Reduced Precision Strategies for Deep Learning: 3DGAN Use Case

CERN openlab Technical Workshop 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]

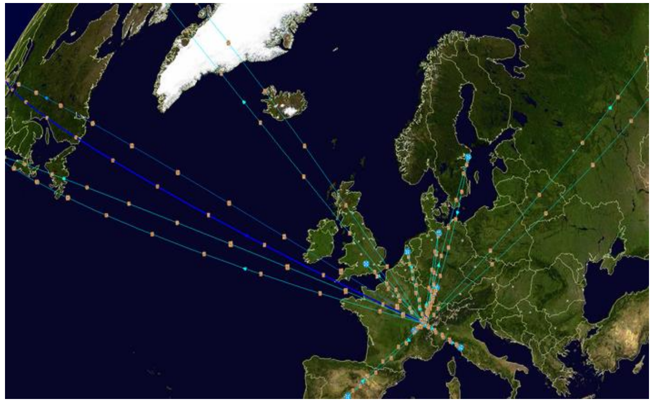
10.03.2021

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**

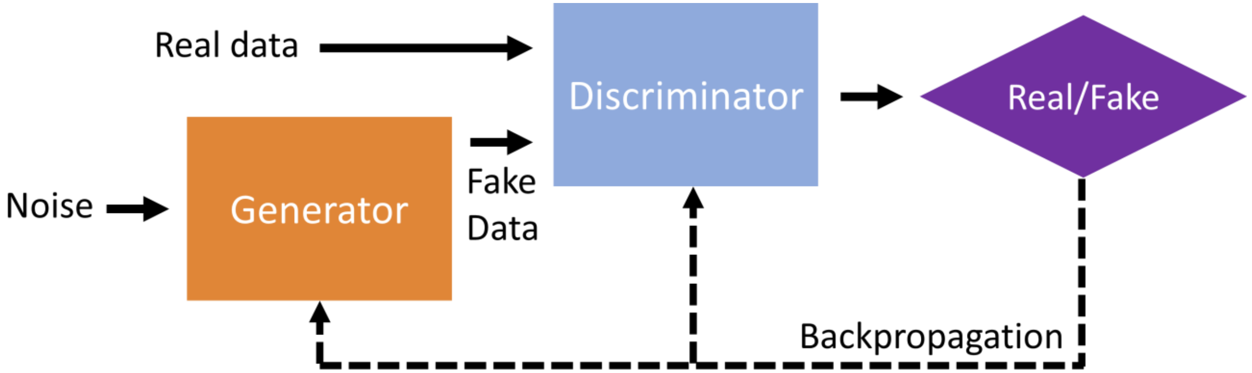


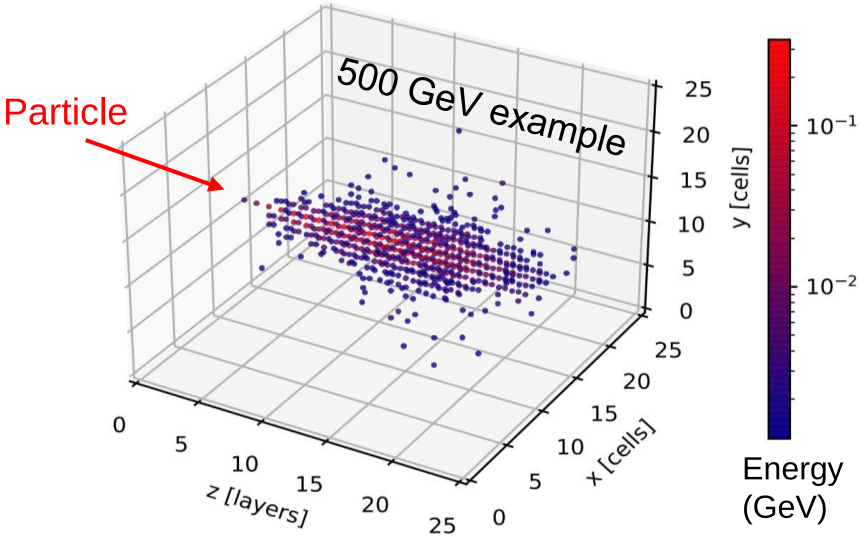
*A Roadmap for HEP Software and Computing R&D for the 2020s
https://doi.org/10.1007/s41781-018-0018-8

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

Why Reduced Precision?

