Application of Machine Learning to Fast Detector Simulation

3D Generative Adversarial Networks for High Energy Physics Calorimeter Simulation

Gulrukh Khattak
Supervised by
Federico Carminati
Sofia Vallecorsa

12/4/2019

Overview

- Dataset
- Evaluation
  - Analysis
  - Physics related quantities
  - Optimization function
  - Image Quality tests
- Future Plans
  - Evolutionary Approach
  - Challanges
  - Implementation

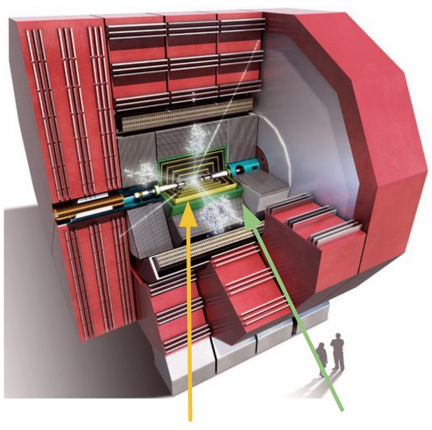
Data set

Energy deposition in calorimeter as pixel intensities for a 3D image

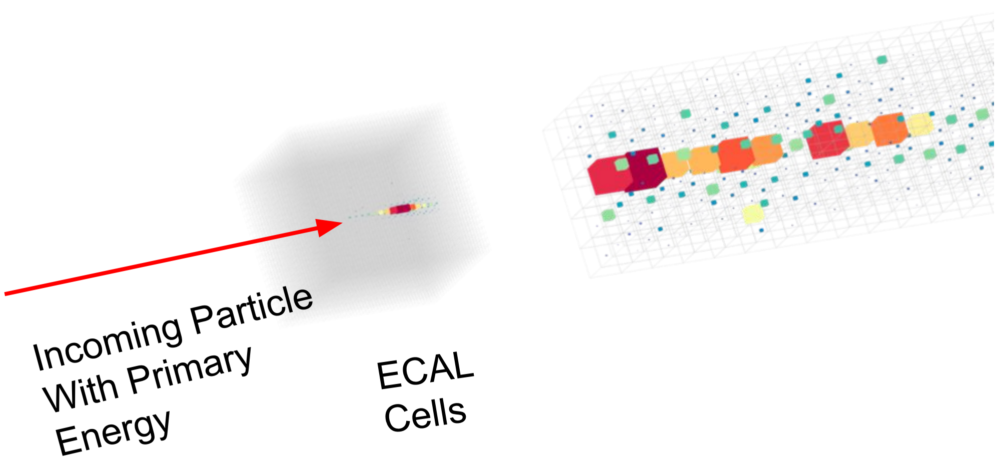
Data set

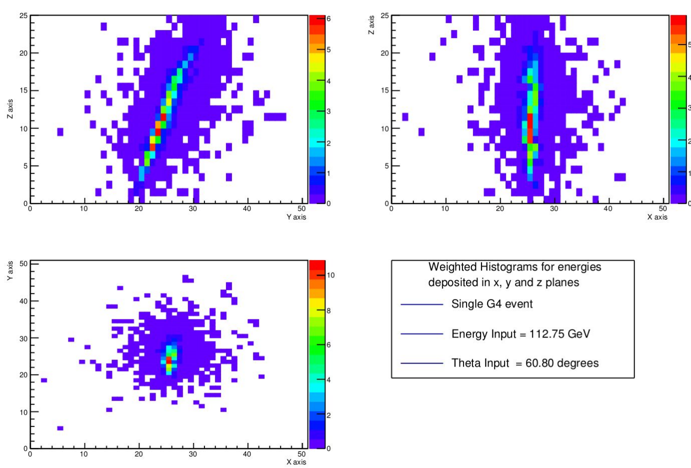
CLIC Calorimeter

- Compact Linear Collider CLIC: **Proposed** linear particle accelerator [6]
- **Open data set** developed for ML applications: Events as selected cells around the barycenter of particle showers
- 200,000 Electron events from 10 to 500 GeV simulated with Geant 4 [7] : Event → 25 x 25 x 25 image → 15, 625 cells
- 120,000 Electron events from 100 to 200 GeV simulated with Geant 4: Event → 51 x 51 x 25 image → 65, 025 cells
- Detector response as 3D images
  - Images are sparse
  - Intensities cover a large spectrum over seven orders of magnitude



