

Status on ML@ZHH: Flavor Tagging and Overlay Removal

ILD Analysis/Software Meeting | 2025/1/15



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG



I.A Flavor Tagging Updates

- Starting point (M. Meyer):
 - ParticleNet implemented
 - Using ILD@ILC500 DBD **full-simulation** flavortag datasets (6-jets of same flavor)
 - 2M jets per flavor, 10M total; much less than what PartT is commonly trained on (e.g. JetClass dataset, 100 M in total)

jet constituents: coordinates

$\Delta\eta, \Delta\Phi$

jet constituents: features

$\Delta\eta, \Delta\Phi$

$\log(p_T), \log(E), \log(p_T/p_T^{\text{jet}}), \log(E/E^{\text{jet}}),$
 $\vec{p}^{\text{track}} \cdot \vec{p}^{\text{jet}/p_{\text{jet}}}$

ΔR

q

isElectron, isMuon, isChargedHadron,
isNeutralHadron, isPhoton

impact parameter & significances

track used in PV?

lepton related variables

pid variables

$E_{\text{HCAL}}/E_{\text{HCAL}+\text{ECAL}}$

χ^2/ndf

28 input features

secondary vertices: coordinates

$\Delta\eta, \Delta\Phi$

secondary vertices: features

$\Delta\eta, \Delta\Phi$

$\log(p_T), E_{\text{SV}}/E_{\text{jet}}, E_{\text{SV}}$

η

m_{SV}

$N_{\text{tracks in SV}}$

χ^2/ndf

impact parameters & significances

$\cos(\text{flight direction}_{\text{SV}}, \vec{p}_{\text{SV}})$

14 input features

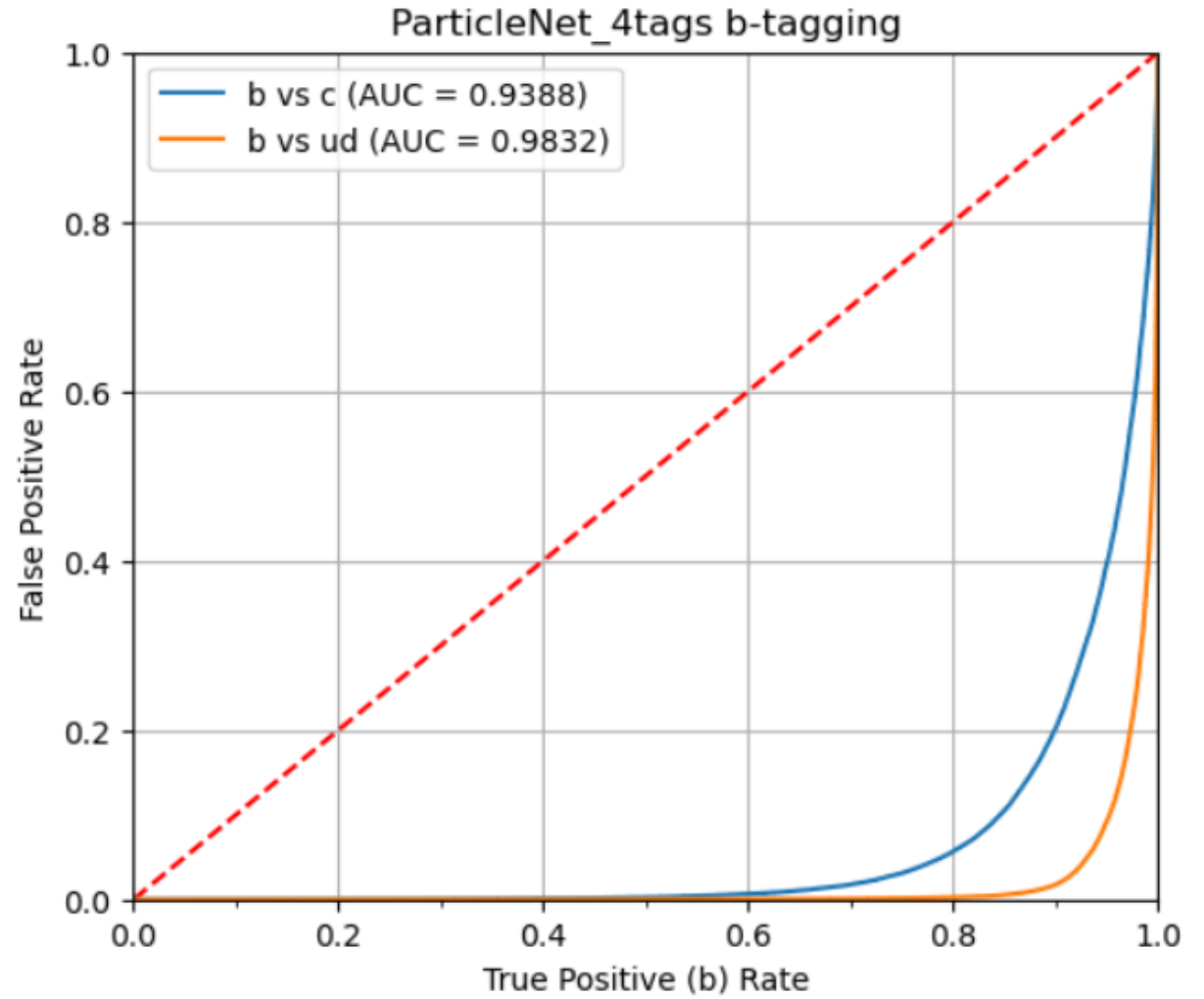
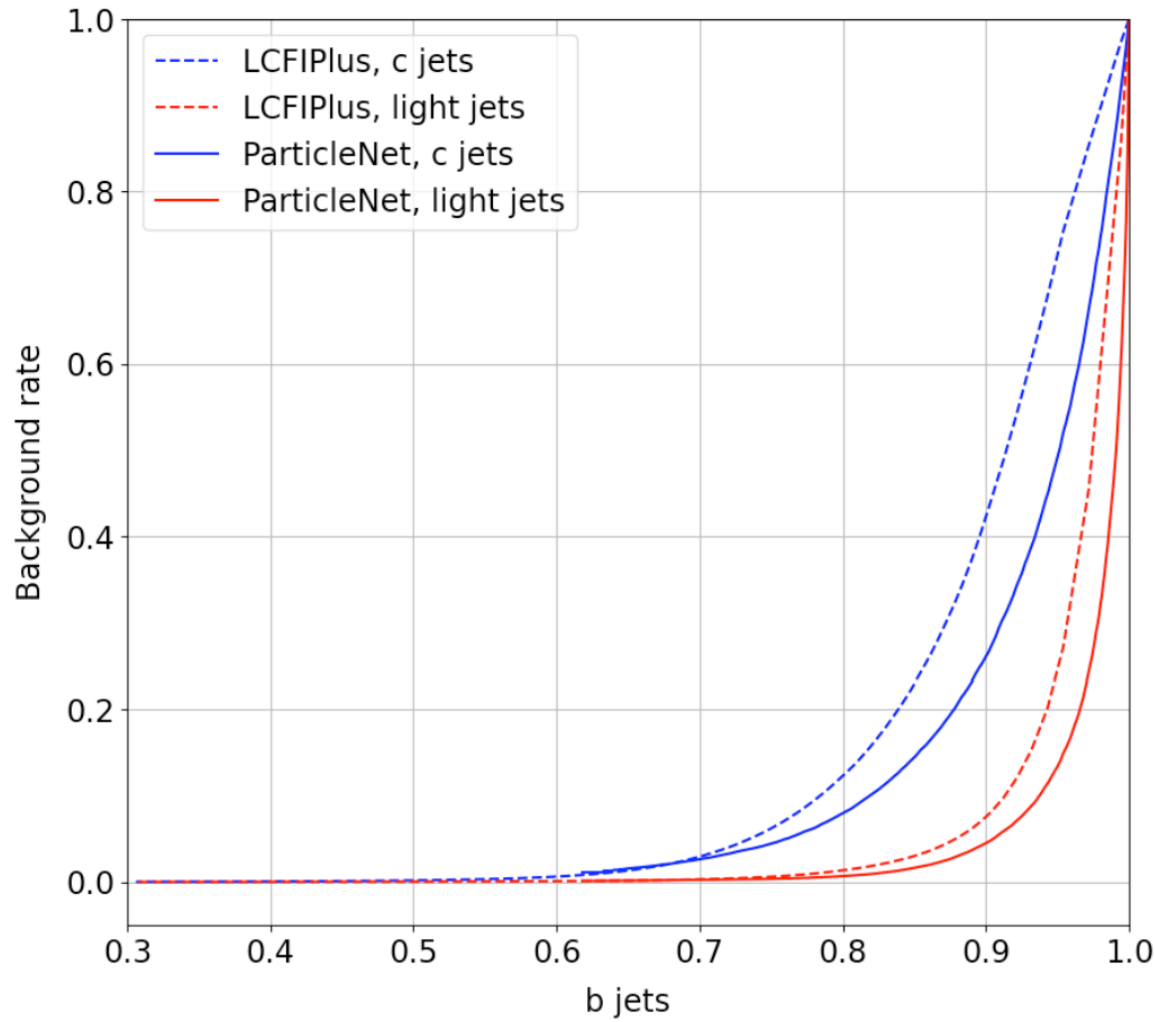
2 SVs & all jet constituents
considered, no ordering of inputs

