Hands on flavor tagging

Fernando Abudinén,
Jochen Gemmler, Stefano Lacaprara

October 29, 2019

INFN TRIESTE
Istituto Nazionale di Fisica Nucleare
Sezione di Trieste

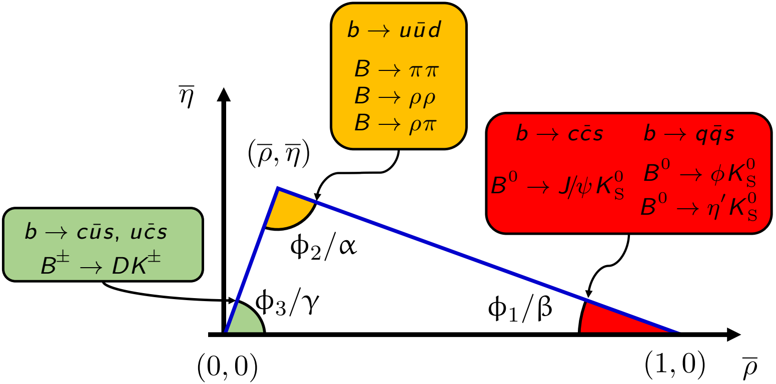
What is flavor tagging?

Flavor tagging ≙ determine the quark content of flavored hadrons.

⇒ In this talk: focus on determining the b-quark content of neutral B mesons.

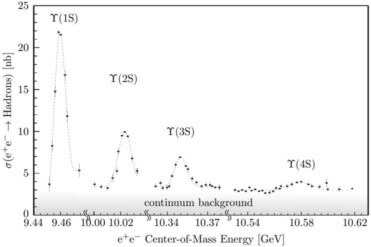
- Many measurements to determine parameters of flavor-physics require flavor tagging of B^0-mesons,

⇒ especially those of CP-asymmetries in decays to CP-eigenstates.

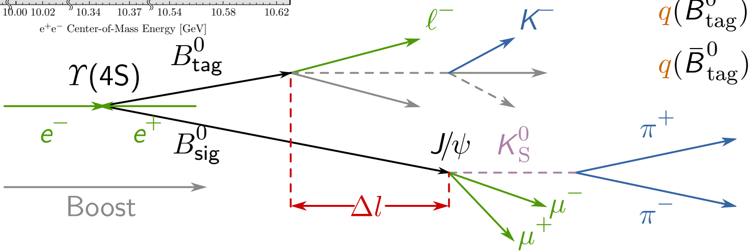


- The flavor info. has also separation power against continuum.

⇒ Can be exploited in many kinds of B-meson analyses.



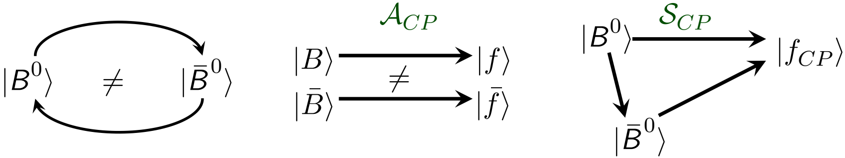
- Schoolbook example: $B^0_{\text{sig}} \to J/\psi K^0_S$.



$$\Rightarrow \Delta t = \frac{\Delta l}{\langle\beta\gamma\rangle c} \text{ since } B^0\bar{B}^0 \text{ at rest in } \Upsilon(4S) \text{ frame}$$

$$\mathcal{P}^{\text{Sig}}(\Delta t, q) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \left[1 + q\left(\mathcal{A}_{CP}\cos(\Delta m \Delta t) + \mathcal{S}_{CP}\sin(\Delta m \Delta t)\right)\right].$$

Time-dependent CP-Analysis



- E.g.: ${\mathcal{A}_{CP}}^{J/\Psi K_S^0} = 0, \qquad {\mathcal{S}_{CP}}^{J/\Psi K_S^0} = \sin(2\phi_1)$
- Flavor-tagging is possible only for a fraction ε of events.
- A fraction w of them is wrongly classified $\Rightarrow$

$$\mathcal{P}^{\text{Obs}}(\Delta t, q, \varepsilon, w) = \varepsilon \cdot \left[(1-w) \cdot \mathcal{P}^{\text{Sig}}(\Delta t, q) + w \cdot \mathcal{P}^{\text{Sig}}(\Delta t, -q)\right]$$

$$= \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \varepsilon \left[1 + q \cdot (1-2w)\left(\mathcal{A}_{CP}\cos(\Delta m \Delta t) + \mathcal{S}_{CP}\sin(\Delta m \Delta t)\right)\right]$$

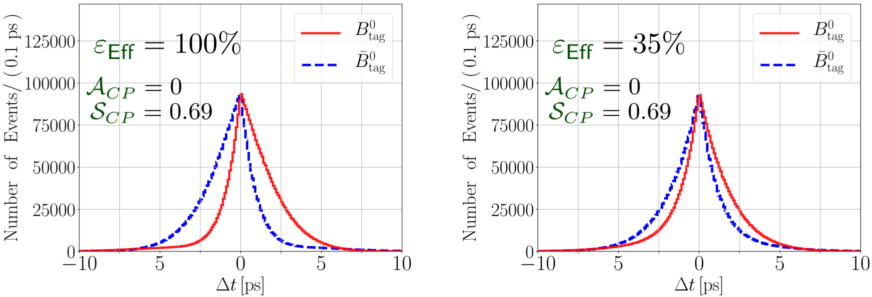
- Dilution factor: $\quad r \equiv |1-2w|$.

Effective Efficiency

- Statistical uncertainty on measured CP asymmetries: $\delta\mathcal{A}_{CP} \propto \frac{1}{\sqrt{N_{\text{tag}}}\cdot(1-2w)}$.
- Introduce N_{eff} (effective number of tagged events) and effective tagging efficiency ε_{Eff}:

$$\Rightarrow \delta\mathcal{A}_{CP} \propto \frac{1}{\sqrt{N_{\text{eff}}}} = \frac{1}{\sqrt{\varepsilon_{\text{Eff}}\cdot N_{\text{total}}}} \Rightarrow \varepsilon_{\text{Eff}} = \frac{N_{\text{tag}}}{N_{\text{total}}}\cdot(1-2w)^2 = \varepsilon\cdot r^2.$$

- Usually divide events in r-bins: $\Rightarrow \varepsilon_{\text{Eff}} = \sum_i \varepsilon_i \cdot r_i^2$.



How to tag flavor?

Many decays of neutral B mesons provide particles tagging the B-flavor.

