CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation

DPG Spring Meeting 2024

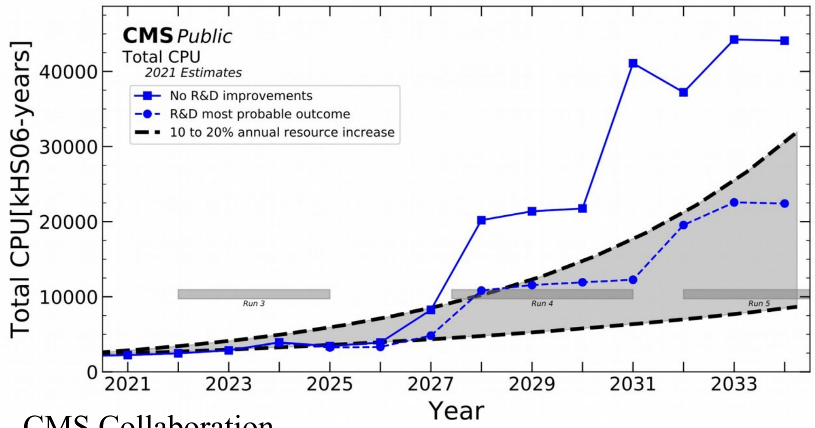
E. Buhmann[2], S. Diefenbacher[3], E. Eren[1], F. Gaede[1], G. Kasieczka[2], W. Korcari[2], **A. Korol[1*]**, K. Krüger[1], P. McKeown[1]

[1] Deutsches Elektronen-Synchrotron, DESY
[2] University of Hamburg, UHH
[3] Lawrence Berkeley National Laboratory, LBNL
*anatolii.korol@desy.de

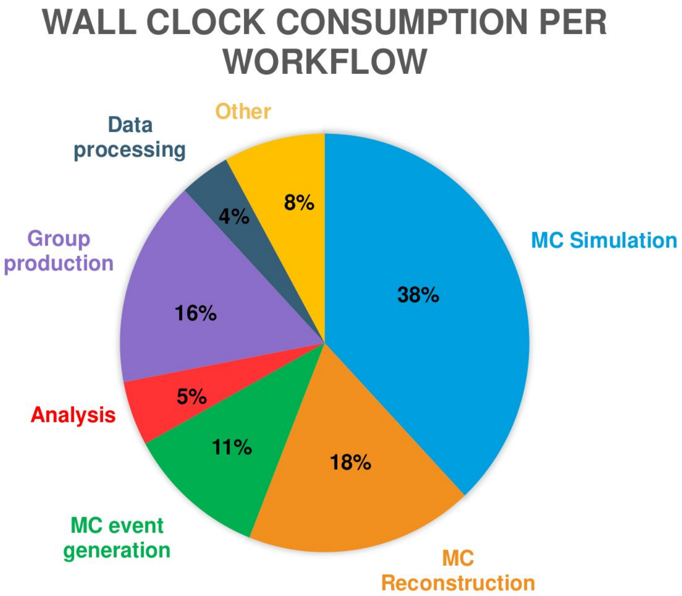
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE
HELMHOLTZ

Introduction

Time-consuming Simulations



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



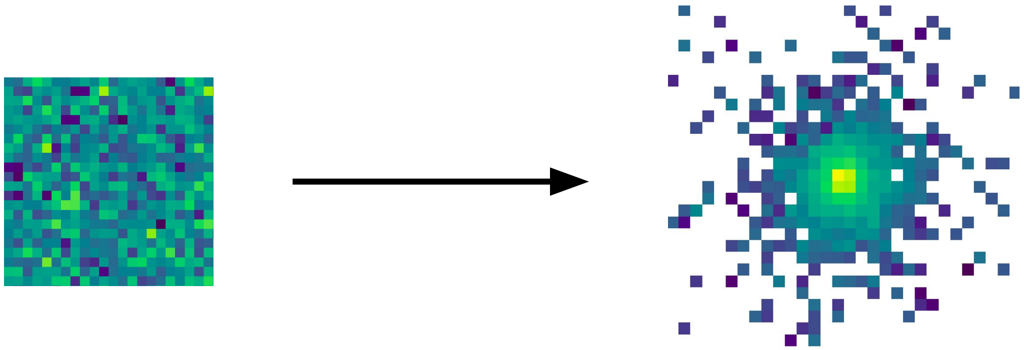
D. Costanzo, J. Catmore, ATLAS Computing update, LHCC meeting , 2019

Goal: replace (or augment) most intensive part of detector simulation (calorimeters simulation) with a faster generator, based on state-of-the-art machine learning techniques

Introduction

Generative Models

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

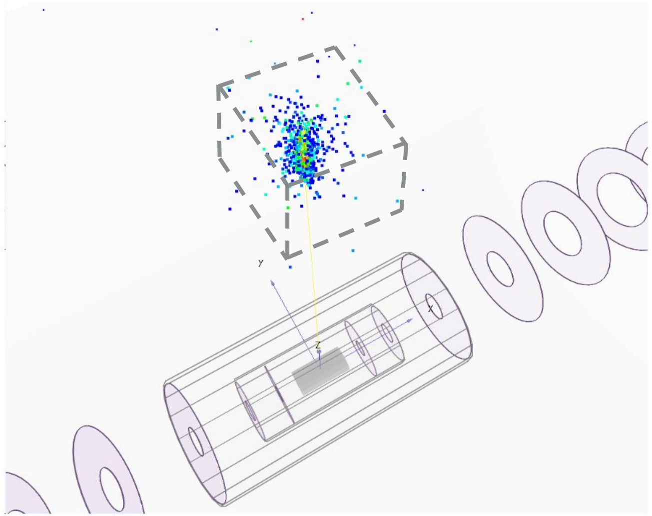


- There exist numerous generative models
  - Generative Adversarial Networks (GANs)
  - Flow-based models
  - Autoencoders (AE), e.g. BiB-AE
  - Diffuison Models (DMs)

