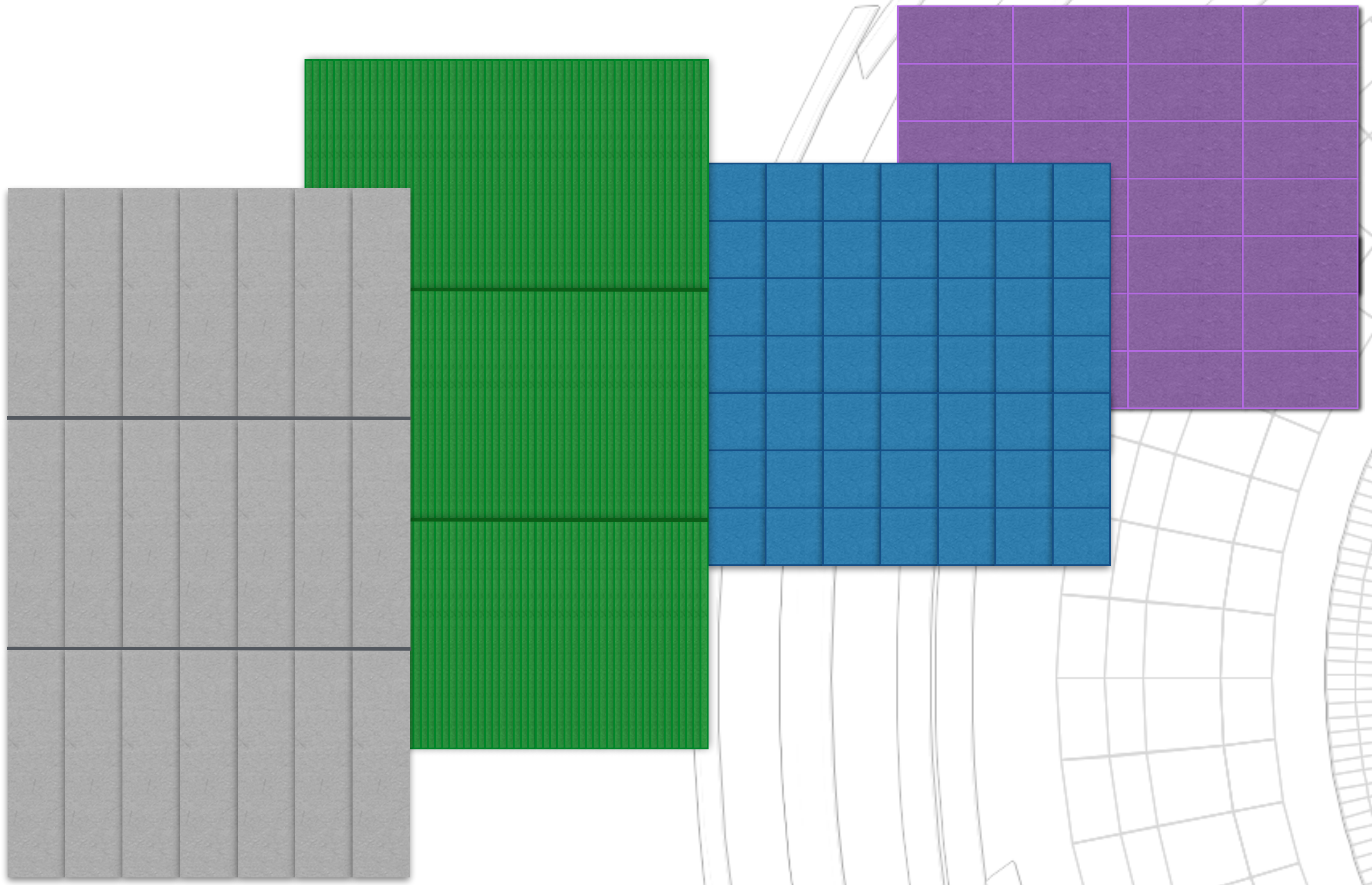
Deep Generative Models for Fast Simulation in ATLAS

Graeme Stewart (CERN), Aishik Ghosh, David Rousseau (LAL, Orsay), Kyle Cranmer (NYU), Michele Faucci Giannelli, Serena Palazzo (University of Edinburgh), Stefan Gadatsch, Tobias Golling, Johnny Raine, Dalila Salamani, Slava Voloshynovskiy (UniGe), Gilles Louppe (ULiège)

on behalf of ATLAS

IML Workshop
17 April 2019

LABORATOIRE DE L'ACCÉLÉRATEUR LINÉAIRE

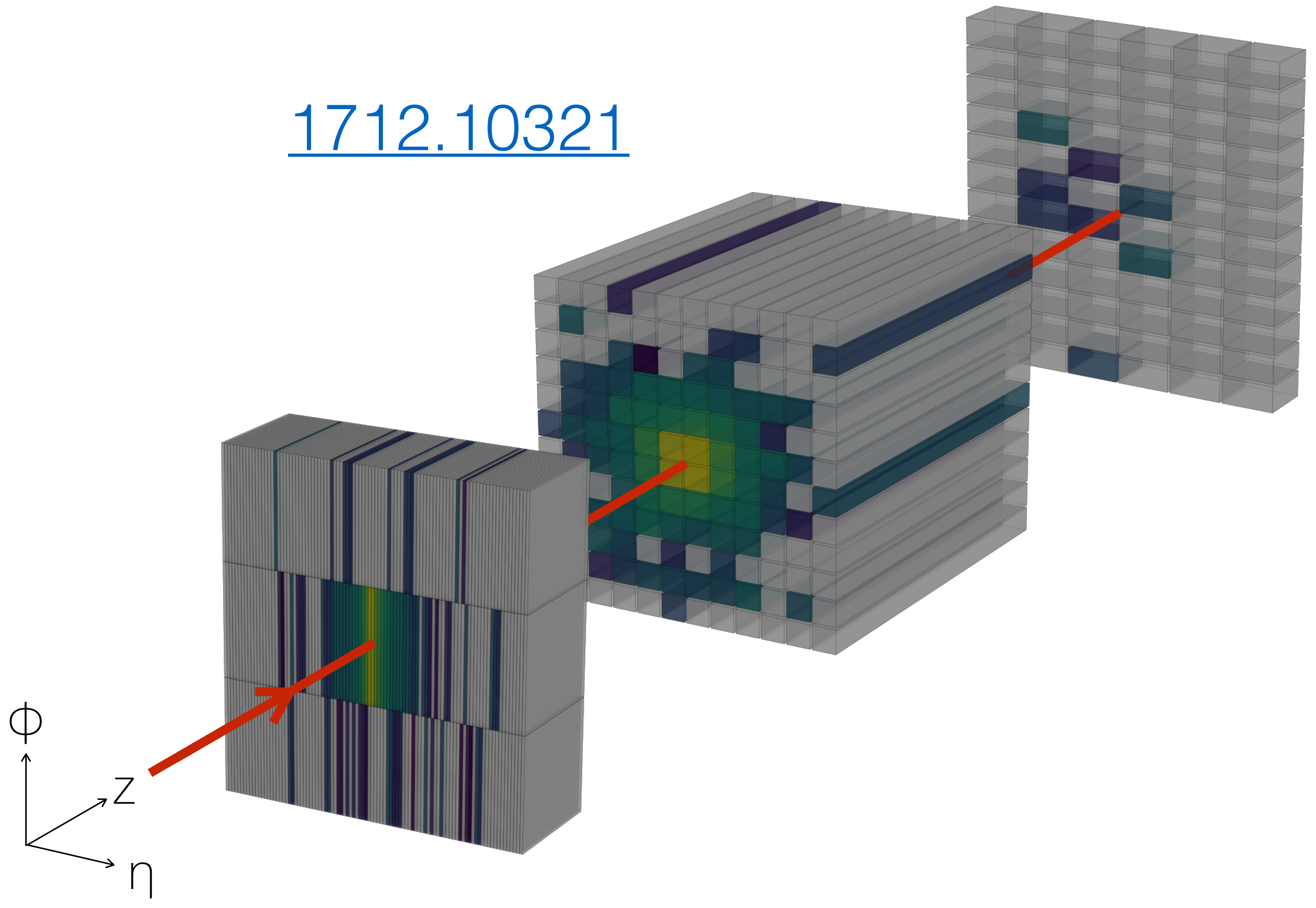


ATLAS EXPERIMENT

Generative Models for EM Shower Simulation

CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

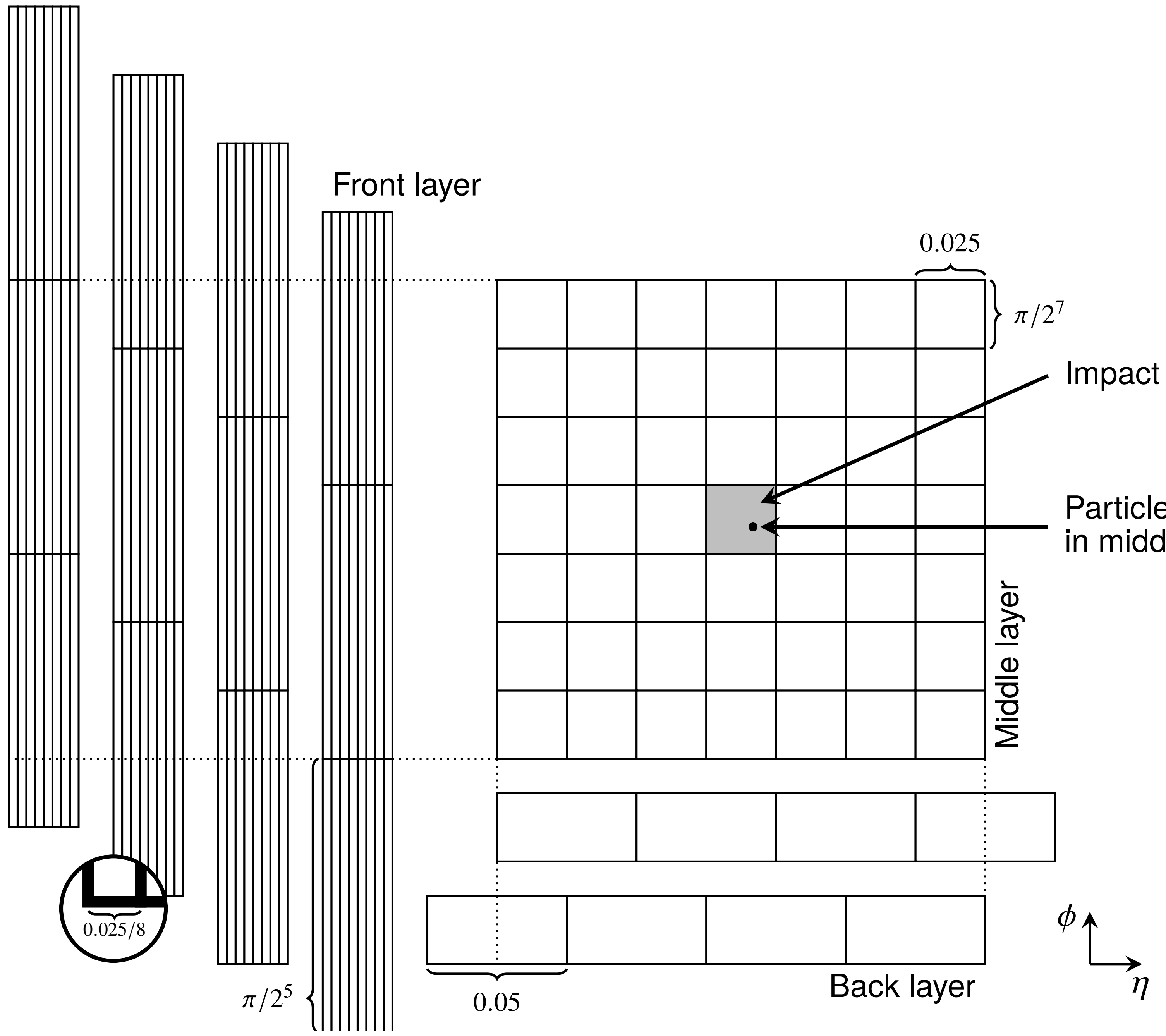
[1712.10321](https://arxiv.org/abs/1712.10321)

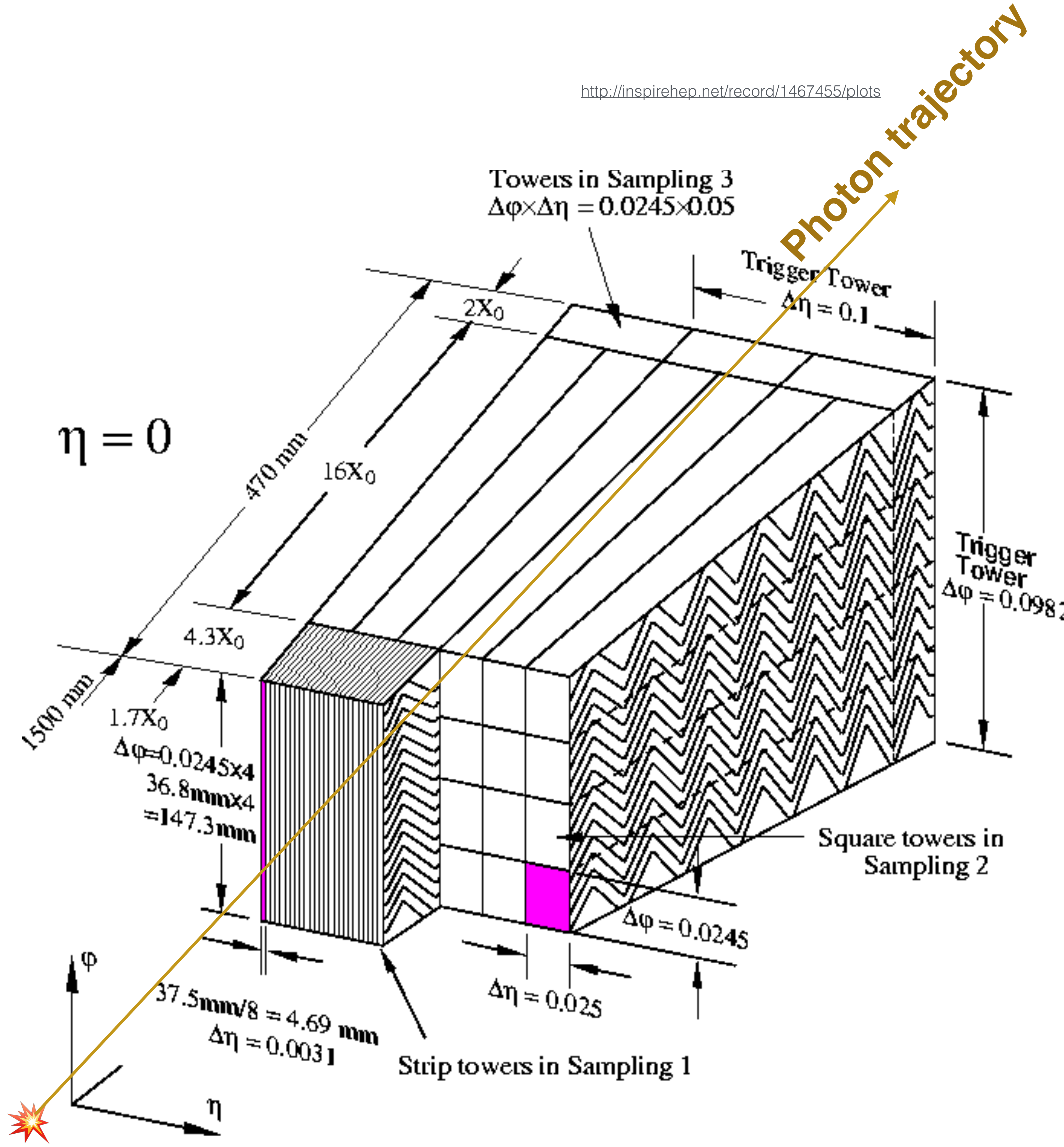


- CaloGAN showed that it is possible to simulate EM showers for a detector like ATLAS using GANs
- Since then you've seen many GANs for particle physics
- What's so special about these Generative Models (GAN, VAE) presented in this talk?

