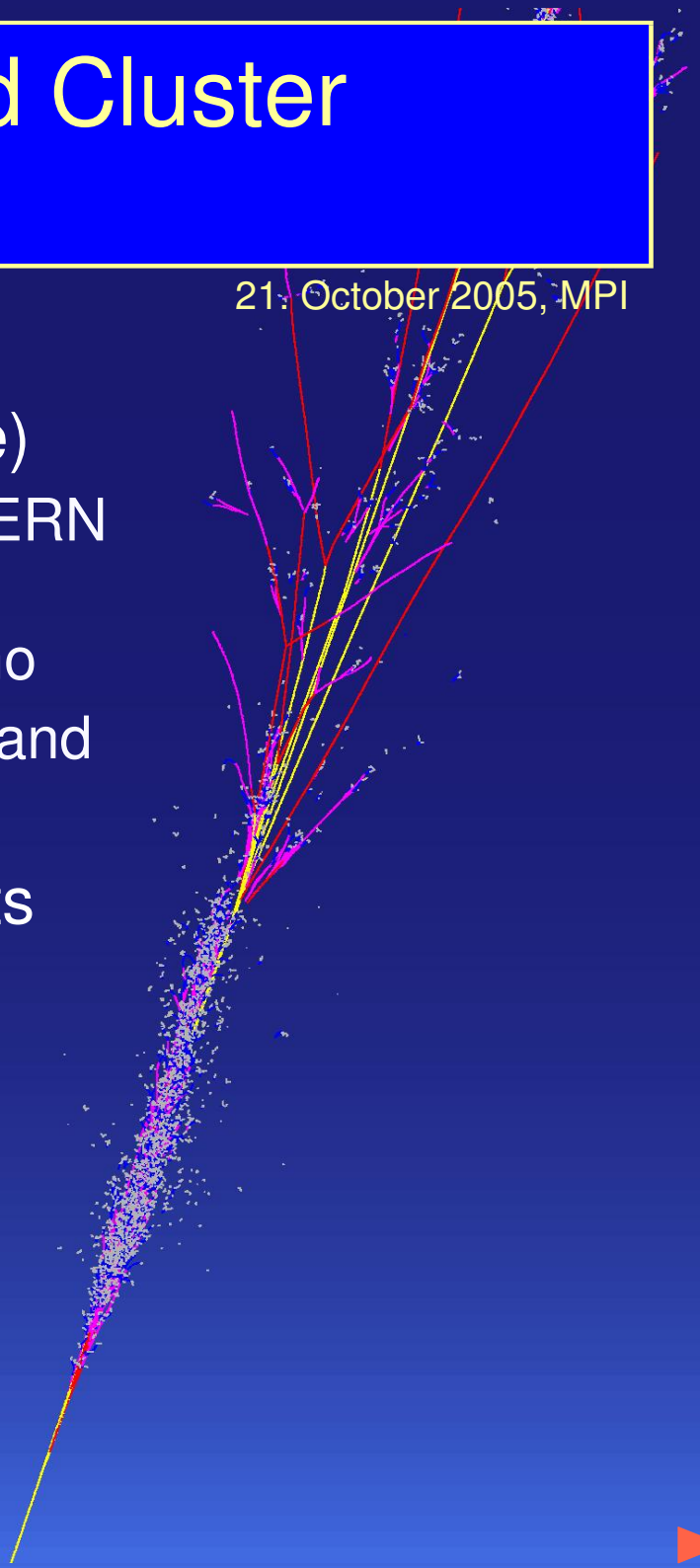
Hadronic Weights and Cluster Classification

- Status of Single Pion Production (done)
  - simulated files from RZG o.k. – from CERN buggy
  - digitized files – no news since Gargnano
  - reconstructed files – found some bugs and re-reconstructed all
- Preliminary results from Calibration Hits
  - definition of energy/eta regions
  - cluster classification
  - weights
  - application of weights
- Implementation in athena
- Conclusions



- 400 k single pions (half positive, half negative) generated in batches of 40 k for 1, 3, 5, 10, 20, 50, 100, 200, 500, 1000 GeV with flat η ($|\eta| < 5$) and flat ϕ ($|\phi| < \pi$)
- production was done at Munich and CERN from 31. July to 13. August based on `10.3.0` with some patches mainly found by Kai (see ATLAS internal web page http://www.cern.ch/menke/cgi-bin/hec/postrome.sh for details)
- each simulated file contains 50 events with normal hits and calibration hits (including FCal)
- naming for the files is like for Rome (same run numbers as for the P*N*P series) but with the prefix postrome instead of rome
- the root and log files can be found in my castor area at CERN: `/castor/cern.ch/user/m/menke/postrome.0040[78][0-9].simul.P{[0-9],10}P_Single{211,211minus}`
- All files simulated at CERN miss some FCal calibration hits (although same release, tags etc.)
- GOOD runs are 0040{72,77,78,81,82,83,84,85,86,87,89}

