
CERN

Generative Models in High Energy Physics

Fast Simulation examples

Sofia Vallecorsa

CERN openlab AI and Quantum Research

CERN IT Department

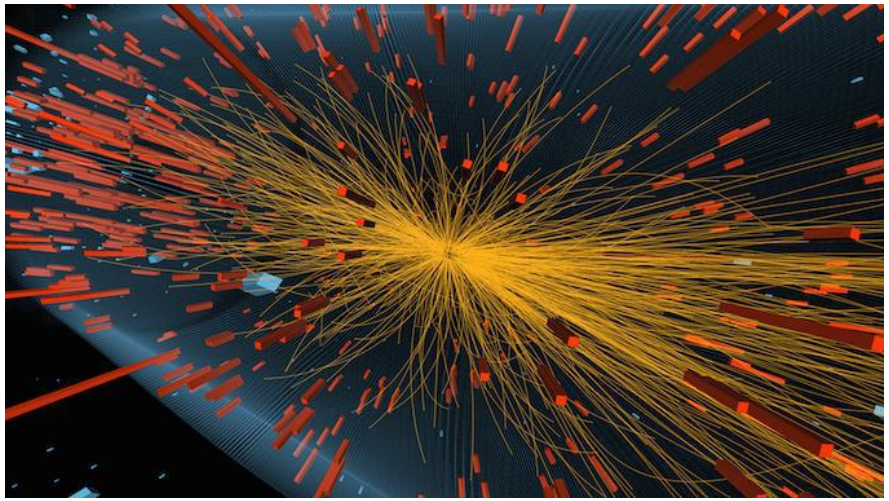
Content

- Introduction
- Architecture Trends
- Results Validation
- Computing Resources
- Summary and Conclusion

... a short teaser !

Deep Generative Models

- **Shallow** models learn **simple** internal representations
- **Deep** models allow higher levels of **abstractions** and improve **generalisation**
- **Multiple applications**
  - **Simulation**
  - Anomaly Detection
  - Data manipulation
- **Different use cases have different requirements:**
  - Fast inference
  - Real time training capability
  - Fast training for large optimizations



Simulation applications

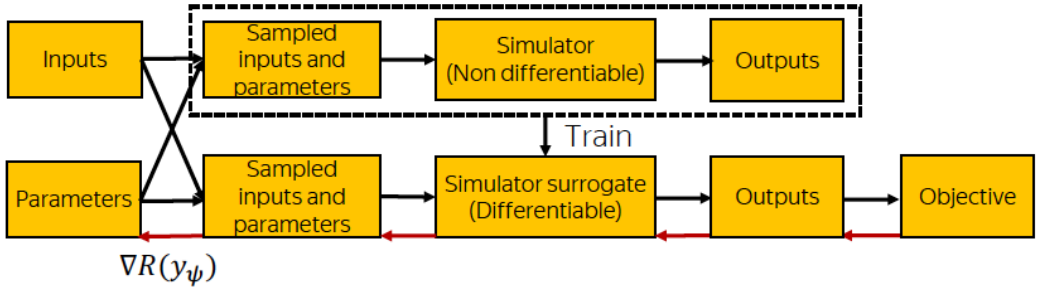
S. Shirobokov *et al., Workshop on Real World Experiment Design and Active Learning at ICML* 2020, arxiv:2002.04632

Fast detector simulation ..

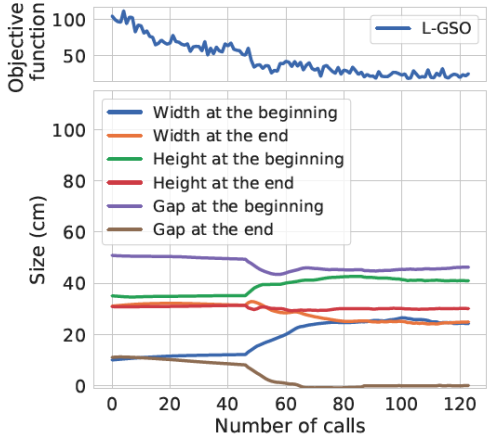
- Examples in ATLAS, CMS, LHCb, ALICE.
- Mostly calorimeters, but also RICH, TPCs..

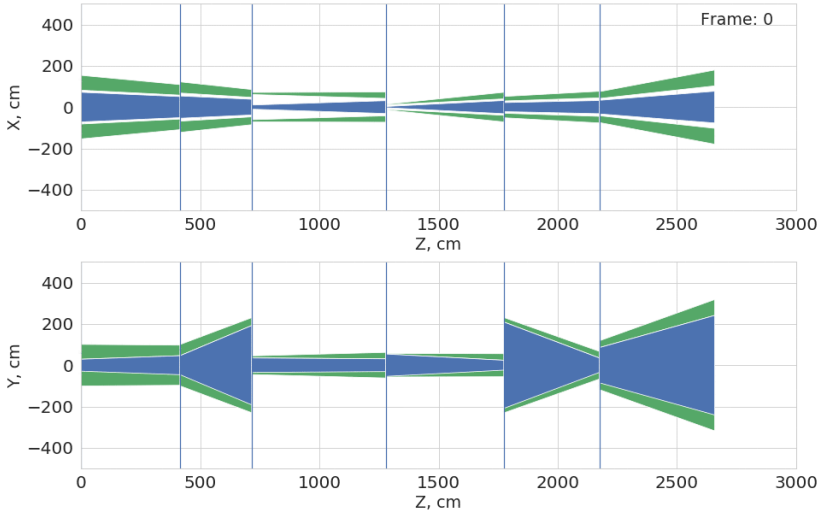
but also:

- Detector design
- Optimisation
- Domain Adaptation
- Monte Carlo agreement to real data



Magnet system for muon beam optimisation

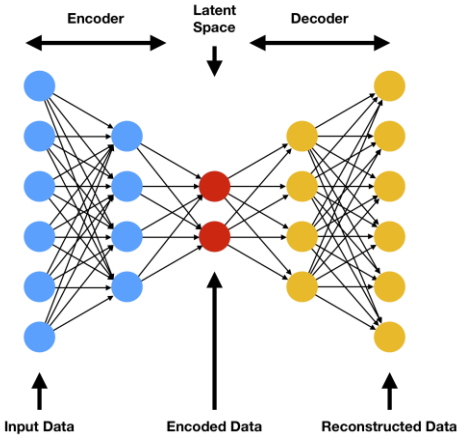


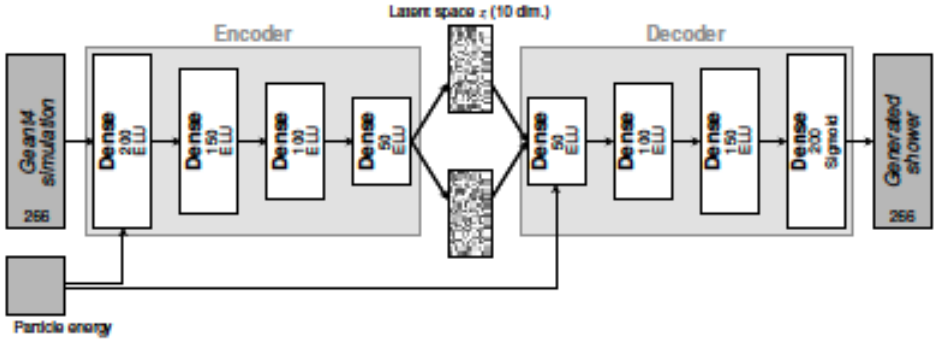


Multiple Architectures

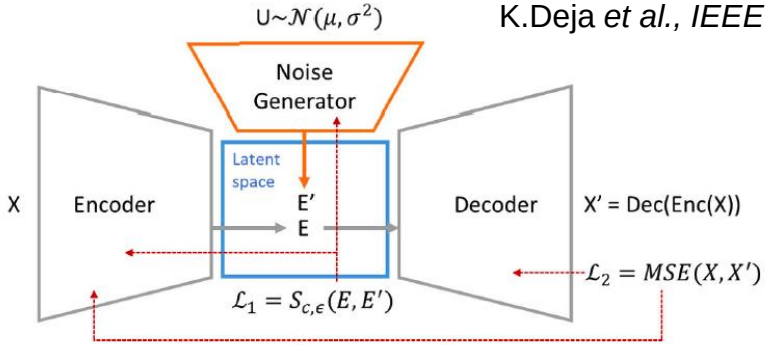
Variational AutoEncoders

- Variational (KL Divergence), arxiv:1312.6114
- Wasserstein (MMD), arxiv:1711.01558
- Sinkhorn arxiv:1810.01118

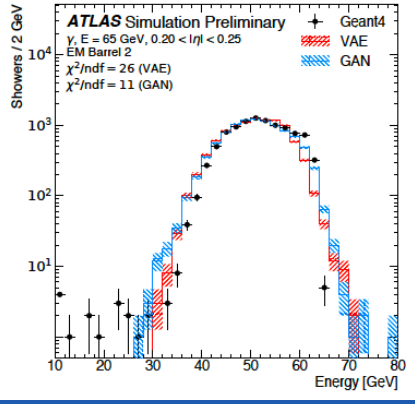


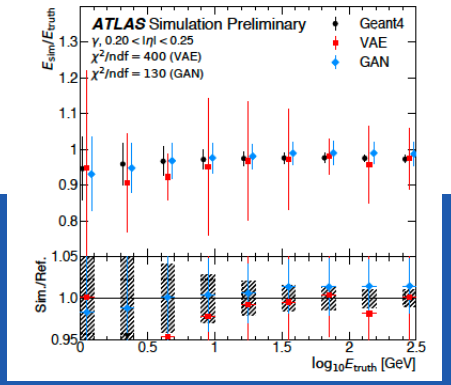


A. Ghosh,*Journal of Physics: Conference Series*. Vol. 1525. No. 1. IOP, 2020.



K.Deja *et al., IEEE Access* 2020.

