Generative Modeling with Diffusion Neural Networks for Fast Simulation of Electromagnetic Showers in the International Large Detector

DPG Spring Meeting 2023

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
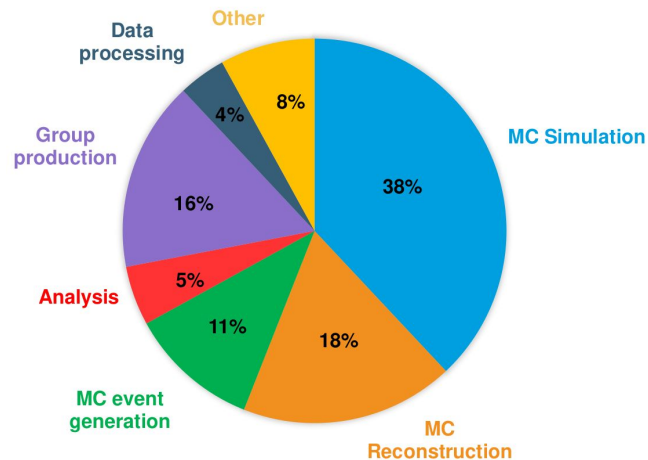
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE
HELMHOLTZ

Problem Definition
Time-consuming Simulations

The most computationally expensive step in the simulation pipeline of a typical HEP experiment is **(MC Simulation)** the detailed modeling of the full complexity of physics processes that govern the motion and evolution of particle showers inside calorimeters.

WALL CLOCK CONSUMPTION PER WORKFLOW

- Other: 8%
- MC Simulation: 38%
- MC Reconstruction: 18%
- MC event generation: 11%
- Analysis: 5%
- Group production: 16%
- Data processing: 4%

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

Problem Definition
The Strain on HEP Computing Resources

- Projected computing resources required far outstrip what will be available
  - E.g HL-LHC
- Future lepton colliders also benefit from much faster MC

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

This work attack the most intensive part of detector simulation – Calorimeter Simulation.
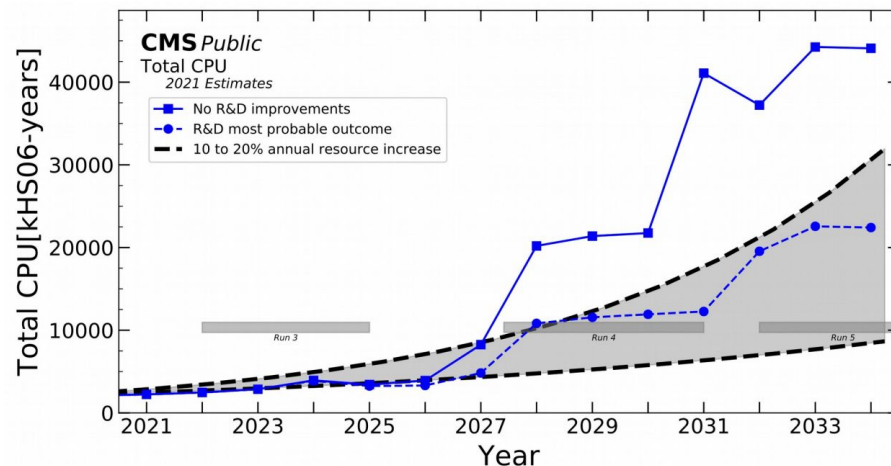
CMS Collaboration,
Offline and Computing Public Results (2021)
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults

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
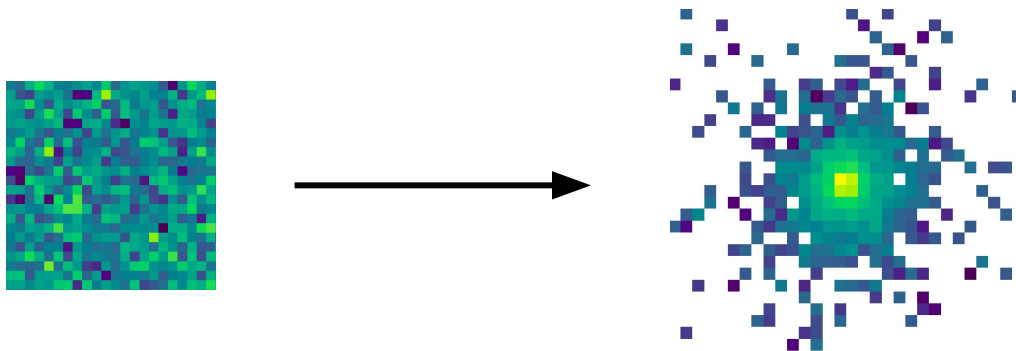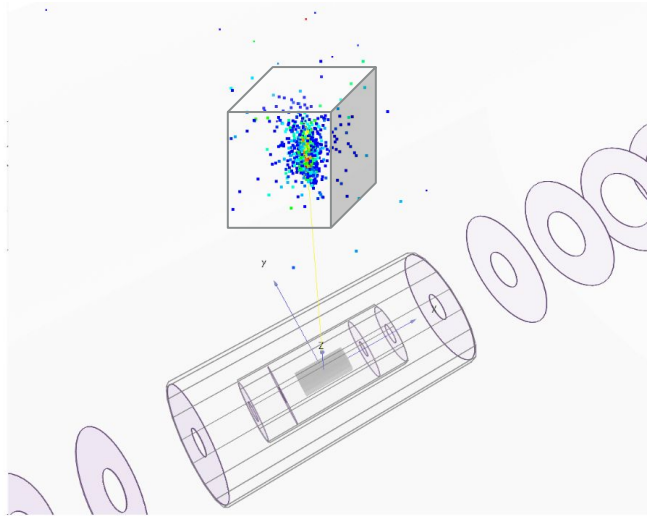

- There exist numerous generative models

  - Generative Adversarial Network (GANs)
  - Flow-based models

  - Variational Autoencoders (VAEs)
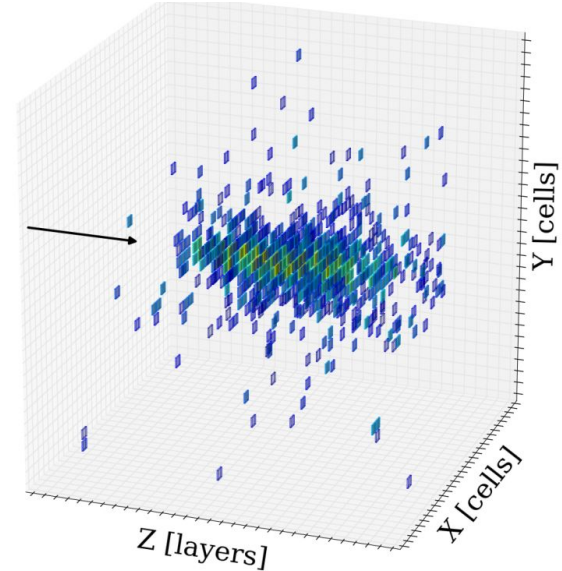 - Denoising Diffusion Probabilistic models (DDPs)

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.

