Evolution of fast simulation using Generative Adversarial Network in ATLAS and beyond

Rui Zhang

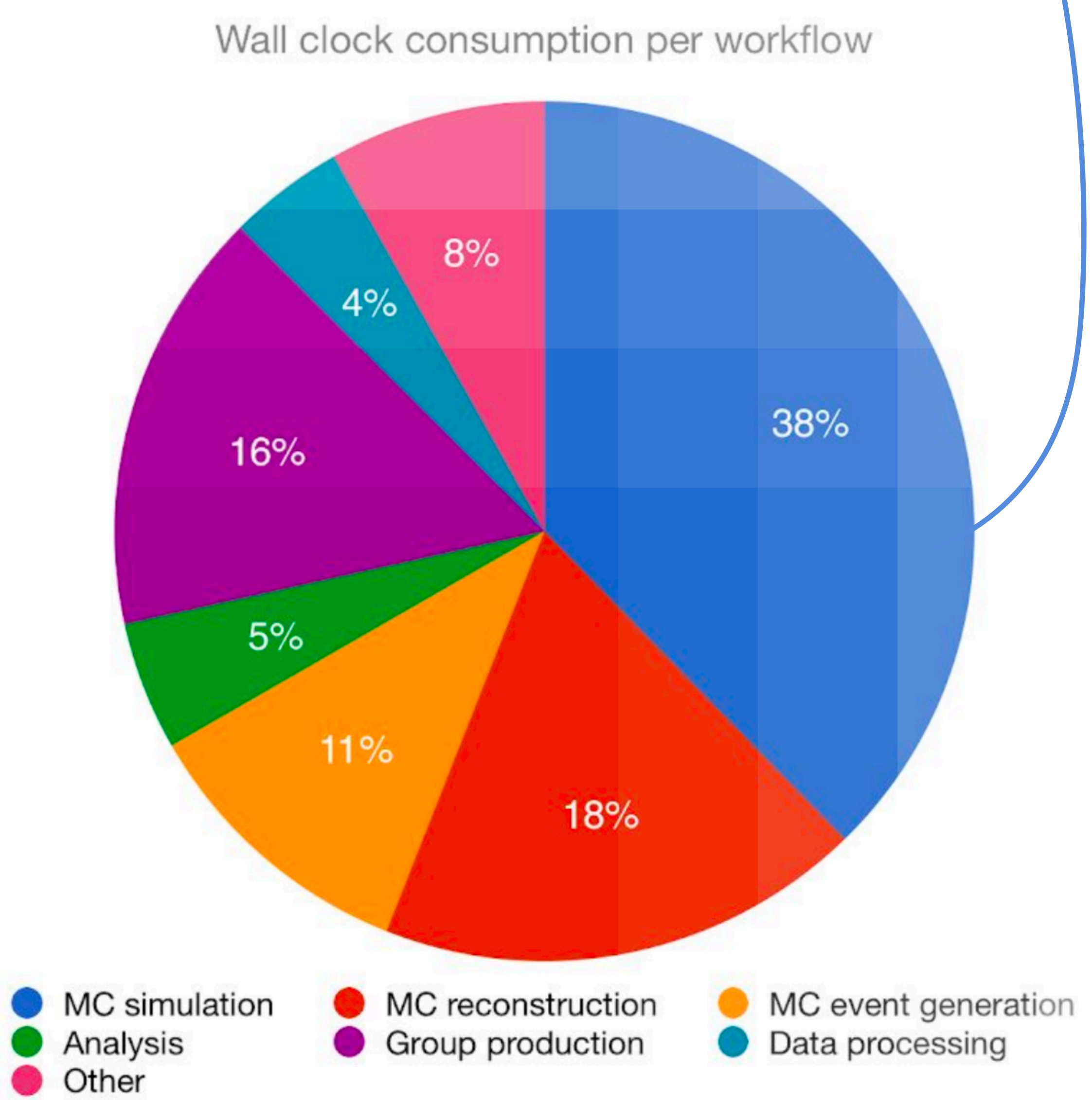
University of Wisconsin-Madison, Wisconsin

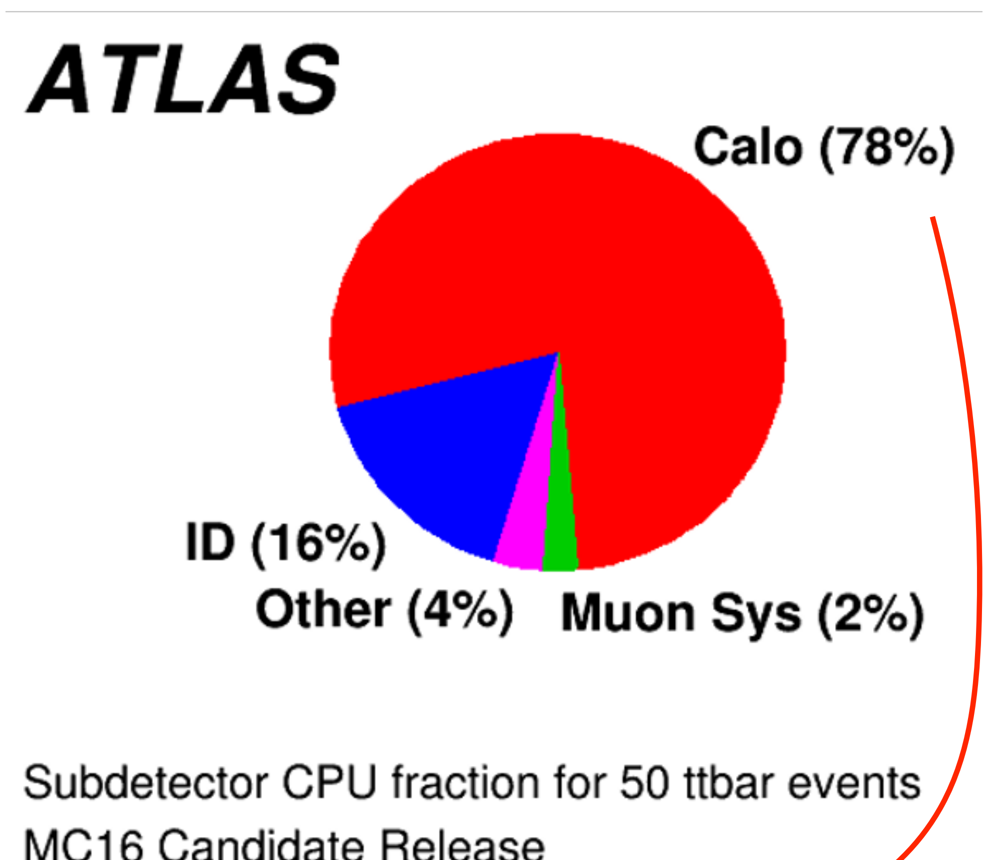
Quantum Computing and Machine Learning Workshop

青岛 11-14 Aug 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



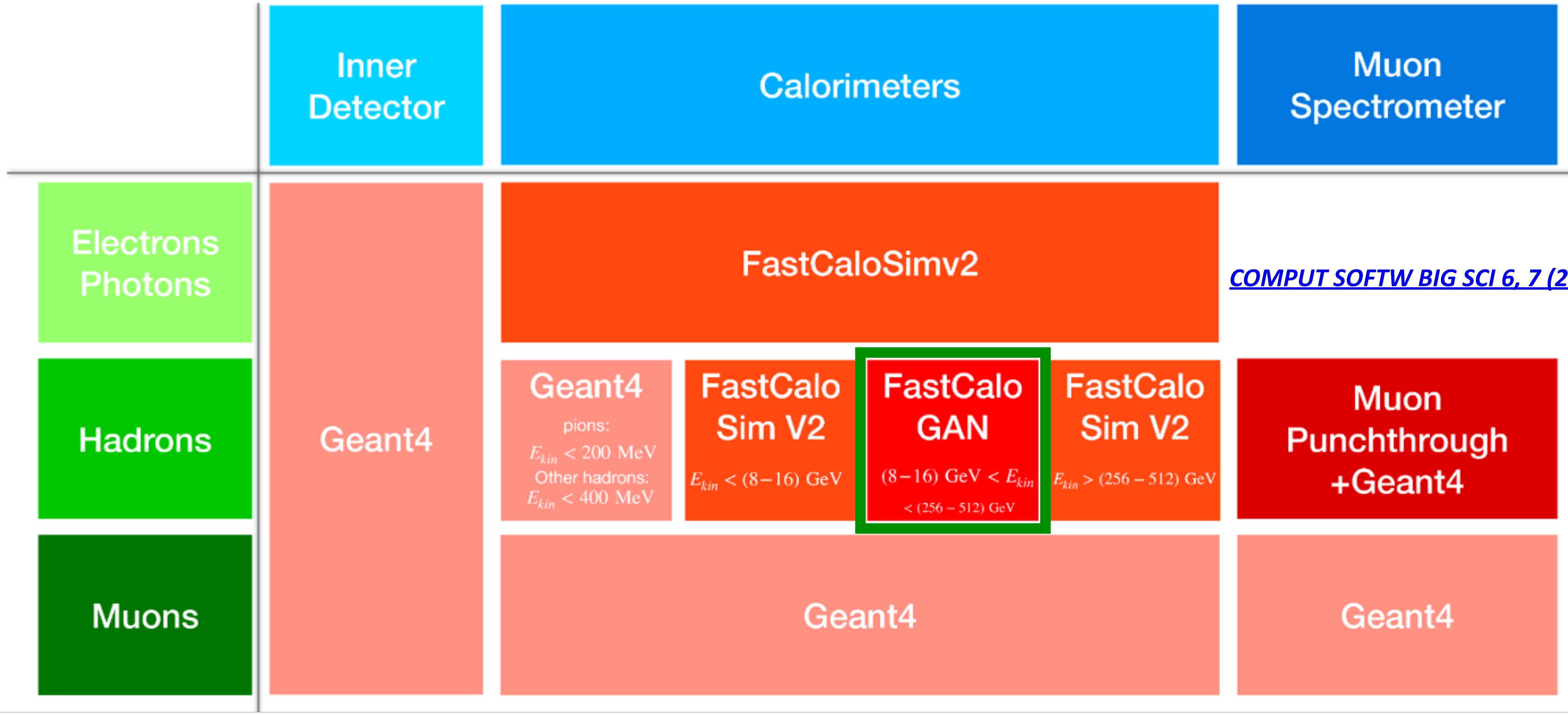


The Geant4 simulation is dominated by the simulation of the calorimeters

Motivation for fast simulation

- More MC samples will be needed in High Luminosity 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 better
- Based on AFII, ATLAS developed the next generation for fast calorimeter simulation, called AtlFast3 (AF3)
  - COMPUT SOFTW BIG SCI 6, 7 (2022)

FastCaloGAN in AtlFast3



COMPUT SOFTW BIG SCI 6, 7 (2022)

- AtlFast3 employs two techniques: **FastCaloSim V2** and **FastCaloGAN**. They are complementary in different part of detector simulation.
- FastCaloGAN has been evolved in the past two years for higher precision. Now FastCaloGAN is responsible for larger parts.

Strategy of FastCaloGAN

- To simulate all showers in the detector we need to parametrise photons, electrons and pions
  - Hadrons share the pion parametrisation with corrections for mass
- ATLAS calorimeter system is divided in 100 slices in η
  - In each slice we have 15 energy points from 256 MeV to 4 TeV (in powers of two)

Limited by ATLAS Fast Sim. historical strategy

- For each particle and η slice we define a GAN → 300 GANs
- Train the GANs on voxelised hits
  - Because cell structure is not homogeneous and would require different GAN architectures, the voxelisation reduces the differences and allows for a single, generic, GAN structure
- Select best epoch based on total energy distribution of each sample

