(Quantum) Machine Learning for Calorimeter Simulations

DESY-CMS Meeting

Florian Rehm [CERN openlab, RWTH Aachen]

Sofia Vallecorsa [CERN openlab], Kerstin Borras [DESY, RWTH Aachen], Dirk Krücker [DESY],
Vikram Saletore [Intel], Hans Pabst [Intel], Adel Chaibi [Intel],

29.04.2021

Content

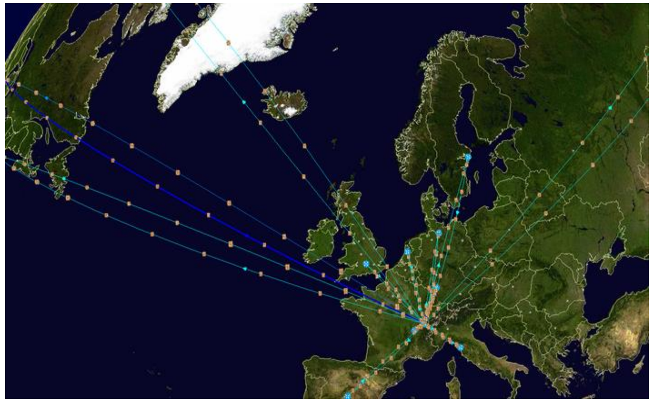
- Quick introduction to computing constraints of calorimeter simulation
- Speeding up simulations using machine learning
  - Generative Adversarial Networks (GAN)
  - Physics validation
- Reduced precision computing
- Quantum computing
  - Possible advantages of quantum computing over classical computing
  - First steps: 1D QGAN
  - Recent and future work: 2D QGAN

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**

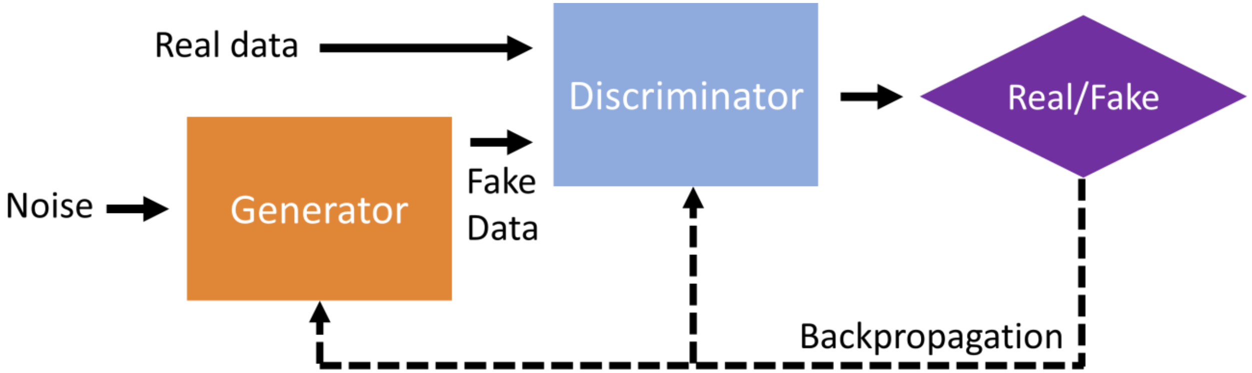


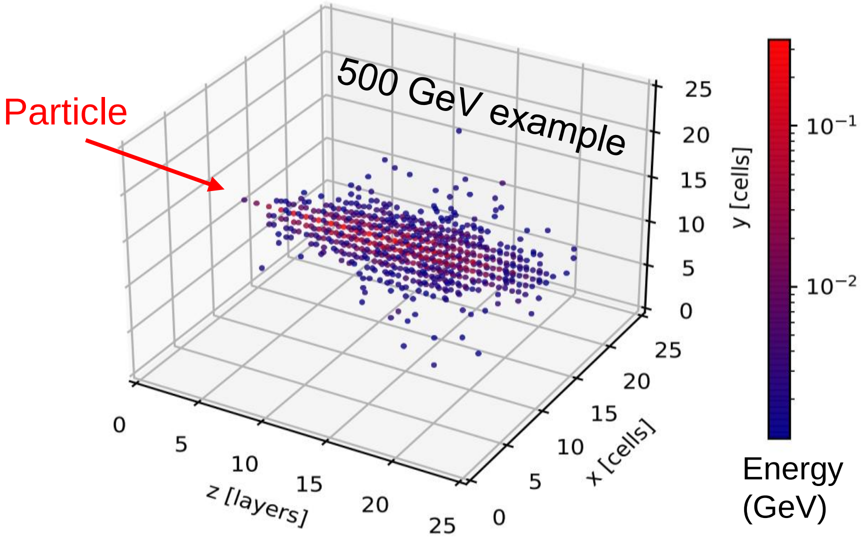
*A Roadmap for HEP Software and Computing R&D for the 2020s https://doi.org/10.1007/s41781-018-0018-8

Generative Adversarial Networks

3DGAN

- Train two networks (Generator & Discriminator) in a minmax game
- We want to further decrease the computational resources



