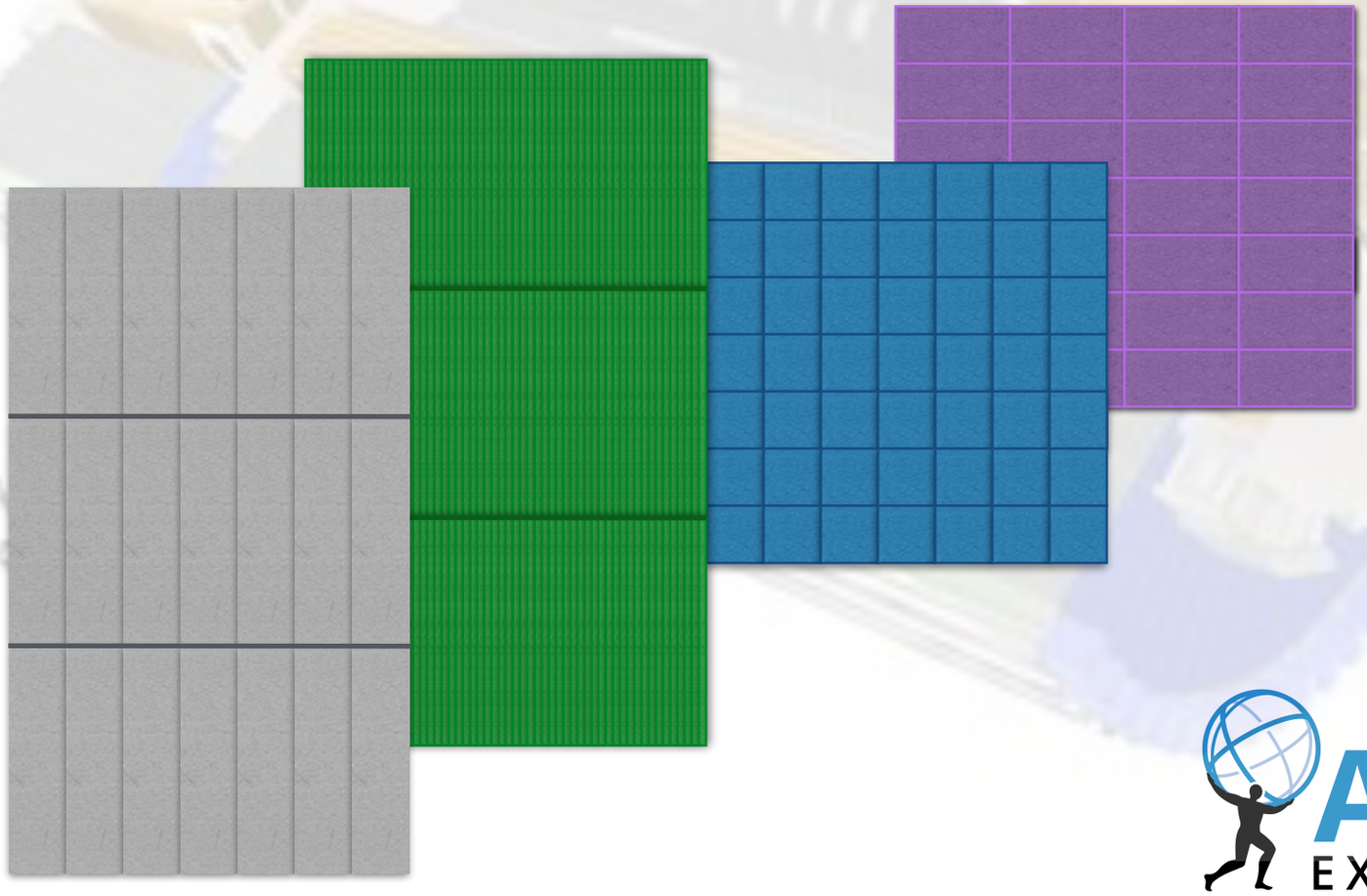
Simulating electromagnetic showers in ATLAS calorimeter with Generative Adversarial Network

Aishik Ghosh

Supervisor: David Rousseau

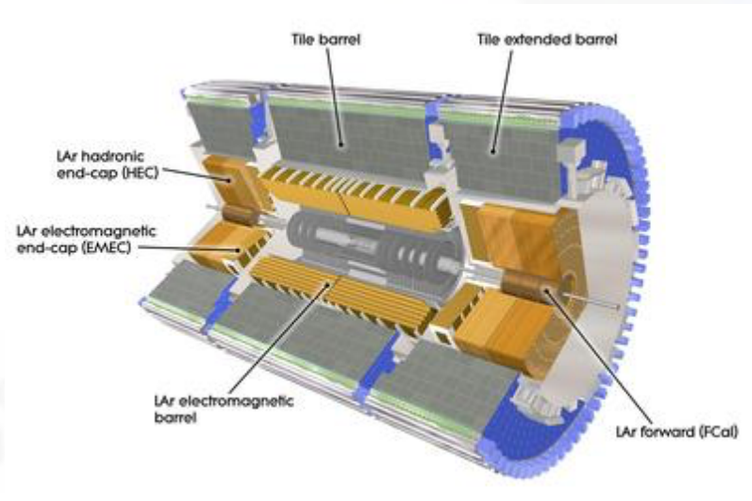
IN2P3 ML Workshop

29 March 2018

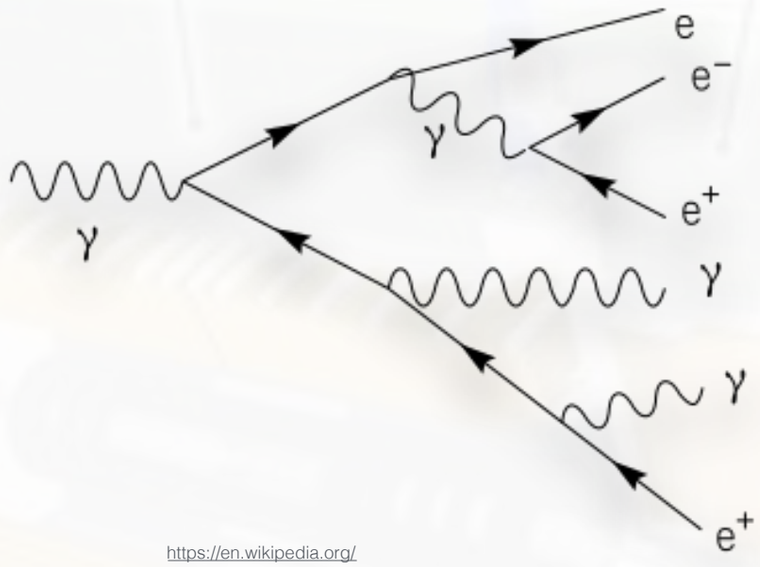


Physics Motivation

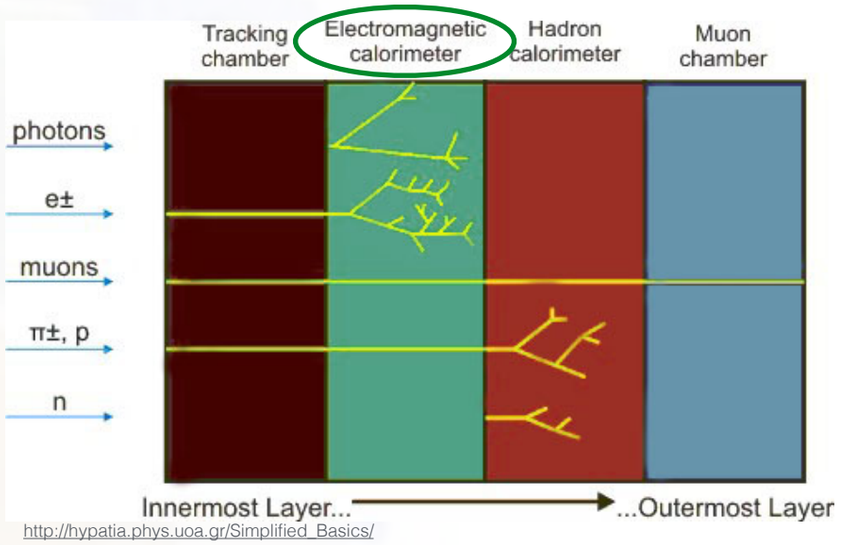
Photons deposit energy into the calorimeter cells as a shower



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



https://en.wikipedia.org/

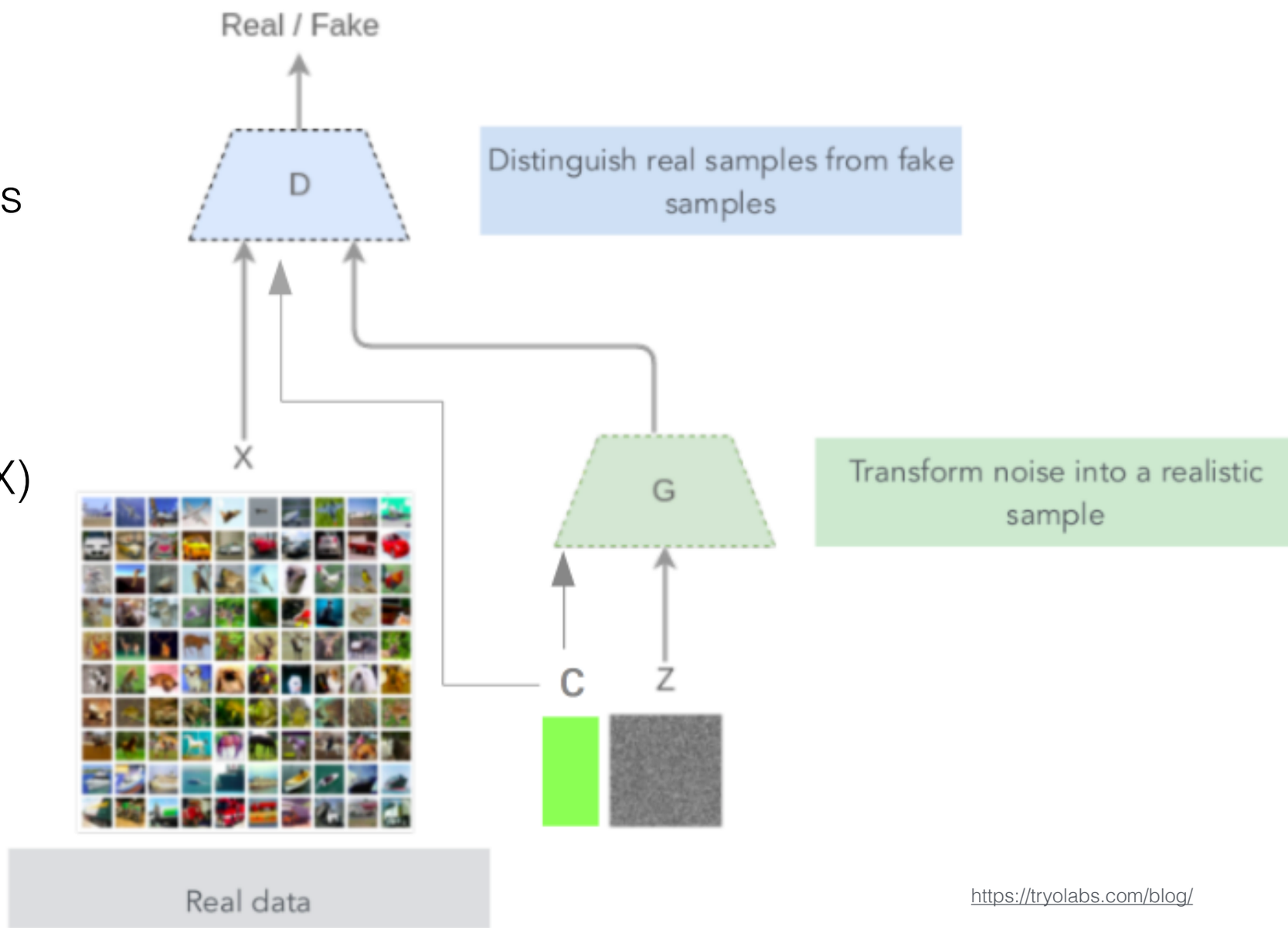


http://hypatia.phys.uoa.gr/Simplified_Basics/

- Currently >50 % of ATLAS computing time spent on shower simulation
- DNNs based generators could significantly speed up the process
    - An alternative to FastCaloSim Parameterisation, Frozen Showers approaches
- Training outside Athena (full ATLAS software framework), apply inside Athena
    - Use state-of-the-art DNN packages, GPU farms for training
    - Use Lightweight Trained Neural Network to integrate into Athena

What are GANs?

- Generative neural network based model to output images
- G: Generative network takes noise (Z) as input, outputs an image
- D: Discriminator Network learns to differentiate real (X) and fake images (from G)
- D helps G learn the distribution of real data
- Can produce conditional output based on C (e.g. Particle Energy)



https://tryolabs.com/blog/

The Calorimeter

2-D Axis: ϕ vs η

Particle goes through 4 layers in this order:

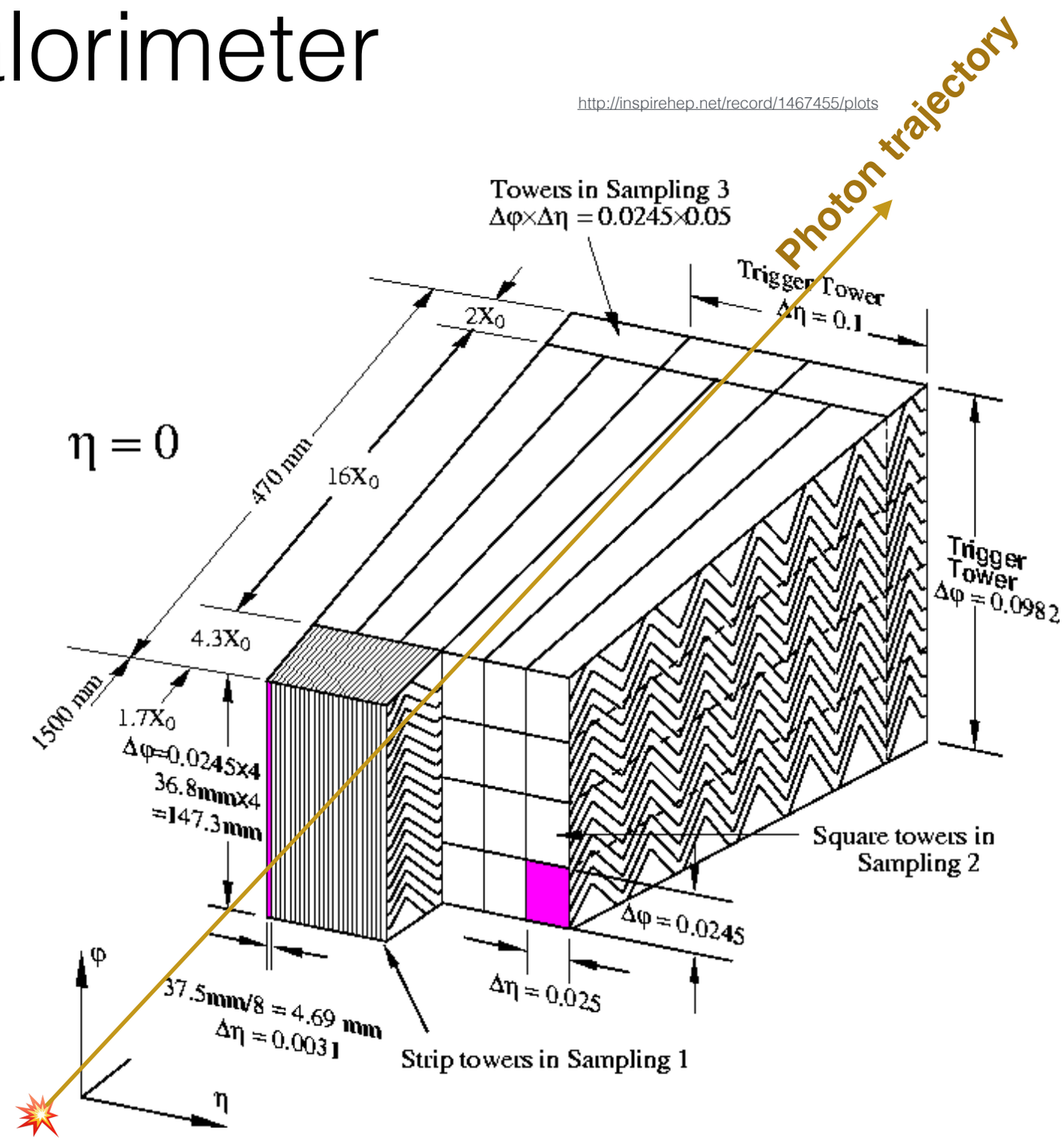
0. Pre-Sampler : Some energy deposit

1. Strips: Very <u>granular in η;</u> more energy deposit

2. Middle: Thickest layer, <u>maximum energy deposit</u>

3. Back: Little Energy deposits

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



The Calorimeter

2-D Axis: ϕ vs η

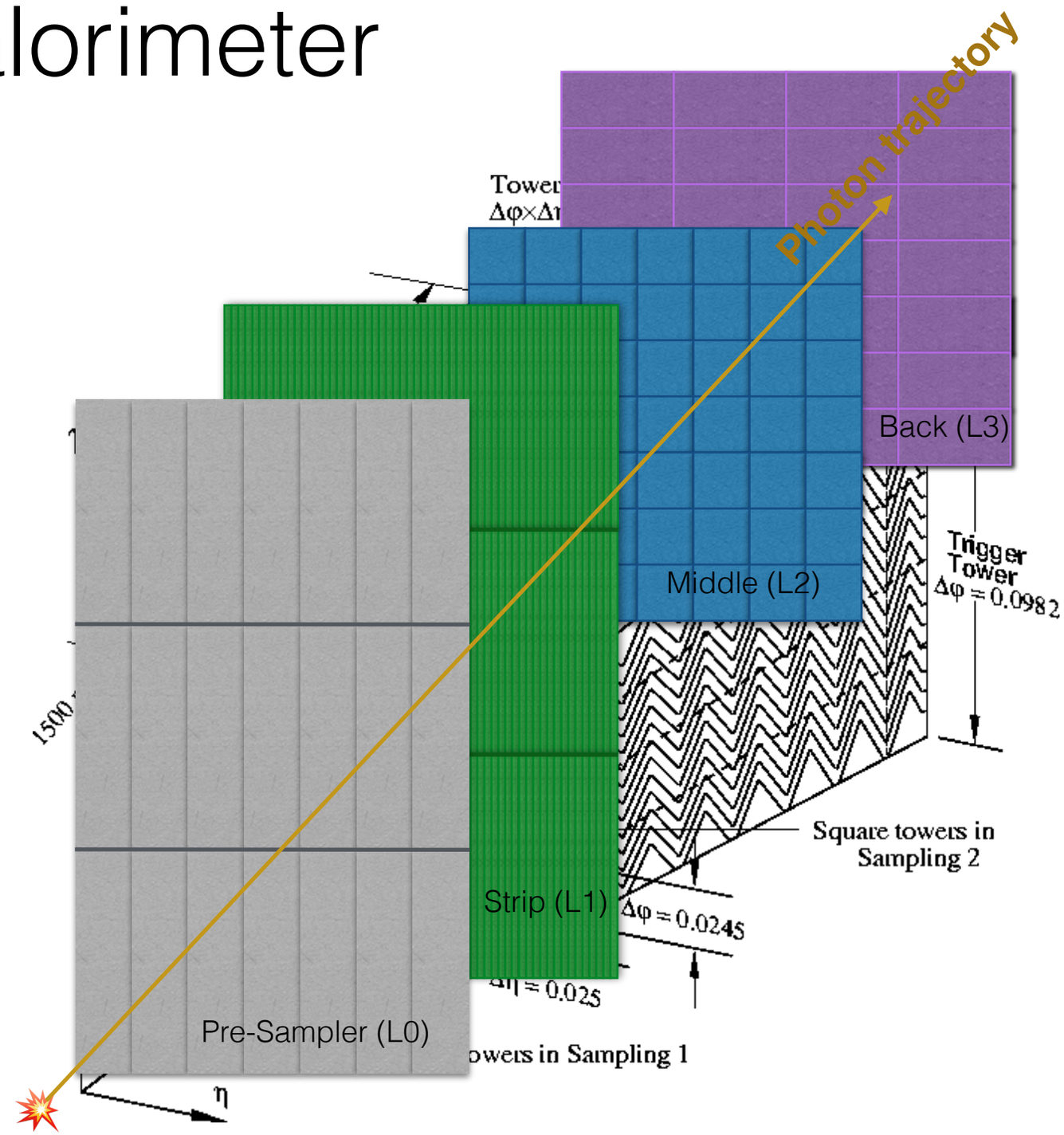
Particle goes through 4 layers in this order:

0. Pre-Sampler : Some energy deposit

1. Strips: Very <u>granular in η;</u> more energy deposit

2. Middle: Thickest layer, <u>maximum energy deposit</u>

3. Back: Little Energy deposits



The Calorimeter

2-D Axis: ϕ vs η

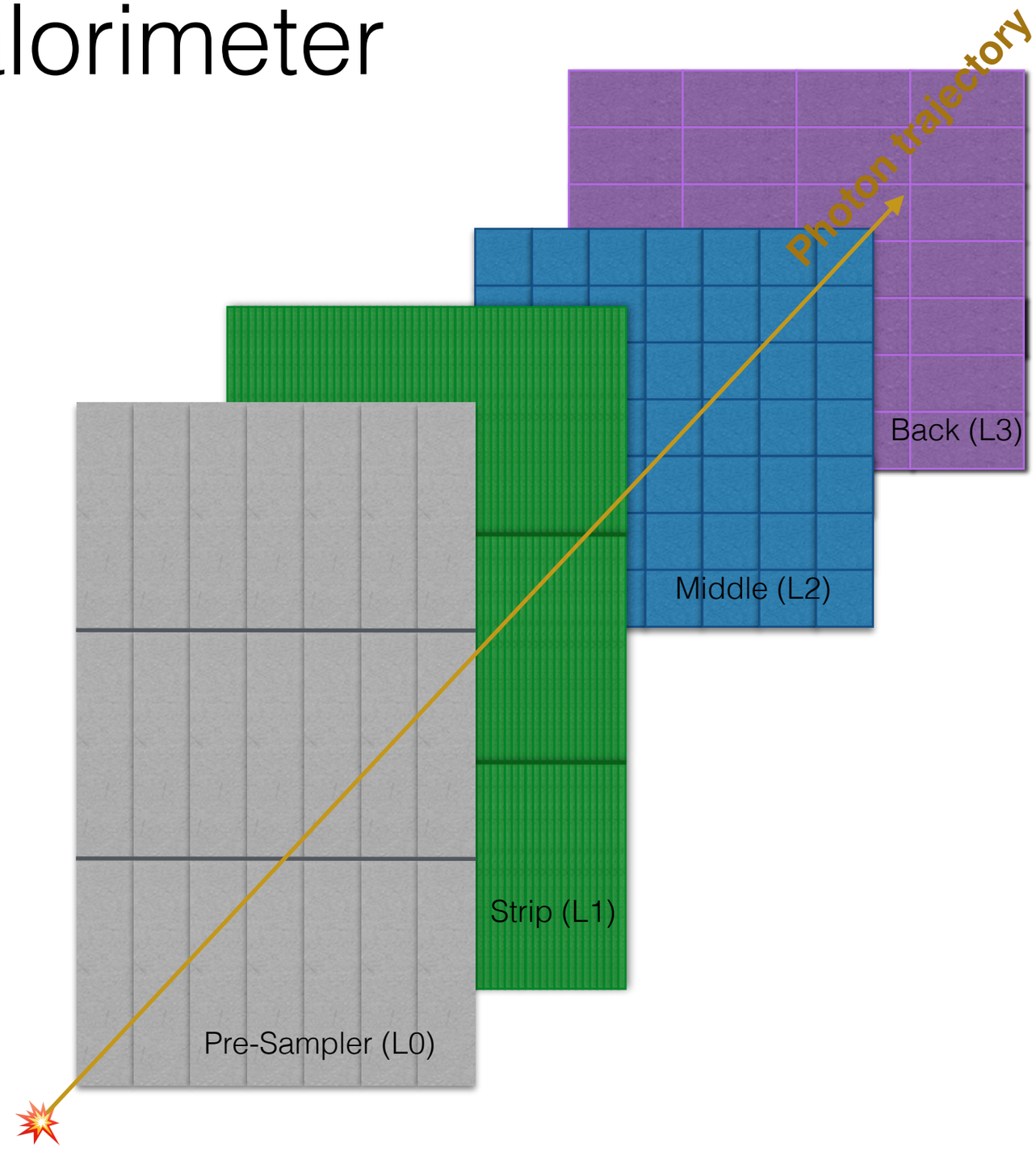
Particle goes through 4 layers in this order:

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



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

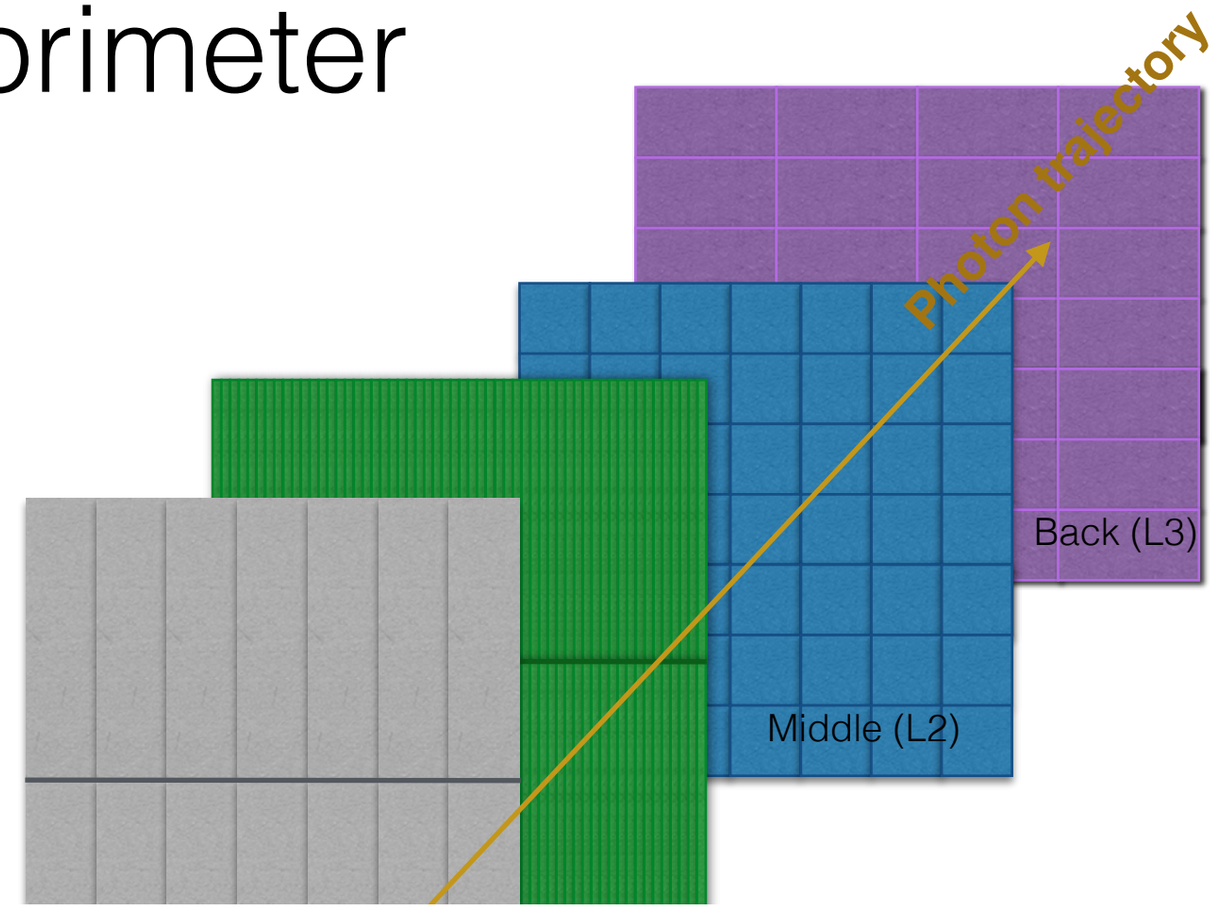
2. Middle: Thickest layer, maximum energy deposit

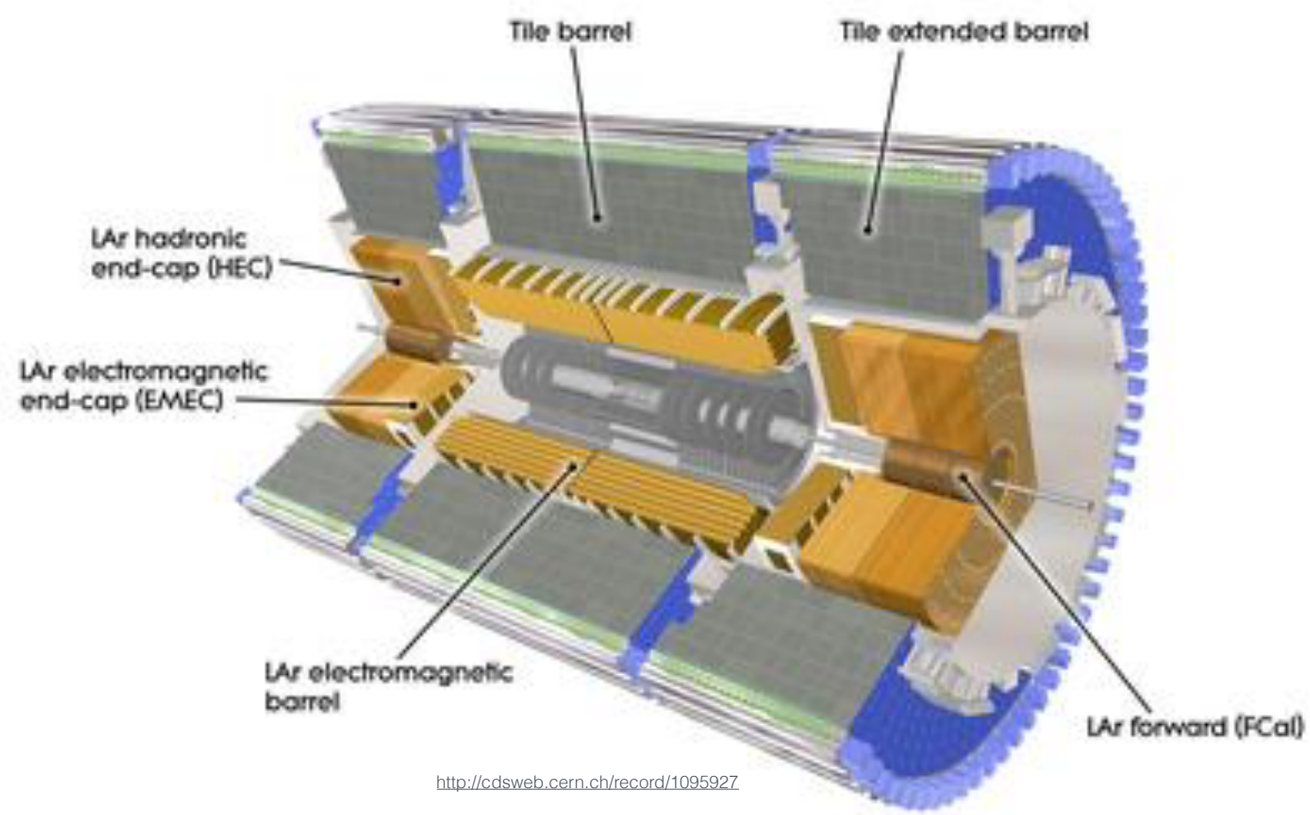
3. Back: Little Energy deposits

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

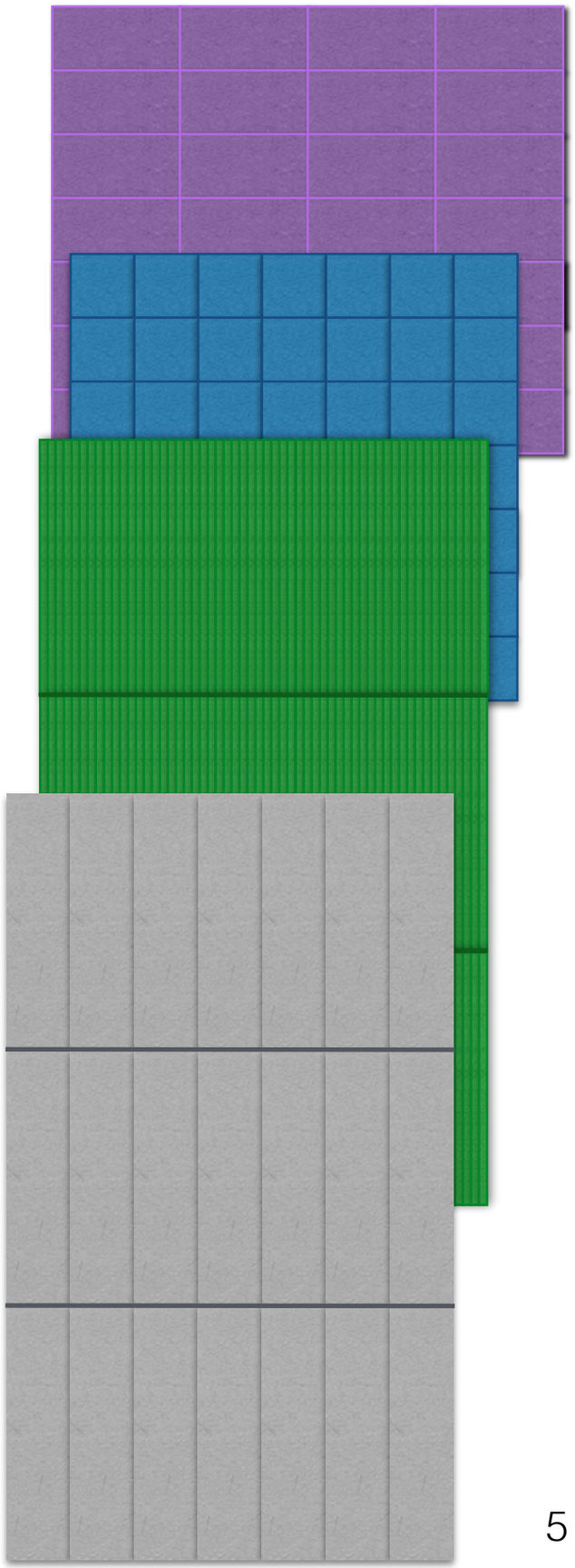
Cells not granular enough to see intricate details of shower pattern





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

Our 'Images'



Calo Cell → Pixel

Energy deposited in Calo Cell → Pixel Intensity

Convert 266 Cells from 4 Calo Layers into 1-D Numpy Array Input for GAN

GAN should catch the correlations between cells (and correlations between calo layers)

Can it learn the 3D distribution without any spatial information?

