Application of Particle Transformer to quark flavor tagging in the ILC project

2024.07.09 Risako TAGAMI[1], Taikan SUEHARA[2], Masaya ISHINO[2]

1:UTokyo, 2:UTokyo ICEPP

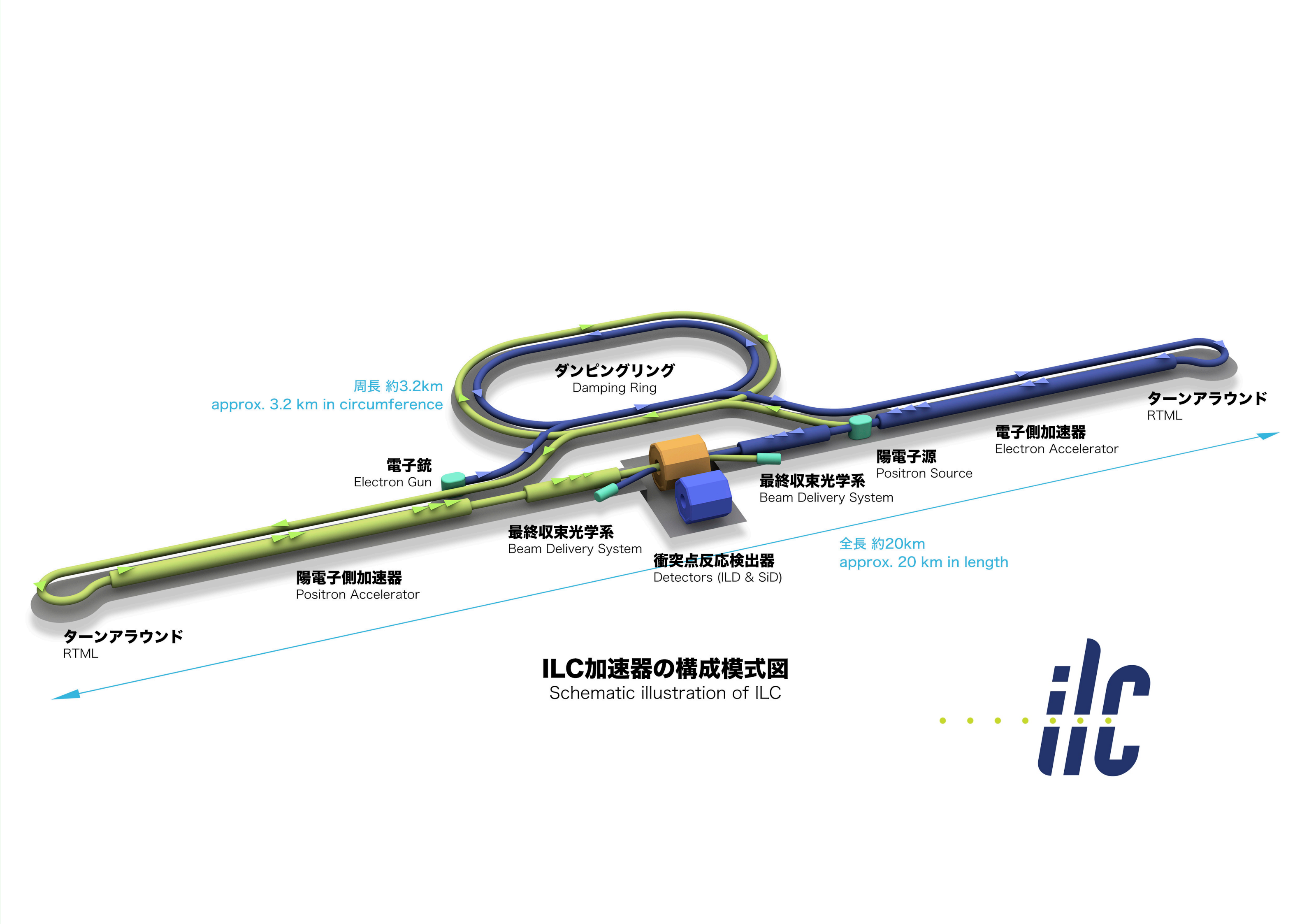
Higgs Factories

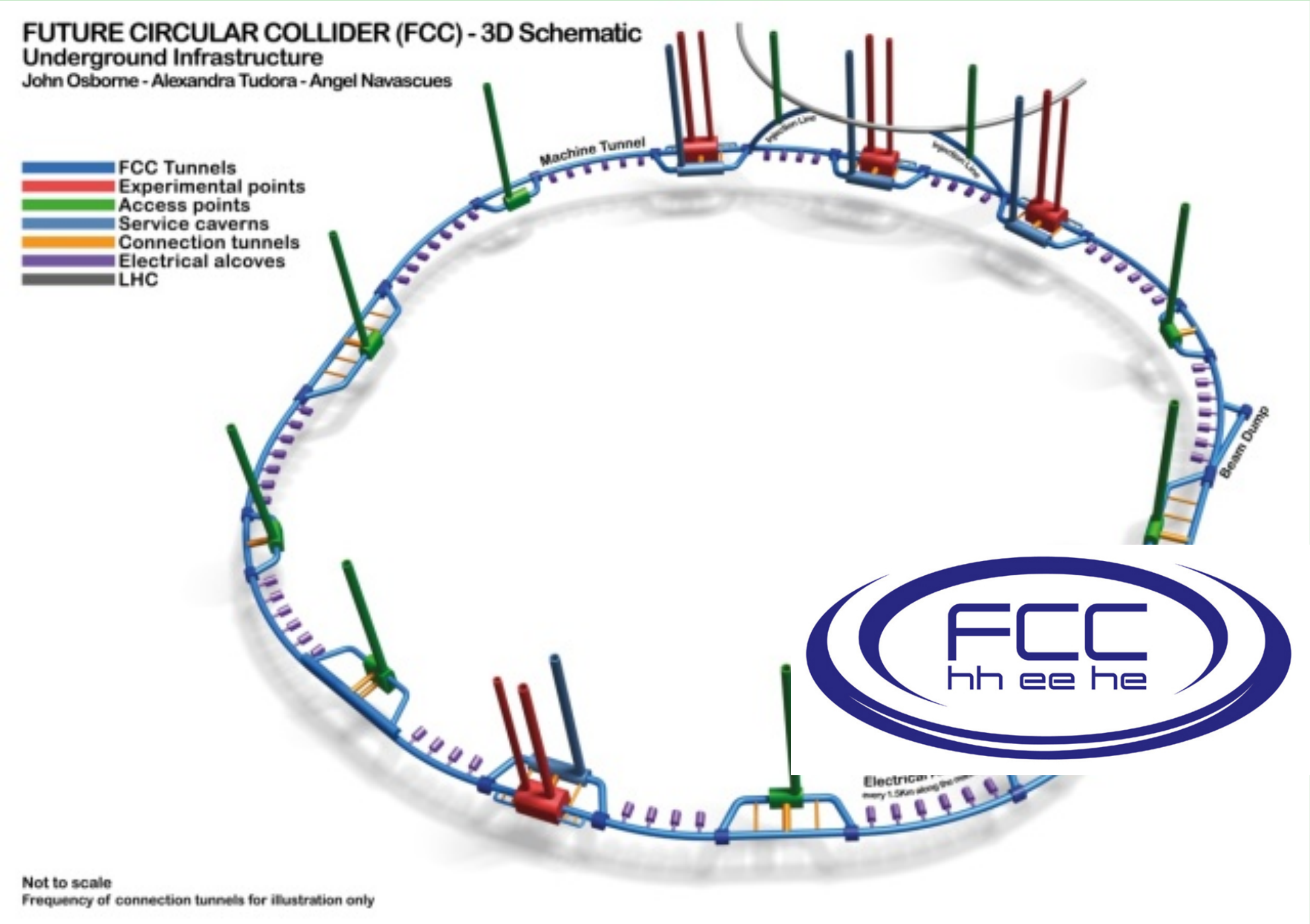
- The search for new physics by **precise measurement of Higgs** is expected.

  ➡The consensus among particle physicists is to build the Higgs Factory as the next-generation accelerator.（cf. European Strategy, Snowmass）

- There are several e+e- Higgs Factories currently under consideration.

  - ILC（Japan）
  - FCCee（CERN）
  - …





Physics fo Higgs Particle and flavor tagging

With more precise measurements of Higgs, the effects of SUSY and many other new TeV physics models can be seen.

