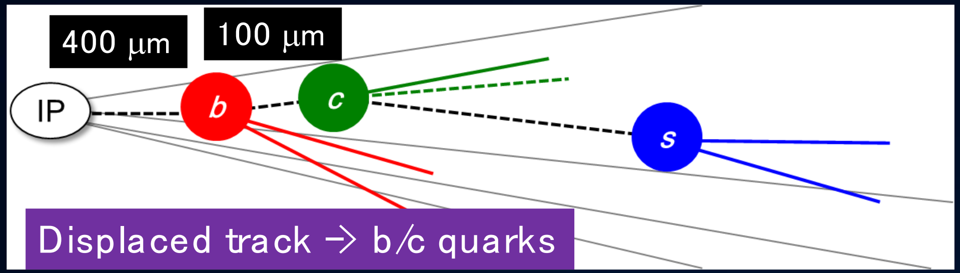
Application of Particle Transformer in Quark Flavor Tagging

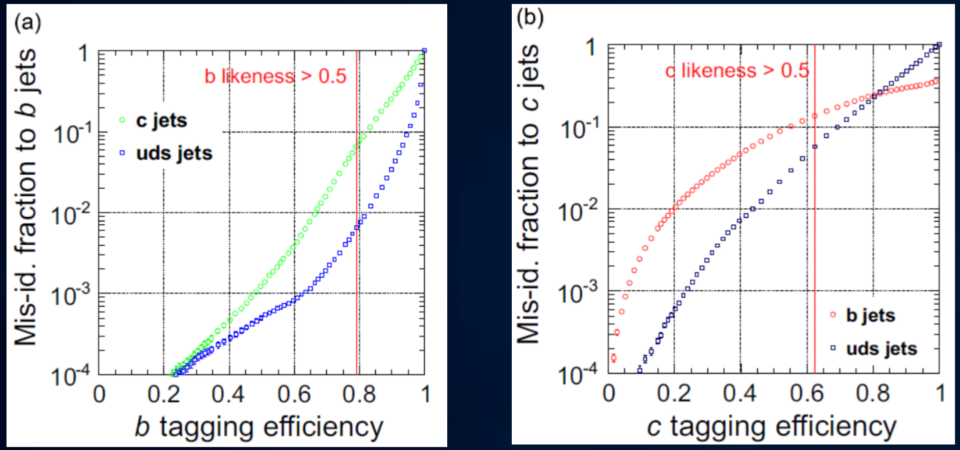
Presenter: Lai Gui

Date: 2023-08-01

Background

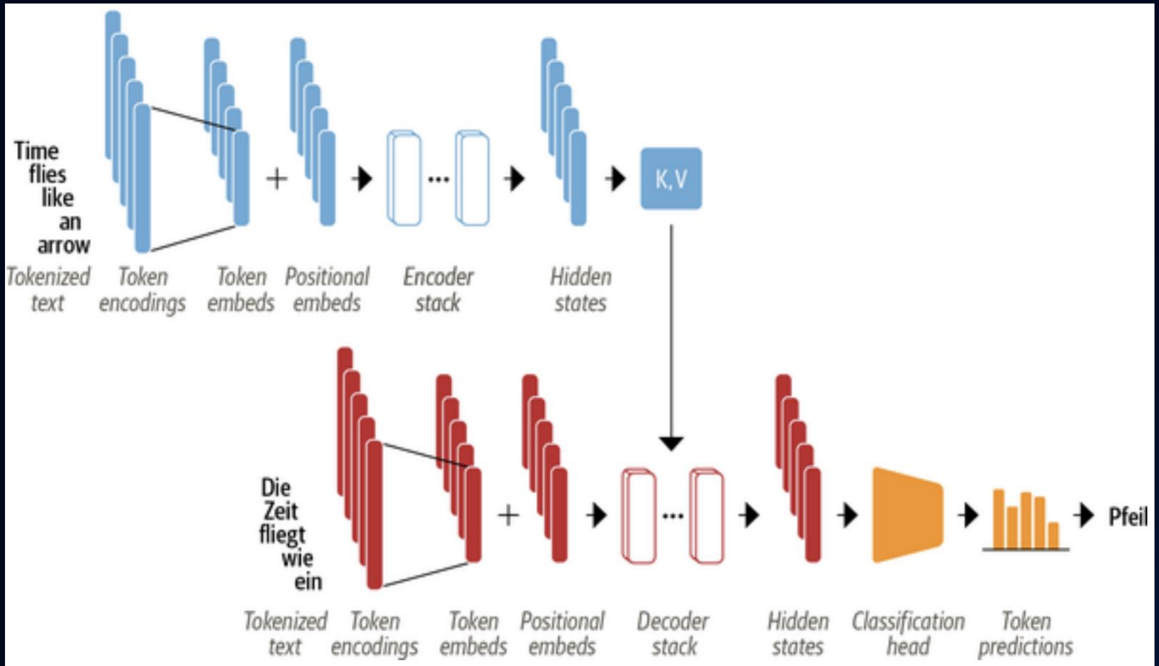
- Precise measurements instrumentation and reconstruction software are essential for the ILC programme.
- Various frameworks have been developed for jet flavour identification.
- LCFIPlus (published 2013)[1] was successful in vertex finding, jet clustering and flavour tagging.
- Reached a reasonable performance of:
  - b-tag: 80% eff., 10% c / 1% uds acceptance;
  - c-tag: 50% eff., 10% b / 2% uds acceptance.

