AtlFast3,
the ATLAS fast simulation tool in Run3

Rui Zhang, on behalf of the ATLAS Collaboration

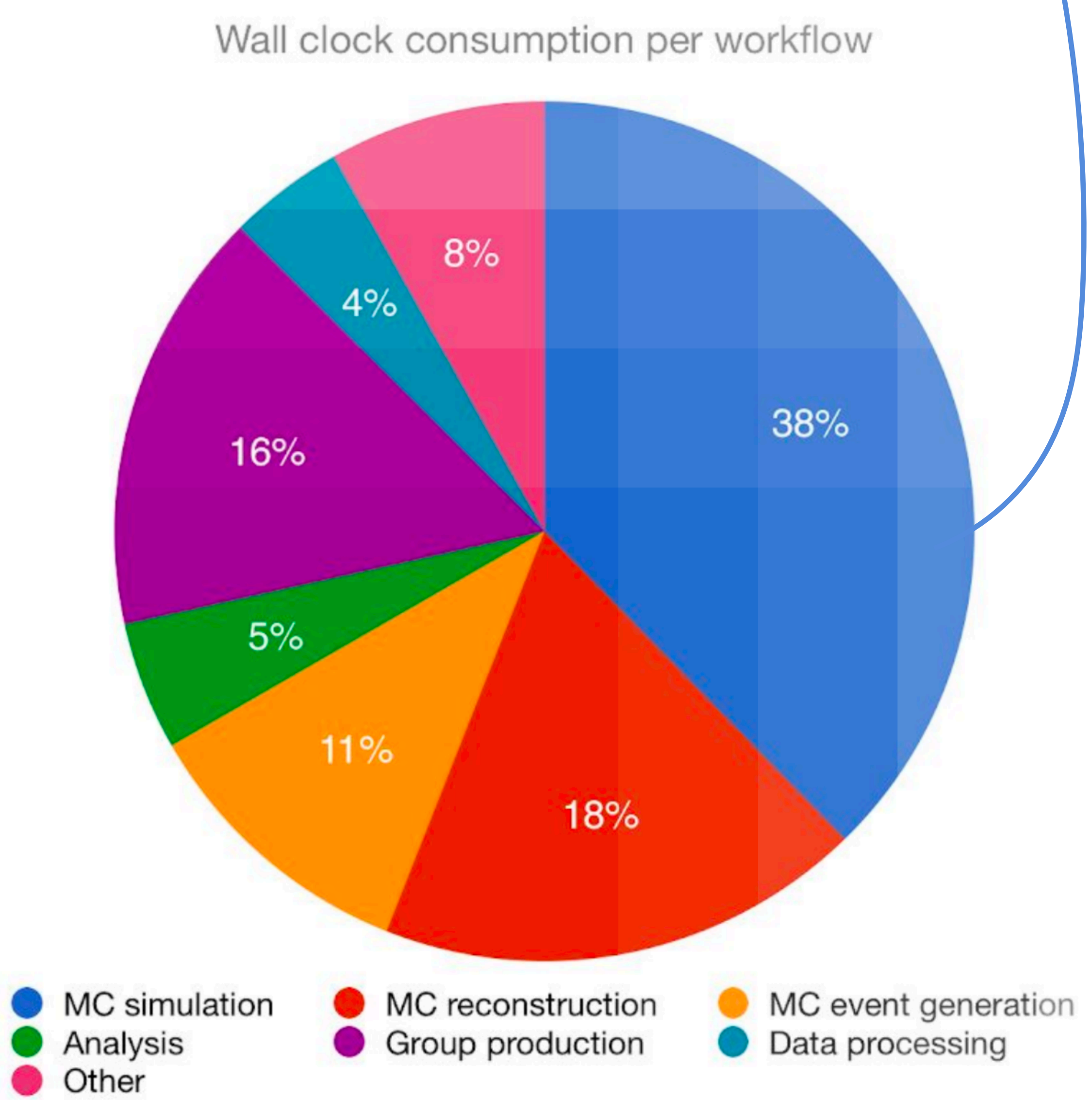
International Conference on New Frontiers in Physics (ICNFP 2023)

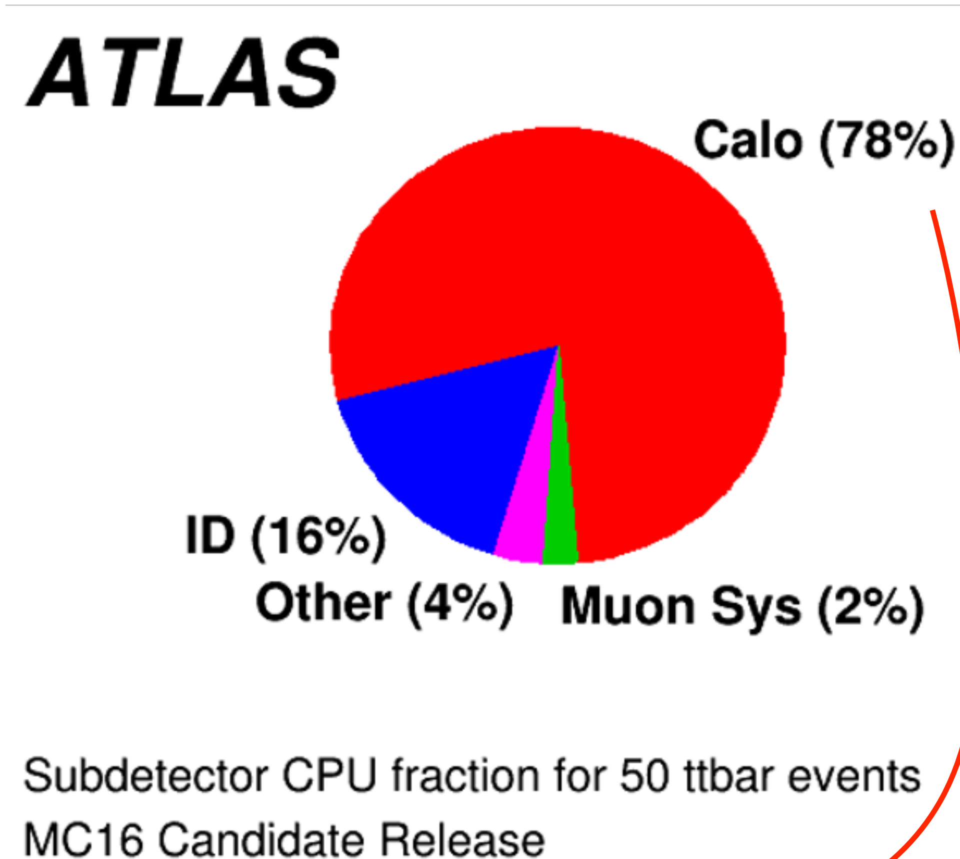
University of Wisconsin-Madison, Wisconsin

July 10—23, 2023

Simulation in ATLAS

Monte Carlo (MC) production takes ~70% of the GRID CPU time in ATLAS: dominated by MC full simulation done in Geant4





