Benchmark of GANs for Fast HEP Calorimeter Simulations

GRID 2021

Florian Rehm [CERN openlab, RWTH Aachen]

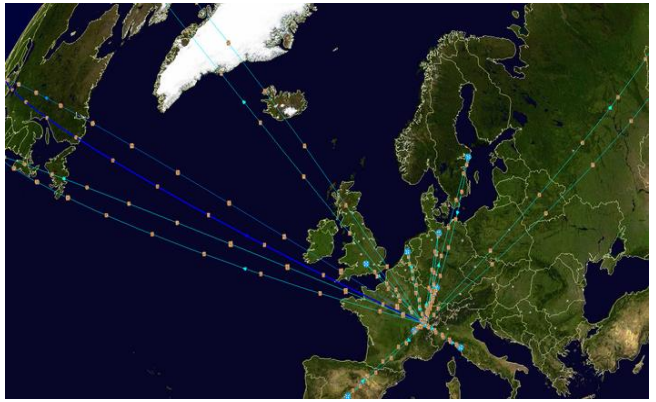
Sofia Vallecorsa [CERN openlab], Kerstin Borras [DESY, RWTH Aachen], Dirk Krücker [DESY]

Calorimeter Simulations

- Calorimeter detectors measure the energy of particles
- Calorimeter simulations are based on Geant4
- Simulations use about 50% of the resources of the worldwide LHC grid
- LHC high luminosity phase requires 100 times more simulated data *

→ Develop a new approach which occupies less resources

→ Employ deep learning

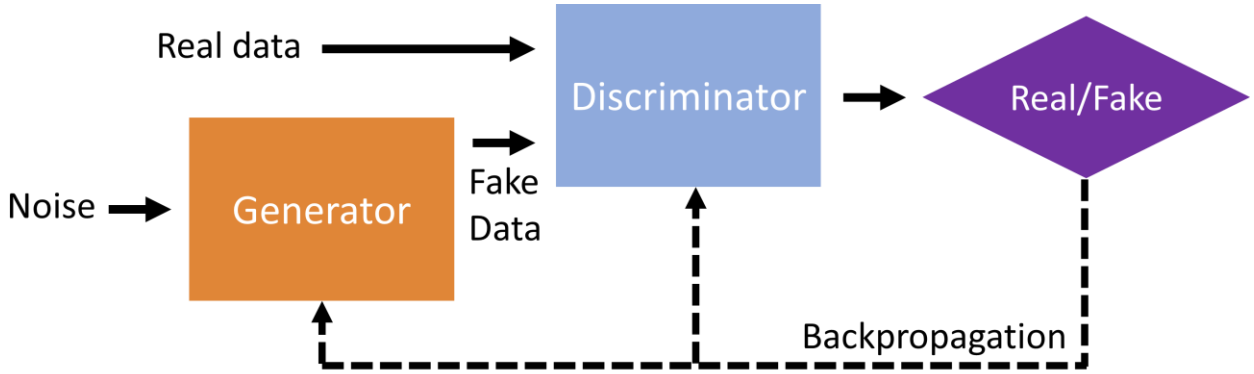


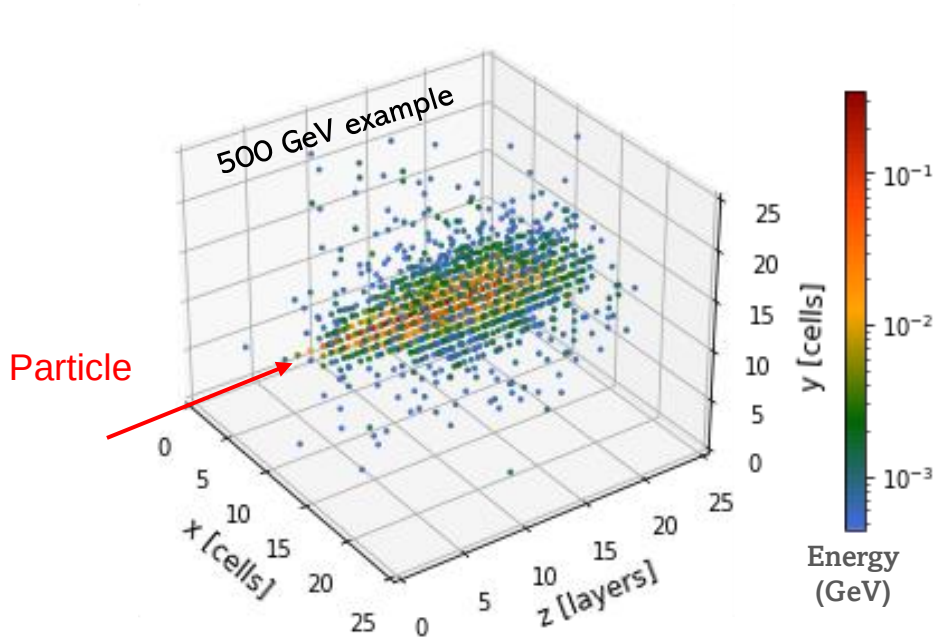
*A Roadmap for HEP Software and Computing R&D for the 2020s https://doi.org/10.1007/s41781-018-0018-8

