The Optimal Use of Segmentation for Sampling Calorimeters

Ryan Milton

University of California, Riverside

CPAD Workshop 2023

November 7, 2023

arXiv:2310.04442

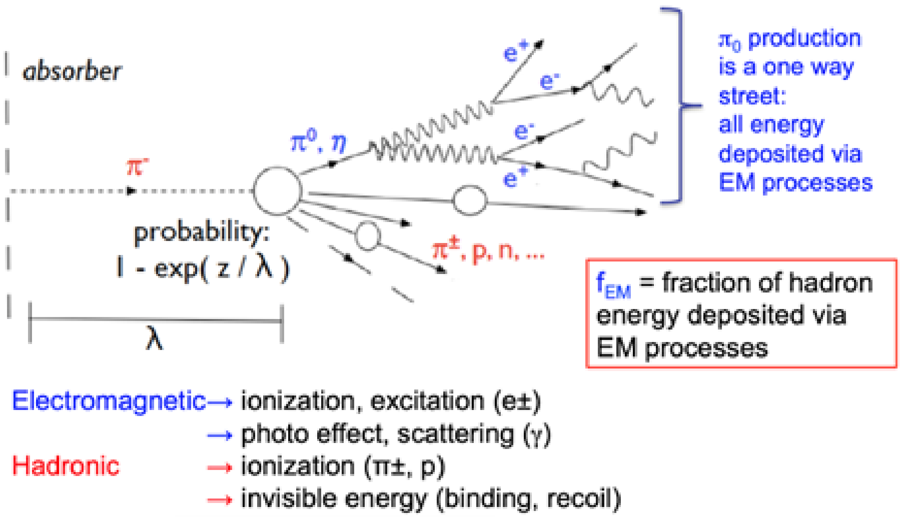
DOE grant award number DE-SC0022355

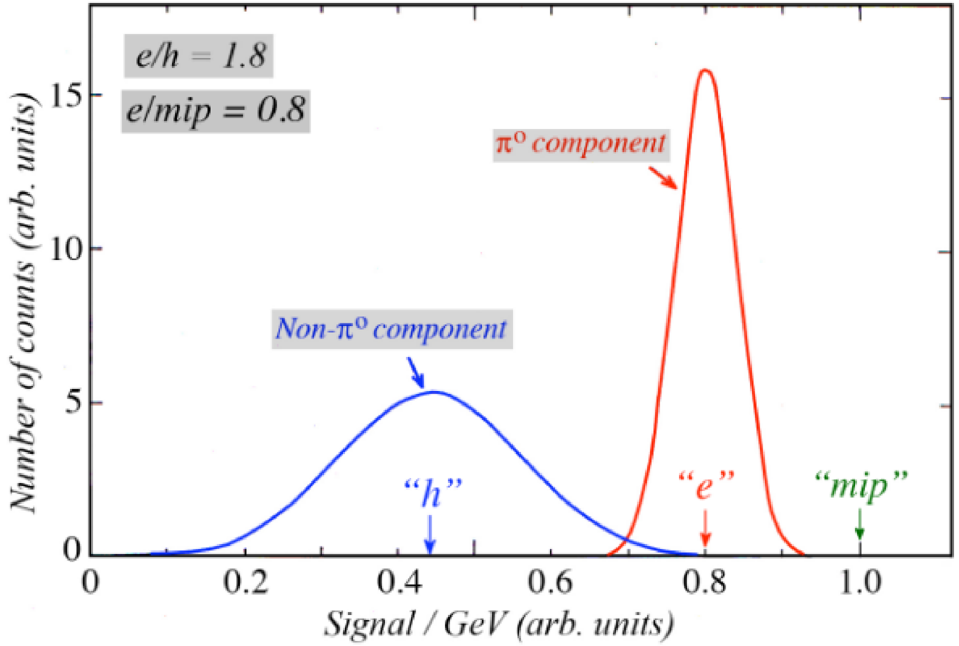
Overview

- Non-compensation in hadronic calorimeters
- Methods to address non-compensation
  - Software compensation algorithms
  - ML based reconstruction
- Using ML based reconstruction to inform design of ePIC forward HCal
  - DeepSets & GNN models
  - Longitudinal segmentation
  - Varying input features for ML model

Non-compensation in hadronic calorimeters

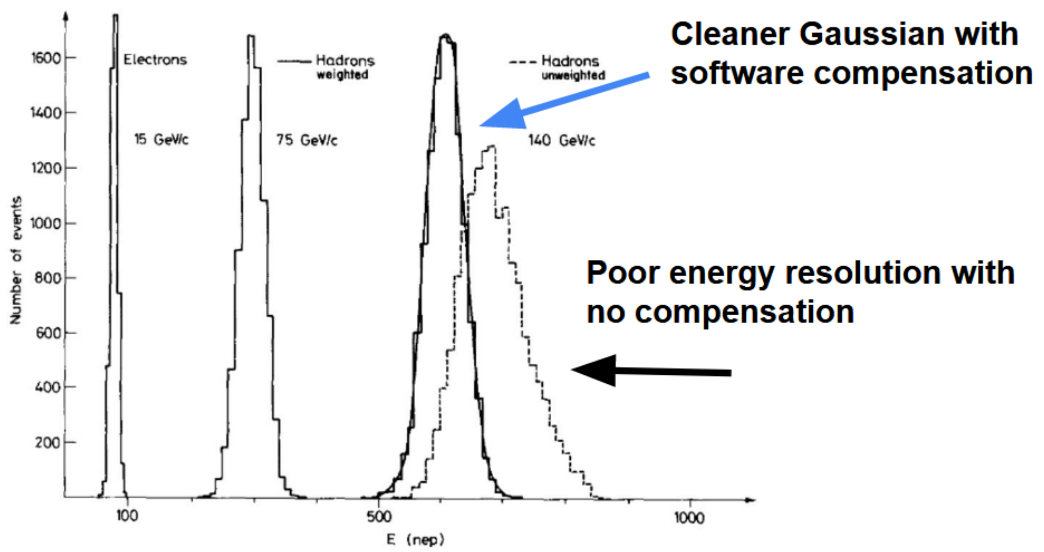
- Hadronic showers have EM and hadronic components
- EM component usually has larger response in calorimeter
  - e/h ≠ 1
  - Leads to deterioration of energy resolution
- One possible solution: Using specific absorbers and scintillators
  - Imposes strict requirements on material





