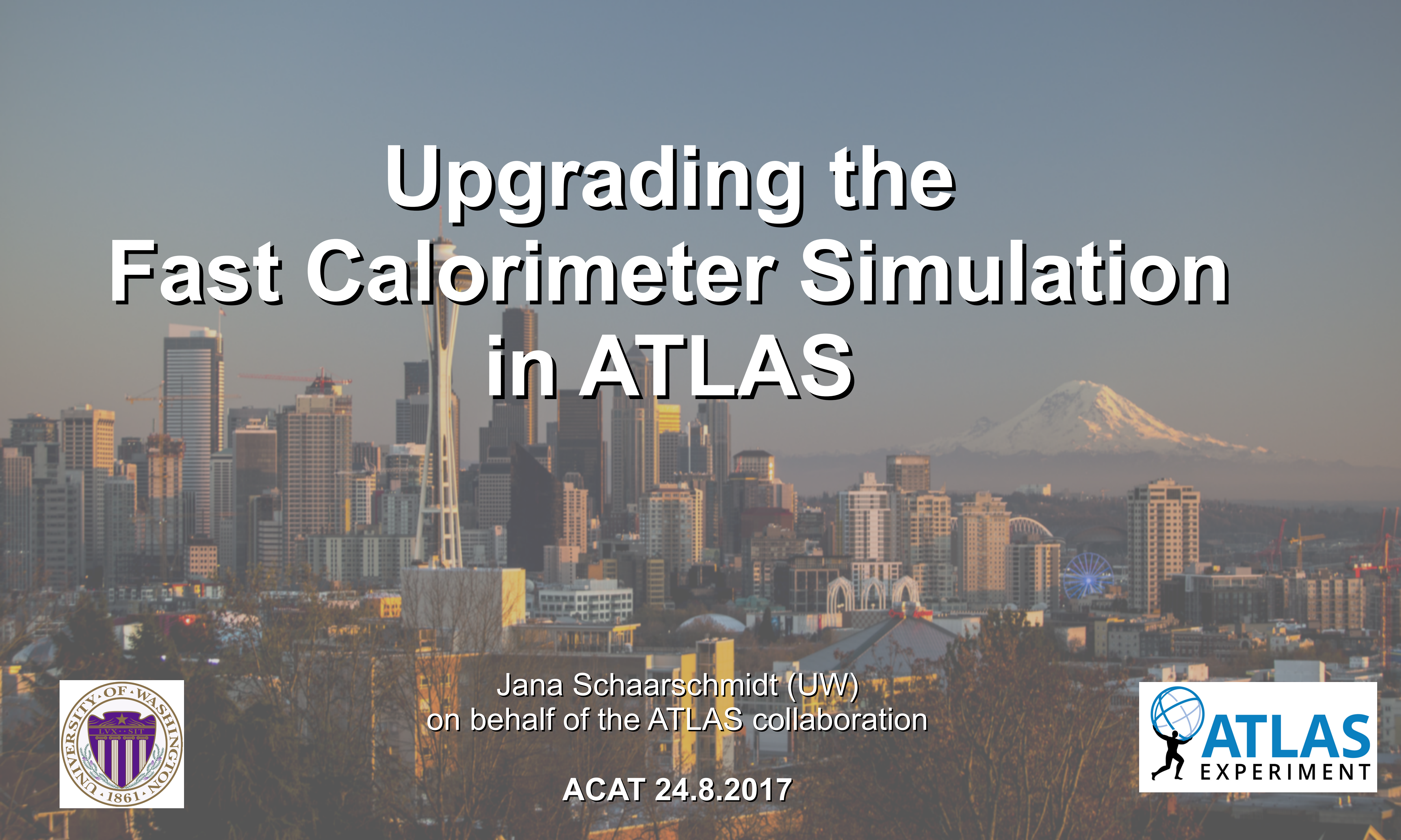
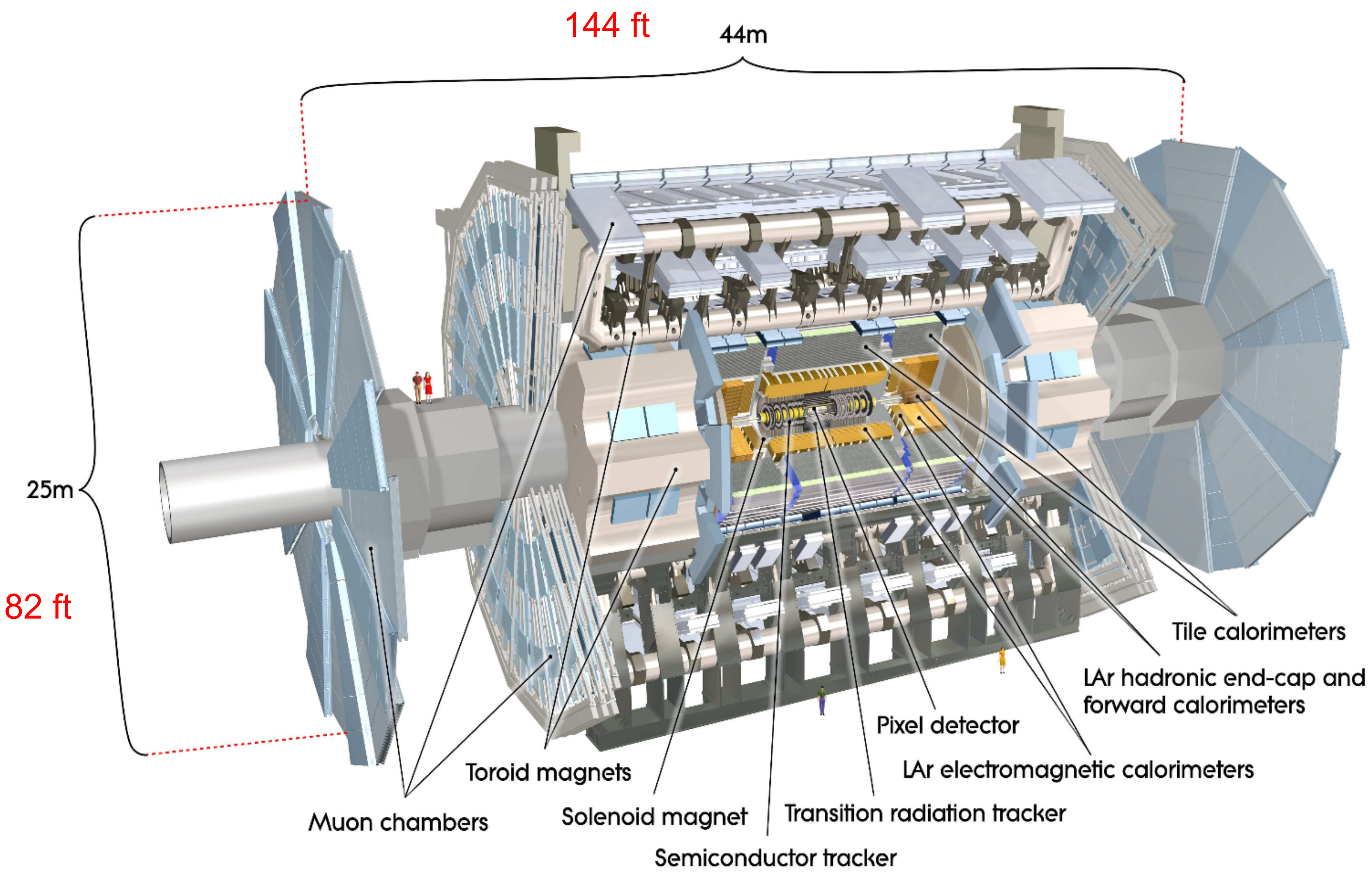


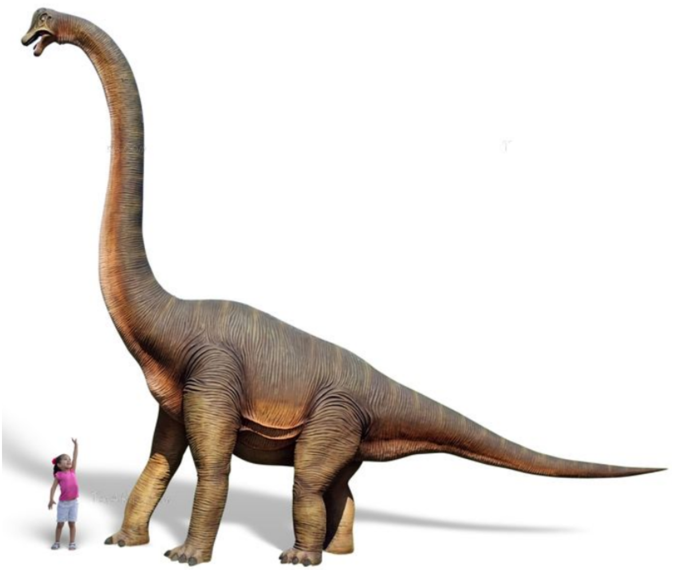
Upgrading the Fast Calorimeter Simulation in ATLAS

Jana Schaarschmidt (UW)
on behalf of the ATLAS collaboration

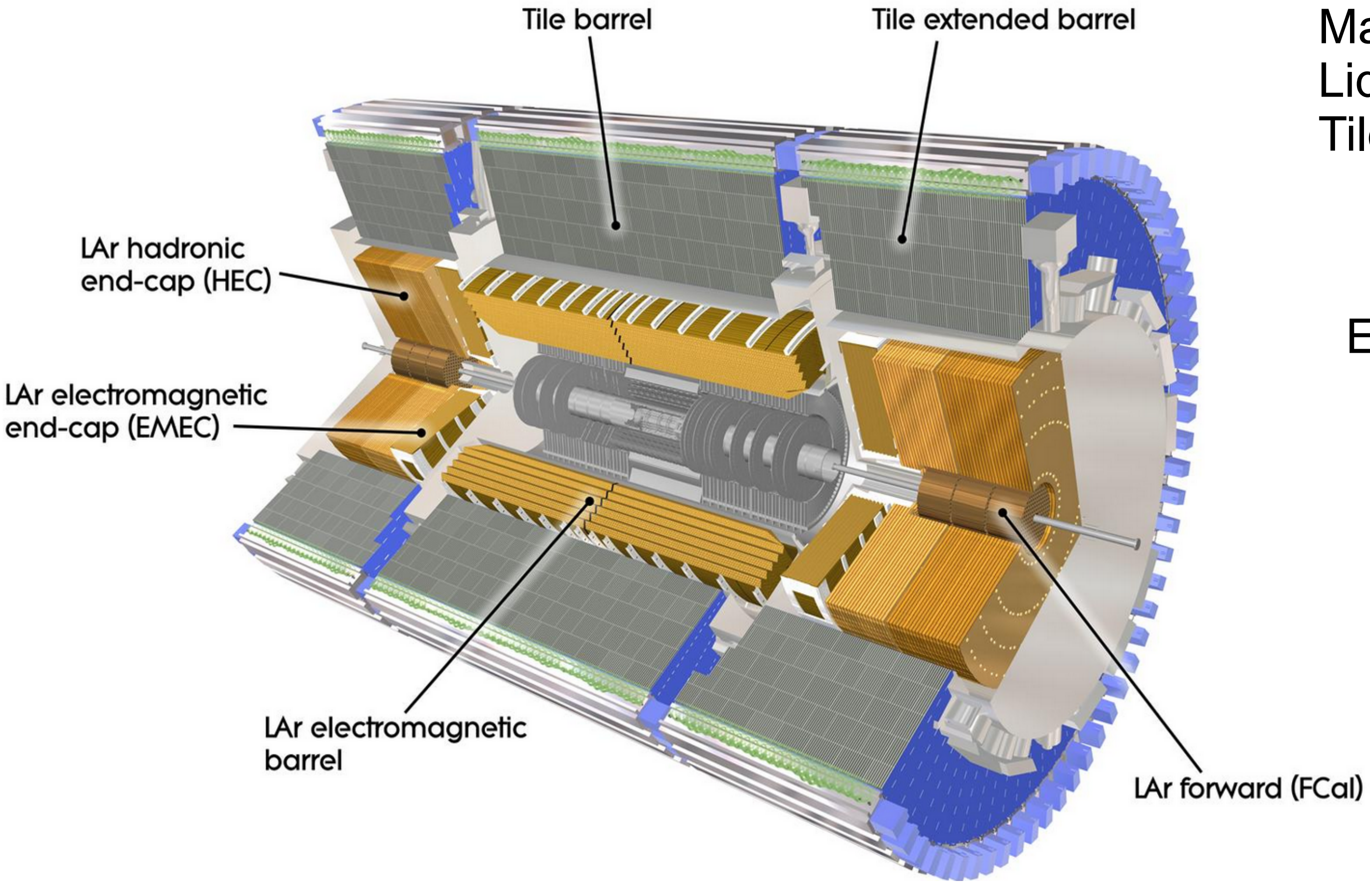
ACAT 24.8.2017

ATLAS Detector at the LHC





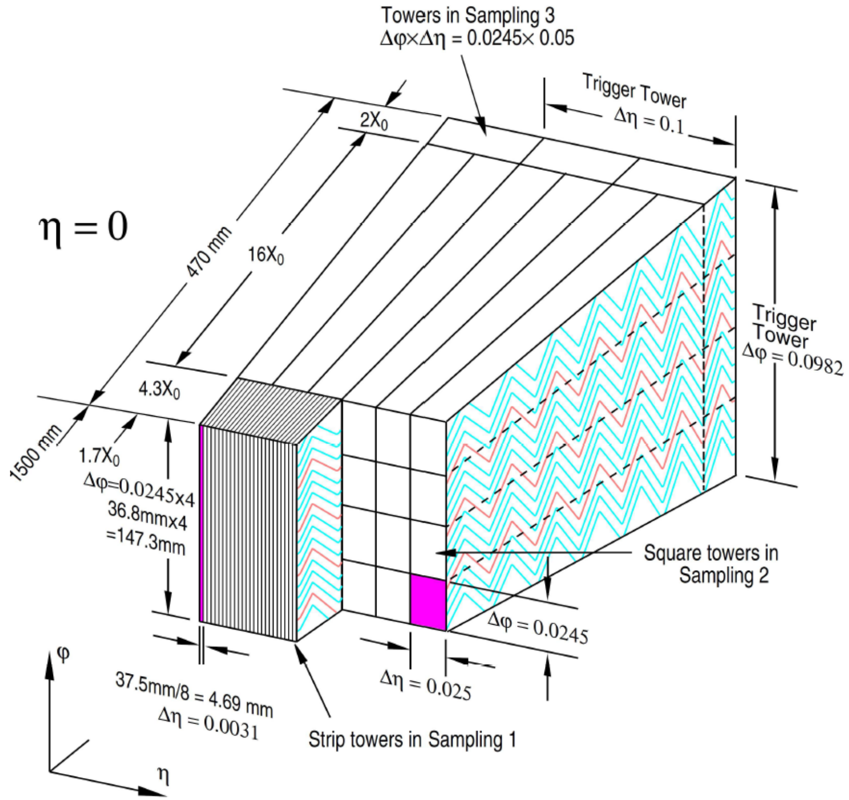
The ATLAS Calorimeter System



Covering |η|<4.9

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

EM Barrel:



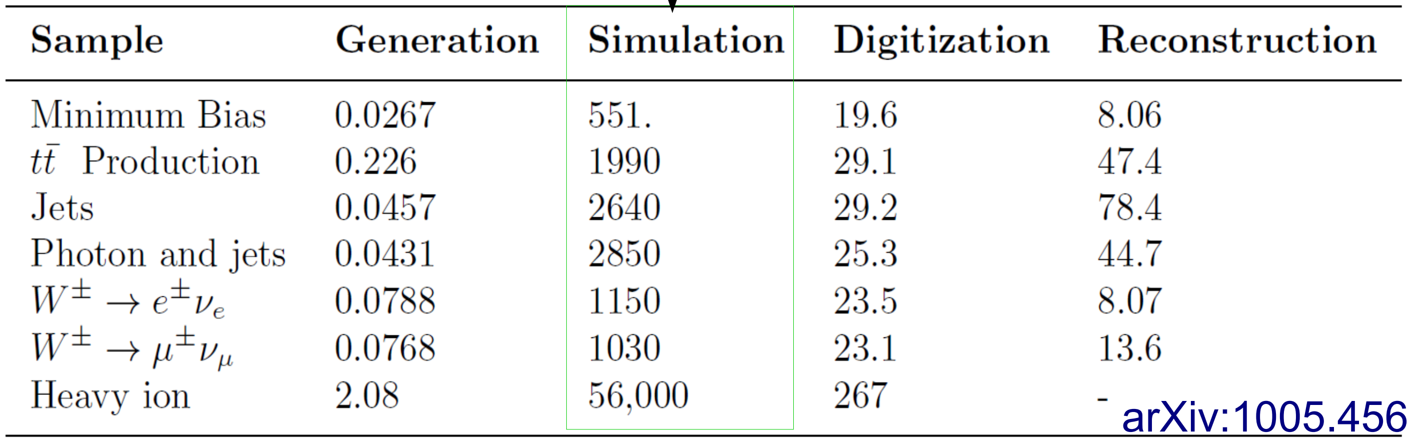
\# Readout channels: ~190 k in total
\# Samplings (layers of cells): 24

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

Crucial for photons & electrons, jets and missing energy reconstruction

The Need for Fast Simulation

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

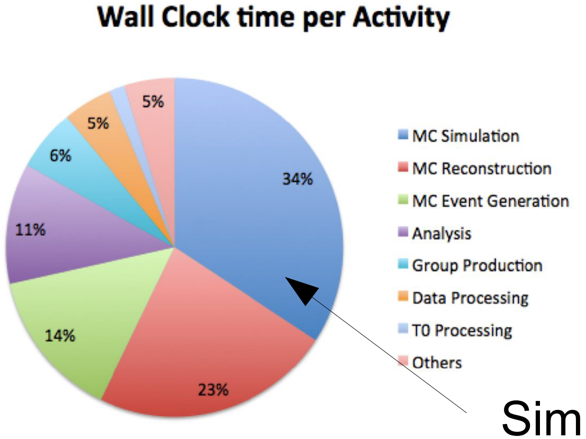


Sample	Generation	Simulation	Digitization	Reconstruction
Minimum Bias	0.0267	551.	19.6	8.06
$t\bar{t}$ Production	0.226	1990	29.1	47.4
Jets	0.0457	2640	29.2	78.4
Photon and jets	0.0431	2850	25.3	44.7
$W^{\pm} \rightarrow e^{\pm}\nu_e$	0.0788	1150	23.5	8.07
$W^{\pm} \rightarrow \mu^{\pm}\nu_{\mu}$	0.0768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

arXiv:1005.4568

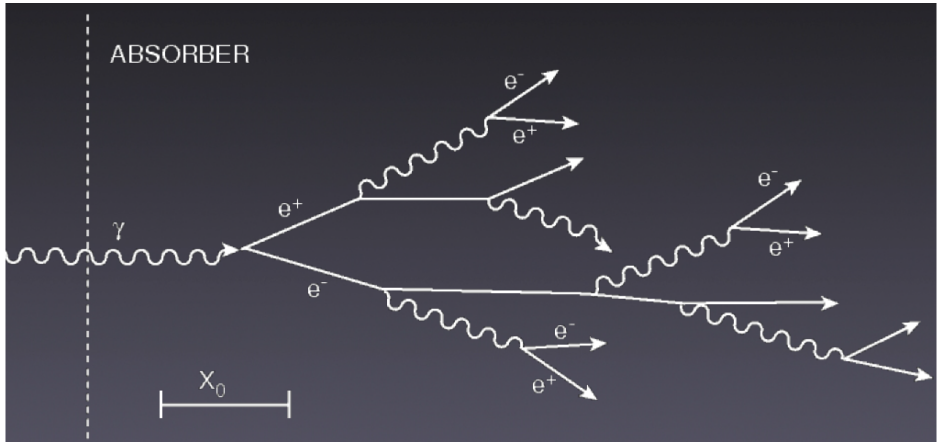
85% of the simulation time is spent in the calorimeters (showering)

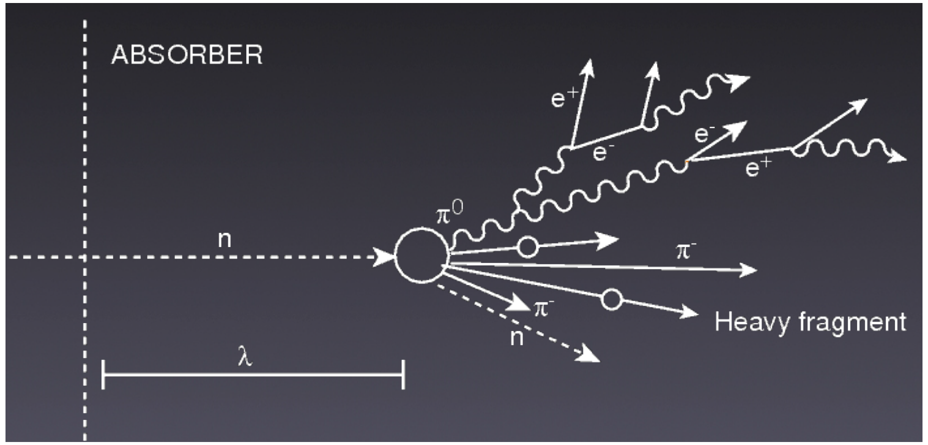
Grid usage 2016:



Simulation!

EM shower:





Hadronic shower:

→ *Wider, slower, larger fluctuations than EM showers*

- Geant4 is slow, but most accurate. It is the ultimate reference for simulation.
- But ATLAS relies on Fast Simulation, even more in the future: The resources/funding does 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.

The Simulation Principle (FastCaloSim)

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

Electrons and Photons:	For the electromagnetic showers
Charged Pions:	For the hadronic showers

Geant4 single particle simulation → Parametrizations → Fast Simulation

Some technical details:

- single particle simulation without calo noise and cross talk

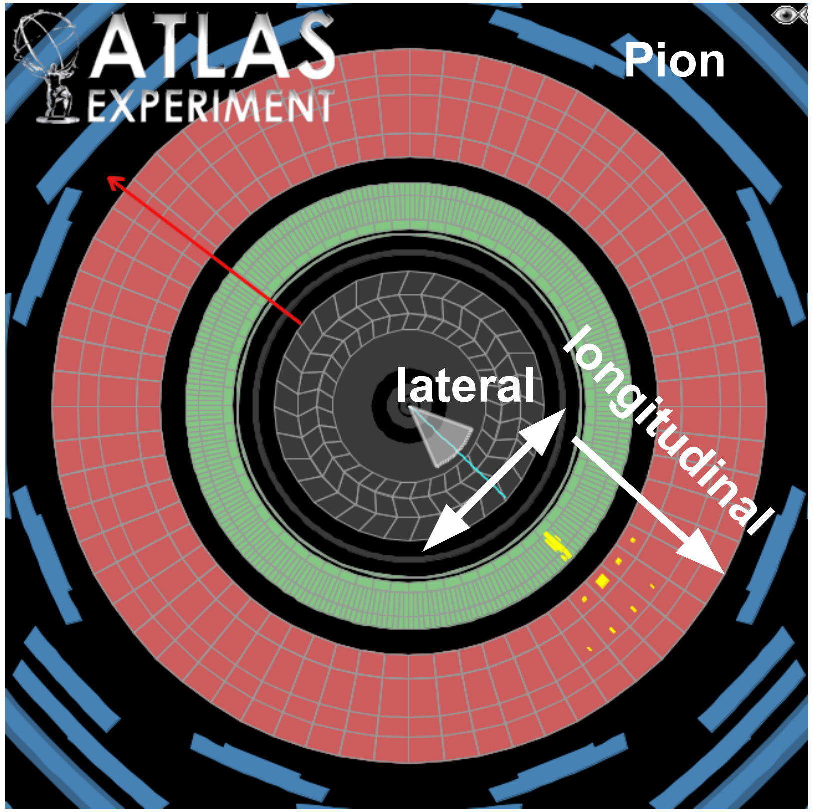
- single particles starting at the calorimeter surface
  (and not at the detector centre)

- Eta grid: 100 bins in size of 0.05 covering 0-5.0
  (averaging over left and right detector sides)

- Energy grid: 17 bins from 60 MeV – 4.2 TeV
  (exponential spacing)

→ 5k parametrizations

- Muons rarely shower → use full simulation (Geant4)



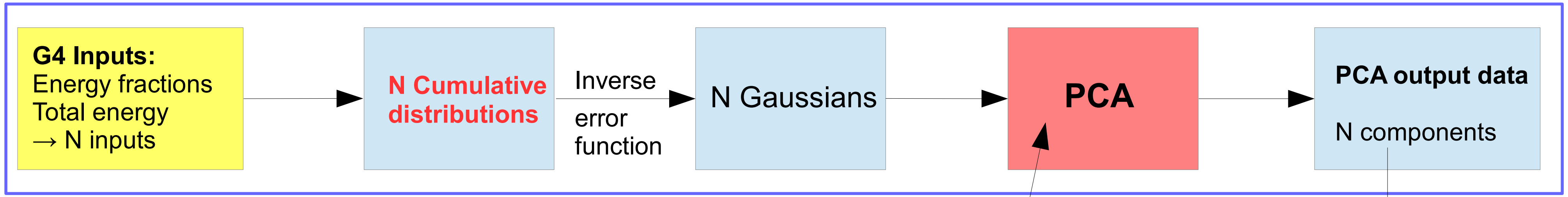
Longitudinal Energy Parametrisation

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

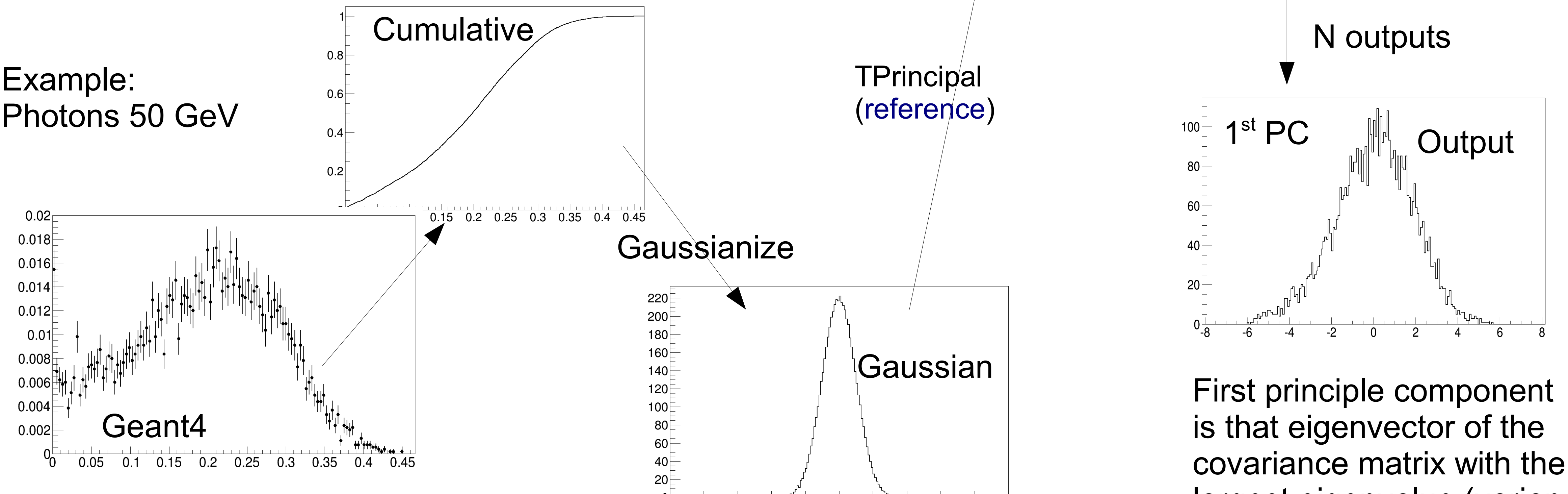
Inverse error function

N Gaussians

PCA

PCA output data

N components



Example:
Photons 50 GeV

TPrincipal
(reference)

E fraction in cell layer 1

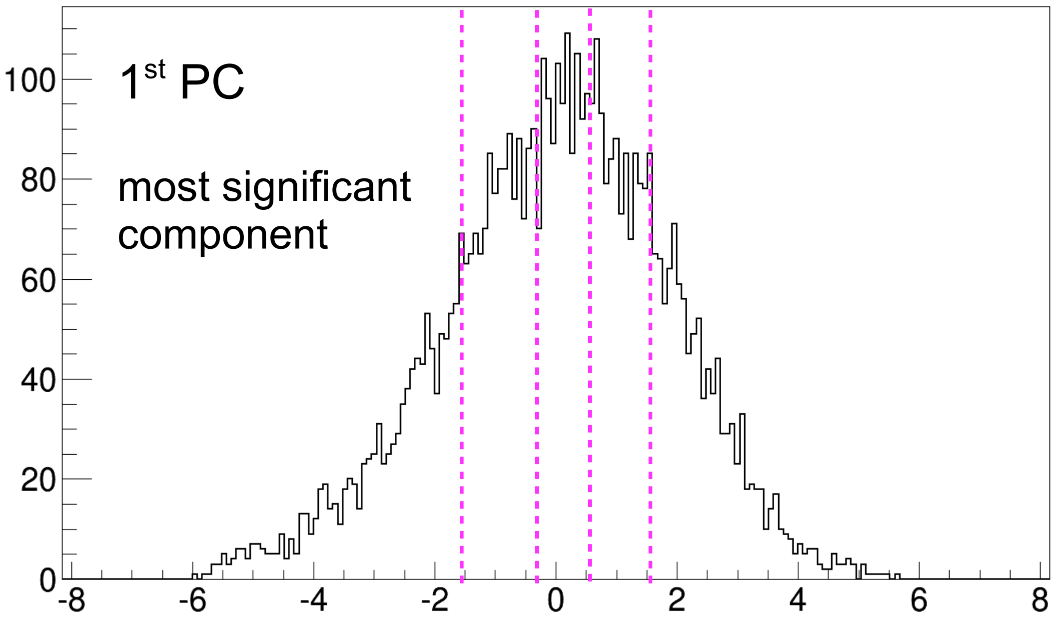
First principle component is that eigenvector of the covariance matrix with the largest eigenvalue (variance)

Longitudinal Energy Parametrisation

The first and second principal component are used to divide the input data into quantiles:

These “PCA bins”
are also used to derive
the shape parametrisation.

The 1st PCA chain is used
only to derive this binning.



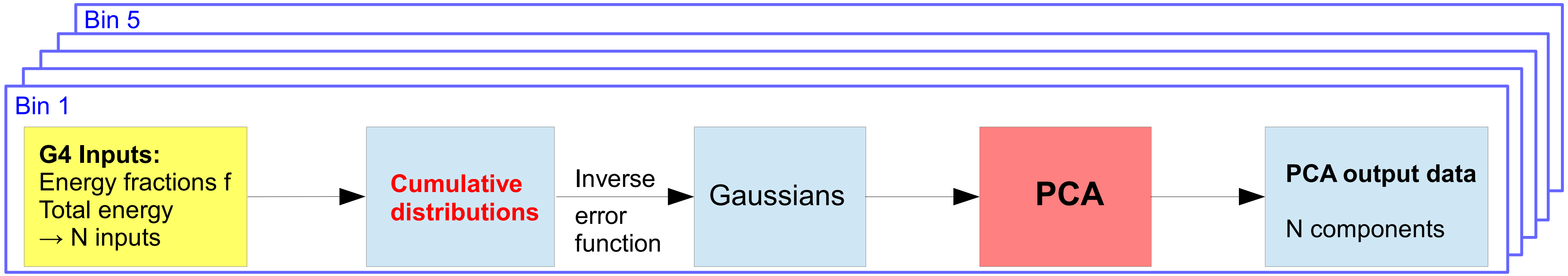
2nd binning in the 2nd
component (not shown)

→ typically 10 bins

→ 50k parametrizations
in total

Showers get classified

2nd PCA chain (another PCA, but now in each of the “PCA bins from the 1st chain”):



G4 Inputs:
Energy fractions f
Total energy
→ N inputs

Cumulative distributions

Inverse error function

Gaussians

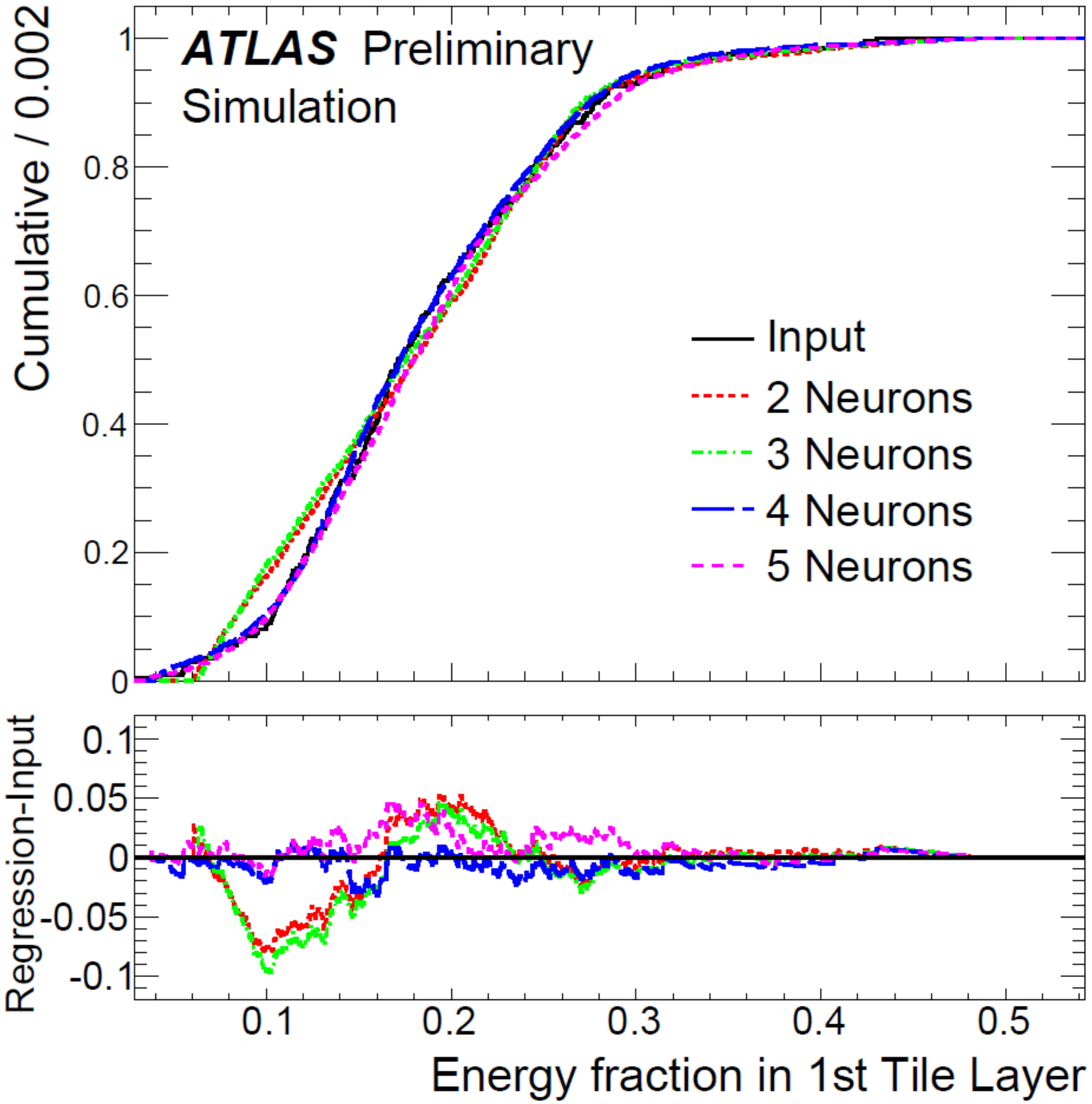
PCA

PCA output data

N components

Output in each bin is a set of N linearly uncorrelated, Gaussian-shaped distributions
To store: Cumulative distributions, PCA matrices, mean and RMS of the output Gaussians