Software

Train using **Keras** (high-level wrapper) with **Tensorflow** backend

Train on **1 GPU at CC Lyon**

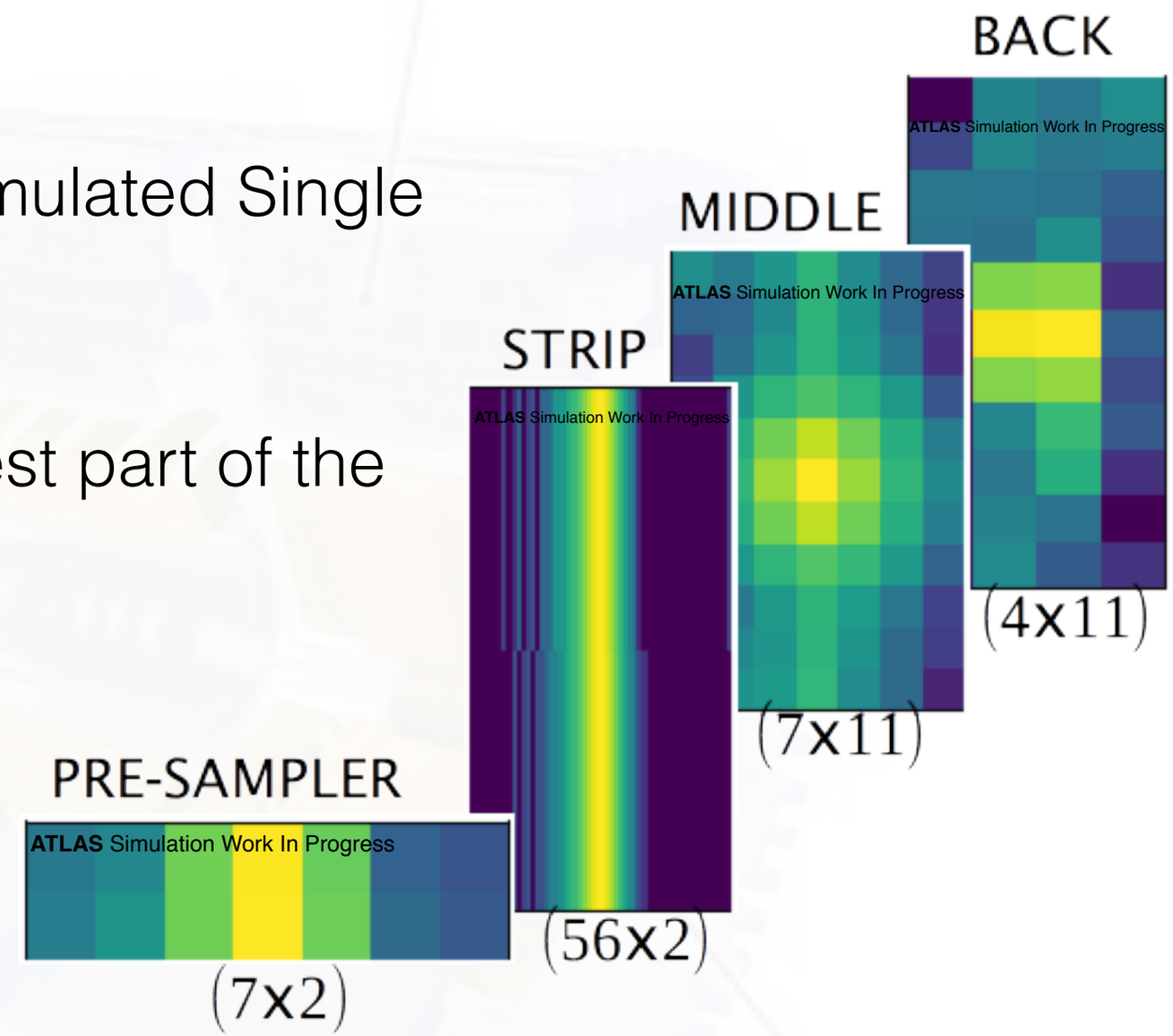
Plan to apply using **Lightweight Trained Neural Network**:

Converter: Keras → C++ framework

"If you have the flexibility to run any framework in your production environment, this package is *not* for you. **If you want to apply a network you've trained with Keras in a 6M line C++ production framework** that's only updated twice a year, you'll find this package very useful."
-from lwtnn Github page

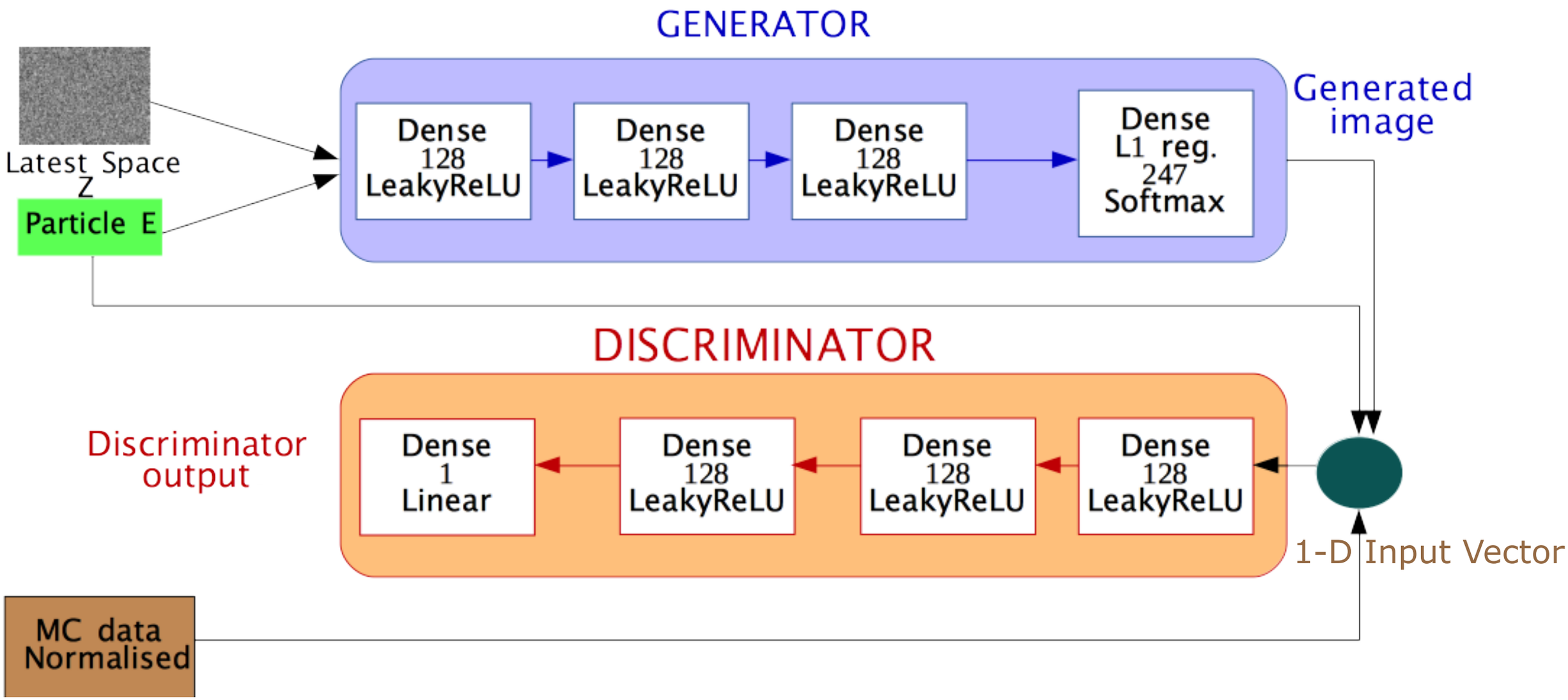
Dataset

- Train on **Geant4 Monte-Carlo** simulated Single Photon shower data
    - Use only the 'barrel' (simplest part of the calorimeter)
- GAN Training time: ~ 1h30m
- GPU speed: 2x over CPU
- GAN Generation Time: 1-2 seconds for 36k images



On the basis of Paul Klein's internship work with David Rousseau, Gilles Louppe, Kyle Cranmer at LAL & CERN, 2017

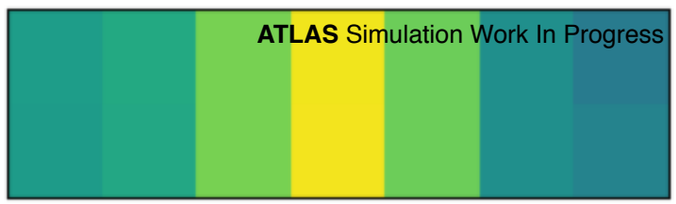
Our WGAN-GP Architecture



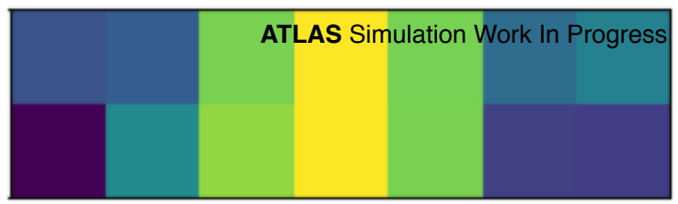
Total Trainable Parameters: G=103543, D=65025

Average Images

Geant4 MC

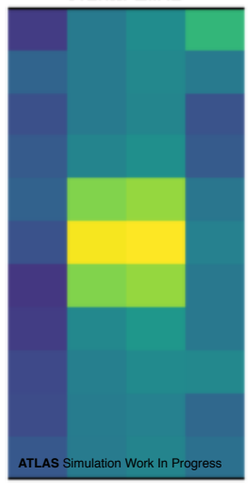


WGAN

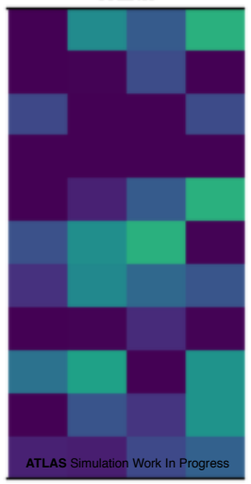


PRE-SAMPLER

Geant4 MC



WGAN

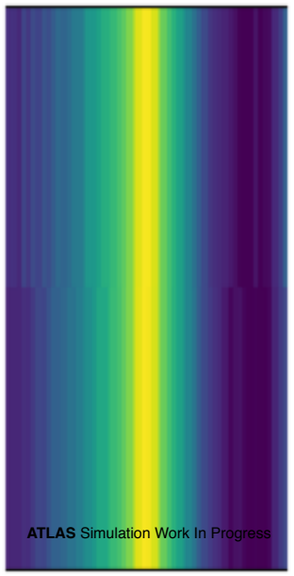


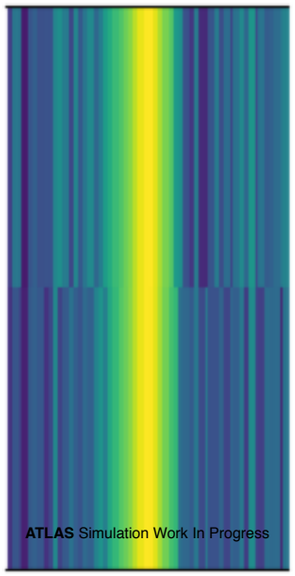
BACK

Geant4 MC

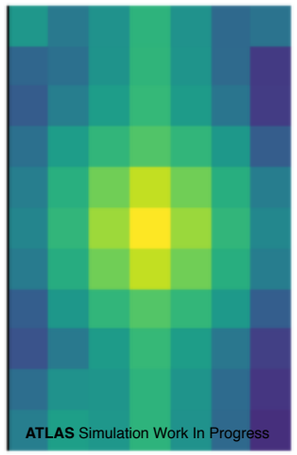
WGAN

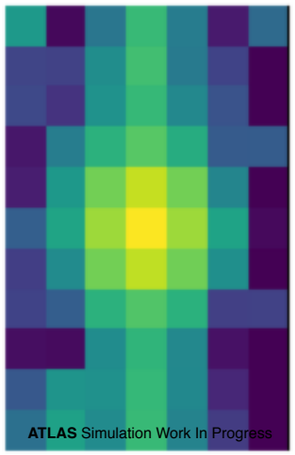
STRIP





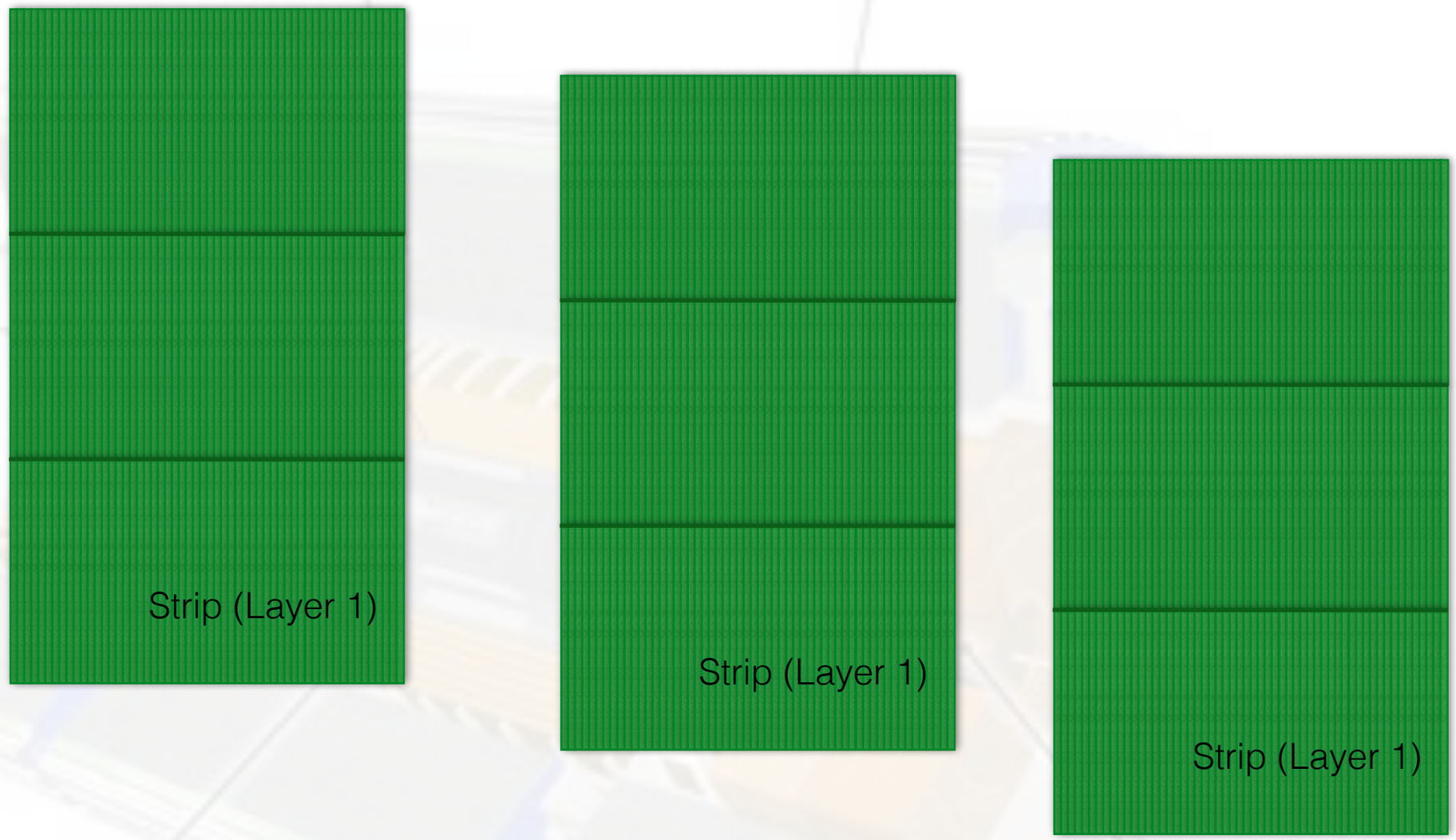
MIDDLE





Warning: pixel sizes don’t correspond to actual width of the cells

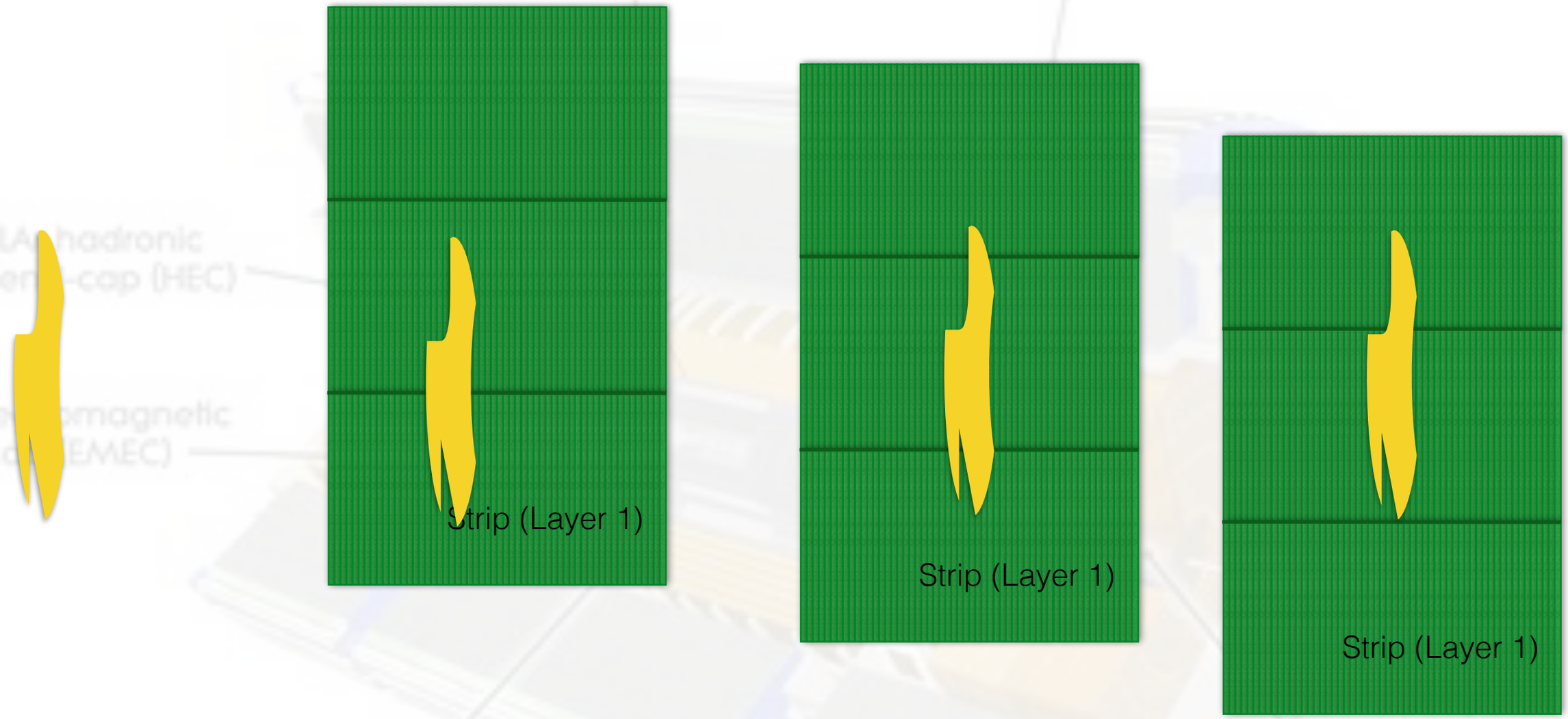
Underlying Geometry Complication



Images are coming from different parts of the detector, with different alignment between calo layers

Same shower pattern, different image!

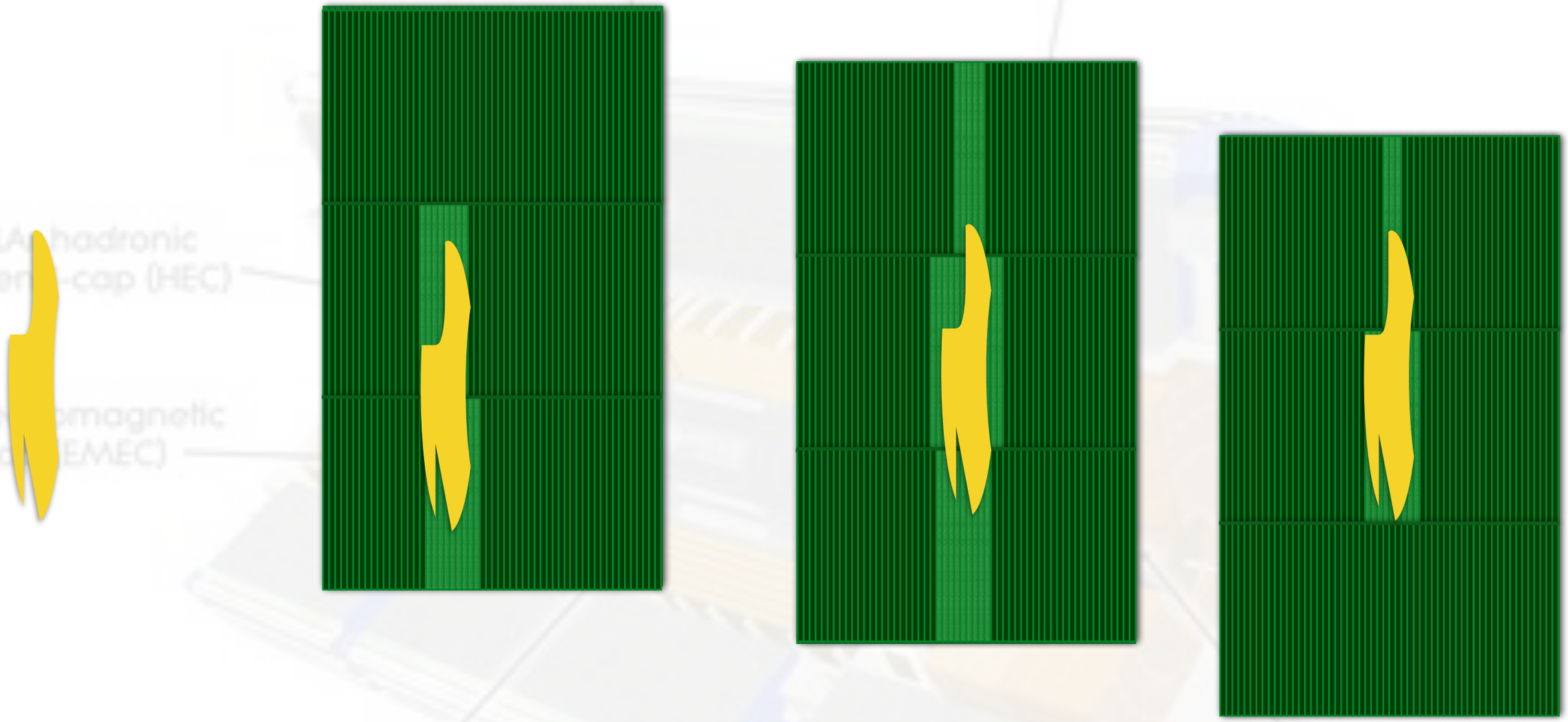
Underlying Geometry Complication



Images are coming from different parts of the detector, with different alignment between calo layers

Same shower pattern, different image!

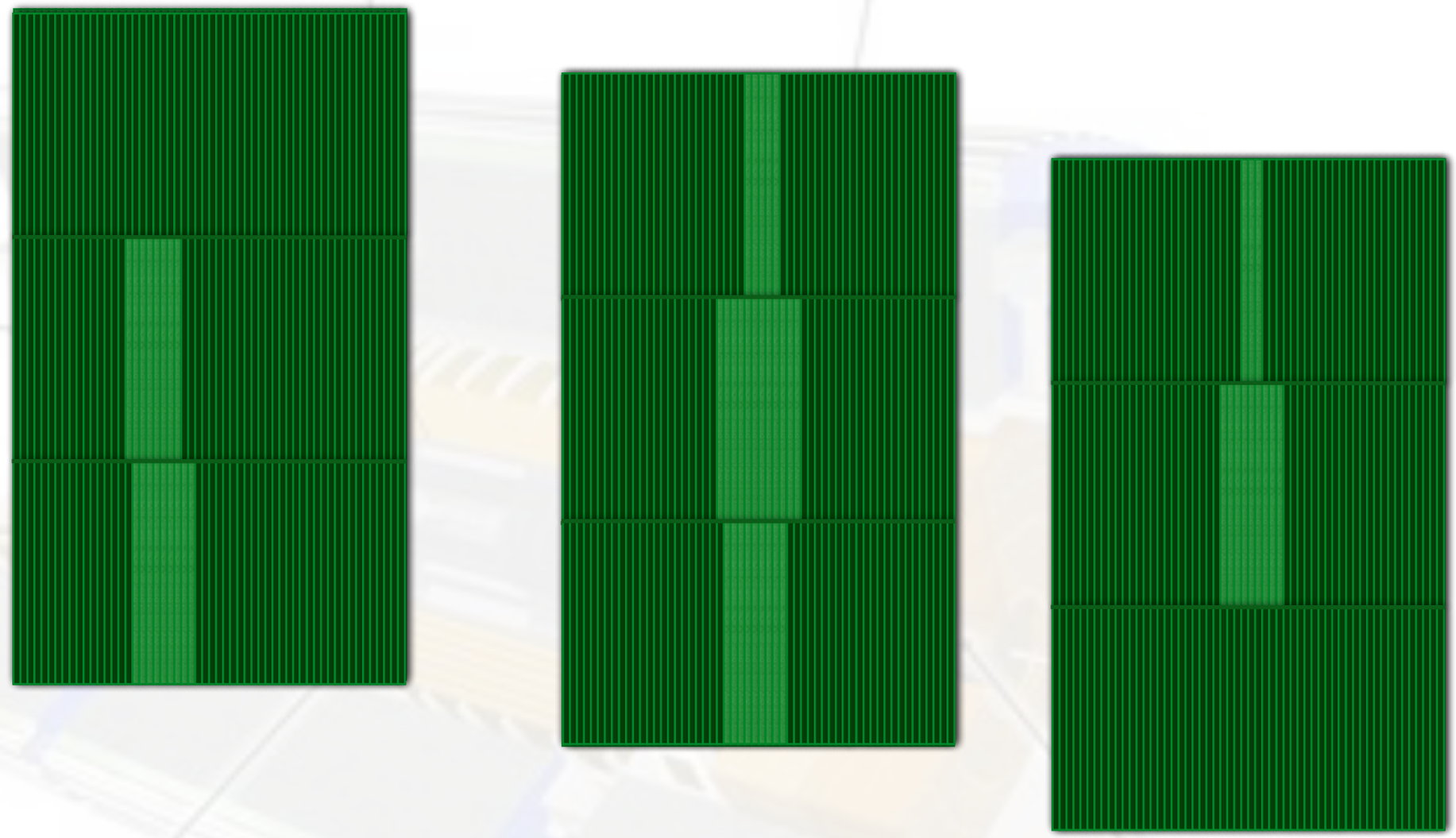
Underlying Geometry Complication



Images are coming from different parts of the detector, with different alignment between calo layers

Same shower pattern, different image!

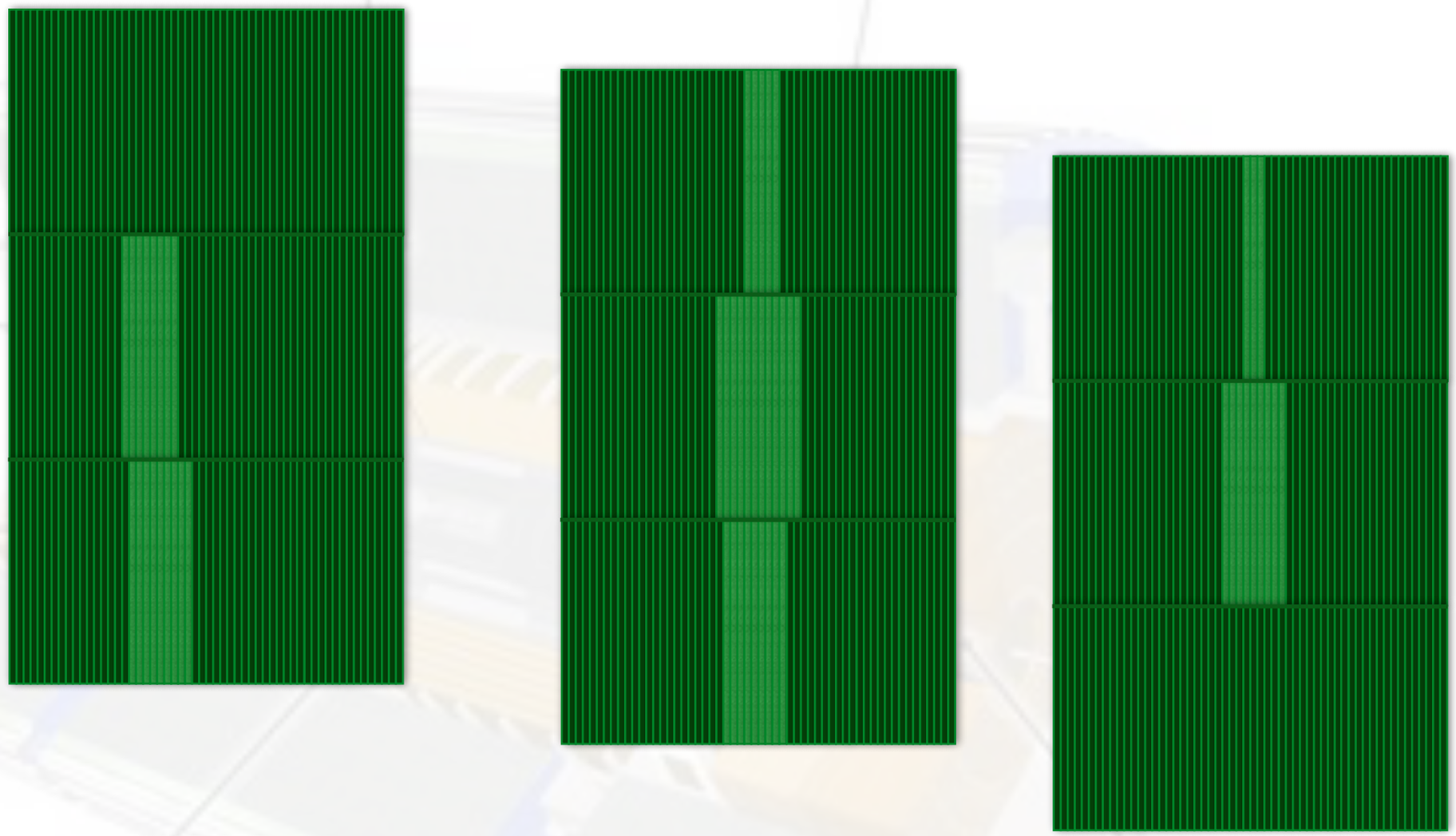
Underlying Geometry Complication



Images are coming from different parts of the detector, with different alignment between calo layers

Same shower pattern, different image!

Underlying Geometry Complication



Train separately on images coming from the same configuration of alignment

⇒ Smaller training dataset

Note: CaloGAN did not see this problem because particles shoot at single impact point in calorimeter

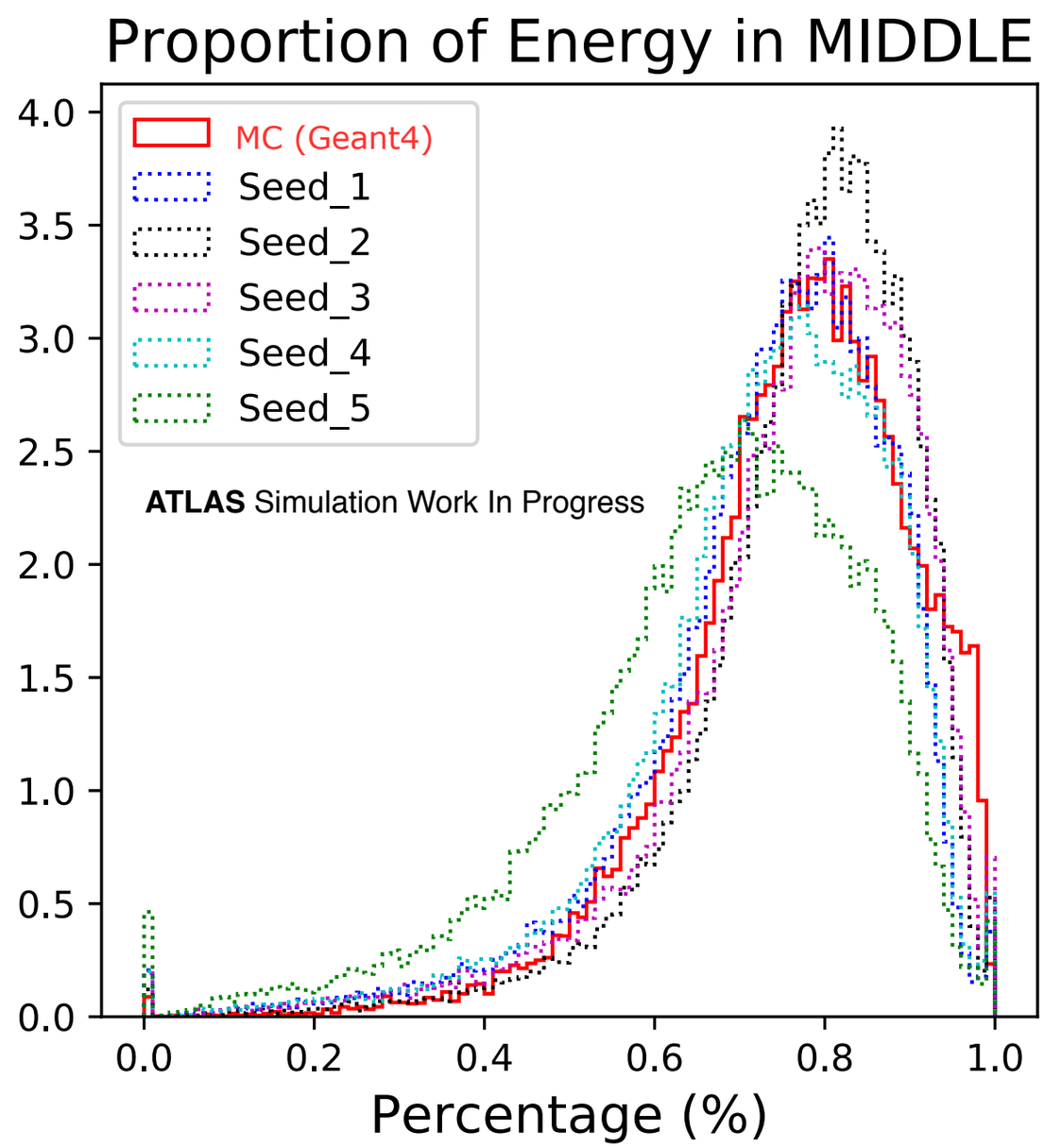
Variation in Training for Random Seed

Training Seed Affects:
1. Initial random weights
2. Random noise input (Z) for generator
3. Batch Selection

Additional Randomness from GPU parallelisation

Do we pick the best one or do we average over 5?

Different training seeds

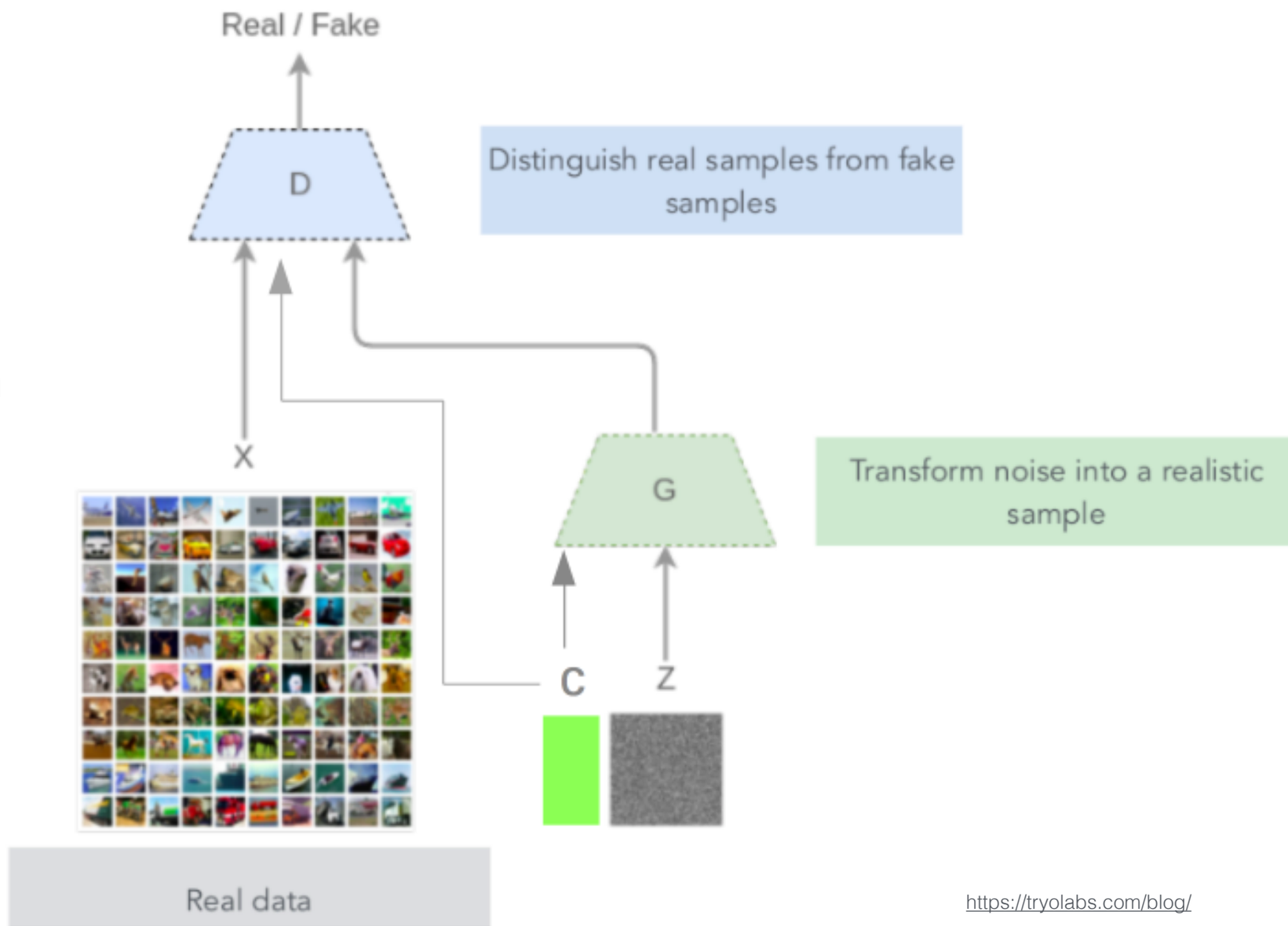


Initial training seed has significant impact on results. **Average over different seed** training to for Hyper-parameter optimisation

Remember the GAN game?

