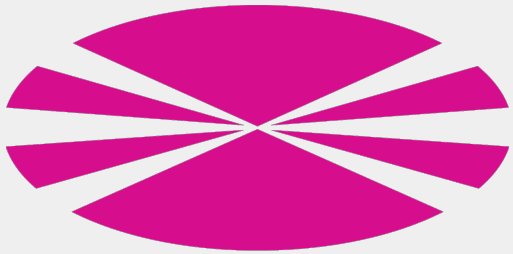
Inclusive Flavour Tagging at the LHCb experiment

John Wendel

XVI CPAN DAYS
November 20, 2024



MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
UNIÓN EUROPEA
FONDO EUROPEO DE DESARROLLO REGIONAL
"Una manera de hacer Europa"
AGENCIA ESTATAL DE INVESTIGACIÓN
INTALENT
UDC - INDITEX
XUNTA DE GALICIA

CP violation

- Our (observable) universe consists mostly of matter, very low amount of anti-matter
- Early universe: Equal amount of matter and anti-matter
  $\rightarrow$ Why did that change?

CP violation

- Our (observable) universe consists mostly of matter, very low amount of anti-matter
- Early universe: Equal amount of matter and anti-matter
  $\rightarrow$ Why did that change?

CP Violation:

- Describes asymmetric behaviour of particles if charge (C) and parity (P) are conjugated
  $\rightarrow$ Different behaviour of particles and anti-particles
- Included in the standard model of particle physics, but not as large as observed
- Neutral particles containing b quarks (B^0 mesons) experience large CP violation effects

Time dependent CP asymmetries

- Neutral mesons are oscillating between their particle (B^0) and anti-particle state ($\bar{B}^0$) state $\rightarrow$ flavour
- Time dependent CP asymmetries that are analyzed by LHCb:

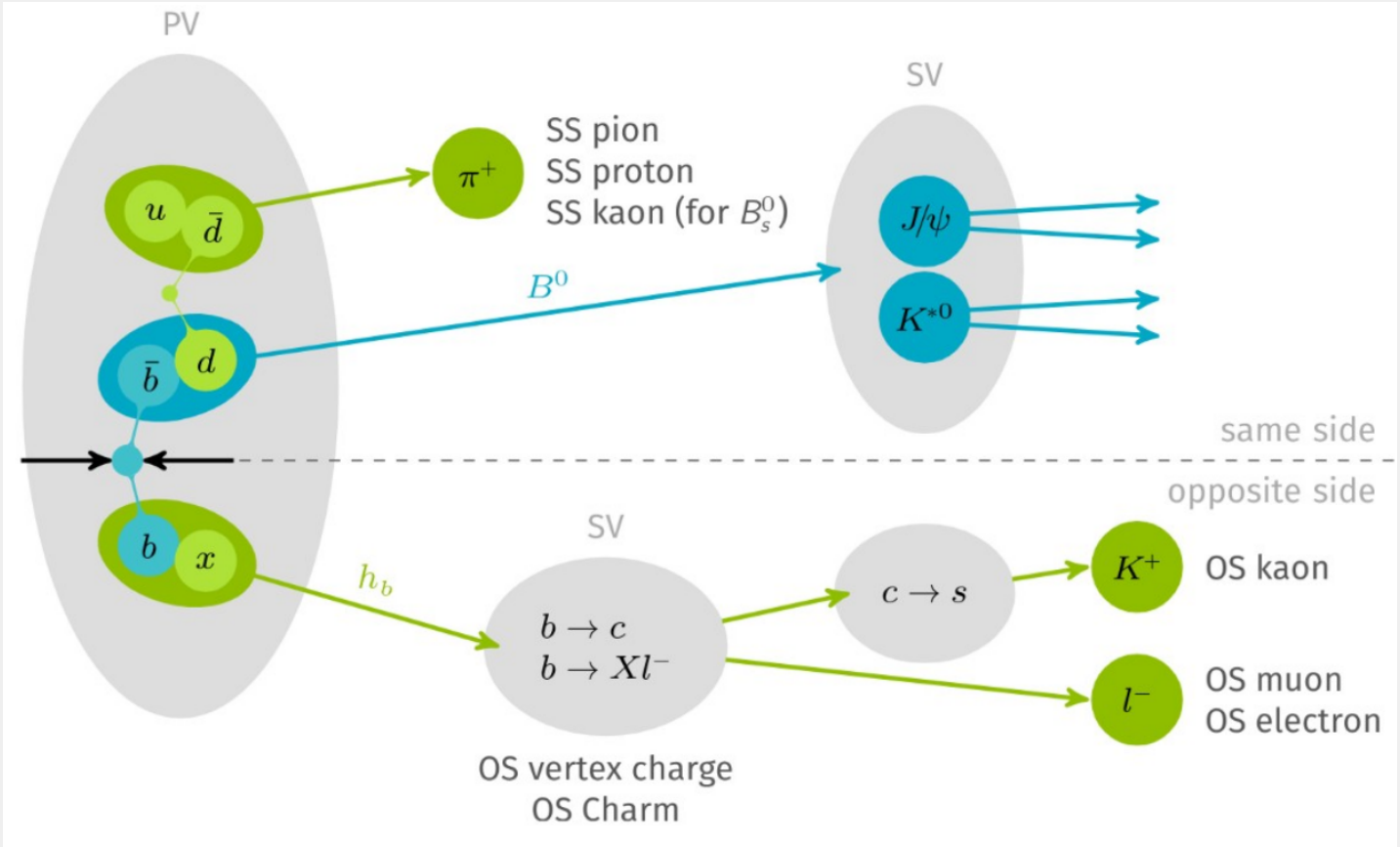
$$A_{CP}(t) = \frac{N(\bar{B}^0 \rightarrow f)(t) - N(B^0 \rightarrow f)(t)}{N(\bar{B}^0 \rightarrow f)(t) + N(B^0 \rightarrow f)(t)}$$

