CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation

CaloChallenge Workshop

Material based on: arXiv:2305.04847v1

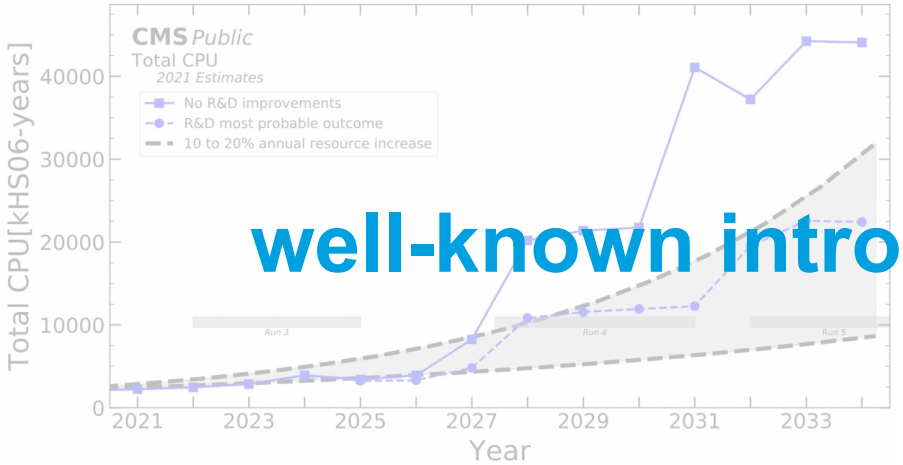
Erik Buhmann, Sascha Diefenbacher, Engin Eren, Frank Gaede, Gregor Kasieczka, William Korcari, **Anatolii Korol**[1], Katja Krüger, Peter McKeown, Lennart Rustige

[1] **Deutsches Elektronen-Synchrotron**, anatolii.korol@desy.de

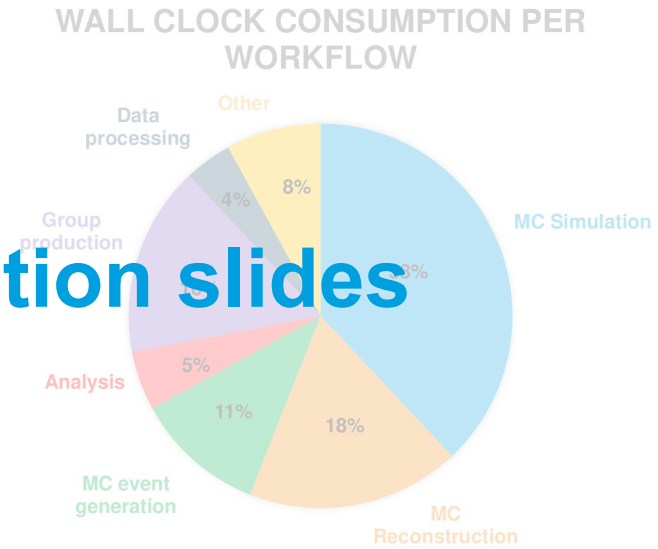
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE
HELMHOLTZ

Problem Definition

Time-consuming Simulations



CMS Collaboration,
Offline and Computing Public Results (2021)
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults



D. Costanzo, J. Catmore, ATLAS
Computing update, LHCC meeting , 2019

well-known introduction slides

Goal: replace (or augment) simulation steps with a faster powerful generator, based on state-of-the-art machine learning techniques

Generative Models

Overview

- Generative Model is just a function that maps random numbers to some structure
- In most cases the structure is an **image representation** of the electromagnetic shower (EM shower) in the calorimeter

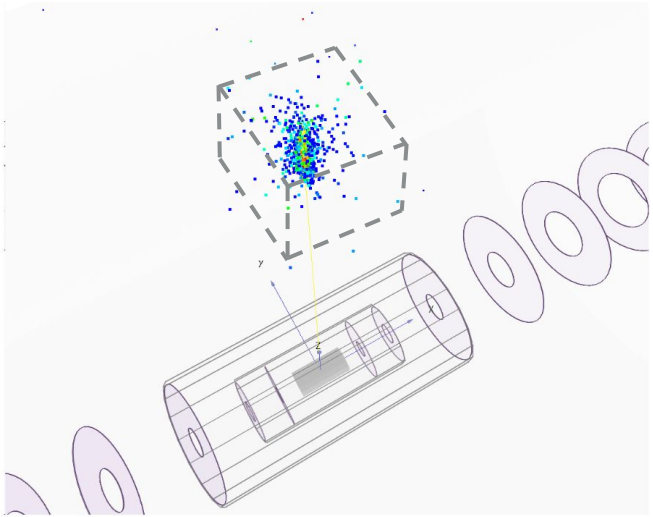
well-known introduction slides

- There exist numerous generative models
  - Generative Adversarial Networks (GANs)
  - Flow-based models
  - Autoencoders (AE), e.g. BiB-AE
  - Denoising Diffusion Probabilistic models (DDPs)

