CaloClouds: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation

Scientific Computing Seminar
DESY, 23.05.2025

E. Buhmann[2], H. Day-Hall[1], 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

Detailed Simulations is Crucial for Connecting Theory with Experimental Observations

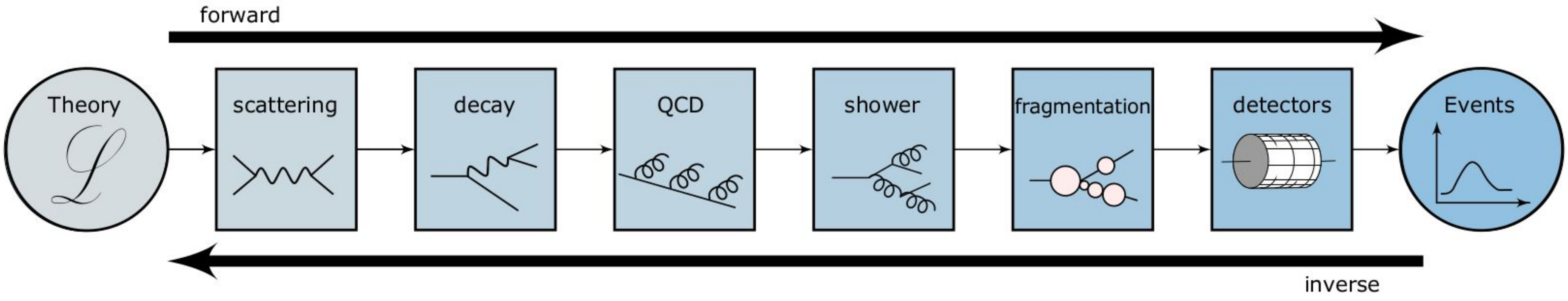


Figure from:
Machine learning and LHC event generation,
DOI: 10.21468/SciPostPhys.14.4.079

Detailed Simulations is Resource Intensive Task

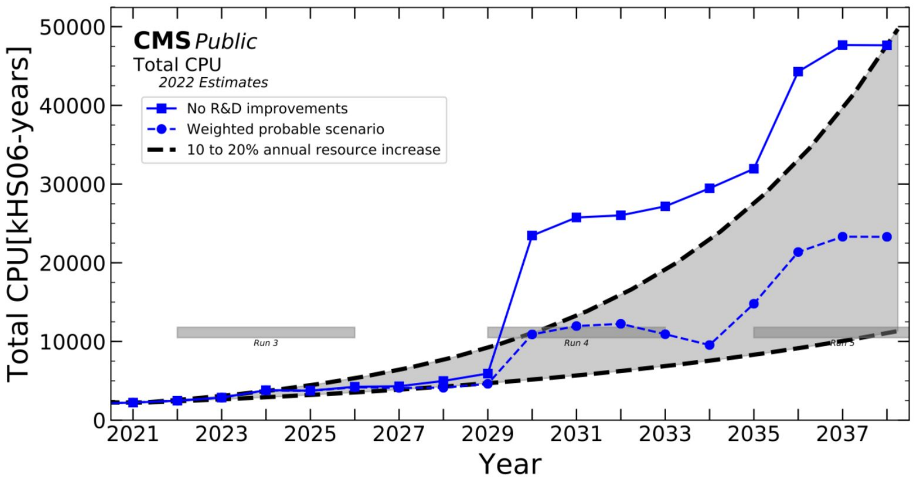


Figure from:
CMS Phase-2 Computing Model: Update Document,
http://cds.cern.ch/record/2815292/files/NOTE2022_008.pdf

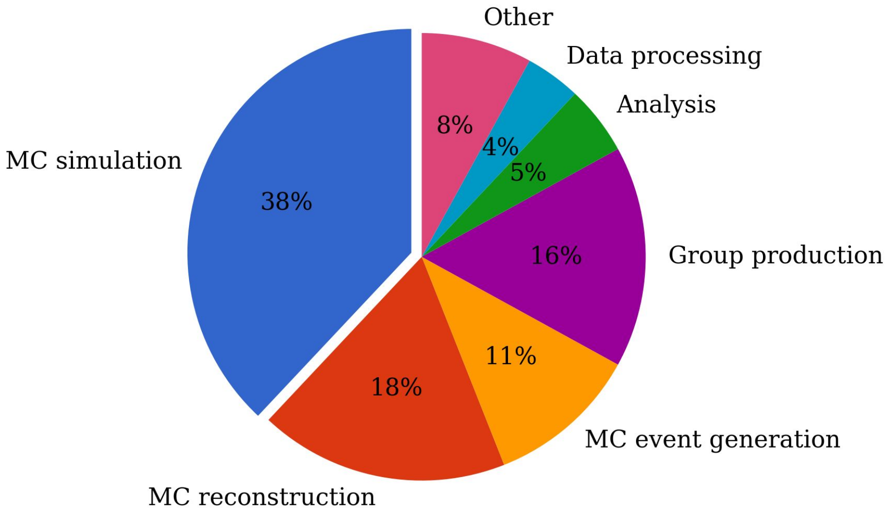
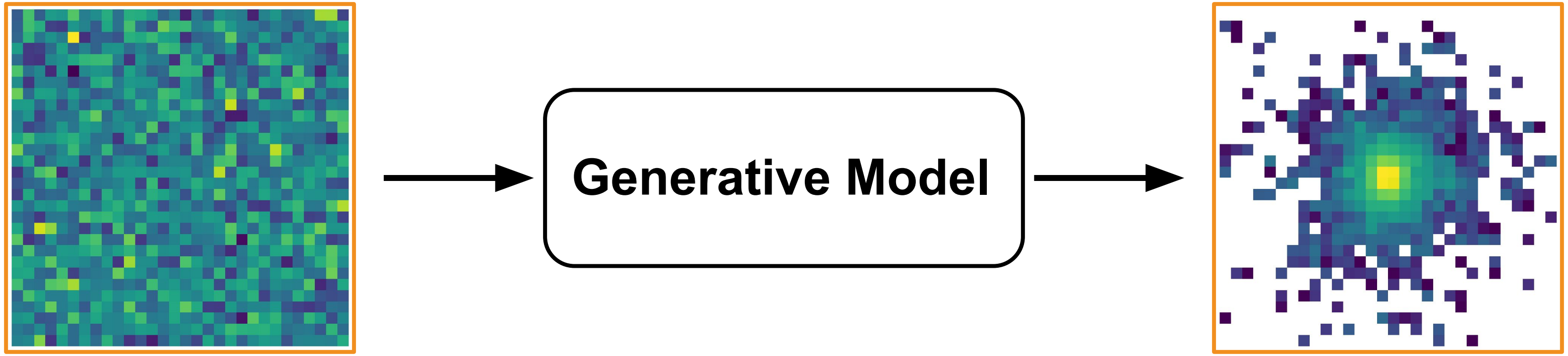


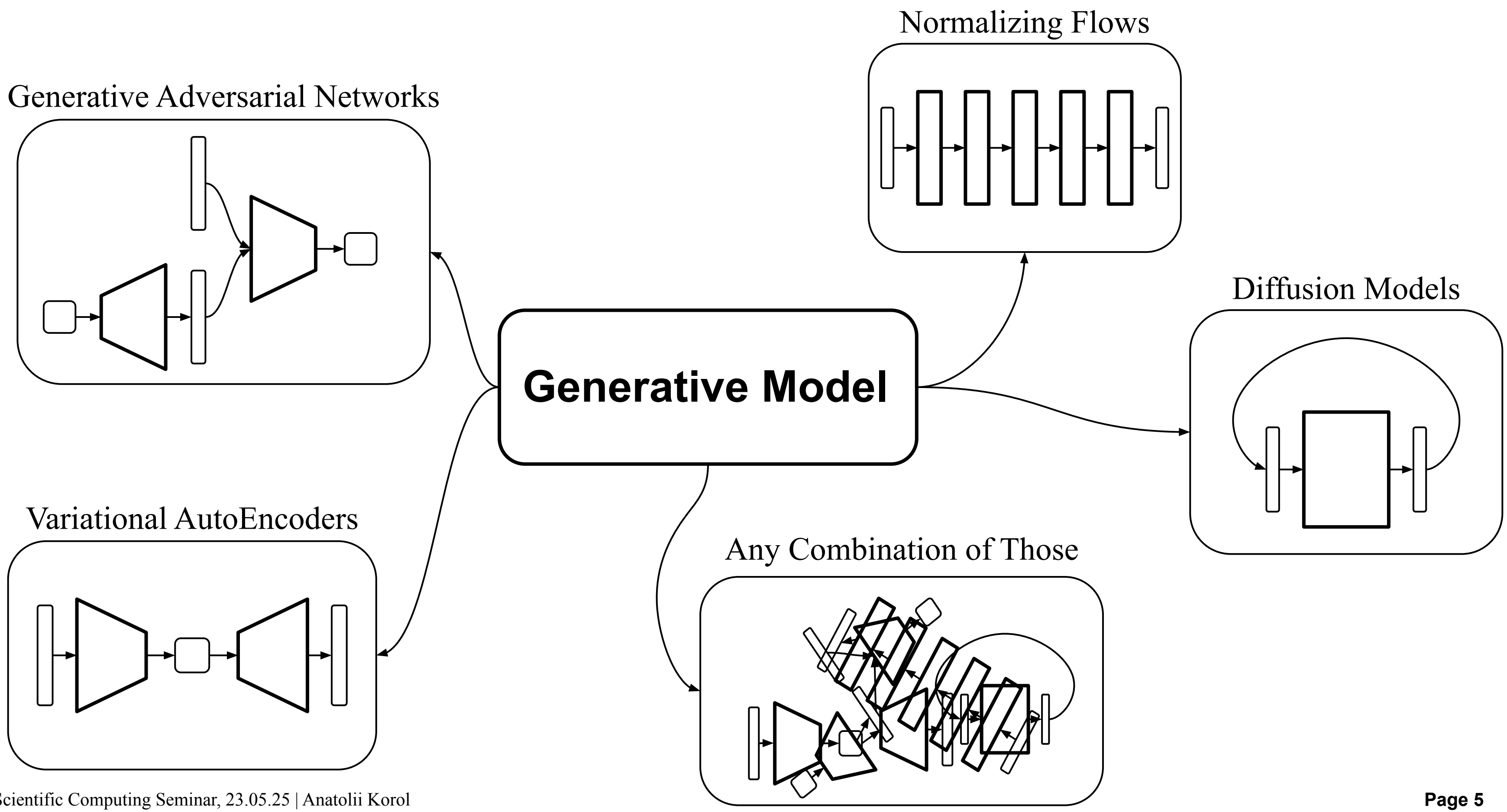
Figure adapted from:
ATLAS HL-LHC Computing Conceptual Design Report,
https://cds.cern.ch/record/2729668/files/LHCC-G-178.pdf

Goal: replace (or augment) most intensive part of simulation with a faster generator based on **generative machine learning**

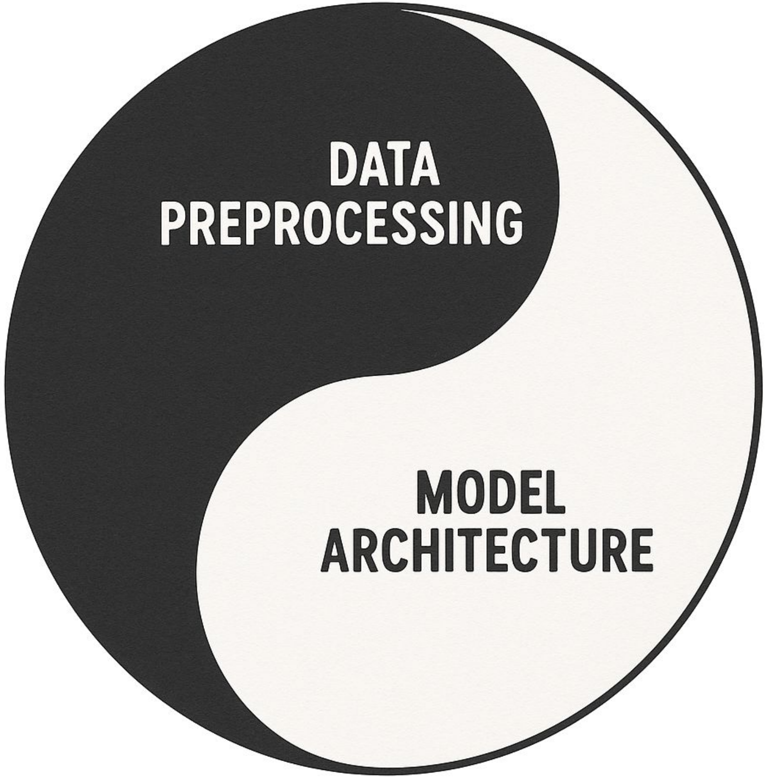
A Generative Model Is “Just a Function” That Maps Random Noise to Some Structure



Many Different Generative Models Exist Choose Any You Like

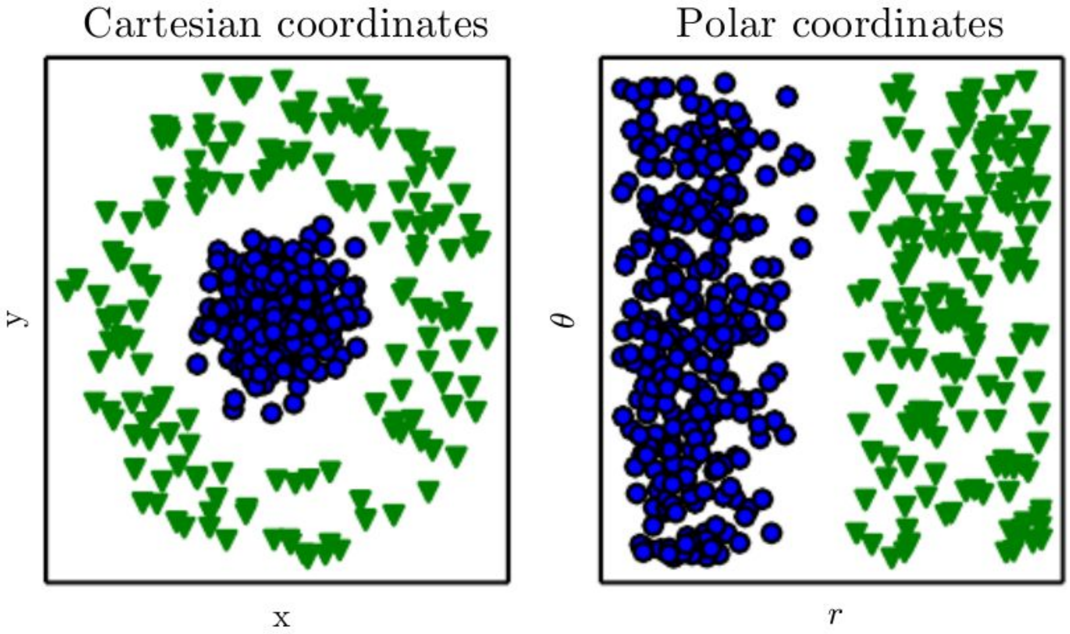


No Matter How Fancy Your Model Is the Performance Depends Heavily on the Representation of the Data It is Given



