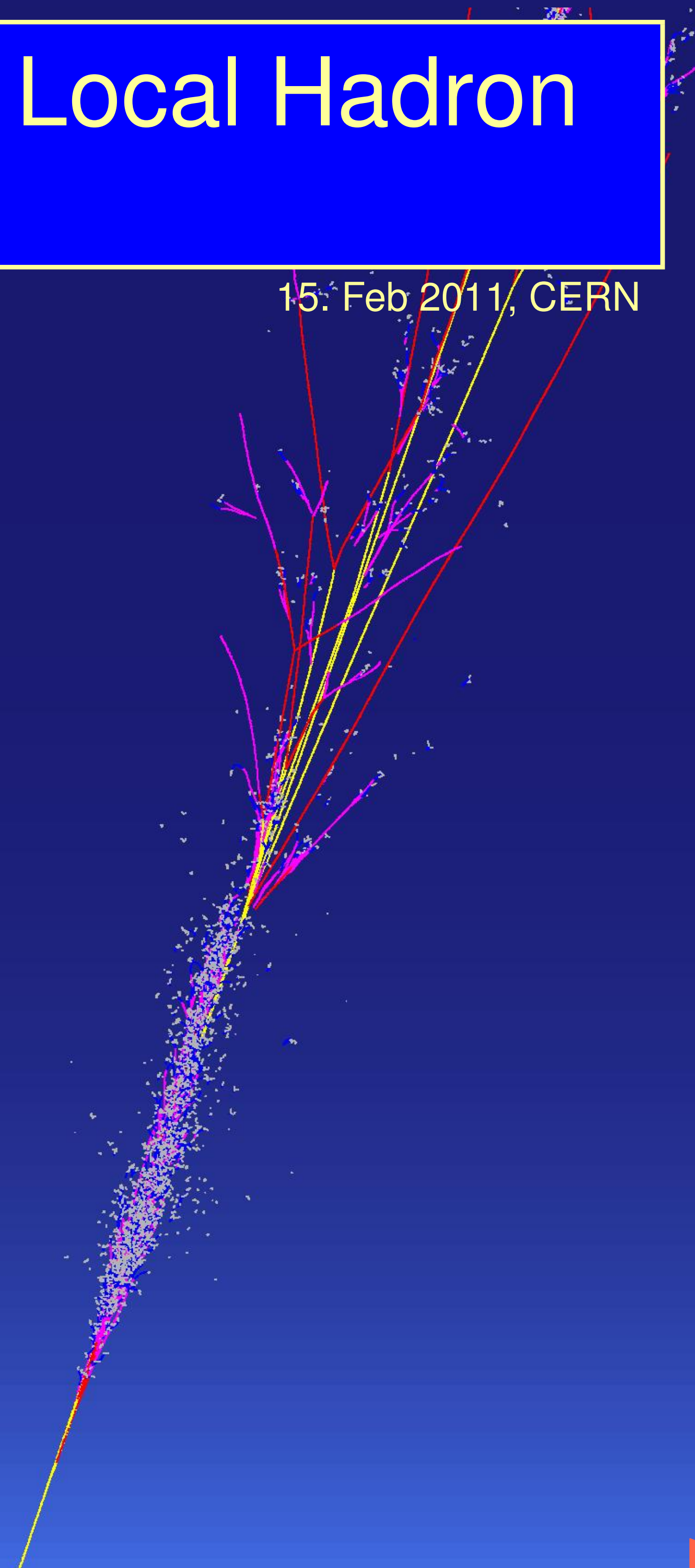
Topological Clustering and Local Hadron Calibration

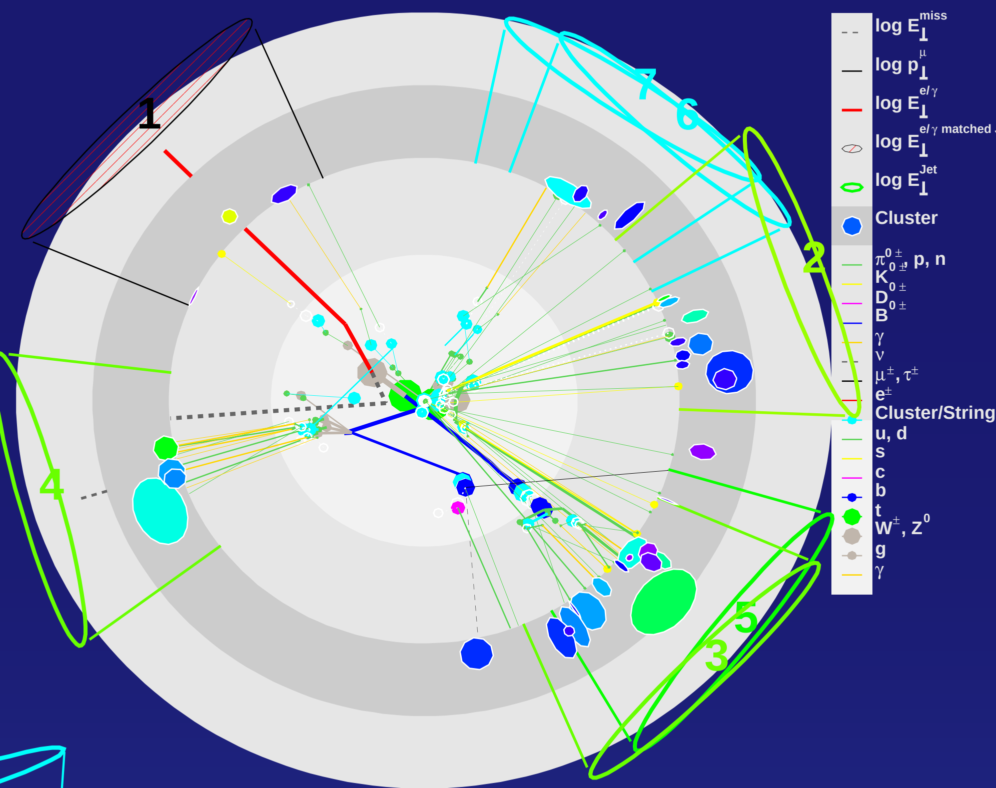
Particle Flow Meeting   Sven Menke, MPP München   15. Feb 2011, CERN

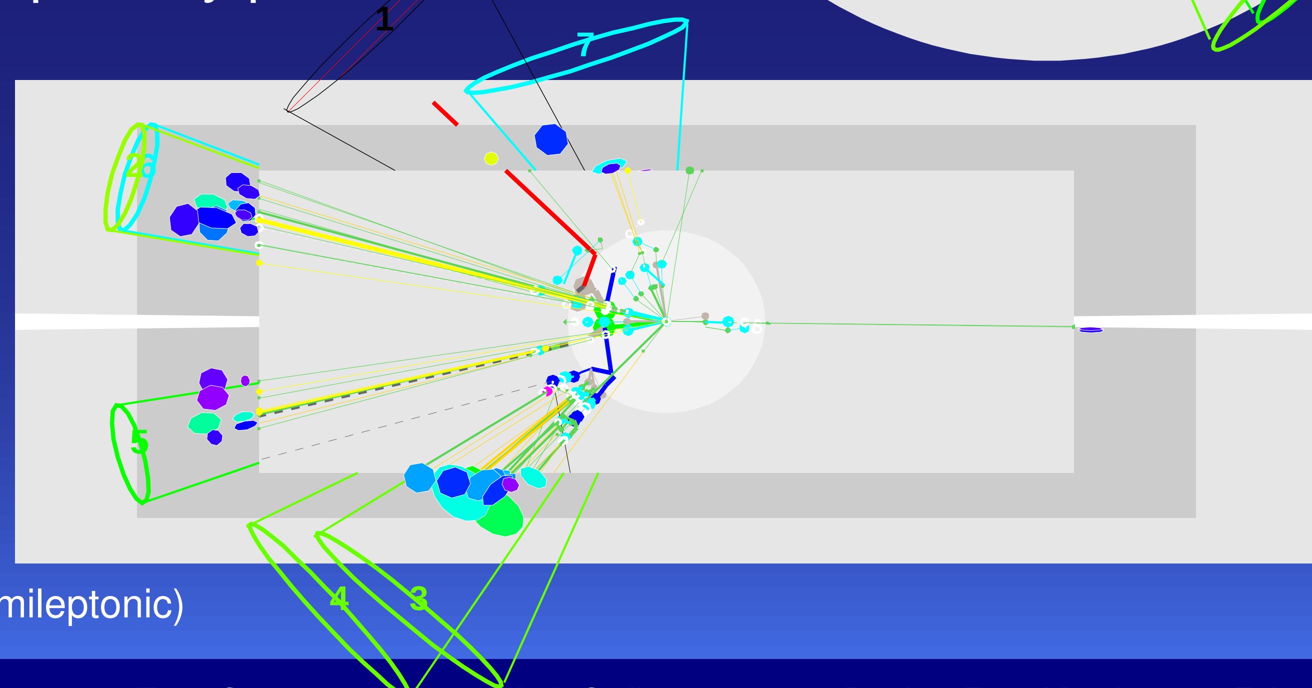


- Motivation
- Topological Clusters
  - cluster making
  - noise thresholds
  - cluster splitting
- Cluster Moments
- Local Hadron Calibration
  - classification
  - cell weighting
  - out-of-cluster corrections
  - dead material corrections
  - jet level corrections
- Topological Clusters as input to jets
- Conclusions

Motivation

- Aim is to have best possible response to hadrons and electrons in physics channels like $t\bar{t} \to Wb\,Wb \to l\nu j_b\, jjj_b$
  - pseudo event display in $r - \phi$ and $r - z$ illustrates this
  - use calorimeter objects calibrated to stable particle level to form jets which point back to primary partons

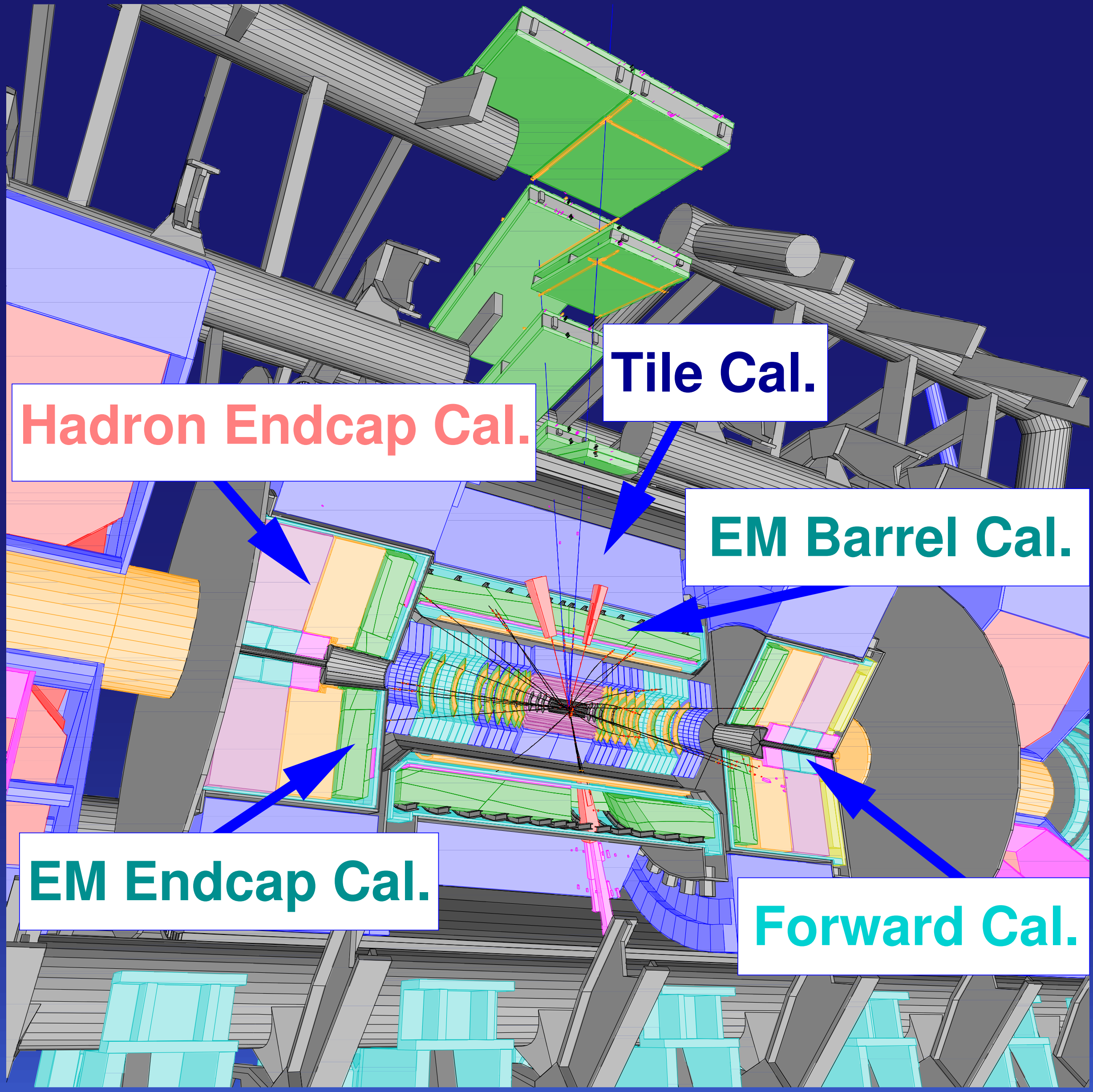




MC@NLO $t\bar{t}$ Event (semileptonic)

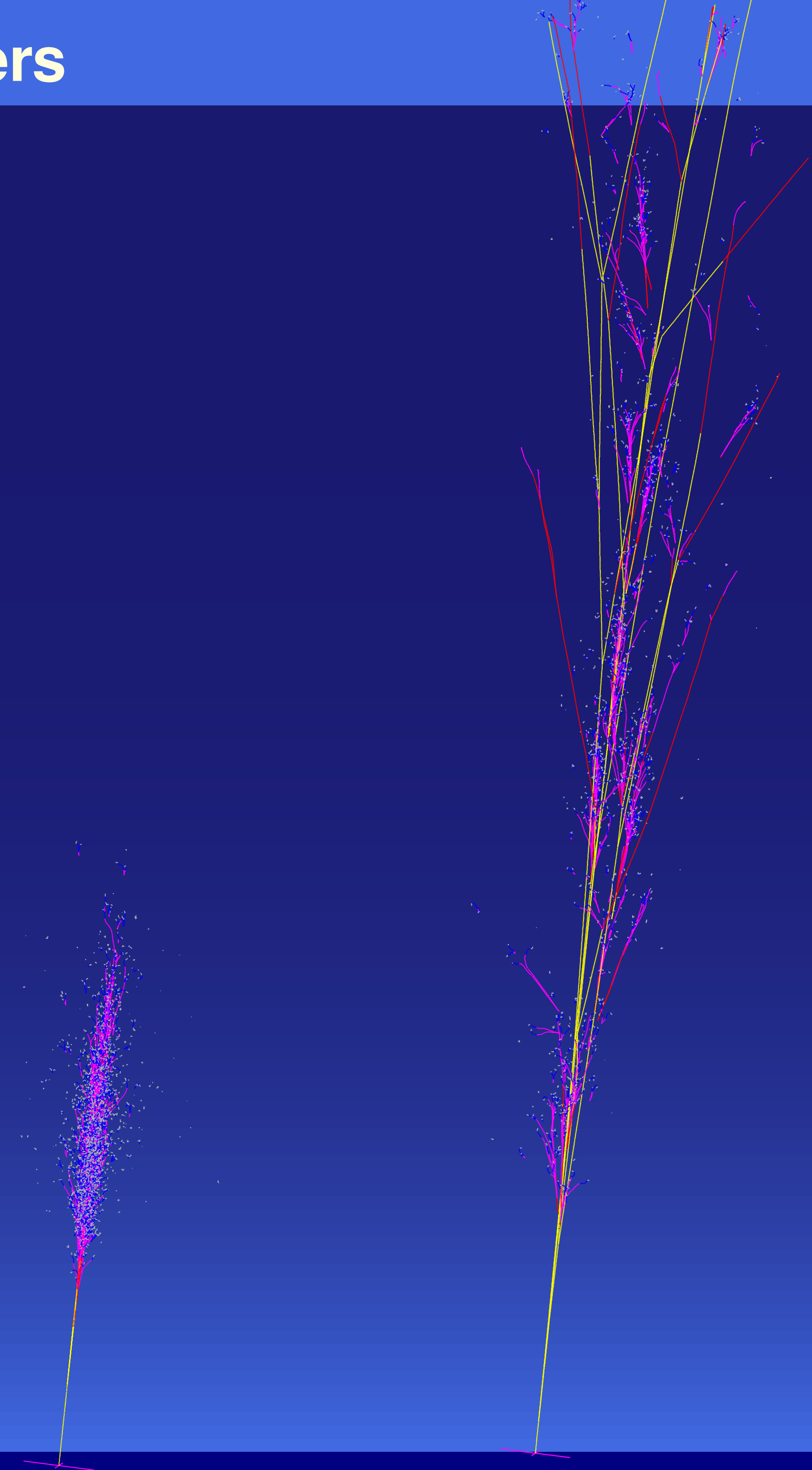
ATLAS Calorimeters

- Layout of the ATLAS Calorimeters
- EM LAr-Pb accordion calorimeter
  - Barrel (EMB): $|\eta| < 1.4$
  - End-cap (EMEC): $1.375 < |\eta| < 3.2$
- Hadron calorimeters
  - Barrel (Tile): Scint.-Steel $|\eta| < 1.7$
  - End-cap (HEC): LAr-Cu $1.5 < |\eta| < 3.2$
- Forward calorimeter (FCal) $3.2 < |\eta| < 4.9$
  - FCal1: LAr-Cu
  - FCal2&3: LAr-W



