Hadronic shower substructure reconstruction with GNN

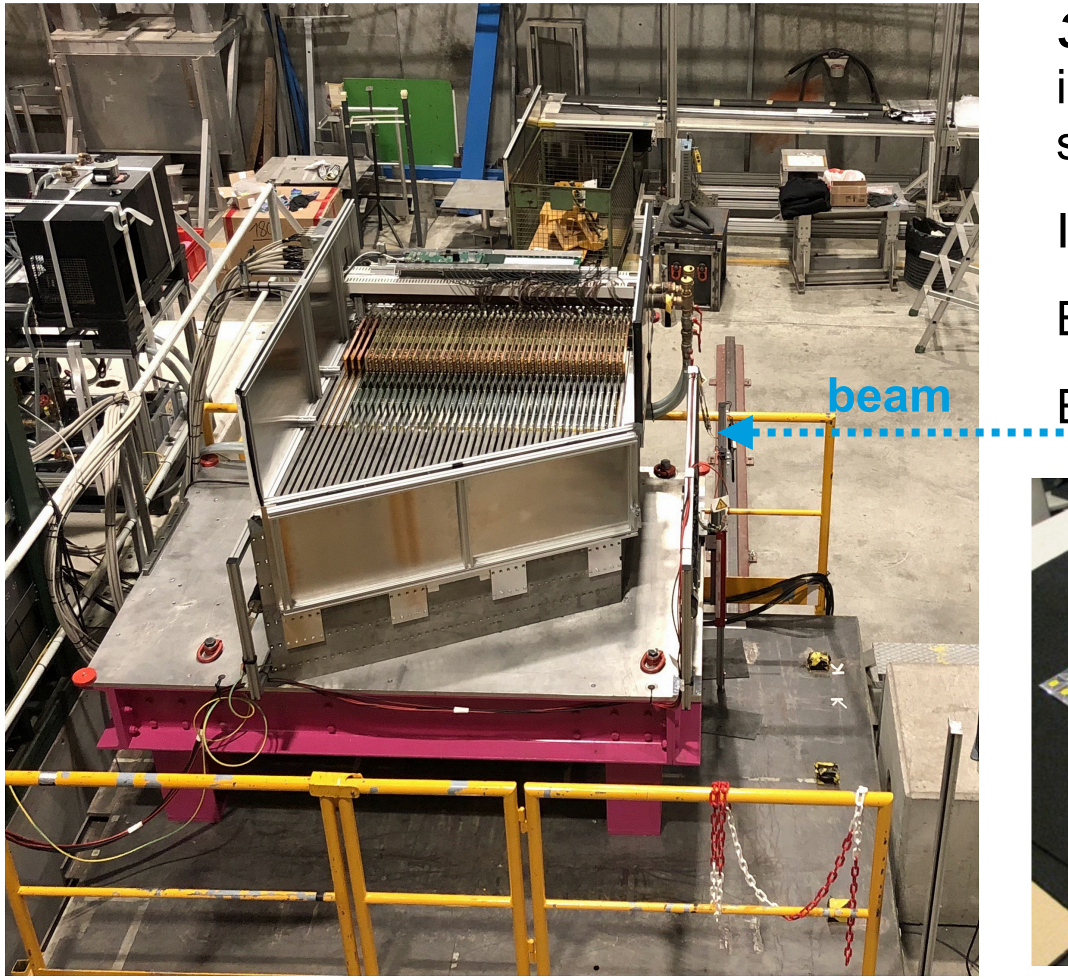
CALICE Collaboration Meeting

Vladimir Bocharnikov (DESY),
10 Sep 2021

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

CALICE AHCAL

Test beam prototype.

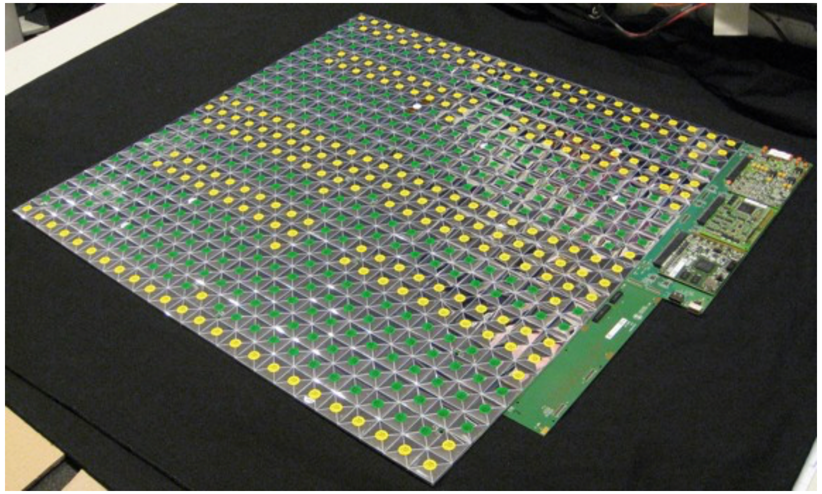


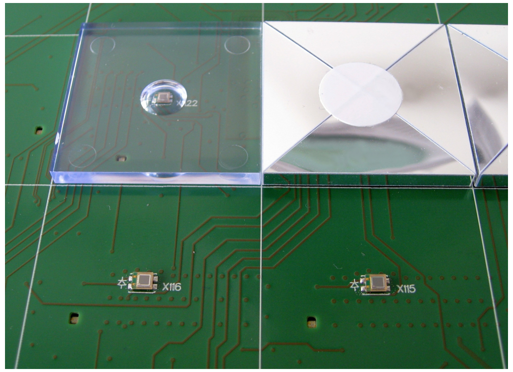
***39* active layers** of *24x24* scintillator tiles (*3x3* cm^2 each) with individual SiPM readout. Active layers alternate with ~*2 cm* steel absorber.

