Summer Student Presentation

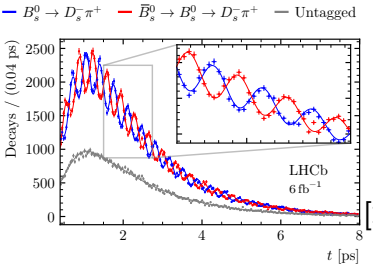
A new approach for Primary Vertex association using Machine Learning techniques

Nico Schiel
(30.06.-05.09.)
Supervisors: Laurent Dufour, Mark Waterlaat

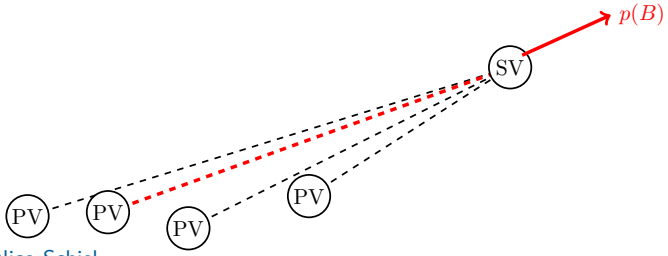
19.08.2025

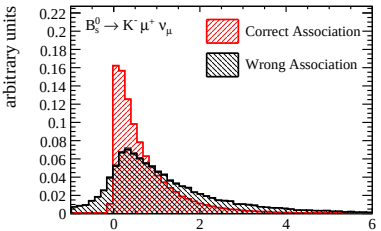
Why PV Association?

- Lifetime Measurement and semileptonic decays essential for LHCb physics program
- E.g. Time-dependent CP Violation
- Need distance of flight to reconstruct lifetime
- Need correct PV association
- $\mu = 5.4$ PVs per bunch crossing



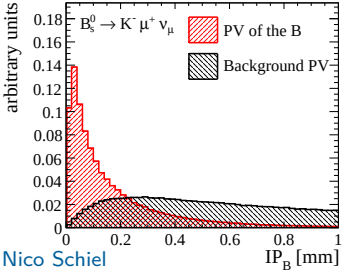
[2]

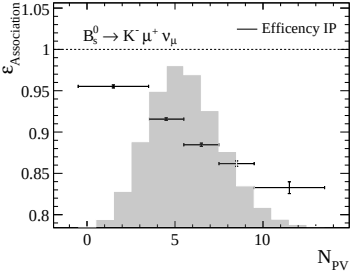


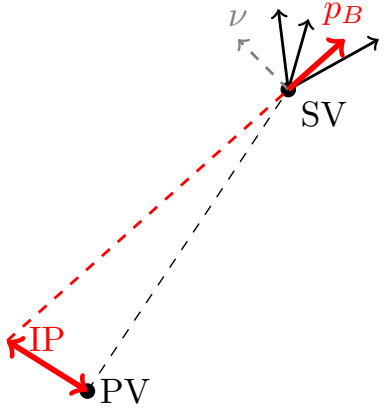


The current approach of PV Association

- Minimize IP
- Efficiencies $B_s^0 \to K^- \mu^+ \nu$ $\varepsilon_{\text{Association}} = 90.21 \pm 0.02\,\%$
- Efficiencies $B^0 \to K^{*0} J/\Psi$ $\varepsilon_{\text{Association}} = 99.23 \pm 0.02\,\%$
- Goal : **Improve PV association**
- **Implement**, **train**, and **test** different **ML Classifier** and **inputs** and implement it **into the LHCb software**