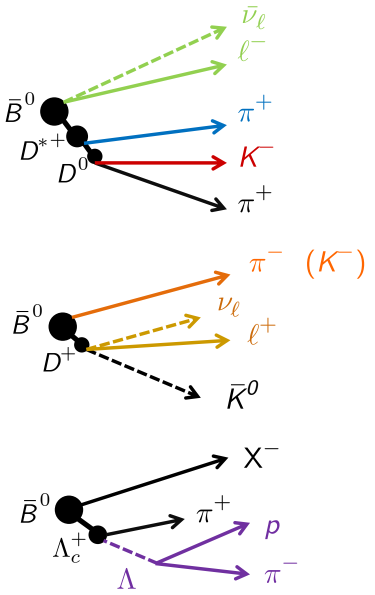
$\Rightarrow$ Use μ, e, $K^{\pm}$, $\pi^{\pm}$ ($\hat{=}$ single tracks) and Λ ($\hat{=}$ pairs of tracks).

$\Rightarrow$ Classify particles according to the decay:

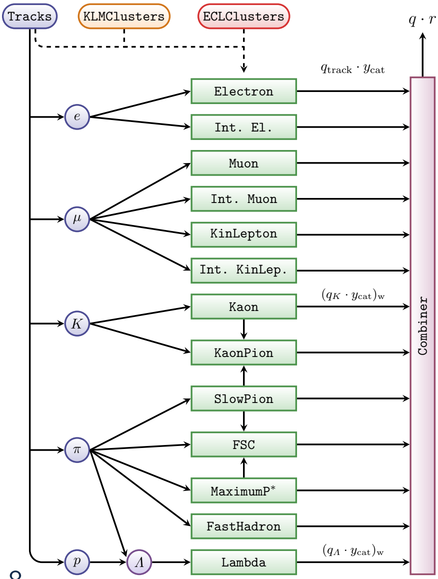
- Primary **electrons** and **muons** from $b \to \ell^{-} \bar{\nu}_{\ell}\, c(u)$.
- Intermediate **electrons** and **muons** from $b \to c \to \ell^{+} \nu_{\ell}\, s(d)$.
- **Kaons** and **lambdas** from $b \to c \to s$.
- **Slow pions** from D^{*+} produced via $b \to c$.
- Primary **kaons** and **pions** from $b \to \pi^{-}(K^{-})\, c(u)$.

Flavor-tagging categories

Categories	Targets
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
KinLepton	ℓ^-
Intermediate KinLepton	ℓ^+
Kaon	K^-
KaonPion	K^-, π^+
SlowPion	π^+
FastHadron	π^-, K^-
MaximumP	ℓ^-, π^-
Fast-Slow-Correlation	ℓ^-, π^+
Lambda	Λ
Total= 13	



Category-based algorithm



PID: combined and individual likelihoods.

Kin.: simple: p, p_{t}, $\cos\theta$, d_0, z_0, ...
elaborated: recoil mass, E_W^{90}, miss. p^*, $\cos\theta_{\text{THRUST}}$, ...

- Total: 220 Inputs. Unique variables: 108.

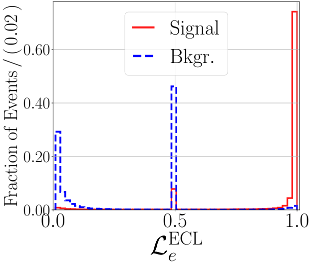
$\Rightarrow$ Calculated only once.

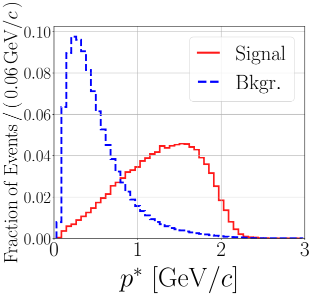
Boxes: Multi-variate Methods.

Default: FBDT.

x-check: 3-layer Perceptron (Only Combiner) FANN Library.

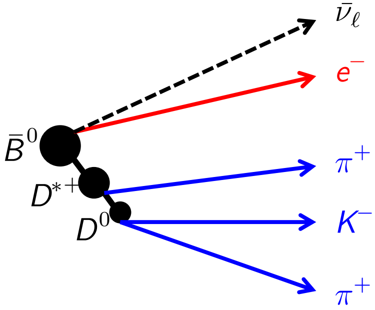
Tagging variables



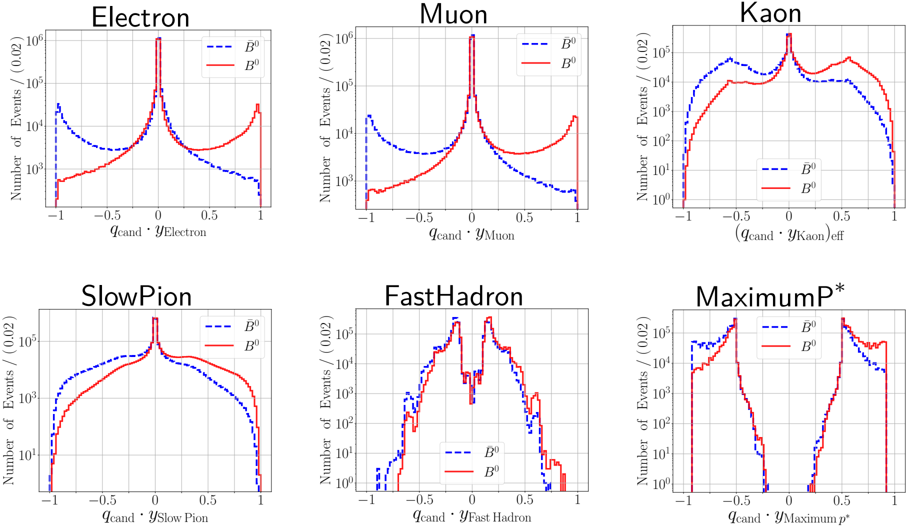


Two types:

- Particle identification (PID)
- Kinematic



Individual categories



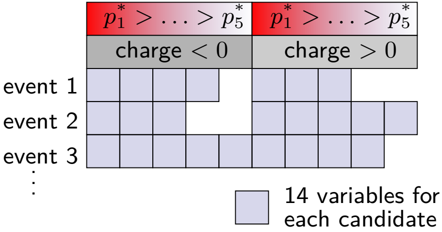
Deep-learning algorithm

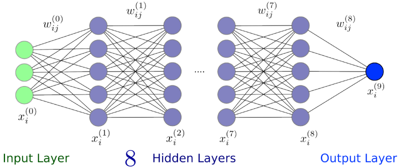
- 10 tracks at maximum
- sorted by momentum and grouped by charge.