- Requires knowledge of B flavour at the time of production
- Production flavour different from decay flavour

$\rightarrow$ **Flavour Tagging algorithms**

Example event that uses flavour tagging

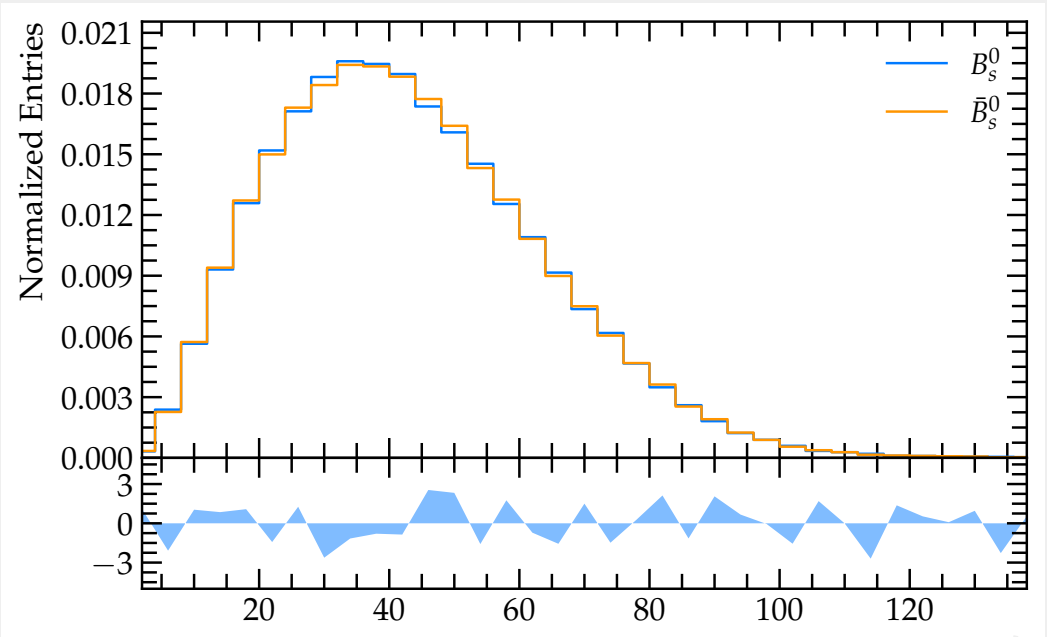
- Flavour Tagging uses charge correlations of particles that are connected to the signal B^0 meson
- Use that quarks are produced in $q - \bar{q}$ pairs



Used data sets

- Different processes in the Same Side fragmentation for B^0 and B_s^0
- 2 different Inclusive Flavour Taggers (IFT)
  - B_s^0 model trained on $B_s^0 \to J/\psi\phi$, calibrated on $B_s^0 \to D_s^- \pi^+$
  - B^0 model trained and calibrated on $B^0 \to J/\psi K^{*0}$
- For each model:
  - Simulation and Data 2016-2018
  - Creating results for each year individually

- Train events:
  - Approx. 2 million events per mode
  - Each event contains around 40 tracks on average



