H2020 MSCA COFUND
G.A. 754496

Sezione di
Roma Tor Vergata

Tuning

Michele Faucci Giannelli,
On behalf of the ATLAS collaboration

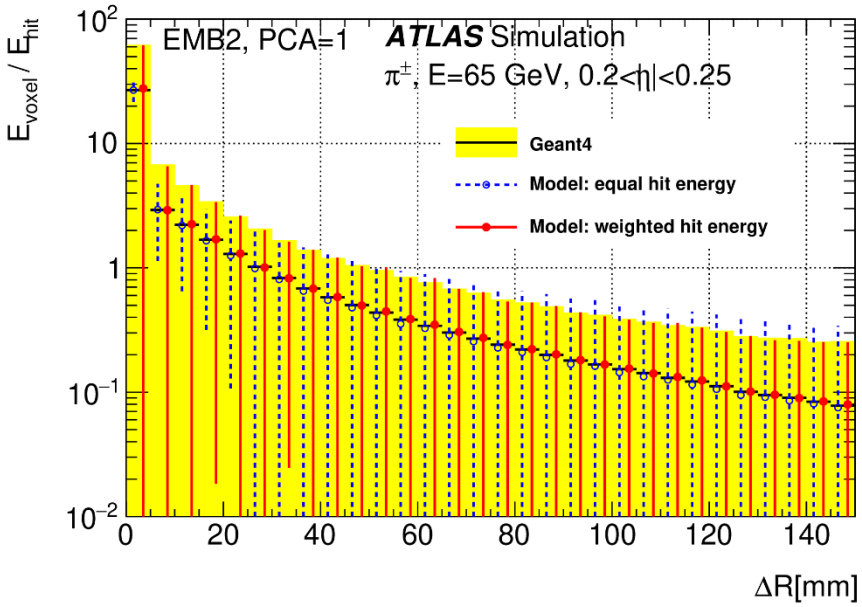
LPCC Fast Detector Simulation
22-11-2021

Tuning

- How the extraction and tuning of parameters for the fast simulation are done?

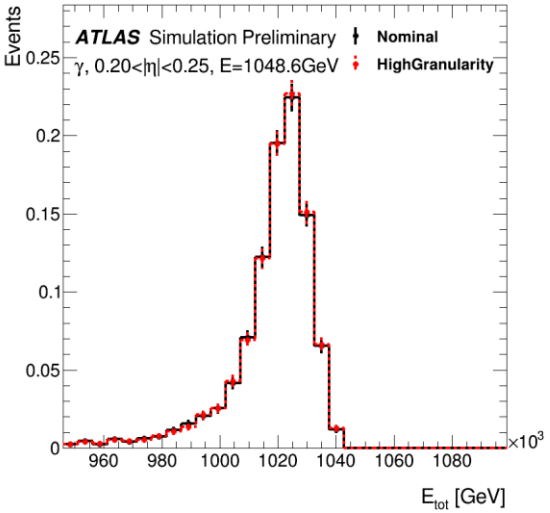
Tuning parameters: FCSV2

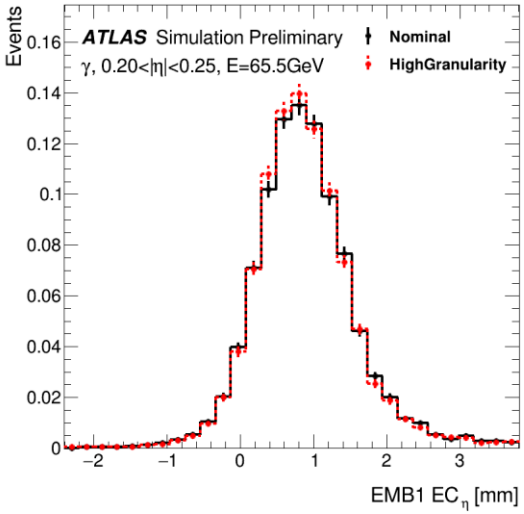
- In FCSV2 we can tune:
  - the number of PCA used to categorise the events
  - the number of hits produced in the simulation
  - the energy of such hits (used only for pions)



