

Search for lepton flavour violation in $B^0 \rightarrow \rho^0 \tau^\pm \ell^\mp$ decays at Belle and Belle II and tracking performance studies

PhD defence

Isaac-Paul Consigny

Supervisors: Justine Serrano, Giampiero Mancinelli

24 September 2026

Centre de Physique des Particules de Marseille
Aix-Marseille Université, CNRS/IN2P3



Outline

Theoretical context

Experimental setup: Belle and Belle II

Tracking performance study

Search for $B^0 \rightarrow \rho^0 \tau^\pm \ell^\mp$

An alternative tagging approach: graFEI

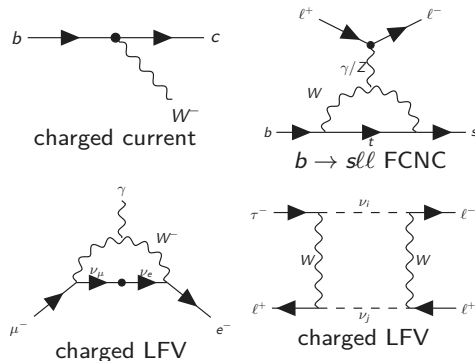
Theoretical context

- Isaac-Paul Consigny –
- $B^0 \rightarrow \rho^0 \tau \ell$
- at Belle and Belle II

	three generations of matter (fermions)			interactions / force carriers (bosons)	
QUARKS	I $\approx 2.16 \text{ MeV}/c^2$ $\frac{2}{3}$ $\frac{1}{2}$ up	II $\approx 1.273 \text{ GeV}/c^2$ $\frac{2}{3}$ $\frac{1}{2}$ charm	III $\approx 172.57 \text{ GeV}/c^2$ $\frac{2}{3}$ $\frac{1}{2}$ top	 gluon	$\approx 125.2 \text{ GeV}/c^2$ 0 0 0 higgs
	$\approx 4.7 \text{ MeV}/c^2$ $-\frac{1}{3}$ $\frac{1}{2}$ down	$\approx 93.5 \text{ MeV}/c^2$ $-\frac{1}{3}$ $\frac{1}{2}$ strange	$\approx 4.183 \text{ GeV}/c^2$ $-\frac{1}{3}$ $\frac{1}{2}$ bottom	 photon	SCALAR BOSONS GAUGE BOSONS VECTOR BOSONS
	$\approx 0.511 \text{ MeV}/c^2$ $-\frac{1}{2}$ $\frac{1}{2}$ electron	$\approx 105.66 \text{ MeV}/c^2$ $-\frac{1}{2}$ $\frac{1}{2}$ muon	$\approx 1.77693 \text{ GeV}/c^2$ $-\frac{1}{2}$ $\frac{1}{2}$ tau	$\approx 91.188 \text{ GeV}/c^2$ 0 1 Z boson	
$< 0.8 \text{ eV}/c^2$ 0 $\frac{1}{2}$ electron neutrino	$< 0.17 \text{ MeV}/c^2$ 0 $\frac{1}{2}$ muon neutrino	$< 18.2 \text{ MeV}/c^2$ 0 $\frac{1}{2}$ tau neutrino	$\approx 80.3692 \text{ GeV}/c^2$ ± 1 1 W boson		

Flavour in the Standard Model

- ▶ **Quark flavour** not conserved: $W^\pm$ currents, up-type $\Leftrightarrow$ down-type
- ▶ **FCNC** (Flavour Changing Neutral Current), suppressed in the SM $\Rightarrow$ sensitive to new physics
- ▶ **Lepton flavour** conserved in the SM
- ▶ **Neutrino oscillations** (1998) $\Rightarrow$ neutrinos have a mass $\Rightarrow$ **Lepton Flavour Violation (LFV)** in the neutral sector
- ▶ **Charged LFV** then possible through neutrino loops, but $\mathcal{B}(\mu \rightarrow e \gamma)_{\text{SM}} \sim 10^{-50}$
- ▶ **Any observation of charged LFV would be a clear sign of new physics.**



Quark FCNC vs. charged LFV

$b \rightarrow s \ell^+ \ell^-$: loop level, suppressed but allowed
 $b \rightarrow s \tau \ell$: forbidden at any observable level

Looking for charged LFV beyond the Standard Model

► Flavour anomalies:

- $R_{D^{(*)}} = \frac{\mathcal{B}(B \rightarrow D^{(*)} \tau \nu)}{\mathcal{B}(B \rightarrow D^{(*)} \ell \nu)}, \quad \ell = e, \mu$
 $\Rightarrow 3.7\sigma$ above SM
- first evidence of $B^+ \rightarrow K^+ \nu \bar{\nu}$ at Belle II,
 $\mathcal{B} \sim 2.7\sigma$ above SM

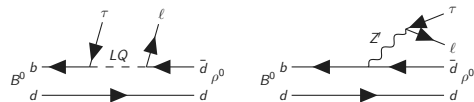
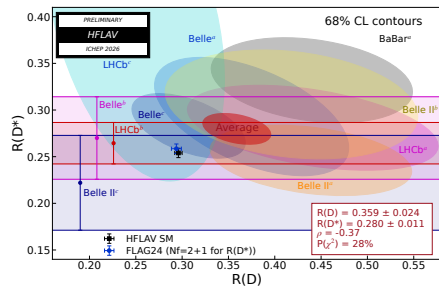
► New Physics mediators: **Leptoquarks** and **Z'**

→ Can address $R_{D^{(*)}}$ and $B \rightarrow K^{(*)} \nu \bar{\nu}$ discrepancies

→ often couple preferentially to τ

► Predicted $b \rightarrow d \tau \ell$ branching fractions up to 10^{-7} [Bečirević, Jaffredo et al., PRD 110, 075004 (2024)]

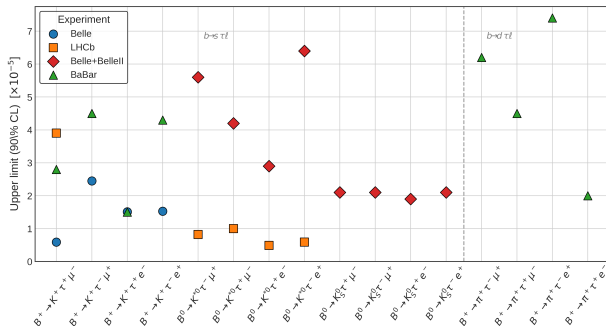
► $B^0 \rightarrow \rho^0 \tau \ell$ interesting because FCNC $b \rightarrow d$ transition and τ involved



possible NP diagrams for $B^0 \rightarrow \rho^0 \tau \ell$

Experimental status of LFV

- ▶ μ and τ sectors: $\mathcal{B}(\mu \rightarrow e\gamma) < 10^{-13}$ (MEG II), $\mathcal{B}(\tau \rightarrow \ell\gamma, 3\ell) < 10^{-8}$ (Belle, Belle II, LHC)
- ▶ B LFV decays to light leptons: limits at 10^{-9}
- ▶ B LFV decays with a τ : limits at 10^{-5} because of the undetectable neutrinos
- ▶ $b \rightarrow s\tau\ell$: $B \rightarrow K^{(*)}\tau\ell$ at Belle, Belle II, LHCb
- ▶ $b \rightarrow d\tau\ell$: only $B^+ \rightarrow \pi^+\tau\ell$ by BaBar, $2.0\text{--}7.4 \times 10^{-5}$



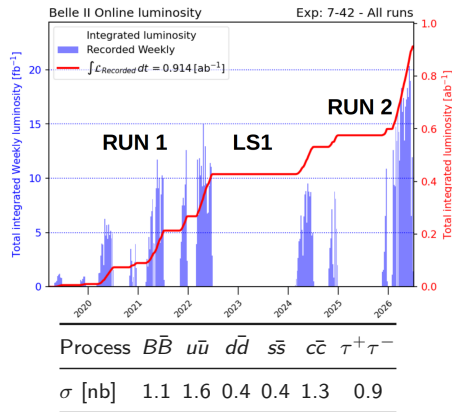
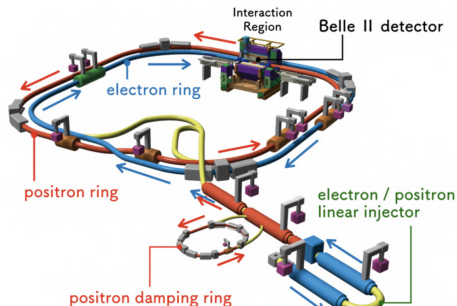
This thesis

First search for $B^0 \rightarrow \rho^0 \tau^\pm \ell^\mp$, $\ell = e, \mu$, at Belle and Belle II

Belle and Belle II

From KEKB to SuperKEKB

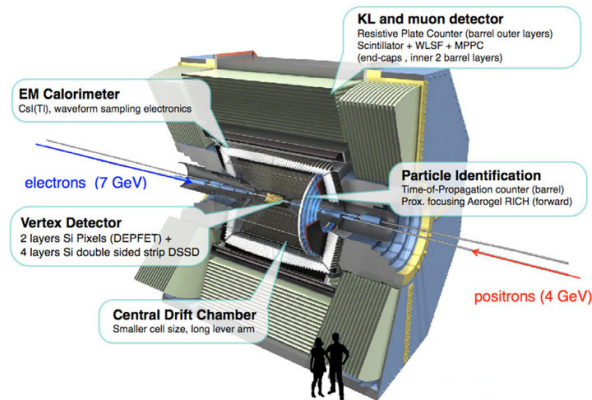
- ▶ KEKB (1999–2010) → SuperKEKB (2019–): e^+e^- colliders
- ▶ $\sqrt{s} = 10.58 \text{ GeV} \Rightarrow \Upsilon(4S) \rightarrow B\bar{B}$
- ▶ **Nano-beam scheme:** very small beam sizes at the interaction point
- ▶ peak luminosity $L_{\text{peak}} = 5.3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
target $\Rightarrow \sim 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$



- ▶ 711 fb^{-1} from Belle and 365 fb^{-1} from Belle II Run 1, total: $\sim 1.160 \times 10^9 B\bar{B}$

The Belle and Belle II detectors

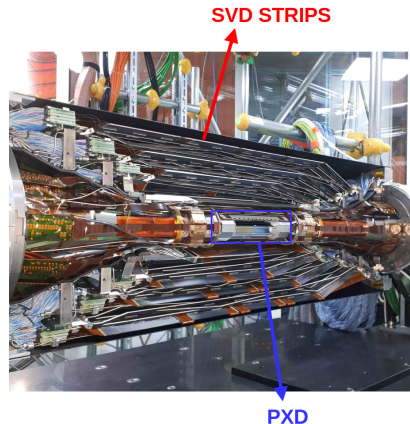
- ▶ Belle II: upgraded version of Belle, upgraded to tackle $\times 30$ more instantaneous luminosity and increased background
- ▶ Precise **vertexing and tracking**: Vertex detector
- ▶ Excellent **particle identification** (PID) based on Cherenkov radiation, calorimetry, and muon detection
- ▶ **Hermetic** detector with nearly full 4π coverage and precisely known **initial state** $\Rightarrow$ **the event's missing energy can be reconstructed** – essential for modes with neutrinos