Input: PID variables, momentum, azim. and polar angles, impact params., hit counts in SVD, PXD, CDC. Total = 140 Inputs.

MVA: Multi-layer percetron.

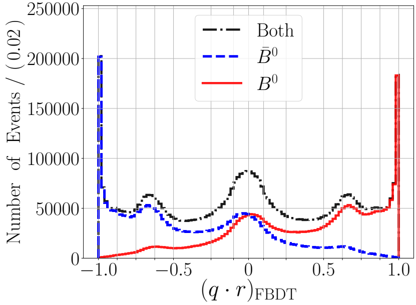
Libry.: Tensor-Flow ⇒ GPUs.





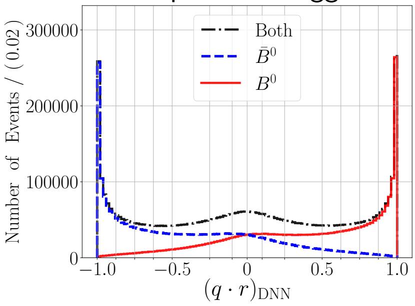
Standard flavor taggers on $J/\psi K_S^0$ simulation

Category-based Tagger



$$\varepsilon_{\text{Eff}} = (36.04 \pm 0.04)\%$$
$$\Delta\varepsilon_{\text{Eff}} = (\ 0.39 \pm 0.07)\%$$

Deep-neural Tagger



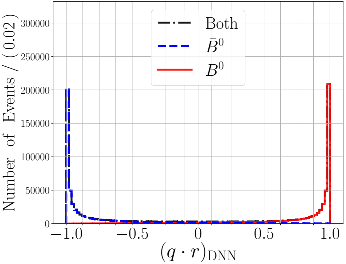
$$\varepsilon_{\text{Eff}} = (38.75 \pm 0.04)\%$$
$$\Delta\varepsilon_{\text{Eff}} = (\ 0.44 \pm 0.07)\%$$

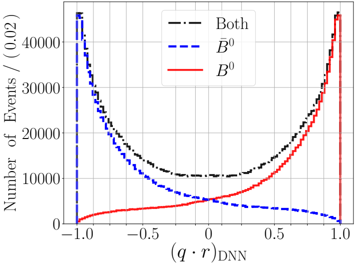
Samples from MC12b

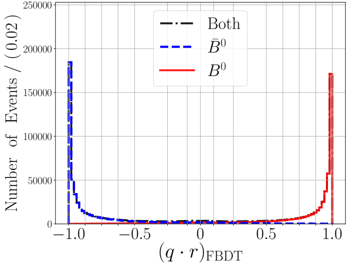
Test for specific categories

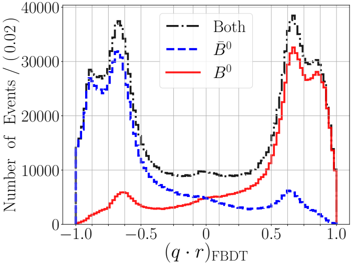
Events with primary leptons

Events with Kaons and no primary leptons









Comparison with data

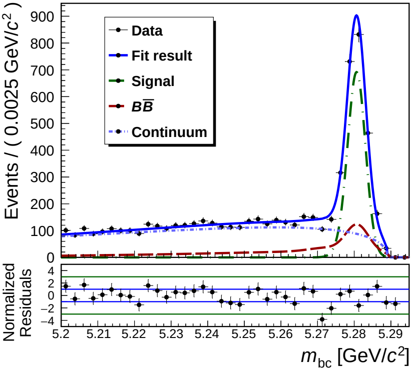
Issue: Output of standard algorithms shows data/MC mismatch.

Solution: (temporary) input variables with large data/MC mismatch are singled out.

Study:

- Reconstruct and select B^0 candidates on data (Proc9 and Bucket7 with 5.15 fb^{-1}) and on simulation (160 fb^{-1} of generic MC12b).
- Consider $B^0_{\rm sig}$-decay modes with large $\mathcal{B}$ (total $\sim 0.55\%$): $B^0_{\rm sig} \to D^{*-}(\pi^+, \rho^+, a_1^+)$ and $B^0_{\rm sig} \to D^-(\pi^+, \rho^+)$.
- Develop statistical model and perform $s\mathcal{P}lot$ to obtain distributions of input and output variables for signal and backgrounds on data.

Fit to data



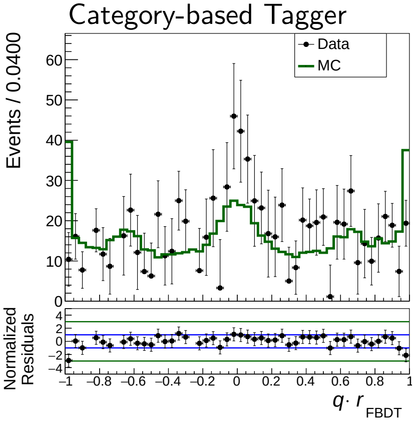
- See B2GM slides.
- $|\Delta E| < 0.05$ GeV.

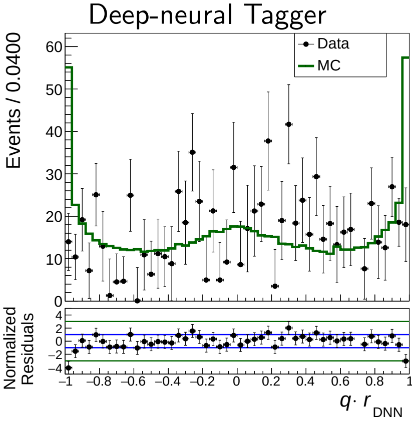
Parameter	Fit result
Signal yield	1725 ± 42
Cont. yield	3105 ± 69
μ_C [MeV]	1.03 ± 0.07
σ_C	0.96 ± 0.03
End point [GeV]	5.2894 ± 0.0003

$\Rightarrow$ Fix μ_C, σ_C and continuum endpoint to perform $s\mathcal{P}lot$.

