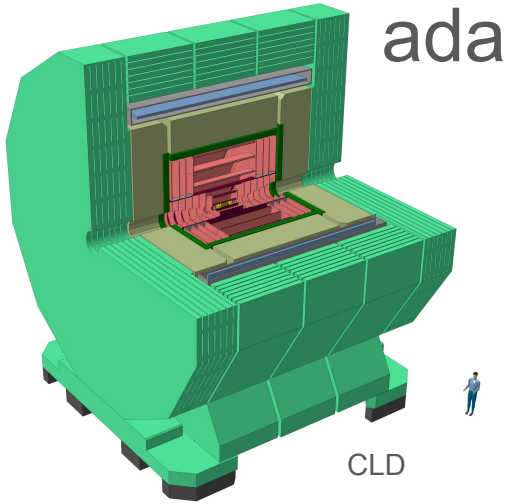
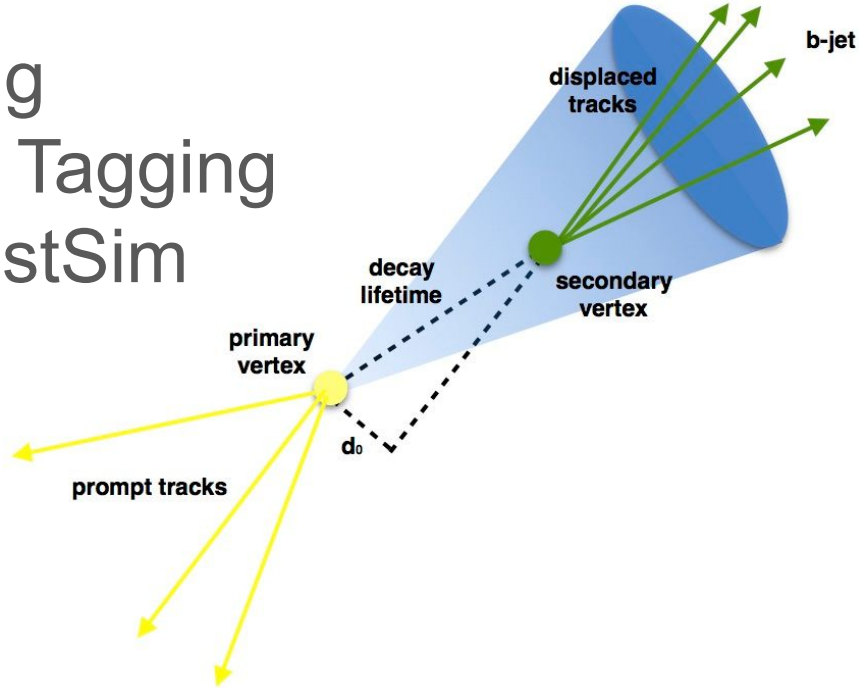
FullSim Jet Flavor Tagging