The Right Data Representation Can Turn an Impossible Problem into an Easy One

impossible task for linear model



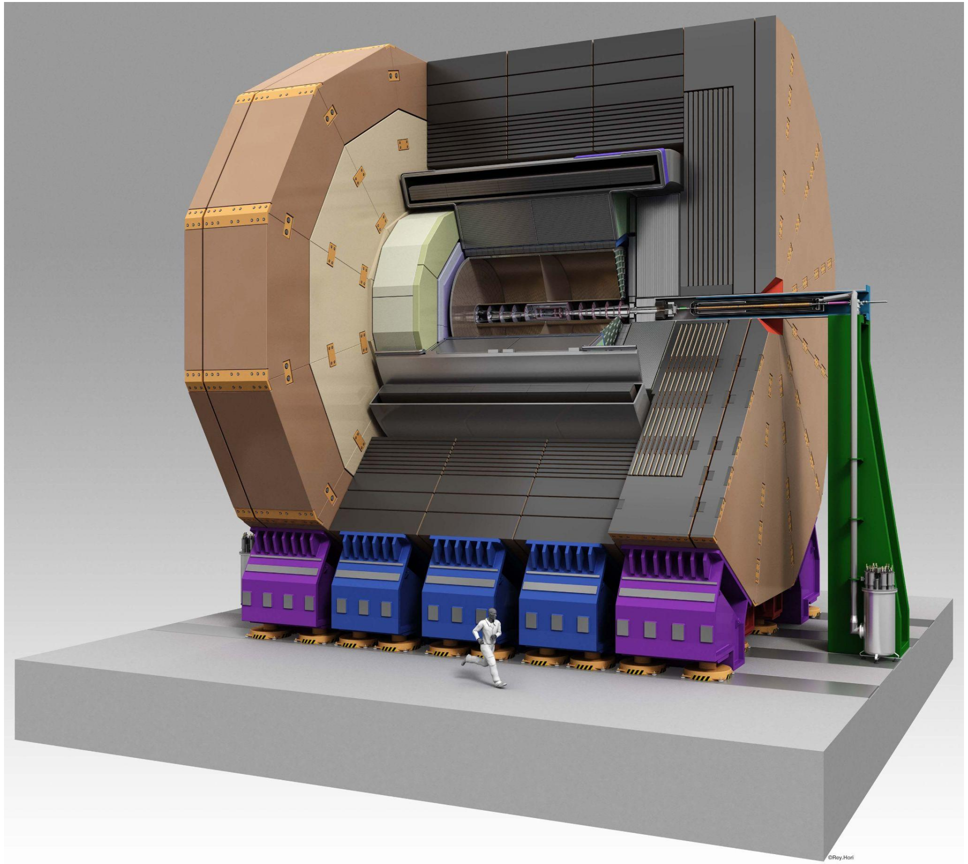
easy to solve with vertical line

Figure from:
Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville,
https://www.deeplearningbook.org/

Target: Generative Model for Electromagnetic Showers

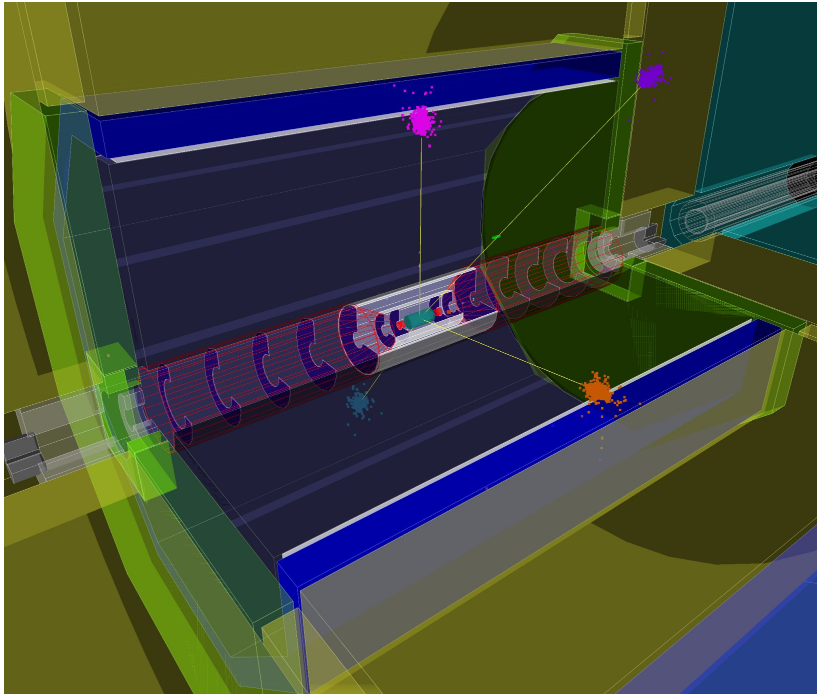
Case study: International Large Detector (ILD) concept

- Detailed and realistic simulation model
- Optimized for Particle Flow (PandoraPFA):
    - High granularity calorimeters
- Same technologies are widely planned for future experiments: e.g. HL-LHC, e^+e^- Higgs Factories
- Presents challenges for a realistic use case

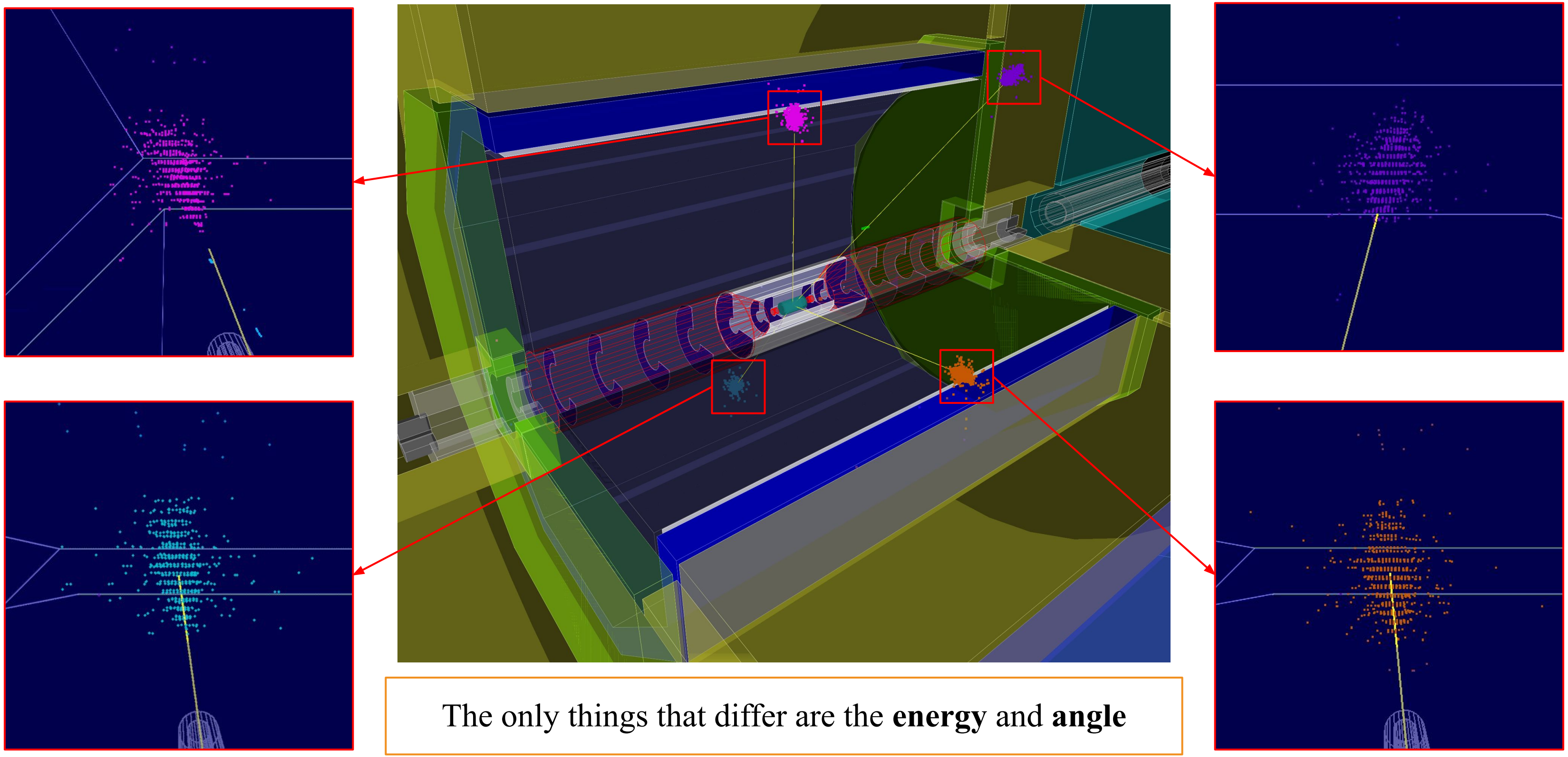


Dataset Preparation

Exploiting Geometrical Symmetries of the Detector



Shower Development is Consistent Across Regions with Identical Material Structure

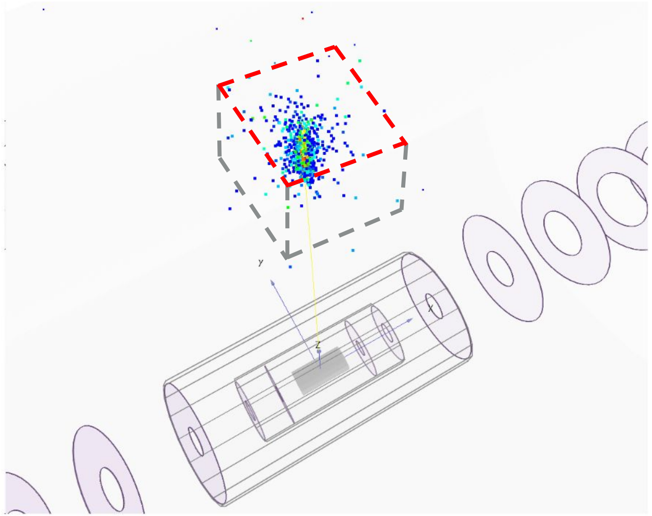


The only things that differ are the **energy** and **angle**

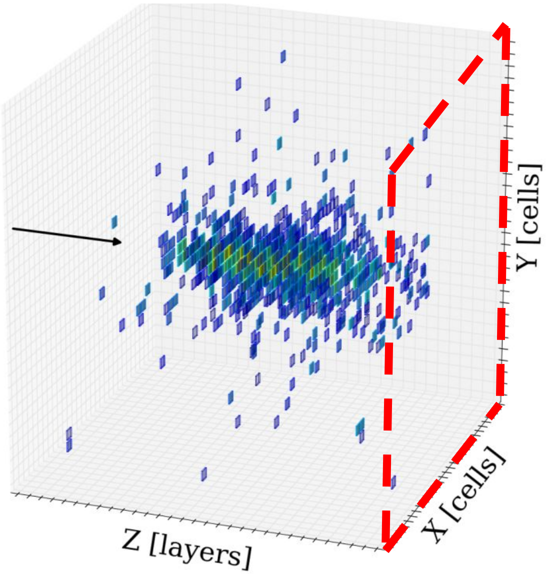
Image Representation

Previous Studies Relay on Fixed Grid Representation

60 GeV photon shower in the ILD ECAL



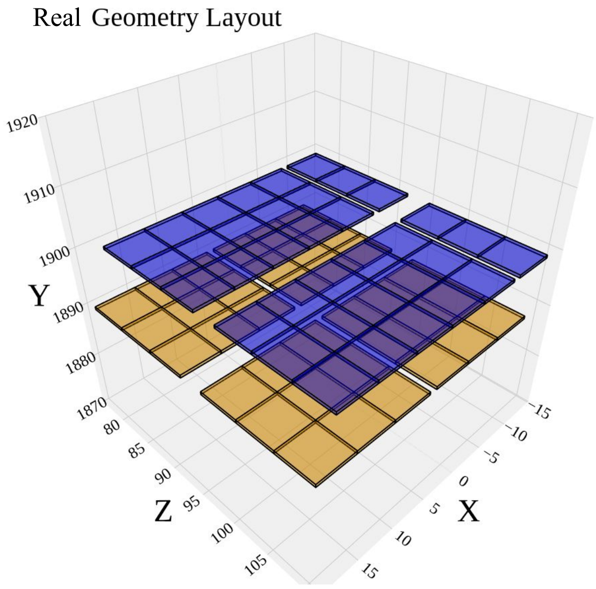
Regular grid 30x30x30

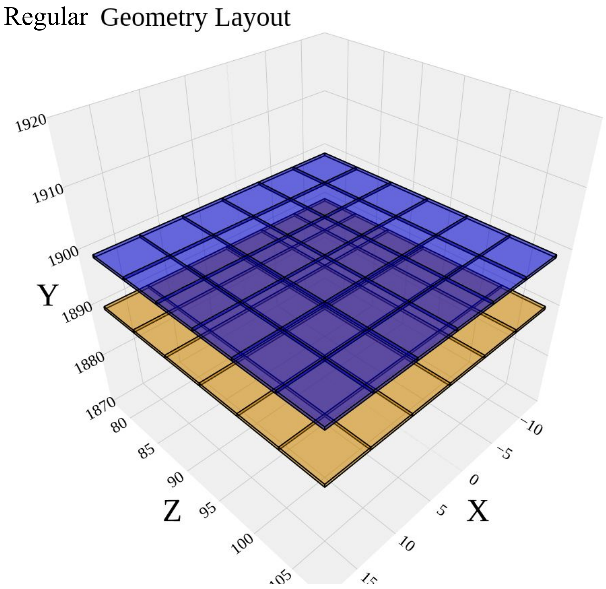


One to one mapping from detector cells to a regular grid

Problems with Image Representation of the EM Showers

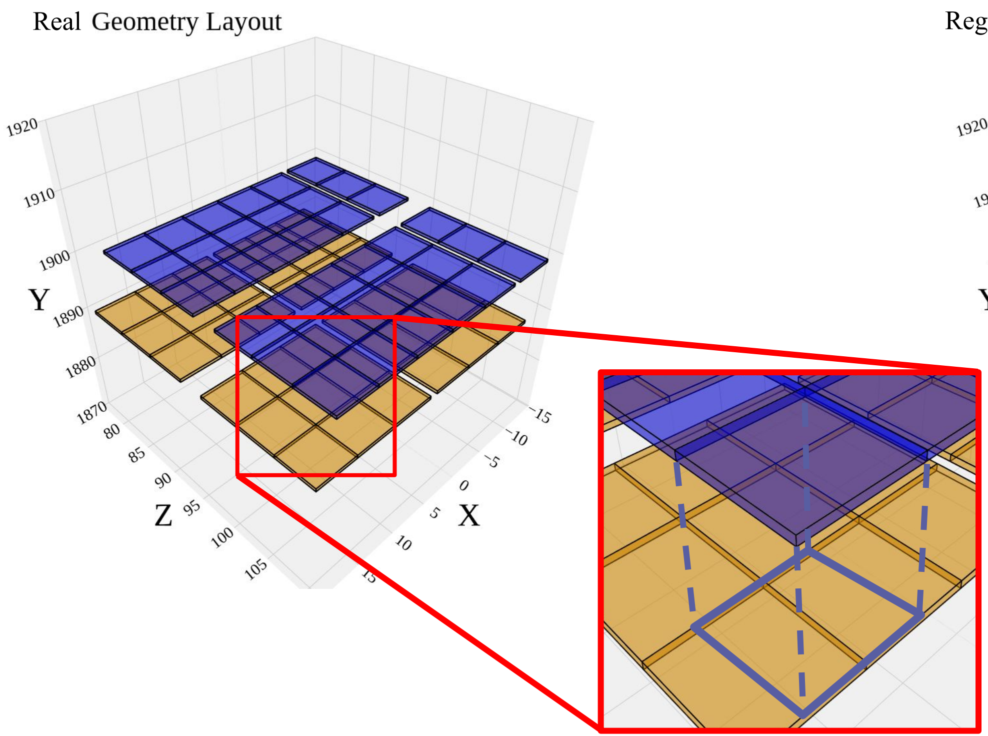
ILD ECAL Layers Structure

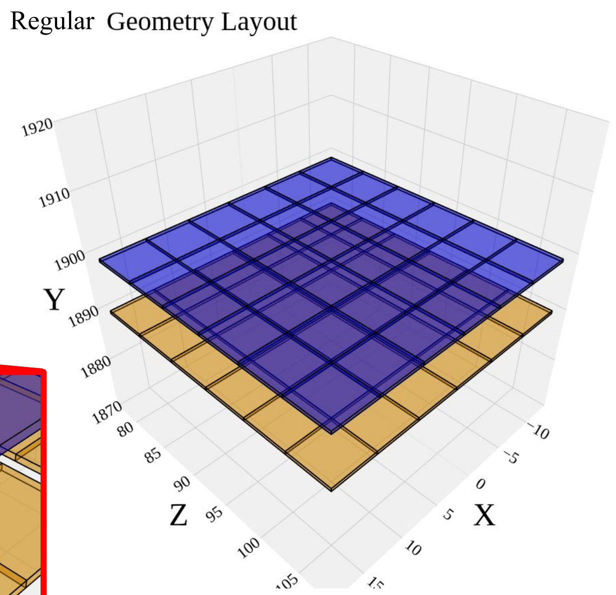




Problems with Image Representation of the EM Showers

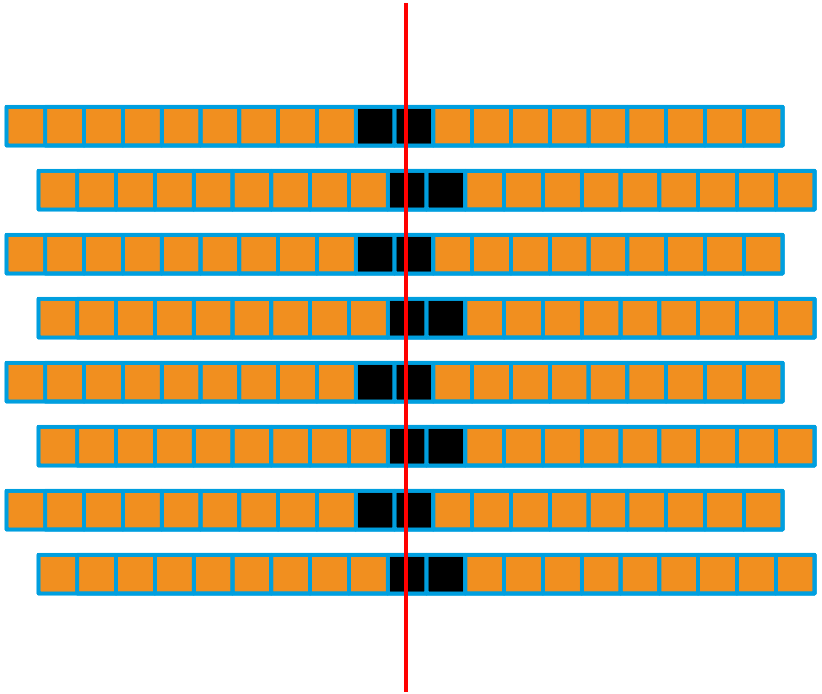
ILD ECAL Layers Structure

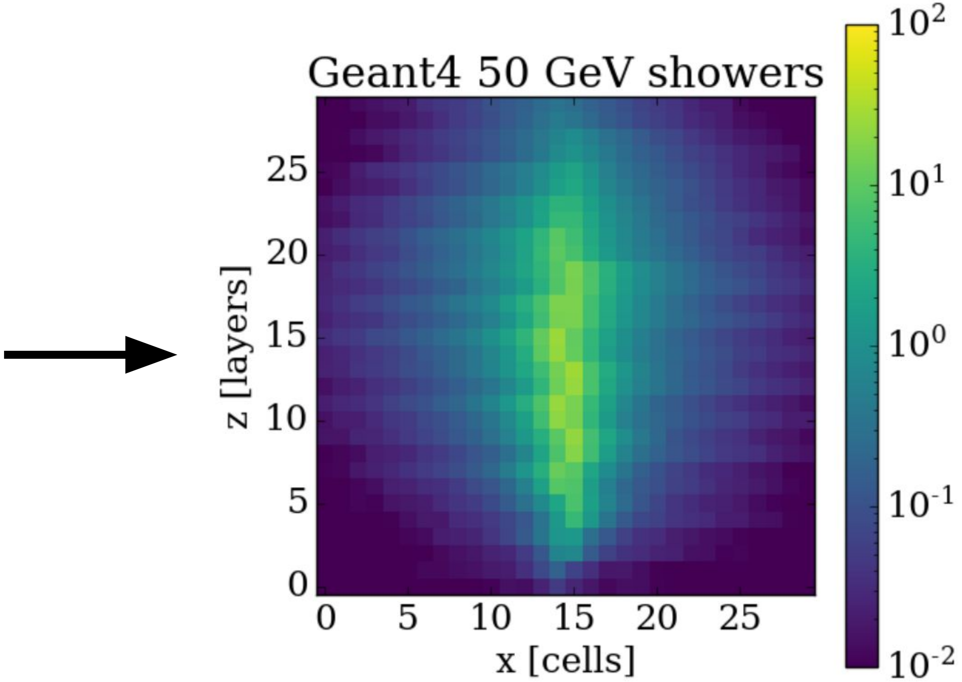




Problems with Image Representation of the EM Showers

Staggering Effect



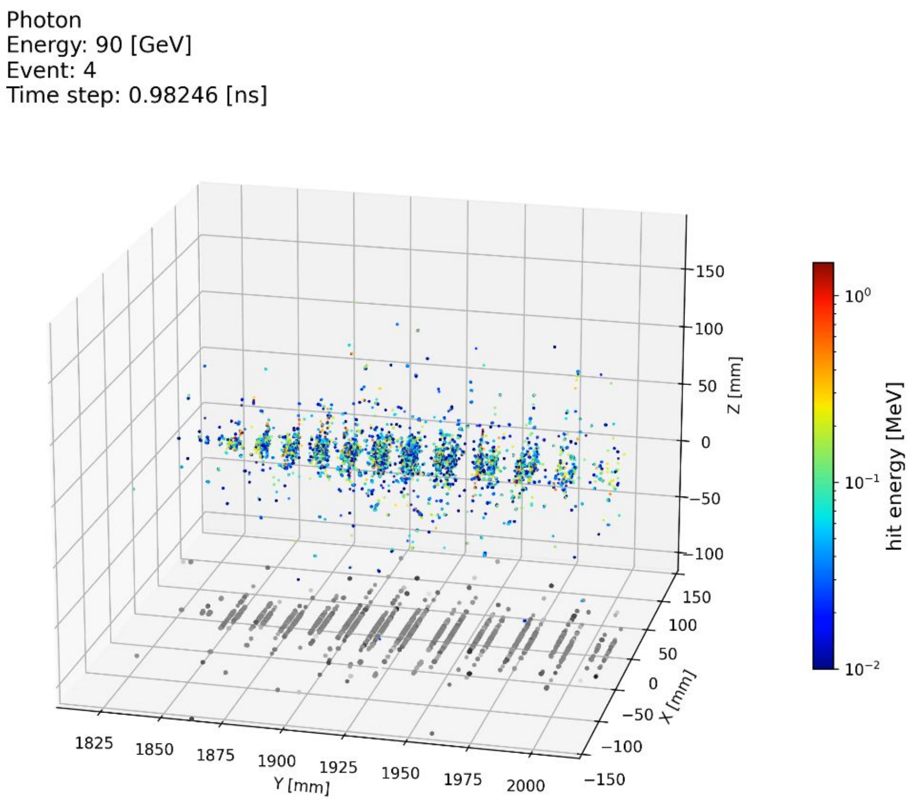


Models have to learn not only EM shower properties, but also geometry “artifacts”, like **staggering effect**

Point Cloud Representation of the EM Showers

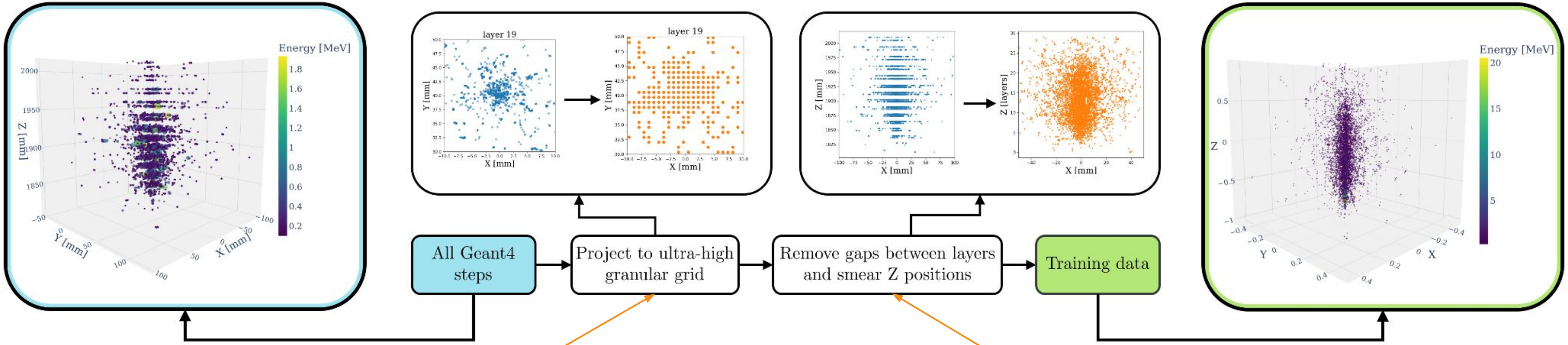
To address potential issues from irregular (realistic) cell geometry, one could use much higher resolution

- GEANT4 Steps, ultimate resolution
- Detached from detector layer geometry
- **Require preprocessing step to reduce number of spacepoints**



Point Cloud Representation of the EM Showers

Data Preprocessing

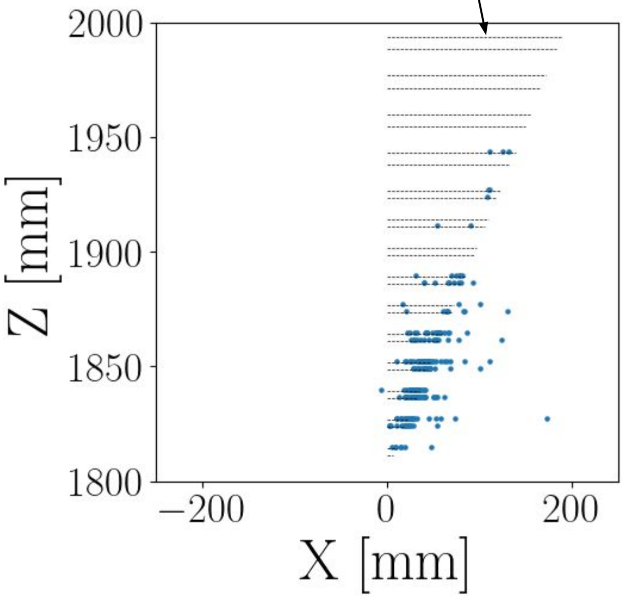


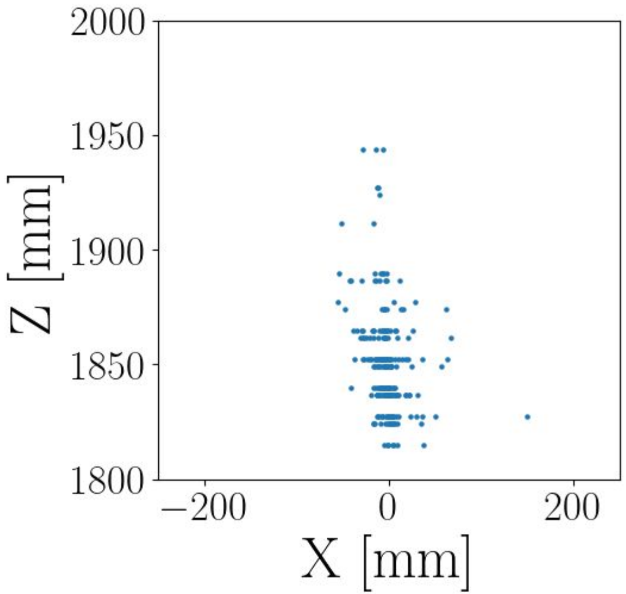
Reduces number of points per shower while maintaining sufficient resolution for translation invariance

For a better training objective, continuous space is easier to learn than quantized one

Simple Trick to Simplify the Objective

- Applying a layer-wise **geometrical offset** to showers entering at an angle, aligning them as if they had an orthogonal impact





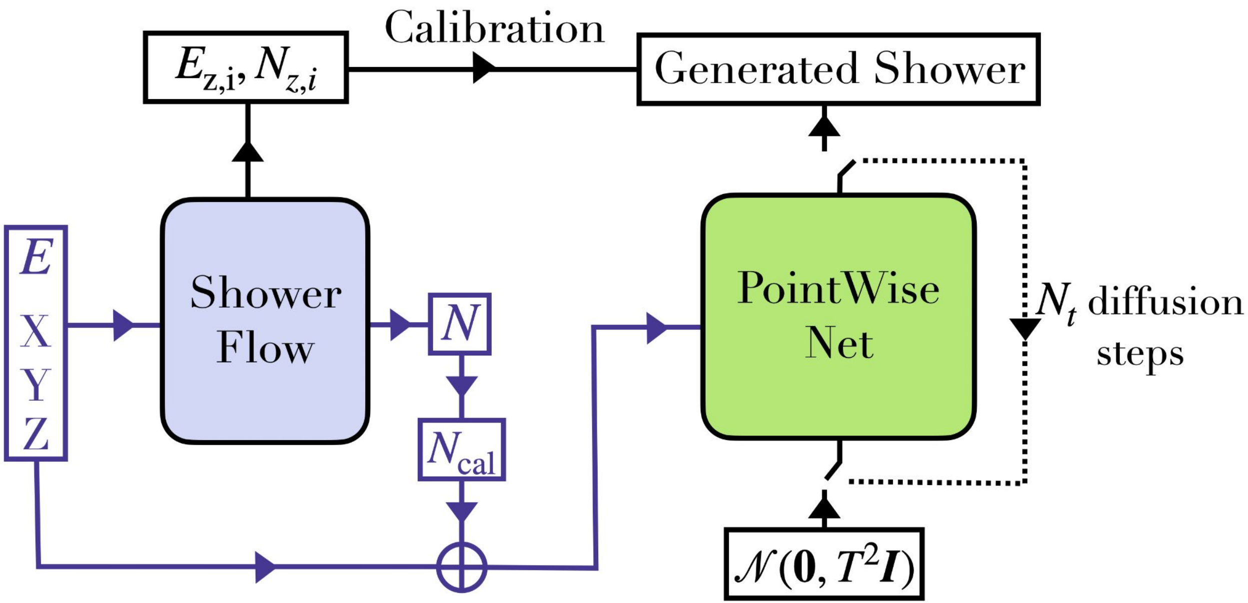
Advantages: smaller box and a simplified training objective

