

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

Risako Tagami¹, Taikan Suehara, and Masaya Ishino

¹rtagami@icepp.s.u-tokyo.ac.jp

Abstract. International Linear Collider (ILC) is a next-generation e^+e^- linear collider to explore Beyond-Standard-Models by precise measurements of Higgs bosons. Jet flavor tagging plays a vital role in the ILC project by identification of $H \rightarrow b\bar{b}, c\bar{c}, gg, s\bar{s}$ to measure Higgs coupling constants and of $HH \rightarrow b\bar{b}b\bar{b}$ and $b\bar{b}WW$ which are the main channels to measure the Higgs self-coupling constant.

Jet flavor tagging relies on a large amount of jet information such as particle momenta, energies, and impact parameters, obtained from trajectories of particles within a jet. Since jet flavor tagging is a classification task based on massive amounts of information, machine learning techniques have been utilized for faster and more efficient analysis for the last several decades.

Particle Transformer (ParT) is a machine learning model based on Transformer architecture developed for jet analysis, including jet flavor tagging. In this study, we apply ParT to ILD full simulation data to improve the efficiency of jet flavor tagging.

Our research focused on evaluating the performance of ParT compared to that of the previously used flavor tagging software, LCFIPlus. We will also report the status of the performance of strange tagging on ILD full simulation dataset using ParT, which can be applied to the analysis of Higgs-strange coupling.

1 Introduction

After the discovery of the Higgs boson, the search for beyond the Standard Model through precise measurement of Higgs has been expected. There is a consensus among particle physicists to build Higgs Factories as the next-generation accelerator. There are several e^+e^- Higgs Factories currently under consideration. ILC project is one of the Higgs Factory projects.

With more precise measurements of Higgs, the effects of SUSY and other new TeV physics models can be seen. For instance, if there is a new physics phenomenon at 1 TeV, the offset from the Standard Model will be about 6% so that accuracy of about 1% will be required. In the ILC project, the accuracy of $H \rightarrow b\bar{b}$ measurements is expected to be $\sim 1\%$. For precise measurements of Higgs, especially Higgs coupling constants of $H \rightarrow b\bar{b}, c\bar{c}, gg, s\bar{s}$, the performance of flavor tagging should be improved.

Jet flavor tagging is a classification task using the information of each jet such as impact parameters of tracks and number of vertices. Bottom jets have two secondary vertices, whereas charm jets have one secondary vertex. Strange jets have more strange hadron like

This work was carried out in the framework of the ILD Concept Group.

Kaon, whereas light jets have less. Those differences between jets should be detected by jet flavor tagging.

In ILC or CLIC studies, LCFIPlus[1] has been used for flavor tagging. This is a flavor tagging software using machine learning technique of boosted decision tree (BDT) classification. Recently, a research group in CERN presented a new model for flavor tagging called ParticleNet[2] using Graph Neural Network (GNN) model. As below, ParticleNet achieved better performance[3] than LCFIPlus, although there are differences between them in terms of the datasets used and their detectors (ILC and FCC-ee).

Table 1. The comparison of the performance of b -jet flavor tagging between using LCFIPlus for ILD full simulation dataset and using ParticleNet for FCC-ee fast simulation dataset.

Method	b -tag 80% efficiency	
	c -bkg acceptance	uds -bkg acceptance
LCFIPlus	6.3 %	0.79 %
ParticleNet	0.40 %	< 0.01 %

Lately, Particle Transformer (ParT)[4] has been applied to LHC fast simulation data. The performance of event tagging of fast simulation dataset (JetClass[4]) has been greatly improved by using ParT. In this study, we apply ParT to ILD full simulation dataset to aim at improving the performance of flavor tagging.