Implementing
Transformer-based Tagging
adapted from FastSim

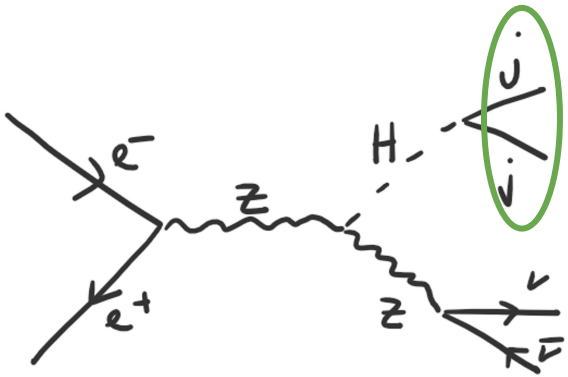
26. June 2024



CLD

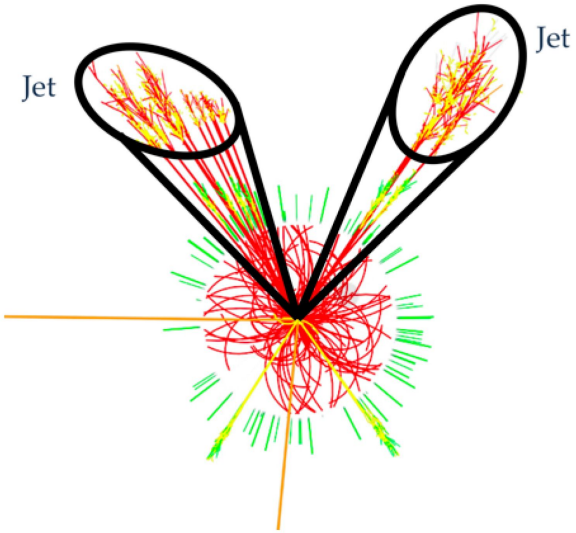


Introduction



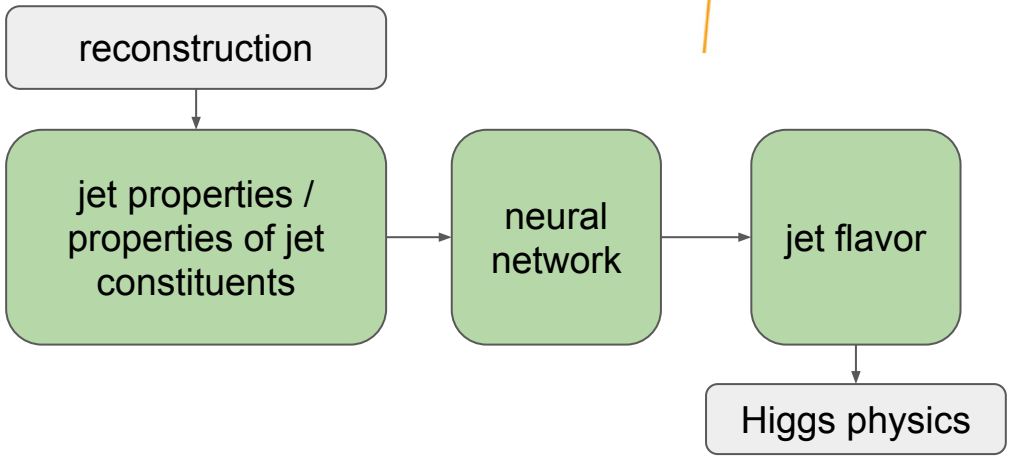
j = jet

jet = narrow cone of many particles produced due to hadronization

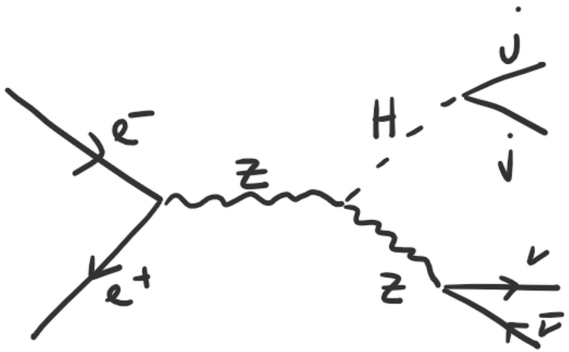


The jet can originate from different sources, e.g.: quarks (u,d,s,c,b) or leptons (τ) or gluons (g)

Jet flavor tagging = which particle caused the jet?



Simulation and data



We use data with 2 jets only!

Jet flavor tagging is done jet based, not event based

~ 3800 jets per flavor

- FastSim on IDEA detector for E=240 GeV

  Tagging done successfully ✅

- FullSim on CLD detector for E=240 GeV (generated by Brieuc)

  Tagging → my task now!

`/eos/experiment/fcc/prod/fcc/ee/test_spring2024/240gev/H*/CLD_o2_v05/rec/*/*/H*_rec_*.root`

The idea of the project

1. Extracting descriptions of jet properties and properties of their constituents (e.g. jet momentum, momentum of jet constituents, d0, z0…)
2. Comparison of FullSim parameters to FastSim parameters
3. Extraction of additional information from FullSim (secondary vertices, V0…)
4. Retraining the ParticleTransformer (neural network) on FullSim data
5. Evaluating the performance
6. Improving the neural network?
7. Implementation of tagger into edm4hep

The idea of the project

today's topic

1. Extracting descriptions of jet properties and properties of their constituents (e.g. jet momentum, momentum of jet constituents, d0, z0…)
2. Comparison of FullSim parameters to FastSim parameters
3. Extraction of additional information from FullSim (secondary vertices, V0…)
4. Retraining the ParticleTransformer (neural network) on FullSim data
5. Evaluating the performance
6. Improving the neural network?
7. Implementation of tagger into edm4hep