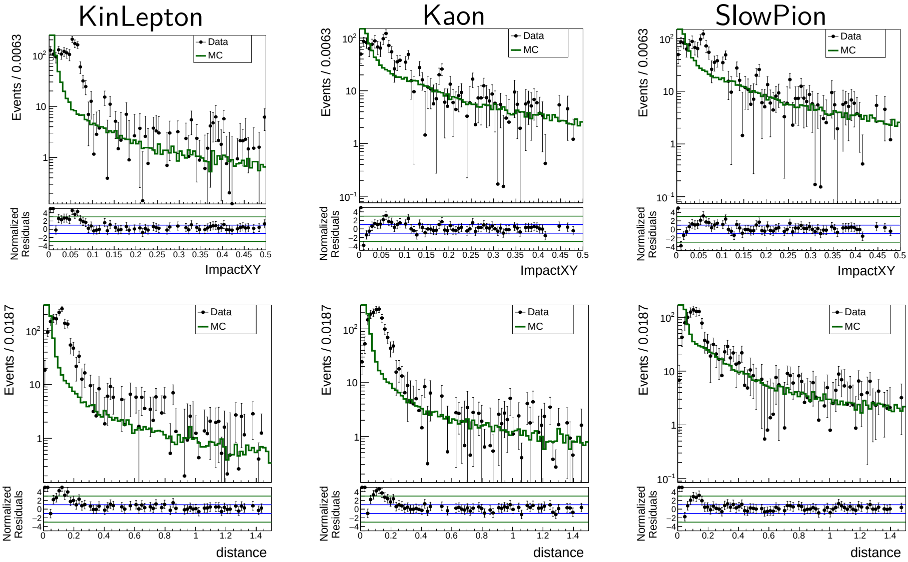
$\Rightarrow$ Compare distributions on data and on MC.

Standard flavor taggers on signal





Impact Parameters



Summary of findings

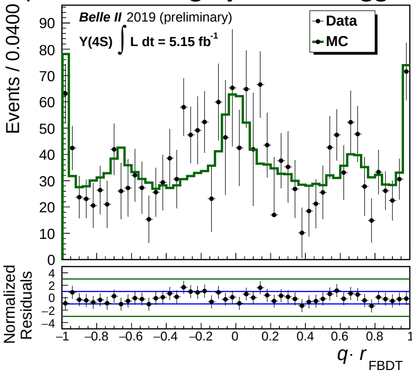
- Variables with large data/MC mismatch:
  d_0 and z_0.
- Variables with moderate data/MC mismatch:
  $\mathrm{d}E/\mathrm{d}x$ PID using CDC and SVD (specially for primary muons), p-value of tracks, and CDC hits.
- Combined PID variables have good data/MC agreement (apart from peak at about 0.2).
- All other input variables have good data/MC agreement, in particular PID variables using single subdetectors (ARICH, TOP, ECL, KLM) ☺.

⇒ Train the category-based tagger avoiding d_0 and z_0 for all candidates, and $\mathrm{d}E/\mathrm{d}x$ and p-value for muons, electrons, slow pions and fast hadrons.

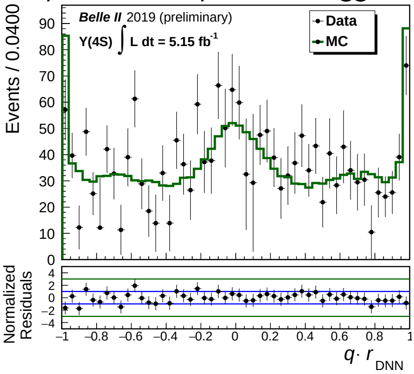
⇒ Train the deep-neural tagger avoiding d_0, z_0, p-value and CDC-hits for all candidates.

