CERN

ML/AI methods for calorimeter simulation

November 9th, 2021

Sofia Vallecorsa - CERN openlab AI and Quantum Research, CERN IT Department

Deep Generative Models

Generative Models learn a probability distribution from the training dataset and produce a new set of examples that belong to the same distribution.

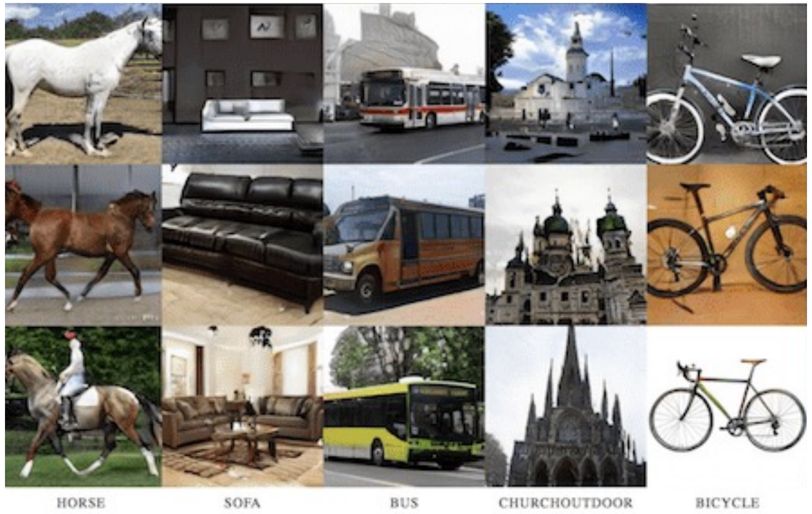
Deep models allow higher levels of **abstractions** and improve **generalization** wrt to **shallow models**

Multiple applications in Simulation, Anomaly Detection, Data manipulation, Data Augmentation

A variety of models:

- Generative Adversarial Networks
- (Variational) Auto Encoders
- Auto-regressive models
- Normalizing flows
- …

Ex. Synthetic image generation



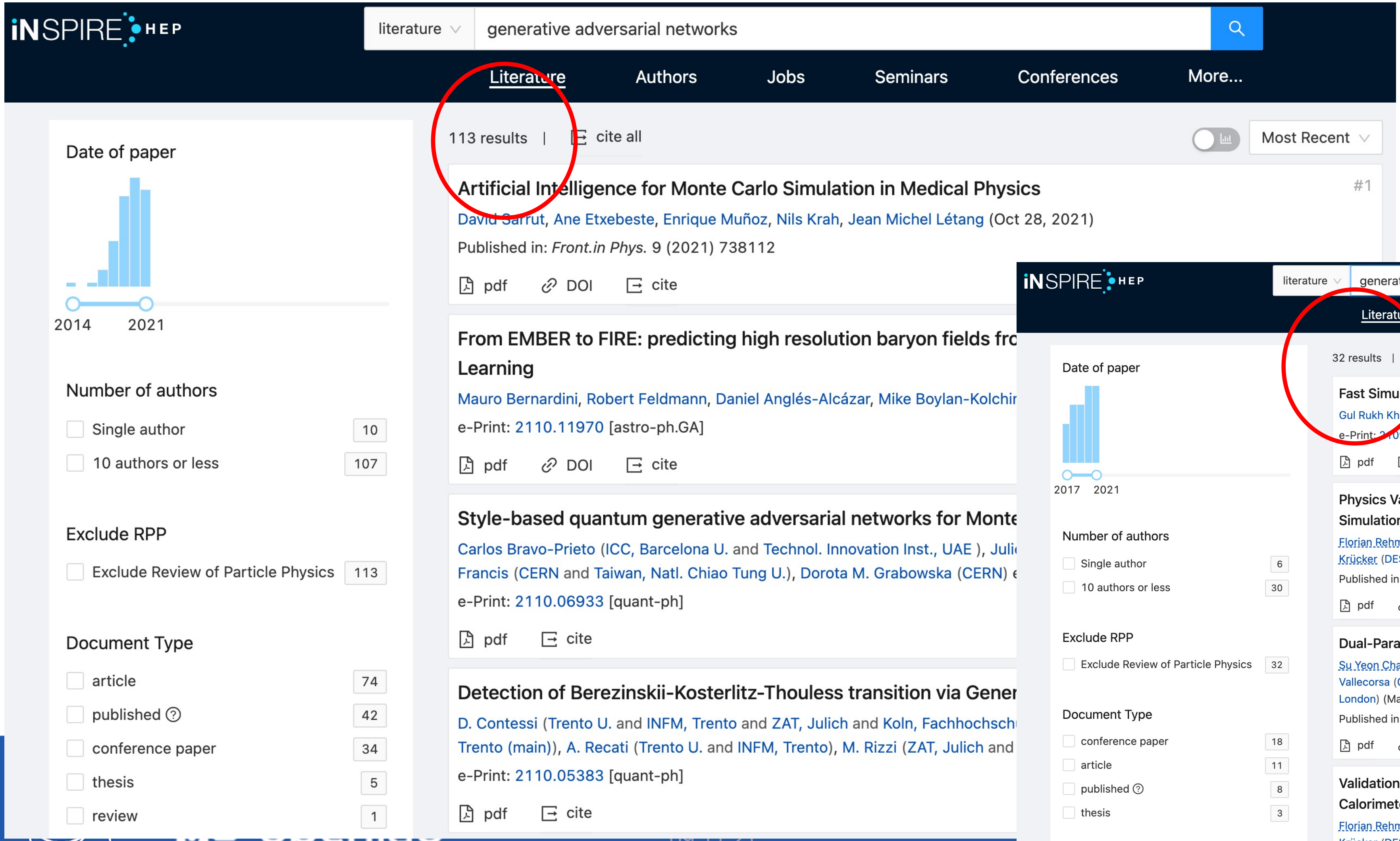
https://arxiv.org/abs/1710.10196

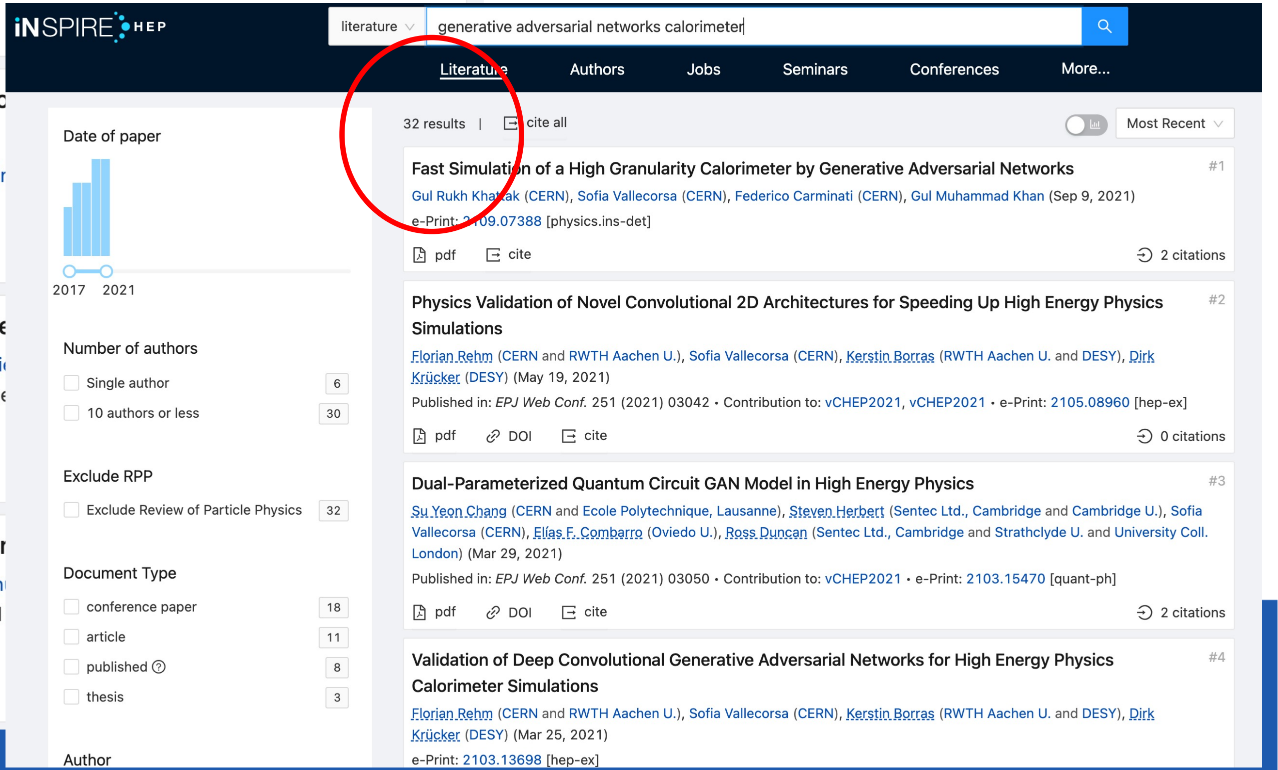
Ex. Text to image translation

‘Small blue bird with black wings’ →
‘Small yellow bird with black wings’