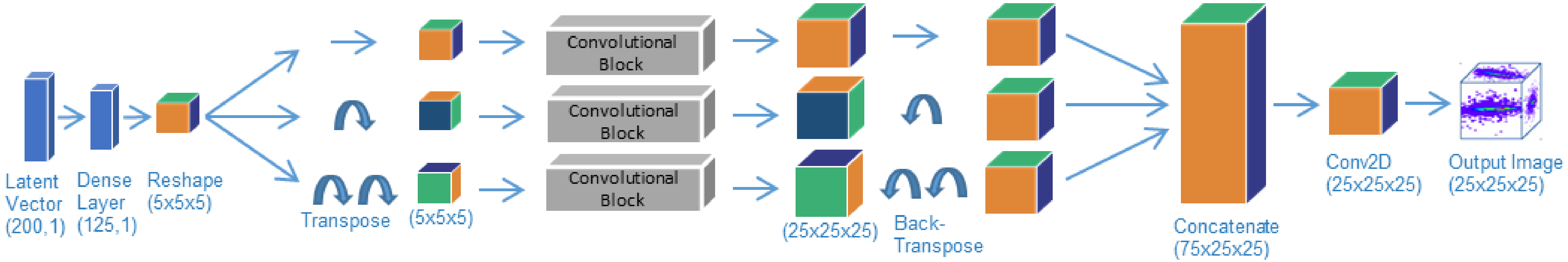


- 200 000 3D shower images with granularity 25x25x25
- Energies between 2-500 GeV

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



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

Computational Evaluation

Model:	Number of Parameters	Inference Time [s]	Speed up vs Geant4
Conv3D	965 000	16.8	11 100x
Conv2D	2 055 000	4.9 (3.4x faster)	38 000x

- New Conv2D model has more than double as much parameters as the previous Conv3D model
- Inference of Conv2D model is faster
- → **38 000x speed up** vs Geant4 simulation

