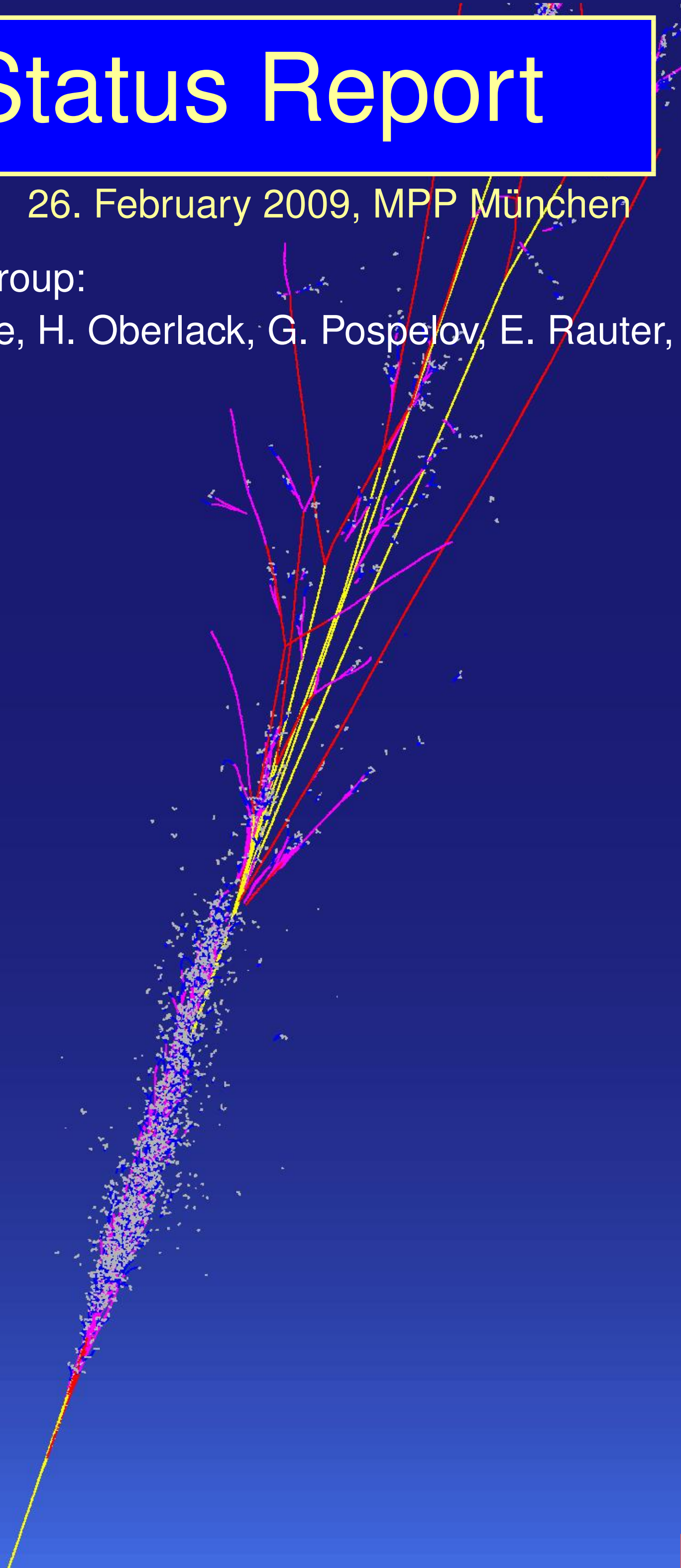
Local Hadron Calibration Status Report

LArg meeting | Sven Menke, MPP München | 26. February 2009, MPP München

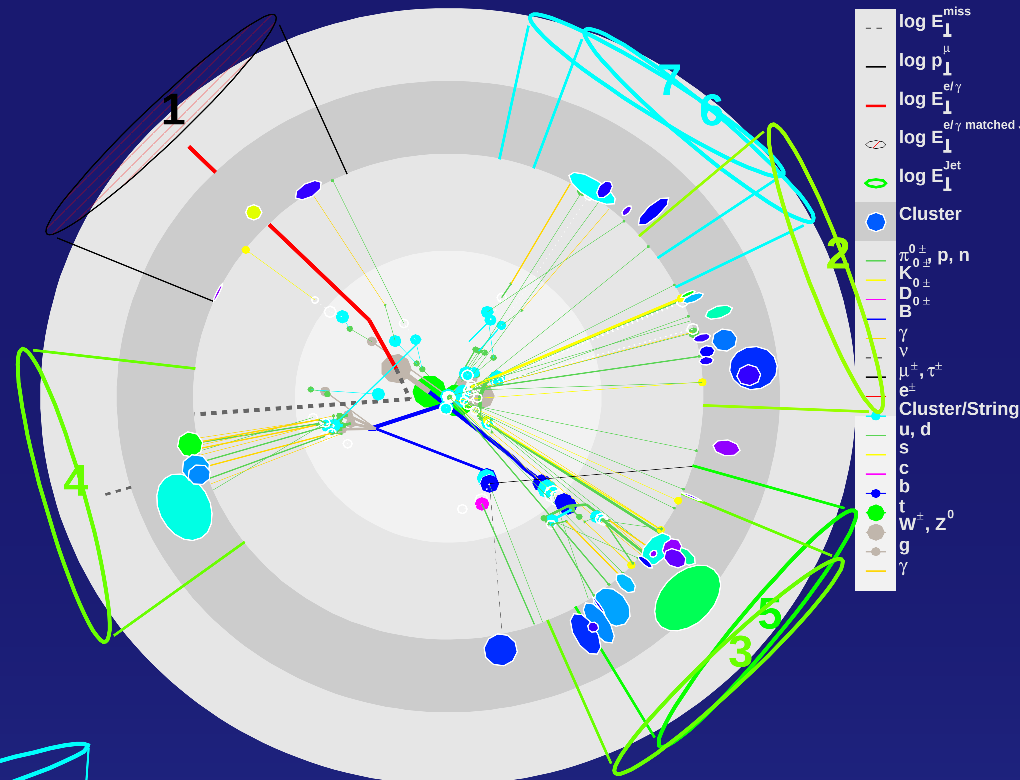
on behalf of the MPP Munich HEC Group:
T. Barillari, J. Erdmann, P. Giovannini, A. Jantsch, A. Kiryunin, S. Menke, H. Oberlack, G. Pospelov, E. Rauter, D. Salihagic, P. Schacht

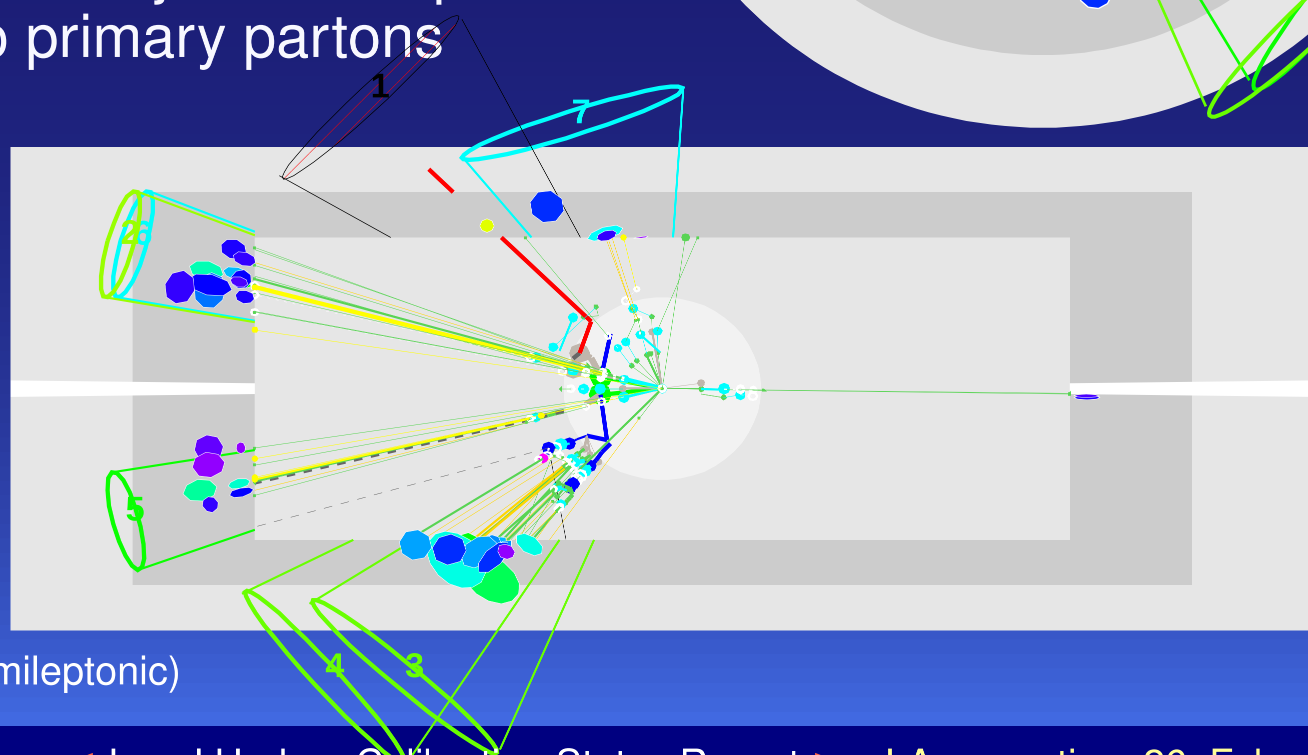
- Introduction
- Local hadron calibration
  - Classification
  - Cell weighting
  - Out-of-cluster corrections
  - Dead-material corrections
  - Jet-level corrections
- Status of new constants for `QGSP_BERT`
- Algorithm improvements
  - Particle ID for calibration hits
  - Truth particle extrapolation
- Readiness for First Data
  - Pile-up, noise, bad cells
  - in-situ strategies
  - Application to top events