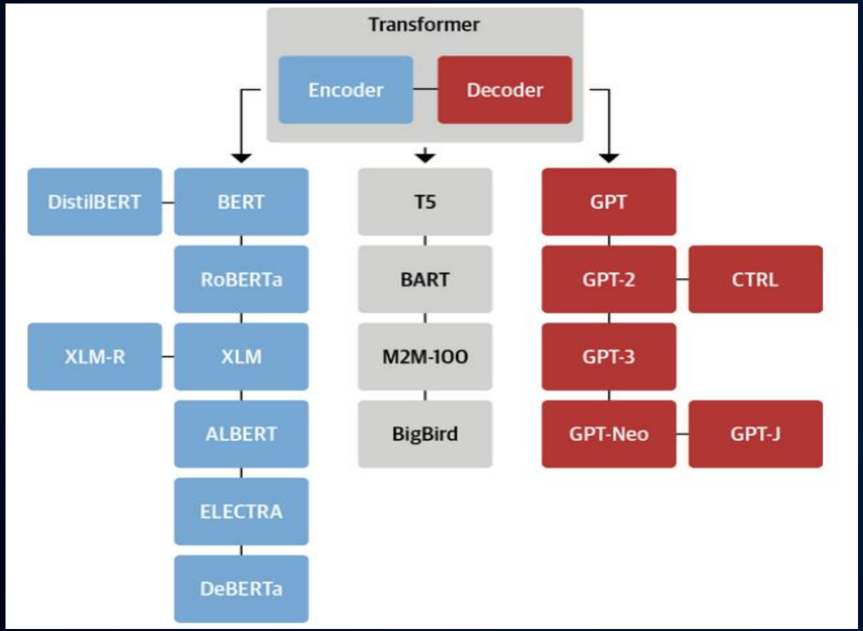




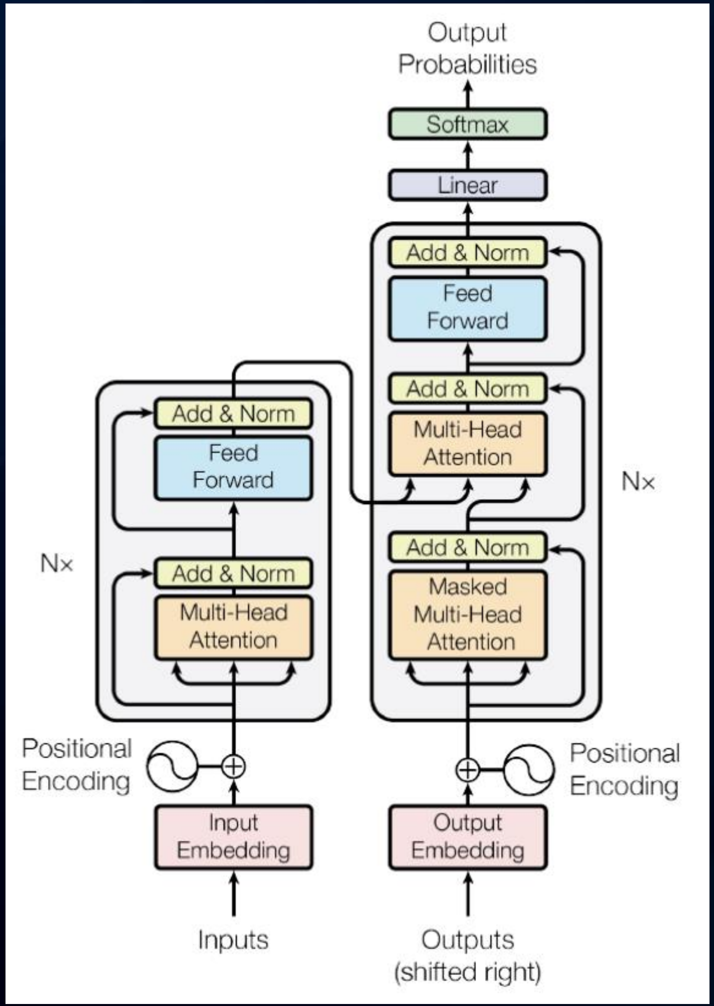
Transformer

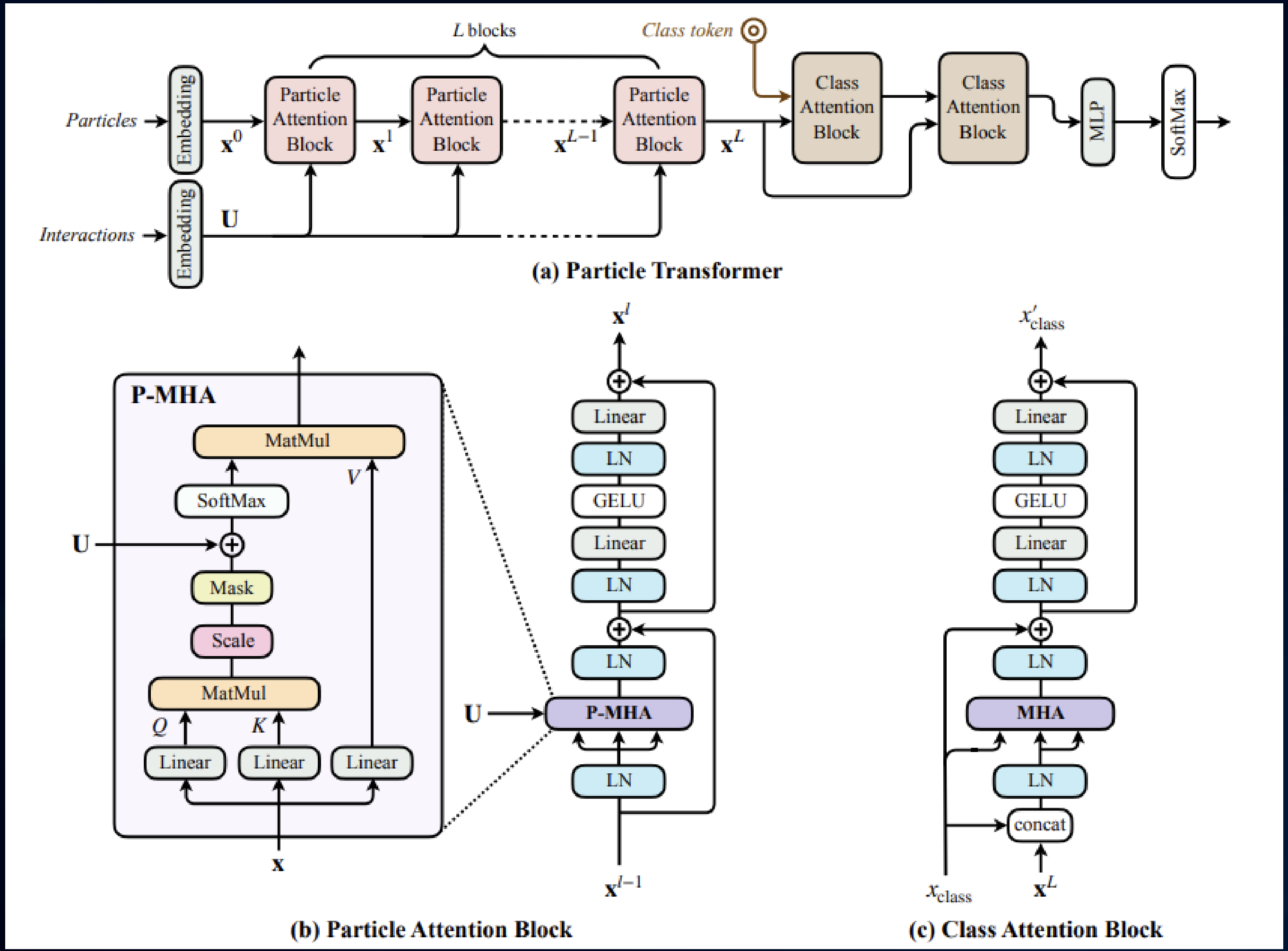
- Input is converted by the *Encoder* into a sequence of *hidden states* that is consisted of *Token Embeds* and *Positional Embeds*.
- This *hidden state* is then processed through layers of *Self-Attention* and *Feed-Forward* neural networks.
- The *Self-Attention* mechanism calculates the relative importance of each token relative to all