2 Flavor tagging of ILD full simulation dataset using ParT

2.1 Dataset

In this study, the full detector simulation events for International Large Detector (ILD)[5], one of the ILC detector concepts, are utilized. The ILD has trackers and calorimeters. The trackers consist of silicon vertex detector, silicon inner and outer tracker and main Time Projection Chamber (TPC) at barrel region, while silicon-only tracking at forward region. The calorimeters consist of electromagnetic calorimeter which is composed of tungsten absorber and highly-granular silicon pad sensors with cell size of $5 \times 5 \text{ mm}^2$, and hadron calorimeter which is composed of steel absorber and scintillator tile sensors with $3 \times 3 \text{ cm}^2$ cells. End-cap and forward calorimeters are also highly segmented with similar configurations. Outside the hadron calorimeter, superconducting solenoid with 3.5 Tesla magnetic field and muon detector as well as return yoke are equipped in this order.

The software framework (iLCSoft) consists of Geant4 based detector simulation with the ILD detector setup (DDsim), digitizer of trackers and calorimeter hits emulating detector effects and a tracking software as low-level reconstruction. For the reconstruction of particles and jets, standard Particle Flow Algorithm (PandoraPFA) and Durham jet clustering algorithm are utilized, respectively. Particles inside each jet are used to produce input variables for the flavor tagging network.

The samples used in this study are $e^+e^- \rightarrow q\bar{q}$ events at 91 GeV center-of-mass energy (fixed energy), 1M jets in total and $e^+e^- \rightarrow ZH \rightarrow \nu\bar{\nu}q\bar{q}$ at 250 GeV (variable energy, their average is $\sim 60 \text{ GeV}$), 1M jets in total.

They are divided into training samples (85%), validation samples (5%) and test samples (15%).

2.2 Model

The architecture of ParT is as Figure 1.

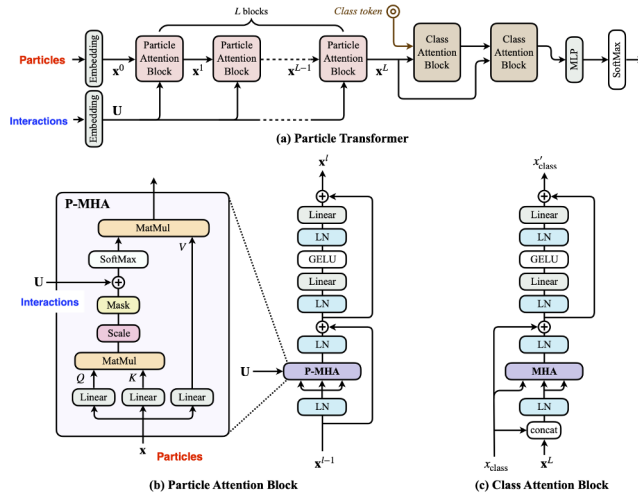


Figure 1. The architecture of Particle Transformer's model. The original image is from [4].

The basic structure of ParT is same as Transformer. The difference between standard Transformer and ParT is that ParT has "interactions." Several physical variables are calculated by forming pairs of tracks within jets using their four-momenta and added as bias in the middle of self-attention.

2.3 Input variables

Input variables (features and interactions) are listed in Table 2 and Table 3.

2.4 Result

Performance of three category (*b*, *c*, *d*) flavor tagging on ILD full simulation dataset (at 91 GeV) using ParT is evaluated and compared to the performance using LCFIPlus, and the results are summarized in Table 4.

The ParT model is modified to adjust to the ILD full simulation dataset. For instance, features are separated into tracks and neutrals before embedding (linear, Gaussian Error Linear Unit (GELU) activation function) and combined before self-attention block, and several variables of features are standardized by sigmoid function before training. The performance on ILD full simulation dataset (at 250 GeV) with those modification is also listed on the Table 4.

The performance of *b*-tagging and *c*-tagging is greatly improved by using ParT, about 4.8 times in *c*-background acceptance of *b*-tagging. We also achieved a ~ 2.7 times improvement in performance with the modifications. It was found that ParT is also effective for ILD full simulation dataset.

