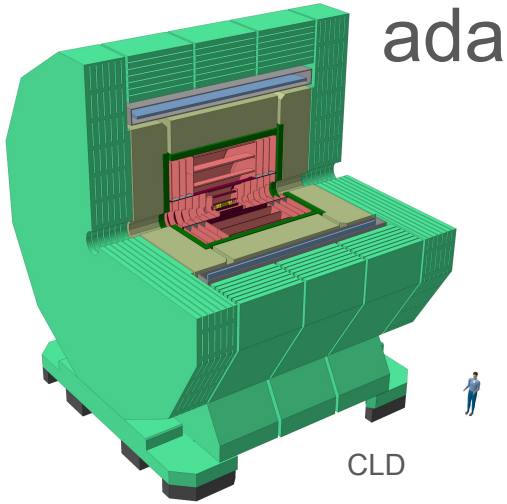
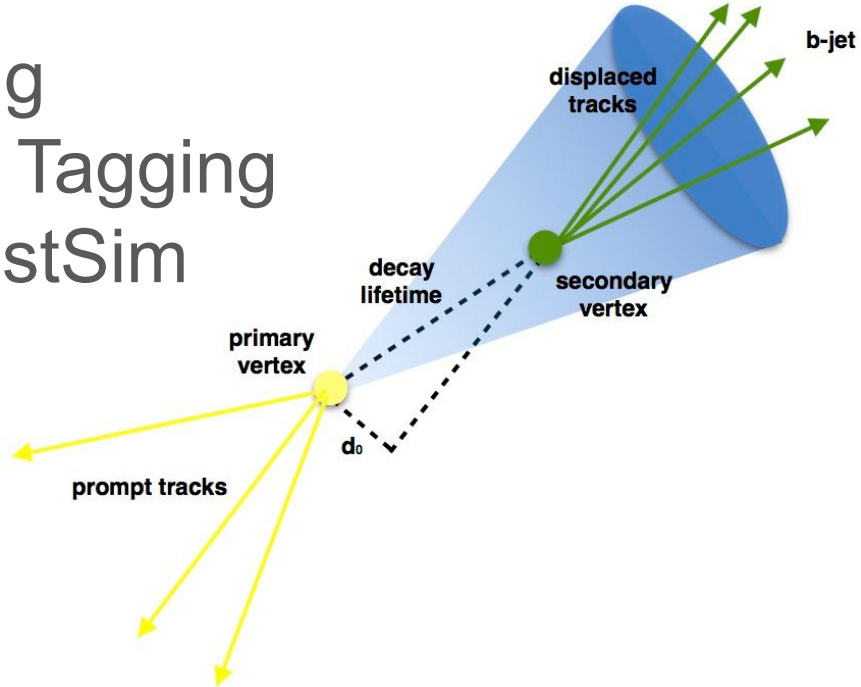
FullSim Jet Flavor Tagging