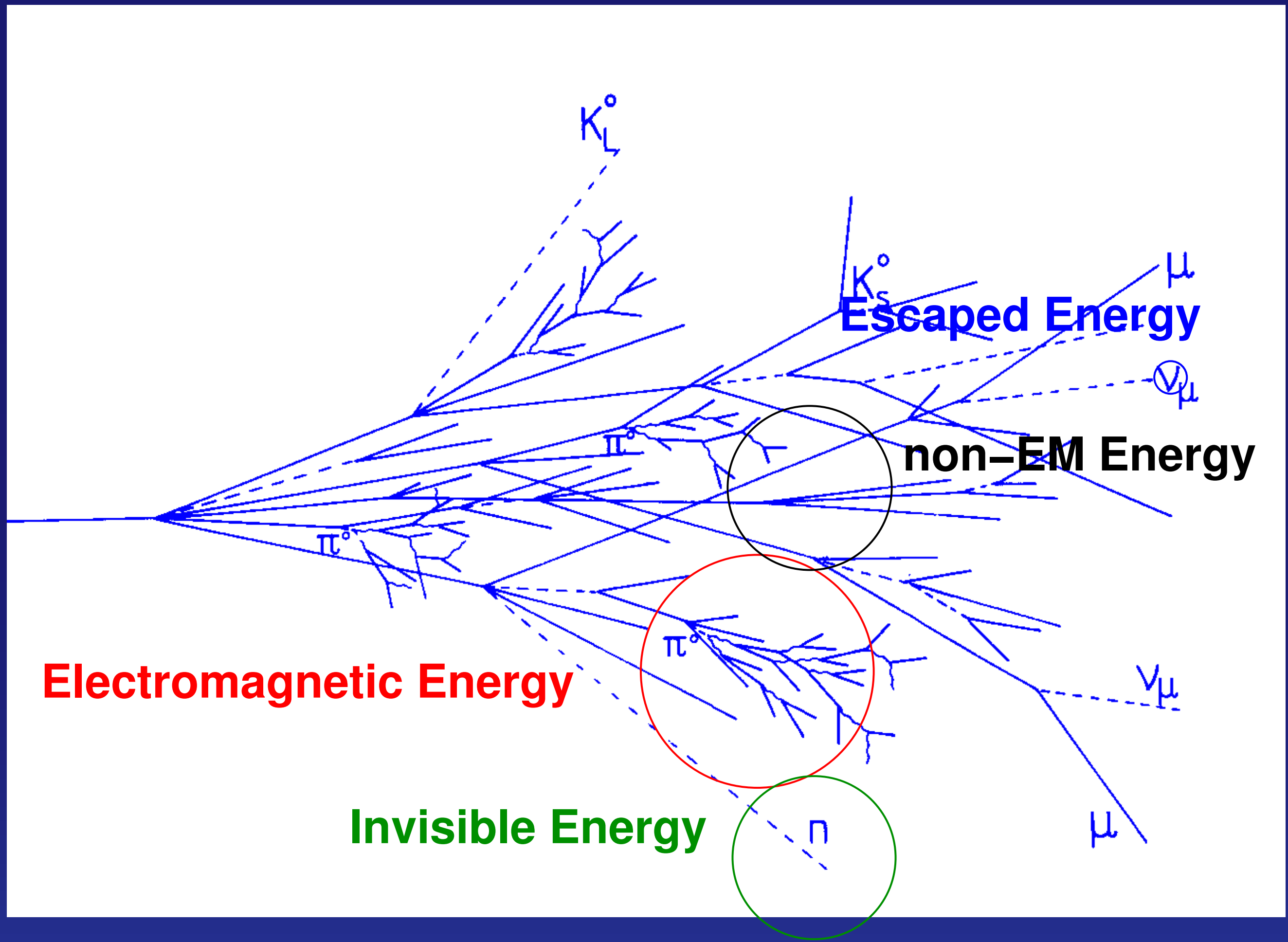
Electromagnetic vs. Hadronic Showers

- An electromagnetic shower
  - consists of visible EM energy only
  - is very compact ($X_0 \simeq 2\,\mathrm{cm}$)
  - can be simulated with high precision since mostly electromagnetic processes need to be calculated
  - allows high accuracy calibration (see talk by Stathes for details)
- A hadronic shower
  - consists of EM and hadronic energy (some invisible)
  - is very large ($\lambda_0 \simeq 20\,\mathrm{cm}$)
  - is difficult to simulate since it involves many QCD processes
  - limits the accuracy for calibration (mostly due to large fluctuations)
- The examples show $50\,\mathrm{GeV}$ showers of an electron (left) and a pion (right) in iron



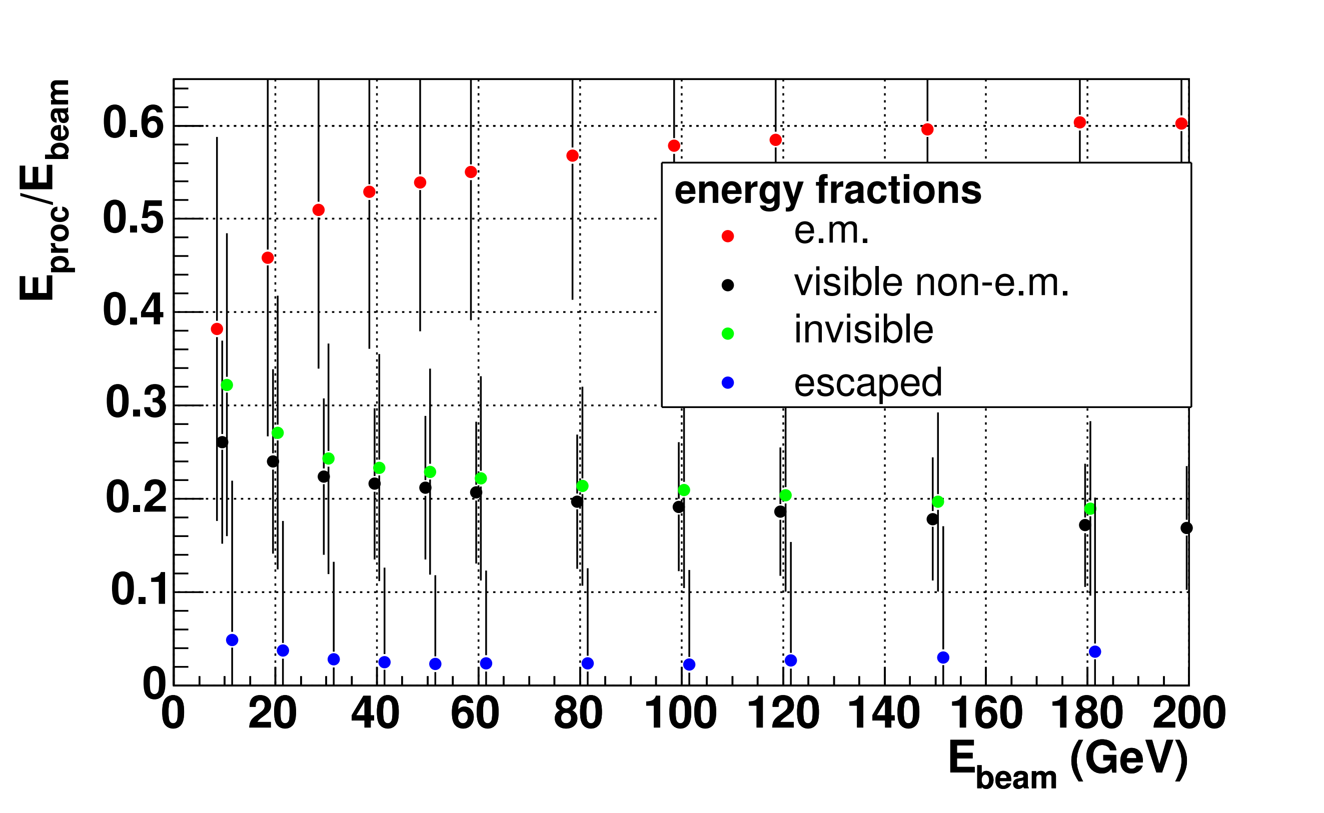
Hadron Calorimetry in ATLAS

- A hadronic shower consists of
  - EM energy (e.g. $\pi^0 \to \gamma\gamma$) $O(50\,\%)$
  - visible non-EM energy (e.g. $\mathrm{d}E/\mathrm{d}x$ from $\pi^{\pm}$,$\mu^{\pm}$, etc.) $O(25\,\%)$
  - invisible energy (e.g. breakup of nuclei and nuclear excitation) $O(25\,\%)$
  - escaped energy (e.g. ν) $O(2\,\%)$
- each fraction is energy dependent and subject to large fluctuations



- invisible energy is the main source of the non-compensating nature of hadron calorimeters
- hadronic calibration has to account for the invisible and escaped energy and deposits in dead material and ignored calorimeter parts

► From a Geant4 simulation of EMEC and HEC:



- EM energy strongly anti-correlated with visible non-EM energy
- visible non-EM energy strongly correlated with invisible energy

► need to separate EM part of the shower from the non-EM part

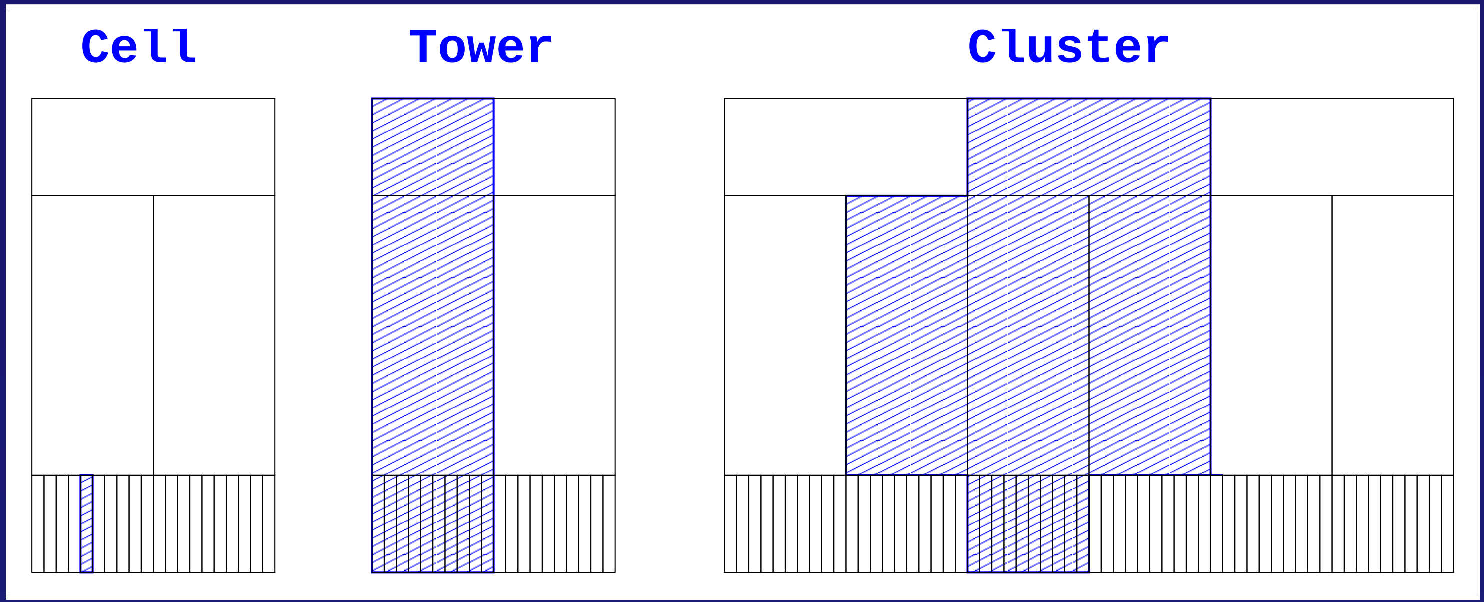
► apply a weight to the non-EM part to compensate invisible energy

► How to separate EM fraction from non-EM fraction?

- $X_0 \ll \lambda \simeq 20$ cm
- high energy density in a cell denotes high EM activity
- low energy density in a cell corresponds to hadronic activity
- apply weights as function of energy density

Calorimeter Reconstruction

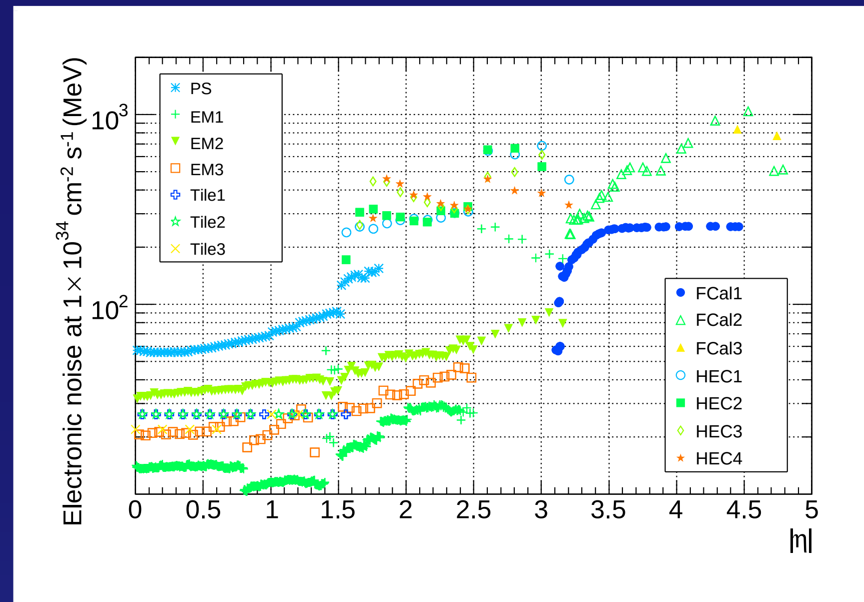
- The cell is the smallest reco object
  - all ATLAS calorimeters together provide 187652 cells
  - each cell provides mainly the raw reconstructed energy in MeV



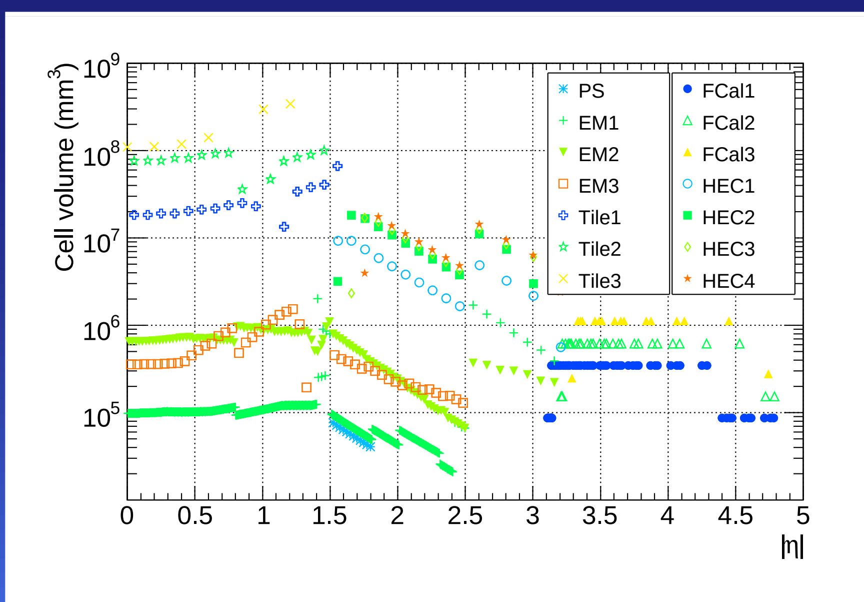
- A tower is a group of cells (or even a group of fractions of cells) in a fixed $\Delta\eta \times \Delta\phi$ grid over some or all samplings
  - contains the sum of cell (fraction) energies and the center of the grid square (η and ϕ) as members
  - in use in ATLAS are 65536 LAr EM only `LArTower`s with $\Delta\eta \times \Delta\phi = 0.025 \times 2\pi/256$
  - and 6400 `CaloTower`s including all calorimeters with with $\Delta\eta \times \Delta\phi = 0.1 \times 2\pi/64$
- A cluster is a group of cells (or even fraction of cells) formed around a seed cell
  - is the main reco object for calorimetry
  - with either a fixed size in $\Delta\eta \times \Delta\phi$ (sliding window used for electrons/photons)
  - or variable borders based on the significance of the cells (topo cluster used for hadrons/jets/MET)
  - contains lots of data members based on weighted cell members for energy, position and shape

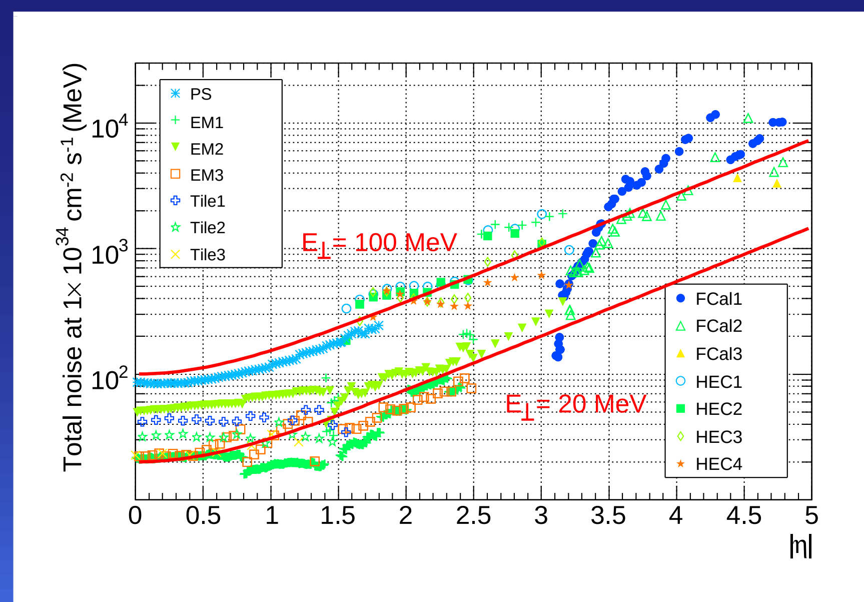
Clusters

- Cluster algorithms need to serve multiple purposes
  - suppress noise (electronics noise and pile-up)
  - keep electromagnetic showers in one cluster
  - separate multiple signals which are close by
  - work on very different sub-systems



- Plots on the right and below show large variations in η for
  - electronics noise at high luminosity ($\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-2}$) ($\sim 10 - 10^3$ MeV)
  - total noise at high luminosity ($\sim 2 - 10^4$ MeV)
  - cell volume ($\sim 2 \cdot 10^4 - 3 \cdot 10^8$, mm^3)





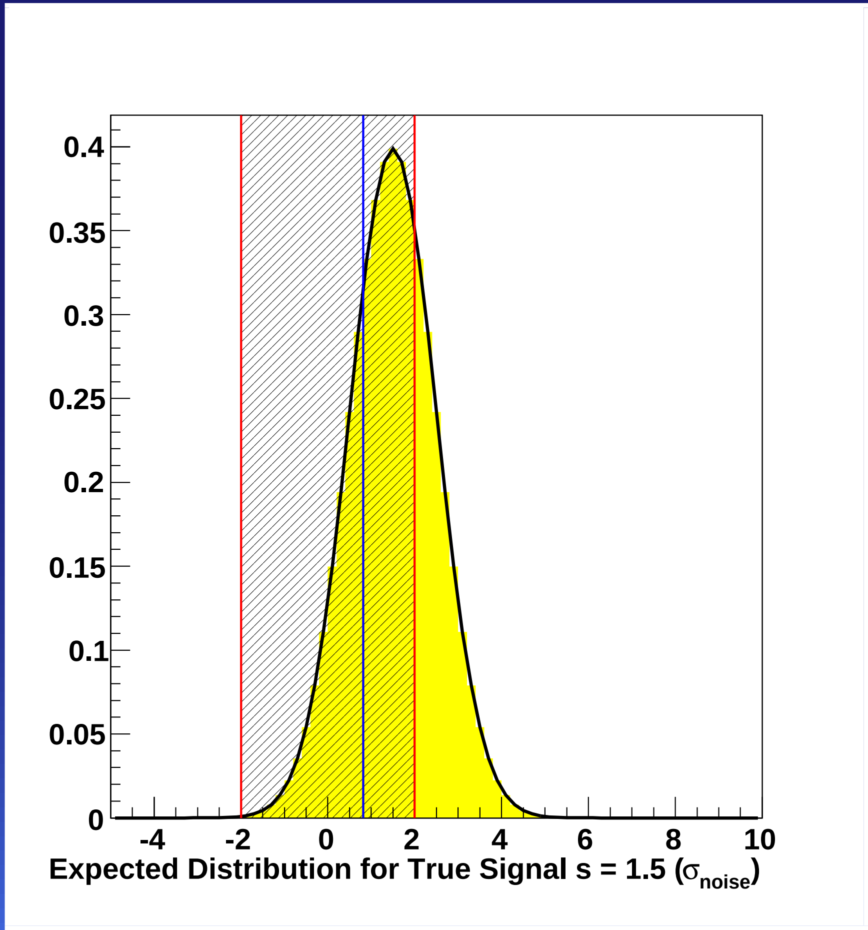
Topological Clusters

- Cluster Making
  - form clusters around seed cells with $|E_{\text{seed}}| > 4(\sigma_{\text{elec-noise}} \oplus \sigma_{\text{pile-up-noise}})$
  - expand clusters around neighbor cells with $|E_{\text{neigh}}| > 2\sigma$
  - include perimeter cells with $|E_{\text{cell}}| > 0\sigma$
  - merge clusters if they share a neighbor cell
  - expansion is driven by neighbors in $3D$: usually 8 neighbors in the same layer ($2D$) plus cells overlapping in η and ϕ with central cell in next and previous layer (just 2 if granularity would be the same)
- Cluster Splitting
  - search for local maxima in cell energy with $E_{seed} > 500\,\text{MeV}$ in all clustered cells in EM-samplings (HAD-samplings secondary)
  - re-cluster around local maxima with same neighbor driven algorithm but no thresholds and no merging
  - cells at cluster borders are shared with energy and distance dependent weights

Noise Bias ► $|E_{\text{cell}}| > 2\,\sigma_{\text{noise}}$

- Default method for $\not{E}_{\perp}$ group was global symmetric cell level cut $|E_{\text{cell}}| > 2\,\sigma_{\text{noise}}$
  - this is o.k. for no expected signal (no bias, reasonable resolution)
  - also o.k. for large signals since they will be accepted (including their noise)
  - a bias $O(-0.6\,\sigma_{\text{noise}})$ is introduced for small signals and tails of large signals (i.e. $E_{\text{cell}} = O(\sigma_{\text{noise}})$) which makes the bias signal dependent
- The plot on the right illustrates this bias.
- Shown is the expected distribution of a small signal ($1.5\,\sigma_{\text{noise}}$) in the presence of noise
- The shaded area shows the region where the measured value is replaced by 0
- The blue line shows the average reconstructed value

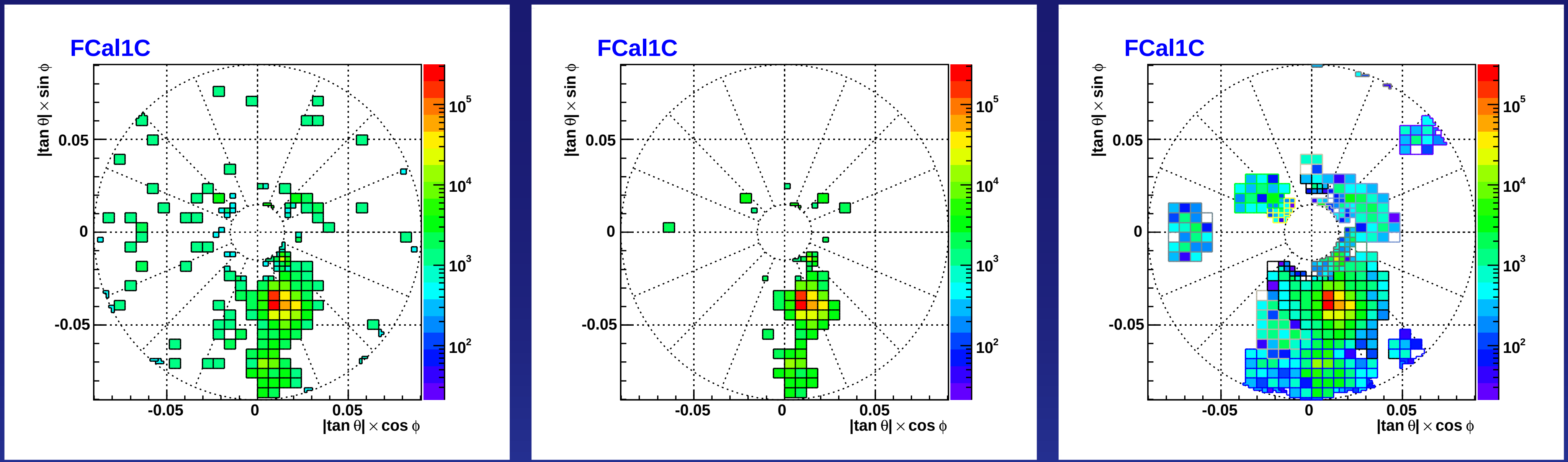
True value (σ_{noise})	Bias (σ_{noise})
0.0	0.00
1.0	-0.60
1.5	-0.69
2.0	-0.60
3.0	-0.23
4.0	-0.04



Topological Cluster Example

- look at di-jet MC sample including electronics noise with activity in the forward region
- plots show $|E_{\text{cell}}|$ on a color coded log-scale in MeV in the first (EM) FCal sampling for one event

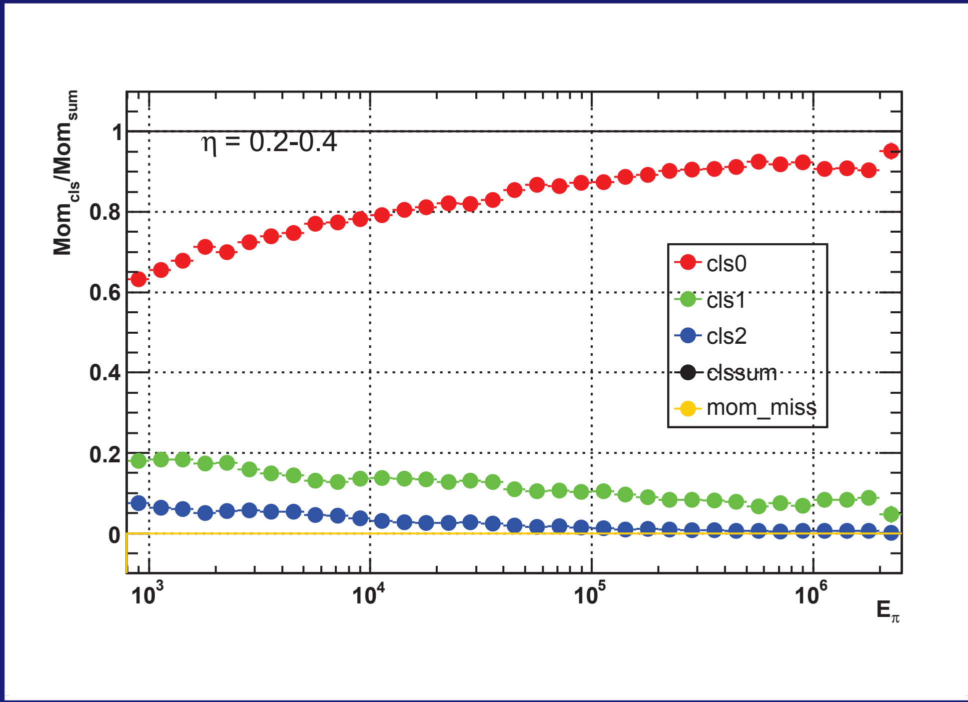
$|E| > 2\,\sigma_{\text{noise}}$ $\quad$ $|E| > 4\,\sigma_{\text{noise}}$ $\quad$ $4/2/0$ topological clusters

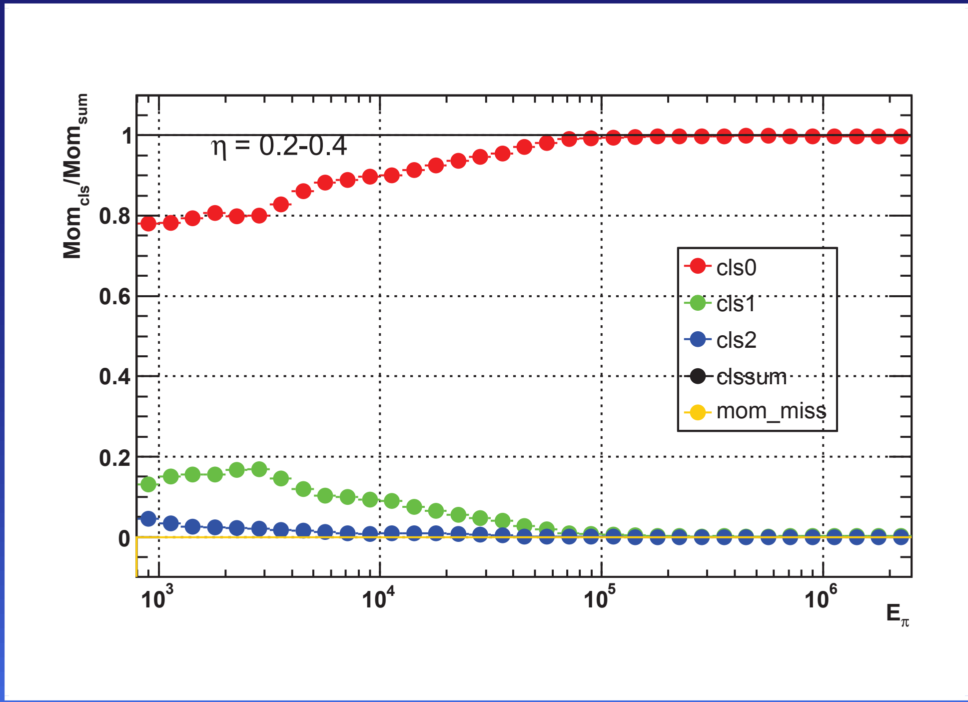


- $2\,\sigma$ cut is removing cells from the signal region
- $4\,\sigma$ cut shows seeds for the cluster maker
- after clustering all cells in the signal regions are kept
- cluster splitter finds hot spots

Topological Clusters

- Number of relevant clusters per particle
  - there can be more than 1 cluster in a cone around the original pion direction
  - but the number of relevant clusters (fraction of energy) is small
  - top plot shows calibration hit energy fraction in the 3 leading clusters for charged pions vs. the pion energy in the barrel
  - bottom plot shows the same for neutral pions
  - for $E > 2\,\mathrm{GeV}$ more than 90% or the energy are in the 2 leading clusters (charged pions)
  - for neutral pions significant energy only in leading 2 clusters and only in 1 if photons can not be resolved





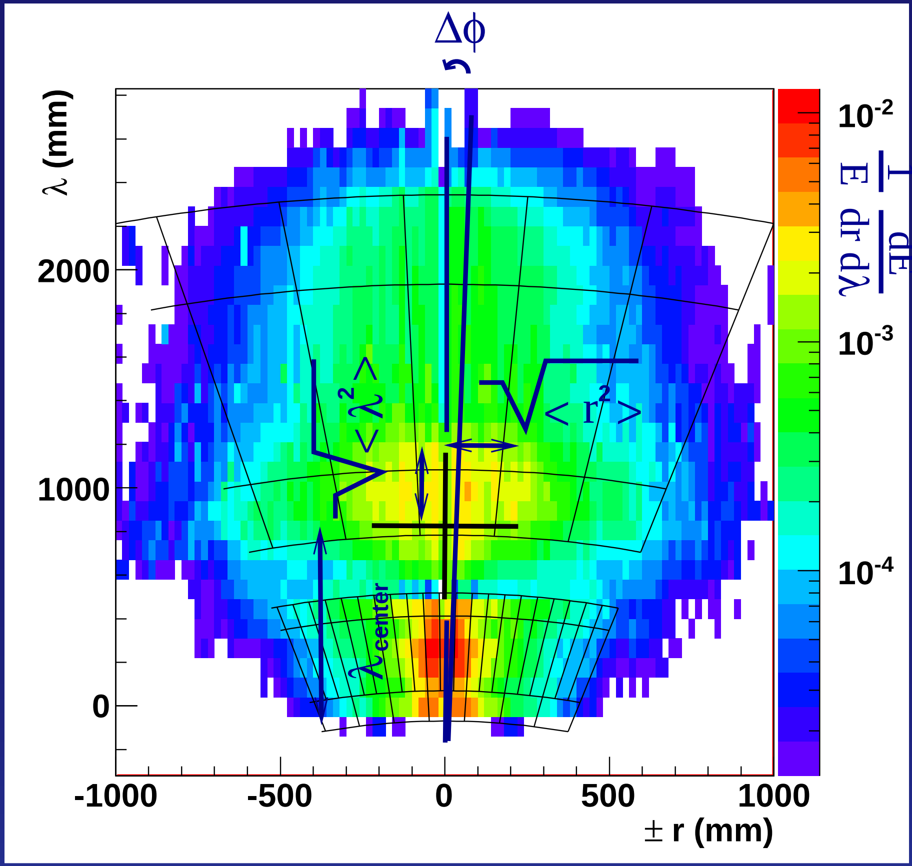
Cluster Moments

- shape variables calculated from the positive cells in a cluster
- first a principal value analysis is run on the cluster cells
  - provides centroid 3 major axes of the shower
- angles of the major axis w.r.t. IP-shower-center direction are calculated
- other shape quantities defined by moments of the form

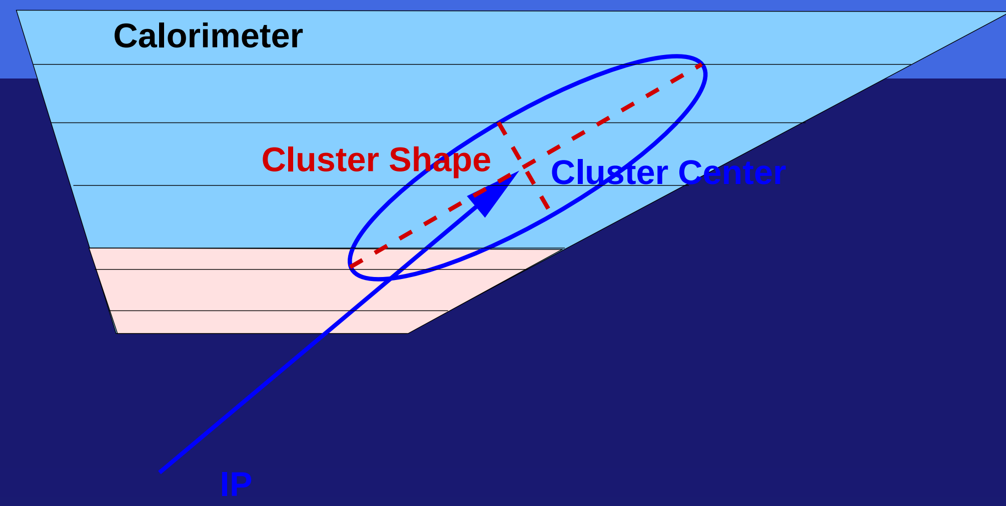
$$\langle x^n \rangle = \frac{1}{E_{\mathrm{norm}}} \times \sum_{\{i|E_i>0\}} E_i\, x_i^n, \text{ with}$$

$$E_{\mathrm{norm}} = \sum_{\{i|E_i>0\}} E_i.$$

- typical choices for x: $\rho = E/V$, r, λ



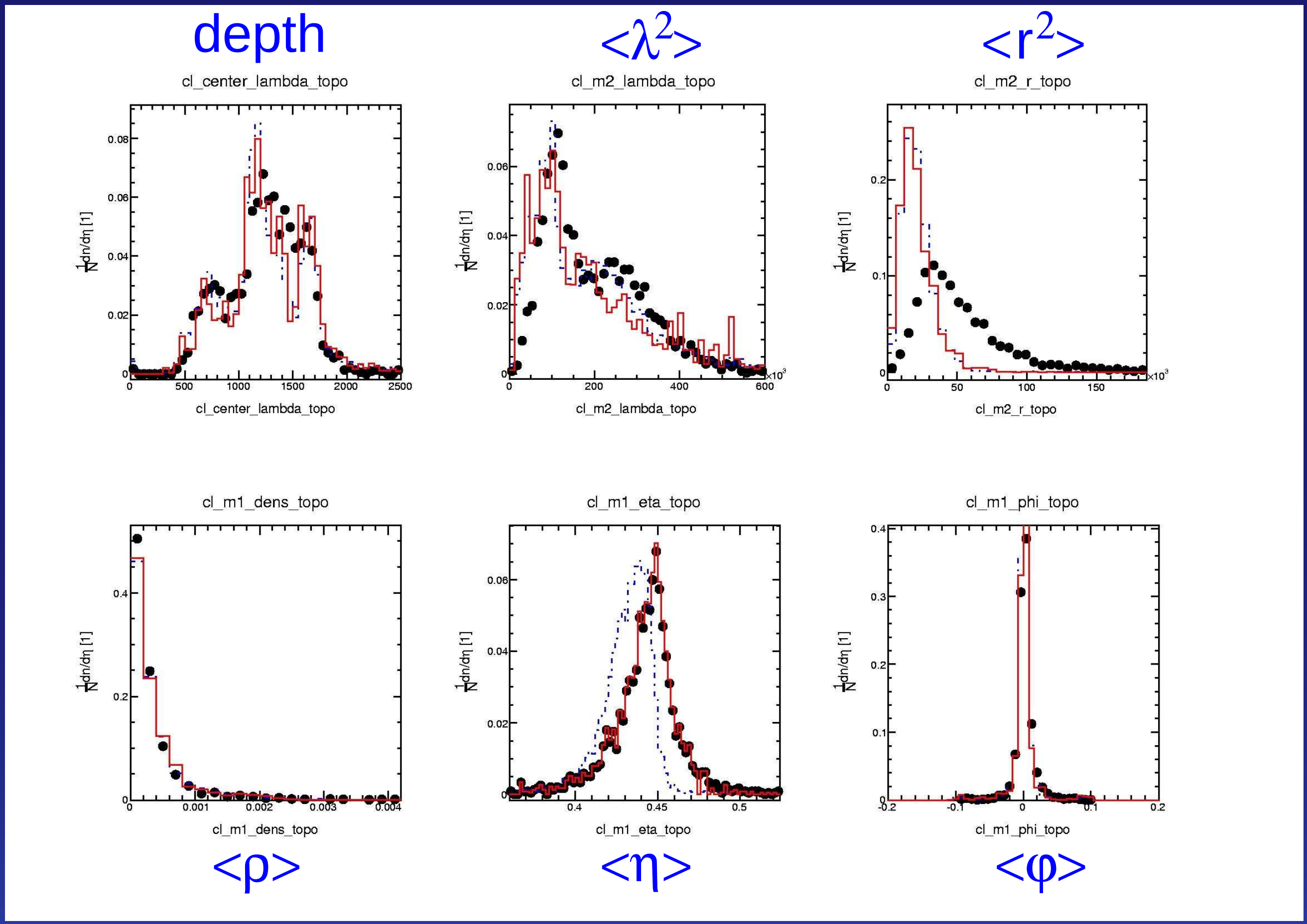
Cluster Moments



- 11 most popular moments are on AOD
  - LATERAL normalized second lateral moment
  - LONGITUDINAL normalized second longitudinal moment
  - SECOND_R the width squared of the cluster
  - SECOND_LAMBDA the length squared of the
  - CENTER_LAMBDA the cluster center depth in the calorimeter
  - CENTER_MAG the distance IP - cluster center
  - FIRST_ENG_DENS the first moment of $\rho = E/V$
  - ENG_FRAC_MAX the ratio of the hottest cell energy over the cluster energy
  - ISOLATION fraction of cells neighbouring the perimeter cells of the cluster which are not included in other clusters
  - ENG_BAD_CELLS energy stored in cells flagged as bad
  - N_BAD_CELLS number of cells flagged as bad