Generative Adversarial Networks

3DGAN

- Train two networks (Generator & Discriminator) in a minmax game





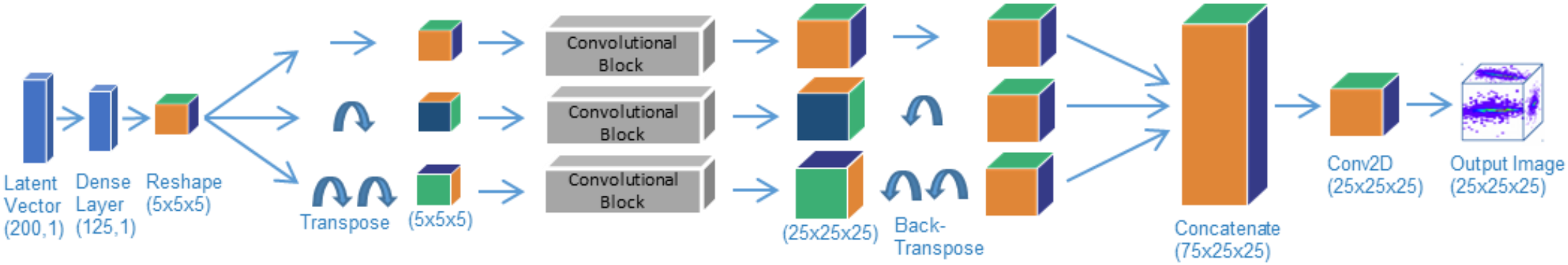
- 200 000 3D shower images with granularity 25x25x25
- Energies between 2-500 GeV

New Conv2D Generator Architecture

- Conv3D layers are computational demanding
- Conv3D layers are not yet supported in less than 32bit precision

  → Creating neural network consisting only of Conv2D layers

  → New Conv2D architectures

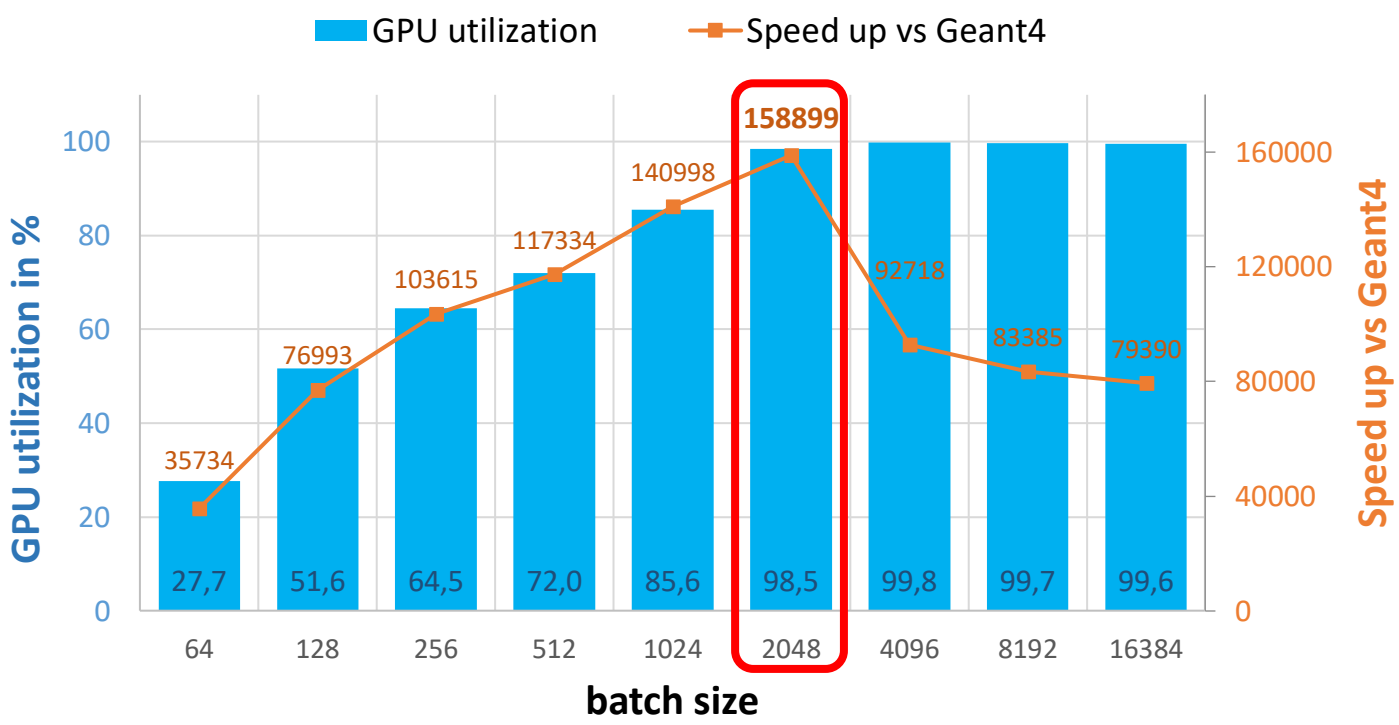


→ Solve 3D image problems with only 2D convolutional layers

New Conv2D Model Evaluation

Computational Evaluation 1

Conv2D Model Inference



- Geant4 takes 17s to reproduce 1 shower *
- Conv2D model **158 000x** faster
  - GPU: Nvidia Tesla V100
  - Optimized for reaching highest inference times