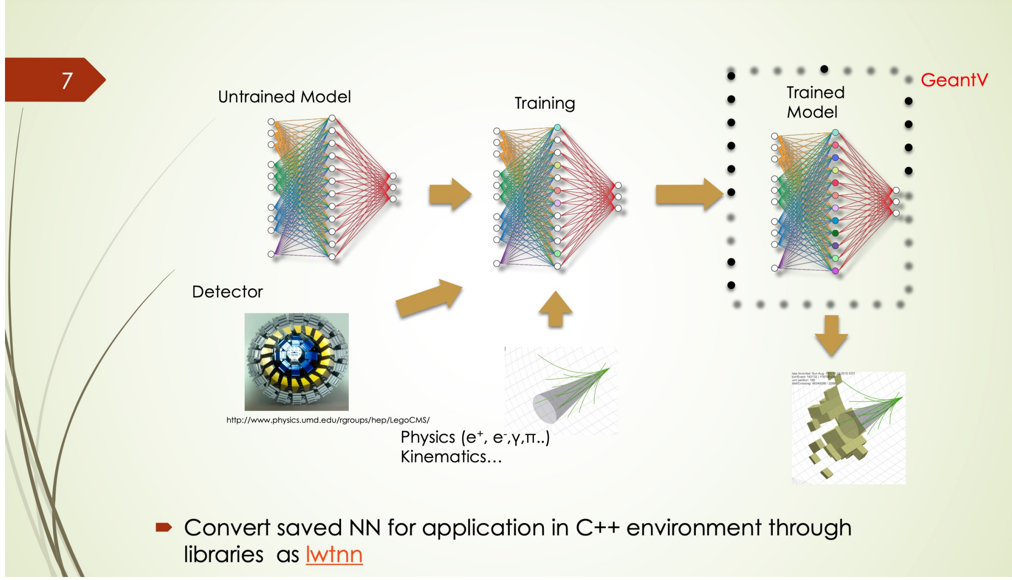
https://arxiv.org/pdf/1605.05396.pdf

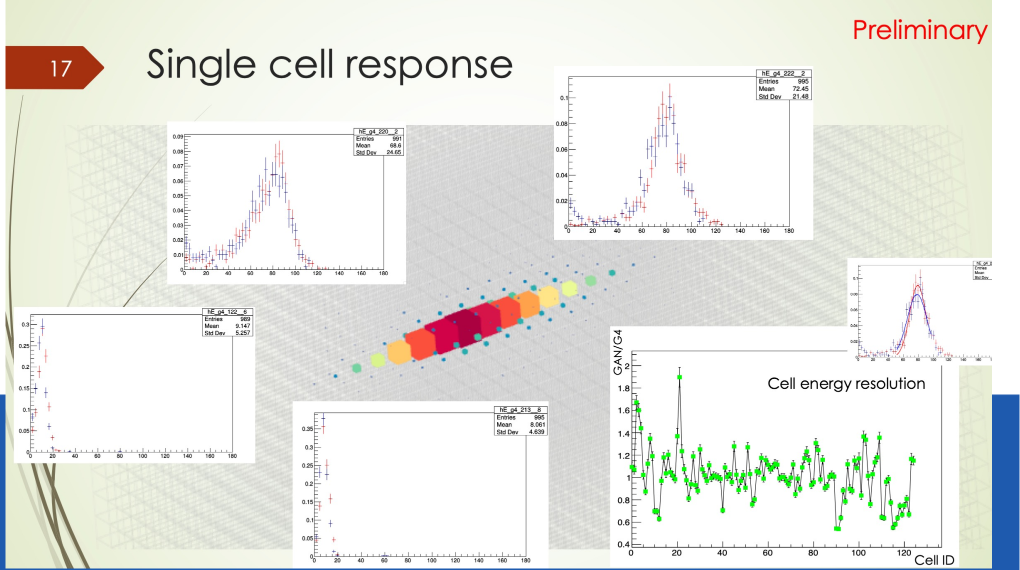
Very popular models





Back in 2017.. @ ACAT





CaloGAN

Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis

Luke de Oliveira[a], Michela Paganini[a,b], and Benjamin Nachman[a]

[a] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA, 94720, USA*
[b] *Department of Physics, Yale University, New Haven, CT 06520, USA*

E-mail: lukedeoliveira@lbl.gov, michela.paganini@yale.edu, bnachman@cern.ch

ABSTRACT: We provide a bridge between generative modeling in the Machine Learning community and simulated physical processes in High Energy Particle Physics by applying a novel Generative Adversarial Network (GAN) architecture to the production of *jet images* – 2D representations of energy depositions from particles interacting with a calorimeter. We propose a simple architecture, the Location-Aware Generative Adversarial Network, that learns to produce realistic radiation patterns from simulated high energy particle collisions. The pixel intensities of GAN-generated images faithfully span over many orders of magnitude and exhibit the desired low-dimensional physical properties (*i.e.*, jet mass, n-subjettiness, etc.). We shed light on limitations, and provide a novel empirical validation of image quality and validity of GAN-produced simulations of the natural world. This work provides a base for further explorations of GANs for use in faster simulation in High Energy Particle Physics.

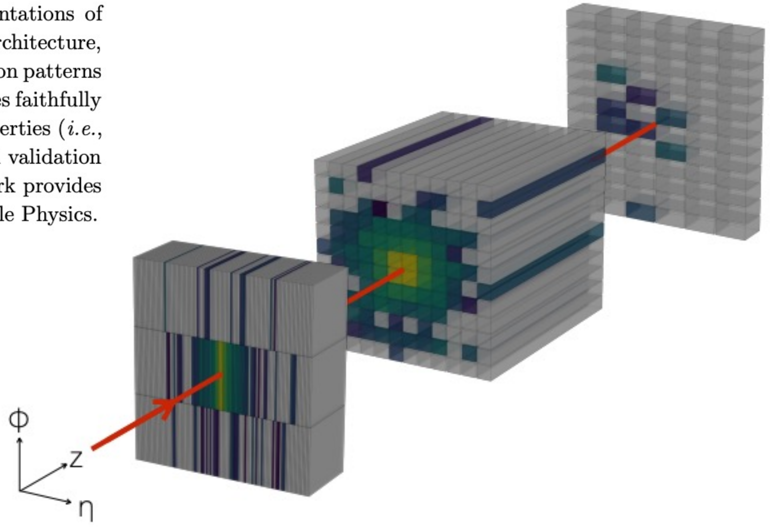
CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

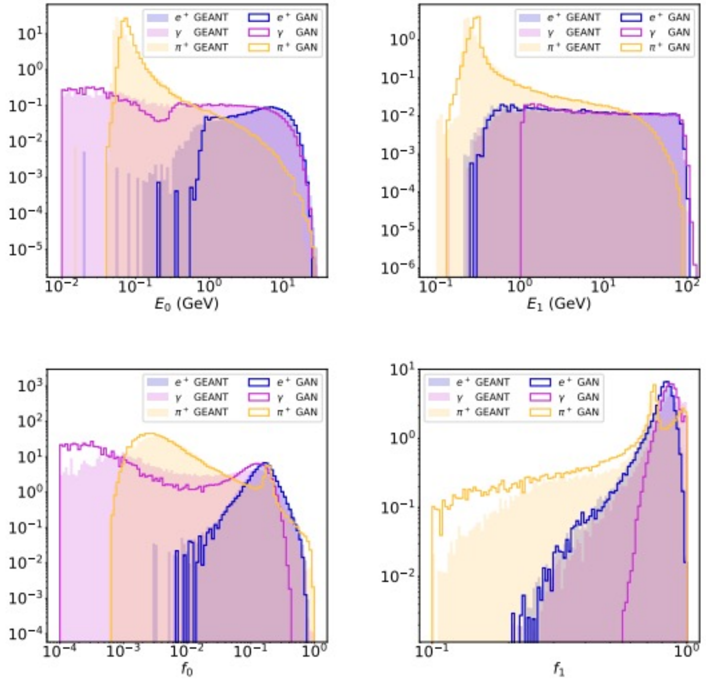
Michela Paganini,[1,2,*] Luke de Oliveira,[2,†] and Benjamin Nachman[2,‡]

[1] *Department of Physics, Yale University, New Haven, CT 06520, USA*
[2] *Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA*

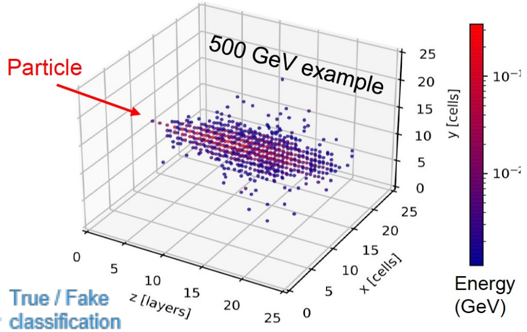
(Dated: January 1, 2018)

The precise modeling of subatomic particle interactions and propagation through matter is paramount for the advancement of nuclear and particle physics searches and precision measurements. The most computationally expensive step in the simulation pipeline of a typical experiment at the Large Hadron Collider (LHC) is the detailed modeling of the full complexity of physics processes that govern the motion and evolution of particle showers inside calorimeters. We introduce CaloGAN, a new fast simulation technique based on generative adversarial networks (GANs). We apply these neural networks to the modeling of electromagnetic showers in a longitudinally segmented calorimeter, and achieve speedup factors comparable to or better than existing full simulation techniques on CPU ($100\times$-$1000\times$) and even faster on GPU (up to $\sim 10^5\times$). There are still challenges for achieving precision across the entire phase space, but our solution can reproduce a variety of geometric shower shape properties of photons, positrons and charged pions. This represents a significant stepping stone toward a full neural network-based detector simulation that could save significant computing time and enable many analyses now and in the future.