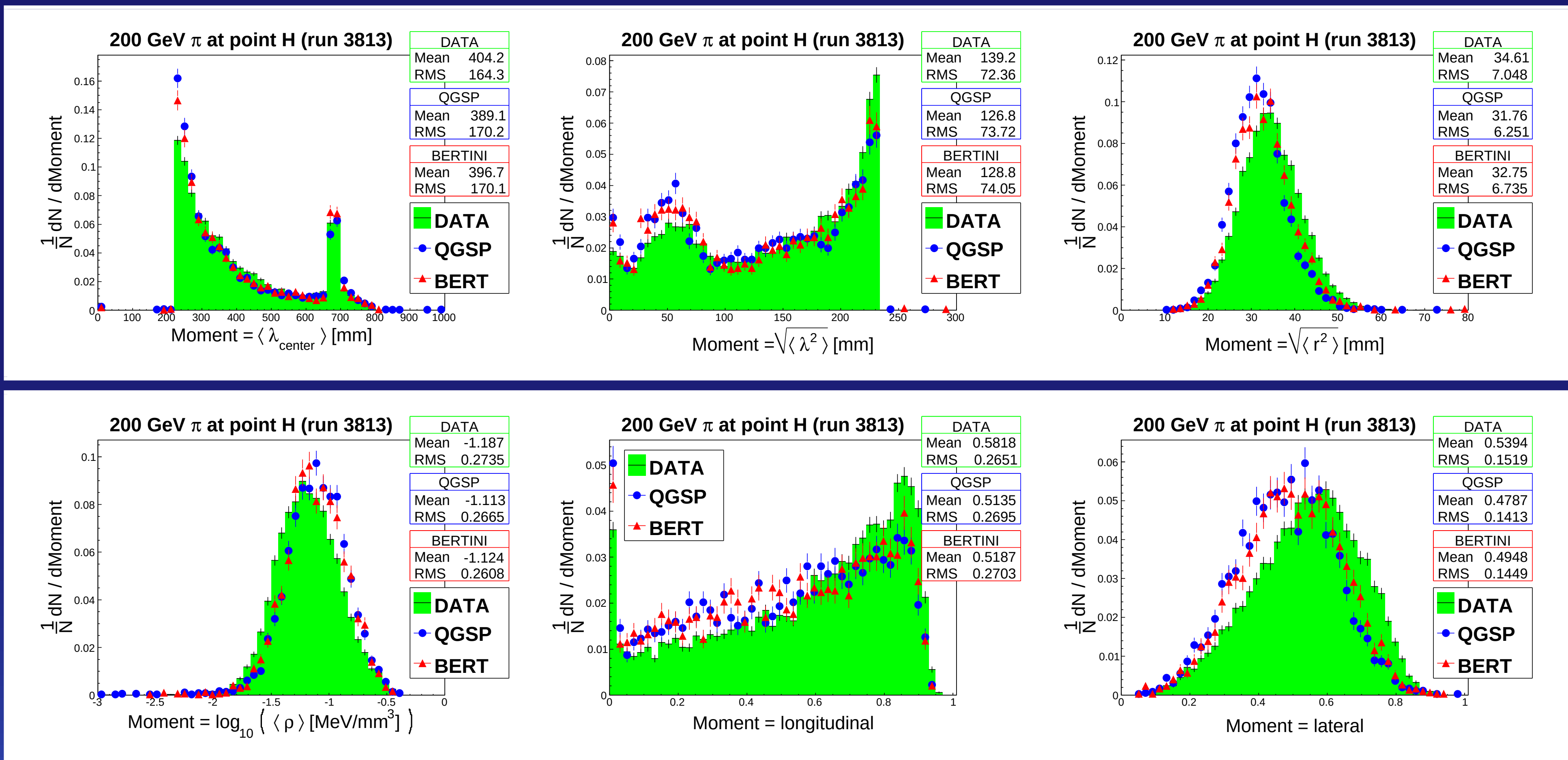
- other important moments available on ESD are
  - CENTER_X/Y/Z the position of the cluster
  - ENG_FRAC_EM the fraction of cluster energy in EM samplings
  - ENG_FRAC_CORE the fraction of cluster energy in the leading cells in every sampling
  - DELTA_PHI/THETA/ALPHA angular deviations of the shower axis from IP-cluster-center axis
  - ENG_CALIB_* 17 of the 22 new moments of calibration hit energies associated to the cluster (in simulations with calibration hits; these are also on AOD)

- look at cluster moments for 20 GeV pions from 2004 barrel test beam data (black points) and compare to G4 simulation (dashed blue lines)
- differences in $\langle\eta\rangle$ might be due to simplified beam trajectories in simulations
  - compare also with η-reweighted distributions (red)
- shower depth and energy density in good agreement
- $\langle r^2\rangle$ shows no agreement at all



- very important to use only moments which are well described
- validation of default athena algorithms with test beam data is crucial

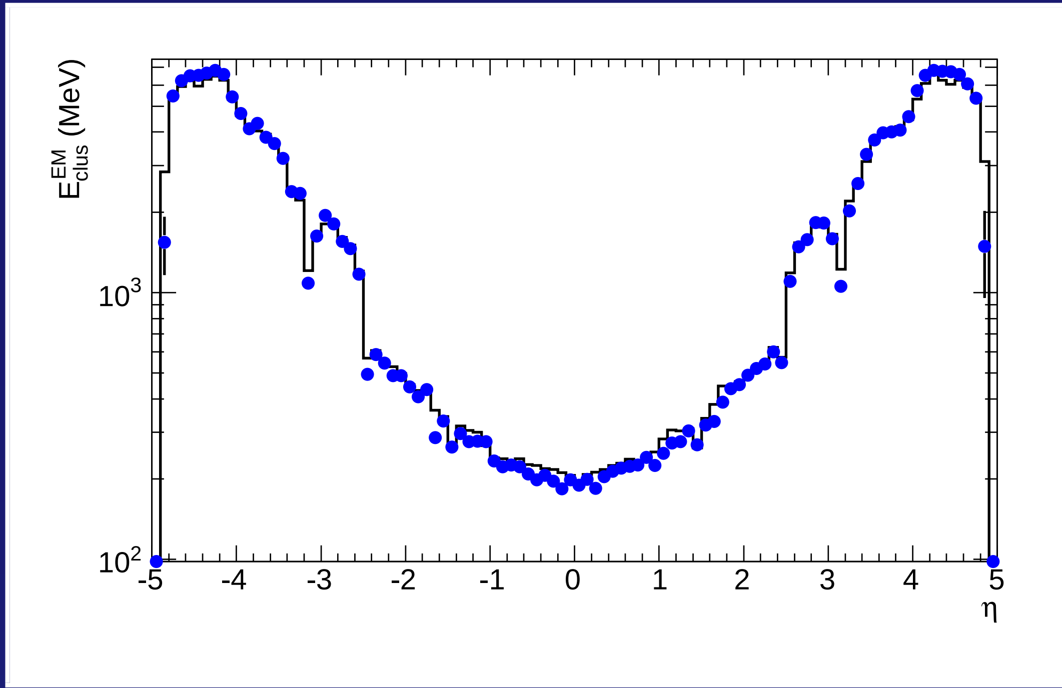
- 200 GeV pions from 2004 endcap test beam data in the FCal region (solid green histogrms) and G4 QGSP (blue) and QGSP_BERT (red) simulations)

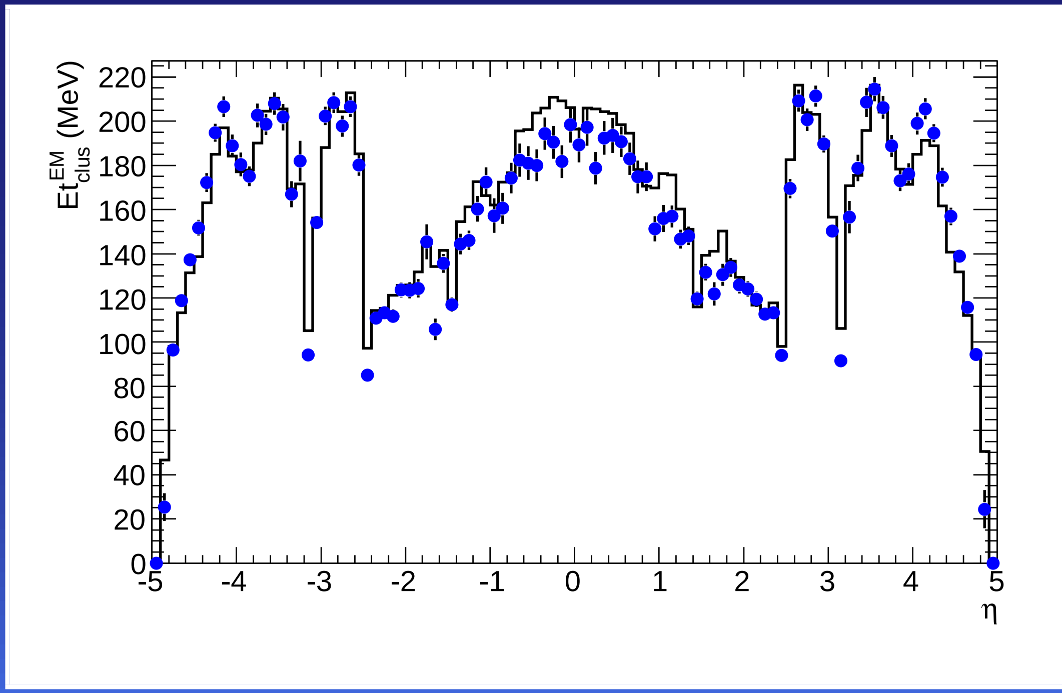


- best description for λ_{center} and $\langle\rho\rangle$
- largest deviations in LATERAL and LONGITUDINAL
- QGSP_BERT slightly better than QGSP

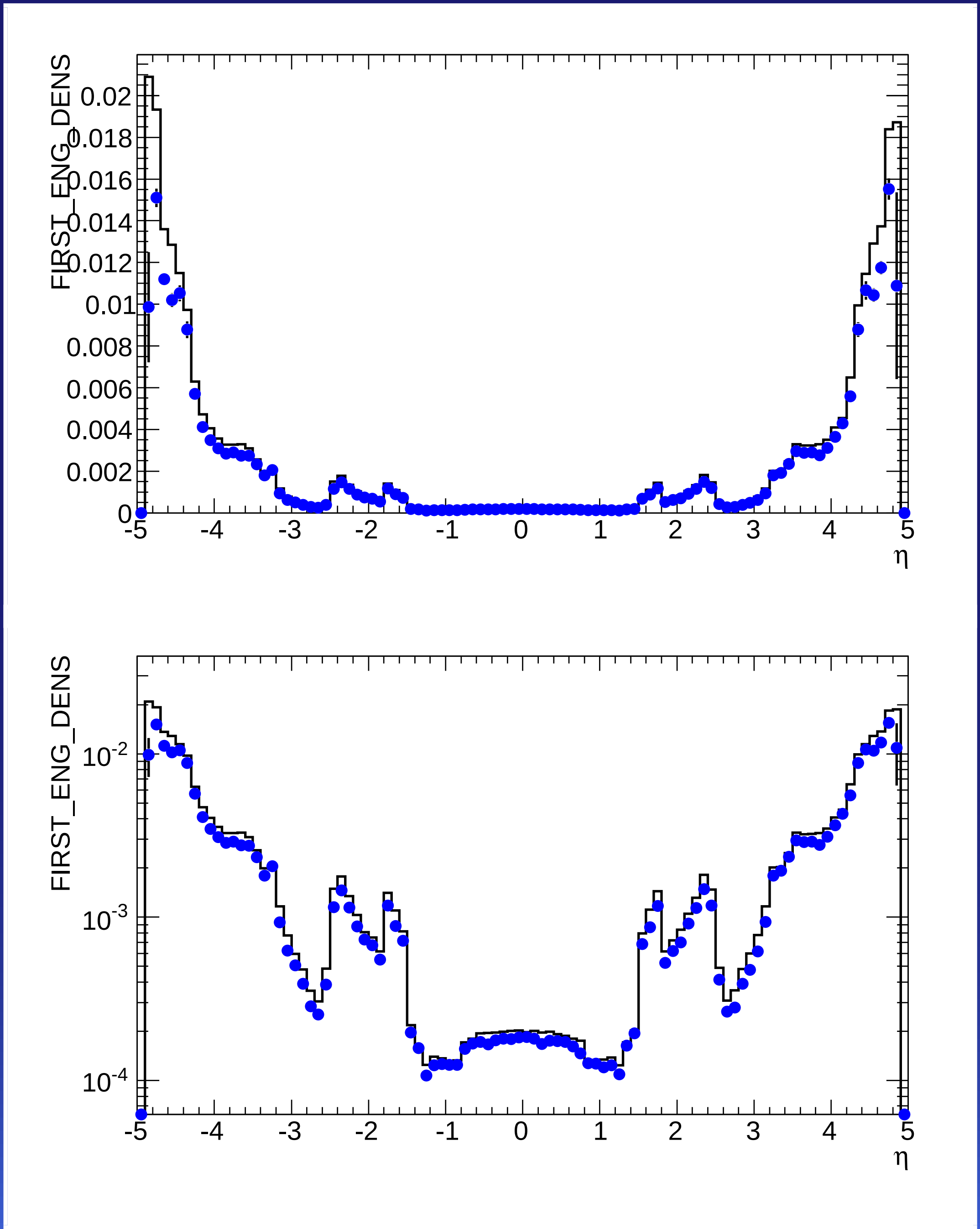
Cluster Moments ► $\sqrt{s} = 900$ GeV data

- Averages of currently used moments in local hadronic calibration from MinBias data and MC
  - `FIRST_ENG_DENS`
  - `CENTER_LAMBDA`
  - `ISOLATION`
- Averages of moments we'd like to use in local hadronic calibration from MinBias data @ 900 GeV and MC
  - `SECOND_LAMBDA`
  - `LONGITUDINAL`
  - `SECOND_R`
  - `LATERAL`
- Caveat: average cluster energies are small – see plots on the right for E (top) and $E_\perp$ (bottom)
- We need moment comparisons for higher energetic clusters
- In the following: plots on top (bottom) lin (log) scale



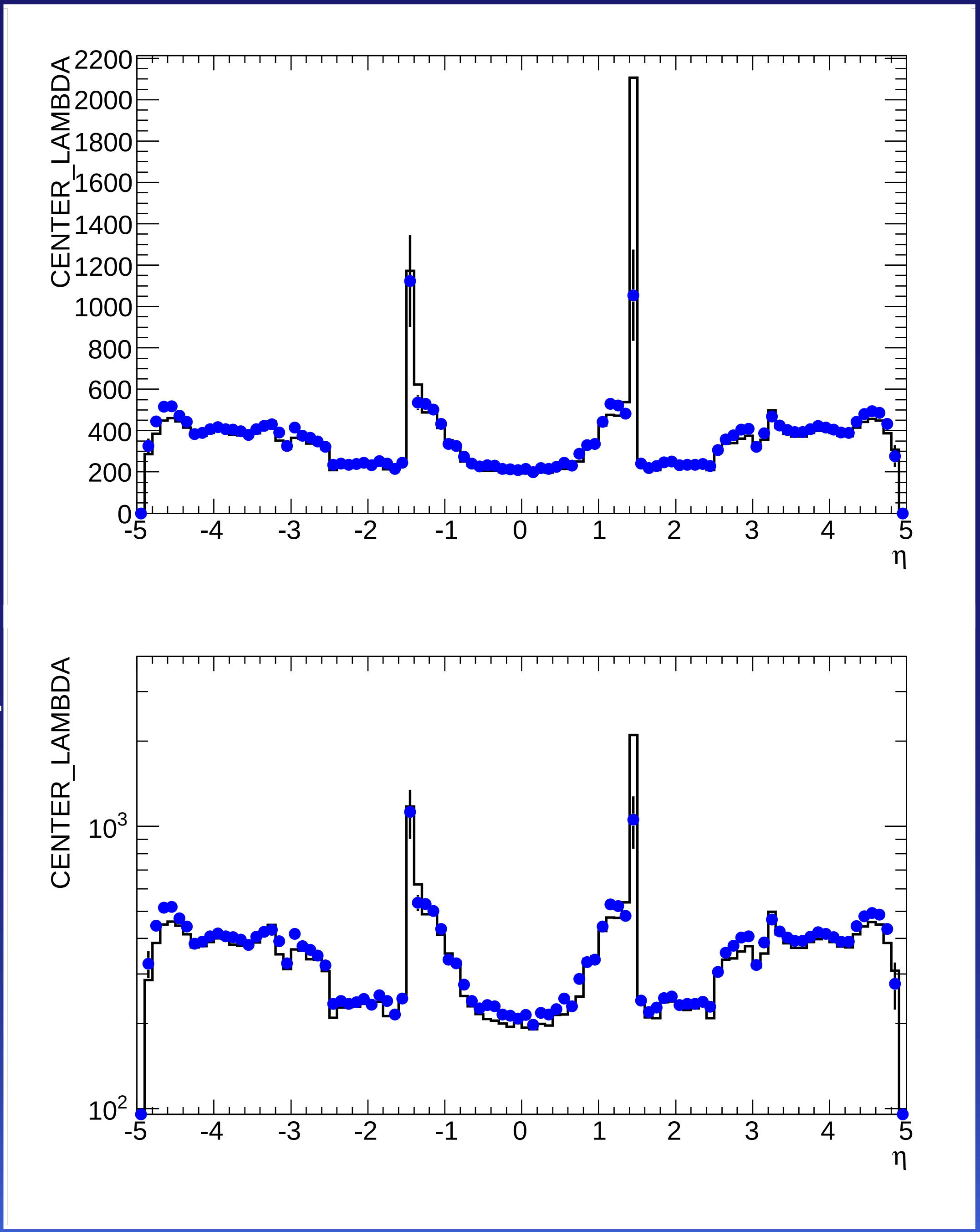


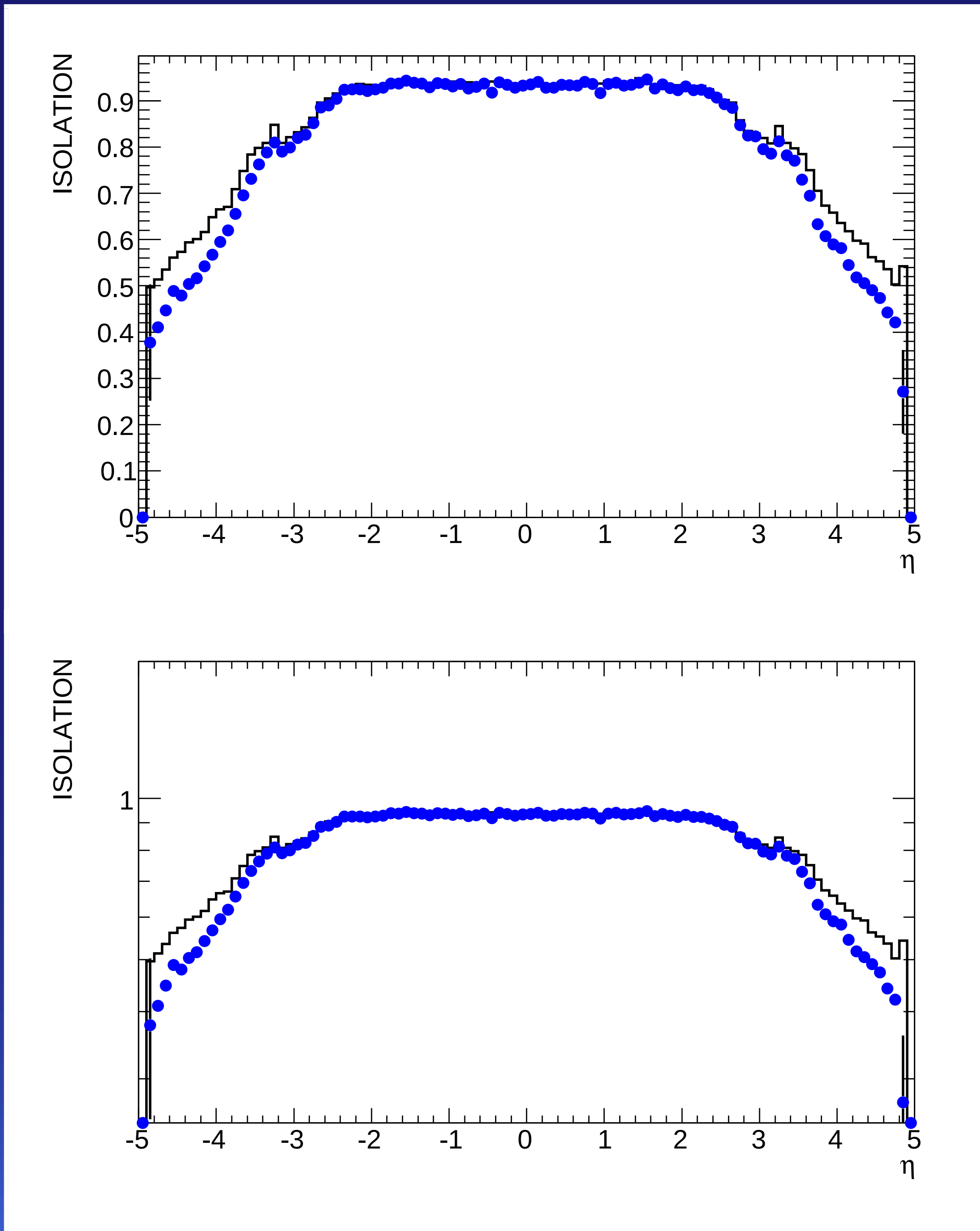
- this moment is used in the classification
- data slightly less dense than MC
- more so in most inner FCal region
- could indicate some very soft physics missing in MC
- Andrey sees larger discrepancy in TB



