technische universität dortmund

FSP LHCb
Erforschung von Universum und Materie

e5 experimentelle physik 5
teilchenphysik & medizinphysik

Inclusive Flavour Tagging at LHCb

Jonah Blank

on behalf of the LHCb Collaboration

Lepton Photon Conference
Madison, USA,
August 28, 2025

What's on the menu?

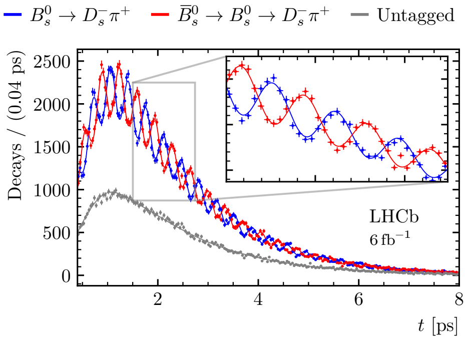
- Flavour Tagging at LHCb
- The Inclusive Flavour Tagger - concept and implementation
- IFT - training and results
- IFT - calibration and performance
- Validation studies
- Summary and outlook

What's on the menu?

- **Flavour Tagging at LHCb**
- The Inclusive Flavour Tagger - concept and implementation
- IFT - training and results
- IFT - calibration and performance
- Validation studies
- Summary and outlook

Flavour Tagging at LHCb - Why?

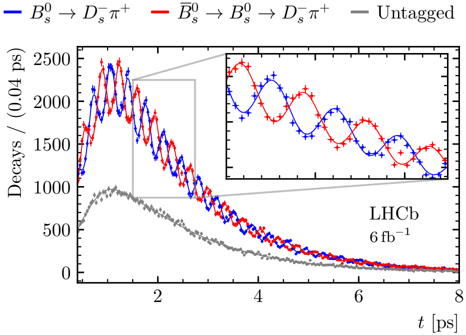
- Neutral B mesons can oscillate between particle and anti-particle
- Measurements of mixing $\Delta m_{(s)}$ & time-dependent CP asymmetry parameters (e.g. ϕ_s)



Nat. Phys. 18, 1–5 (2022)

Flavour Tagging at LHCb - Why?

- Neutral B mesons can oscillate between particle and anti-particle
- Measurements of mixing $\Delta m_{(s)}$ & time-dependent CP asymmetry parameters (e.g. ϕ_s)
- Knowledge about initial production flavour required to resolve oscillation

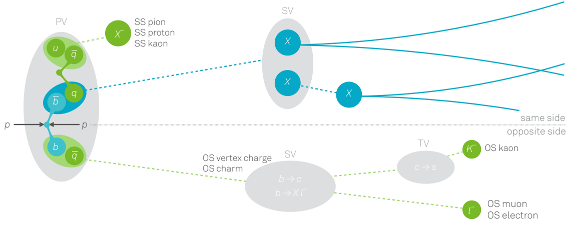


Nat. Phys. 18, 1–5 (2022)

Flavour Tagging at LHCb - Current status

8 available particle taggers based on physics cases:

- Opposite Side (OS): using hadronization and decay of correlated partner b
- Same Side (SS): using fragmentation partner of signal B



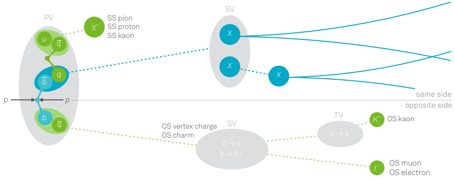
What's on the menu?

- Flavour Tagging at LHCb
- **The Inclusive Flavour Tagger - concept and implementation**
- IFT - training and results
- IFT - calibration and performance
- Validation studies
- Summary and outlook

The Inclusive Flavour Tagger (IFT)

Limits of classical taggers:

- Each designed for specific final state particles
  → events that don't fulfill requirements cannot be tagged
- Only a subset of tracks/vertices is used
  → correct selection not guaranteed
  → correlations in the event are lost

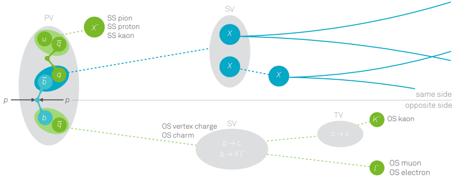


)

The Inclusive Flavour Tagger (IFT)

