Fellini
Fellowship for Innovation at INFN

H2020 MSCA COFUND
G.A. 754496

INFN
Sezione di
Roma Tor Vergata

Reusability

Michele Faucci Giannelli,
On behalf of the ATLAS collaboration

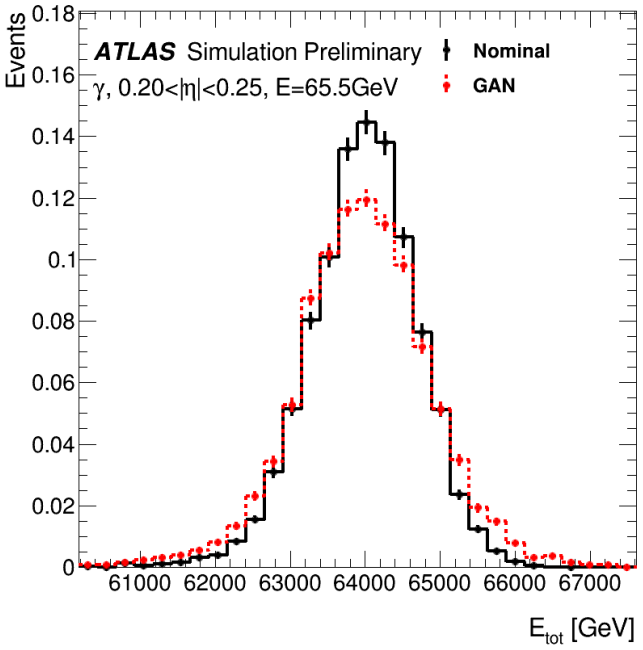
LPCC Fast Detector Simulation
23-11-2021

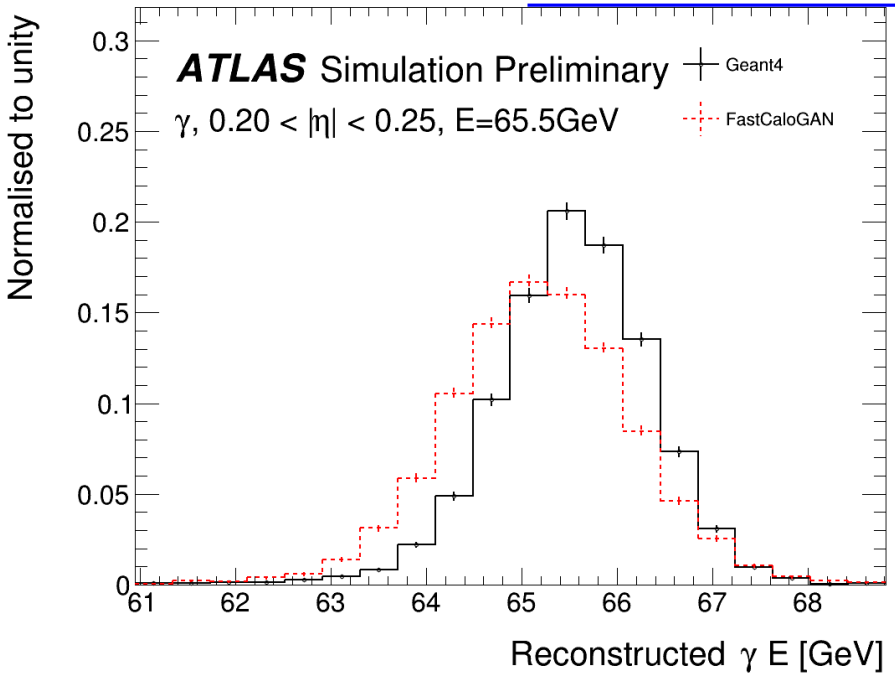
Lessons learned

- Are there any valuable lessons to be learnt by others?
  - Fast simulation is difficult, years are required to achieve high performance and many small effects, very experiment specific, must be considered
  - ML tools are now mature and can be considered as valid alternatives to classical methods
  - So far, ML tools were limited more by physics effects that required time to be understood than ML itself (models, HPO, pre-processing strategy).
  - More in general, validation can require a long time (ML training time << validation time)