Optimized flavor taggers on signal

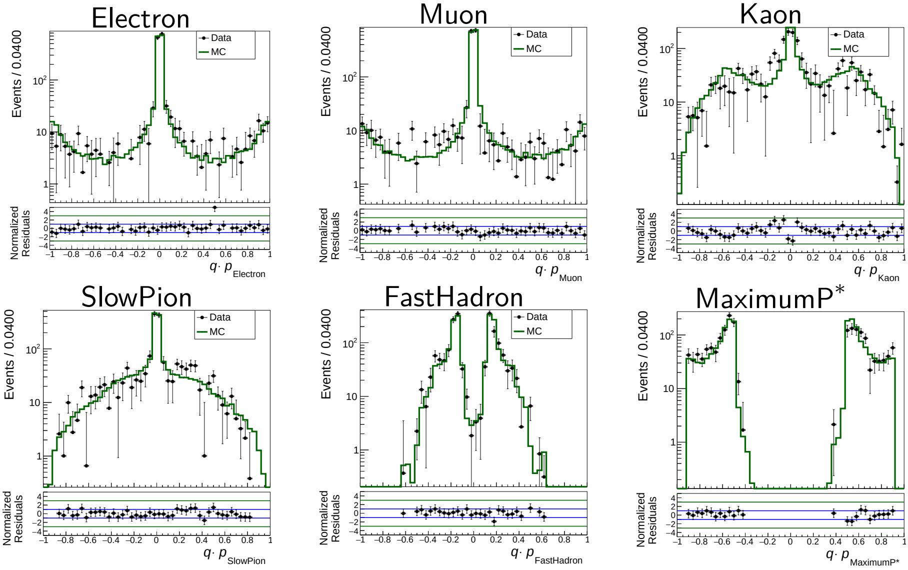
Optimized category-based Tagger



Optimized deep-neural Tagger

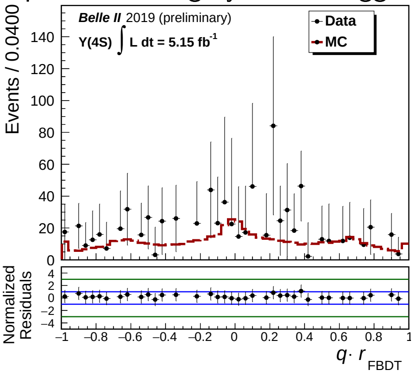


Optimized flavor tagger

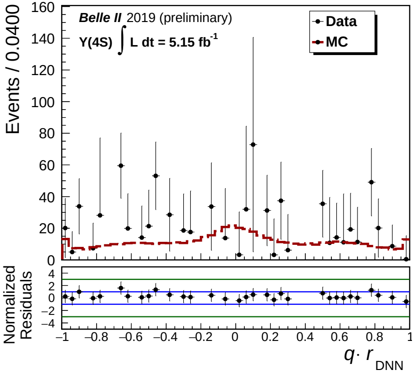


Optimized flavor taggers on $B\bar{B}$ + scf

Optimized category-based Tagger

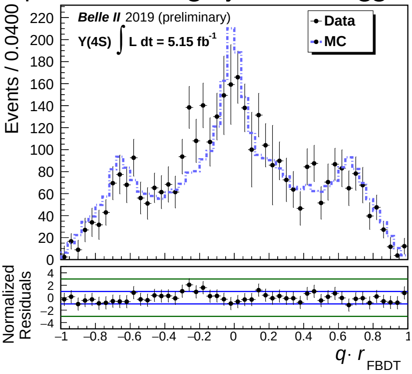


Optimized deep-neural Tagger

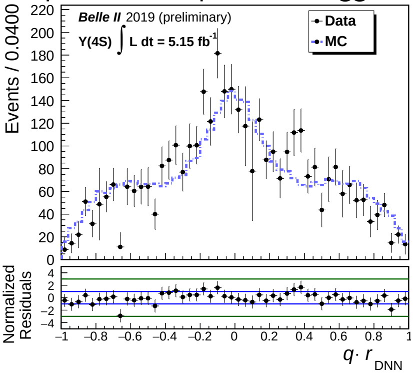


Optimized flavor taggers on continuum

Optimized category-based Tagger

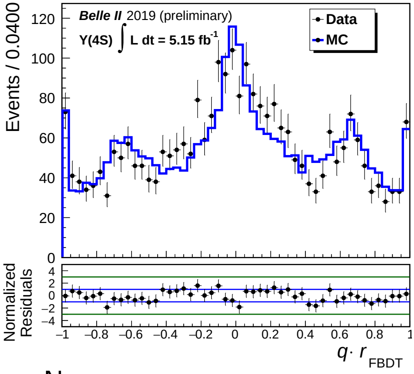


Optimized deep-neural Tagger

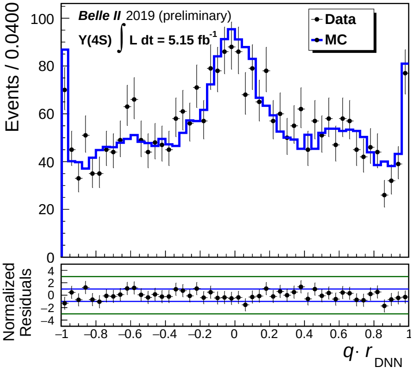


Require additionally $m_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$.

Optimized category-based Tagger



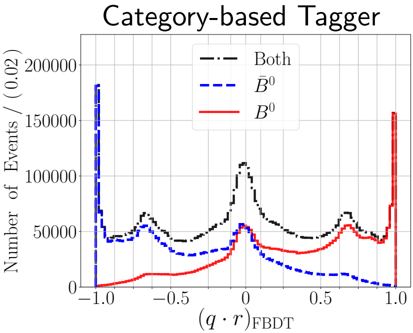
Optimized deep-neural Tagger

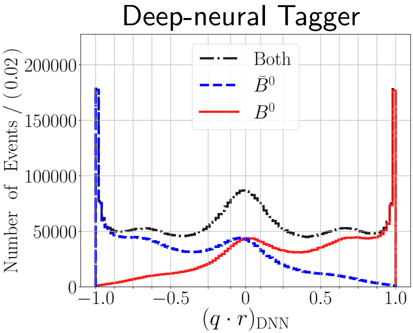


Note:

- Fraction of continuum events on data larger than on MC.
- Discrepancy in continuum description for $m_{\mathrm{bc}} < 5.27\,\mathrm{GeV}/c^2$.

Optimized flavor taggers on $J/\psi K_S^0$ simulation





Category-based Tagger:

$$\varepsilon_{\mathrm{Eff}} = (33.61 \pm 0.03)\%$$
$$\Delta\varepsilon_{\mathrm{Eff}} = (\ 0.64 \pm 0.07)\%$$

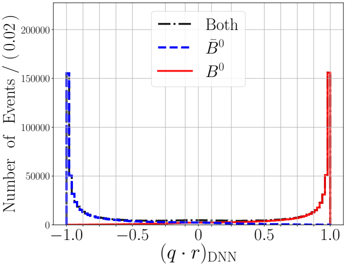
Deep-neural Tagger:

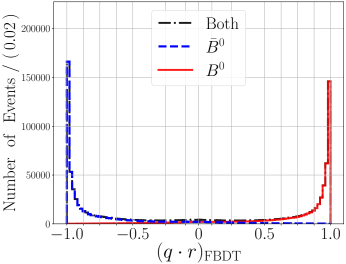
$$\varepsilon_{\mathrm{Eff}} = (33.84 \pm 0.03)\%$$
$$\Delta\varepsilon_{\mathrm{Eff}} = (\text{-}1.30 \pm 0.07)\%$$

Samples from MC12b

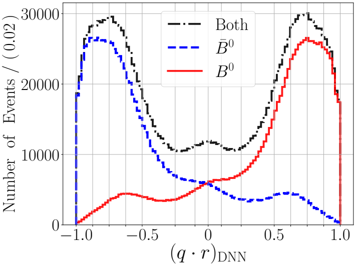
Test for specific categories

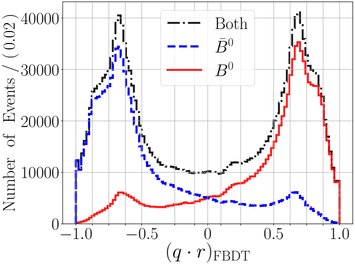
Events with primary leptons



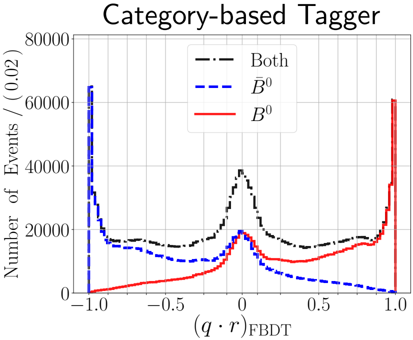


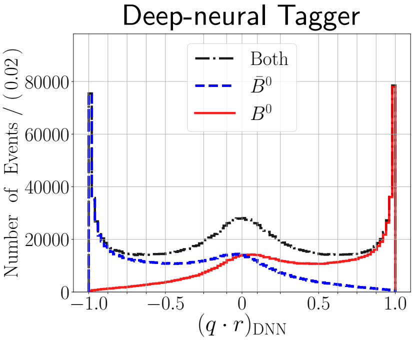
Events with Kaons and no primary leptons