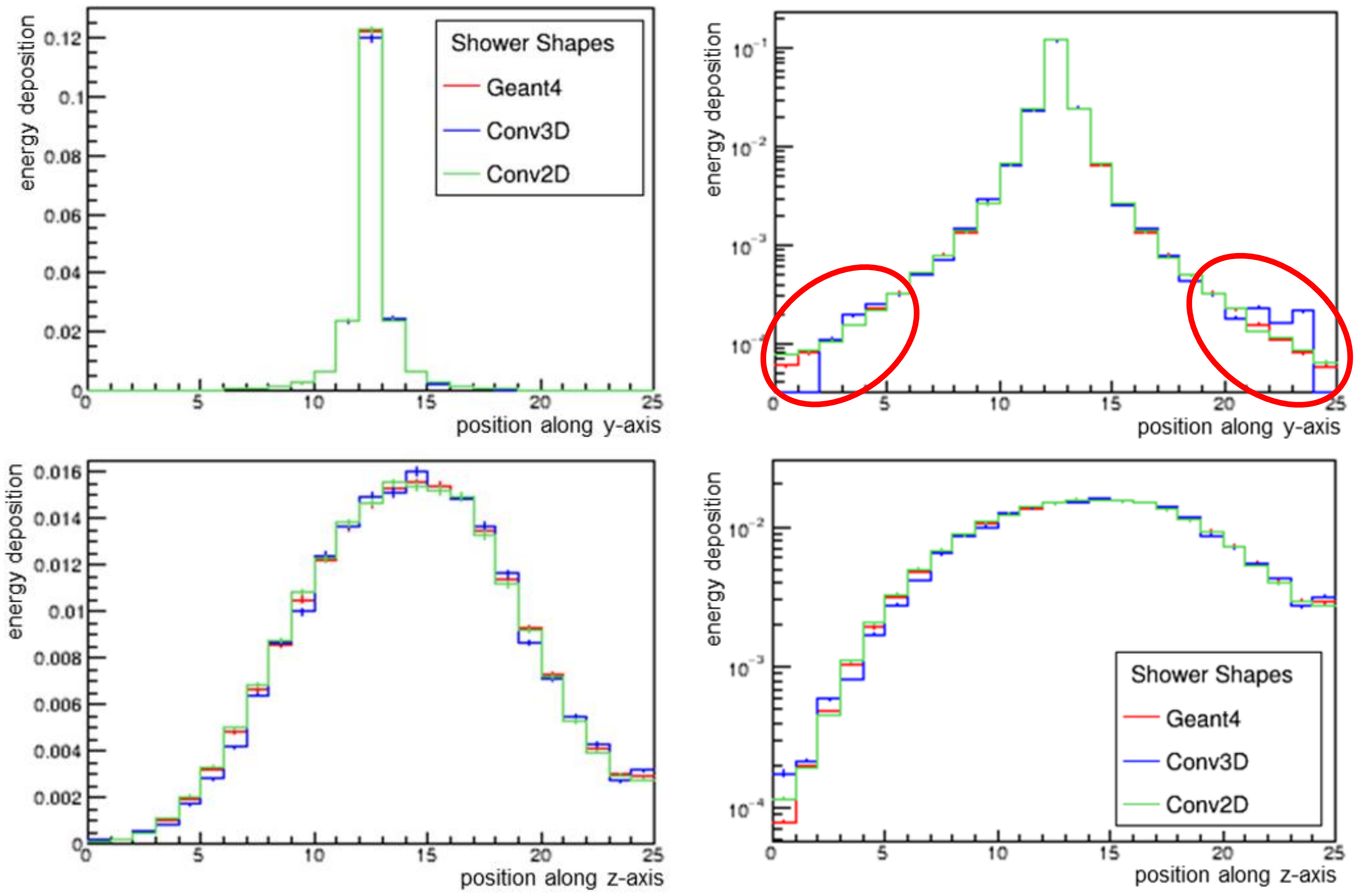
Physics Evaluation

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	0.048
Conv2D	**0.027**

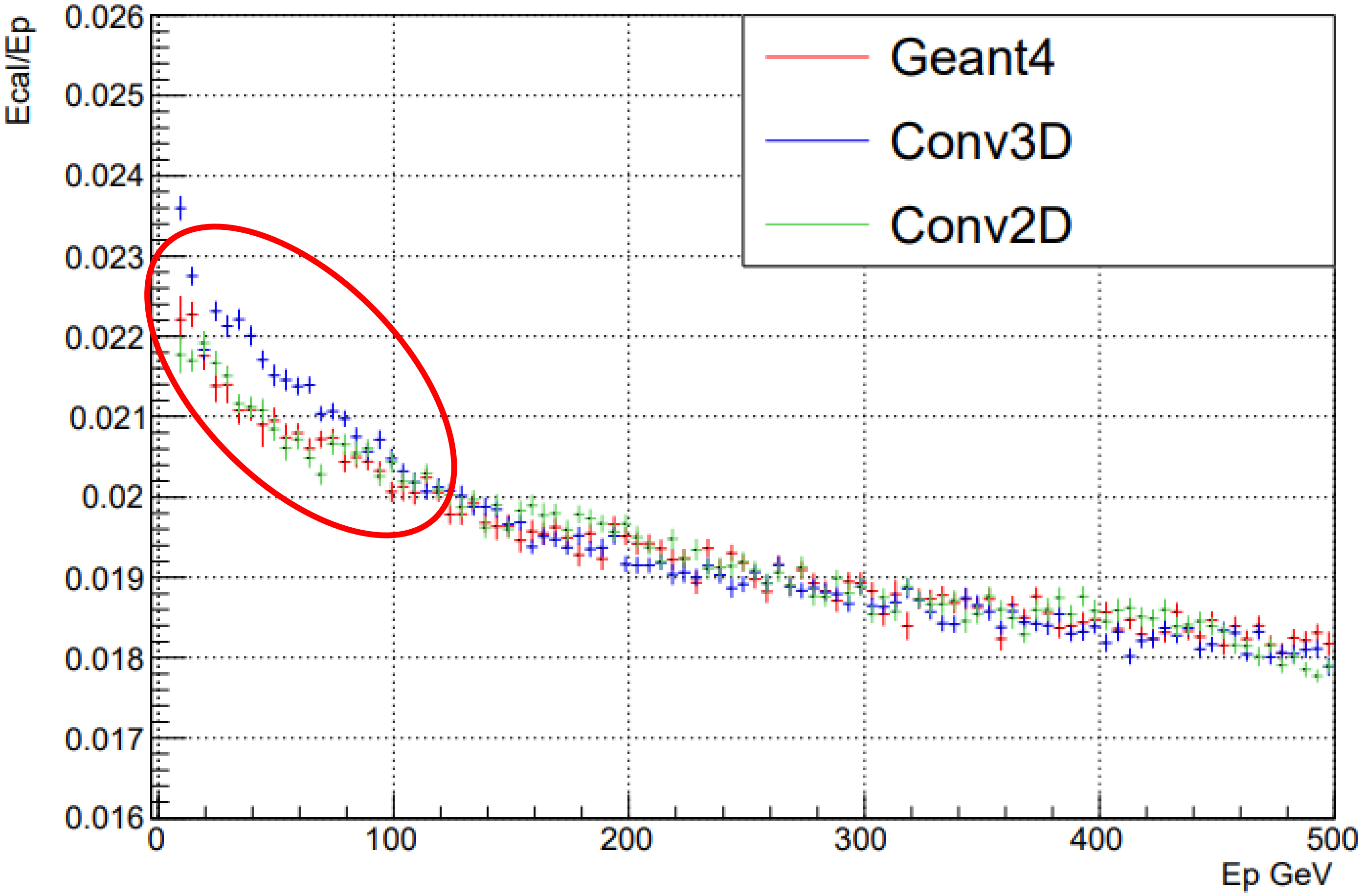
- Projection of the shower along the different axis
  - Conv2D performs better along the tails



Physics Evaluation

- Ratio between the total measured energy ECAL and the initial particle energy E_p