3 Strange tagging

We also report current status of strange jet tagging of ILD full simulation dataset using ParT.

Table 2. Input variables (features).

Category	Variable	definition
Impact Parameter	d_0	transverse impact parameter value
	z_0	longitudinal impact parameter value
	ip2D	2D impact parameter value
	ip3D	3D impact parameter value
Jet Distance	d_j	displacement of tracks from the line passing interaction point with direction of their jets
Particle ID	Muon	if the particle is a muon ($ pid == 13$)
	Electron	if the particle is an electron ($ pid == 11$)
	Gamma	if the particle is a gamma ($ pid == 22$)
	Charged Hadron	if the particle is a charged hadron
	Neutral Hadron	if the particle is a neutral hadron
Kinematic	type	PDG ID reconstructed by each Particle ID
	Charge	electric charge of the particle
	$\log \frac{E}{E_{jet}}$	logarithm of the particle's energy relative to the jet energy
	$\Delta\theta$	$\theta_{track} - \theta_{jet}$
Track Errors	$\Delta\phi$	$\phi_{track} - \phi_{jet}$
	σ	each element of covariant matrix

* Impact parameter, Jet Distance and Track Errors are only valid for charged particles. They are set to be -9 for neutral particles to fill in the input data.

Table 3. Input variables (interactions). i and j each represent one of a pair of particles.

Variable	definition
$\log(\Delta R)$	logarithm of $\sqrt{(\eta_i - \eta_j)^2 + ((\phi_i - \phi_j + \pi) \% (2\pi) - \pi)^2}$
$\log(kt)$	logarithm of $\sin \theta_{ij} * p_{\min}$ ($p_{\min} = \min(p_i, p_j)$)
$\log(z)$	logarithm of $\frac{p_{\min}^T}{p_i^T + p_j^T}$ ($p_{\min}^T = \min(p_i^T, p_j^T)$)
$\log(\text{inv.mass})$	logarithm of invariant mass

Table 4. The performances of b -tagging and c -tagging. In the top row and the bottom of the table, ILD full simulation dataset (at 250 GeV, 1M jets) is used, while in the middle row, ILD full simulation dataset (at 91 GeV, 1M jets) is used.

Method	b -tag 80% efficiency		c -tag 50% efficiency	
	c -bkg acceptance	d -bkg acceptance	b -bkg acceptance	d -bkg acceptance
LCFIPlus (at 250 GeV)	6.3 %	0.79 %	7.4 %	1.2 %
ParT (at 91 GeV)	1.3 %	0.25 %	1.0 %	0.43 %
ParT (at 250 GeV, modified)	0.48 %	0.14 %	0.86 %	0.34 %

There are only slight differences in kinematics between strange jet and light jets. Their main difference is the ratio of each particle in jets shown in Table 5. In strange jets, there are more strange hadrons like Kaon.

It is common to use particle identification (ID) of particles in jets for tagging strange jets. We use a new particle ID module called Comprehensive PID (CPID)[6] to improve the performance of strange tagging. With CPID, particles are split by momentum range of 1-100 GeV into 12 momentum bins and their IDs are determined by BDT with the information

Table 5. The ratios of track particles with momenta bigger than 5 GeV.

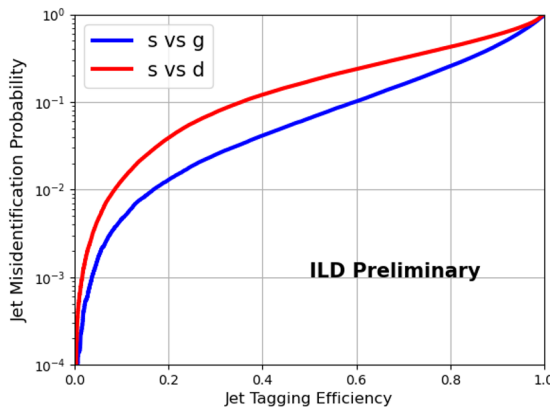
Particle	$H \rightarrow s\bar{s}$	$H \rightarrow gg$	$H \rightarrow d\bar{d}$
Kaon	27.1 %	16.2 %	10.7 %
Pion	63.1 %	69.2 %	79.2 %
Proton	7.2 %	11.5 %	7.1 %
Electron	2.1 %	2.3 %	2.5 %
Muon	0.4 %	0.9 %	0.4 %