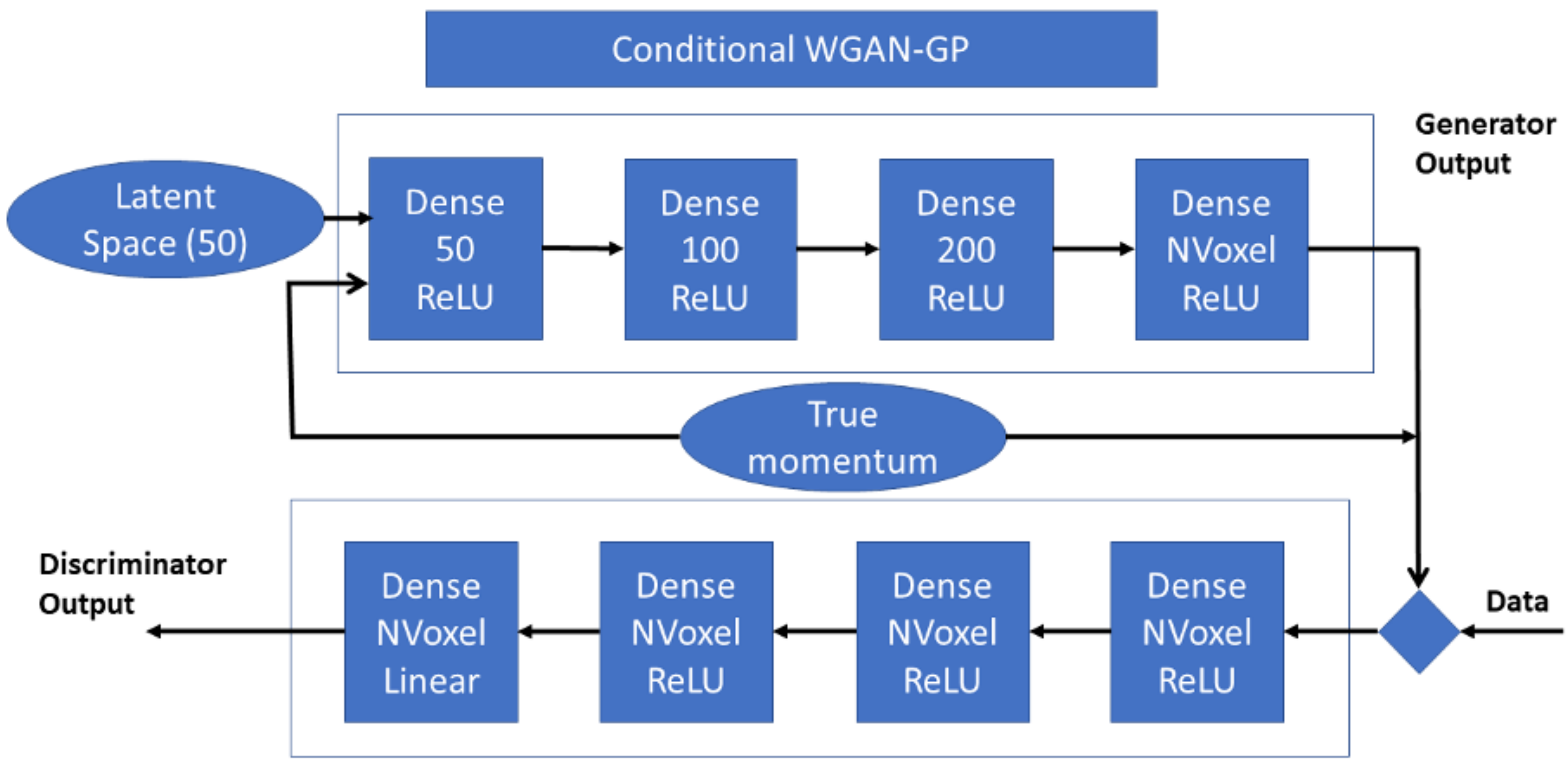
Machine learning part

- Convert the selected generator into LWTNN
- Simulate hits in Athena inverting voxelisation
- Compare with Geant4 using high-level observables for single-particle and di-jet samples after reconstruction

