Lepton ID in jets

Leonhard Reichenbach

CERN / University of Bonn

ECFA Higgs Factories: 2nd Topical Meeting on Reconstruction
July 11, 2023

SPONSORED BY THE

Federal Ministry of Education and Research

What will I talk about

Lepton ID (in jets):

- Here: separation of electrons, muons and charged hadrons
- Jets: b-jets from $H \to b\bar{b}$ (from ZH @ 250 GeV)
- Everything in ilcsoft/key4hep based full sim ILD DST files
- Work done with the ILD model but applicable to lots of other proposed concepts (with imaging calorimeters)

What will I talk about

Lepton ID (in jets):

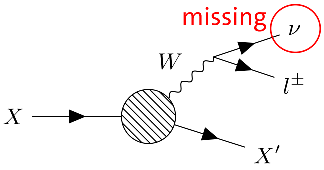
- Here: separation of electrons, muons and charged hadrons
- Jets: *b*-jets from $H \to b\bar{b}$ (from ZH @ 250 GeV)
- Everything in ilcsoft/key4hep based full sim ILD DST files
- Work done with the ILD model but applicable to lots of other proposed concepts (with imaging calorimeters)

Agenda:

- Why did I look at lepton ID?
- Have we already done this?
- How can we identify leptons?
- And how well can we do it?
- What else can we use lepton ID for?

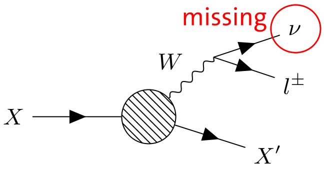
Why do lepton ID (in jets)?

- Normally we are interested in isolated leptons with high momenta e.g. from $Z \to ee/\mu\mu$ or $H \to \mu\mu$
- But $H \to b\bar{b}$ happens much more often ($\sim 60\%$ of decays)
- In $\sim 2/3$ of those decays at least one semi-leptonic B/D decay causing missing neutrino energy



Why do lepton ID (in jets)?

- Normally we are interested in isolated leptons with high momenta e.g. from $Z \to ee/\mu\mu$ or $H \to \mu\mu$
- But $H \to b\bar{b}$ happens much more often ($\sim 60\%$ of decays)
- In $\sim 2/3$ of those decays at least one semi-leptonic B/D decay causing missing neutrino energy



- The missing neutrino energy can be corrected if the e/μ from the decay is found
- See: talk by Y. Radkhorrami at last years ECFA WS and [arXiv: *2111.14775*]
- Also many other possible applications for lepton ID (later)

Have we already done this?

Yes:

- We already classify charged particles into electrons, muons and hadrons in Pandora PFA
- There are also other algorithms in ILD reconstruction like the `LikelihoodPID` processor
- Some crude lepton ID is already used in LCFIPlus as a flavor tag input
- And LICH which is very similar to what I will show but it requires full hit information

Have we already done this?

Yes:

- We already classify charged particles into electrons, muons and hadrons in Pandora PFA
- There are also other algorithms in ILD reconstruction like the `LikelihoodPID` processor
- Some crude lepton ID is already used in LCFIPlus as a flavor tag input
- And LICH which is very similar to what I will show but it requires full hit information

What is different/new?

- Offer a more modern but realistic lepton ID usable on DST files i.e. only re-use existing information
- Train on leptons from ”physical” events instead of single particles
- Provide something that is easy to change and extend

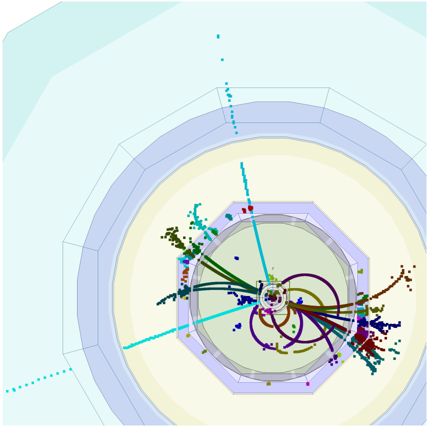
How to identify particles

with imaging calorimeters

Can be really simple e.g.

- electron: (well-formed) ecal cluster + track
- muon: track + line shaped ecal and hcal clusters (+ muon system hits)
- charged hadron: track + (less-regular) ecal + hcal cluster
- photon: like electron but without a track
- neutral hadron: like a charged hadron without a track

But sometimes less obvious, especially in crowded environments (jets)



$ZH \to \mu^+\mu^- b\bar{b}$ event at ILD

Pandora PFA

What does it do?

- Takes in all tracks, calorimeter hits and the detector geometry
- Forms calorimeter clusters and matches them to tracks
- Does this iteratively for different particle type hypotheses
- Returns particle flow objects (PFOs) i.e. clusters (+ tracks) classified into one of the 5 particle categories ($e, \mu, \pi^{\pm}, \gamma, n$)

Pandora PFA

What does it do?