Implementing
Transformer-based Tagging
adapted from FastSim

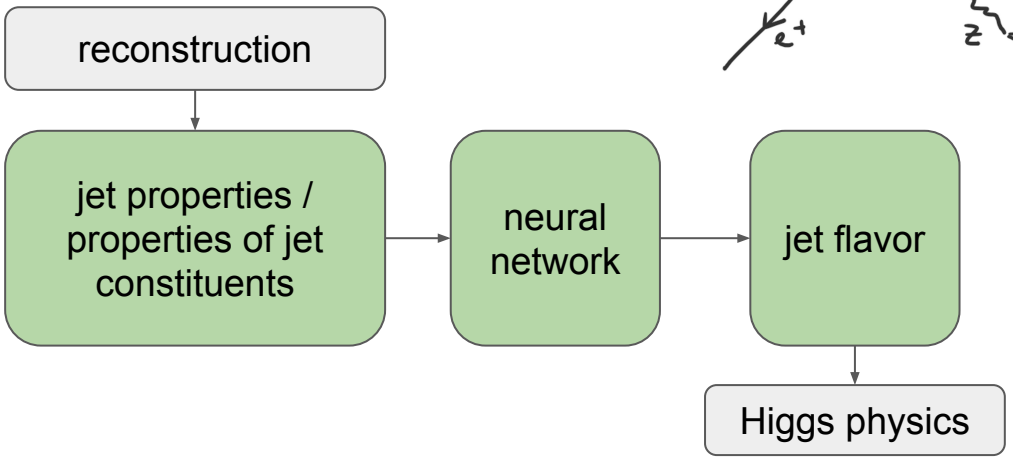
24. July 2024

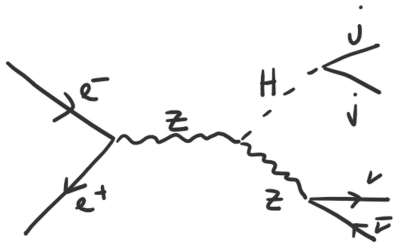


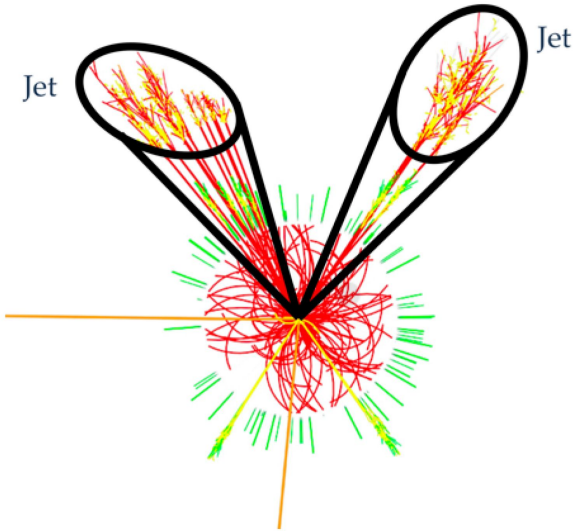
CLD



Reminder







Presentation end of June:

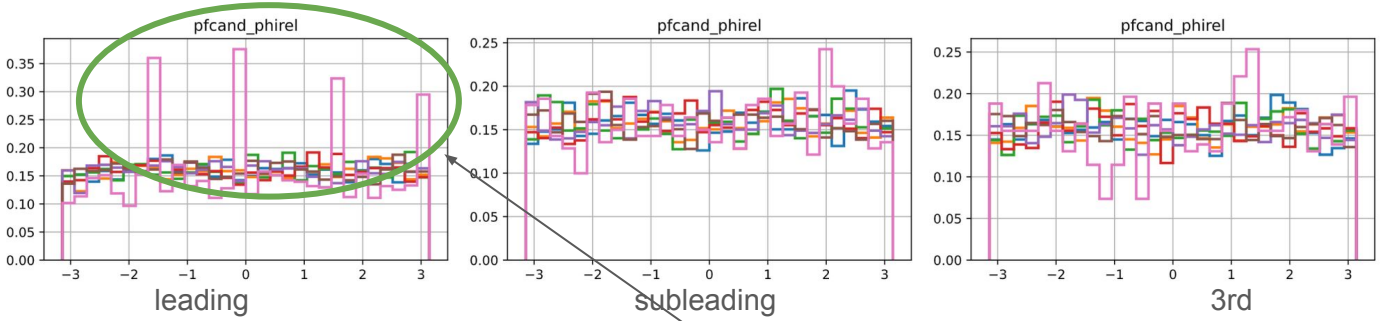
- first comparison between FastSim IDEA and FullSim CLD
- comparison of different Higgs channels in FullSim
- we have seen discrepancies in Δφ

Today's Agenda

1. Explanation of Δφ discrepancies
2. FastSim CLD vs. FullSim CLD & FullSim in different Higgs channels with more data (Thanks Brieuc!)
3. ML Training: Jet Tagging on FullSim CLD data 
4. Investigation on lost charged particles & unassigned tracks to PFOs

