Deep Learning for Accelerating High Energy Physics Simulations

DPG-Frühjahrstagung 2021

Florian Rehm [CERN openlab, RWTH Aachen]

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

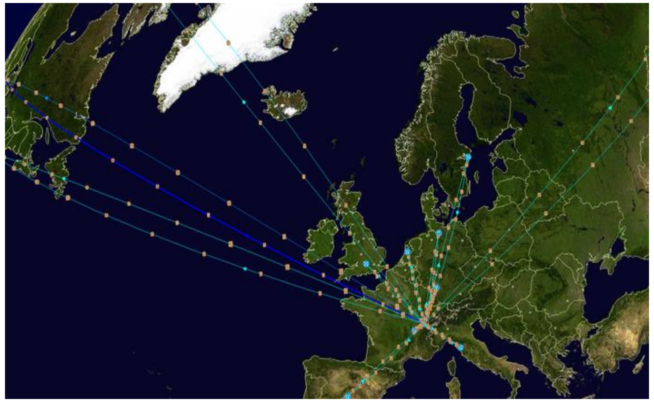
18.03.2021

Calorimeter Simulations

- Calorimeter detectors measure the energy of particles
- Calorimeter simulations are based on Geant4
- Geant4 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**

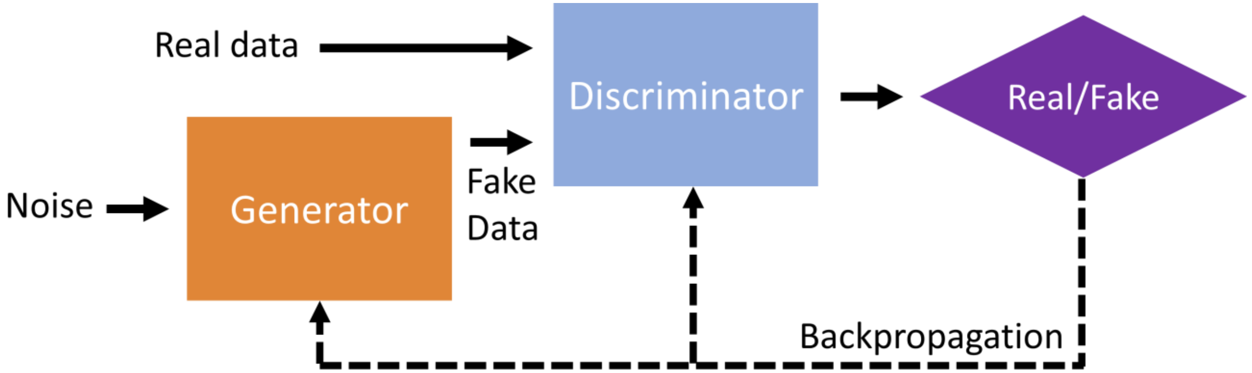


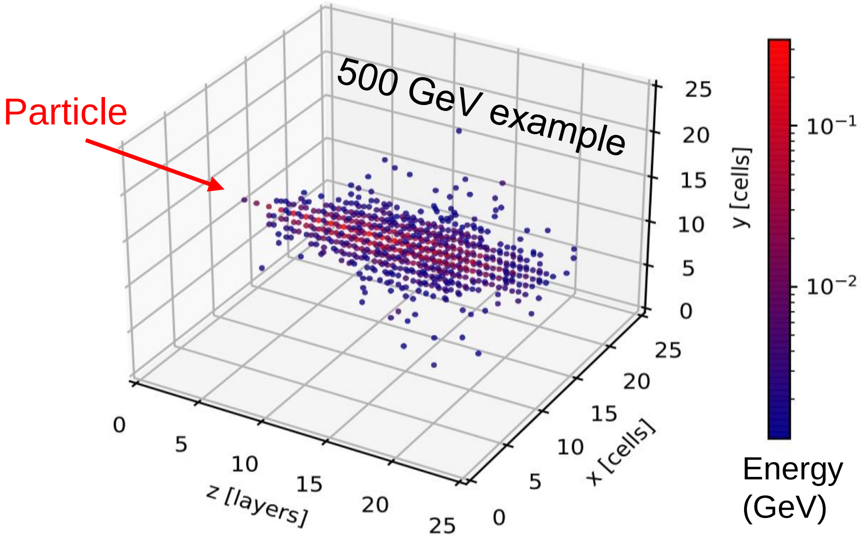
*A Roadmap for HEP Software and Computing R&D for the 2020s
https://doi.org/10.1007/s41781-018-0018-8