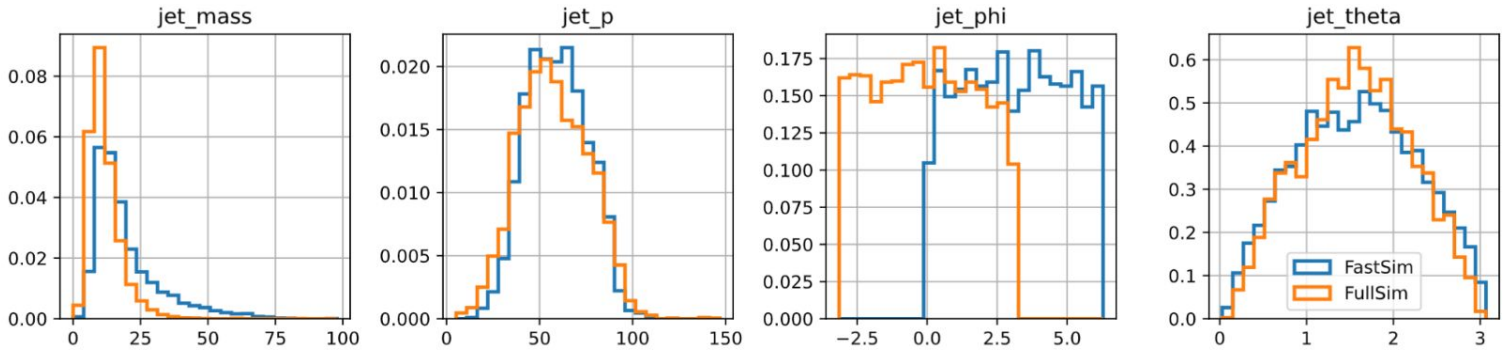
work in progress

Full vs. FastSim observables

- extraction of the the same jet (constituent) properties in FullSim as in FastSim
- 58 overservables in total
- Analyzing H(bb) events (3800 jets originating from b quarks in Full & FastSim)

- modifications made my me (compared to FastSim):
  - time of flight of jet constituents not available in CLD -> filled with 0
  - dN/dx of jet constituents not available in CLD -> filled with 0
  - SIP significance value: FastSim: -9, FullSim: -200 (placement outside of distribution)

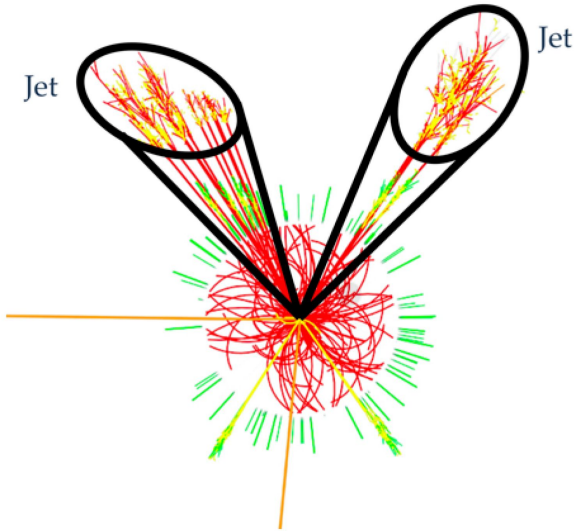
Full vs. FastSim: jet kinematics



- more lower mass events and less higher mass events in FullSim
- similar distributions for momentum and θ
- different convention in φ (FastSim: [0, 2π], FullSim: [-π, π])