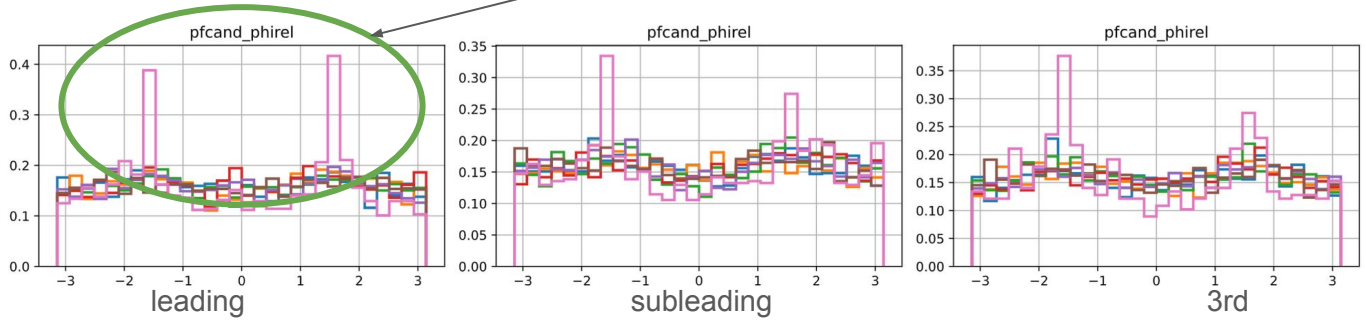
* peak at 0 due to single particle jets

1. Δφ discrepancies in H → ττ



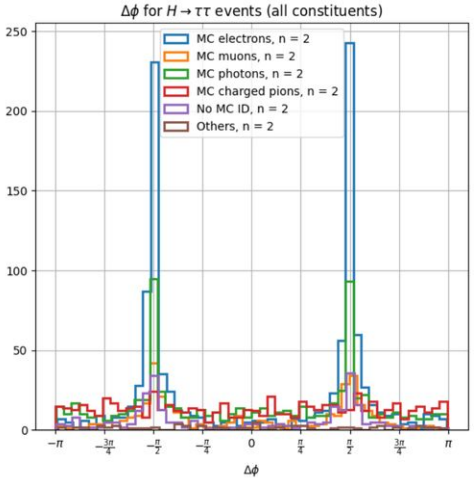
Δφ for charged constituents. Pink is τ

Peaks at ±π/2?



Δφ for photon constituents

Explanation for peaks at ±π/2: Bremstrahlung



For n=2, the peak occurs for electron and photons

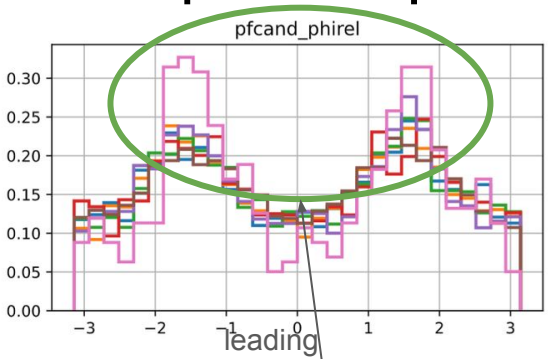
$$\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau \rightarrow e^- + \bar{\nu}_e + \nu_\tau + \gamma$$

Electron emits a Bremsstrahlung photon
⟶ same θ angle due to magnetic field along z
⟶ φ at interaction point also very similar

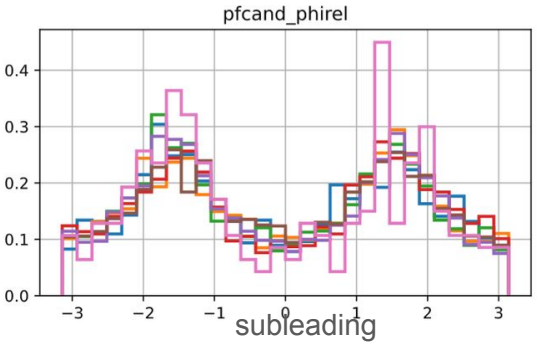
$\Delta\varphi = \arctan(p_x/p_y)$ after rotation of constituent of φ and θ jet angles
⟶ if $p_x << p_y$ ⟶π/2

Find mathematical deviation of ±π/2 limit attached on indico

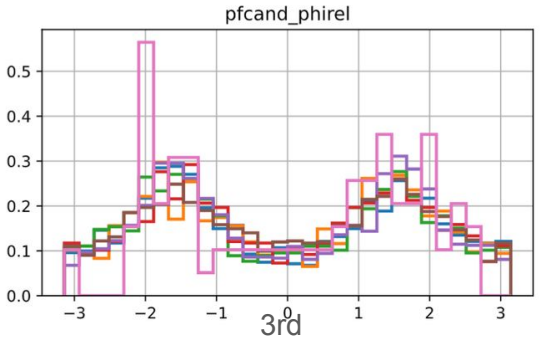
1. Δφ discrepancies in H → ττ



leading



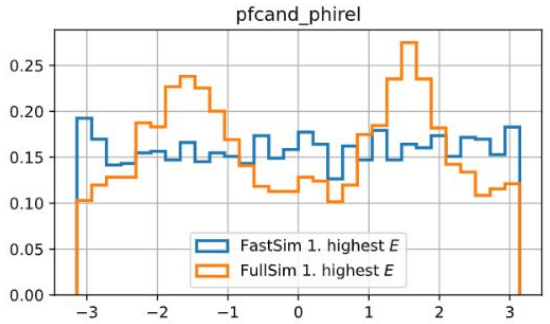
subleading



3rd

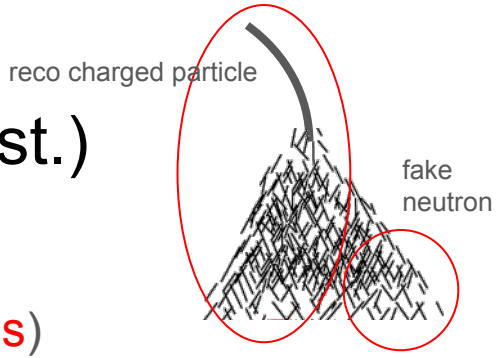
Δφ for neutral constituents

Where do these peaks come from?