the other tokens in the input sequence (Outperforms traditional RNN and CNN).
- The *Decoder* then outputs one token at a time, and this token is then added to the input to generate the next context iteratively.





Comparison between regular Transformer and Particle Transformer

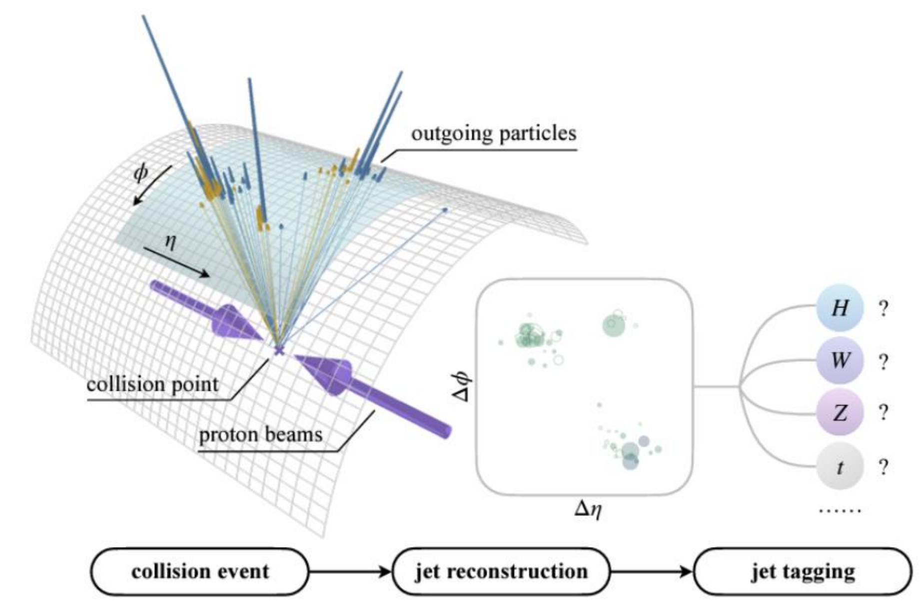




Note: MHA – MultiHeadAttention
P-MHA – Augmented version of MHA by Particle Transformer that involves Interactions Embeddings instead of Positional Embeddings