Model Architecture

CaloClouds3

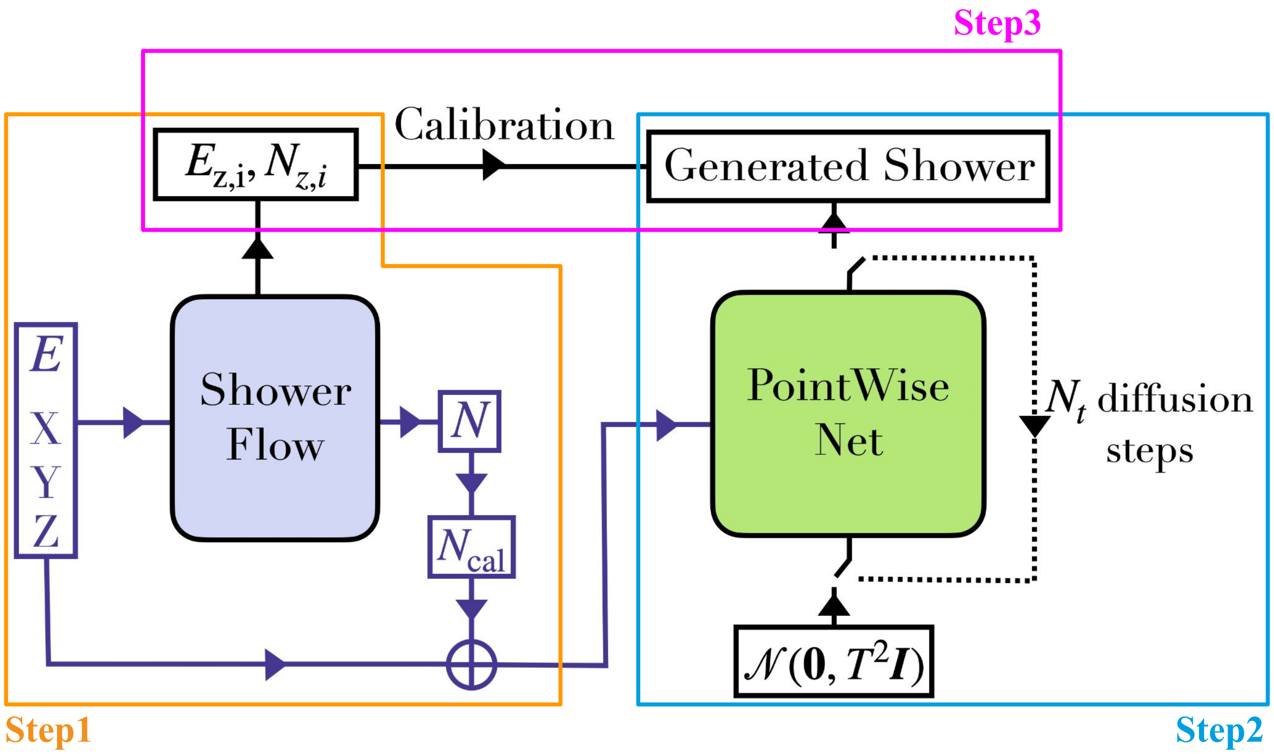
CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation, Buhmann, AK, et al. 2023, **arXiv:2305.04847**