If D is trained too little, it will get fooled by G too easily, so G won't improve

Don't want to train D too much before every generator training



https://tryolabs.com/blog/

Training Ratio:

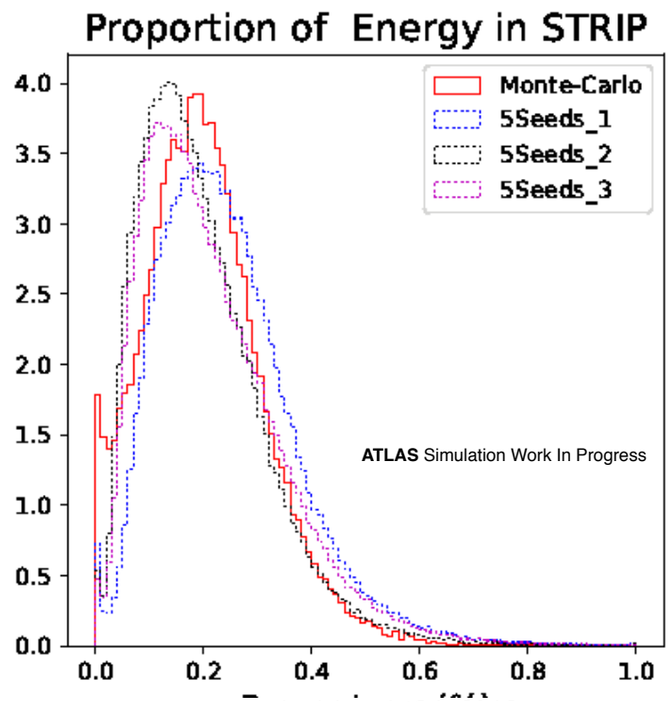
Number of discriminator training iterations for each generator iteration

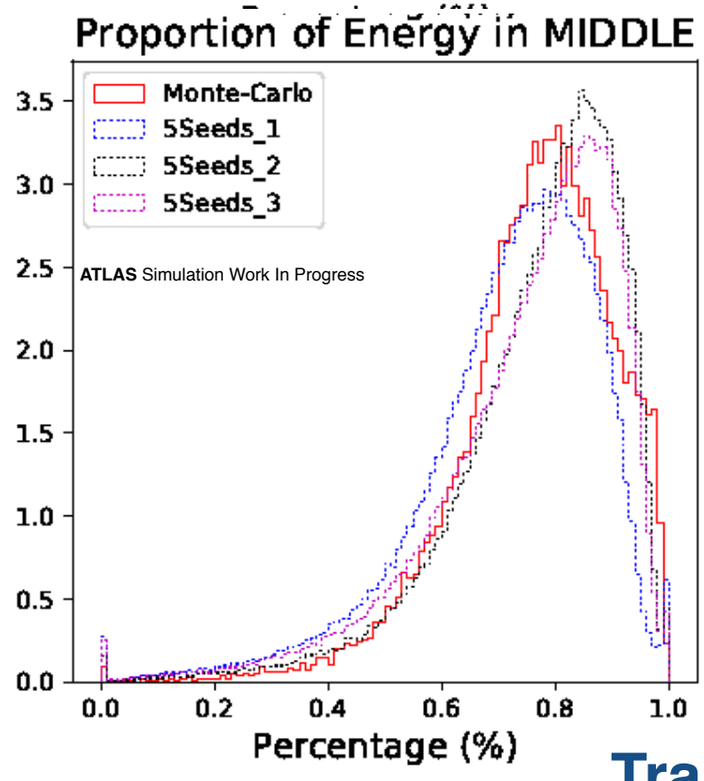
Hyperparameter Optimisation: Training Ratio

Same hyper-params,
3 sets of 5 separately trained GAN datasets

1,5,10 Ratio; combine 5 trained GANs for each

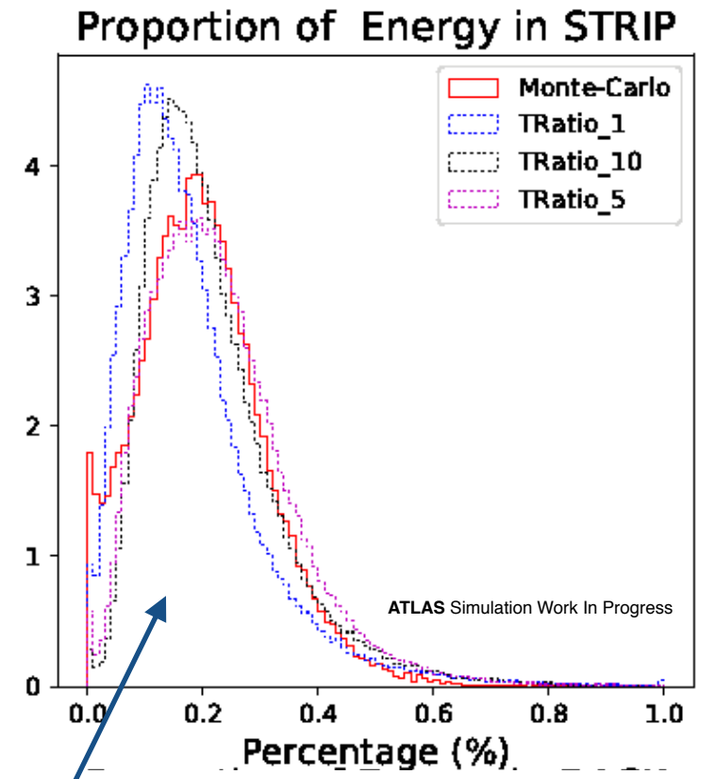
Random Variations in performance

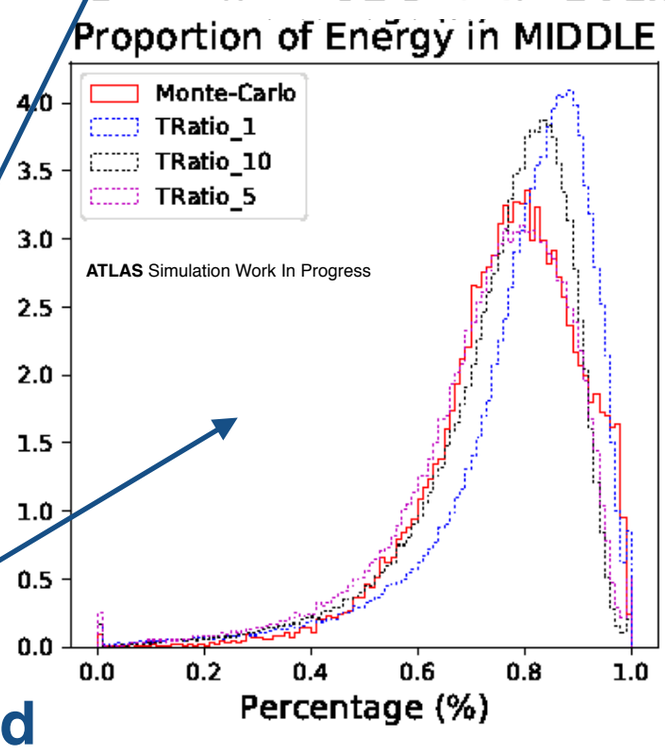




Monte-Carlo is from Geant4

Non-random improvement in performance





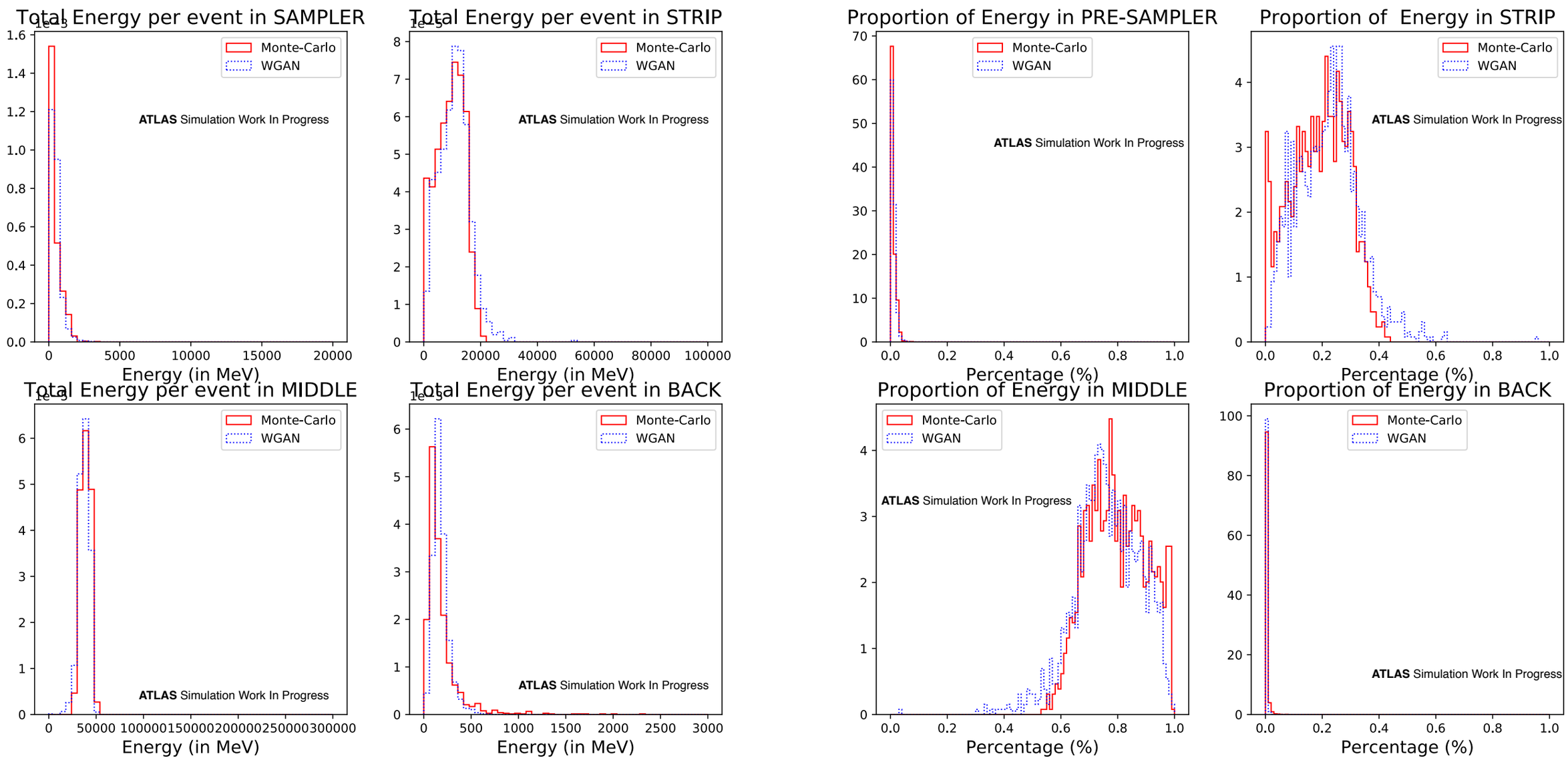
Training Ratio of 5 is good

Energy Distribution in Each Calo Layer

Total Energy

Fraction of Energy

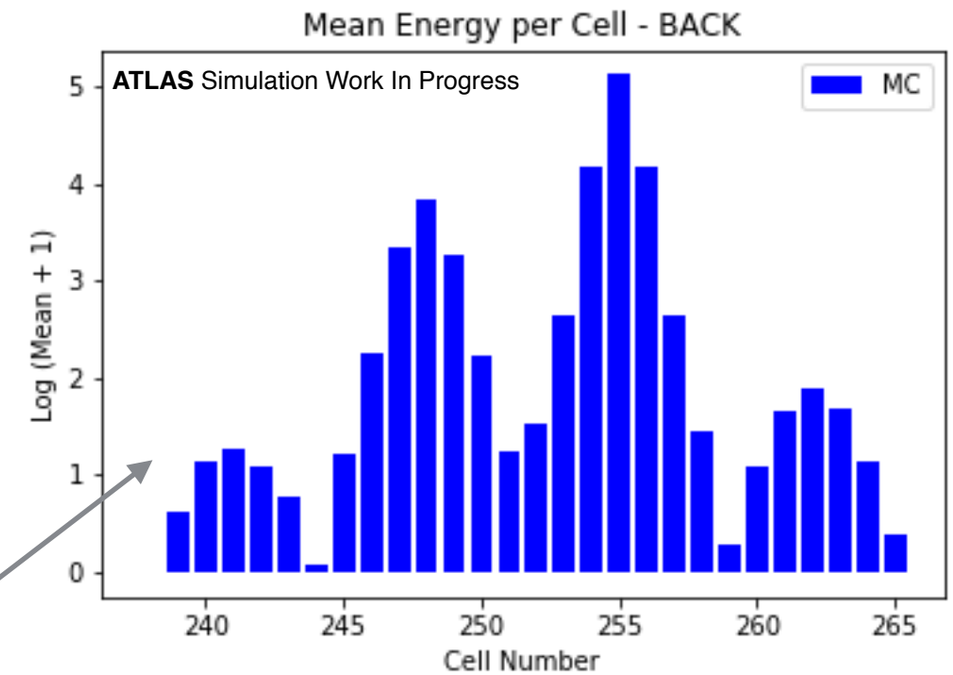
Monte-Carlo is from Geant4



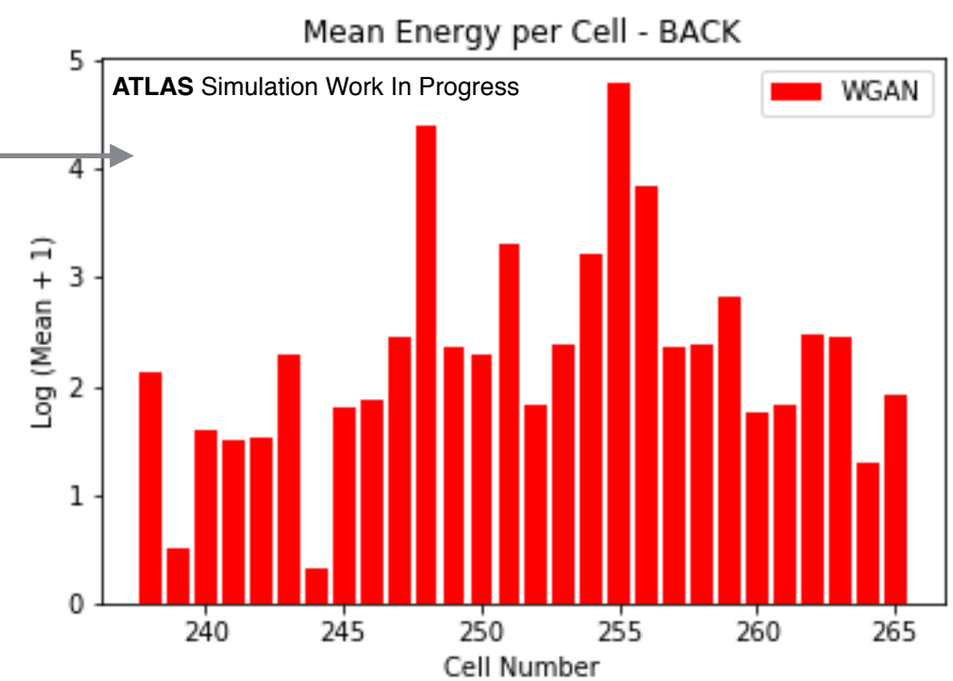
Energy Per Cell in BACK (L3)

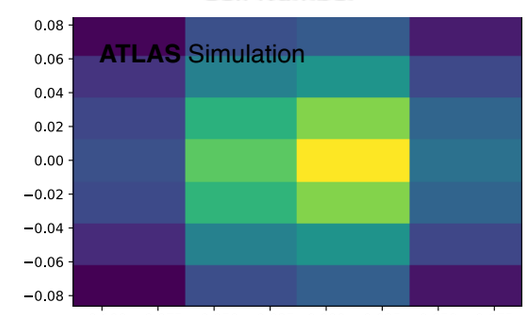
Monte-Carlo is from Geant4

The 4 peaks, periodic in 7 are now perfectly understandable

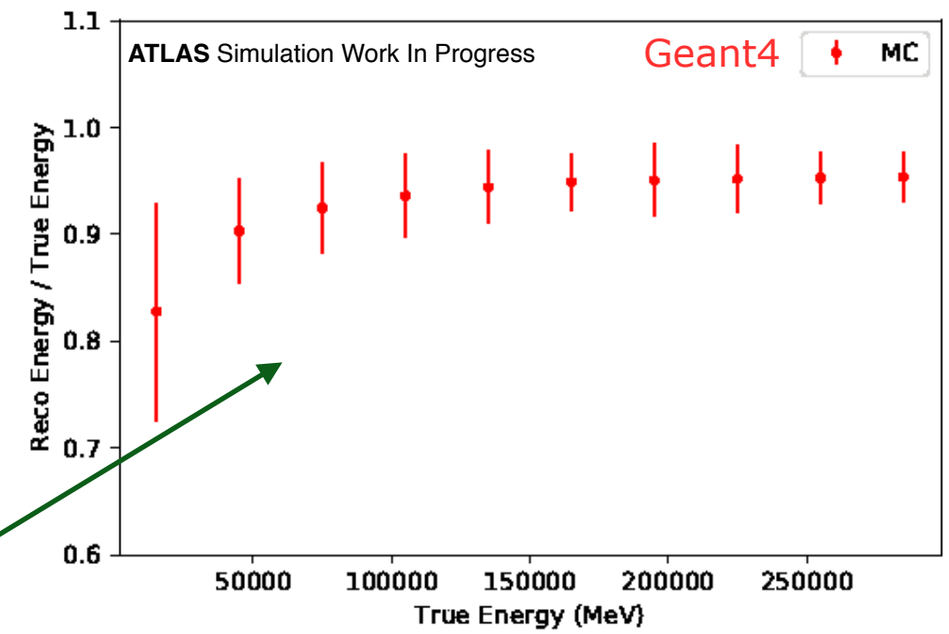


GAN doesn't learn this structure well





Energy resolution of Detector



The energy resolution of the detector gets better at higher energies:

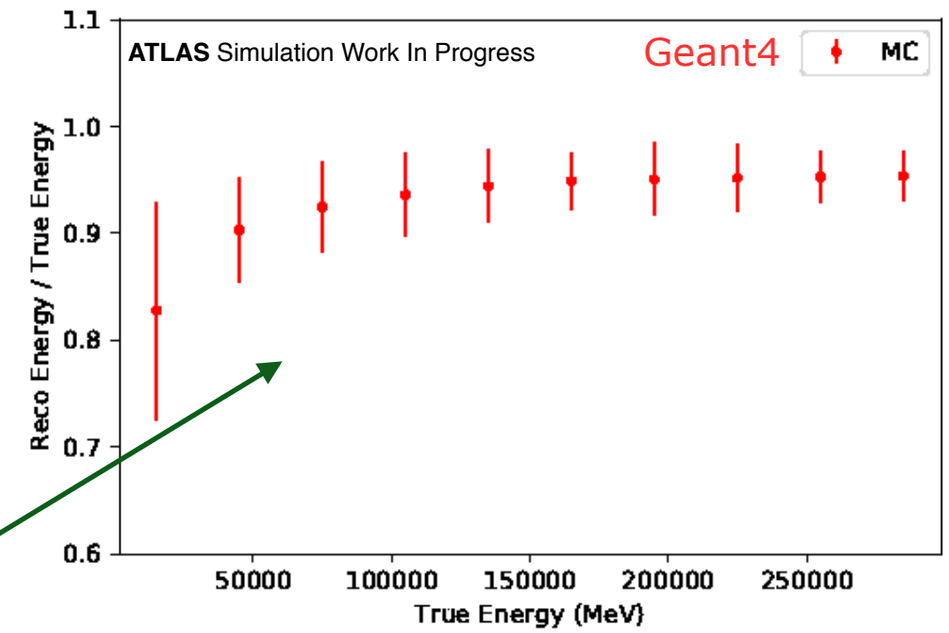
$$\sigma_E \sim 10\% \sqrt{E}$$

Can the GAN mimic this?

Energy resolution of Detector

To reproduce the energy resolution of the detector:

- Condition on True Energy of Particle
- 'Normalise' cell energies to by true energy of particle, **(Σ(cell_energies) ≠ 1 anymore)**
- Change activation function in the final layer from softmax to relu



$\sigma_E \sim 10\% \sqrt{E}$

The error bars are the **standard deviation in each bin**

Softmax:

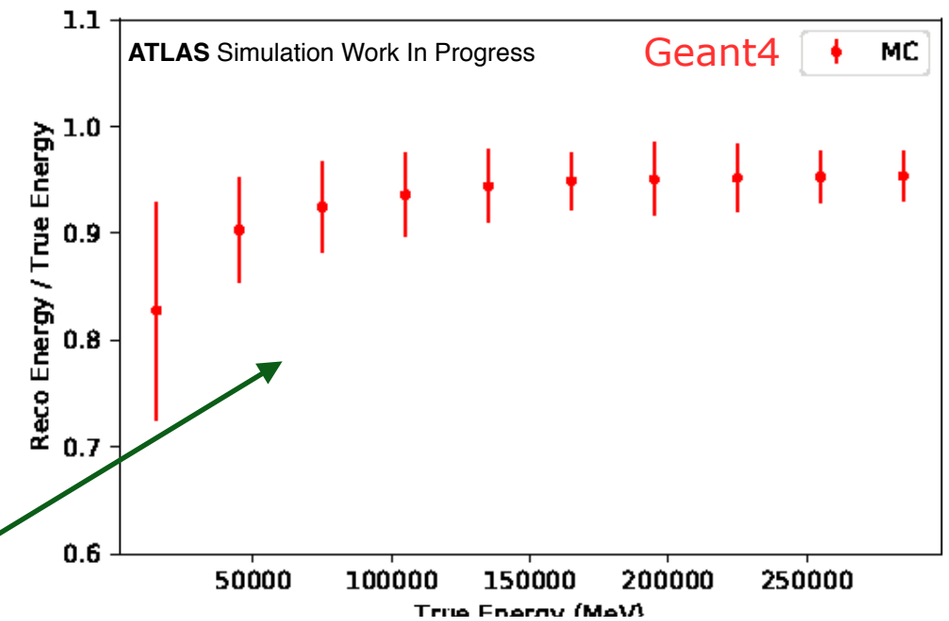
$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

Enforces normalisation of output $(\Sigma = 1)$

Energy resolution of Detector

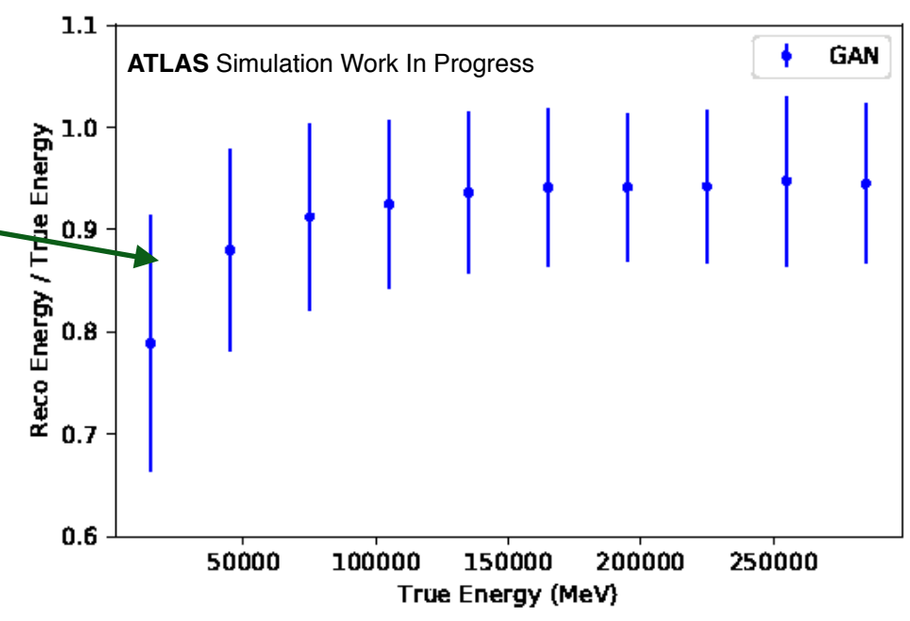
To reproduce the energy resolution of the detector:

- Condition on True Energy of Particle
- 'Normalise' cell energies to by true energy of particle, **(Σ(cell_energies) ≠ 1 anymore)**
- Change activation function in the final layer from softmax to relu



$\sigma_E \sim 10\% \sqrt{E}$

The error bars are the **standard deviation in each bin**



Softmax:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

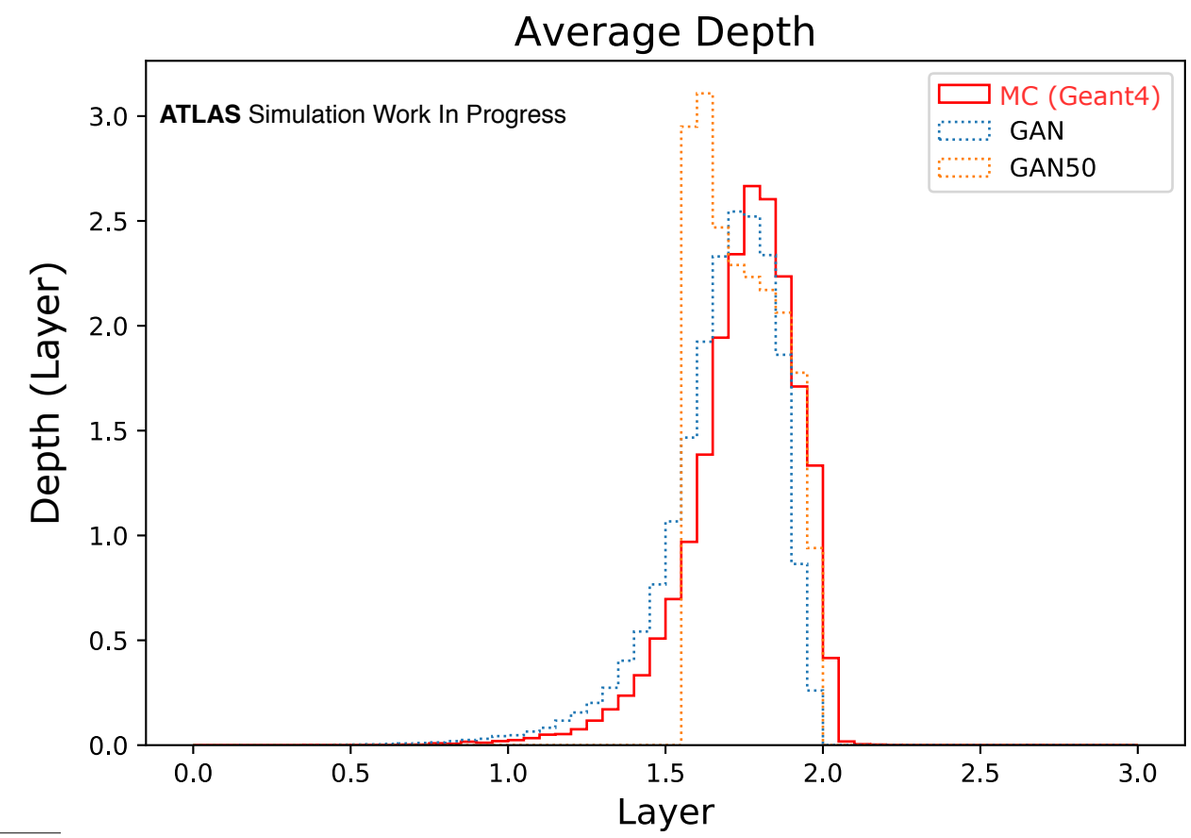
Enforces normalisation of output $(\Sigma = 1)$

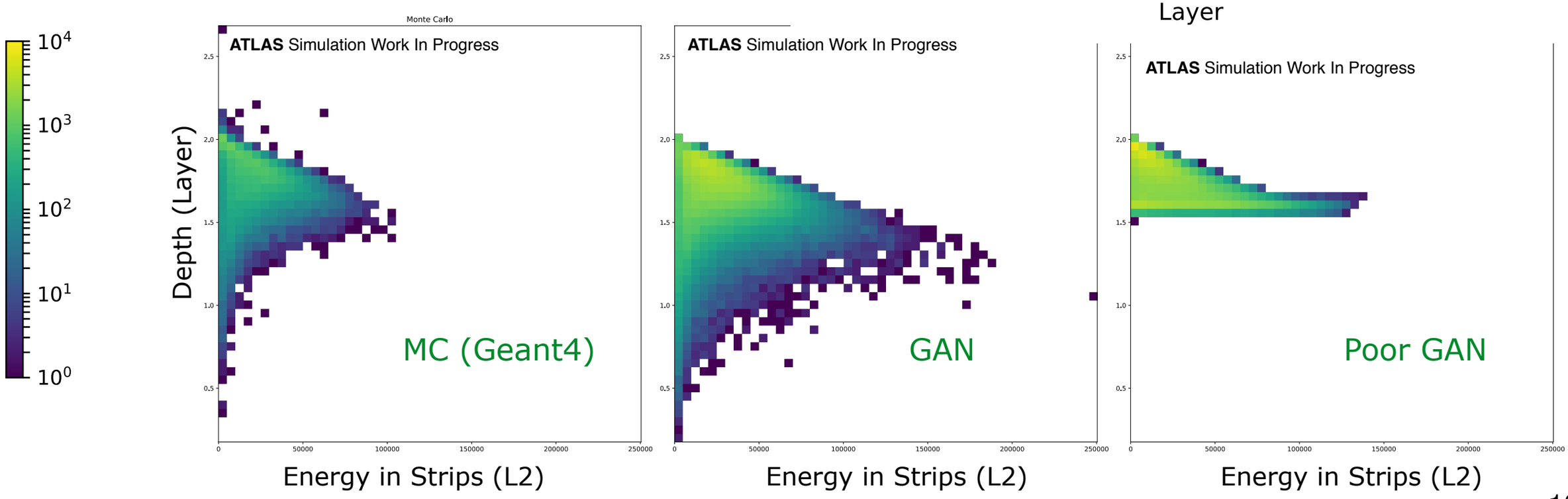
GAN reproduces the mean
cannot yet reproduce $\sigma_E \sim 10\% \sqrt{E}$

Investigation of Physics Variables: Depth

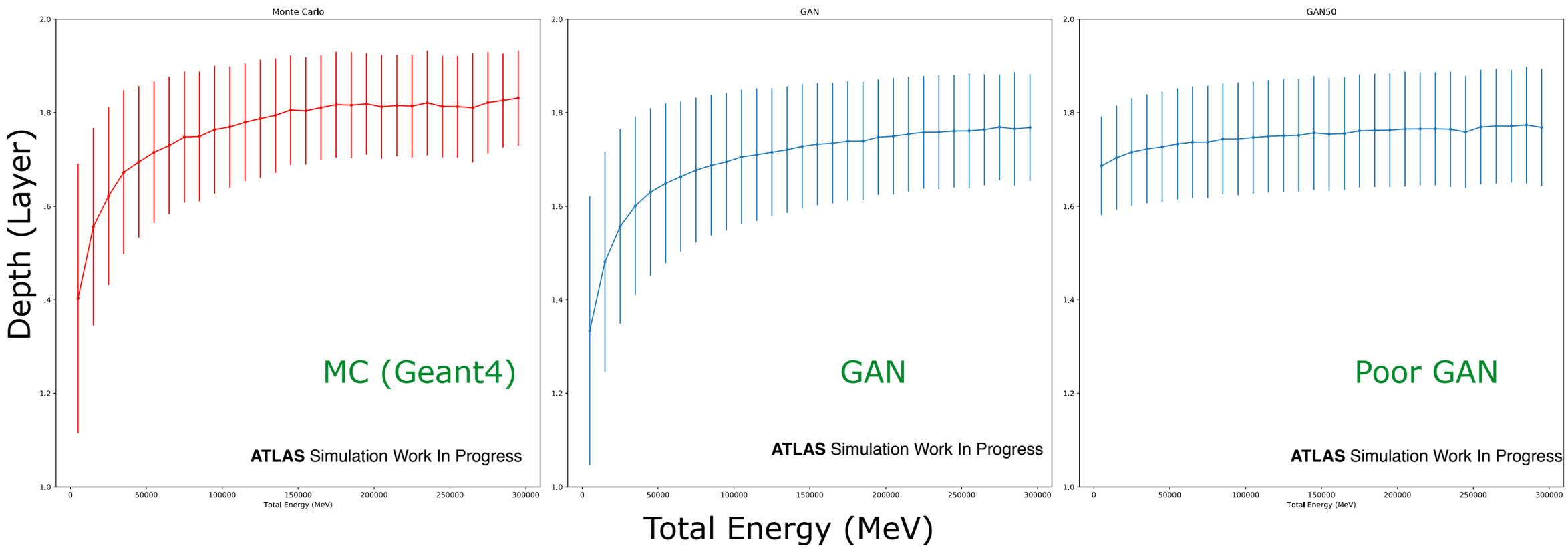
With our new Master 1 intern Jose-Antonio Mendez we are looking into new validation plots for depth

Comparing MC to GAN and ‘poor GAN’ trained for only 50 Epoch





Profile of Depth



The GAN **reproduces this non-trivial shape** of depth vs total energy very well, but is **shifted downwards**

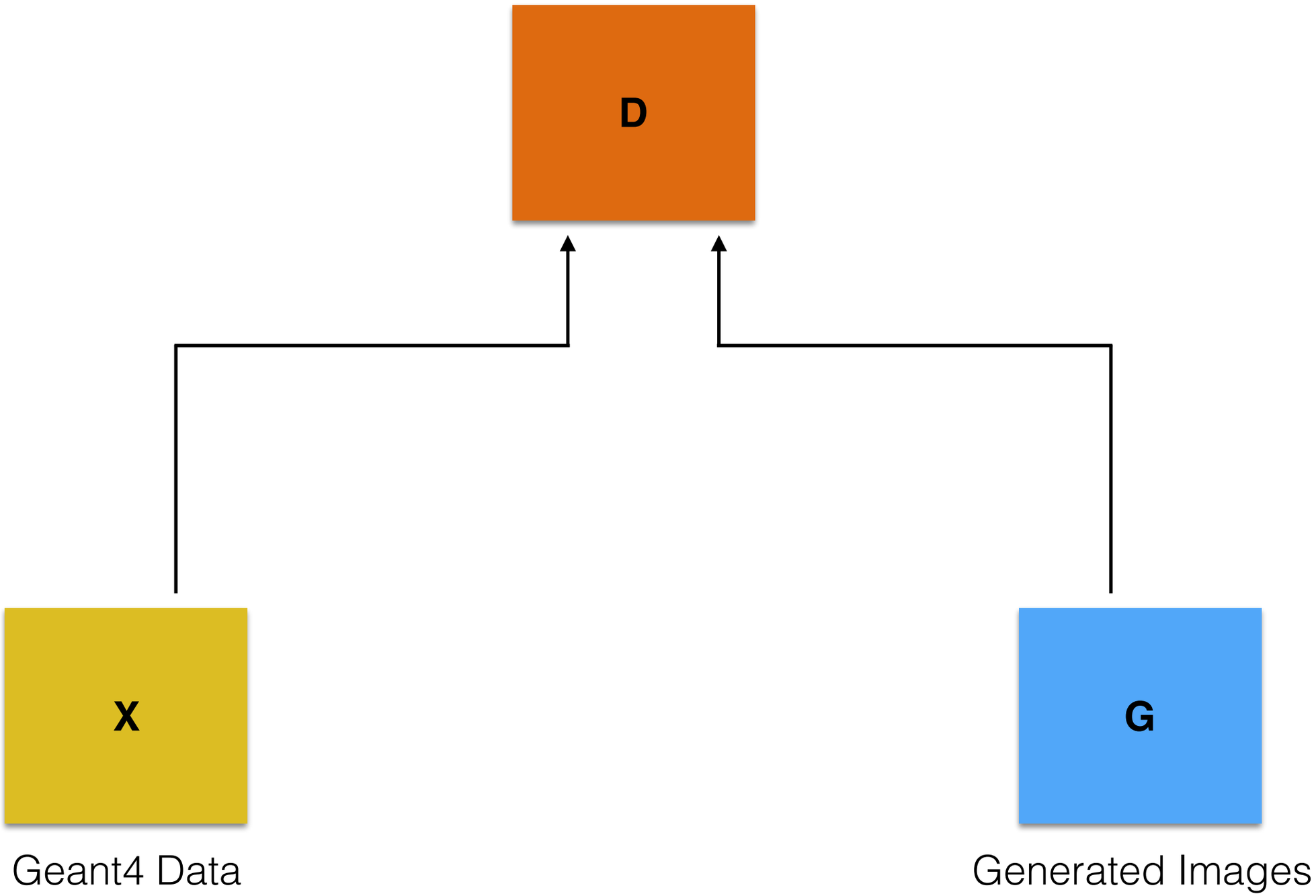
The error bars are the **Standard Deviation in each bin**

(Statistical Errors are negligible)

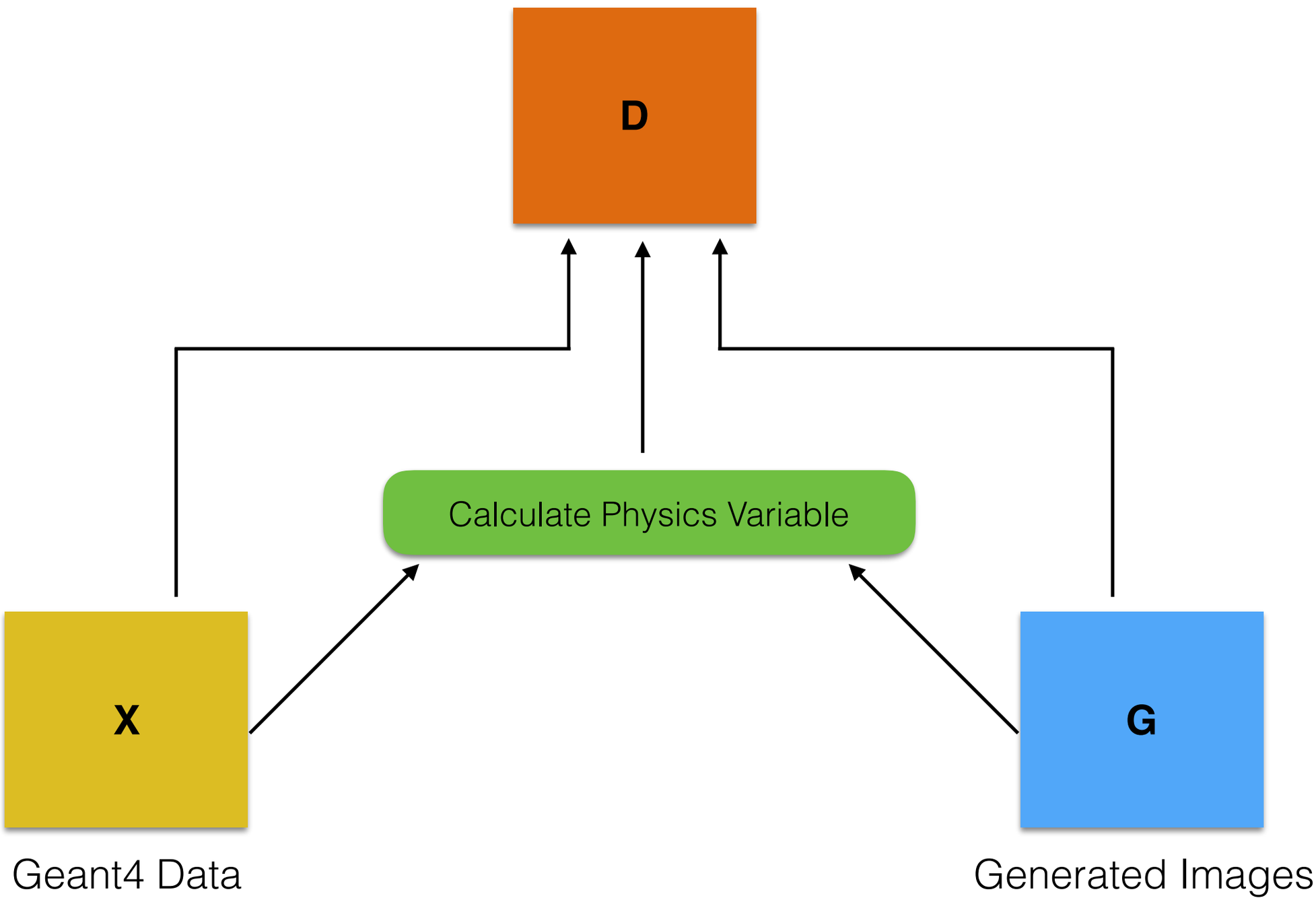
Next Plans

- Train GAN with knowledge about the alignment of calo layers (One-hot vector encoding) or use η/ϕ info
- Combine datasets to have more statistics, different Energy and η points
- Try new techniques:
    - Improve discriminator performance
    - Training Tricks: Train on log(cell_E), learning rate of G, Gradient penalty, non-RELU activations
    - Add physics variables in GAN training (see slide)
- Try to train on more 'High Resolution', high granularity images to avoid the complications of calorimeter geometry
    - Harder to train on high resolution, but accuracy only needed at less granular 'cell level'
    - Use Progressive GAN with High Resolution images, other GAN flavours