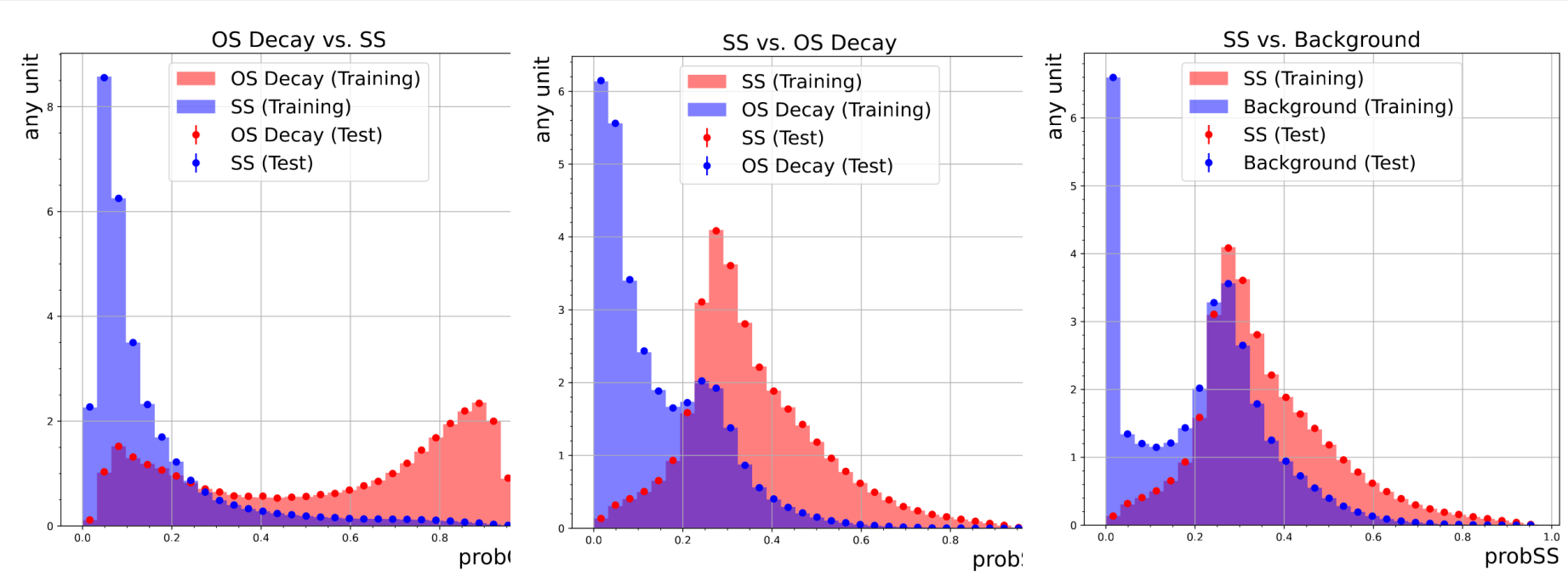
First Step: Track Classification

- Categorize each track with a BDT before training the IFT NN
  - $\rightarrow$ Reduces complexicity of the problem $\rightarrow$ gain performance

- 4 track categories
  - SS Fragmentation
  - OS Decay
  - OS Fragmentation
  - Background
- Each year and mode separately
- Tracks weighted by category during training due to imbalanced amount of tracks in each category

input features
PID(e, μ, K, p)
ProbNN($e, \mu, k, p, \pi, ghost$)
p, p_T
Charge, VeloCharge
BPVIP, BPVIPCHI2
SumBDT_ult
IP_trMother, TRCHI2DOF
diff_z
diff_eta
cos_diff_phi
Pproj

First Step: Track Classification

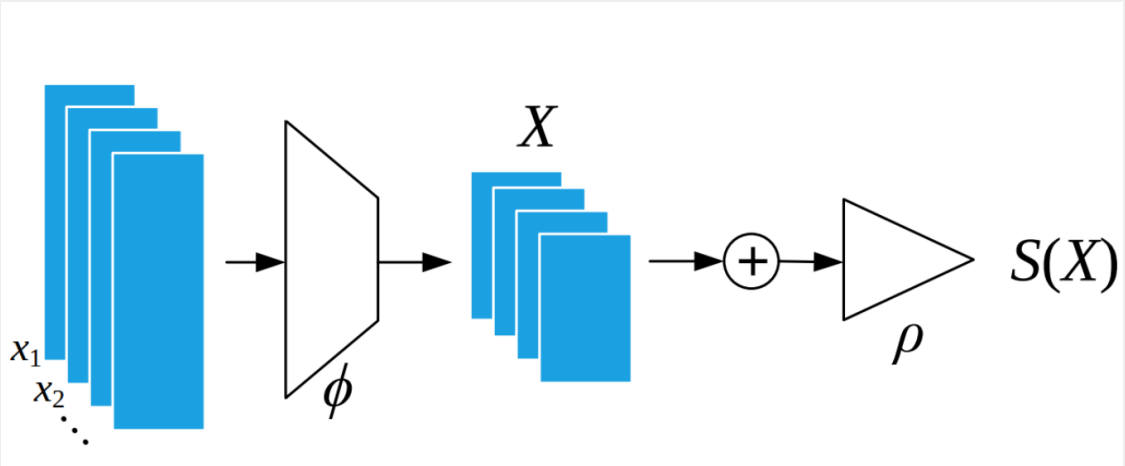


Few examples of BDT outputs
Distributions similar for all years, modes and categories

DeepSet Neural Network

- New approach to tag *B* flavour
- Neural Network that uses features of all tracks (inclusive) of an event (except signal)

- Combination of 2 NN:
  - ϕ: combines n tracks to matrix *X*
  - ρ: tagging decision based on *X*



- Arbitrary number of tracks
  - No cutting/ filling of tracks
  - No need to order tracks

DeepSet Training

- Training of Neural Network weights and biases
- Minimizing binary cross-entropy loss function via backpropagation