- Conv2D performs better for energies below 100 GeV



Reduced Precision Computing

Reduced Precision Computing

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

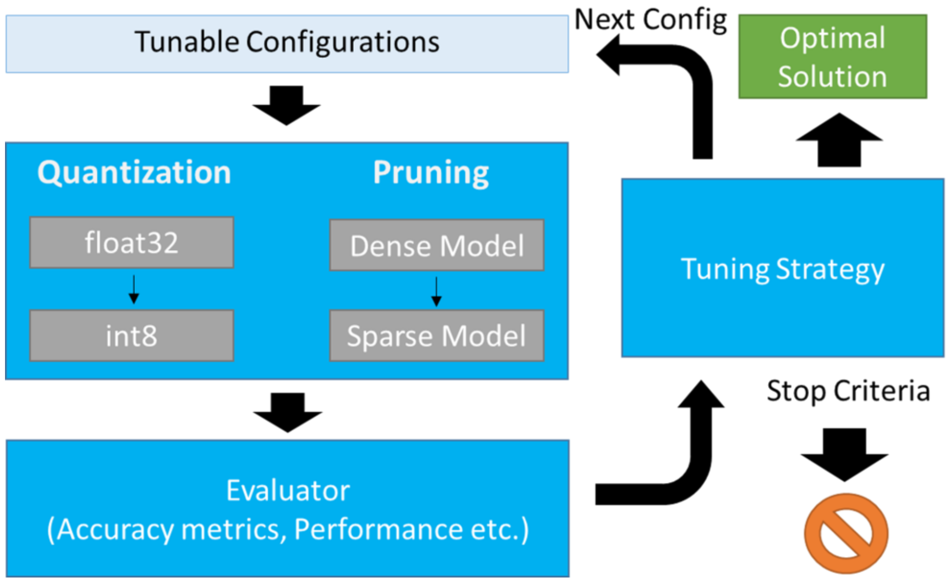
Float32
4 byte
Max Number: $3.4 * 10^{38}$

→

Int8
1 byte
Max Number: 255

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

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

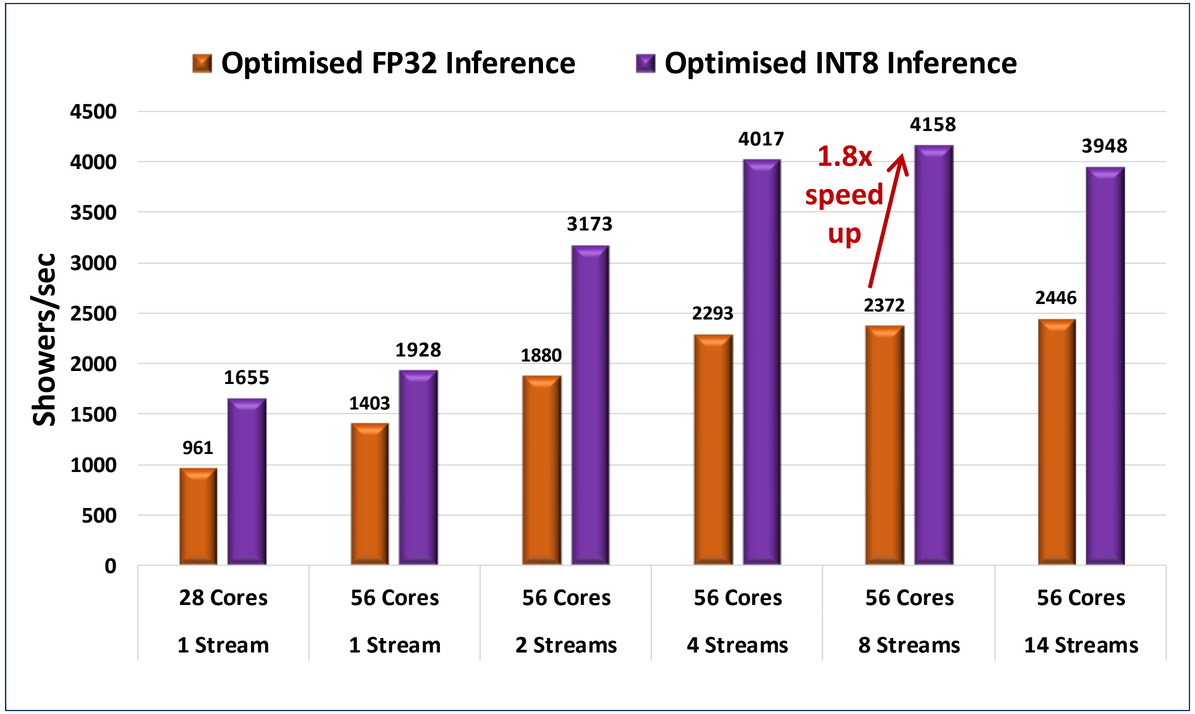


- Reference Tool: TensorFlow Lite

https://www.tensorflow.org/lite

Computational Evaluation

(of iLoT model)