In total: ***~22000* channels, *~4 λ*, *~38*X_0**

Beam particles: **muons, electrons, pions**

Energy range: ***10-200 GeV***

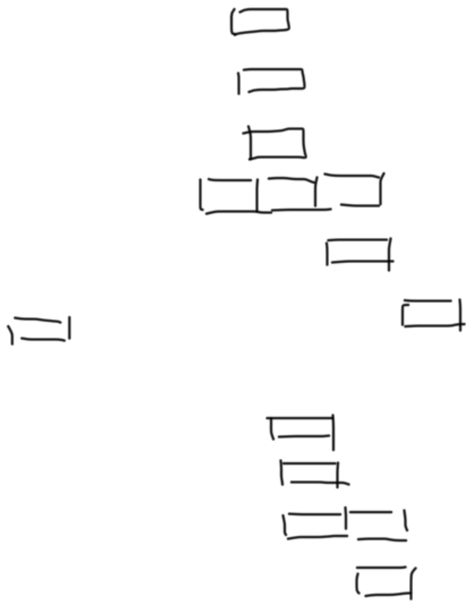




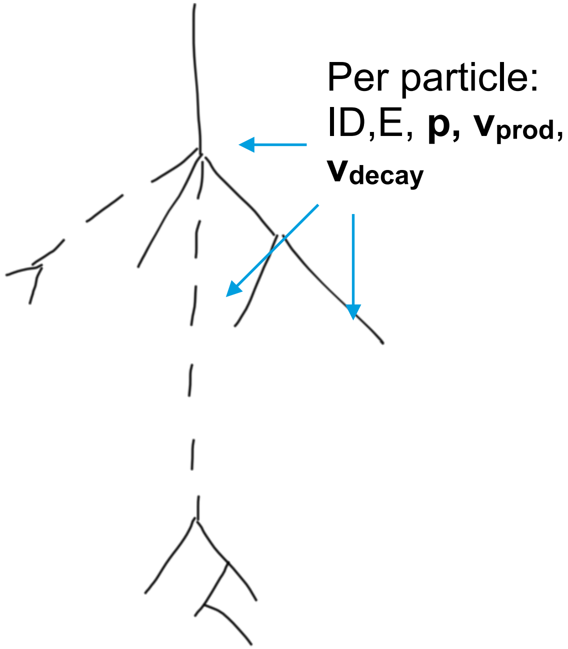
Calorimeter vision for hadronic showers

Ultimate goal and general approach

Set of hits in highly granular calorimeter

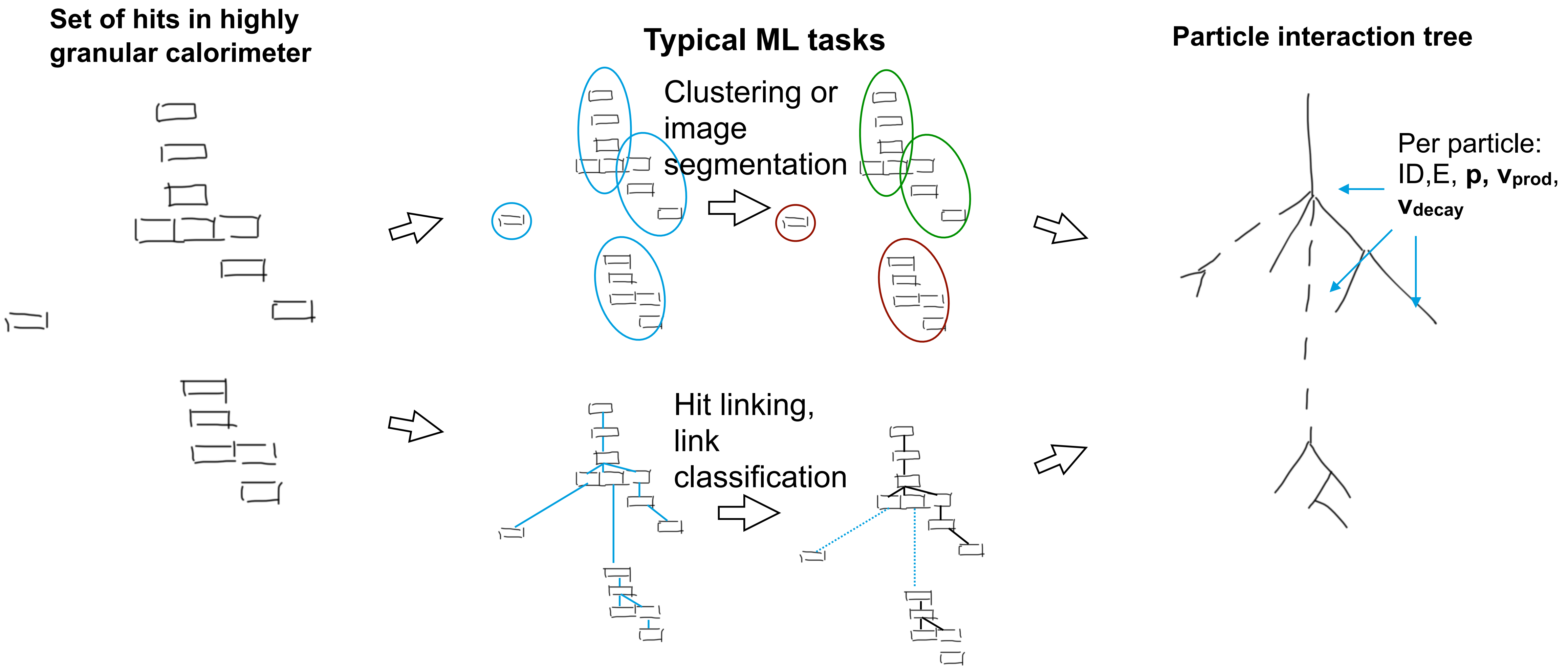


Particle interaction tree



Calorimeter vision for hadronic showers

Ultimate goal and general approach

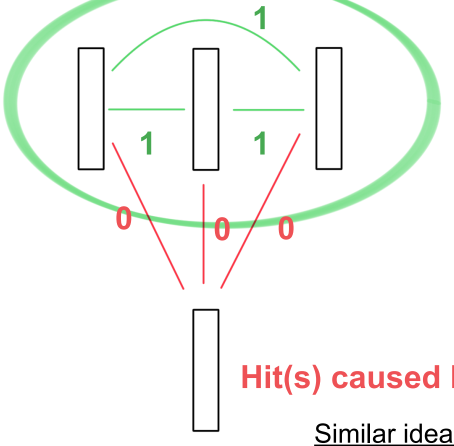


Graph representation of calorimeter event

First steps

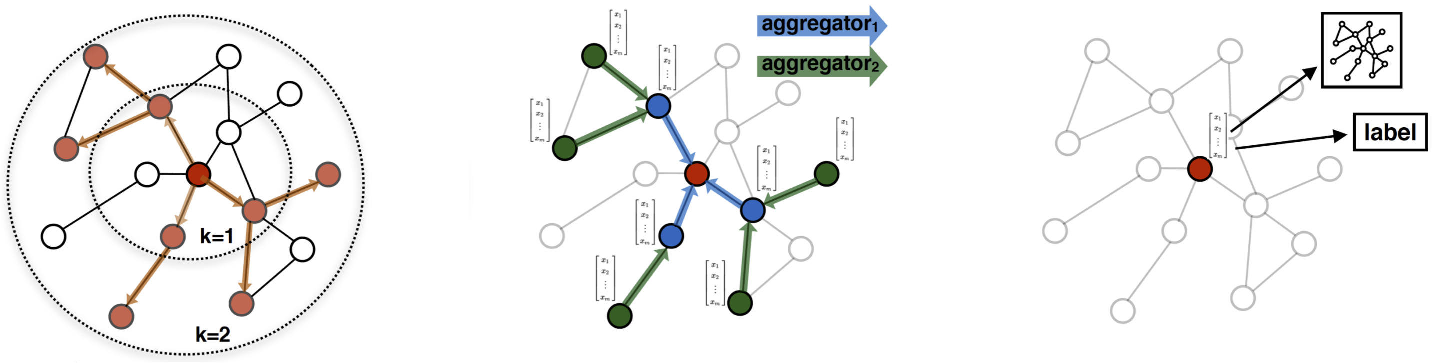
Event graph:

- Nodes - calorimeter hits
- Node features - position, energy, (time)
  - — Edges - neighbours within distance < R_{max} (Radius graph)
  - — Edge weights - 1 if pair of hits belong to same **fundamental object** (e/m sub-shower, track), otherwise 0
- ML **objective** - **predict edge weights** given the radius graph of event



Similar idea is discussed for CMS tracker and HGCAL

GraphSAGE (**SA**mple and aggre**G**at**E**) architecture (Graph neural network model (GNN)):



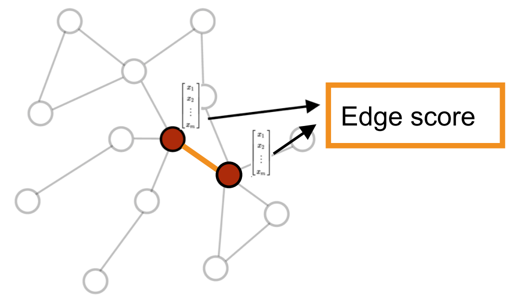
Sample neighbourhood of graph nodes

Aggregate feature information from neighbours

Get graph context embeddings for node using aggregated information

Custom output layer

+



Predict edge score for each pair of connected nodes using embedded features

Fundamental objects in hadronic showers

Shower components

<table>
<tr><td colspan="8">Hadronic shower</td></tr>
<tr><td colspan="4">MIP particles
Mostly charged hadrons</td><td>E/m showering particles
$e^{\pm}, \gamma, \pi^0, \eta$</td><td colspan="3">Challenging particles
n, K_0L, ν, …</td></tr>
<tr><td colspan="3">With shower</td><td>W/o shower</td><td rowspan="4">Topology:
e/m shower</td><td colspan="3" rowspan="2">Topology:
Hadronic shower/escape</td></tr>
<tr><td colspan="3">Topology:
MIP track (before shower) +
Hadronic shower</td><td rowspan="3">Topology:
MIP track</td></tr>
<tr><td>MIP particles</td><td>E/m showering particles</td><td>Challenging particles</td><td>MIP particles</td><td>E/m showering particles</td><td>Challenging particles</td></tr>
<tr><td>(…)</td><td>(…)</td><td>(…)</td><td>(…)</td><td>(…)</td><td>(…)</td></tr>
</table>

Fundamental objects in hadronic showers

Shower components

<table>
<tr><td colspan="8">Hadronic shower</td></tr>
<tr><td colspan="4">MIP particles
Mostly charged hadrons</td><td>E/m showering particles
e$^{\pm}$,γ, π^0, η</td><td colspan="3">Challenging particles
n, K_0L, ν, …</td></tr>
<tr><td colspan="3">With shower</td><td>W/o shower</td><td rowspan="4">Topology:
e/m shower</td><td colspan="3" rowspan="2">Topology:
Hadronic shower/escape</td></tr>
<tr><td colspan="3">Topology:
MIP track (before shower) +
Hadronic shower</td><td rowspan="3">Topology:
MIP track</td></tr>
<tr><td>MIP particles</td><td>E/m showering particles</td><td>Challenging particles</td><td>MIP particles</td><td>E/m showering particles</td><td>Challenging particles</td></tr>
<tr><td>(…)</td><td>(…)</td><td>(…)</td><td>(…)</td><td>(…)</td><td>(…)</td></tr>
</table>

Truth information from Monte-Carlo

Algorithm to find truth e/m objects

Simulations

Geant4 (v10.03.p02) QGSP_BERT_HP using CALICE AHCAL geometry

Pure energy deposition in cells (before digitalisation and reconstruction)

Truth electromagnetic sub-shower definition:

"**Electromagnetic**" particles: $e^{\pm}$, γ, π^0, η

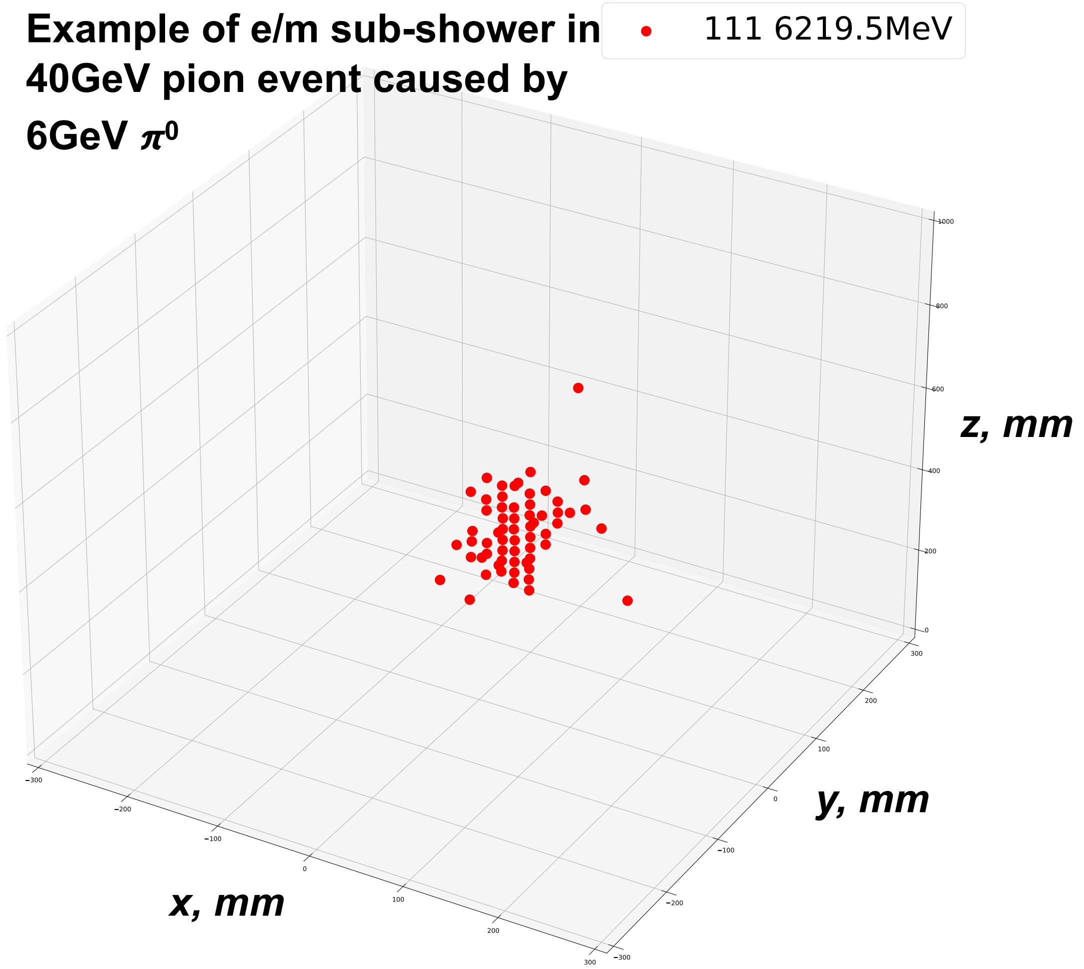
Energy threshold - *0.1GeV* (arbitrary now)