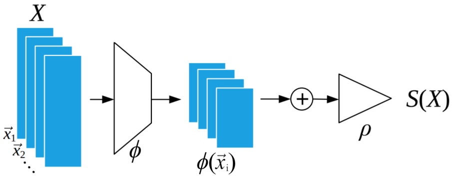
Limits of classical taggers:

- Each designed for specific final state particles
  → events that don't fulfill requirements cannot be tagged
- Only a subset of tracks/vertices is used
  → correct selection not guaranteed
  → correlations in the event are lost
- Idea: One tagger using all tracks of the underlying event (DOI:10.1088/1742-6596/762/1/012045:)



The Inclusive Flavour Tagger (IFT)

- Different implementations under investigation
  → Today: The DeepSet IFT
- Consists of two Neural Networks (NNs) ϕ and ρ
  - ϕ processes input matrix X
  - Permutation invariant aggregation of $\phi(\vec{x}_i)$ into single vector
  - Input for ρ producing tagging decision
- Advantages: arbitrary number of tracks possible, no ordering necessary



Adapted from: arXiV: 1703.06114

What's on the menu?

- Flavour Tagging at LHCb
- The Inclusive Flavour Tagger - concept and implementation
- **IFT - training and results**
- IFT - calibration and performance
- Validation studies
- Summary and outlook

The datasets

- 2 classifiers for B^0 and B_s due to different SS fragmentation
- Training on MC simulation with conditions 2016-2018
  → supervised training possible due to true flavour information
- Calibration and validation on real data

B meson	training MC	calibration data	validation studies
B^0	$B^0 \to J/\psi K^*(892)^0$	$B^0 \to J/\psi K^+\pi^-$	$B^0 \to J/\psi K_S^0$
B_s^0	$B_s^0 \to J/\psi\phi(1020)$	$B_s^0 \to D_s^-\pi^+$	$B_s^0 \to J/\psi K^+K^-$

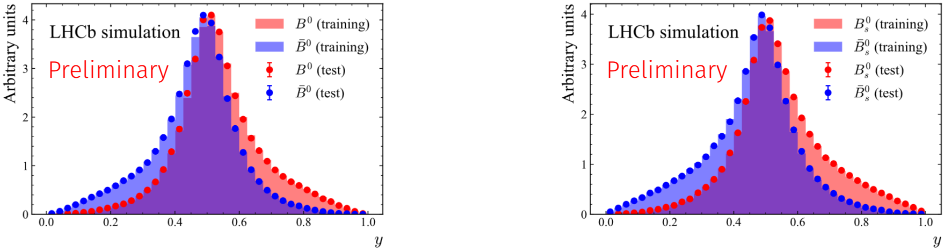
The training pipeline

Preprocessing:

- Multiclass BDT to preclassify each track into 4 categories: SS Fragmentation, OS Fragmentation, OS Decay, unrelated tracks (BKG)

IFT training:

- DeepSet NN takes same features as the BDT + BDT outputs & classical tagger decisions



What's on the menu?

- Flavour Tagging at LHCb
- The Inclusive Flavour Tagger - concept and implementation
- IFT - training and results
- **IFT - calibration and performance**
- Validation studies
- Summary and outlook

The mistag probability

- Output of NN: $y \in [0,1]$ - probability estimate that particle is a B
- Used to define output decision d and mistag estimation η as

$$d = \begin{cases} +1 & \text{if } y > 0.5, \\ -1 & \text{if } y < 0.5 \end{cases} \qquad \eta = \begin{cases} 1-y & \text{if } y > 0.5, \\ y & \text{if } y < 0.5 \end{cases}$$

- η not necessarily describing measured data well
  → calibration on real decays

Calibration of mistag probability

- Calibrated mistag fraction ω parametrized as linear function in η:

$$\omega^{B/\bar{B}}(\eta) = p_0 \pm p_1 \cdot (\eta - \bar{\eta}) \pm \frac{1}{2} \cdot \Delta\omega,$$

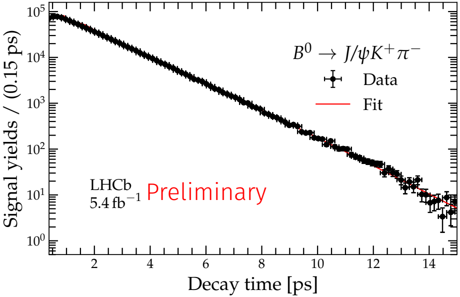
with the average mistag estimate on the training sample $\bar{\eta}$ and

