# Full Event Interpretation B-Tagging Algorithm at Belle II


Meihong Liu 刘美宏

Jilin University

on behalf of the Belle II collaboration


Quantum Computing and Machine Learning Workshop

August 8, 2024 Plenary session @ Changchun


liumh19@fudan.edu.cn

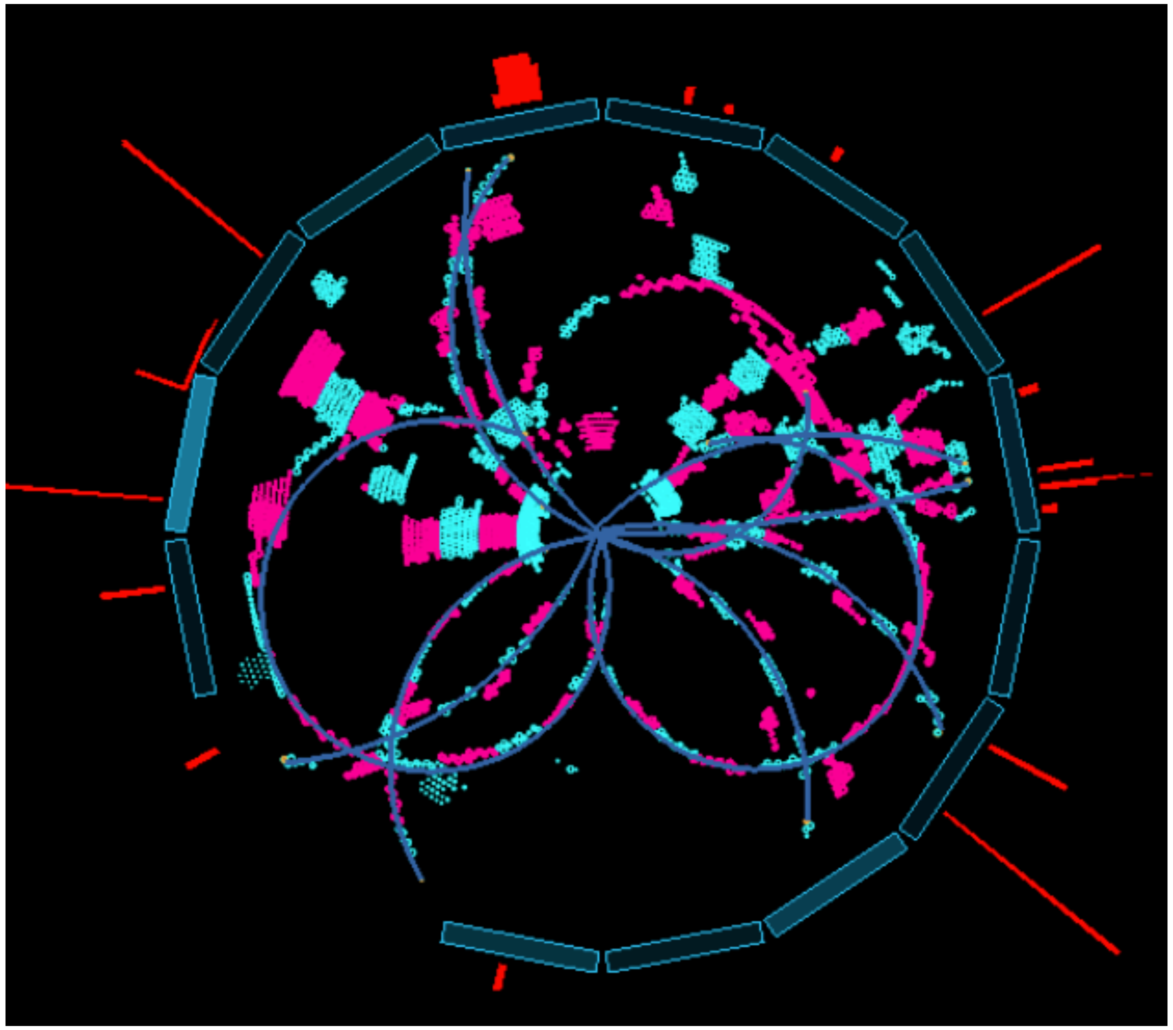
# Outline

- **Why?** Purpose of B-tagging
- **How?** How we do at Belle II
- **What?** Tool: FEI
- Usage of FEI
- Calibration&improvement
- GraFEI

**Main players in B-physics**

Belle (II), BaBar → B-mesons in $e^+e^-$ collisions

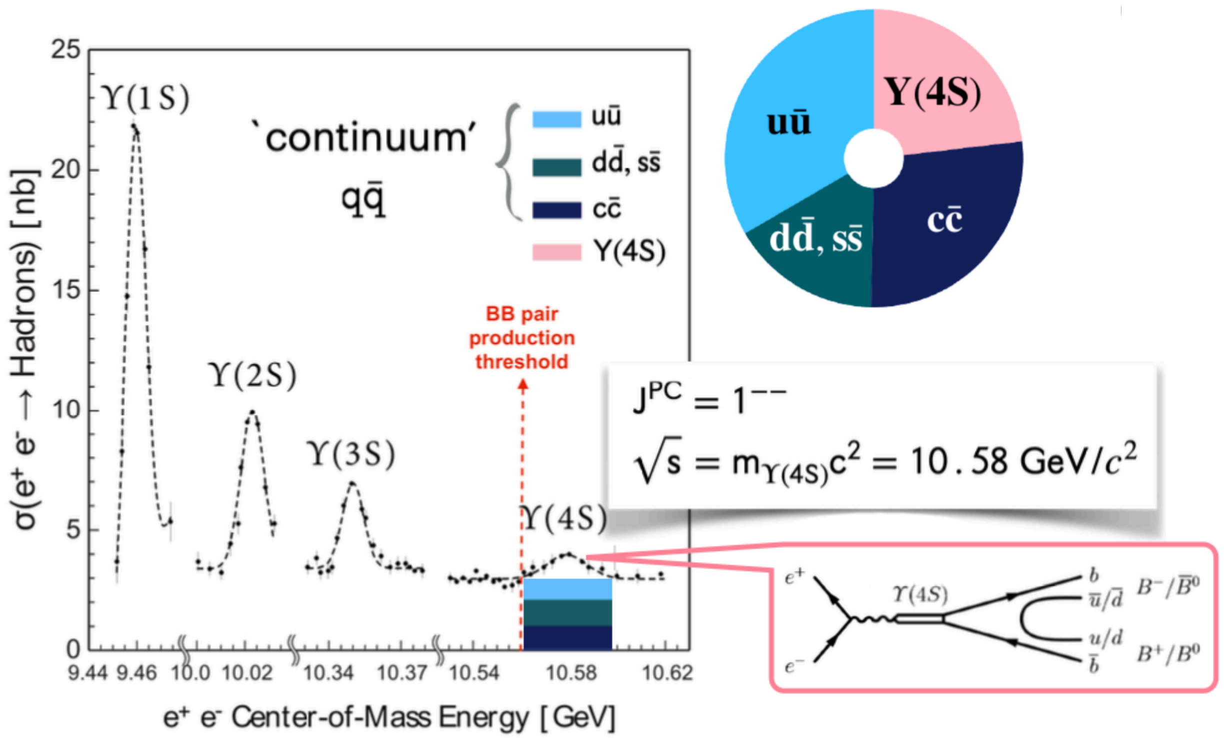
LHCb → *b*-flavored hadrons in *pp* collisions

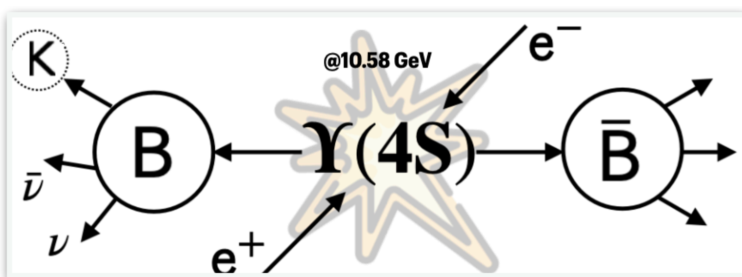


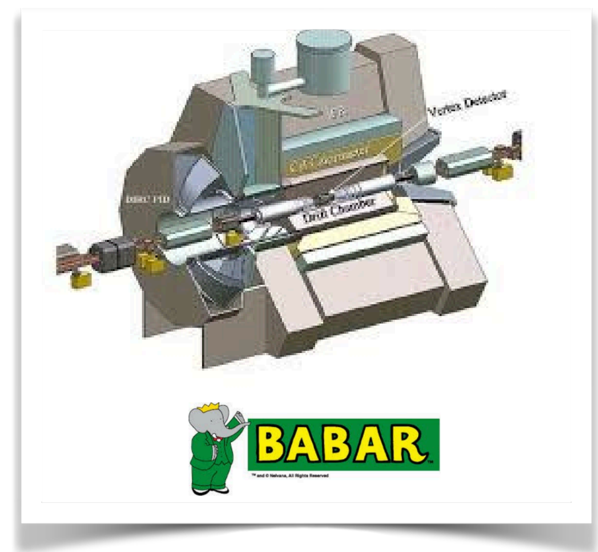
**Average 11 tracks per event in Belle (II)**

# B factory— Collide $e^+e^-$ at $\Upsilon(4S)$?

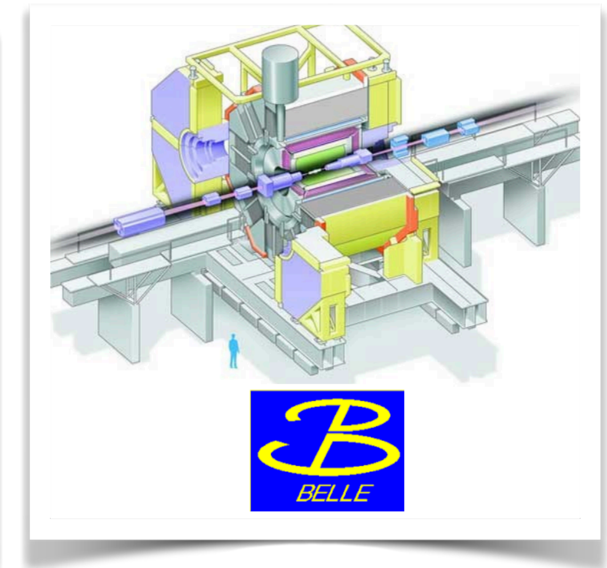
$1^{st}$ generation B factory



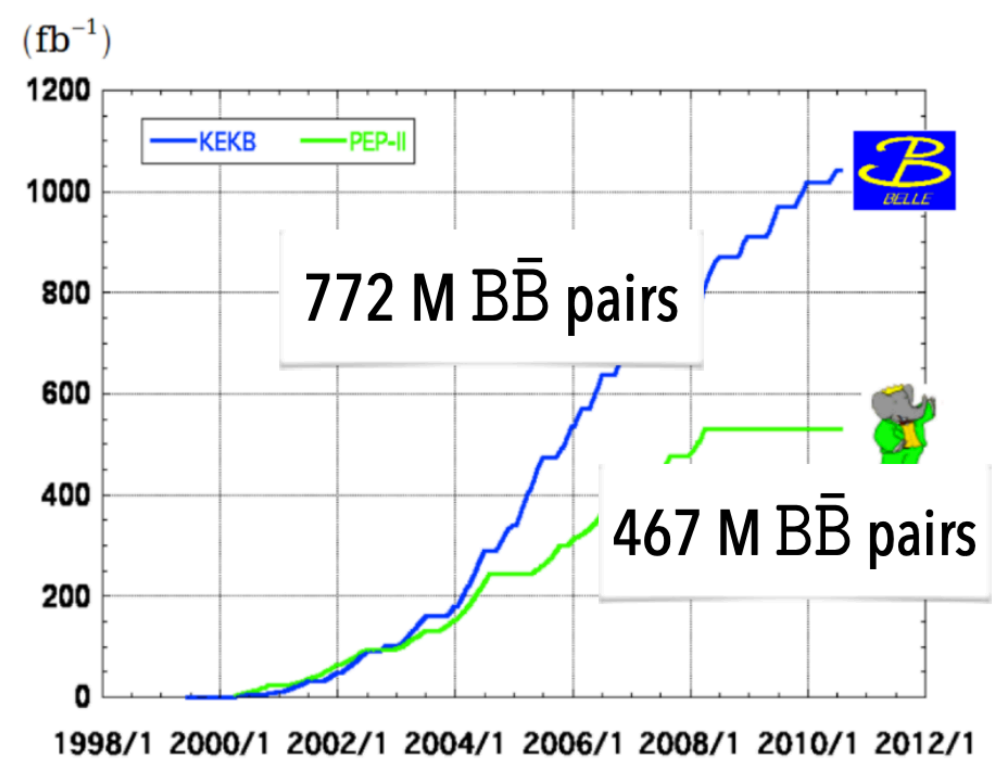




SLAC - PEP-II collider
[1999-2008]




KEK – KEKB collider
[1999-2010]

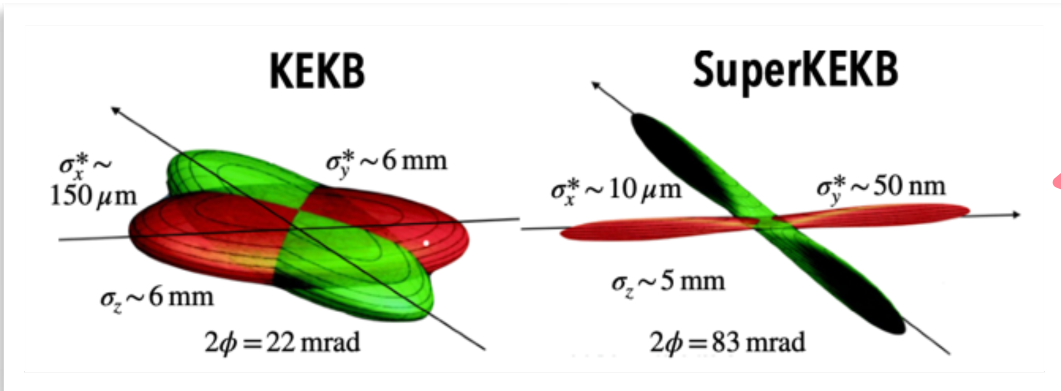


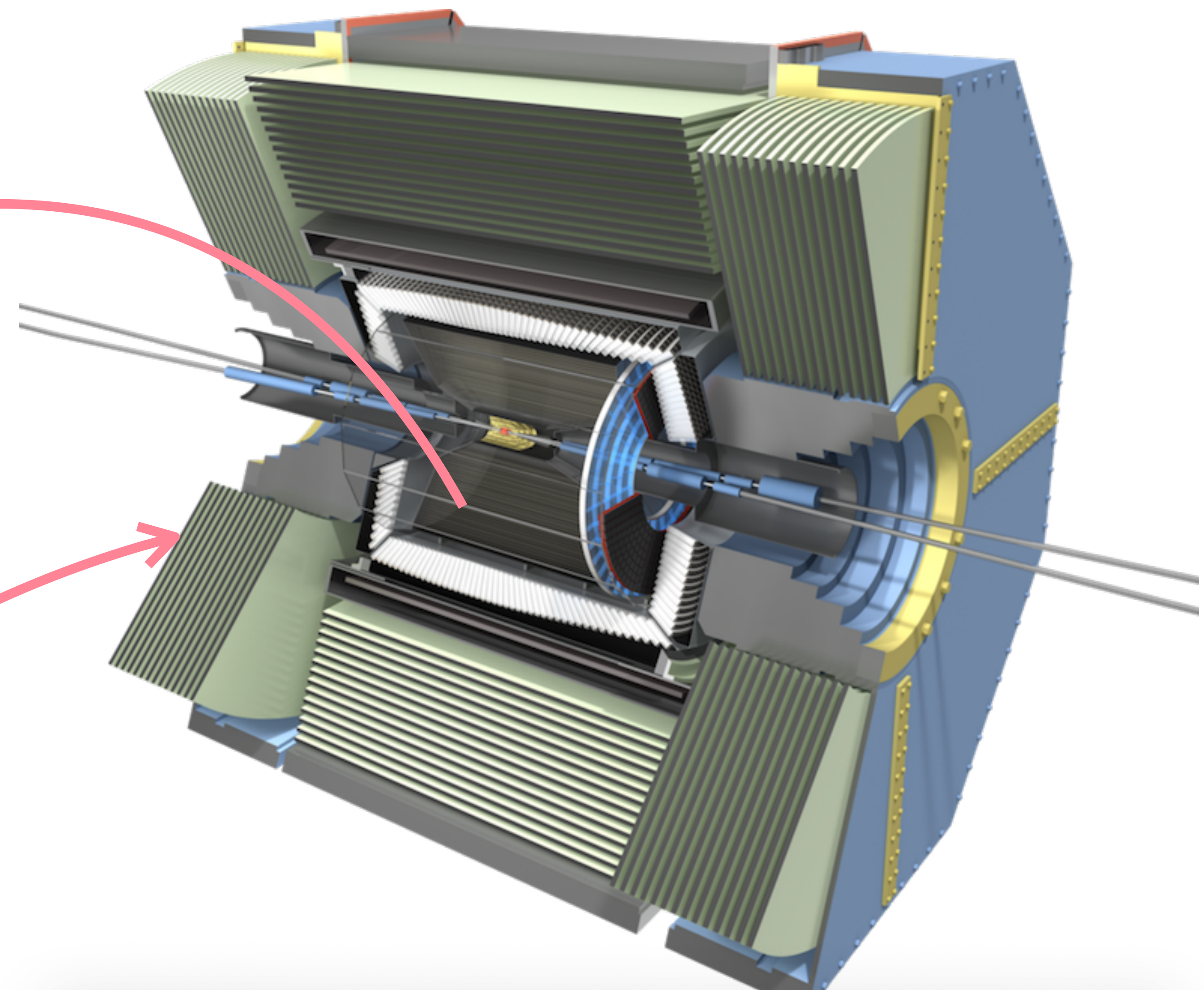

- $\Upsilon(4S)$ – cross section of $B\bar{B}$ one quarter of continuum:

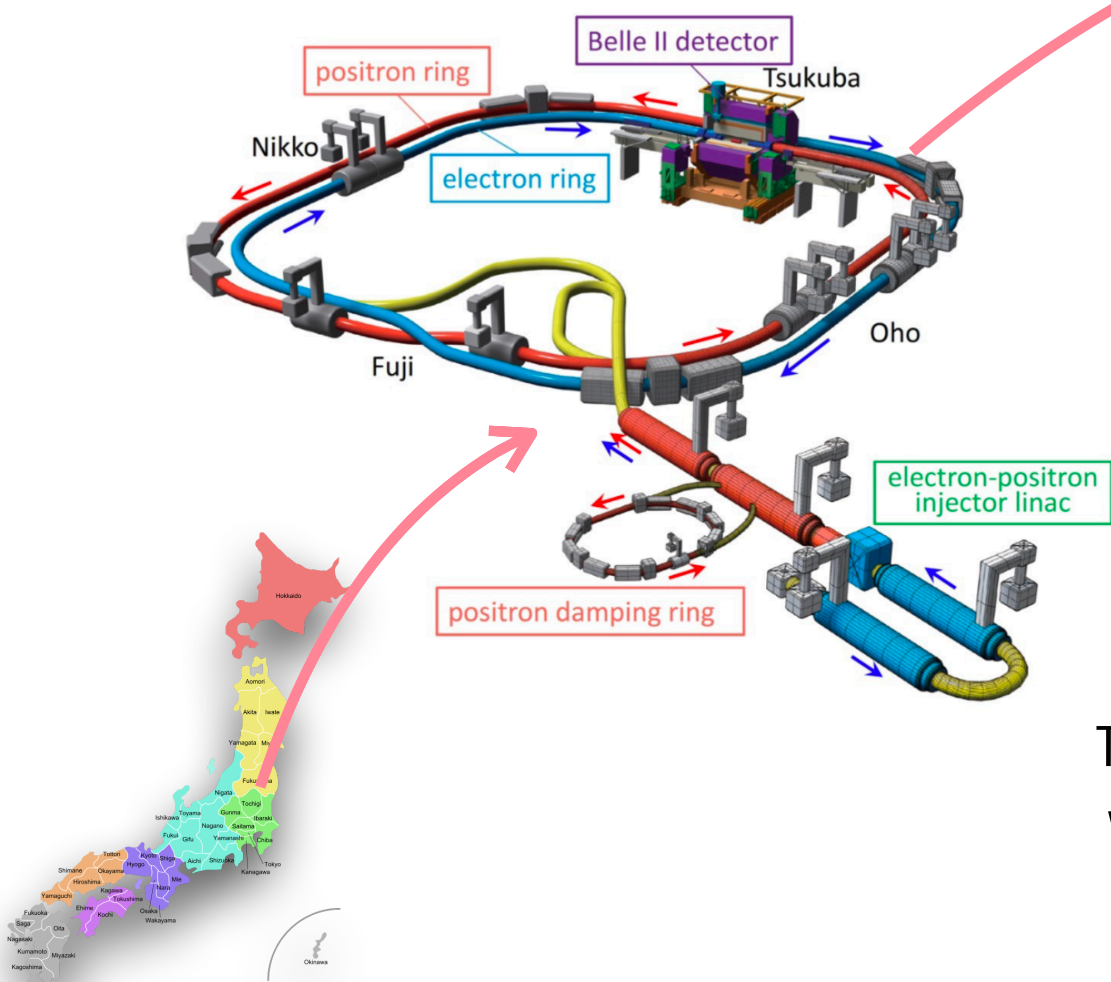
  $e^+e^- \to q\bar{q}, q = u, d, s, c$
- $e^+e^-$constrained kinematics and **no other particles** at threshold
- Known initial kinematics and good hermeticity: possible to fully reconstruct events with invisible particles

# Belle (II)—Collide $e^+e^-$ at $\Upsilon(4S)$

$2^{nd}$ generation B factory






Designed luminosity: $6 \times 10^{35}$ $cm^{-2}$ $s^{-1}$
Currently: $4.7 \times 10^{34}$ $cm^{-2}$ $s^{-1}$ > 2 $fb^{-1}$ per day!