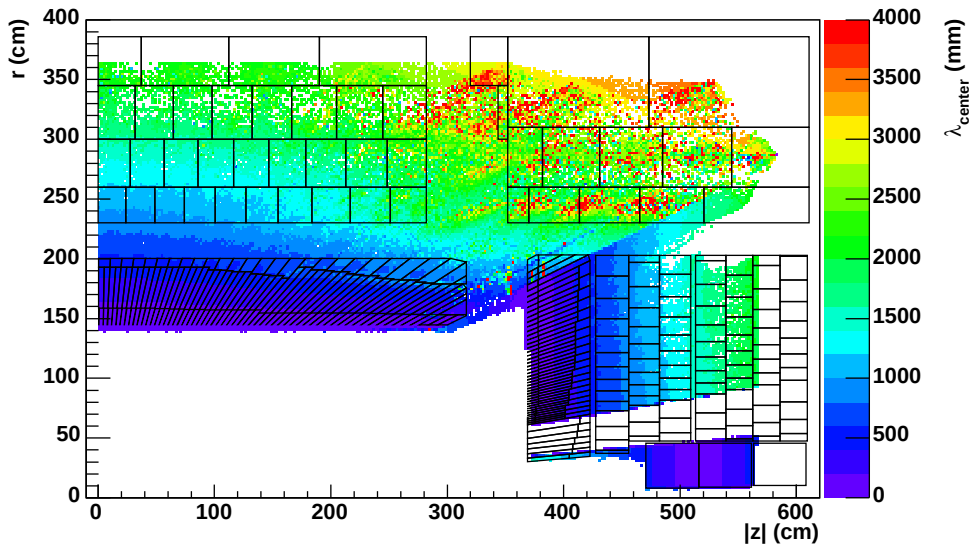
- nothing new since Gargnano
- digitization was done with the same `10.3.0` based release at CERN and Munich in batches of 10 files (i.e. 500 events) per job
- each of the 800 digit files is 700 MB large due to the simulation of noise in the calorimeters
- the root and log files can also be found in my castor area at CERN:
  `/castor/cern.ch/user/m/menke/postrome.0040[78][0-9].digitcalib.P{[0-9],10}P_Single{211,211minus}`

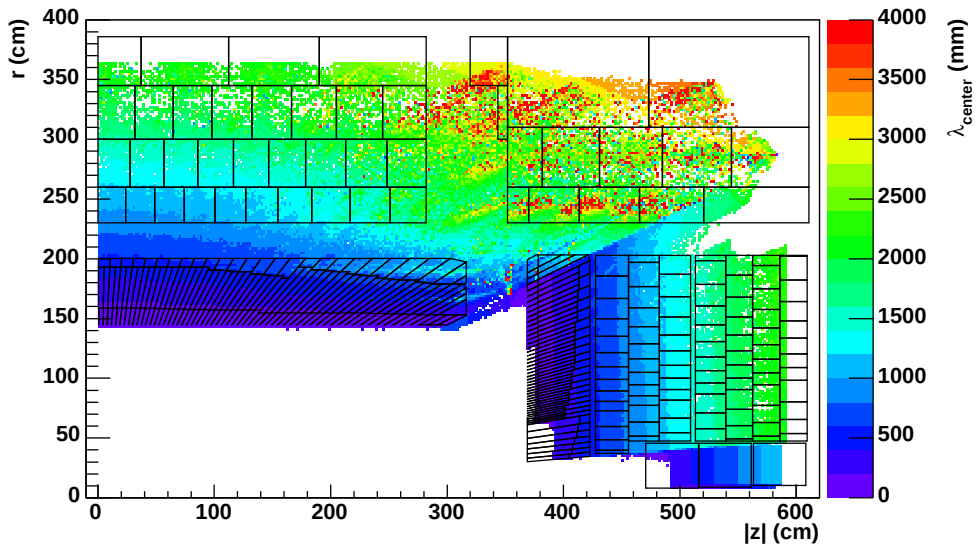
- I found some bugs in the treatment of calibration hits and inconsistent use of some tags relevant for the moments after Gargnano
- reconstruction was re-done with `atlrel_2` from September 19th plus `CaloCalibHitRec-00-00-12` and `CaloRec-02-06-21` at CERN
- the optional 5 new moments were enabled for topo clusters
- each reco file is based on one digitcalib file (i.e. 500 events)
- the root files contain the `CBNT` with `truth`, `topo`-cluster, and `EMtopo`-cluster blocks including all cell details for each cluster and calibration hit info for each cell inside each `topo`-cluster.
- they contain also a separate dead material ntuple which is also used for the plots I show in the second half of the talk
- as the simulated and digitized files the root and log files can be found in my castor area at CERN:
  `/castor/cern.ch/user/m/menke/postrome_rel2.0040[78][0-9].reco.P{[0-9],10}P_Single{211,211minus}`
- the old reco files are still in castor with the original names; the new ones are called `postrome_rel2` instead of just `postrome`

- Analysis is based on roughly half the files (since about half the files were simulated at CERN and those are buggy)
- For $1\,\mathrm{GeV}$ and $5\,\mathrm{GeV}$ both sets were simulated at CERN – all other energies have at least one set simulated in Munich
- all histograms are made in 25 separate $|\eta|$ regions in steps of $\Delta|\eta| = 0.2$ from $0.0 \leq |\eta| < 0.2$ to $4.8 \leq |\eta| < 5.0$, where η is either the first moment over η of the positive cells inside the current cluster or the η of the generated pion.
- 10 bins for the 10 generated energies are used for the final performance control histograms
- cluster classification histograms are made in 5 logarithmic cluster energy intervals ($0\,\mathrm{GeV} < E_0 < 1\,\mathrm{GeV} < E_1 < 4\,\mathrm{GeV} < E_2 < 16\,\mathrm{GeV} < E_3 < 64\,\mathrm{GeV} < E_4$)
- cell weight histograms are made for each calorimeter sampling with 20 bins in $\log_{10}\left(E_{\mathrm{clus}} \times [1\,\mathrm{MeV}^{-1}]\right)$ from $E_{\mathrm{clus}} = 100\,\mathrm{MeV}$ to $1\,\mathrm{TeV}$ times 20 bins in $\log_{10}\left(\rho_{\mathrm{cell}} \times [1\,\mathrm{mm}^3\,\mathrm{MeV}^{-1}]\right)$ from $\rho_{\mathrm{cell}} = 10^{-7}\,\mathrm{MeV\,mm^{-3}}$ to $10\,\mathrm{MeV\,mm^{-3}}$