Add Physics Variables in Training



Add Physics Variables in Training

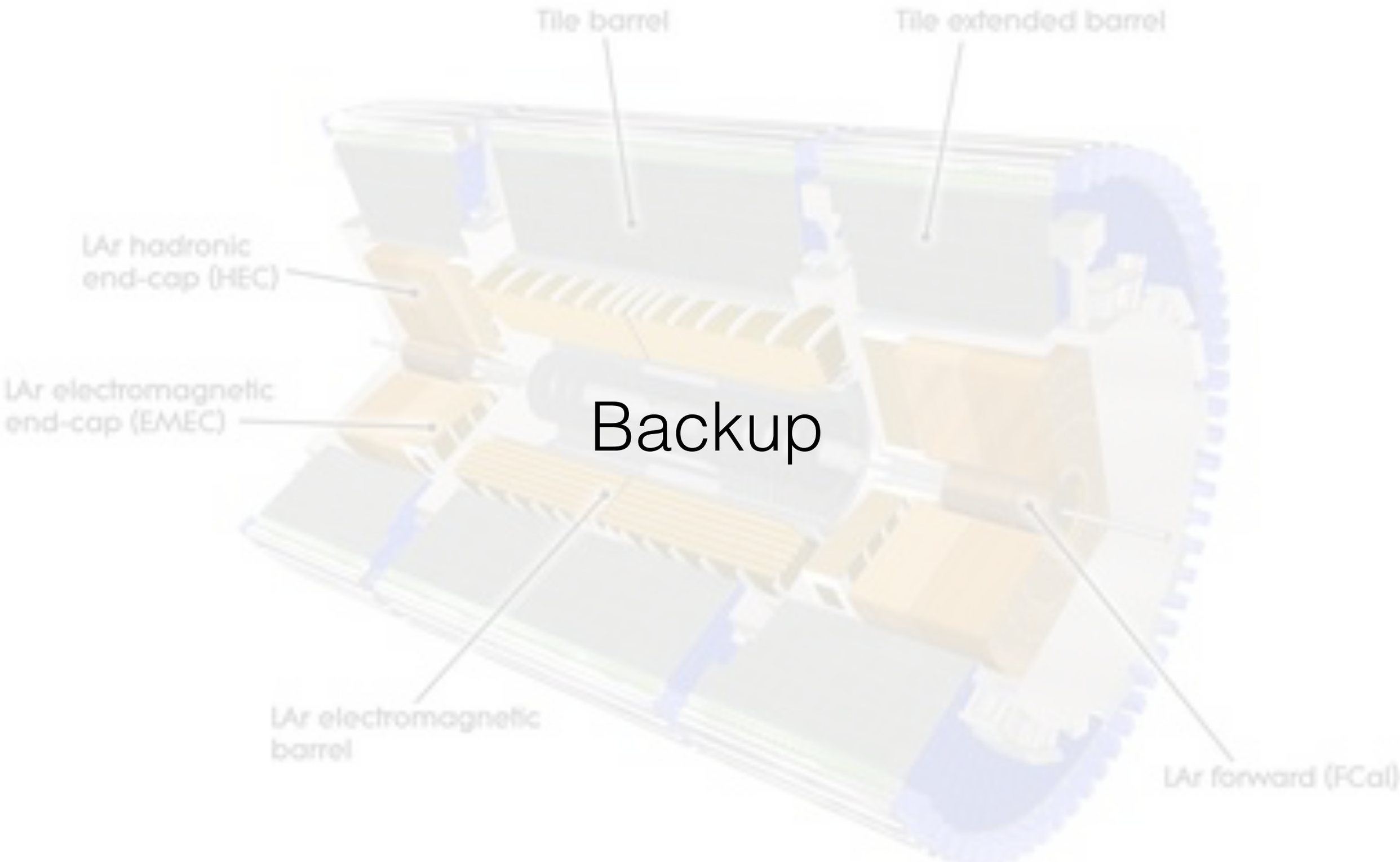


Help the discriminator see physics

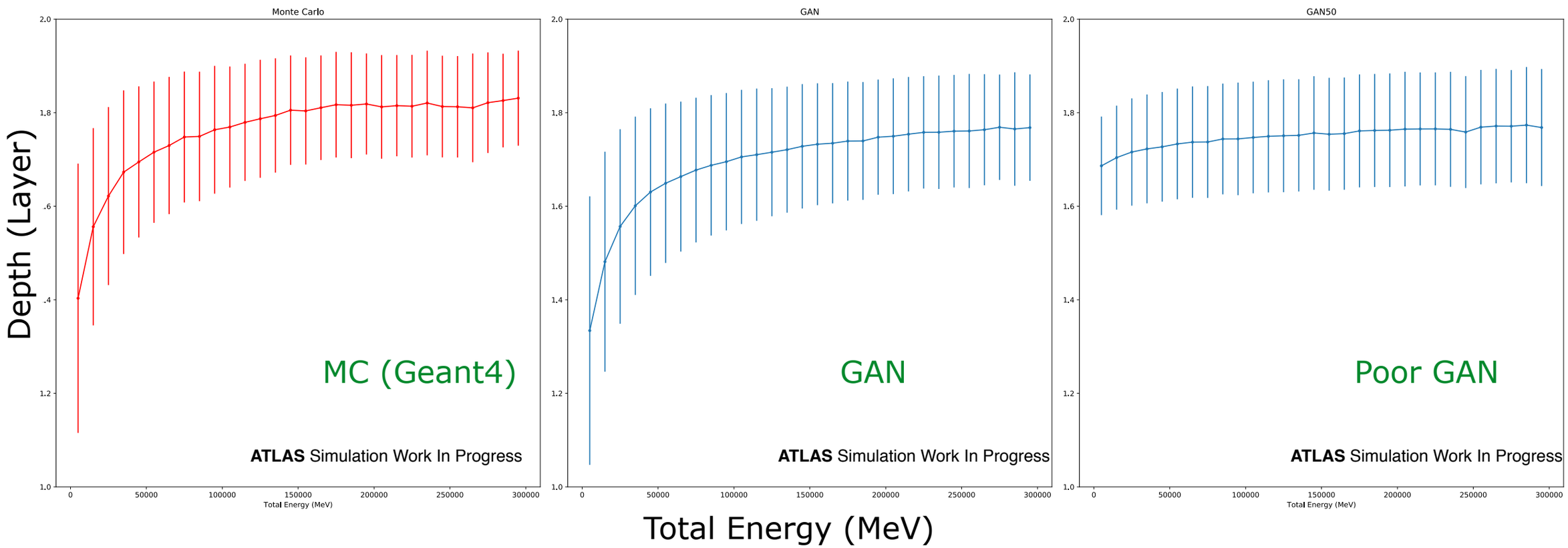
Conclusion

- Understanding the data structure remains essential for Machine Learning tasks
- Motivation to move to more granular, ‘High Resolution’ images
- Comparisons need to be made with FastCaloSim, VAE, which are other methods of fast simulation of EM showers
- GPU gives 2x speed up compared to CPU-only training
    - Might improve for larger Neural Nets, more high definition images

Backup

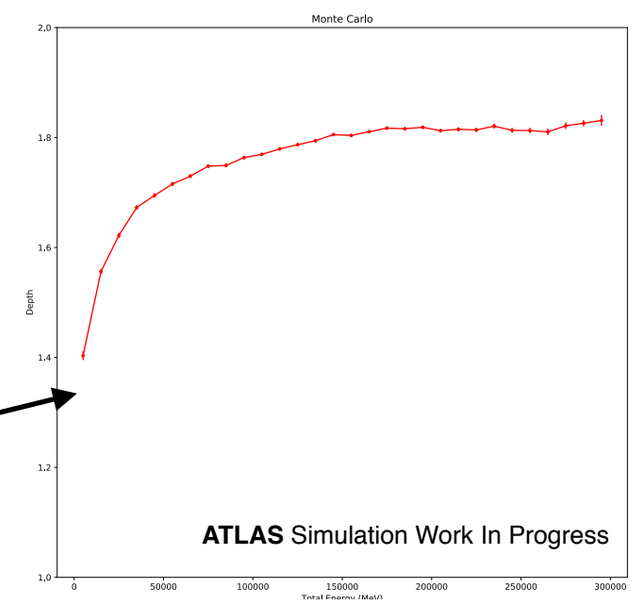

Tile barrel
Tile extended barrel
LAr hadronic end-cap (HEC)
LAr electromagnetic end-cap (EMEC)
LAr electromagnetic barrel
LAr forward (FCal)

Profile of Depth



The GAN **reproduces this non-trivial shape** of depth vs total energy very well, but is **shifted downwards**

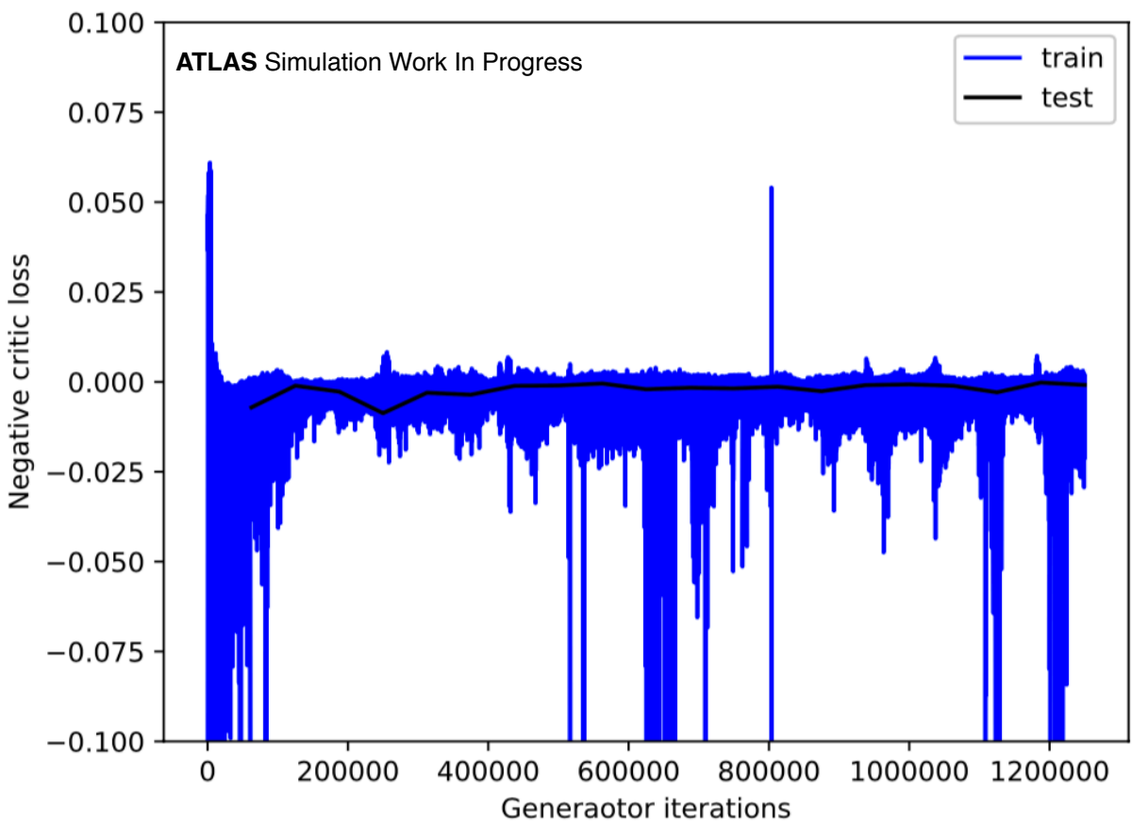
The error bars are the **Standard Deviation in each bin**



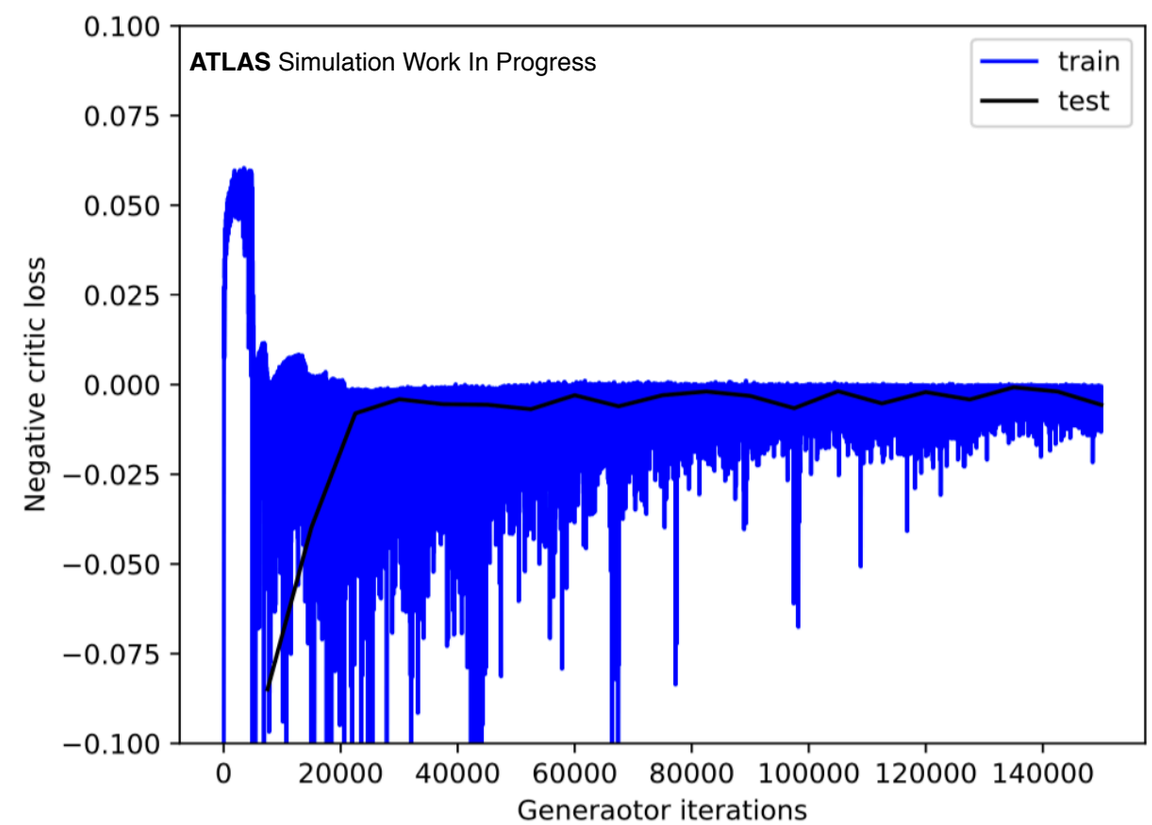
Statistical Errors are negligible

Mixed Config: Critic Loss

10k Events



Only 10k/8 Events



‘Configurations’

- The calo layers are one behind the other : PreSampler, Strip, Middle, Back. But they are not aligned in the same way for every event. Since they have different widths, the exact same EM shower would look different at different parts of the calorimeter.

- If we select a 7x7 cluster in the middle layer, then in front of it, the strip could have 4 different ‘configurations’. See. For a fixed middle 7x7, there could be 4 different ways the strip cells are aligned w.r.t the middle cells. In config 0 and 7, 2 cells of the strip can cover all the 7 cells of the middle layer in phi (y-axis). Whereas for the other 2 configurations 3 cells are required.

- A similar case for the back layer’s alignment w.r.t. the middle layer means that there are 2 distinct ways the back might be aligned.

- 4 (strips in phi) * 2 (back in eta) = 8 different arrangements of the cells in the calorimeter.

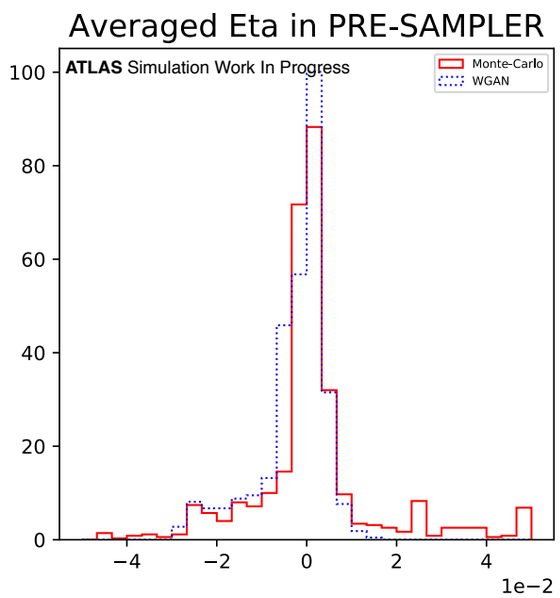
- The Pre-Sampler behaves the same as Strip in phi, and behaves the same as Middle in eta. The Back behaves the same as Middle in phi. 8 strip cells have the same width as 1 Middle cell in eta, so there is no problem of misalignment.

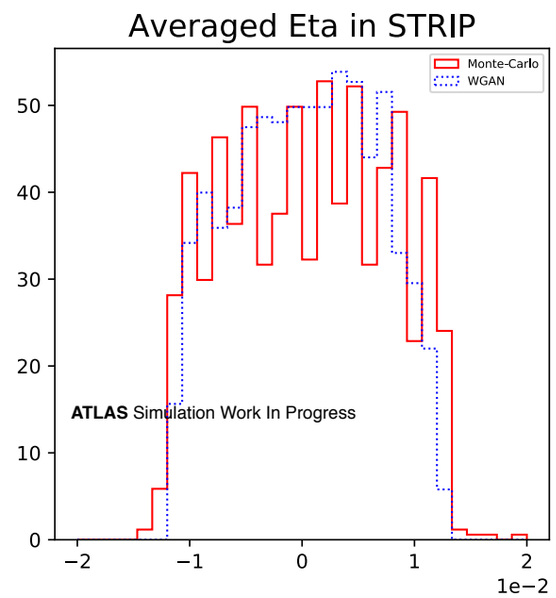
- Configurations 0,1,2,3 are the same as 4,5,6,7 respectively for Strips/Pre-Sampler. All configurations in the set {0,1,2,3} are equivalent for the Back, all configurations in the set {4,5,6,7} are equivalent, but configurations from different sets are not equivalent to each other for the Back (e.g. config 0 not equivalent to config 4).

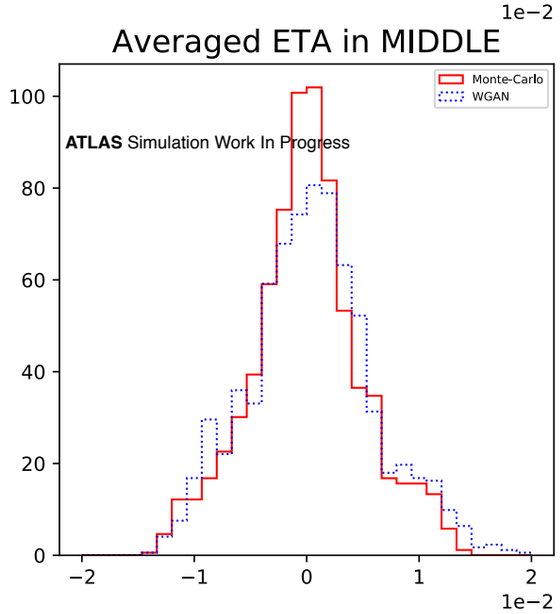
- The configurations in the Strip are periodic every 4 cells in phi, and the back are periodic every 2 cells in eta in the EM calorimeter.

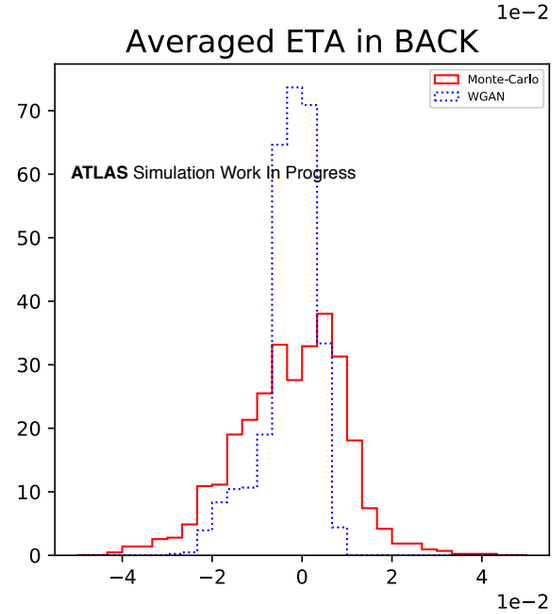
Eta/Phi for Config 0

Average Eta

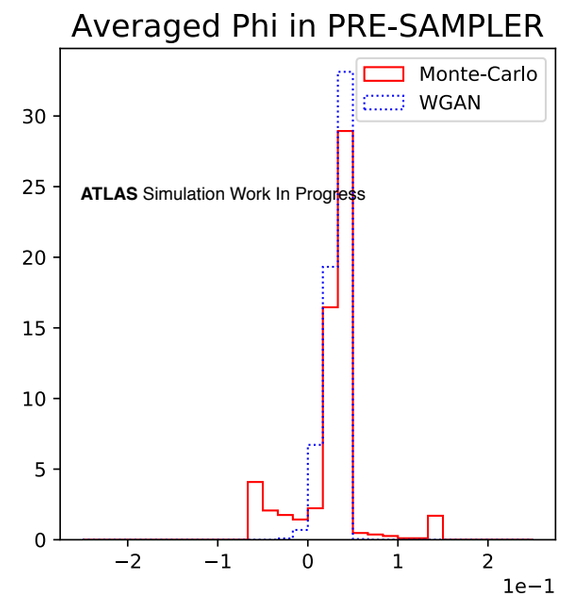


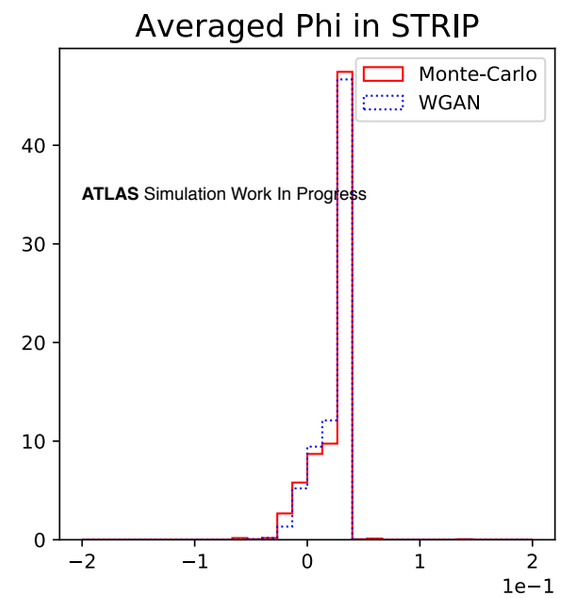


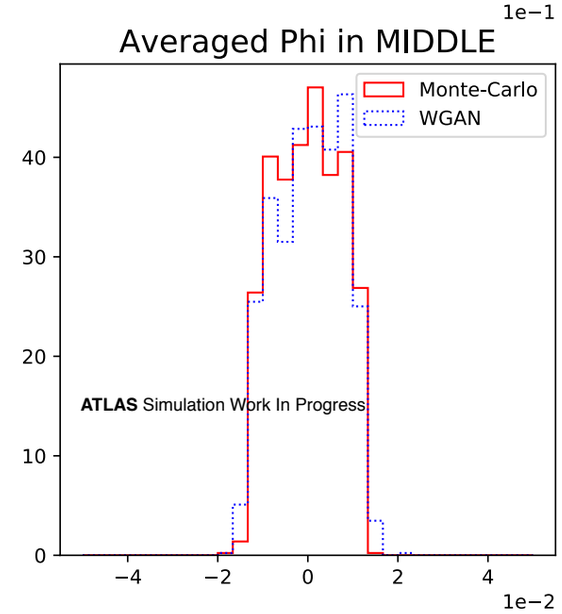


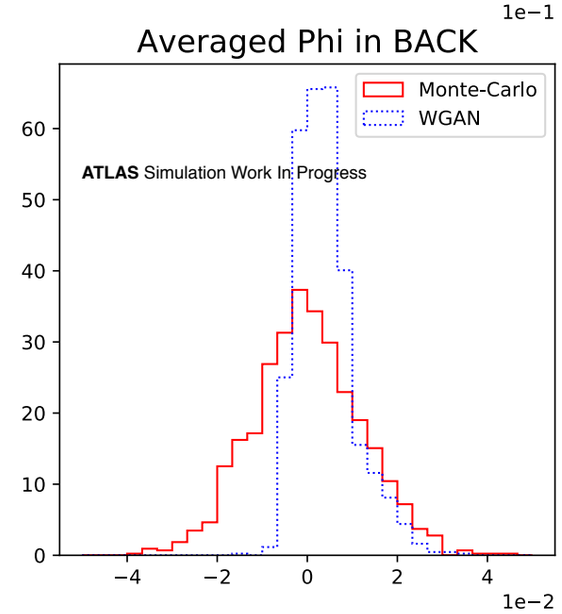


Average Phi



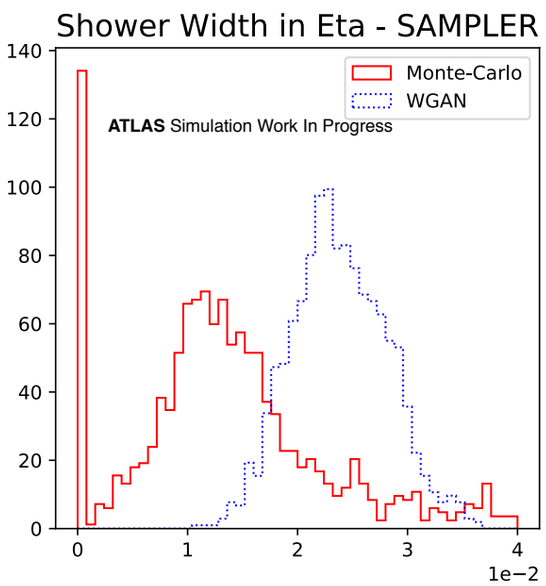


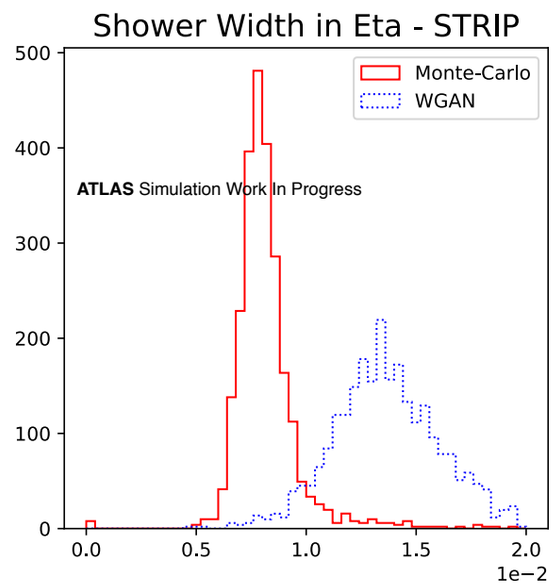


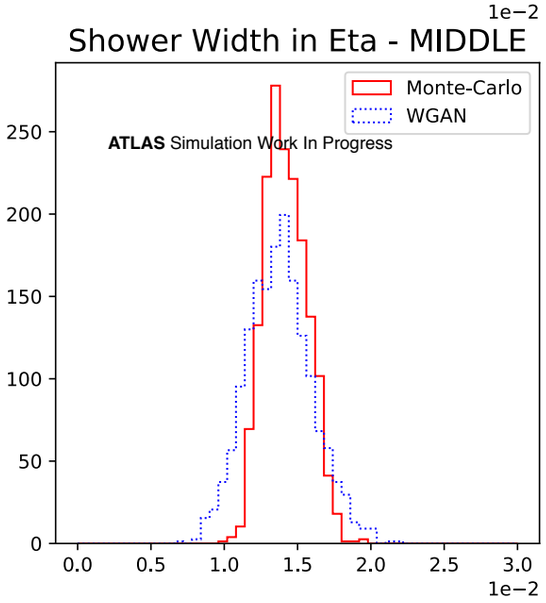


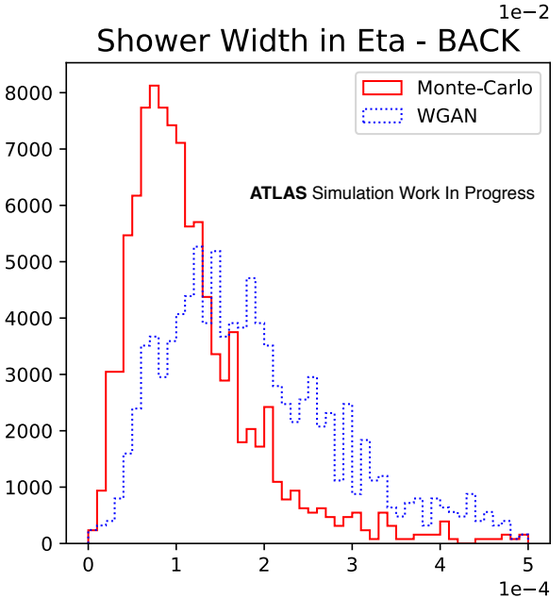
Width for Config 0

Eta Width

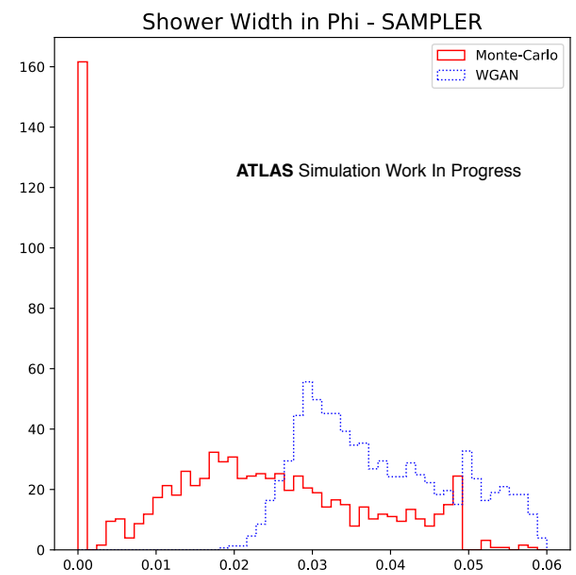


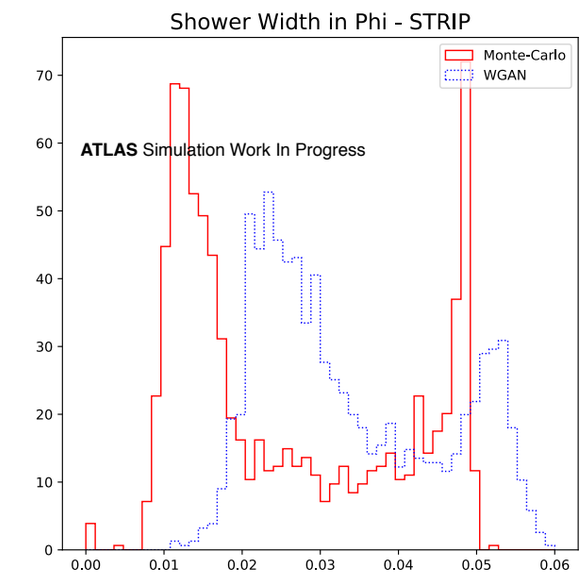


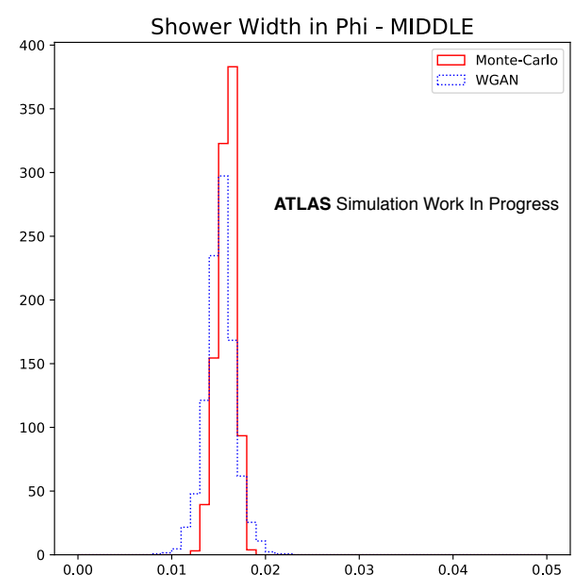


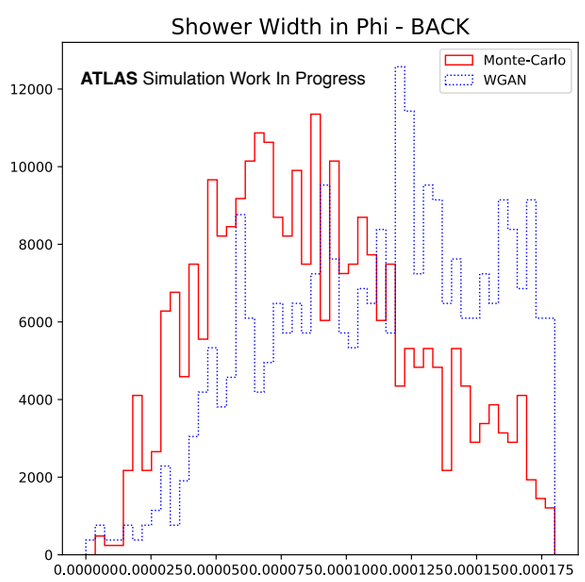


Phi Width

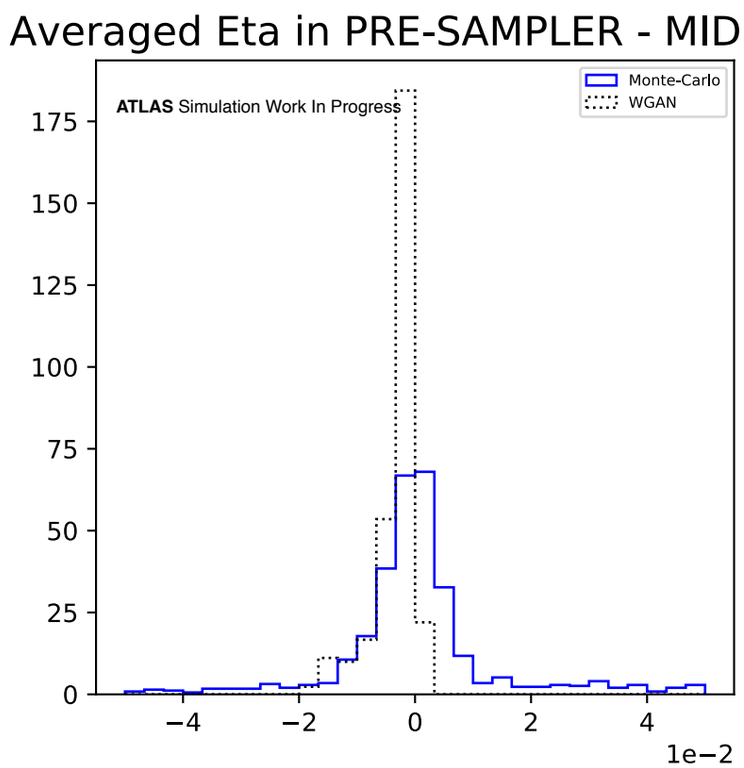


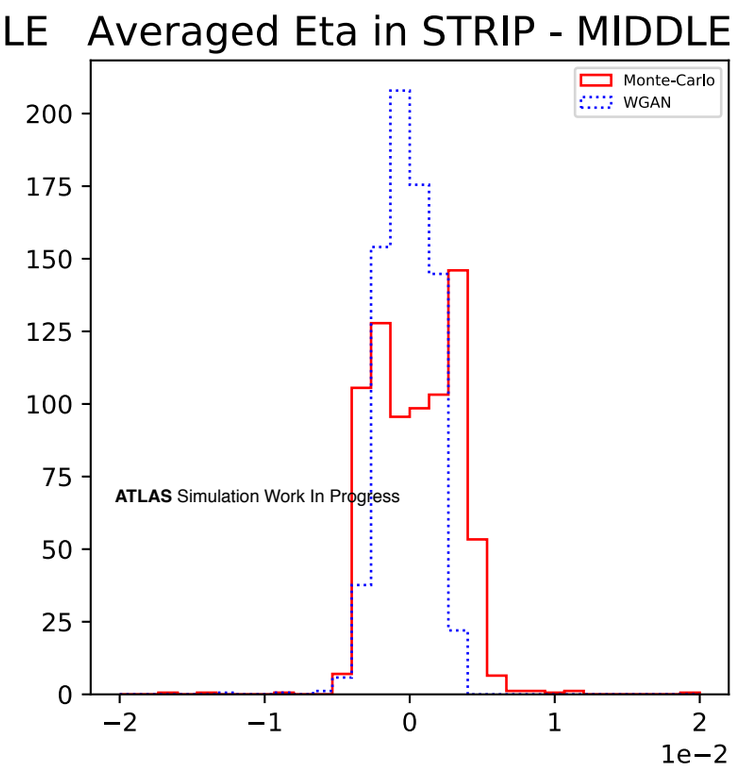


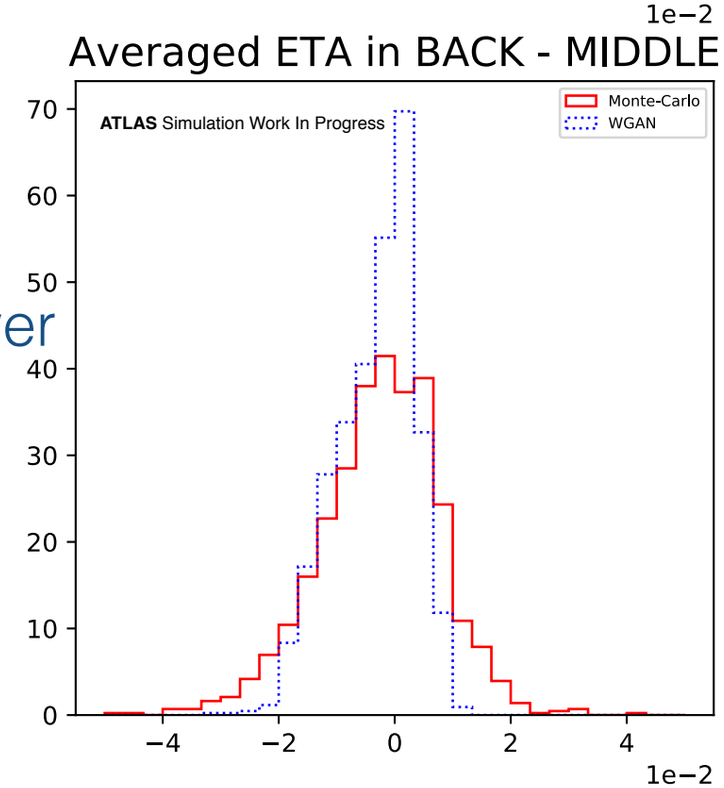




Layer-to-Layer Correlation (Config 0)





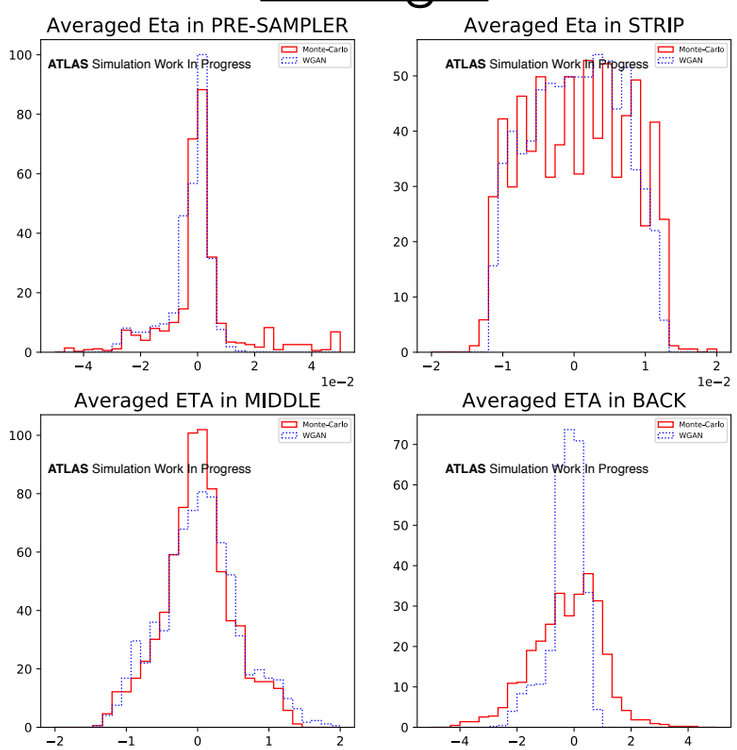


Compute:
Average Eta in Layer - Average Eta in Middle Layer

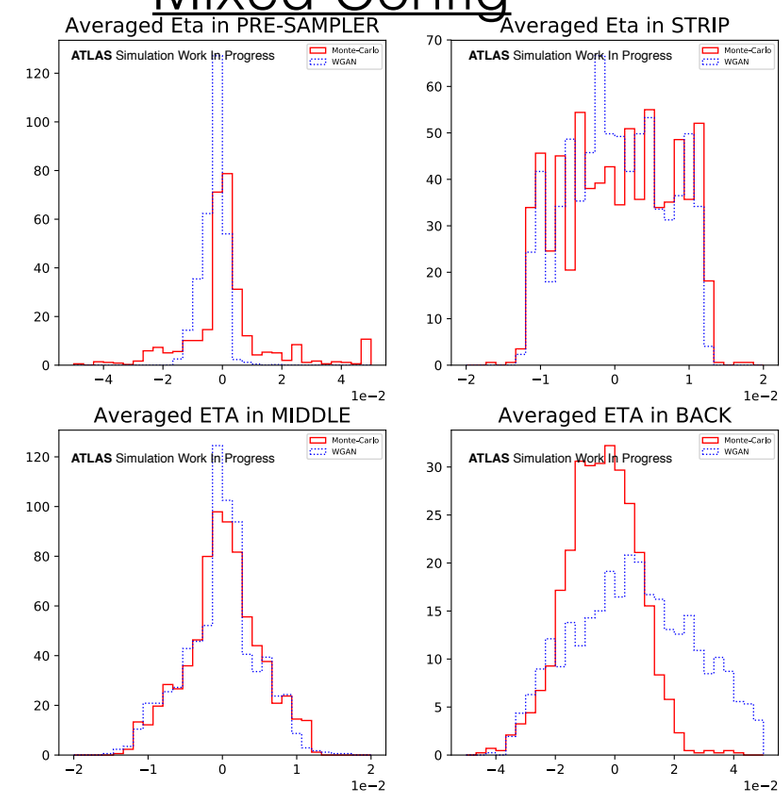
Compare with Mix of 8 Configs (same number of events)