Generative Adversarial Networks

3DGAN

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



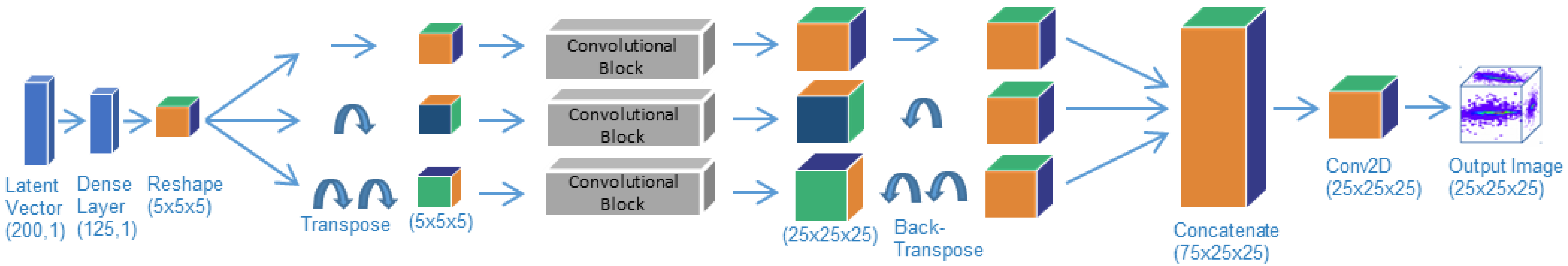


- 200 000 3D shower images with granularity 25x25x25
- Energies between 2-500 GeV

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]	GPU Utilization [%]	Training Time per Epoch [min]
Conv3D	965 000	16.8	93.15	258
Conv2D	2 055 000	4.9 (3.4x faster)	21.75 (4.3x less)	40 (6.45x faster)

- Conv2D model has more than double as much parameters as the Conv3D model and inference is much faster

- Conv2D model has lower inference time **and** lower GPU utilization
  → Potential to increase speed up to a factor of 12x when using multiple streams