Introduction

- 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

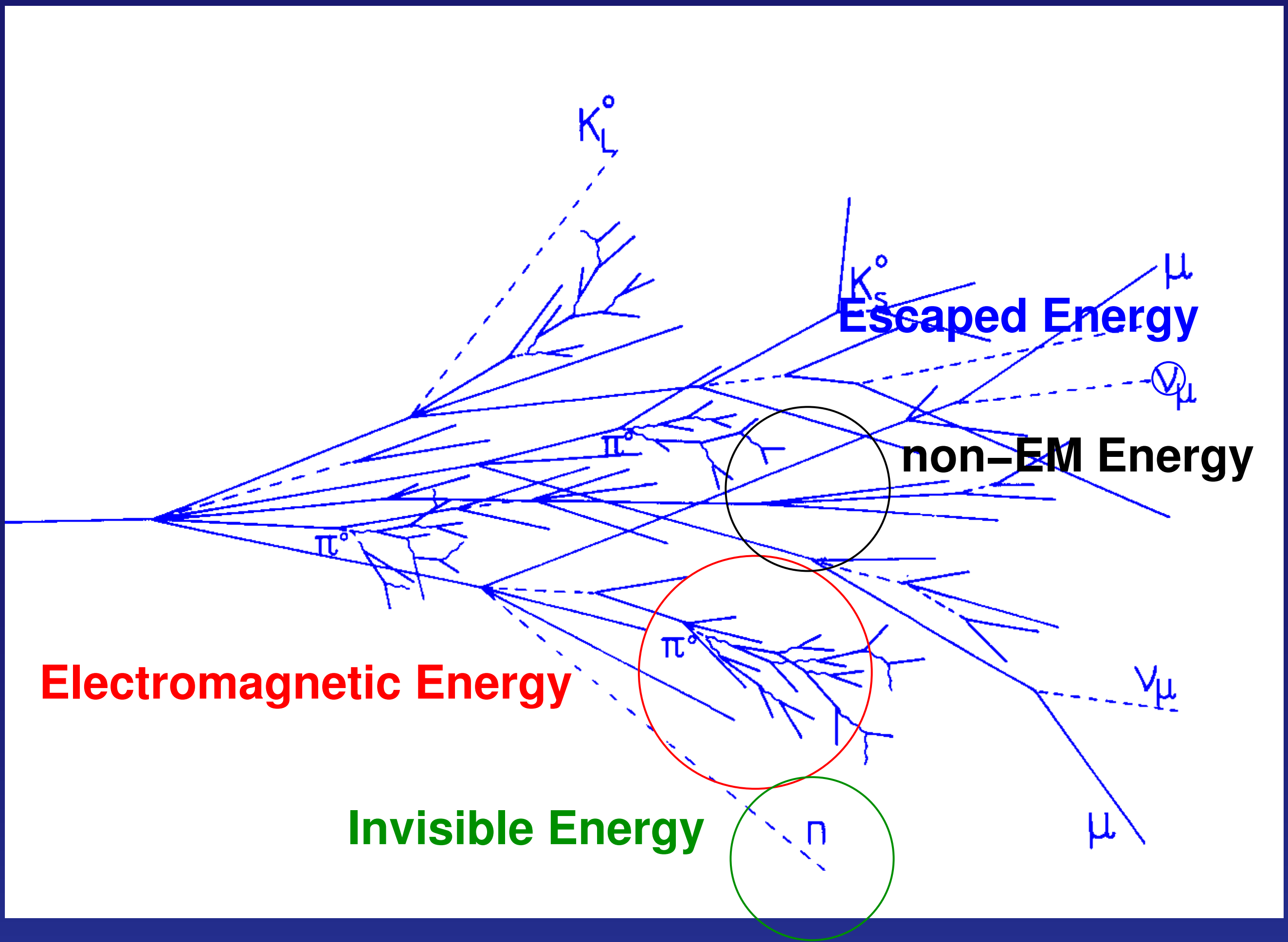




MC@NLO $t\bar{t}$ Event (semileptonic)

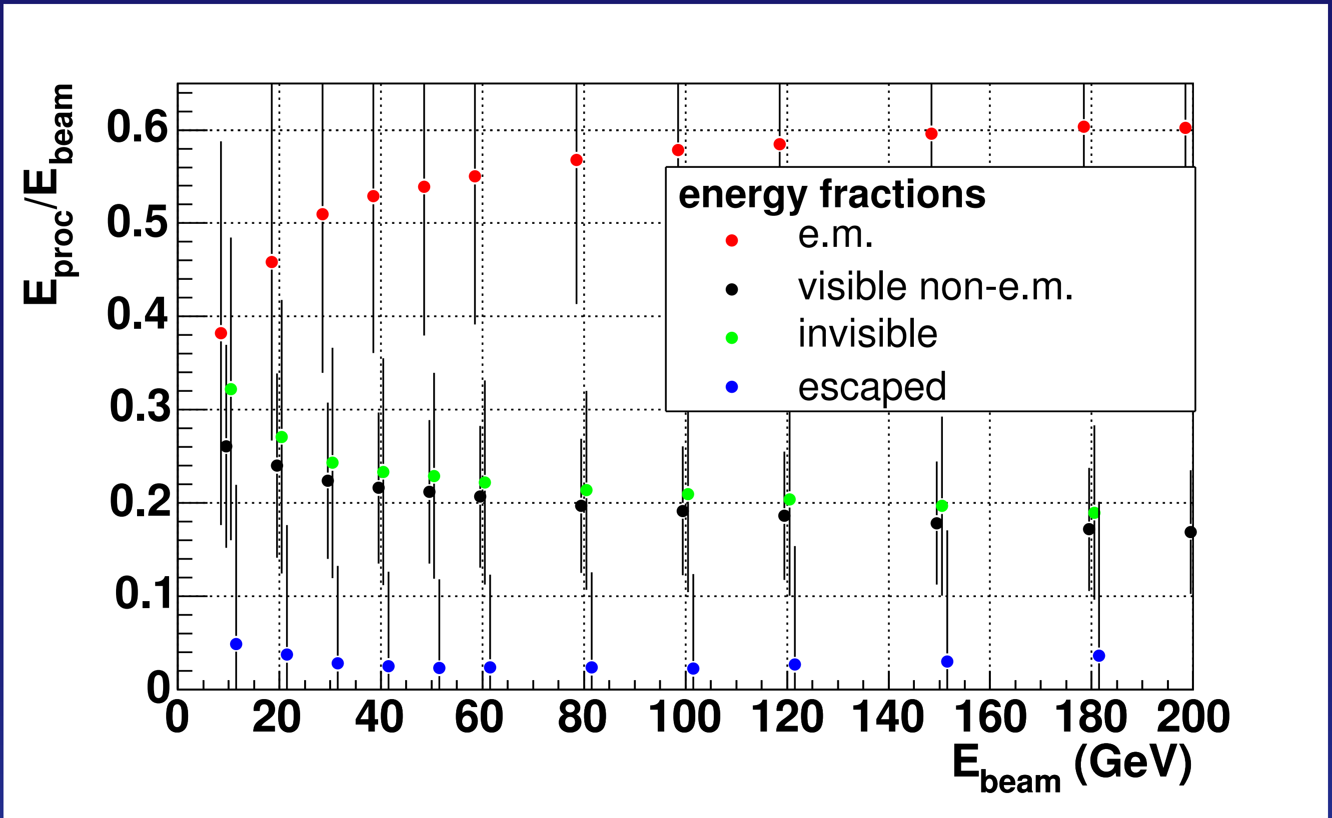
Hadron Calorimetry in ATLAS

- A hadronic shower consists of
  - EM energy (e.g. $\pi^0 \to \gamma\gamma$) $O(50\,\%)$
  - visible non-EM energy (e.g. dE/dx 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:

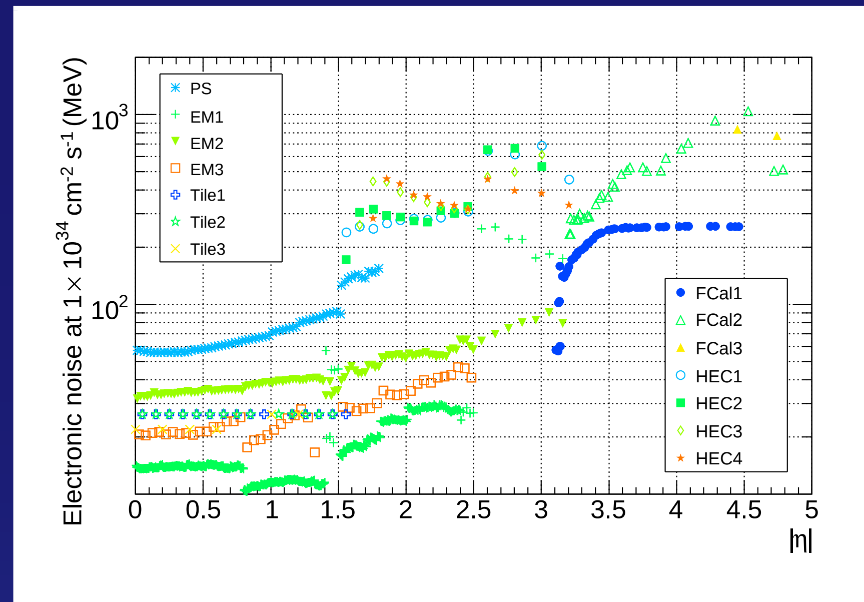


- EM energy strongly anti-correlated with visible non-EM energy
- visible non-EM energy strongly correlated with invisible energy

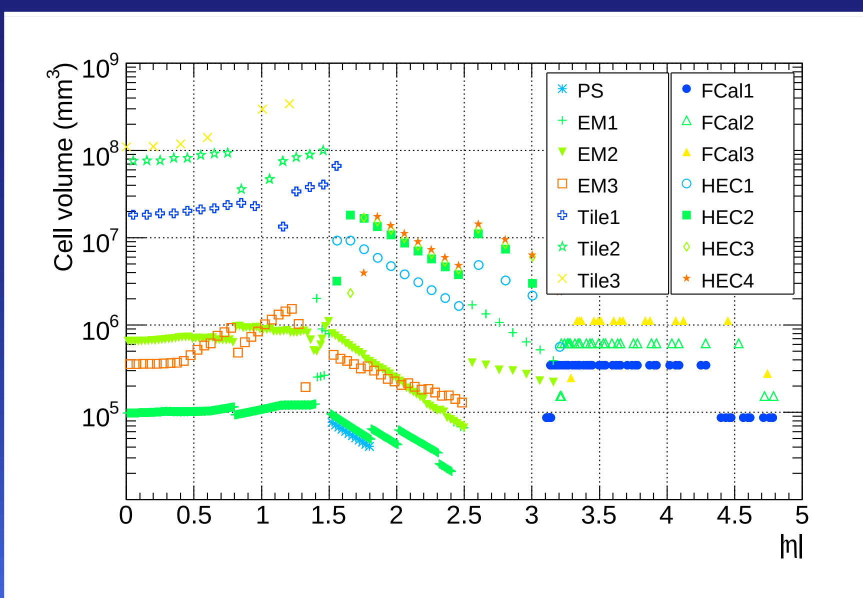
- need to separate EM part of the shower from the non-EM part
- apply a weight to the non-EM part to compensate invisible energy

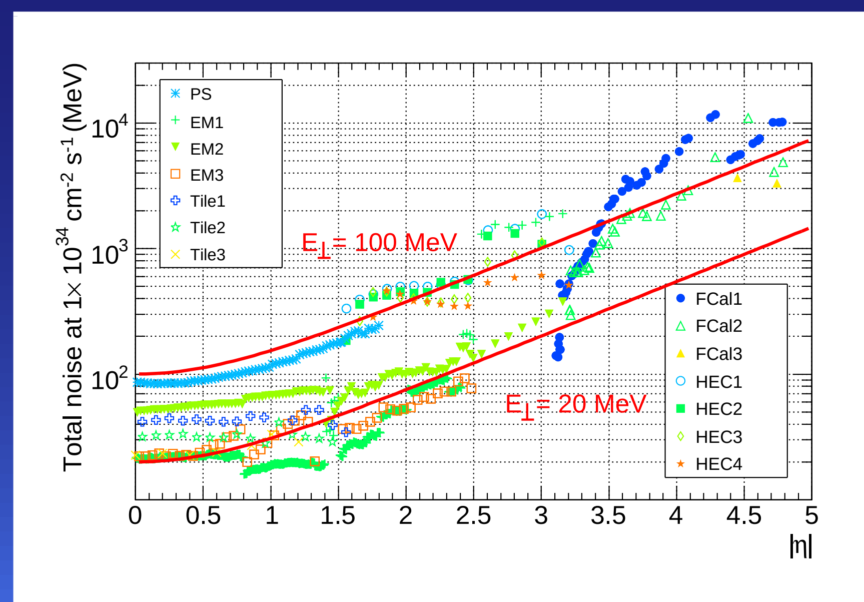
- How to separate EM fraction from non-EM fraction?
  - $X_0 \ll \lambda \simeq 20\,\mathrm{cm}$
  - high energy density in a cell denotes high EM activity
  - low energy density in a cell corresponds to hadronic activity
  - apply weights as function of energy density

- 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)



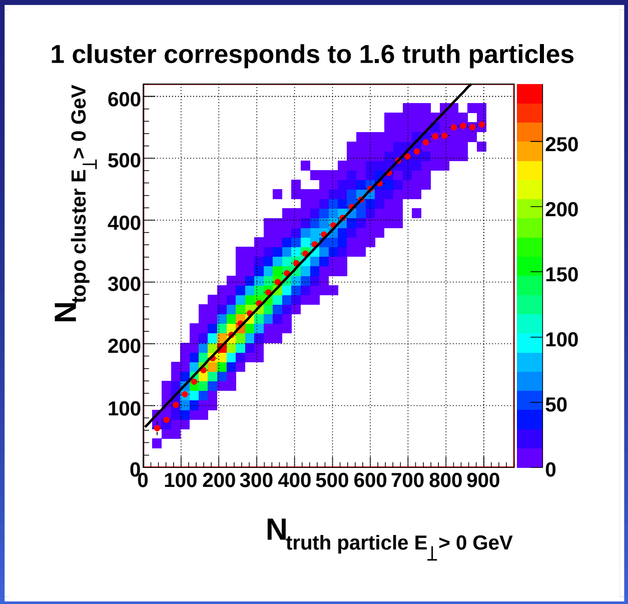


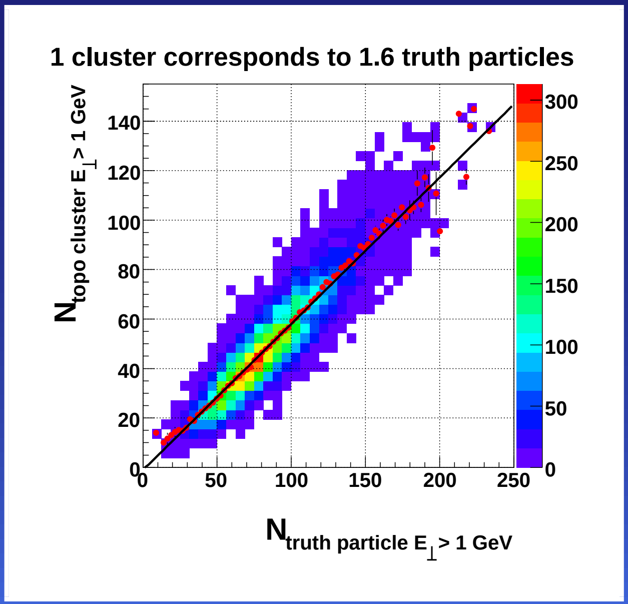
Topological Clusters

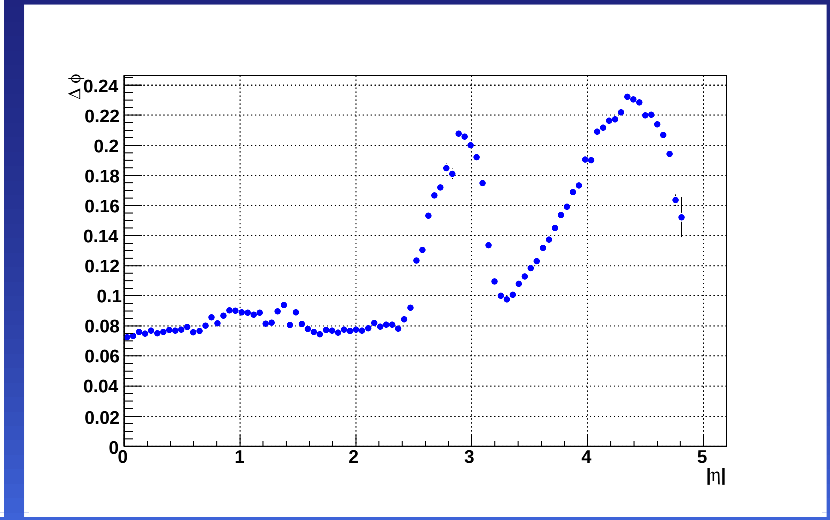
- Cluster Making
  - form clusters around seed cells with $|E_{\text{seed}}| > 4(\sigma_{\text{elec-noise}} \oplus \sigma_{\text{pile-up-noise}})$
  - expand clusters around neighbor cells with $|E_{\text{neigh}}| > 2\sigma$
  - include perimeter cells with $|E_{\text{cell}}| > 0\sigma$
  - merge clusters 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

