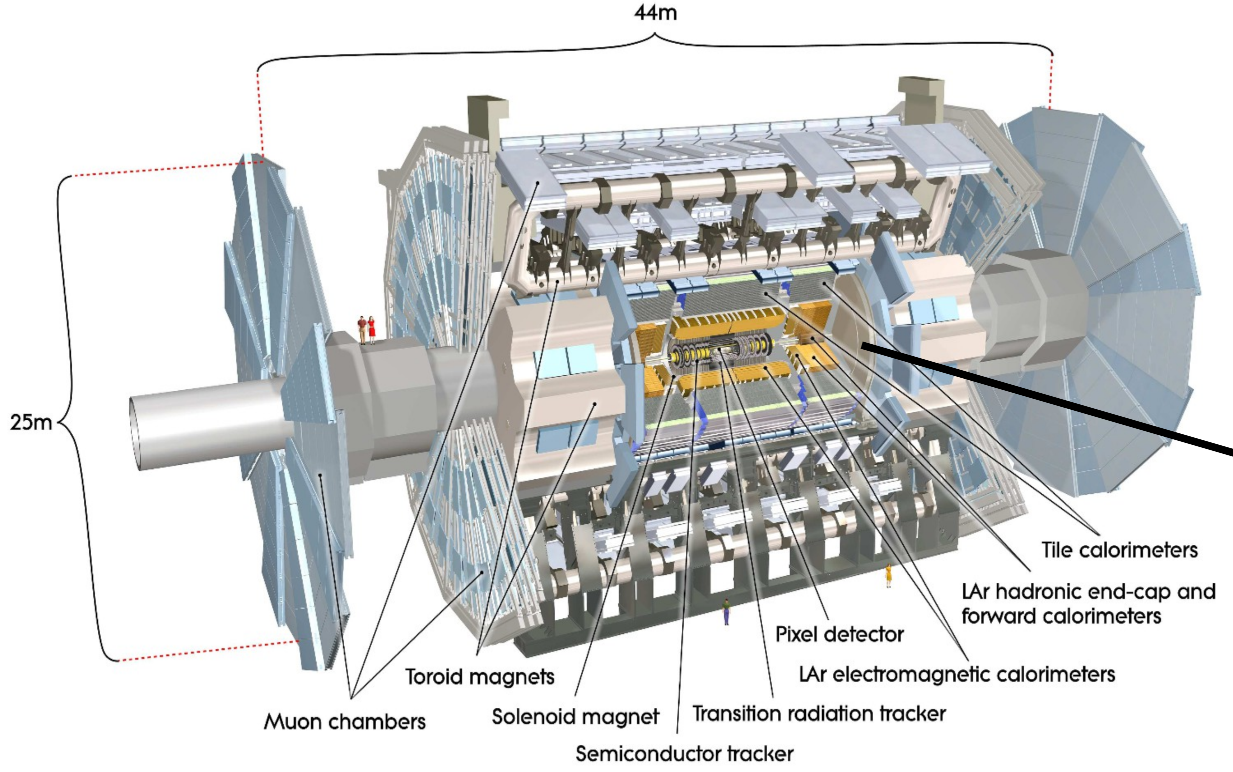
The new ATLAS Fast Calorimeter Simulation

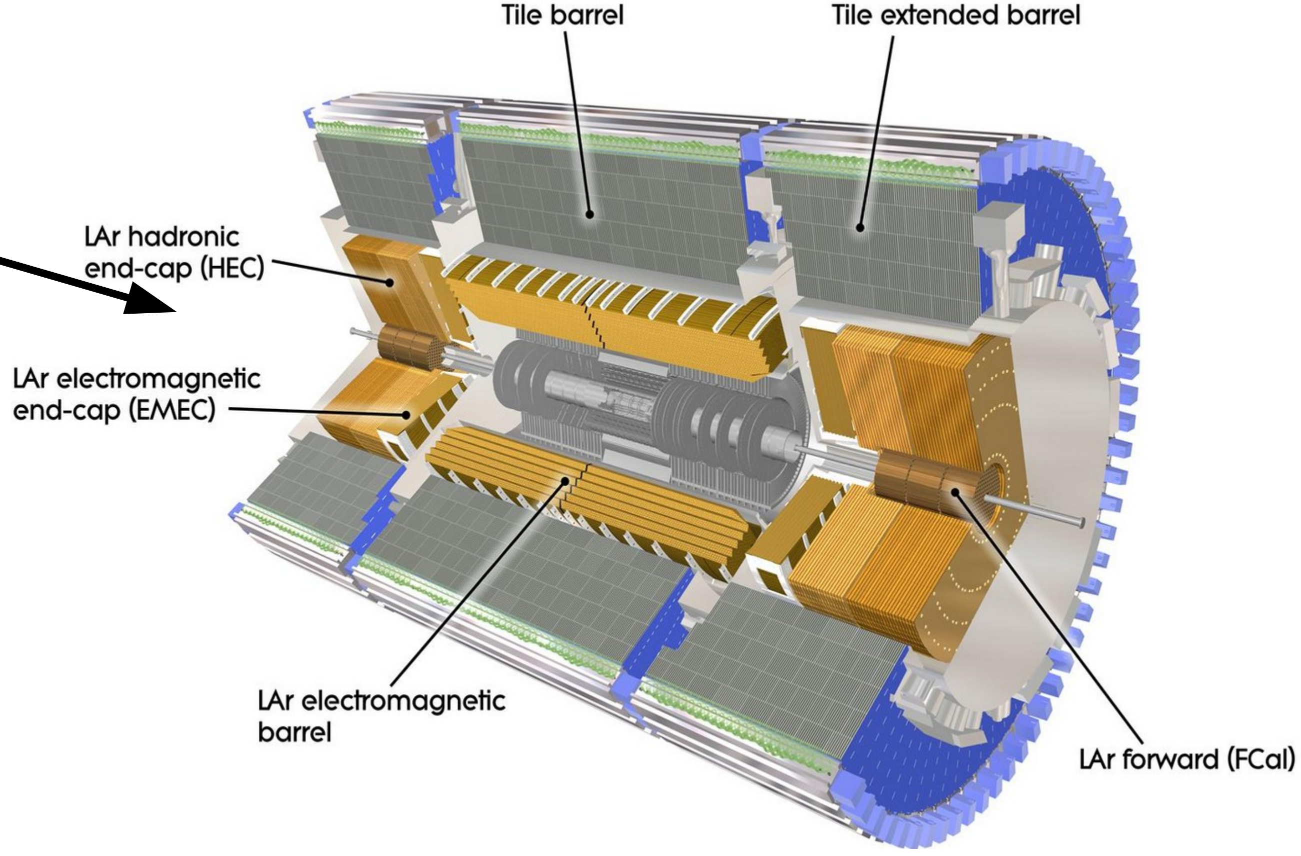
Jana Schaarschmidt (University of Washington)
on behalf of the ATLAS collaboration

CHEP Sofia, 9th July 2018

The ATLAS detector & calorimeter



The ATLAS calorimeter system:



Readout channels: ~190 k in total
Samplings (layers of cells): 24

System	EM Barrel	EM EndCap	Hadronic EndCap	FCAL	TileCal
#Channels	110k	64k	5.6k	3.5k	9.8k

Materials:

Liquid Argon + Lead, or copper or tungsten
Tile Cal: Steel + plastic

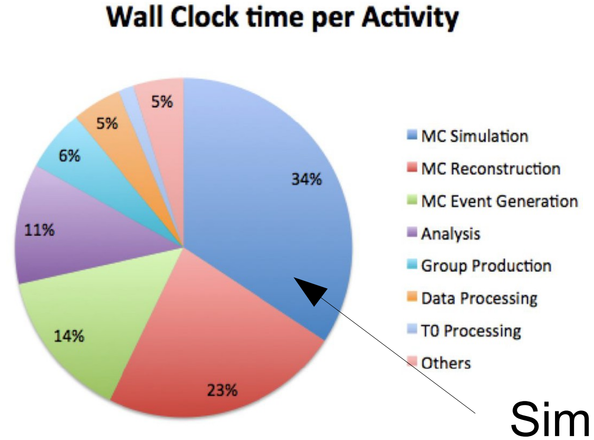
The need for fast simulation

Typical times in s: 2010, “full” simulation with Geant4 (ref)

Sample	Generation	Simulation	Digitization	Reconstruction
Minimum Bias	0.0267	551.	19.6	8.06
$t\bar{t}$ Production	0.226	1990	29.1	47.4
Jets	0.0457	2640	29.2	78.4
Photon and jets	0.0431	2850	25.3	44.7
$W^{\pm} \rightarrow e^{\pm}\nu_e$	0.0788	1150	23.5	8.07
$W^{\pm} \rightarrow \mu^{\pm}\nu_\mu$	0.0768	1030	23.1	13.6
Heavy ion	2.08	56,000	267	-

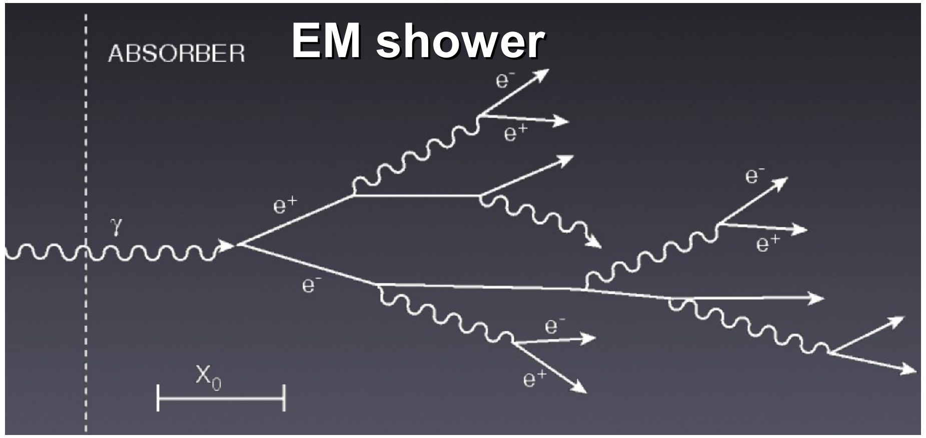
arXiv:1005.4568

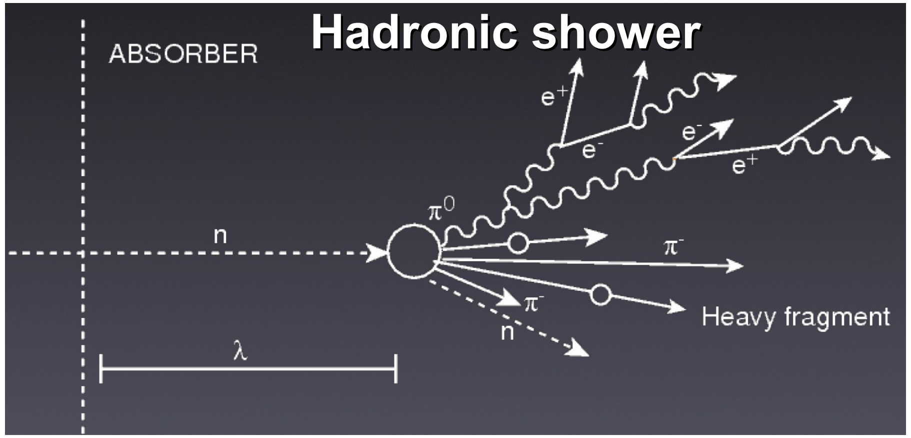
Grid usage 2016:



Simulation!

85% of the simulation time is spent in the calorimeters





Wider, slower, larger fluctuations than EM showers

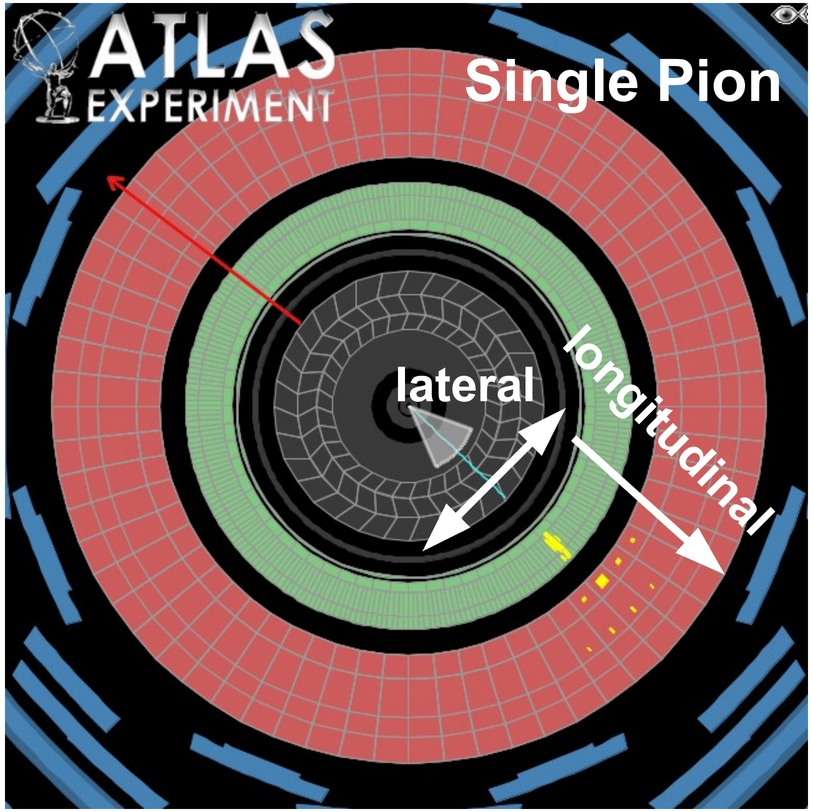
- Geant4 is slow, but most accurate. It is the ultimate reference for simulation
- ATLAS relies on fast simulation, even more in the future: The resources do not scale with our MC needs!
- Now ~50% of all MC events in ATLAS are fast simulated. But gains in speed come at the cost of accuracy. Ultimate goal is that fast simulation becomes so good, that it can be used for (almost) any process.

Fast Calorimeter Simulation: Introduction

Parametrized calorimeter response of single particles, based on the Geant4 simulation, derived on a fine grid of energy and eta, separated into longitudinal and lateral components.

Electrons and Photons:	To decribe electromagnetic showers
Charged Pions:	To describe any hadronic shower

Geant4 single particle simulation → Parametrizations → Fast Simulation → Validation



- Need **5100 samples** to derive all parametrisations:
  - 17 points in energy: 64 MeV – 4 TeV
  - 100 bins in eta: covering full detector $0 < |\eta| < 5$
  - 3 particle types: $e^{\pm}$, γ, $\pi^{\pm}$

In the end, the entire parametrisation needs to be loaded in memory. Should be limited < 2 GB.

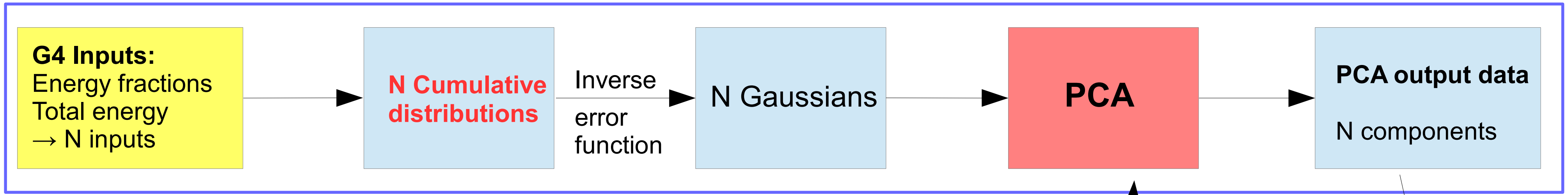
The Longitudinal Energy Parametrisation

Energy deposit in each calorimeter layer along the shower axis and total energy

Problem: The energy deposits in the various layers are correlated with each other

Transformation to uncorrelated set of variables with principal component analysis, to reduce complexity

1st PCA chain:



G4 Inputs:
Energy fractions
Total energy
→ N inputs

N Cumulative distributions

Inverse error function

N Gaussians

PCA

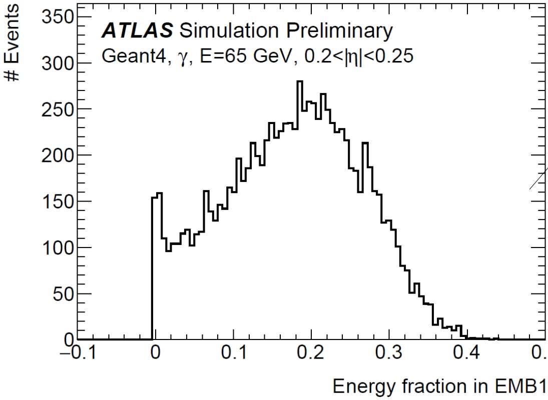
PCA output data

N components

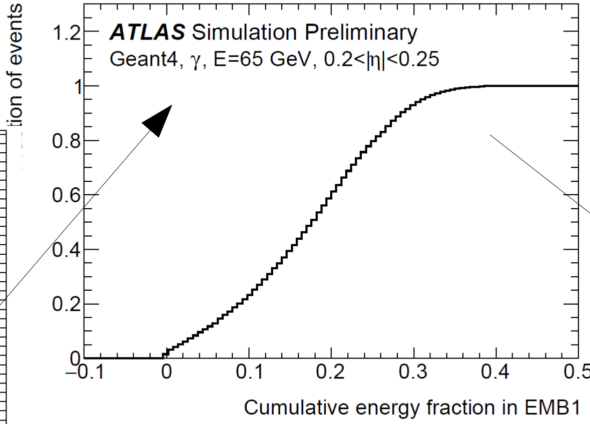
TPrincipal (reference)

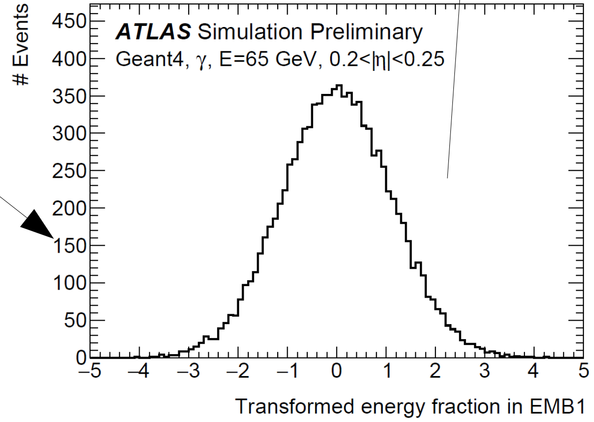
N outputs

Example:
Photons 65 GeV

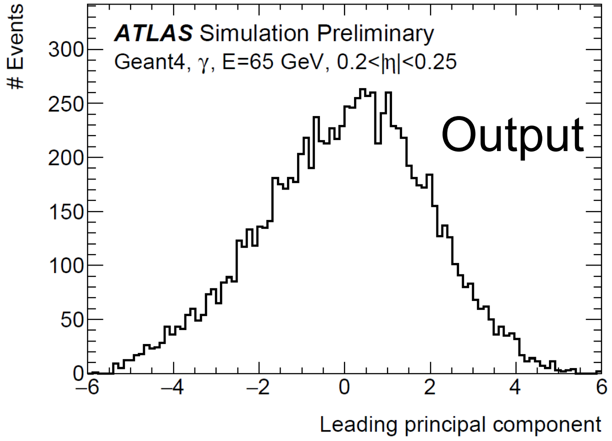


Cumulative





Gaussian

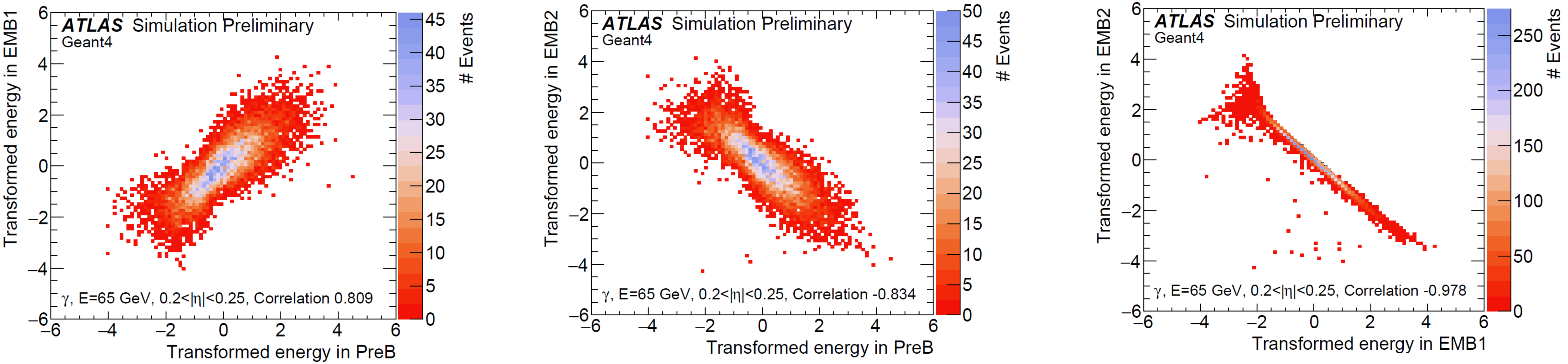


Output

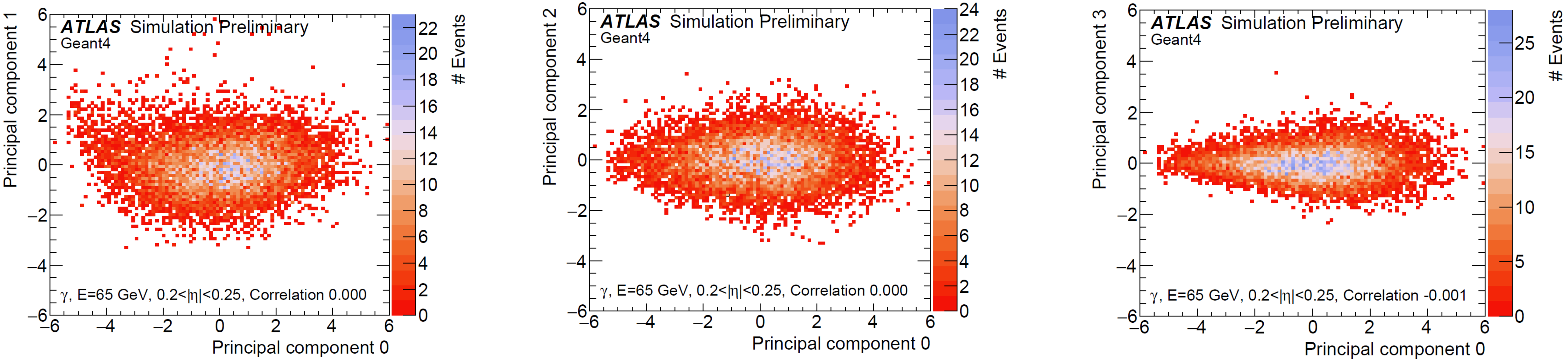
First principle component is that eigenvector of the covariance matrix with the largest eigenvalue (variance)

The Longitudinal Energy Parametrisation

Correlations between energies before PCA rotation, here for 65 GeV photons 0.2<|η|<0.25 :

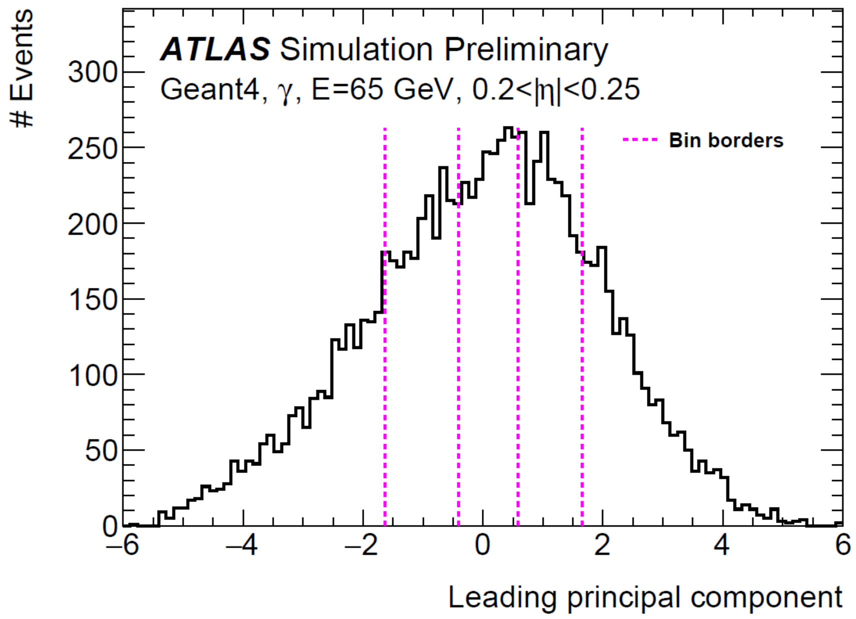


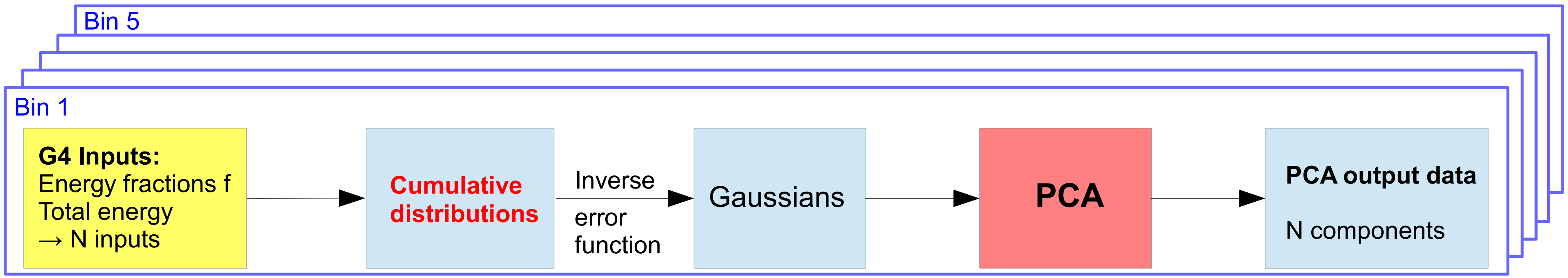
Correlations between energies after PCA rotation:



The Longitudinal Energy Parametrisation

- The leading principal component is used to divide the input data into quantiles („PCA bins“)
- In each such bin, showers have similar features
- These “PCA bins” are also used to derive the shape parametrisation.
- In each „PCA bin“, another PCA rotation is perfomed to get even better decorrelation:





Bin 5

Bin 1

G4 Inputs:
Energy fractions f
Total energy
→ N inputs

Cumulative distributions

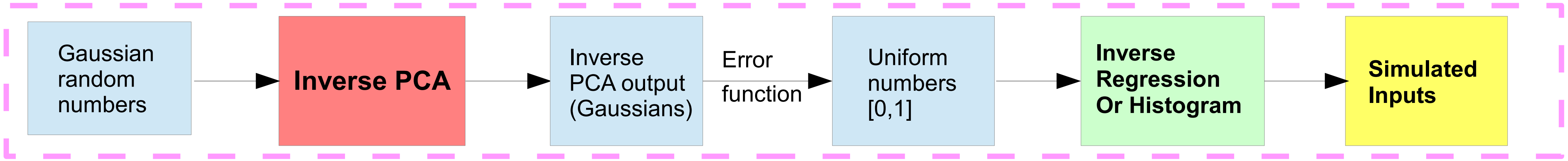
Inverse error function

Gaussians

PCA

PCA output data
N components

During simulation, this chain is performed back-wards:



Gaussian random numbers

Inverse PCA

Inverse PCA output (Gaussians)

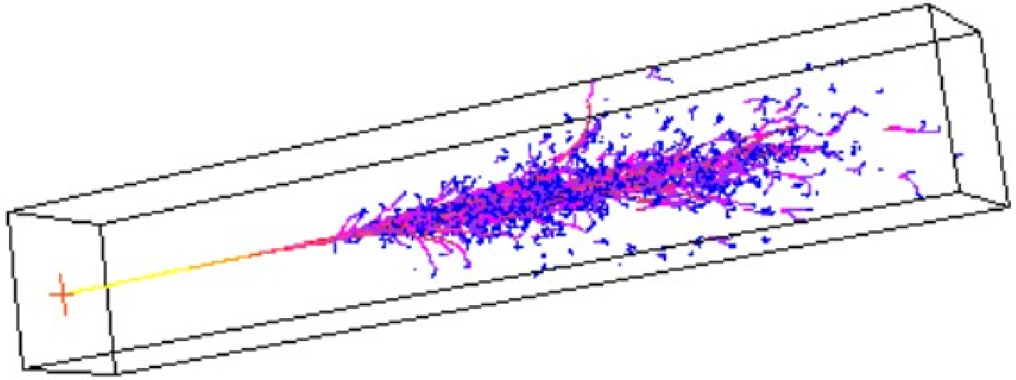
Error function

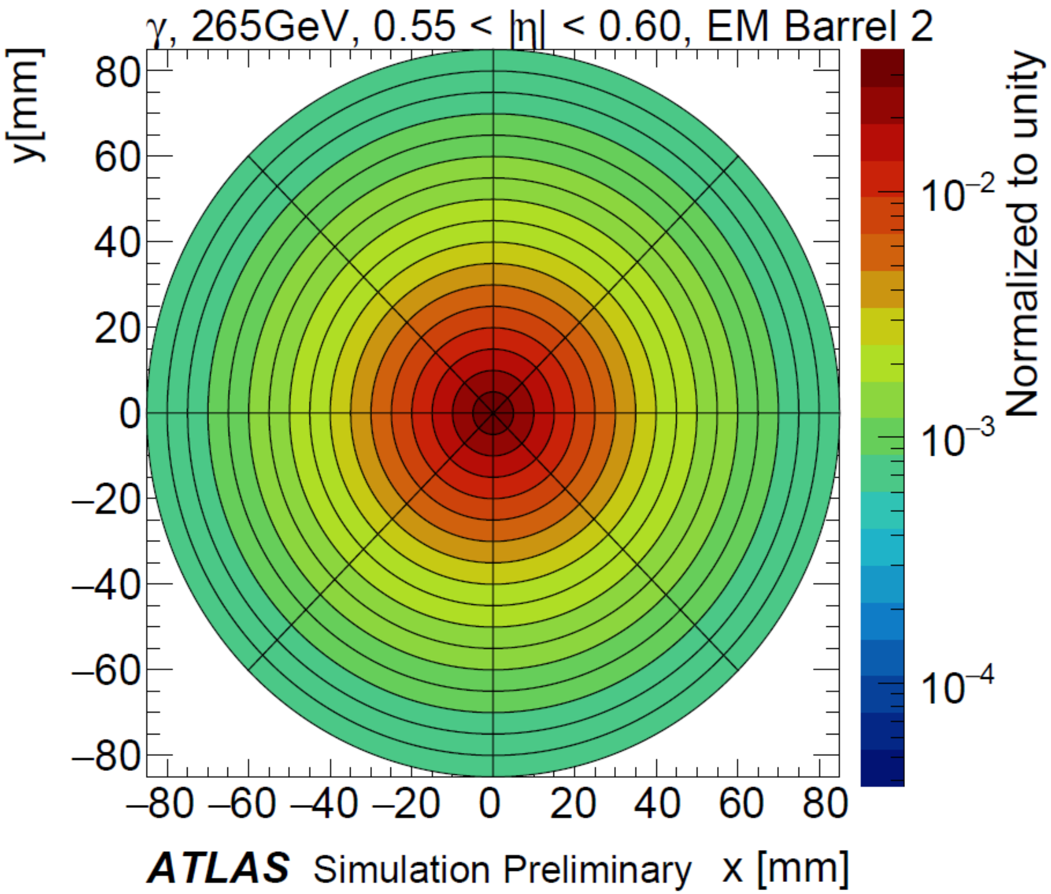
Uniform numbers [0,1]

Inverse Regression Or Histogram

Simulated Inputs

The Lateral Energy Parametrisation („Shape“)





- Shower shape:

  - Most energies in the center (close to the shower axis)
  - Energy tails extending perpendicular to the axis

- The shape parametrisation is based on Geant4 HITs.