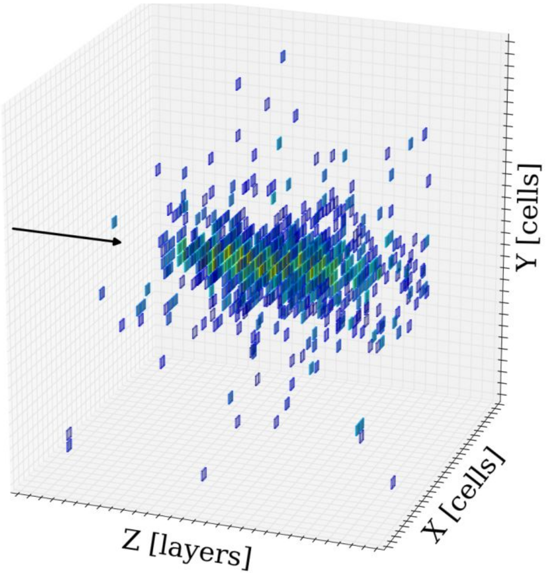
Image Representation of the EM Showers

ILD Detector

A simulated 60 GeV photon shower in the ILD detector



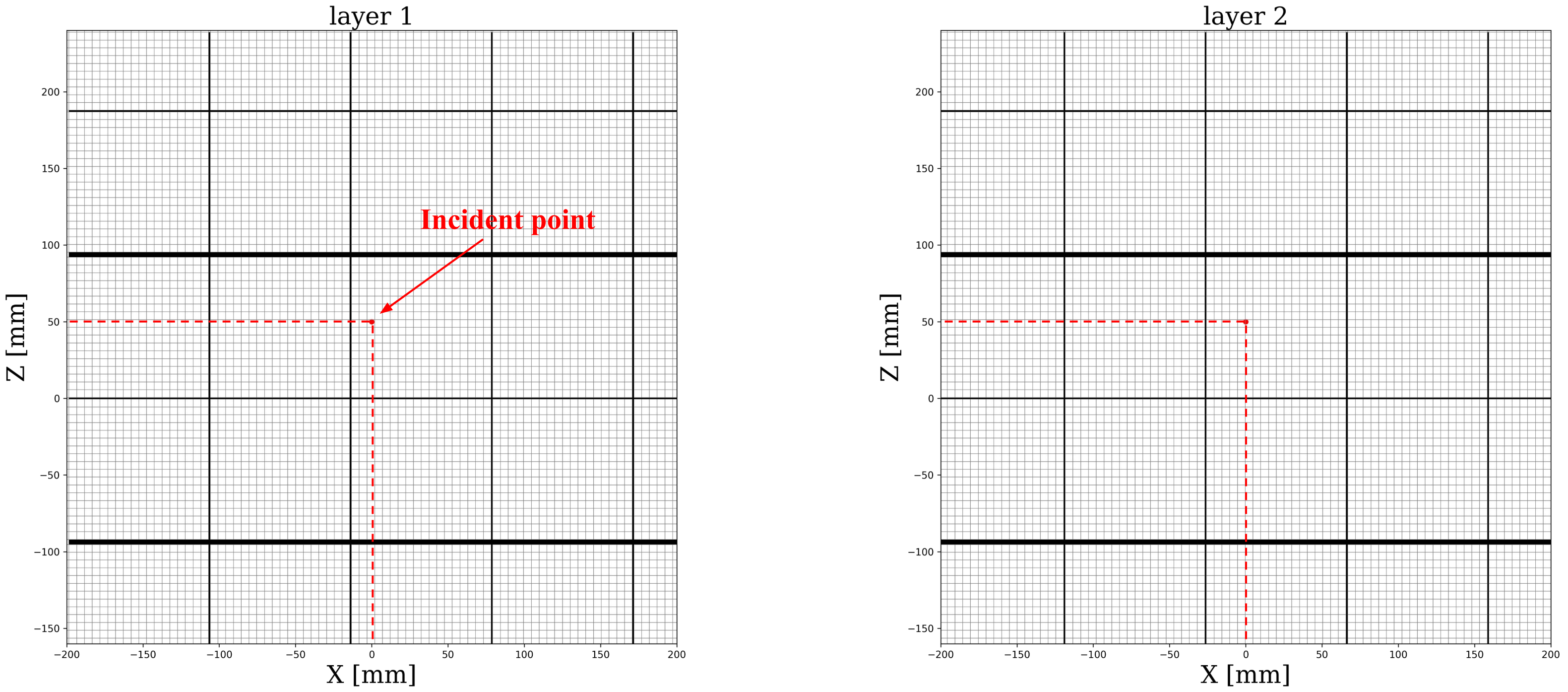
Regular grid 30x30x30



One to one mapping from detector geometry to a regular grid

Problems with Image Representation of the EM Showers

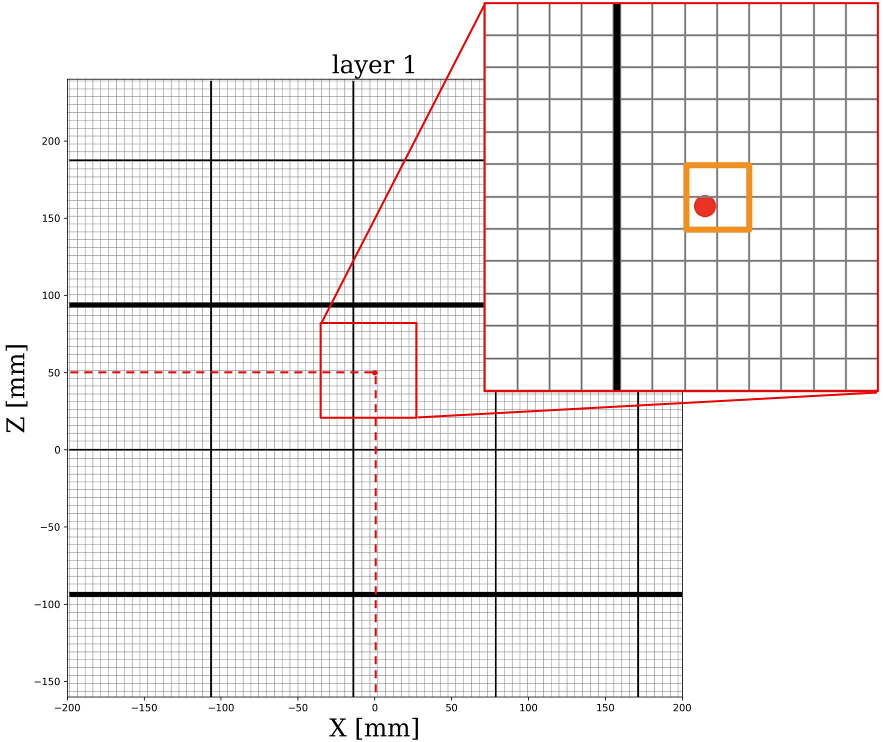
ILD Detector, ECAL Layers Structure

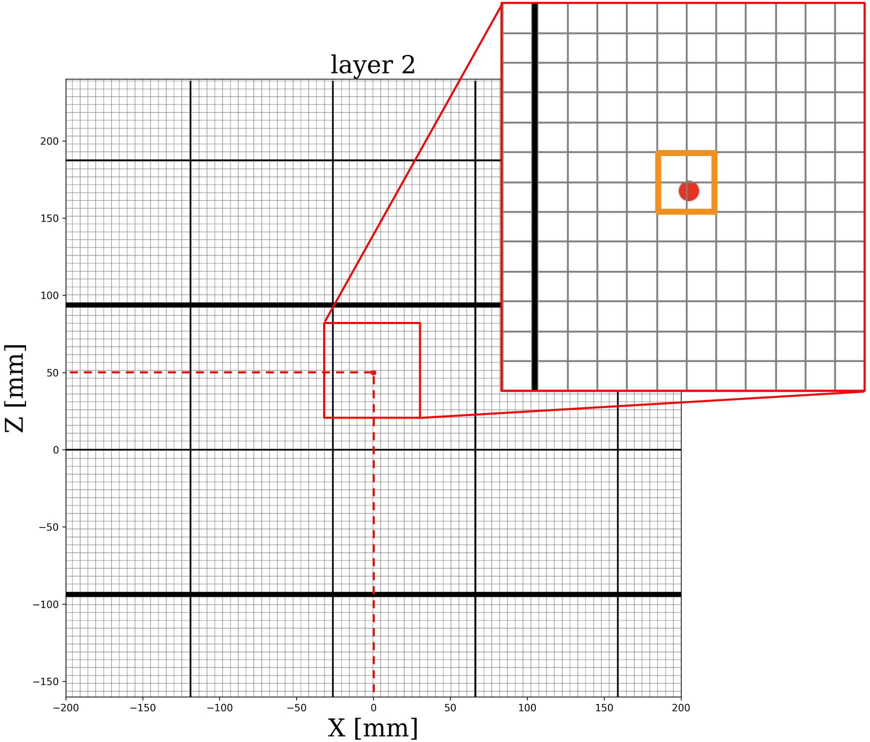


White squares represent **active cells**. Black lines are wafers, construction gaps, etc. (**not active material**)