http://clicdp.web.cern.ch/





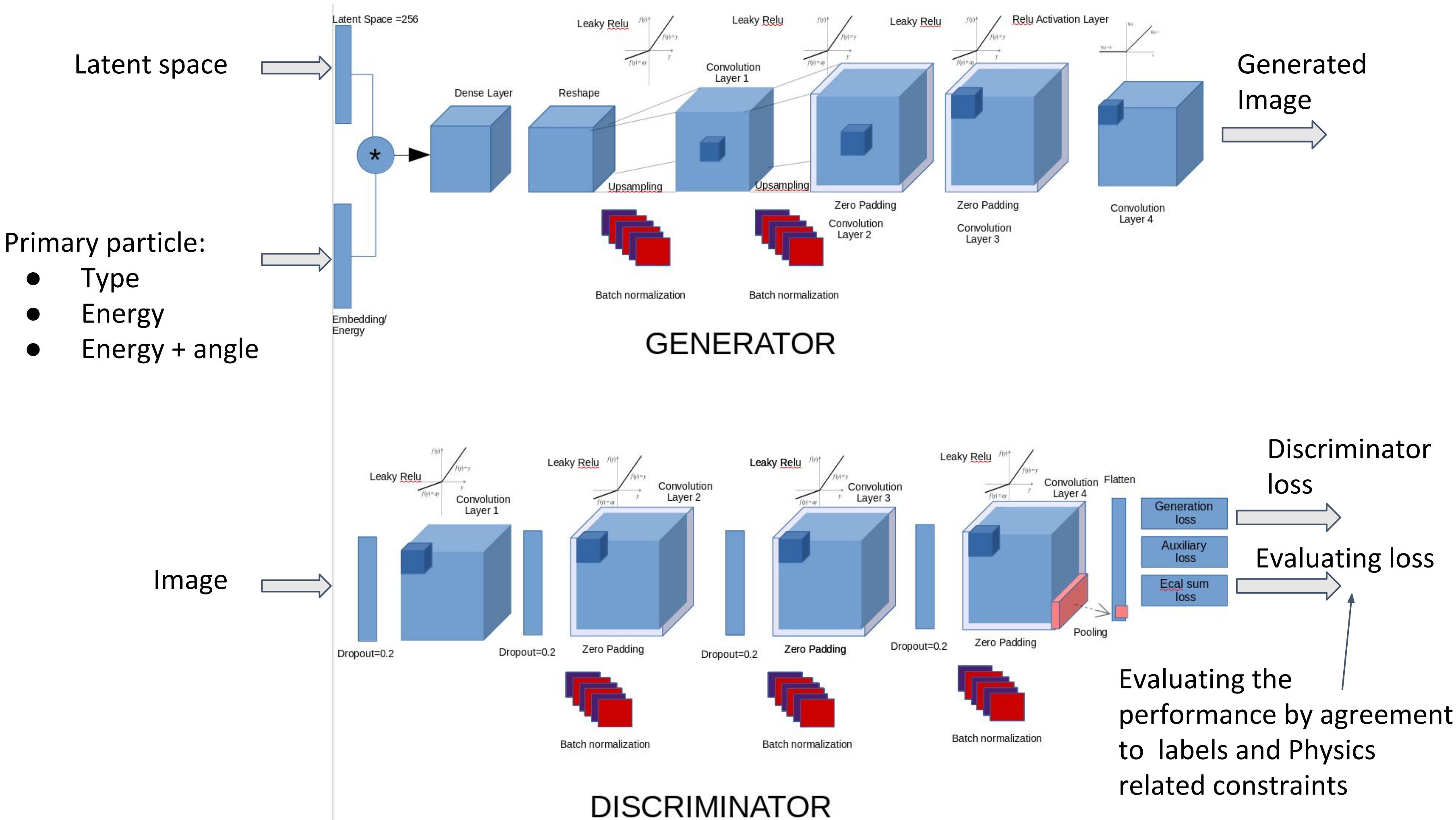
http://cds.cern.ch/record/2254048

3DGAN

*Three dimensional **G**enerative **A**dversarial **N**etworks*

3DGAN

Architecture



Evaluation

*Three dimensional **G**enerative **A**dversarial **N**etworks*

Detailed Analysis

Evaluating and tuning performance

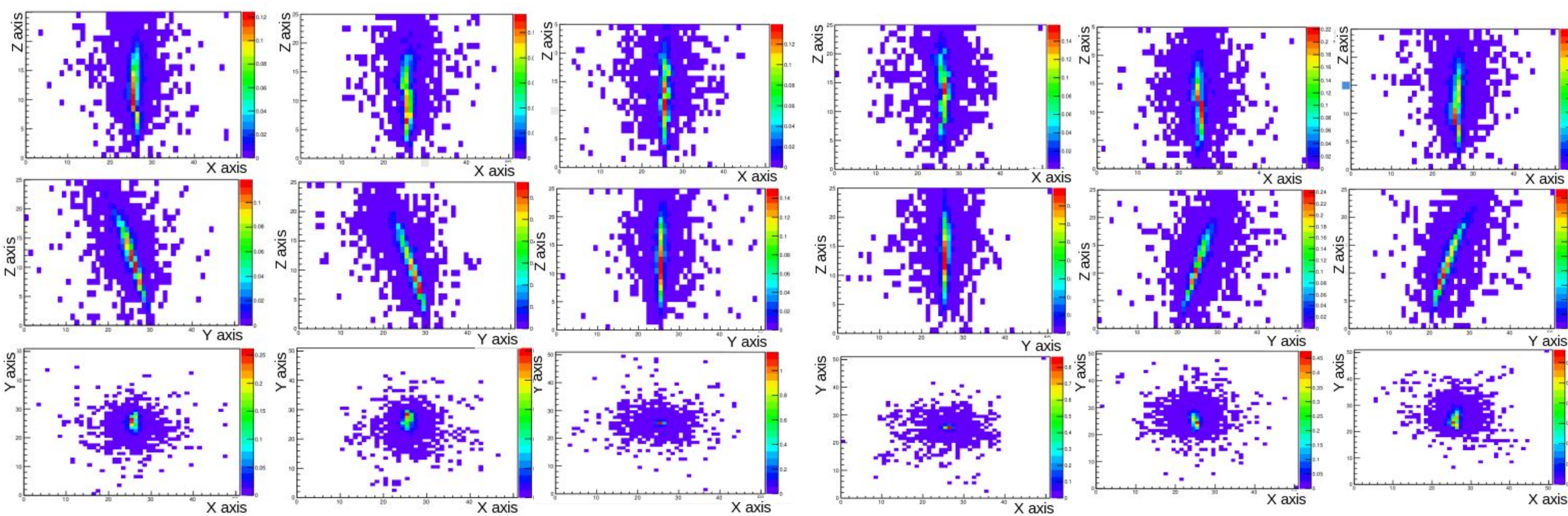
- Multiple criterion
- Detailed GAN vs GEANT4 comparison (More than 200 Plots!) for multiple features:
    - Shower Shapes
    - Sampling Fraction
    - Discriminator Primary Energy regression (dense layer)
    - Position of max energy deposition
    - Hits above a threshold (0.0003 GeV)
    - fraction of energy deposited in different parts of the shower
    - Angle measured from shower
    - Discriminator real/fake probabilities (dense layer)
    - Shower moments
    - Sparsity
    - Cell energy histograms
    - Correlation between different quantities
    - Projections in x, y, z planes to access visually
- Image Quality Analysis