Eta

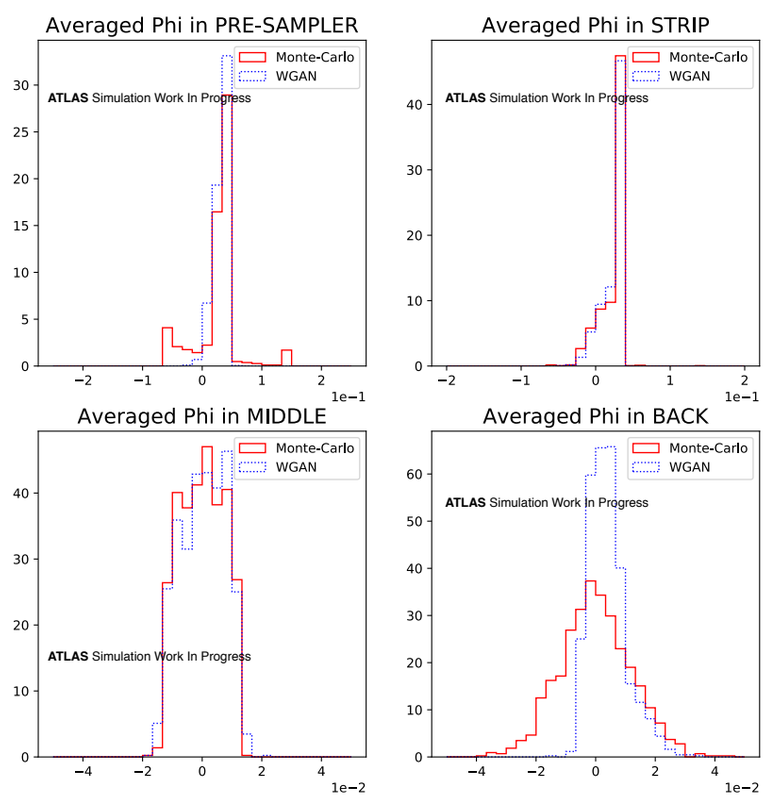
Config 0

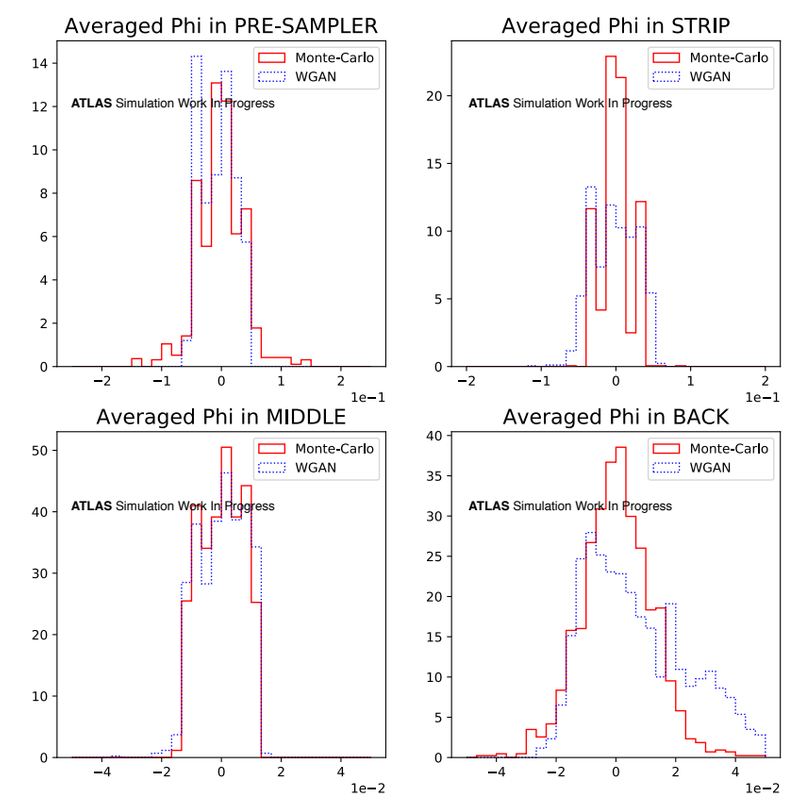


Mixed Config



Phi



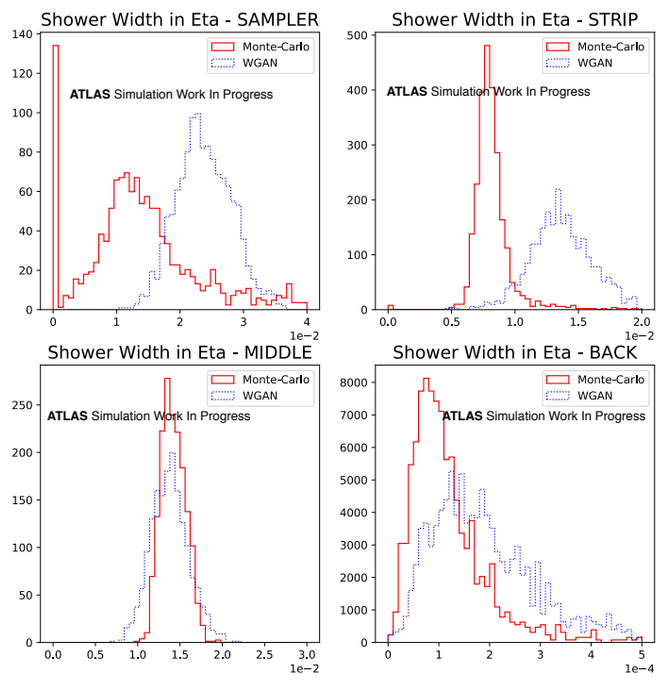


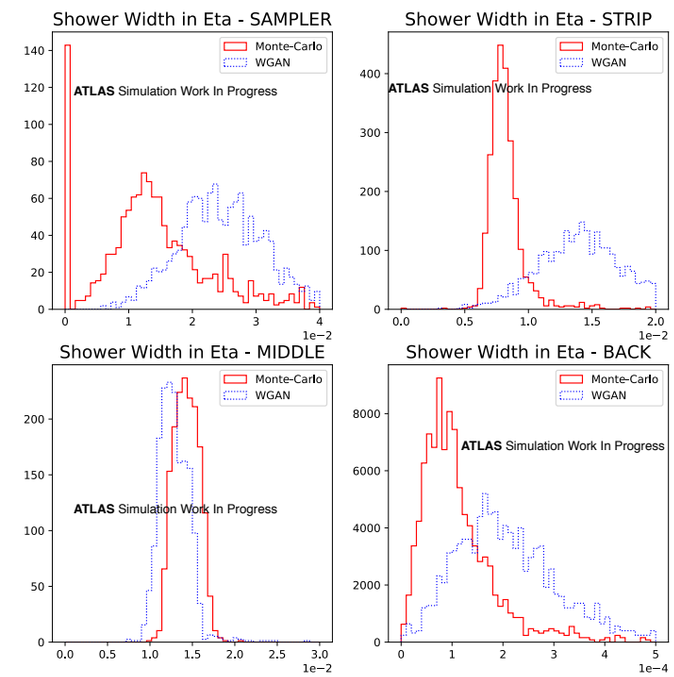
Compare with Mixed Configs (same number of events)

Config 0

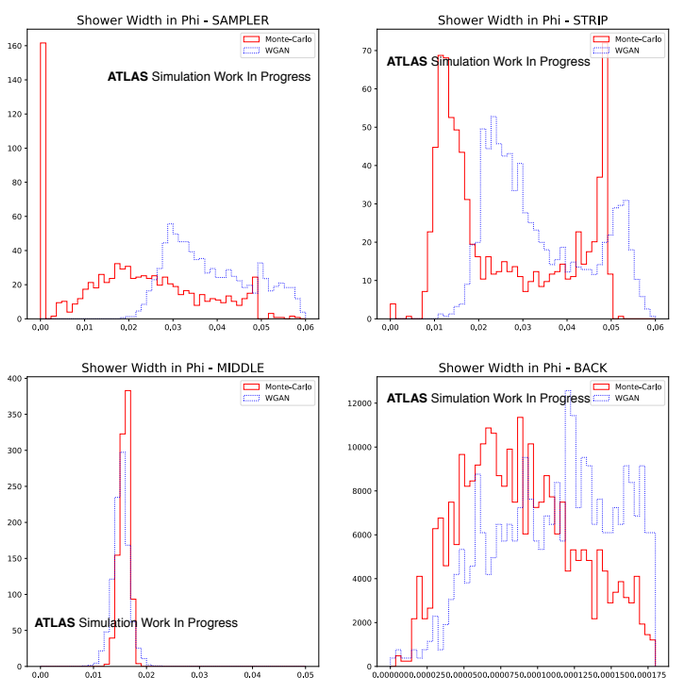
Mixed Config

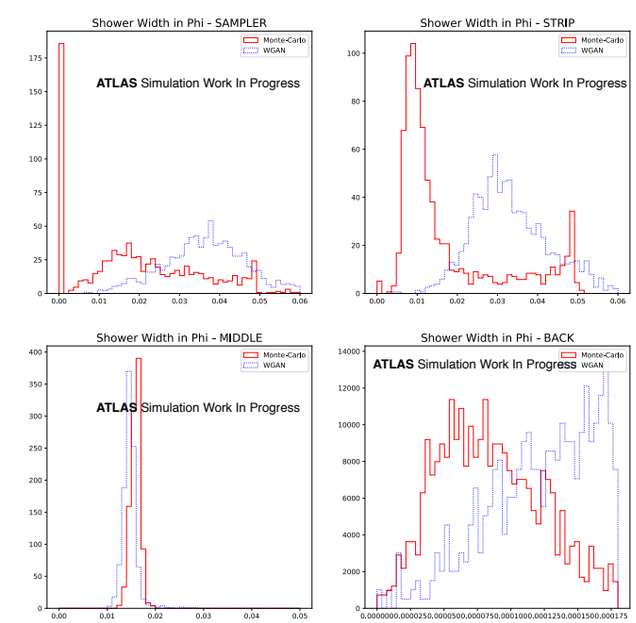
Eta Width





Phi Width

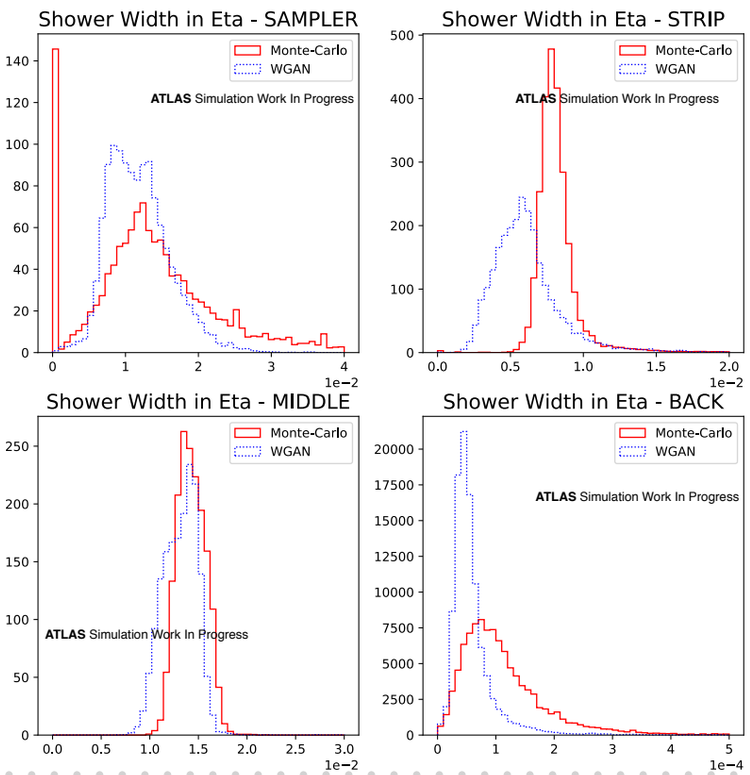




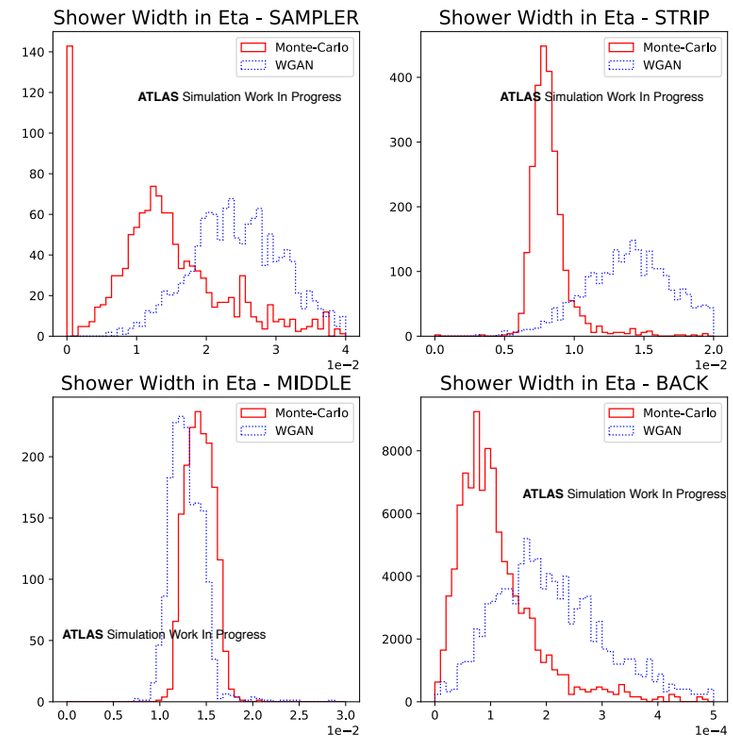
Mixed Config: All 10k Events vs 10k/8 Events

Eta Width

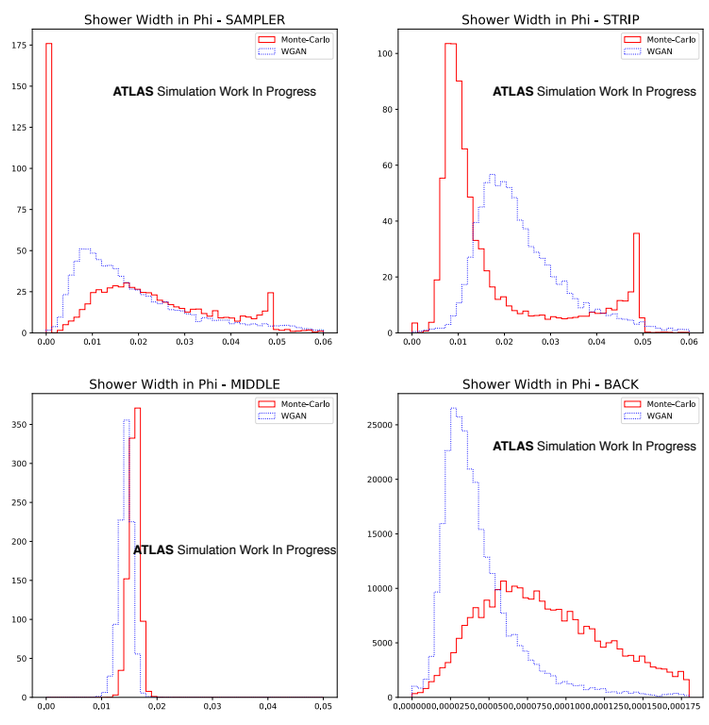
10k Events

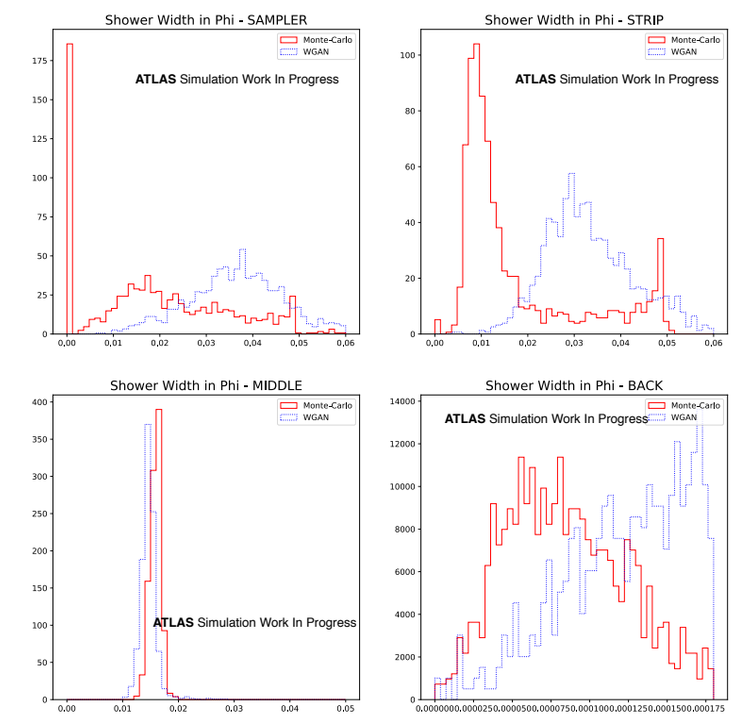


Only 10k/8 Events



More statistics helps GAN learn widths (but no other distributions)



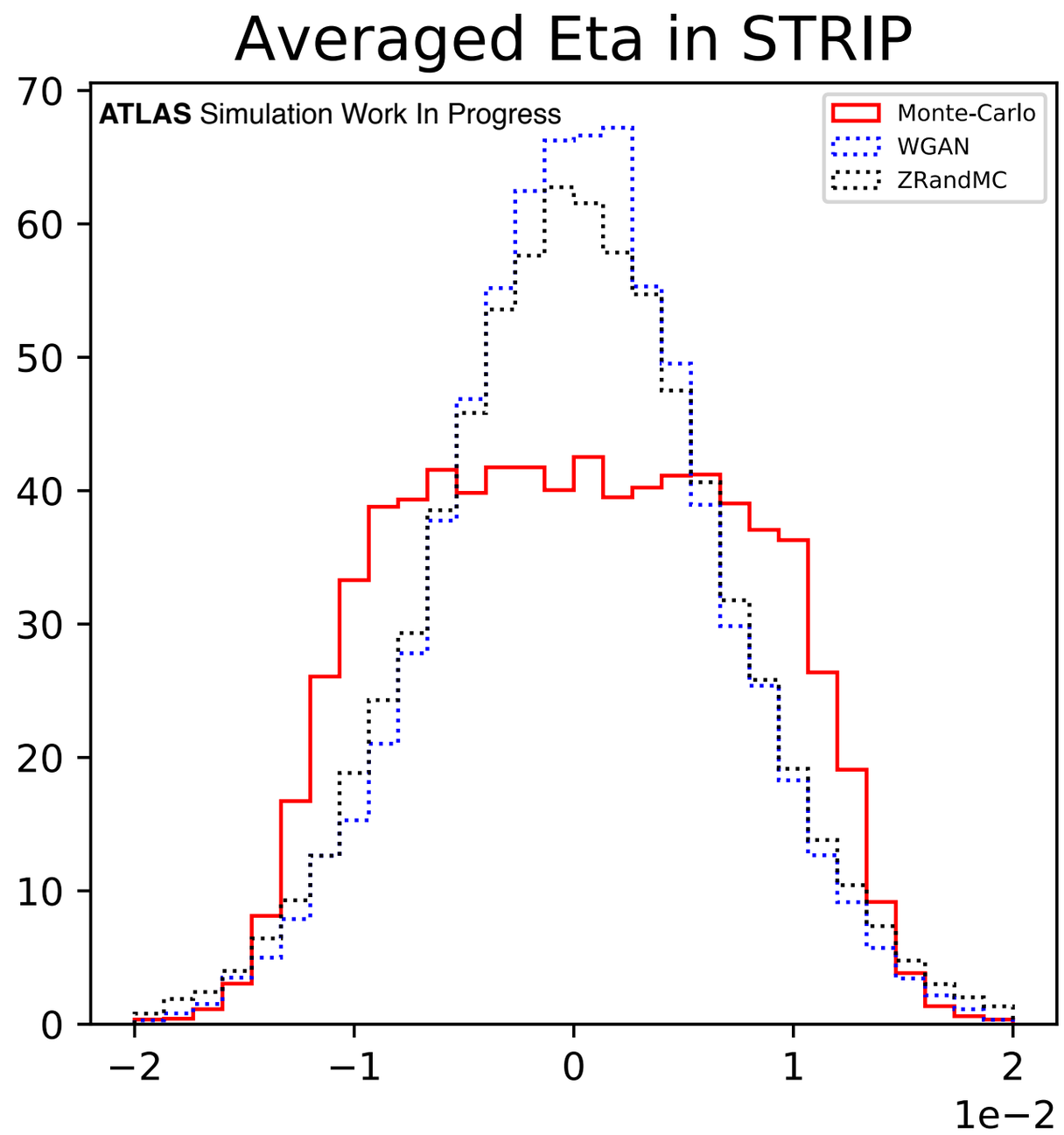


Unfair comparison for GAN in xAOD dataset

	Red	Blue	**Black**
Cell Energy (E)	MC	GAN	MC
Cell η	From same event	From an MC event	From different MC event
E-η Correspondence	Correct	N/A	Randomised

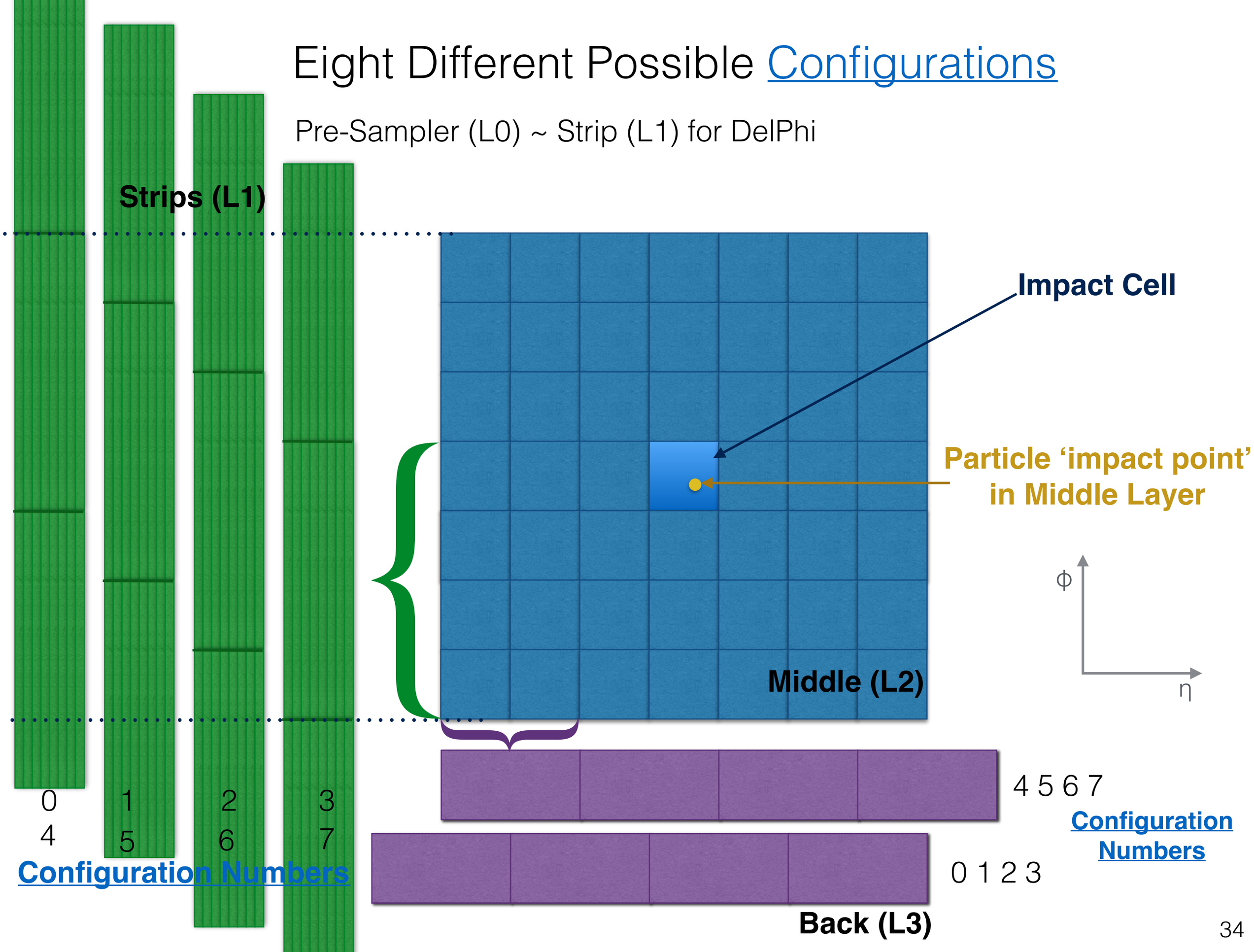
The previous analysis (see 1, 2) did not account for this inherent advantage of MC over GAN

GAN only knows Energy, no η/ϕ info. For validation, η/ϕ came in from MC data in previous analysis

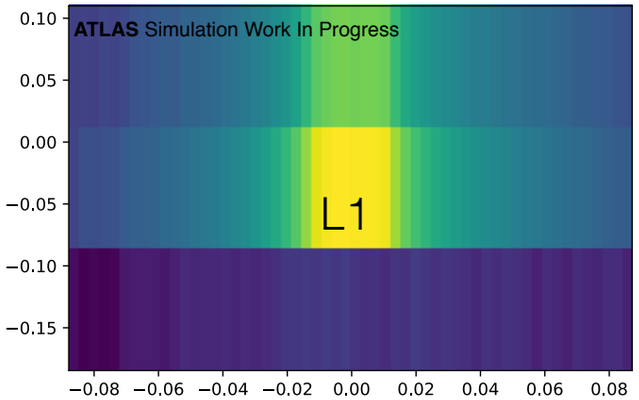


Eight Different Possible Configurations

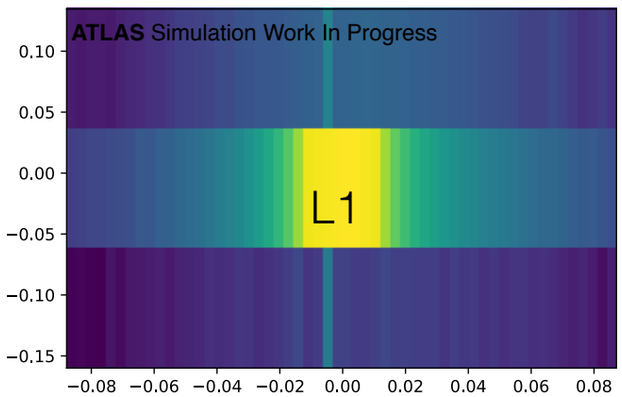
Pre-Sampler (L0) ~ Strip (L1) for DelPhi



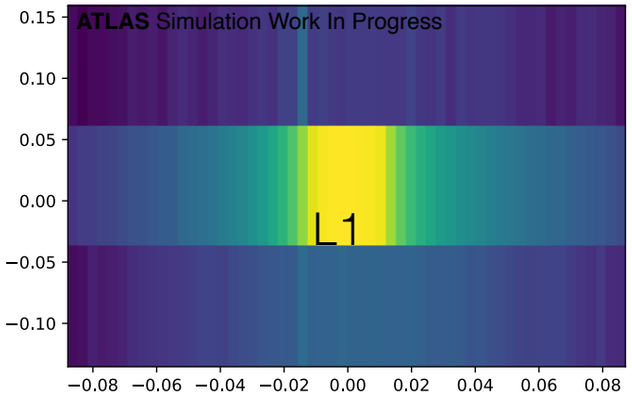
ATLAS Simulation



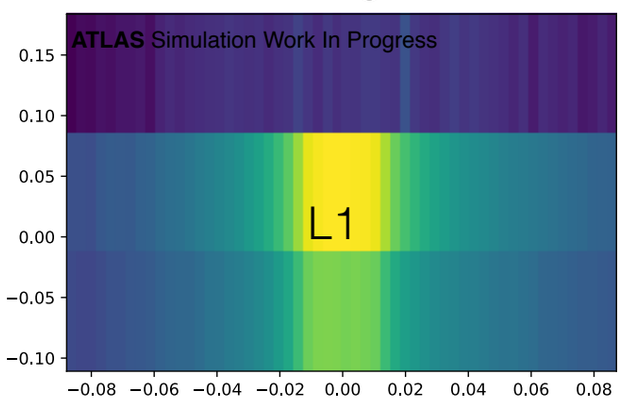
Config 7



Config 6

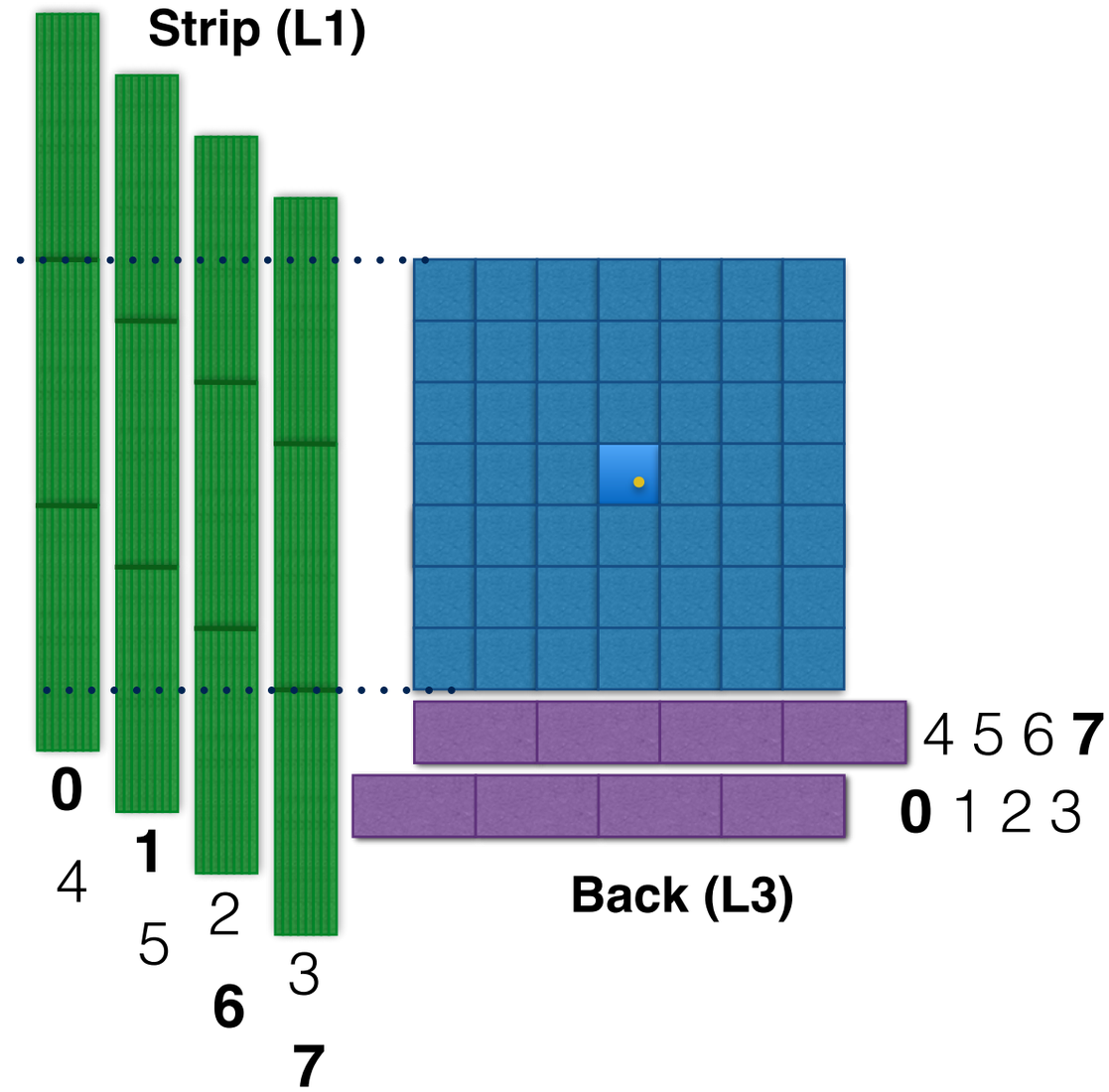


Config 1



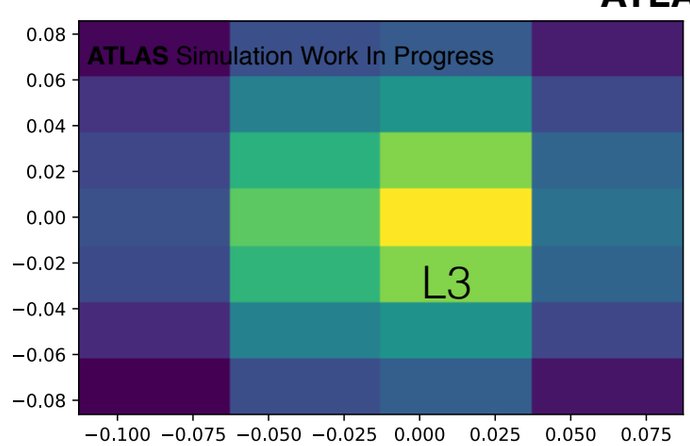
Config 0

Strip (L1)

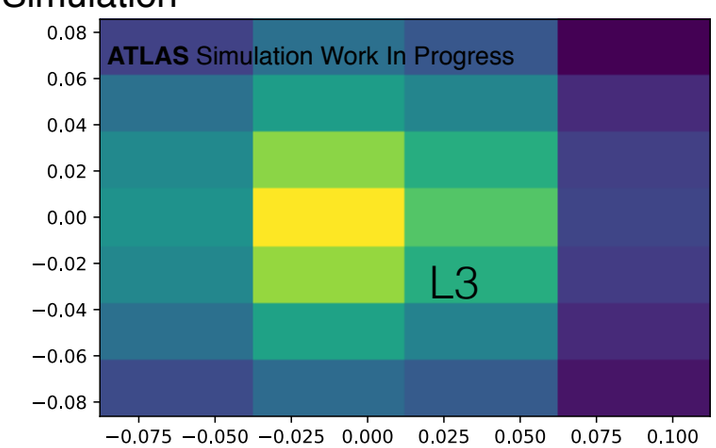


Back (L3)

ATLAS Simulation



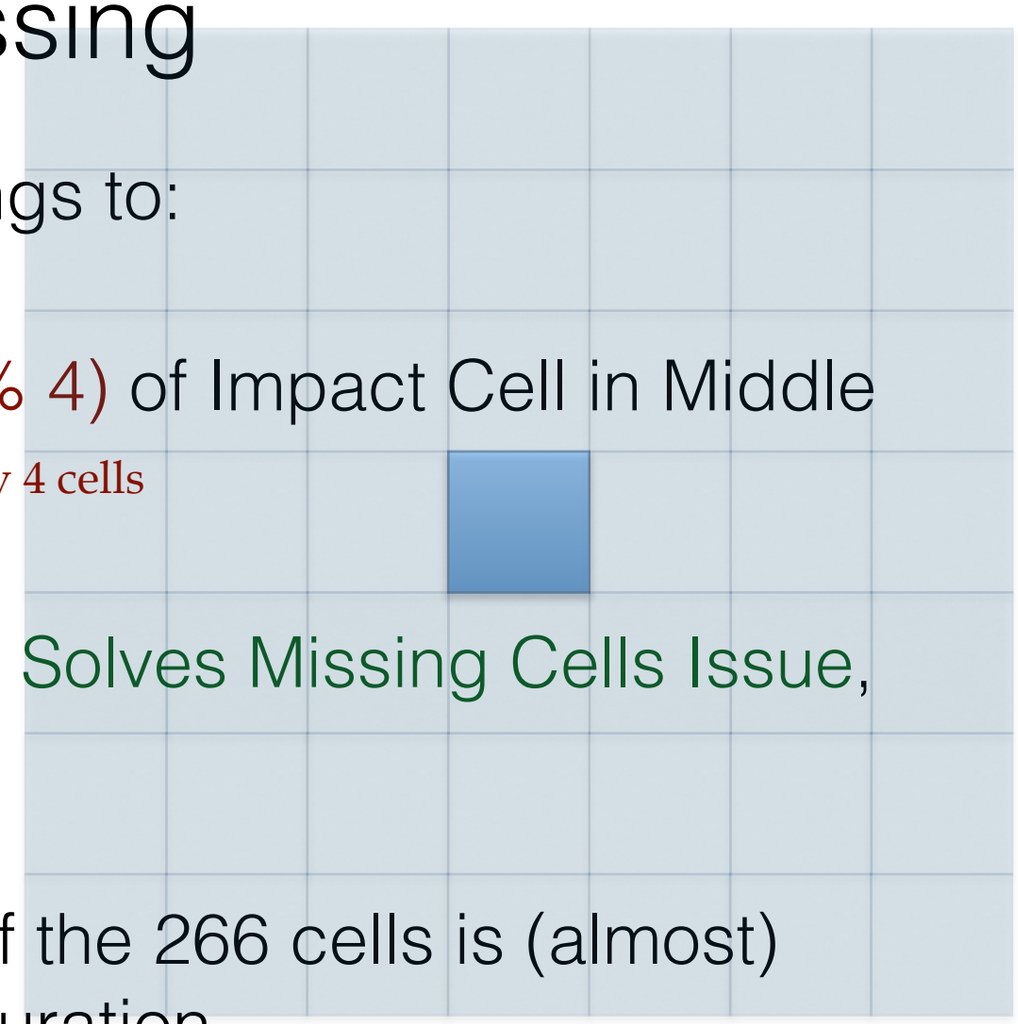
Config 0



Config 7

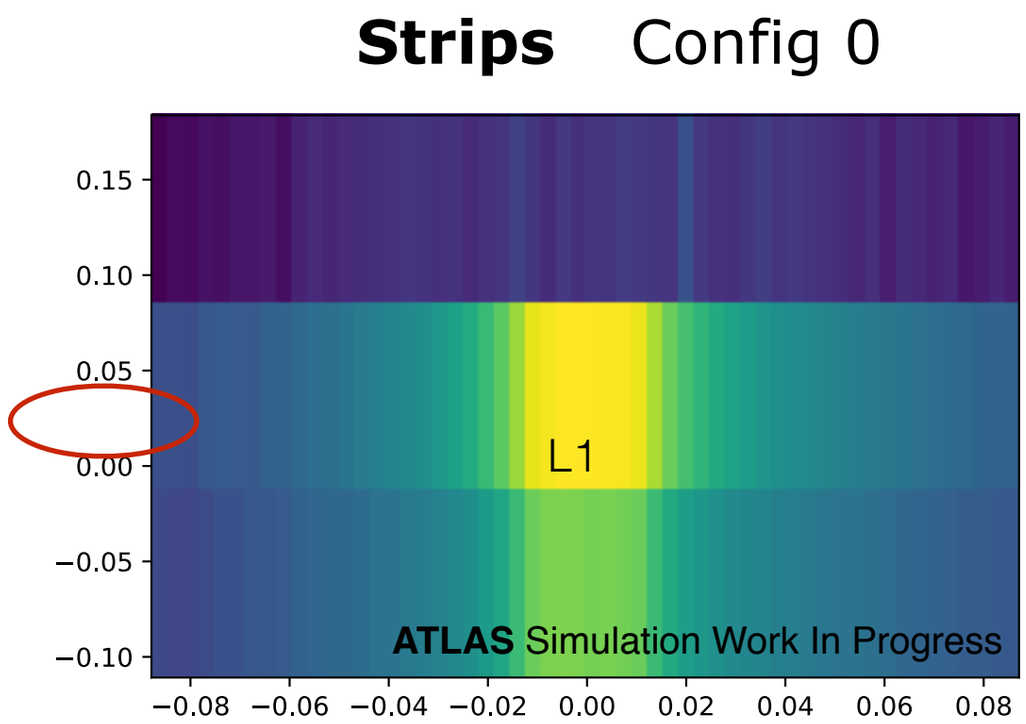
So How Can We Train On These Images? : Pre-Processing