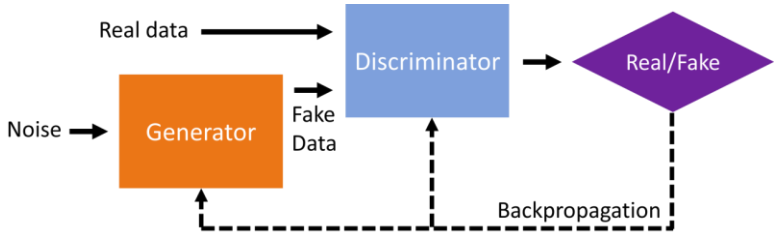


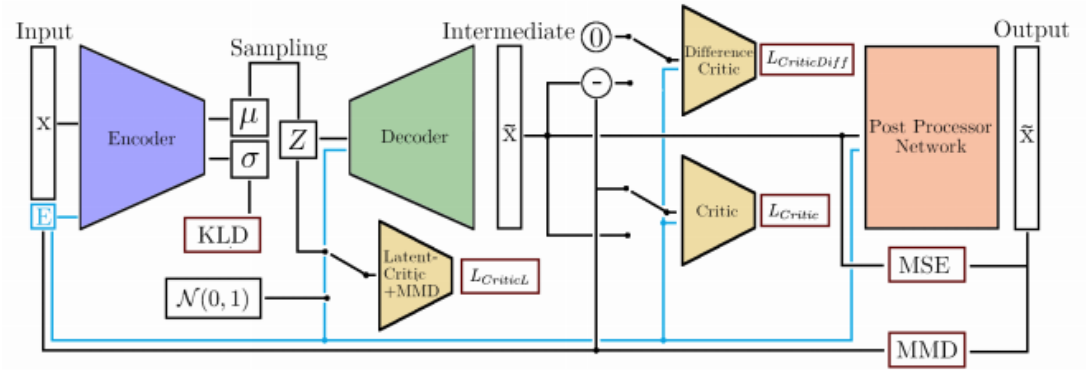


	MNIST		HEP	
Model	FID	Sinkhorn	MAE	Sinkhorn
cond VAE	6.61	30.13	23.13 (±65.53)	14.92
cond WAE (MMD)	34.73	30.44	43.54 (±55.23)	34.46
cond e2eSAE (ours)	**4.11**	**24.92**	**13.50 (±29.82)**	**7.91**
cond DCGAN	0.93	22.23	68.27 (±180.45)	6.95
original data	0.33	0	6.59	2.89

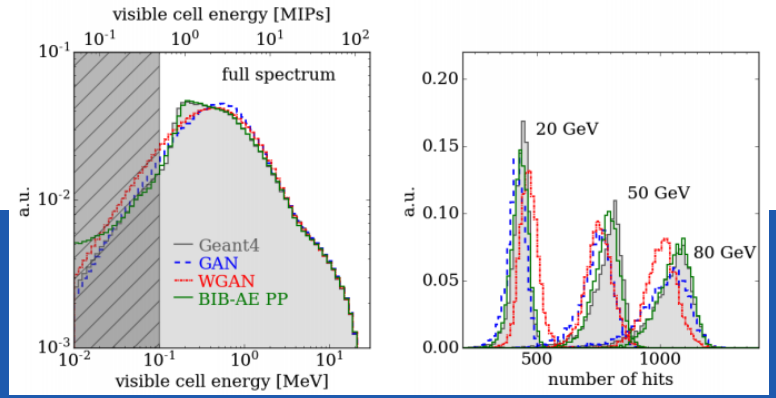
Generative Adversarial Networks

- Vanilla, Conditional, Auxiliary Classifier GAN
- Wasserstein, Cramer (arxiv:1704.00028)
- Complex topologies (ex. BiGAN, BIB-AE)

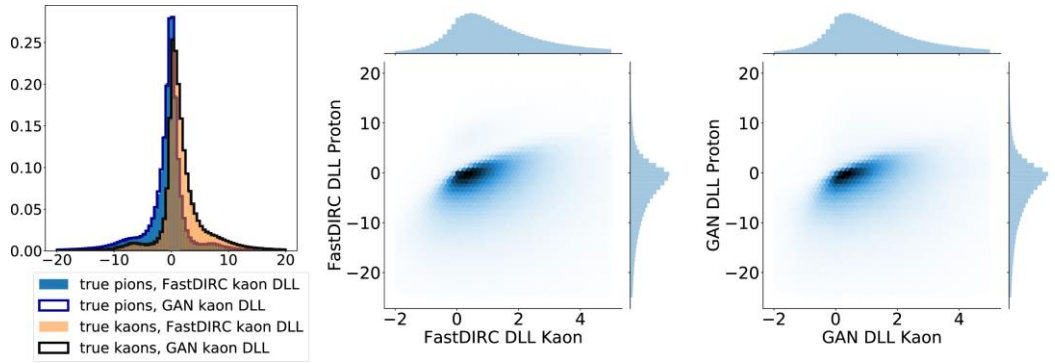




E. Buhman *et al.*, *arxiv:2005.05334*.



Wasserstein GAN for Cherenkov detector simulation:
D. Derkach *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 952 (2020): 161804.