GAN generated events

G4 vs. GAN events 2D histograms for energy deposited in x, y, z planes

- GAN vs. G4 with same Primary Particle energy and angle

111.07 GeV and 115.54 Degrees	147.49 GeV and 87.83 Degrees	188.95 GeV and 62.96 Degrees

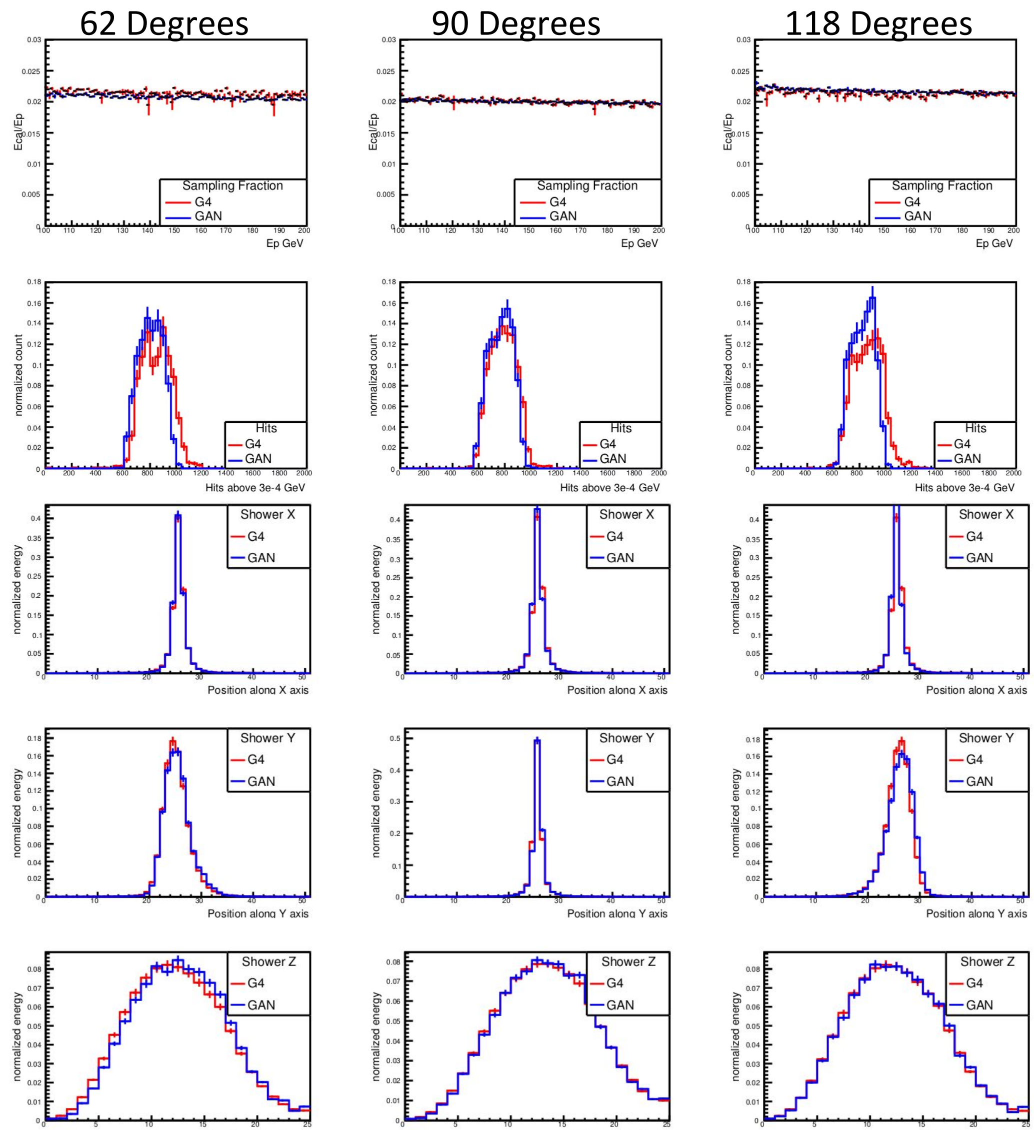


G4 GAN G4 GAN G4 GAN

Physics Quantities

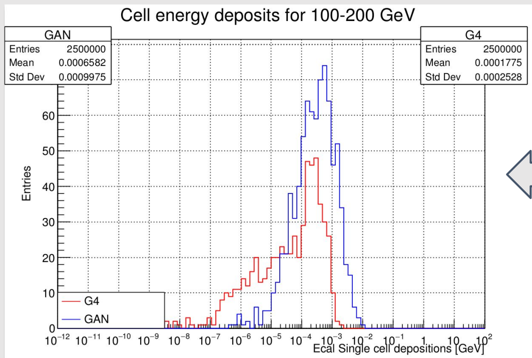
For primary particle energy 100-200 GeV and angle in bins around 62, 90 and 118 Degrees

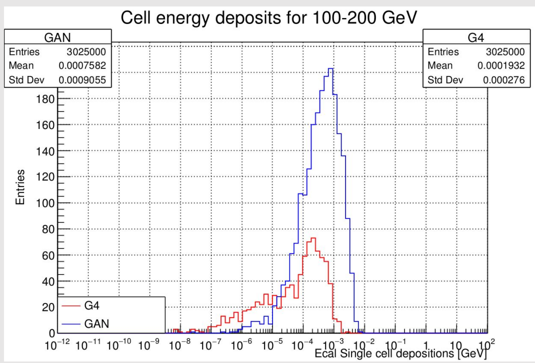
- Sampling Fraction
- Hits
- Shower Shapes:
  - Energy deposited along x, y and z axis

