Full Event Interpretation at Belle II

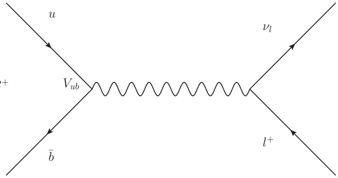
Belle II

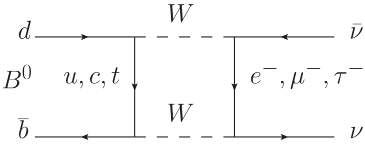
Alexander von Humboldt
Stiftung / Foundation

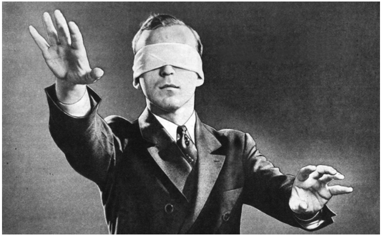
KIT
Karlsruher Institut für Technologie

Why we need Full Event Interpretation

- Interesting physics can be obtained from several challenging modes with missing neutrinos ($B \to D^{(*)}\tau\nu$, $B \to l\nu$, $B \to X_u l\nu$, $B \to h\nu\bar{\nu}$)

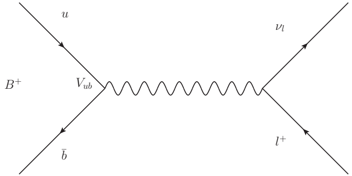


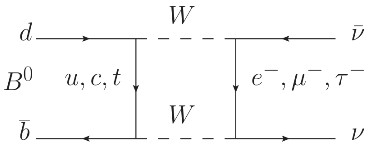




Why we need Full Event Interpretation

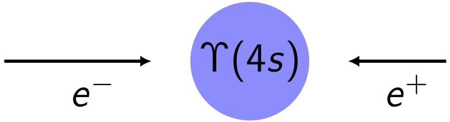
- Interesting physics can be obtained from several challenging modes with missing neutrinos ($B \to D^{(*)}\tau\nu$, $B \to l\nu$, $B \to X_u l\nu$, $B \to h\nu\bar{\nu}$)





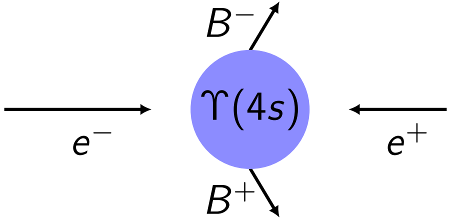
Tag-side B reconstruction

- Collide e^+ and e^- at the energy to make $\Upsilon(4S)$ particles



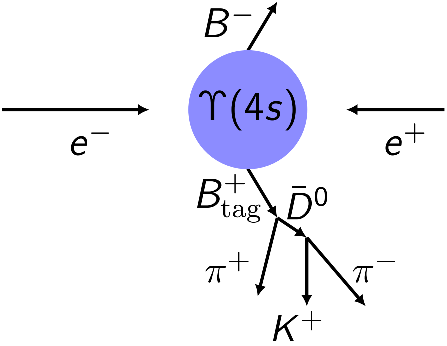
Tag-side B reconstruction

- Collide e^+ and e^- at the energy to make $\Upsilon(4S)$ particles
- $\Upsilon(4S)$ decays to B^+B^- and $B^0\bar{B}^0$ 96% of the time.



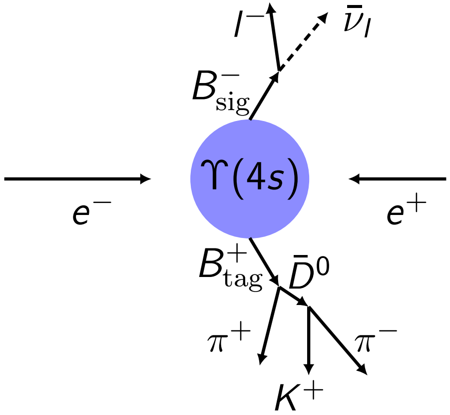
Tag-side B reconstruction

- Collide e^+ and e^- at the energy to make $\Upsilon(4S)$ particles
- $\Upsilon(4S)$ decays to B^+B^- and $B^0\bar{B}^0$ 96% of the time.
- Reconstruct one B meson as tag-side (B_{tag}) hadronic or SL