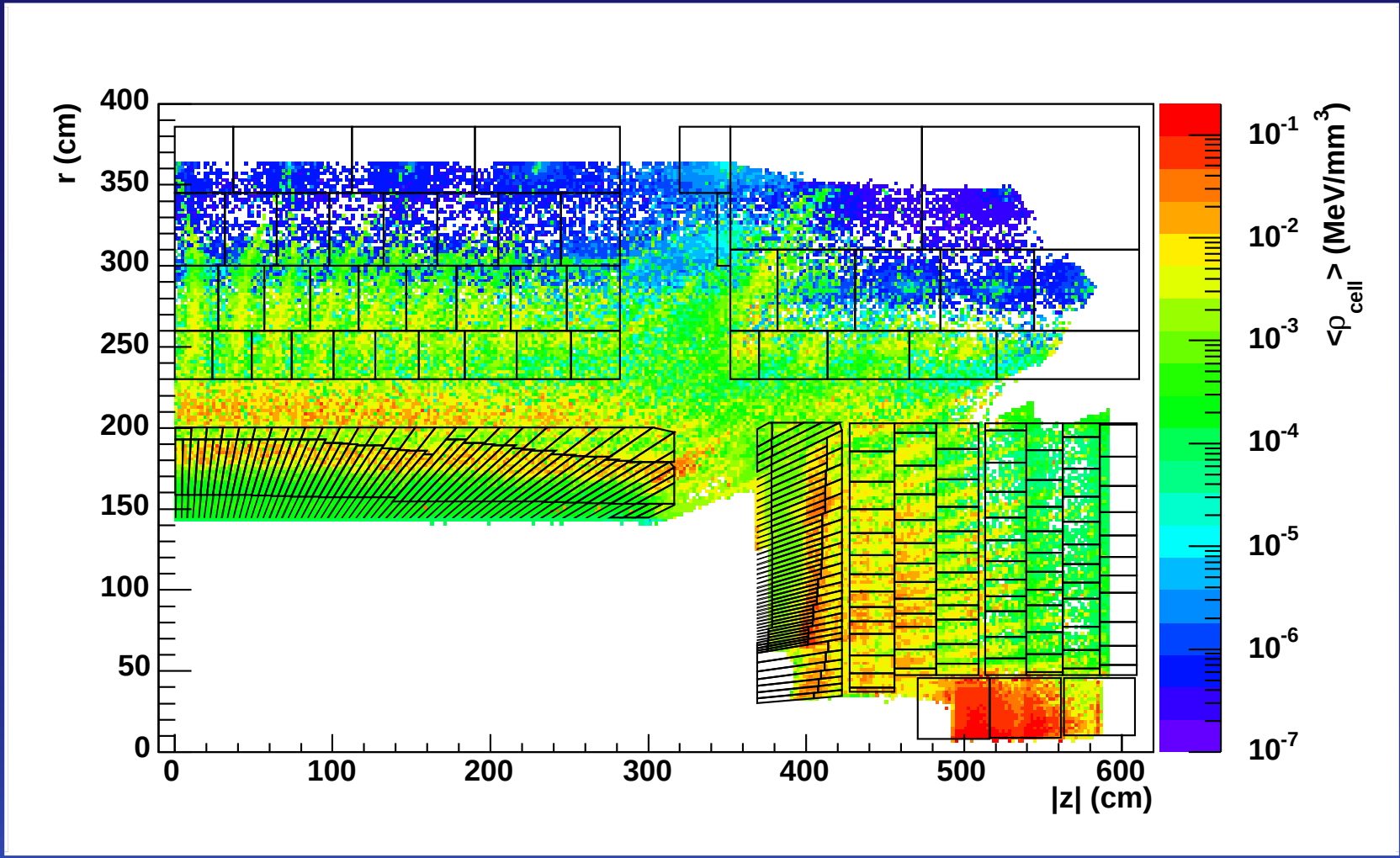
- needed to characterize clusters in order to classify them as electromagnetic or hadronic
- based on all cell members of the cluster with positive energy
- by default the following 13 moments are calculated:
  - x, y, z-position of the cluster centroid
  - first moments in η and ϕ
  - deviations of the cluster principal axis from IP-axis in θ, ϕ and absolute (α)
  - second moments in r and λ, with r (λ) being the radial (longitudinal) cell distances from the shower axis (center)
  - depth of the shower center (λ)
  - normalized lateral and longitudinal moments
- optionally the following 5 moments can be calculated:
  - first and second moment in energy density ρ
  - energy fraction in EM calorimeters
  - energy fraction in most energetic cell
  - energy fraction in sum of most energetic cells per sampling

- λ_{center}: depth of the (energy weighted) center of the cluster for 200 k single pions from 3 GeV to 1 TeV
- left plot: Status from Gargnano with tag mismatch resulting in some geometry shifts:
  - $\lambda_{center} = 0$ for endcap inner wheel
  - λ_{center} w.r.t. FCal2 for forward region
  - cell position is on front face for LAr; center for Tile
- right plot: New reco files (postrome_rel2):
  - λ_{center} o.k. for endcap inner wheel and FCal
  - cell position is in cell center for all calos