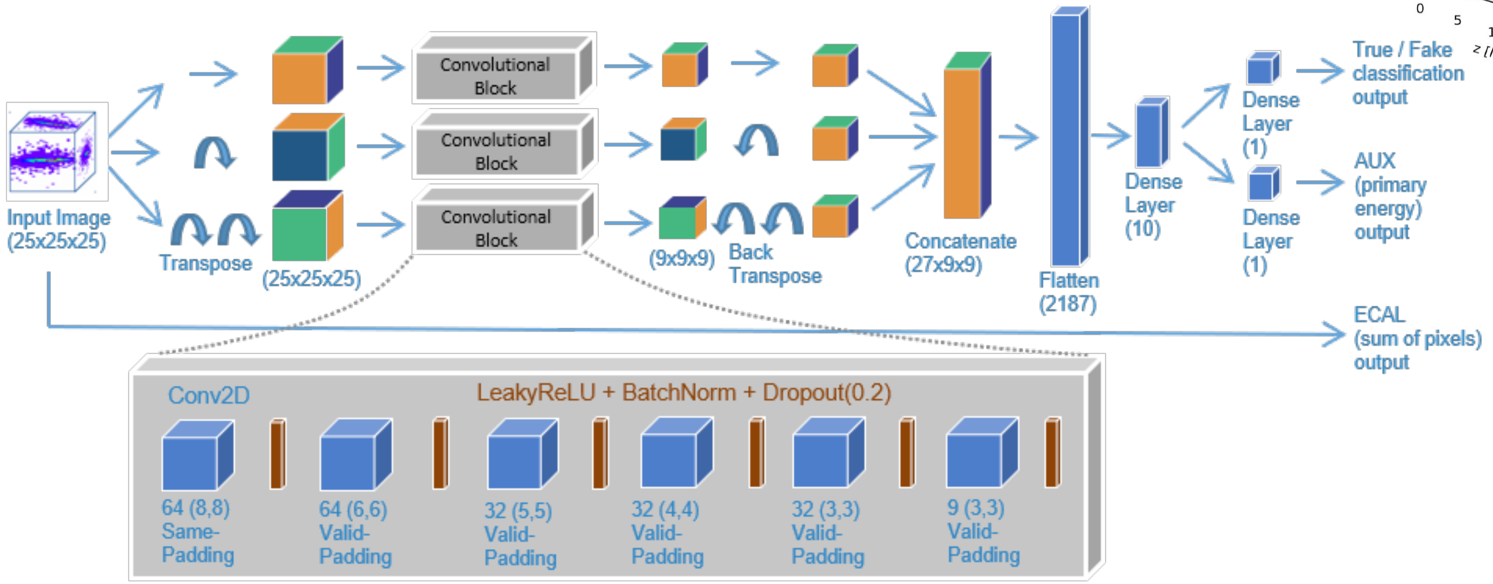


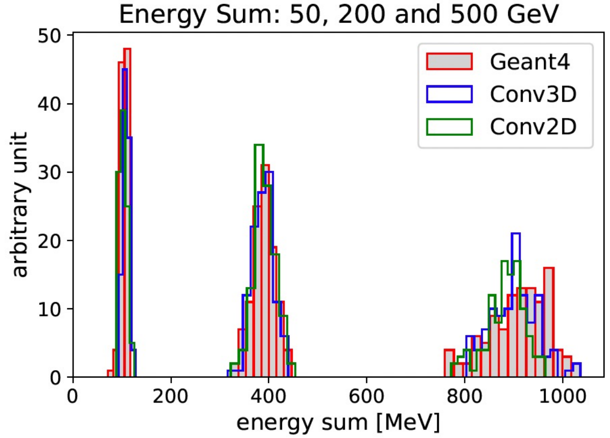


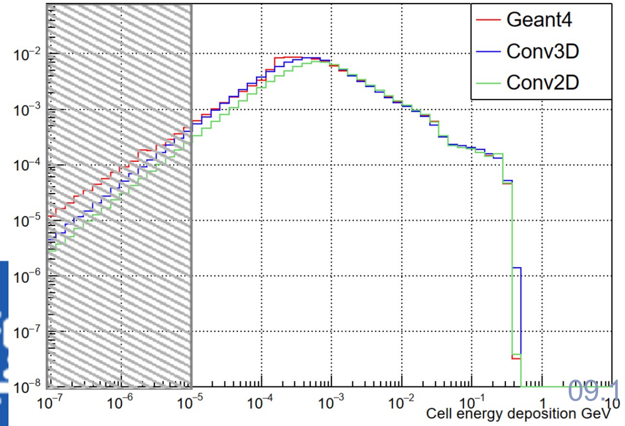
3DGAN 2.0

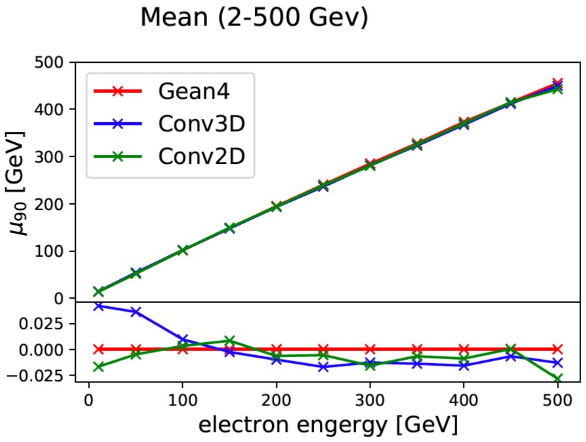


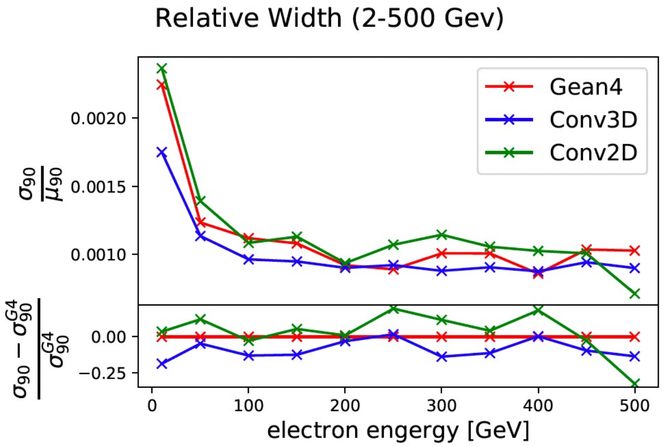
Rehm, Florian, et al. "Physics Validation of Novel Convolutional 2D Architectures for Speeding Up High Energy Physics Simulations." *arXiv preprint arXiv:2105.08960* (2021).





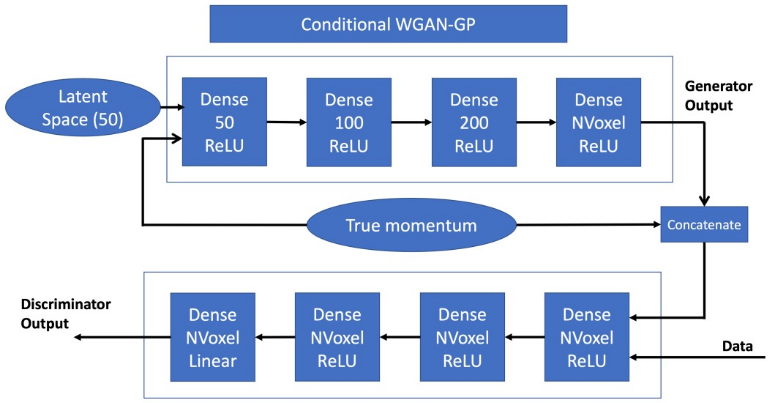


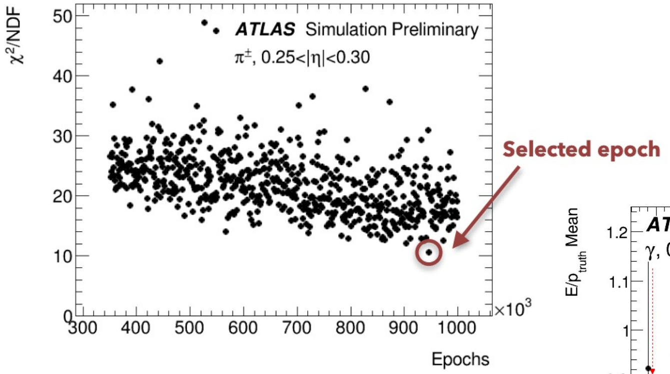


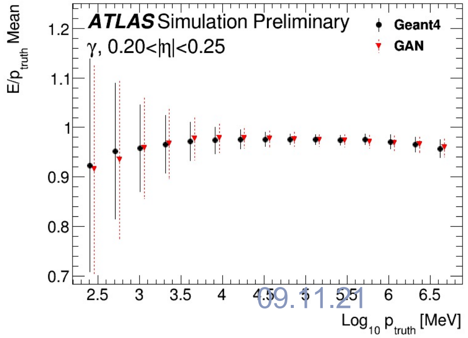


Recent models (I)

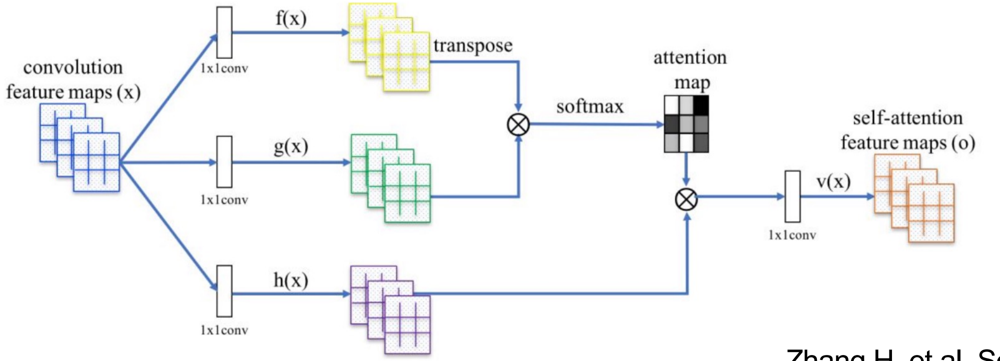
FastCaloGAN: 300 GANs for the **full ATLAS calo**
part of AtlFast3 (J.F. Beirer, ML4Jets2021)



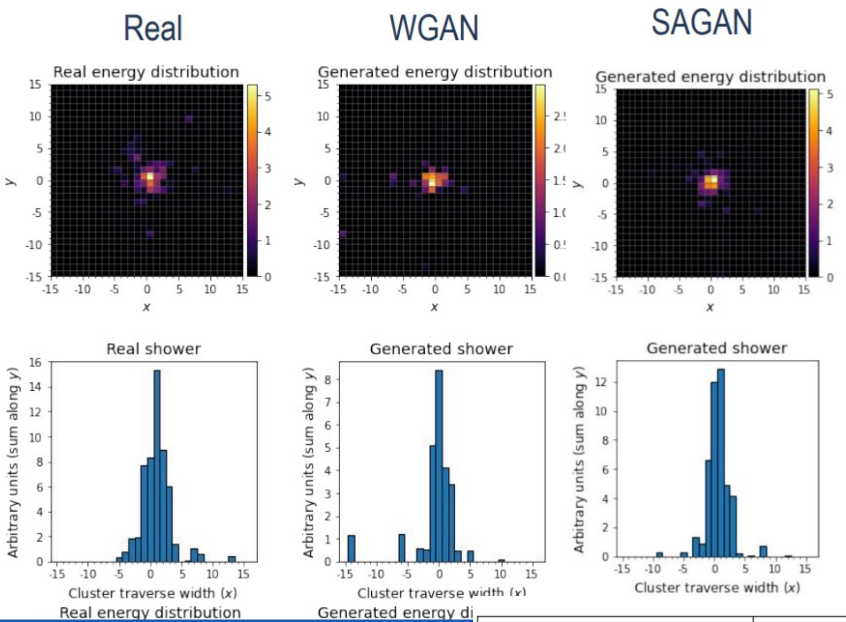




Self-Attention GANs for LHCb calorimeter
F. Ratnikov, A. Rogachev:
https://indico.cern.ch/event/948465/contributions/4324135



