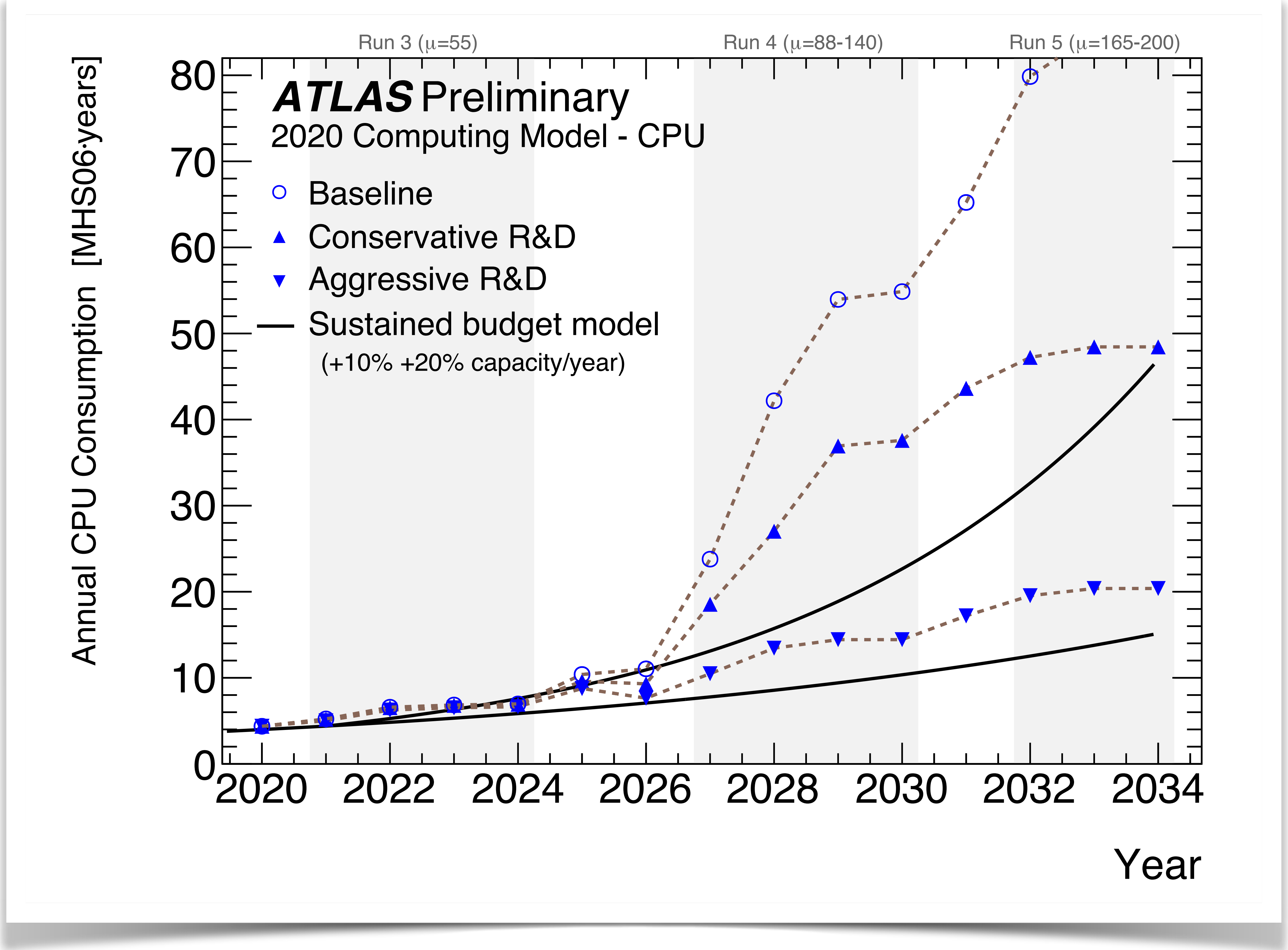
ATLFAST3: THE NEXT GENERATION OF FAST SIMULATION IN ATLAS

Hasib Ahmed (University of Edinburgh)
on behalf of ATLAS Collaboration

Challenges in Simulation: Run 3 & beyond

- Simulation of the ATLAS detector with Geant4 is CPU intensive.
- ~90% spent in shower simulation i.e. Calorimeter simulation.
- The CPU requirement will increase due to the increased luminosity and pileup in Run 3 & HL-LHC.
- In Run 3, > 50% of all events will be simulated with fast simulation increasing to > 75% in Run 4 to mitigate this.
- Beyond Run 3 fast Inner Detector (ID) simulation along with fast digitization and fast reconstruction will be required. [2]
- Utilize the inherent parallelism of fast calorimeter simulation with GPUs. [3]



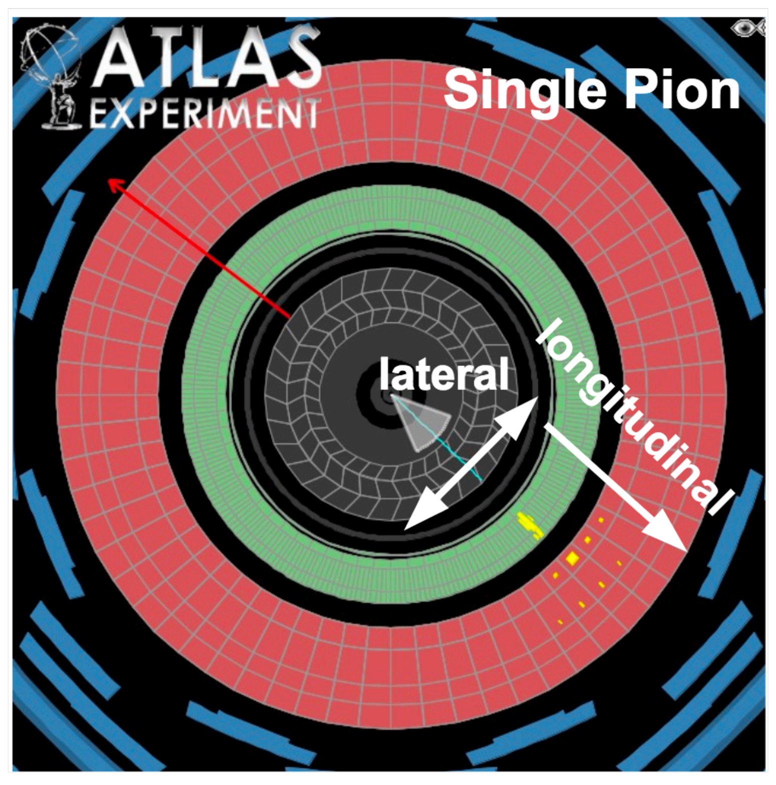
[1] HS06 benchmark

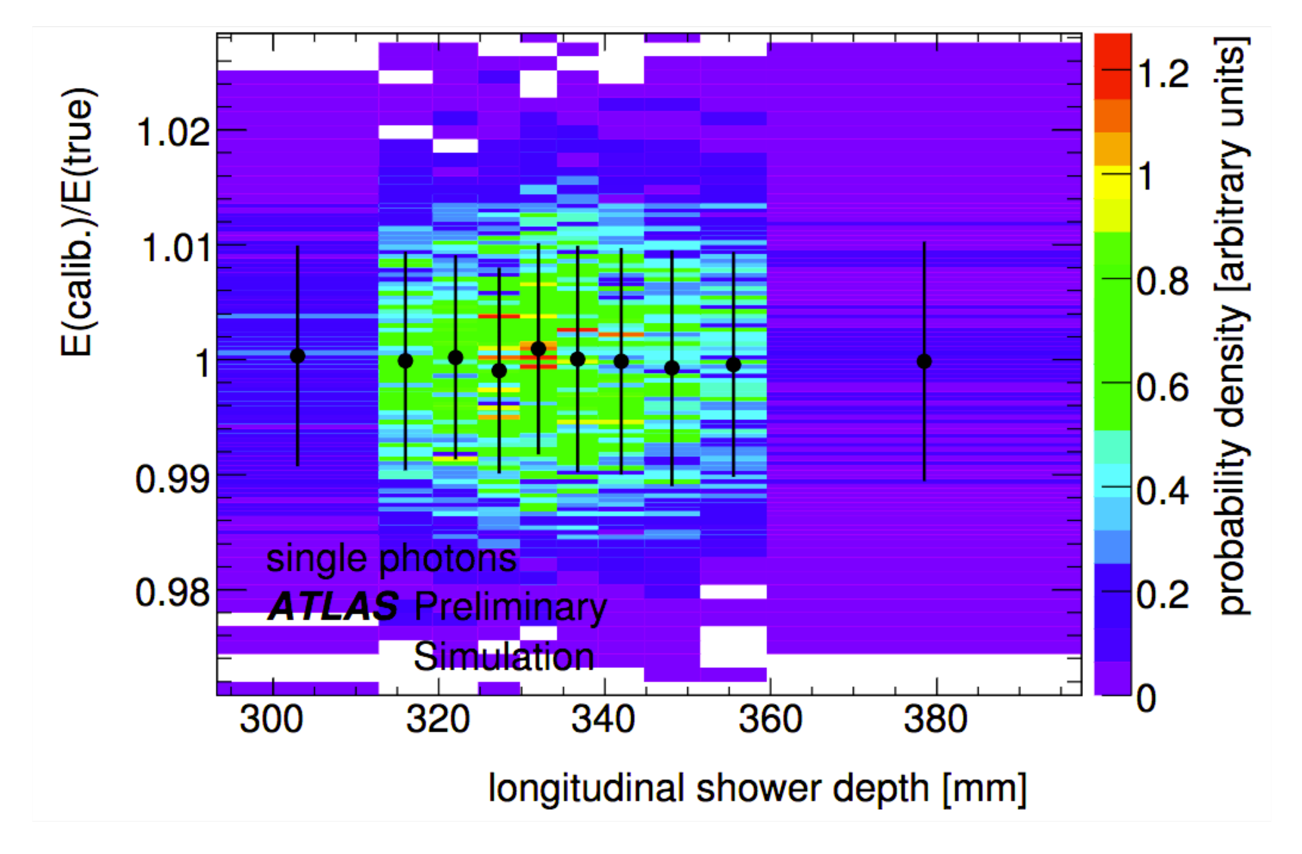
[2] See talk on Fast Simulation Chain

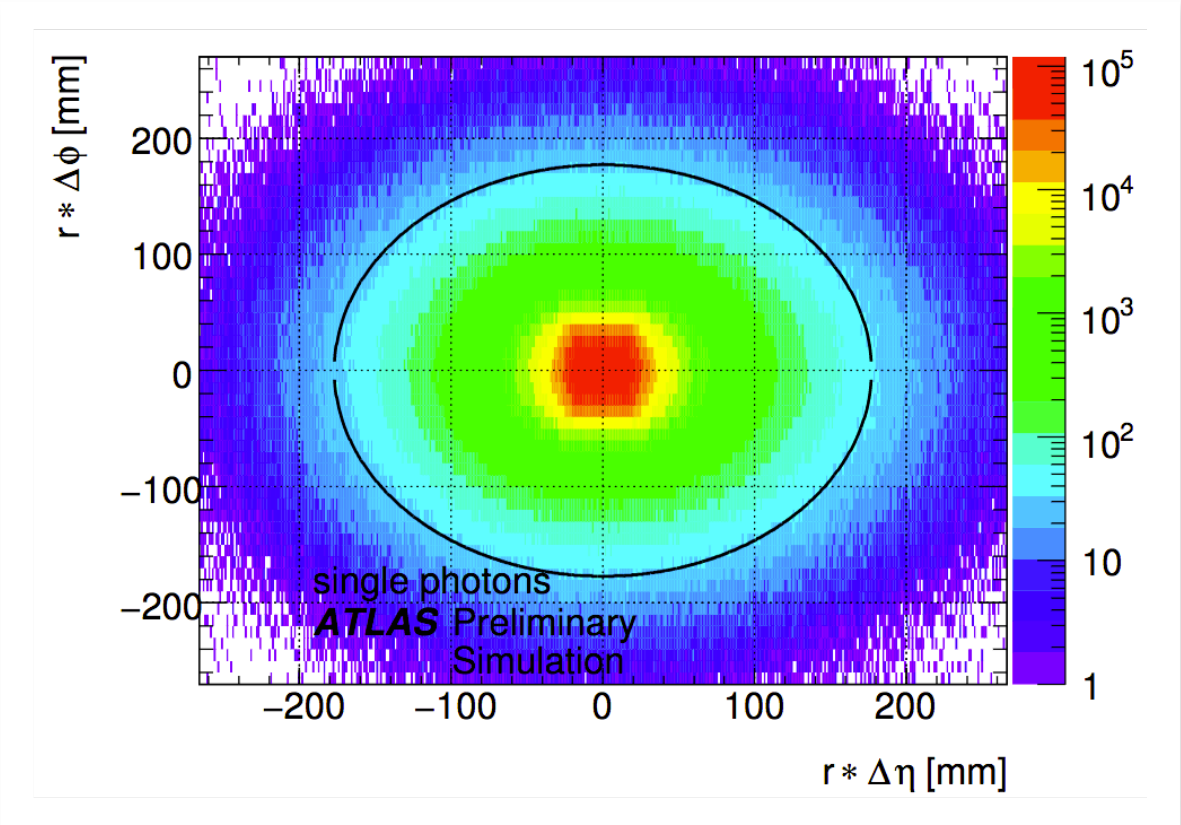
[3] See talk on Porting Parametrized Calorimeter Simulation to GPU

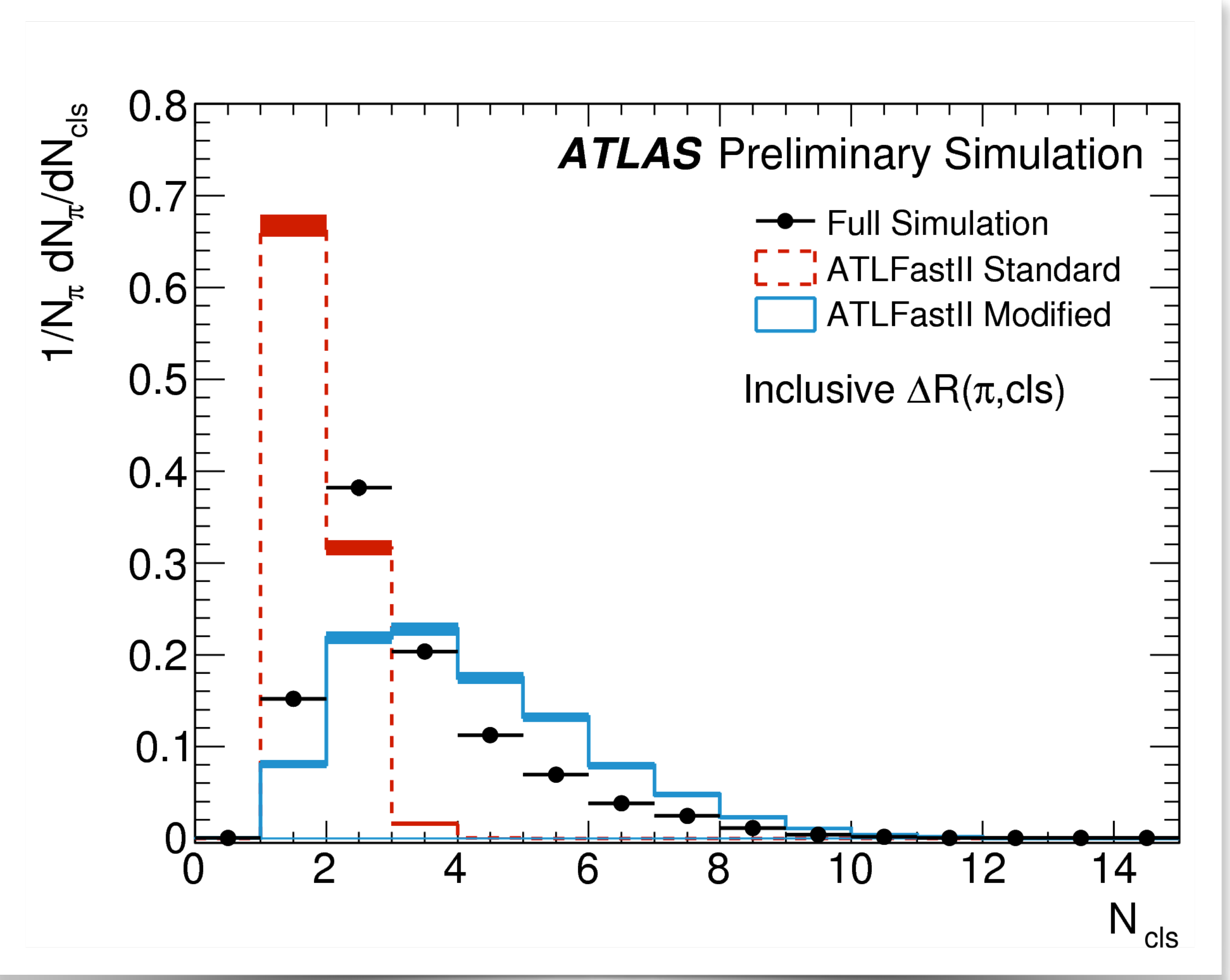
Fast simulation in Run 1 & Run 2: ATLFastII (AF2)

- AF2 - a parametrized calorimeter simulation is used in ATLAS during Run 1 and Run 2.
- e/γ and π is used for electromagnetic and hadronic shower parametrization respectively.
- Longitudinal shower: energy vs shower depth and correlation between layers
- Lateral shower: Average shower profile from a fitted radial symmetric function for each layer.
- Good average shower description but complex variables e.g. jet substructure is not well modeled.
- No lateral parametrization for Forward Calorimeter (FCal), particles escaping calorimeter volume (*punch through*)
- In Run 2, ~50% of all simulation were done in AF2.



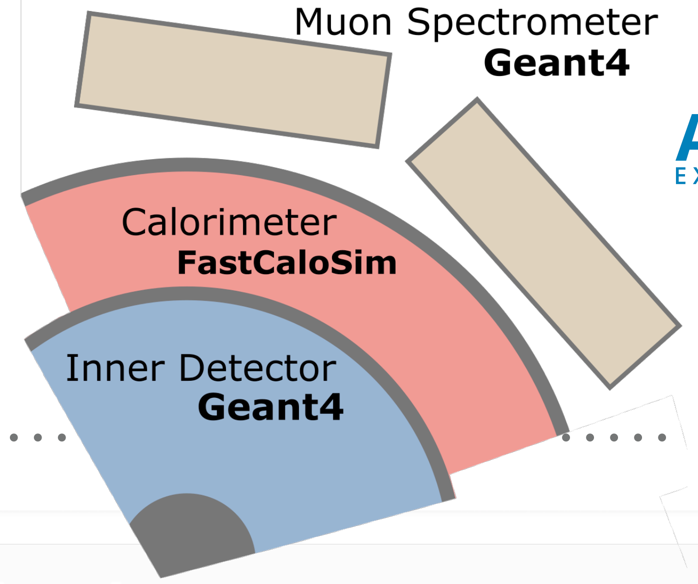




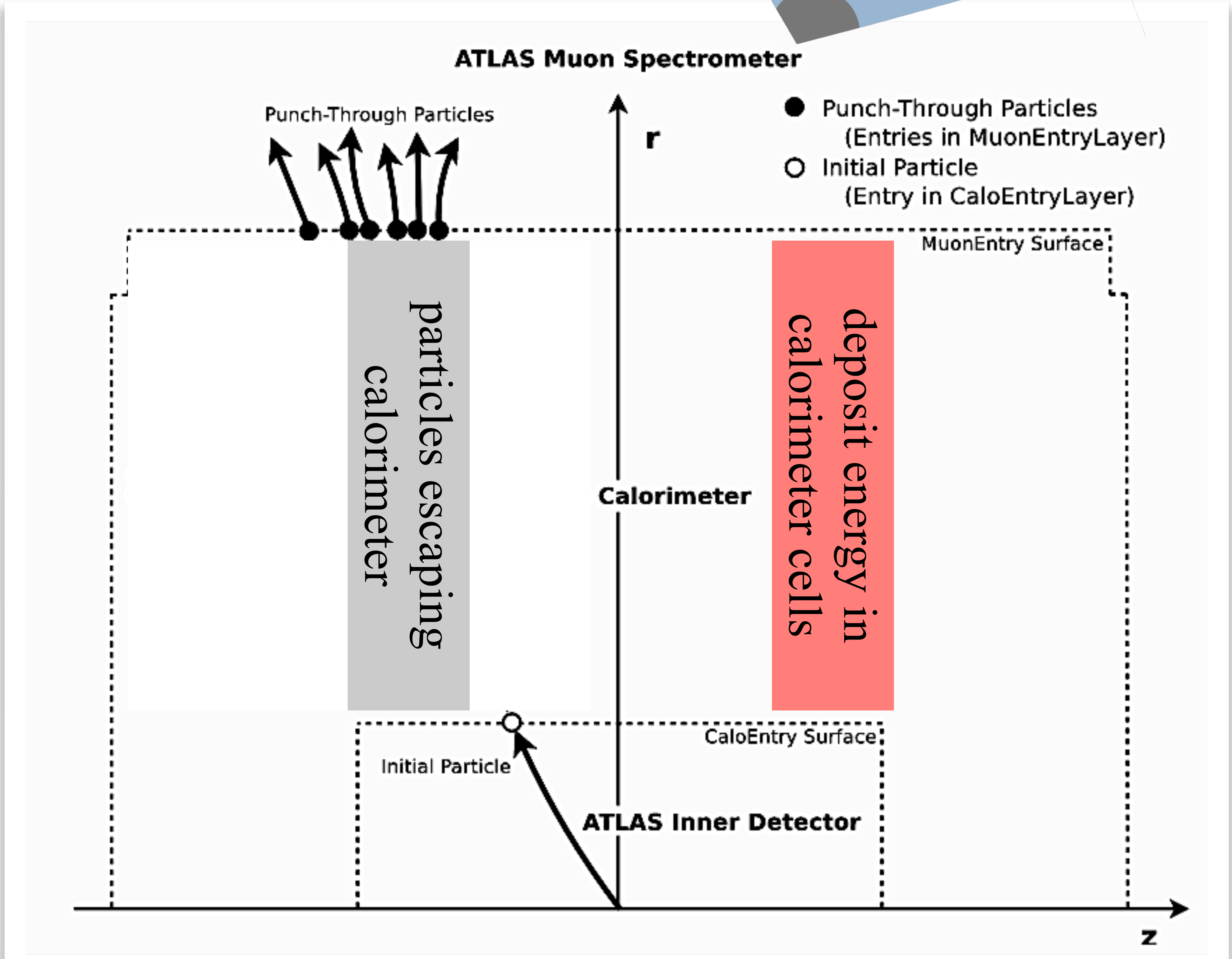


AF2 is tuned to data instead of Geant4 - requires a separate set of calibrations for reconstructed objects

Fast simulation in Run 3 & beyond: ATLFast 3 (AF3)



- AF3 improves physics performance significantly over AF2 and will meet the fast simulation needs of ATLAS for Run 3.
- AF3 uses two distinct approach of shower generation (includes FCal):
  - Parametrization based modeling - FastCaloSim V2 (FCSV2)
  - Generative Adversarial Network (GAN) based modeling - FastCaloGAN (FCSGAN)
- Dedicated parametrization for punch through particles - particles escaping calorimeter volume.



FastCaloSim V2	Used for electron, photon and low or high energy hadrons
FastCaloGAN	Used for medium energy hadrons
Punch Through	Used for simulating particles that exit the calorimeter and enters Muon Systems (MS)
Geant4	Used for simulating very low energy hadrons in the Calorimeter, all particles in the MS and ID.

- AF3 provides a speed gain by a factor of -
  - O(500) for calorimeter only simulation
  - O(10) for full detector simulation

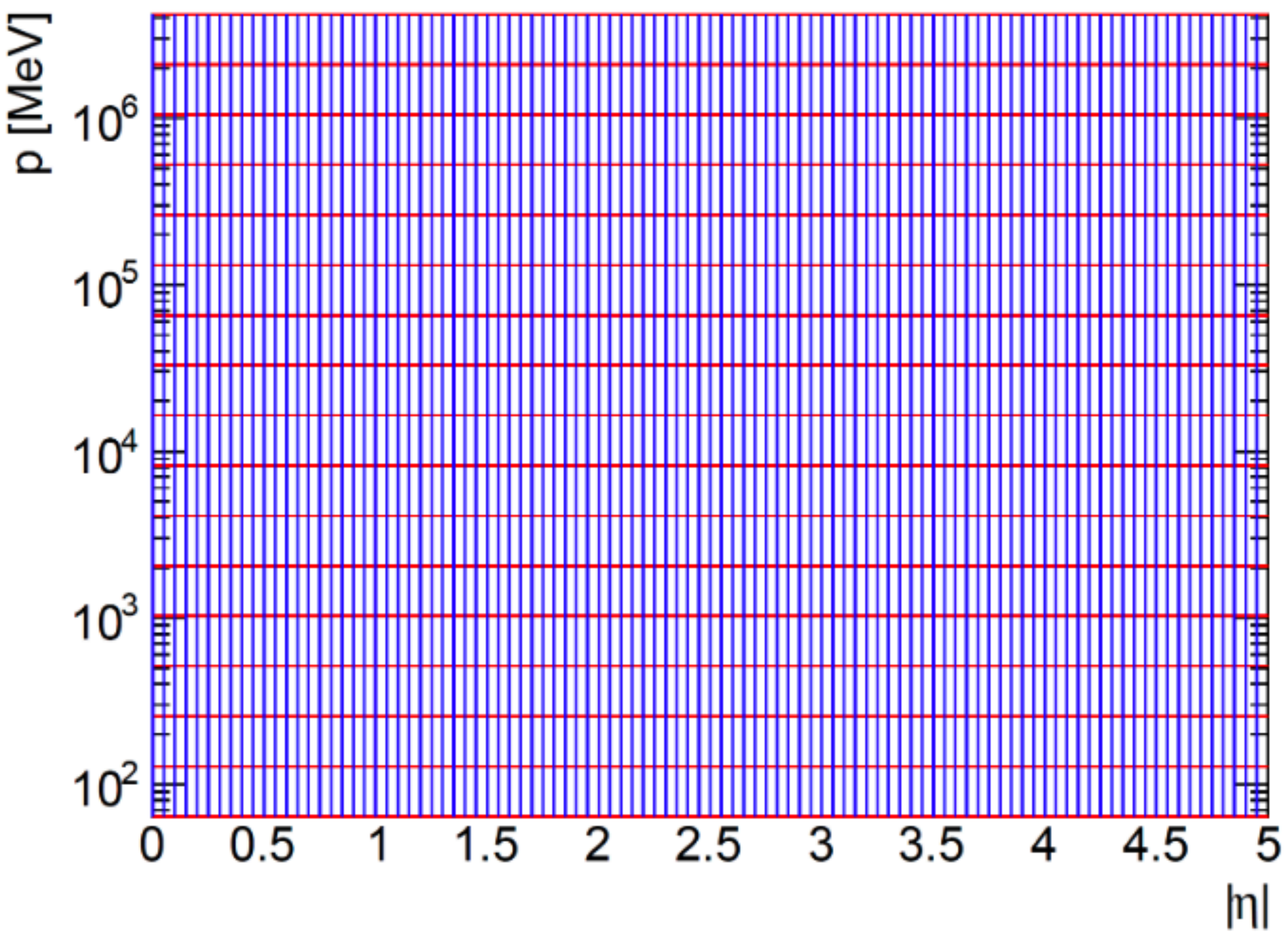
AF3 targets achieving identical modeling to Geant4 requiring only one set of calibrations

Input datasets for AF3 modeling

Geant4 simulated single particles generated at the calorimeter surface is used for modeling AF3

17 discrete points:

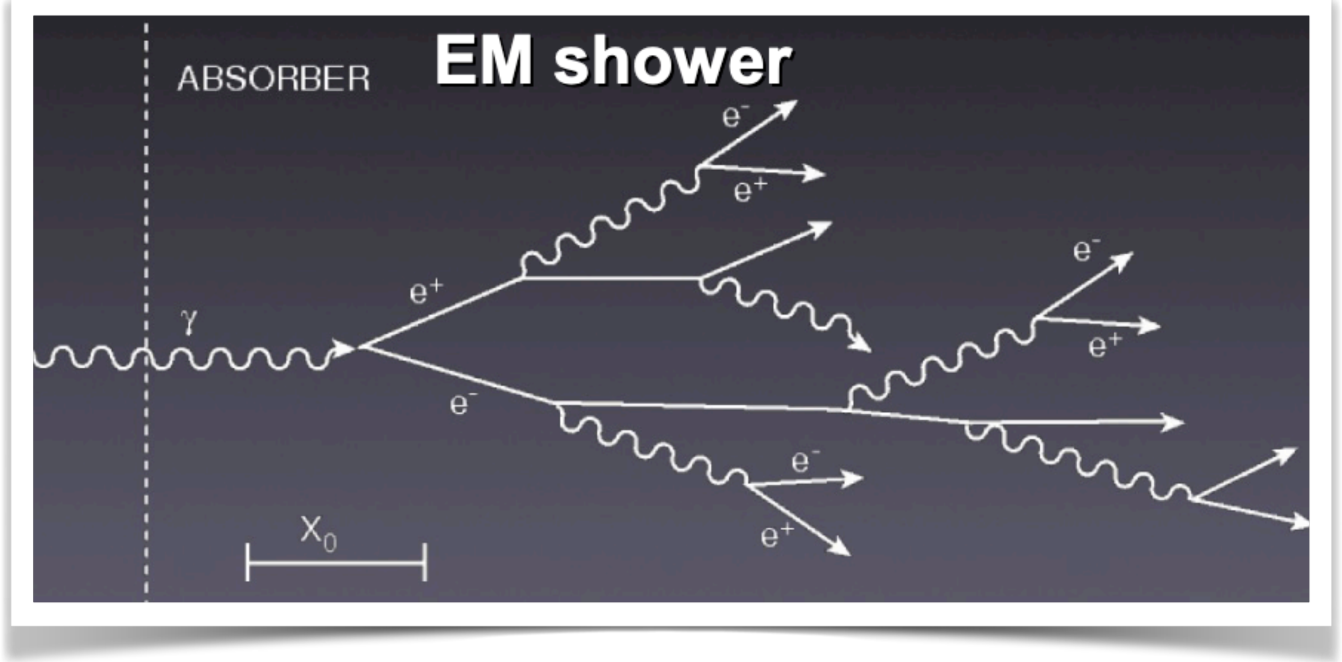
64MeV - 4.2 TeV (log spacing)



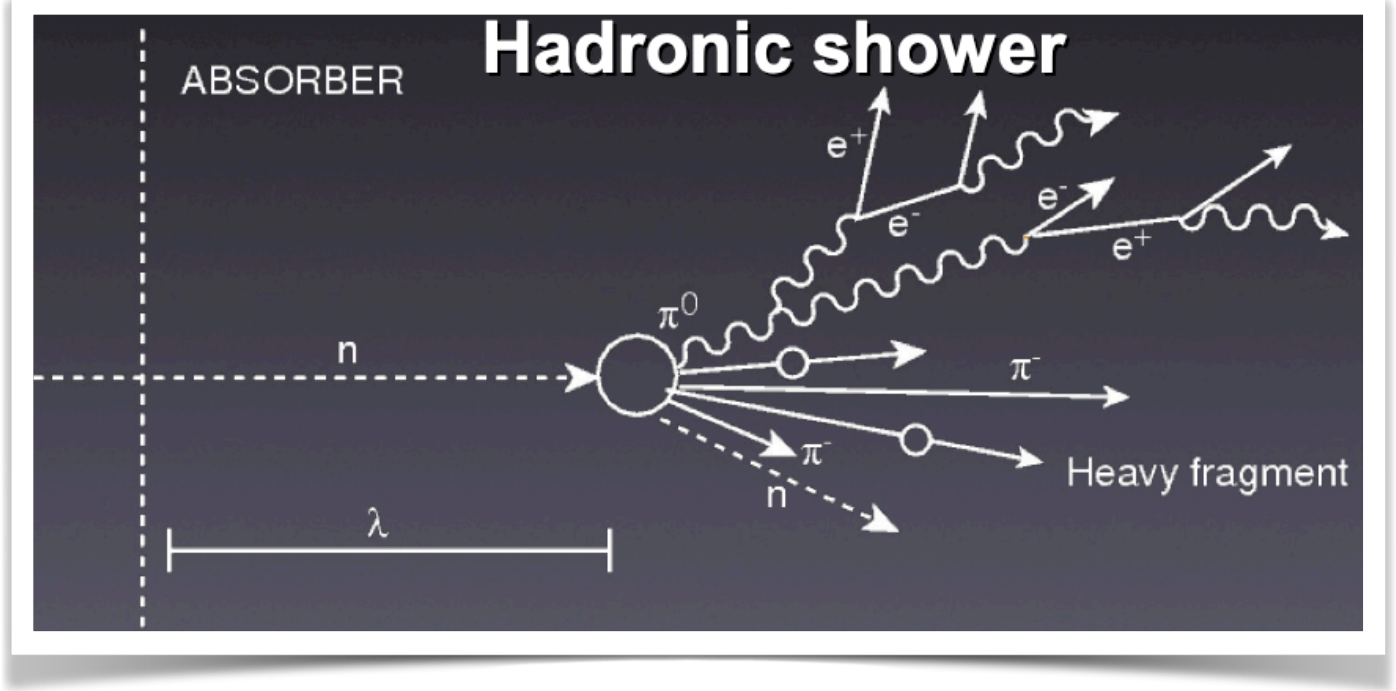
100 bins of size 0.05 covering $0 \leqslant |\eta| \leqslant 5$

Photons: for photon shower

Electrons ($e^{\pm}$): for electron shower



Pions ($\pi^{\pm}$): for hadronic shower



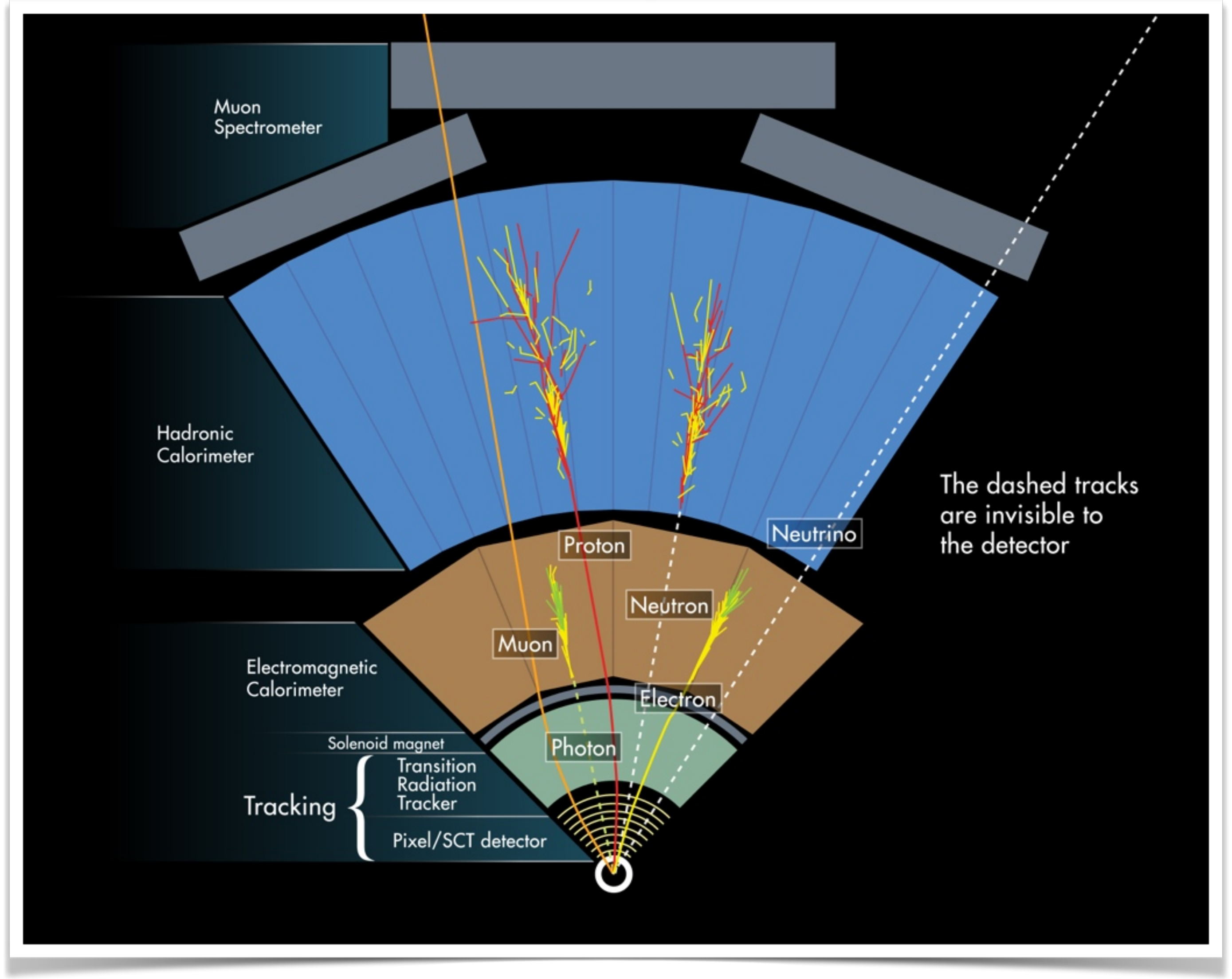
Detail G4 steps (granular energy deposit)

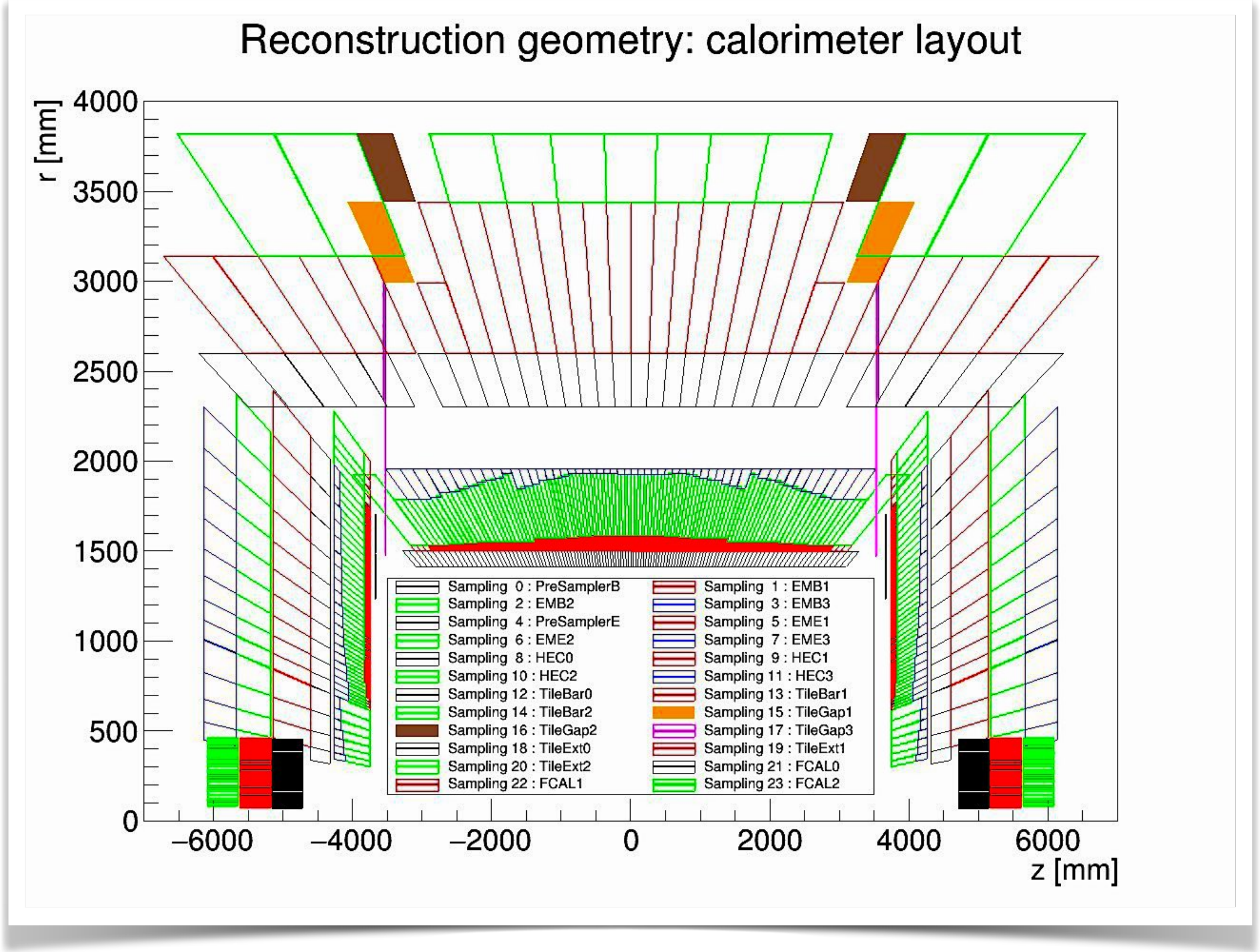
No primary vertex smearing in simulation

Noise, cross-talk between neighboring cells and bad cells turned off in digitization.