Memory Optimisation



Number of weights: 3n+1 (n #neurons)

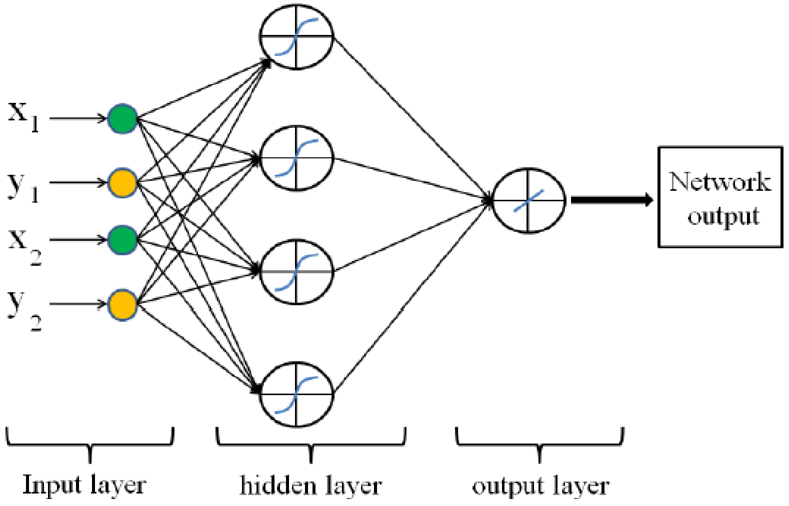
The entire parametrisation is loaded into memory.

The memory space is limited to 2GB (standard CERN grid machines).

FastCaloSim will run together with other programs in the FastChain → 1GB available.

We don't store histograms,
we store MLP regression weights.

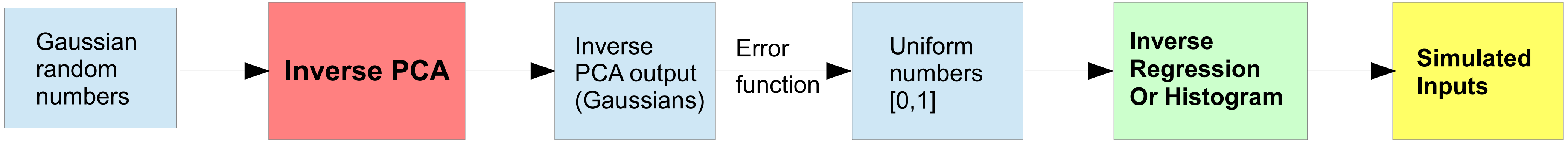
Training a multilayer perceptron (MLP) with 1 hidden layer and as few neurons a possible (iteration).



Longitudinal Energy Simulation

3. Step: **Simulation**

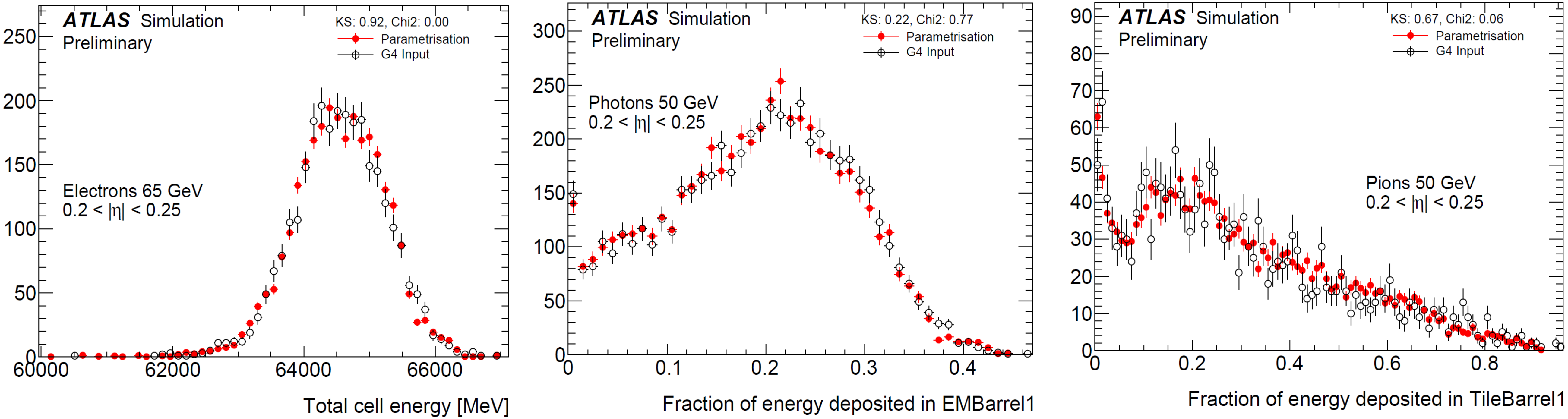
Reverse chain in a given PCA bin:



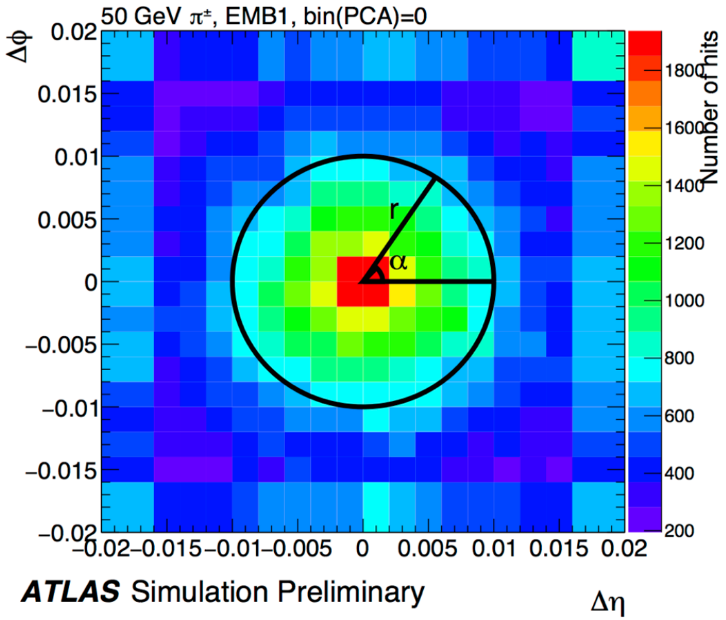
Gaussian random numbers → **Inverse PCA** → Inverse PCA output (Gaussians) → Error function → Uniform numbers [0,1] → **Inverse Regression Or Histogram** → **Simulated Inputs**

For each simulation call, the PCA bin is randomly determined, all PCA bins have the same probability by construction.

Validation (simulation only, no digi or reco):



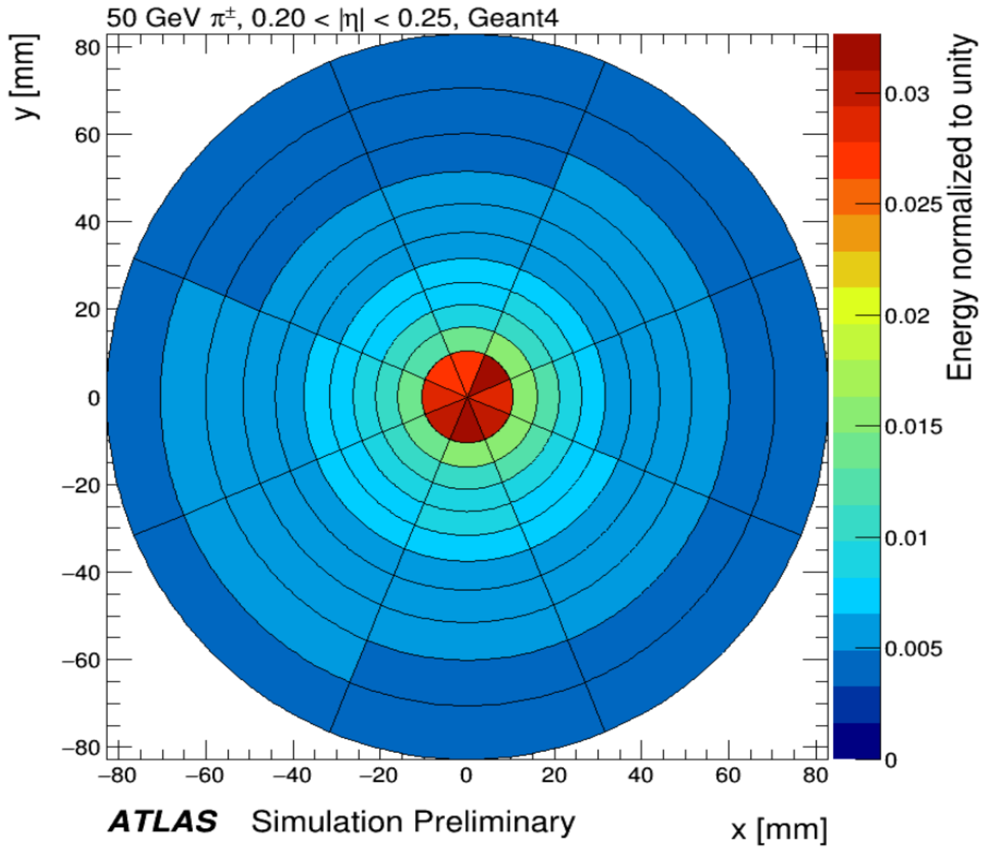
Lateral Energy Parametrisation



2D lateral shower shape (perpendicular to shower axis) described as a function of

- distance from shower axis
- angle around shower axis

Summed shower shape from thousands of single particles (50 GeV pions) in one PCA bin:



Convenient binning of the energy is crucial for the parametrisation

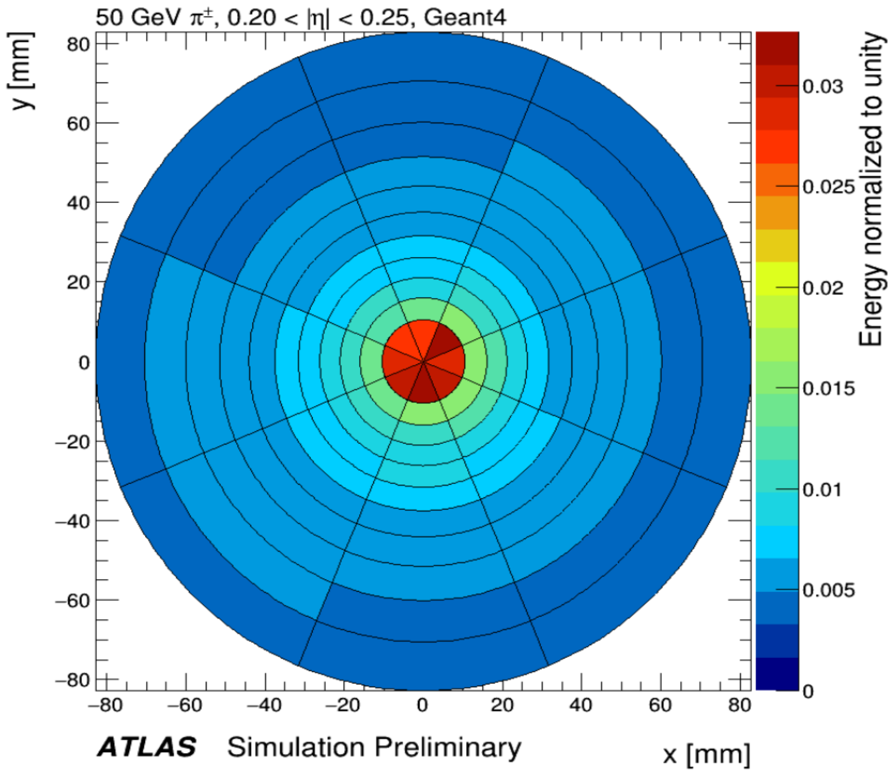
The energy deposits from Geant4 are given as “hits”

Hit: Energy deposit at a given position

Hit grid: 1mm distance in all directions

Binning decided to have roughly same number of hits per bin.

Lateral Energy Parametrisation



2D Histograms stored and loaded into memory (to be replaced by MLP regression as well)

2D energy histogram treated like a PDF (probability density function)

- Randomly sample hit positions from that PDF (do this n times → N_{hits})
- Assign energy to hits: $E_{hit} = E_{layer} / N_{hits}$

The number of hits is an important parameter, calculated from expected energy resolution:

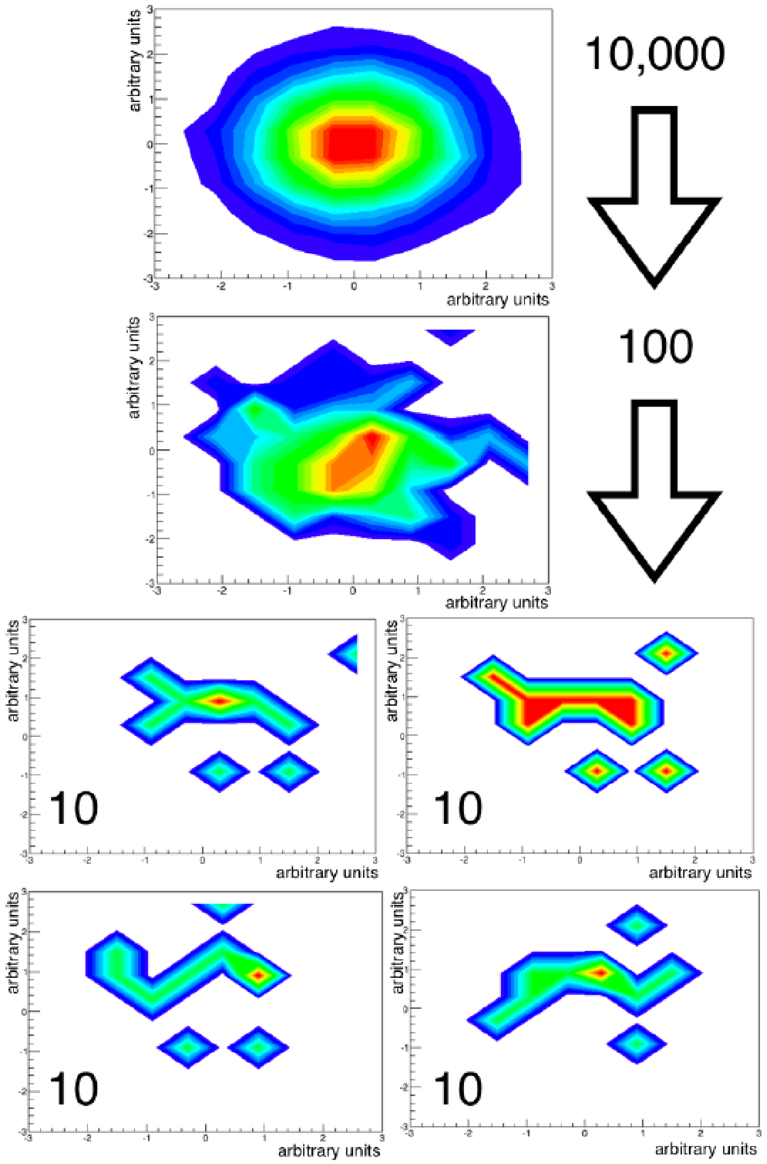
$\mathrm{sqrt}(N_{hits})/N_{hits} = \alpha / \mathrm{sqrt}(E)$ (sampling term of the energy resolution, α depends on the layer)

The final step is to assign the sampled hits and energies to the actual calorimeter cells.

Use simplified geometry and hashing for the cell look-up.

Increasing Hadronic Shower Fluctuations

Reducing the number of hits to increase fluctuations:



In hadronic showers, deposited energy can vary a lot from shower to shower:

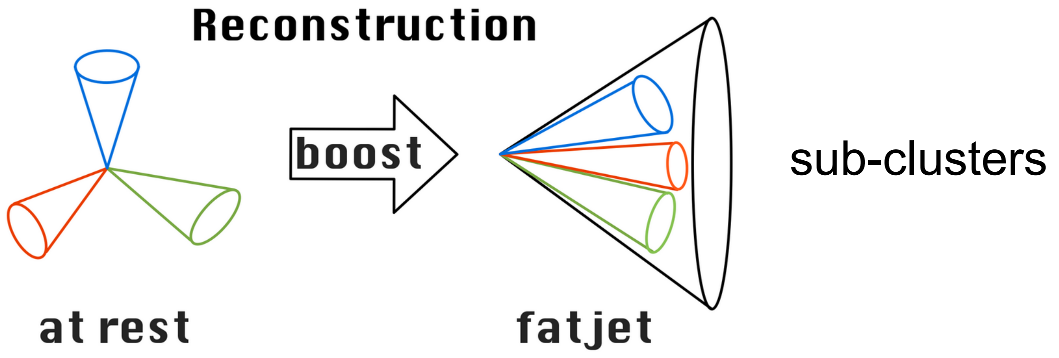
- presence of invisible energy (eg. nuclear excitation)
- varying fraction of EM shower component

A realistic simulation has to reproduce such fluctuations.
A complete theory is difficult, therefore the FastCaloSim approach is to increase energy fluctuations by

- reducing the number of hits sampled from the average shower shape, such that energy sub-clusters are reconstructed

- or even tuning the number of hits such that we reproduce G4 fluctuation (ie. RMS of the energy as a function of distance from shower axis)

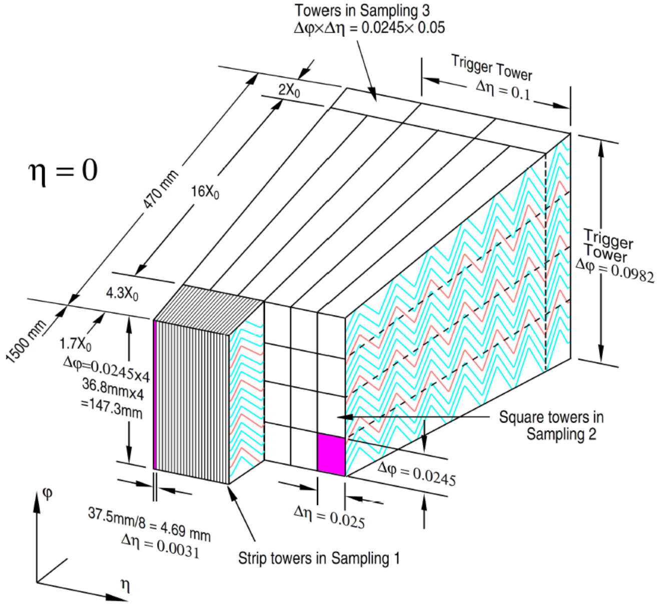
- Presence of sub-clusters heavily linked with ability to model boosted tolopogies:

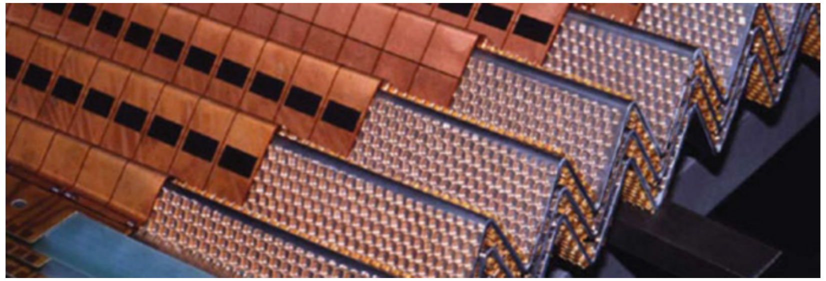


Simplified Geometry and FCAL Geometry

Final step of the simulation: Assigning hit energies to calorimeter cells (`CaloCellContainer` object)

- Simplified geometry tool used in the Liquid Argon calorimeter parts:
  Returns cell identifier given a hit position, based on the assumption that the cells are cuboids (neglecting accordion)
- Much faster than loading the full geometry as used in G4

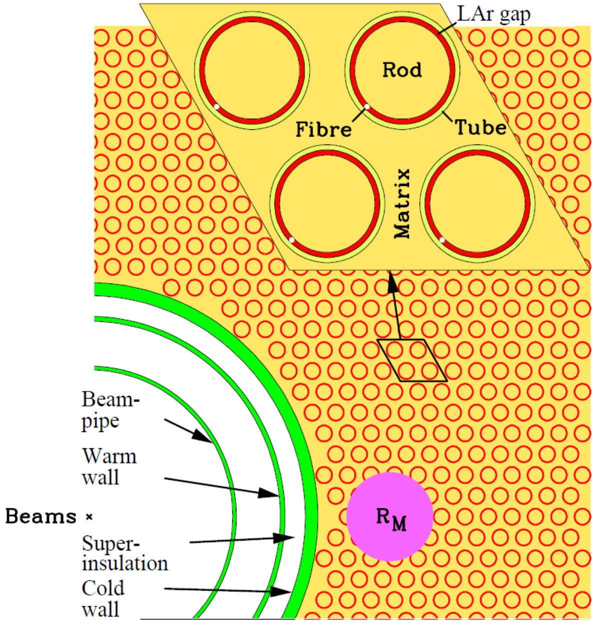


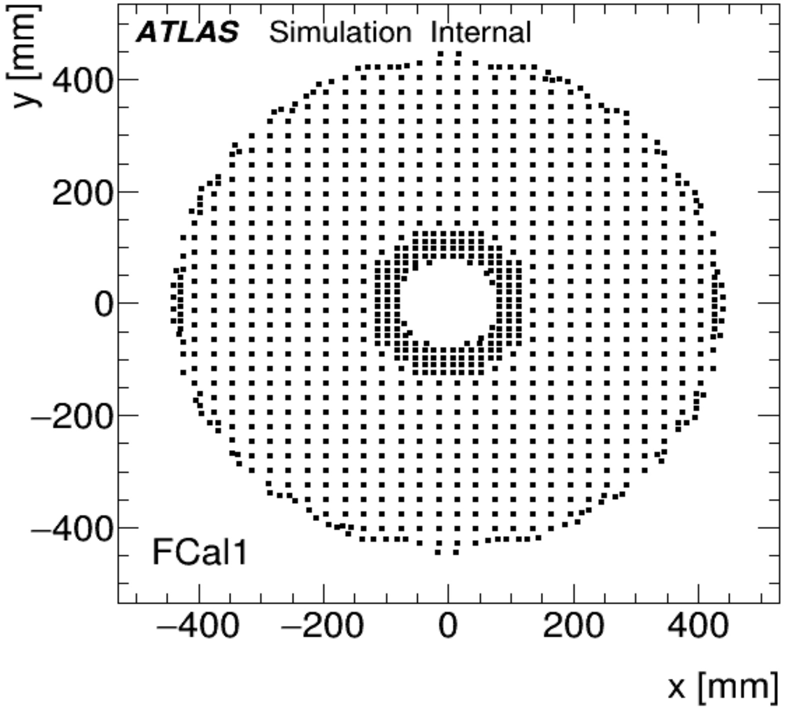


Accordion structure to avoid projective azimuthal cracks that would increase the constant term.

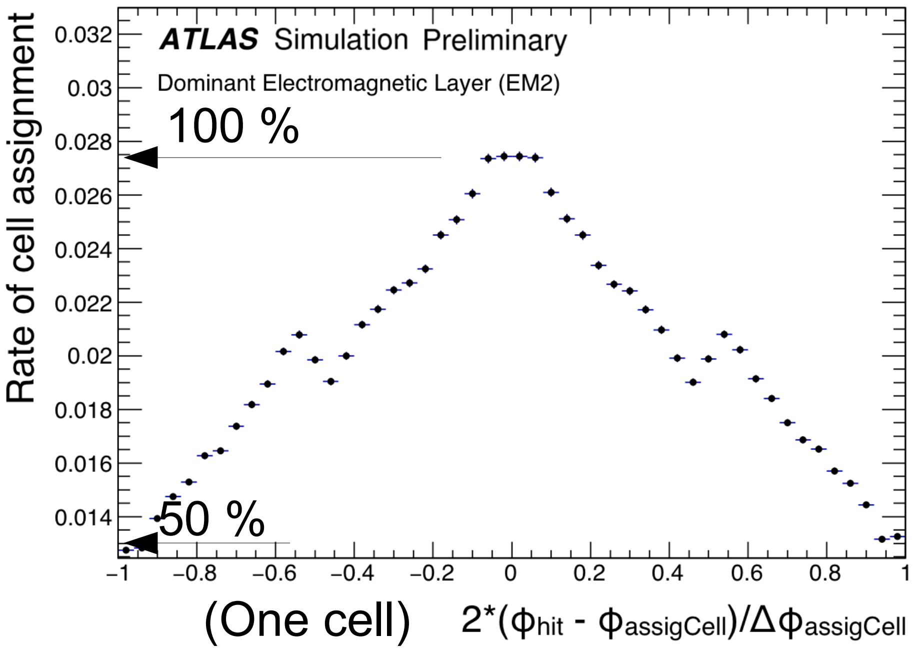
- FCAL geometry is different, cylinders arranged in rhombus-like patterns, no cuboids
- Use exact geometry information

FCAL readout channel positions:





Emulating the Accordion Structure



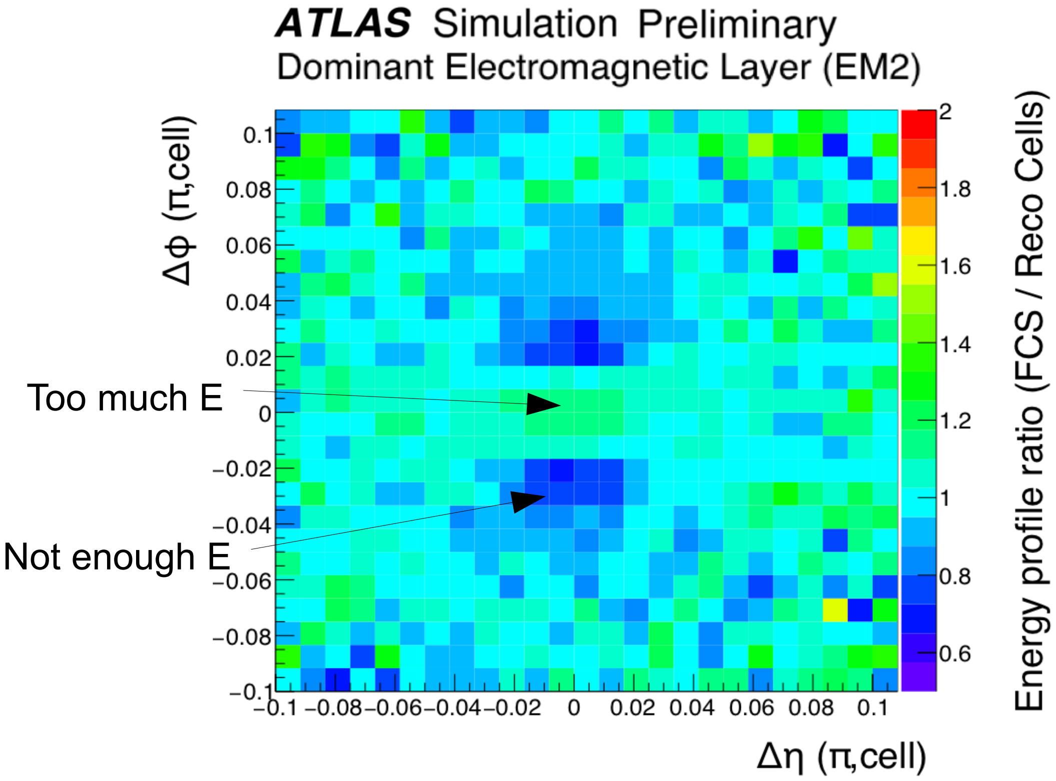
“Wiggle function” (hit displacement):

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

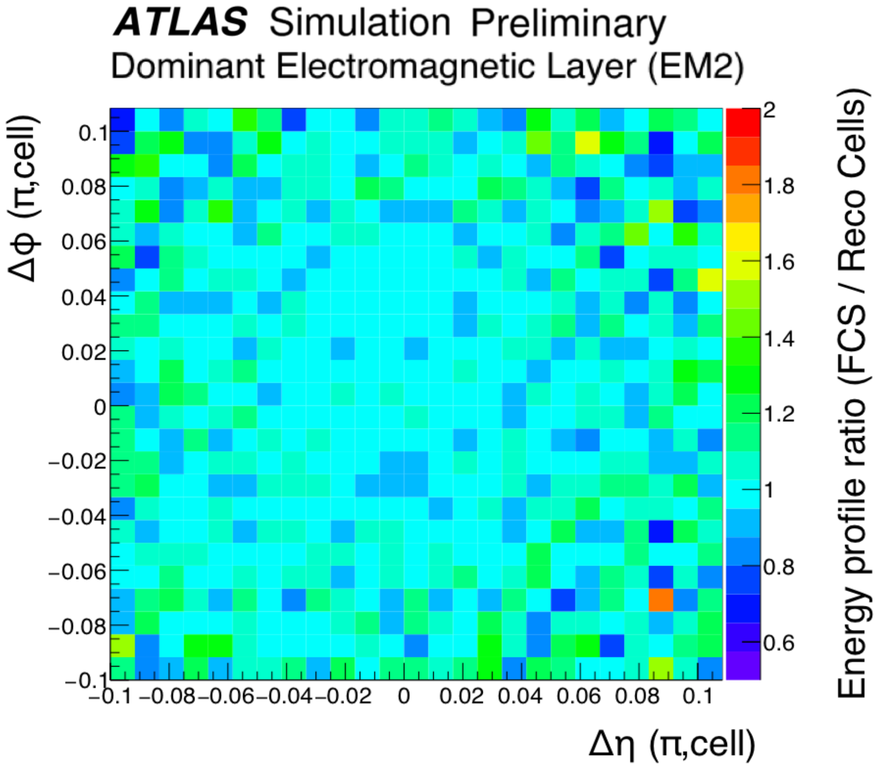
→ Good closure compared to the Geant4 inputs, and nice flatness in phi, when wiggle function is activated

Ratio of energies as a function of distance of the hit from the pion shower center:

Without wiggle



With wiggle

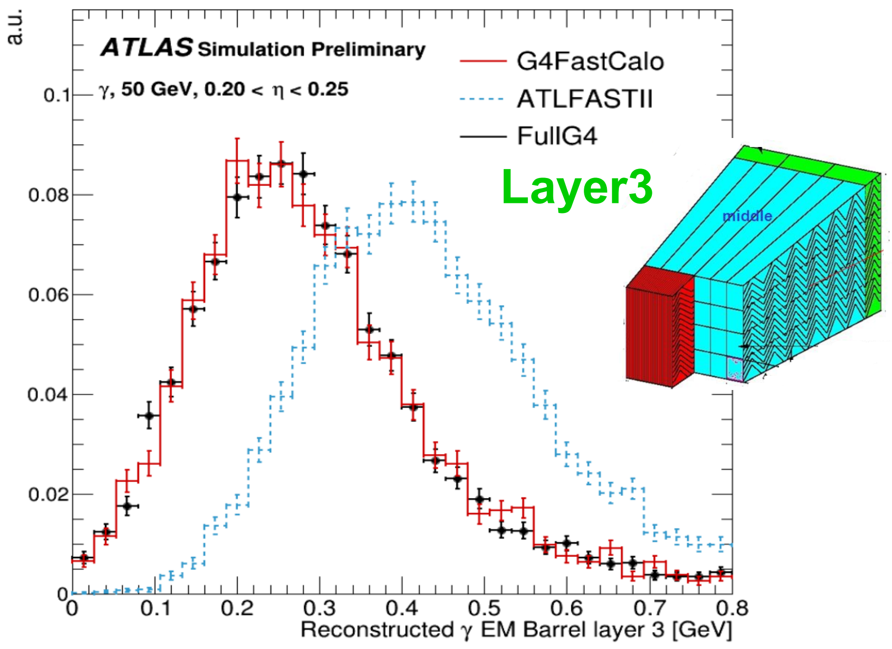


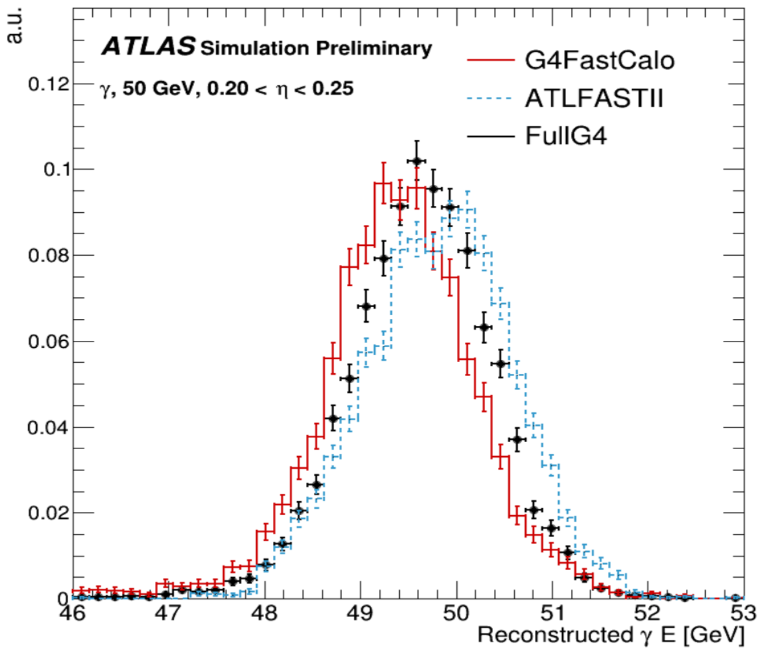
Reco cell: Geant4 cell
FCS cell: assigned cell

Prototype Validation: EM Showers (Single Photons 50 GeV)

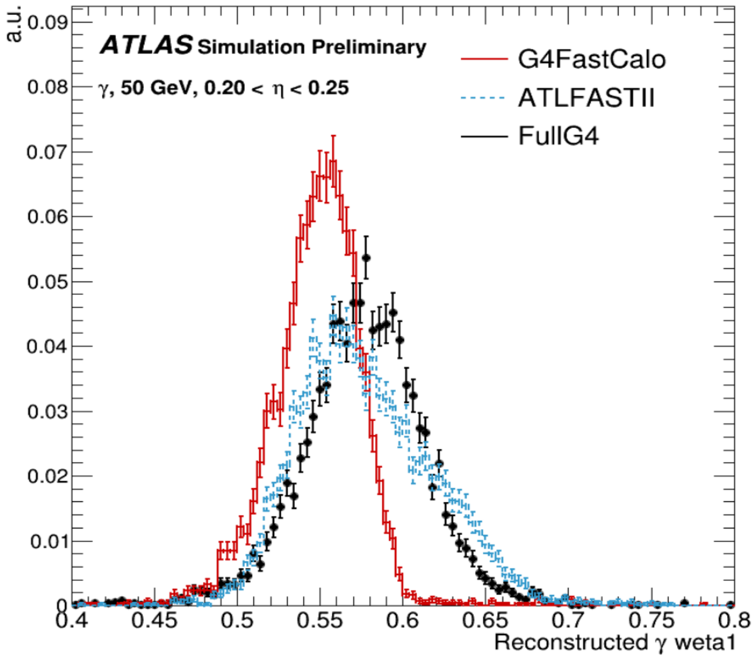
After full chain: Simulation → Digitization → Reconstruction

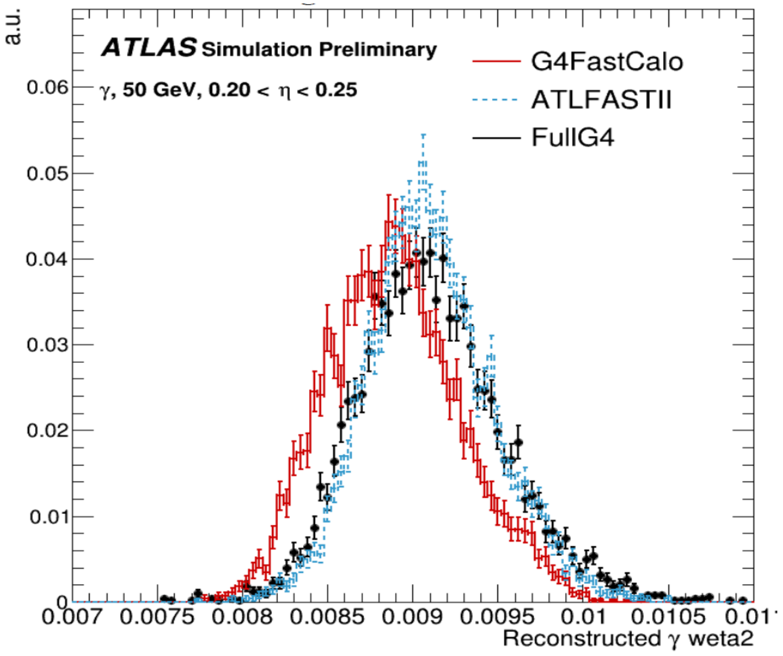
Legend:
Improved FCS
Previous FCS version
Geant4





Reconstructed energy depends on correct modelling of the shape





weta1:

Energy weighted shower width in layer1 in ±3 strips around center

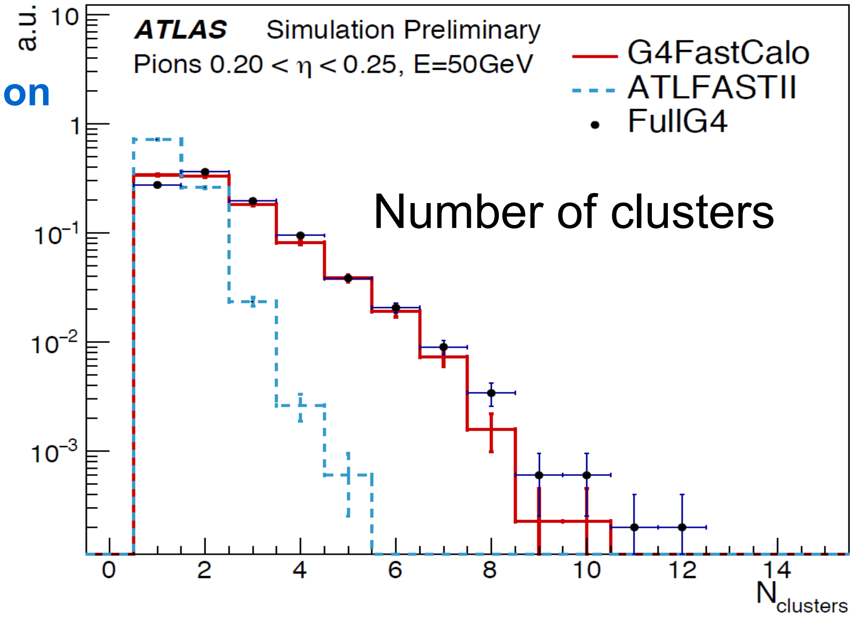
weta2:

Energy weighted lateral width in layer2 in 3x5 cell window

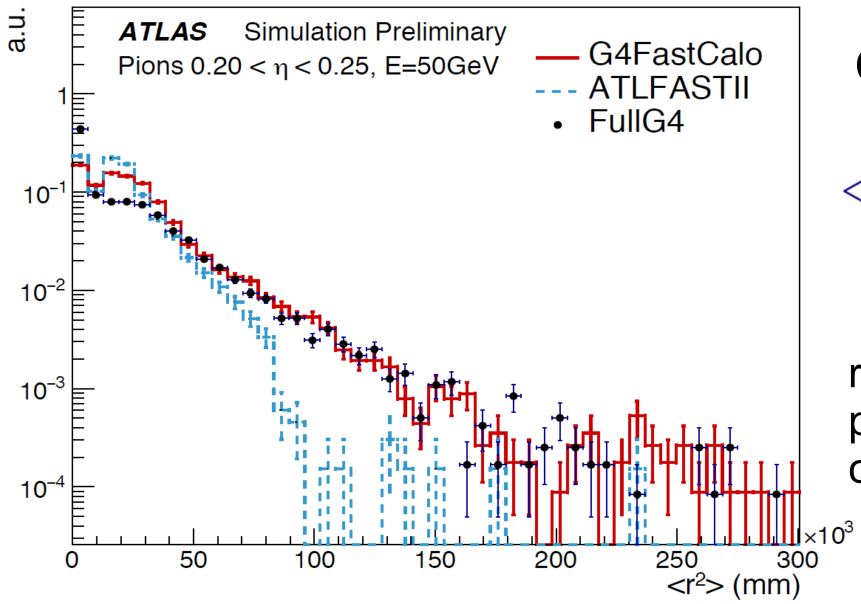
$$\omega_{\eta 2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i}\right)^2}$$

Prototype Validation: Hadronic Showers (Single Pions 50 GeV)

Legend:
Improved FCS
Previous FCS version
Geant4



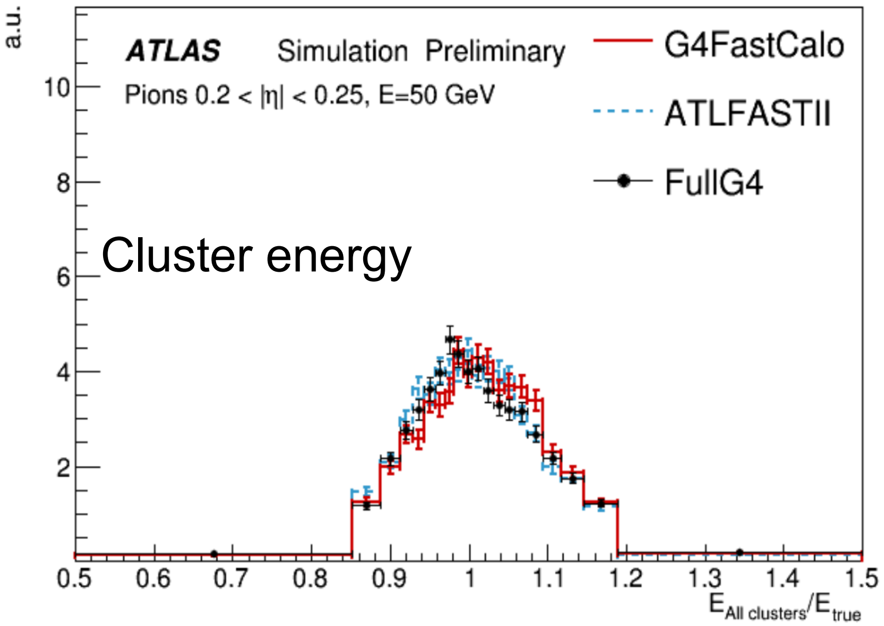
Number of clusters



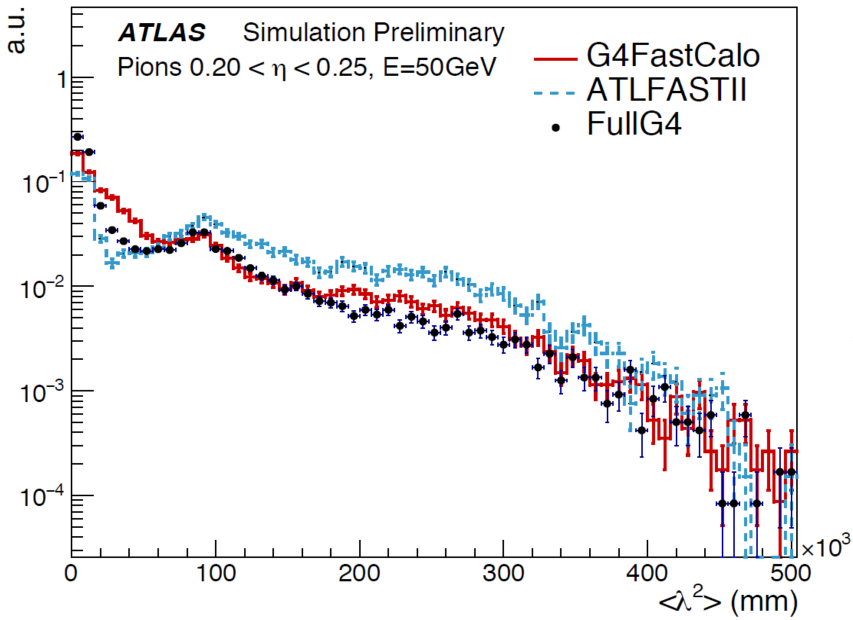
Cluster moments:

$$< x^n > = \frac{\sum_{E_i>0} E_i x_i^n}{\sum_{E_i>0} E_i}$$

r: Energy weighted perpendicular distance of cell from shower axis



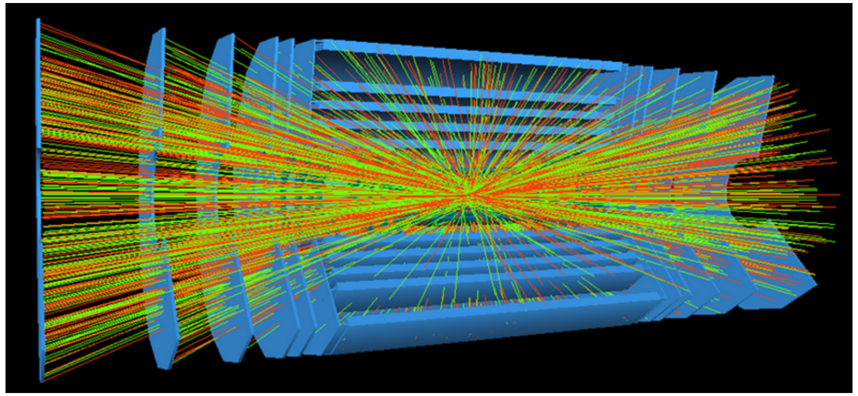
Cluster energy



λ: Energy weighted distance of cell from the shower center along shower axis

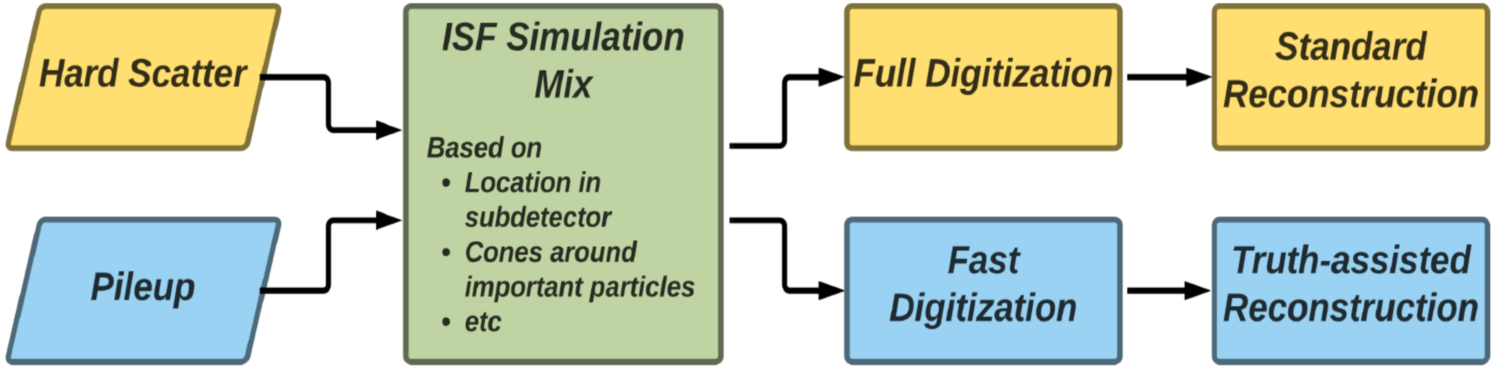
The ATLAS Fast Chain

In Run-3 and with high lumi-LHC, fast simulation will constitute the majority of simulation, but it alone will not be enough.
Digitization and reconstruction also have to be simplified
(and those don't scale linearly).



(230 pile-up events)

FastChain is a **flexible** set of tools to speed up the entire production chain:



Tools under development:

- FastCaloSim
- Fast Tracking (FATRAS)
- Fast Digitisation in Pixel, SCT & TRT
- Pile-up on the fly, out-of-time pile-up emulation
- Truth-assisted reconstruction

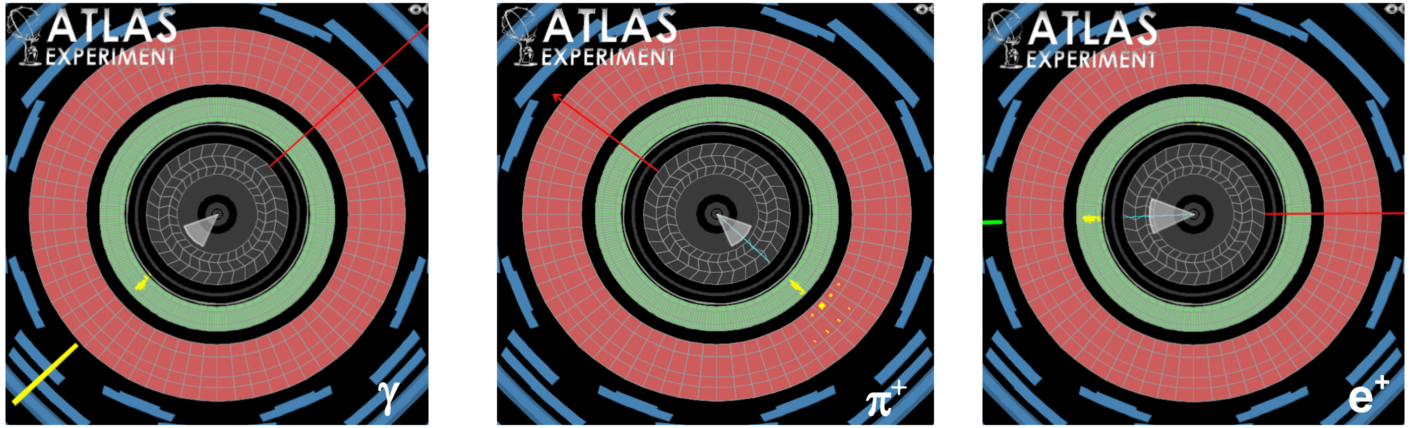
(see LPCC talk for more details)

Possible configurations:

- Full simulation of hard scatter (HS),
  fast generation+simulation+digi+reco for pile-up (PU)
- Combine all fast tools and only simulate particles if detected in an area of interest (analysis-dependent)

Anticipate speed-up of several orders of magnitude!

Summary



Fast Calorimeter Simulation:

- Crucial for ATLAS.
  Even more so in the future.

- Based on single particle parametrisations

- Using ML techniques (PCA, MLP)

- Prototype fully integrated in ATLAS infrastructure

- First validations show improvements over the previous version, though problems with EM shower shapes

- About 20 times faster than Geant4

- Pending: Performance on real physics processes (will it work for boosted objects?)

- Timescale: End of 2017

Fast Chain:

- Flexible set of tools to speed up the entire production chain, currently in development and validation ongoing

- Timescale: End of 2018

Backup

Performance of the Previous FastCaloSim

“old” FastCaloSim

Averaged simulation times per event in seconds:

Sample	Full G4 Sim	Fast G4 Sim	ATLFast-II	ATLFast-IIF
Minimum Bias	551	246	31.2	2.13
$t\bar{t}$	1990	757	101	7.41
Jets	2640	832	93.6	7.68
Photons, jets	2850	639	71.4	5.67
$W \rightarrow e\nu$	1150	447	57.0	4.09
$W \rightarrow \mu\nu$	1030	438	55.1	4.13

Precalculated showers
(“Frozen showers”)

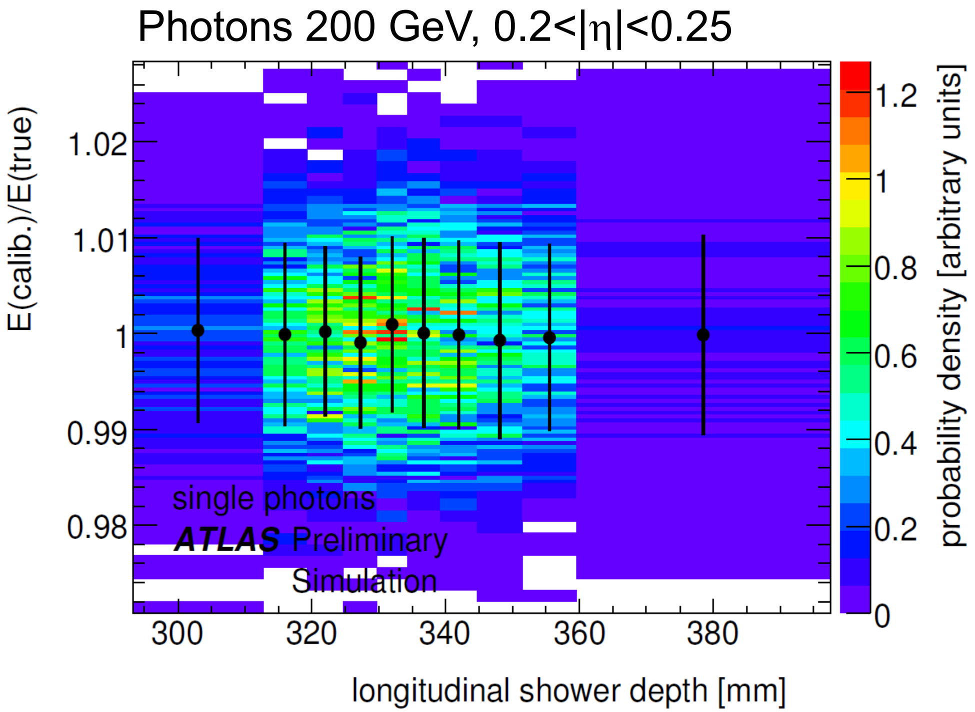
FastCaloSim+Fast Tracking Simulation

(from ATL-PHYS-PUB-2010-013)

The previous FastCaloSim

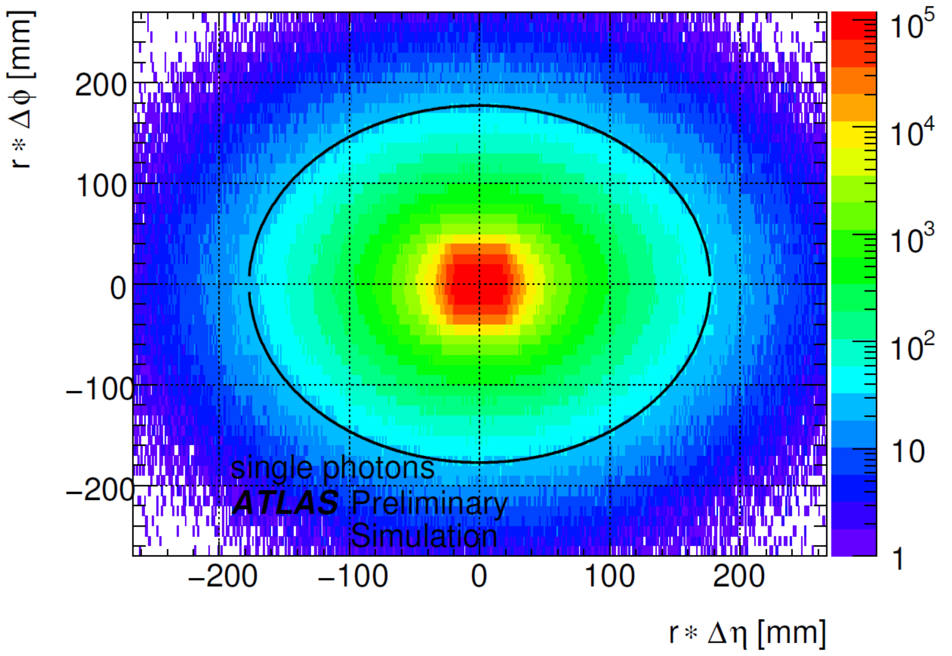
Longitudinal energy parametrisation:

- 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

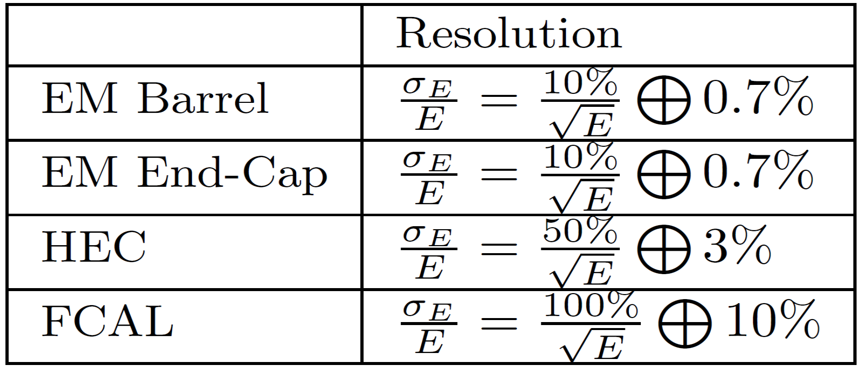
$$\frac{\Delta E}{E} = \frac{\alpha}{\sqrt{E}} \oplus \beta \oplus \frac{\gamma}{E}$$

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	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \bigoplus 10\%$