of reconstructed particle including time of flights and dE/dx . As a result, we can obtain particle's probabilities for each species ; kaon, pion, proton, electron and muon. The accuracy of particle ID obtained by CPID is as Table 6. The dataset used is ILD full simulation dataset at 250 GeV with 1M jets.

Table 6. Accuracy rate of CPID in $H \rightarrow s\bar{s}$ events with momenta bigger than 5 GeV. Accuracy is defined by the proportion of tracks that CPID (made into 0 or 1 by setting the probability of the Kaon, Pion, Proton, Electron and Muon such that the highest value is 1, and the others are 0) and truth match.

Particle	Accuracy
Kaon	0.74
Pion	0.89
Proton	0.76
Electron	0.38
Muon	0.40

We use CPID scores of those five particles and also usual inputs in Table 2 and Table 3 for strange tagging. The current performance of strange tagging is as Figure 2.

**Figure 2.** The performances of strange tagging of ILD full simulation dataset (at 250 GeV, 1M jets) using ParT. Six category (b , c , d , g , u , s) flavor tagging was performed. Blue line is g -background acceptance, and red line is d -background acceptance.

The g -background acceptance in s -tagging 80% efficiency is 25.7% and the d -background acceptance in s -tagging 80% efficiency is 42.7%. The performance of strange tagging on ILD

full simulation dataset is lower than that on FCC-ee dataset, but they are different in terms of their detectors and fast/full simulation. The difference of performance is currently under investigation.

4 Discussion and Conclusion

Particle Transformer is also effective for b , c jet flavor tagging on ILD full simulation dataset. The performance of b -tagging (c -background) with ParT is about five times better than with LCFIPlus, and c -background acceptance is 0.48% with b -tagging 80% efficiency.

Strange jet tagging with ParT using Comprehensive PID is also performed. The performance of s -tagging 80% efficiency (g -background) is about 25%. We will try to improve the performance, using CPID for particles with their momenta less than 1 GeV.

5 Acknowledgements

We would like to thank the LCC generator working group and the ILD software working group for providing the simulation and reconstruction tools and producing the Monte Carlo samples used in this study. This work has benefited from computing services provided by the ILC Virtual Organization, supported by the national resource providers of the EGI Federation and the Open Science GRID.

References

- [1] T. Suehara, T. Tanabe, LCFIPlus: A framework for jet analysis in linear collider studies, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **808**, 109-116 (2016). <http://dx.doi.org/10.1016/j.nima.2015.11.054>
- [2] H. Qu, L. Gouskos, Jet tagging via particle clouds, Physical Review D **101** (2020). <http://dx.doi.org/10.1103/PhysRevD.101.056019>
- [3] S.M. Bedeschi Franco, Gouskos Loukas, Jet flavour tagging for future colliders with fast simulation, The European Physical Journal C (2022). <https://doi.org/10.1140/epjc/s10052-022-10609-1>
- [4] H. Qu, C. Li, S. Qian, Particle transformer for jet tagging (2024), 2202.03772, arXiv:2202.03772 [hep-ph]
- [5] The ILD Collaboration, International Large Detector: Interim Design Report (2020), arXiv:2003.01116 [physics.ins-det]
- [6] U. Einhaus, CPID: A Comprehensive Particle Identification Framework for Future e^+e^- Colliders (2023), 2307.15635, arXiv:2307.15635 [hep-ex]