Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Generative Modeling of Hadronic Showers in the ILD Calorimeters

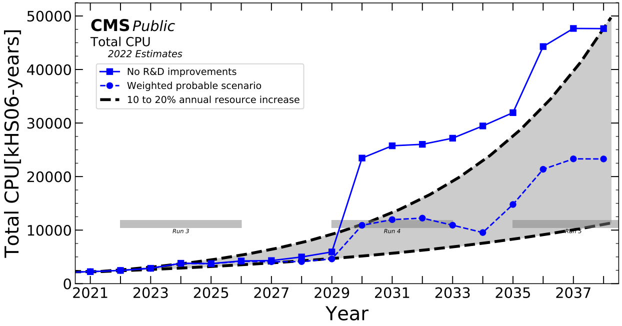
Thorsten Buss, Henry Day-Hall, Frank Gaede, Gregor Kasieczka, William Korcari
Anatolii Korol, Katja Krüger, Martina Mozzanica, Lorenzo Valente

March 12, 2025
KISS Annual Meeting

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

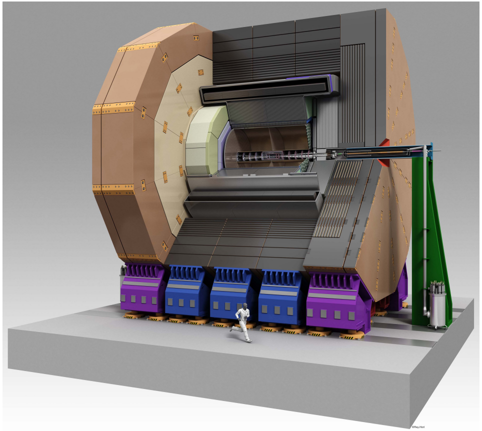


[1] CMS Offline Software and Computing. *CMS Phase-2 Computing Model: Update Document*. 2022. URL: https://cds.cern.ch/record/2815292

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:
  - pion showers in ECAL and HCAL
  - uniform distribution of incident energies

$$10\ \mathrm{GeV} \leq E_{\mathrm{inc}} \leq 90\ \mathrm{GeV}$$

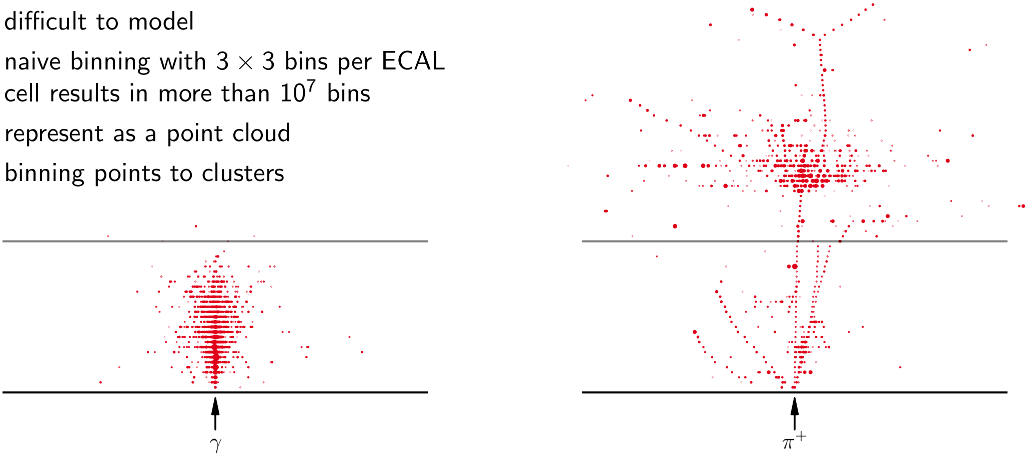


[2] Ties Behnke et al. *The International Linear Collider Technical Design Report - Volume 4: Detectors*. 2013. arXiv: 1306.6329

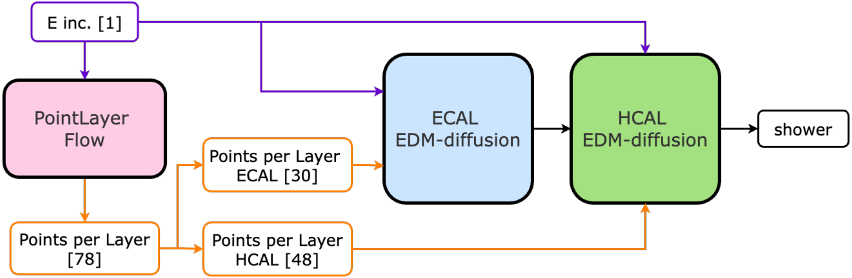
[3] ILD Concept Group. *International Large Detector: Interim Design Report*. 2020. arXiv: 2003.01116