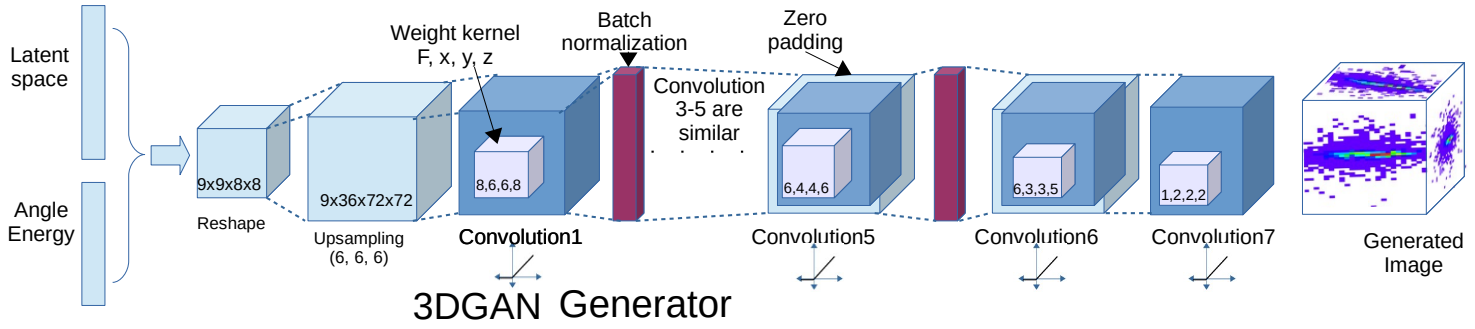
Detector response as images

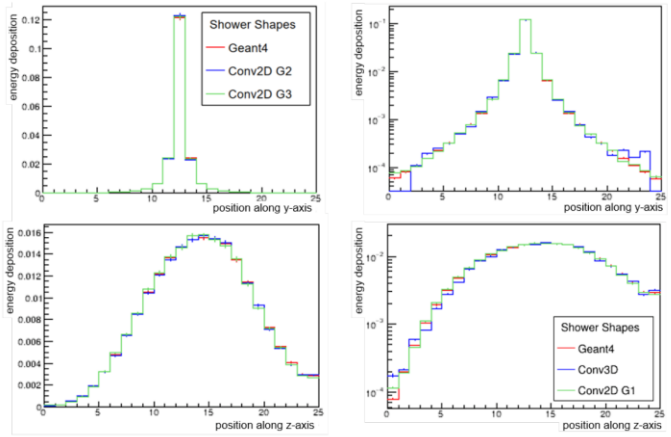
G. Khattak *et al.*, ICMLA 2019

F. Rehm *et al.*, AAAI 2020, arXiv:2103.13698

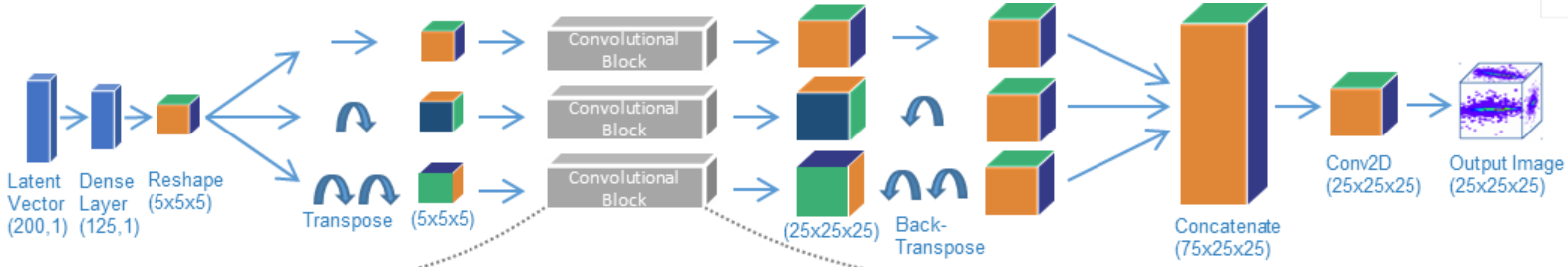
Convolutional layers for 51x51x25 pixels image: sparse, large dynamic range

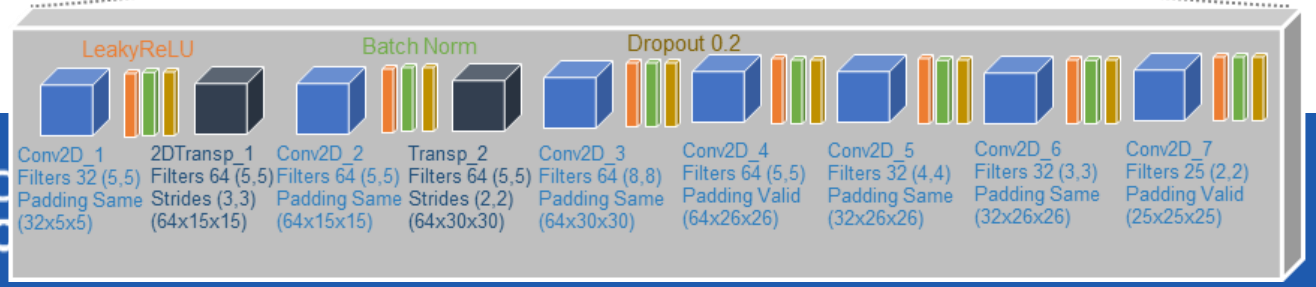
3 Conv.

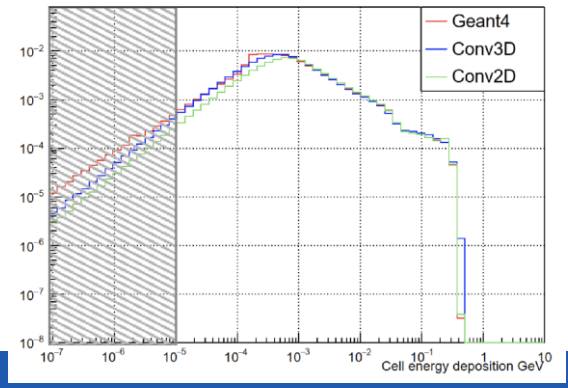




2D Conv:
2x faster &
Lower mem footprint!







Detector response as graphs

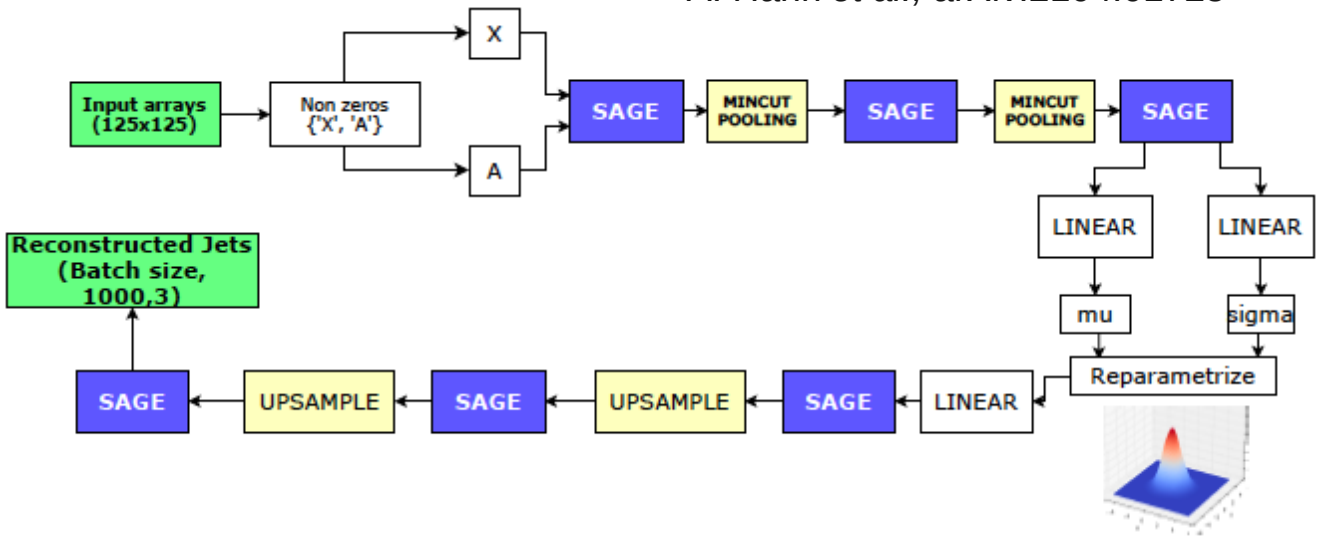
Calorimeter energy deposits as graphs

Data as a **graph of connected hits**

Connect hits using **geometric** constraints

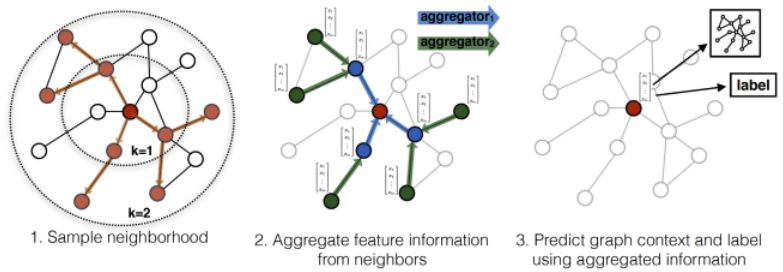
Embedding requires **large graphs** (~10^5 nodes)

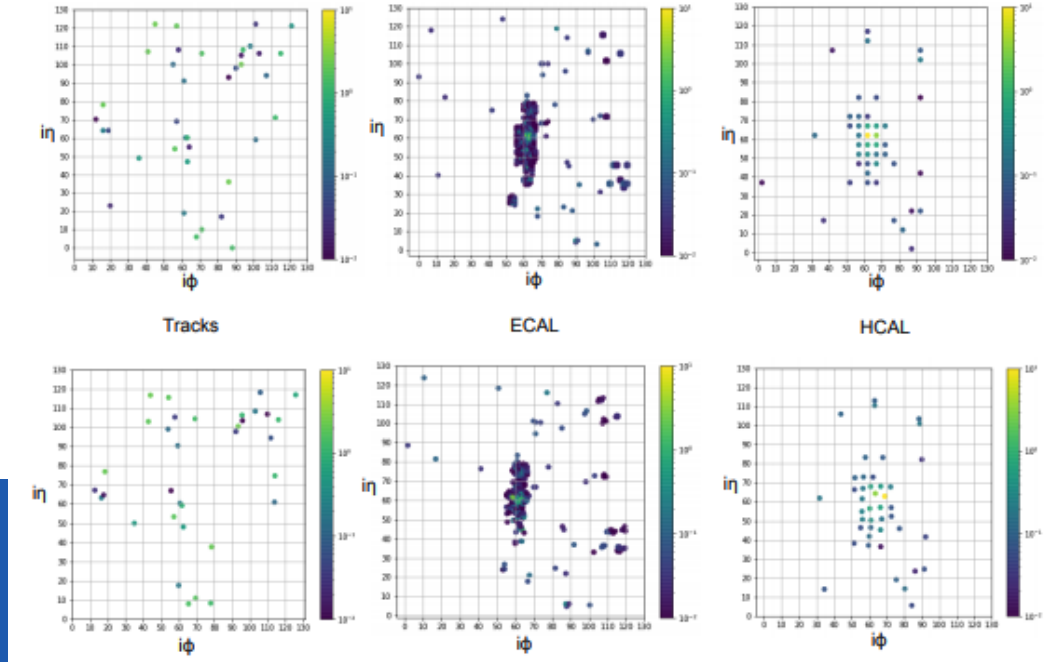
A. Hariri *et al.*, arXiv:2104.01725



GrapSAGE: W. Hamilton, R. Ying, J. Leskovec, NIPS (2017), arXiv:1706.02216:

Inductive approach to generate embedding for unseen nodes





Detector Size..