Particle Transformer (ParT)

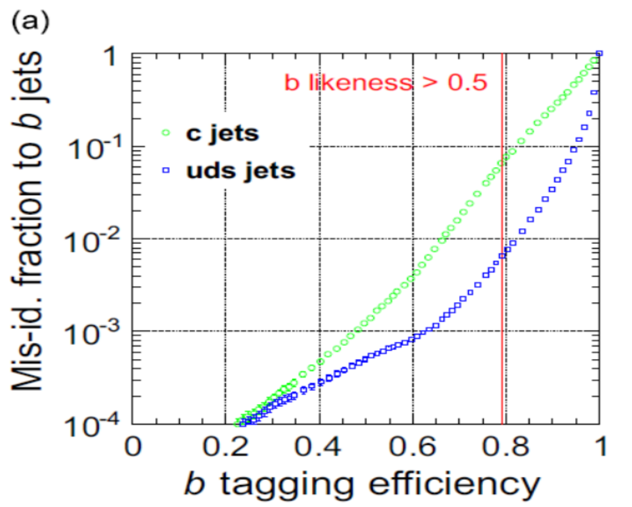
- A new Transformer-based architecture for Jet tagging, published in 2022[2].
- It analyses the readings collected after collision events to reconstruct jets. (Illustration of CERN LHC p-p collisions)
- Surpasses the performance of previous architectures by a large margin. Values below are rejection ratio (inverse of acceptance ratio).

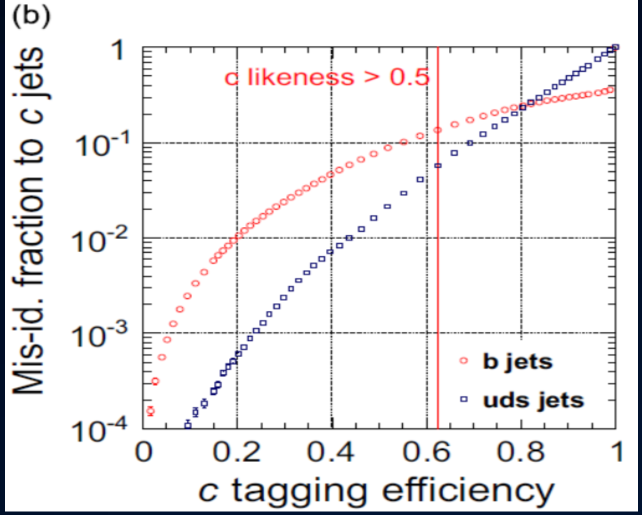


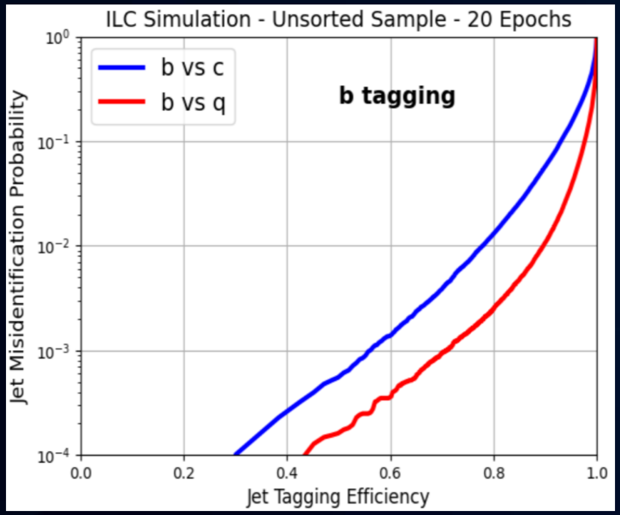
	All classes Accuracy	All classes AUC	$H \to b\bar{b}$ Rej$_{50\%}$	$H \to c\bar{c}$ Rej$_{50\%}$	$H \to gg$ Rej$_{50\%}$	$H \to 4q$ Rej$_{50\%}$	$H \to \ell\nu qq'$ Rej$_{99\%}$	$t \to bqq'$ Rej$_{50\%}$	$t \to b\ell\nu$ Rej$_{99.5\%}$	$W \to qq'$ Rej$_{50\%}$	$Z \to q\bar{q}$ Rej$_{50\%}$
PFN	0.772	0.9714	2924	841	75	198	265	797	721	189	159
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	2304	241	204
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT	**0.861**	**0.9877**	**10638**	**4149**	**123**	**1864**	**5479**	**32787**	**15873**	**543**	**402**
ParT (plain)	0.849	0.9859	9569	2911	112	1185	3868	17699	12987	384	311

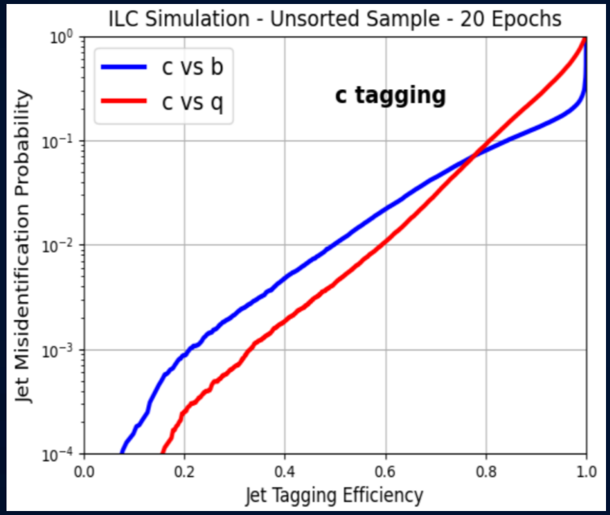
Application of ParT to ILC data

- Jet tagging performance is greatly improved by ParT immediately:
  - b-tag: 80% eff., 1.29% c / 0.247% uds acceptance;
  - c-tag: 50% eff., 1.02% b / 0.428% uds acceptance.
- The performance is improved by 4.05 – 9.80 times compared to LCFIPlus with the same set of data.
- Can this performance to be further improved?