Jets ► Input

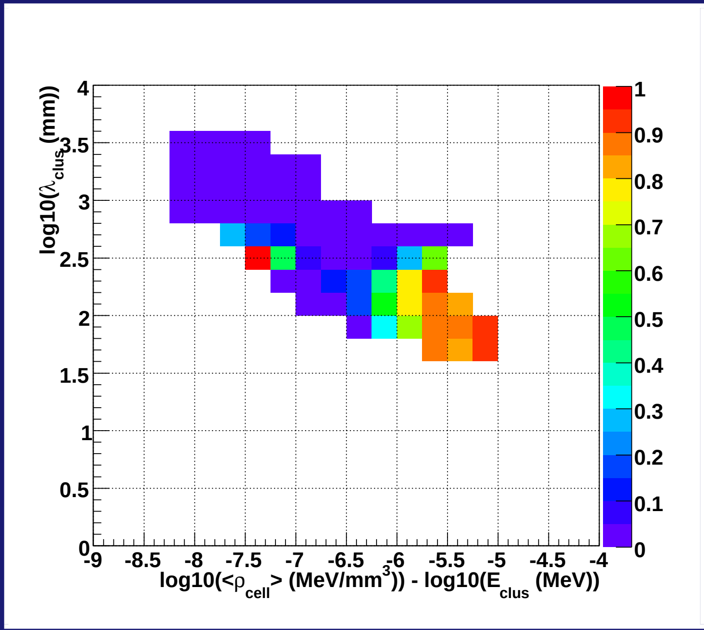
- Pro's & Con's of towers and topo clusters as jet input
- Towers
  - \+ have always the same fixed size $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$
  - \+ have no seed – all cells end up in towers
  - – do not provide noise or pile-up suppression
  - – do not contain showers
- Topo Clusters
  - \+ provide efficient noise and pile-up suppression
  - \+ correspond to individual hadrons
  - – need studies for the effects of noise thresholds ► work started by M. Simonyan
  - – typically have detector region dependent size $r \sim 0.1 - 0.2$



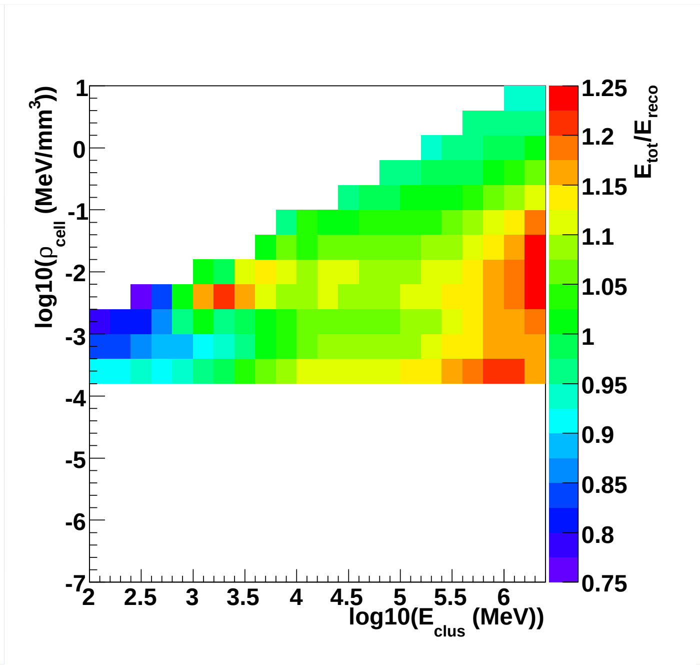




Local Hadron Calibration



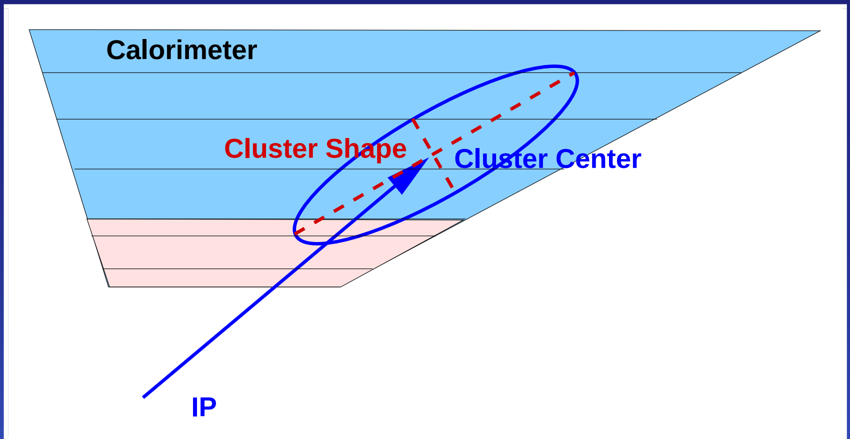
- Classify and calibrate topo clusters to hadron-level
- Classification
  - use shower shape variables (cluster moments) like shower depth and (weighted) energy density of the cell constituents
  - em showers are less deep and have higher average energy density than had showers
  - make a cut on probability 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: $w_i = \langle E_{\mathrm{true}} / E_{\mathrm{reco}} \rangle$, $i = \mathrm{bin\#}(E_{\mathrm{cluster}}, E_{\mathrm{cell}} / V_{\mathrm{cell}})$
  - 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)

9 most popular moments are on AOD

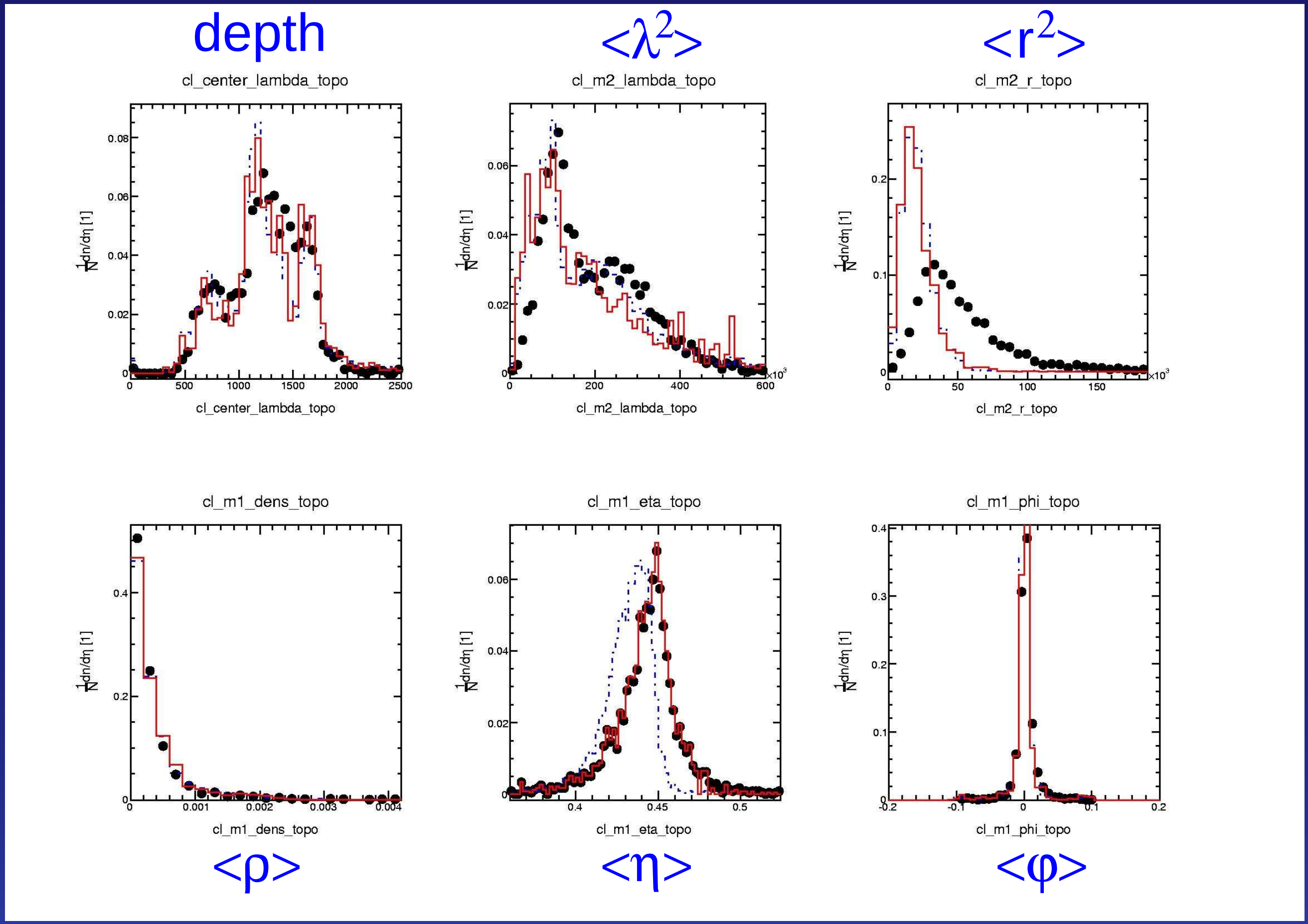
- `LATERAL` normalized second lateral moment
- `LONGITUDINAL` normalized second longitudinal moment
- `SECOND_R` the width squared of the cluster
- `SECOND_LAMBDA` the length squared of the
- `CENTER_LAMBDA` the cluster center depth in the calorimeter
- `CENTER_MAG` the distance IP - cluster center
- `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



other important moments available on ESD are

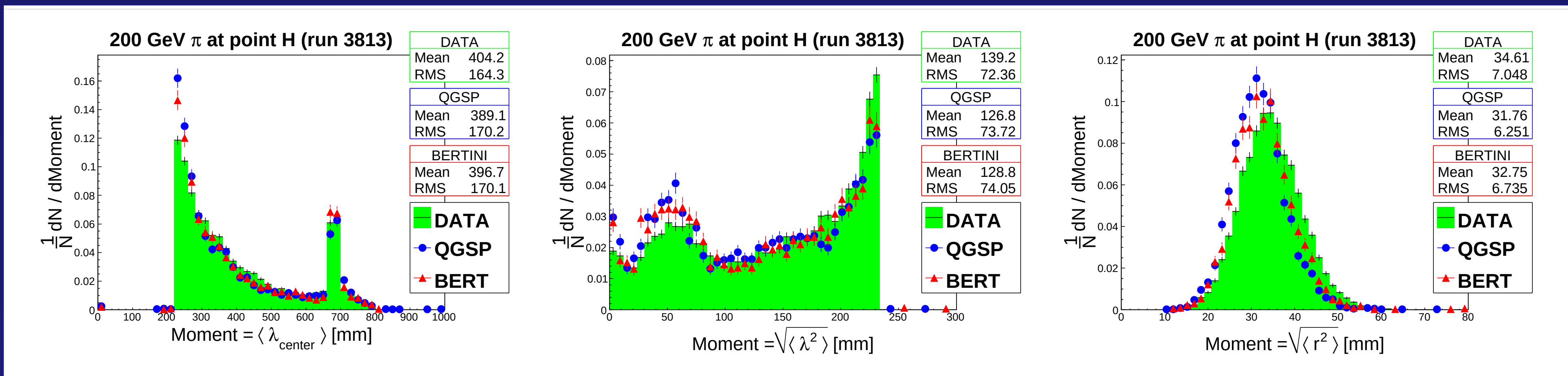
- `CENTER_X/Y/Z` the position of the cluster
- `ENG_FRAC_EM` the fraction of cluster energy in EM samplings
- `ENG_FRAC_CORE` the fraction of cluster energy in the leading cells in every sampling
- `DELTA_PHI/THETA/ALPHA` angular deviations of the shower axis from IP-cluster-center axis
- `ENG_CALIB_*` 13 of the 19 new moments of calibration hit energies associated to the cluster (in simulations with calibration hits; these are also on AOD)

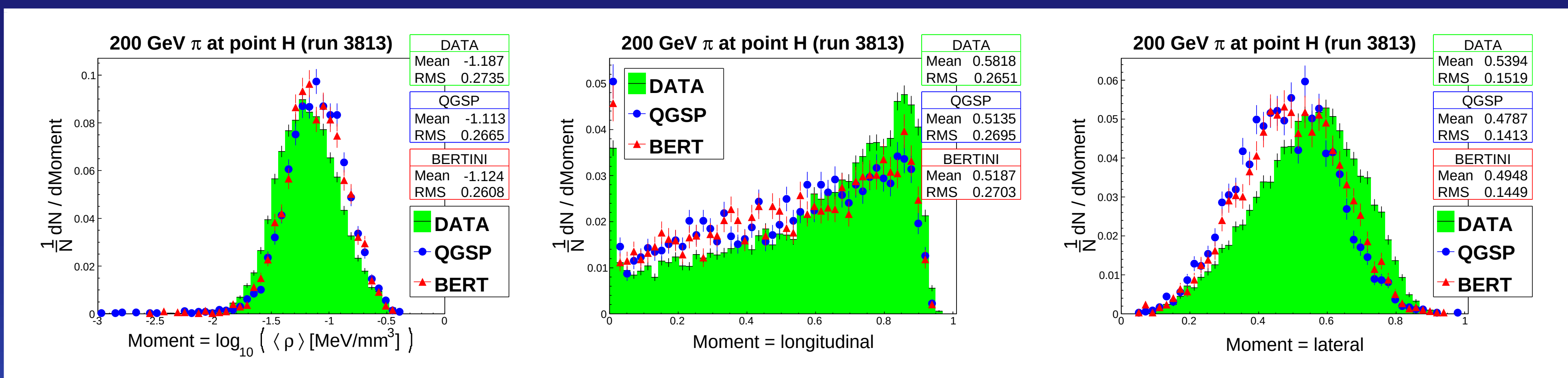
- look at cluster moments for 20 GeV pions from 2004 barrel test beam data (black points) and compare to G4 simulation (dashed blue lines)
- differences in $\langle \eta \rangle$ might be due to simplified beam trajectories in simulations
  - compare also with η-reweighted distributions (red)
- shower 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)

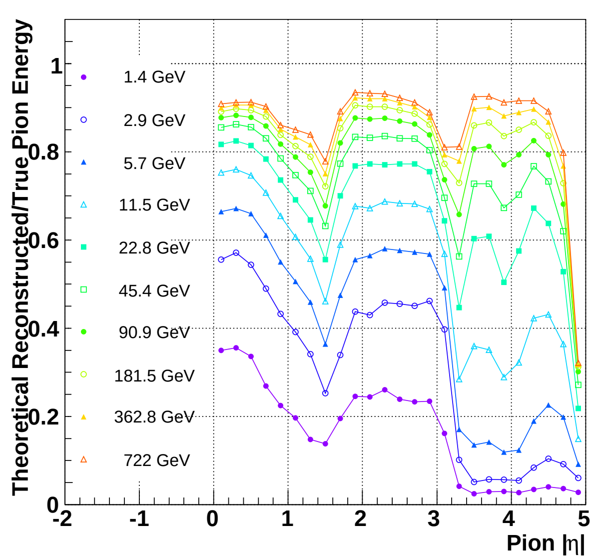


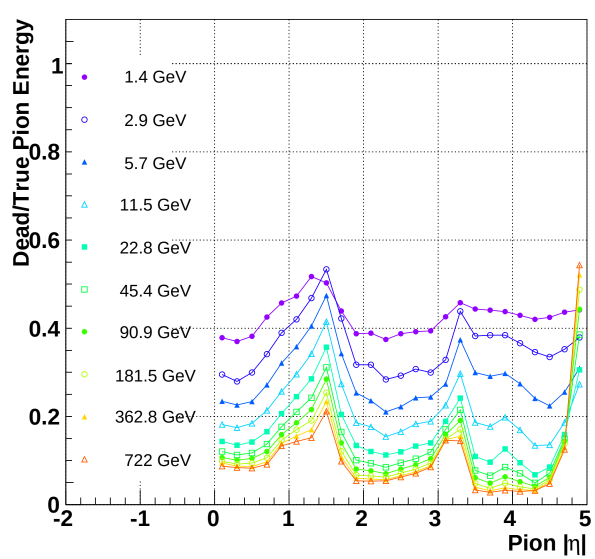


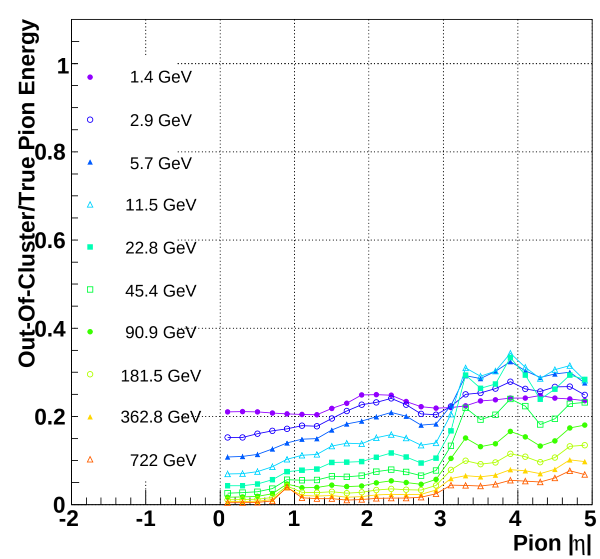
- best description for λ_{center} and $\langle\rho\rangle$
- largest deviations in `LATERAL` and `LONGITUDINAL`
- `QGSP_BERT` slightly better than `QGSP`

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

- New athena based weight extraction
  - new athena package CaloLocalHadCalib to
    ► extract classification, cell weights and out-of-cluster corrections from (private) ESDs with calibration hits
    ► Algorithms present are: GetLCClassification, GetLCWeights, GetLCOutOfCluster
    ► produce one set of histograms per athena job; merged later if needed
    ► package is in CVS and in release 14.5.0
- Refinement of hadronic weights with CaloLocalHadCalib and 14.2.21
  - based on $\sim 6 \cdot 10^6$ single pions ($\pi+$,π^-,π^0) produced as mc08.10741[012] on the grid
  - no noise cuts in cell selection
  - use inversion method ($E_{\mathrm{true}}/\langle E_{\mathrm{rec}}\rangle$instead of $\langle E_{\mathrm{true}}/E_{\mathrm{rec}}\rangle$)
  - include TileGap1, TileGap2

► Refinement of out-of-cluster corrections

- store corrections relative to cluster energy on EM-scale
- correct for out-of-cluster energy assigned to clusters only
  ► with help of recently added `CaloCalibHitClusterMoments`
  `ENG_CALIB_TOT`: total calib hit energy for a cluster
  `ENG_CALIB_OUT_M`: medium ($\Delta\alpha < 0.5$) associated out-of-cluster calibration hits in proportion to `ENG_CALIB_TOT` of all matching clusters
  `ENG_CALIB_OUT_L/T`: same for loose ($\Delta\alpha < 1.0$) and tight ($\Delta\alpha < 0.3$) association – disabled by default
- the new moments are available automatically for each dataset with calibration hits on ESD/AOD/DPD
- similarly for dead-material energy Gena implemented new calibration-hit-based moments
  `ENG_CALIB_DEAD_TOT`: dead material energy assigned to all clusters inside the medium cone (see above) with relative weight $\sqrt{E}\exp(-\Delta R/0.2)$.
- detailed studies are possible with the following additional moments defined by Gena:
  `ENG_CALIB_EMB0/EME0/TILEG3`: calibration hit energy in pre-sampler and tile gap scintillator inside clusters
  `ENG_CALIB_DEAD_EMB0/TILE0/TILEG3/EME0/HEC0/FCAL/LEAKAGE/UNCLASS`: associated dead material energy according to the 8 different regions

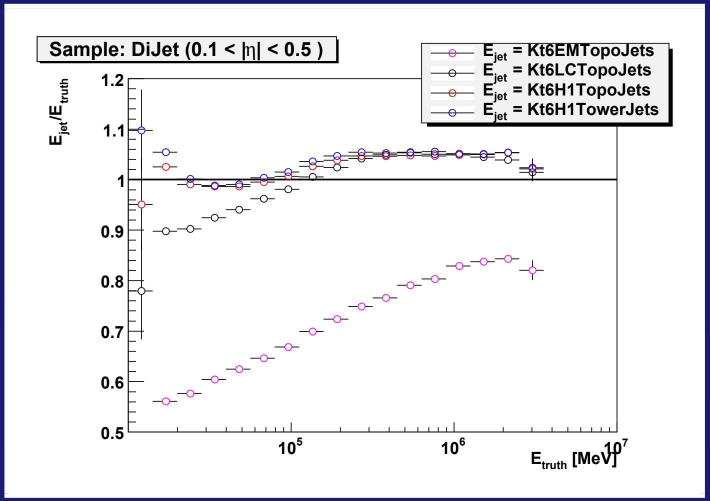
► Refinement of dead material corrections

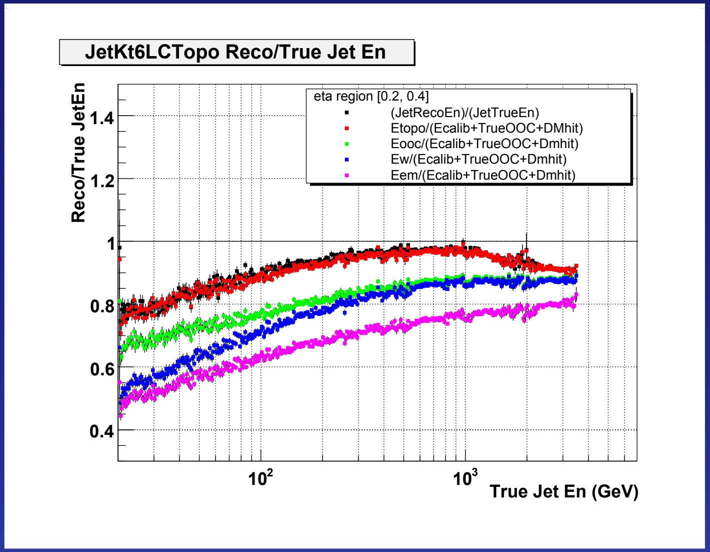
- correct for dead material energy assigned to clusters only
  ► with Gena's assignment of dead material hits to clusters
- treat presamplers as dead-material
  ► simplifies the separation of out-of-cluster and dead-material corrections
- effect of bias in dead-material corrections understood:
  binning in noise containing quantity creates bias for cut > 0
  ► use reco vs. truth instead of truth vs. reco for profile
- make leakage correction explicit instead of implicit

► Expected effects

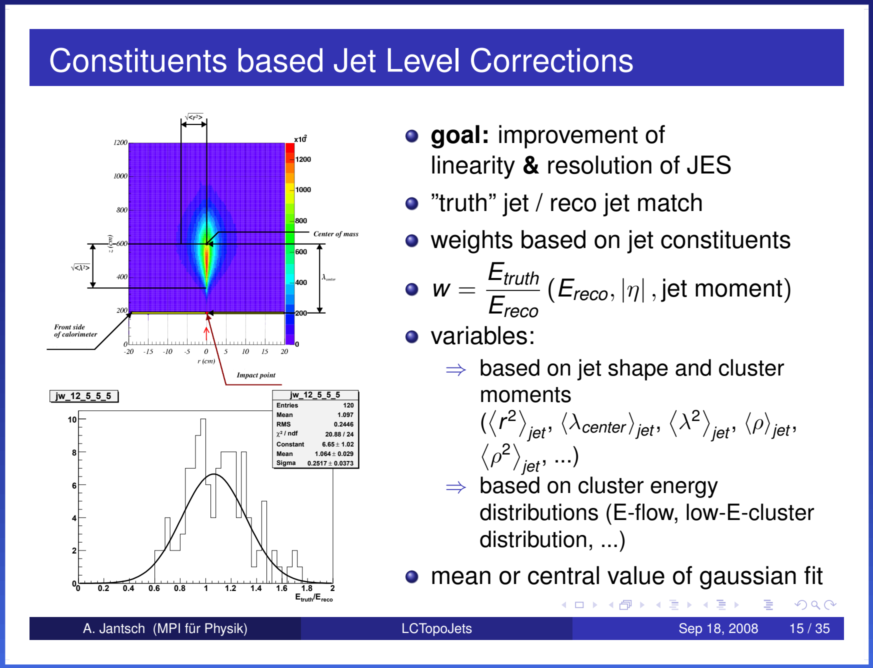
- better weighting performance due to correct simulation
- smaller out-of-cluster and dead-material corrections since we correct for assignable stuff only (Discussion at Ringberg)
- better defined base for jet-level corrections

- **Linearity for truth-jet-matched di-jet events in the barrel**
  - compared are em-scale, (old) local hadron calibrated scale, global H1 weight scale
  - overshoot at high energies due to changed physics list
    - ► solved with new constants
  - undershoot at low energies for local hadron calibration due to lost particles without correlation to calorimeter signals
    - ► confirmed with calibration hit plot from P. Giovannini





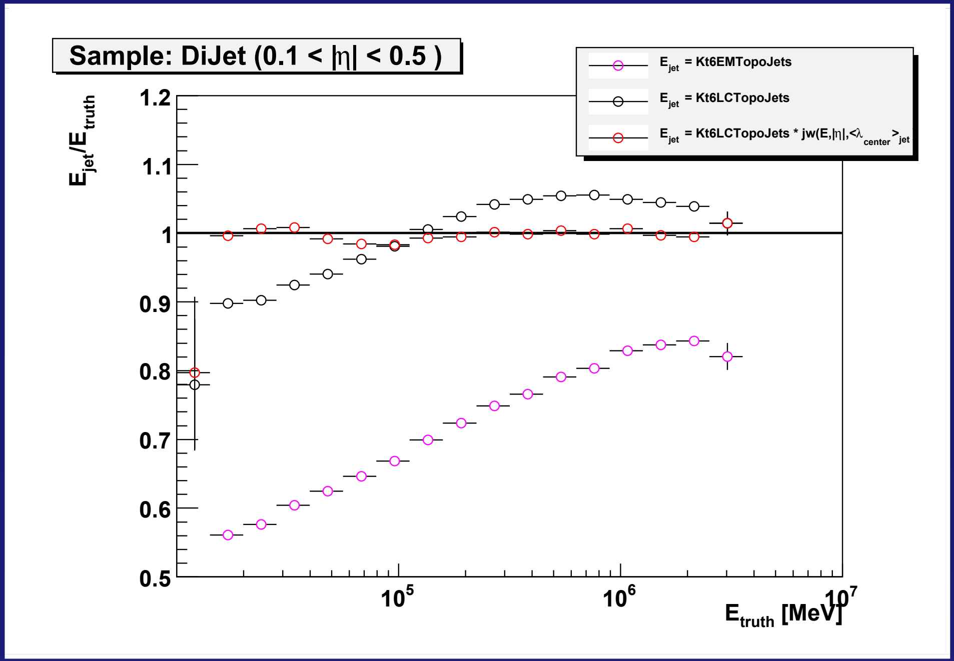
- **Jet-level corrections based on constituents need to be applied**

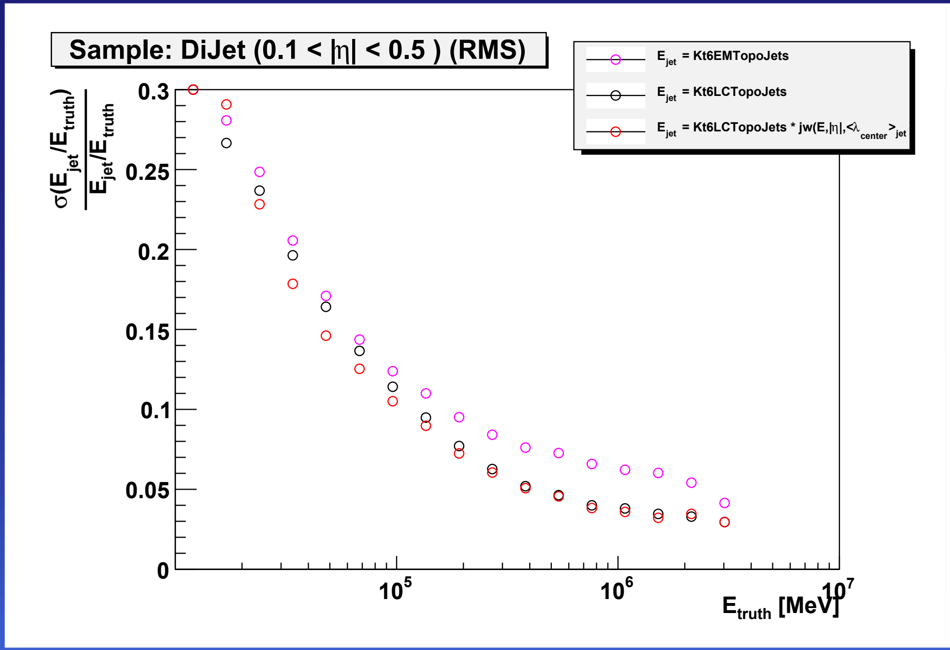

Constituents based Jet Level Corrections

- **goal:** improvement of linearity **&** resolution of JES
- "truth" jet / reco jet match
- weights based on jet constituents
- $w = \frac{E_{truth}}{E_{reco}}(E_{reco}, |\eta|, \text{jet moment})$
- variables:
  - ⇒ based on jet shape and cluster moments $(\langle r^2\rangle_{jet}, \langle\lambda_{center}\rangle_{jet}, \langle\lambda^2\rangle_{jet}, \langle\rho\rangle_{jet}, \langle\rho^2\rangle_{jet}, ...)$
  - ⇒ based on cluster energy distributions (E-flow, low-E-cluster distribution, ...)
- mean or central value of gaussian fit

A. Jantsch (MPI für Physik) LCTopoJets Sep 18, 2008 15 / 35

- **Linearity for truth-jet-matched di-jet events in the barrel**
  - compared are em-scale, (old) local hadron calibrated scale, local hadronic + jet-level-corrected scale
    - ► linearity is restored within 2%
- **Resolution for truth-jet-matched di-jet events in the barrel**
  - again em-scale, (old) local hadron calibrated scale, local hadronic + jet-level-corrected scale
    - ► resolution improves for low jet energies ($E < 100$ GeV) where corrections are largest
- **Code will be in CVS soon**
  - modular to choose different constituent based jet-moment
  - constants will become conditions data





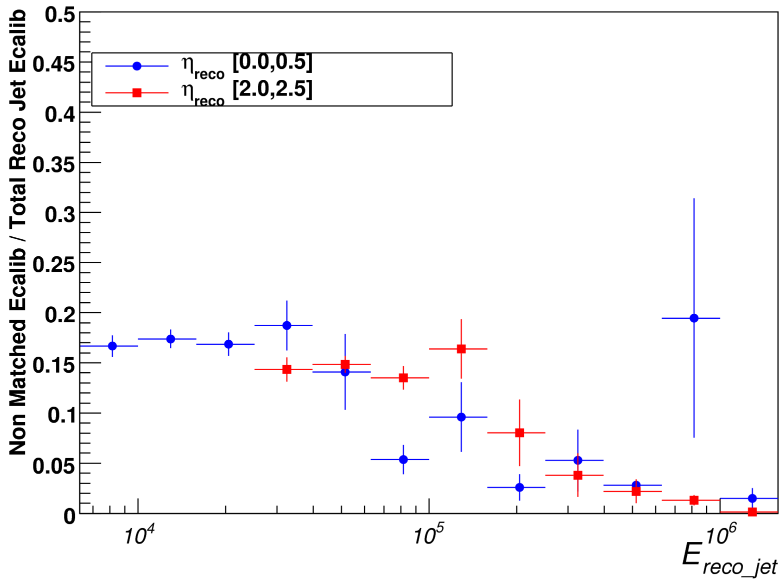
- On my wishlist since a long time:
  - if we'd know exactly which primary generator particle caused which calibration hit, we'd have no ambiguities in jet-truth-matching!
  - deficits in the jet-truth-matching can be made visible
  - actual true expected energy can be derived for clusters, jets, MET, ...
- Gena modified 10 `athena` packages to keep `ParentID` for every `G4Track` ► provides new method `CaloCalibrationHit::particleID()` returning the barcode of the primary particle causing the hit
- Performance price

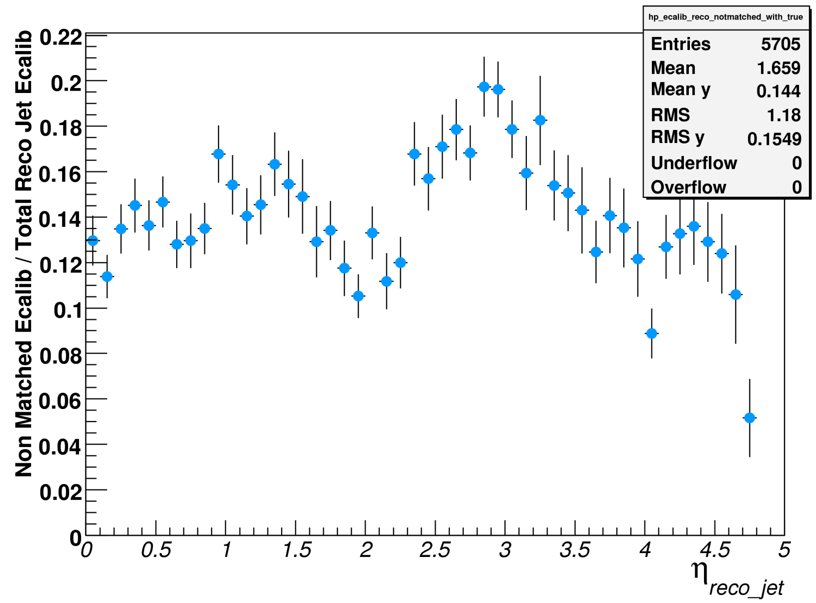
	standard 14.2.21	hits with ParticleID	
average time per event, sec	2139.7 ± 187.6	2320.2 ± 173.1	~ 1.08 ↑
memory per event, Mb	694.1	684.1	~ 1.0
simul size, Mb/event	2.35	5.68	~ 2.4 ↑
av.number of DM hits per event	40770 ± 6277	245500 ± 95160	~ 6.0 ↑
av.number of active+inactive hits	70840 ± 18660	176300 ± 62100	~ 2.5 ↑

J4

► about the same factors for all `JX` samples
► CPU time increases by $\sim 10\%$
► disk size per simulated file increases by $\sim 140\%$ or $3.3\,\mathrm{Mb/ev}$
► disk size per digitized file increases also by $3.3\,\mathrm{Mb/ev}$

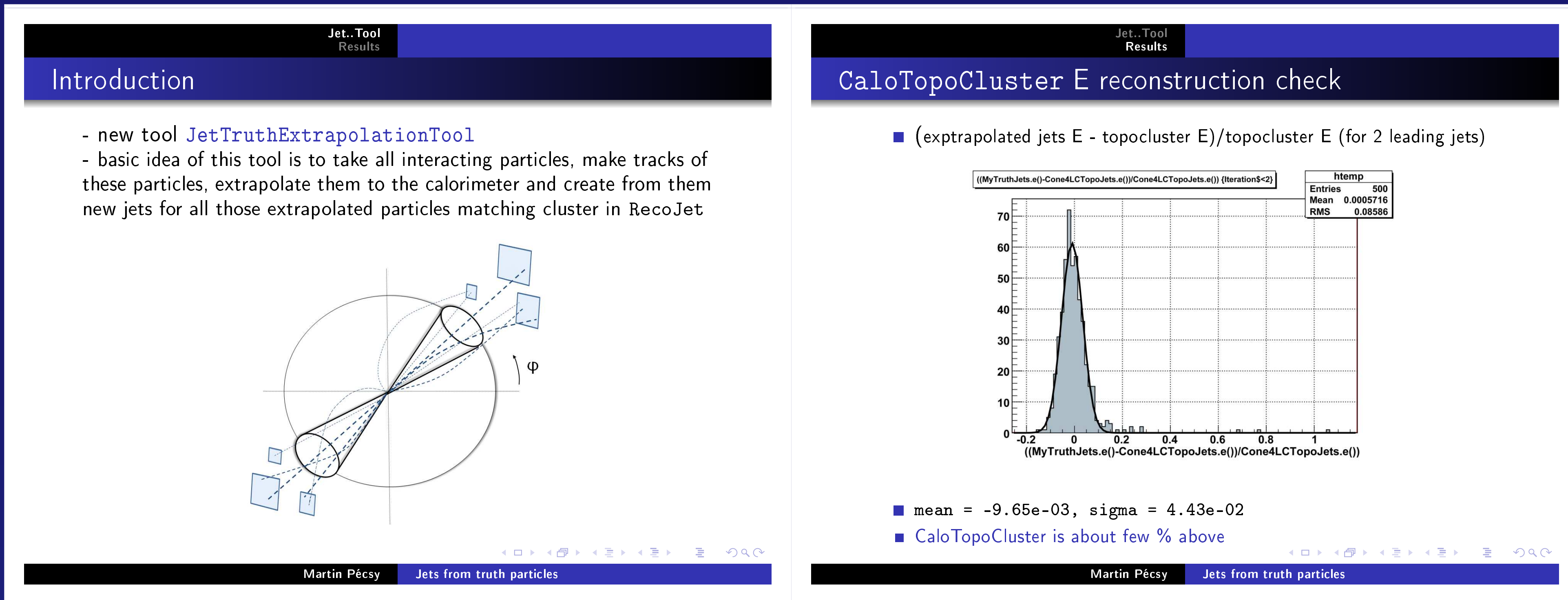
Foreign calibration energy of Cone4LC *reco* matched jet





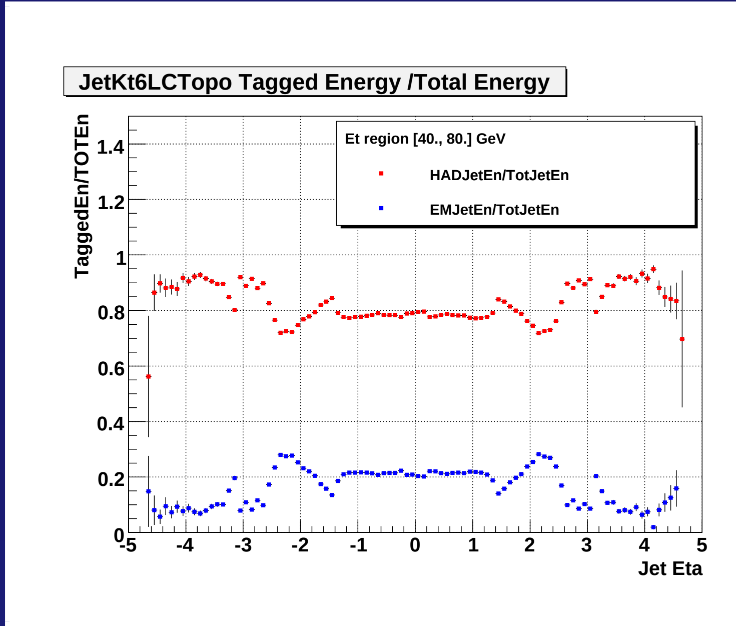
Average ratio of calibration energy inside *reco* jet which doesn't belong to the particles of true matched ($\Delta R<0.3$) jet as a function of energy of reco jet (left), eta of reco jet (right).

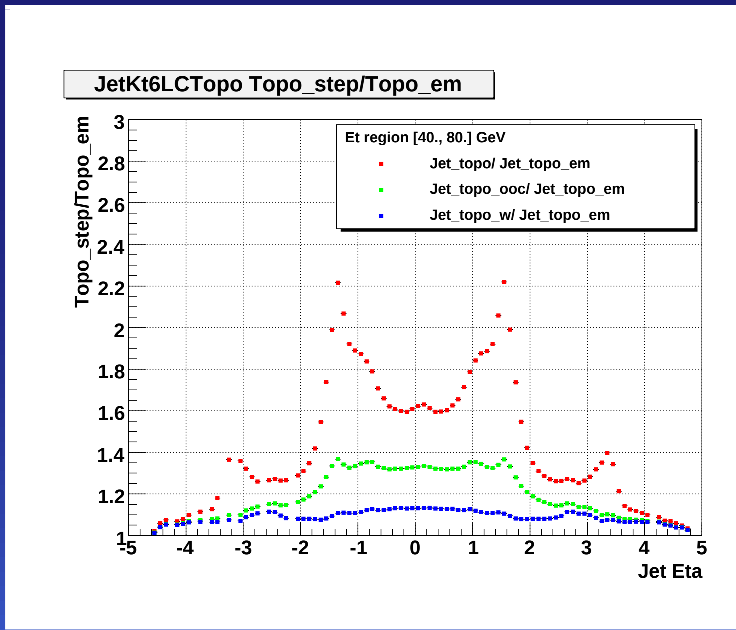
Gennady Pospelov, MPI Munich *hadronic calibration meeting, October 8, 2008* *02*



- Extrapolation tool can be used when no `ParticleID` is present
- Current status
  - code `JetParticleExtrapolationTool` is in CVS in `JetSimTools-00-01-04`
  - will be used in rel `15.0.0` to have less ambigous truth matching

- jet-level checks with early data on QCD events
- take leading jets and plot amount of tagged hadronic and tagged electromagnetic energy per jet vs. η
  - tests the classification
- take leading jets and compare ratio of jet energy after each calibration step over em-scale energy vs.η
  - tests each calibration step separately
- compare with di-jet simulations to provide feedback
  - classification probes impact of deficits in moment description
  - η-structure is sensitive to dead material
  - relative calibration steps probe physics lists
- feedback loop: modify simluation/digitization (sometimes just adding cross-talk as seen in CTB2004 is enough) and compare again ...

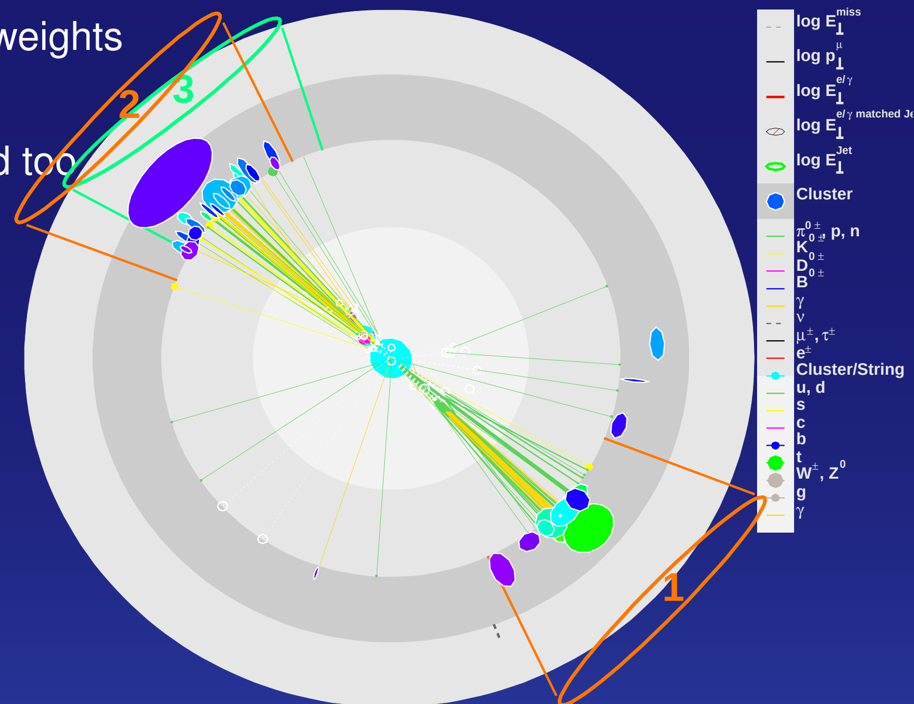




Conclusions

Local Hadron Calibration Constants

- $\sim 6 \cdot 10^6$ single pions ($\pi+$,π^-,π^0) `mc08.10741[012]` have been used to produce new constants
  - uploaded in conditions database (thanks Walter!)
- checks of physics lists and moments done on CTB2004 H6 and H8 (P. Speckmayer & J. Erdmann)
  - continued by A. Kiryunin
- production of constants is now completely `athena` based in package `CaloLocalHadCalib`
- some small refinements for classification and weights
- cleaner definition for out-of-cluster corrections
- dead-material corrections will soon be updated too



Jet-level corrections

- proposed new package from A. Jantsch
- corrects for losses not directly correlated with individual constituents
- restores linearity and improves resolution

New Algorithms

- code by G. Pospelov to attach `ParticleID` to `CalibrationHit` exists
  - is in CVS already and can be used in release `15.0.0` during simulation with `CalibrationHits`
- even without `CalibrationHits` and `ParticleID` improved truth matching with code from M. Pecsy

Early Data

- modularity of calibration approach leads to good data/MC tests (P. Giovannini)
- application to top-physics (see talk by T. Barillari) for in-situ performance checks