*Geant4 simulation time from a previous measurement in 2018: Book title: “High Performance Computing”, Springer, Chapter: “Distributed Training of Generative Adversarial Networks for Fast Detector Simulation”

Computational Evaluation 2

Comparison Conv2D vs Conv3D

Model	Number of Parameters	Speed up vs Geant4	GPU Utilization
Conv3D	752 000	6 200x *(not optimized)*	78.75%
Conv2D	2 055 000	158 000x	98.50%

- Inference of Conv2D model is faster
  - However, Conv3D inference not optimized for batch size on GPU
- New Conv2D model has more than double as much parameters as the previous Conv3D model
  - → Potential for reaching higher physics accuracy

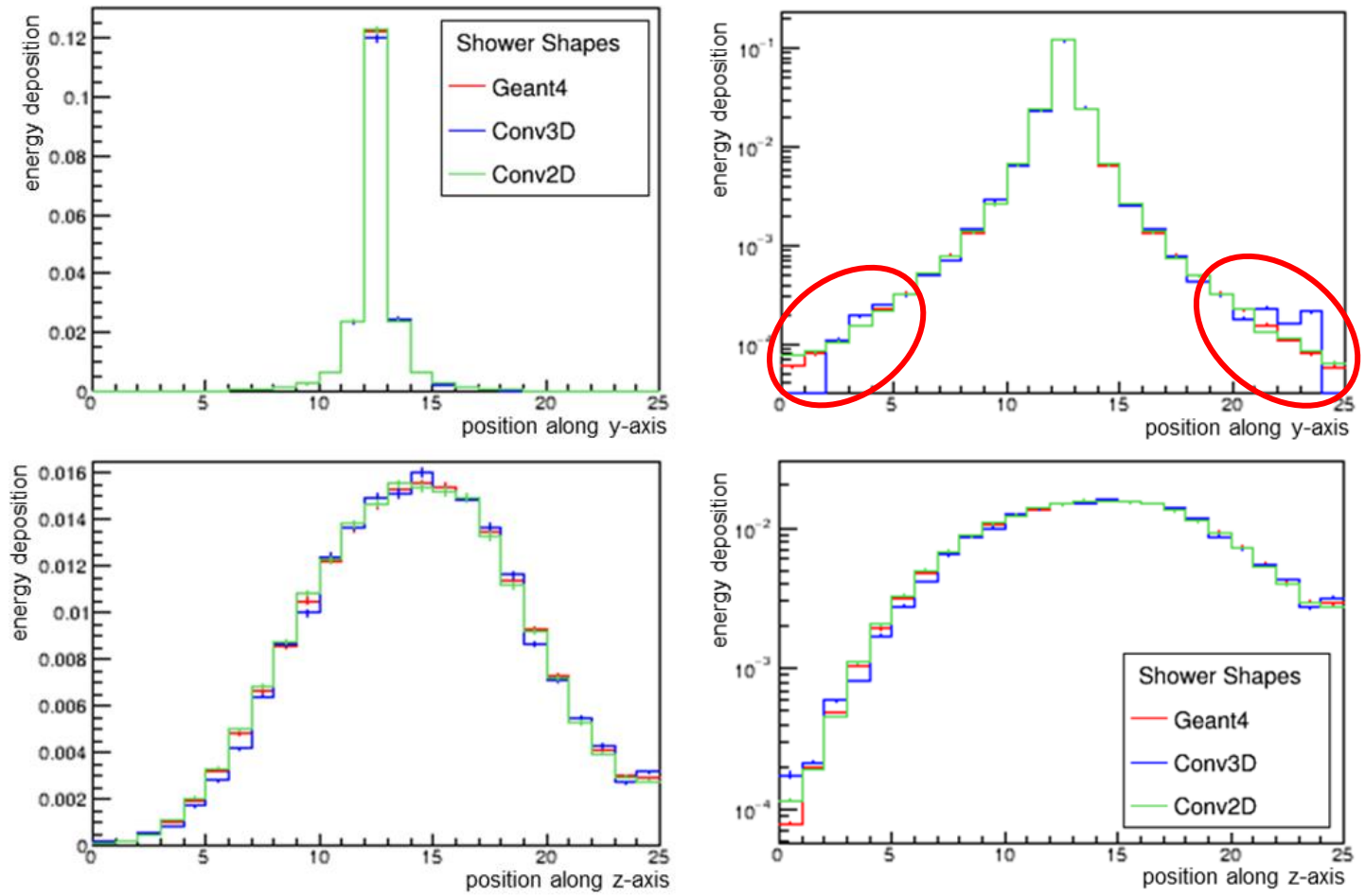
Physics Evaluation 1

Shower Shapes

- Cell wise the Mean Squared Error (MSE) between GAN and validation data along the three canonical axes:

Model	MSE (Lower is better)
Conv3D	0.065
Conv2D	**0.027**

- Projection of the shower along the different axes
  - Conv2D performs better along the tails

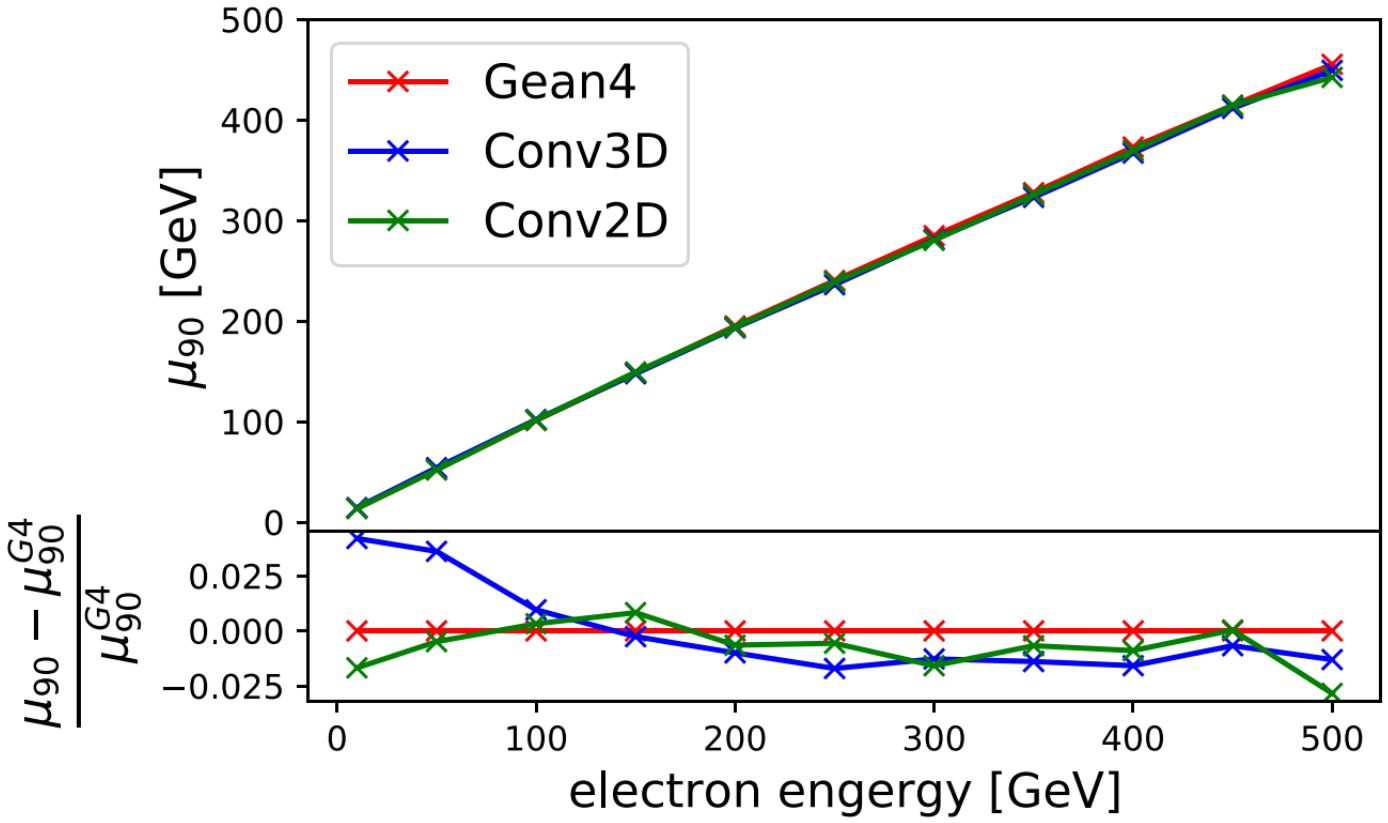


Physics Evaluation 2

Deposited Mean Energy

- Deposited energy with respect to the incident electron energy
- μ_{90}: 90% core of the distributions

→ Conv2D model performs better for lower energies



Reduced Precision Computing

Reduced Precision Computing

- Quantization: Converting a number from a higher to a lower format
  - E.g. from float32 to int8

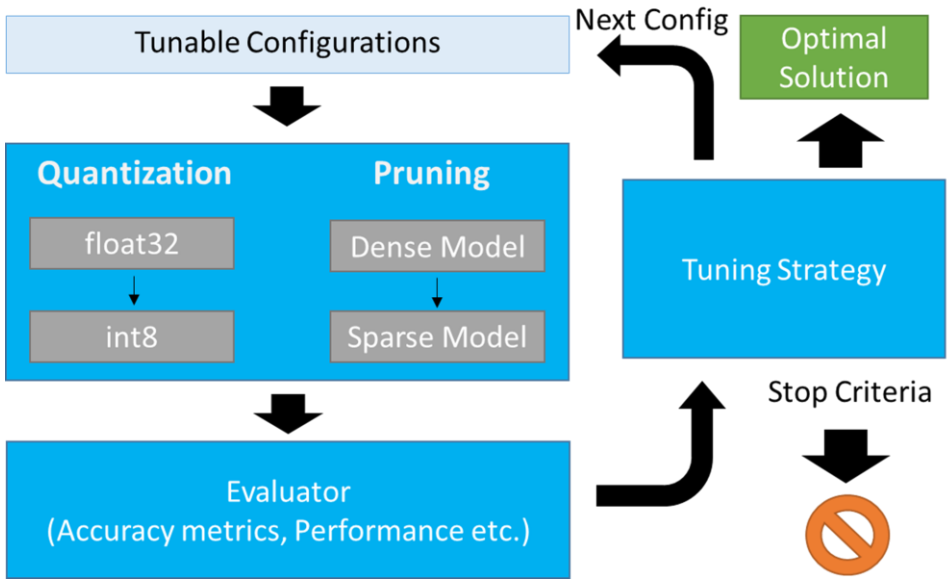
Float32		Int8
4 byte	→	1 byte
Max Number: $3.4 * 10^{38}$		Max Number: 255