If MC particle is "electromagnetic", all it's "electromagnetic" daughters compose e/m shower and removed from further consideration

Corresponding simulated hits compose sub-shower,

0.5MIP cut: $E_{hit}>0.25MeV$

Example of e/m sub-shower in 40GeV pion event caused by 6GeV π^0



MC history for **ionising particles** is more complicated to easily define individual objects (tracks). Work in progress

Datasets and model parameters

Edge score model

Train&test dataset:

- ~6000 MC event graphs (50/50 split)
  - Pure energy deposition in calorimeter cells (before digitalisation and reconstruction)
  - **10-100 GeV pion** samples

Evaluation dataset

- ~550K tb data events
  - Reconstructed hits
  - **10,20,30,40,60,80 GeV pion** samples

Run #	61264	61269	61384-5	61274	61259	61279
E, GeV	10	20	30	40	60	80
particle	π^-	π^-	π^-	π^-	π^-	π^-
N_{events}	96937	94632	94116	98856	96085	98473
N_{events}(PID)	74066	77638	81041	82759	80865	84124

➡ Radius graphs with calorimeter hit nodes (x,y,z,E_{hit}) R_{max} = ***59 mm***

Model:

GraphSAGE GNN

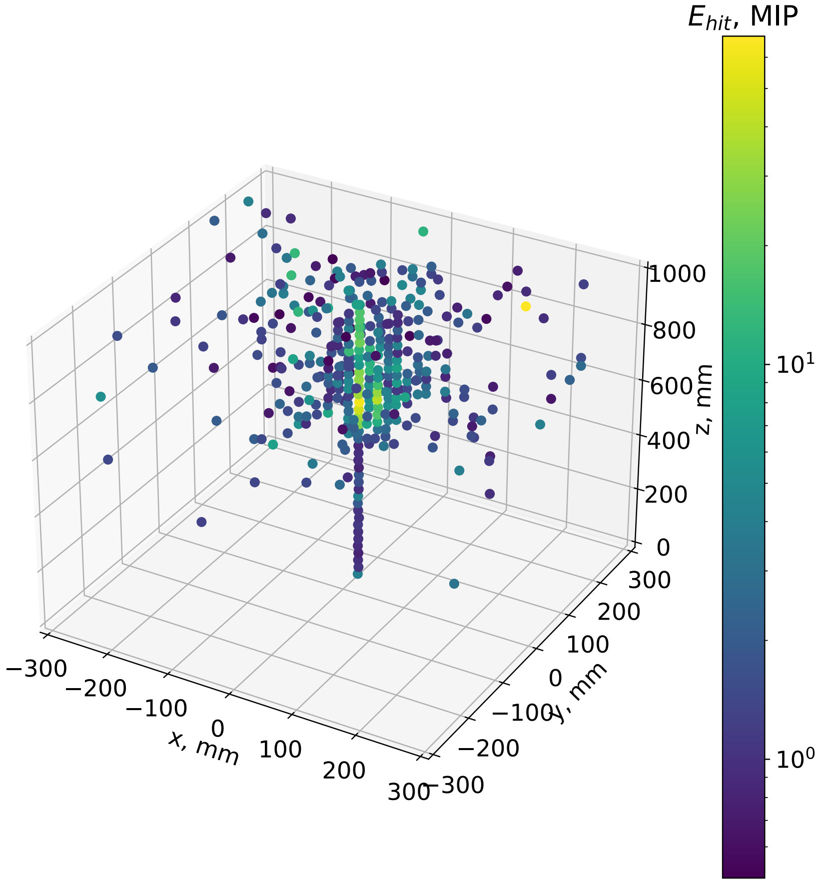
8 layers with 16 hidden channels + 1 linear output layer to convert node embeddings to edge scores