- New simulation approach

  → Use hardware as efficiently as possible

  → 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

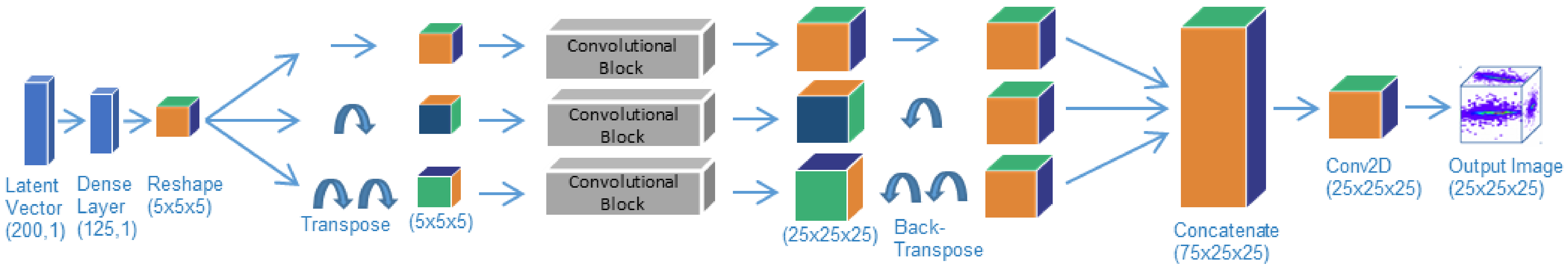
- Reduced precision computation reduces memory and bandwidth occupation

  → Speed-up the simulations and lower memory requirements

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



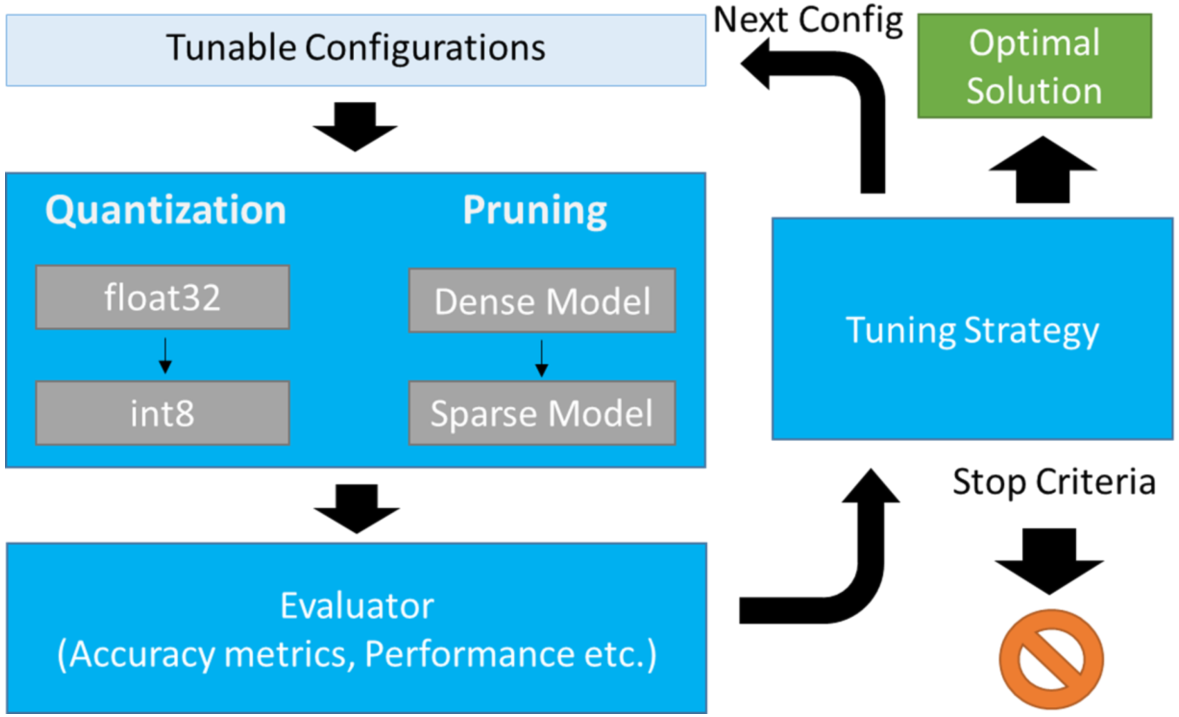
- **2.1x speed-up** due to transition from Conv3D to Conv2D model