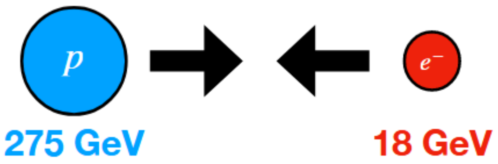
Methods to address non-compensation

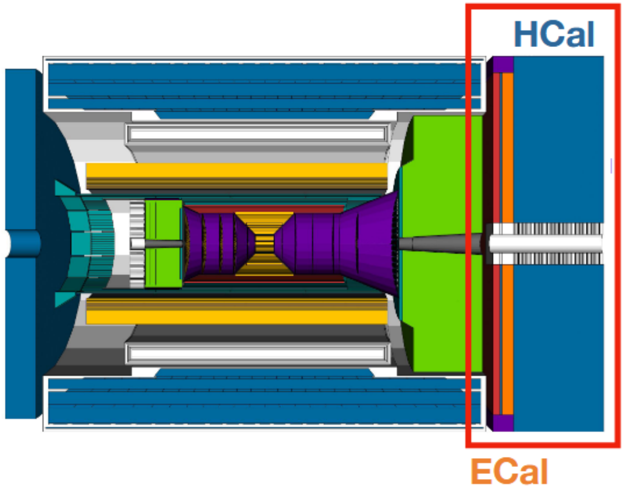
- Assigning weights to EM and hadronic energies in cells, event-by-event
  - Example ([CERN, 1980]): $E_{cell,weighted} = E_{cell}(1 - \frac{.03}{\sqrt{E_{total}}} \cdot E_{cell})$
  - Example ([CALICE, 2012]): $E \propto \sum_i E_{HCAL,i}\, \omega_i$, ω_i is energy density dependent weight
- AI/ML-based reconstruction
  - Seen in ATLAS
- **These methods are employed & optimized after detector construction and data-taking**



Detector & Simulation info

- ePIC forward ECal + HCal
  - Reminder: HCal 4 layers W/Sc, 60 layers Fe/Sc, 5 cm x 5 cm granularity
- Use ML based reconstruction to optimize the performance and design of HCal
- Employing these methods before detector construction
- Simulate the ECal + HCal in DD4hep and deploy ML methods





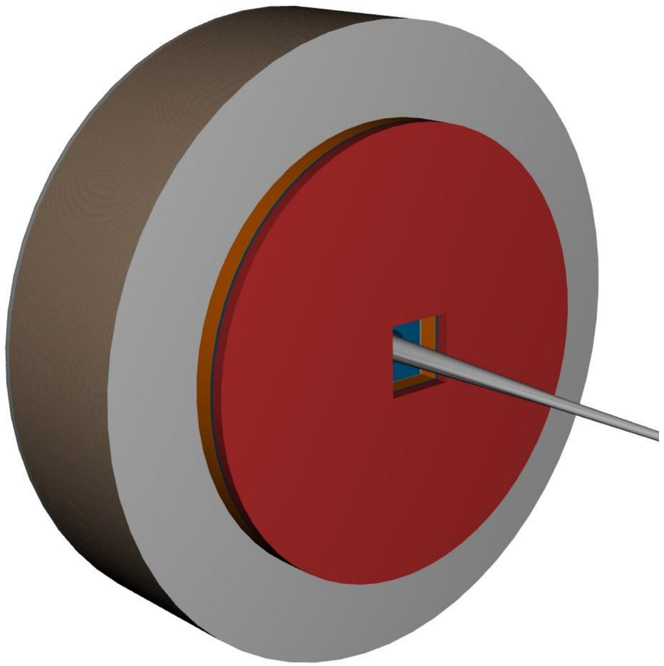
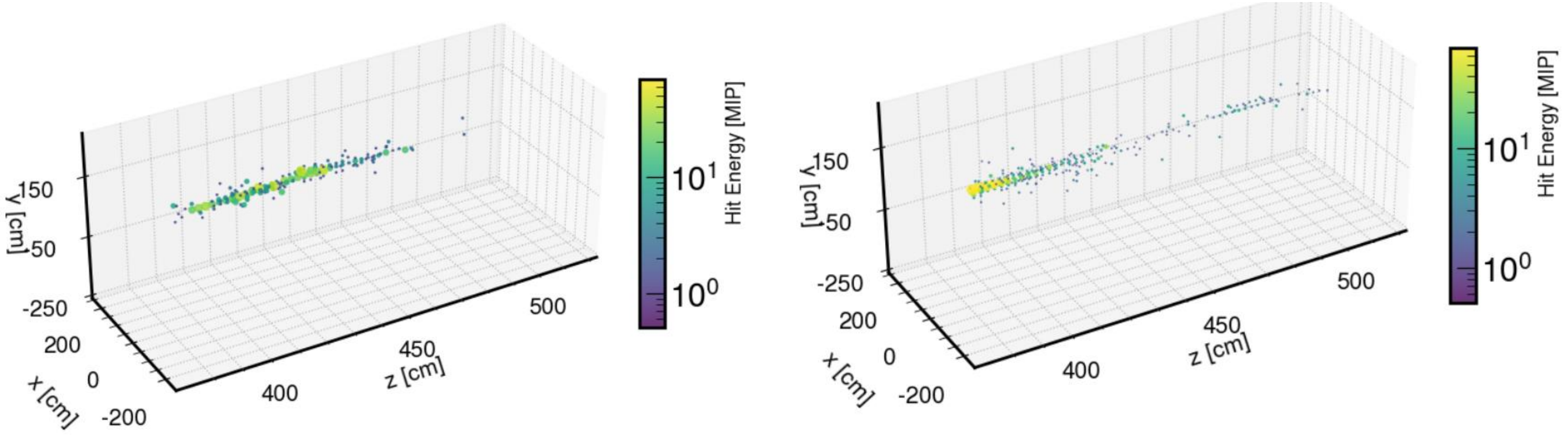