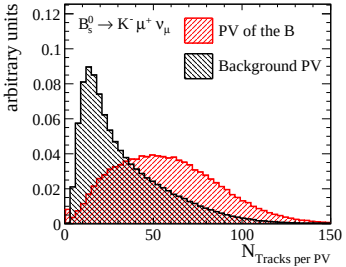


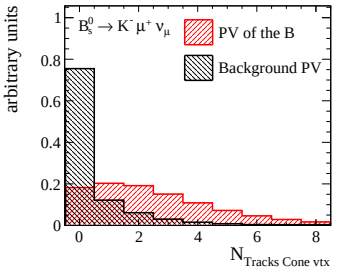


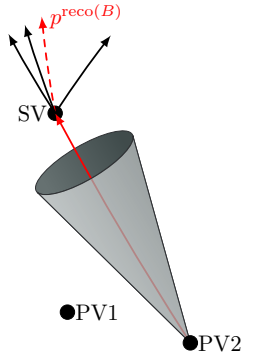
How to Improve PV Association?

Sensitive Features

- Test multiple Features
- Large differences Sig. vs Bkg.
- Variables uncorrelated to distance of flight and corrected mass
- Most sensitive Features:
  - IP
  - $N_{\text{Tracks per PV}}$
  - $N_{\text{Tracks Cone VTX}}$
  - $\eta(B)$
  - $p_{\text{T}}(B)$
  - N_{PV}



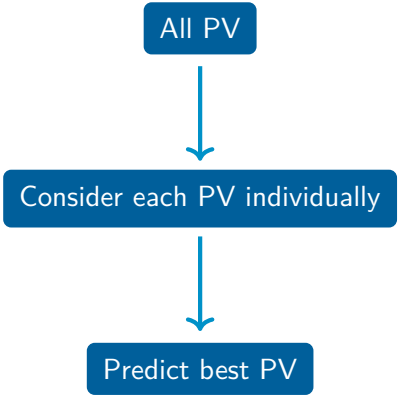




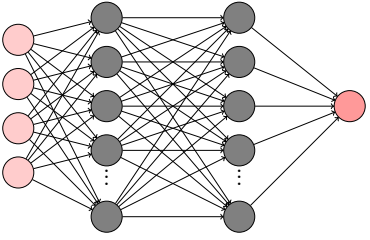
$$(\Delta R)^2 = (\Delta\eta)^2 + (\Delta\phi)^2$$

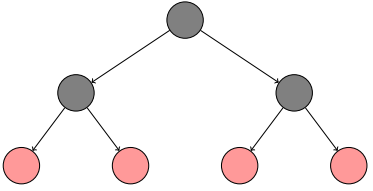
How to Improve PV Association?

Binary Classifier



- My best PV is the one with the highest classifier score
- Goal: Usable for all decay modes of $B_{(s)}$ hadrons
- Simultaneous training on multiple samples, including partially and fully reconstructed decays
- All studies in MC simulations
- Remove splitted and merged and not truth matched PVs

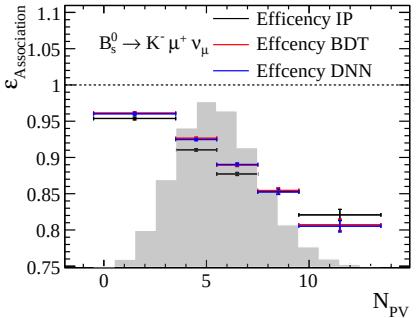




Performance of the Binary Classifier

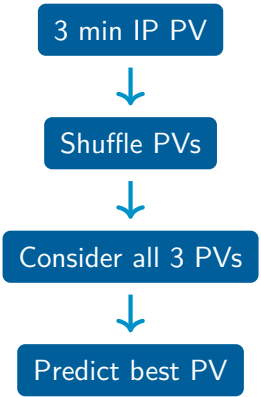
	$B^0 \to K^{*0} J/\Psi$	$B^0 \to K_S J/\Psi$	$B^+ \to K^+ J/\Psi$	$B_s^0 \to K^- \mu^+ \nu$
$\varepsilon_{\text{Assosiation}}$ IP	99.2 %	99.3 %	99.3 %	90.2 %
$\varepsilon_{\text{Assosiation}}$ BDT	99.1 %	99.3 %	99.2 %	91.8 %
$\varepsilon_{\text{Assosiation}}$ DNN	99.3 %	99.2 %	99.2 %	91.7 %