FastCaloGAN: full calorimeter fast simulation using 300 GANs

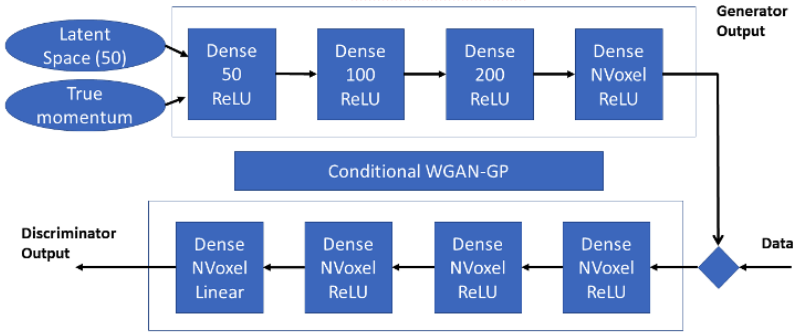
100 η slices (0.05 wide)

- 15 energy points
- Electrons, Photons, Pions

Generated hits in Athena

The GAN

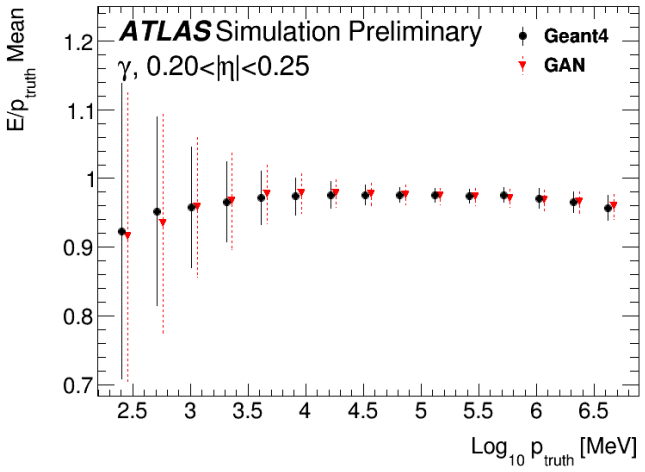
- Train the GANs on voxelised hits

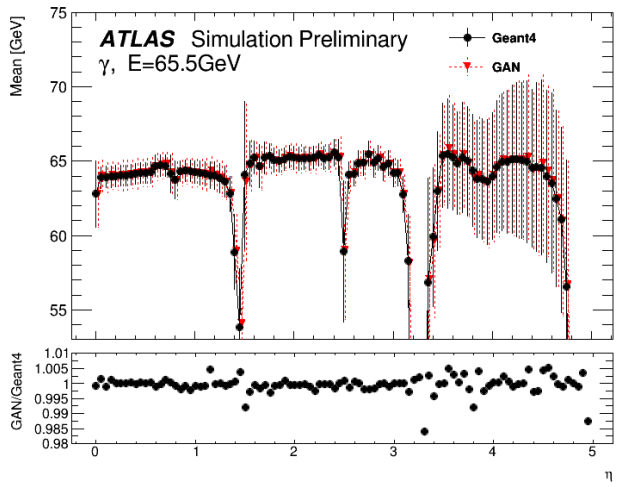


G	50 (Input latent Space), 50, 100, 200, NVoxel (pid and η dependent)
D	NVoxel, NVoxel, NVoxel, NVoxel, 1
Activation function	ReLU (in all layers)
Optimiser	Adam [21]
Learning Rate	10^{-4}
β	0.5
Batchsize	128
Training ratio (D/G)	5
Gradient penalty λ	10

NVoxel depends on PID and η

06/04/2021 Michele Faucci Giannelli 3





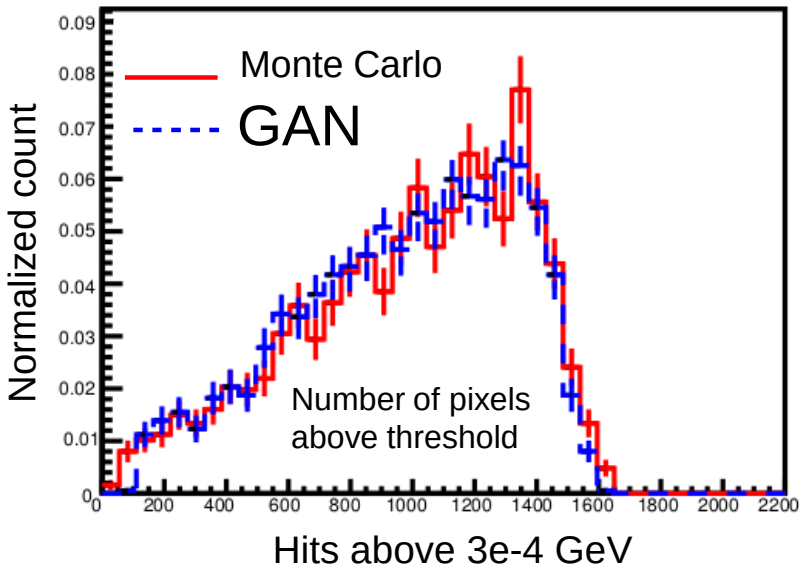
Need to take **training time** into account!

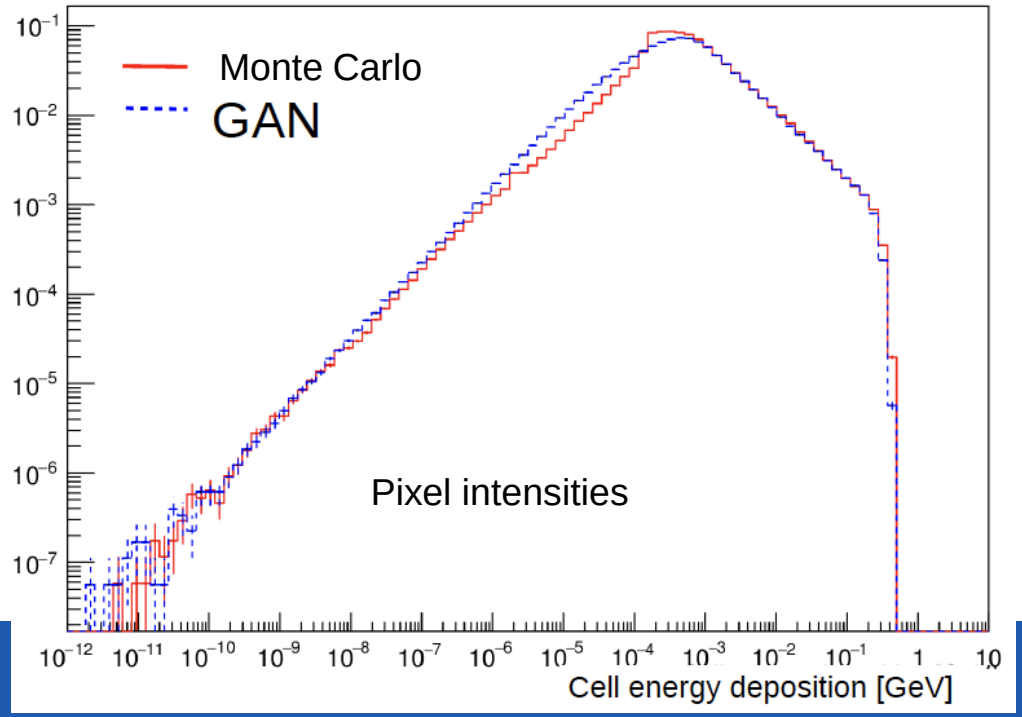
Results validation

Physics validation

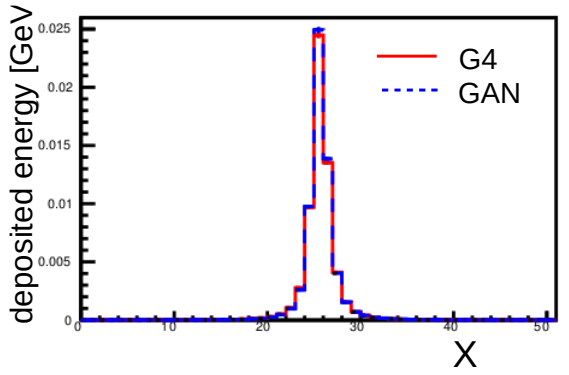
- Compare GAN images against Monte Carlo
- **Depending on the application need few percent accurate representation** of all relevant physics variables

