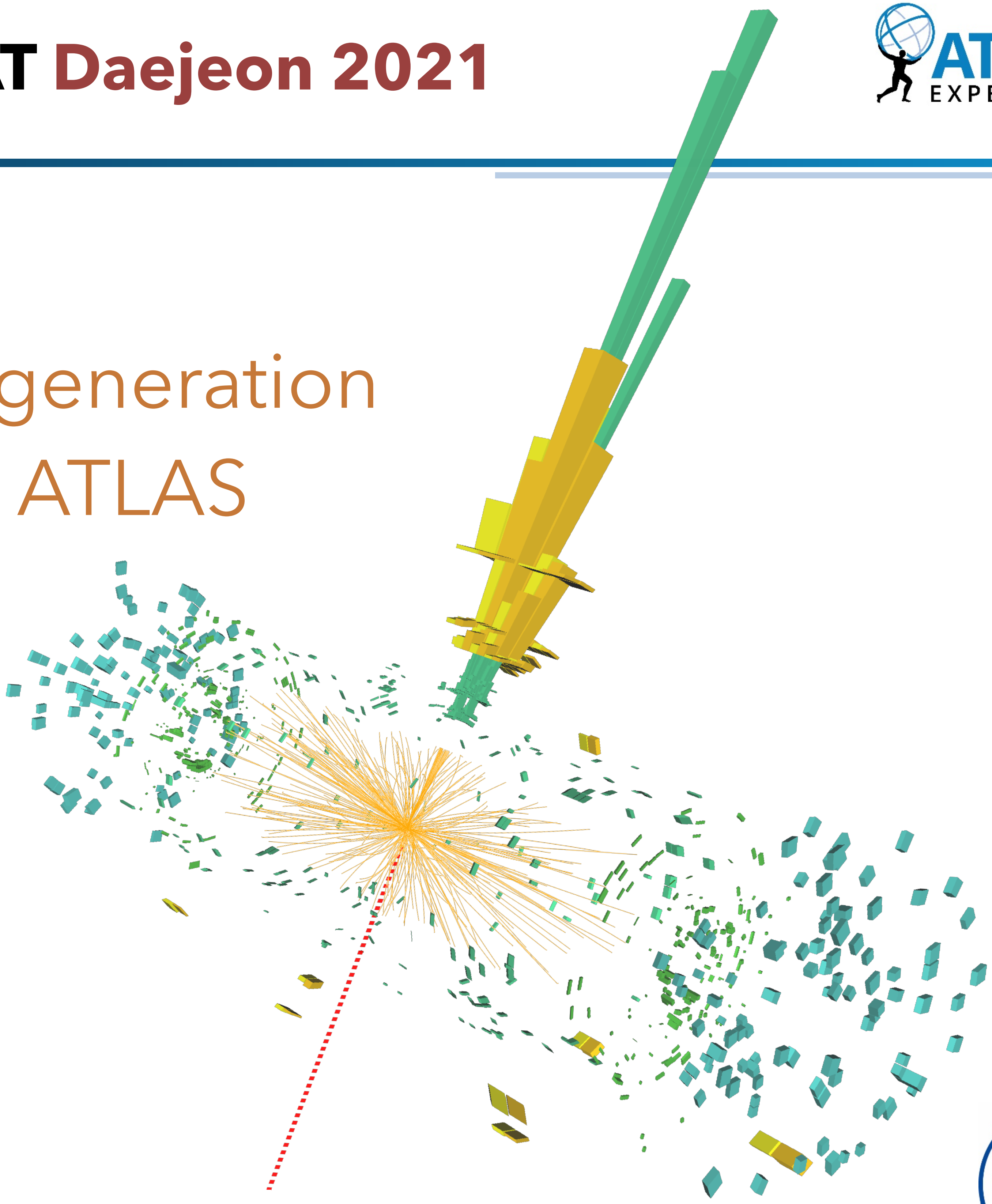
GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

ACAT Daejeon 2021

AtlFast3: The next generation of fast simulation in ATLAS

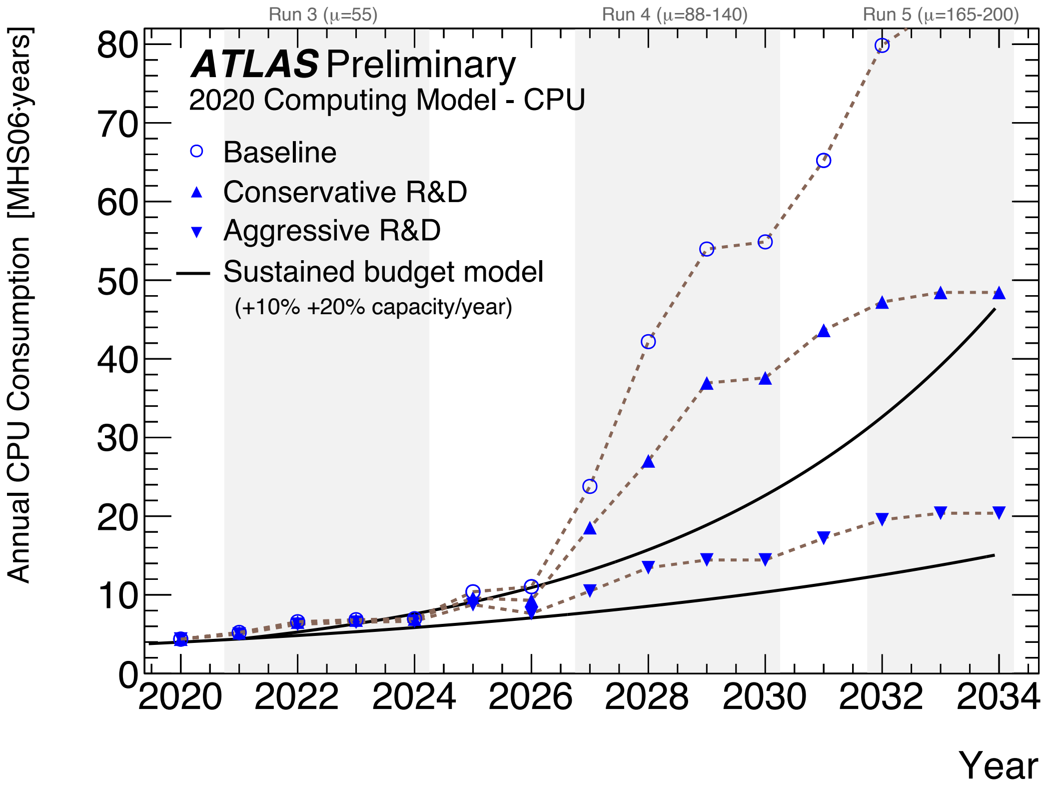


Joshua F. Beirer [1, 2]
On behalf of the ATLAS Collaboration

[1] CERN
[2] Georg-August-Universität Göttingen

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

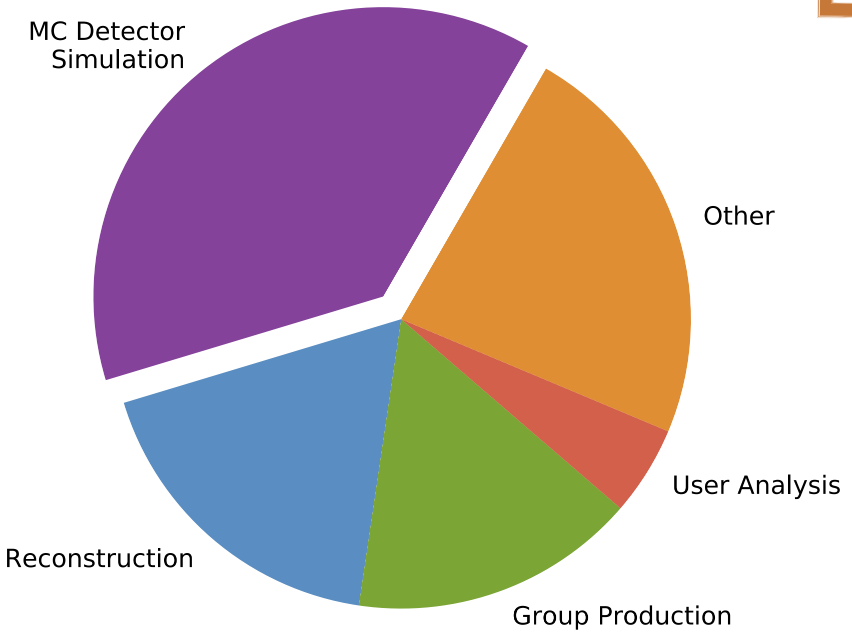
Fast Simulation in ATLAS: Motivation



- ATLAS expects significant increase in annual CPU consumption in the coming years
- Aggressive R&D required in order to keep up in Run3 and HL-LHC

In Run 2: 32 out of 52 billion (~60%) events simulated with AtlFastII

- Around 40% of our computing power is used for MC detector simulation
- ~80-90% of detector simulation spent on calorimeter simulation
- Fast and accurate shower simulation crucial

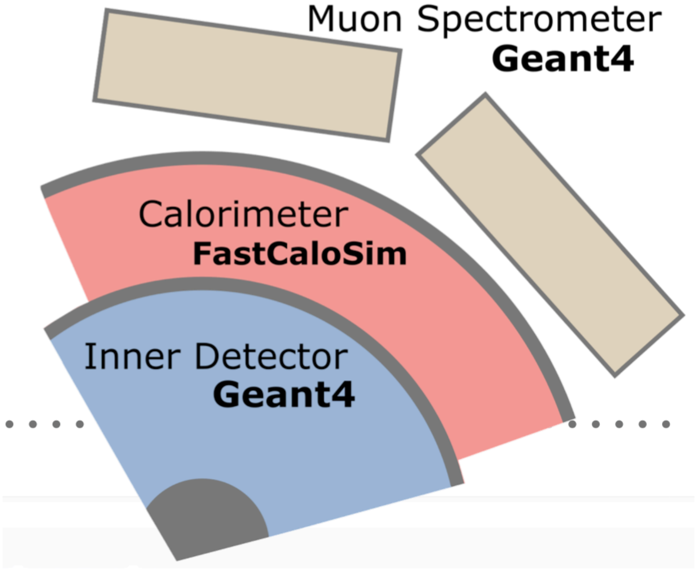


Fast Simulation Usage Targets:

- Run 3: >75 % events (analysis required to justify use of full simulation)
- Run 4: >90 % events

Overview

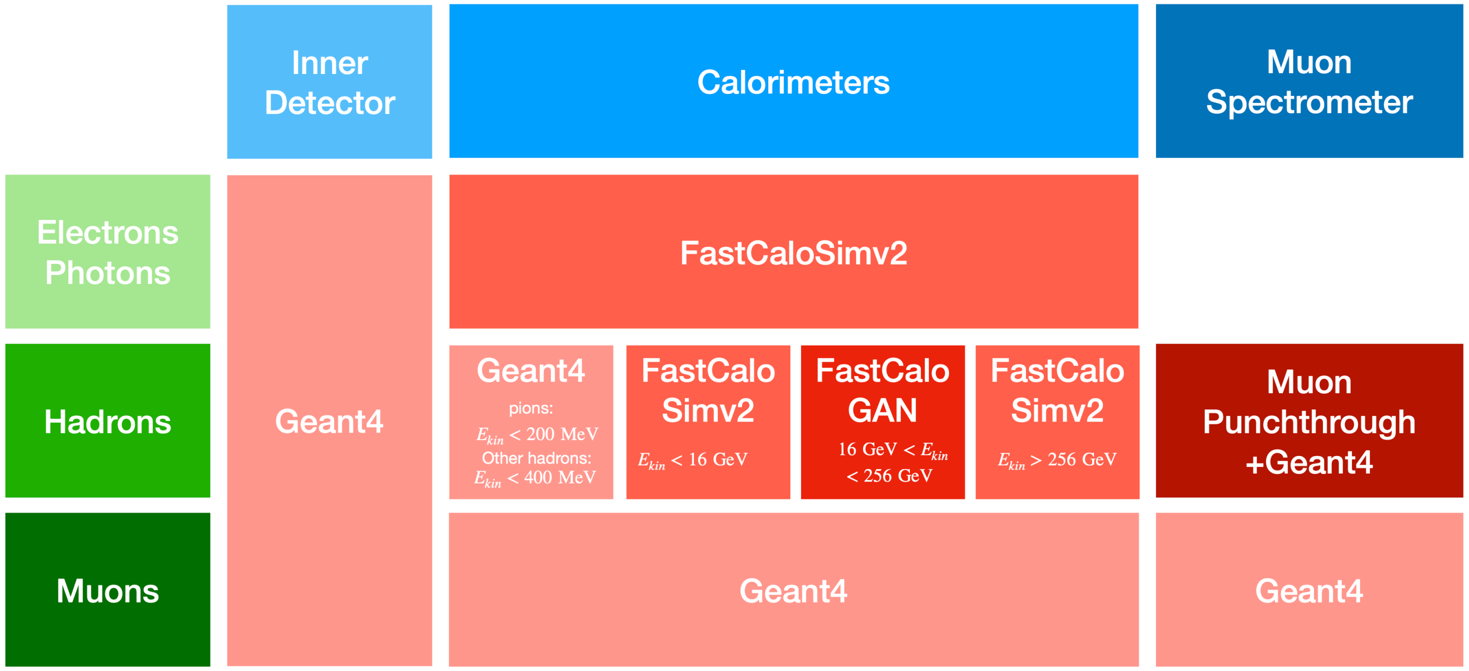
- Atlfast-III (AF3) is the successor of the Atlfast-II (AF2) simulator
- Full simulation of the ID and parametrised simulation of the calorimeter
- AF3 implements two distinct approaches of shower generation:
  - **FastCaloSimV2:** parametrised modelling
  - **FastCaloGAN:** Generative Adversarial Network
- Dedicated parametrisation for punch through particles



AF3 already used for reprocessing Run-2 events!

arXiv:2109.02551

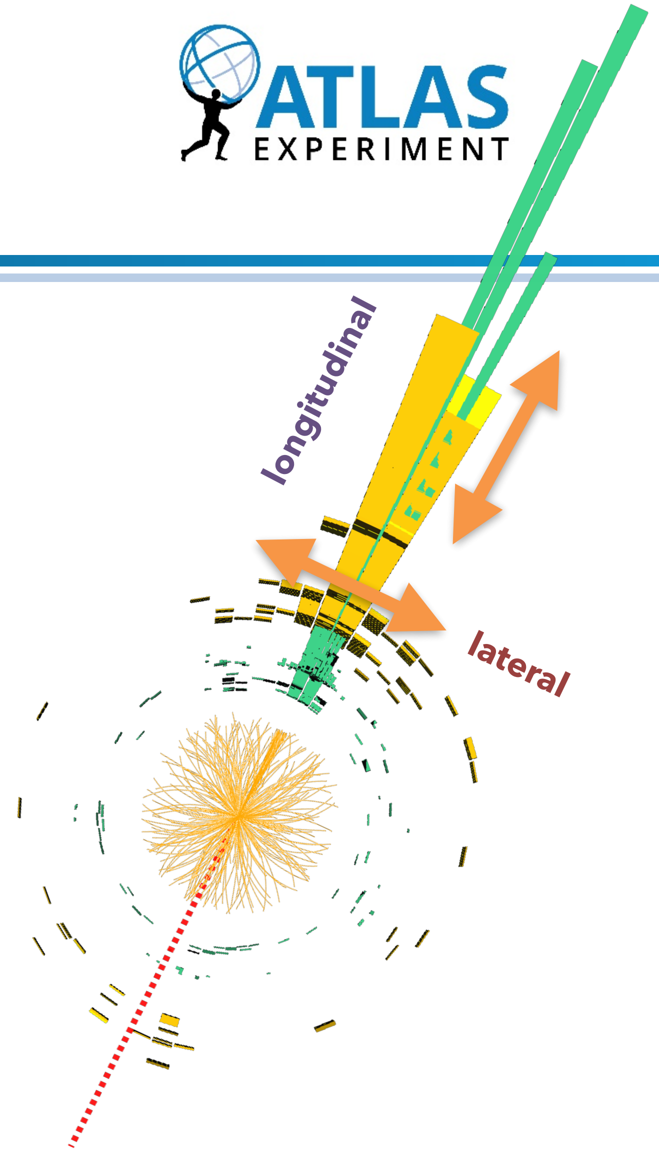
SCAN ME!



	Inner Detector	Calorimeters				Muon Spectrometer
Electrons Photons	Geant4	FastCaloSimv2				
Hadrons		Geant4 pions: $E_{kin} < 200$ MeV Other hadrons: $E_{kin} < 400$ MeV	FastCalo Simv2 $E_{kin} < 16$ GeV	FastCalo GAN 16 GeV $< E_{kin} <$ 256 GeV	FastCalo Simv2 $E_{kin} > 256$ GeV	Muon Punchthrough +Geant4
Muons		Geant4				Geant4

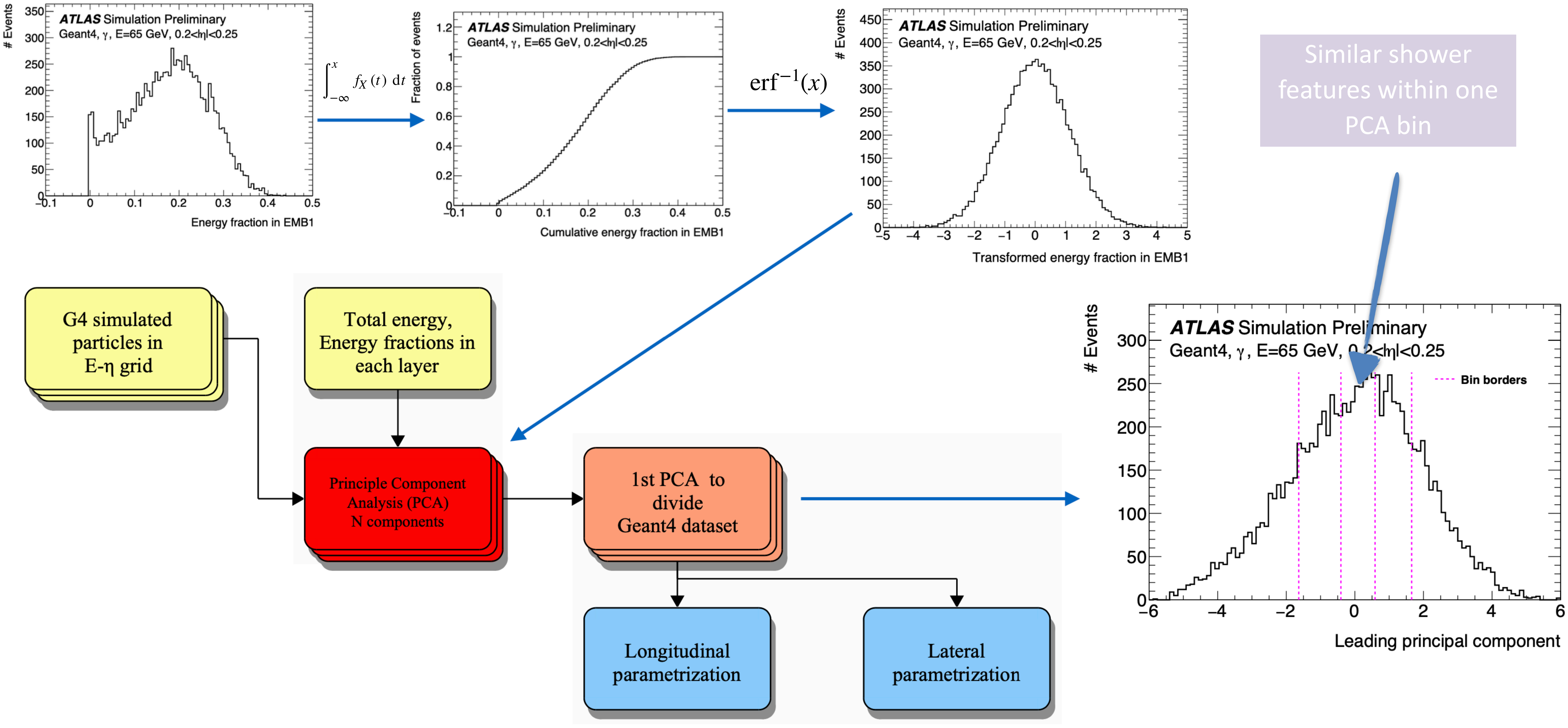
Introduction to FCSV2

- Use electrons and photons for electromagnetic showers and pions for hadronic showers
- Separate parametrisation of **longitudinal** and **lateral** shower development
- Geant4 spatial energy deposits ('hits') grouped into volumes ('voxels')
- 17 bins of energy (64MeV - 4TeV) and 100 bins of |η| (0 - 5.0)
- Energy deposits in layers are **highly correlated**, difficult to model
- Classify showers based on depth on the interaction point
- **Principal Component Analysis** (PCA) to de-correlate layers



Hadronic showers: wider, slower, and larger fluctuations

Energy Decorrelation using PCA



Similar shower features within one PCA bin

G4 simulated particles in E-η grid

Total energy, Energy fractions in each layer

Principle Component Analysis (PCA) N components

1st PCA to divide Geant4 dataset

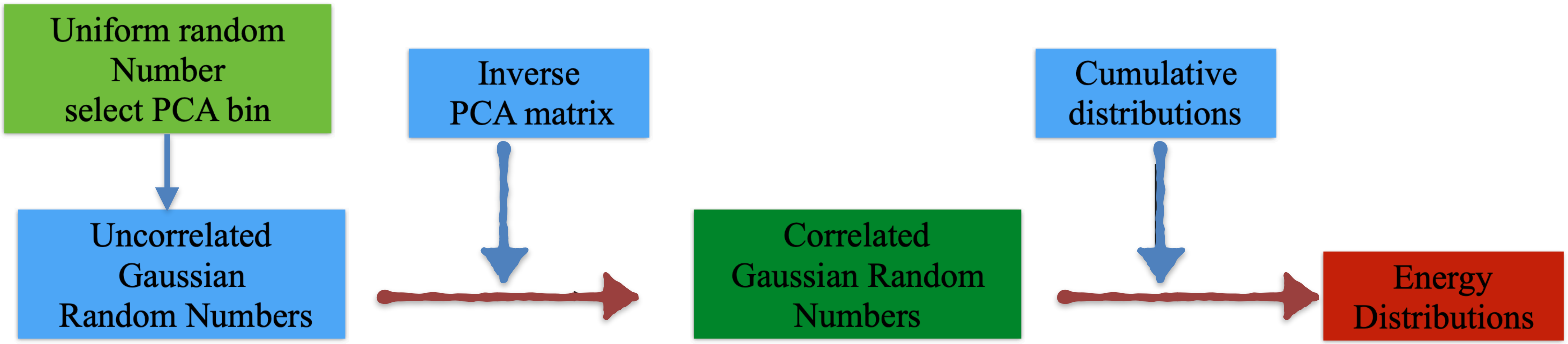
Longitudinal parametrization

Lateral parametrization

- Additional PCA on each bin of 1st PCA to achieve further decorrelation
- Store cumulative energy fractions, mean and RMS of gaussians and PCA matrix

Energy Parametrisation and Interpolation

During simulation, chain is performed backwards:



Uniform random Number select PCA bin

Uncorrelated Gaussian Random Numbers

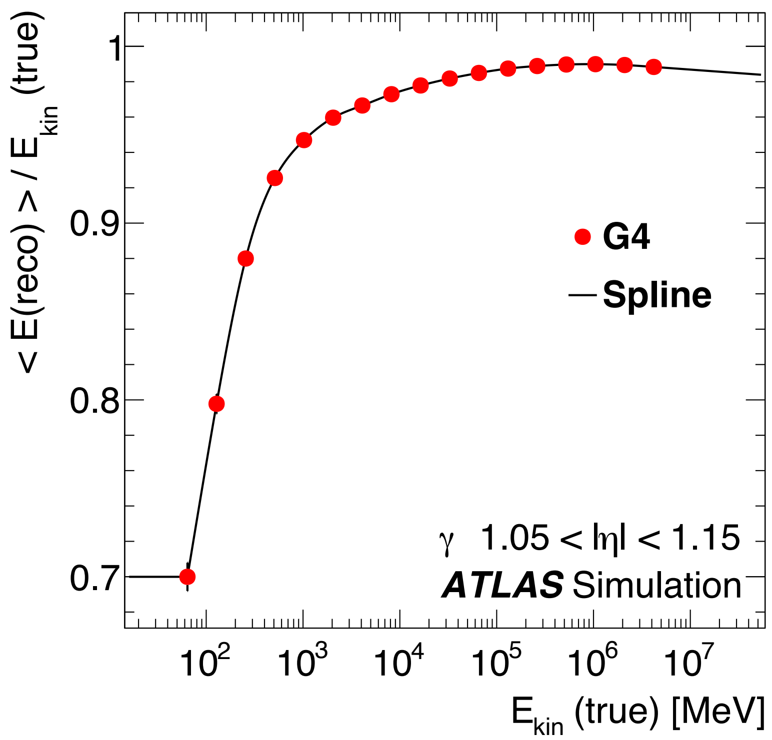
Inverse PCA matrix

Correlated Gaussian Random Numbers

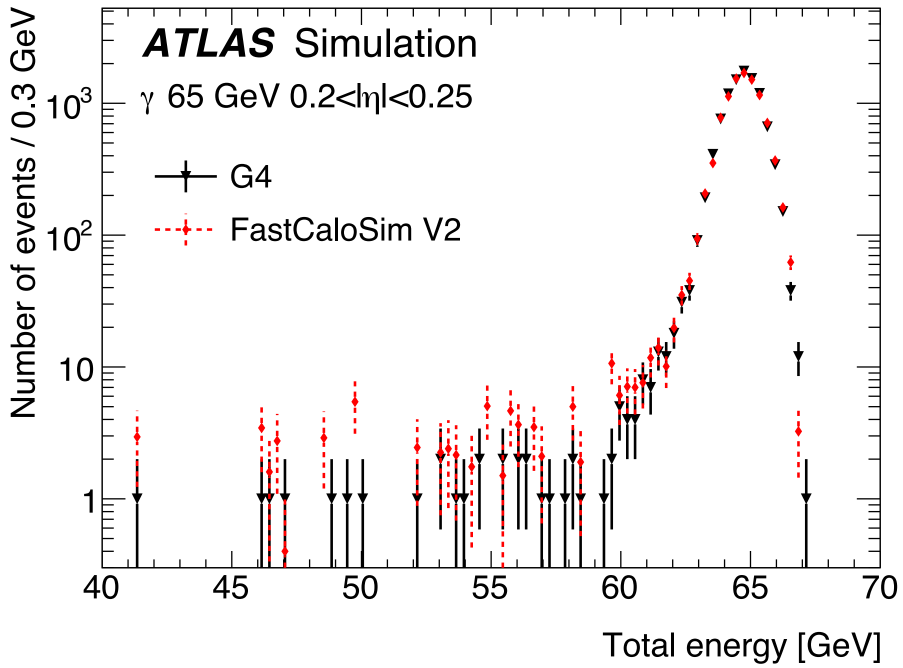
Cumulative distributions

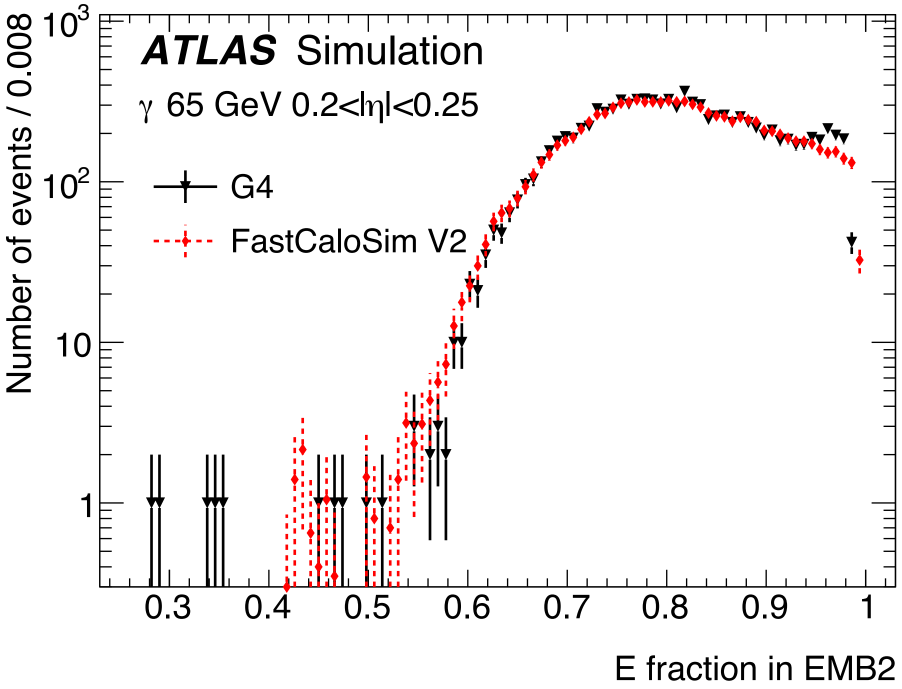
Energy Distributions

Interpolate energy response using splines:

