Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

DESY.

CaloHadronicVAE : Fast Point Cloud Generation via Latent Diffusion

Martina Mozzanica
Phd student at the 3rd year
Universität Hamburg
martina.mozzanica@uni-hamburg.de

The Need for Fast Simulation

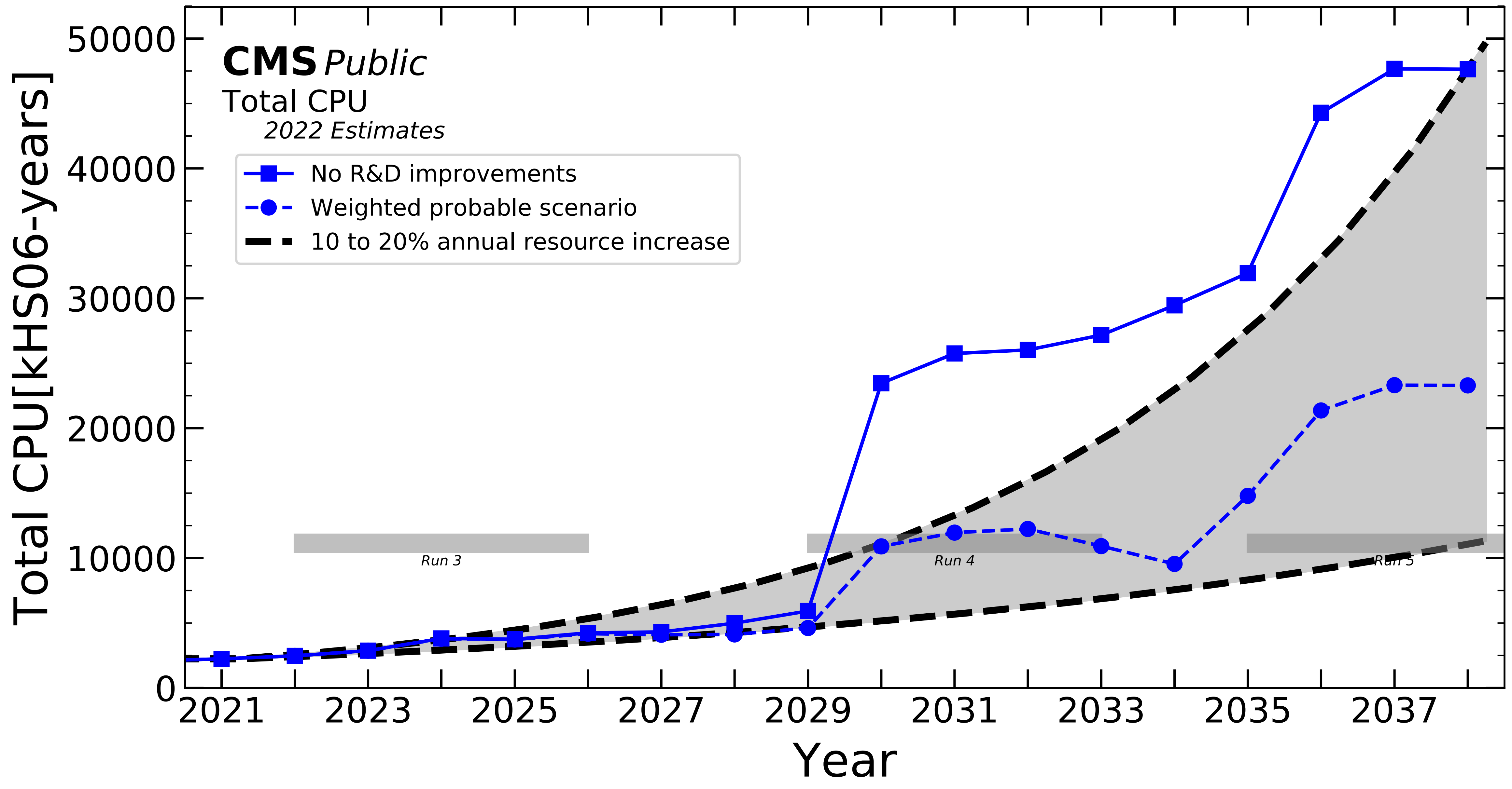
- **Simulation is fundamental** to compare theory to measurements

- Simulations will have to match the statistics of collected data x10

- Computational requirements expected to **exceed available resources** soon

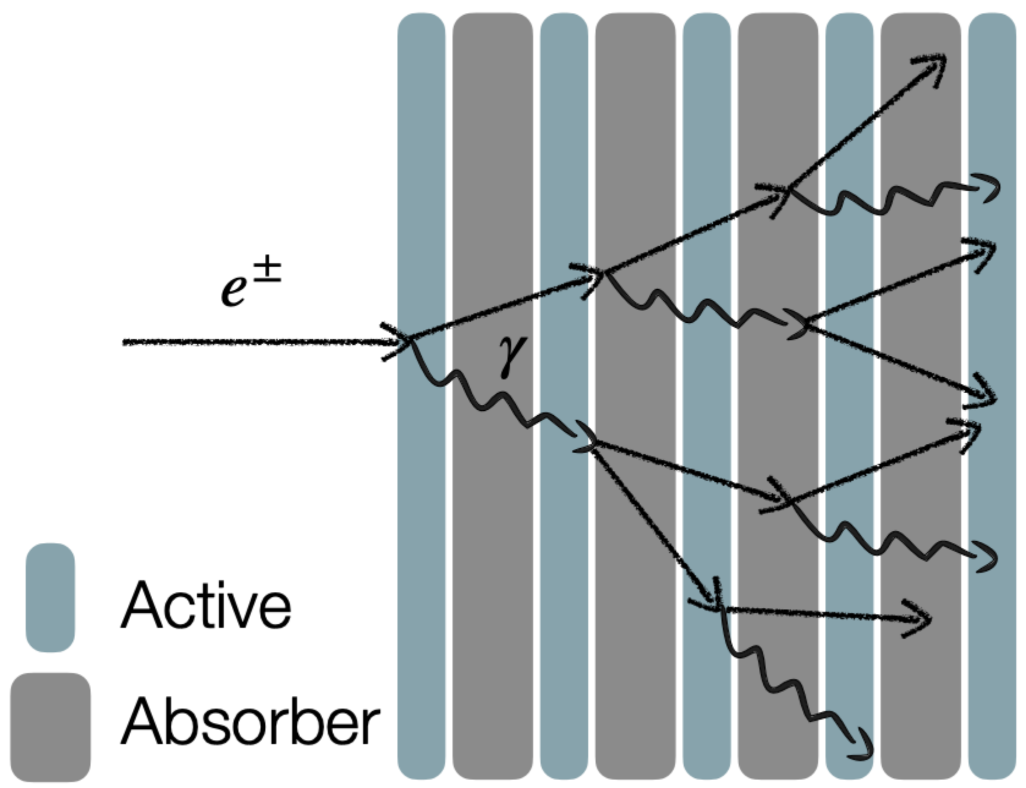
- **Detector simulation** is the most expensive block of the chain

- Without R&D, increase in computing time exceeds the one in resources



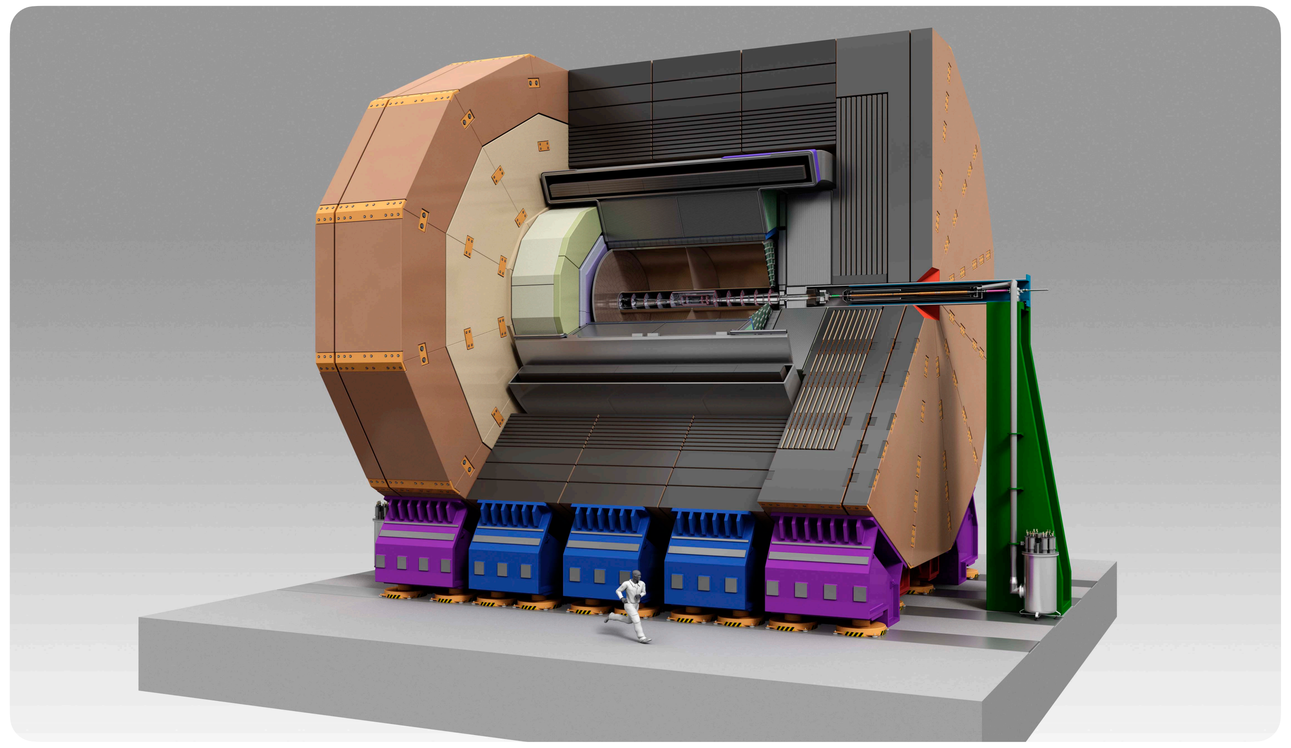
CMS Collaboration, Offline and Computing Public Results (2021)