  - Close-by hits merged to reduce computation time
  - Hits saved in ntuple format to be used to derive histograms

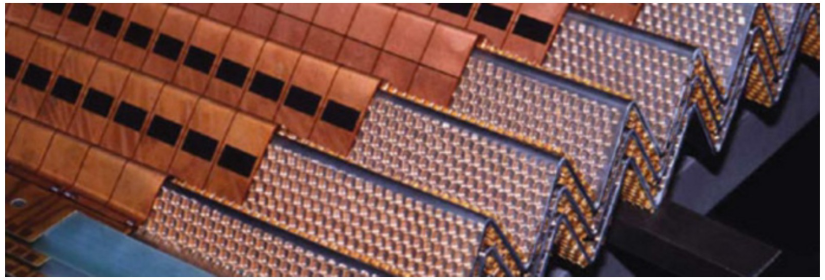
- These 2D histograms act as probability density functions during the fast simulation: Fast sim hits are randomly sampled from it

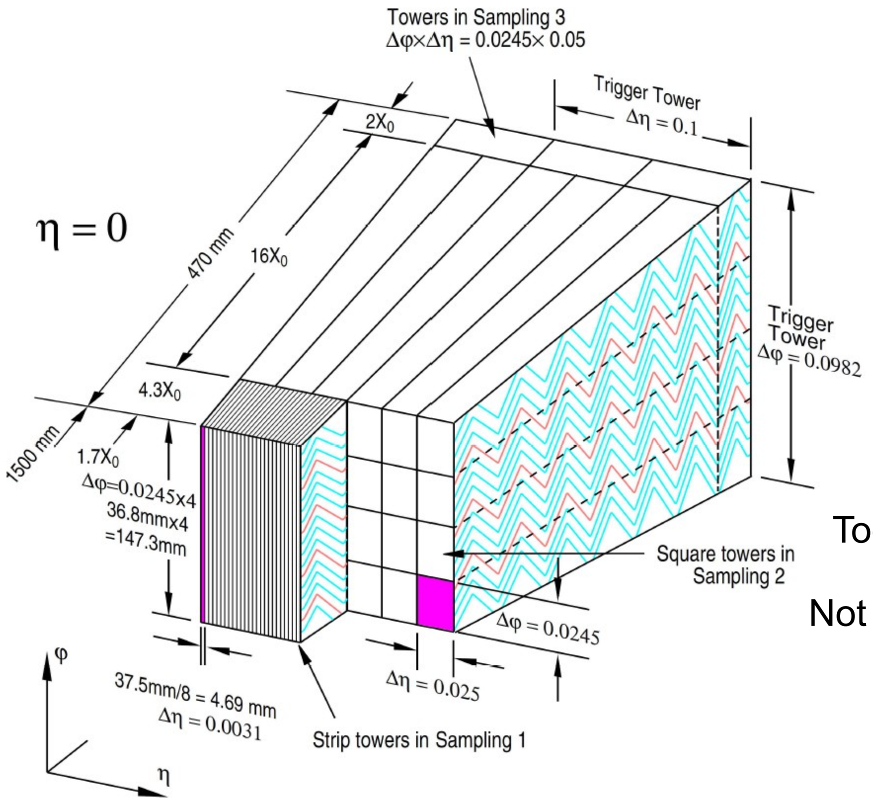
- 2D histogram stored per layer and per PCA bin

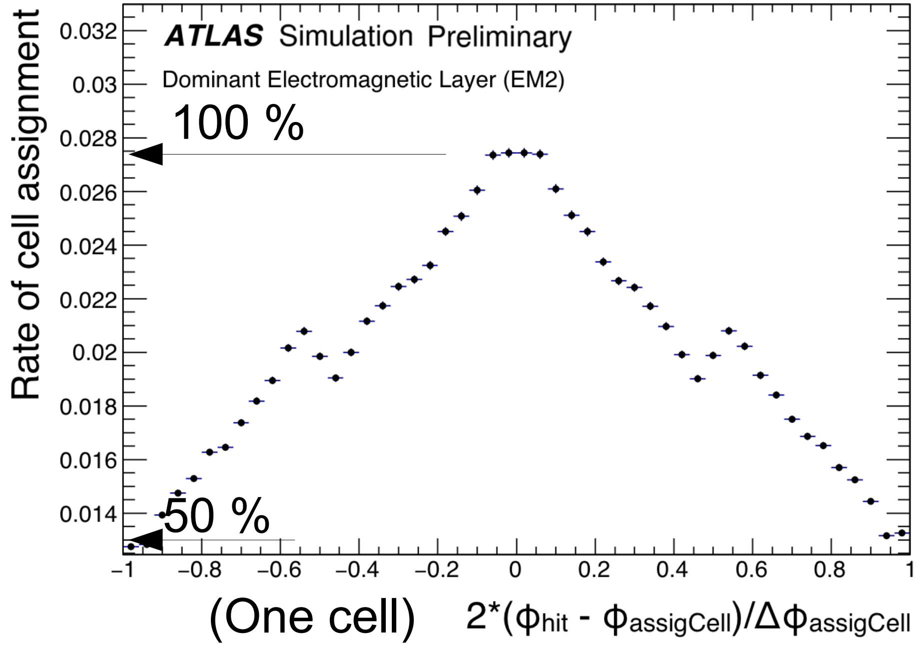
- Spline and regression techniques can be used to reduce memory

Emulating the Accordion Structure („Wiggle“)

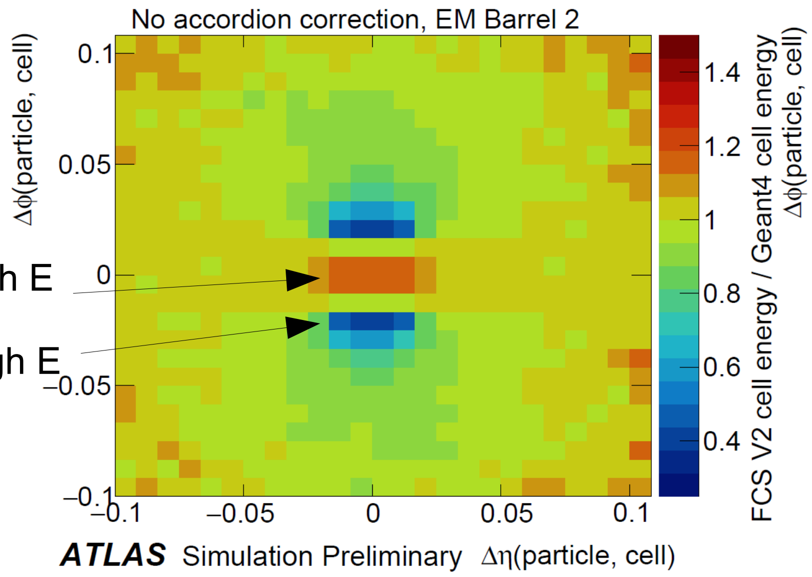
Probability function describing the chance that the energy belongs to this cell (if <1, some chance it belongs to a neighbour cell)







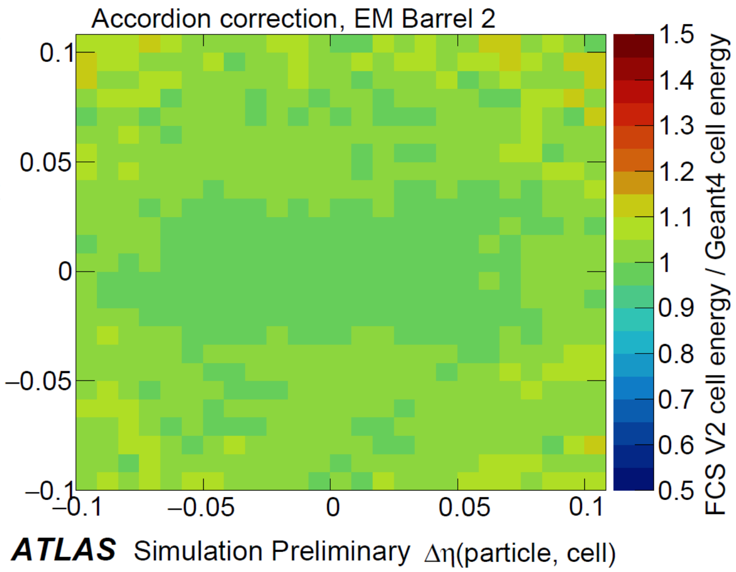
Without wiggle:



Too much E

Not enough E

With wiggle:



The Simulation Steps

Start of the event: Loop over all truth particles

- Per particle: Incoming **truth particle** with a given PDG-ID, eta and kinetic energy
- Grep the **right parametrisation** that fits these kinematics
- Throw a **random PCA bin**
- Perform the **energy simulation** in that bin → output: total energy E, energy per layer
  - For each layer: **Sample N hit positions** from the shape histogram
    - For each hit: Assign a **hit energy** E/N and match it to a physical **calorimeter cell**
    - If the hit belongs to a layer that has the accordion structure, determine a **random wiggle factor**

At the end of the event: Pass the CaloCellContainer to **digitization** and then **reconstruction**

All these steps are implemented in the Integrated ATLAS simulation framework (ISF)

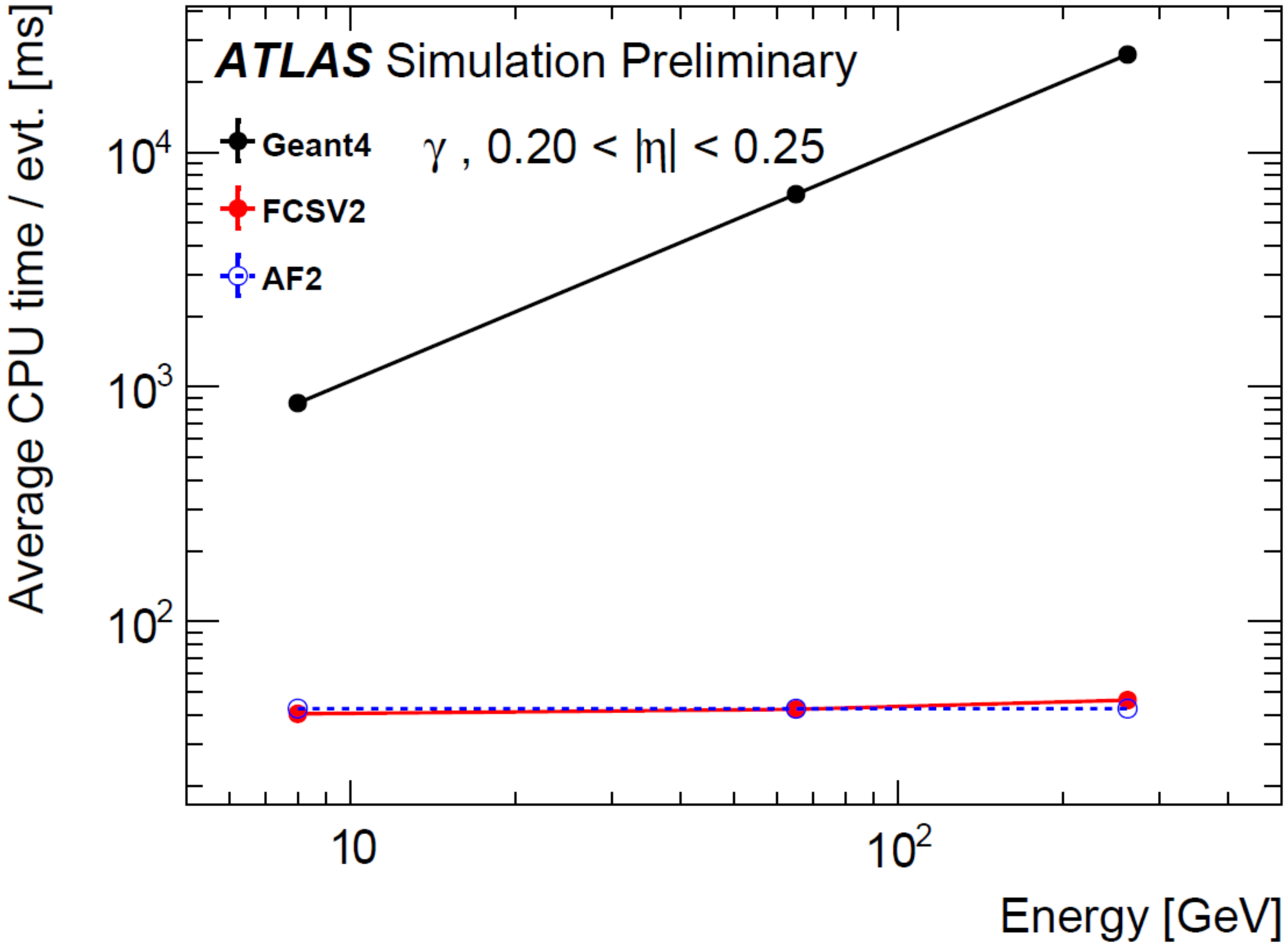
The Timing Performance & Validation Strategy

How fast is FastCaloSim?

