

CALOCLOUDS3: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation

Thorsten Buss^{a,b} Henry Day-Hall^{a,1} Frank Gaede^a Gregor Kasieczka^b Katja Krüger^a
Anatolii Korol^a Thomas Madlener^a Peter McKeown^c Martina Mozzanica^b Lorenzo
Valente^b

^a*Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*

^b*Institut für Experimentalphysik, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany*

^c*CERN, 1211 Geneva 23, Switzerland*

E-mail: henry.day-hall@desy.de

ABSTRACT: We present CALOCLOUDS3, a model for the fast simulation of photon showers in the barrel of a high granularity detector. This iteration demonstrates for the first time how a pointcloud model can employ angular conditioning to replicate photons at all incident angles. Showers produced by this model can be used across the whole detector barrel, due to specially produced position agnostic training data. With this flexibility, the model is usable in a full simulation and reconstruction chain, which offers a further handle for evaluating physics performance of the model. As inference time is a crucial consideration for a generative model, the pre-processing and hyperparameters are aggressively optimised, achieving a speed up factor of two orders of magnitude over GEANT4 at inference.

KEYWORDS: Detector modelling and simulations I, Simulation methods and programs, Calorimeter methods

¹Corresponding author. Authorship is alphabetical.

Contents

1	Introduction and Motivation	1
2	Dataset	3
2.1	Detector	3
2.2	Generation and preprocessing	4
3	CALOCLOUDS Architecture	5
3.1	Previous CALOCLOUDS2 model	5
3.2	Upgrade methodology	5
3.3	CALOCLOUDS3 architecture	7
4	Physics Accuracy	9
4.1	Impact of the changes from CALOCLOUDS2 to CALOCLOUDS3	9
4.2	Performance of CALOCLOUDS3 at random angles and fixed energies	12
4.3	Angular reconstruction	14
5	Di-Photon Separation	17
6	Inference Speed	17
7	Conclusions	19
	Appendices	19
A	CALOCLOUDS2 Architecture	19
B	Further plots of all angles at fixed energies	21
C	Angular error in PCA	22
8	Acknowledgements	23

1 Introduction and Motivation

Comparison between events observed in particle colliders and predictions made by theoretical models is an inverse problem. We wish to assess the likelihood of the theory given the observations, but it is only possible to simulate samples of likely observations, given the theory. Ensuring the statistical uncertainty of the samples does not dominate the analysis requires a vast quantity of simulation. In experimental particle physics the detailed simulation of the response to particles in calorimeter detectors is the single most resource demanding step of such large Monte Carlo

simulations. Generating this with the computational resources available requires the use of fast simulation methods, such as generative Machine Learning (ML) methods [1]. Some of those include generative adversarial networks (GANs) [2–13], variational autoencoders (VAEs) and their variants [13–19], generative pre-trained transformer (GPT) style models [20], normalizing flows and diffusion models [18, 21–42].

In this paper, we present the latest of a series of models designed for the fast simulation of photon showers in a high granularity calorimeter. This is the latest of the CALOCLOUDS series of models, which represent the showers to be modeled as pointclouds. Two ML techniques are used; a normalising flow combined with a diffusion model. For a review covering other possible architectures, see [43].

GEANT4 [44] is a particle transport Monte Carlo toolkit for the detailed simulation of the interaction of particles with matter based on a detailed modelling of the underlying physics. It is used by the majority of experiments in particle physics but requires large compute resources, especially in calorimeter simulation. Fast simulation aims to closely approximate the output of GEANT4, at greater speed than is possible with the detailed Monte Carlo techniques employed by GEANT4.

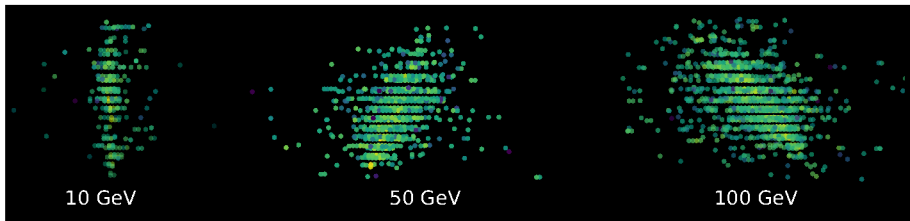


Figure 1. Photon showers, with varying incident angle, in the top barrel segment of the ILD SiECAL as simulated by GEANT4. The incident energy of the photon gun is indicated beneath each shower.

This paper describes the development of the current model from its predecessor. We aim to both justify the conclusion that the model is an optimal example of its architecture, and provide perspectives useful to optimising other models.

The previous model, CALOCLOUDS2, was only capable of simulating photons incident at a single angle and position. Incident angle conditioning, along with location agnostic training data, allows a single model to be used for all photons entering the calorimeter, which simplifies training, and potentially occupies less memory at inference time. The octagonal barrel shape and magnetic field, which is aligned with the beam axis, break the rotational symmetry about the shower axis, therefore the model must be conditioned on both θ and ϕ directions separately.

Clearly this upgrade is needed for the model to be used in realistic simulation, but it also helps to judge the fidelity with which the physics is modelled. Evaluating the success of a generative model is often challenging. We know the model must replicate the shower development of GEANT4, but it is difficult to decide on an acceptable precision. Even GEANT4 produces some changes between versions [45], and perfectly replicating GEANT4 would likely be as slow as GEANT4. With angular conditioning, CALOCLOUDS3 is viable to be used in a realistic detector simulation; full GEANT4 simulation is performed until a photon reaches the calorimeter. Assuming the photon satisfies the appropriate kinematic and geometrical constraints, it is then removed from the detailed full

simulation, and replaced by the corresponding output of CALOCLOUDS3. We then run a standard reconstruction on these simulated events, allowing for realistic physics benchmarking of the ML model and thereby providing the ultimate test of the model performance.

While inference speed for CALOCLOUDS2 was already ≈ 40 times faster than GEANT4 for 100 GeV photons, the model hyperparameters were not chosen in a systematic manner, and were inspired by choices made in other domains. Like many ML tools, diffusion models were originally developed for use on natural images [46]. Normalising flows were initially proposed in 2013 [47], but without the involvement of neural networks. Later variants that did use neural networks popularised the method [48, 49], and were also developed for image processing. This has influenced standard hyperparameter choices, so it is likely that a fundamentally different task like ours might benefit from significant changes to these choices.

In section 2 we describe the example detector we have chosen to demonstrate our architecture on, and the dataset generation is detailed. In section 3 the new CALOCLOUDS3 model is defined. In sections 4 and 5 we compare physics accuracy between the previous iteration, CALOCLOUDS2, and the new model, CALOCLOUDS3. After this, a snapshot of the newly attained capacity for full reconstruction is offered in section 5, however this is covered in more depth in a companion paper. In the penultimate section, section 6, the improvements in inference speed between CALOCLOUDS2 and CALOCLOUDS3 are demonstrated. Finally, we draw the conclusions in section 7.

2 Dataset

While the techniques and model architecture described here is expected to be a good solution for photon simulation in any high granularity calorimeter, we require a concrete example for training and testing. First, section 2.1 specifies the example detector chosen to generate the data. Then section 2.2 describes the preprocessing used to optimally interface the data and the model.

2.1 Detector

The International Large Detector (ILD) [50] is chosen as an example, as it has a calorimeter typical of a Higgs Factory detector. Higgs factories are designed to do precision physics, colliding fundamental particles (often electron-positron pairs) at energies in a tight range around the di-Higgs production mass. The ILD is designed with a normal range of sub-detectors; a vertex detector, a tracking detector, a time projection chamber, calorimeters and a muon system. However, in this study we are focused only on the calorimeters. Specifically, we consider the silicon variant of the electromagnetic calorimeter (denoted SiECAL in the ILD report [50], see section 5.1.2.4). Overall, the barrel is constructed of 8 flat segments, forming an octagon. As all segments have rotational symmetry, all simulated data is focused on the top segment. Within each segment are 30 sensitive layers, composed of active Si-cells of 5×5 mm, with a thickness of 0.5mm, interspersed at every other layer with passive Tungsten absorbers. The Tungsten absorbers in the first 20 layers are 2.1 mm thick, while the final 10 layers have thicker 4.2 mm absorbers. Such a high granularity calorimeter would provide sufficient detail to accurately reconstruct photons as low as 0.1 GeV, and while photons of energy $\lesssim 10$ GeV do not take so long to generate with GEANT4, replacing the higher energies with a faster technique would be essential. A magnetic field of 3.5 T parallel to

the beam encompass the detector, and the curvature this causes in charged particles is seen in the shower development.

2.2 Generation and preprocessing

Photons between 1 and 127 GeV were generated using GEANT4, a slightly expanded range from the energies considered in CALOCLOUDS2. In the framework used here, a fast model is only utilised (triggered) once a photon in an appropriate energy range reaches the surface of the calorimeter, therefore any photons that start showering before the surface of the calorimeter will not be simulated with the fast simulation model¹ therefore it is important to generate the datasets used to train the model without any pollution for early showering photons². To avoid any early showering, the photons in the training set are inserted into the simulation at the surface of the calorimeter. The angle of each photon is consistent with having originated from the Interaction Point (IP), having $30 < \theta < 150$ and $80 < \phi < 110$, sampled to have an even density on the upper barrel segment. GEANT4 then records the hits of the shower in the sensitive layers. The training data includes 3×10^6 events, with a test set of 7×10^4 which was held back.

As the objective is to use the same shower model for the entire detector, the training data must not contain specific information about support structures or dead zones. Additionally, the edges of each segment of the hexagonal barrel join at a 135° angle, meaning each layer starts at differing positions and calorimeter tiles are somewhat staggered between layers. We would prefer that the shower model does not learn these localised effects. For this reason, the detector is regularised; all supporting structure that interrupts the plane of the layers is removed, and the cells are aligned to form an identical grid in each layer. The shower model is then trained on this, location agnostic, regularised geometry. At inference time, the effects of the detector geometry at the relevant position are recovered by projecting the showers produced by the model back into a realistic detector map, as described in the next subsection 3.3.

A local coordinate system for the shower is defined with the z axis going through the sensitive layers, both the x and y coordinates are parallel to the surface of the layers, with the x coordinate also parallel to the beam direction.

The centre of each layer is shifted in the x and y directions to remove the shower's tilt. This shift corresponds to the rotation that would align the incident photon with the z axis, so the size of the shift is calculated from the depth of the layer d_{layer} the momentum of the incident particle \vec{p}_{inc} ; in the x direction, the shift is $\Delta x = d_{\text{layer}} \tan(\sin^{-1}(\vec{p}_{\text{inc},x}/|\vec{p}_{\text{inc}}|))$, and likewise for the y direction. To first approximation, the shifted showers are now a similar oblong shape, somewhat akin to one created by an upward moving particle. This preprocessing helps the diffusion model to learn the details of backscattering, rather than being forced to focus on getting the gross shape right, which is known from the direction of the incident particle.

Finally, all 30 layers of the detector are trimmed to a box of 100×100 cells, cutting to ± 250 mm about the shower axis. To strike a balance between the fidelity of all GEANT4 hits, and computational efficiency, the energy in each layer is binned into a grid of 500×500 bins (each cell is split into

¹Potentially, some secondary photons from the shower of an early showering photon could reach the calorimeter surface with sufficient energy to trigger the fast simulation, but this would anyway have the same physics as a low energy primary photon, and so is covered by the dataset.

²For further description of how the model is triggered in the context of a complete event simulation, see section 3.3.

5×5 bins). This is a slightly coarser granularity than that of CALOCLOUDS2, and was implemented because lowering the number of points was seen to have minimal impact on the kinematics after projection, while improving the computational efficiency of the model. As with CALOCLOUDS2, in each event, this grid is given a uniform random offset of up to half a cell, to reduce discretisation effects. A point with the total energy contained in the bin is created at the location of the highest energy point in the bin. This creates a set of up to 6,000 points per shower.

	CALOCLOUDS2	CALOCLOUDS3
Photon energy range	10 to 100 GeV	1 to 127 GeV
Photon direction	Directly upwards	Uniformly sampled in upper barrel segment
Detector geometry	Real geometry with support structure and staggered cells	Regularised geometry: no gaps in layers for supports
Layer size (in mm)	400 mm \times 400 mm	500 mm \times 500 mm
Bin granularity (per cell)	6 \times 6 bins per cell	5 \times 5 bins per cell
Bin to point conversion	Point placed at centre of bin	Point placed at highest-energy location in bin

Table 1. Summary of updates to training data for CALOCLOUDS3

3 CALOCLOUDS Architecture

A wide variety of approaches to fast calorimeter simulation have been attempted. CALOCLOUDS explores a point based approach. This has proven particularly advantageous for high granularity calorimeters, where cell activity is typically sparse and detailed, meaning that grid based approaches would require a prohibitively large number of voxels and parametrised approaches would struggle to replicate the fine structure.

3.1 Previous CALOCLOUDS2 model

A starting point of the existing CALOCLOUDS2 model is taken. A complete description, along with a description of the training data, is available in the existing paper [51], or a somewhat abbreviated description can be found in appendix A.

The key points for the following discussion are that the CALOCLOUDS models are composed of a normalising flow, referred to as SHOWERFLOW, which determines the overall structure of each photon shower, and a diffusion model, which generates each point. Both models are conditioned to create behaviour appropriate for the incident photon. The previous CALOCLOUDS2 model was only capable of simulating photons at a fixed position, and the model’s hyperparameters were adopted from standard practice for these models.

3.2 Upgrade methodology

In this section we describe two elements that significantly shaped choices made in developing CALOCLOUDS3.

Methodology for angular conditioning. Conditioning on the incident angle is a relatively simple change. In CALOCLOUDS2 the normalising flow was conditioned on E_{inc} , the incident particle energy, in CALOCLOUDS3 it will be conditioned on $(E_{inc}, \vec{p}_{inc}/|\vec{p}_{inc}|)$, the incident energy and normalised momentum direction. Likewise, the diffusion model of CALOCLOUDS3 will be conditioned on $(E_{inc}, \vec{p}_{inc}/|\vec{p}_{inc}|)$. Conditioning the diffusion model on the number of points in the shower, used in CALOCLOUDS2, was found to not have significant effect and therefore removed.

The decision to use Cartesian coordinates for the direction, rather than a pair of angles, was driven by a desire to avoid degeneracy in the input specifications. Consider a local polar angle, measuring the deviation of the incident photon from one that travels at 90° to the surface of the layers. When this local polar angle is 0 there are an infinite number of azimuthal angles specifying the same direction, which is challenging for interpolation. We do observe some minor instability at $(0, 0, 1)$, as this point is singular. Occasionally, this leads to unphysical events from the SHOWERFLOW (with many negative energies or number of points) but it is rare enough (around 0.1% of the samples) to be handled by rejection sampling.

Methodology for efficiency improvements. In CALOCLOUDS2, the inference time of the model was dominated by the normalising flow, SHOWERFLOW. While the number of parameters in both the normalising flow and the diffusion model have been reduced for CALOCLOUDS3, the focus has been on the flow model.

In order to limit the computational resources (and therefore the carbon footprint) required to investigate variations of normalising flow model, some toy models, representing expected data characteristics were employed. These small scale experiments focused on investigating two potential properties of a probability distribution function (pdf); sub manifolds and first or second order discontinuities. Physics data might be expected to contain sub manifolds, as the number of output parameters generally exceeds the input parameters, and would be expected to be approximately smooth, with few first order discontinuities. For example, the pdf of the energy in each layer of the detector would be expected to vary smoothly with incident particle energy, and have an approximate submanifold defined by adjacent layers having similar energies. Toy datasets were created from various pdfs, with and without the aforementioned properties. By observing the training and goodness of fit for normalising flow models with a similar architecture on those toy datasets, we could economically obtain indications of what might be expected in more complex cases. While this is far from conclusive, it helped guide development of the more expensive training process of the full model.

All expected characteristics of the physics data favoured the use of simple normalising flow models. Reducing the parameter space of our SHOWERFLOW model, it was seen that the number of parameters could be safely reduced by a factor of 5 (but that a factor of 10 was detrimental to performance). In a similar fashion, the number of parameters in the diffusion model was reduced by a factor of 4.

While the toy models showed no overwhelming failure to learn due to manifold overfitting, submanifolds did somewhat impair training stability. Identifying and removing submanifolds in the normalising flow of our physical model still offered a marginal performance increase. Comparing the training process of SHOWERFLOW in CALOCLOUDS2 and CALOCLOUDS3 (figure 2), there are indications that a more stable training is taking place in CALOCLOUDS3.

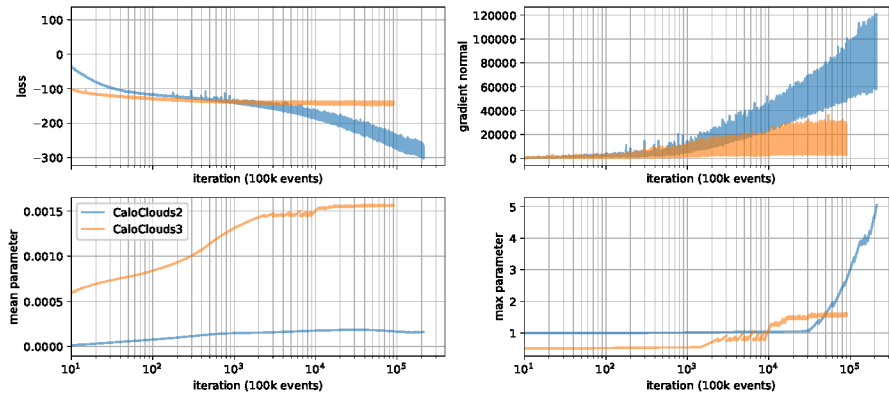


Figure 2. Differences observed while training the SHOWERFLOW model for CALOCLOUDS2 and CALOCLOUDS3. In the top left, we observe that the training loss of CALOCLOUDS2 never plateaus. In the top right, we see that the gradient normal of CALOCLOUDS3 stabilised during the training, but the gradient normal of CALOCLOUDS2 continues to increase. The relative magnitudes of the mean and maximum parameters in CALOCLOUDS3 remain in fairly constant relationship.

3.3 CALOCLOUDS3 architecture

This section describes the dataset and finalised design choices for the CALOCLOUDS3 model. It highlights changes made since the previous CALOCLOUDS2 model.

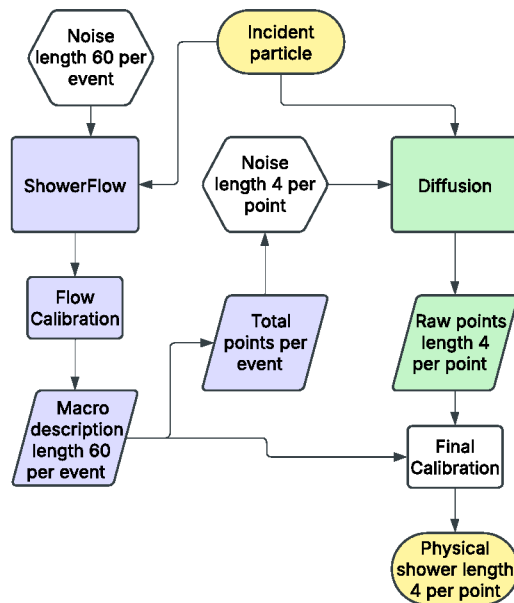


Figure 3. The model architecture of CALOCLOUDS3. Parts belonging to SHOWERFLOW are in purple, and those belonging to the diffusion model are in green. A full description is provided in section 3.3.

As with CALOCLOUDS2, the model is composed of a normalising flow, referred to as SHOWER-

FLOW, (shown in purple in figure 3) and a distilled diffusion model (shown in green in figure 3). SHOWERFLOW is now limited to distributions along the shower axis, specifically the points and energy per layer, and the diffusion model is still used to create individual points. Both are calibrated by $(E_{inc}, \vec{p}_{inc}/|\vec{p}_{inc}|)$, the incident energy and normalised momentum direction of the incident photon. The decision to use Cartesian coordinates (rather than angles) is detailed in section 3.2.

In CALOCLOUDS3, the information predicted by SHOWERFLOW is

- 1 to 30 The number of points in each layer divided by a fixed scaling factor of 7864 which rescales the majority of the distribution to be between 0 and 1.
- 30 to 60 The visible energy in each layer, divided by a fixed scaling factor of 3.4 that rescales the majority of the distribution to be between 0 and 1.

The reductions from the design of SHOWERFLOW in CALOCLOUDS2 (see appendix A) remove redundancies that created submanifolds in the normalising flow. The flow interleaves 12 affine couplings and 2 spline couplings. This has been reduced from the 60 affine couplings and 10 spline couplings used in CALOCLOUDS2, as a much simpler flow was seen to be sufficient to capture the complexity of the distribution, while being faster at inference time and more stable to train. A variant of SHOWERFLOW that predicted the log of the energy and points per layer was experimented with, but it was found to be unstable, and so not ultimately pursued.

The CoG of the shower is no longer predicted by SHOWERFLOW, the reasoning for which is discussed later in this section.

The SHOWERFLOW model is trained using an Adam [52] optimiser, with a learning rate of 0.0002. The batch size is 62 and a linear scheduler is used. Due to the high volume of training data, overfitting is not observed.

This output is reverted to physical values in the *Flow Calibration*, which yields the number of points in each layer, and the energy in each layer.

Then, like in CALOCLOUDS2, a diffusion model which has been distilled to a single step³ is used to generate individual points as an iid distribution. Unlike CALOCLOUDS2, the diffusion model is no longer conditioned on the total number of points. Removing this conditioning was seen to have no impact on the physics performance of the model; the conditioning is the same as for SHOWERFLOW (incident particle energy and direction). Raw points express their energy on a log scale, such that the energy spectrum of the points covers all real numbers. These raw points are calibrated in the *Final Calibration*. Starting from the first layer the photon reaches and working outwards:

1. The number of points predicted by SHOWERFLOW for this layer are selected from the unused points generated from the diffusion model, starting with those generated at the base of the shower. The selected points are moved along the shower axis to sit at the centre of the layer.
2. The total energy of this layer is rescaled to match the energy predicted by SHOWERFLOW.

This yields a shower in physical units, which can be projected back into detector cells. Two fixed scale factors are applied to correct for bias in occupancy and cell energy. This is simplified from CALOCLOUDS2, where occupancy was corrected with a polynomial fit.

³As with CALOCLOUDS2, distillation of the diffusion model is seen to have no adverse impact on the physics performance of the model.

Projection to detector geometry. On a basic level, detector readout can be mimicked by converting the energy deposits back to detector coordinates and allocating them to cells using records of the cell and layer boundaries. Energy in each cell is summed, and energy that falls into gaps between sensitive material is abandoned. This is done when comparing CALOCLOUDS2 to CALOCLOUDS3, as CALOCLOUDS2 does not have the angular conditioning capability required for use in full reconstruction.

When working with only CALOCLOUDS3 a more extensive framework can also facilitate reconstruction. The DD4HEP [53] toolkit exists to provide detector descriptions. Built on top of this, the DDSIM and DDML [54] libraries collectively run simulations. A full GEANT4 simulation can be run, or an augmented GEANT4 simulation, in which particles identified as suitable for a fast simulation model will be delegated to that fast simulation.

To use a fast simulation model in this manner, it must be compiled. This can be achieved using TORCHSCRIPT [55], which compiles the models in a language designed to be cross compatible with C++. Once the model has been compiled, triggers are defined using DDSIM's python interface, to replace photons in the full simulation that CALOCLOUDS3 is able to simulate.

Once simulation, using GEANT4 or GEANT4 and CALOCLOUDS3, is complete, the energy deposits created are projected back into the detector geometry supplied by DD4HEP to mimic real detector readout, and reconstruction tools are applied.

Removal of Centre of Gravity (CoG) calibration. Typically, the CoG is close to the incident particle axis i.e. the shower axis. When the CoG does deviate significantly from the shower axis, it has been shifted by high energy secondary particles scattered at wide angles. As the diffusion model draws points iid, these correlated structures cannot be replicated, and so the tails of the CoG distribution are not well modelled by the diffusion model.

Previously, in CALOCLOUDS2, SHOWERFLOW predicted values for the CoG perpendicular to the shower axis, and all points in the shower were shifted to match in calibration. This did achieve a good CoG distribution, however it is unphysical because the tails of the GEANT4 CoG distribution are not formed by shifting the bulk of the points, but by correlated, off centre, clusters of hits.

Given this objection, it was decided that the x and y centre of gravity that was naturally yielded by the diffusion model would be left as is. It will be seen in the results section that this only marginally degrades the centre of gravity profile in x and y .

4 Physics Accuracy

To evaluate the model, first a number of kinematics are compared on the unreconstructed detector output. Having significantly altered the model to allow angular conditioning and improve speed, we compare kinematics between CALOCLOUDS2 and CALOCLOUDS3 to verify that the performance has not been impaired. Then the performance of CALOCLOUDS3 is investigated at fixed energies, as varying the incident energy of the photon has a significant impact on the shower behaviour. Finally, the replication of shower angle is investigated.

4.1 Impact of the changes from CALOCLOUDS2 to CALOCLOUDS3

The fidelity of photons with orthogonal impact, between 10 and 90 GeV, is compared between CALOCLOUDS2, CALOCLOUDS3 and GEANT4, since this is the only region in which CALOCLOUDS2

can produce output. The results are projected into a realistic detector map, as described at the end of section 3.3, and key shower observables calculated – these observables are shown in figures 4, 5 and 8.

In the comparison, CALOCLOUDS2 is being trained on photons generated the same way as those used as benchmark, including all detector geometry effects, while the training dataset of CALOCLOUDS3 is as described in section 2.2.

This does give CALOCLOUDS2 an advantage in detector geometry, but this is unavoidable as CALOCLOUDS2 requires orthogonal photons for training. Barring one small change, it is trained in the same way as it was in its debut paper [56]. That is, photons with orthogonal impact, between 10 and 90 GeV, with slightly higher granularity and a realistic detector, including all supporting structure. However, when generating the previous dataset for training CALOCLOUDS2, the photon gun was placed just inside the calorimeter, before the first sensitive layer, but after the first absorber. This was an error, and it caused the first sensitive layer to be overpopulated, so the dataset for this work has been regenerated with the gun pulled back, to sit properly outside the calorimeter.

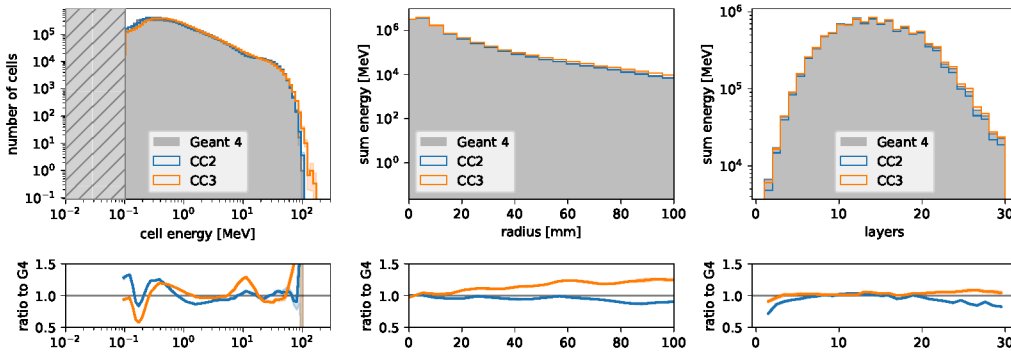


Figure 4. Energy distributions of CALOCLOUDS2 (CC2) and CALOCLOUDS3 (CC3), in comparison to GEANT4. Top row are the distributions themselves, averaged over three different seeds, with error bars from the standard deviation. Bottom row are ratio plots between each model and GEANT4, with errors propagated from the distributions. Left, histogram of the energy of each cell, with the MIP cut marked as a hatched box. Centre, radial profile of the energy. Right, distribution of the energy through the layers.

Figure 4 focuses on the energy profiles. The leftmost plot is a histogram with the energy of each cell, the far left of which is blocked off to represent the MIP (minimally ionising particle) cut. Cells with less than half the energy of a MIP are conventionally removed to reduce issues of electronic noise. Just to the right of this, at ≈ 0.2 MeV, the MIP peak can be seen. While neither model quite replicates the MIP peak, CALOCLOUDS2 is a little closer. On the right hand of this first plot, CALOCLOUDS3 is seen to slightly overstate the high energy tail of the cell energies. Overall, this plot is perhaps most impacted by the reduction in granularity of the training data, however the effect is still marginal.

The central plot of figure 4 shows the radial profile of the energy. One can observe a slight dip at 0 radius because the impact point is in the centre of a cell, so the first bin is only half a cell wide. Overall, both models replicate the GEANT4 data well, with CALOCLOUDS2 producing a marginally flatter ratio.

The right hand plot of figure 4 shows the energy per layer. This plot shows an alternating

pattern, as the silicon sensor layers are constructed in pairs, and mounted back to back on a tungsten absorber layer. CALOCLOUDS3 replicates the distribution the best, achieving a remarkably flat ratio throughout, although both models perform well.

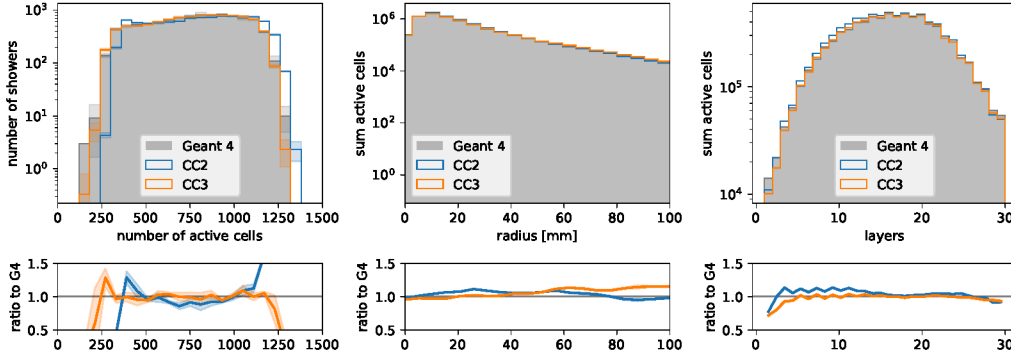


Figure 5. Occupancy distributions of CALOCLOUDS2 (CC2) and CALOCLOUDS3 (CC3), in comparison to GEANT4. Top row are the distributions themselves, averaged over three different seeds, with error bars defined by the standard deviation. Bottom row are ratio plots between each model and GEANT4, with errors propagated from the distributions. Left, histogram of the occupancy of the calorimeter in each shower. Centre, the number of active cells in concentric rings about the shower axis. Right, the number of active cells in each layer.

Figure 5 focuses on the occupancy profiles. The leftmost plot is a histogram with the occupancy of the calorimeter in each shower. As we simulate a continuous range of energies (10 GeV to 90 GeV), this distribution has a relatively flat profile. Both models are seen to perform well, but CALOCLOUDS2 has a slight tendency to overestimate the occupancy.

In the centre plot of figure 5, the number of active cells in concentric rings about the shower axis is shown. Towards 0 radius, bins encompass a smaller volume of the detector, creating a dip in the number of active cells (hence this isn't a true occupancy). In this metric, both models are excellent, with CALOCLOUDS3 recreating the centre of the distribution the best.

The rightmost of the three plots shows the number of active cells per layer (or the occupancy in each layer). The ability of the models to replicate this pattern is defined by the SHOWERFLOW component, which specifies the points in each layer. In this plot a slight pattern from the alternating absorber placement can be seen, in the same way as seen for the energy in figure 4, figure 5 shows an excess of hits in every other layer. The change in absorber thickness at layer 20 is also somewhat visible here. Both models replicate these effects well, but CALOCLOUDS3 has a flatter ratio to GEANT4.

Figure 6 shows the centre of gravity of the two models against the centre of gravity produced by GEANT4. Here, we expect worse performance in the perpendicular direction from CALOCLOUDS3, since it has had the centre of gravity conditioning from SHOWERFLOW removed, as per the reasoning given in section 3.3, and additionally, it does not have the advantage of training on the specific detector location with supporting structure. The first two plots, left and centre, show the perpendicular directions. As expected, CALOCLOUDS2 matches these better, with CALOCLOUDS3 slightly underestimating the width of the distribution, and not quite replicating the offset.

On the right, the centre of gravity along the shower axis is shown, expressed as a location

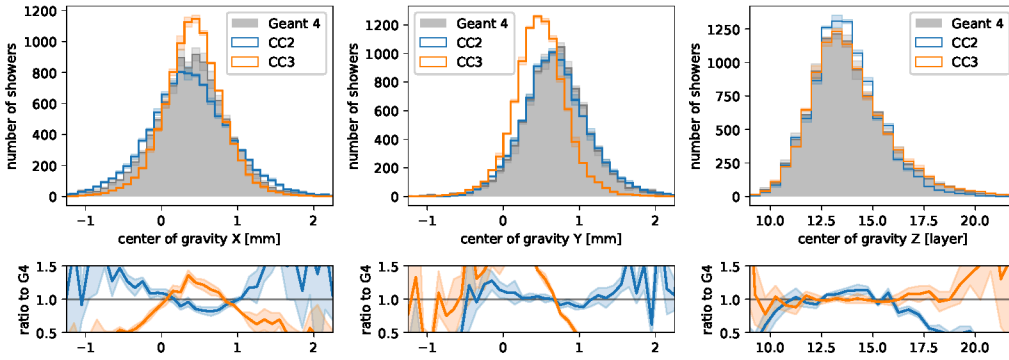


Figure 6. Centre of Gravity for showers from CALOCLOUDS2 (CC2) and CALOCLOUDS3 (CC3), in comparison to GEANT4. Top row are the distributions themselves, averaged over three different seeds, with error bars defined by the standard deviation. Bottom row are ratio plots between each model and GEANT4, with errors propagated from the distributions. Left and centre are the centre of gravity in x and y respectively, the directions perpendicular to the shower axis. Right is the centre of gravity parallel to the shower axis, expressed as a position in the 30 layers.

in the 30 layers. In this CALOCLOUDS3 provides a marginally better fit, but both models perform acceptably.

	CALOCLOUDS2	CALOCLOUDS3
total visible energy	$1.52(52) \times 10^{-3}$	$3.75(77) \times 10^{-3}$
energy per cell	$2.42(4) \times 10^{-3}$	$4.54(6) \times 10^{-3}$
energy per layer	$2.17(17) \times 10^{-4}$	$4.01(96) \times 10^{-5}$
energy per unit radius	$0.64(16) \times 10^{-4}$	$3.15(35) \times 10^{-4}$
total occupancy	$1.42(14) \times 10^{-2}$	$0.59(29) \times 10^{-3}$
occupancy per layer	$2.12(12) \times 10^{-4}$	$9.55(93) \times 10^{-5}$
occupancy per unit radius	$1.33(12) \times 10^{-4}$	$2.91(13) \times 10^{-4}$
center of gravity in X	$4.38(79) \times 10^{-3}$	$0.96(11) \times 10^{-2}$
center of gravity in Y	$2.20(48) \times 10^{-3}$	$3.05(19) \times 10^{-2}$
center of gravity in Z	$5.00(75) \times 10^{-3}$	$0.90(31) \times 10^{-3}$

Table 2. All metrics in this sub-section are summarised by calculating a Jensen Shannon divergence between GEANT4 and each model. The distance is calculated on histograms that combine all energies used, shown in appendix section B, as figures 16 and 17. An epsilon of 10^{-10} is added to empty bins to prevent division by zero. Errors have been propagated from the histograms. The value in bold is the better of the two scores.

Since the changes from CALOCLOUDS2 to CALOCLOUDS3 are significant, involving both a reduction in size, and the extension of the model to cover incident particles of all angles, It is satisfactory that the metrics in table 2 indicate comparable performance. Overall it appears that the upgrades to CALOCLOUDS3 have not adversely impacted the performance of the model.

4.2 Performance of CALOCLOUDS3 at random angles and fixed energies

Having established that CALOCLOUDS3 can replicate photons travelling along the z axis, the next assessment looks at photons all across a barrel segment to evaluate behaviour of CALOCLOUDS3

across all angles. The following observables are computed at the reconstruction level, after the model has been integrated into the simulation pipeline.

We now fix the energy of the incident particle in increments of 10 GeV from 10 GeV to 100 GeV. For legibility the histograms only present in full 10 GeV, 50 GeV and 100 GeV. All other energies are included in summary statistics plotting in figure 9 and plotted collectively in figure 17.