Belle II

The tracking system at Belle II

- ▶ Charged particles $\Rightarrow$ hits in PXD, SVD, CDC $\Rightarrow$ **tracks**
- ▶ **PXD**: Pixel detector, $\sigma \sim 10\mu m$
- ▶ **SVD**: double sided silicon strips detector, $\sigma \sim 20\mu m$
- ▶ **CDC**: Central Drift Chamber, $\sigma \sim 100\mu m$
- ▶ **Magnet**: 1.5 T solenoid $\Rightarrow$ track curvature $\Rightarrow$ momentum measurement



- ▶ The SVD also **timestamps its hits** with a resolution of ~ 3 ns: **track time**

Tracking performance study

Motivation: reduce the fake tracks rate

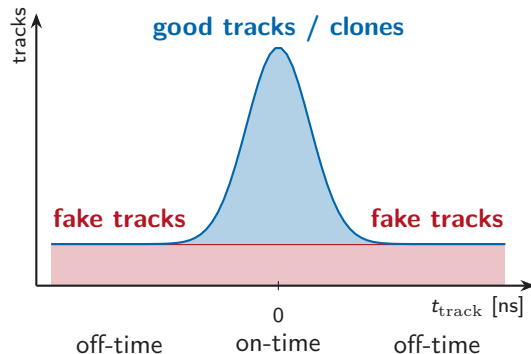
- ▶ **Fake tracks:** tracks not from the collision: beam background hits, wrong hit combinations, *clones* of real tracks
- ▶ **Higher luminosity** $\Rightarrow$ higher occupancy and beam background $\Rightarrow$ more fake tracks

Track time

$$t_{\text{track}} = \langle t_{\text{SVD clusters on track}} \rangle - \text{SVDEvent } T_0$$

SVDEvent T_0 = average time of the SVD clusters on all the tracks

- ▶ Real tracks are on-time, fakes (except clones) are off-time
- ▶ **Goal:** Find a selection on $|t_{\text{track}}|$ to suppress fake tracks with a limited impact on the tracking efficiency
- ▶ **Tracking efficiency:** probability to reconstruct a track from a real particle in the acceptance



Fake tracks rate and tracking efficiency measurements: tag-and-probe method

- ▶ Measurements in $e^+e^- \rightarrow \tau^+\tau^-$ events, $\tau^\pm \rightarrow \ell^\pm \nu \bar{\nu}$ vs. $\tau^\mp \rightarrow (3\pi)^\mp \nu$
- ▶ Clean environment, jet-like topology, large momentum range

Tracking efficiency

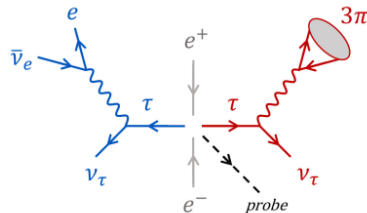
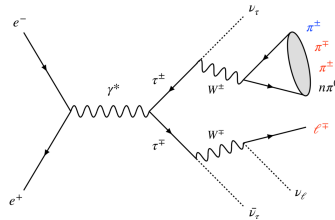
- ▶ **Tag (N3):** 3 tracks: lepton + 2 pions, total charge ± 1
- ▶ **Probe (N4):** a 4th track restoring charge conservation

$$\varepsilon \times A = \frac{N_4}{N_3 + N_4}$$

Fake-track rate

- ▶ **Tag (N4):** 4 tracks: lepton + 3 pions
- ▶ **Probe (N5):** the 5th track is a fake

$$r_{\text{fake}} = \frac{N_5}{N_4 + N_5}$$



Fake-track rate and efficiency vs. track time selection

Without track time selection

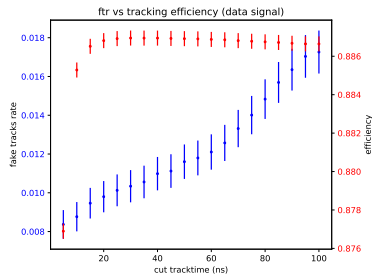
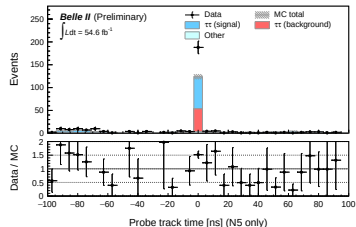
	data (signal)	MC (signal)
Fake-track rate	$2.01 \pm 0.12\%$	$1.62 \pm 0.06\%$
Tracking efficiency	$88.92 \pm 0.04\%$	$88.79 \pm 0.04\%$

Fake rate vs tracking efficiency

The track time selection **strongly reduces** the fake-track rate without affecting the efficiency.

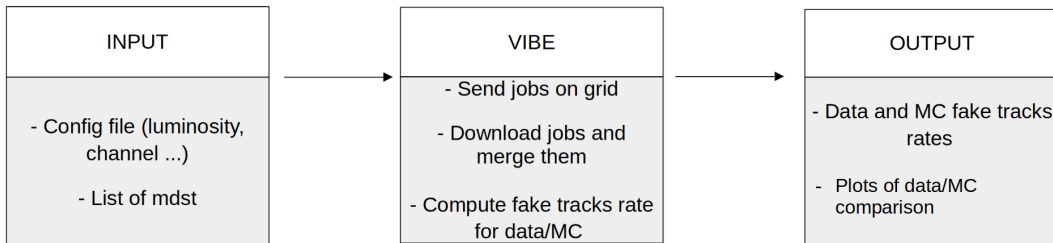
Selection $|t_{\text{track}}| < 20 \text{ ns}$

Fake-track rate reduced by a **factor ≈ 2** ($2.01 \rightarrow 0.98\%$), efficiency loss of only **0.24 points** ($88.92 \rightarrow 88.68\%$).



VIBE: automating the fake-track rate measurement

- ▶ Validation Interface for the Belle II Experiment: Python framework to execute tasks automatically as their requirements are met
- ▶ My contribution: a complete **fake-track rate** new feature: from ntuple production to measurement and data/MC plots
- ▶ Inputs: sample list + configuration (luminosities, selections, variables)
- ▶ Fully automated *and* restartable from any step
- ▶ First end-to-end VIBE feature; used for all results shown here and as a basis for future features



$B^0 \rightarrow \rho^0 \tau \ell$ analysis

Analysis strategy

Reconstruction

Background suppression

Signal extraction

Systematic uncertainties

Results

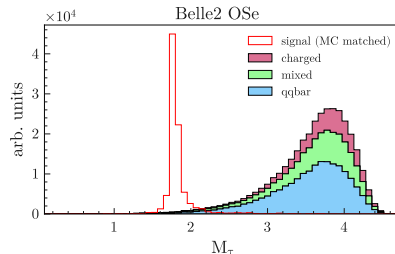
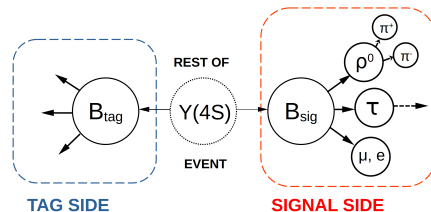
Analysis strategy

- ▶ **Hadronic B -tagging** $\Rightarrow$ flavour
- ▶ The B_{sig} has the **opposite flavour** of the B_{tag} flavour
- ▶ Four modes to analyse: (OS, SS) $\times$ (e, μ) – OS/SS: opposite/same sign between the prompt lepton and the b quark of the B_{sig}
- ▶ Signal extraction from a fit to the τ **recoil mass**:

$$M_\tau^2 = m_B^2 + m_{\rho\ell}^2 - 2(E_{\text{beam}}^* E_{\rho\ell}^* + |\vec{p}_{B_{\text{tag}}}^*| |\vec{p}_{\rho\ell}^*| \cos \theta^*)$$

- ▶ Derive upper limit at 90% CL on the branching fraction:

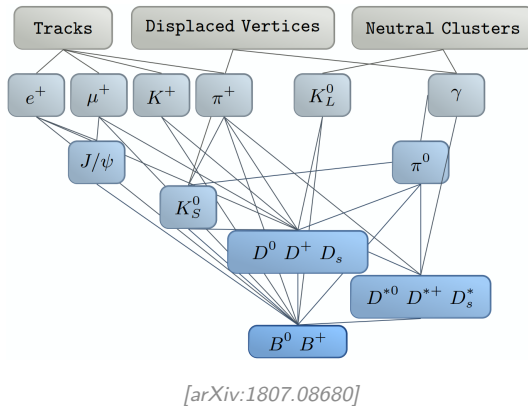
$$\mathcal{B}(B^0 \rightarrow \rho^0 \tau \ell) = \frac{N_{\text{sig}}}{2 \varepsilon_{\text{sig}} N_{B\bar{B}} f_{00} \mathcal{B}(\rho^0 \rightarrow \pi\pi)}$$



- ▶ charged: $\Upsilon(4S) \rightarrow B^+ B^-$; mixed: $\rightarrow B^0 \bar{B}^0$
- ▶ qqbar: $e^+ e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ continuum

Hadronic tagging: the Full Event Interpretation (FEI)

- Fully reconstruct one B (B_{tag}) in $\sim 10\,000$ exclusive hadronic decay chains
- Hierarchical reconstruction:
tracks and clusters $\rightarrow$ final-state particles $\rightarrow$ intermediate states $\rightarrow B$; a multivariate classifier at each stage
- Output: B_{tag} candidates with a signal probability $P_{\text{sig}}^{\text{FEI}}$
- Efficiency $\sim 1\%$, but no missing energy on the tag side $\Rightarrow$ strong constraint on the B_{tag} four-momentum
- Calibrated on data with $B^0 \rightarrow D^- \pi^+$, $B \rightarrow X \ell \nu$ control samples by the performance group

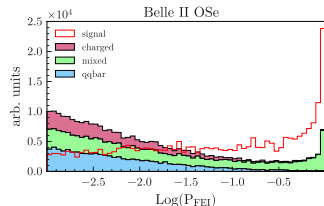


Tag-side selection

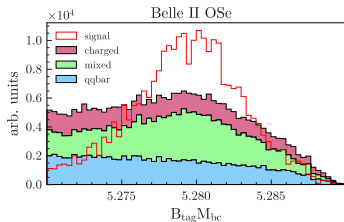
► Hadronic FEI, quality requirements on:

- $M_{bc} = \sqrt{E_{\text{beam}}^2 - |\vec{p}_{B_{\text{tag}}}|^2}$
- $\Delta E = E_{B_{\text{tag}}} - E_{\text{beam}}$
- $P_{\text{sig}}^{\text{FEI}}$: FEI signal probability
- $\cos_{\text{TBTO}}$: angle between the B_{tag} decay axis and the rest of the event axis

