Generating Accurate Showers in Highly Granular Calorimeters Using Convolutional Normalizing Flows

Thorsten Buss, Sascha Diefenbacher, Frank Gaede, Gregor Kasieczka, Claudius Krause, David Shih

thorsten.buss@uni-hamburg.de

Institut für Experimentalphysik
Universität Hamburg

November 6, 2023

UHH Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

DESY.

KISS

Detector Simulation

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

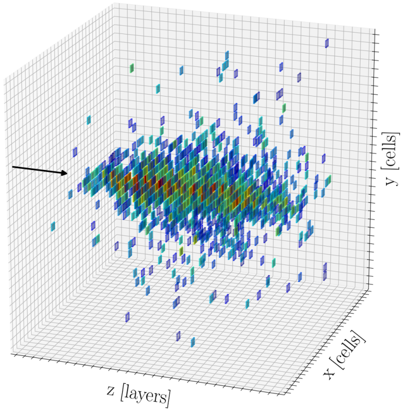
  $\rightarrow$ need for speeding up detector simulation
- generative neuronal networks learn distributions and can sample from them
- work flow:
  - simulate small amounts of data using slow monte carlo
  - train generative model on these data
  - draw large amounts of data from fast ML model

International Large Detector (ILD)

- proposed detector for the ILC
- has two sampling calorimeters
- electromagnetic calorimeter
  - 30 layers, 5mm x 5mm cells
- hadronic calorimeter
  - 48 layers, 30mm x 30mm cells

dataset “getting high”[1]:

- photon showers in ECAL
- bounding box: 30x30 cells
- 30x30x30 voxels

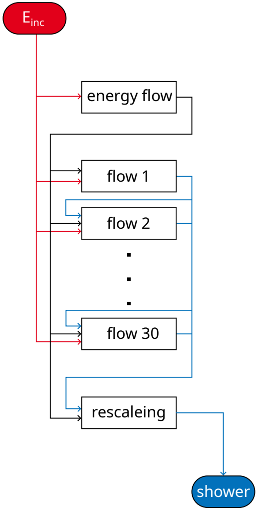


[1] Erik Buhmann et al. *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed*. 2021. arXiv: 2005.05334.

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

Architecture

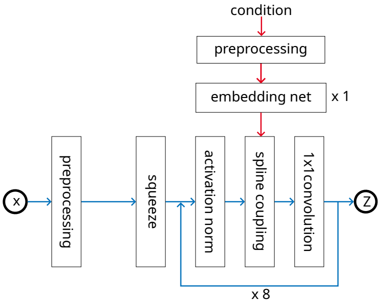
- based on CaloFlow[3] and L2LFlows[4]
- one energy distribution flow
  - learns distribution of layer energies
  - conditioned on incident energy
- 30 causal flows
  - learn shower shape in layer
  - conditioned on
    - incident energy
    - layer energy
    - previous layers
- generation
  - sample layer energies using energy distribution flow
  - sample shower shape using causal flows
  - rescale voxel energies



[3] Claudius Krause and David Shih. *CaloFlow: Fast and Accurate Generation of Calorimeter Showers with Normalizing Flows*. 2021. arXiv: 2106.05285.

[4] Sascha Diefenbacher et al. *L2LFlows: Generating High-Fidelity 3D Calorimeter Images*. 2023. arXiv: 2302.11594.

Flows



- energy distribution flow
  - masked autoregressive flow[5]
- causal flows
  - spline coupling flow[6]
  - convolutional U-Nets[7] as sub networks
  - architecture similar to Glow[8]
- training
  - apply gradient clipping
  - apply wight decay
  - use One Cycle scheduler
- features in energy spectrum are smeared out
  - → apply element-with function to get them back

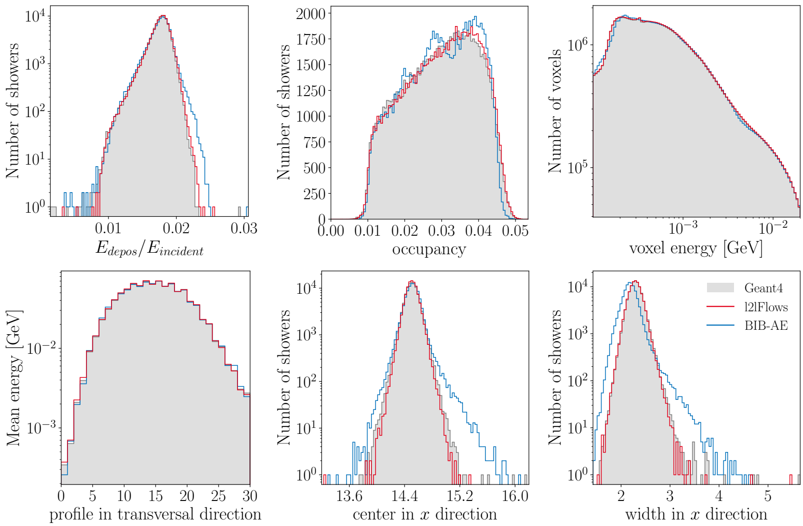
[5] Mathieu Germain et al. *MADE: Masked Autoencoder for Distribution Estimation*. 2015. arXiv: 1502.03509.