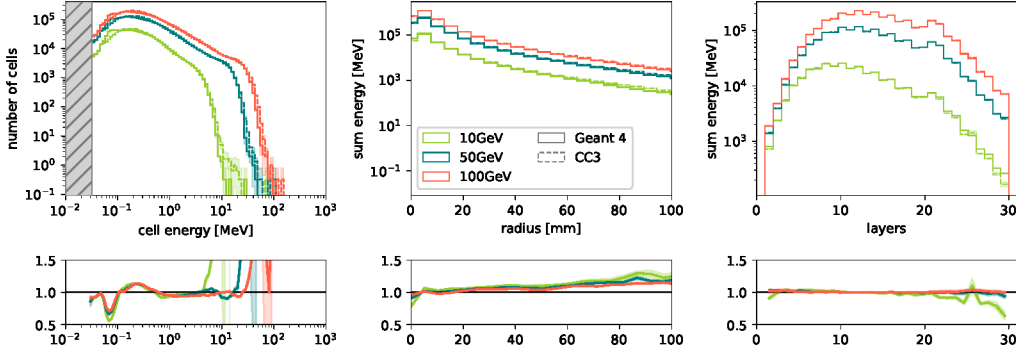


Figure 7. Energy spectrum of CALOCLOUDS3 (CC3) at random angles and fixed incident energies. For legibility, the plot only shows three of the ten photon incident energies tested; 10 GeV, 50 GeV and 100 GeV. Layout is the same as in figure 4.

Figure 7 shows examples of the different energy distributions. The leftmost plot shows the per-cell energy distribution, with the far left blocked off by the MIP cut applied in reconstruction. For the majority of the spectrum, all three energies produce a flat ratio, only somewhat overestimating the tail of high energy cells.

The central plot of figure 7 shows the radial profile of the energy, and in the right hand plot the energy per layer is seen. At all energies, these distributions are well replicated, but the ratio is seen to improve somewhat at higher incident particle energies.

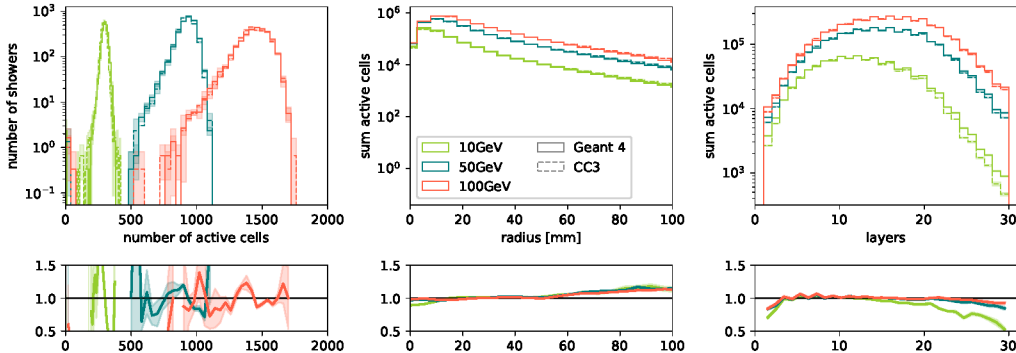


Figure 8. Occupancy spectrum of CALOCLOUDS3 (CC3) at random angles and fixed energies. For legibility, the plot only shows three of the ten photon incident energies tested; 10 GeV, 50 GeV and 100 GeV. Layout is the same as in figure 5.

Figure 8, shows examples of the occupancy spectrum of CALOCLOUDS3 at random angles and fixed energies. The leftmost plot is a histogram of total occupancy in each shower, while all three energies show reasonable agreement, it's clear that the model performs better at higher energies.

The central plot of figure 8 shows the radial profile of the occupancy, and in the right hand plot the occupancy per layer can be seen. At all energies, these distributions are well replicated, with a more pronounced deviation at 10 GeV in the occupancy per layer.

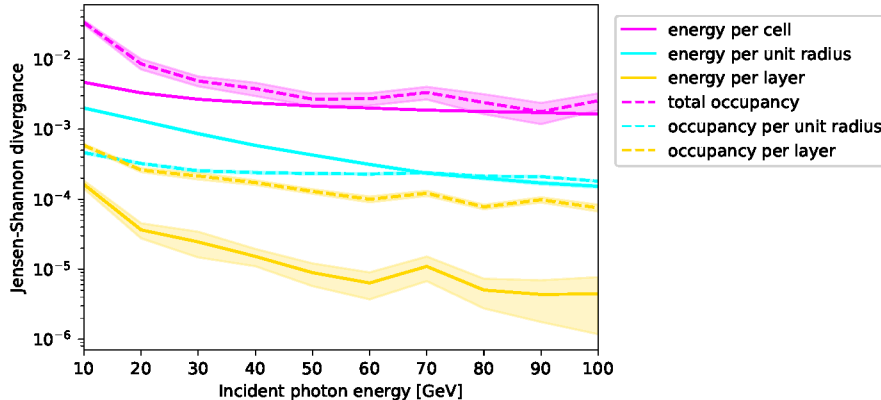


Figure 9. Kinematic performance of CALOCLOUDS3 at random angles and fixed energies. The Jensen Shannon divergence between GEANT4 and CALOCLOUDS3 for the energy spectrums (see figure 7) at each incident energy and the occupancy spectrums (see figure 8) are plotted against the incident photon energy.

	energy per cell	energy per layer	energy per unit radius	occupancy per layer	occupancy per unit radius	total occupancy
CALOCLOUDS3	$2.06(1) \times 10^{-3}$	$0.50(11) \times 10^{-5}$	$2.95(4) \times 10^{-4}$	$1.13(4) \times 10^{-4}$	$2.18(2) \times 10^{-4}$	$3.62(26) \times 10^{-3}$

Table 3. All of the plots in this sub-section are summarised by calculating a Jensen Shannon distance between GEANT4 and each model. An epsilon of 10^{-10} is added to empty bins to prevent division by zero. Errors have been propagated from the histograms.

For these comparisons we make two forms of summary statistic. The first is the Jensen Shannon distance between GEANT4 and CALOCLOUDS3 at each energy separately. This is plotted against the incident energy in figure 9. In all cases, the clear conclusion is that CALOCLOUDS3 performs best at higher energies. The main reason for this is that the diffusion model makes an iid assumption about the points it is creating, and this iid assumption is closest to the truth at higher energies. In figure 1, examples of showers produced by GEANT4 at 10 GeV, 50 GeV and 100 GeV are given. Examining these, more structure can be seen at 10 GeV, but by 100 GeV this structure is obscured by the number of hits. As CALOCLOUDS3 is incapable of replicating tracks and other internal hit correlations, it performs best at higher energies where such things have been washed out.

In addition, in table 3, we calculate a Jensen Shannon distance between the histograms of all energy increments combined. The histograms themselves can be found in the appendix section B.

Overall, the kinematics seem well matched in this dataset representing all angles.

4.3 Angular reconstruction

One further point that must be investigated is the replication of the internal angle of the shower. The internal angle of a particle shower is defined by the direction that would be reconstructed for that

shower's incident particle, without any assumptions on the origin of the incident particle. This is distinct from the angle of the incident particle that would be reconstructed with the assumption that the incident particle originated from the interaction point at the centre of the detector. For searches for long lived particles, such as a dark matter particle, the reconstruction of the internal angle is particularly important.

The dataset used for this comparison is the same as in section 4.2, but all ten energies are used in combination.

A common method for calculating the internal angle of the shower is to take the first component of an energy weighted PCA (principle component analysis) applied to all the reconstructed hits in the shower. When calculating the internal angle by PCA of hits, we observed systematic biases in the angle calculated in this way. This bias is caused by a combination of factors; undetected areas of secondary particles before the first layer of the detector, and the difference in path length of secondary particles raveling as shallow and steep angles. Further discussion, and illustration of these biases can be found in the appendix section C.

The bias can be quantified as an angular error;

$$\text{angular error} = \cos^{-1} \left(\frac{\vec{p}_{\text{PCA}} \cdot \vec{p}_{\text{inc}}}{|\vec{p}_{\text{PCA}}| |\vec{p}_{\text{inc}}|} \right).$$

Where \vec{p}_{PCA} is the direction of the photon reconstructed by PCA, and \vec{p}_{inc} is the true direction of the incident photon. Or alternatively, broken down into a ϕ component that rotates in the plane perpendicular the beam axis and a θ component that measures the angle to the beam axis;

$$\text{error in } \phi = \tan^{-1} \left(\frac{\vec{p}_{\text{PCA};x}}{\vec{p}_{\text{PCA};y}} \right) - \tan^{-1} \left(\frac{\vec{p}_{\text{inc};x}}{\vec{p}_{\text{inc};y}} \right), \quad \text{error in } \theta = \cos^{-1} \left(\frac{\vec{p}_{\text{PCA};z}}{|\vec{p}_{\text{PCA}}|} \right) - \cos^{-1} \left(\frac{\vec{p}_{\text{inc};z}}{|\vec{p}_{\text{inc}}|} \right).$$

As the bias on this error is due to off-axis secondary hits, the simplest way to remove it was to do a PCA with only the highest energy hits in the shower. To rephrase this, rather than calculating the PCA on all hits in the shower, a limited percentage of the highest energy hits are selected, and a PCA is calculated on this subset. A scan over the percentage used is shown in figure 10. It is seen that not only does the error itself decrease, but the variance of the error is also reduced. With this we justify our preference for calculating the internal angle of the shower using the highest 4% of hit energies in the shower.