http://inspirehep.net/record/1467455/plots

First efforts to simulate the real, present day, irregular, coarse granularity ATLAS calorimeter with Generative Models



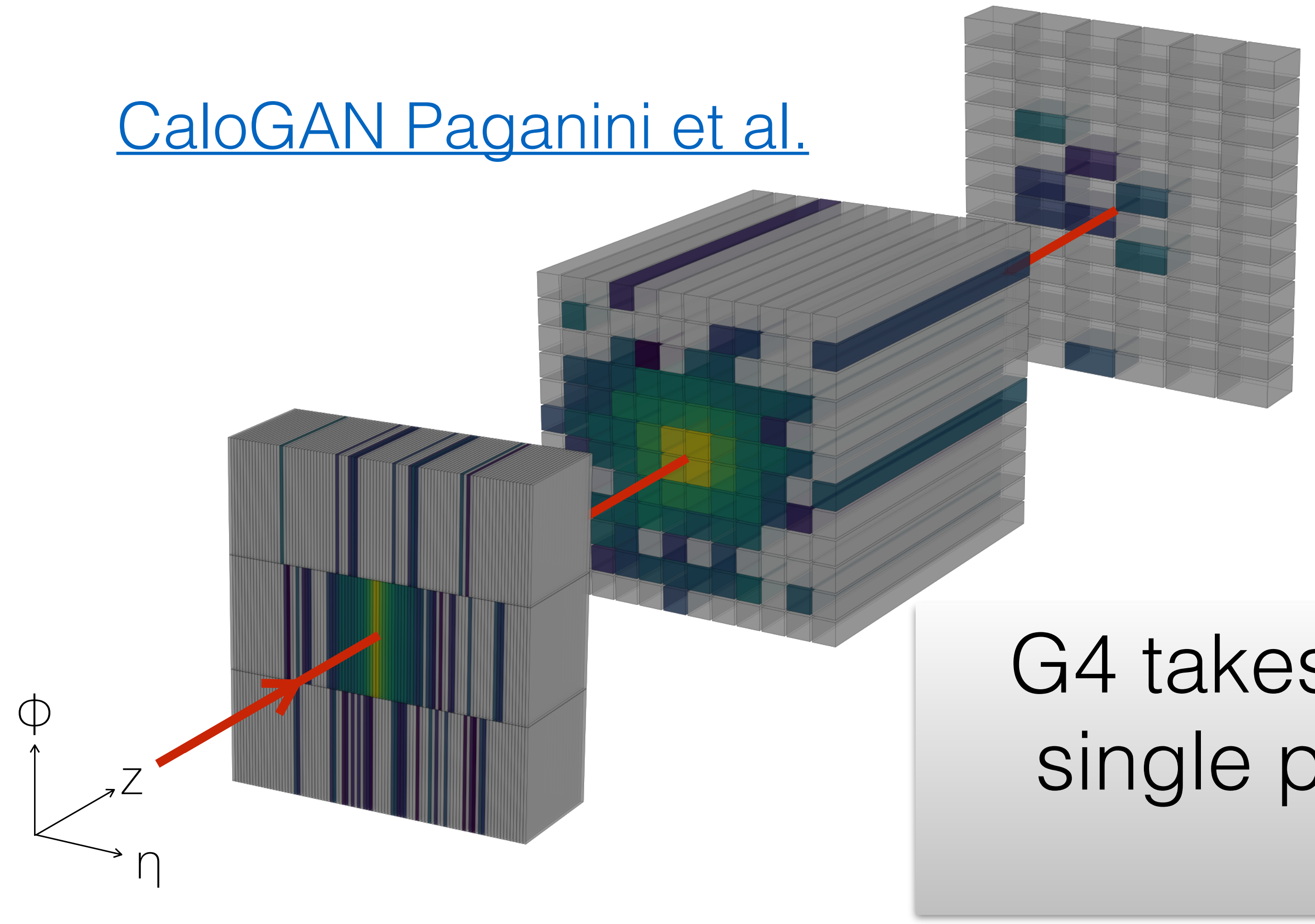


Train and validate using G4 simulations for the ATLAS geometry

Generative Models for EM Shower Simulation **in ATLAS**

CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

[CaloGAN Paganini et al.]



G4 takes ~10 seconds per shower for a 65 GeV single photon shower (more for higher energy)
Too slow!

The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim

The ATLAS Collaboration

7 Conclusion

The fast calorimeter simulation FastCaloSim has been developed in order to reduce the simulation time in the ATLAS calorimeter system from several minutes to a few seconds per event, using a parametrization model for the longitudinal and lateral shower development of photons, electrons and charged pions. The

ATL-PHYS-PUB-2010-013
18 October 2010

ATL-SOFT-PUB-2018-002

10th July 2018

The new Fast Calorimeter Simulation in ATLAS

The ATLAS Collaboration

during the simulation jobs. A prototype is being tested and validated, and it has shown significant improvements in the modelling of cluster-level variables in electromagnetic and hadronic showers.

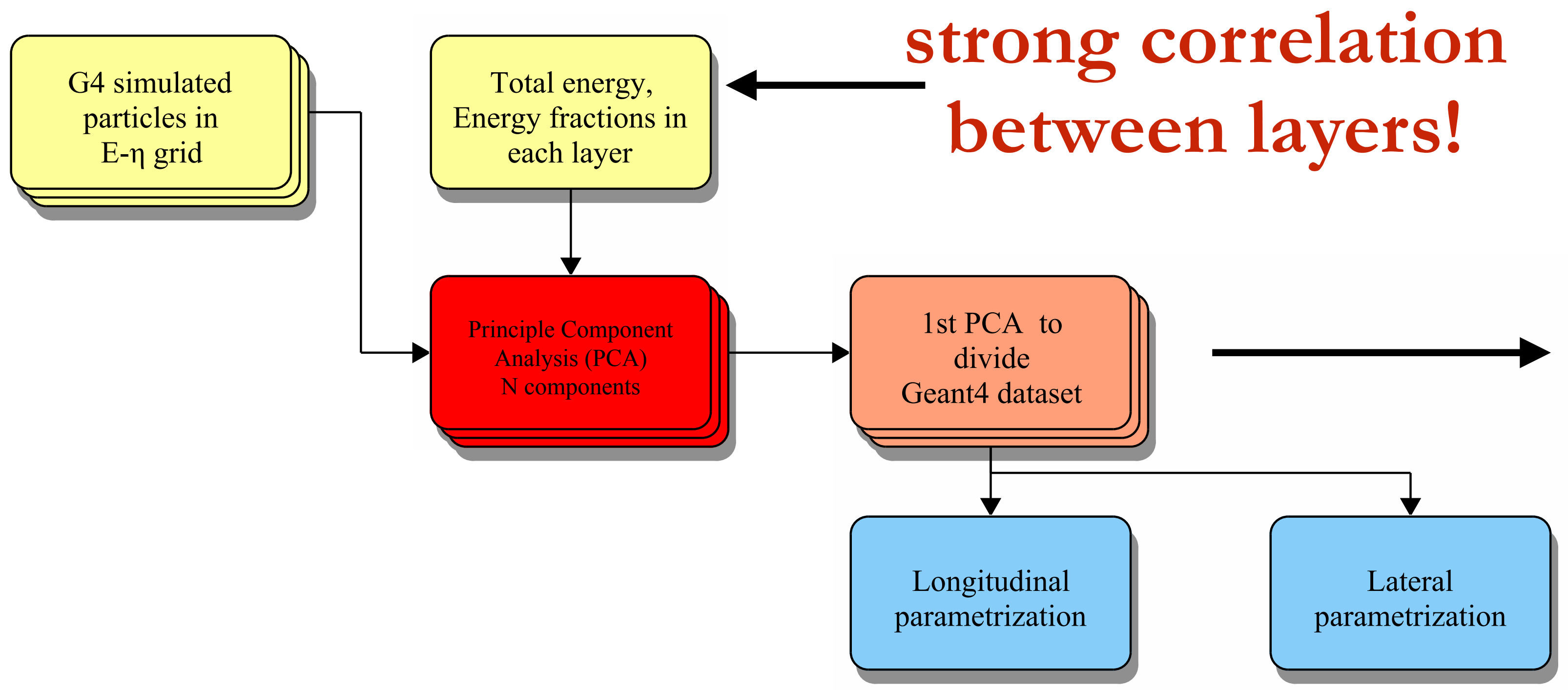
- ATLAS already using fast simulation techniques for years!
- Trade-off between slow accurate G4 and fast less accurate FastCaloSim V1
- New FastCaloSim V2 using some ML techniques already in advanced state of development

First efforts to simulate the real, present day, irregular, coarse granularity ATLAS Calorimeter with Generative Models

(Only small eta region in barrel for photons right now)

FastCaloSim V2 already using Machine Learning at various points in the chain

Still have to parameterise separately for η slices, energies, interpolation …

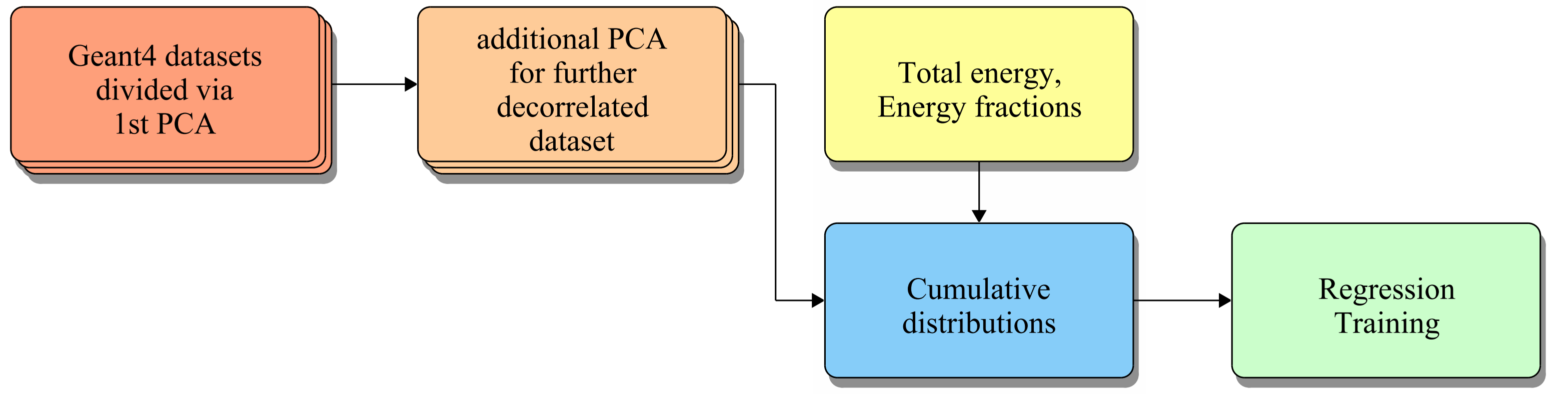


Human designed parameterisation techniques being developed for many years -> A **high benchmark** against which to compare GAN / VAE performance

Validation cross-check frameworks already in place for FastCaloSim including variables defined by the EGamma group: same level of scrutiny for all fast simulation approaches.

Need to get all distributions right simultaneously, **average distributions might look right** but must **verify also the distributions per energy point / section of the calorimeter**

Additional PCA transformation to further decorrelation



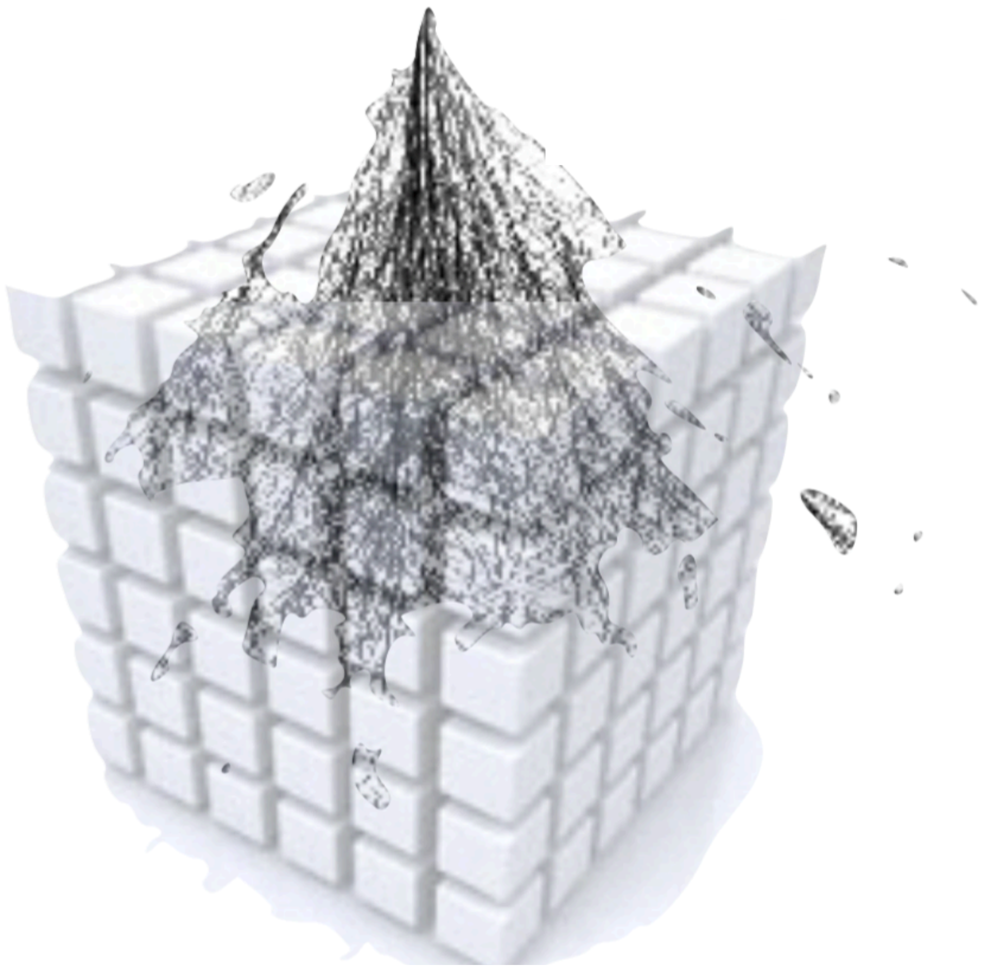
Introduction (Finally)