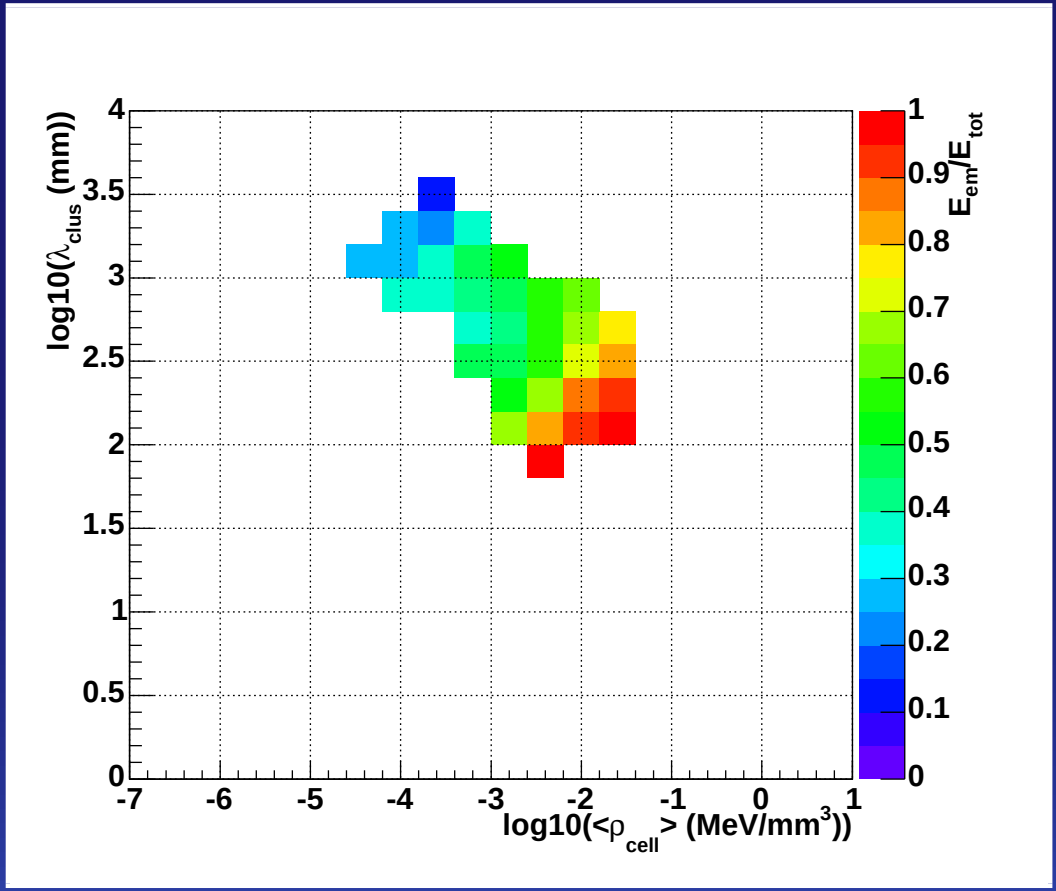


- $\langle \rho_{\text{cell}} \rangle$: energy weighted average (first moment) of cell energy densities inside the cluster for 200 k single pions from 3 GeV to 1 TeV



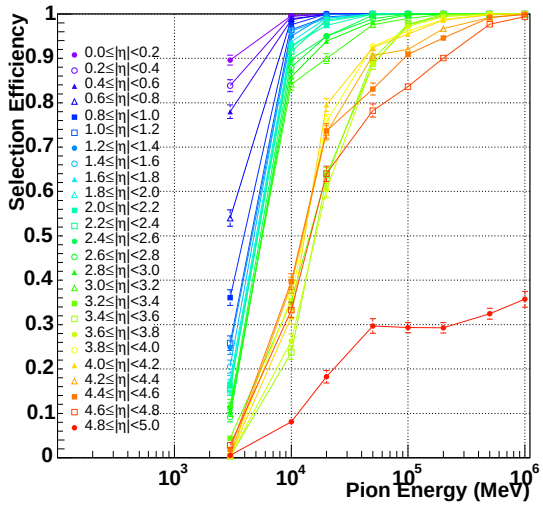
- Cluster Classification is made in 2D
- Currently I favor two moments for the binning:

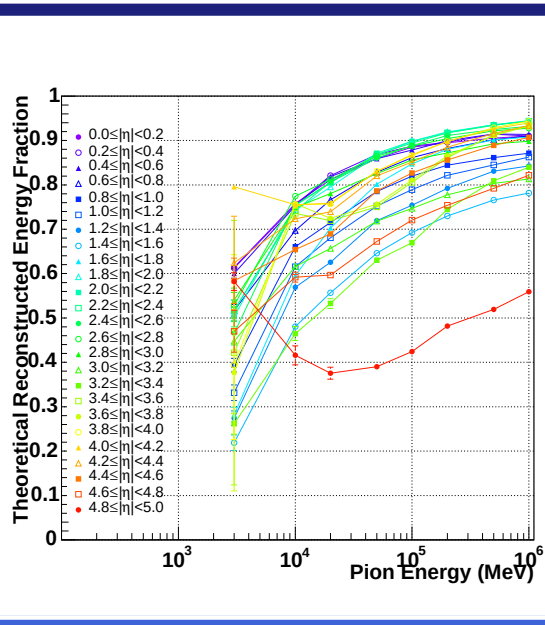
  - the shower depth λ_{clus} (i.e. the distance of the shower center from the calorimeter front along the shower direction); deep showers tend to be more hadronic in nature
  - the first moment over the cell densities $\langle \rho_{\text{cell}} \rangle$ (i.e. the energy weighted average over the cell densities); high average densities for a given cluster energy indicate e.m. activity
  - the color coded z-axis shows the e.m. fraction of the cluster energy
  - the example plot is for the region $2.0 \leq |\eta| < 2.2$ and $4\,\text{GeV} \leq E_{\text{clus}} < 16\,\text{GeV}$



- Hadronic weighting can be made for clusters with e.m. fractions below 85 %. Cutting at 80 % would degrade the weighting

