed



$\text{AUC} = 0.59$ — separation limited by the physical similarity of the two event types

Feature Importance — Stage 2

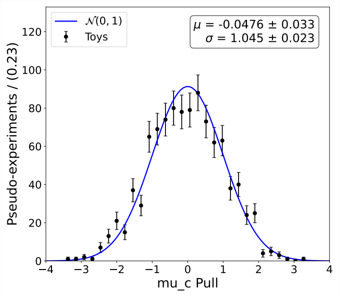


Mean absolute SHAP values, 13 input variables

- `nLambdas` and `nChargedPions` dominate
- consistent with charge conservation: $B^{\pm} \to$ odd, $B^0 \to$ even number of charged particles
- multiplicities outrank kinematics throughout

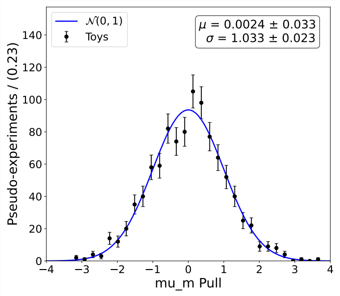
Toy Studies — Fit Validation

1000 pseudo-experiments generated from the fitted model and refitted; $\mathrm{pull}_i = (\hat{\mu}_i - \mu_i^{\mathrm{true}})/\sigma_{\hat{\mu}_i}$



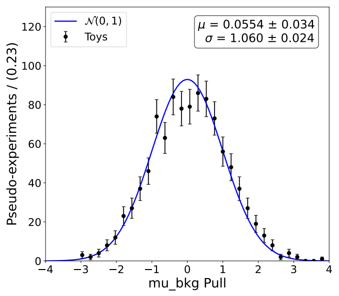
μ_{c} — charged

mean -0.048, width 1.045



μ_{m} — mixed

mean 0.002, width 1.033

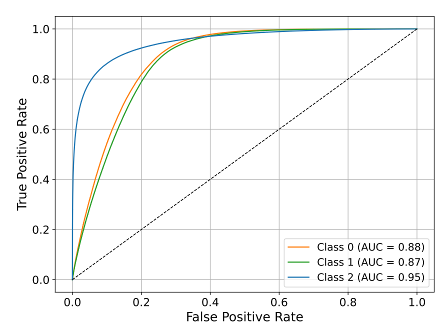


μ_{bkg} — background

mean 0.055, width 1.060

Means consistent with zero, widths within a few percent of unity $\Rightarrow$ fit procedure unbiased. The background parameter deviates most, consistent with the residual mismodelling of its template.

Three-Class Classifier



Metric	Value
Precision	0.587
Recall	0.629
F_1	0.602
ROC	0.902
κ	0.506
F_1 (charged vs. mixed)	0.595
ROC (charged vs. mixed)	0.559

Class Good background separation, weak charged/mixed separation — and the event-shape variables it needs for the former are the mismodelled ones

Stage 2 Architecture Scan

Shape	Layers	Max width	KL divergence	Accuracy	AUC
bottleneck	2	128	0.0472	0.5592	0.5861
diamond	3	64	0.0468	0.5601	0.5858
linear	**2**	**32**	**0.0466**	**0.5598**	**0.5851**
linear	5	32	0.0464	0.5600	0.5850
diamond	2	64	0.0462	0.5600	0.5853

120 configurations: 5 shapes $\times$ layer counts $\{2, 3, 4, 5, 7\}$ $\times$ max widths $\{32, 64, 128, 256\}$

Ranks 1 and 2 discarded: background output too flat to fit without constraining its normalisation to $\pm 20\%$

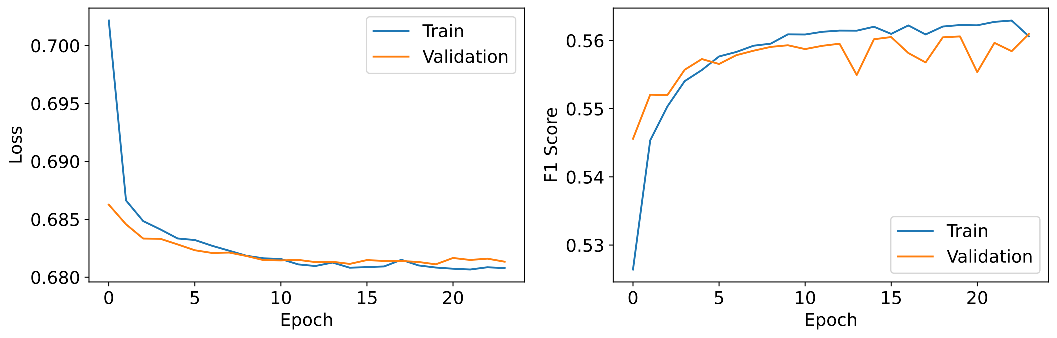
Input Variables

Category	Variables
Particle multiplicities	`nTracksCut`, `nChargedPions`, `nChargedKaons`, `nElectrons`, `nMuons`, `nProtons`, `nKshorts`, `nLambdas`
Event kinematics	`Pt_leading`, `myECL`, `myChargedECL`, `visibleEnergyOfEventCMS`, `missingEnergyOfEventCMS`
Event shape	`thrust`, `thrustAxisCosTheta`, `sphericity`, `aplanarity`, `foxWolframR2`, `cleoConeThrust1-8`

Stage 1 uses all 26; Stage 2 drops the 13 event-shape variables

Training — Stage 2

- Training and validation loss converge smoothly and track each other
- Validation F_1 fluctuates — expected given the small separation between the two classes
- No significant overfitting; early stopping after $\sim$20 epochs



The Belle II Environment

- B mesons nearly at rest
  $\Rightarrow$ isotropic decays;
  continuum ($\sim$80% of events)
  is jet-like

$B\bar{B}$ — isotropic   continuum — jet-like