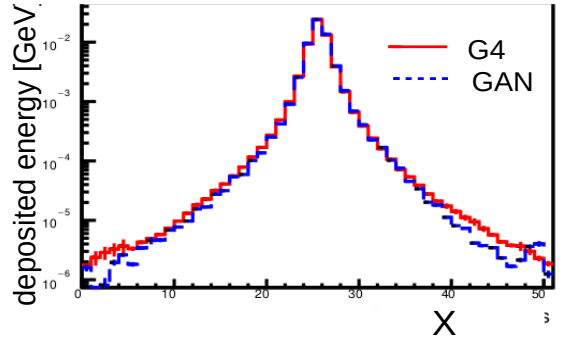
G. Khattak *et al.*, ICMLA 2019

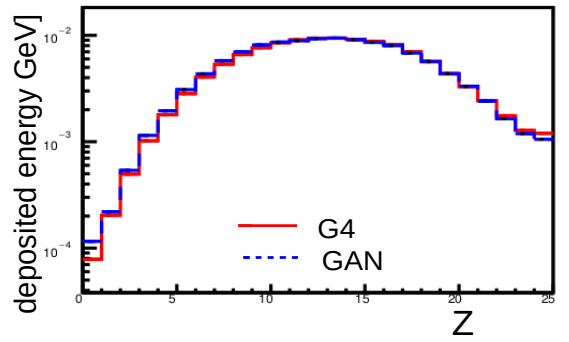




Energy patterns along detector axes





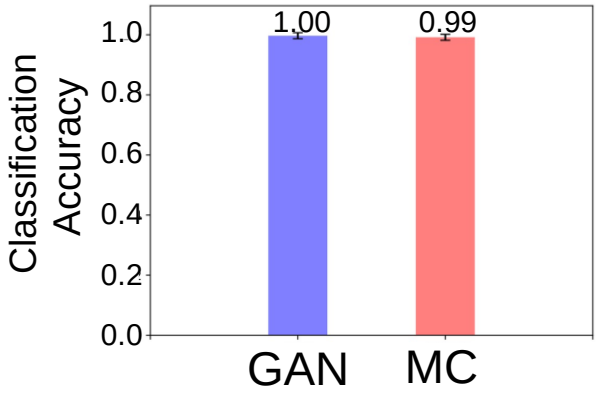


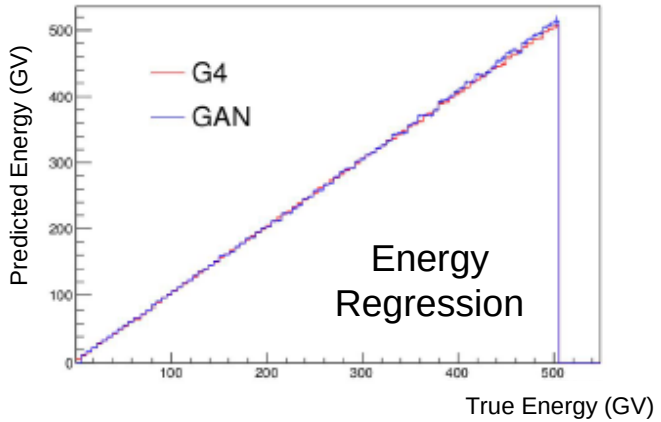
Physics validation (II)

G. Khattak *et al.*, ICMLA 2019

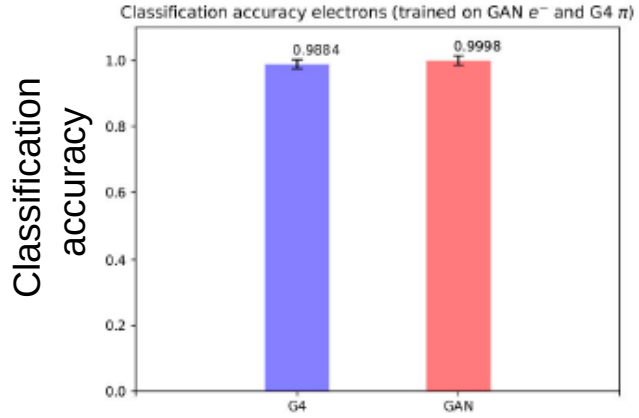
- **Triforce*** DNN has been developed to distinguish different kind of particles and measure their energy

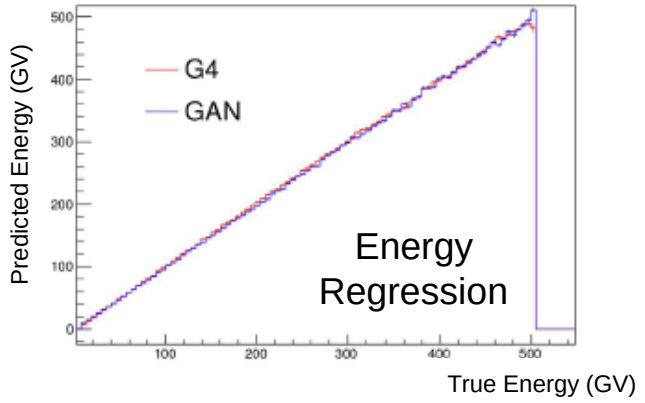
Trained on G4





Trained on GAN





*D. Belayneh et al.,"Calorimetry with deep learning: Particle simulation and reconstruction for collider physics," 2019, https://inspirehep.net/literature/1770936

Systematic effects and interpretability

Systematics: image similarity

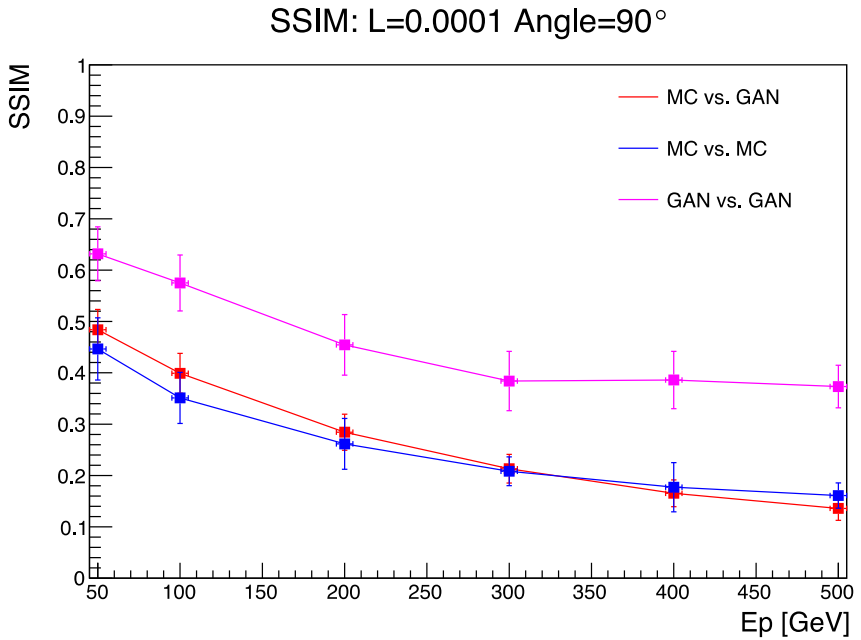
GAN can exhibit **mode-collapse** or **mode-drop**

How much **diversity** in the generated sample?

- Use the **Structural Similarity Index**

$$\mathrm{SSIM}(\boldsymbol{x},\boldsymbol{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

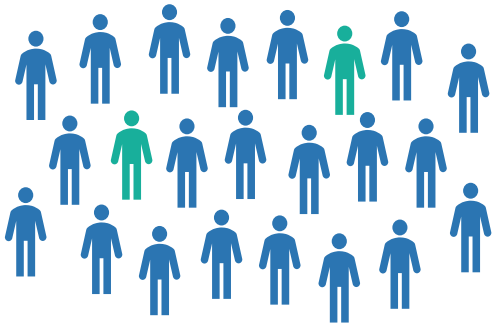
where $\boldsymbol{x}, \boldsymbol{y}$ are two samples to be compared



$$\mathrm{SSIM}(\boldsymbol{x},\boldsymbol{y}) = 1 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$$

- Calculated on sliding windows, then averaged.
- Ours is a 3D problem: SSIM computed in ***xy* plane**, 3rd dimension is **channel**
- Adjust C1-C2 to the pixel dynamic range

Systematics: Support size



Empirical evidence of the GAN **low support size** (Arora and Sanjeev, 2017)

- Learnt distribution not representative enough

Birthday paradox test (Brink, 2012):

How many people need to be in one room so that P(*at least two people have same birthday*) > 0.5 ?

- 365 days in a year → 23 people is enough

Generalized problem:

How many samples is it necessary to generate to have P*(at least one pair of duplicates among the samples)* > 0.5 ?