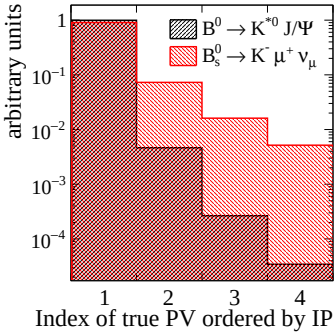
- Same performance for fully reconstructed decays compared to minimum IP method
- Improvement for partially reconstructed decays
- BDT and DNN performance similar



How to Improve PV Association?

Mulitclass Classifier





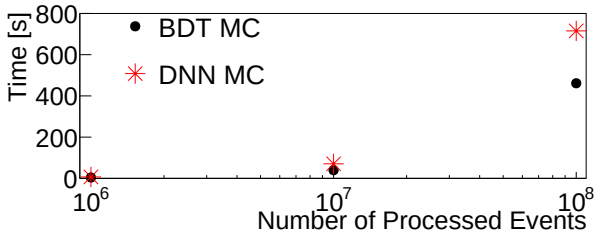
- Train on 3 PVs simultaneously
- Use the 3 PVs with min IP
- Assumption is false for
  - 0.98% partially reconstructed events
  - $< 0.1\%$ fully reconstructed events
- Model determines my best PV
- Same samples as for the binary classifier

Performance of the Multiclass Classifiers

	$B^0 \to K^{*0} J/\Psi$	$B^0 \to K_S J/\Psi$	$B^+ \to K^+ J/\Psi$	$B_s^0 \to K^- \mu^+ \nu$
$\varepsilon_{\text{Assosiation}}$ IP	99.2 %	99.3 %	99.2 %	90.2 %
$\varepsilon_{\text{Assosiation}}$ BDT MC	99.5 %	99.5 %	99.6 %	93.2 %
$\varepsilon_{\text{Assosiation}}$ DNN MC	99.3 %	99.3 %	99.4 %	91.9 %

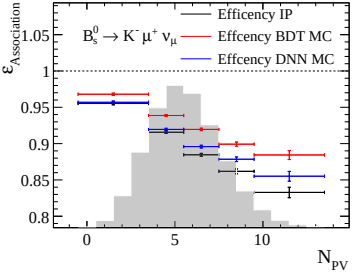
- Multiclass classifier performs better than binary
- BDT performs better than DNN
- BDT faster training than DNN
- BDT faster inference than DNN

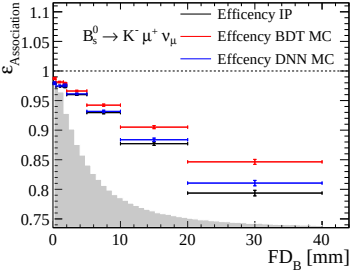
Timing the inference with ONNX Runtime CPU



Performance of the Multiclass BDT

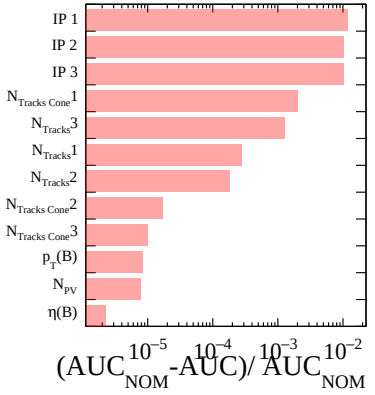
- Distributions more uniform
- Still differential dependencies
- IP still most important variable for classification
- Track distribution variables important