Hadronic Showers in ILD Calorimeters

- hadronic showers are complex
- difficult to model
- naive binning with 3×3 bins per ECAL cell results in more than 10^7 bins
- represent as a point cloud
- binning points to clusters



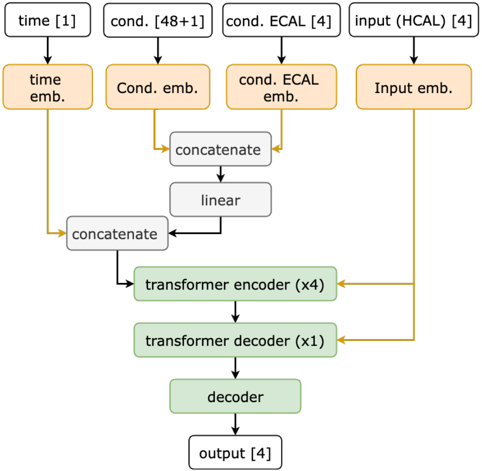
PionClouds



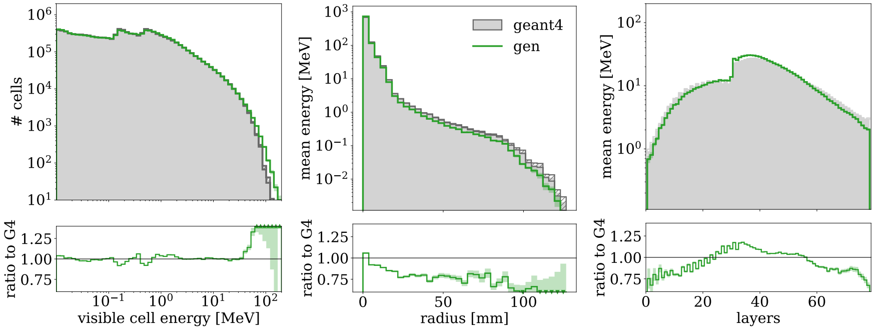
- generative model for hadronic showers
- based on point clouds
- score matching diffusion model
- no calibration necessary

HCAL Network Architecture

- modeling point interactions with attention mechanism
- four encoder layers and one decoder layer
- Fourier Layers help to learn high-frequency features
- conditional inputs:
  - diffusion time
  - number of points per layer
  - incident energy
  - ECAL point cloud
- inputs:
  - HCAL point cloud
- outputs:
  - score function

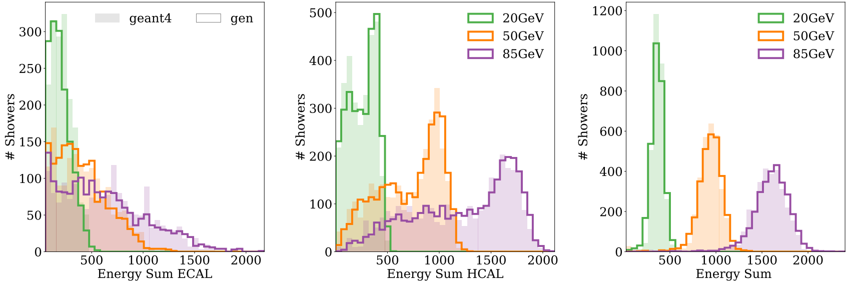


Resuts Point Level



- point level observables
- good agreement with MC
- uniformly distributed incident energies
- left: cell energy spectrum
- middle: radial energy profile
- right: longitudinal energy profile

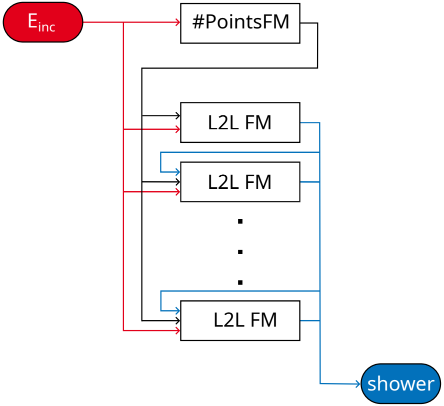
Resuts Shower Level



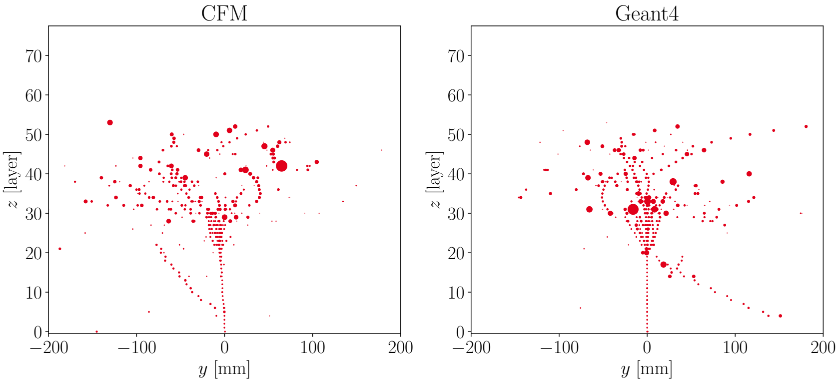
- visible energy sum
- good agreement with MC
- left: energy deposited in ECAL
- middle: energy deposited in HCAL
- right: energy deposited in both