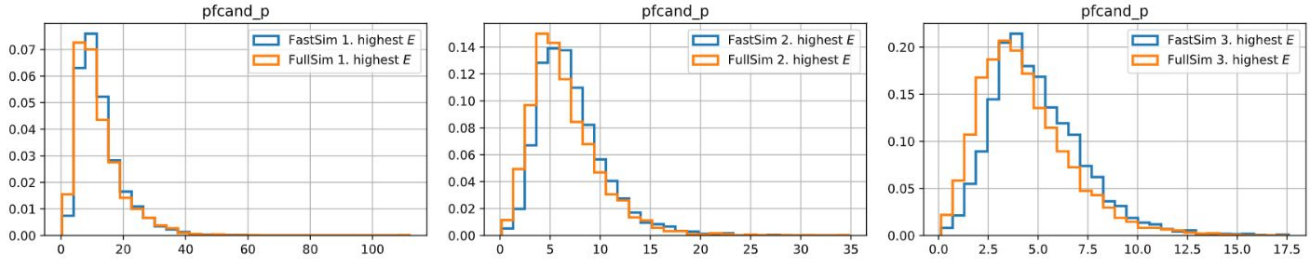
Full vs. FastSim: jet constituents

- Separation into charged and neutral particles
- Analysis of the three leading particles (highest contribution to jet energy)

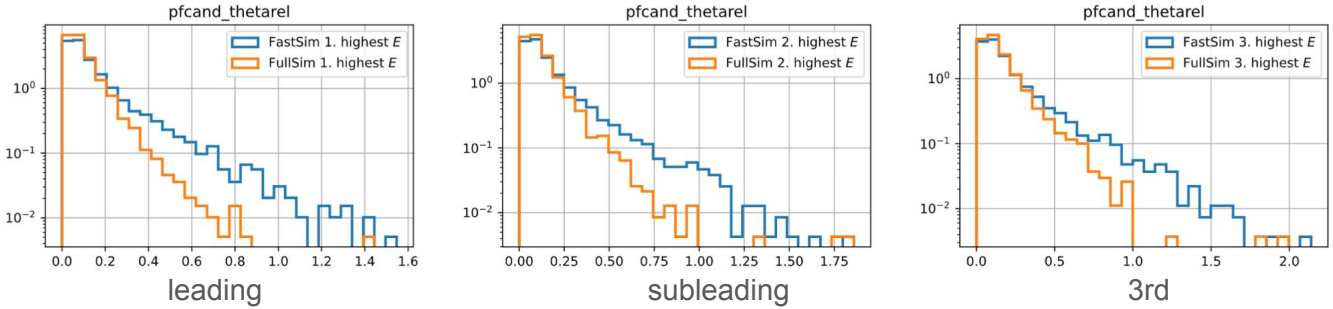


Full vs. FastSim: jet constituents - charged particles (kinematics)

- good agreement in most parameters, e.g. momentum

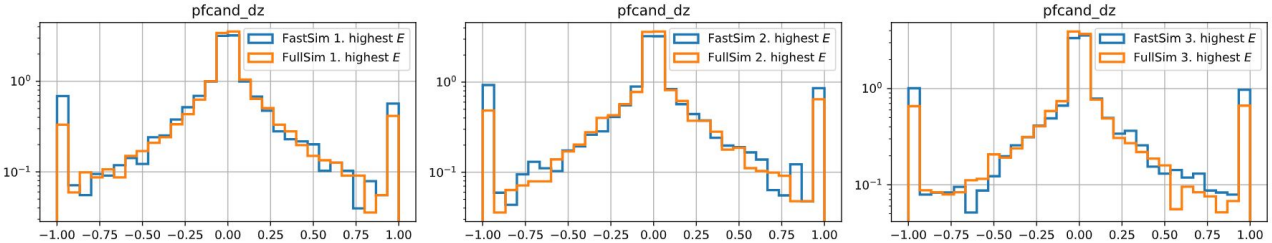


- different tails in $\Delta\theta$ (to be investigated)

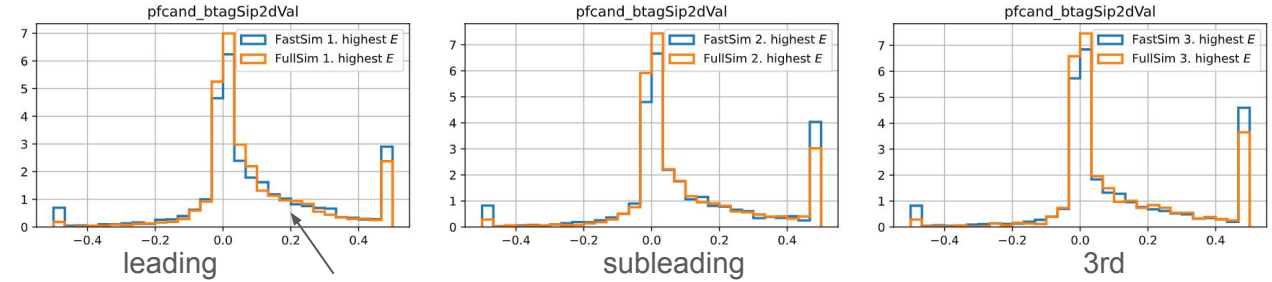


Full vs. FastSim: jet constituents - charged particles (displacements)