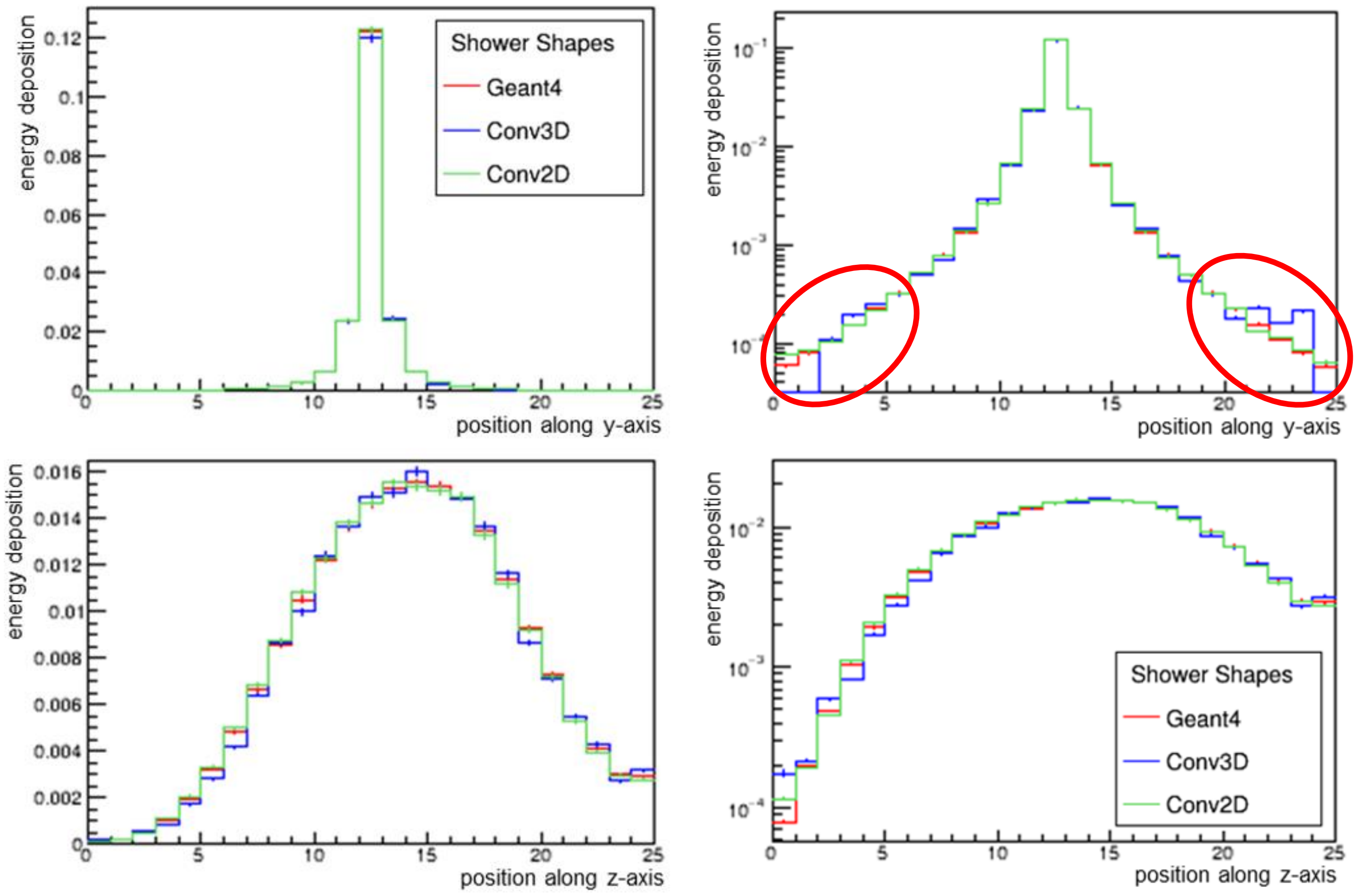
Physics Evaluation

Shower Shapes

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

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

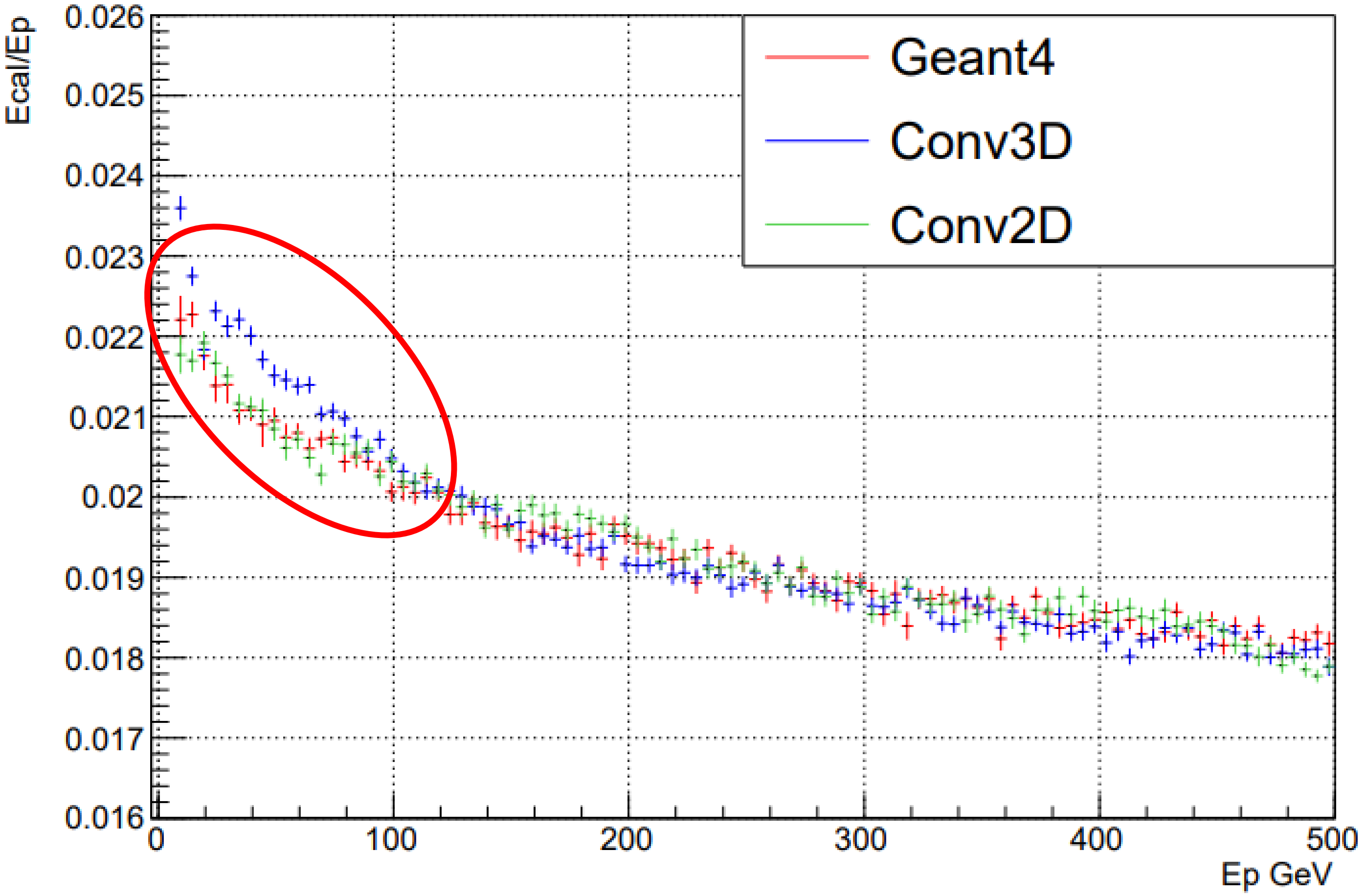