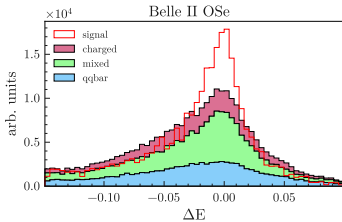
► Several B_{tag} candidates: keep the highest $P_{\text{sig}}^{\text{FEI}}$



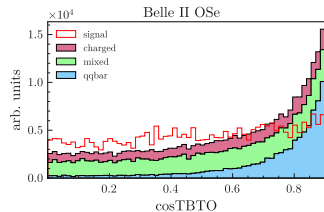
$$P_{\text{sig}}^{\text{FEI}} > 10^{-3}$$



$$M_{bc} > 5.27 \text{ GeV}/c^2$$



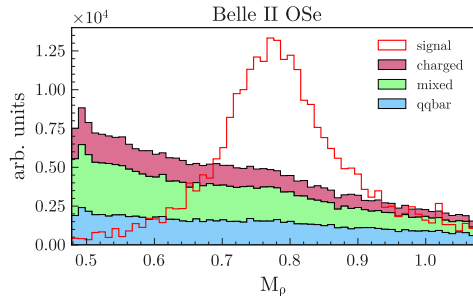
$$-0.15 < \Delta E < 0.10 \text{ GeV}$$



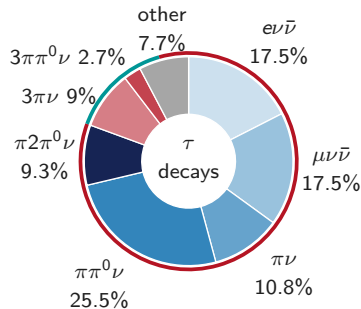
$$\cos_{\text{TBTO}} < 0.9$$

Signal-side reconstruction: $\rho^0 \ell$ and τ prong

- ▶ Reconstruction of π^+ and π^- from the ρ^0 , and of the prompt lepton ℓ
- ▶ Wide ρ^0 mass window to account for large ρ^0 natural width



- ▶ Reconstruction of **1 track t_τ from the τ decay**, for background rejection purpose – not used in the τ recoil-mass measurement
- ▶ **1-prong** $\approx 85\%$ of τ decays
- ▶ **3-prong** $\approx 15\%$ of τ decays



Event-level preselection

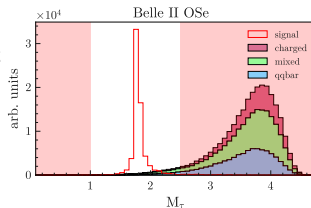
Preselection:

- Sphericity event shape variable, to reject continuum
- M_τ in the range $[1.0, 2.5] \text{ GeV}/c^2$
- $N_{\text{charged}}^{\text{ROE}}$: number of charged tracks in the ROE
- $N_{\text{neutral}}^{\text{ROE}}$: number of neutral clusters in the ROE

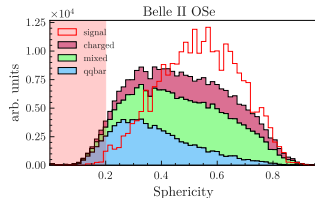
ROE (Rest Of Event): all tracks and clusters not used to reconstruct B_{tag} , the $\rho^0 \ell$ system or the τ prong.

For signal events:

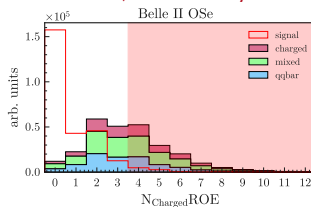
- $\tau \rightarrow 1\text{-prong} \Rightarrow 0$ track in the ROE
- $\tau \rightarrow 3\text{-prong} \Rightarrow 2$ tracks



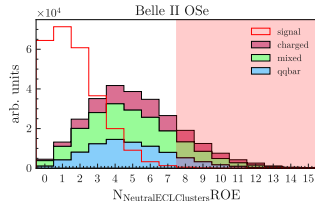
$1.0 < M_\tau < 2.5 \text{ GeV}/c^2$



$\text{Sphericity} > 0.2$



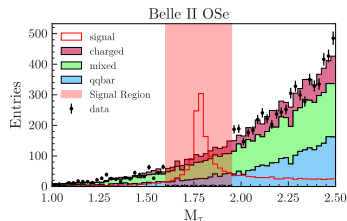
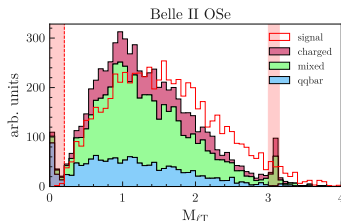
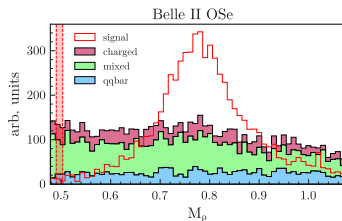
$N_{\text{charged}}^{\text{ROE}} < 4$



$N_{\text{neutral}}^{\text{ROE}} < 8$

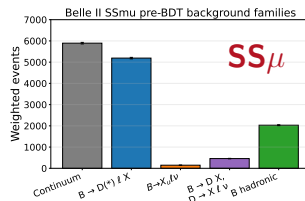
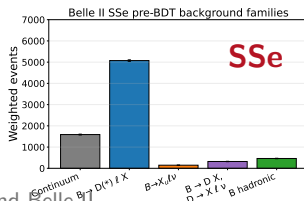
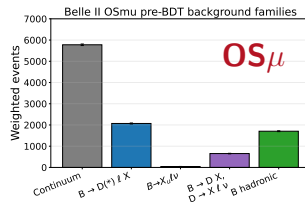
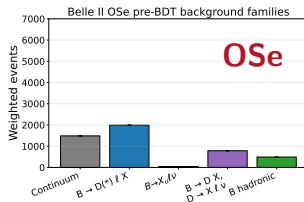
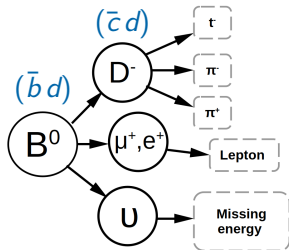
Veto and blinding

- ▶ K_S^0 veto: $M_{\pi\pi} \notin [0.490, 0.503] \text{ GeV}/c^2$
- ▶ $J/\psi \rightarrow \ell\ell$ veto: $M_{\ell\ell} \notin [3.00, 3.15] \text{ GeV}/c^2$ (lepton + τ prong)
- ▶ Photon conversions (electron modes): $M_{\ell\tau} > 0.2 \text{ GeV}/c^2$
- ▶ **Blinded region** $M_\tau \in [1.6, 1.95] \text{ GeV}/c^2$ (92–97% of the signal); sidebands used for data/MC checks and background model



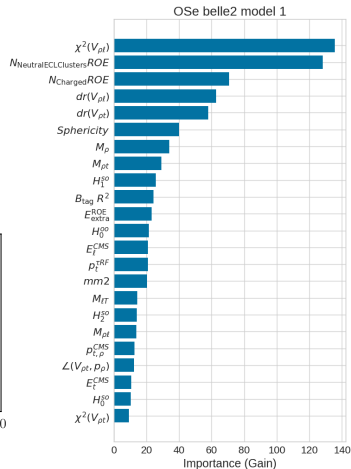
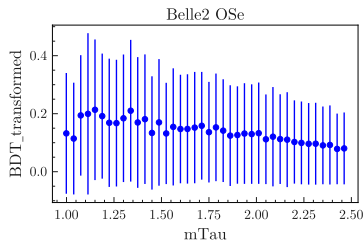
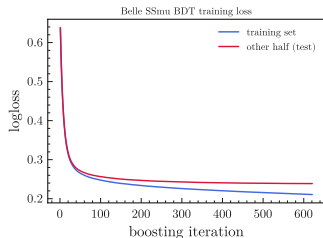
Main background sources

- ▶ Main sources of background at this stage of the selection: **continuum** and $B \rightarrow D^{(*)} \ell \nu$
- ▶ **SS modes** $\Rightarrow$ more semileptonic $B \rightarrow D^{(*)} \ell \nu$ background
- ▶ **Muon modes** $\Rightarrow$ more **continuum** through mis-ID of pions as muons



Boosted Decision Trees (BDT) selection

- **Goal:** reject all remaining background types
- **Input variables:** invariant masses, ℓ/t_τ energies, ROE multiplicities and extra energy, event-shape variables, vertex-fit variables
- **Loss curves** show that overtraining is not affecting the BDT performance
- **Background profile plots** show no sculpting in the signal region



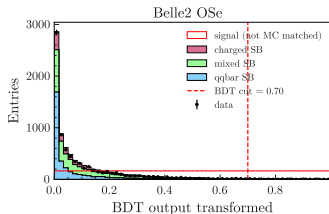
BDT selection optimisation

- Maximise the Punzi figure of merit in the signal region $M_\tau \in [1.6, 1.95] \text{ GeV}/c^2$:

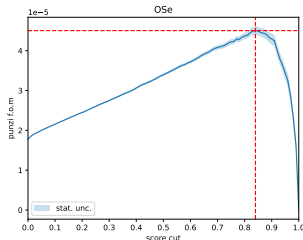
$$\text{Punzi} = \frac{\varepsilon_{\text{sig}}}{a/2 + \sqrt{N_{\text{bkg}}}}, \quad a = 3$$

- Cut capped at 0.70 (Belle II) and 0.90 (Belle) to keep enough background in data sidebands to guarantee a stable fit

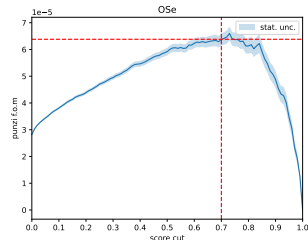
Belle II OSe



Belle OSe



Belle II OSe



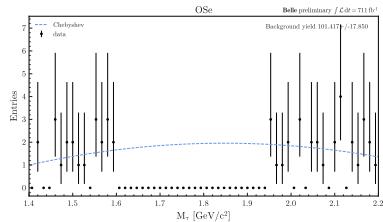
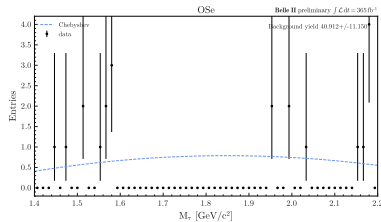
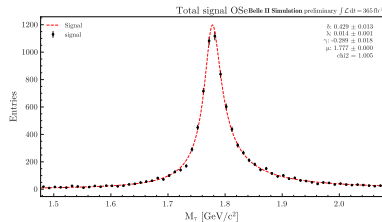
Selection summary

- ▶ If several candidates pass the selection, one is picked randomly – multiplicity left after the BDT: 1.13–1.19 (Belle II), 1.21–1.53 (Belle) candidates per event
- ▶ Signal efficiencies and expected background from MC in the fit region $M_\tau \in [1.4, 2.2] \text{ GeV}/c^2$ (background scaled to the data luminosity):