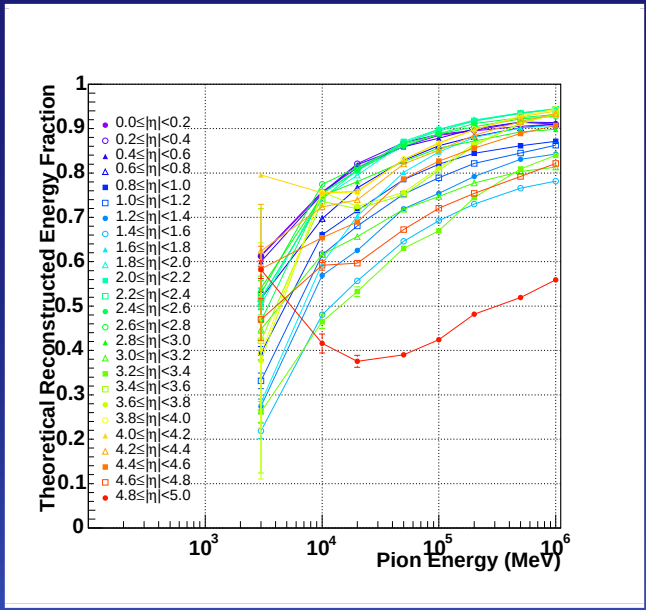
- The nominator for the weights is the sum of all calibration energies for a given cell
- The sum of all calibration hits (times the cluster-weight for the cell) of all cells in all clusters in an event is the total deposited energy that can be restored by weighting
- The ratio of the average of this theoretical energy over the momentum of the simulated pion is shown in the lower plot for all 25 $|\eta|$-regions as function of the energy of the simulated pion
- Below $\sim 10\,\mathrm{GeV}$ less than 90 % of the pions leave a reconstructible energy deposition in the calorimeters (upper plot shows efficiency)
- Below $\sim 10\,\mathrm{GeV}$ less than 70 % of the pion energy is inside the selected calorimeter cells.
- Even at highest energies it never exceeds 95 %
- The fraction degrades in the gap region $1.0 < |\eta| < 1.8$ and the forward region $3.0 < |\eta|$

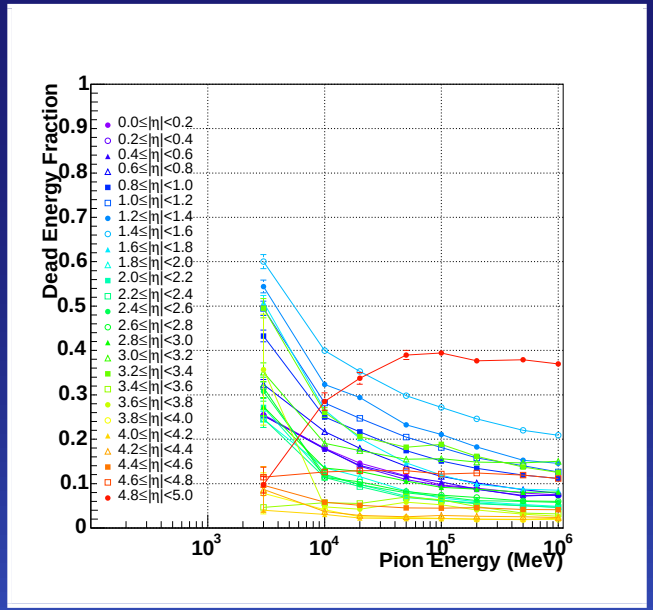


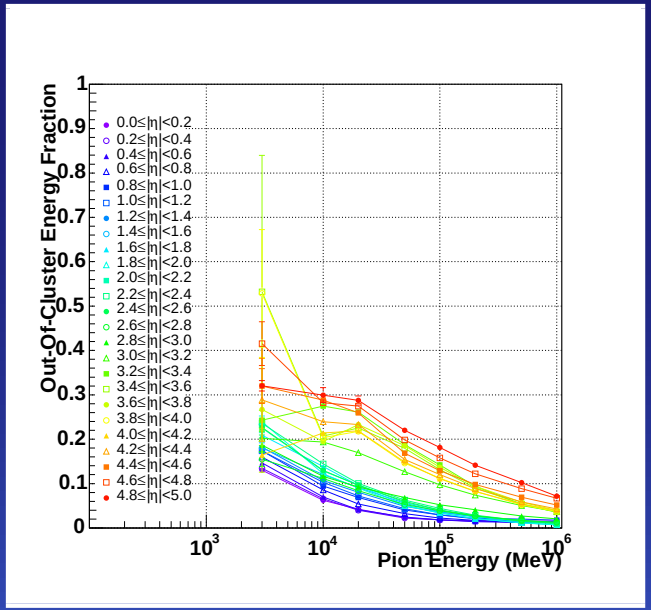


Preliminary Results from Calibration Hits ► Total Energy Deposits

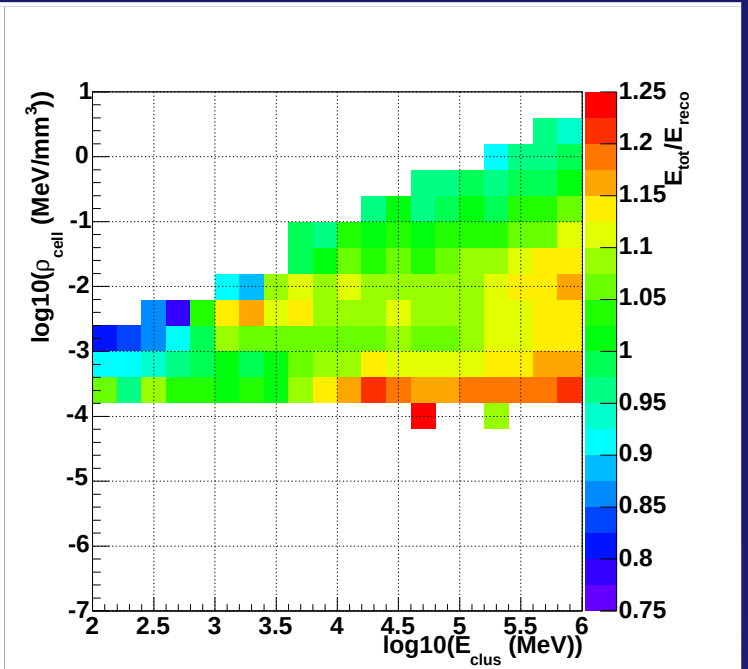
- Only the total sum of all calibration hits inside the clusters can be regained by weighting (left plot, repeated from previous slide)
- The deposits in dead material i.e. outside the calorimeters (middle plot)
- And inside the calorimeters but outside the clusters (right plot) need additional corrections

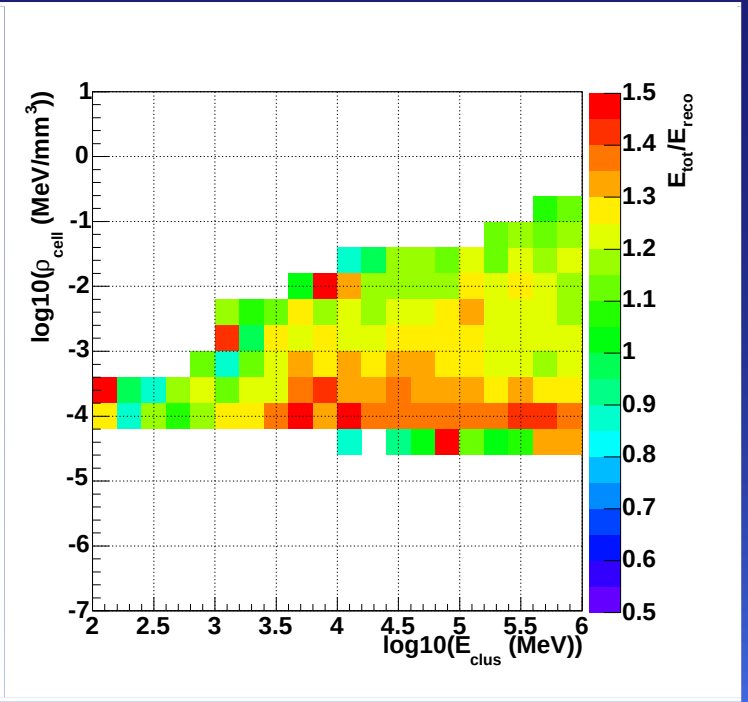




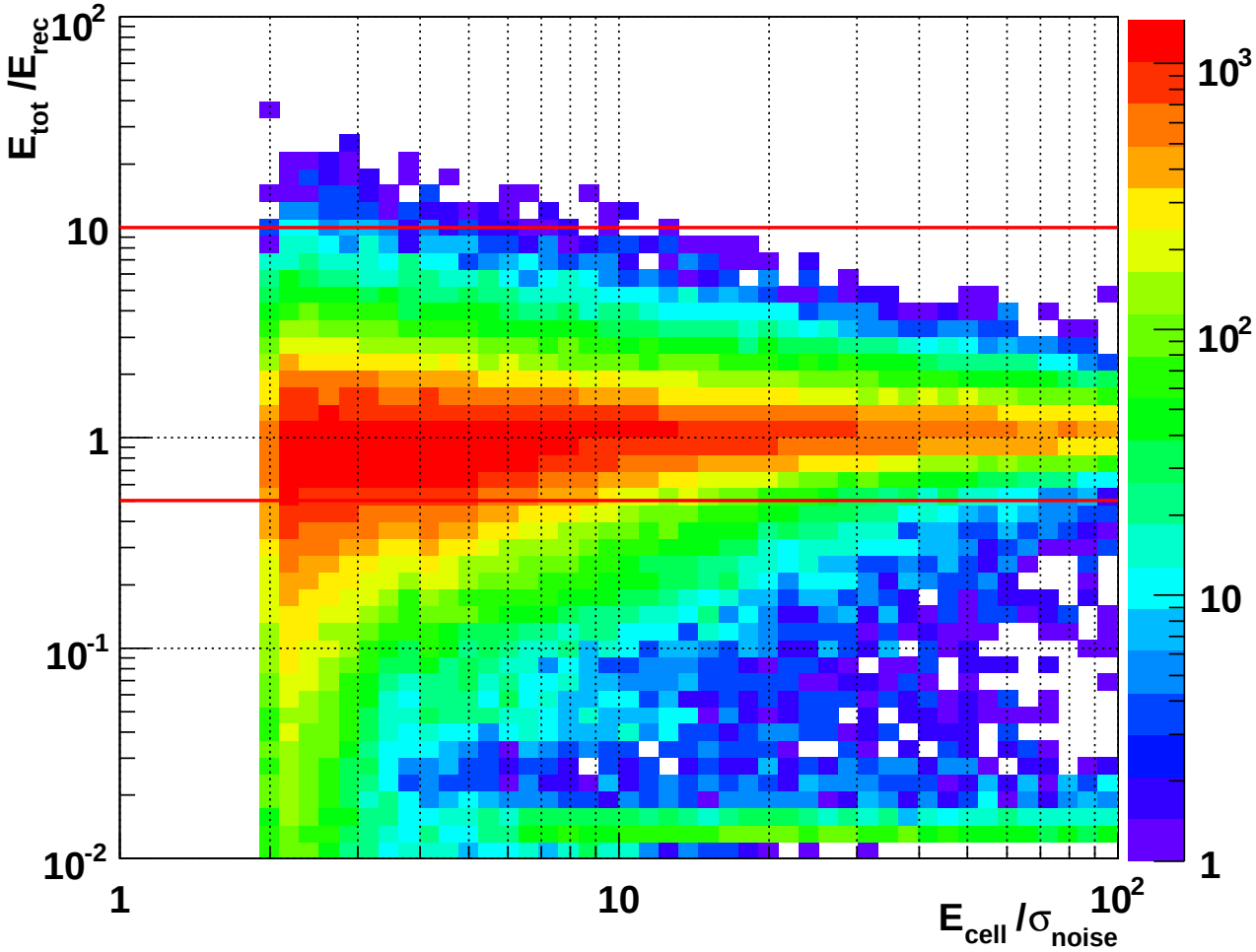


- Cell weights are done in 2D as function of the logarithm of cluster energy E_{clus} and the logarithm of the cell energy density ρ_{cell}
- Only cells with $E_{\text{cell}} > 2\sigma_{\text{noise}}$ are used to define weights – it does not make sense to weight noise
- Only cells with $0.5 < E_{\text{tot}}/E_{\text{cell}} < 10$ are considered for the weights – values outside this range are still dominated by noise (see next slide)
- Each sampling and $|\eta|$-region gets its own cell weights
- The examples shown are for $2.0 \leq |\eta| < 2.2$