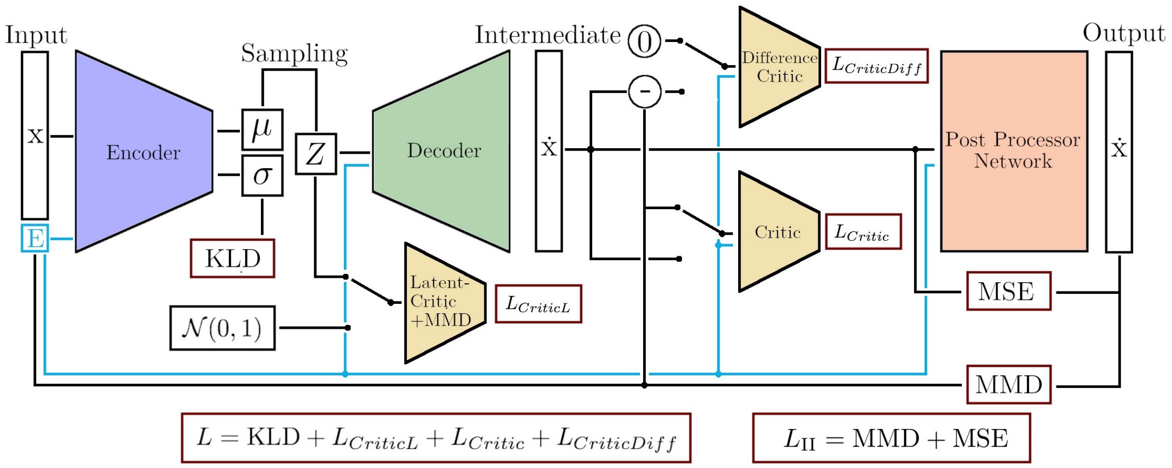
Zhang H. et al. Self-attention generative adversarial networks //International conference on machine learning. – PMLR, 2019 C. 7354-7363.

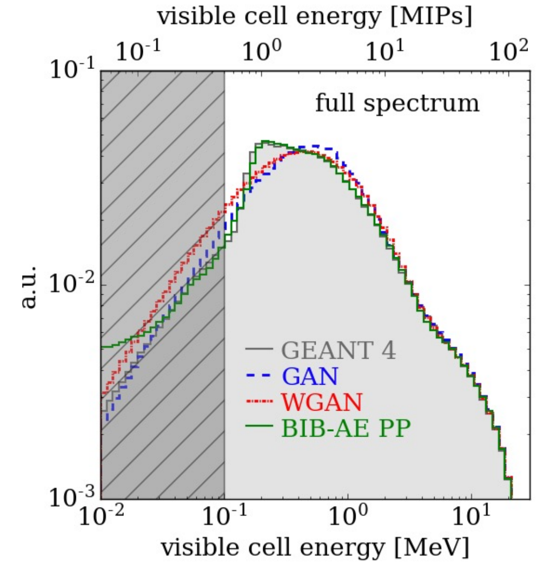


Model	Physics PRD-AUC	Raw Images PRD-AUC
WGAN	0.936	0.971
SAGAN+SN D	0.895	0.901
SAGAN+SN G and D	0.948	0.975

Recent models (II)

GAN – AutoEncoder hybrid

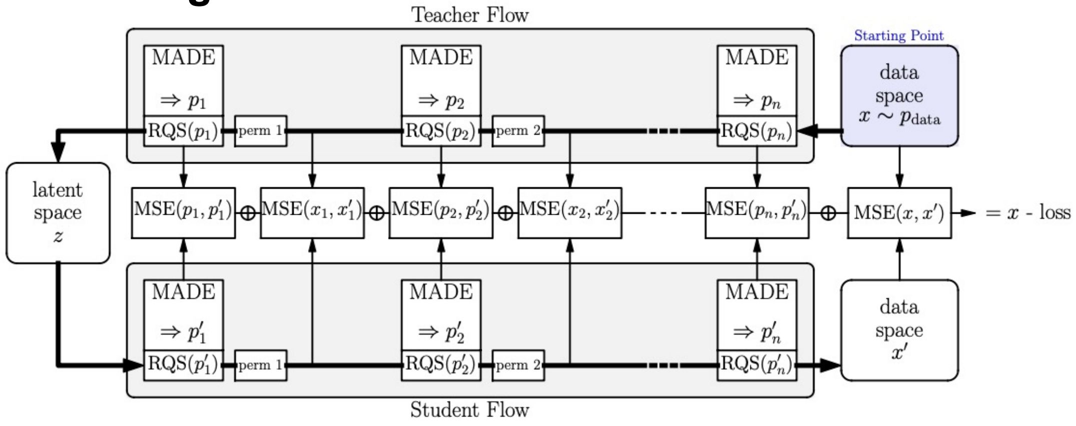


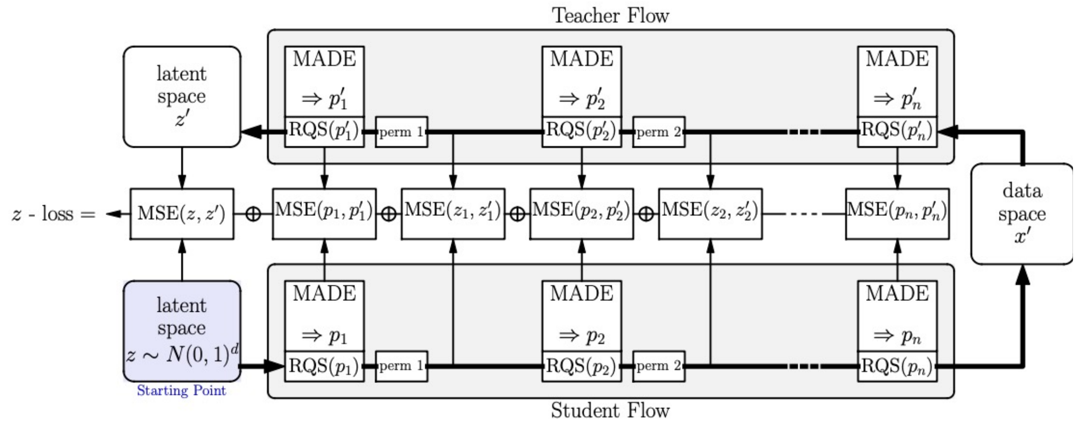


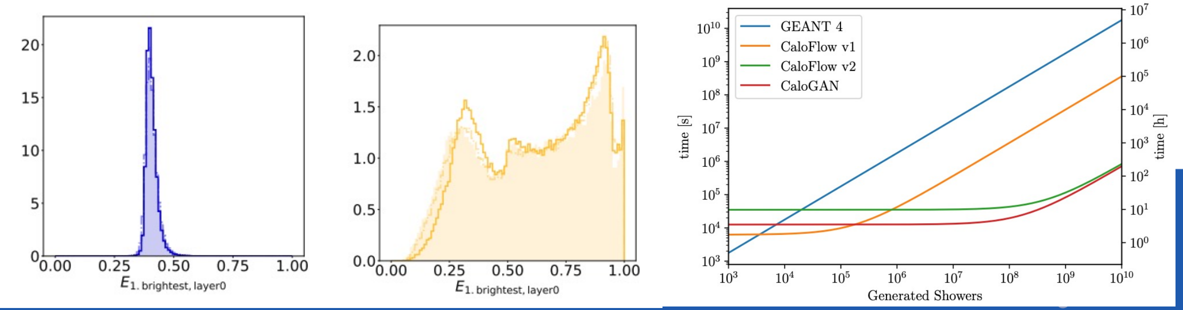
Buhmann, Erik, et al. "Getting high: high fidelity simulation of high granularity calorimeters with high speed." *Computing and Software for Big Science* 5.1 (2021): 1-17.

Normalizing Flows

Krause, Claudius, and David Shih. "CaloFlow II: Even Faster and Still Accurate Generation of Calorimeter Showers with Normalizing Flows." *arXiv:2110.11377*







HSF simulation : https://indico.cern.ch/event/1089895/

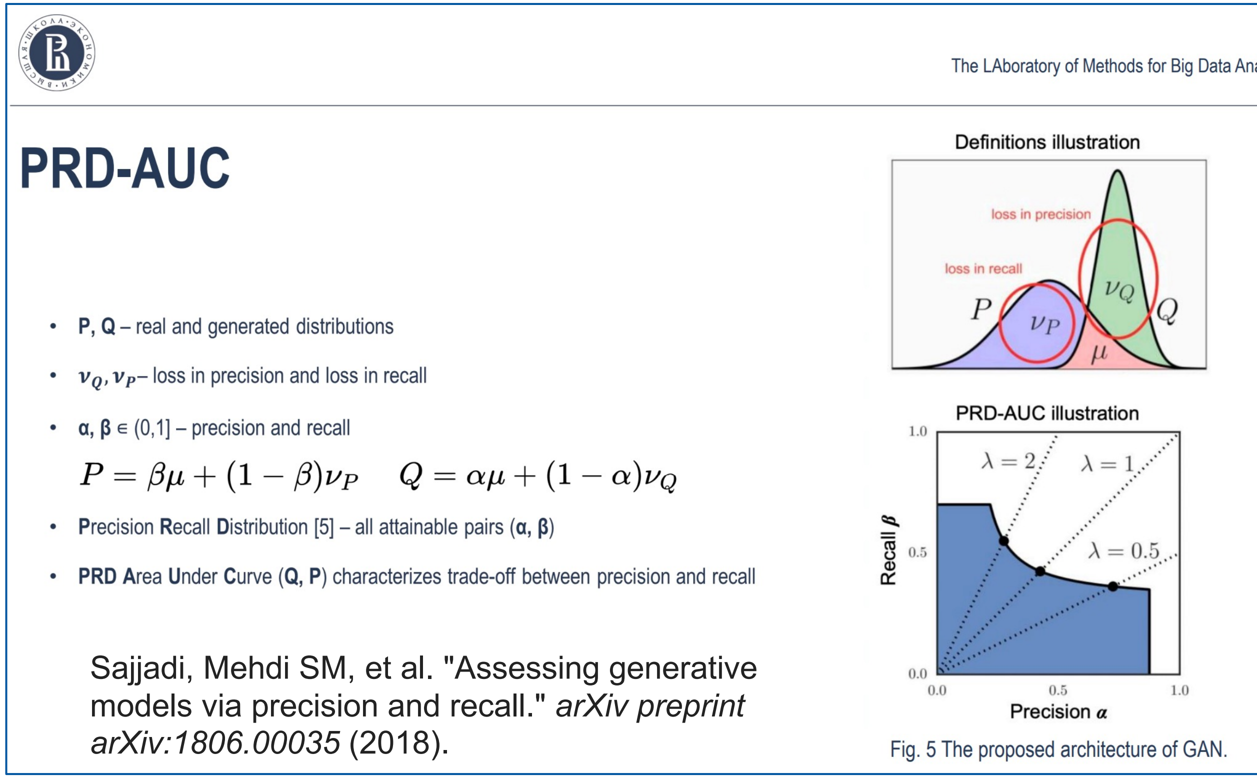
The situation today

- **Deep Learning-based fast simulation is a reality**
  - Large number of prototypes for different experiments
- **Need to bring it to production level**
  - Establish **validation process** and evaluate systematics
    - Metrics, benchmarks, ..
  - Design **integration** in (fast) simulation frameworks
    - Work already ongoing in Geant4 (D. Salamani, A. Zaborowska)
  - Evaluate **computing resources**
    - Fair comparison to state-of-the-art, resource budget
  - **Generalisation**
    - How to move beyond the ”one use case - one prototype” approach?