Parametrization based - FastCaloSim V2

- Instead of simulating particle interactions, directly parametrize the detector response of single particles entering the calorimeter system.
- Parametrize the single particle shower development in longitudinal (energy) and lateral (shape) directions.
- Use the parametrization at simulation step to deposit energy in calorimeter cells using simplified geometry.

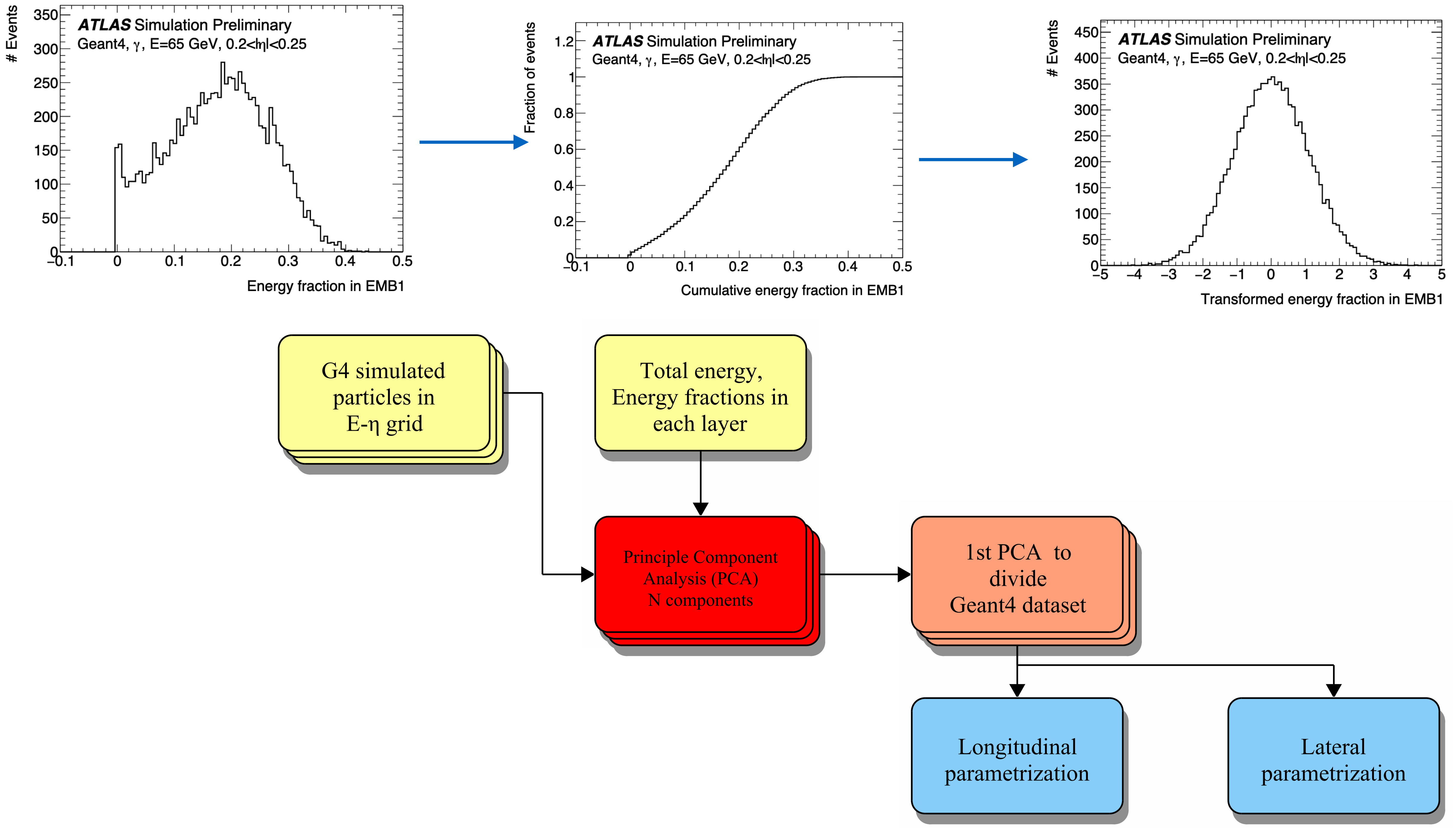


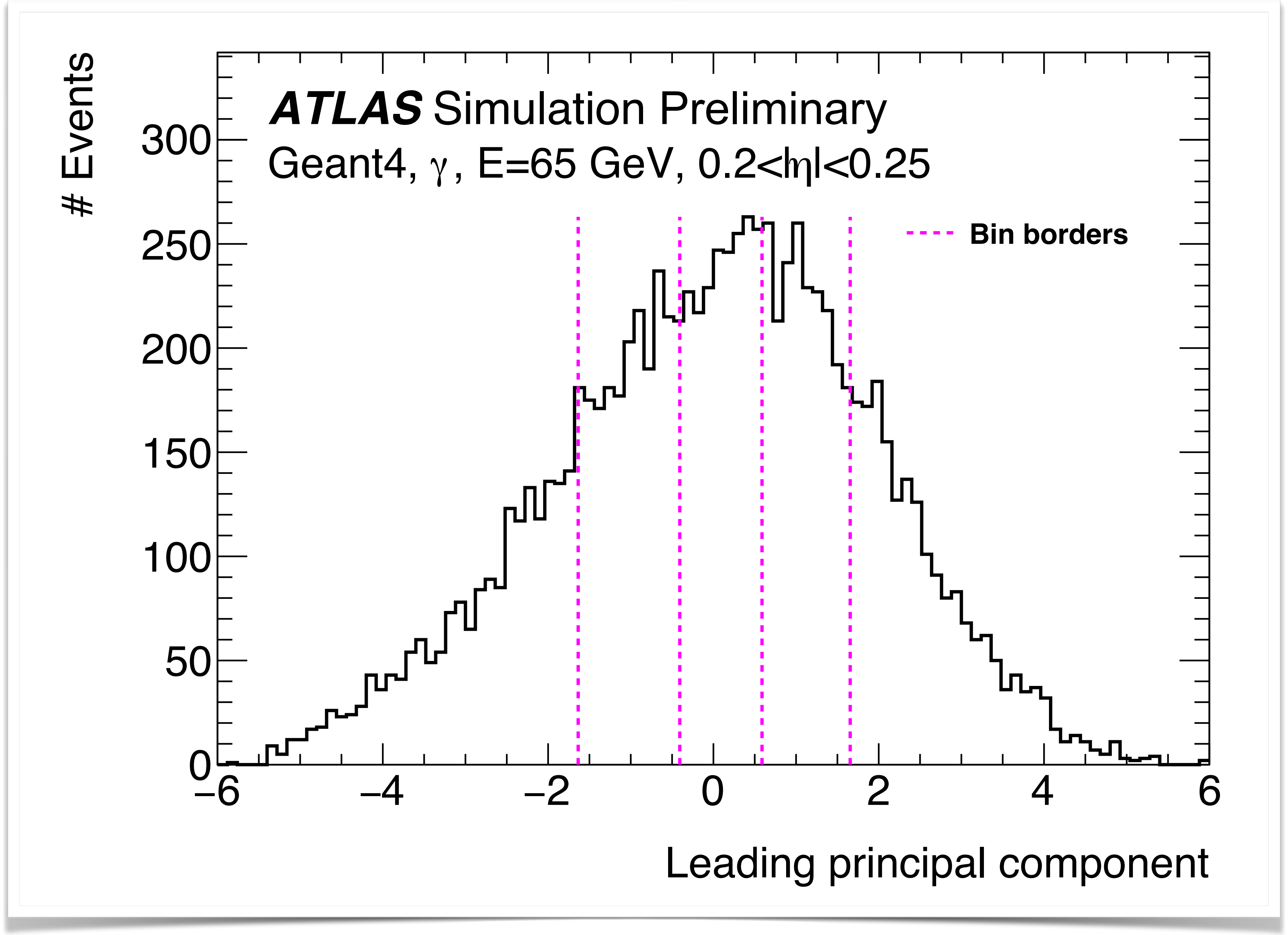


- The energy in each sampling layer is highly correlated.
- Classify showers based on the depth on the interaction point (i.e. depth at where a particle initiates the shower)
- The longitudinal and lateral parametrization is done for each for the shower type, for each calorimeter layers.

Shower Classification

- To remove the energy correlation between layers - single particles are classified based on its depth on interaction point.
- The energy fraction of each layer, total energy for all particles are used to perform a Principal Component Analysis (PCA).

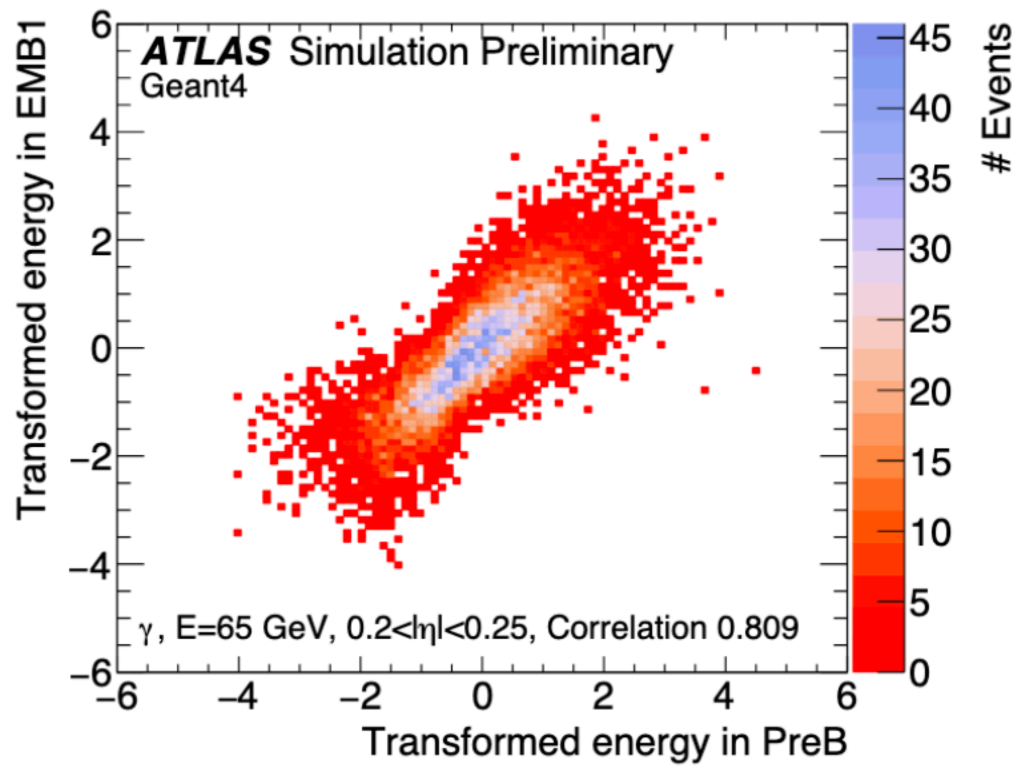




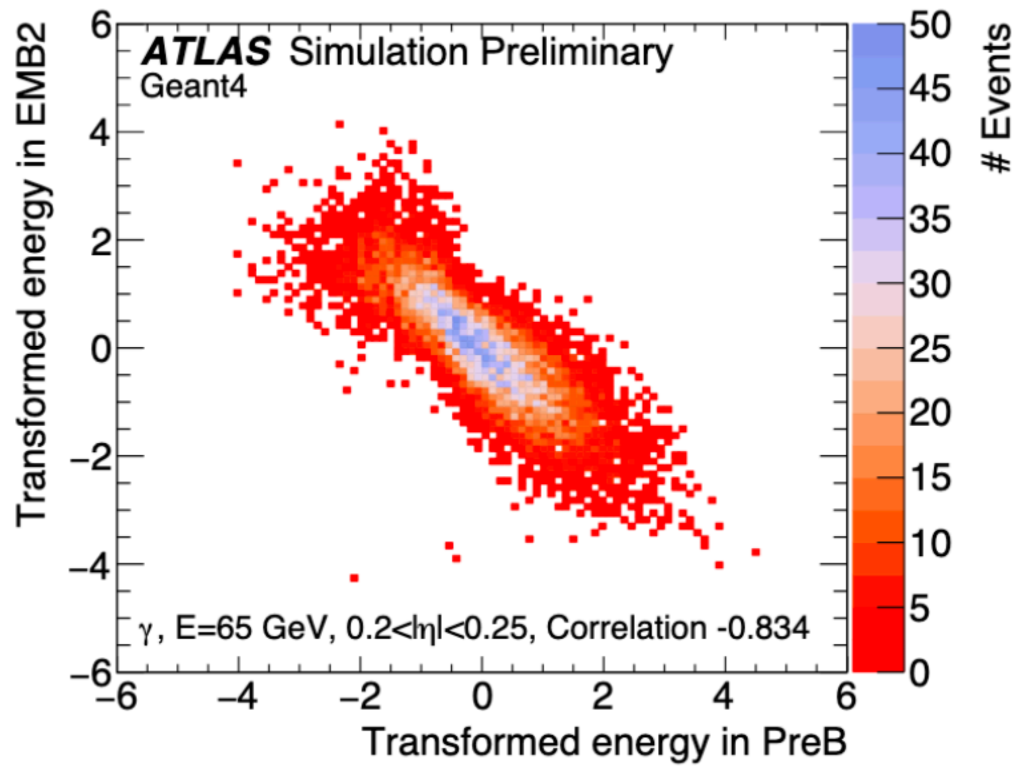
Leading principal component is used to divide the particles in quantiles

Validation of shower classification - energy decorrelation

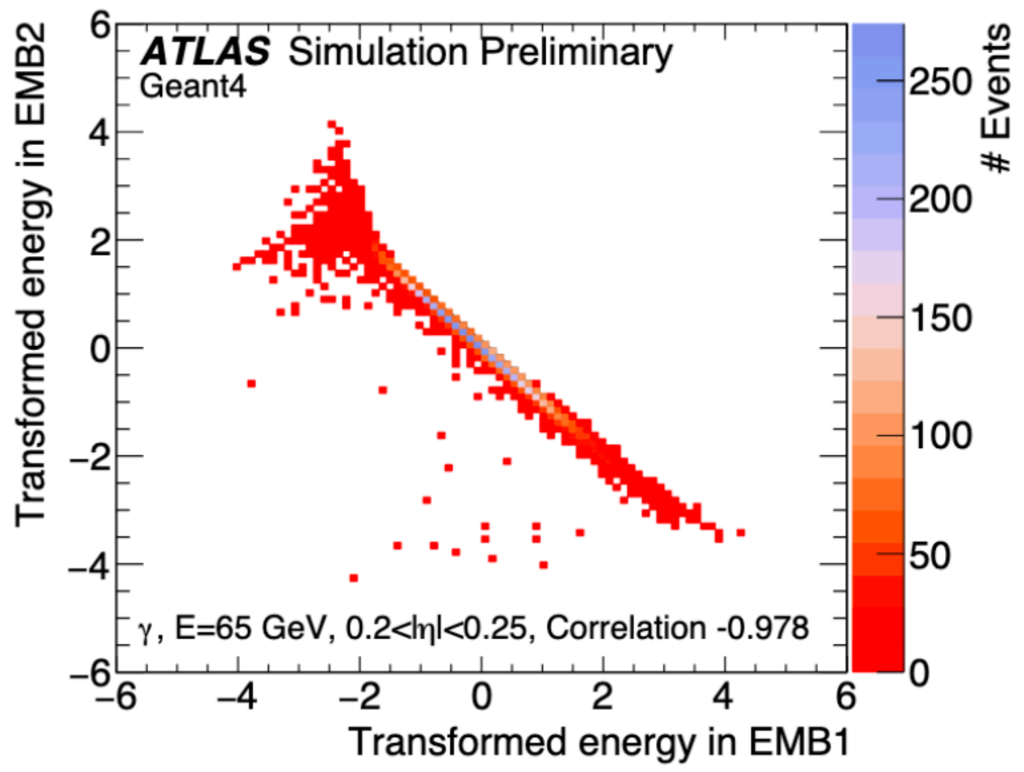
Before PCA:



(a) Presampler vs EM Barrel 1

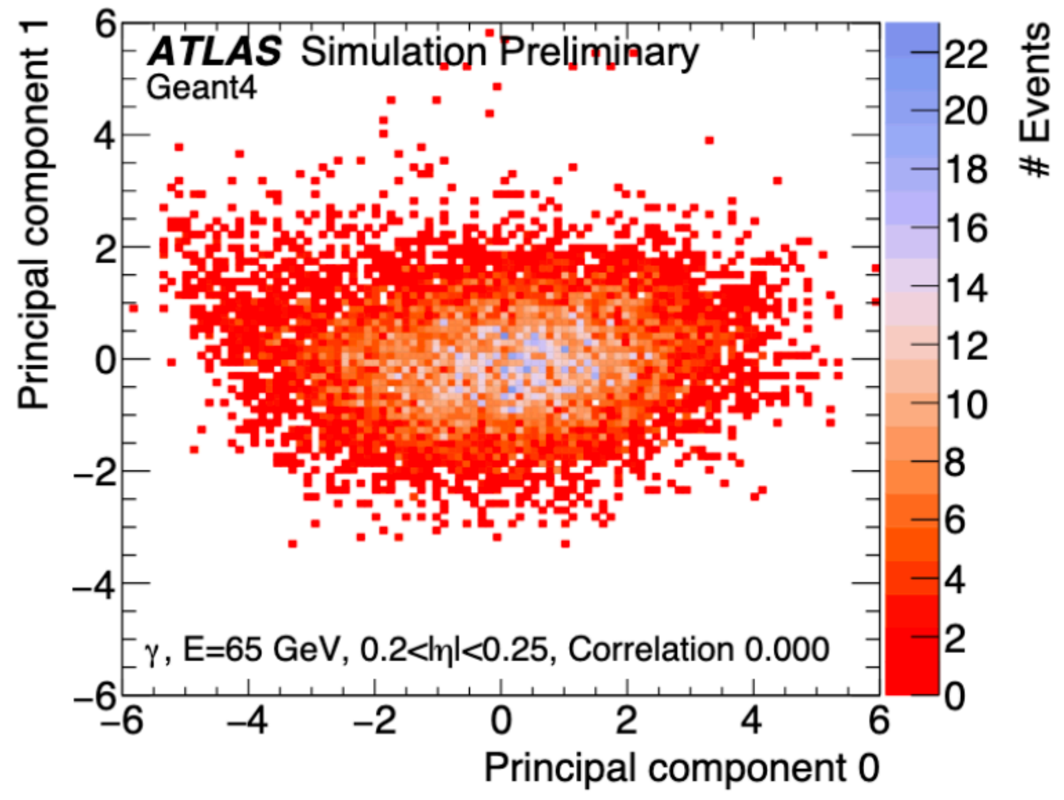


(b) Presampler vs EM Barrel 2

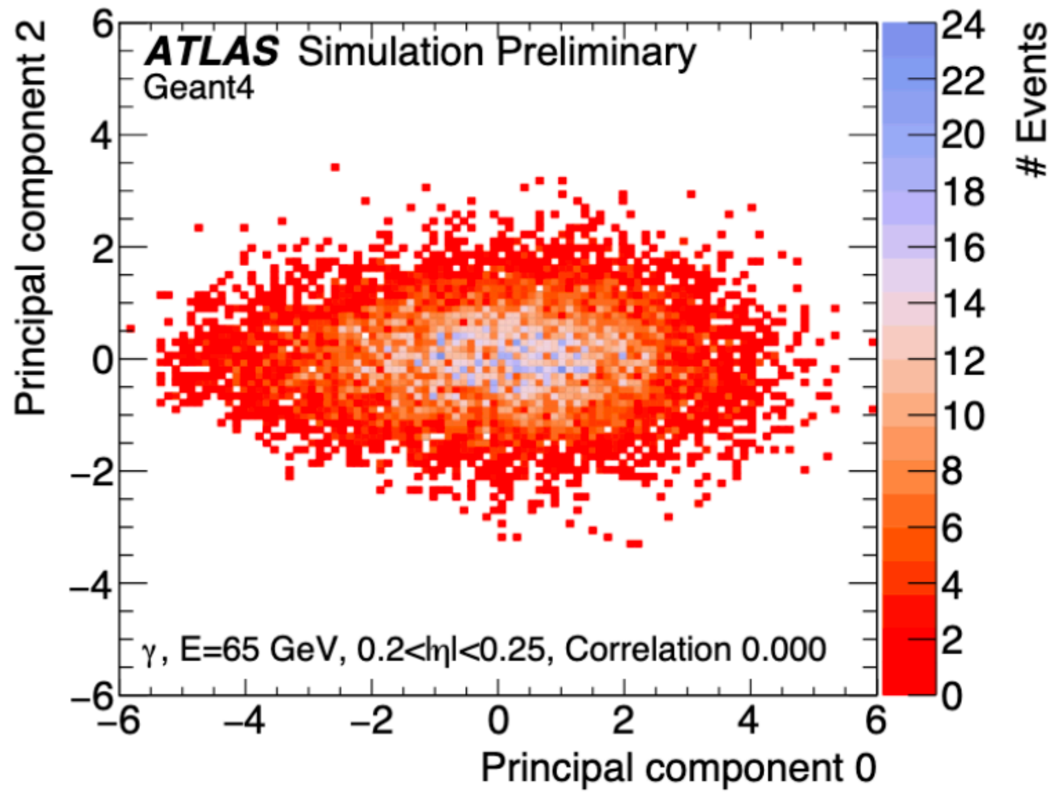


(c) EM Barrel 1 vs EM Barrel 2

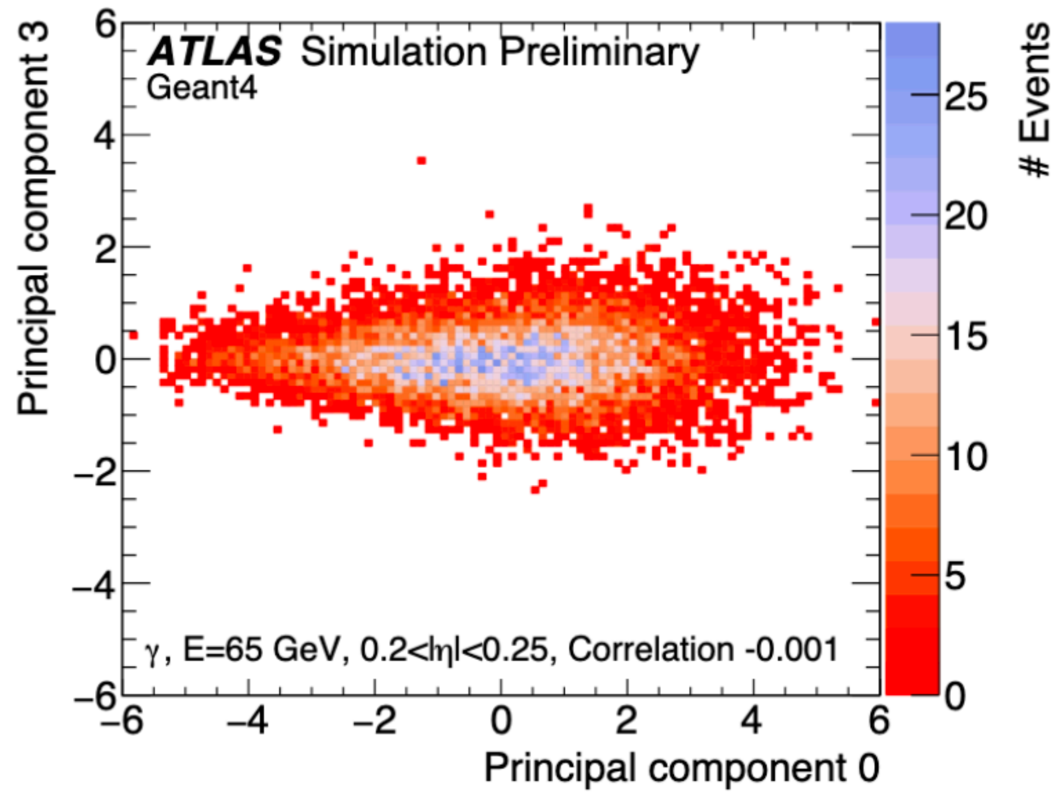
After PCA:



(a) First vs second PC



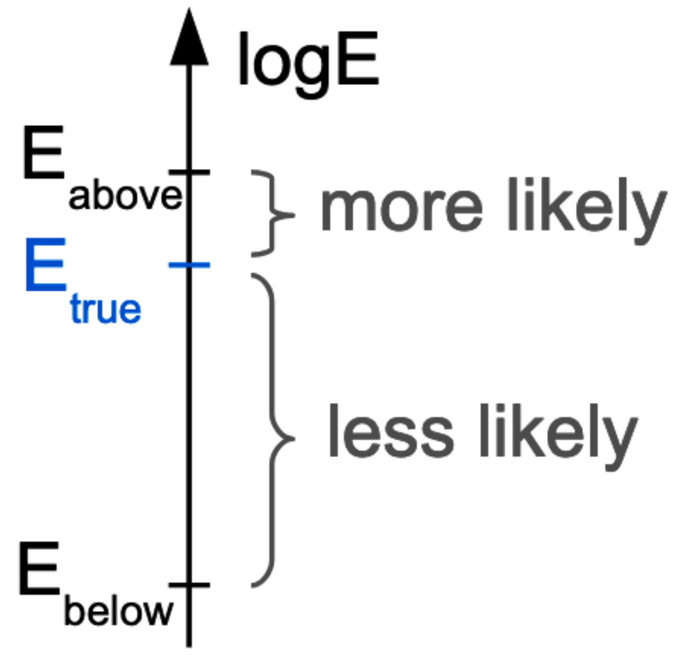
(b) First vs third PC

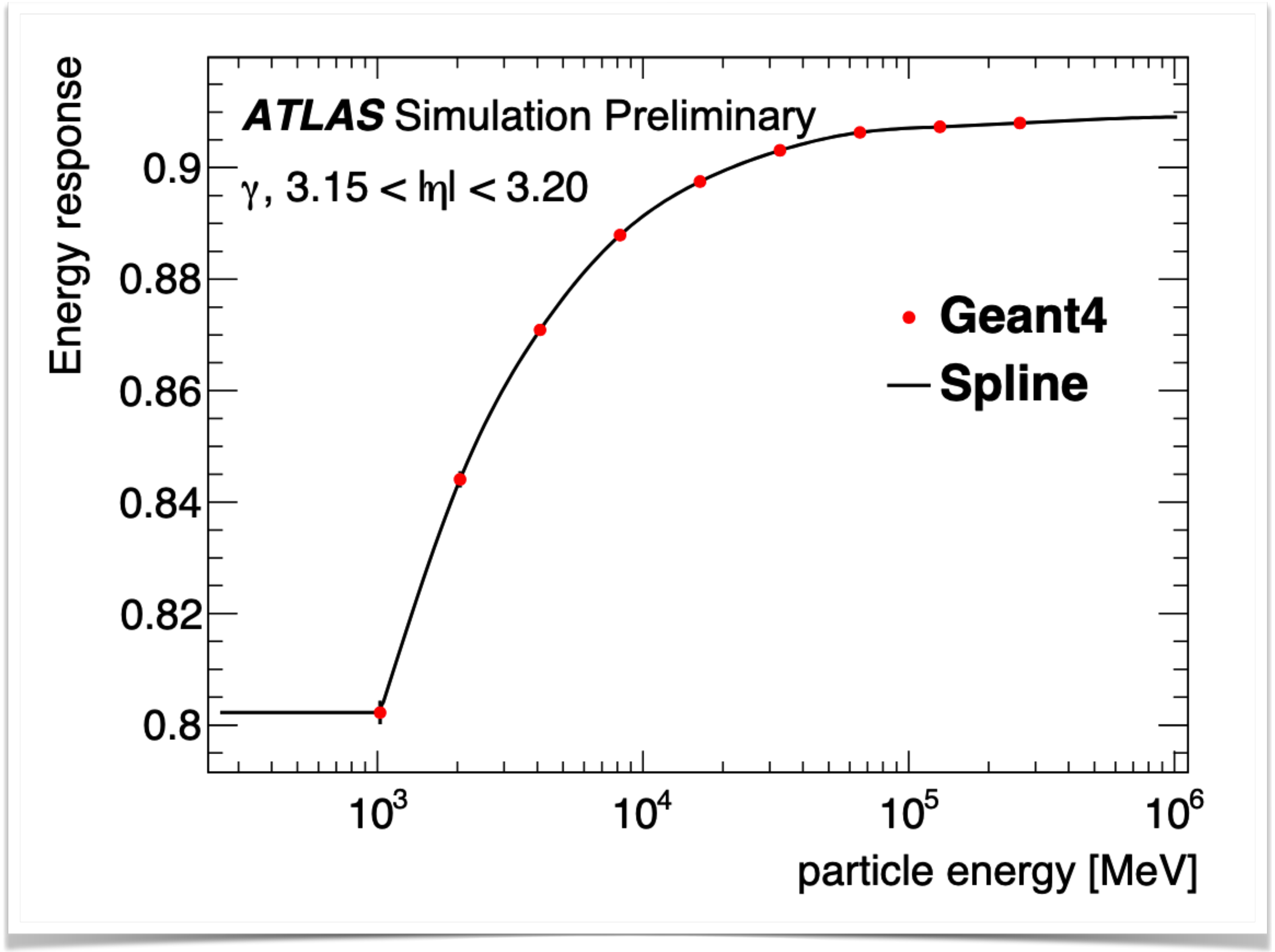


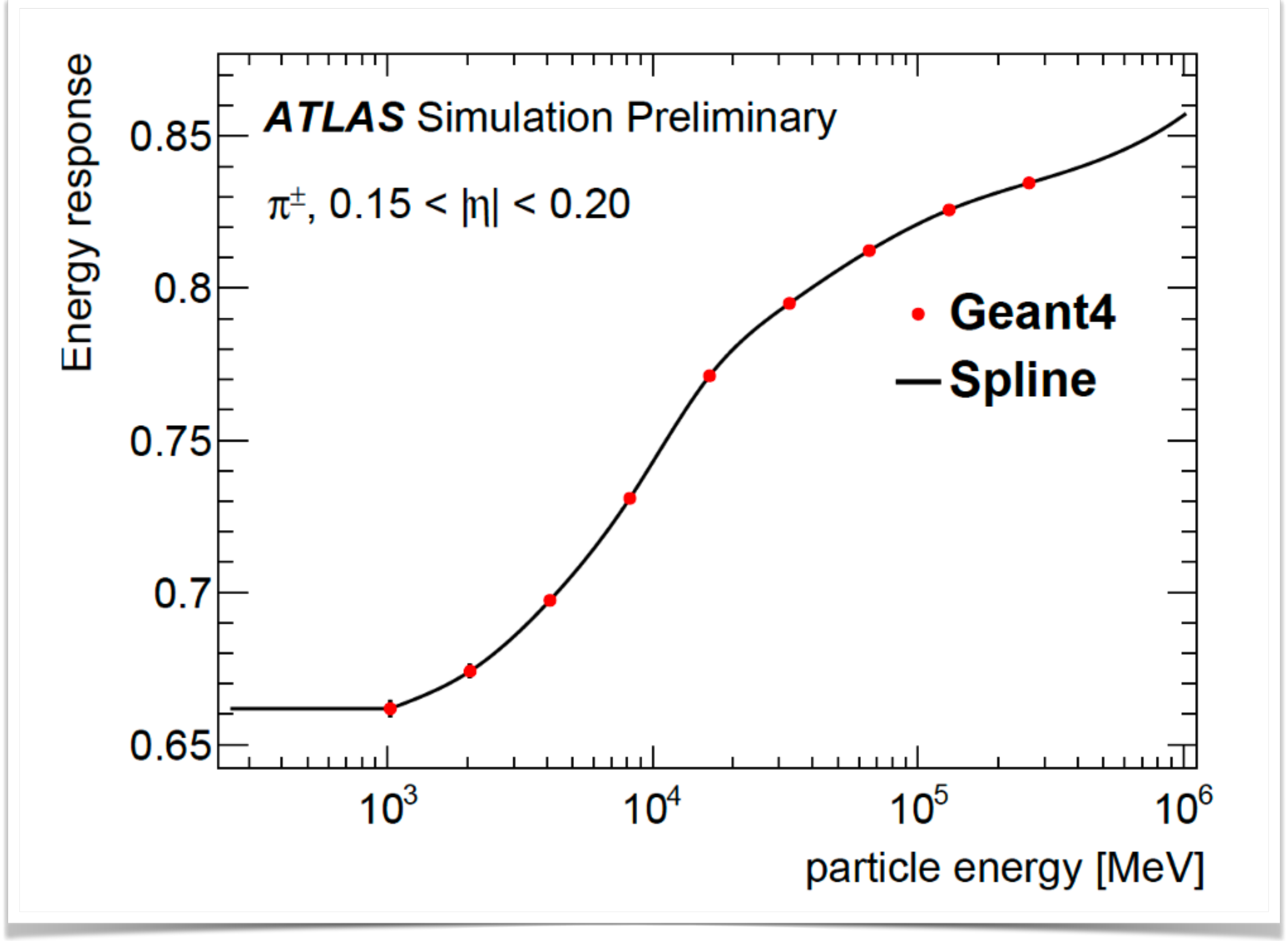
(c) First vs fourth PC

Energy Parametrization & Interpolation

- Additional PCA on each bins of 1st Principal Component and the cumulative distributions, mean & RMS of the gaussians along with the PCA matrix is saved for energy parametrization - for the 17 discrete points.
- A piece-wise polynomial (spline) is used to fit the 17 energy points for interpolation.
- During simulation the parametrization is randomly selected based on the logarithm distance of E_{true} from parametrization grid.