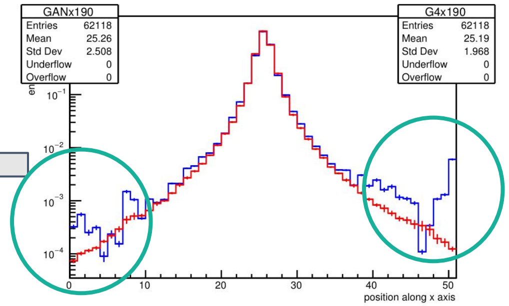


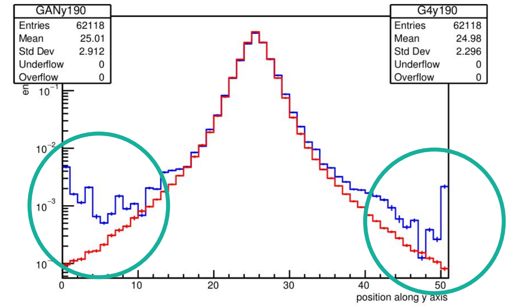
CERN openlab

Disagreement in tails for the shower shapes in transverse direction

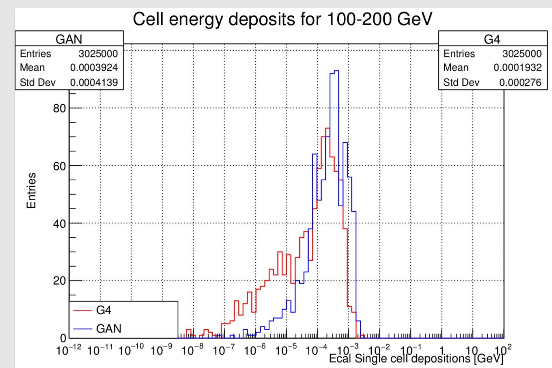


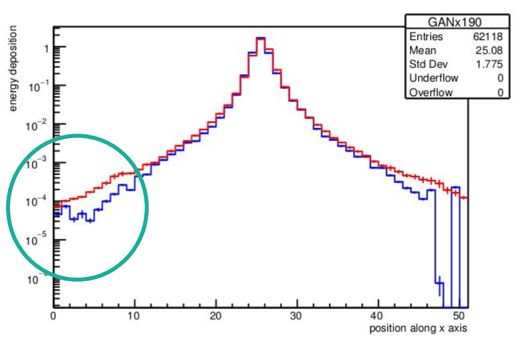


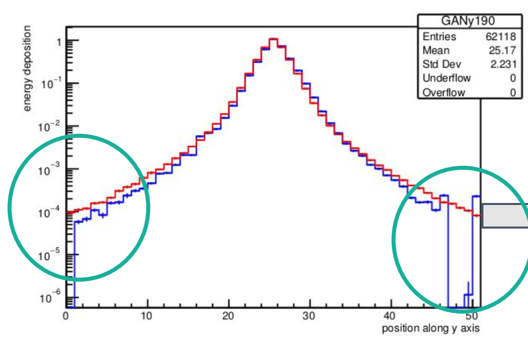


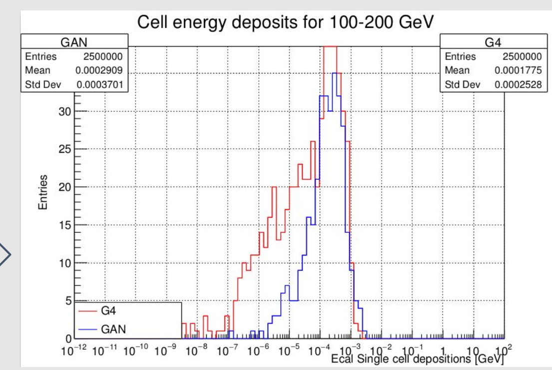


- Much less energy deposited in tails
- The intensity is also lower





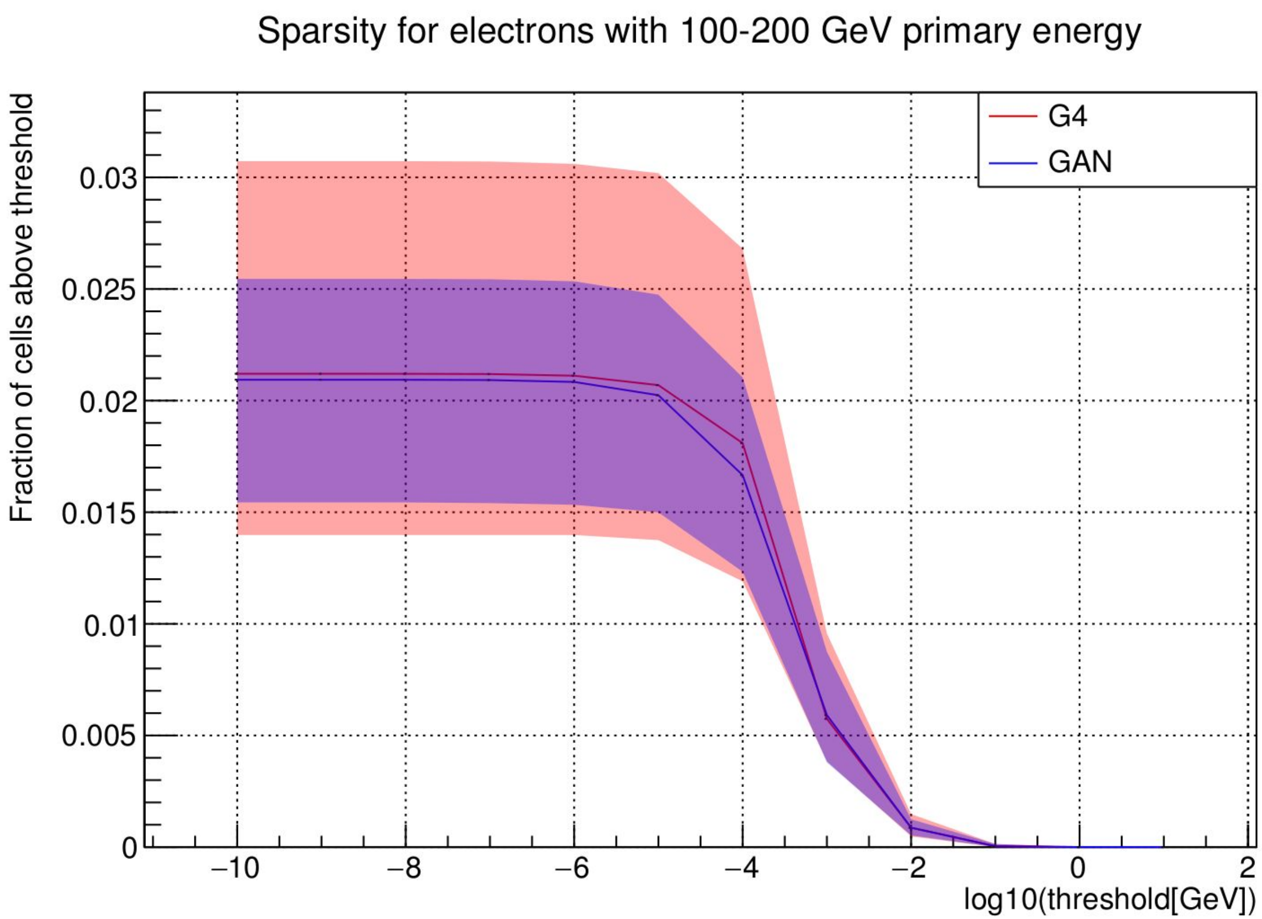




Sparsity

GAN vs. G4

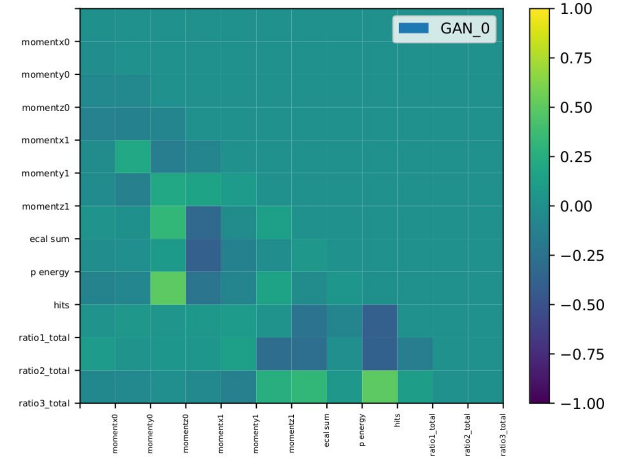
GAN generated images are less sparse but are compatible within error bars