Prediction of edge scores

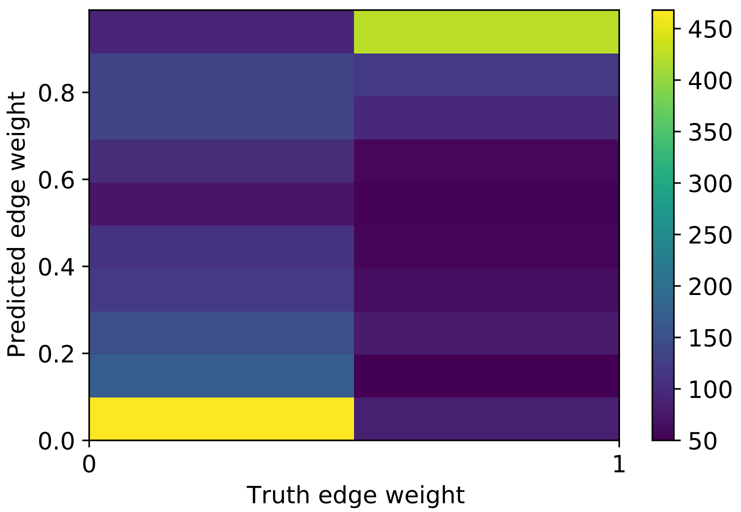
Binary cross entropy loss

Example of output for test event

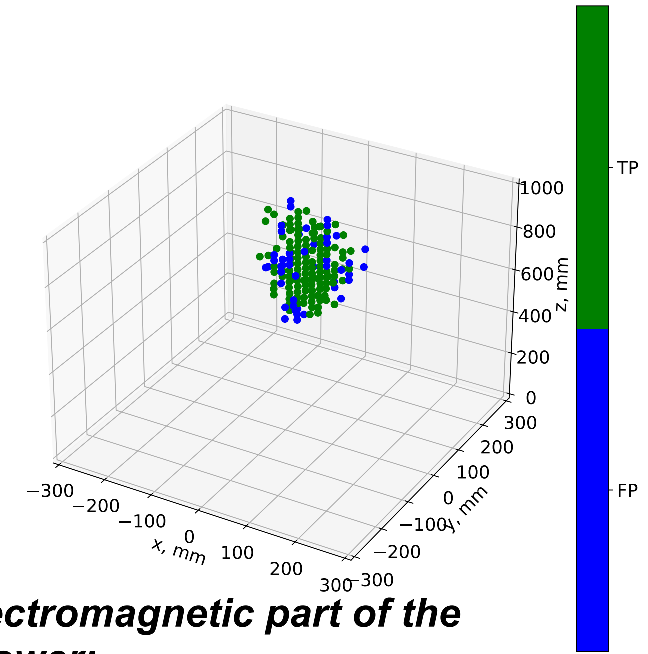
Preliminary results for single test event



2650 graph edges

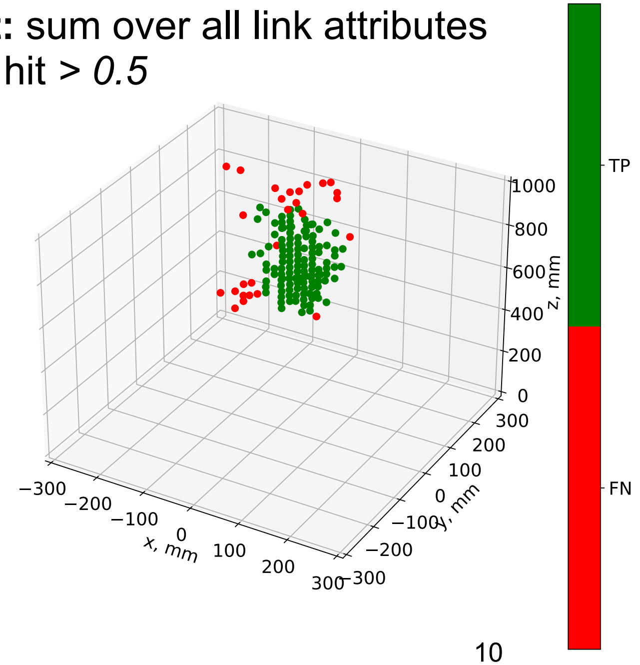


Work in progress…



Electromagnetic part of the shower:

Cut: sum over all link attributes per hit > *0.5*

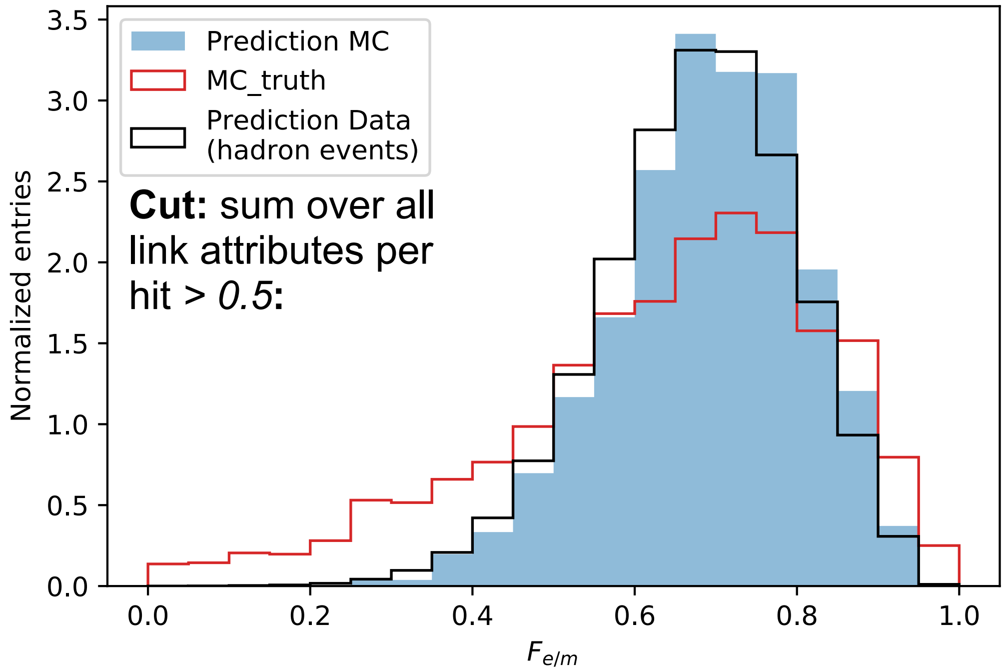


Electromagnetic fraction of hadronic showers

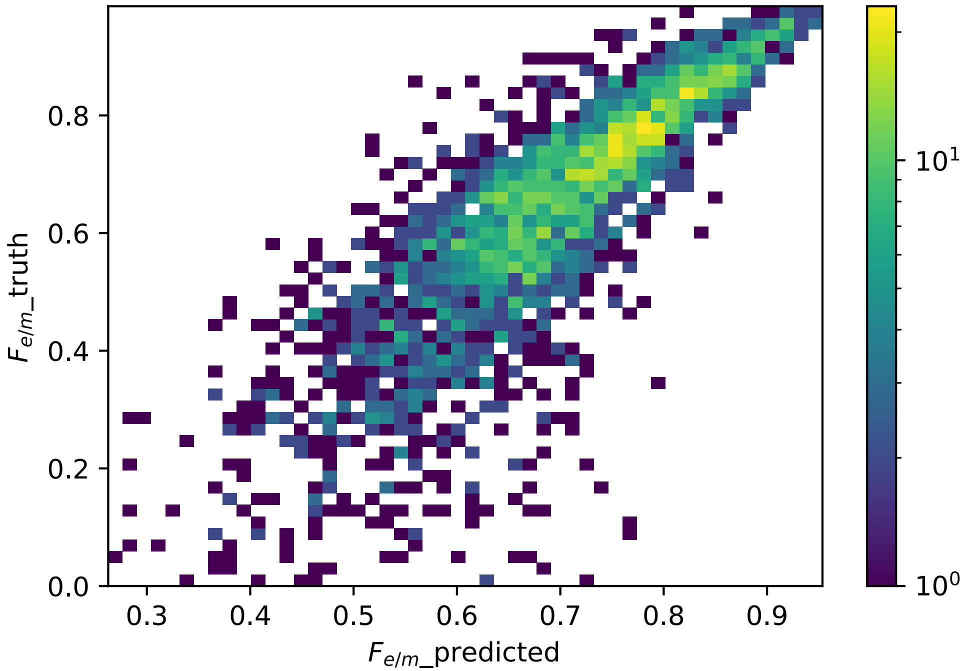
Preliminary results for 10,20,30,40,60,80 GeV pions