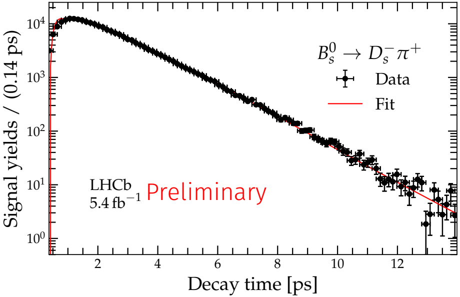
$$\Delta\omega(\eta) = \Delta p_0 \pm \Delta p_1 \cdot (\eta - \bar{\eta})$$

accounting for potentially asymmetric calibration parameters for different flavours due to threshold at $y = 0.5$

The decay time fit

- Using flavour-specific decay channels $B^0 \to J/\psi K^+\pi^-$ ($B_s^0 \to D_s^- \pi^+$) for $B^0(B_s^0)$-IFT
  → flavour at decay is known
- Unbinned likelihood fit to decay time τ, including time resolution σ_τ, to extract measured mistag from dilution $D = 1 - 2\omega$





Tagging combination and performance

- Flavour Tagging performance evaluated by tagging power:

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}\langle D^2\rangle = \varepsilon_{\text{tag}}\langle(1-2\omega)^2\rangle,$$

- Higher tagging power corresponds to higher statistics
  → increasing it directly translates to lower statistic uncertainty $\sigma^{\text{stat}} \propto \frac{1}{\sqrt{\varepsilon_{\text{eff}} N}}$

Tagging combination and performance

- Flavour Tagging performance evaluated by tagging power:

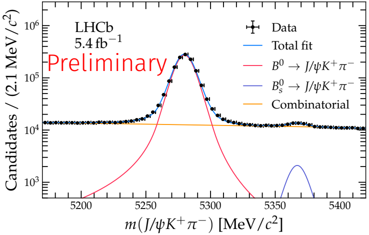
$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}} \langle D^2 \rangle = \varepsilon_{\text{tag}} \langle (1 - 2\omega)^2 \rangle,$$

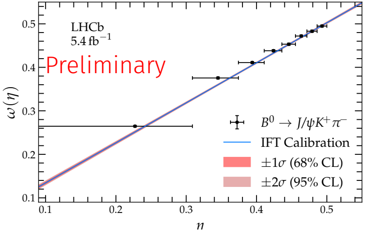
- Higher tagging power corresponds to higher statistics
  → increasing it directly translates to lower statistic uncertainty $\sigma^{\text{stat}} \propto \frac{1}{\sqrt{\varepsilon_{\text{eff}} N}}$
- Comparing IFT performance to combination of SS and OS taggers

Calibration and performance: B^0-IFT

- Loose selection criteria for $B^0 \to J/\psi K^+\pi^-$
- Remaining combinatorial background is statistically subtracted using the $sPlot$ technique

Year	Tagger	$\varepsilon_{\rm tag}(\%)$	$\langle D^2\rangle$	$\varepsilon_{\rm tag}\langle D^2\rangle(\%)$
2016	SS and OS	87.030 ± 0.058	0.0466 ± 0.0010	4.05 ± 0.09
	IFT	95.789 ± 0.035	0.0582 ± 0.0011	5.58 ± 0.10
2017	SS and OS	87.208 ± 0.058	0.0485 ± 0.0010	4.23 ± 0.09
	IFT	95.024 ± 0.038	0.0617 ± 0.0011	5.87 ± 0.10
2018	SS and OS	87.631 ± 0.052	0.0493 ± 0.0010	4.32 ± 0.09
	IFT	95.172 ± 0.034	0.0615 ± 0.0010	5.85 ± 0.10



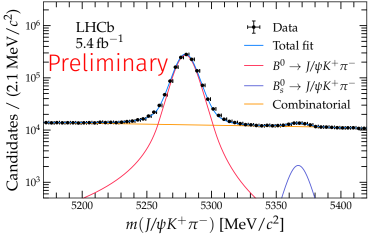


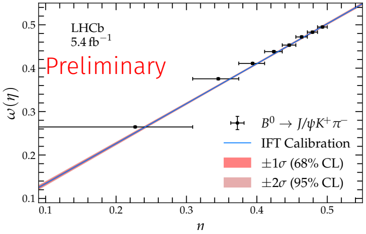
Calibration and performance: B^0-IFT

- Loose selection criteria for $B^0 \to J/\psi K^+\pi^-$
- Remaining combinatorial background is statistically subtracted using the $sPlot$ technique