Figure 11 shows the error on the internal angle of CALOCLOUDS3 and GEANT4, considering both the calculation utilising 100% of the hits, and the highest 4% of hit energies. The error in both directions, and in the absolute angle, is considerably improved by taking the highest 4% of hit energies.

When using only the highest 4% of hit energies, CALOCLOUDS3 produces an internal angle spectrum that is a much better match for GEANT4. This would indicate that while the tilt of the highest energy hits is well replicated, the lower energy hits do not have such a good angular placement in CALOCLOUDS3. Particularly striking, is the double peak in the θ angular error of CALOCLOUDS3 when calculating with 100% of the hits. The double peak is caused by backscattered hits tilting the shower towards the beam direction. Plots that better illustrate this concept can be seen in the appendix section C. The difference in the θ error spectrum specifically indicates that the backscattering hits are incorrectly placed by CALOCLOUDS3.

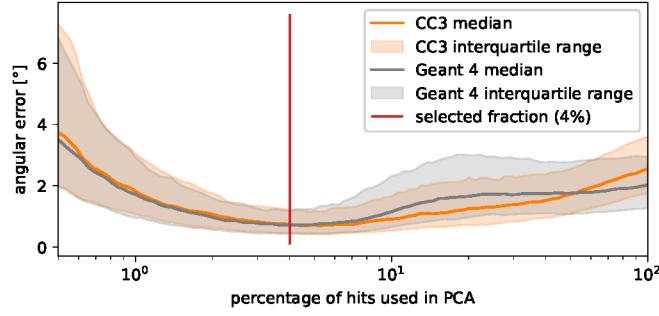


Figure 10. Error of the internal angle as a percentage of highest energy hits used in the energy weighted PCA calculation. The angular error is the angular distance between the direction of the first PCA component, and the true direction of the incident particle.

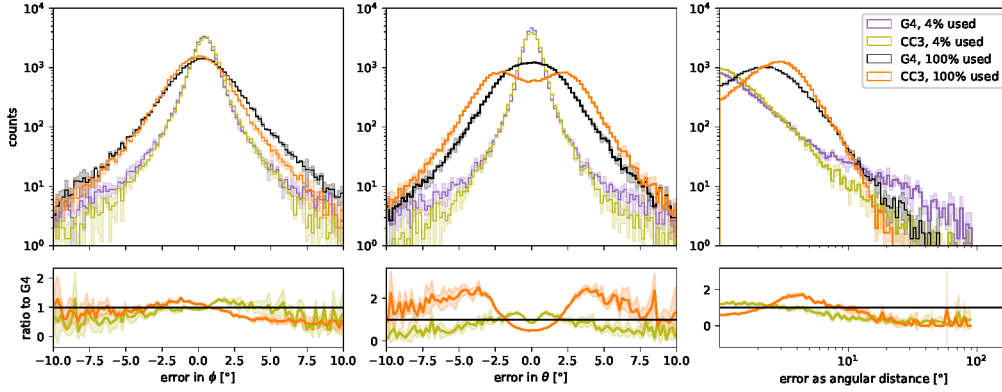


Figure 11. Angular errors of CALOCLOUDS3 compared to GEANT4. The left hand plot compares error in the azimuthal direction about the detector barrel, and the central plot compares error in the polar direction about the detector barrel (angle to the beam). The right hand plot compares the angular distance between the reconstructed internal angle and the true direction of the incident particle.

Despite the issues with modelling the angle of the lower energy hits, the agreement between CALOCLOUDS3 and GEANT4 is relatively good when calculating the internal angle using the classical PCA on 100% of the hits as can be seen in the ratio. When using the improved metric for internal angle, the agreement with GEANT4 becomes excellent. Summary statistics are shown in table 4. As the internal angle is a key feature in searches of long lived particles, replication of this metric provides additional use cases.

	ϕ	θ	angular error
4% used	$1.96(21) \times 10^{-3}$	$3.76(26) \times 10^{-3}$	$7.75(65) \times 10^{-3}$
100% used	$5.02(33) \times 10^{-3}$	$3.80(8) \times 10^{-2}$	$1.49(7) \times 10^{-2}$

Table 4. Jensen Shannon distances between the angular error spectrum as seen in figures 11 of CALOCLOUDS3 and GEANT4.

5 Di-Photon Separation

While replicating the kinematics is promising, it doesn't address the true purpose of a fast model, which is to replace the full simulation in a reconstruction. The only reliable way to evaluate a fast simulation model is to compare the output of a reconstruction applied to showers simulated with the fast model to the output of a reconstruction applied to `GEANT4` samples.

Reconstruction performance on a wider variety of metrics, across a range of fast simulation models, will be presented in a companion paper to this one. To validate this model with a basic example, we show the performance of `CALOCLOUDS3` on di-photon separation.

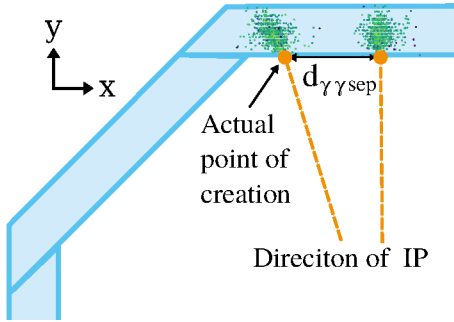


Figure 12. Schematic, describing the setup used to test di-photon separation. Two photons are created at random points (depicted as orange dots) at the surface of the ECAL (depicted in blue). Their direction is consistent with photons originating from the IP. The distance, $d_{\gamma\gamma sep}$, is measured at the surface of the calorimeter, where the photons are created.

The di-photon separation is a measure of how far apart two photon showers must be to both be resolved in reconstruction. It is computed by firing a pair of photons with the same energy at random distances in close proximity into a barrel segment, then using standard reconstruction software in `KEY4HEP` [57] the event is reconstructed and the number of photons are counted. A schematic of this, defining the distance used, can be seen in figure 12.

By basic physical constraints, two photons with exactly the same direction cannot be distinguished from one high energy photon. As the separation between a pair of photons grows, the likelihood of reconstructing both photons grows, until the likelihood of reconstructing two photons becomes the product of the likelihood of reconstructing each individual photon. The progression from at most a single reconstructed photon to two reconstructed photons depends strongly on the radial profile of the shower, and will only match `GEANT4` if it is described correctly.

In figure 13, we can see that `CALOCLOUDS3` reproduces the behaviour of `GEANT4` to within the fluctuation expected from the errors. In an analysis performed on samples generated by `CALOCLOUDS3` we have confidence that the photon multiplicity would be correctly reconstructed.

6 Inference Speed

Faster inference than `GEANT4` is a key objective. To fairly compare, we must run in the same software environment on the same hardware.

As detailed in section 3.3, as `CALOCLOUDS3` has angular conditioning, it can be run in a realistic simulation toolchain. Purely for timing purposes `CALOCLOUDS2` has also been compiled into the DDML stack; the angular conditioning is simply ignored. We use the embedding of the two models in order to benchmark the inference speed, using showers only at orthogonal impact.

The results are shown in Figure 14. This timing is done on a compiled model and includes hit placement into the detector geometry. Unfortunately, it is not currently possible to utilise a GPU

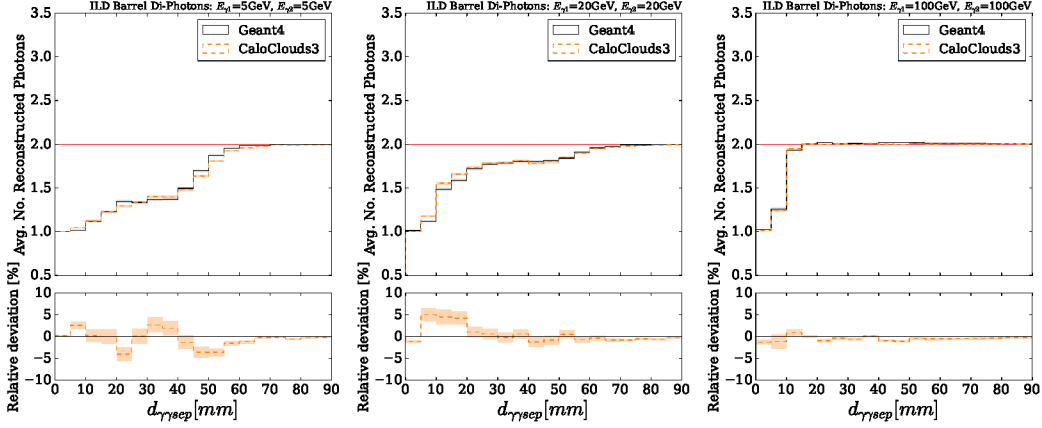


Figure 13. The reconstructed multiplicity of two photons fired into the same barrel segment, as modelled in CALOCLOUDS3 and GEANT4. Three different energy are shown, from left to right: 5 GeV, 20 GeV and 100 GeV.

in this setup, but this matches the current realistic use pattern, in which fast simulation is mainly done on CPU. The CPU used for our tests was a AMD EPYC 7513 32-Core Processor from DESY’s Maxwell [58] cluster.