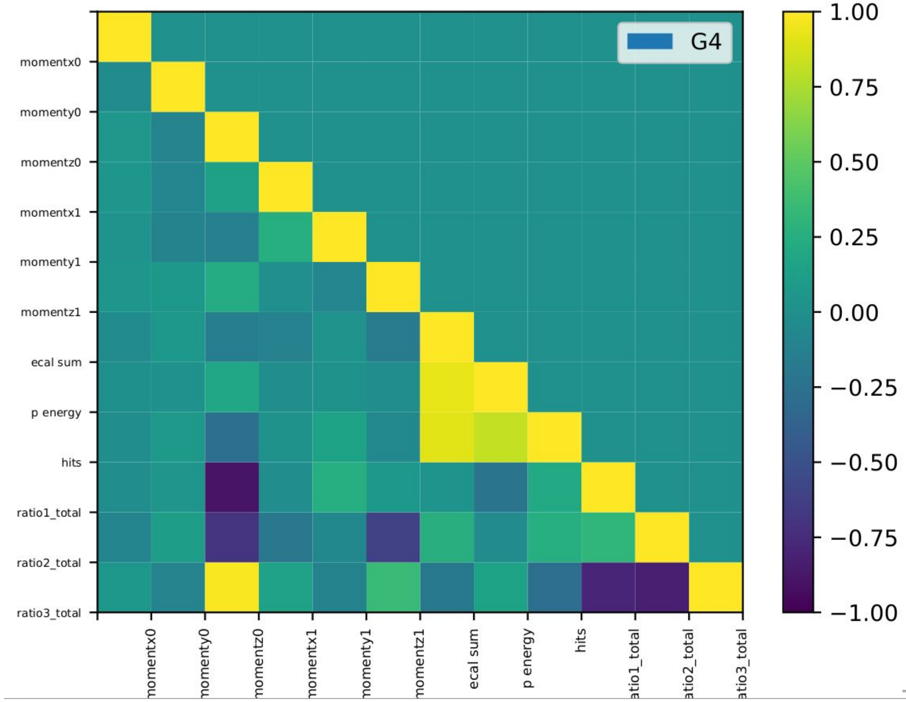


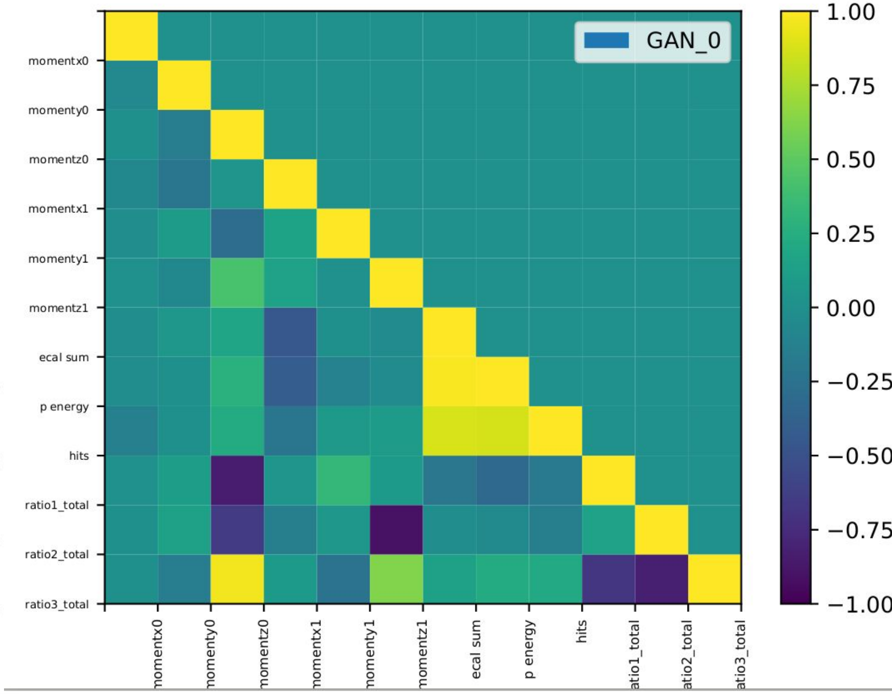
CERN openlab

Correlations

G4 vs. GAN

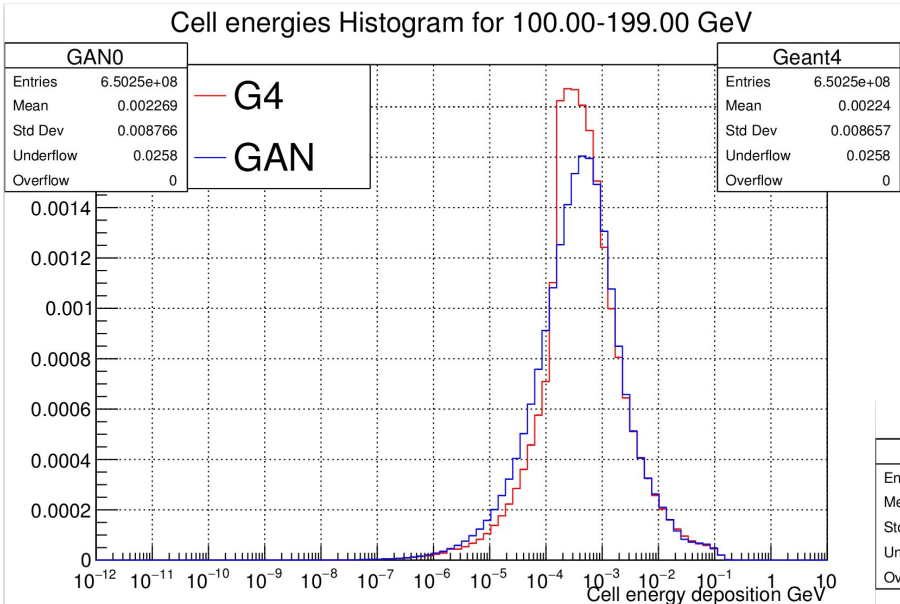


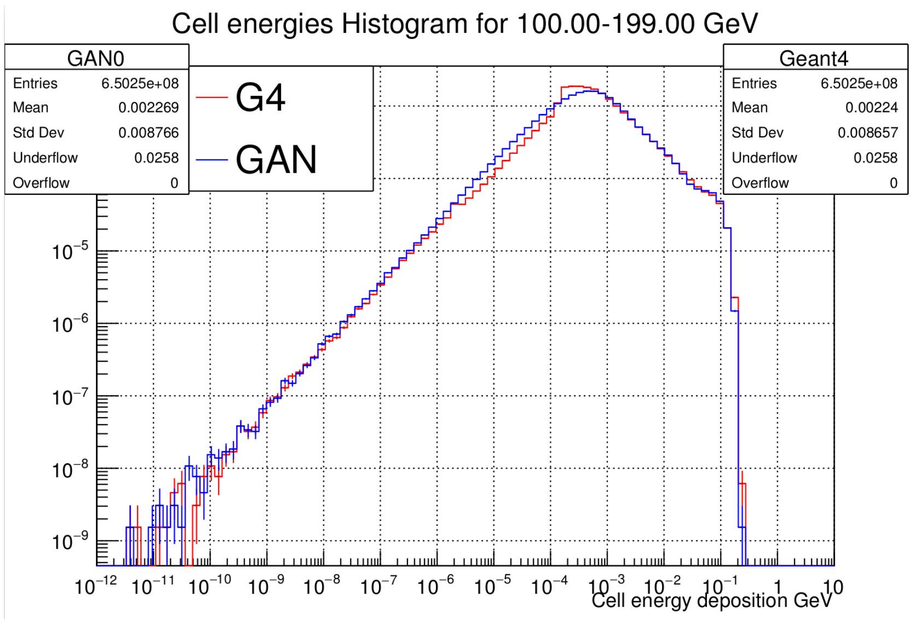




Cell energies

Energy deposited in individual cells





Reasonable agreement to G4 across seven orders of magnitude

CERN openlab

Optimization Function

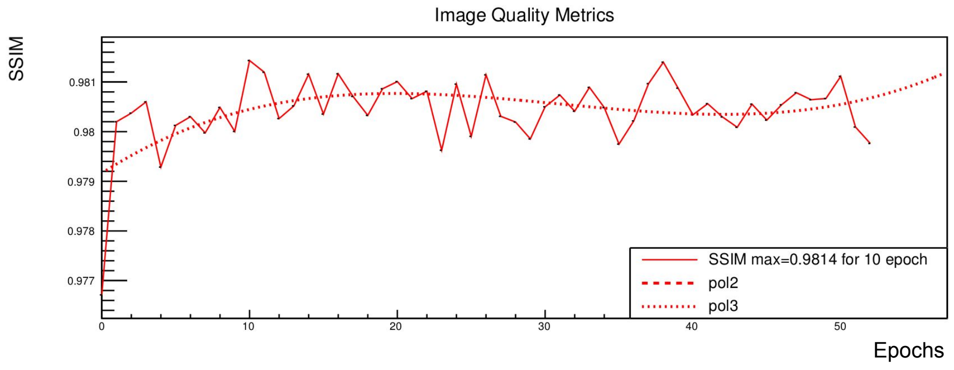
Validation

- GAN is a Minmax game with multiple objectives thus it is very difficult to assess performance on the basis of losses
- A single figure to assess performance
  - Compare results of different hyperparameter settings
  - Assess best weights from training results
- Figure of merit to takes into account mean relative errors for:
  - Shower Shapes
  - Moments
  - Measured angles
  - Sampling Fraction
- Wasserstein Distance between these quantities (work in progress)
  - Wasserstein distance for more than one dimension is NP hard
  - Gromove Wasserstein is being investigated

Image Quality Analysis

Structural Similarity Index, MSCN Coefficients

- Structural Similarity Index or **SSIM [9]** is used to assess **similarity** between images commonly used in denoising applications
- For **GAN** SSIM has also been used to measure **diversity** between generated images
- SSIM was computed for images from same energy and angle bin:
  - SSIM **G4** to **G4** = ~0.983
  - SSIM **GAN** to **GAN** = ~0.981
  - SSIM **G4** to **GAN** = ~0.98



SSIM as training progresses

Future Plans……..

Generalization

Generalize 3DGAN

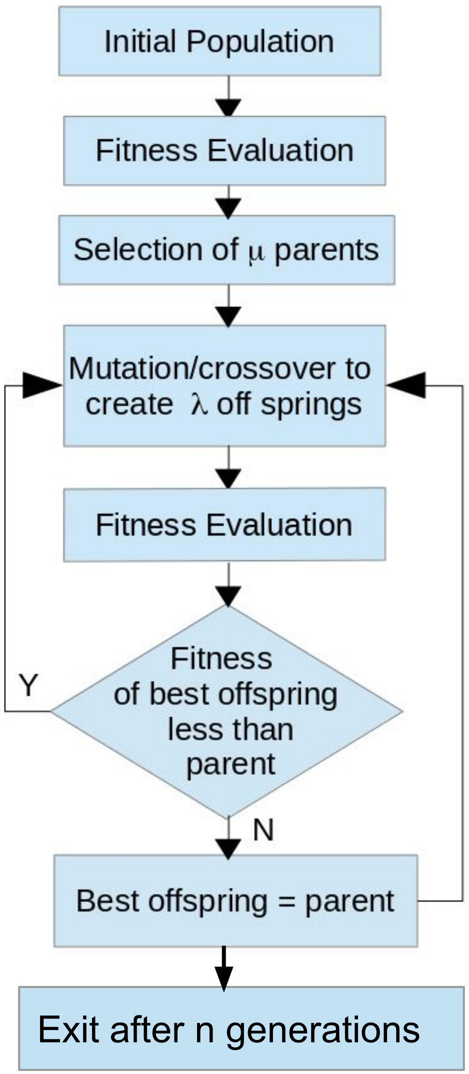
Generalization Challenges

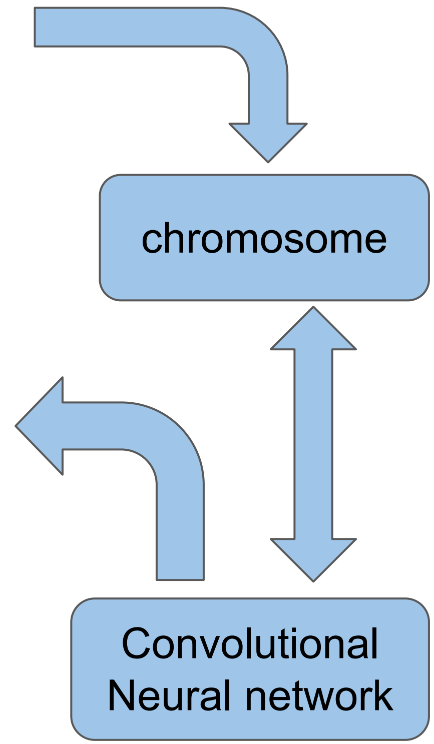
- Generalize the approach so that 3DGAN can be trained and tuned automatically to data from different types of calorimeter
- Challanges
  - Potentially large number of possible configuration
    - Training related Hyper parameters:
      - e.g learning rates, loss weights, optimizer, batch size, latent size etc.
    - Architecture related parameters:
      - Number of layers, number of filters, filter sizes, use of dropout, batch normalization, pooling and upsampling
  - Long training times
- Proposed approaches:
  - Hyper parameter scan using distributed training [11]
  - Training and Optimization at the same time using evolution

Evolutionary Approach

Genetic Algorithm to train and optimize neural networks at the same time

- **Genetic Algorithm** will be used to train a Neural Network
- **Generic Tool** for different detectors
- **Global** instead of local minima
- **Complex and indirect cost** functions are possible
- **Highly Scalable**





CERN openlab

Challanges

GA for GAN

- Network Size:
  - Trainable parameters in **millions** for 3DGAN model
  - Deep GA [12] has been able to train successfully over four million parameters
- Big Data:
  - Large data for evaluation on entire training data.
  - **Batch training** has been successfully implemented LEEA [13].
- Required Resources:
  - Due to network and data sizes **HPC resources** will be essential.
- Architecture Optimization:
  - Design mechanism to offer greater flexibility
  - **Non conventional connections** between convolutional filter nodes EXACT [14].
  - **Indirect Weight encoding** similar to HyperNEAT [15]
- Adversarial Training:
  - Simultaneous training of **two networks** competing with each other
  - Evolutionary approach is frequently used in **multiplayer games**
- Inexact solution:
  - Evolutionary approach can discover **neighbourhood of global optima** but cannot arrive at **exact solution** of a local search
  - A **hybrid approach** can incorporate SGD as a callback after no improvement for predefined number of generations to help arrive at exact solution.

Implementation

Step by step implementation

- This approach will be implemented in three phases:
- Initially we will reduce the problem complexity and focus on genetic training by:
  - Converting 3D data to 2D images, slicing the detector volume along the direction of particle propagation.
  - focusing on the discriminator network only
  - Fixed architecture
- During a second phase architecture hyper-parameters will also be encoded in chromosomes and optimised
  - Test indirect encoding for weight encoding to reduce the number of parameters and stabilize training.
- Finally the complete GAN scenario will be implemented.

References

[1] Weinzierl, S. (2000). Introduction to Monte Carlo Methods. Website http://arxiv.org/abs/hep-ph/0006269 (Accessed 27.10.18)

[2] The Wordwide LHC computing grid, https://home.cern/about/computing/worldwide-lhc-computing-grid, accessed 27.10.18.

[3] I. J. Goodfellow, "On distinguishability criteria for estimating generative models," ArXiv e-prints, Dec. 2014

[4] A. Odena, C. Olah, and J. Shlens, "Conditional Image Synthesis With Auxiliary Classifier GANs," ArXiv eprints, Oct. 2016

[5] J.Wu, C. Zhang, T.Xue, W.T.Freeman, J.B.Tennen Baum, "Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, NIPS, 2016

[6] CERN Accelerating Science, https://home.cern/ , accessed 27.10.18.

[7] S. Agostinelli et al. GEANT4: A Simulation toolkit. Nucl. Instrum. Meth. , A506:250–303, 2003.

[8] I. J. Goodfellow, et al. Generative Adversarial Networks. ArXiv e-prints , June 2014.

References

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: from error visibility to structural similarity “. *Trans. Img. Proc.* 13, 4 (April 2004), 600-612. DOI=http://dx.doi.org/10.1109/TIP.2003.819861

[10] L. Song, C. Chen, Y. Xu, G. Xue and Y. Zhou, "Blind image quality assessment based on a new feature of nature scene statistics," *2014 IEEE Visual Communications and Image Processing Conference*, Valletta, 2014, pp. 37-40.

[11] S. Vallecorsa, D. Moise, F. Carminati. IEEE HiPC 2018.

[12] F. Petroski Such. et. al. “Deep Neuroevolution: Genetic Algorithms Are a competitive Alternative for Training Deep Neural Networks for Reinforcement Learning”. 2018. arXiv:1712.06567

[13] G. Morse, K. O. Stanley. “Simple Evolutionary Optimization Can Rival Stochastic Gradient Descent in Neural Networks”. GECCO 2016

[14] T. Desell. “Large Scale Evolution of Convolutional Neural Networks Using Volunteer Computing”. IEEE eScience. 2017.

[15] K. O. Stanley, D. D’ Ambrosio, J. Gauci. “A Hypercube-Based Indirect Encoding for Evolving Large-Scale Neural Networks” Artificial Life Journal, MIT Press, 2009.

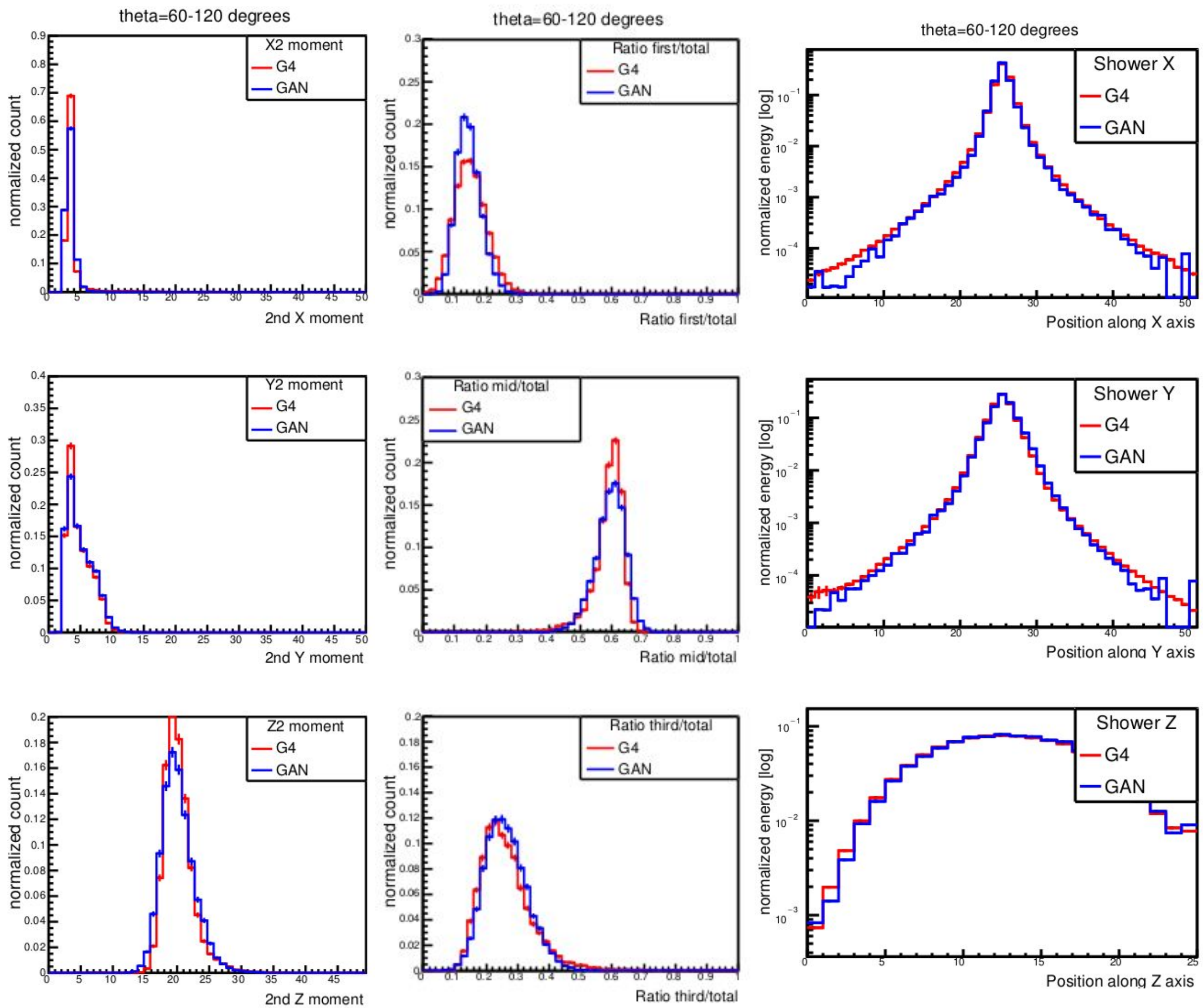
Thank you !!!

Bonus Slides

Physics Quantities

Primary particle energy 100-200 GeV and angle 60-120 degrees

- shower width
- Fraction of energy deposited in first, second and third part of shower to total
- Shower shapes in log



CERN openlab

Indirect Weight Encoding

Evolving Filters

- Convolutional Neural networks use the idea that filter kernels can be used to extract interesting features from an image.
- Digital filters can perform operations like smoothing and edge detection etc.
- Image is convolved with a filter kernel
- The weights in a kernel are meaningful in their relation with other weights and its position
- If the filter size is changed the trained weight is not meaningful any more
- Thus instead of direct coding, indirect coding can be used. Thus the underlying relation of weights will be learnt instead of fixed values
- **Hyper Neat** uses indirect coding for neural network evolutionary training as Compositional Pattern Producing Networks **CPPNs**
- Same concept can be extended to CNN

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

−1	−1	−1	−1	0	1
0	0	0	−1	0	1
1	1	1	−1	0	1

Prewitt

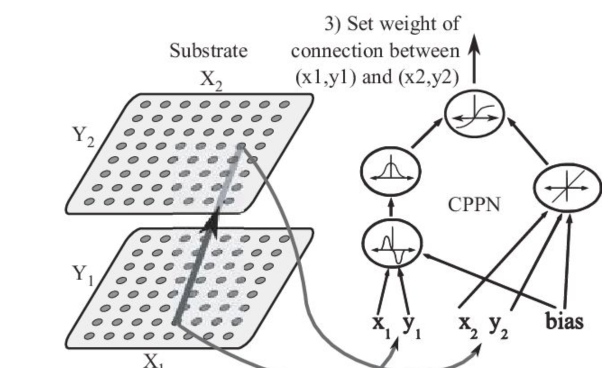
−1	0	0	−1
0	1	1	0

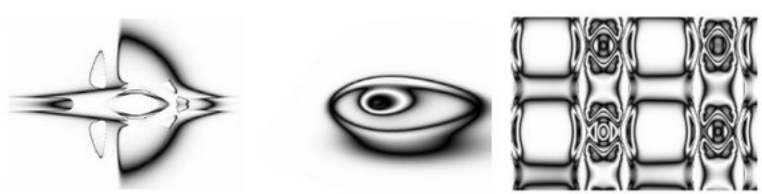
Roberts

−1	−2	−1	−1	0	1
0	0	0	−2	0	2
1	2	1	−1	0	1

Sobel

Some commonly used edge detection filters





(a) Symmetry (b) Imperfect Symmetry (c) Repetition with Variation

HyperNEAT CPPN

Asynchronous Update

Genetic Training

- A pool of random parent chromosomes is created
- Fitness for each parent is evaluated
- Offsprings are obtained (mutation and/or cross over)
- Each thread evaluates a single child
- As soon as child is evaluated the fitness is compared with lowest fitness of current pool
    - If child fitness is same or higher the child replaces the parent with lowest fitness
- Different children can take different time for evaluation
- No need for communication between threads
- There will be no loss in performance with number of threads