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



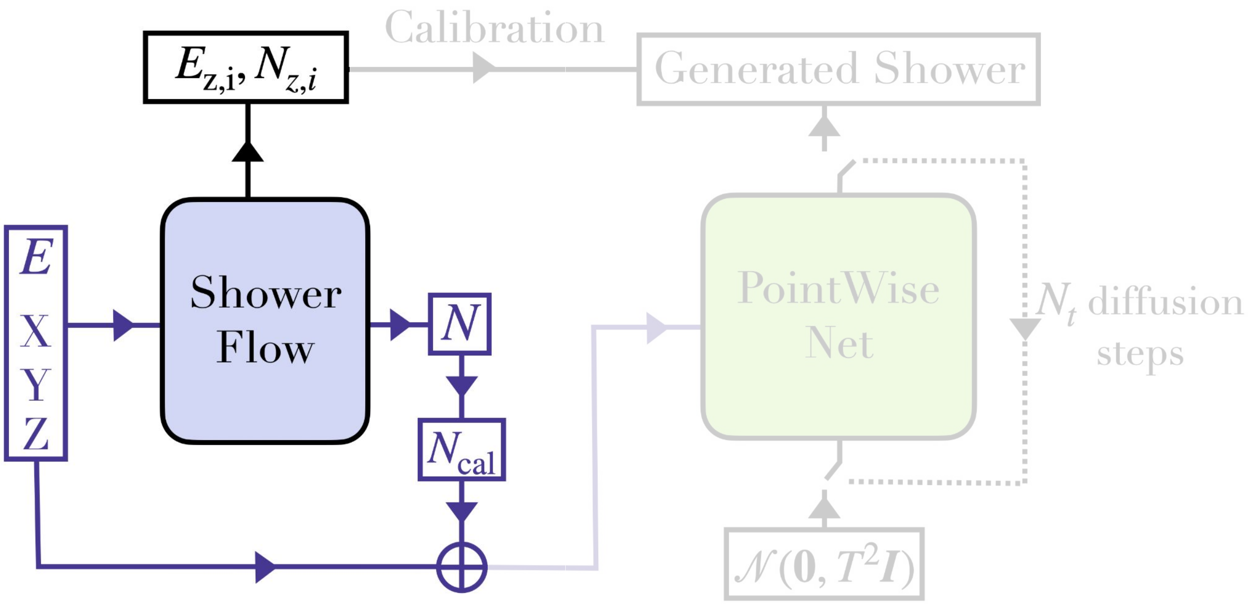
CaloClouds3

3 Steps Pipeline



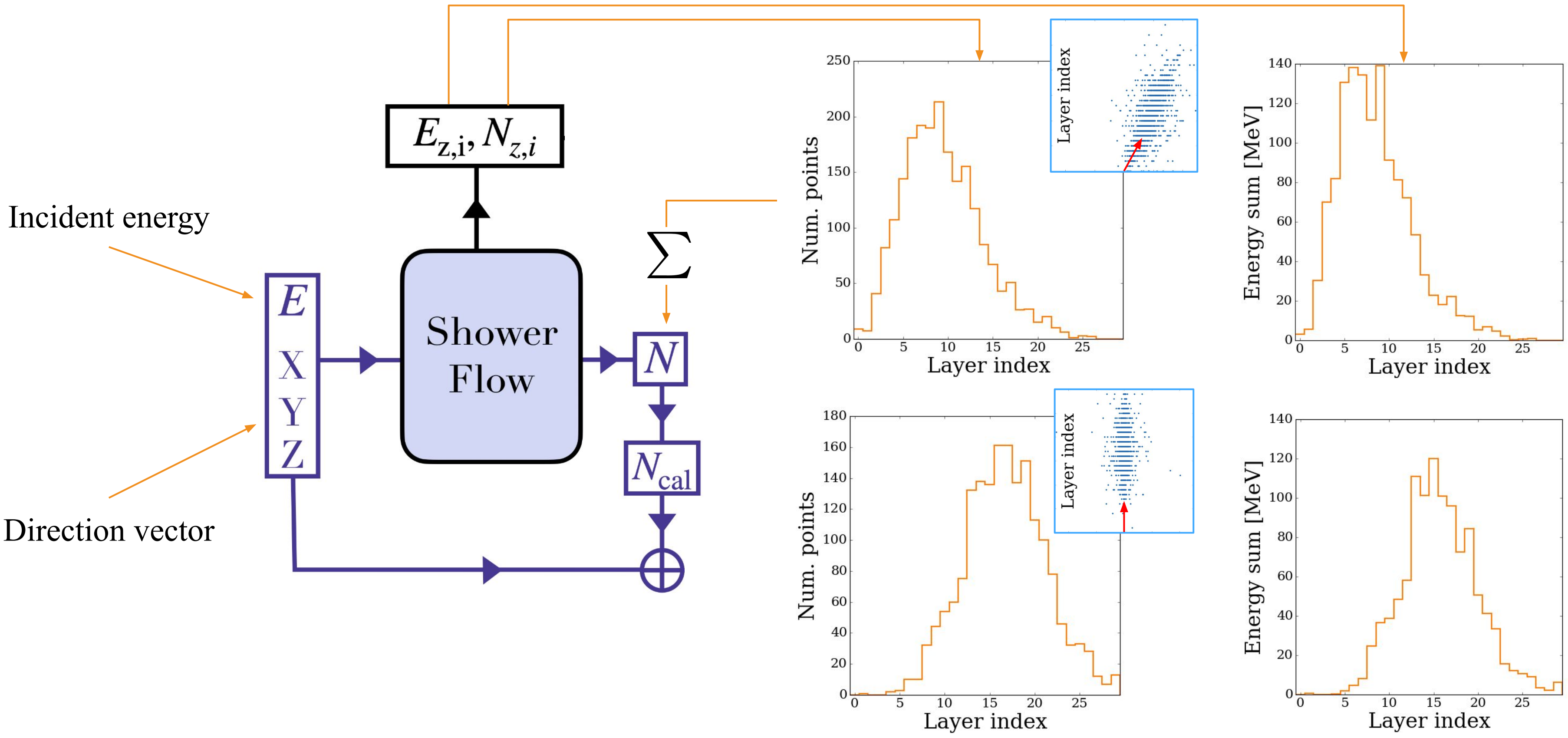
CaloClouds3

Step1: Normalizing Flow Model



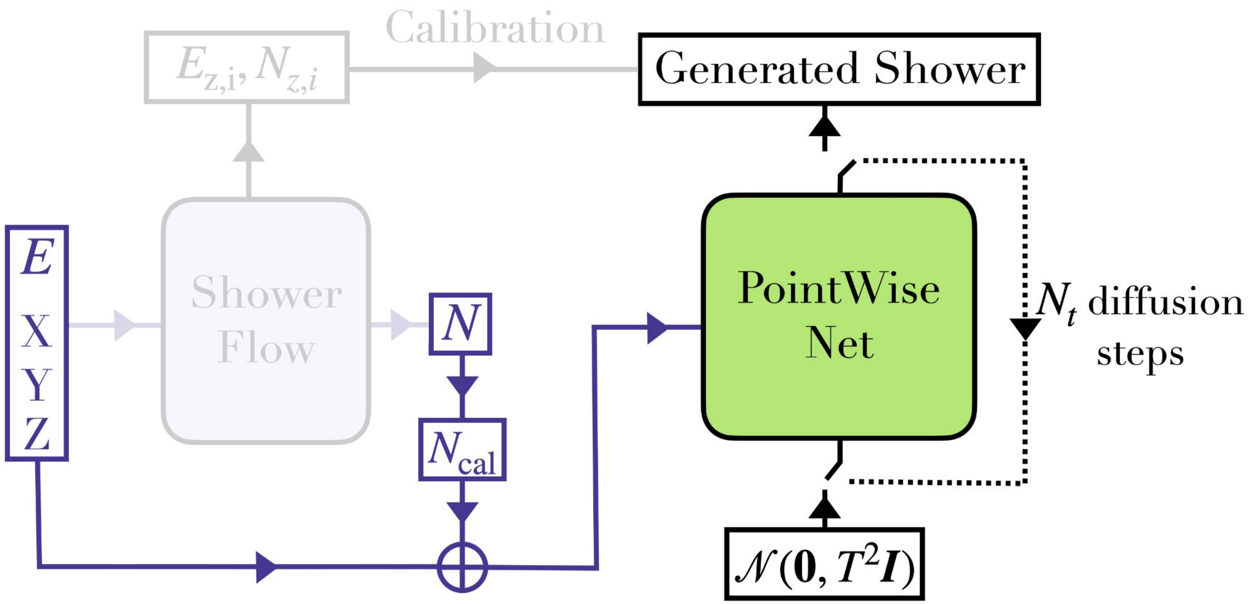
CaloClouds3

Normalizing Flow Model Generates Low Level Shower Observables



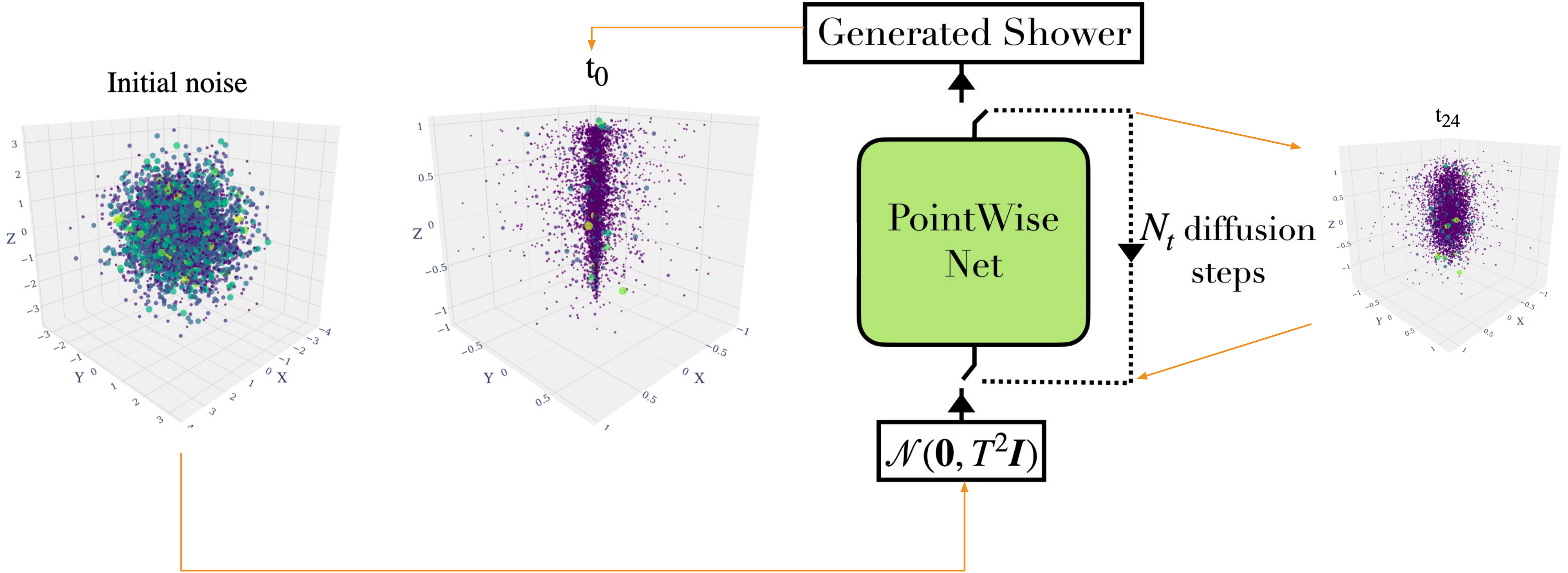
CaloClouds3

Step2: Diffusion Model



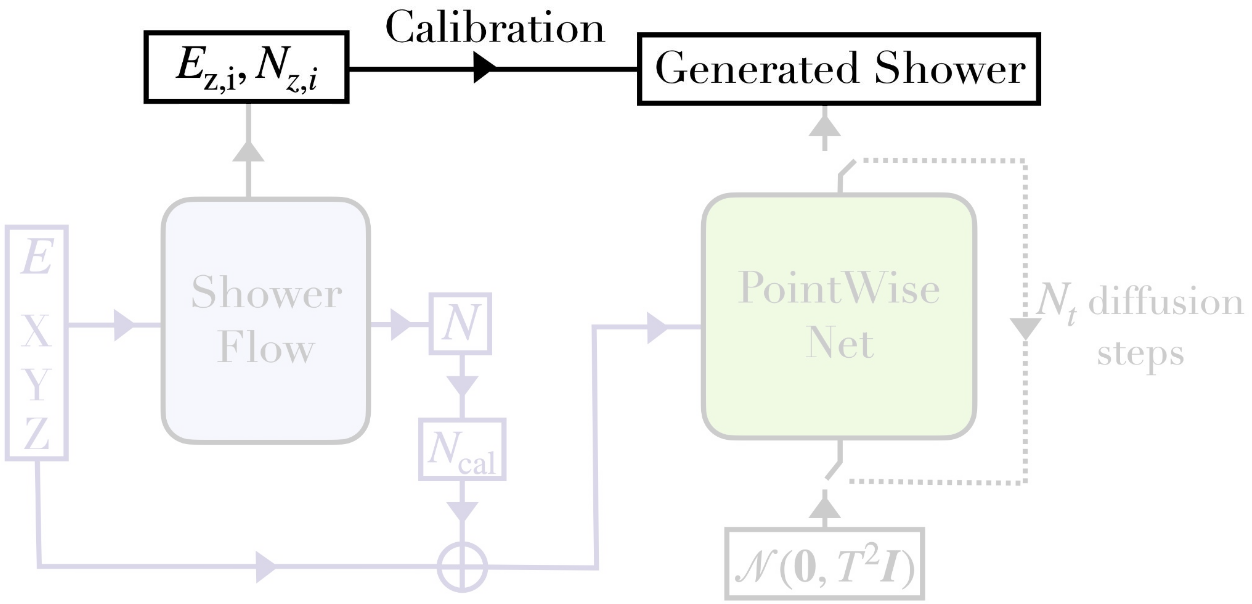
CaloClouds3

Step2: Diffusion Model



CaloClouds3

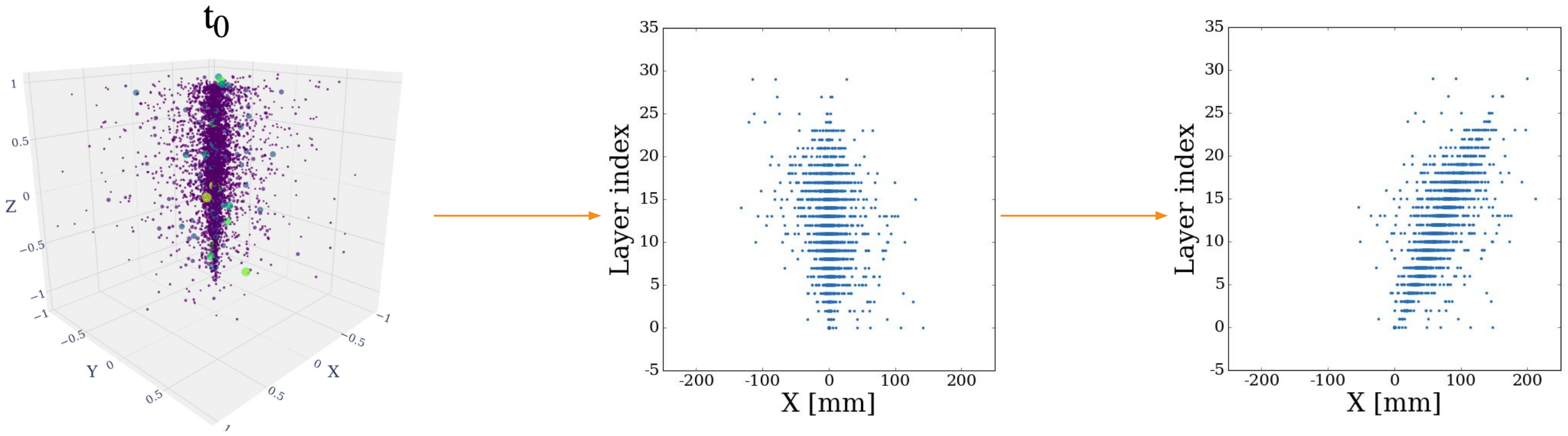
Step3: Calibration and Postprocessing



CaloClouds3

Step3: Calibration and Postprocessing

$$\boxed{E_{z,i}, N_{z,i}} \xrightarrow{\text{Calibration}} \boxed{\text{Generated Shower}}$$



Integration into Simulation Chain

DDFastShowerML

Generic library for running ML-based fast sim models in DD4hep

Algorithm 1 Pseudocode illustrating the order of operations for the core components of the *DDFastShowerML* library.

```
if Trigger.checkTrigger(track) == True then
    Kill full simulation of particle
    localDir = Geometry.getLocalDir(track)
    inputs = Model.prepareInputs(track, localDir)
    outputs = Inference.runInference(inputs)
    localSPs = Model.convertOutput(track, localDir, outputs)
    globalSPs = Geometry.localToGlobal(track, localSP)
    for (sp in globalSPs) do
        HitMaker.makeHit(sp, track)
    end for
else
    Full simulation of particle with GEANT4
end if
```

Algorithm from:
Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models,
Peter McKeown,
DOI: 10.3204/PUBDB-2024-01825

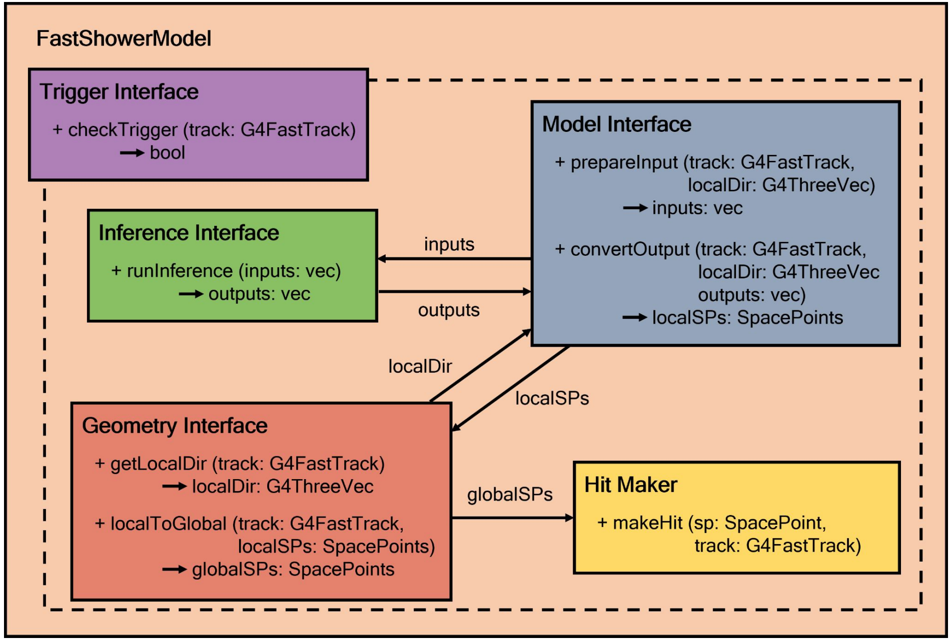


Figure from:
Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models,
Peter McKeown,
DOI: 10.3204/PUBDB-2024-01825

DDFastShowerML

Generic library for running ML-based fast sim models in DD4hep

Algorithm 1 Pseudocode illustrating the order of operations for the core components of the *DDFastShowerML* library.

```
if Trigger.checkTrigger(track) == True then
    Kill full simulation of particle
    localDir = Geometry.getLocalDir(track)
    inputs = Model.prepareInputs(track, localDir)
    outputs = Inference.runInference(inputs)
    localSPs = Model.convertOutput(track, localDir, outputs)
    globalSPs = Geometry.localToGlobal(track, localSP)
    for (sp in globalSPs) do
        HitMaker.makeHit(sp, track)
    end for
else
    Full simulation of particle with GEANT4
end if
```

Algorithm from:
Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models,
Peter McKeown,
DOI: 10.3204/PUBDB-2024-01825

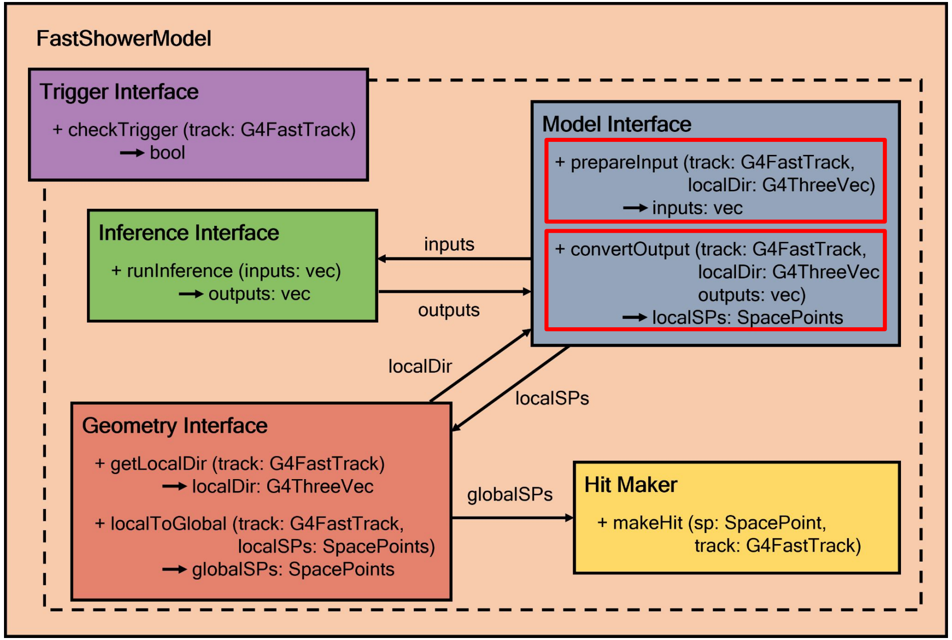


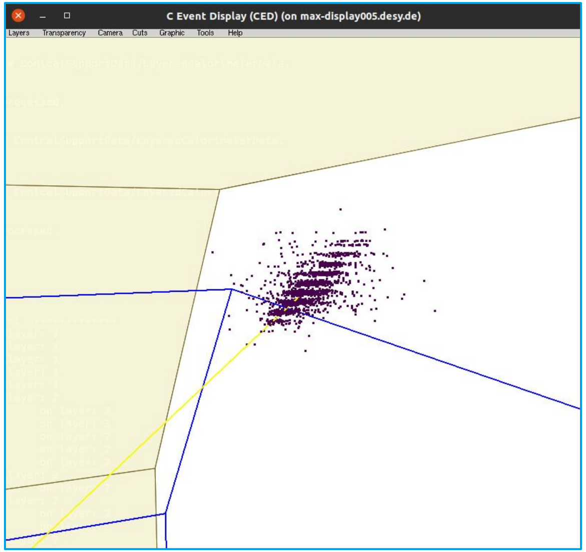
Figure from:
Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models,
Peter McKeown,
DOI: 10.3204/PUBDB-2024-01825

DDFastShowerML

Generic library for running ML-based fast sim models in DD4hep

PyTorch
Scripted fast sim model

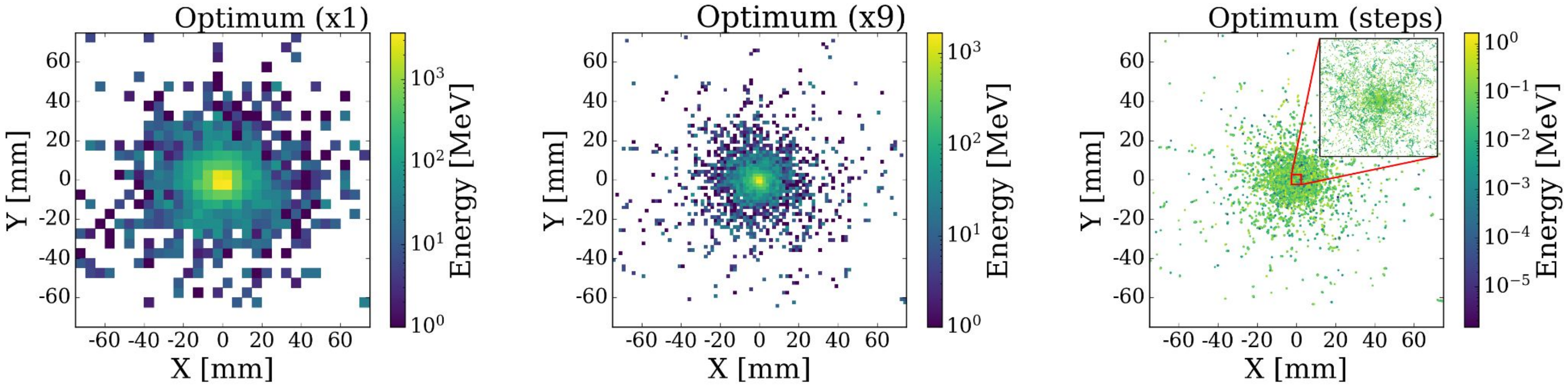
DDFastShowerML
https://gitlab.desy.de/ilcsoft/ddfastshowerml



50 GeV photon shower generated with CaloClouds3 in the ILD ECAL

Benchmarks

Optimal Generators Reference



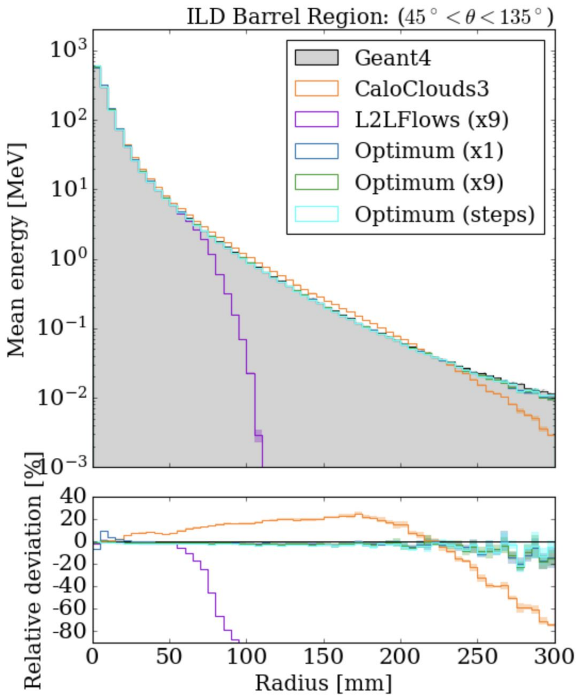
- Cell-level readout
- Common approach, used by e.g. BIB-AE
- Regular grid

- x3 lateral resolution, x9 cells
- Used by L2LFlows model
- Regular grid

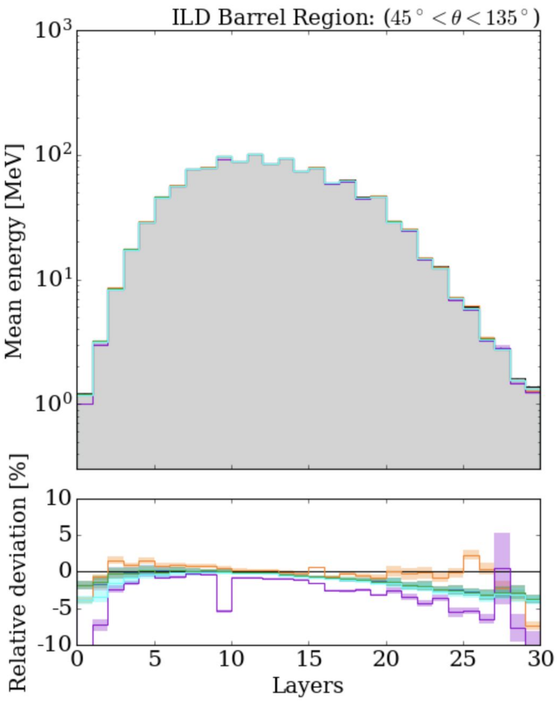
- Geant4 steps, highest resolution
- Used by CaloClouds model (with additional clustering)
- Point cloud representation

Single Particle Performance

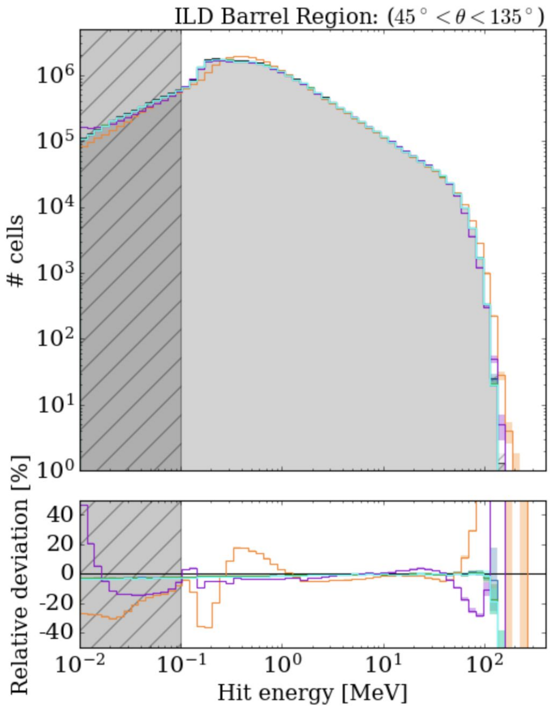
Radial Profile



Longitudinal Profile



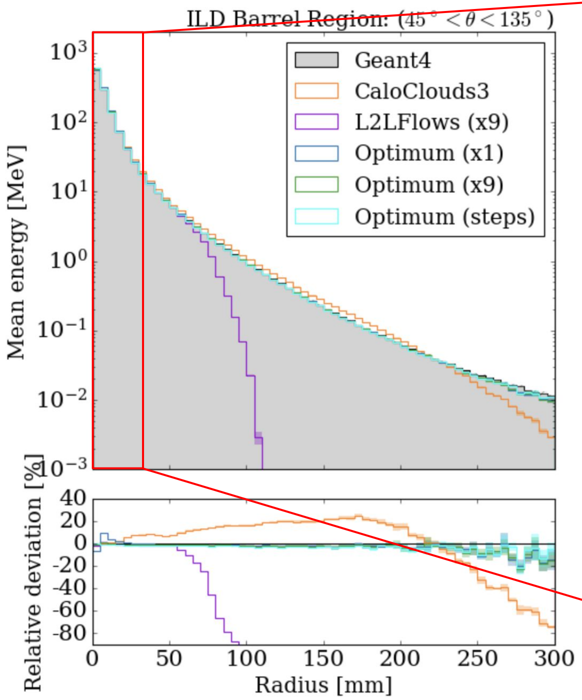
Hit Energy Spectrum

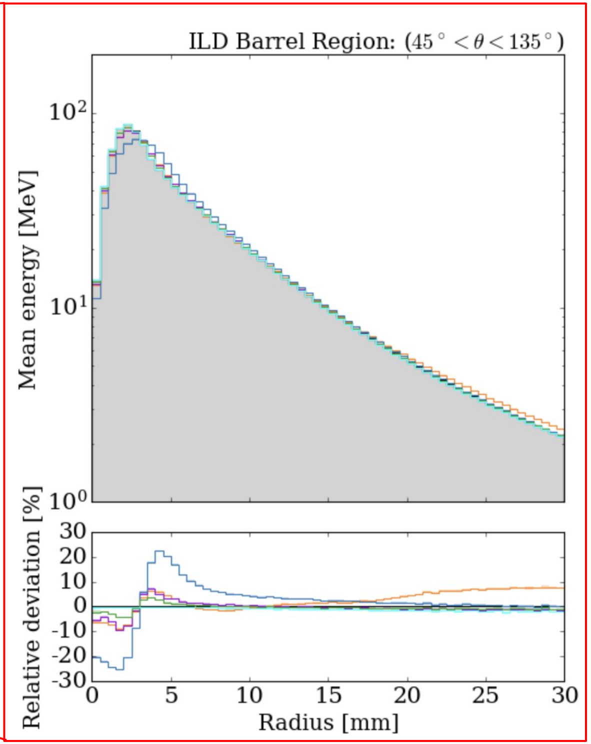


Single Particle Performance

Radial Profile

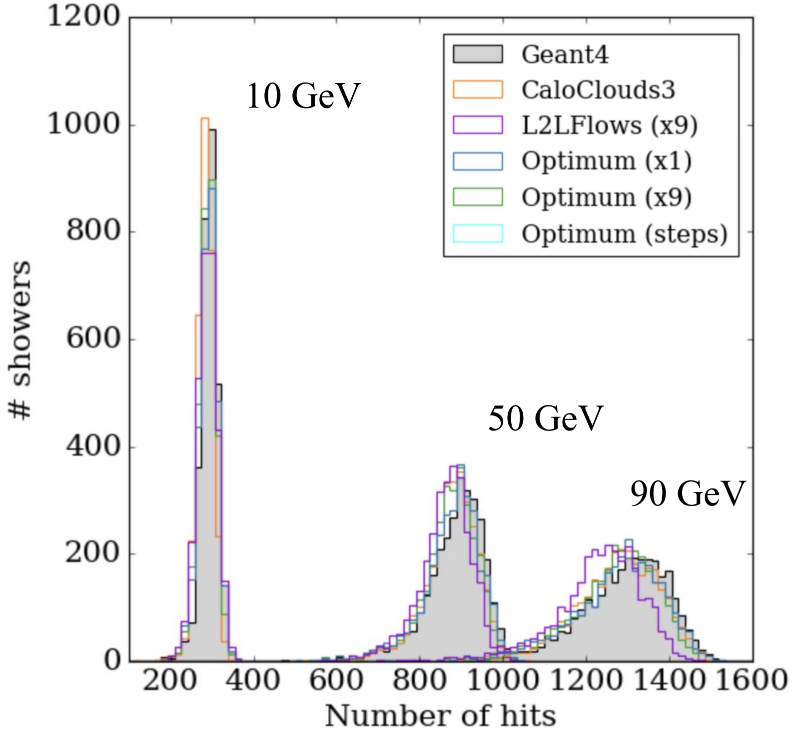
> 90% of the shower content



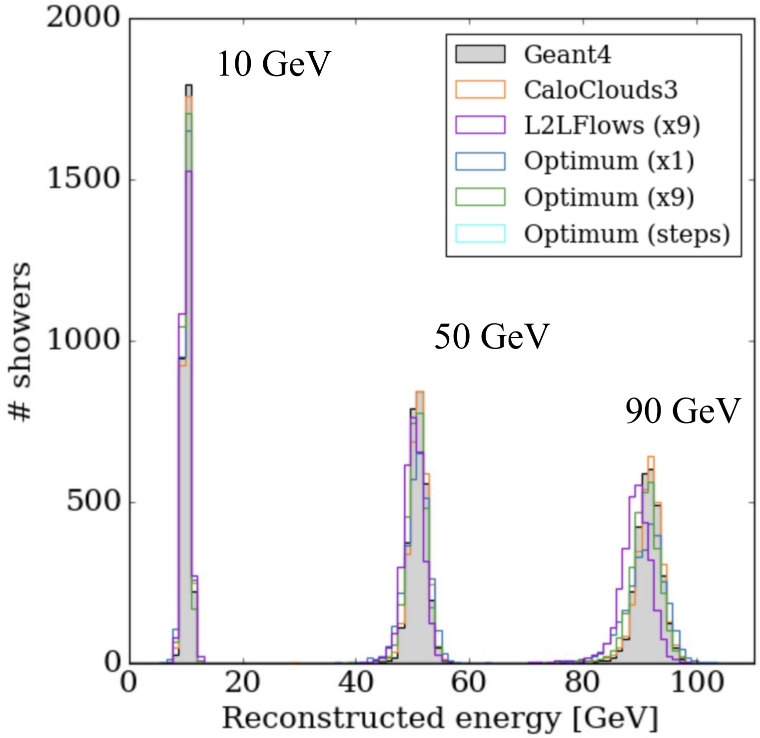


Single Particle Performance

Occupancy

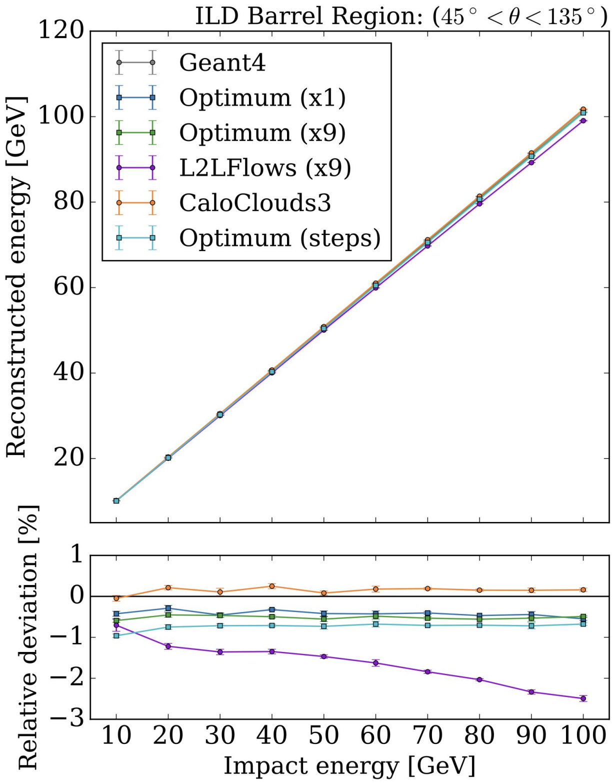


Energy Distribution

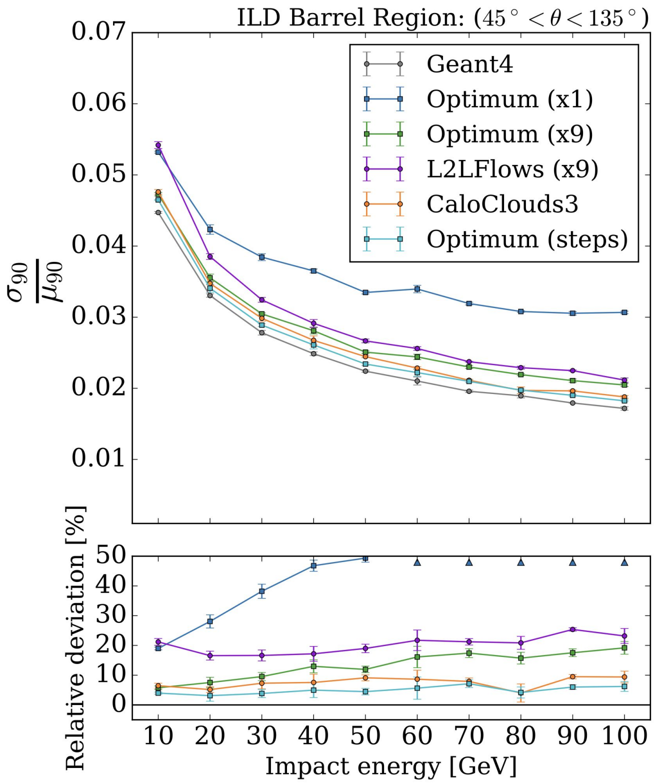


Single Particle Performance

Linearity

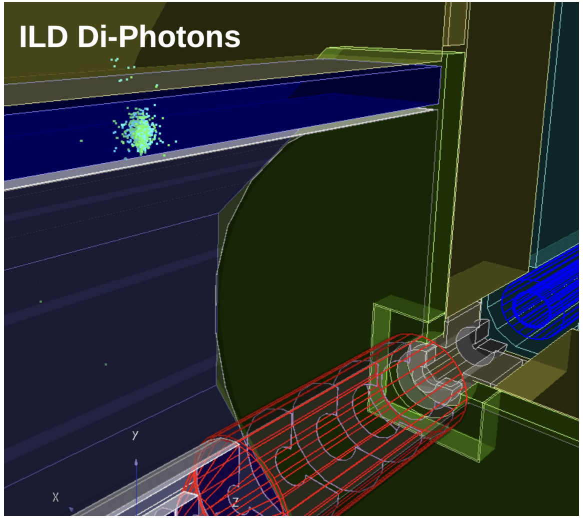


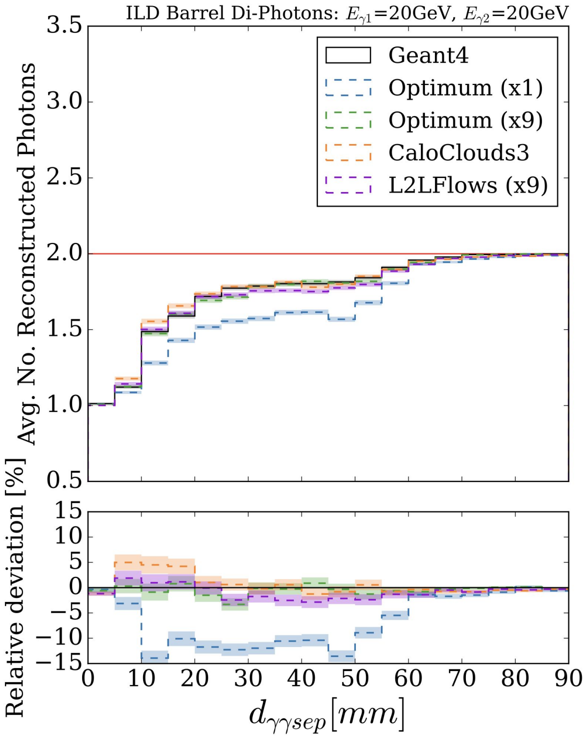
Energy Resolution



Di-Photons Reconstruction Benchmark

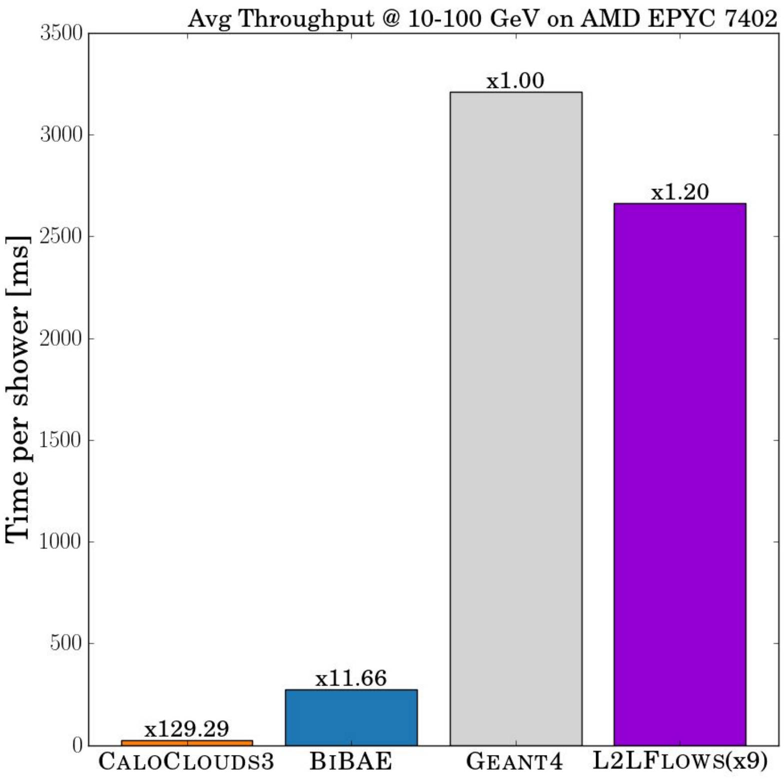
Courtesy P.McKeown



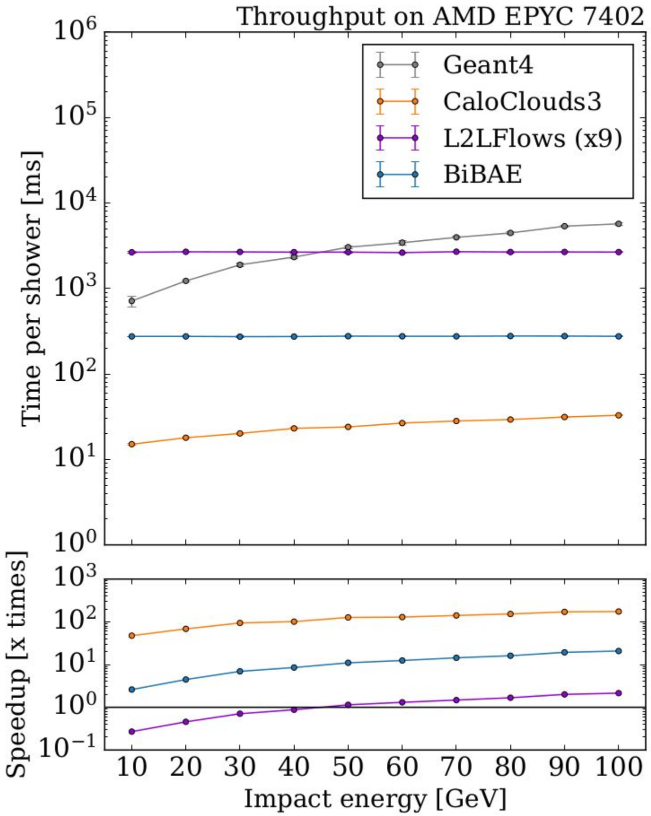


Di-Photon Reconstruction Benchmark provides a direct **physically relevant** quantification of model performance

Timing



Avg Throughput @ 10-100 GeV on AMD EPYC 7402
Time per shower [ms]
x1.00
x1.20
x11.66
x129.29
CaloClouds3
BiBAE
Geant4
L2LFlows(x9)

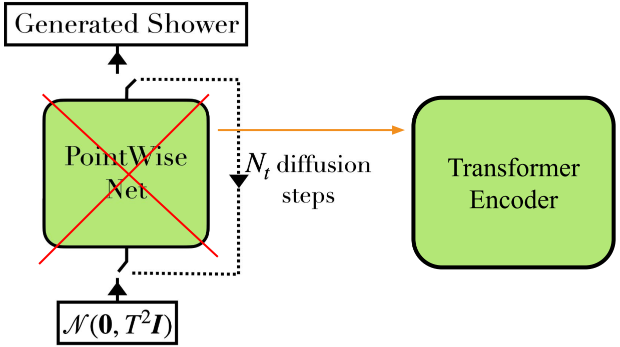


Throughput on AMD EPYC 7402
Geant4
CaloClouds3
L2LFlows (x9)
BiBAE
Time per shower [ms]
Speedup [x times]
Impact energy [GeV]

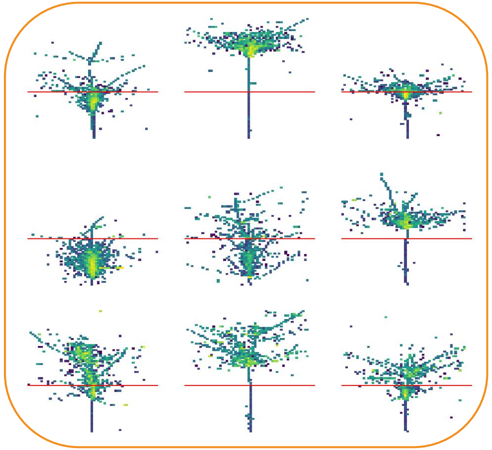
Possible Extensions

Hadron Showers

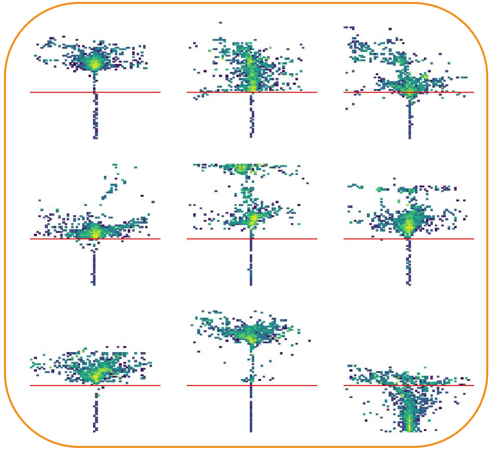
POC for Pion Showers in Combined ECAL + HCAL



Geant4



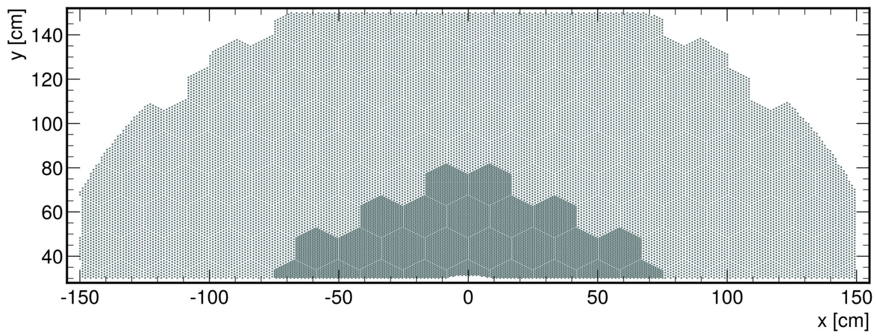
PionClouds

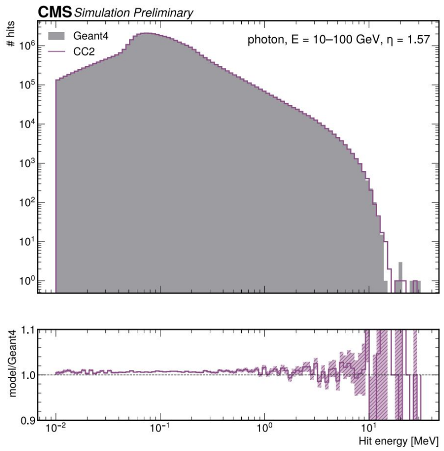


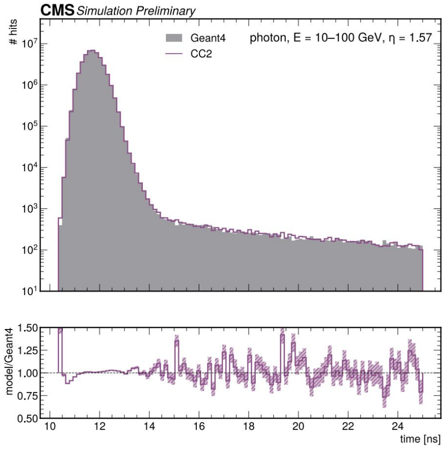
Different Detectors

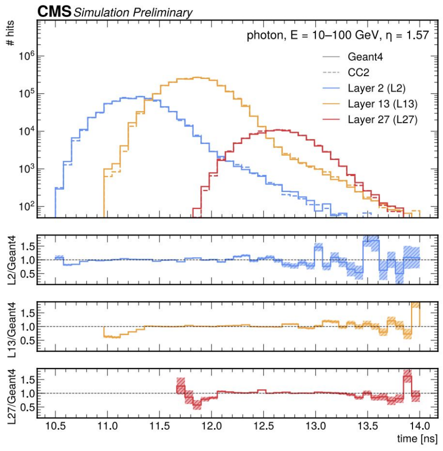
Adopted for CMS-HGCAL by W. Korcari

It's about time: a Point Cloud Generative Model for the CMS High Granularity Calorimeter, CMS Collaboration. 2025, CMS-DP-2025-016, CERN-CMS-DP-2025-016, **CDS:2932517**









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**

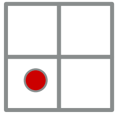
- Traditional fixed grid representations suffer from geometry artifacts, limiting model generalization
- CaloClouds introduces a data representation paradigm that addresses this challenge
- CaloClouds offers a general solution that is easily adaptable to different detector geometries, e.g. CMS-HGCAL
- CaloClouds3 outperforms previous state-of-the-art models while being approximately ~10x faster, and ~100x faster than GEANT4 @ 10-100 GeV range on the same hardware

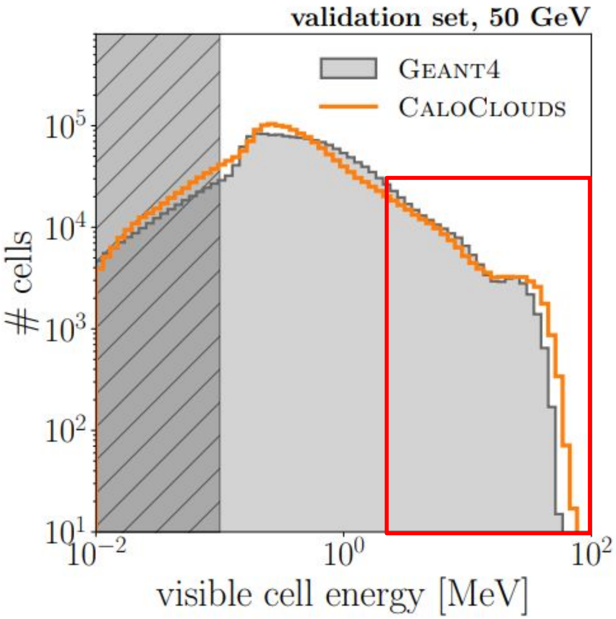
BACKUP SLIDES

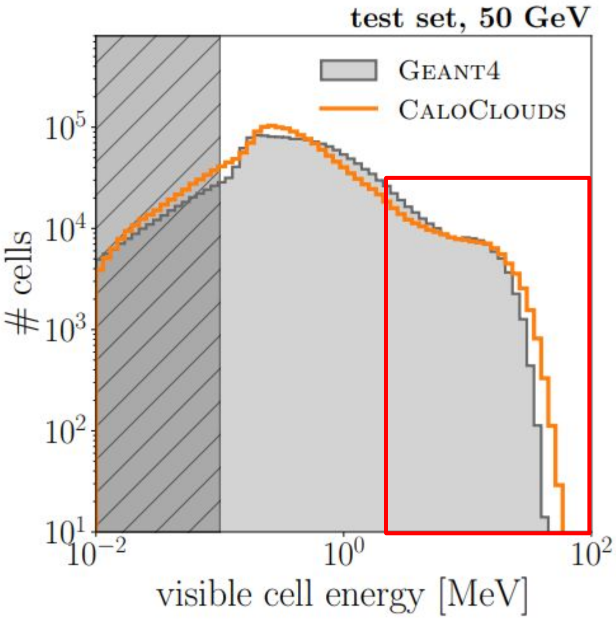
Physics Observables at Different Positions

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

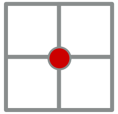
validation incident point pos.







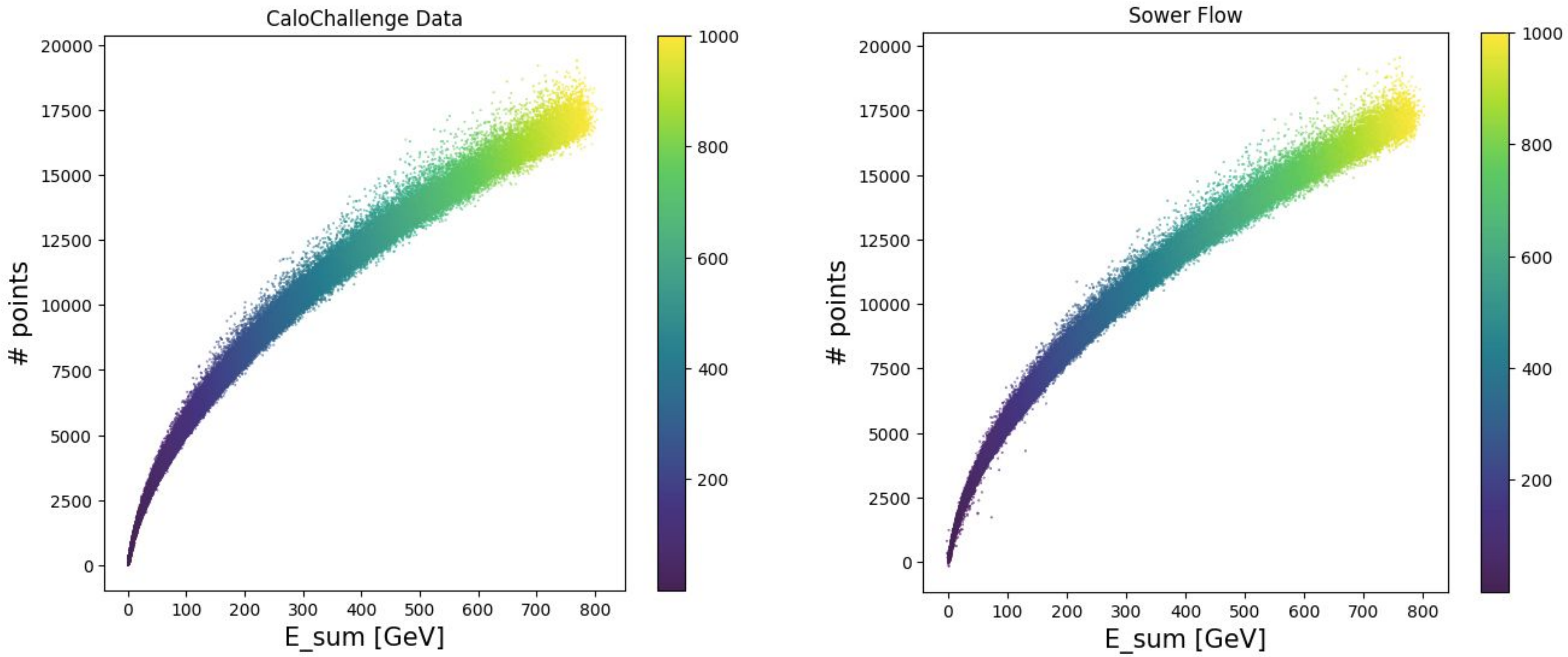
test incident point pos.



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

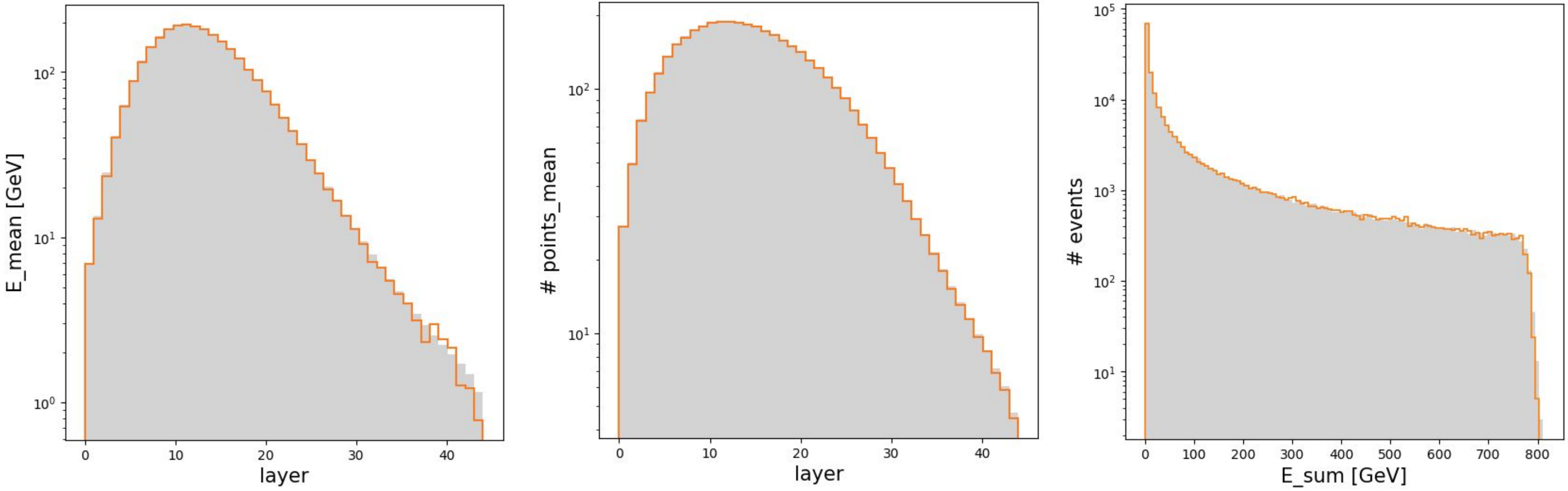
Shower Flow

Results

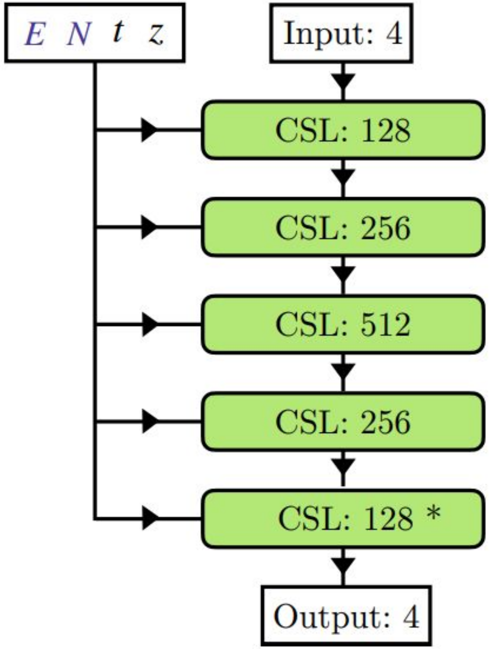


Shower Flow

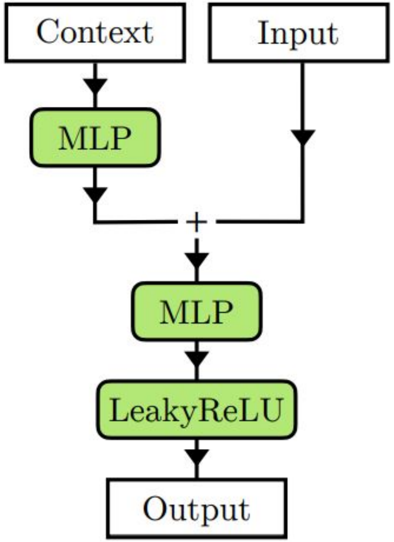
Results



PointWise Net



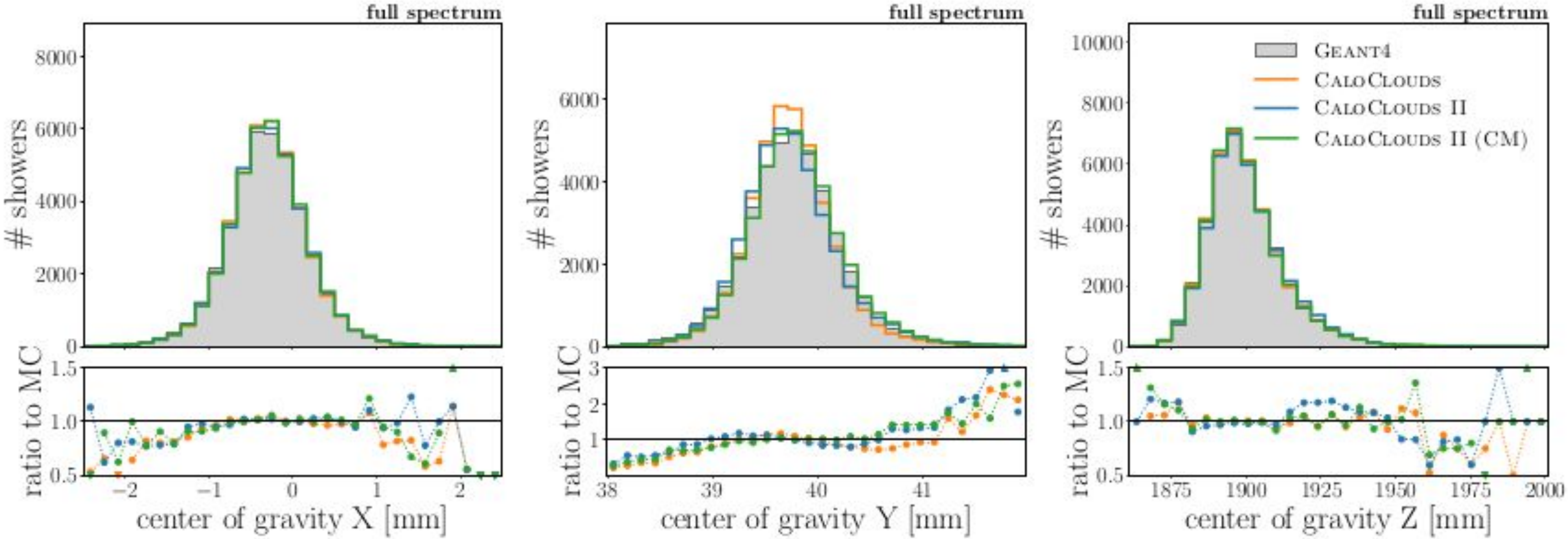
(a) PointWise Net



(b) ConcatSquash Layer (CSL)

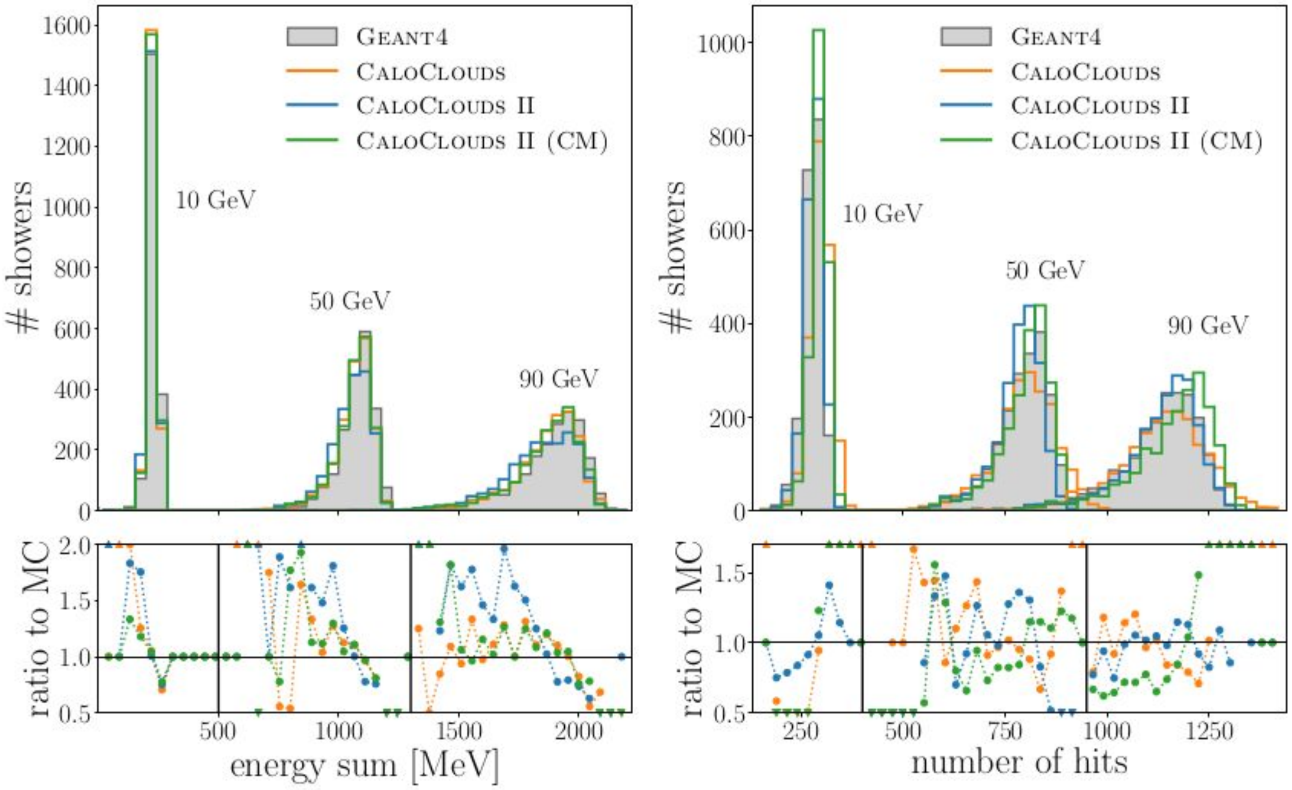
Center of Gravity

CaloClouds2



Visible Energy and Occupancy

CaloClouds2

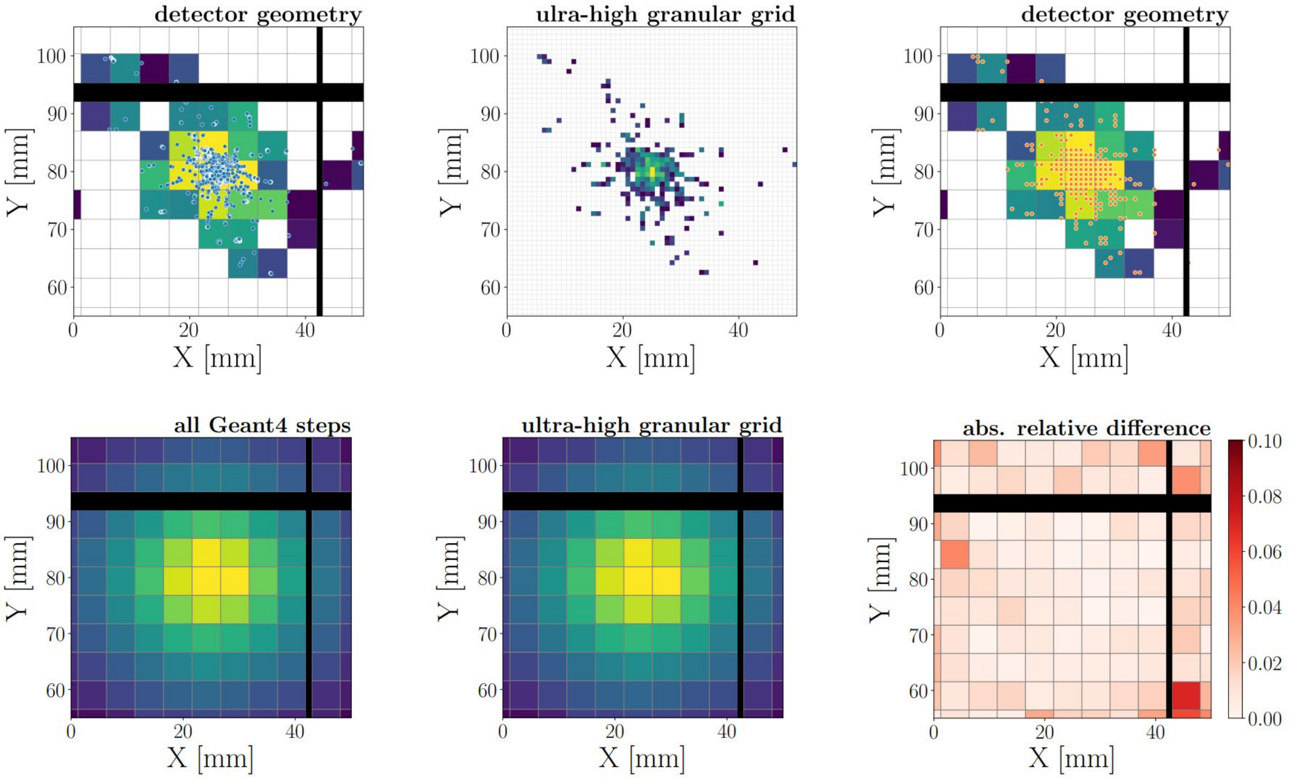


Point Cloud Representation of the EM Showers

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