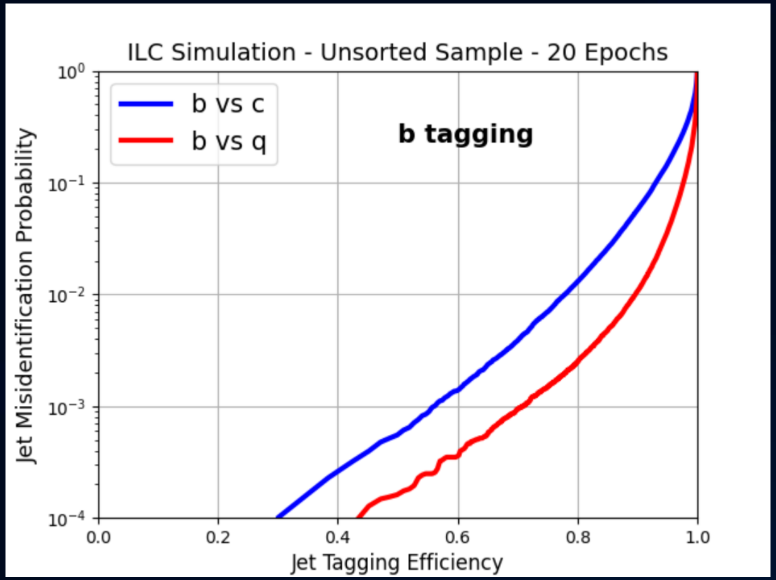


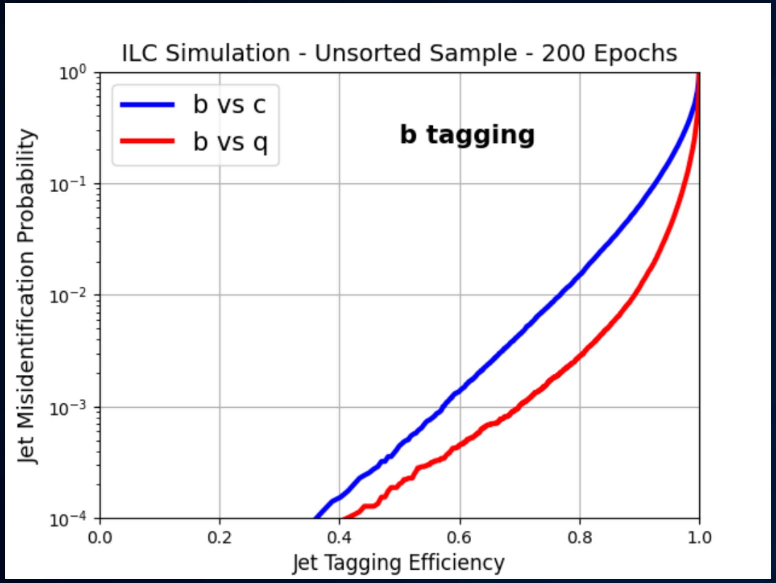




Training parameters - epochs

- 20-epoch training takes 3 hours
- 200-epoch training takes 30 hours
- No significant improvement in tagging efficiency at 0.6 or 0.8 efficiency
- Difference towards 0.4 efficiency might due to random fluctuation



