CaloClouds 3

Fast photon showers

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] DESY, [2] University of Hamburg [3] CERN
Munich, 12.03.2025

HELMHOLTZ

Overview

- Conceptual design and objectives

Overview

- Conceptual design and objectives
- Data format

Overview

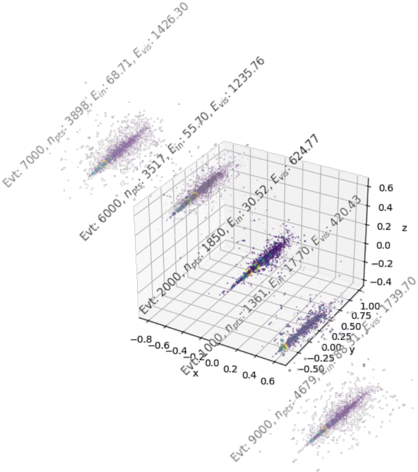
- Conceptual design and objectives
- Data format
- Network architecture

Overview

- Conceptual design and objectives
- Data format
- Network architecture
- Performance of the model

CONCEPTUAL DESIGN AND OBJECTIVES

The Objective



A generative model for detector response to photon showers.

- Produce realistic distributions.
- Must reconstruct like G4 events.
- Detector has a high granularity calorimeter, we need high spatial resolution.
- Model must be light weight, without too many parameters.
- Inference must be fast.

Comparison to industrial generative objectives

We use techniques from image generation, but need something else.

	Image generation	**Detector simulation**
Primary objective	Singular samples	Distribution over samples
Training data	Potentially limited	Excess available
Statistical properties	Discontinuities, irregular	smooth "physical" distributions
Submanifolds	Theorised, maybe subjective	Certain, quantifiable
Model size	May be large	Must be small

Reconstructions considered

Beyond just comparing the distributions of kinematics in the training data, we need a metric that reflects the end goal. The simulation should give the same results as G4 under reconstruction.

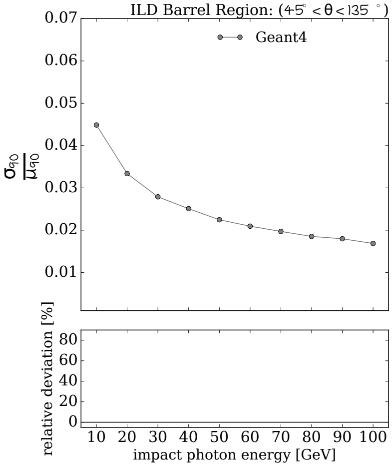
- The **energy** of a reconstructed photon should be the same.
- The **resolution**, $\frac{\sigma_{90}}{\mu_{90}}$, of a reconstructed photon should be the same.
- The **multiplicity** of reconstructed photons should be the same.

Reconstructions considered

Resolution of a reconstructed photon

- σ_{90} is the width of the distribution of reconstructed energies at 90% of the true energy.
- μ_{90} is the mean of the distribution of reconstructed energies at 90% of the true energy.

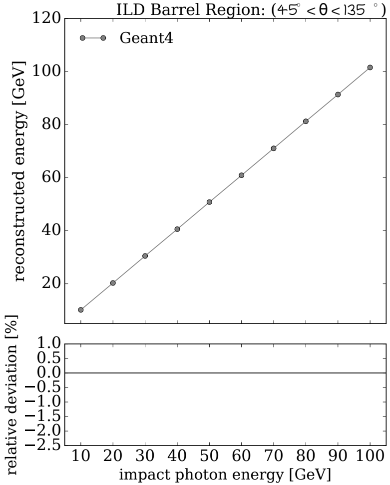
The resolution, $\frac{\sigma_{90}}{\mu_{90}}$, falls with increasing energy using a G4 simulation.



Reconstructions considered

Energy of a reconstructed photon

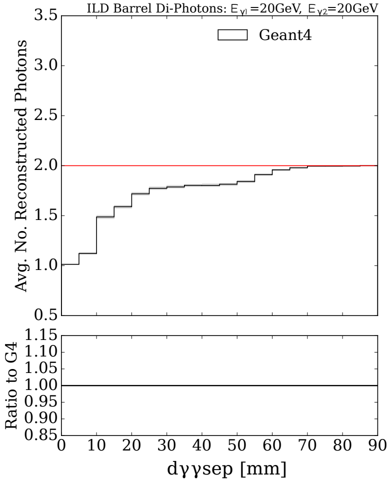
Using a G4 simulation reconstructed energy of the photon is very linear with the true energy of the photon.



Reconstructions considered

Multiplicity of reconstructed photons

The number of reconstructed photons varies with separation. We must reproduce the quite non linear behaviour of the G4 simulation.



DATA FORMAT

Grids v.s. Clouds

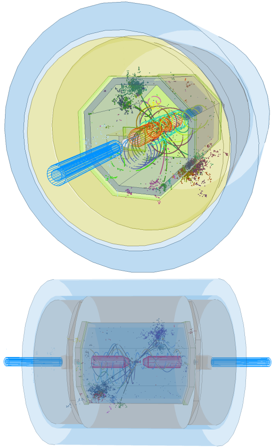
A taxonomy of fast detector simulations should be done by the way the data is represented;
This is not an comprehensive list (apologies if your favourite model isn't shown).

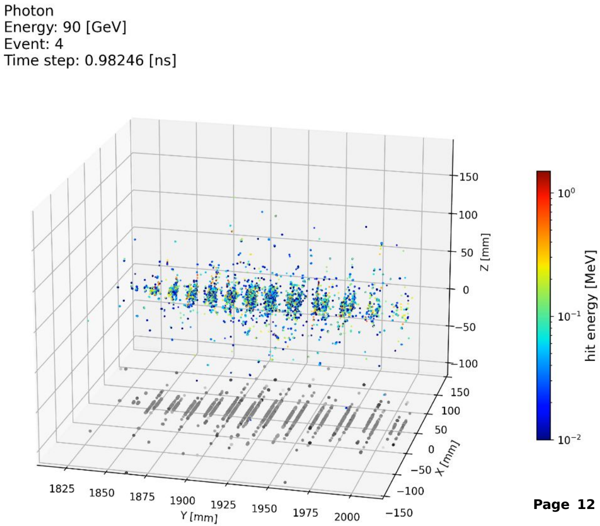
- Grid based
  - Grid is fixed to location in detector
    - ATLAS fast sim.
  - Grid is fixed to shower axis
    - All the models of CaloChallange.
    - Models being developed by CERN quest.
    - L2L flows

- Gridless distributions
  - Point clouds
    - CaloClouds (I, II and 3)
  - Parametrised distributions
    - DELPHIS (kind of)
    - Old CMS fast sim.
- Post reconstruction (jets, etc)
  - CMS new fast sim.
  - PIPPIN fast simulation and reconstruction.

Our training data

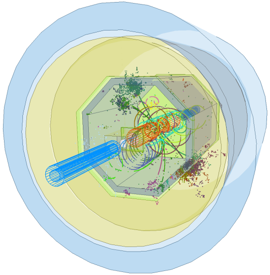
Photon showers in the ILC EM calorimeter

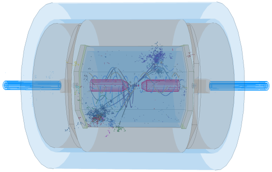




Our training data

Geometry of the ILC EM calorimeter



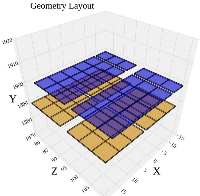


- The ILC EM calorimeter is a high granularity calorimeter.
- Designed to do precision work for a Higgs factory.
- Calorimeter naturally has supports, creating blind spots.
- Each module contains two PCBs, such that layers are grouped in pairs.
- Modules are intentionally offset, such that cells are staggered one layer to the next.

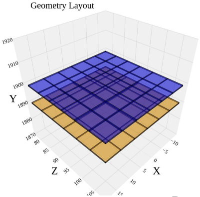
Our training data
Regularising the geometry

We want one model for the whole calorimeter, but the irregular geometry is position dependant.

- Gaps due to supporting structures are removed.
- The modules are aligned, so each cell is centred above the one radially below.



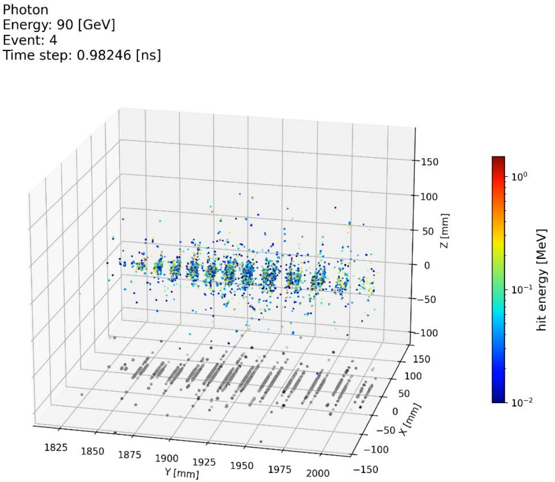
⇓ *Regularise* ⇓



Reduced input multiplicity

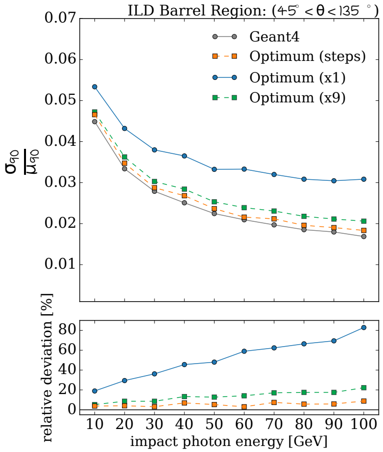
G4 gives more details than just which cell was hit; approximately 40k “steps” per event.

- More steps reduces the impact of moving the shower to different cells.
- Training on all G4 steps is not feasible.
- We must discover a safe, but feasible, granularity.



Reduced input multiplicity

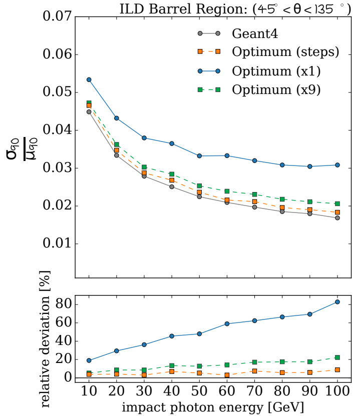
Impact on performance metrics

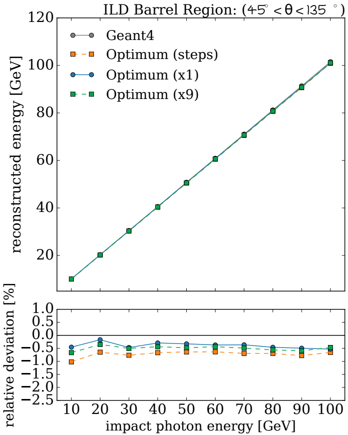


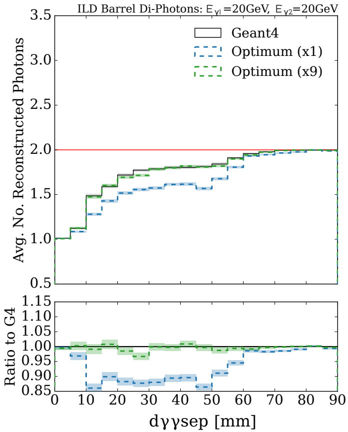
- Run G4 with the regular detector geometry.
- Place hits into regular grids.
  - Optimum (x1) is a grid of the regular detector cells.
  - Optimum (x9) is a grid with 9 time the fidelity of Optimum (x1).
  - Optimum (steps) is a sample without the grid.
- Project the hits back into real geometries and evaluate performance.

Reduced input multiplicity

Impact on performance metrics

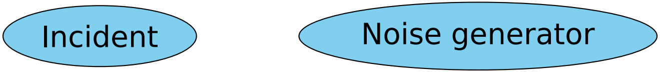




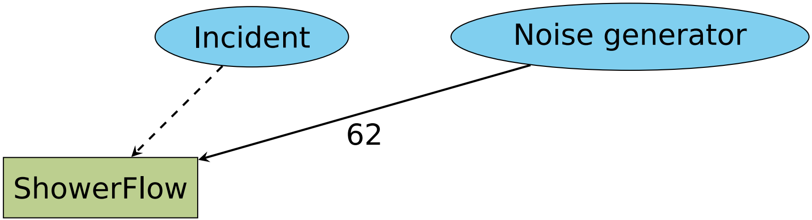


NETWORK ARCHITECTURE

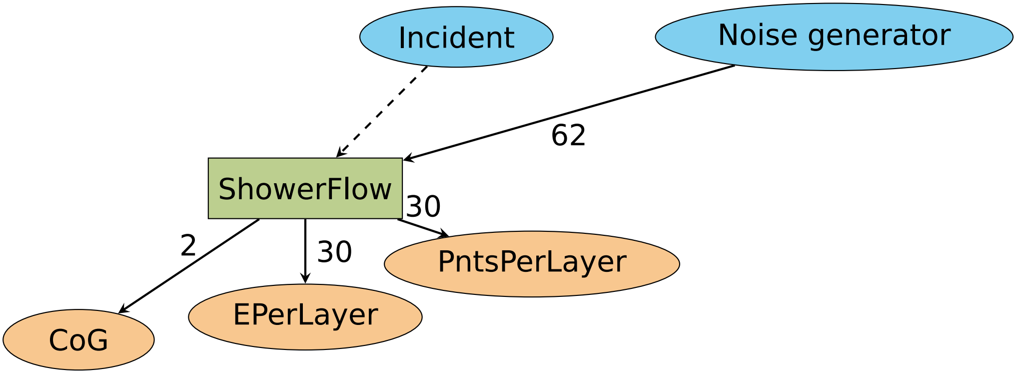
Architecture of CaloClouds 3



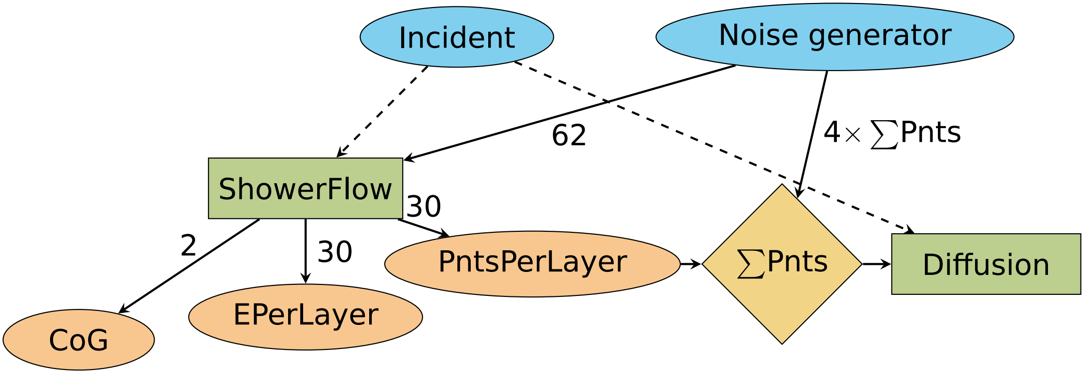
Architecture of CaloClouds 3



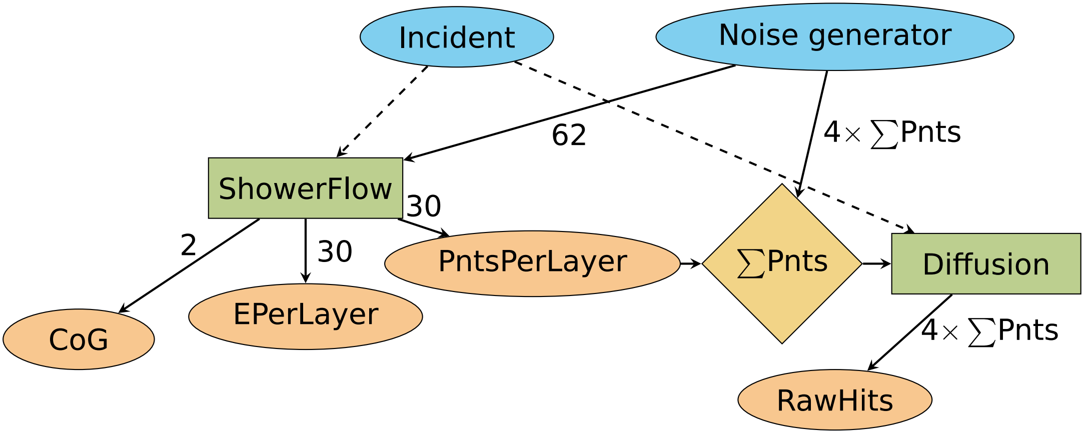
Architecture of CaloClouds 3



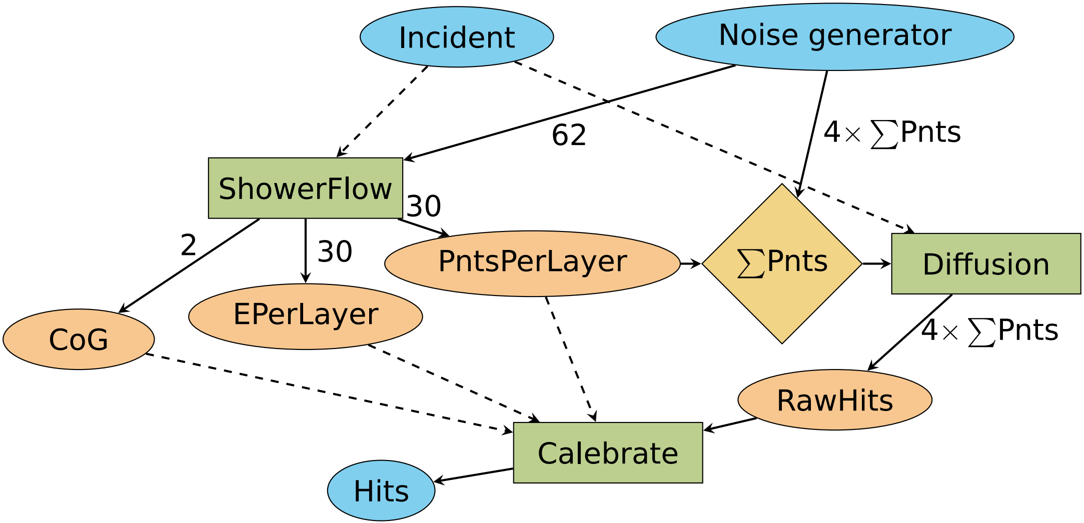
Architecture of CaloClouds 3



Architecture of CaloClouds 3

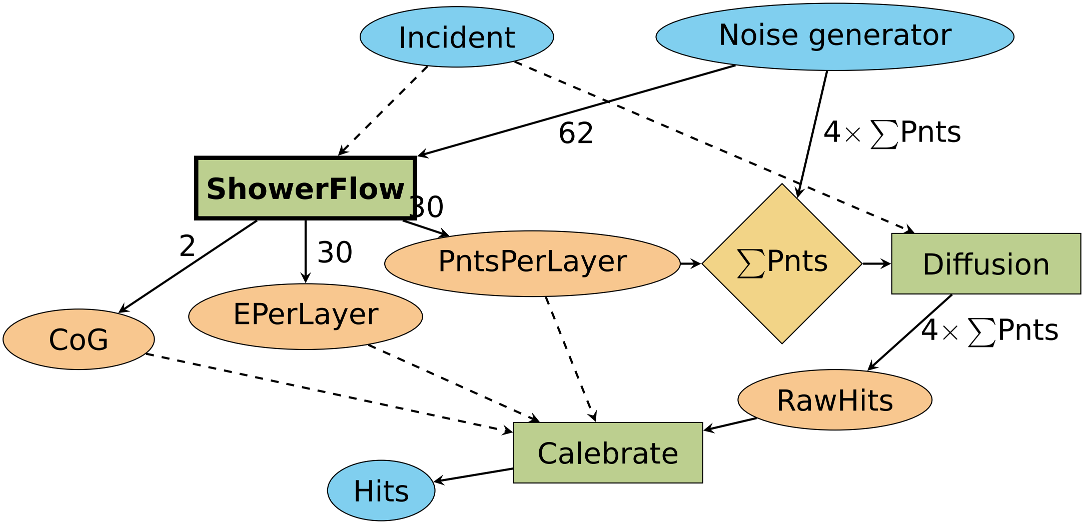


Architecture of CaloClouds 3



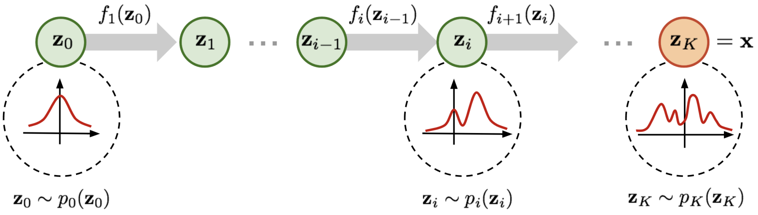
Architecture of CaloClouds 3

ShowerFlow



Architecture of CaloClouds 3

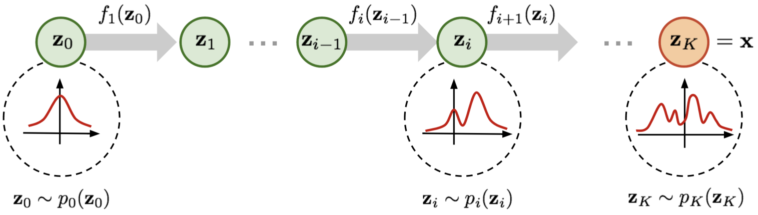
ShowerFlow



A normalising flow, with transforms conditioned on the input.

Architecture of CaloClouds 3

ShowerFlow

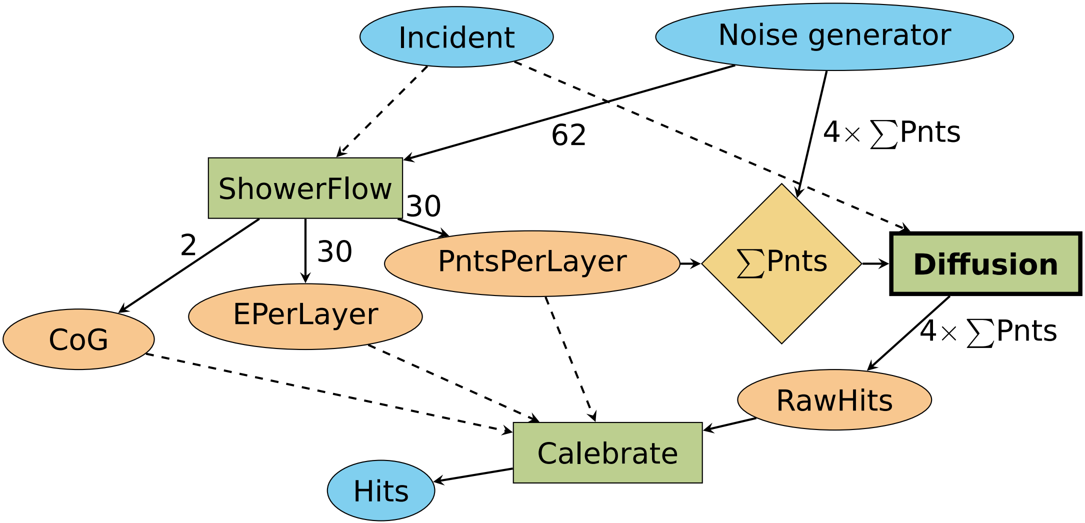


A normalising flow, with transforms conditioned on the input.

- Input is a vector of noise length 62.
- Transforms include 14 permutations, 12 affine couplings and 2 splines (interleaved).
- The output is center of gravity in x and y, energy per layer and points per layer.

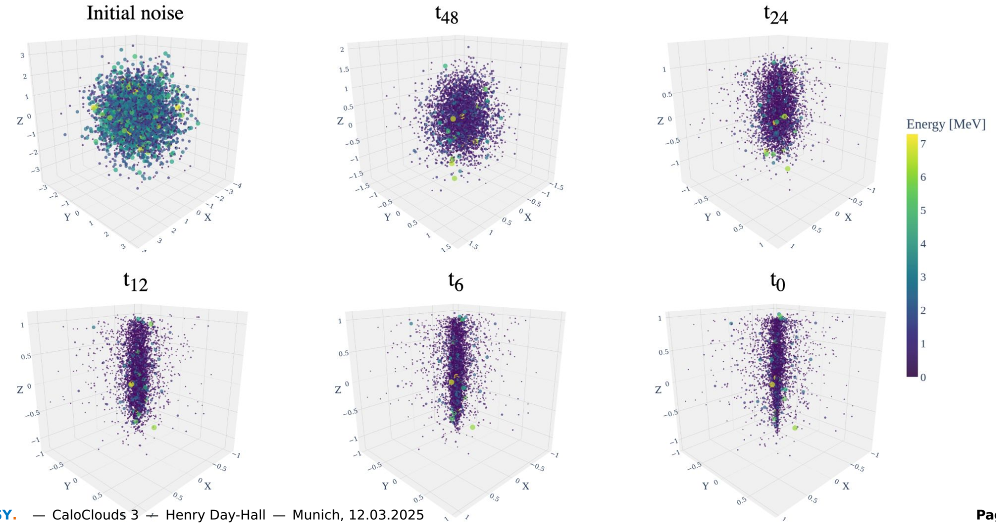
Architecture of CaloClouds 3

Diffusion model



Architecture of CaloClouds 3

Diffusion model

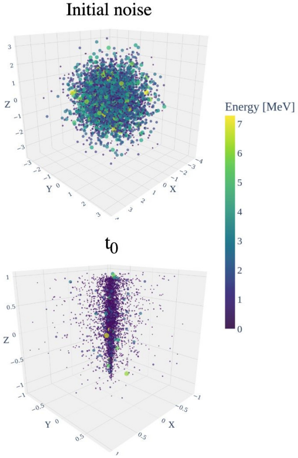


Architecture of CaloClouds 3

Diffusion model

A diffusion model, conditioned on the incident particle.

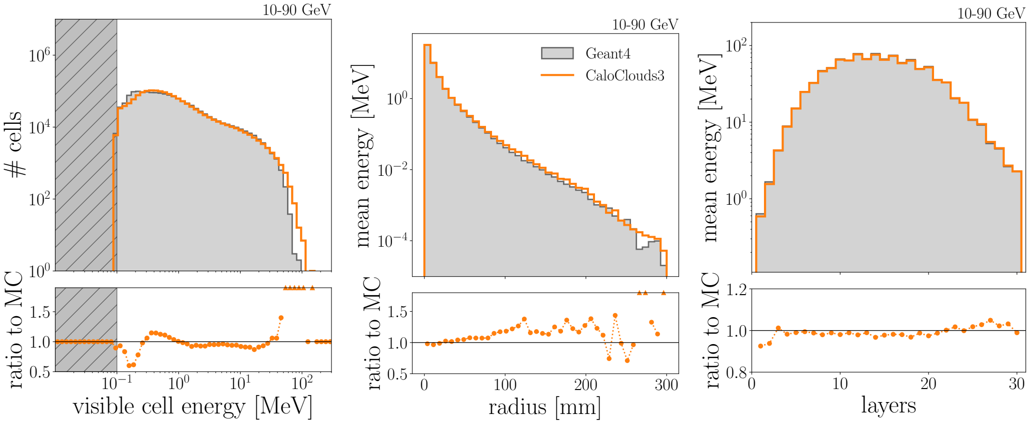
- k-Diffusion in the style of Karras et al., 2022
- Each hit (4 vector) is a single draw, can draw batches.
- Points are iid (the diffusion model has no latent space).
- The model is distilled for speed.



PERFORMANCE OF MODEL

Performance of model

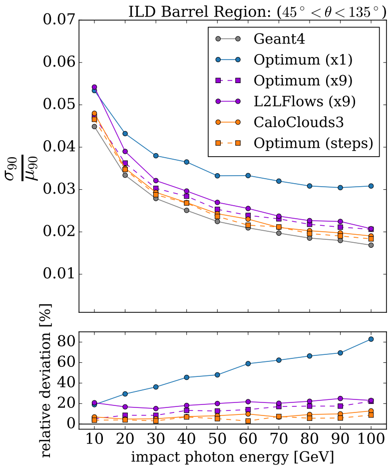
Against the training data



Performance of the model

Resolution of a reconstructed photon

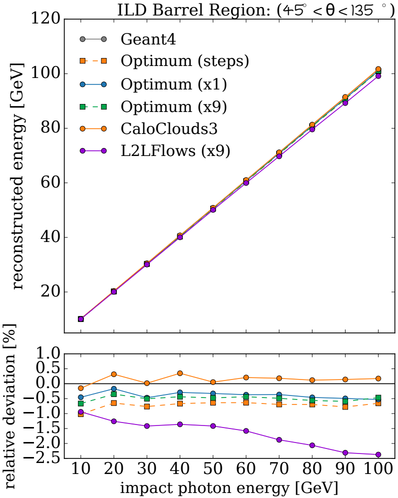
- CaloClouds 3 reproduces the resolution of the photon showers almost as well as is possible with the training data.
- Remaining discrepancies are due to the irregular nature of the detector.



Performance of the model

Energy of a reconstructed photon

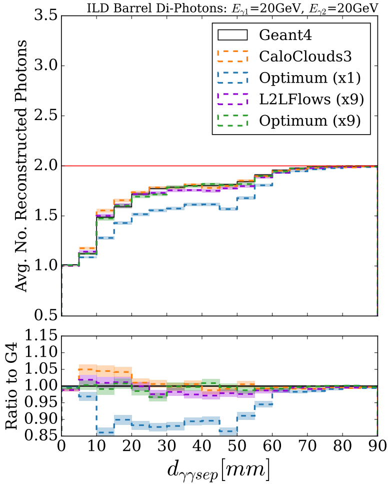
- CaloClouds 3 receives a constant correction factor to match the simulated energy to the expected energy after projecting back into a real geometry.
- This correction factor makes up for energy loss from the simulated shower in dead regions. It moves all CaloClouds 3 ratios upwards by the same amount.
- Importantly, CaloClouds 3 remains linear in reconstructed energy.



Performance of the model

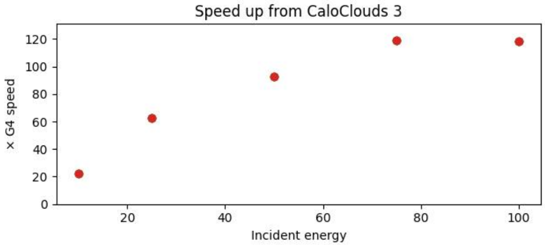
Multiplicity of reconstructed photons

- Reconstructions of CaloClouds 3 obtain the same number of photons as the photons are separated.
- Devotion from the G4 behaviour is very minor, and almost within errors.



Performance of the model

Inference speed



While the speed up from this model does vary with photon energy, it is highly efficient across a wide range of relevant energies.

Conclusions and further work

- CaloClouds 3 is a fast generative model using diffusion and normalising flows for photon showers in a high granularity calorimeter.
- It reproduces the kinematics of it’s training data with great accuracy.
- Objects produced upon reconstruction are nearly indistinguishable from those produced by G4.

Thank you!

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Henry Day-Hall
0000-0002-7881-2506
FTX-SFT
henry.day-hall@desy.de
+49–15510497087
arxiv:2309.05704

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As there
some unprocessed data that should have been added to the final pag
extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page will g
because LaTeX now knows how many pages to expect for this documen