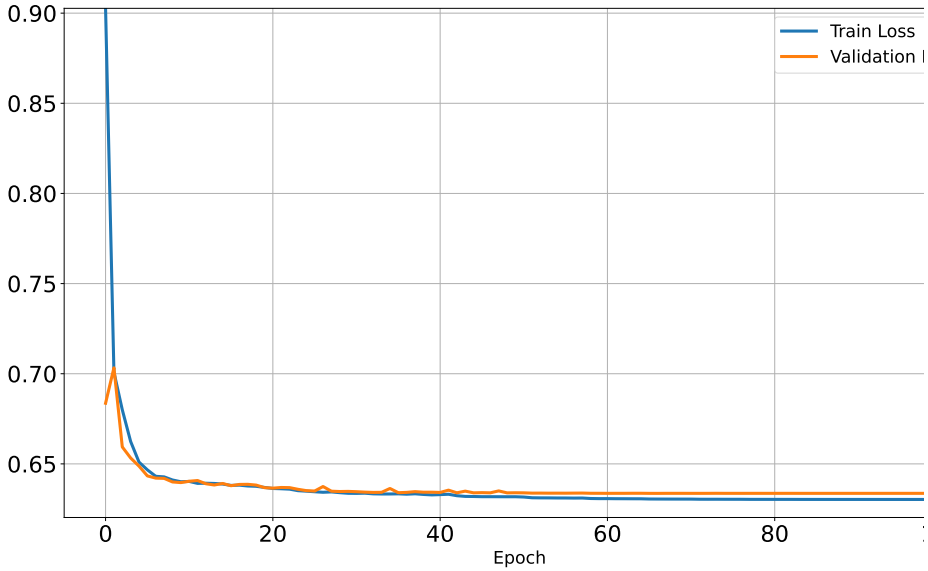
$$L = -\frac{1}{N}\sum_{i=1}^{N} y_{i,true} \cdot log(y_i) + (1 - y_{i,true}) \cdot log(1 - y_i)$$

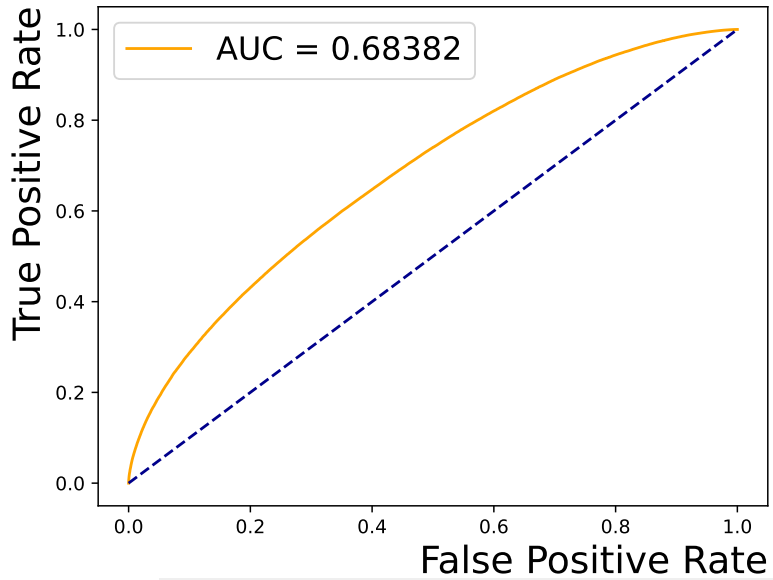
- Trained on all years (2016-2018) combined
- Same features as track classification BDT plus
  - 4 BDT outputs
  - 8 Classical Taggers (previous flavour tagging algorithms)