Calorimeters and ILD



Generative Models for Particle Physics Data

- A calorimeter is a detector which fully absorbs the particles
- The signals produced are a measure for the energy of the particle

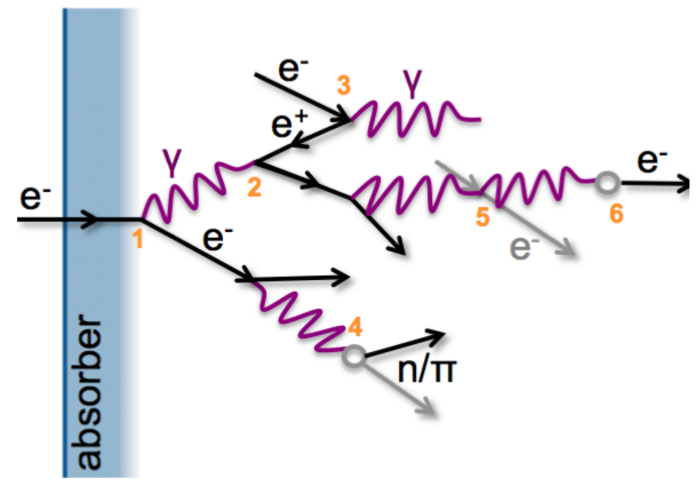

The International Linear Collider: A Global Project

- International Large Detector (ILD) concept for the International Linear Collider (ILC)
- High-Granularity sampling calorimeters:
  - ECal: Si-W - 5mm x 5mm - 30 layers
  - HCal: Sci-Fe - 30mm x 30mm - 48 layers

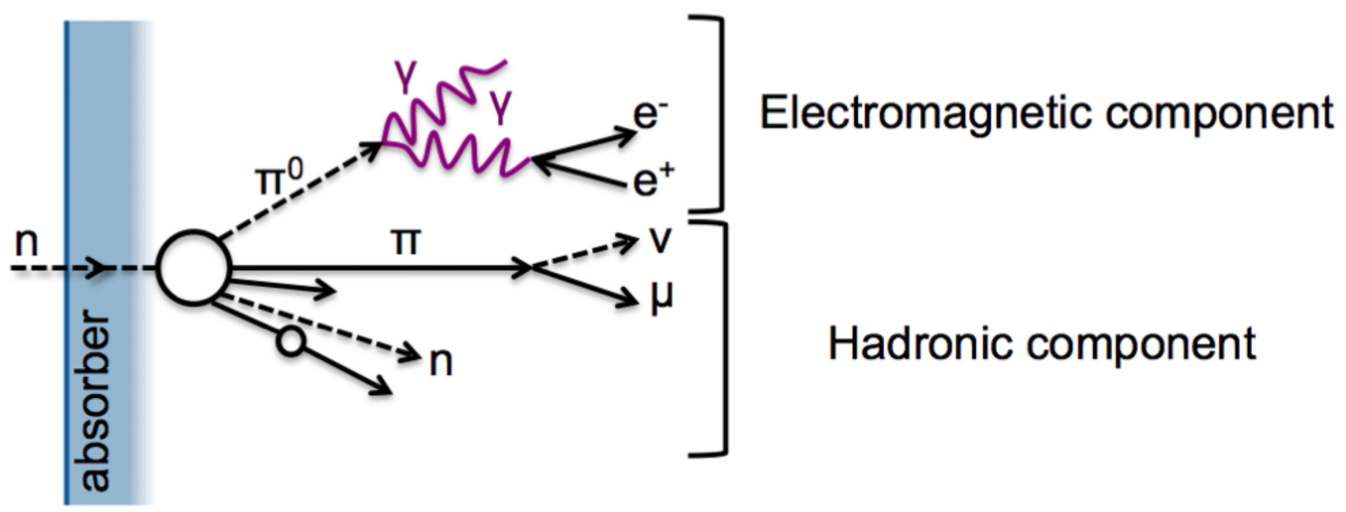
High granularity calorimeter ⟶ High fidelity simulation