	ε_{sig}		N_{bkg} left	
	Belle II	Belle	Belle II	Belle
OSe	0.048%	0.053%	63	181
SSe	0.048%	0.041%	94	86
OS μ	0.044%	0.038%	121	94
SS μ	0.046%	0.035%	205	158

Fit procedure

- ▶ The τ recoil mass is fitted in $[1.4, 2.2] \text{ GeV}/c^2$
- ▶ **Signal:** **Johnson** PDF fitted on signal MC
- ▶ **Background:** **Second-order Chebyshev** polynomial fitted in the data sidebands to create toy datasets for the expected ULs



OSe

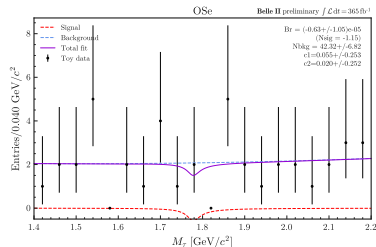
Total fit

- ▶ The branching fraction is extracted with an unbinned extended maximum-likelihood fit performed **simultaneously** on the Belle and Belle II samples:

$$\mathcal{L}(M_\tau) = \frac{e^{-(n_{\text{sig}} + n_{\text{bkg}})}}{N!} \prod_{i=1}^N \left(n_{\text{sig}} \mathcal{P}_{\text{sig}}(M_\tau^i) + n_{\text{bkg}} \mathcal{P}_{\text{bkg}}(M_\tau^i) \right) \cdot \prod_s \text{Gauss}(s, \sigma_s)$$

- ▶ $n_{\text{sig}} = 2 \mathcal{B} \varepsilon_{\text{sig}} f_{00} N_{B\bar{B}}$
 f_{00} : fraction of $B^0 \bar{B}^0$ in $\Upsilon(4S)$ decays
- ▶ Signal PDF parameters **constrained** to their signal MC values
- ▶ Free parameters: branching fraction, background yields and Chebyshev coefficients
- ▶ **Systematics** enter as Gaussian-constrained nuisance parameters

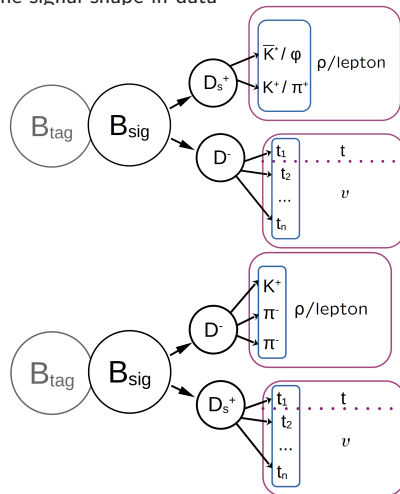
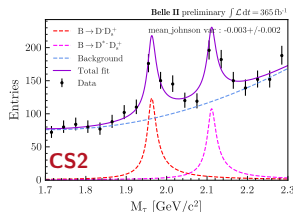
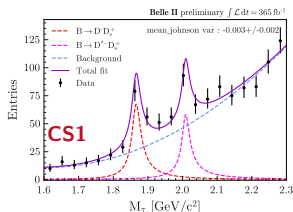
Toy dataset OSe



Control samples

► **Control samples** used to validate the BDT efficiency and the signal shape in data

- **CS1:** $B^0 \rightarrow D^- D_s^+ (\rightarrow \bar{K}^{*0} K^+ / \phi \pi^+)$
- $\bar{K}^{*0} K^+ / \phi \pi^+$ mimics the $\rho \ell$ system
 - D^- / D_s^+ mimics the τ decay
- **CS2:** $B^0 \rightarrow D_s^+ D^- (\rightarrow K^+ \pi^- \pi^-)$
- $K^+ \pi^- \pi^-$ mimics the $\rho \ell$ system
 - D_s^+ / D^- mimics the τ decay



Systematic uncertainties

Signal efficiency

- ▶ FEI calibration, PID, tracking, MC statistics $\Rightarrow \mathcal{O}(2\text{--}10\%)$
- ▶ BDT: with control channels $B^0 \rightarrow D^- D_s^+ \Rightarrow$ **13–33%** Belle, **16–30%** Belle II

Signal shape

- ▶ With control channels $B^0 \rightarrow D^- D_s^+ \Rightarrow (0.1\%, 17\%)$ Belle, $(0.2\%, 24\%)$ Belle II for (μ, λ)

Background shape

- ▶ Order-2 vs. order-3 polynomial $\Rightarrow$ up to $\sim 0.1 \times 10^{-5}$ on $\mathcal{B}$

Others

- ▶ Number of $B\bar{B}$ pairs, fraction of $\Upsilon(4S) \rightarrow B^0 \bar{B}^0$ (f_{00})

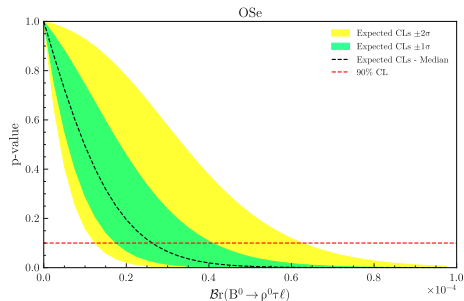
Expected upper limits

- The analysis is still blinded
- Asymptotic CL_s method, 90% CL, on toys from the data sideband fits

	ε_{sig} (Belle II)	ε_{sig} (Belle)	$\mathcal{B}_{\text{exp}}^{\text{UL}}$ (10^{-5})
OS ϵ	0.048%	0.053%	2.63 (2.56)
OS μ	0.044%	0.038%	2.96 (2.85)
SS ϵ	0.048%	0.041%	2.59 (2.48)
SS μ	0.046%	0.035%	3.93 (3.62)

With (without) systematic uncertainties.

Comparable to the BaBar $B^+ \rightarrow \pi^+ \tau \ell$ limits ($2\text{--}7 \times 10^{-5}$), the only existing $b \rightarrow d \tau \ell$ constraints.



Recasting the expected limits

- ▶ Signal generated in **phase space**; a NP interaction would reshape the q^2 and angular distributions
- ▶ Representing NP with an EFT:

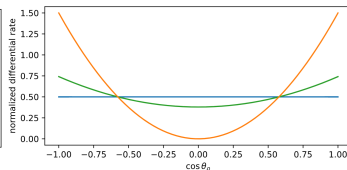
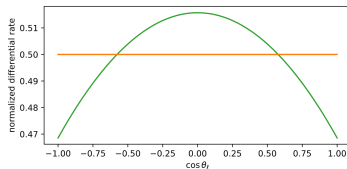
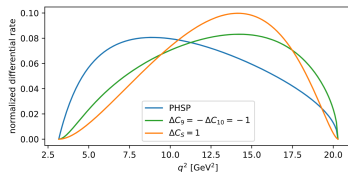
$$\mathcal{H}_{\text{eff}}^{b \rightarrow d \tau \ell} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{td}^* \sum_{i=9,10,S,P} (C_i O_i + C'_i O'_i) + \text{h.c.}$$

Wilson coefficients: $C_{9(10)}$: (axial) vector, $C_{S(P)}$: (pseudo) scalar

- ▶ **Left-handed** model: $C_9 = -C_{10} \neq 0$; **scalar** model: $C_S = -C_P \neq 0$

$\text{UL} \times 10^{-5}$	PHSP	Left-handed	Scalar
OS ℓ	2.63	2.74	2.89
SS ℓ	2.59	2.80	2.93
OS μ	2.96	3.04	3.13
SS μ	3.93	4.08	4.24

- ▶ $q^2 = M_{\ell\tau}^2$
- ▶ θ_ℓ : lepton / B flight directions angle in $\tau\ell$ rest frame
- ▶ θ_ρ : ρ pion / B flight directions angle in the ρ rest frame



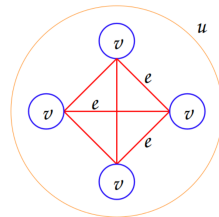
Analysis summary and outlook

- ▶ First search for $B^0 \rightarrow \rho^0 \tau^\pm \ell^\mp$ ($\ell = e, \mu$), $b \rightarrow d \tau \ell$ sector, with Belle + Belle II Run 1 (1.08 ab^{-1})
- ▶ Expected 90% CL upper limits: $2.6\text{--}3.9 \times 10^{-5}$, comparable to existing $b \rightarrow d \tau \ell$ constraints
- ▶ Status: under Belle II working-group review; unblinding and observed limits with the same procedure; paper in preparation
- ▶ Outlook for B tagging: FEI **+40%** hadronic tagging efficiency in the latest reprocessing, **alternative approaches**: inclusive tagging, deep-learning based tagging are under study

An alternative tagging approach: graFEI

GraFEI: graph-based Full Event Interpretation

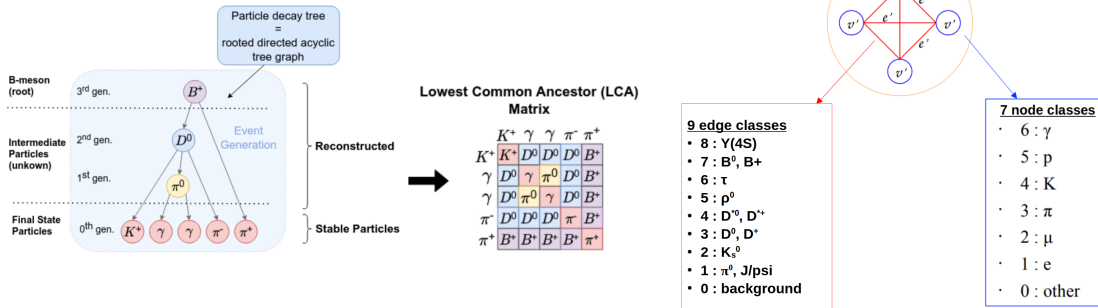
- ▶ **GraFEI**: algorithm developed at Belle II
- ▶ **Graph Neural Network** inferring the *full* decay tree from the $\Upsilon(4S)$ using the reconstructed final-state particles
- ▶ **Input graph**:
 - node features = particles (momentum, charge, PID, dr , dz , cluster shape...)
 - edge features = pairwise relations (angles, distance of closest approach)
- ▶ Three different types of neural network are used: **edge**, **node** and the **global** layers



Input graph

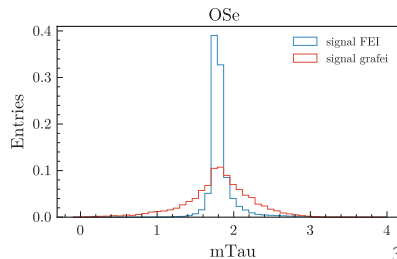
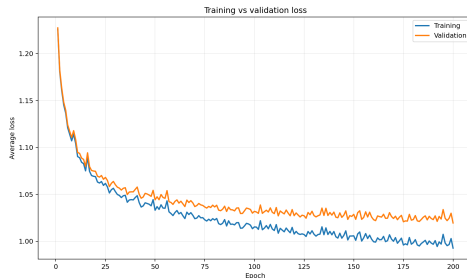
Representing an event with a matrix

- ▶ The full decay tree can be represented as a **Lowest Common Ancestor** matrix (LCA)
- ▶ **Output of the graFEI:** the predicted LCA matrix + the mass hypothesis of each final-state particle $\Rightarrow$ full representation of the event



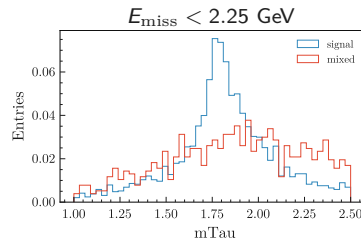
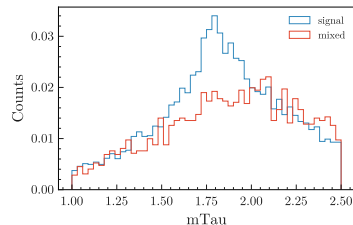
graFEI: training and reconstruction of $B^0 \rightarrow \rho^0 \tau e$

- ▶ **Training** on 2 million events of signal MC
- ▶ **Signal side** ($\ell, t_\tau, \pi^+, \pi^-$) particles are selected from the predicted LCA pattern and mass hypotheses
- ▶ **Tag side inclusive** from the remaining particles
- ▶ Signal efficiency $\approx 6\%$ vs. 0.5% with the FEI, but **much worse M_τ resolution** (no exclusive tag)



Backgrounds with graFEI

- ▶ Background $\times 26\text{--}66$ larger than with the FEI
- ▶ Signal is very "background-like"
- ▶ **Reject** semileptonic tag-side decays using the predicted tag LCA matrix and $E_{\text{miss}} < 2.25$ GeV: better resolution
- ▶ -30% signal (mostly wrongly reconstructed events),
 -15% $B^0 \bar{B}^0$ background

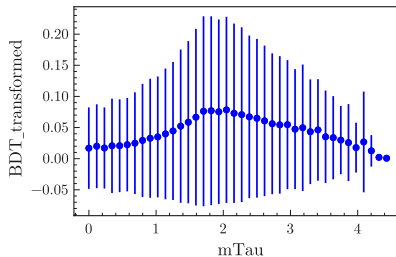


graFEI: BDT attempt and conclusions

- ▶ BDT trained as in the nominal analysis
- ▶ Clear **sculpting of M_{τ}** by the BDT in background: not usable for a fit; not solved within the available time

Conclusions

- ▶ Current graFEI not suited to analyses needing an accurate tag-side B : inclusive tag $\Rightarrow$ poor recoil-mass resolution
- ▶ Training on signal MC likely biases the reconstruction of background events and complicates calibration
- ▶ Directions: train on generic MC, reconstruct intermediate states, transformer-based architectures and recursive graph blocks



Conclusion

$B^0 \rightarrow \rho^0 \tau^\pm \ell^\mp$ analysis

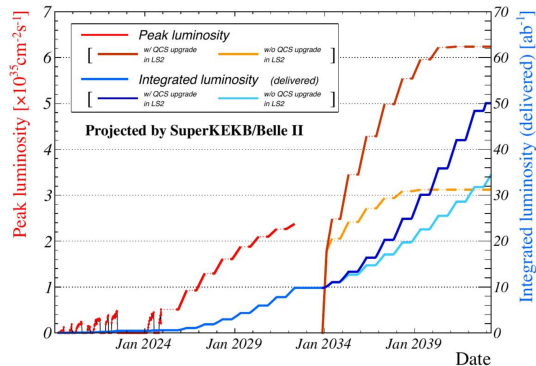
- First search of this mode
- Unblinding and publication to follow

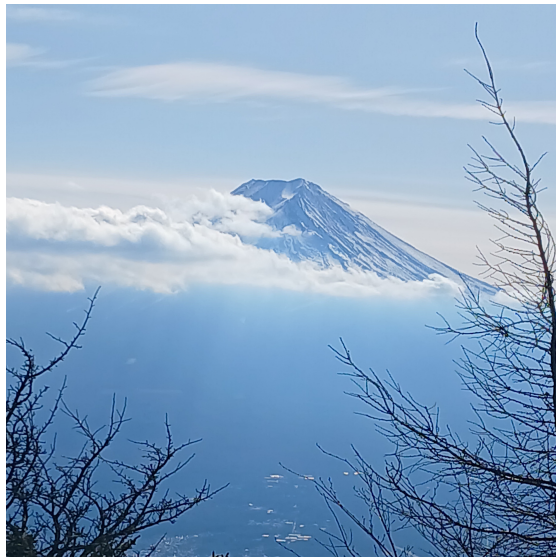
Outlook

- Belle II is collecting data, now close to $\approx 1 \text{ ab}^{-1}$
- Belle II should be upgraded with enhanced performances
- Should reach 50 ab^{-1} by 2043

Tracking performance

- Established a new selection on the track time
- Fake-track measurement automated in VIBE



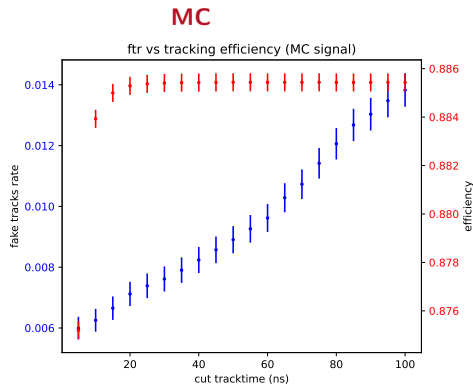
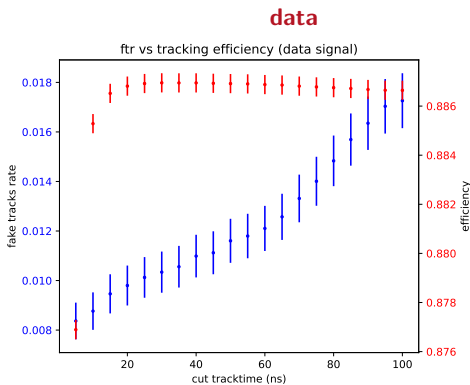


ご清聴ありがとうございました！

Backup

Fake-track rate vs tracking efficiency: $|t_{\text{track}}|$ selection scan

- ▶ The fake-track rates measured in data and in MC **differ in absolute value**, but follow the **same trend** with the $|t_{\text{track}}|$ selection



Thrust and $\cos T_{BTO}$

- ▶ **Thrust** (centre-of-mass frame):

$$T = \max_{\hat{n}} \sum_i |\vec{p}_i \cdot \hat{n}| / \sum_i |\vec{p}_i|$$

thrust axis $\hat{n}_T$: direction along which the momenta are most aligned

$T \rightarrow 1$: two back-to-back jets, $T \rightarrow \frac{1}{2}$: spherical event

- ▶ $\cos T_{BTO}$: $|\cos|$ of the angle between the thrust axes of the B_{tag} daughters and of the rest of the event

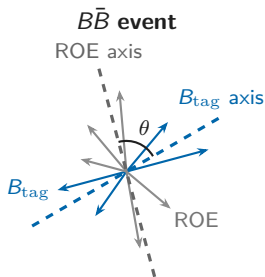
- ▶ **$B\bar{B}$** : both B almost at rest ($p^* \approx 0.34 \text{ GeV}/c$) $\Rightarrow$ isotropic, independent decays $\Rightarrow$ flat

- ▶ **Continuum** $q\bar{q}$: two jets $\Rightarrow$ fake B_{tag} in one jet, rest of event = the other jet $\Rightarrow$ peak at 1

- ▶ $\cos T_{BTO} < 0.9$: removes much of the continuum for $\approx 10\%$ signal loss

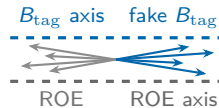
- ▶ **Sphericity** $S = \frac{3}{2}(\lambda_2 + \lambda_3)$, λ_i eigenvalues of $S^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i |\vec{p}_i|^2$: $S \rightarrow 0$ jet-like ($q\bar{q}$), $S \rightarrow 1$ isotropic ($B\bar{B}$)

- ▶ **KSFW moments**: from $H_\ell = \sum_{i,j} |\vec{p}_i| |\vec{p}_j| P_\ell(\cos \theta_{ij})$, $R_\ell = H_\ell / H_0$; separate signal-side and ROE particles, plus the missing mass



axes uncorrelated
 $\cos T_{BTO}$ flat

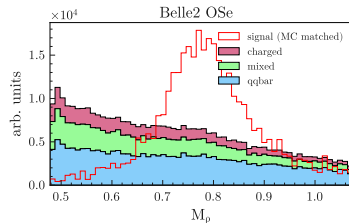
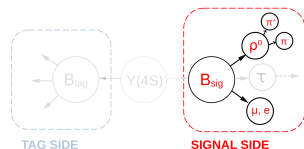
$q\bar{q}$ event



axes aligned
 $\cos T_{BTO} \approx 1$

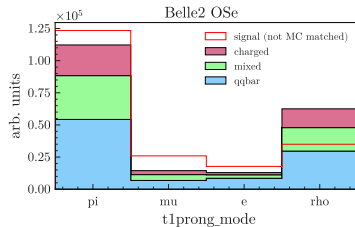
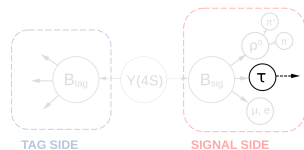
Signal-side reconstruction: $\rho^0 \ell$

- ▶ Reconstruct the $B \rightarrow \rho^0 \ell$ system.
- ▶ Track selection: $d_r < 0.5$ cm, $|d_z| < 5$ cm, CDC acceptance, and $n_{\text{CDC hits}} > 20$ for pions.
- ▶ Particle identification:
 - pionID > 0.6 ,
 - electronID > 0.9 ,
 - muonID > 0.9 .
- ▶ $|M_{\pi\pi} - M_{\rho^0}| < 298 \text{ MeV}/c^2$, with $M_{\rho^0} = 775 \text{ MeV}/c^2$.
- ▶ Bremsstrahlung recovery ($E_\gamma > 50 \text{ MeV}$) in electron modes.



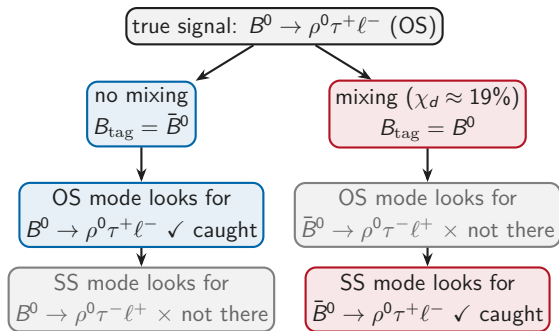
Signal-side reconstruction: τ prong

- ▶ Reconstruct one track from the τ decay for background rejection only
- ▶ Track selection: $d_r < 0.5$ cm, $|d_z| < 5$ cm.
- ▶ Four accepted τ -prong hypotheses:
 - e: electronID > 0.9 ,
 - π : pionID > 0.6 ,
 - μ : muonID > 0.9 ,
 - $\rho^\pm \rightarrow \pi^\pm \pi^0$: pionID > 0.6 ,
 $|M_{\pi^\pm \pi^0} - M_{\rho^\pm}| < 298 \text{ MeV}/c^2$
 $|\Delta M_{\pi^0}| < 10 \text{ MeV}/c^2$.
- ▶ The τ modes are not separated in the analysis.



$B^0-\bar{B}^0$ mixing and the OS/SS categories

- ▶ The B_{tag} flavour fixes which decay is looked for; for $B_{\text{tag}} = B^0$: OS mode $\rightarrow \bar{B}^0 \rightarrow \rho^0 \tau^- \ell^+$, SS mode $\rightarrow \bar{B}^0 \rightarrow \rho^0 \tau^+ \ell^-$
- ▶ Mixed true-OS event ($B^0 \rightarrow \rho^0 \tau^+ \ell^-$ with $B_{\text{tag}} = B^0$): the OS mode looks for a final state that *is not there* – the **SS mode catches it**
- ▶ **No double counting**: for one B_{tag} flavour the two chains need opposite lepton charges $\Rightarrow$ at most one mode catches an event
- ▶ **Double counting** can happen in case of misreconstruction between the prompt lepton and the τ prong, independant from mixing
- ▶ OS $\leftrightarrow$ SS **cross-feed**, in the signal MC: 3.6–6.0% before the BDT, 1.8–3.0% after

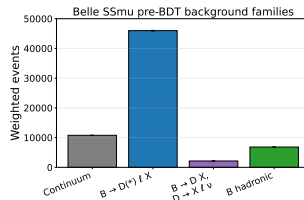
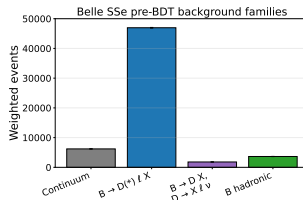
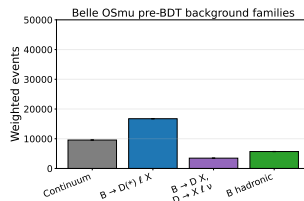
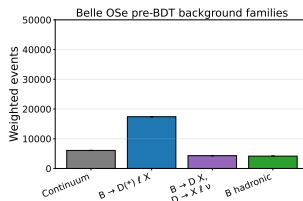


Background yields family before the BDT - Belle

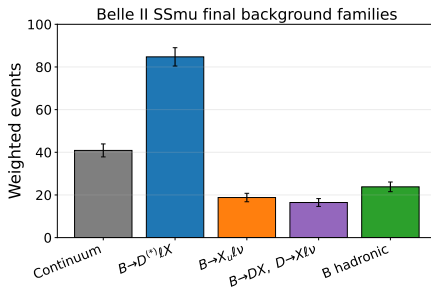
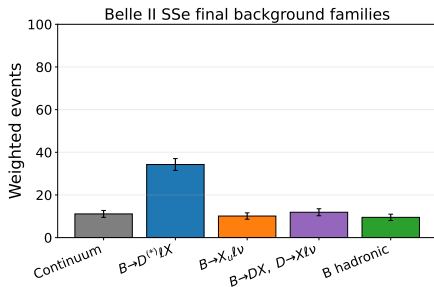
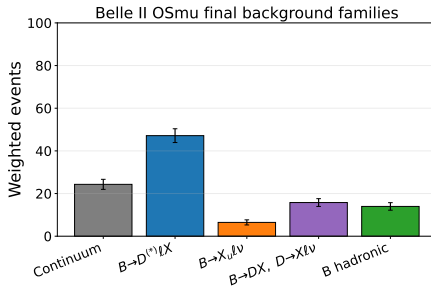
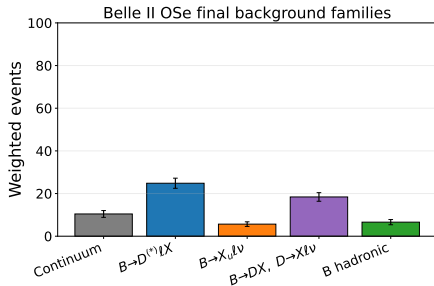
Background family	OSe		OSmu		SSe		SSmu	
	Belle II	Belle	Belle II	Belle	Belle II	Belle	Belle II	Belle
Continuum (evts/ fb^{-1})	4.1	8.5	15.8	13.5	4.4	8.7	16.1	15.2
All semileptonic (evts/ fb^{-1})	7.7	30.5	7.6	28.4	15.9	68.6	15.9	67.7

► ~ 4 times more
semi-leptonic background
in Belle

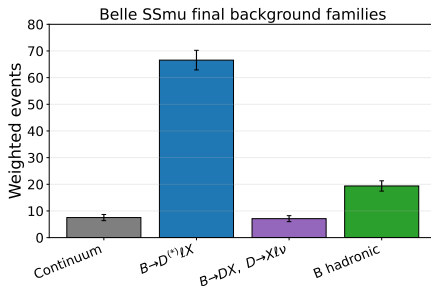
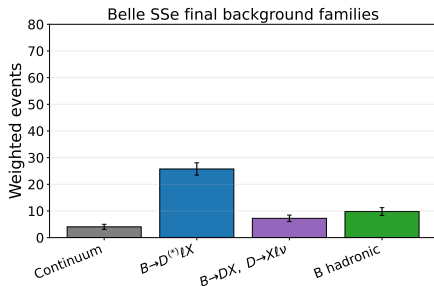
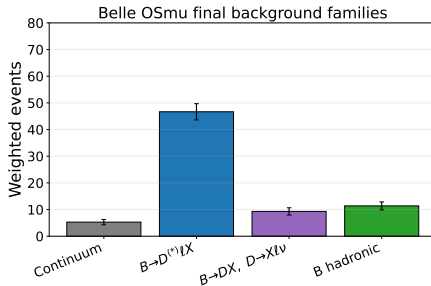
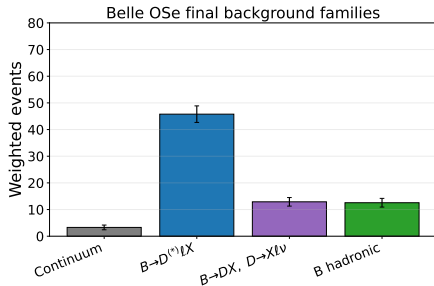
► ~ 2 times more continuum
background in Belle
electron modes



Background families after the BDT: Belle II



Background families after the BDT: Belle



BDT training procedure

Training samples: MC split 50/50 into halves A and B; model 1 trained on A, model 2 on B $\Rightarrow$ (**cross-evaluation**).

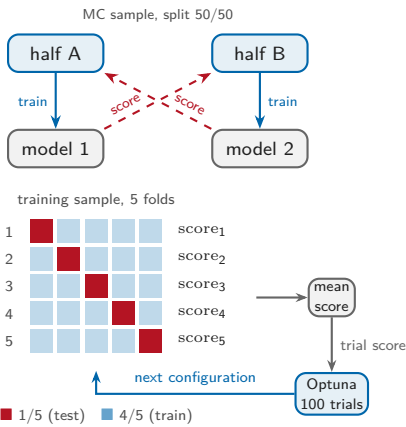
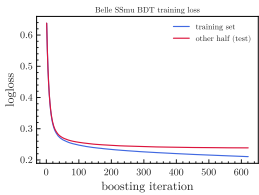
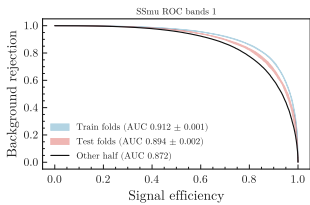
Hyperparameter optimisation: Optuna, 100 trials, independently per model

- ▶ One XGBoost configuration per trial: `n_estimators` 250–1000, `max_depth` 1–5
- ▶ 5 stratified folds: train on 4/5, evaluate on the held-out 1/5;

$$\text{score} = \frac{\text{AUC}_{\text{test}}}{\mathcal{L}_{\text{test}}} \left(\frac{\text{AUC}_{\text{test}}}{\text{AUC}_{\text{train}}} \right)^2 \frac{\mathcal{L}_{\text{train}}}{\mathcal{L}_{\text{test}}} \quad (\mathcal{L} = \text{log-loss})$$

rewards held-out performance, **penalises the train–test gap** (overtraining)

- ▶ Trial



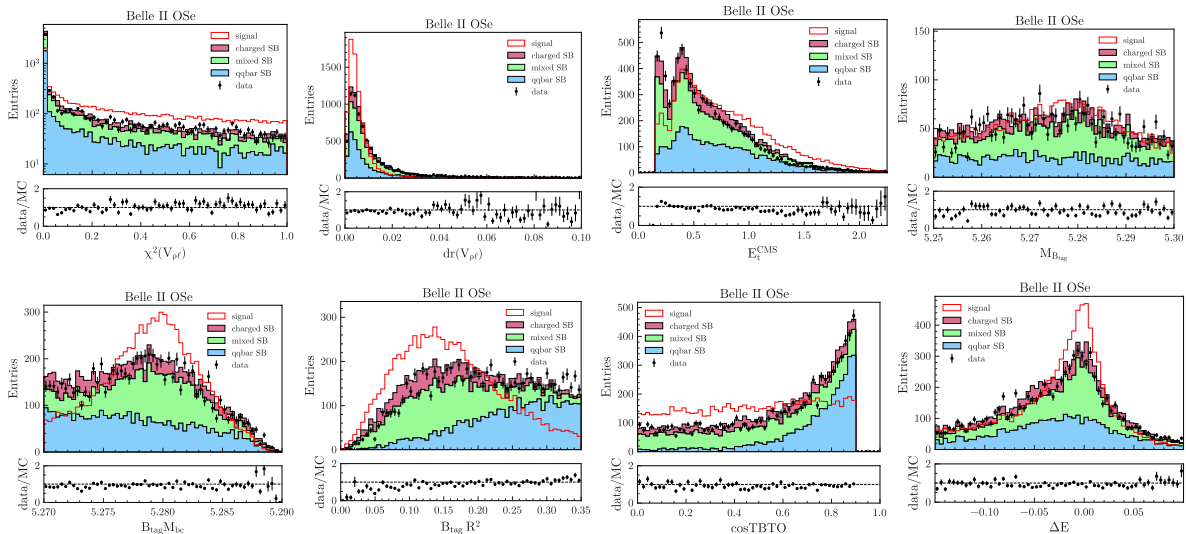
Final training – best configuration trained on the full training sample.

Belle vs Belle II: selection performance

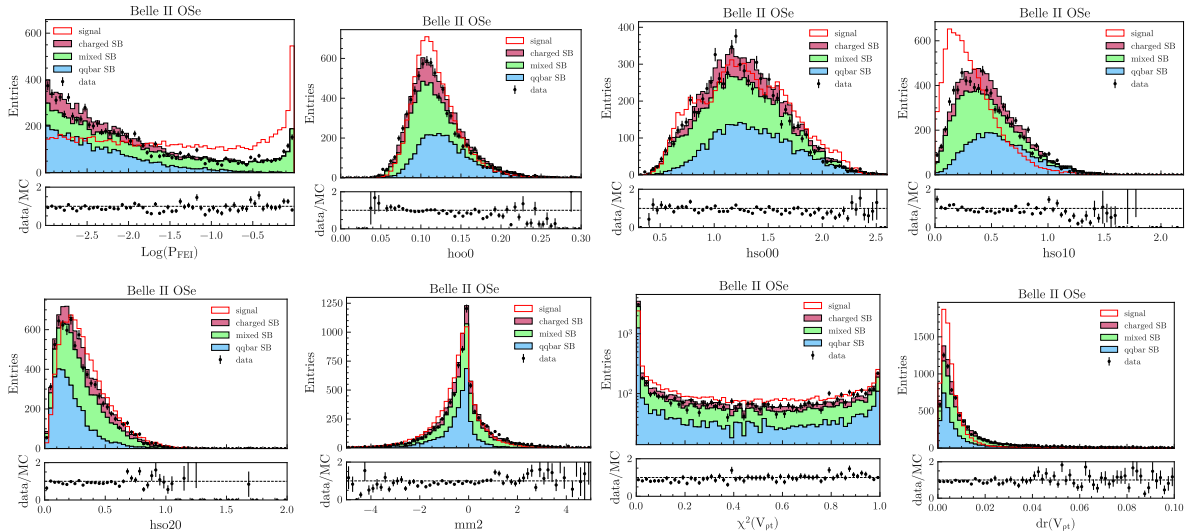
after the BDT	OS _e	OS _{μ}	SS _e	SS _{μ}
ϵ_{sig} Belle II [%]	0.048	0.044	0.048	0.046
ϵ_{sig} Belle [%]	0.053	0.038	0.041	0.035
$N_{\text{bkg}}/\text{ab}^{-1}$ Belle II	173	332	258	562
$N_{\text{bkg}}/\text{ab}^{-1}$ Belle	255	132	121	222
BDT cut (Belle II / Belle)	0.70/0.84	0.70/0.88	0.70/0.88	0.70/0.88

- ▶ Per ab^{-1} and *before* the BDT, Belle II is **2–6 $\times$ cleaner** in the electron modes ($1.8\times$ in OS _{μ}): better tag purity and π/K separation
- ▶ But Belle keeps $\approx 1.6\times$ more signal at that stage
- ▶ *After* the BDT the comparison is driven by the **cut choice**, capped at 0.70 for Belle II to keep enough sideband events for a stable fit, against 0.84–0.88 for Belle

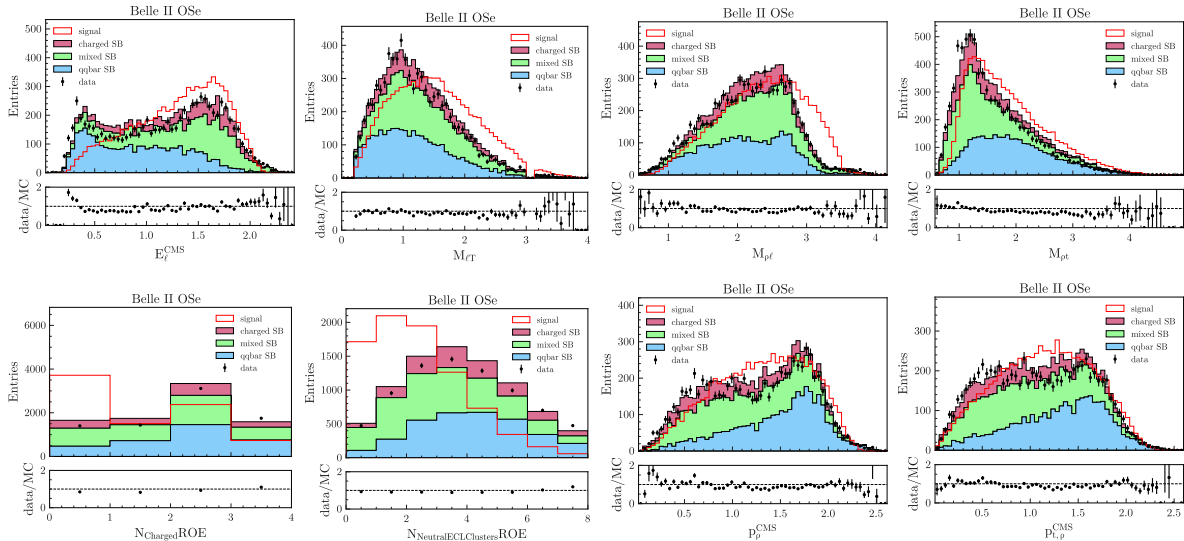
BDT input variables: data/MC in sidebands (1/4)



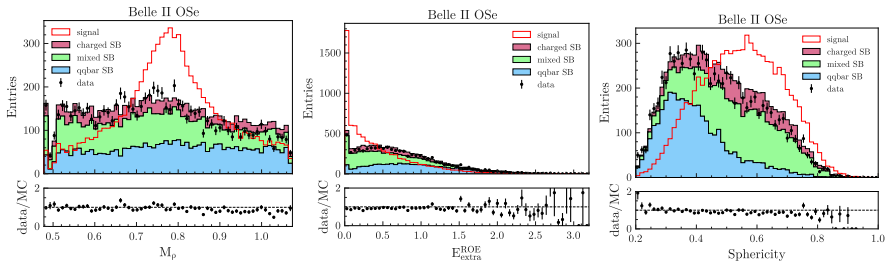
BDT input variables: data/MC in sidebands (2/4)



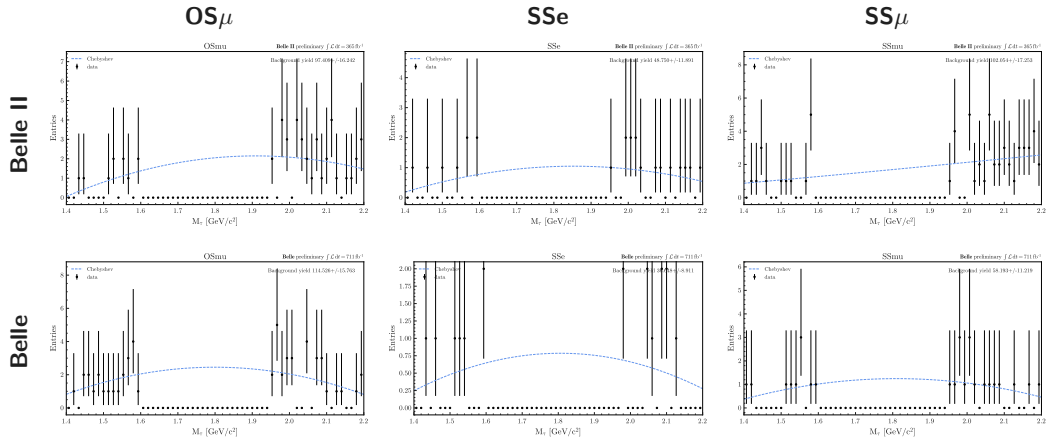
BDT input variables: data/MC in sidebands (3/4)



BDT input variables: data/MC in sidebands (4/4)

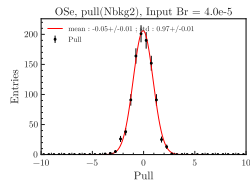
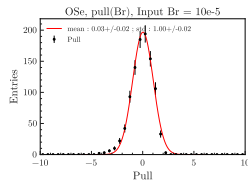
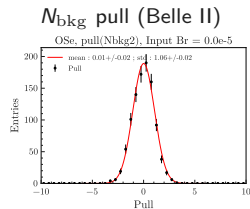
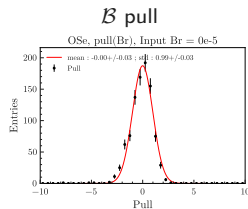
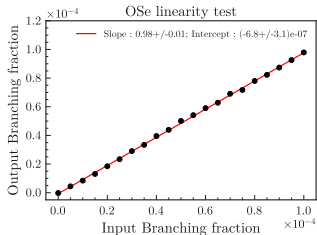


Data sideband fits: other modes



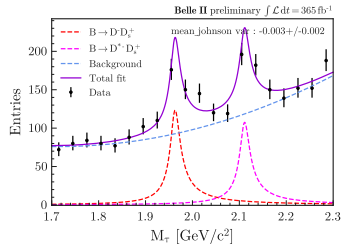
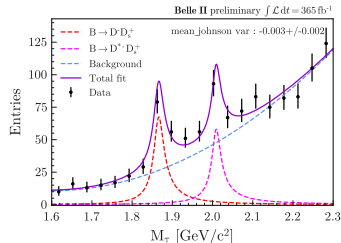
Fit validation

- ▶ 1000 toy datasets per injected branching fraction ($0 \rightarrow 1 \times 10^{-4}$), each fitted with the nominal procedure
- ▶ Linear response slope compatible with 1, intercept nearly compatible with 0 $\Rightarrow$ **no bias**
- ▶ pulls mean compatible with 0 and width compatible with 1 $\Rightarrow$ **uncertainties are well estimated**



Signal shape systematics

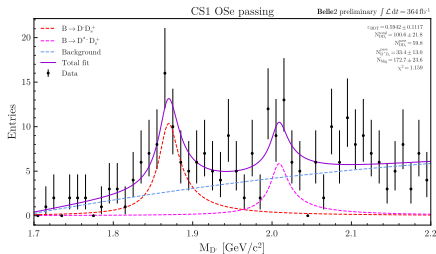
- ▶ Simultaneous fit of CS1 and CS2 samples, common fudge factor on the mean ($\delta\mu$) / width (f) of the Johnson PDF
- ▶ $B^0 \rightarrow D^{*-} D_s^+$ (CS1) and $B^0 \rightarrow D^- D_s^{*+}$ (CS2) are also reconstructed and fitted
- ▶ Fit signal MC $B^0 \rightarrow D^- D_s^+$ with a Johnson PDF
- ▶ **Mean parameter:** define $\mu' = \mu + \delta\mu$
 μ is fixed to the fitted value in signal MC, $\delta\mu$ is free to vary
- ▶ **Resolution parameter:** define $\lambda' = \lambda \times f$
 λ is fixed to the fitted value in signal MC, f is free to vary
- ▶ Fit data with this modified PDF
- ▶ Assign $\max(|\delta\mu|, \sigma_{\delta\mu})$ as syst on μ
- ▶ Assign $\max(|1 - f|, \sigma_f)$ as syst on λ
- ▶ Belle II: $\delta\mu = 0.2\%$, $\delta\lambda = 24\%$, Belle: $\delta\mu = 0.1\%$, $\delta\lambda = 17\%$.