- **1.8x** speedup due to quantization

- Total speedup of **68 000x** versus Monte Carlo

Model	Speedup vs Monte Carlo
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

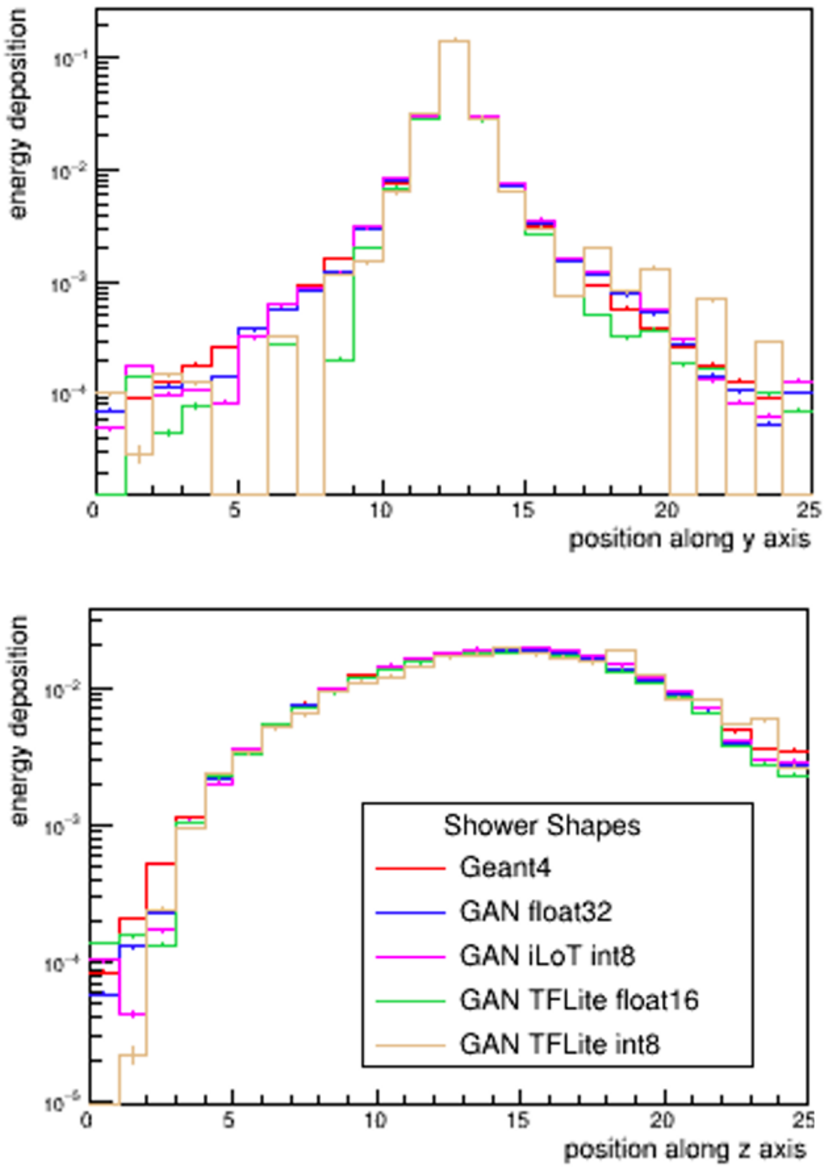
Shower Shapes

- 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 Machine Learning

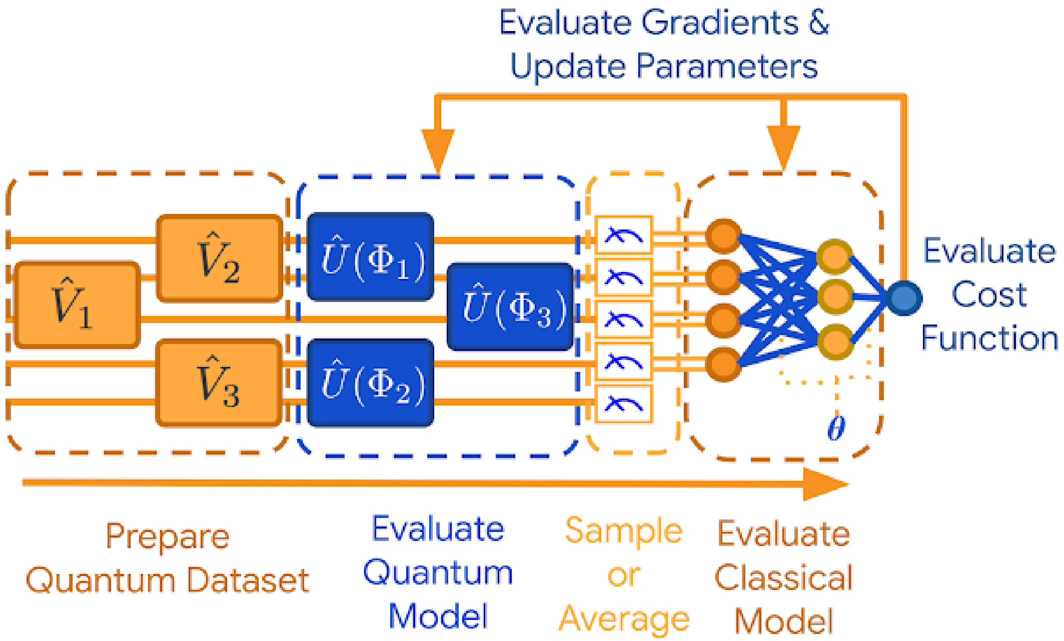
- New Conv2D network: **3.4x** speed up + better physics accuracy
- Quantized iLoT model: **1.8x** speed up + good accuracy