- To precisely measure the coupling constants such as H→bb, cc, gg, ss, etc., the performance of **flavor tagging** needs to be improved.

$$\kappa = \frac{g_x}{g_x^{SM}} = 1 + \Delta\kappa$$

$$\Delta\kappa \sim \mathcal{O}(\frac{v^2}{\Lambda^2})$$

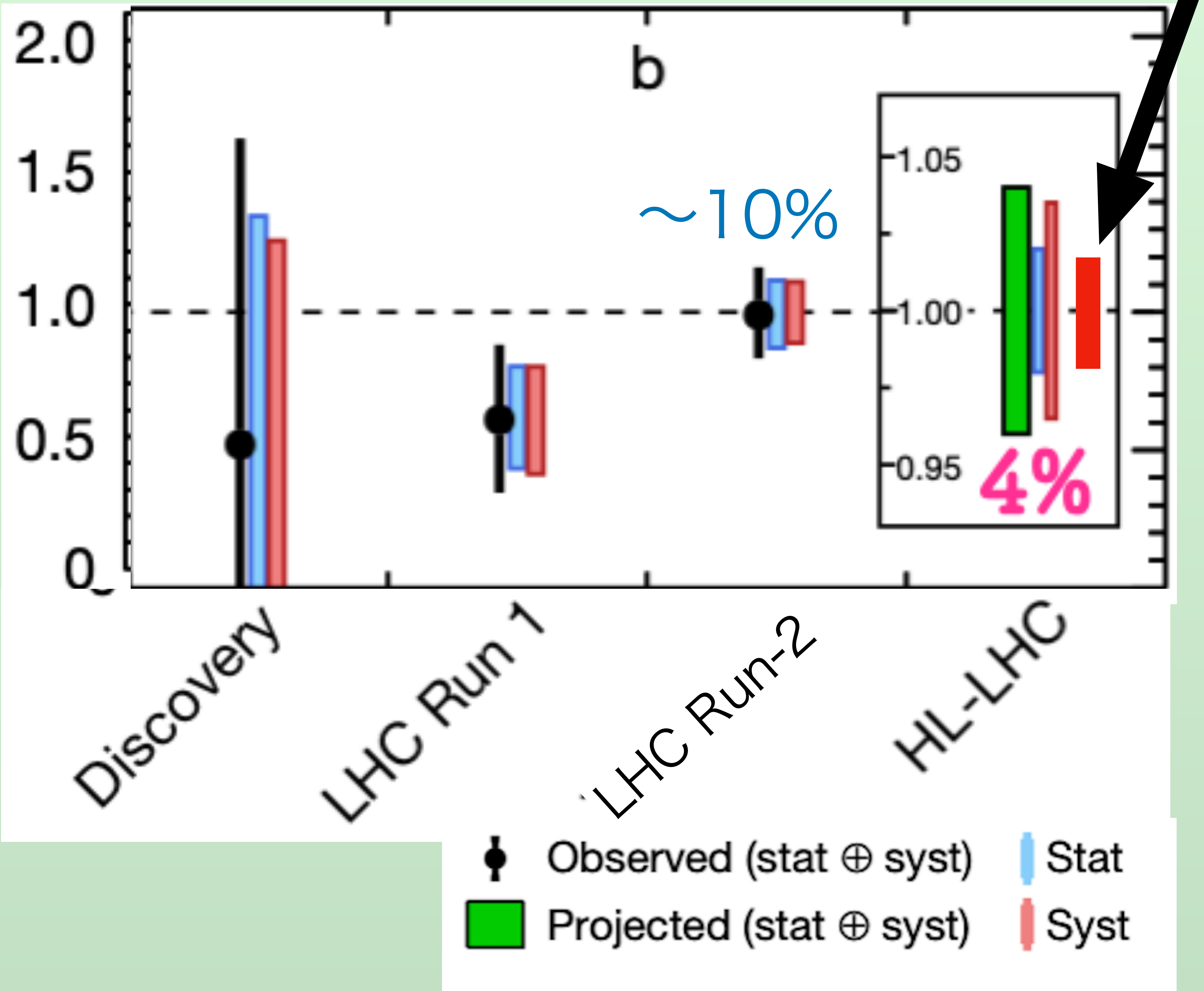
e.g. new physics at 1TeV

→ expected ～6% offset

→ Accuracy of about 1% required

ILC

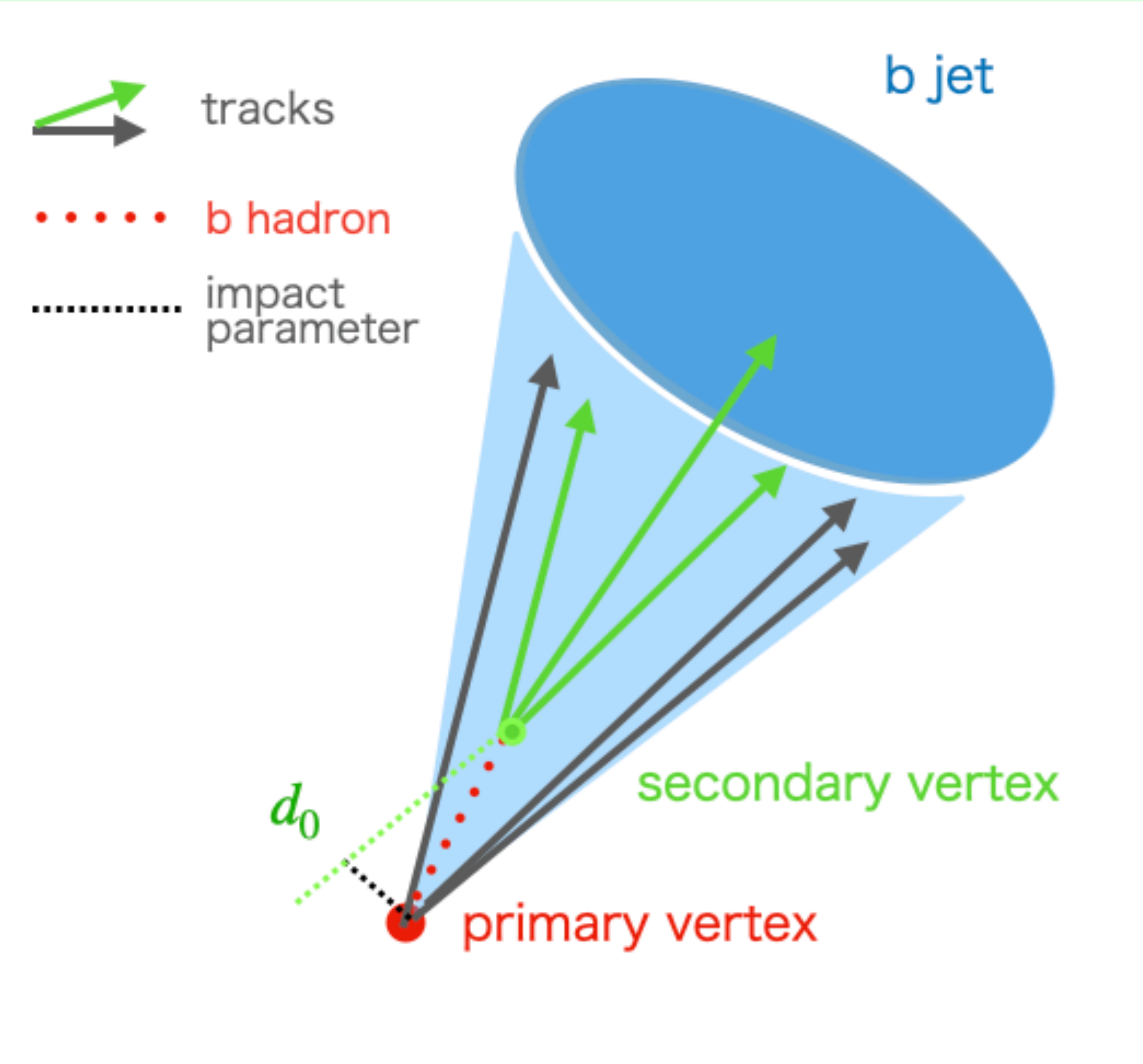
(～1%)



Nature 607, 60-68 (2022)