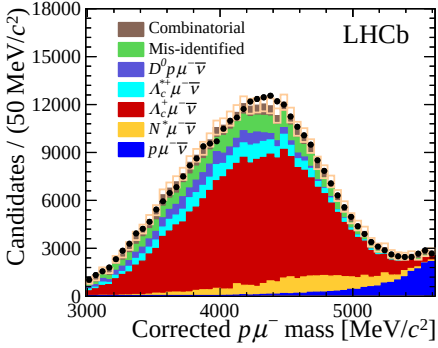
Legend: min IP, BDT MC, DNN MC

Feature Importance first output



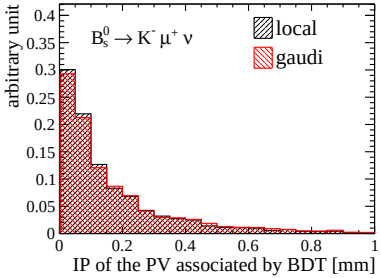
Additional Input Features

- Further variables improve performance of model
  - Distance of flight
  - Corrected mass (partially reconstructed decays)
- **Problem** often analyses perform fits to these or correlated variables
  - PV association introduces bias in background if variables included in training
- Efficiencies for partially reconstructed $B_s^0 \to K^- \mu^+ \nu_\mu$ with corrected mass
  $\varepsilon_{\text{Association}} = 95.8\,\%$
- In the following model without biasing variables



[1]

Implementation into LHCb software



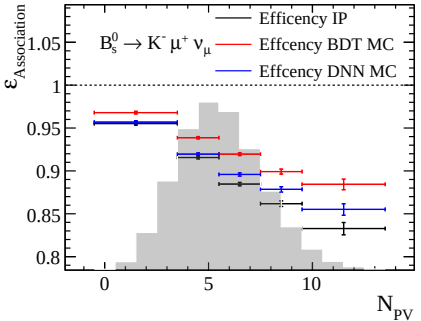

```
from PyConf.Algorithms import PV_Association_BMesons_BDTMulticlass

def main():
    ...
    out = PV_Association_BMesons_BDTMulticlass(
                    InputParticles = InputParticles,
                    InputVertices  = InputVertices)
    OutputParticles    = out.OutputParticles
    OutputProbability  = out.OutputProbability
    ...
```

- Goal: Implement into LHCb software (in `Rec/Tr/PatPV`)
- Runs and gives promising first results
- Trying to write merge request before I leave
- Algorithm does all the work for you
  - Calculation of input variables
  - Application of the MVA model via ONNX Runtime
  - Gives out particles with new associated PV and probability
- Easily adjustable due to ONNX

Summary and Outlook

- Significantly improve PV association
  - Especially for partially reconstructed decay modes on average 3 %
  - Larger improvements differentially
- Best performance by Multiclass BDT
- Started implementation into LHCb software
- Outlook:
  - Fine-tuning of the model
  - Validate method
  - Add splitted vertices to training
  - Finalise implementation into LHCb software
  - Look at Upgrade II conditions




```
from PyConf.Algorithms import PV_Association_BMesons_BDTMulticlass

def main():
    ...
    out = PV_Association_BMesons_BDTMulticlass(
                    InputParticles = InputParticles,
                    InputVertices  = InputVertices)
    OutputParticles     = out.OutputParticles
    OutputProbability   = out.OutputProbability
    ...
```

Thank you for your attention!

And for the great time during the Summer School!!!

References

[1] LHCb Collaboration. "Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays". In: *Nature Phys.* 11 (2015), pp. 743–747. DOI: 10.1038/nphys3415. arXiv: 1504.01568 [hep-ex].

[2] LHCb Collaboration. "Precise determination of the B_s^0–$\overline{B}_s^0$ oscillation frequency". In: *Nature Phys.* 18.1 (2022), pp. 1–5. DOI: 10.1038/s41567-021-01394-x. arXiv: 2104.04421 [hep-ex].