- **(The answer)^2** = estimate of the support size

Birthday paradox for GANs

Original birthday paradox problem

- Days in a year – **finite set** of possible values with discrete uniform distribution
- **Unique duplicates definition** – people born on the same day

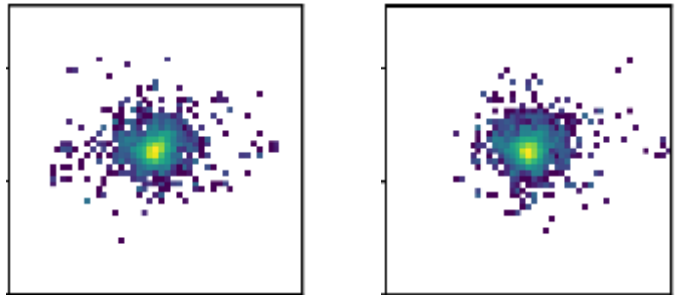
Exact duplicates

July 4 July 4

GAN distribution

- Images – pixels of **continuous values**
- **Multivariate continuous distribution** → occurrence of exact duplicates has zero probability
- Duplicates as “similar enough” images

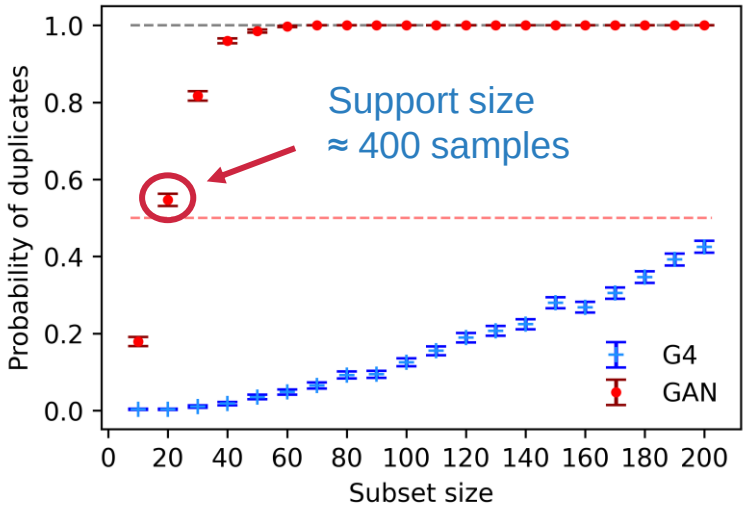
Not exact duplicates
But similar enough?



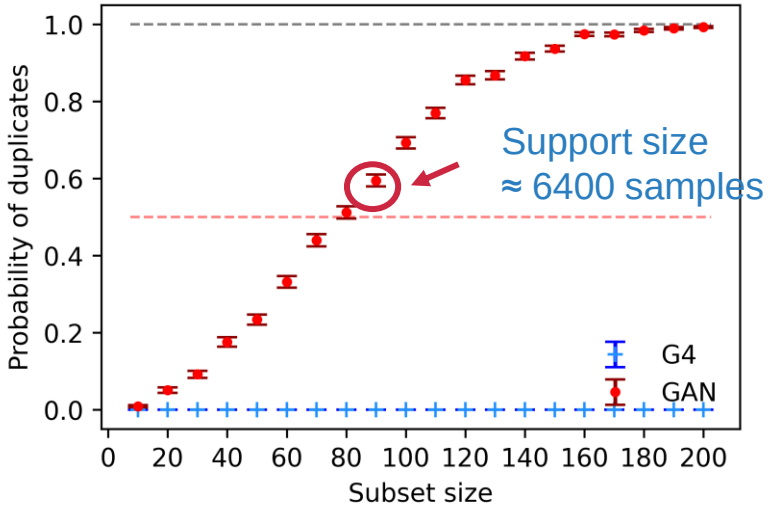
Similarity metrics depend on the use case and data type

Support size estimates

- GAN samples significantly more similar → **smaller** support size
- Test depends strongly on **duplicates definition**



Energy-based duplicate definition

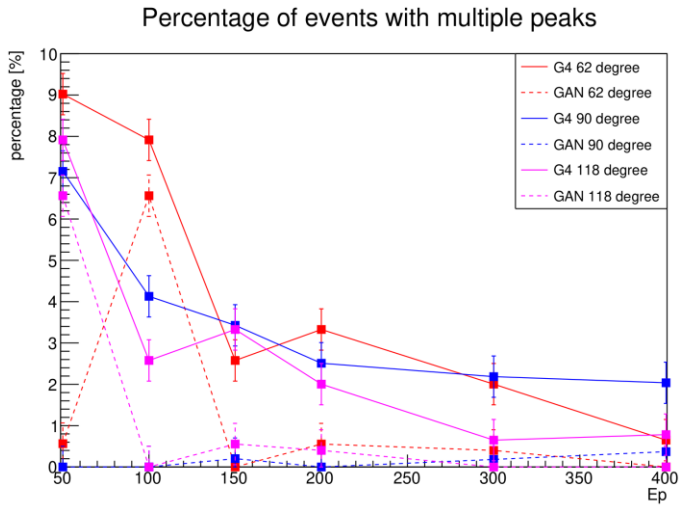


SSIM –based duplicate definition

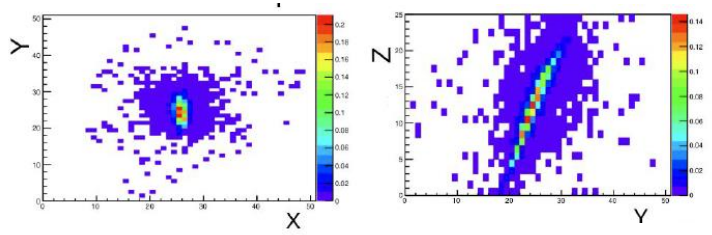
Not adapted to our problem?

Systematics: rare events

In some cases it is important to reproduce correctly the topology and occurrence of rare even

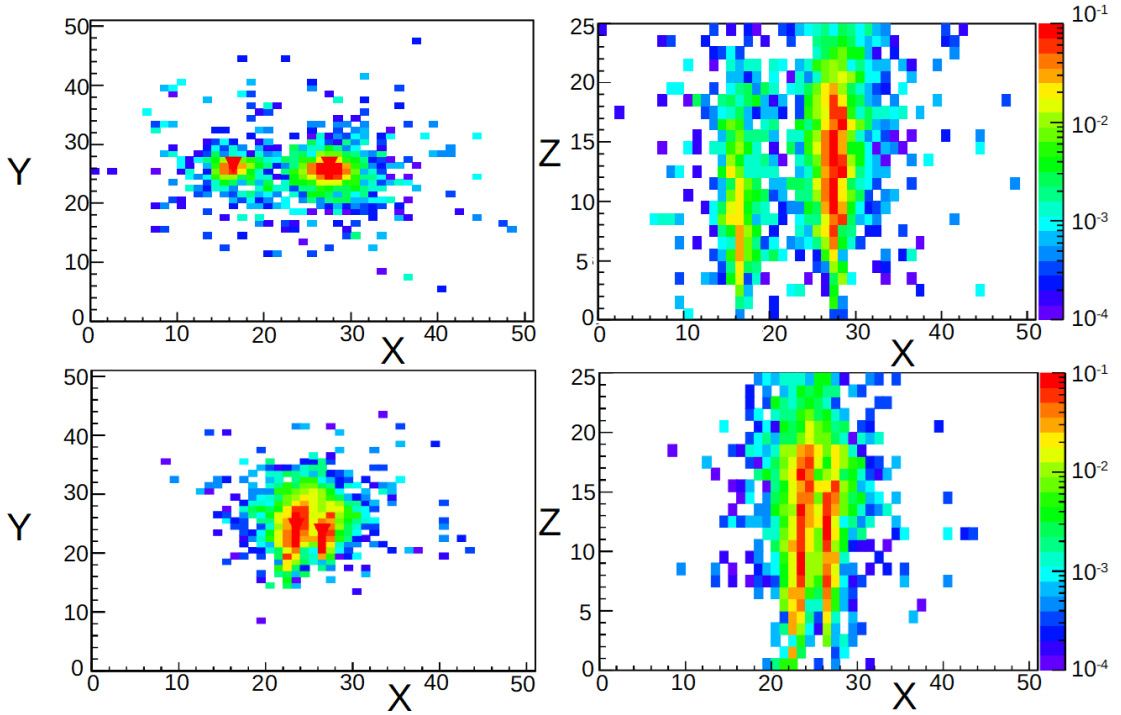


"Standard"



MC

GAN



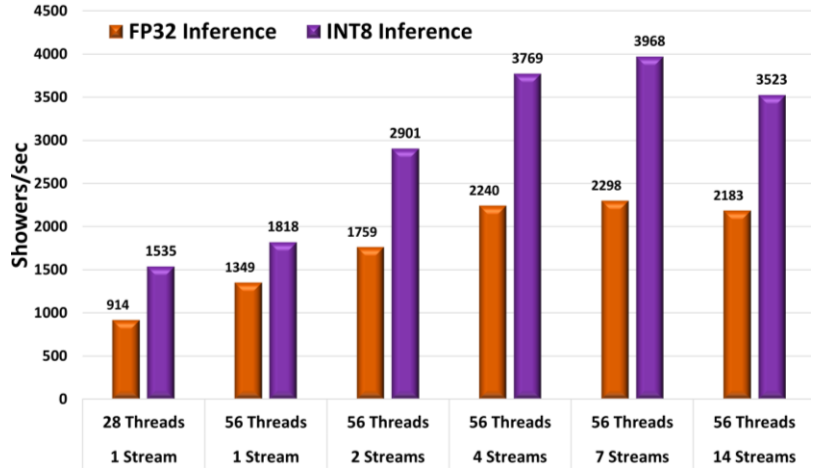
Computing resources

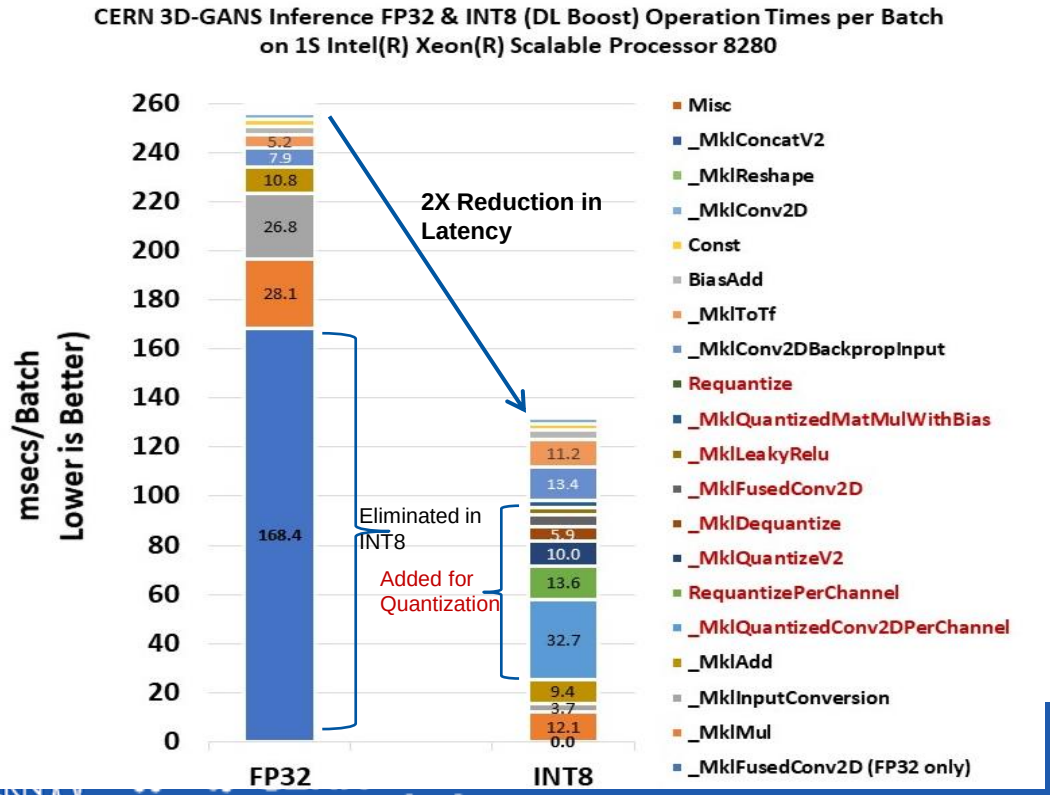
Faster then Monte Carlo?

F. Rehm, ICPRAM2021
in collaboration with Intel

Post training quantization (INT 8) using Intel DLBoost and iLoT tool

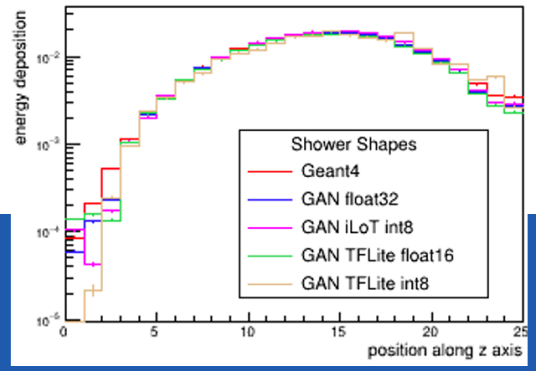
FP32: 3DGAN is **38000x faster** than Monte Carlo (on Intel Xeon processors)
INT8: quantized 3DGAN is **68000x faster** than Monte Carlo (FP32)

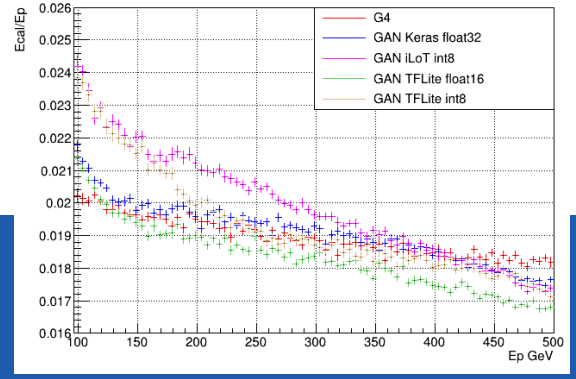




Reduced data representation reduce inference time but reduce phsyics performance

- Need ad-hoc optimisation strategy

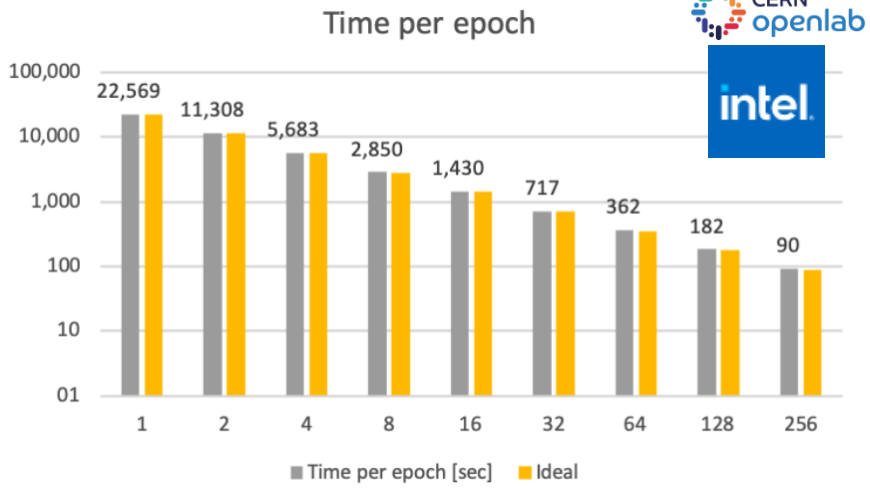


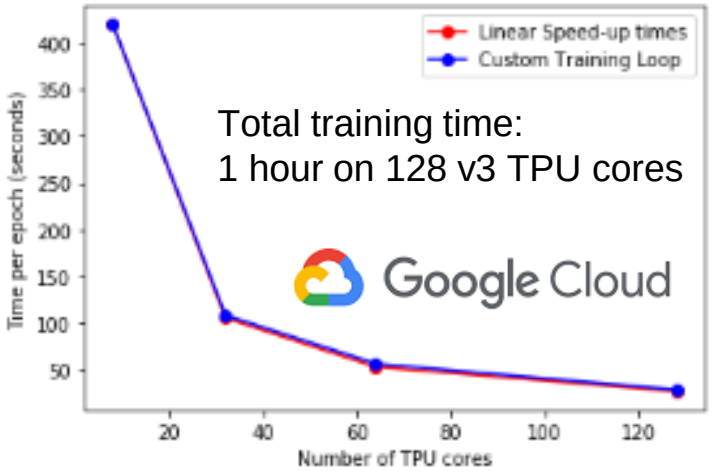


Training time

- Training the 3D convolutional GAN model (3M parameters) takes about 7 days on a GPU
  - **Distributed training** is essential
  - Need to keep **physics under control**
- Tested different data parallel approach on different hardware on **HPC** and **Cloud**

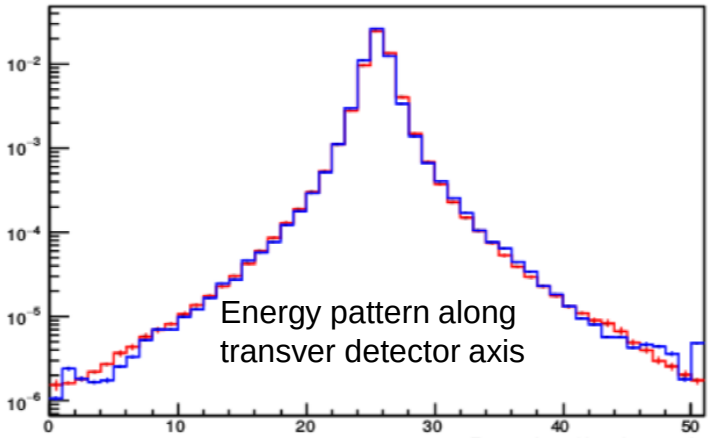
Total training time: 3 hours on 256 Intel Xeons





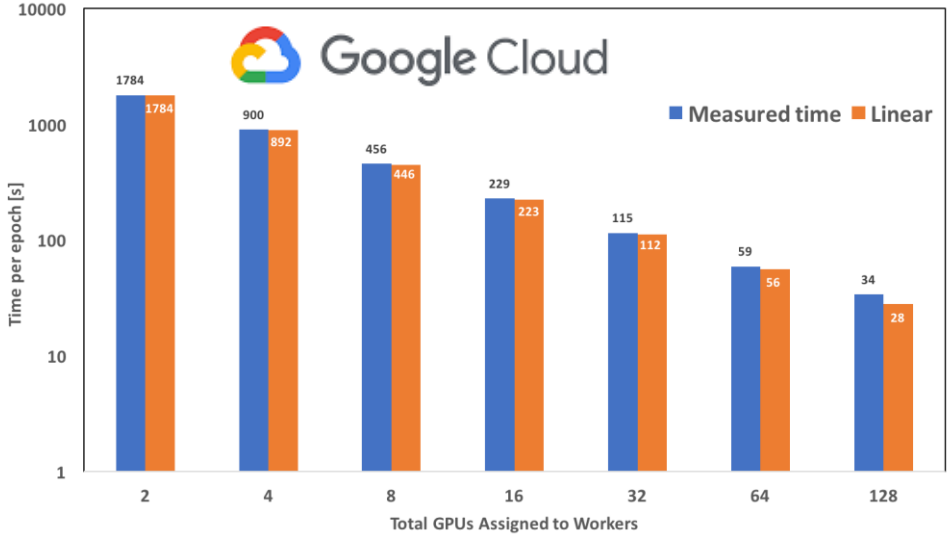
Total training time:
1 hour on 128 v3 TPU cores

Google Cloud



Energy pattern along
transver detector axis

Total training time:
1 hour on 128 V100 GPUs



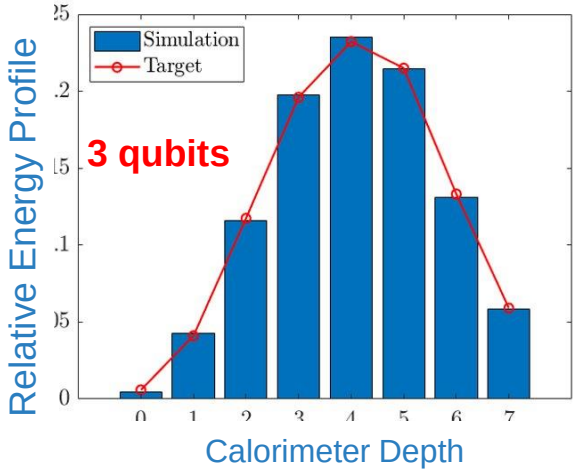
Summary

- High Energy Physics experiments are heavily investigating Generative Models for **fast simulation**
  - High level of **customization**
  - Increasing complexity
- Domain-specific knowledge is key
  - Model design and architecture **otpimisation**
  - Results **validation**
  - Models **interpretability** and **performance systematic studies**
- **Efficient** use of **computing resources** broadens the scope

Teaser: Quantum Generative Adversarial Networks

Explore quantum advantage in terms of:

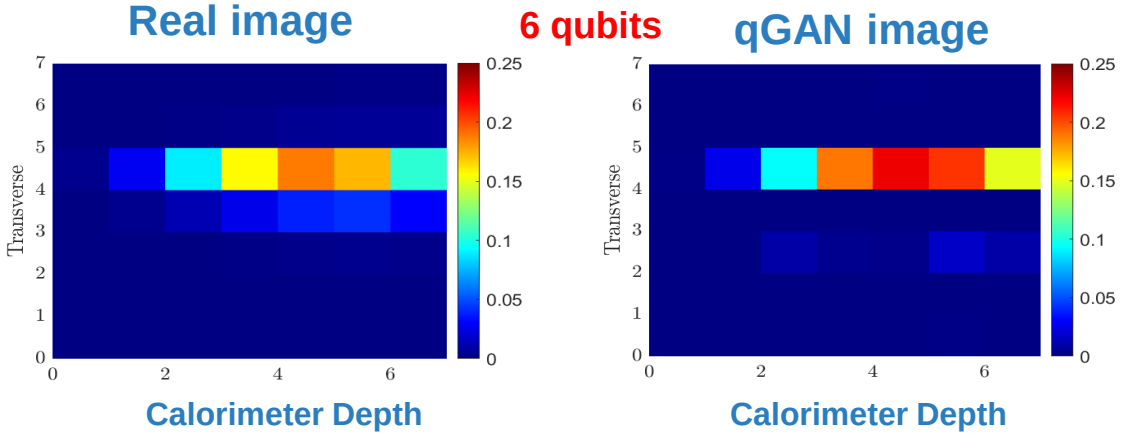
- Compressed data representation in quantum states
- **Faster** training with **smaller number of parameters**
- Support space of the learned distribution



Simplify 3DGAN simulation problem

1D & 2D energy profiles from detector

Train a **hybrid classical-quantum** GAN



QUANTUM TECHNOLOGY INITIATIVE

S. Y. Chang *et al.*,
arXiv:2103.15470
arXiv:2101.11132

Thanks!

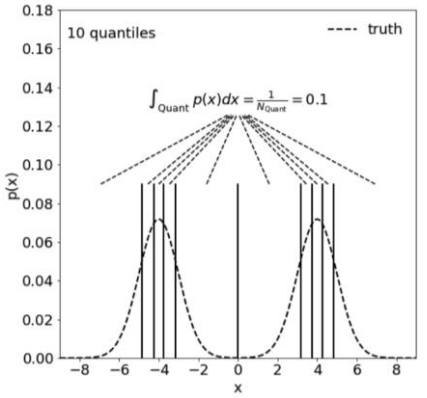
Sofia.Vallecorsa@cern.ch

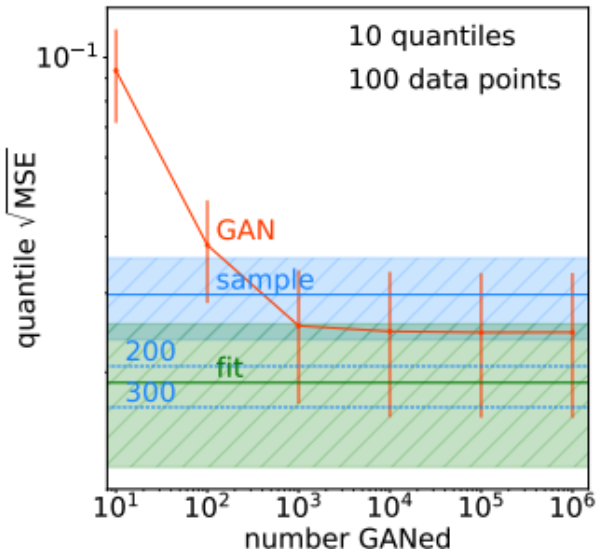
https://openlab.cern/
https://home.cern/

A. Butter *et al.*, arxiv:2008.06545

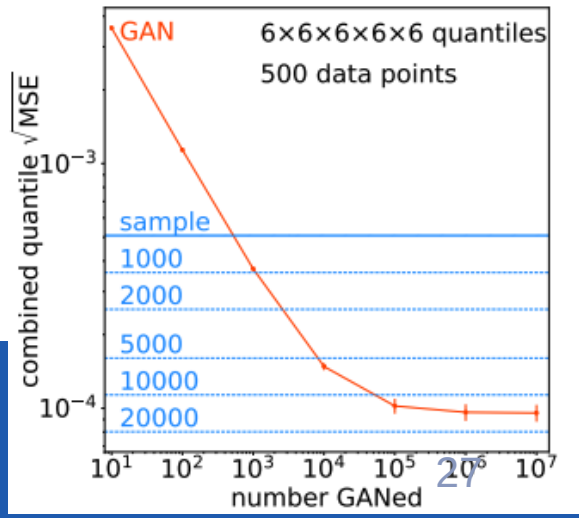
Systematics: increasing statistics

- If a GAN is trained on **N** data points, how many **new** points can be drawn?
- GAN can describe distribution better than training data
- Needs 10,000 GAN points to match 150 true points
- In terms of **information**:
  - **sample**: only data points
  - **fit**: data + true function
  - **GAN**: data + smooth, continuous function





Generalisation to multi-dimensional problem



Most physics data sets described by continuous function → GAN can interpolate