# Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation with Diffusion Point Clouds

CHEP, Krakow, Poland
22.10.2024

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

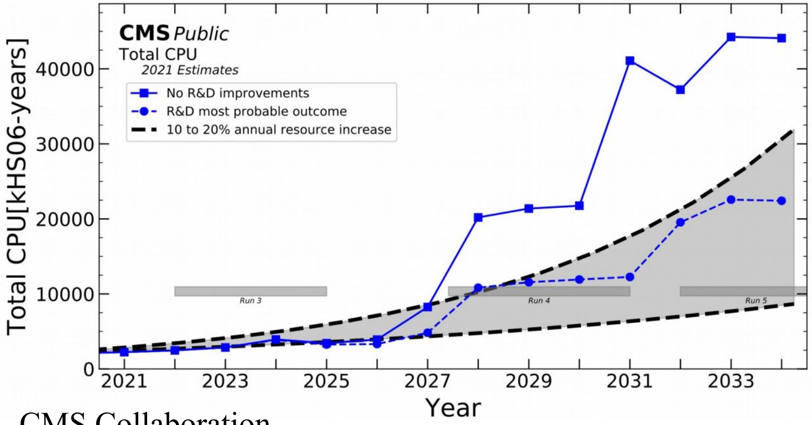
[1] Deutsches Elektronen-Synchrotron, DESY
[2] University of Hamburg, UHH
[3] CERN

*anatolii.korol@desy.de

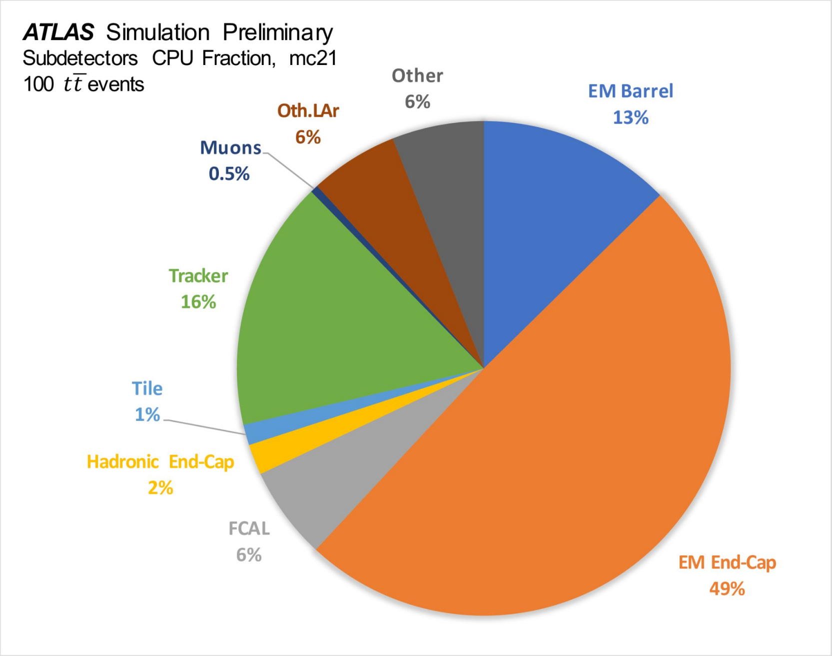
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE
HELMHOLTZ

# Introduction

## Time-consuming Simulations



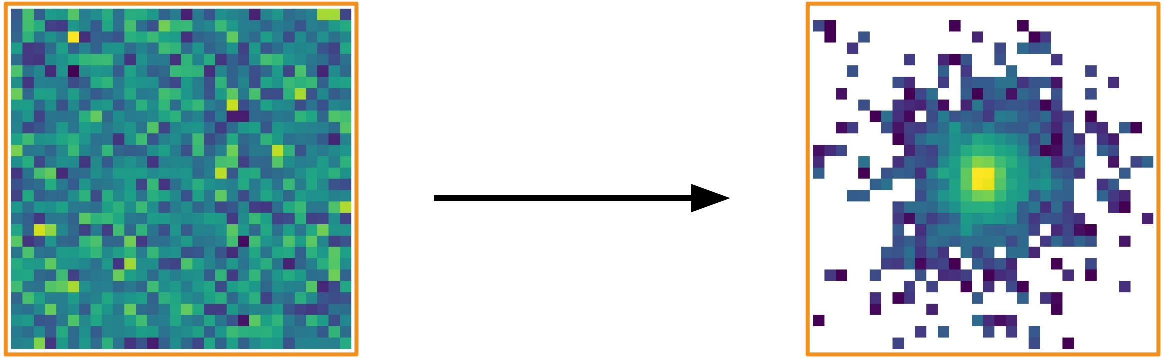

CMS Collaboration,
Offline and Computing Public Results (2021)
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults




**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

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

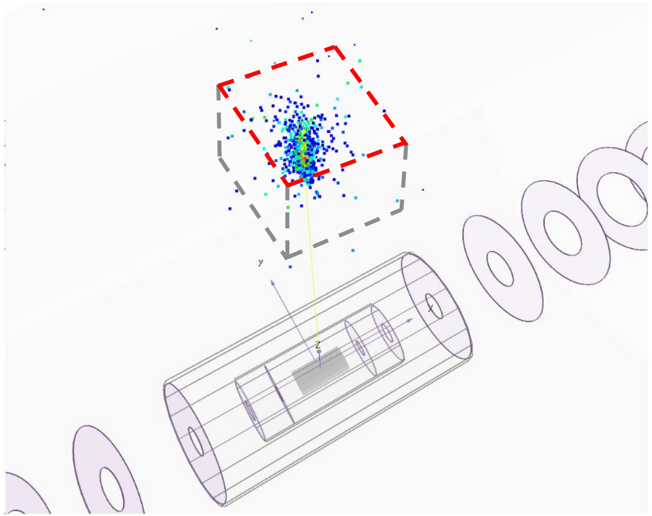


- There exist numerous generative models
  - Generative Adversarial Networks (GANs)
  - Flow-based models
  - Autoencoders (AE), e.g. BiB-AE
  - Diffusion Models (DMs)

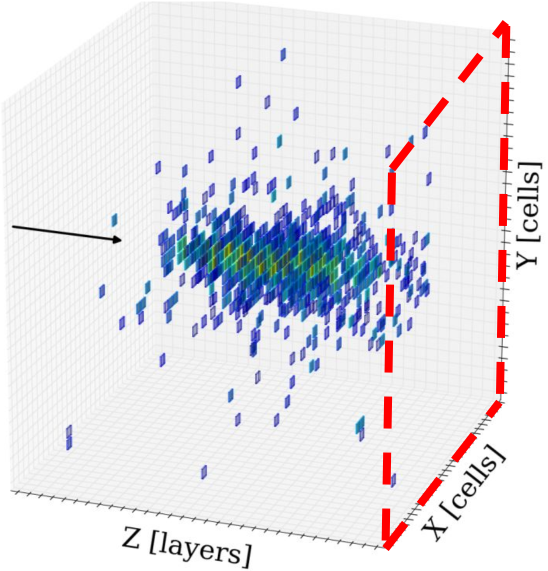
# Image Representation of the EM Showers