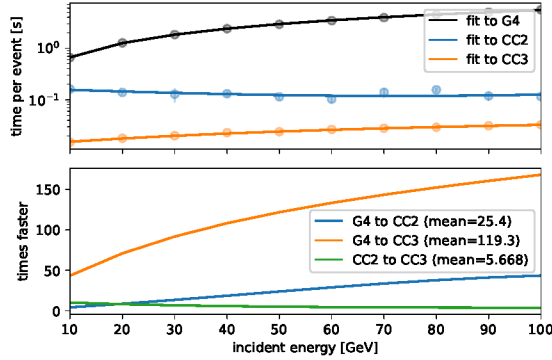


Figure 14. A comparison of the timing of CALOCLOUDS3 and CALOCLOUDS2 with respect to GEANT4. The upper plot includes error bars for each point, calculated as the standard deviation of 3 runs, however only the error bars for CALOCLOUDS2 are large enough to be visible, the reasons for this are unknown. The lines in the upper plot are a linear fit, and in the lower plot, the ratios of the fits are given. All timings are performed on a single thread of a single CPU, in the same cluster environment. In the legend, the mean speed increase across all incident energies is provided.

From this comparison, we see that CALOCLOUDS3 is on average over 6 times faster than CALOCLOUDS2, and two orders of magnitude faster than GEANT4. GEANT4 and CALOCLOUDS3 have roughly linear time complexity against the energy of the incident particles, which reflects the linear increase in number of points in the shower. A significant portion of the time for a CALOCLOUDS2 shower is spent in the SHOWERFLOW model, resulting in a large base time, even for low energy showers. By reducing the SHOWERFLOW model to the essentials, CALOCLOUDS3 has a much more economical base time and so continues to be advantageous at low energies. It is also seen that the

variation on the inference time of CALOCLOUDS2 is significantly larger than that of CALOCLOUDS3 or GEANT4. We did not anticipate this behaviour from CALOCLOUDS2, and do not know what causes it.

7 Conclusions

In this paper we have described the final iteration of the CALOCLOUDS series, CALOCLOUDS3. Within the constraints of this architecture, the hyperparameters of this model provide fastest inference without compromising accuracy. The improvements therefore facilitate more analysis work than otherwise possible, or potentially a reduction in the carbon footprint of the simulation.

By training the model in a location agnostic manner, and conditioning on incident angle in addition to energy, this model is applicable across the barrel region of the whole detector. It remains able to accurately recreate the kinematics of the photon shower with these changes, and the reconstructed metrics from the CALOCLOUDS3 fast simulation are compatible with those of a GEANT4 simulation, demonstrating that this model is ready to replace GEANT4 in a simulation chain.

Finally, this work demonstrates two broader concepts. Firstly, representing data as a point cloud offers significant advantages in high granularity calorimeter simulations; no other approach remains efficient in the sparse data space, while capturing the details of the high granularity calorimeter. Secondly, as a great majority of the models used in physics were designed for natural image processing, optimal hyper parameter choices may differ significantly from those typically used.

If the detector design itself were modified, we would expect ideal hyperparameters choices to also change, but in other respects the model is agnostic to the specifics. In the same vein, this architecture could be applied to the end-cap, requiring only an additional dataset, and some minimal hyperparameter adjustments.

In future work, we aim to apply these techniques to other common particles observed in the detector, and to investigate the potential for reconstruction using the physics knowledge of a fast simulation model.

Appendices

A CALOCLOUDS2 Architecture

In this appendix, we briefly describe the penultimate version, CALOCLOUDS2⁴. A complete description, along with a description of the training data, is available in the existing paper [51].

The model itself is composed of a normalising flow, referred to as SHOWERFLOW, (shown in purple in figure 15) and a distilled diffusion model (shown in green in figure 15).

SHOWERFLOW finds the macro features of the shower, and the distribution along the shower axis. It is conditioned on the incident energy of the photon, to produce a flow that transforms a vector of Gaussian noise into a vector whose elements predict the following 65 variables:

⁴Previously written as CALOCLOUDS II

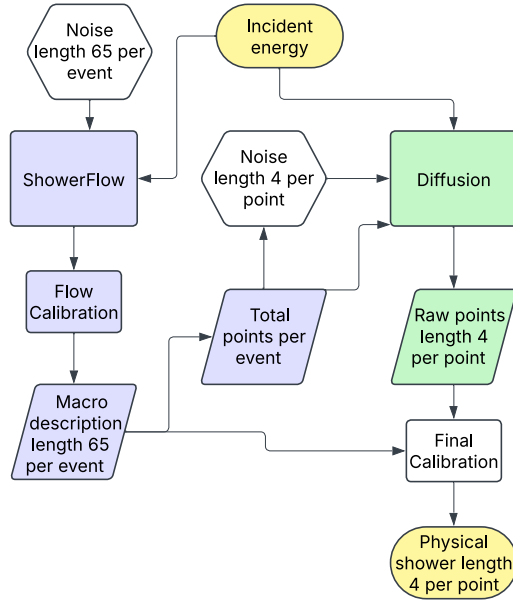


Figure 15. The model architecture of CALOCLOUDS2. Parts belonging to SHOWERFLOW are in purple, and those belonging to the diffusion model are in green. A brief description of the model is provided in appendix A, for further details see [51].

- 1 The total number of points in the shower, rescaled so that the dataset is roughly between 0 and 1.
- 2 The total visible energy of the shower, rescaled so that the dataset is roughly between 0 and 1, and clipped to a minimum of 40 MeV.
- 3 to 5 The Center of Gravity (CoG) of the shower in the x, y, z directions, rescaled to have a mean of 0 and standard deviation of 1.
- 6 to 35 The number of points in each layer divided by the maximum number of points in any layer, such that all numbers are between 0 and 1 inclusive.
- 36 to 65 The energy in each layer, divided by the maximum energy in any layer, such that all numbers are between 0 and 1 inclusive.

This output is used to calculate physical values in the *Flow Calibration*. The diffusion model is then conditioned on the incident particle energy and the total number of points predicted by SHOWERFLOW. The conditioned diffusion model creates a set of raw points, drawn from an iid⁵ distribution. This diffusion model has been distilled such that it generates points in a single network evaluation. Distillation was not seen to cause any significant loss of accuracy as shown in [56].

The raw points created by the diffusion model are calibrated in the *Final Calibration*. Starting from the first detector layer the photon encounters and working outwards:

⁵independent and identically distributed

1. The number of points predicted by SHOWERFLOW for this layer is selected and moved to the correct position along the shower axis.
2. The total energy of this layer is rescaled to match the energy predicted by SHOWERFLOW.

Once all layers are finished, the points are offset to match the CoG in x and y predicted by SHOWERFLOW. After projecting the energy into the detector cells, the occupancy of a shower generated by the model is not well matched to the GEANT4 result. A correction factor is applied to counter this; this correction factor modifies the number of hits requested from the diffusion model to using a polynomial function of the number of hits predicted by SHOWERFLOW. Finally, one scalar correction factor for the energy of the points is applied.

B Further plots of all angles at fixed energies

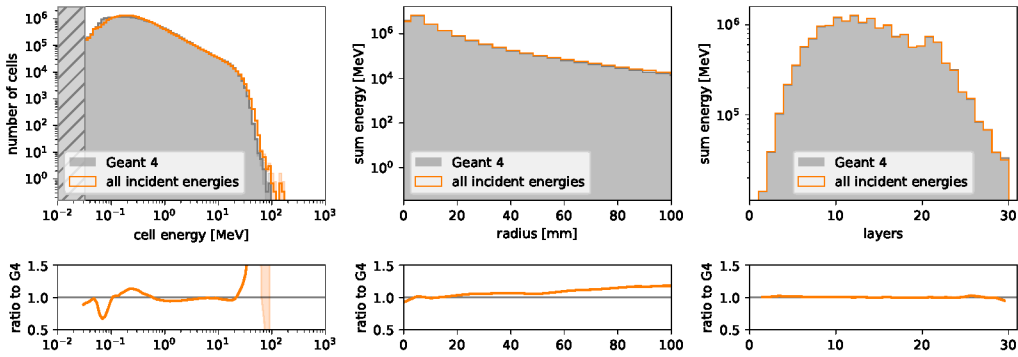


Figure 16. Energy spectrum of CALOCLOUDS3 at random angles, with all 10 energy increments combined. Layout is the same as in figure 4.

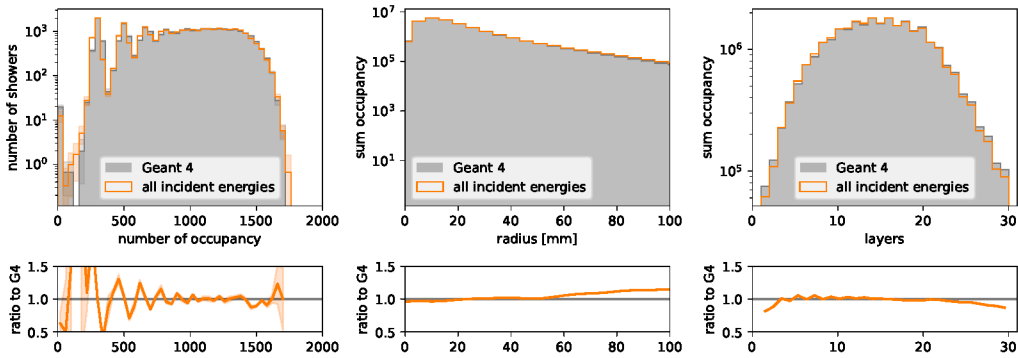


Figure 17. Occupancy spectrum of CALOCLOUDS3 at random angles, with all 10 energy increments combined. Layout is the same as in figure 5.