AF2: Former version of FastCaloSim
FastCaloSim V2: Improved version.

→ FastCaloSim is factor 10-25 faster than G4

Caveat: This is simulation only!
Particles are generated at calo surface
(no inner detector simulation).

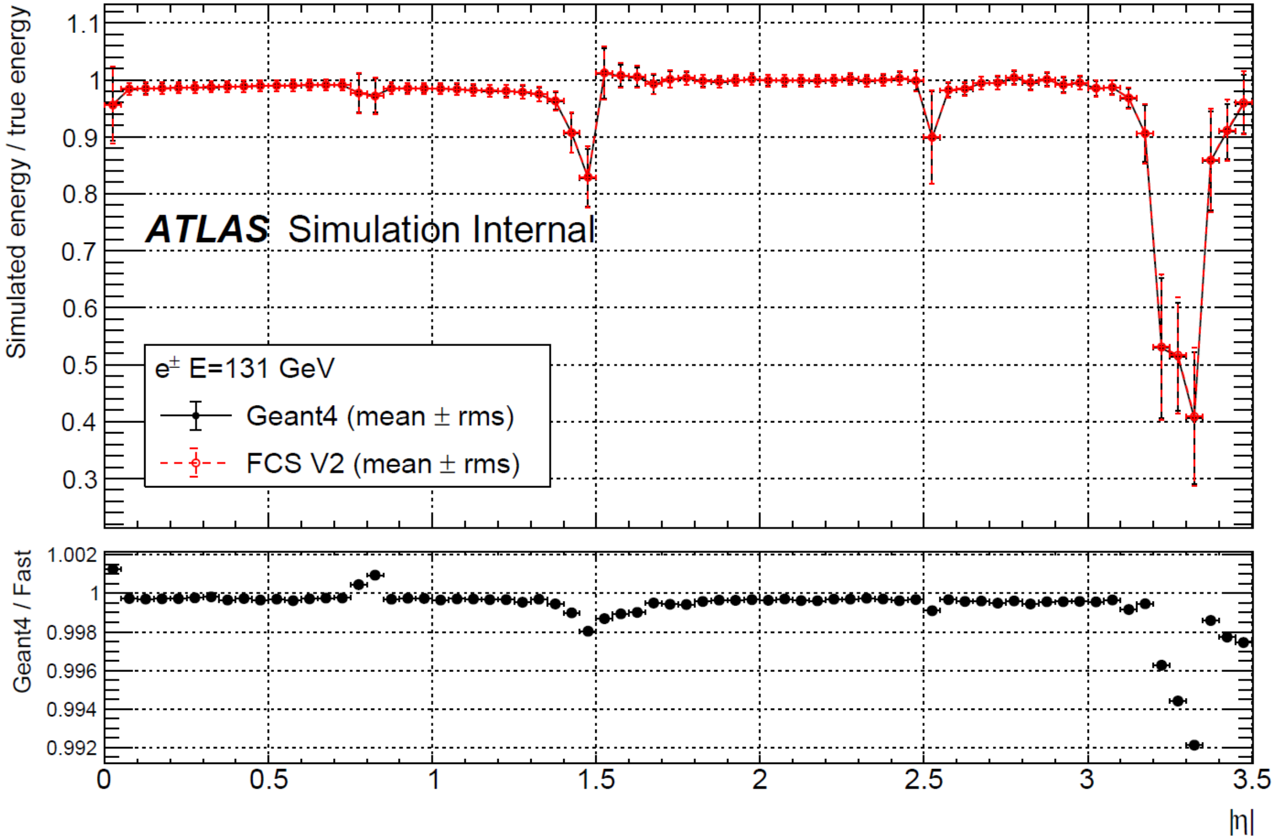


Validation strategy:

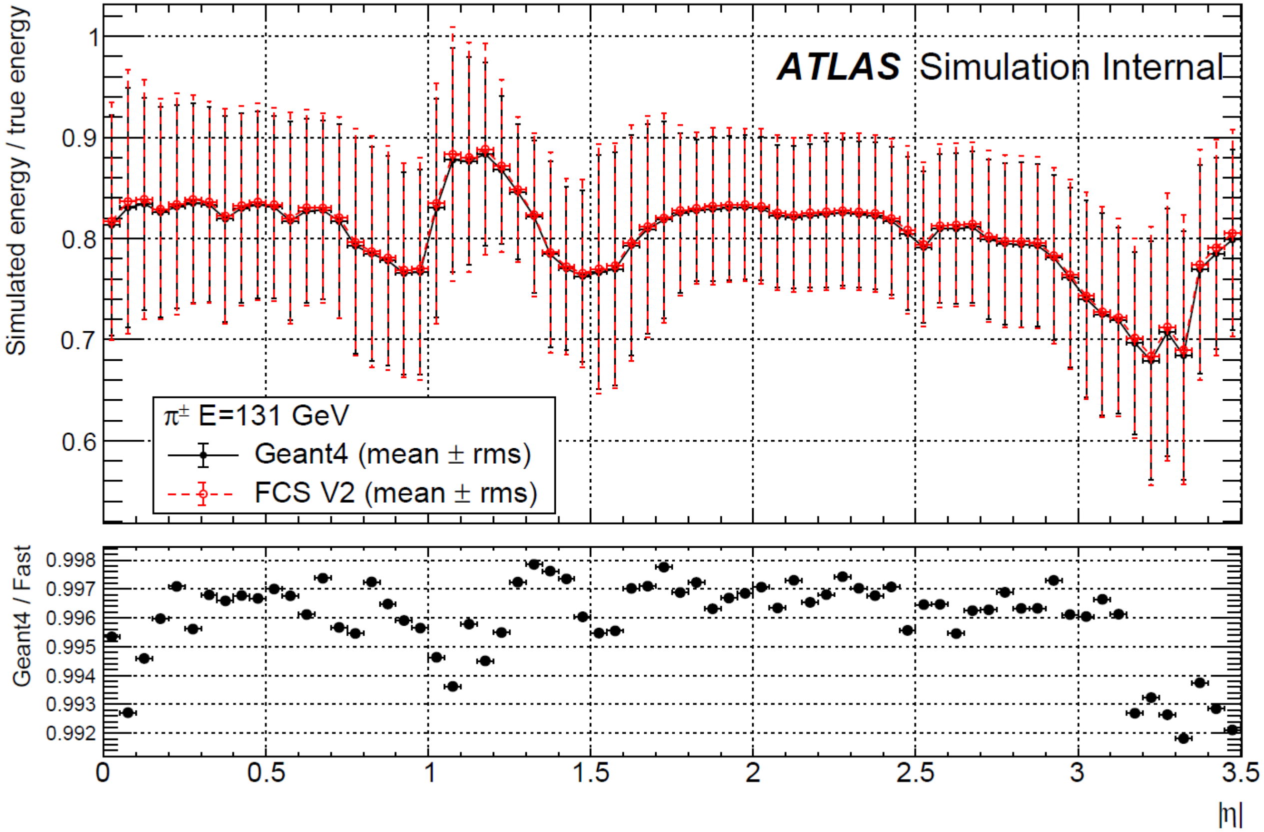
- On-the-fly toy simulation when parametrizations are produced. This is fast!
- Comparisons of single particle simulations with Geant4 and FCS
- Comparison of (multi-particle) physics simulation with Geant4 and FCS

Validation of the energy response

Electrons:



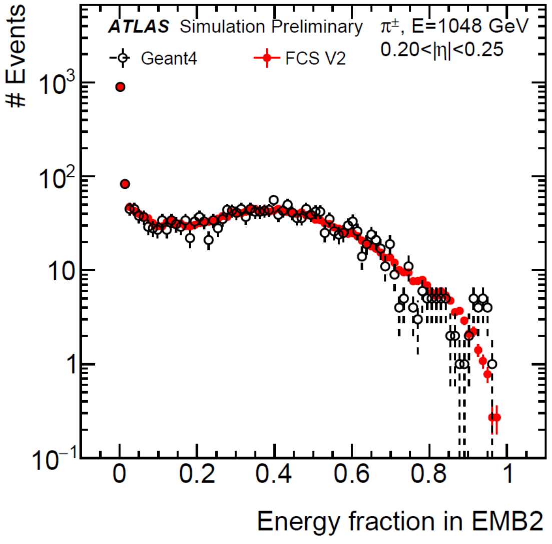
Pions:

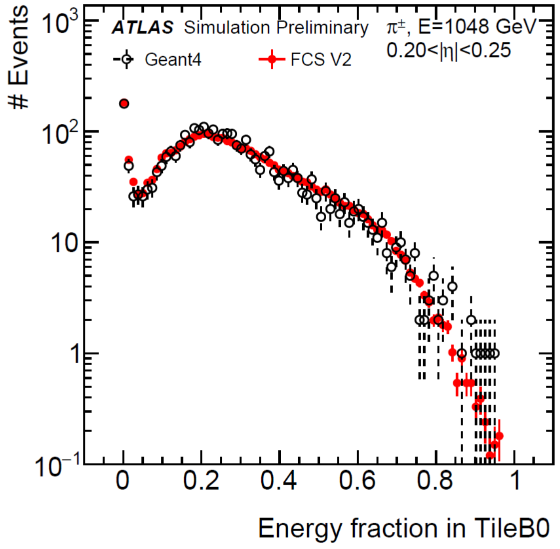


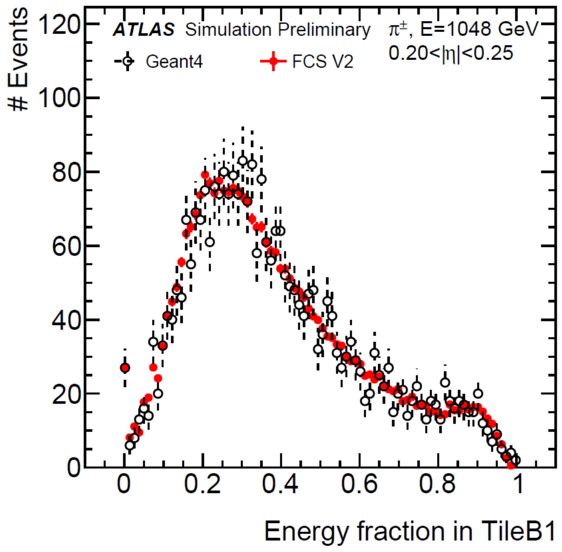
- Egamma showers are more narrow, so more sensitive to the detector geometry changes
- Total energy response agrees remarkably well between G4 and new FastCaloSim
- Even if correlations between layers are not well modelled for difficult eta regions, the total energy is still well reproduced

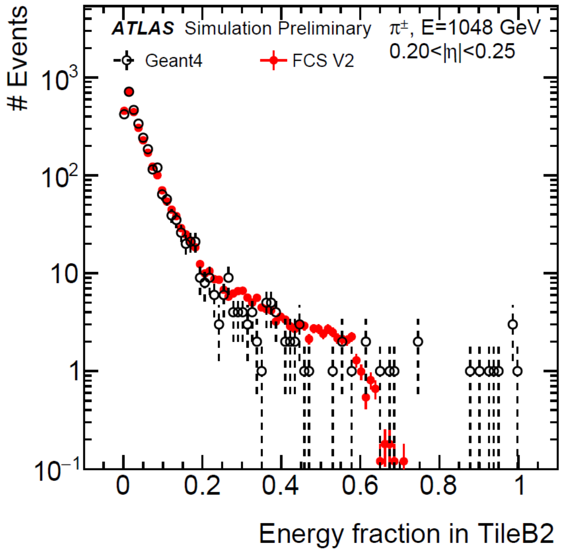
Validation of the energy response

Energy deposited in each layer, 1 TeV central pions:



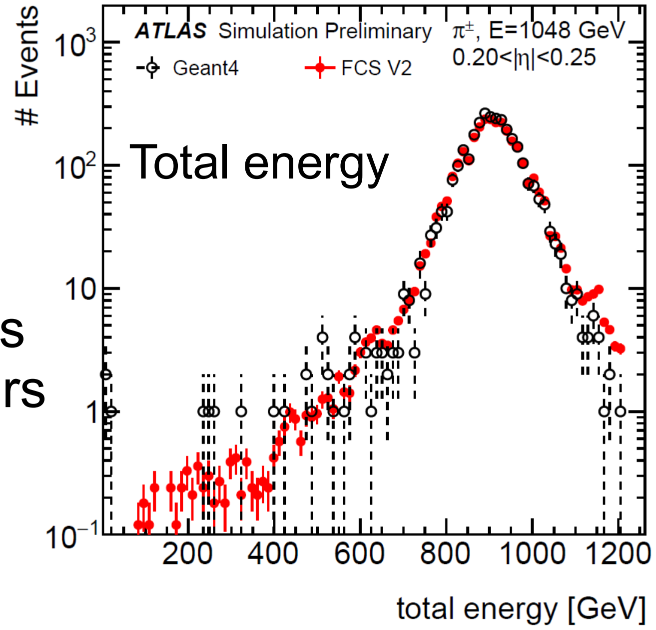




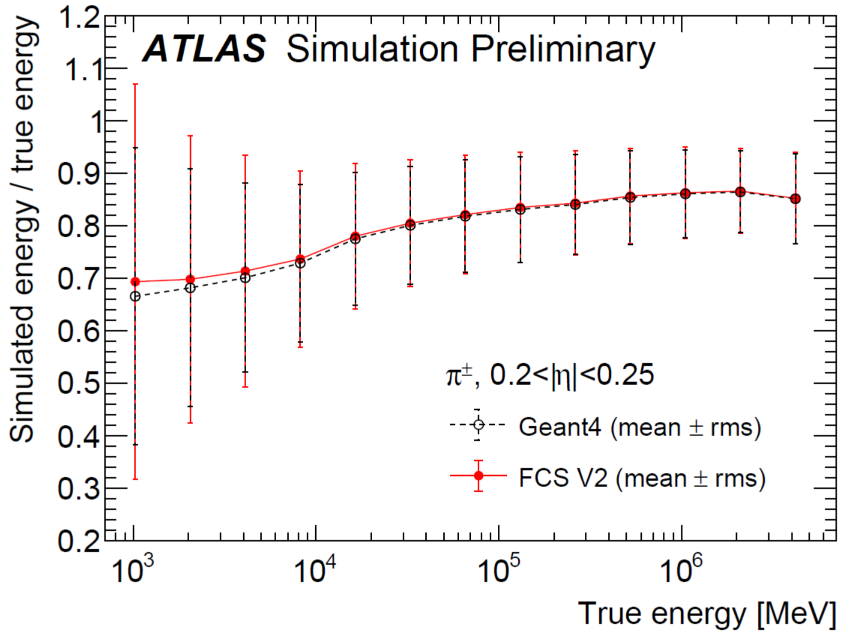


Energy in all layers well modelled.

For transition regions between subdetectors still some problems (not shown here).

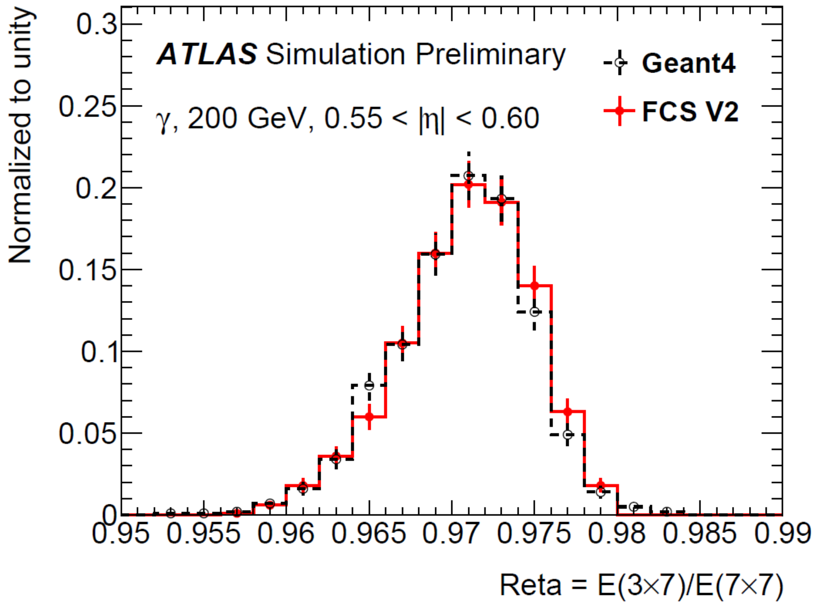


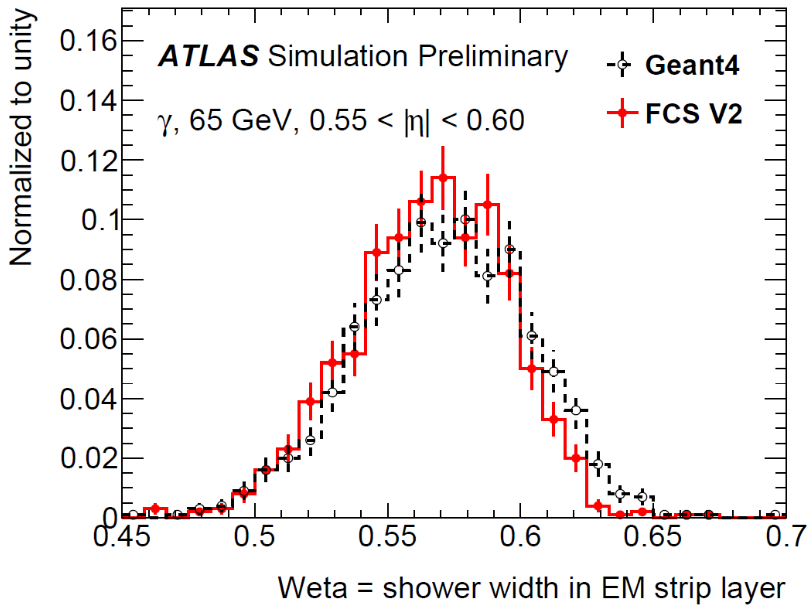
Energy response, central pions, scan over energy →

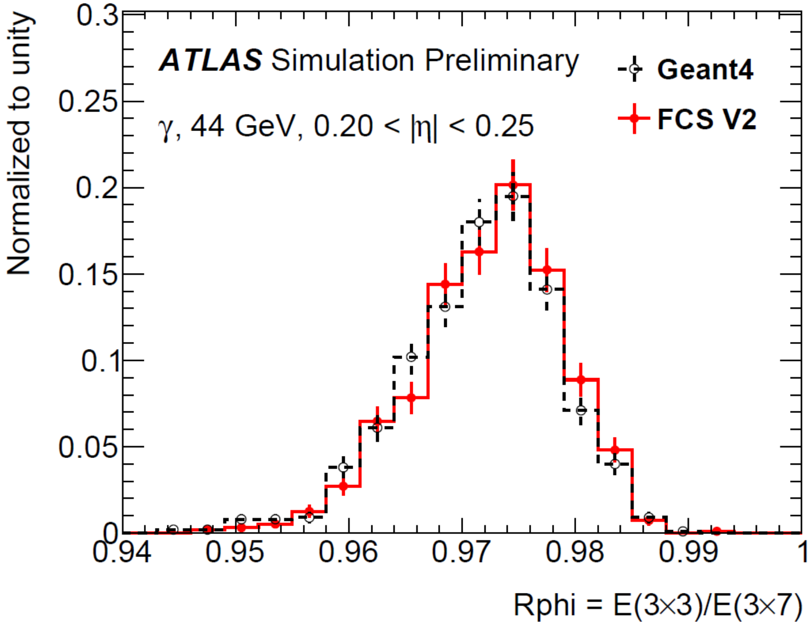


Validation of single particles (after reconstruction)

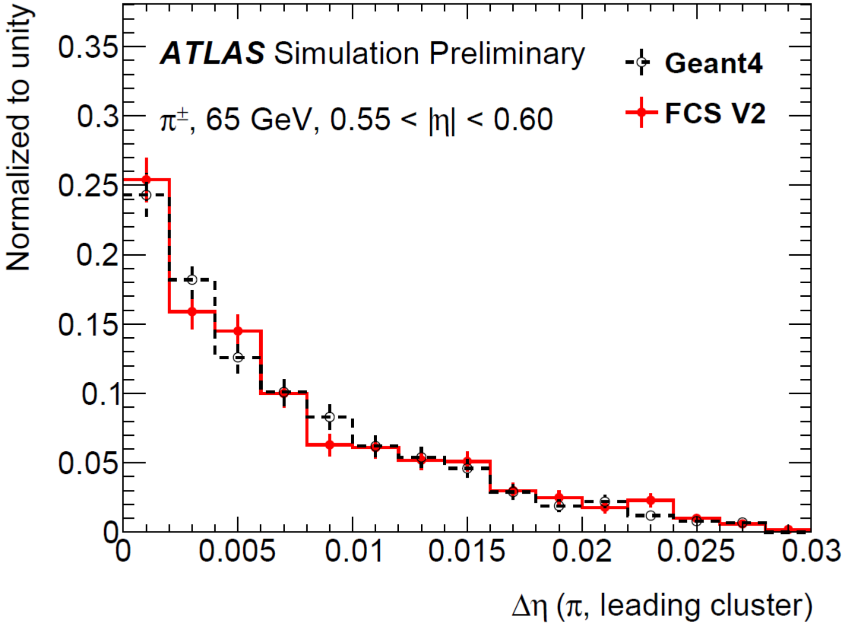
Photons, 0.2<|η|<0.25:

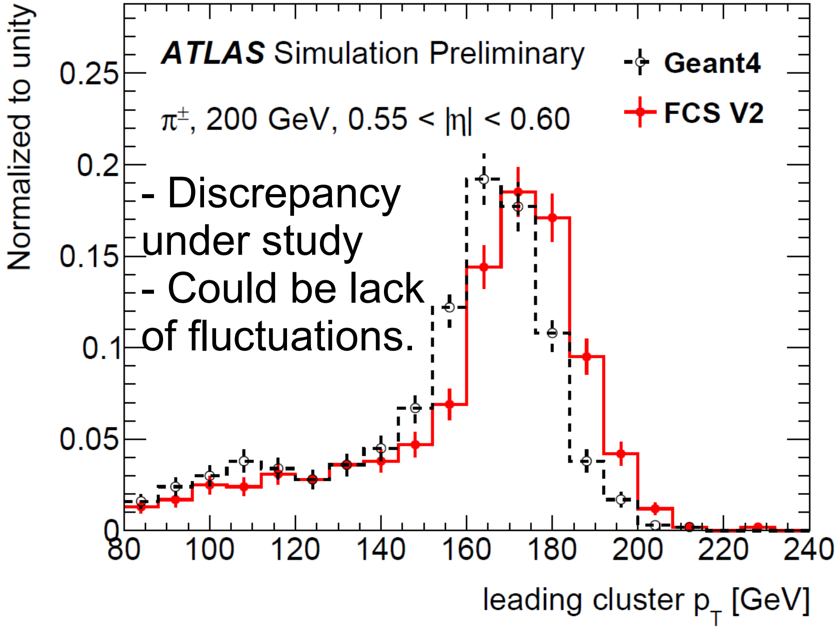




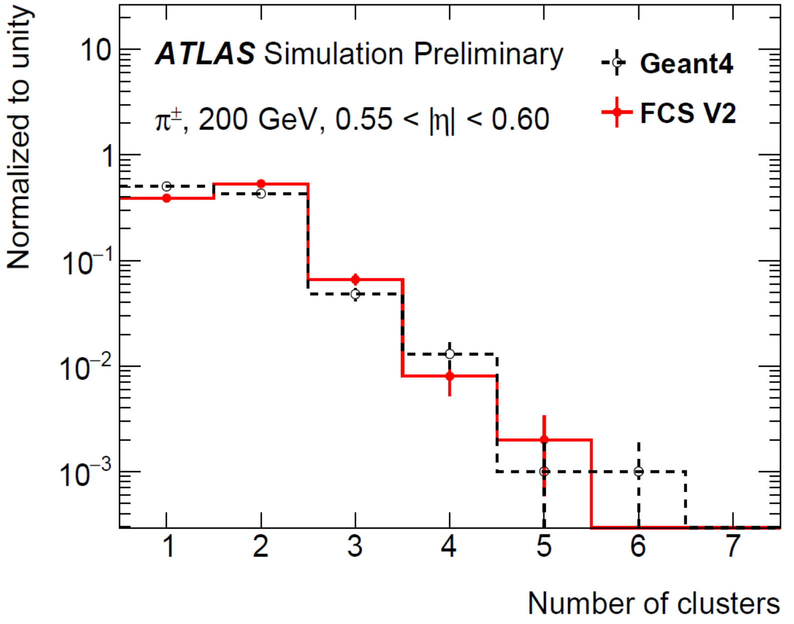


Pions, 0.2<|η|<0.25:

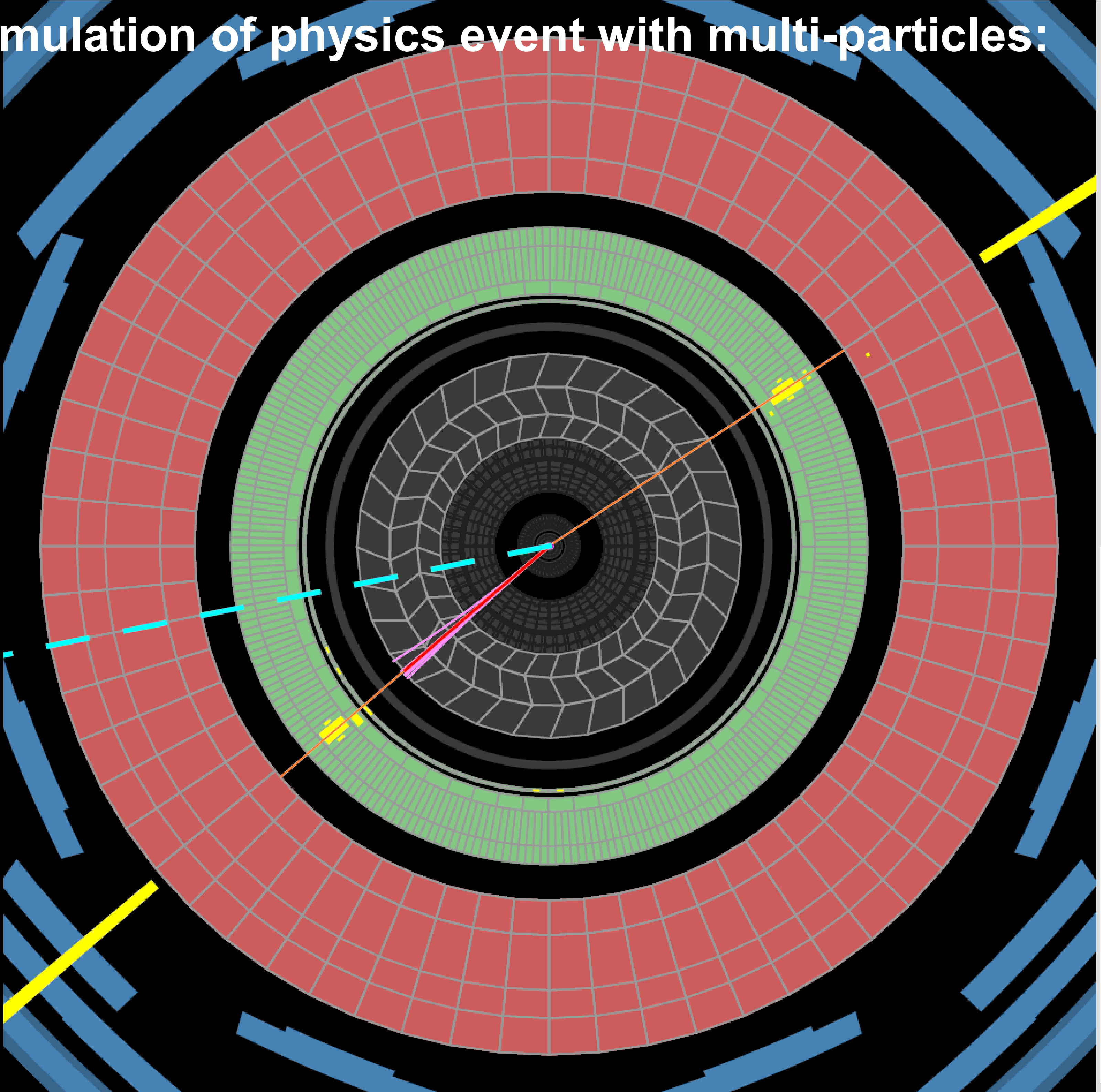




- Discrepancy under study
- Could be lack of fluctuations.



Simulation of physics event with multi-particles:



ATLAS
EXPERIMENT

FCS V2, H -> γγ MC

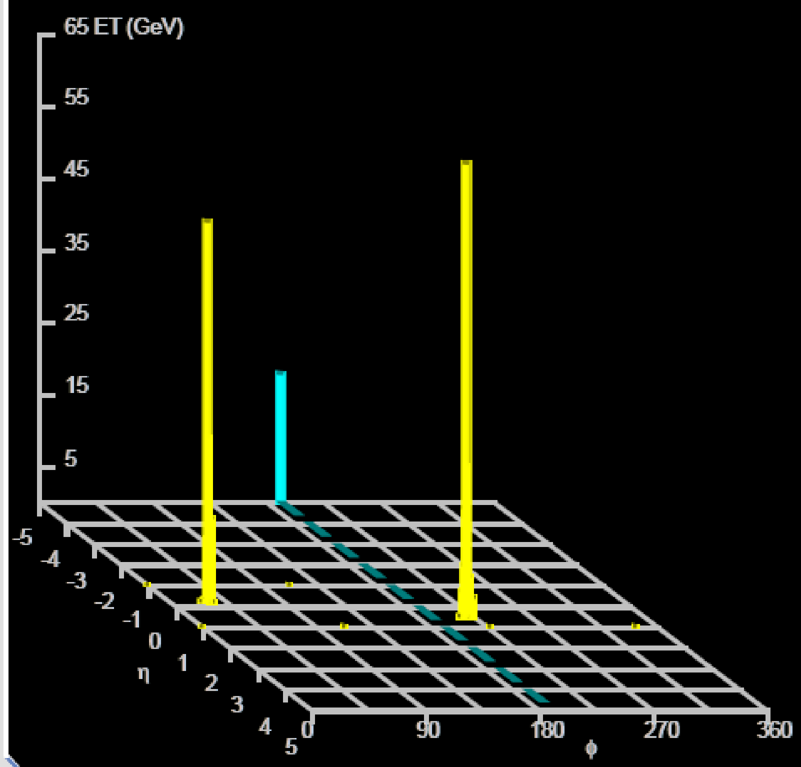
Reconstructed photon

Reconstructed track

MET

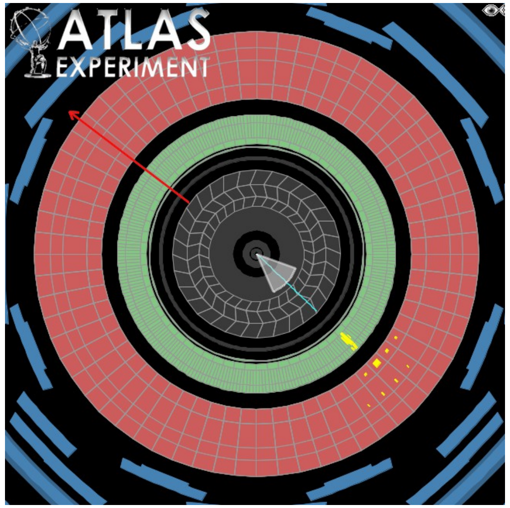
Simulated charged particle

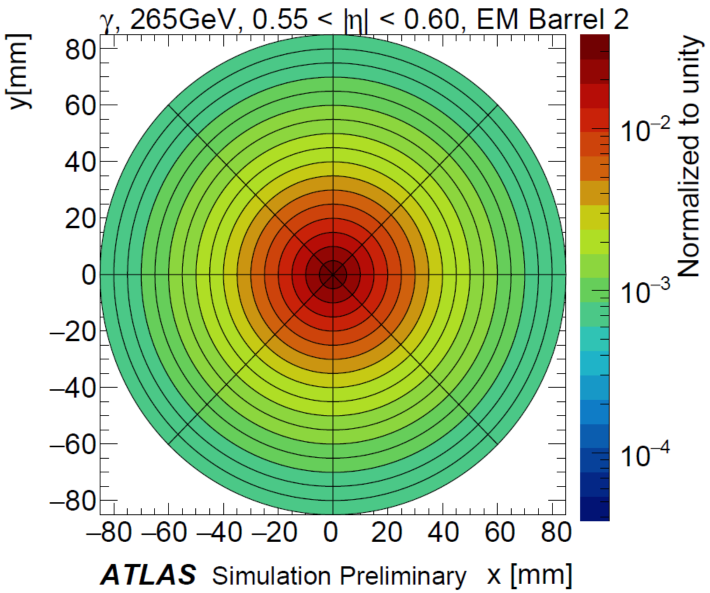
Simulated neutral particle

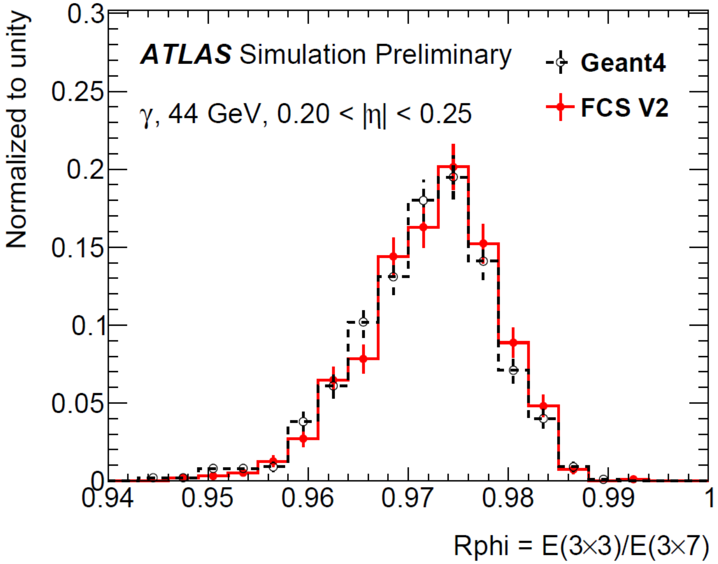


Conclusions

- Fast simulation is crucial for the future of ATLAS
- The showering in the calorimeter is the bottle neck for simulation
- FastCaloSim is based on parametrizations derived from G4
- Improved FastCaloSim: Based on latest G4 simulations, more modern, faster, more accurate!
- Hopefully will be used for the next big MC production campaign in ATLAS