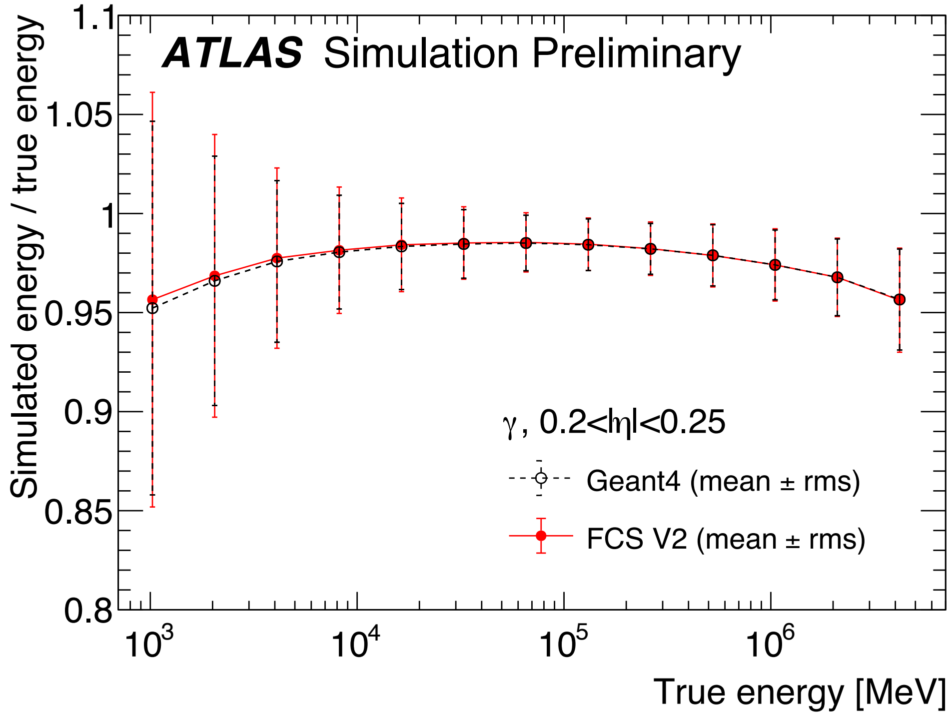


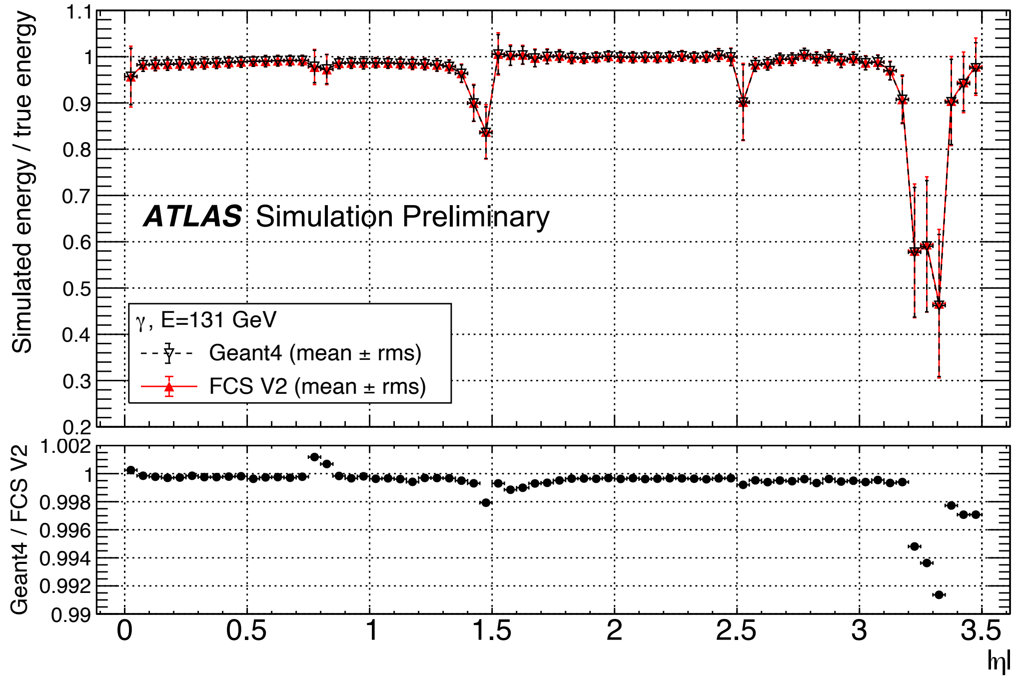
Energy Parametrisation: Toy Validation





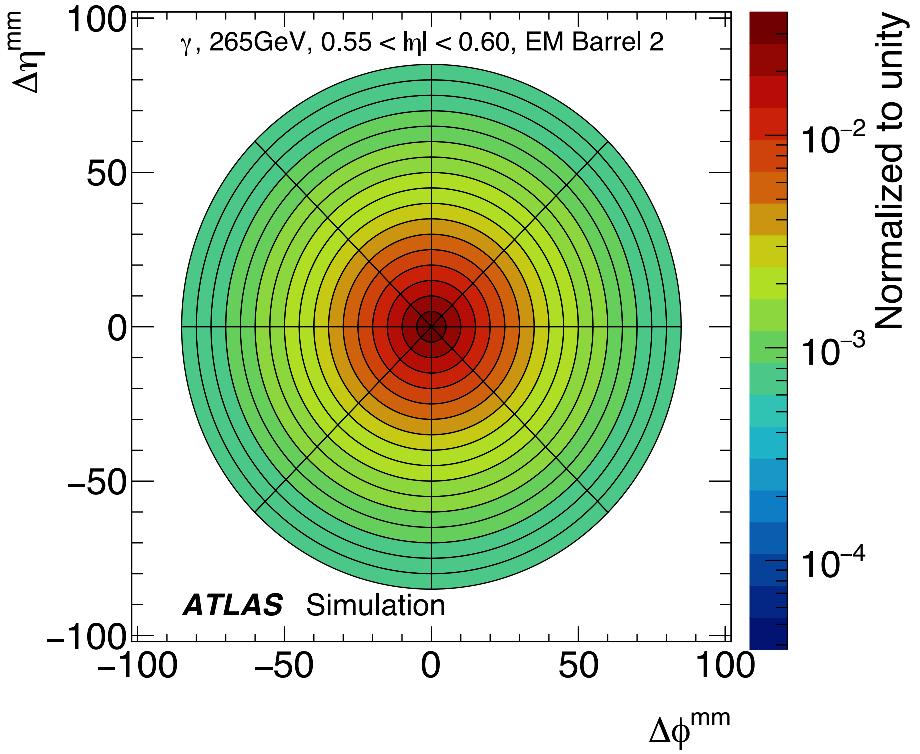
Excellent agreement in toy validation



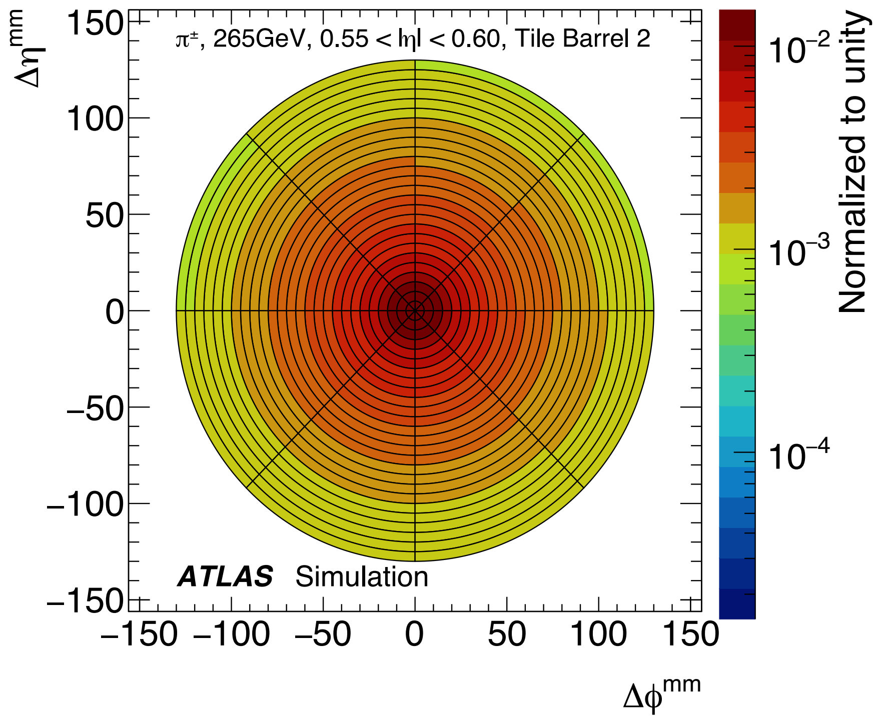


Lateral Shape Parametrisation

Photons



Pions

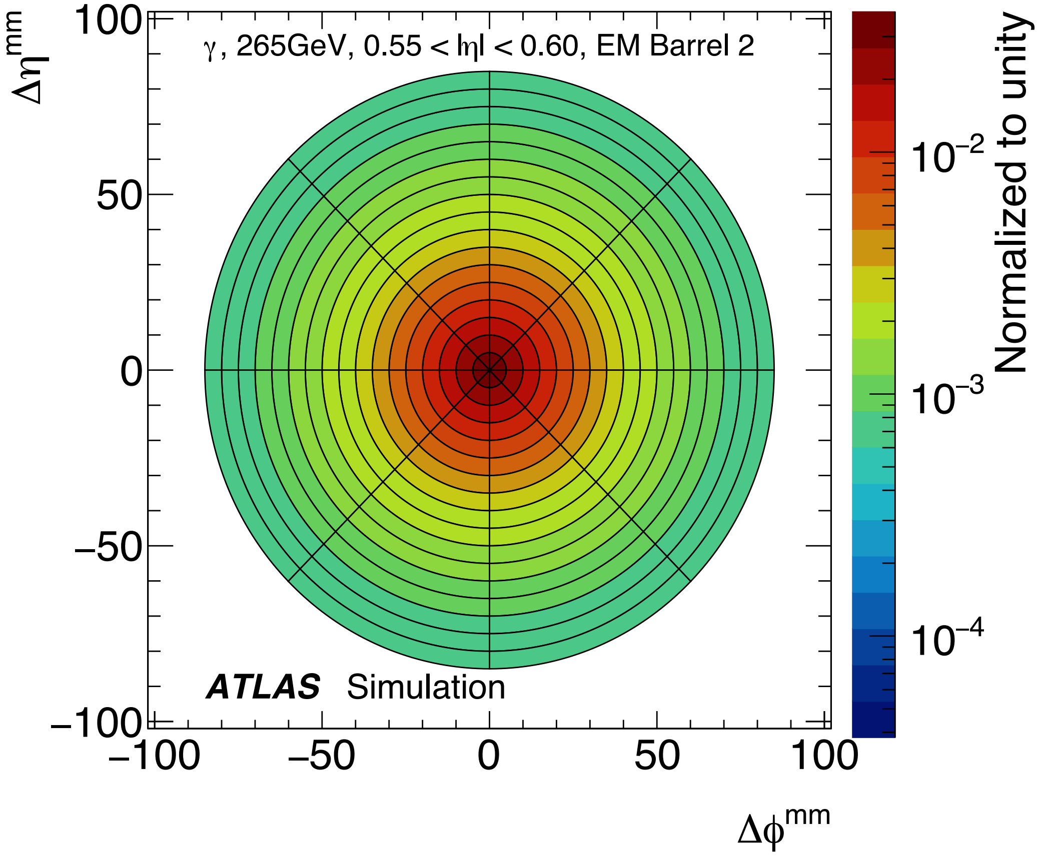


- Average energy distribution in lateral direction parametrised over radial distance containing 99.5% of total energy and 8-bins in angular direction
- Parametrization for each particle, energy, eta, calorimeter layers and bins of 1st PCA
- During simulation, randomly generate quantised energy deposits (hits) from 2D shape histograms (PDFs)
- To each hit, assign hit energy

$$E_{\text{hit}} = \frac{E_{\text{layer}}}{N_{\text{hits}}^{\text{layer}}} \times w$$

Weight dependent on radial position of hit

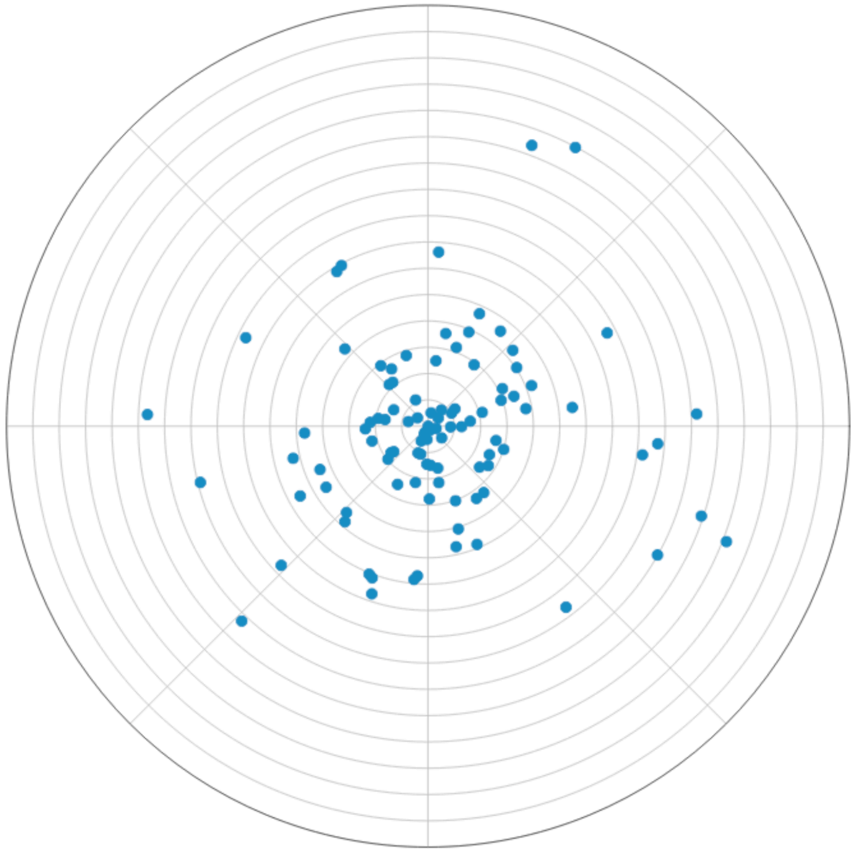
Random Fluctuations



Sampling uncertainty (e.g. EMB2)

$$\sigma_E \approx \frac{10\,\%}{\sqrt{E}}$$

$$N_{\text{hits}}^{\text{layer}} \sim Poisson\left(\frac{1}{\sigma_E^2}\right)$$

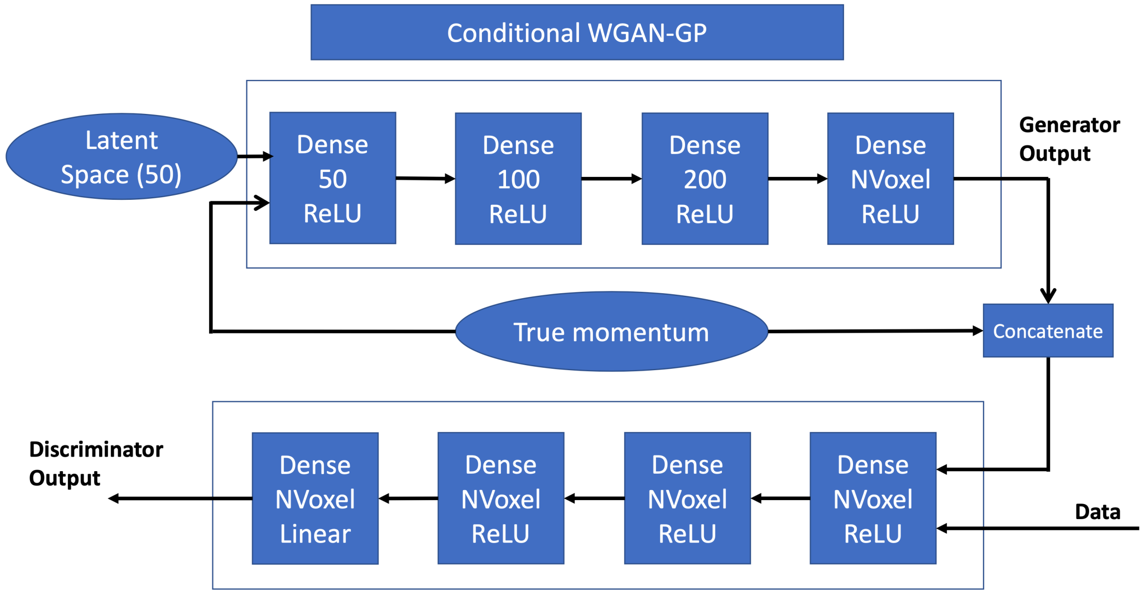


- Individual showers different than average showers
- Modelling of fluctuations extremely important, especially for pions

Statistical fluctuation of drawing N hits tuned to match expected calorimeter sampling uncertainty

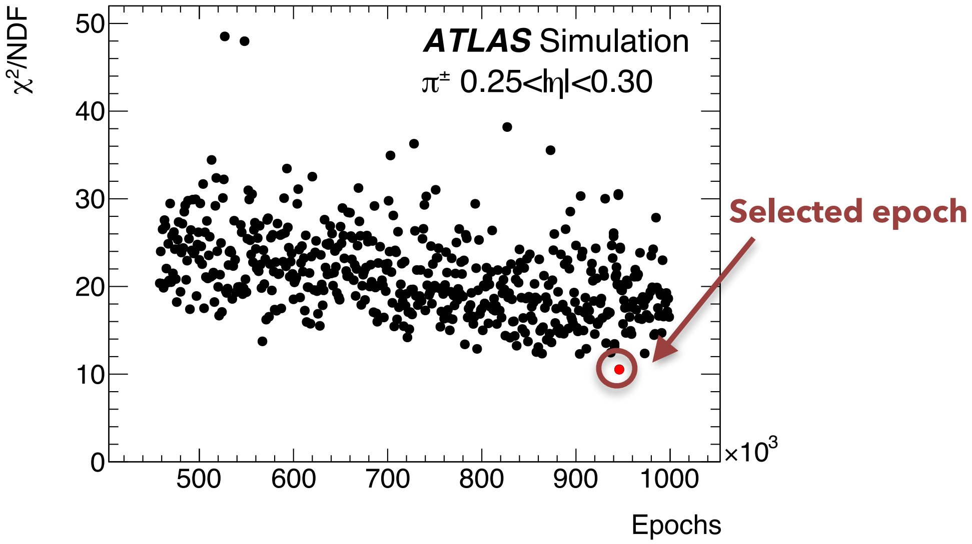
FastCaloGAN in a Nutshell

- FastCaloGAN based on **WGAN-GP** algorithm which offers more stable training compared to conventional GANs
- Electrons, photons and pions used to train the network
- **One GAN is trained for each of the 100 bins** in |η| (0 - 5.0) and conditioned on **truth momentum**
- Total of 300 GANs to cover full detector region
- FastCaloGAN uses same input hits as FCS and bins them in voxels similar to FCSV2 shape histograms
- GAN **trained to reproduce voxels and energies in the layer as well as total energy** in one single step
- Each GAN trained for 1M epochs with a checkpoint saved every 1K epochs
- GAN with **best epoch** is used to generate hits which are deposited in corresponding voxels



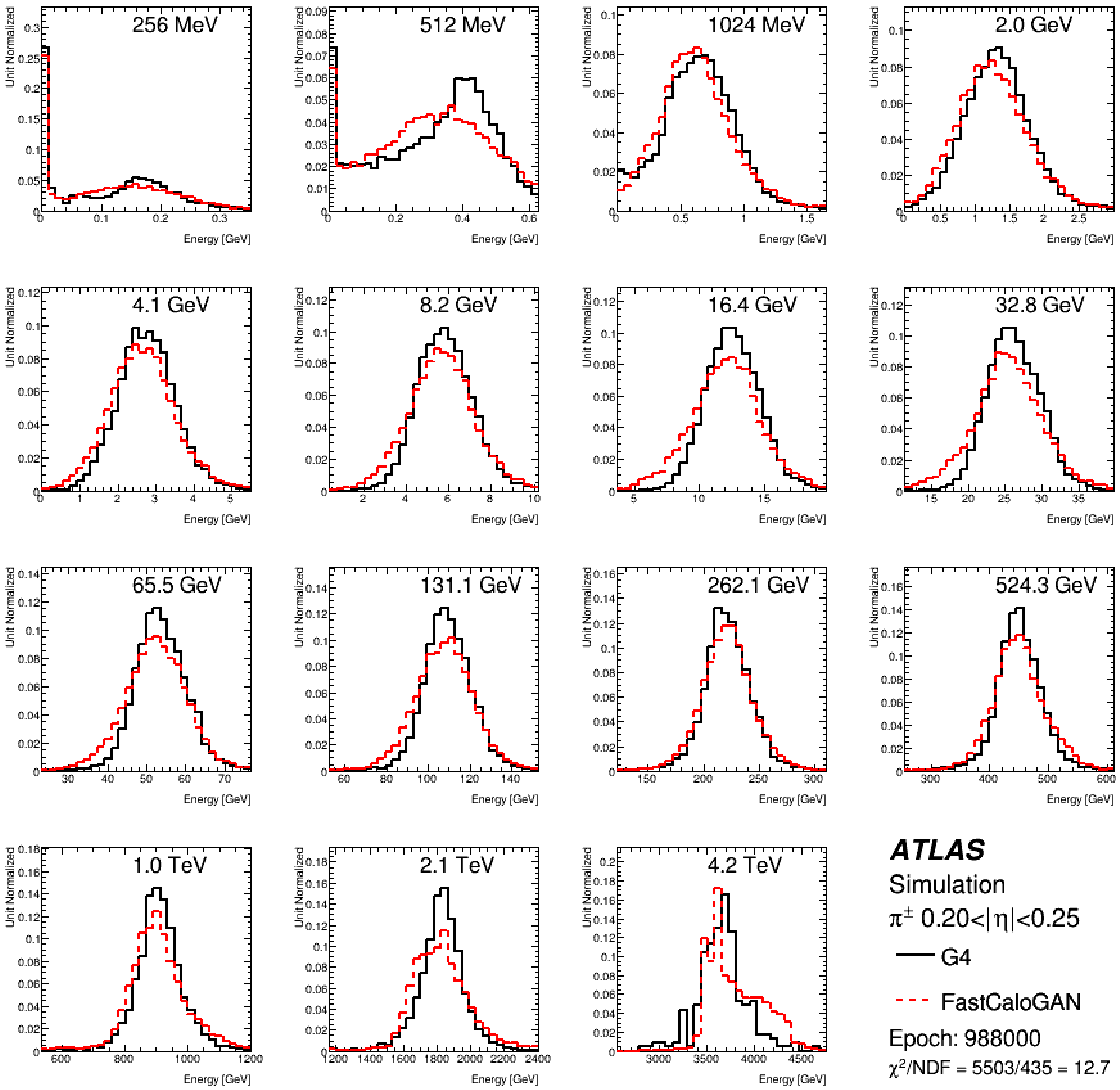
NVoxel	Number of voxels
Generator nodes	50, 50, 100, 200, NVoxel
Discriminator nodes	NVoxel, NVoxel, NVoxel, NVoxel, 1
Activation function	ReLU
Optimizer	Adam [50]
Learning rate	10^{-4}
$\beta 1$	0.5
$\beta 2$	0.999
Batch size	128
Training ratio (D/G)	5
Gradient penalty (λ)	10

WGAN-GP parameters



Selected epoch

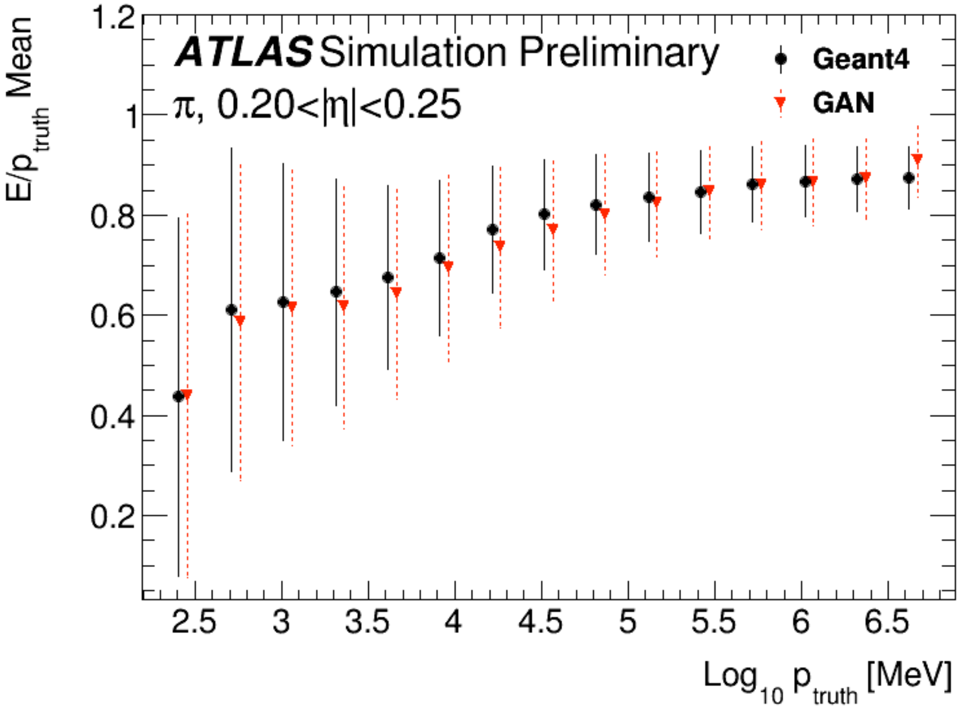
FastCaloGAN Performance

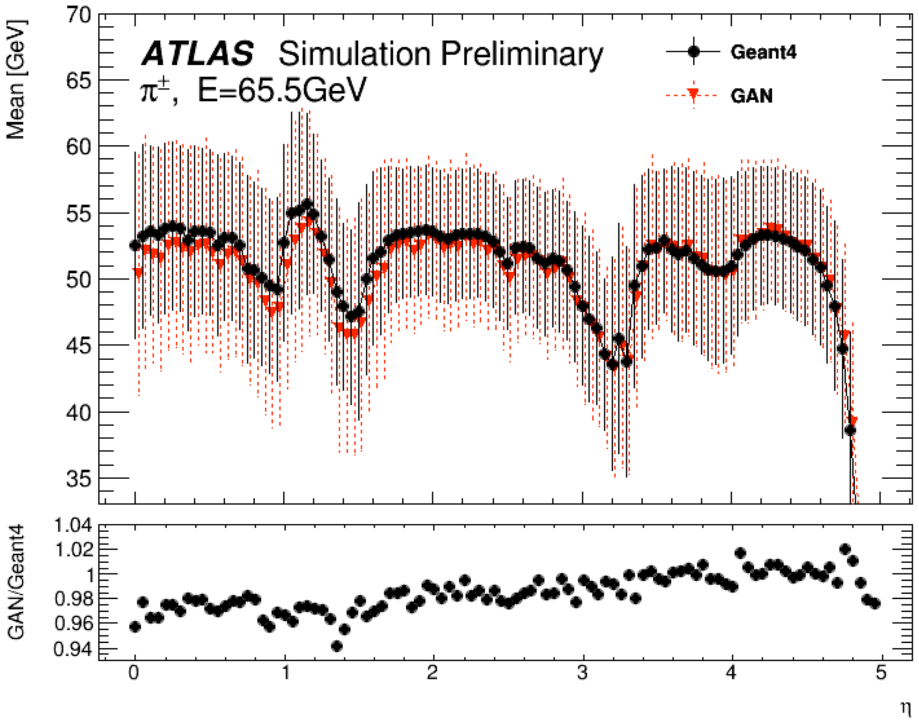


Good performance for total energy simulation across large energy range

See here for an animation showing the learning process of the GAN throughout the training

FastCaloGAN Performance

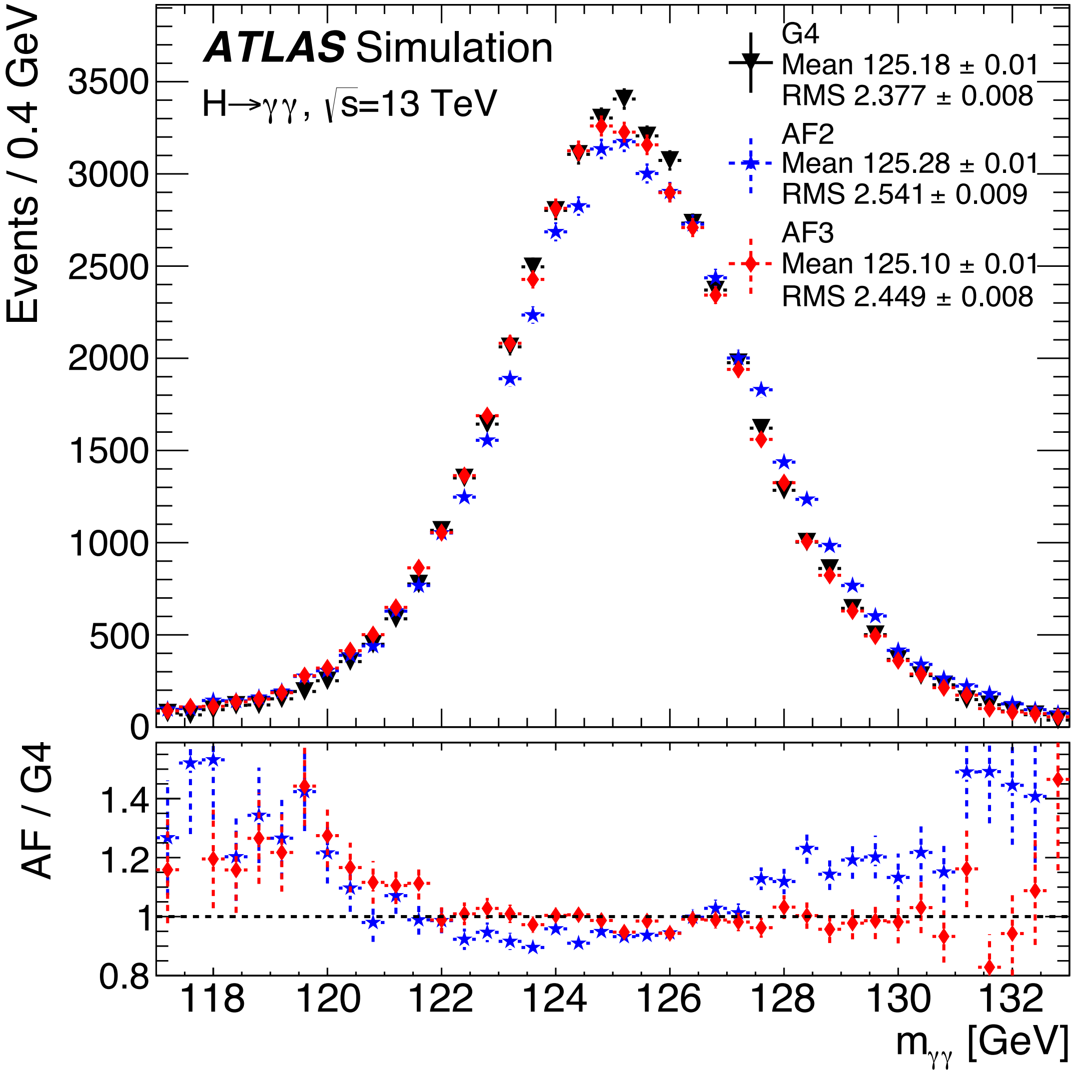


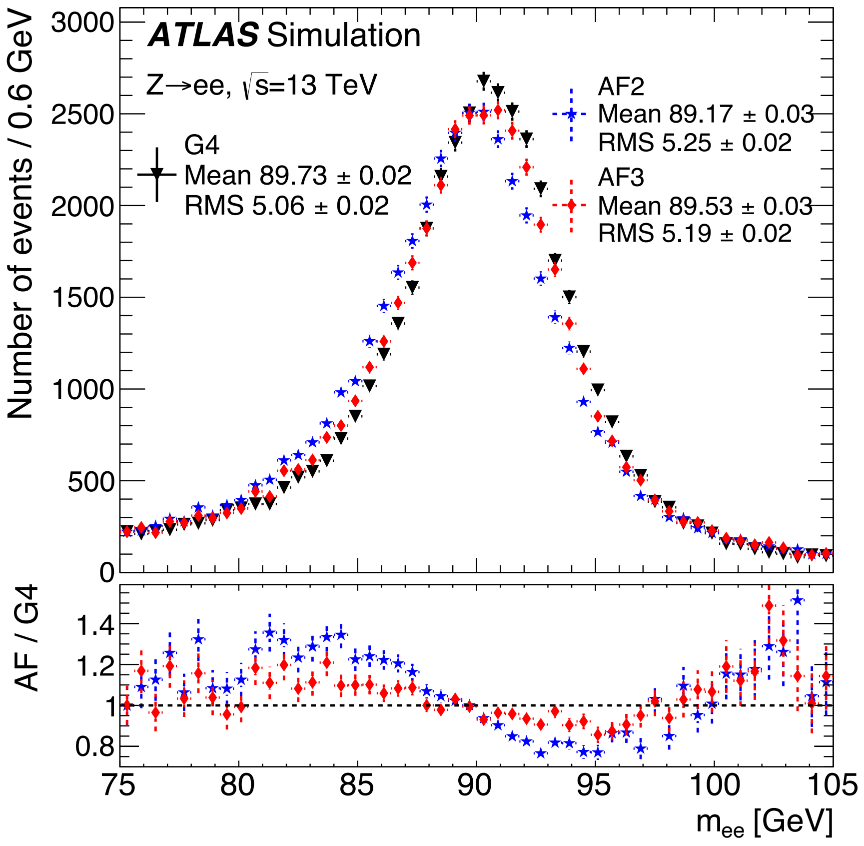


- FastCaloGAN shows better modelling compared to FCSV2 for hadrons in **medium energy range**
- AF3 uses FastCaloGAN for hadron showers in the range of $16\,\mathrm{GeV} \leq E_{\mathrm{kin}} \leq 256\,\mathrm{GeV}$
- Total energy of FastCaloGAN scaled to fit total energy of FCSV2 to allow smooth transition between simulation flavours

Full implementation in ATLAS simulation infrastructure

Performance: e/γ

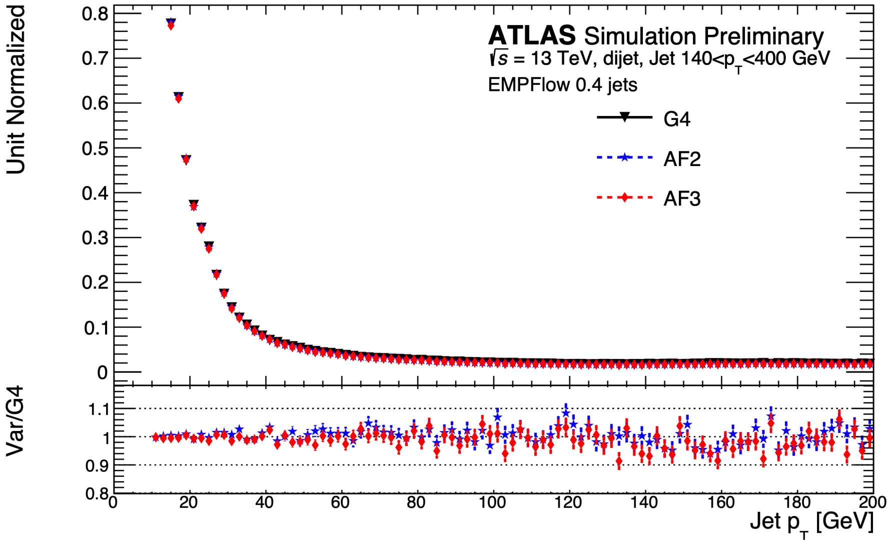




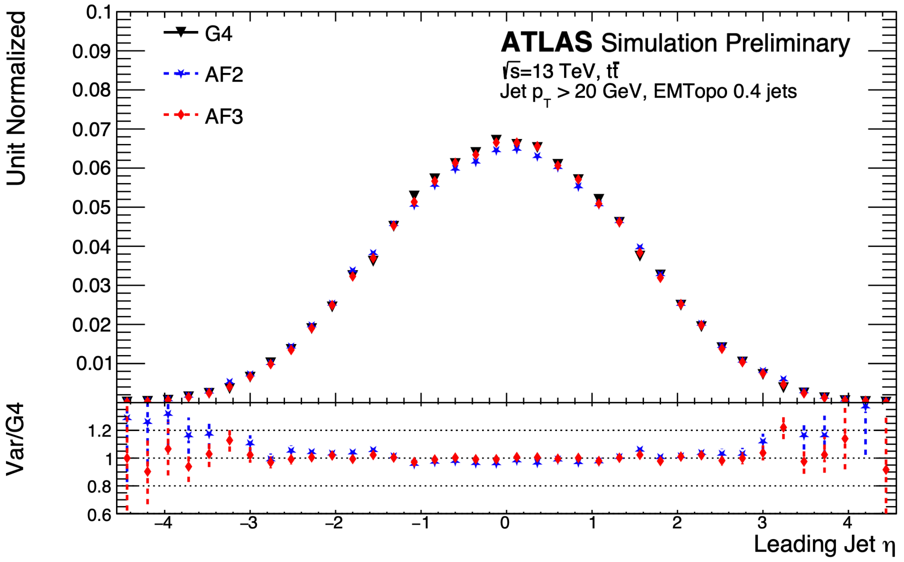
- Electron/photon candidates reconstructed from topological clusters of energy deposits in EM calorimeter
- Reconstructed $H \rightarrow \gamma\gamma$ and $Z \rightarrow ee$ mass shows large improvement compared to AF2

Performance: Jet kinematics

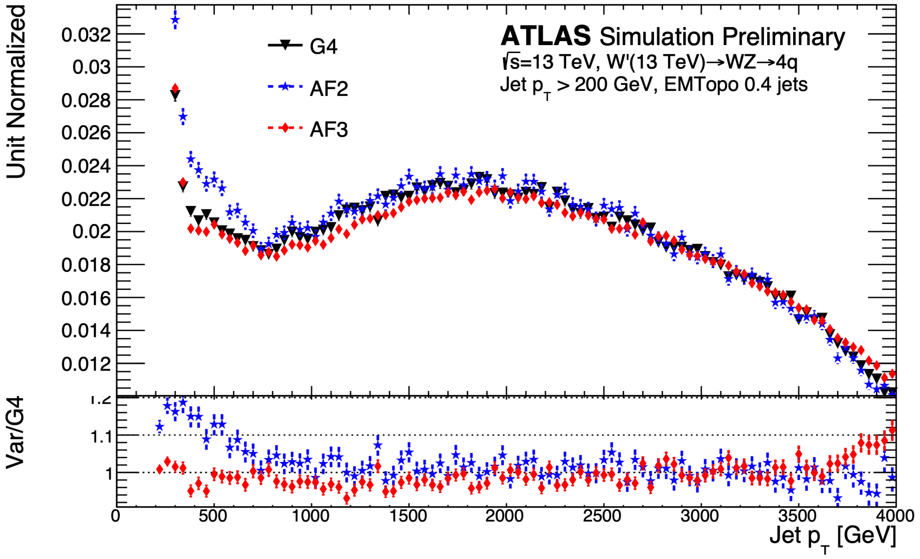
Jet p_T in di-jet jets reconstructed with EMPFlow 0.4 jets



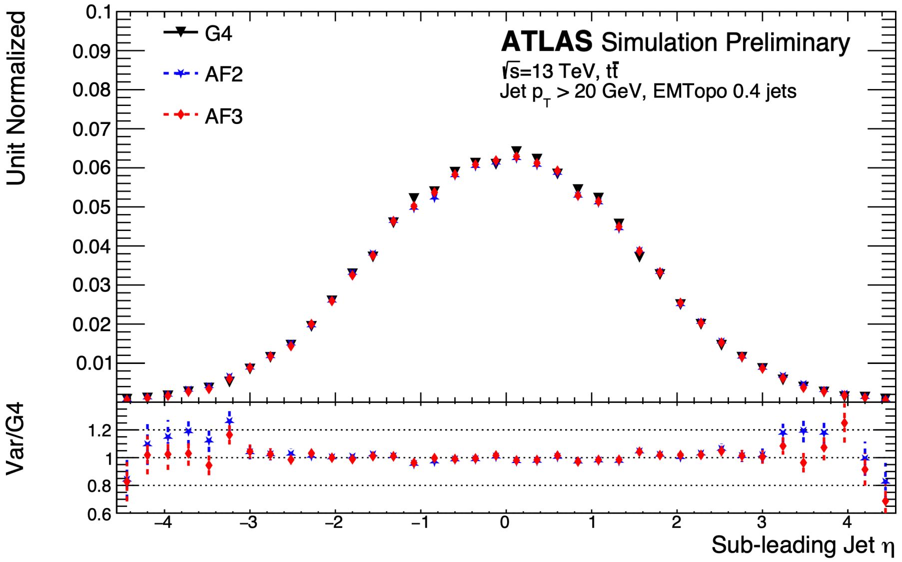
Leading jet η in $t\bar{t}$ sample reconstructed with EMTopo 0.4 jets



Jet p_T in $W'(13\ \text{TeV}) \rightarrow WZ \rightarrow 4q$ sample, reconstructed with EMPFlow 0.4 jets



Sub-leading jet η in $t\bar{t}$ sample reconstructed with EMTopo 0.4 jets



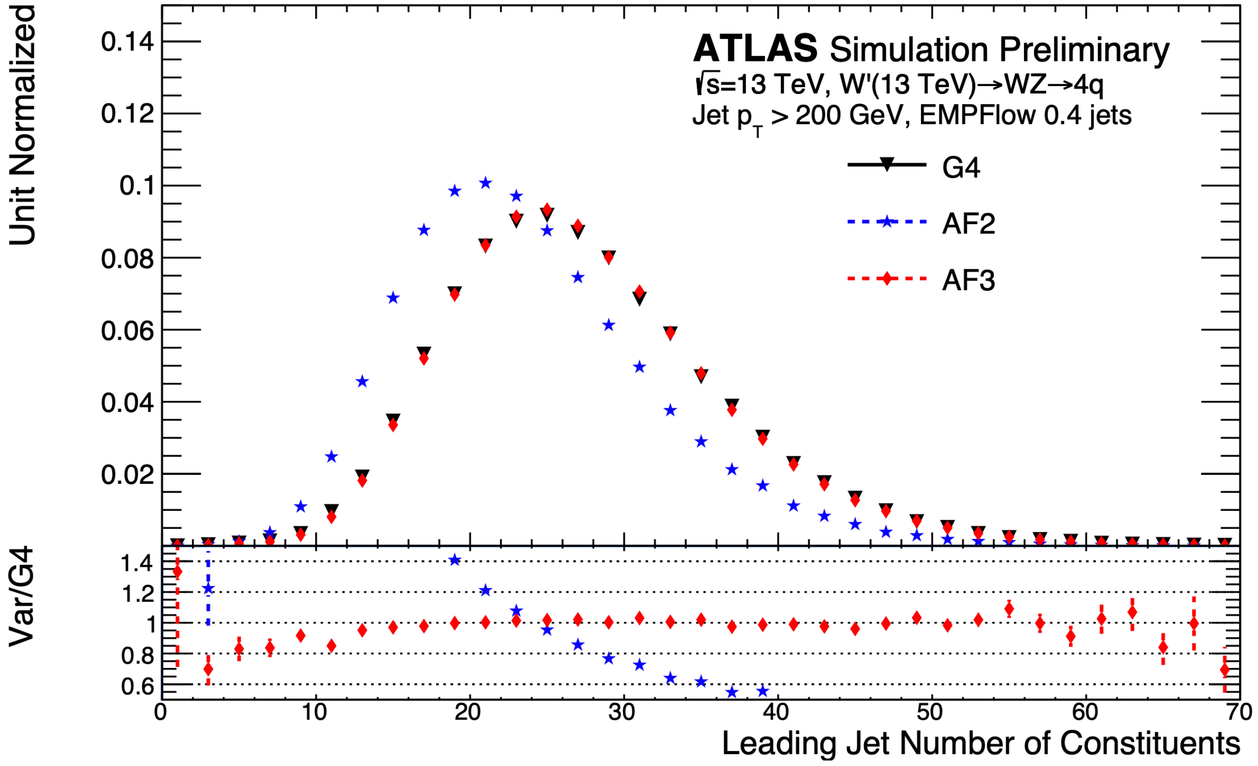
- Good modelling of jets kinematics in a broad range of physics applications
- Improved modelling for high-p_T jets compared to AF2
- Improved modelling for $|\eta| > 3.0$ as a result of dedicated parametrisation in forward calorimeter

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

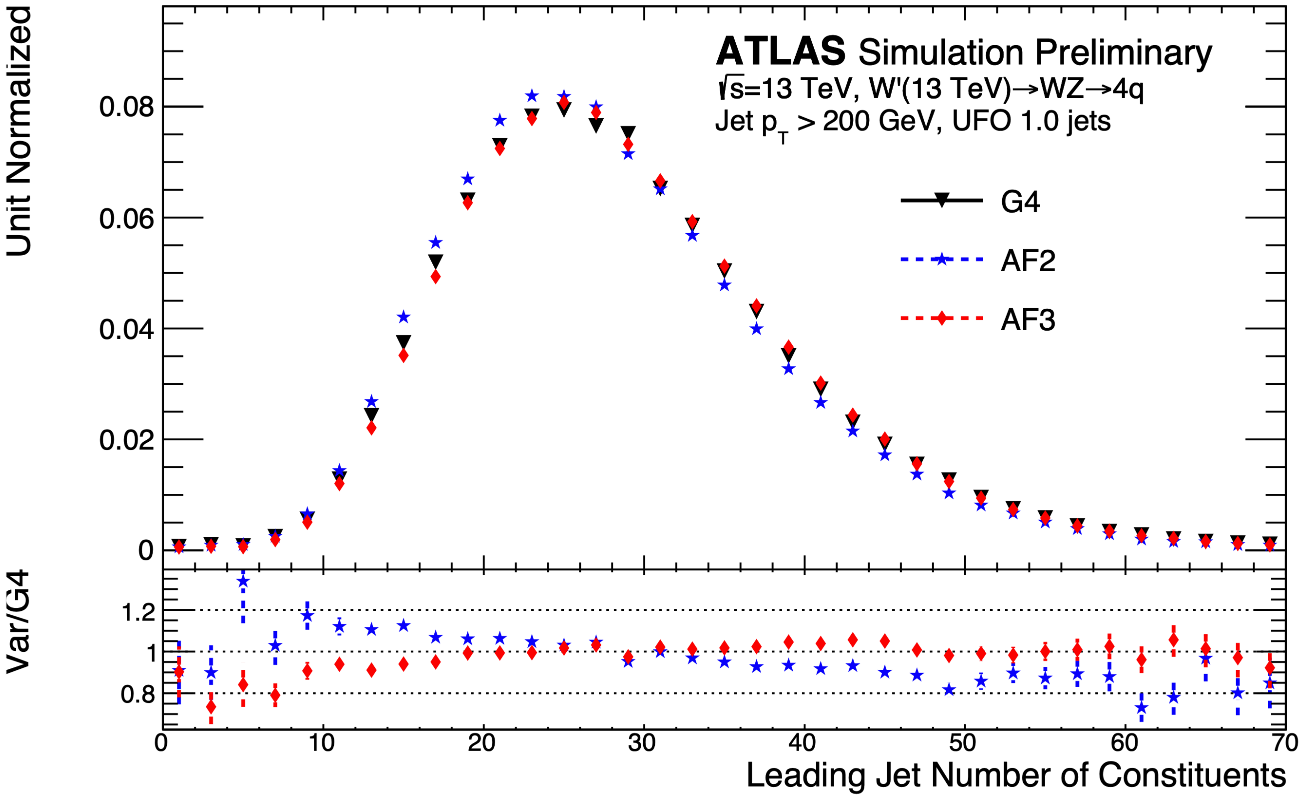
Performance: Jet constituents

Number of constituents of p_T-leading jet in $W'(13\,\text{TeV}) \rightarrow WZ \rightarrow 4q$ sample

EMPFlow 0.4 jets



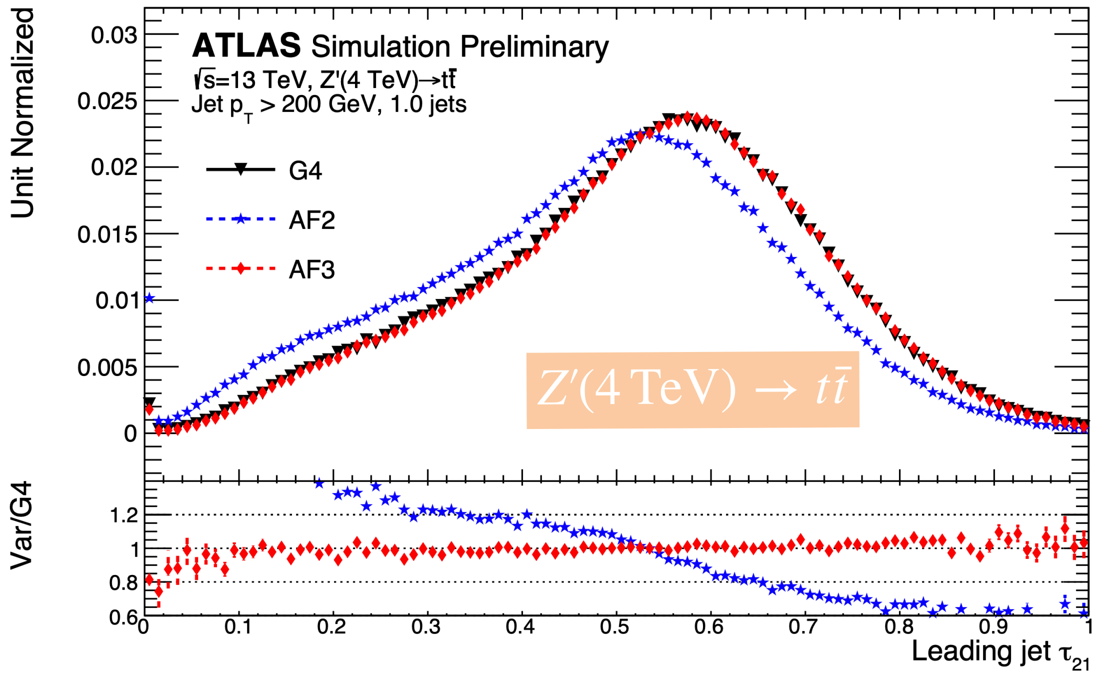
UFO 1.0 jets



Significant improvements in the modelling of number of constituents

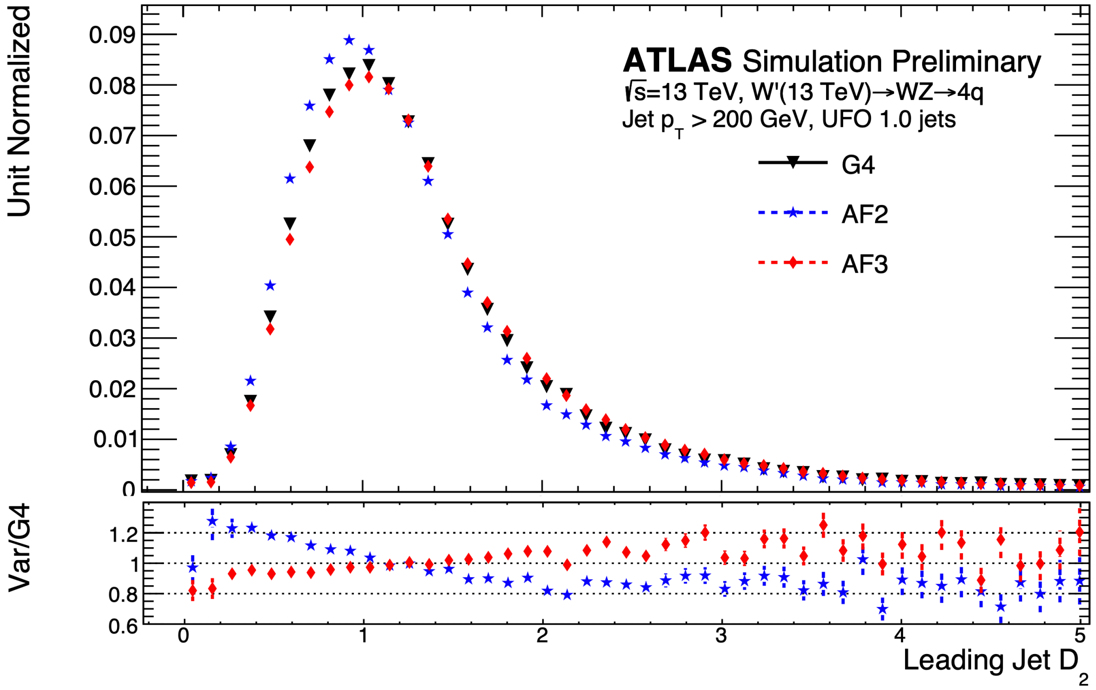
Performance: Jet substructure

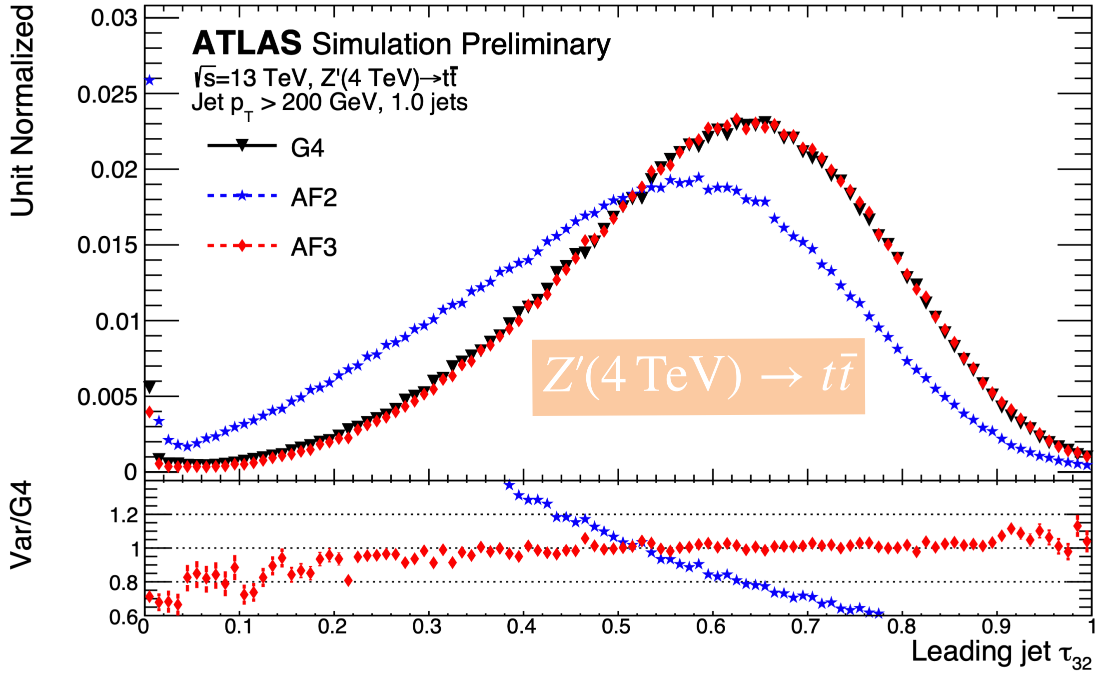
Substructure variables with UFO 1.0 jets



$Z'(4\,\text{TeV}) \rightarrow t\bar{t}$

$W'(13\,\text{TeV}) \rightarrow WZ \rightarrow 4q$



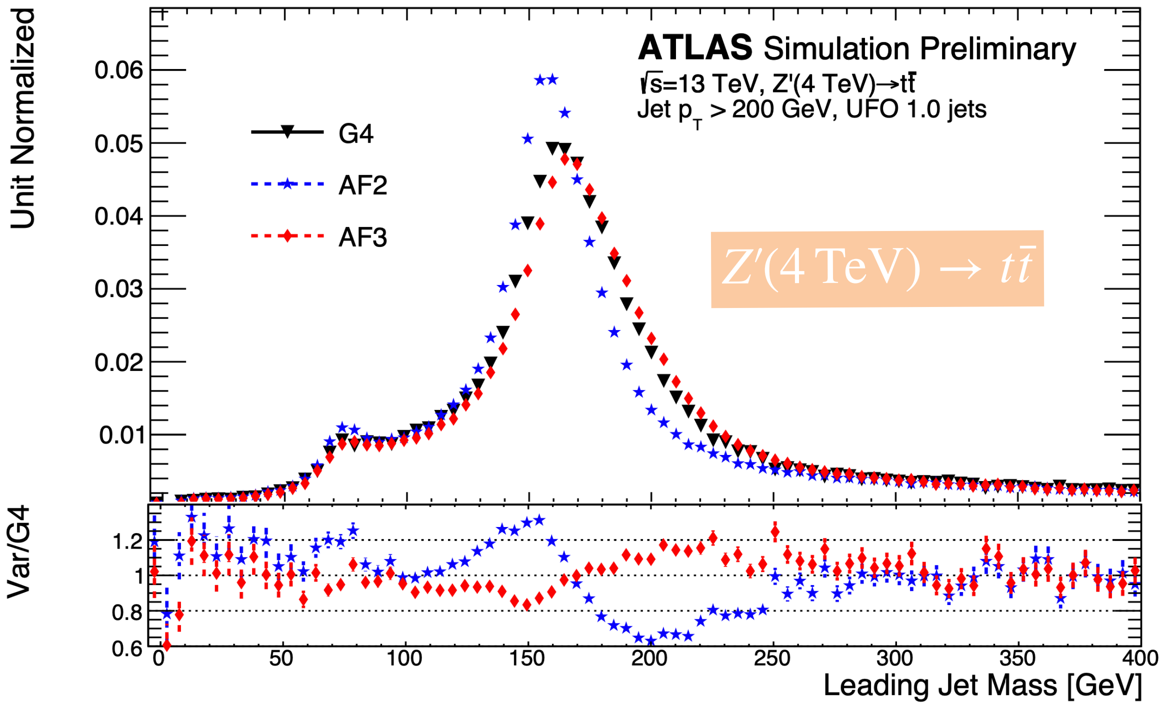


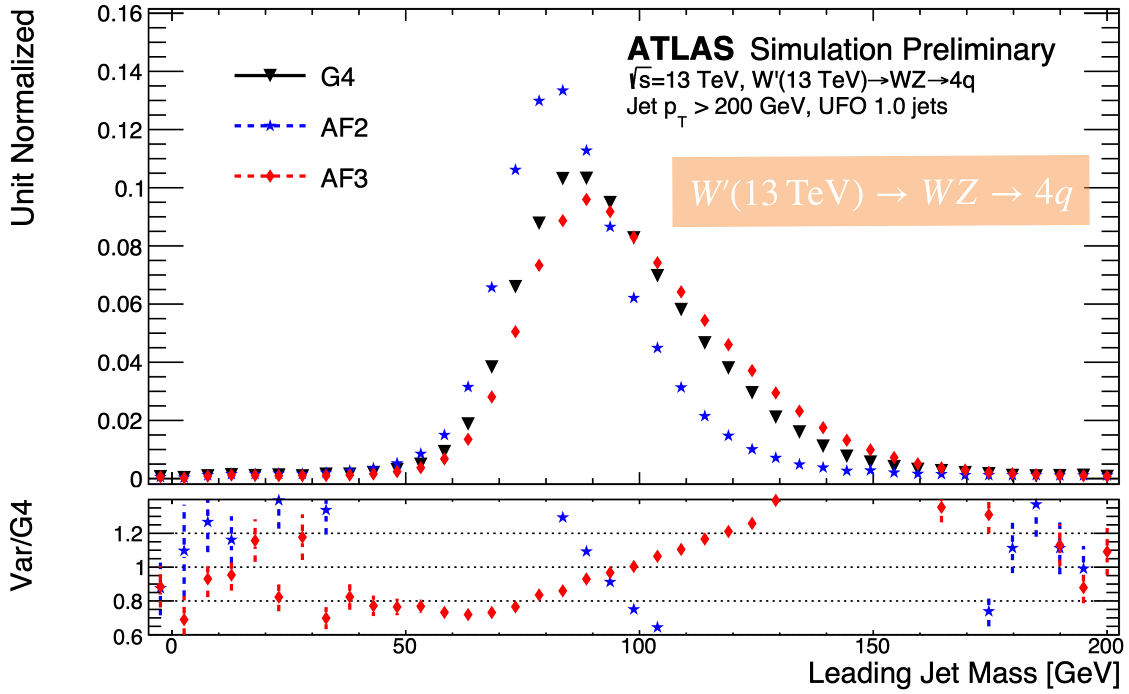
$Z'(4\,\text{TeV}) \rightarrow t\bar{t}$

Improvements in modelling of substructure variables will allow more analysis to use fast simulation in ATLAS!