Backup

Event Selection and Object Definition

	$B^0 \to K^{*0} J/\Psi$	$B^0 \to K_S J/\Psi$	$B^+ \to K^+ J/\Psi$	$B_s^0 \to K^+ \mu^- \nu$
$m_{J/\Psi}$ [MeV]	$[2847, 3347]$			-
m_K [MeV]	$[826, 966]$	$[300, 700]$	-	-
m_B [MeV]	$[5150, 5340]$		$[5000, 5600]$	$[0, 5500]$
$\frac{\chi^2_B}{\text{NDF}_B}$	< 20			
$\frac{\chi^2_{K/J/D}}{\text{NDF}_{K/J/D}}$	< 16			

- Final state particles truth matched to mother particle
- $J/\Psi \to \mu^+ \mu^-$

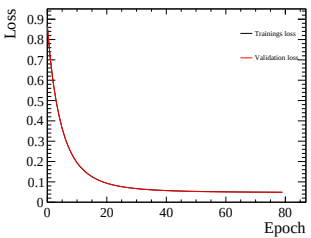
- $K_s \to \pi^+ \pi^-$
- $K^{*0} \to K^+ \pi^-$
- $p_T > 400\,\text{MeV}$

- $p > 4000\,\text{Mev}$
- $1.8 < \eta < 5.0$
- $p_B > 1000\,\text{MeV}$

Model BDT and DNN Binary

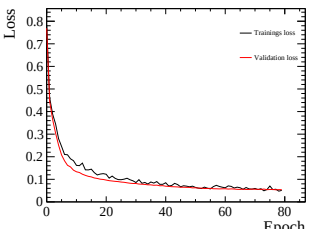
BDT

- Gradient boosted decision tree
- Depth = 3
- Maximum number of leaves = 30
- Maximum Number of trees = 80
- Learning rate = 0.1



DNN

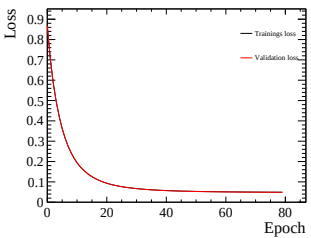
- Deep Neutral Network
- Min max input scaling
- Tried different architectures
  - Relu activation function
  - Sigmoid output activation function
  - Layer nodes : 64 - 32 - 16 - 1
  - Dropout : 0.1 - 0.0 - 0



Model BDT and DNN MC

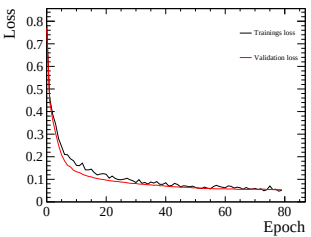
BDT

- Gradient boosted decision tree
- Depth = 3
- Maximum number of leaves = 30
- Maximum Number of trees per class = 80
- Learning rate = 0.1



DNN

- Deep Neutral Network
- Min max input scaling
- Tried different architectures
  - Relu activation function
  - Softmax output activation function
  - Layer nodes : 128 - 64 - 32 - 3
  - Dropout : 0.2 - 0.1 - 0



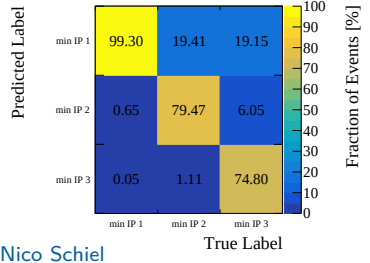
Efficiencies with Uncertainties

	$B^0 \to K^{*0} J/\Psi$	$B^0 \to K_S J/\Psi$	$B^+ \to K^+ J/\Psi$	$B_s^0 \to K^- \mu^+ \nu$
$\varepsilon_{\text{Assosiation}}$ IP	$99.21 \pm 0.02\,\%$	$99.33 \pm 0.08\,\%$	$99.29 \pm 0.01\,\%$	$90.23 \pm 0.03\,\%$
$\varepsilon_{\text{Assosiation}}$ BDT	$99.11 \pm 0.02\,\%$	$99.35 \pm 0.08\,\%$	$99.22 \pm 0.01\,\%$	$91.84 \pm 0.02\,\%$
$\varepsilon_{\text{Assosiation}}$ DNN	$99.31 \pm 0.02\,\%$	$99.24 \pm 0.08\,\%$	$99.82 \pm 0.01\,\%$	$91.79 \pm 0.02\,\%$
$\varepsilon_{\text{Assosiation}}$ BDT MC	$99.58 \pm 0.02\,\%$	$99.51 \pm 0.08\,\%$	$99.62 \pm 0.01\,\%$	$93.26 \pm 0.02\,\%$
$\varepsilon_{\text{Assosiation}}$ DNN MC	$99.37 \pm 0.03\,\%$	$99.32 \pm 0.08\,\%$	$99.40 \pm 0.01\,\%$	$91.91 \pm 0.02\,\%$