L2LFlowMatching

- generative model for hadronic showers
- based on point clouds
- conditional flow matching model
- point interactions modeled with attention mechanism
- layer are generated sequentially
  - #PointsFlow generates number of points per layer
  - Inductive Flow model generates point positions
  - conditioning on up to fore previous layers
  - reduces number of points in the attention mechanism

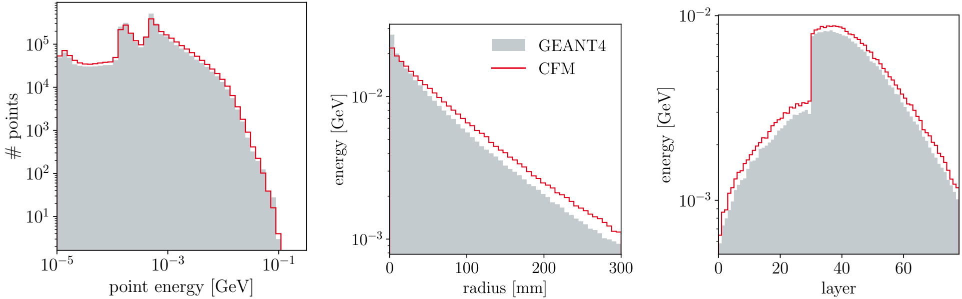


Preliminary Results



- number of points per layer from real data
- good agreement on individual shower level
- left: generated shower
- right: real shower

Preliminary Results Point Level



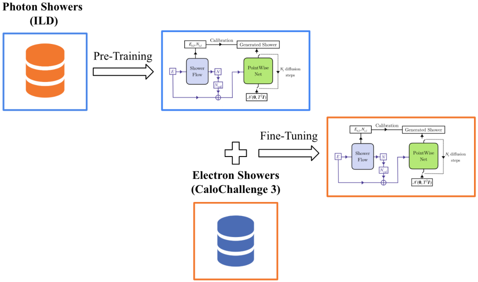
- point level observables
- number of points per layer from real data

- left: point energy spectrum
- middle: radial energy profile
- right: longitudinal energy profile

Domain Adaptation

Basic Assumption: Shared properties between domains facilitates knowledge transfer.

- photon showers $\rightarrow$ electron showers
- photon showers: CaloClouds dataset
- electron showers: CaloChallenge 3[4]
- model: CaloClouds II[5]

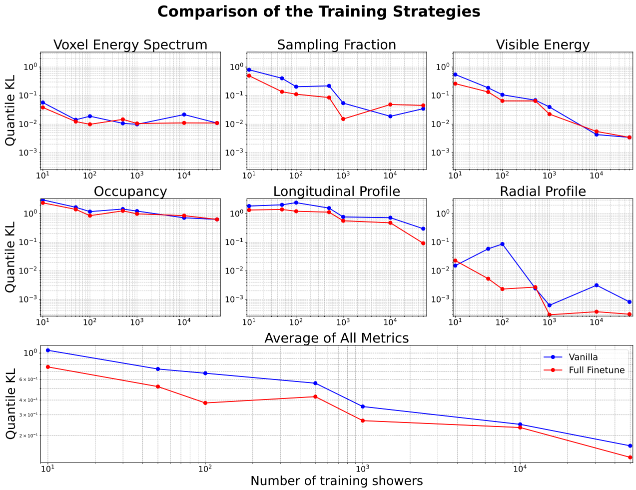


[4] Claudius Krause et al. *CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation*. 2024. arXiv: 2410.21611

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

Preliminary Results

- Kullback-Leibler divergence (KL)
- six heigh level observables
- average over all six kl values
- same metric used for model selection
- advantage of domain adaptation



Summary

- hadronic showers are complex
- high granularity of ILD calorimeters
- modeling based on point clouds
- attention mechanism for point interactions
- PionClouds
  - score matching diffusion model
  - no calibration necessary
  - good agreement with MC
- L2LFlowMatching
  - conditional flow matching model
  - layer are generated sequentially
  - good agreement on individual shower level
- domain adaptation
  - reusing knowledge
  - improves performance for smaller datasets