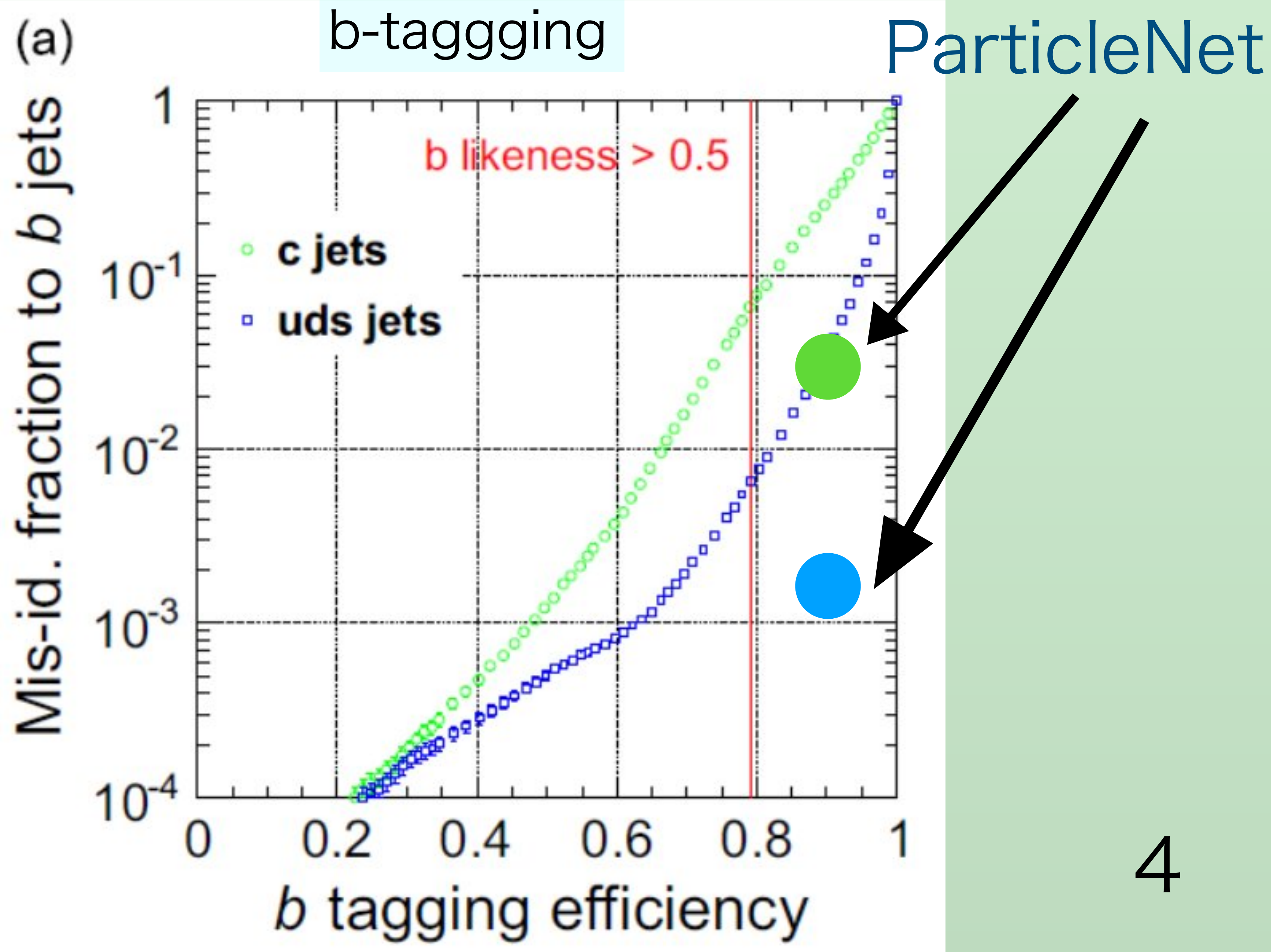
H→bb : **LHC Run-2**(precision) ～10% → **ILC** ～1%

Flavor tagging for Higgs factories



- For flavor tagging, the software **LCFIPlus** (published 2013) has been used in ILC/CLIC studies.
  - Flavor tagging using machine learning techniques (BDT)
  - b-tag: ~80%eff., 10% c / 1% uds mis-ID
  - c-tag: ~50%eff., 10% b / 2% uds mis-ID
- Recently FCCee's group reported this ~10 times better performance.
  - Flavor tagging using **ParticleNet** (GNN)
  - the dataset used was fast simulation
- **Particle Transformer** (ParT) research is currently being conducted by a group at the LHC

  ➡Trying to improve the performance of flavor tagging by applying **ParT** to **full simulation data of ILC**



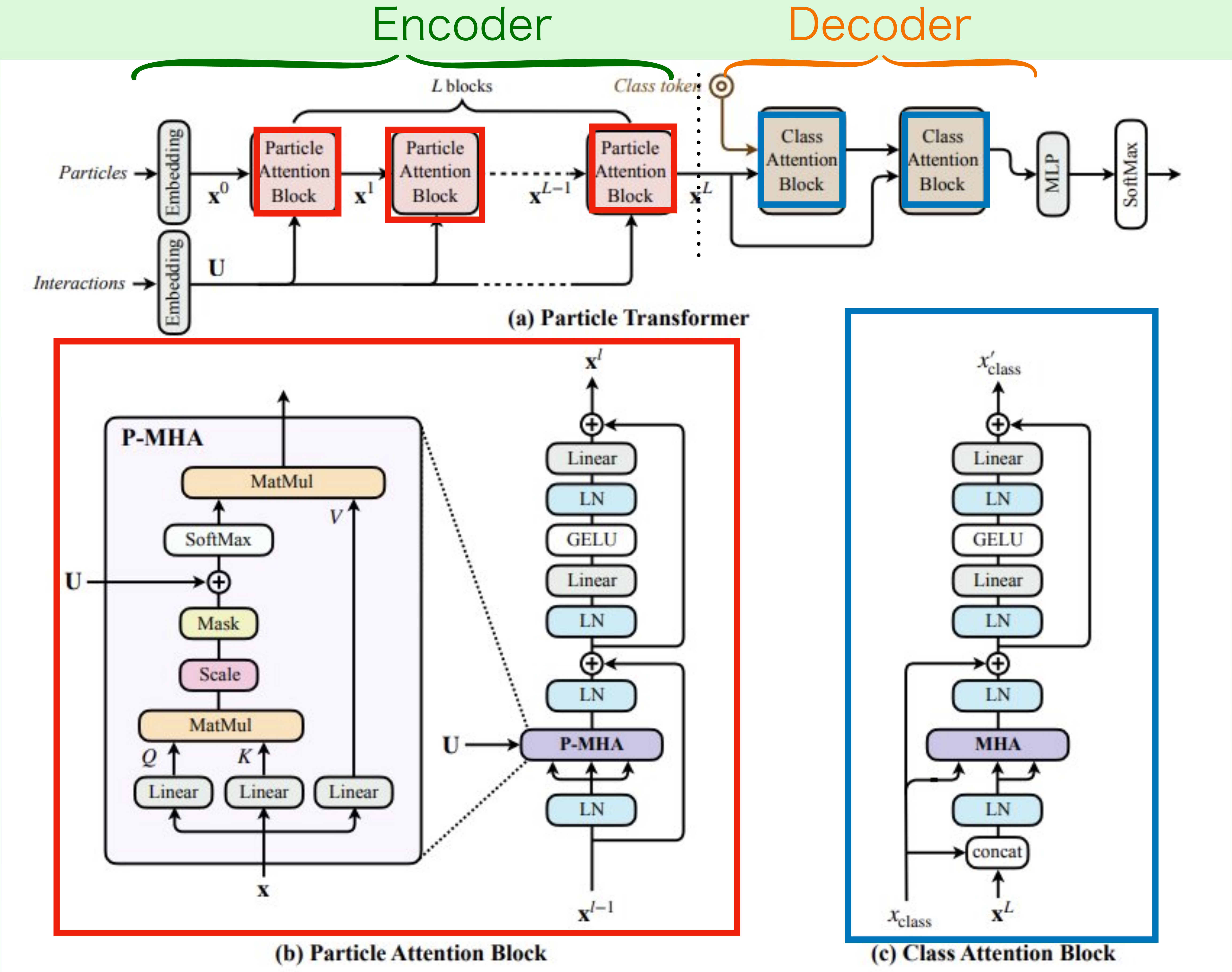
Performance of LCFIPlus

Particle Transformer (ParT)

- **ParT** is a modified Transformer model for Jet research (published in 2022.)
  - Considering the nature of Jet, input the physical quantity calculated from the quaternion momentum of two particles to Multihead attention.
- ParT has surpassed the performance of ParticleNet, which has been the highest-performing (arXiv: 2202.03772) 。

Event classification for JetClass

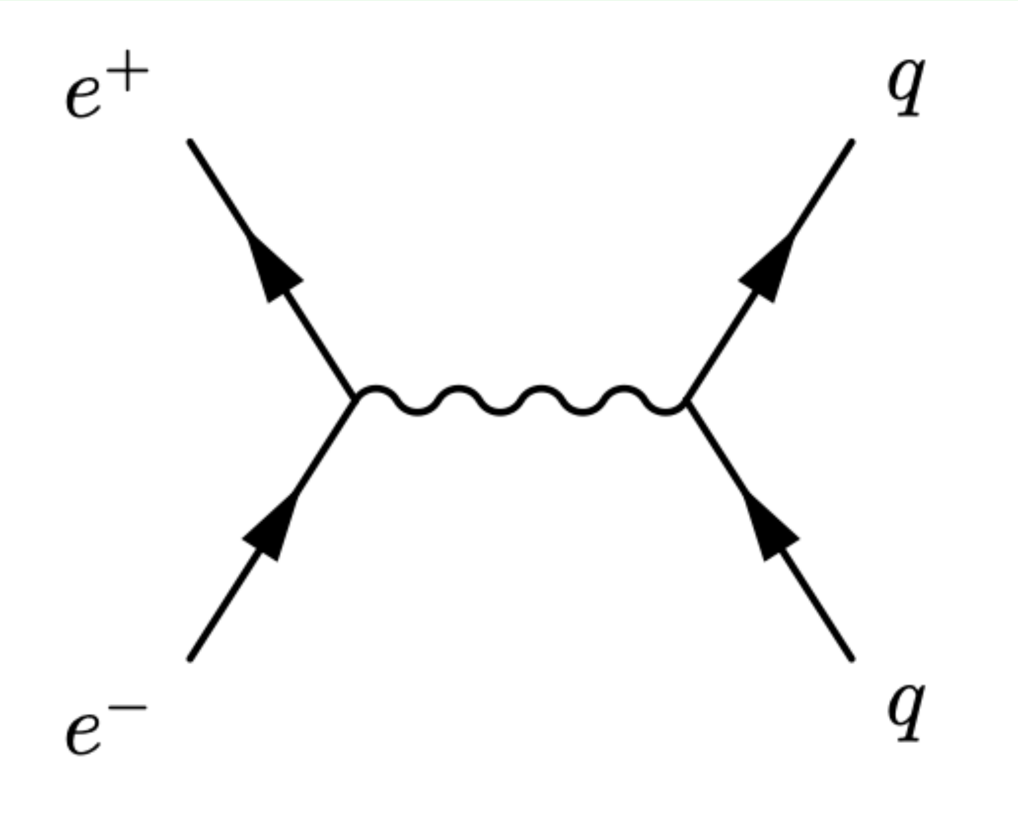
Event	H→bb Rej. 50%	H→cc Rej. 50%
Particle Net	0.013 %	0.04 %
ParT	0.0094%	0.024%