Implemented input features for jet flavor tagging / M. Meyer

I.A Flavor Tagging Updates

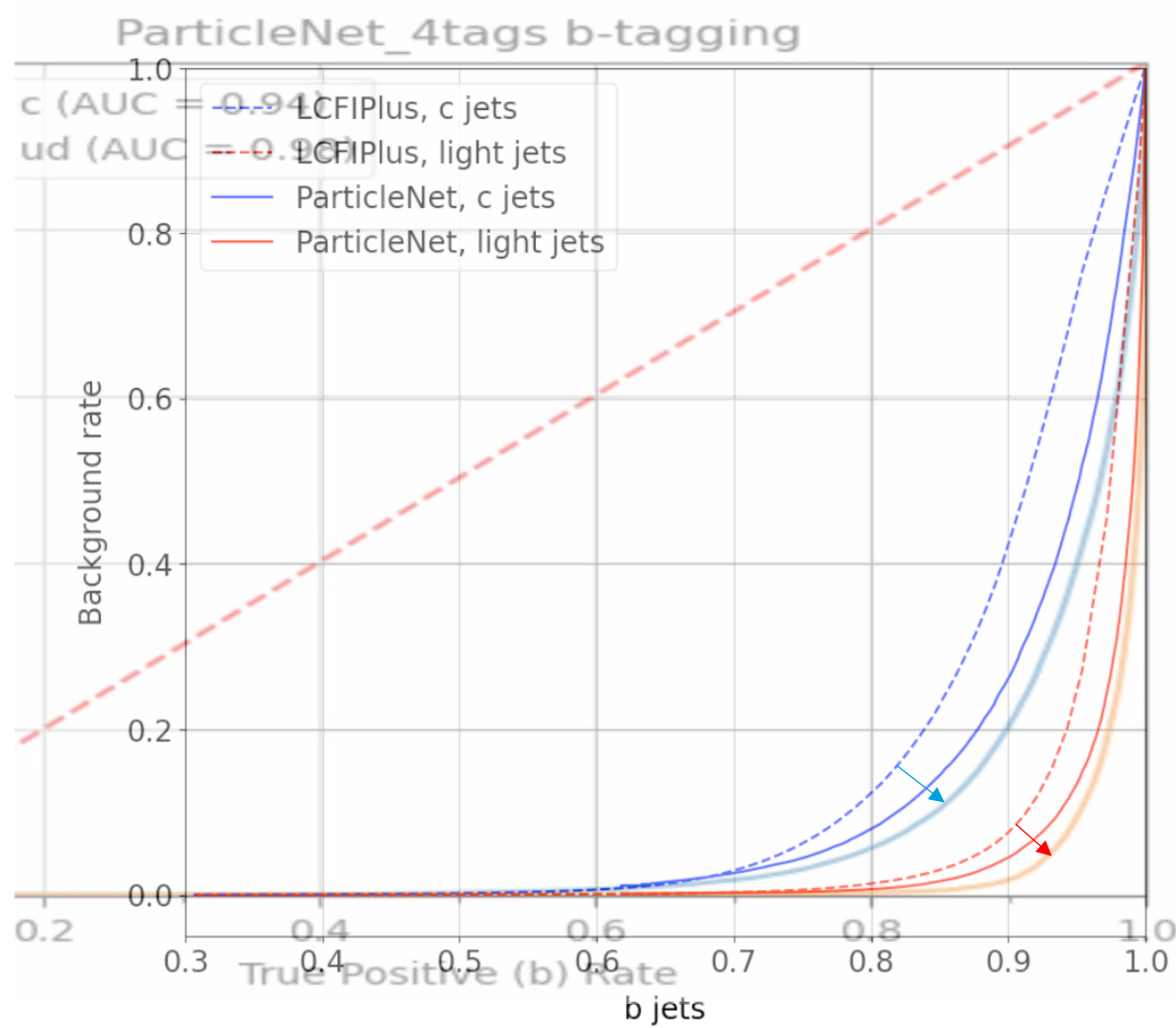
- Extended existing FlavorTag framework
 - Using PyTorch data loaders and transformations:
 - Customizable transformations of input features and labels
 - Configurable sampling and loss reweighting; automatic oversampling of the minority class, etc.
 - Can now support more architectures: ParticleTransformer added ([PELICAN](#) in the future?)
 - Configuration now more in-line with the [weaver](#) tool
 - Overhauled data conversion from ROOT to HDF5
 - Faster deployment due to multiprocessing
 - Jupyter notebook examples, documentation
 - Added [ComprehensivePID](#) as input (combines dEdx, TOF, cluster shapes; by Uli Einhaus) and s-tag [however, following plots are still without CPID]
- Supported by Uli Einhaus and Thomas Madlener