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

DeepSet Training Results

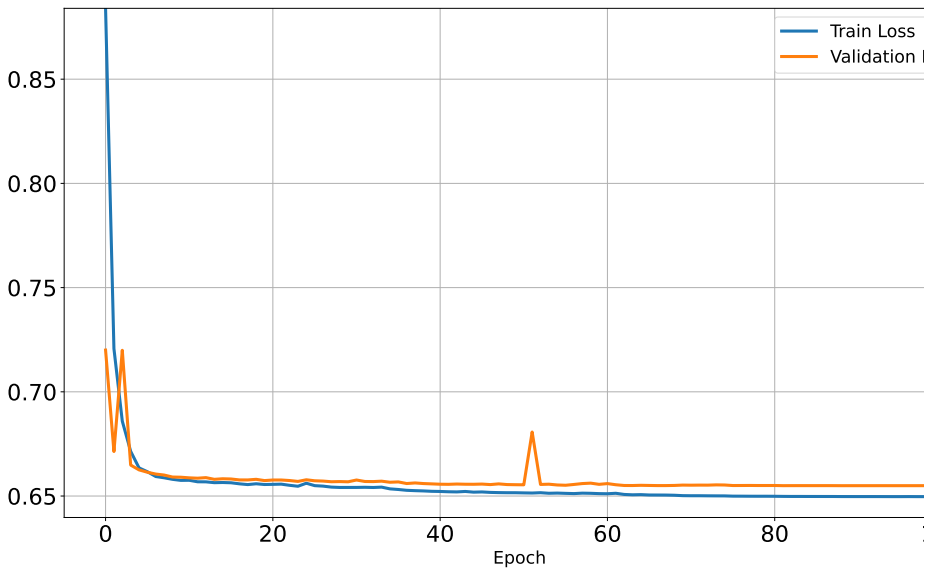
$$B_s^0 \to J/\psi\phi$$

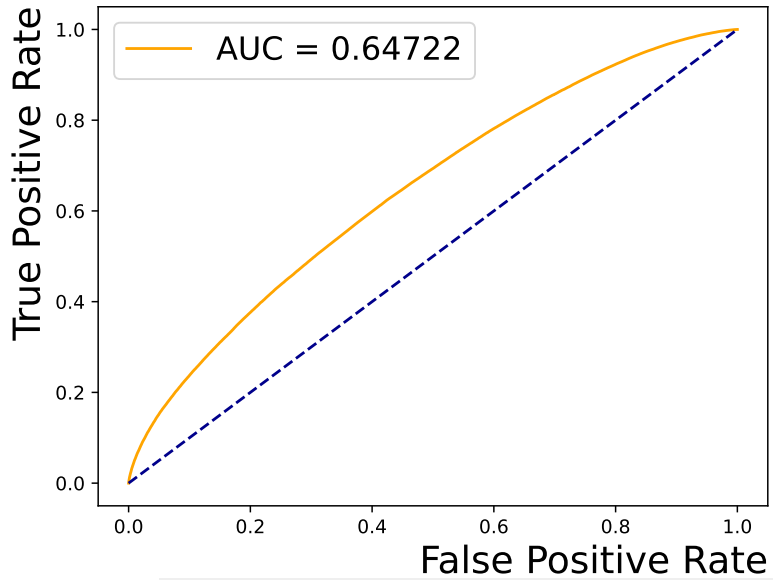




DeepSet Training Results

$$B^0 \to J/\psi K^{*0}$$





Mistag Calibration

- Calibrate mistag estimation η to a true mistag probability $\omega(\eta)$ assuming linear dependence
- Calibration performed on Data, use decay time of B meson to obtain production flavour
  - Only possible due to flavour specific decay modes

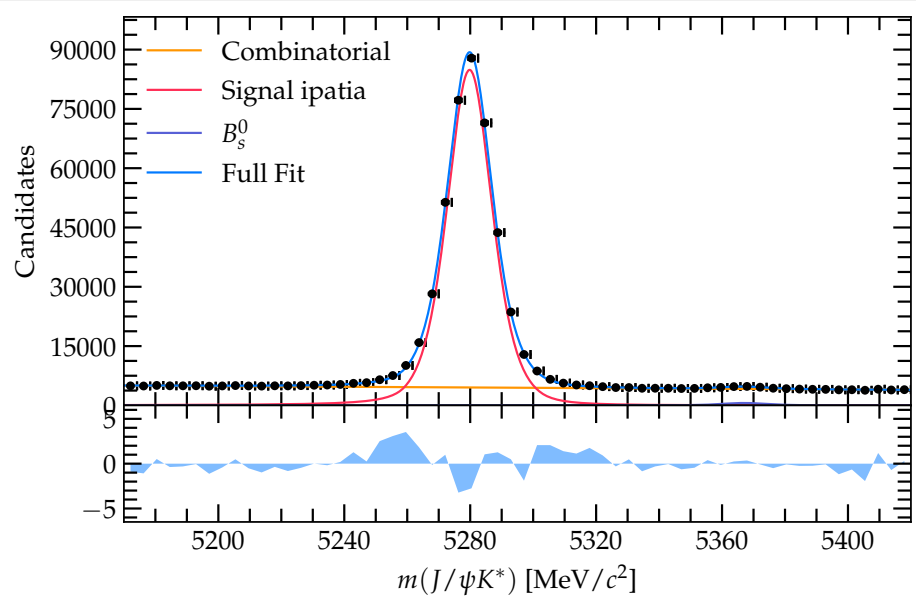
$$P(t|\sigma_t) = \epsilon(t)\left[\Gamma(t') * G(t - t'|\sigma_t)\right]$$