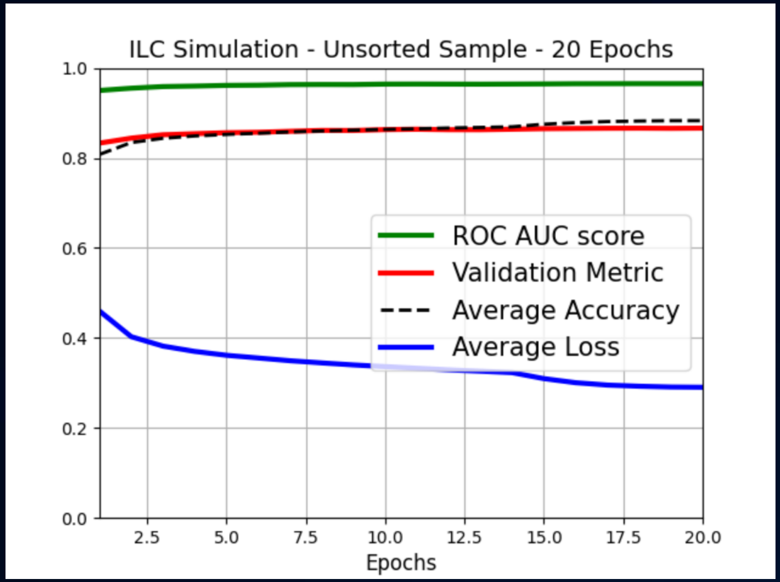


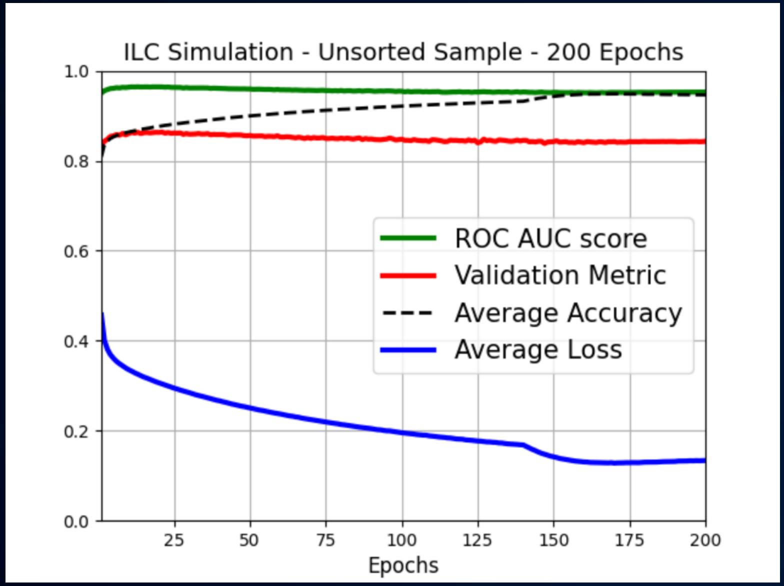
Training parameters - epochs

- Both ROC AUC score and Validation Metric reaches a maximum around 20 epochs

- The Average Accuracy and Average Loss still witness an improvement, but not significant in the analysis result – overtraining after 20 epochs