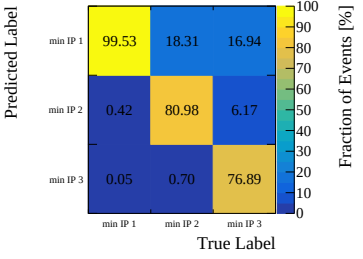
Differences Binary and Multiclass

- Divide sample into classes
  - minIP 1 = True PV is reconstructed one with the smallest IP
  - minIP 2 = True PV is reconstructed one with second-smallest IP
  - minIP 3 = True PV is reconstructed one with third-smallest IP
- Binary confuses minIP 2 and minIP 3 with minIP 0
- Multiclass much less confusion

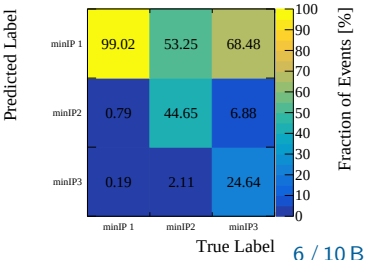
DNN MC



BDT MC



BDT Binary

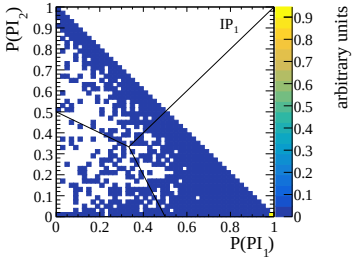


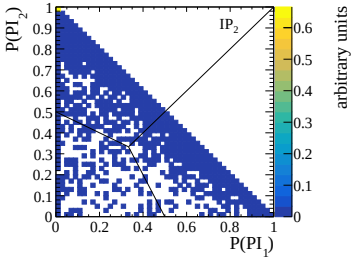
Output of the Multiclass BDT

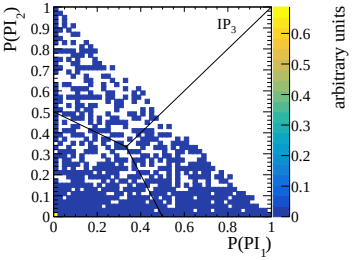
- Output scaled by softmax $\text{softmax}(\vec{x}, i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$