$2^{nd}$ generation B-factory based on the **nanobeam** scheme.
The upgrade required a substantial redesign of the Belle II detector, whose performance is challenged by radiation damage and higher background (**design luminosity is x40 higher**). The aim is to guarantee equal or better performance than Belle @ KEKB.

# Why we need Full Event Interpretation?

- Important physics can be obtained from several challenging modes with **missing neutrinos**, either from B meson decays or originating from tau leptons.

$B \to D^{(*)}\tau\nu$

$B \to \tau\tau$

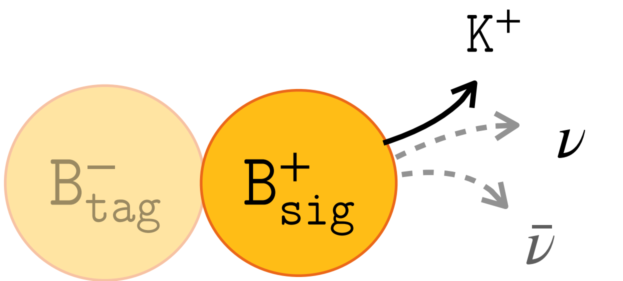
$B \to \tau\nu$

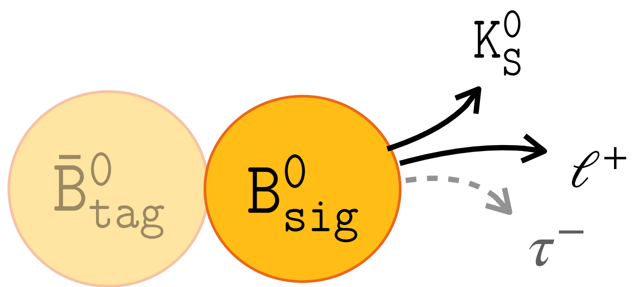
$B \to K\nu\bar{\nu}$

$B \to K\tau\tau$

$B \to \tau\ell$

$B \to K_{(S)}\tau\ell$

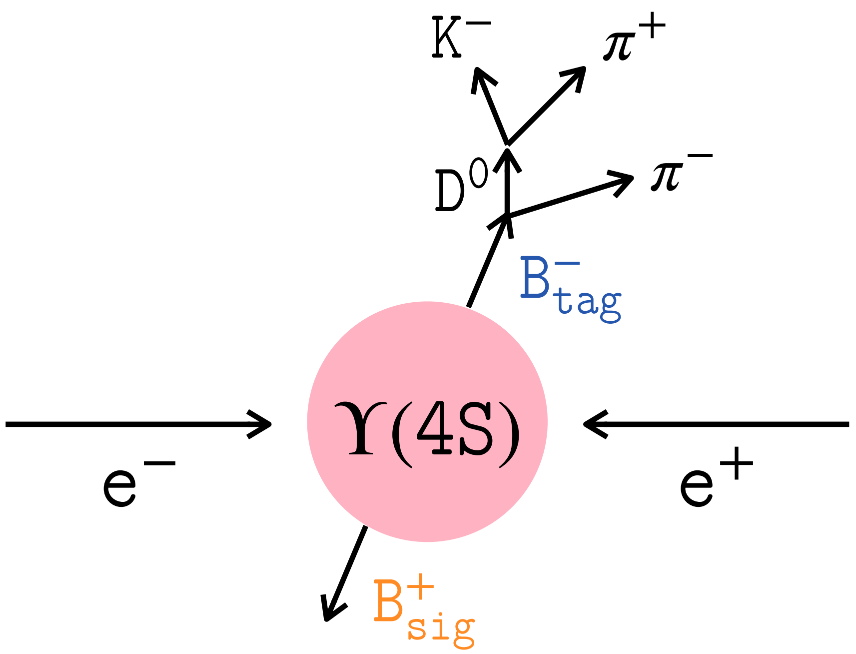



# Belle II experiment - 2 B's and nothing else

- $e^+e^-$ collision at $\Upsilon(4S)$ resonance
- A pair of $B\bar{B}$ is produced at threshold → **low backgrounds**
- $\Upsilon(4S) \rightarrow B^+B^-, B^0\bar{B}^0$ with $\mathscr{B}$~100%
- Reconstruct one of the B-mesons in either semileptonic or hadronic decay chains ($B^-_{tag}$)
- Flavour constraint: $B^-_{tag}$ → $B^+_{sig}$

Example of mode with hadronic $B_{tag}$



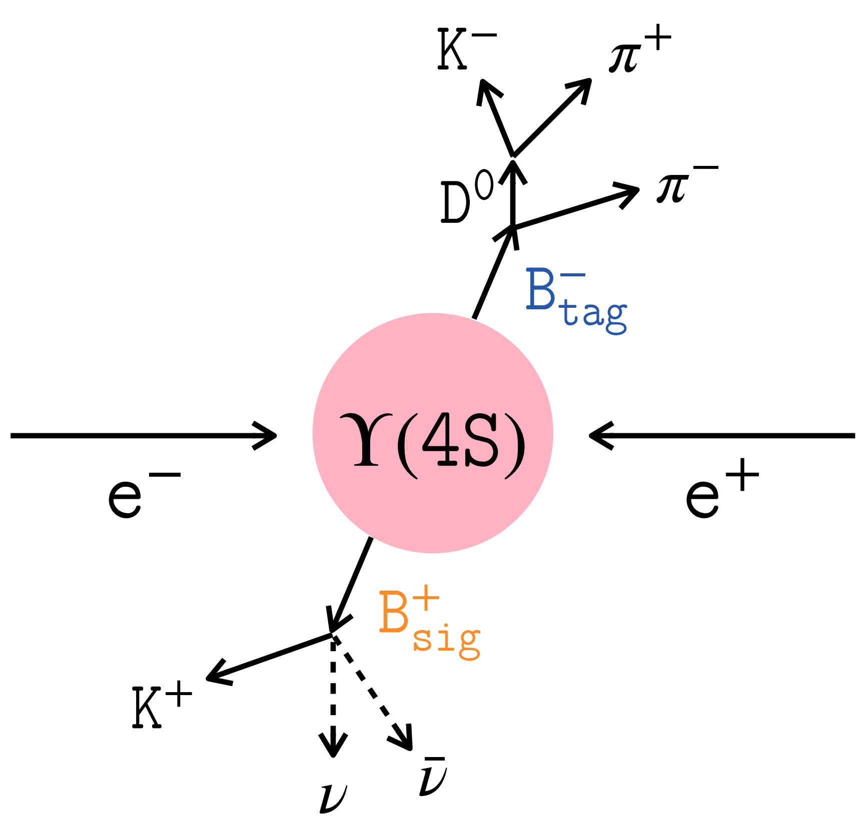
# Belle II experiment - 2 B's and nothing else

- $e^+e^-$ collision at $\Upsilon(4S)$ resonance
- A pair of $B\bar{B}$ is produced at threshold → **low backgrounds**
- $\Upsilon(4S) \rightarrow B^+B^-, B^0\bar{B}^0$ with $\mathscr{B}$~100%
- Reconstruct one of the B-mesons in either semileptonic or hadronic decay chains ($B^-_{tag}$)
- Flavour constraint: $B^-_{tag} \rightarrow B^+_{sig}$
- Kinematically constrained system with hadronically tagged event:

$$p_{e^+} + p_{e^-} = p_{B_{sig}} + p_{\bar{B}_{tag}}$$

Get $B_{sig}$ momentum even with multiple missing neutrinos

Example of mode with hadronic $B_{tag}$



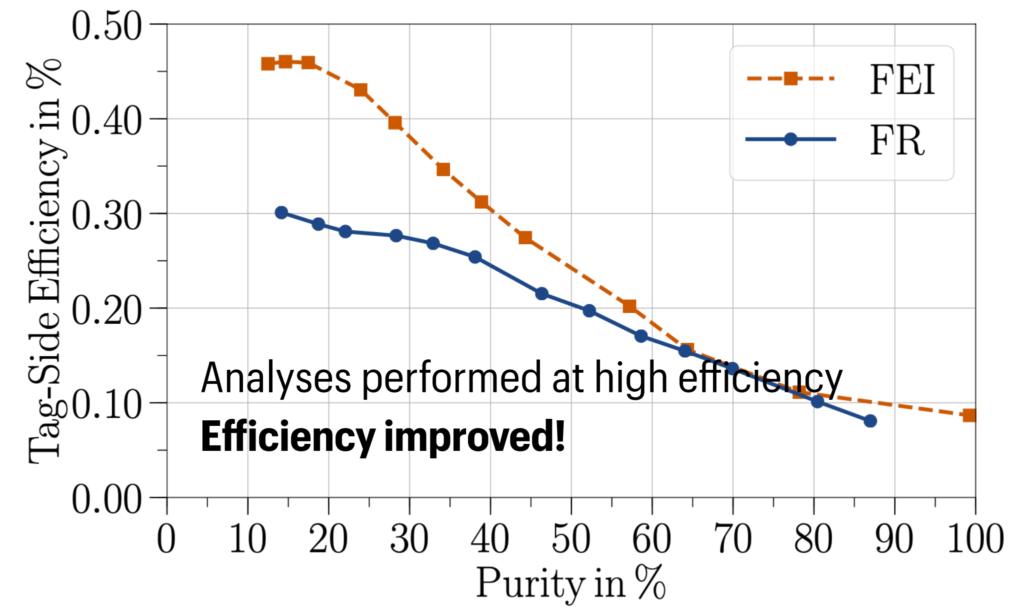
# What is Full Event Interpretation (FEI)?

- Flexible multivariate tagging algorithm developed for B-meson reconstruction in Belle II

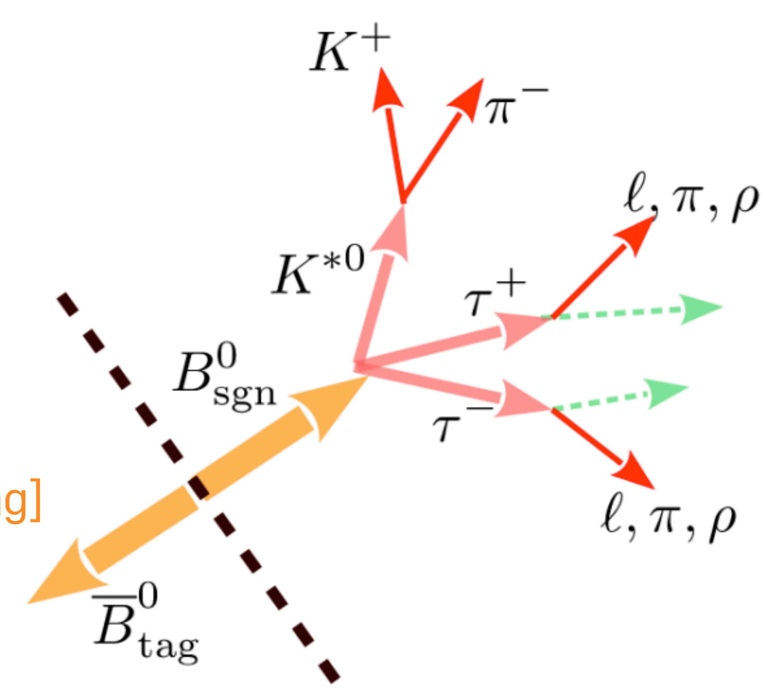
[ Keck, T. et al. Comput. Softw. Big. Sci. (2019) 3: 6 ]

- **Task**: Correctly identifying one B decay ($B^{-}_{tag}$) allows detailed investigation of the other B ($B^{+}_{sig}$)
- **Use in B-physics**: Especially useful when studying modes with missing energy (modes with one or more neutrinos, specific dark matter searches)
- Can be used on Belle data set
- Successor of the Belle **F**ull **R**econstruction (FR)

[ Feindt, M. et al. Nucl.Instrum.Meth.A 654 (2011) 432-440 ]




Analysis example $B^0 \rightarrow K^{*0}\tau^+\tau^-$



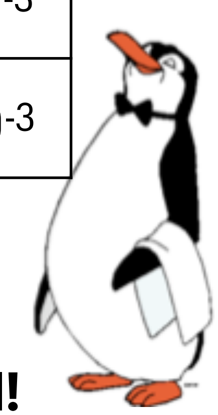

[Hadronic B tagging]

@90%CL

| | Dataset | Tag Algorithm | $\mathscr{B}^{UL}$ |
|---|---|---|---|
| [PRD 108 L011102 (2023)] BELLE | 711 $fb^{-1}$ | FR | $3.1\times10^{-3}$ |
| [New release in ICHEP 2024] Belle II | 362 $fb^{-1}$ | FEI | $1.8\times10^{-3}$ |

**Almost 4× improvement**

**Dominant improvement source is FEI!**



# How does it work?

- FEI uses hierarchical approach to reconstruct $\mathcal{O}(200)$ decay channels via $\mathcal{O}(10^4)$ decay chains

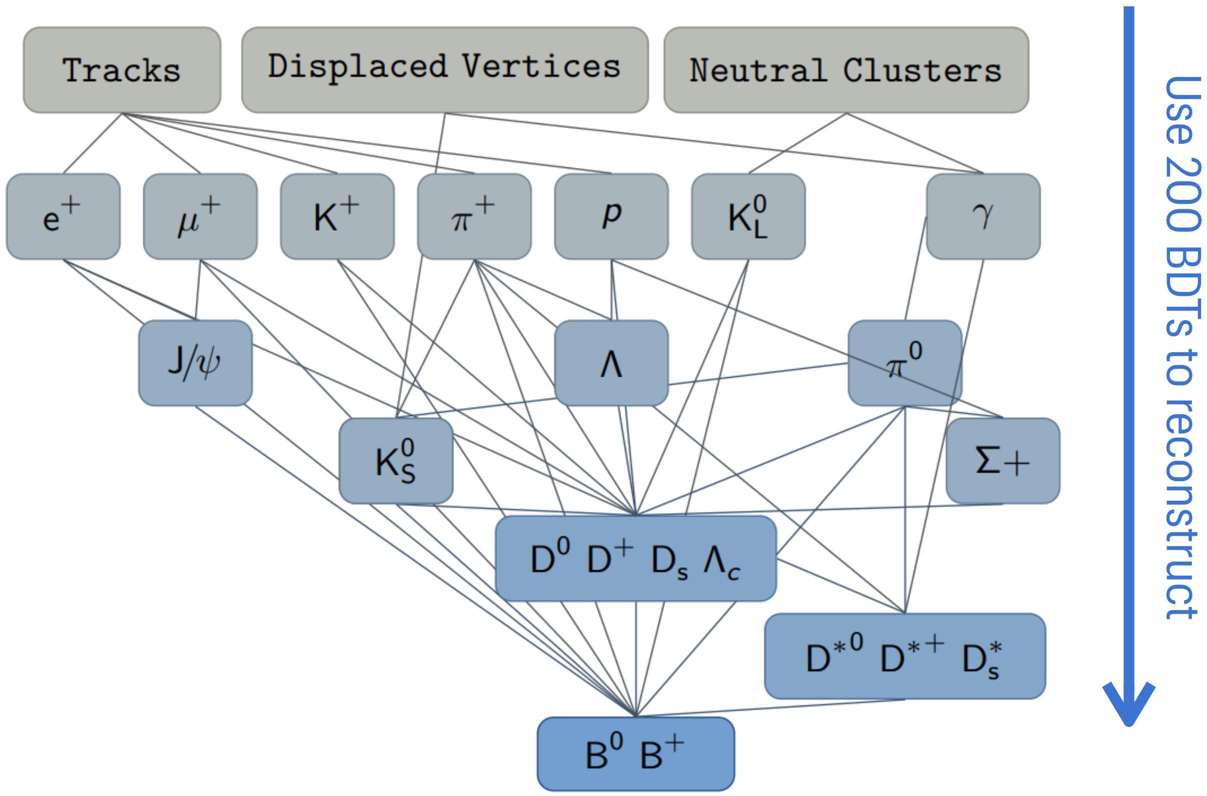
Tracks, neutral clusters and displaced vertices

↓

Combined into intermediate states

↓

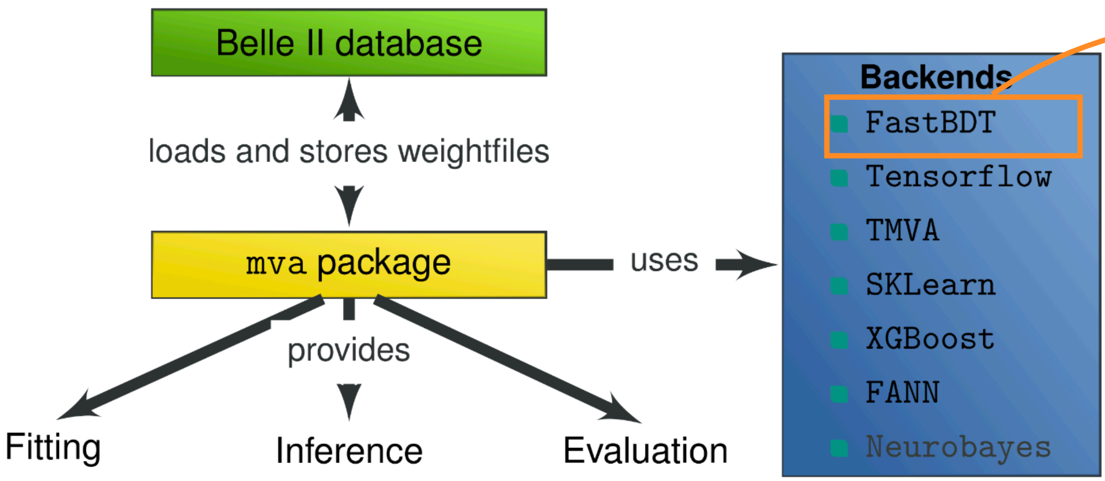
B meson



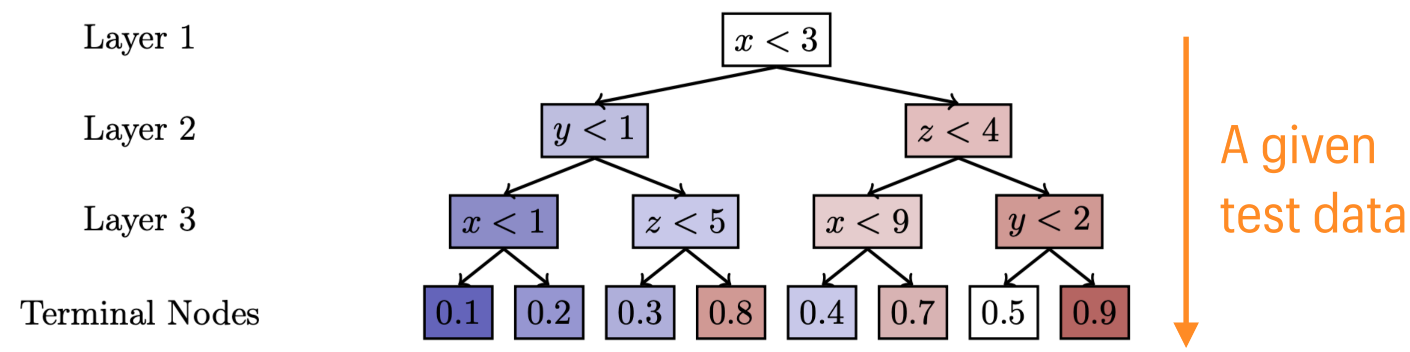

- Each unique particle has its own multivariate classifier which quantifies the correctness of reconstruction based on input features such as four-momentum, vertexing information...
- Training inputs: kinematic variables of the decay chains, such as invariant mass, momentum...
- Training output:
  - List of tag candidates
  - A probability to have correct reconstruction (signal probability)

# Training model: fastBDT

- The algorithm was originally designed for the FEI to speed up the training and application phase.




A **speed-optimized** and **cache-friendly** implementation of stochastic gradient-boosted decision trees for multivariate classification.




- At each node of the tree **a binary decision** is made until a terminal node is reached.
- Probability of test data to be **signal** (number stated in terminal node layer) is signal-fraction of all training data-points, which ended up in the same terminal node.

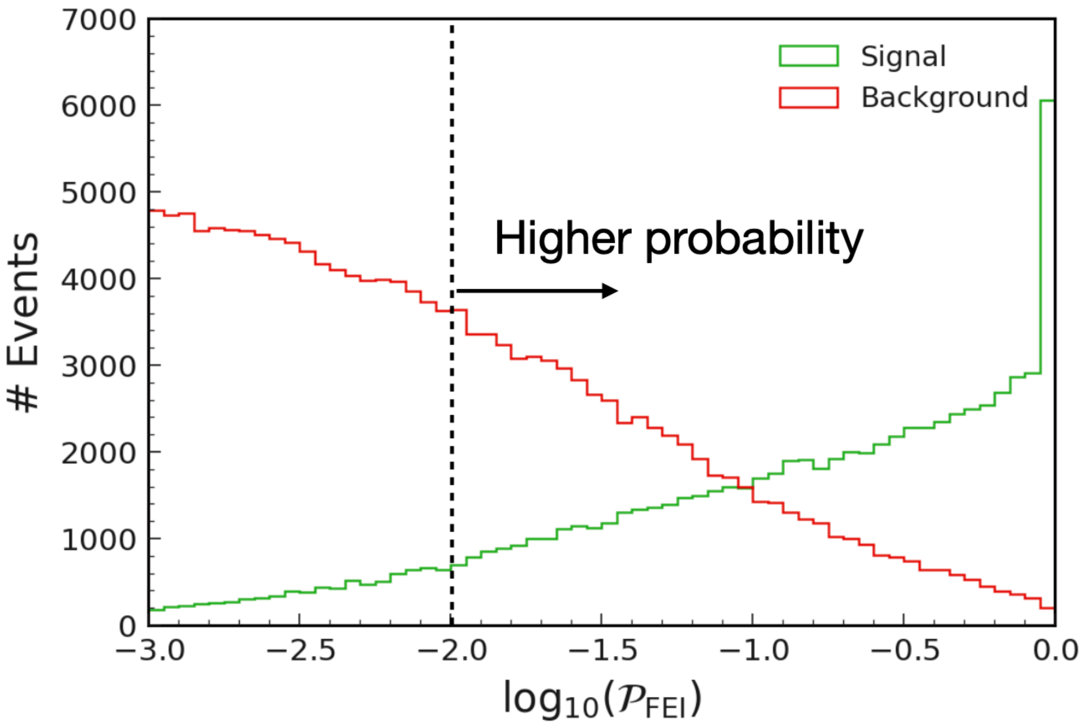
- **Gain an order of magnitude in execution time** by optimizing mainly the implementation of the algorithm.
- Most of time when using fastBDT is spent during the extraction of necessary features, therefore no further significant speedups can be achieved by employing a different method.

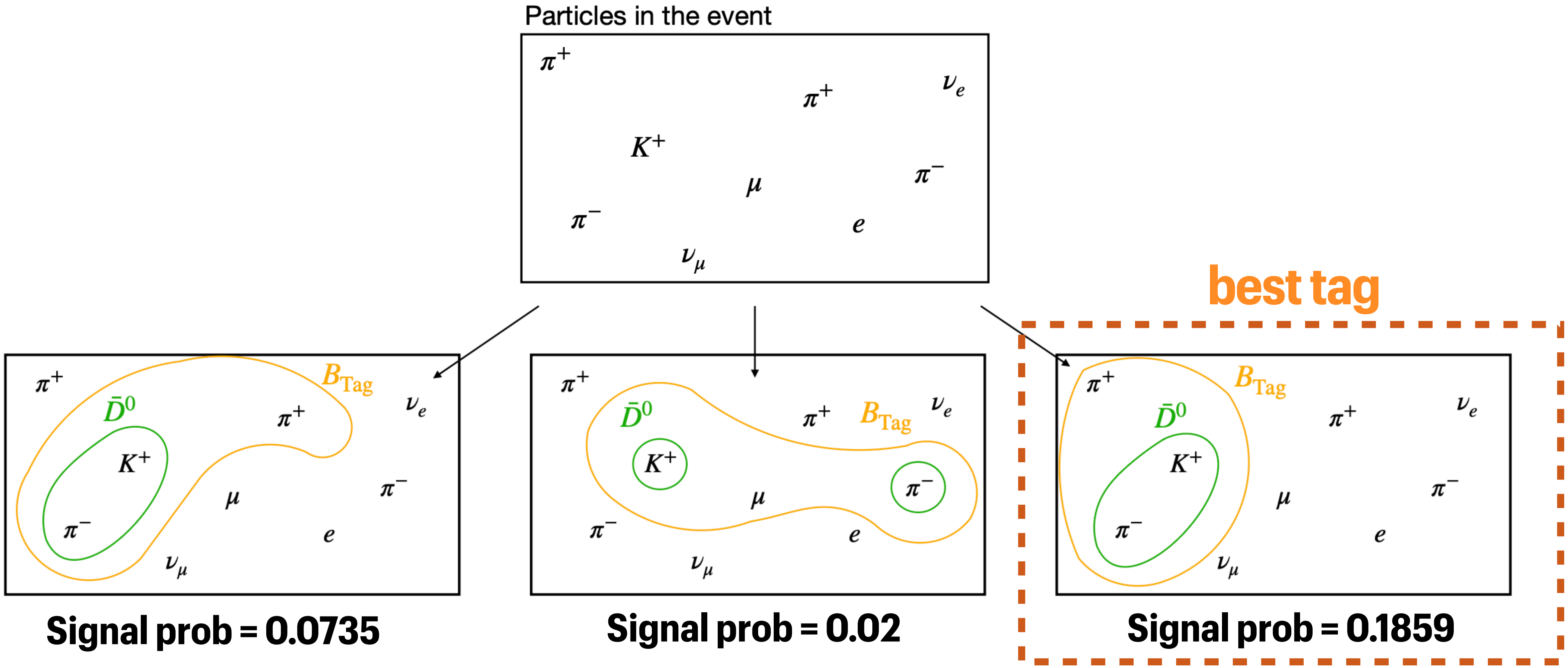
ArXiv: http://arxiv.org/abs/1609.06119

# How do we select good and best tag?

**Signal probability**

Enhance your purity based on selection on the signal probability

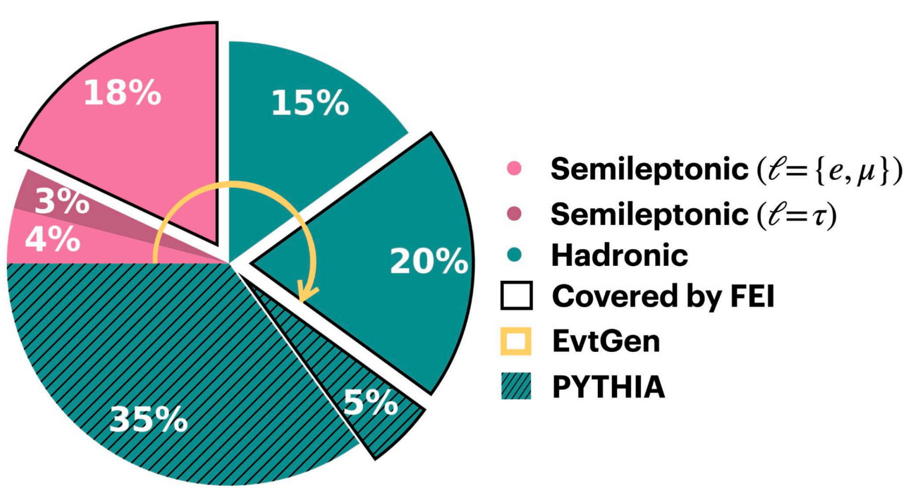


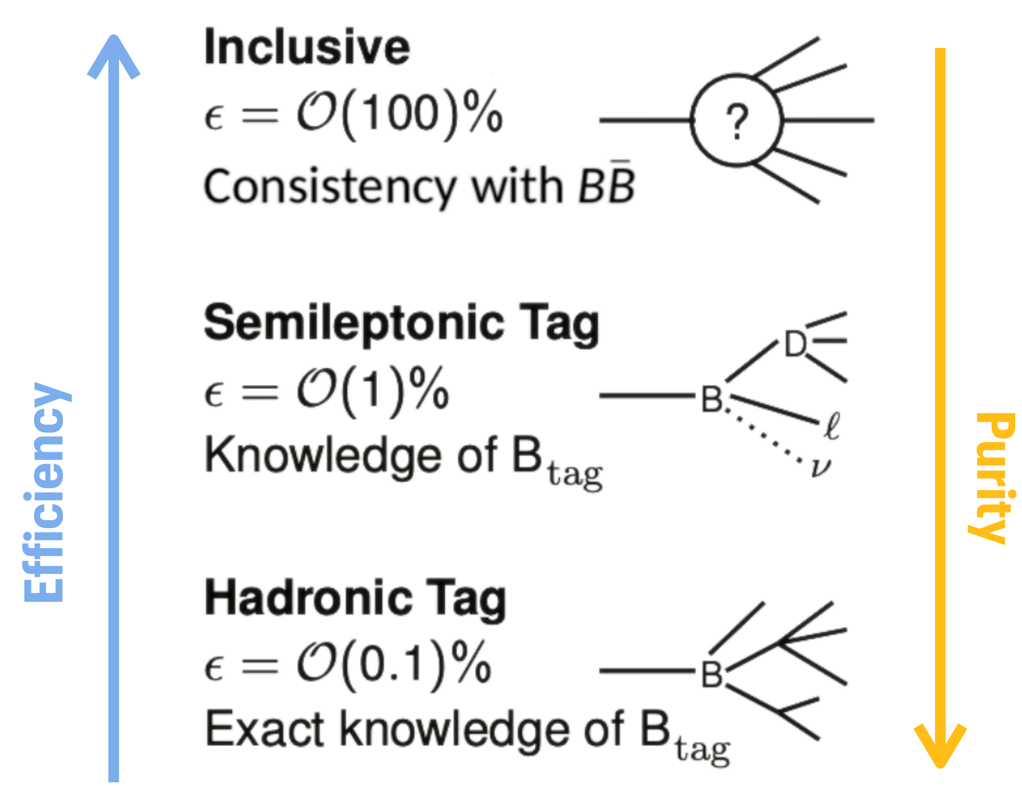



# Tagging Techniques at Belle II

- **High efficiency**: fraction of events that are identified as a tag
- **High Purity**: fraction of identified tags that are "correct"
- Good kinematic information: minimise missing/fake


- Trade-off between efficiency, purity, and knowledge of missing kinematics
- Generic FEI techniques include reconstruction of the B-meson candidate with
  - Semileptonic Tagging
  - Hadronic Tagging



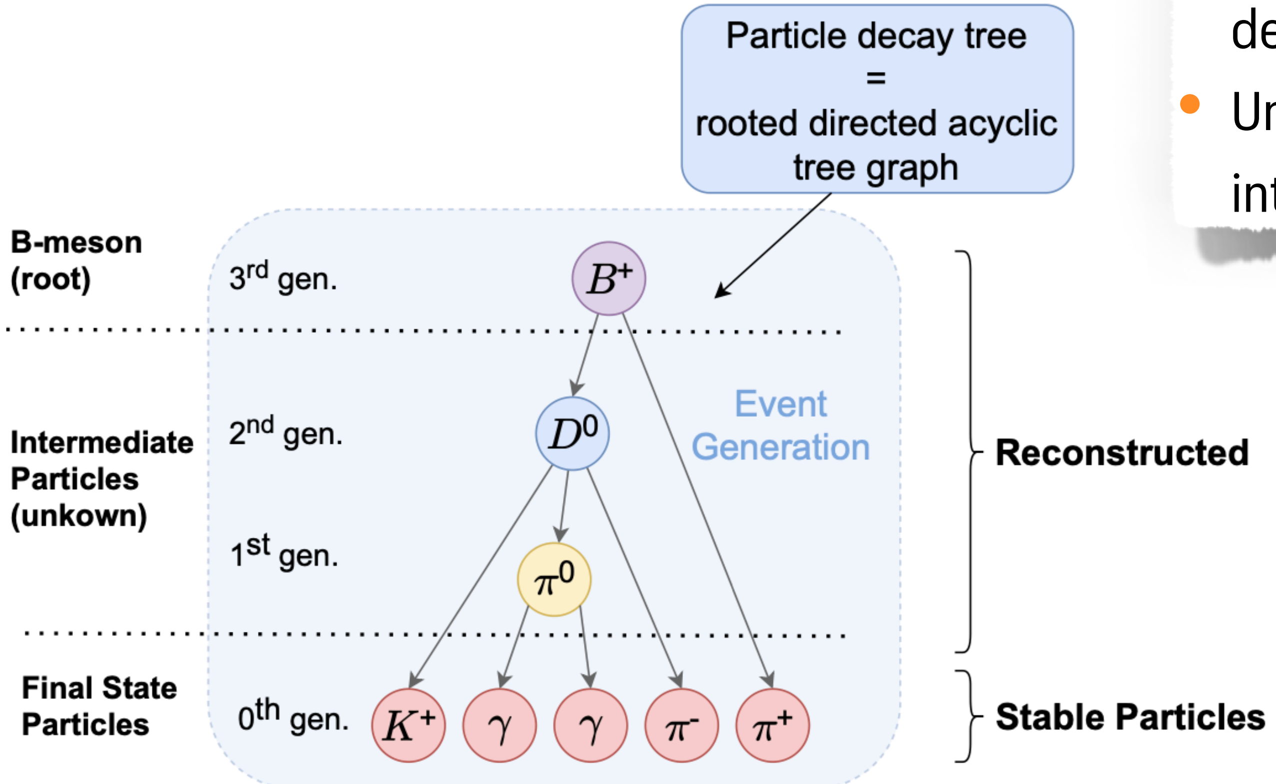
# Graph-based Full Event Interpretation

[ACAT_2022_proceeding]

**Full Event Interpretation** FEI: decay modes are hard-coded

Particle decays are naturally described by rooted directed acyclic tree graphs

- Goal: develop **graph-based Full Event Interpretation (graFEI)** to **inclusively** reconstruct tag B meson (Increase the tagging efficiency)
- Proof of concept: Learning tree structures from leaves for particle decay reconstruction, *Kahn et al 2022*




- Only final-state particles are detected by experiments
- Unknown total number of intermediate particles

- LCA of two nodes is defined as the **farthest** node from the root that is an ancestor of both nodes.

**Adjacency Matrix**

| | $B^+$ | $D^0$ | $\pi^0$ | $K^+$ | $\gamma$ | $\gamma$ | $\pi^-$ | $\pi^+$ |
|---|---|---|---|---|---|---|---|---|
| $B^+$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| $D^0$ | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| $\pi^0$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| $K^+$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\gamma$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $\gamma$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $\pi^-$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\pi^+$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Lowest Common Ancestor (LCA) Matrix**

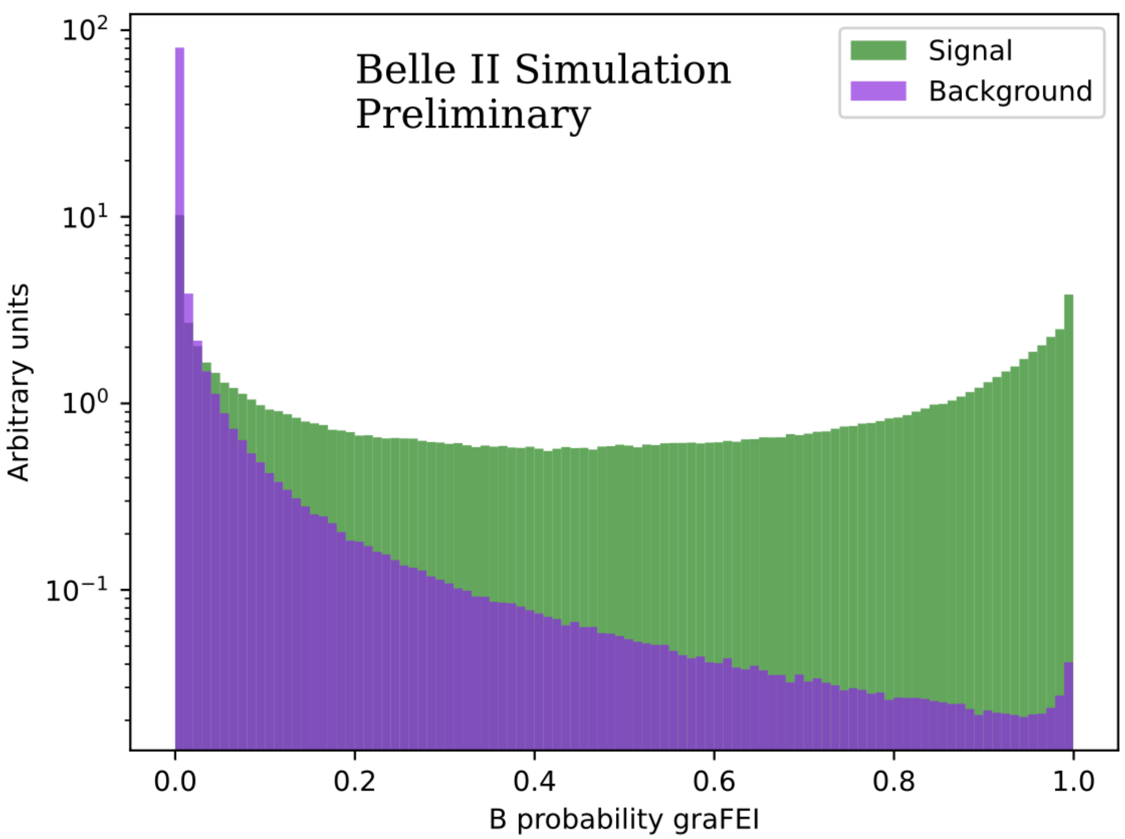
| | $K^+$ | $\gamma$ | $\gamma$ | $\pi^-$ | $\pi^+$ |
|---|---|---|---|---|---|
| $K^+$ | $K^+$ | $D^0$ | $D^0$ | $D^0$ | $B^+$ |
| $\gamma$ | $D^0$ | $\gamma$ | $\pi^0$ | $D^0$ | $B^+$ |
| $\gamma$ | $D^0$ | $\pi^0$ | $\gamma$ | $D^0$ | $B^+$ |
| $\pi^-$ | $D^0$ | $D^0$ | $D^0$ | $\pi^-$ | $B^+$ |
| $\pi^+$ | $B^+$ | $B^+$ | $B^+$ | $B^+$ | $\pi^+$ |

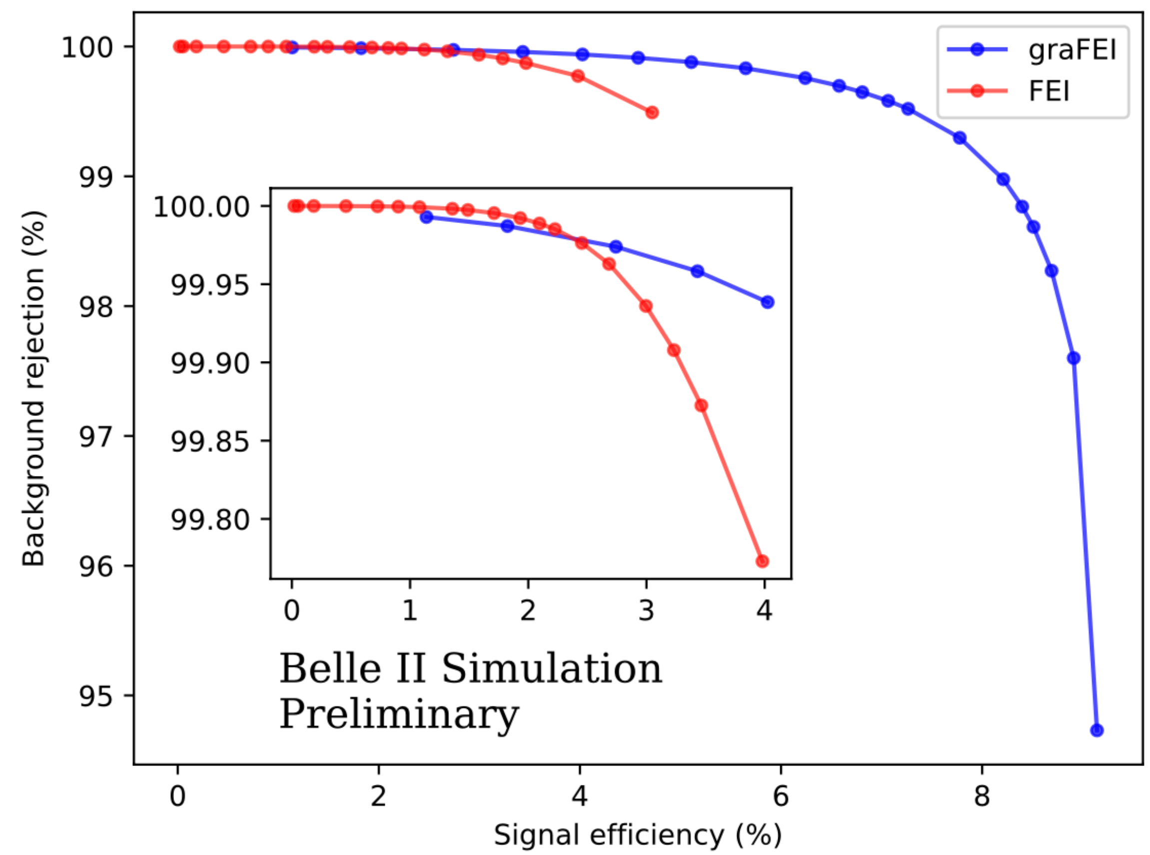
# Comparison with FEI

Having a **definition of "B probability"** analogous to FEI is needed

- Each LCA element has a corresponding probability of belonging to the predicted class given by the model
- Arithmetic mean of class probabilities defined as B probability

$$\text{LCA} = \begin{pmatrix} 0 & 3 & 5 \\ 3 & 0 & 5 \\ 5 & 5 & 0 \end{pmatrix} \longleftrightarrow \begin{pmatrix} 0 & 0.62 & 0.31 \\ 0.62 & 0 & 0.76 \\ 0.31 & 0.76 & 0 \end{pmatrix} \rightarrow 0.563$$




graFEI: maximum efficiency 9.1%, background rejection 94.7%
FEI: maximum efficiency 4.7%, background rejection 99.5%

# Summary

- FEI algorithm has been used in many analyses in Belle and Belle II:

  $R_{D^{(*)}}$ [PRL 124 161803 2020], $B^+ \to \ell^+\nu_\ell\gamma$ [PRD 98 112016 2018], $B^+ \to K^+\tau^\pm\ell^\mp$ [PRL 130 261802 2023], $B^+ \to K^+\nu\bar{\nu}$ [PRD 109 112006 2024], $B^0 \to K^{*0}\tau^+\tau^-$, $B^0 \to K_S^0\tau^\pm\ell^\mp$ [released in ICHEP 2024] ……

- Overall improvement of hadronic FEI
  - Updated decay model for 11 most efficient decay modes

    0.65 → 0.81 :  25% ⬆️ in calibration factor
  - Training with the new MC

    56% → 63% :  12% ⬆️ in purity
  - Loosen the preselection and mass-constraint $\pi^0$

    0.93% → 1.13% :  21% ⬆️ in efficiency

  Belle II is measuring more relevant modes of hadronic FEI.

- A novel approach - Graph-based Full Event Interpretation (GraFEI) is developed and will be used in more analyses.

# Backup

# FEI performance in data: current status

*Hadronic tag ($B^+$) as an example*

Calculated directly on data

- Calibration factor : **65%**

$$\frac{\text{Signal yield in data}}{\text{Signal yield in MC}}$$

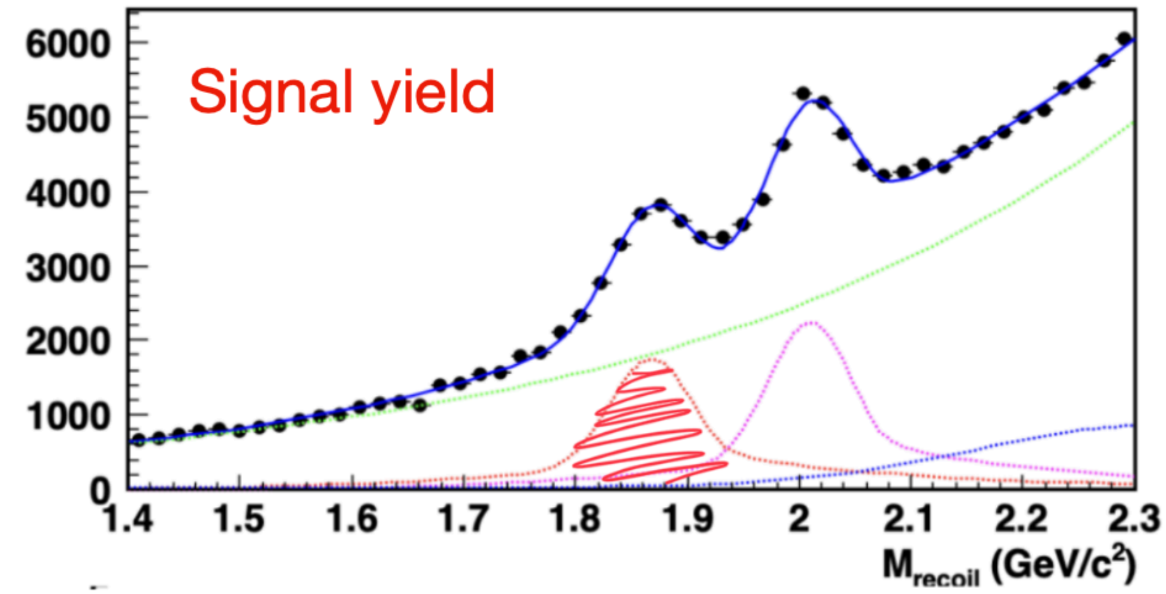
- Purity: **56%**

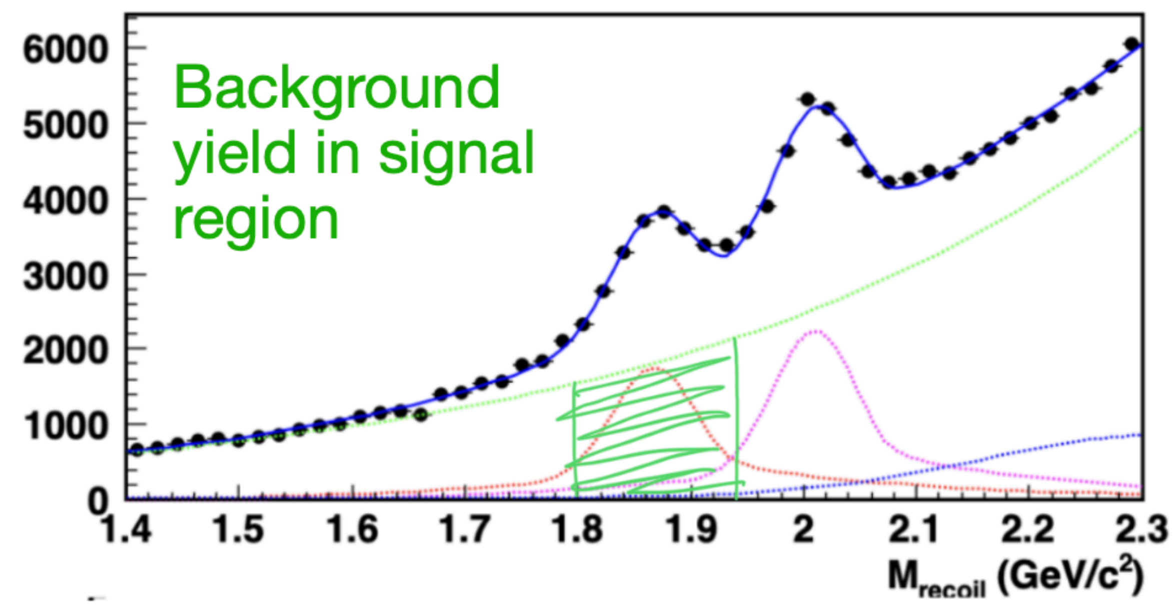
$$\frac{\text{Signal yield}}{\text{Signal yield} + \text{Background yield in signal region}}$$

- Efficiency: **0.93%**

$$\frac{\text{Signal yield}}{n_{BB} \cdot BF_{B\rightarrow D\pi} \cdot \epsilon_{\pi}}$$

$n_{BB}$ ↓ $392.5 \times 10^6$ ; $BF_{B\rightarrow D\pi}$ ↓ PDG ; $\epsilon_{\pi}$ ↓ 90%



# FEI performance in data: current status

*Hadronic tag ($B^+$) as an example*

Calculated directly on data

- Calibration factor : 65%

$$\frac{\text{Signal yield in data}}{\text{Signal yield in MC}}$$

- Purity

$$\frac{\text{Signal yield}}{\text{Signal yield} + \text{Background yield in signal region}}$$

- Efficiency

$$\frac{\text{Signal yield}}{n_{BB} \cdot BF_{B\to D\pi} \cdot \epsilon_{\pi}}$$

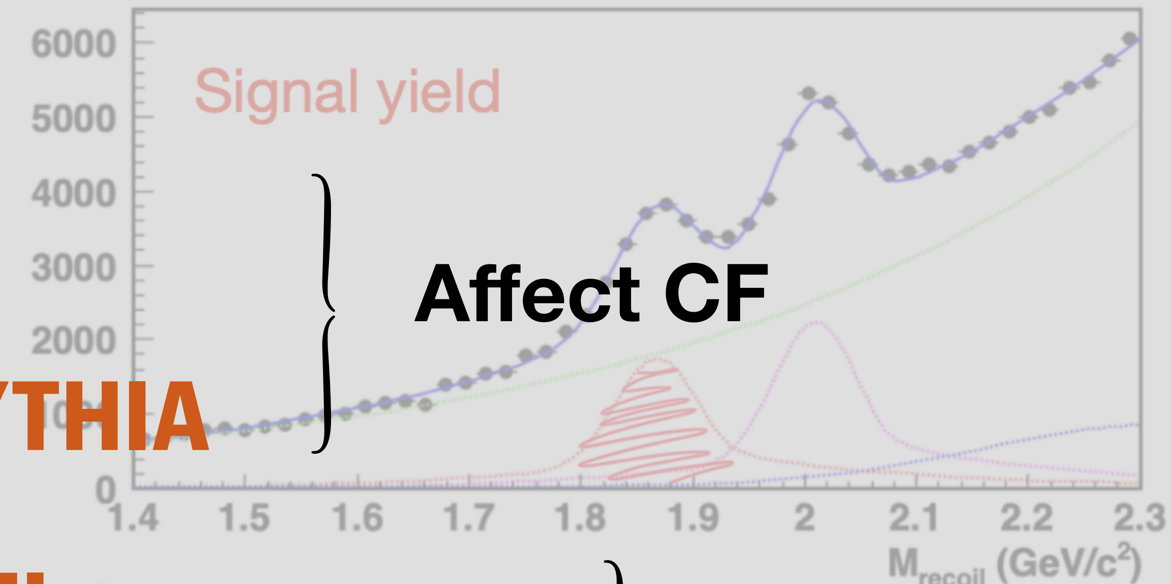
$392.5 \times 10^6$ PDG 90%

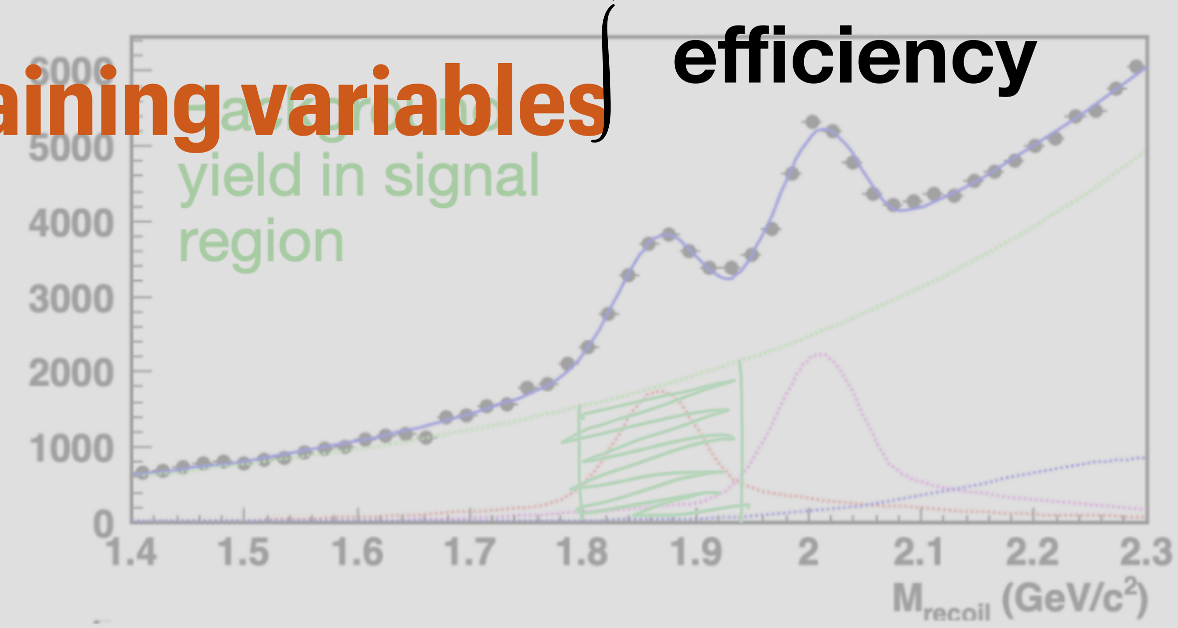
- ▶ **Wrong/outdated BFs in MC**
- ▶ **Half of the MC is unknown: PYTHIA**

**Affect CF**

- ▶ **Loose selections applied in FEI**
- ▶ **Suboptimal choice of input training variables**

**Affect purity & efficiency**

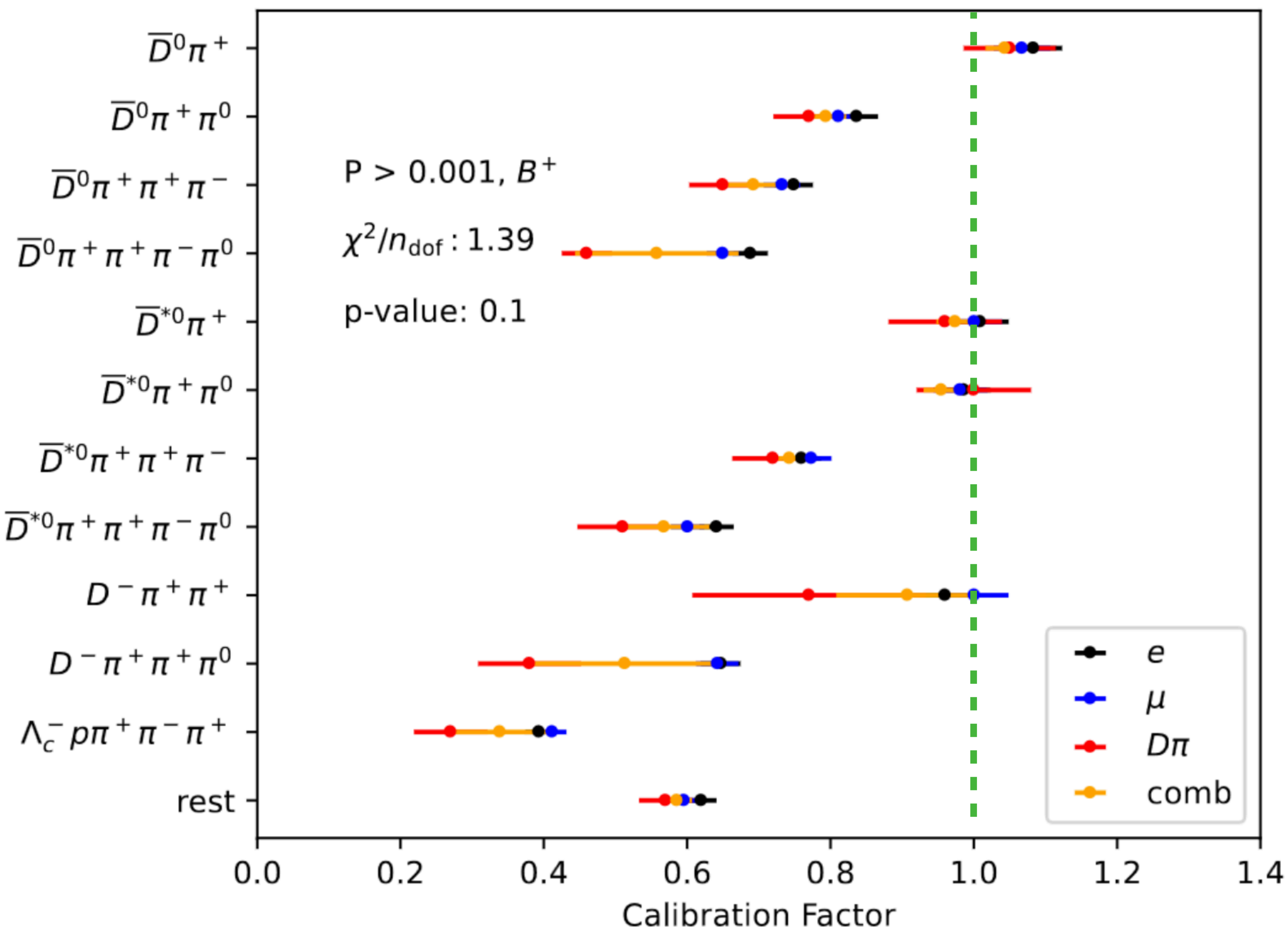


# Calibration

*Hadronic tag as an example*

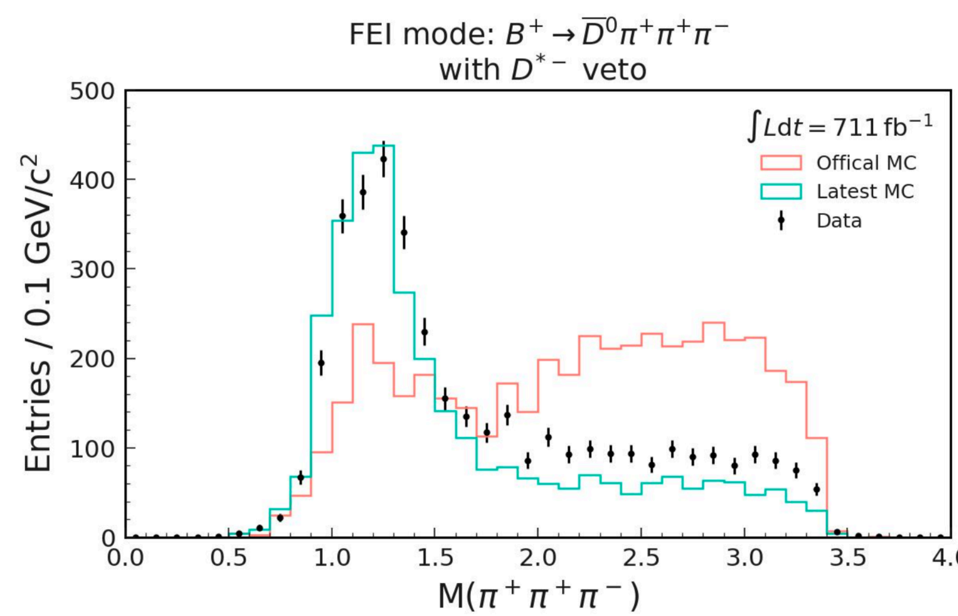
- Two independent control samples are adopted as signal-sides to calibrate the data-MC difference of B tagging
- Calibration factors (CFs) are calculated as ratio of signal yields of data and MC
- Good agreement of CFs despite two orthogonal signal-sides




Two samples are combined to get final CF

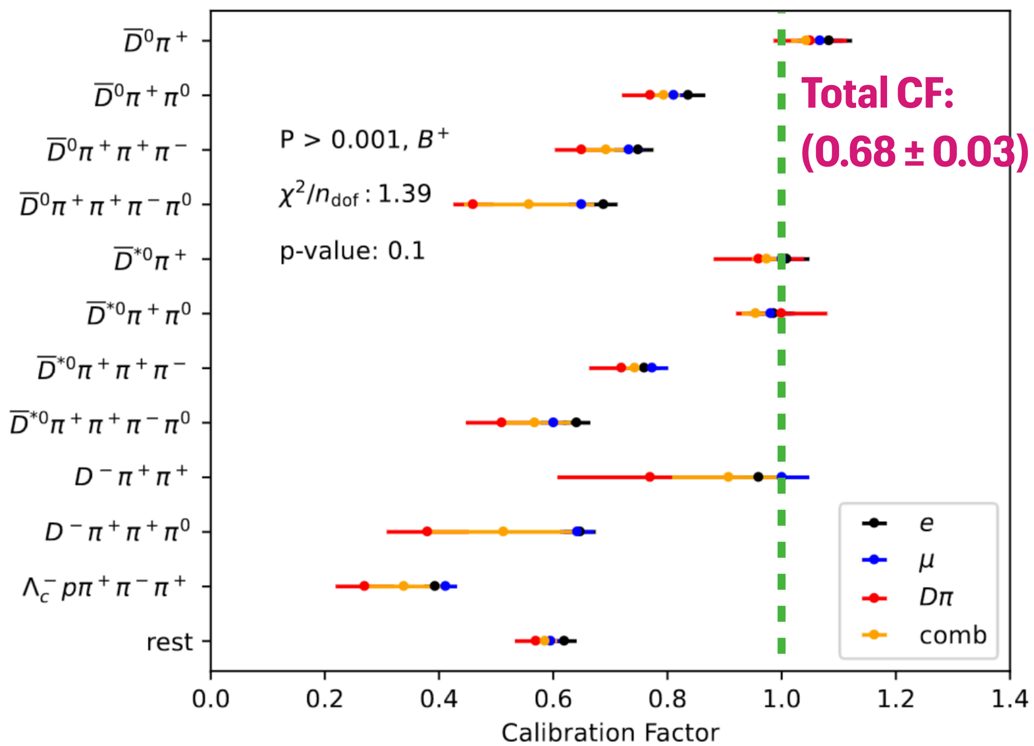
**Total CF:**
**(0.68 ± 0.03)**

Far from 1?
Discrepancy in data and simulation?

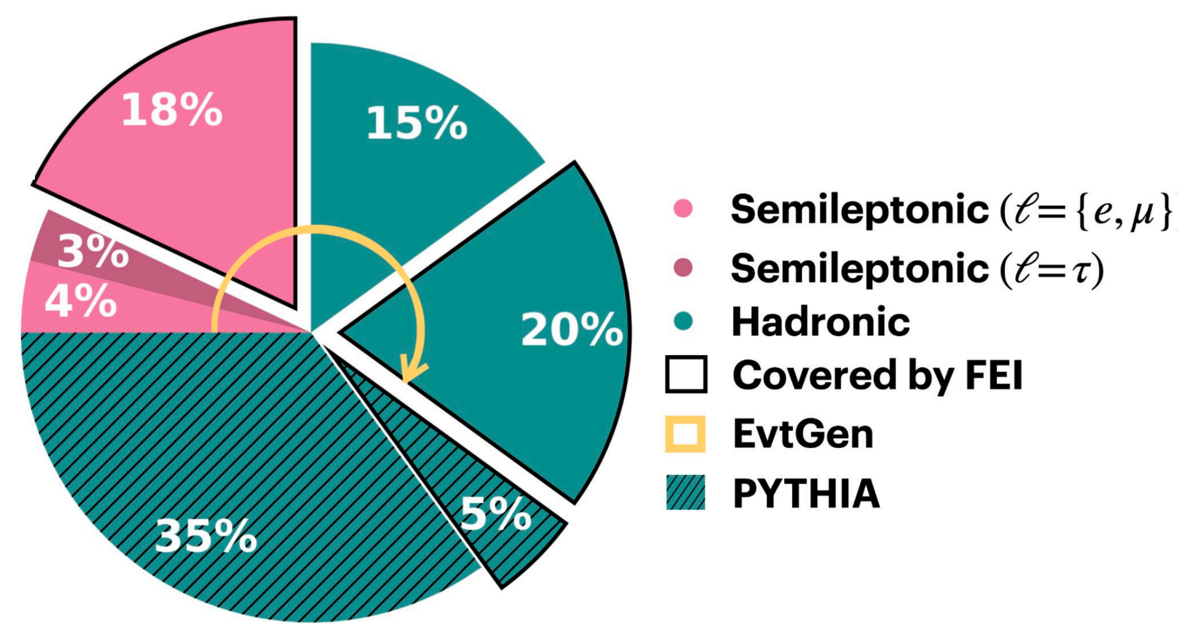


# Can we do better?

*Hadronic tag ($B^+$) as an example*




Why MC has such large discrepancy with data for hadronic tag?




- Branching fractions of hadronic B decays ~75%
- **Only half of it is measured** and the rest is generated by PYTHIA

Most of the known measurements are **performed with small data sets**

⇒ Large statistical uncertainties.

Poor MC (significantly different from reality/data)

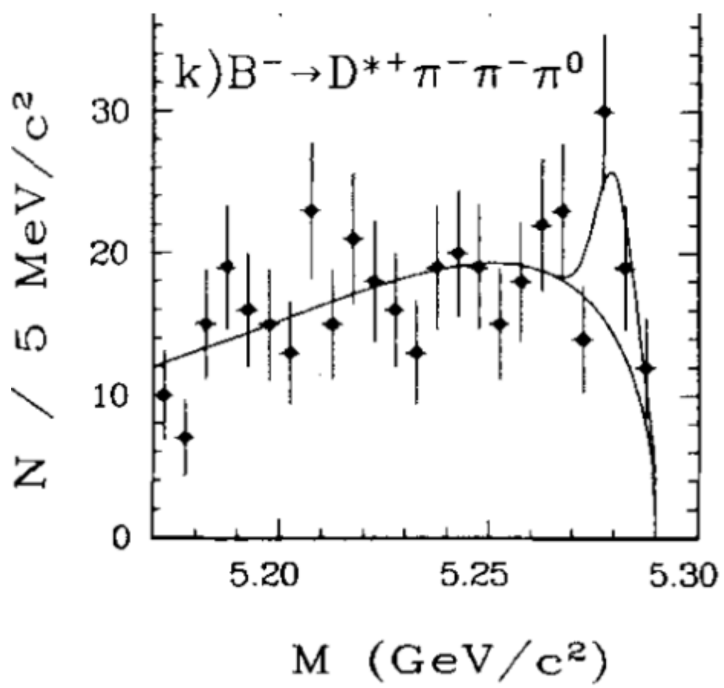
⇒ Poor hadronic B tagging

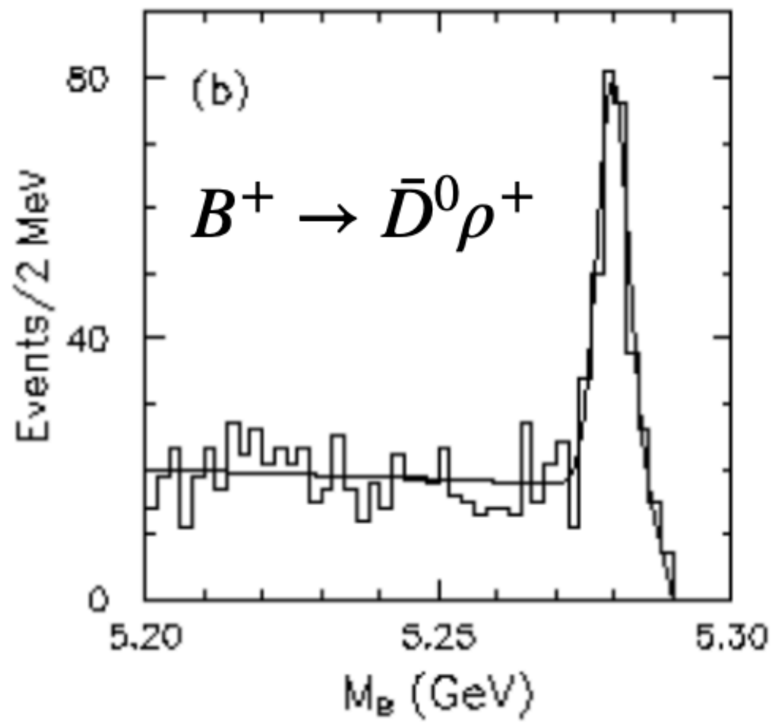
Understanding $B \to D^{(*)}h$ decays is essential for B tagging

# Decays in hadronic B-tagging

Understanding $B \to D^{(*)}h$ decays is essential for B tagging

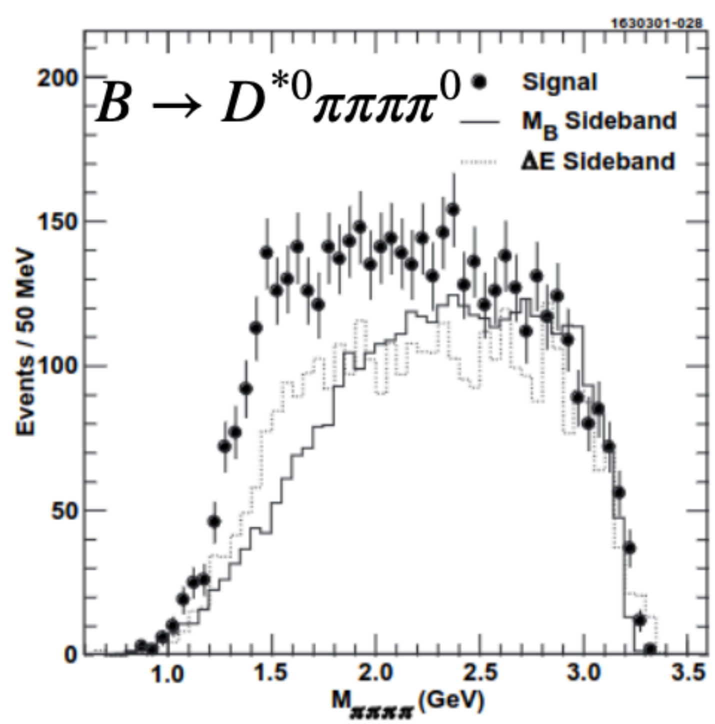
ARGUS, 229 $pb^{-1}$
33 years ago
$M_{bc}$ fit
$\mathcal{B}$ = (1.5±0.7)%
47% uncertainty!
[Z.Phys.C 48 (1990) 543-552]

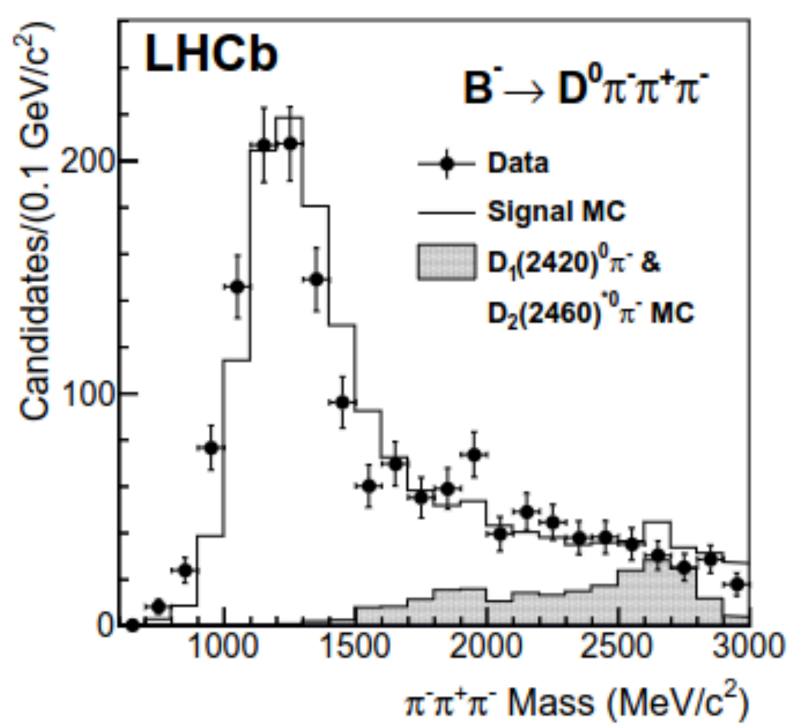



CLEO, 0.89 $fb^{-1}$
29 years ago
$M_{bc}$ fit
$\mathcal{B}$ = (1.34±0.18)%
13% uncertainty!
[PRD 50 (1994) 43-68]

CLEO, 9 $fb^{-1}$
22 years ago
$M_{bc}$ fit
$\mathcal{B}$ = (1.8±0.4)%
22% uncertainty!
But model? ⇒ ρ′?
[PRD 64 (2001) 092001]




LHCb, 35 $fb^{-1}$
12 years ago
But $\mathcal{B}(B^+ \to \bar{D}^0 a_1^+)$
not provided!
[PRD 84 (2011) 092001]

# Decays in hadronic B-tagging

Understanding $B \to D^{(*)}h$ decays is essential for B tagging

For decays with higher multipilicity, we need to know the decay kinematics

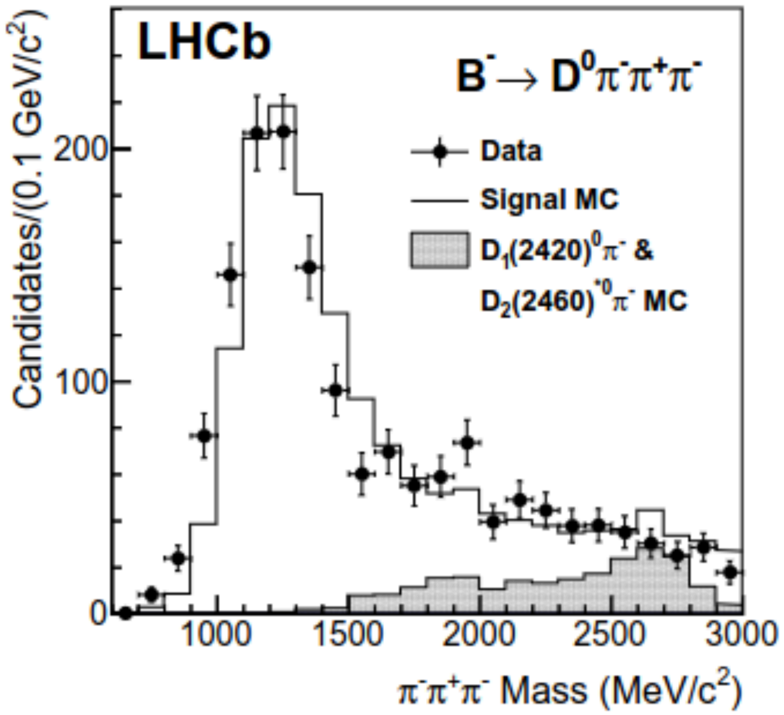
In MC, modelled as a coherent sum of decays through many intermediate resonances.

Inclusive $D^0\pi^-\pi^+\pi^-$

Measured:

$$\frac{\mathscr{B}\left(B^- \to D^0\pi^-\pi^+\pi^-\right)}{\mathscr{B}\left(B^- \to D^0\pi^-\right)} = 1.27 \pm 0.06 \pm 0.11$$

- But LHCb does not explicitly provide information on $a_1^+$...
- we are left with $\mathscr{B}(B^+ \to \bar{D}^0 a_1^+)$ = (0.4 ± 0.4)% and $\mathscr{B}(B^+ \to \bar{D}^0\pi^+\rho^0)$ = (0.4 ± 0.3)% from CLEO (1992, 212 $pb^{-1}$) in PDG.



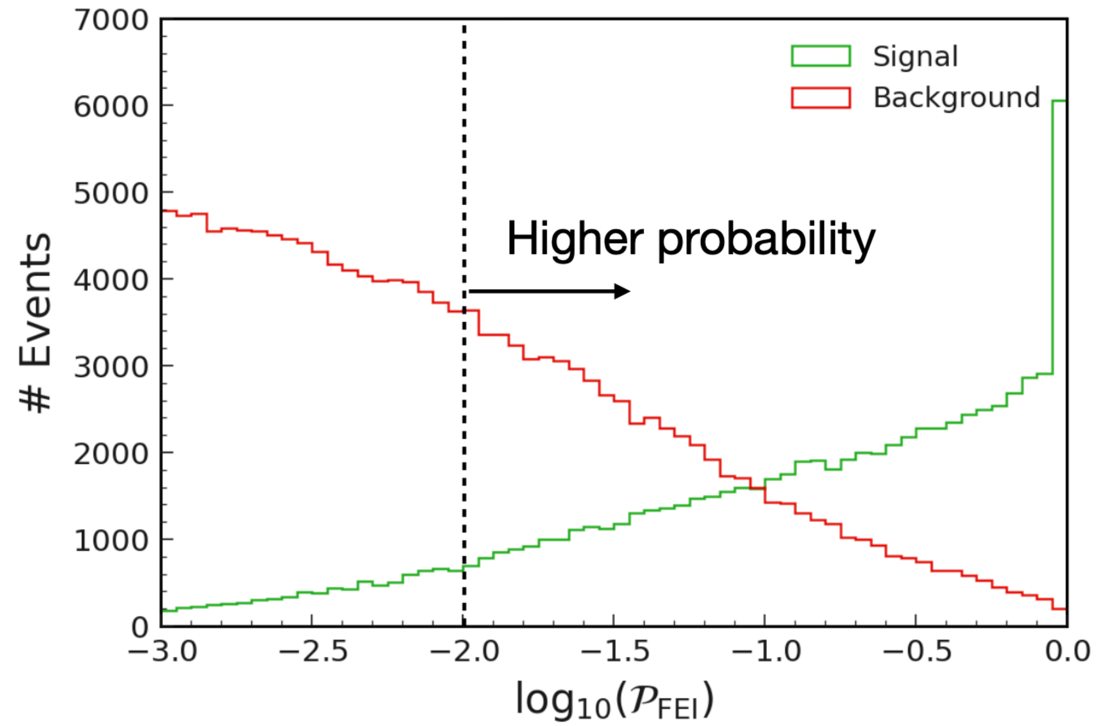

LHCb, 35 $fb^{-1}$
12 years ago
But $\mathscr{B}(B^+ \to \bar{D}^0 a_1^+)$
not provided!

[PRD 84 (2011) 092001]

# How do we select good tags?

**Signal probability**

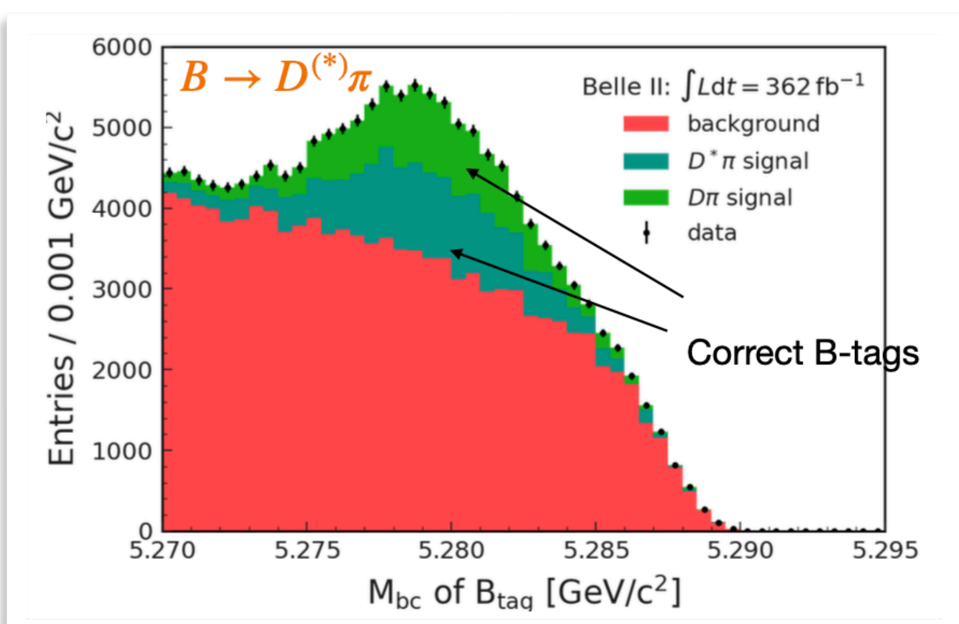
Enhance your purity based on selection on the signal probability




**Hadronic Tag**

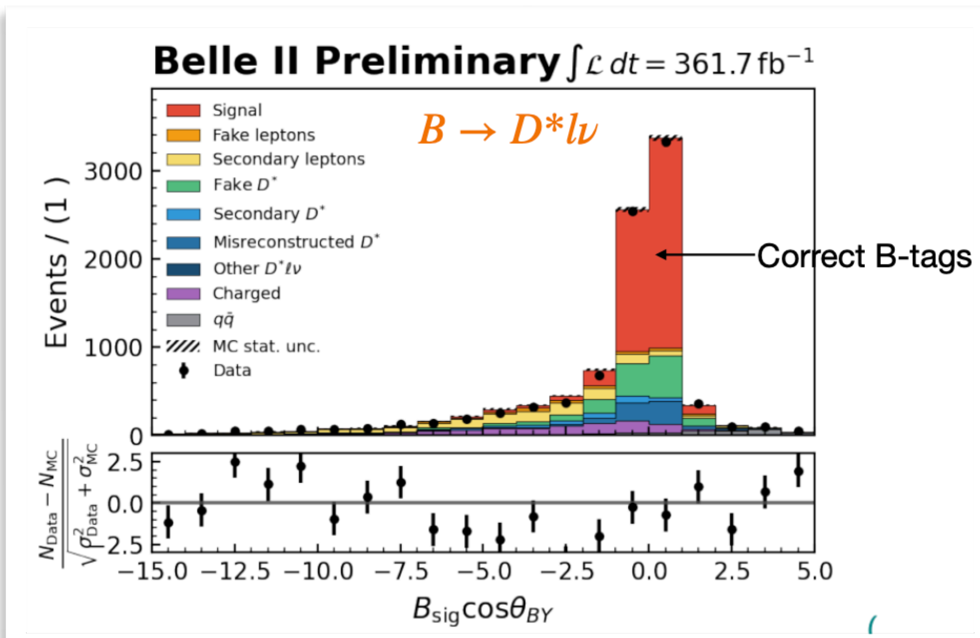
$$M_{bc} = \sqrt{(\sqrt{s}/2)^2 - \vec{p}_B^{*2}}$$ ~ B mass

$$\Delta E = E_B^* - \sqrt{s}/2$$ ~ 0




**Semileptonic Tag**

$$\cos\theta_{BY} \in [-1, 1]$$



# Decays in hadronic B-tagging

Implement first, and then validate

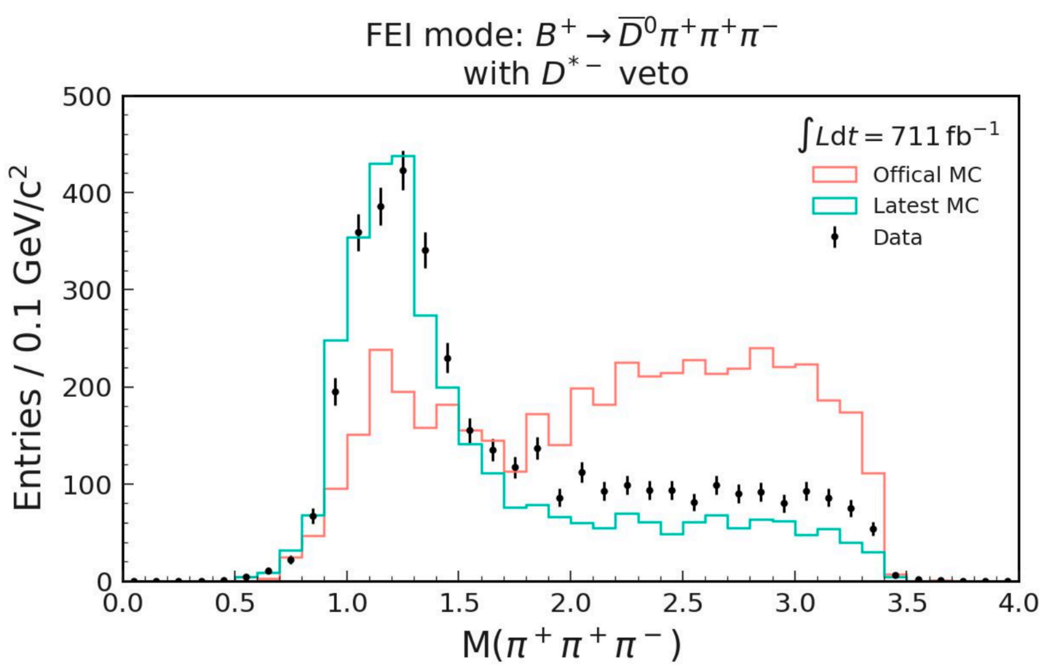
For decays with higher multiplicity, we need to know the decay model for MC.

Not necessarily the complete amplitude with interferences,
but something simple to set in MC,
i.e., intermediate resonances.

| $B^+$ FEI mode | Contribution | $\mathcal{B}^{\text{Belle}}(\%)$ | $\mathcal{B}^{\text{Belle II}}(\%)$ | $\mathcal{B}^{\text{proposed}}(\%)$ |
|---|---|---|---|---|
| $\overline{D}^0\pi^+\pi^-\pi^+$ | $\overline{D}^0\pi^+\pi^-\pi^+$ (NR) | 0.46 | 0.51 | 0 |
| | $\overline{D}^0\rho^0\pi^+$ | 0.39 | 0.42 | 0 |
| | $\overline{D}^0 a_1^+$ | 0.18 | 0.26 | 0.58 |
| | $\overline{D}_1^0\pi^+$ | 0.04 | 0.04 | 0.08 |
| | $\overline{D}_1^{\prime 0}\pi^+$ | 0.03 | 0.02 | 0.03 |
| | $\overline{D}_2^{*0}\pi^+$ | 0.01 | 0.01 | 0.01 |
| | $D^{*-}\pi^+\pi^+$ | - | 0.09 | 0 |
| | | **1.11** | **1.36** | **0.70** |

With the help of control sample $B \to D\pi^+$ (high signal-side purity), we validated our model via the $B_{tag}$ reconstruction:

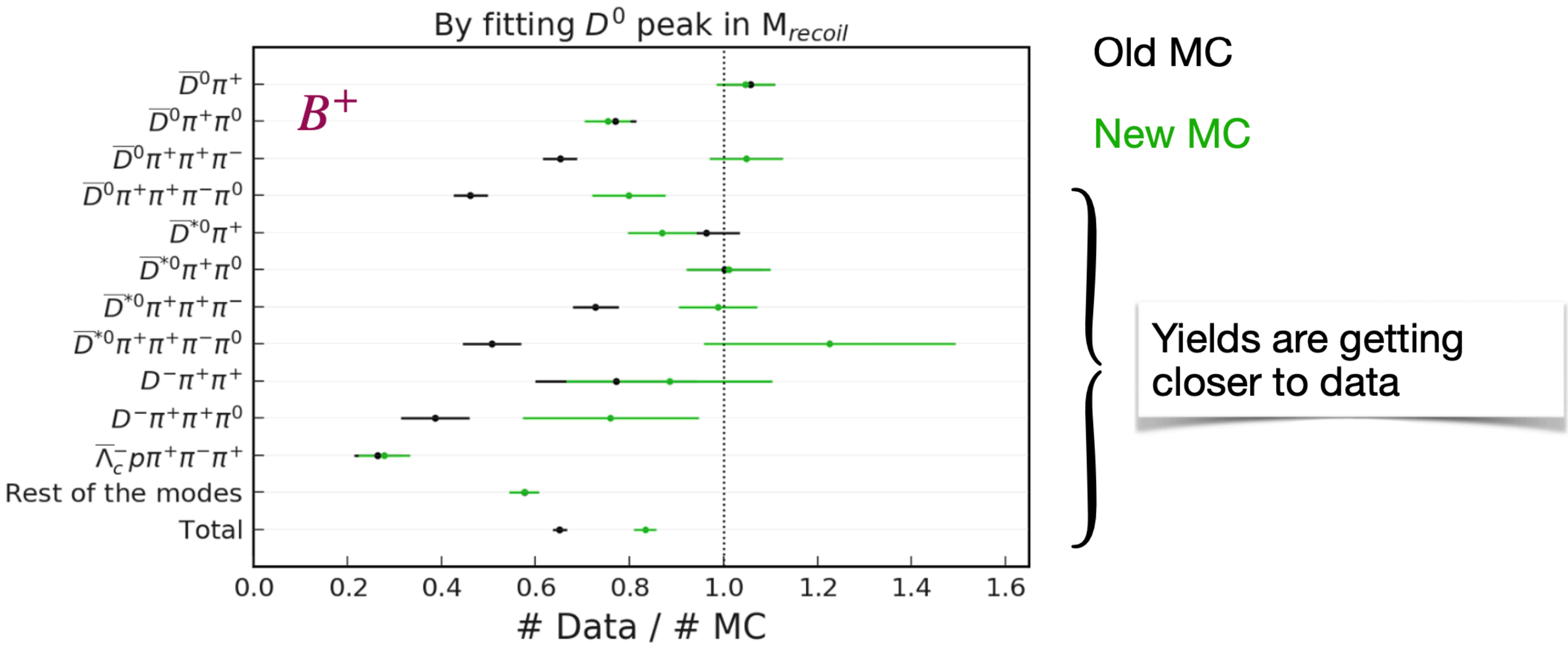
This not only improved the calibration factors of B-tagging, but also provided more realistic decay kinematics to train on, providing better purity.



# Updated CF

**Reminder**

*MC is first modified based on our best understanding.*
*Dπ sample is used to validate.*

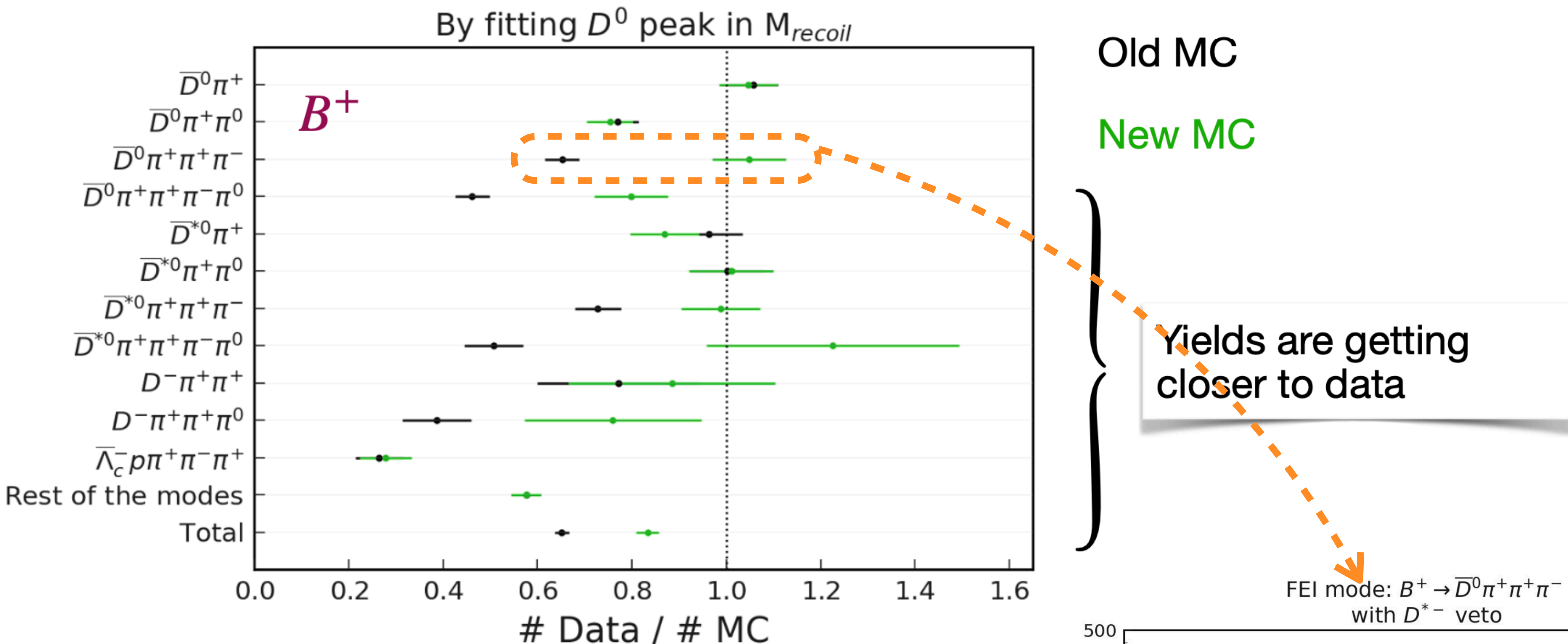


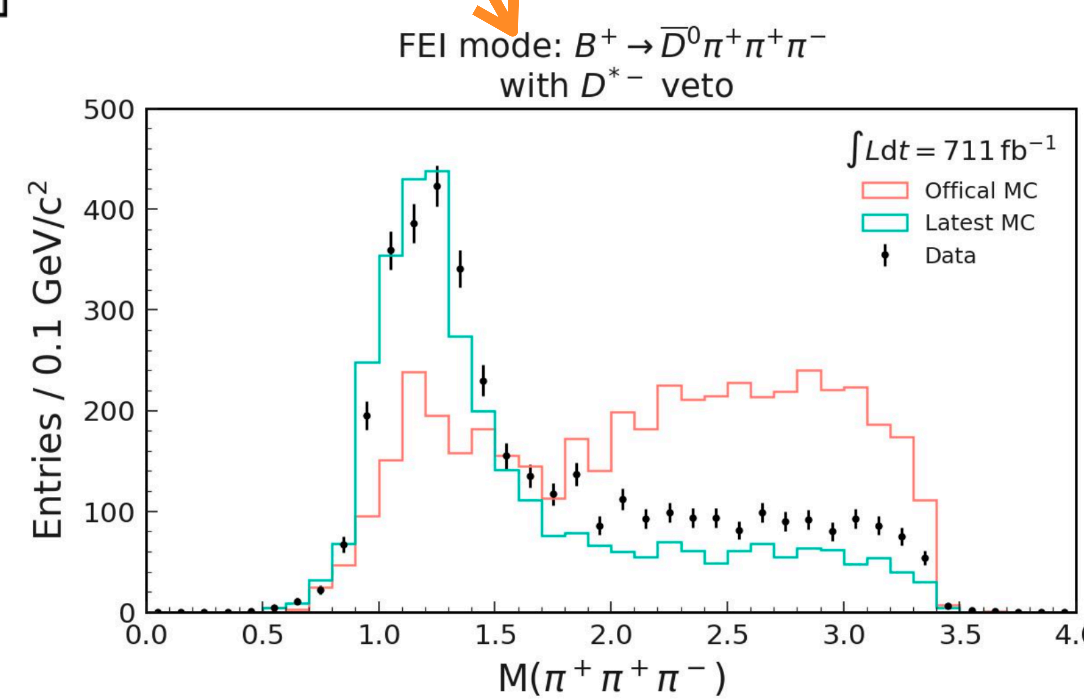

Overall calibration factor: 65% ➡ 83%
For the top 10 decay modes: 68% ➡️ 92%

# Updated CF

**Reminder**

*MC is first modified based on our best understanding.*
*Dπ sample is used to validate.*



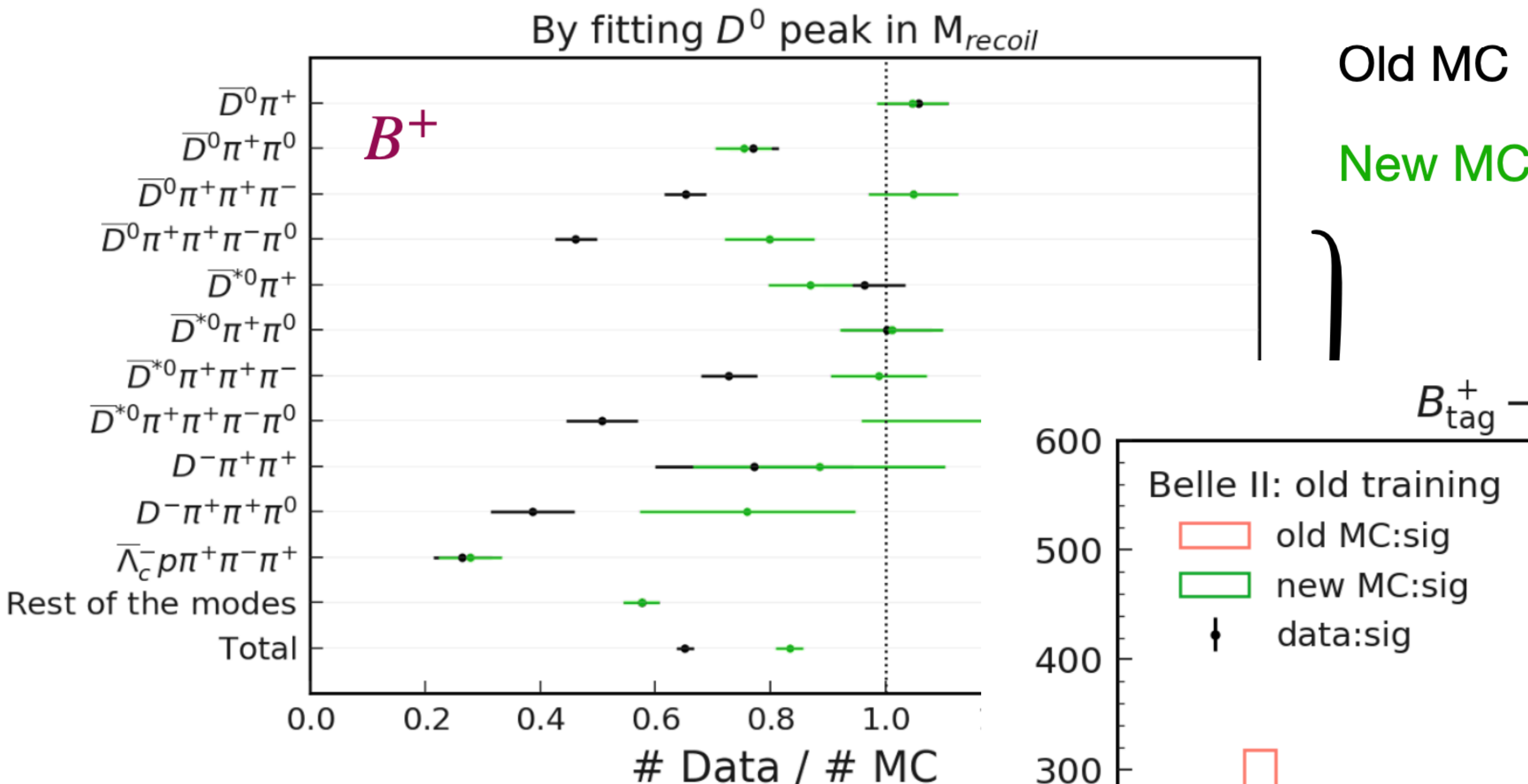



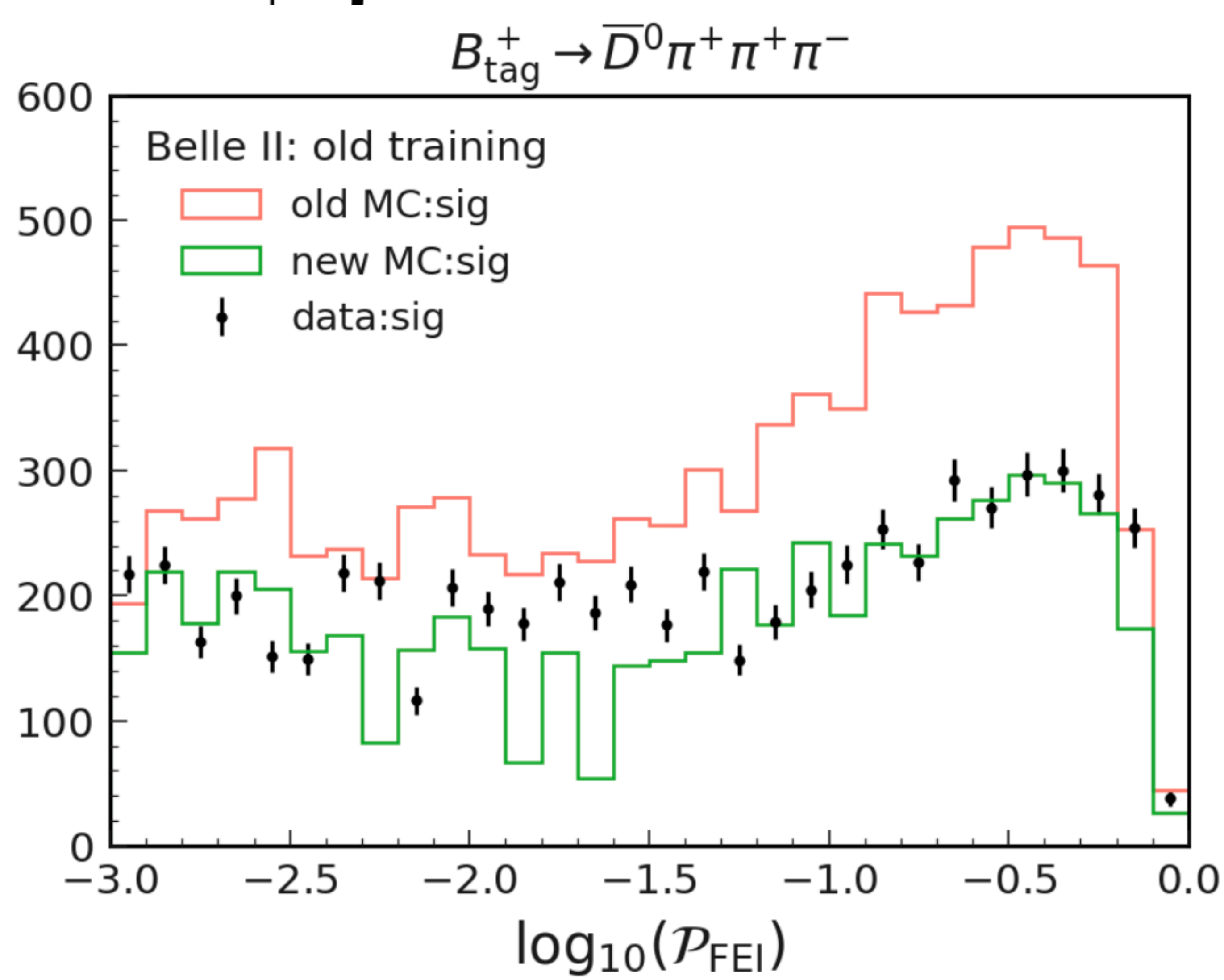

Overall calibration factor: 65% ➡ 83%
For the top 10 decay modes: 68% ➡️ 92%

# Updated CF

**Reminder**

*MC is first modified based on our best understanding.*
*$D\pi$ sample is used to validate.*



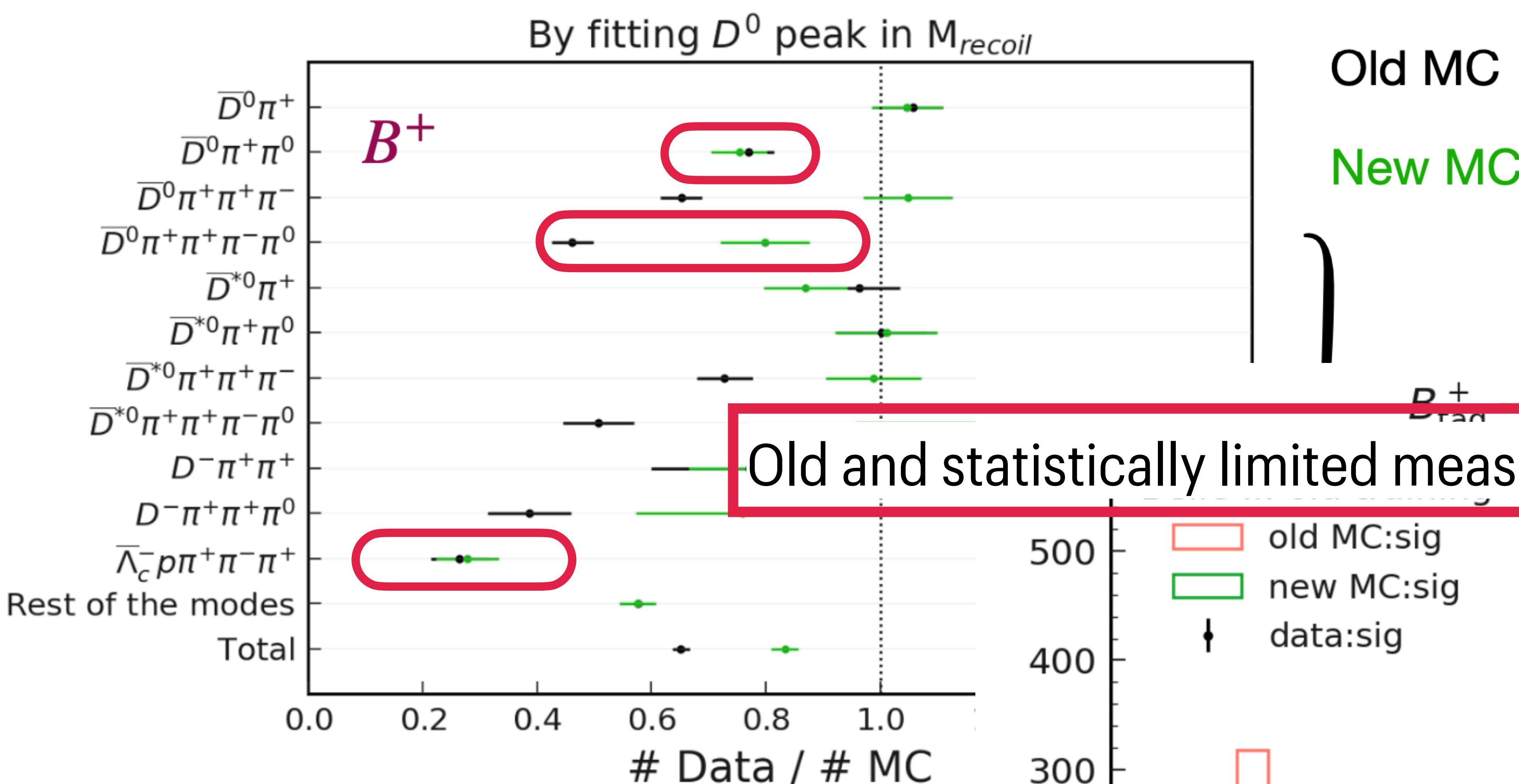




BDT output is closer to data

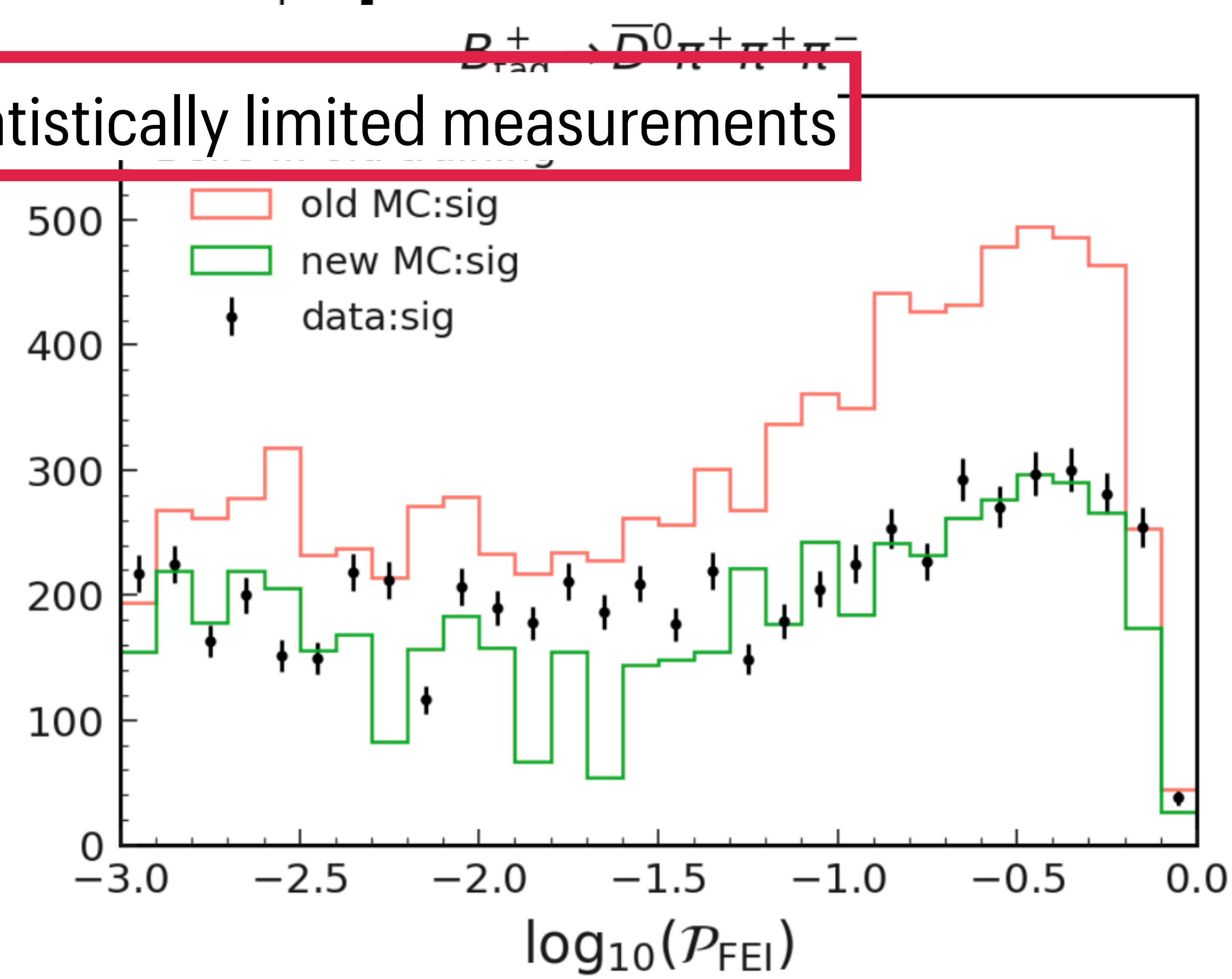
# Updated CF

**Reminder**

*MC is first modified based on our best understanding.*
*Dπ sample is used to validate.*




Old and statistically limited measurements




BDT output is closer to data

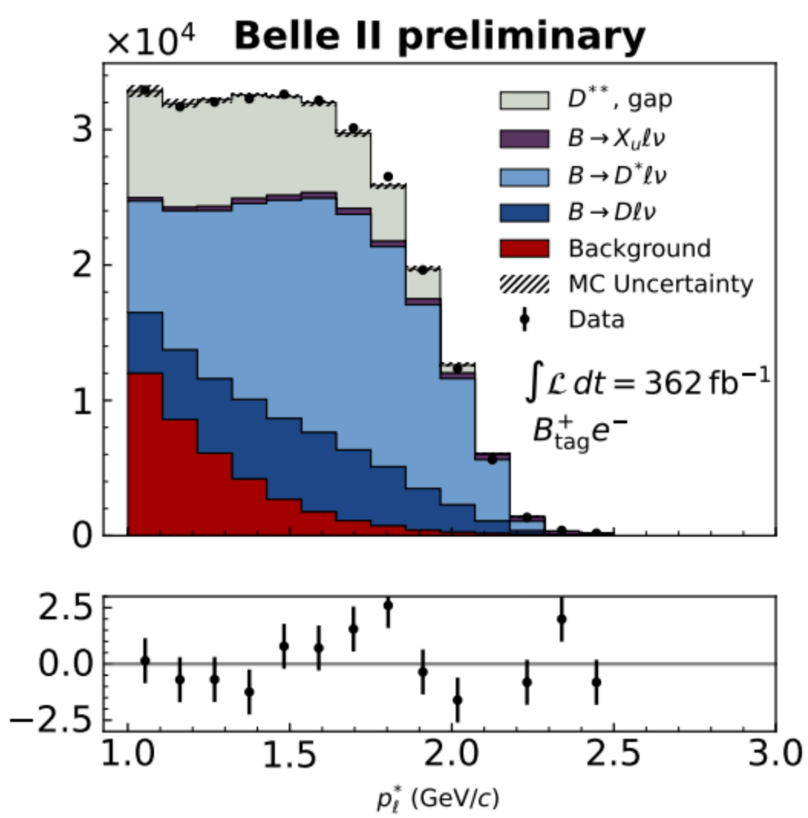
# Calibration

*Hadronic tag as an example*

- Two independent control samples are adopted as signal-sides to calibrate the data-MC difference of B tagging
- Calibration factors (CFs) are calculated as ratio of signal yields of data and MC

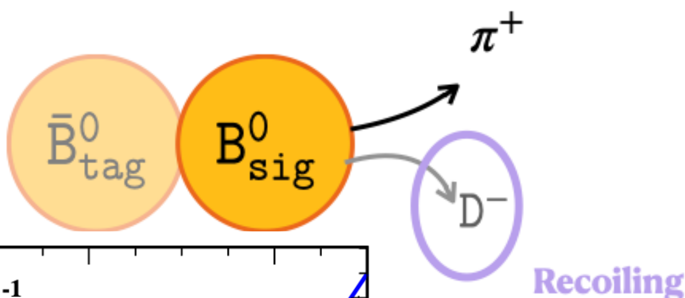
$\mathrm{B} \rightarrow \mathrm{D}^{(*)}\ell\nu$

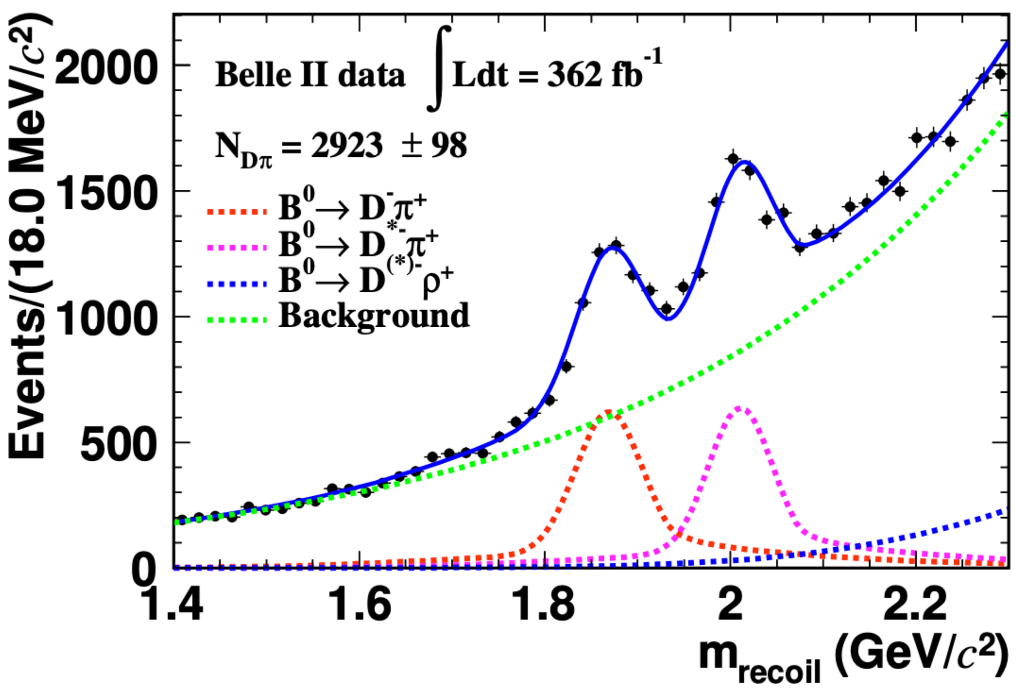
- Fit to lepton momentum in B rest frame: $\mathrm{p}_{\ell}^{*}$
- Yield: ~$10^5$, High statistics, low purity
- no peaking observable ~ dependent on background modeling




$\mathrm{B} \rightarrow \mathrm{D}^{(*)}\pi$

- Fit to recoiling system against $\mathrm{B}_{\mathrm{tag}}\pi$: $\mathrm{M}_{\mathrm{recoil/D}}$
- Yield: ~$10^4$, low statistics, high purity
- peaking observable ~ correct $\mathrm{B}_{\mathrm{tag}}$ events will peak at $\mathrm{D}^{(*)}$

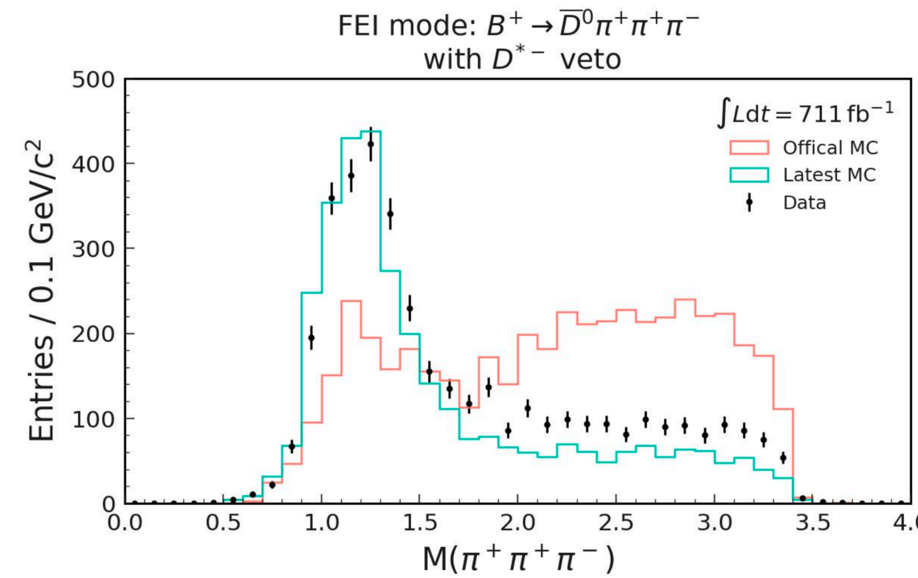


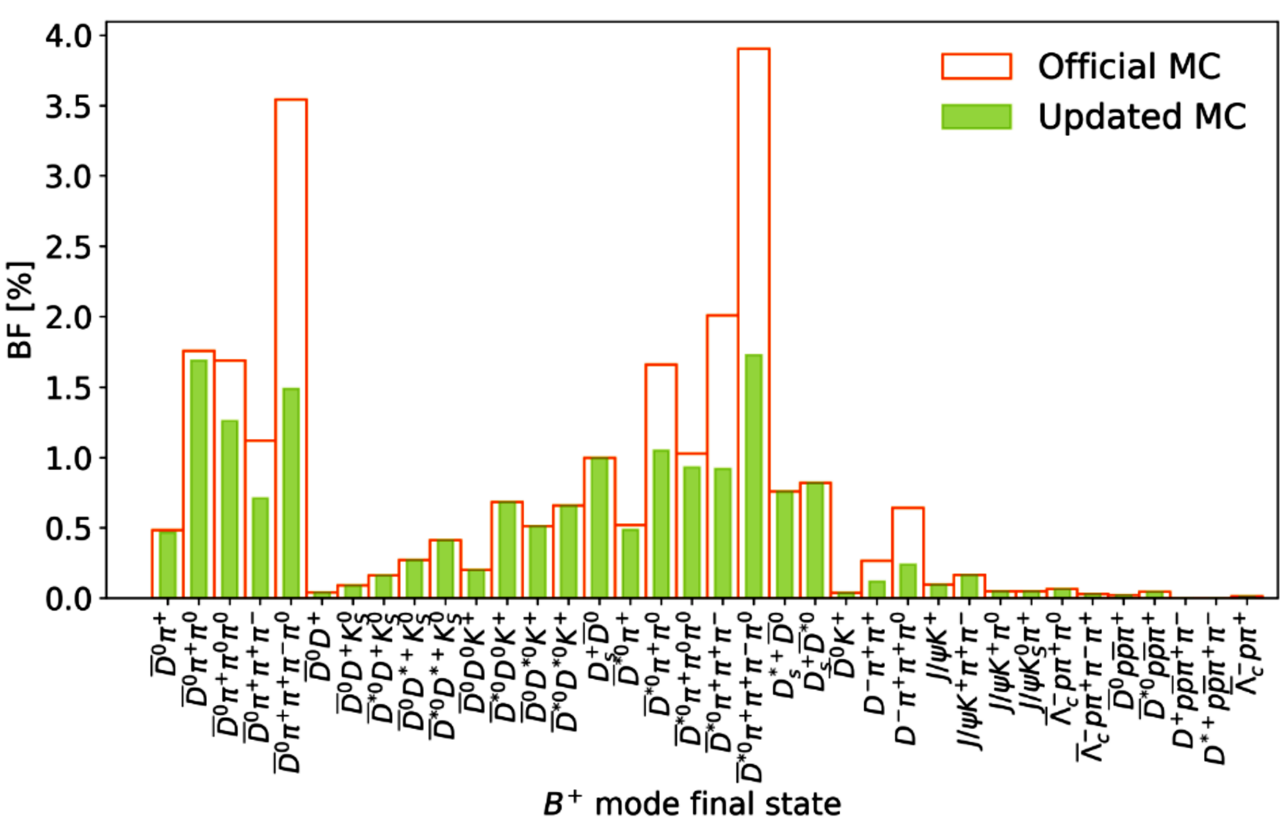



# Can we do better?

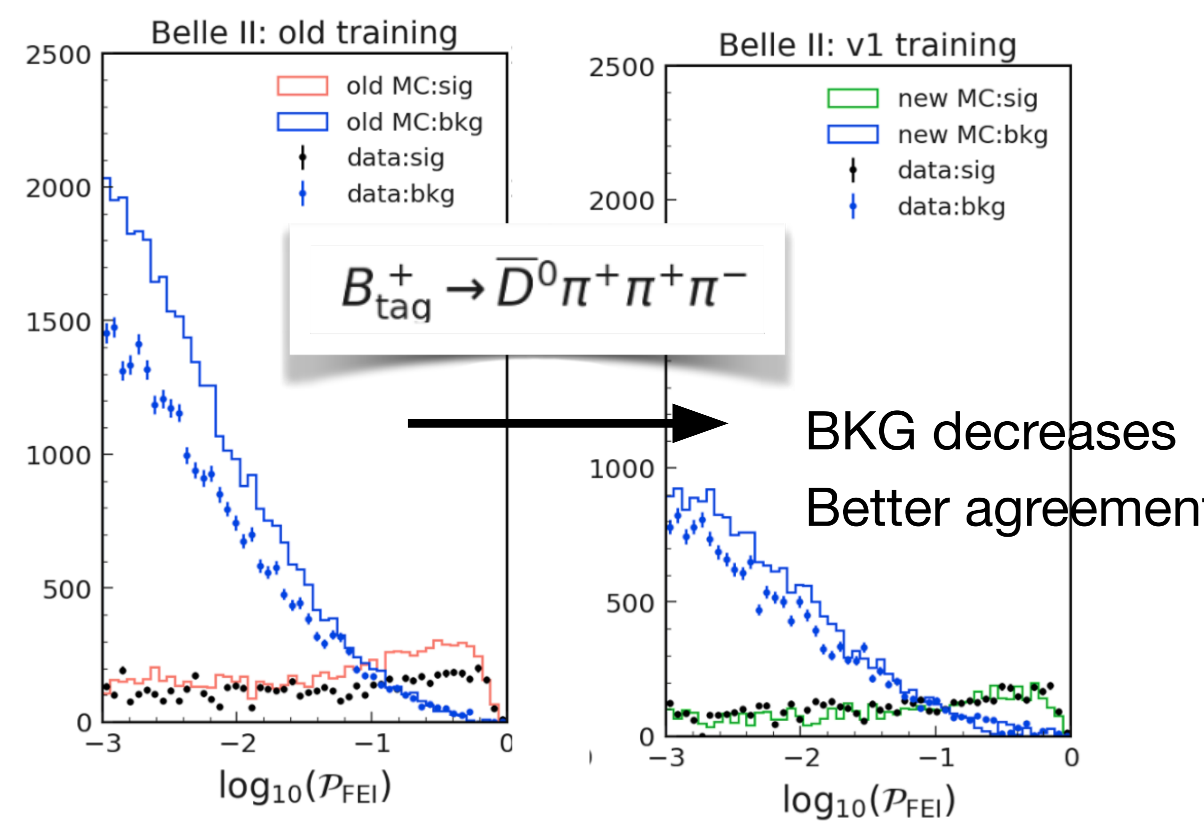
**Overall improvement of hadronic FEI**

- Updated decay model for 11 most efficient decay modes
  **0.65** → 0.81 : 25% ⬆️ in calibration factor
- Training with the new MC
  56% → 63% : 12% ⬆️ in purity
- Loosen the preselection and mass-constraint $\pi^0$
  0.93% → 1.13% : 21% ⬆️ in efficiency




**Training with new MC**



[ACAT_2022_proceeding]

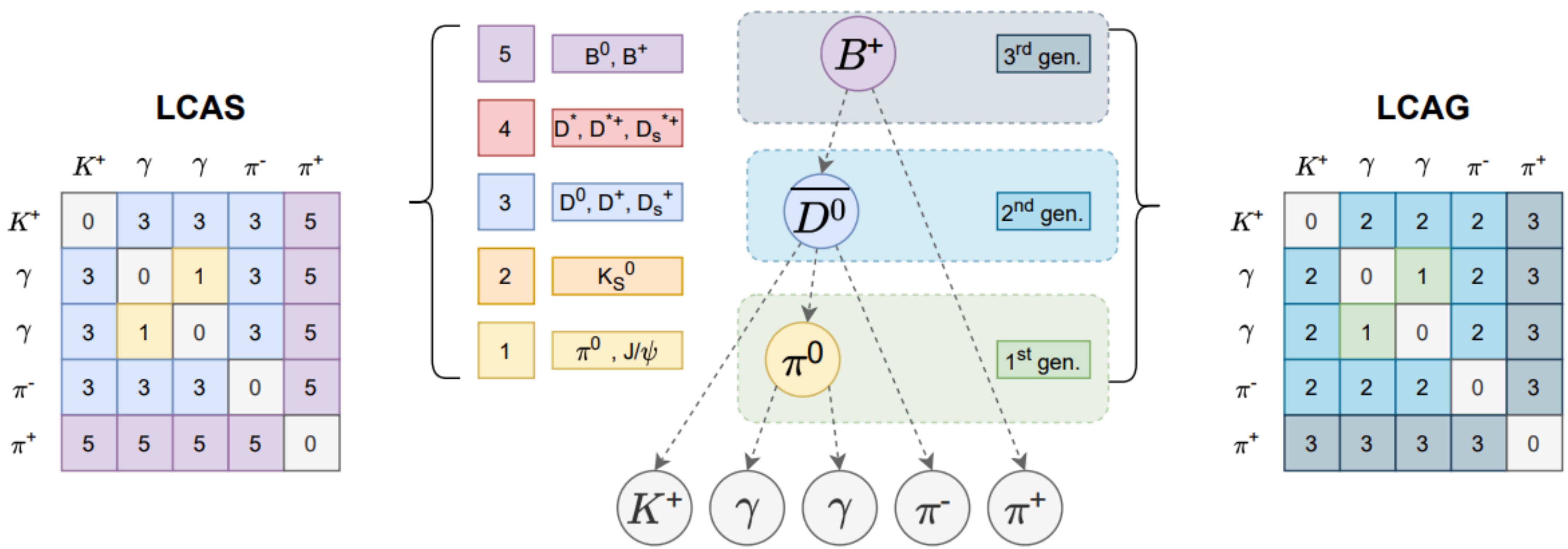
# Graph-based Full Event Interpretation

**Full Event Interpretation** FEI: decay modes are hard-coded

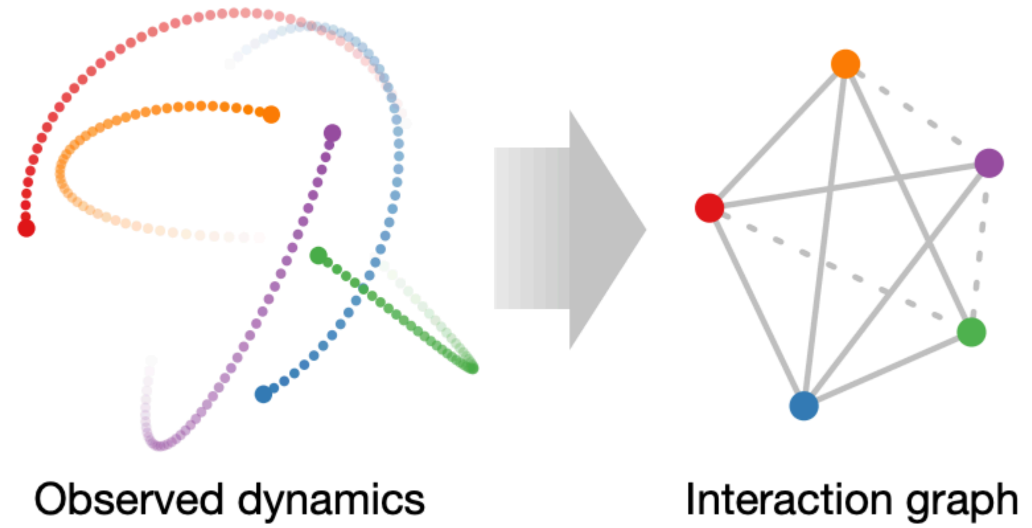
Particle decays are naturally described by rooted directed acyclic tree graphs

Increase the tagging efficiency

- Goal: develop **graph-based Full Event Interpretation (graFEI)** to **inclusively** reconstruct tag B meson
- Proof of concept: Learning tree structures from leaves for particle decay reconstruction, *Kahn et al 2022*

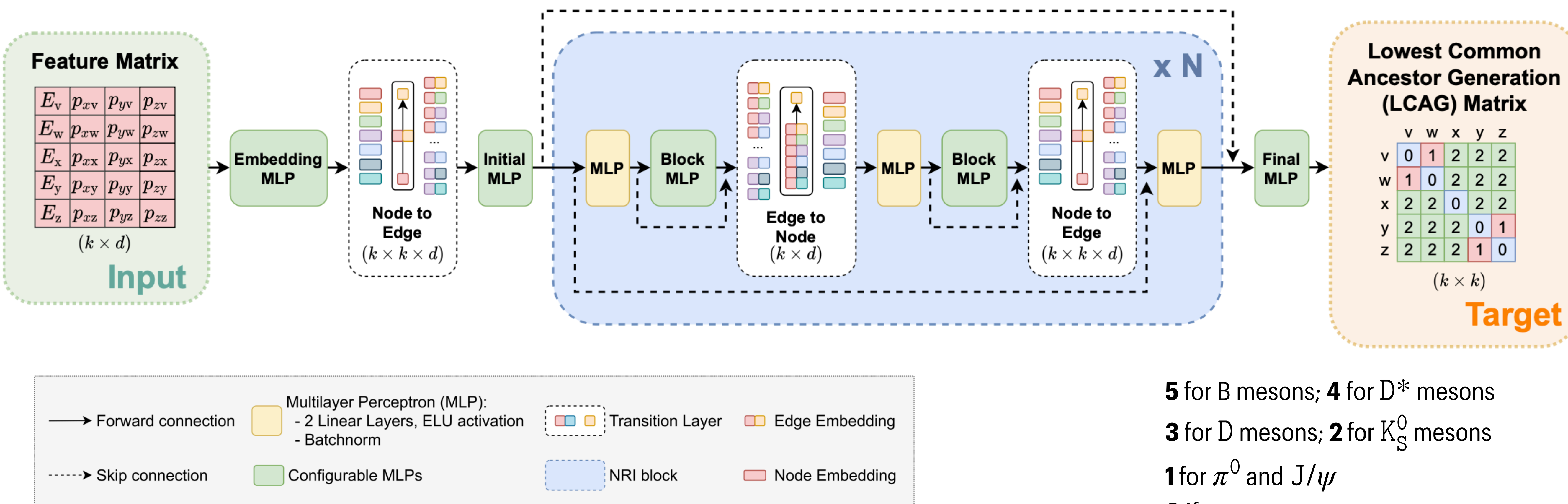


# Training of **graFEI**




- Neural Relational Inference (NRI) model [arXiv:1802.04687]
- Dataset generated with Phasespace library
- 4-momentum used as input feature
- Average 47.7 % perfectly predicted LCAG on Phasespace dataset (60.9 % for decays with up to 10 leaves, 94.2 % up to 6 leaves)

final-state particles




**5** for B mesons; **4** for $D^*$ mesons

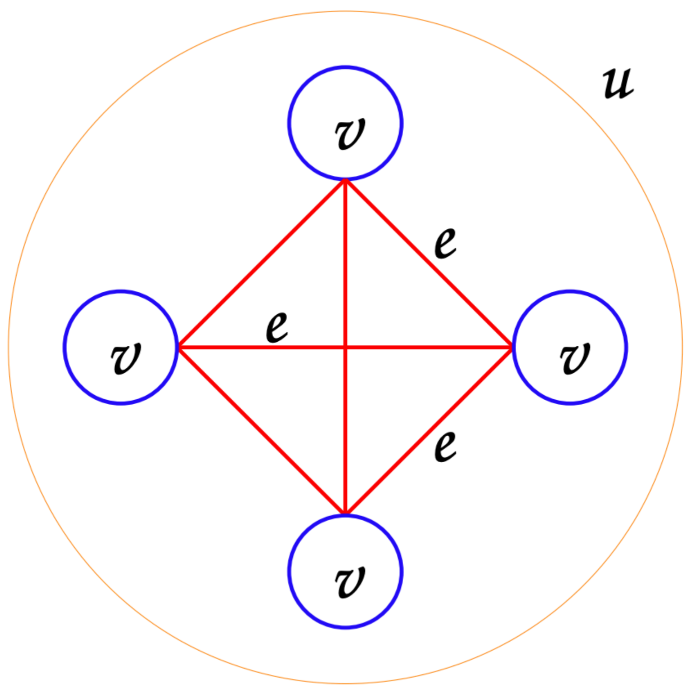
**3** for D mesons; **2** for $K_S^0$ mesons

**1** for $\pi^0$ and $J/\psi$

**0** if a common ancestor can not be identified

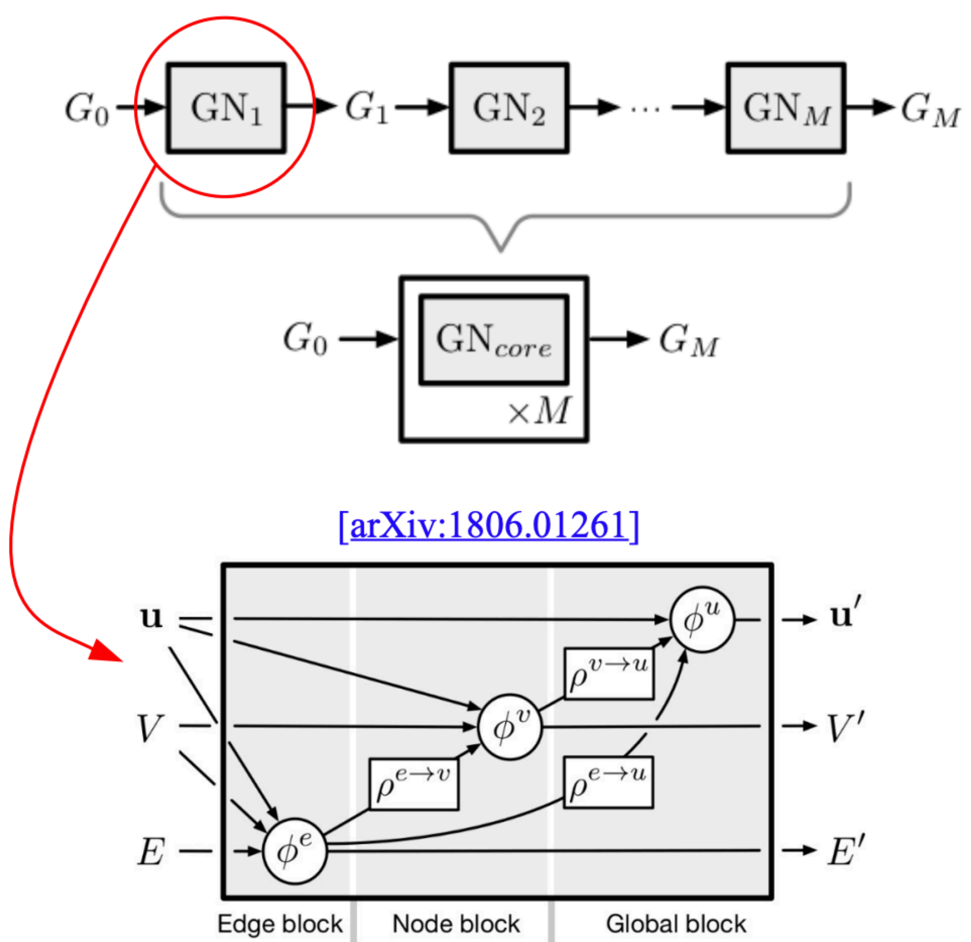
# Model of **graFEI**

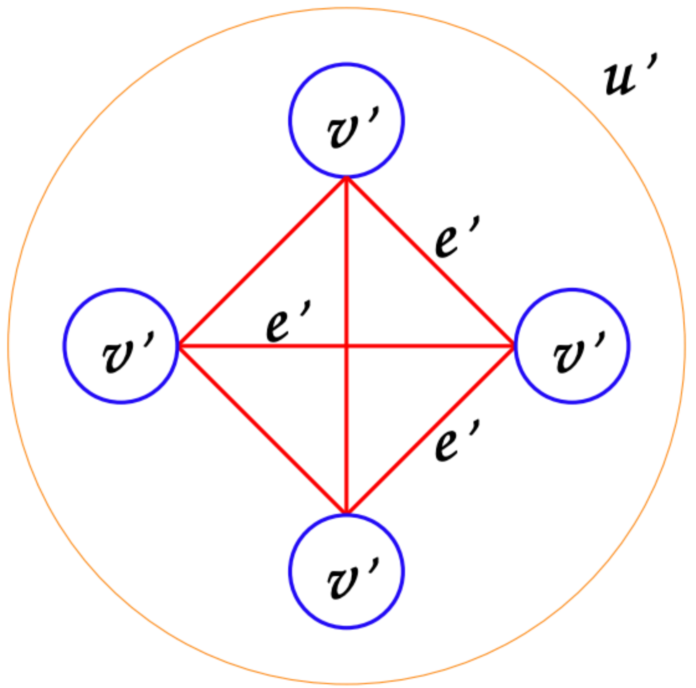
- We input a fully connected graph, output graph has same structure with updated attributes
- **LCA matrix** predicted as training target via edge labels classification, **particle IDs** via node labels classification



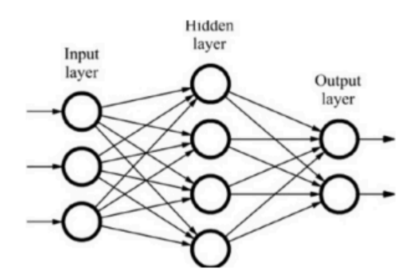

u = global features
v = node features
e = edge features

input is a fully-connected graph where nodes represent the final-state particles detected by the experiment.




$\phi =$




$$\rho = \frac{1}{N} \sum_i x^i$$