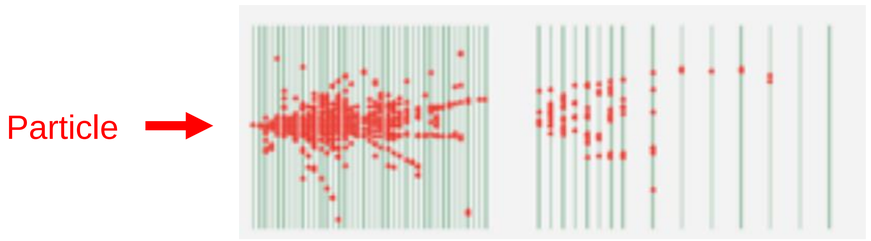
Reduced Precision Strategies for Deep Learning: A GAN Use Case from High Energy Physics

CERN openlab

RWTH AACHEN UNIVERSITY

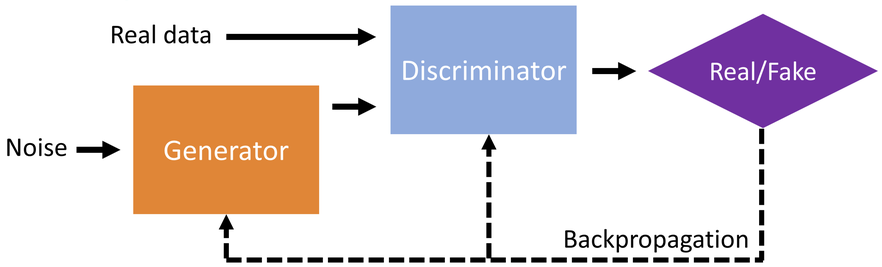
intel

Electromagnetic Calorimeter Simulations



Generative Adversarial Networks (GANs)

to replace traditional Monte Carlo Geant4 simulations



Our Generator Model consists out of 2D convolutional layers to solve a 3D problem. Building block: [Conv2D→LeakyReLU→BatchNorm]

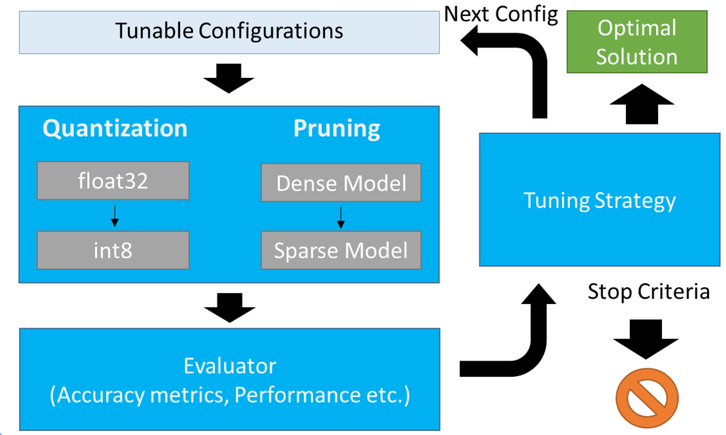
Quantization

Converting a number from higher to lower precision

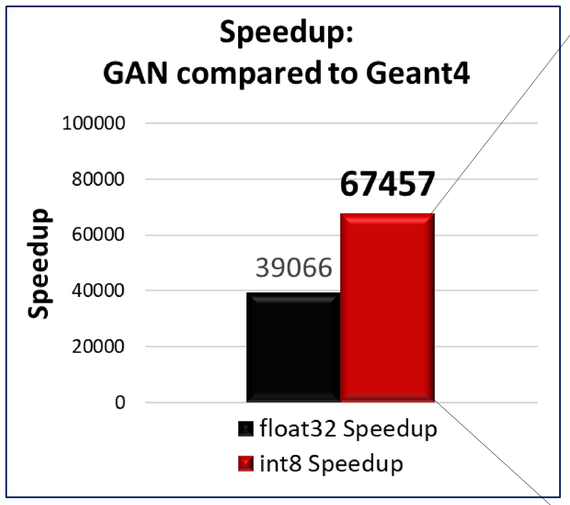
Float32	Int8
4 byte	1 byte
Max Number: $3.4 * 10^{38}$	Max Number: 255

- Geant4 → float32 GAN: 39000x speedup
- Geant4 → int8 GAN: **67000x** speedup
- Maintaining physics accuracy

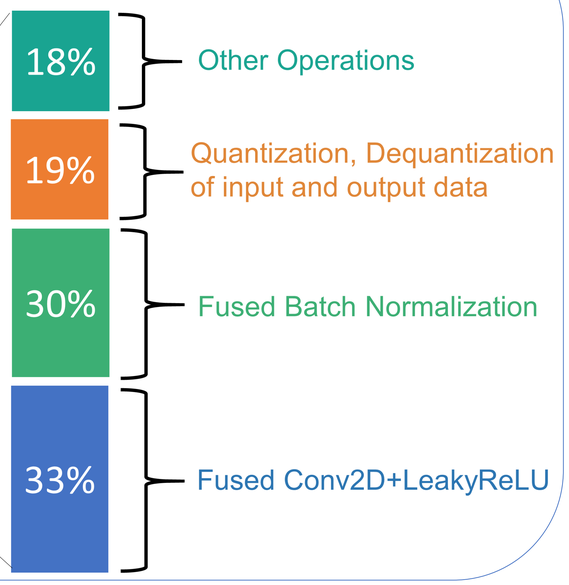
Intel Low Precision Optimization Tool (iLoT)



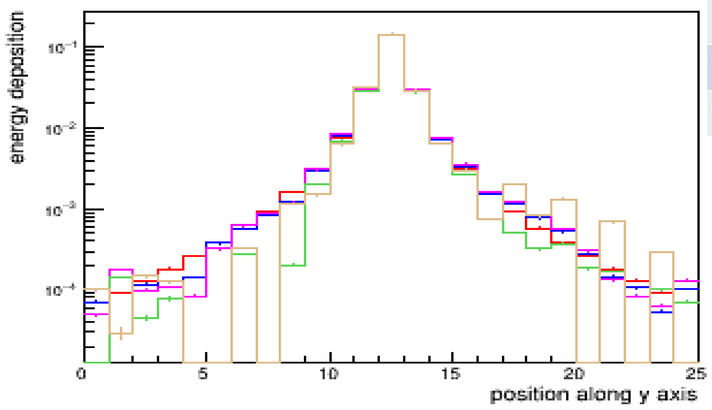
Computational Evaluation



→ Speedup due to quantization: 1.7x



Physics Evaluation



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

→ Lower uncertainty ≙ better accuracy

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