Tuning parameters: GAN/VAE

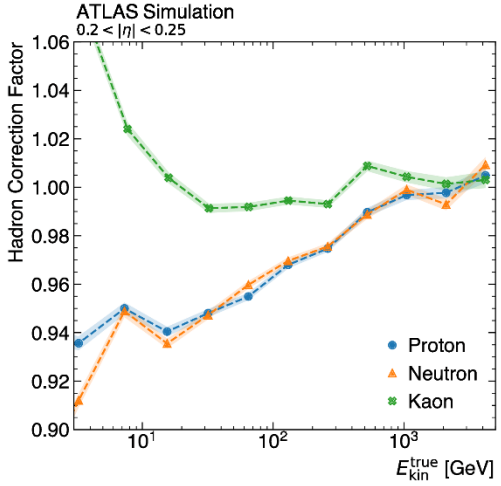
- In FastCaloGAN/VAE there are three sets of parameters:
  - Voxelisation
  - GANs hyper parameters, training strategy, etc
  - Hit generation strategy to map voxels to cells

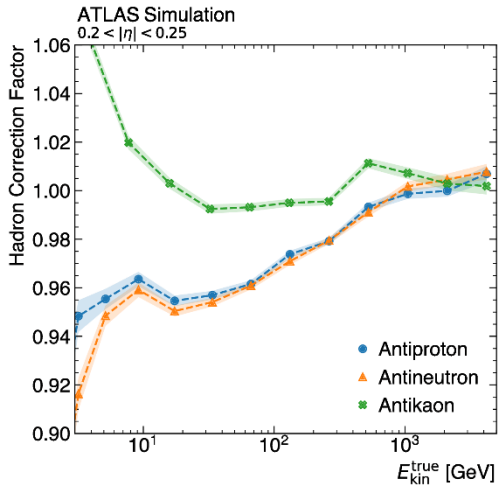




Tuning parameters: AF3

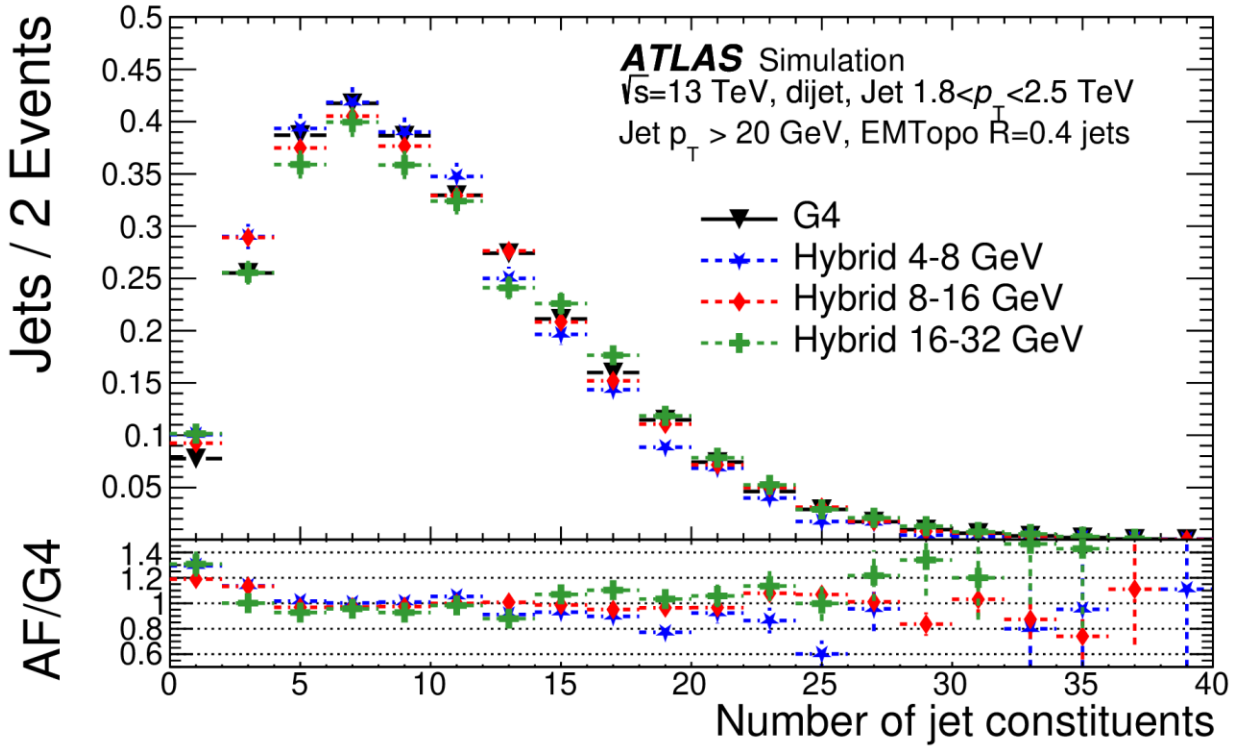
- We tested several options for all the parameters and validated them at different stages
  - Post-voxelisation (making sure there is only a minimal loss of information)
  - Stand-alone simulations
  - Full validation with single particles and physics samples
- Since the number of parameters is very large and the full chain is very time expensive, informed decision are taken at each step
- We also apply corrections for other hadrons by rescaling the energy of the pions