I.B Flavor Tagging - ParticleNet



Lower right is better

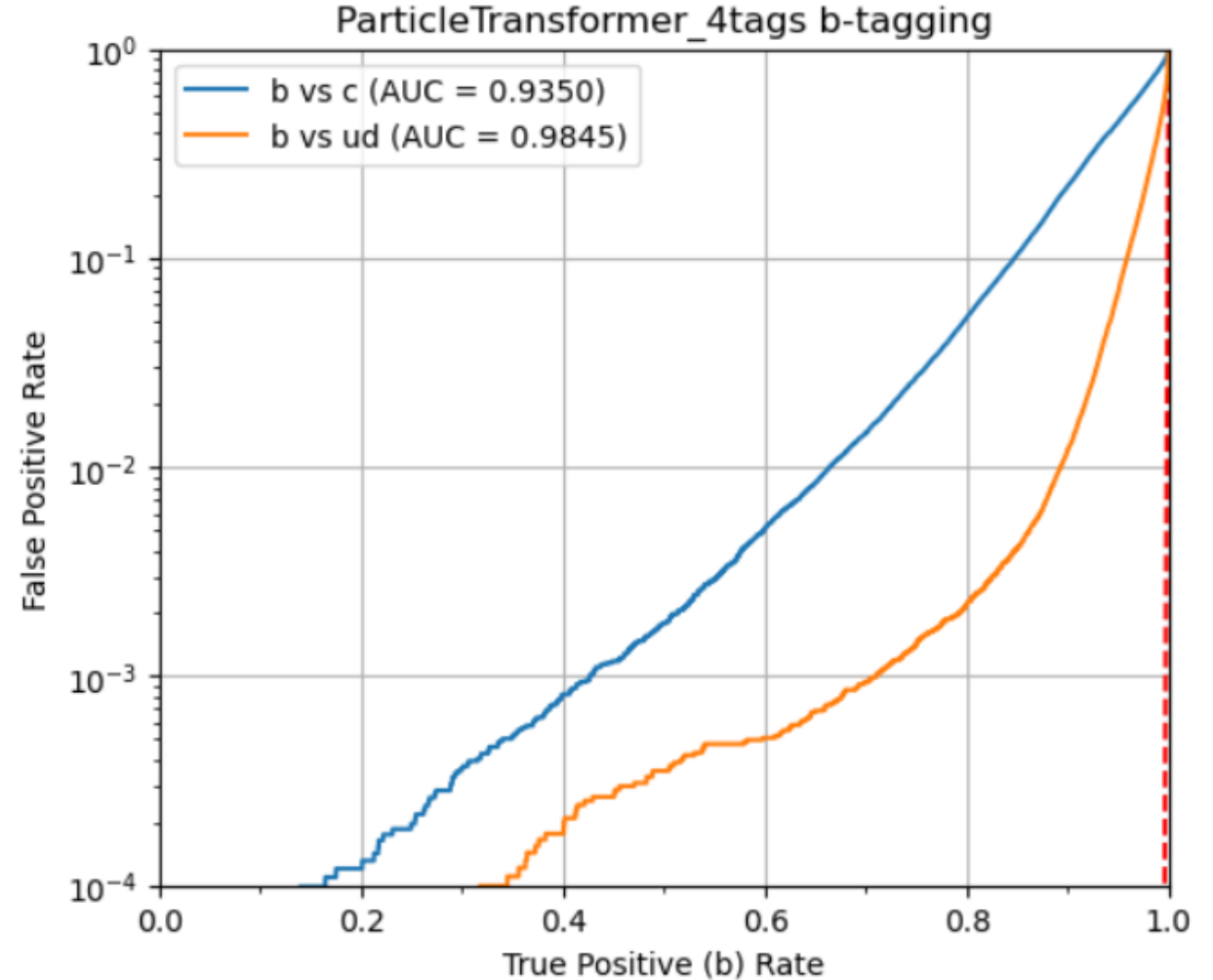
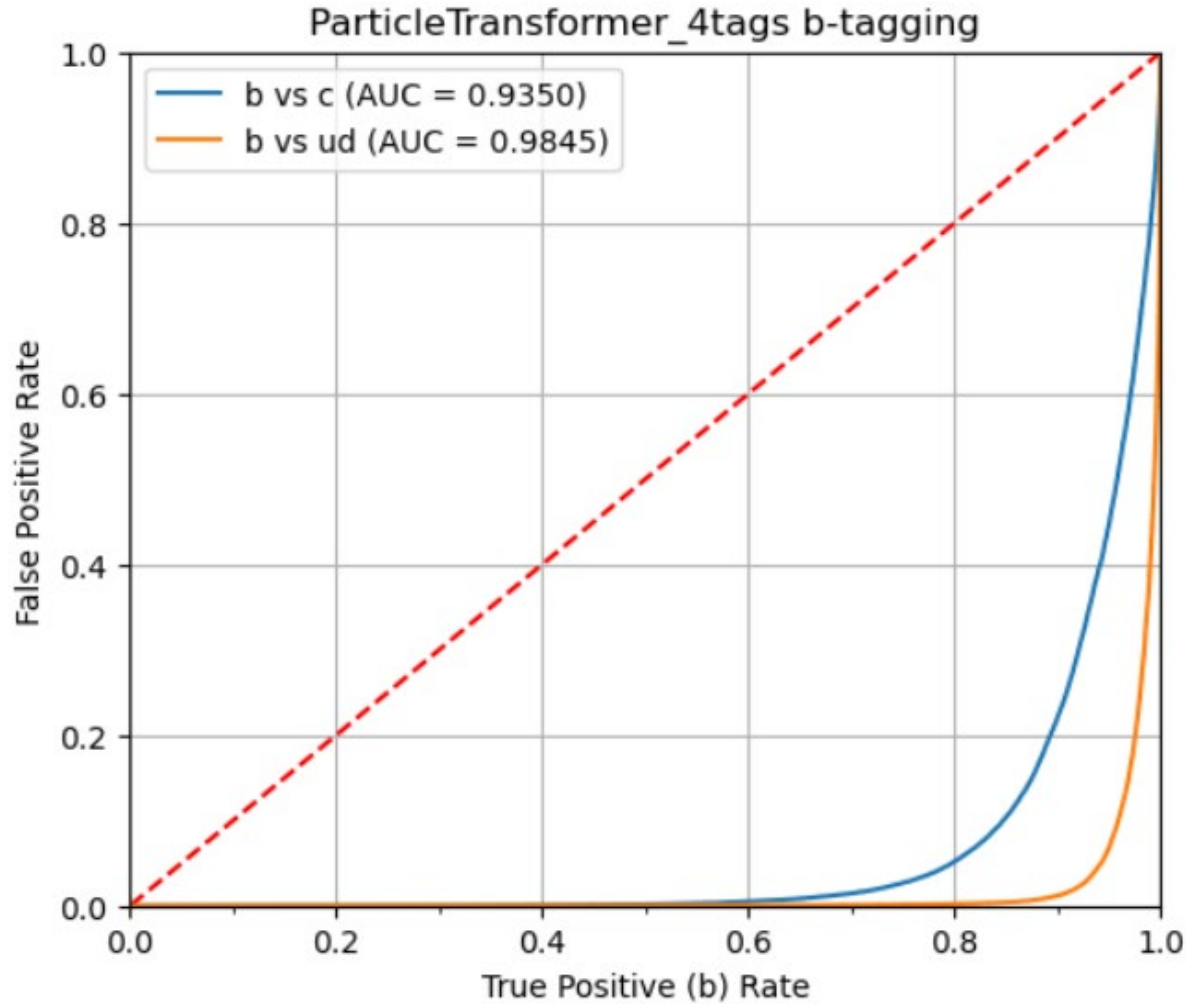
I.B Flavor Tagging - ParticleNet