Year	Tagger	$\varepsilon_{\text{tag}}(\%)$	$\langle D^2\rangle$	$\varepsilon_{\text{tag}}\langle D^2\rangle(\%)$
2016	SS and OS	87.030 ± 0.058	0.0466 ± 0.0010	4.05 ± 0.09
	IFT	95.789 ± 0.035	0.0582 ± 0.0011	5.58 ± 0.10
2017	SS and OS	87.208 ± 0.058	0.0485 ± 0.0010	4.23 ± 0.09
	IFT	95.024 ± 0.038	0.0617 ± 0.0011	5.87 ± 0.10
2018	SS and OS	87.631 ± 0.052	0.0493 ± 0.0010	4.32 ± 0.09
	IFT	95.172 ± 0.034	0.0615 ± 0.0010	5.85 ± 0.10

- **relative performance increase $> 35\%$ over SS and OS**

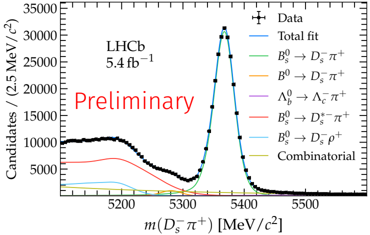


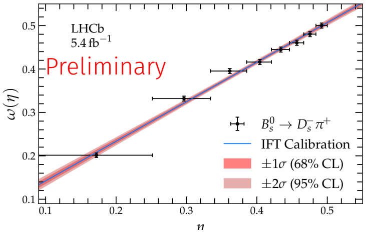


Calibration and performance: B_s^0-IFT

B_s^0-IFT performance on $B_s^0 \to D_s^- \pi^+$:

Year	Tagger	$\varepsilon_{\text{tag}}(\%)$	$\langle D^2 \rangle$	$\varepsilon_{\text{tag}}\langle D^2 \rangle(\%)$
2016	SS and OS	73.81 ± 0.16	0.0877 ± 0.0054	6.47 ± 0.40
	IFT	87.97 ± 0.13	0.0882 ± 0.0050	7.76 ± 0.44
2017	SS and OS	79.05 ± 0.15	0.0758 ± 0.0043	5.99 ± 0.34
	IFT	96.87 ± 0.06	0.0747 ± 0.0038	7.23 ± 0.37
2018	SS and OS	74.34 ± 0.15	0.0819 ± 0.0042	6.09 ± 0.31
	IFT	95.49 ± 0.07	0.0760 ± 0.0036	7.25 ± 0.34



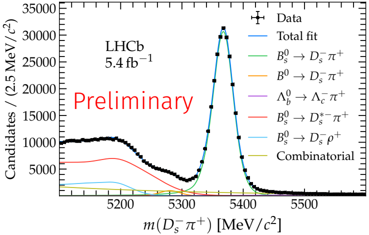


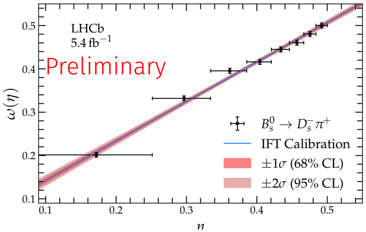
Calibration and performance: $\mathrm{B_s^0}$-IFT

B_s^0-IFT performance on $B_s^0 \to D_s^- \pi^+$:

Year	Tagger	$\varepsilon_{\text{tag}}(\%)$	$\langle D^2 \rangle$	$\varepsilon_{\text{tag}}\langle D^2 \rangle(\%)$
2016	SS and OS	73.81 ± 0.16	0.0877 ± 0.0054	6.47 ± 0.40
	IFT	87.97 ± 0.13	0.0882 ± 0.0050	7.76 ± 0.44
2017	SS and OS	79.05 ± 0.15	0.0758 ± 0.0043	5.99 ± 0.34
	IFT	96.87 ± 0.06	0.0747 ± 0.0038	7.23 ± 0.37
2018	SS and OS	74.34 ± 0.15	0.0819 ± 0.0042	6.09 ± 0.31
	IFT	95.49 ± 0.07	0.0760 ± 0.0036	7.25 ± 0.34

- **relative performance increase $> 20\%$ over SSand OS**



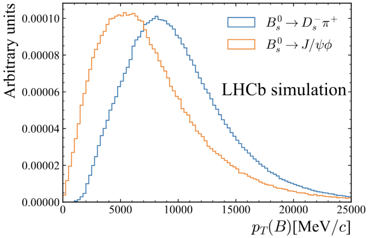


What's on the menu?

- Flavour Tagging at LHCb
- The Inclusive Flavour Tagger - concept and implementation
- IFT - training and results
- IFT - calibration and performance
- **Validation studies**
- Summary and outlook

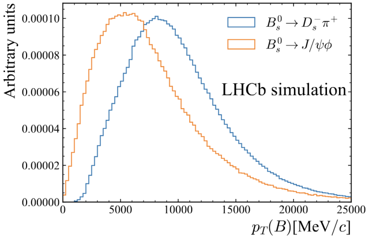
B_s^0-IFT portability: $\mathrm{B_s^0} \to \mathrm{J}/\psi \mathrm{K^+K^-}$

- Decay channel used for weak mixing phase ϕ_s-analysis (DOI:10.1103/PhysRevLett.132.051802)
- Calibration using kinematically reweighted $B_s^0 \to D_s^- \pi^+$ decays



B_s^0-IFT portability: $\mathrm{B}_s^0 \to \mathrm{J}/\psi \mathrm{K}^+\mathrm{K}^-$

- Decay channel used for weak mixing phase ϕ_s-analysis (DOI:10.1103/PhysRevLett.132.051802)
- Calibration using kinematically reweighted $B_s^0 \to D_s^- \pi^+$ decays
- Relative improvement comparable to the one in $B_s^0 \to D_s^- \pi^+$
- **Blind fit in ϕ_s shows 10% reduction in uncertainty, consistent with improvement in ε_{eff}**



Year	Tagger	$\varepsilon_{\text{tag}}(\%)$	$\langle D^2 \rangle$	$\varepsilon_{\text{tag}}\langle D^2 \rangle(\%)$
2016	SS and OS	74.91 ± 0.14	0.0623 ± 0.0039	4.67 ± 0.29
	IFT	91.18 ± 0.09	0.0609 ± 0.0035	5.55 ± 0.32
2017	SS and OS	76.18 ± 0.14	0.0582 ± 0.0034	4.44 ± 0.26
	IFT	93.18 ± 0.08	0.0578 ± 0.0030	5.38 ± 0.28
2018	SS and OS	72.31 ± 0.13	0.0628 ± 0.0033	4.54 ± 0.24
	IFT	90.32 ± 0.09	0.0604 ± 0.0029	5.46 ± 0.26

Portability systematics

- Taggers calibrated on flavour-specific decays & ported to other channels e.g. for CP measurement
- introduces systematic uncertainty due to different kinematics affecting calibration
  → found dominant in some analyses
- Uncertainty determined for both B^0- and B_s^0-IFT on MC

Portability systematics

- Taggers calibrated on flavour-specific decays & ported to other channels e.g. for CP measurement
- introduces systematic uncertainty due to different kinematics affecting calibration
  → found dominant in some analyses
- Uncertainty determined for both B^0- and B_s^0-IFT on MC
- found comparable to or smaller than for SSand OS combination

What's on the menu?

- Flavour Tagging at LHCb
- The Inclusive Flavour Tagger - concept and implementation
- IFT - training and results
- IFT - calibration and performance
- Validation studies
- **Summary and outlook**

Summary and outlook

- Development of a new DeepSet based inclusive flavour tagger with data and simulation from Run 2
- Performance gain of $35\%(20\%)$ with respect to combination of classical taggers in $B^0(B_s^0)$ decay modes
- Improvement confirmed in blinded ϕ_s fit
- Portability systematic uncertainties comparable with respect to SS and OS

Summary and outlook

- Development of a new DeepSet based inclusive flavour tagger with data and simulation from Run 2
- Performance gain of $35\%(20\%)$ with respect to combination of classical taggers in $B^0(B_s^0)$ decay modes
- Improvement confirmed in blinded ϕ_s fit
- Portability systematic uncertainties comparable with respect to SS and OS

Outlook:

- Run 2 analyses in progress that will use the IFT in their measurement
- IFT for Run 3 in development