Cluster Moments ► CENTER_LAMBDA

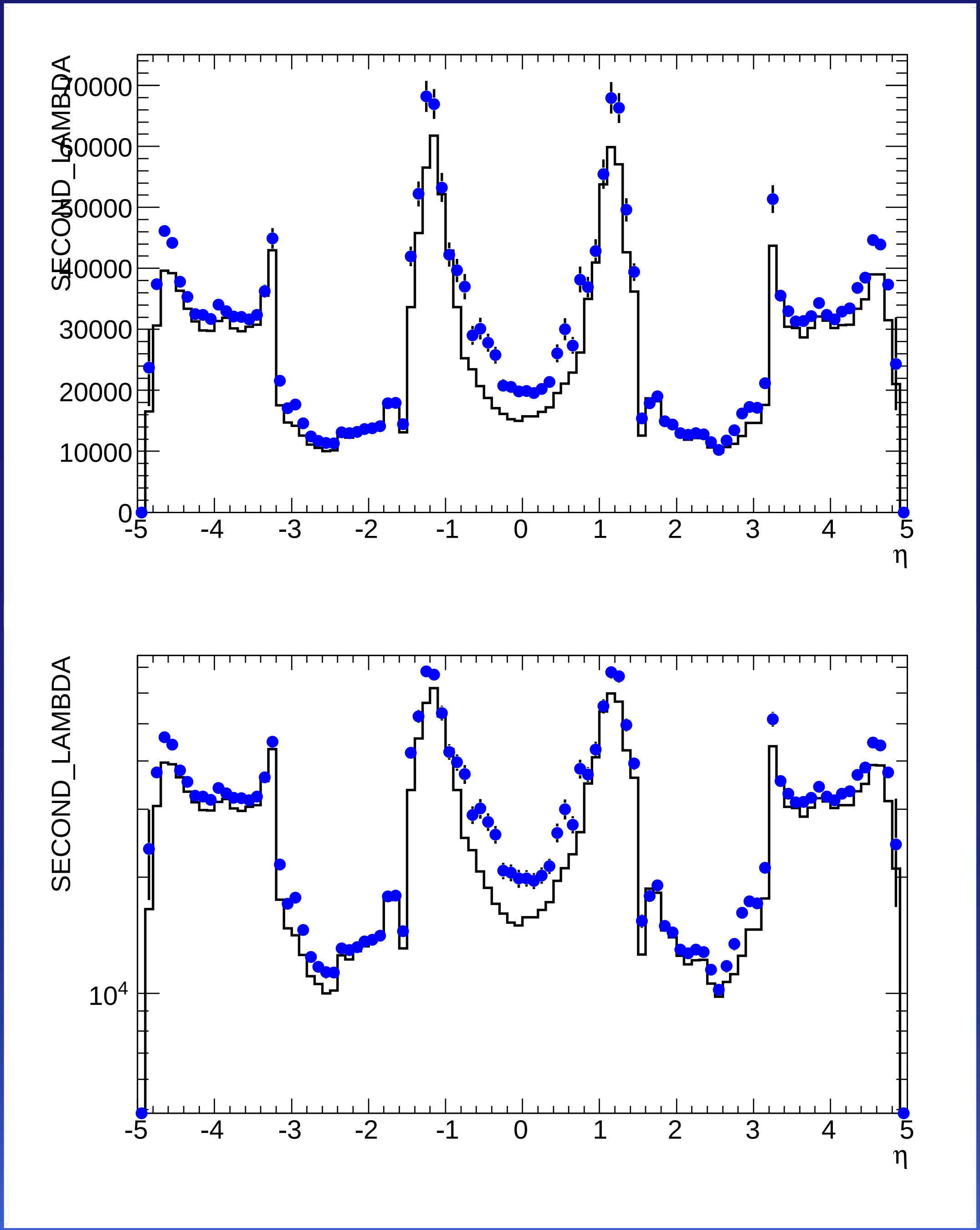
- this moment is used in the classification and out-of-cluster corrections
- in general very good agreement between data and MC
- showers are a little bit deeper in the calo in data
- no specific η-region
- could indicate too compact phsyics list in MC





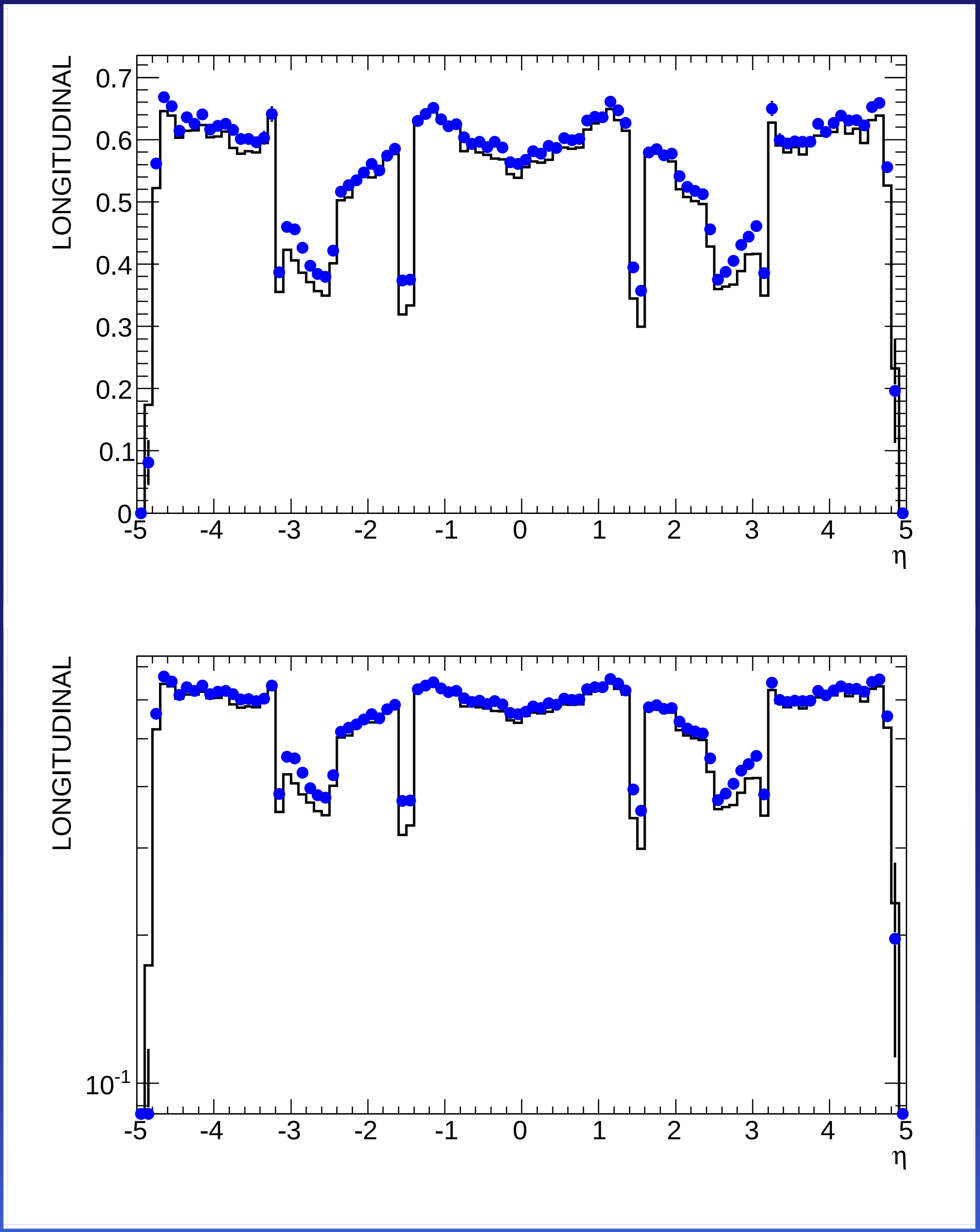
- this moment is used in the out-of-cluster corrections
- in general very good agreement between data and MC
- except for forward region
- data is less isolated than MC
- could be missing soft QCD processes in MC

Cluster Moments ► `SECOND_LAMBDA`

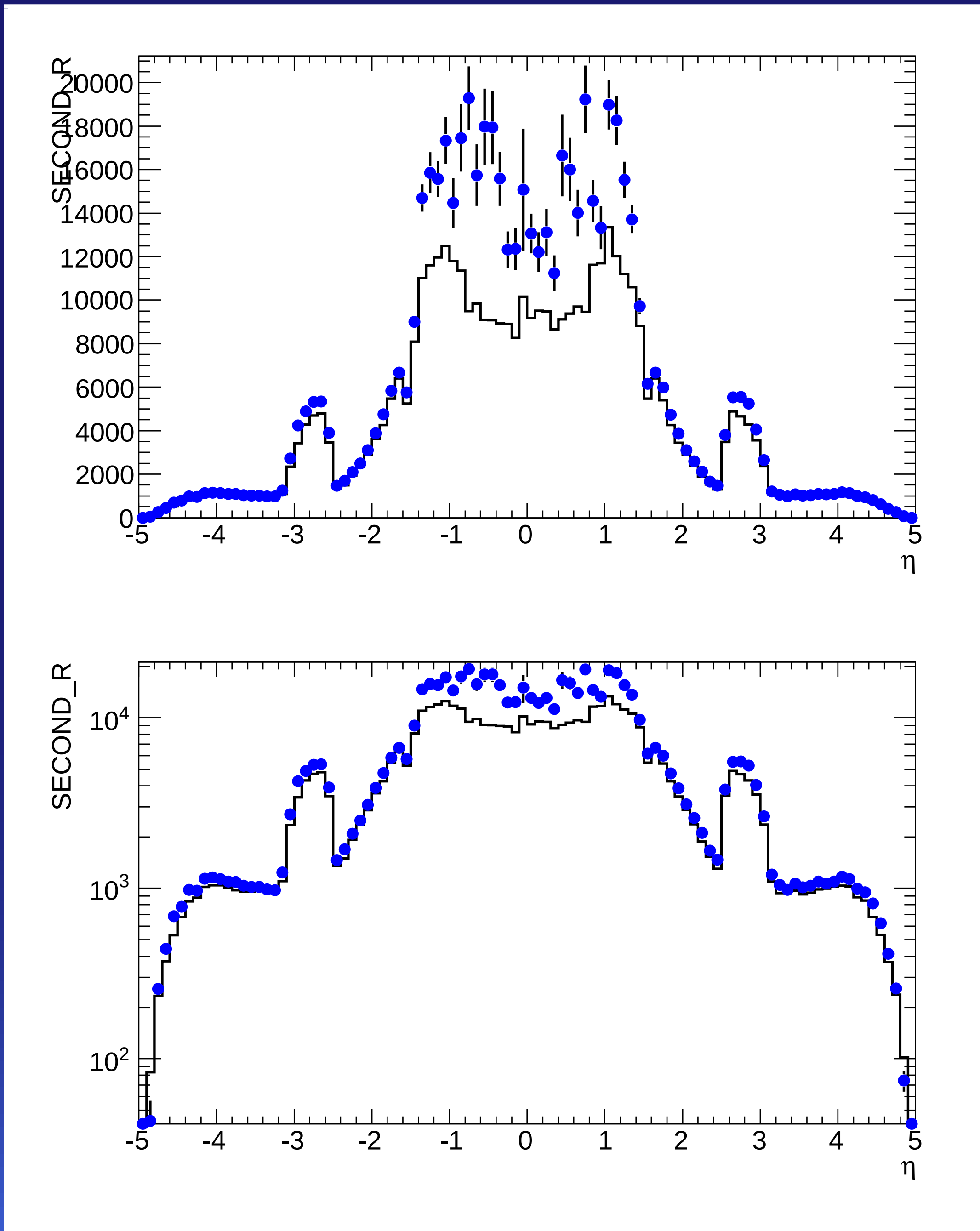


- in general very good agreement between data and MC
- except central barrel region where data has significantly longer showers
- largest effect for $|\eta| < 1$
- could be noise description/treatment in `Tile`

- potentially good moment to separate em/had
- in general very good agreement between data and MC
- bigger deviations in crack regions
- data shows flatter showers than MC



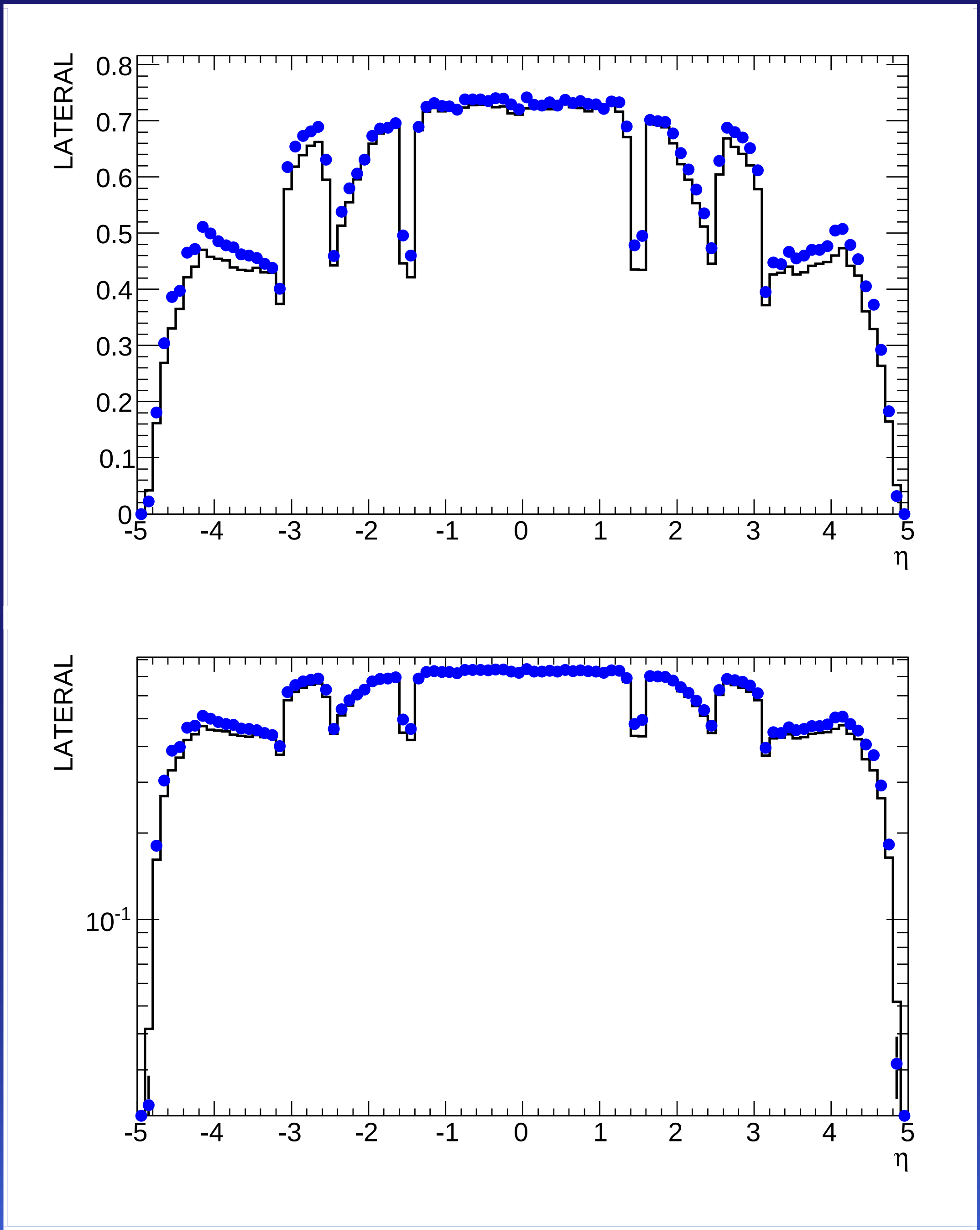
Cluster Moments ► SECOND_R



- surprisingly good agreement between data and MC
- except in central barrel
- overshooting data for $|\eta| < 1.5$ hints again at `Tile`

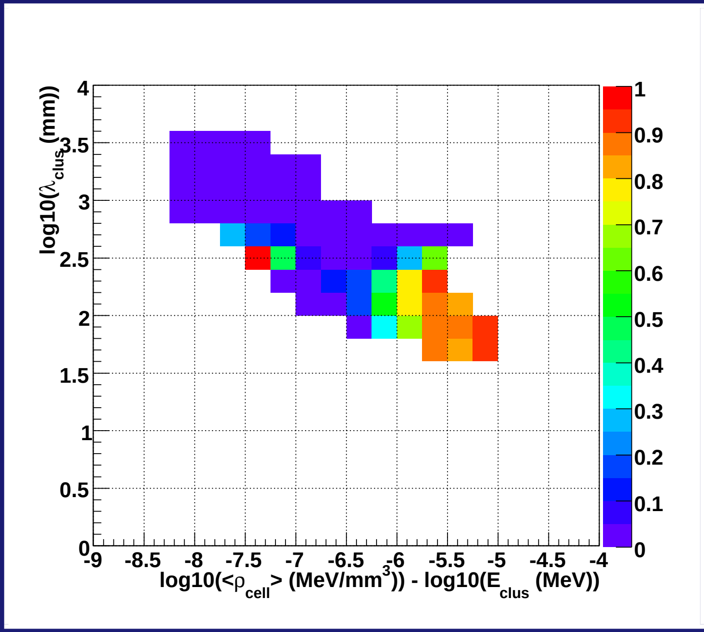
Cluster Moments ► LATERAL

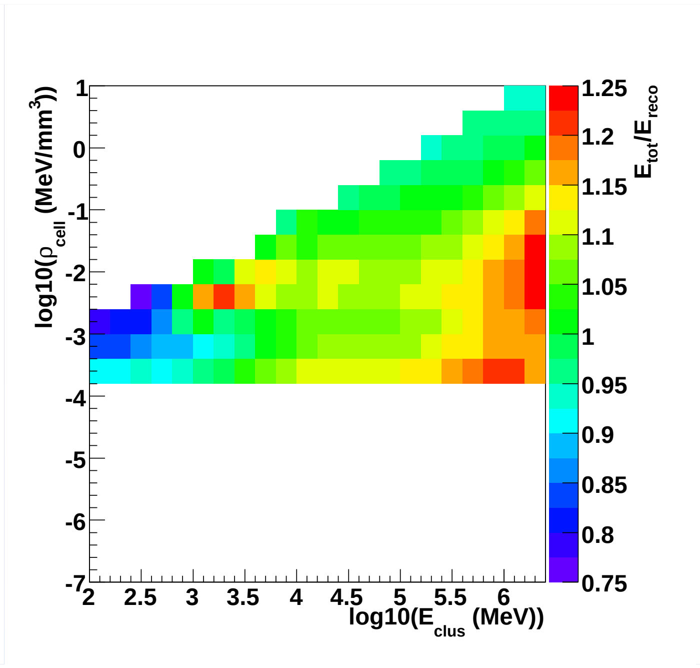
- like LONGITUDINAL moment potentially good moment to separate em/had
- in general very good agreement between data and MC
- bigger deviations in crack regions
- data shows flatter showers than MC



Local Hadron Calibration

- Classify and calibrate topo clusters to hadron-level
- Classification
  - use shower shape variables (cluster moments) like shower depth and (weighted) energy density of the cell constituents
  - em showers are less deep and have higher average energy density than had showers
  - make a cut on probability ratio to observe a neutral over a charged pion in a given bin derived from single pion simulations (right plot)





- Calibration
  - cell weights are applied to clusters classified as hadronic
  - derive cell weights from Geant4 true energy (calibration hits) including invisible energy and absorber deposits and reconstructed cell energy for each η region and layer: $w_i = \langle E_{\text{true}} / E_{\text{reco}} \rangle$, $i = \text{bin\#}(E_{\text{cluster}}, E_{\text{cell}} / V_{\text{cell}})$
  - example weights in main sampling of EM calorimeter for $2.0 < |\eta| < 2.2$
- Correct for dead material and out-of-cluster deposits for clusters classified as hadronic and electromagnetic (corrections differ)

Local Hadron Calibration ► Energy Corrections

► Cell weights

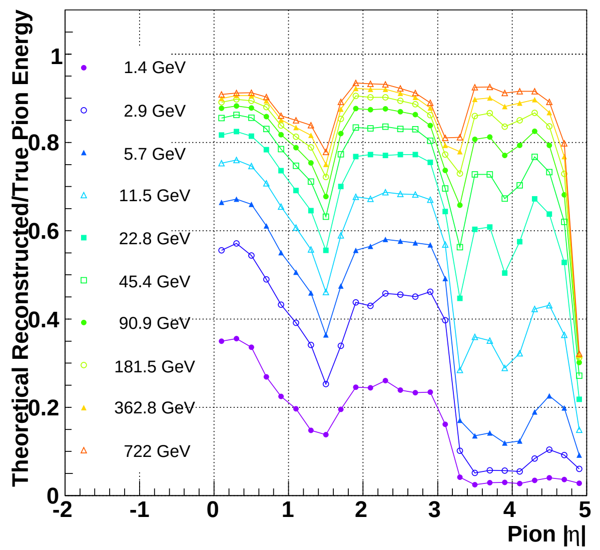
- account for the non-compensation of the calorimeters

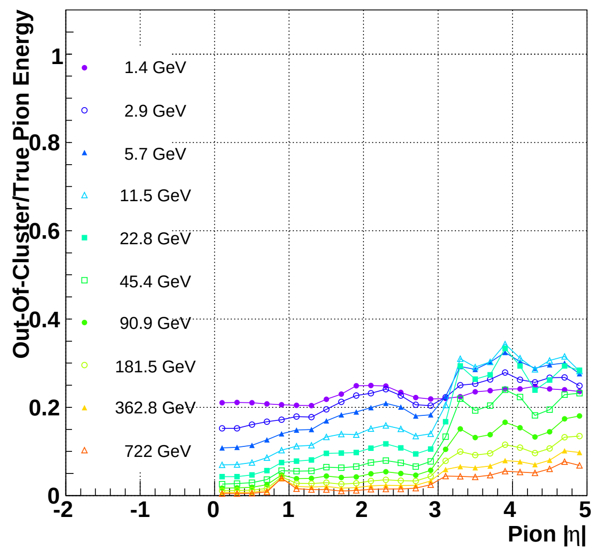
► Out-Of-Cluster Corrections

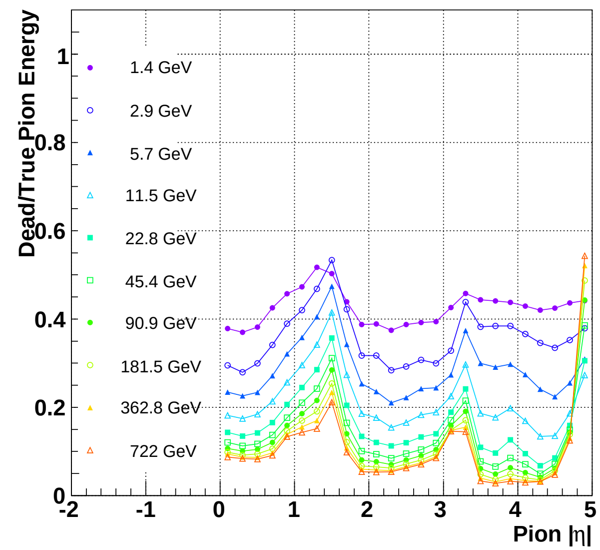
- recover lost energy inside the calorimeters due to noise thresholds

► Dead-Material Corrections

- recover lost energy outside the calorimeters







► Cell Weights

- can be defined non-ambiguously from calibration hits and reconstructed cell energy

► Out-Of-Cluster & Dead-Material corrections

- need assignment algorithm of nearby calibration hits to clusters
- can correct only those cases where a signal cluster is present
  - ► jets need additional corrections for lost low energetic particles

- use separate charged and neutral single pion samples and fill all found clusters in 2D histograms in $\log_{10}\lambda$ and $\log_{10}\langle\rho\rangle$
- define individual weights for each bin i:
  $w^i_{\pi^0} = N^i_{\pi^0}/N_{\pi^0}$
  $w^i_{\pi^\pm} = N^i_{\pi^\pm}/N_{\pi^\pm}$
- take a-priori probability to get a neutral pion instead of a charged pion of $1/3$
- combine separate weights bin by bin to create 2D probability table:
  $p^i_{\pi^0} = \frac{w^i_{\pi^0}}{w^i_{\pi^0}+2w_{\pi^\pm}}$
- classify each cluster as em that falls into a bin with $p^i_{\pi^0} > 0.5$
- since we have typically more than one cluster per pion and the ratio is different for charged and neutral pions, each cluster enters with a weight $E_{\mathrm{CalibHit}}/\sum E_{\mathrm{CalibHit}}$, such that every pion (or event) contributes with total weight 1 and the leading cluster contributes most

Corrections

- After the classification clusters were treated either as electromagnetic or as hadronic
- This is o.k. for very clear assignments (em-probabilities above 90% or below 10%)
- Often clusters are classified much closer to 50%
- Here similar clusters could end up with completely different weights depending on few percent-points in the probability
- The classification now stores the moment `EM_PROBABILITY` (ESD only) and subsequent weighting steps can use it to gradually switch from em to had weights:
  - $w = p_{\mathrm{EM}} w_{\mathrm{EM}} + (1 - p_{\mathrm{EM}}) w_{\mathrm{HAD}}$
- This is default since `15.5.4`

Performance of single pions

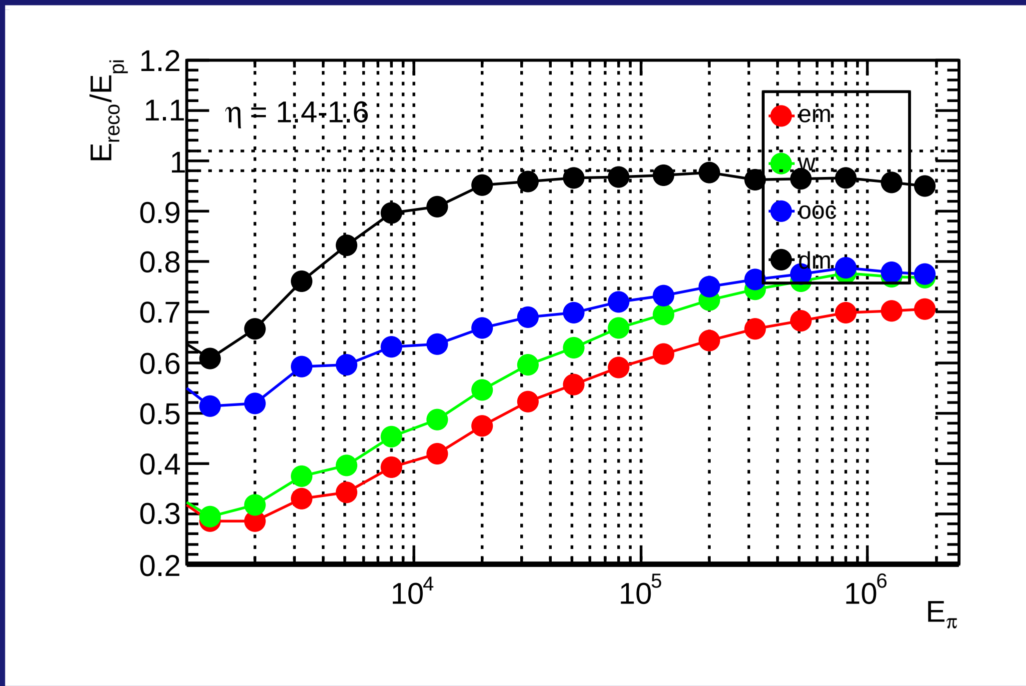
Plots produced by G. Pospelov with

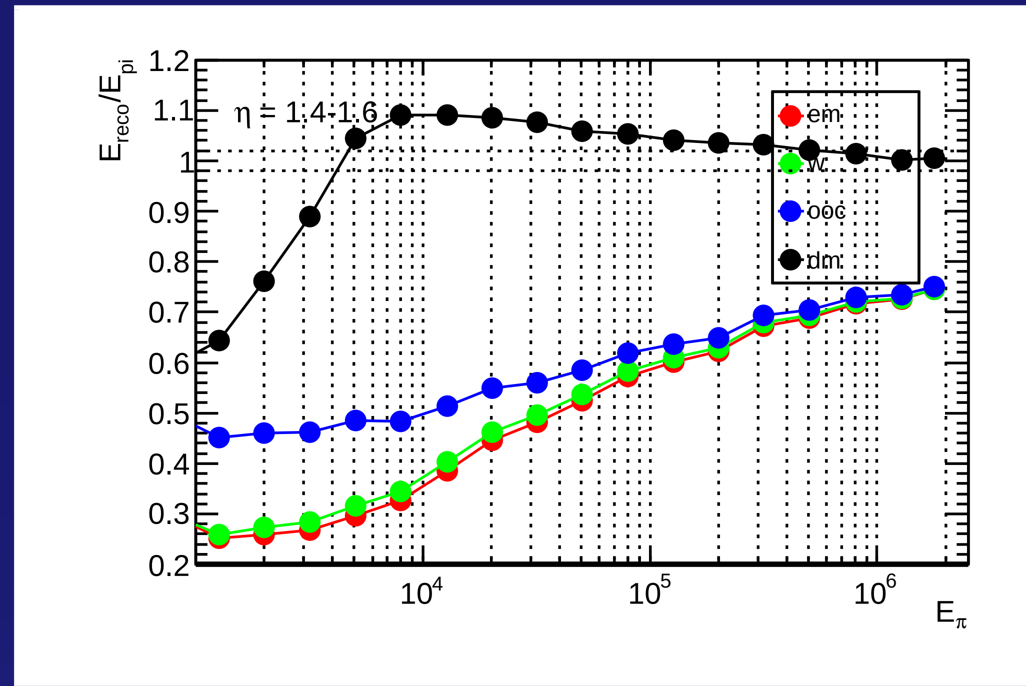
`CaloLocalHadCalib/GetLCSinglePionsPerf`

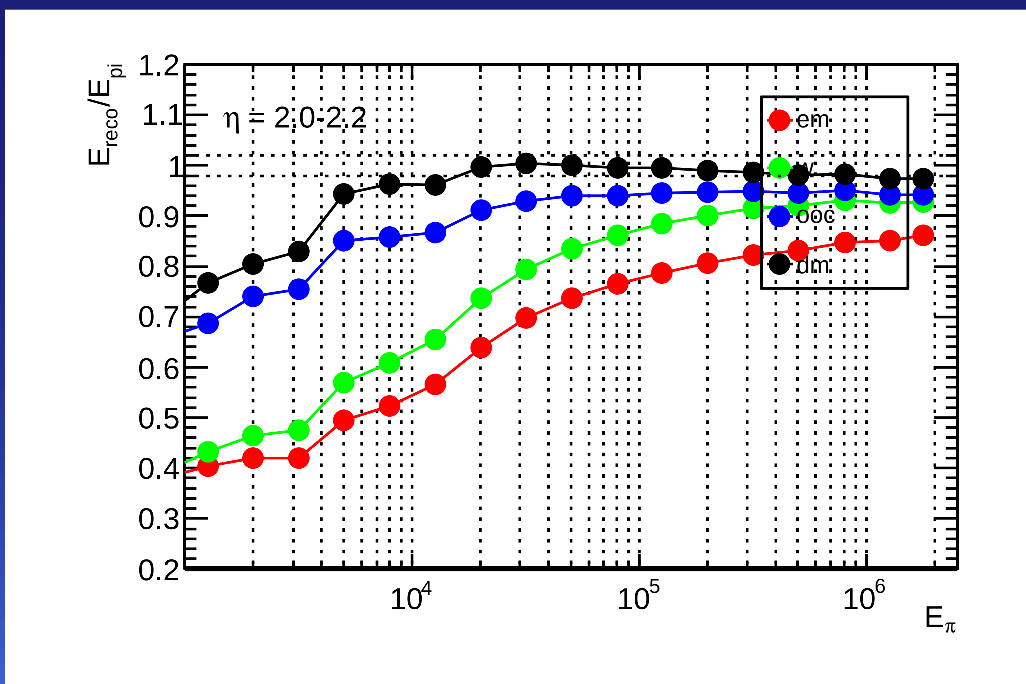
- Linearity/resolution for $\pi^{\pm}$ (left) and π^{0} (right) in `15.5.4`
- Examples are for $1.4 < |\eta| < 1.6$ (top) and $2.0 < |\eta| < 2.2$ (bottom)
- Other regions show similar tendency
- good linearity above $10\,\mathrm{GeV}$ for charged pions
- small overshoots for neutral pions in regions with large corrections
- improved resolution especially in crack regions

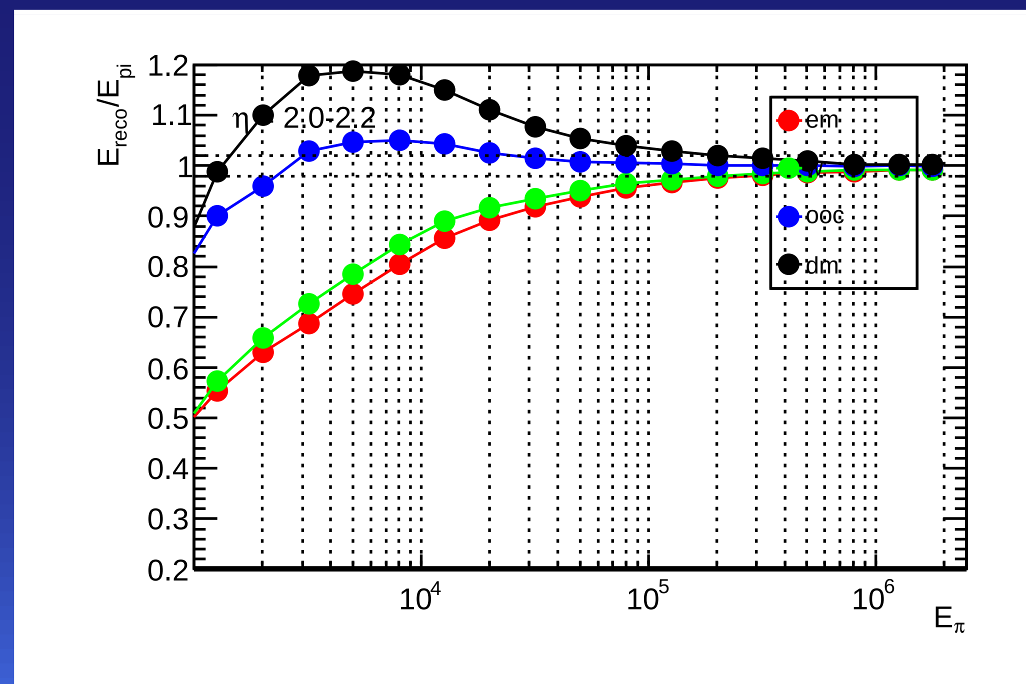
Performance of single pions ► $\pi^{\pm/0}$ Linearity

G. Pospelov





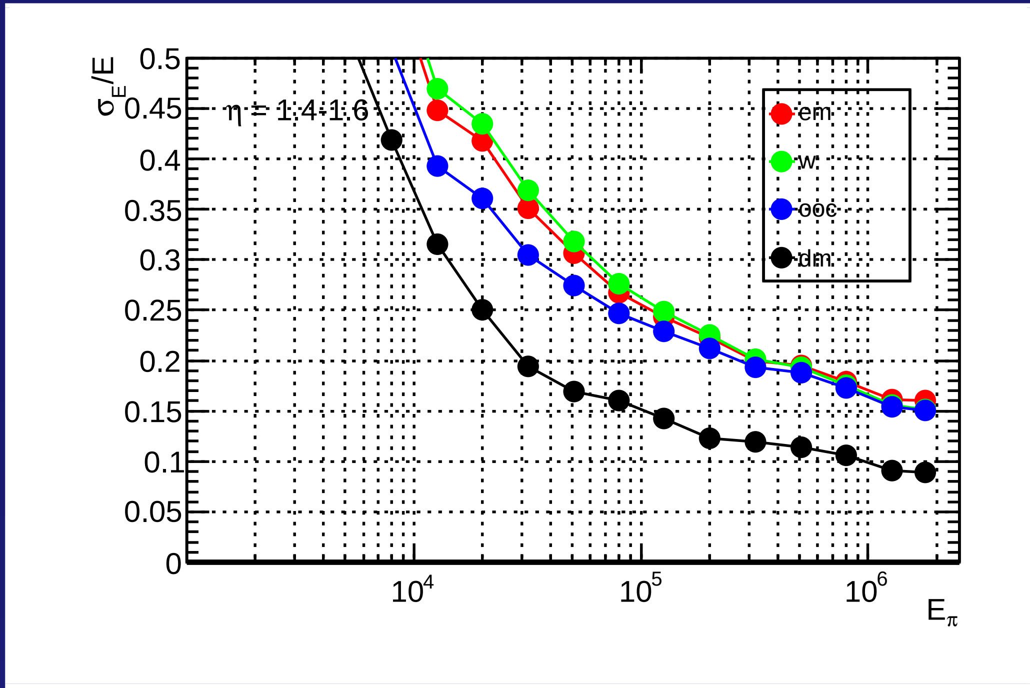


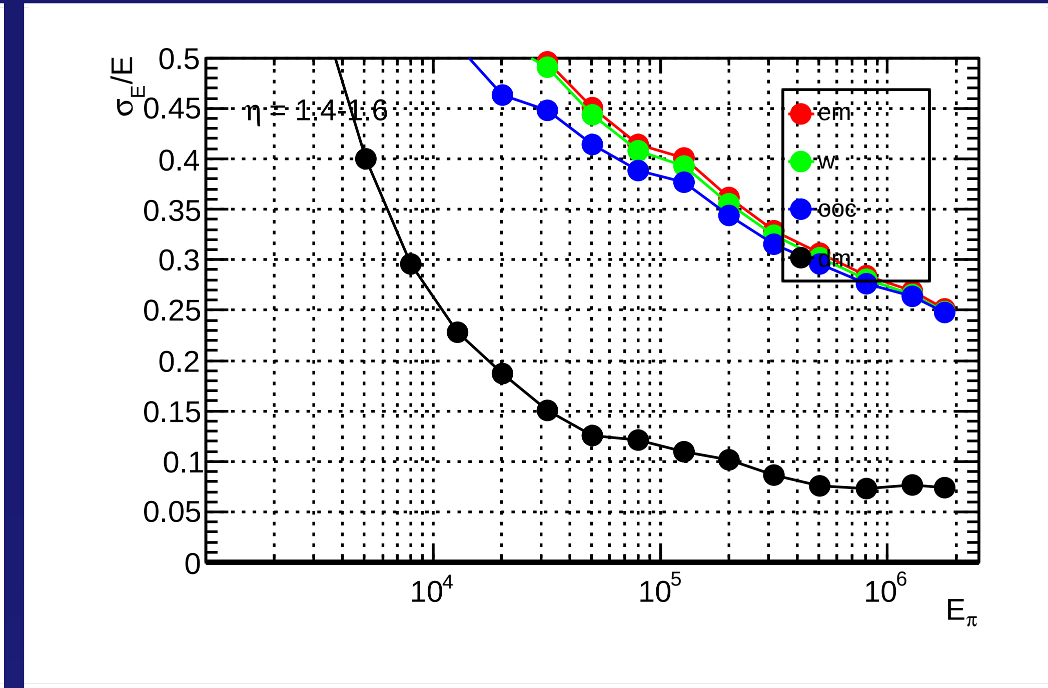


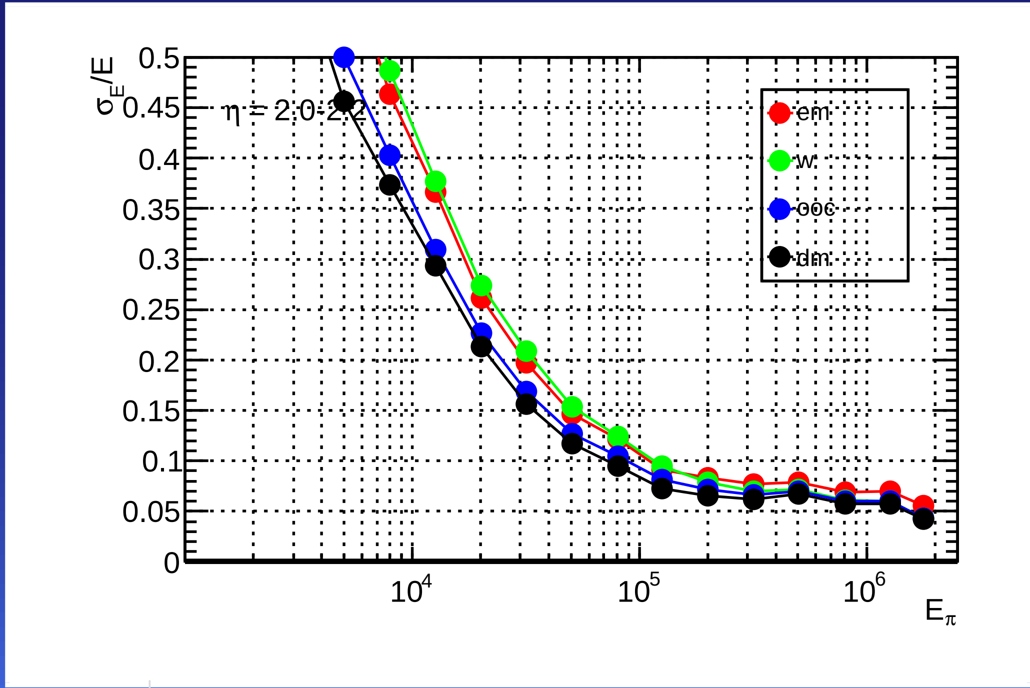
$\pi^{\pm}$ π^{0}

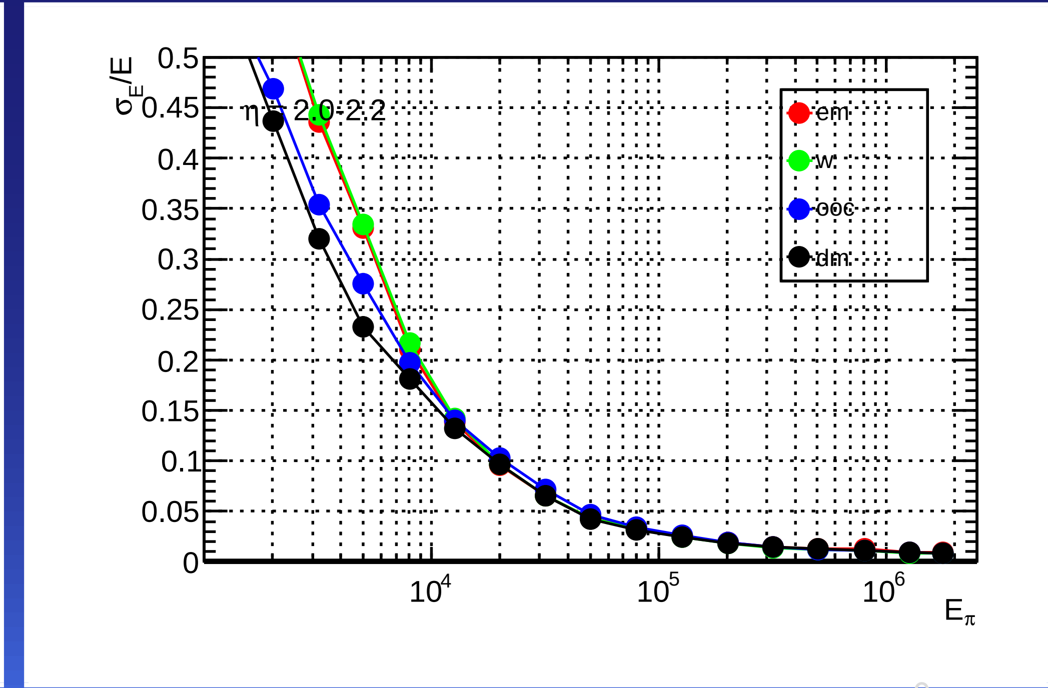
Performance of single pions ► $\pi^{\pm/0}$ Resolution

G. Pospelov







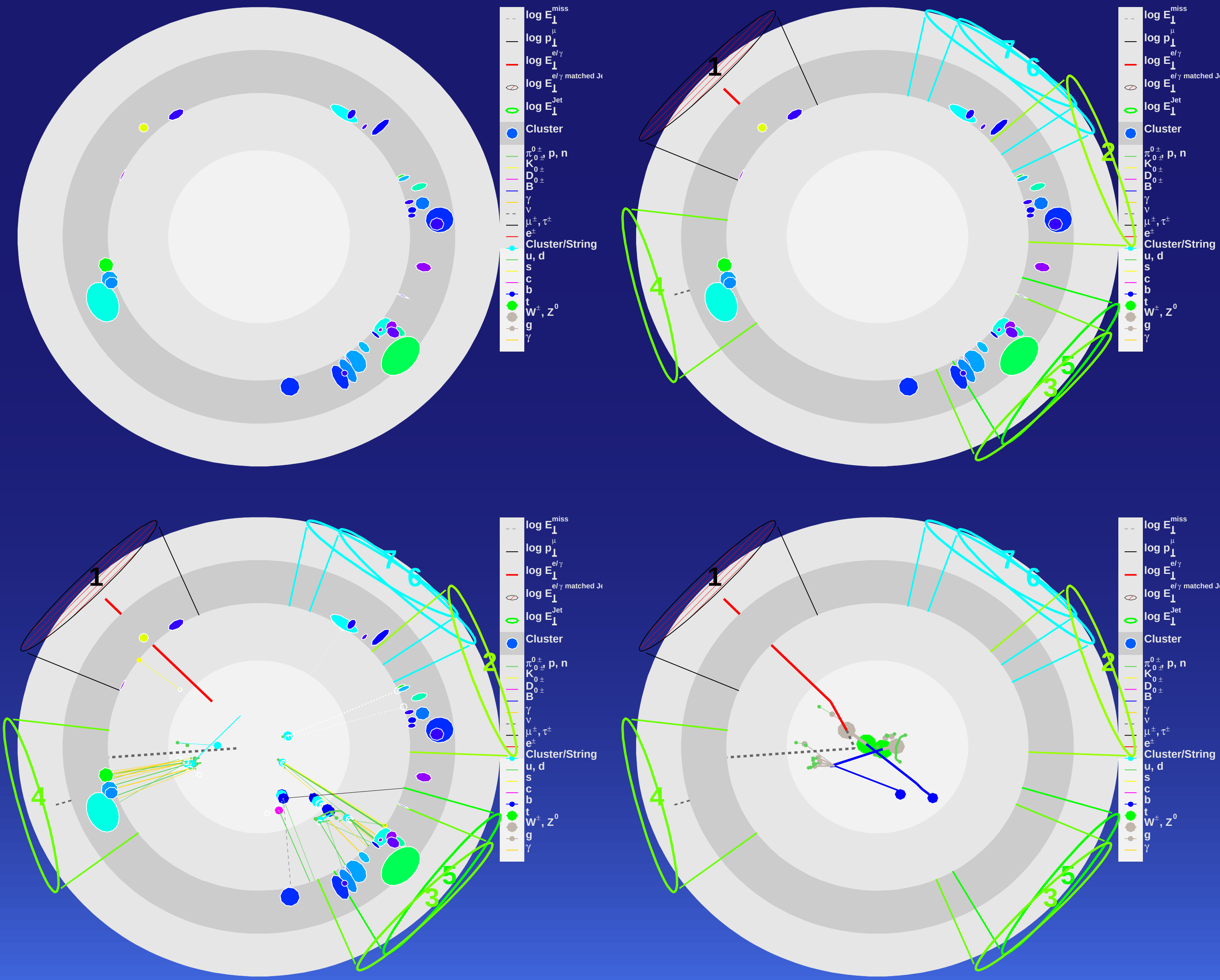


$\pi^\pm$ π^0

Topo Clusters as Input to Jets

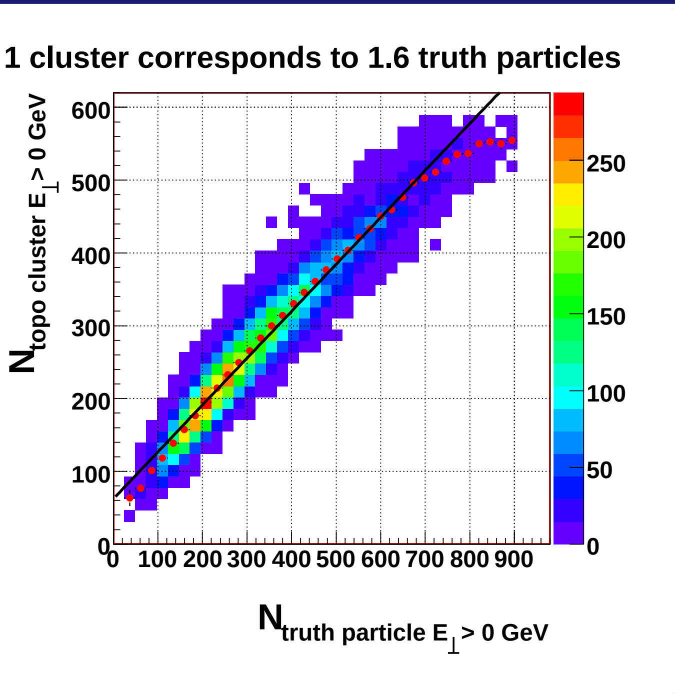
- Jet reconstruction and calibration can be divided in 4 steps

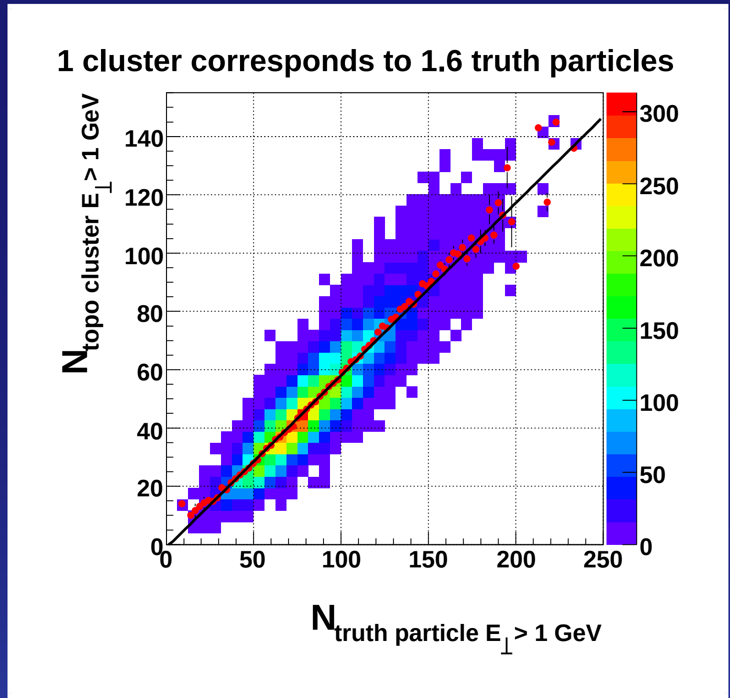
1. calorimeter tower/cluster reconstruction
2. jet making
3. jet calibration from calorimeter to particle scale
4. jet calibration from particle scale to the parton scale

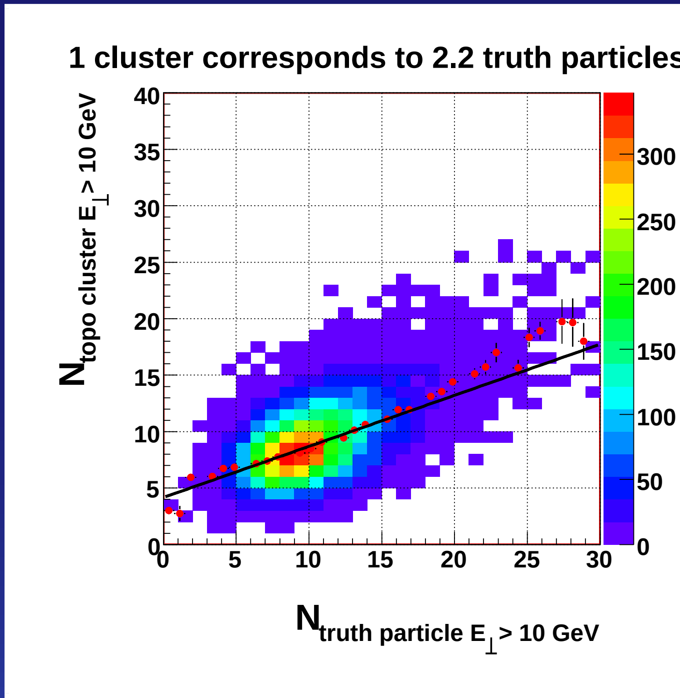


Jet Input

- Cluster particle correspondance
  - 1.6 stable particles for 1 cluster @ $E_T > 0$
  - 1.6 stable particles for 1 cluster @ $E_T > 1\,\text{GeV}$
  - 2.2 stable particles for 1 cluster @ $E_T > 10\,\text{GeV}$ (mainly due to merged photons from $\pi^0 \to \gamma\gamma$)



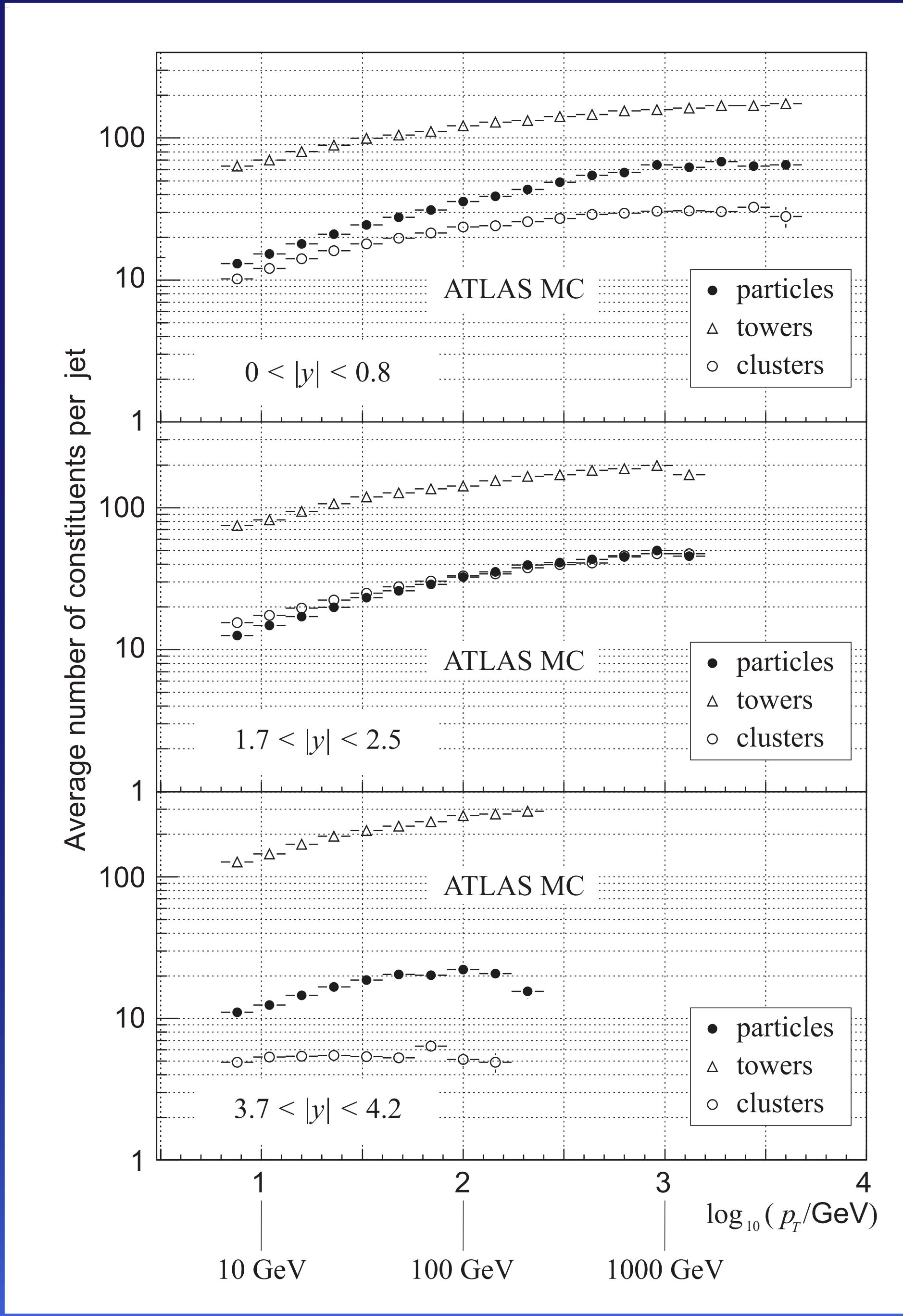




Jet Input ► Number of Constituents

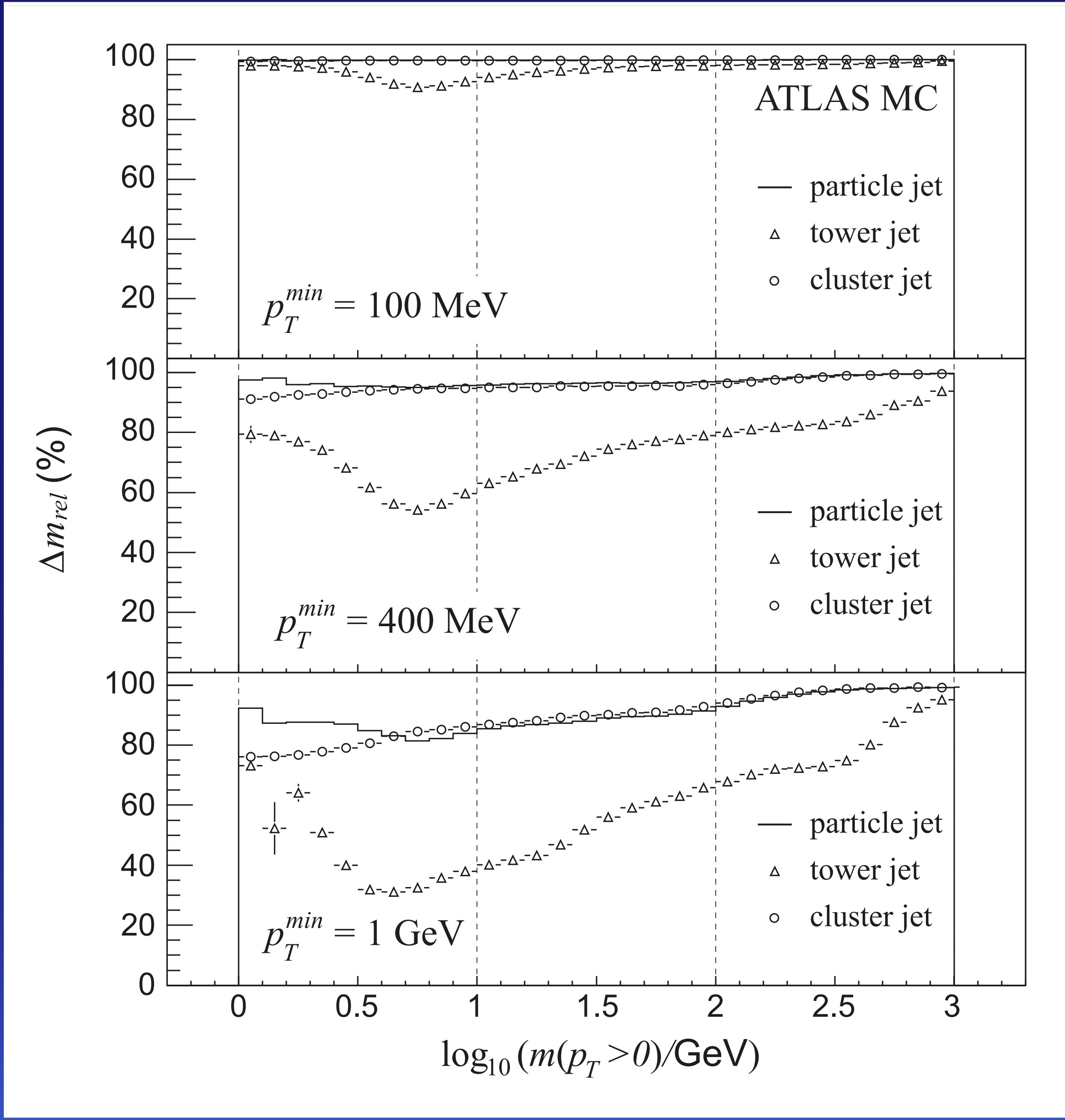
► Number of Constitutents per jet

- CSC book; di-jet MC; `Kt6` jets
- stable particles vs. topo clusters vs. towers
- clusters much closer to truth than towers



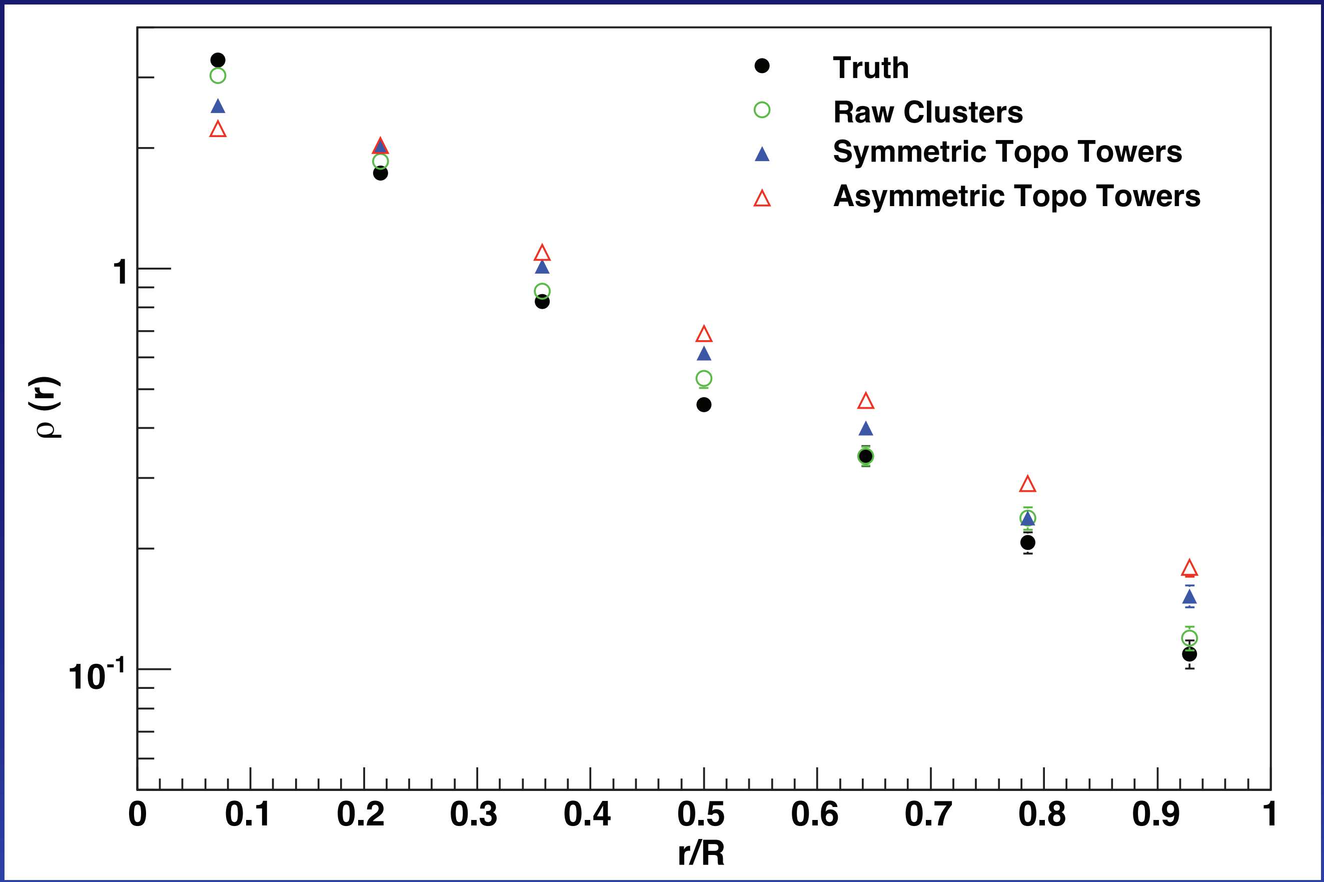
Jet Input ► Jet Mass

- Jet Mass for the same choice of jet inputs
  - CSC book; di-jet MC; `Kt6` jets
  - stable particles vs. topo clusters vs. towers
  - again cluster jets much closer to truth than towers



- Jet Shape (i.e. radial energy flow)

  - ATL-PHYS-INT-2009-099
  - stable particles vs. topo clusters vs. topo towers
  - again cluster jets closer to truth than towers
  - but topo towers already better than towers (unfortunately plain towers not plotted)



Constituents based Jet Level Corrections (Method)

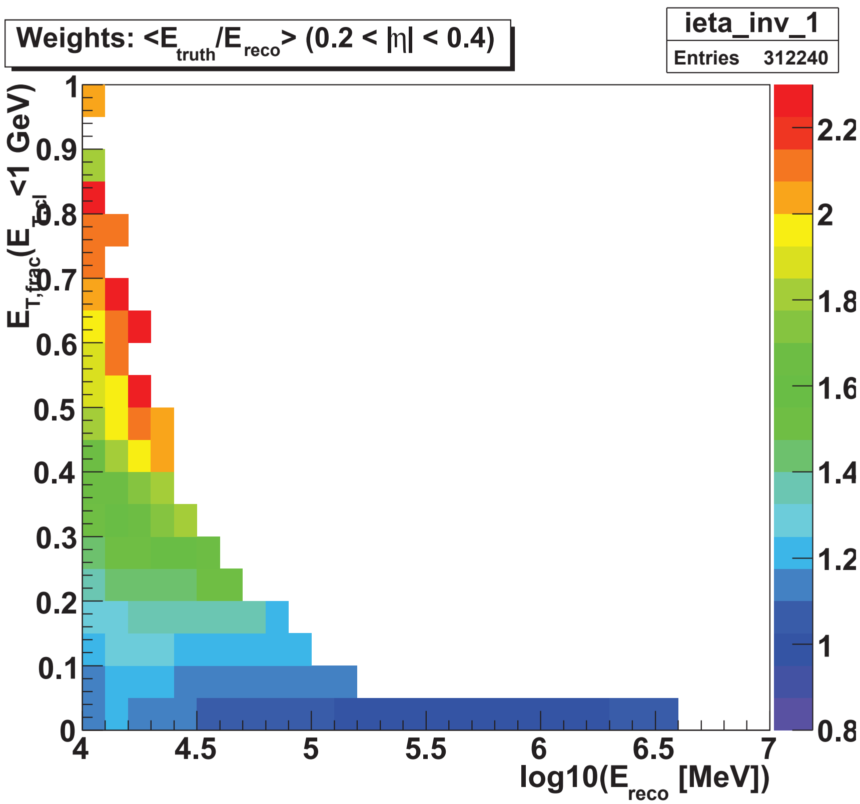
Idea:

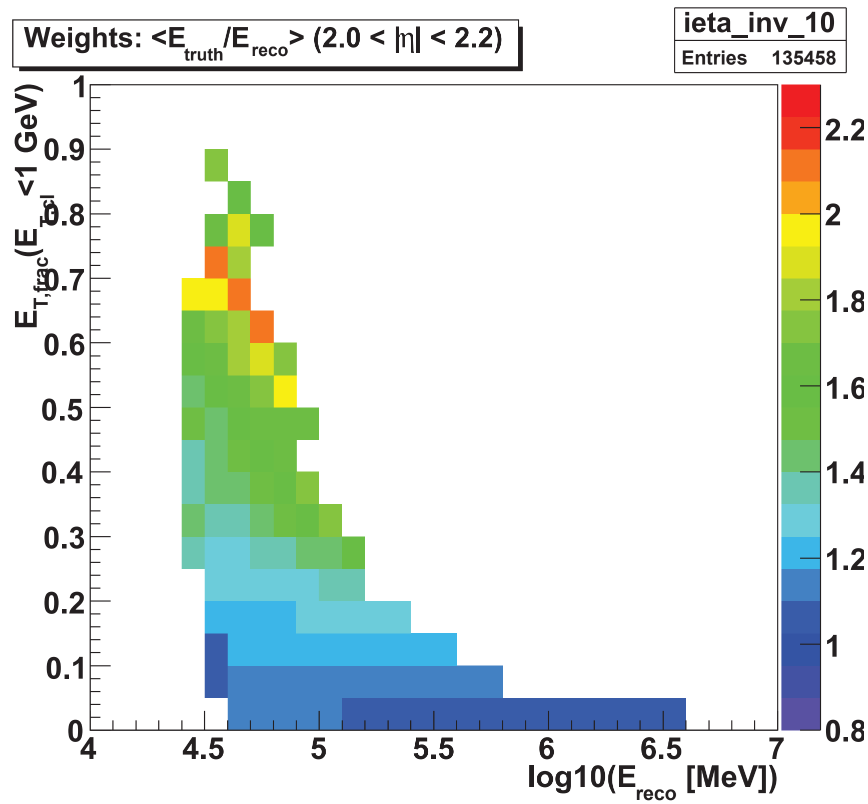
- Monte Carlo based Jet Level Corrections
- What does a jet know about its missing constituents?
- Jet shape/sub-structure information:
  - $\Rightarrow$ jet moments (JM) based on jet constituents can be used
- Matched truth/reco jet pairs: $\Delta R_{min}^{match} < 0.3$ & $p_T^{true} > 20\ GeV$
- Isolated reco jets: $\Delta R_{min}^{iso} > 1.0$ w.r.t. clostest reco or truth jet
  - $\Rightarrow$ anti-K_T4 jets
  - $\Rightarrow$ inversion method used for energy bins
- The jet moment used here is:

$$E_{T,frac,cl\,(E_{T,cl}<1\ GeV)} = \frac{1}{E_{T,jet}} \sum\nolimits_{cl\,(E_{T,cl}<1\ GeV)} E_{T,cl}$$

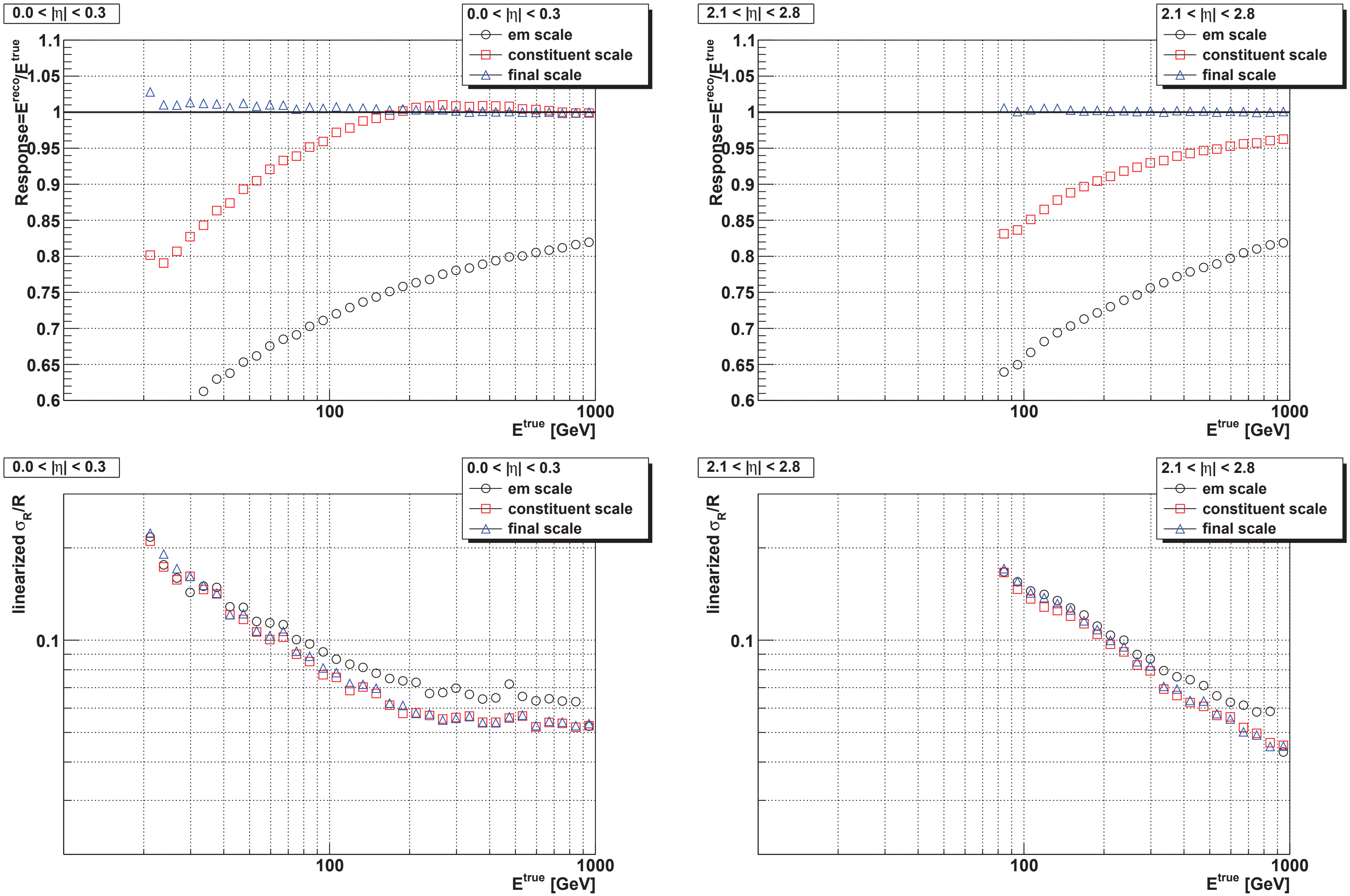
Constituents based Jet Level Corrections (Method)

Weights: $\left\langle \frac{E_{truth}}{E_{reco}} \right\rangle$ in bins of $\log10(E_{\mathbf{reco}})$, $|\eta|$ & $E_{T,frac,cl}\,(E_{T,cl} < \mathbf{1}\ GeV)$





QCD DiJet (J0-J8)



Conclusions

- Topo Clusters preserve particle picture, suppress noise and follow showers
- Cluster moments allow classification
- Local Hadron Calibration based on cell weights, cluster shapes
- Modular approach to treat one effect at a time
  - non-compensation
  - losses due to noise thresholds
  - losses due to non-instrumented material
- Clusters as input to jets preserve jet shapes
- Allow for constituent based jet energy scale corrections
- Suitable for quark and gluon jets
- Next steps:
  - Provide systematic error tool for clusters and LCTopo jets
  - Particle ID (`CalibrationHit` "knows" its parent particle)
  - Try jet MC to derive Local Had Calib weights instead of Single Pion MC