→ Total **68 000x** speed up vs Geant4
\+ maintaining physics accuracy

Quantum Computing

Advantages of Quantum Computing

- Use of quantum properties
  - Entanglement
  - Superposition
- Hope to solve problems faster and / or more accurately
- "Quantum Advantage" not yet reached
  - ➢ Initial investigations: aimed at understanding advantages and challenges
- Hybrid models:
  - Quantum circuit + classical optimization



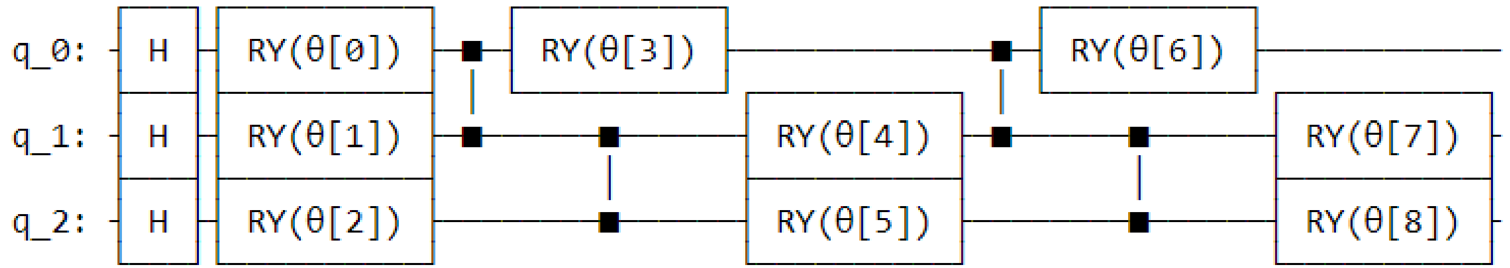
https://ai.googleblog.com/2020/03/announcing-tensorflow-quantum-open.html

Hybrid Quantum GAN Model

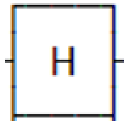
QGAN 1D

- Hybrid GAN model:
  - quantum generator
  - classical discriminator
  - classical optimization
- Only 1D 8-pixel images
  - 3 qubits ($2^3 = 8$) in quantum generator circuit
- Run on quantum simulator

Quantum Generator Circuit:

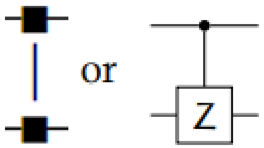


Used Gates:



$$= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

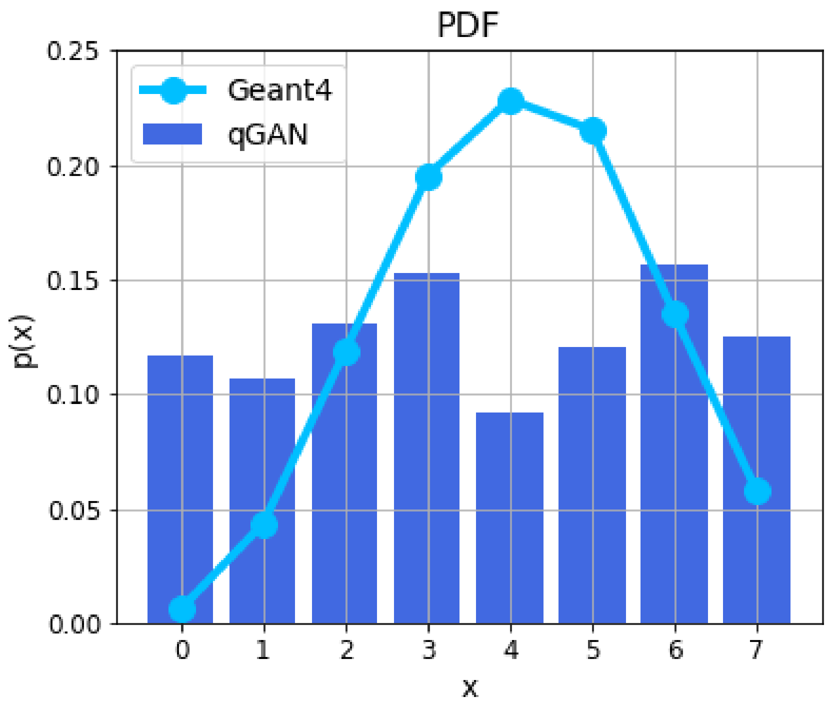
$$R_y(\theta) = \begin{pmatrix}\cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right)\\ \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right)\end{pmatrix}$$