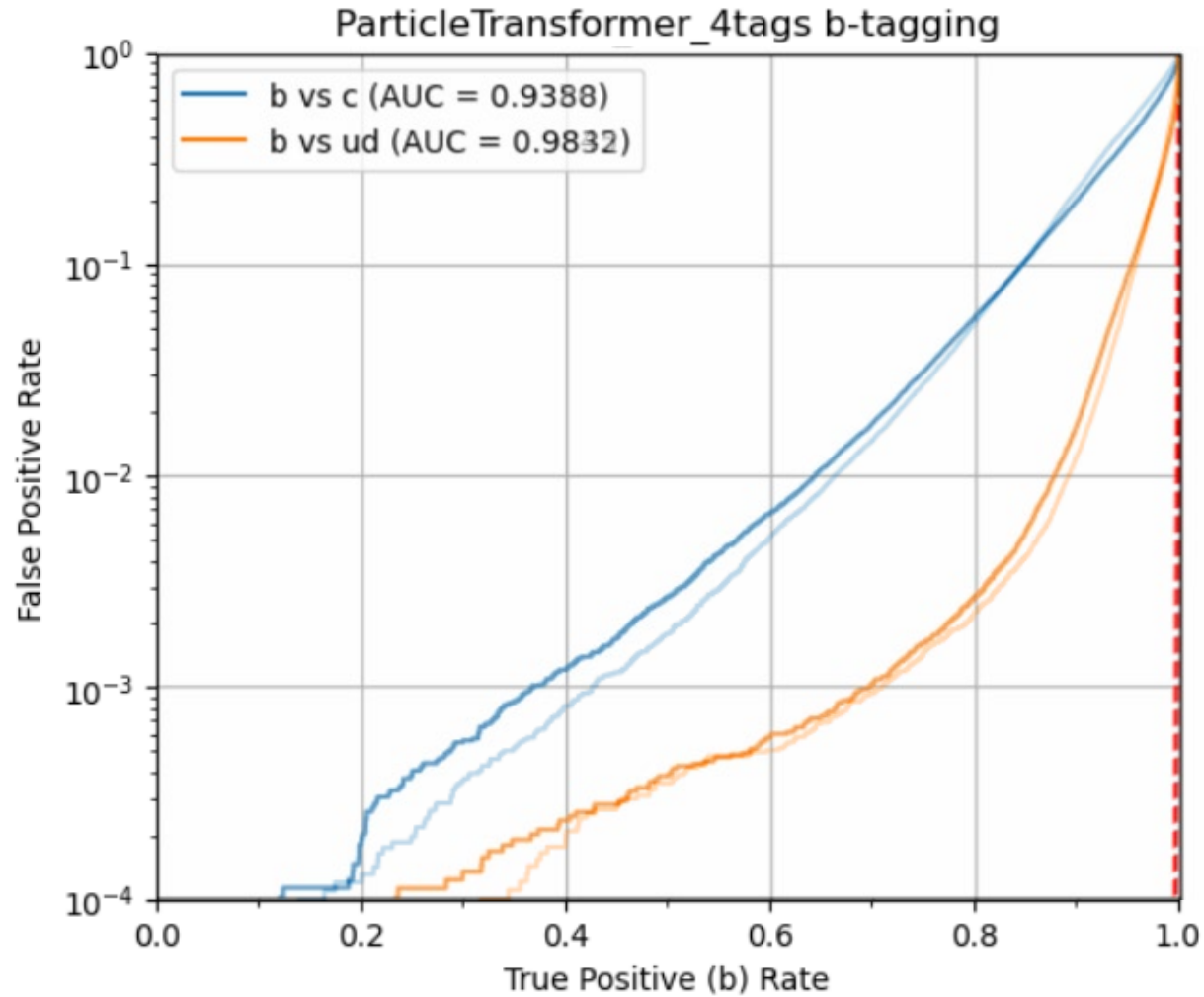
Comparison of b-vs-c at
ILD@ILC500

ParticleNet (current): light-blue
LCFIPlus: dark-dashed blue

I.C Flavor Tagging – ParticleTransformer



I.D Flavor Tagging – ParT vs ParticleNet



Light color: ParT
Dark color: ParticleNet

I.E Flavor Tagging – Outlook

- Analyze physics performance on signal/background datasets
 - ✓ Fast simulation using SGV on 500 GeV flavortag, ZHH and ZZH datasets for comparison on “common ground” (flavortag sample from older MC production)
 - Find b-tagging working point that was used in last ZHH analysis for comparison
 - **ca. 10% improvement per jet** anticipated
 - Target: for ZHH analysis, maximum efficiency (4 b jets) at approx. same background rejection is desired