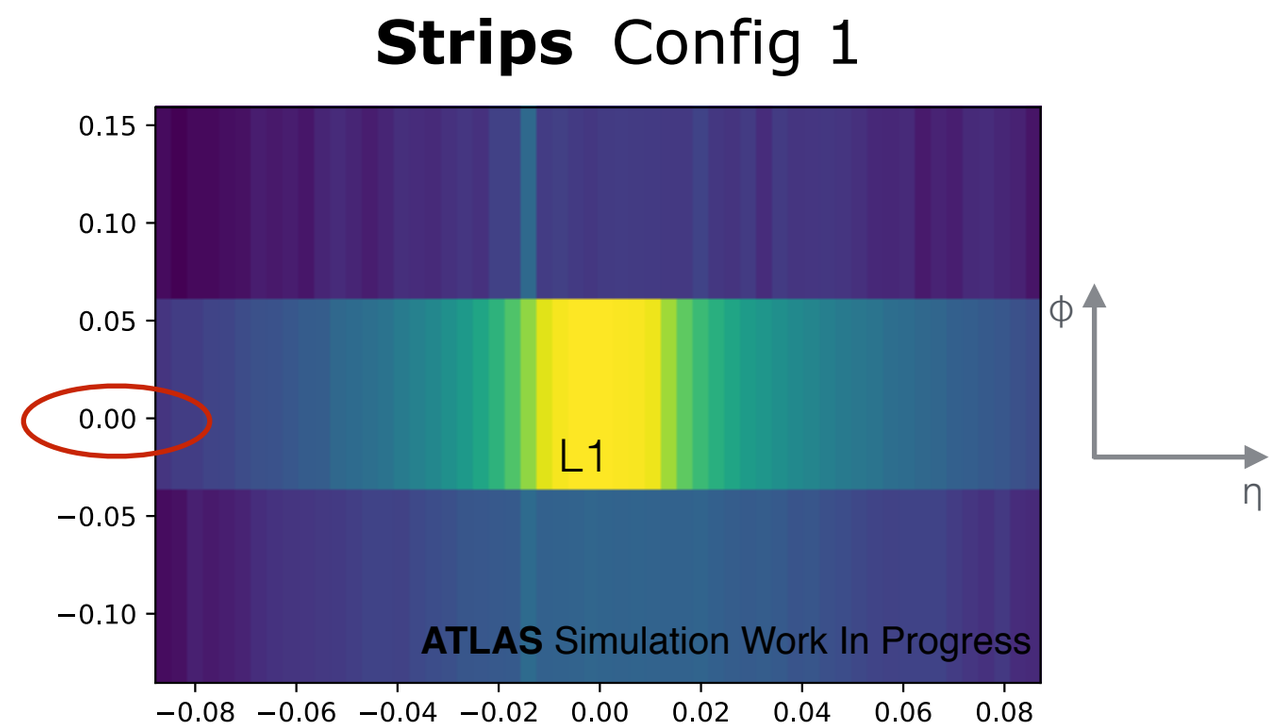
- Identify which config each event belongs to:
  - Back periodic every 2 cells
  - Using (eta_index % 2), (phi_index % 4) of Impact Cell in Middle Layer
  - Strip periodic every 4 cells
  - Create 8 grids of η, ϕ for 8 configs : Solves Missing Cells Issue, brings back η=0, ϕ=0
  - Verify that DelEta, DelPhi for each of the 266 cells is (almost) identical for all events of one configuration
- **Train separately** for each of the 8 configuration (but low stats: 10k/8 events)

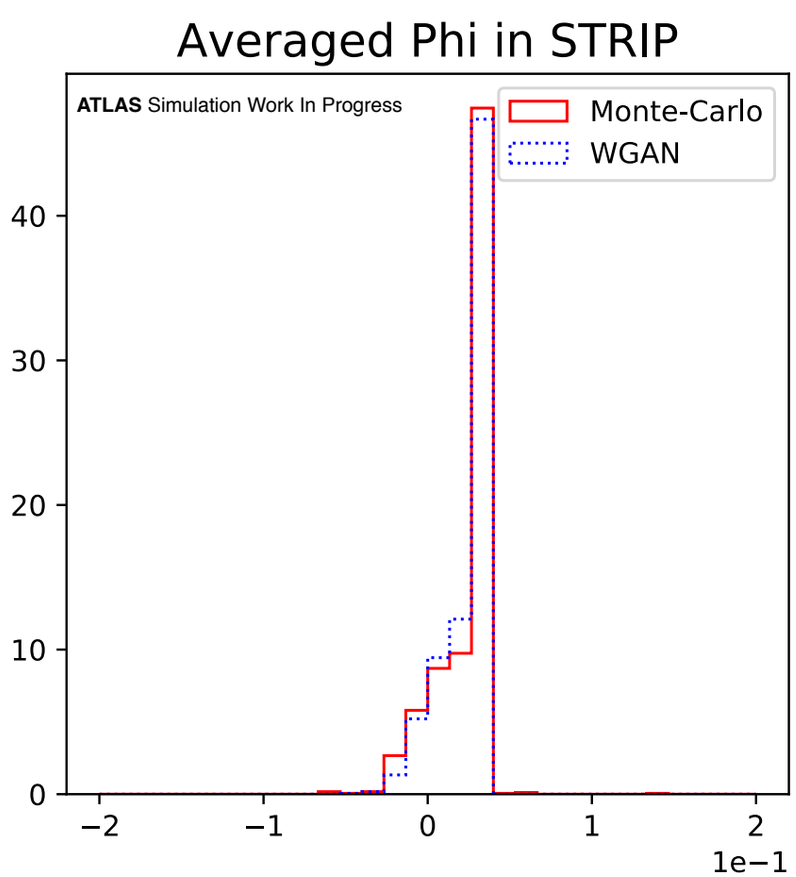


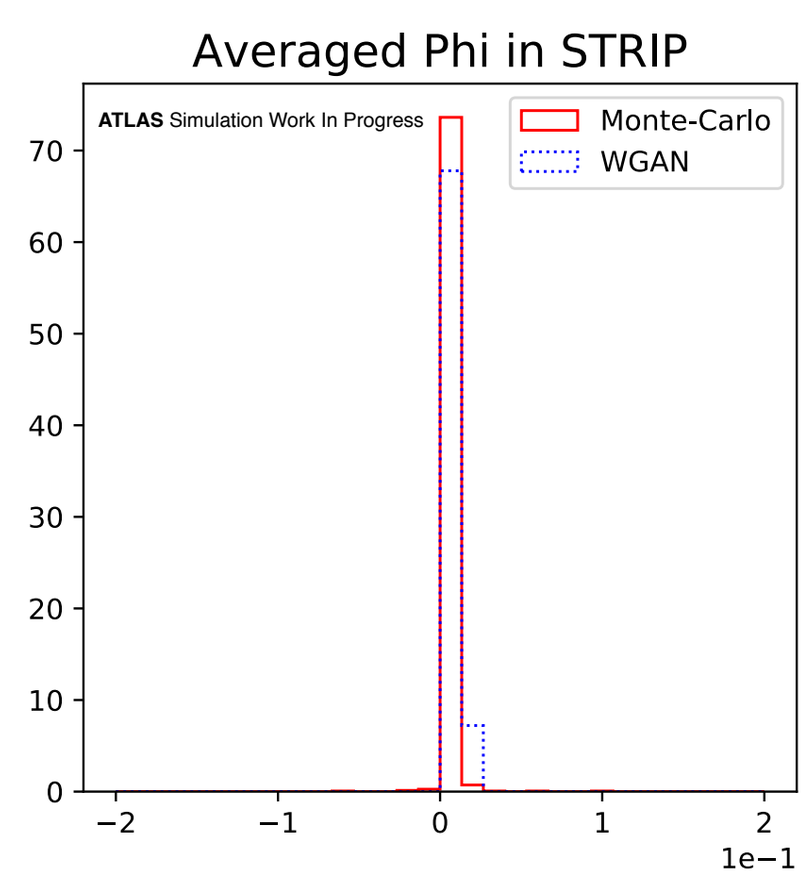
To Do:

- Tell the GAN about 8 configurations: one-hot encoding



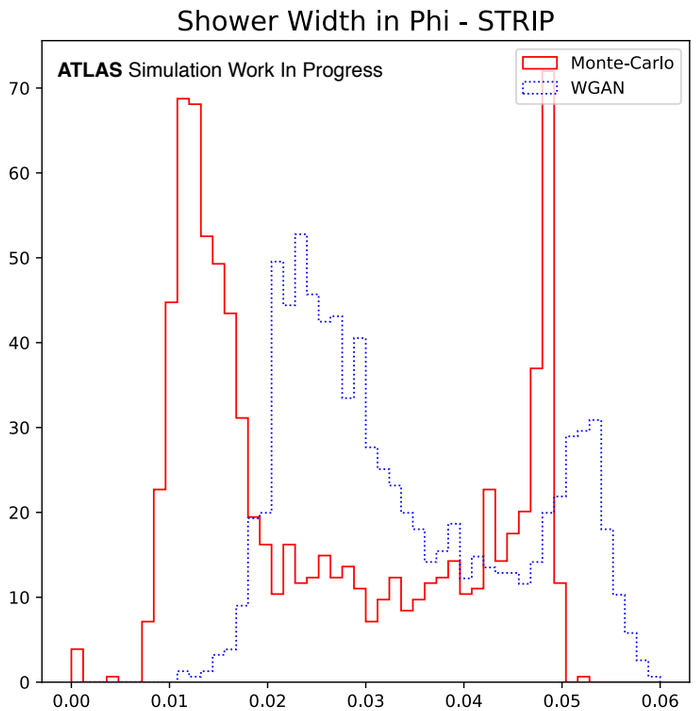




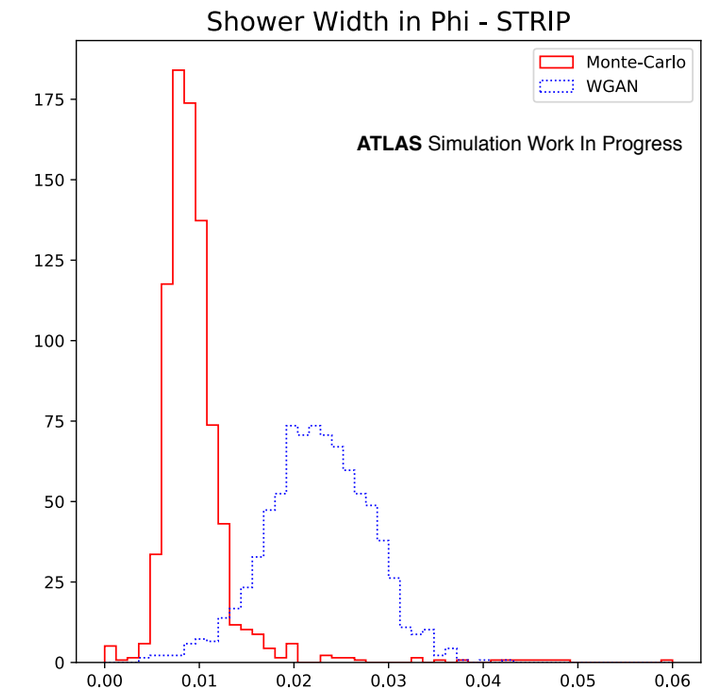


The two GANs learn their respective distributions well

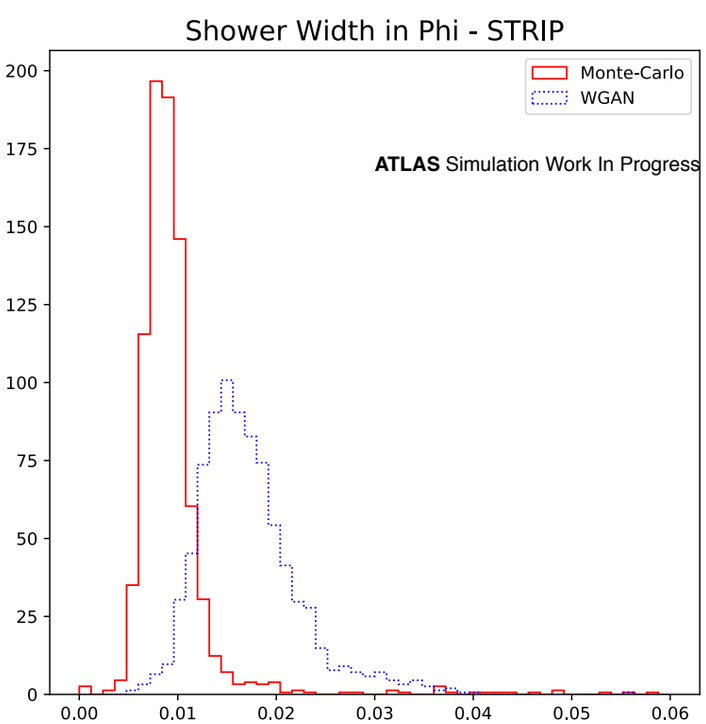
Double Peak in ϕ Width in Strips



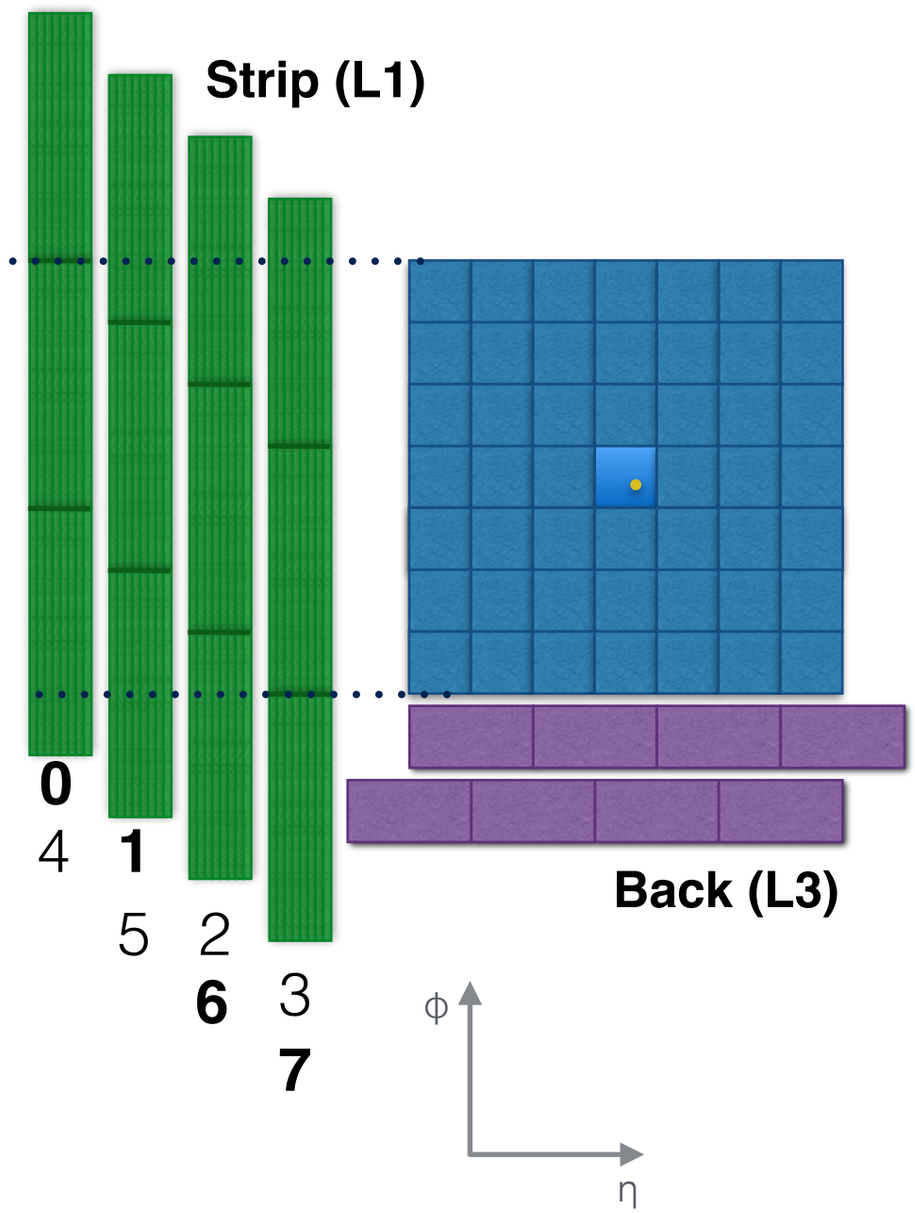
Config 0

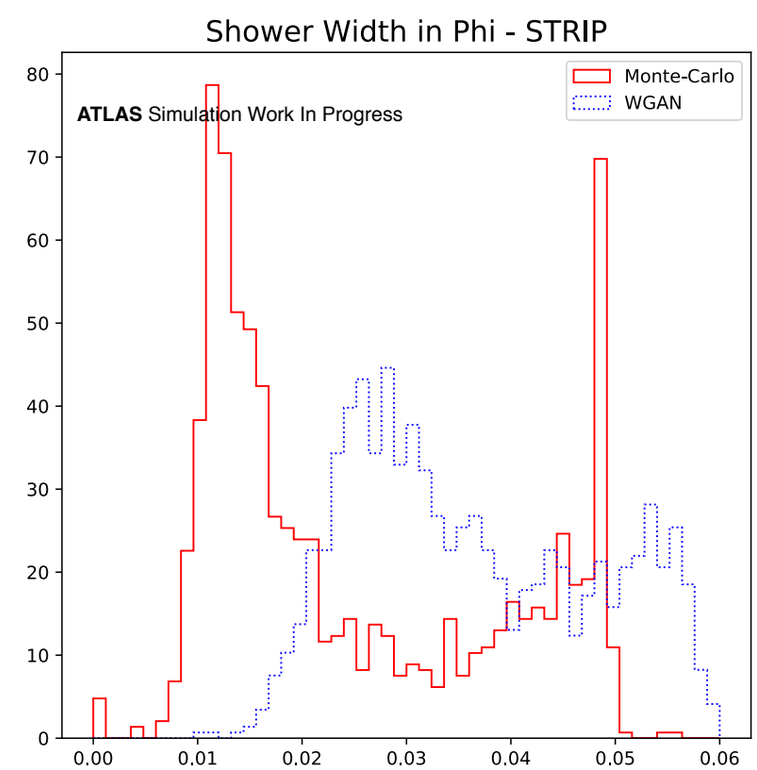


Config 1



Config 6





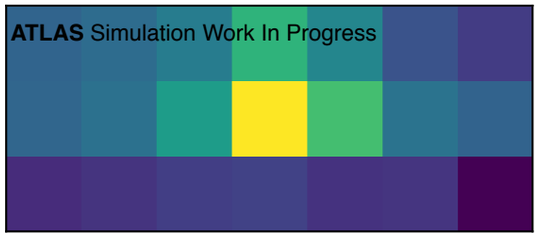
Config 7

Double peaks in config {0,7}, single peak in {1,6}

Training on Config 0

Average Images

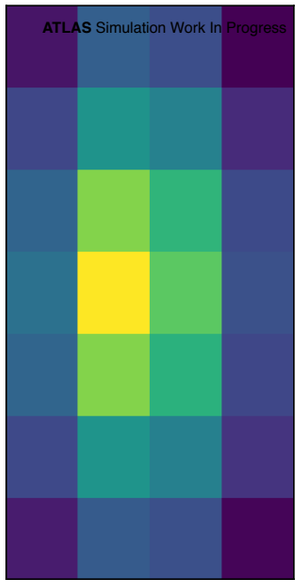
Monte-Carlo



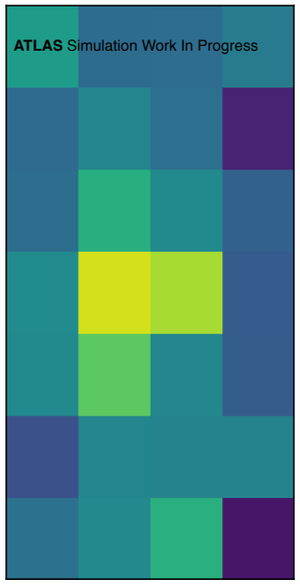
WGAN

PRE-SAMPLER

Monte-Carlo



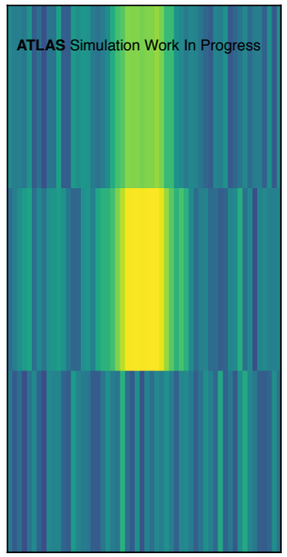
WGAN



BACK

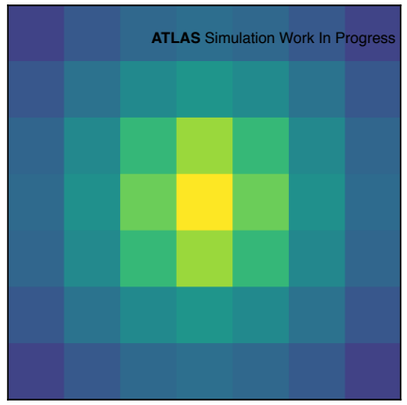
Monte-Carlo

WGAN

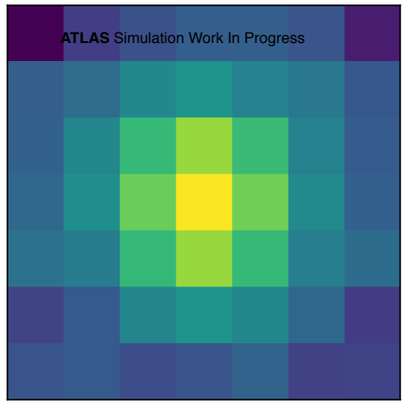


STRIP

Monte-Carlo



WGAN



MIDDLE

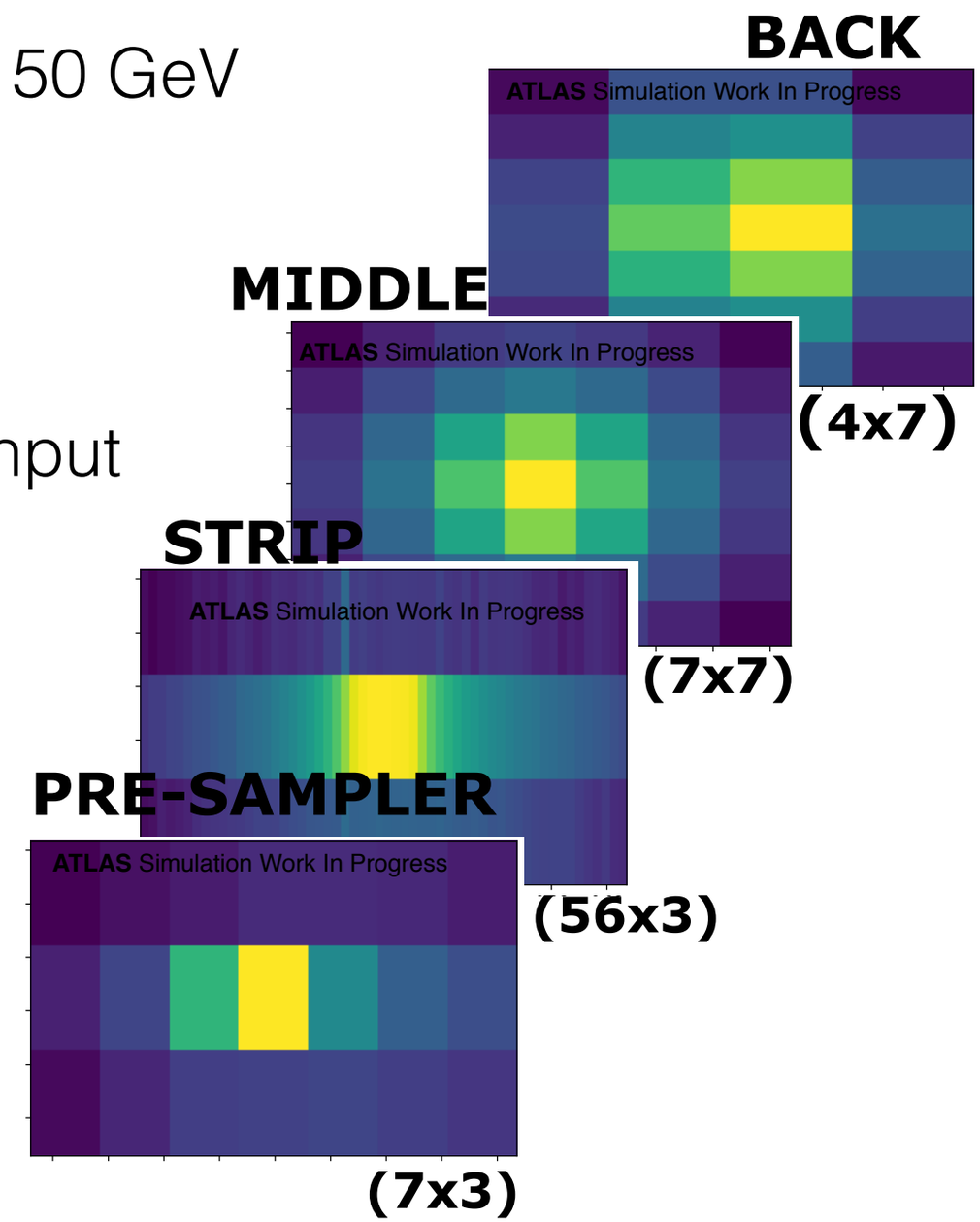
Warning: pixel sizes don't correspond to actual width of the cells

Improvement over mixed configurations training

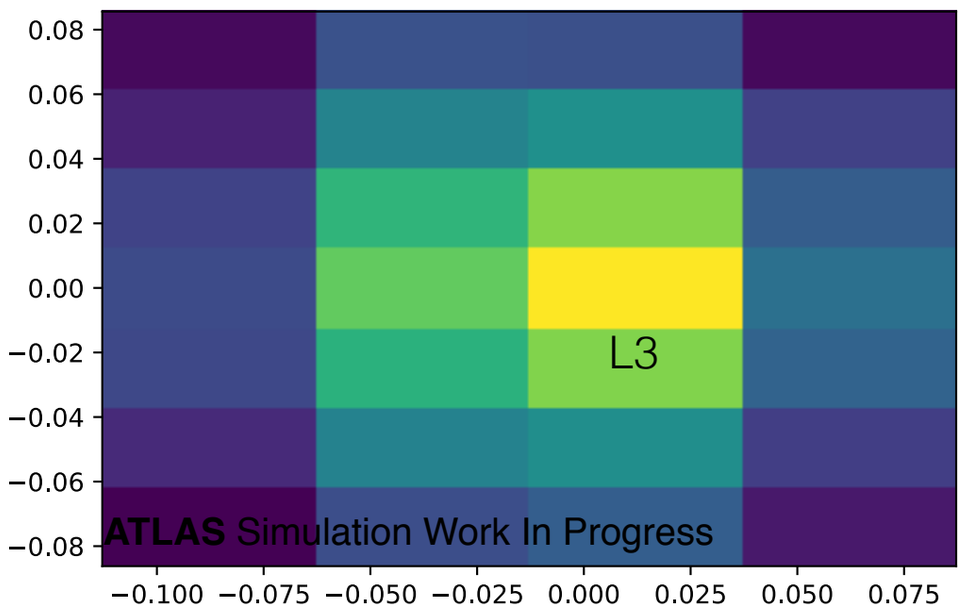
Dataset & Software

Taking over from Paul Klein, 2017

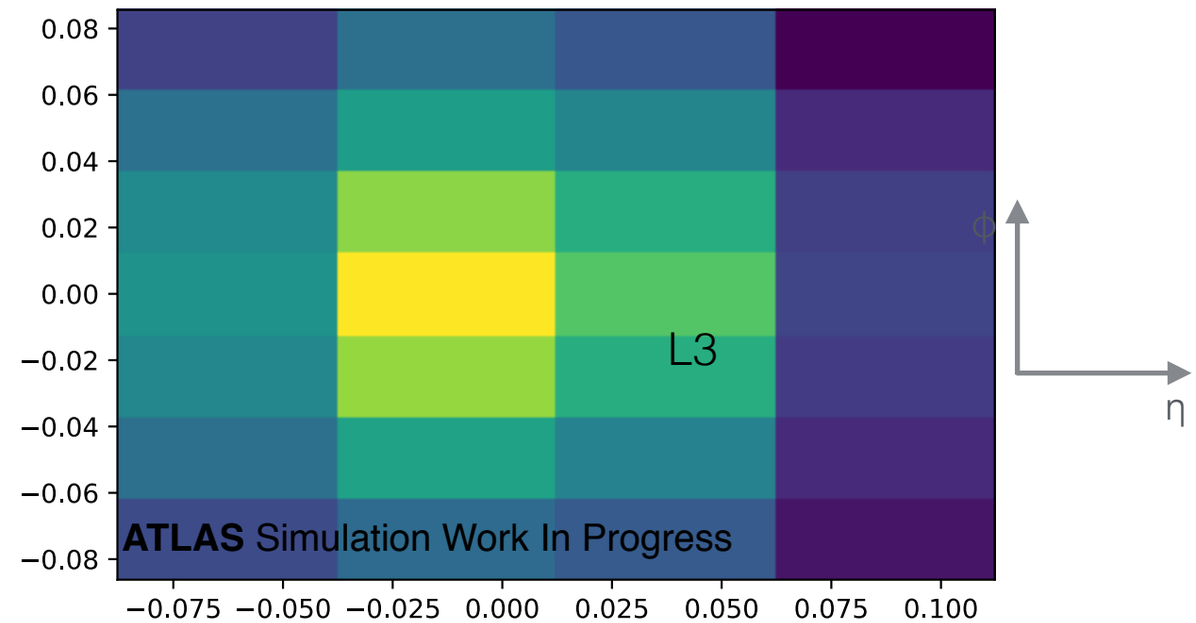
- Geant4 simulated single photon showers at 50 GeV
- $|\eta| \in [0.20, 0.25]$
- 266 Cells (Covering 4 layers) as 1-D array input
- 10k Events
- Train using: **Keras with Tensorflow**
- GAN Training time: ~ 2h50m
- Training Platform: **1 GPU at CC Lyon**
- GAN Generation Time: 1-2 seconds for 10k events