- very good agreement

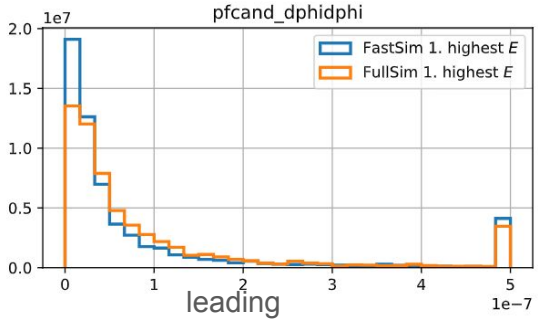


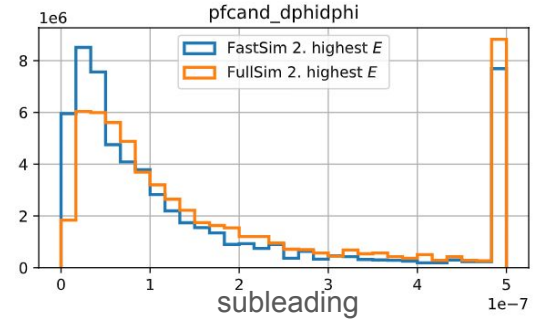
- expected asymmetry from H(bb) events

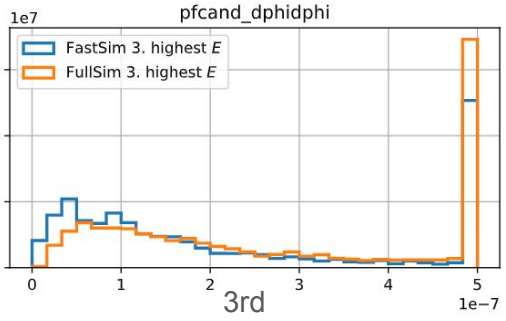


Full vs. FastSim: jet const. - charged particles: cov matrix

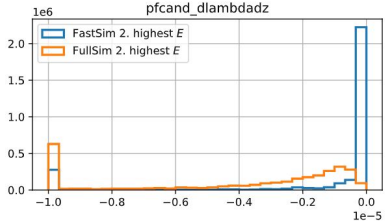
- good agreement among some parameters:

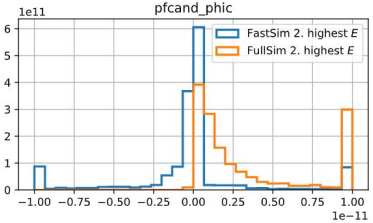






- discrepancy among others:
  - longer tails in FullSim
  - sign conventions?





lambda ⟶ tan(λ)

c ⟶ p_t

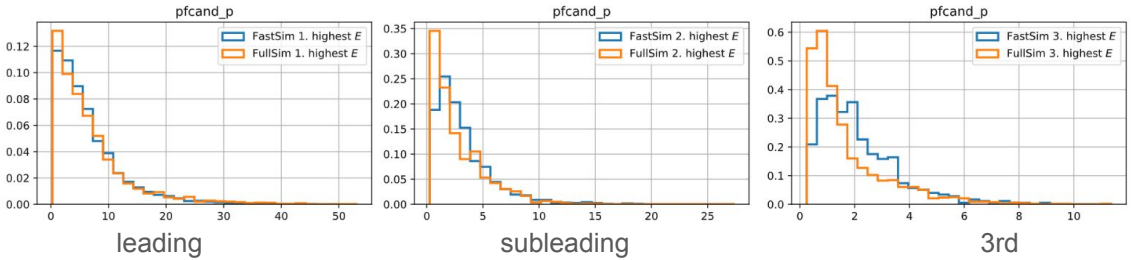
Full vs. FastSim: jet with charged particles: shift from IP (0,0,0) to PV