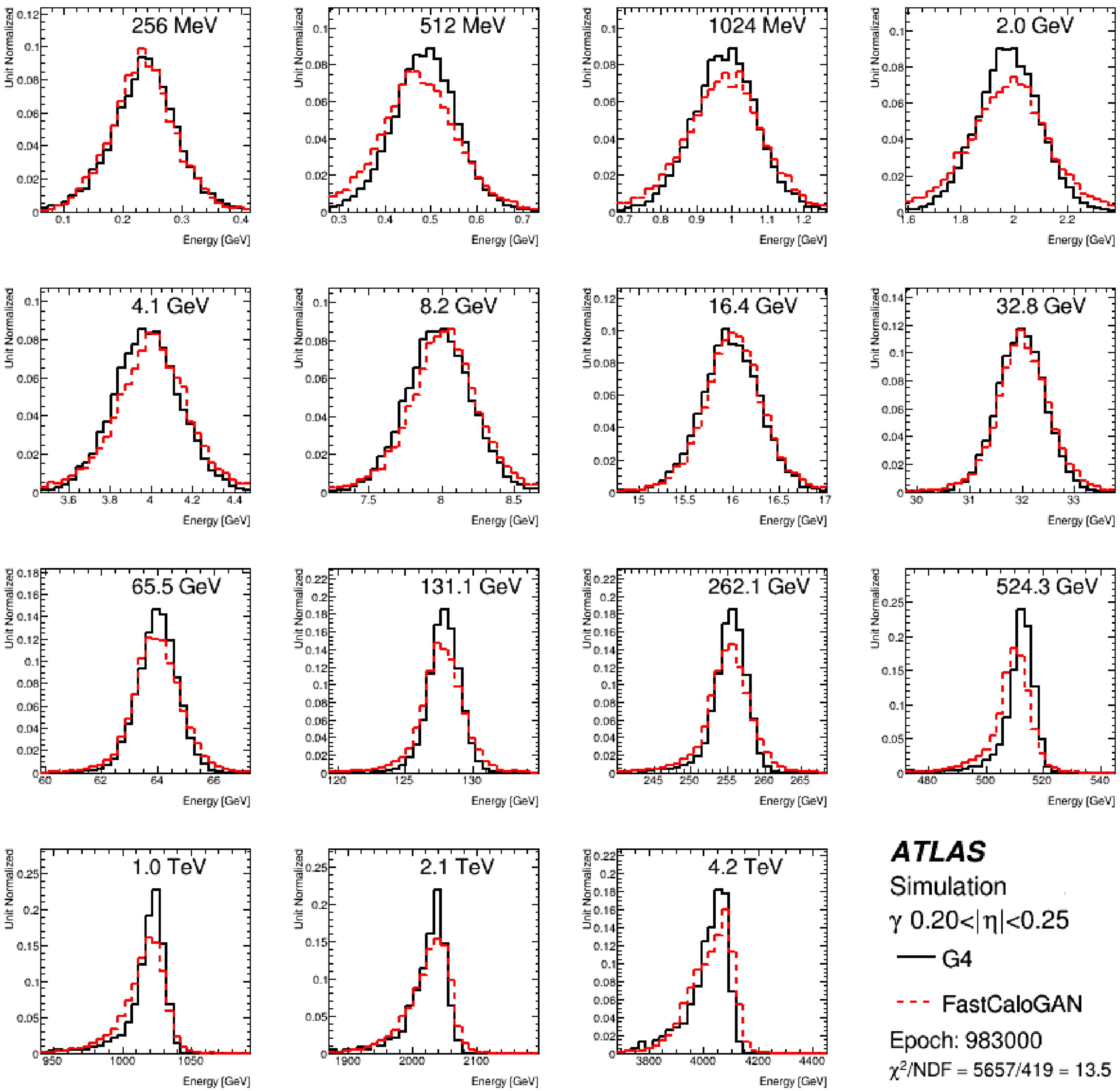
Can be used by ATLAS

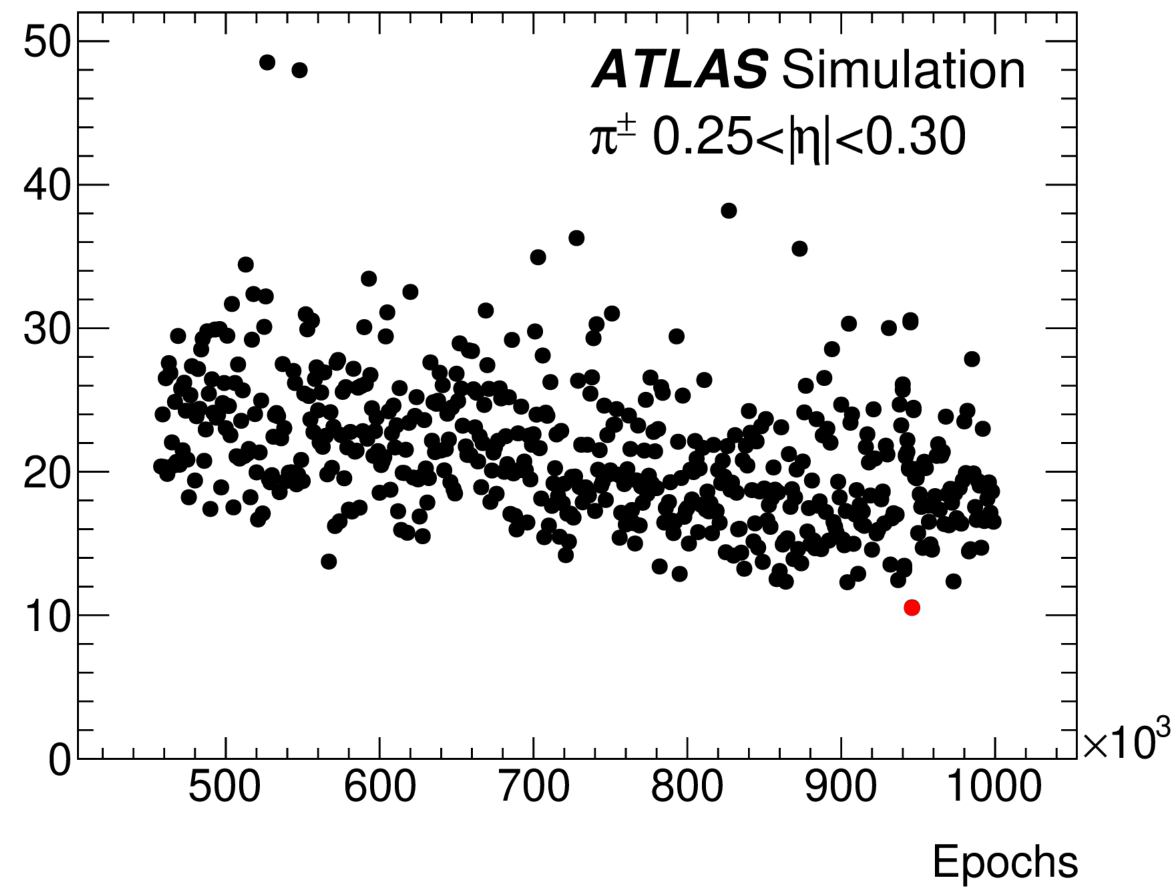
FastCaloGAN

- 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

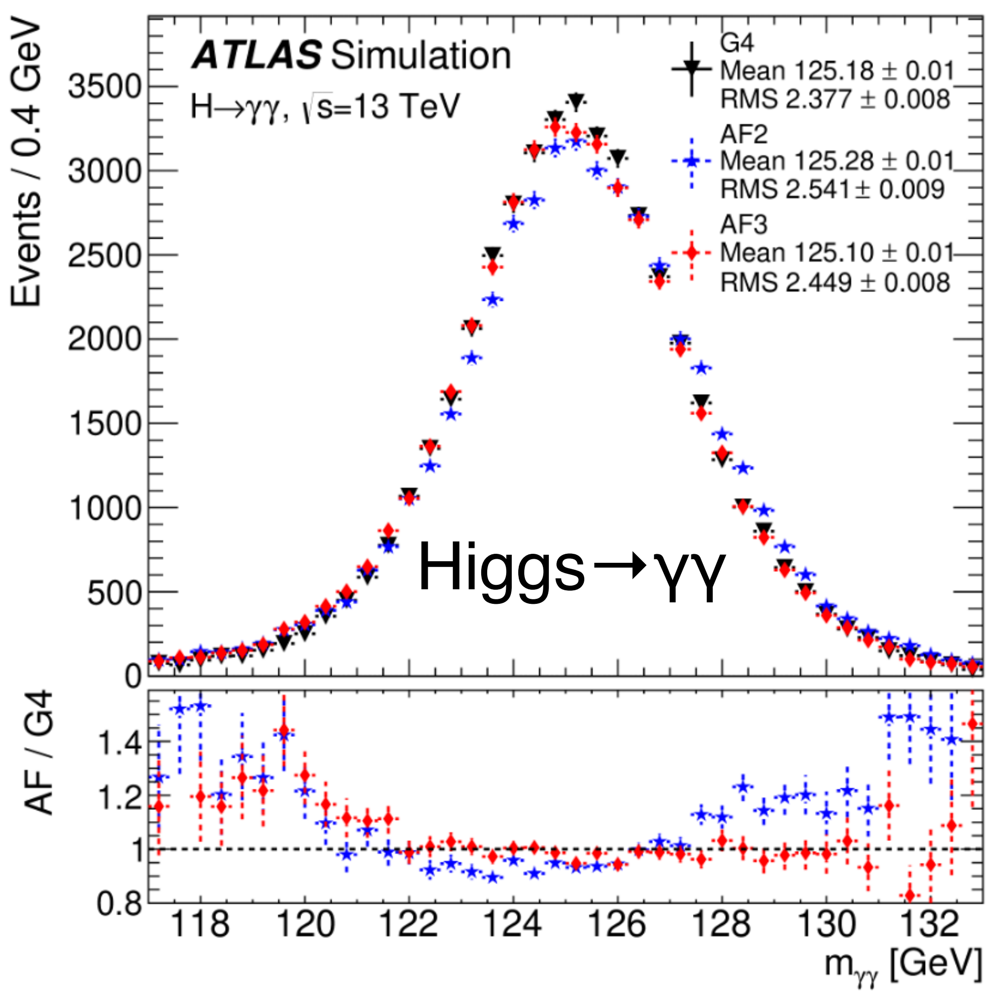


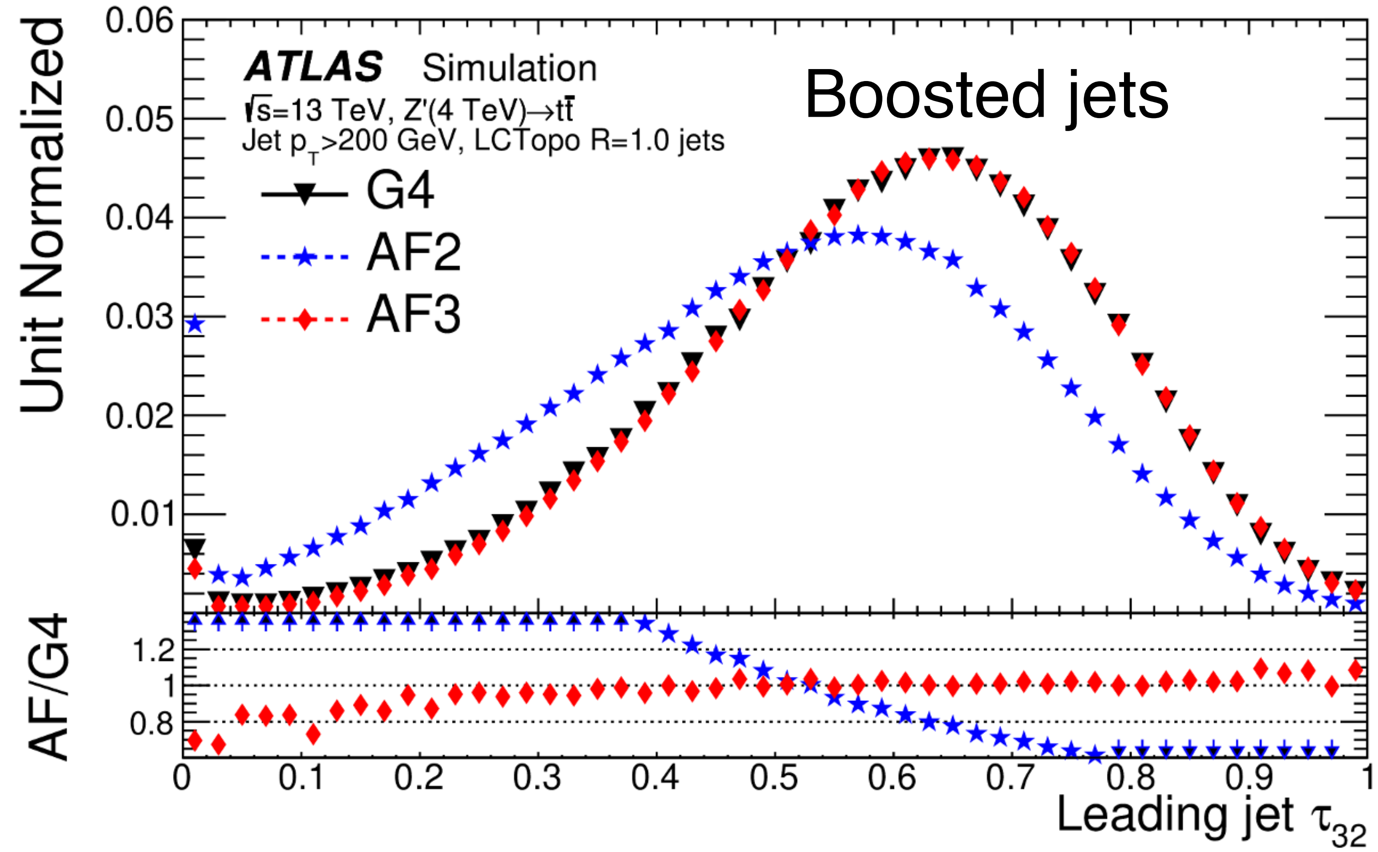
Training

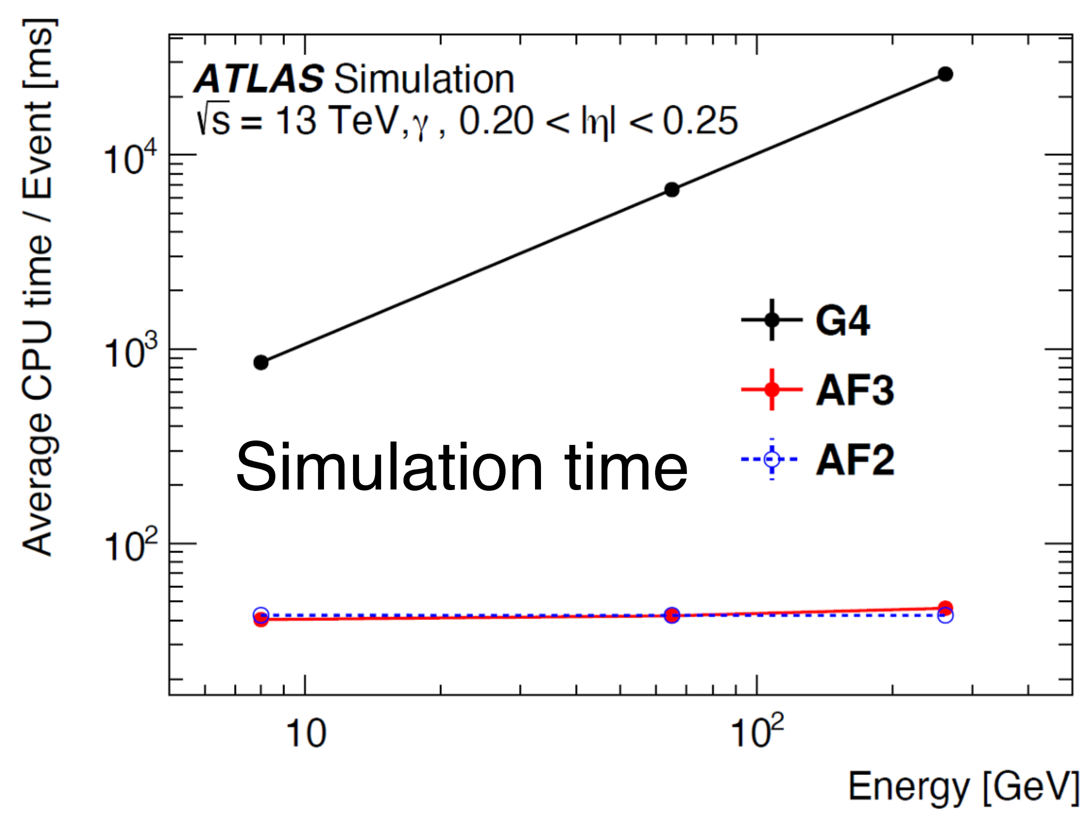

256 MeV
512 MeV
1024 MeV
2.0 GeV
4.1 GeV
8.2 GeV
16.4 GeV
32.8 GeV
65.5 GeV
131.1 GeV
262.1 GeV
524.3 GeV
1.0 TeV
2.1 TeV
4.2 TeV
Unit Normalized
Energy [GeV]
ATLAS
Simulation
γ 0.20<|η|<0.25
G4
FastCaloGAN
Epoch: 983000
χ²/NDF = 5657/419 = 13.5


ATLAS Simulation
π± 0.25<|η|<0.30
50
40
30
20
10
0
500
600
700
800
900
1000
×10³
Epochs