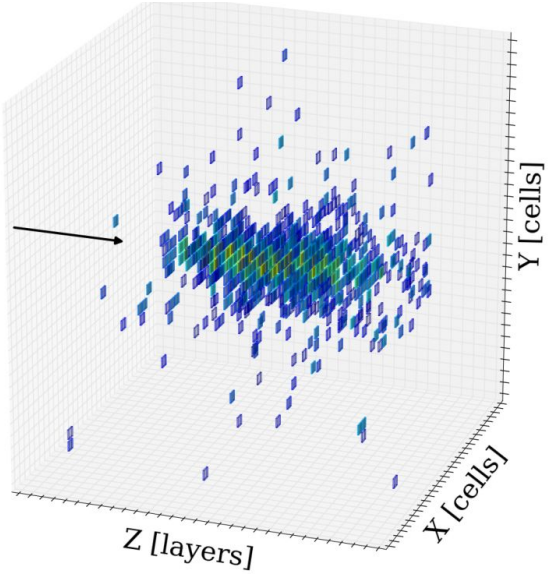
Image Representation of the EM Showers

ILD Detector

A simulated 60 GeV photon shower in the ILD detector



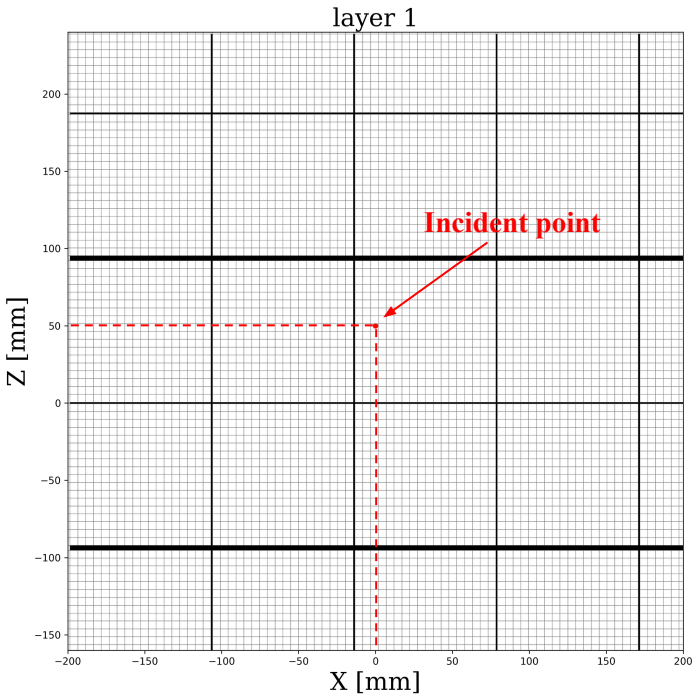
Regular grid 30x30x30

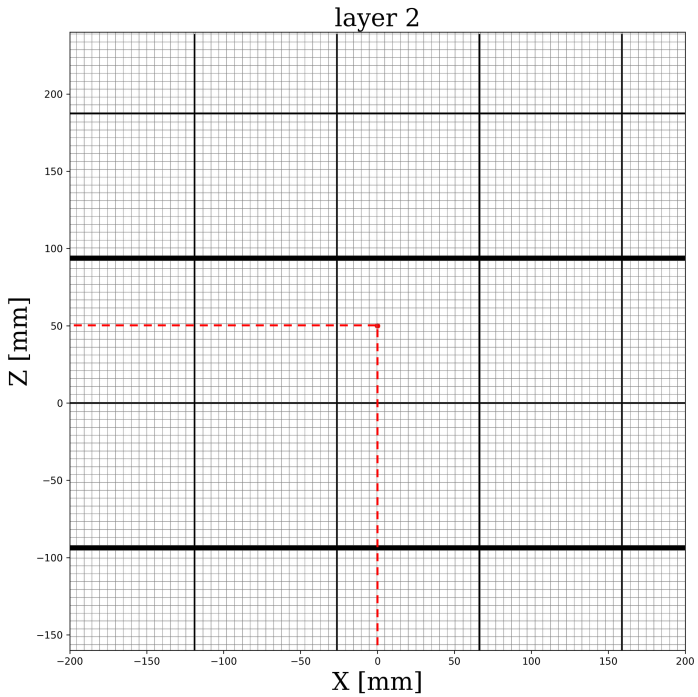


One to one mapping from detector geometry to a regular grid

Image Representation of the EM Showers

ILD Detector, ECAL Layers Structure

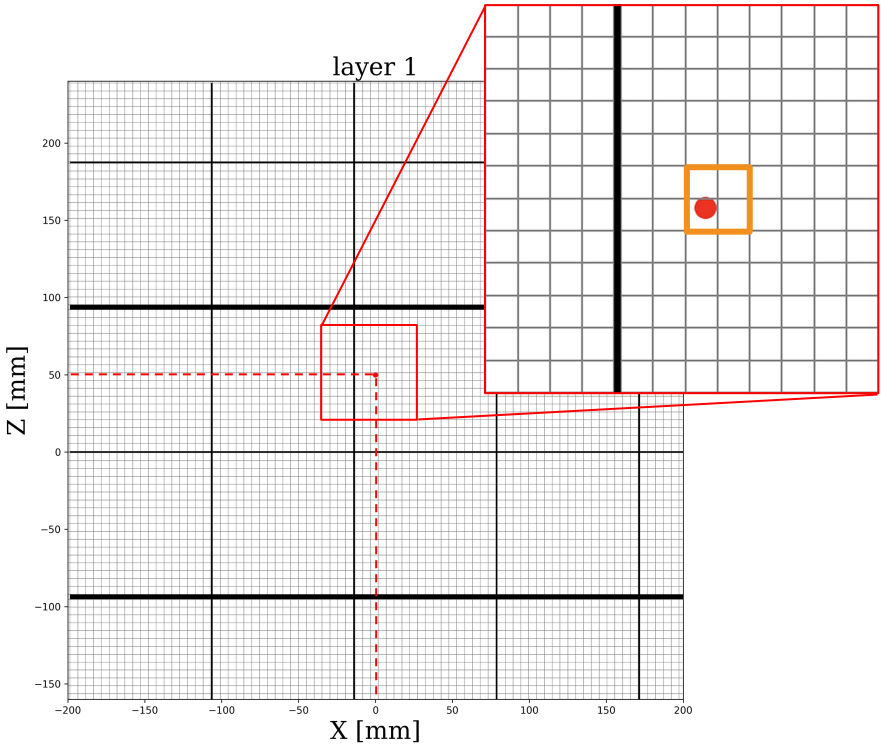


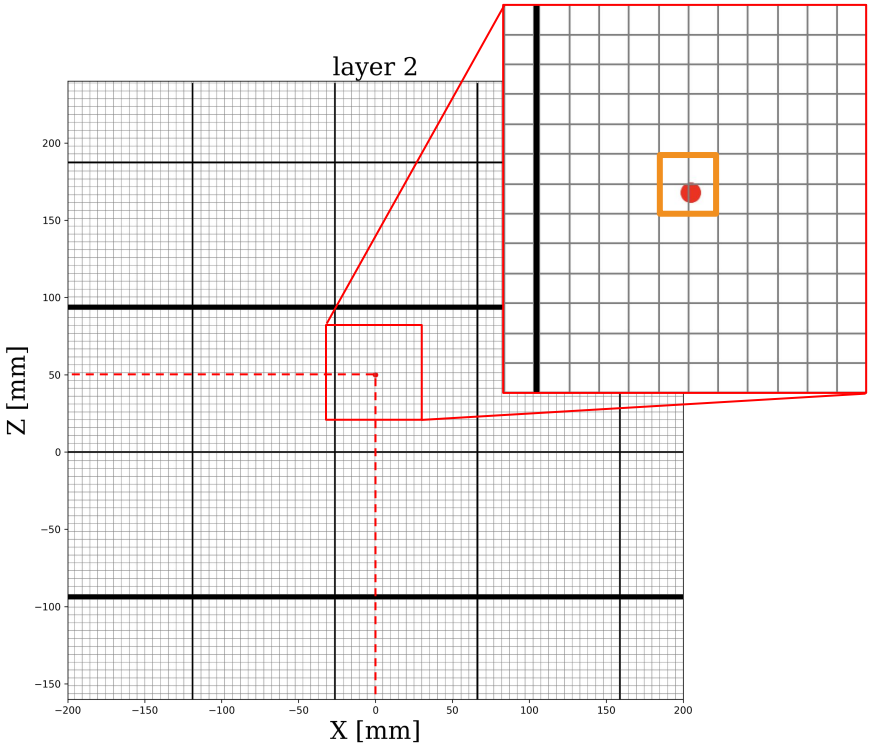


White squares represent **active cells**. Black lines are wafers, construction gaps, etc. (**not active material**)

Image Representation of the EM Showers

ILD Detector, ECAL Layers Structure

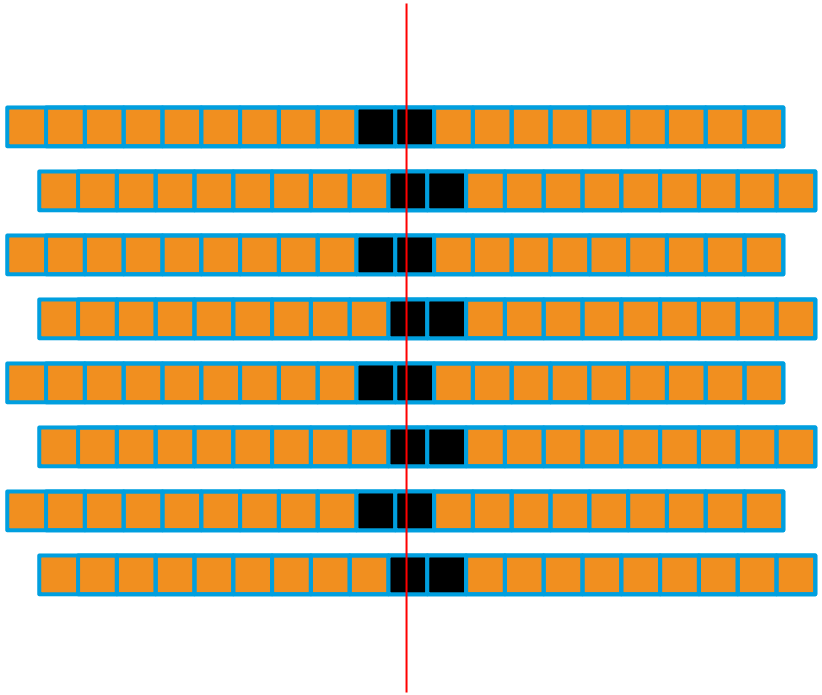


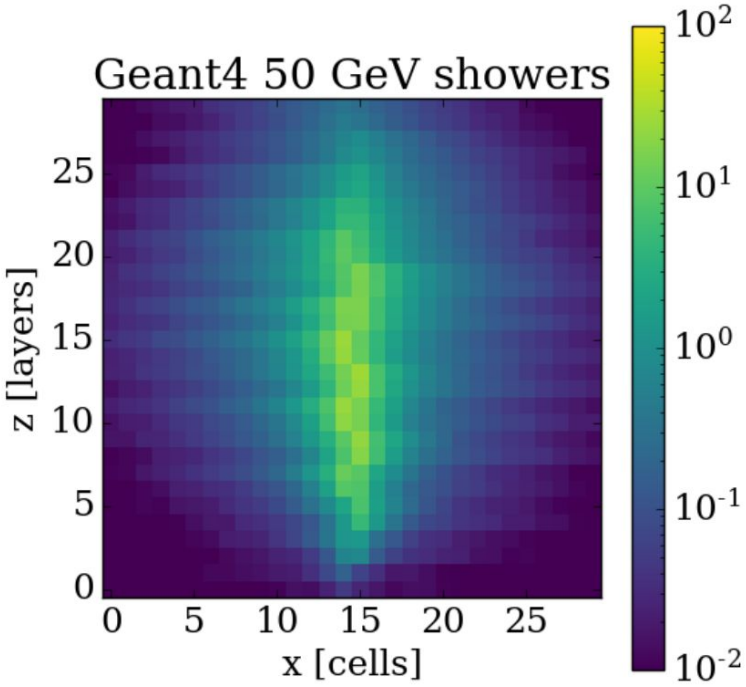


White squares represent **active cells**. Black lines are wafers, construction gaps, etc. (**not active material**)

Image Representation of the EM Showers

ILD Detector, ECAL Layers Structure, Staggering Effect





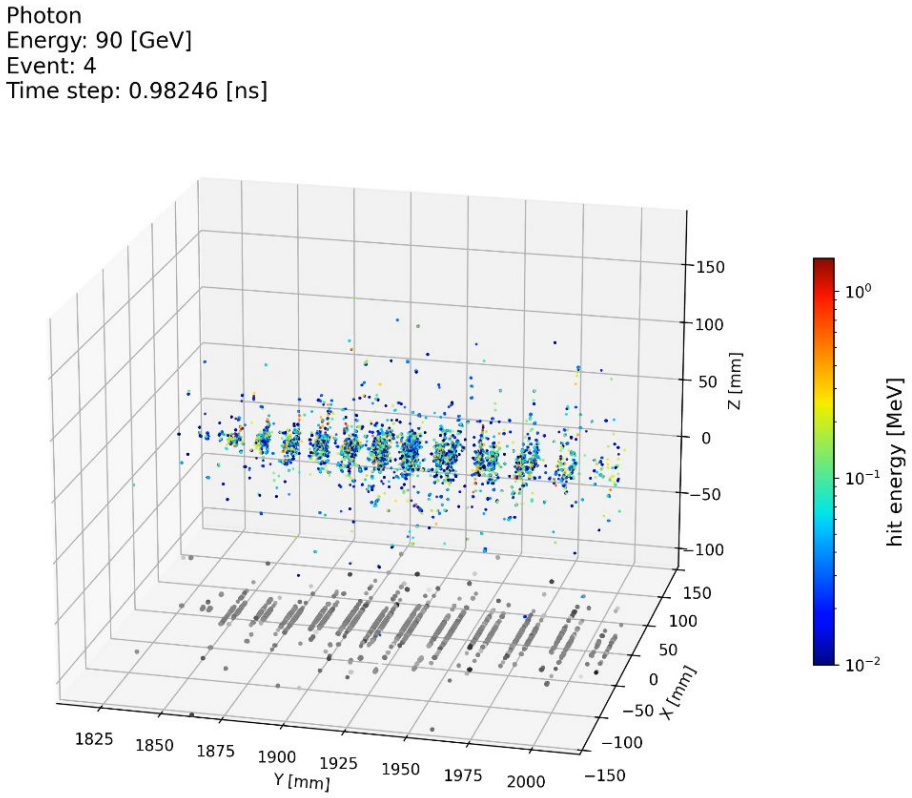
Models have to learn not only EM shower properties, but also geometry “artifacts”, like **staggering effect**