Δφ at Hbb FastSim vs. FullSim

Explanation for peaks at ±π/2 (neutral const.)



- particles with similar θ and similar φ as jet
- these are wrongly reconstructed neutrons! (fake neutrons)
- **high energetic charged particles get reconstructed with an additional neutron**
- high energetic charged particles are leading momentum of the jet ⟶ dominate direction (and angles) of jet

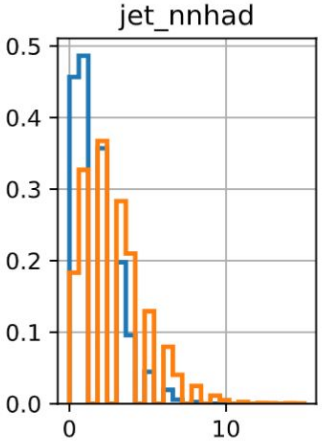
Example of fake neutron event in n=2 jet:

(same) parents: [15 15]

MC PID: [-211 -211]

reco pid: [-211 2112]

momentum: [59.51 4.67]



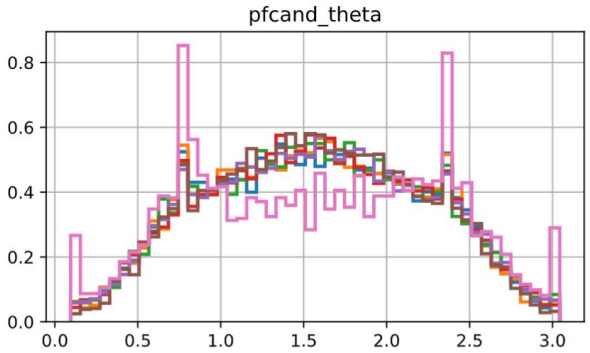
blue: FastSim,
orange FullSim:
higher neutral hadron multiplicity in FullSim

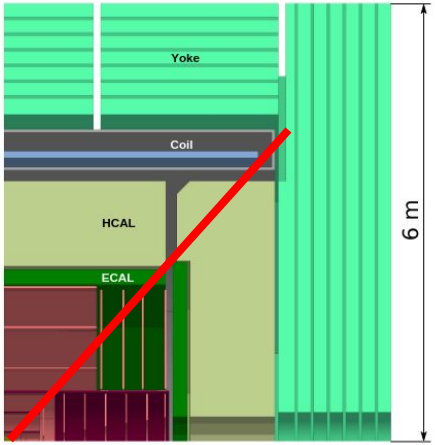
2. FastSim CLD vs. FullSim CLD

- find all plots in attached pdf of this meeting
- all plots very similar to FastSim IDEA I have shown the last time

& FullSim in different Higgs channels with more data

- instead of 3800 jets per channel, now 10000 per channel
- find plots in attached pdf
- neutral const: peaks at θ=π/4 and 3π/4
  - CLD geometry?