AtlFast3 performance







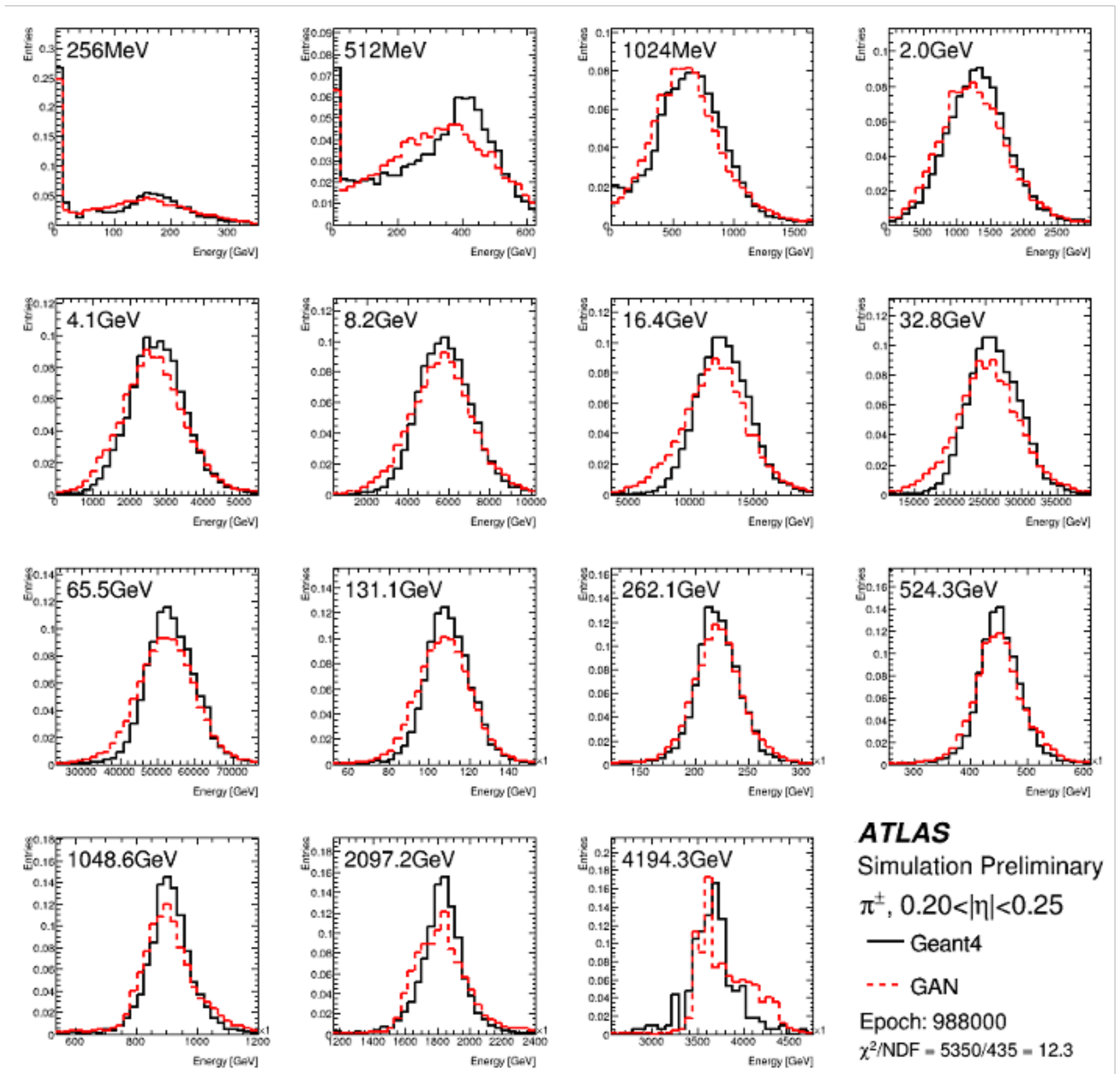
Improvements for Run3
FastCaloGAN V2

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

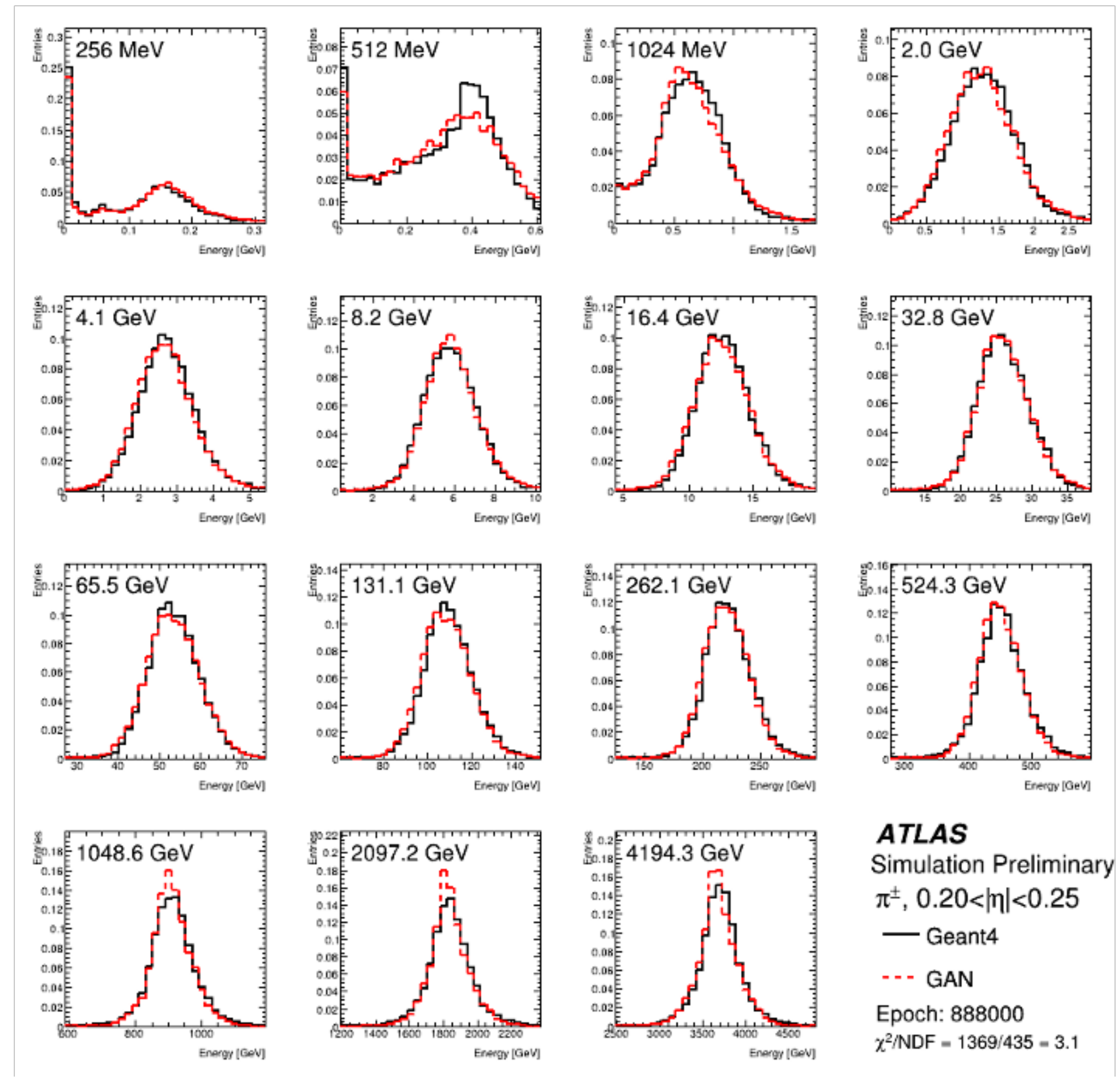
Best GAN for pions: V1 vs V2

V1



χ^2/NDF = 12.3

V2

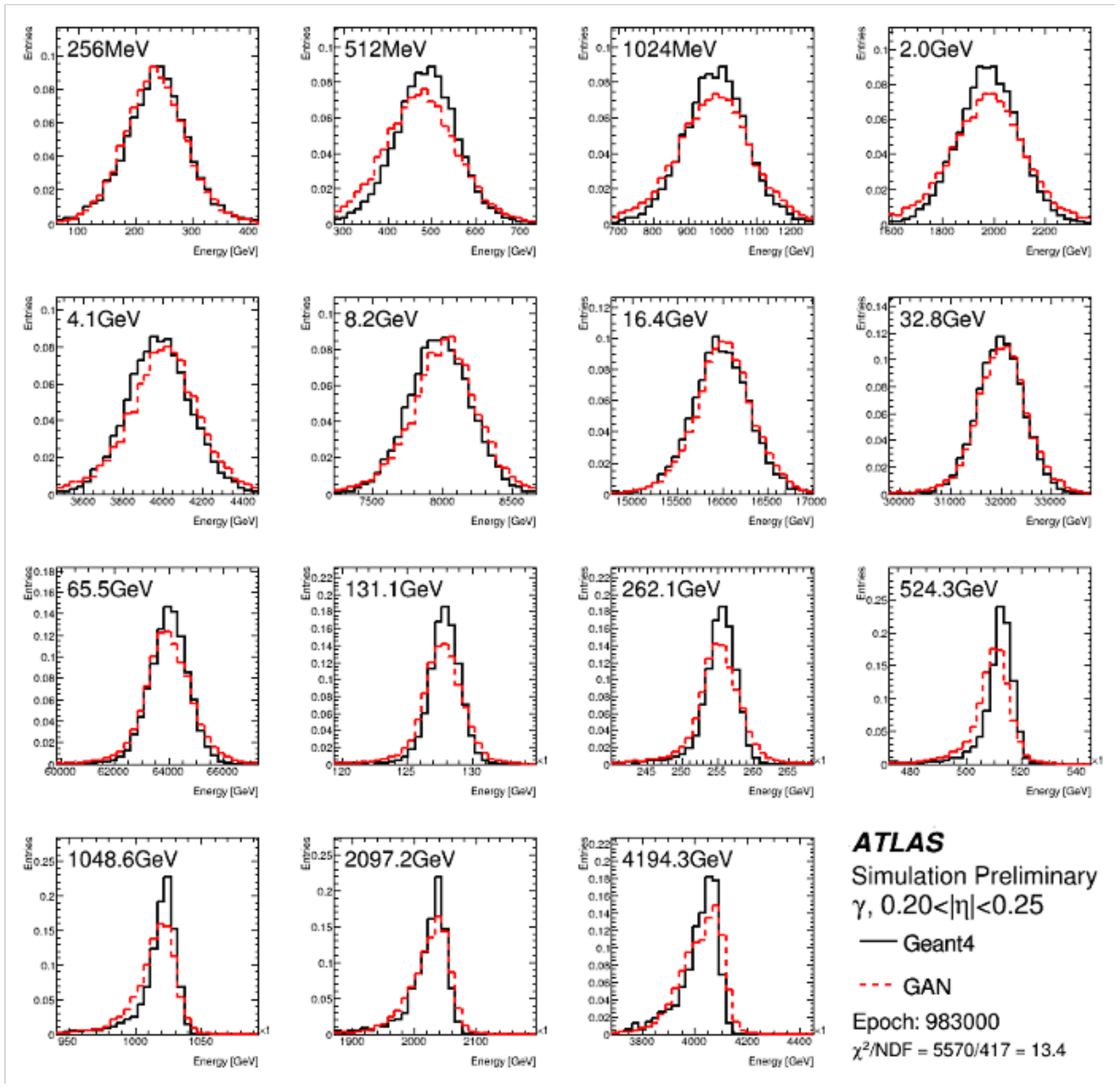


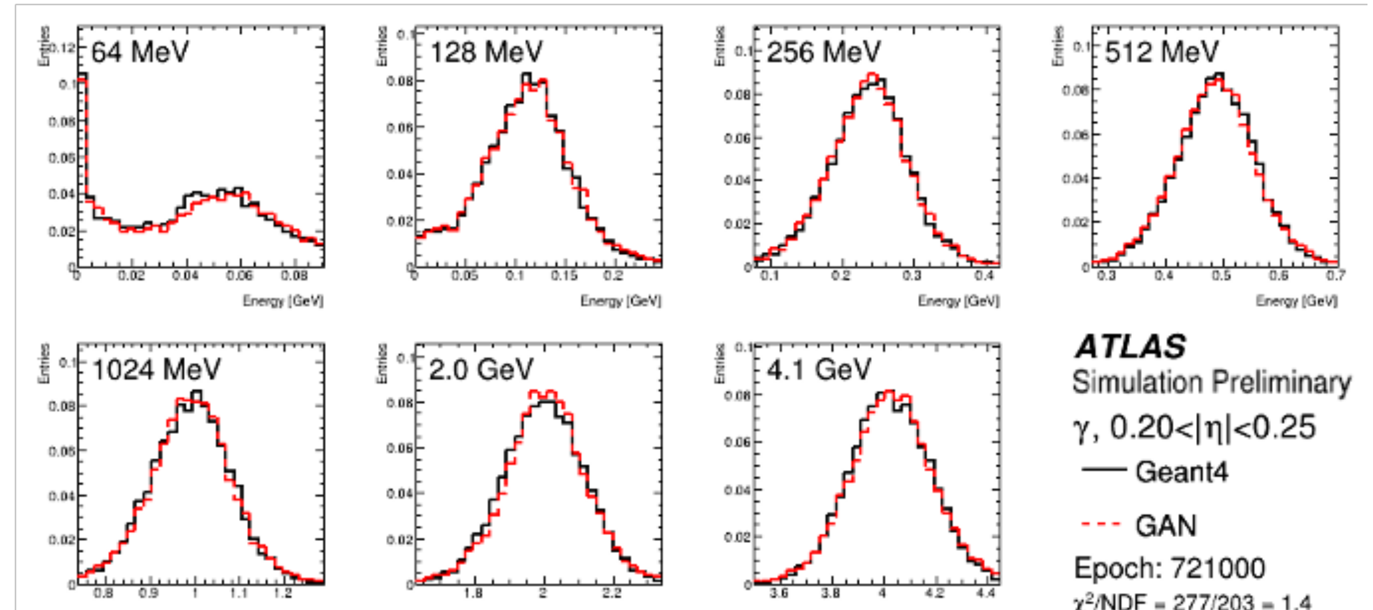
χ^2/NDF = 3.1

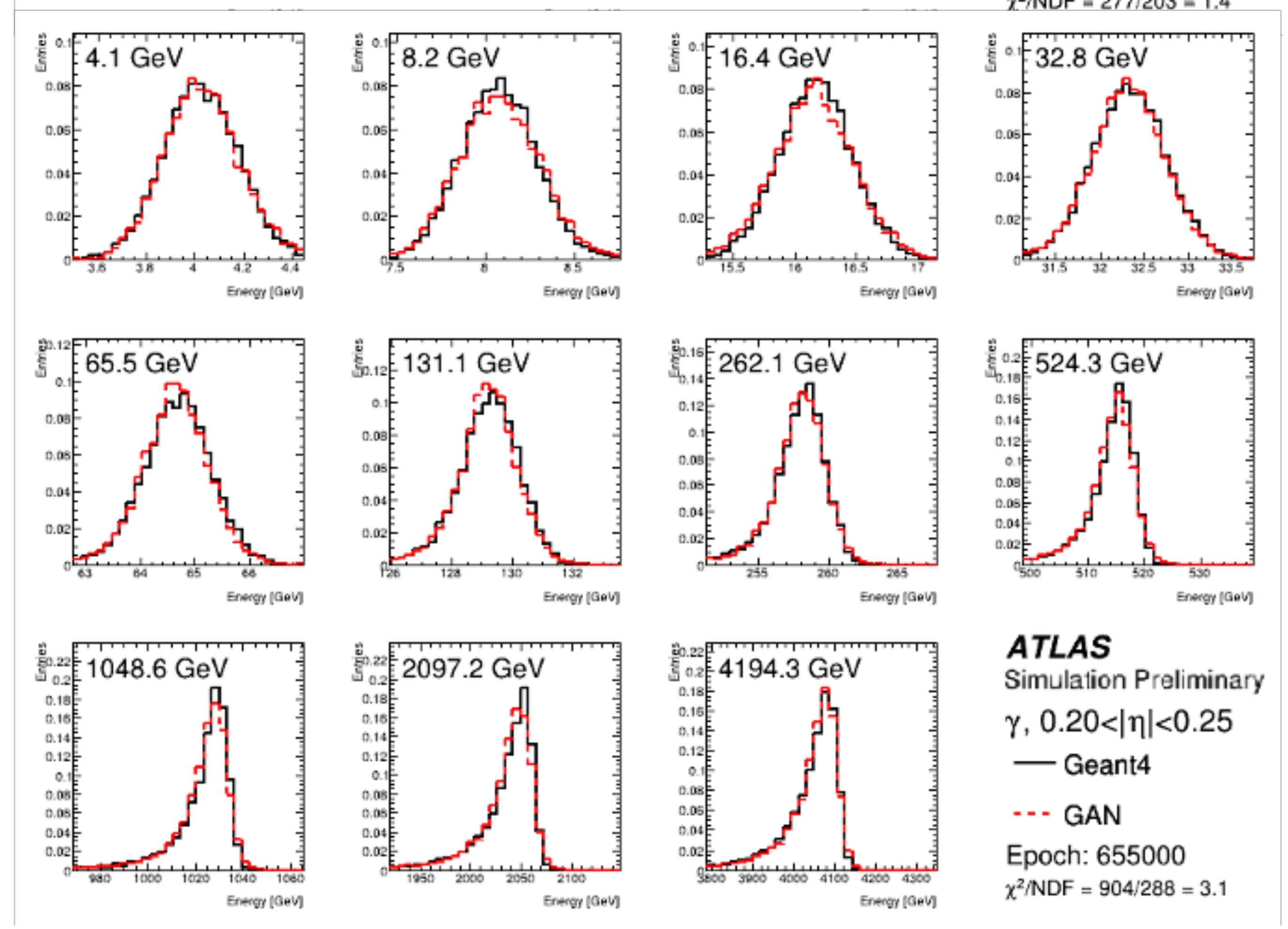