Standard flavor taggers on $J/\psi K_S^0$ Belle MC





$\varepsilon_{\text{Eff}} = (34.44 \pm 0.06)\%$

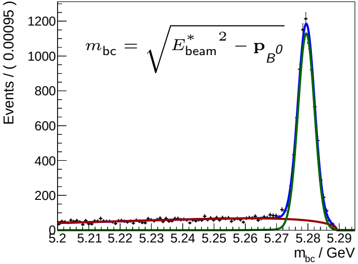
$\Delta\varepsilon_{\text{Eff}} = (\text{-}0.09 \pm 0.11)\%$

$\varepsilon_{\text{Eff}} = (34.67 \pm 0.06)\%$

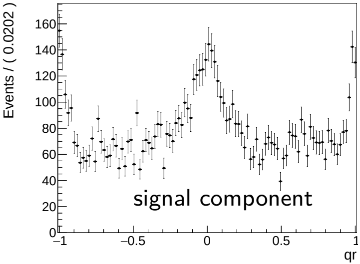
$\Delta\varepsilon_{\text{Eff}} = (\ 0.65 \pm 0.11)\%$

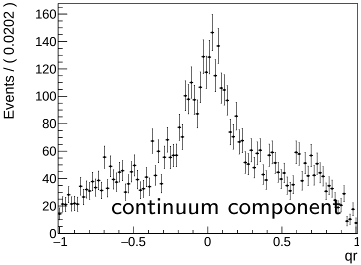
Using B2BII.

sPlot analysis on Belle Data



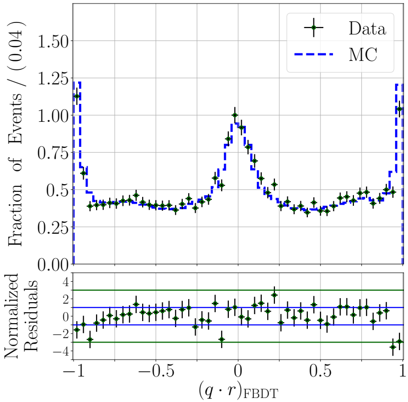
- Splot performed with converted Belle data using m_{bc} as conrol variable.
- Full Belle 0.8 ab^{-1} $B^0 \rightarrow J/\psi K^0_S$





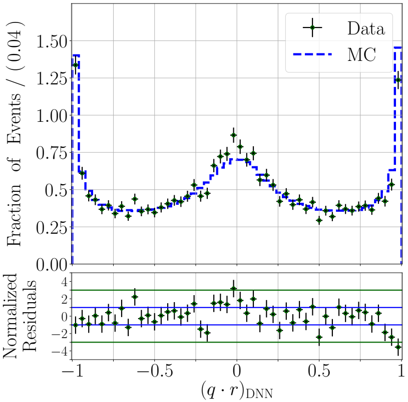
- Belle Data distribution for signal component (see excercises):

Category-based tagger



$\varepsilon_{\text{Eff}} = (34.2 \pm 0.5(\text{stat}))\,\%$

Deep-neural tagger



$\varepsilon_{\text{Eff}} = (34.0 \pm 0.5(\text{stat}))\,\%$

Performance: take-home numbers

Using Proc9, Bucket7 and MC12b ,
taking $B^0_{\rm sig} \to \nu_\tau \bar{\nu}_\tau$, $J/\psi K^0_S$, and $D^{(*)-}(\pi^+, \rho^+)$ into account:

- Standard cat-based tagger reaches $\varepsilon_{\sf Eff} = 37\%$ and standard DNN tagger $\varepsilon_{\sf Eff} = 40\%$ on simulation.

$\Rightarrow$ Large data/MC mismatch (especially in highest r-bin).

- Optimized taggers reach both about 35% on simulation.

$\Rightarrow$ Good data/MC agreement in all r-bins.

Using `B2BII`:

- Standard flavor taggers reach both about $\varepsilon_{\sf Eff} = 34\%$ on Belle simulation and data.
  (Belle reached 30% and BaBar 33%)
- Output distributions show good data/MC agreement.

Goal: Wrong-tag fractions (and other parameters) obtained from the output of the flavor taggers have to match the true ones.

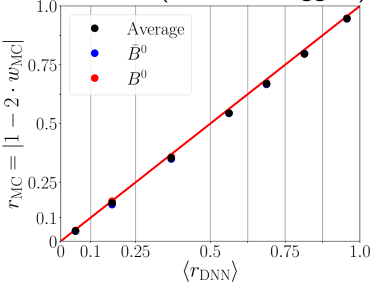
On simulation:

- Calculate the wrong-tag fraction w_{MC} using the true flavor q_{MC} and the dilution factor $r_{\mathrm{MC}} = |1 - 2w_{\mathrm{MC}}|$ (for each r-bin).
- Calculate the average $\langle r_{\mathrm{tagger}} \rangle = \langle |(q \cdot r)_{\mathrm{tagger}}| \rangle$ and compare in each r-bin).

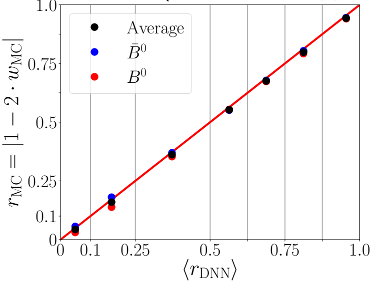
Calibration test on MC

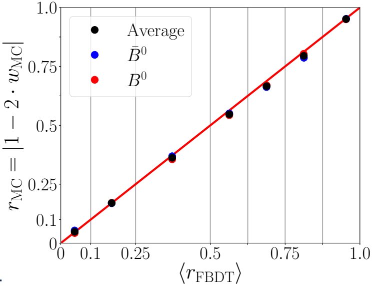
$B^0_{\rm sig} \to \nu\bar{\nu}$ (train) , $J/\psi K^0_{\rm S}$ (test)

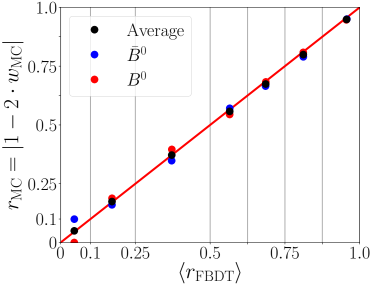
Belle MC (Standard taggers)



Belle II MC (Optimized taggers)







On data:

- Measure w_{data} by analyzing a flavor-specific B^0_{sig} decay. In this case, the signal Δt pdf. is given by

$$\mathcal{P}^{\mathsf{Obs}}(\Delta t, q_{\mathrm{sig}}, q_{\mathrm{tag}}, \varepsilon, w) = \frac{e^{\frac{-|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}}\varepsilon\left[1 - q_{\mathrm{sig}} \cdot q_{\mathrm{tag}} \cdot (1-2w) \cdot \cos(\Delta m \Delta t)\right].$$

$$\Rightarrow \quad \frac{\mathcal{P}^{\mathsf{Obs}}_{\mathsf{OF}}(\Delta t) - \mathcal{P}^{\mathsf{Obs}}_{\mathsf{SF}}(\Delta t)}{\mathcal{P}^{\mathsf{Obs}}_{\mathsf{OF}}(\Delta t) + \mathcal{P}^{\mathsf{Obs}}_{\mathsf{SF}}(\Delta t)} = (1-2w) \cdot \cos(\Delta m \Delta t).$$

OF: $q_{\mathrm{sig}} \neq q_{\mathrm{tag}}$

SF: $q_{\mathrm{sig}} = q_{\mathrm{tag}}$

Or time-integrated:

$$\mathcal{P}^{\mathsf{Obs}}(q_{\mathrm{sig}}, q_{\mathrm{tag}}, \varepsilon, w) = \frac{\varepsilon}{2}\left[1 - q_{\mathrm{sig}} \cdot q_{\mathrm{tag}} \cdot (1-2w) \cdot (1-2\chi_d)\right].$$

$$\Rightarrow \quad \frac{\mathcal{P}^{\mathsf{Obs}}_{\mathsf{SF}}}{\mathcal{P}^{\mathsf{Obs}}_{\mathsf{OF}} + \mathcal{P}^{\mathsf{Obs}}_{\mathsf{SF}}} = \frac{1-(1-2w)(1-2\chi_d)}{2} \hat{=} \chi_d^{\mathsf{Obs}} \qquad \Rightarrow w = \frac{\chi_d^{\mathsf{Obs}} - \chi_d}{1-2\chi_d}.$$

Excercises

- Go to `https://stash.desy.de/users/abudinen/repos/handsonexercises/browse`.
- Login with your DESY credentials!
- Login at kekcc.
- Follow instructions in the `README` file.
- I prepared one example for Belle II data/MC and for Belle data/MC using `B2BII`.

- So far considered $w = w_{B^0} = w_{\bar{B}^0}$ and $\varepsilon = \varepsilon_{B^0} = \varepsilon_{\bar{B}^0}$. But they can be different.

$$\Rightarrow w = \frac{w_{B^0}+w_{\bar{B}^0}}{2},\ \Delta w = w_{B^0} - w_{\bar{B}^0},$$

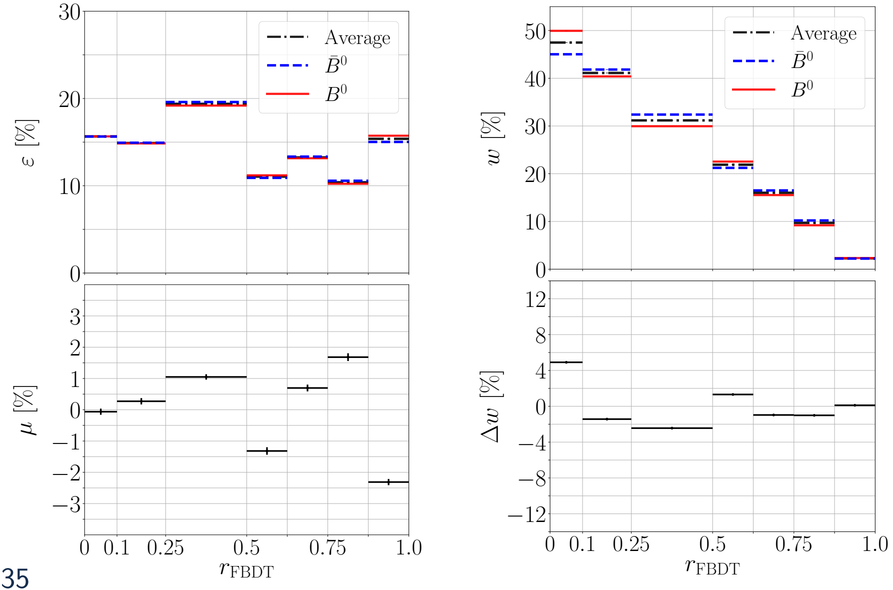
$$\Rightarrow \varepsilon = \frac{\varepsilon_{B^0}+\varepsilon_{\bar{B}^0}}{2},\ \ \mu = \frac{\varepsilon_{B^0}-\varepsilon_{\bar{B}^0}}{2\cdot\varepsilon}.$$

$$\Rightarrow \mathcal{P}^{\mathsf{Obs}}(\Delta t, q_{\text{sig}}, q_{\text{tag}}, \varepsilon, \mu, w, \Delta w) =$$

$$\frac{e^{\frac{-|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}}\varepsilon\Bigg[1 - q_{\text{tag}}\cdot\Delta w + q_{\text{tag}}\cdot\mu\cdot(1-2w)$$

$$-q_{\text{sig}}\cdot\left[q_{\text{tag}}\cdot(1-2w) + \mu\cdot(1-q_{\text{tag}}\Delta w)\right]\cdot\cos(\Delta m\Delta t)\Bigg].$$

Optimized cat.-based tagger on $J/\psi K^0_S$ MC12b



ϕ_1 from $b \to c\bar{c}s$ decays

Belle PR Lett 108 171802

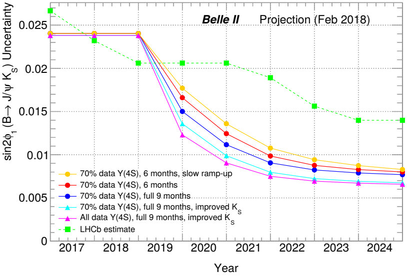
$$\mathcal{S}_{CP} = 0.667 \pm 0.023 \pm 0.012$$
$$\mathcal{A}_{CP} = 0.006 \pm 0.016 \pm 0.012$$

Systematic effect	$\delta\mathcal{S}_{CP}$	$\delta\mathcal{A}_{CP}$
Δt resolution	± 0.007	± 0.001
Vertexing	± 0.007	± 0.007
Tag-side interference	± 0.001	± 0.008
Flavor tagging	± 0.004	± 0.003
Possible fit bias	± 0.004	± 0.005
Signal fraction	± 0.004	± 0.002
Background Δt	± 0.001	< 0.001
Physics parameters	± 0.001	< 0.001
Total Belle	± 0.012	± 0.012