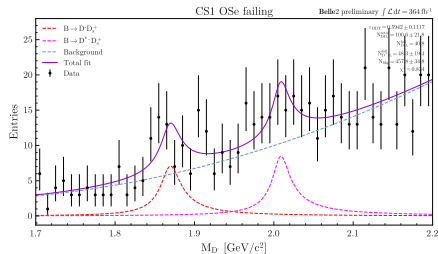
Systematics on the BDT efficiency - I

- Choose the control-channel BDT cut reproducing the nominal signal relative efficiency
- Simultaneous fit of the pass/fail categories, sharing the efficiency ε_{BDT} and $N_{\text{signal}}^{\text{tot}}$ as common parameters
- $N_{\text{pass}} = \varepsilon_{\text{BDT}} \cdot N_{\text{signal}}^{\text{tot}}$, $N_{\text{fail}} = (1 - \varepsilon_{\text{BDT}}) \cdot N_{\text{signal}}^{\text{tot}}$

CS1 data passing BDT cut (Belle II, OSe)



CS1 data failing BDT cut (Belle II, OSe)



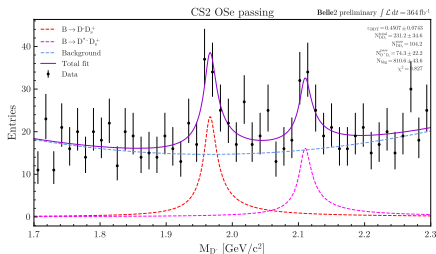
Systematics on the BDT efficiency - II

→ Same simultaneous pass/fail fit for CS2, $R_i = \varepsilon_{\text{BDT},i}^{\text{data}} / \varepsilon_{\text{BDT},i}^{\text{MC}}$, $i \in \{\text{CS1}, \text{CS2}\}$

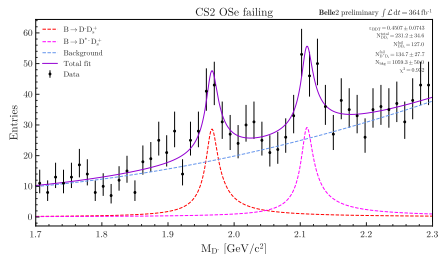
→ Combine CS1, CS2 with an inverse-variance weighted average of R_{CS1} , R_{CS2} :

$$R_{\text{data/MC}}^{\varepsilon_{\text{BDT}}} = \frac{\sum_i w_i R_i}{\sum_i w_i}, \quad \sigma_{R_{\text{data/MC}}^{\varepsilon_{\text{BDT}}}} = \frac{1}{\sqrt{\sum_i w_i}}, \quad w_i = \frac{1}{\sigma_{R_i}^2}.$$

CS2 data passing BDT cut (Belle II, OSe)



CS2 data failing BDT cut (Belle II, OSe)



Other systematics

Background shape:

- ▶ Fit large bootstrap ensembles of the toy generated from the data sidebands with 2nd and 3rd order Chebychev polynomials
- ▶ Assign as systematic uncertainty the difference between the average fitted branching fractions
 $\rightarrow 0.050, 0.041, 0.009, 0.113 (\times 10^{-5})$ for OSe, OS μ , SSe, SS μ

Tag side (FEI):

- ▶ Vary the FEI calibration weights within their quoted uncertainties, quadratic sum of stat. and syst. components, and reevaluate efficiency
- ▶ Assign the larger of the up/down variations as systematic
 $\rightarrow \sim 3.2 - 3.4\%$ for Belle II, $\sim 2.2 - 2.3\%$ for Belle

Particle identification:

- ▶ Vary the PID weights (lepton and hadron ID separately) and reevaluate efficiency
- ▶ Assign the largest relative variation as systematic
 $\rightarrow \sim 2.4 - 2.9\%$ (LID) and $\sim 5.3 - 5.5\%$ (HadID) for Belle II,
 $\rightarrow \sim 2.0 - 2.2\%$ (LID) and $\sim 8.3 - 9.7\%$ (HadID) for Belle

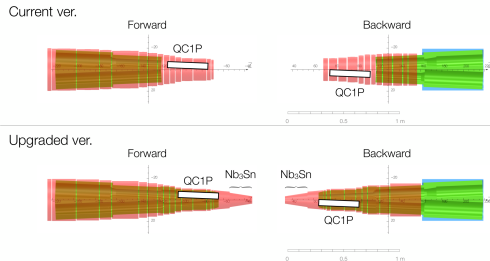
Summary of systematic uncertainties

Belle	OSe	OS μ	SSe	SS μ
Lepton ID	2.04%	2.19%	2.21%	2.16%
Hadron ID	8.65%	8.28%	9.71%	8.93%
FEI	2.28%	2.31%	2.33%	2.34%
BDT	13.37%	19.40%	17.46%	32.53%
Total efficiency	16.26%	21.36%	20.27%	33.90%
Signal width λ	17%	17%	17%	17%
Signal mean μ	0.5%	0.5%	0.5%	0.5%
Belle II	OSe	OS μ	SSe	SS μ
Lepton ID	2.94%	2.43%	2.88%	2.41%
Hadron ID	5.26%	5.31%	5.48%	5.29%
FEI	3.39%	3.30%	3.17%	3.30%
BDT	16.18%	16.08%	29.98%	21.06%
Total efficiency	17.63%	17.46%	30.80%	22.12%
Signal width λ	24%	24%	24%	24%
Signal mean μ	0.3%	0.3%	0.3%	0.3%
Background shape, both exp. [10^{-5} on $\mathcal{B}$]	0.050	0.041	0.009	0.113

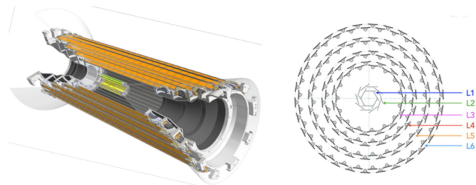
Belle II and SuperKEKB upgrades

- ▶ Goal $6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ and $50 \text{ ab}^{-1} \Rightarrow$ interaction region redesigned in **LS2**
- ▶ New superconducting final-focus magnets (QCS): the layout **reorganises the space around the IP** and sets the volume left for the vertex detector
- ▶ **Sudden beam losses** (SBL) in the LER, with quenches of the final-focus magnets, limited recent operation
- ▶ **CDC**: ageing and gain loss under study; replacement or upgrade considered for LS2
- ▶ **PXD and SVD** cannot stand the expected background occupancies $\Rightarrow$ replaced by VTX, an all-pixel CMOS detector

[arXiv:2406.19421], [arXiv:2503.24155]



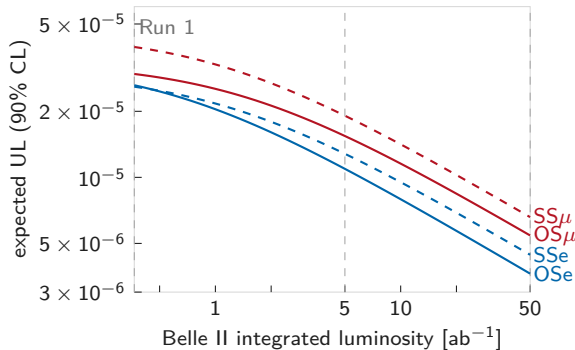
current (top) vs. new (bottom) QCS configuration



VTX baseline layout

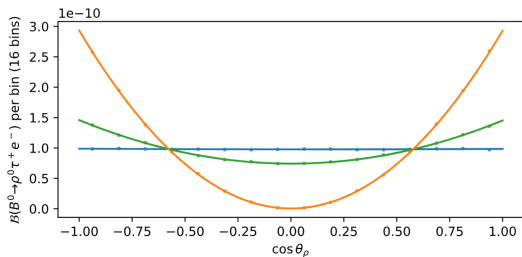
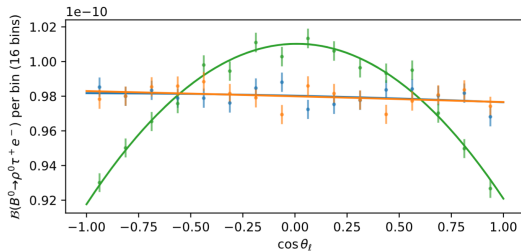
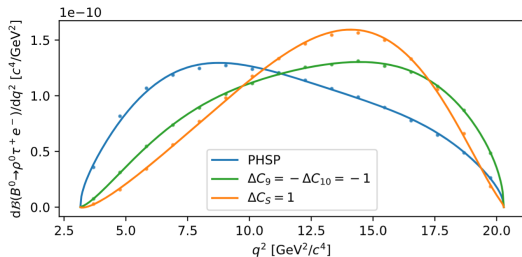
Expected upper limits with more Belle II data

- ▶ Same analysis; Belle II luminosity scaled, Belle 711 fb⁻¹ kept in the combination
- ▶ Background-dominated (60–200 events per mode) ⇒ Belle II sensitivity $\propto \sqrt{L}$
- ▶ Systematics neglected (stat.-dominated today)
- ▶ $\approx 7\times$ **better at 50 ab⁻¹**, still above the $\sim 10^{-7}$ predictions



UL [10 ⁻⁵]	OSe	SSe	OS _μ	SS _μ
today	2.6	2.6	3.0	3.9
5 ab ⁻¹	1.1	1.3	1.5	1.9
50 ab ⁻¹	0.36	0.44	0.54	0.66

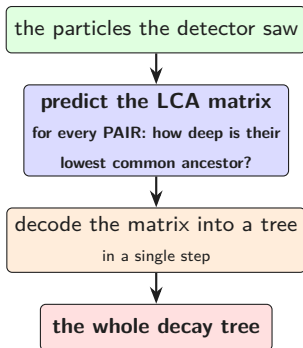
NP recast: fits of the phase-space distributions



- Points: theory distributions for $B^0 \rightarrow \rho^0 \tau^+ e^-$, 500 000 toy events per model
- Lines: fitted shapes using polynomials

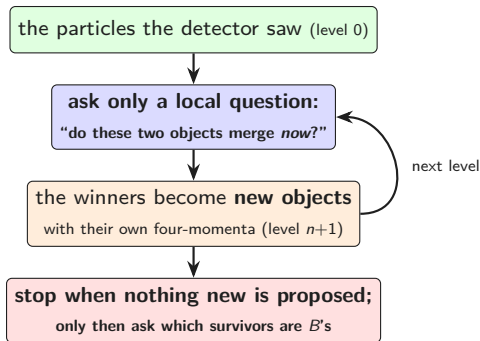
TREX, a transformer-recursive B tagging algorithm

graFEI: one shot



never usable as a tagging algorithm:
trained on *specific signal events* to
find
the signal LCA sub-matrix and flag it.
No clean tag side was ever con-
structed.

T-Rex: one generation at a time



the tag side IS the output:
every survivor is a fully built candidate
with its own four-momentum, on
events
with no signal channel assumed.