Best GAN for photons: V1 vs V2

V2

V1





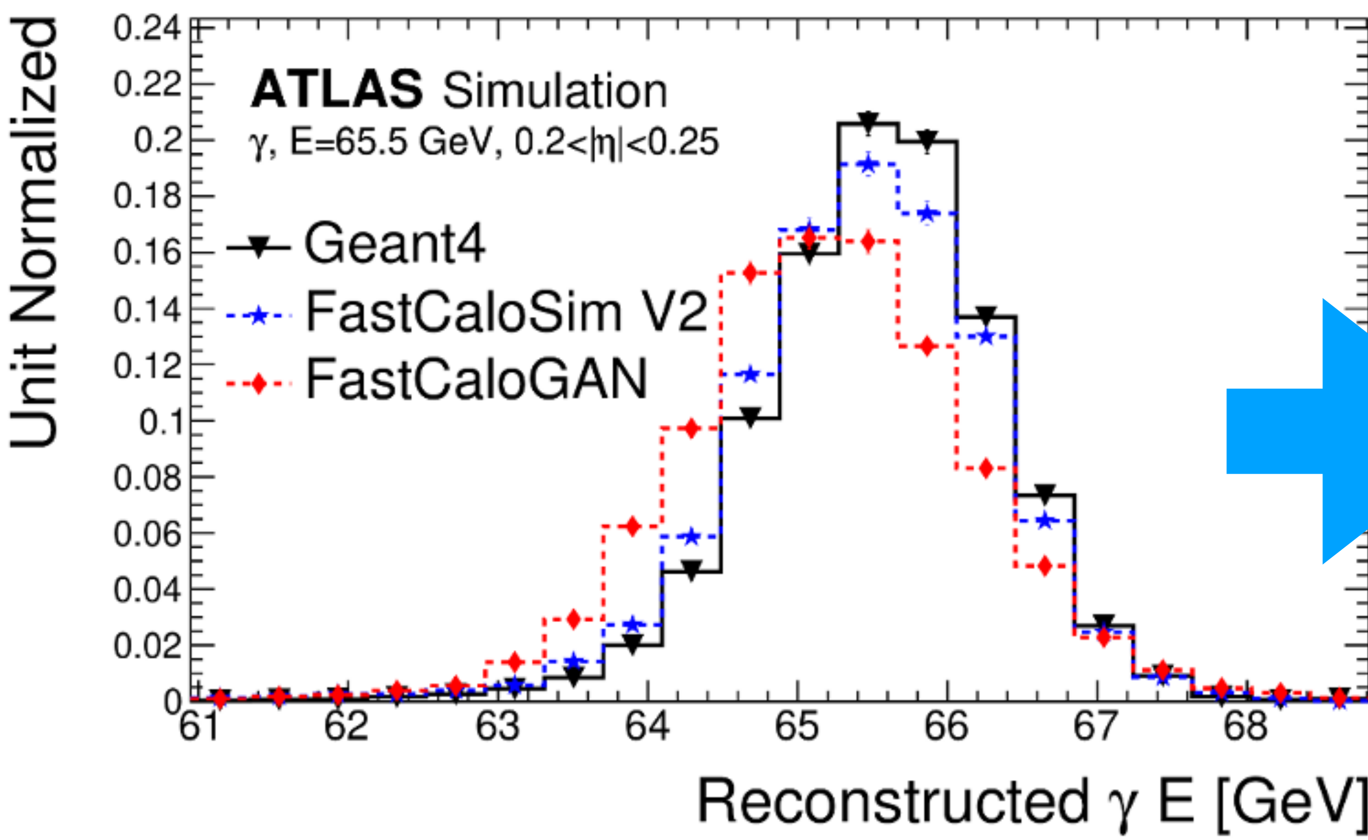


χ^2/NDF = 12.3

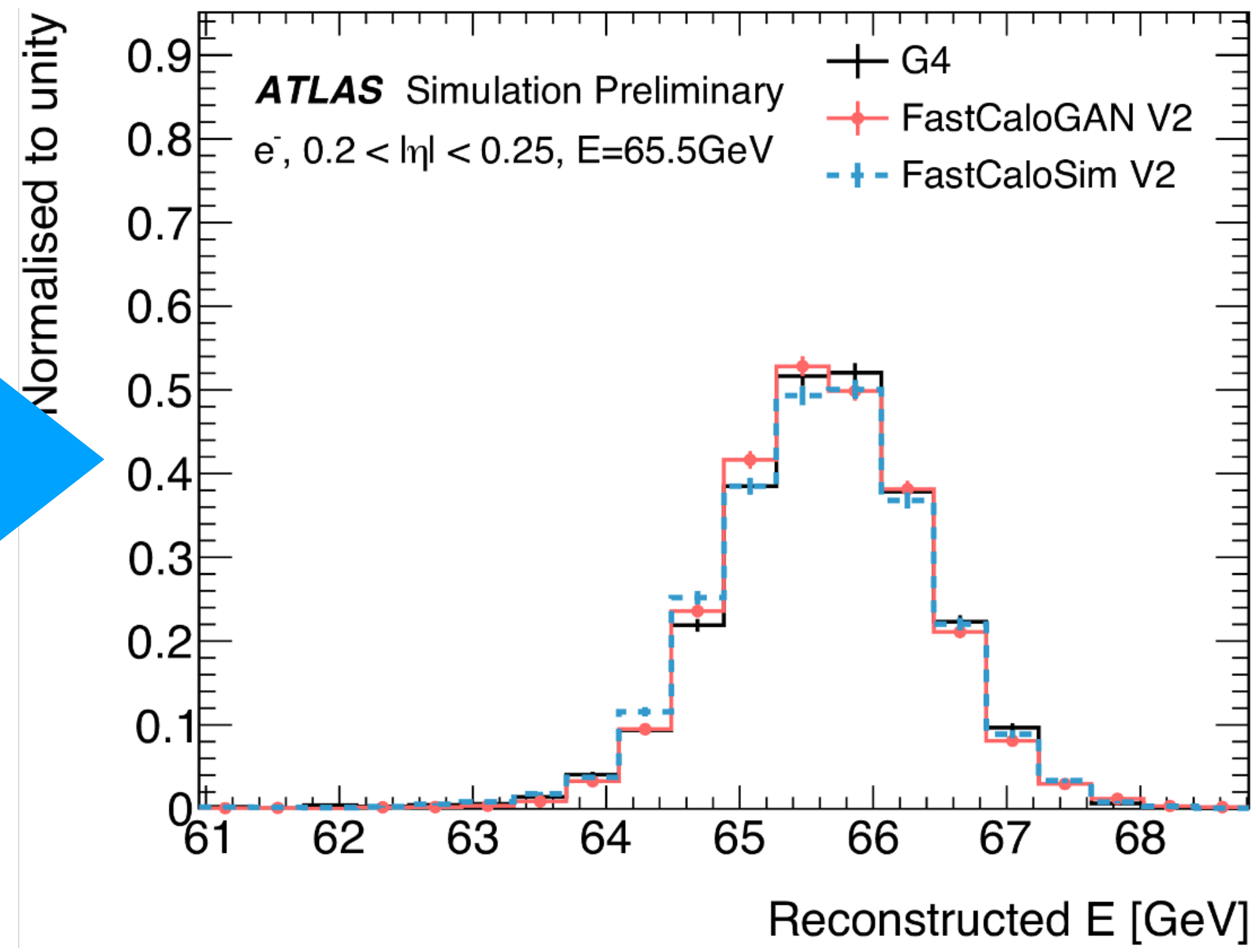
Single particle performance: V1 vs V2

- Photon energy

V1



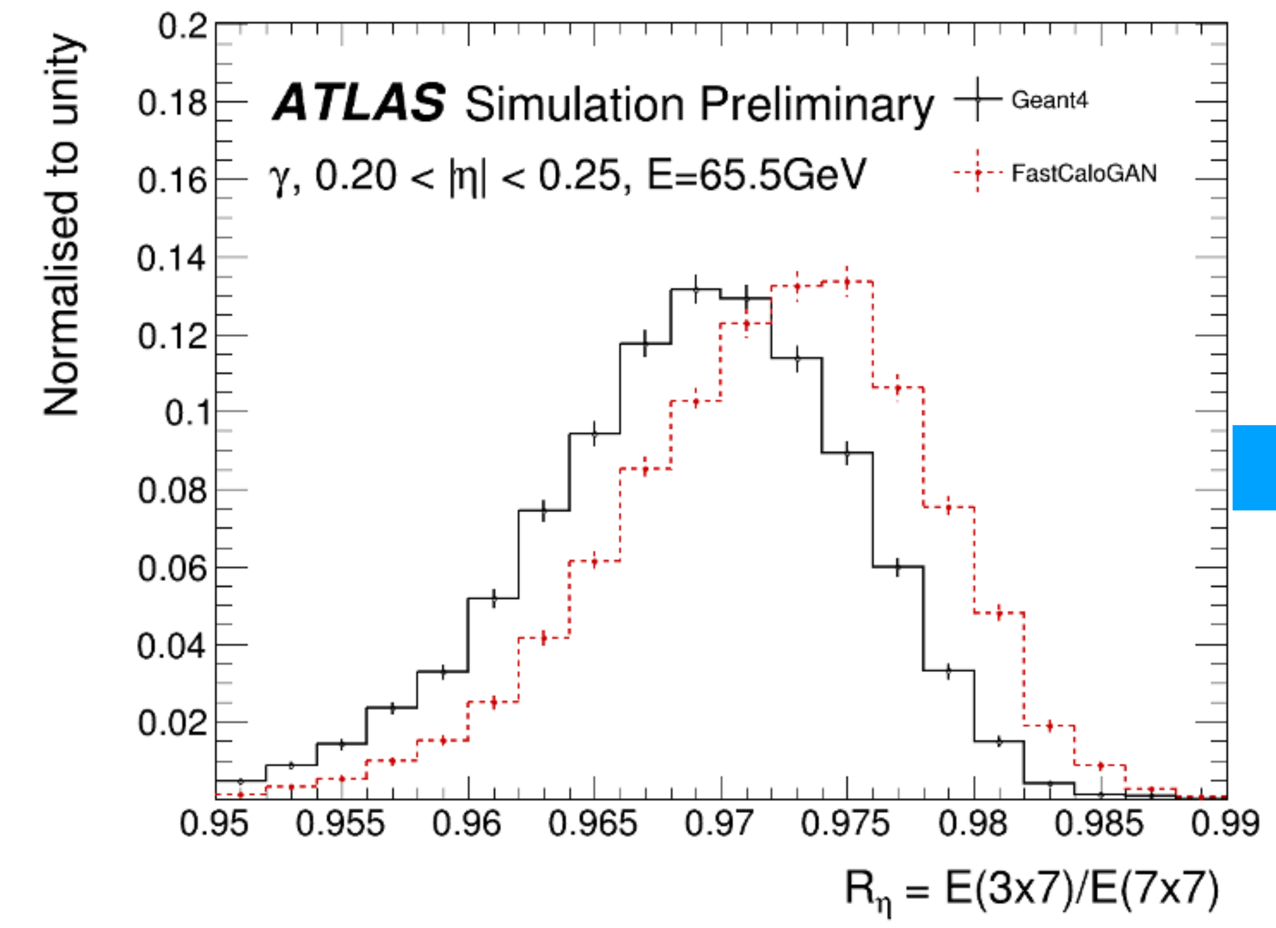
V2



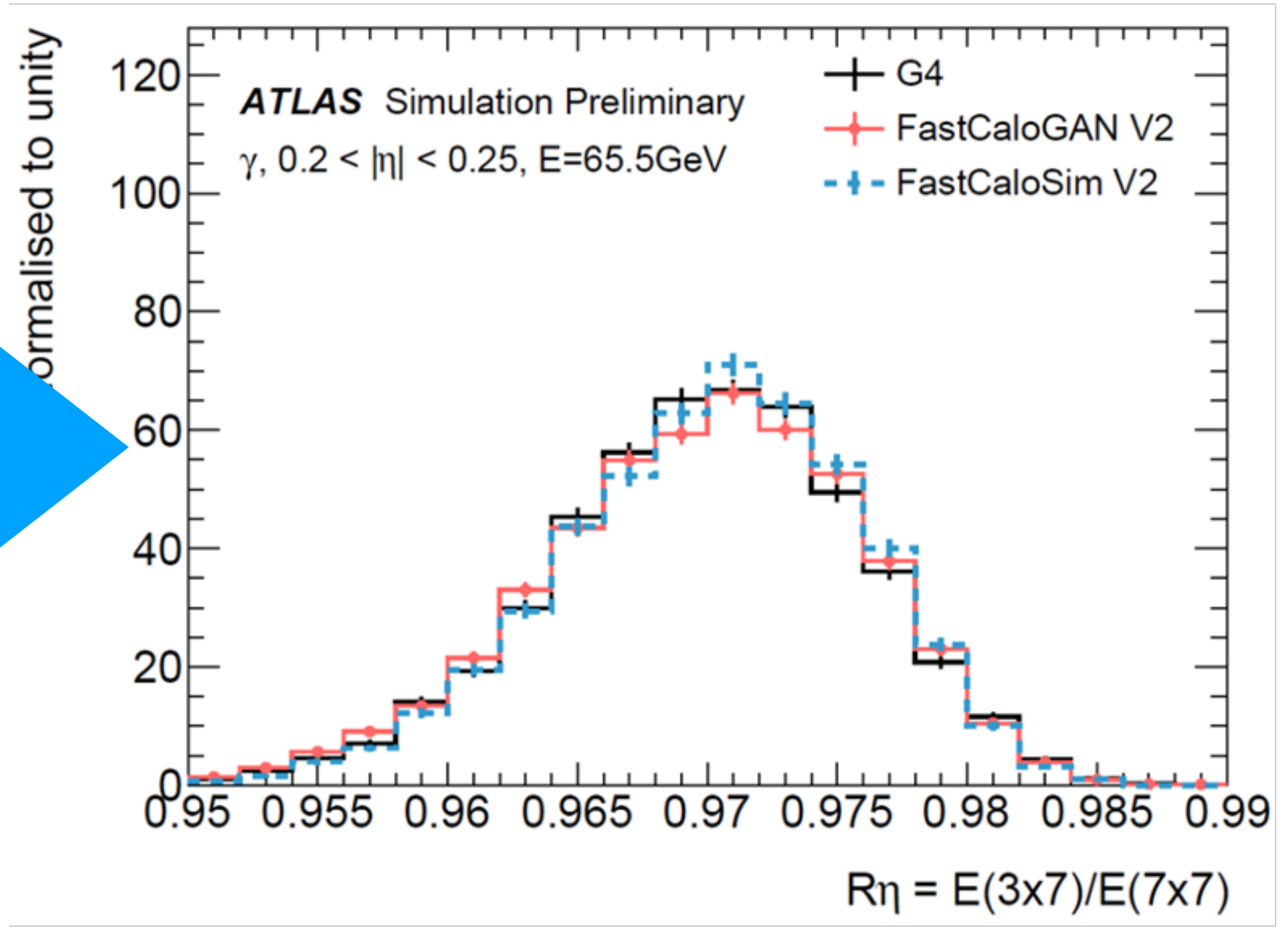
Single particle performance: V1 vs V2

- Photon shower shape

V1



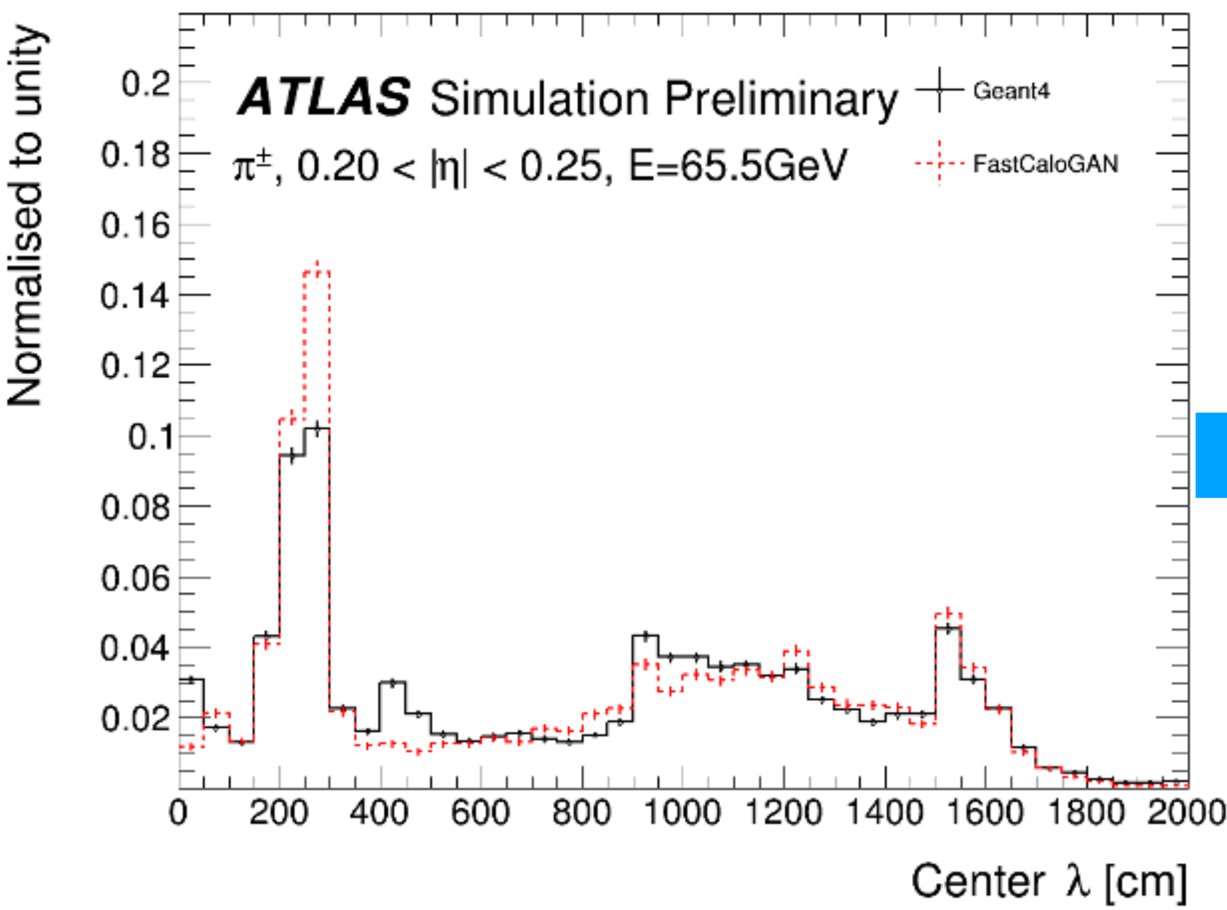