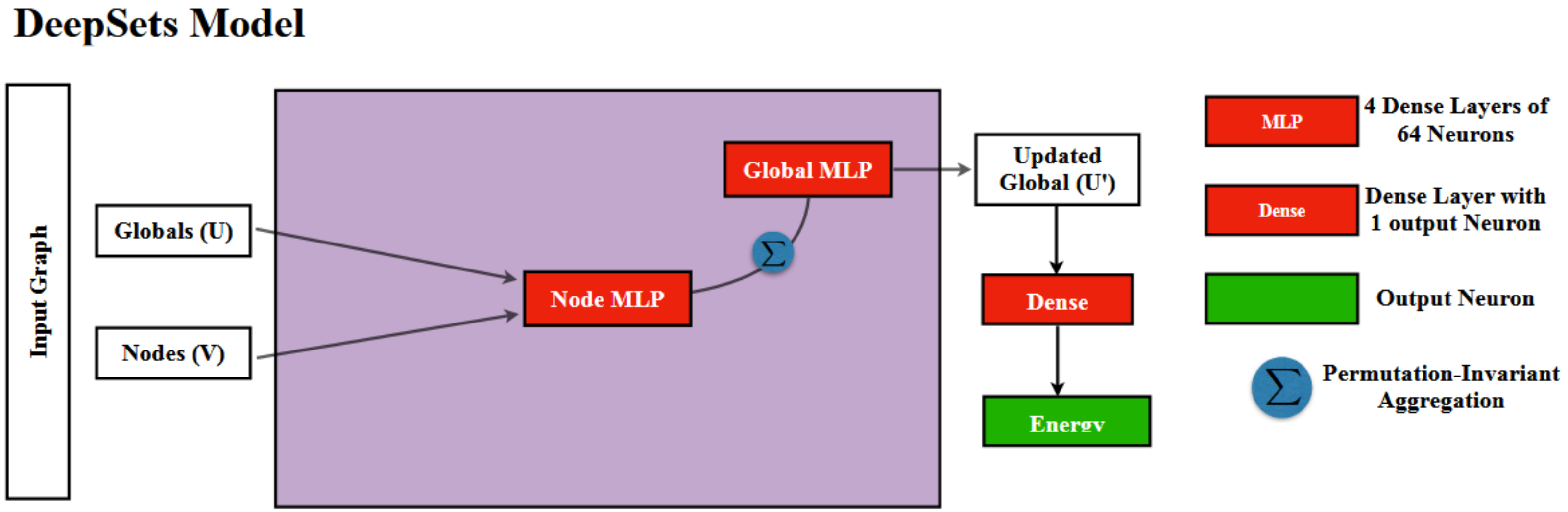
Figure courtesy of ePIC

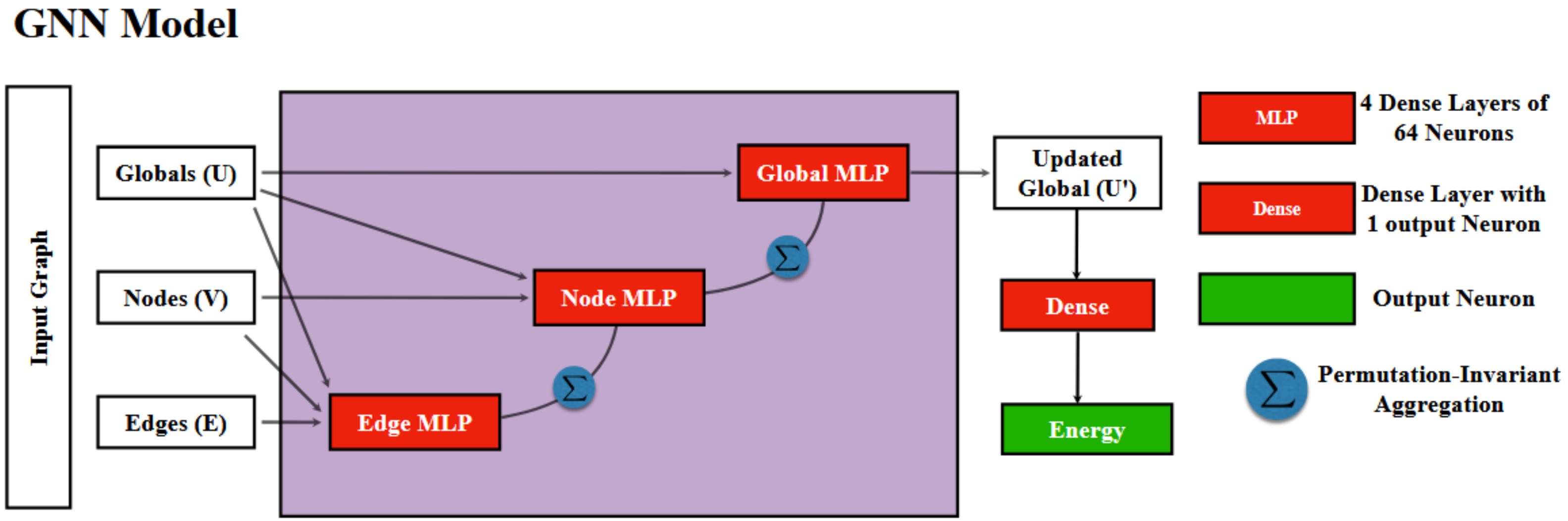
Machine learning model

- $\mathcal{O}(100 - 1000)$ cell hits per shower → point clouds
- Establish a model to predict P_{gen} given cell information
    - Use DeepSets & GNNs