Performance: Jet mass

Leading jet mass with UFO 1.0 jets





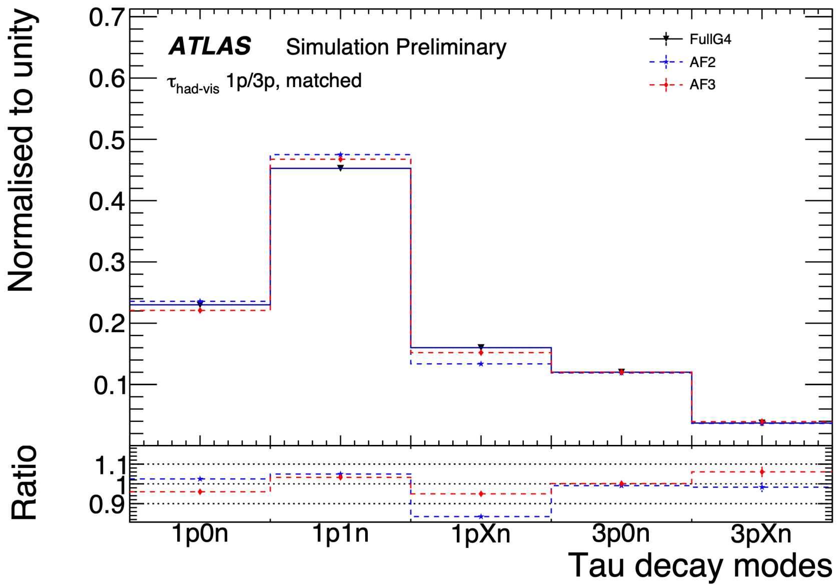
- Significant improvements in modelling leading jet mass in high energetic jets compared to AF2
- Room for improvement to correct remaining mismodelling, especially in the tails of the distribution

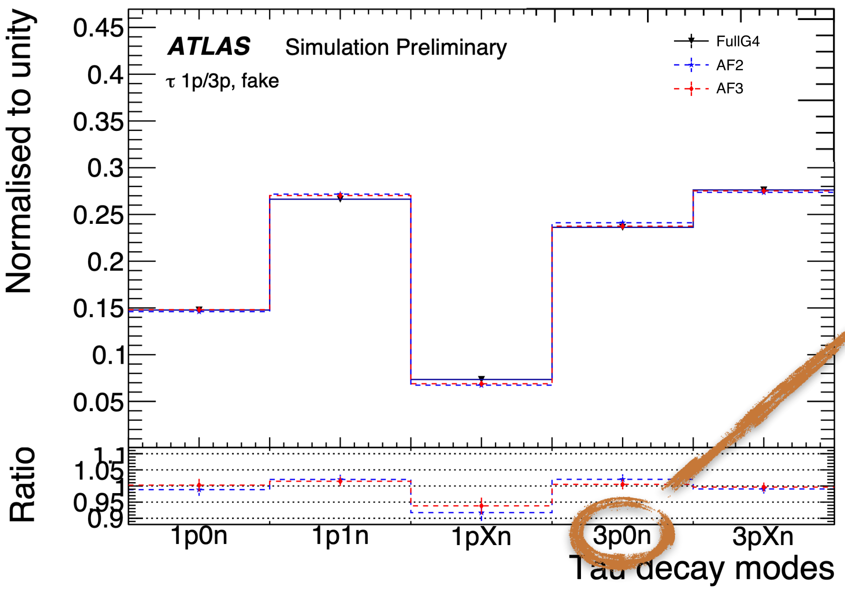
Performance: Hadronic τ leptons

True τ's

Fake τ's

Tau decay classification

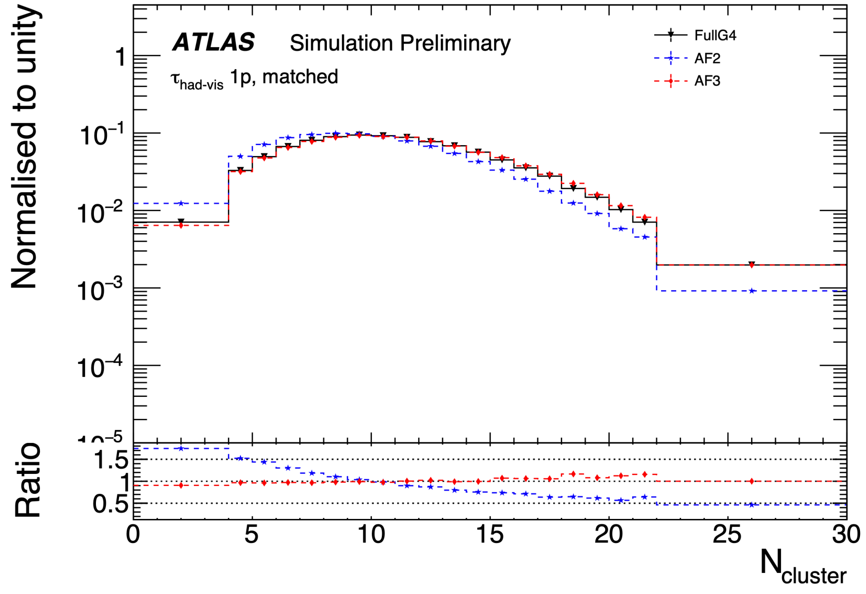


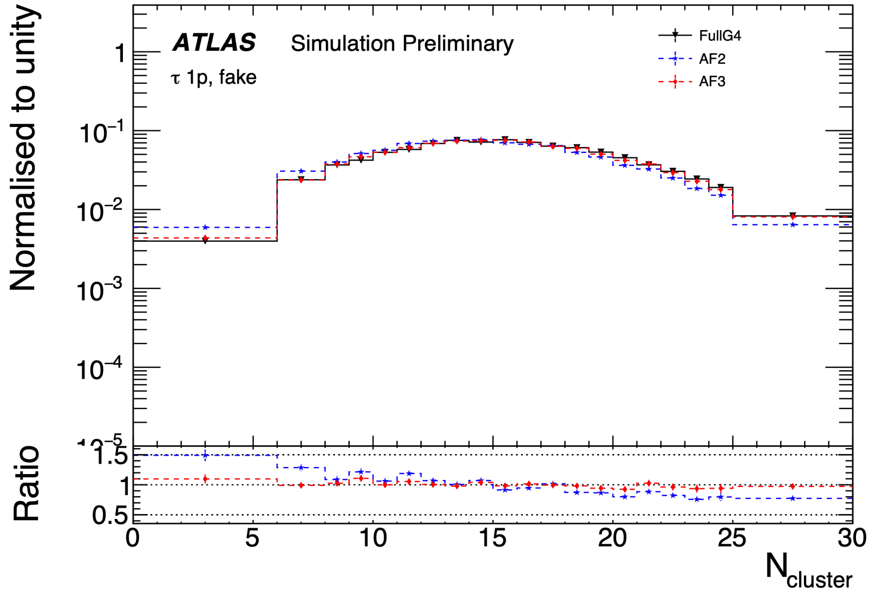


τ decay to three charged particles ("3-prong") and 0 neutrals

Validation with $Z \rightarrow \tau\tau$ Drell-Yan sample in 2.0-2.5 TeV mass range

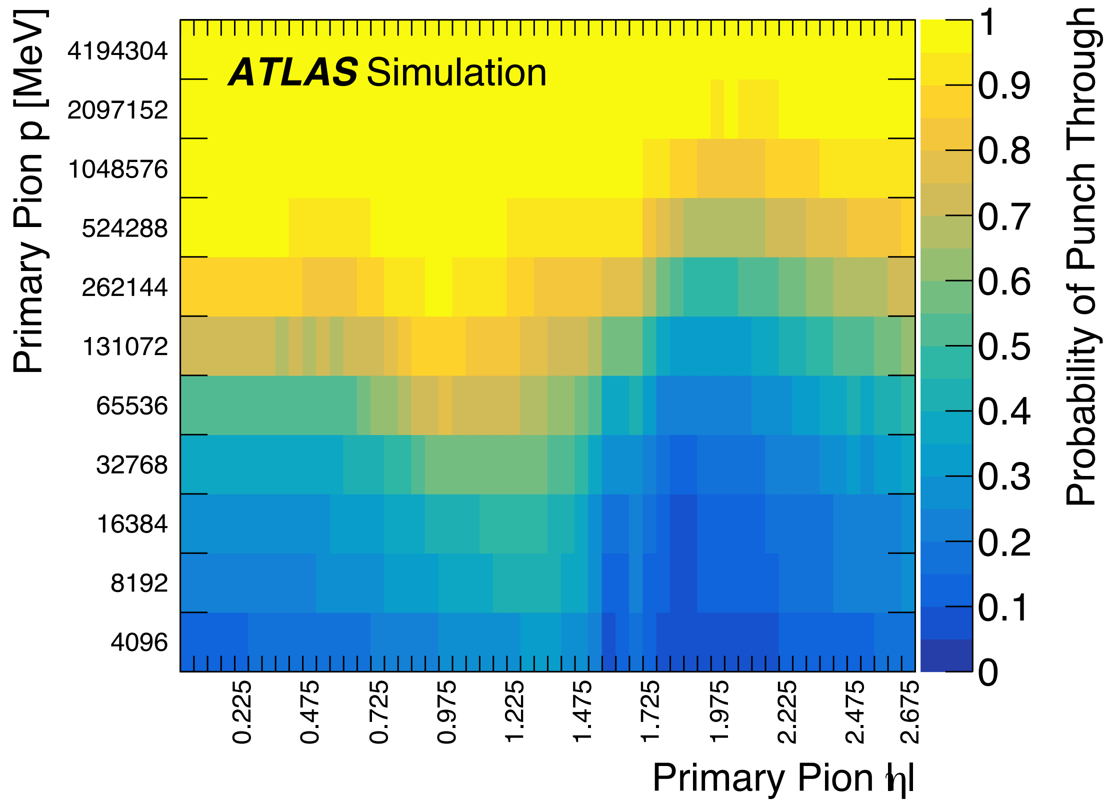
Number of clusters

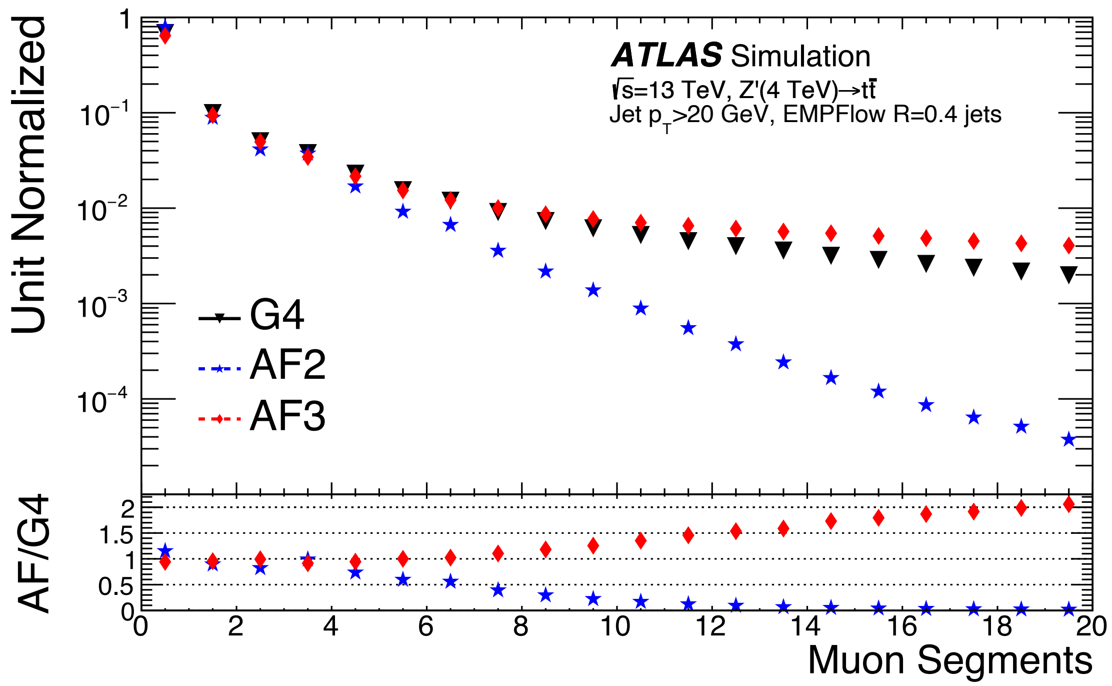




Good performance for true and fake τ's in AF3!

Performance: Punch-through muons





- Secondary particles in hadronic showers can escape calorimeter to the Muon Spectrometer
- Muon punch-through particles important for modelling background
- AF3 parametrises number of secondary particles and kinematics
- During simulation, secondary particles are created and passed to Geant4

Summary

- Encompassing complex parametrised and deep learning algorithms, **AF3 is the state of the art fast simulation in ATLAS** and able to simulate a broad range of physics processes with high precision
- AF3 provides **significant improvements in physics performance** compared to AF2 while giving a speedup of $\mathcal{O}(10)$ compared to Geant4
- Improvements include better modelling of jet masses and jet constituents as well as jet substructure, better e/γ simulation and more
- AF3 was used for the re-processing of **~7 billion Run 2 events**
- Many more improvements expected for Run 3 and beyond

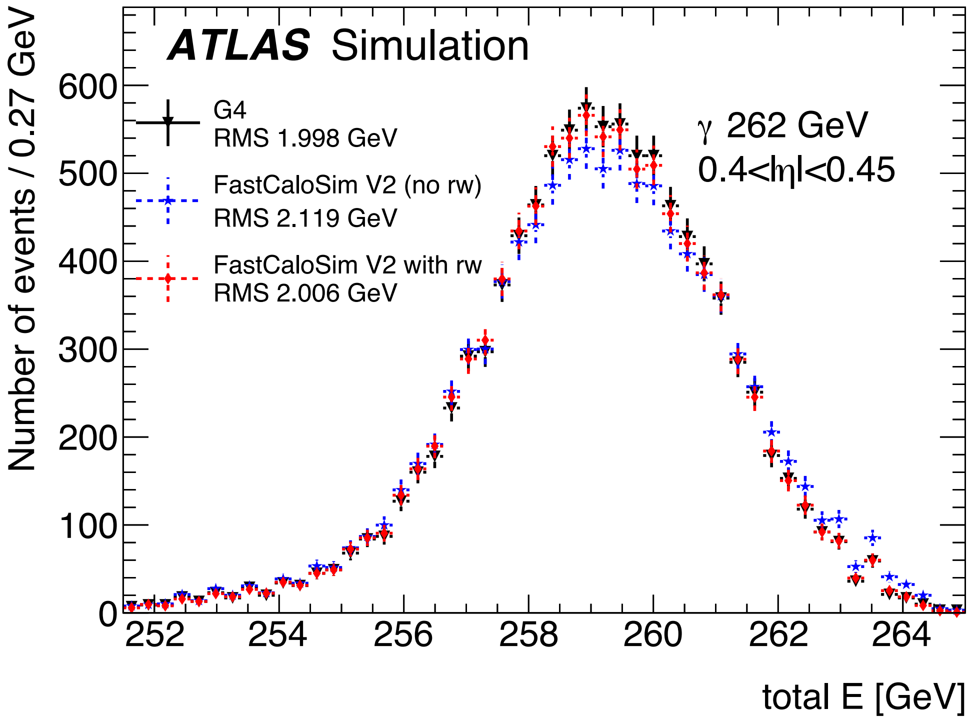
Thank you for your attention and stay tuned!

Backup

Energy Response Corrections

Energy Resolution Correction

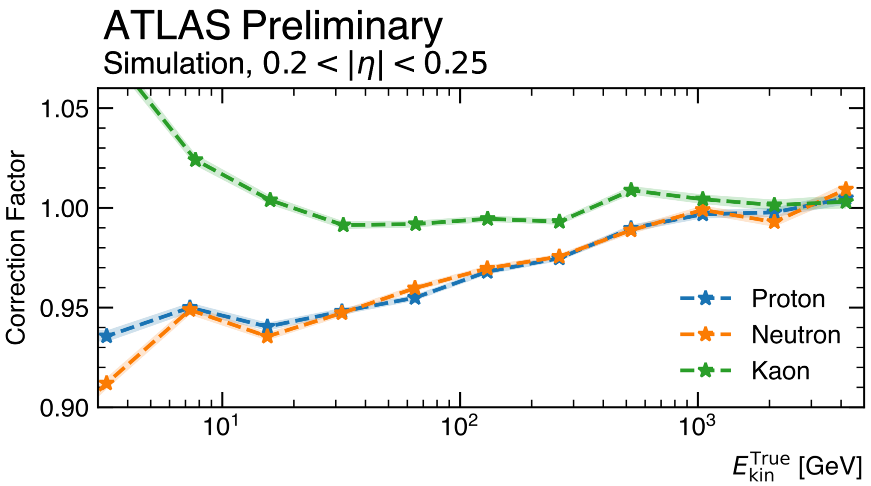
Probabilistic reweighting to reject simulated energies that are far off from G4 distribution using PDF derived from simulated over expected energy

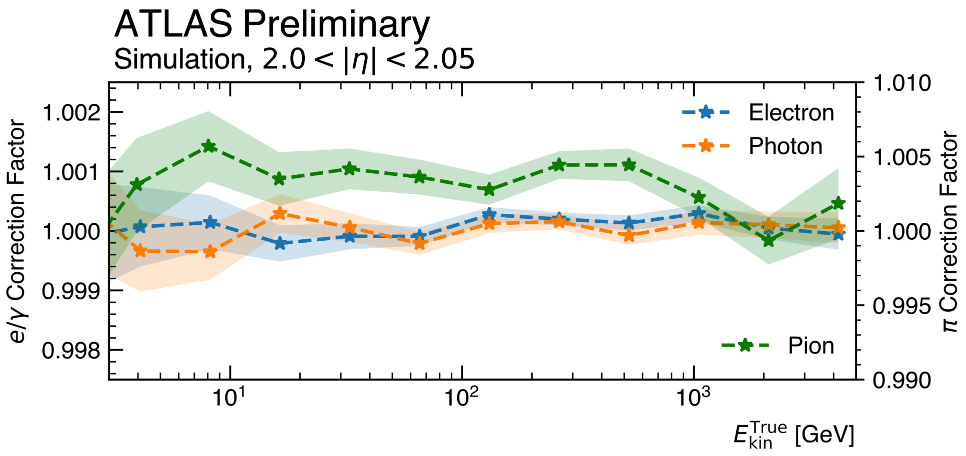


Energy Mean Correction

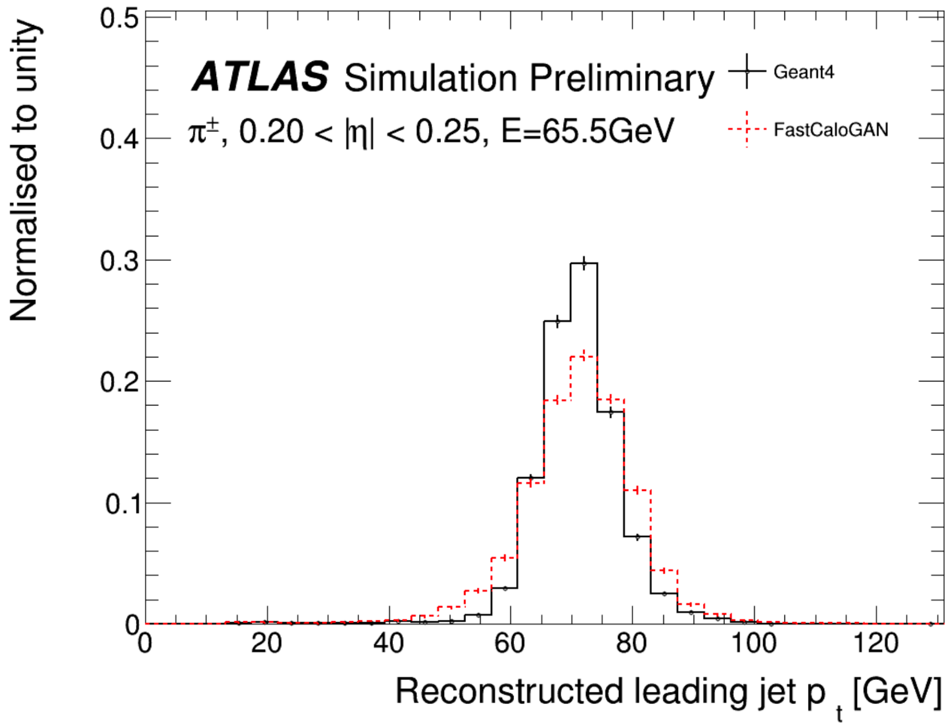
Remaining energy response differences corrected using fudge factors:

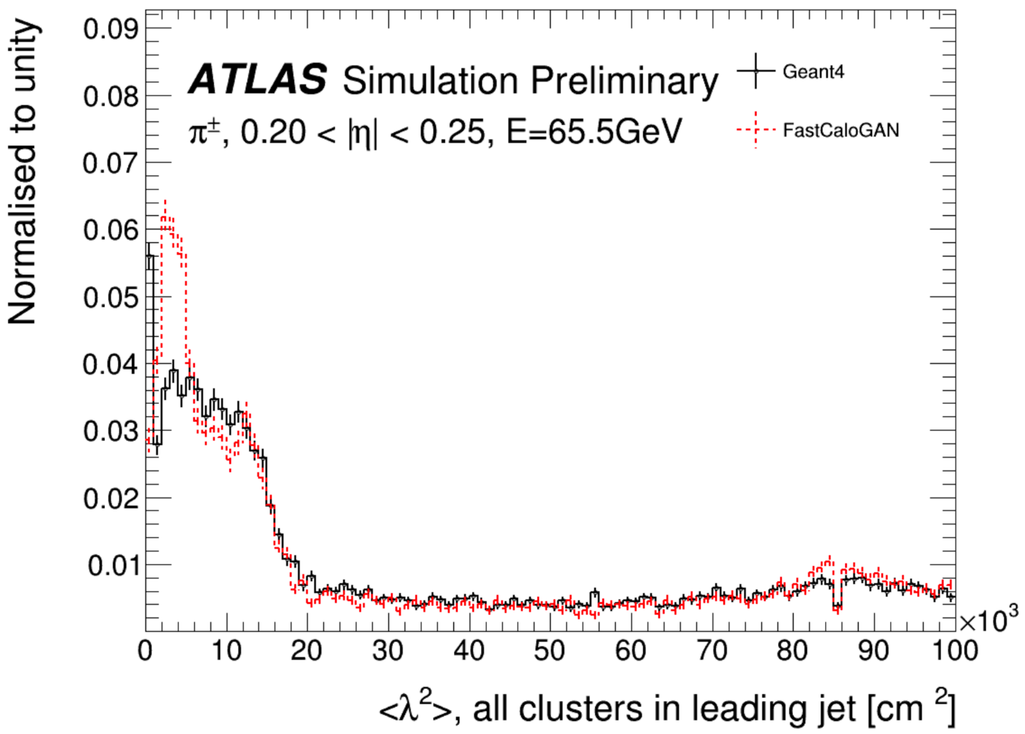
- Hadrons not included in parametrisation (protons, neutrons, kaons and anti-particles)
- Intrinsic energy differences in e, γ and π resulting from digitisation and reconstruction effects and an imperfect parametrisation