Application ParT for ILD datasets

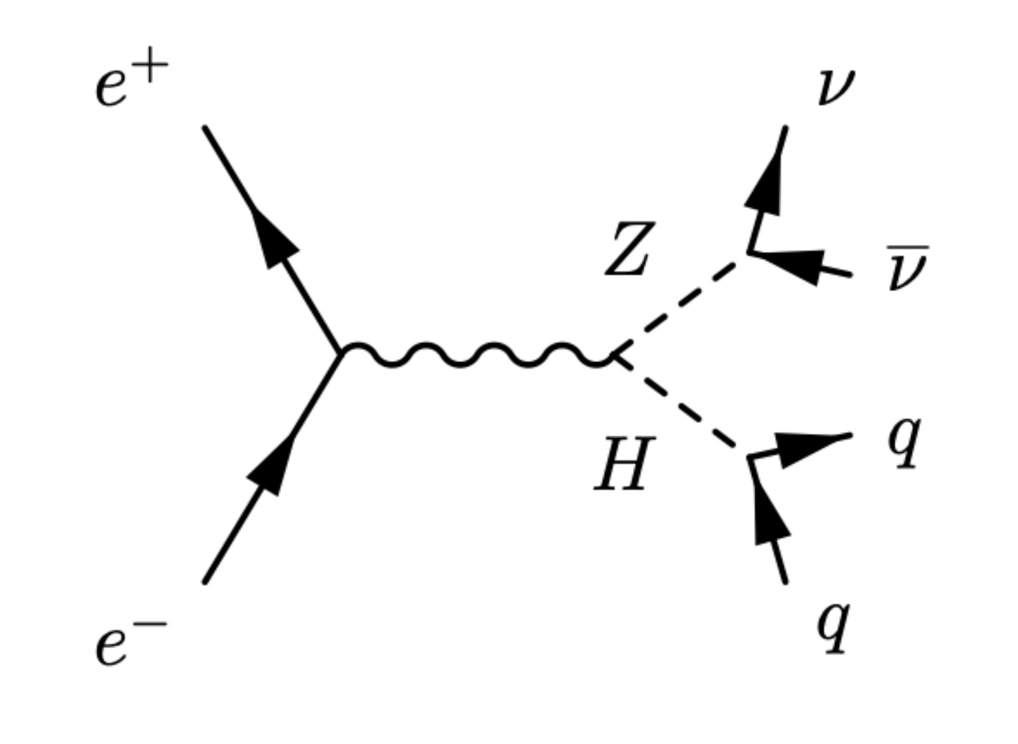
Dataset

- The dataset used for this study was the **ILD full simulation dataset**.
  - e+e- → Z → qq (at 91 GeV, 1M jets)
    (Same as used in the LCFIPlus study)
  - e+e- → ZH (H→ qq) (at 250 GeV, 1M jets)