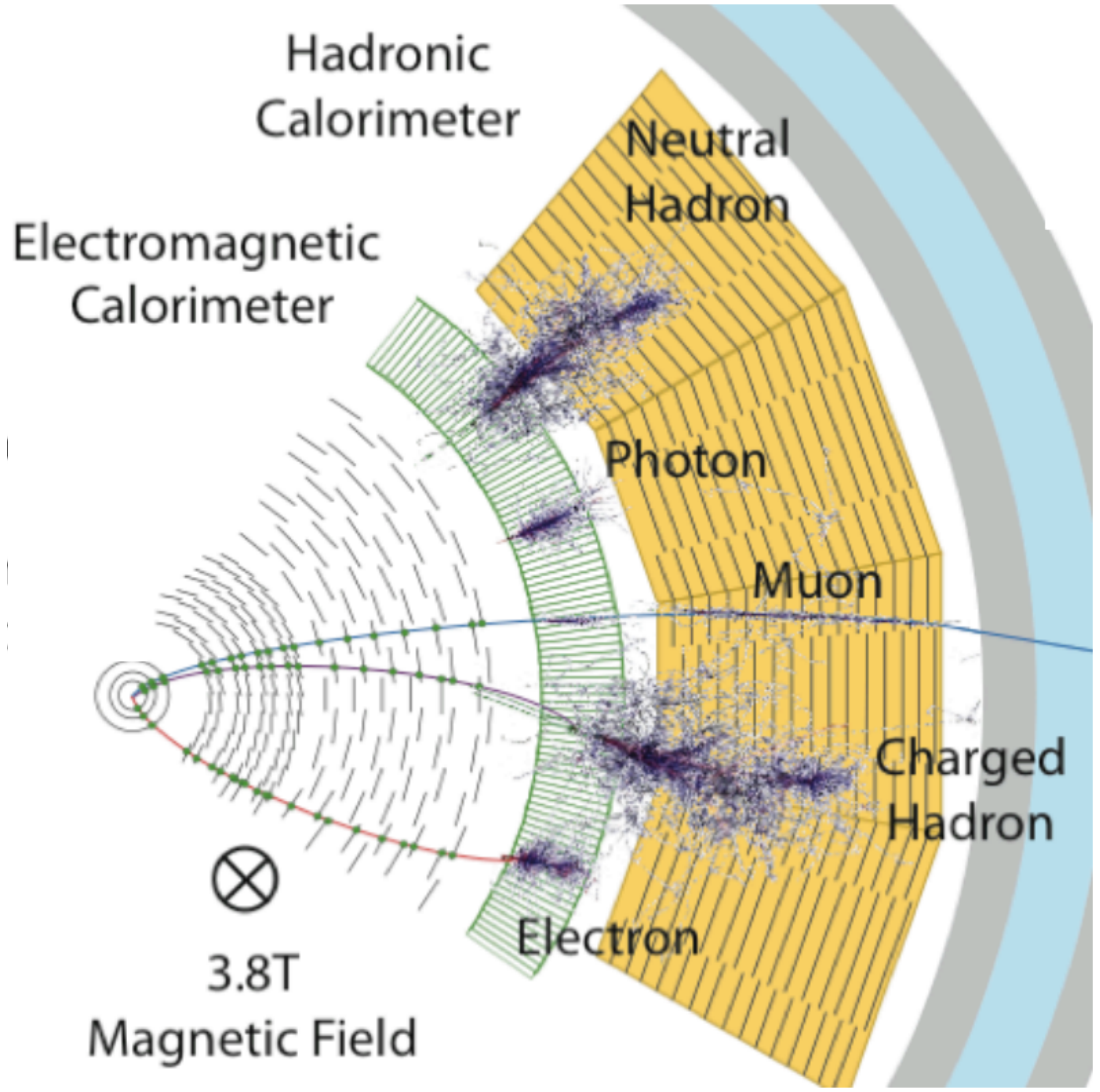
Electromagnetic and hadronic showers

Electromagnetic showers: mostly Bremsstrahlung and pair-production



Hadronic Showers: complex, non-perturbative phenomenology



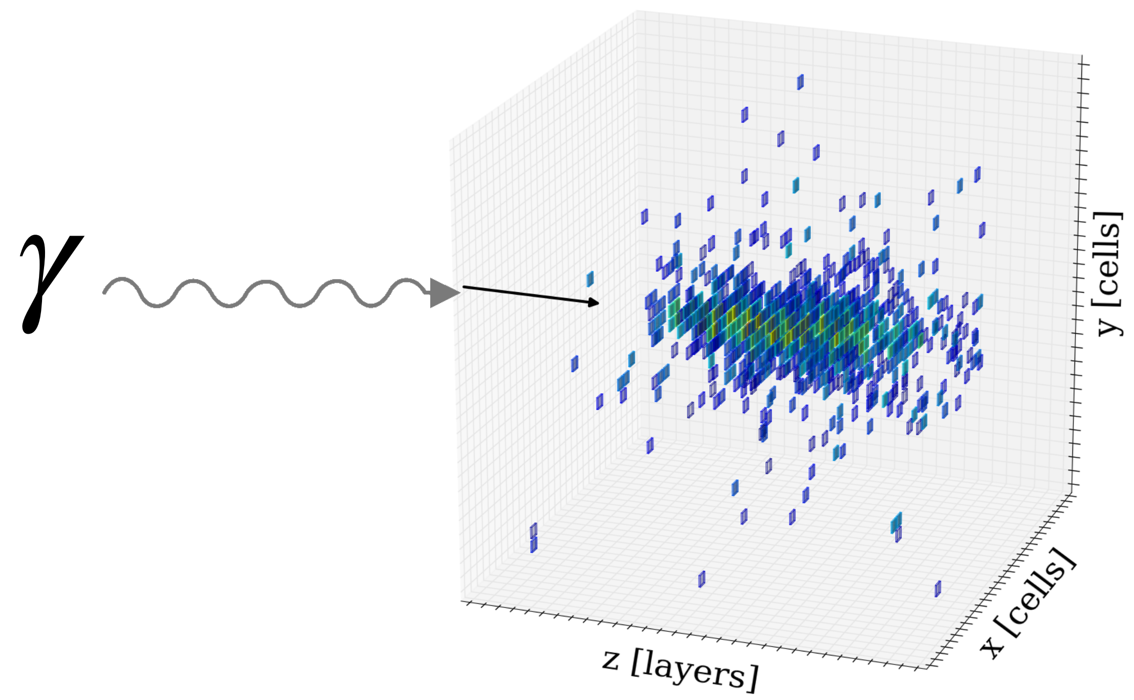


Instrumentation in HEP Calorimetry

Data Representation of Showers

Fixed Grid

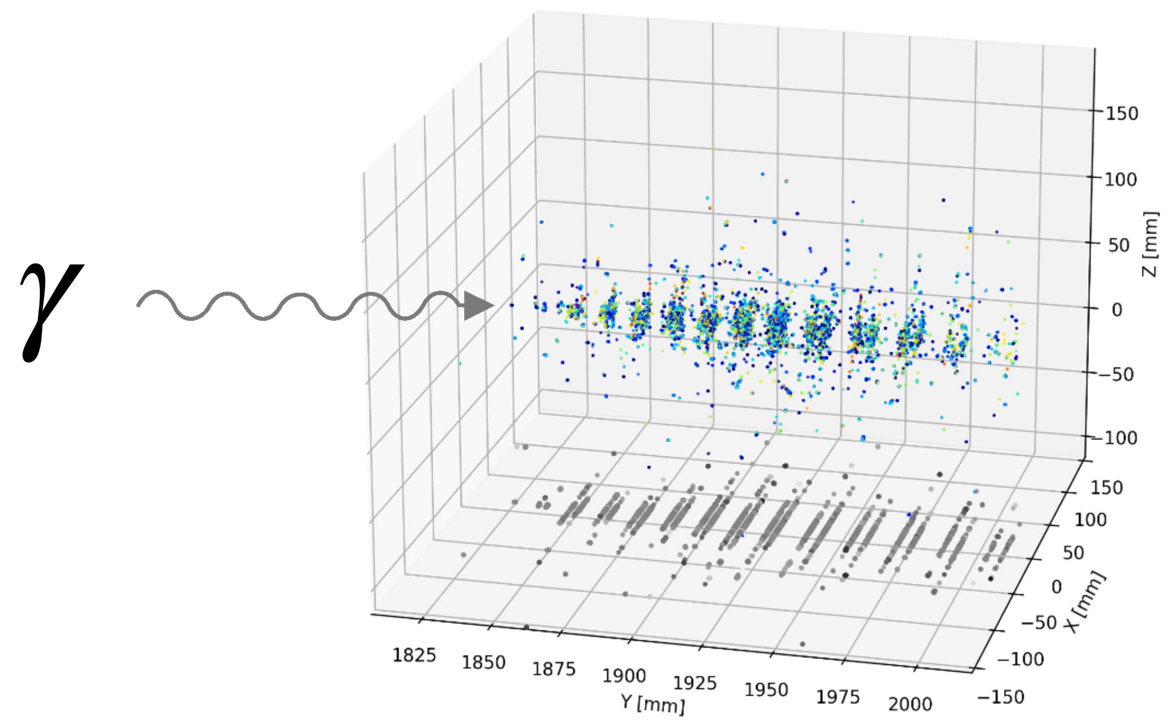
- 3D array filled with energy values
- entries correspond to calorimeter cells
- only c.a. 4% of cells are non-zero



Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed

Point Clouds

- variable-length, permutation-invariant sets
- more economically represented
- only generation of non-zero points

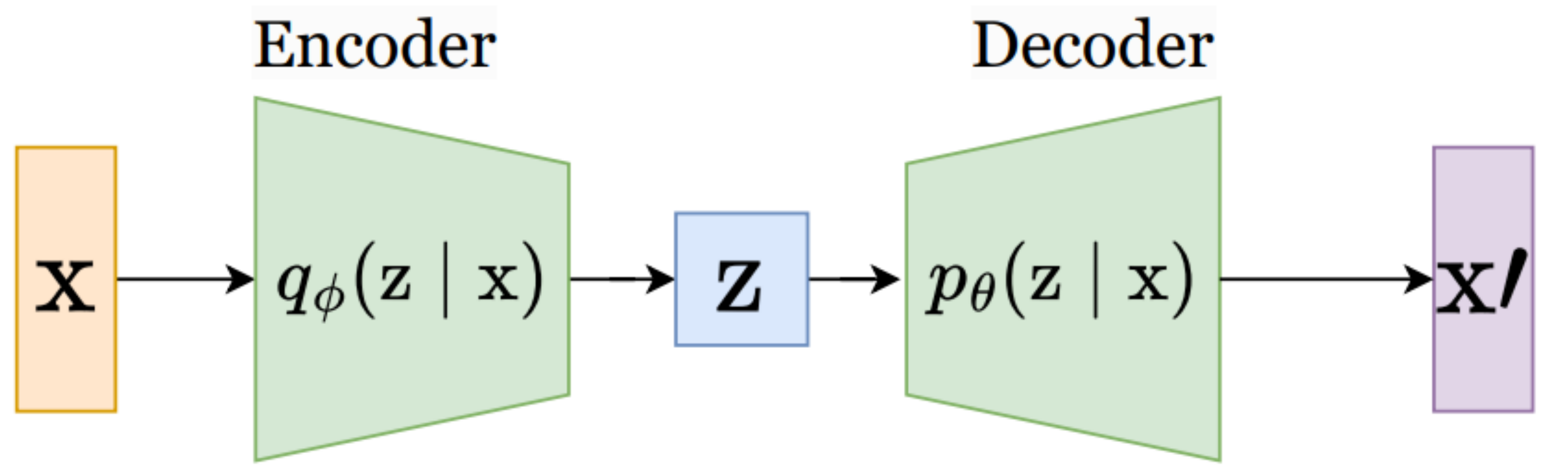


CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation

Multiple points per cell & cell-geometry independence

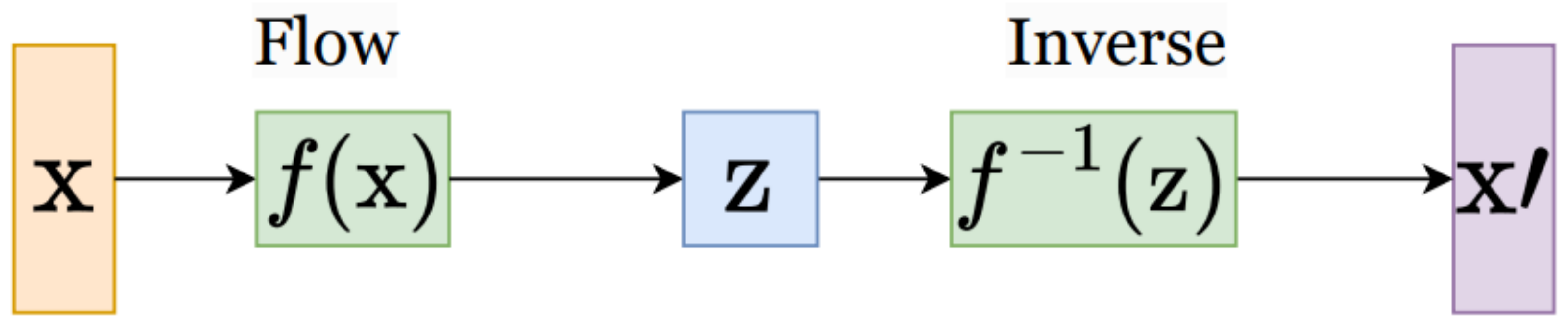
Some of the Best Generative Methods for Fast Sim

Variational AutoEncoders



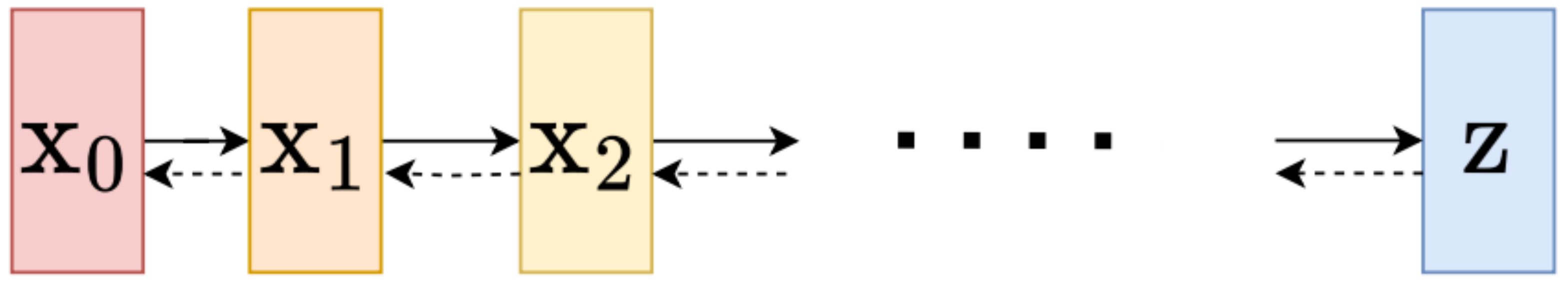
VAEs learn a compressed representation of the data to generate new samples