Physics Evaluation

Sampling Fraction

- 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

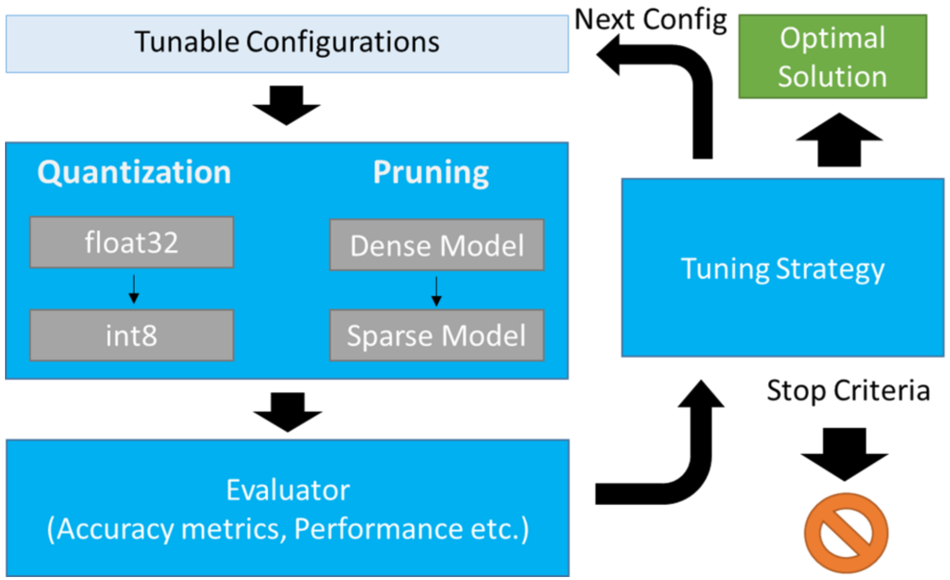
- Quantization: Converting a number from a higher to a lower format
  - E.g. from float32 to int8