- Takes in all tracks, calorimeter hits and the detector geometry
- Forms calorimeter clusters and matches them to tracks
- Does this iteratively for different particle type hypotheses
- Returns particle flow objects (PFOs) i.e. clusters (+ tracks) classified into one of the 5 particle categories ($e, \mu, \pi^{\pm}, \gamma, n$)

Example: dedicated muon ID (simplified)

- Extrapolate track helix through the calorimeters and (flipped) through the solenoid
- Try to match to a sufficiently muon-like yoke-hit cluster
- ‘Swim’ back through the calorimeters collecting minimum ionizing hits along the way
- Call this a muon and remove the track and the hits from the pool

Pandora PFA

What does it do?

- Takes in all tracks, calorimeter hits and the detector geometry
- Forms calorimeter clusters and matches them to tracks
- Does this iteratively for different particle type hypotheses
- Returns particle flow objects (PFOs) i.e. clusters (+ tracks) classified into one of the 5 particle categories $(e, \mu, \pi^{\pm}, \gamma, n)$

Example: dedicated muon ID (simplified)

- Extrapolate track helix through the calorimeters and (flipped) through the solenoid
- Try to match to a sufficiently muon-like yoke-hit cluster
- ‘Swim’ back through the calorimeters collecting minimum ionizing hits along the way
- Call this a muon and remove the track and the hits from the pool

This influences all other attempts to do particle ID further down the chain!

ILD DST PID content

Multiple sources from processors run during reconstruction

- `MarlinReco/AddClusterProperties`: uses `MarlinUtil/WeightedPoints3D` to treat the cluster hits as an energy-weighted 3D point cloud
- Calculates position, direction, squared and quartic energy sums and 4-momentum covariance matrix
- Getters for multiple other parameters of this point cloud
- A lot of them can be recovered from the cov-matrix etc.

ILD DST PID content

Multiple sources from processors run during reconstruction

- `MarlinReco/AddClusterProperties`: uses `MarlinUtil/WeightedPoints3D` to treat the cluster hits as an energy-weighted 3D point cloud
- Calculates position, direction, squared and quartic energy sums and 4-momentum covariance matrix
- Getters for multiple other parameters of this point cloud
- A lot of them can be recovered from the cov-matrix etc.

- `MarlinReco/ComputeShowerShapes`: uses `MarlinUtil/ClusterShapes`
- Fits a 3D shower profile to the cluster and calculates other classic shower shape variables e.g. expected EM-shower length
- Partially used as input for the old `LikelihoodPID` processor

ILD DST PID content

Multiple sources from processors run during reconstruction

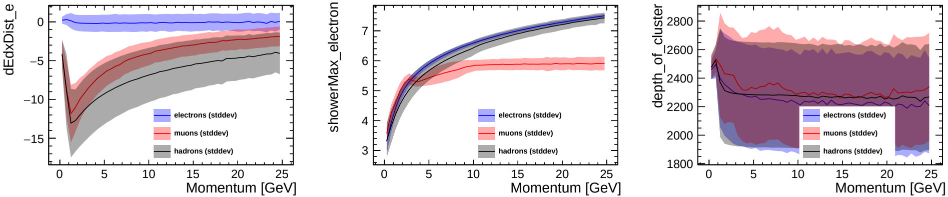
- `MarlinReco/AddClusterProperties`: uses `MarlinUtil/WeightedPoints3D` to treat the cluster hits as an energy-weighted 3D point cloud
- Calculates position, direction, squared and quartic energy sums and 4-momentum covariance matrix
- Getters for multiple other parameters of this point cloud
- A lot of them can be recovered from the cov-matrix etc.

- `MarlinReco/ComputeShowerShapes`: uses `MarlinUtil/ClusterShapes`
- Fits a 3D shower profile to the cluster and calculates other classic shower shape variables e.g. expected EM-shower length
- Partially used as input for the old `LikelihoodPID` processor
- `MarlinReco/Compute_dEdx`: calculates dE/dx distances for different particle hypotheses and stores them with the PFOs

LeptonID processor

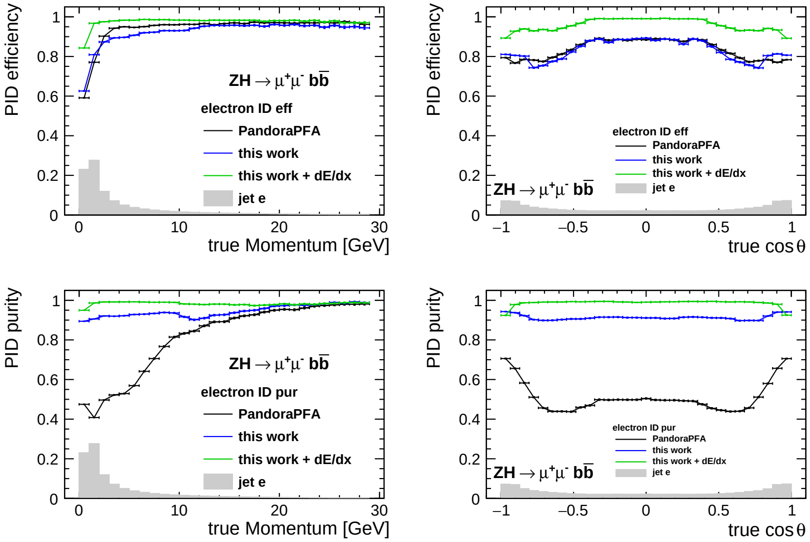
- Build a BDT based classifier using pre-computed calorimeter shapes as inputs
- Classify into three classes: electrons, muons and charged hadrons
- Train on b-jets from $H \to b\bar{b}$ (from ZH @ 250 GeV)
- Goal: refine Pandora PFA 'PID' output with an easy to adapt system