Work in progress…

Electromagnetic fraction



Prediction vs truth correlation



- Higher MPV for F_{em} than expected
  - ➡ Non-e/m contributions to the hits are not taken into account
- Predictions for data and MC show good agreement
- Less pronounced tails for F_{em} prediction than for MC truth

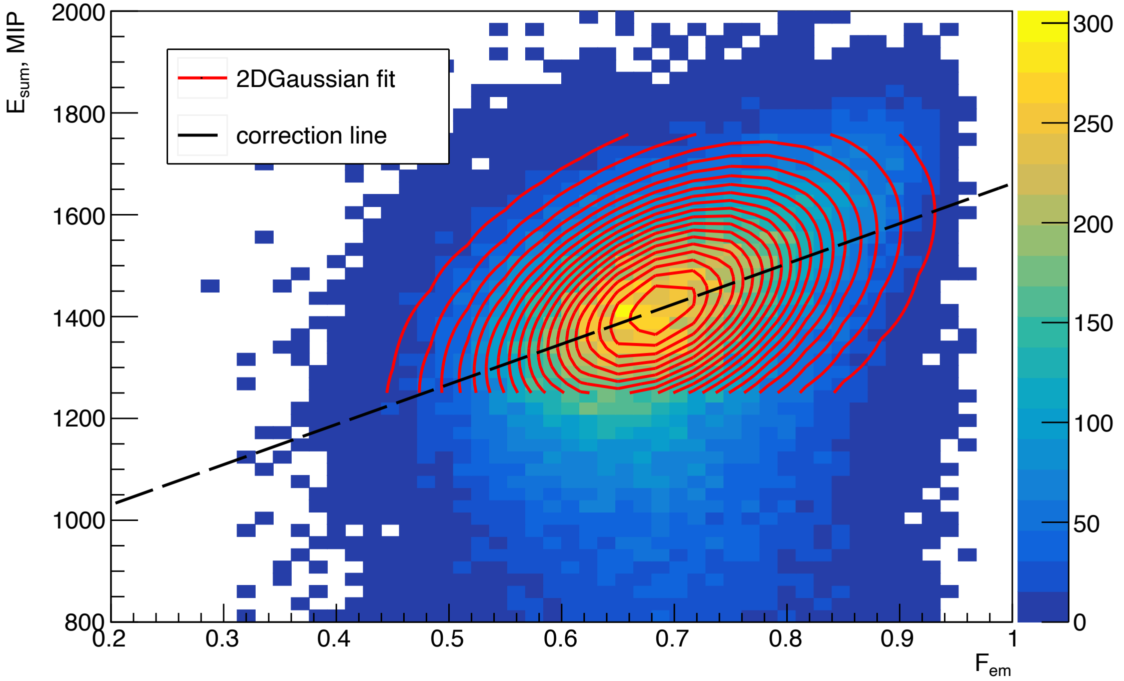
- Satisfying correlation of prediction vs truth

Energy correction

Using reconstructed e/m fraction on test beam data

Work in progress…

Correlation example for 40 GeV pion



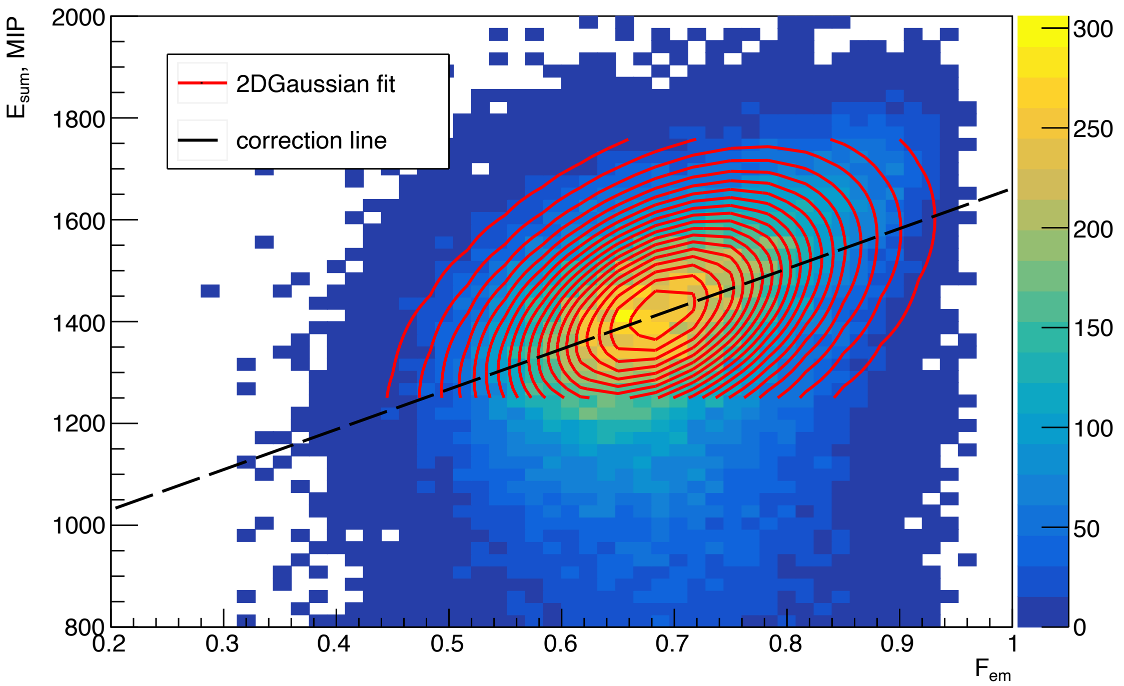
- Well pronounced correlation between E_{sum} and F_{em} observed for all energies
- For each energy point simultaneous gaussian fit is performed to extract the correction line
  - To minimise leakage effects, range for E_{sum} is set $[-\sigma, +2\sigma]$