LHCb-PAPER-2025-024 submitted to arXiV

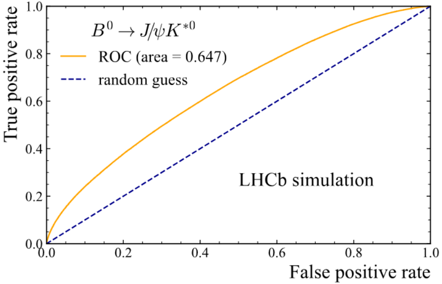
BACKUP

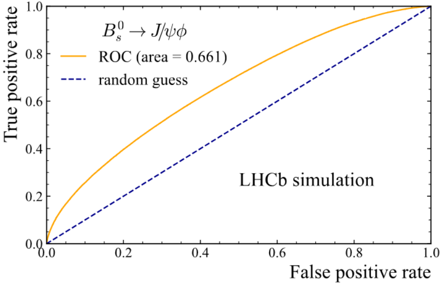
Input features

- charge
- momentum, transverse momentum
- χ^2 per degree of freedom (track fit)
- ionisation discharge inside the Vertex Locator (VELO)
- particle identification info (electron, muon, pion, kaon, proton, ghost)
- IP of a track to the decay vertex of the signal B
- BDT based isolation variable
- IP and χ^2_{IP}
- IP of track with respect to the PV of the signal B
- difference between pseudorapidity of track and signal B
- difference between z-coordinate of track and signal B
- cosine of difference between the azimuth angle of the track and signal B
- projection of track momentum on the momentum direction of signal B
- 8 single particle taggers

NN architecture & ROC

- 1 input layer, 3 hidden layers, 1 output layer
- 3980 trainable parameters
- Activation function: ReLU for intermediate and sigmoid for output layer
- Loss: binary cross-entropy





LHCb-PAPER-2025-024 submitted to arXiV

Combination of taggers

Procedure:

- Trained taggers calibrated individually on real data
- Combination of d_i and ω_i of individual taggers, assuming no correlation
- Second calibration of combined mistag $\omega_{\text{comb}}(\eta_{\text{comb}})$

Combination:

$$P(b) = \frac{p(b)}{p(b) + p(\bar{b})} = 1 - P(\bar{b})$$

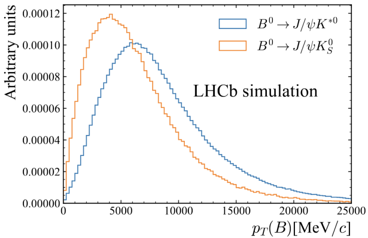
$$p(b/\bar{b}) = \prod_i \left(\frac{1 + d_i}{2} \mp d_i(1 - \omega_i) \right)$$

$$d_{comb} = \pm 1 \text{ if } P(b) \lessgtr P(\bar{b})$$

$$\eta_{comb} = \begin{cases} 1 - P(b) \text{ if } P(b) > P(\bar{b}) \\ P(b) \text{ if } P(b) < P(\bar{b}) \end{cases}$$

B^0-IFT portability: $\mathrm{B^0 \to J/\psi K^0_S}$

- Decay channel used for CP parameter $\sin(2\beta)$ analysis
- Here: simplified selection, using only $J/\psi \to \mu^+\mu^-$
- Calibration using kinematically reweighted $B^0 \to J/\psi K^+\pi^-$ decays
- Even larger improvement due to lower tagging efficiency in SS and OS
  $\rightarrow$ naive selection affects classical taggers more than IFT



Year	Tagger	$\varepsilon_{\text{tag}}(\%)$	$\langle D^2\rangle$	$\varepsilon_{\text{tag}}\langle D^2\rangle(\%)$
2016	SS and OS	69.18 ± 0.35	0.0473 ± 0.0011	3.27 ± 0.08
	IFT	95.42 ± 0.16	0.0563 ± 0.0012	5.37 ± 0.12
2017	SS and OS	68.65 ± 0.34	0.0482 ± 0.0011	3.31 ± 0.08
	IFT	98.53 ± 0.09	0.0571 ± 0.0012	5.63 ± 0.12
2018	SS and OS	69.58 ± 0.31	0.0487 ± 0.0010	3.38 ± 0.07
	IFT	96.41 ± 0.13	0.0573 ± 0.0011	5.52 ± 0.11

LHCb-PAPER-2025-024 submitted to arXiV