II Overlay Removal

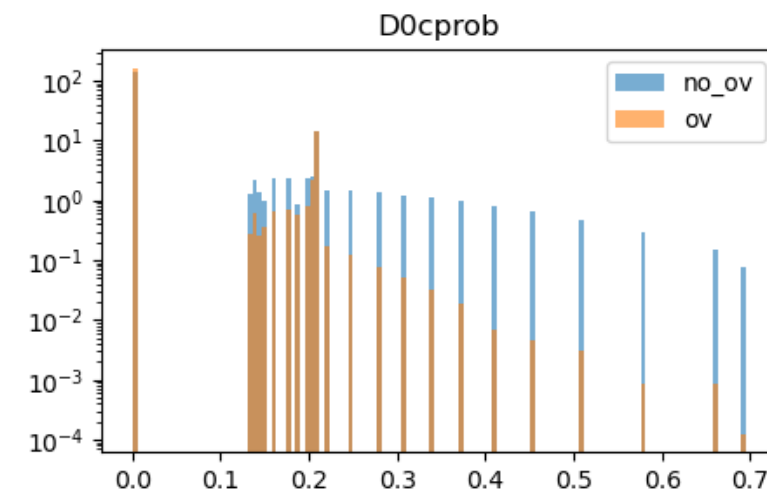
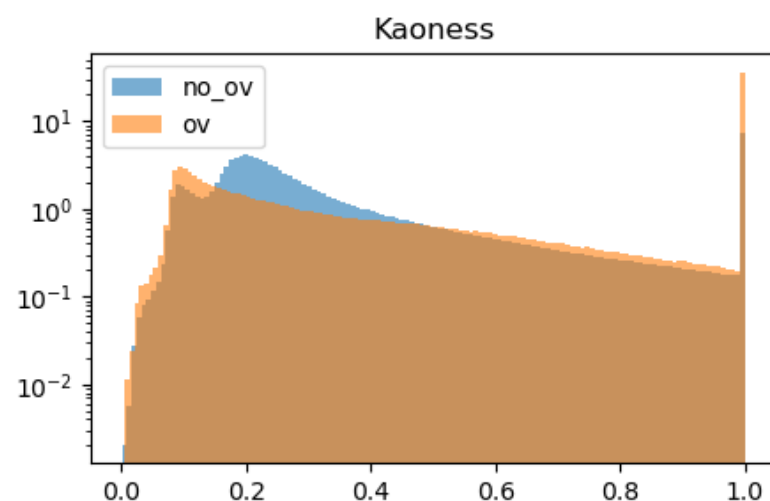
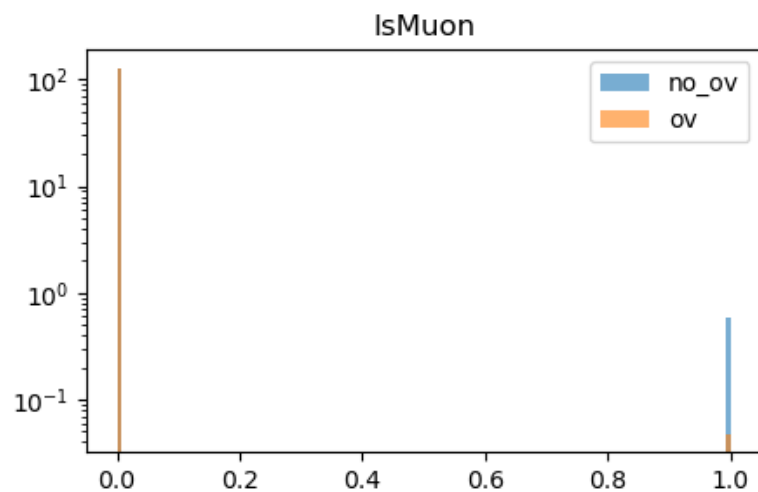
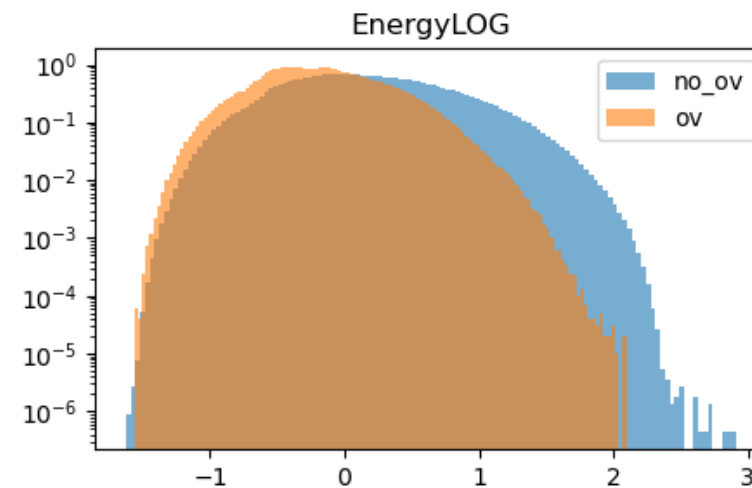
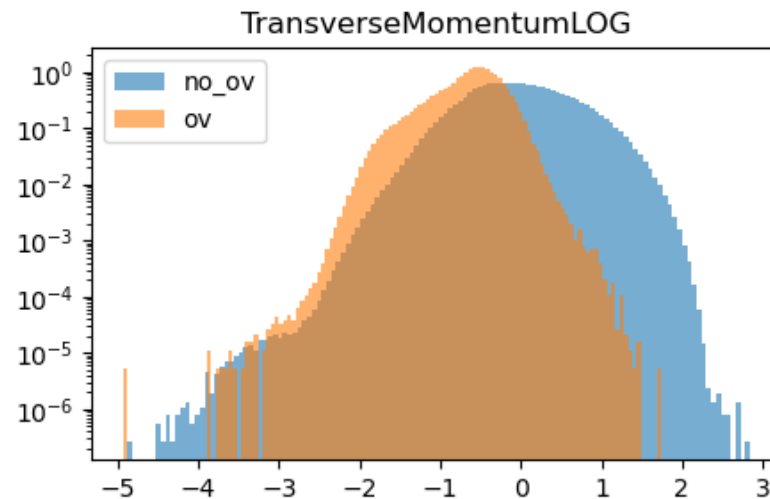
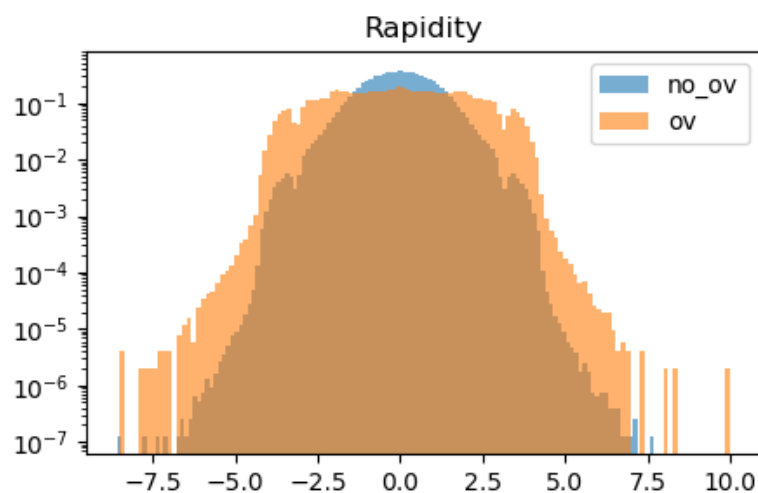
- Idea: use same input data as flavor tagging dataset
- At the moment: some jet-related input features removed
 - Relative phi, rapidity etc.
- Added a label `PFOisOverlay`
- Tested some toy models
- Sent framework and data to Dimitris for cross-check / more ideas

