CaloHadronic : a diffusion model for the generation of hadronic showers

Thorsten Buss, Frank Gaede, Gregor Kasieczka, Anatolii Korol,
Katja Krüger, Peter McKeown and Martina Mozzanica

Paper: https://arxiv.org/pdf/2506.21720

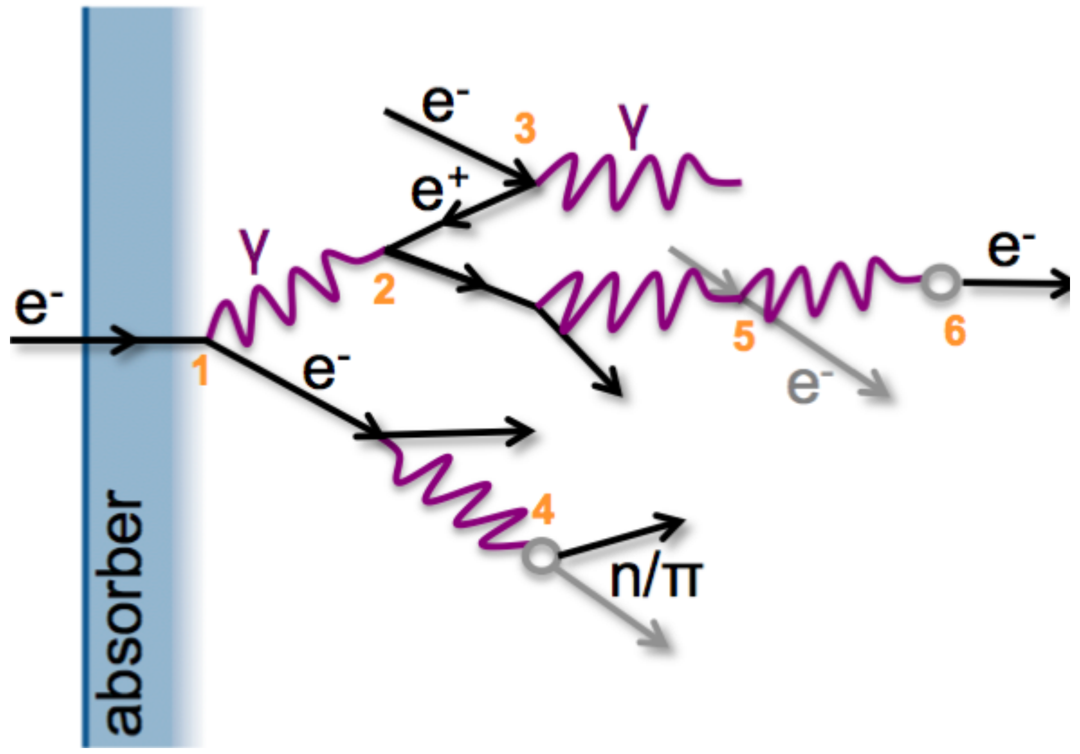
8th Round table on Deep Learning @ DESY

martina.mozzanica@uni-hamburg.de

Generative model for em and hadronic showers

CaloClouds II [1] ⟶ Fast & accurate point clouds based generative model applied **only to electromagnetic showers**

[1] CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation
E. Buhmann et al: arxiv: 2309.05704



Now

↓

HADRONIC showers

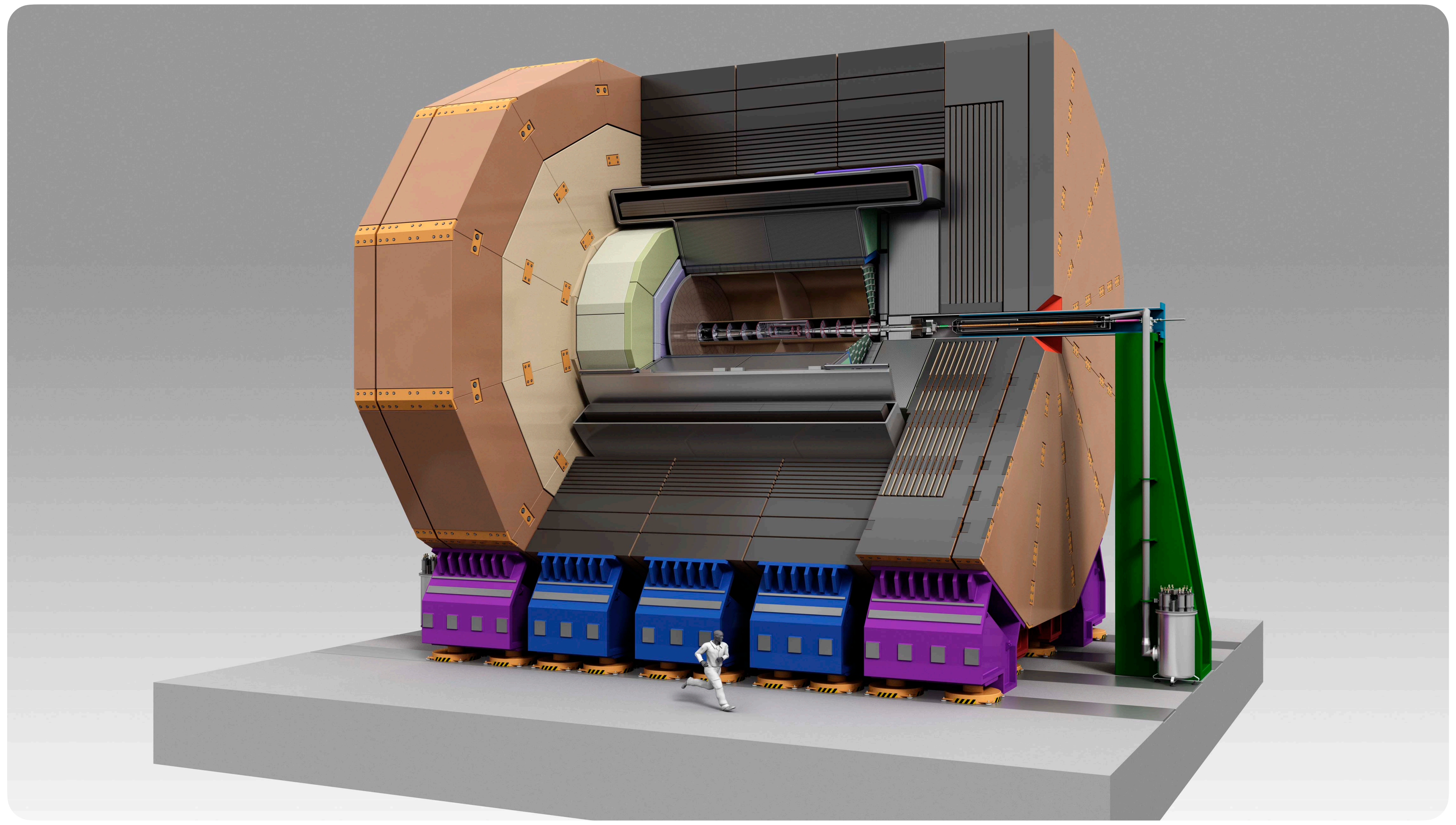
Dealing with 10-90 GeV π^+ showers
two components: hadronic and electromagnetic

Hadron showers have **much larger:**

- **spatial extension** -> interaction length larger than radiation length
- **lateral dispersion** -> large transverse energy transfers in nuclear reactions

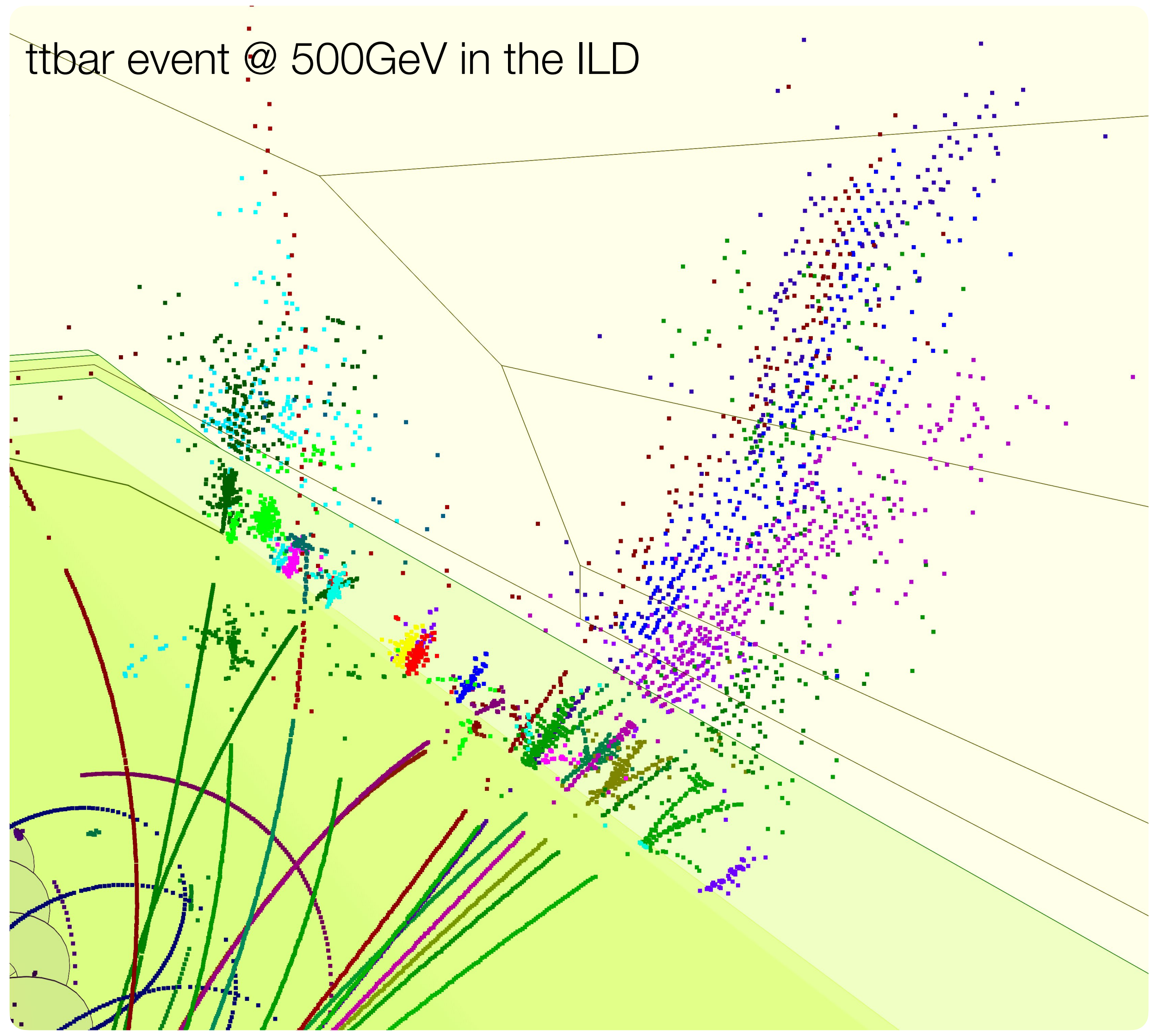
ILD calorimeter

The International Linear Collider: A Global Project [arxiv: 1903.01629]



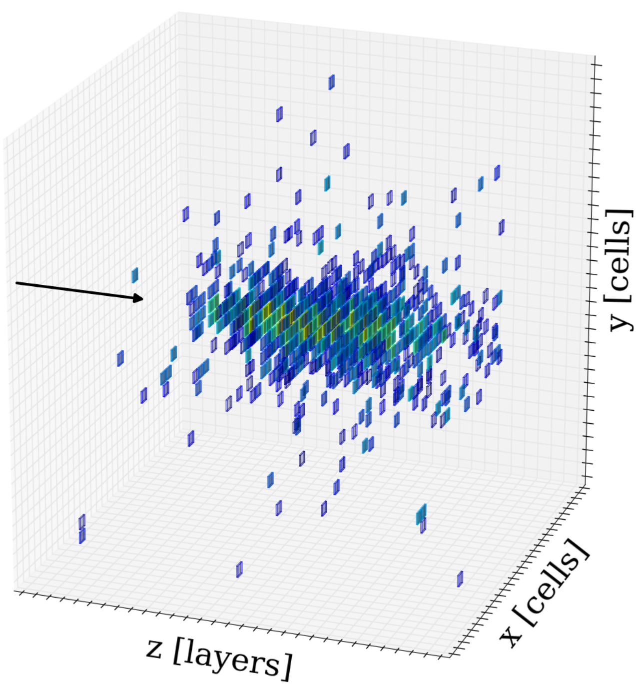
- International Large Detector (ILD) concept for the International Linear Collider (ILC)
- High-Granularity calorimeters:
  - ECal: Si-W - 5mm x 5mm - 30 layers
  - HCal: Sci-Fe - 30mm x 30mm - 48 layers

High granularity calorimeter ⟶ High fidelity simulation

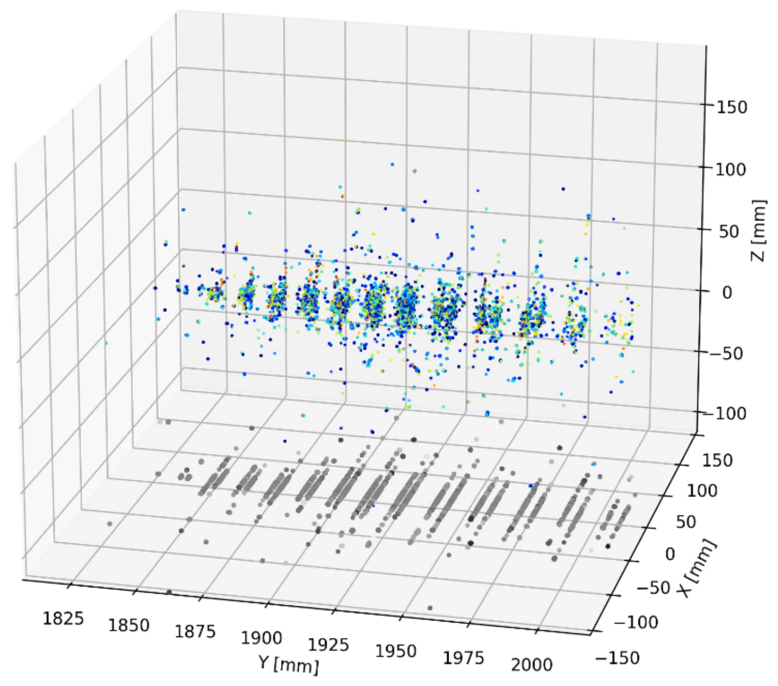


The International Linear Collider: A Global Project [arxiv: 1903.01629]

EM shower as 3D image



EM shower as point cloud



point clouds

Multiple points per cell & cell-geometry independence

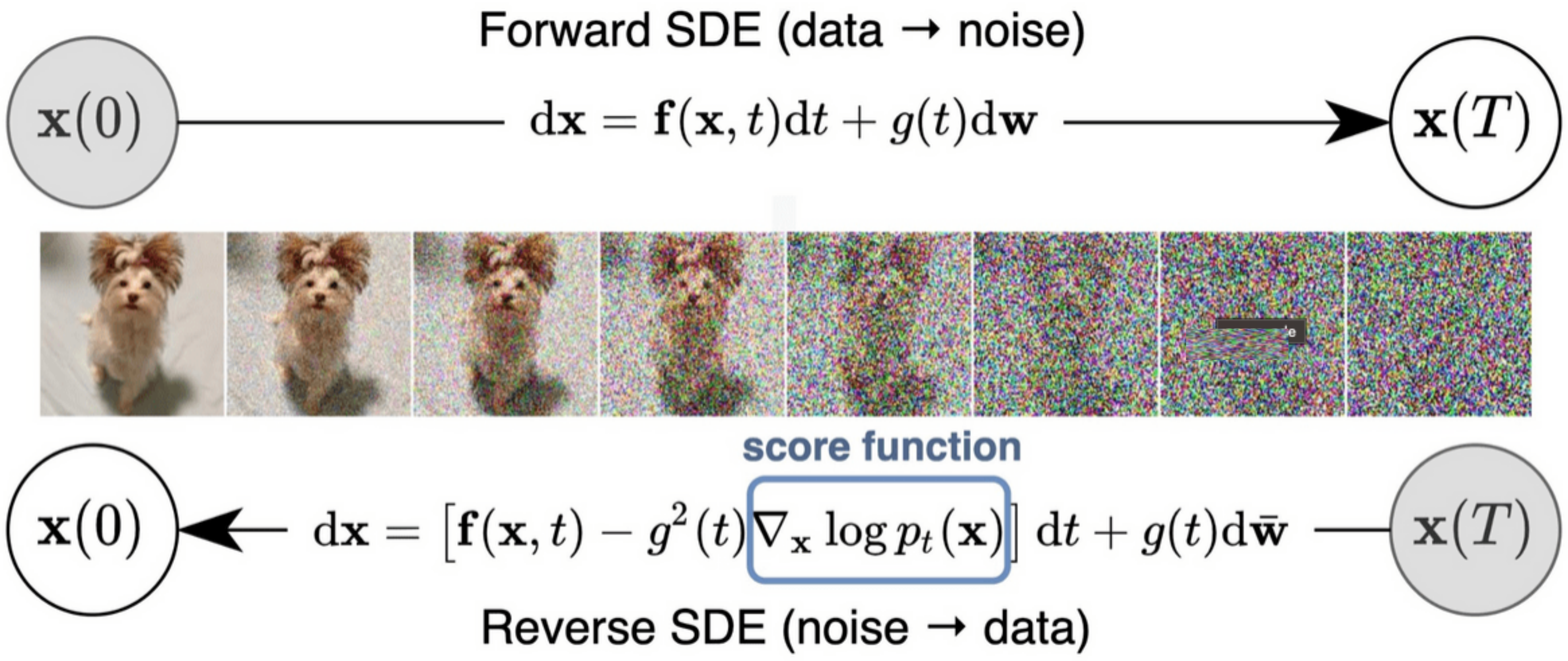
Score-based continuous-time diffusion [2]

Continuous Time:

- this process is modeled by a prescribed stochastic differential equation (SDE)

Score-based:

- score-matching: modelling the gradient of the log probability density function
- langevin dynamics: iterative process that draws samples from a distribution using only its score function



To compute the reverse SDE, we need to estimate the score function:

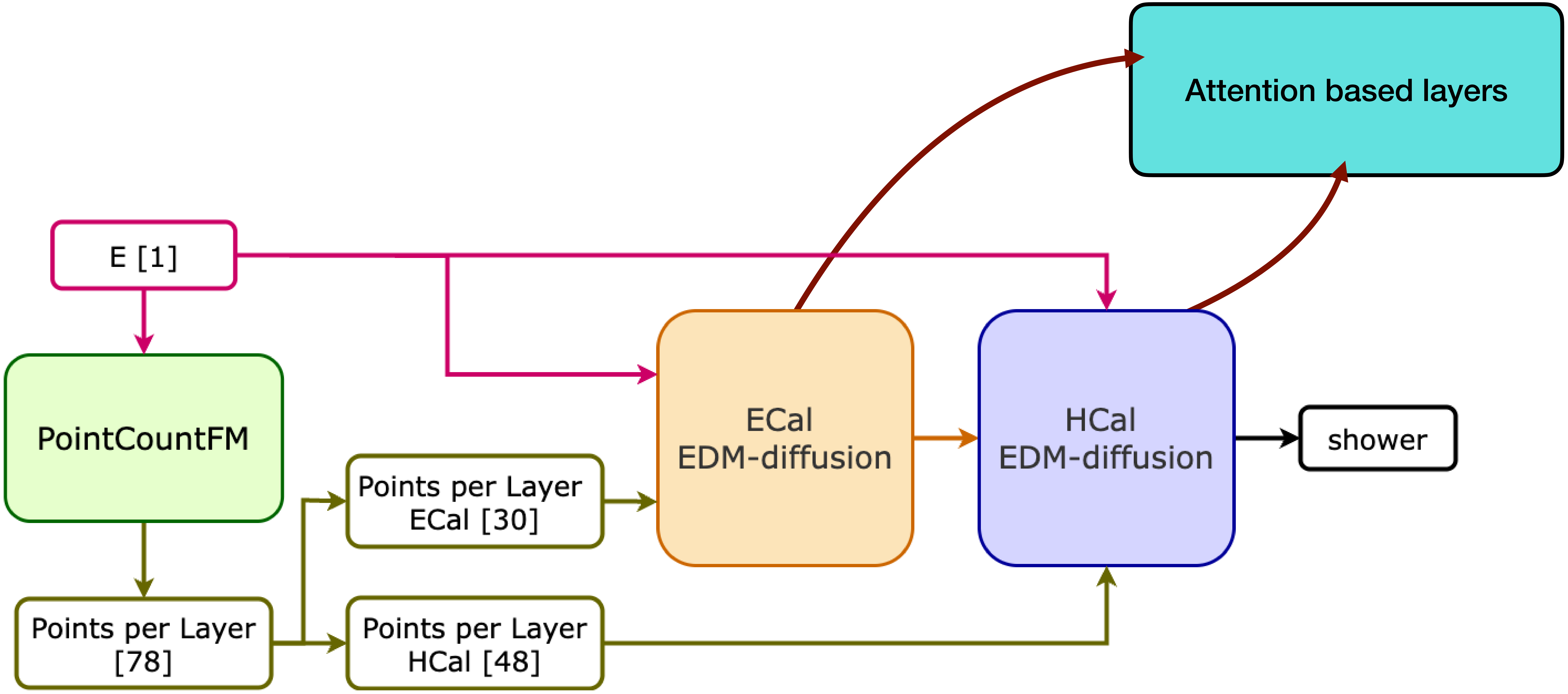
$$\mathbb{E}_{t \in \mathcal{U}(0,T)} \mathbb{E}_{p_t(\mathbf{x})} [\lambda(t) \| \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x}, t) \|_2^2]$$

score-based model

[2] **Score-based Continuous-time Discrete Diffusion Models**
Haoran Sun et al: arxiv: 2211.16750

CaloHadronic

- input: 4D point cloud
- conditioning on: incident energy (E) and number of points per layer
- calibrating on: points per layer

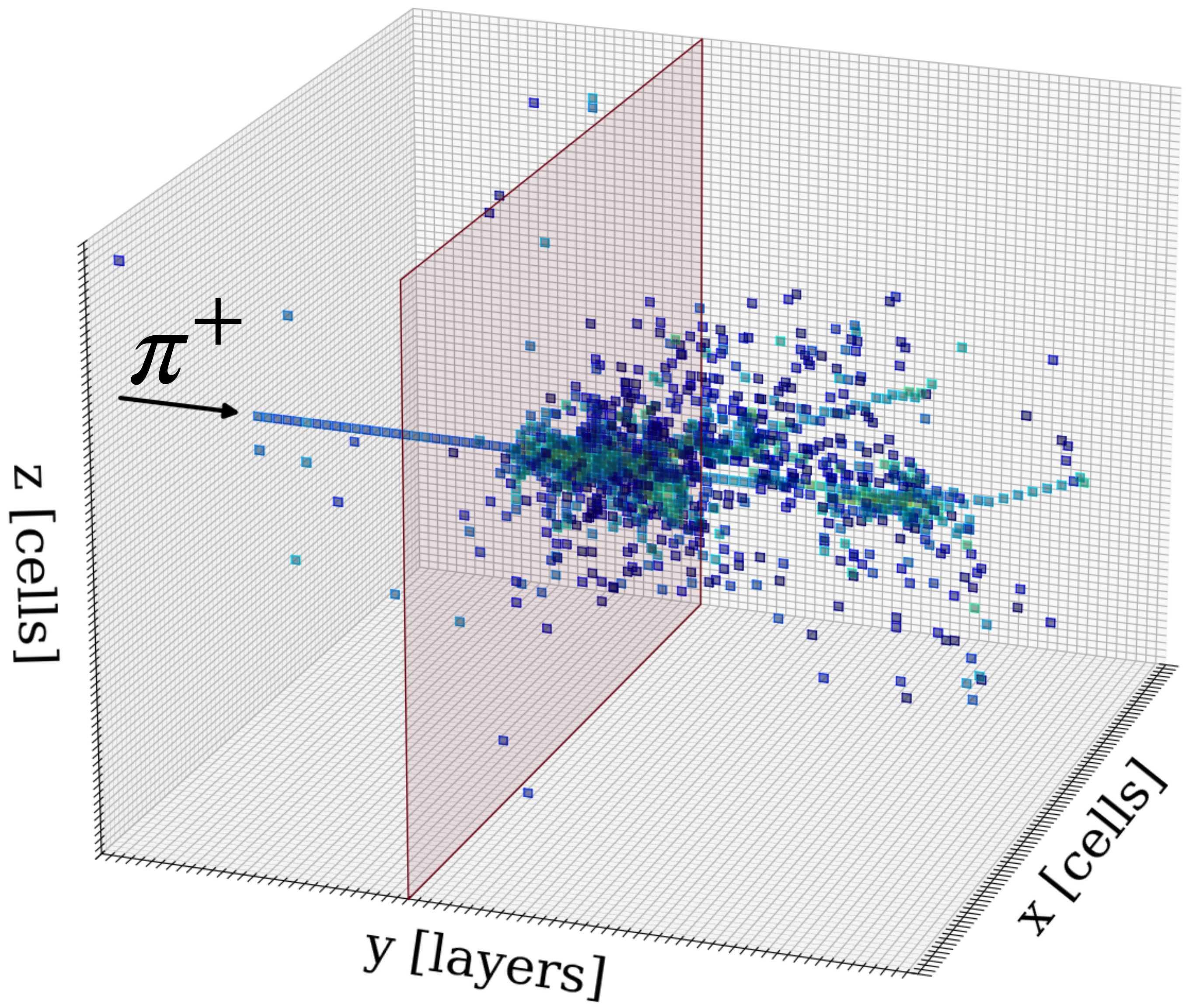


Sampling

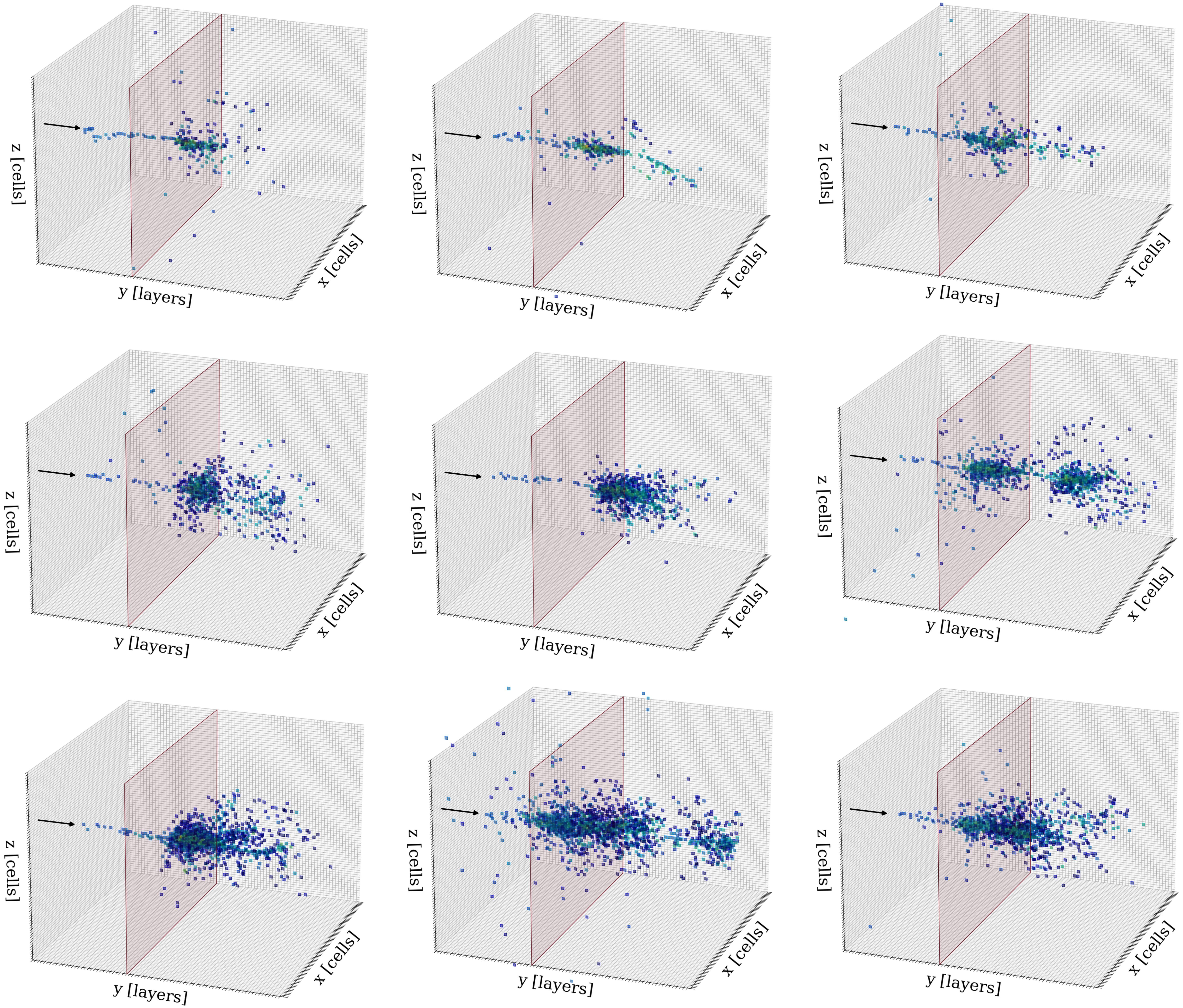
Results

π⁺ showers

Geant4 Simulation



CaloHadronic



15 GeV

50 GeV

85 GeV

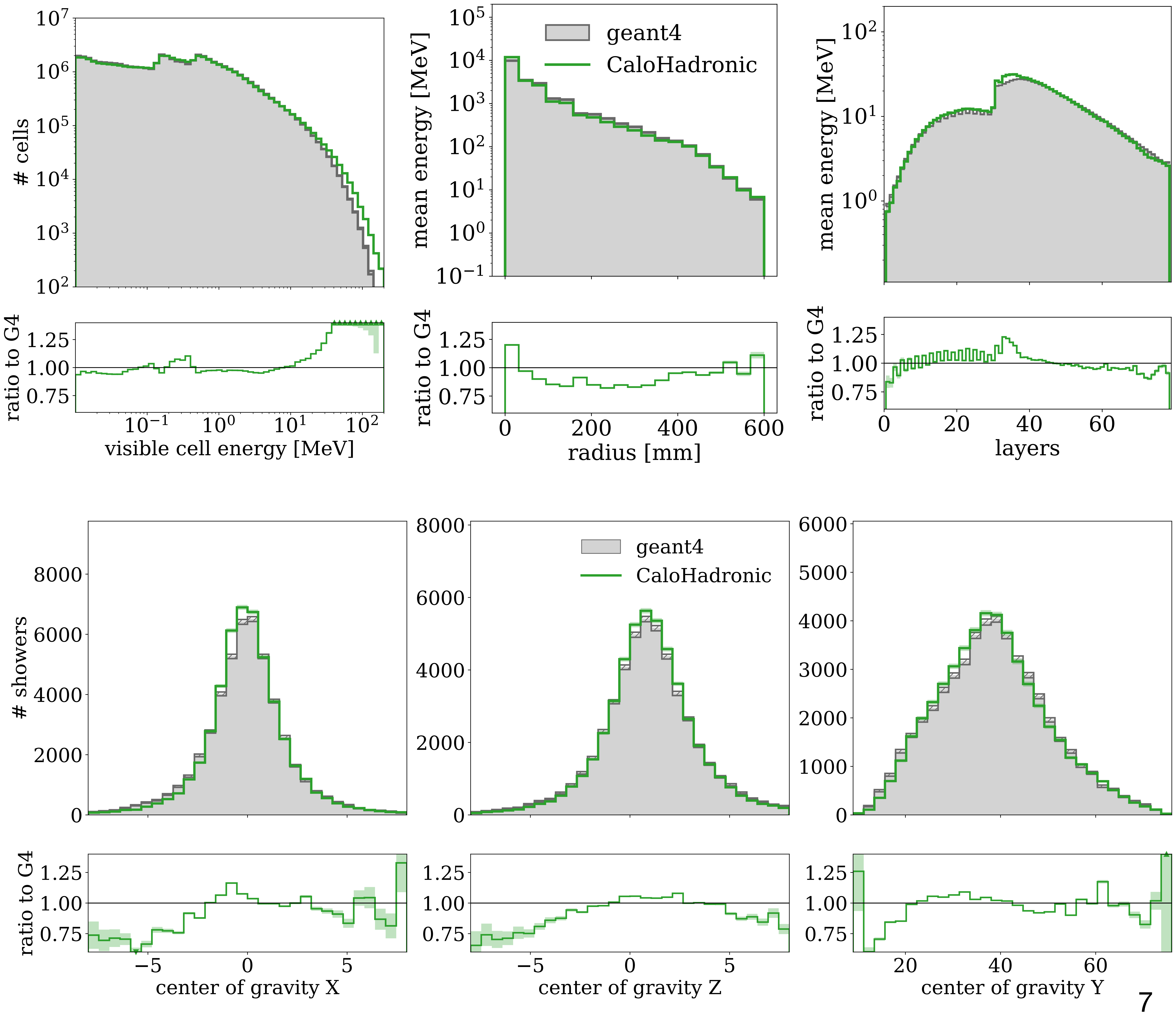
Results

10-90 GeV π⁺ showers

50 000 showers

All evaluations with point cloud showers are projected to a regular cell grid:

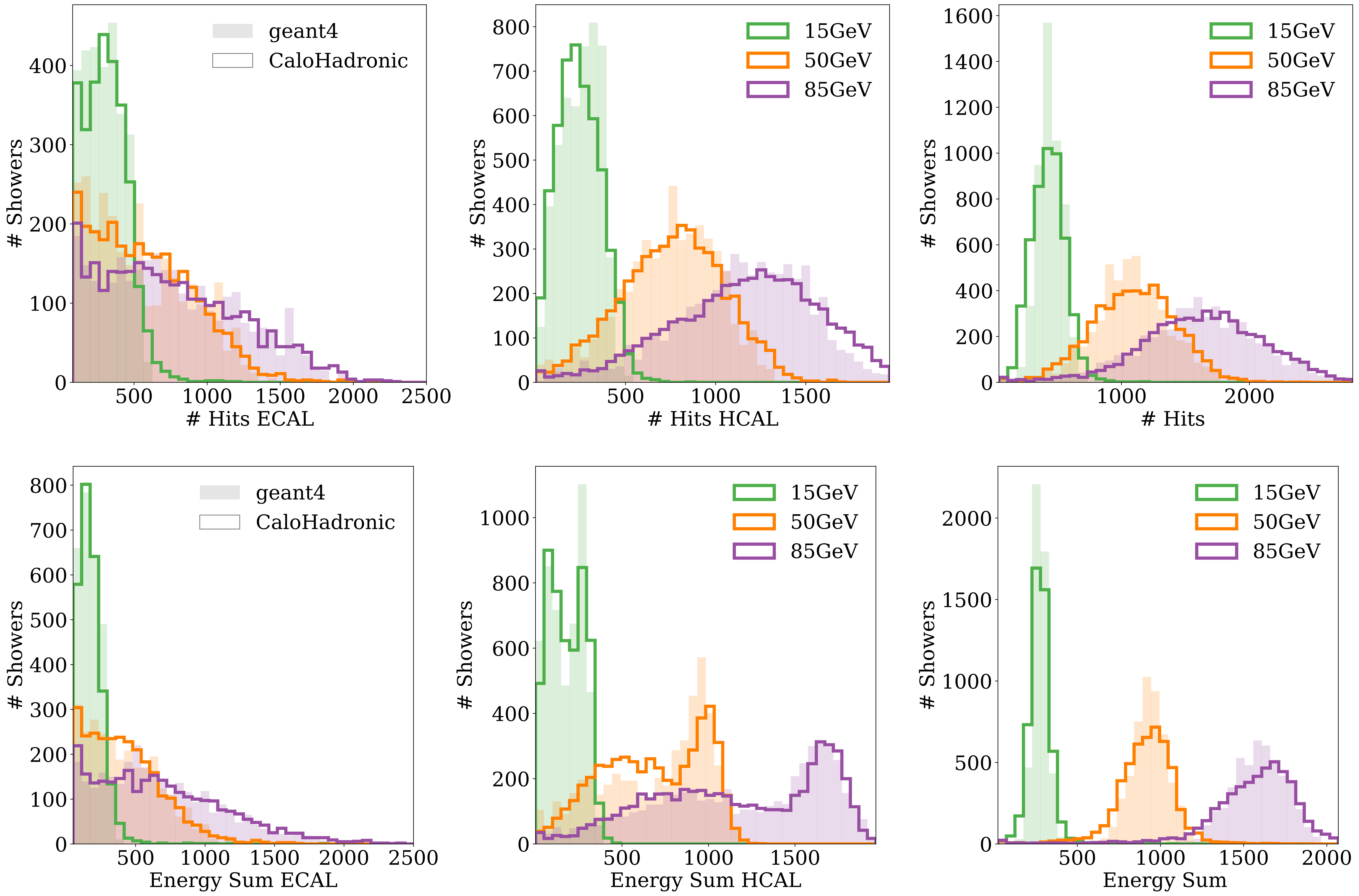
ECAL: 5 mm x 30 layers x 5 mm
HCAL: 30 mm x 48 layers x 30 mm



Results

15, 50 and 85 GeV π^+ showers

5 000 showers each



All evaluations with point cloud showers are projected to a regular cell grid:

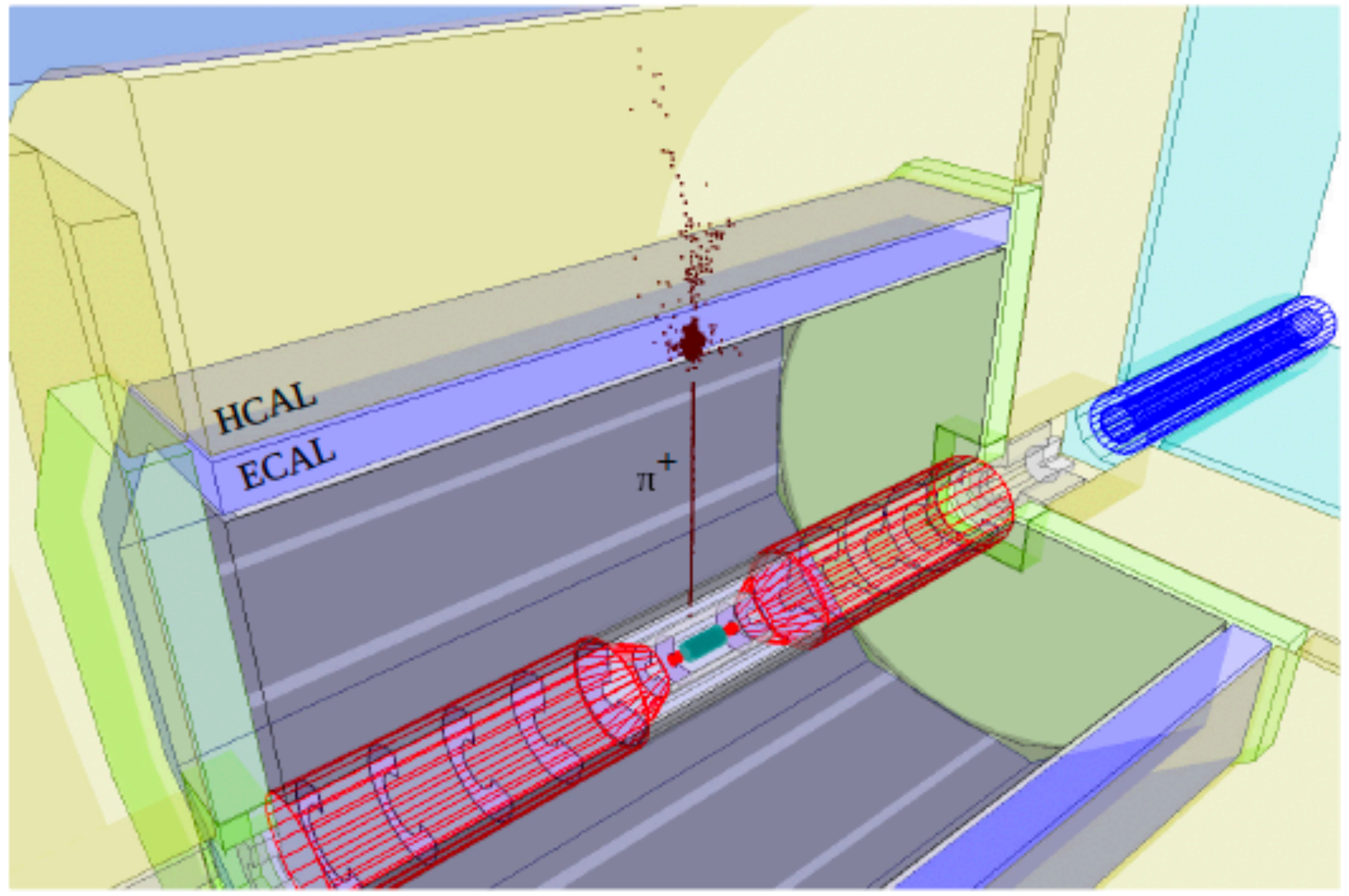
ECAL: 5 mm x 30 layers x 5 mm
HCAL: 30 mm x 48 layers x 30 mm

Fast calorimeter shower must ultimately be interfaced with the reconstruction chain

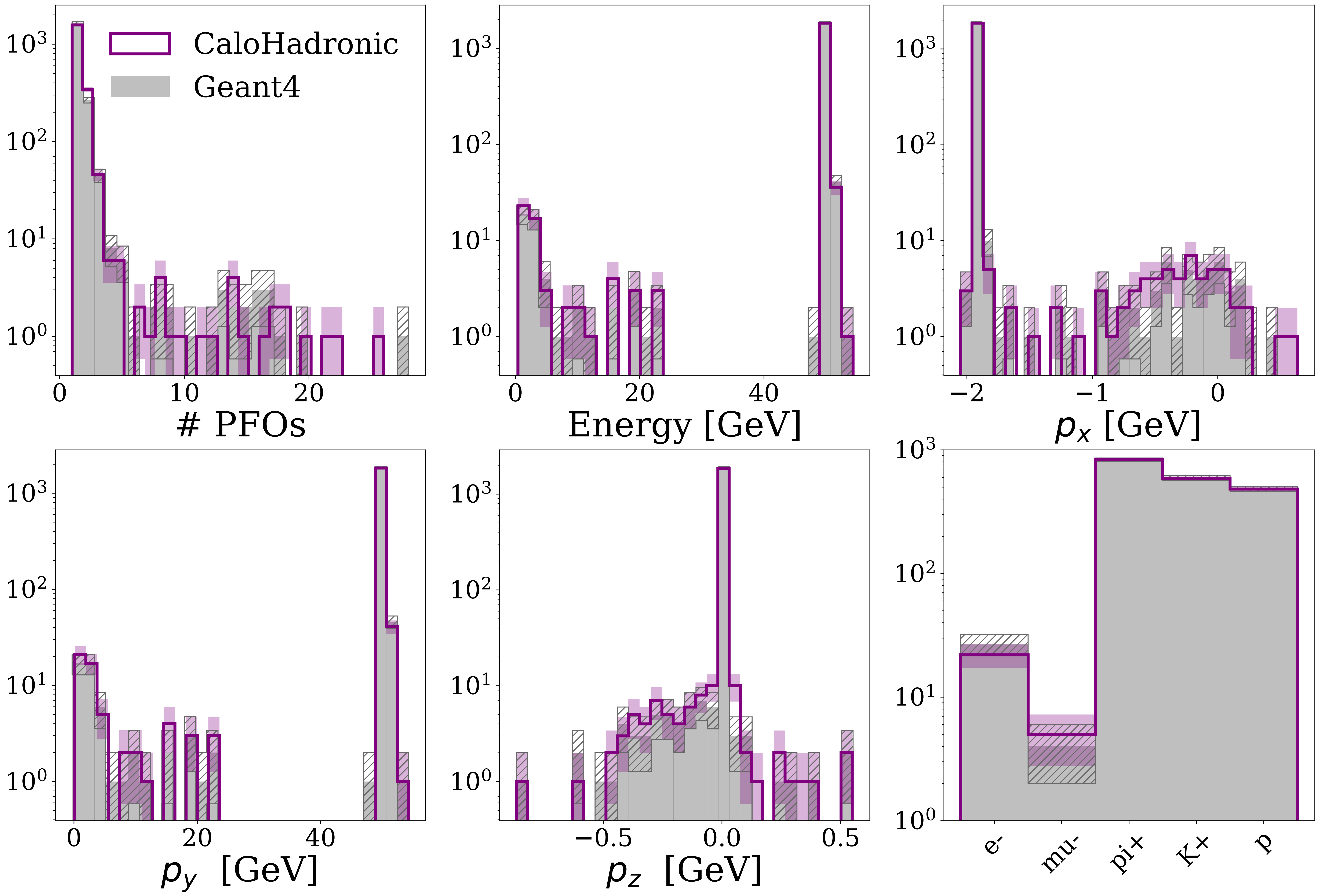
Reconstruction with Pandora Particle Flow

50 GeV π⁺ showers

- 2 000 showers



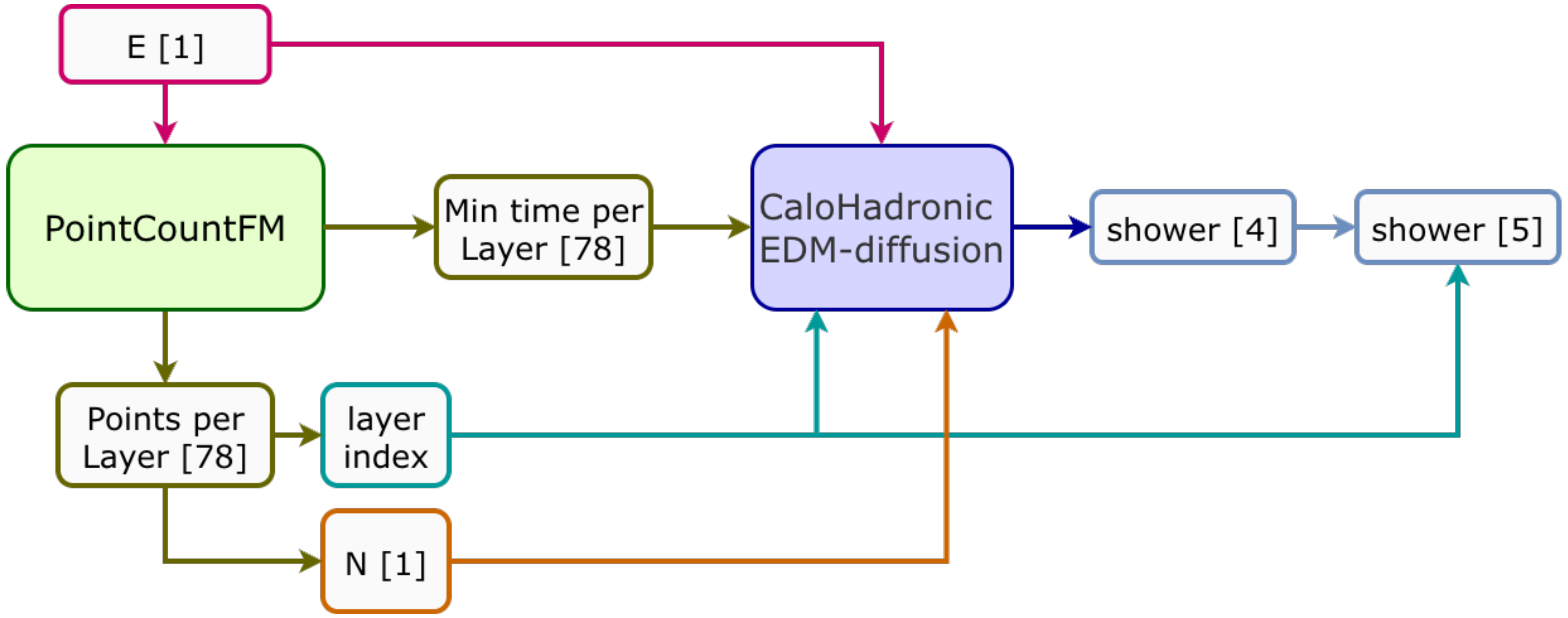
- allows performance validation in realistic detector geometry and in high-level physics observables



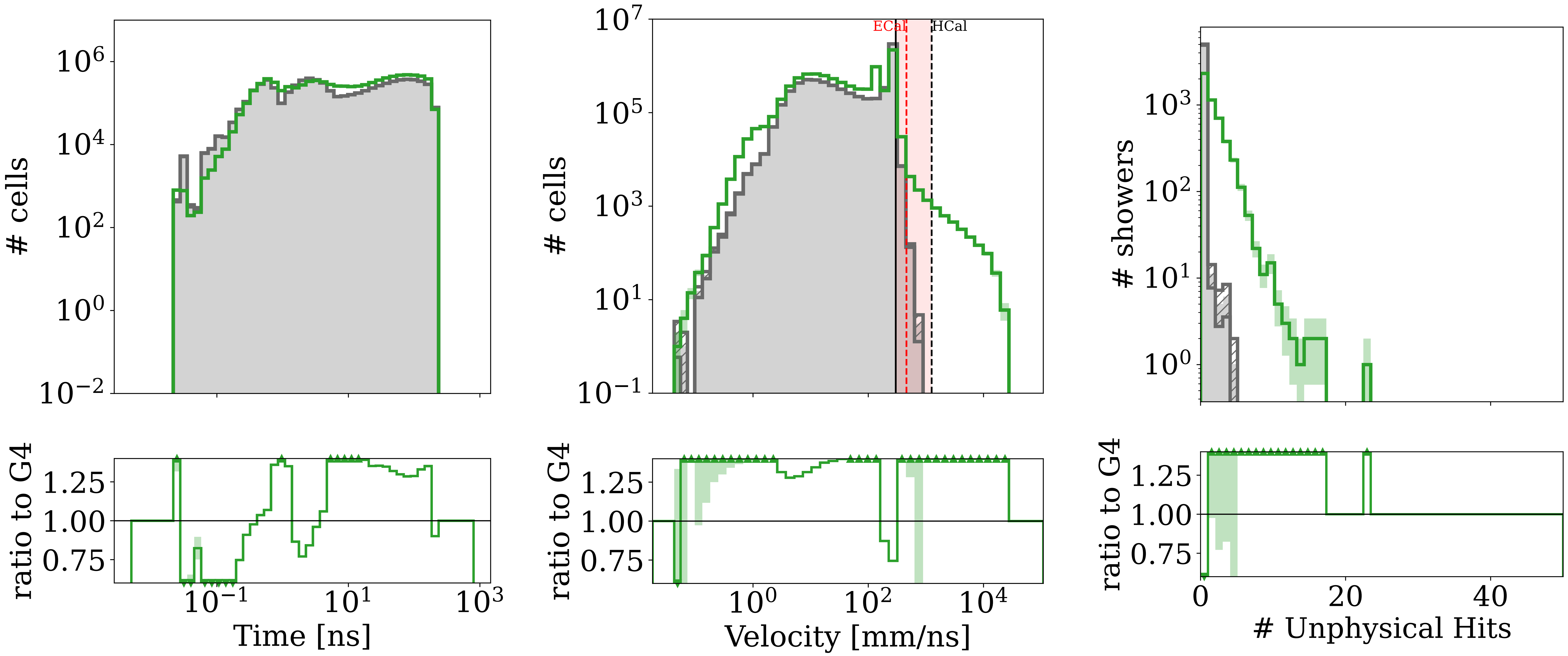
Particle Flow Objects (PFOs): list of reconstructed objects with particle's four-momentum and ID

Updates

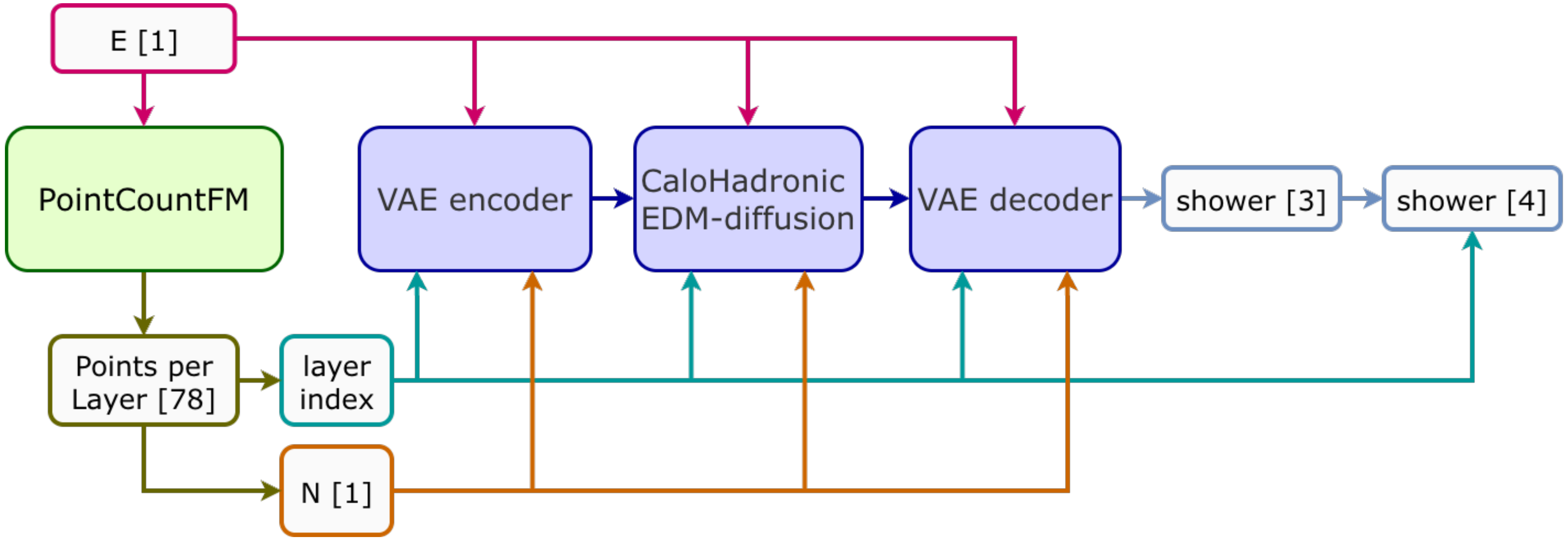
Updates: time feature



5k, 50GeV

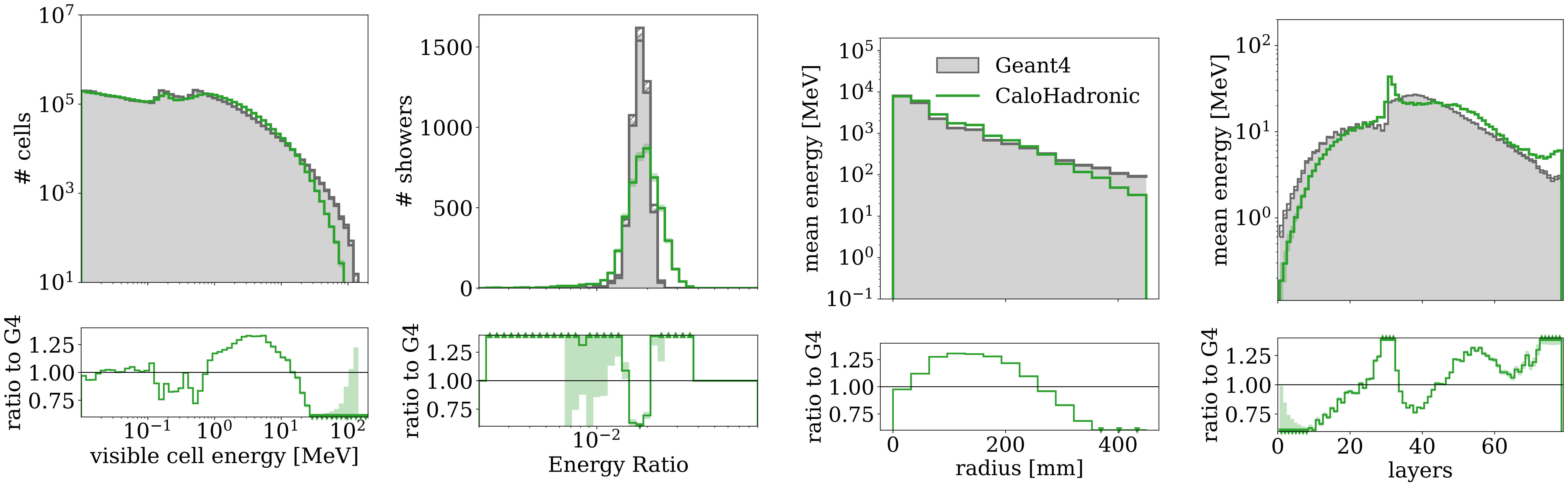


Updates: latent diffusion



~30 times faster than CaloHadronic

5k, 50GeV



Thank you