CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Higher Resolution and Angular Conditioning for Normalizing-Flow-based Generation of Calorimeter Showers

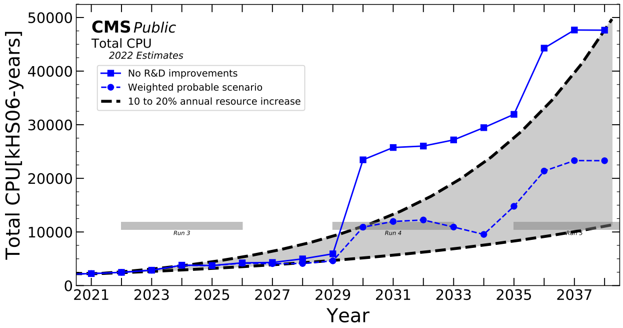
Thorsten Buss, Frank Gaede, Gregor Kasieczka, Anatolii Korol, Peter McKeown

November 5, 2024
ML4Jets 2024

thorsten.buss@uni-hamburg.de

Detector Simulation

- Monte Carlo (MC) necessary to compare theory and measurements
- computational requirements expected to exceed available resources soon
- detector simulation most expensive part of simulation chain

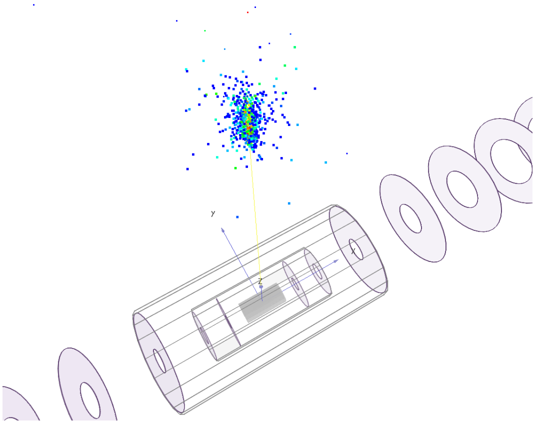


1 CMS Offline Software and Computing. *CMS Phase-2 Computing Model: Update Document*. 2022. URL: https://cds.cern.ch/record/2815292

International Large Detector (ILD)

- proposed detector for the International Linear Collider ILC
- has two sampling calorimeters
- electromagnetic calorimeter (ECAL)
  - 30 layers, 5mm x 5mm cells
- hadronic calorimeter (HCAL)
  - 48 layers, 30mm x 30mm cells
- dataset:
  - photon showers in ECAL
  - uniform distribution of incident energies

$$1 \text{ GeV} \leq E_{\text{inc}} \leq 127 \text{ GeV}$$



[2] Erik Buhmann et al. *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed*. 2021. arXiv: 2005.05334

[3] ILD Concept Group. *International Large Detector: Interim Design Report*. 2020. arXiv: 2003.01116

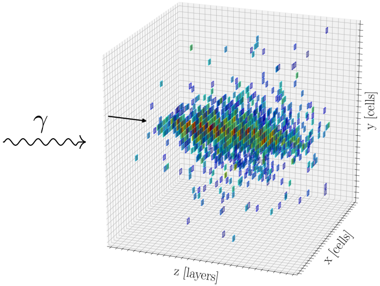
Data Representation of Showers

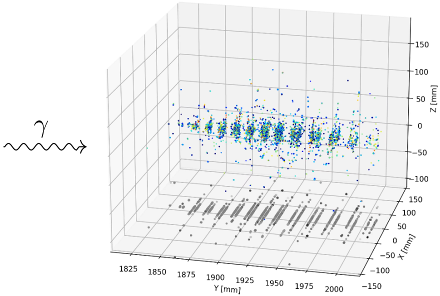
Fixed Grid

- 3D array filled with energy values
- entries correspond to calorimeter cells
- allows for convolutional networks
- needs bounding box

Point Clouds

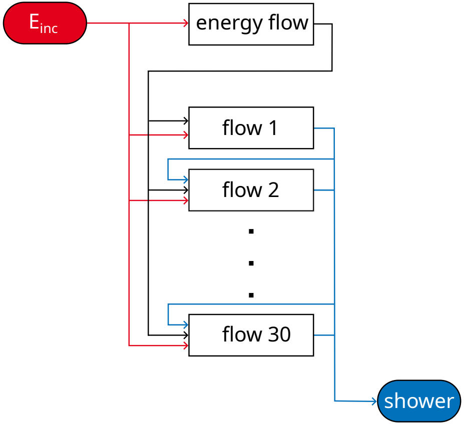
- variable-length, permutation-invariant sets
- only c.a. 4% of cells are non-zero
- more economically represented
- only generation of non-zero points





Convolutional L2LFlows[4]

- one energy distribution flow
  - learns distribution of layer energies
  - conditioned on incident energy
  - masked autoregressive flow[5]
- 30 causal flows
  - learn shower shape in layer
  - conditioned on
    - incident energy
    - layer energy
    - previous layers
  - Glow-like[6] architecture with U-Nets



[4] Thorsten Buss et al. *Convolutional L2LFlows: generating accurate showers in highly granular calorimeters using convolutional normalizing flows*. 2024. arXiv: 2405.20407

[5] Mathieu Germain et al. *MADE: Masked Autoencoder for Distribution Estimation*. 2015. arXiv: 1502.03509

[6] Diederik P. Kingma and Prafulla Dhariwal. *Glow: Generative Flow with Invertible 1x1 Convolutions*. 2018. arXiv: 1807.03039

Integration into Full Simulation

Geometry Dependensy

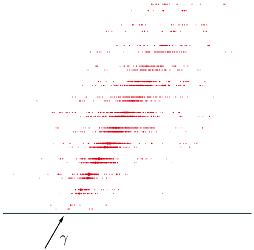
- training on single incident point
  - → decrease in performance when shifted
- solution[7]:
  - training with 9× higher granularity ($90 \times 90 \times 30$)
  - removing detector irregularities Thur. A.K.

Angular Conditioning

- training on single incident angle
- solution[8]:
  - conditioning on incident angle

Integration

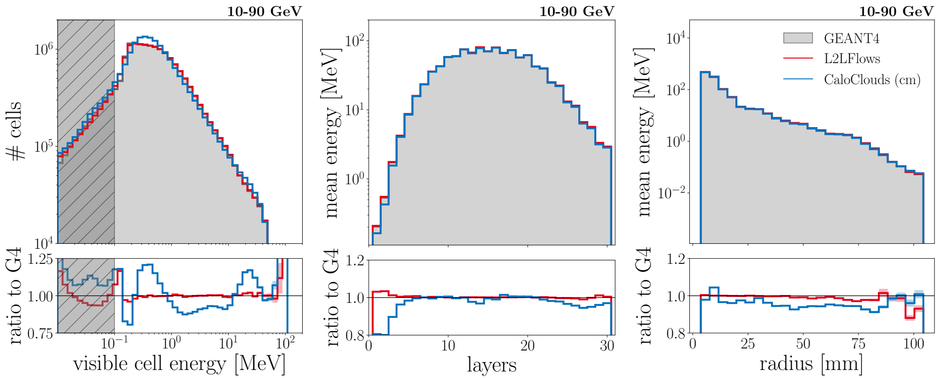
1. Pytorch model Thur. P.M.
2. export to Torchscript and load from C++
3. integrate into simulation chain using DDFastShowerML



[7] Erik Buhmann et al. *CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation*. 2023. arXiv: 2309.05704

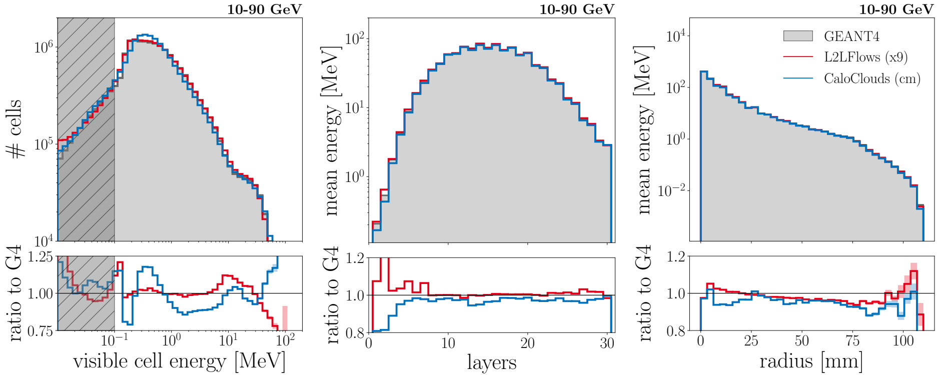
[8] Sascha Diefenbacher et al. *New angles on fast calorimeter shower simulation*. 2023. arXiv: 2303.18150

Results with Box Cut



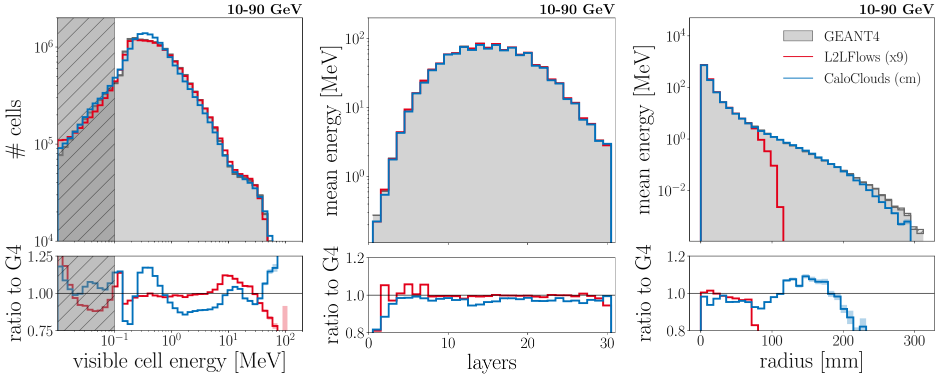
- evaluation at same incident point
- evaluated with 30x30 cell box cut
- good agreement with data

Shifting the Showers



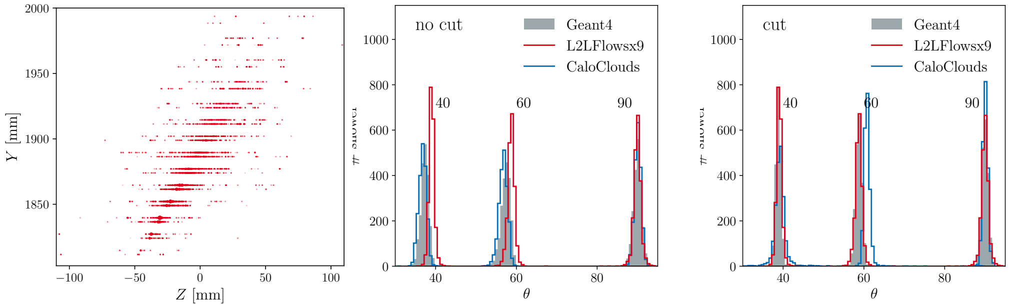
- shift the showers in the calorimeter
- still apply 30x30 box cut
- need to train L2LFlows with nine times higher granularity

No Box Cut



- shift the showers in the calorimeter
- no box cut applied

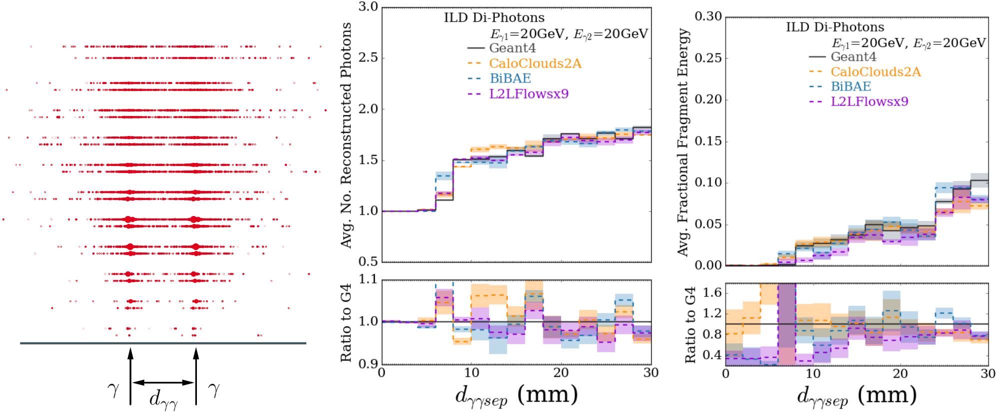
Angular Conditioning



- L2LFlows: good agreement within bounding box
- CaloClouds: not restricted to bounding box

Di-Photon Separation

Can the detector separate two photons?



- using ILD reconstruction software
- possible due to Integration into full simulation

Speedup over GEANT4

- comparison of generation times
- hardware: Intel® Xeon® E5-2640
- #threads: 1
- on GPU speed up of $\times$150 to $\times$1800

Simulator	Batch size	time [ms]	speed up
GEANT4	1	3915	x1.0
CaloClouds II		652	x6.0
CaloClouds (cm)		84	x46.6
L2LFlows		1203	x3.3
L2LFlows (x9)		3713	x1.1
L2LFlows	100	371	x10.6
L2LFlows (x9)		2453	x1.6

timing on single CPU thread

Summary

General

- higher granularity allows for shifting of showers
- angular conditioning allows for varying incident angles

L2LFlows

- very higher fidelity generation of shower core

CaloClouds

- no bounding box necessary
- very fast inference

Integration

- allows for more involved studies
- running reconstruction software on generated data