Point Cloud Representation of the EM Showers

GEANT4 Steps

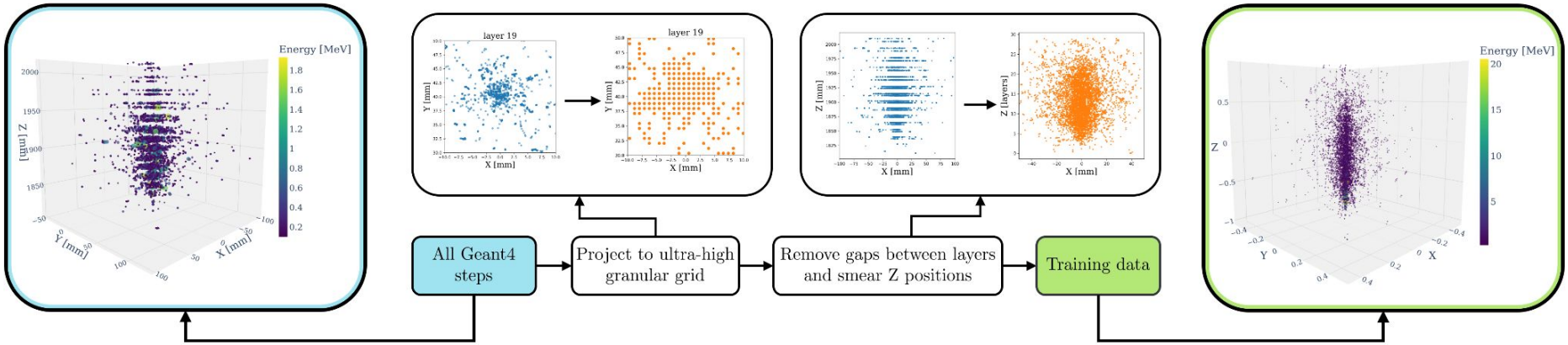
A way to overcome potential issues from irregular (realistic) cell geometry would be to use much higher granularity/resolution

- All G4 interactions, ultimate resolution
- Detached from detector layer geometry
- Too many points to generate, ~40k per shower (need pre-processing step to reduce number of spacepoints)



Point Cloud Representation of the EM Showers

Data Preprocessing

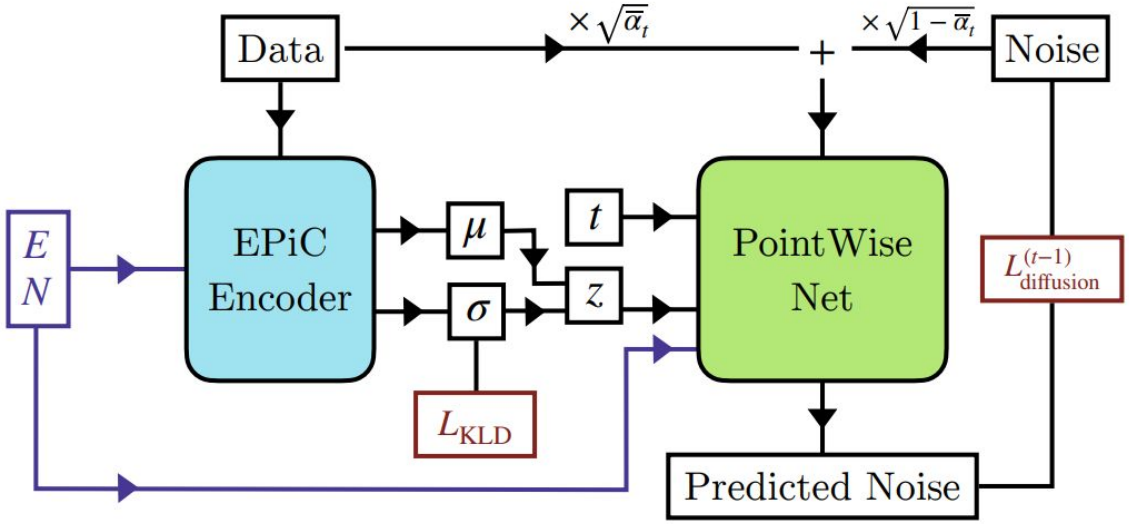


Number of points reduced to ~6k per shower, high enough resolution to move the shower in different place without harming physical properties of the shower

Point Cloud + Diffusion Model

Model Overview, Training

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation
arXiv:2305.04847v1



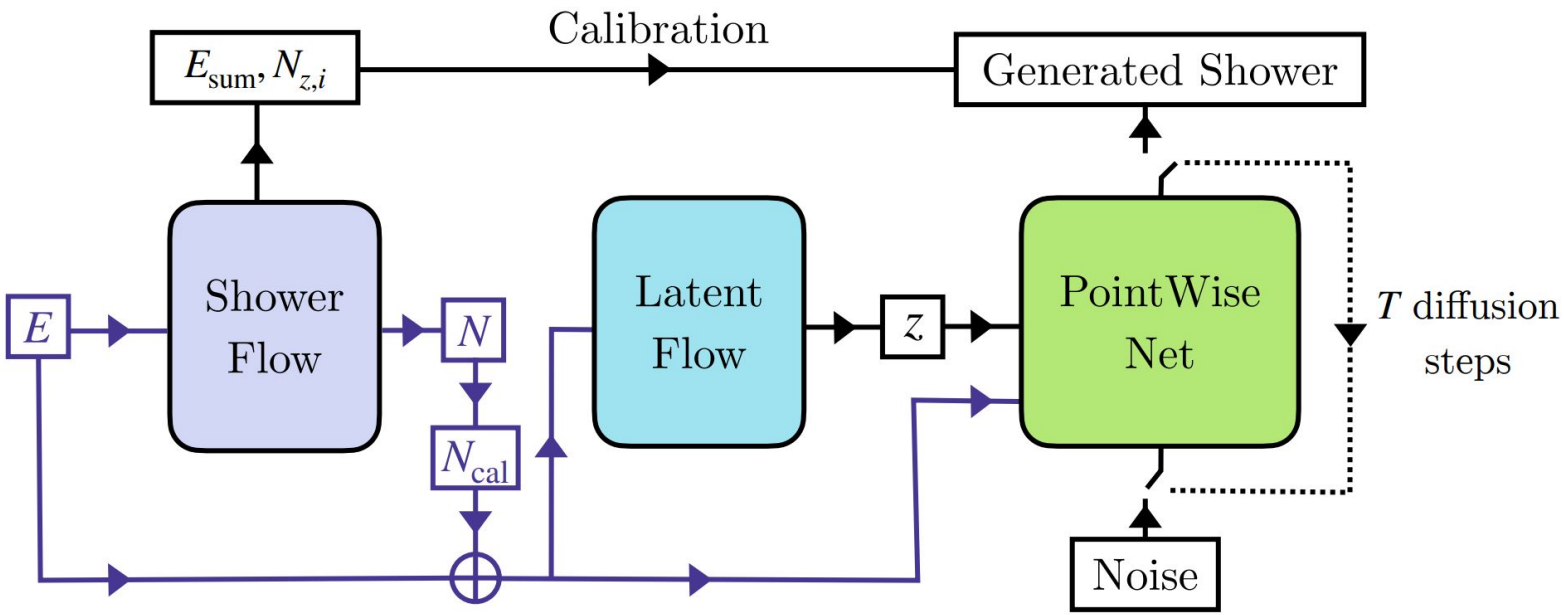
(a) Training at random time step t

- GANs and VAEs convert noise from some simple distribution to a data sample
- **DMs learn to gradually denoise data starting from noise**