- we can calculate the primary vertex (PV) of jets with charged particles
- apply correction from (0,0,0) to PV for particles with tracks
- correction applied to:
  - d0 (transverse impact parameter)
  - z0 (longitudinal impact parameter)
  - φ of jet constituent
  - θ of jet constituent
- NO correction for:
  - covariance matrix
  - relative angles (Δφ, Δθ)
- correction to be applied in the future

(showing neutral hadrons, but it looks similar for photons)

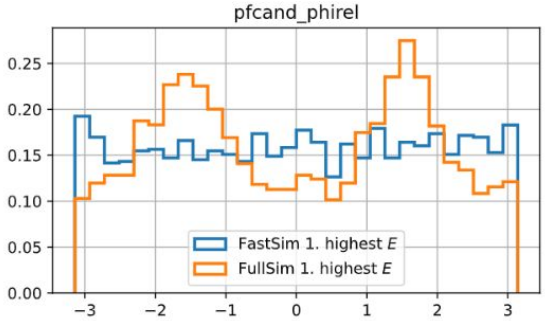
Full vs. FastSim: jet constituents - neutral particles (kinematics)

- good agreement in leading particle, less accordance afterwards



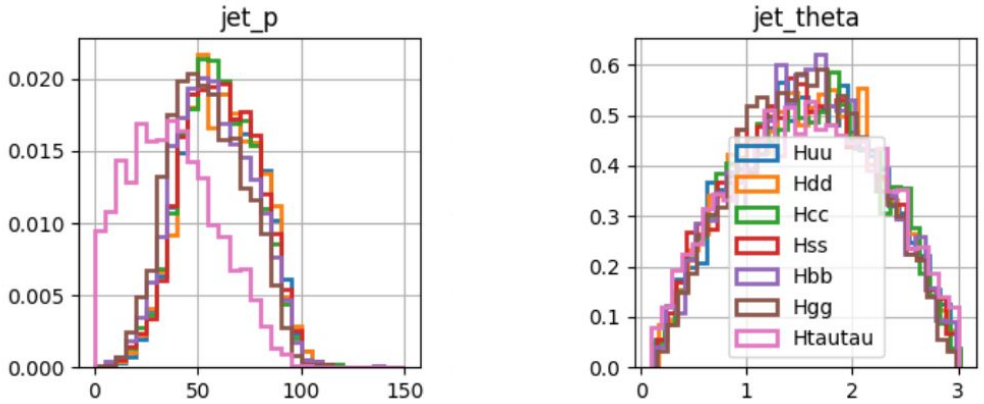
leading subleading 3rd

- Discrepancy in Δφ: Where do the peaks at $\pm\pi/2$ come from?



H to uu/dd/ss/cc/bb/ττ - kinematics

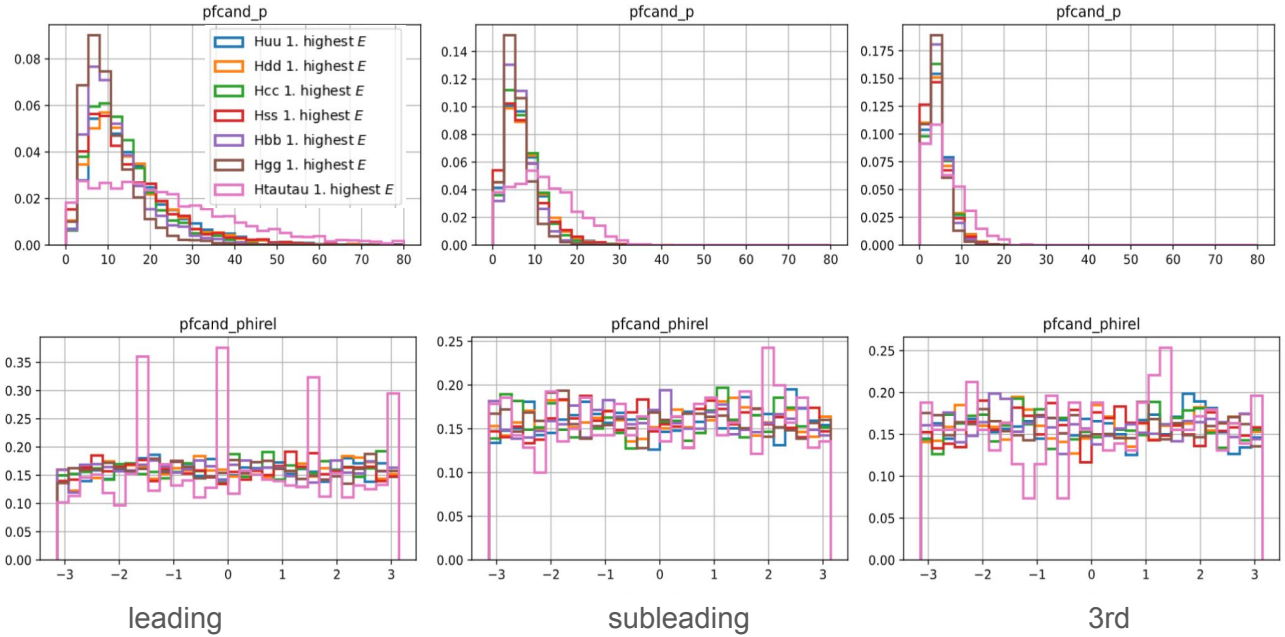
- comparison of all other Higgs decay channels in FullSim instead of comparison between all FullSim/FastSim pairs