- Hence 20 epochs of training is selected to be the standard for future training

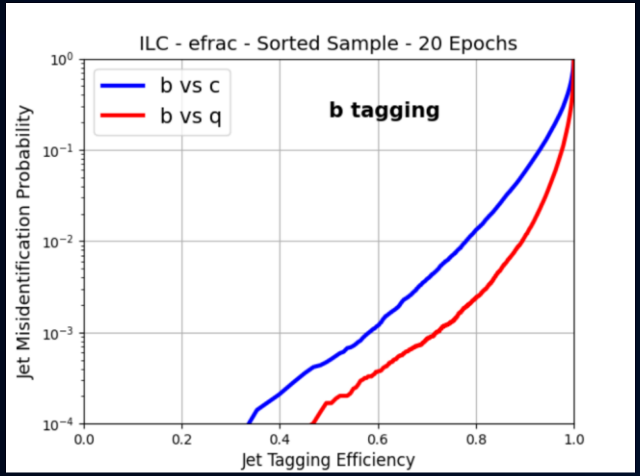


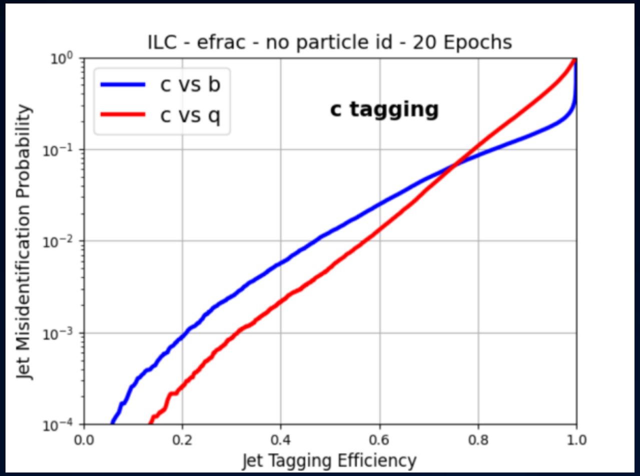


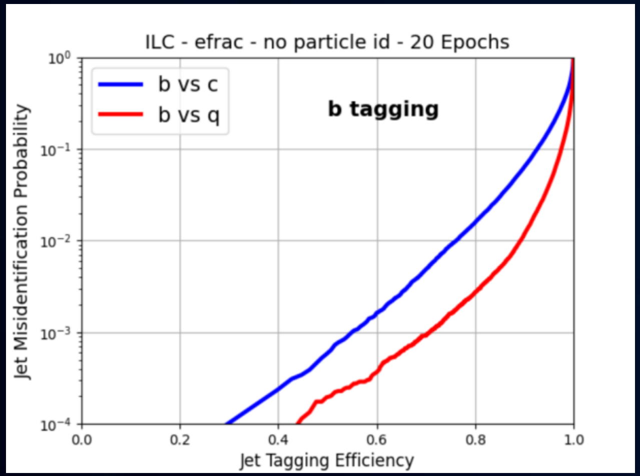
Training parameters – Particle ID

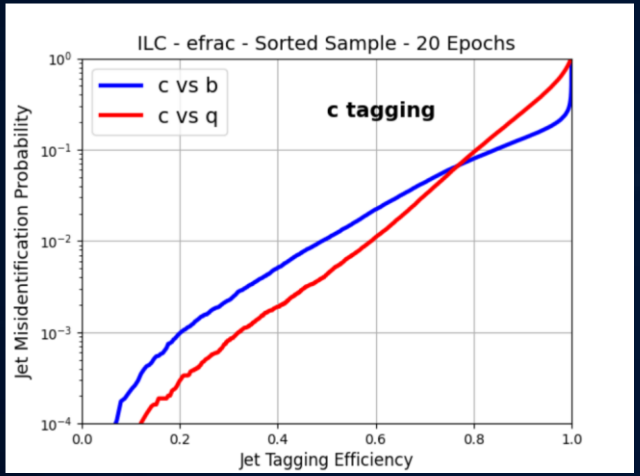
- With Particle ID:
  - b-tag: 80% eff., 1.32% c / 0.237% uds acceptance;
  - c-tag: 50% eff., 1.06% b / 0.429% uds acceptance.
- Without Particle ID:
  - b-tag: 80% eff., 1.57% c / 0.272% uds acceptance;
  - c-tag: 50% eff., 1.24% b / 0.507% uds acceptance.

- Particle ID improves tagging performance





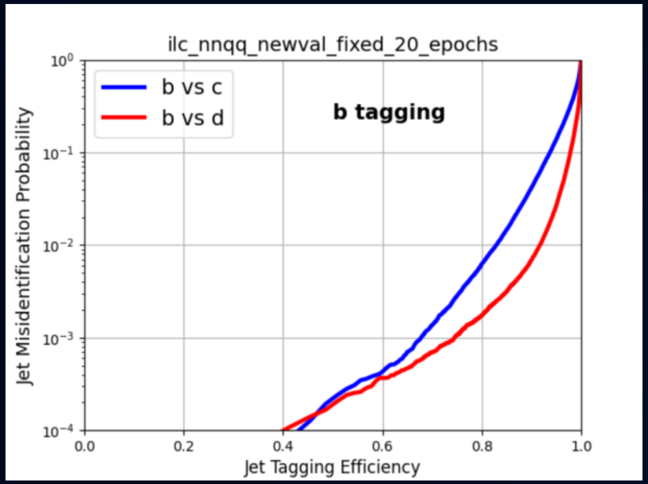


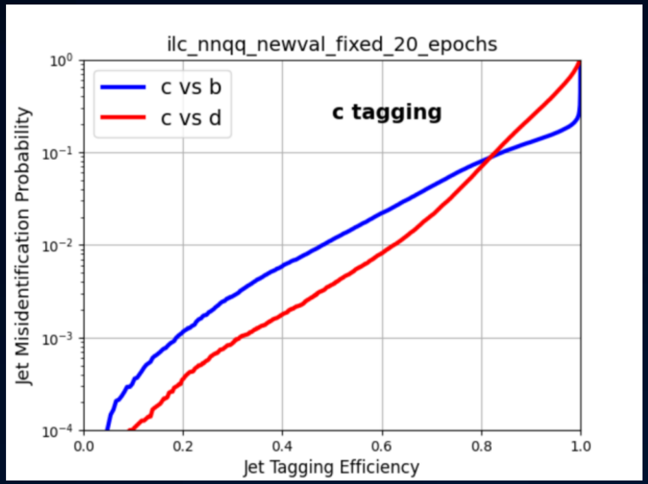


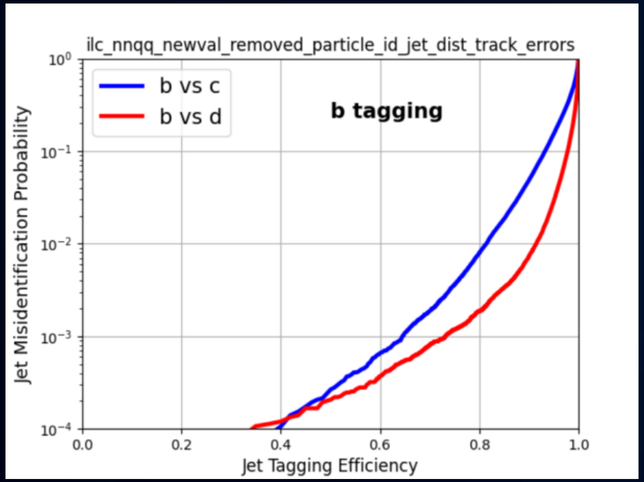
Training parameters – Jet Distance Values and Track Errors

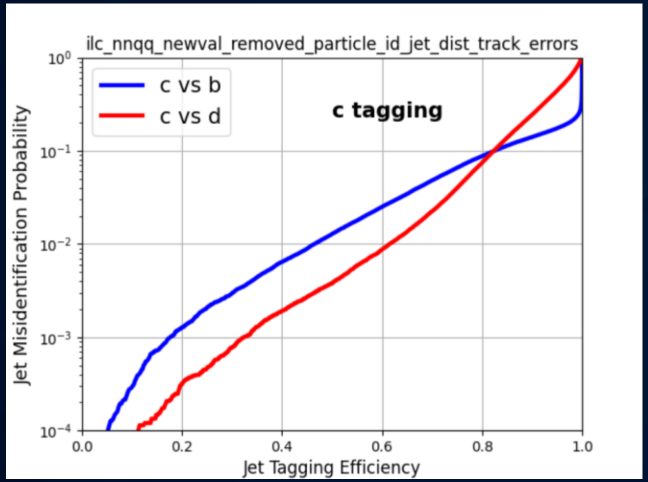
- With Jet Distance Values and Track Errors :
  - b-tag: 80% eff., 0.623% c / 0.174% uds acceptance;
  - c-tag: 50% eff., 1.14% b / 0.372% uds acceptance.
- Without Jet Distance Values and Track Errors :
  - b-tag: 80% eff., 0.794% c / 0.187% uds acceptance;
  - c-tag: 50% eff., 1.28% b / 0.380% uds acceptance.