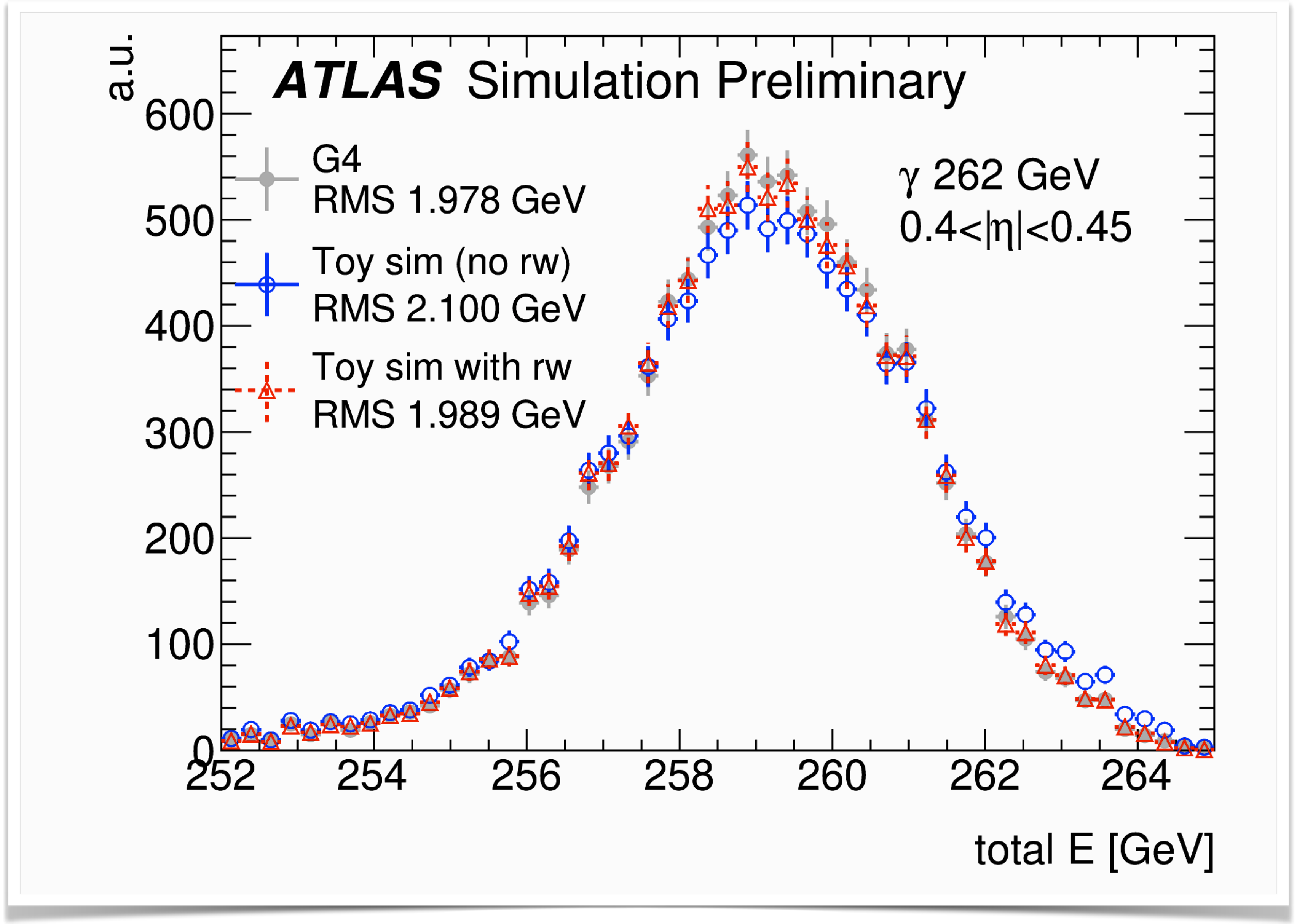


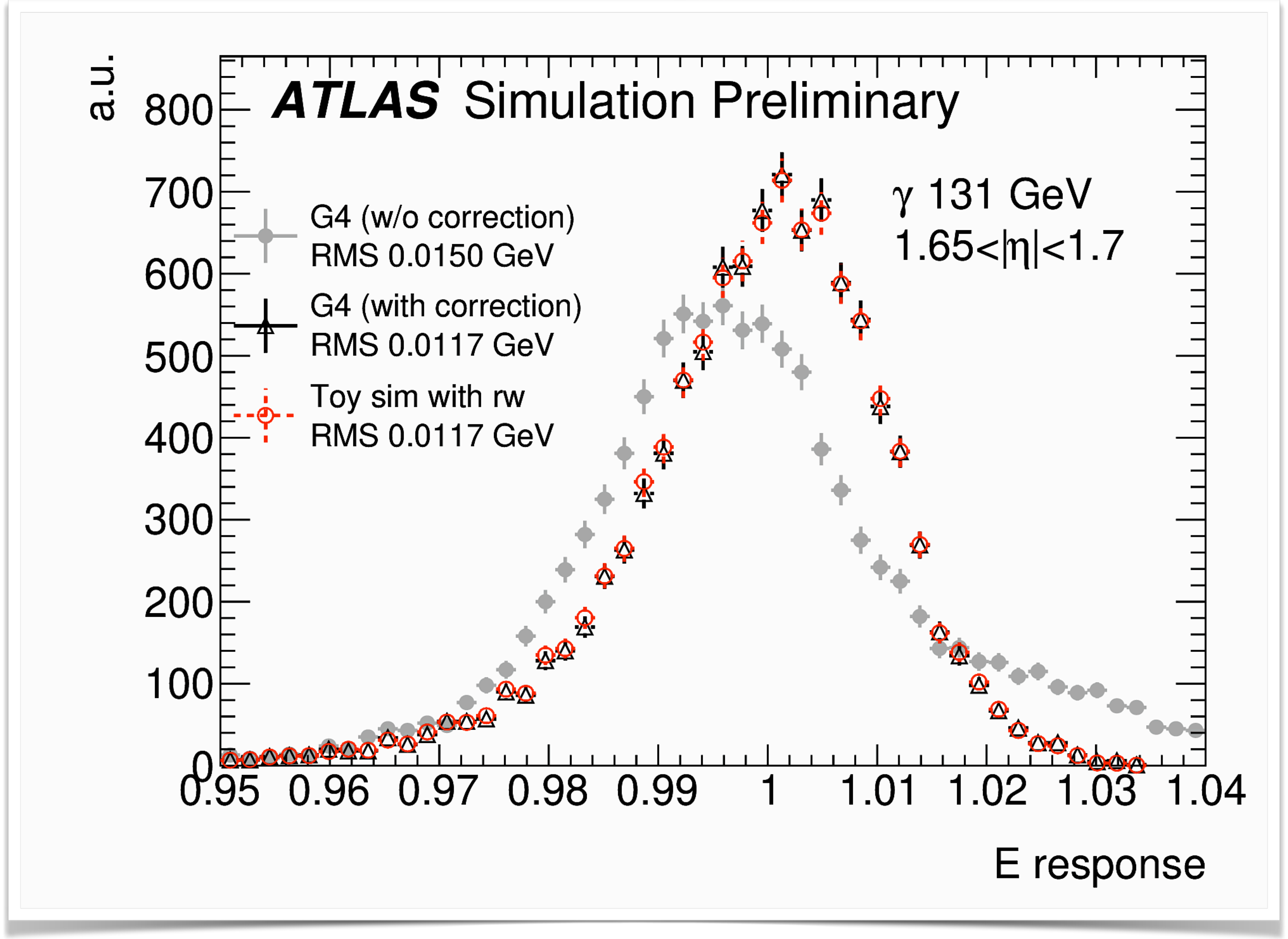




Corrections to improve energy resolution

- Probabilistic reweighting to reject simulated energy far off from the G4 distribution - using a PDF derived from simulated over expected energy.





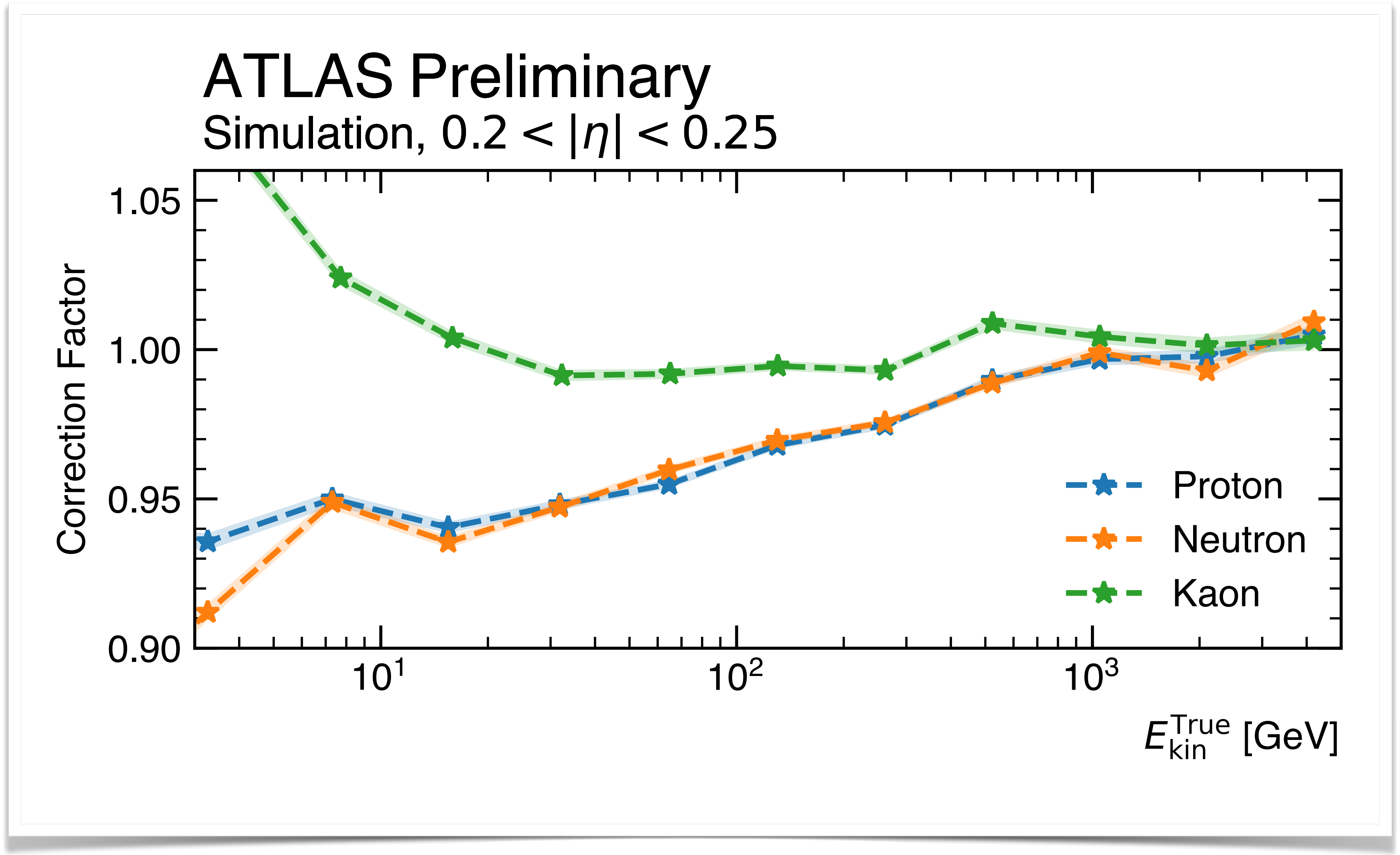
- There is a modulation of energy in the phi direction in the input G4 due to accordion structure in Liquid Argon (LAr) calorimeter which is not modeled in AF3.
- The Geant4 inputs are corrected before parametrization to “flatten” the phi modulation.

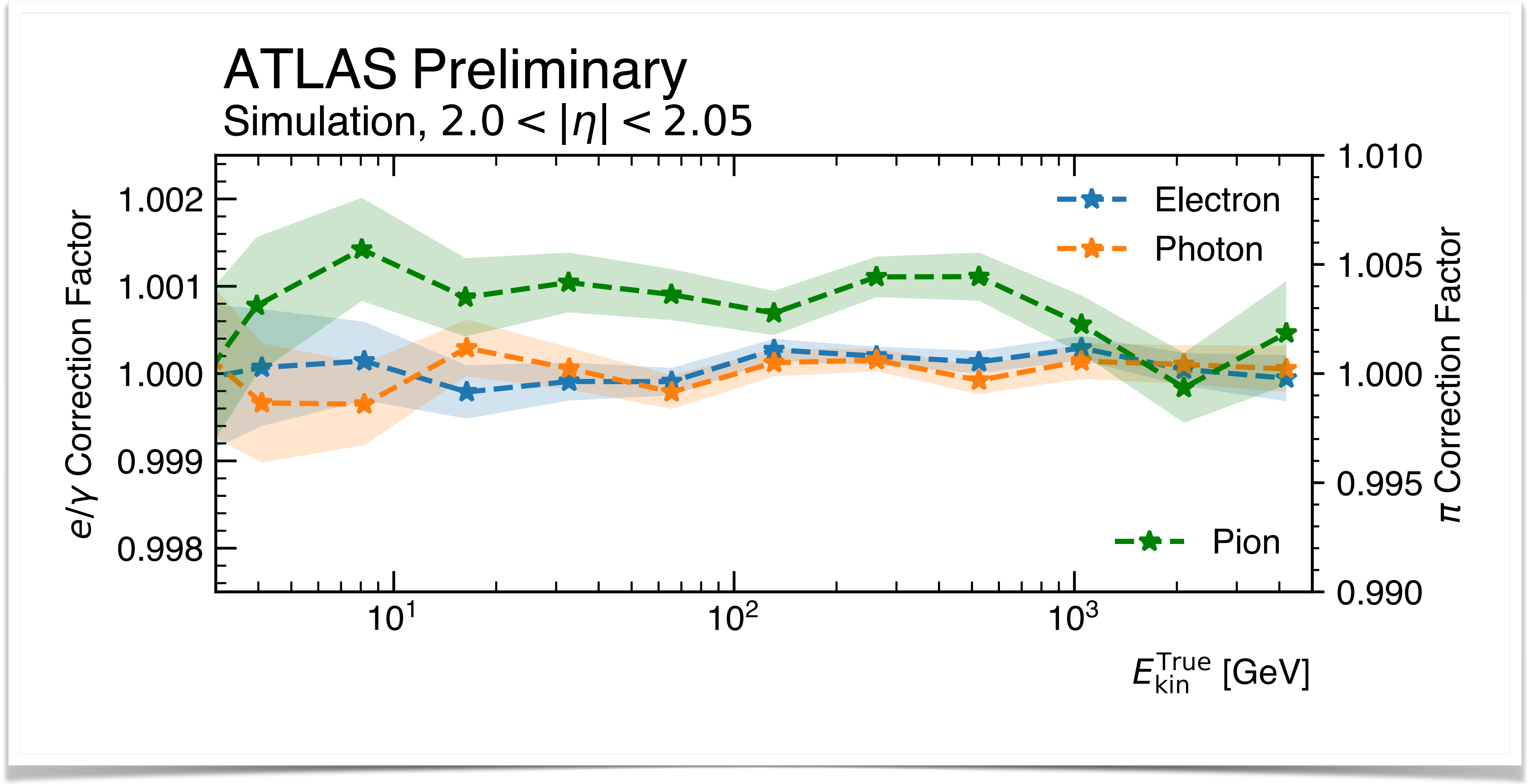
Hasib Ahmed(U Edinburgh)

Corrections to improve mean energy

- Small residual differences in energy response simulation in electron , photon and pion:
  - derive correction factors $\bar{E}_{\mathrm{G4}}/\bar{E}_{\mathrm{AF3}}$ for each energy and η points with linear interpolation in between
  - for photons & electrons the corrections are applied if the correction factor is statistically significant

- Hadron showers simulated with pion parametrization has an intrinsic energy difference:
  - derive $\bar{E}^{\mathrm{Hadron}}_{\mathrm{G4}}/\bar{E}^{\pi}_{\mathrm{G4}}$ correction factors scaled by $E^{\pi}_{\mathrm{kin,true}}/E^{\mathrm{Hadron}}_{\mathrm{kin,true}}$
  - the correction factors are linearly interpolated in between the discrete energy points

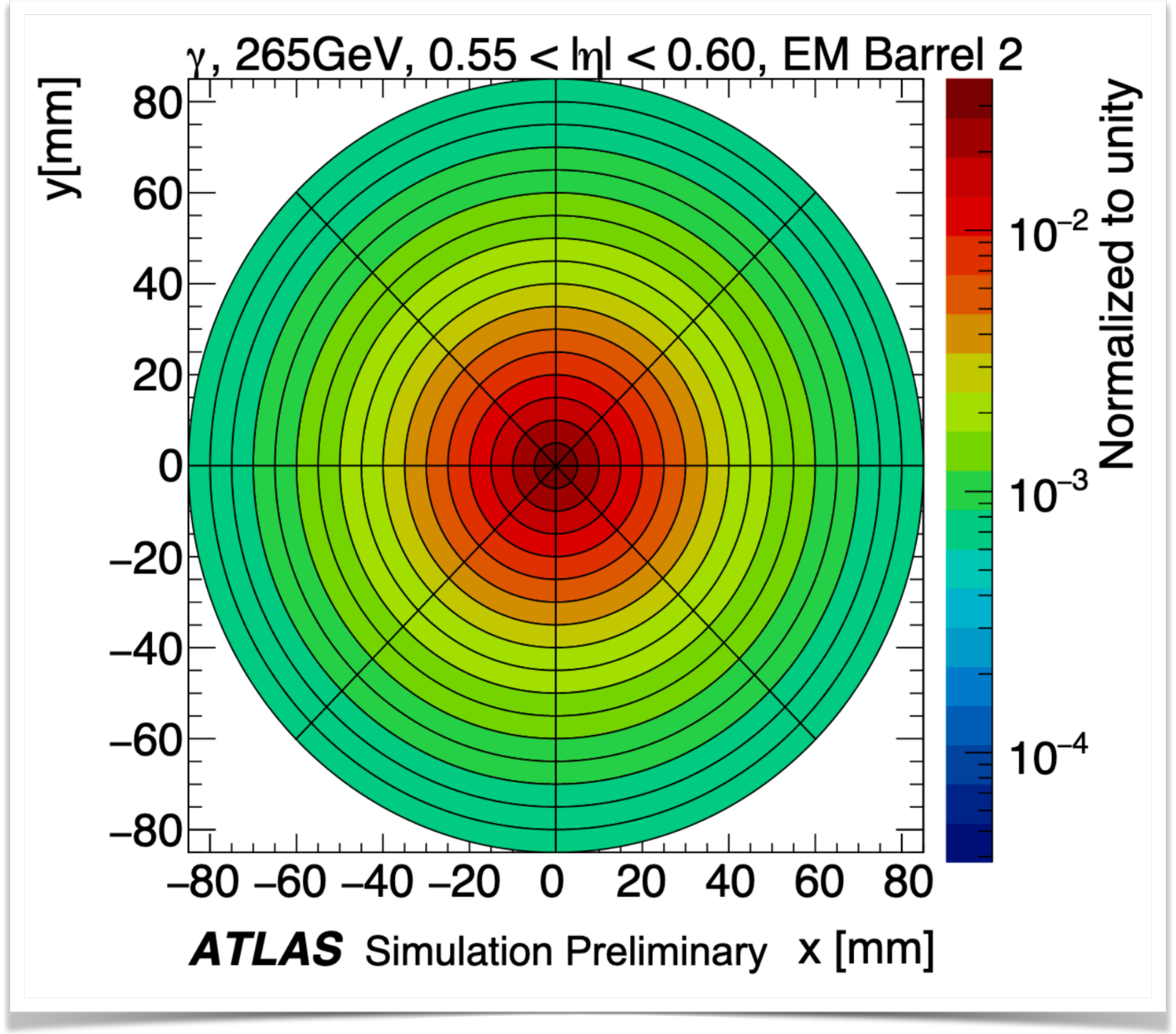




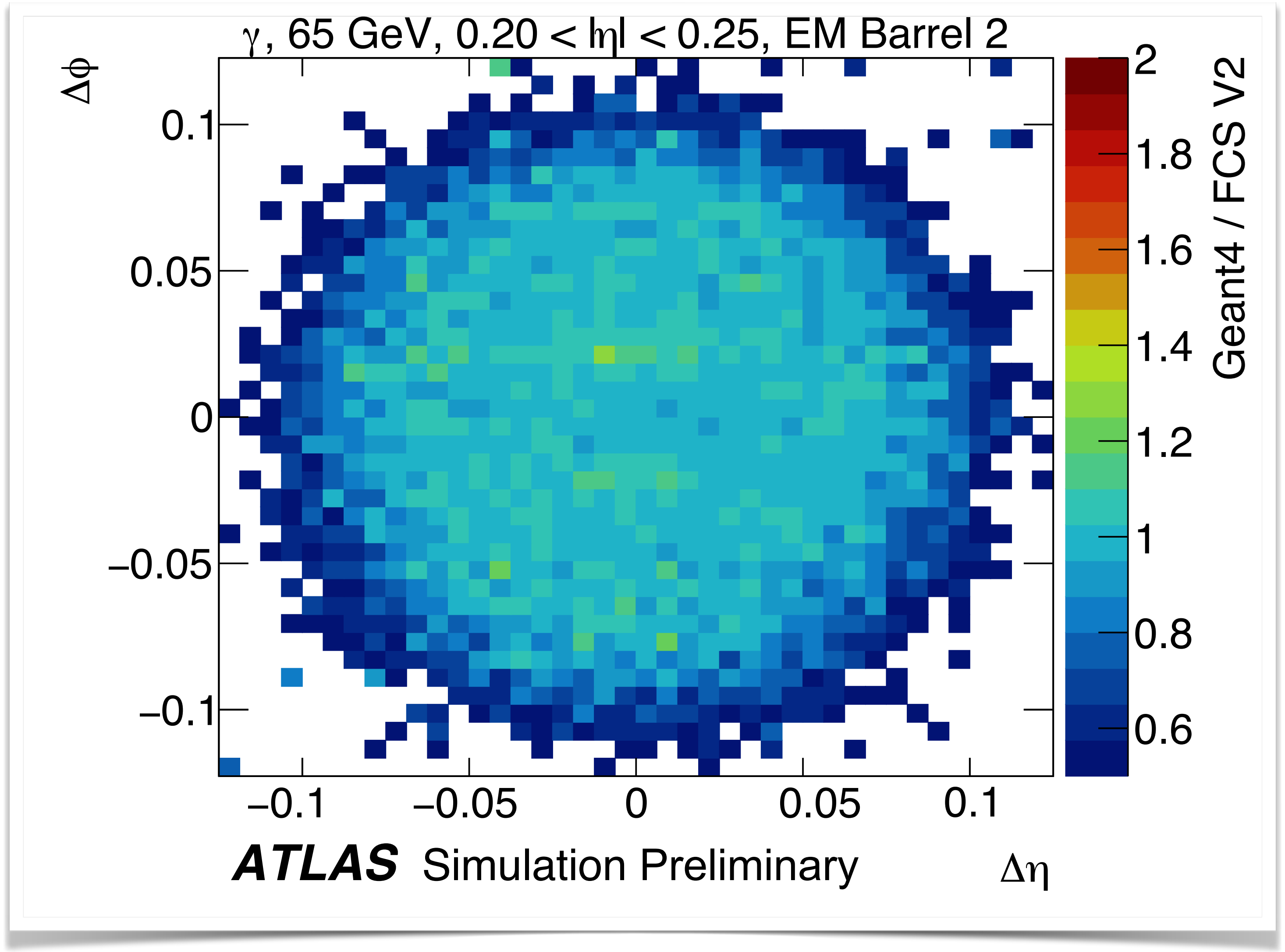
Hasib Ahmed(U Edinburgh)

Lateral shower shape parametrization

- The distribution of energy in lateral direction averaged over many showers is parametrized over a certain radial distance (r) containing 99.5% of the total energy and 8-bins in the angular direction (α).
- The bin size (1 or 5mm) in the radial direction is coarser compared to G4 steps but finer compared calorimeter cell size in each layer.
- Shower centers are corrected by average longitudinal depth of energy deposits in each PCA bin.
- This parametrization is done for each layer and PCA bin for each parametrization grid point.
- These 2D histograms are used as PDF during simulation to randomly generate quantized energy deposits (hits)

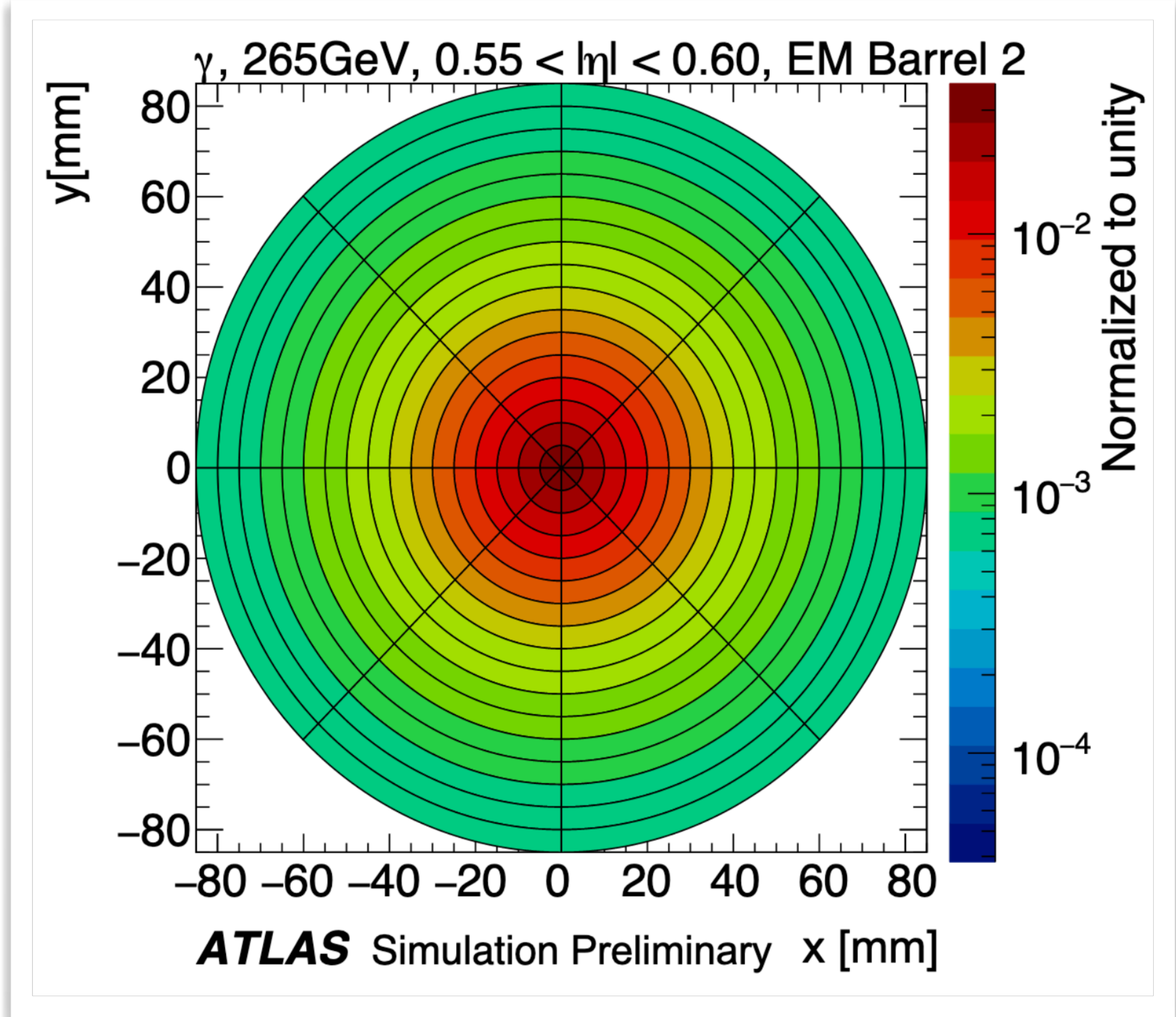


toy simulation



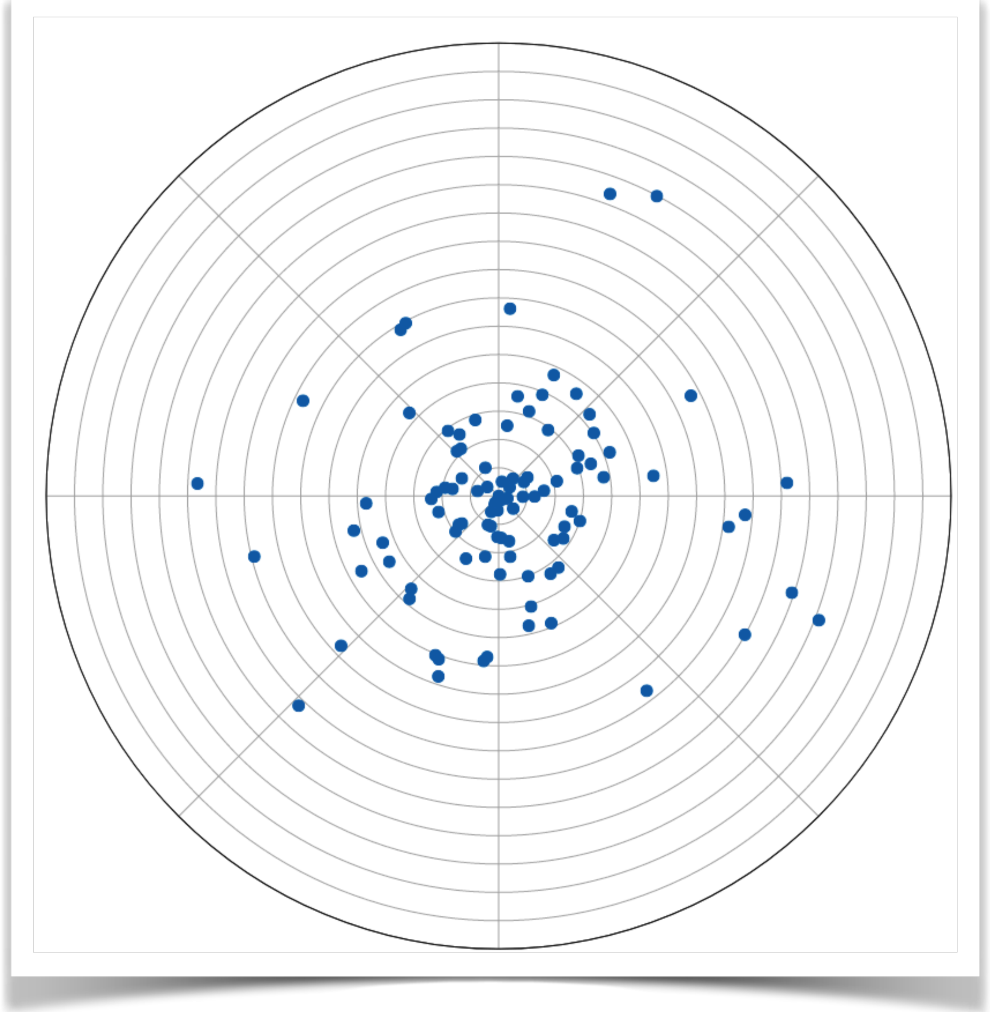
Simulation of lateral shower

- Generation of shower is a stochastic process with the average shower gives the PDF.
- Energy is deposited using N_{hits} of equal energy.
- N_{hits} is calculated such that it gives the same poisson RMS as the resolution of the calorimeter layer.



$$\sigma_E/E = a/\sqrt{E/\mathrm{GeV}} \oplus c$$

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



$$E_{hit} = E_{layer}/N_{hits}$$

Calorimeter technology	Constant term c	Stochastic term a
LAr EM barrel and endcap	0.2%	10.1%
Tile	5.5%	56.4%
LAr hadronic endcap	0	76.2%
FCal	3.5%	28.5%