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

Int8
1 byte
Max Number: 255

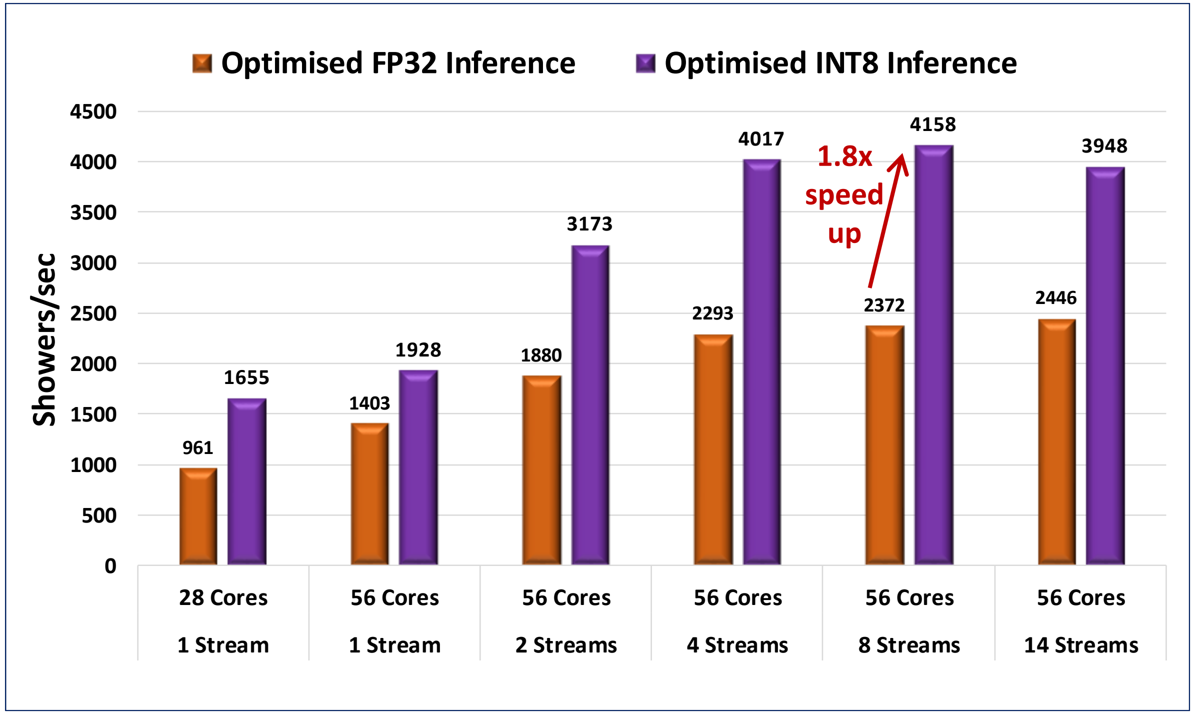
- Quantization Tool:
  Intel Low Precision Optimization Tool (iLoT)
  https://github.com/intel/lp-opt-tool

- 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

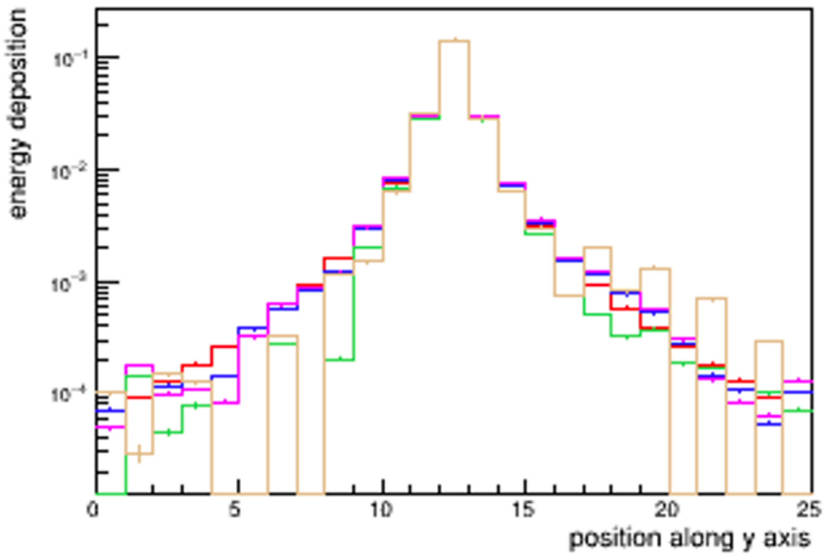
Shower Shapes

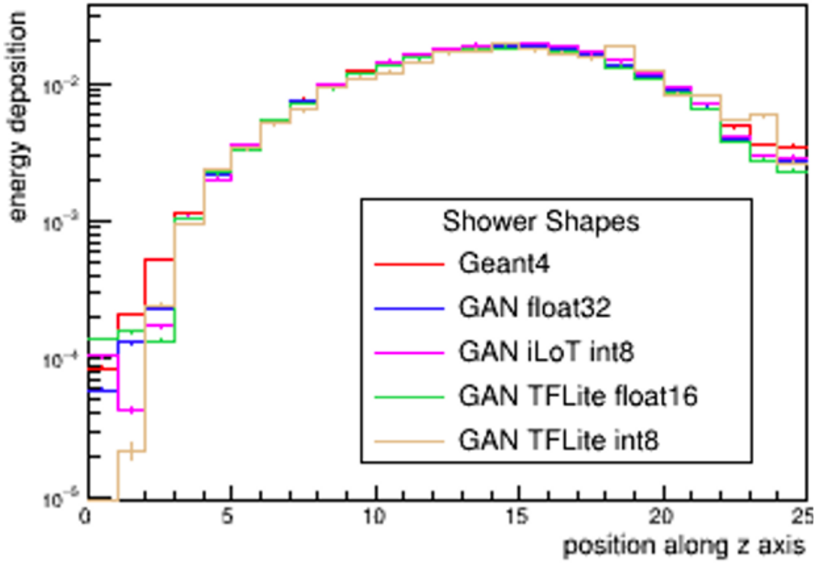
- Mean Squared Error (MSE) between GAN and validation data

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

- 2.1x speed up with new Conv2D network
- Better physics accuracy with Conv2D network
- 1.8x speed up with quantized iLoT model + good accuracy

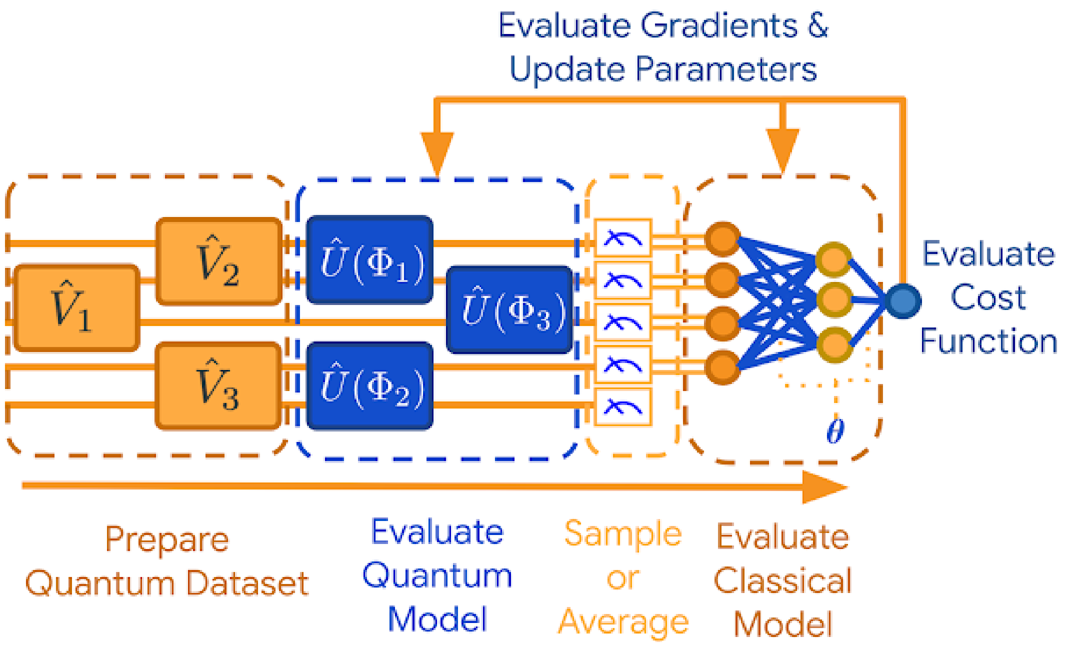
→ Total **68 000x** speed up vs Monte Carlo

Future Work
Quantum Computing

Quantum Computing

General Introduction

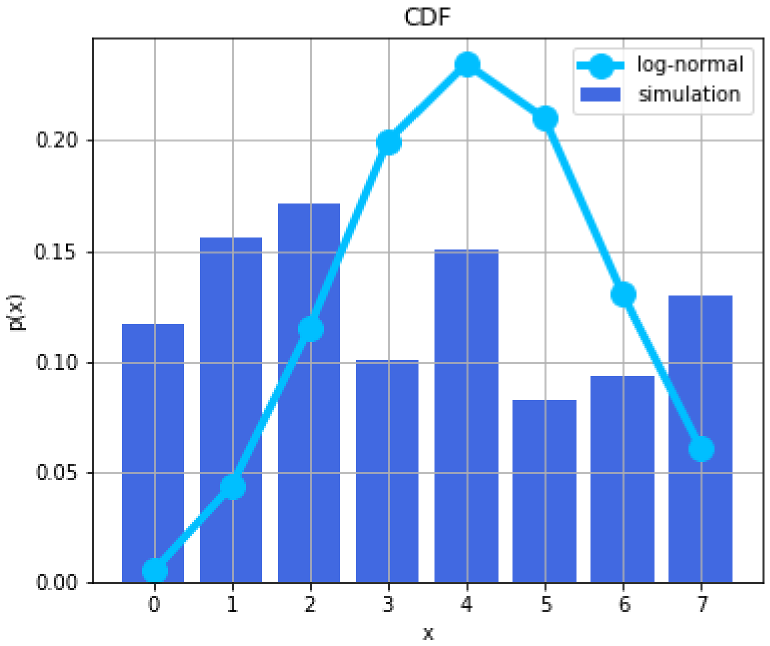
- Use of quantum mechanical properties
  - Entanglement
  - Superposition
- Hope to solve problems faster and / or more accurate
- "Quantum Advantage" not yet reached
  - ➢ We want to start initial tests
- Hybrid model:
  - quantum generator and classical discriminator