Energy correction

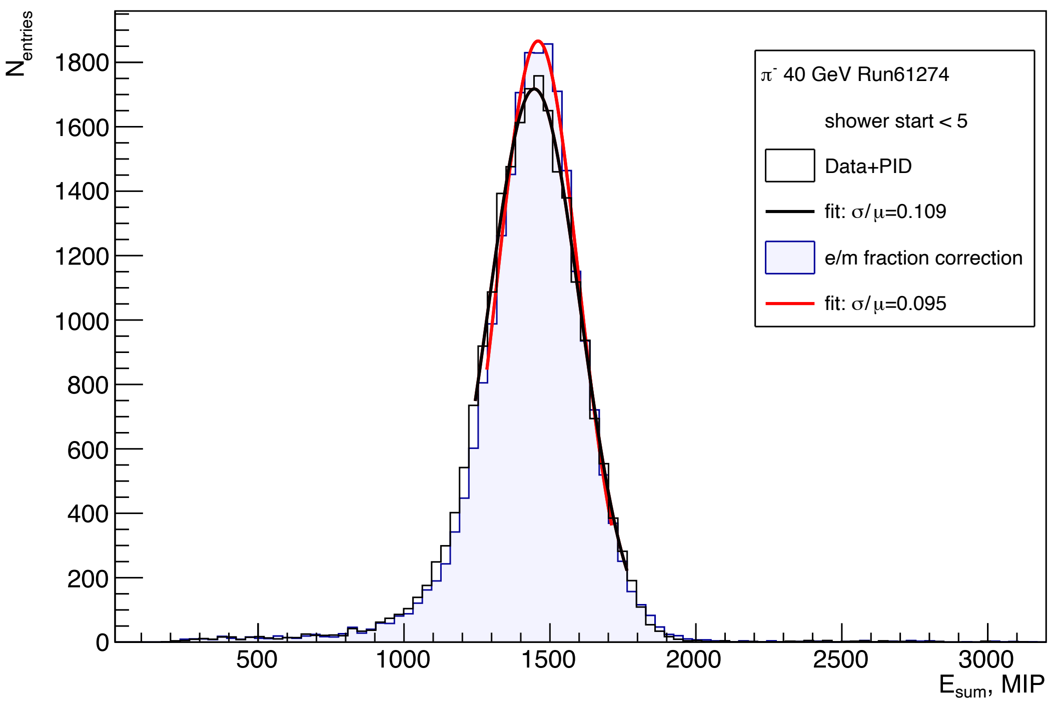
Using reconstructed e/m fraction on test beam data

Work in progress…

Correlation example for 40 GeV pion



Correction example for 40 GeV pion



- Well pronounced correlation between E_{sum} and F_{em} observed for all energies
- For each energy point simultaneous gaussian fit is performed to extract the correction line
    - To minimise leakage effects, range for E_{sum} is set $[-\sigma,+2\sigma]$

- Linear $E_{sum}(F_{em})$ correction:

$E_{cor} = C^*E_{sum}$

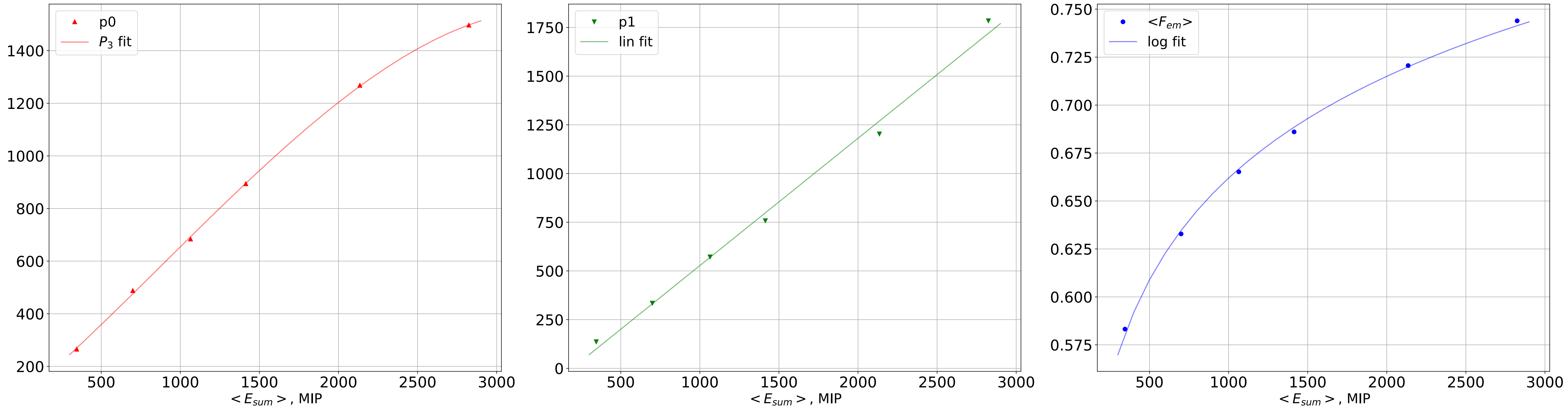
$C = \langle F_{em}\rangle/(p_1 \cdot F_{em} + p_0)$

Unified correction

Getting P_{beam}-independent correction

Work in progress…

Correction parameters as a function of $\langle E_{sum} \rangle$:

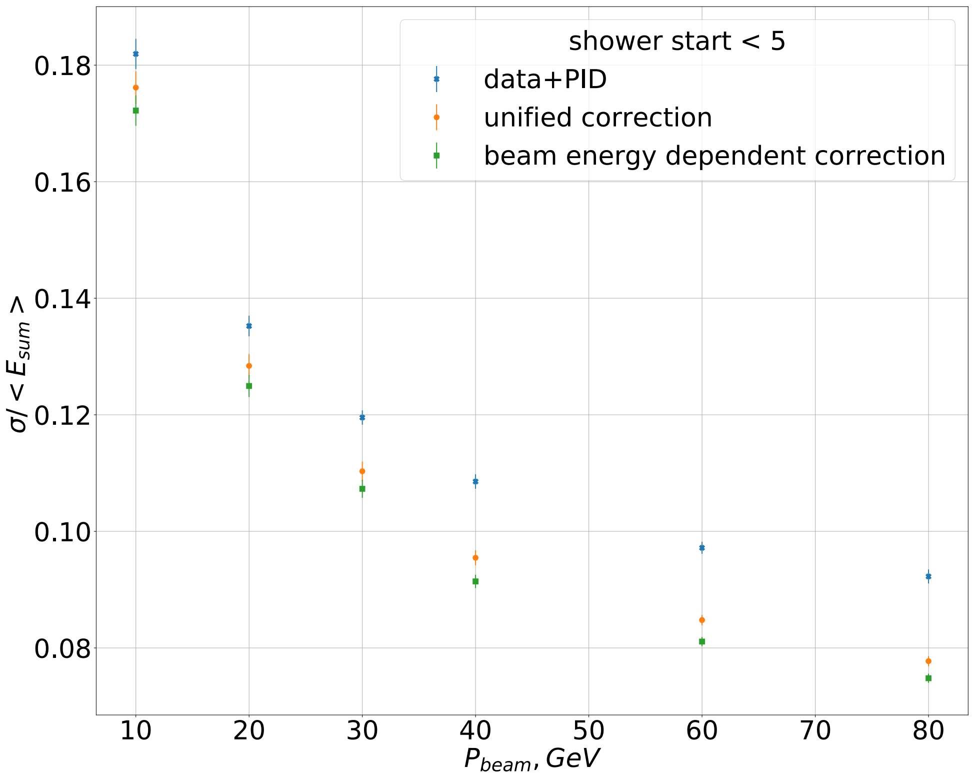


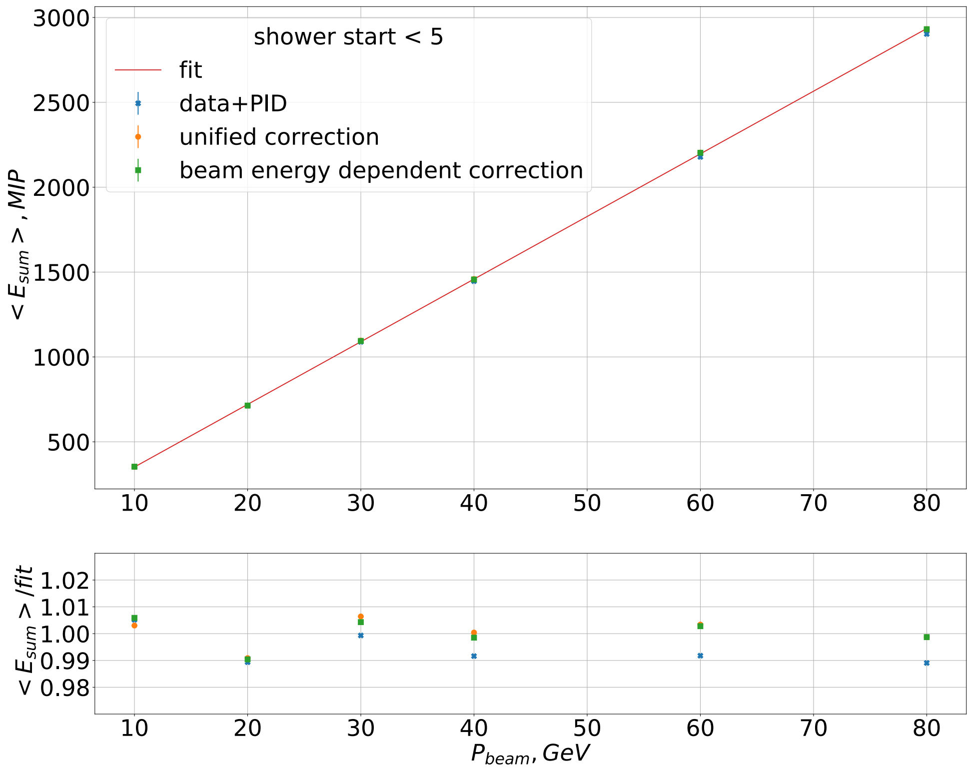
- p_0,p_1 and $\langle F_{em} \rangle$ are calculated for each event from the observed energy using resulting fits
  - More energy points need to be included to check the overfitting
  - Parameter uncertainties are not taken into account

Energy resolution and linearity

10-80 GeV data

Work in progress…





- Resolution improvement:
  - ~6-23% for run energy dependent correction
  - ~3-19% for unified correction
- No meaningful fit results with standard parametrisation yet

- Linear response after correction
  - within 1%

Summary & outlook

- [x] Simulations provide detailed information about hadronic shower sub-structure
- [x] GNN shows promising results to act as tool for reconstructing individual particles within hadronic showers
  - [x] GNN reconstruction of electromagnetic part within hadronic shower gives physically meaningful output
  - [ ] Include reconstruction of corresponding hit energy fractions
  - [ ] Individual particle separation method is under development
  - [ ] Repeat study for reconstruction of ionising tracks within hadronic showers
- [x] New software compensation method using reconstructed e/m fraction with GNN
  - [x] Energy resolution improvement with good linearity
  - [ ] To be parametrised and compared with simulations and existing software compensation methods

Backup slides

Constructing e/m link graph

Input and training target

Input graph

- Radius graph with calorimeter hit nodes (x,y,z,E) *Rmax=**59mm***
- Links - matrix of shape (Nlinks, 2) represent whether his are closer than Rmax, have no attribute

Hit_i	…	…	…	…
Hit_j	…	…	…	…

N_{links}

Target for e/m sub-showers

- For all e/m objects construct it's own radius graph and link attribute equal 1
- Update full event graph with new link atributes

0	1	…	1	0

N_{links}

E/m showers

Algorithm to find truth objects

"**Electromagnetic**" particles: $e^{\pm}$,γ, π^0, η

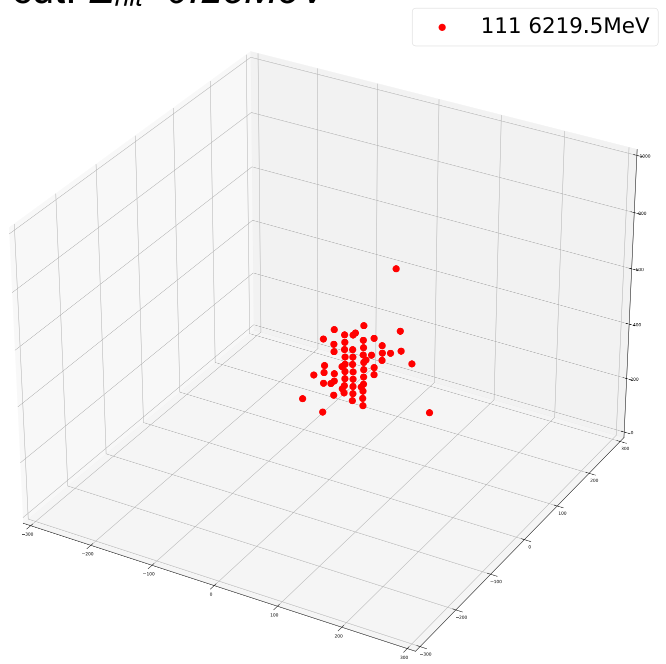
Energy threshold - *0.1GeV* (arbitrary now)

If MC particle is "electromagnetic", all it's "electromagnetic" daughters compose e/m shower and removed from further consideration

Corresponding simulated hits compose sub-shower,

0.5MIP cut: $E_{hit}>0.25MeV$

•



"**Ionising**" particles: $\pi^{\pm}$,$p^{\pm}$, $\mu^{\pm}$

Energy threshold - *0.1GeV* (arbitrary now)

If MC particle is "Ionising", all it's "Ionising" daughters compose ion shower (not individual tracks)

Corresponding hits are collected from SimCalorimeterHit collection ($E_{hit}>0.25MeV$)