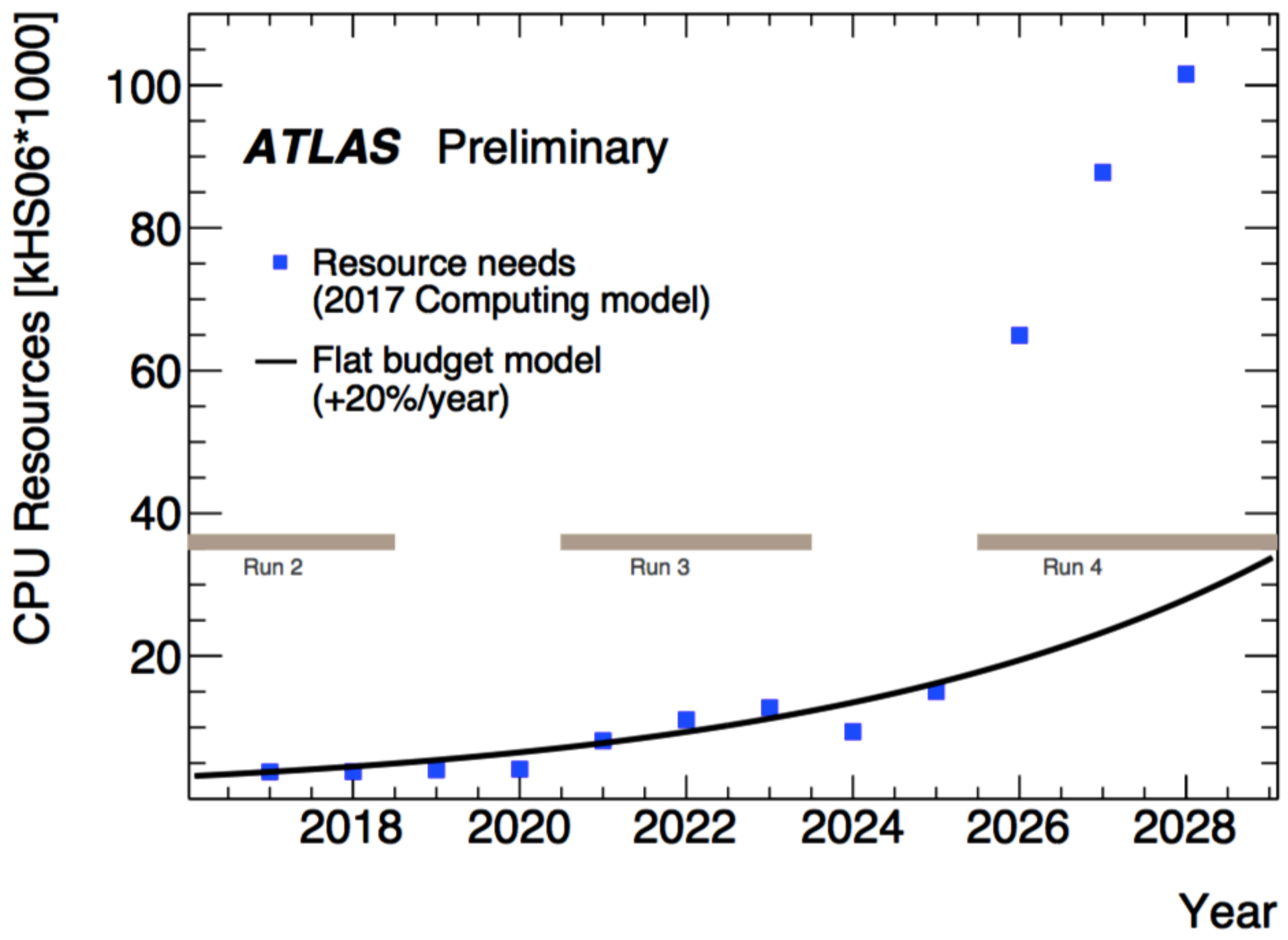
<u>Is it possible to one day…? :</u>

Simulate showers 100-1000x *faster* than Geant4

Less human time intensive, *higher accuracy* than current fast simulation methods

Have it run inside Athena (ATLAS C++ software) and be *less resource* hungry than current fast simulation methods



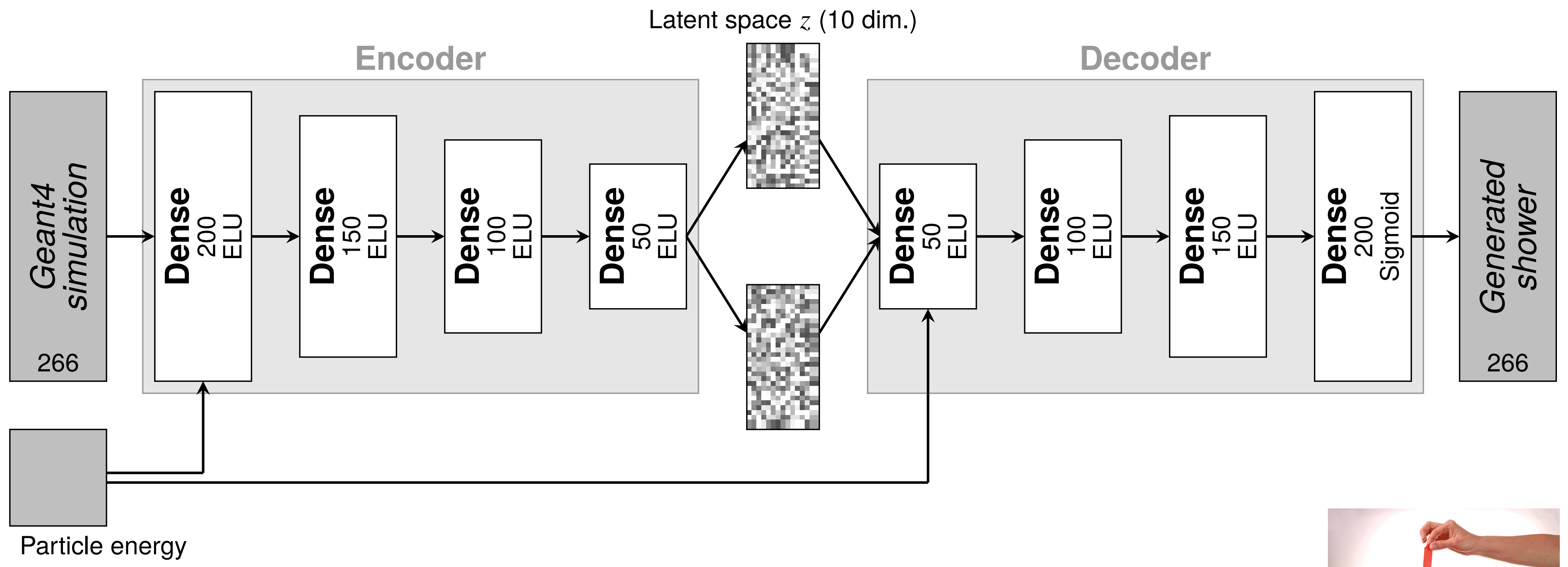


Geant4 requires significant resources with ~75% spent in shower simulation i.e. Calorimeter simulation

Imperative to develop fast shower simulations compared to Geant4

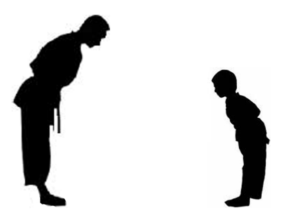
PubNote: VAE and GAN

VAE:

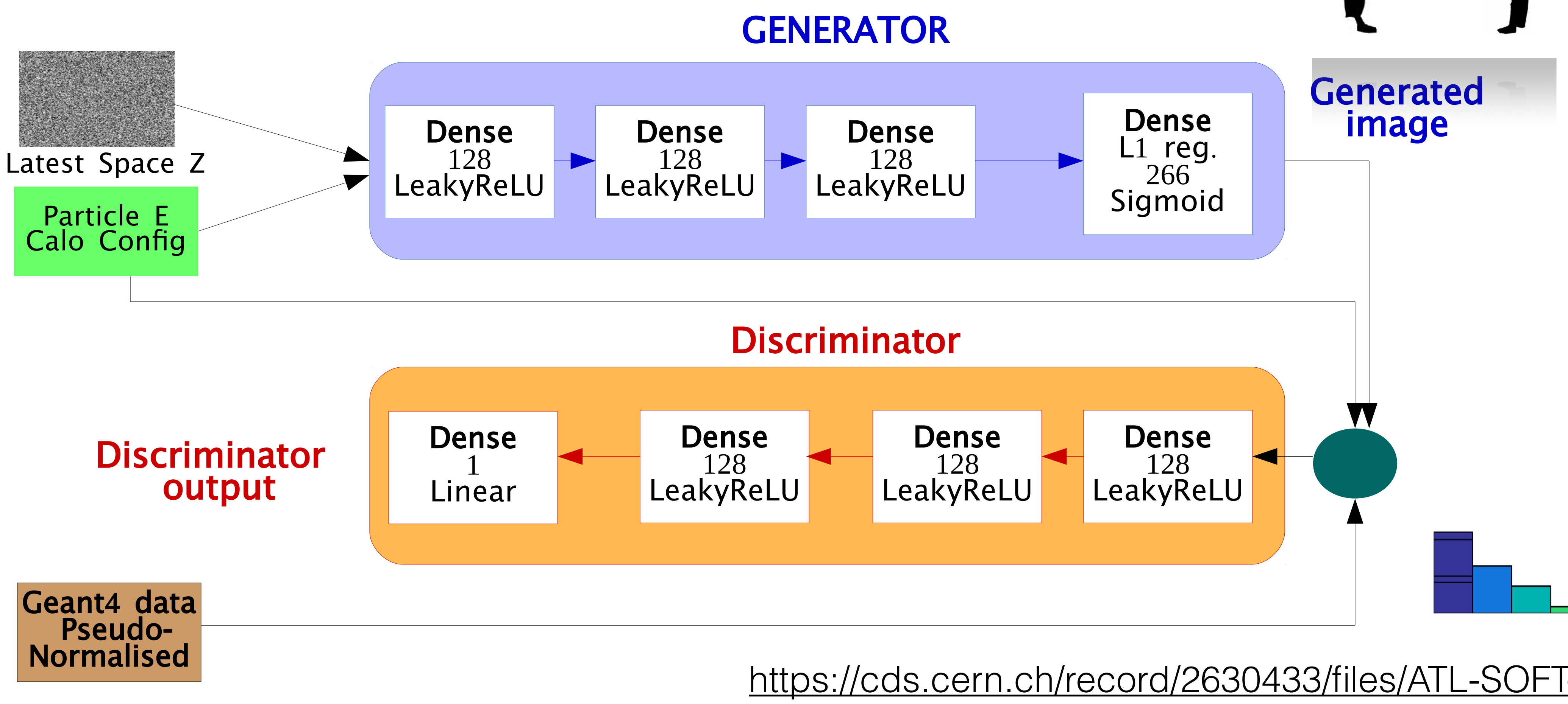


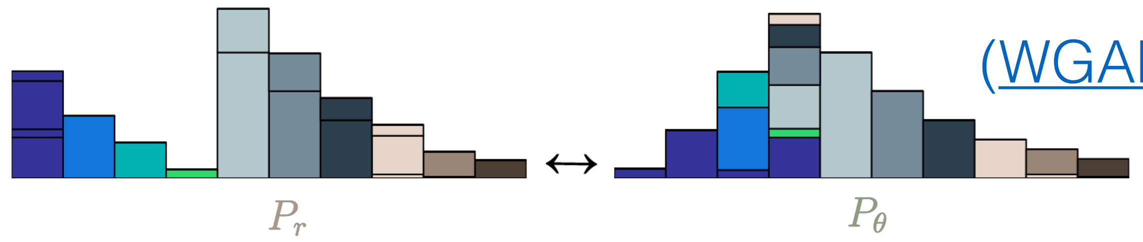
100 epochs, 2 mins, CPU

50k ‘epochs’, 7 hours training, 1 GPU



GAN:





Training dataset:

- Single **photon** samples from Geant4
- 88000 events
- 9 energy points : {1, 2, 4, 8, 16, 32, 65, 131, 262} GeV
- **0.20 < |η| < 0.25**
- 4 electromagnetic calorimeter layers

Data preprocessing

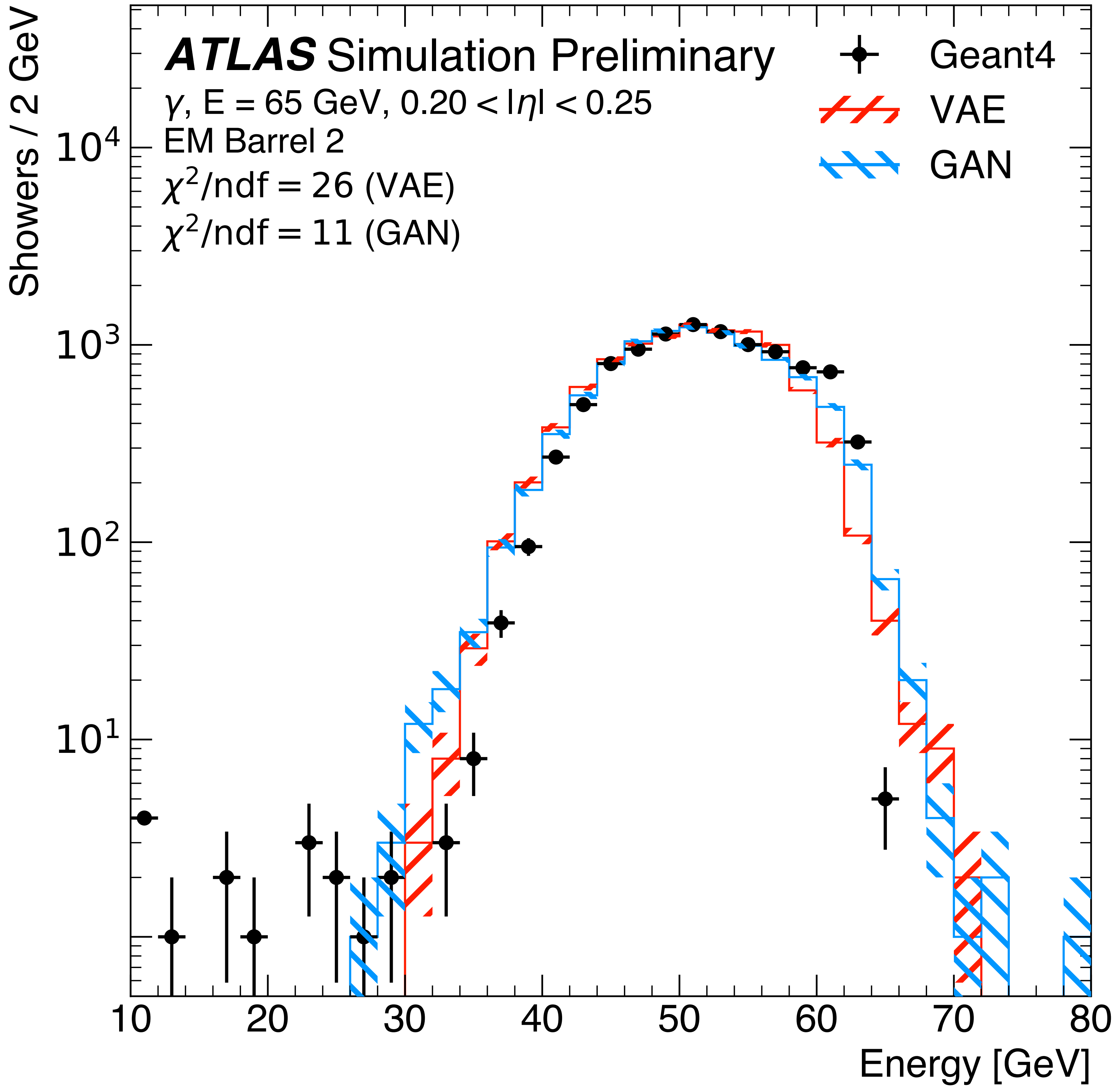
- Negative energies set to 0
- Mirror η < 0

([WGAN-GP](), Improved WGAN-GP nightmare on Keras!)