- Upper plot is for EMEC Layer 2
- Lower plot is for HEC Layer 1
- The asymptotic approach of the weights to $w = 1$ with increase in ρ_{cell} is visible
- The differences between layers and subsystems is much larger than the energy dependency

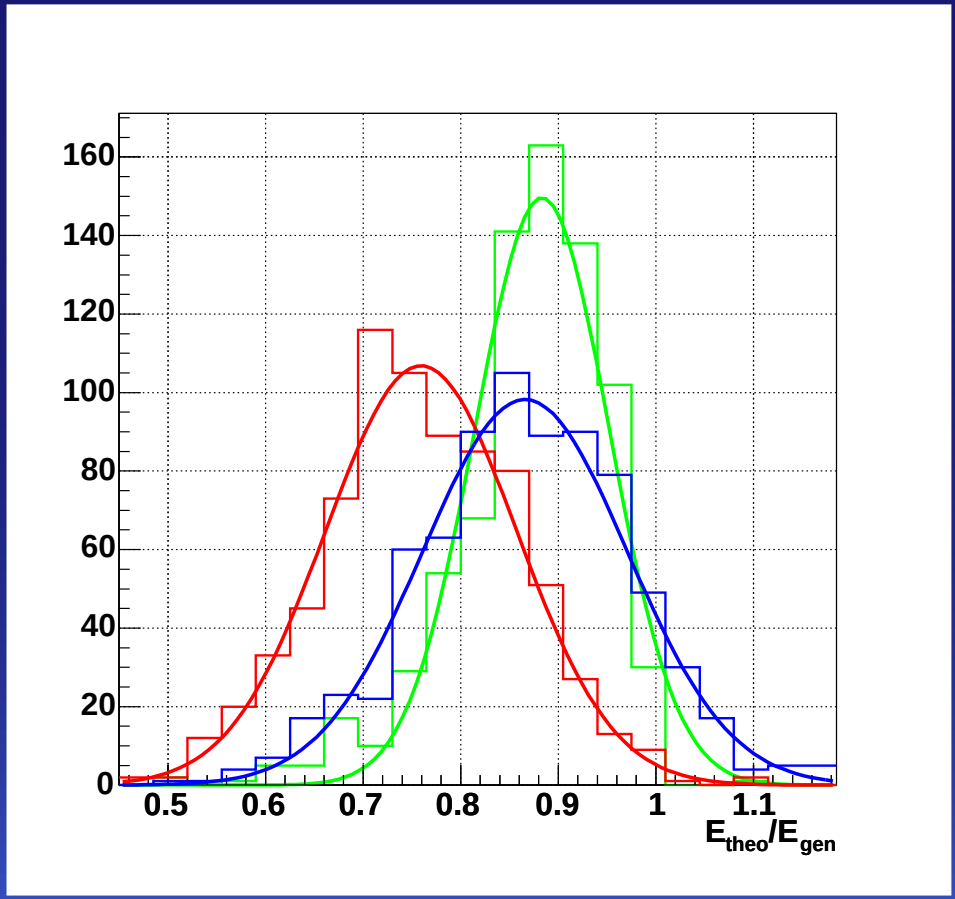




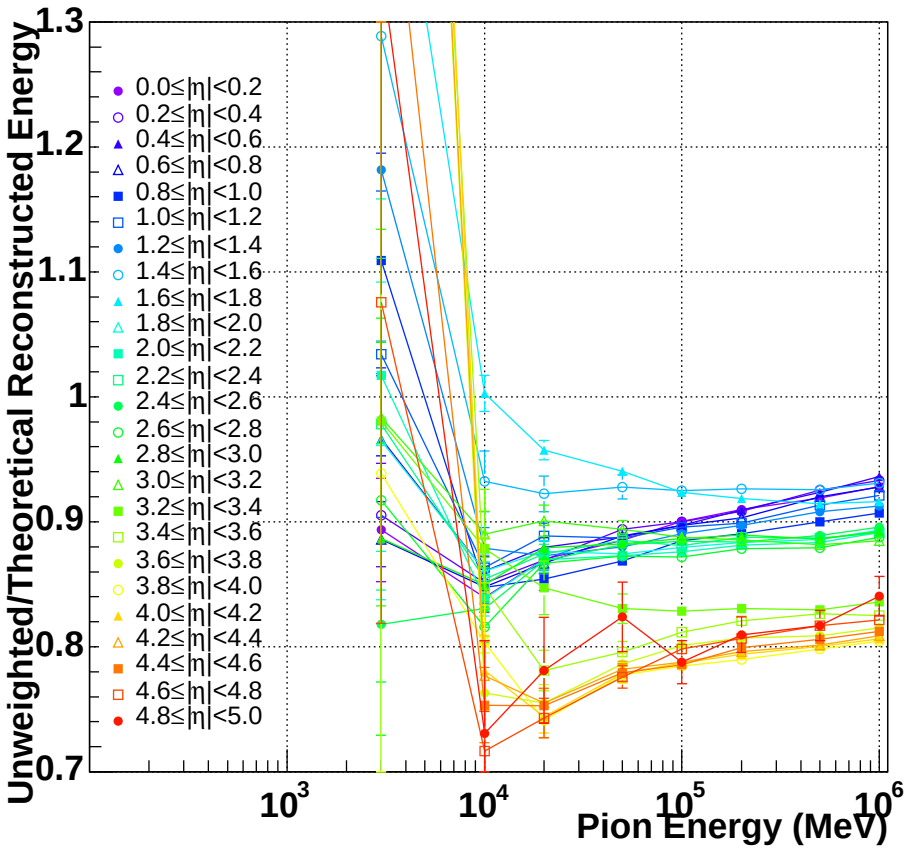
- Only cells with $0.5 < E_{\text{tot}}/E_{\text{cell}} < 10$ are considered for the weights – values outside this range are still dominated by noise
- plot shows $\langle E_{\text{tot}}/E_{\text{cell}} \rangle$ vs. $E_{\text{cell}}/\sigma_{\text{noise}}$