Problems with Image Representation of the EM Showers

ILD Detector, ECAL Layers Structure

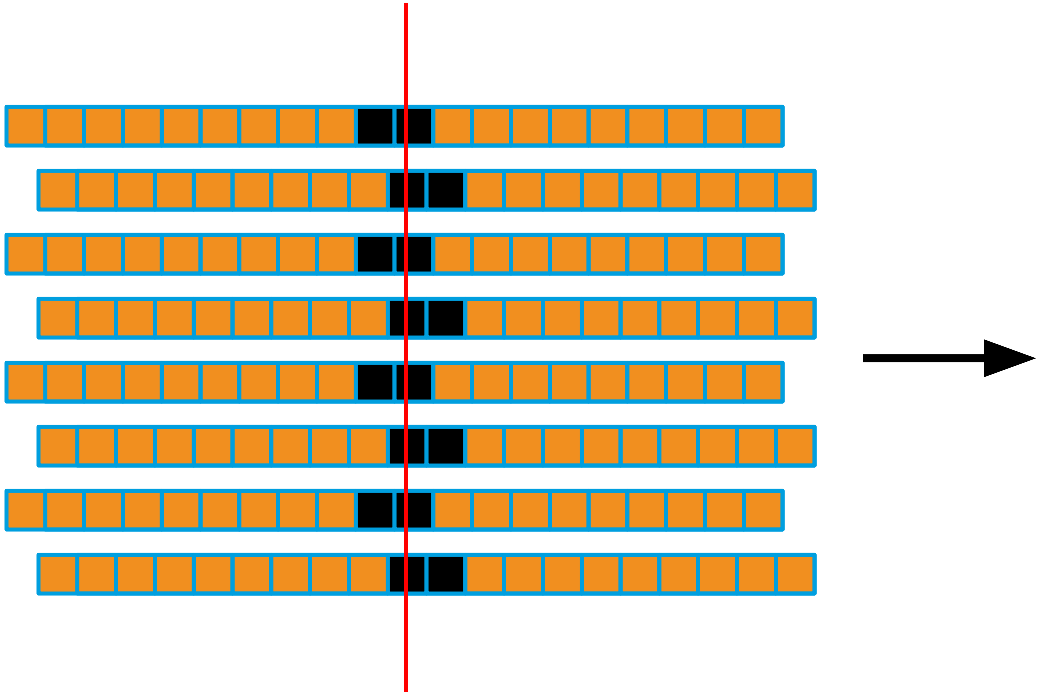


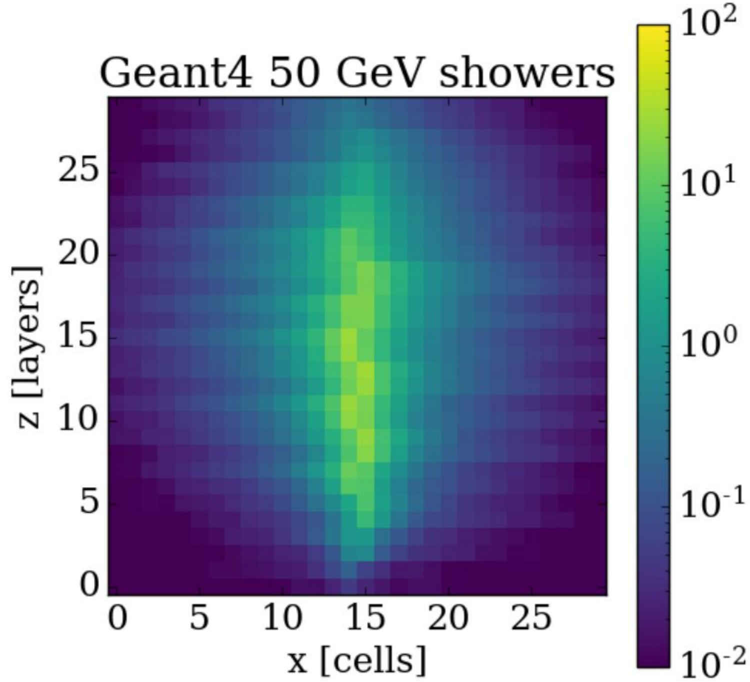


White squares represent **active cells**. Black lines are wafers, construction gaps, etc. (**not active material**)

Problems with Image Representation of the EM Showers

ILD Detector, ECAL Layers Structure, Staggering Effect





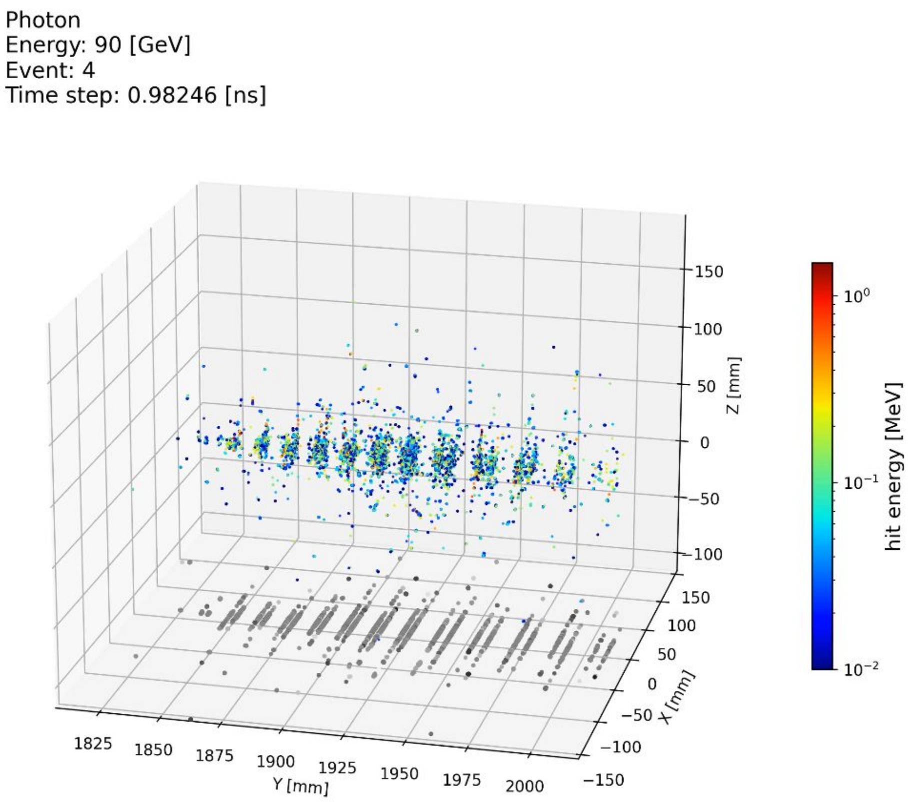
Models have to learn not only EM shower properties, but also geometry “artifacts”, like **staggering effect**

Point Cloud Representation of the EM Showers

GEANT4 Steps

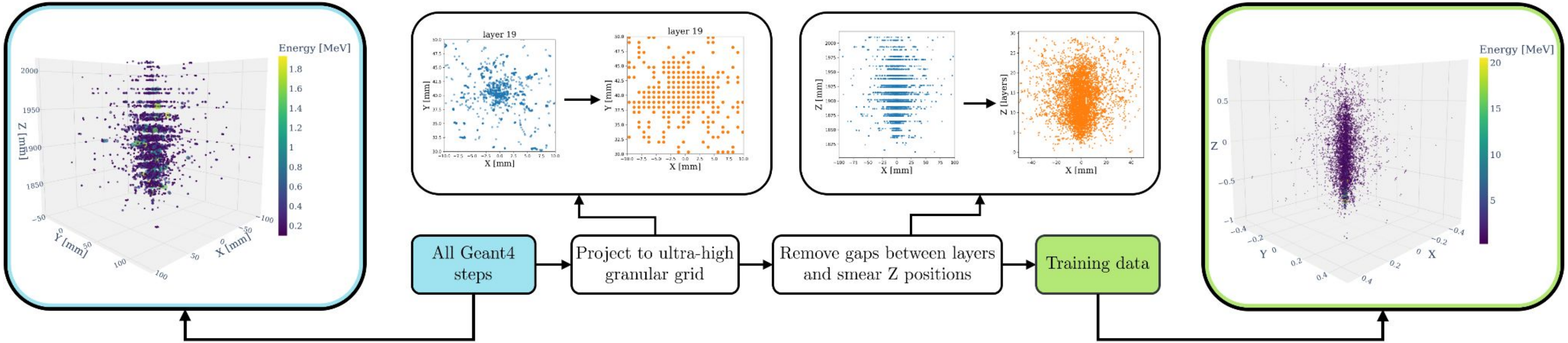
A way to overcome potential issues from irregular (realistic) cell geometry would be to use much higher granularity/resolution