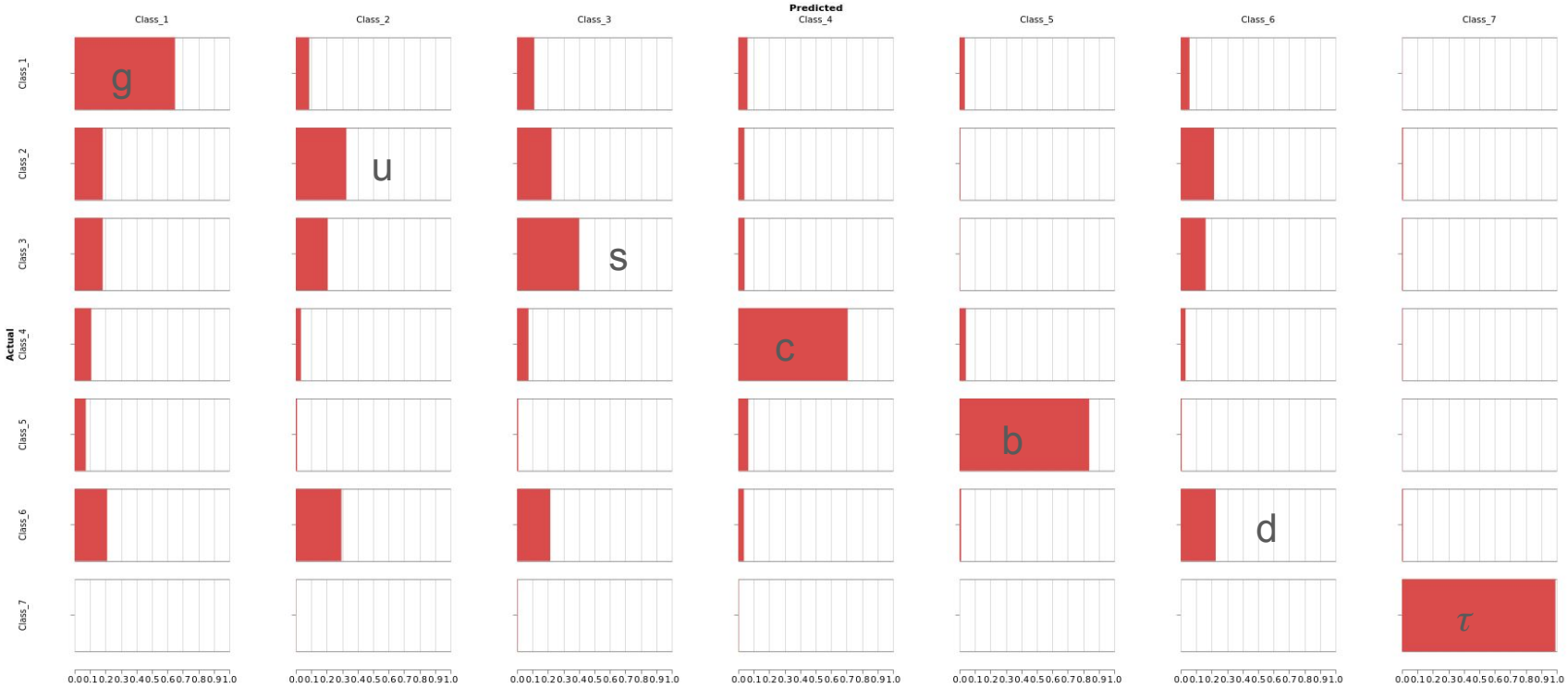




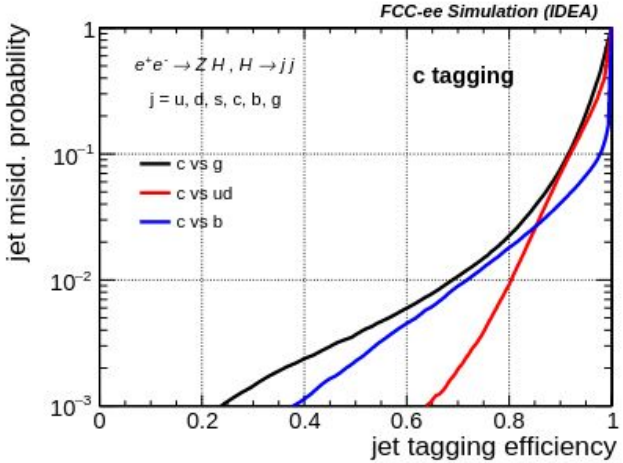
CLD - A Detector Concept for the FCC-ee

3. ML Training: Jet Tagging on FullSim CLD data

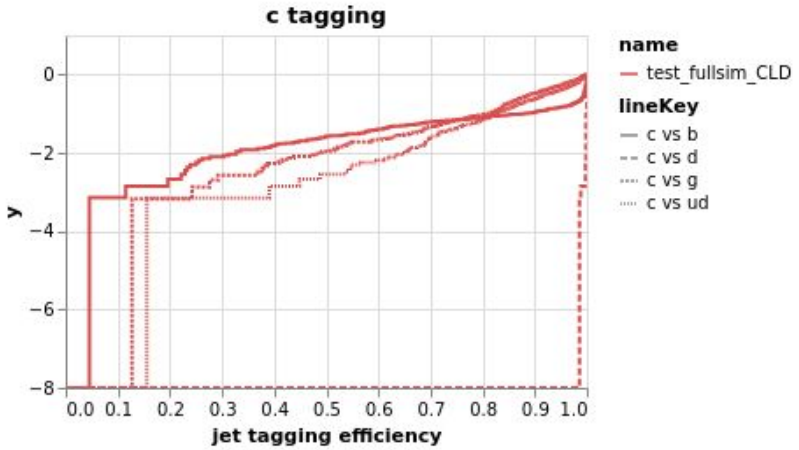
First training on FullSim CLD data - working pipeline!