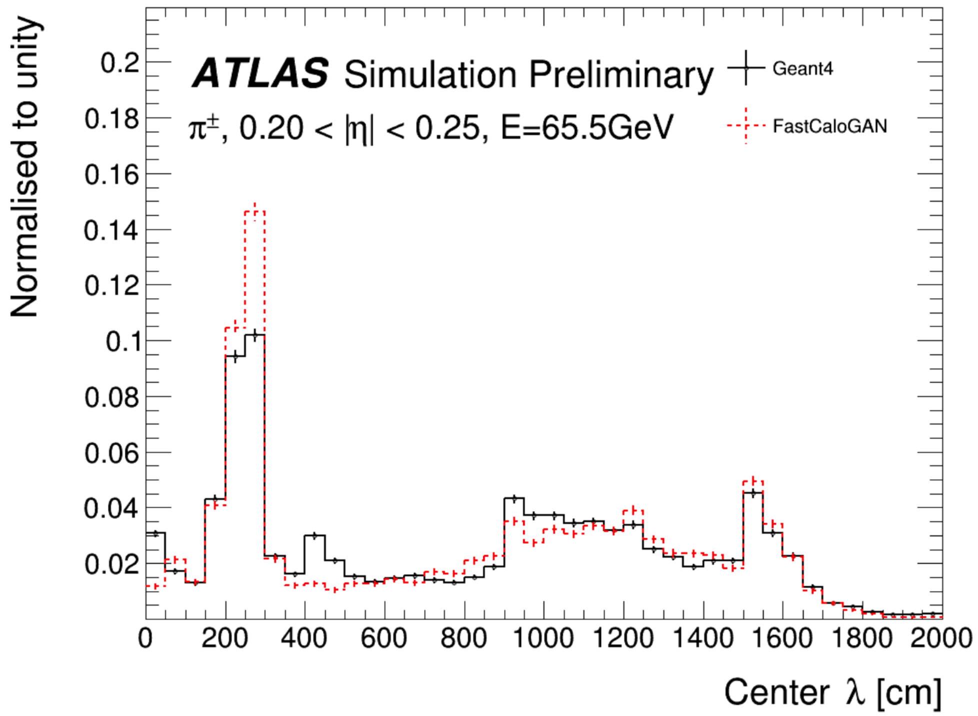


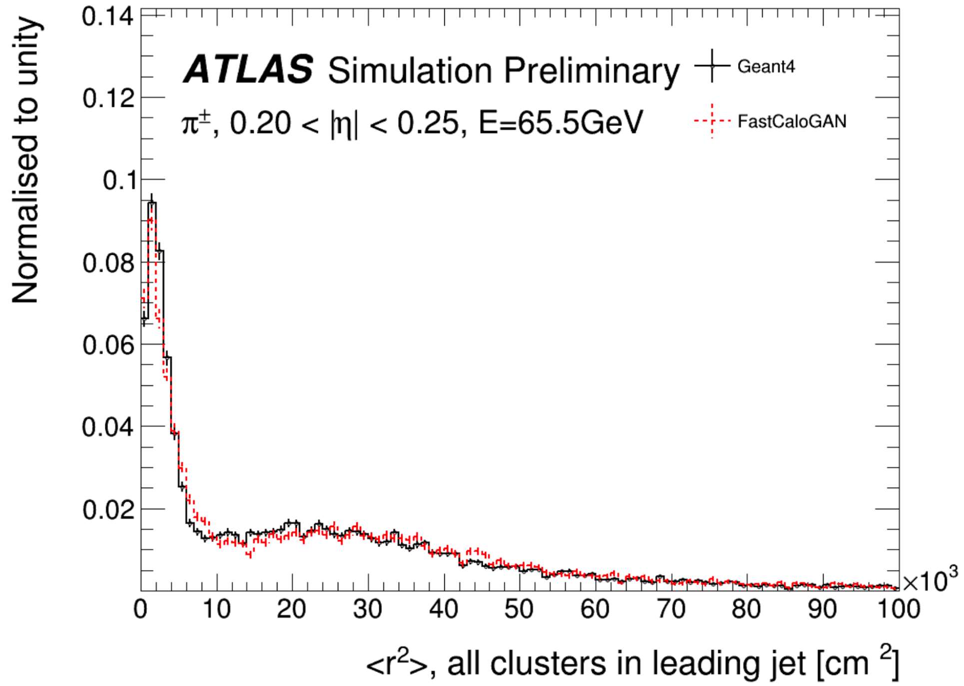


FastCaloGAN Performance





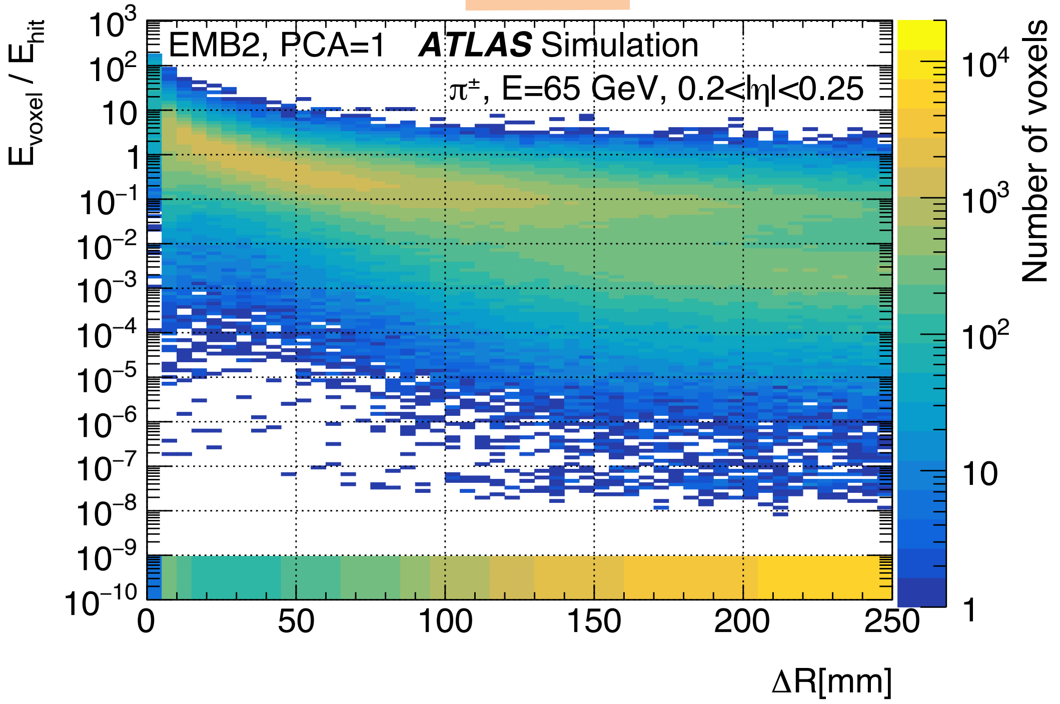




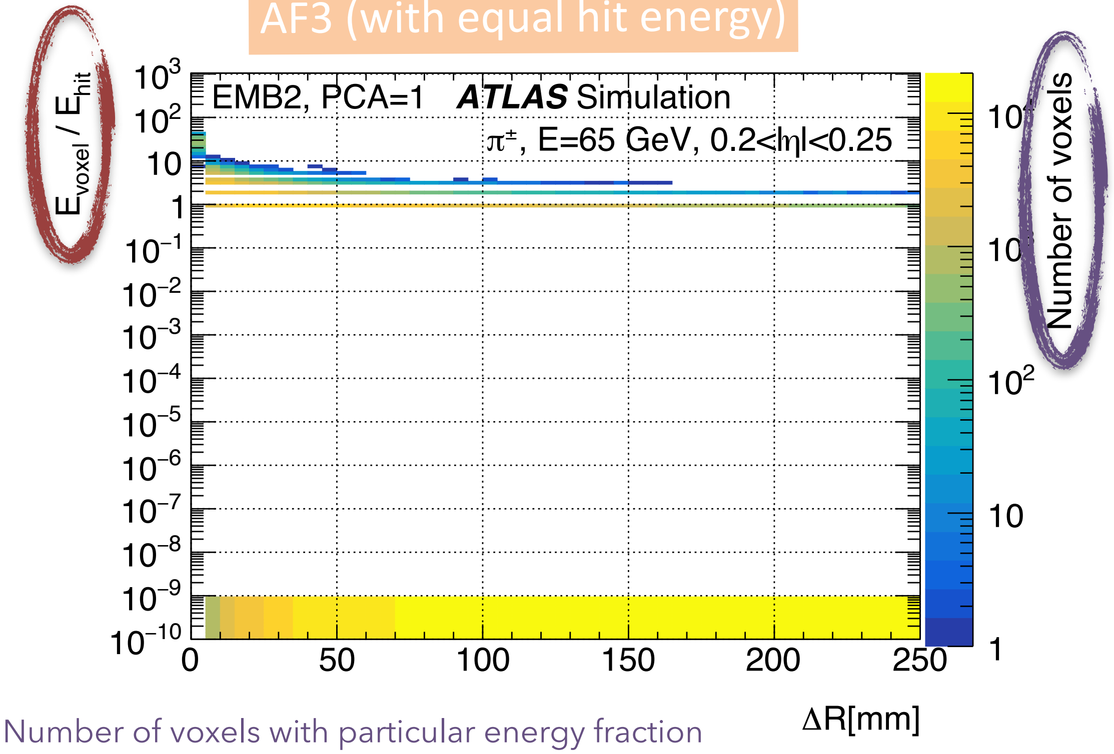
Cluster moments reproduced accurately

Weighted Hit Simulation for Hadrons

Geant4



AF3 (with equal hit energy)



Number of voxels with particular energy fraction

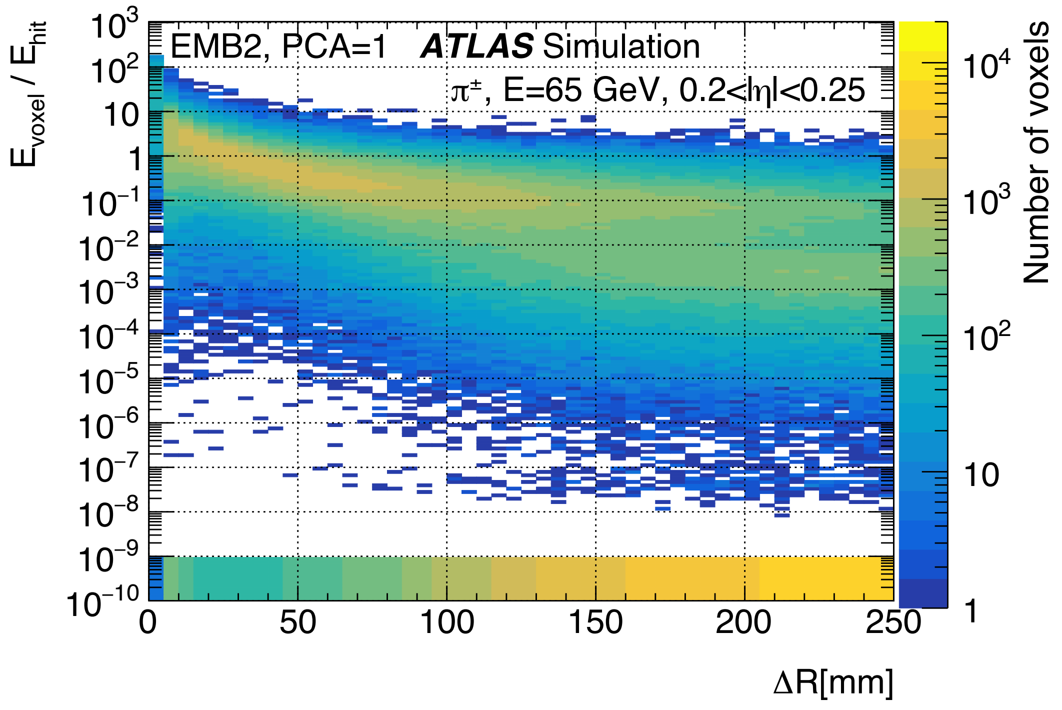
Energy fractions inside voxels along ΔR

- Large stochastic terms (>30%) in hadronic calorimeter layers lead to large energy deposits (100 - 300 MeV) for hits with equal energy
- Few hits far away from shower centre can create clusters
- Low energy clusters introduce mismodelling in total number of clusters

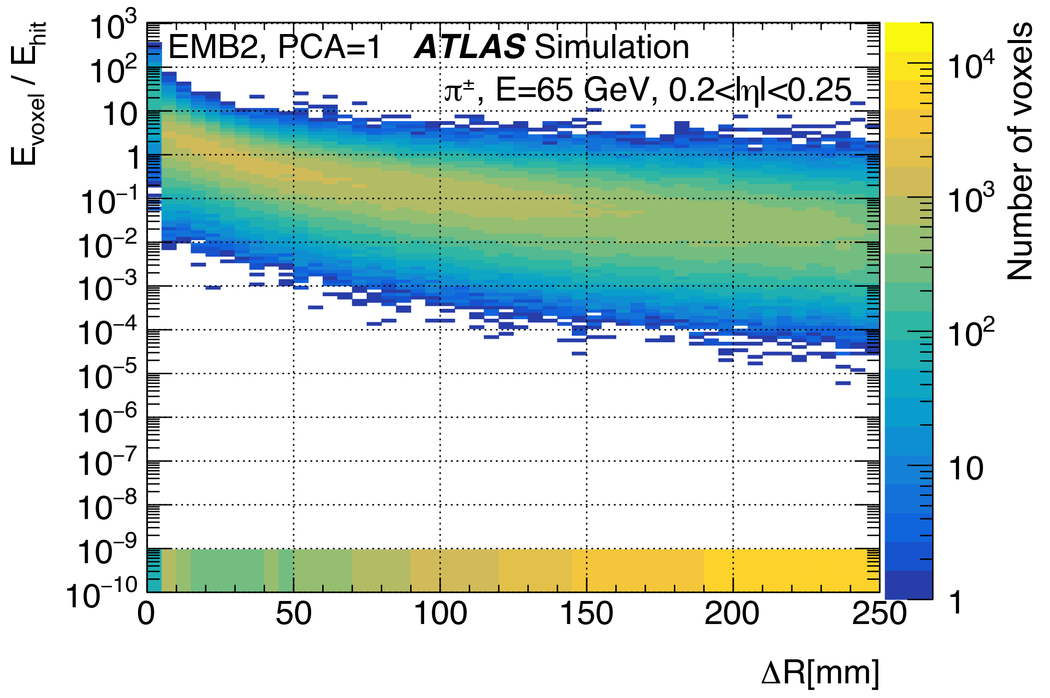
Geant4: far from shower centre most voxels have energy < E_{hit}

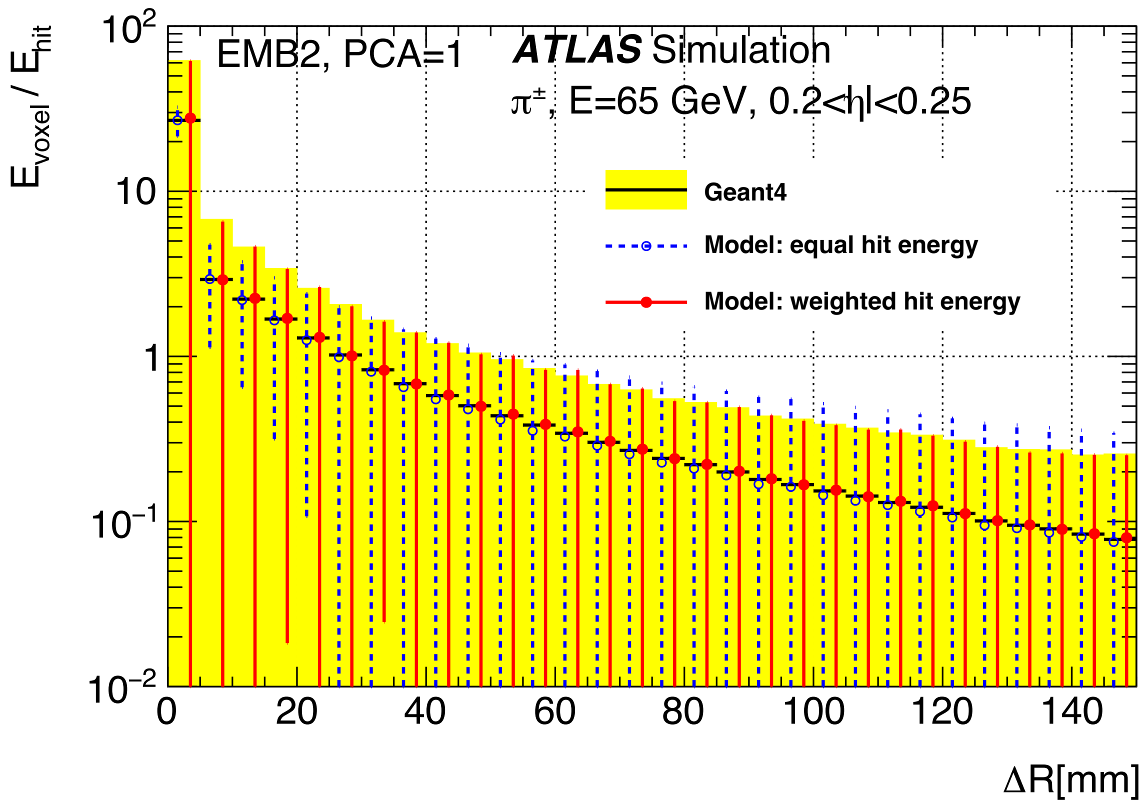
Weighted Hit Simulation for Hadrons

Geant4



AF3 (with weighted hits)

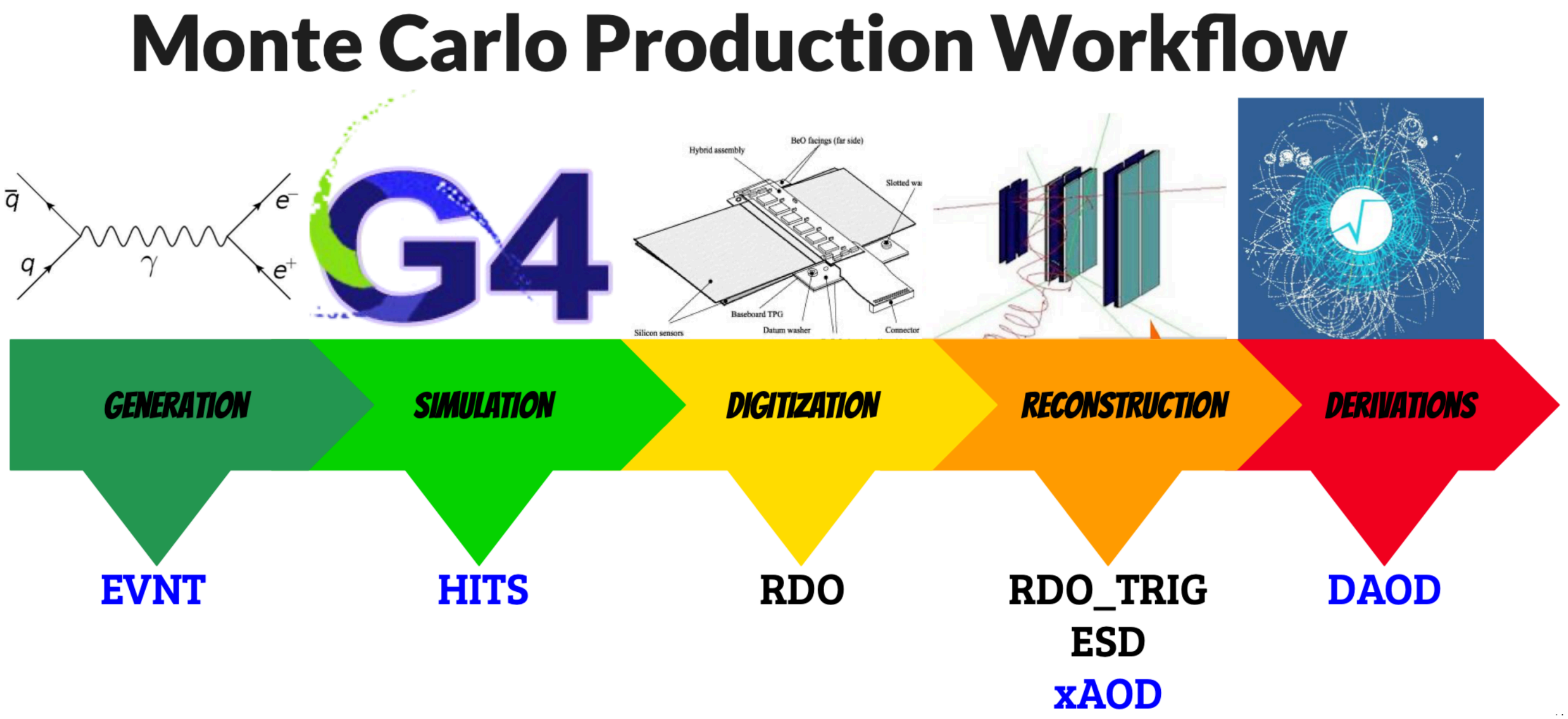


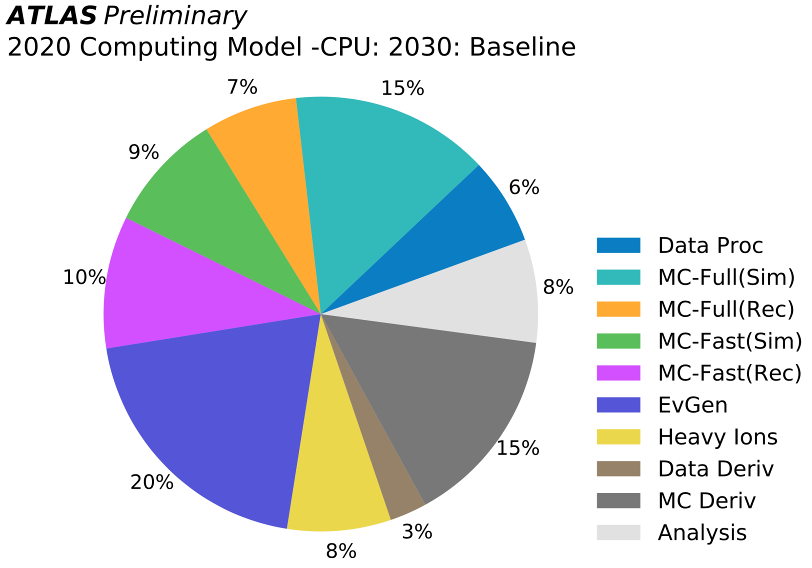


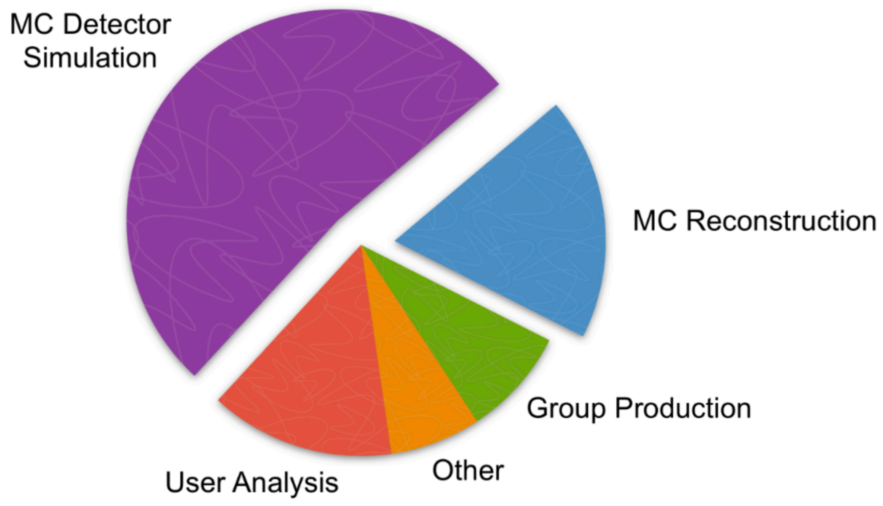
- Equal hit energies reproduce mean well but fail to reproduce RMS
- Introduce weights to change RMS of each ΔR bin to reproduce RMS observed in Geant4

Significant improvement in modelling of hadronic showers

Monte Carlo Production Workflow

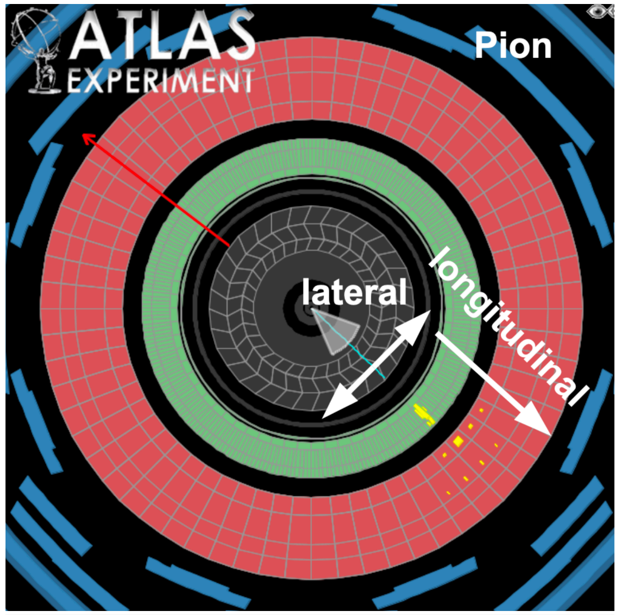


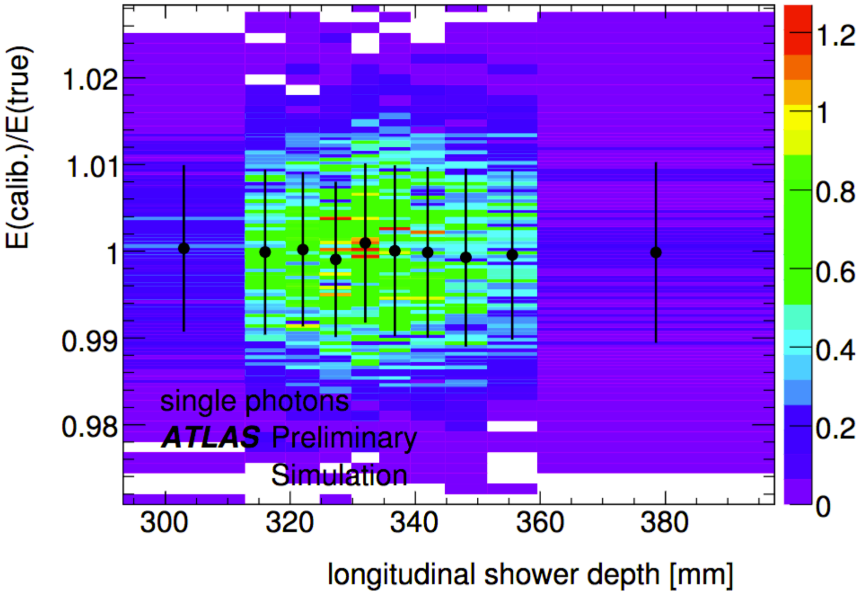


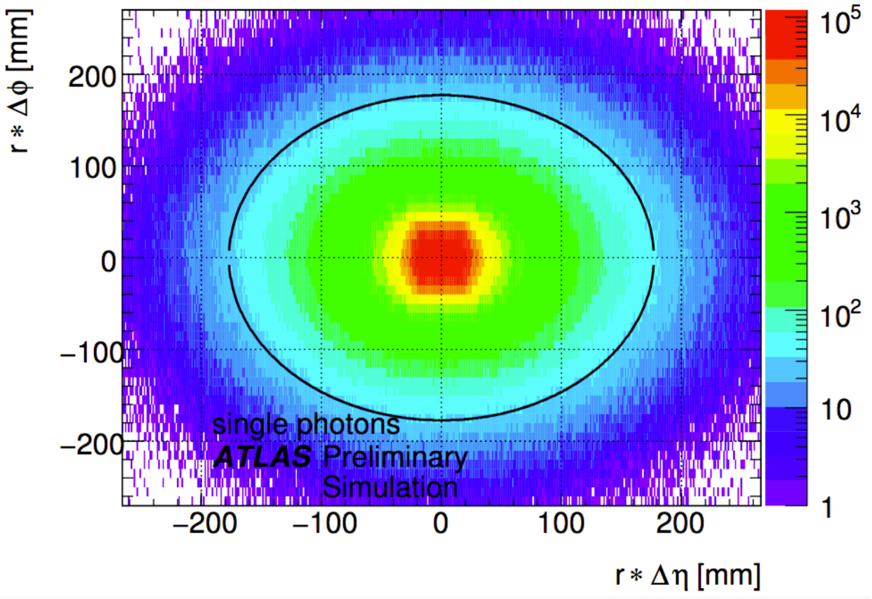


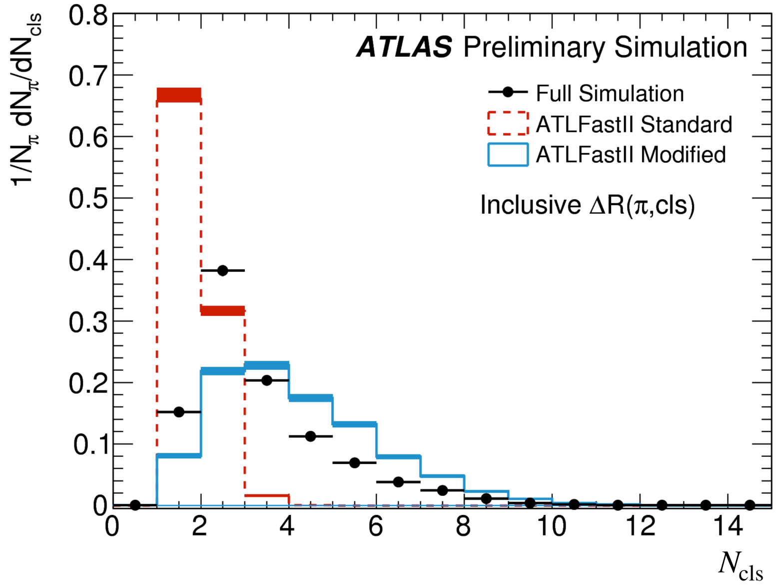
Fast simulation in Run 1 & Run II (AF2)

- AF2: parametrised calorimeter simulation used in ATLAS during Run 1 and Run 2
- e/γ and π used to parametrise EM and hadronic showers
- Simulate:
  1. **Longitudinal shower development:** E vs. d_{shower} and correlation between layers
  2. **Lateral shower development:** Average shower profile from radial symmetric function for each layer
- Good average shower description but complex variables not well modelled (e.g. jet substructure)
- No lateral parametrisation for Forward Calorimeter (FCal)
- No simulation of particles escaping calorimeter volume (punch through particles)



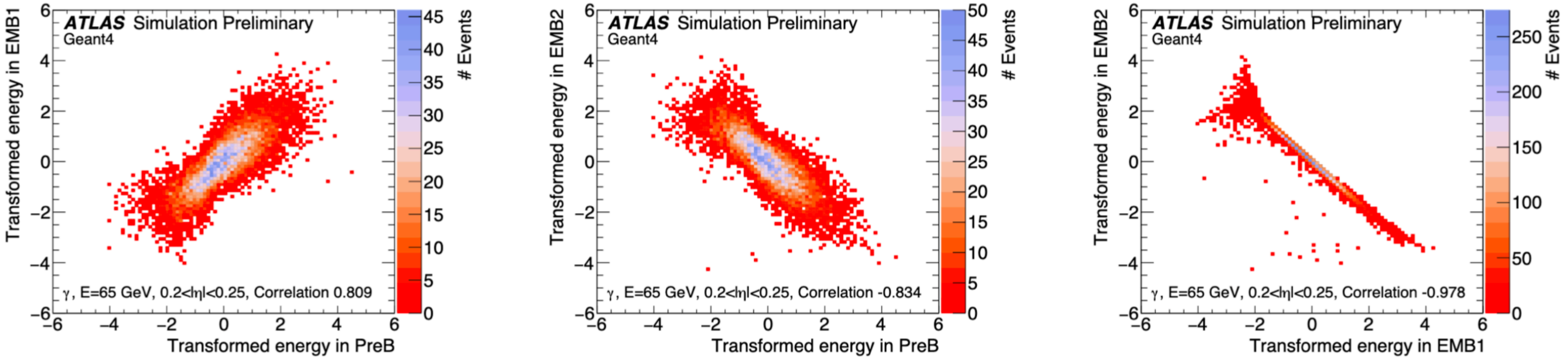






PCA Rotation

Before PCA:

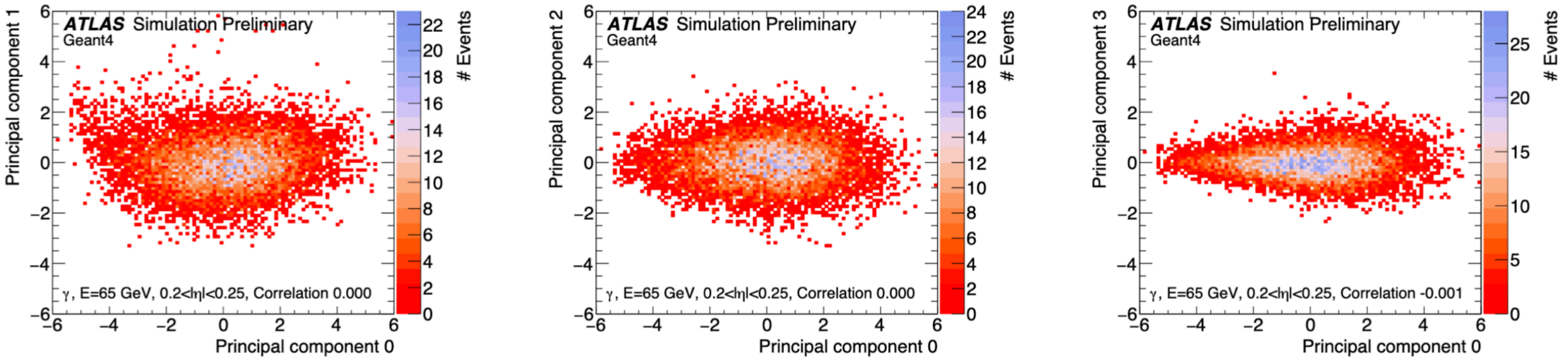


(a) Presampler vs EM Barrel 1

(b) Presampler vs EM Barrel 2

(c) EM Barrel 1 vs EM Barrel 2

After PCA:



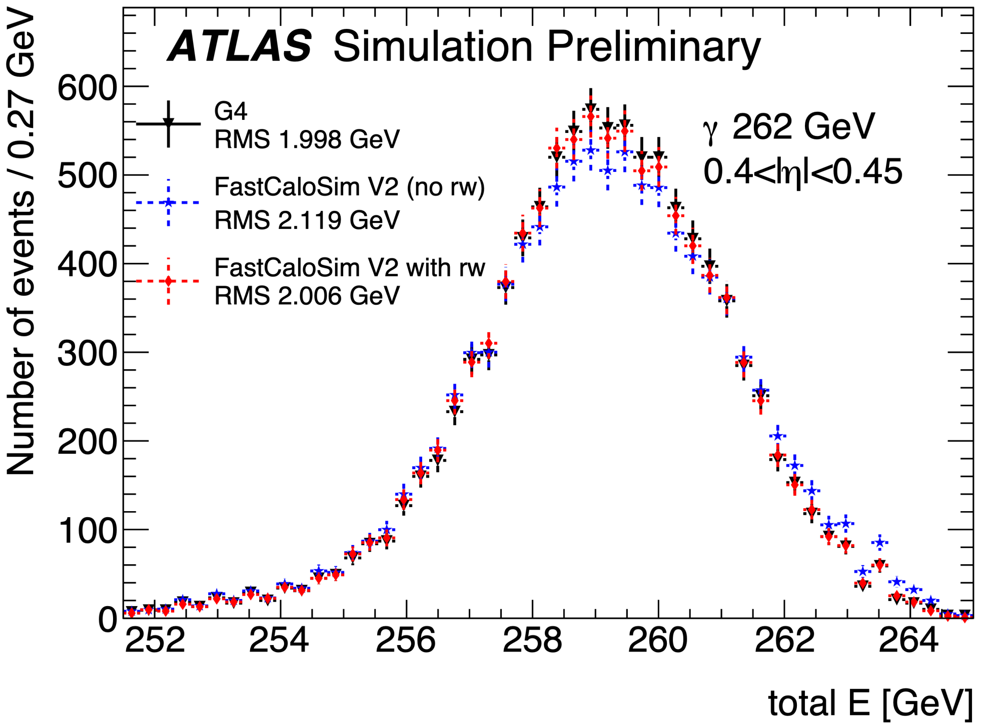
(a) First vs second PC

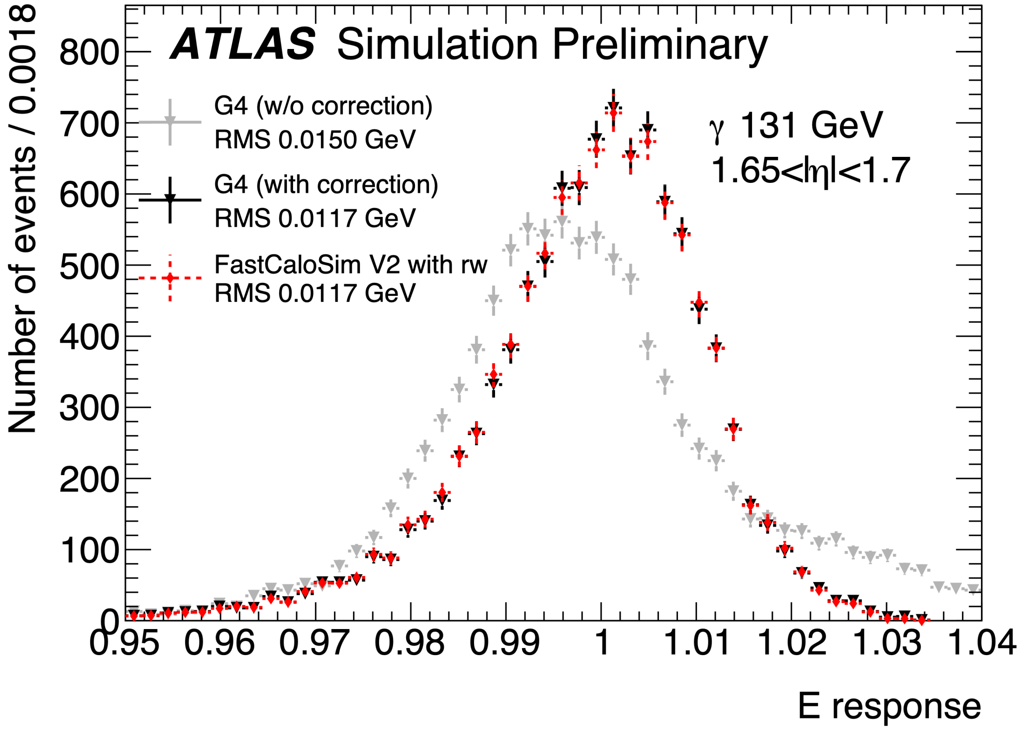
(b) First vs third PC

(c) First vs fourth PC

Energy Resolution Corrections

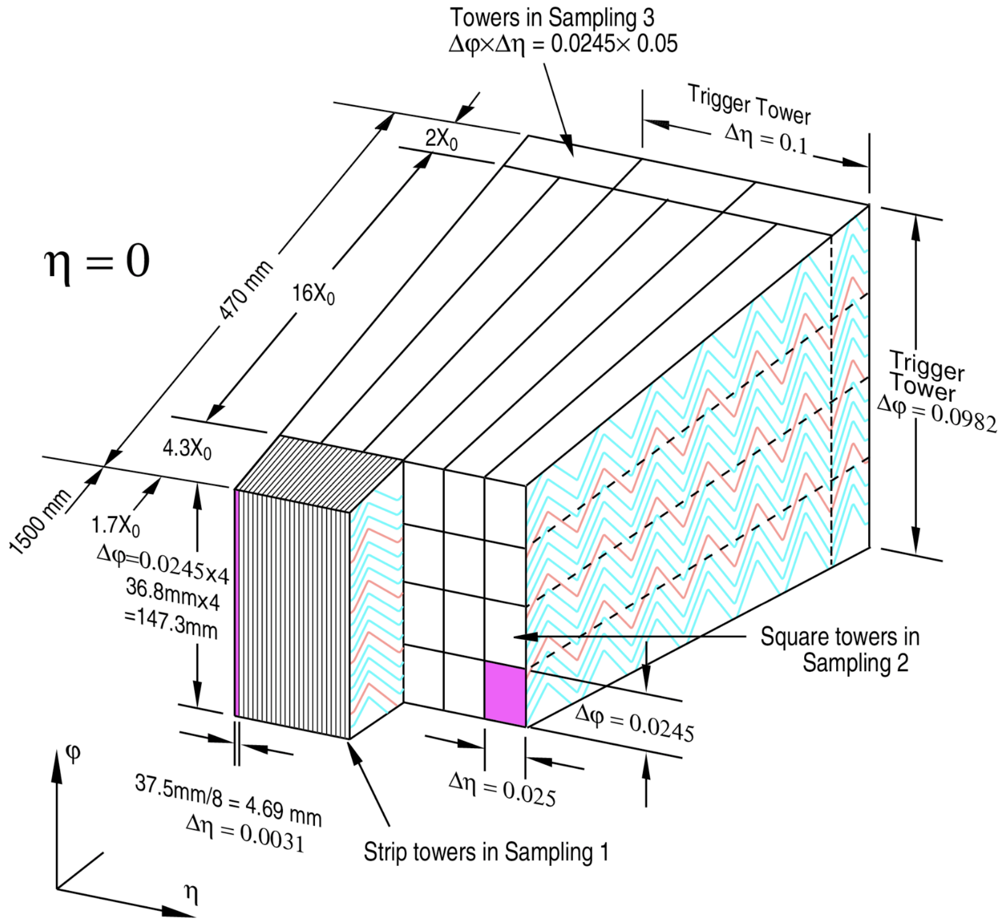
Probabilistic reweighting to reject simulated energies that are far off from G4 distribution using PDF derived from simulated over expected energy

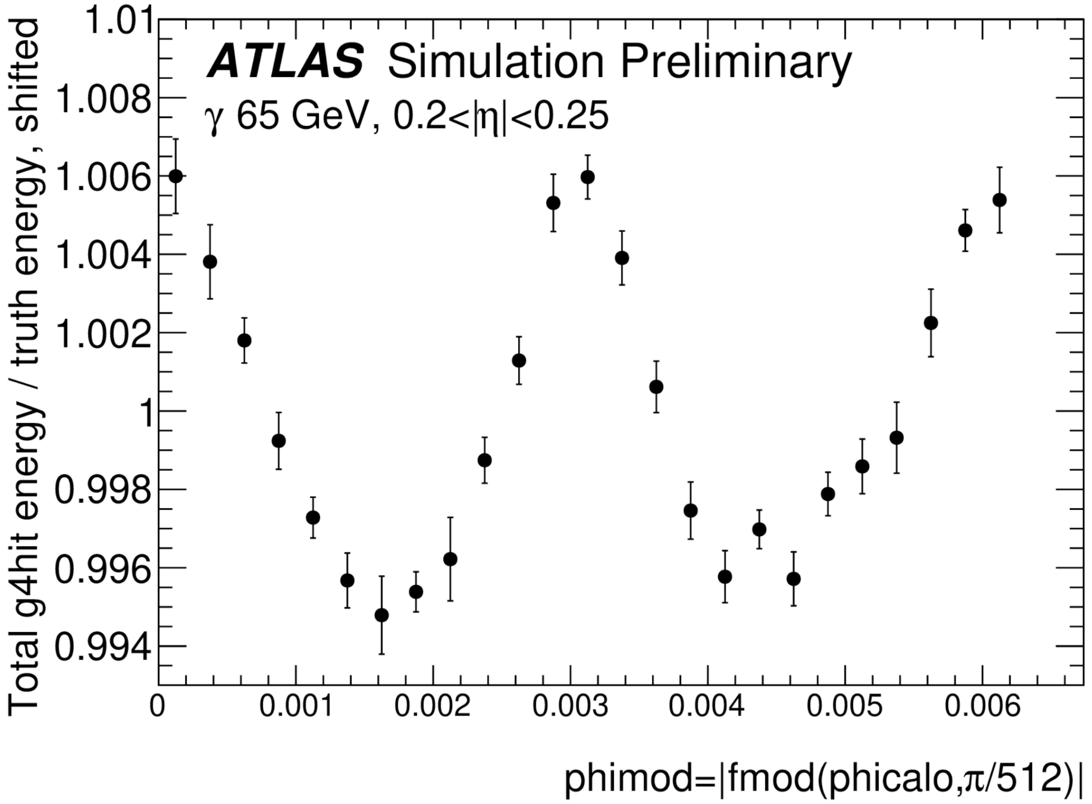




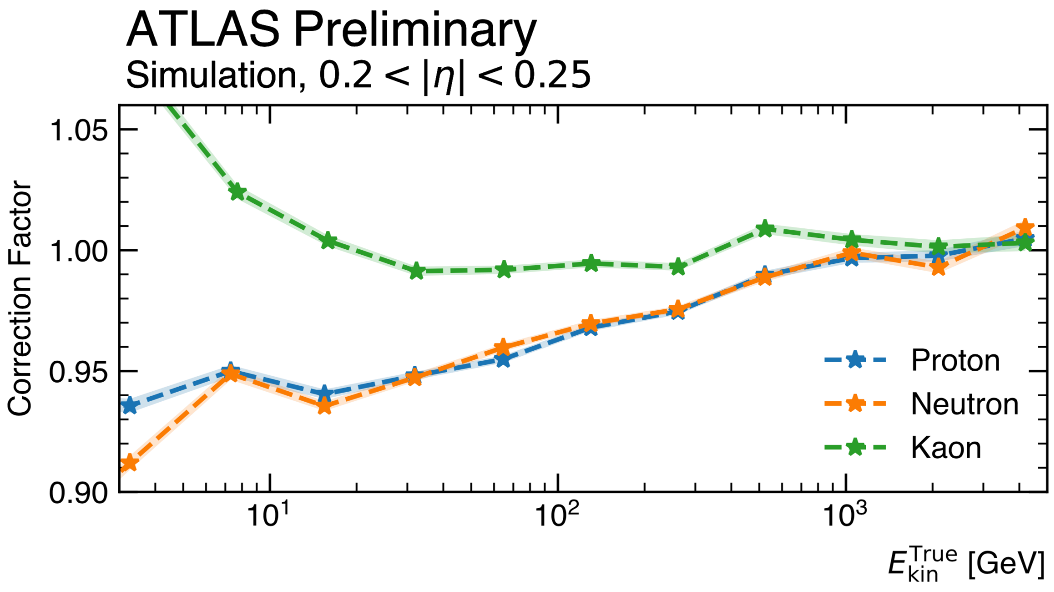
- modulation of energy in ϕ direction in input G4 due to Liquird Argon (LAr) accordion structure
- G4 uses calibration that removes modulation, while AF3 would use a calibration without modulation effect
- Geant4 inputs corrected before parametrisation to remove ϕ modulation

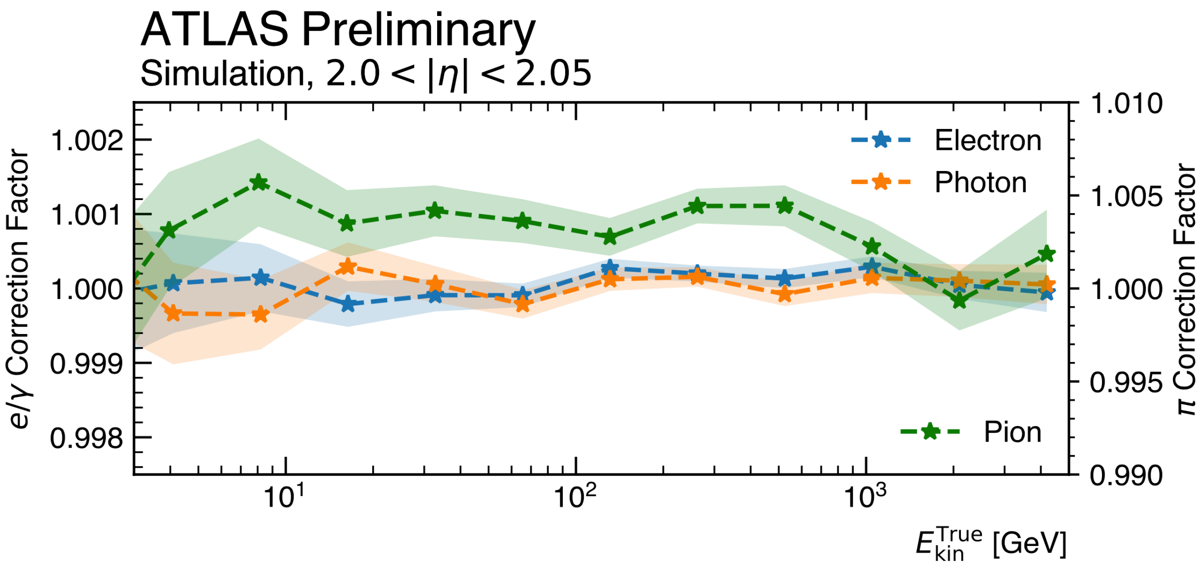
ϕ Modulation





Energy Response Corrections





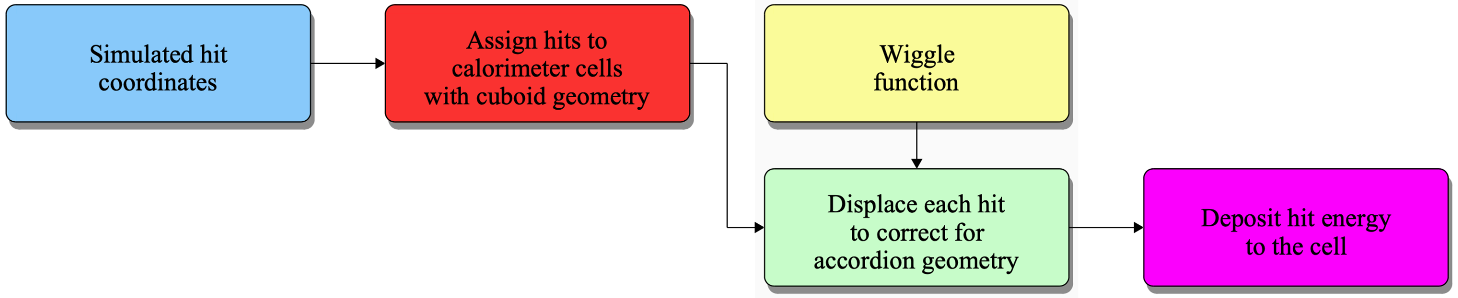
Remaining energy response differences corrected using fudge factors:

- Hadrons not included in parametrisation (protons, neutrons, kaons and anti-particles) corrected using $\bar{E}^{\mathrm{Hadron}}_{\mathrm{G4}}/\bar{E}^{\pi}_{\mathrm{G4}}$ fudge factors
- Intrinsic energy differences in e, γ and π resulting from digitisation and reconstruction effects and an imperfect parametrisation corrected with $\bar{E}_{\mathrm{G4}}/\bar{E}_{\mathrm{AF3}}$ fudge factors

Linear interpolation in between energy points

Hit to Cell Assignment

- Hits are assigned to cells which are assumed to have a simplified cuboid geometry
- EM calorimeter has accordion structure
- Wiggle function to displace hits to neighbouring cells with a certain probability



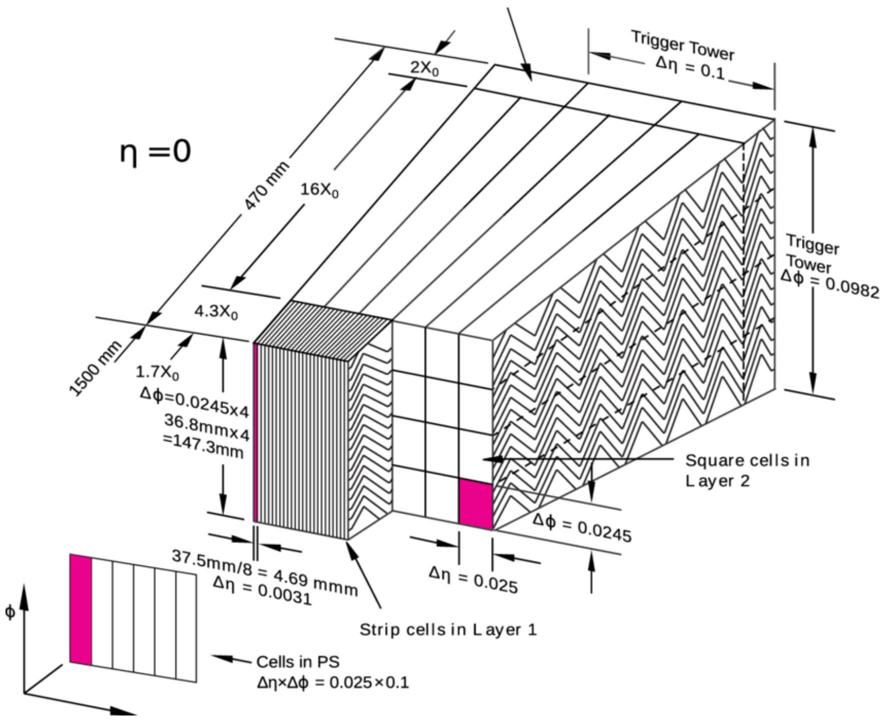
Simulated hit coordinates

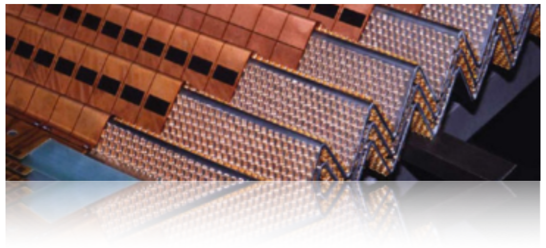
Assign hits to calorimeter cells with cuboid geometry

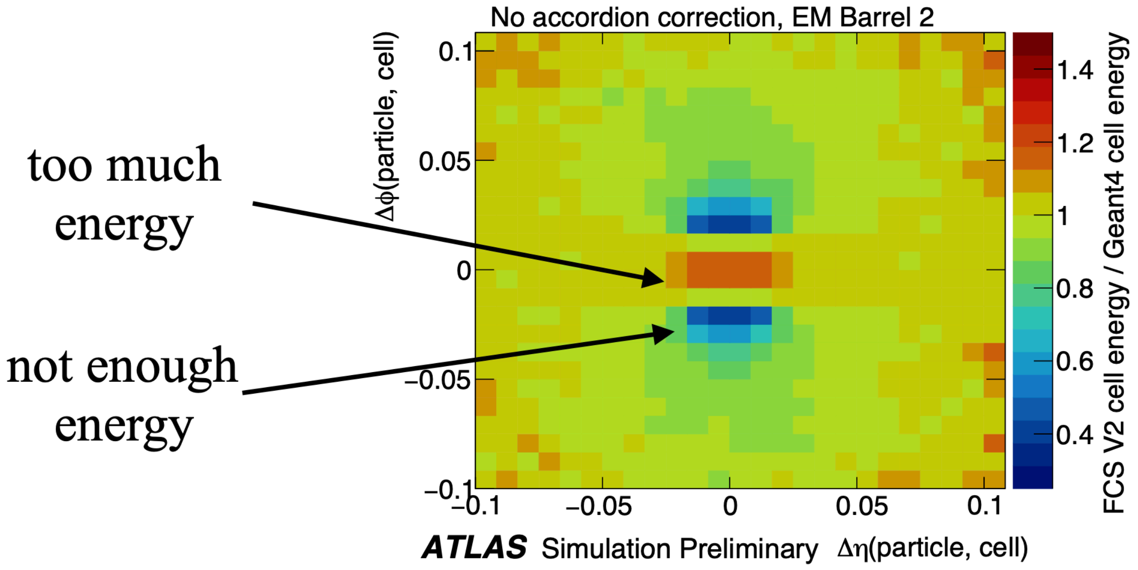
Wiggle function

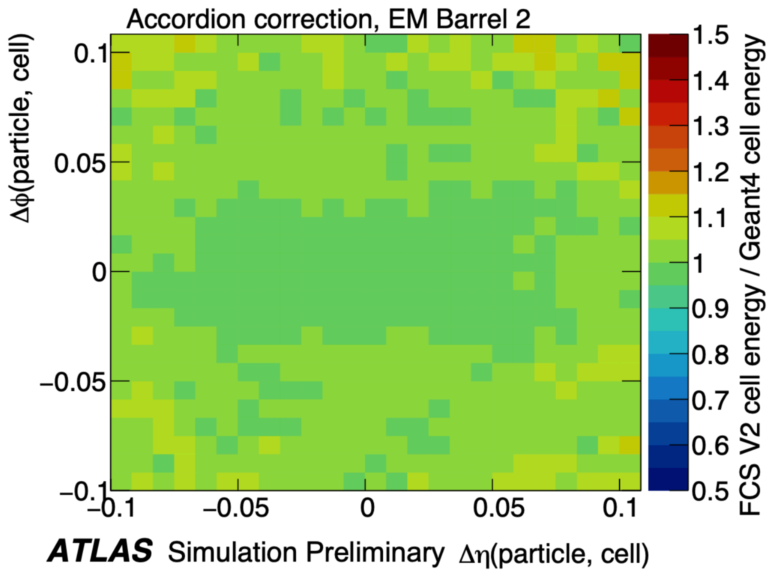
Displace each hit to correct for accordion geometry

Deposit hit energy to the cell

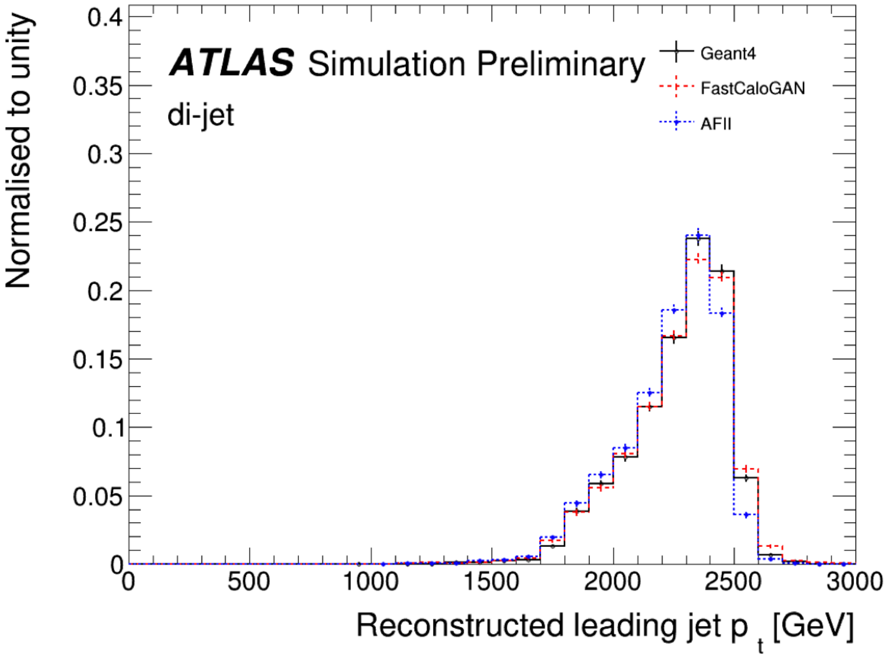


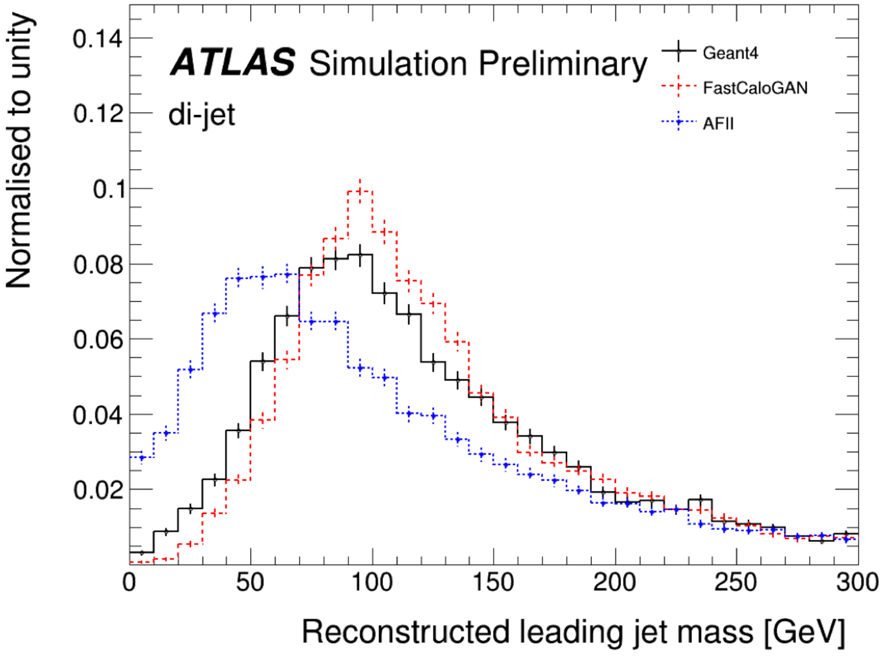


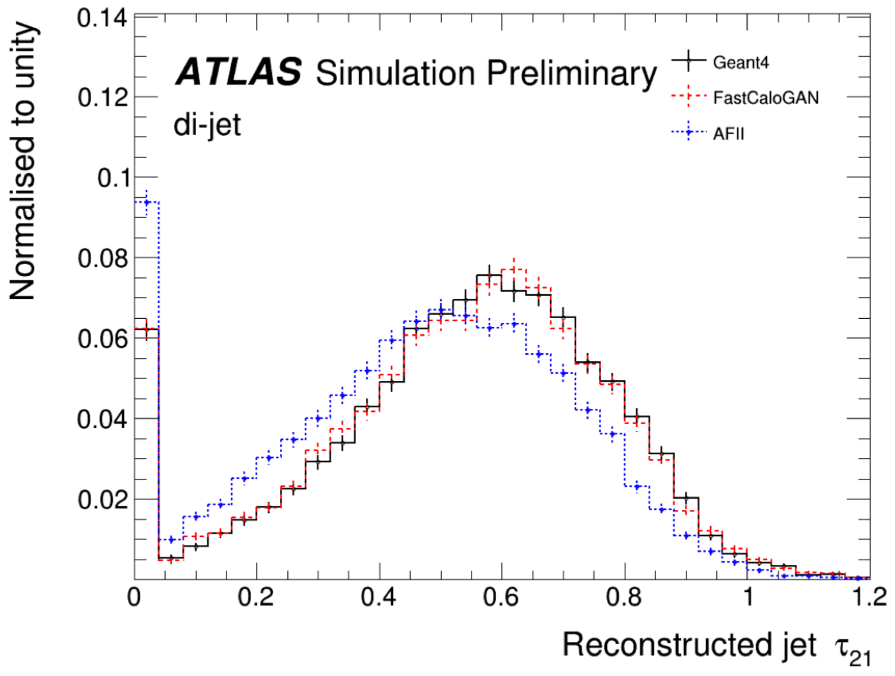

No accordion correction, EM Barrel 2

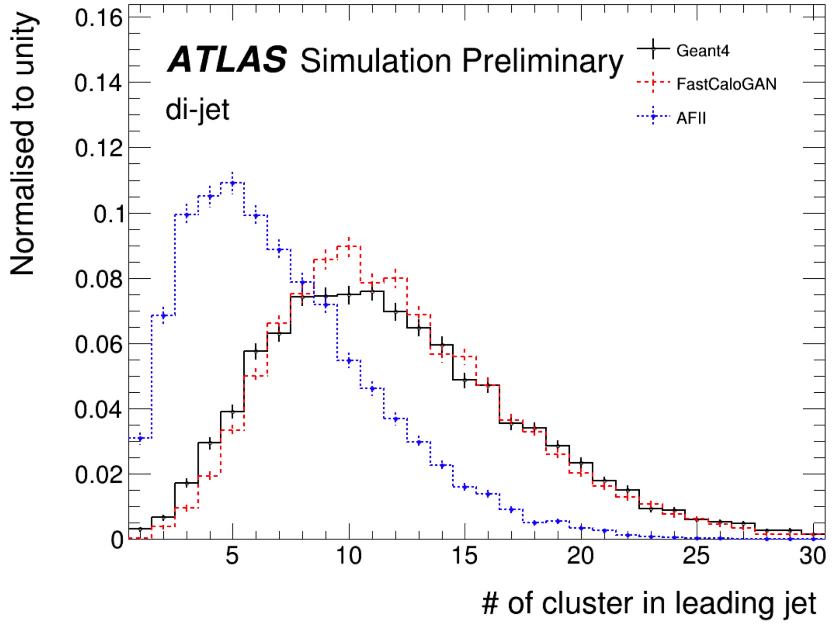

Accordion correction, EM Barrel 2

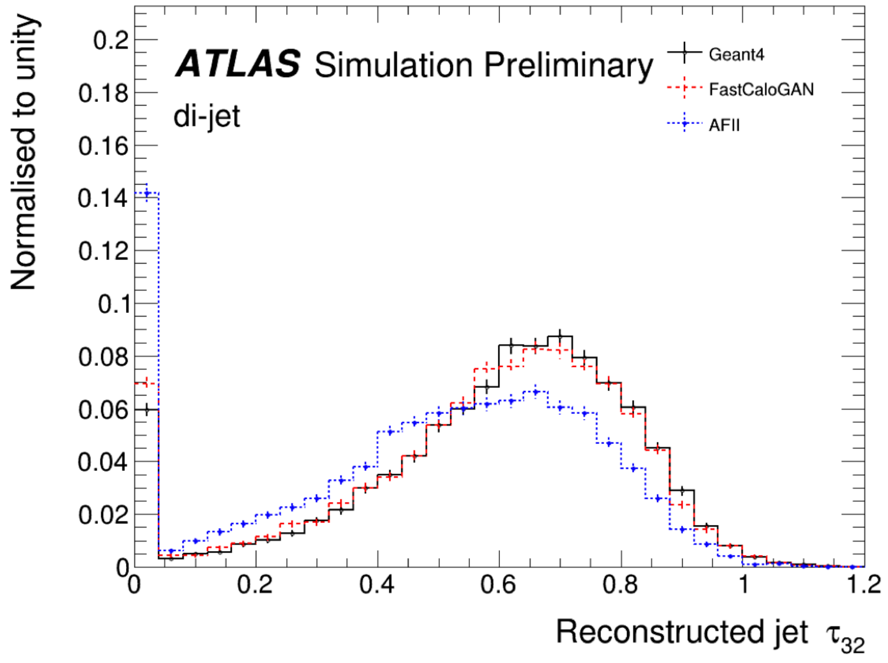
FastCaloGAN: Jet Performance

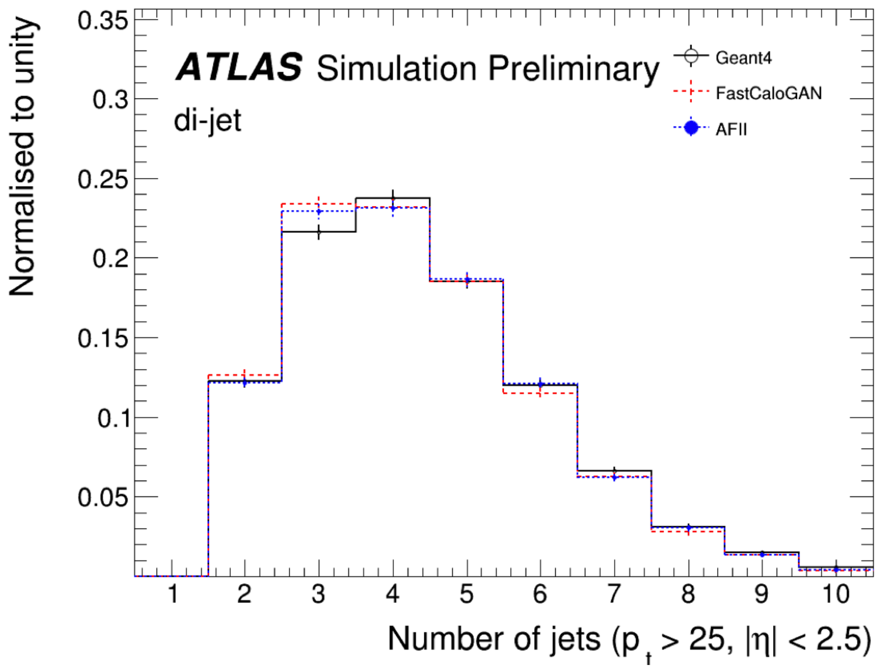




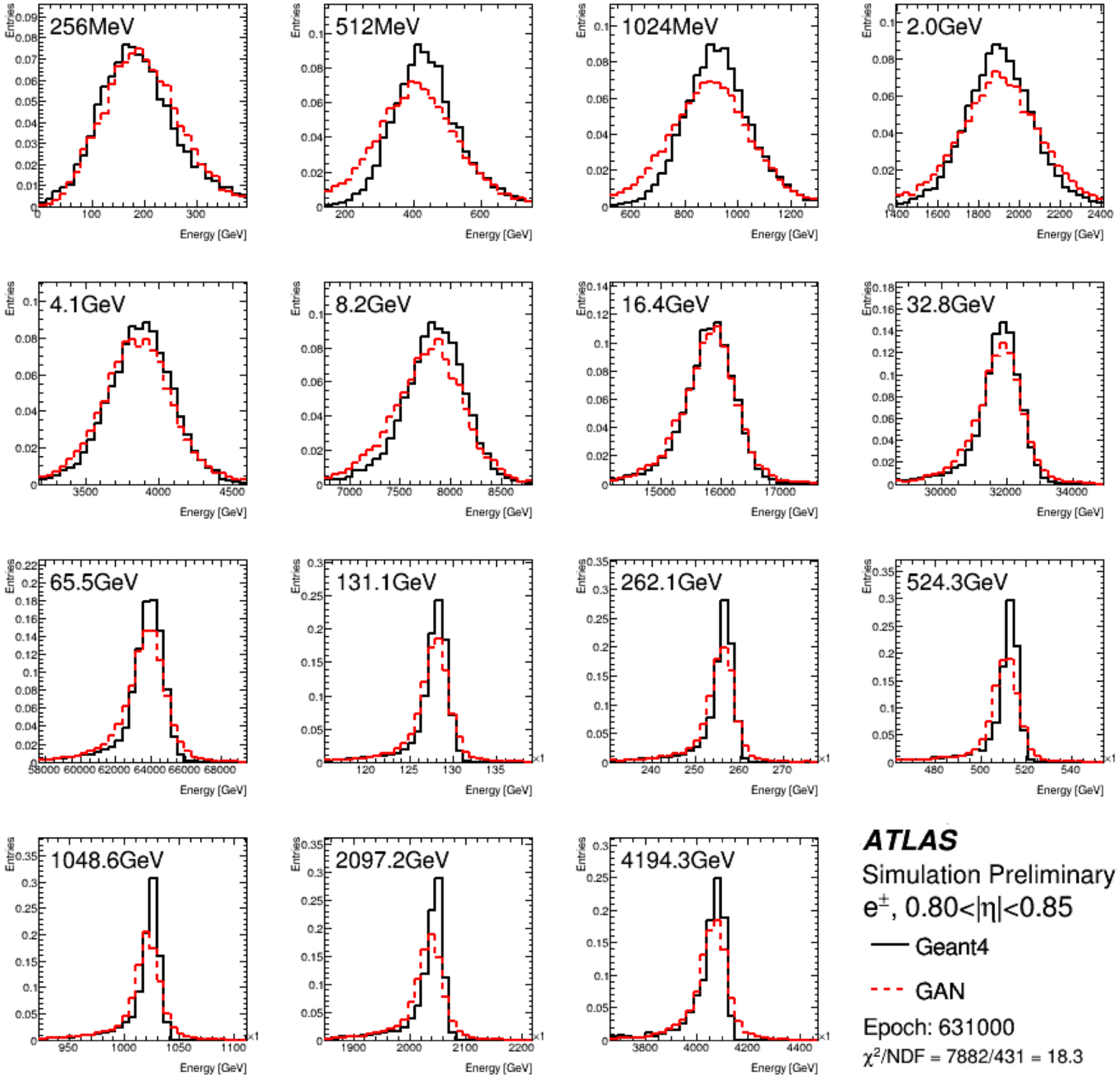


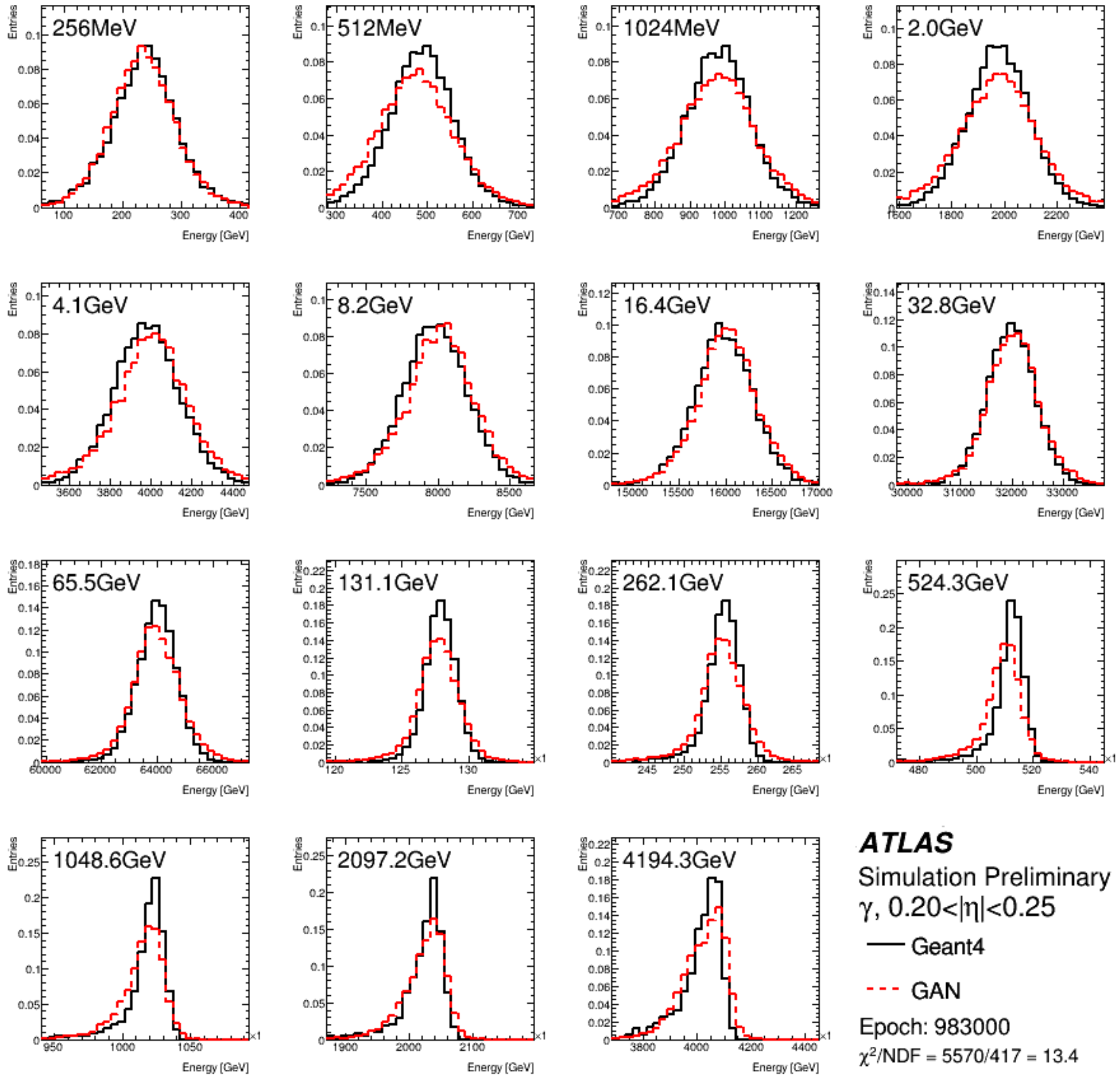




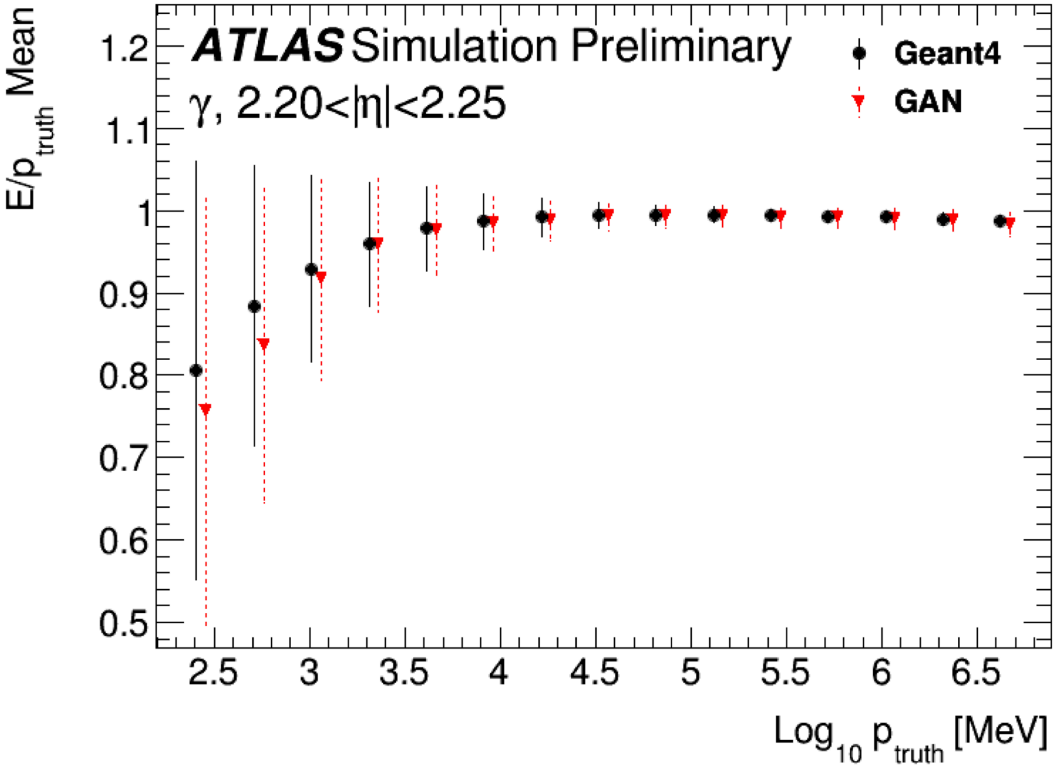


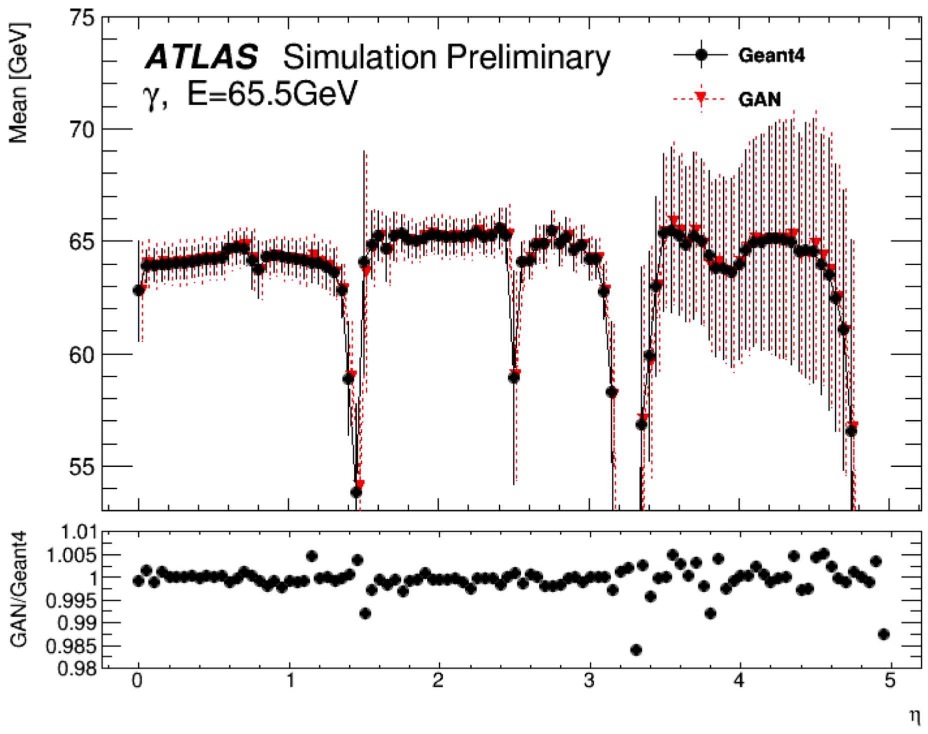
FastCaloGAN: e/gamma performance

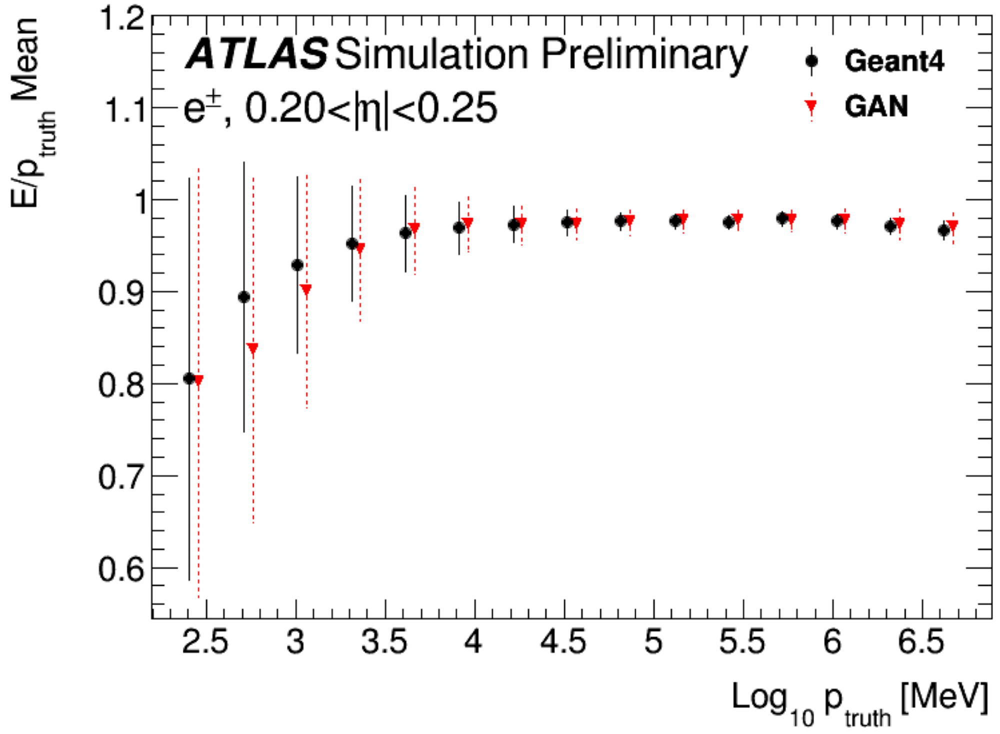


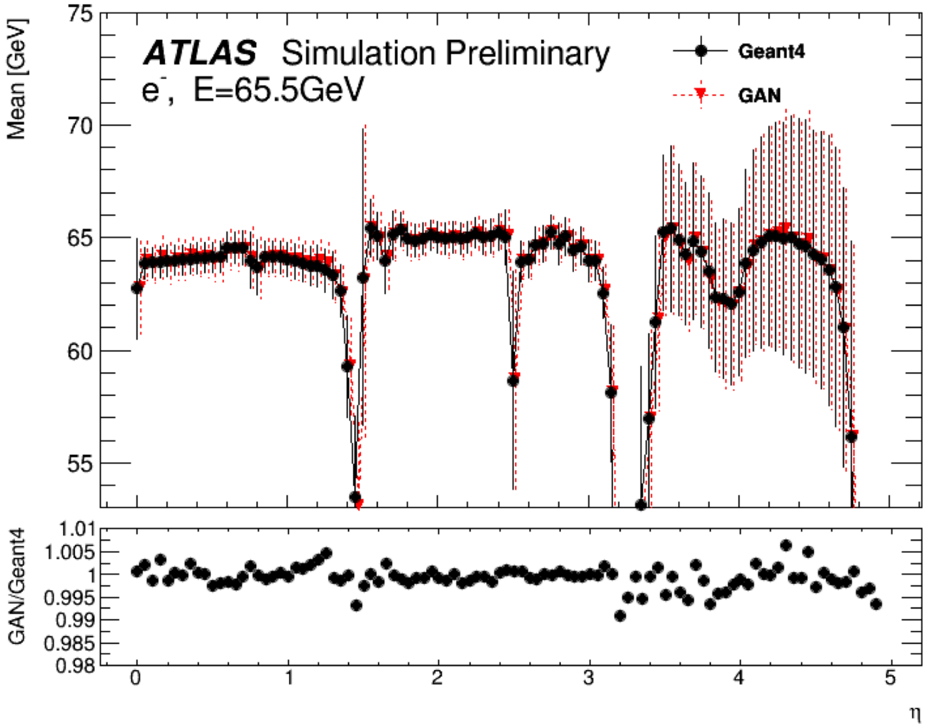


FastCaloGAN: e/gamma performance

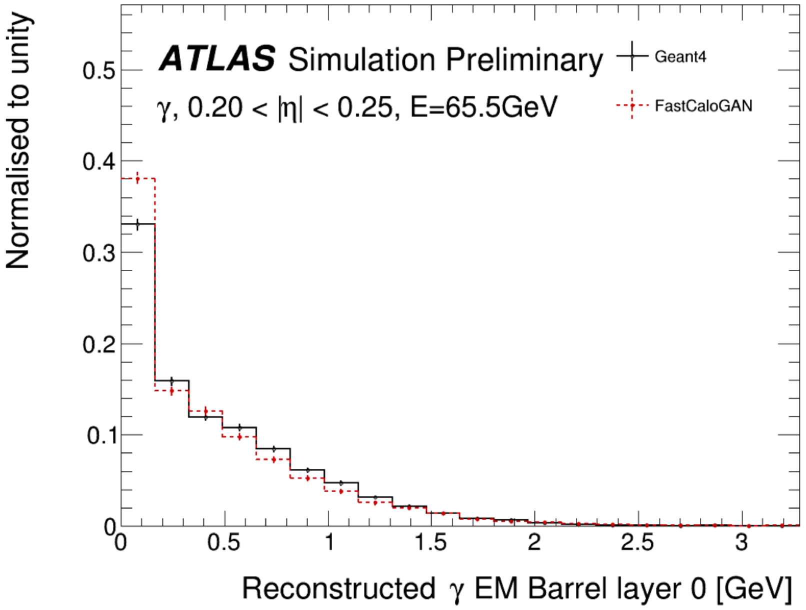


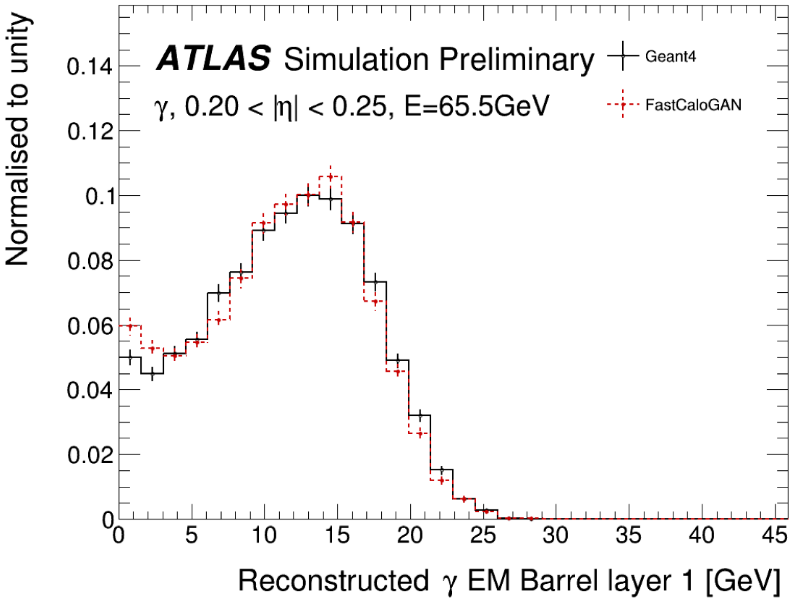


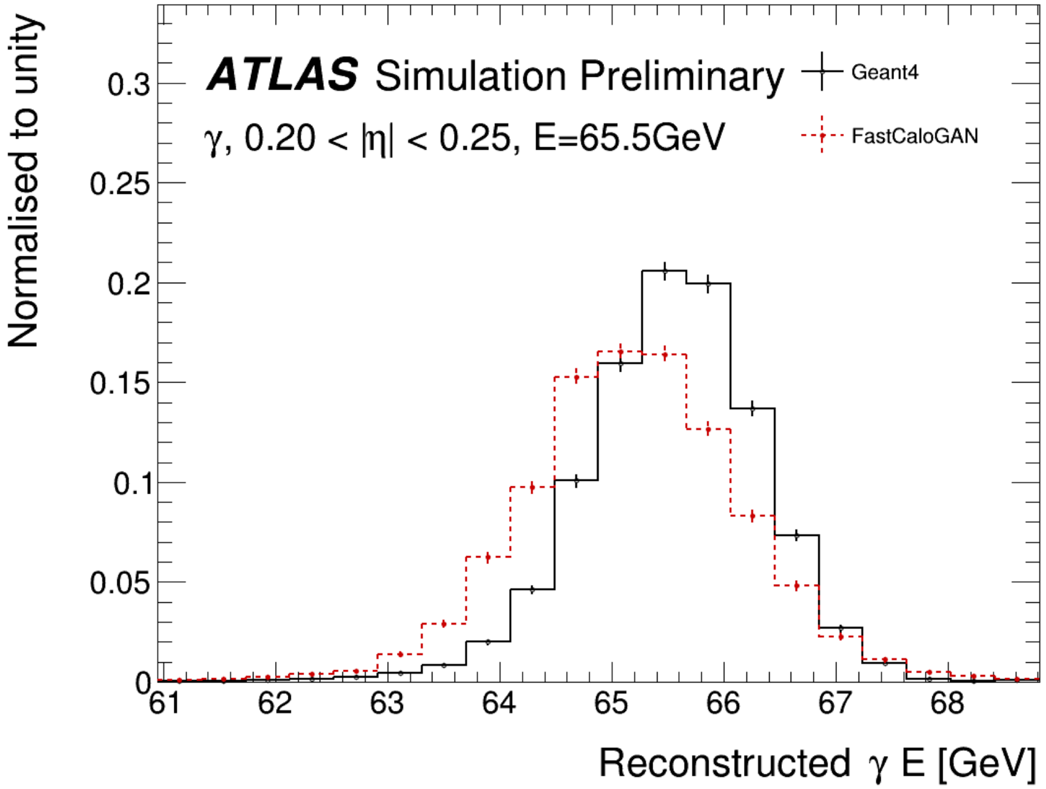


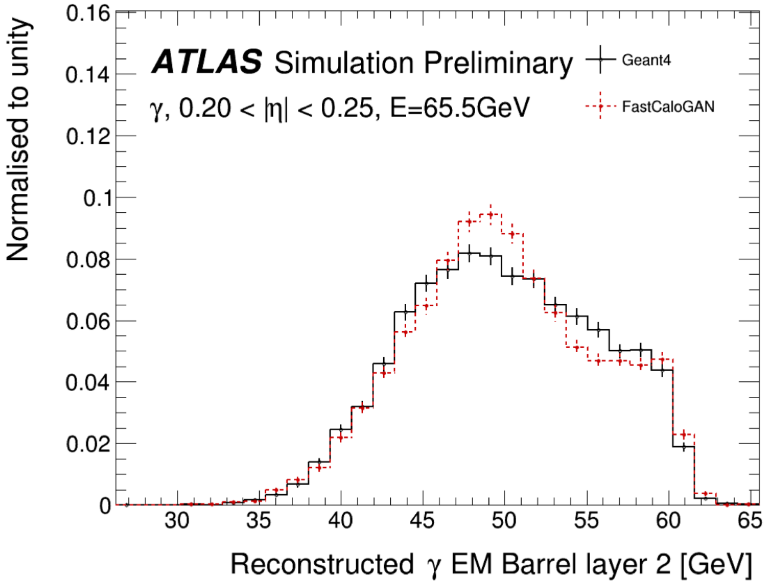


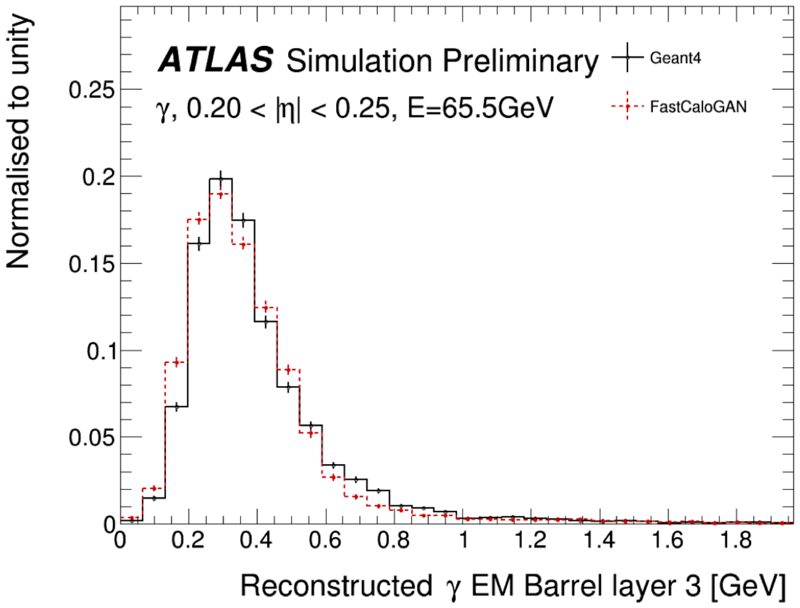
FastCaloGAN: e/gamma performance



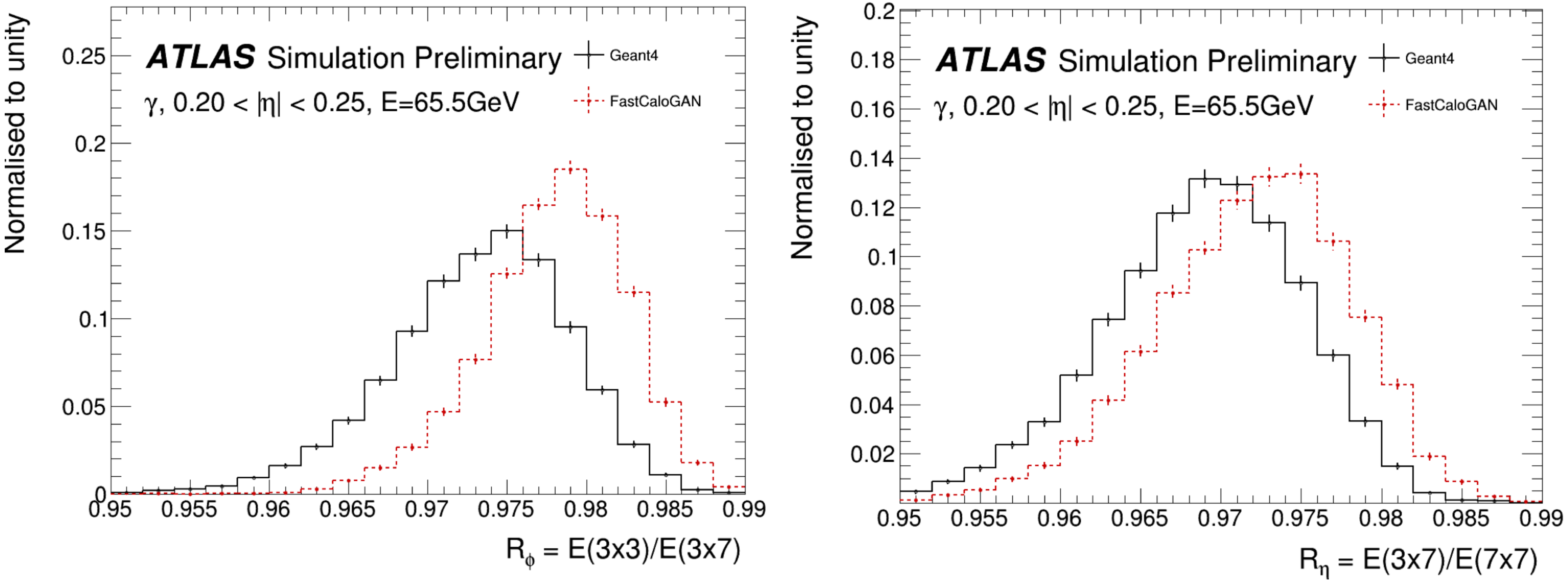




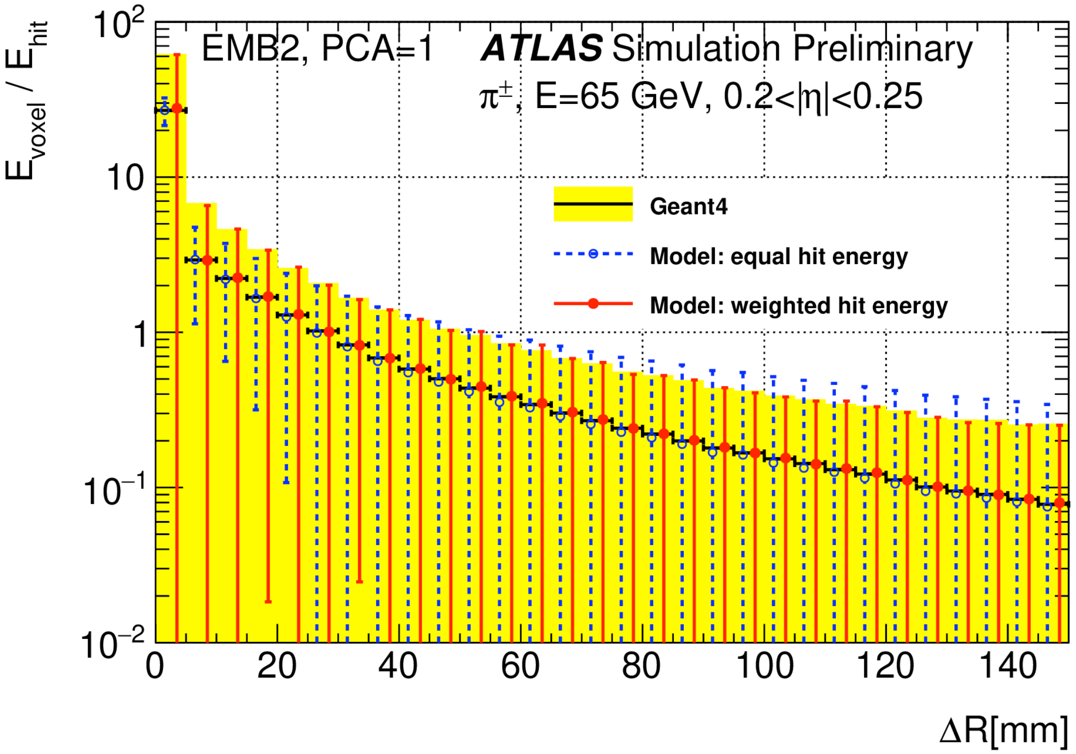




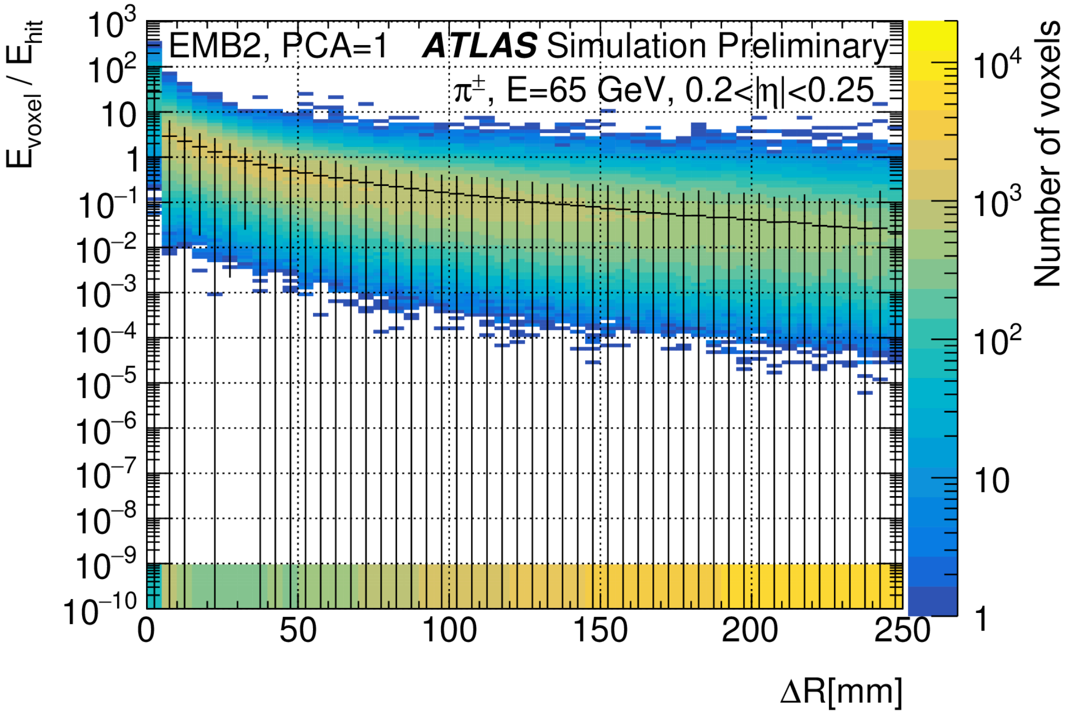
FastCaloGAN: e/gamma performance



Weighted Hit Simulation for Hadrons



AF3 (with weighted hits)

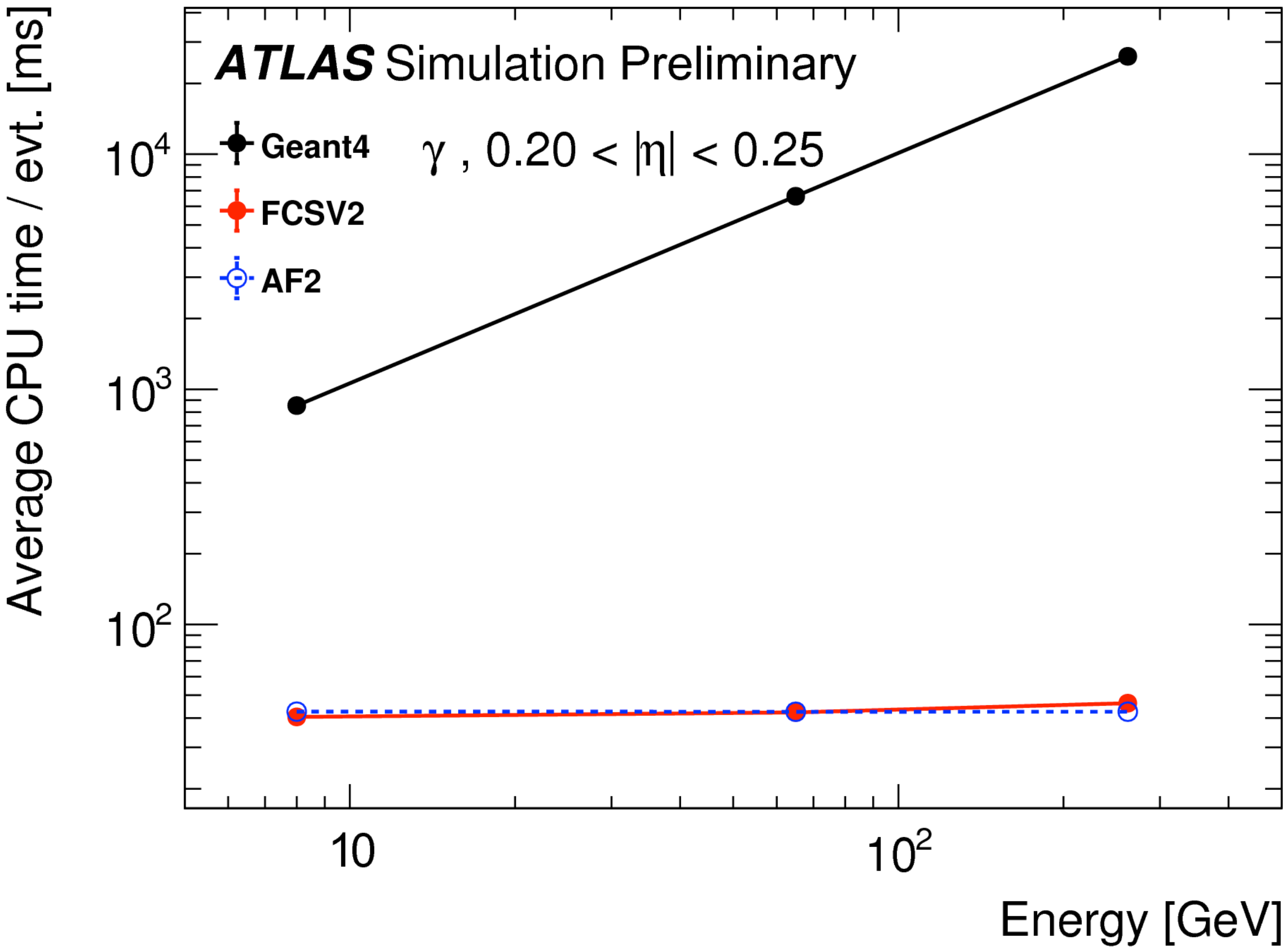


- Select smaller RMS of two poisson distributions reproducing:
  - ➡ Fraction of events with E = 0
  - ➡ RMS of Geant4 distribution
- Calculate weight as $w = \frac{\langle E_{\text{voxel}}/E_{\text{hit}}\rangle}{N_{\text{Poisson}}}$ with

  $N_{\text{Poisson}} = (\text{RMS}_{\text{Poisson}}/\lambda_{\text{Poisson}})^{-2}$ and re-calculate energy

  $E'_{\text{hit}} = E_{\text{hit}} \cdot w$
- If $\text{RMS}_{\text{Poisson}} < \text{RMS}_{\text{G4}}$: add additional fluctuations by smearing $\text{N}_{\text{Poisson}}$ (e^{s} as smearing function, where s is gaussian random number)

Simulation Time in AFII and FastCaloSimV2

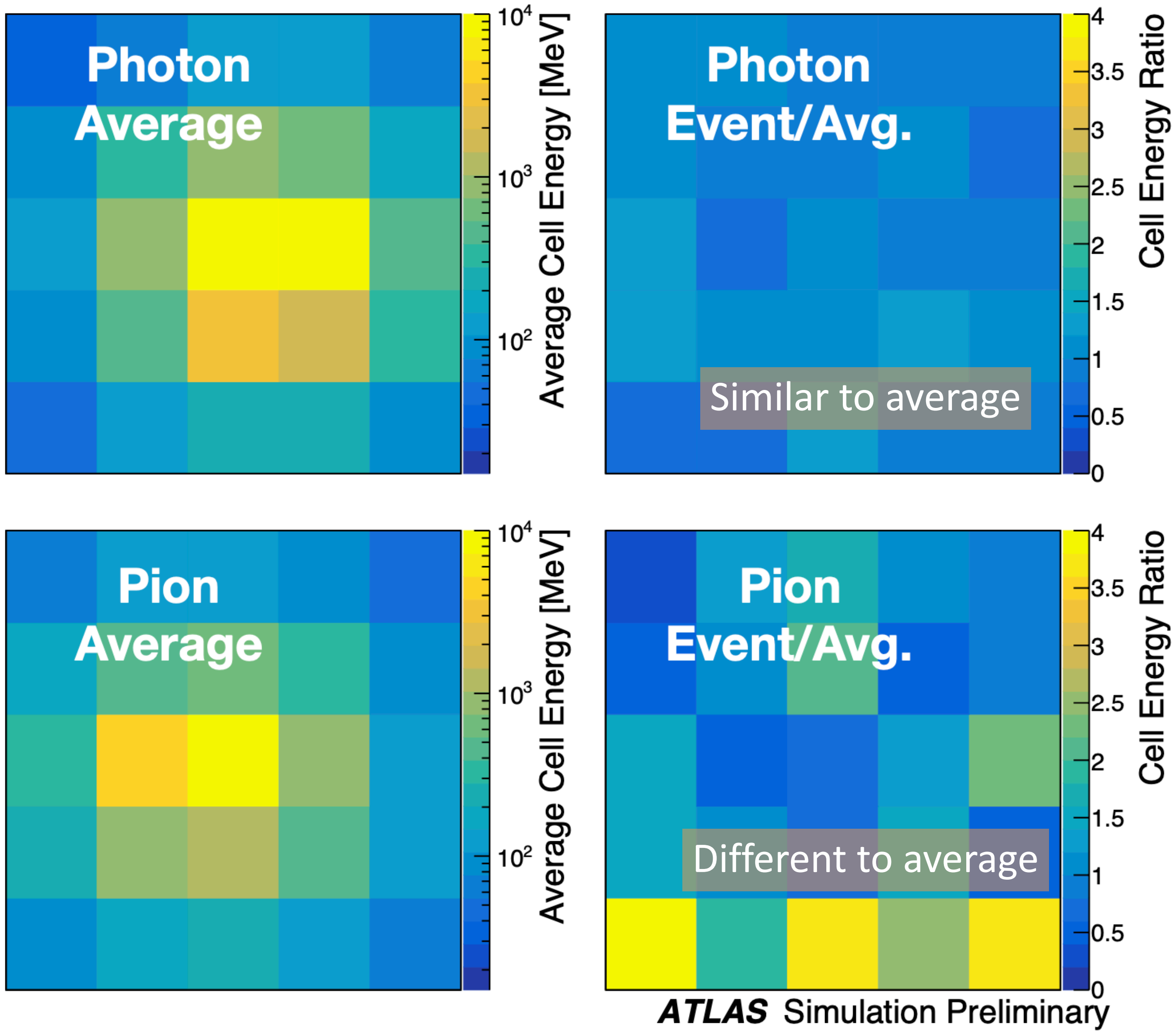


Particles are generated on calorimeter surface

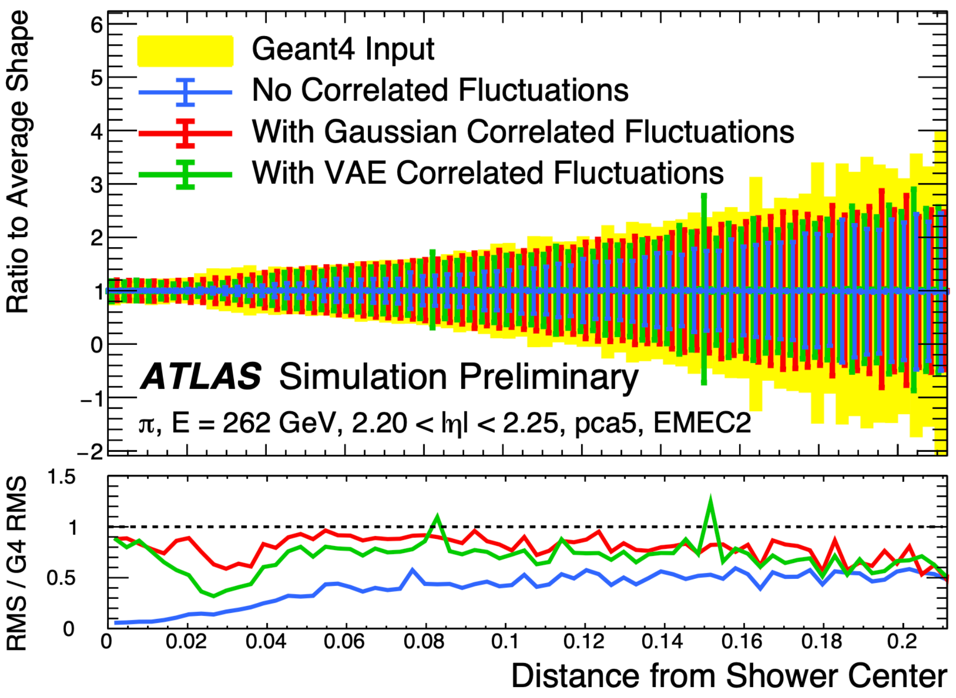
A factor of ~ 10 -25 times faster than Geant4

Work-in-Progress: Correlated Fluctuations

Random fluctuations do not include long-range correlations of energy fluctuations across cells:



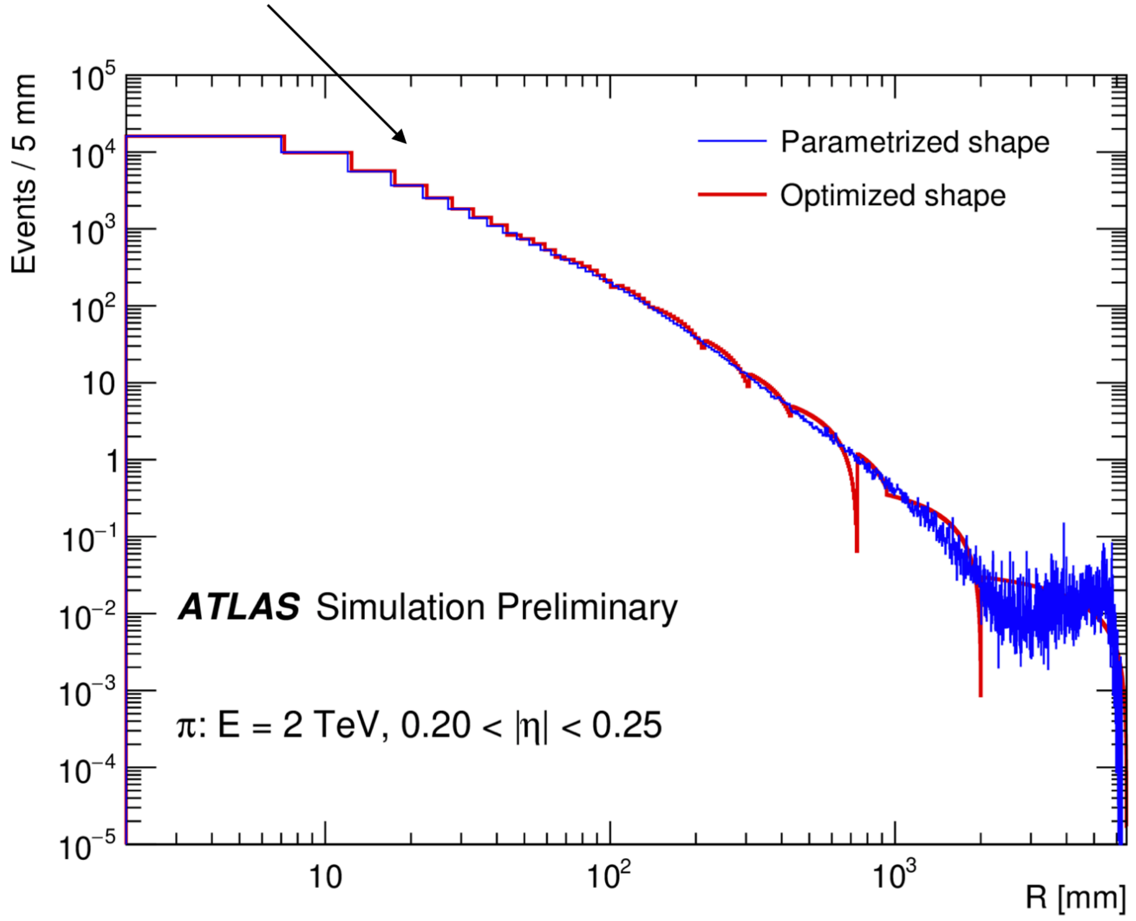
Energy ratios of single Geant4 events to their average shapes in a grid of 5x5 calorimeter cells for EMB2



- Developed model of correlated fluctuations with multivariate gaussians as well as Variational Auto Encoders (VAE)
- Proof of concept promising
- Expected to improve shape modelling

Memory optimisation

shower shape in radial direction



Parametrized shape:
- 1200+ bins of 5mm
- 43 Kbytes in memory

Optimized shape:
- 15 non-uniform bins in the same range with interpolation inside each bin
- 16-bit for bin edges
- 32-bit for bin content
- 353 bytes in memory!

- Current parametrisation file (~1.4Gb) occupies large amount of memory
- Need efficient memory optimisation for bulk production in Run 3
- Idea is to replace high memory consuming ROOT::TH2 classes with efficient custom classes
- Developed evolutionary algorithm to
  1. Perform optimised re-binning
  2. Perform optimised interpolation between bins
  3. Represent bin content/edges in reduced (e.g. 8-,16-bit) formats.
- Remains to be implemented