Example of physics limitation

ATL-SOFT-PUB-2020-006

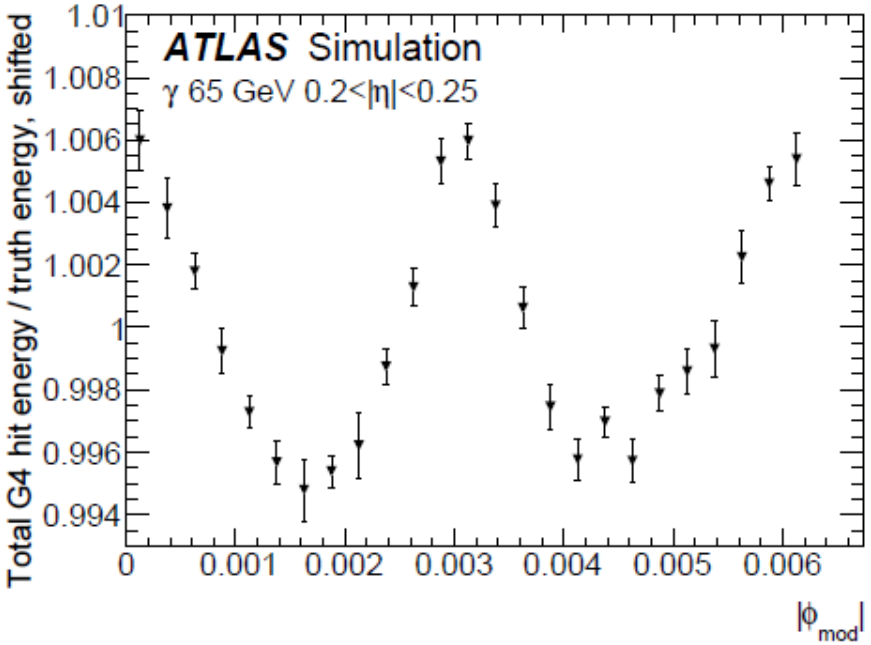


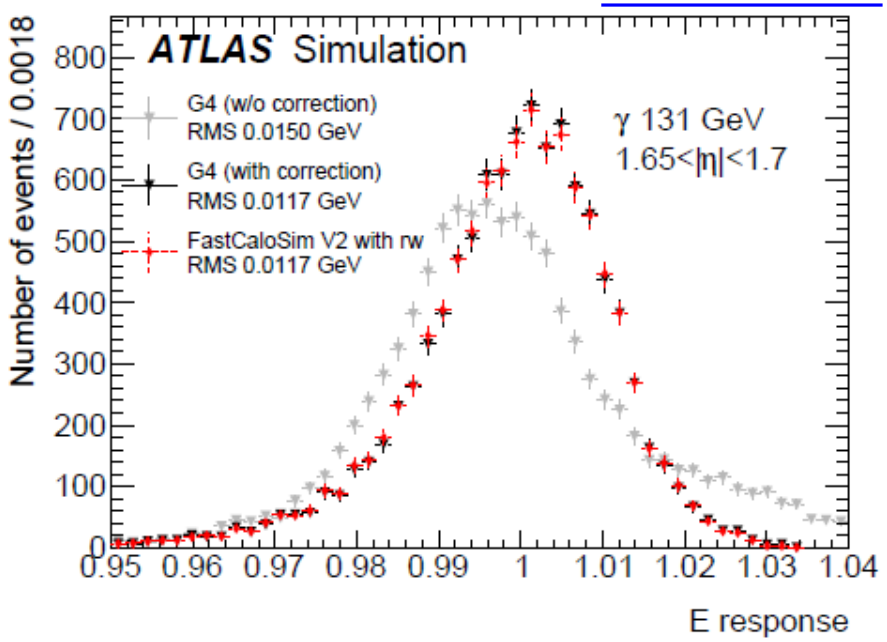


- GANs learn the input distribution correctly but after simulation and reconstruction a shift in energy was observed
- Problem identified in an missing energy correction that needs to be applied to the training sample

Accordion effect

SIMU-2018-04





- The accordion structure result in an energy response that depends on the phi position of the shower
- This result in a correction that must be applied to the training sample so that a model (FCSV2 or GAN) can produce the correct energy response

Re-use

- Tools which can be reused?
  - The use of voxelisation, which is detector dependent, makes both FCSV2 and ML tools rather independent on the detector so they could be re-used by other experiments
    - The voxelisation and the inference would need to be detector specific

Unsuccessful ideas

- Maybe some ideas have failed for your detectors (or parts of it) and it can be useful to others?
  - Training a model on the lateral shape distributions for each layer/pca instead of saving them as 2D histogram did not work
  - Voxels perform better than cells:
    - GAN at cell level did not perform as well as FastCaloGAN using voxels
    - VAE also saw a significant improvement going from cells to voxels
  - Normalisation is crucial, without a normalisation strategy no approach works (FCSV2, GAN or VAE)
  - Conditioning GANs on phiMod did not work, nor trying to have them predict the longitudinal position of the hits in a layer
    - Lesson: don’t make the model more complex than it should be and factorise different problems

Collaborations

- What are your plans, maybe there is room for collaboration?
  - We will keep developing classical and ML based fast calorimeter simulations
  - Classical tools and the approach developed for FCSV2 may be used by other experiment, especially those focussing on e/gammas
  - We are publishing the voxelised inputs files used to train the GANs used in AF3 on the CERN OpenData
  - The FastCaloGAN code is already available on zenodo, any contribution from the community is welcome
  - FCSV2 code is also in an open GitLab repository