3. Room for improvement



[Jet Flavour Tagging for Future Colliders with Fast Simulation]



e.g. 10e-2 mistagging rate for c vs ud:

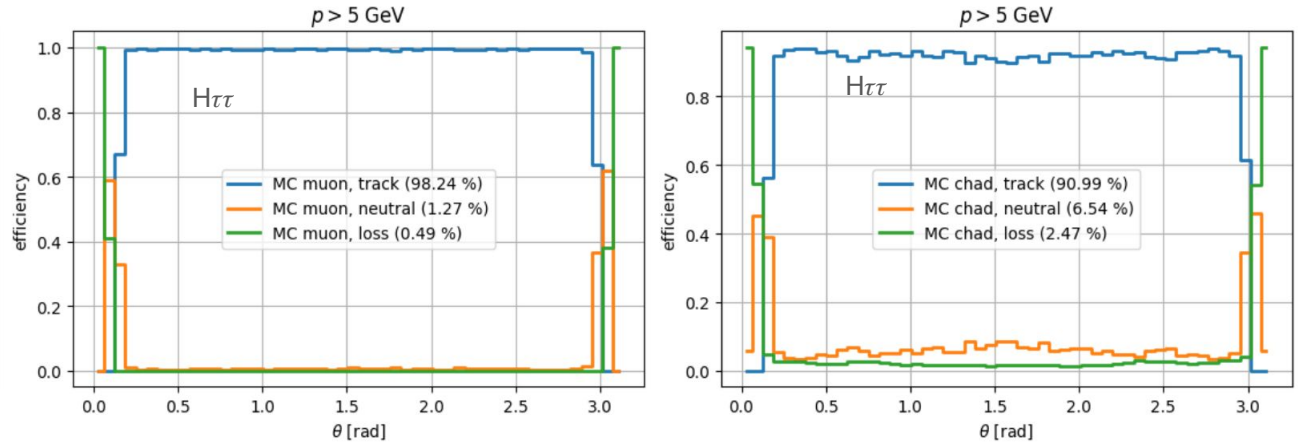
- FastSim: 80% efficiency
- FullSim: 65% efficiency

3. Reasons for suboptimal jet tagging performance on FullSim?

- possible issue: charged particles get reconstructed as neutrals due to **unassigned track to PFOs**
- issue already presented by Maria Cepeda (Tau Polarisation Discussion)
- existing issue on github from Leonhard Reichenbach on track cluster associations
- Two questions:
  - How many charged particles are not reconstructed as PFOs? (loss)
  - How many tracks are not correctly assigned to the PFO of charged particles and are therefore reconstructed as neutrals? (neutral)

4. Track efficiency vs. PFO track efficiency

- **PFO track efficiency** = what fraction of MC charged particles have a PFO with an associated track?
- Comparison of muons (don’t use PFO) vs. charged hadrons (use PFO) suggests that tracks might be reconstructed but not assigned to a PFO -> **needs to be checked!**



4. Lost charged particles & unassigned tracks to PFOs

Biggest issue: overall low PFO track efficiency:

- Hbb: 74.22 %
- Hgg: 77.34 %
- H$\tau\tau$: 91.46 %

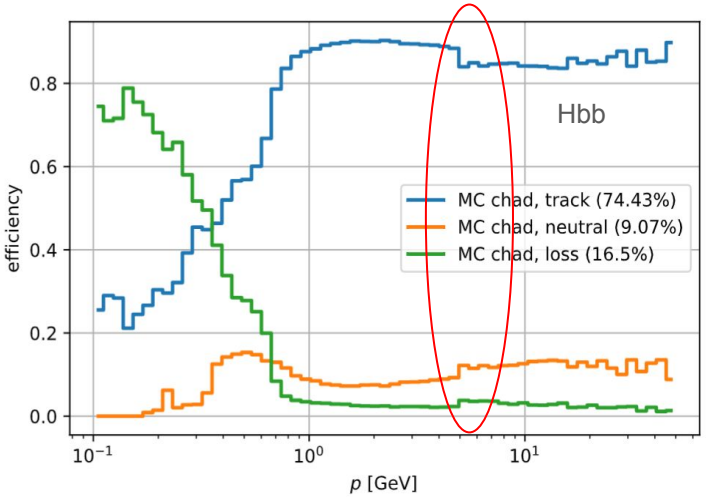
$\longrightarrow$ other MC particles either lost (not recovered as PFOs) or reconstructed as PFOs without tracks (neutrals)

$\longrightarrow$ Hbb/Hgg/H$\tau\tau$ charged particles which are reconstructed as neutrals (= unassigned track): 9.13% / 6.60% / 6.16%