    - Provide model with varying cell and segmentation information
    - Cell data → predicted momentum of incident particle



Model schematics

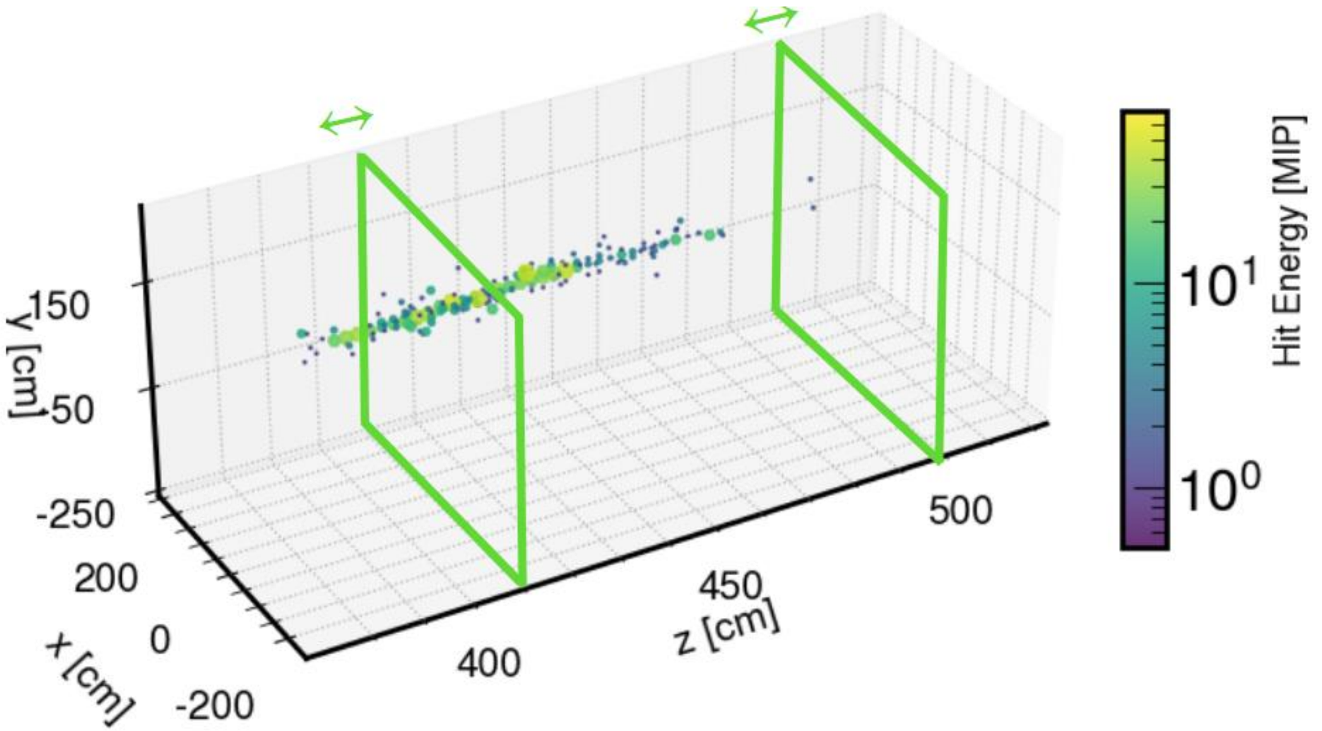




- In theory, DeepSets can learn everything a GNN can
- Encode geometric information directly in the GNN