The Geant4 simulation is dominated by the simulation of the calorimeters

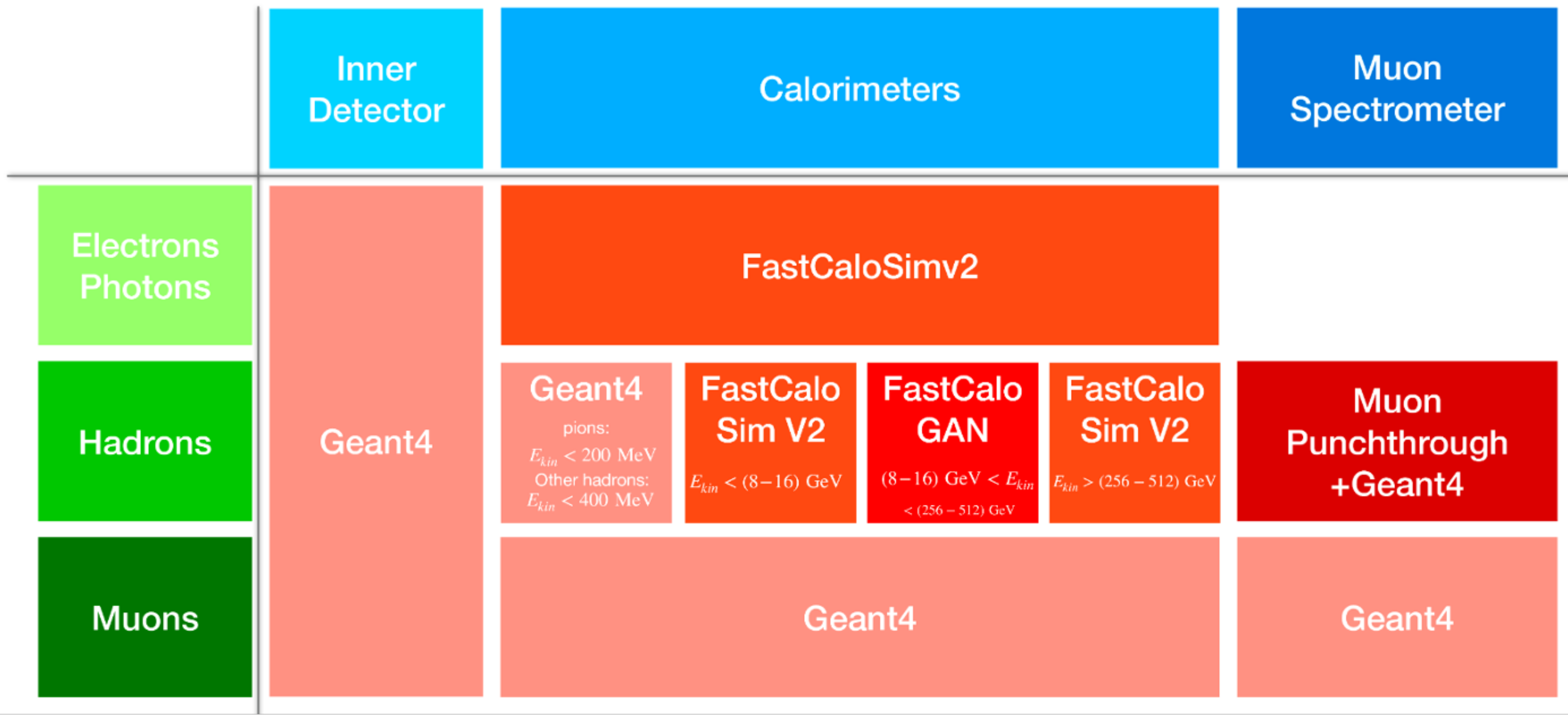
Motivation for fast simulation

- More MC samples will be needed in HL-LHC but limited computing resources
  - ➜ Make MC production fast
- The bottleneck of the MC production chain is the simulation of showering in calorimeters
  - ➜ Make calorimeter simulation fast
- ATLAS developed a fast calorimeter simulation in Run2 (AFII) but it does not reproduce data as well as Full Simulation
  - ➜ Make calorimeter simulation fast and better
- Based on AFII, ATLAS developed the next generation for fast calorimeter simulation, called AtlFast3 (AF3)
  - COMPUT SOFTW BIG SCI 6, 7 (2022)

AF3 in Run 2

[*COMPUT SOFTW BIG SCI 6, 7 (2022)*]

AF3 configuration



- AF3 employs two techniques: FastCaloSim V2 and FastCaloGAN. They are complementary in different part of detector simulation.
- FastCaloSim V2 parametrises showers in the longitudinal (distance from the interaction point) and lateral (angular spread about the interaction point, unrolled barrel view) directions under different conditions; FastCaloGAN uses machine learning models to generate showers.

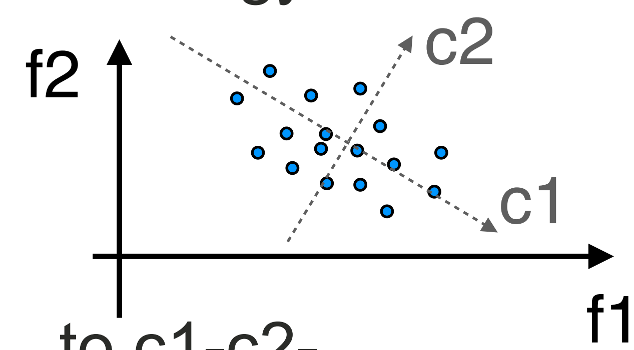
AF3 strategy

- Simulate photons, electrons and pions to represent all particles interacting in calorimeters
  - All hadrons share the pion parametrisation with a correction for the mass
- Divide calorimeter in 100 slices in $|\eta|$
  - In each slice we studied 17 energy points from 64 MeV to 4 TeV (in powers of two)
- For each particle/energy/$|\eta|$ point a Geant4 sample (10k) is produced at the calorimeter surface
  - Noise and other imperfections are removed to parametrise on “perfect” calorimeter showers
- The FastCaloSimV2 and FastCaloGAN parametrisations are derived from these samples
  - Both strategies define their own custom voxels to group the calorimeter hits to avoid handling the complex and non homogeneous calorimeter structure
- At simulation time, hits are produced in the calorimeter based on the chosen parametrisation for that particle
  - Then additional corrections are applied to match the precision of Geant4

FastCaloSim V2

- Predict energy fractions in layers (highly correlated) using Gaussian

1. Take a sample of showers, each shower is represented by the energy fractions in each layer: s = (f1, f2, …)



2. Re-parametrise these fractions using Principal Component Analysis (PCA). Effectively change coordinates from f1-f2-… to c1-c2-…

3. Split the sample of showers into PCA bins using the coordinate with most variation. Within each bin, there is a subset of the whole sample of showers.

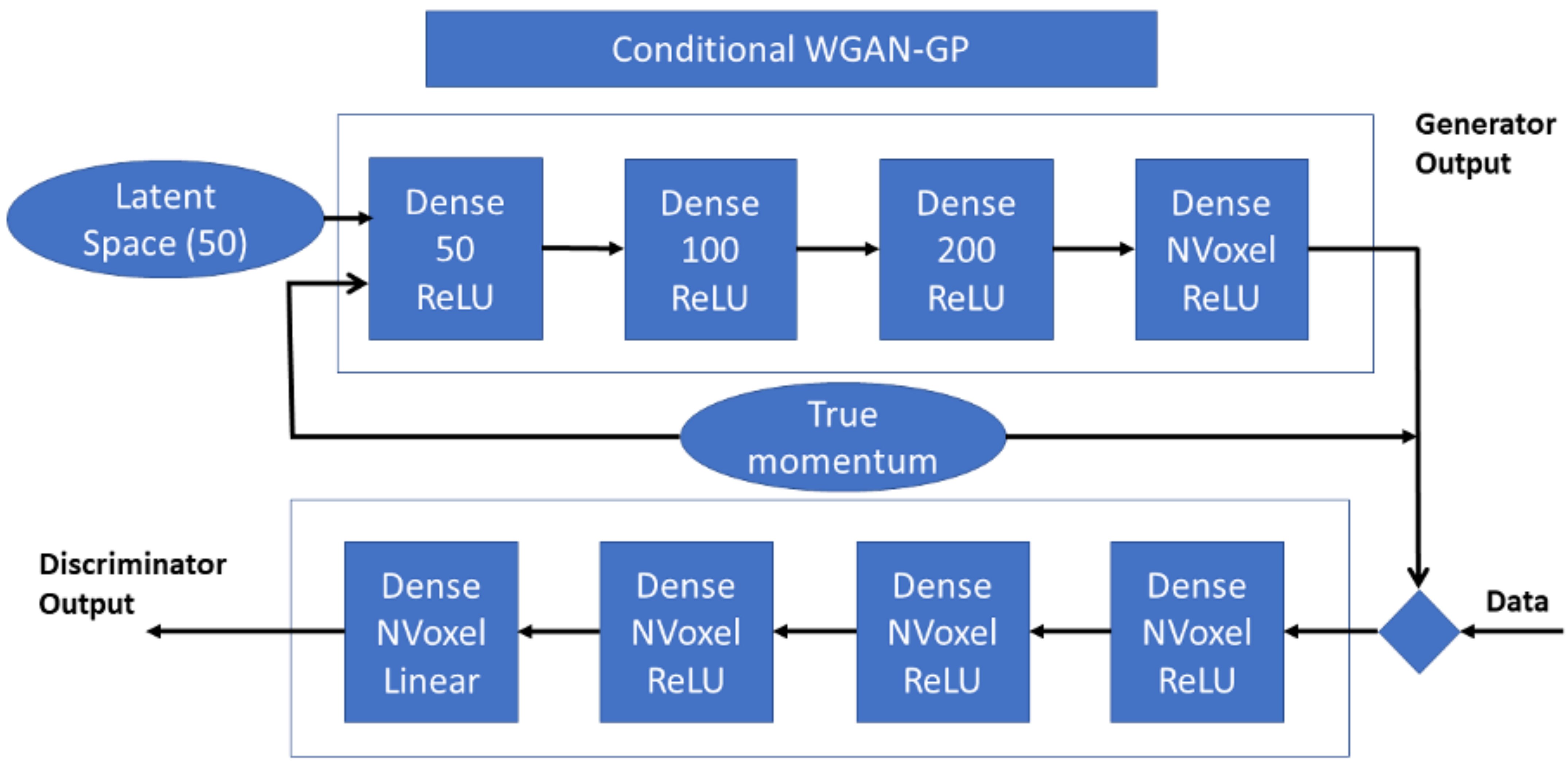
4. For this subset, c1-c2-… are no longer the principle components due to local variations. A 2nd PCA is performed. Output is a set of uncorrelated, Gaussian-shaped distributions.