## ILD Detector

**A simulated 60 GeV photon shower in the ILD detector**



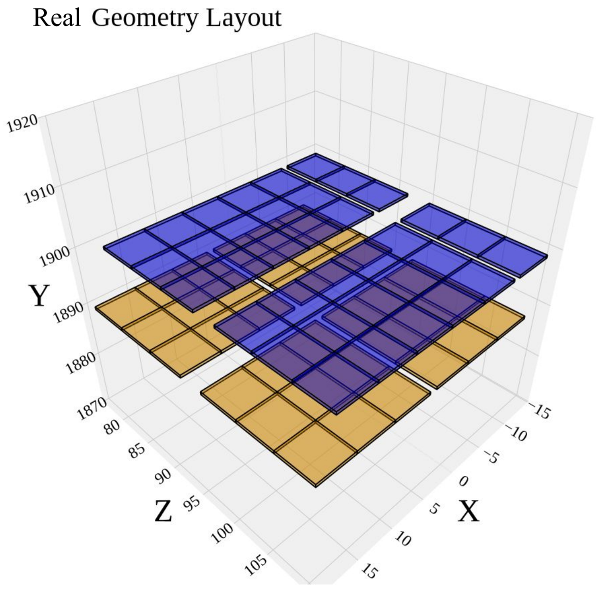
**Regular grid 30x30x30**

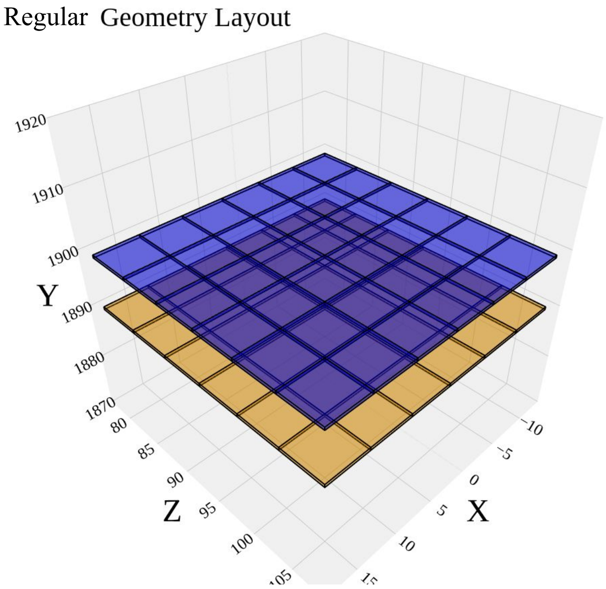



One to one mapping from detector geometry to a regular grid

# Problems with Image Representation of the EM Showers

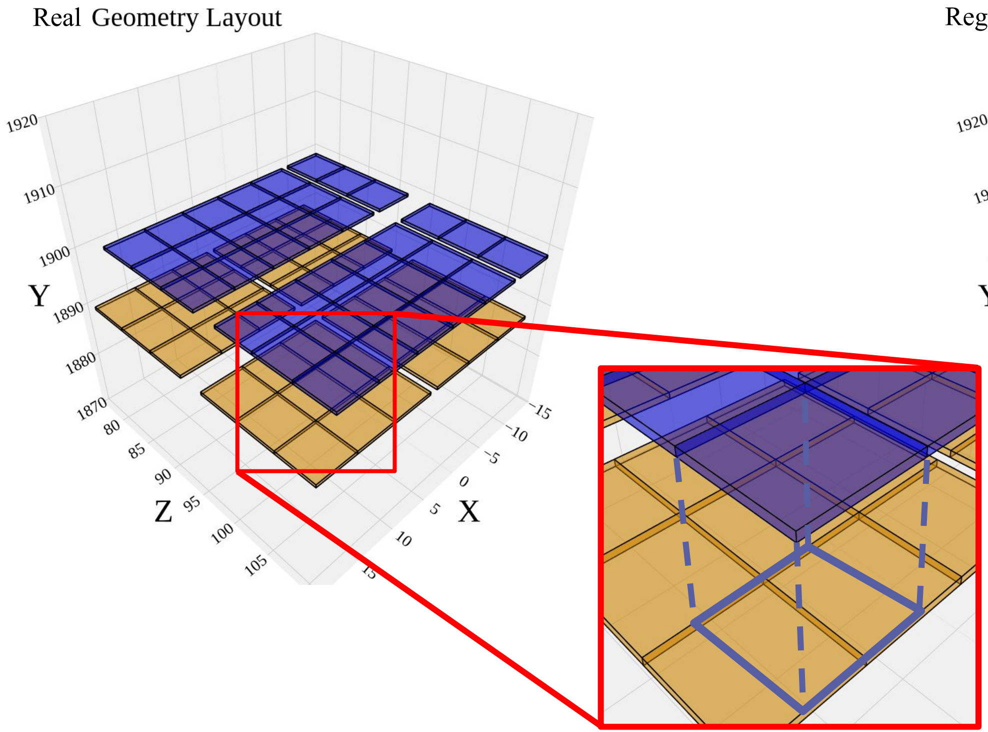
## ILD Detector, ECAL Layers Structure

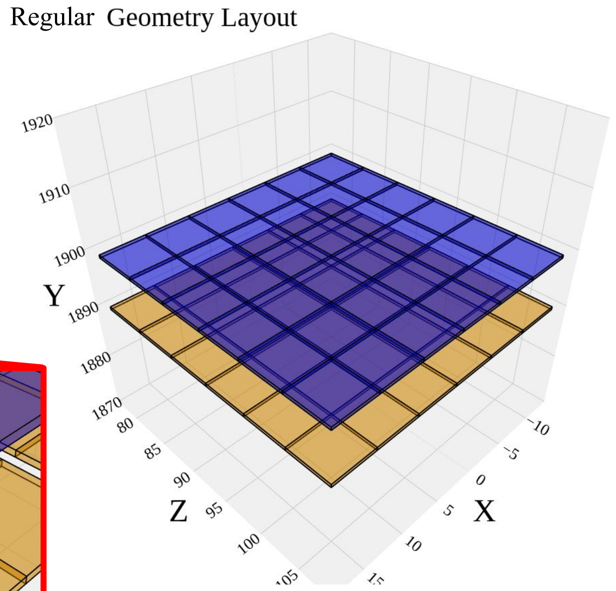


# Problems with Image Representation of the EM Showers

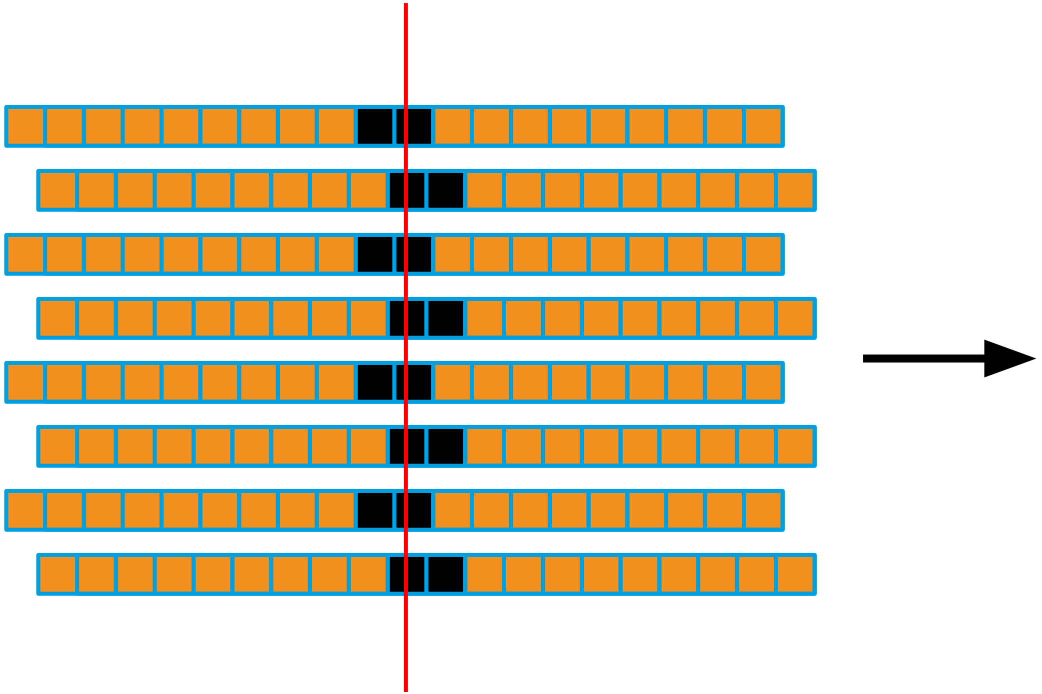
## ILD Detector, ECAL Layers Structure

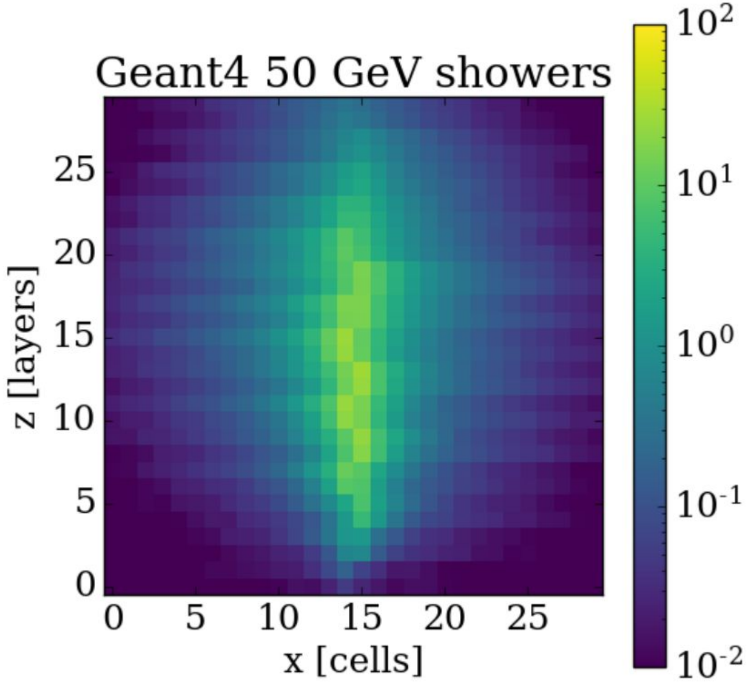