Studying longitudinal segmentations

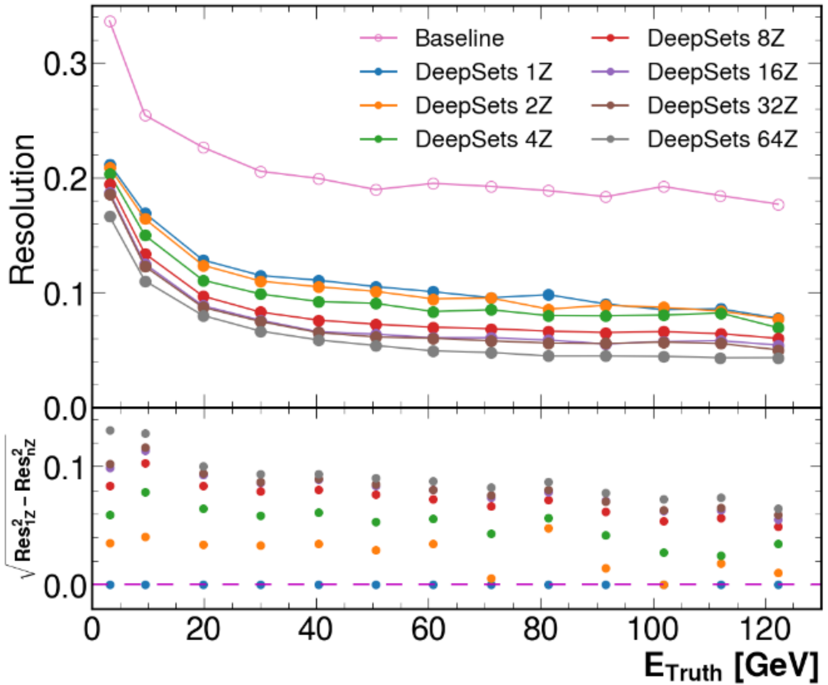
4-Layer Configuration

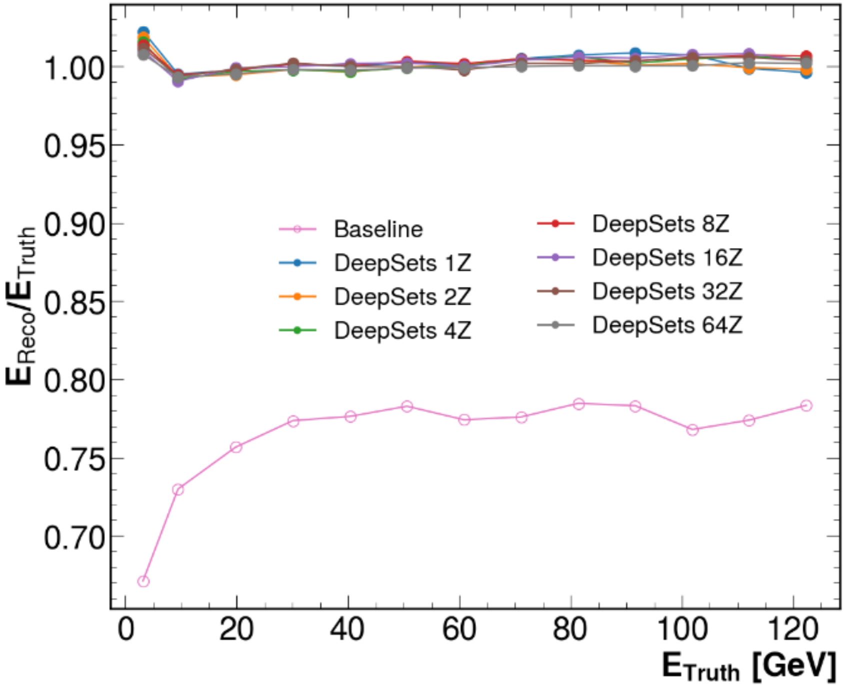


- Segment the calorimeter into 64 layers
- Combine energies of cells with same transverse position, different layers
- Run regression and identify optimal longitudinal configuration
  - Current ePIC forward HCal design has 7 segmentations

Studying longitudinal segmentations

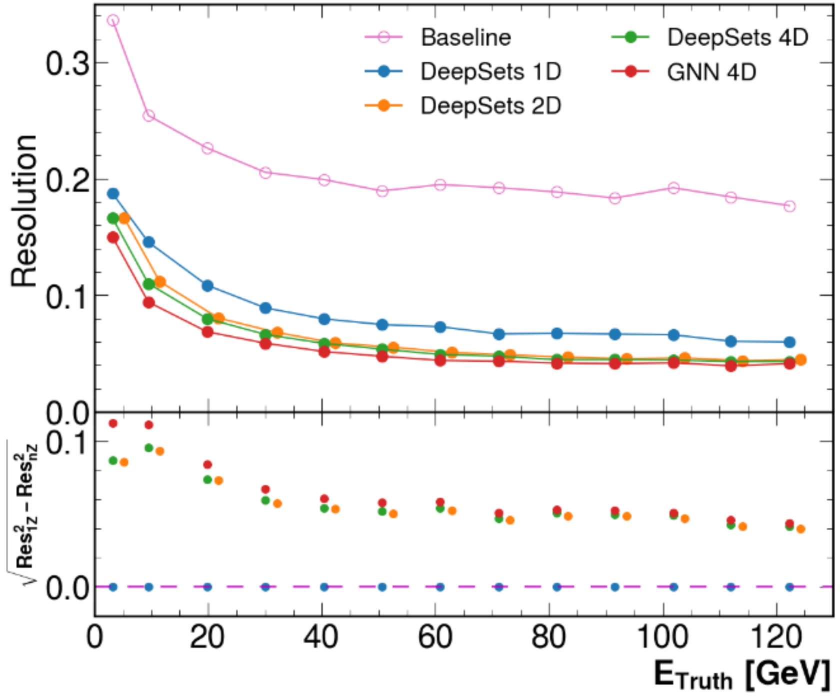
- Resolution = $\sigma(\frac{E_{pred}}{E_{true}})/\mu(\frac{E_{pred}}{E_{true}})$
- Compare with "baseline" energy reconstruction
  - $E = \sum_i \frac{E_{i,HCAL}}{SF_{HCAL}} + \sum_i \frac{E_{i,ECAL}}{SF_{ECAL}}$
- Improvement in energy resolution with more longitudinal segmentations
  - 1Z: $\frac{\sigma}{E} = \frac{0.39}{\sqrt{E}} \oplus .0834$, 64Z: $\frac{\sigma}{E} = \frac{0.30}{\sqrt{E}} \oplus .0325$