Tuning parameters: combining FCSV2 and GANs

- AF3 was then built combining FCSV2 and FastCaloGAN, the most sensitive variable in jets was the number of constituents. Other variables considered too in the final selection making sure that a gain in one variable was offset by another



Tuning ML: FastCaloGAN

- How training is performed?
  - Each of the 300 GANs is trained using 15 single particle samples in the energy range (256 MeV-4 TeV, produced in powers of 2) in a narrow (0.05) eta slice
    - Training was performed on HTCondor V100 cards for 1M iterations taking ~8h
  - A checkpoint is saved every 1k iterations
  - The best checkpoint is selected using the total energy (sum of all voxels)
    - We tested different variables and the total energy, which is quick to calculate, was one of the most difficult to reproduce for the GANs. It was also a good metric to describe the overall quality of the shower produced by the GANs
    - A GAN that can reproduce significantly better the total energy, typically produces showers that have better shapes too

Tuning ML: FastCaloGAN

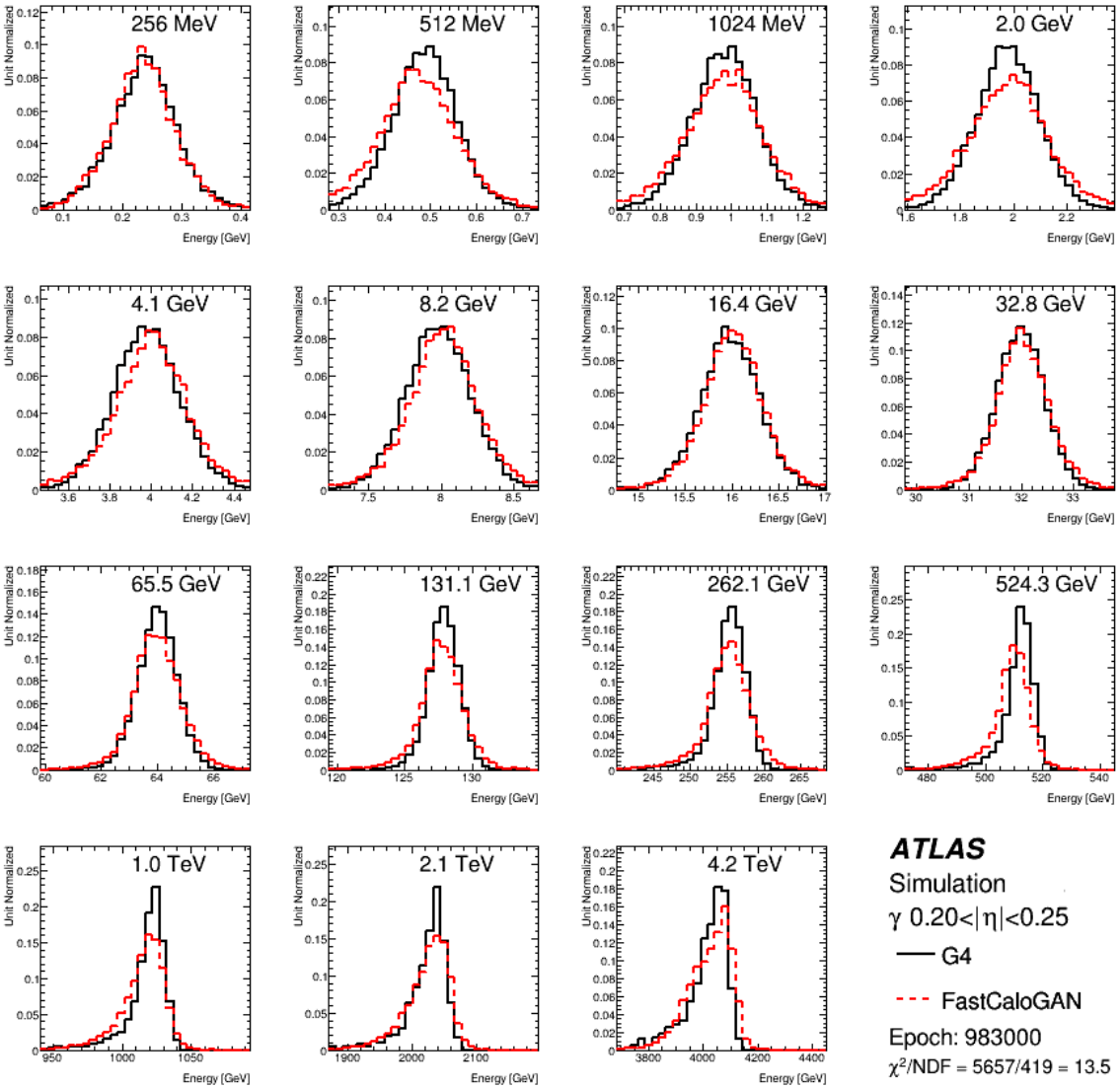
- What hyperparameter tuning approach is adopted?
  - For FastCaloGANv1 there was limited HPO. We tested several combinations of HP until we could successfully train a GAN with “good” performances
  - We later performed a grid search and discovered that most HP had little impact and/or were already well tuned
  - The BatchSize had one of the largest impact; increasing it improves the results. This was expected since we train for a fixed number of iterations and increasing the batchsize essentially means using more events in the training process

Tuning ML: FastCaloGAN

- What metric is used to compare between the performance of the model during the tuning process?
  - We calculate the χ^2 between the input sample and the generated showers (10k events in both, “WW” in TH1.Chi2TestX())
  - The metric is the sum of the 15 χ^2 values; the checkpoint with the lowest total χ^2 is chosen for a model
  - The model with the lowest total χ^2 is considered the best model and its HP preferred over other models

Selection criteria for GANs

SIMU-2018-04



Tuning ML: VAE

- How training is performed?
  - The VAE model is trained as a single model per particle type (all energies 1Gev-1TeV, all eta regions) on GPUs :
    - NVIDIA Titan X Pascal graphics card with a processing power of 3584 cores, each clocked at 1417.
  - The model is trained for 5k epochs and is saved every 100 to test overfitting and convergence

- What hyperparameter tuning approach is adopted?
  - A grid search was performed in parallel on multiple GPUs

Tuning ML: VAE

- What metric is used to compare the performance of the model during the tuning process?
  - There are 2 steps validation:
  - Reconstruction level (using encoder and decoder): Structure Similarity index metric (SSIM) + Jensen–Shannon Divergence (JSD) are used to compare between a validation set of showers and their reconstructed version. These 2 measures are used to directly discard HP sets or to pass to the next validation step. (Note: SSIM is used to compare images and JSD to compare energy distributions of cells/voxels/centroids)
  - Generation level (using only the decoder): here the metric used is the χ^2 of the total energy and energy per layer

GAN architectures in AF3

SIMU-2018-04