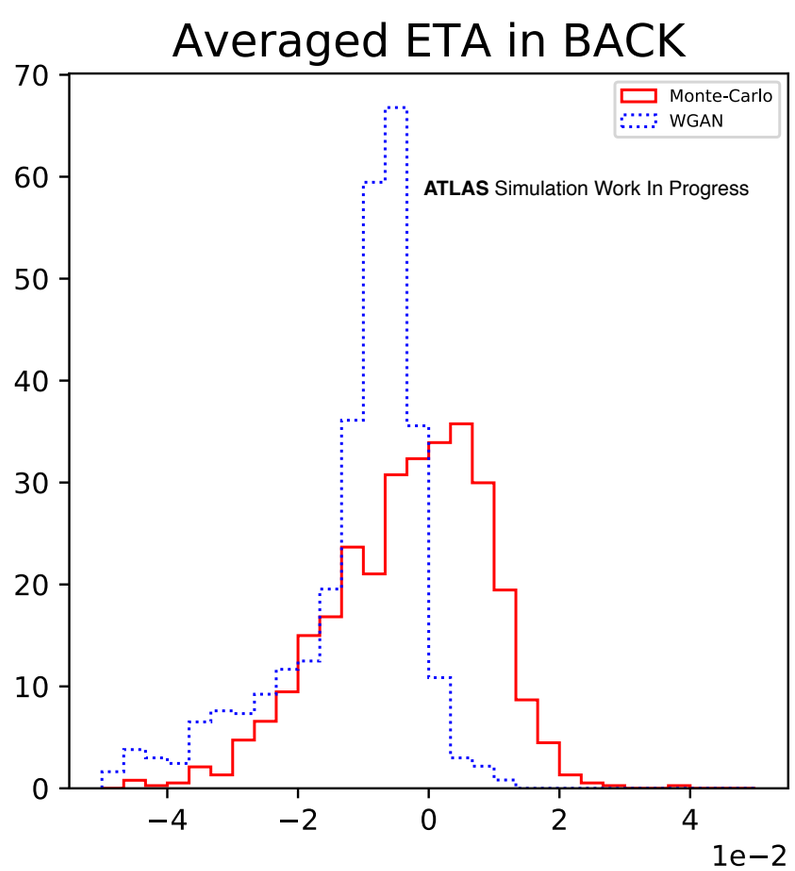


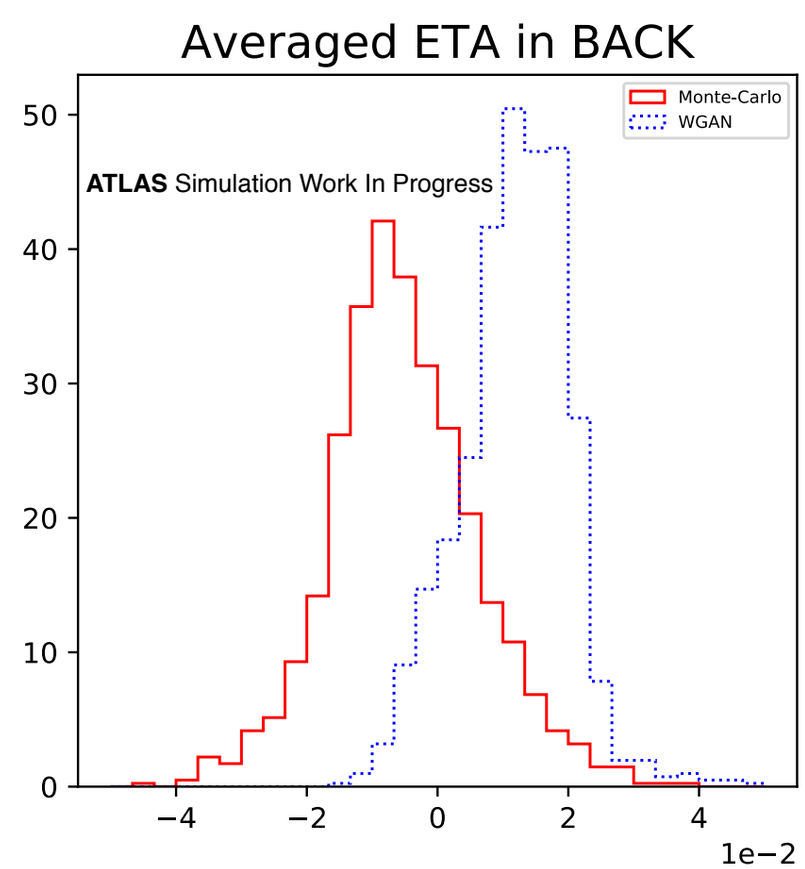
Back Config 1



Back Config 7

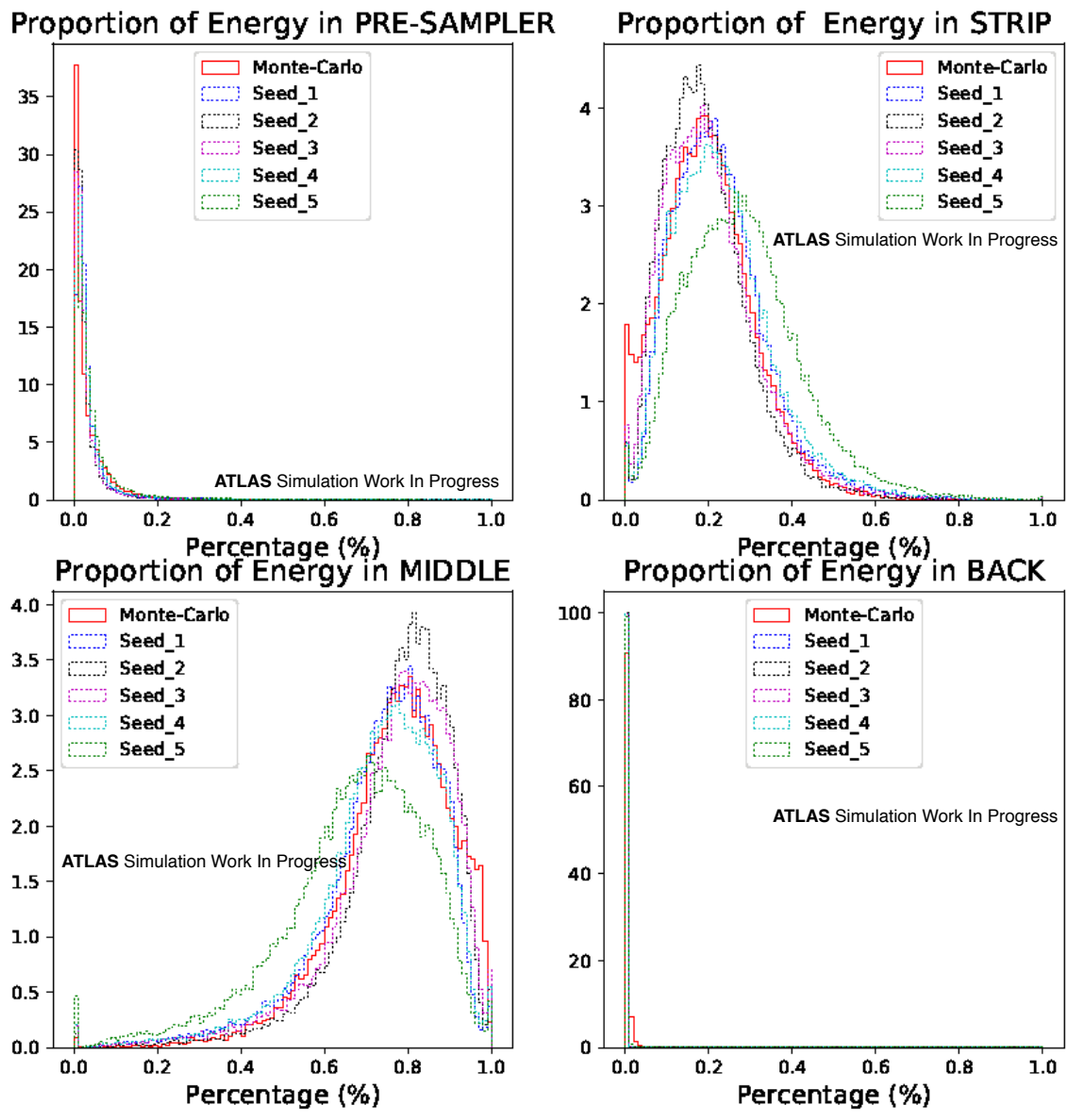




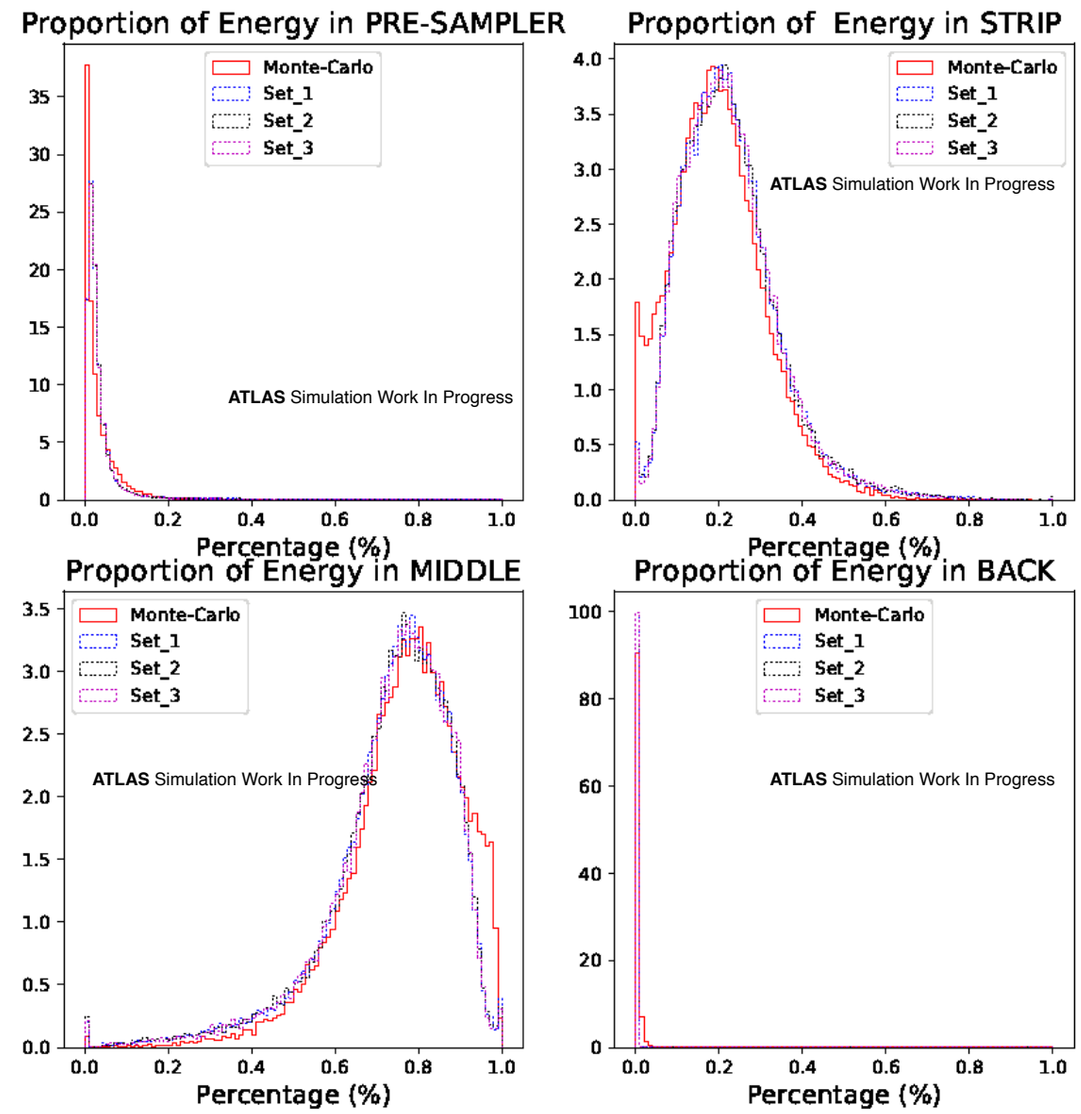


Variation in Training for Random Seed

Different training seeds



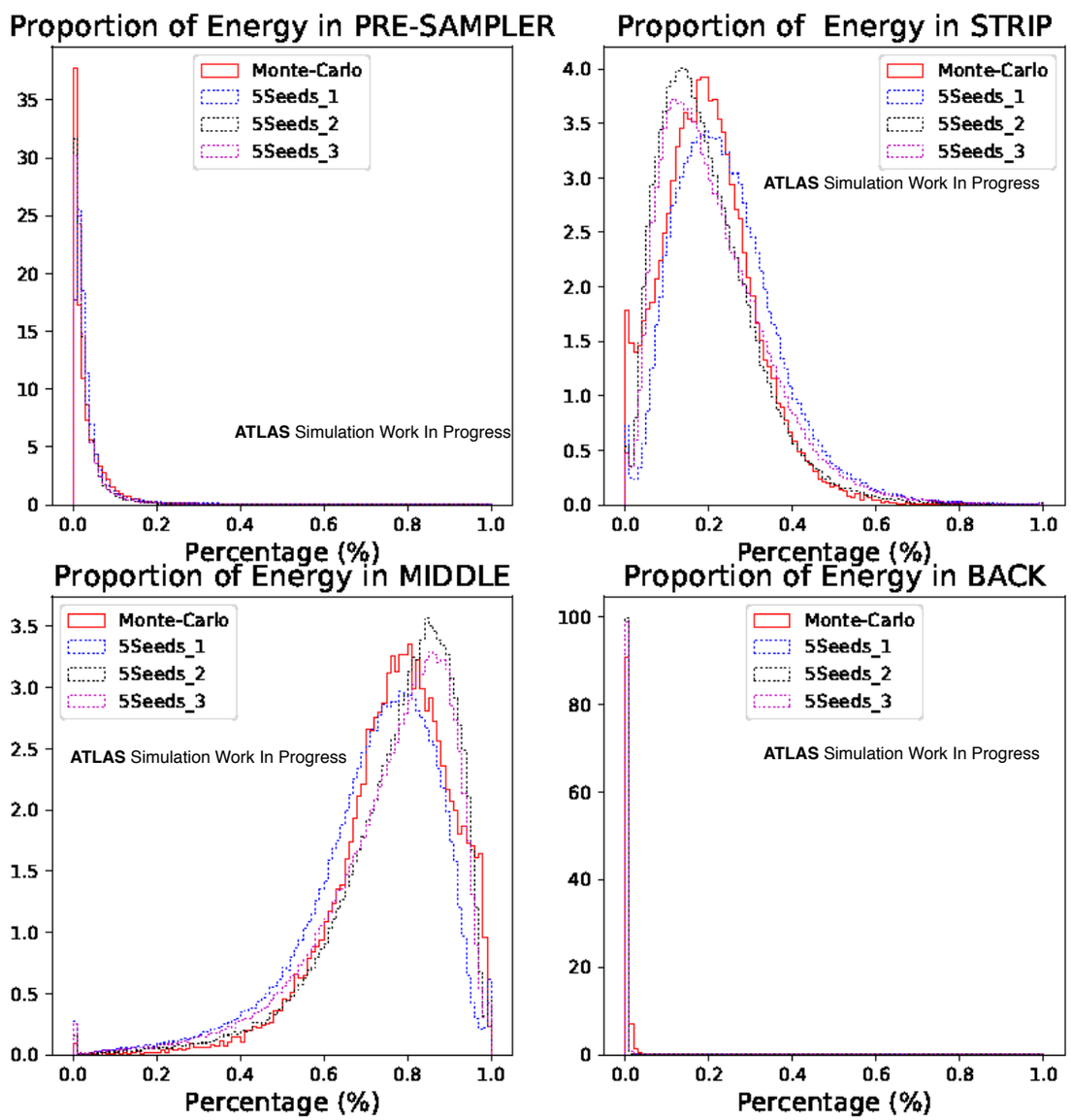
Same training seed;
different generation seeds



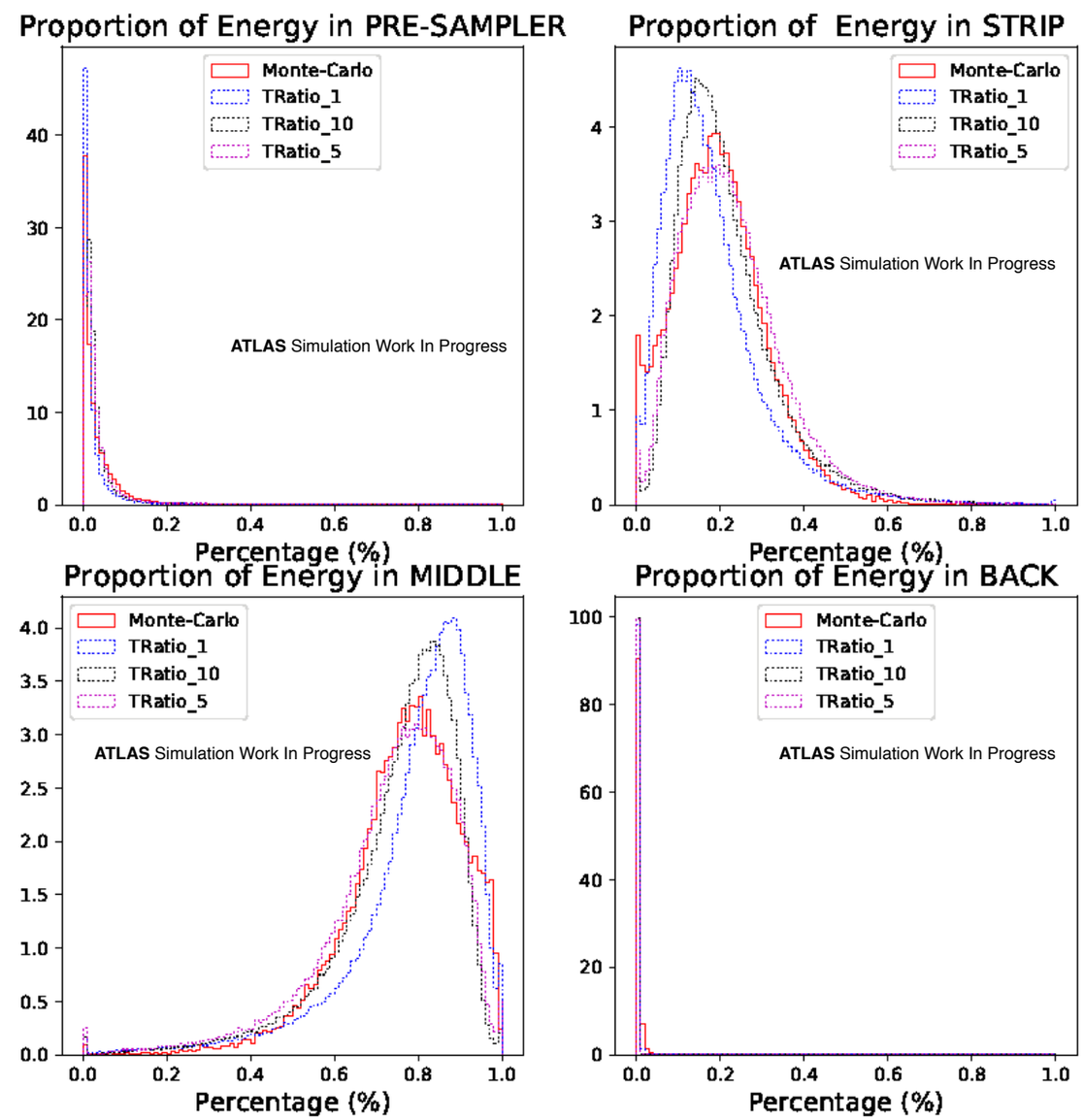
Initial training seed has significant impact on results. Need to average over different seed training to assess the GAN

Hyperparameter Optimisation: Training Ratio

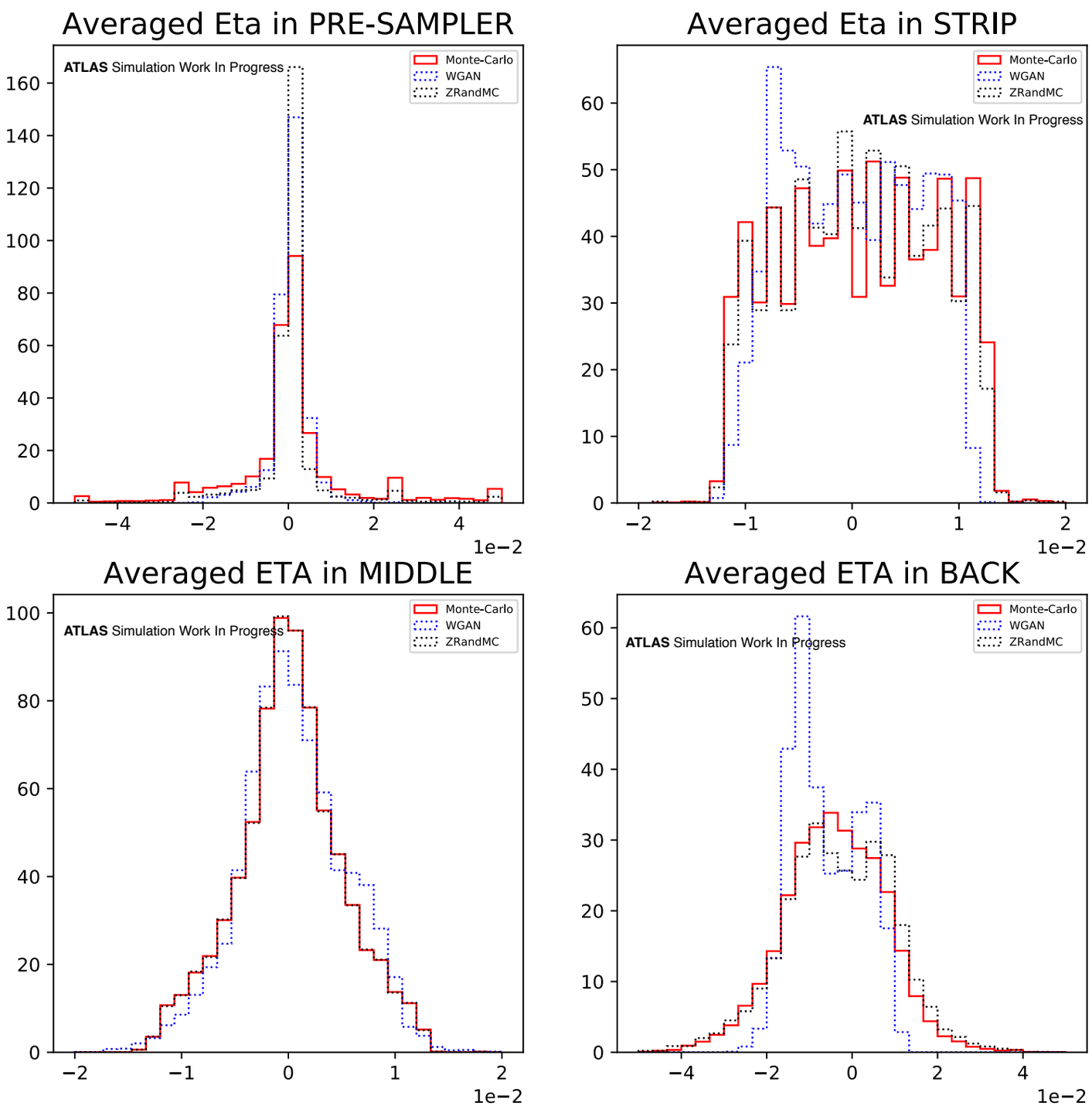
Same hyper-params,
3 sets of 5 separately trained GAN datasets



1,5,10 Ratio; combine 5 trained GANs for each



Unfair comparison for GAN



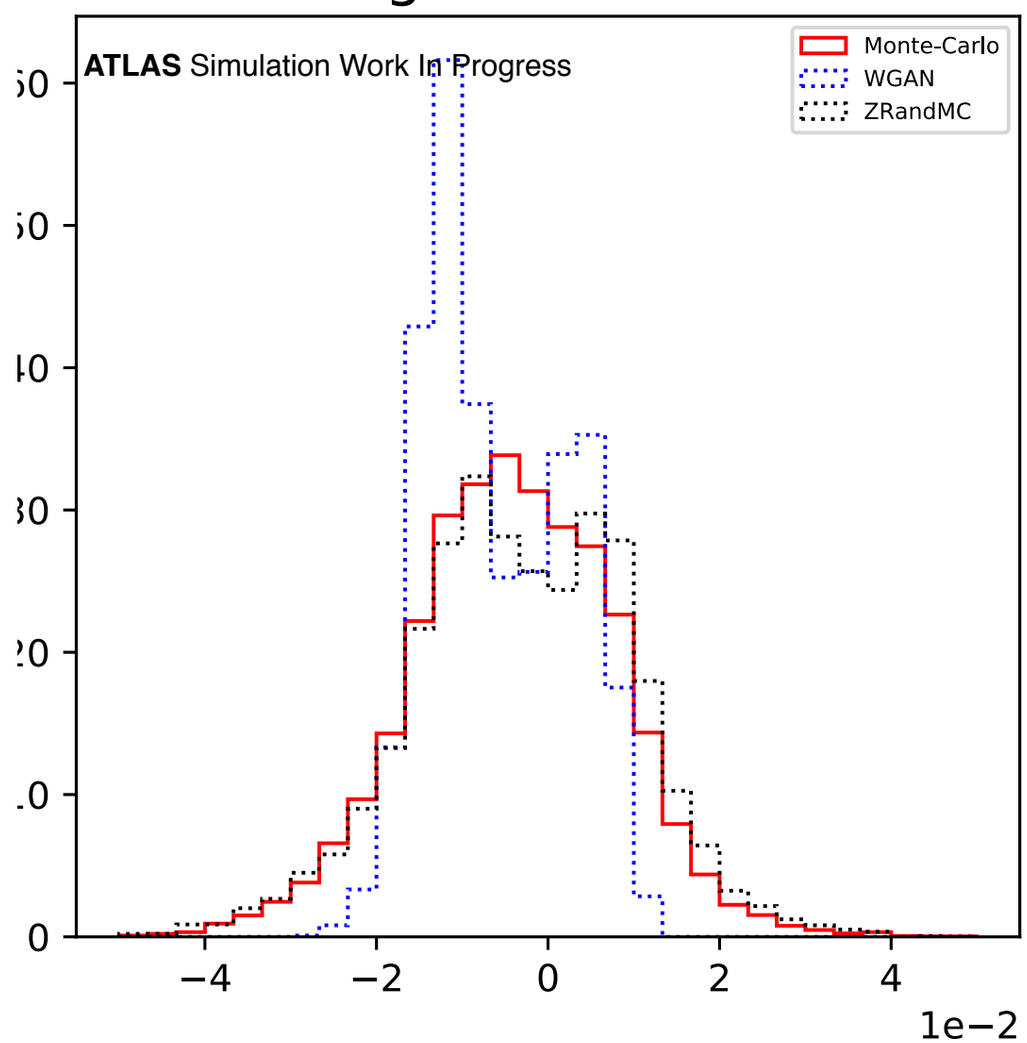
Very clearly that we're mixing these 8 different configurations.

MC Events Plotted with Eta of the Corresponding Events, WGAN plotted with Eta of MC Events, MC Events Plotted with Eta of Random Event

MC vs GAN vs MC

	Red	Blue	**Black**
Cell Energy (E)	MC	GAN	MC
Cell η	From same event	From an MC event	From different MC event
E-η Correspondence	Correct	N/A	Randomised

Averaged ETA in BACK

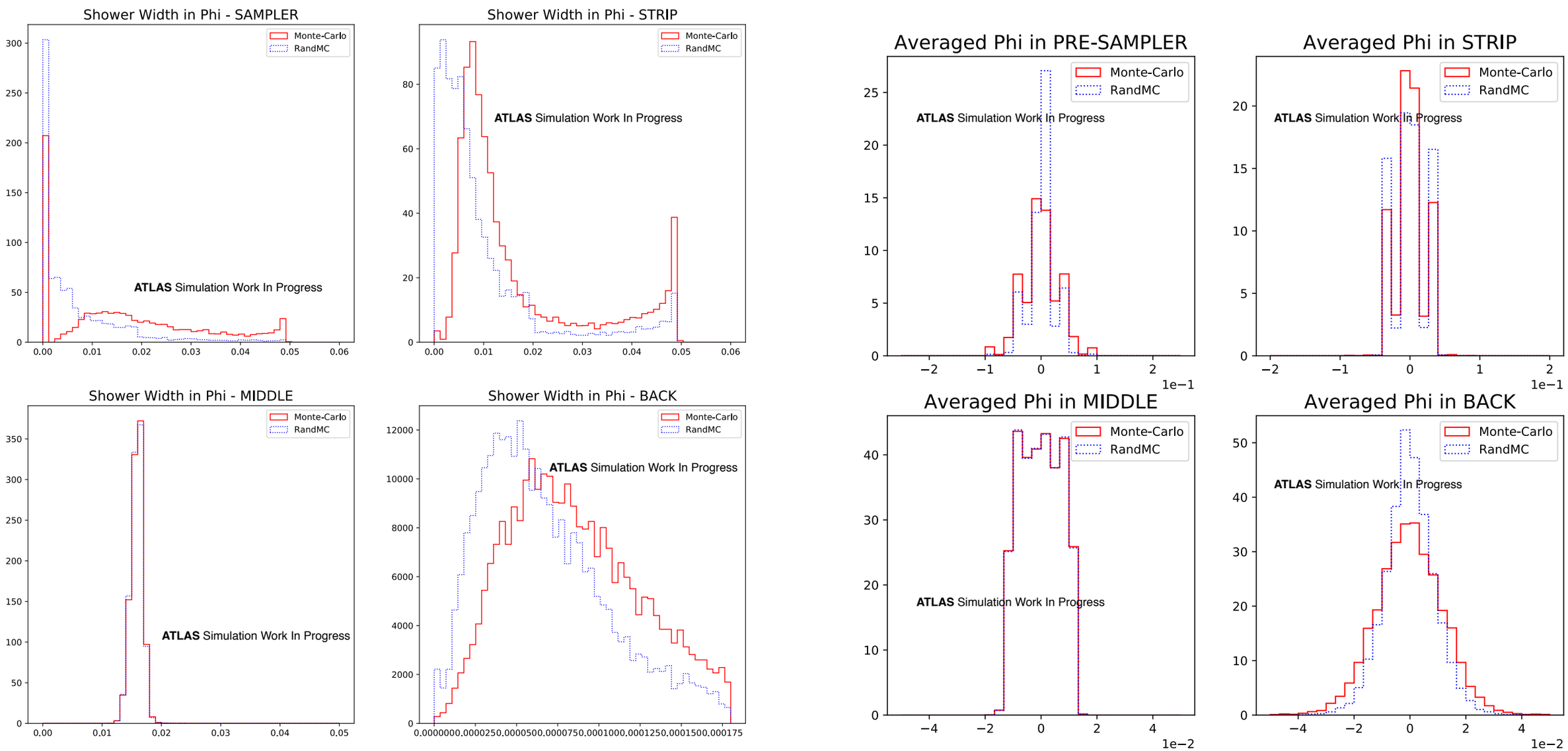


Two peaks for the blue and **black**

MC vs MC

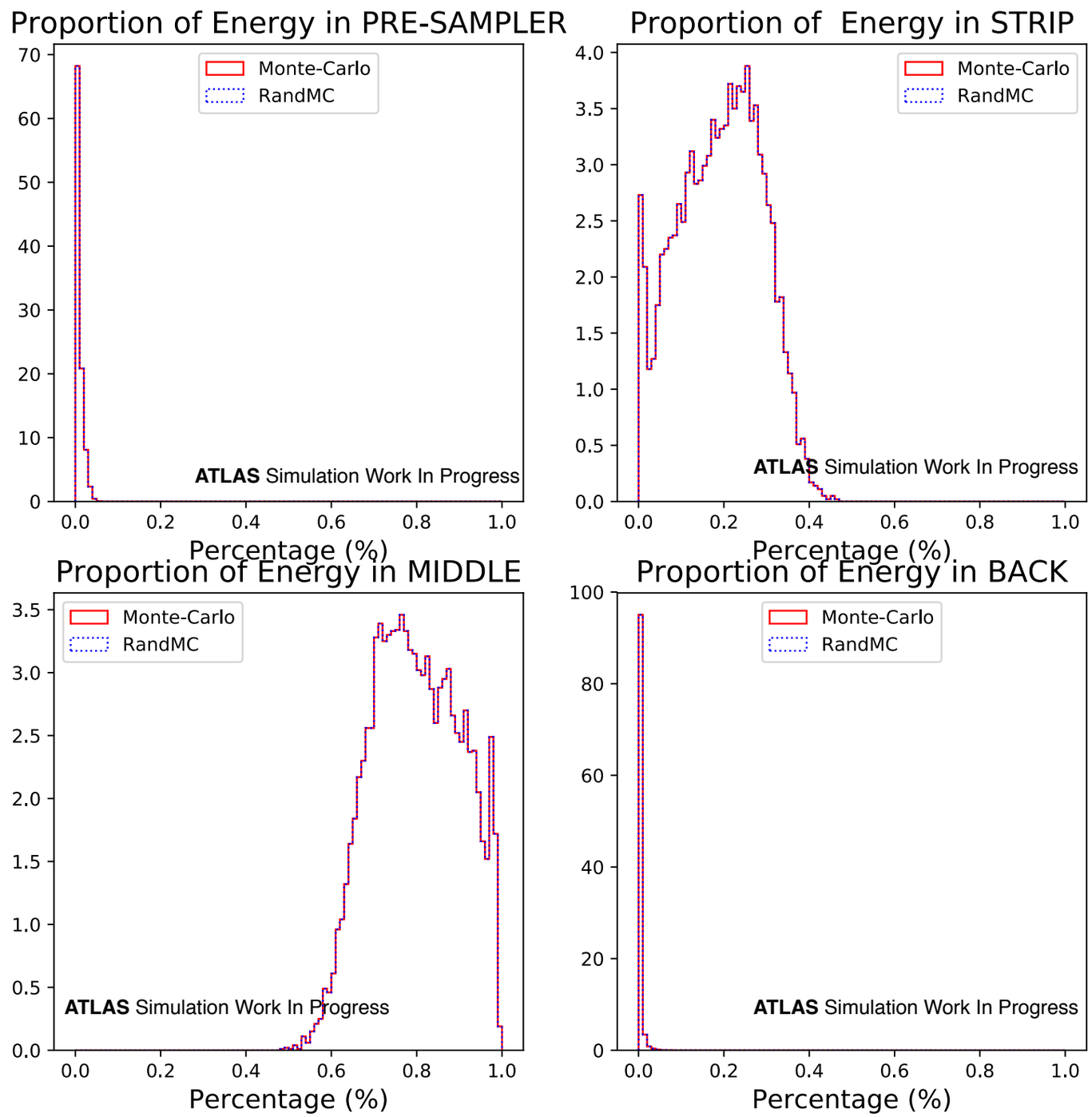
Phi from Corresponding Event vs Phi from another random event

Blue here is also Geant4 data (not GAN!)