Validation metrics

Self-Attention GANs for LHCb calorimeter
F. Ratnikov, A. Rogachev:
https://indico.cern.ch/event/948465/contributions/4324135

- Measure difference between **model and real PDF**
  - Kullback-Leibler Divergence
  - Inception score, Fréchet Inception Distance
  - Maximum Mean Discrepancy
  - Structural Similarity Index
  - ...
- Compare **physics distributions to MC**
- Investigate **different aspects**
  - Mixing and coverage (sample diversity)
  - Saliency
  - Mode collapse or mode dropping
  - Overfitting (has the network memorized samples?)

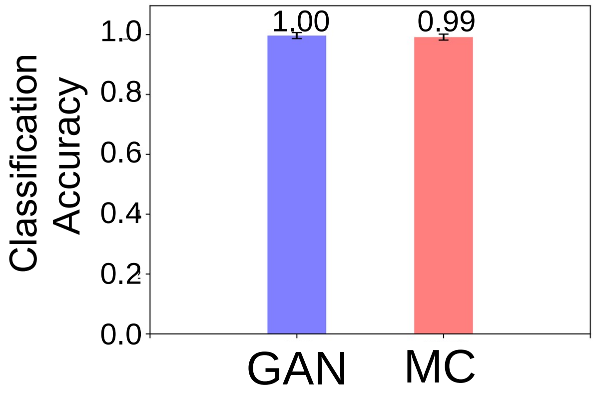


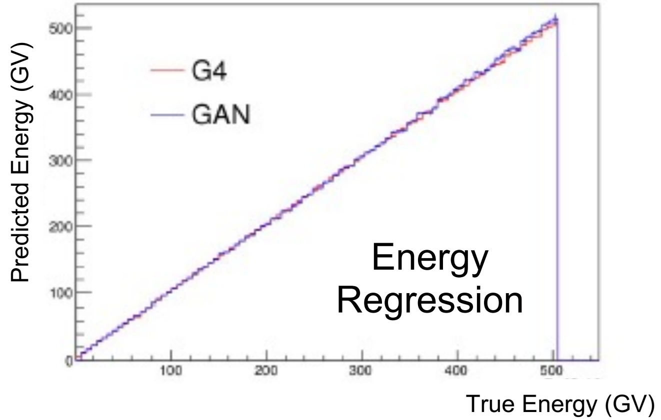
Khattak, Gul Rukh, et al. "High energy physics calorimeter detector simulation using generative adversarial networks with domain related constraints." *IEEE Access* 9 (2021): 108899-108911

Validation through “external” tools

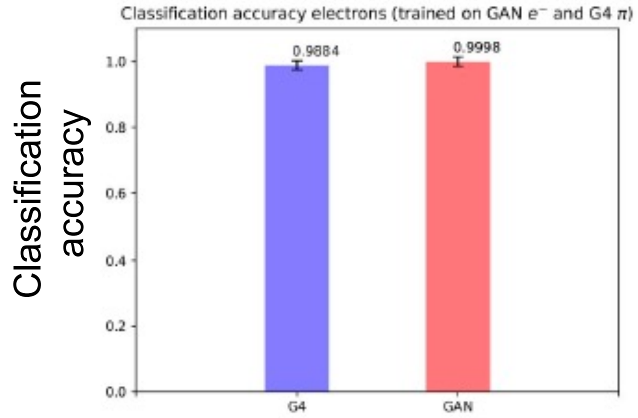
- **Triforce*** DNN has been developed for electron/pion classification and energy regression

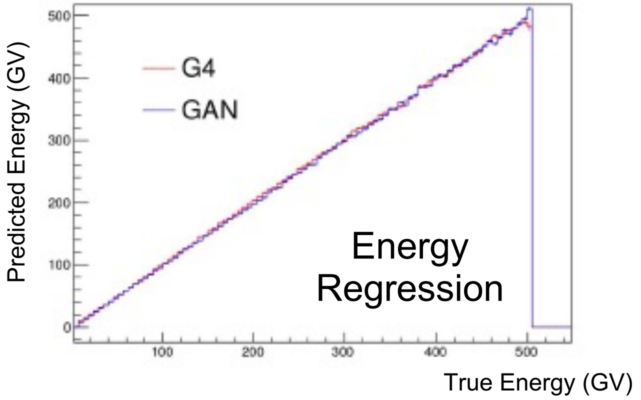
Trained on G4





Trained on GAN



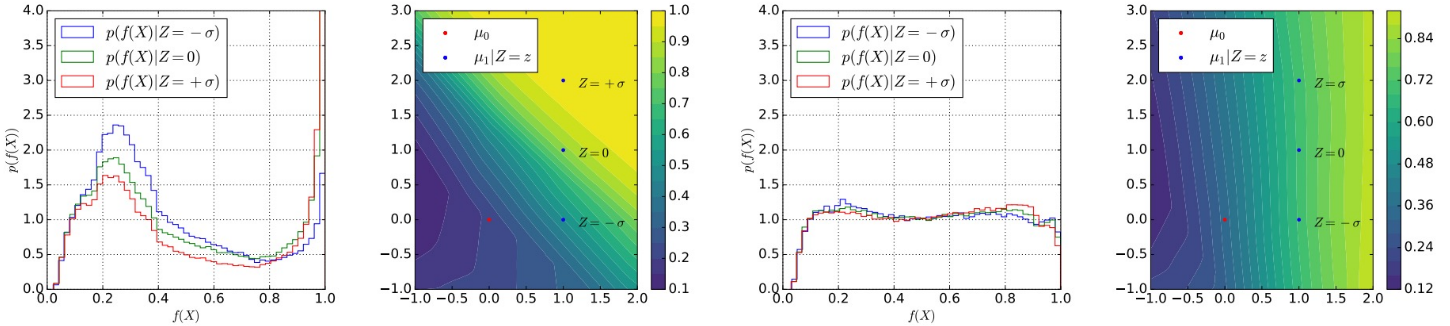


*D. Belayneh et al.,“Calorimetry with deep learning: Particle simulation and reconstruction for collider physics,” 2019, https://inspirehep.net/literature/1770936

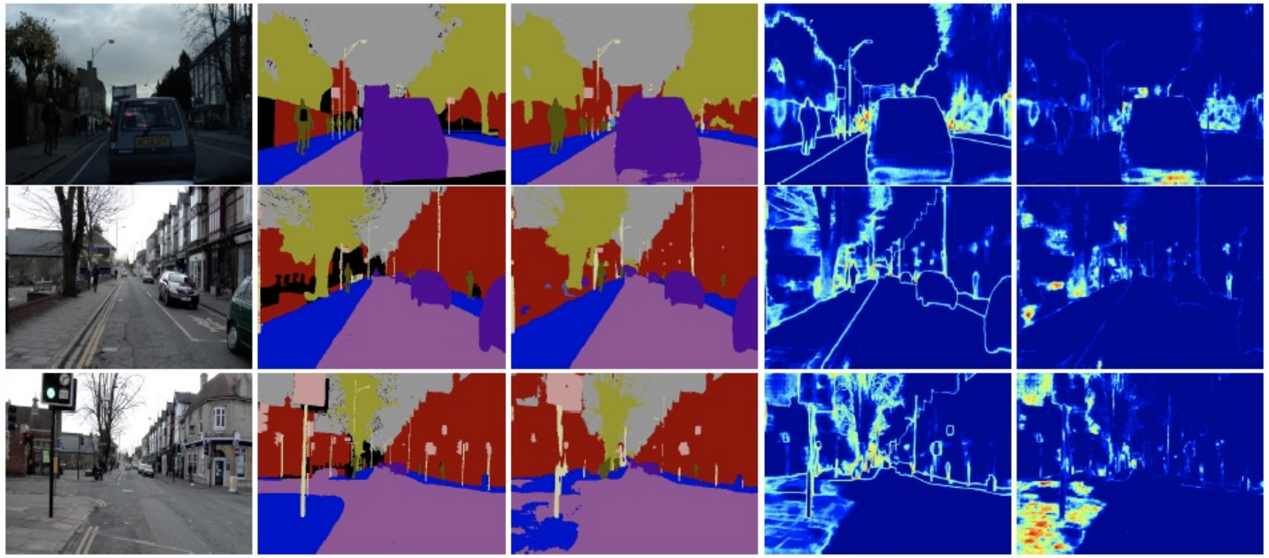
Systematic effects

Louppe, Gilles, Michael Kagan, and Kyle Cranmer. "Learning to pivot with adversarial networks." *arXiv preprint arXiv:1611.01046* (2016).

- Different approaches and techniques
- Research not restricted to generative models



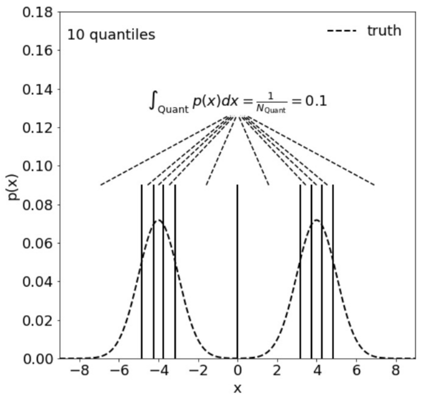
Kendal, Gal, NIPS 2017, https://papers.nips.cc/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf

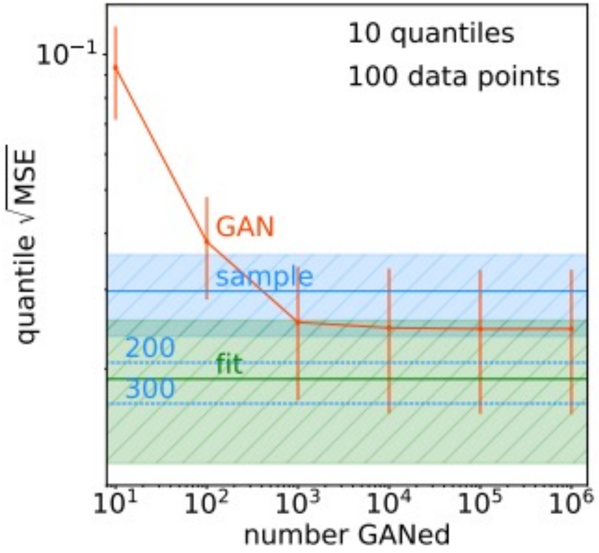


(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

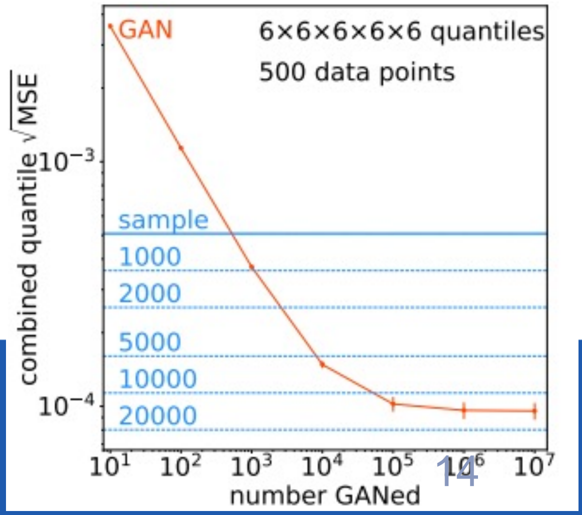
Systematics: training dataset size

- 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

Systematics: image similarity

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

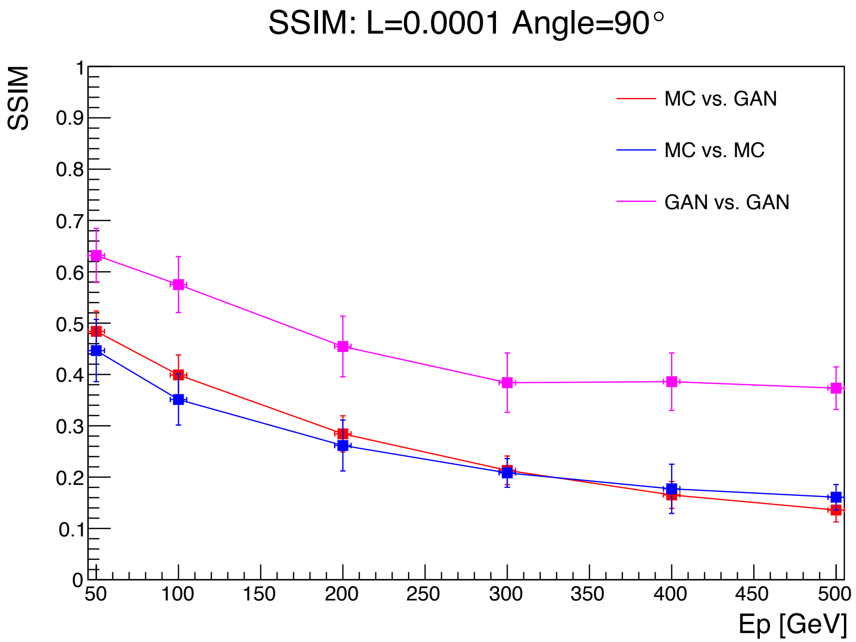
How much **diversity** in the generated sample?

- 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

- 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



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

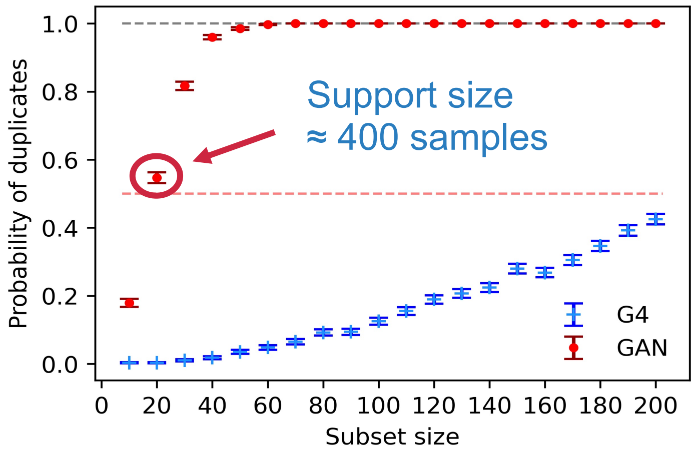
Jaruskova, Kristina, and Sofia Vallecorsa. "Estimating the Support Size of GANs for High Energy Physics Detector Simulation."ML4PS NEURIP202

Systematics: Support size

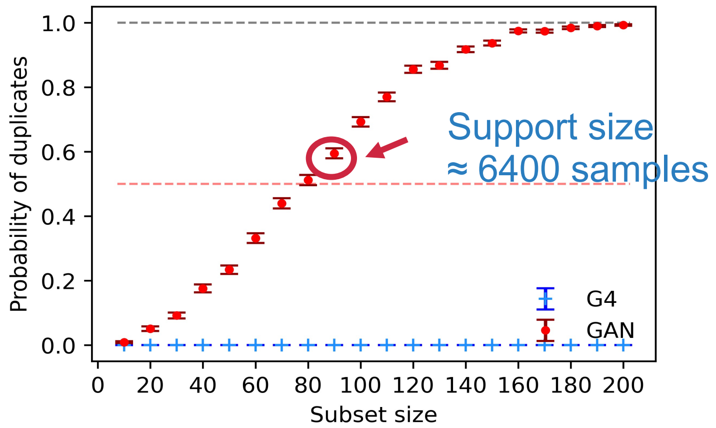
Empirical evidence of the GAN **low support size** (Arora and Sanjeev, 2017)

- Learnt distribution not representative enough
- Use Birthday paradox test to measure GAN support size
- GAN samples significantly more similar → **smaller** support size
- Test depends strongly on **duplicates definition**

Not adapted to our problem?



Energy-based duplicate definition

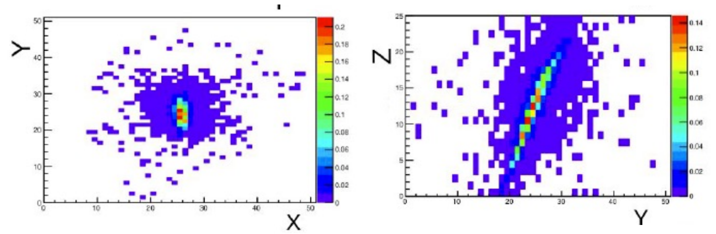


SSIM –based duplicate definition

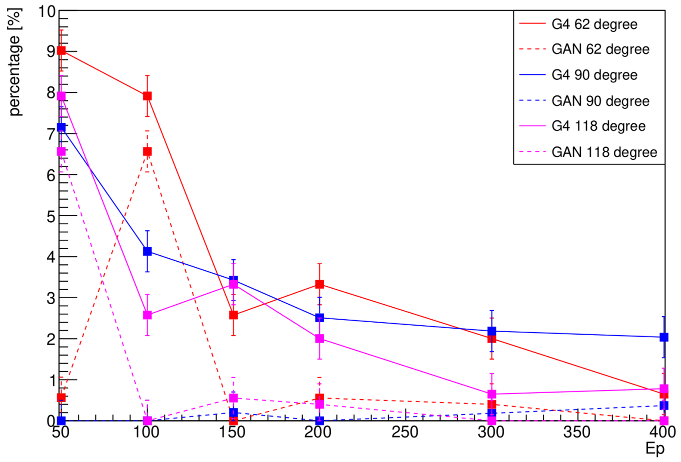
Systematics: rare events

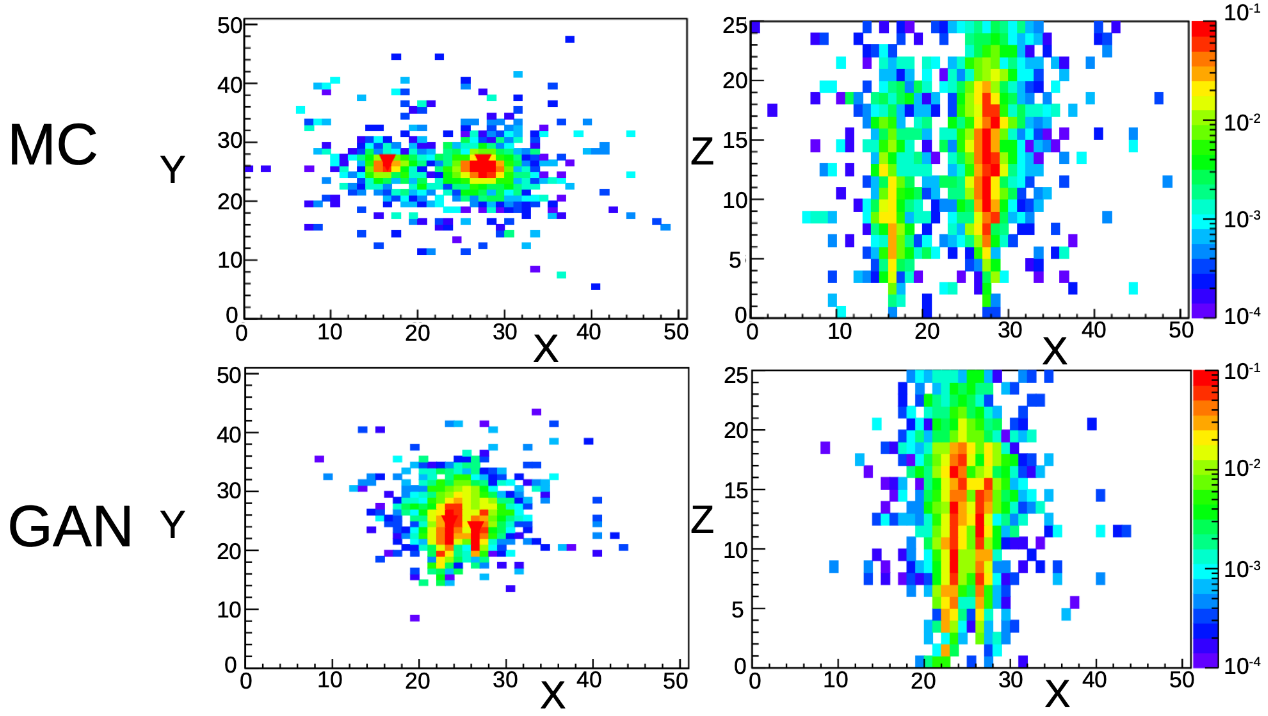
In some cases it is important to reproduce correctly the topology and occurrence of rare events