In section 4.2, we show kinematic performance metrics for CALOCLOUDS3 at three fixed energies, in a dataset with random angles. For completeness, plots containing the combination of all 10 energies are presented in figure 16 and figure 17. It is from these histograms that the metrics in table 3 are calculated.

C Angular error in PCA

In this section, we explore the structure of the error between the calculated internal angle and the true direction of the incident particle as described in section 4.3.

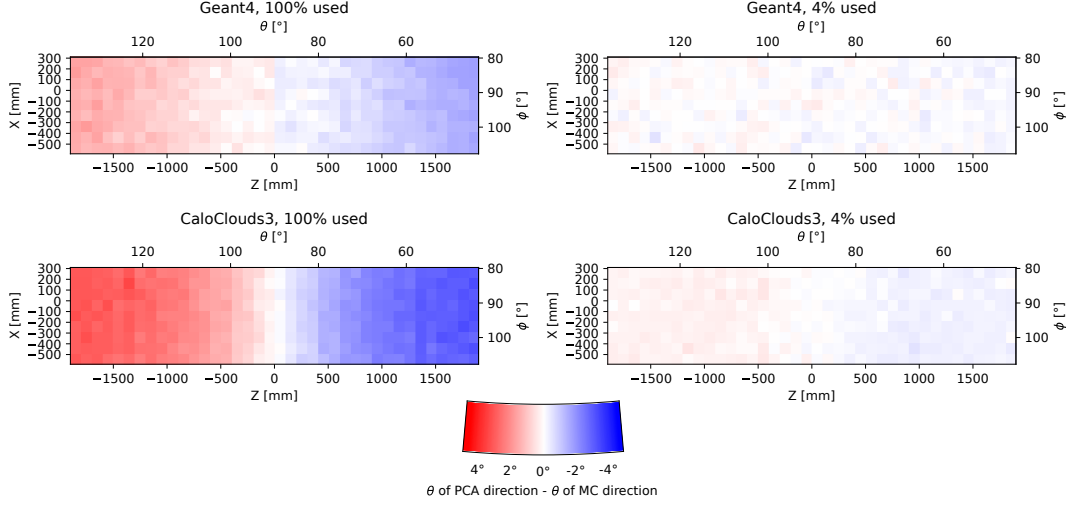


Figure 18. Heatmaps of the bias from the θ direction of the incident particle to the first PCA component. In the top row, the model generating the shower from which the PCA is calculated is `GEANT4`, and in the bottom row the model is `CALOCLOUDS3`. In the left column, 100% of the hits are used, and in the right column, 4% of the hits are used. The two distinct colour areas for `CALOCLOUDS3`, with 100%, create the two distinct peaks in figure 11.

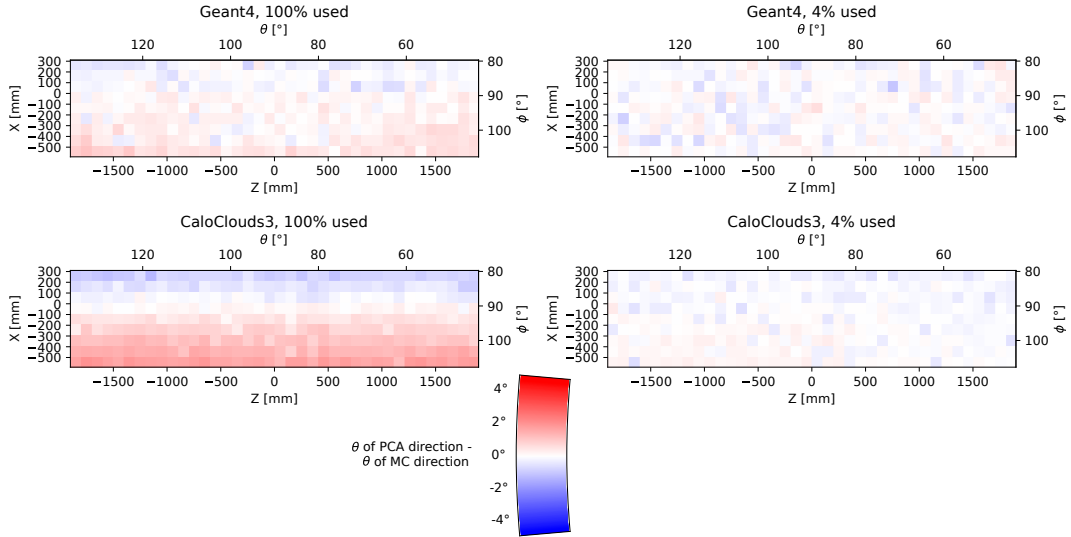


Figure 19. Heatmaps of the bias the ϕ direction of the incident particle to the first PCA component. In the top row, the model generating the shower from which the PCA is calculated is `GEANT4`, and in the bottom row the model is `CALOCLOUDS3`. In the left column, 100% of the hits are used, and in the right column, 4% of the hits are used. Given on the same colour scale as figure 18

Figure 18 shows the bias in the θ direction (polar angle) of the calculated internal angle, for

both GEANT4 and CALOCLOUDS3. In the first column, where 100% of the hits are used, the bias is clearly symmetric; in the positive θ direction, the error is positive, and in the negative θ direction, the error is negative. This is a physical result of the secondary particles encountering more active material at shallower angles, and clipping of the shower at the calorimeter surface, both of which rotate the first PCA component away from the true direction of the incident particle. This artefact is present in both GEANT4 and CALOCLOUDS3, but significantly more pronounced in CALOCLOUDS3, hence it has a double peak in figure 11.

From the second column, we see that when considering the PCA fit to only the 4% of highest energy hits, the effect is almost entirely eliminated.

In a similar way, bias in ϕ can be plotted, shown in figure 19. Here, in the left hand column we see similar angular effects, which bias the shower direction calculated by PCA towards shallower angles than the correct incident particle angle. In the left hand column, with only the highest 4% of hits, we see the backscattering effect vanish.

8 Acknowledgements

We thank Dirk Zerwas for the valuable comments on the manuscript. This research was supported in part by the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. This project has received funding from the European Union’s Horizon 2020 Research and Innovation programme under Grant Agreement No 101004761. We acknowledge support by the Deutsche Forschungsgemeinschaft under Germany’s Excellence Strategy – EXC 2121 Quantum Universe – 390833306 and via the KISS consortium (05D23GU4, 13D22CH5) funded by the German Federal Ministry of Education and Research BMBF in the ErUM-Data action plan. A.K. has received support from the Helmholtz Initiative and Networking Fund’s initiative for refugees as a refugee of the war in Ukraine. P.M. has benefited from support by the CERN Strategic R&D Programme on Technologies for Future Experiments [59].

References

- [1] ATLAS collaboration, *Overview of the atlas distributed computing system*, Tech. Rep. [ATL-SOFT-PROC-2018-009](#), CERN, Geneva (2019), [DOI](#).
- [2] M. Paganini, L. de Oliveira and B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters*, *Phys. Rev. Lett.* **120** (2018) 042003 [[1705.02355](#)].
- [3] M. Paganini, L. de Oliveira and B. Nachman, *CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks*, *Phys. Rev. D* **97** (2018) 014021 [[1712.10321](#)].
- [4] L. de Oliveira, M. Paganini and B. Nachman, *Controlling Physical Attributes in GAN-Accelerated Simulation of Electromagnetic Calorimeters*, *J. Phys. Conf. Ser.* **1085** (2018) 042017 [[1711.08813](#)].
- [5] M. Erdmann, L. Geiger, J. Glombitza and D. Schmidt, *Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks*, *Comput. Softw. Big Sci.* **2** (2018) 4 [[1802.03325](#)].
- [6] P. Musella and F. Pandolfi, *Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks*, *Comput. Softw. Big Sci.* **2** (2018) 8 [[1805.00850](#)].
- [7] M. Erdmann, J. Glombitza and T. Quast, *Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network*, *Comput. Softw. Big Sci.* **3** (2019) 4 [[1807.01954](#)].
- [8] D. Belayneh et al., *Calorimetry with deep learning: particle simulation and reconstruction for collider physics*, *Eur. Phys. J. C* **80** (2020) 688 [[1912.06794](#)].
- [9] A. Butter, S. Diefenbacher, G. Kasieczka, B. Nachman and T. Plehn, *GANplifying event samples*, *SciPost Phys.* **10** (2021) 139 [[2008.06545](#)].
- [10] ATLAS collaboration, *The Fast Simulation Chain in the ATLAS experiment*, *EPJ Web Conf.* **251** (2021) 03012.
- [11] S. Bieringer, A. Butter, S. Diefenbacher, E. Eren, F. Gaede, D. Hundhausen et al., *Calomplification — the power of generative calorimeter models*, *JINST* **17** (2022) P09028 [[2202.07352](#)].
- [12] B. Hashemi, N. Hartmann, S. Sharifzadeh, J. Kahn and T. Kuhr, *Ultra-high-granularity detector simulation with intra-event aware generative adversarial network and self-supervised relational reasoning*, *Nature Commun.* **15** (2024) 4916 [[2303.08046](#)].
- [13] ATLAS collaboration, *Deep Generative Models for Fast Photon Shower Simulation in ATLAS*, *Comput. