# Topological cell clustering and making jets in ATLAS

Amal Vaidya

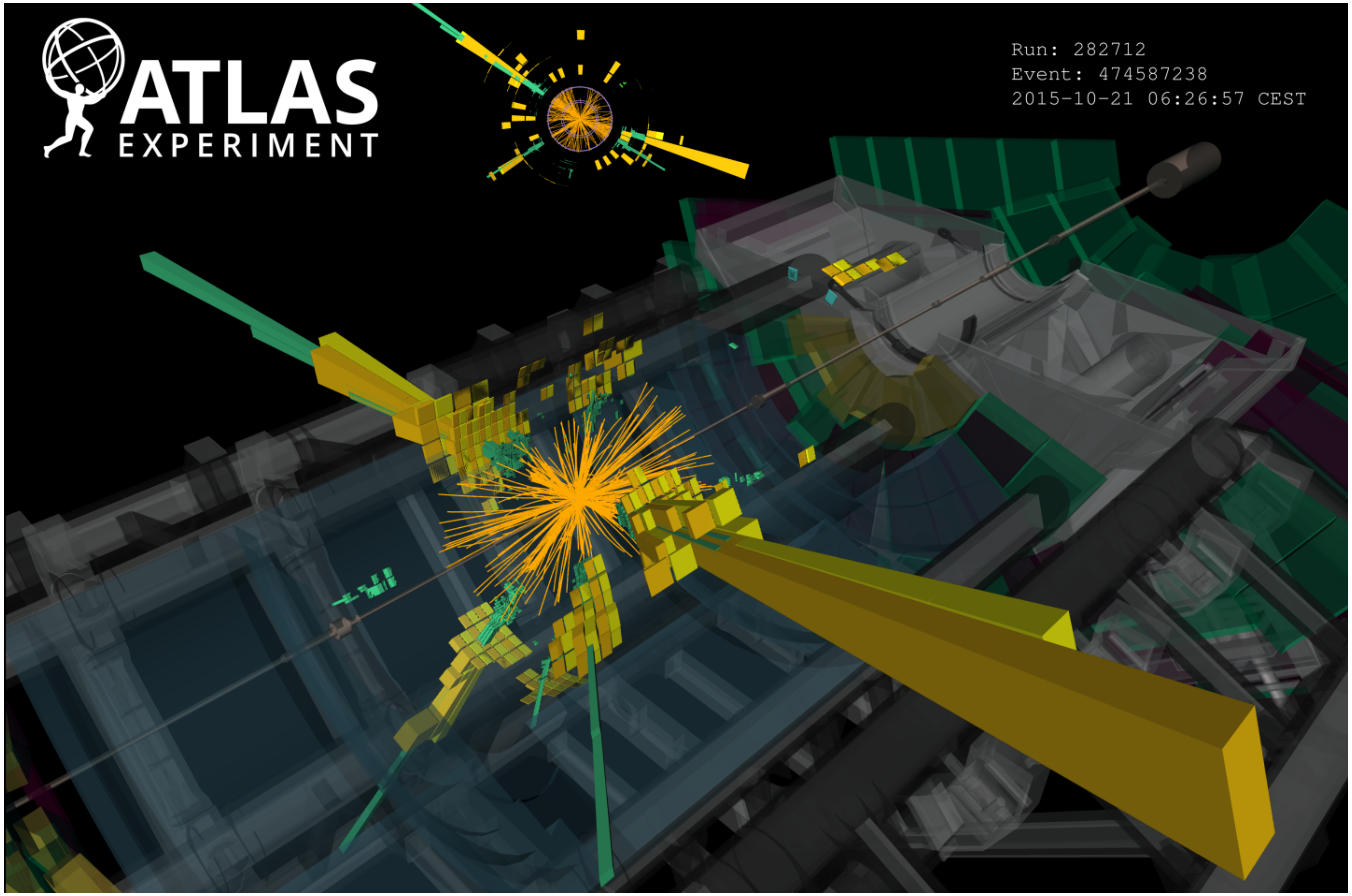
UCL ATLAS meeting
28 Oct 2016

bit of a work in progress…

# Topoclusters: From calo cells to jet constituents

*topological cell clusters - 3d clusters of calorimeter cells*

“an attempt to extract significant signal from from a background of electronic noise and other sources of fluctuations such as pile-up”

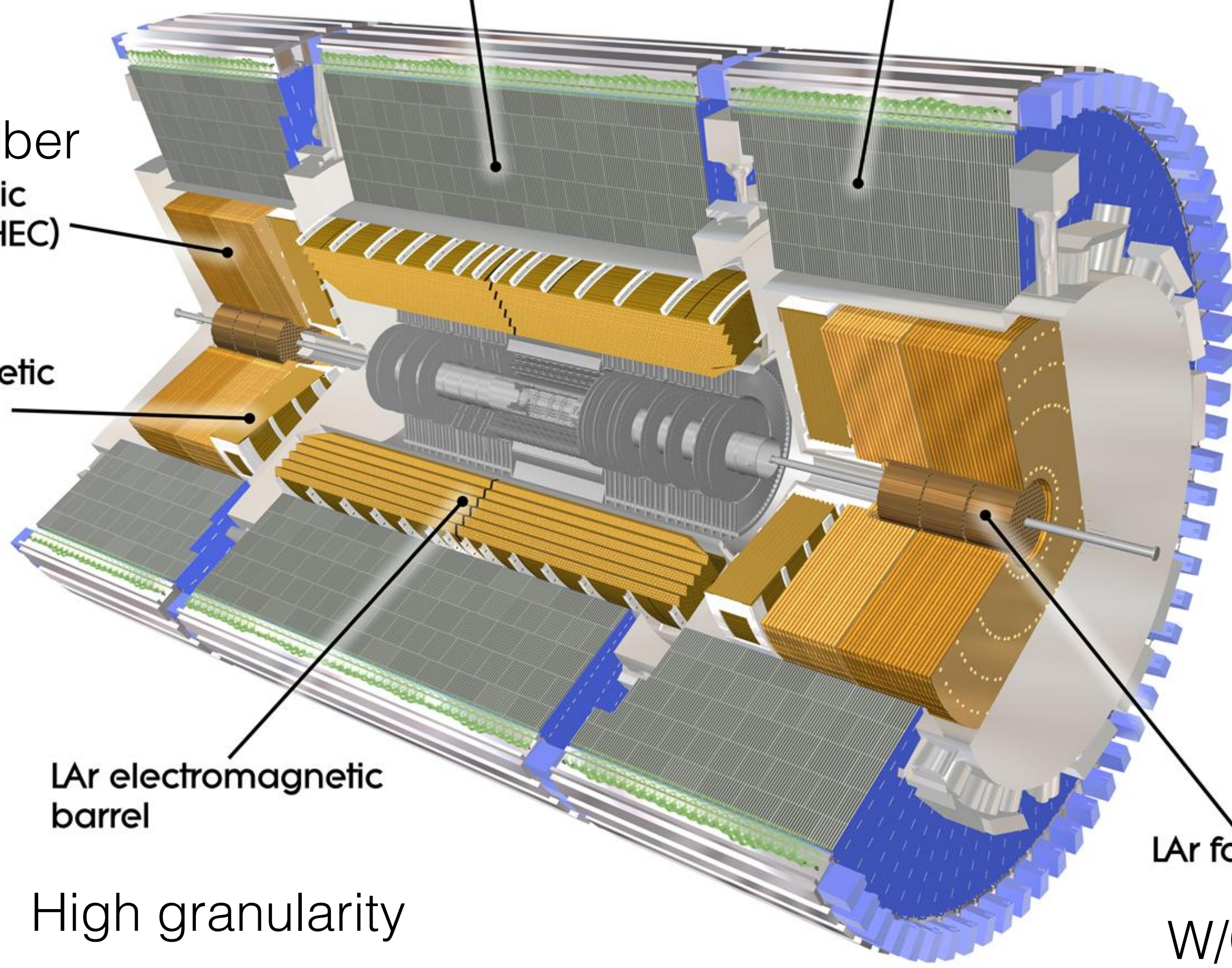


# The ATLAS calorimeter system

Several subsystems with different granularities and technologies

Non-compensating sampling calorimeters: Passive material to cause showering, active material measures energies

Steel absorber and scintillating tiles as active material




Copper absorber

LAr Pre-sampler corrects for energy loss by electrons and photons upstream in the calorimeter

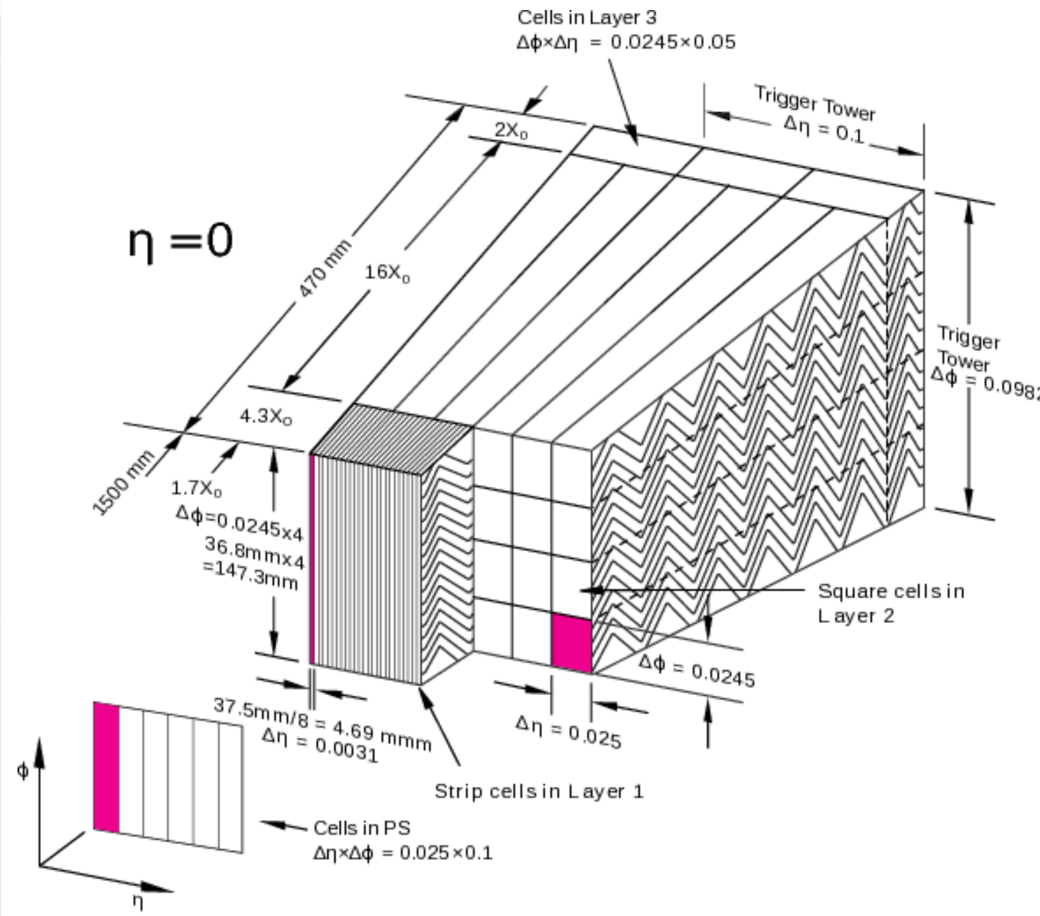
Should provide good containment for EM and hadronic showers: avoid punch through

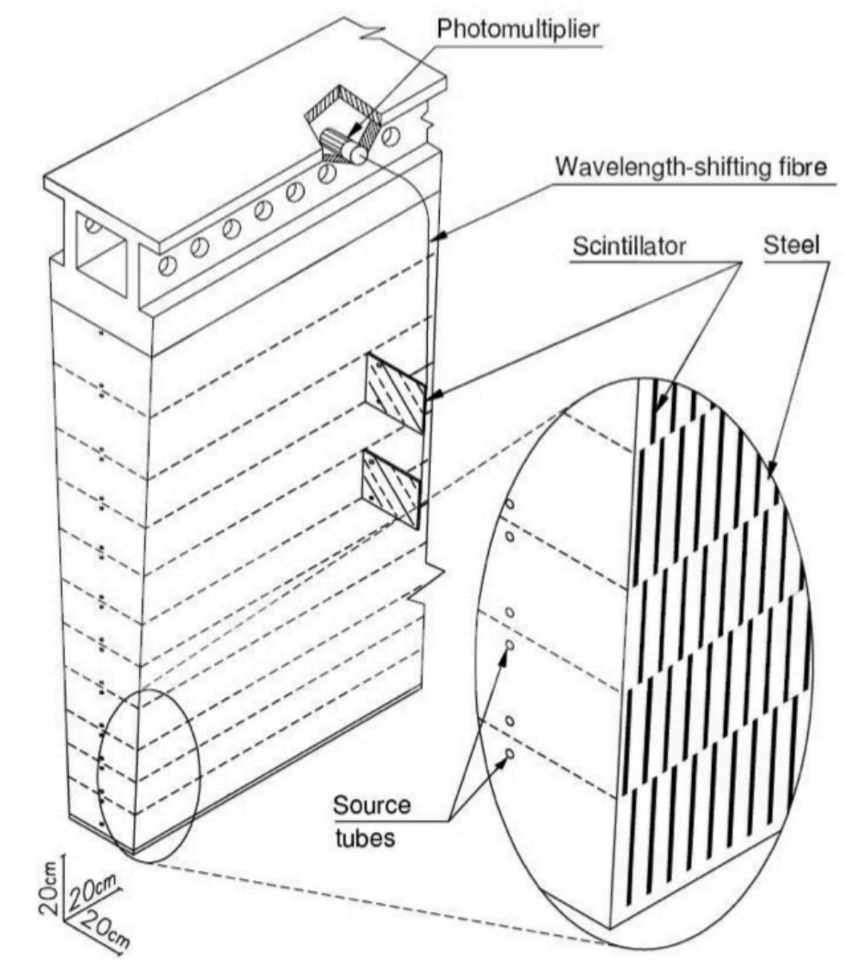
High granularity

W/Cu absorber

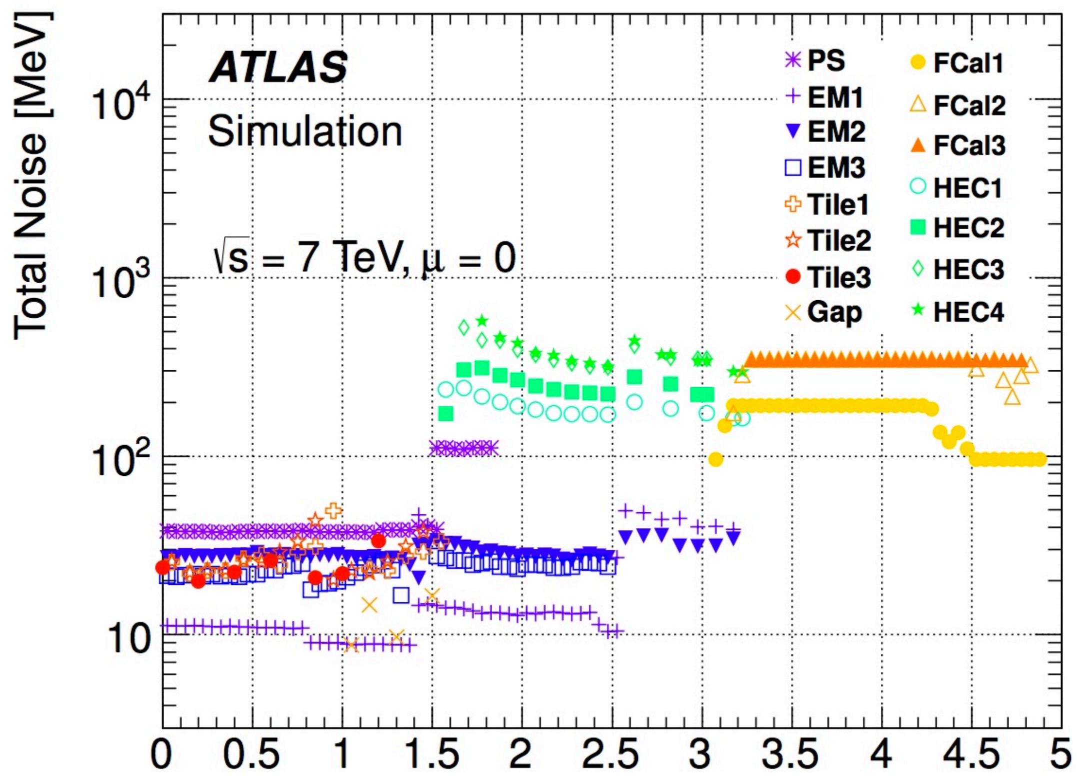
| Calorimeter | Module | Sampling ($S_{calo}$) | $N_{cells}$ | $\eta$-coverage | $\Delta\eta \times \Delta\phi$ |
|---|---|---|---|---|---|
| *Electromagnetic calorimeters* | EMB | | 109 568 | $\lvert\eta\rvert < 1.52$ | |
| | | PreSamplerB | 7 808 | $\lvert\eta\rvert < 1.52$ | $0.025 \times \pi/32$ |
| | | EMB1 | | $\lvert\eta\rvert < 1.4$ | $0.025/8 \times \pi/32$ |
| | | | | $1.4 < \lvert\eta\rvert < 1.475$ | $0.025 \times \pi/128$ |
| | | EMB2 | | $\lvert\eta\rvert < 1.4$ | $0.025 \times \pi/128$ |
| | | | | $1.4 < \lvert\eta\rvert < 1.475$ | $0.075 \times \pi/128$ |
| | | EMB3 | | $\lvert\eta\rvert < 1.35$ | $0.050 \times \pi/128$ |
| | EMEC | | 63 744 | $1.375 < \lvert\eta\rvert < 3.2$ | |
| | | PreSamplerE | 1 536 | $1.5 < \lvert\eta\rvert < 1.8$ | $0.025 \times \pi/32$ |
| | | EME1 | | $1.375 < \lvert\eta\rvert < 1.425$ | $0.050 \times \pi/32$ |
| | | | | $1.425 < \lvert\eta\rvert < 1.5$ | $0.025 \times \pi/32$ |
| | | | | $1.5 < \lvert\eta\rvert < 1.8$ | $0.025/8 \times \pi/32$ |
| | | | | $1.8 < \lvert\eta\rvert < 2.0$ | $0.025/6 \times \pi/32$ |
| | | | | $2.0 < \lvert\eta\rvert < 2.4$ | $0.025/4 \times \pi/32$ |
| | | | | $2.4 < \lvert\eta\rvert < 2.5$ | $0.025 \times \pi/32$ |
| | | | | $2.5 < \lvert\eta\rvert < 3.2$ | $0.1 \times \pi/32$ |
| | | EME2 | | $1.375 < \lvert\eta\rvert < 1.425$ | $0.050 \times \pi/128$ |
| | | | | $1.425 < \lvert\eta\rvert < 2.5$ | $0.025 \times \pi/128$ |
| | | | | $2.5 < \lvert\eta\rvert < 3.2$ | $0.1 \times \pi/128$ |
| | | EME3 | | $1.5 < \lvert\eta\rvert < 2.5$ | $0.050 \times \pi/128$ |
| *Hadronic calorimeters* | Tile (barrel) | | 2 880 | $\lvert\eta\rvert < 1$ | |
| | | TileBar0/1 | | | $0.1 \times \pi/32$ |
| | | TileBar2 | | | $0.2 \times \pi/32$ |
| | Tile (extended barrel) | | 2 304 | $0.8 < \lvert\eta\rvert < 1.7$ | |
| | | TileExt0/1 | | | $0.1 \times \pi/32$ |
| | | TileExt2 | | | $0.2 \times \pi/32$ |
| | HEC | | 5 632 | $1.5 < \lvert\eta\rvert < 3.2$ | |
| | | HEC0/1/2/3 | | $1.5 < \lvert\eta\rvert < 2.5$ | $0.1 \times \pi/32$ |
| | | | | $2.5 < \lvert\eta\rvert < 3.2$ | $0.2 \times \pi/16$ |
| *Forward calorimeters* | FCAL | | 3 524 | $3.1 < \lvert\eta\rvert < 4.9$ | $\Delta x \times \Delta y$ |
| | | FCAL0 | | $3.1 < \lvert\eta\rvert < 3.15$ | 1.5 cm × 1.3 cm |
| | | | | $3.15 < \lvert\eta\rvert < 4.3$ | 3.0 cm × 2.6 cm |
| | | | | $4.3 < \lvert\eta\rvert < 4.83$ | 1.5 cm × 1.3 cm |
| | | FCAL1 | | $3.2 < \lvert\eta\rvert < 3.24$ | 1.7 cm × 2.1 cm |
| | | | | $3.24 < \lvert\eta\rvert < 4.5$ | 3.3 cm × 4.2 cm |
| | | | | $4.5 < \lvert\eta\rvert < 4.81$ | 1.7 cm × 2.1 cm |
| | | FCAL2 | | $3.29 < \lvert\eta\rvert < 3.32$ | 2.7 cm × 2.4 cm |
| | | | | $3.32 < \lvert\eta\rvert < 4.6$ | 5.4 cm × 4.7 cm |
| | | | | $4.6 < \lvert\eta\rvert < 4.75$ | 2.7 cm × 2.4 cm |

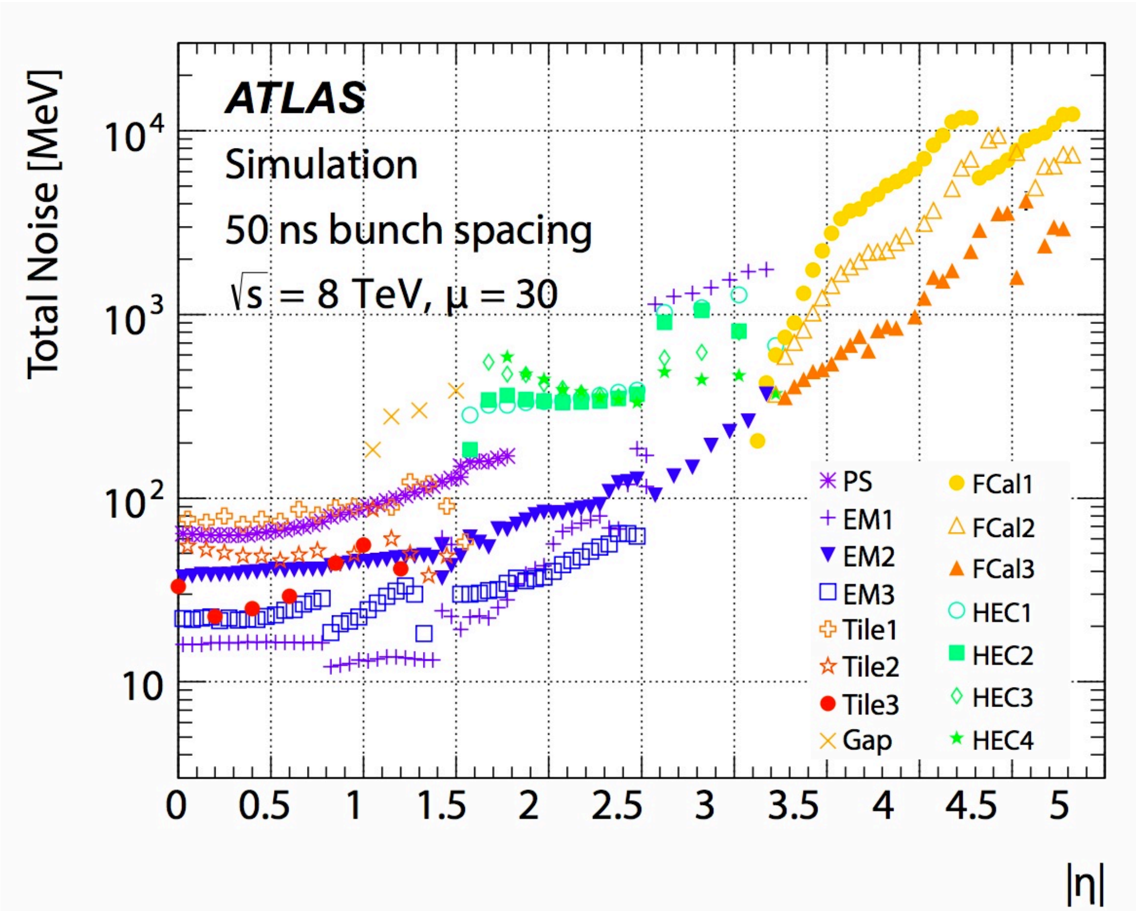
Total 188000 read-out channels!!!

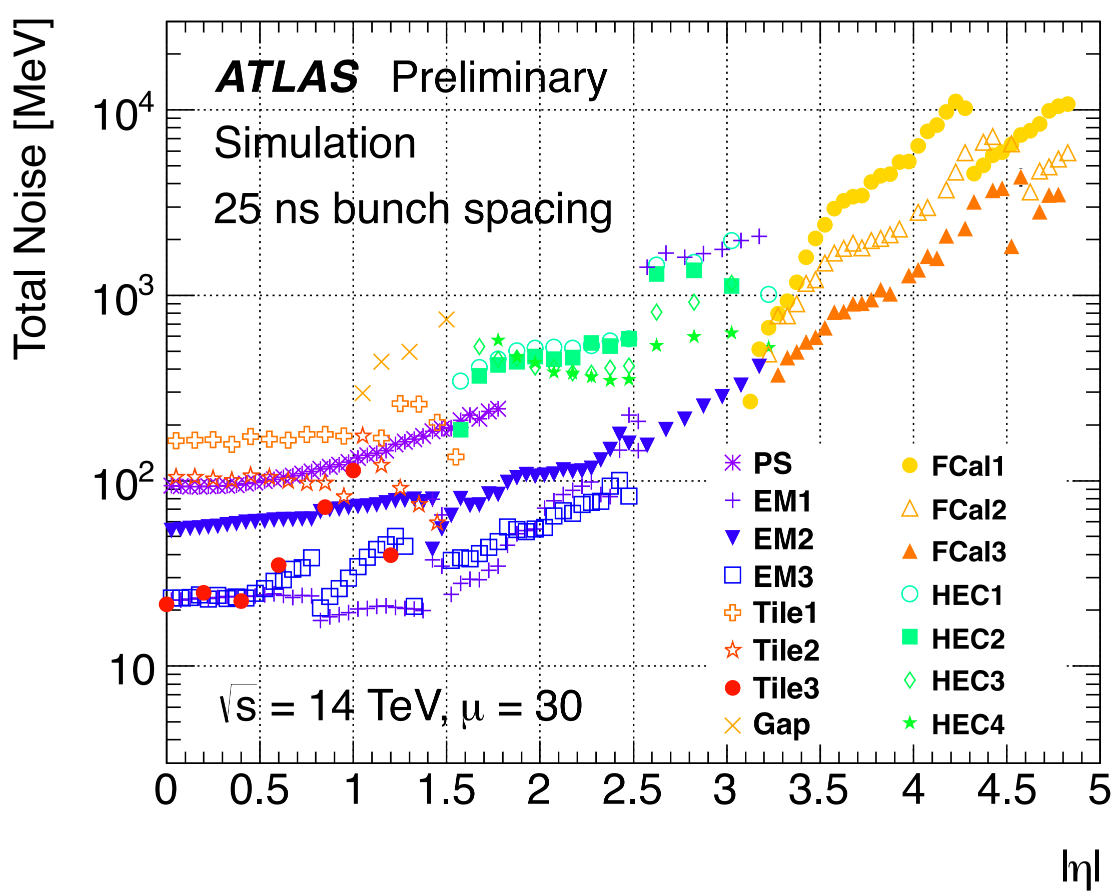


# Noise in the calorimeter



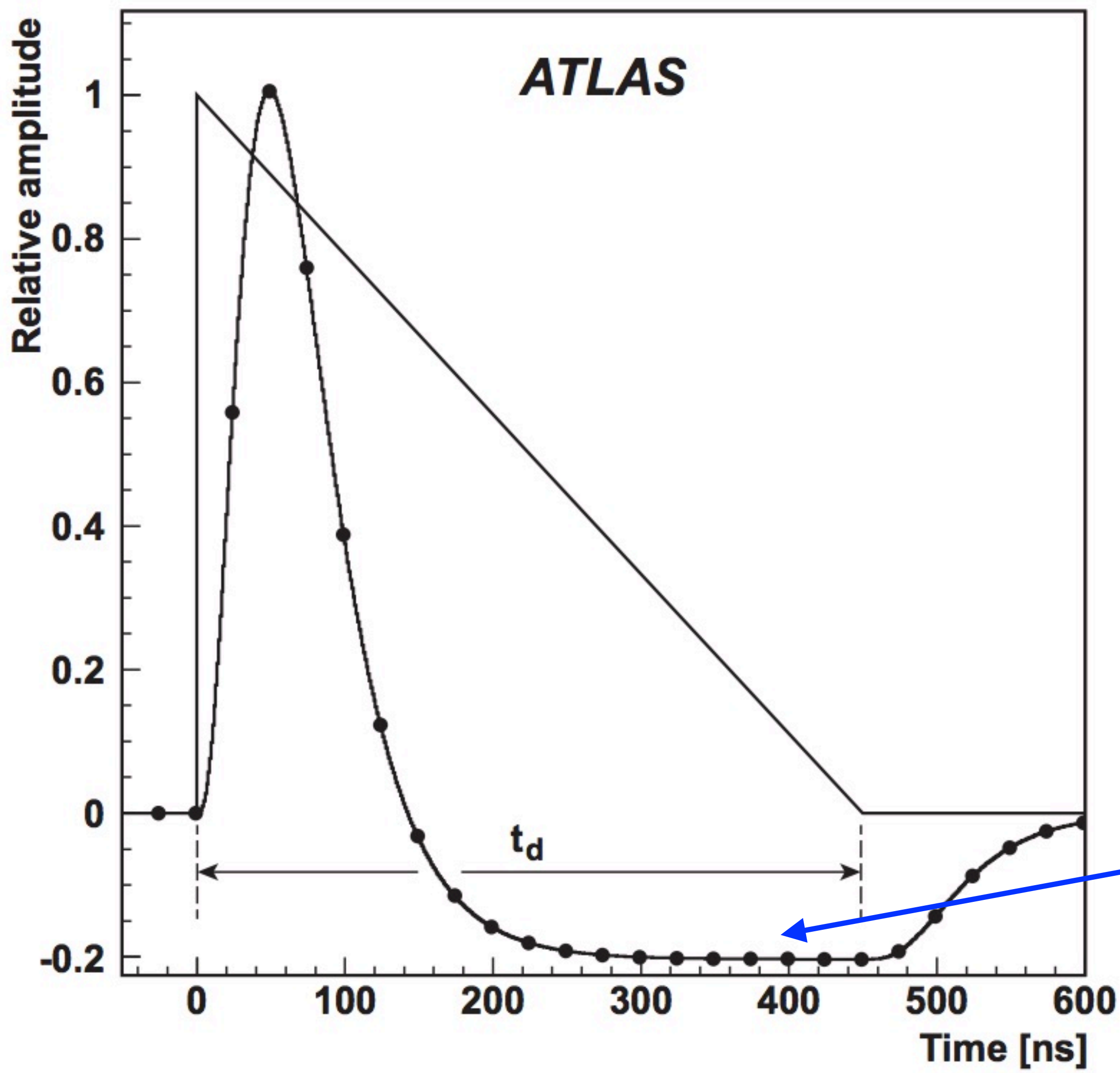




Fluctuations in the cell energies due to electronic noise and extra signals which are induced by pile up

# LAr and pileup noise

Shape of pulse in LAr helps suppress pileup noise!




- Bunch spacing of 25 ns
- Bunches come in "trains" which contain a fixed number of bunches

Out of time pileup: pile up from pre or proceeding bunches contributes to the signal and adds extra noise

The bipolar shape of the signal response can reduce the total effects of pileup as the integral can be net-zero over time.

contribution here stable from bunch-to-bunch fluctuations

leads to cancellation on average of in-time pile-up signal contributions by out-of-time pile-up signal residuals in any given calorimeter cell. (cell can have negative energies)

was optimised for 25 ns bunch spacings: in terms of sampling frequency and amplifier design

However, large fluctuations in the number of interactions per bunch crossing and the difference in energy flow

Tile far less sensitive to out of time pile up

# Topoclustering

## How do we cluster together cells in a sensible way?

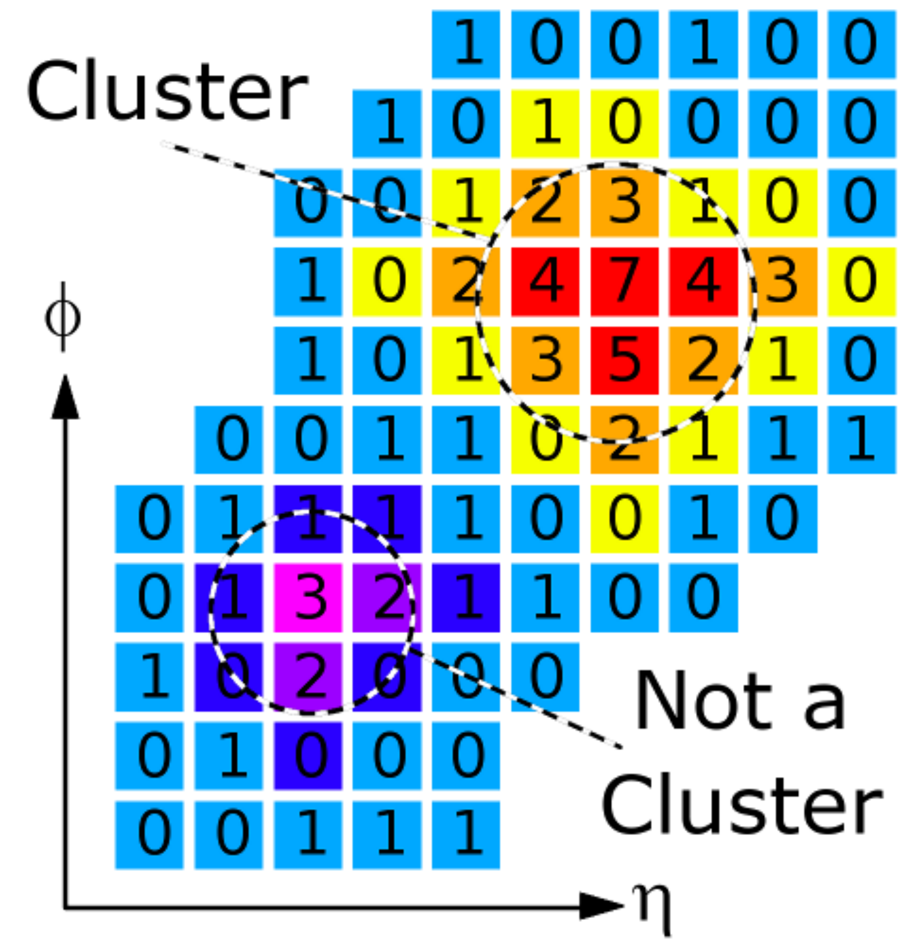
Define the significance of each cell: $$S_{\text{cell}} = \frac{E_{\text{cell}}}{\sigma_{\text{noise,cell}}}$$

Define three threshold parameters: {S,N,P}, in terms of the significance

- S: seed threshold, cells with this a significance > S, default = 4
- N: Neighbour threshold: cells with significance > N, default = 2
- P: Principle cell filter, Default = 0

The algorithm identifies seed cells based on the above threshold (ranked in order of decreasing significance ) each forms a *proto-cluster*

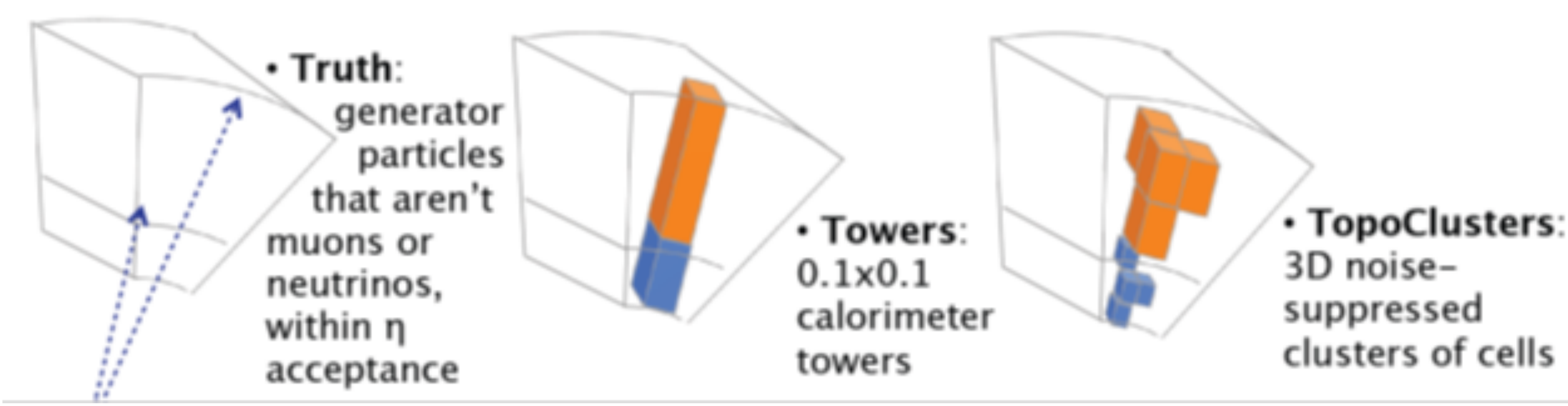
Neighbouring cells are then added to the protocluster if they satisfy The N or P conditions. This stops then the last set of neighbouring cells match the P threshold




proto-clusters which are connected by neighbouring cells satisfying S or N are merged.

Clusters can also be split if they are found to contain local maxima

# Noise suppression - negative energy cells




Implicit noise suppression in this process

**Cells can have negative energy (Bipolar LAr response)**

In determining significance, the absolute value of energy is taken:
Cells with $E < 0$ can contribute to cluster seeding and growth!

A negative seed cell is typically generated by out of time pile up 100 ns before the event and can seed negative clusters
While out of time pile up closer to the event typically generates cells with $E > 0$

Therefore clustering cells with negative energy tends to cancel the random upward fluctuation of some out-of-time pile up -> allowing only positive signals introduces a bias

When constructing physics objects such as jets, only clusters with $E > 0$ are used, since there is not much correlated expected between the energy flow of different bunch crossings

# Topoclustering - Cluster multiplicities

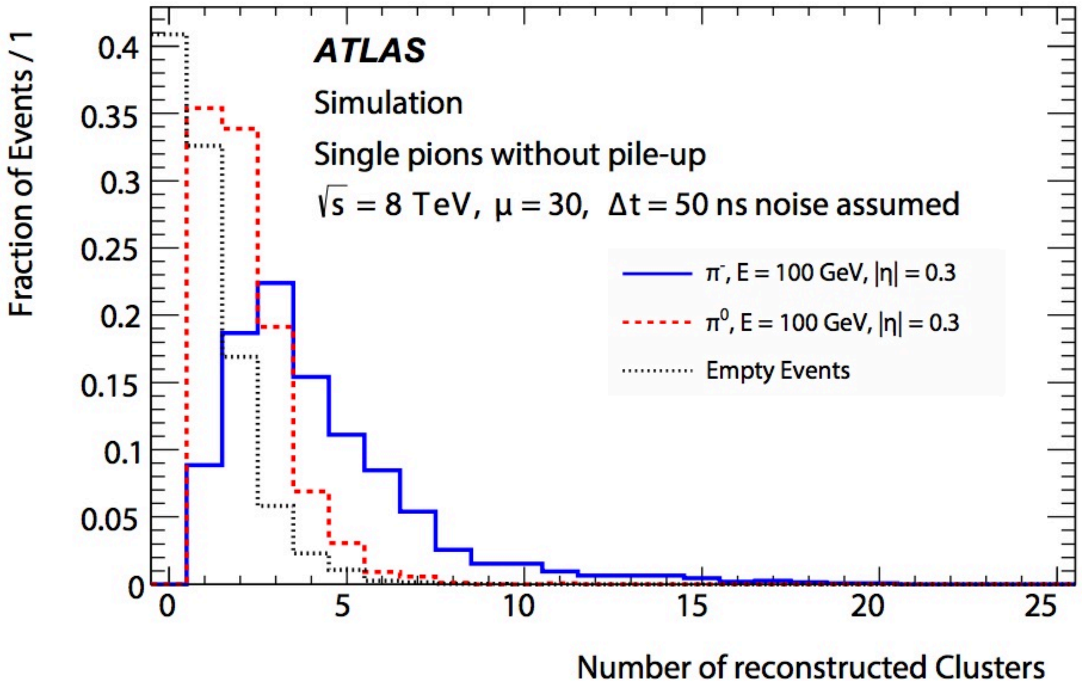
It would be nice to assume that a single cluster corresponds to a single particle shower

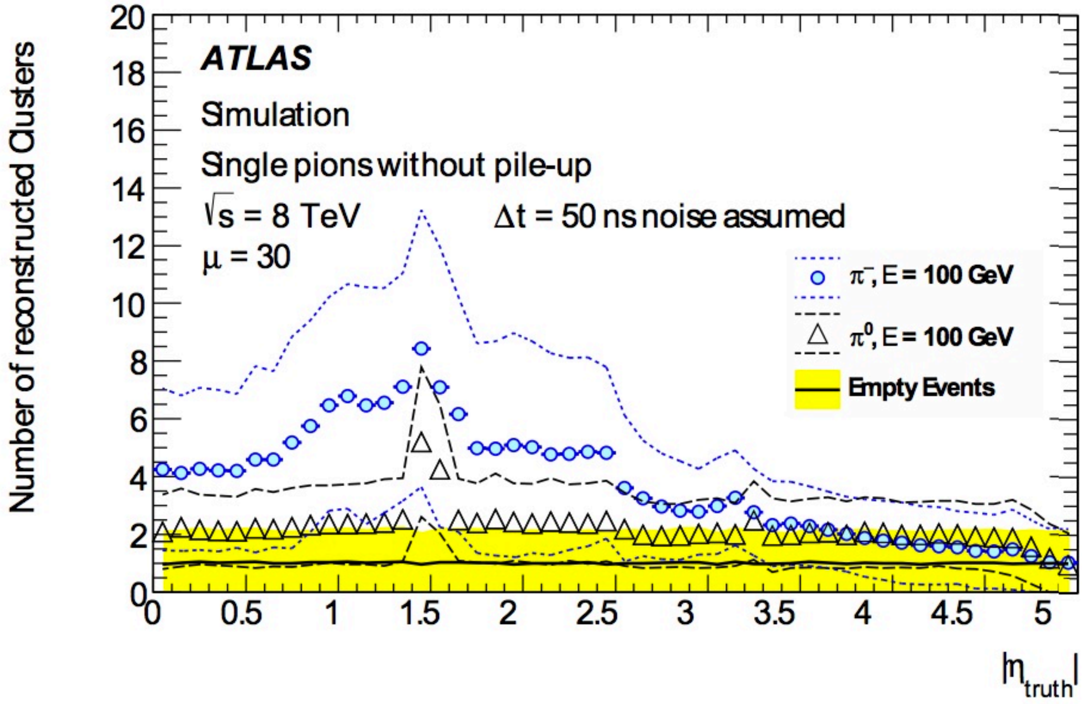
Tends to be true for particles which generate predominantly electromagnetic showers: The angular separation between photons from neutral pion decays is small

Large peak at even for neutral pions:

- incorrect proto-cluster splitting
- Resolving the two photon showers
- Additional clusters generated by noise

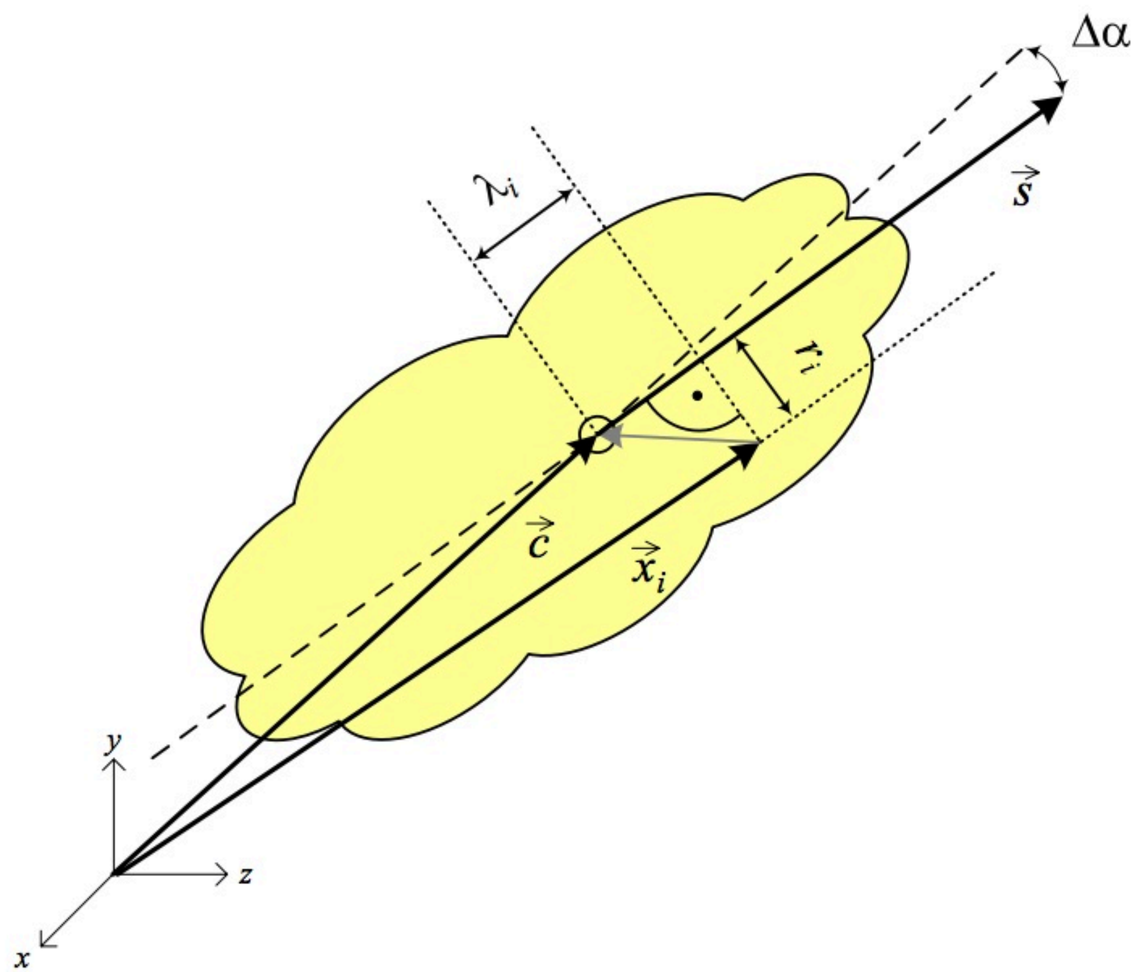
Charged pions typically induce multiple clusters which can be well separated due to inelastic interactions



# Topocluster Moments

Huge amount of information can be extracted based on shape and energy distribution within a cluster




- $\vec{c}$ *centre of gravity of cluster, measured from the nominal vertex* $(x = 0, y = 0, z = 0)$ *in ATLAS*
- $\vec{x}_i$ *geometrical centre of a calorimeter cell in the cluster, measured from the nominal detector centre of ATLAS*
- $\vec{s}$ *particle direction of flight (shower axis)*
- $\Delta\alpha$ *angular distance* $\Delta\alpha = \angle(\vec{c}, \vec{s})$ *between cluster centre of gravity and shower axis* $\vec{s}$
- $\lambda_i$ *distance of cell at* $\vec{x}_i$ *from the cluster centre of gravity measured along shower axis* $\vec{s}$ $(\lambda_i < 0$ *is possible)*
- $r_i$ *radial (shortest) distance of cell at* $\vec{x}_i$ *from shower axis* $\vec{s}$ $(r_i \geq 0)$

Information about distribution of energy, cell significance and topological isolation is also extracted

Important for quality and signal type classification

# Topoclustering - Cluster kinematics

## Physics objects have 4 vectors!!

Important to account for negative cells:

Potential to have cluster position allocated to the wrong hemisphere:
Removing negative cells adds a bias towards upward noise fluctuations

Geometrical weights are calculated for each cell based on their separation from the cluster centre of mass

$$\eta_{\mathrm{clus}} = \frac{\sum_{i=1}^{N_{\mathrm{cell}}} w_{\mathrm{cell},i}^{\mathrm{geo}} \cdot |E_{\mathrm{cell},i}^{\mathrm{EM}}| \cdot \eta_{\mathrm{cell},i}}{\sum_{i=1}^{N_{\mathrm{cell}}} w_{\mathrm{cell},i}^{\mathrm{geo}} \cdot |E_{\mathrm{cell},i}^{\mathrm{EM}}|}$$

$$\phi_{\mathrm{clus}} = \frac{\sum_{i=1}^{N_{\mathrm{cell}}} w_{\mathrm{cell},i}^{\mathrm{geo}} \cdot |E_{\mathrm{cell},i}^{\mathrm{EM}}| \cdot \phi_{\mathrm{cell},i}}{\sum_{i=1}^{N_{\mathrm{cell}}} w_{\mathrm{cell},i}^{\mathrm{geo}} \cdot |E_{\mathrm{cell},i}^{\mathrm{EM}}|}$$

$$E_{\mathrm{clus}}^{\mathrm{EM}} = \sum_{i=1}^{N_{\mathrm{cell}}} w_{\mathrm{cell},i}^{\mathrm{geo}} E_{\mathrm{cell},i}^{\mathrm{EM}}$$

Now, topoclusters are considered massless pseudo-particles:

Distributions of cell energy is mostly lateral and doesn't represent a physically meaningful mass: it's also very sensitive to the noise thresholds used.

It's not correct to simply relate a single cluster to the response of a single particle

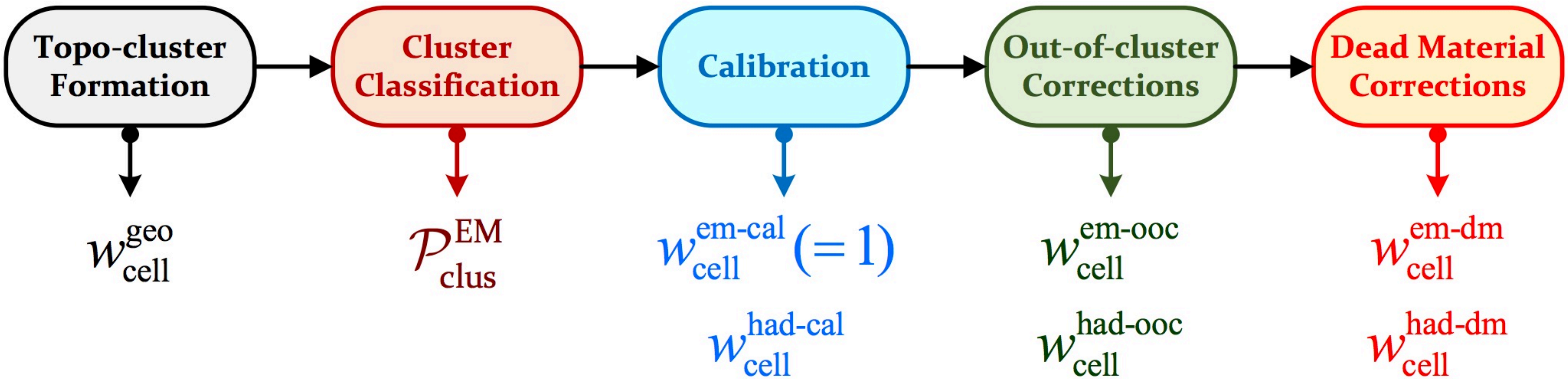
# Classification, calibration, correction

## Not all clusters are made equal…

Local calibration weights (LCW) strategy: uses topocluster moment information to correct for:

- Non-compensating calorimeter response: response for hadrons is smaller for e of same energy
- Signal losses due to clustering
- Signal losses due to energy lost in inactive material

Tested using MC simulations of single pions: required good detector modelling




Crucial to mainting a high resolution of missing transverse energy as well as the transverse momentum of jets

# Cluster classification

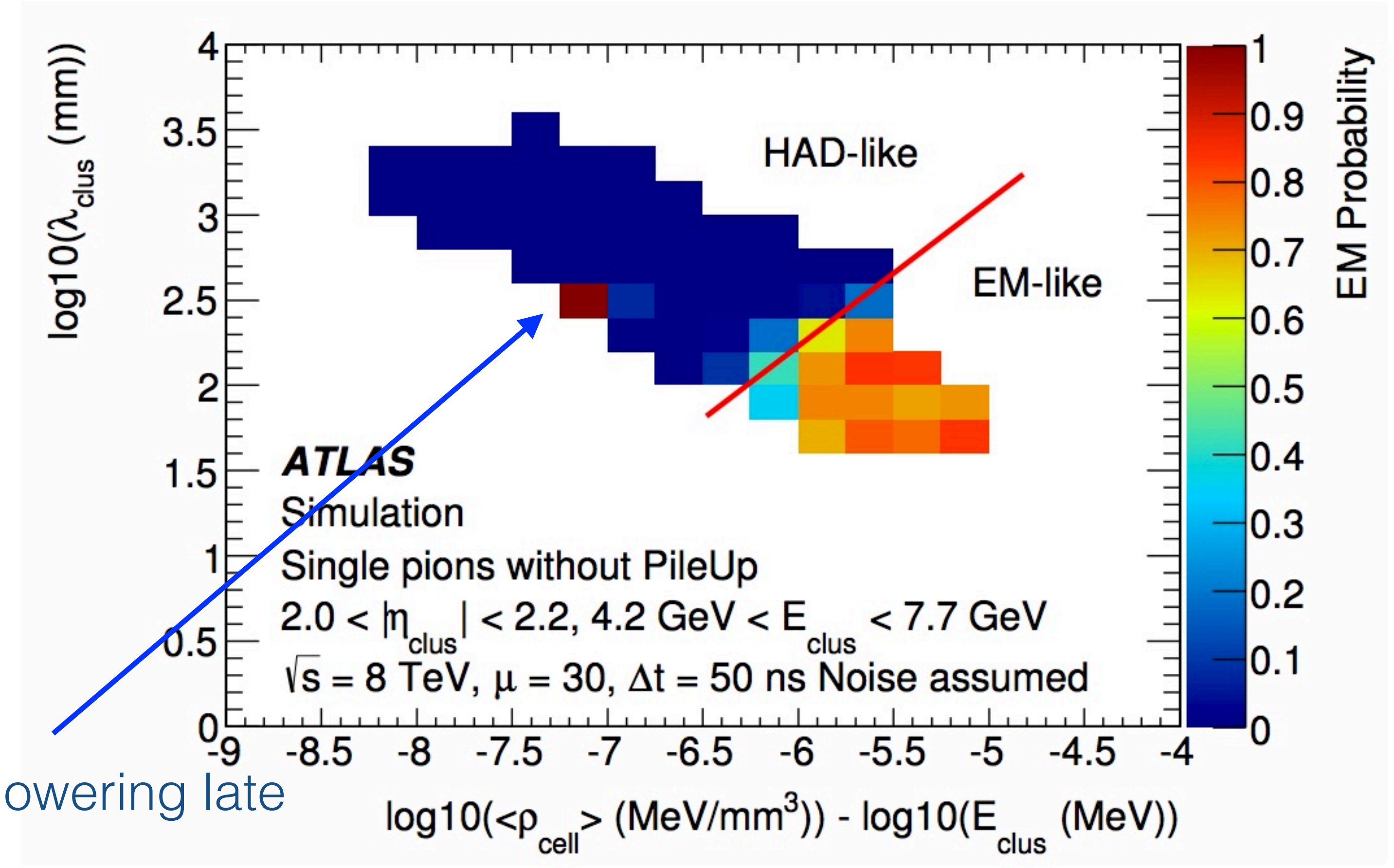
## Calculate the likelihood that the cluster is generated by EM energy deposit

"electromagnetic showers with their compact shower development, early starting point and relatively small intrinsic fluctuations can generate cluster characteristics very different from those generated by hadronic showers."

Hadronic showers:

- tend to develop deeper in the calorimeter
- Have large shower to shower fluctuations
- Have a larger variation of starting points in the calorimeter

The depth (λclus) of the cluster and the average cell signal density (ρclus), determined in bins of the cluster energy and eta have been found to be powerful discriminatory variables




Neutral pions showering late

$$\mathfrak{O}^{\text{class}}_{\text{clus}} = \left\{ E^{\text{EM}}_{\text{clus}},\ \eta_{\text{clus}},\ \log_{10}(\rho_{\text{clus}}/\rho_0) - \log_{10}(E^{\text{EM}}_{\text{clus}}/E_0), \log_{10}(\lambda_{\text{clus}}/\lambda_0) \right\}$$

4 variable discriminant

$$\mathcal{P}^{\text{EM}}_{\text{clus},ijkl} = \frac{\varepsilon^{\pi^0}_{ijkl}}{\varepsilon^{\pi^0}_{ijkl} + 2\varepsilon^{\pi^-}_{ijkl}}$$

likelihood in each bin based on pion efficiency

$$w^{\text{cal}}_{\text{cell}} = \mathcal{P}^{\text{EM}}_{\text{clus}} \cdot w^{\text{em-cal}}_{\text{cell}} + (1 - \mathcal{P}^{\text{EM}}_{\text{clus}}) \cdot w^{\text{had-cal}}_{\text{cell}}$$

# Hadronic calibration

Attempt to establish an average of e/π = 1 for cluster signal

$$w_{\text{cell}} = \frac{E^{\text{dep}}_{\text{cell}}}{E^{\text{EM}}_{\text{cell}}}$$

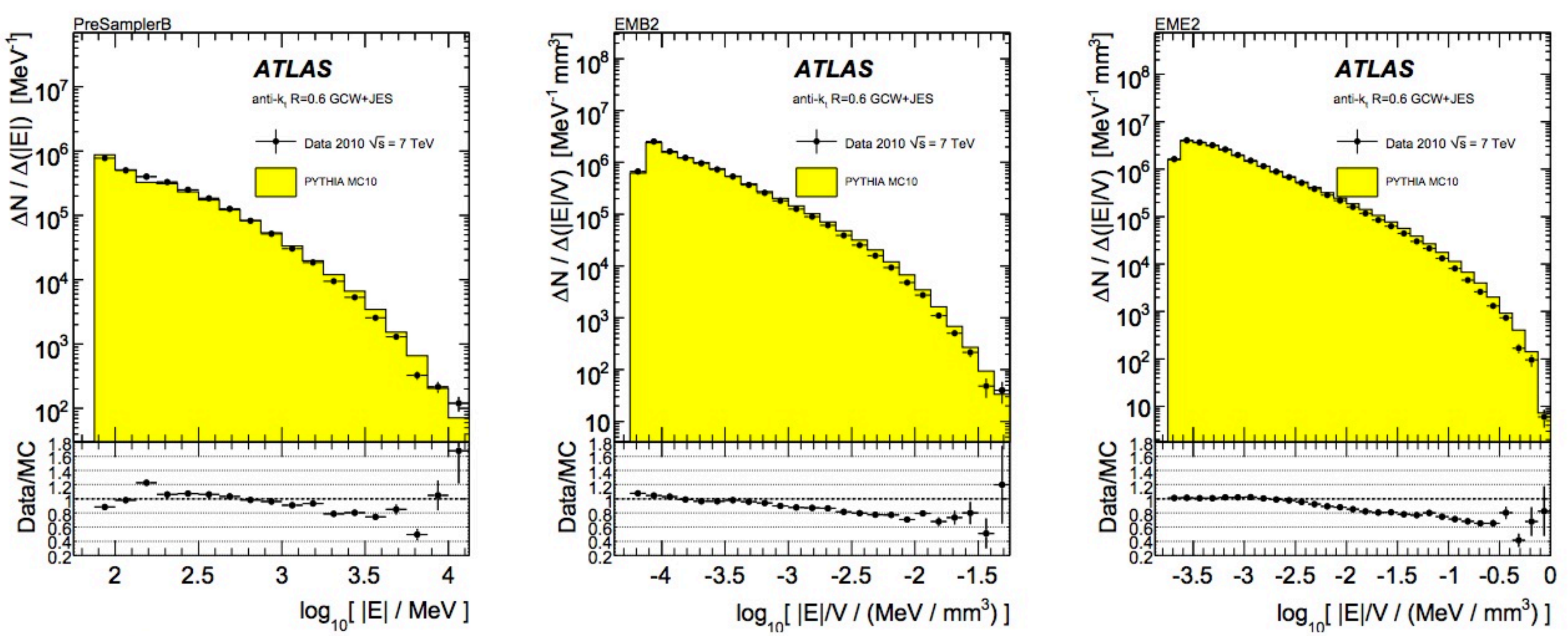
Edep has contributions from energy loss mechanisms which do not contribute to the signal, including nuclear binding energies and escaping energy carried by neutrinos. determine

Weights per cell are calculated from a set of observables

$$\mathfrak{O}^{\text{had-cal}}_{\text{cell}} = \left\{ S_{\text{calo}},\ \eta_{\text{cell}},\ \log_{10}(\rho_{\text{cell}}/\rho_0),\ \log_{10}(E^{\text{EM}}_{\text{clus}}/E_0) \right\}$$

sampling layer of the calorimeter

Requires very good modelling of calorimeter response




(a) $E^{\text{EM}}_{\text{cell}}$ in PreSamplerB (b) $\rho_{\text{cell}}$ in EMB2 (c) $\rho_{\text{cell}}$ in EME2

# Out of cluster correction

## Single particles can generate multiple clusters

+cells with small, true deposited energy may not be added to a cluster due to noise fluctuations

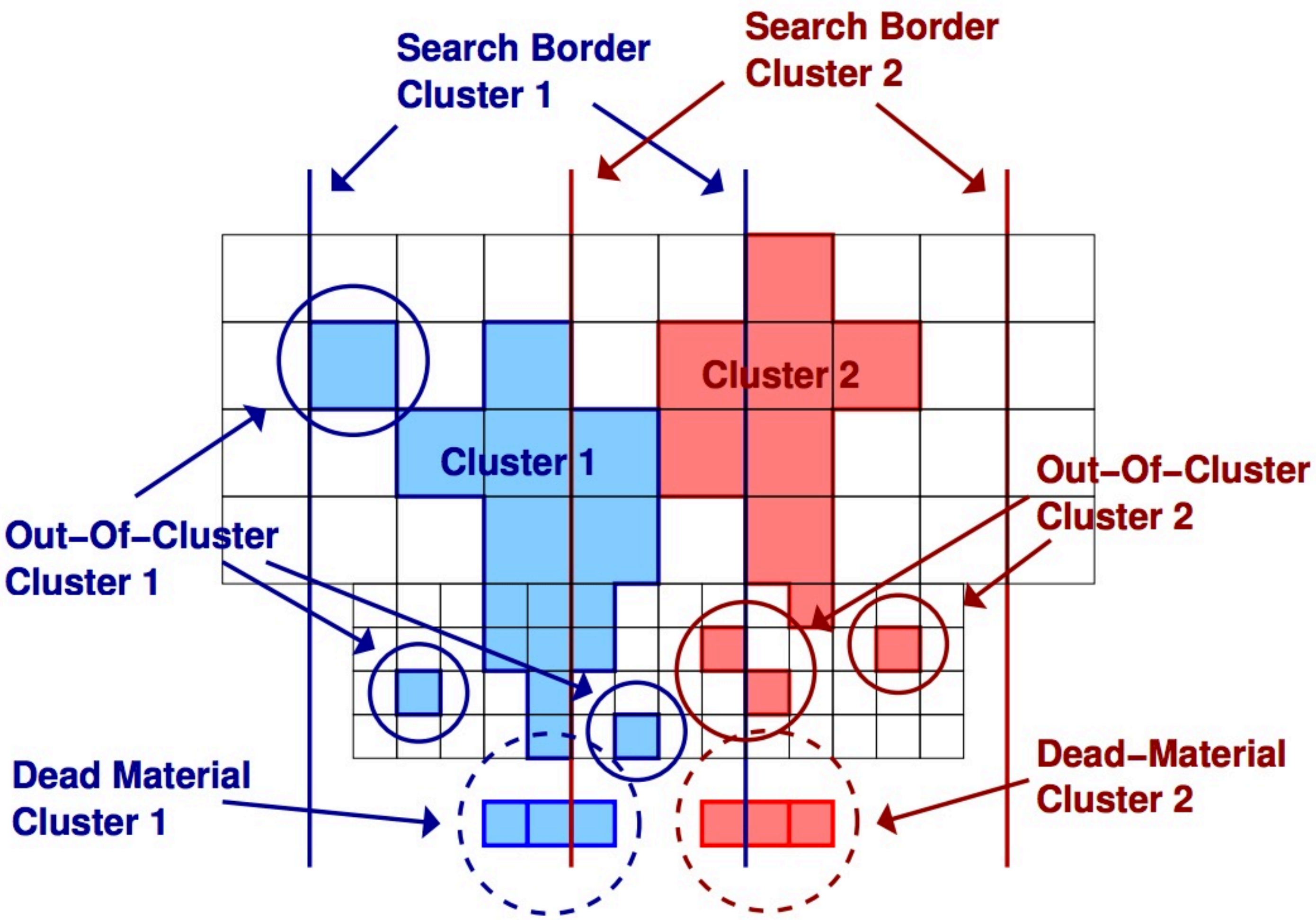
algorithm defining an *out-of-cluster neighbourhood* has been defined to try and recover these cells/ clusters.

- Size of neighbourhood depends on granularity (cluster eta)

$$w_{\text{clus}}^{\text{ooc}} = \frac{E_{\text{clus}}^{\text{ooc}} + E_{\text{clus}}^{\text{dep}}}{E_{\text{clus}}^{\text{dep}}} \quad \text{and} \quad E_{\text{clus}}^{\text{ooc}} = \sum_{i \in \{\text{lost cells}\}} E_{\text{cell,lost},i}^{\text{dep}} \, .$$

- Cell energy can be shared between overlapping clusters
- The correction is different for EM and hardon like clusters, again based on a set of parameters, tested in MC

$$\mathfrak{O}_{\text{clus}}^{\text{ooc}} = \left\{ \eta_{\text{clus}},\ \log_{10}(E_{\text{clus}}^{\text{EM}}/E_0),\ \log_{10}(\lambda_{\text{clus}}/\lambda_0) \right\}$$




Clusters in dense environments need a smaller correction since neighbouring clusters can capture less cells. The isolation moment is therefore also used to determine the sale of the correction

# Dead material correction

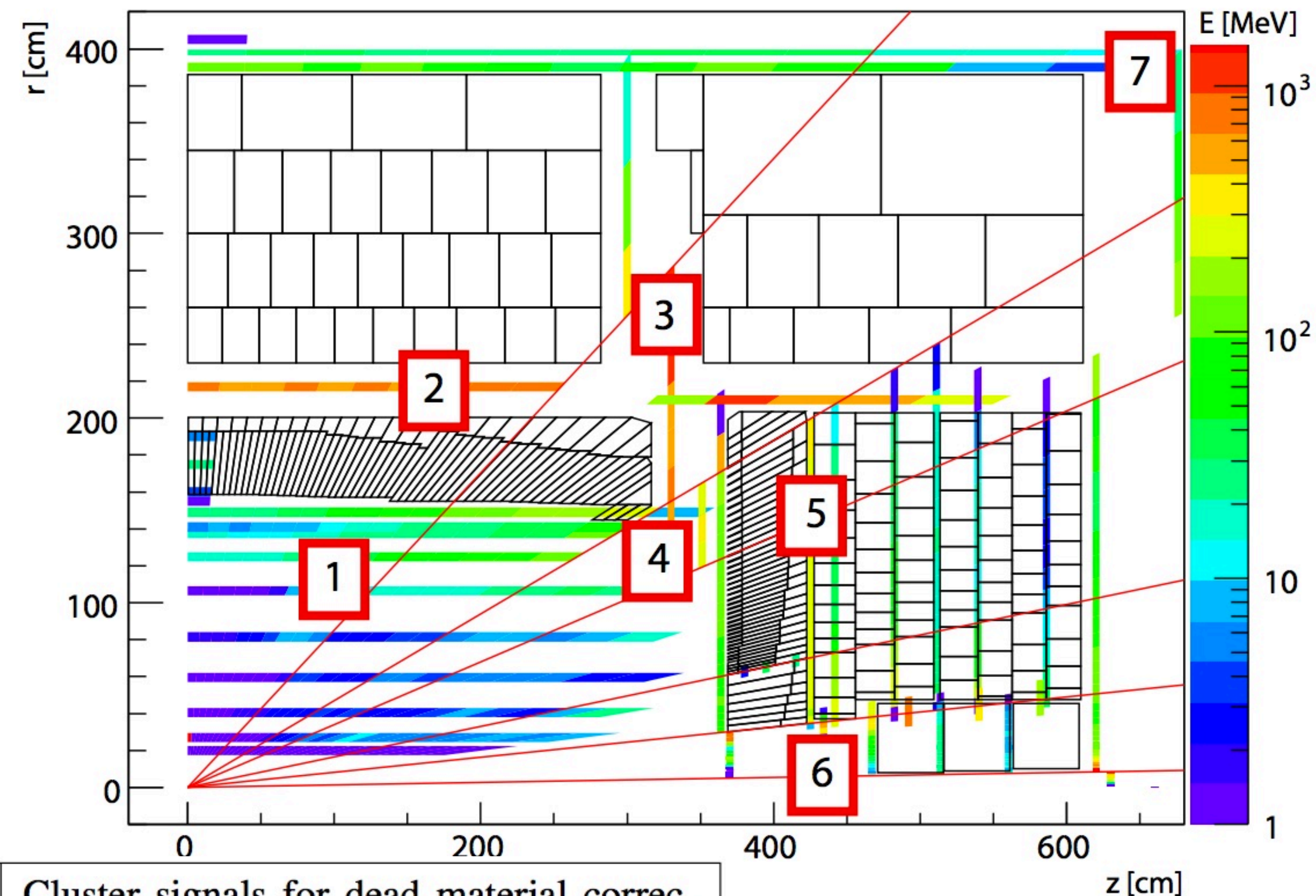
Energy deposited in detector dead material is accounted for

Detector simulation where dead material is divided into virtual cells (8 distinct regions)

Single pion MC samples are used to determine the expected correction, again binned in a set of cluster observables

$$\mathfrak{O}^{\mathrm{dm}}_{\mathrm{clus}} = \left\{ \eta_{\mathrm{clus}},\ \log_{10}(E^{\mathrm{EM}}_{\mathrm{clus}}/E_0),\ \lambda_{\mathrm{clus}} \right\}$$

In certain layers, where available, Pre sampler and measured energies are also used.




| Regions | Description | Cluster signals for dead material correction |
|---|---|---|
| 1 | In front of EMB | Energy in PreSamplerB |
| 2 | Between EMB and Tile | Energies in last layer of EMB and first layer of Tile |
| 3 | In front of Tile gap scintillators | Energy in Tile gap scintillators |
| 4 | In front of EMEC | Energy in PreSamplerE |
| 5 | Between EMEC and HEC | Energies in last layer of EMEC and first layer of HEC |
| 6 | In front of FCAL | Energy in first FCAL module |
| 7 | Behind calorimeters | Energy in last layer of hadronic calorimeters and $\mathfrak{O}^{\mathrm{dm}}_{\mathrm{clus}}$ given in Eq. (25) |
| 8 | Everywhere else | |

# + more

Not the whole story

Cluster kinematics are recalculated after the full calibration

Many further things that could be added

- Topo-cluster performance in terms of jets
- Jet algorithms
- Further corrections in Jet calibrations
- Performance tests
- Thresholds: optimisation w.r.t pile-up
- Many Data/MC plots
- Missing Et
- Cell level noise calibration