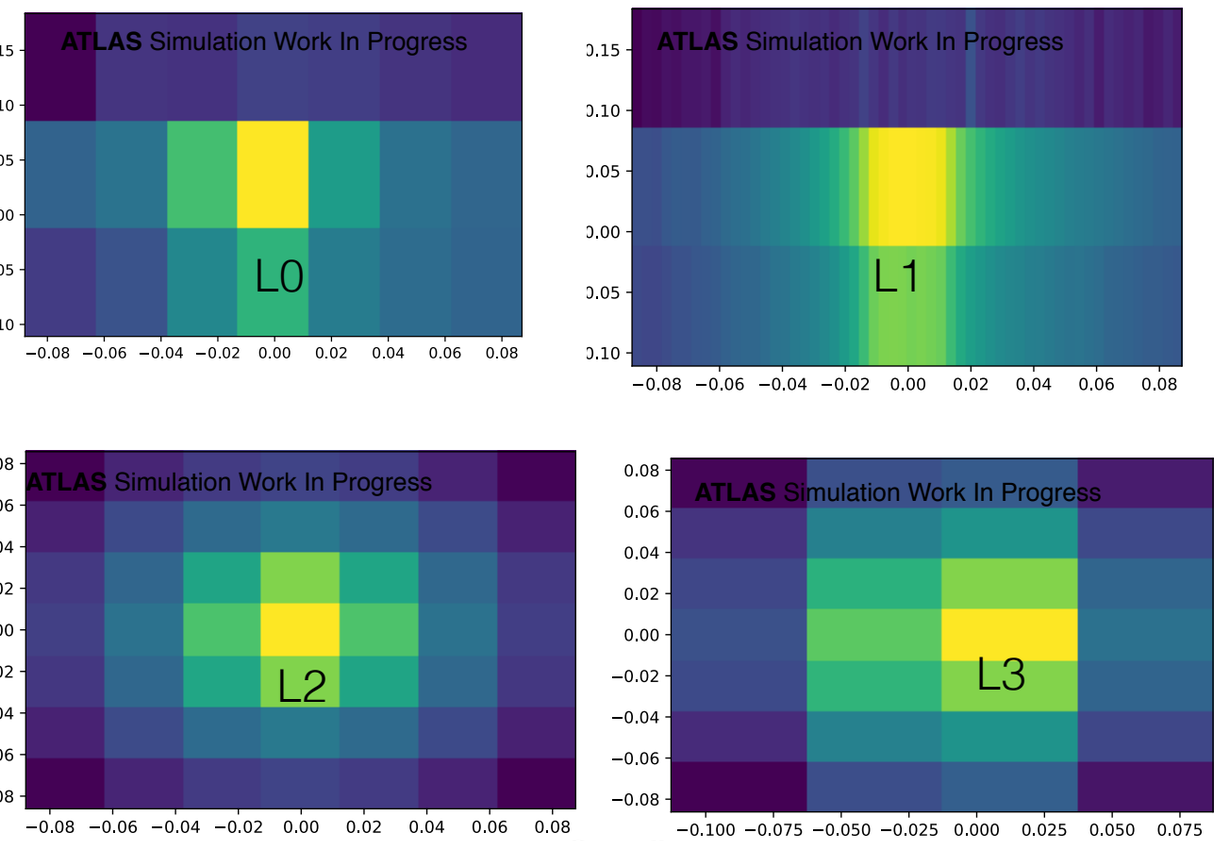


Distributions for Middle Layer are almost perfect

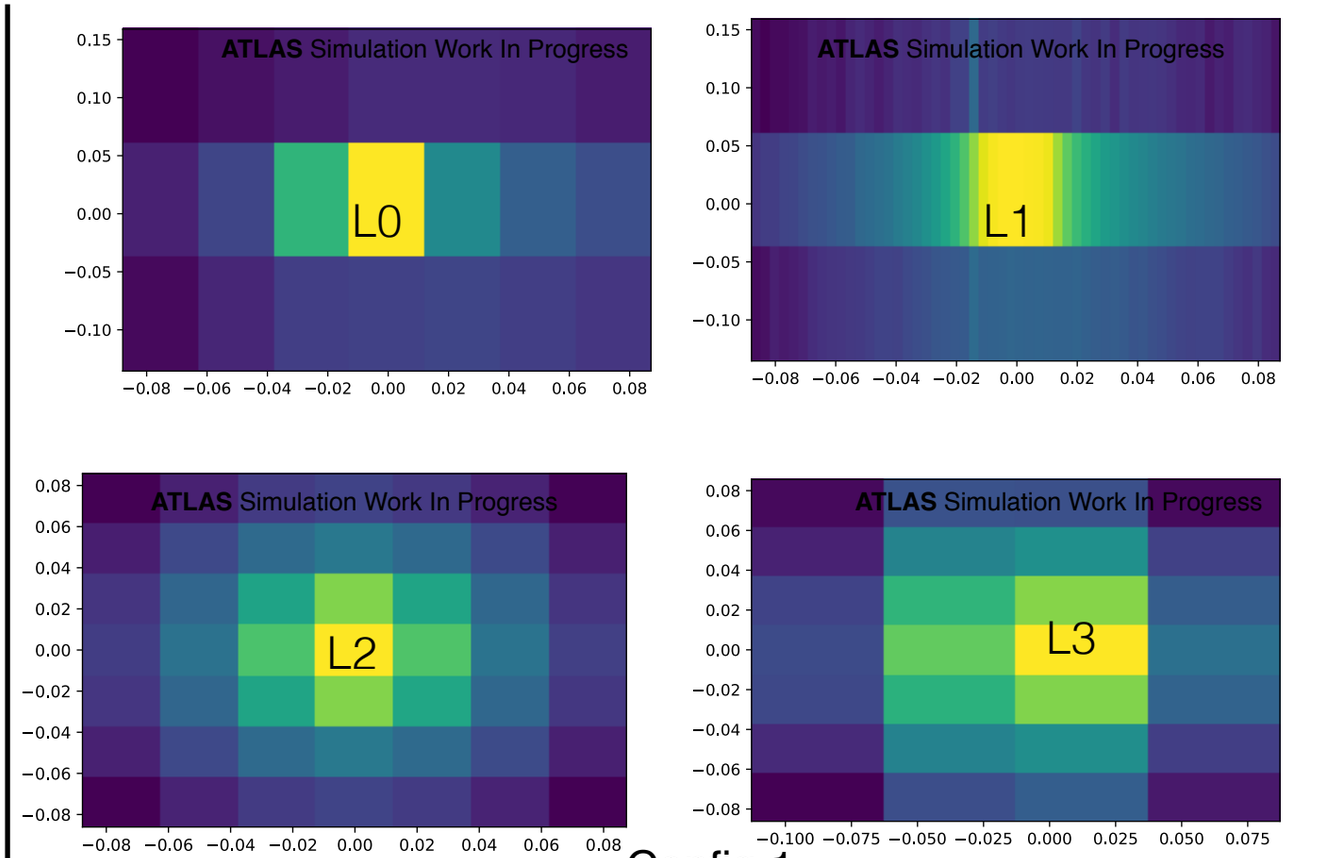
MC vs MC



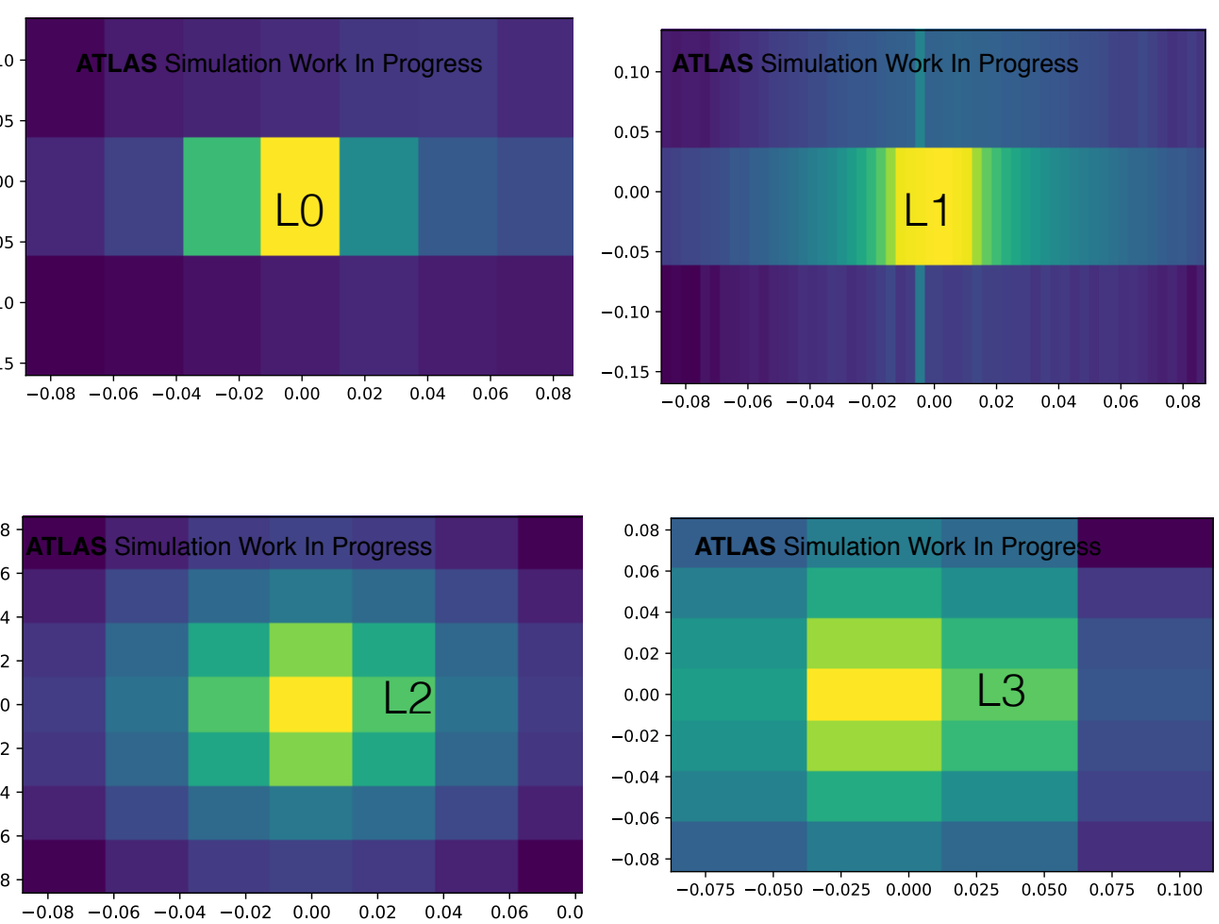
Energy Distributions Match Perfectly



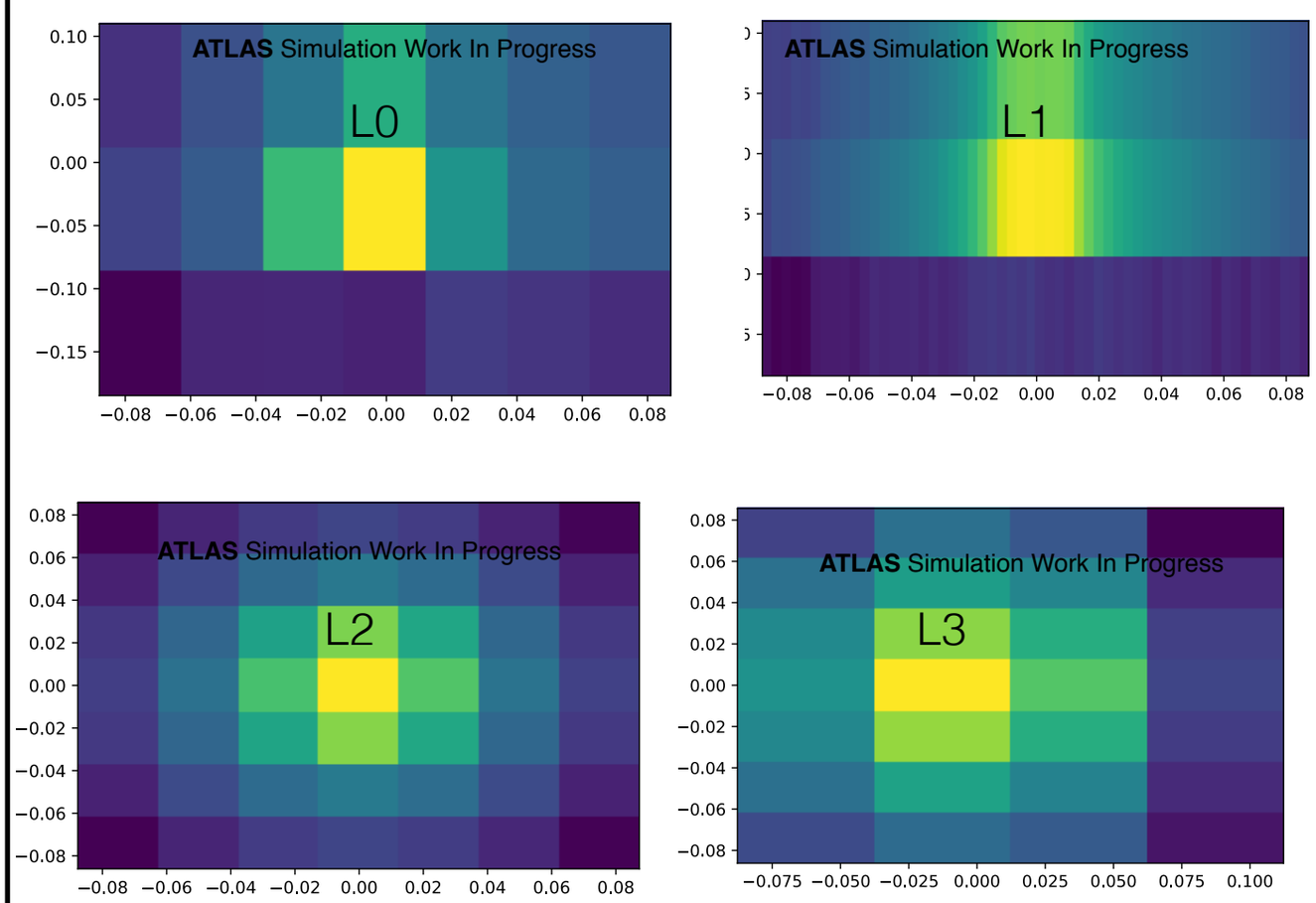
Config 0



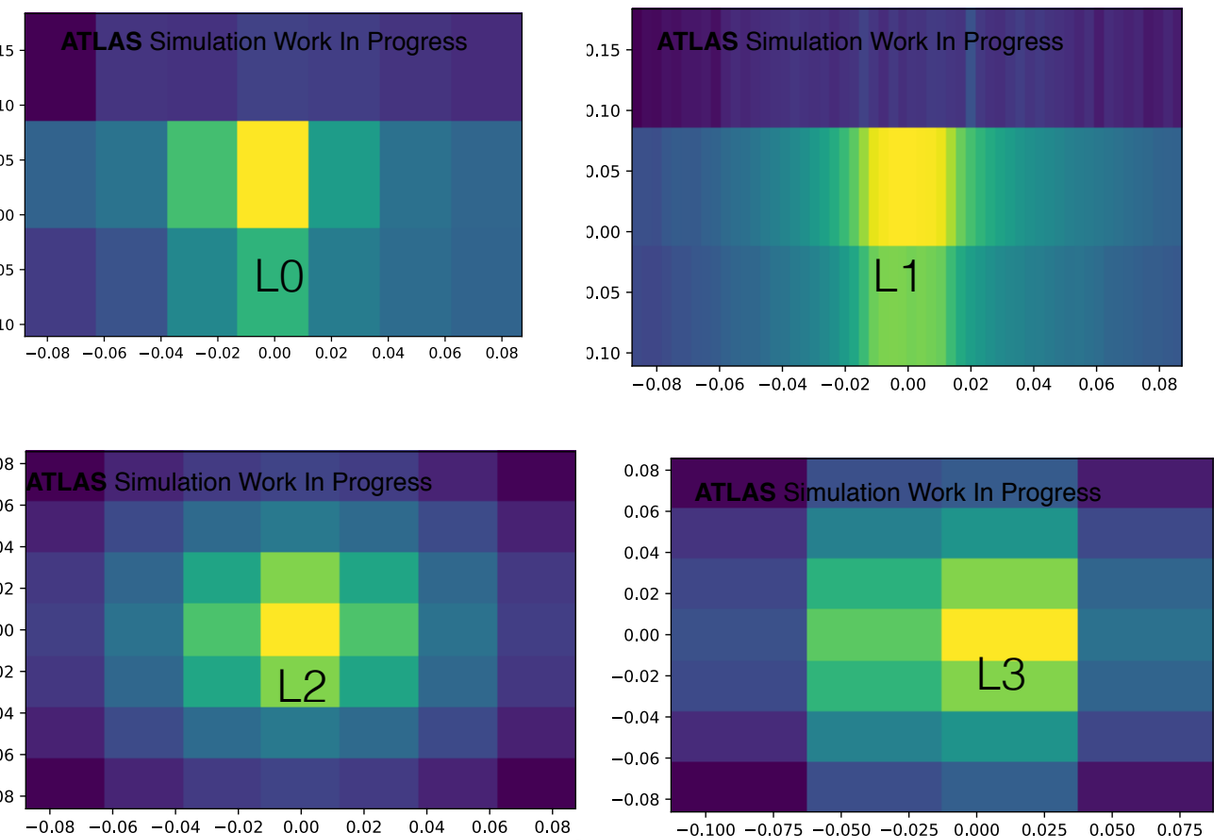
Config 1



Config 6

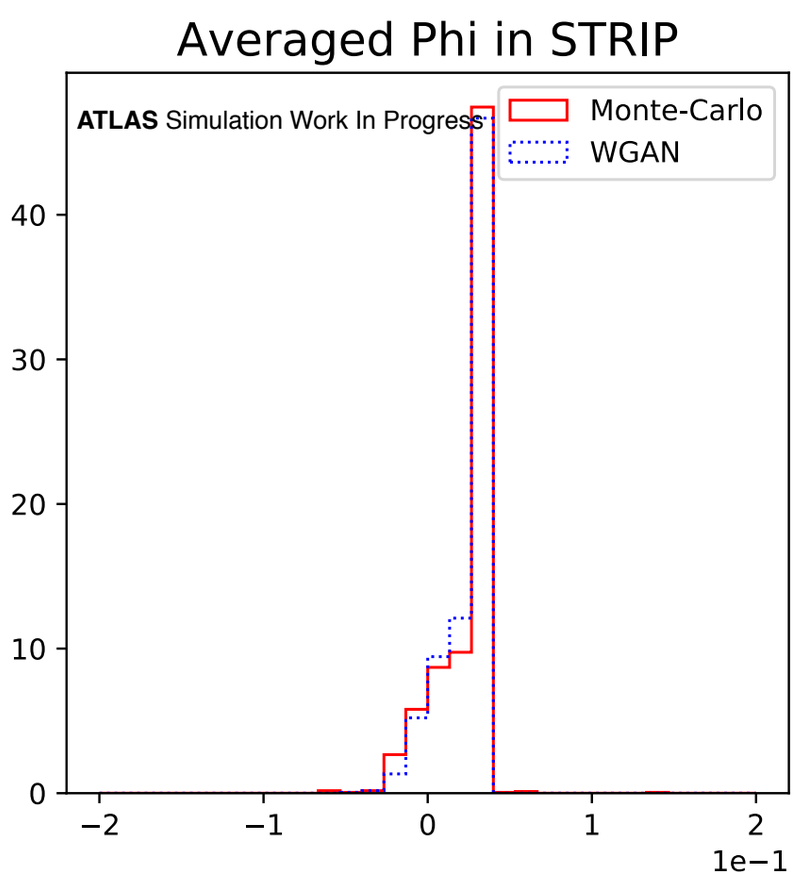


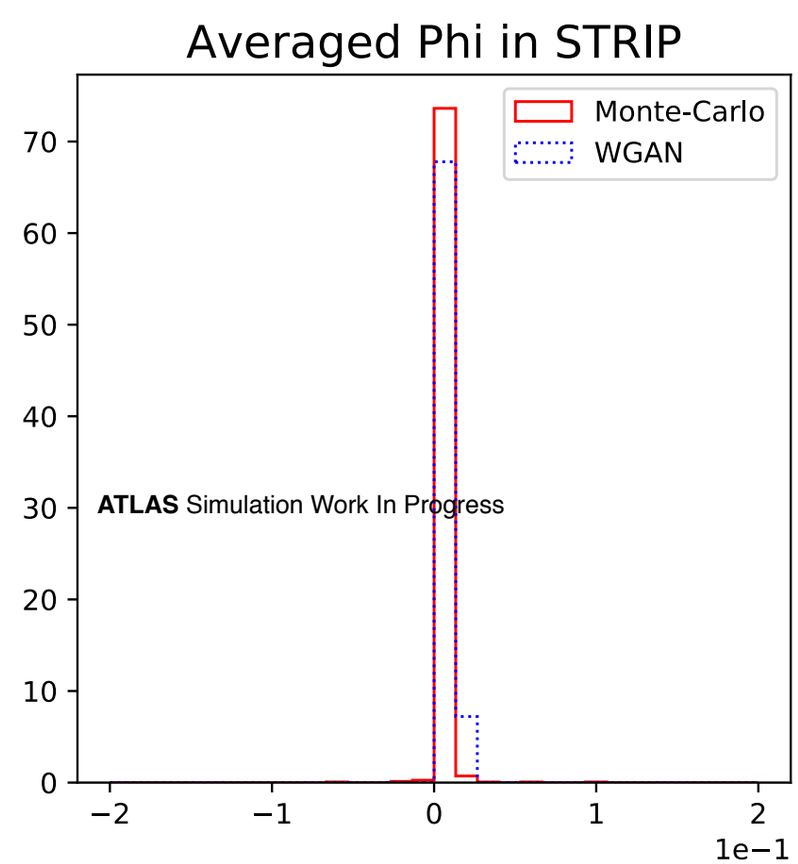
Config 7

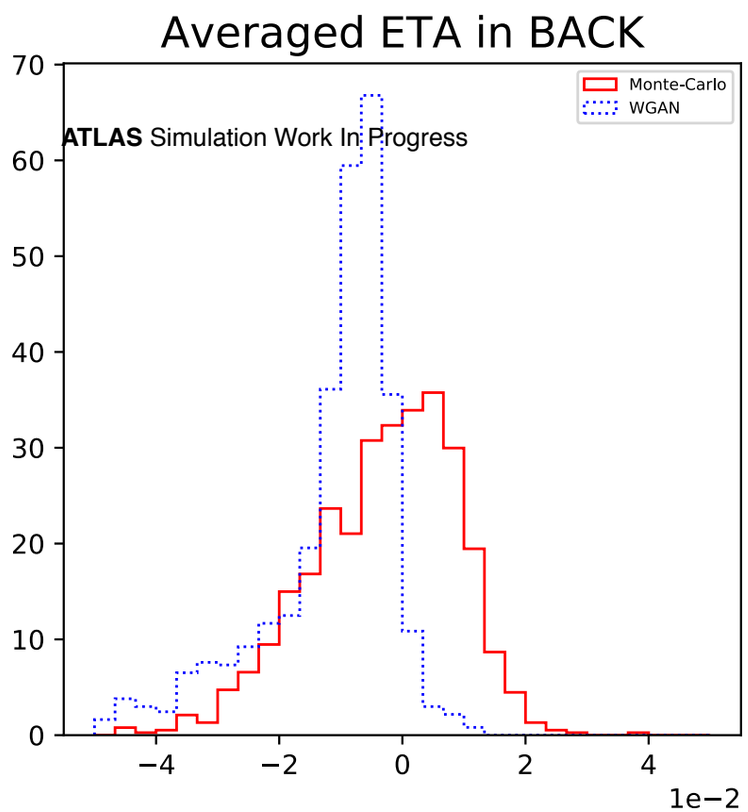


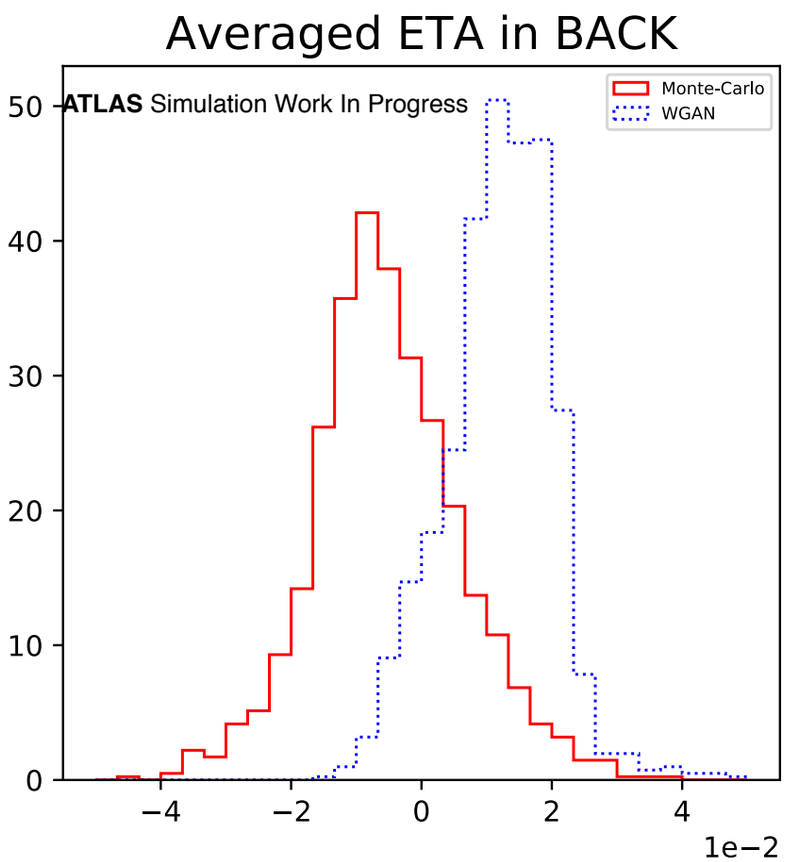
Config 0

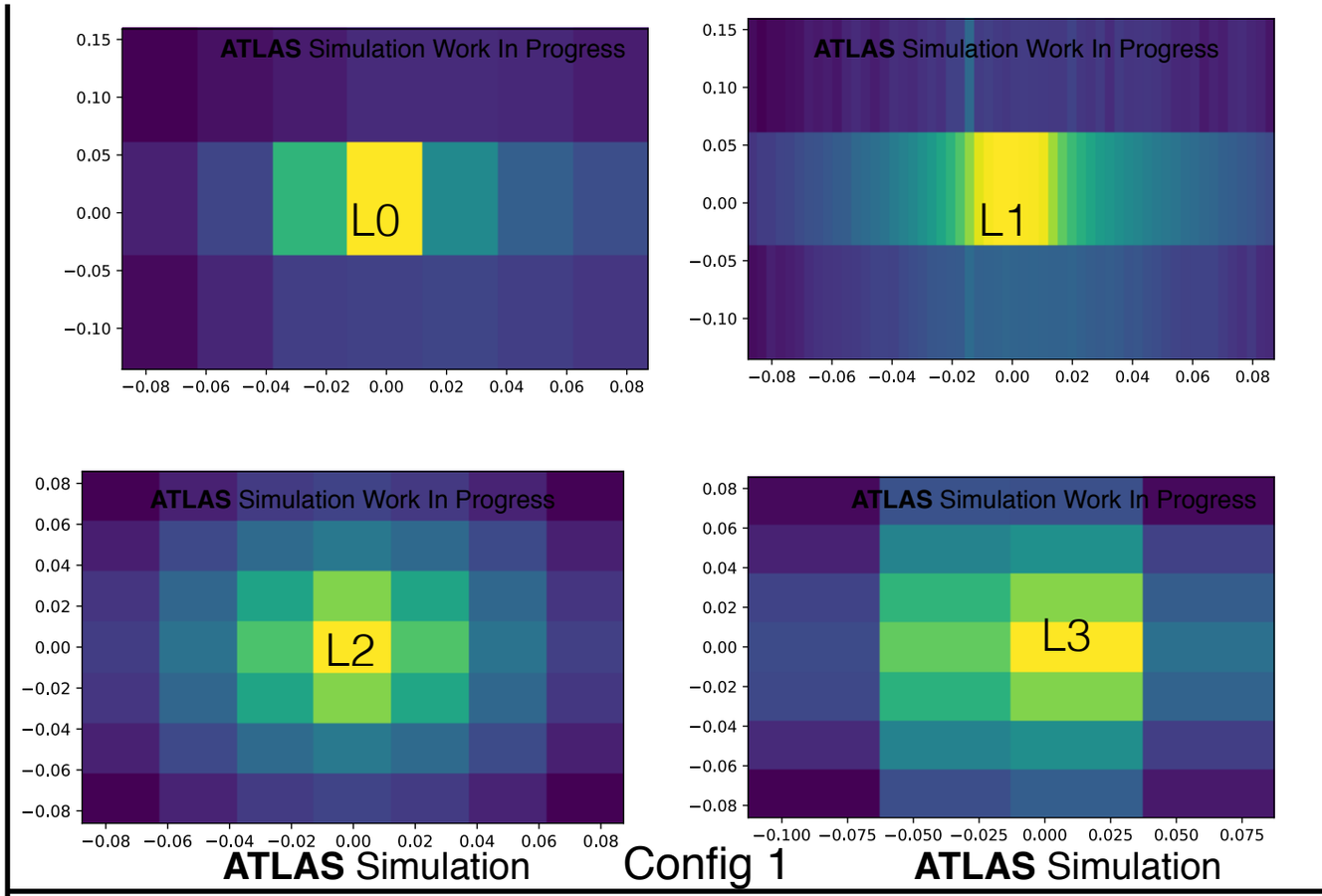
Config 1



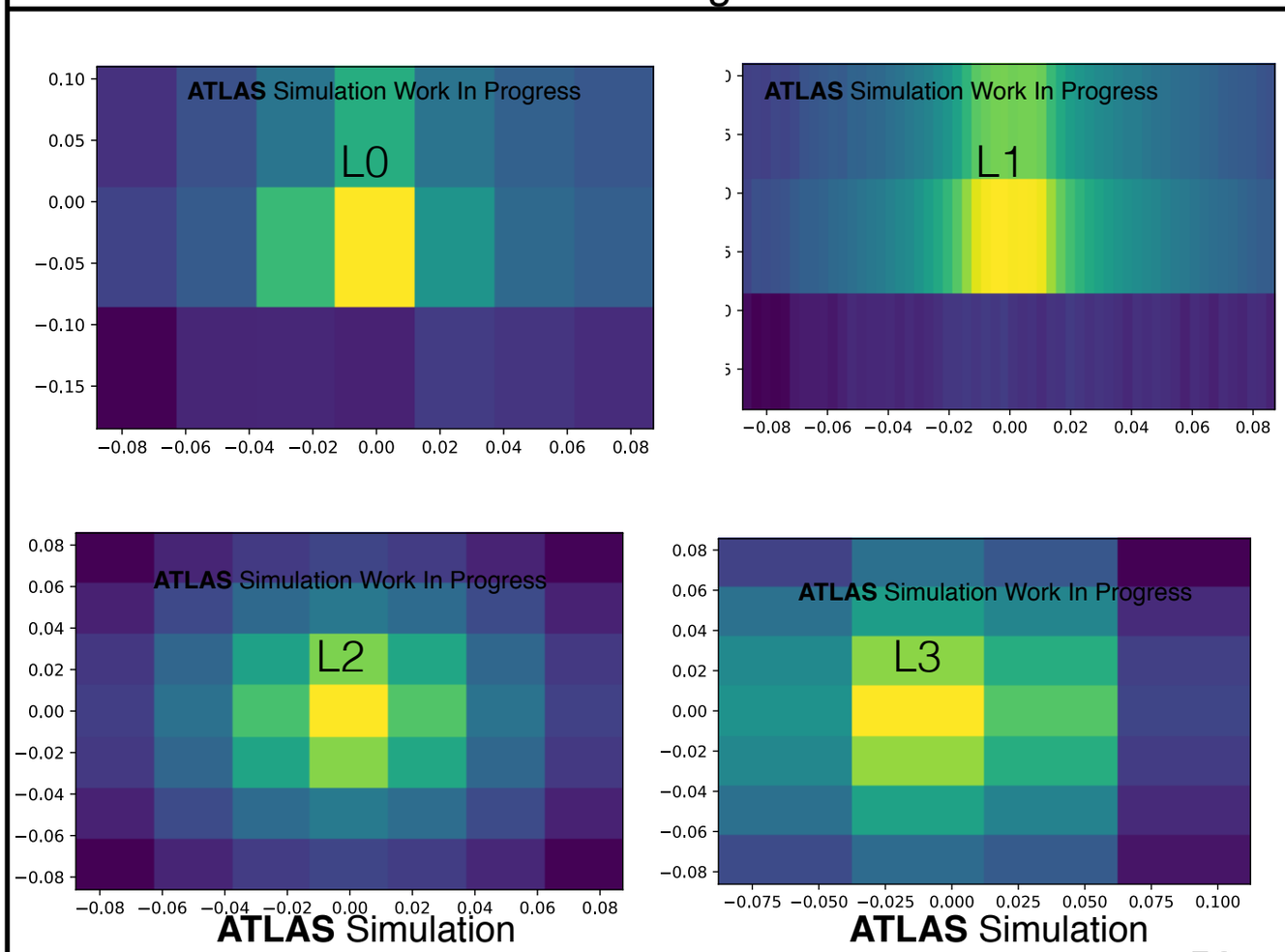




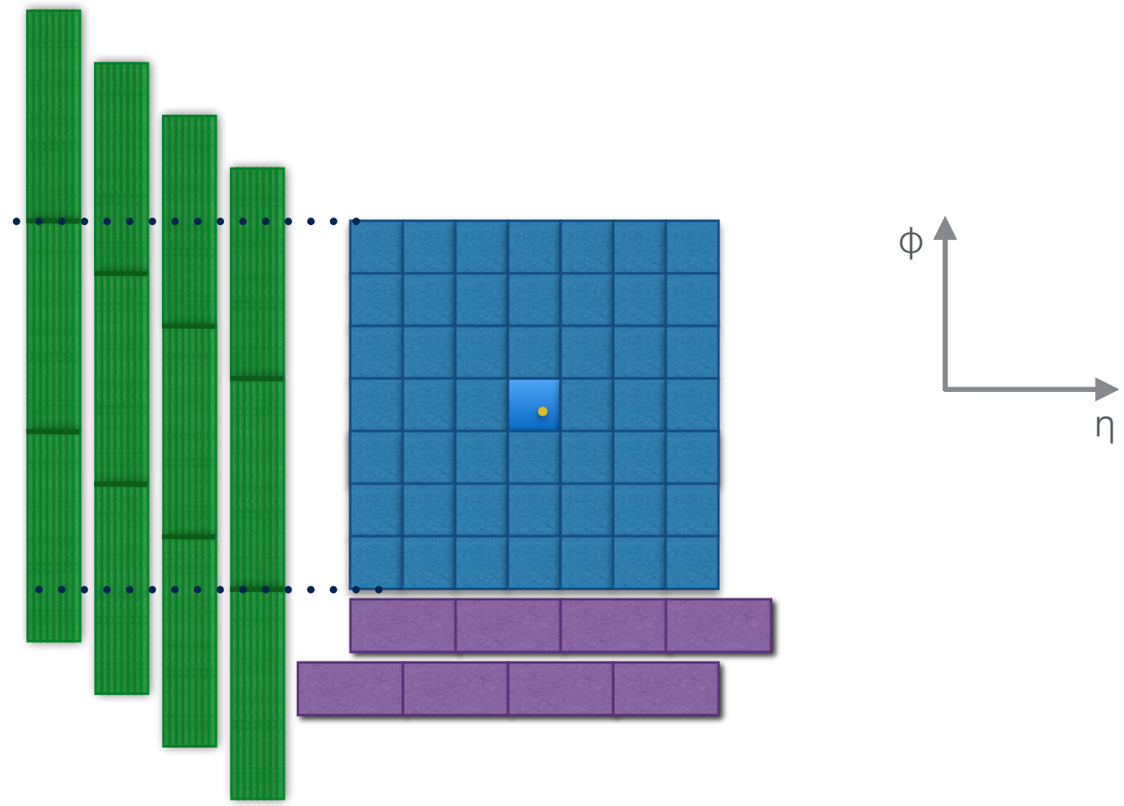




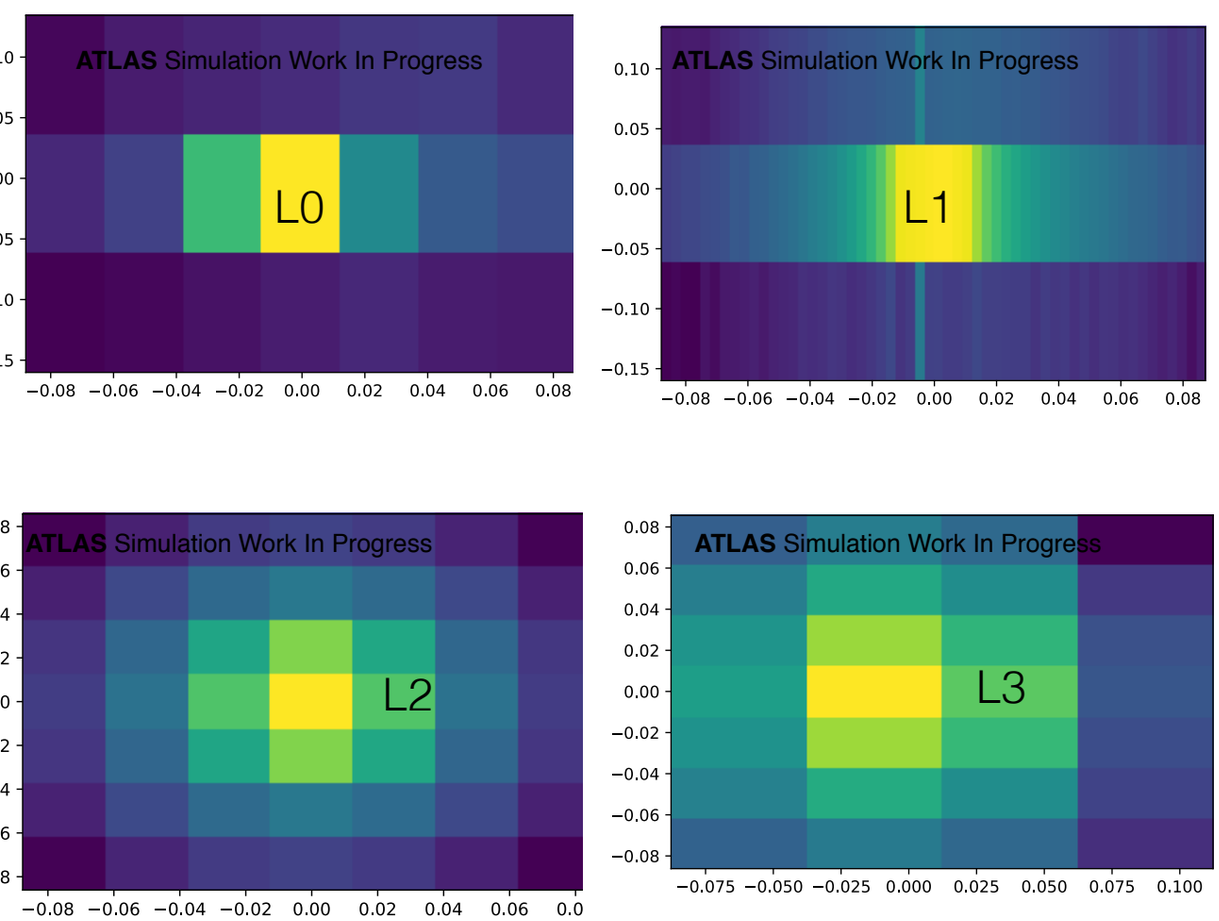
Config 1



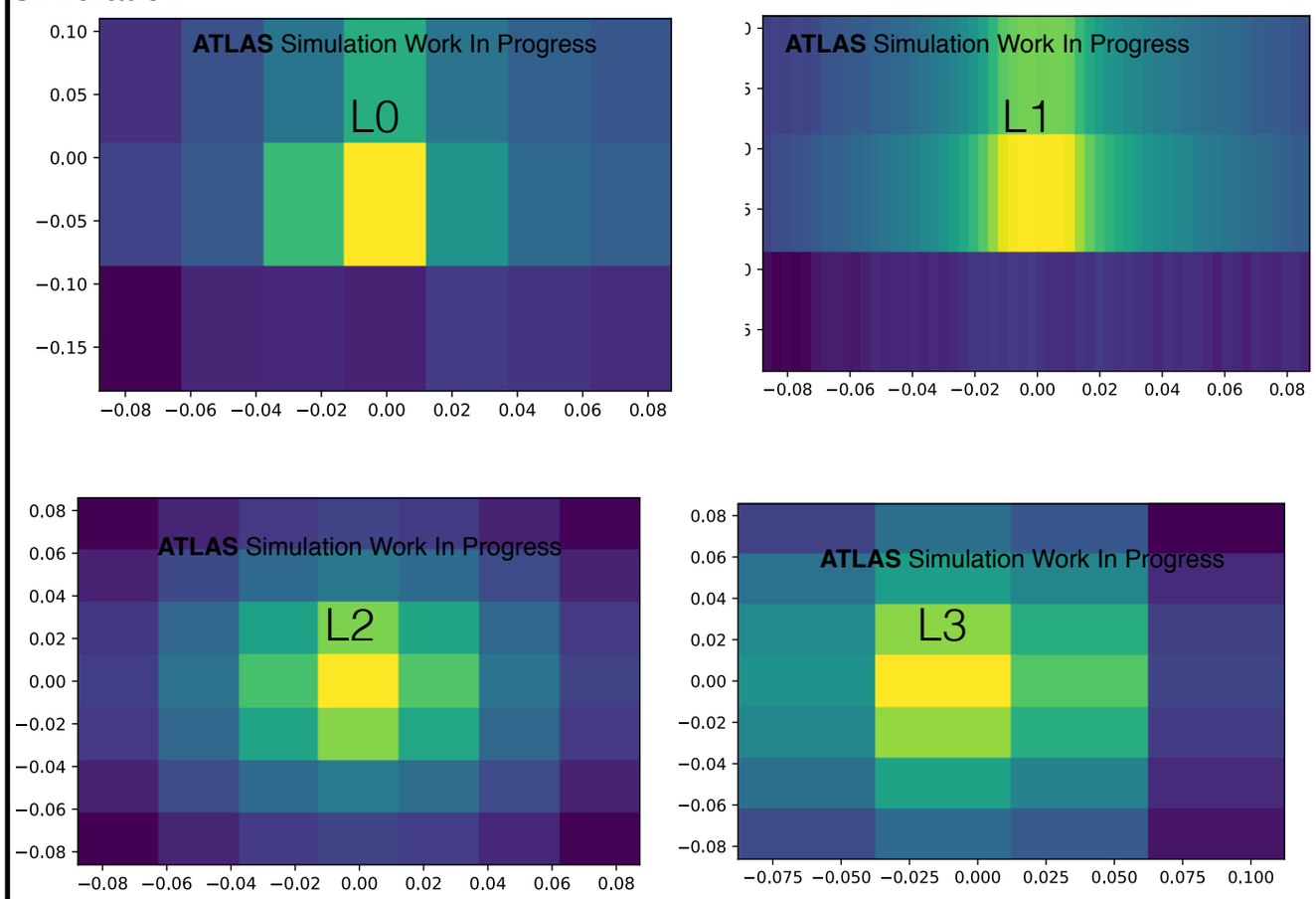
Config 7



ATLAS Simulation

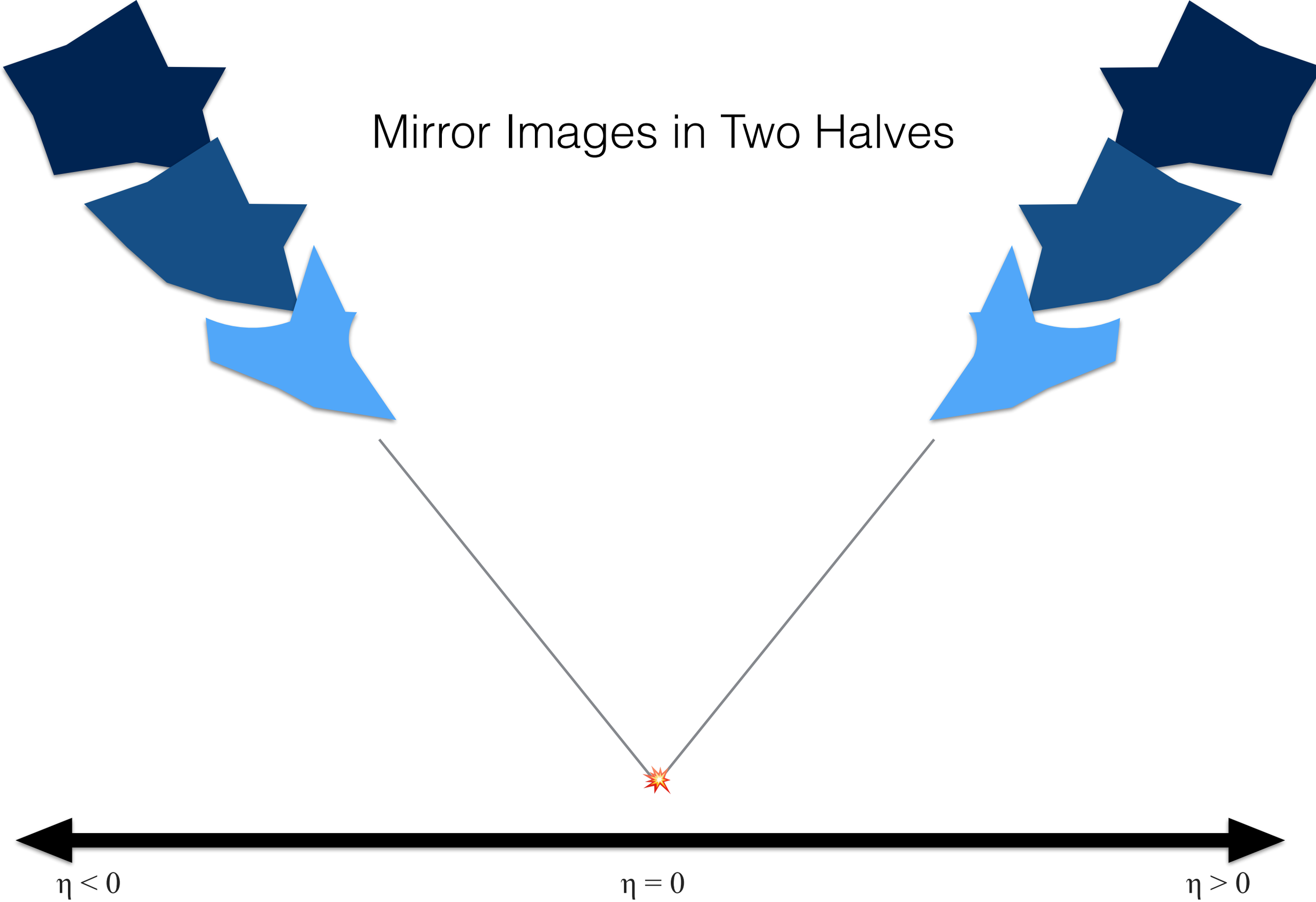


Config 6



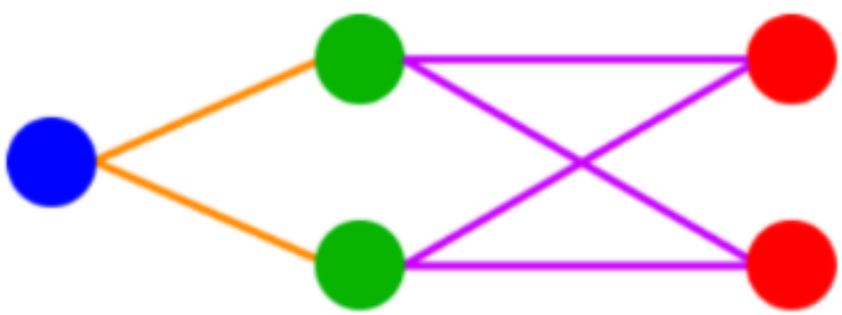
Config 7

Mirror Images in Two Halves



What are Neural Networks?

- Just a lot of linear combinations followed by non-linear activation:



$f_{NN} = \sigma(b_2 + W_2\sigma(b_1 + W_1 x))$

$$\text{Input} : x = \begin{bmatrix} x_{1,1} \\ x_{2,1} \end{bmatrix}$$

$$\text{Weight} : W_1 = \begin{bmatrix} W_{1,1} & W_{1,2} \\ W_{2,1} & W_{2,2} \end{bmatrix}$$

$$\text{Bias} : b_1 = \begin{bmatrix} b_{1,1} \\ b_{2,1} \end{bmatrix}$$

$$\text{Activation} : \sigma(x) = \tanh(x) \ \ \text{(as example)}$$

Activation is applied elementwise!

From Stefan Wunsch

- We can have multiple outputs: