CaloClouds 3

Fast photon showers

Thorsten Buss[1,2], **Henry Day-Hall**[1], Frank Gaede[1],
Gregor Kasieczka[2], Katja Krüger[1], Anatolii Korol[1],
Thomas Madlener[1], Peter McKeown[3], Martina Mozzanica[2],
Lorenzo Valente[2]

[1] DESY, [2] University of Hamburg [3] CERN

ML Round Table, 14.11.2025

HELMHOLTZ

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Overview

- Objectives and dataset

Overview

- Objectives and dataset
- Network architecture

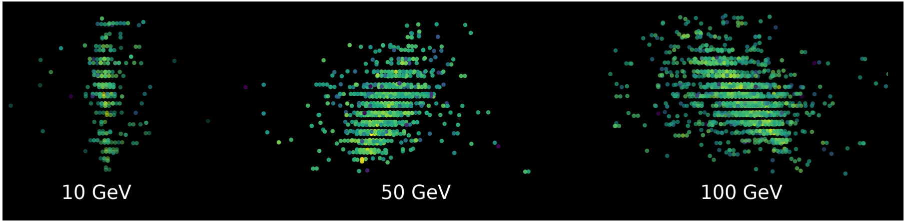
Overview

- Objectives and dataset
- Network architecture
- Performance

OBJECTIVES AND DATASET

The Objective

Give me this. . . fast!

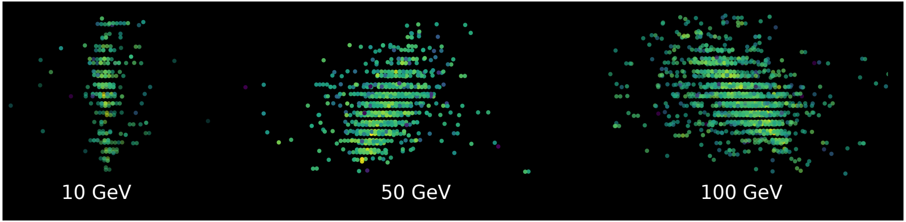


How is a photon recorded in a high granularity calorimeter?

(e.g. 10^8 cells, compared to ATLAS with $\approx 10^5$ cells)

The Objective

Give me this. . . fast!



How is a photon recorded in a high granularity calorimeter?

(e.g. 10^8 cells, compared to ATLAS with $\approx 10^5$ cells)

Normal sim → Photon reaches calorimeter

Normal sim → Other particles (G4)

Photon reaches calorimeter → G4 (slowly) simulates photon in calorimeter

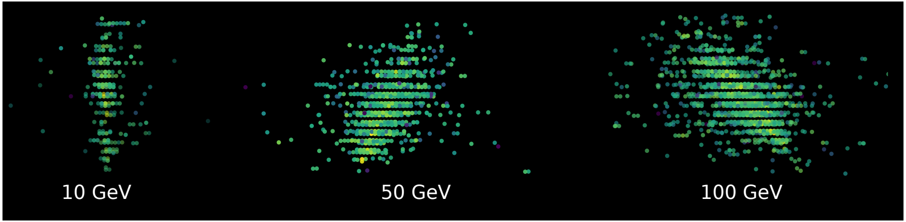
Other particles (G4) → Normal reco.

G4 (slowly) simulates photon in calorimeter → Normal reco.

Takes too long for the amount of simulated data required.

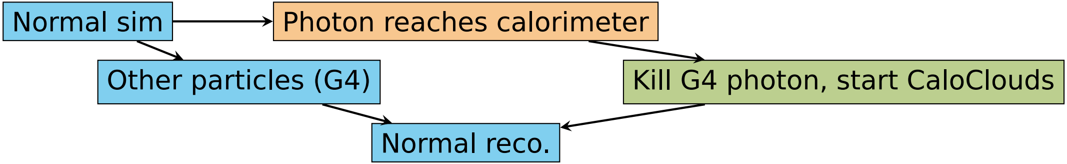
The Objective

Give me this. . . fast!



How is a photon recorded in a high granularity calorimeter?

(e.g. 10^8 cells, compared to ATLAS with $\approx 10^5$ cells)



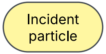
Need to be as fast as possible at inference (and small in memory).

NETWORK ARCHITECTURE

Network architecture

Normalising flow and diffusion model

Incident particle is specified by (e, p_x, p_y, p_z)



(x, y, z) handled later.

Network architecture

Normalising flow and diffusion model

ShowerFlow is a normalising flow - accurate, but can't do high dimensions.

Predicts

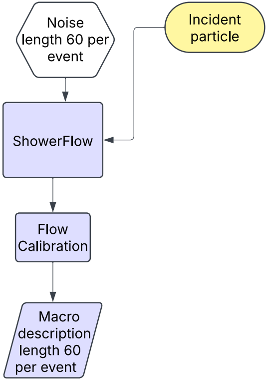
$N_{\text{layer 1}}, N_{\text{layer 2}} \cdots N_{\text{layer 30}}$

where N is points in layer

and

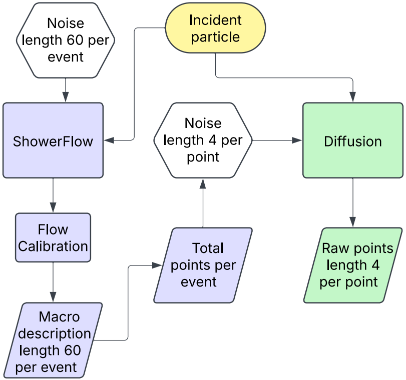
$E_{\text{layer 1}}, E_{\text{layer 2}} \cdots E_{\text{layer 30}}$

where E is total energy in layer



Network architecture

Normalising flow and diffusion model

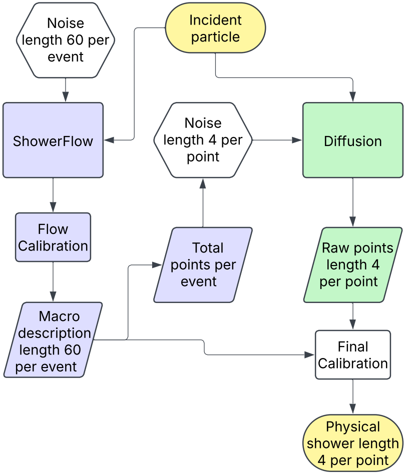


Diffusion model creates batch of iid points.

(e_1, x_1, y_1, z_1),
(e_2, x_2, y_2, z_2),
$\dots$
$(e_{\sum N}, x_{\sum N}, y_{\sum N}, z_{\sum N})$

Network architecture

Process of generation



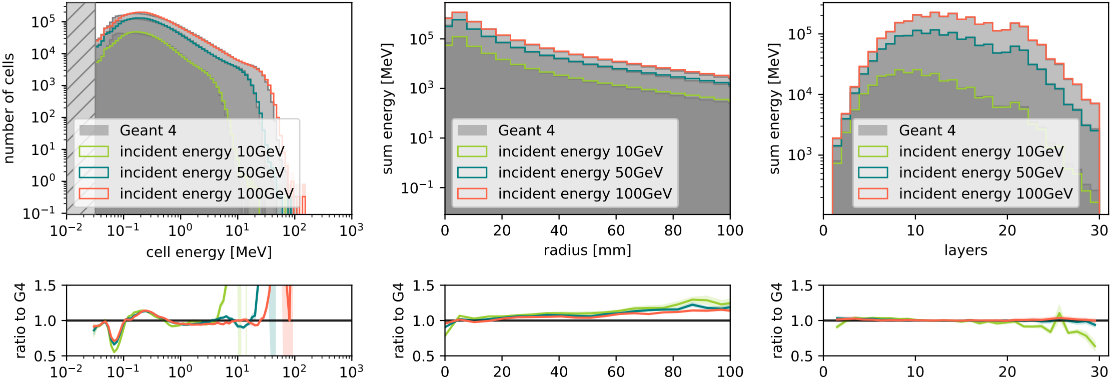
Split the points between layers, as dictated by ShowerFlow.

Enforce energy per layer predicted by ShowerFlow.

PERFORMANCE

Performance

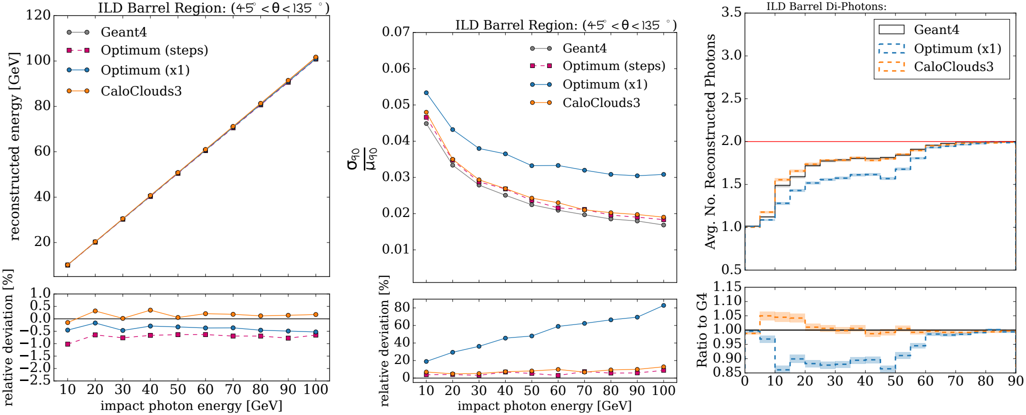
Kinematics at fixed energies



Good results across a range of energies.
High energies are better, because iid assumption holds better.

Reconstructed quantities

After the simulated calorimeter is processed by standard reconstruction algorithms



CaloClouds3 agrees as well as is possible given the limitations of the training data.

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 is reconstruction ready, and performs fast and accurately.

Further work

- Inversion of the flow models for a potential reconstruction tool.

In other models, handling multiple particle types collectively is explored, see CaloHadronic, PanShower, OmniJet-α.

JOINT ML JOURNAL CLUB

https://indico.desy.de/event/51323/

Thank you!

Contact

Deutsches Elektronen-
Synchrotron DESY

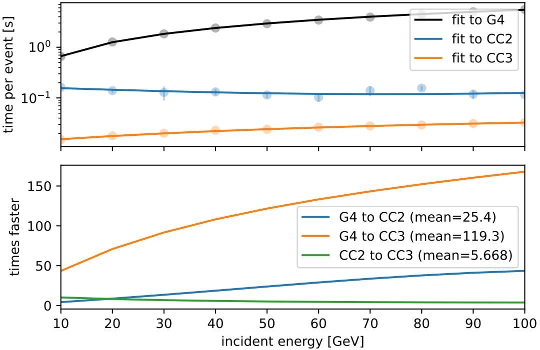
ftx.desy.de

Henry Day-Hall
0000-0002-7881-2506
FTX-SFT
henry.day-hall@desy.de
https://github.com/FLC-QU-hep/CaloClouds-3

BACKUP

Inference speed

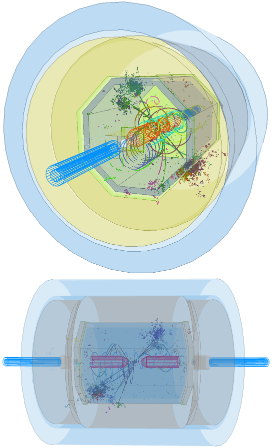
CPU, including projection and reconstruction



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

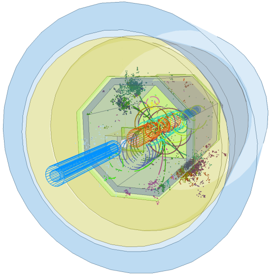
Chosen detector example

Geometry of the ILC EM calorimeter



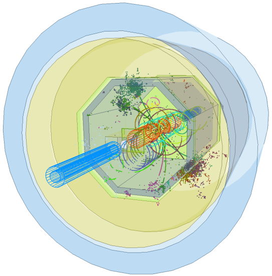
- The ILC EM calorimeter is a high granularity calorimeter, we consider SiECAL variant.
- Calorimeter naturally has supports, creating blind spots.
- Layers of sensitive silicon detectors are sandwiched either side of passive Tungsten absorber, creating alternating patterns.
- Silicon cells are $5 \times 5 \times 0.5$ mm, Tungsten is 2.1 mm in first 20 layers, then 4.2 mm in last 10.
- Octagonal structure means layers start at varying offset, such that cells are staggered one layer to the next.
- Magnet field of 3.5 T.

PreProcessing
Regularising the geometry

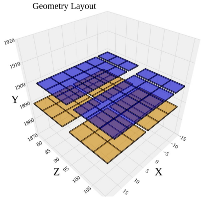


Take the ILD (SiECAL) as an example. It has a octagonal ECAL with 30 layers.
We want one model for the whole calorimeter, so we need to make some position agnostic training data.

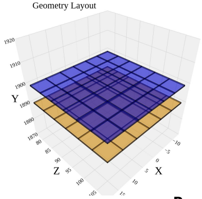
PreProcessing
Regularising the geometry



Take the ILD (SiECAL) as an example. It has a octagonal ECAL with 30 layers.
We want one model for the whole calorimeter, so we need to make some position agnostic training data.



$\Downarrow$ *Regularise* $\Downarrow$



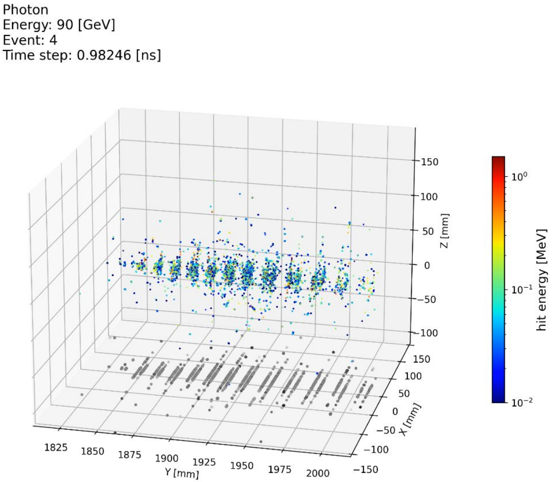
- Gaps due to supporting structures are removed.
- The modules are aligned, so each cell is centred above the one radially below.

End up with ≈ 1000 of (e_i, x_i, y_i, z_i) per event.

Reduced input multiplicity

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

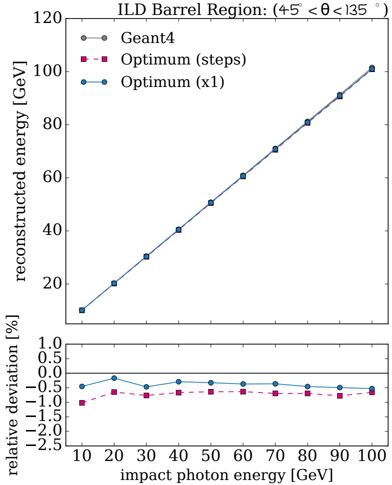
- More steps reduces the impact of moving the shower to different cells.
- Training on all G4 steps is not feasible, we find 6k a good compromise.
- Each cell is divided into 5×5 subcells, all energy is placed at location of highest energy hit.



Reconstructed quantities

Energy of a reconstructed photon

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

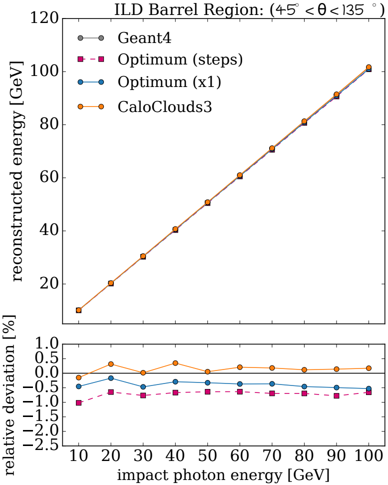


Reconstructed quantities

Energy of a reconstructed photon

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

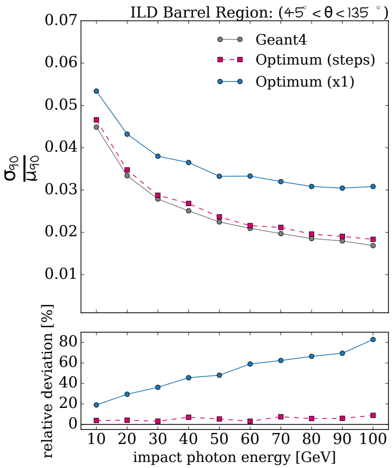


Reconstructed quantities

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.



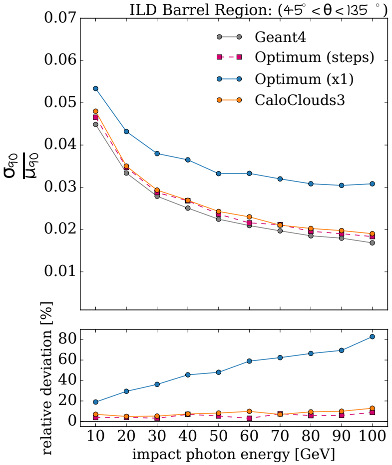
Reconstructed quantities

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.

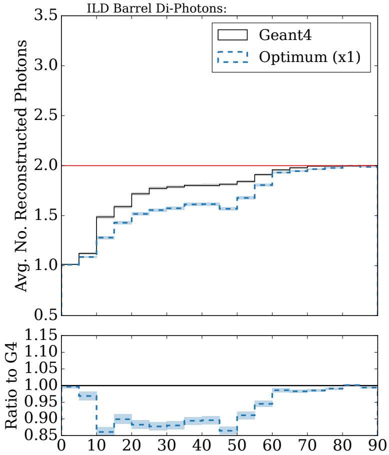
The resolution, $\frac{\sigma_{90}}{\mu_{90}}$, falls with increasing energy using a G4 simulation.

- CaloClouds 3 reproduces the resolution of the photon showers almost as well as is possible with the training data.



Reconstructed quantities
Multiplicity of reconstructed photons

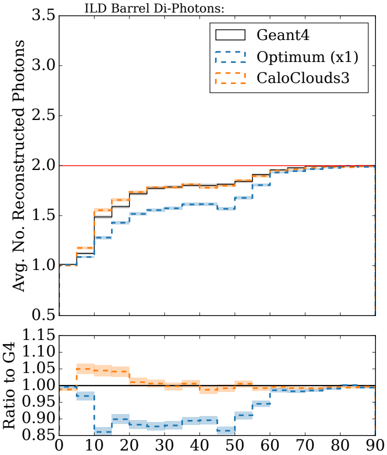
The number of reconstructed photons varies with separation. We must reproduce the quite non linear behaviour of the G4 simulation.



Reconstructed quantities
Multiplicity of reconstructed photons

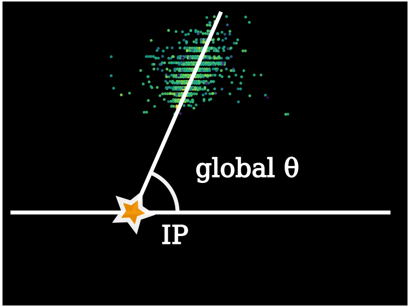
The number of reconstructed photons varies with separation. We must reproduce the quite non linear behaviour of the G4 simulation.

- Devotion from the G4 behaviour is very minor, and almost within errors.



Reconstructed direction

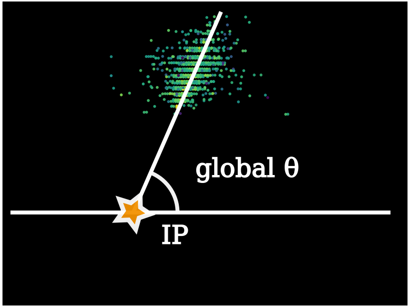
What is an internal angle?

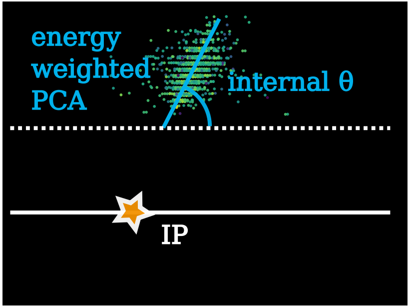


Global particle direction assumes it came from the IP.
Detector is big compared to the shower $\rightarrow$ shower shape has little impact.

Reconstructed direction

What is an internal angle?



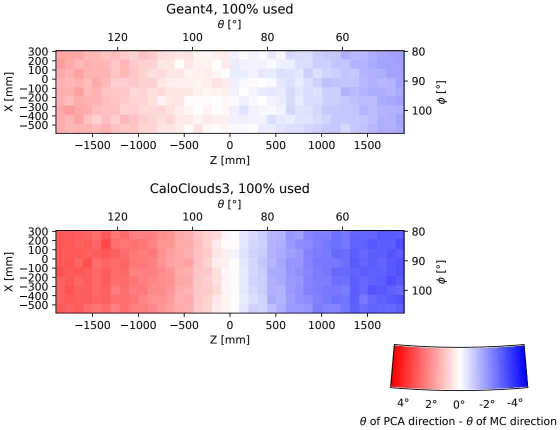


Global particle direction assumes it came from the IP.
Detector is big compared to the shower → shower shape has little impact.

For LLPs, this assumption doesn't hold.
Particles direction can be estimated from the distribution of the points.

Reconstructed direction

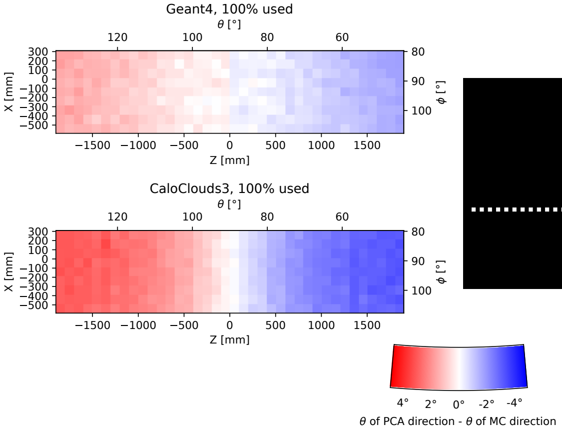
A bias in reconstruction

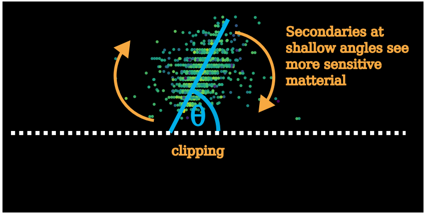


Both have a bias in the reconstructed θ, CaloClouds 3 is worse than G4.

Reconstructed direction

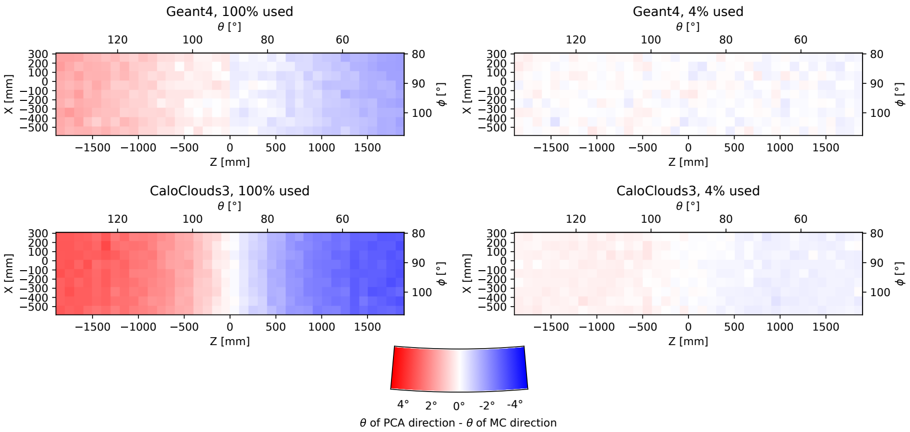
A bias in reconstruction





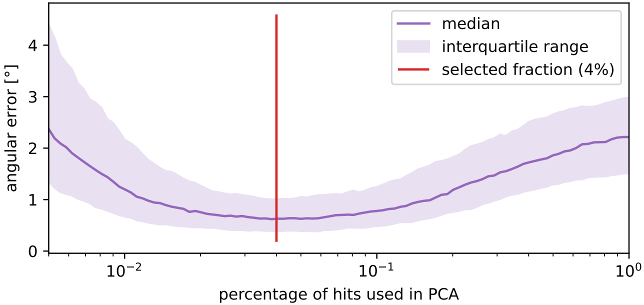
Speculative...

Reconstructed direction
Improved internal angle



Reconstructed direction

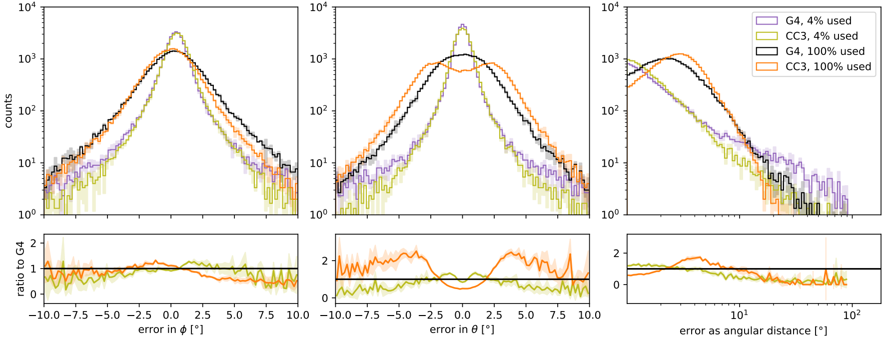
Improving the reconstruction



By restricting the PCA calculation to use only the 4% of hits with higher energy, the direction calculated closer to the incident particle.

Reconstructing internal angles

With all points, and highest energy 4%



Agreement is poor for all hits, but good when using the improved metric, with the 4% of highest energy hits.