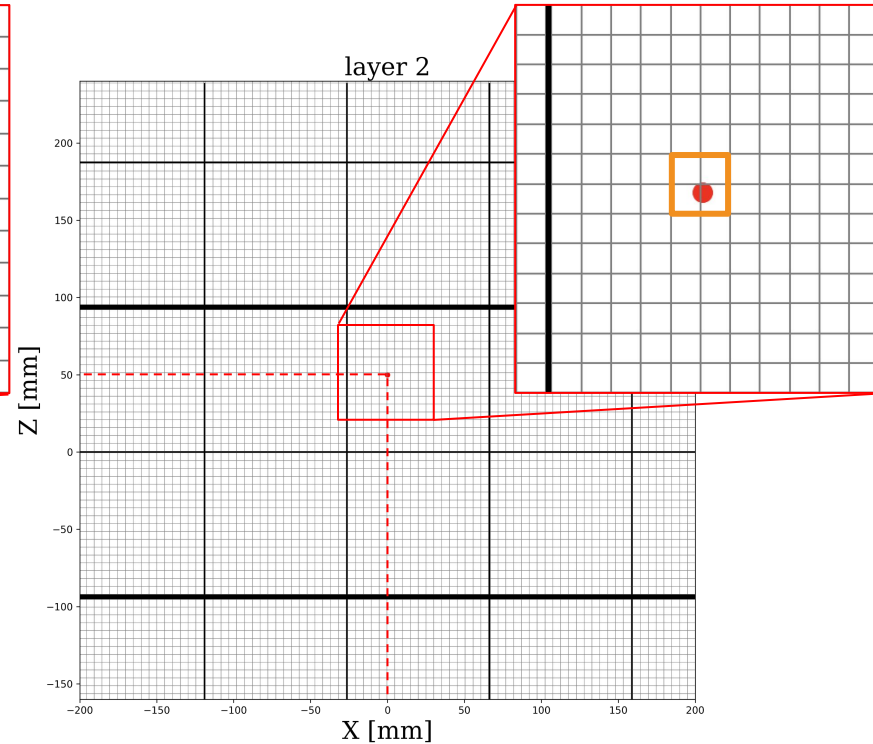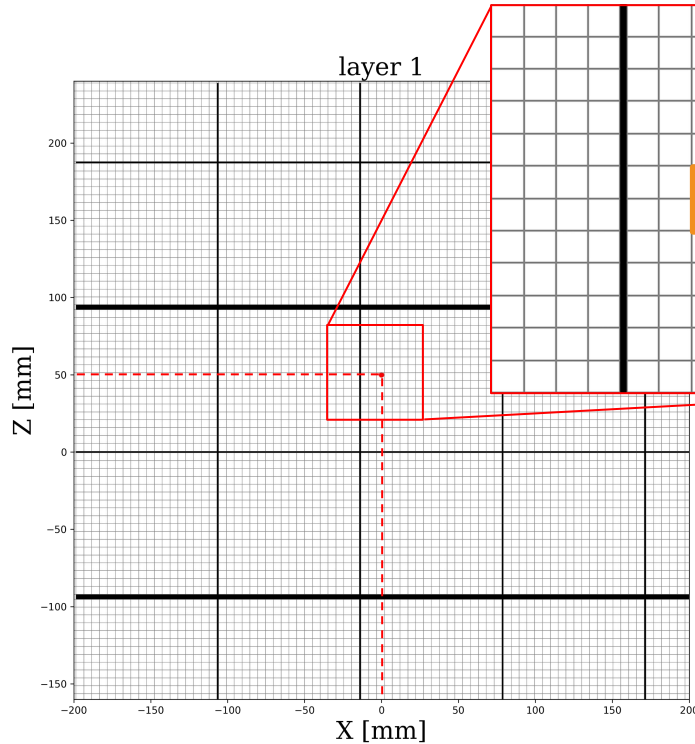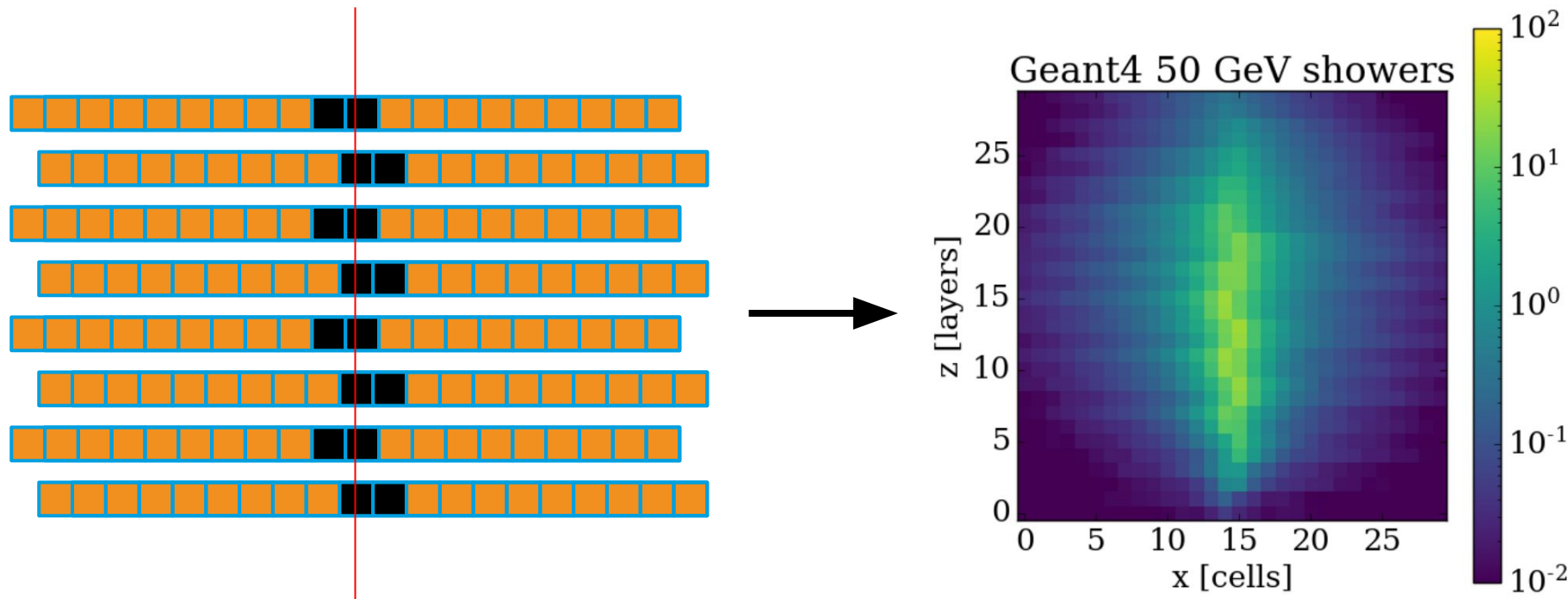# Image representation of the EM Showers
ILD Detector, ECAL Layers Structure
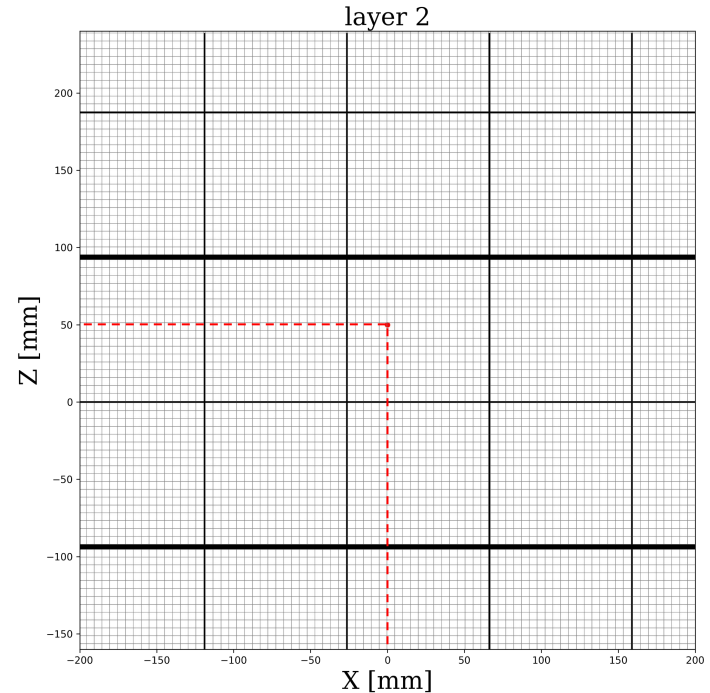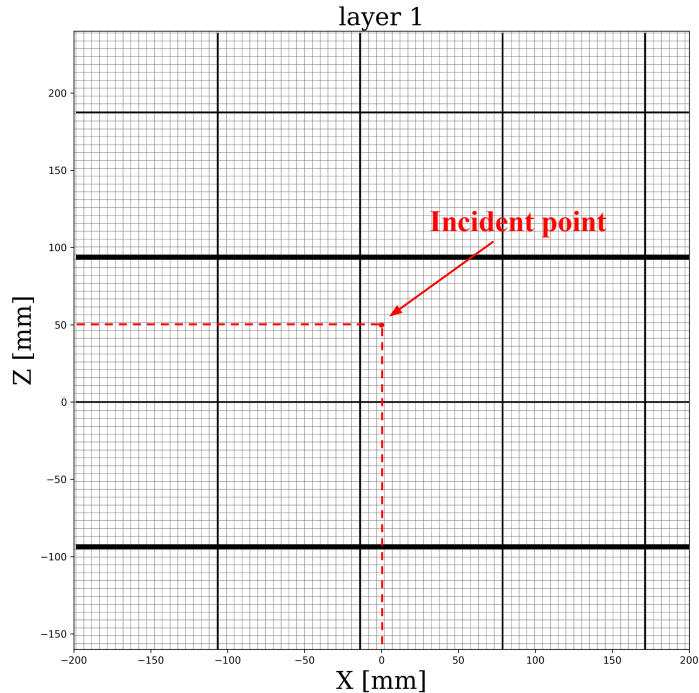
White squares represent **active cells**. Black lines are wafers, construction gaps, etc. (**not active material**).

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

Geant4 50 GeV showers

Models have to learn not only EM shower properties, but also geometry "artifacts", like **staggering effect.**
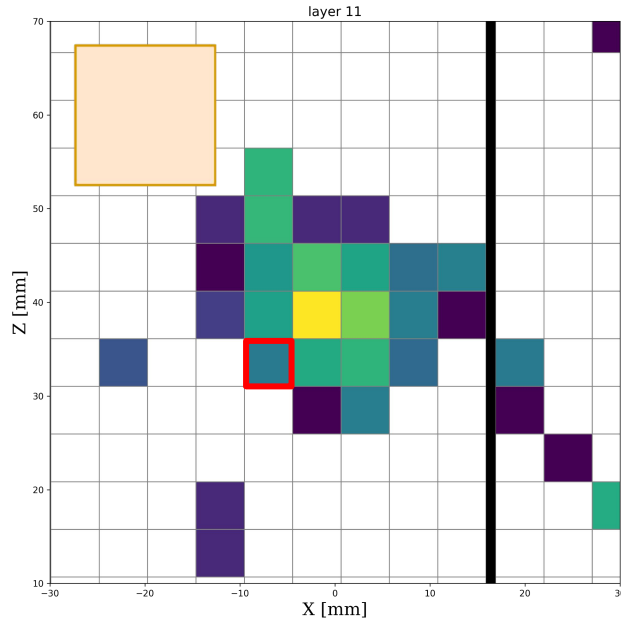
Point Clouds representation of the EM Showers

GEANT4 Steps

- All G4 interactions, highest possible resolution

- Detached from detector layer geometry
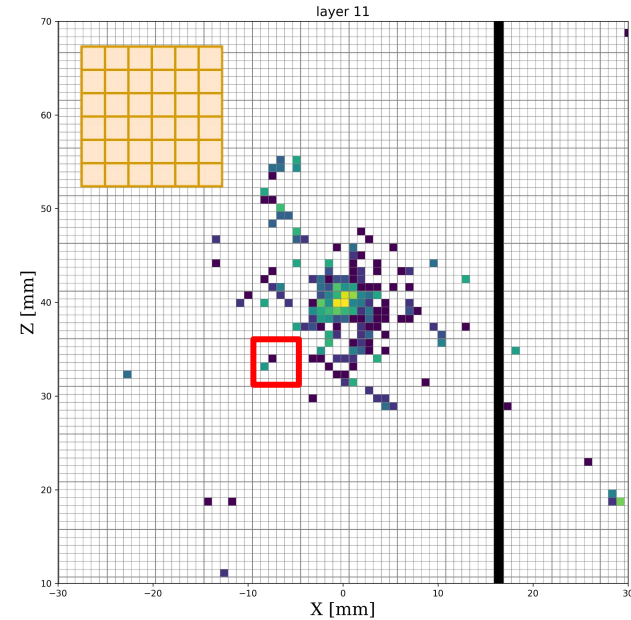
- Too many points to generate (~40k per shower)

Photon
Energy: 90 [GeV]
Event: 4
Time step: 0.98246 [ns]
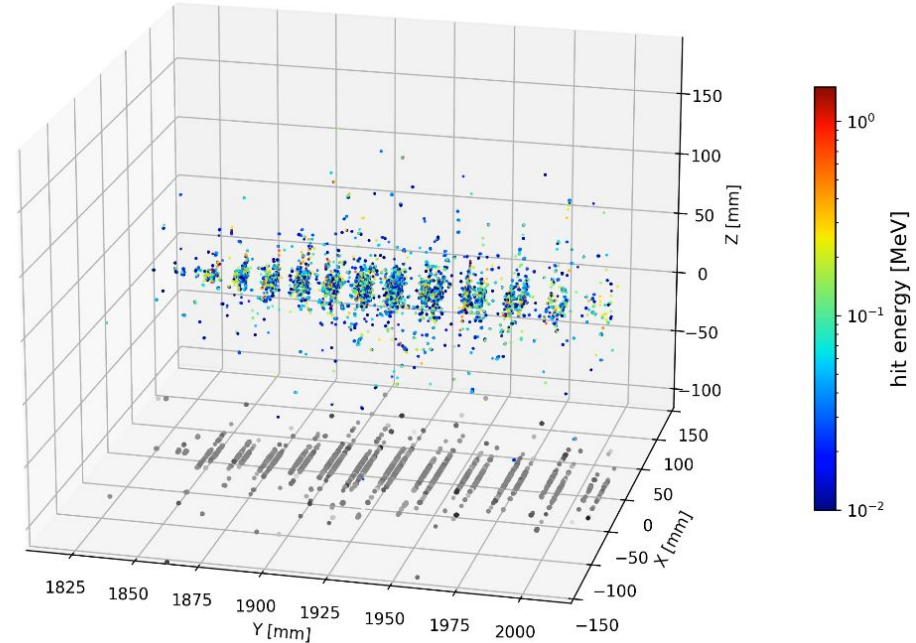
Point Clouds representation of the EM Showers

Artificially Increased Granularity, Cell Split 6x6
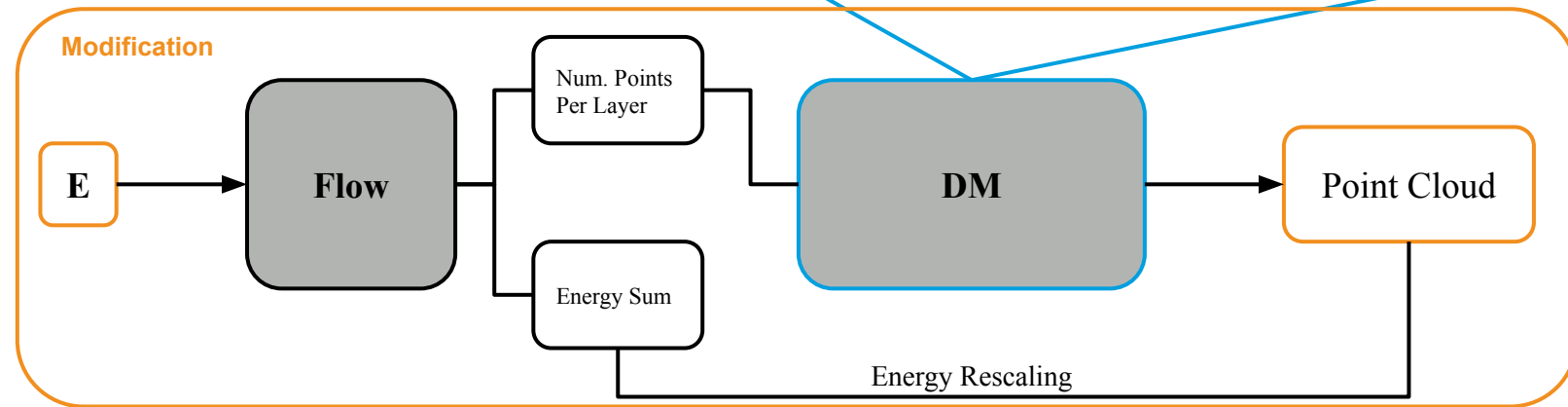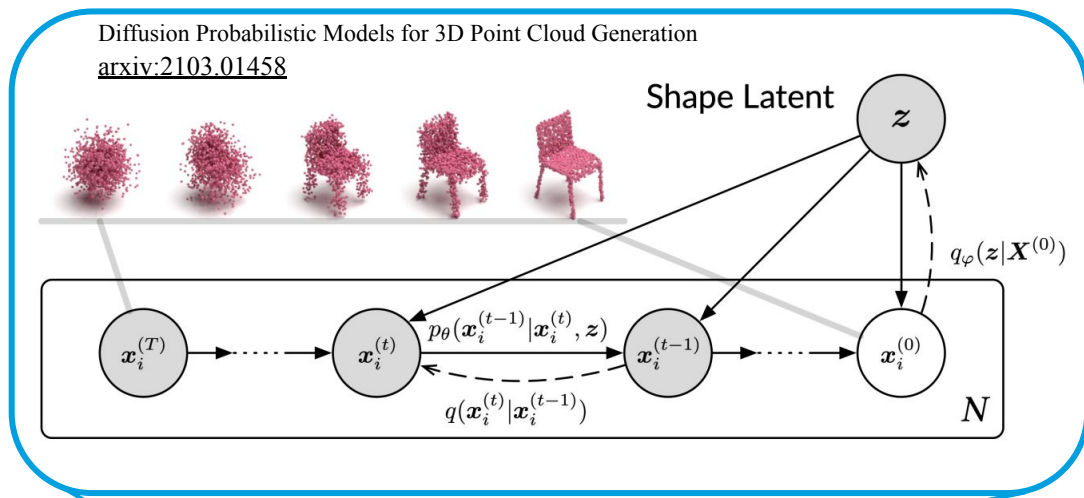

Original cell size ~ 5 mm per side

~ 0.83 mm per side

Number of points reduced to ~5k per shower, high enough resolution to move the shower in different place without harming physical properties of the shower.
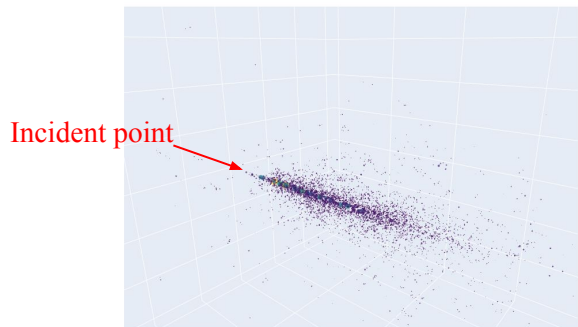
Diffusion Model

Model Overview

- GANs and VAEs convert noise from some simple distribution to a data sample

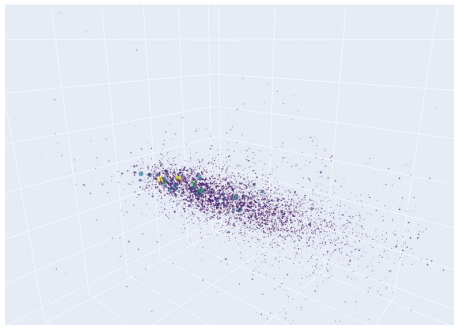- **DMs learn to gradually denoise data starting from noise**

Diffusion Probabilistic Models for 3D Point Cloud Generation
arxiv:2103.01458

Shape Latent z

$q_\varphi(z|X^{(0)})$

$p_\theta(x_i^{(t-1)}|x_i^{(t)}, z)$

$x_i^{(T)}$ ... $x_i^{(t)}$ $x_i^{(t-1)}$ ... $x_i^{(0)}$

$q(x_i^{(t)}|x_i^{(t-1)})$

N

Modification

E → **Flow** → Num. Points Per Layer / Energy Sum → **DM** → Point Cloud

Energy Rescaling

Point Clouds + Diffusion Model

Forward Diffusion Process

Step 0

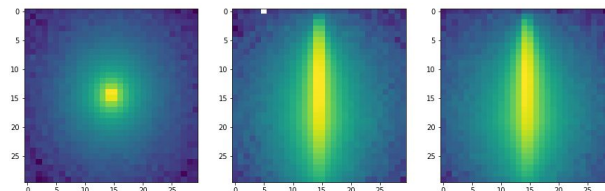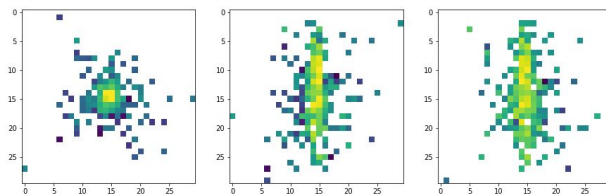Incident point

Step 2

Step 4
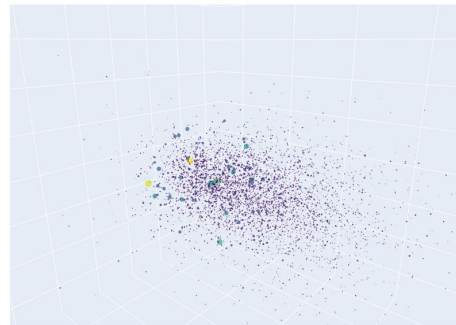
Step 10
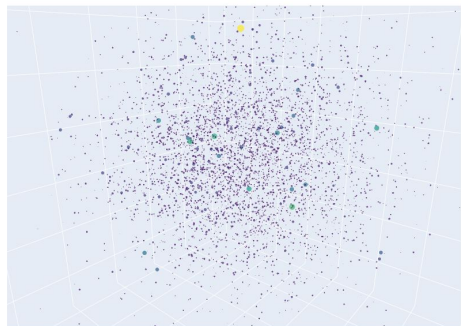
Step 25

Step 50

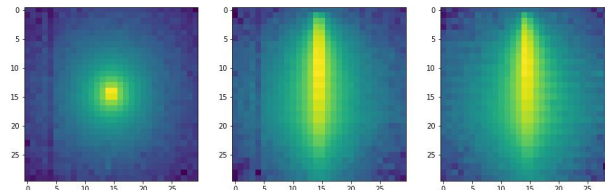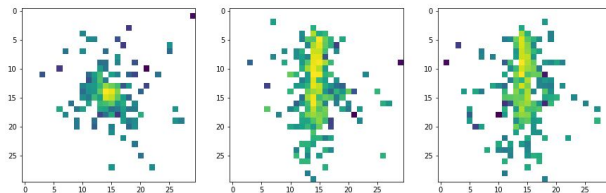# Point Clouds + Diffusion Model

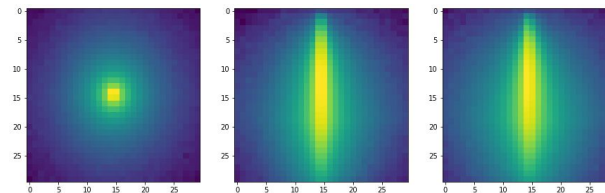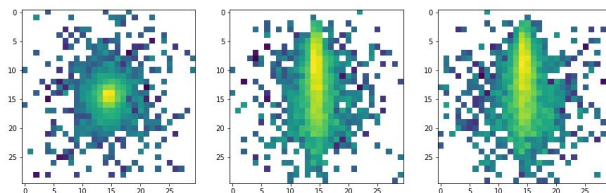## Results, Projected Images

Single Event   2k Events Overlay

DM 10 GeV

G4 10 GeV

DM 90 GeV

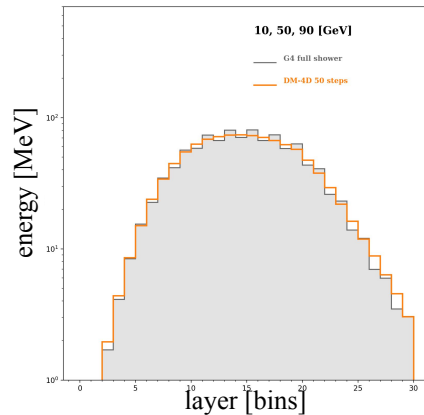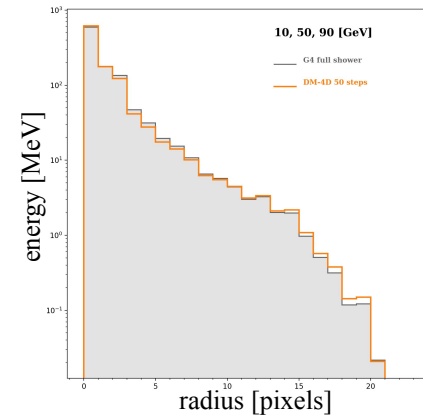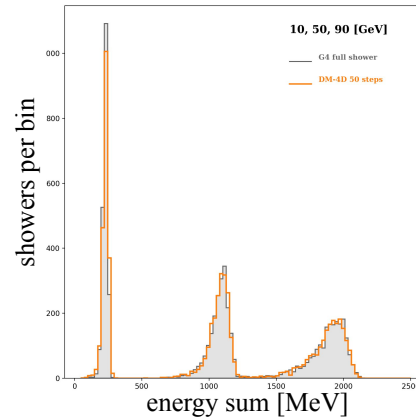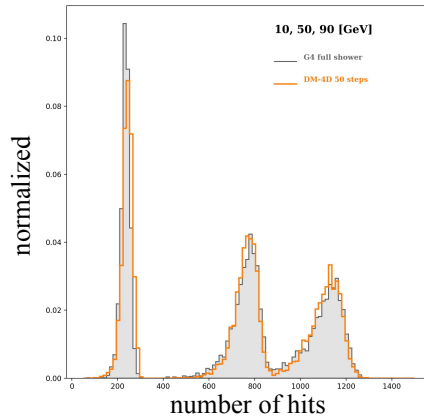**G4** 90 GeV
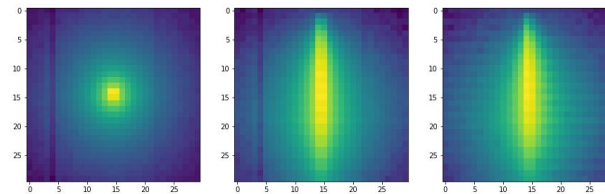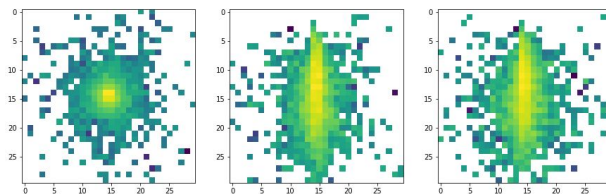
Point Clouds + Diffusion Model

Results, Distributions
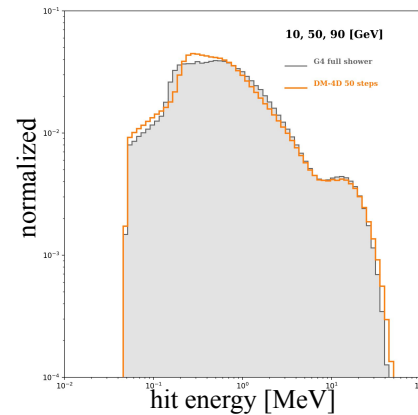
Summary

- 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**

BACKUP SLIDES

Point Clouds + Diffusion Model

Results, Potential Speed-up

	CPU 10-100 [GeV] | Speed-up	GPU 10-100 [GeV] | Speed-up
GEANT4	~4000 [ms] per shower	–
BiBAE	~400 [ms] per shower | ~**x10**	~1.5 [ms] per shower | ~**x2800**
DDP	~400 [ms] per shower | ~**x10**	~100 [ms] per shower | ~x40