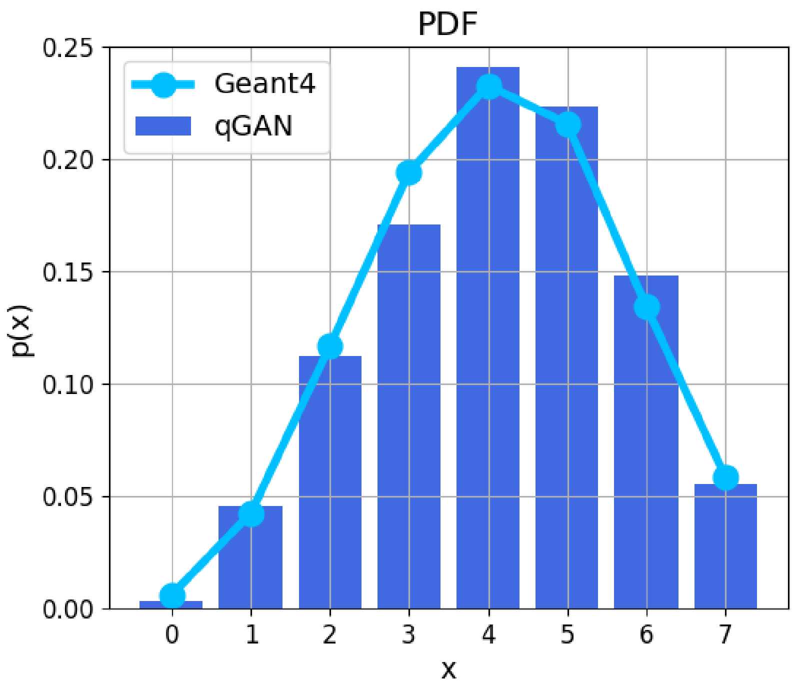
$$= CZ = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}$$

Hybrid QGAN 1D

First Results



Random Initialization



Trained Model

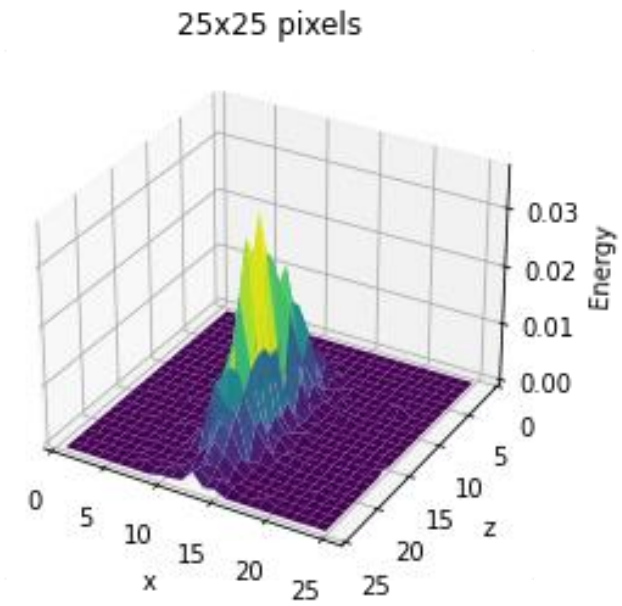
→ Good Results

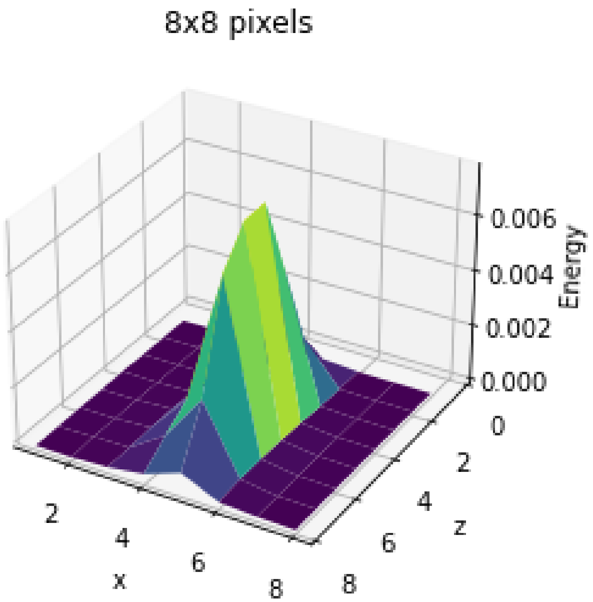
→ Moving on to more complicated data representation

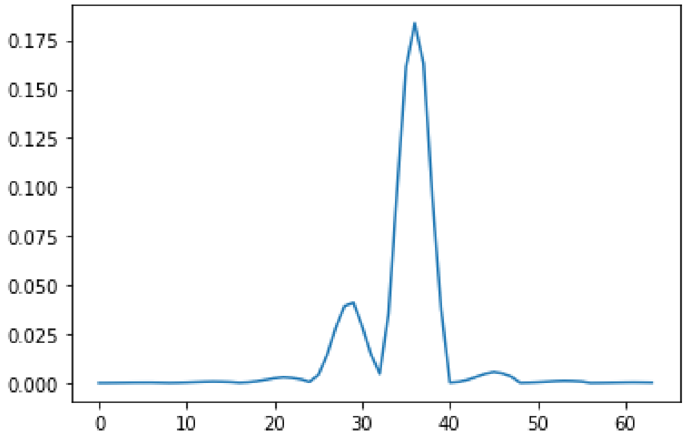
Hybrid QGAN 2D

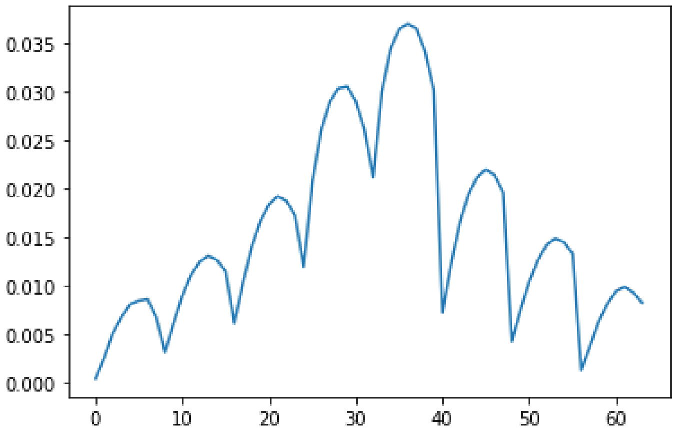
Training Data Downsampling

- $8 * 8$ pixels → 64 pixels in total
- 1D stacking
- Applying logarithm





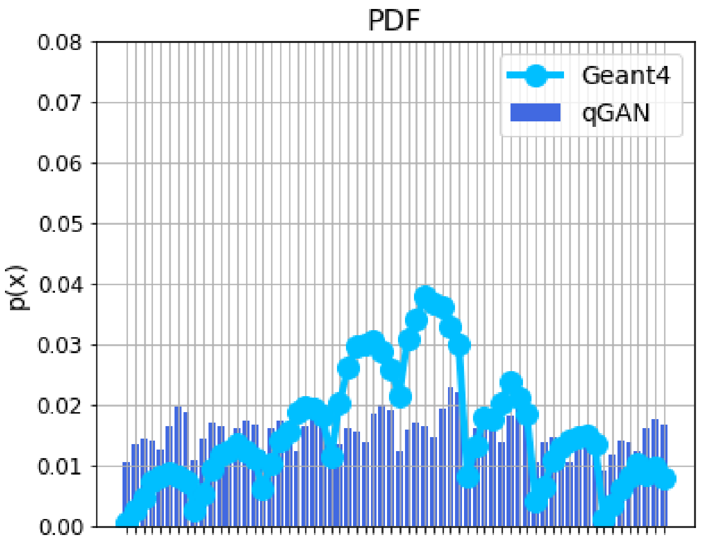




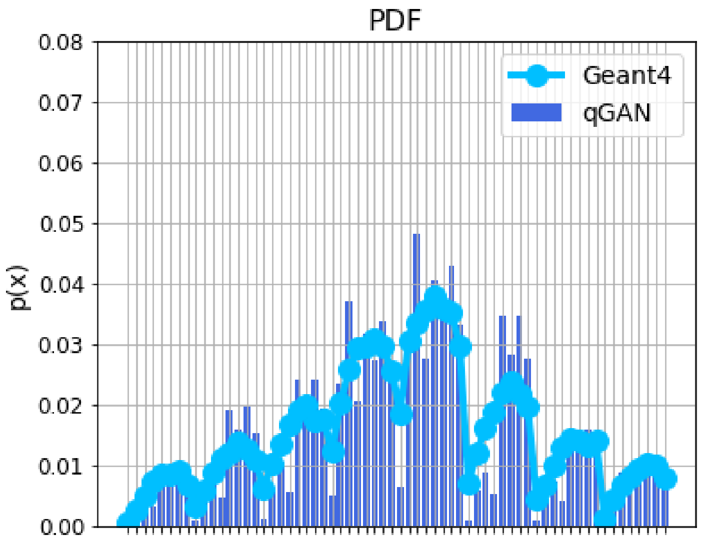
Hybrid QGAN 2D

First Results

- $8 * 8 = 64$ pixels
- 6 qubits ($2^6 = 64$) in quantum generator
- Run on quantum simulator



Random Initialization



Trained Model

→ Promising results

→ Further optimization

Quantum Computing: Future Work

- Main goal: Make calorimeter more realistic
  - Higher number of pixels
  - Conditional case: different primary particle energies
  - Different generator circuit architectures
  - Optimize generator circuit to maximum express ability
  - Different data encoding strategies
  - Discriminator as quantum circuit
  - Error mitigation of quantum simulations with noise
  - …
  - …

- Later: Run on a real quantum device

QUESTIONS?