V2



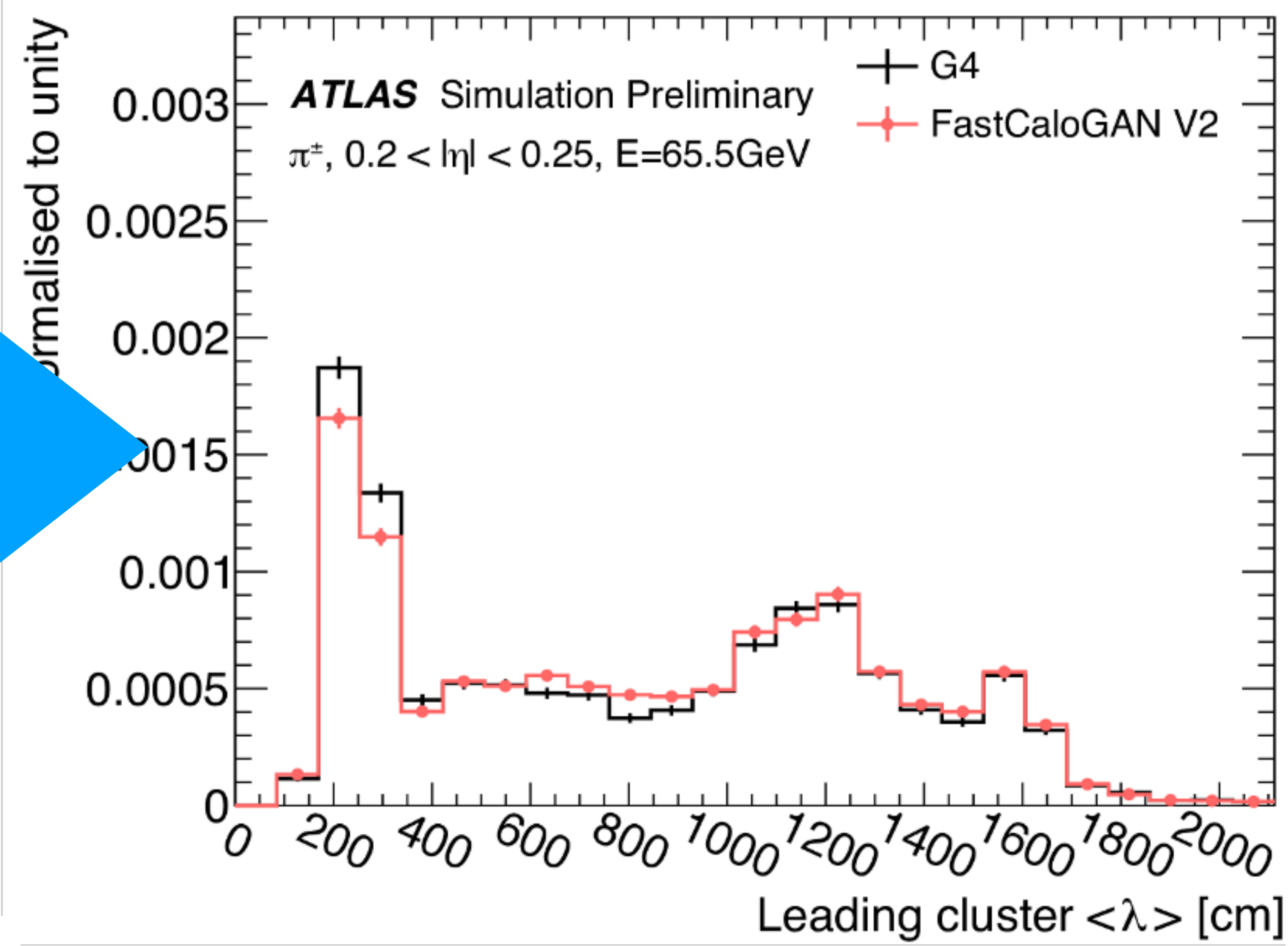
Single particle performance: V1 vs V2

- Pion shower position

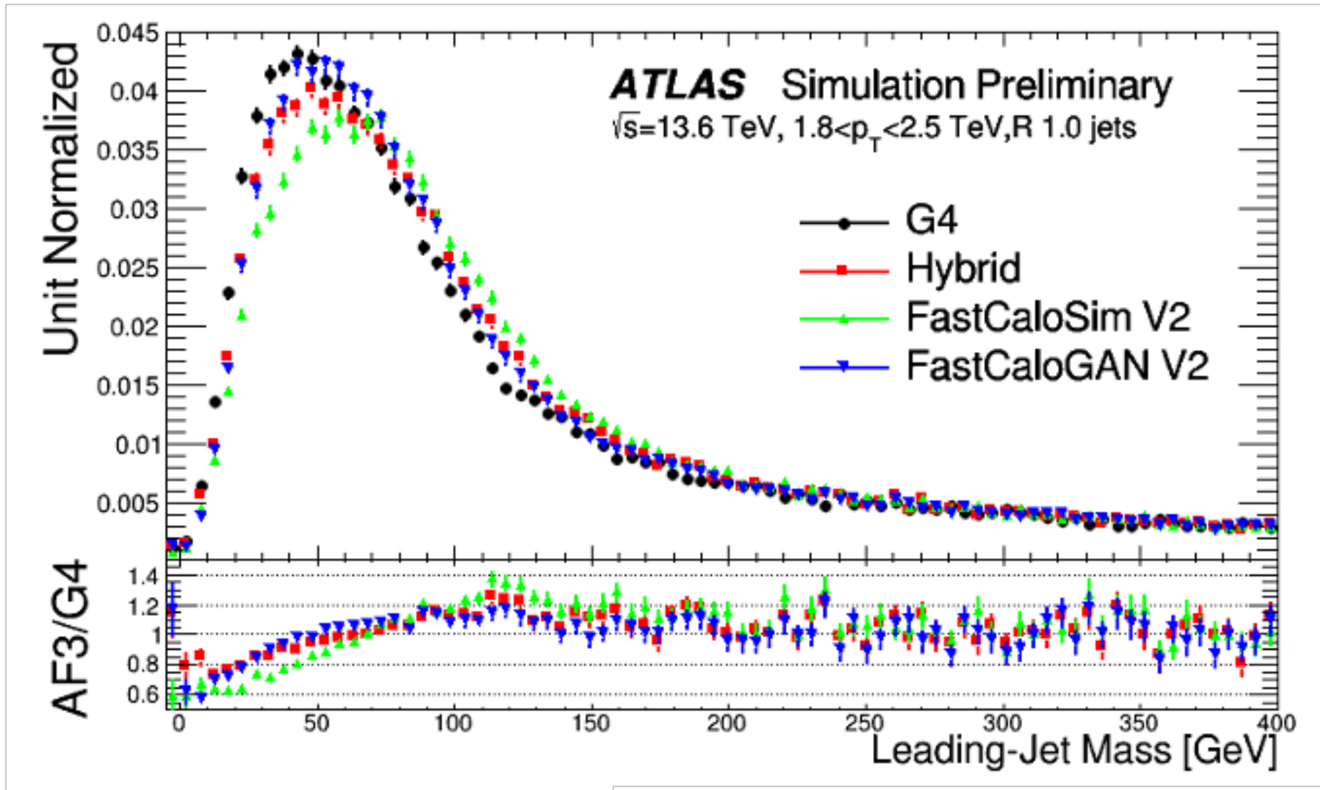
V1

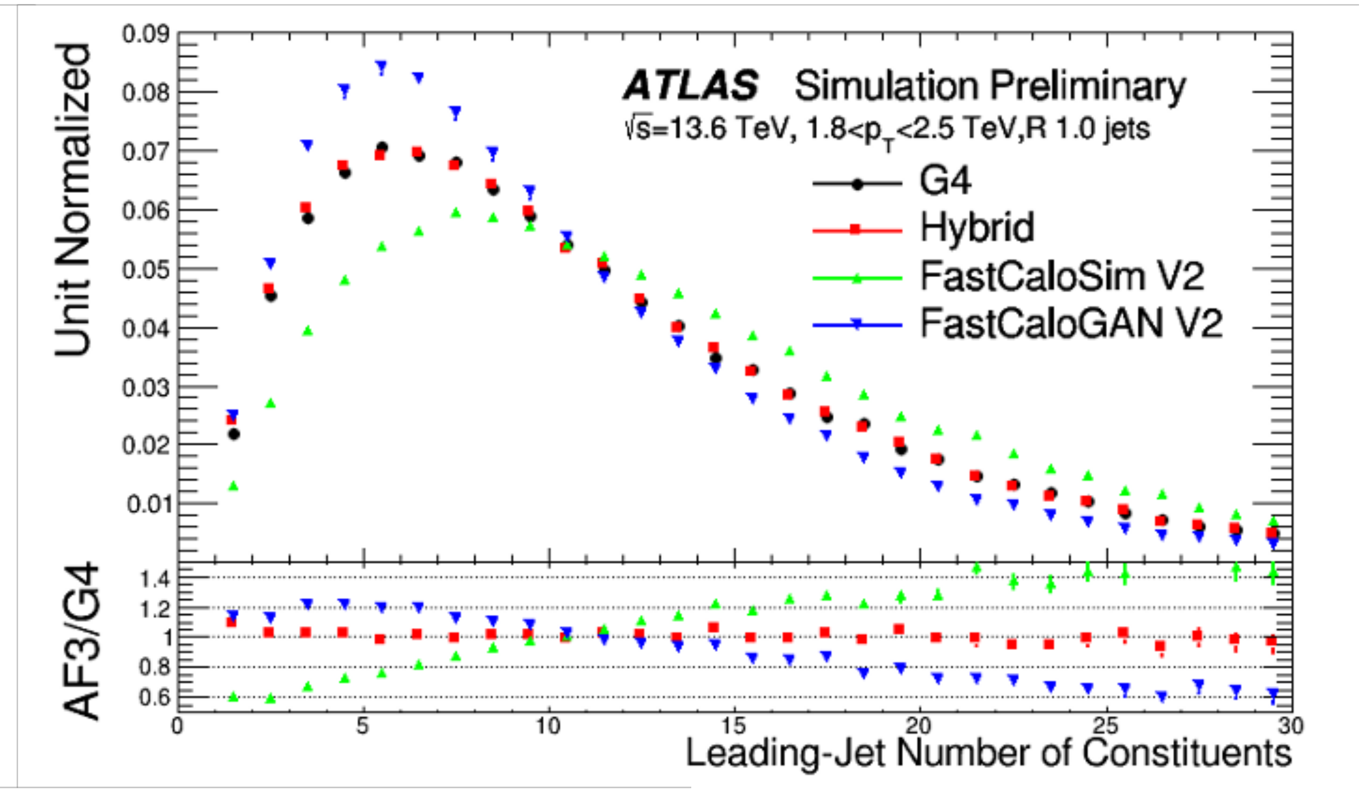


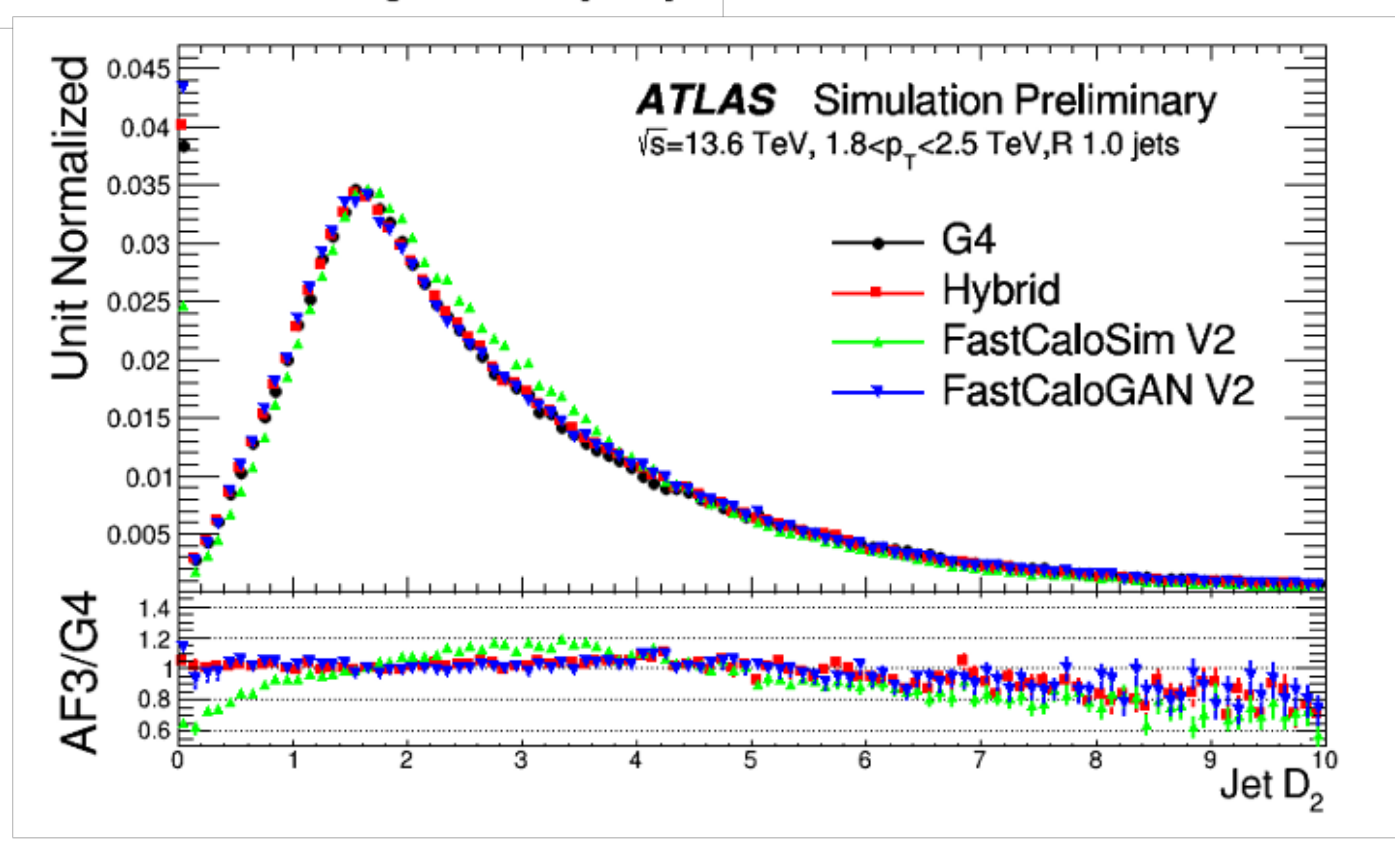
V2



Jet performance

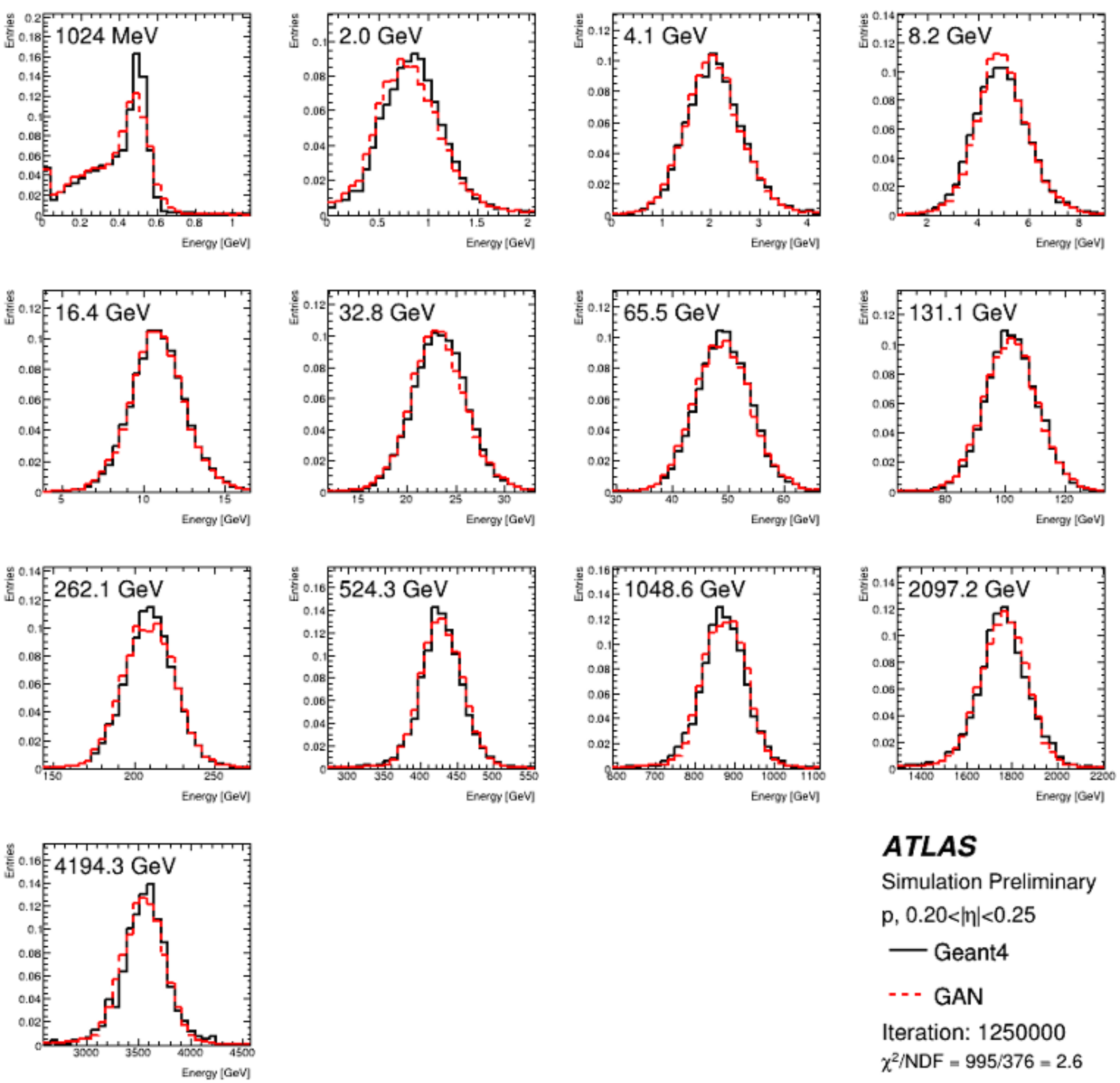


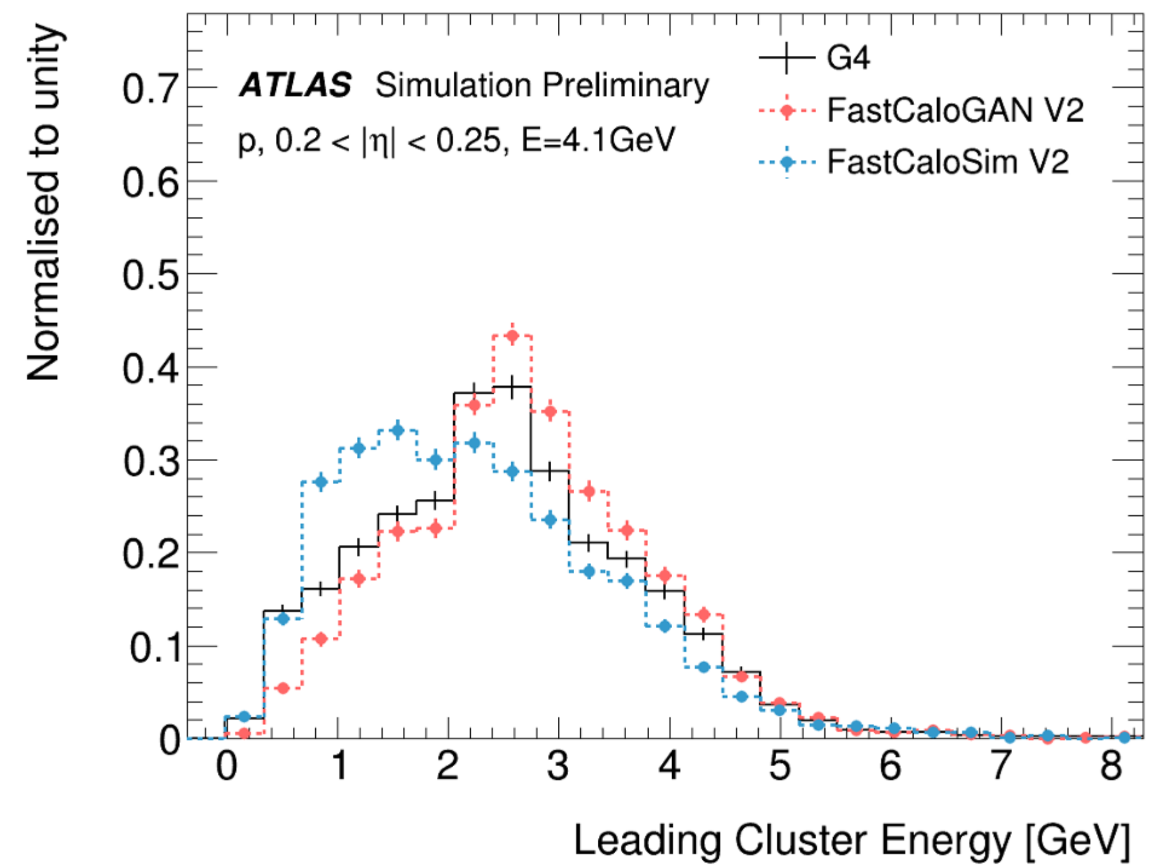




Introduce proton GAN

- FastCaloGAN performance is much better than FastCaloSim V2





FastCaloGAN V1 legacy

- The GAN code is now public on zenodo
- The datasets used to train two GANs (one pion and one photon) were also made public in the CERN OPENData
- This will allow to get the wider community help ATLAS with better generative models
- The datasets are used in the #calochallenge, the latest ML challenge proposed to the HEP and ML communities