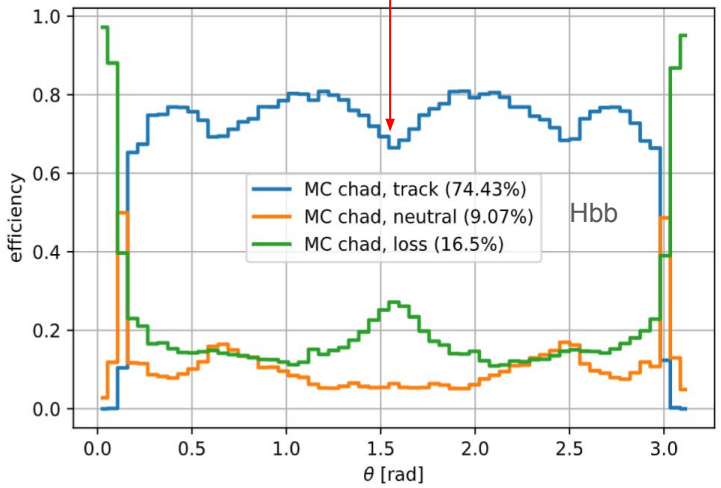
4. Lost charged particles & unassigned tracks to PFOs

Other issues:

What happens at 5 GeV?

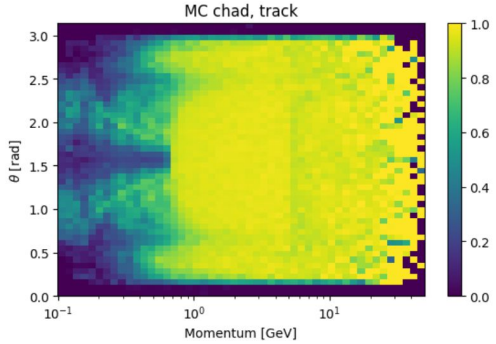


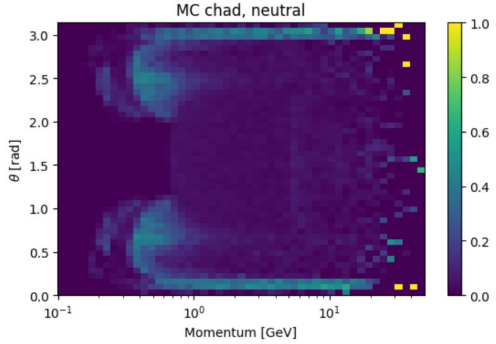
Why are spiraling charged particles lost?

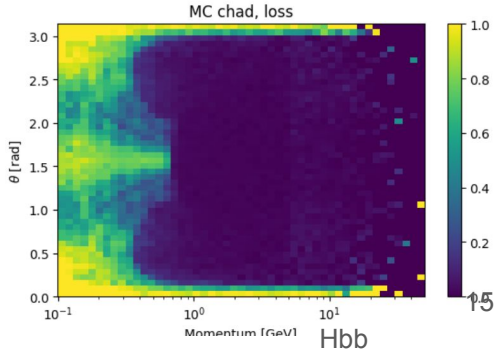


4. Lost charged particles & unassigned tracks to PFOs

- lost PFO tracks at low momenta over all angles (below 0.7 Gev)
- regions of end-caps are visible
- many lost PFO tracks at low momenta perpendicular to the beam axis -> **spiraling particles**?
- neutrals in between loss and track

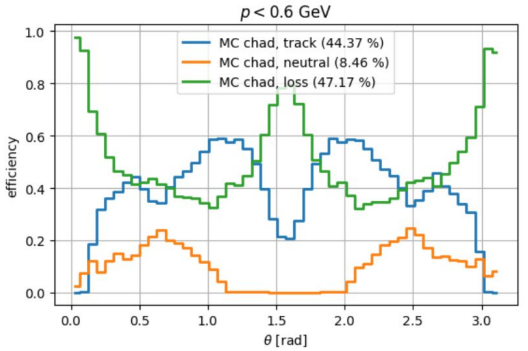


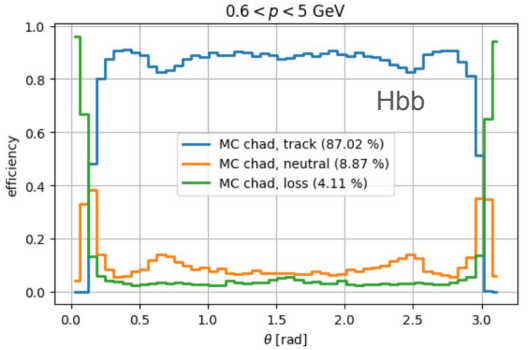


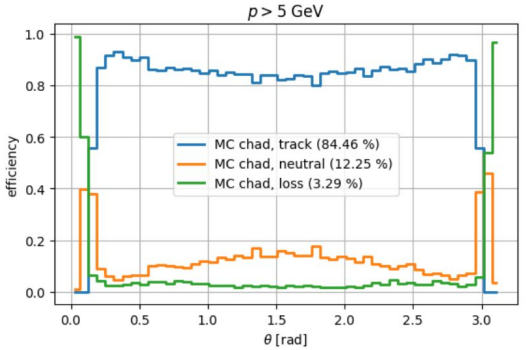


4. Lost charged particles & unassigned tracks to PFOs

- even at high momentum (p>5 GeV) the PFO track efficiency is only 84.46% for Hbb and 12.25% of the tracks are lost
- at low momentum high loss perpendicular to the beam axis