- PID efficiency $= \frac{\text{correctly identified}}{\text{identifiable}}$
- PID purity $= \frac{\text{correctly identified}}{\text{identified as}}$
- Identifiable:
  - PFO has one track and one cluster
  - track/cluster were caused by the same MC particle to at least 50%
  - 50% of all hits from the MC particle are assigned to the track/cluster



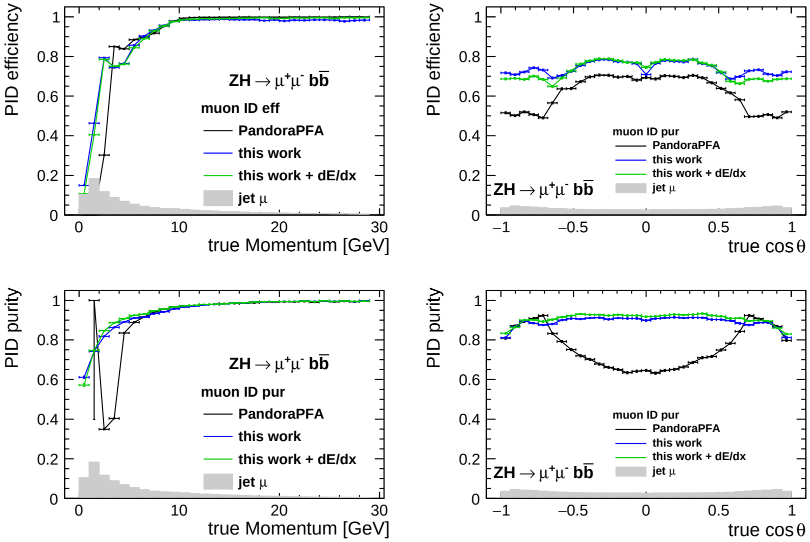
Results: electrons

- Efficiency is similar to Pandora PFA
- Purity at low momenta is much better
- Highlight: dE/dx significantly improves performance for electrons at low momenta



Results: muons

- Overall better efficiency, especially in forward region
- Better purity in barrel region
- Pandora PFA muon tag too dirty at low momenta



Results: summary

- Overall performance for $H \to b\bar{b}$ jet leptons:
  - e: 95% efficiency and 98% purity with dE/dx
  - e: 82% efficiency and 92% purity without dE/dx
  - μ: 74% efficiency and 89% purity
- Having dE/dx information available improves electron identification significantly
- Using full reconstruction data instead of just DST will offer further improvement
- Promising results motivating to build a more production ready solution out of this. (see Uli's talk tomorrow morning)

9:55 AM **Proposal for a Particle ID Framework** 25m

Speaker: Ulrich Einhaus

Results: summary

- Overall performance for $H \rightarrow b\bar{b}$ jet leptons:
  - e: 95% efficiency and 98% purity with dE/dx
  - e: 82% efficiency and 92% purity without dE/dx
  - μ: 74% efficiency and 89% purity
- Having dE/dx information available improves electron identification significantly
- Using full reconstruction data instead of just DST will offer further improvement
- Promising results motivating to build a more production ready solution out of this. (see Uli's talk tomorrow morning)

9:55 AM **Proposal for a Particle ID Framework** 25m

Speaker: Ulrich Einhaus

The `LeptonID` processor is now available in MarlinReco!

Improvements and applications

Possible improvements

- Train multiple classifiers for different θ and momentum bins
- Use full hit information to investigate other shape variables
- Revisit Pandora PFA cluster assignment for muons

Improvements and applications

Possible improvements

- Train multiple classifiers for different θ and momentum bins
- Use full hit information to investigate other shape variables
- Revisit Pandora PFA cluster assignment for muons

Future applications?

- Identification of leptons from semi-leptonic decays of heavy flavor
- Pandora PFA optimisation, e.g. try out what shapes work and propagate that knowledge back, add ML-based algorithms
- Flavor tagging
- Refitting of tracks with adapted hypothesis, e.g. Gaussian sum filter (GSF) for electrons

Summary

- Lepton ID in jets using imaging calorimeters is feasible and already performs well
- dE/dx (and possibly dN/dx?) capabilities enable perfect ID of electrons at low momenta
- We could still exploit the capabilities of imaging calorimeters further, e.g. by improving Pandora PFA

This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA).

SPONSORED BY THE

Federal Ministry
of Education
and Research