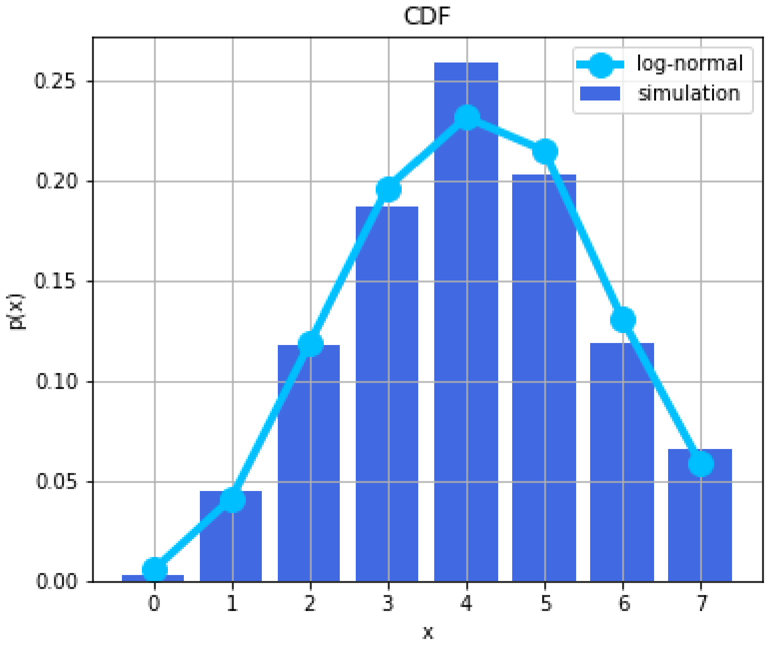
https://ai.googleblog.com/2020/03/announcing-tensorflow-quantum-open.html

Hybrid QGAN

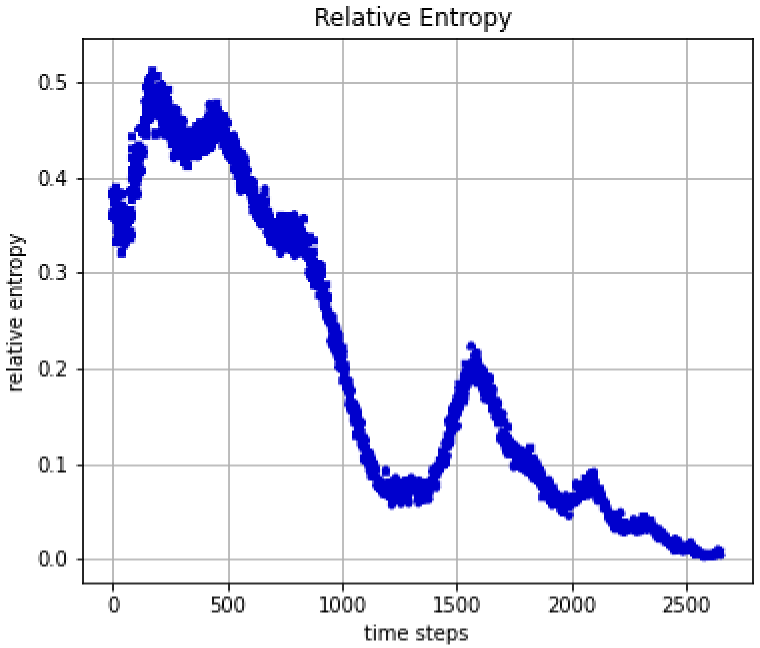
First Test Results



Random Initialization



Trained Model



"How similar the two distributions are"

QUESTIONS?