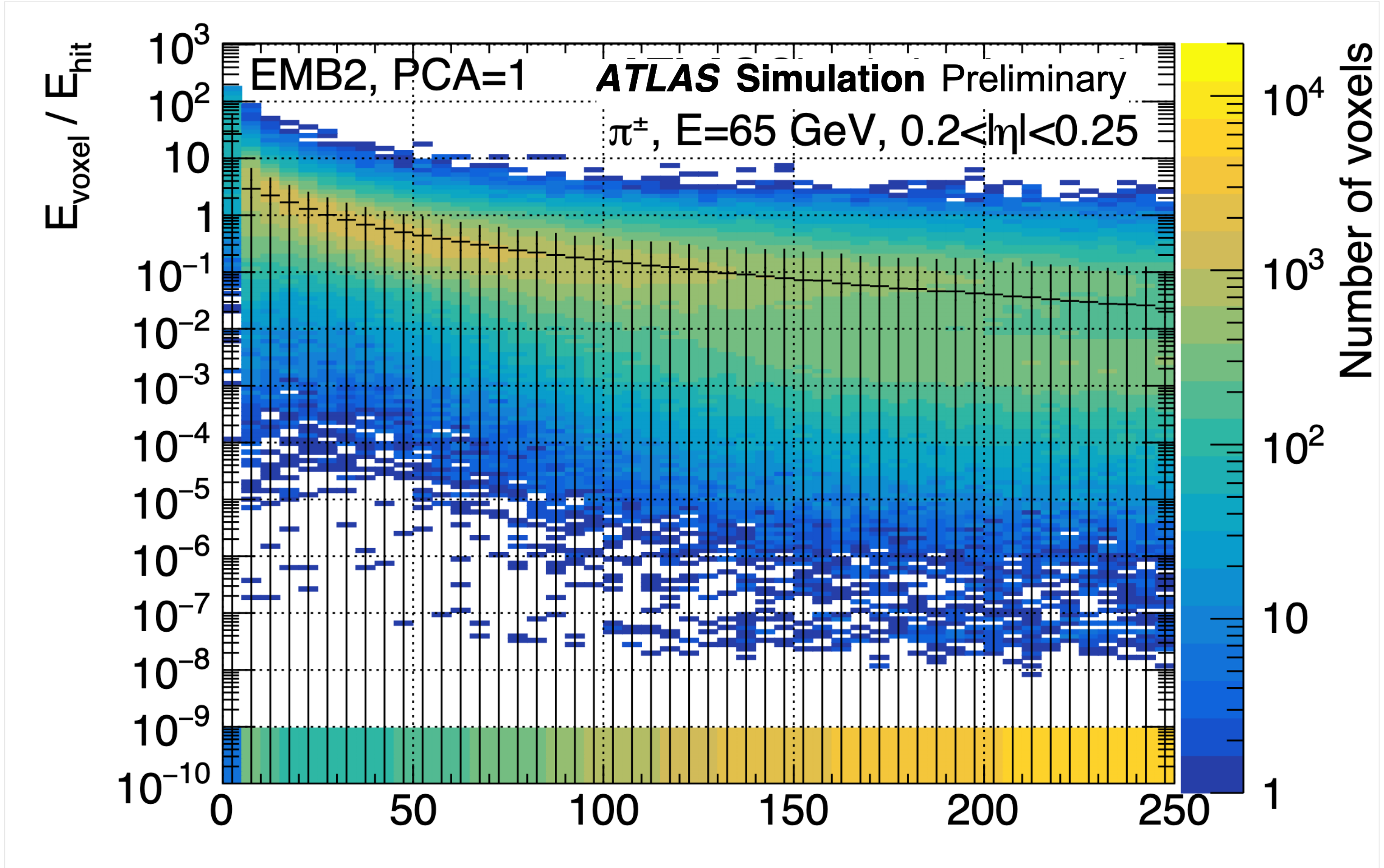
Hadronic showers have larger intrinsic fluctuations and the stochastic terms are calculated for each layer and η region

Calorimeter	Stochastic term a
EM	30 - 40%
Tile	50 - 60%
Hadronic endcap	60 - 80%
FCal	80 - 100%

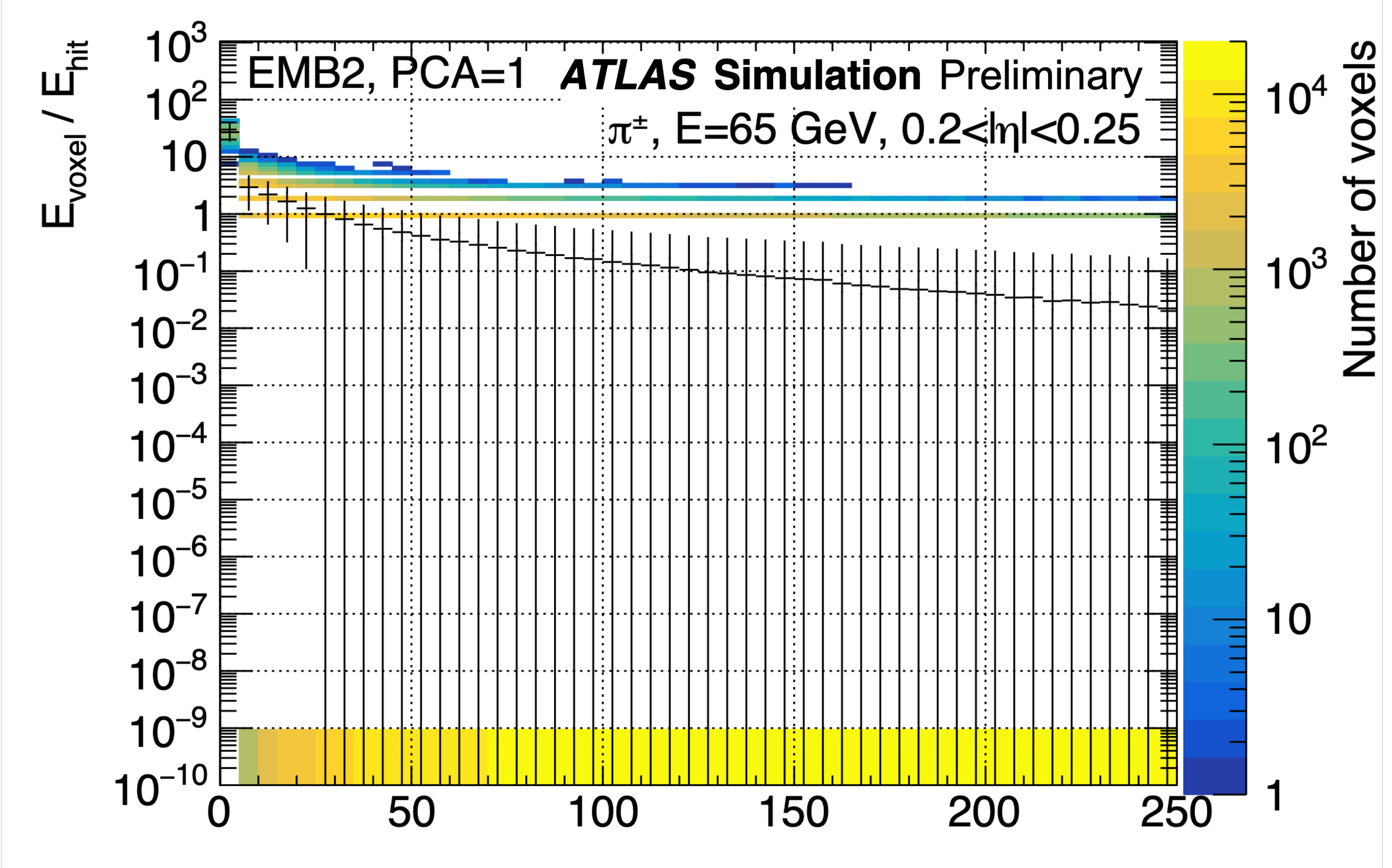
This model with equal energy hits works well for EM sowers but require hit reweighting for Hadronic showers.

Weighted hit simulation for hadrons (1)

- Hadronic calorimeter layers have large stochastic terms (> 30%) leading to large energy deposits (100 - 300 MeV) for hits with equal energy.
- Even only few hits at far away from the shower center have large probability to create clusters.
- These low energy clusters introduce mismodeling in the total number of clusters.



Geant4



AF3 with equal hit energy

E_{voxel} - *bins in average shower histogram*

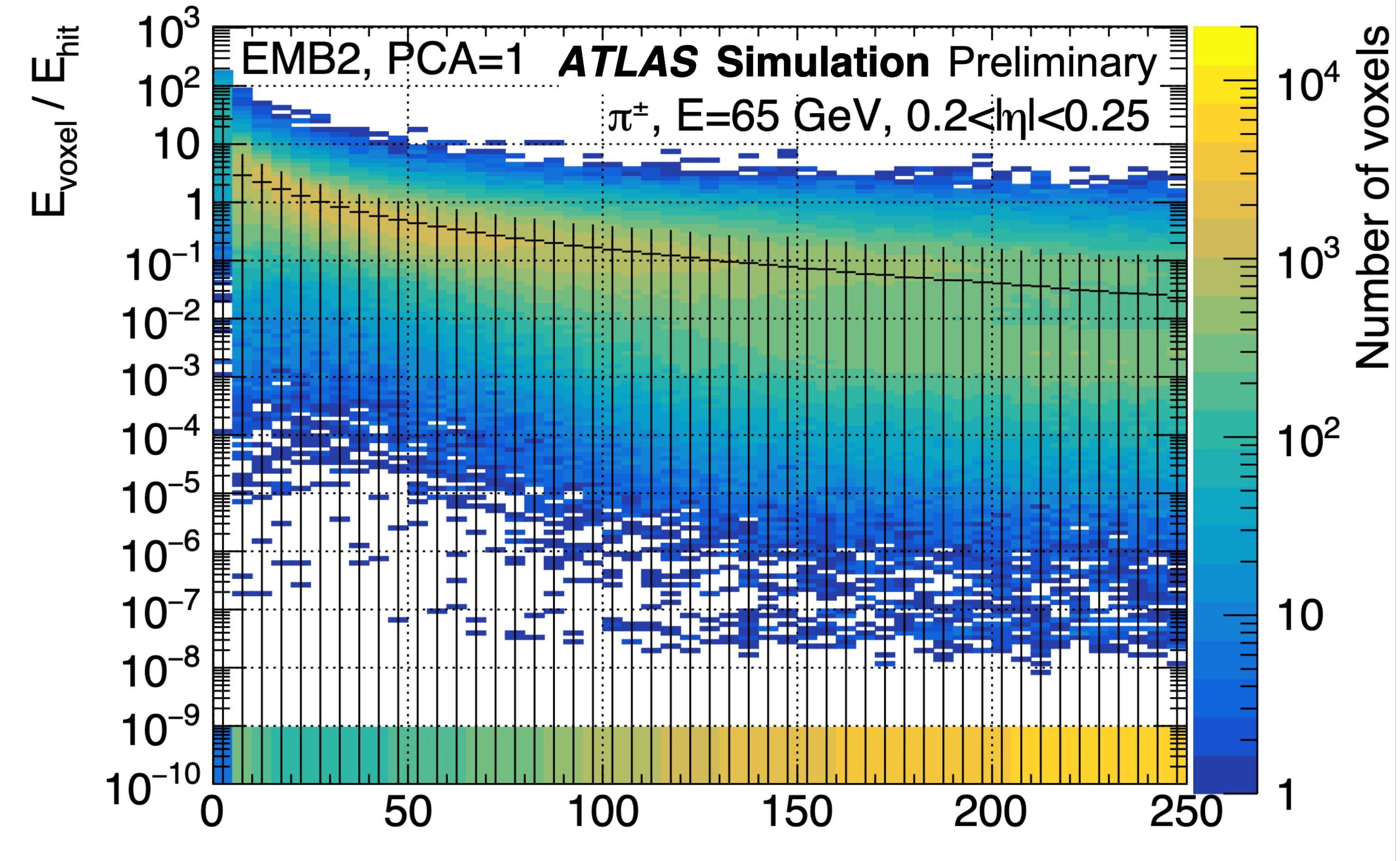
E_{voxel} / E_{hit} - *energy fraction in each bin of avg. shower*

ΔR [mm] - *radial distance from shower center in mm unit*

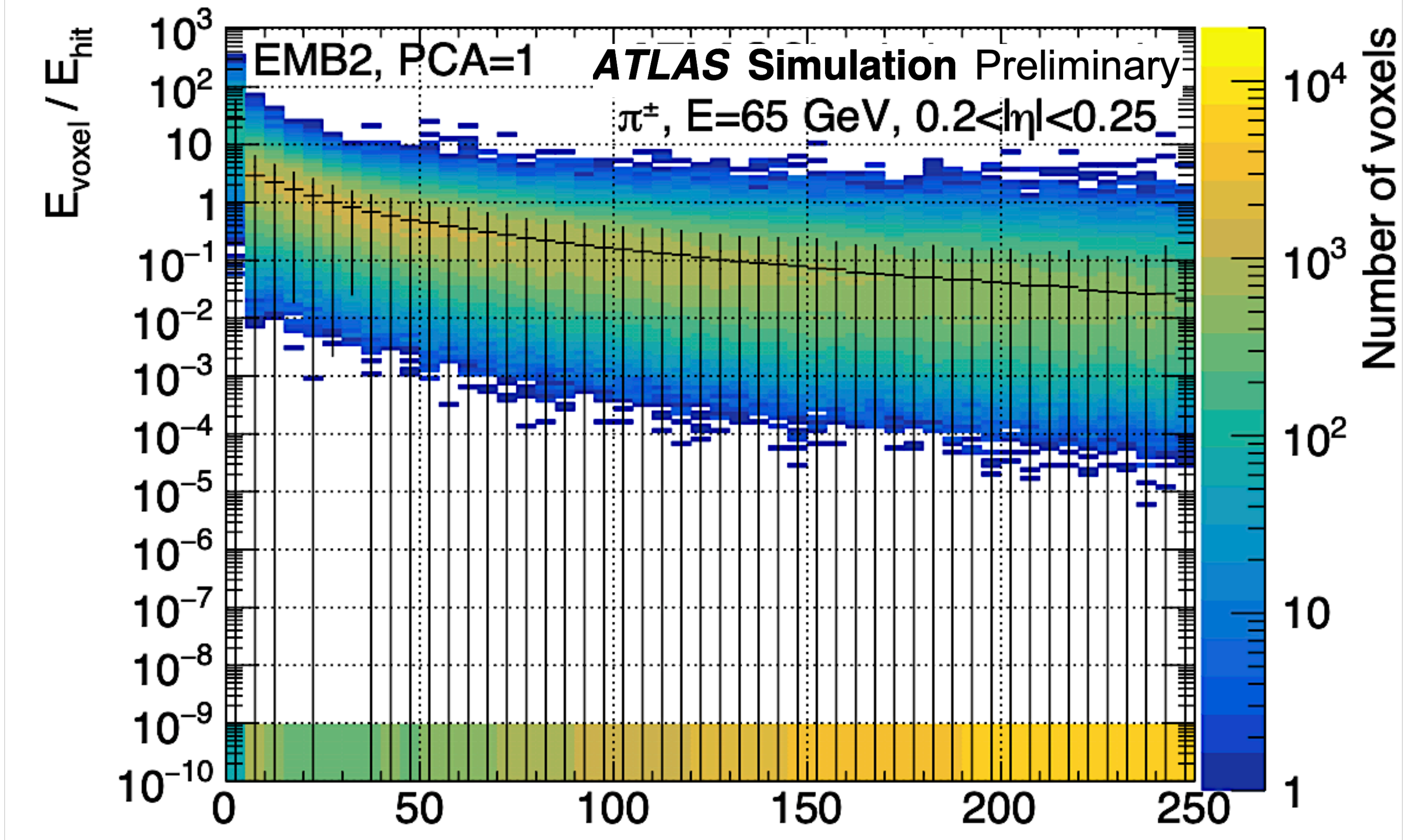
Equal hit energy deposition creates large number of clusters away from the center of the shower not observed in Geant4!

Weighted hit simulation for hadrons (2)

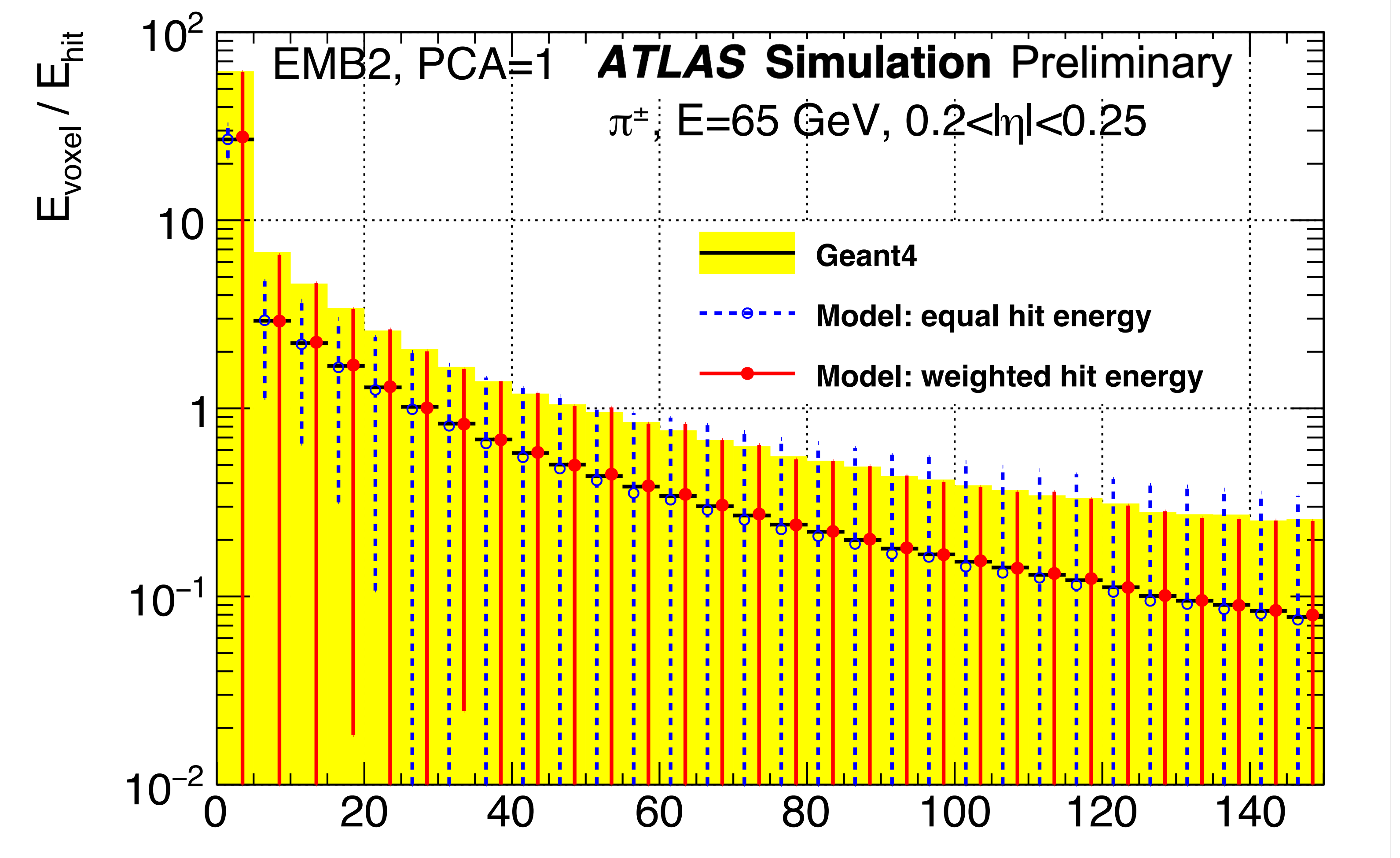
- ✦ The equal hit energy model can reproduce the mean well but not the RMS.
- ✦ Introduce weights to change the RMS of each bin to reproduce the RMS of the Geant4 distribution.
- ✦ Additional smearing is applied to include unaccounted fluctuations.



Geant4



AF3 with weighted hits



comparison of mean & RMS

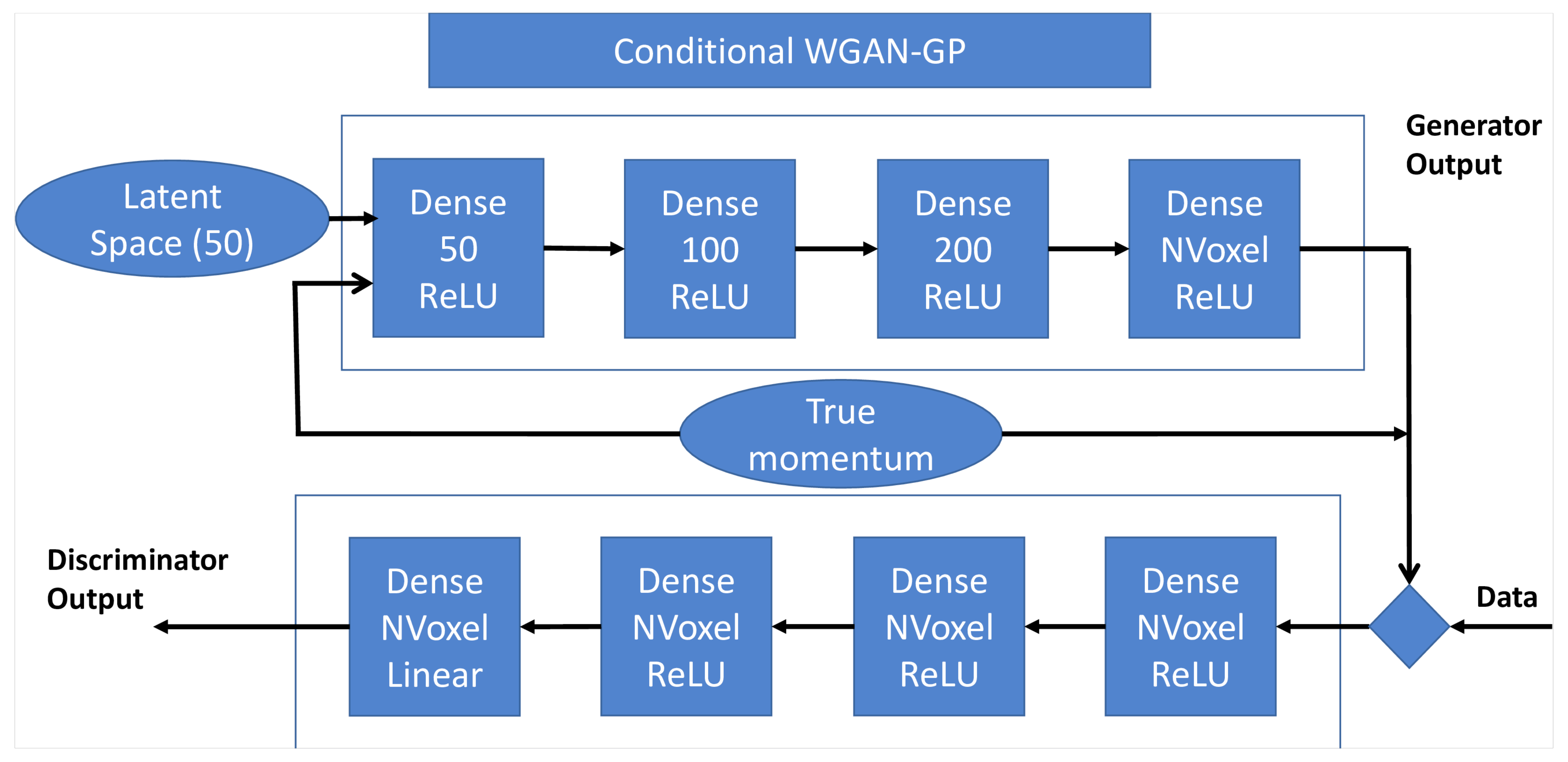
Weighted hit model significantly improves modeling of hadron showers!

Neural Network based- FastCaloGAN

ATL-SOFT-PUB-2020-006

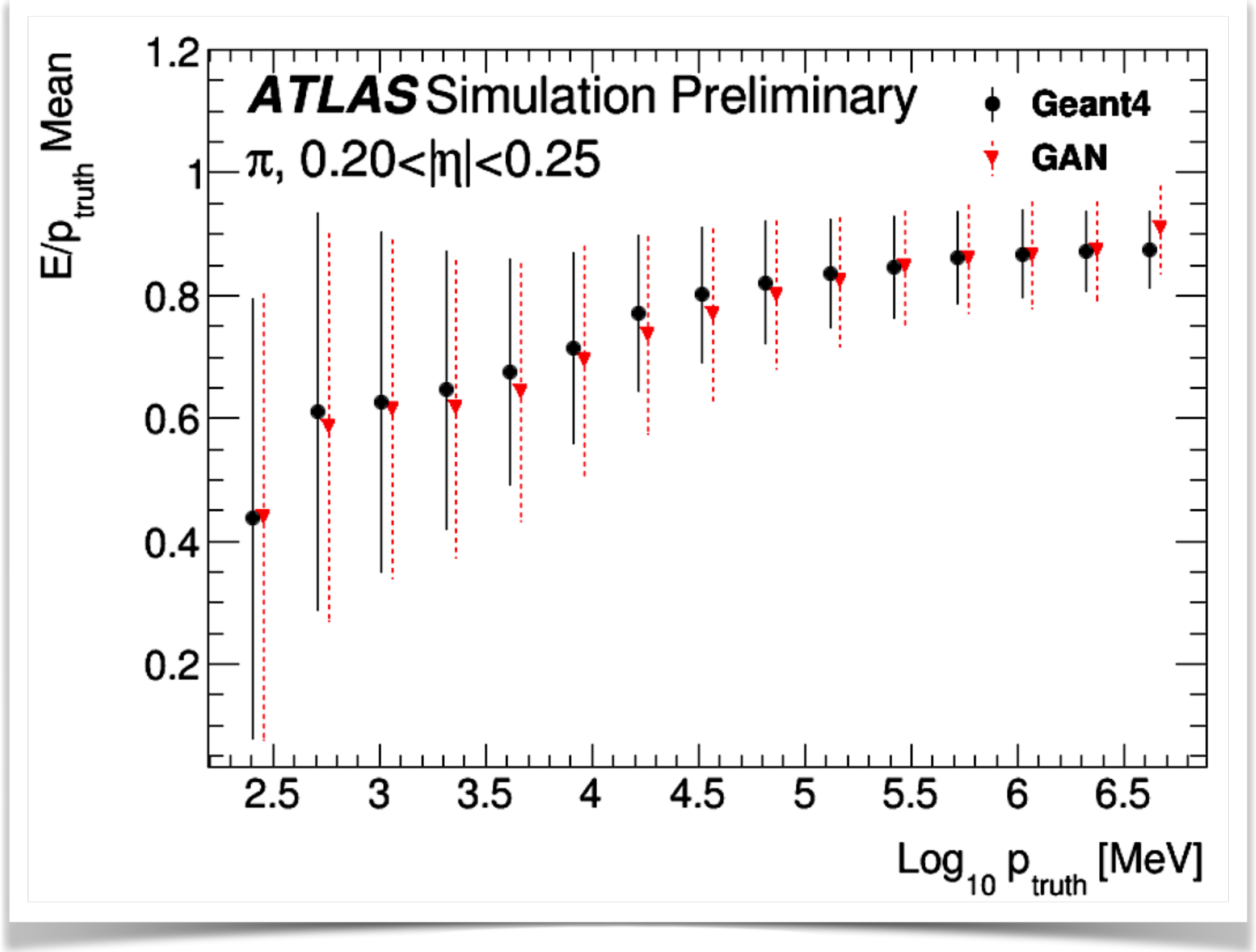
- Generative Adversarial Network (GAN) to simulate shower generation in the entire calorimeter - providing both longitudinal (including correlation between layers) and lateral shower modeling.
- The hits are voxelized in the same frame of reference (r,α) as in FCS V2 shape parametrization - optimized for each particle and η bins.
- Wasserstein loss with gradient penalty (WGAN-GP) is used, conditioned on the truth momentum and trained for each η slice but inclusive in energy - resulting 100 GANs for pions.
- Each GAN is trained for 1M epochs with a checkpoint saved every 1K epochs.
- At simulation step the GAN with best epoch is used to generate hits which are deposited in the corresponding voxels.

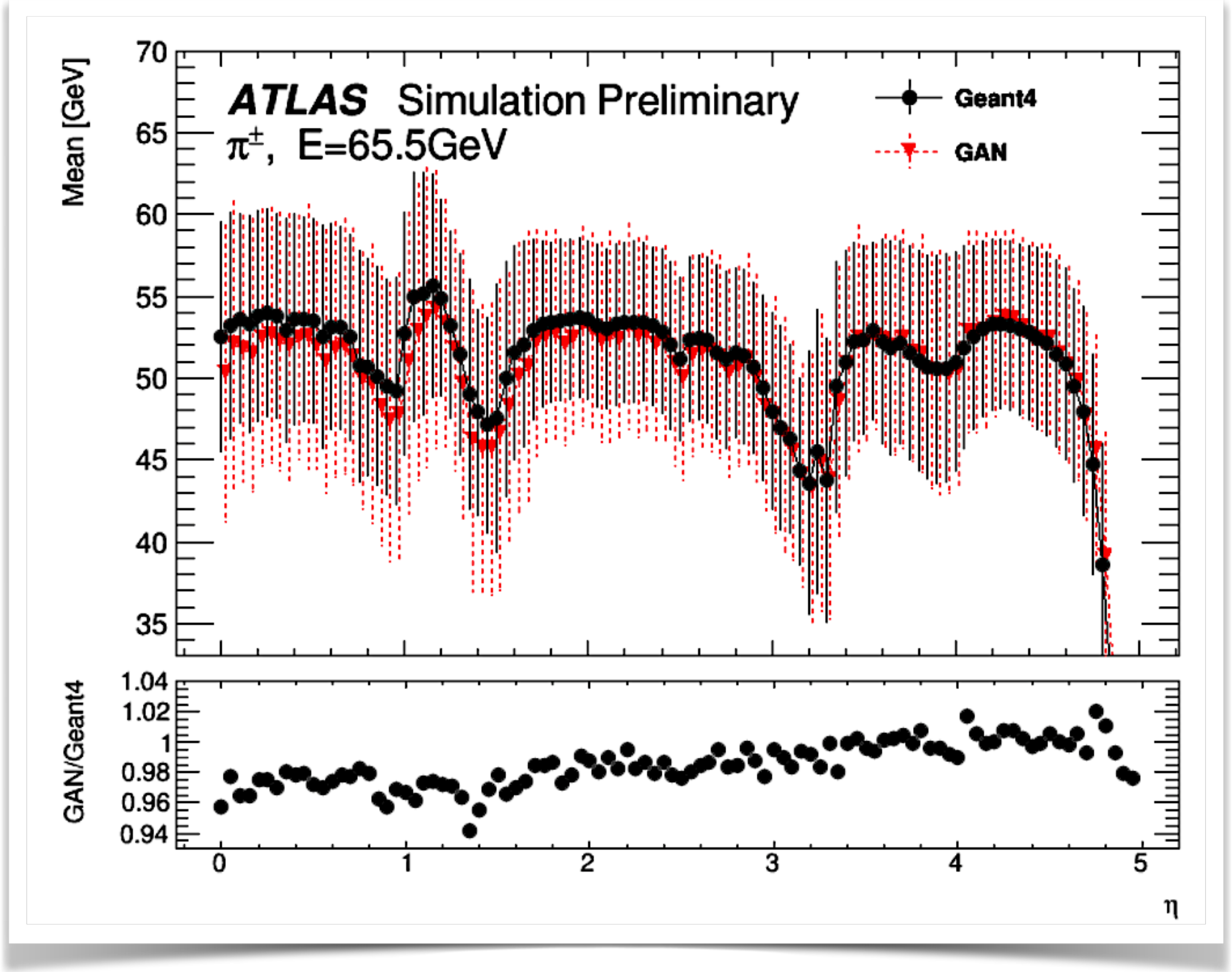
G	50 (Input latent Space), 50, 100, 200, NVoxel
D	NVoxel, NVoxel, NVoxel, NVoxel, 1
Activation function	ReLU (in all layers)
Optimiser	Adam [20]
Learning Rate	10^{-4}
β	0.5
Batchsize	128
Training ratio (D/G)	5
Gradient penalty λ	10



FCSGAN Performance

- FastCaloGAN shows better modeling compared to FCS V2 for hadrons in the medium energy range.
- The exact threshold is determined based on single cluster and jet properties.
- AF3 uses FCSGAN for hadron showers in the range: 16 GeV ≤ E_{kin} ≤ 256 GeV
- The total energy of the FastCaloGAN is scaled to the energy of FCS V2 - allows smooth transition between the two simulation flavors.

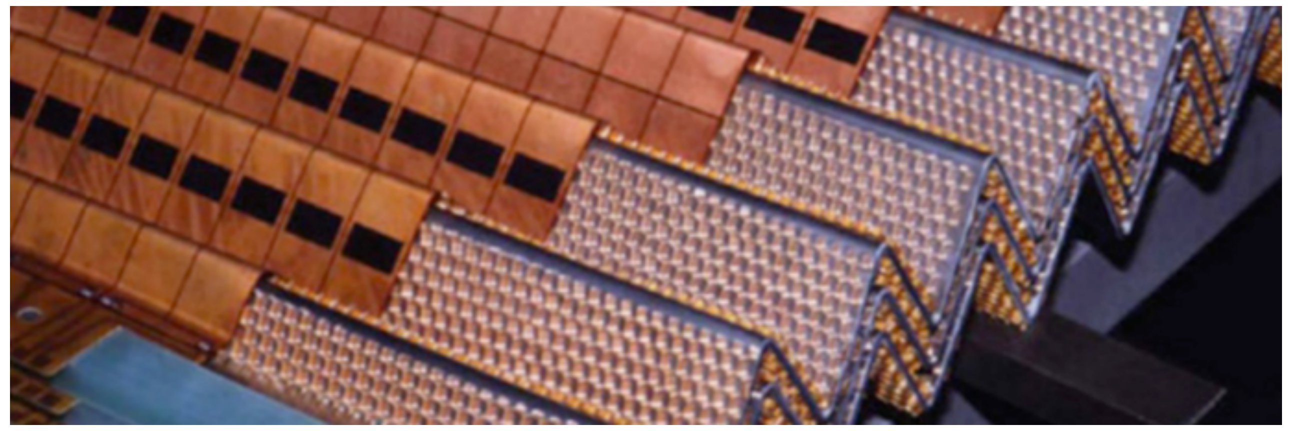




Fully implemented in the ATLAS simulation infrastructure and will be used as part of AF3 for sample production!

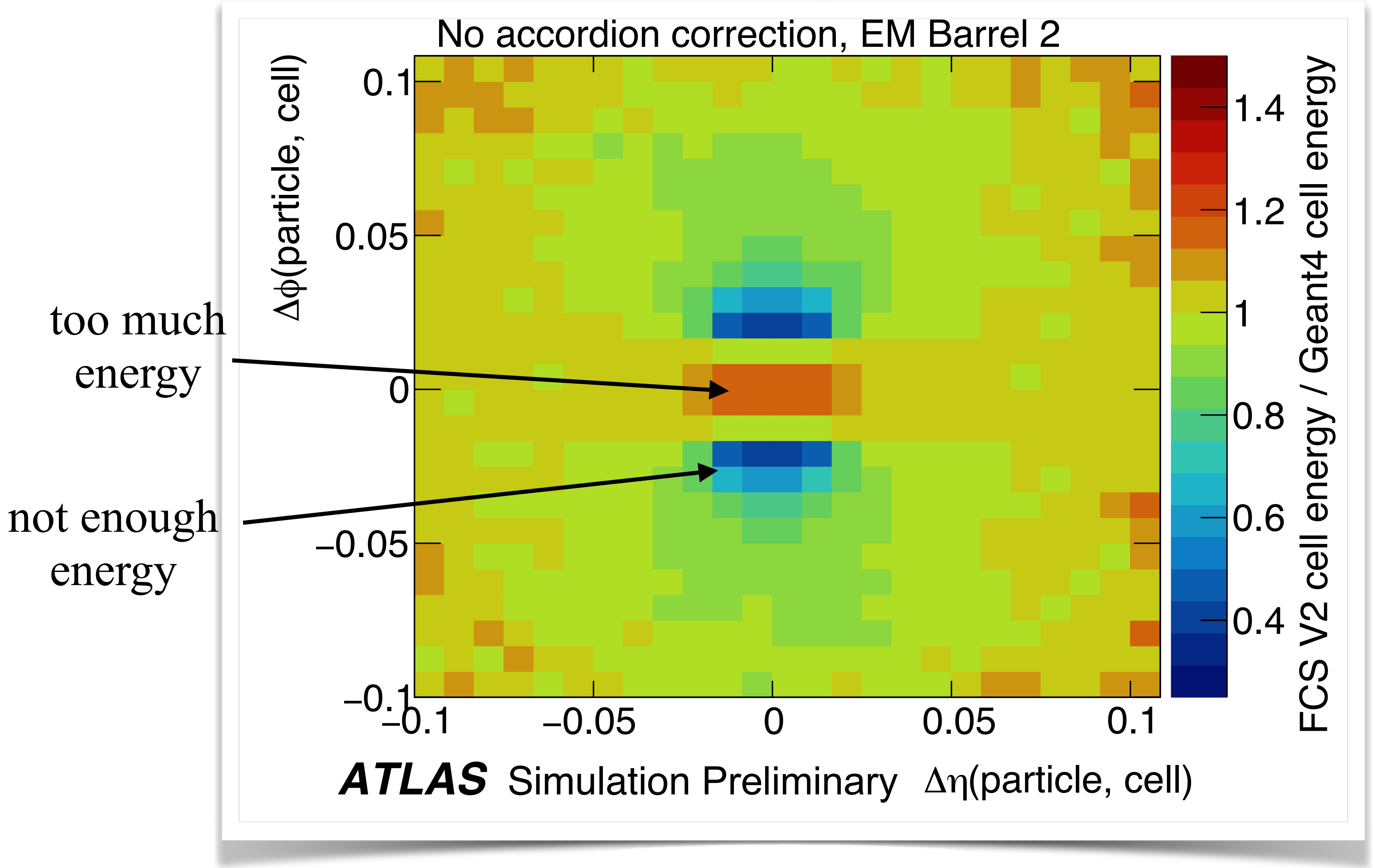
Assigned quantized energy (hits) to calorimeter cells

- Simulated hits (from FCS V2 or FCSGAN) are assigned to cells assuming simplified cuboid geometry.
- Derive a probability density function (PDF) from the difference of cell assignment efficiency calculated in Geant4 and AF3.
- Use the PDF to randomly assign a displacement to a hit before assigning to a cell.

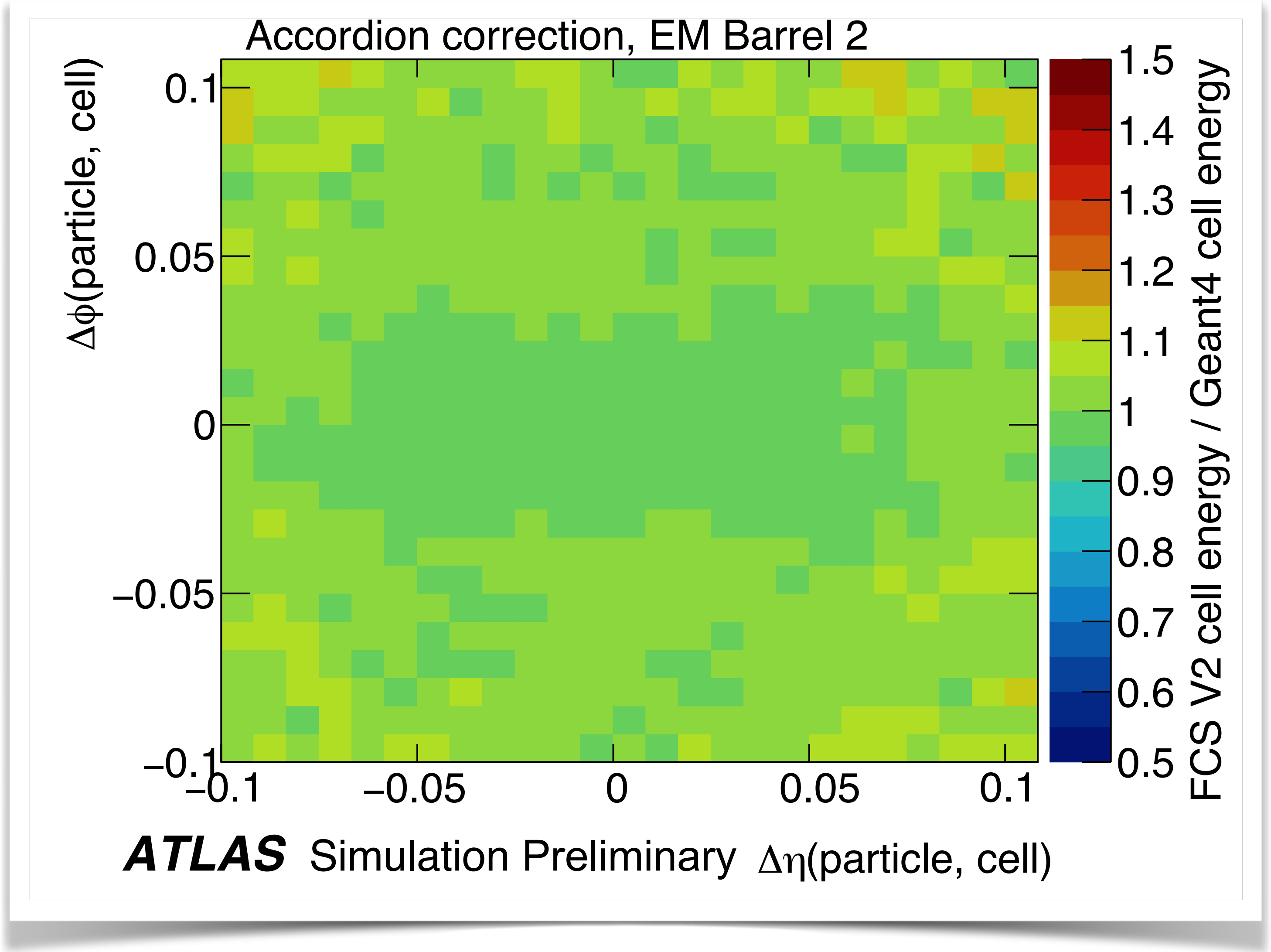


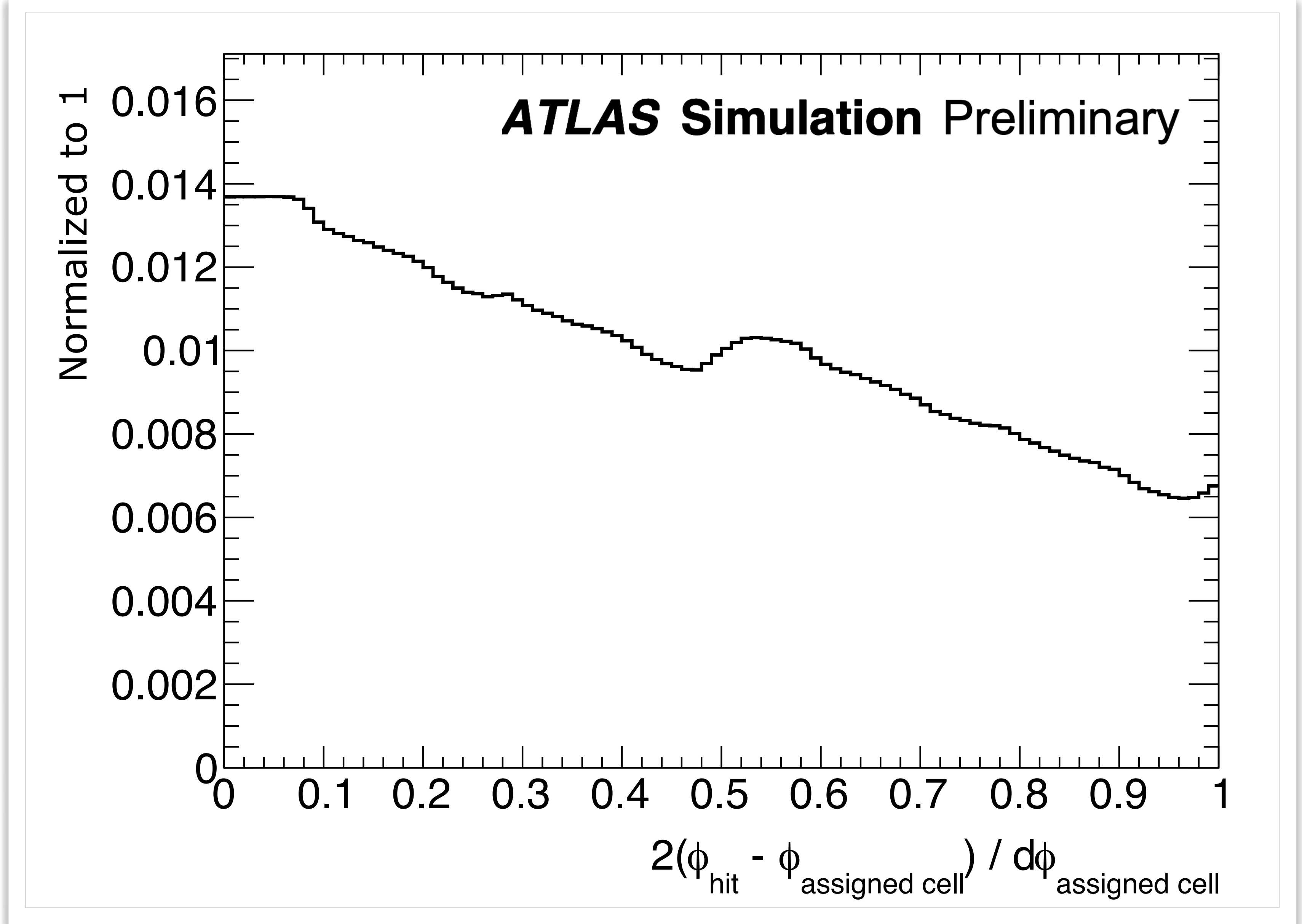
Probability function describing the chance that the energy belongs to this cell

Without correction:



With correcton:

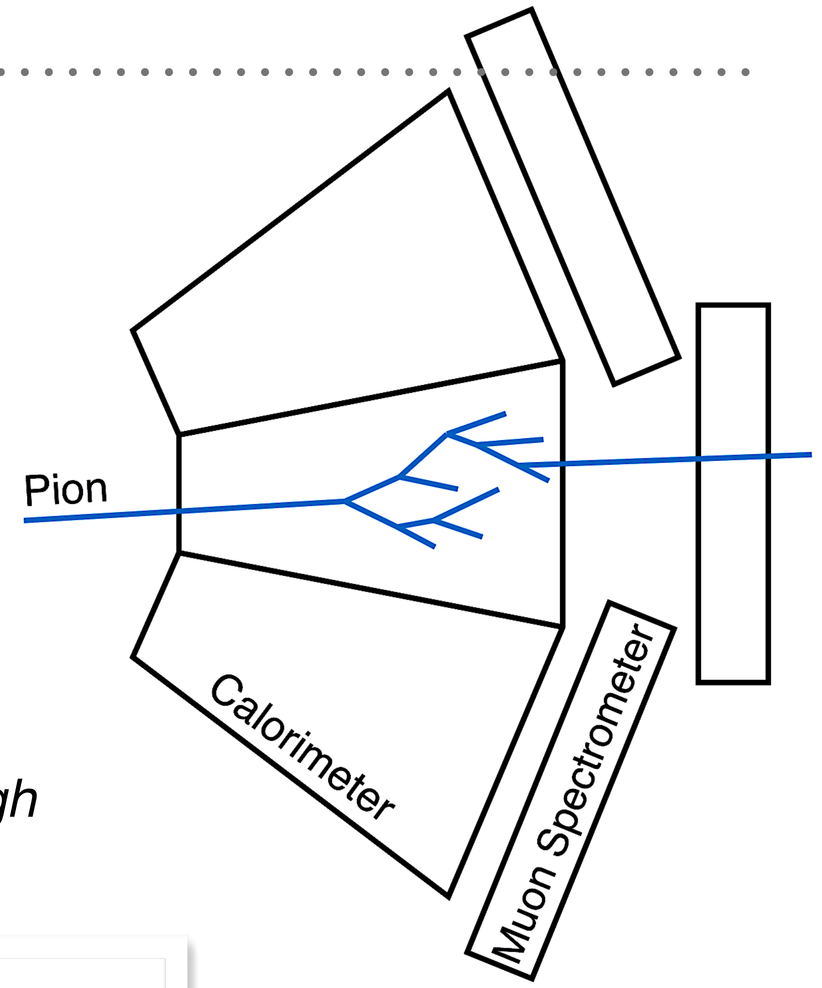




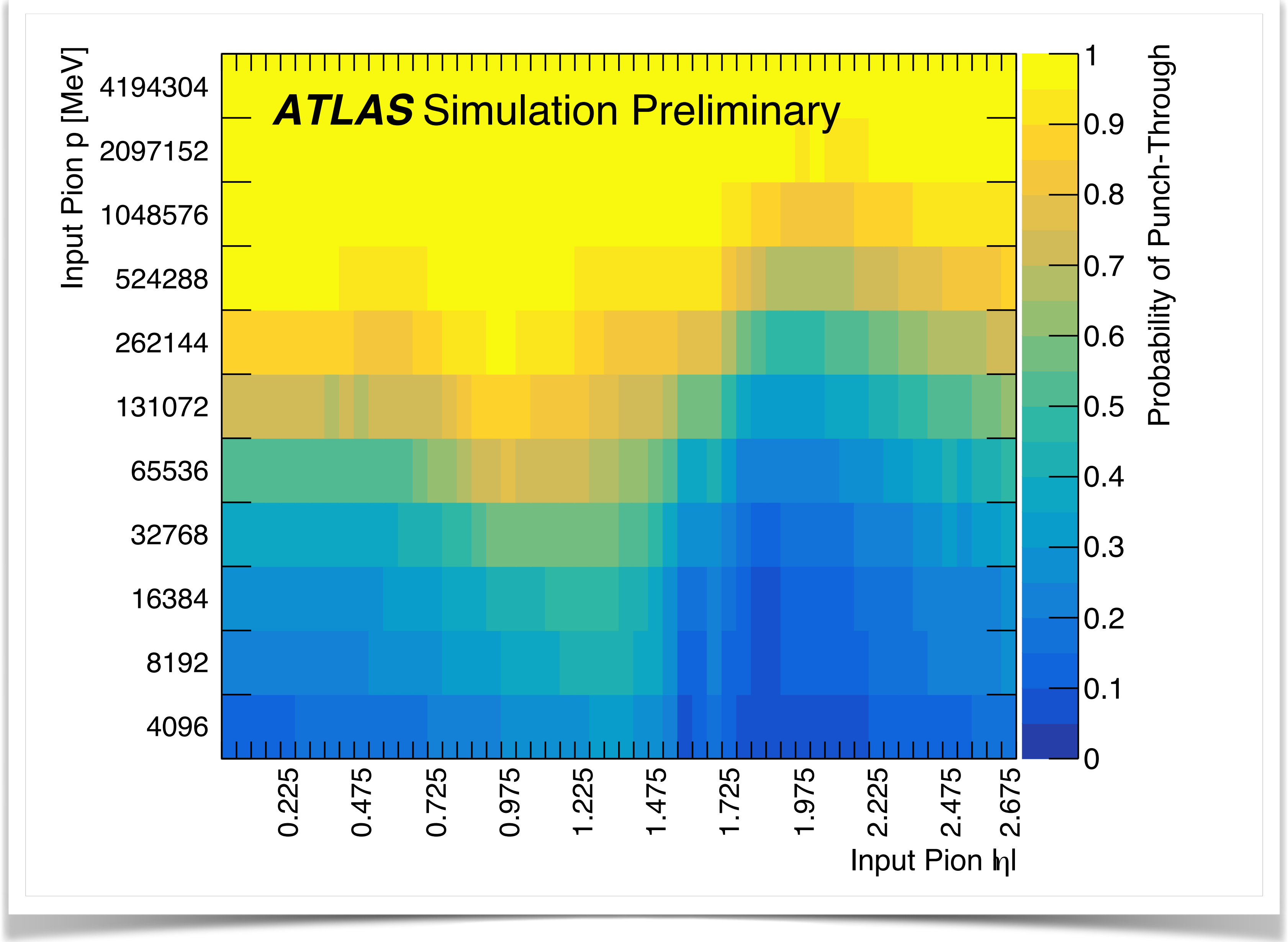
Closure with Geant4 with the correction applied!

Parametrization of particles escaping to the muon systems

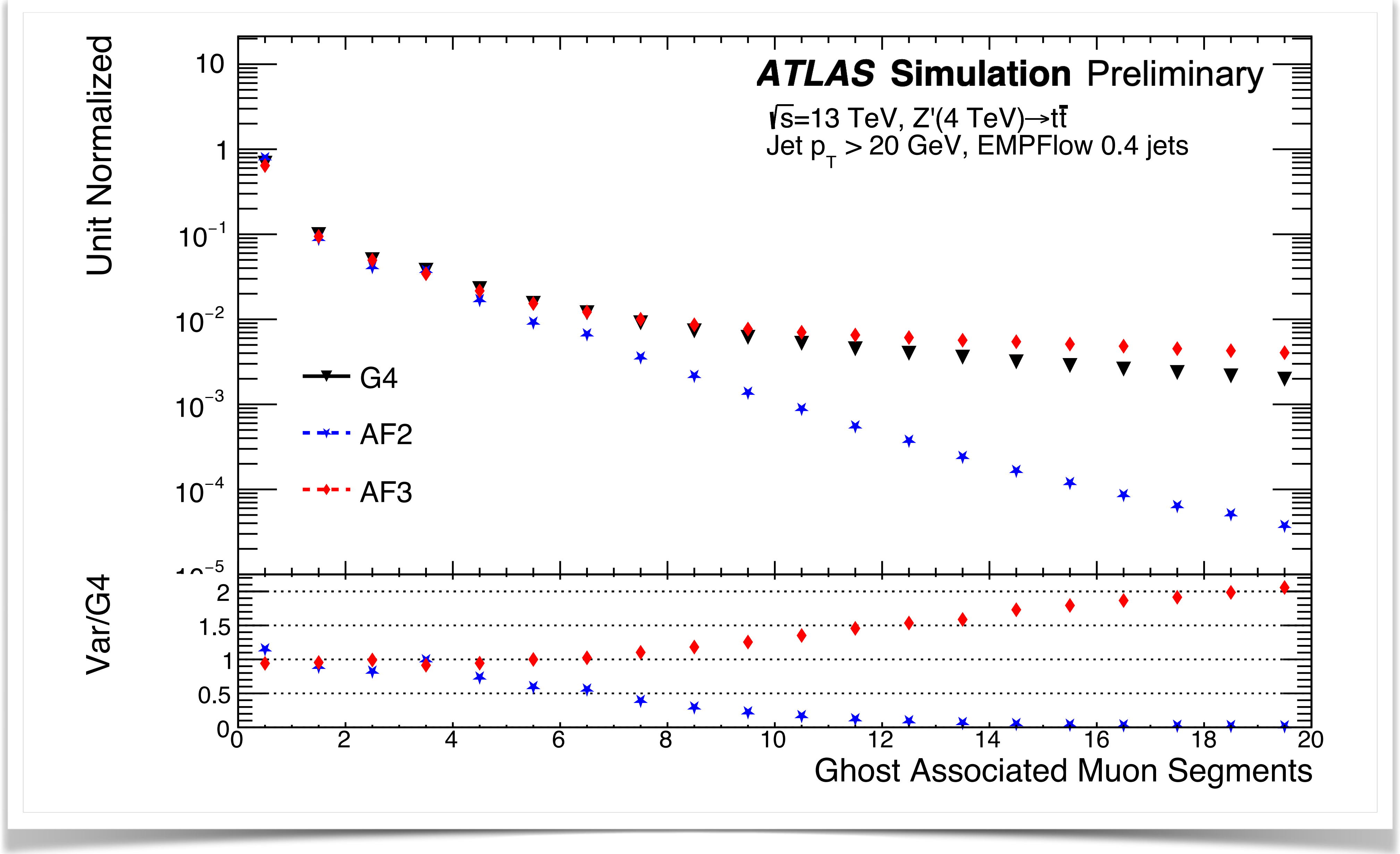
- Particles that punch through to the MS are reconstructed as a fake muon.
- AF3 includes a dedicated parametrization to model the secondary particles (e,γ, π, μ, p)
- Depending on the momentum and η for a pion entering the calorimeter volume, the punch through particles are generated and passed to Geant4.



probability of a single pion to produce at least one punch-through particle of at least 50 MeV



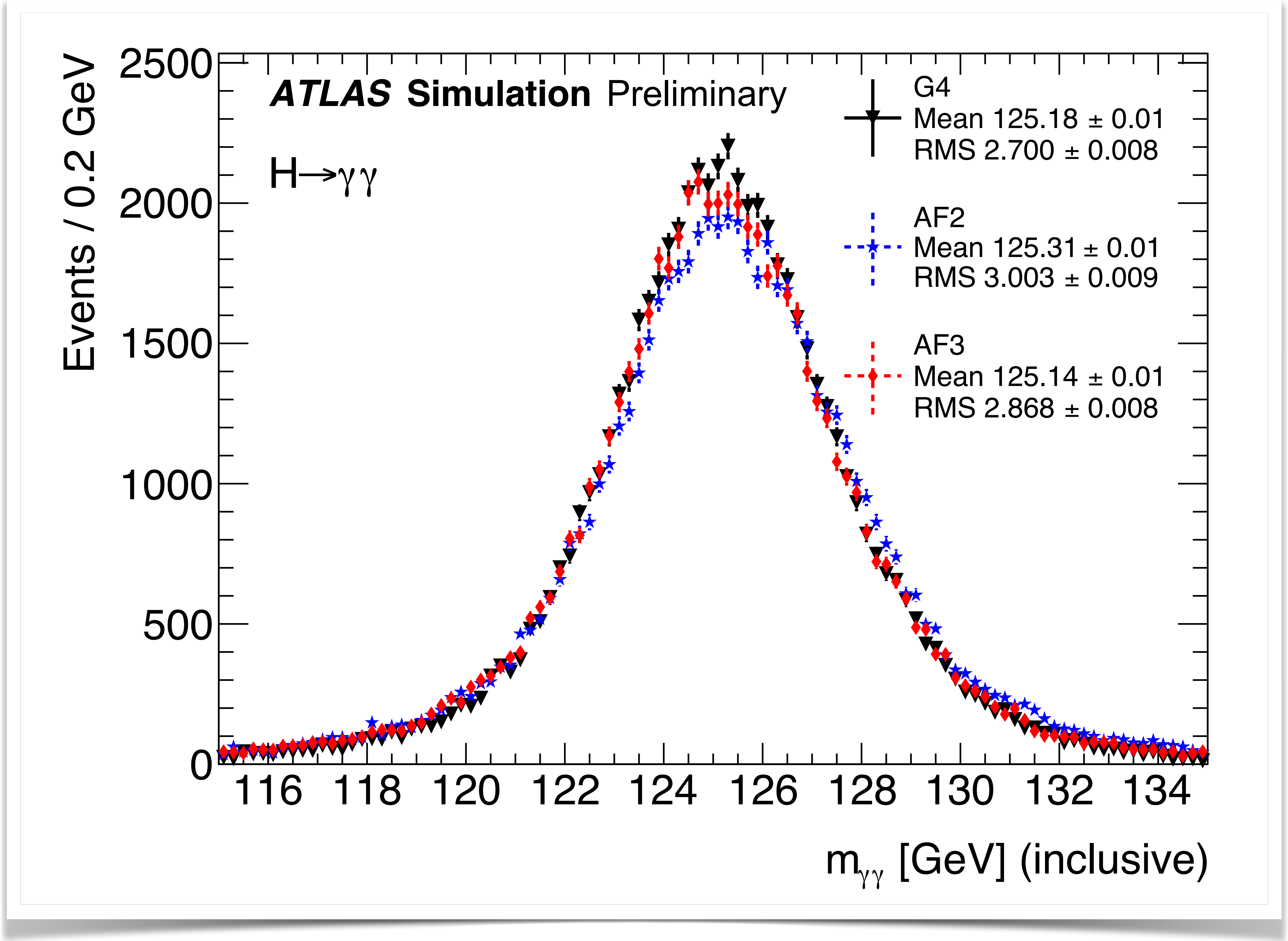
muon segments results from particles punching through the calorimeter as well as real muons

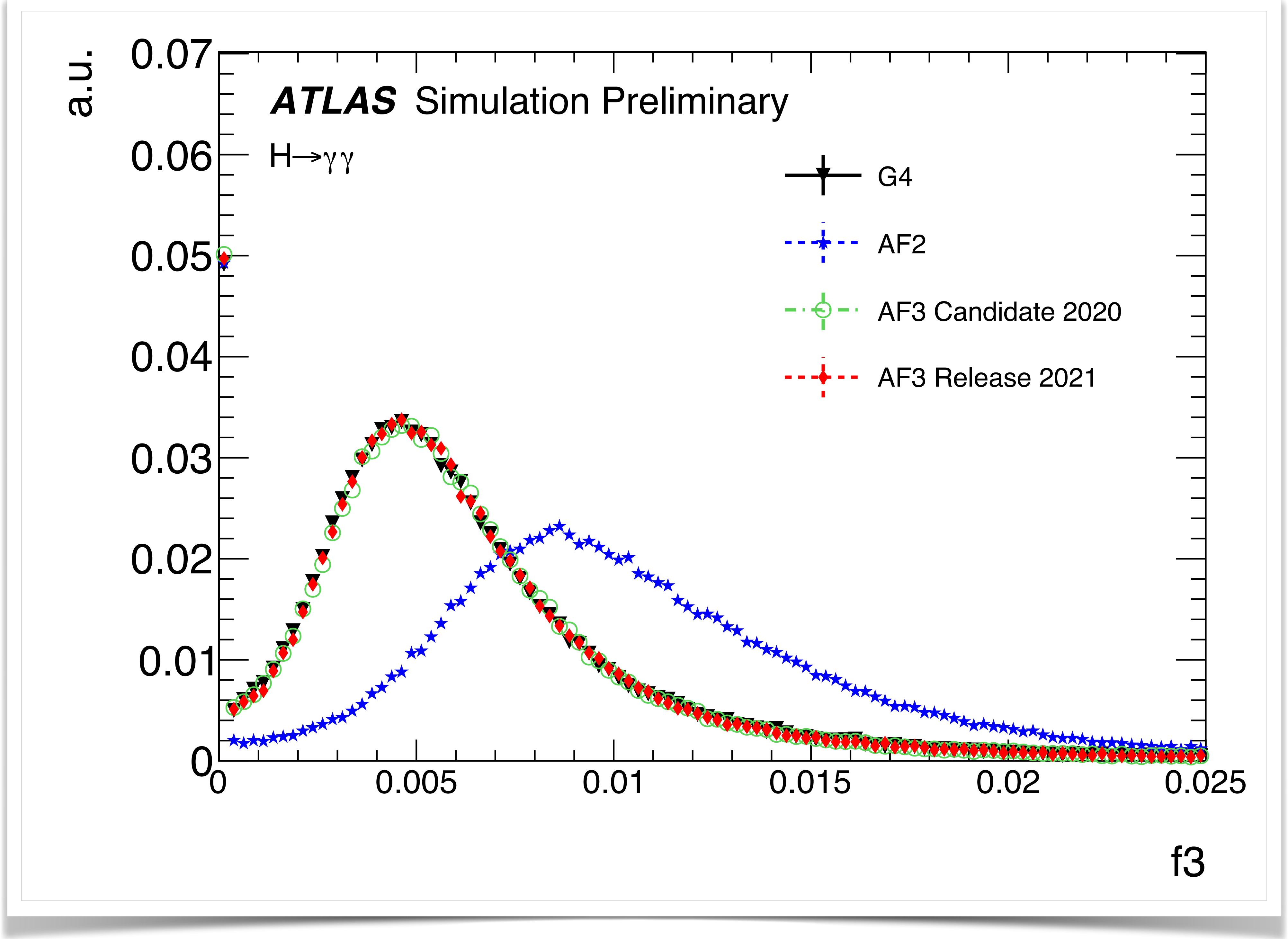


Improves modeling of muons significantly compared to AF2!

Performance of AF3: reconstructed photons & electrons

- Photons and electrons are reconstructed from clusters of energy deposits in EM calorimeter.
- The objects are selected with identification criteria with high purity as used in physics analyses.



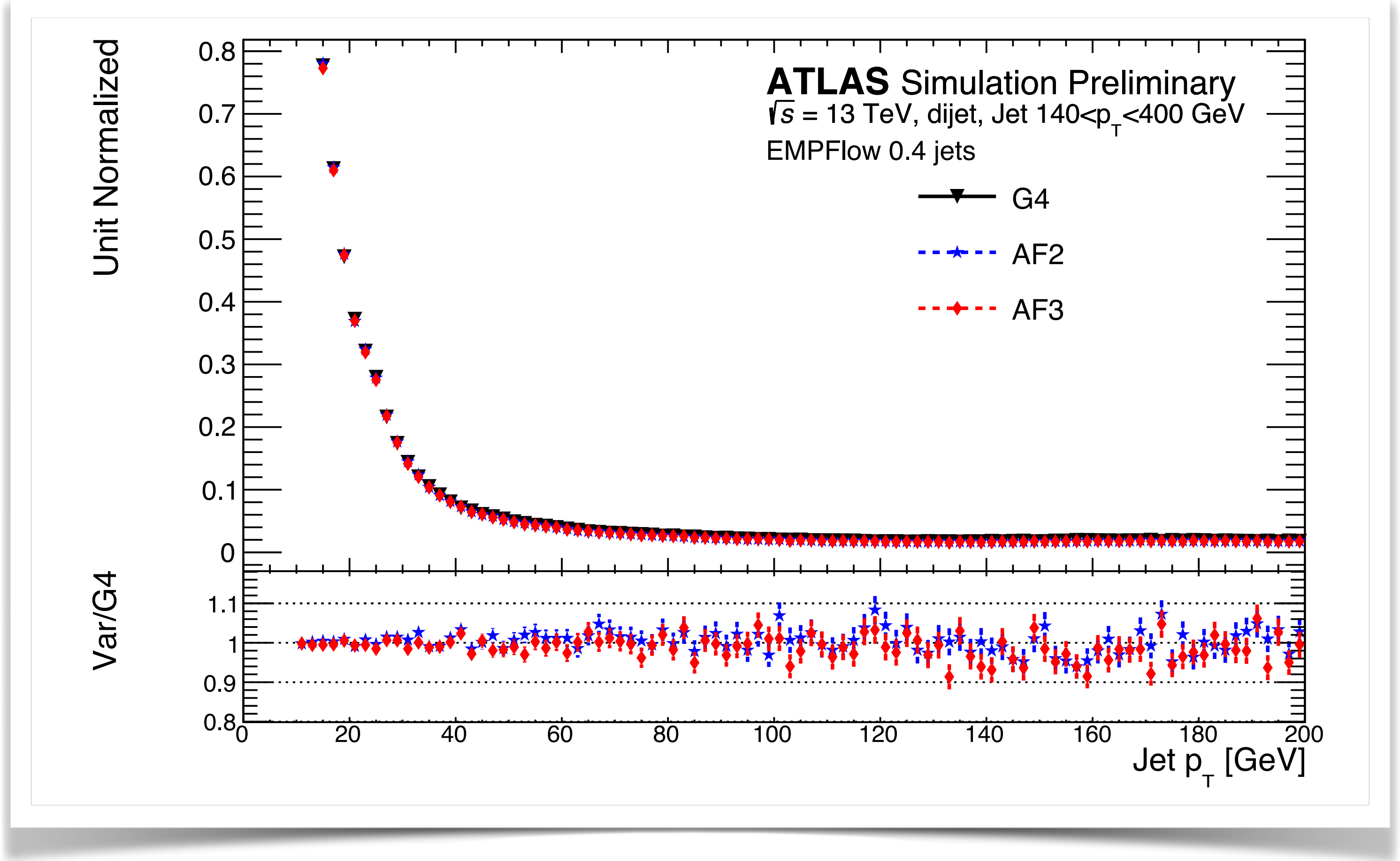


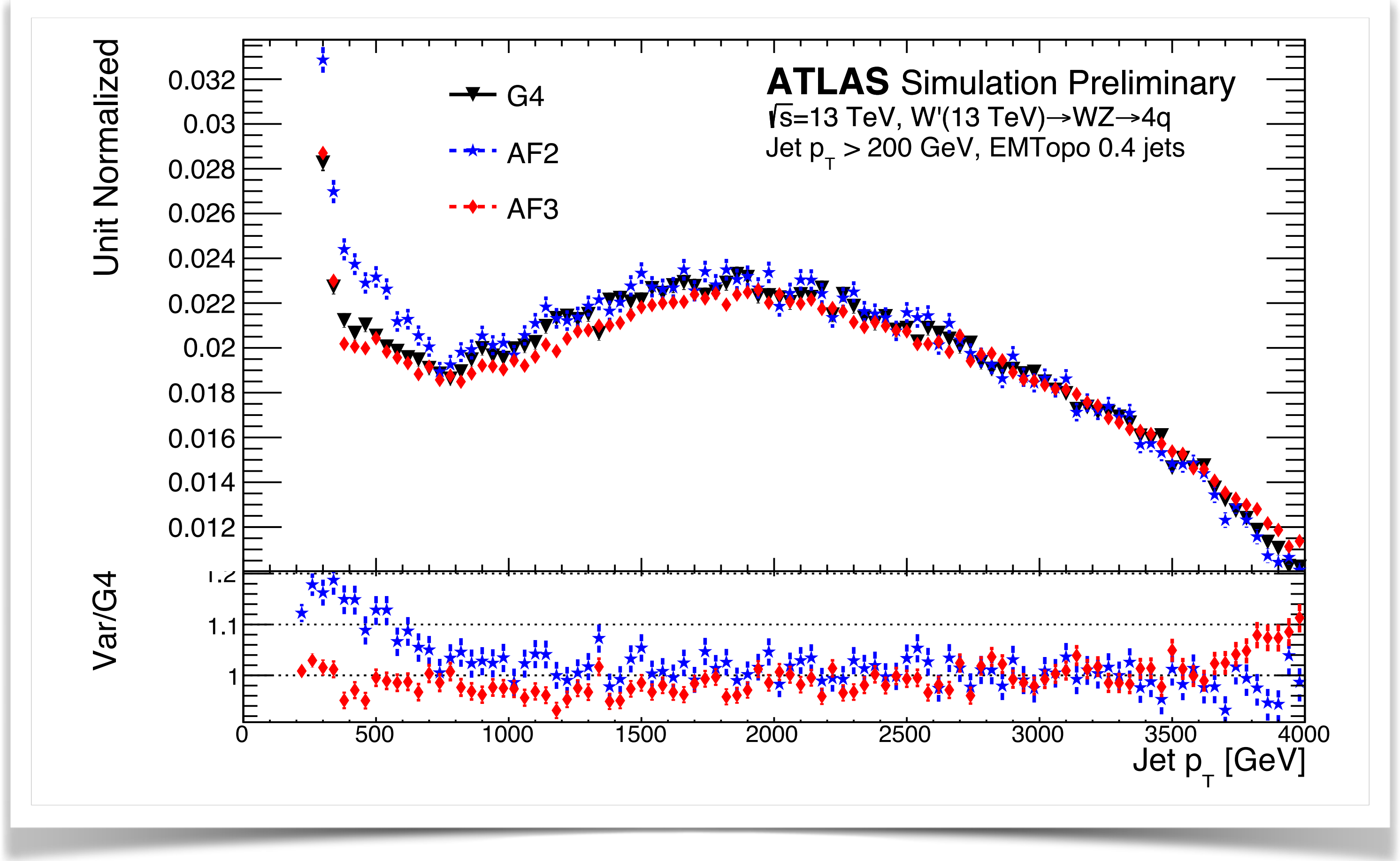
Invariant mass of Higgs (left) and energy fraction in sampling 3 of the EM calorimeter (right) in H→γγ events

Very good modeling for all electron/photon variables!

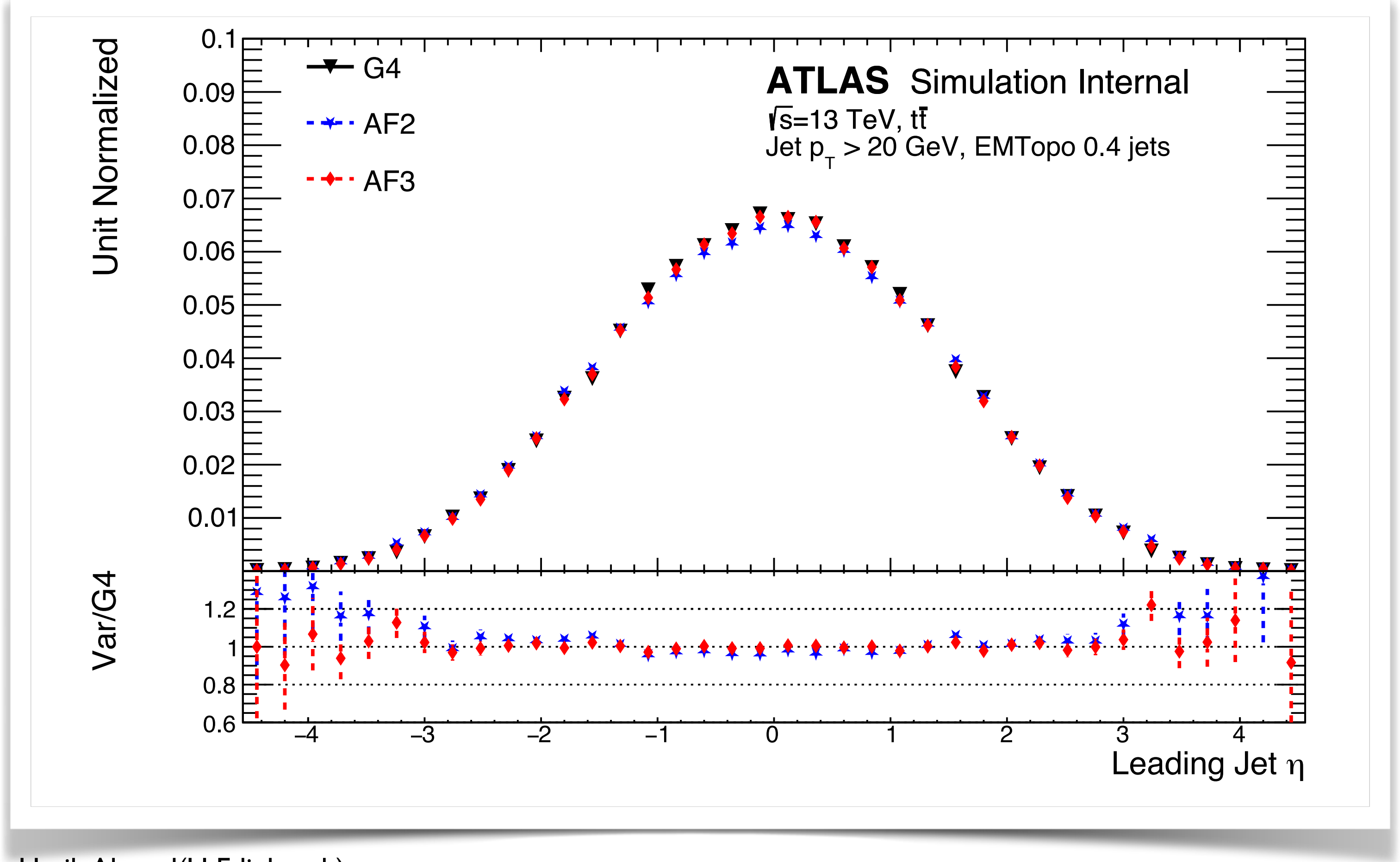
Performance of AF3: kinematics of reconstructed jet

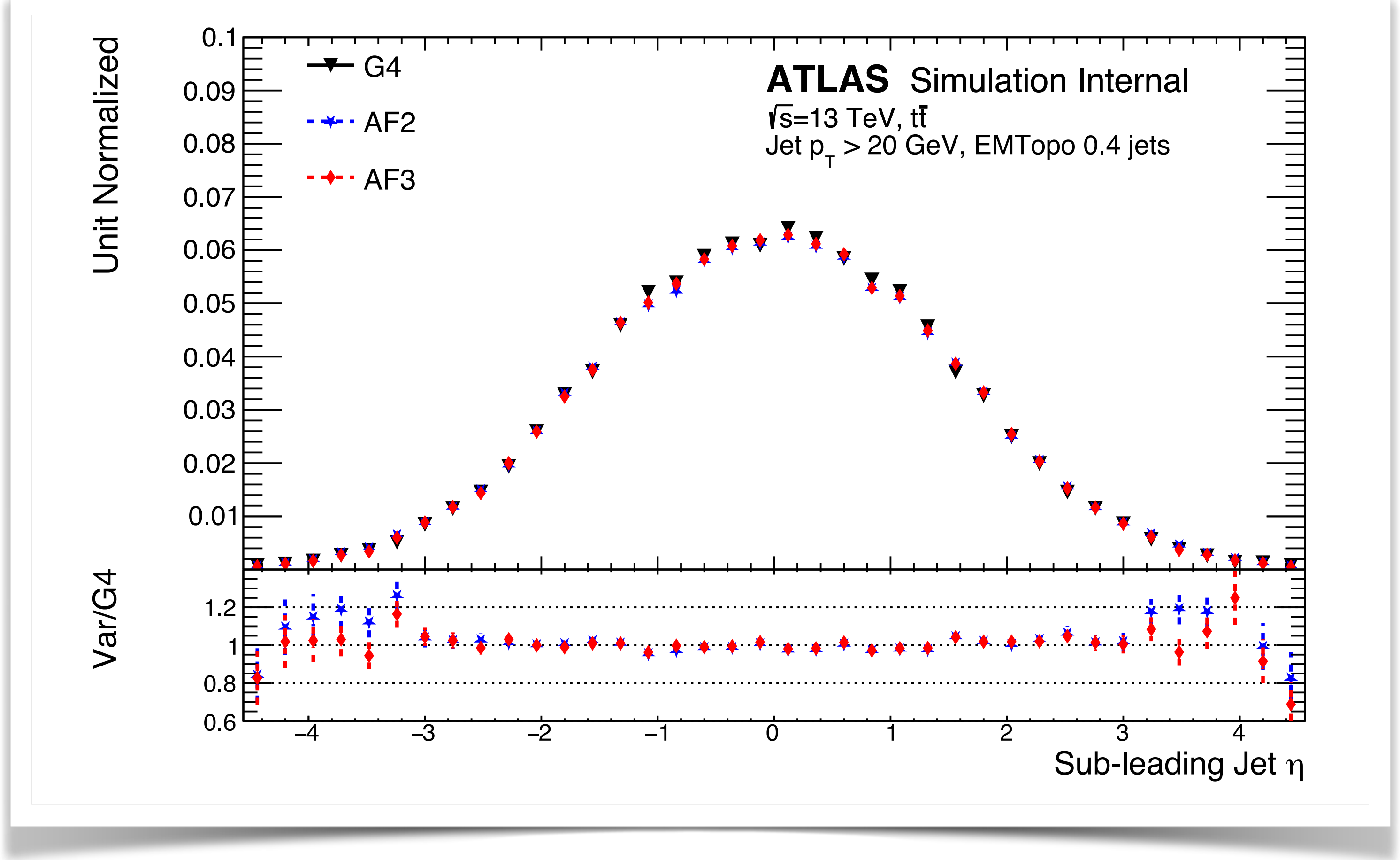
- Good modeling of jet kinematics for jets of cone 0.4 reconstructed with EMPFlow or EMTopo algorithms
- Jets with pT > 200 GeV shows better agreement in AF3 compared to AF2
- Dedicated parametrization in forward calorimeter also improves the modeling for |η| > 3





all jets pT distribution - (left) for pT < 200 GeV in dijet and (right) for pT > 200 GeV in W′(13TeV) → W Z → 4q

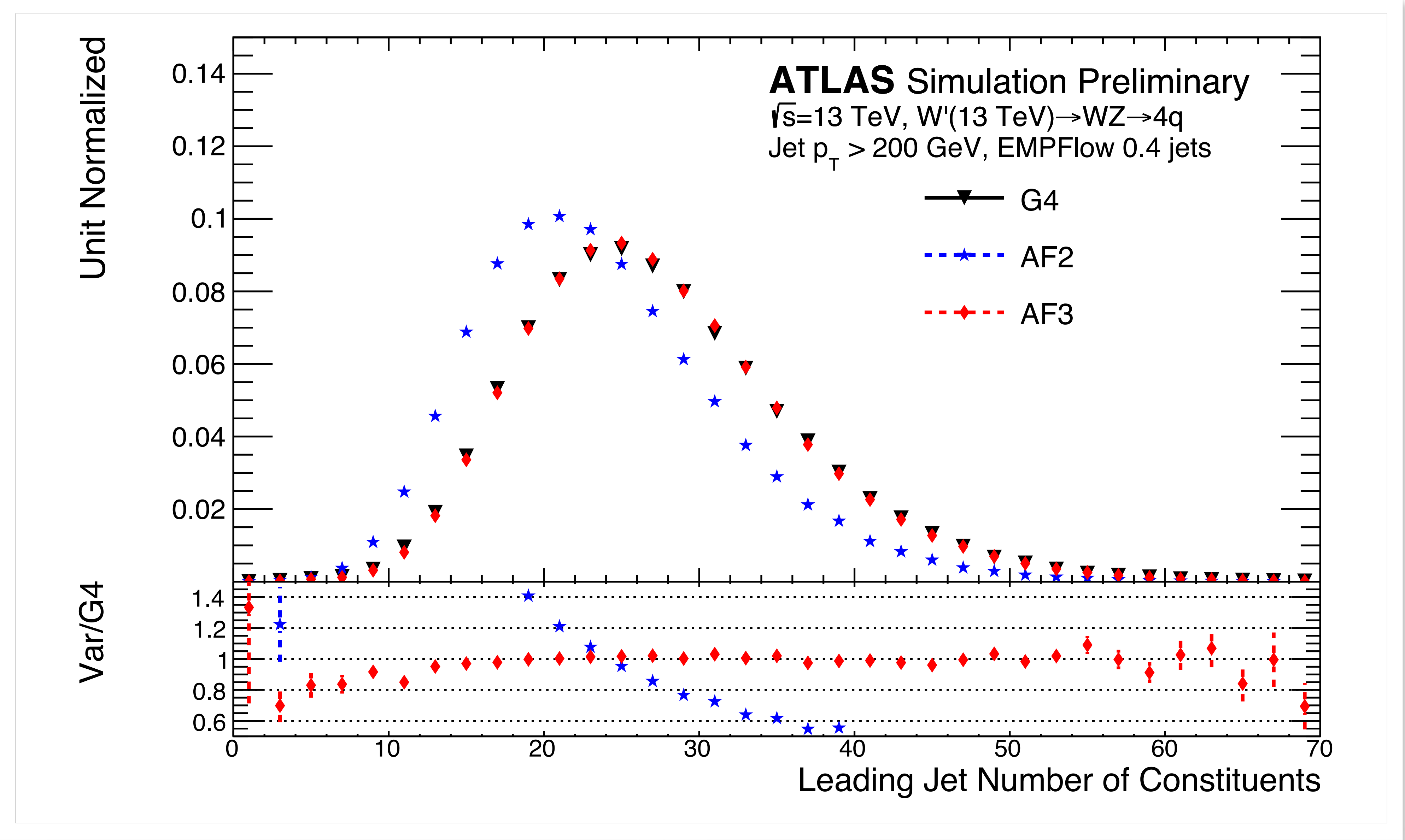


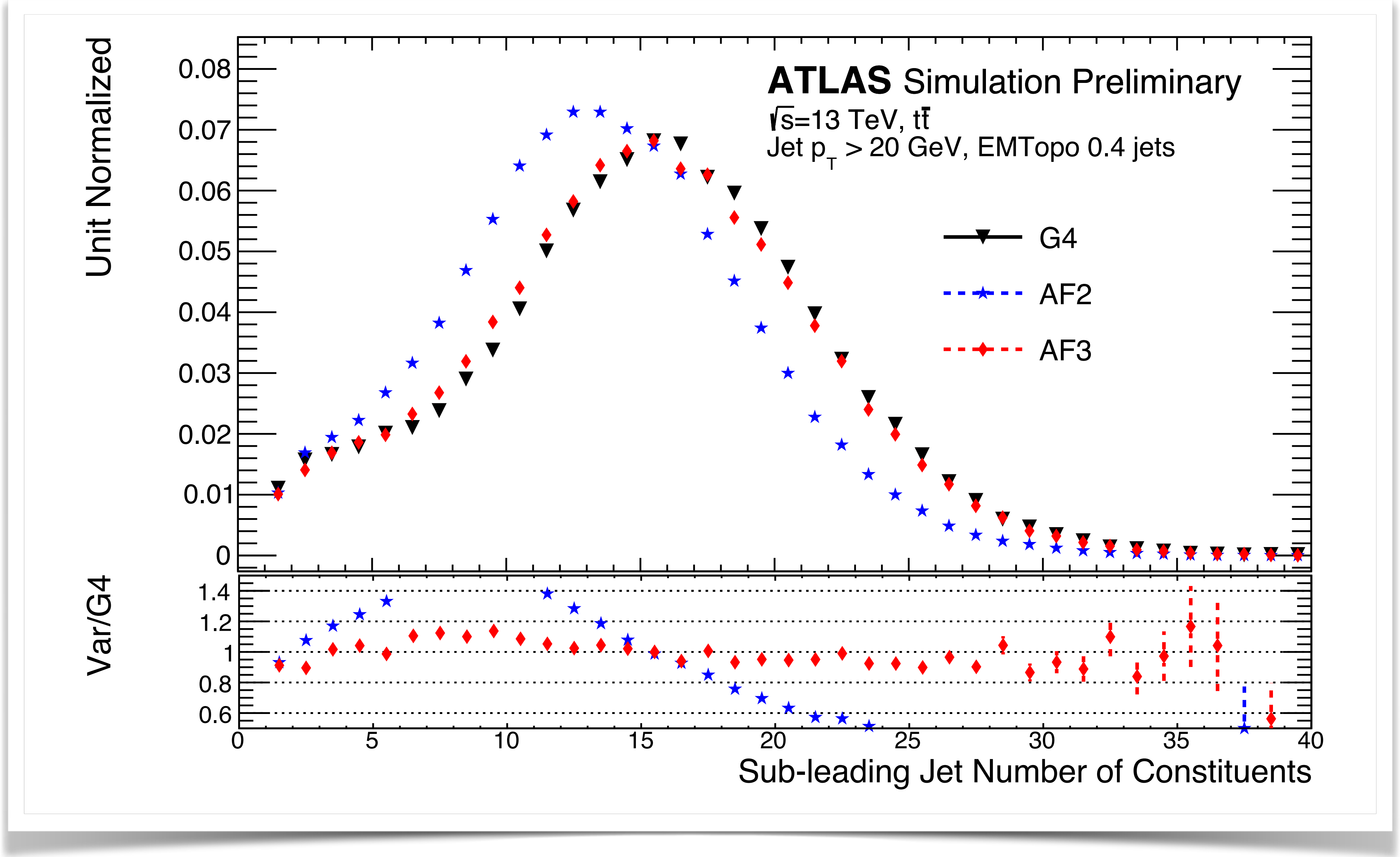


jet η distribution - (left) for leading and (right) sub-leading jet in ttbar

Jet performance: small-R jet number of clusters

- Number of constituents inside a jet of cone 0.4 for leading (pT > 200 GeV) and sub-leading (pT > 20 GeV)
- Jets reconstructed with EMPFlow algorithm



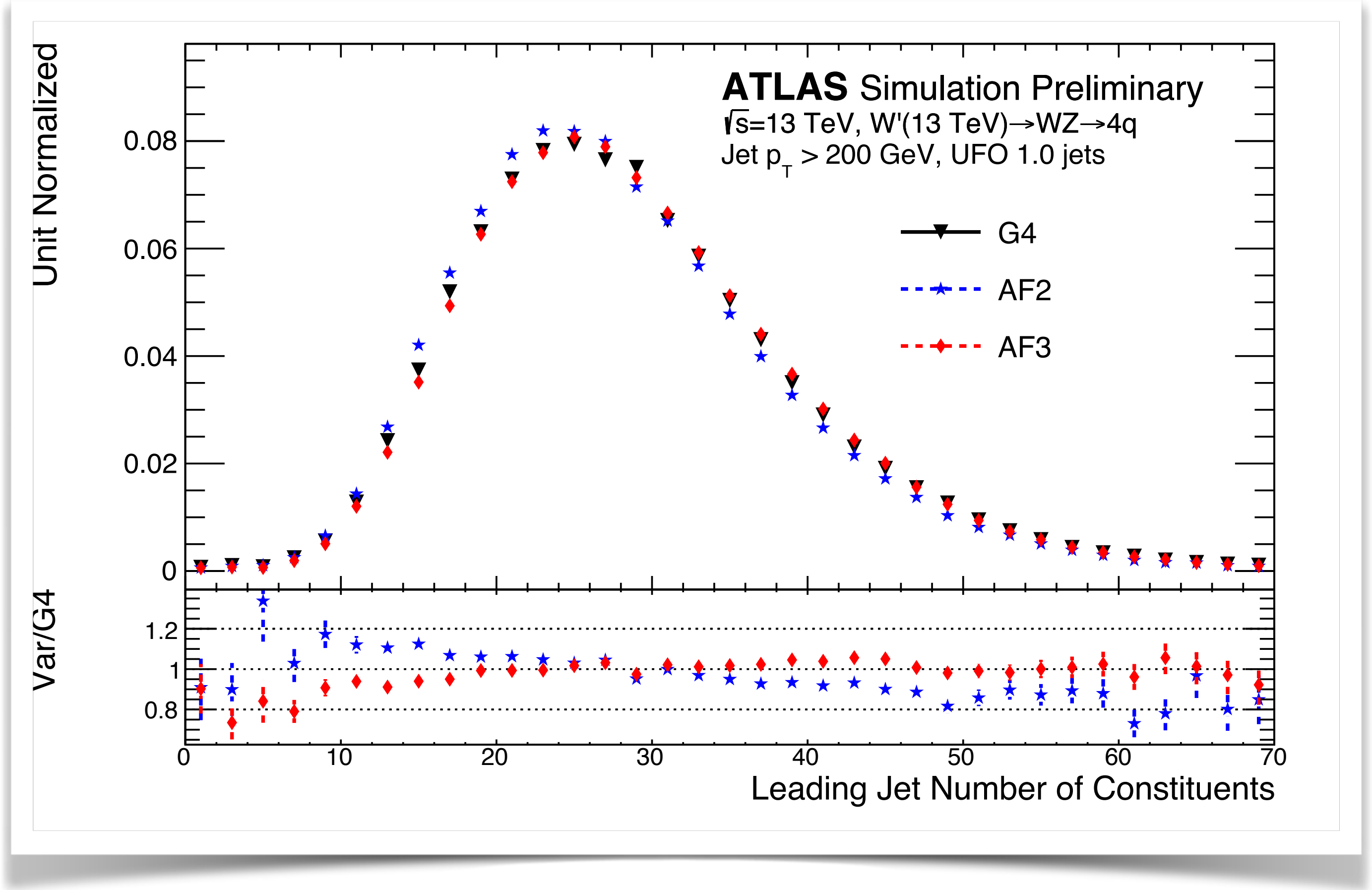


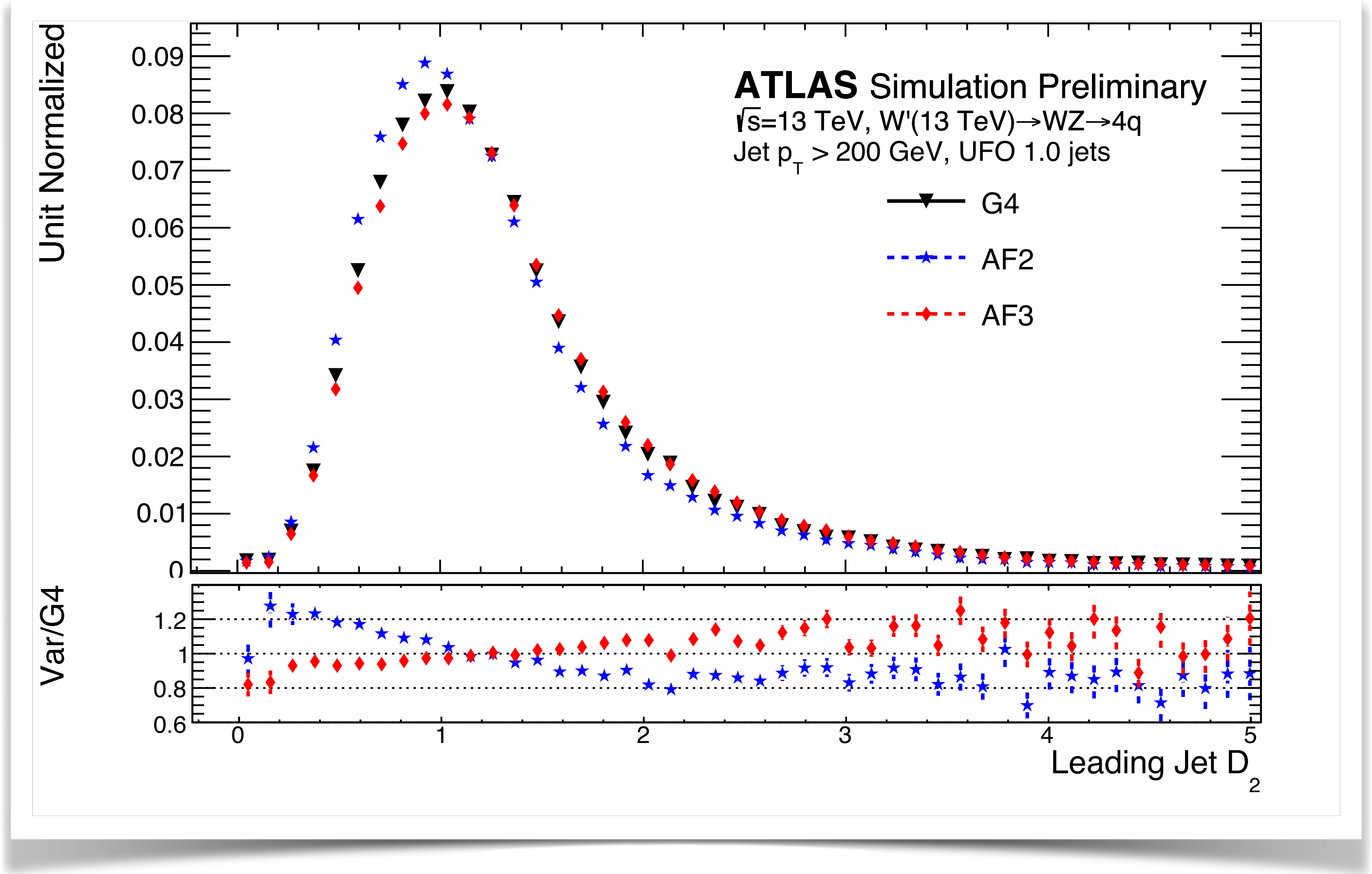
Number of constituents - (left) leading jet in W′(13TeV) → W Z → 4q and (right) sub-leading jet in ttbar events

Significant improvement over AF2 leading to improvements in other observables!

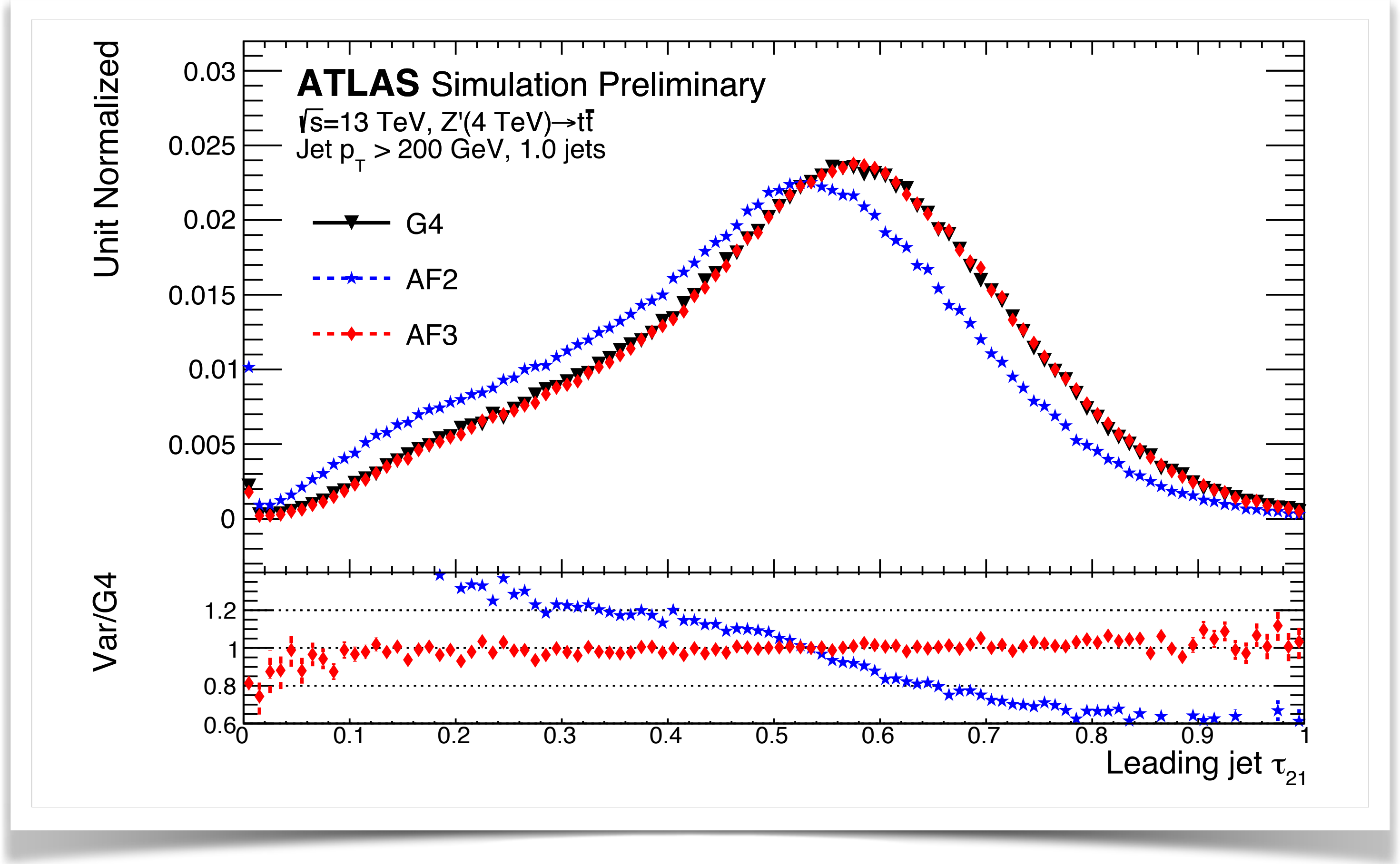
Jet performance: large-R jet substructures

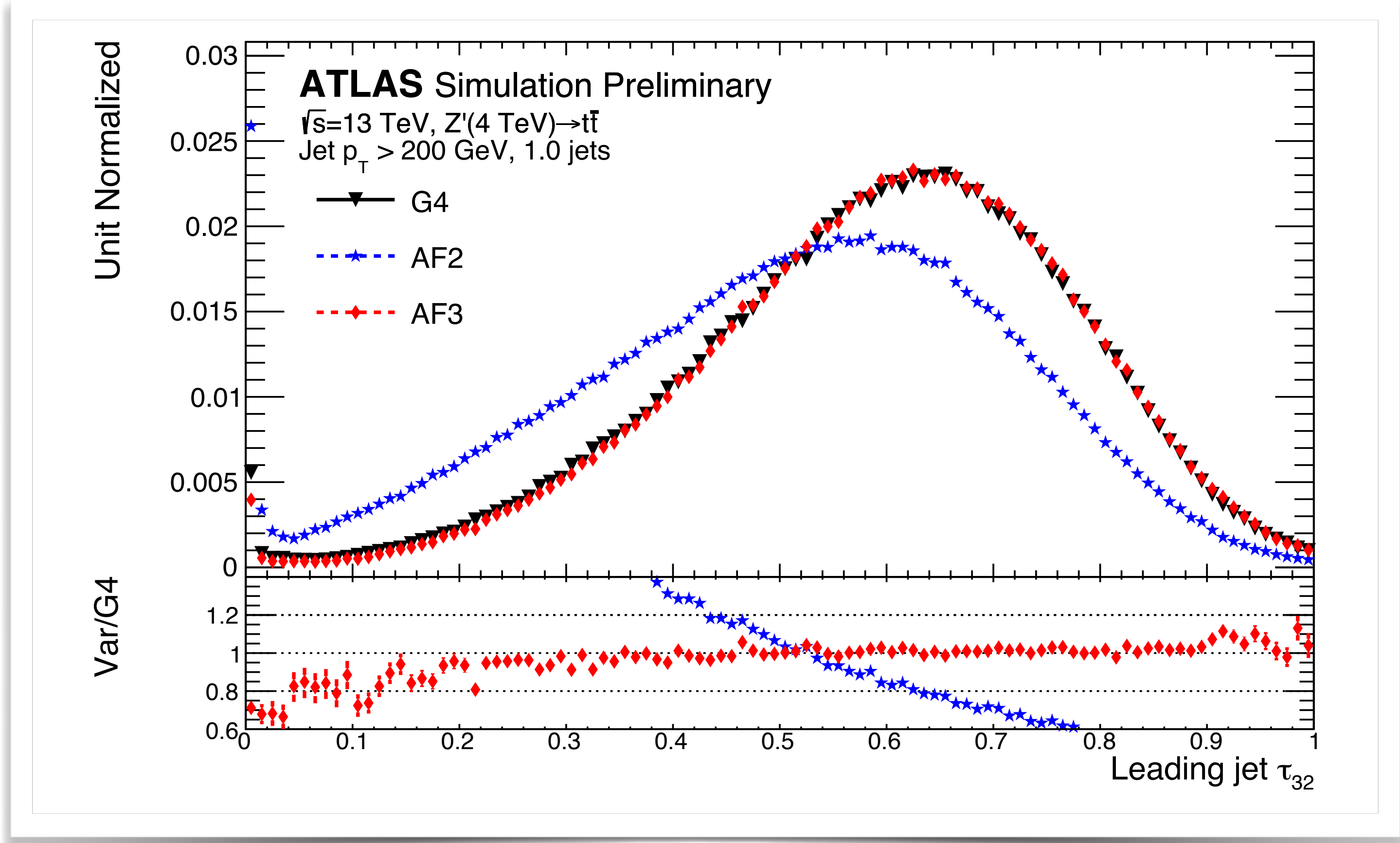
- Jet substructure variables for high energetic jets inside a cone of 1.0
- Reconstructed with trimmed UFO or LCTopo algorithm





Number of constituents - (left) and dipolarity (right) in $W'(13TeV) \rightarrow W Z \rightarrow 4q$ events



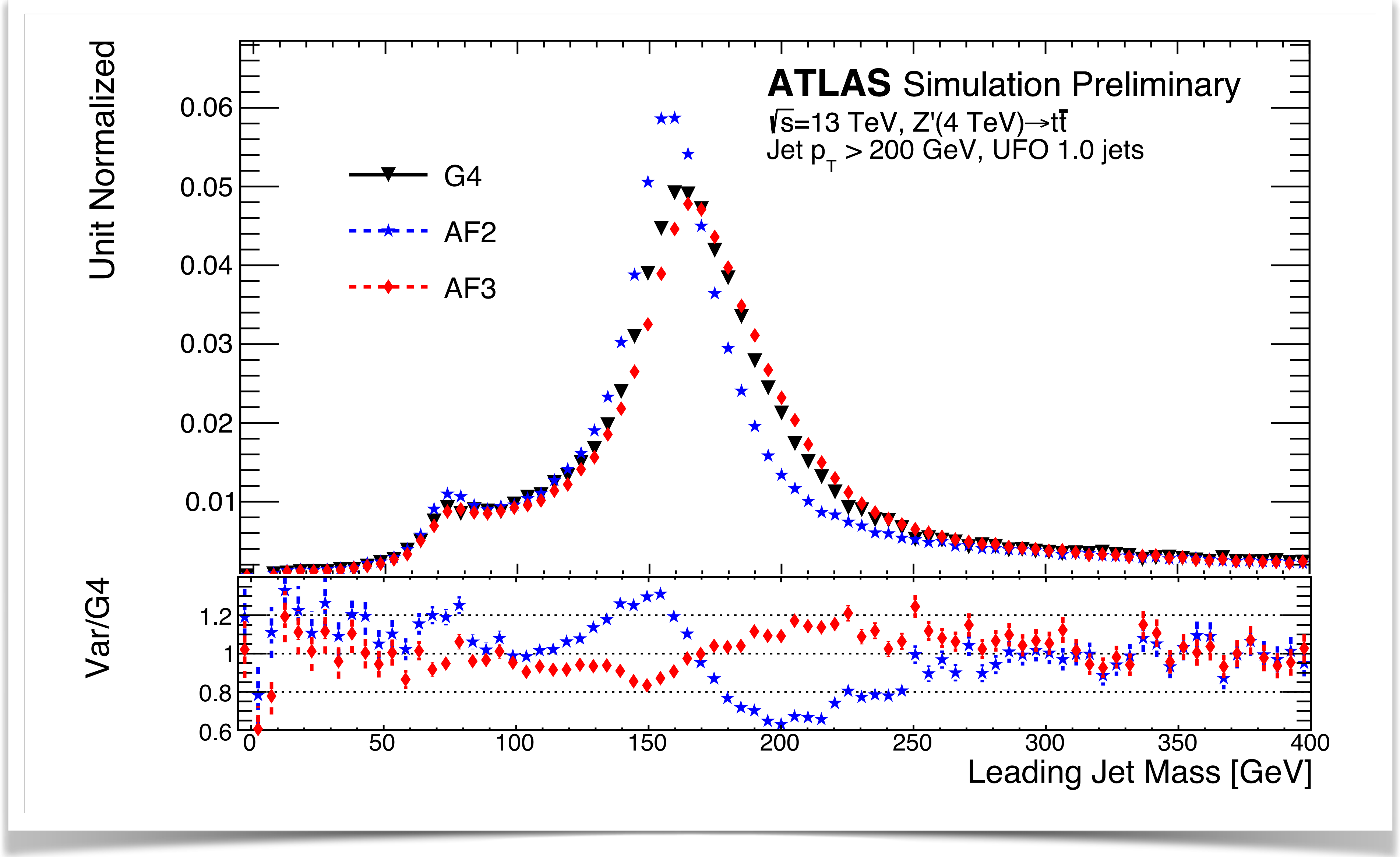


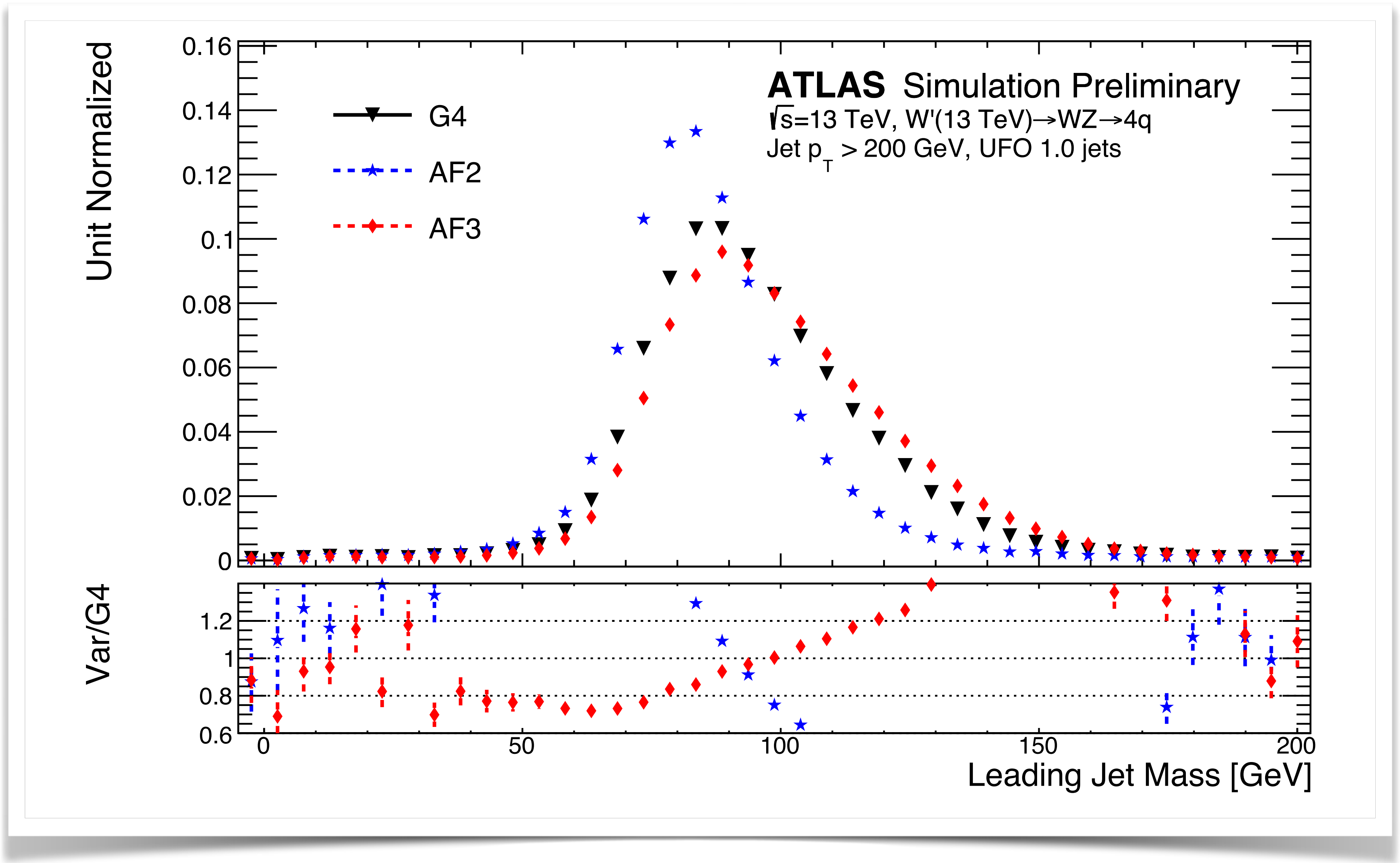
sub-jetiness variables - τ_{21}, (left) and τ_{32} (right) in $Z'(4TeV) \rightarrow t\bar{t}$ events

Improvements of these variables in AF3 over AF2 will allow more analyses to use fast simulation!

Jet performance: scope for improvement

- High energetic jets inside a cone of 1.0 reconstructed with trimmed UFO algorithm
- AF3 shows some discrepancy for jet mass - although improves upon AF2
- The discrepancies are in the tails of the high energetic jets



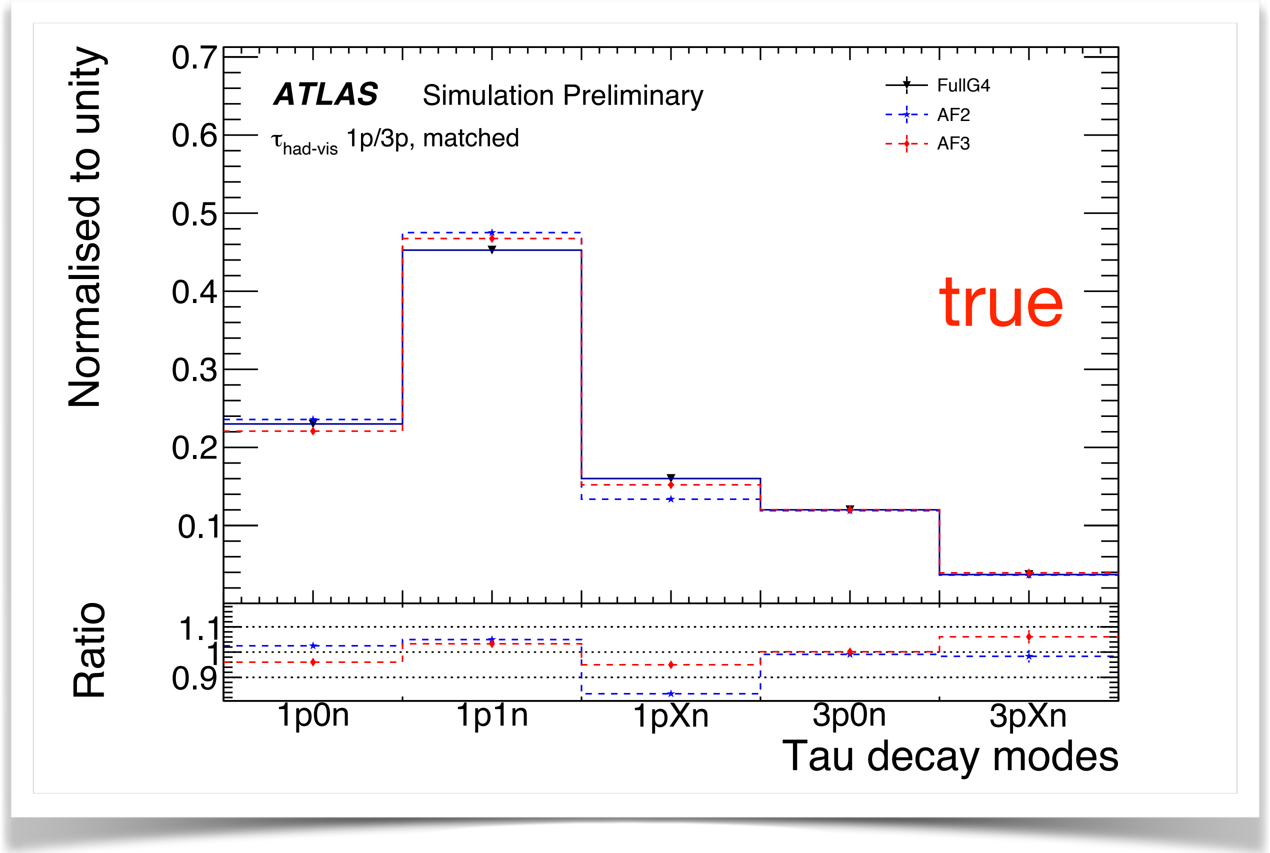


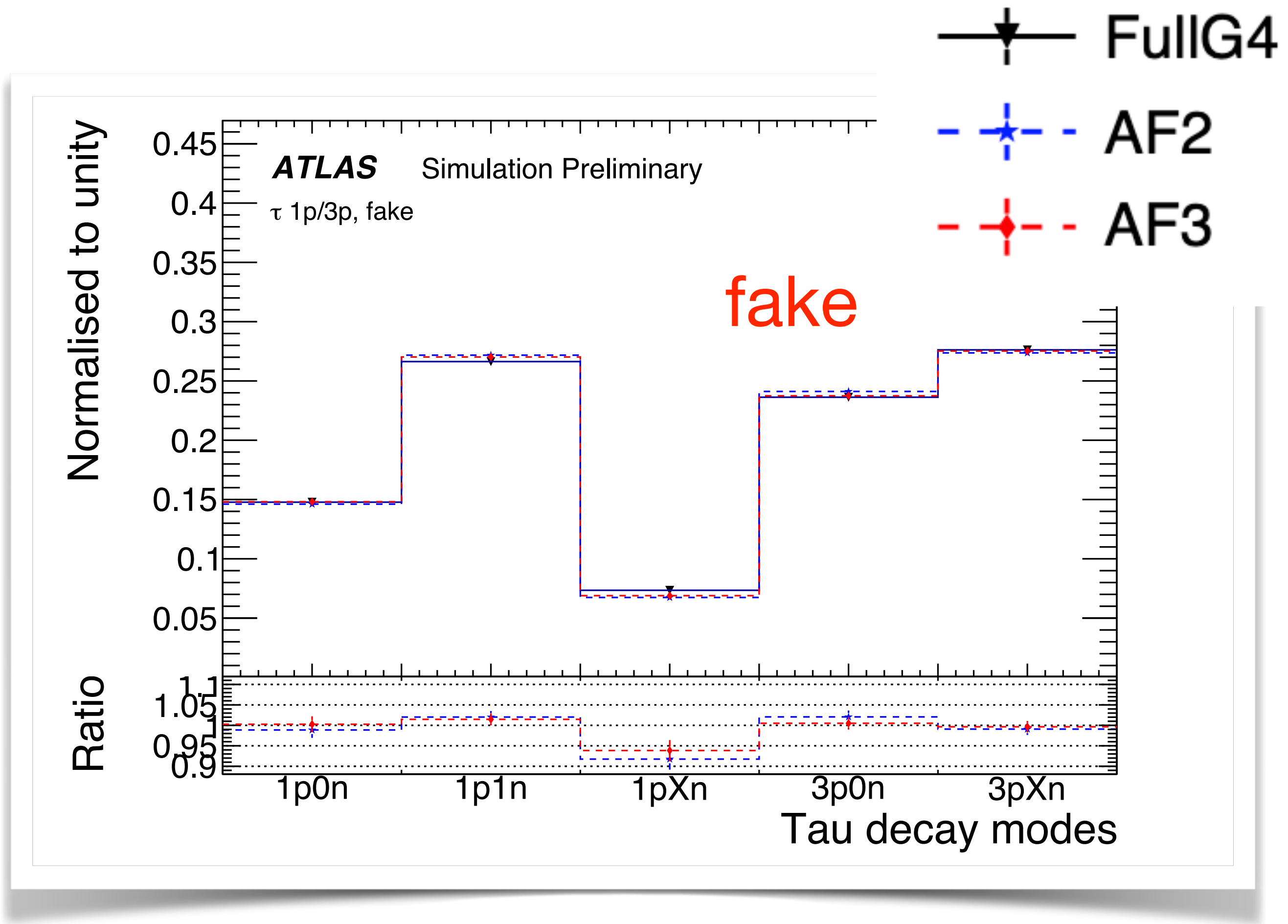
leading jet mass - (left) in $Z'(4TeV) \rightarrow t\bar{t}$ and (right) in $W'(13TeV) \rightarrow W Z \rightarrow 4q$ events

Most physics analyses are not affected!

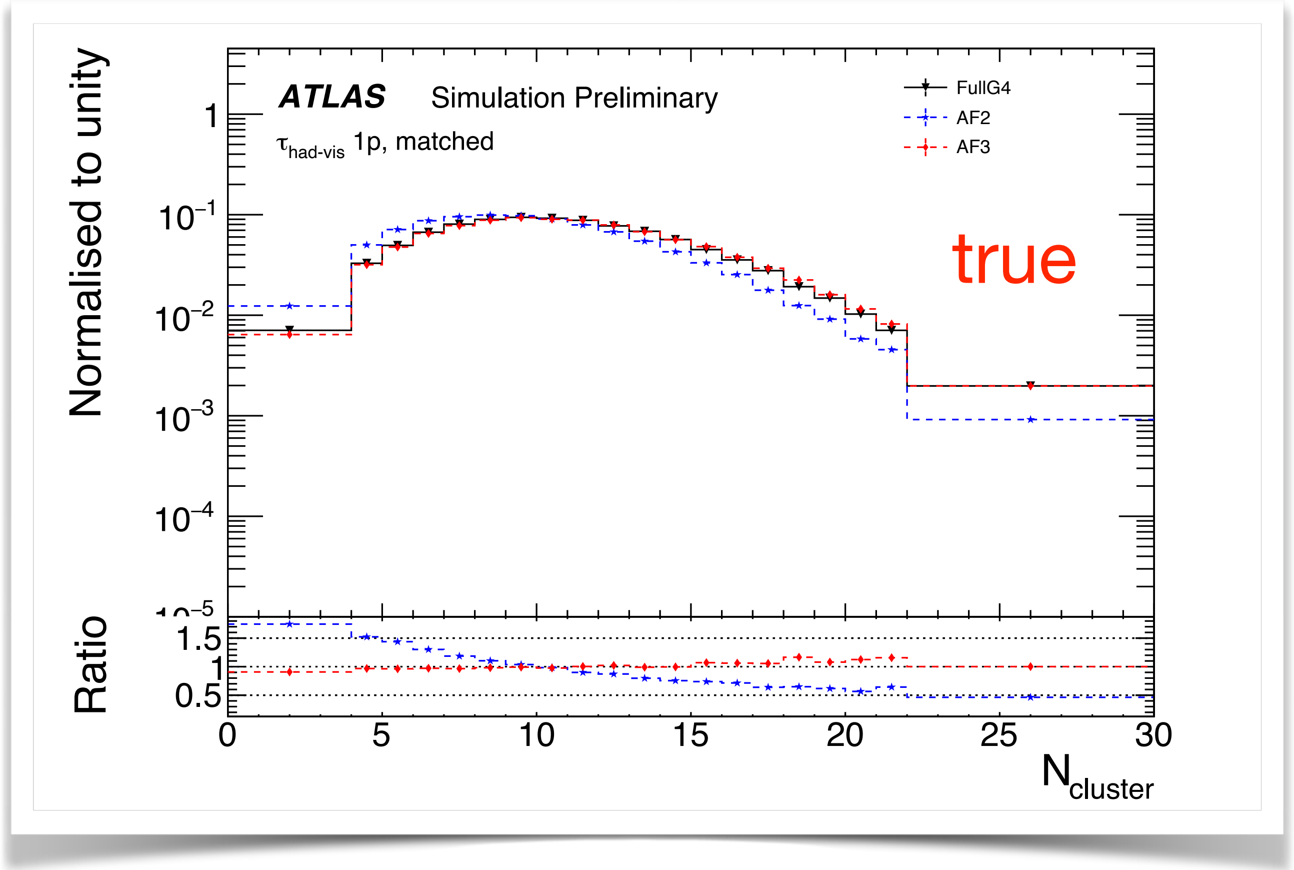
Performance of AF3: reconstructed hadronic taus

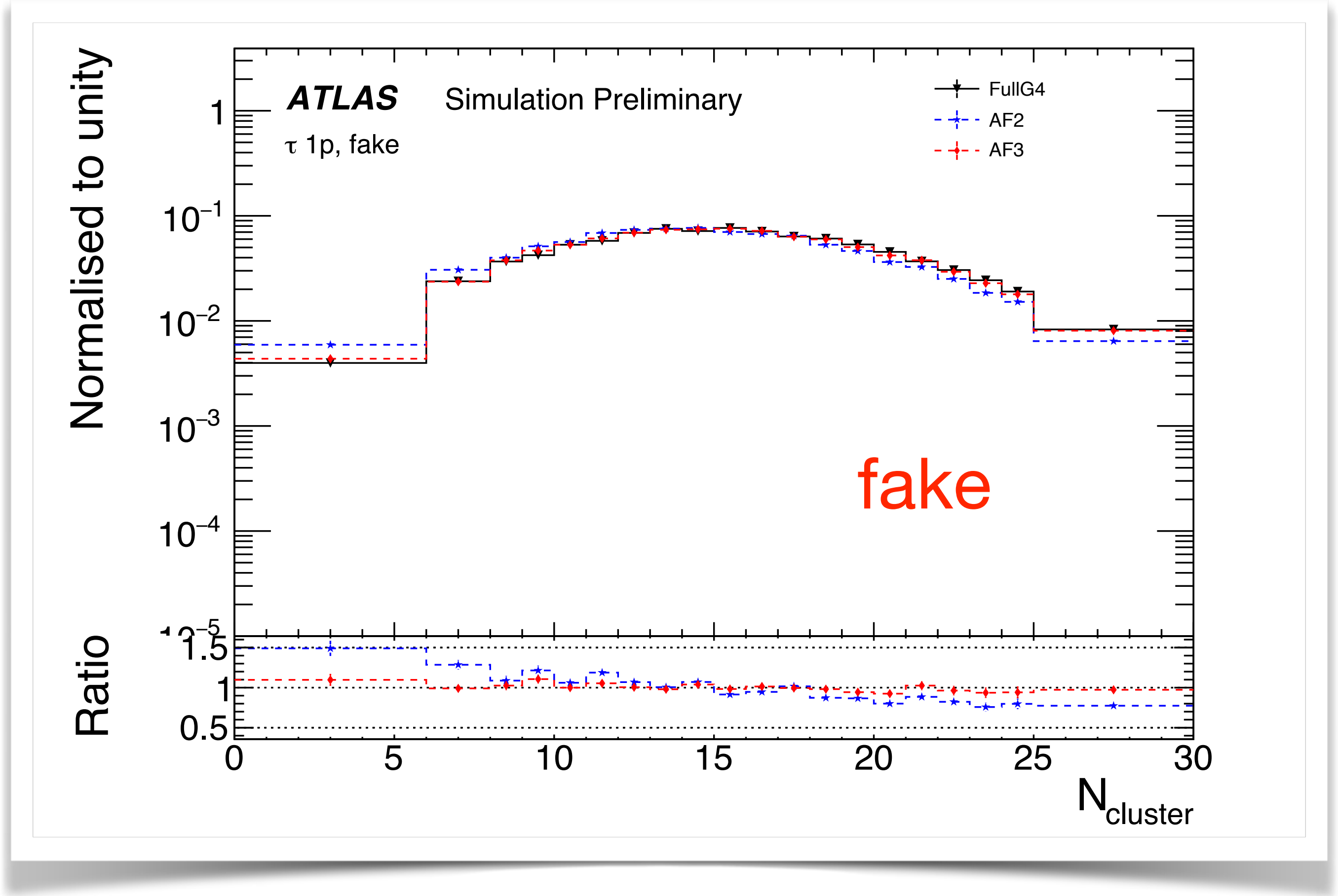
- Hadronically decaying τ-lepton is reconstructed using BDT algorithm and matched to truth taus





tau decay to one or two charged particles (1p or 3p) and multiple neutral particles (0n, 1n, Xn, etc.) for (left) true taus and (right) fake taus in a $Z \rightarrow \tau\tau$ Drell-Yan sample with a mass of 2.0-2.5 TeV.





number of clusters for (left) true taus and (right) fake taus in a $Z \rightarrow \tau\tau$ Drell-Yan sample with a mass of 2.0-2.5 TeV.

AF3 shows good performance for both true and fake taus !

Summary

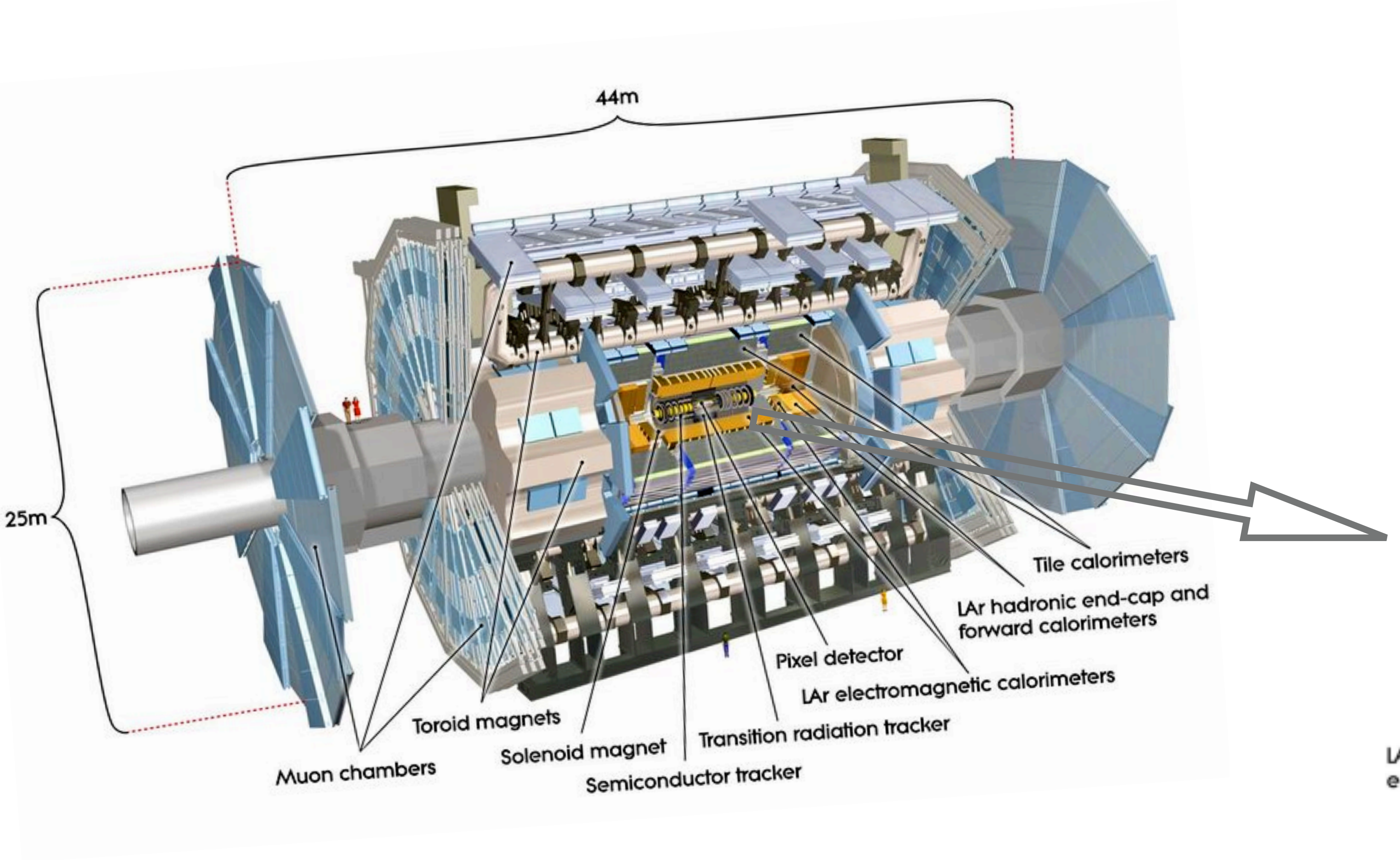
- AF3 is the next generation of fast simulation in ATLAS - successfully deploying complex parametrized and deep learning algorithms.
- AF3 achieved very good modeling for all reconstructed observables compared to Geant4 even for complex variables such as jet substructure.
- The CPU performance of AF3 is only limited by the ID simulation (Geant4), but a factor of O(10) speed up is sufficient to meet the CPU needs for Run 3.
- ATLAS will use AF3 to re-simulate ~7 billion events from Run 2.
- An update of the current AF3 version in expected for Run 3 - current performance seems sufficient to produce a large fraction of ATLAS Run 3 Monte Carlo events.

Thank you!

BACKUP

ATLAS Calorimeter and shower generation

Sampling calorimeter covering |η| < 4.9

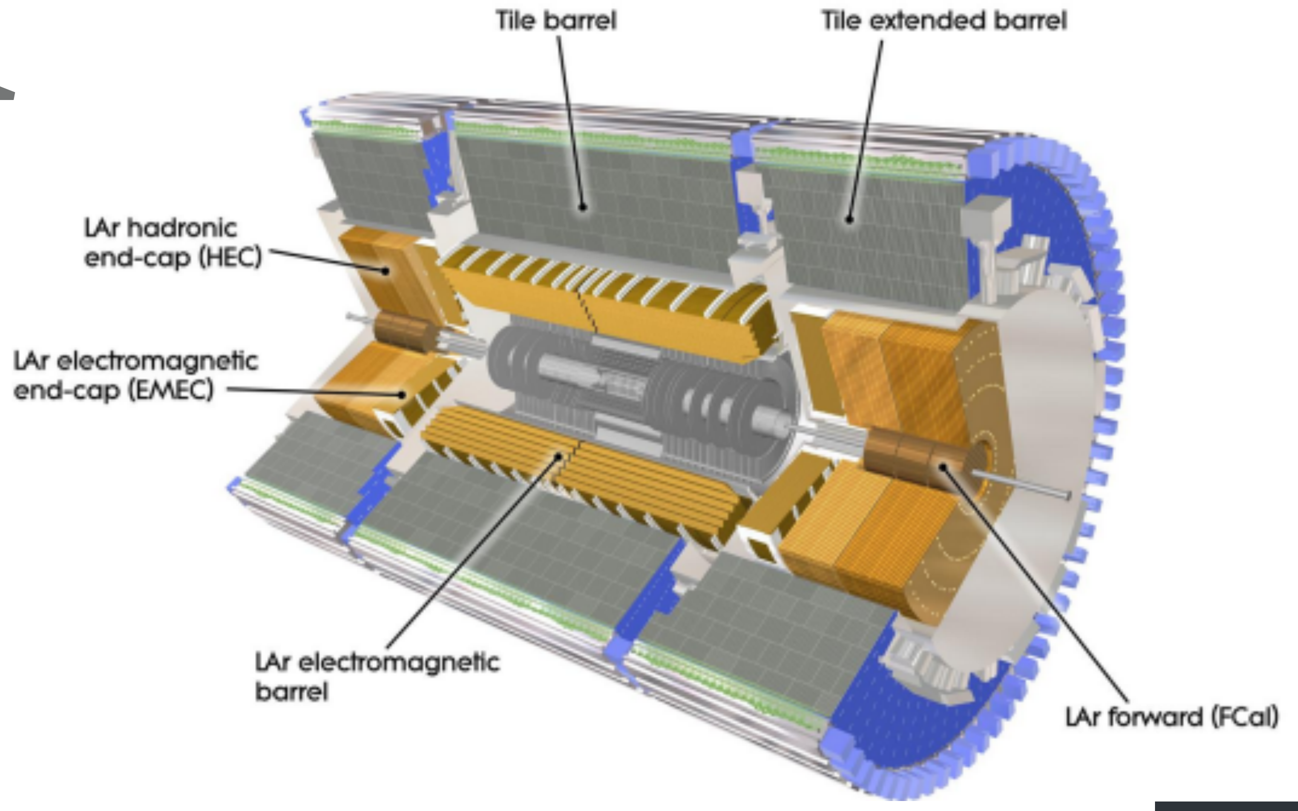


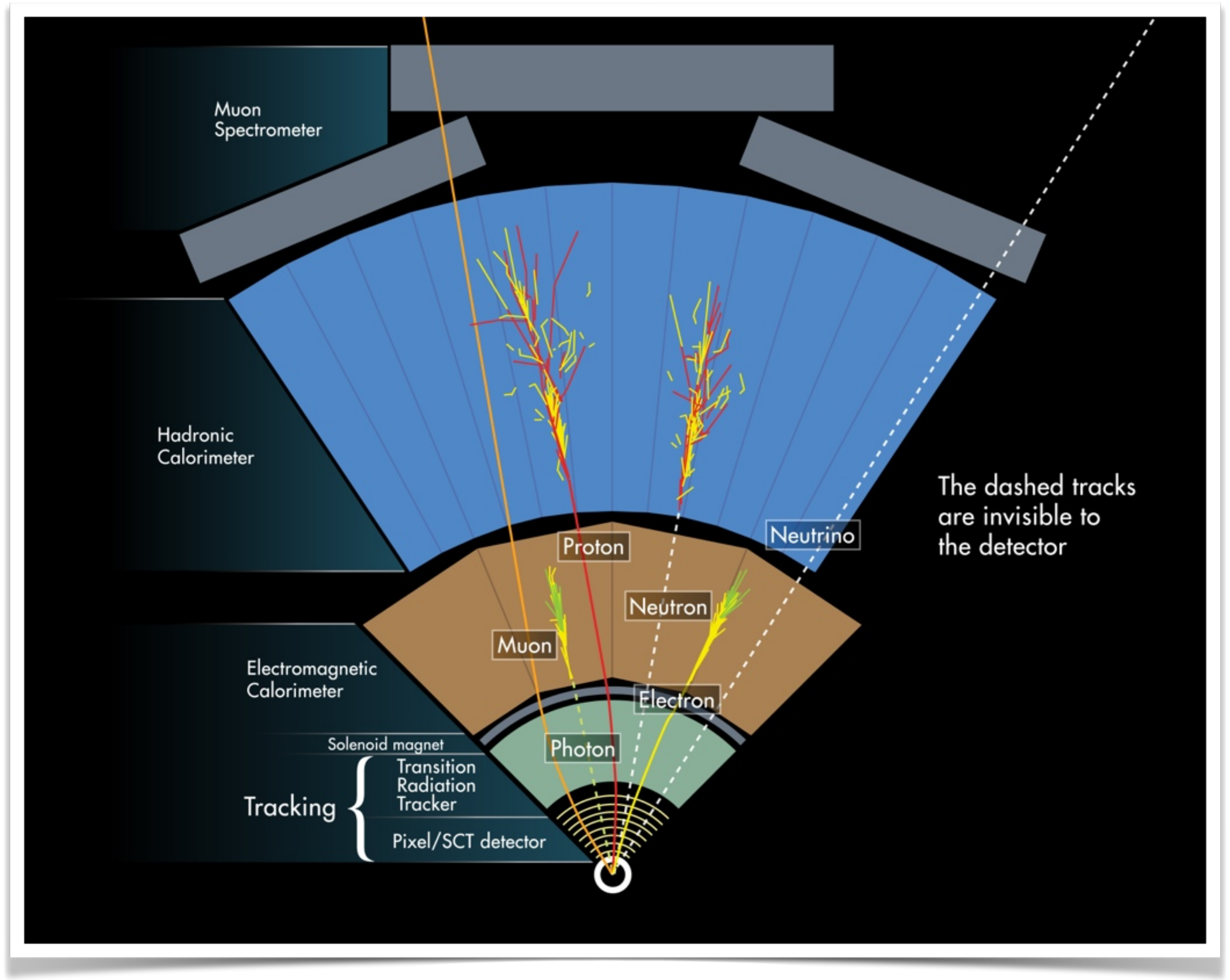
Electromagnetic (EM) Cal:

- Liquid Argon (active)
- Pb/Cu/Tungsten (absorber)

Hadronic/Tile Cal:

- Scintillating tiles (active)
- Steel (absorber)

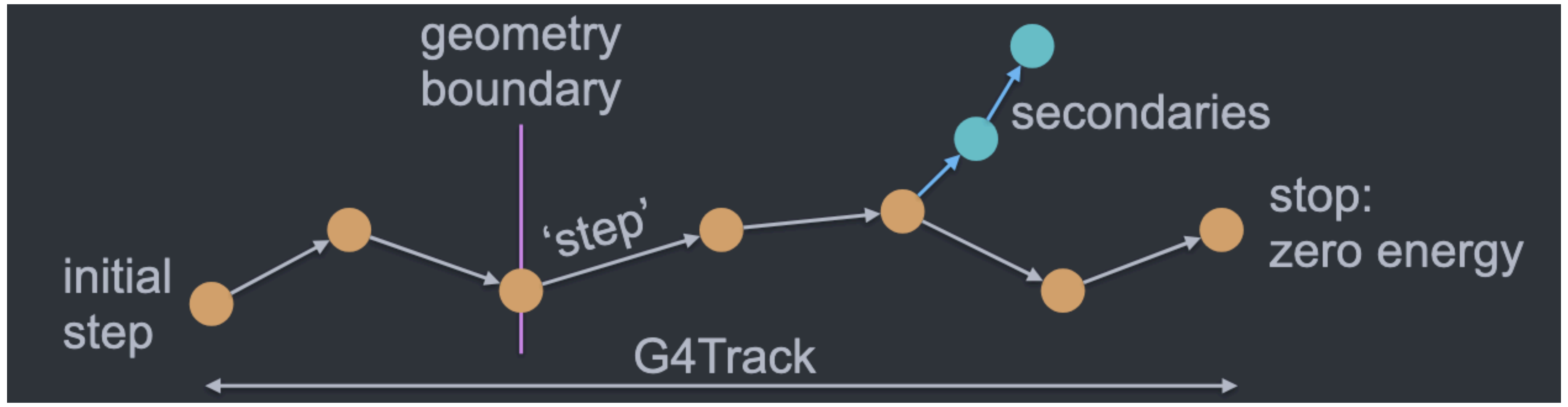




Simulation in Geant4 with each Geant4 process responsible for the smallest unit called 'step'

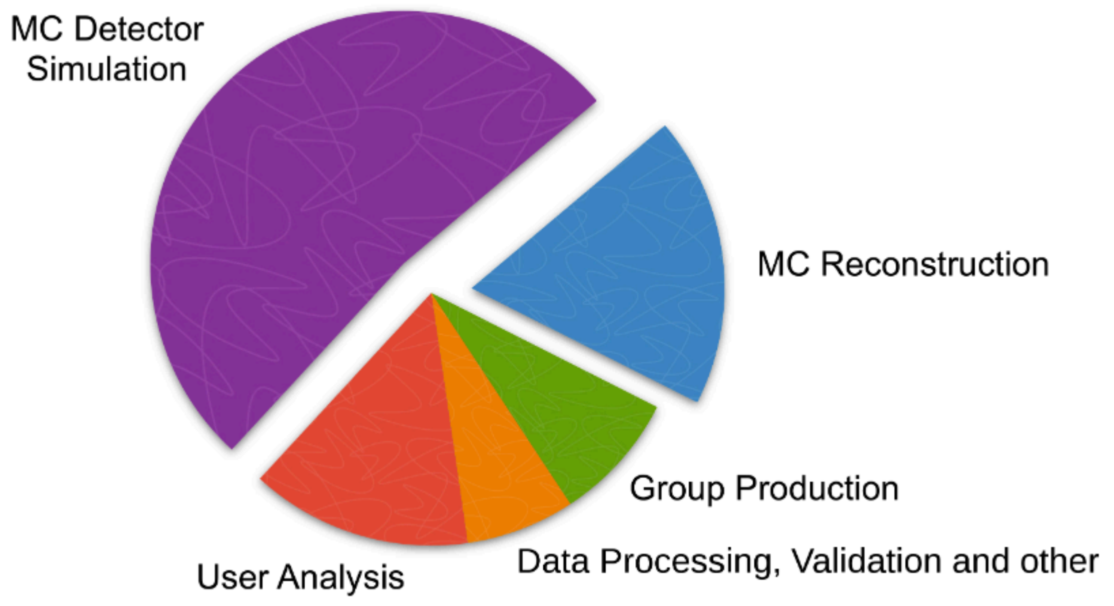
Total readout channels: ~190 k
Number of layers: 24

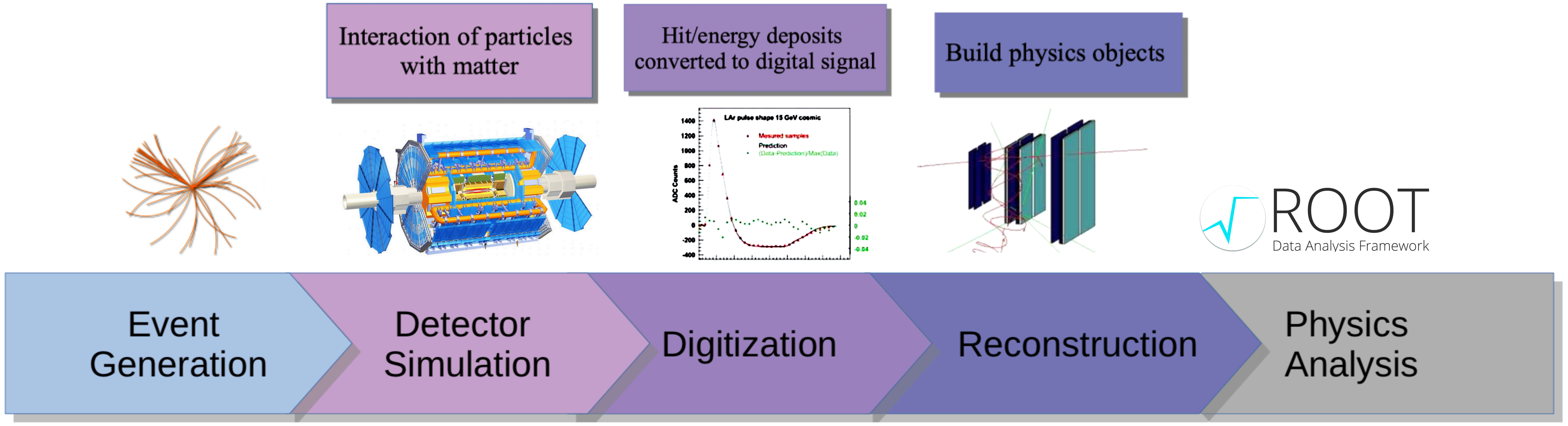
System	EM Barrel	EM EC	Hadronic EC	FCAL	Tile
#Channels	110k	64k	5.6k	3.5k	9.8k

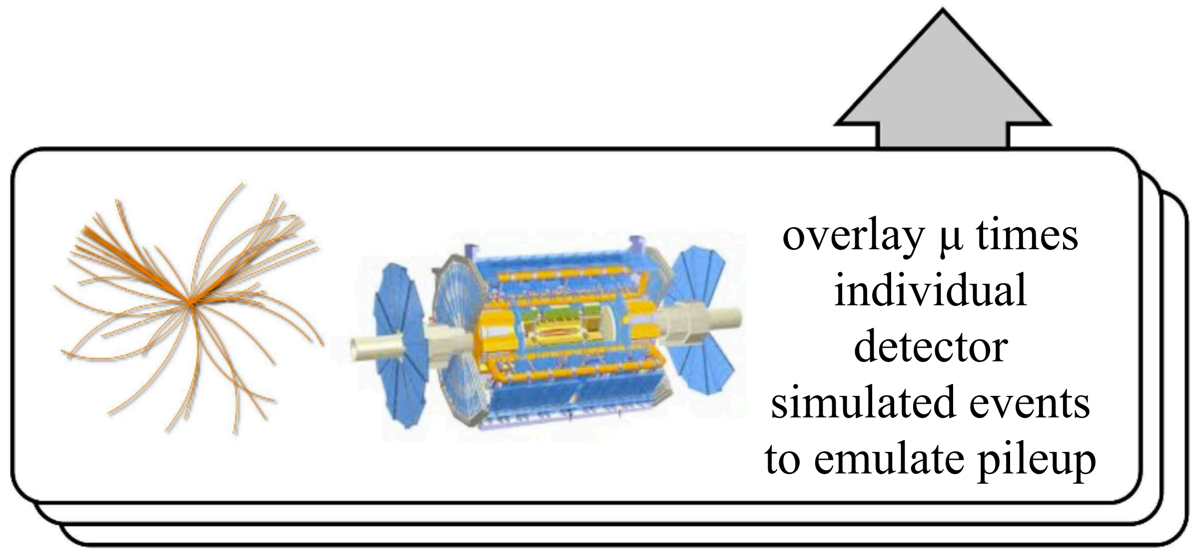


No. of steps ∝ simulation time

ATLAS Monte Carlo production steps







Detector Simulation:

- Dense hit content in inner trackers
- Larger sub-detectors means longer simulation time

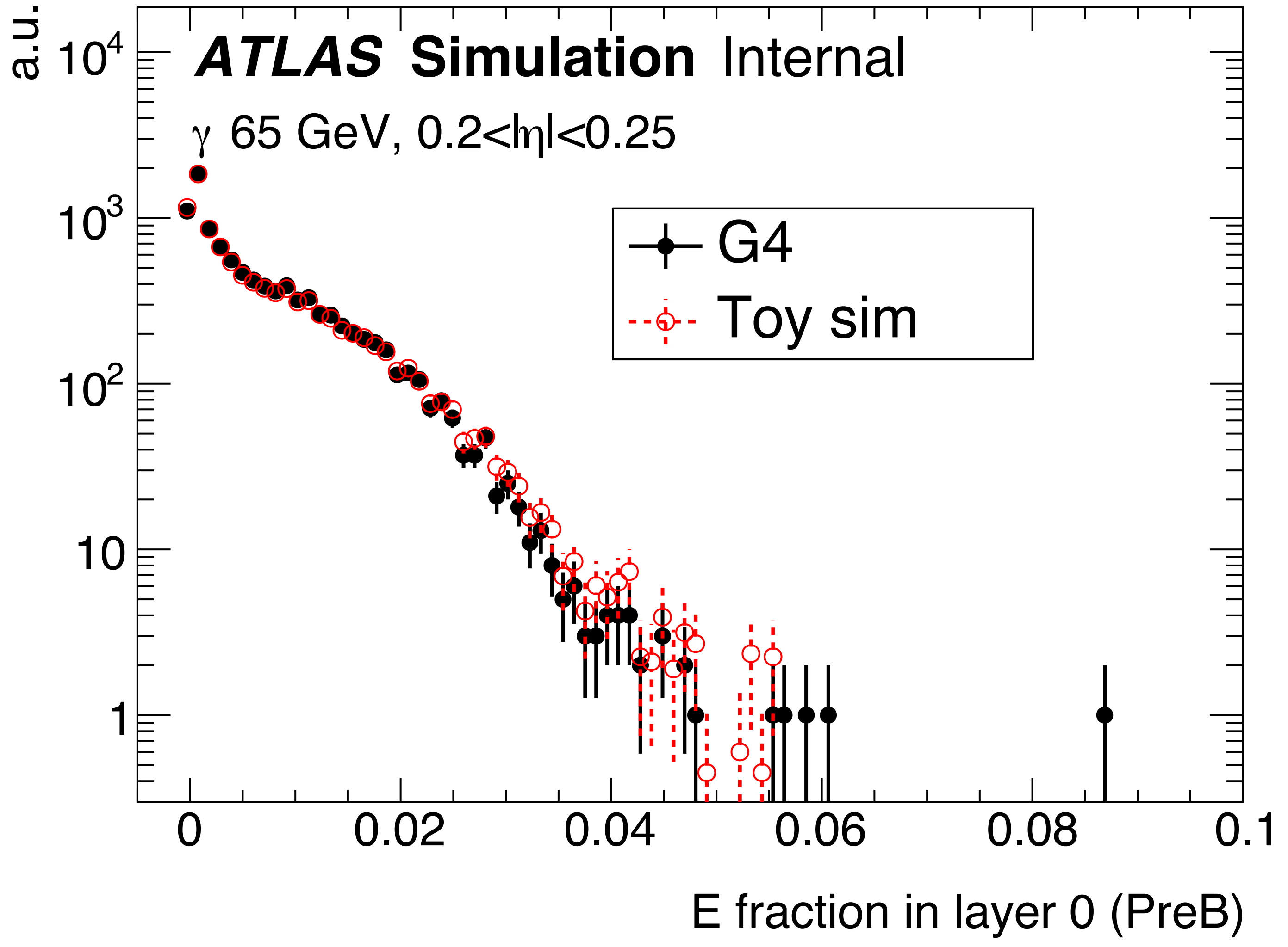
Digitization:

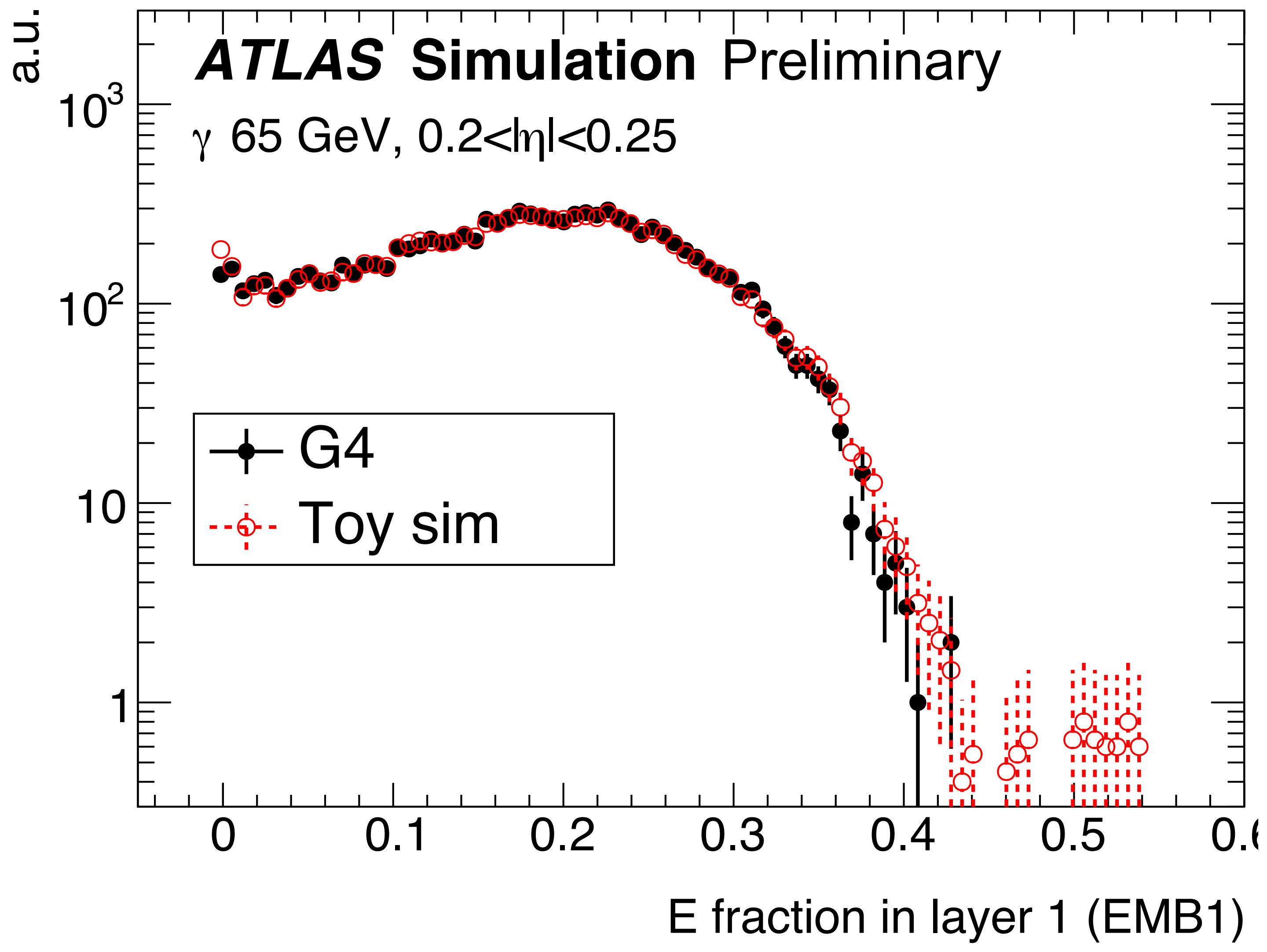
- Large number of inner tracker readout channels
- Complex modeling of readout emulation

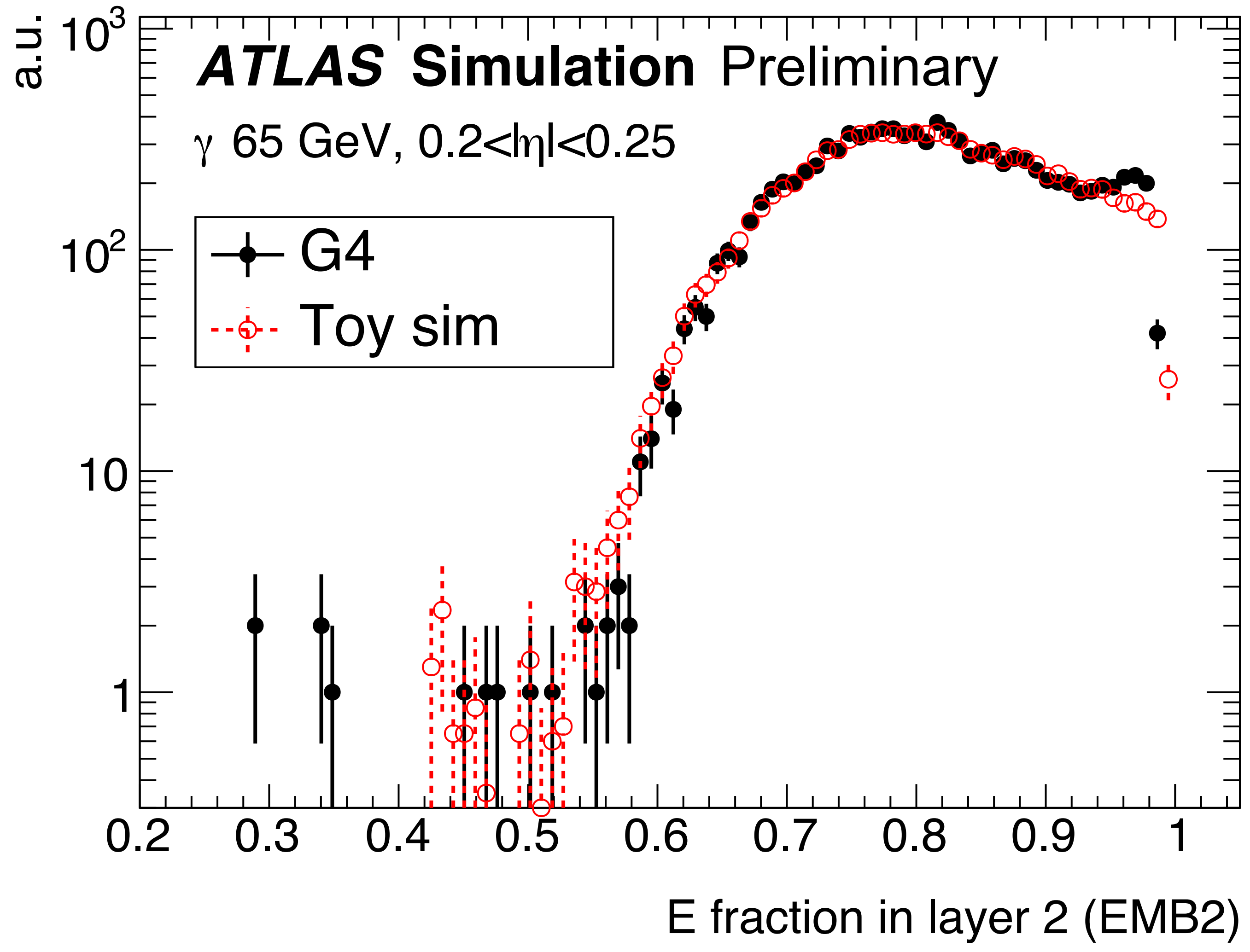
Reconstruction:

- Pattern recognition (combinatorics) function of average pileup

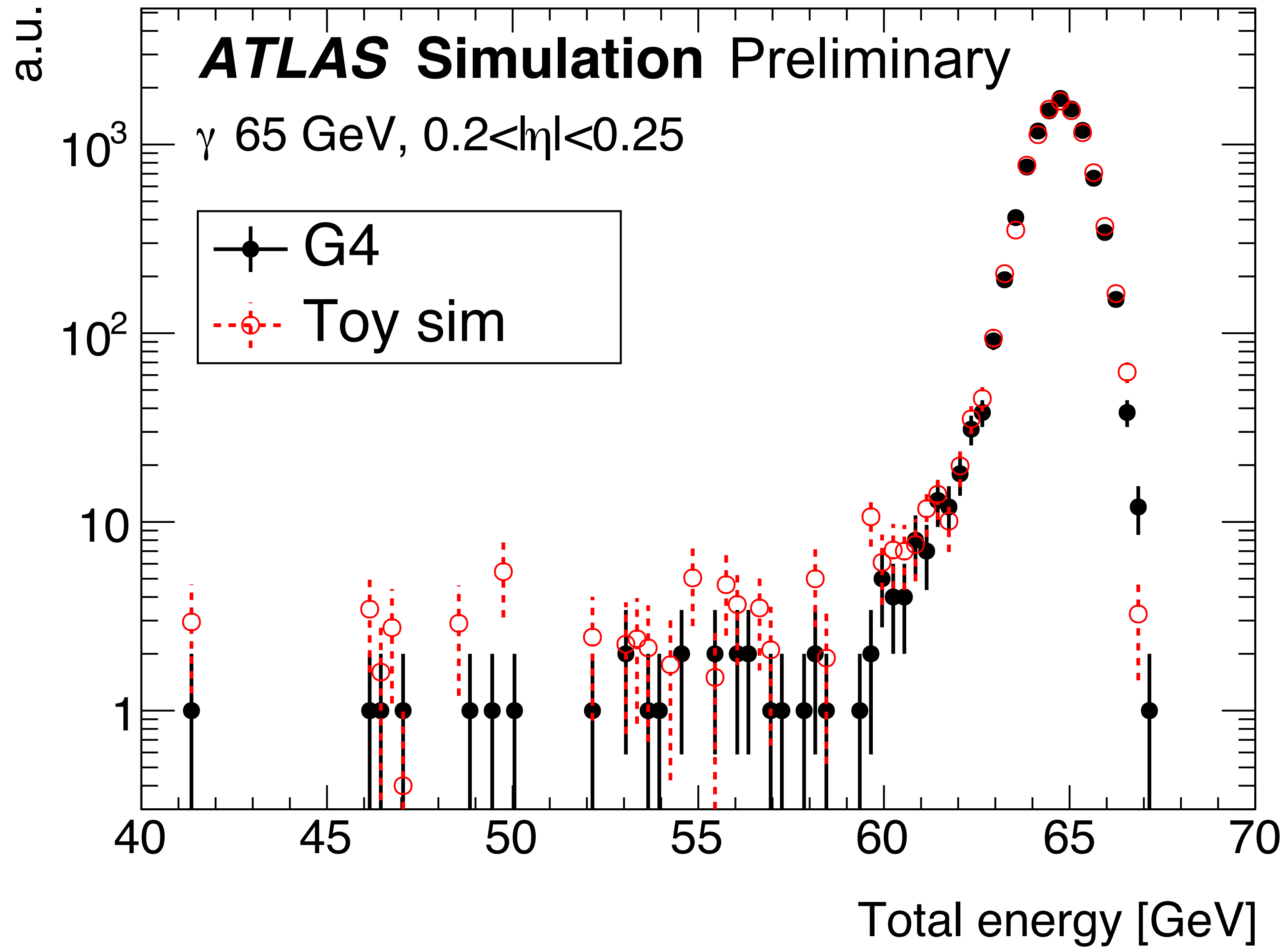
Energy Parametrization: Toy validation







- Small mismodeling compared to G4 is observed
- The mean is shifted & RMS is larger than that of G4

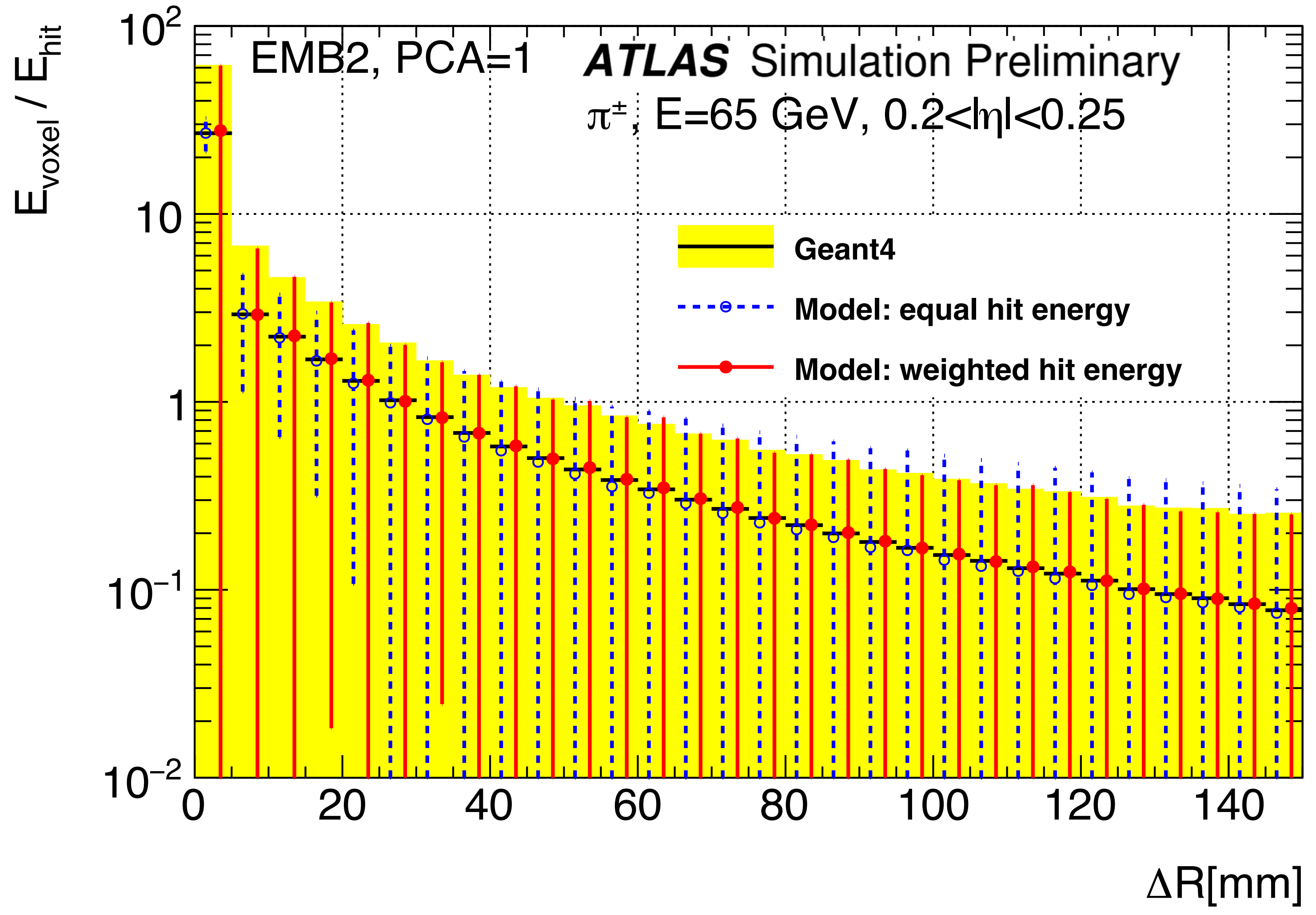


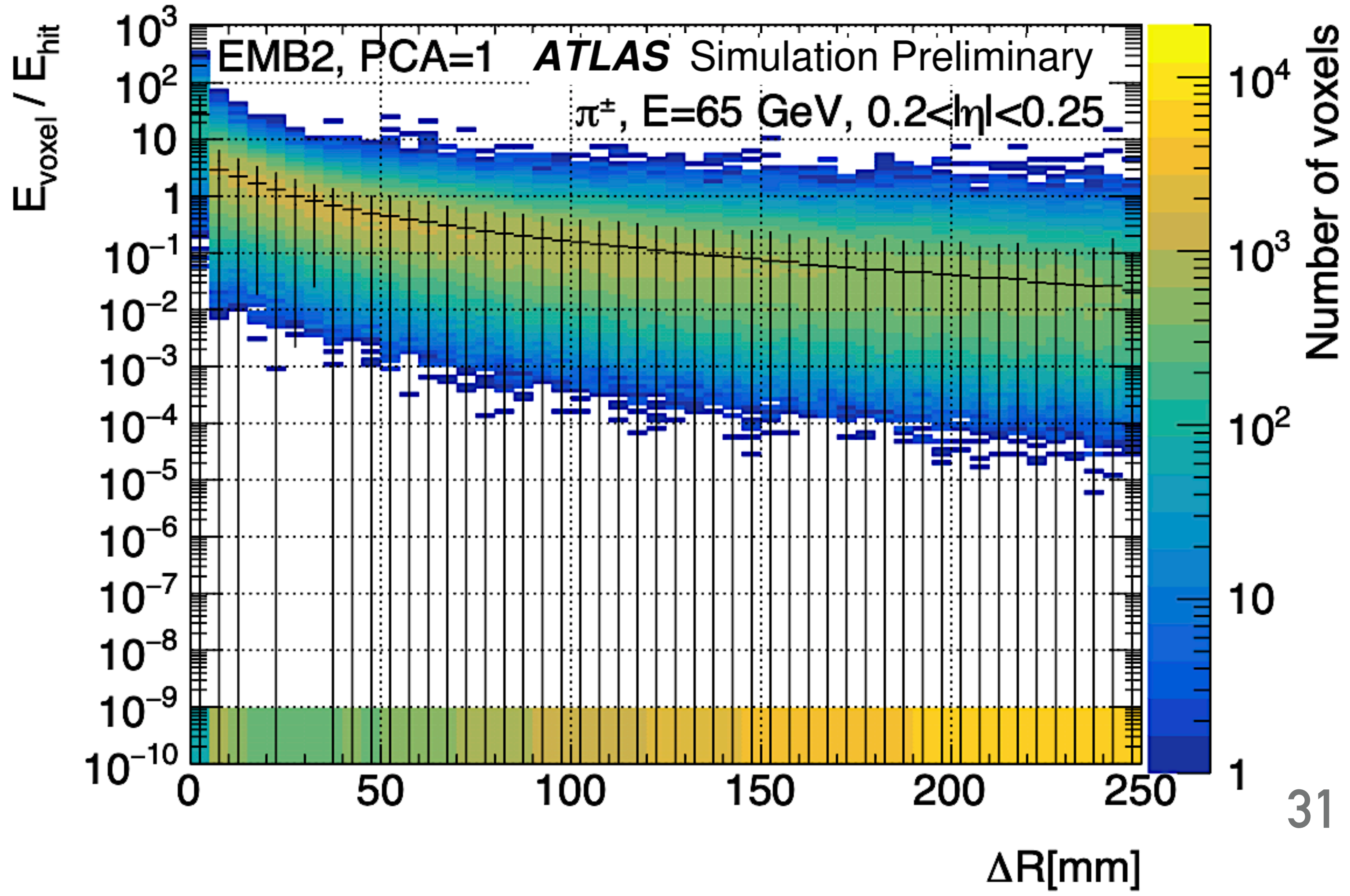
Weighted hit simulation for hadrons (2)

- Take the smaller RMS of the two poisson distributions ($RMS_{Poisson}$) reproducing: i)fraction of events with E = 0 ii) RMS of G4 distribution
- Calculate the weight as $w = \langle E_{voxel}/E_{hit} \rangle / N_{Poisson}$ with $N_{Poisson} = 1/(RMS_{Poisson}/\lambda_{Poisson})^2$
- $E'_{hit} = E_{hit} \times w$
- An extra smearing function (e^s with s being a RandGauss) is introduced for cases $RMS_{Poisson} < RMS_{G4}$
- Calculate the unaccounted fluctuation as: $RMS^2_{smearing} = RMS^2_{G4} - RMS^2_{Poisson}$
- The RMS of the smearing function e^s is then matched to $RMS_{smearing}/\sqrt{N_{Poisson}}$ by adjusting the sigma of the gaussian distribution to draw the random number s
- $E''_{hit} = E_{hit} \times w \times e^s$

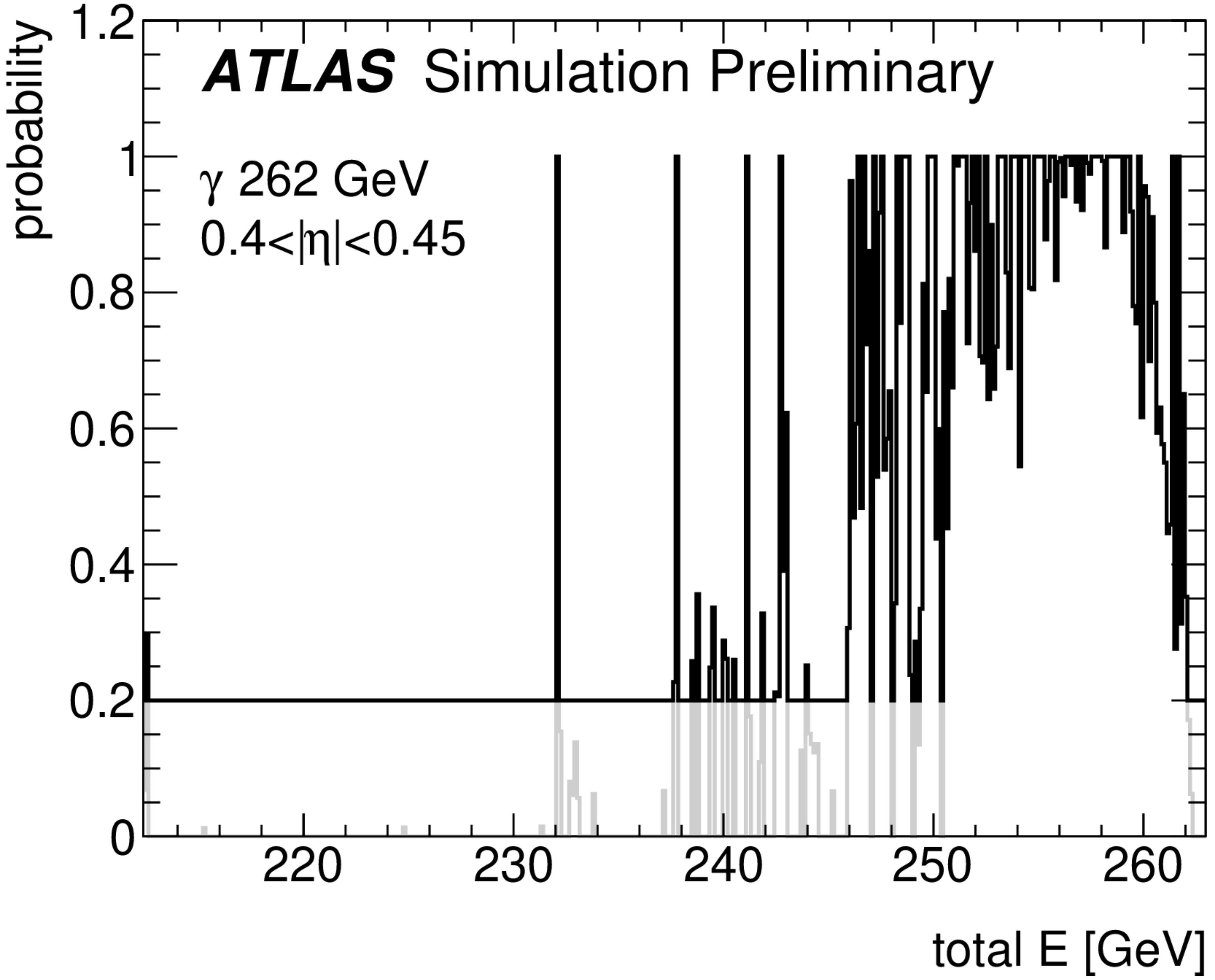
The resulting weight and sigma of the gaussian distribution for smearing are stored as a function of ΔR[mm] and used in simulation to improve energy deposits.

AF3 with equal hit energy deposition:

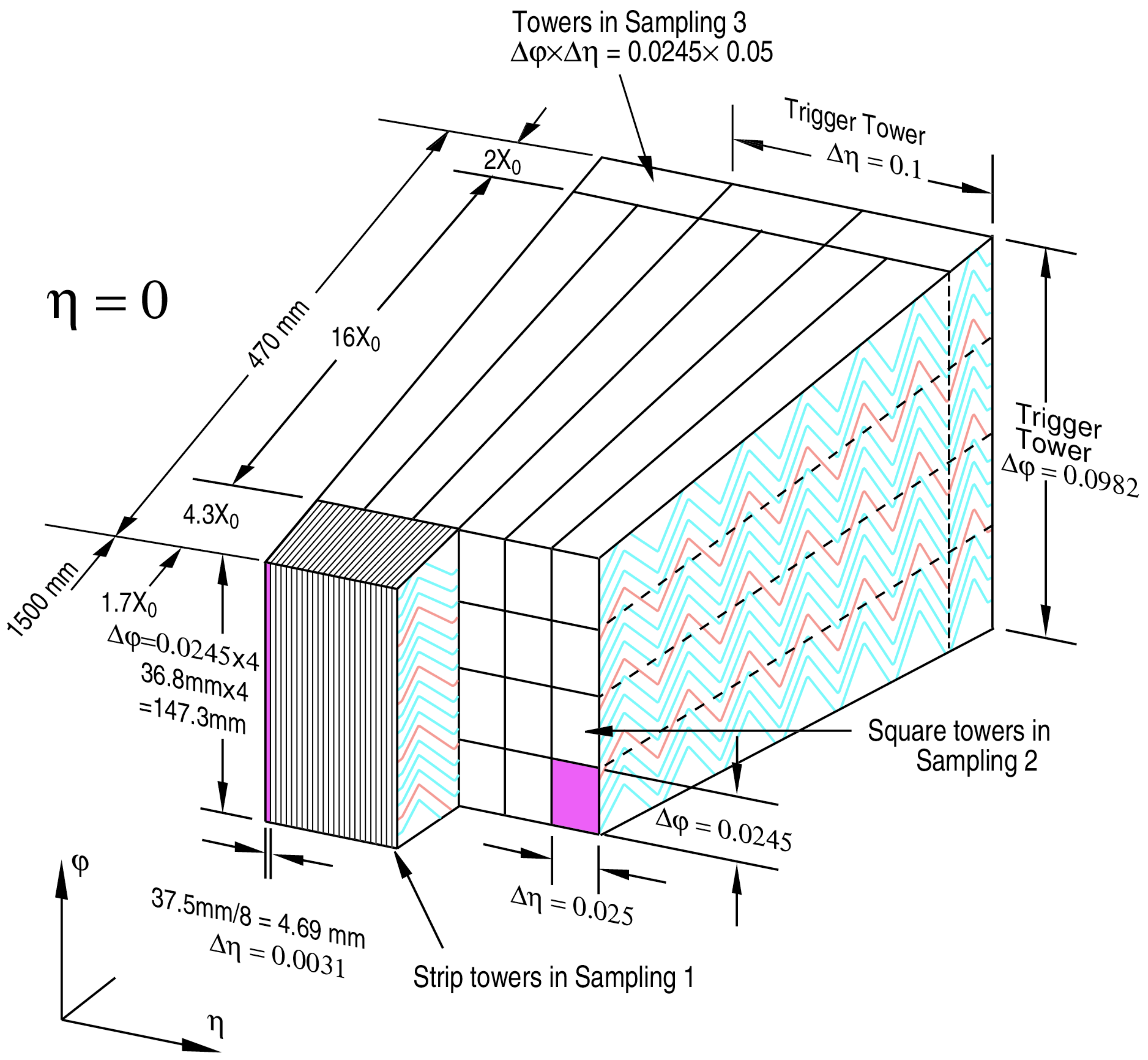


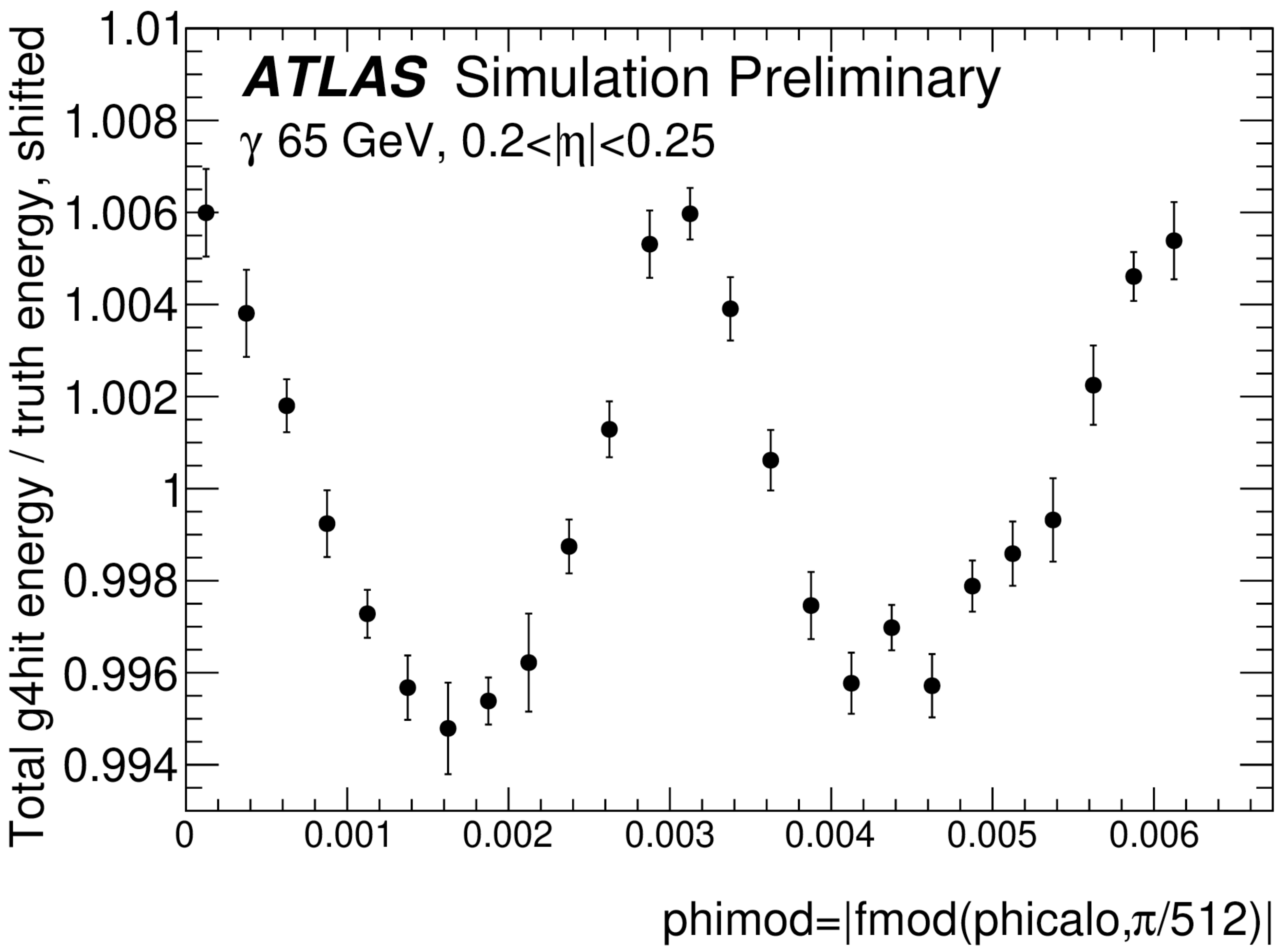


PDF for probabilistic reweighting



Phi modulation





Best epoch selection

For every checkpoint, 10k events are generated foreach energy point. The χ_2 is then evaluated between the binned distributions of the GAN and the training sample. The total χ_2 for a checkpoint is the sum of the $15\chi_2$. The checkpoint with the lowes tχ_2 is chosen for each GAN