- 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

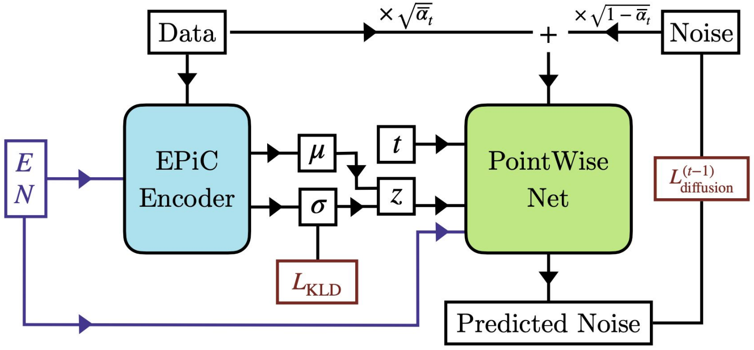


Number of points reduced to ~6k per shower, high enough resolution to move the shower in different place without harming physical properties of the shower

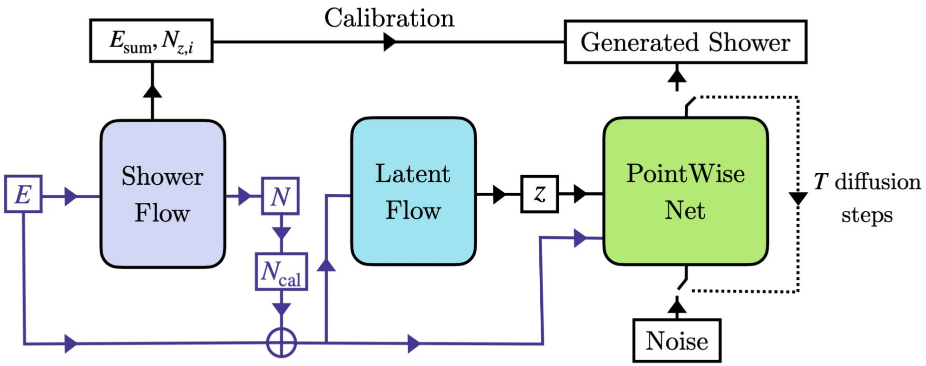
Point Cloud + Diffusion Model

CaloClouds, Model Overview

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2305.04847**



(a) Training at random time step t

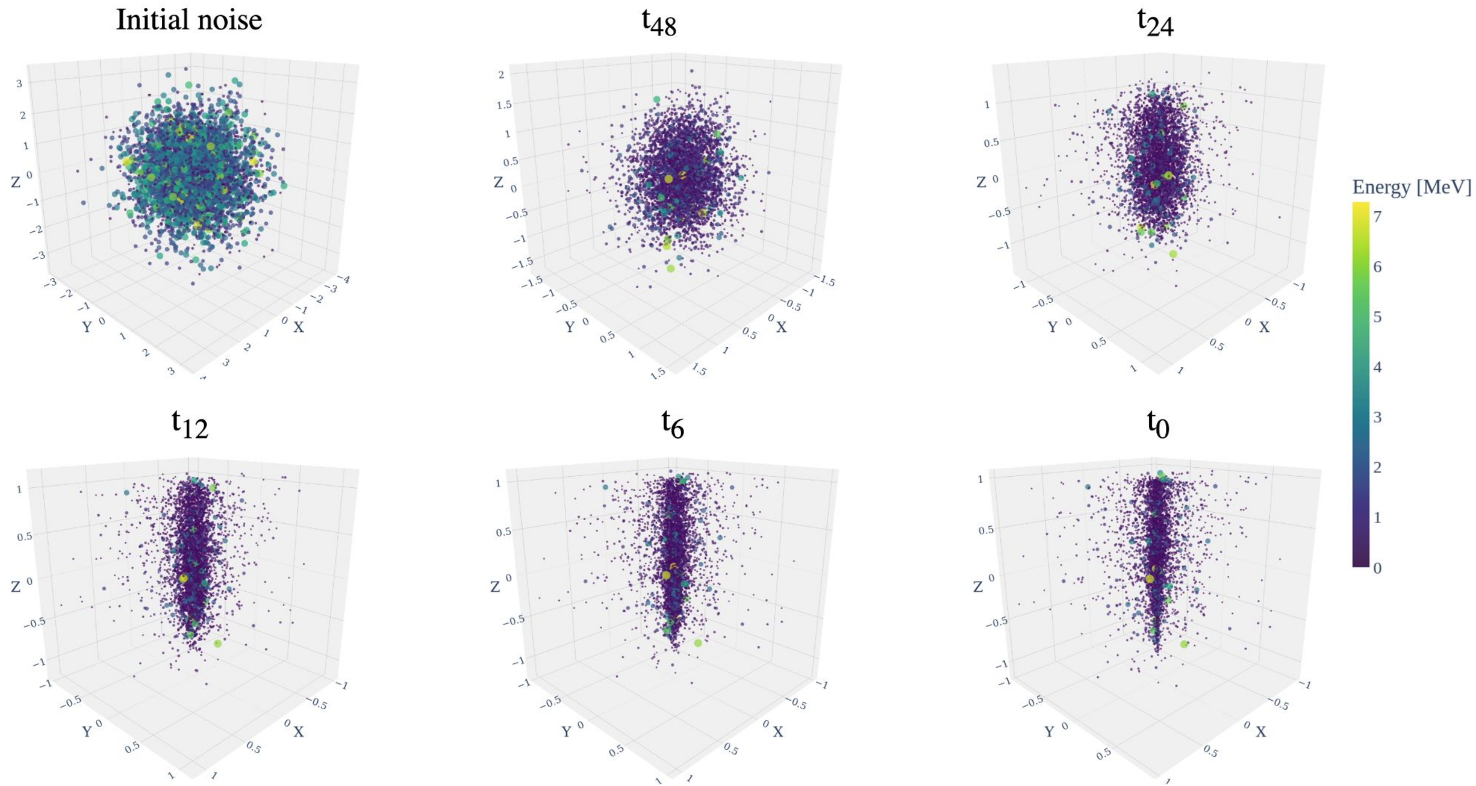


(b) Sampling with reverse diffusion through all time steps T

- 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

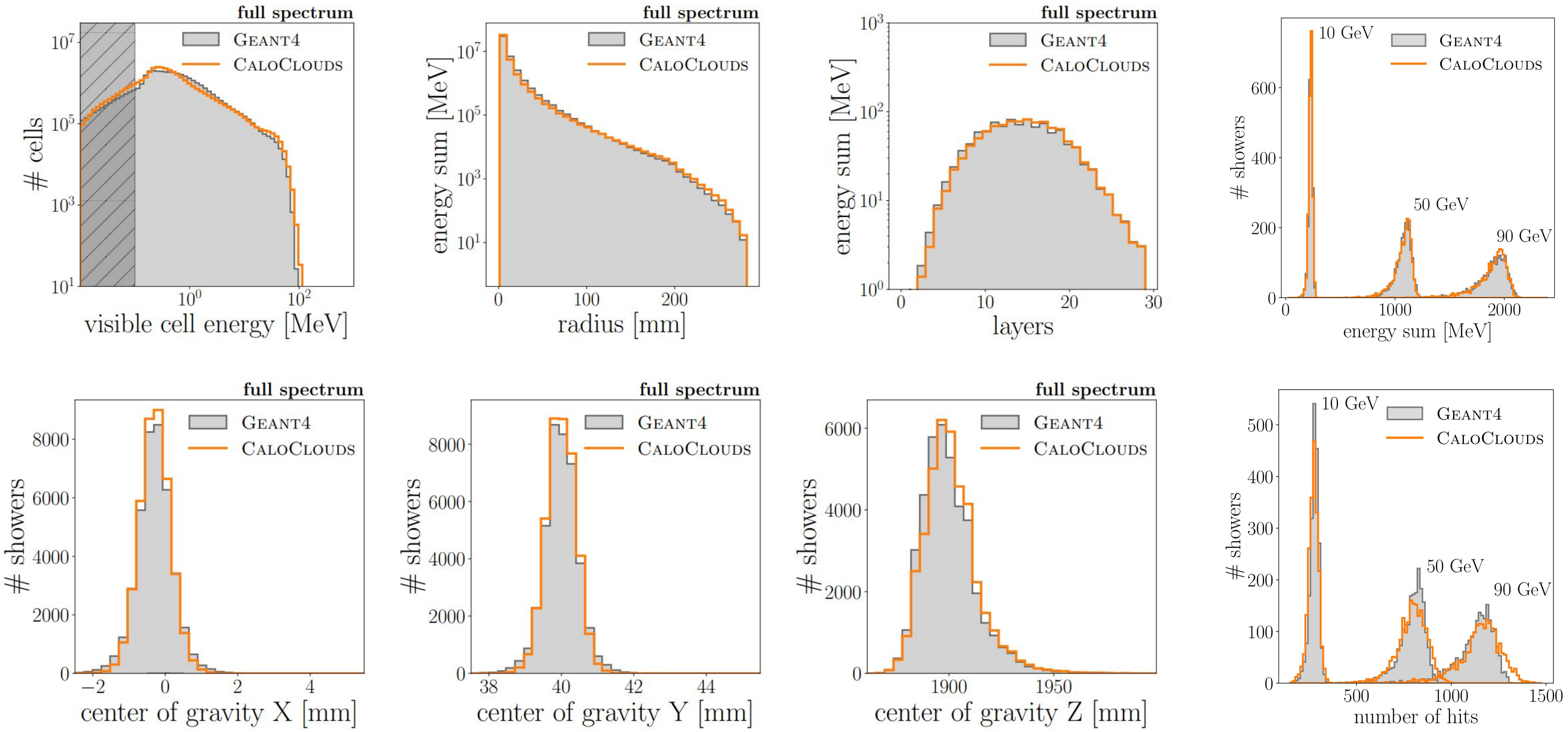
Reverse Diffusion Process



Point Cloud + Diffusion Model

Results

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2305.04847**

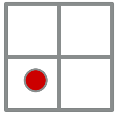


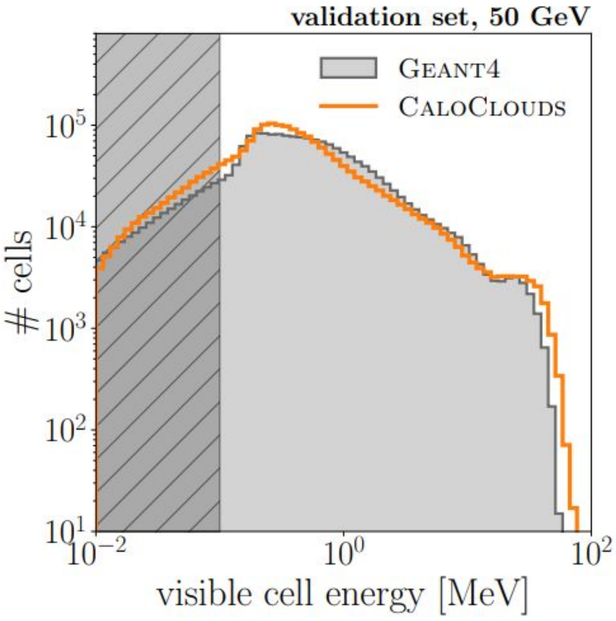
Point Cloud + Diffusion Model

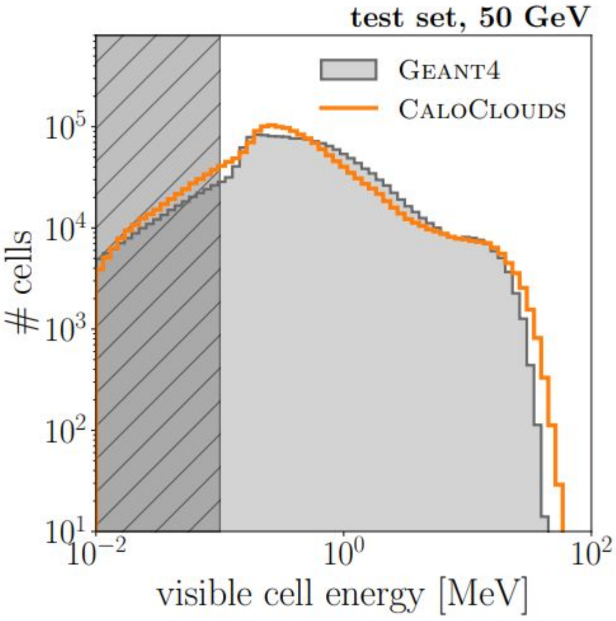
Results

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2305.04847**

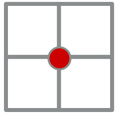
validation incident point pos.







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

Point Cloud + Diffusion Model

Results

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2305.04847**

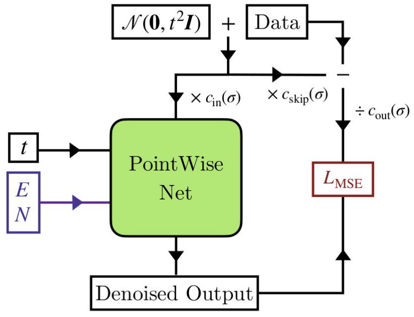
Hardware	Simulator	Time / Shower [ms]	Speed-up
CPU	GEANT4	4082 ± 170	$\times 1$
	CALOCLOUDS	3509 ± 220	$\times 1.2$
GPU	CALOCLOUDS	38 ± 3	$\times 107$

Not impressive inference time

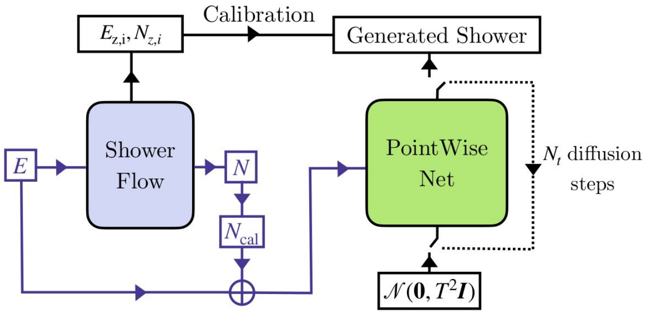
Point Cloud + Diffusion Model

CaloClouds II, Model Overview

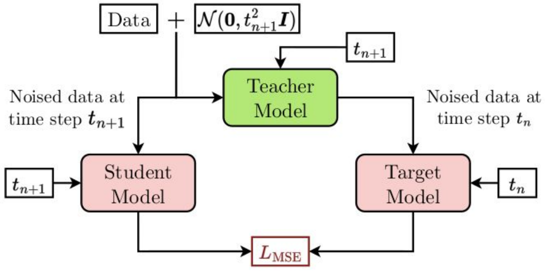
CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2309.05704**



(a) Training



(b) Sampling



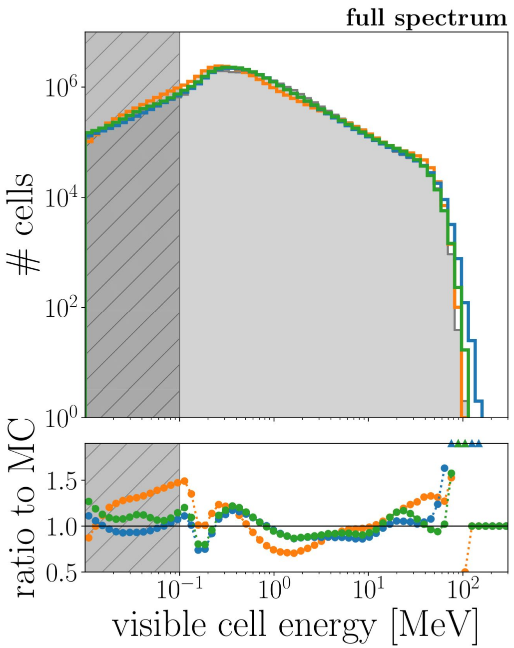
Consistency Distillation

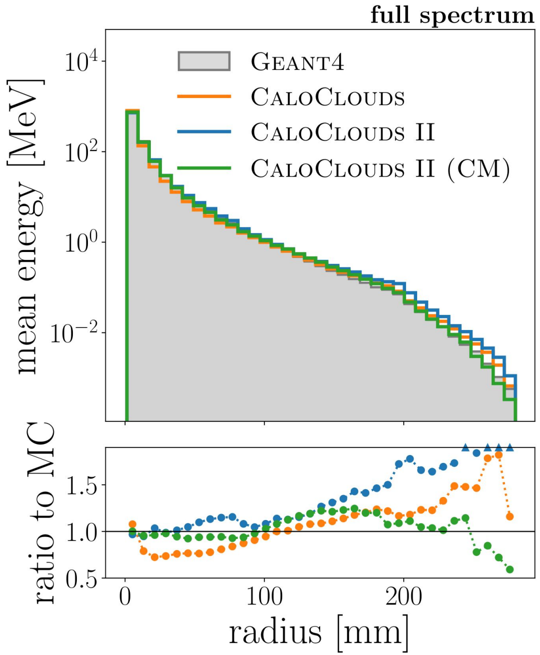
Modified version of CaloClouds version + Consistency Distillation → significantly reduced inference time

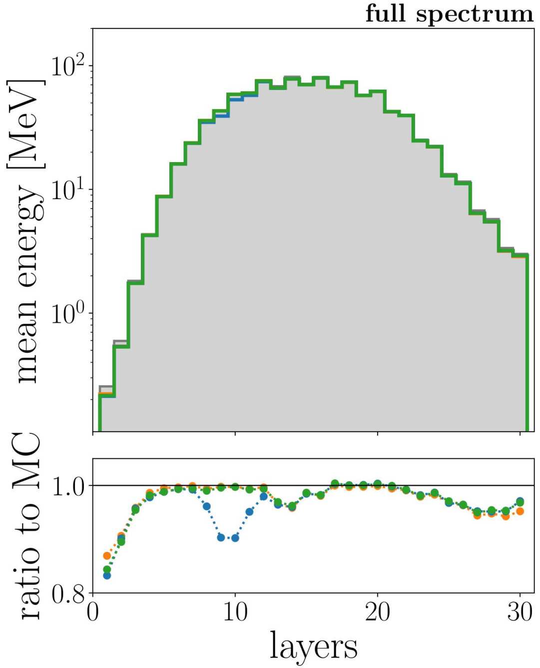
Point Cloud + Diffusion Model

Results

CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2309.05704**







Point Cloud + Diffusion Model

Results

CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2309.05704**

Hardware	Simulator	NFE	Batch Size	Time / Shower [ms]	Speed-up
CPU	GEANT4			3914.80 ± 74.09	×1
	CALOCLOUDS	100	1	3146.71 ± 31.66	×1.2
	CALOCLOUDS II	25	1	651.68 ± 4.21	×6.0
	CALOCLOUDS II (CM)	1	1	84.35 ± 0.22	×46
GPU	CALOCLOUDS	100	64	24.91 ± 0.72	×157
	CALOCLOUDS II	25	64	6.12 ± 0.13	×640
	CALOCLOUDS II (CM)	1	64	2.09 ± 0.13	×1873

Summary

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2305.04847**

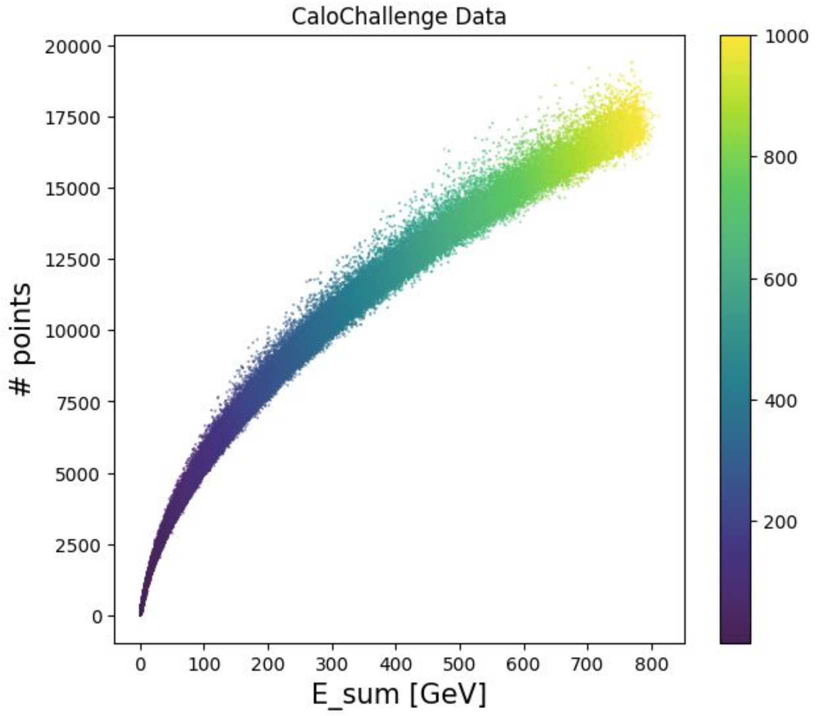
CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2309.05704**

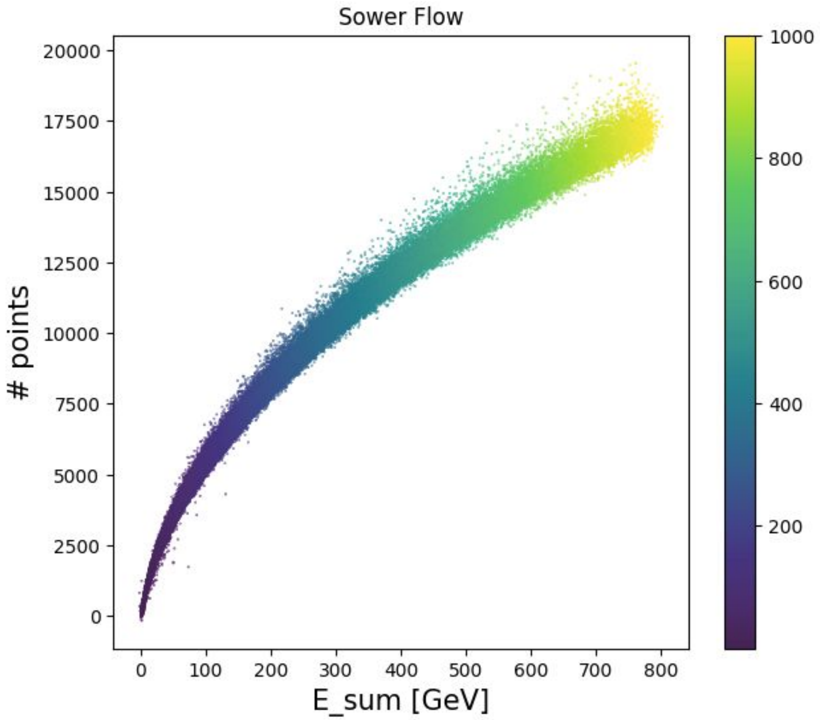
- Investigated new generative model architecture for generating EM showers data
- Fidelity of the physical properties learned well, but still have to be improved to achieve better agreement with GEANT4
- The combination of **Point Clouds** representation of EM showers and **Diffusion Model** looks promising as **a setup for easy integration into the simulation pipeline**
- Next steps: extend functionality (e.g. angular conditioning), study full physics benchmarks

BACKUP SLIDES

Shower Flow

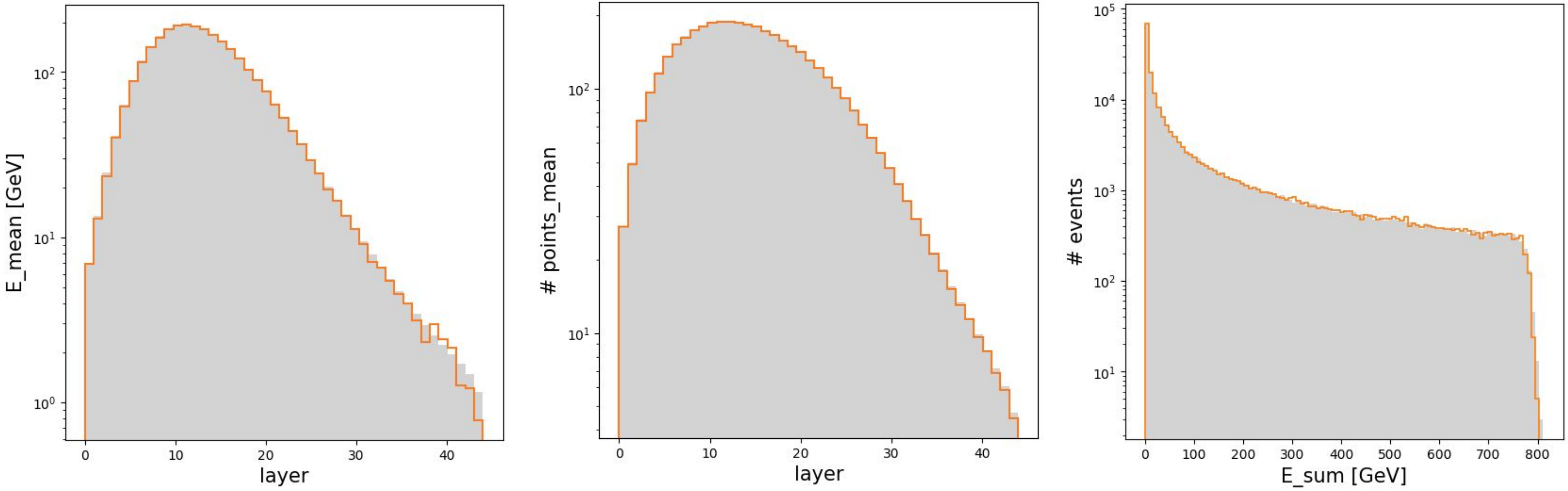
Results





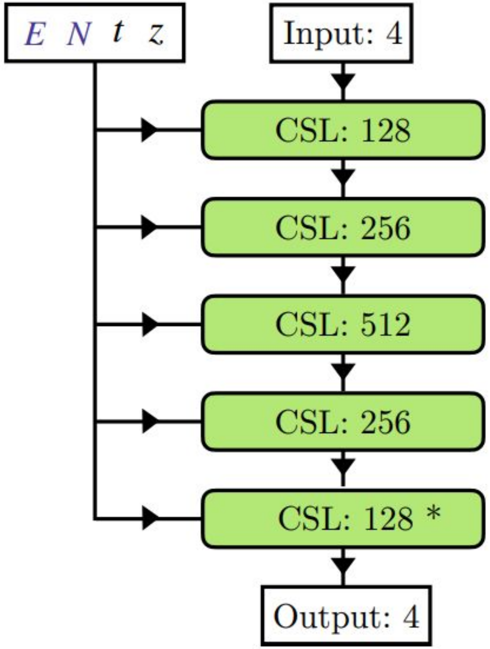
Shower Flow

Results

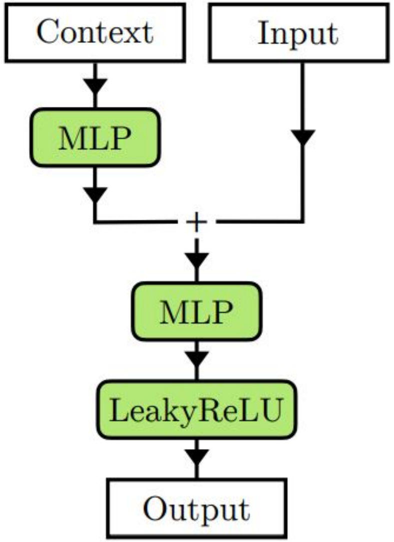


Point Cloud + Diffusion Model

PointWise Net



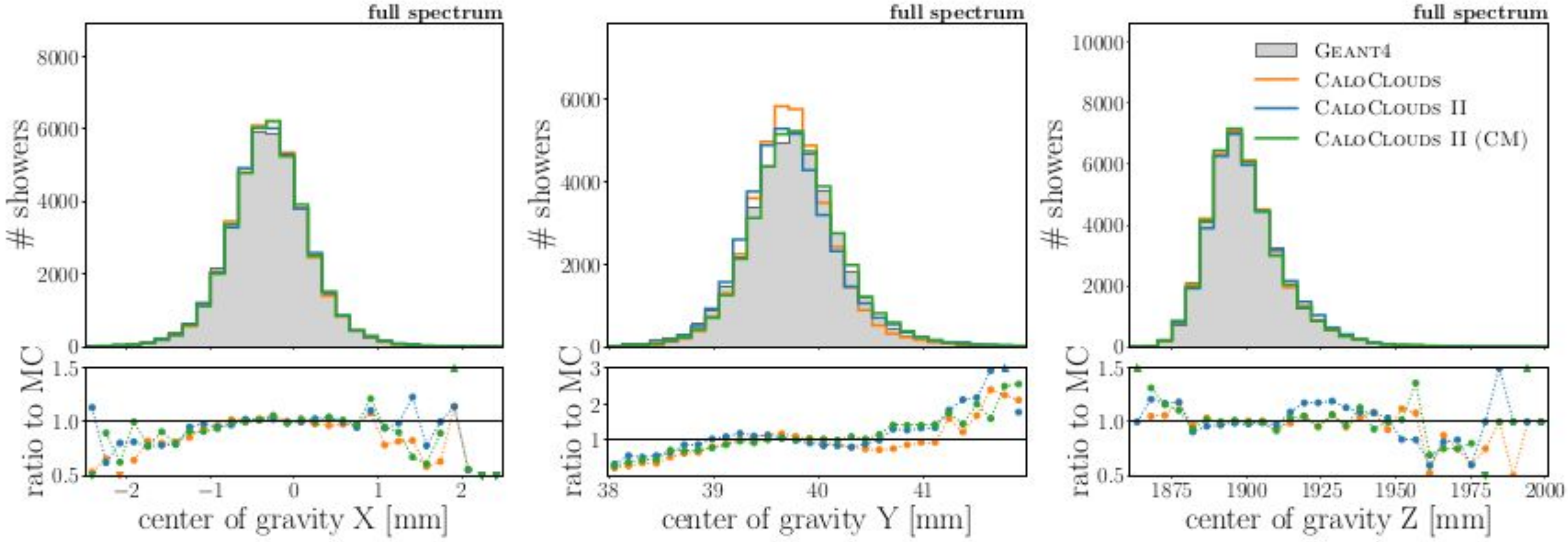
(a) PointWise Net



(b) ConcatSquash Layer (CSL)

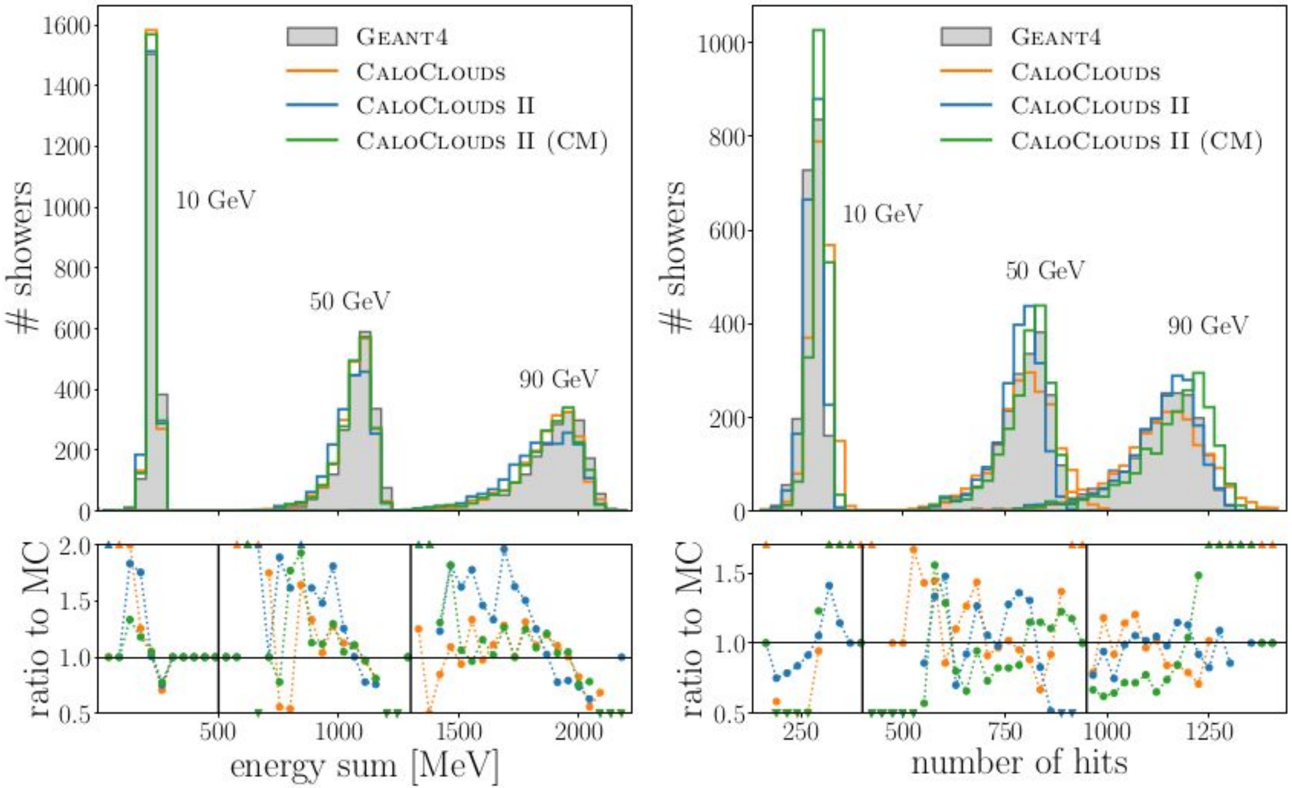
Point Cloud + Diffusion Model

Results, Position of the Center of Gravity



Point Cloud + Diffusion Model

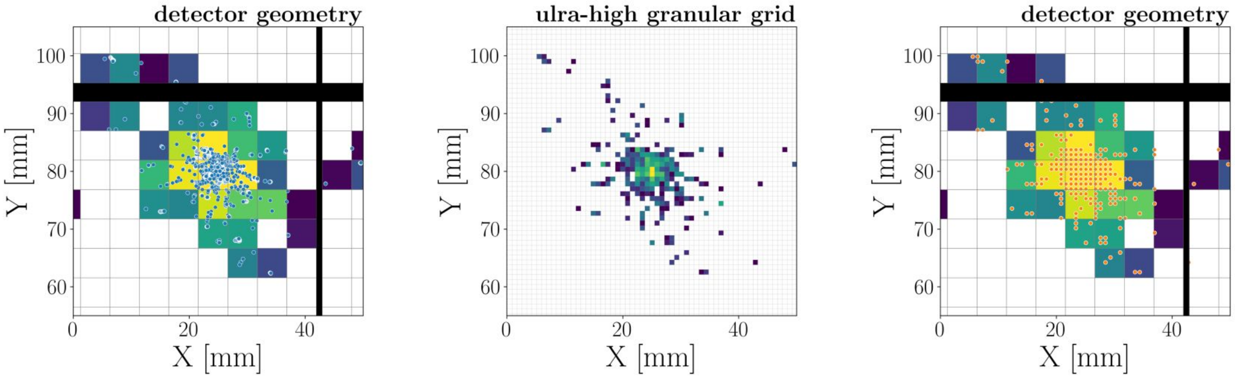
Results, Visible Energy and the Number of Hits



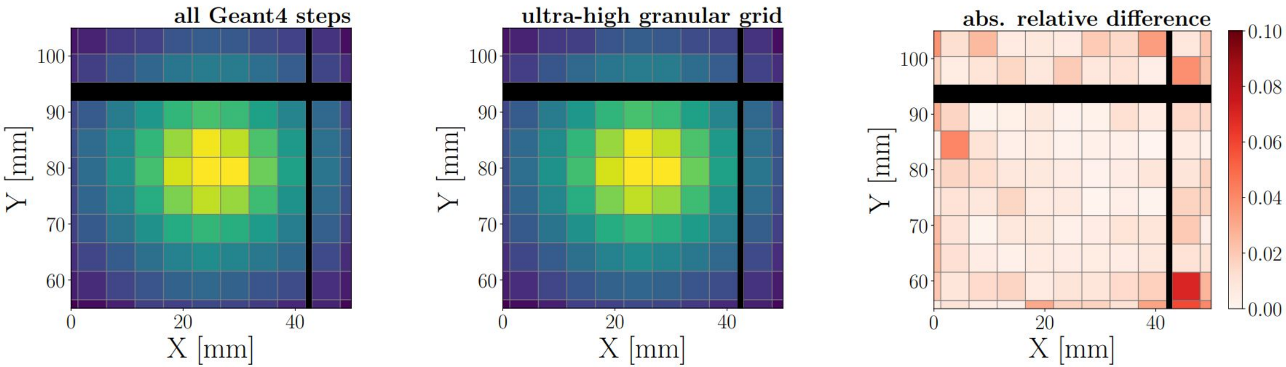
Point Cloud Representation of the EM Showers

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