- agreement between different quarks, τ behaves different

H to uu/dd/ss/cc/bb/ττ - charged constituents (kinematics)

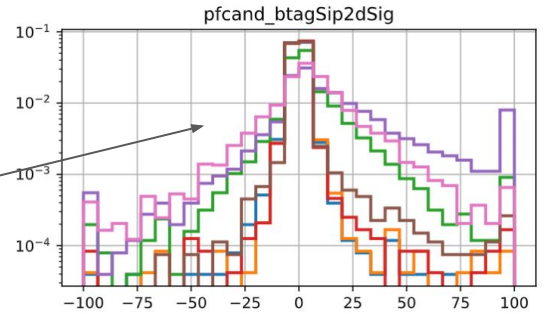
- agreement between channels

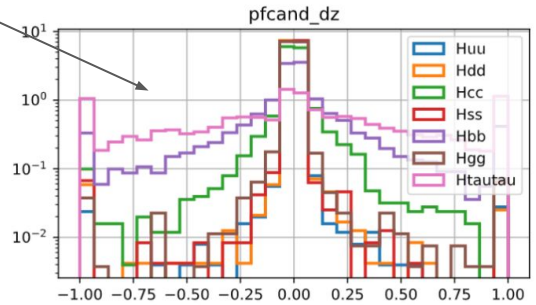


leading subleading 3rd

H to uu/dd/ss/cc/bb/ττ - charged constituents (displacements)

- very good consistency
- asymmetry in 2D SIP for b and c jets (and τ)
- larger displacement for b and c jets (and τ)
- covariance matrices all in accordance