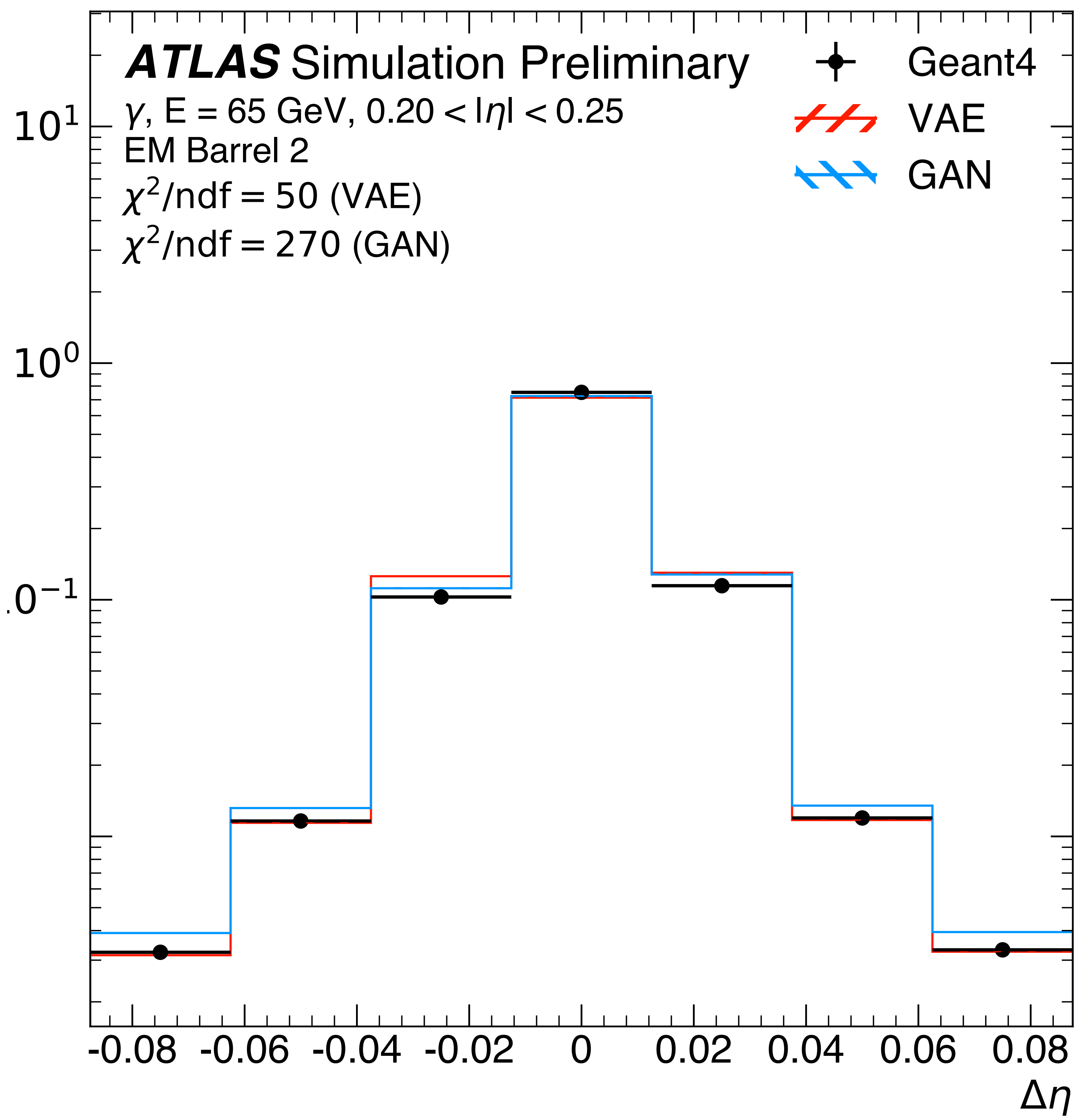
https://cds.cern.ch/record/2630433/files/ATL-SOFT-PUB-2018-001.pdf

2018 Results(1/3)

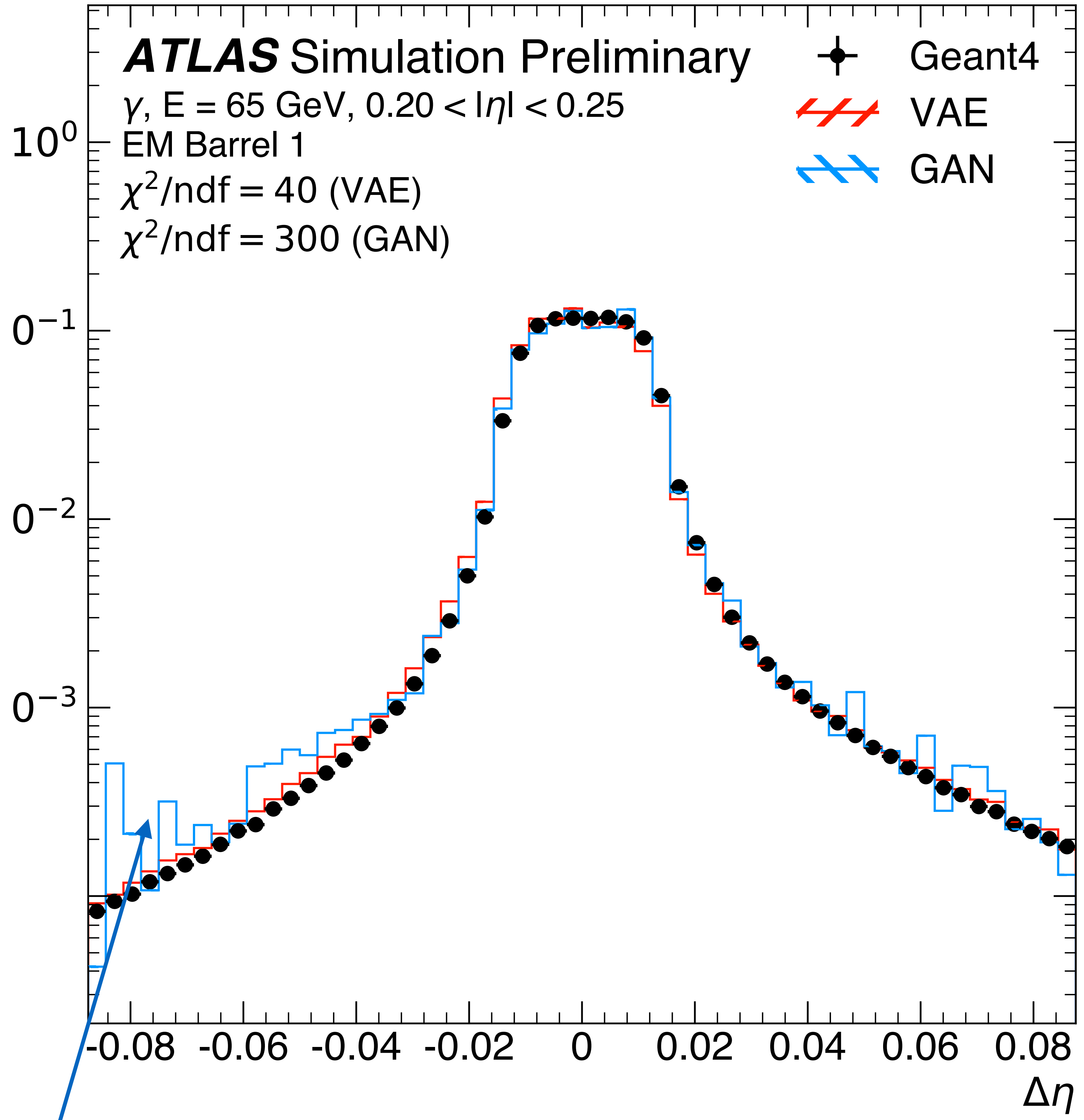
From summer PubNote 2018



Energy in Middle Layer



Average η in Middle

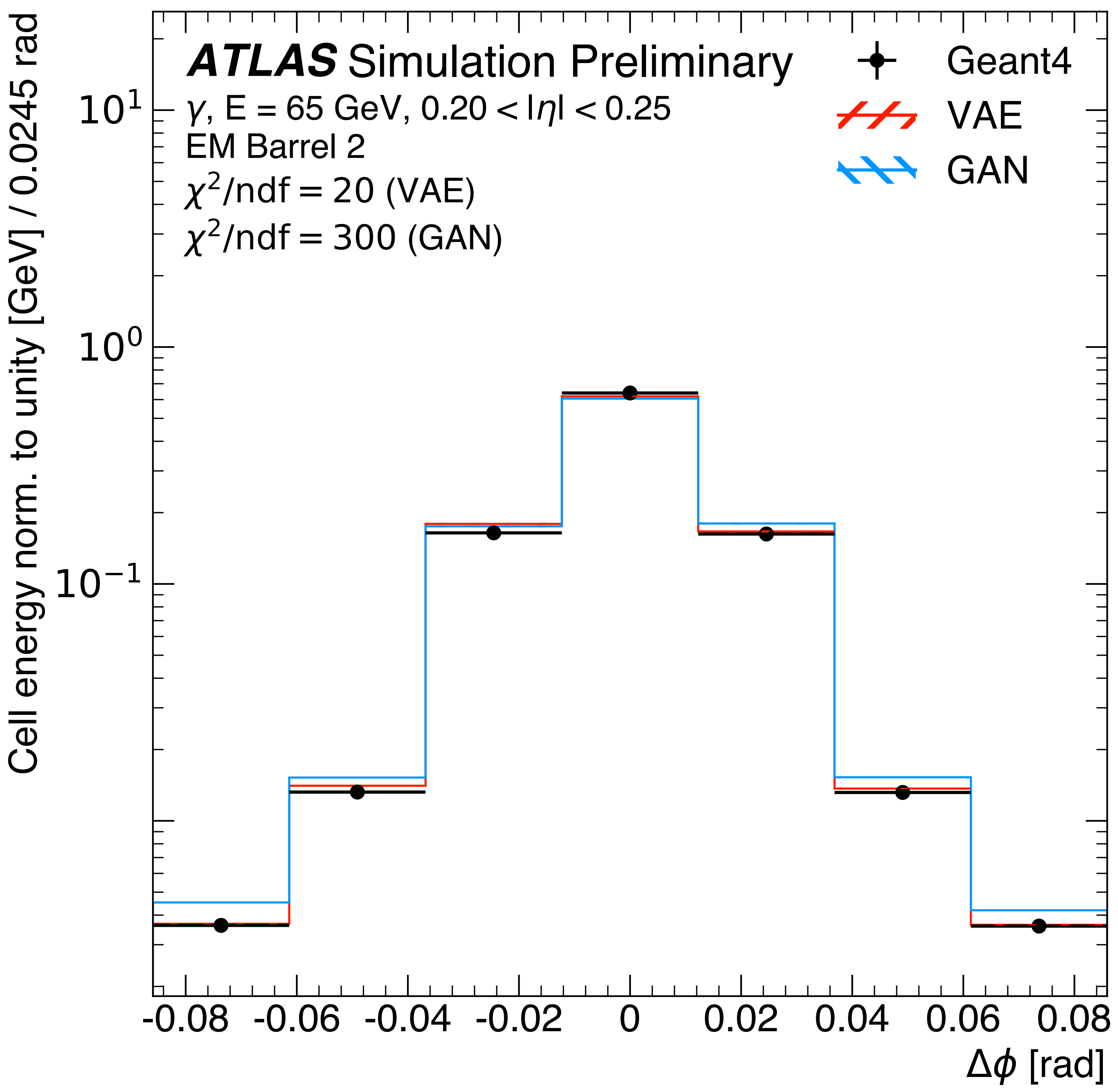


Average η in Strip

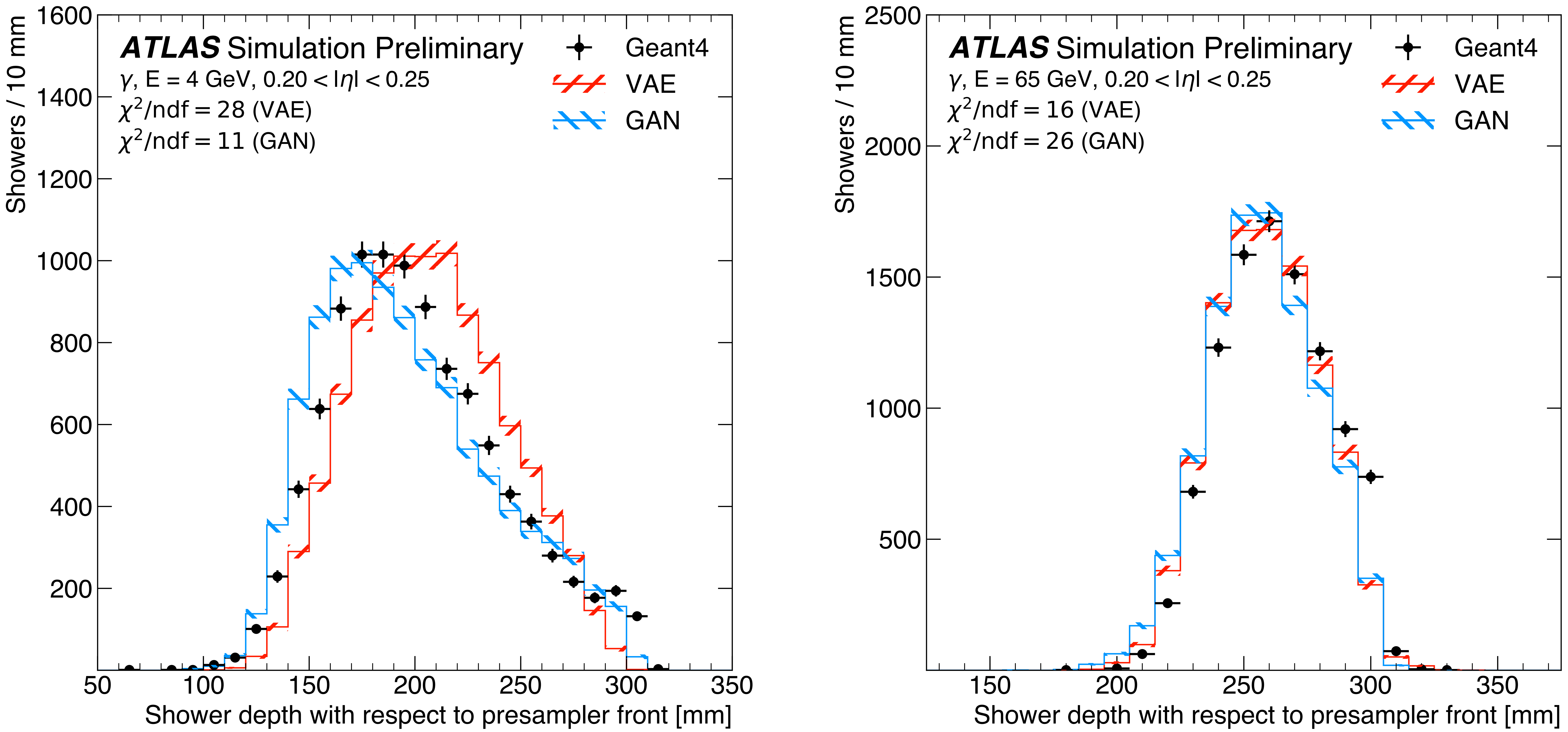
GAN: Fluctuations due to small training size, fixed since PubNote. 4% -> 50% of dataset by removing momentum from Adam optimiser, lowering number of epochs

2018 Results(2/3)

From [summer PubNote 2018]



Average ϕ in Middle

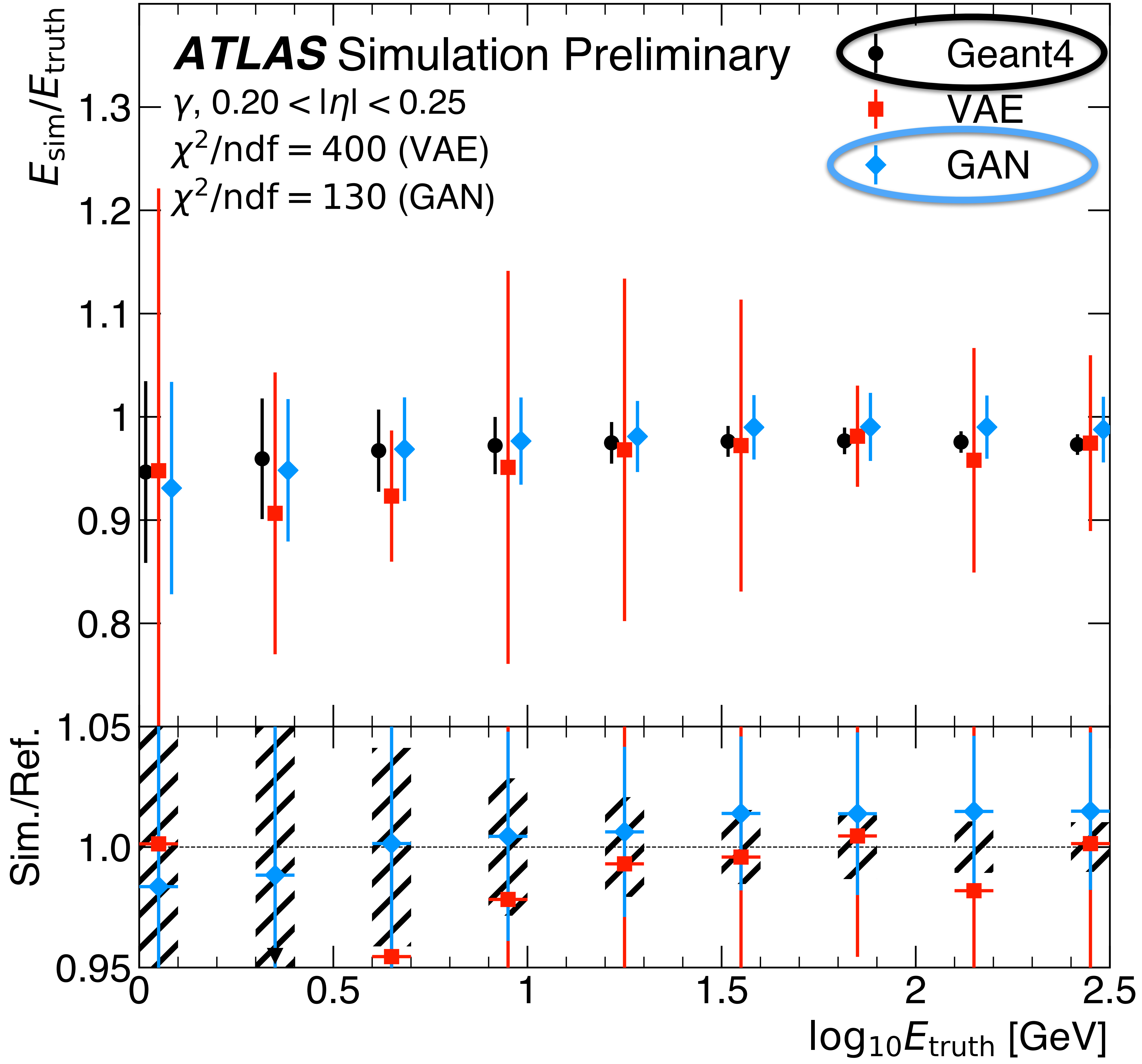


(a) 4 GeV

(b) 65 GeV

Shower Depth

2018 Results(3/3)



η, ϕ, other distributions not so bad but for total energy…

GAN gets the means but not the **widths** of the energies

Critic can't see the difference in real and fake images.