   - To store: Cumulative distributions, PCA matrices, mean and RMS of the output Gaussians

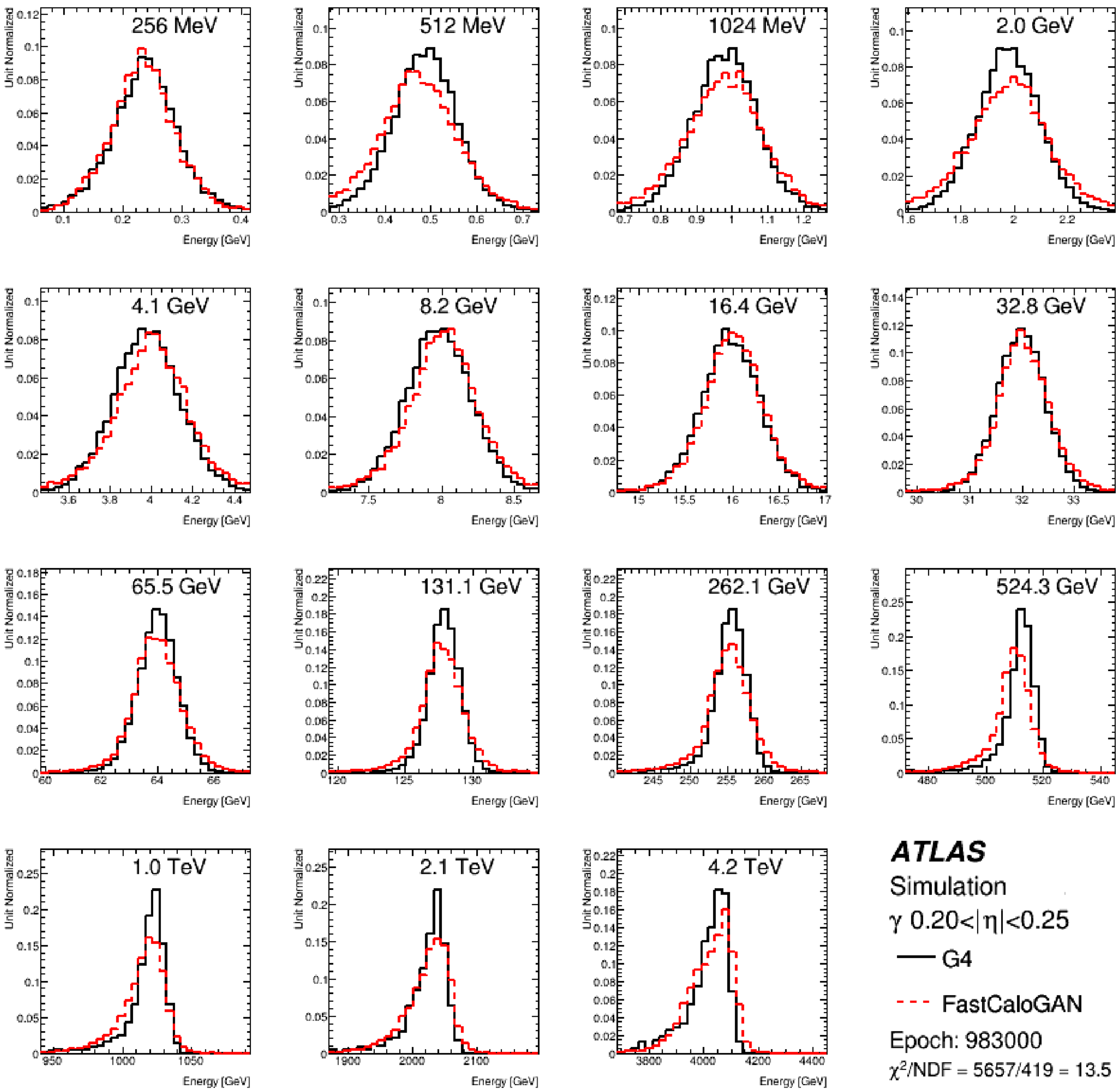
- Shower shape is parametrised and stored in a given calorimeter layer and PCA bin