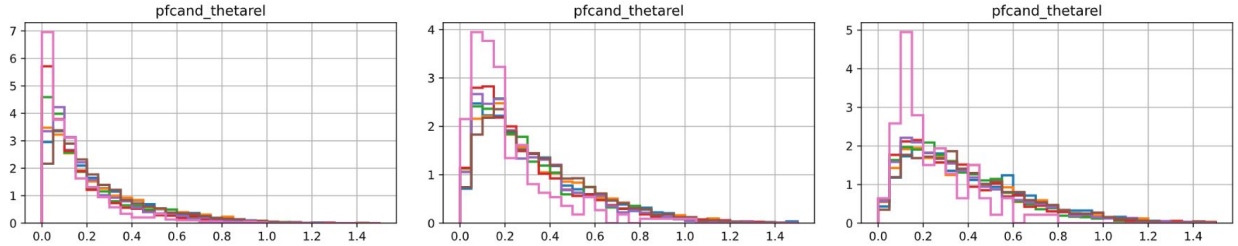




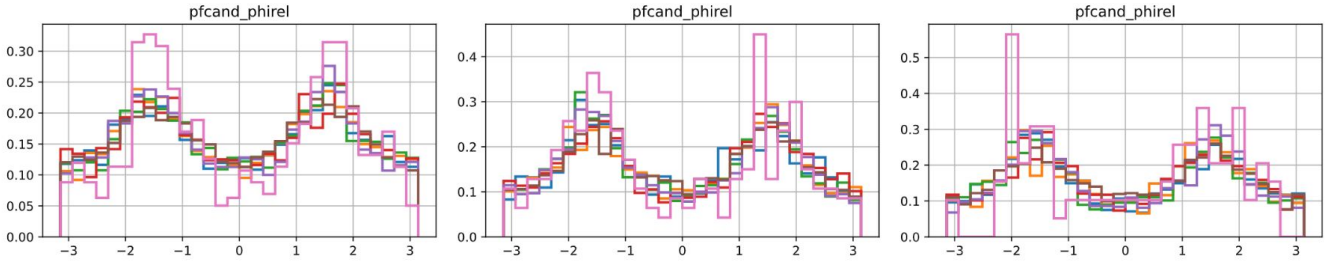
(showing neutral hadrons, but it looks similar for photons)

H to uu/dd/ss/cc/bb/$\tau\tau$ - neutral constituents (kinematics)

- good agreement in most parameters

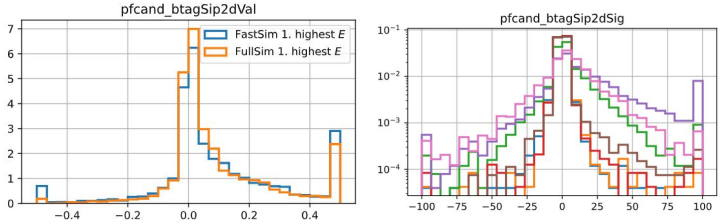


- peaks at $\pm\pi/2$ for $\Delta\varphi$; need for explanation



leading subleading 3rd

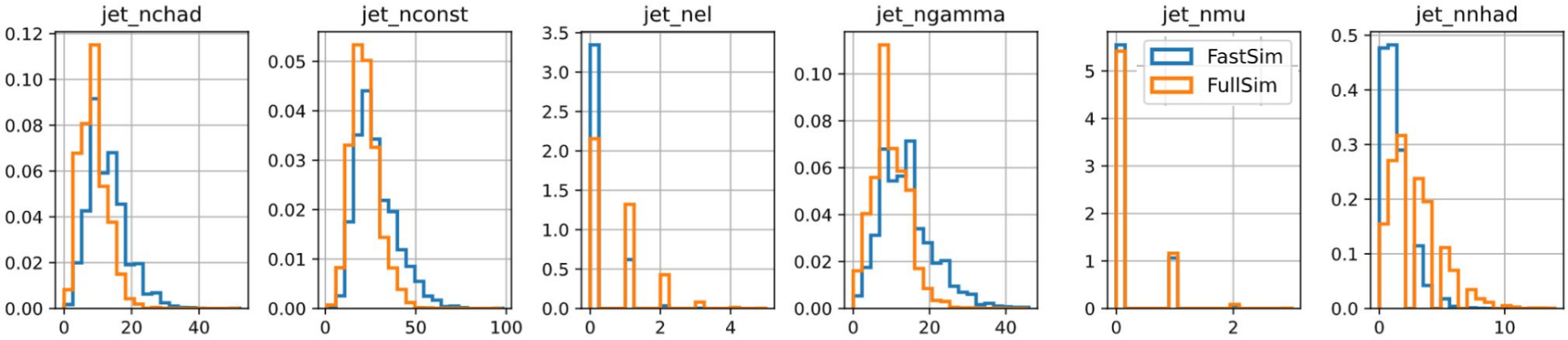
Summary



- overall good agreement between FastSim IDEA and FullSim CLD (tested in H(bb) channel)
- overall good agreement between different Higgs channels in FullSim
- Open tasks/questions:
  - What's happening with Δφ for neutral particles in FullSim? (peaks at $\pm\pi/2$)
  - some discrepancies in covariance matrix between Full and FastSim
  - Task: Apply correction of PV shift to covariance matrix and relative angles
- Next tasks:
  - Better statistics with more FullSim CLD data
  - comparison between FastSim CLD and FullSim CLD
  - Adding more information, e.g. secondary vertices
  - Evaluation at different energy (E=365 GeV)
  - Training the neural network with large FullSim CLD dataset to compare tagger performance with FastSim

Backup slides

Full vs. FastSim: particle multiplicities in jet



Dip in 3D SIP sign. but not in 2D SIP sign.