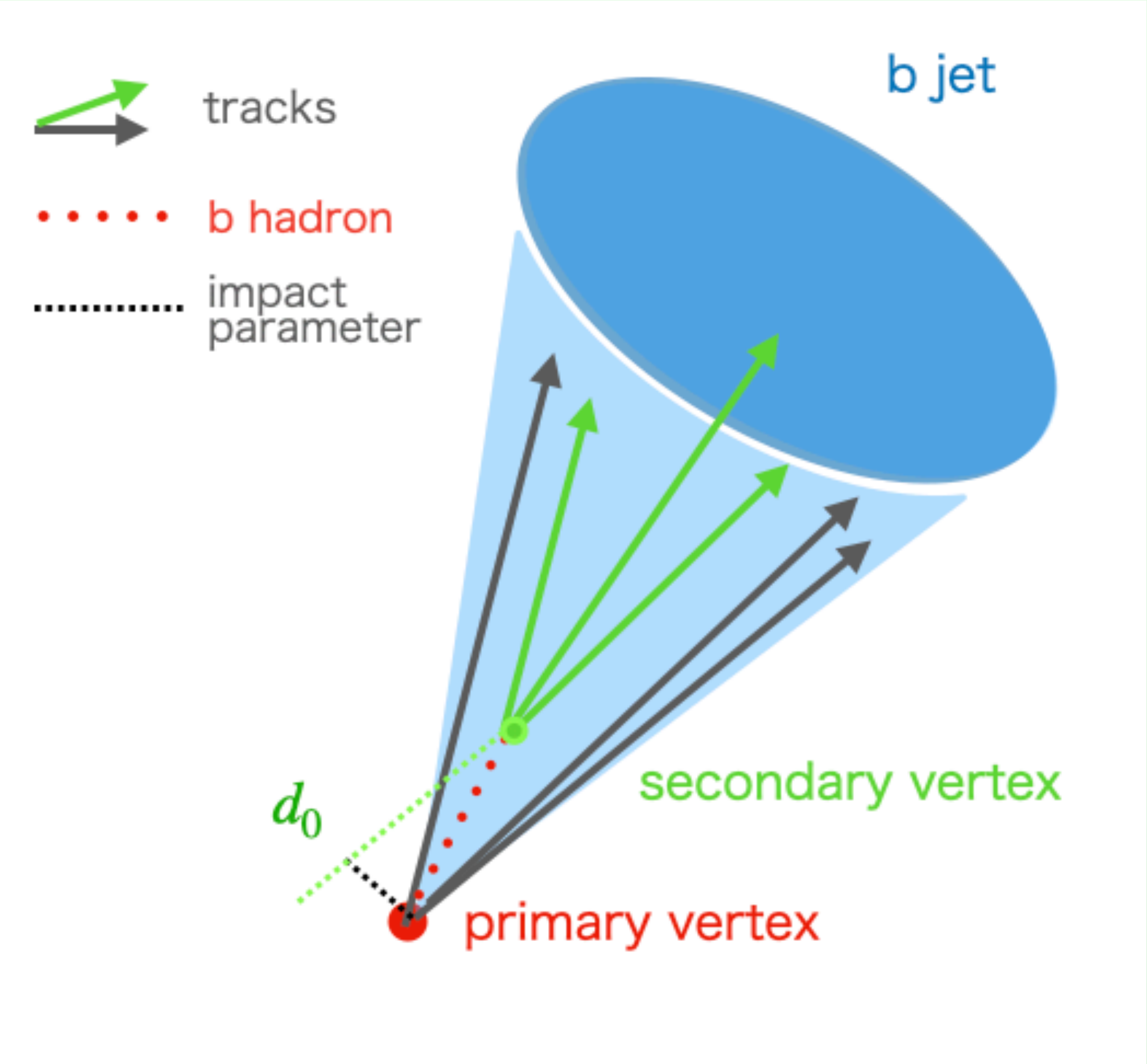
q = b,c,u,d,s
ν = neutrino

- training 80%, validation 5%, test 15%

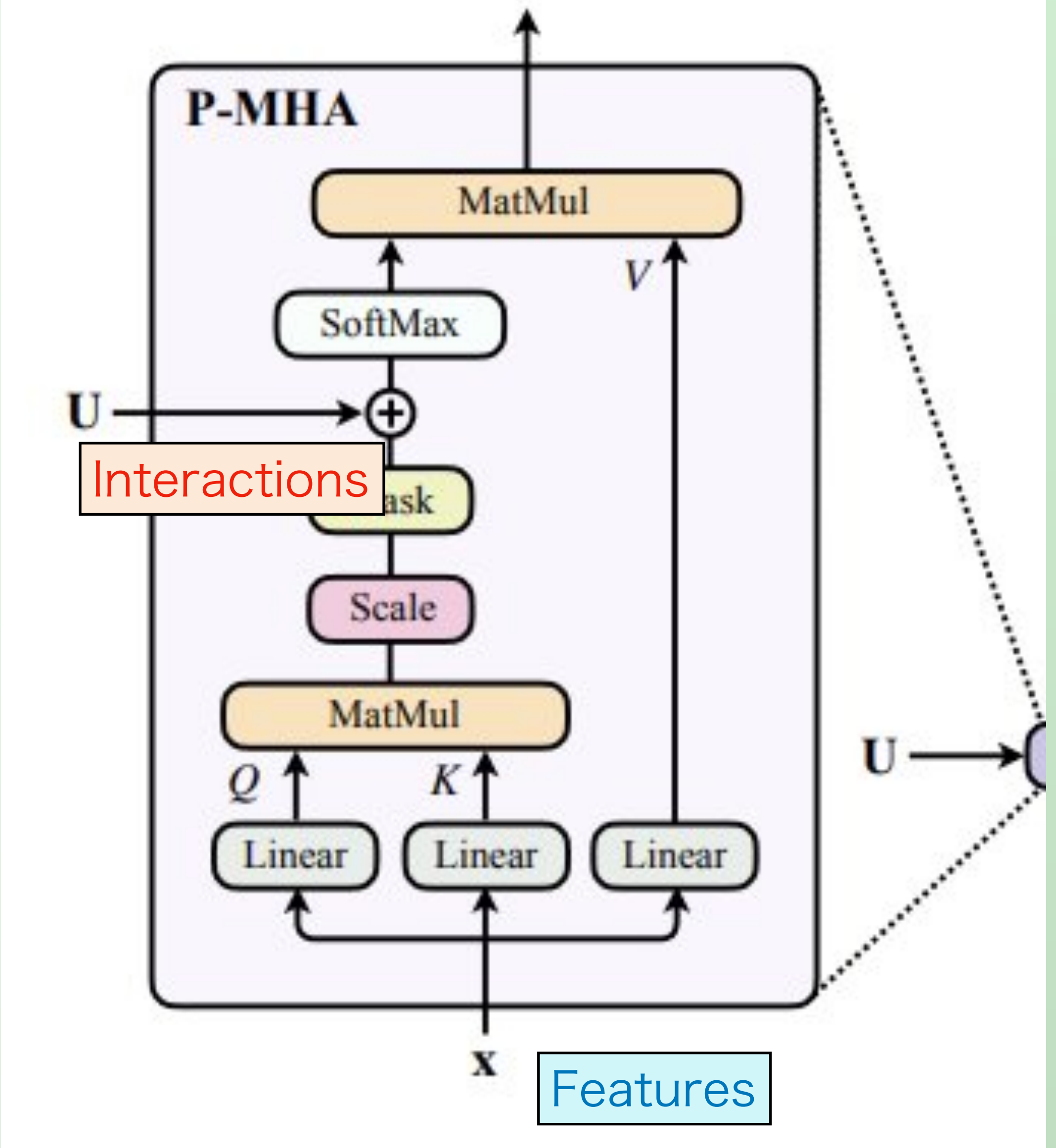


Input variables

- **Features** (for each track/neutral)
  - Impact Parameter (6): Distance between primary vertex and track (2D/3D)
  - Particle ID (6) : Each particle's character is expressed as 0 or 1. (e, mu, charged hadron, gamma, neutral hadron)
  - Kinematic (4) : particle energy/jet energy etc.
  - Track Errors (15) : covariant matrix
  - Jet Distance (2) : Distance between jet axis and each track (2D/3D)



- **Interactions**
  - Several variables calculated in pairs using quaternion momentum are listed as input variables
  - Add as MASK in the middle of attention



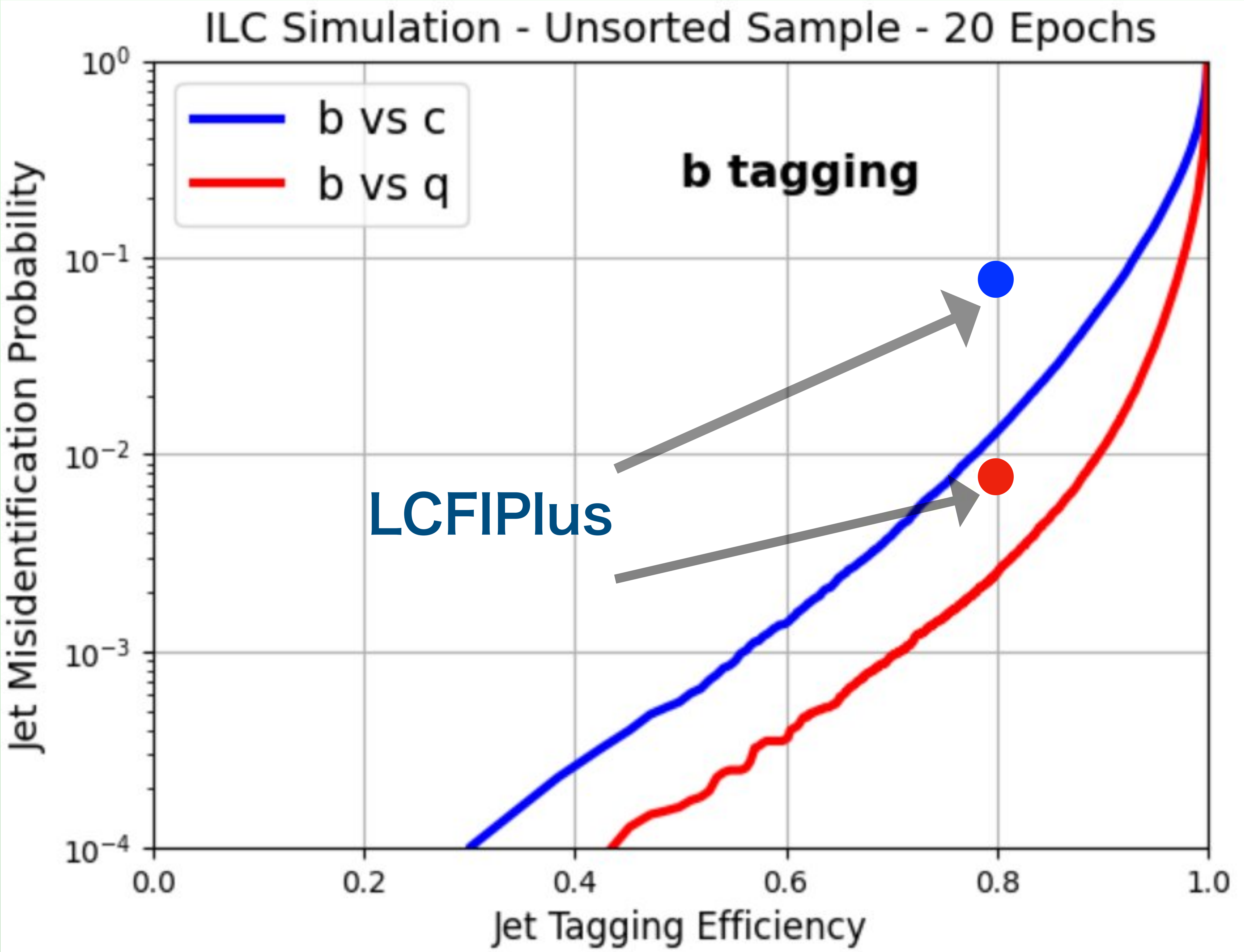
Compare LCFIPlus and ParT (ILD full simulation)

- 91 GeV data from ILD was used.
- The performance is greatly improved over LCFIPlus.

About 7.8 times

	b-tag 80% eff.		c-tag 80% eff.	
Method	c-bkg acceptance	uds-bkg acceptance	b-bkg acceptance	uds-bkg acceptance
LCFIPlus	10%	1%	10%	2%
ParT	1.29%	0.25%	1.02%	0.43%