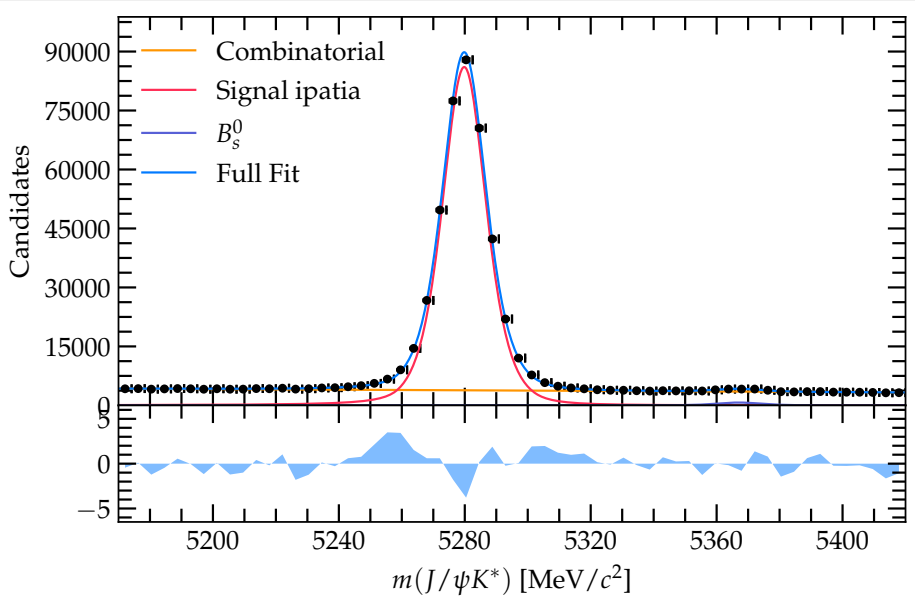
$$e^{-\Gamma t}\left[(1 - q^{mix}\Delta\omega(\eta))\cosh(\tfrac{1}{2}\Delta\Gamma t) + q^{mix}(1 - 2\omega(\eta))\cos(\Delta m t)\right]$$

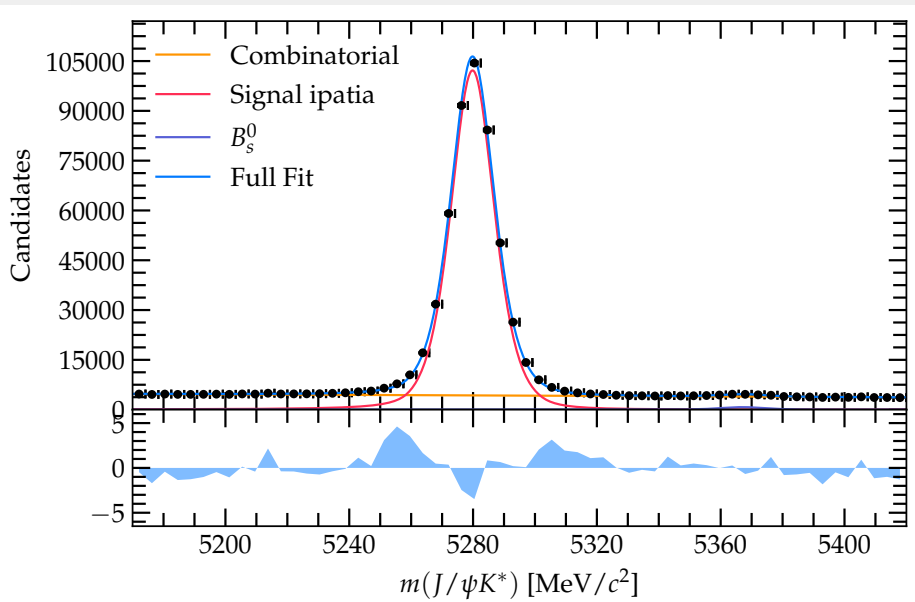
where $\Delta\Gamma_{s,d}$, $\Gamma_{s,d}$ and $m_{s,d}$ are gaussian constrained

- sWeights for statistical background reduction

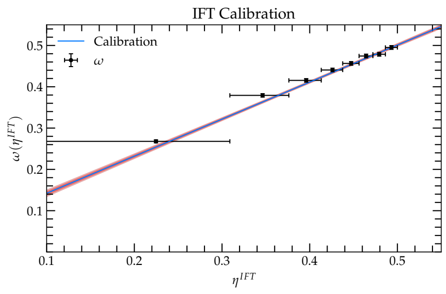
$B^0 \to J/\psi K^{*0}$ MASS FIT

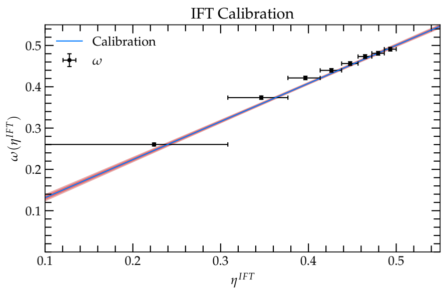


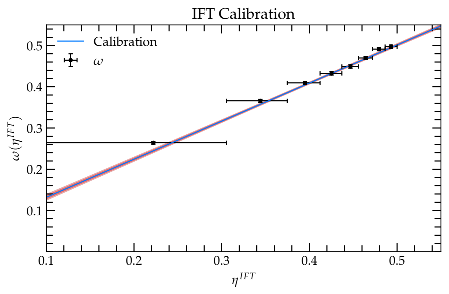




$B^0 \to J/\psi K^{*0}$ IFT calibration







Tagging Performance

- Evaluate performance by using tagging power (statistical degradation of CP asymmetries)