isOverlay

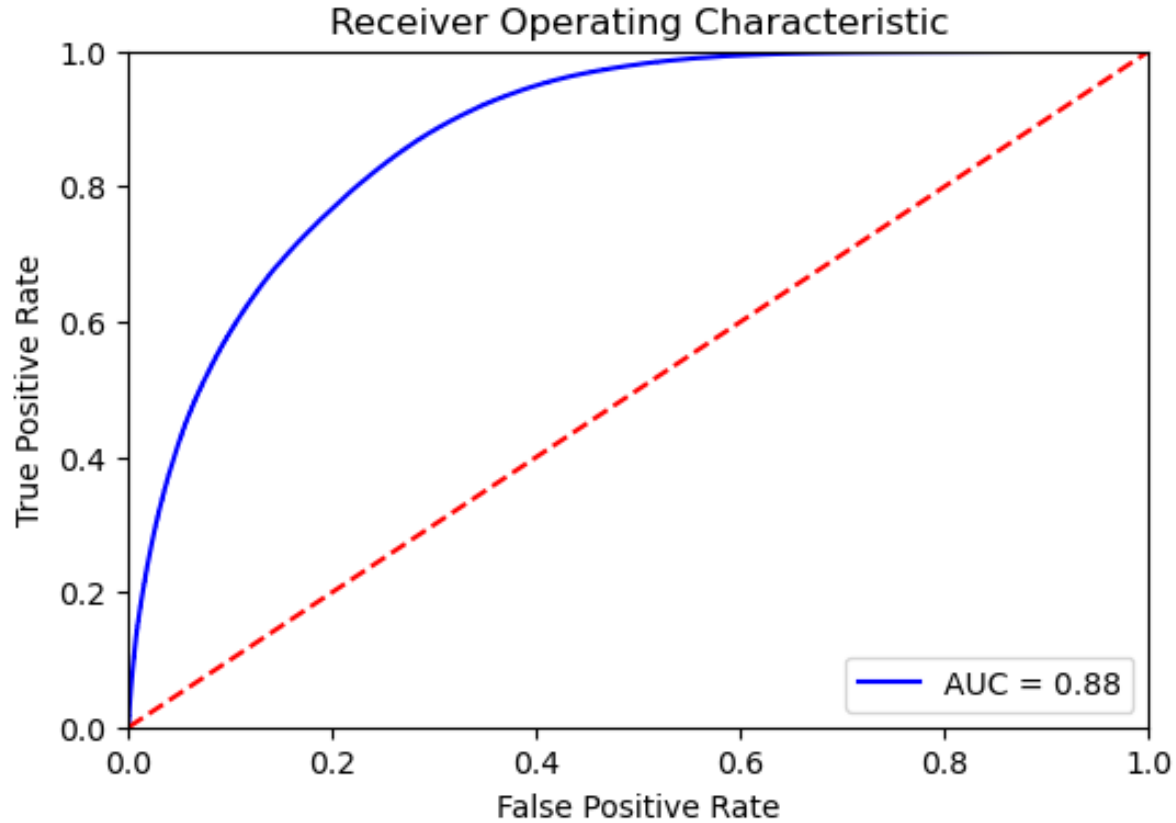


isOverlay:
0: 94.63%
1: 5.37%

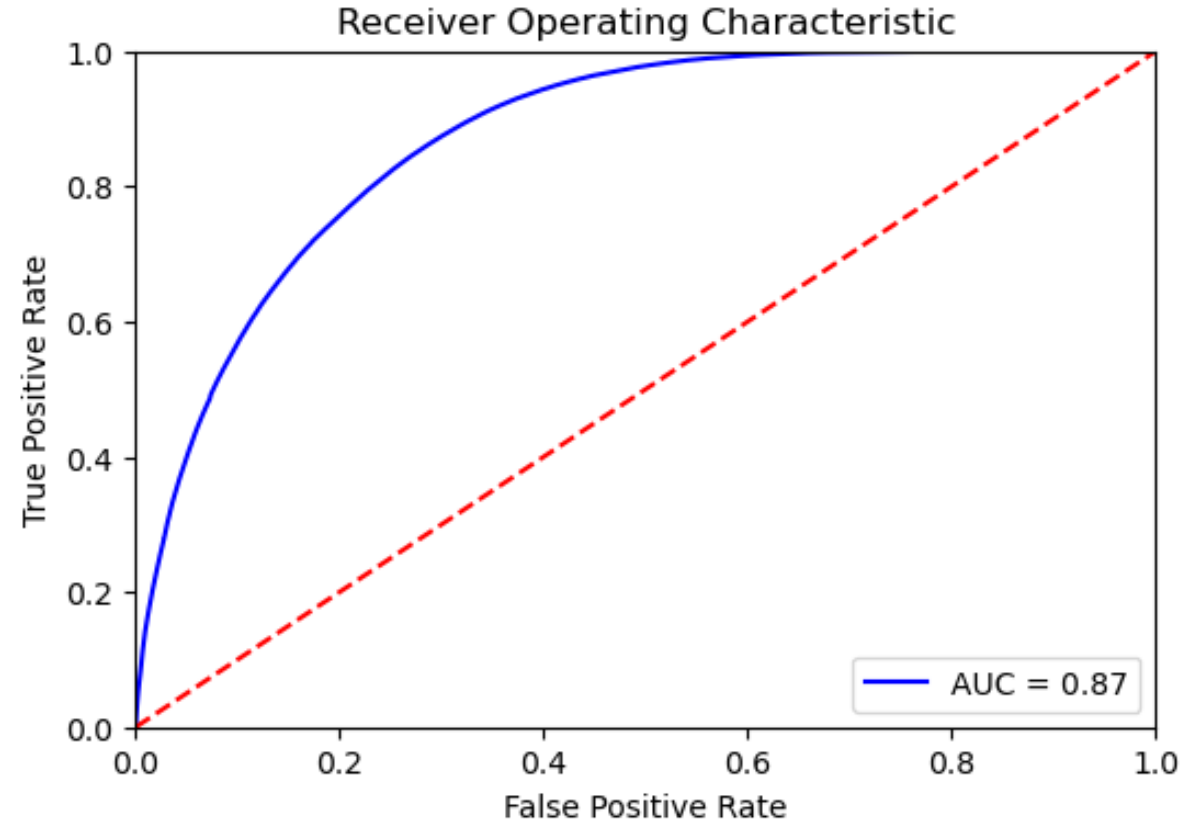
II Overlay Removal – Example features (standardized)



II Overlay Removal



MLP@ PyTorch



GradientBoostingClassifier @ sklearn
(very similar performance just with
logistic regression)

II Overlay Removal

➤ Next steps:

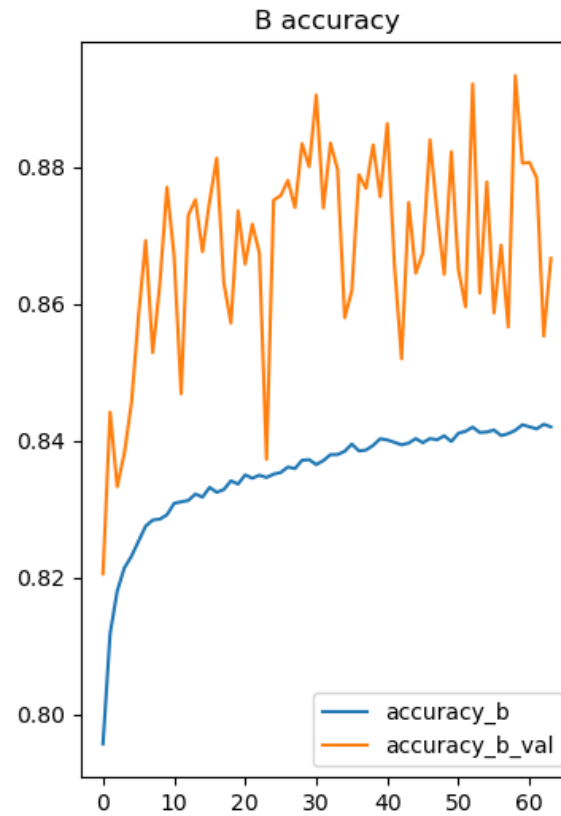
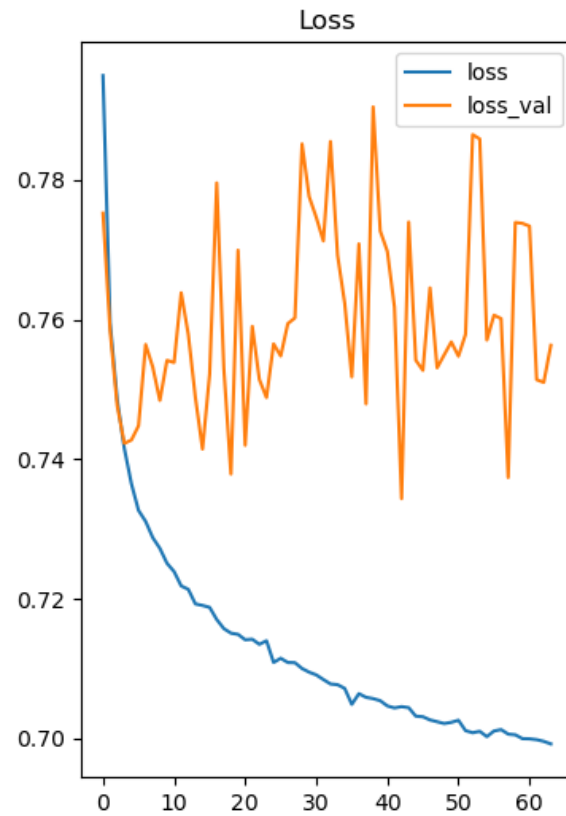
- Compare to “traditional way” of overlay removal (jet clustering to beam jets)
 - ROC curve
 - Physical distributions, influence on physics analysis

➤ Caveats / open questions:

- What about IRC safety, esp. when using features that are based on jet information?

Backup

Flavor Tagging – ParticleNet



Flavor Tagging – ParticleTransformer