$\Rightarrow$ Sum of all outputs is one

- Model in general very sure about its classification

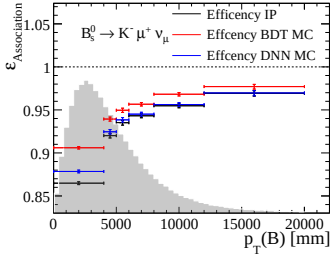


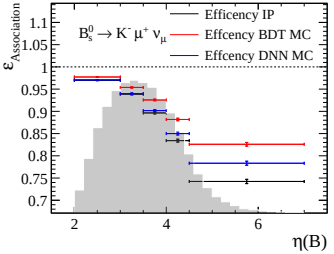


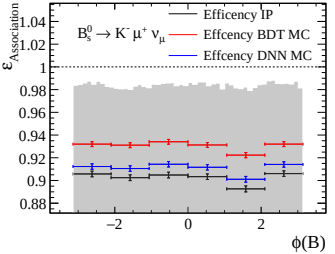


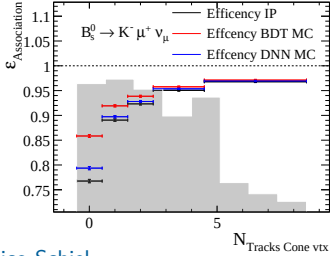
Differential Distribution

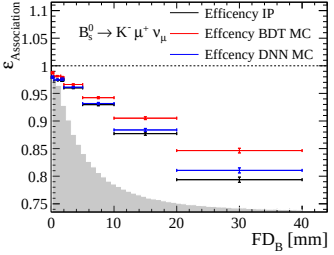
$B_s^0 \to K^- \mu^+ \nu_\mu$

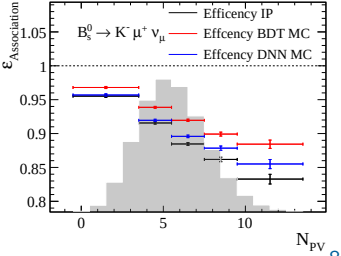






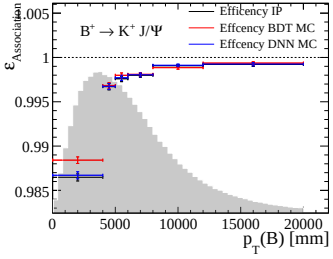


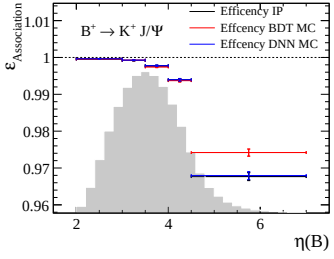


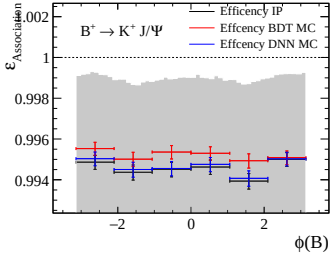


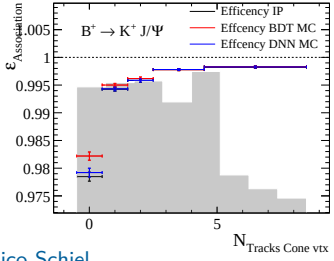
Differential Distribution

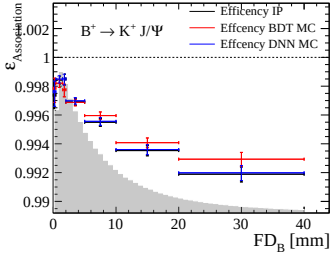
$B^+ \to K^+ J/\Psi(\mu^+\mu^-)$

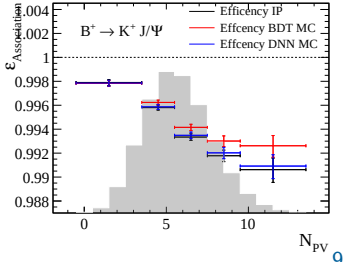




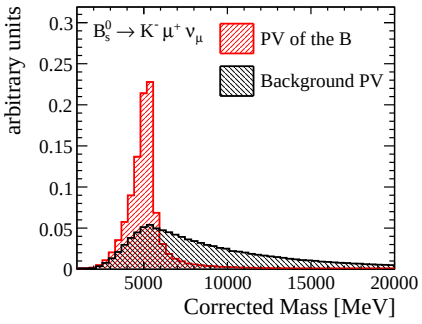








Corrected Mass



- Account for missing momentum e.g. from neutrino
- Assume B travels in direction given by the connecting linr of PV and SV
- $p_{\mathrm{T,\,FD}}$ is momentum of the B orthogonal to the axis connecting PV and SV
- Assume this is missing transversal momentum
- Cannot account for missing longitudinal momentum
- $m^{\mathrm{corr}} = \sqrt{(m^{\mathrm{reco}})^2 + p^2_{\mathrm{T,\,FD}}} + p_{\mathrm{T,FD}}$