- Quantization Tool:
  Intel Low Precision Optimization Tool (iLoT)
  - Now renamed to LPOT

https://github.com/intel/lp-opt-tool

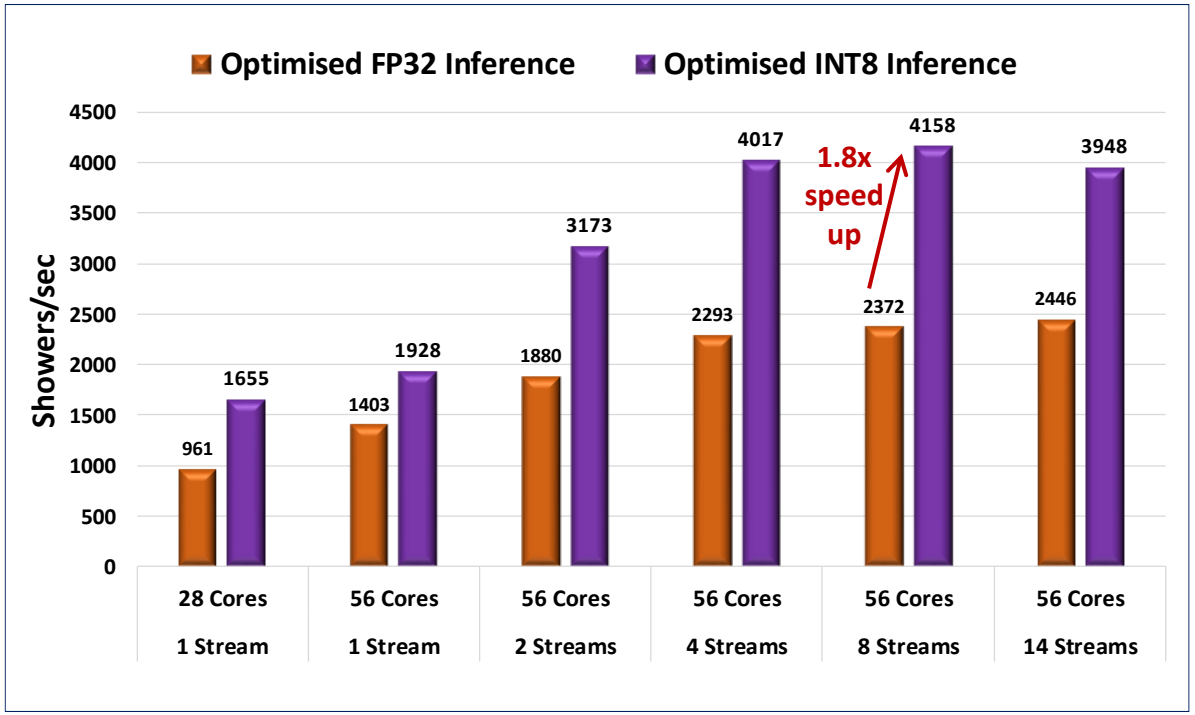
- Reference Tool: TensorFlow Lite

https://www.tensorflow.org/lite



Computational Evaluation

(of iLoT model)



- **1.8x** speed up due to quantization

Test run on CPU: Intel 2S Xeon Processor 8280 with Cascade Lake architecture and 28 cores (56 virtual cores or threads)