Point Cloud + Diffusion Model

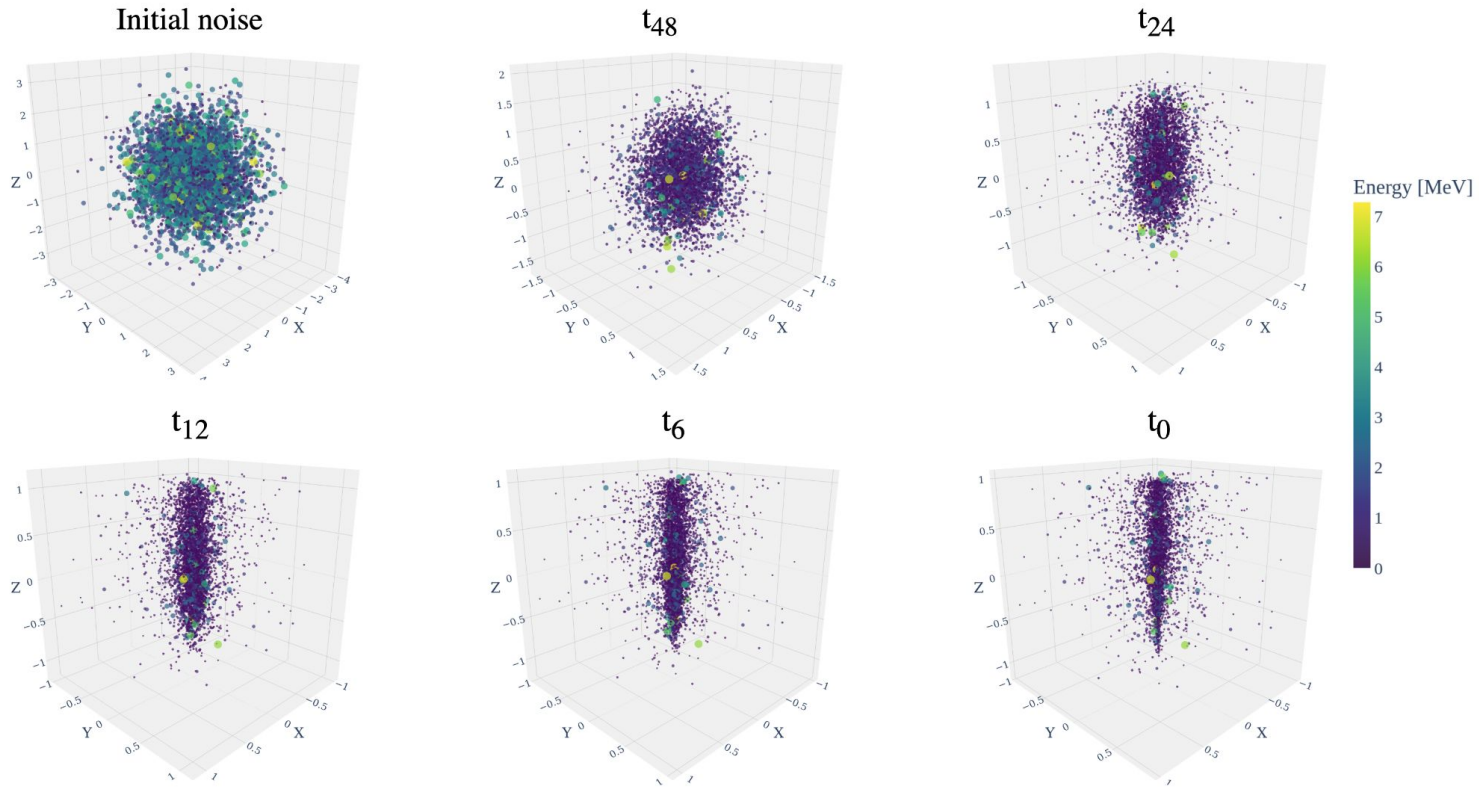
Model Overview, Inference

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation
arXiv:2305.04847v1



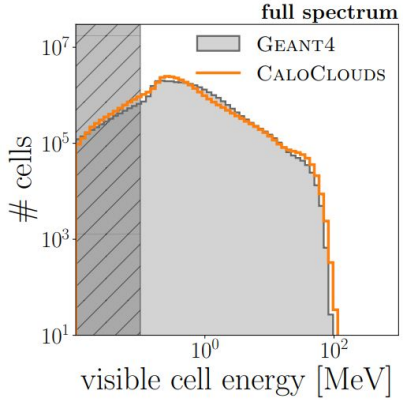
Point Cloud + Diffusion Model

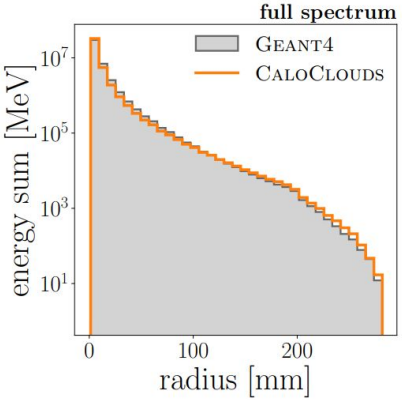
Reverse Diffusion Process

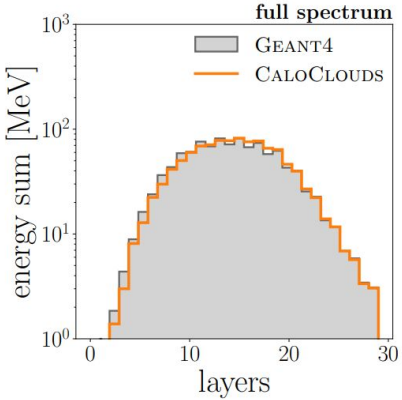


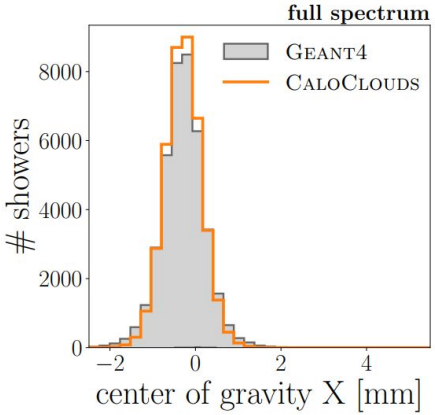
Point Cloud + Diffusion Model

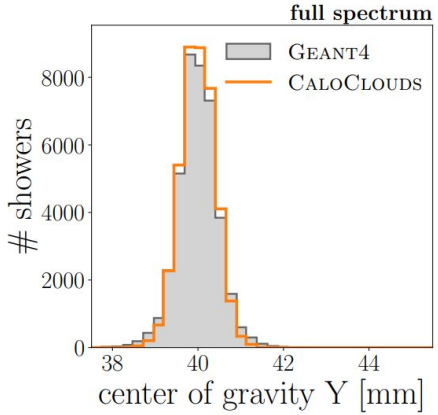
Results

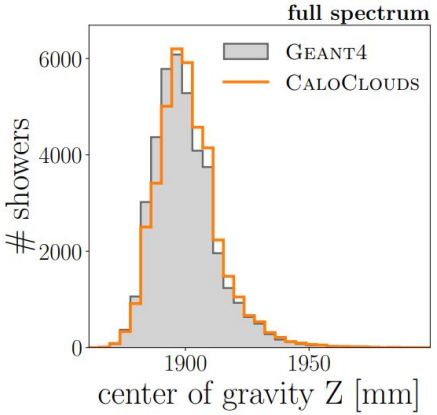






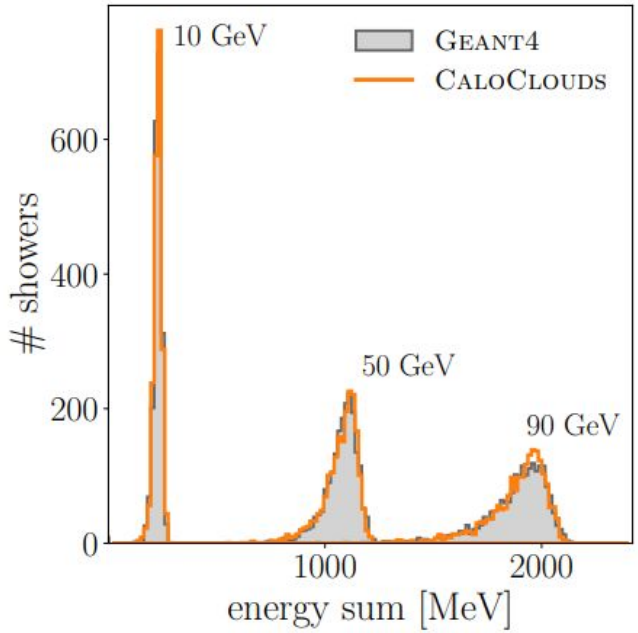


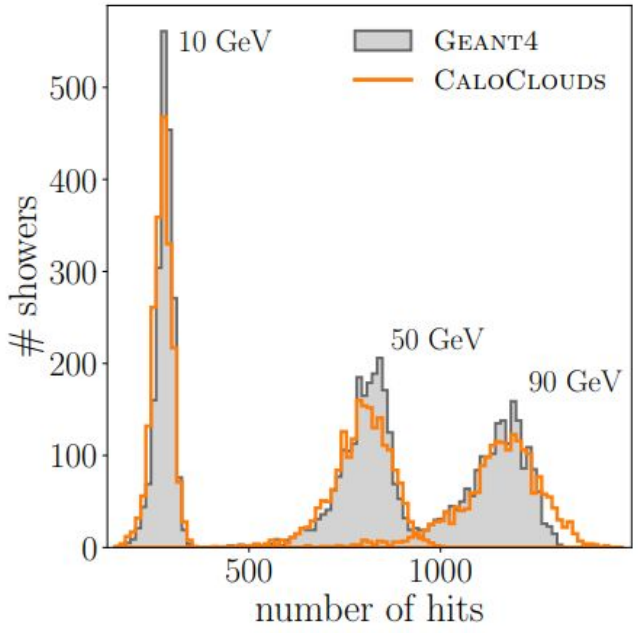




Point Cloud + Diffusion Model

Results

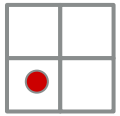


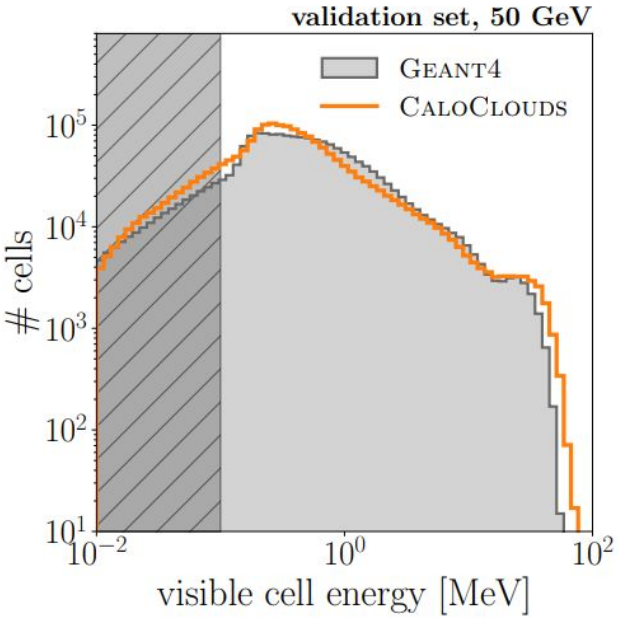


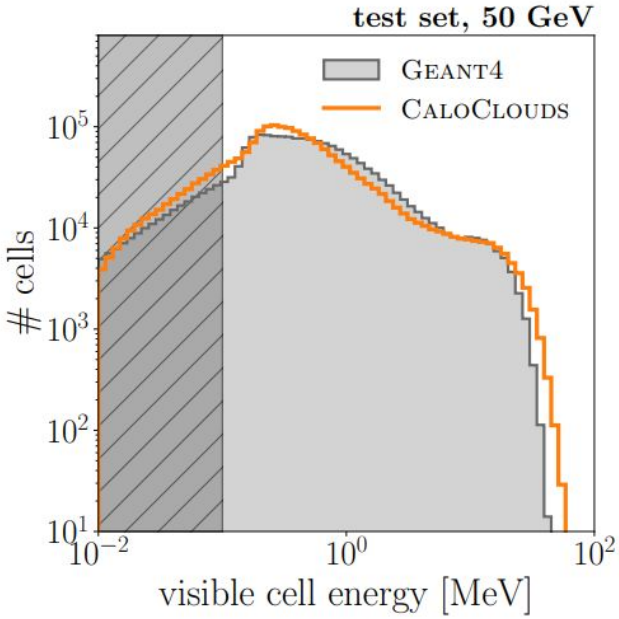
Point Cloud + Diffusion Model

Results

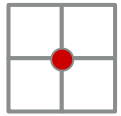
validation incident point pos.