CaloChallenge and CaloShowerGAN

Current landscape

- A dedicated workshop to discuss what worked and what did not
  - 44 registered, 25 in person
  - 12 contribution divided in 4 categories

GANs
Generative Adversarial Networks

VAEs
Variational AutoEncoder

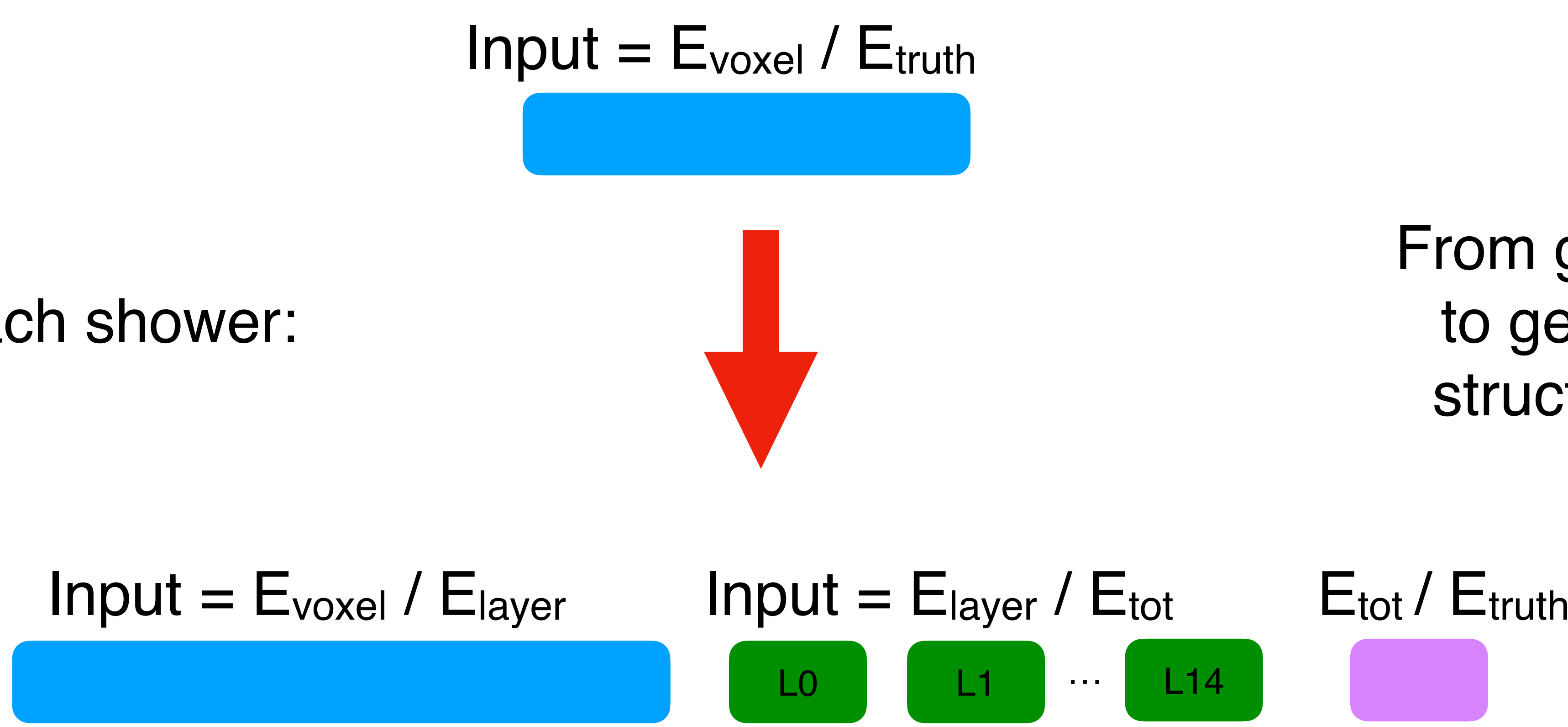
NFs
Normalising Flows

DFs
Diffusion models

CaloShowerGAN

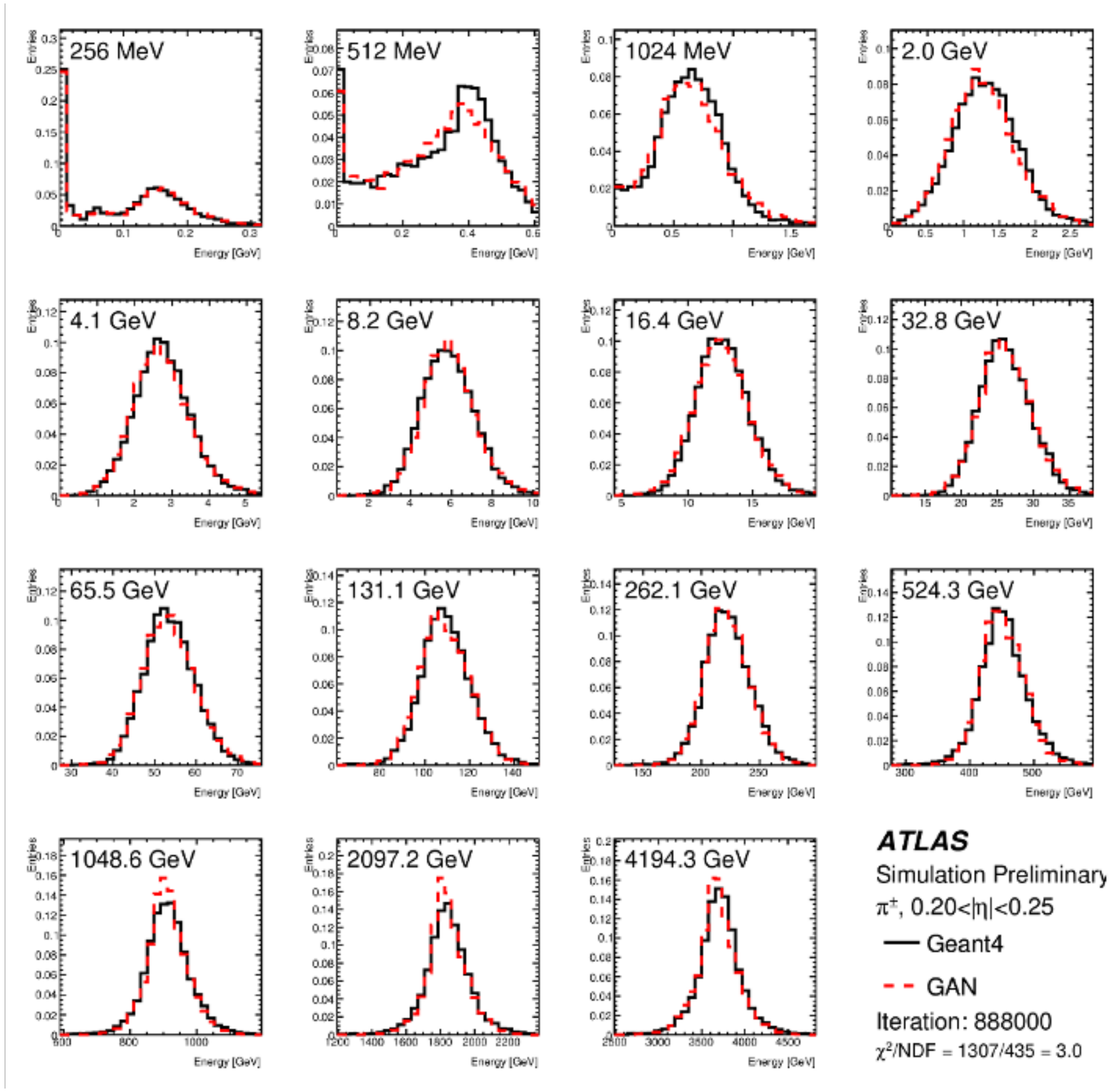
- Further hyperparameter and architecture optimisations than FastCaloGANv2
- Exploits a new normalisation; instead of normalising all voxels by the nominal energy, the normalisation is done by layer. This is the method employed by the VAE method that had the best performance at the CaloChallenge Workshop

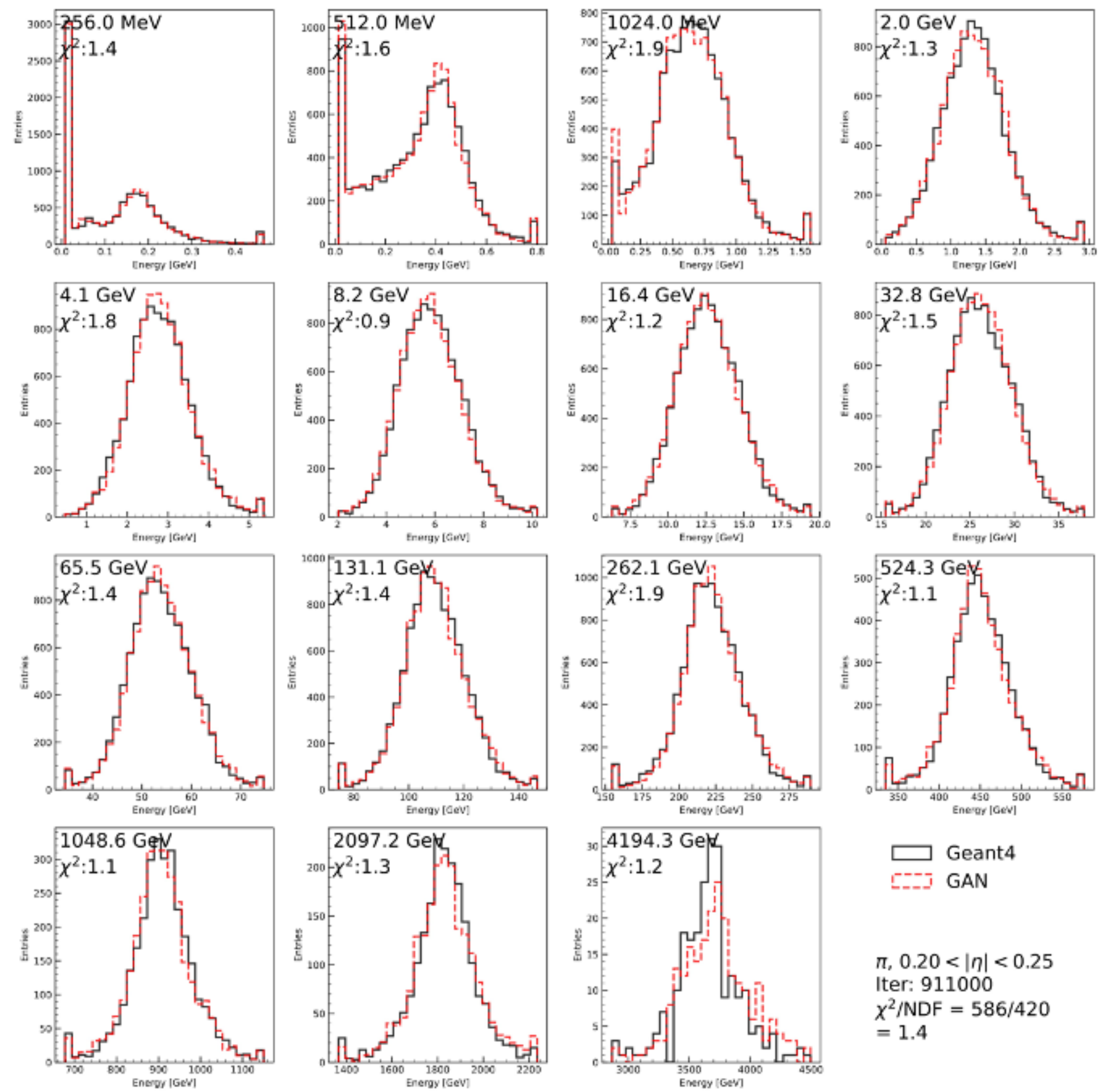
Each shower:



From generating all voxel energies to generating more constrained structure of a vector of numbers

Pions





χ^2/NDF = 3.0

χ^2/NDF = 1.4

Conclusion

- AI for simulation has now entered the realm of production
  - It has the advantage of a few magnitude speedup
- FastCaloGAN V2 showed large improvement from the already deployed V1
  - This is crucial to allow a wider use of fast simulation required to match the designed high luminosity in Run3 and beyond
- Other techniques besides GAN also show promising performance
  - The #calochallenge and the workshop bring together diverse techniques and experts in generative modelling, providing a comprehensive overview of various approaches
- The trend involves increasingly substituting "traditional" methods with AI, harnessing the untapped potential within the AI/ML approaches.

Thank you for your attention!