- Jet Distance Values and Track Errors improves tagging performance





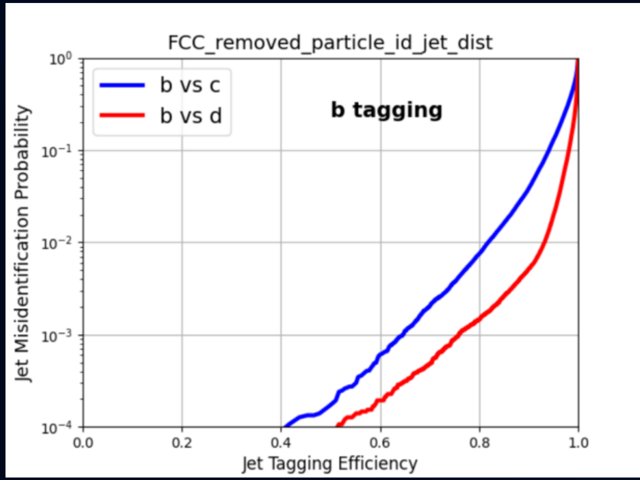


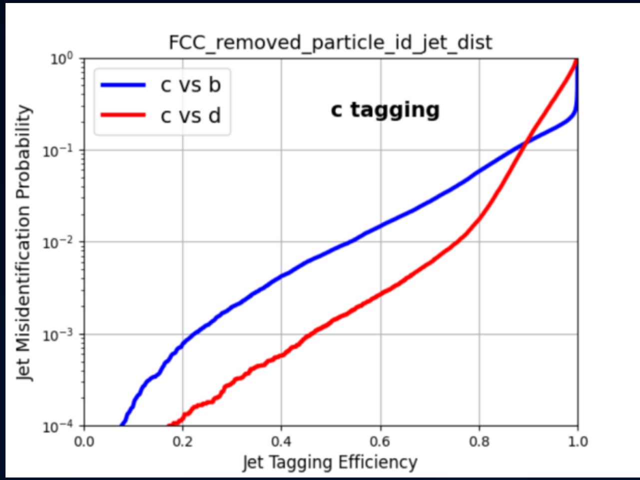


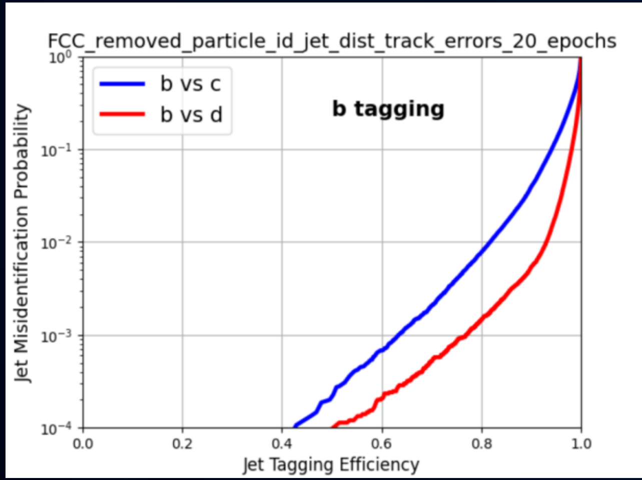
Training parameters – Track Errors

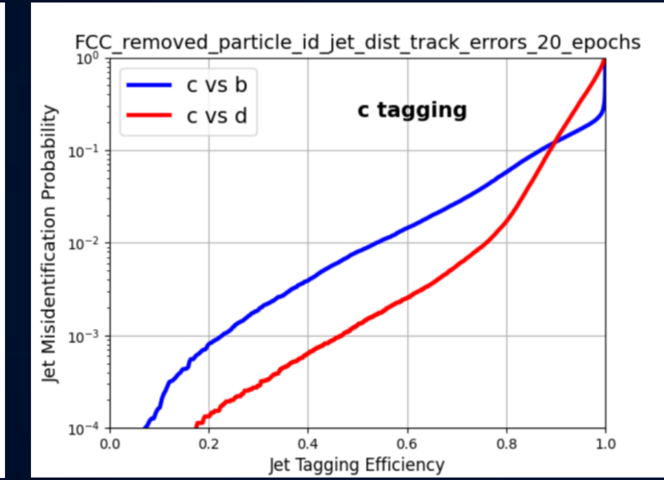
- With Track Errors:
  - b-tag: 80% eff., 0.747% c / 0.145% uds acceptance;
  - c-tag: 50% eff., 0.797% b / 0.131% uds acceptance.
- Without Track Errors:
  - b-tag: 80% eff., 0.773% c / 0.146% uds acceptance;
  - c-tag: 50% eff., 0.799% b / 0.130% uds acceptance.

- Track Errors does not affect tagging performance significantly.









Training parameters – More to be confirmed

Reference List

[1] https://doi.org/10.1016/j.nima.2015.11.054

[2] https://arxiv.org/abs/2202.03772