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

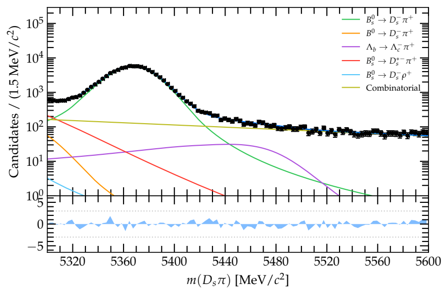
$$\sigma^{\text{stat}}(\text{CP asym}) \propto \frac{1}{\sqrt{\epsilon_{\text{eff}} N}}$$

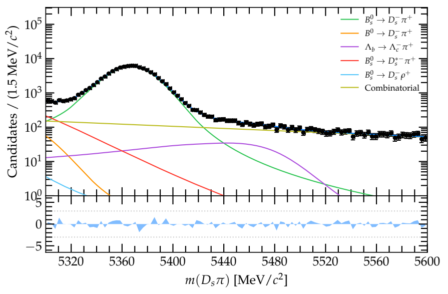
with $D = (1-2\omega)$ as the dilution and $\epsilon_{\text{tag}} = \frac{N_{\text{wrong}} + N_{\text{right}}}{N_{\text{wrong}} + N_{\text{right}} + N_{\text{untag}}}$ as the tagging efficiency

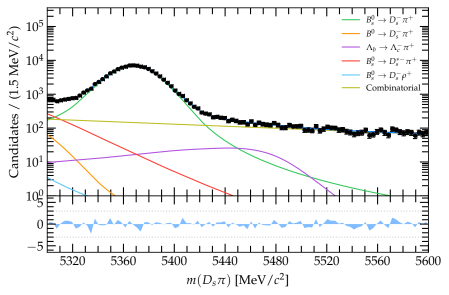
Tagging power $B^0 \to J/\psi K^{*0}$

- No exact numbers since still in review
- Classic tagger combination $\approx 4\% - 4.3\%$ of tagging power
- Average relative improvement of IFT of $\approx 35\%$ with respect to classical tagger combination

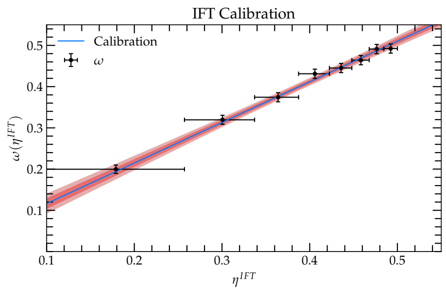
$B_s^0 \to D_s^- \pi^+$ MASS FIT

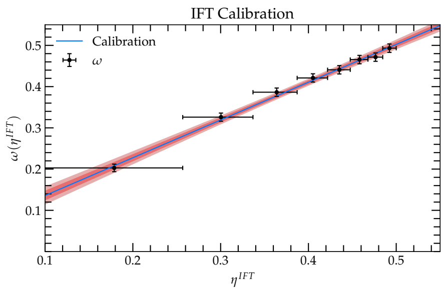


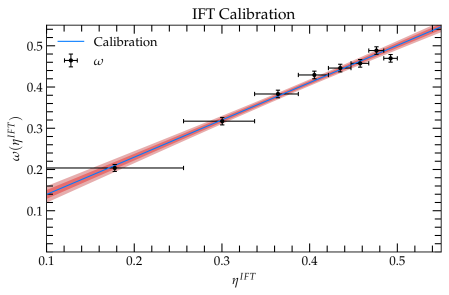




$B_s^0 \rightarrow D_s^- \pi^+$ IFT calibration







Tagging power $B_s^0 \to D_s^- \pi^+$

- No exact numbers since still in review
- For comparison:
  - Classic tagger combination $\approx 6\% - 6.5\%$ of tagging power
  - Average relative improvement of IFT of $\approx 20\%$ with respect to classical tagger combination
- Same improvement can be seen in tagging power on $B_s^0 \to J/\psi\phi$ and stat. uncertainty of *CP*-parameter ϕ_s

Inclusive Flavour Tagging in Run3

- IFT still in development, however:
  - Luminosity increased drastically by increasing Pile-Up
  - Run2 shows strong nPV dependency if tagging power
  - $\rightarrow$ Tagging power most likely much lower than in Run2
- LHCb Data 2024 approx as much as Run2 combined

$$\sigma^{stat}(\text{CP asym}) \propto \frac{1}{\sqrt{\epsilon_{eff} N}}$$

Hard to tell improvement in *CP* parameters since N increased but ϵ_{eff} will decrease, optimizations will take some time

Conclusion & what to expect

Conclusion

- Developing a Neural Network based Inclusive Flavour Tagger
- Relative improvement of 20% (35%) on $B_s^0(B^0)$ with respect to previous Flavour Tagger combination method (classical taggers)
- The IFT includes track categorization BDT and information from classical taggers
  - $\rightarrow$ Both improve the IFT by a bit ($\approx$ 5% at maximum), tough not mandatory to see improvement of IFT w.r.t. classical taggers