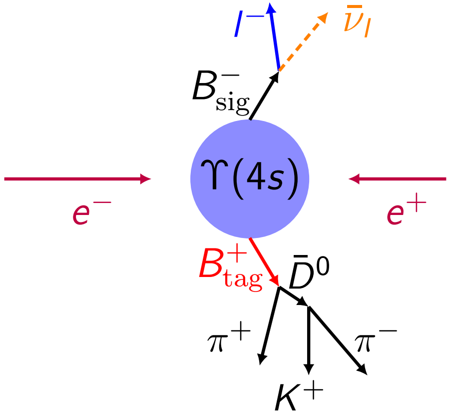
Tag-side B reconstruction

- Collide e^+ and e^- at the energy to make $\Upsilon(4S)$ particles
- $\Upsilon(4S)$ decays to B^+B^- and $B^0\bar{B}^0$ 96% of the time.
- Reconstruct one B meson as tag-side (B_{tag}) hadronic or SL
- Study remaining B meson as signal (B_{sig})



Tag-side B reconstruction

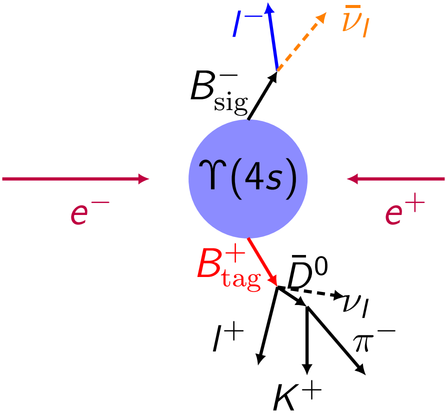
- Collide e^+ and e^- at the energy to make $\Upsilon(4S)$ particles
- $\Upsilon(4S)$ decays to B^+B^- and $B^0\bar{B}^0$ 96% of the time.
- Reconstruct one B meson as tag-side (B_{tag}) hadronic or SL
- Study remaining B meson as signal (B_{sig})
- Flavour constraints:
  $B^+_{\mathrm{tag}} \implies B^-_{\mathrm{sig}}$
  Kinematic constraints:
  $p_\nu = p_{e^+e^-} - p_{\mu^-} - p_{B^+}$



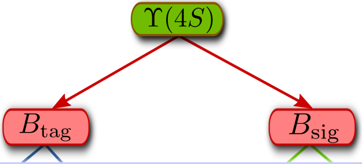
Tag-side B reconstruction

- Collide e^+ and e^- at the energy to make $\Upsilon(4S)$ particles
- $\Upsilon(4S)$ decays to B^+B^- and $B^0\bar{B}^0$ 96% of the time.
- Reconstruct one B meson as tag-side (B_{tag}) hadronic or SL
- Study remaining B meson as signal (B_{sig})
- Flavour constraints:
  $B^+_{\mathrm{tag}} \implies B^-_{\mathrm{sig}}$
  Kinematic constraints:

  $p_\nu = p_{e^+e^-} - p_{\mu^-} - p_{B^+}$



Which tag-side reconstruction?



Tagging techniques

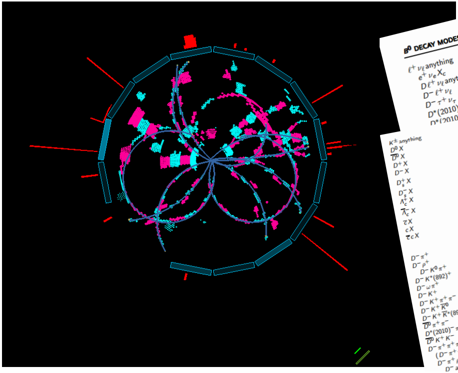
Purity →

← Efficiency

Inclusive	Semileptonic	Hadronic
$B \to \text{anything}$	$B \to D^{(*)}\ell\nu_\ell$	$B \to \text{hadrons}$
$\epsilon \approx \mathcal{O}(100\%)$	$\epsilon \approx \mathcal{O}(1\%)$	$\epsilon \approx \mathcal{O}(0.1\%)$
Very large statistics; Also very large background	Mid-range reconstruction efficiency; Less information about B_{tag} due to neutrino	Cleaner sample Knowledge of $p(B_{\text{sig}})$; Lower tag-side efficiency

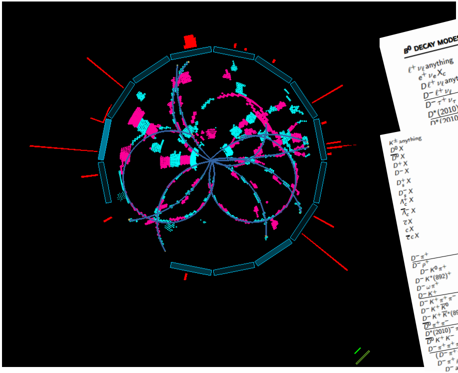
The Task

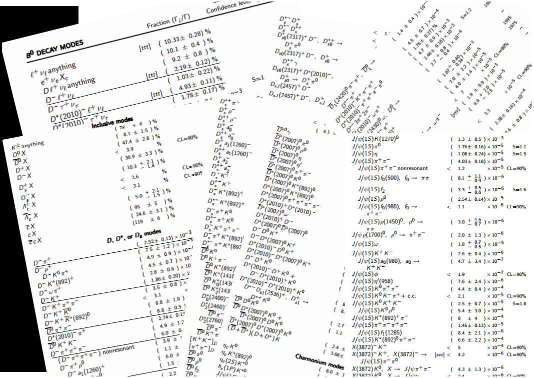
- Must reconstruct a substantial number of modes (e.g O(50) hadronic modes of B mesons)
- Face combinatorics which scale as the factorial of the number of tracks $\implies$ Require selections early on



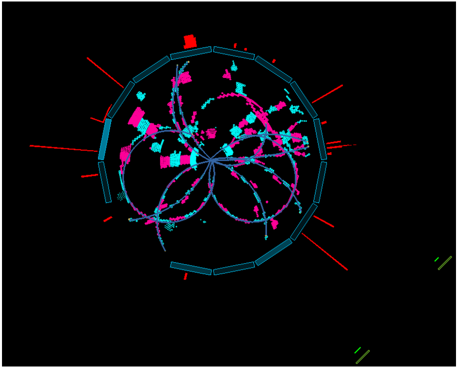
The Task

- Must reconstruct a substantial number of modes (e.g O(50) hadronic modes of B mesons)
- Face combinatorics which scale as the factorial of the number of tracks $\implies$ Require selections early on





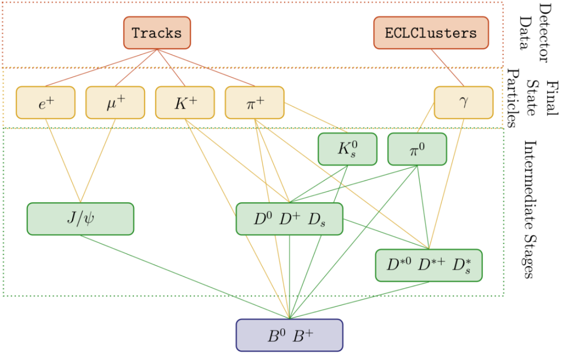
Combinatorics



- $\sim$10 tracks in this event
- Let's assume 5 positively charged and 5 negatively charged.
- Now lets reconstruct $D^0 \to K^-\pi^+\pi^+\pi^-$
- $\binom{5}{2}^2 = 100$ possible combinations
- Reconstructing $B^+ \to (D^0 \to K^-\pi^+\pi^+\pi^-)\pi^+$ introduces $\binom{3}{1} \times 100 = 300$ combinations.

The Full Event Interpretation

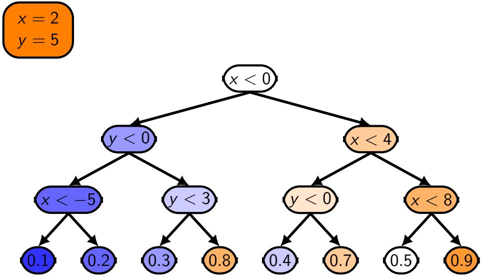
- Utilises O(200) decay channels with classifiers (BDTs) trained for each.
- Reconstructs O(10000) unique decays chains in six stages.

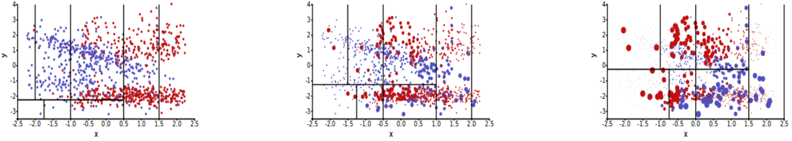


arxiv1807.08680, Keck, T. et al.   Keck, T., PhD Thesis

Classifiers

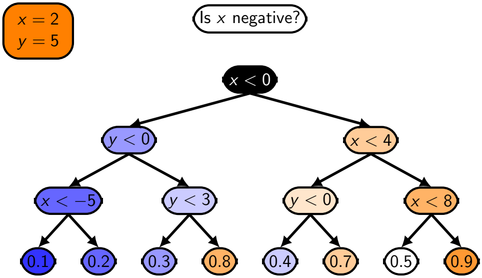
- Describe decay channels with features x_i.
- Labelled training datasets obtained from simulation.
- Train classfiers for each channel, $\mathcal{P}(x_i)$
- The FEI by default uses Boosted Decision Trees.

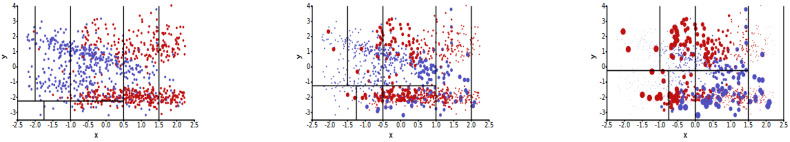




Classifiers

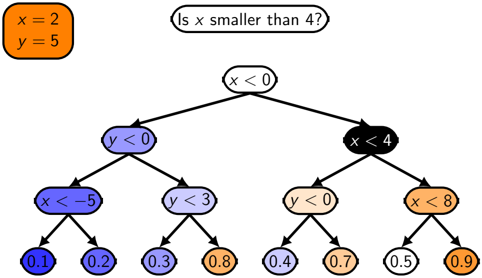
- Describe decay channels with features x_i.
- Labelled training datasets obtained from simulation.
- Train classfiers for each channel, $\mathcal{P}(x_i)$
- The FEI by default uses Boosted Decision Trees.

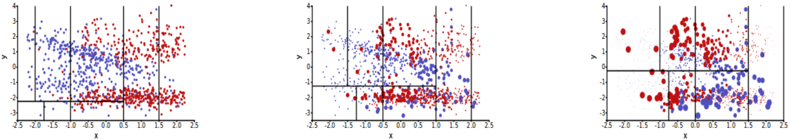




Classifiers

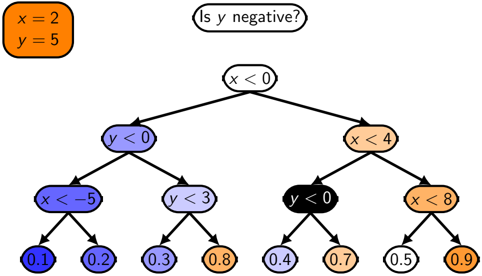
- Describe decay channels with features x_i.
- Labelled training datasets obtained from simulation.
- Train classfiers for each channel, $\mathcal{P}(x_i)$
- The FEI by default uses Boosted Decision Trees.

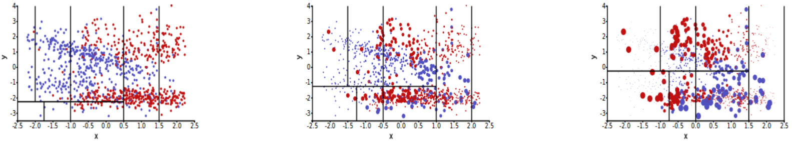




Classifiers

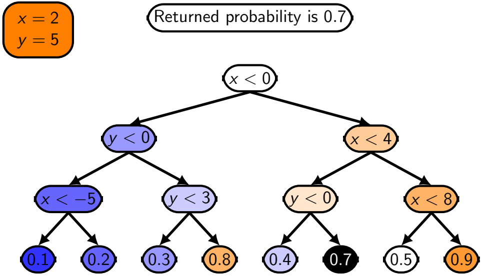
- Describe decay channels with features x_i.
- Labelled training datasets obtained from simulation.
- Train classfiers for each channel, $\mathcal{P}(x_i)$
- The FEI by default uses Boosted Decision Trees.

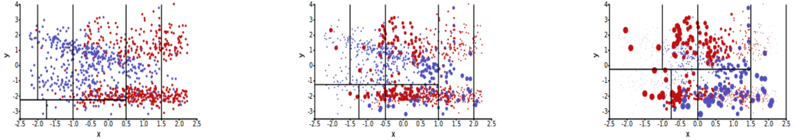




Classifiers

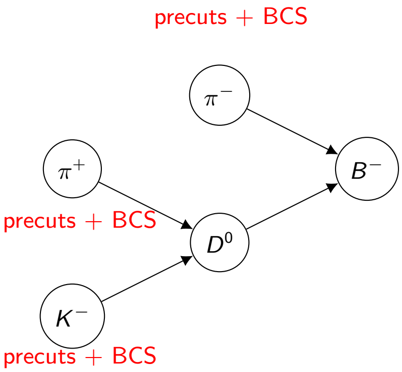
- Describe decay channels with features x_i.
- Labelled training datasets obtained from simulation.
- Train classfiers for each channel, $\mathcal{P}(x_i)$
- The FEI by default uses Boosted Decision Trees.





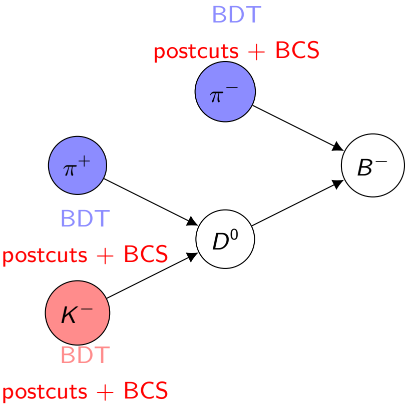
The Algorithm

- Particle candidates assigned from tracks and clusters after a precuts + Best Candidate Selection (BCS).



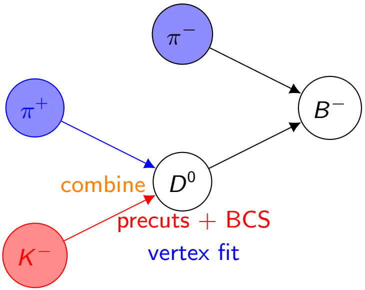
The Algorithm

- Particle candidates assigned from tracks and clusters after a precuts + Best Candidate Selection (BCS).
- For each particle a pre-trained BDT is applied and post cuts + BCS are made.



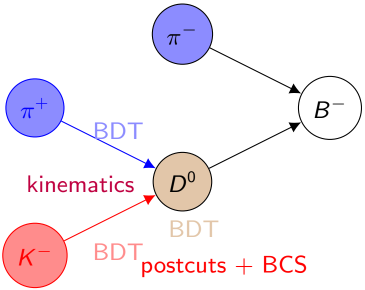
The Algorithm

- Particle candidates assigned from tracks and clusters after a precuts + Best Candidate Selection (BCS).
- For each particle a pre-trained BDT is applied and post cuts + BCS are made.
- Stable particles are combined to reconstruct decays of intermediate particles. After precuts + BCS a vertex fit is performed.



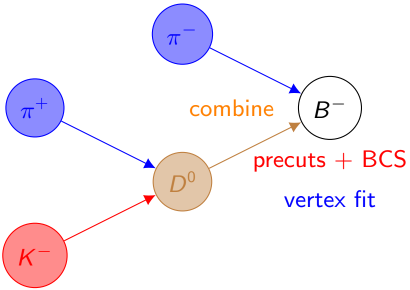
The Algorithm

- Particle candidates assigned from tracks and clusters after a precuts + Best Candidate Selection (BCS).
- For each particle a pre-trained BDT is applied and post cuts + BCS are made.
- Stable particles are combined to reconstruct decays of intermediate particles. After precuts + BCS a vertex fit is performed.
- Intermediate classifiers use daughter kinematics and classifiers.



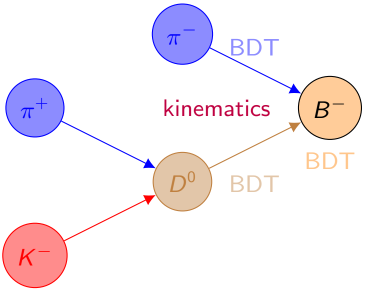
The Algorithm

- Particle candidates assigned from tracks and clusters after a precuts + Best Candidate Selection (BCS).
- For each particle a pre-trained BDT is applied and post cuts + BCS are made.
- Stable particles are combined to reconstruct decays of intermediate particles. After precuts + BCS a vertex fit is performed.
- Intermediate classifiers use daughter kinematics and classifiers.
- Intermediates and stable particles are combined into a B candidate.

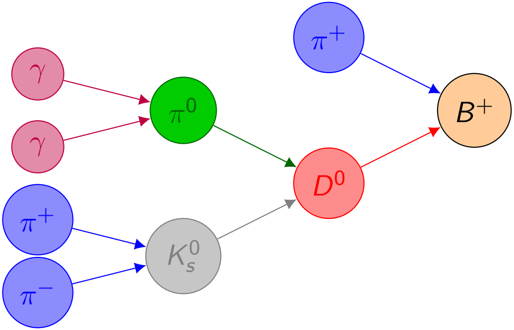


The Algorithm

- Particle candidates assigned from tracks and clusters after a precuts + Best Candidate Selection (BCS).
- For each particle a pre-trained BDT is applied and post cuts + BCS are made.
- Stable particles are combined to reconstruct decays of intermediate particles. After precuts + BCS a vertex fit is performed.
- Intermediate classifiers use daughter kinematics and classifiers.
- Intermediates and stable particles are combined into a B candidate.
- B classifier takes daughter classifiers and kinematics as inputs.

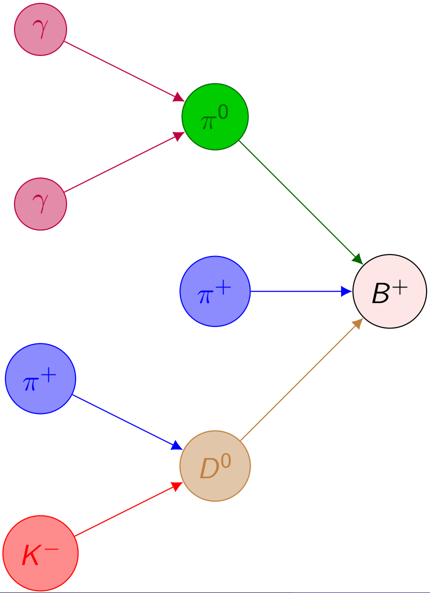


The Algorithm



- Same $B^+ \to D^0\pi^+$ classifier.
- Different decay chain as $D^0 \to K_s^0\pi^0$.
- $D^0 \to K_s^0\pi^0$ has its own classifier.

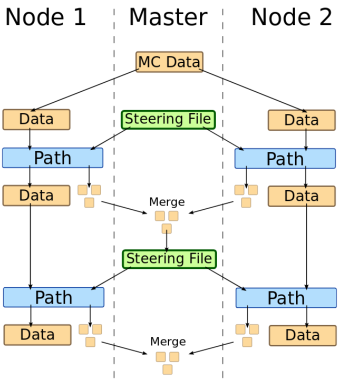
The Algorithm



- Different $B^+ \to D^0\pi^+\pi^0$ decay with its own classifier.
- Original D decay chain as $D^0 \to K^-\pi^+$.

Training the FEI

- Both training and application phases can be distributed via a map reduce approach.
- For training:
  - $\mathcal{O}(100\text{M})$ simulated $\Upsilon(4S) \to B\bar{B}$ events
  - Monte carlo is partitioned and processed at different nodes.
  - At each of the reconstruction phases training data is generated.
  - Training data of each stage is subsquently merged and classifiers trained.

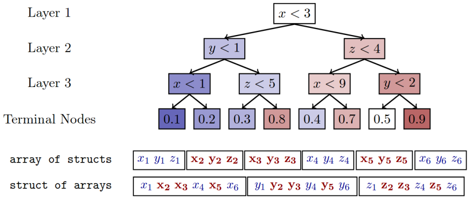


Need for speed

- Utilise FastBDT:
  - Computes cumulative probability histograms (CPH) of nodes in the same level simultaneously.
  - Stores data as an array of structs.

  arxiv1609.06119, Keck, T.

- Utilise FastFit (GitHub link):
  - Uses eigen libraries to gain from vectorisation.
  - Overall factor of 2.7 speed up in the FEI

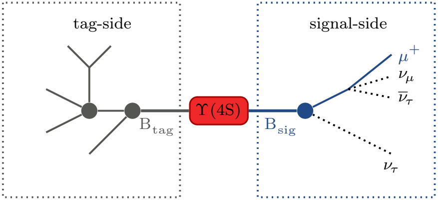


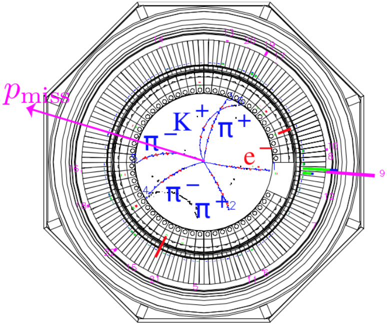
Task	Training	Application
read/write DataStore	30	0
vertex fitting	26	38
particle combination	19	27
classifier inference	11	15
training data & monitoring	6	0
best candidate selection	3	6
other	5	14

- In application 38% of the time is spent on vertex fitting, 27% on particle combination and 15% on classifier inference.

Specific vs Generic FEI

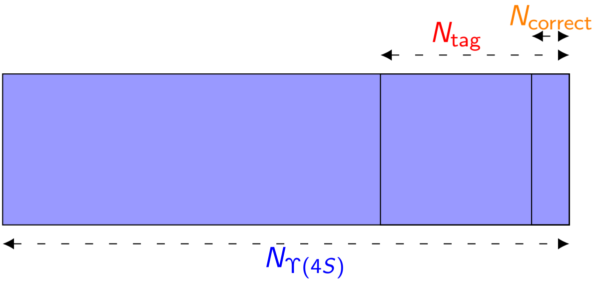
- **Generic FEI** - Reconstruct signal after reconstructing a tag-side B candidate.
- **Specific FEI** - Reconstruct a tag-side B candidate after reconstructing signal





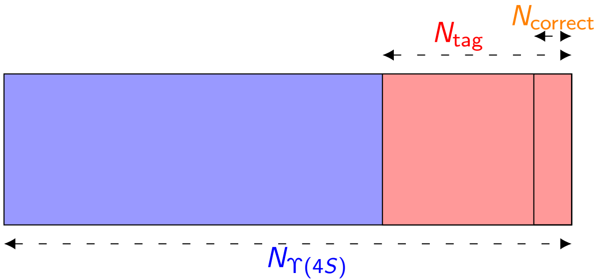
$B^+_{tag} \rightarrow \bar{D}^0\pi^+\pi^+\pi^-$ $\hookrightarrow K^+\pi^-$ $B^-_{sig} \rightarrow \tau^-\bar{\nu}_\tau$ $\hookrightarrow e^-\bar{\nu}_e\nu_\tau$

How does one quantify tagging performance?



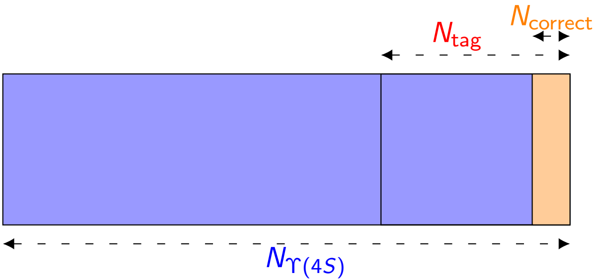
- **tagging efficiency** $= N_{tag} / N_{\Upsilon(4S)}$

How does one quantify tagging performance?



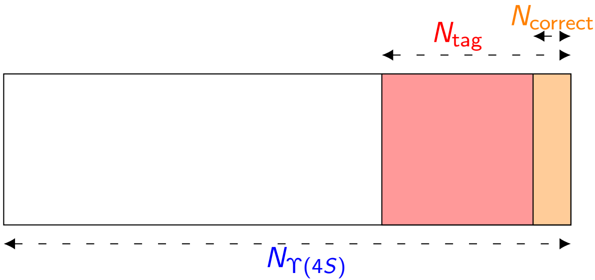
- **tagging efficiency** $= N_{\text{tag}} / N_{\Upsilon(4S)}$

How does one quantify tagging performance?



- **tagging efficiency** $= N_{\text{tag}} / N_{\Upsilon(4S)}$
- **tag-side efficiency** $= N_{\text{correct}} / N_{\Upsilon(4S)}$

How does one quantify tagging performance?



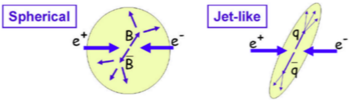
- **tagging efficiency** $= N_{\text{tag}} / N_{\Upsilon(4S)}$
- **tag-side efficiency** $= N_{\text{correct}} / N_{\Upsilon(4S)}$
- **purity** $= N_{\text{correct}} / N_{\text{tag}}$

Tagging performance in Belle data

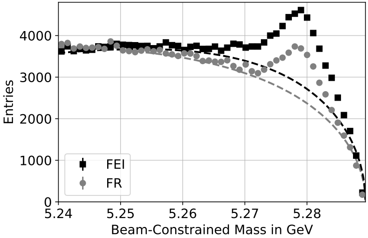
$$m_{bc} = \sqrt{E_B^2 - p_B^2}$$

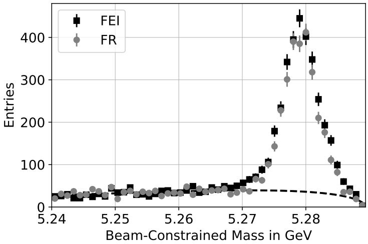
$$E_B = \sqrt{s}/2$$

Different event topologies



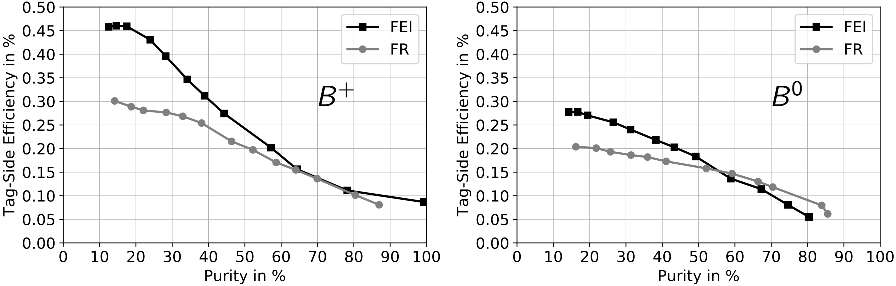
B-mesons   Continuum





Tagging performance

Tag-side efficiency again purity in Belle data



Maximum tag-side efficiency

Tag	FR	SER	FEI Belle MC	FEI Belle II MC
Hadronic B^+	0.28%	0.4%	0.76%	0.66%
Hadronic B^0	0.18%	0.2%	0.46%	0.38%
SL B^+	0.31%	0.3%	1.80%	1.45%
SL B^0	0.34%	0.6%	2.04%	1.94%

FR = Full Reconstruction (Belle Algorithm), SER = Semi-Exclusive Reconstuction (BaBar Algorithm)

Conclusion

- The Full Event Interpretation (FEI) is an algorithm for tag-side B reconstruction at Belle 2.
- It trains O(200) decay channel classifiers which are used in the reconstruction of O(10000) decay chains.
- The FEI outperforms its predecessors with a higher tag-side efficiency.
- The FEI is an essential to the Belle II physics program and resolving the B physics anomalies.

References

The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment - Thomas Keck et al. https://arxiv.org/abs/1807.08680

Machine learning algorithms for the Belle II experiment and their validation on Belle data - Thomas Keck https://publikationen.bibliothek.kit.edu/1000078149

Analysis Software and Full Event Interpretation for the Belle II Experiment - Christian Pulvermacher http://ekp-invenio.physik.uni-karlsruhe.de/record/48741

FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification - Thomas Keck https://arxiv.org/abs/1609.06119

Tutorial

- Log in to kekcc with port forwarding:
  ssh username@login.cc.kek.jp -L 8300:localhost:8300
- Note that you should choose a different port.
- Clone repository at https://stash.desy.de/users/sutclw/repos/feitutorial/browse:
  git clone ssh://git@stash.desy.de:7999/~sutclw/feitutorial.git
- Within the tutorial directory run: source setup_basf2_rel3.sh
- Run jupyter note book with the following command:
  jupyter-notebook --port 8300 --no-browser
- If problems see:
  https://confluence.desy.de/display/BI/Running+Jupyter+Notebook+on+KEKCC