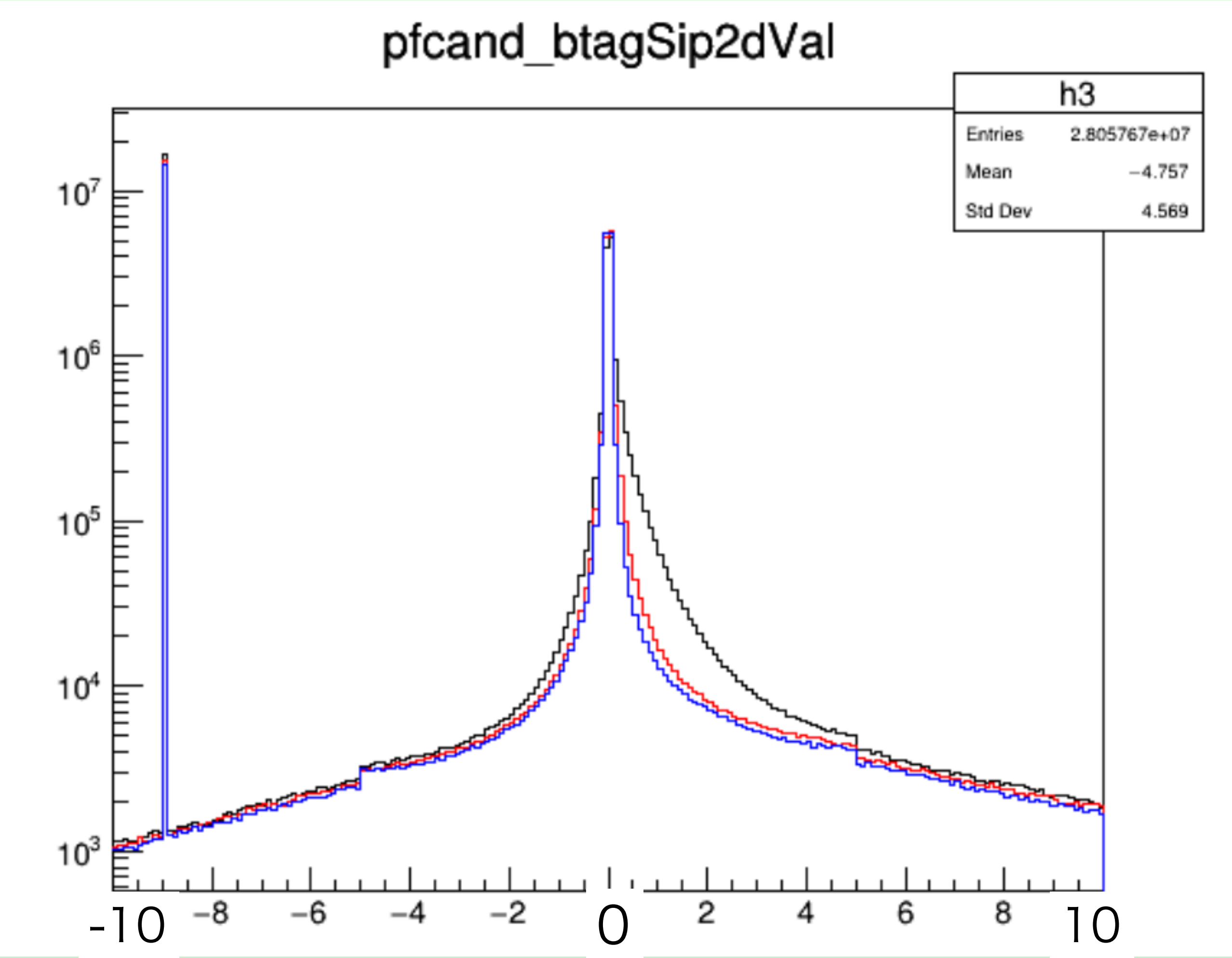
Performance of ParT



Handling of neutral particles (input node)

- Neutral Particle has been set to -9 for track among the many features variables.
- To avoid embedding (linear, GELU) mixed with Track particles, we performed embedding separately before training, and observed a performance improvement of ~8%.

Neutral's data is gathered to -9

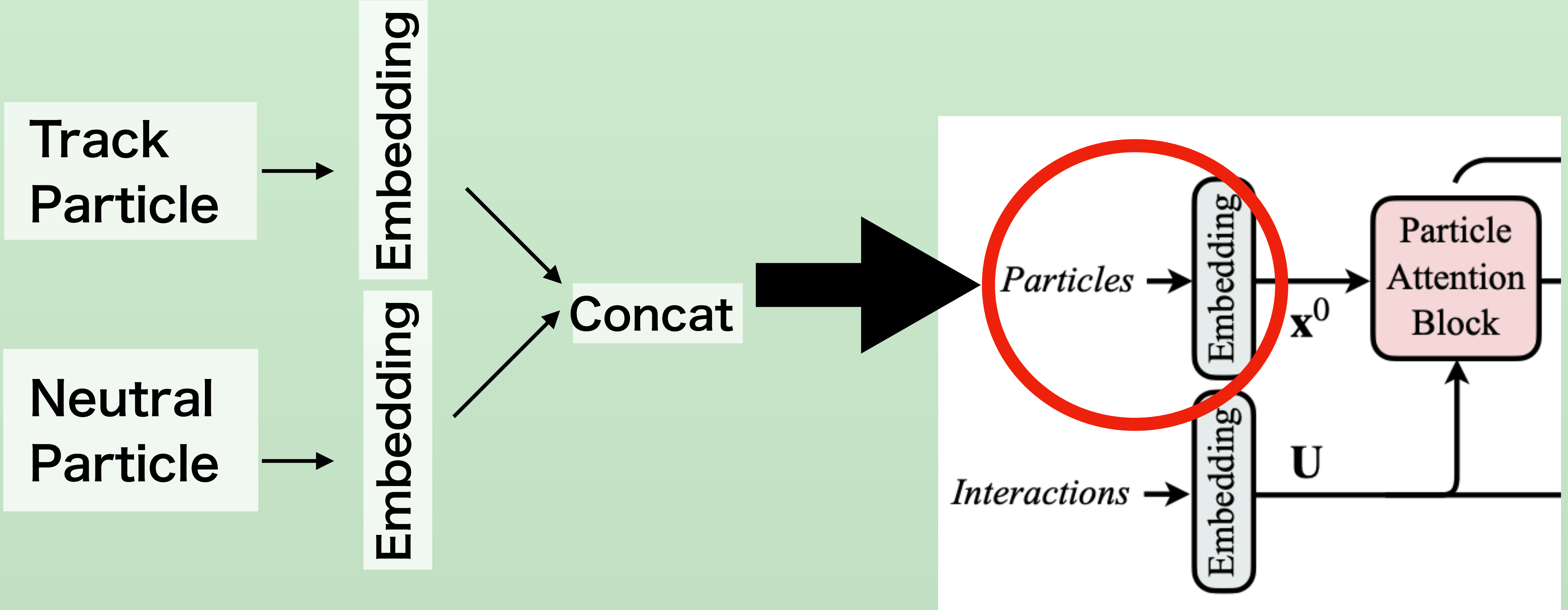


black ; b,   red ; c,   blue ; d

Learning for ILD data

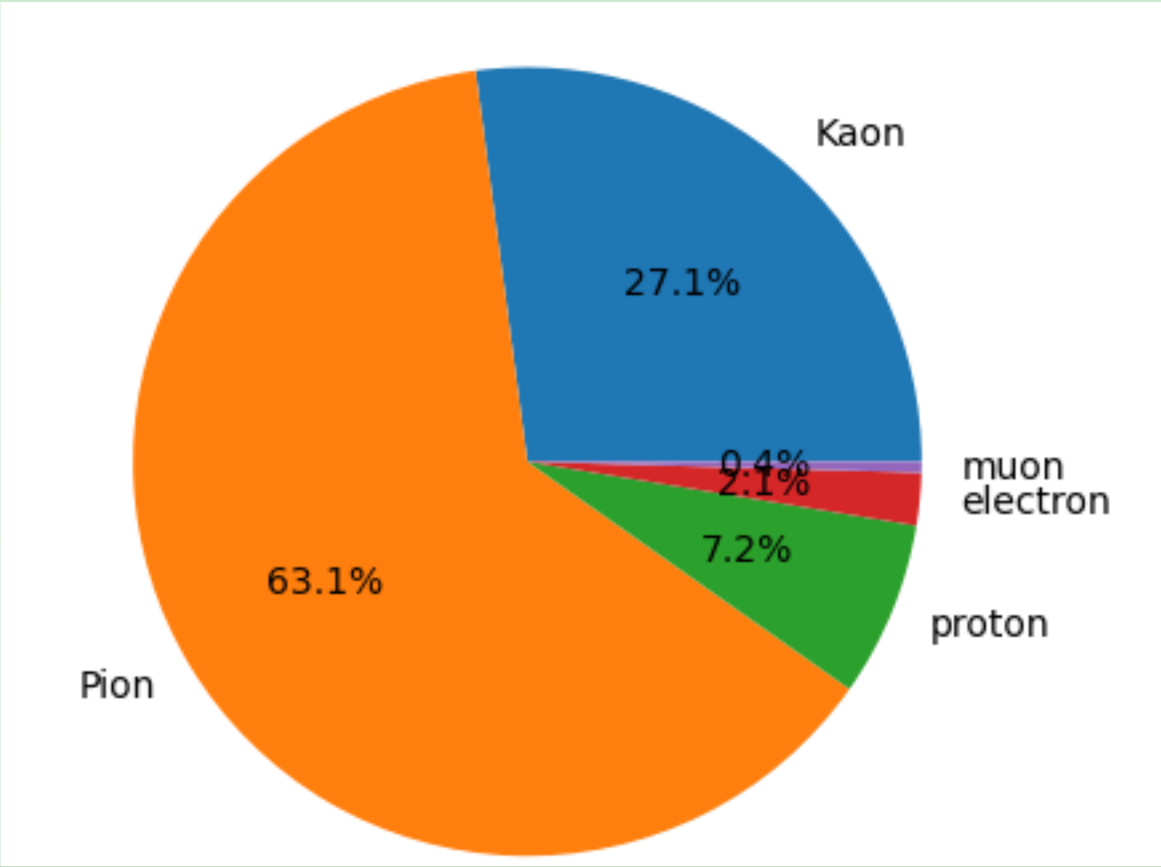
	b-tag 80% eff. c-bkg acceptance (%)	c-tag 80% eff. b-bkg acceptance (%)
Without dividing	0.518	6.60
Dividing and embedding	0.476	6.20