"Standard"



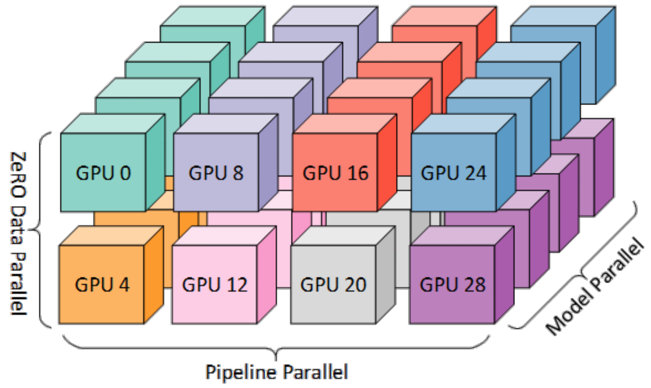
Percentage of events with multiple peaks





Computing resources

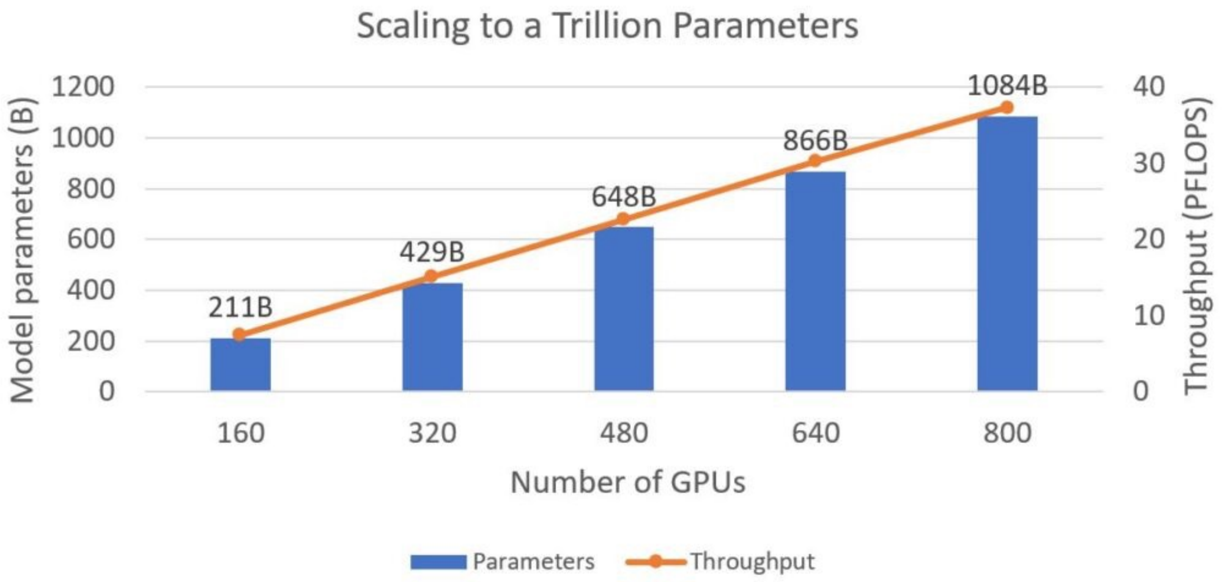
Access to large scale resources essential to model development



https://www.microsoft.com/en-us/research/blog/deepspeed-extreme-scale-model-training-for-everyone/

DeepSpeed and ZeRO-2 on Microsoft Azure

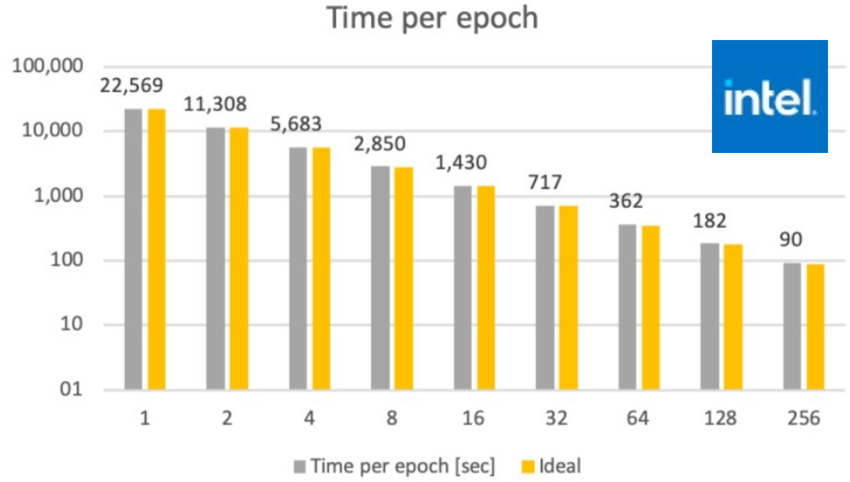
Hybrid parallel strategies, **Reduced precision** representation , **Hardware-aware** optimization enable **extreme scaling**



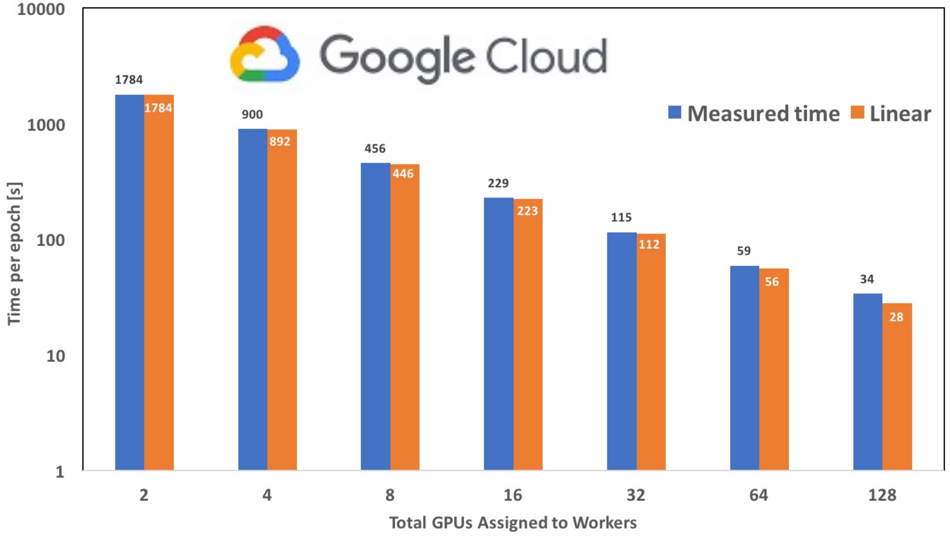
Reducing training time

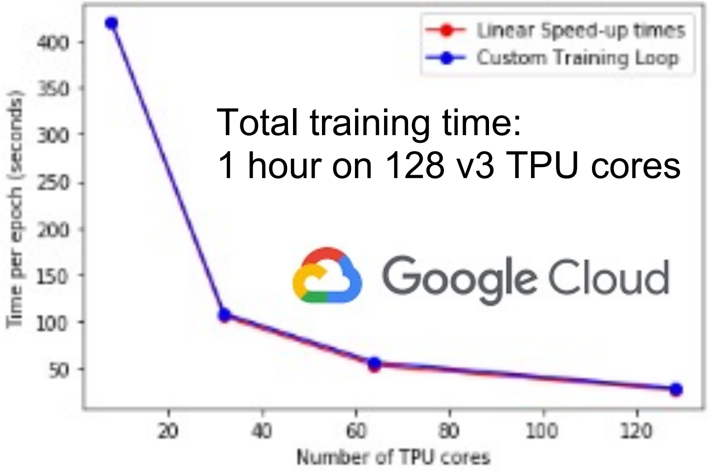
- Training the 3D convolutional GAN model (3M parameters) takes about **7 days on a V100 GPU**
- Tested different data parallel approach on different hardware on **HPC** and **Cloud**