Normalizing Flows



Flow-based models directly transform the data distribution

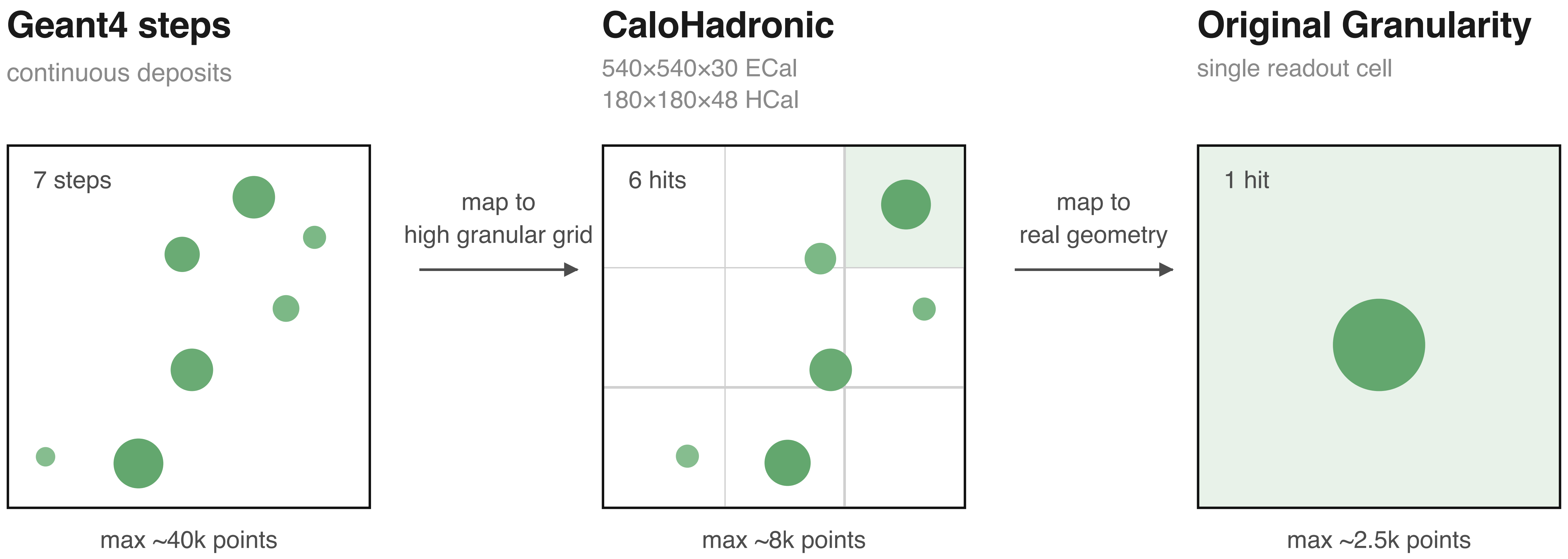
Diffusion Models

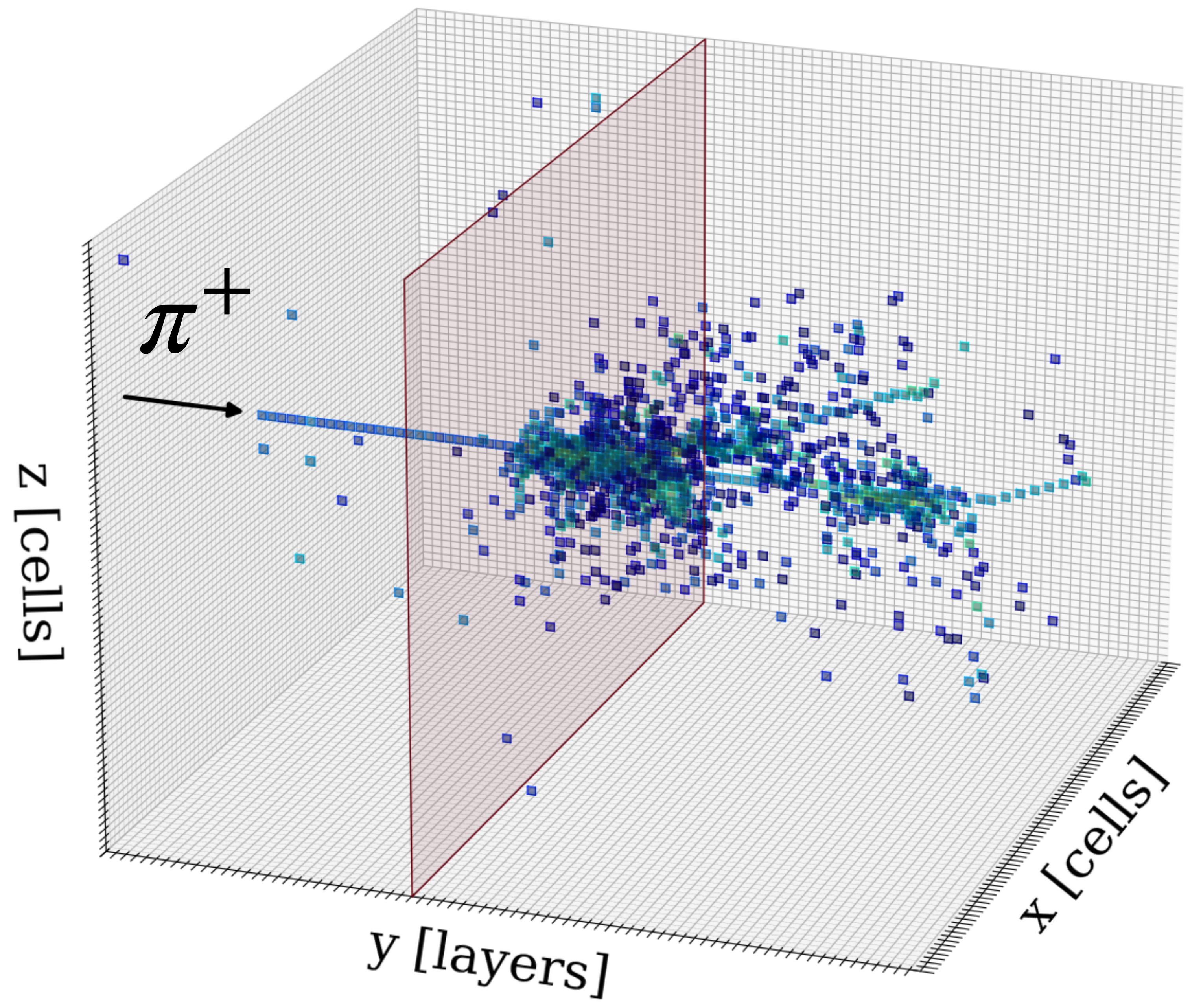


Diffusion models gradually add noise then reverse the process to generate data

Dataset

- Only **hadronic showers** are considered (the harder ones to model)
- **10-90 GeV π^+ showers** with direction perpendicular to the face of the ECal
- The gun positioned at the face of ECal
- The magnetic field is switched off
- **Time feature included** - first time for high granular calorimeters
- **x9** higher granularity