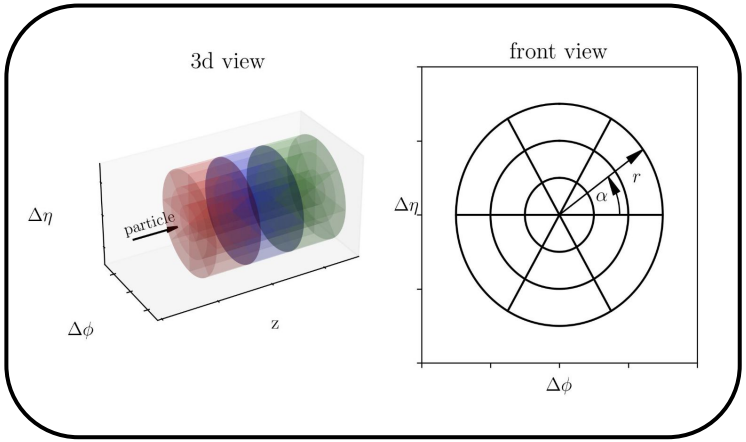
test incident point pos.

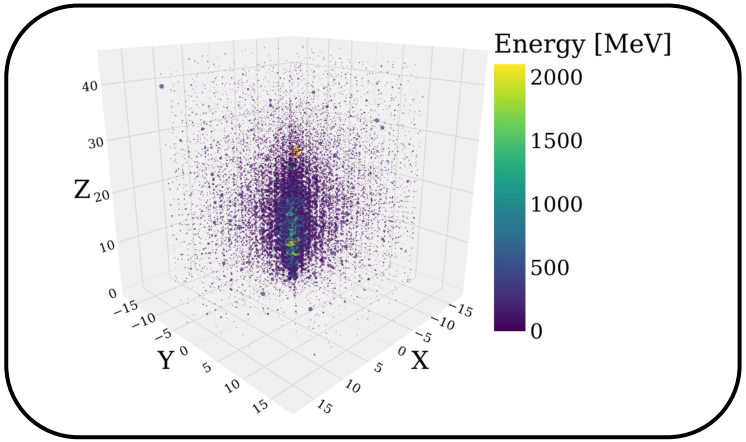


Per-cell energy distribution for the 50 GeV validation (left) data set, created at the same position as the training data set and for a 50 GeV test (right) data set simulated at a different position with the generated point cloud translated to this position

CaloChallenge

Data Preprocessing





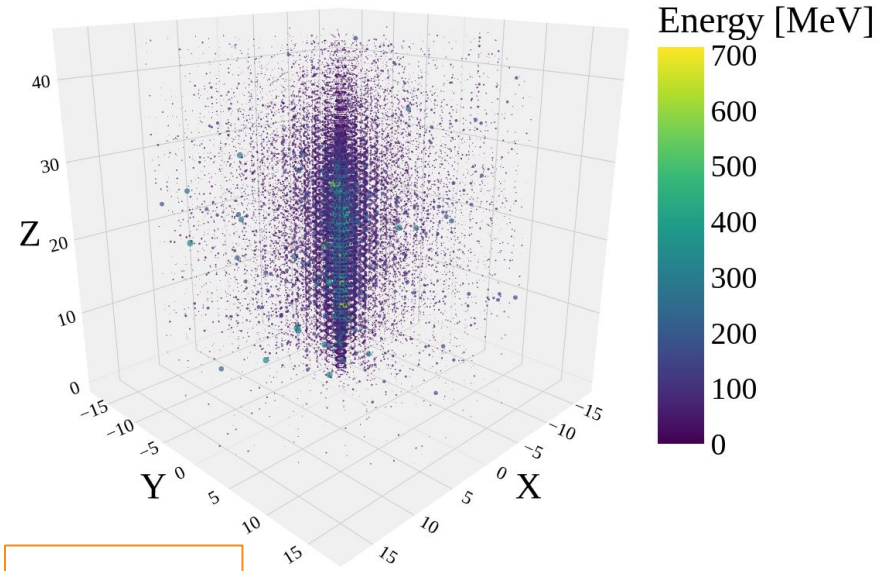
dataset 3, electron, 1 TeV

- Cylindrical to cartesian coordinates transform
- Incident energy normalization, feature scaling
- Coordinates normalization, feature scaling

CaloChallenge

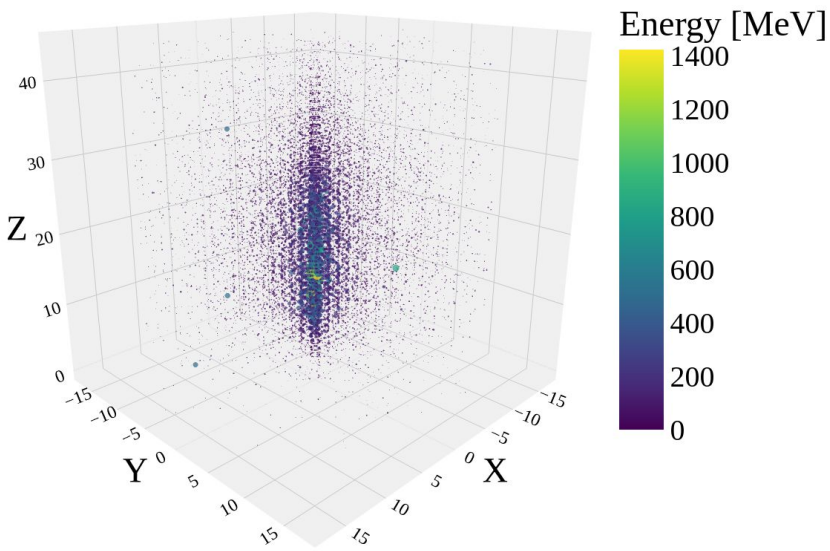
Results, Single Event Comparison

Geant 4



E = 1 TeV

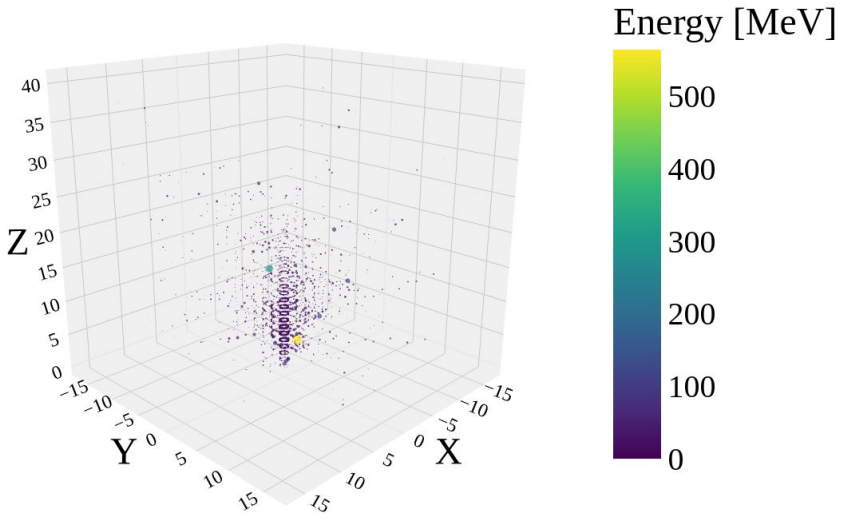
CaloCloud, time stamp: t



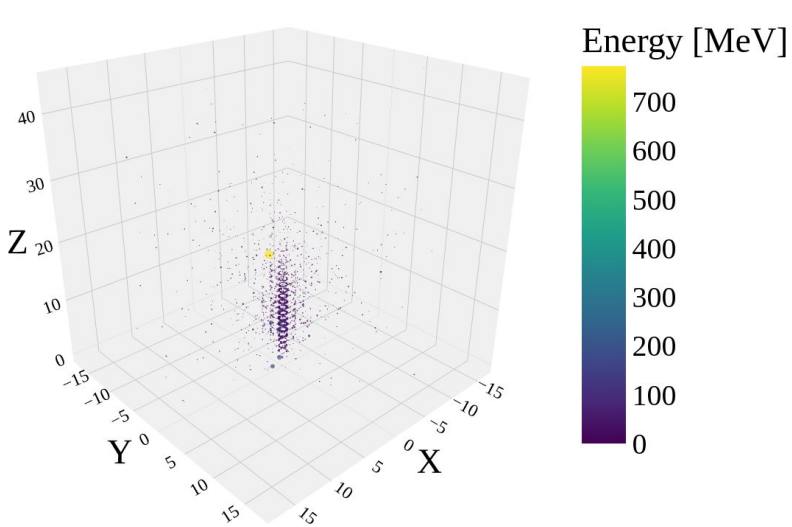
CaloChallenge

Results, Single Event Comparison

Geant 4



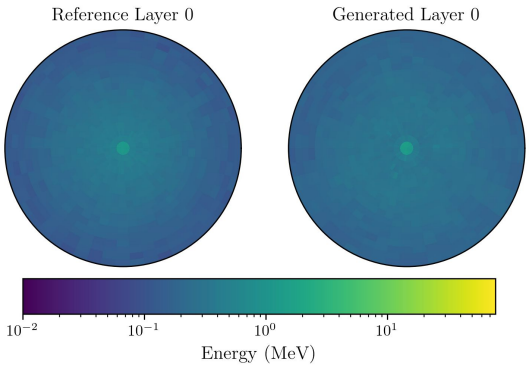
CaloCloud, time stamp: t_0

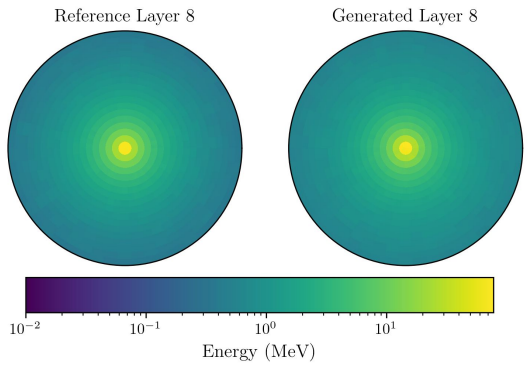


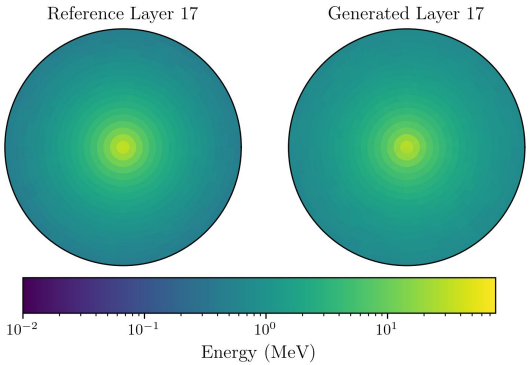
E = 30 GeV

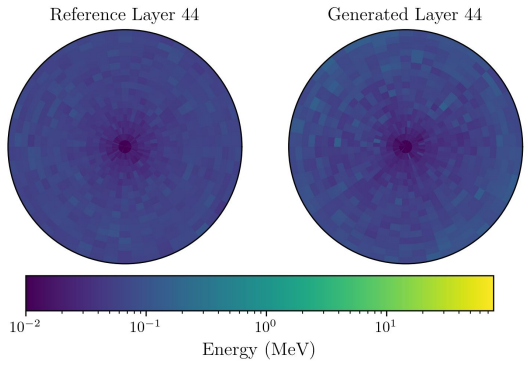
CaloChallenge

Results



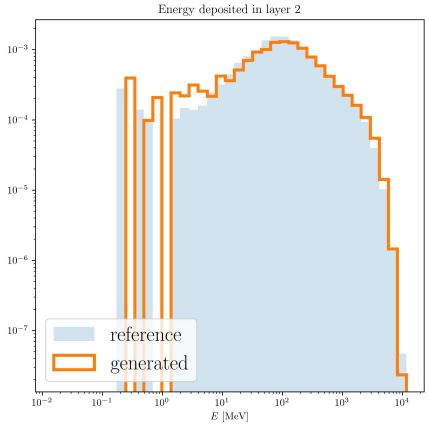


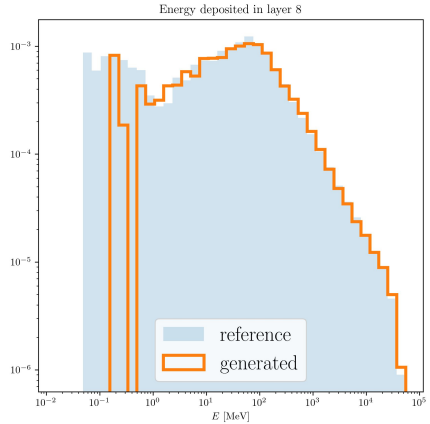


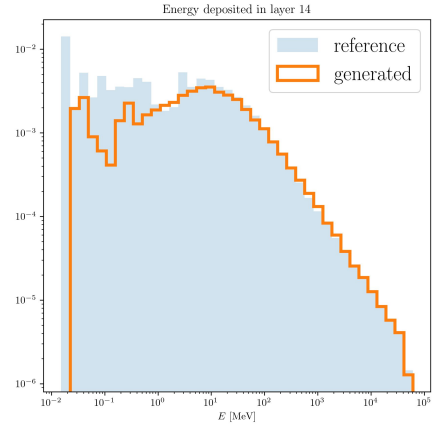


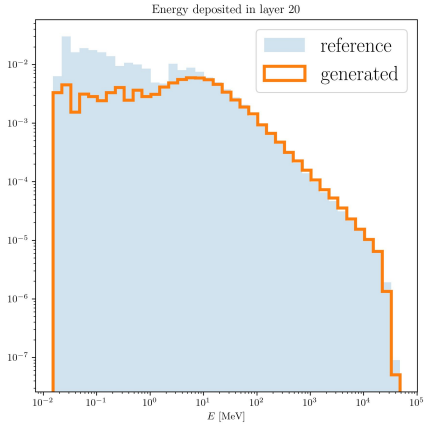
CaloChallenge

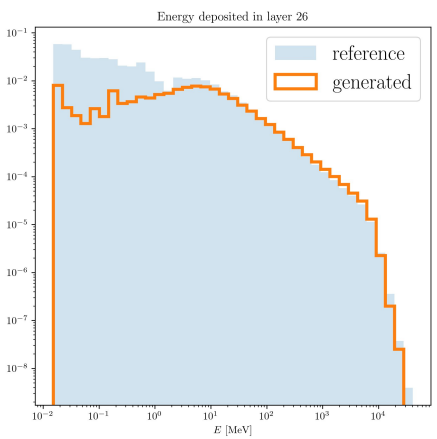
Results

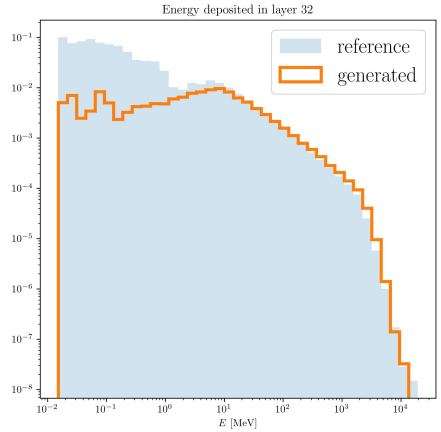






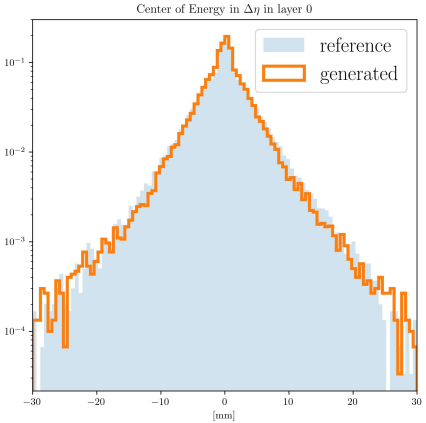


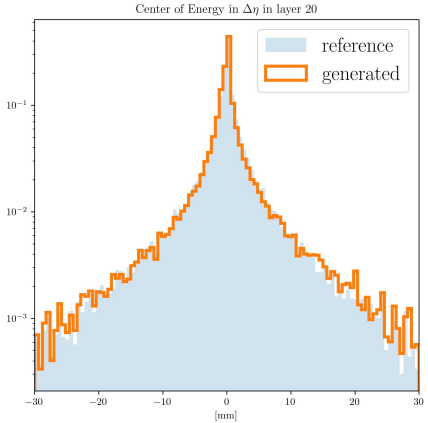


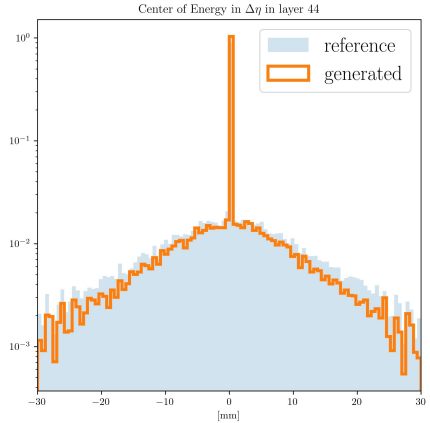


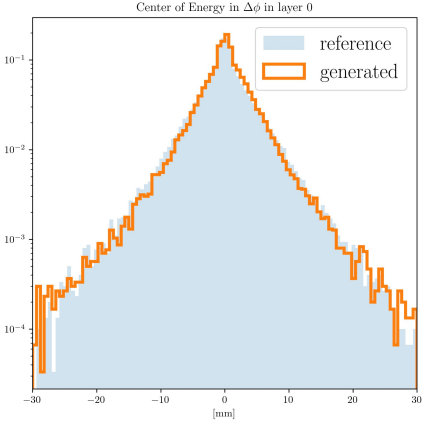
CaloChallenge

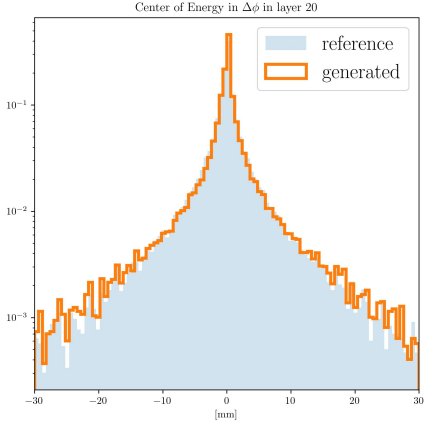
Results

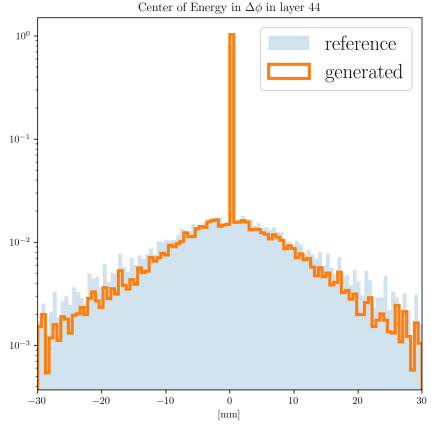






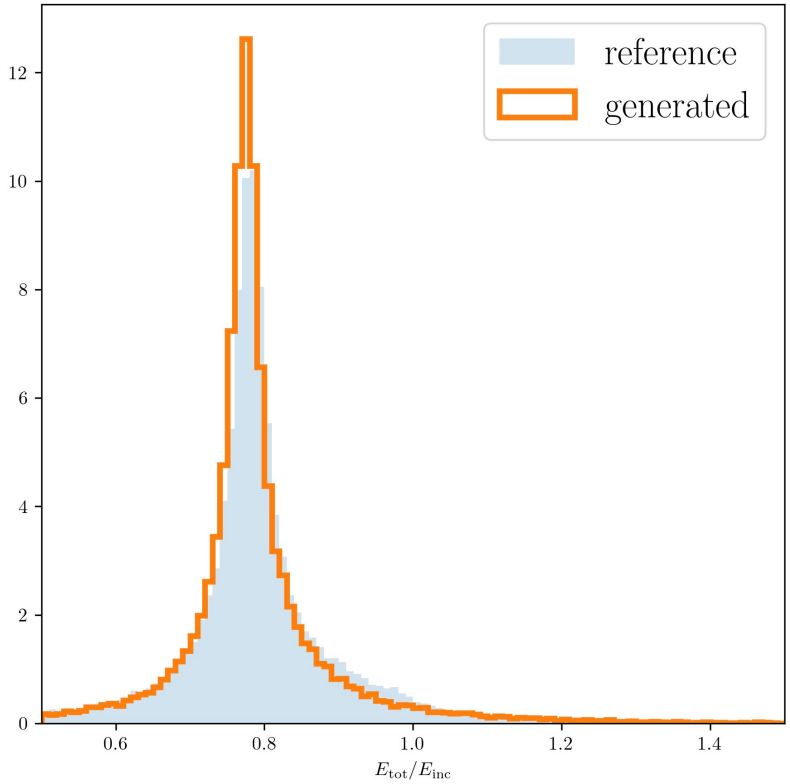


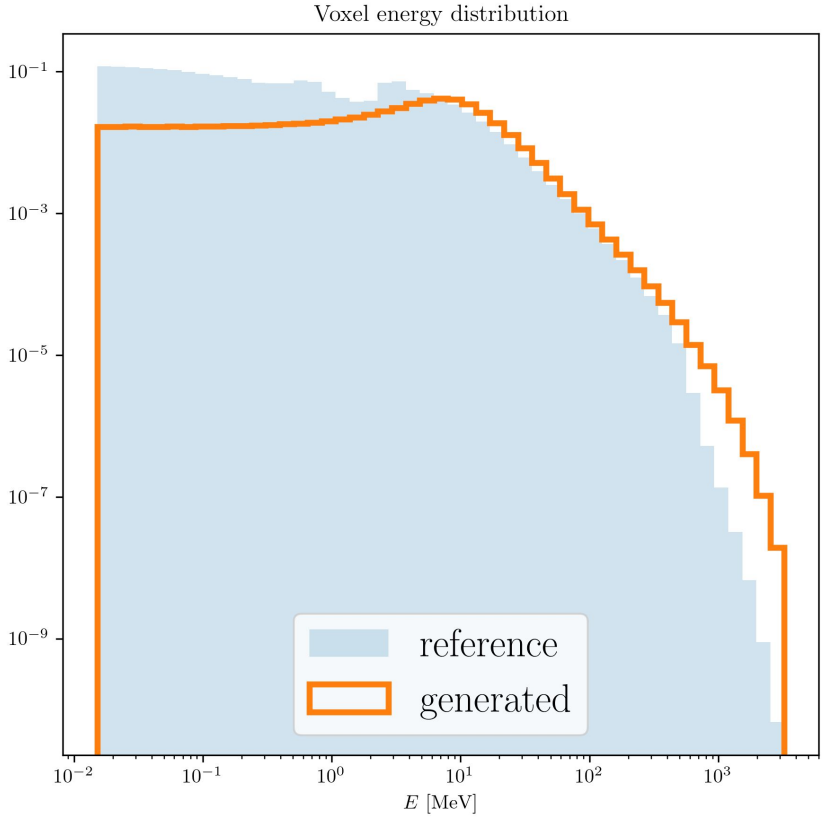




CaloChallenge

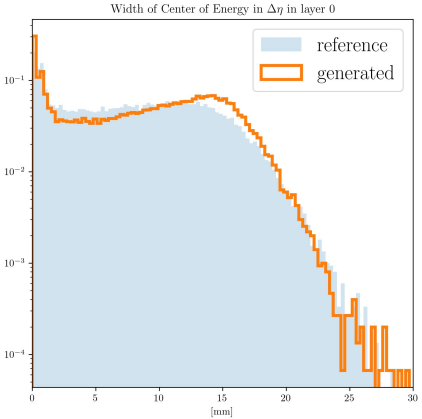
Results

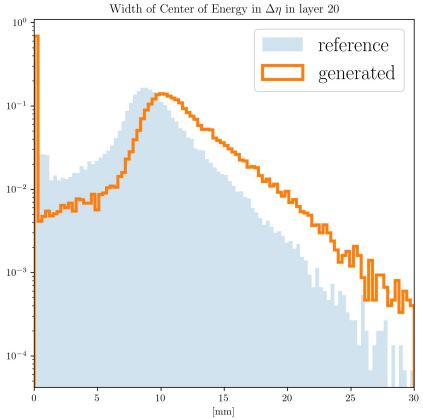


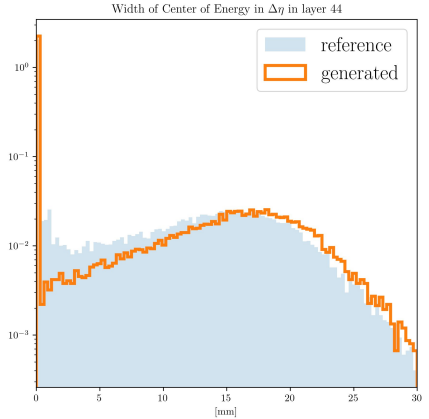


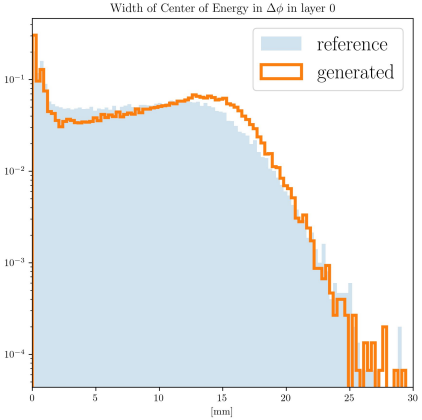
CaloChallenge

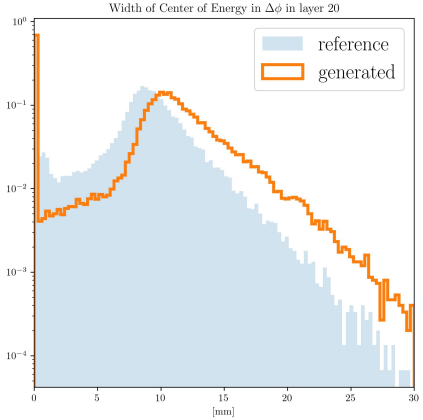
Results

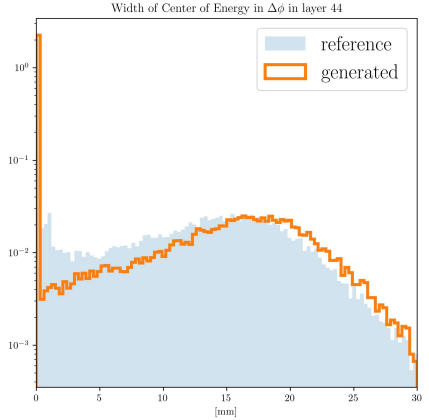












Summary

- Investigated new generative model architecture for generating EM showers data
- Fidelity of the physical properties learned well, but still have to be improved to achieve better agreement with GEANT4
- The combination of **Point Clouds** representation of EM showers and **Diffusion Model** looks promising as **a setup for easy integration into the simulation pipeline**
- Next step: work on timing performance, e.g. via distillation

BACKUP SLIDES

Point Cloud + Diffusion Model

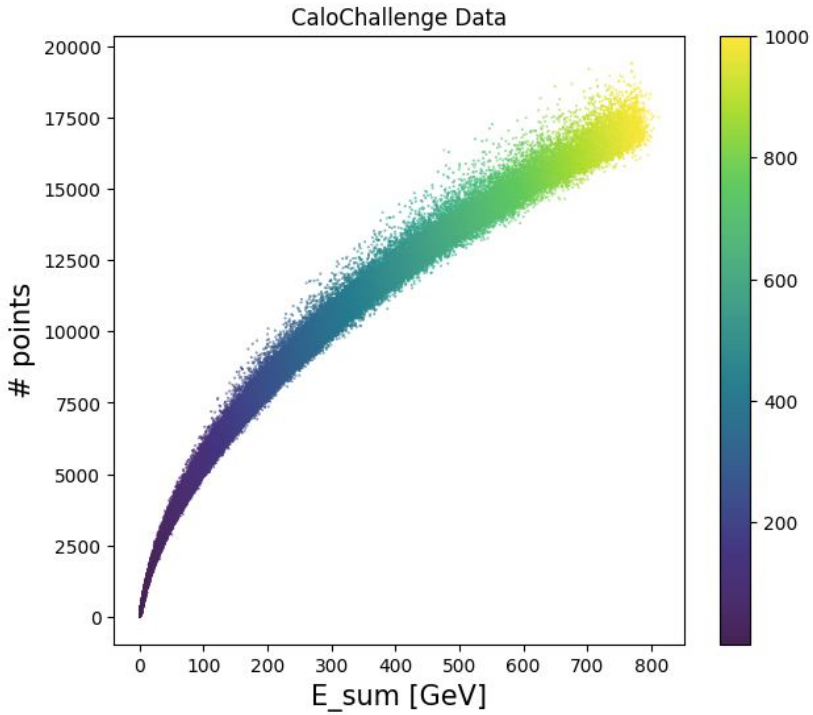
Results, Potential Speed-up

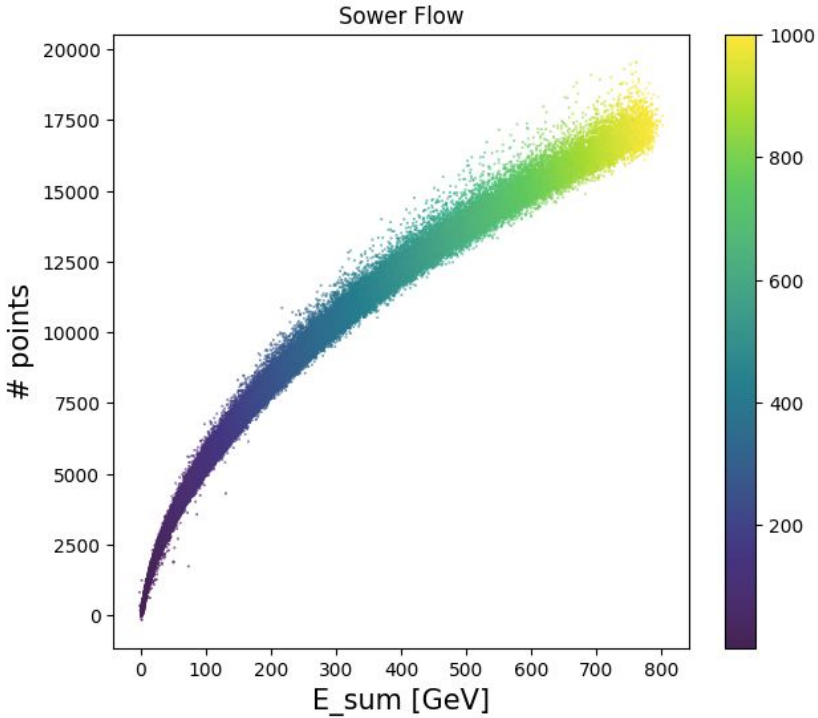
Hardware	Simulator	Time / Shower [ms]	Speed-up
CPU	GEANT4	4082 ± 170	$\times 1$
	CALOCLOUDS	3509 ± 220	$\times 1.2$
GPU	CALOCLOUDS	38 ± 3	$\times 107$

Timing performance needs to be improved in further work

Shower Flow

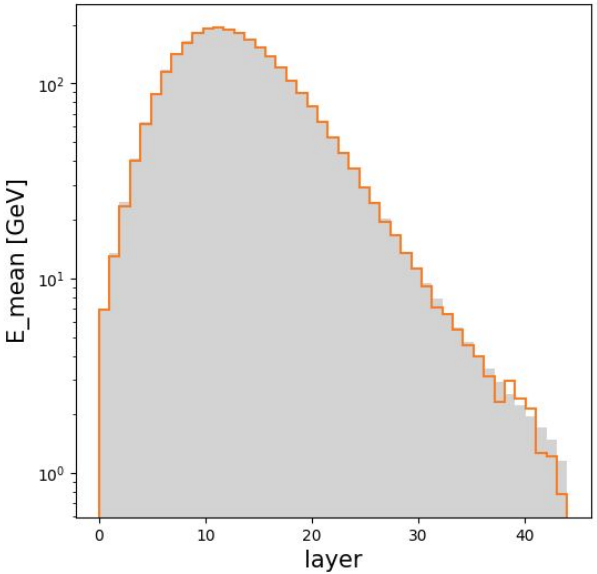
Rusalts

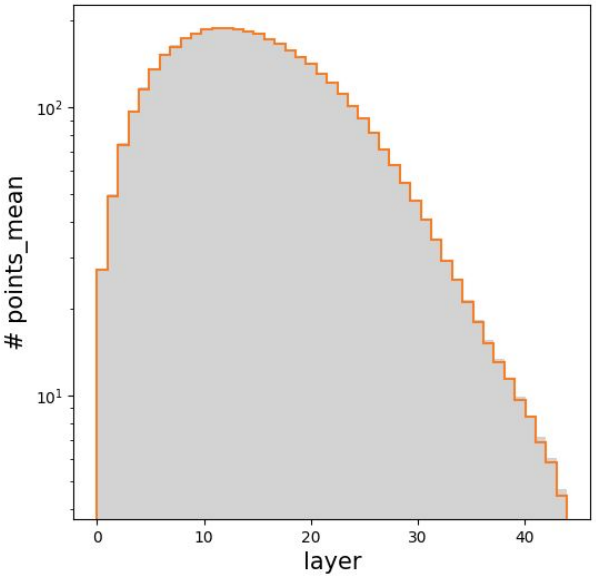


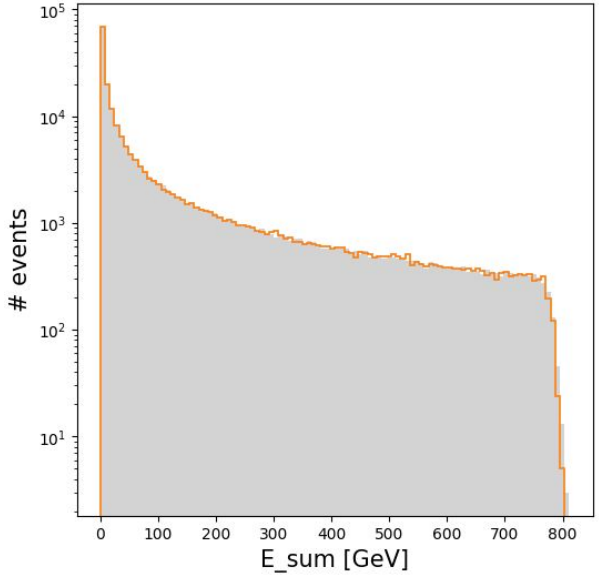


Shower Flow

Rusalts

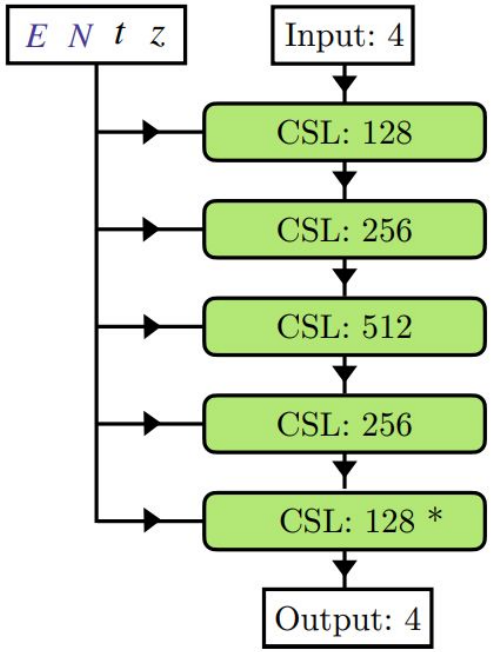




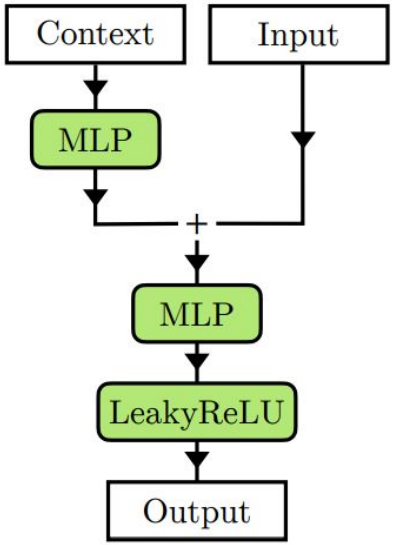


Point Cloud + Diffusion Model

PointWise Net



(a) PointWise Net

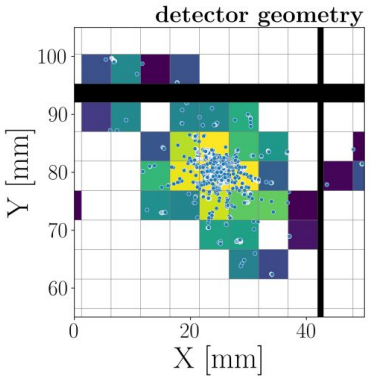


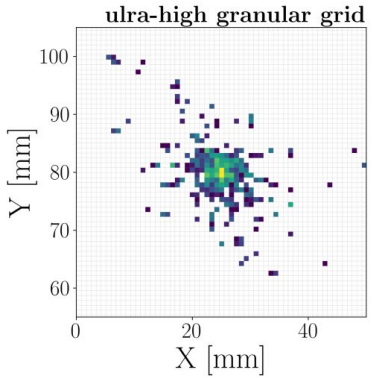
(b) ConcatSquash Layer (CSL)

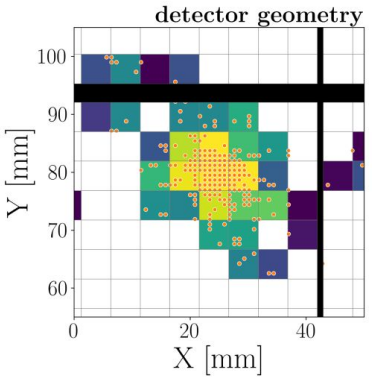
Point Cloud Representation of the EM Showers

Effects of the Pre-Clustering

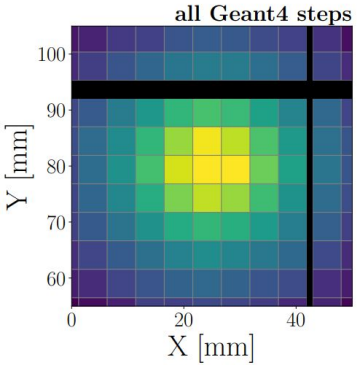
Single event of 90 GeV shower in 21th layer

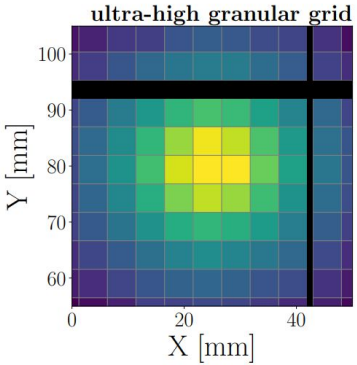


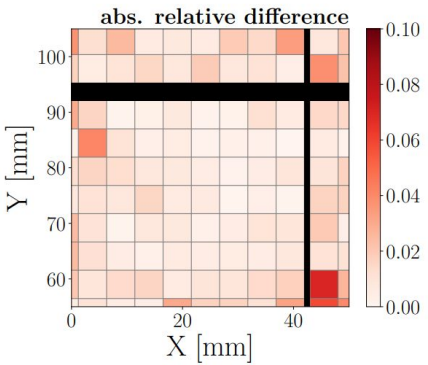




2k events of 90 GeV showers in 21th layer, overlay







Point Cloud Representation of the EM Showers

Effects of the Pre-Clustering