- Total speed up of **68 000x** versus Geant4

Model	Speedup vs Geant4
float32	38 000x
int8	68 000x

- Reduction in model memory size of **2.26x**

Model	Memory [MB]
float32	8.08
int8	3.57

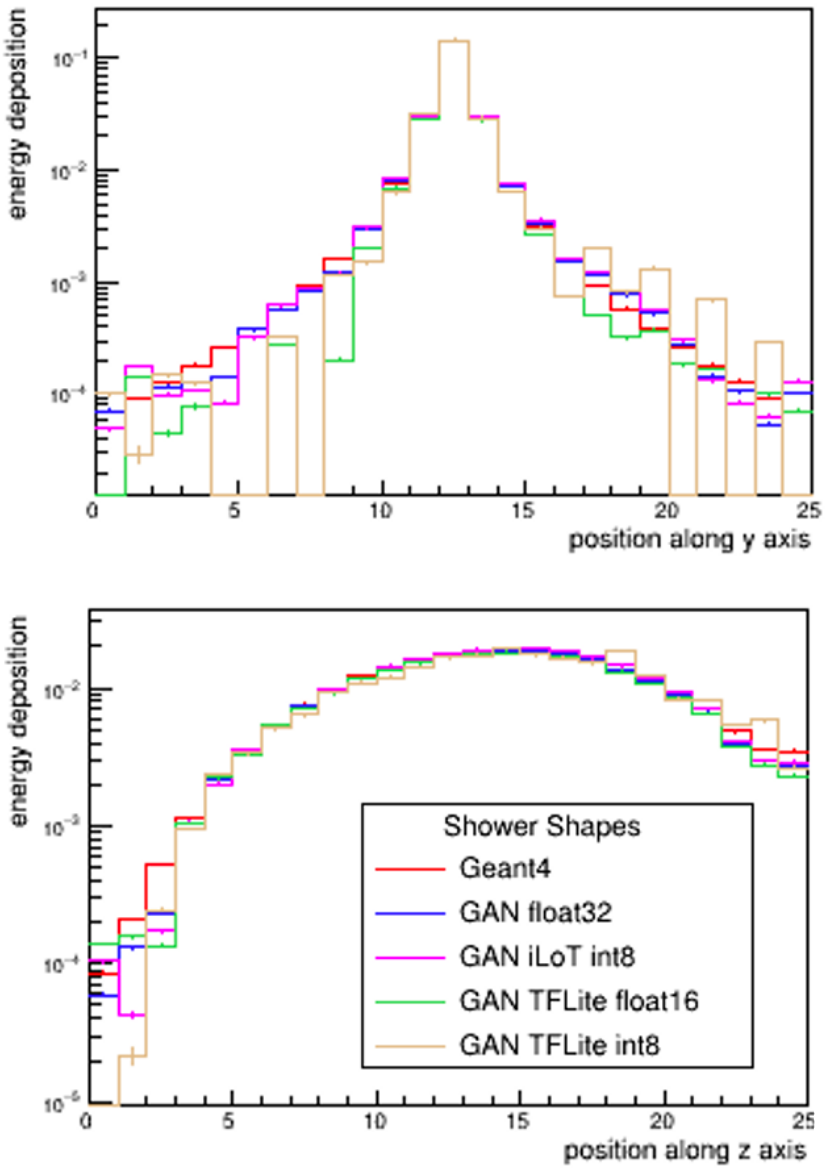
Physics Evaluation

Shower Shapes

- Mean Squared Error (MSE) between GAN and validation data

Model	MSE (Lower is better)
float32	0.061
iLoT int8	**0.053**
TFLite float16	0.253
TFLite int8	0.340

- ➢ iLoT shows a good accuracy
- ➢ TensorFlow Lite performs worse



Summary

- New Conv2D architecture:
  - ✓ **158 000x** speed up vs recent Geant4 simulations
  - ✓ **Better physics accuracy** than previous Conv3D model

- Quantized iLoT int8 model:
  - ✓ Further **1.8x** speed up
  - ✓ Keeping **good physics accuracy**

QUESTIONS?

Benchmark of GANs for Fast HEP Calorimeter Simulations

Florian Rehm [CERN openlab, RWTH Aachen]

Sofia Vallecorsa [CERN openlab], Kerstin Borras [DESY, RWTH Aachen], Dirk Krücker [DESY]