Total training time: 3 hours on 256 Intel Xeons

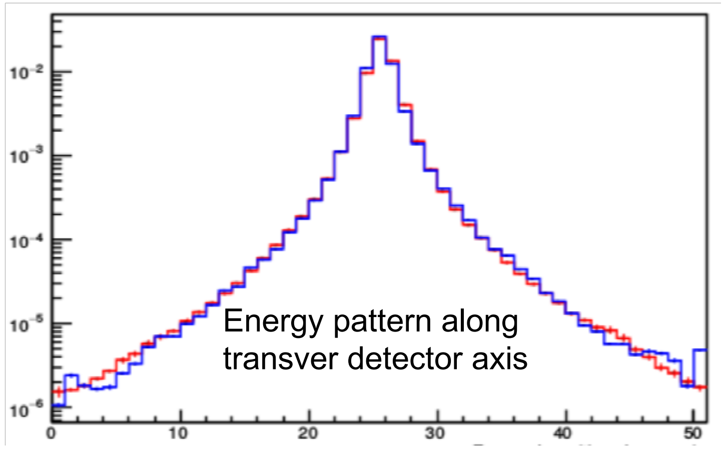


Total training time:
1 hour on 128 V100 GPUs





Total training time:
1 hour on 128 v3 TPU cores

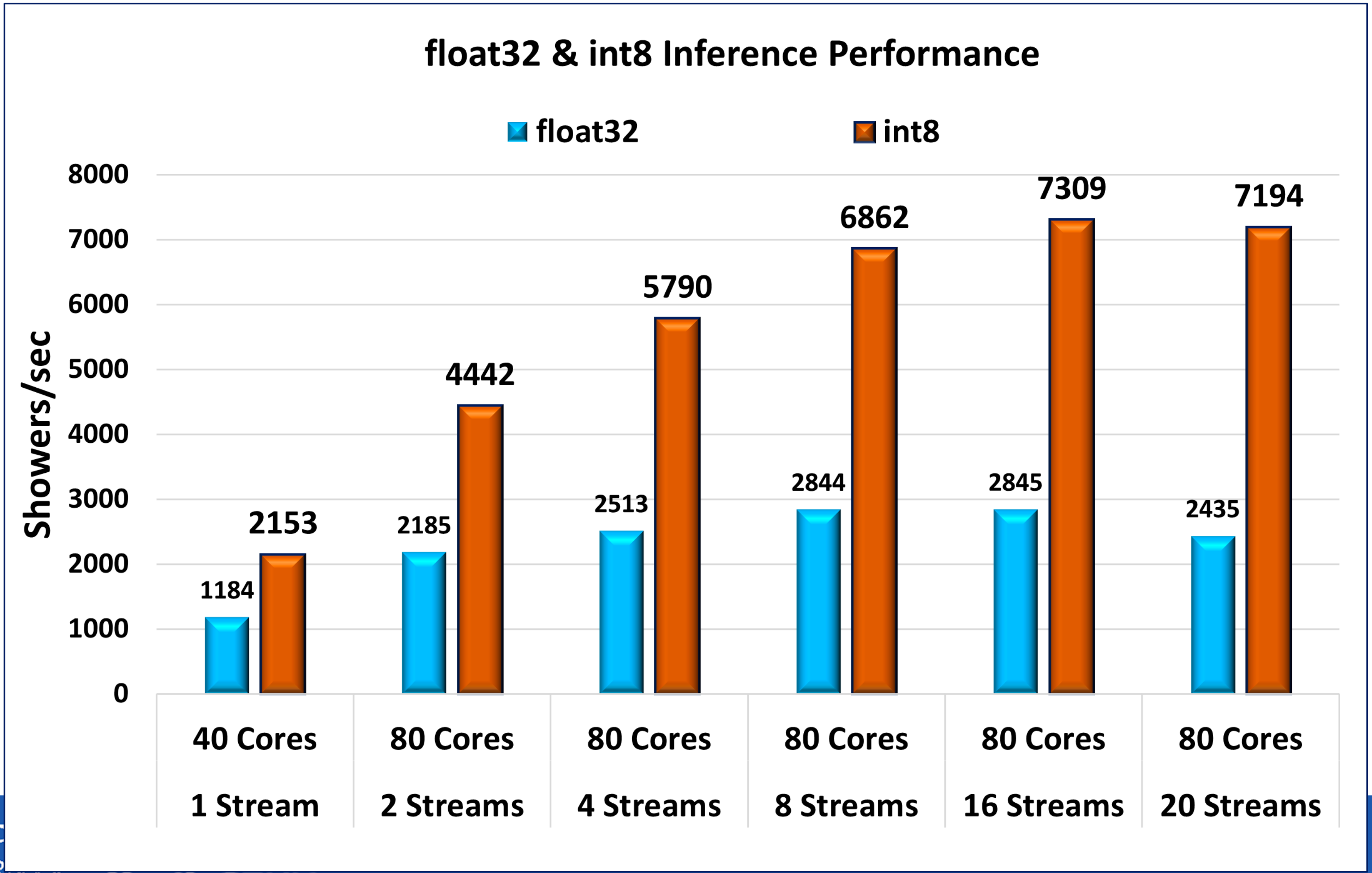


Energy pattern along transver detector axis

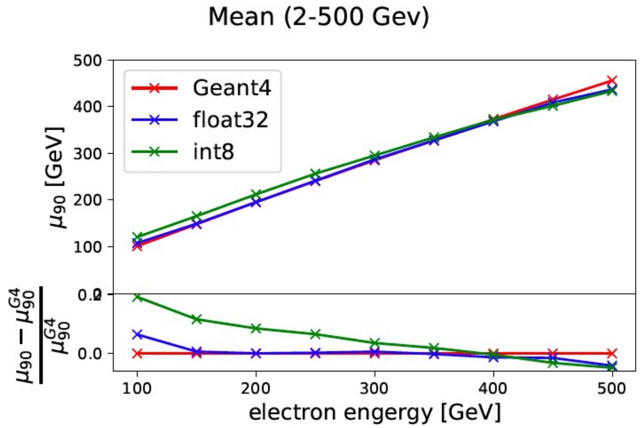
F. Rehm, ICPRAM2021
in collaboration with Intel

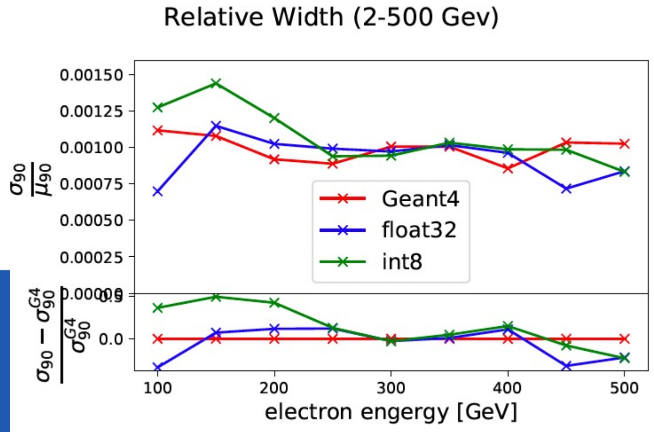
Faster then Monte Carlo?

Post-training quantization (INT 8) using **Intel DLBoost** and **Intel Neural Compressor**



Intel Ice Lake 2S Xeon 8380
Python 3.6.8
Intel optimized TensorFlow 2.3.0
batch size = 128





Development directions

ML/DL have their origins in the studies on the human brain, but today DL doesn't learn like humans do.
Current research in DL tries to improve on this aspects

G. Hinton, Y. Le Cunn, Y. Bengio , AAAI 2020 keynotes, Turing Award Winners Event
https://www.youtube.com/watch?v=UX8OubxsY8w

- New improvements will not be achieved by simply making models **larger and larger**
- **Alternative architectures** and **approaches to learning** :
  - **Few-shots learning**
  - **Self-Supervised Learning**
  - **Meta-Learning**
- **Generalisation** to different data distributions (out-of-distribution generalisation)

openAI GPT-3 as a foundation model

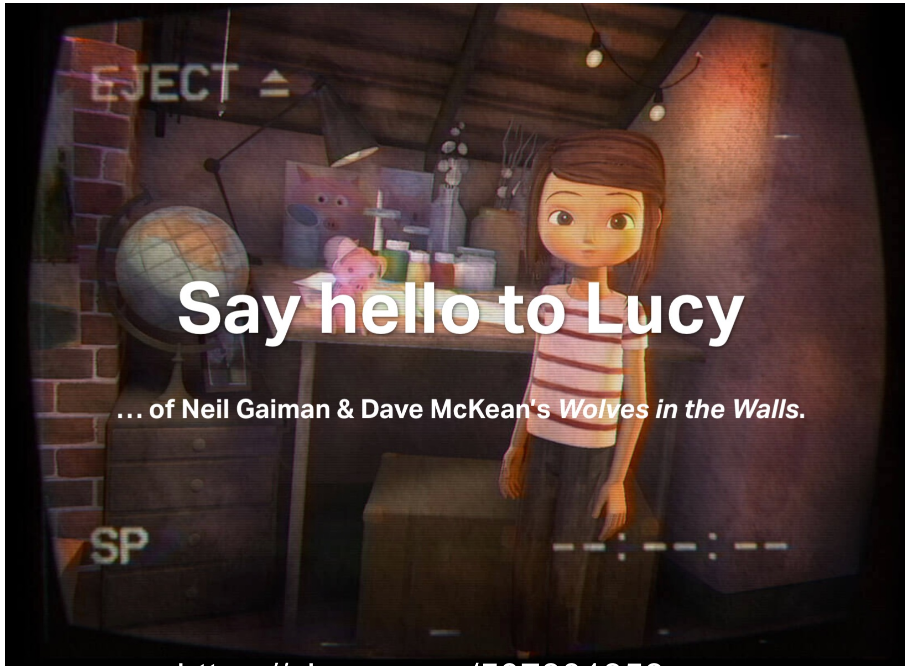
Brown, Tom B., et al. "Language models are few-shot learners." *arXiv preprint arXiv:2005.14165* (2020).

Generative Pretrained Transformer-style autoregressive model

- **175 billion parameters**
- Previously largest model was **Microsoft's Turing NLG**, with 17 billion parameters (Feb. 2020)

Trained with large Internet data sets to perform multiple **downstream tasks**

- A “foundation” model



Can we build **foundation models** for detector simulation?

Bommasani, Rishi, et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021)

Summary

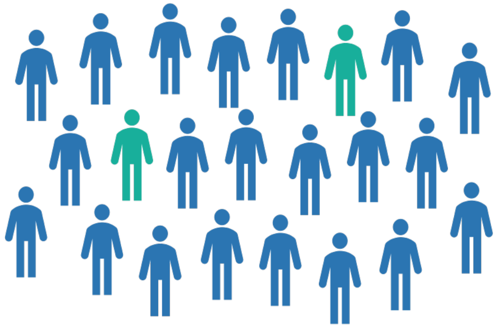
- Research in the past few years has proven **generative models can be used for (fast) simulation**
  - **Large range of applications** beyond detector simulation: direct analysis-level event generation, reconstruction-level features, optimisation,…
- Efforts needed to achieve **full integration** in simulation frameworks
  - Lots of initiative and already some results
- Those are **very exciting times** for Deep Learning
  - We should continue core research on models
  - Follow general research directions and apply them to our field.

Thanks!

Sofia.Vallecorsa@cern.ch

Jaruskova, Kristina, and Sofia Vallecorsa. "Estimating the Support Size of GANs for High Energy Physics Detector Simulation."ML4PS NEURIP202

Systematics: Support size



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

- Learnt distribution not representative enough
- Use Birthday paradox test to measure GAN support size

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

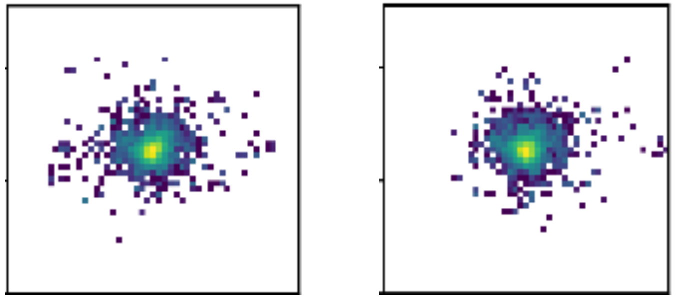
Exact duplicates

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