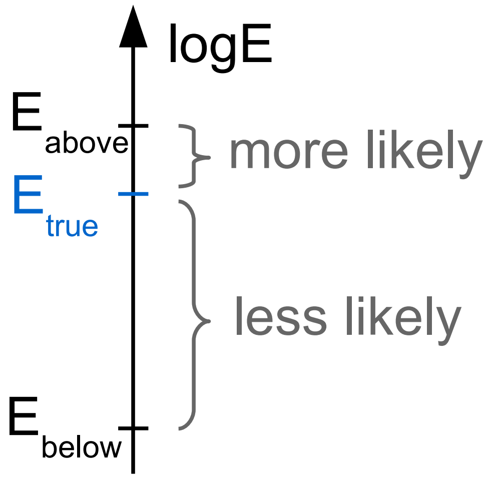
Backup

Interpolation between energy points

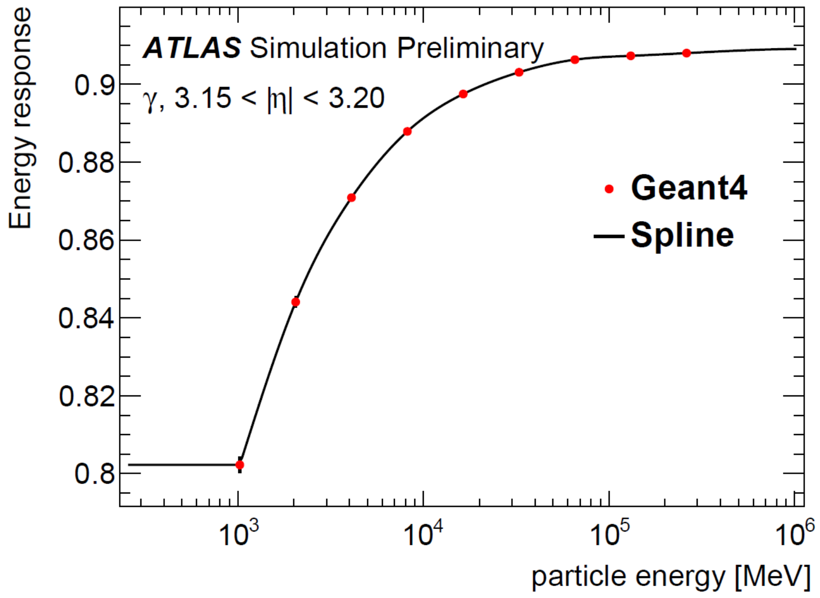
- G4 inputs simulated for fixed energies (17 points on a logarithmic scale, 64 MeV – 4 TeV)
- To simulate a truth particle with any energy value E_{true} :
  - The parametrisation picked is determined randomly, depending on $\log E_{true}$:

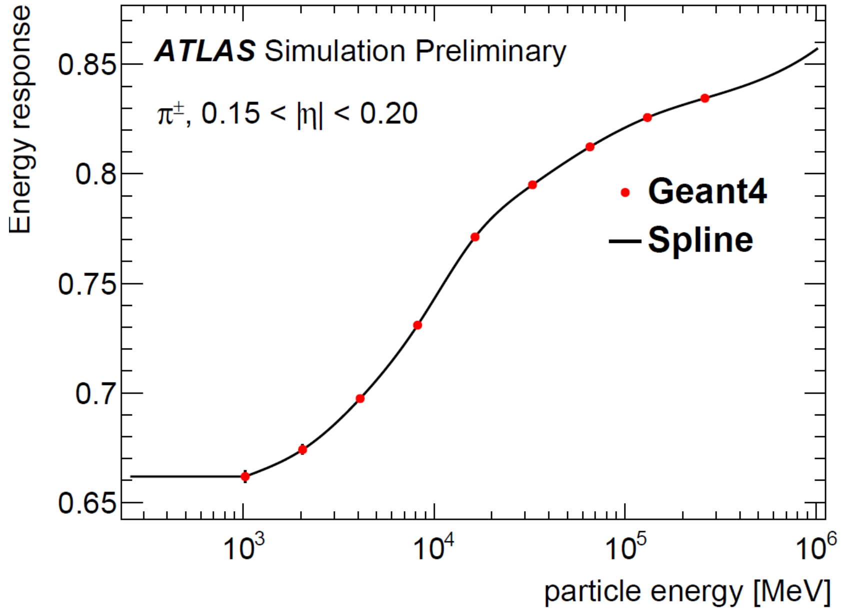
    - throw random number r [0,1]

    - if $((\log E_{true} - \log E_{below}) \,/\, (\log E_{above} - \log E_{below})) > r$ → choose above point
      otherwise, choose below point



  - The total energy response is interpolated using a spline:

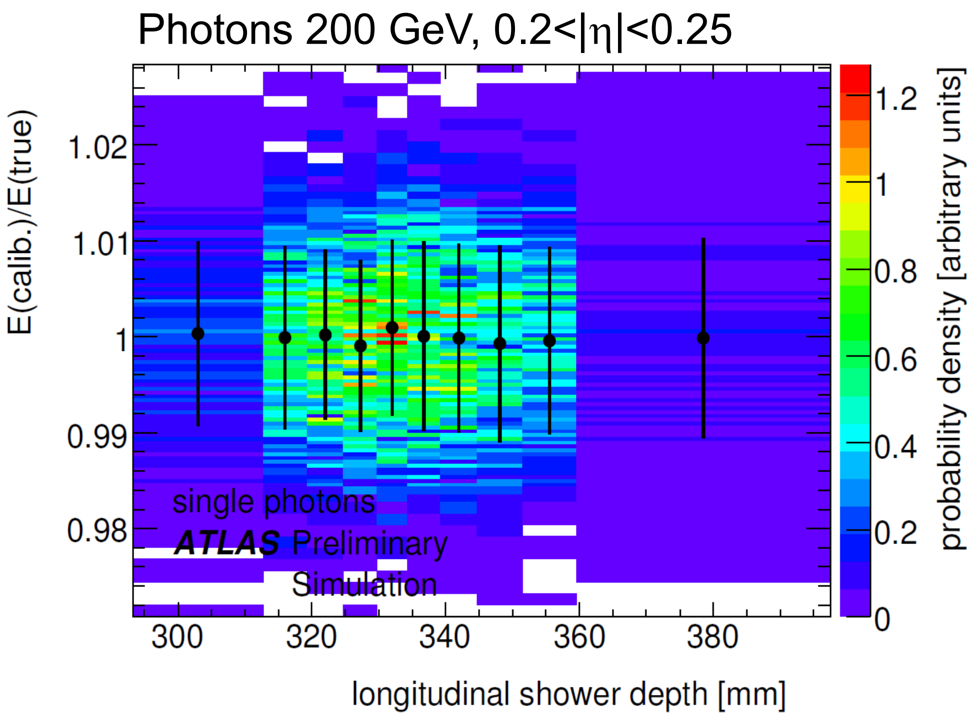




FastCaloSim V1 („old FastCaloSim“)

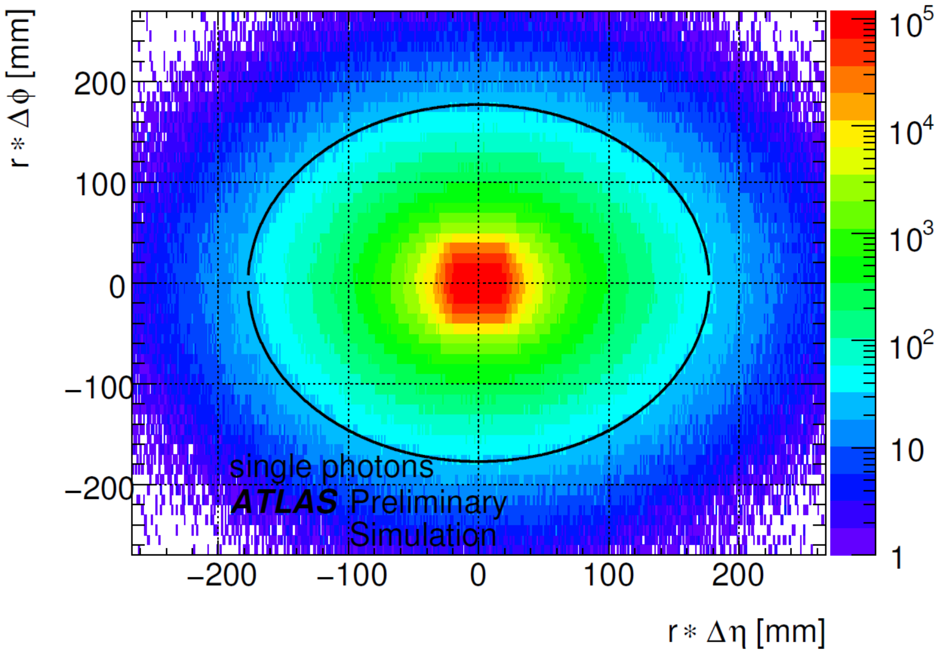
Longitudinal energy parametrisation:

- For each particle, energy and |η| store 2D histograms of energy vs. longitudinal shower depth (distance of the deposit from the calo surface), for total energy and energy fraction per layer
- Correlations between the deposits in each layer stored in correlation matrices
- Simulation: Randomly draw an energy value and energy fractions from the stored 2D histograms



Lateral shower shape parametrization:

- Radial symmetric function centered around the impact point of a particle in the calo layer, (3rd order polynomial function), modified with parameters to describe asymmetries when particles cross the calorimeter not perpendicular to the calo layer surface
- Parameters obtained from a fit to the Geant4 single particle lateral shape in each calo layer, for each particle type, energy, |η|, shower depth bin
- Good average shower description, poor modelling of substructure variables, no explicit FCAL parametrisation



Energy resolution in the calorimeter

$$\frac{\Delta E}{E} = \frac{\alpha}{\sqrt{E}} \oplus \beta \oplus \frac{\gamma}{E}$$

α: Sampling term (choice of active/passive material, fluctuations in number of charged particles passing through active layers)

β: Constant term (cracks, dead material, dominant at high energies)

γ: noise term (electronics, dominant at low energies)

ATLAS calorimeter design resolution:

	Resolution
EM Barrel	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \bigoplus 0.7\%$
EM End-Cap	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \bigoplus 0.7\%$
HEC	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \bigoplus 3\%$
FCAL	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \bigoplus 10\%$