About 8%

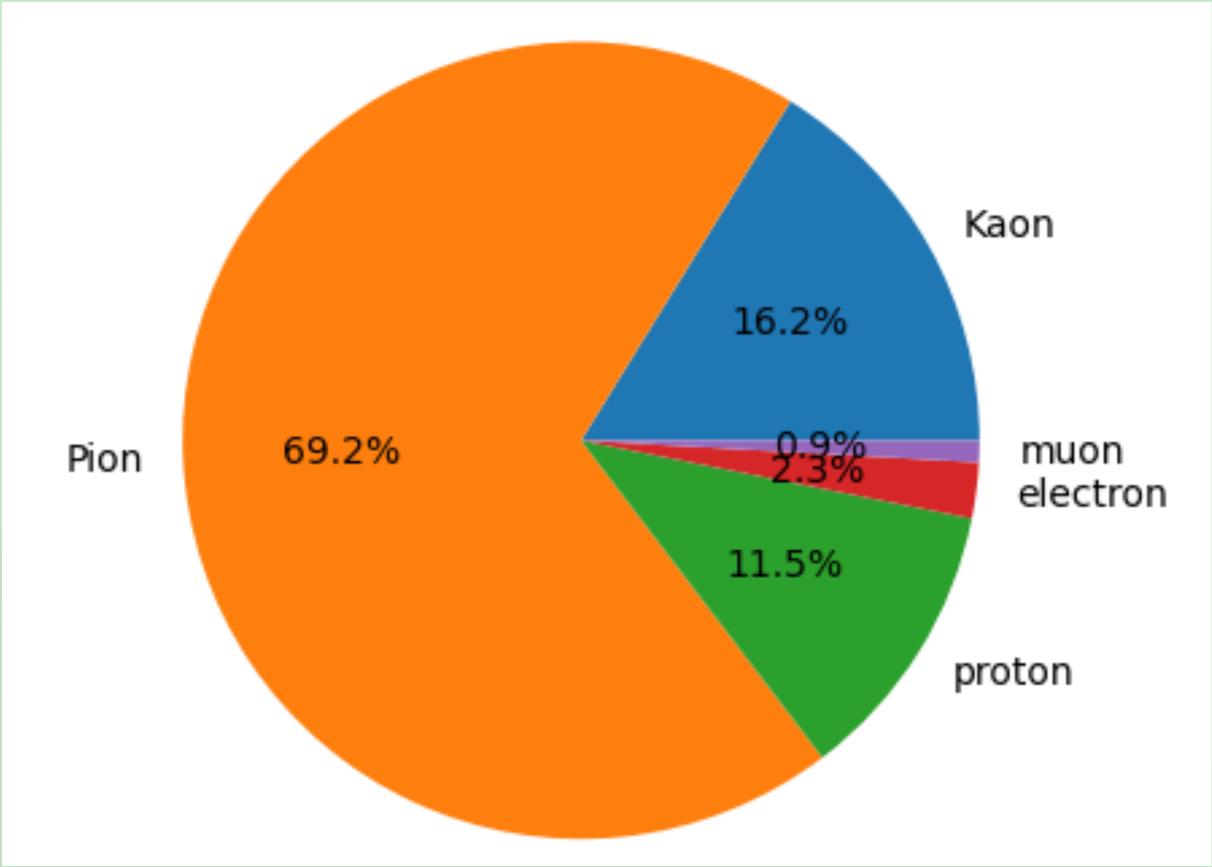


Strange tagging

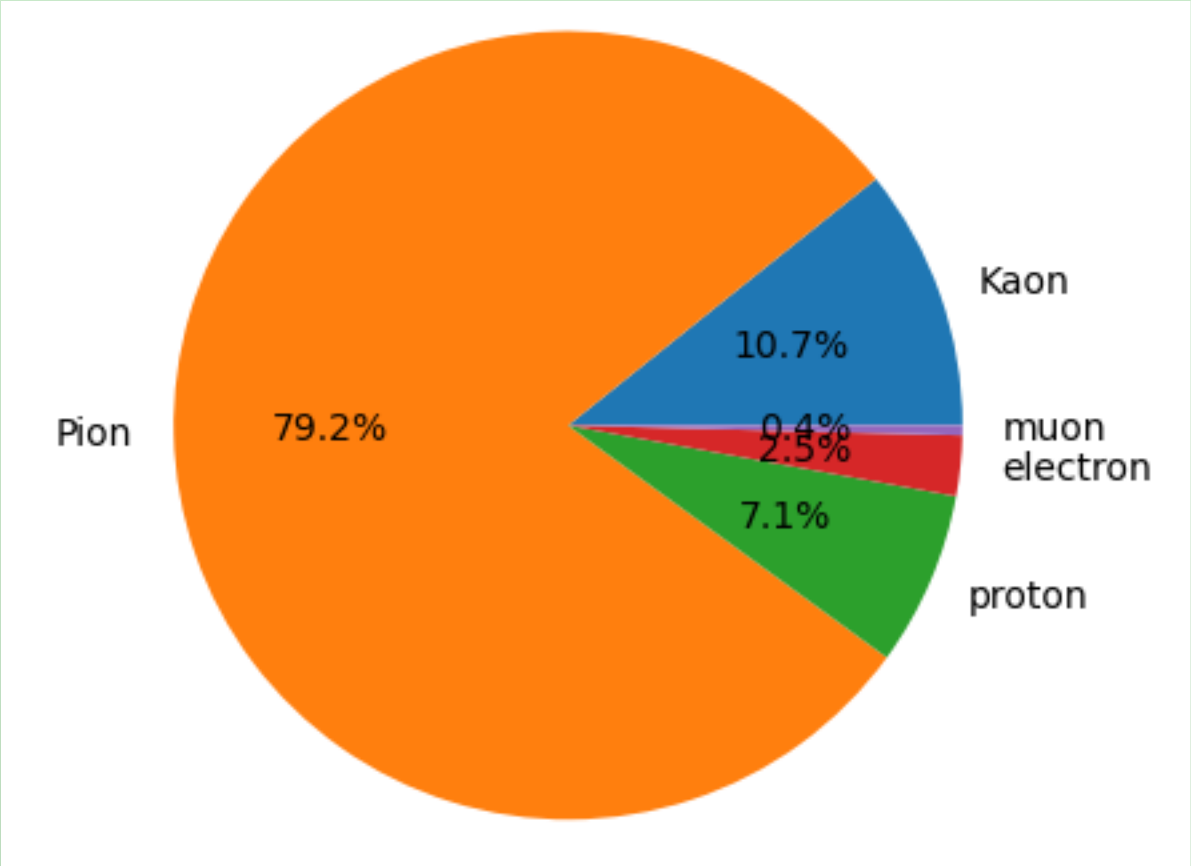
- We also work on to improve the efficiency of **strange jet** tagging by mainly using particle ID of the particles in the jets
- Particle ID
  - Upgrade instant ID to using **Comprehensive PID(CPID)**
  - Particles IDs : electron, muon, kaon, pion, proton



H->ss



H->gg



H->dd

Particle ID (truth) ratio (p>5GeV)

- Strange jets have more Kaons
- Down jets have more Pions

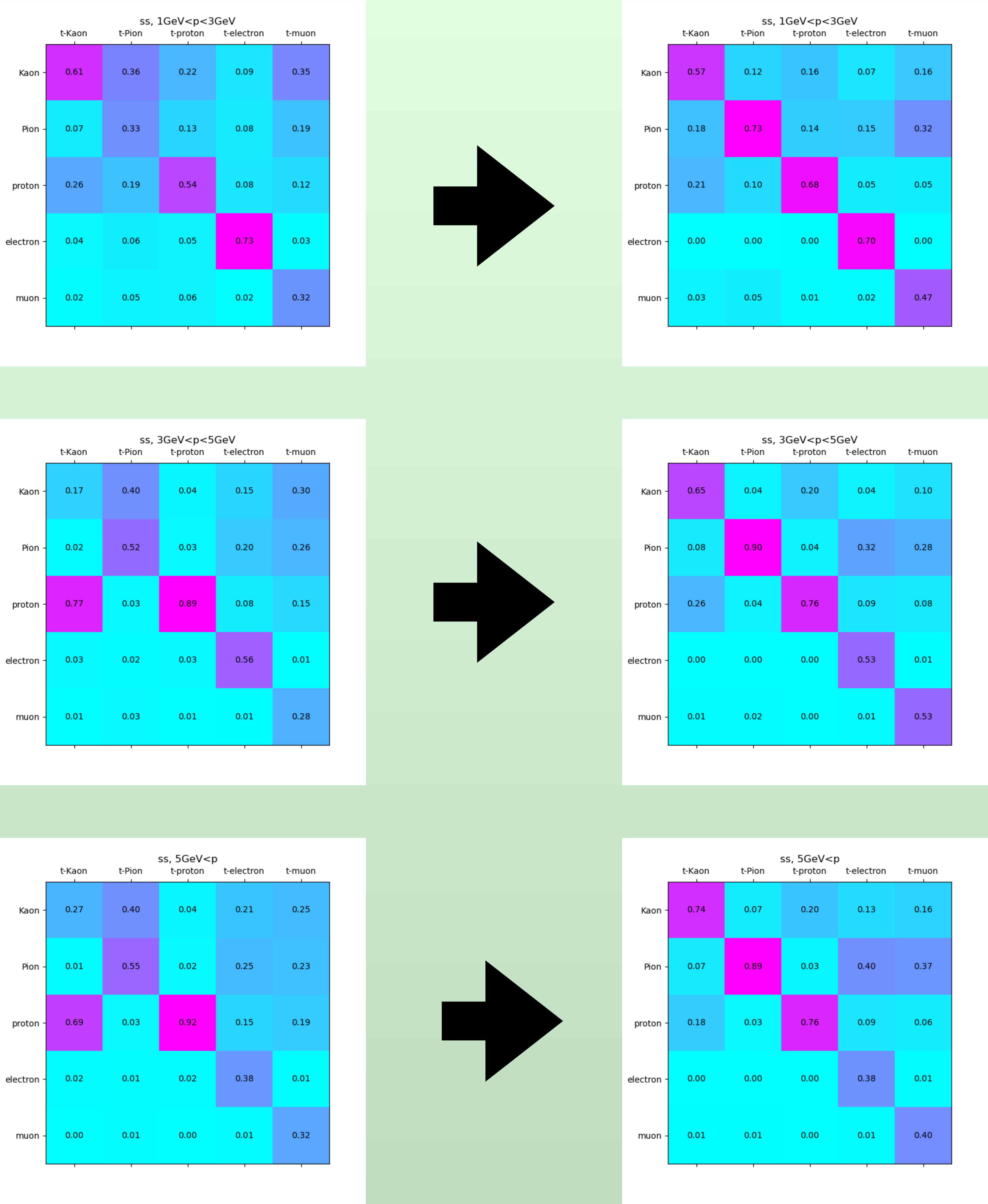
Comprehensive PID

- Central book-keeping, modules for PID observables as well as training & inference
- Split momentum range of 1-100 GeV into 12 momentum bin with separate multi class BDT each
- Input : slcio file, steering files with processor parameters, module parameters
- Output : BDT score for each species hypothesis (-> slcio file)
- For 250 GeV MC production of 2020 (ILD simulation)

https://agenda.linearcollider.org/event/10211/contributions/53839/attachments/39291/61960/2024_01_16_ILDMeeting.pdf