Expectation in Run 3:

- Tagging Power will decrease due to increased number of nPV and nTracks per event
- Hard to tell how IFT and classical taggers will perform

Thank You

Tagger Combination

- Compare IFT to classical taggers

$\rightarrow$ Classical taggers need to be combined

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

$$p(b) = \prod_i \left(\frac{1 + d_i}{2} - d_i(1 - \eta_i) \right)$$

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

$P(b)(P(\bar{b}))$: probability of signal candidate containing a $b(\bar{b})$ quark
d_i: tagging decisions
η_i: mistag estimations

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

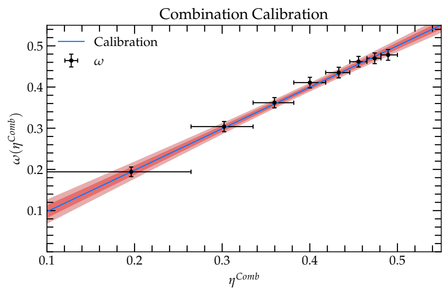
Selection and mass fits: $B_s^0 \to D_s^- \pi^+$

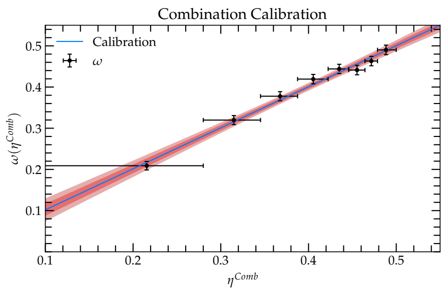
- Selection: trigger and rectangular cuts
- Calibration of decay-time resolution with prompt D_s^- decays
- Generate sWeights to subtract backgrounds in following steps

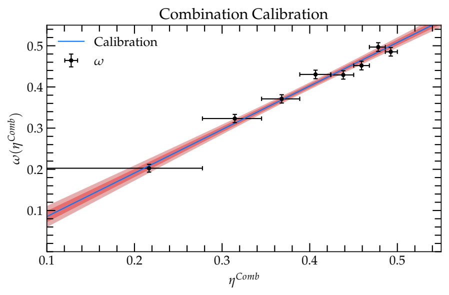
Description	Requirement
HLT1	`B_Hlt1TrackMVADecision_TOS`
	`B_Hlt1TwoTrackMVADecision_TOS`
HLT2	`B_Hlt2Topo2BodyDecision_TOS`
	`B_Hlt2Topo3BodyDecision_TOS`
	`B_Hlt4Topo2BodyDecision_TOS`

Particle	Description	Requirement
B_s^0	χ^2_{IP}	< 16
	t	> 0.2 ps
	$m(D_s^-\pi^+)$	$\in [5000,6000]$ MeV/c^2
	B_{DIRA}	> 0.999
Bachelor π^+	p	> 2 GeV/c
	p_T	> 0.4 GeV/c
	χ^2_{IP}	> 36
	$\chi^2_{track}/nDoF$	< 2
	isMuon	$= 0$
	DLL$(\pi - K)$	< 0
D_s^-	$m(KK\pi)$	$\in [1933,2003]$ MeV/c^2
	χ^2_{IP}	> 4
π^+ from D_s^-	χ^2_{IP}	> 9
	DLL$(\pi - K)$	< 8.0
$K^\pm$ from D_s^-	χ^2_{IP}	> 9
	DLL$(K - \pi)$	> 5.0
D^0 veto	$m(KK)$	< 1840 MeV/c^2
$D^\pm$ veto	DLL$(K - \pi)$	> 10.0
	or $m(K\pi\pi)$	$\notin [1840,1900]$ MeV/c^2
Λ_c veto	DLL$(K - p)$	> 5.0
	or $m(p\pi K)$	$\notin [2255,2315]$ MeV/c^2

$B_s^0 \to D_s^- \pi^+$ COMBINATION CALIBRATION







Selection and mass fits: $B^0 \to J/\psi K^{*0}$

- Selection: trigger and rectangular cuts
- Calibration of decay-time resolution from MC
- Signal shape from MC
- B_s^0 peak shifted by 87.19 MeV

Description	Requirement
HLT1	`B_Hlt1TwoTrackMVADecision_TOS`
	`Jpsi_Hlt1DiMuonHighMassDecision_TOS`
	`Jpsi_Hlt1TrackMuonDecision_TOS`
HLT2	`Jpsi_Hlt2DiMuonDetachedJPsiDecision_TOS`

Particle	Description	Requirement
B^0	t	> 0.15 ps
	$m(J/\psi K^{*0})$	$\in [5150, 5450]$ MeV/c^2
π^-	p_T	> 250 MeV/c^2
	DLL$(\pi - K)$	< 0
	ProbNN(p)	< 0.9
	ProbNN(K)	< 0.9

$B^0 \to J/\psi K^{*0}$ COMBINATION CALIBRATION