Back Up Slides

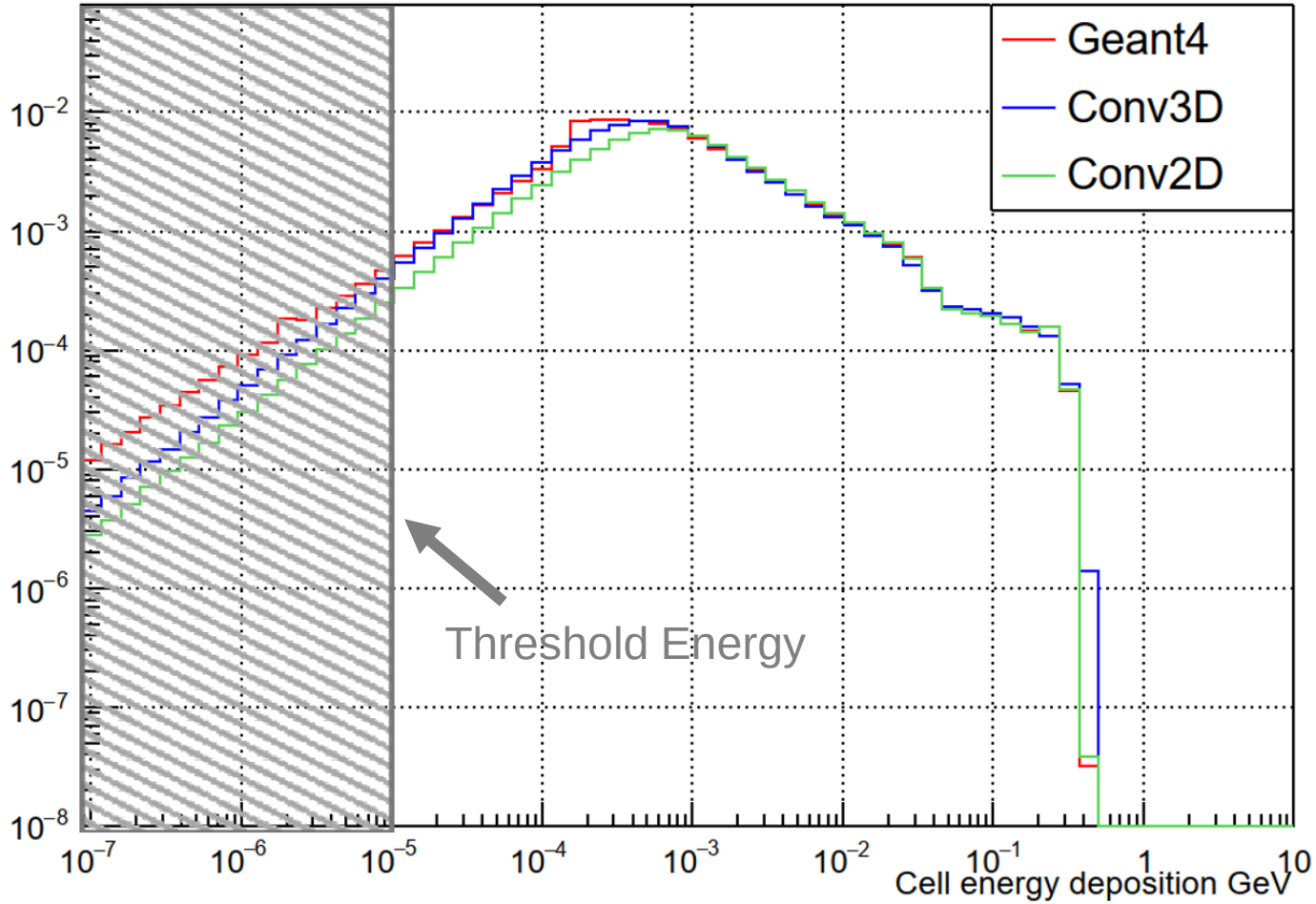
Physics Evaluation 3

Histogram of Single Cell Energy Deposition

- Conv2D model performs slightly worse for lower energy cells

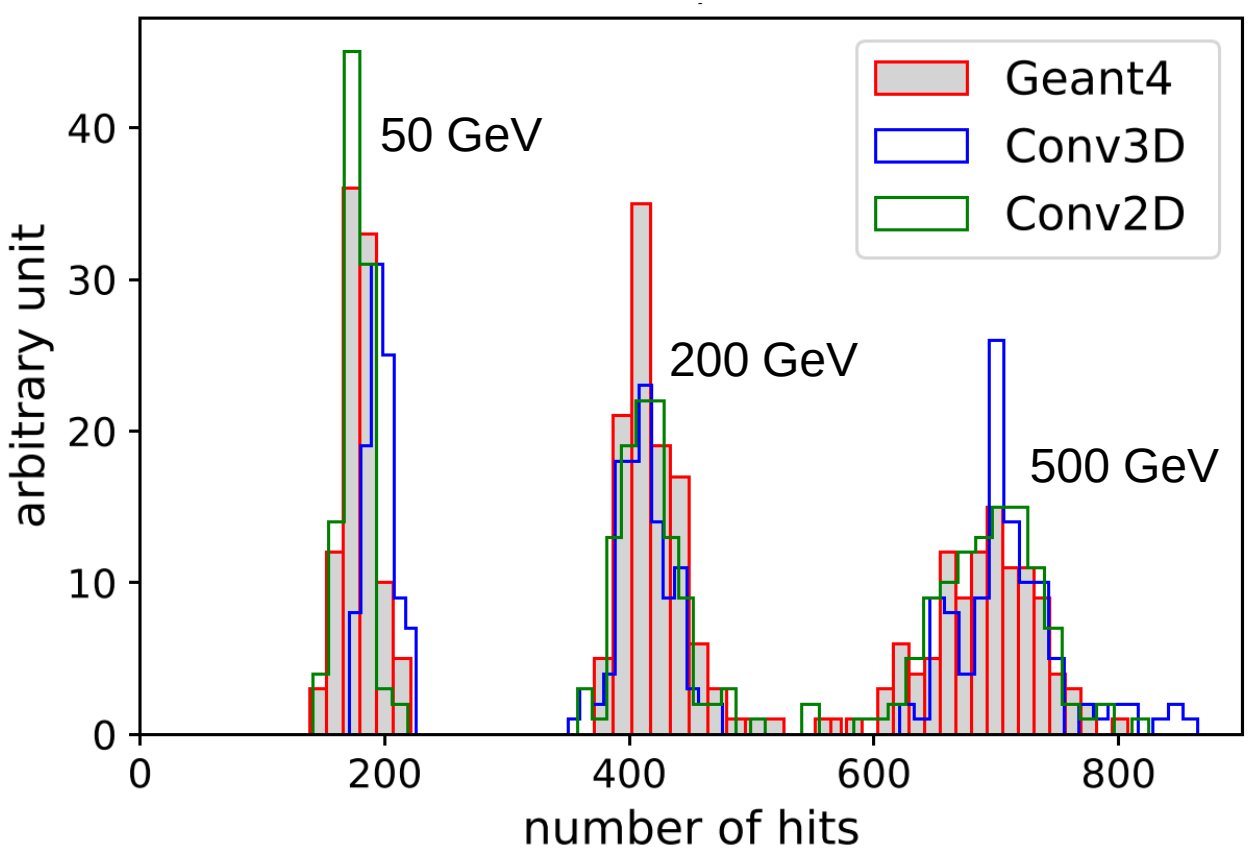
→ Further investigations

- Minimal energy threshold at 10^{-6} GeV because of detector resolution
- Practice: 10^{-6} GeV energy threshold



Physics Evaluation 4

Number of Hits

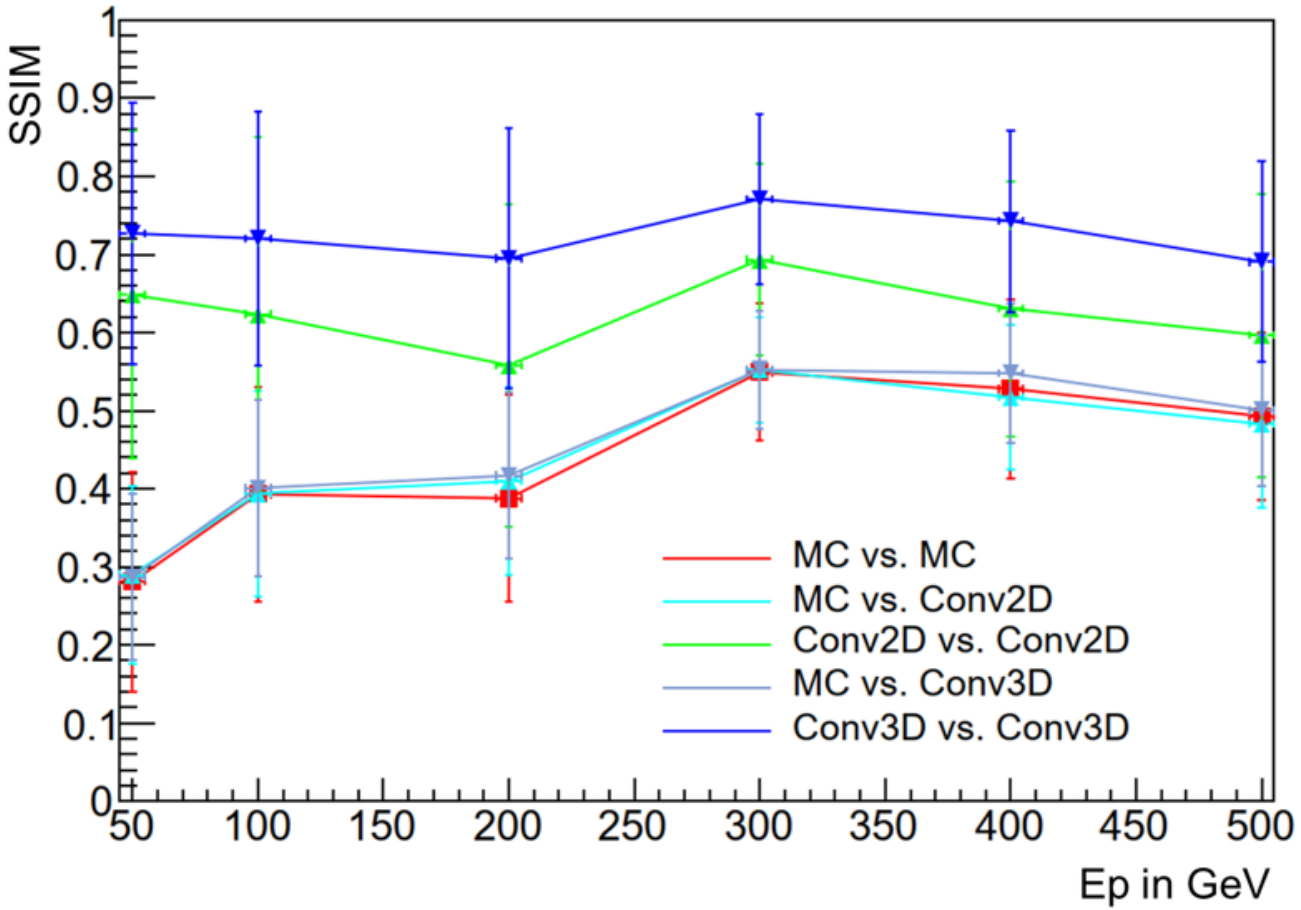


- Number of hits:
  - "How many cells receive energy above a certain threshold"
- Primary particle energies of 50, 200 and 500 GeV
  - Averaging the distributions over ±2 GeV bins

→ Both models show a good performance

Physics Evaluation 5

SSIM



Structural Similarity Index

- Known problem of GANs: Mode collapse
- SSIM estimates the perceptual difference between similar images

→ Conv2D performes better