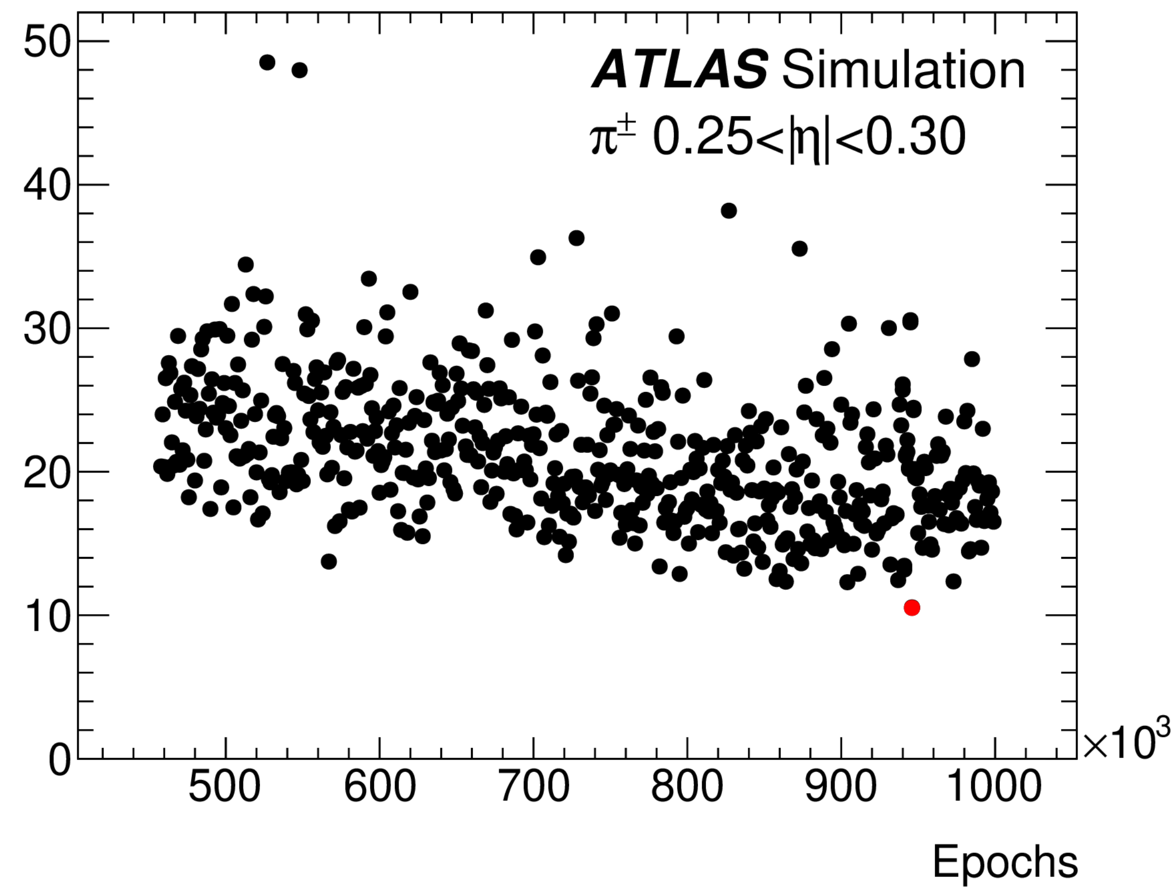
FastCaloGAN

- For each particle and |η|, a GAN is trained on all energies (as conditions)
  - 300 GANs are trained in total
- A similar structure is used for all GANs

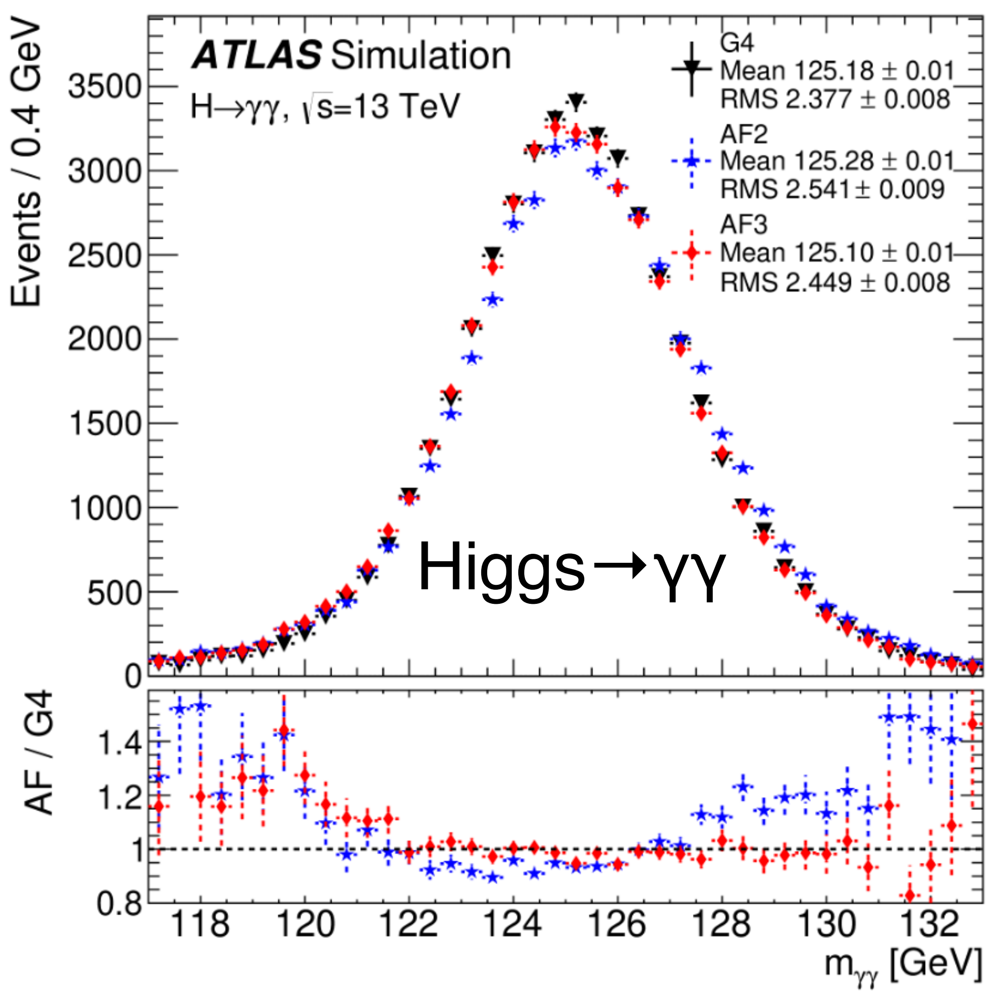


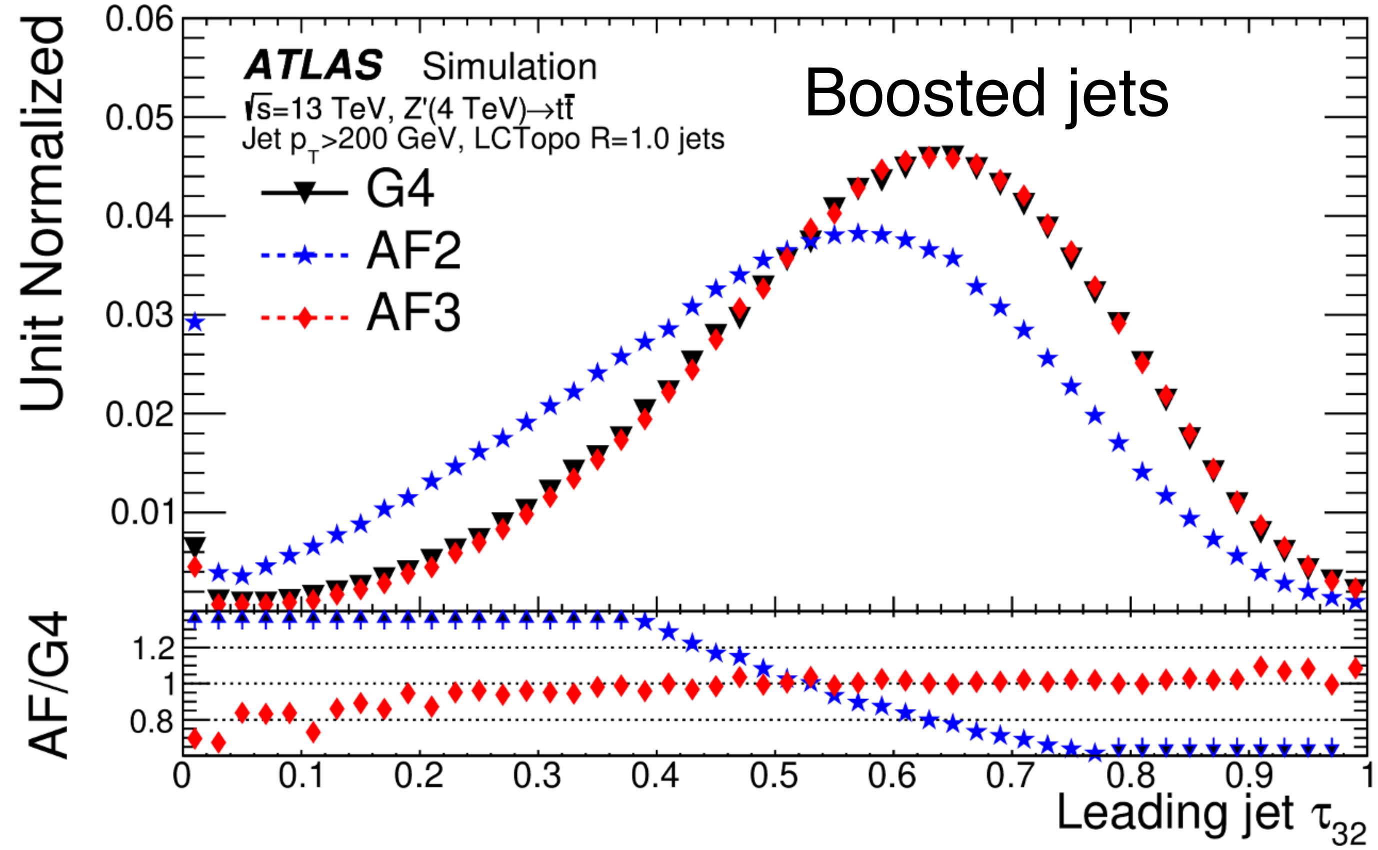
Training

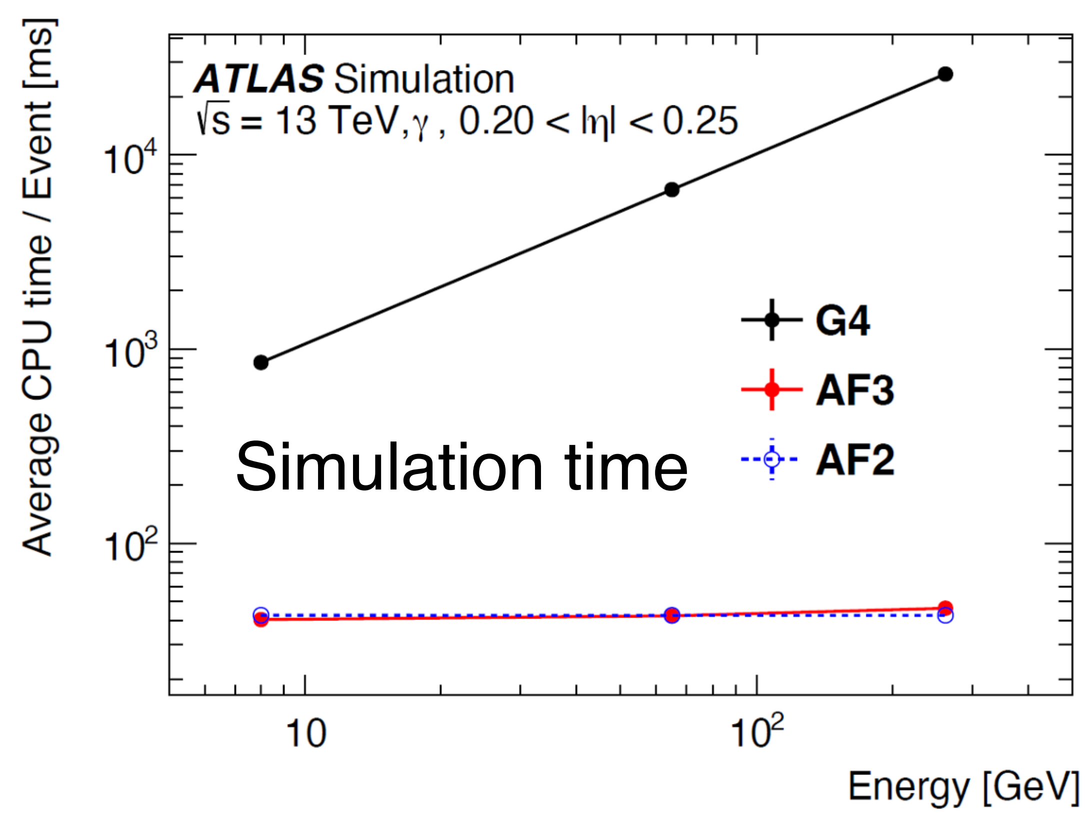

256 MeV
512 MeV
1024 MeV
2.0 GeV
4.1 GeV
8.2 GeV
16.4 GeV
32.8 GeV
65.5 GeV
131.1 GeV
262.1 GeV
524.3 GeV
1.0 TeV
2.1 TeV
4.2 TeV
Unit Normalized
Energy [GeV]
ATLAS
Simulation
γ 0.20<|η|<0.25
G4
FastCaloGAN
Epoch: 983000
χ²/NDF = 5657/419 = 13.5


ATLAS Simulation
π± 0.25<|η|<0.30
×10³
Epochs

AF3 performance







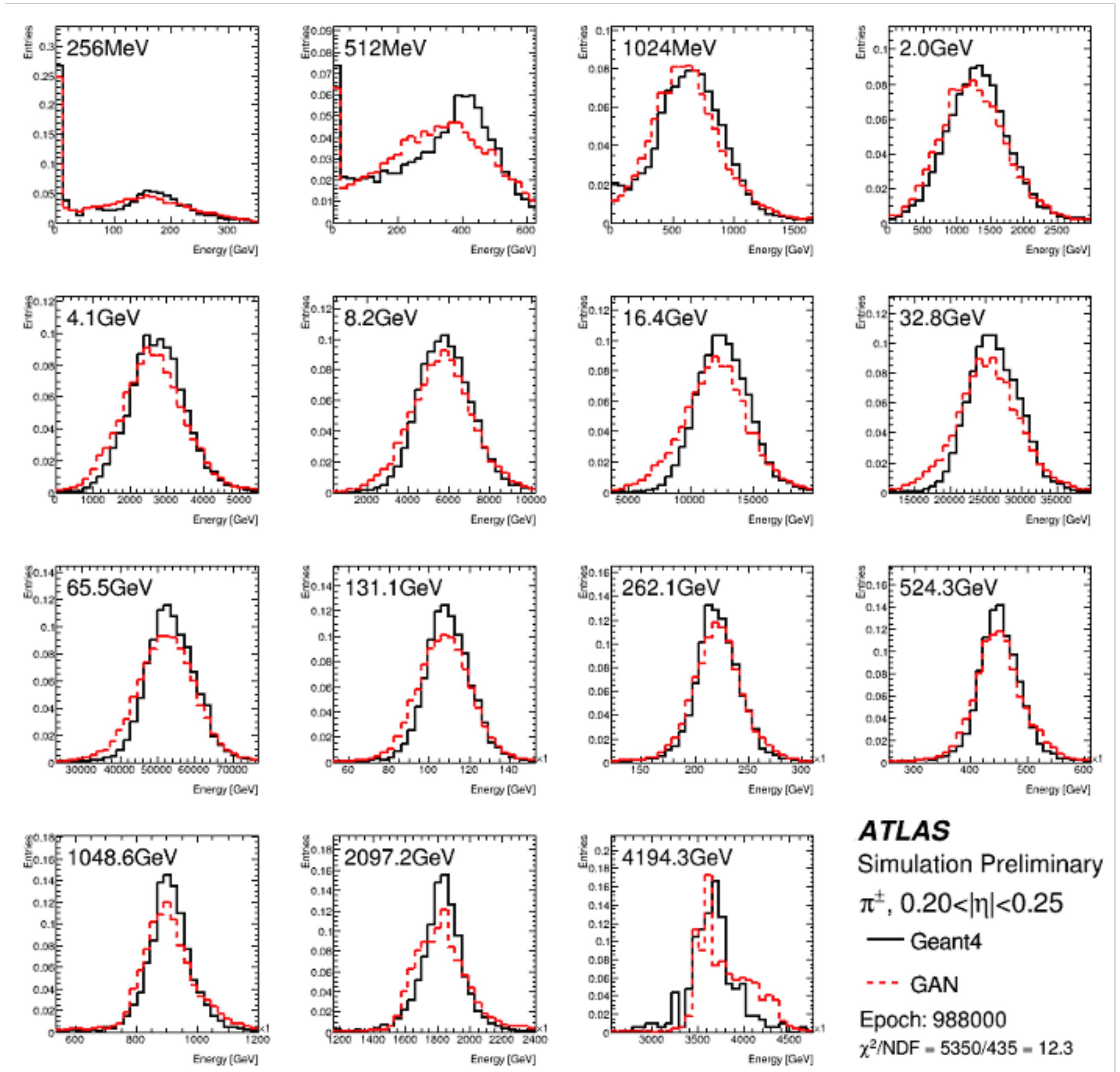
Improvements for Run3
FastCaloGAN V2

FastCaloGAN optimisations

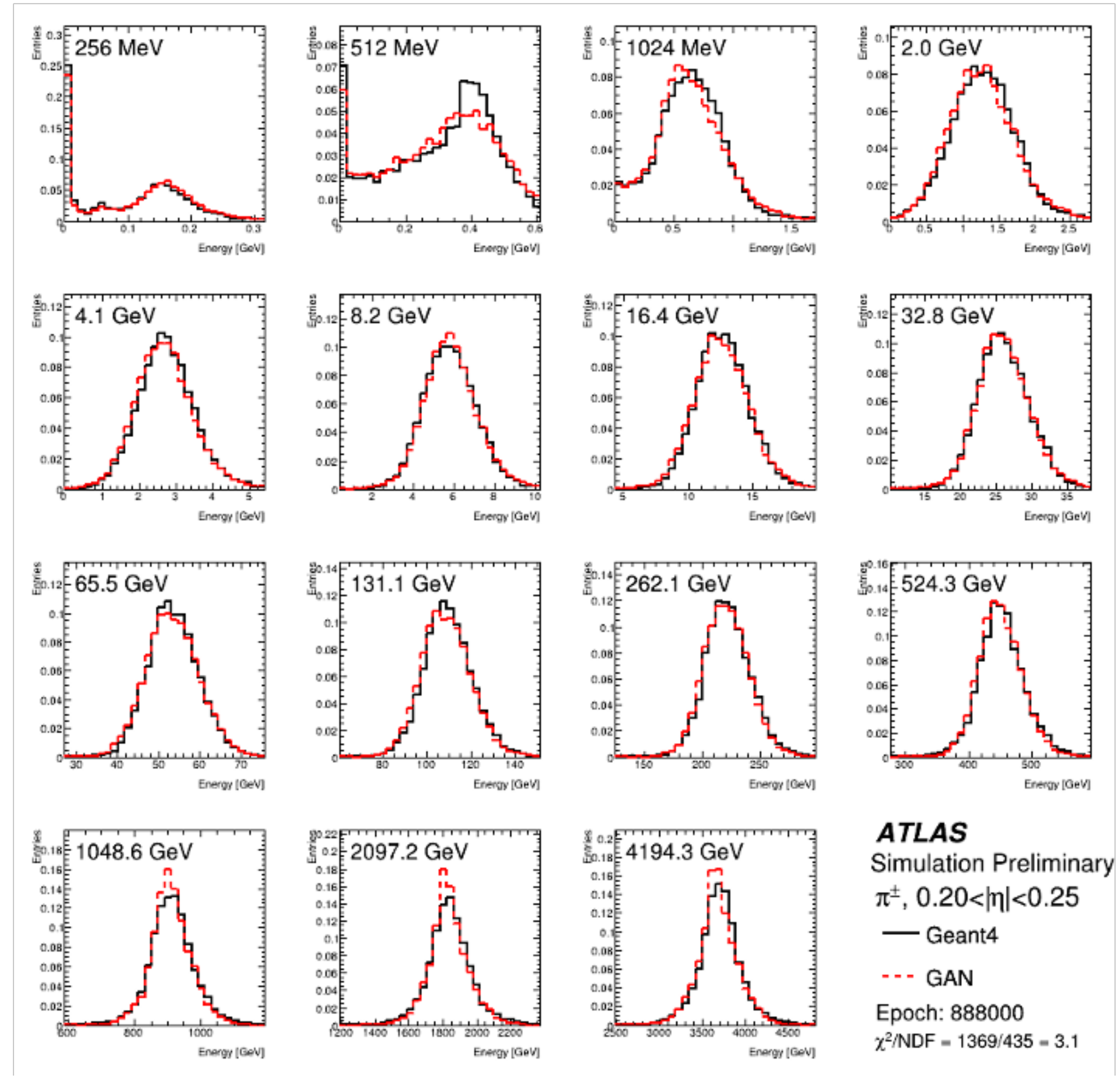
- More granular voxelisation for a more accurate voxel-to-cell energy assignment
  - This is further improved by exploiting energy-independent lateral shower profile
- New TensorFlow provide more stable and faster training
- Change training strategy to two-step training
  - Divide the detector into regions based on groups of $|\eta|$ and train with a single $|\eta|$ for an extended period in each region
  - Train with other $|\eta|$ slices, starting from the best trained model obtained in the first step
- Hyper parameter optimised for each GAN
  - Bigger networks (due to larger input dimensions)
  - High batchsize
  - Swish activation for e/γ (useful in e/γ but not in pions)
  - Split e/γ in high and low energies samples (different behaviour in low and high energies)
    - 2 GANs are trained in each $|\eta|$
- Improved voxel-to-cell energy assignment exploiting energy-independent lateral shower profile

Best GAN for pions: V1 vs V2

V1



V2

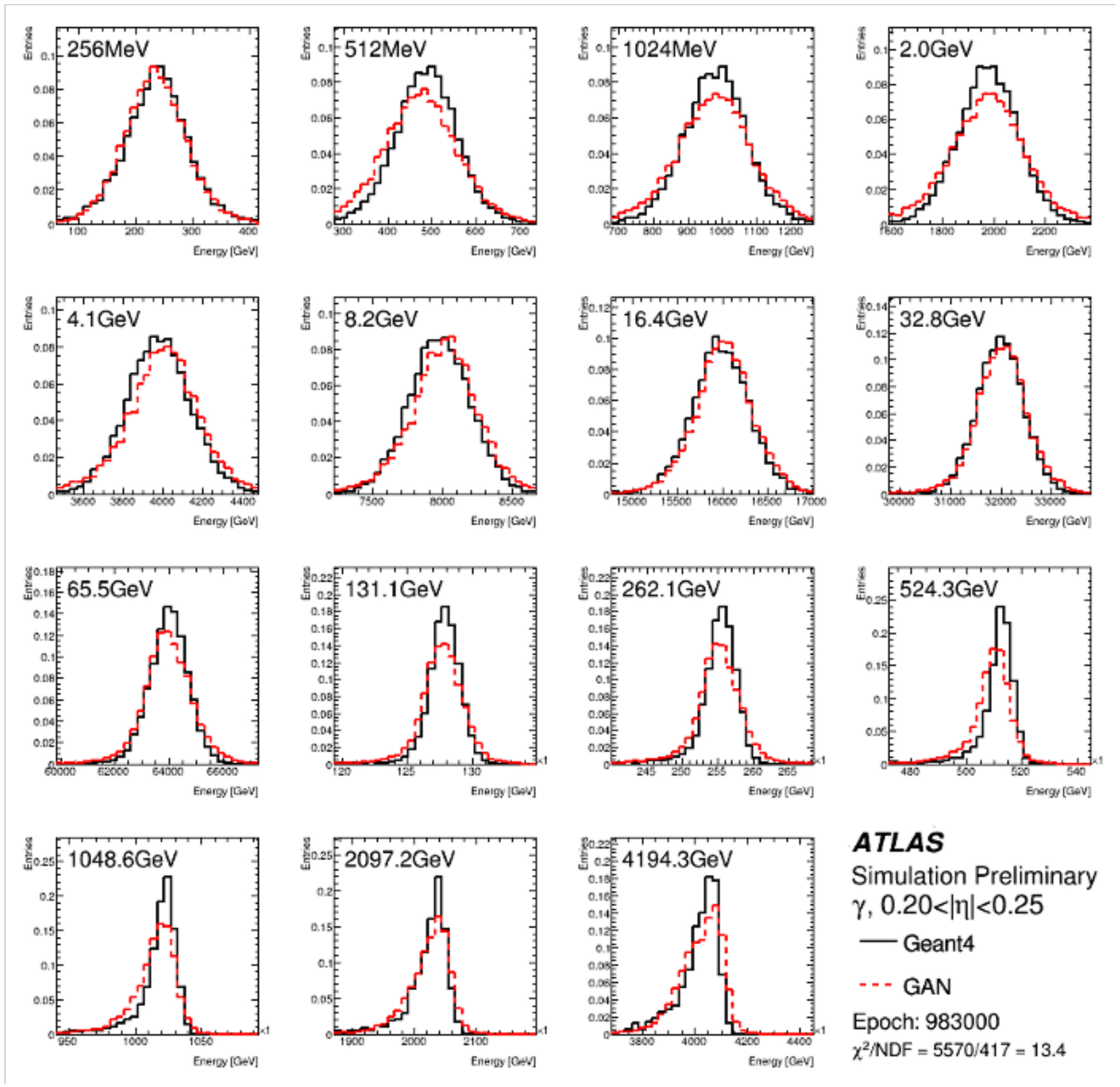


χ^2/NDF = 12.3

χ^2/NDF = 3.1

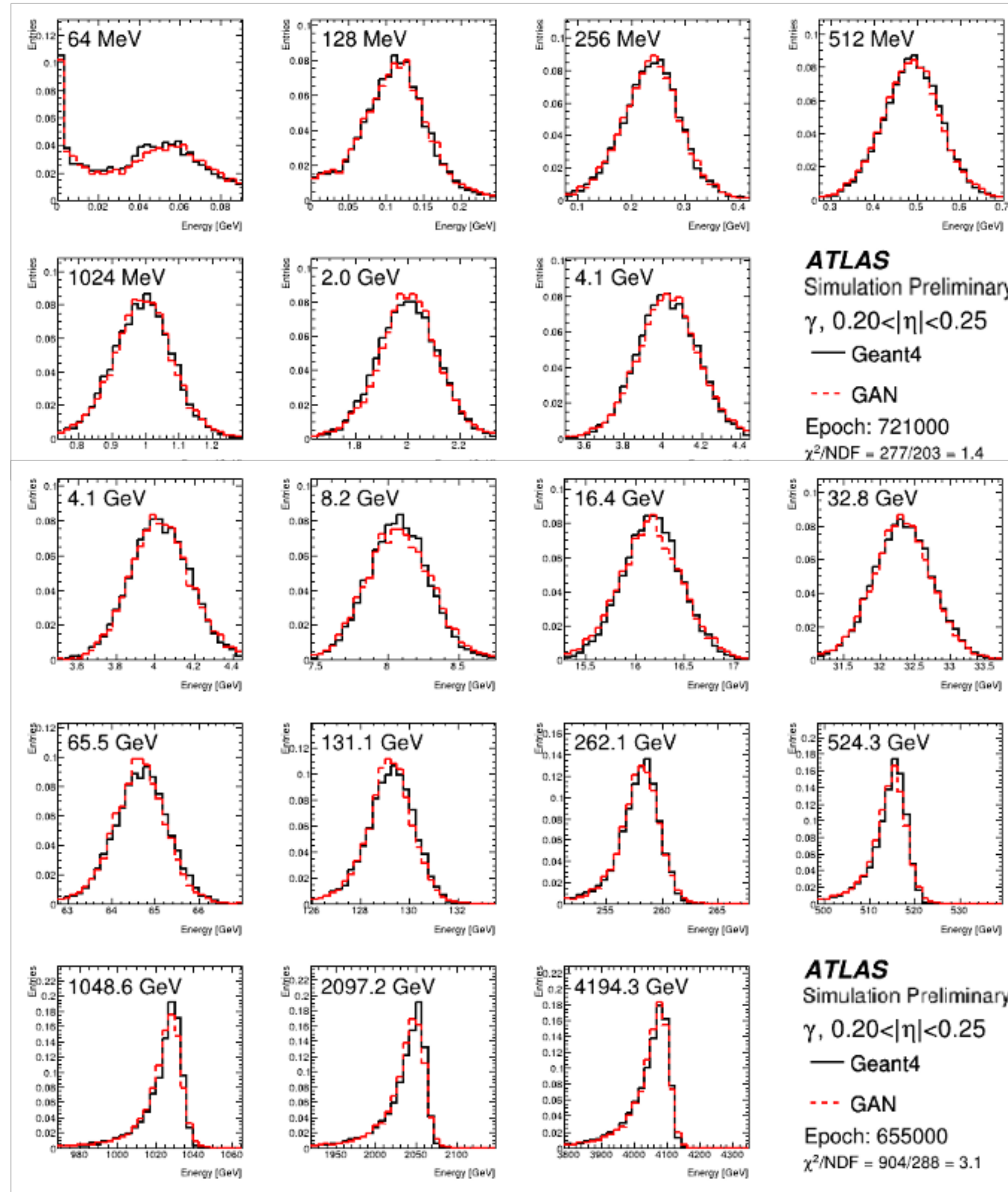
Best GAN for photons: V1 vs V2

V1

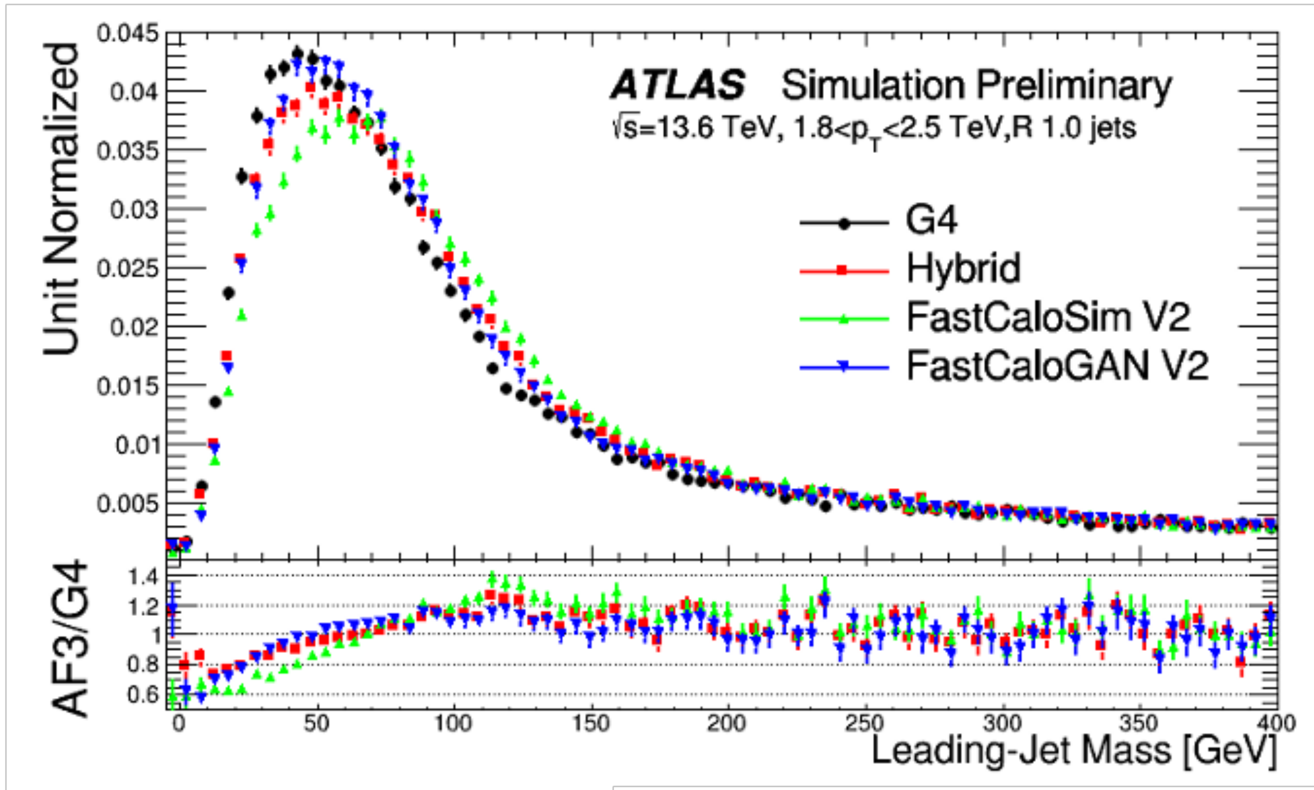


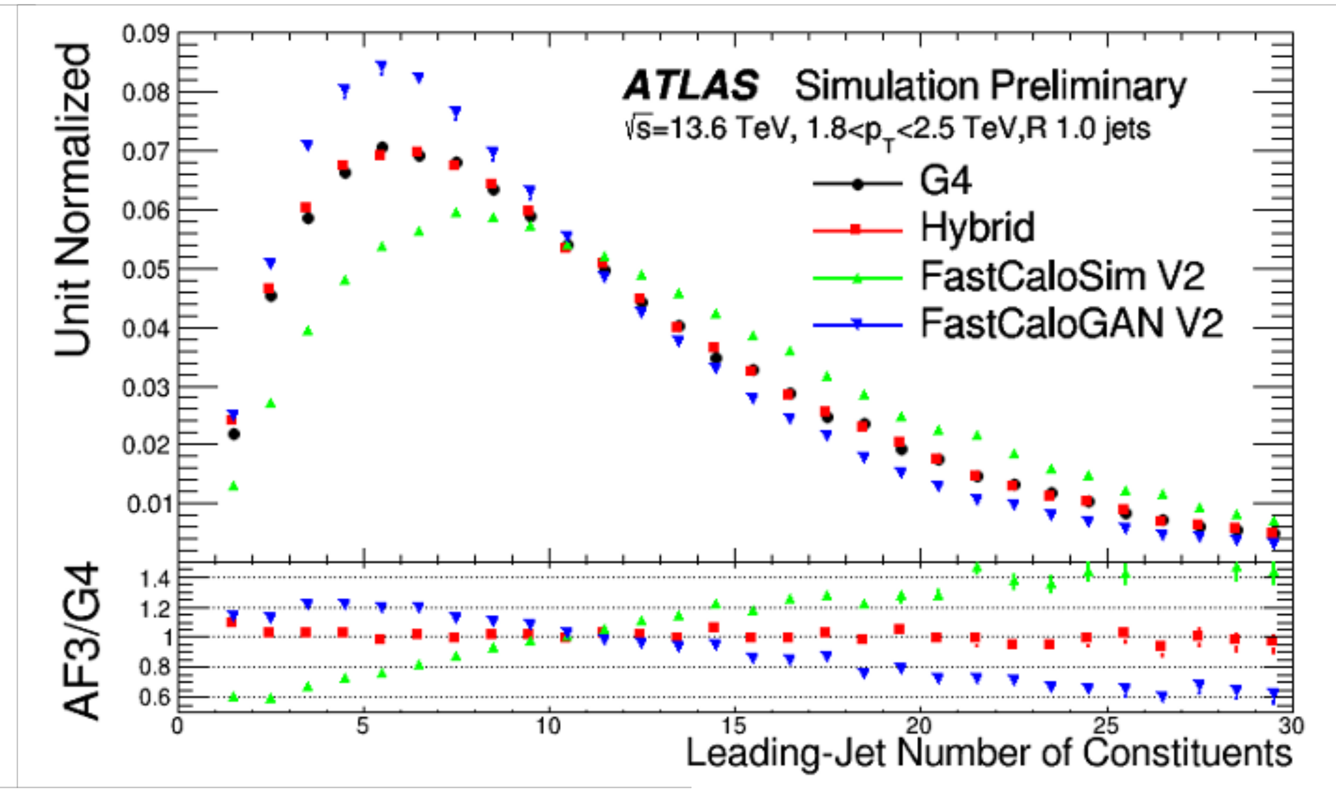
χ^2/NDF = 12.3

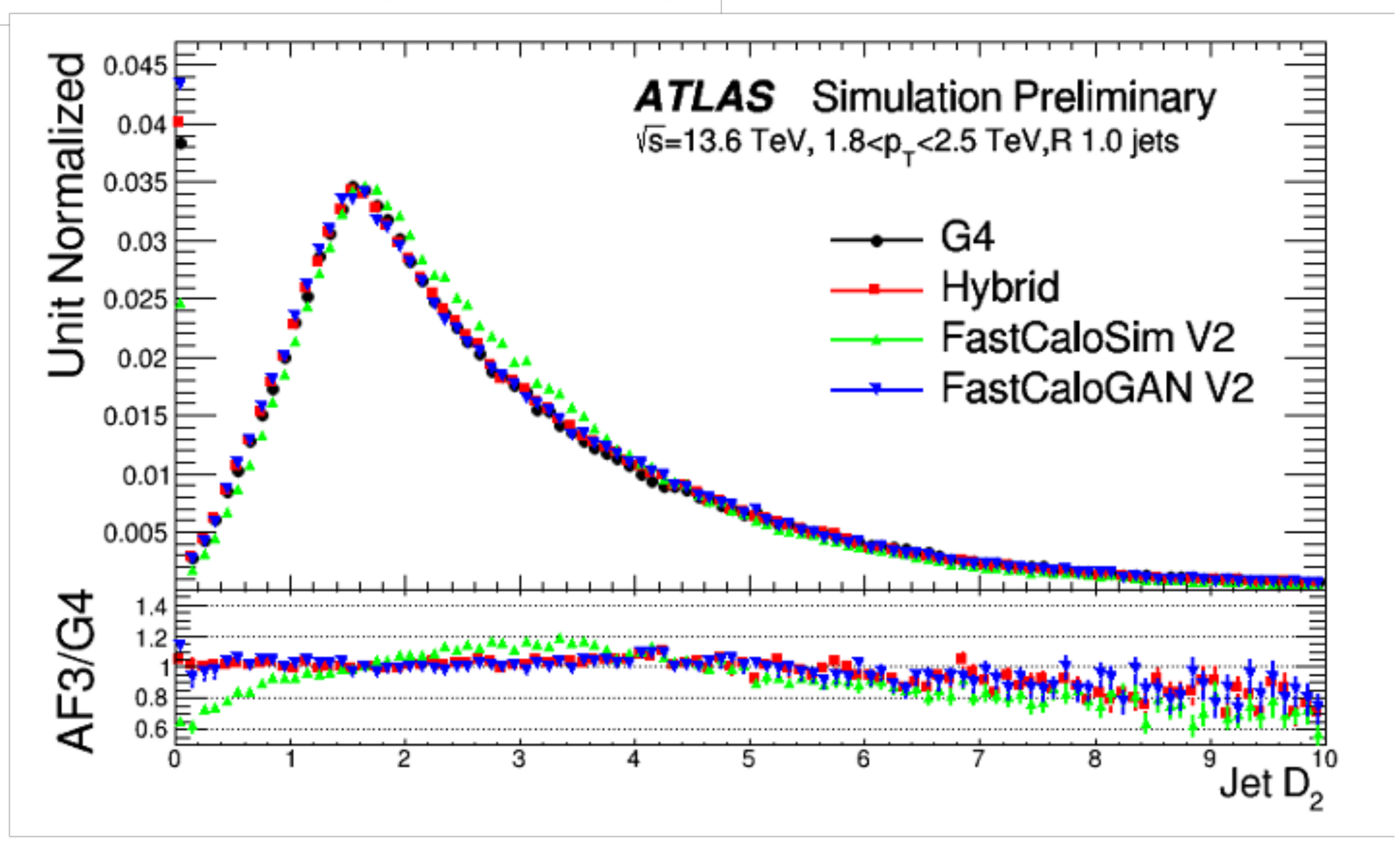
V2



Jet performance in Run3







Conclusion

- AtlFast3 has significant improvements w.r.t. AtlFast2 in reproducing key variables used in analysis
  - This is crucial to allow a wider use of fast simulation required to match the designed luminosity in Run3 and beyond
  - COMPUT SOFTW BIG SCI 6, 7 (2022)
- FastCaloGAN V2 is developed for Run3 and currently under deployment.
- There is still room for improvement beyond Run3 and we will continue to explore new models to further improve the simulation in ATLAS
  - We released the public dataset used for the #calochallenge and will consider adopting tools that will achieve high performance

Thank you for your attention!