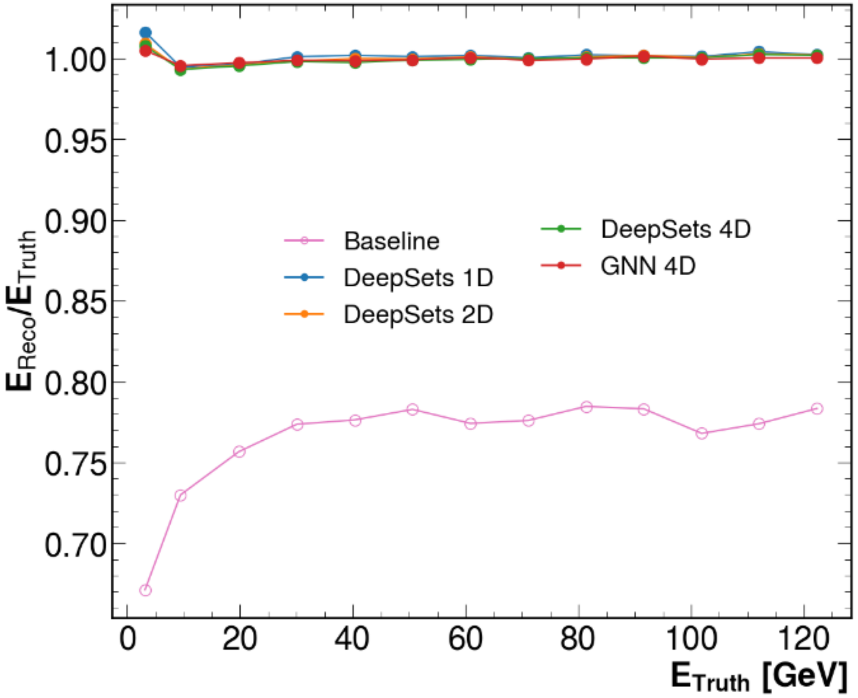




Varying input features for model

- Vary information given to model
  - Train models on cell info: E, E + Z, E + XYZ (1D, 2D, 4D)
  - GNN 4D: Cell energy + cell XYZ + energy of nearby cells
- Biggest improvement after inclusion of Z information, especially at low energies
  - 1D: $\frac{\sigma}{E} = \frac{0.36}{\sqrt{E}} \oplus .0561$, 2D: $\frac{\sigma}{E} = \frac{0.31}{\sqrt{E}} \oplus .0344$





Conclusions & outlook

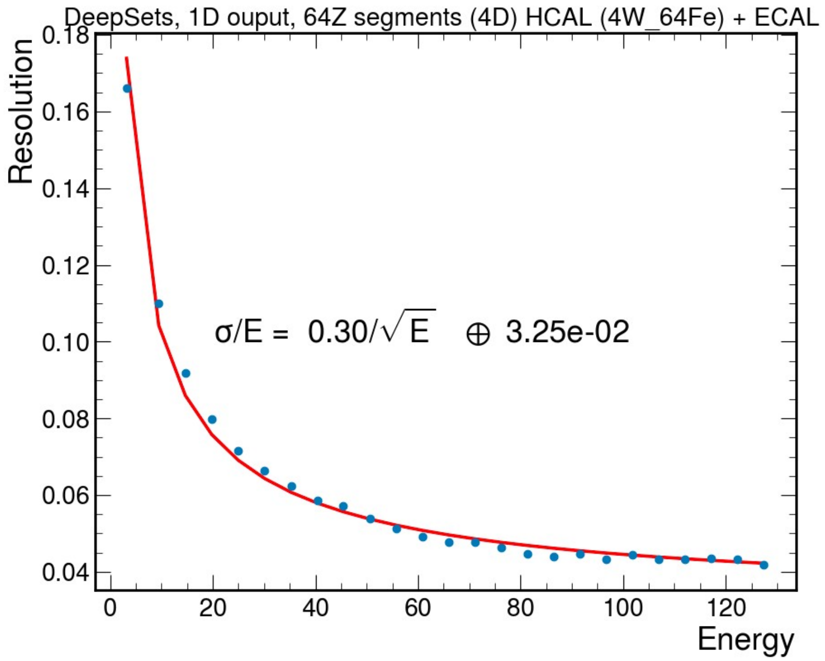
- Employ ML based reconstruction to optimize longitudinal segmentation in ePIC HCal before detector is constructed
- ML reconstruction improves energy resolution of detector compared to simple reconstruction
- Can later incorporate θ and ϕ info to investigate position resolution and HCal's transverse segmentation
- Further studies needed for more complex events including many particles & jets

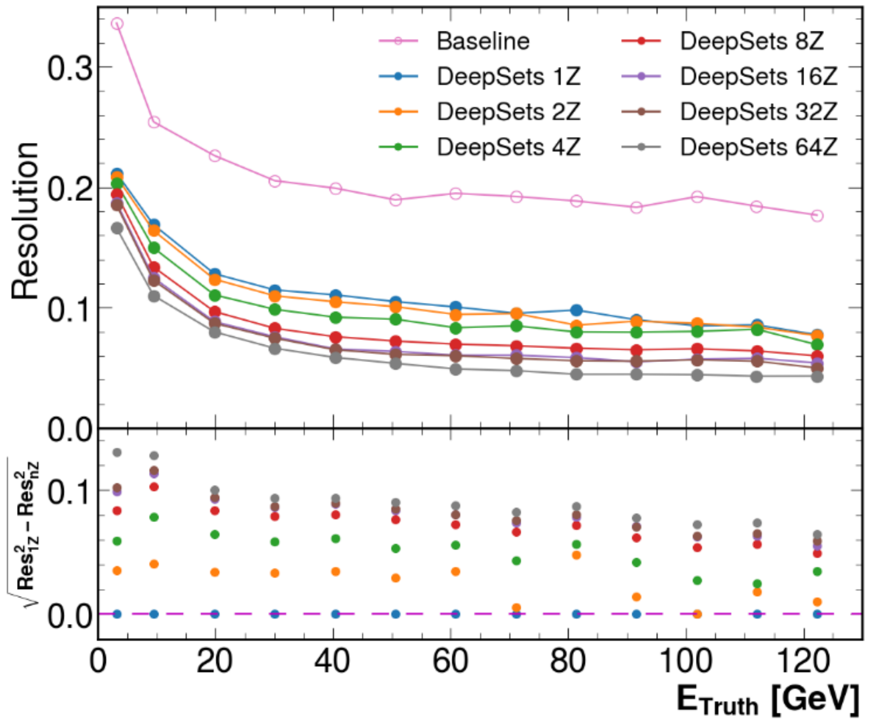
Thank you!