# Problems with Image Representation of the EM Showers

## ILD Detector, ECAL Layers Structure, Staggering Effect





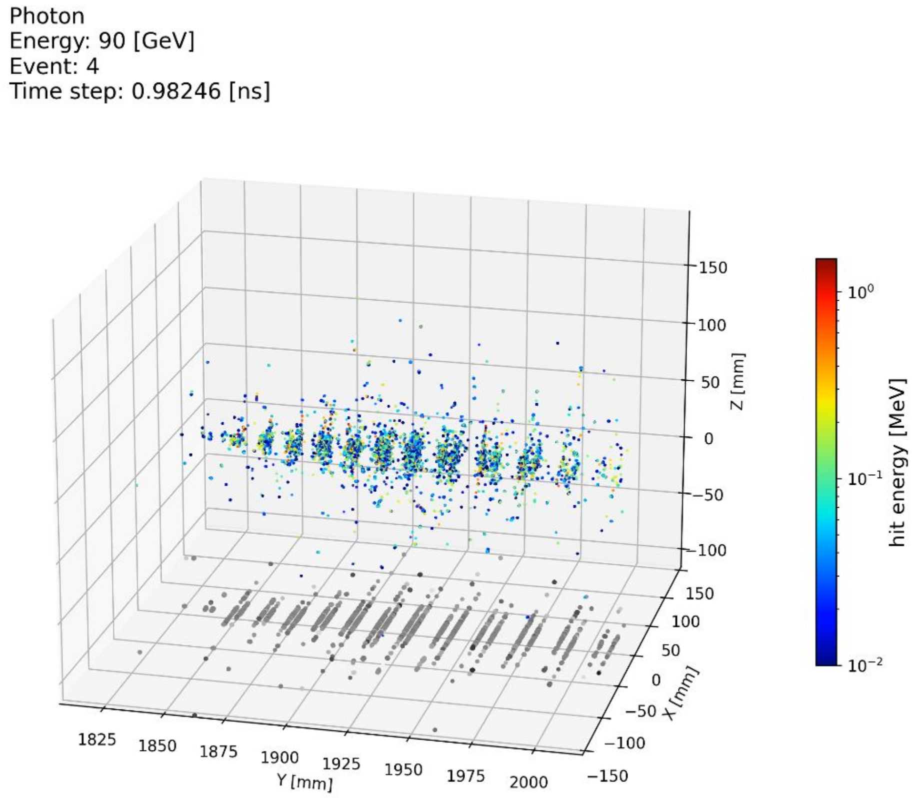

Models have to learn not only EM shower properties, but also geometry “artifacts”, like **staggering effect**

# Point Cloud Representation of the EM Showers

## GEANT4 Steps

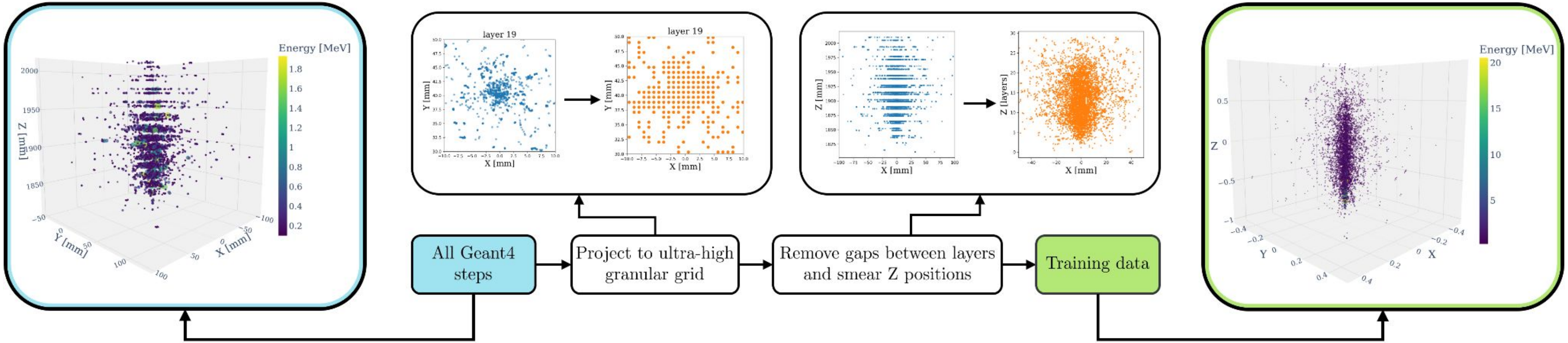
A way to overcome potential issues from irregular (realistic) cell geometry would be to use much higher granularity/resolution

- 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

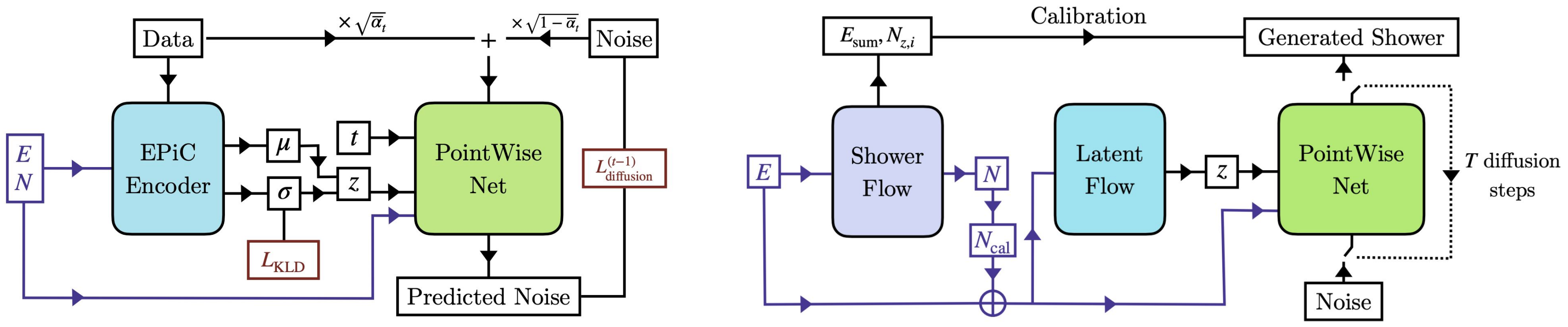



Number of points reduced to ~6k per shower, high enough resolution to move the shower in different place without harming physical properties of the shower

# Point Cloud + Diffusion Model

## CaloClouds, Model Overview

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



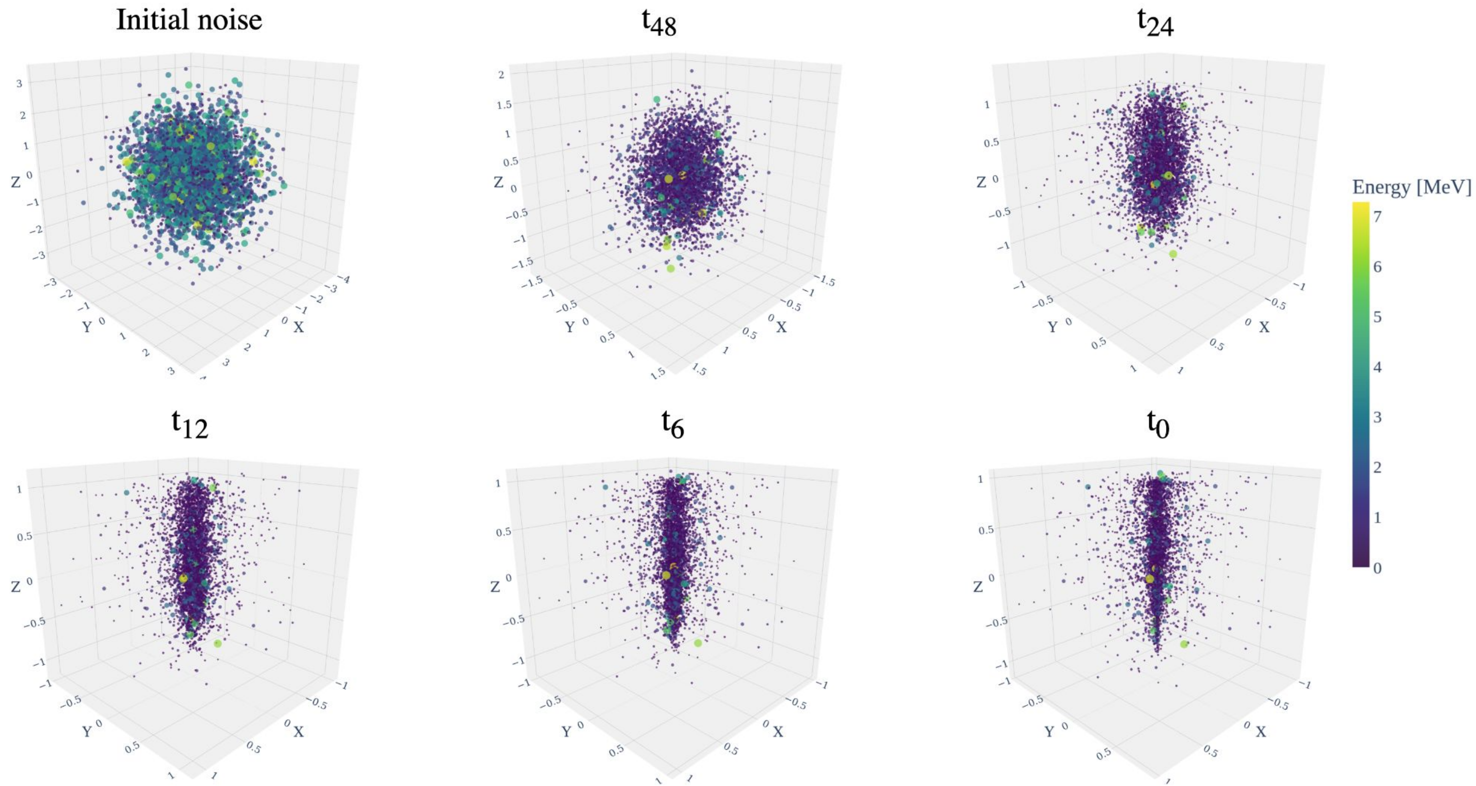

(a) Training at random time step $t$

(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

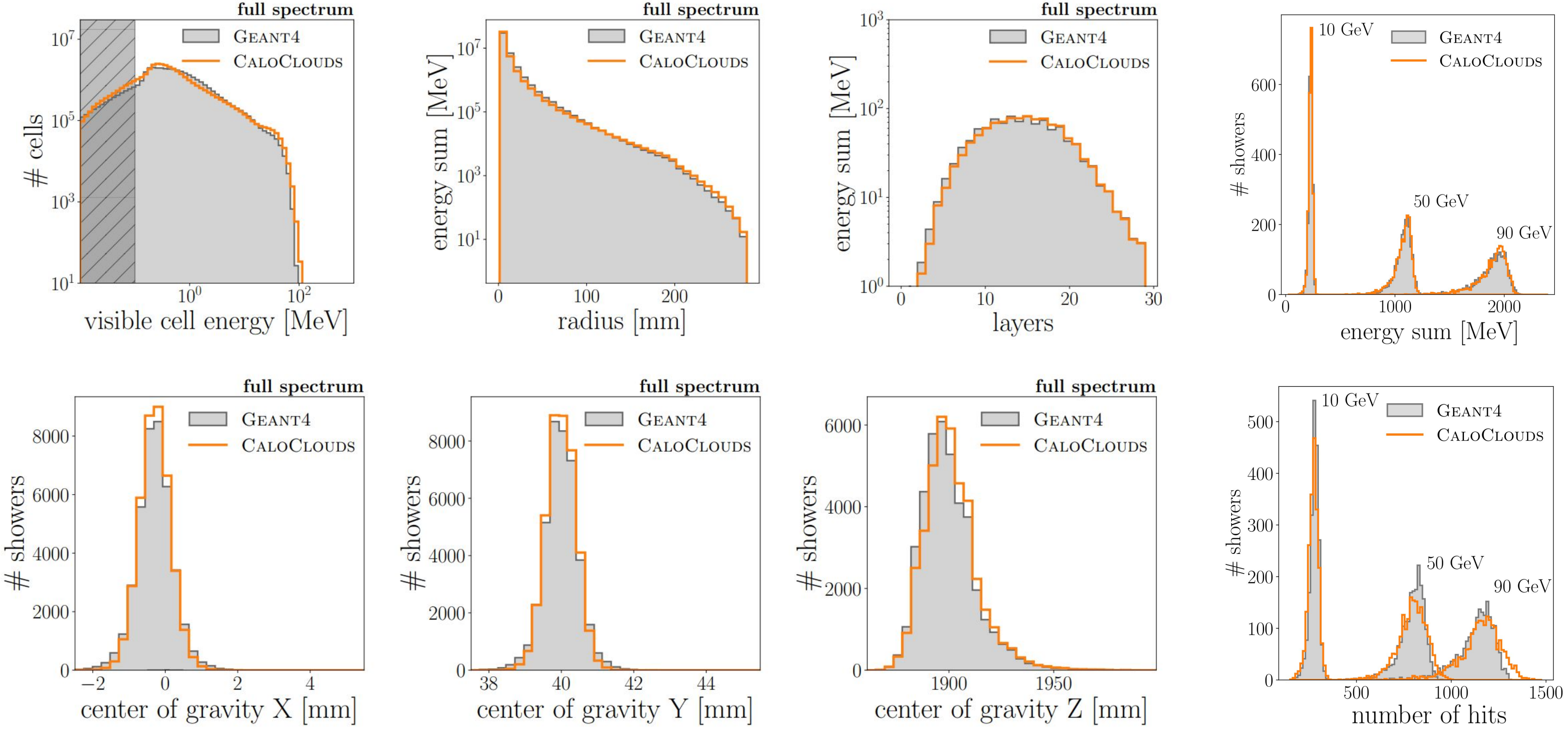
## Reverse Diffusion Process



# Point Cloud + Diffusion Model

## Physics Observables

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

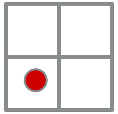


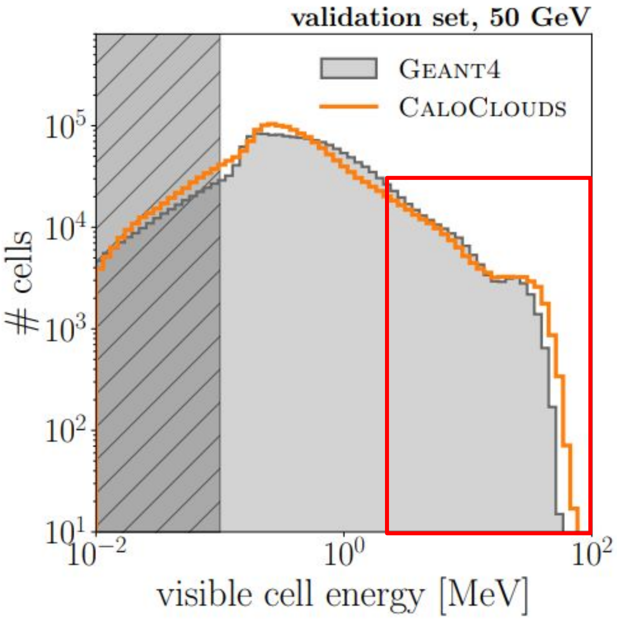
# Point Cloud + Diffusion Model

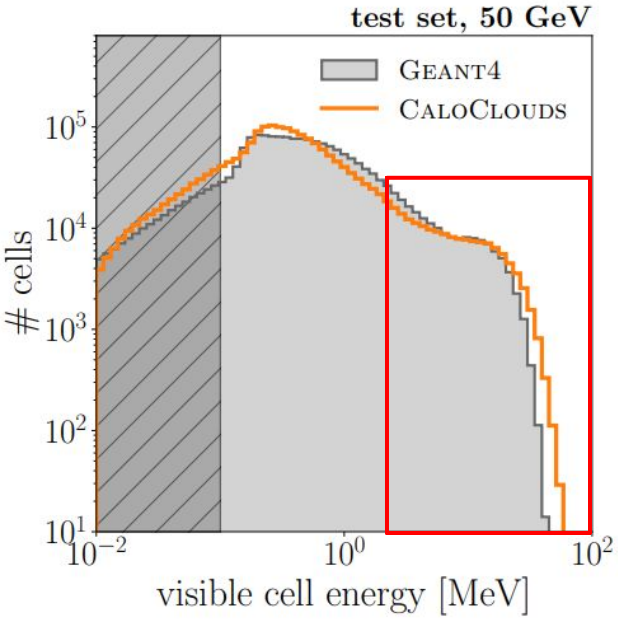
## Physics Observables For Different Positions

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

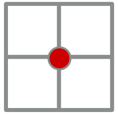
validation incident point pos.






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

## Speedup, CaloClouds I

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

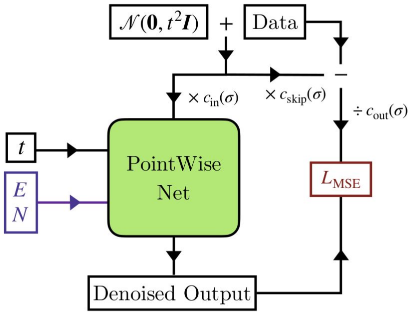
| Hardware | Simulator | Time / Shower [ms] | Speed-up |
|---|---|---|---|
| CPU | GEANT4 | $4082 \pm 170$ | $\times 1$ |
| | CALOCLOUDS | $3509 \pm 220$ | $\times 1.2$ |
| GPU | CALOCLOUDS | $38 \pm 3$ | $\times 107$ |

Not impressive inference time

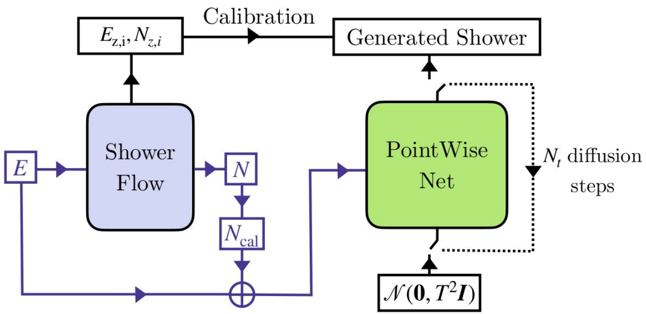
# Point Cloud + Diffusion Model

## CaloClouds II, Model Overview

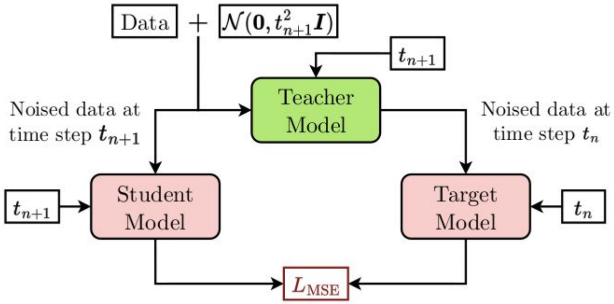
**CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation**, Buhmann, AK, et al. 2023, **arXiv:2309.05704**




(a) Training




(b) Sampling



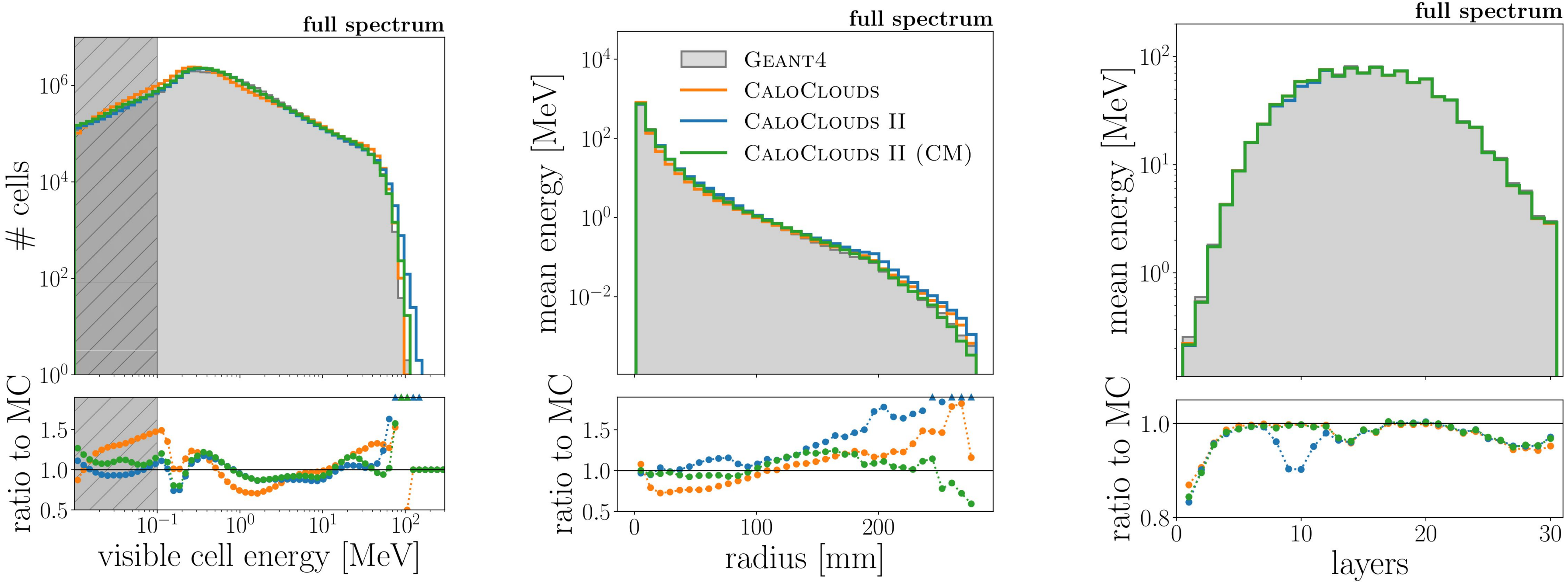

Consistency Distillation

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

# Point Cloud + Diffusion Model

## Physics Observables

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



# Point Cloud + Diffusion Model

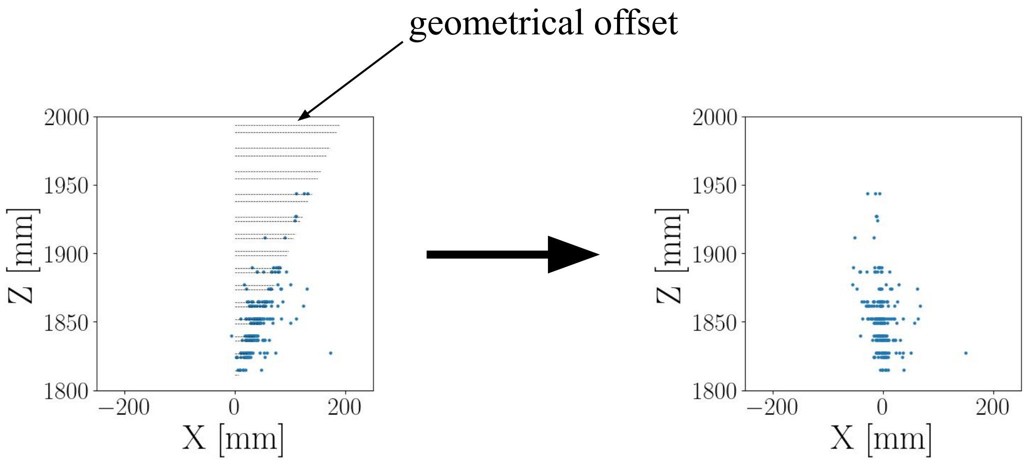
## Speedup, CaloClouds II

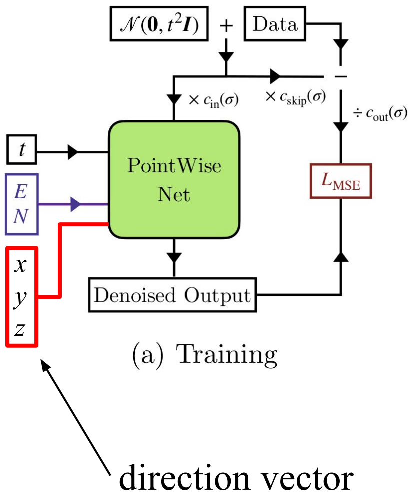
**CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation**, Buhmann, AK, et al. 2023, **arXiv:2309.05704**

| 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 |

# Point Cloud + Diffusion Model

## CaloClouds III, Adding Angular Conditioning



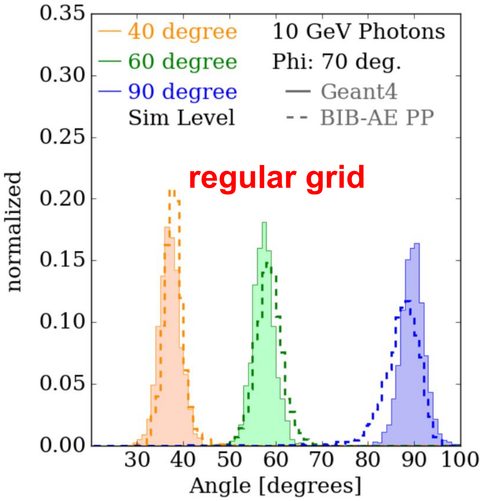




(a) Training

Additional data preprocessing step, another conditioning vector

# Point Cloud + Diffusion Model

## Angle Reconstruction

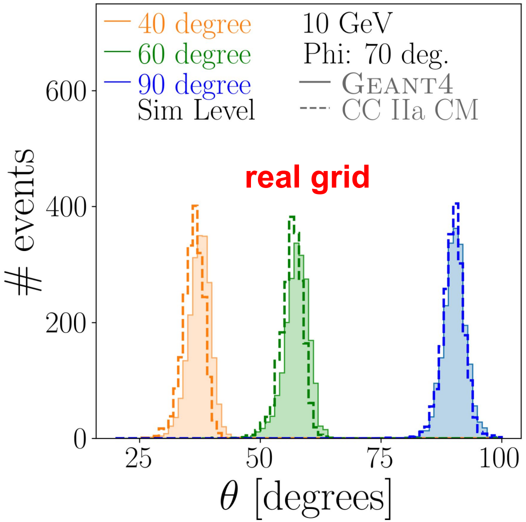



Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models

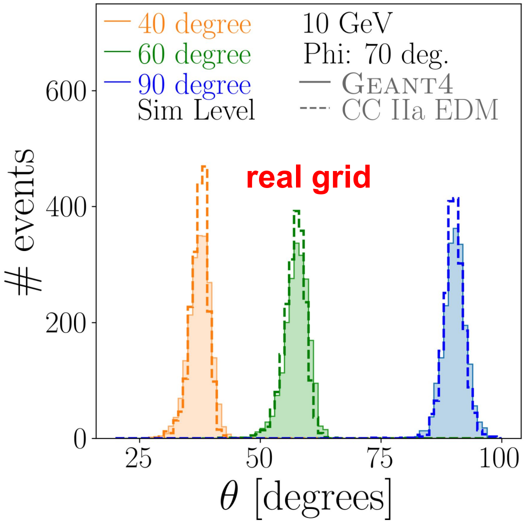
Peter McKeown (Hamburg U.), 2024

DOI: 10.3204/PUBDB-2024-01825

**1 step**

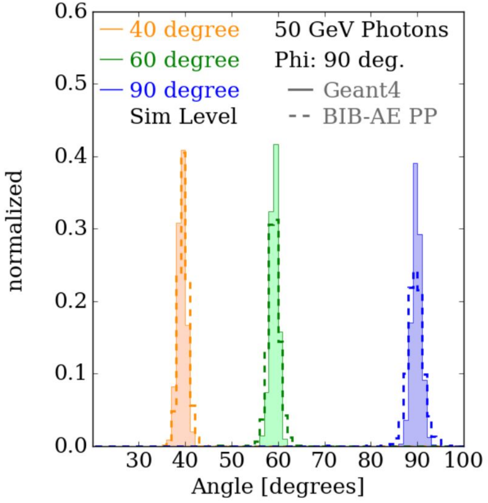



**90 steps**



# Point Cloud + Diffusion Model

## Angle Reconstruction

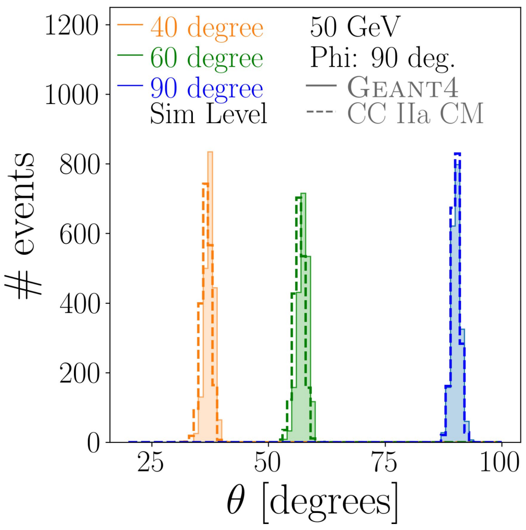



Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models

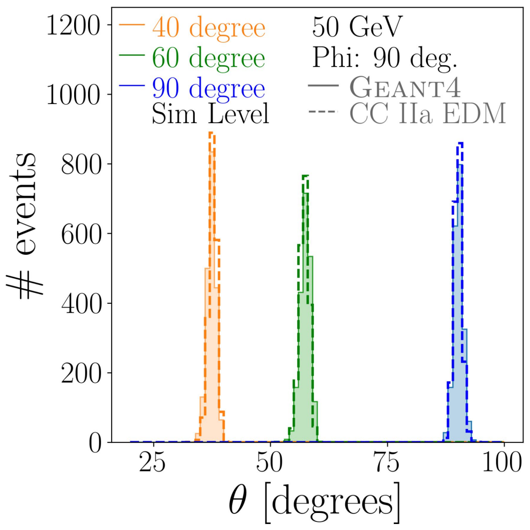
Peter McKeown (Hamburg U.), 2024

DOI: 10.3204/PUBDB-2024-01825

**1 step**

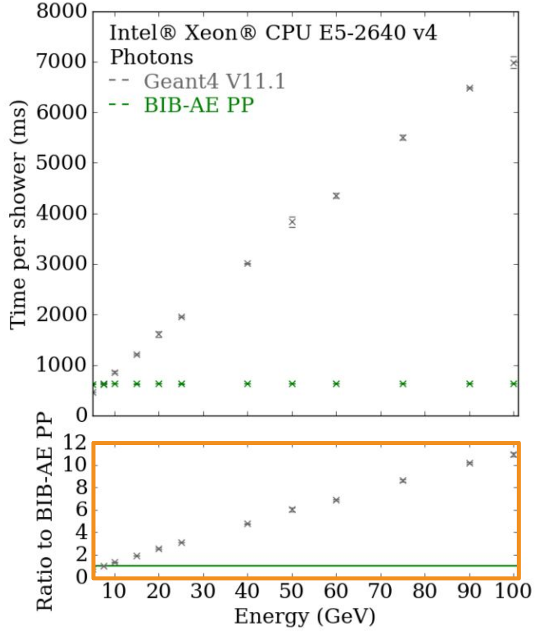


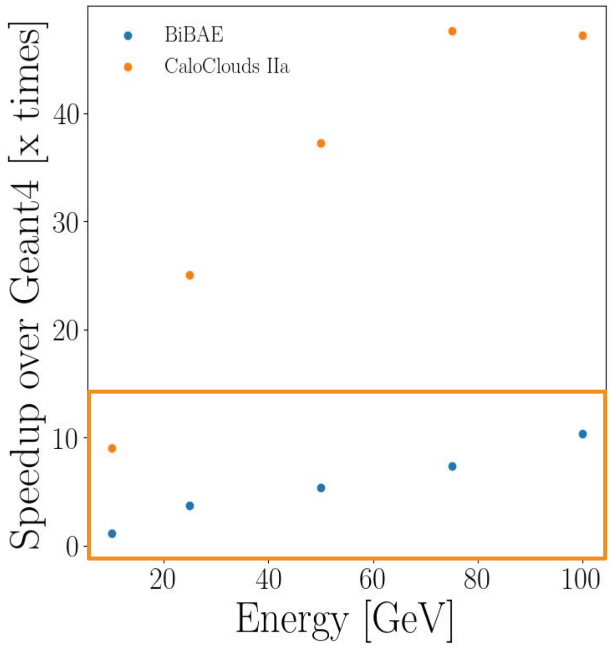

**90 steps**



# Point Cloud + Diffusion Model

## Speedup, CaloClouds III vs BiBAE




Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models

Peter McKeown (Hamburg U.), 2024

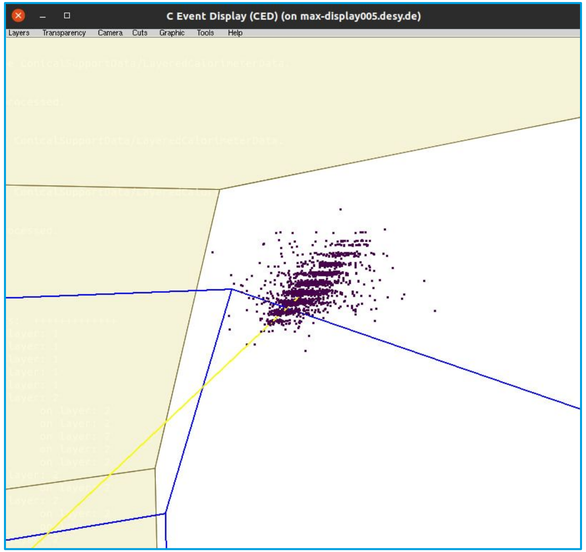
DOI: 10.3204/PUBDB-2024-01825

# Integration into Simulation Chain

## DDFastShowerML

PyTorch
Scripted fast sim model

DDFastShowerML
https://gitlab.desy.de/ilcsoft/ddfastshowerml



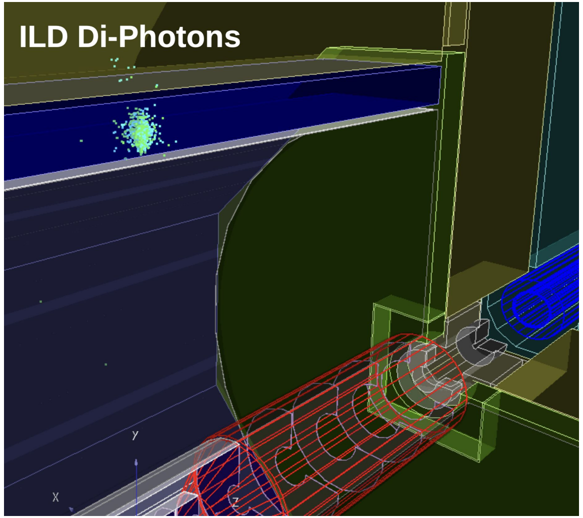

50 GeV photon shower generated with CaloClouds in the ILD ECAL

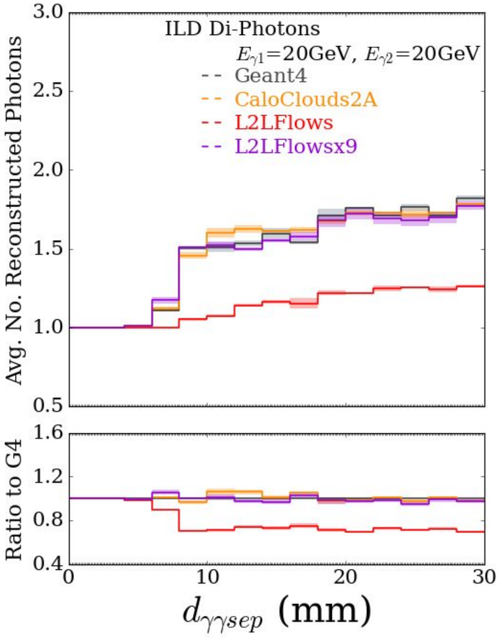
Easy to use library which can be adapted for all types of ML architectures in DD4hep

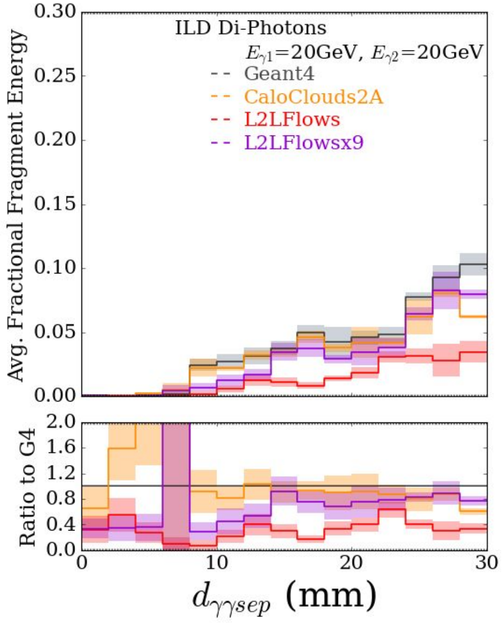
# Di-Photons Reconstruction Benchmark

## Two Photons Orthogonal to the Face of ILD ECAL

Courtesy P.McKeown




Di-Photon Reconstruction Benchmark provides a direct **physically relevant** quantification of model performance

# Summary

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

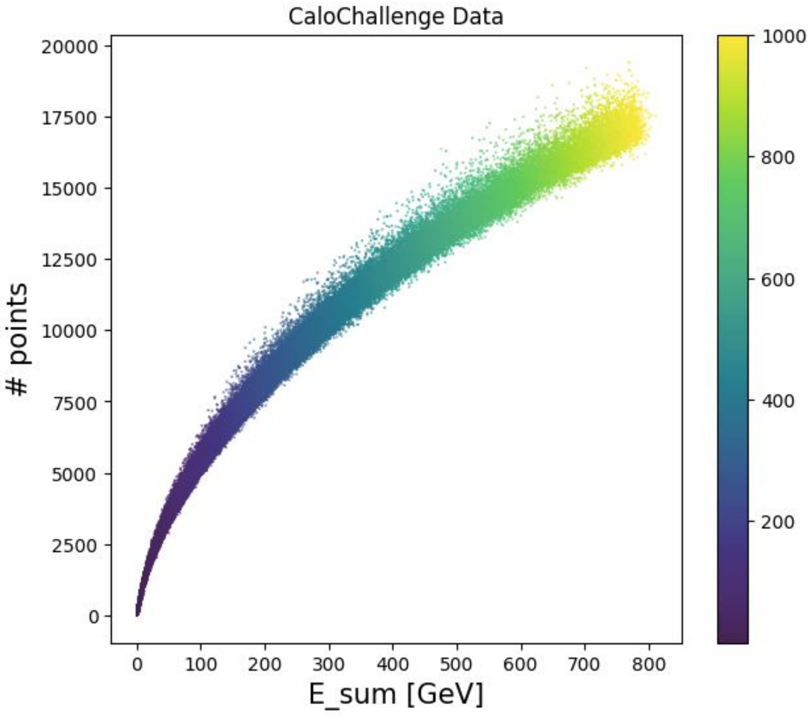
**CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation**, Buhmann, AK, et al. 2023, **arXiv:2309.05704**

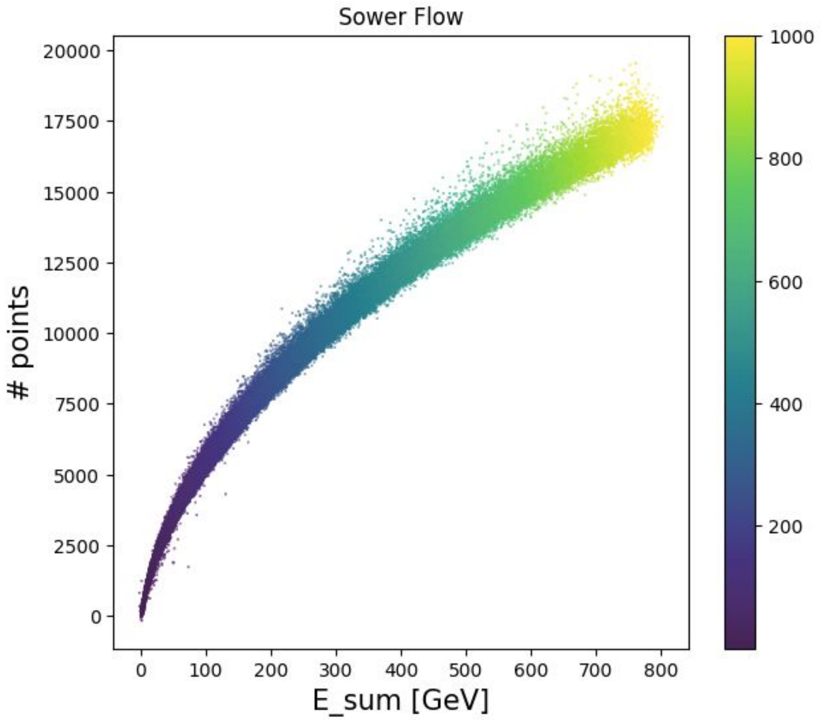
- Investigated new generative model architecture for generating EM showers in highly-granular calorimeters
- The fidelity of CaloClouds' physical properties is competitive with BiBAE, while offering ~x6 faster inference
- Model integrated into existing software ecosystem
- Next steps: more in depth study of physics benchmarks

# BACKUP SLIDES

# Shower Flow

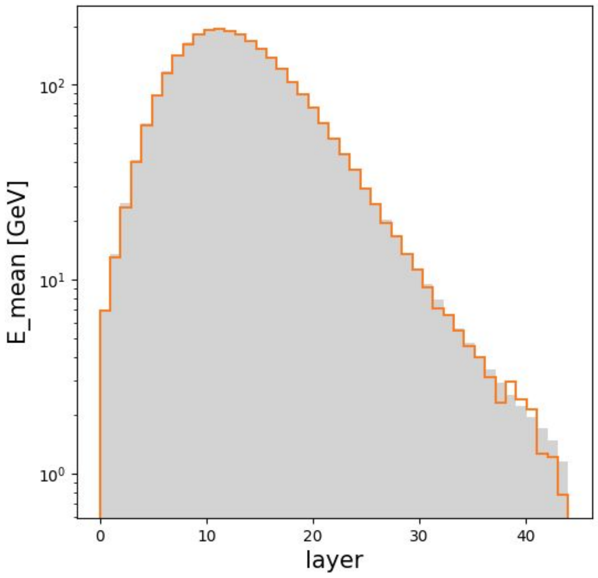
## Results

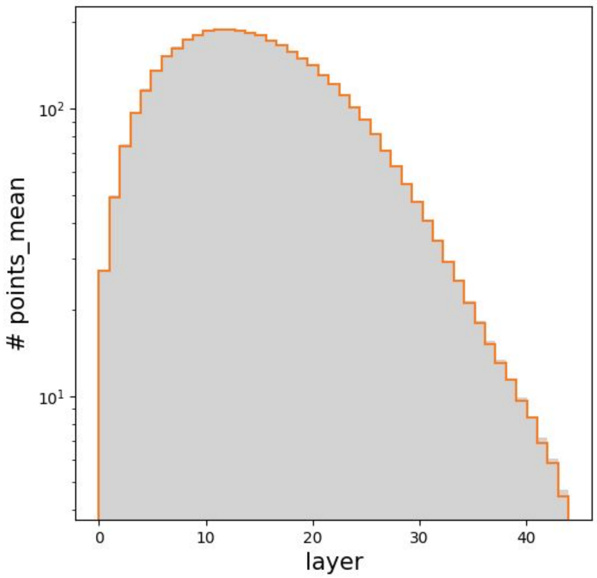


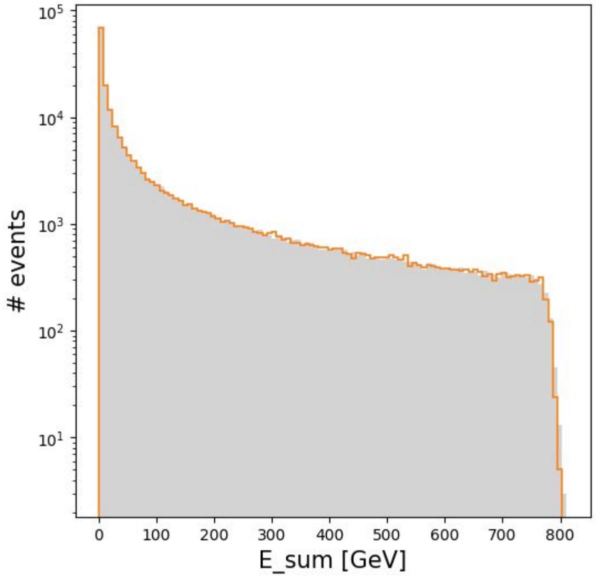



# Shower Flow

## Results

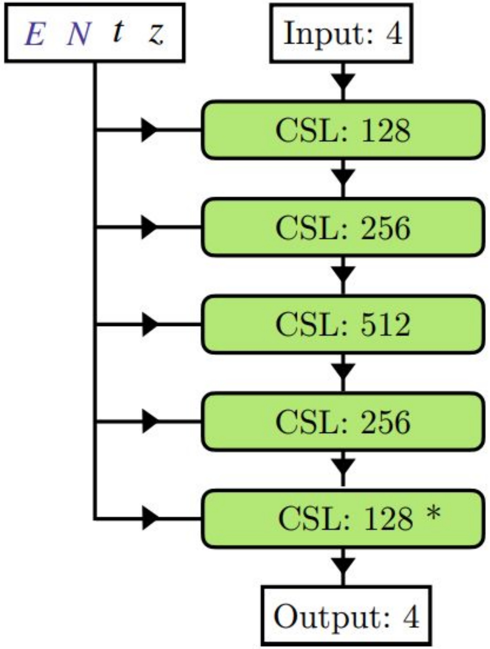


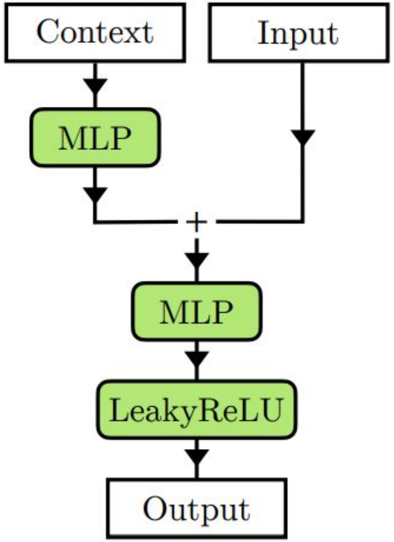



# Point Cloud + Diffusion Model

## PointWise Net



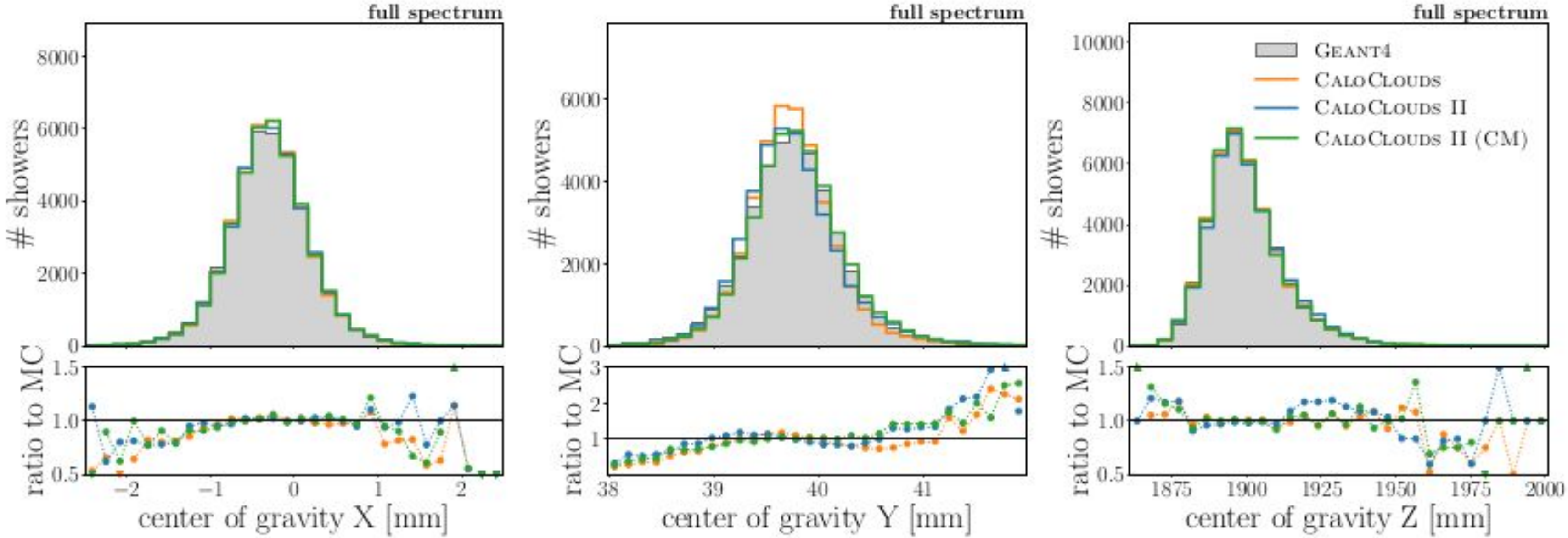

(a) PointWise Net




(b) ConcatSquash Layer (CSL)

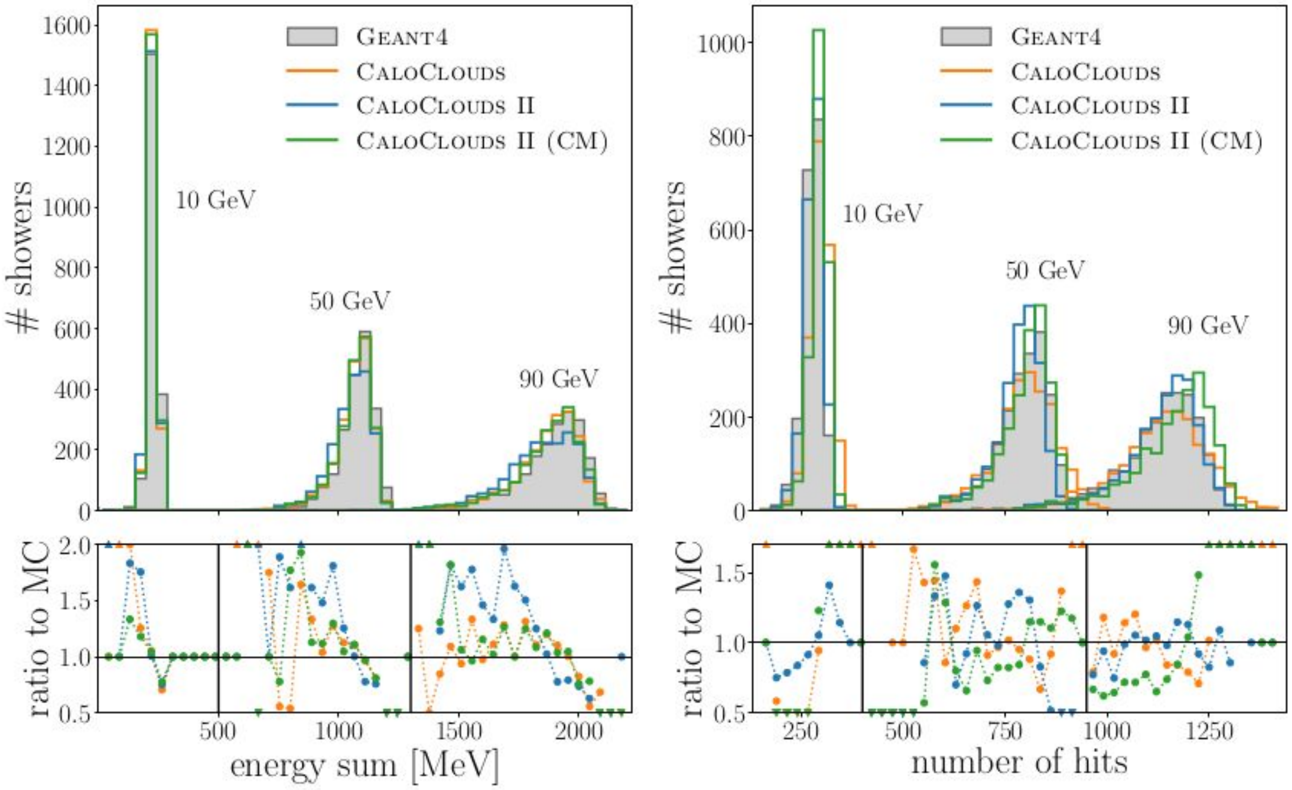
# Point Cloud + Diffusion Model

## Results, Position of the Center of Gravity



# Point Cloud + Diffusion Model

## Results, Visible Energy and the Number of Hits

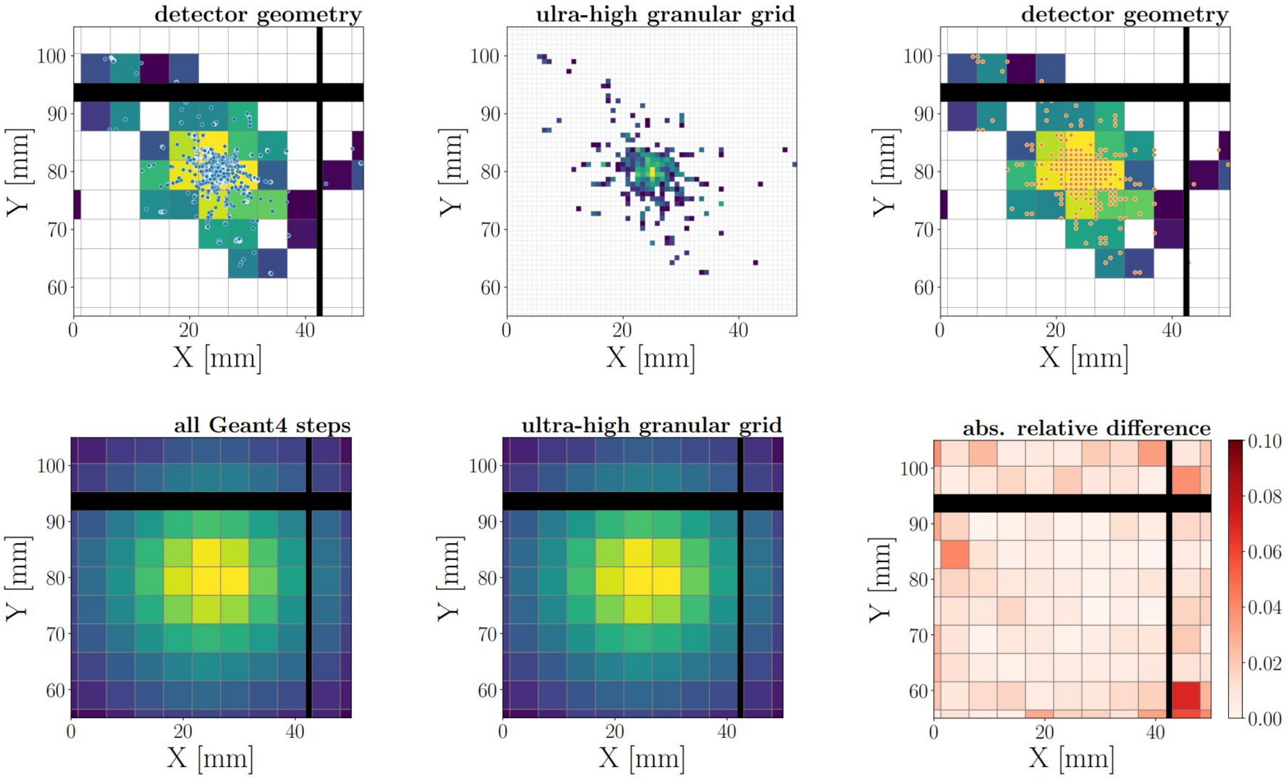


# Point Cloud Representation of the EM Showers

## 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