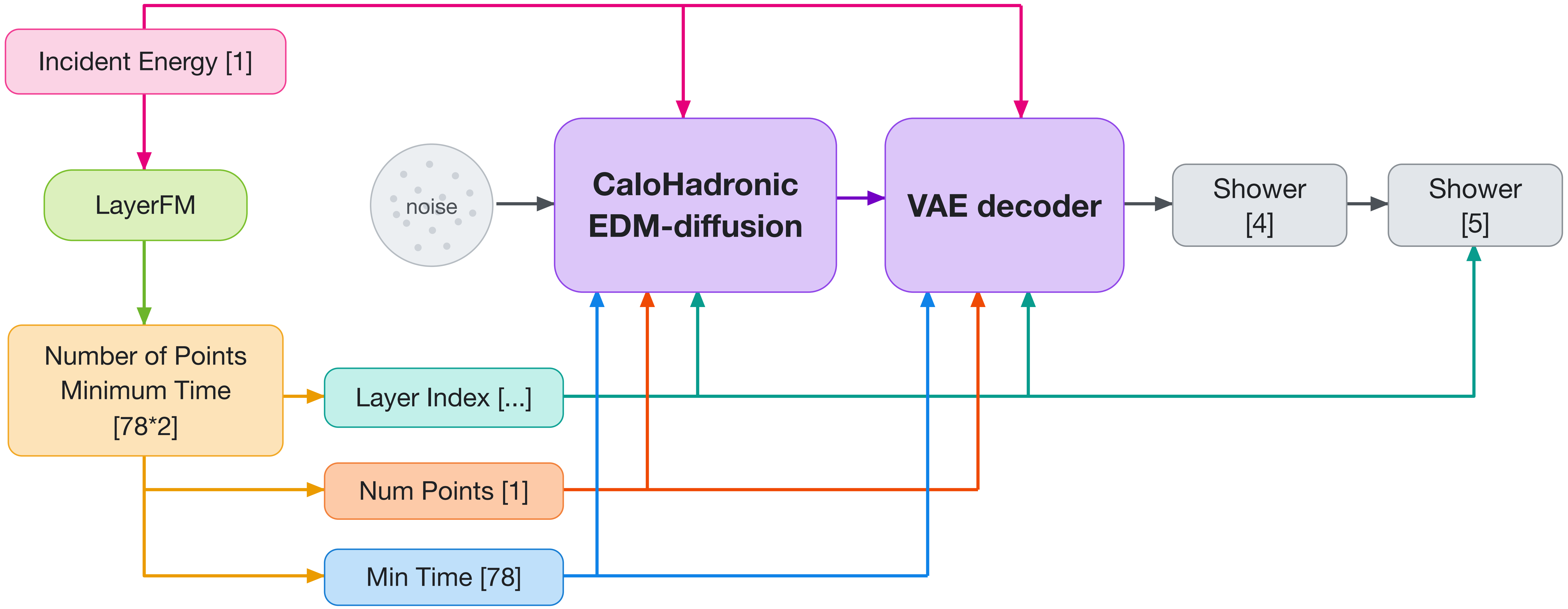




CaloHadronic: a diffusion model for the generation of hadronic showers

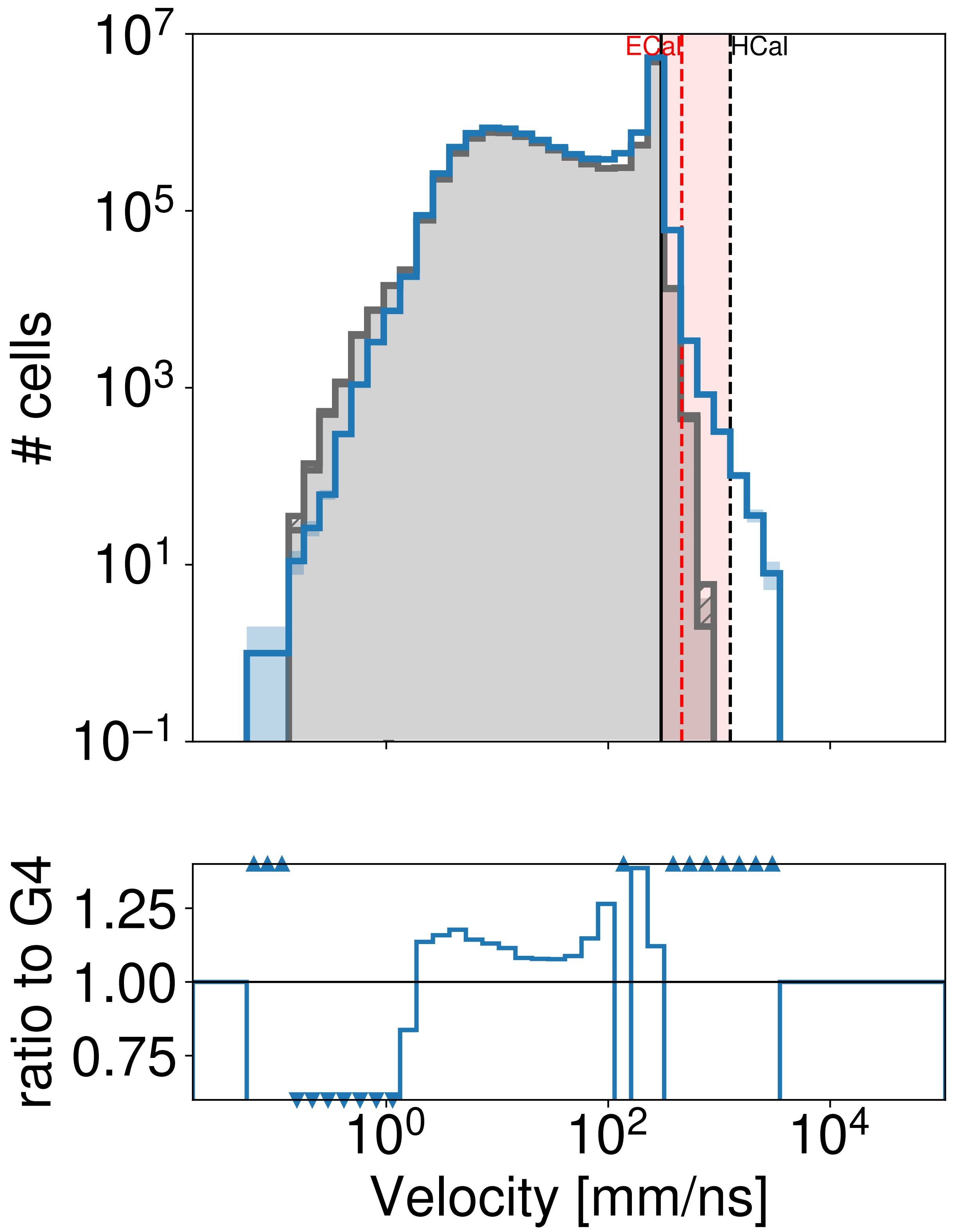
CaloHadronicVAE

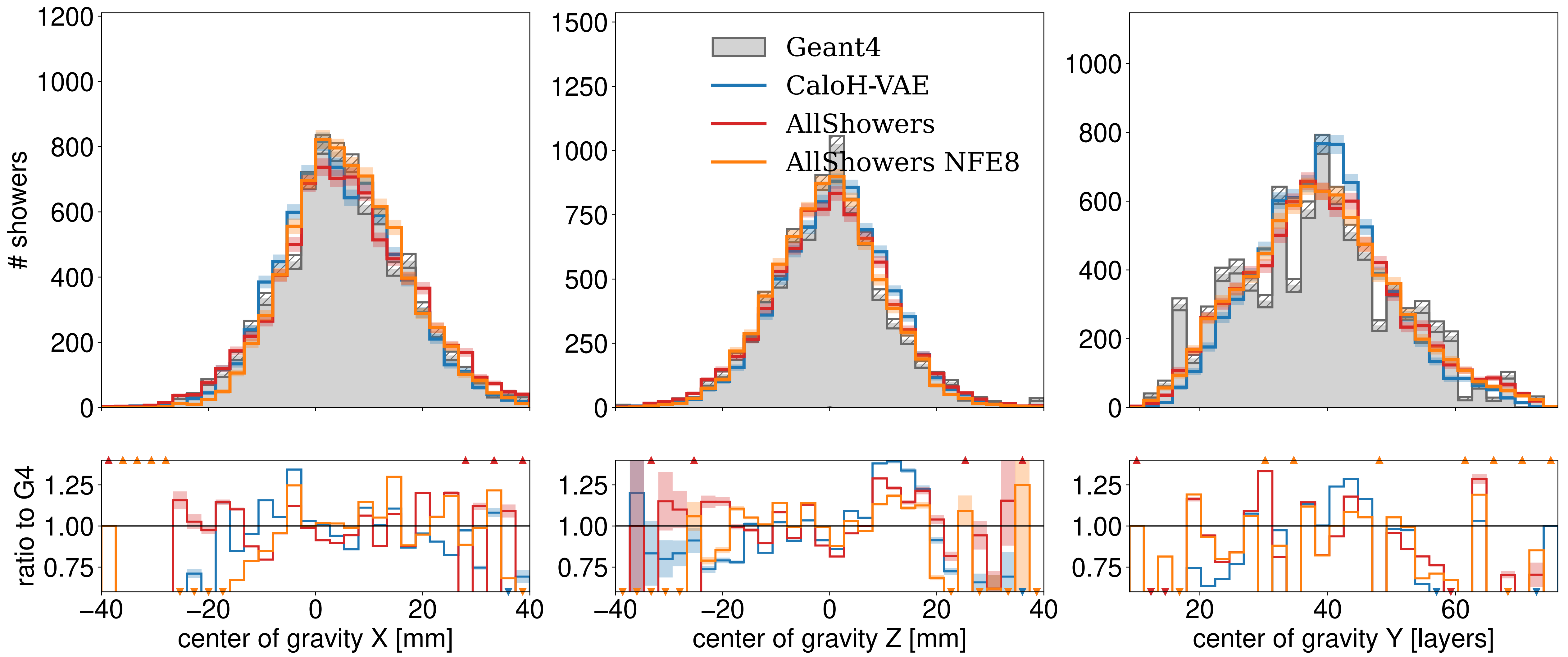
- The model comprises three components: a continuous normalising flow + a VAE + a diffusion model.
- It generates 5D point clouds [x, y, z, energy, time], conditioned on the incident energy.
- The flow generated layer-wise quantities: the number of points per layer (split into total points and layer index) and the minimum time per layer.



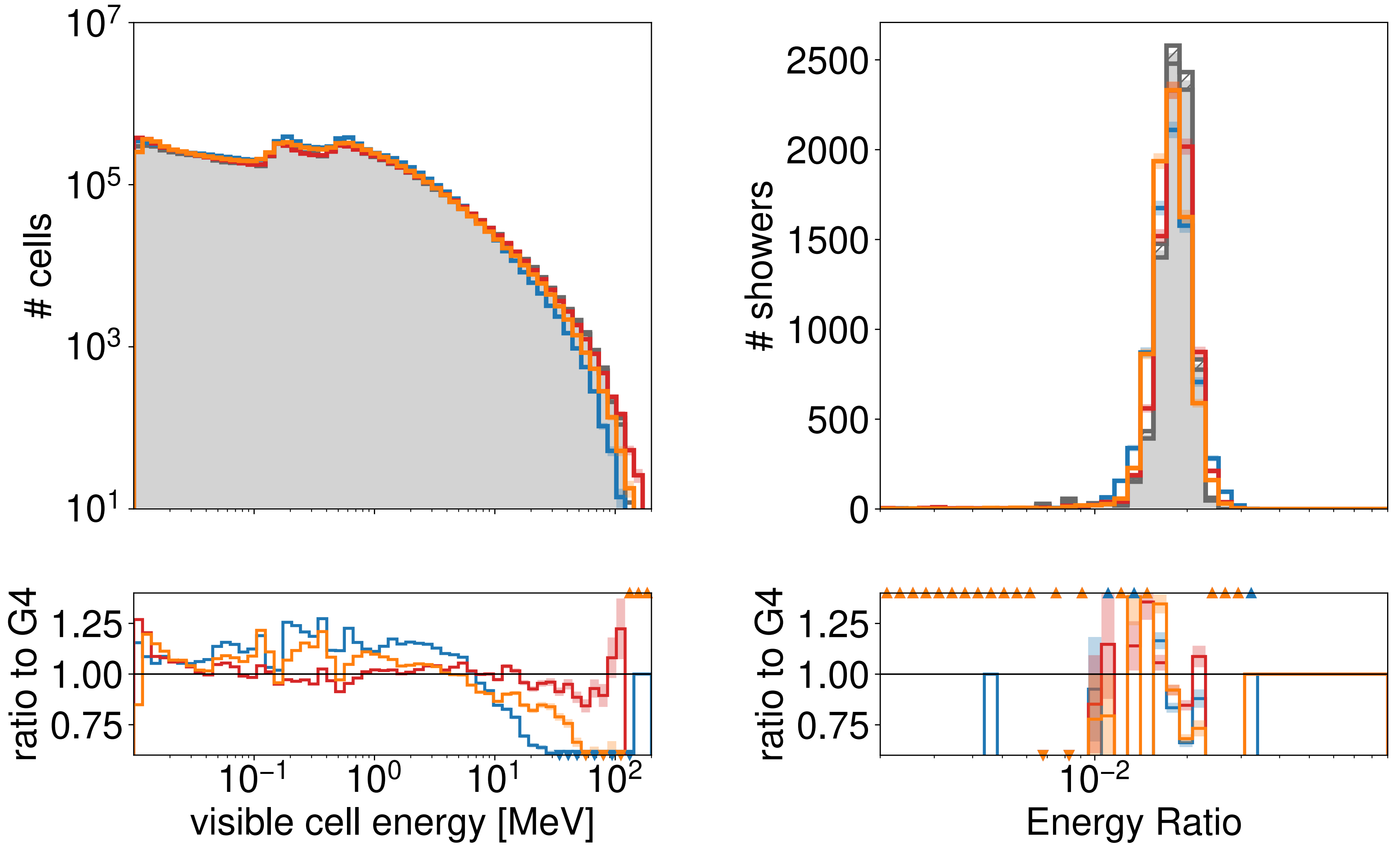
Results

8k, 50 GeV, π^+ showers



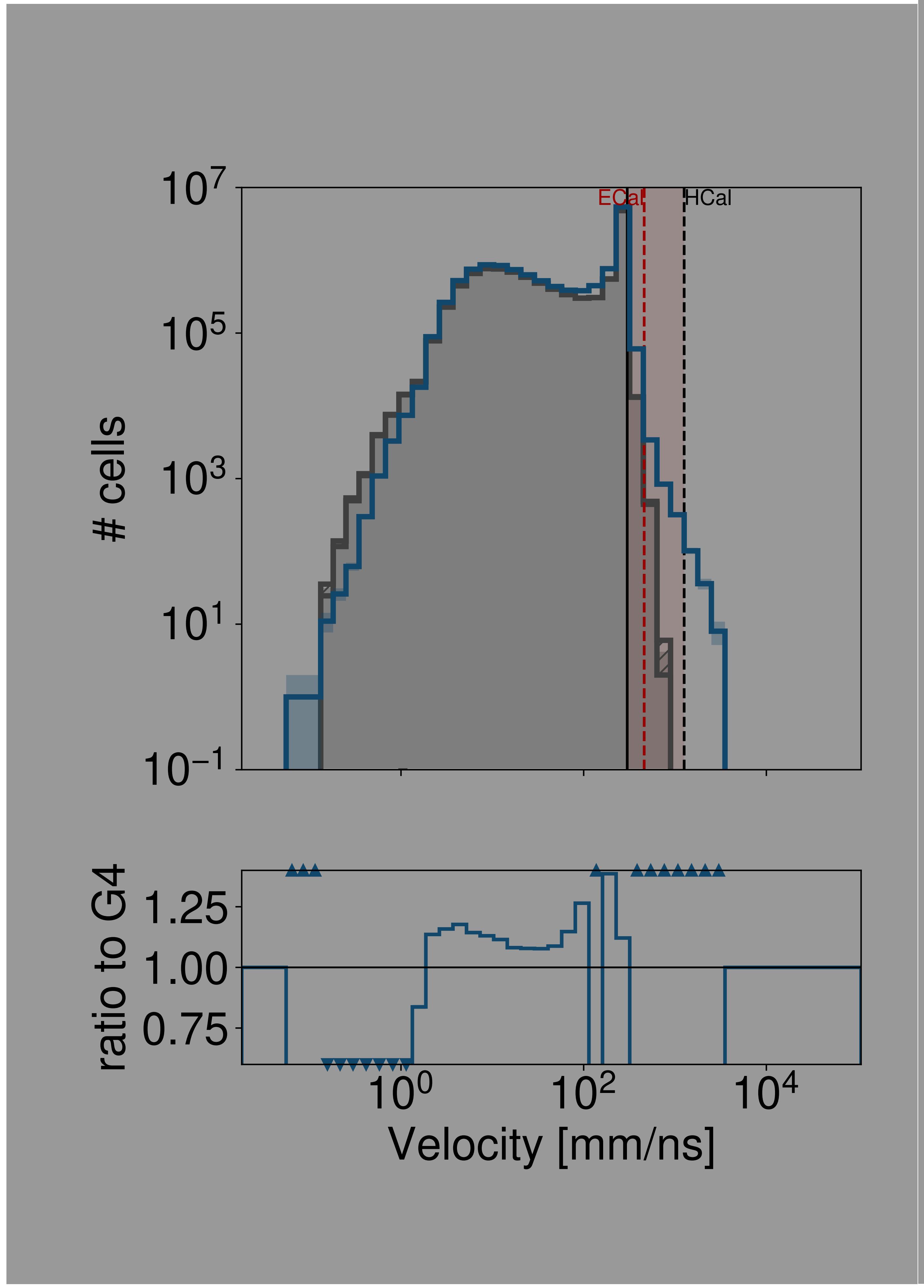


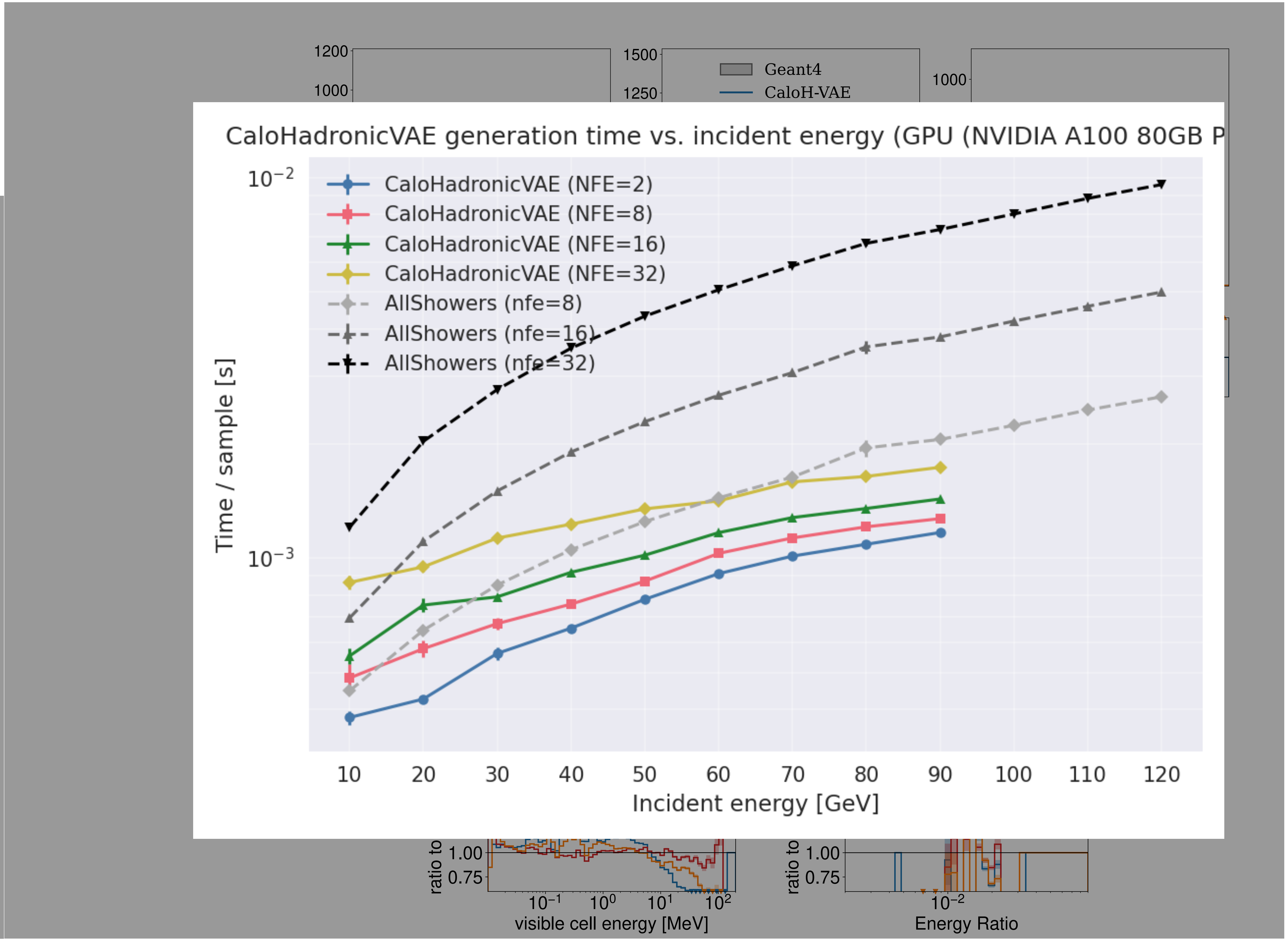
AllShowers: One model for all calorimeter showers



Results

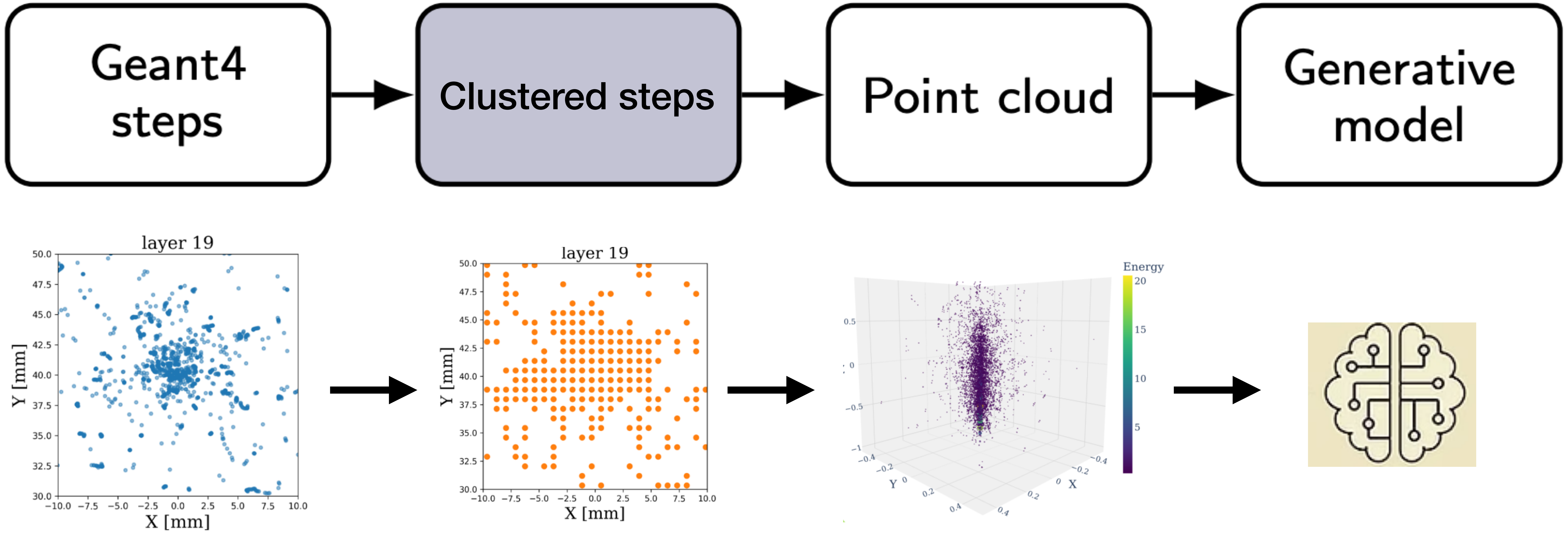
8k, 50 GeV, π^+ showers





Bonus: a library for point clouds clustering

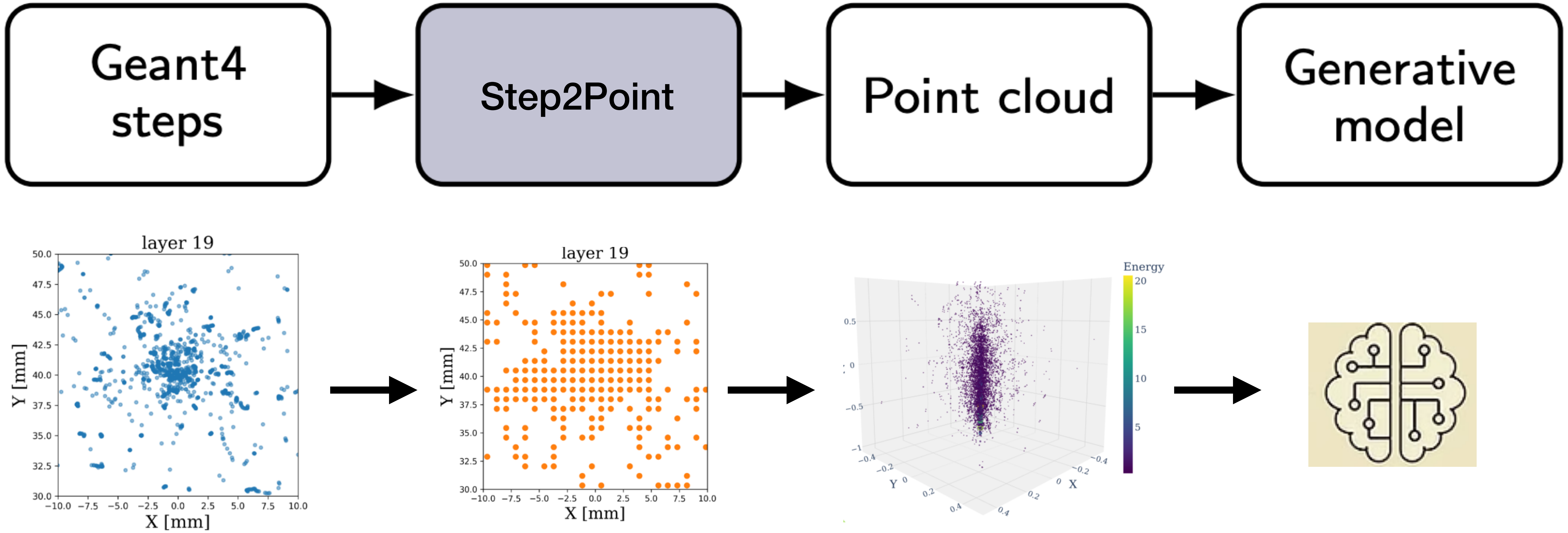
- Point clouds need to be trained on a **higher resolution** than calorimeter cells to avoid grid artifacts.
- In ml-based fast simulation, point cloud models revealed to be the fastest.



EP-SFT
Software Frameworks and Tools

Bonus: a library for point clouds clustering

- Point clouds need to be trained on a **higher resolution** than calorimeter cells to avoid grid artifacts.
- In ml-based fast simulation, point cloud models revealed to be the fastest.



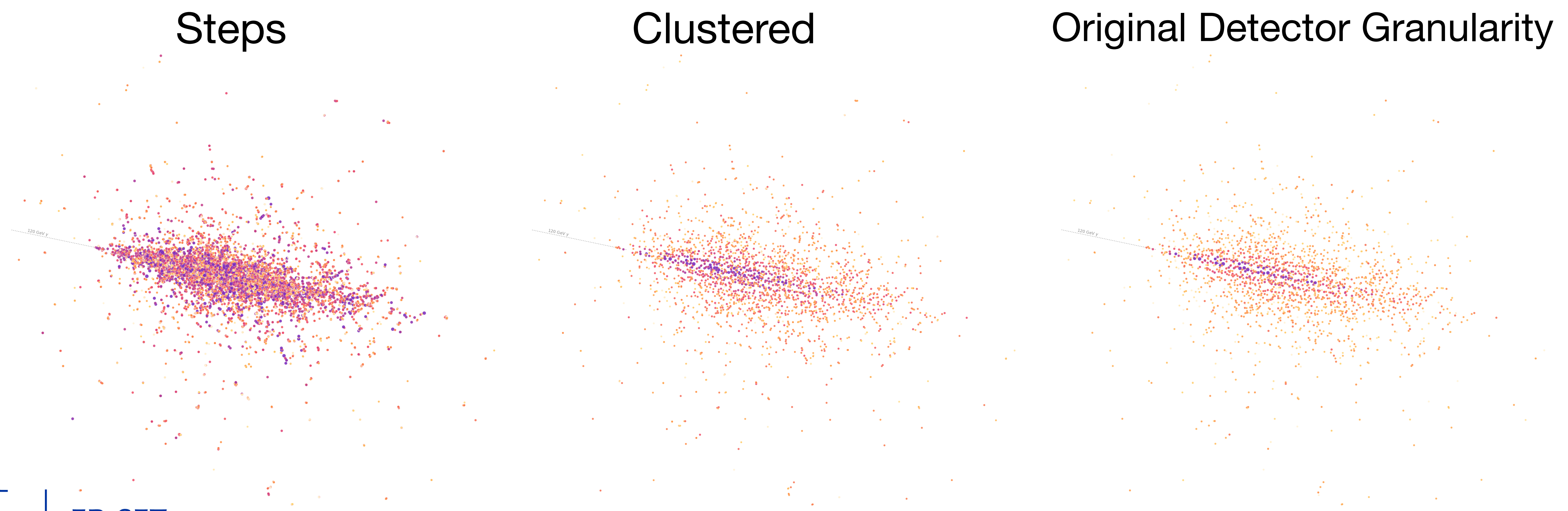
EP-SFT
Software Frameworks and Tools

What is Step2Point?

Goal 1: optimal point cloud representation to find the smallest set of points that preserves the physics of the shower.

Goal 2: we aim to provide infrastructure that should/could work with different detectors.

A library for turning detailed calorimeter shower deposits into **compact point-cloud representations** usable for ML-based fast simulation.



- GitHub repository here
- Webpage here
- CHEP 2026 here

EP-SFT
Software Frameworks and Tools

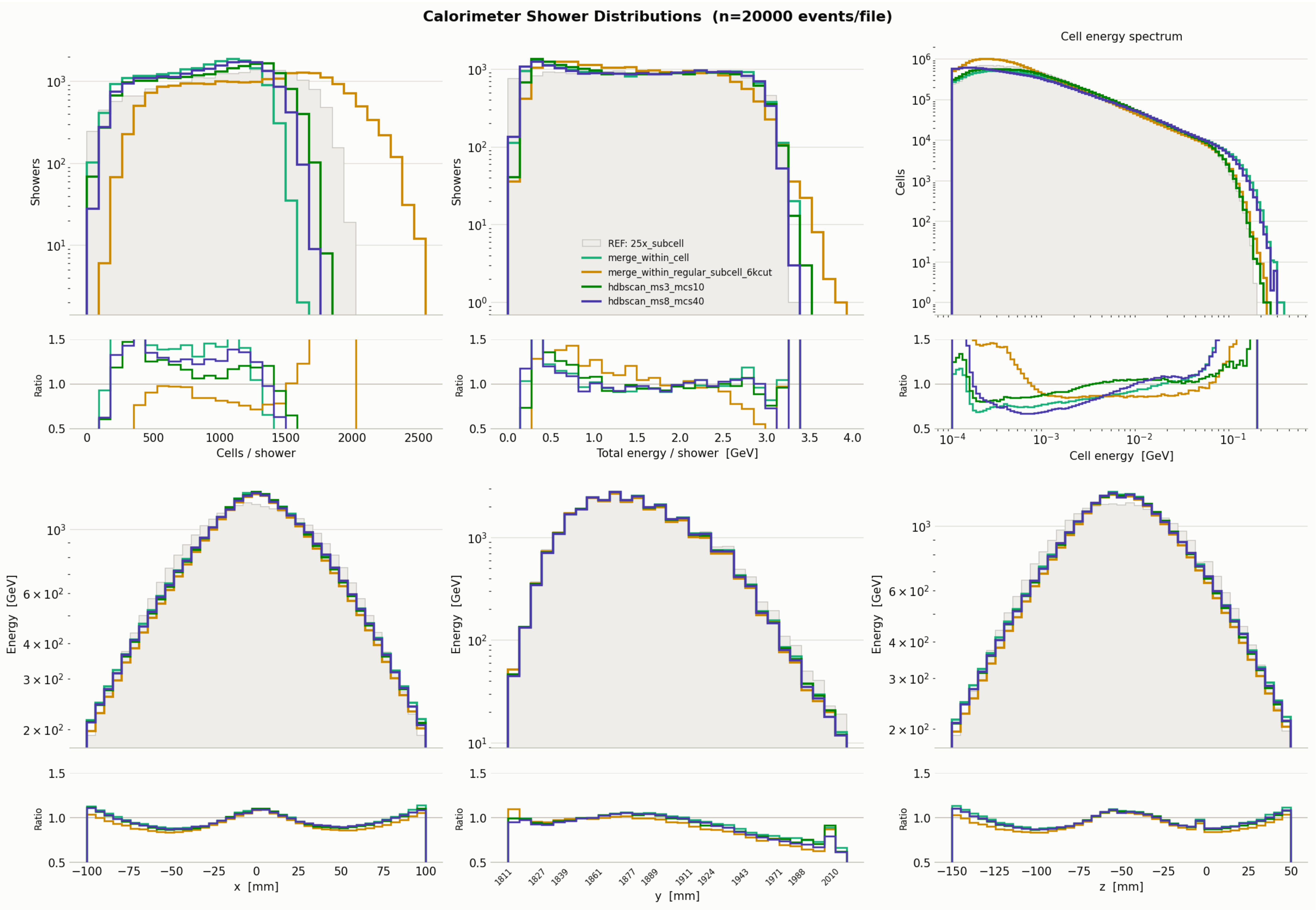
Results on photons in ILD with CaloClouds3

CaloClouds3: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation

- Higher-granularity point clouds enable more efficient training of generative models.

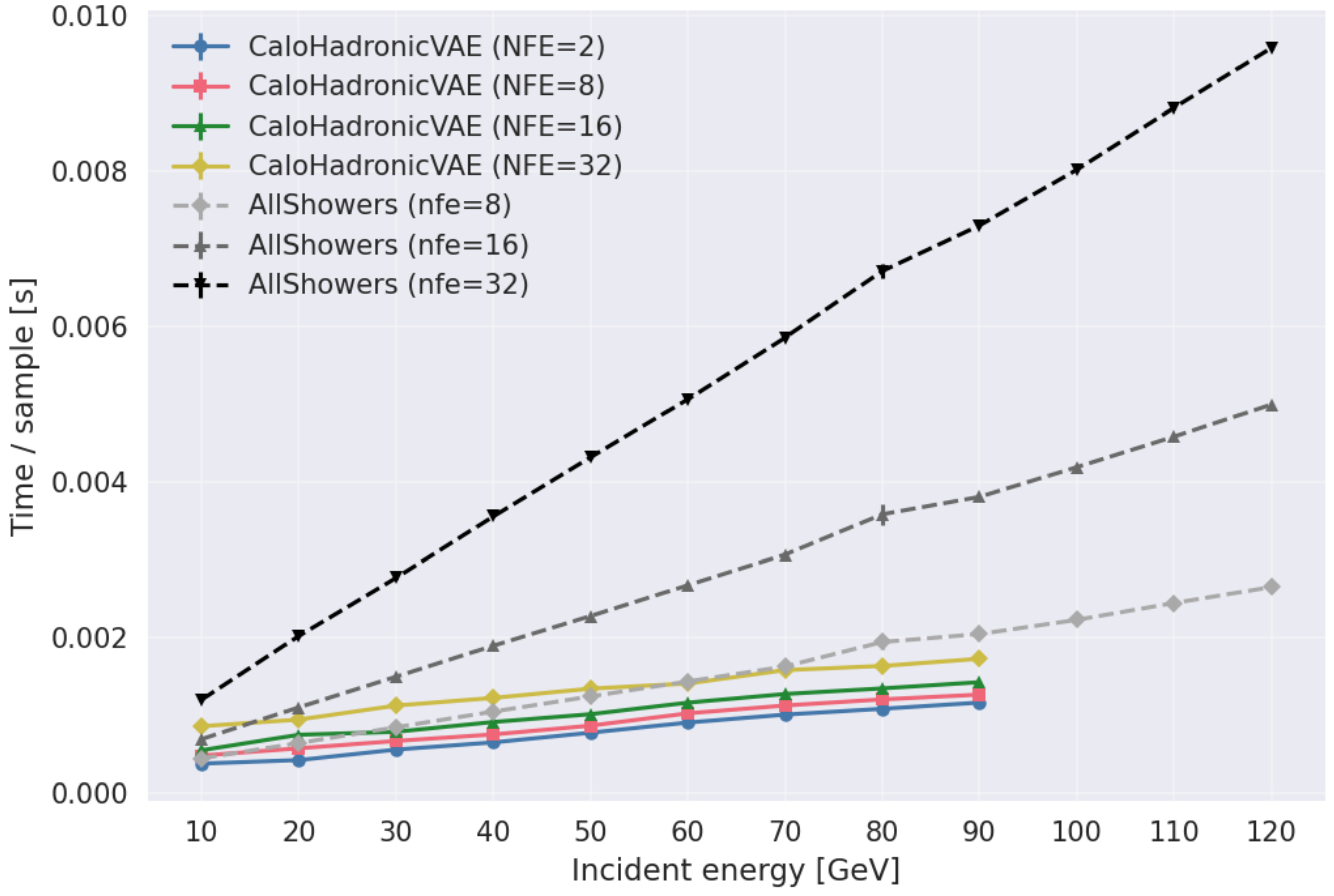
- HDBSCAN clusters points within a cell effectively — independent of cell geometry.

Point Cloud level	Mean number of points	Max. number of points
Cell	1203	2257
Hdbscan ms3	1783	3692
Hdbscan ms8	2332	4854
Subcell	3627	5785

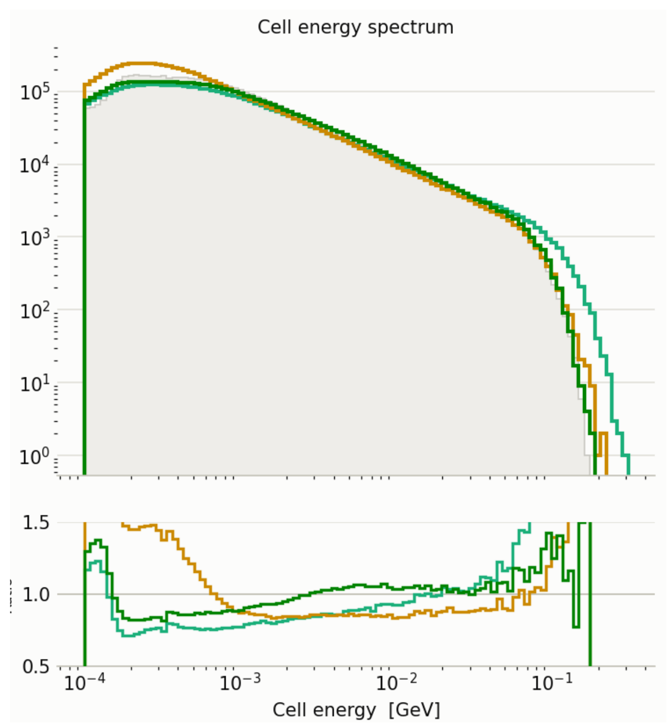


EP-SFT
Software Frameworks and Tools

Summary:



- Need of fast detector simulation
- **Point cloud** based model are the best solution for speed purposes
- **Latent diffusion** enables to have a compact representation of a shower
- Scales very good with the number of points/ incident energy
- Performances can be improved



- StepToPoint: answers the need for a library to convert voxel data to point clouds
- Hdbscan is a valid solution for all cell shapes detector