Backup

Longitudinal segmentation fits

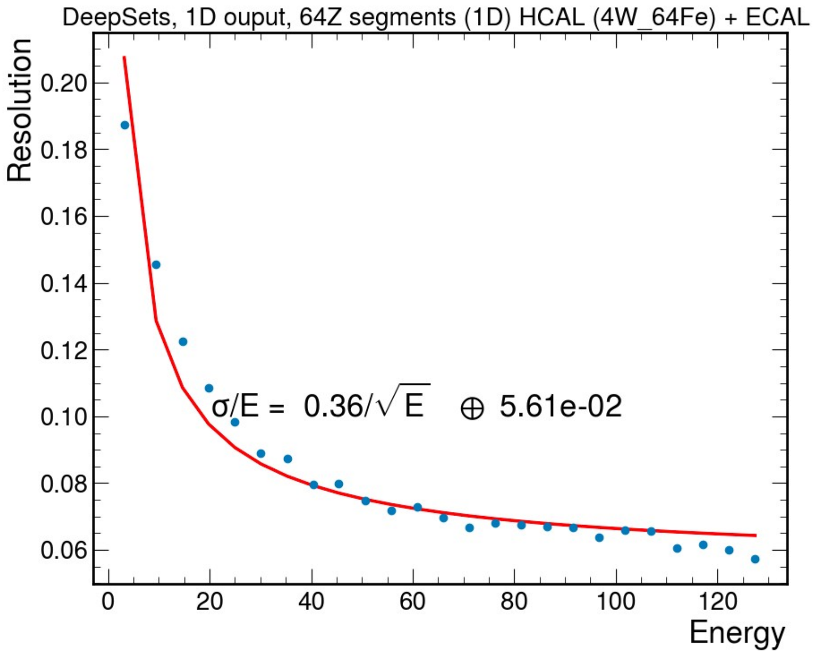
- 1Z: $\frac{\sigma}{E} = \frac{0.39}{\sqrt{E}} \oplus .0834$
- 2Z: $\frac{\sigma}{E} = \frac{0.38}{\sqrt{E}} \oplus .0800$
- 4Z: $\frac{\sigma}{E} = \frac{0.36}{\sqrt{E}} \oplus .0711$
- 8Z: $\frac{\sigma}{E} = \frac{0.34}{\sqrt{E}} \oplus .0541$
- 16Z: $\frac{\sigma}{E} = \frac{0.33}{\sqrt{E}} \oplus .0460$
- 32Z: $\frac{\sigma}{E} = \frac{0.33}{\sqrt{E}} \oplus .0431$
- 64Z: $\frac{\sigma}{E} = \frac{0.30}{\sqrt{E}} \oplus .0325$