Comprehensive PID

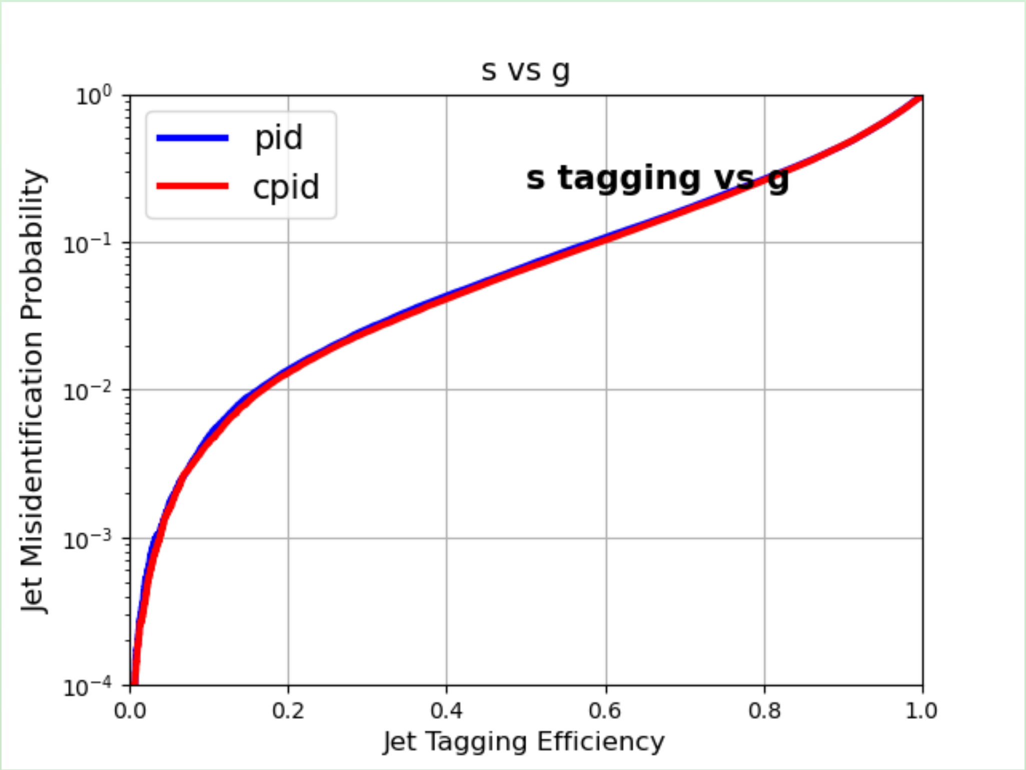
- CPID improves the accuracy of PIDs a lot
- There are not much difference between H->dd and H->ss data except kaon pid and proton pid, so we think we have to make some weights on them



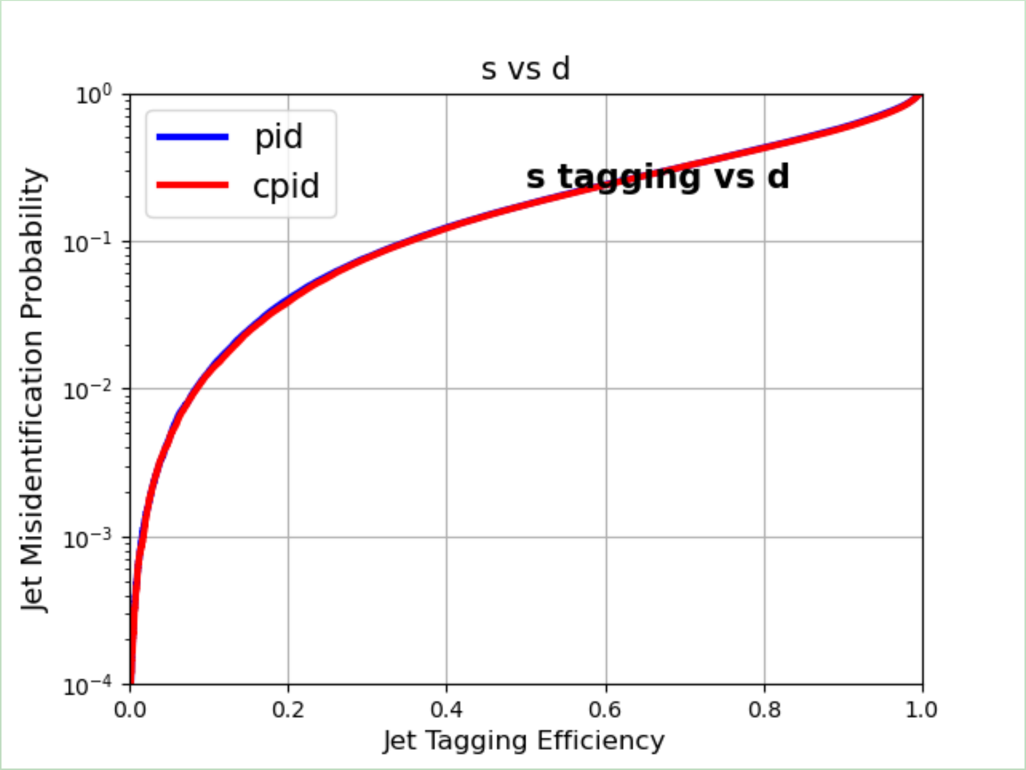
Strange tagging

- The efficiency of strange tagging is below.
- There are lots of contaminations. In addition, there are few differences between previous pids and CPIDs. We are trying to investigate the reasons of them and improve the effs.

	s-tag 80% eff.	
Method	g-bkg acceptance (%)	d-bkg acceptance (%)
Previous PID	26.5%	42.8%
CPID	25.7%	42.7%



s vs g



s vs d

Summary

- **Flavor tagging** is important in the search for new physics through precise measurement of Higgs. Machine learning can be used to improve performance and contribute to the search.
- In this research, **Particle Transformer** with higher performance for flavor tagging was developed by the LHC group and applied to the ILD dataset.
- Particle Transformer is also valid for the ILD datasets. The performance of b-tagging is 8 times better than the conventional software (LCFIPlus).
- We're also trying to improve strange jet tagging by using ParT.

Back up : Input Variables - Features

- Impact Parameter (6):

pfcand_dxy
pfcand_dz
pfcand_btagSip2dVal
pfcand_btagSip2dSig
pfcand_btagSip3dVal
pfcand_btagSip3dSig

*d0/z0 and 2D/3D impact parameters, -9 for neutrals

- Jet Distance(2):

pfcand_btagJetDistVal
pfcand_btagJetDistSig

*Displacement of tracks from line passing IP with direction of jet, -9 for neutrals

- Particle ID (6):

pfcand_isMu
pfcand_isEl
pfcand_isChargedHad
pfcand_isGamma
pfcand_isNeutralHad
pfcand_type

*Not including strange-tagging related variables (TOF, dE/dx etc.)

*Simple PID for ILD, not optimal

- Kinematic (4):

pfcand_erel_log
pfcand_thetarel
pfcand_phirel
pfcand_charge

*Fraction of the particle energy wrt jet energy (log is taken)

- Track Errors(15):

pfcand_dptdpt
pfcand_detadeta
pfcand_dphidphi
pfcand_dxydxy
pfcand_dzdz
pfcand_dxydz
pfcand_dphidxy
pfcand_dlambdadz
pfcand_dxyc
pfcand_dxyctgtheta
pfcand_phic
pfcand_phidz
pfcand_phictgtheta
pfcand_cdz
pfcand_cctgtheta

*Each element of covariant matrix, -9 for neutrals

Backup: Interaction variables

$$\begin{cases} \log(\Delta R) \\ \log(kt) \\ \log(z) \\ \log(\text{inv. mass}) \end{cases}$$

$$z_{ij} = \frac{pt_{min}}{pt_i + pt_j}$$