Reduced Precision Computing

- Quantization Tool: Intel Low Precision Optimization Tool (iLoT)

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

intel

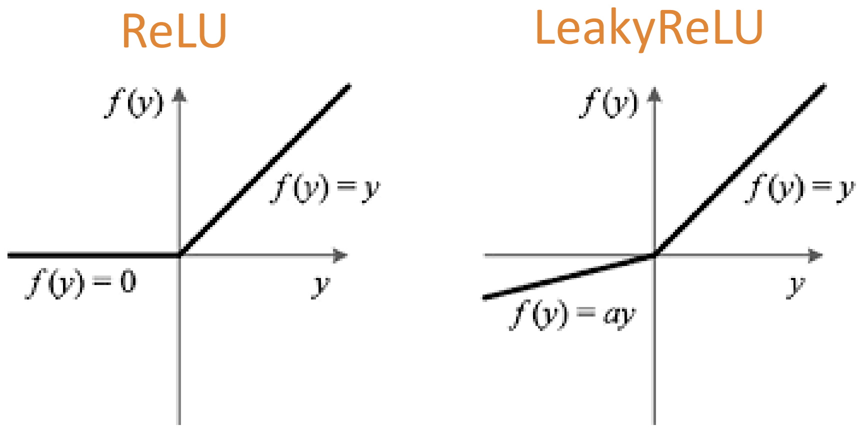


Quantization Problems

- TensorFlow supports no negative quantized values (signed int8)

  → All quantization tools do not support LeakyReLU function

- Needed to be implemented

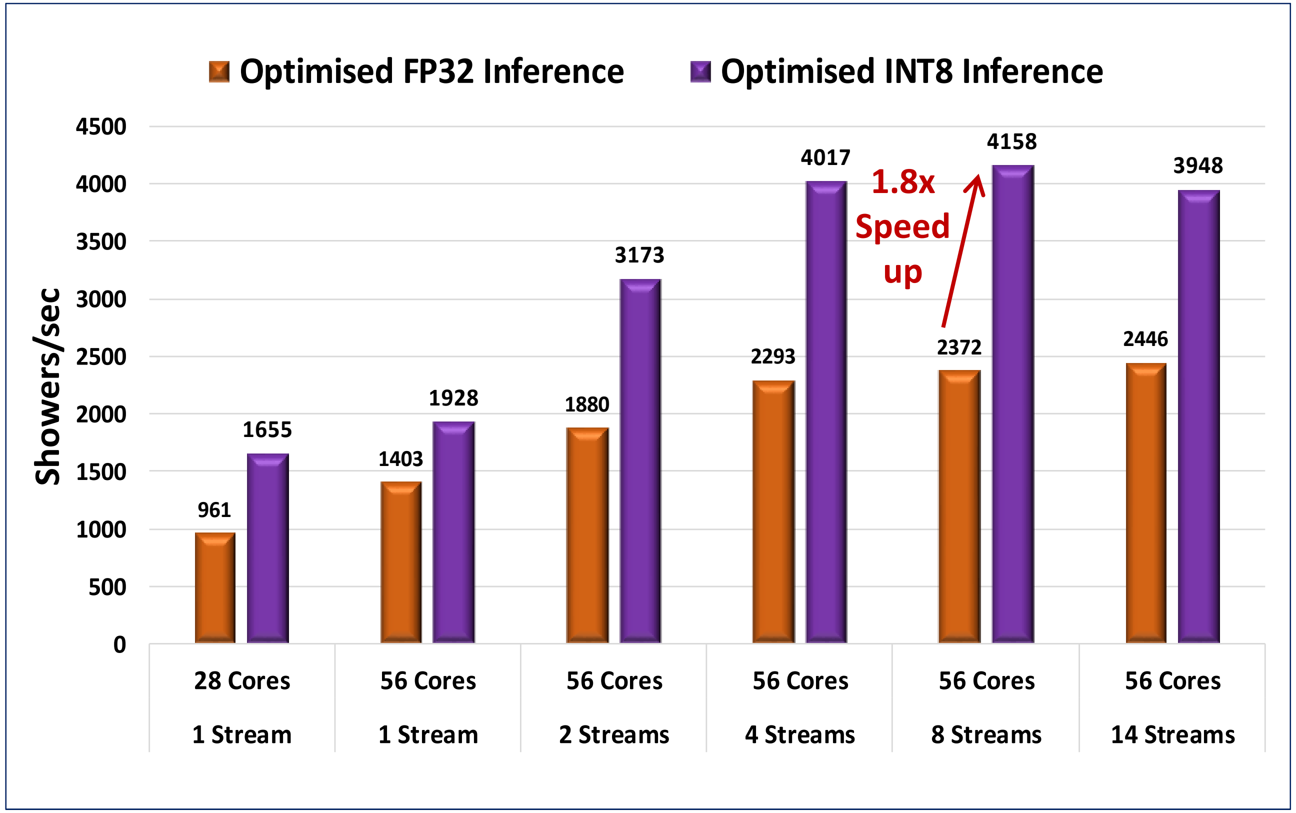


- Reference Tool for accuracy measurement: TensorFlow Lite

  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

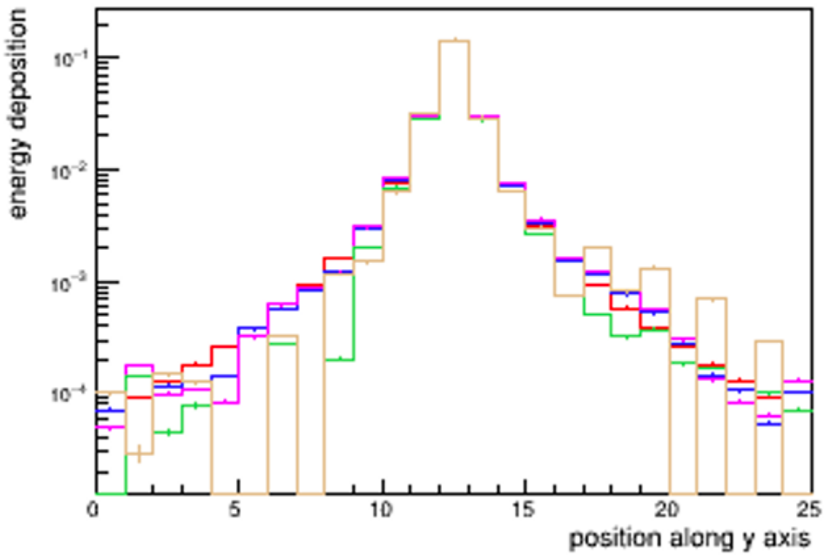
Shower Shapes

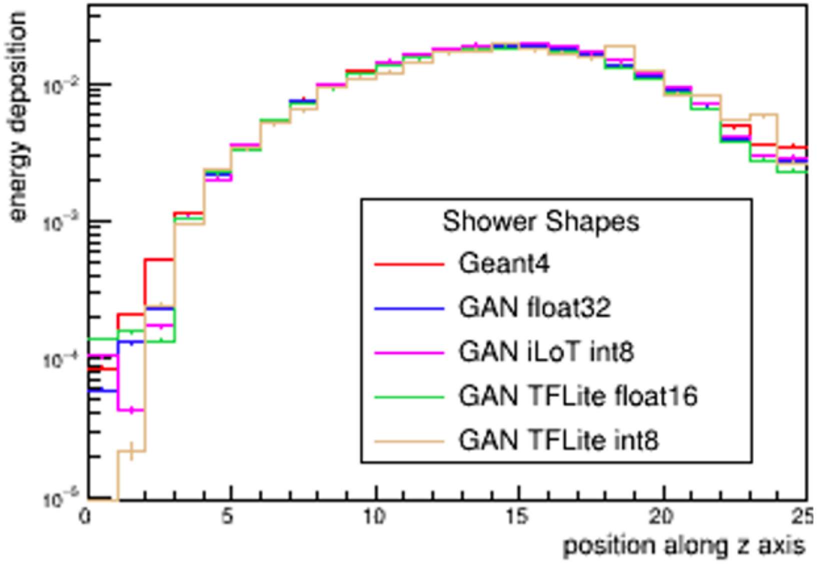
- Mean Squared Error (MSE) between GAN and validation data

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

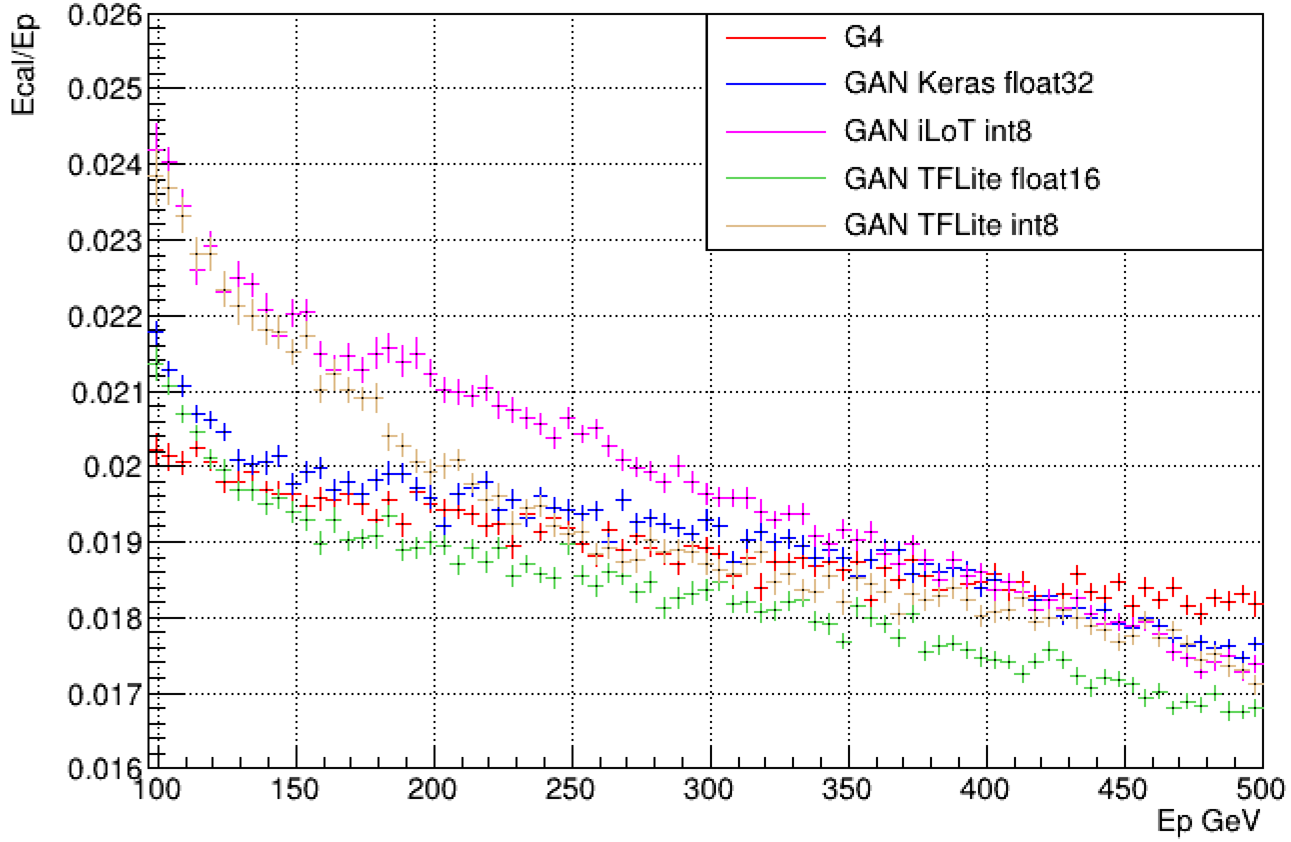




Physics Evaluation

Sampling Fraction

- Ecal: Sum of all image cells ≙ total measured energy
- Ep: Energy of the injected particle into the calorimeter



The quantization does not take this metrics into account

→ More detailed studies needed

→ Define new physics metrics

Summary and Future Plans

- **2.1x** speedup due to conversion from Conv3D to Conv2D
- **1.8x** speedup due to quantization from float32 to int8
- **68 000x** total speedup of quantized GAN versus Geant4 simulation
- **2.24x** less memory with quantized model
- Good physics accuracy for optimization metrics

→ We have successfully accelerated the 3DGAN **inference** process

- 2021: Further optimization on **training** process
  - OneAPI deployment across multiple architectures
  - Introduce ensemble techniques to improve convergence and physics accuracy

QUESTIONS?

Reduced Precision Strategies for Deep Learning: 3DGAN Use Case

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]