- $\sigma_{\text{stat}} = \sigma_{\text{stat}}^{\text{Belle}} \cdot \sqrt{\frac{\mathcal{L}^{\text{Belle}}}{50\ \text{ab}^{-1}}}$
- $\sigma_{\text{sys}} = \sqrt{(\sigma_{\text{red}}^{\text{Belle}})^2 \cdot \frac{\mathcal{L}^{\text{Belle}}}{50\ \text{ab}^{-1}} + (\sigma_{\text{ired}}^{\text{Belle}})^2}$

Belle II with $50\ \text{ab}^{-1}$

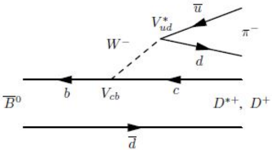
$$\delta\mathcal{S}_{CP} = \pm 0.003 \pm 0.004$$
$$\delta\mathcal{A}_{CP} = \pm 0.002 \pm 0.009$$



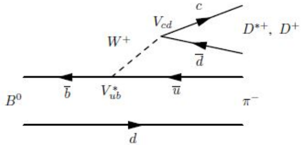
- Precision at Belle II dominated by systematic effects. $\Rightarrow$ effort concentrated in understanding and reducing them.
- Expected precision: $\phi_1^{c\bar{c}s} \lesssim 0.1^\circ$

The tag-side interference effect

PRD 68 034010



CKM-favored amplitude



doubly CKM-suppressed amplitude

$\Rightarrow$ Some hadronic final states possible for B^0 and for $\bar{B}^0$.

$\Rightarrow$ Introduces $r = \left|\frac{V^*_{ub}V_{cb}}{V_{cd}V_{ud}}\right| \approx 0.02$, δ and $\Phi = 2\beta + \gamma$.

- Many categories $\Rightarrow r', \delta'$.
- Belle and BaBar studied it via pseudo-experiments.

Belle: partial cancelation analyzing CP-even and CP-odd final states simulataneously.

$$\mathcal{C}_{\rm fit} = \mathcal{C}_0[1 + 2r'\cos\delta'\{\mathcal{G}\cos(2\beta+\gamma) - \mathcal{S}_0\sin(2\beta+\gamma)\}] - 2r'\sin\delta'\{\mathcal{S}_0\cos(2\beta+\gamma) + \mathcal{G}\sin(2\beta+\gamma)\} \quad (31)$$

$$\mathcal{S}_{\rm fit} = \mathcal{S}_0[1 + 2r'\cos\delta'\mathcal{G}\cos(2\beta+\gamma)] + 2r'\sin\delta'\mathcal{C}_0\cos(2\beta+\gamma), \quad (32)$$

where $\mathcal{G} \equiv 2\,\mathrm{Re}\,\lambda_{CP}/(|\lambda_{CP}|^2+1)$. Note that, with respect to the nominal values, there are both multiplicative and additive corrections which are proportional to $\cos\delta'$ and $\sin\delta'$ respectively. In the limit of a vanishing effective tag-side strong phase difference ($\delta' \to 0$), only the multiplicative corrections remain.

The tag-side interference effect

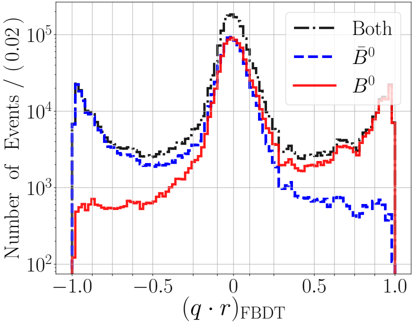
Options:

- Try to measure r', δ'?
- Get rid of it fully by tagging only with primary leptons.

$\Rightarrow$ Reduces ε_{Eff} by about one third.

$\Rightarrow$ Increases the statistical uncertainties on CP-violation parameters by about $\sqrt{3} \approx 1.7$.

Category-based tagger using only the Electron, Muon and KinLepton categories



$$\varepsilon_{\text{Eff}} = (11.99 \pm 0.03)\%$$

$$\Delta\varepsilon_{\text{Eff}} = (\text{-}0.09 \pm 0.07)\%$$

- We have two functional algorithms for flavor tagging on Belle and on Belle II simulation and data.

Belle II: Compromise between eff. efficiency and data/MC agreement avoiding some input variables.

⇒ Temporary until simulation improves.

⇒ Room for improvement as simulation of beam-spot, PID, tracking, continuum background, etc. improves.

- Optimized taggers will be default in `release-04`.
- To be continued: measurement of wrong-tag fractions on data: We need you!
- Test in future combination of DNN and cat.-based algorithms.
- Hope to face challenges at high lumi in not too far future.

Thanks for your attention!

- Train (and test performance) using correctly MC-matched event: `abs(qrMC==1)`.

- Train on samples where $B^0_{\rm sig} \to \nu_\tau \bar{\nu}_\tau$:

⇒ No dependency on performance of MC association (ROE built for B^0 MC particle).

⇒ Channels produced by default without built-in CP violation.

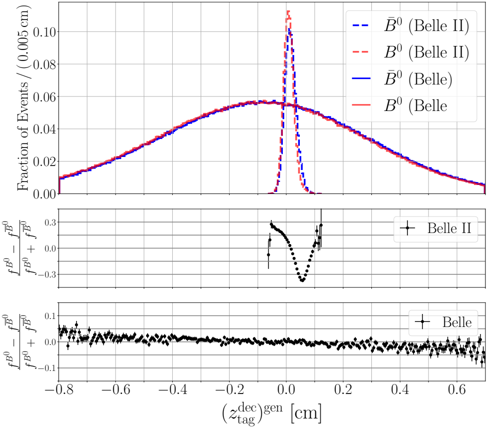
⇒ Same number of ROE tracks as correctly matched $J/\psi K^0_{\rm S}$.

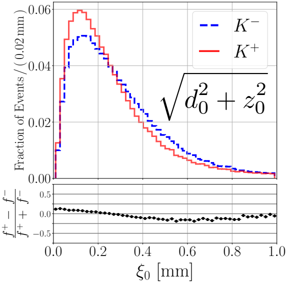
- Train with 10 Mio events (and test with the same number).

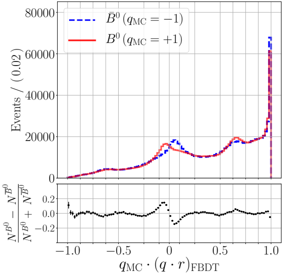
If you want to train yourself, check (`release-04` or newer):

`analysis/release-validation/CPVTools/README.md`

Train without built-in CP!







$\Rightarrow$ We have to train using MC **WITHOUT** built-in CP-violation. Otherwise MVAs learn from it.