Tried training on single high energy point,
Minibatch discrimination, various other tricks. No result.

Trade-Off b/w Distributions and Total Energy

GAN Critic
Gradient Penalty :
10

Plots get worse

Energy gets better

GAN Critic
Gradient Penalty :
1e-13

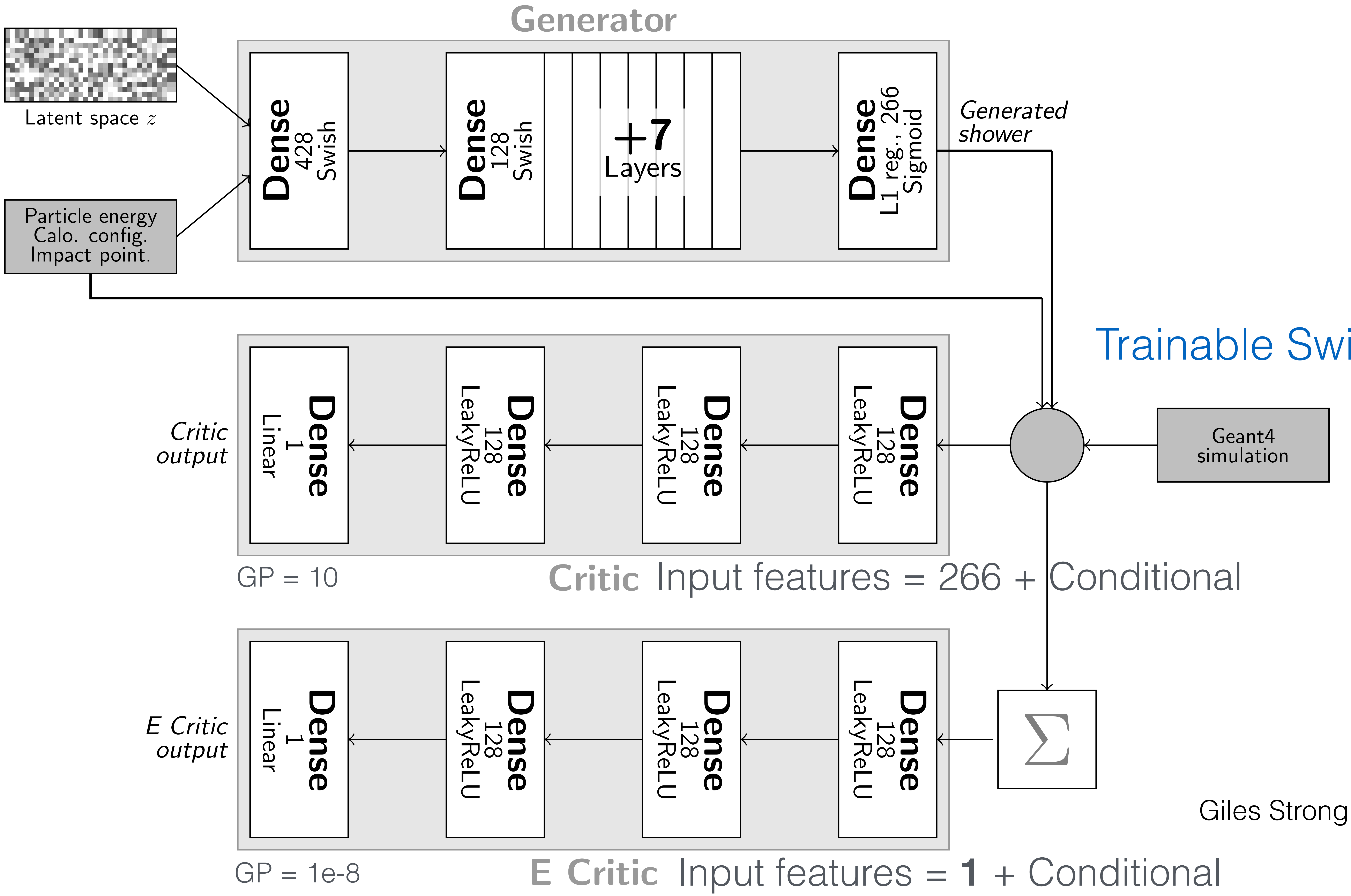
And highly unstable training

$$L_{\mathrm{GAN}} = \underset{\tilde{x}\sim p_{\mathrm{gen}}}{E}[D(\tilde{x})] - \underset{x\sim p_{\mathrm{Geant4}}}{E}[D(x)] + \lambda \underset{\hat{x}\sim p_{\hat{x}}}{E}[(||\Delta_{\hat{x}} D(\hat{x})||_2 - 1)^2]$$

Trade-Off b/w Distributions and Total Energy: How to get the best of both??

"Train the Generator against a Critic of each type!"
-Gilles Louppe (ATLAS ACE), at ATLAS ML Workshop 2018

New GAN Architecture



2 Critics

Deeper Generator needed

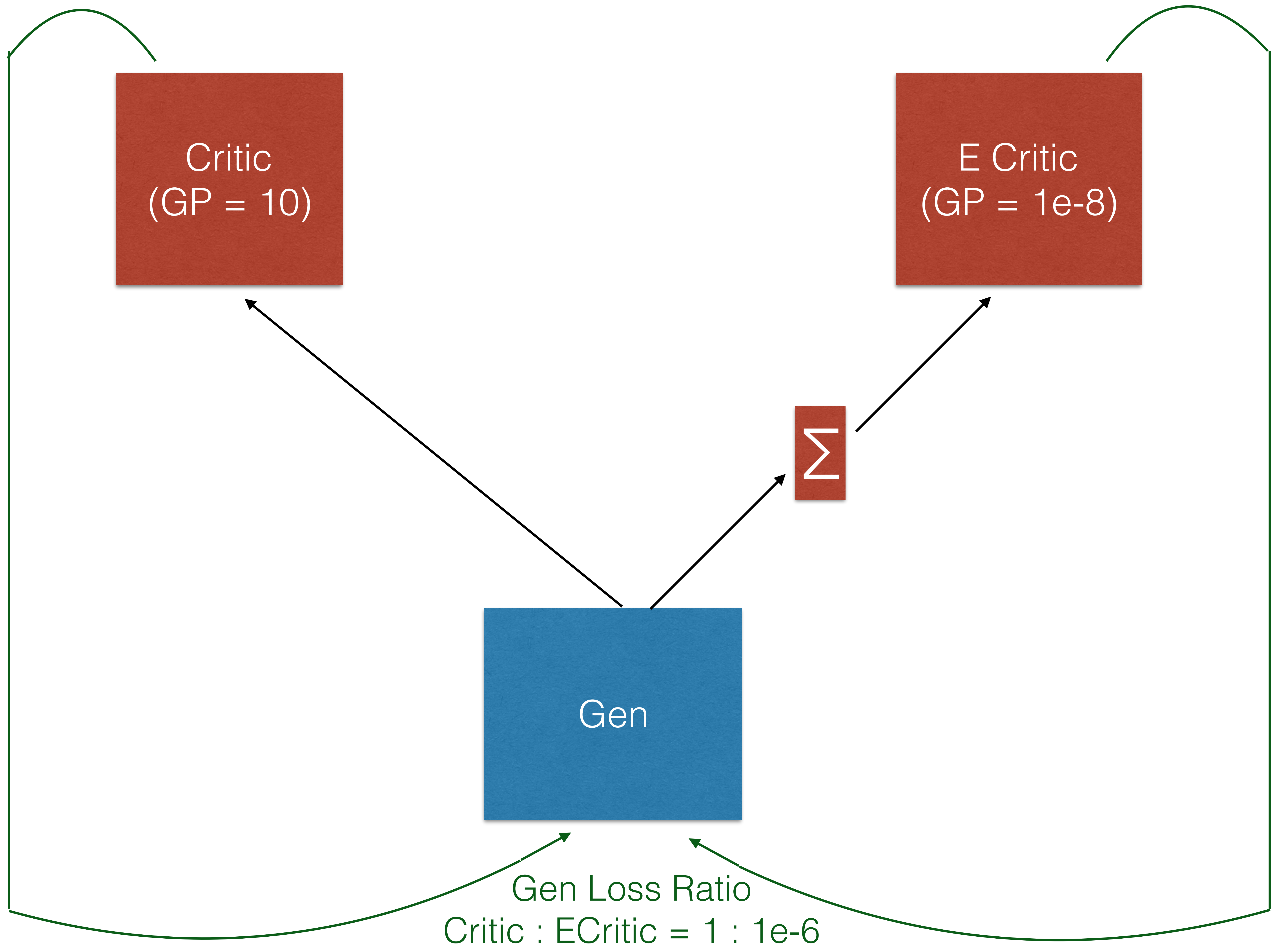
Trainable Swish activation for Generator

$$Swish(x) = x \cdot \mathrm{sigmoid}(\beta x)$$

Swish activation inspired from
Giles Strong's presentation also at AML Workshop 2018

Training time: 2.5 Days on 1 GPU for 1.5k Epochs
Training Size: 44000 events (50% of Dataset)
CPU = 2 x GPU training time at 52% GPU utilisation

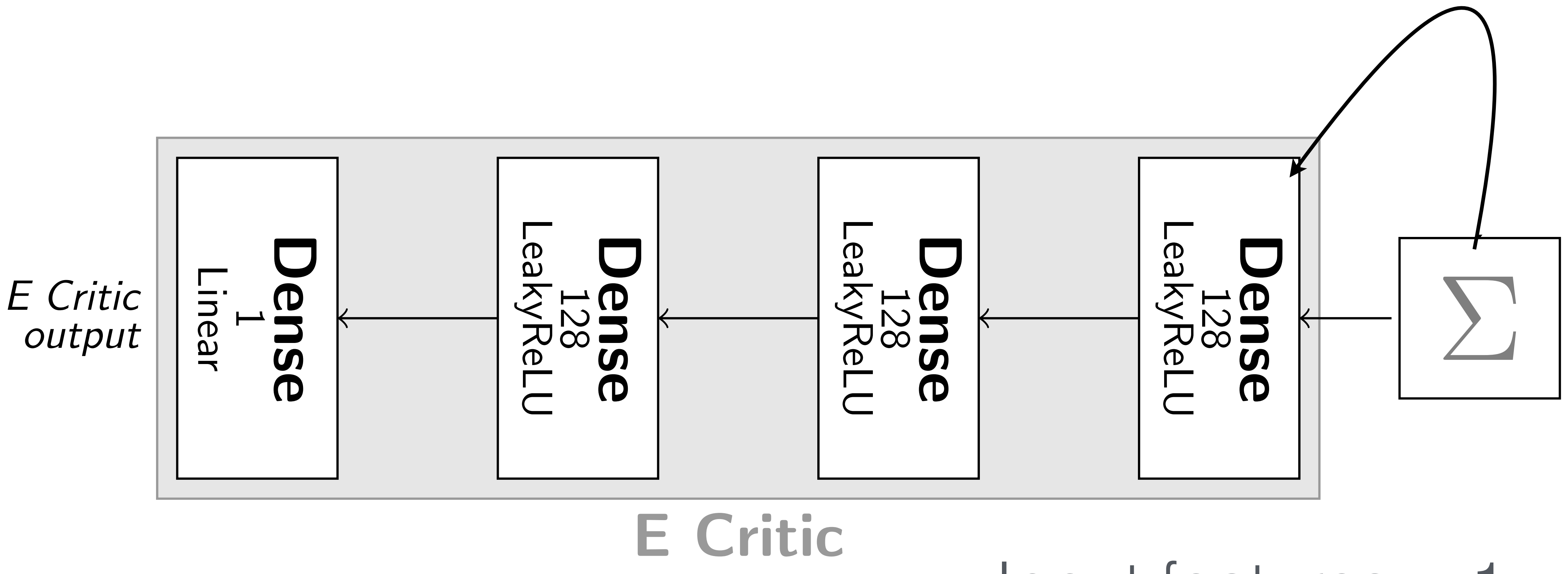
New GAN Architecture



GP: Two Sided Gradient Penalty

Careful: Sum Inside or Outside the Network?

$\boxed{\sum}$ = `Lambda(sumFunc)(m_input_image)`



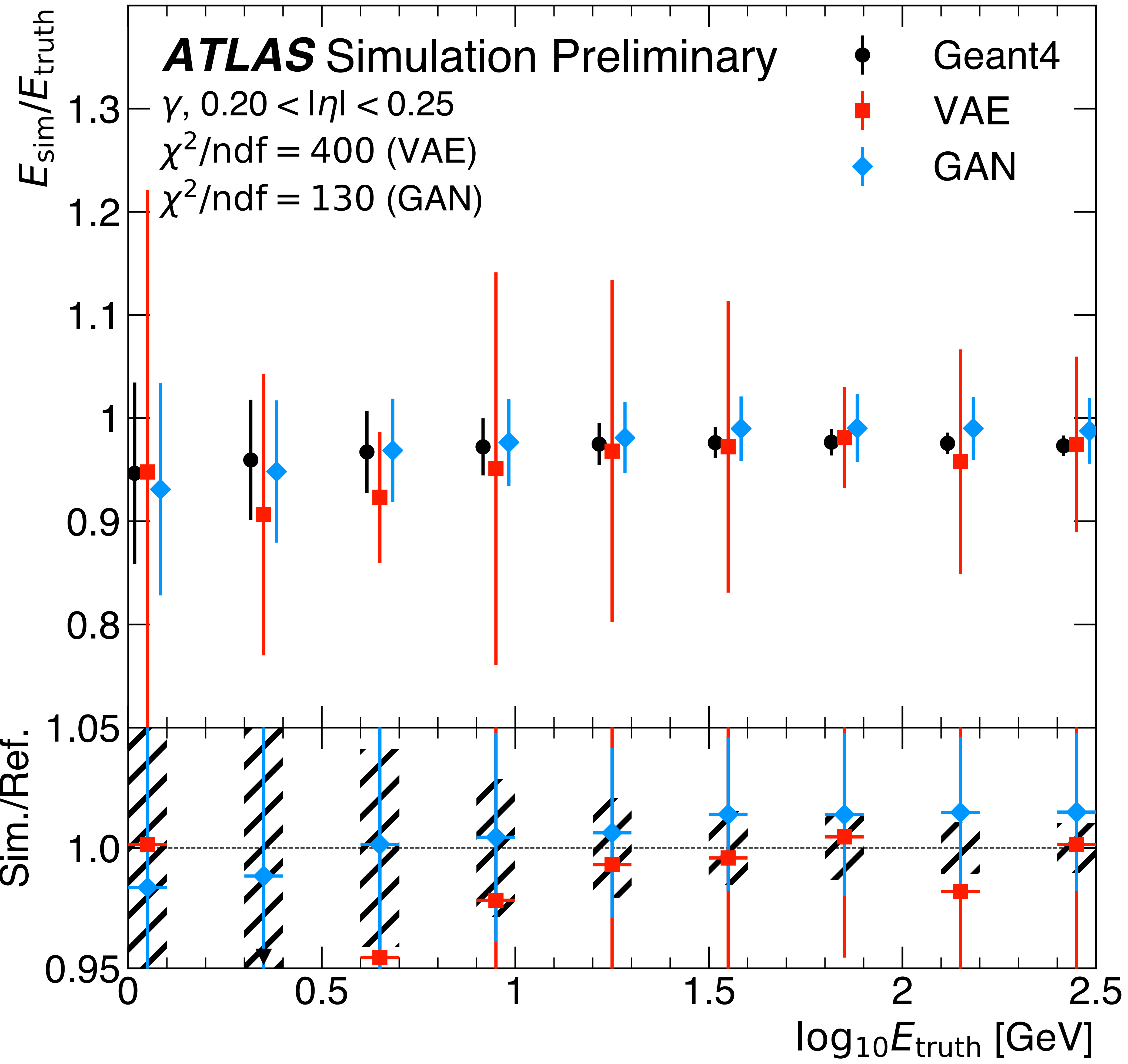
Input features = 1 + Conditional

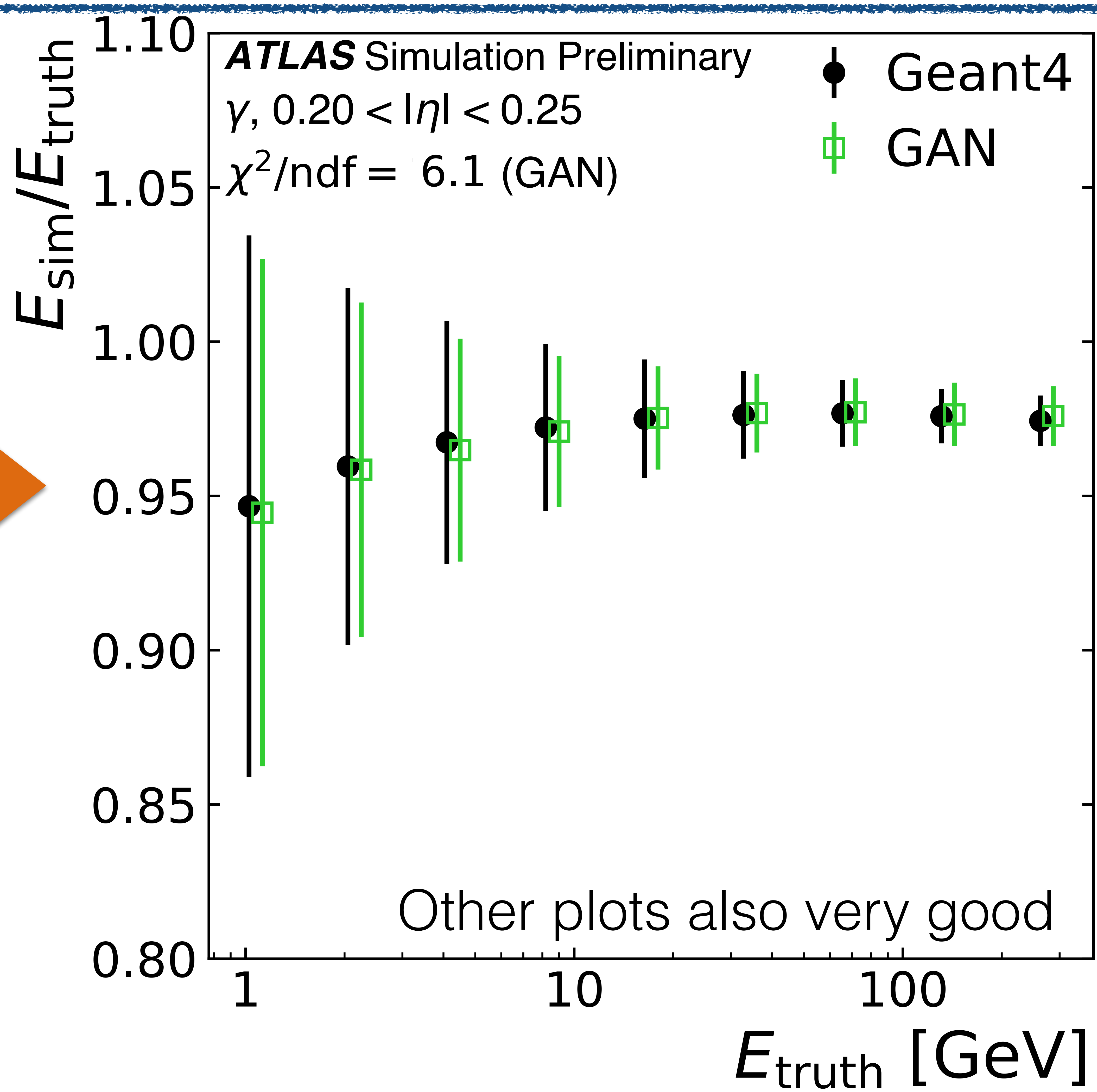
They are not equivalent, need to tune hyper parameters differently

$$L_{\mathrm{GAN}} = \underset{\tilde{x}\sim p_{\mathrm{gen}}}{E}[D(\tilde{x})] - \underset{x\sim p_{\mathrm{Geant4}}}{E}[D(x)] + \lambda \underset{\hat{x}\sim p_{\hat{x}}}{E}[(||\Delta_{\hat{x}} D(\hat{x})||_2 - 1)^2]$$

Gradient Penalty on 1 input vs 266 inputs

GAN: Improved Energy Resolution

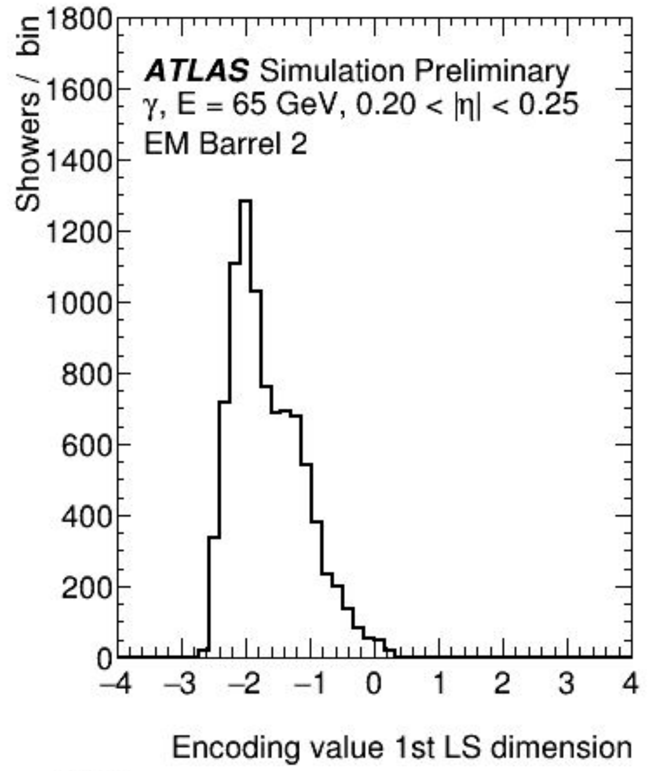


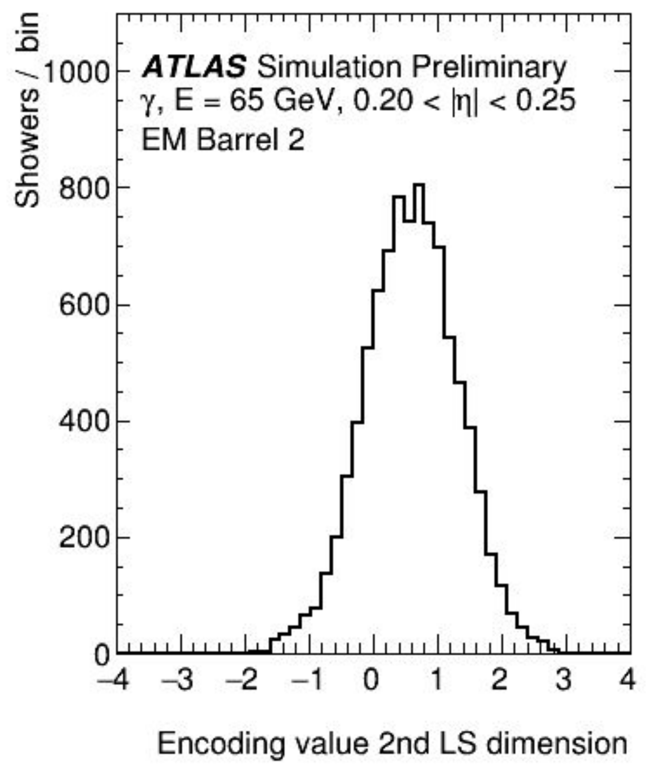


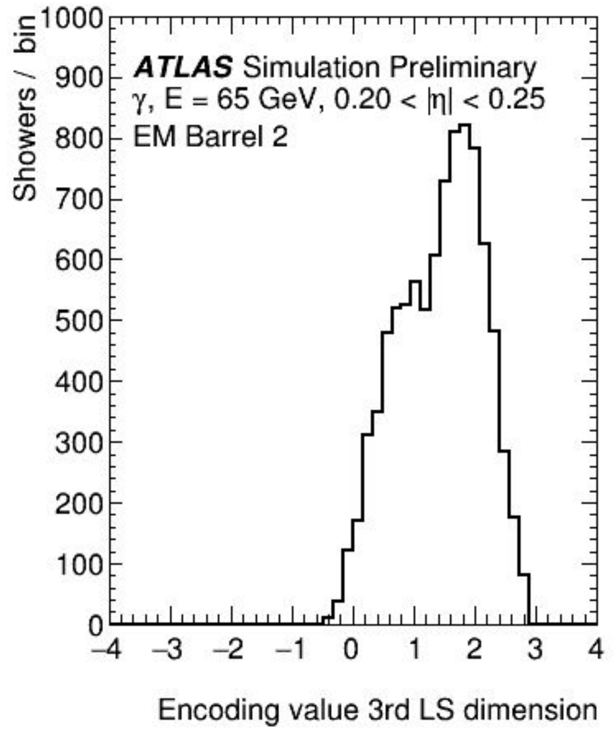
Reference

GAN still about 15% too large at worst case but not fine tuned

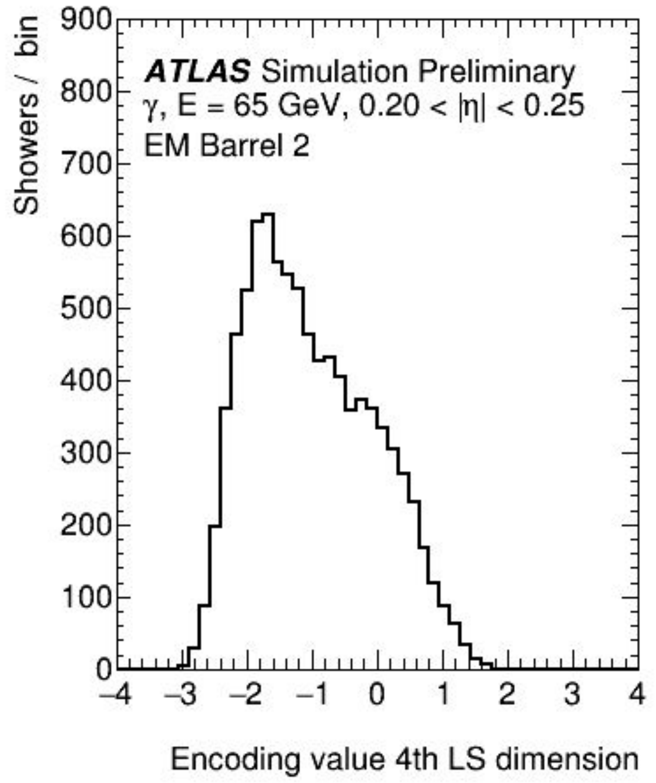
VAE Latent Space

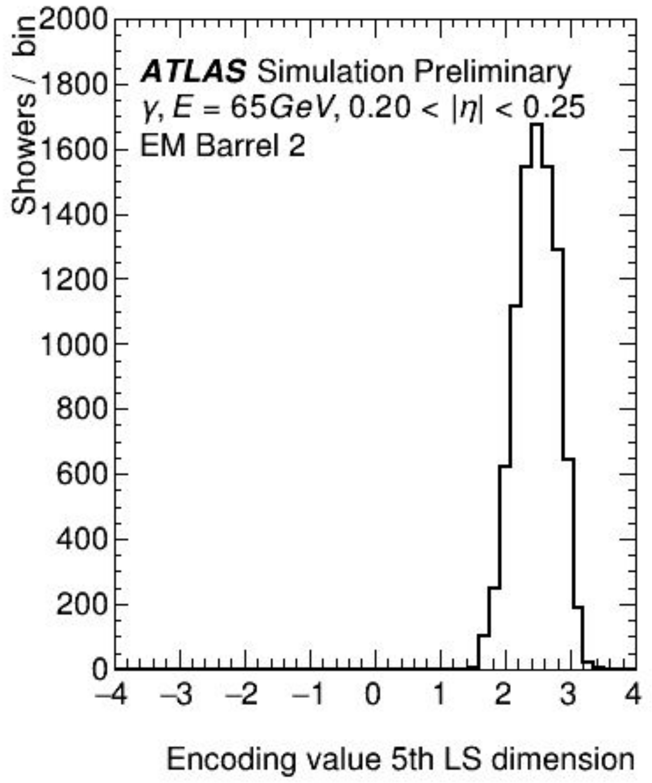






5D Latent Space don't look Gaussian



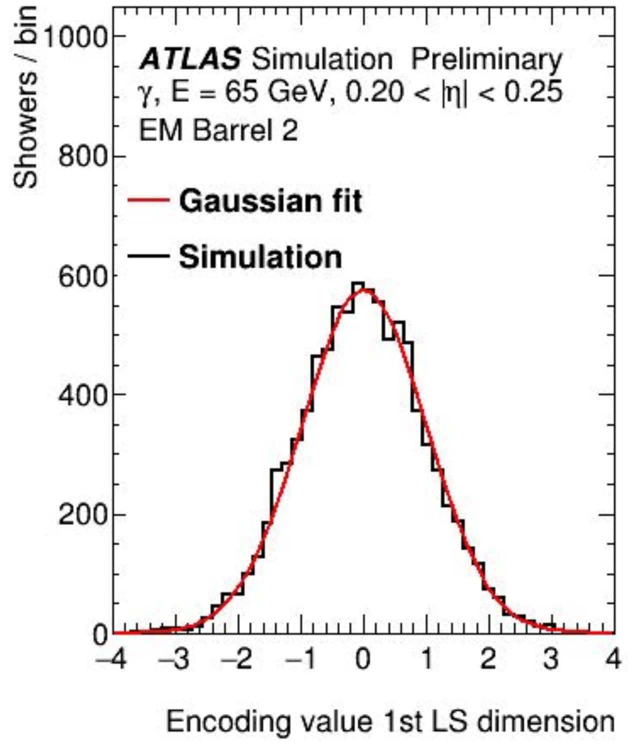


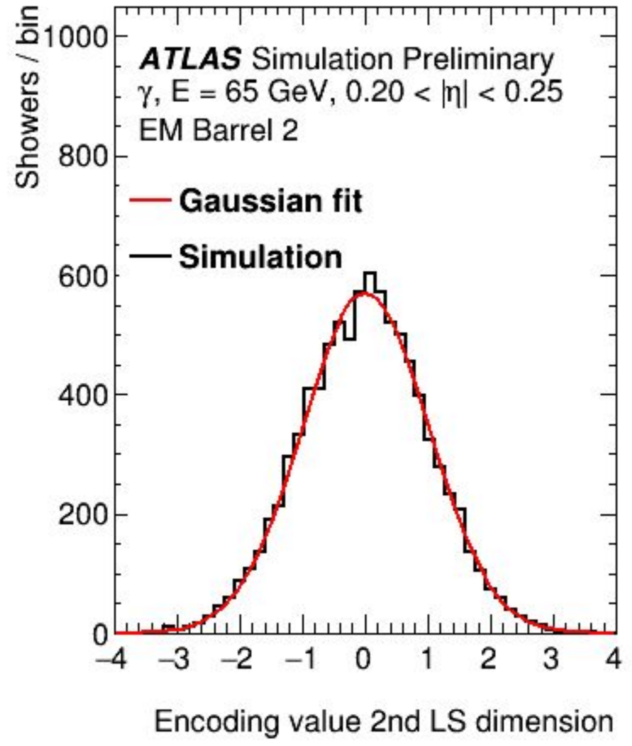
- Input : a variable with some specified ordering (multidimensional tensor)
- Output : (μ,σ) for each element of the input variable conditioned on the previous elements.

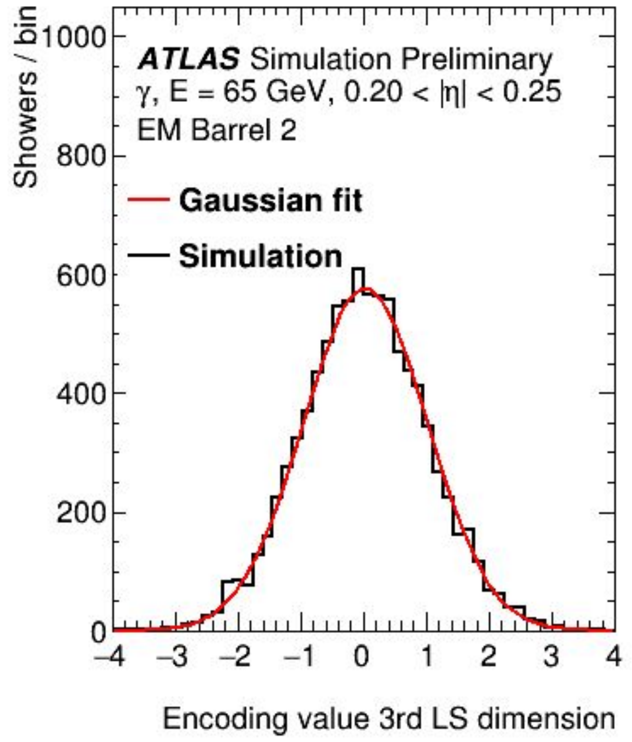
Inverse Autoregressive transformations

a type of Normalizing Flow to make the latent space more Gaussian

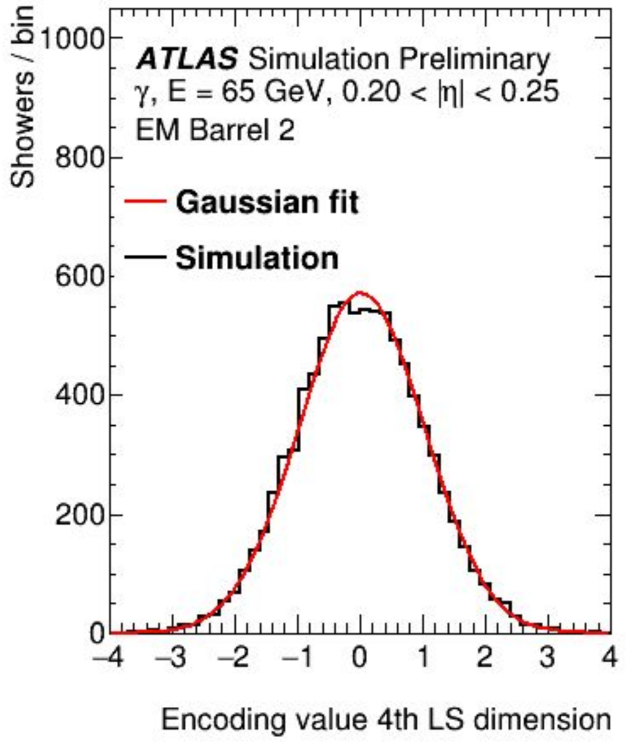
IAF transformations make the latent space distributions more Gaussian like.

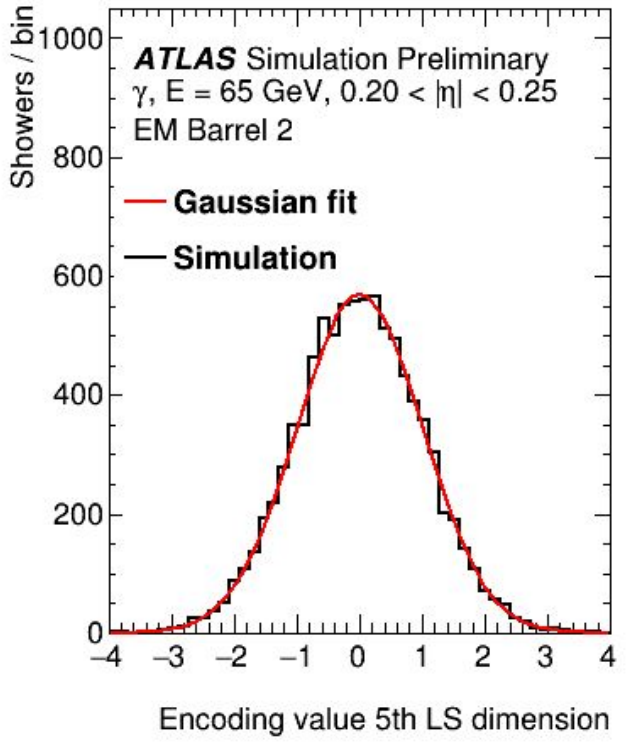






When we use the Decoder as a generator, it will be more correct to sample from a Gaussian distribution, impact on physics under study



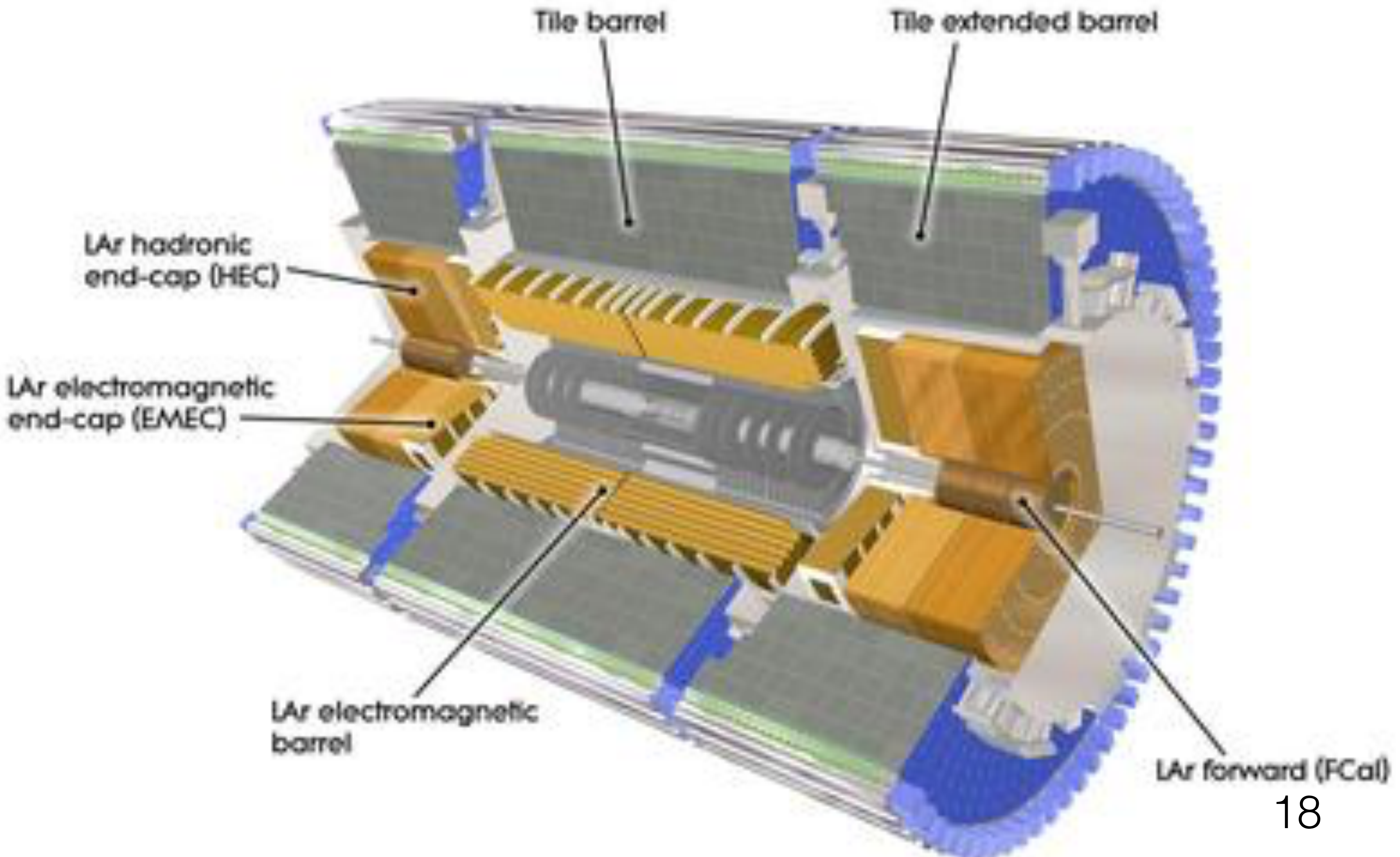


Integration of DNN into ATLAS (C++) Software

- Added Swish activation to [Light Weight Trained Neural Nnetwork package]() (Thanks Dan Guest!)
- Find cell closest to extrapolated position of particle in entrance of Middle Layer (called “Impact Cell”), it’s η, ϕ
- Build 266 cells around it, order in CaloLayer, η, ϕ increasing (mirror η on left half)
- Mimic preprocessing of GAN
- Generate energy with DNN in LWTNN
- Mimic post-processing of GAN, and fill energy into CaloCells
- Validation comparisons of **DNN** with **G4**, **FCS** using standard EGamma definitions (ATLAS internal)
    - Only photons in the barrel

Resource utilisation:

- DNNCaloGAN ~ same speed as FastCaloSimV2
    - LWTNN takes <1 ms per shower
- DNNCaloGAN VmPeak also small, not a concern



CaloGAN

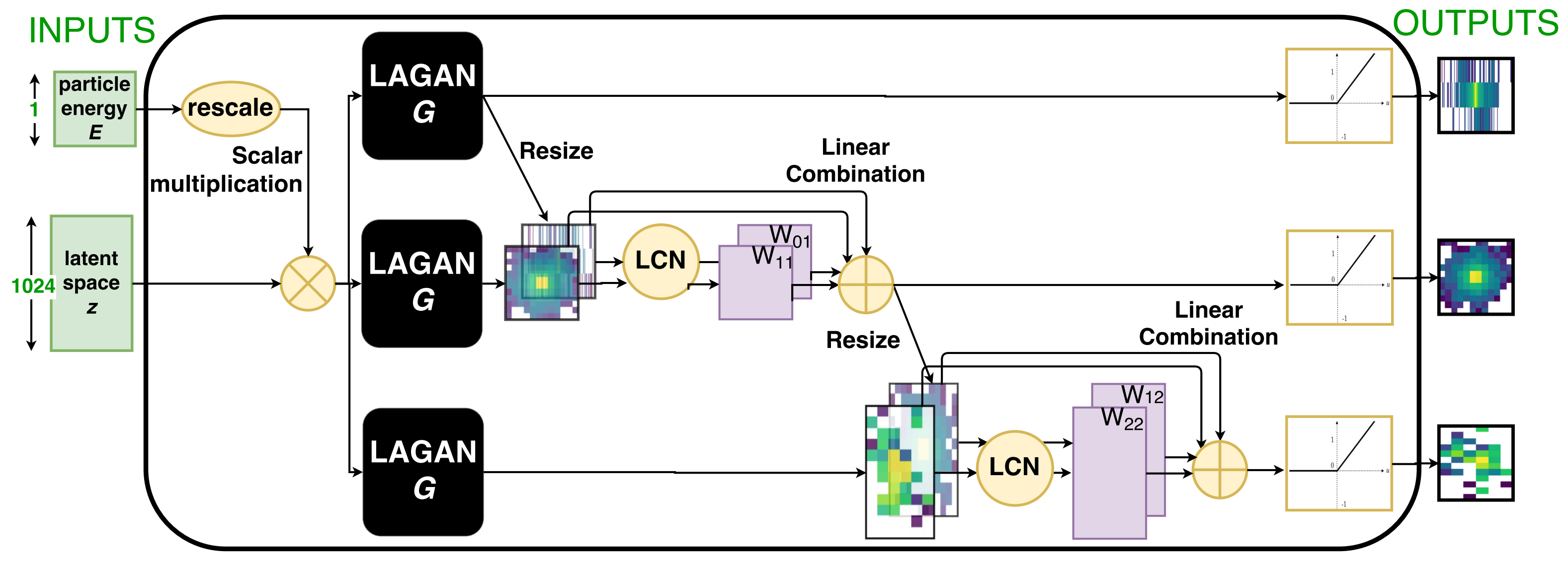


FIG. 4: Composite Generator, illustrating three stream with attentional layer-to-layer dependence.

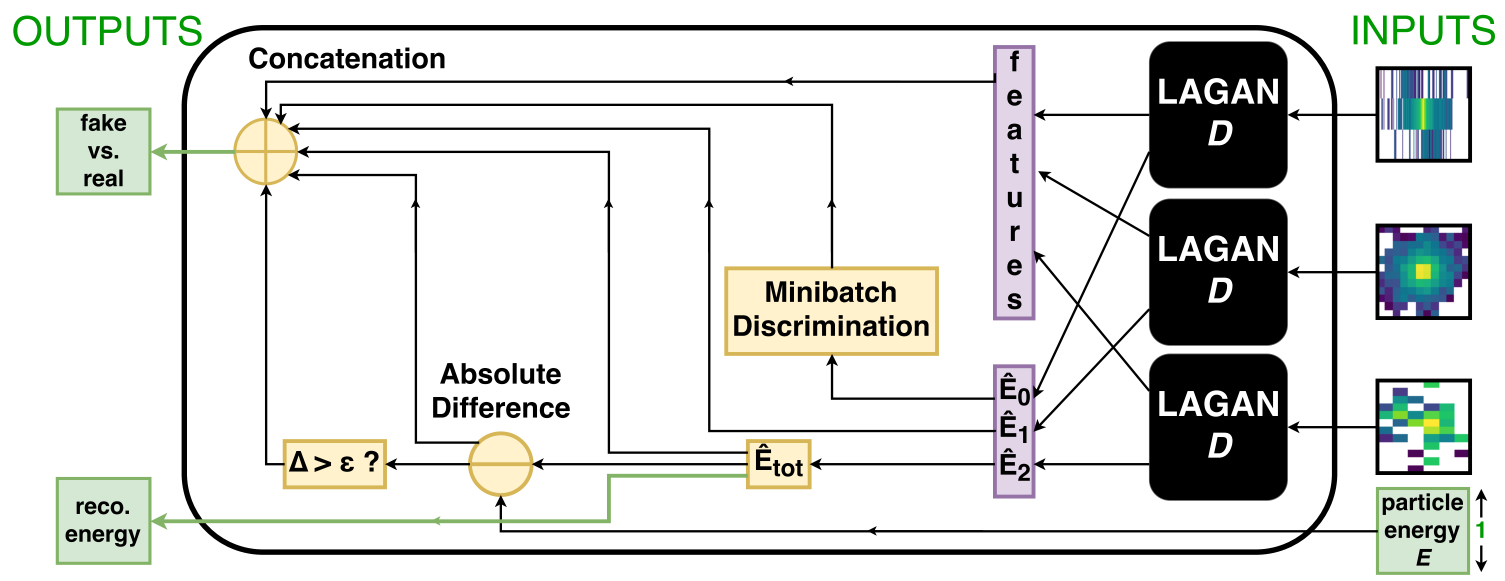


FIG. 5: Composite Discriminator, depicting additional domain specific expressions included in the final feature space.

[1712.10321](https://arxiv.org/abs/1712.10321)

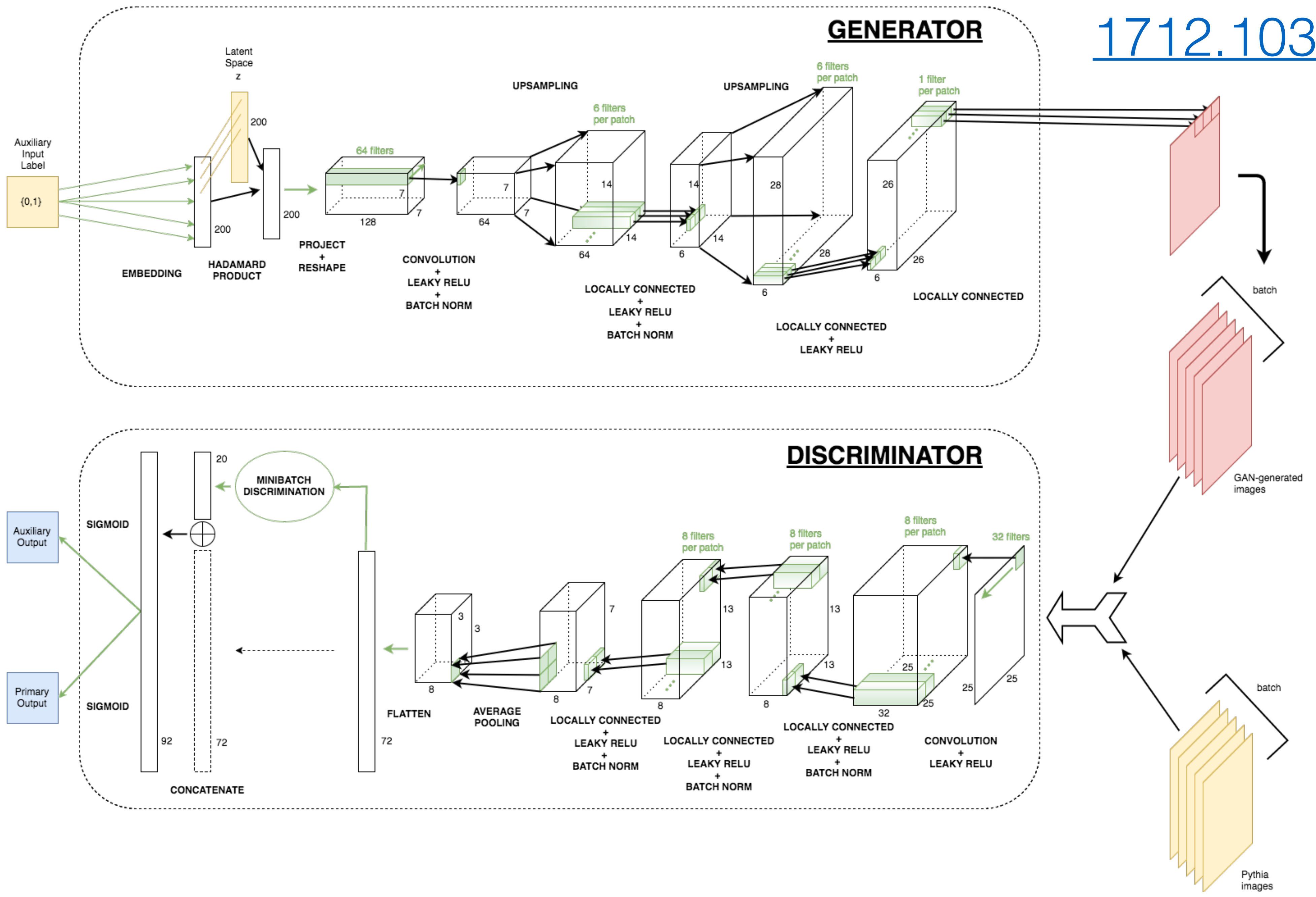


Figure 4: LAGAN architecture

The [CaloGAN](#) architecture (see [conditional version](#)). Now with GANCaloSim, a simpler architecture (Dense layers, 266 cells in 1 vector as input) achieves the various complicated conditionings on physics and calorimeter geometry (although only photons)

Lots of room to scale up architecture with Distributed Deep Learning

[1711.08813](https://arxiv.org/abs/1711.08813)

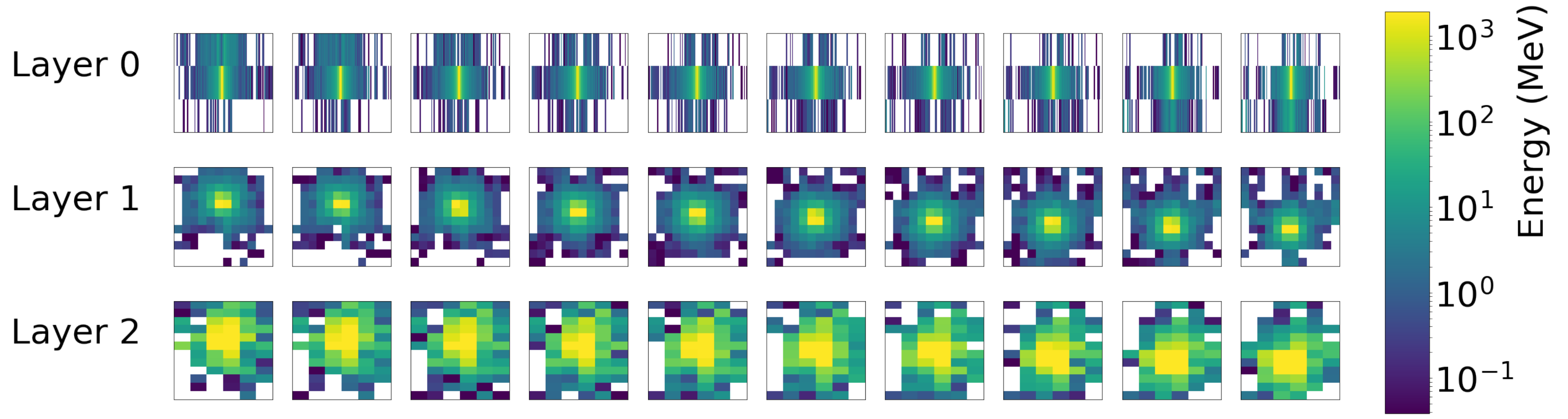


Figure 4. Interpolation across physical range of x_0 as a conditioning latent factor for e^+ showers.

Conclusion

- We have our first working Generative Network in ATLAS Software for our irregular shaped ATLAS Calorimeter (only within barrel for now) !
- Just a flag at run time allows to switch out one trained generative model for another
- Smart detector specific conditioning, preprocessing essential for good results
- No plot that GAN is unable to learn at all, now also getting the energy resolution correct
- GAN interpolates on untrained parameter space (don't need to train on 25GeV to produce 25GeV !)
- Comparisons being made on physics quantities with the in-development FastCaloSim V2 using established, time tested validation framework
- Many ideas to further improve performance: Look at plots at fixed energy points, Improve Strip images (Additional Critic/Grad Penalty/ Convolutions …)
    - Elaborate training (multiple Critics, Convolutions, DDL) ⇨ simple application (dense Generative network with Light Weight Trained Neural Network package in ATLAS Software)

- Future: More granular level data, Larger range in η, Transfer Learning from LHC data …

Backup

The Calorimeter

2-D Axis: ϕ vs η

Particle goes through 4 layers in this order:

0. Pre-Sampler : Some energy deposit

1. Strips: Very granular in η; more energy deposit

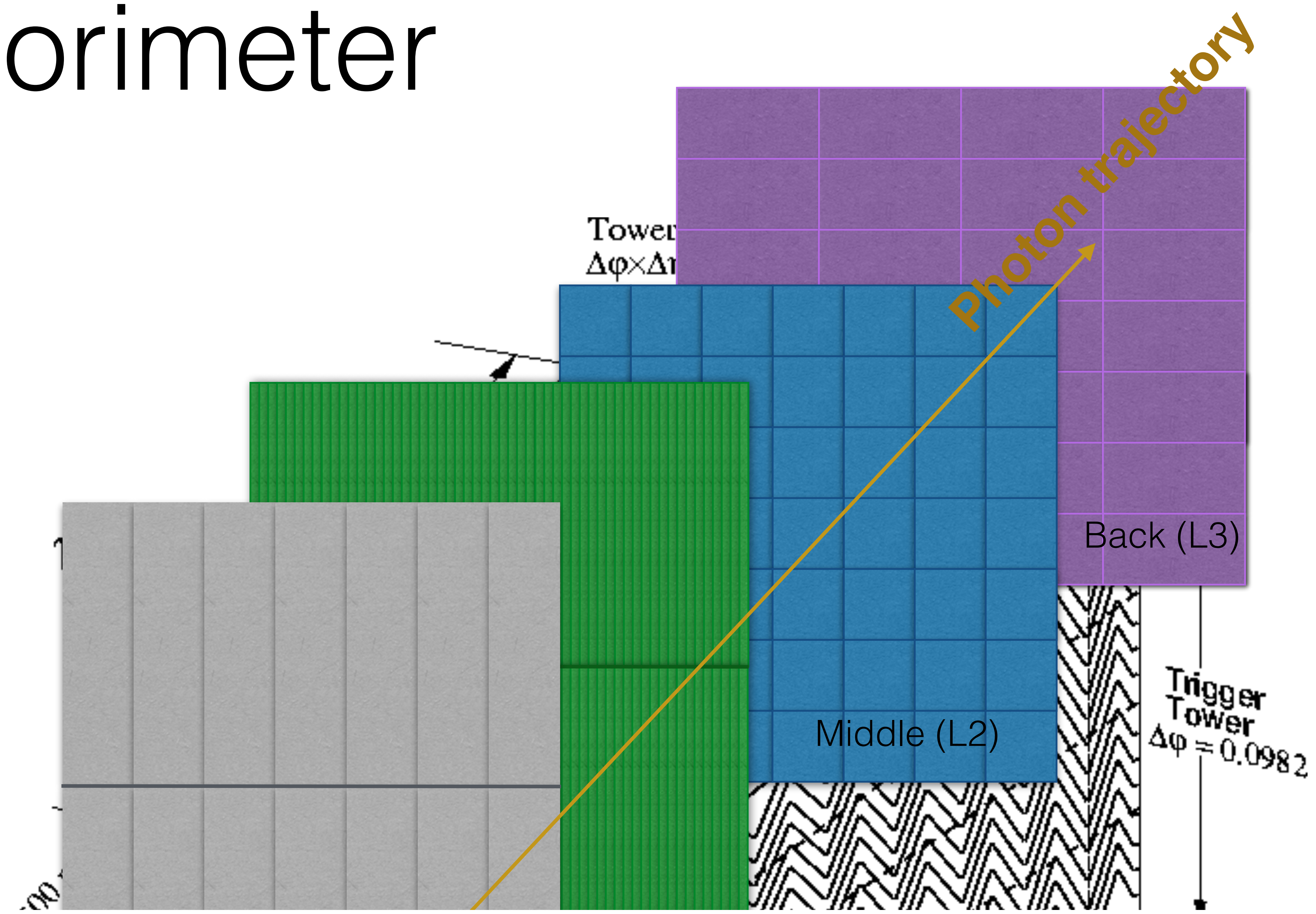
2. Middle: Thickest layer, maximum energy deposit

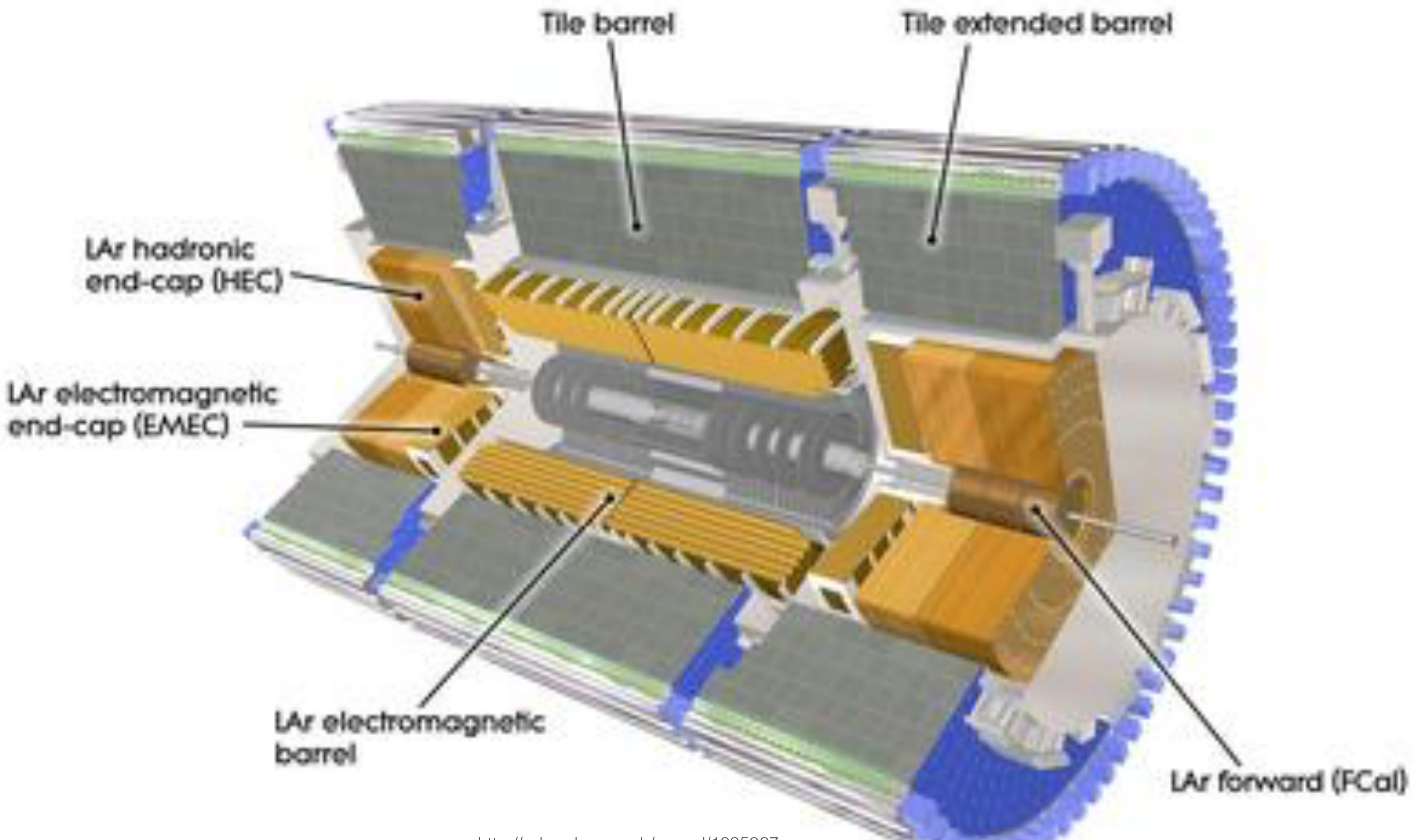
3. Back: Little Energy deposits

Due to misalignment of the two halves of the detector, cells are not perfectly well aligned.

Different widths of cells further complicate the alignment between cells of different layers

Cells not granular enough to see intricate details of shower pattern





http://cdsweb.cern.ch/record/1095927

Backpropagate through Sum?

When you train the Generator

Yes, gradients useful for Generator

When you train the E Critic

No, we don't want to apply gradient penalty to each cell via the Sum function

Treat Sum as independent input feature, not as a sum of the other 266 features