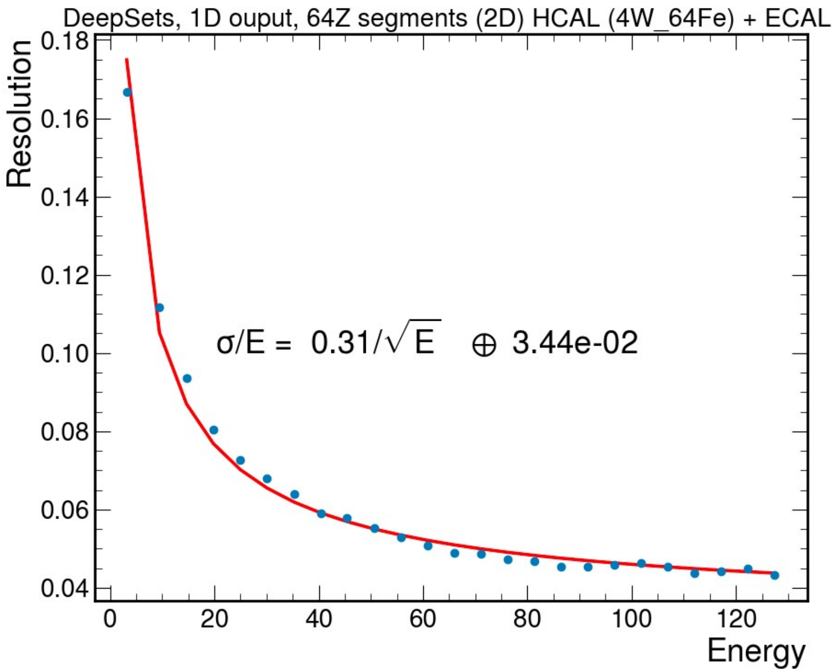


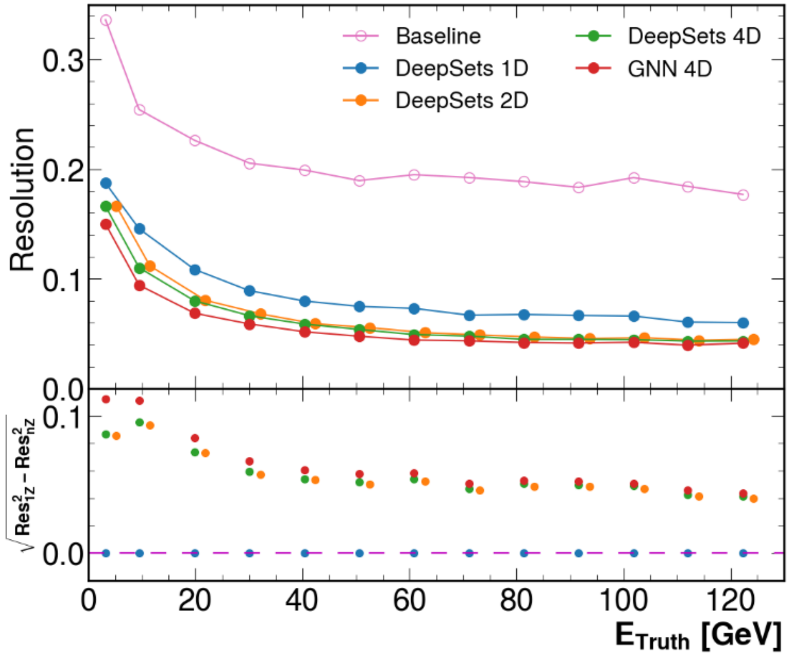


Input feature fits

- 1D: $\frac{\sigma}{E} = \frac{0.36}{\sqrt{E}} \oplus$ .0561 (E)
- 2D: $\frac{\sigma}{E} = \frac{0.31}{\sqrt{E}} \oplus$ .0344 (E + Z)
- 4D: $\frac{\sigma}{E} = \frac{0.30}{\sqrt{E}} \oplus$ .0325 (E + XYZ)
- GNN 4D: $\frac{\sigma}{E} = \frac{0.27}{\sqrt{E}} \oplus$ .0307 (E + XYZ + neighboring cell energy)

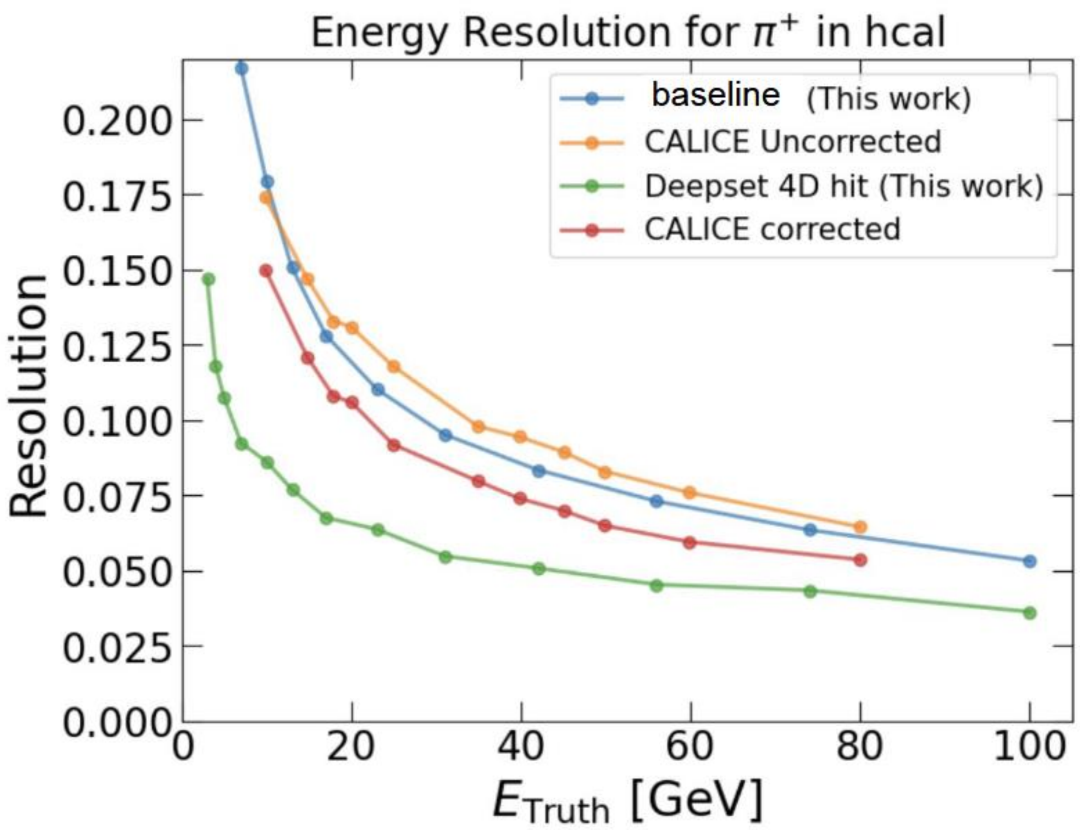






Comparing ML method to CALICE software compensation

- Simulate CALICE Fe/Sc sampling calorimeter (arXiv:1207.4210)
- Apply ML-based reconstruction
- Improvement in energy resolution compared to CALICE software compensation



Typical Gaussian Fit

- Strawman = baseline: $E = \sum_i \frac{E_{i,HCAL}}{SF_{HCAL}} + \sum_i \frac{E_{i,ECAL}}{SF_{ECAL}}$