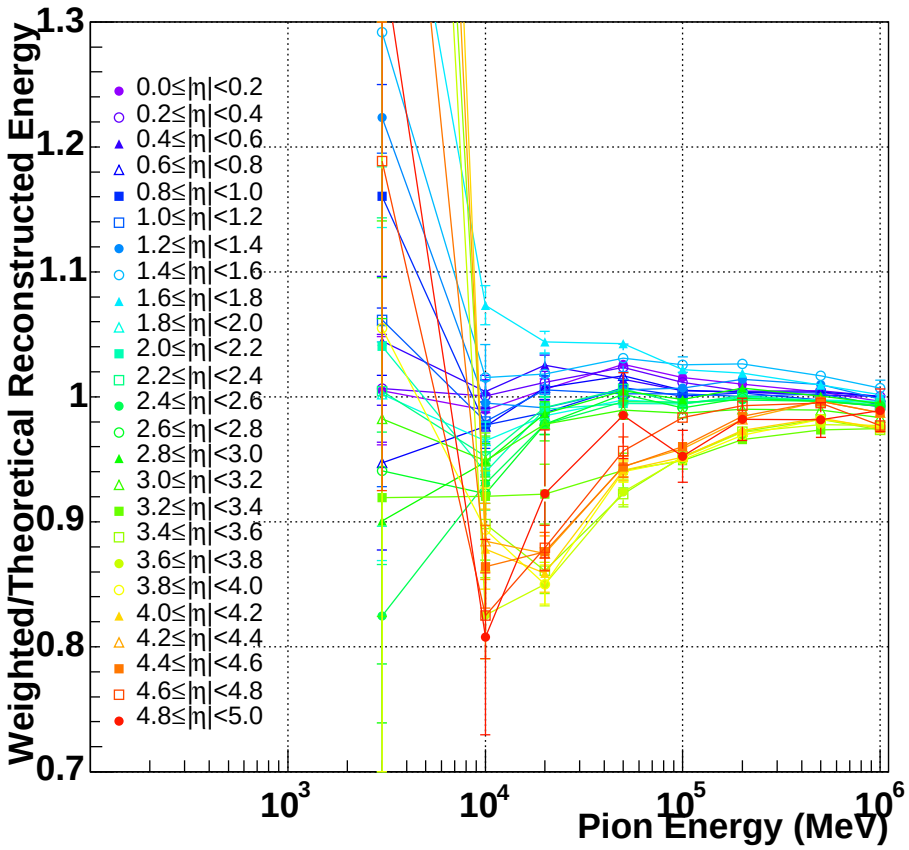


- All clusters with $f_{\text{e.m.}} < 0.9$ are weighted (only $2.4‰$ of all clusters are not weighted)
- The weights are applied to all cells with $|E_{\text{cell}}| > 2\sigma_{\text{noise}}$ (note absolute value here but not for the definition of weights)
- $|E_{\text{cell}}|$ instead of E_{cell} is used in lookup of the weight, but the sign of E_{cell} is preserved after weighting
- Only weights with at least 10 entries in the weight histogram are used
- Plot shows weighting example for 50 GeV single pions at $2.0 \leq |\eta| < 2.2$ with unweighted (red), expected (green), and weighted (blue) reconstructed energy
- Mean is correct after weighting
- Resolution improves only slightly from 12.9 % to 12.1 %



- Plots show ratio of reconstructed over expected energy before (left) and after (right) weighting





- weighting works for energies above 10 GeV
- FCal might need additional weight for cells with noise

Implementation in athena

- Both classification and weighting has been implemented in athena for `11.0.0`
- The number of constants needed for classification is ~ 48000 and for weighting ~ 52000
- They are stored as `TProfile2D` objects in 2 root-files in `CaloClusterCorrection/share`
- Need some database solution in the future
- In the last nightlies before `11.0.0` was built classification and weighting was working
- Will check `11.0.0` as soon as it becomes available

Conclusions

- We finally have 400 k Rome-like single pions with calibration hits
- Some problems with simulation have been discovered since LAr week in Gargnano (files excluded)
- Some problems with reconstruction for LAr geometry and calibration hits have been fixed (redone all reconstructions)
- plots shown are based on the 200 k pions simulated in Munich
- preliminary results shown:
  - dead material and clusterization effects become large for pion energies below $\sim 10\,\mathrm{GeV}$
  - modest η dependencies with worst response in the gap regions
  - cluster classification based on moments can be used to leave clusters with $> 85\,\%$ e.m. energy unchanged
  - weighting works for energies above 10 GeV
- need to include dead material effects next
- clusterization effects probably on `Jet`-level?
- implementation in athena was in time for `11.0.0`