4. Outlook: what to do next

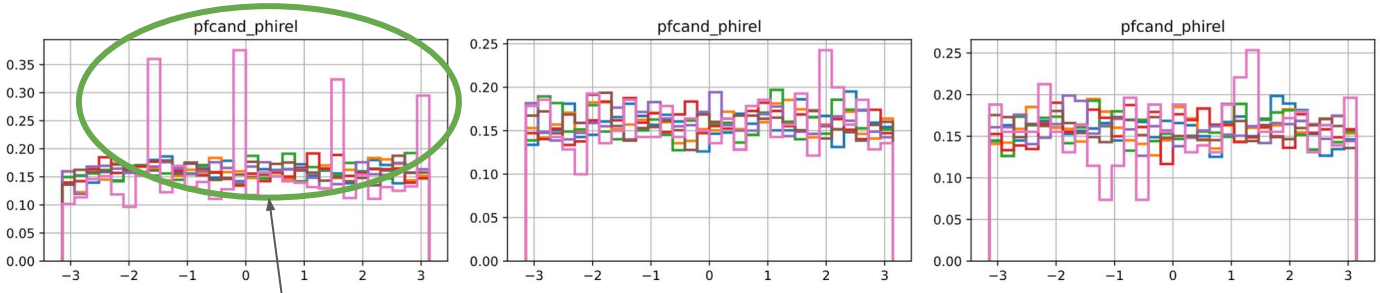
1. Further investigations on the track problem:
   a. PFO track efficiency vs. track efficiency
   b. How much energy are we missing due to lost tracks?
   c. asking the other way round: how high is the fake rate? How much fake energy do we reconstruct?
2. Verify if neutral PFOs from MC particles decrease the tagging performance by manuel assignment of tracks (artificial adjustment)
   a. If track is not assigned to reco particle (MC charged particle), do it by hand
   b. then re-run the training for tagging -> are results better?

Summary

- **Finished comparison** between Full and FastSim observables (input to tagger). See report attached. 
- First **training results** on FullSim CLD data 
- Investigation of **lost charged particles & unassigned tracks to PFOs** as possible source of suboptimal tagging performance
- **Outlook**:
  - check track efficiency vs PFO track efficiency
  - rerun tagger training with artificial correction of unassigned tracks to verify source of current suboptimal tagging performance

Backup slides

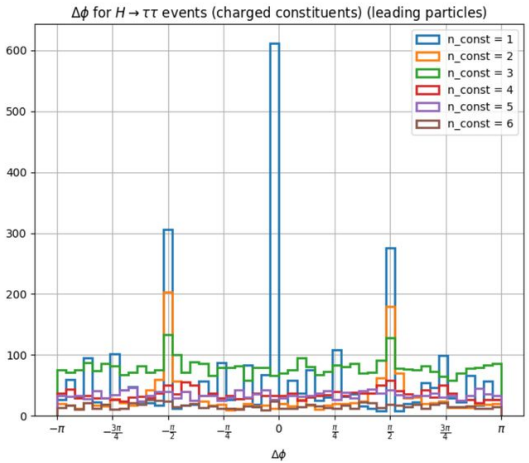
1. Δφ discrepancies in H → ττ

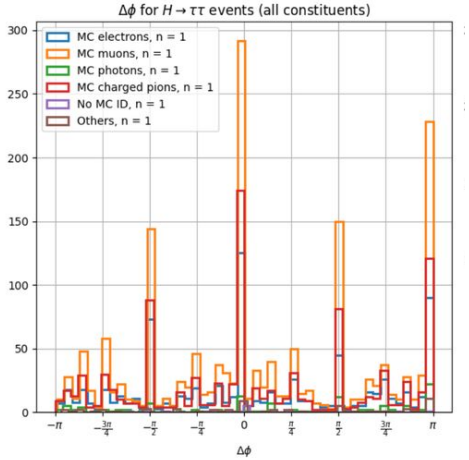


Δφ for charged constituents. Pink is τ

Where do these peaks come from?

Explanation for peak at 0





They come from jets with one constituents with MC PID muon or charged pion.

$$\tau^- \rightarrow \pi^- + \bar{\nu}_\tau$$

$$\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau$$

$$\tau^- \rightarrow \mu^- + \bar{\nu}_\mu + \nu_\tau$$

Relative angle not defined in case of one constituent! -> 0