[6] Conor Durkan et al. *Neural Spline Flows*. 2019. arXiv: 1906.04032.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597.

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

Observables



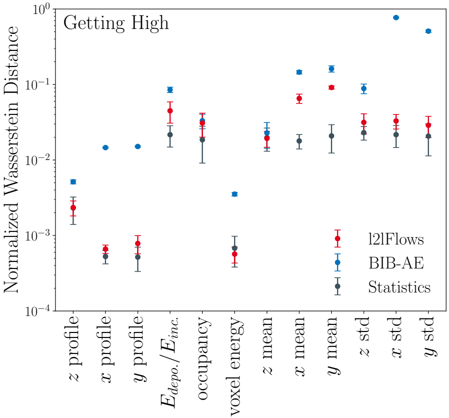
Metrics & Timing

High Level Classifier:
(10 features)

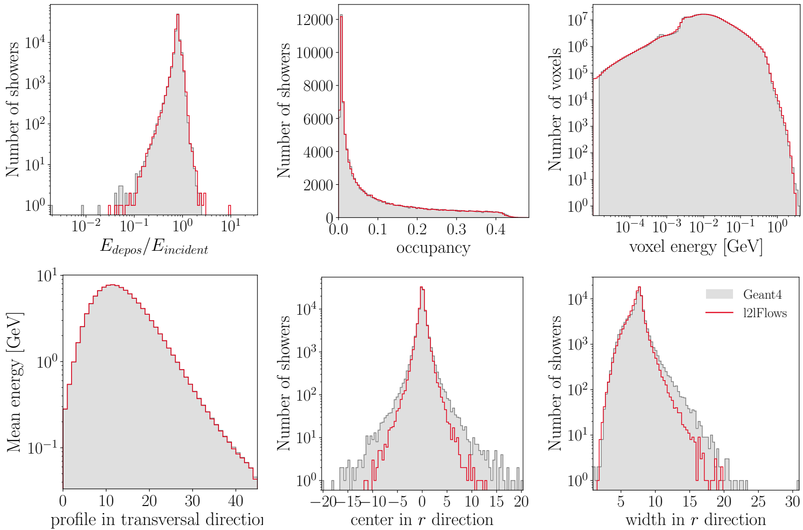
	AUC	JSD
L2LFlows	0.63	0.05
BIB-AE	0.90	0.43

Timing on a single CPU thread:

Simulator	Batch size	time [ms]
GEANT4	1	4081.53
L2LFlows		1202.66
BIB-AE		426.32
L2LFlows	100	417.55
BIB-AE		418.04

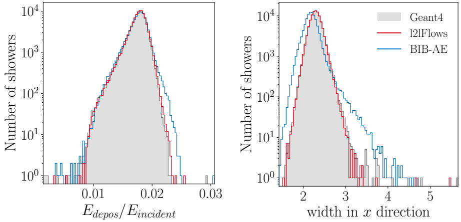


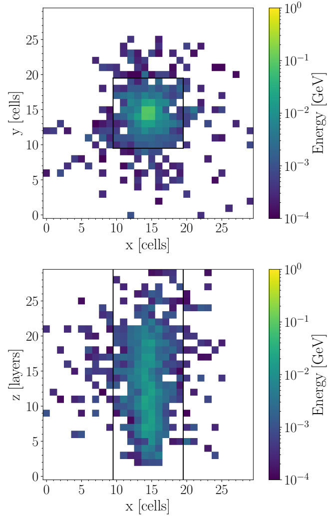
Calo Challenge 3



Summary

- ML generators fast alternative to MC simulations
- the ILD has highly granular calorimeter
  → hard to learn
- convolutional flows scale well with input dimensions
- flows can generate highly accurate showers

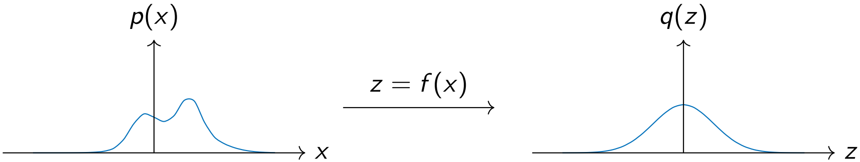




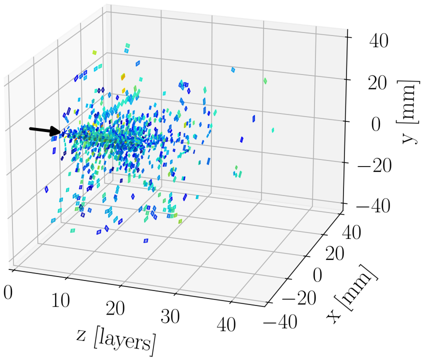
Normalizing Flows

- diffeomorphism between physics space and latent space
- transform physics space distribution into a simple prior distribution
- change of variables formula allows for physics space density estimation
- training: minimize negative log-likelihood
- generation: sample from latent distribution and apply inverse of function

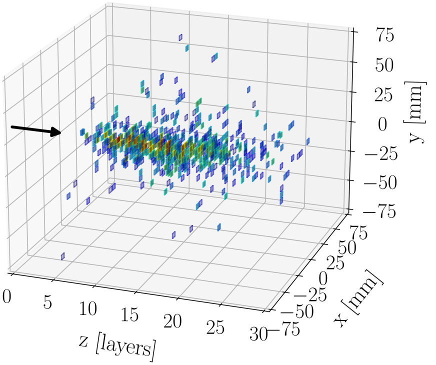
$$p(x) = q(f(x))\,|J(x)| \qquad \mathcal{L} = -\log q(f(x)) - \log|J(x)|$$



Showers

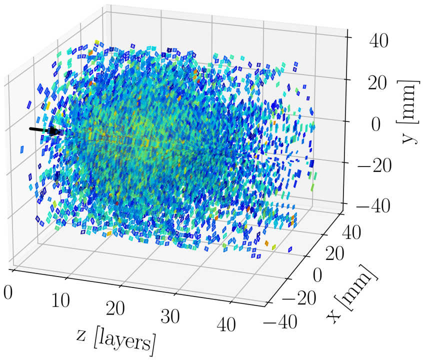


Calo Challenge 3 E_{inc} = 26.6 GeV

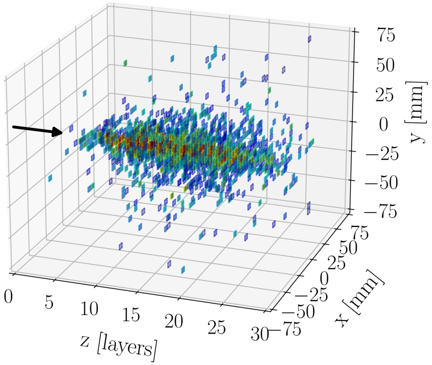


Getting High E_{inc} = 29.3 GeV

Showers

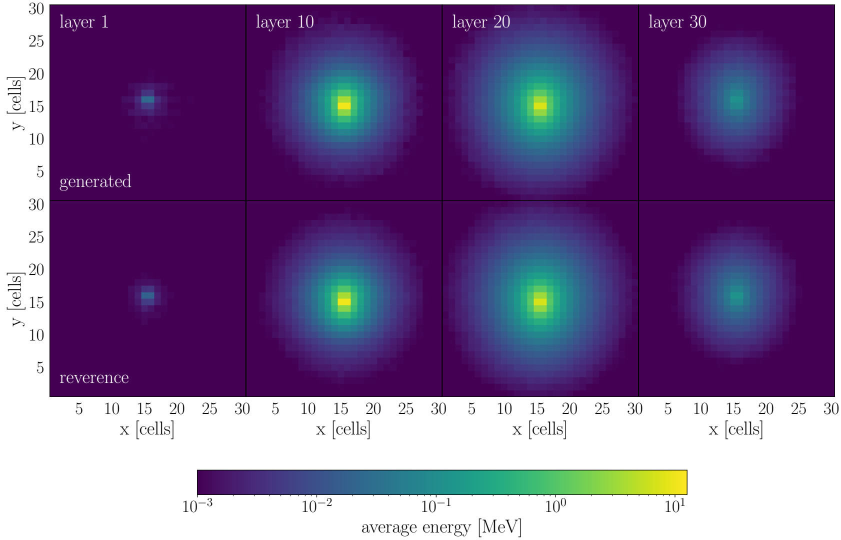


Calo Challenge 3 $E_{inc} = 913.4$ GeV



Getting High $E_{inc} = 81.0$ GeV

Layers


layer 1
layer 10
layer 20
layer 30
generated
reverence
y [cells]
x [cells]
average energy [MeV]
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1}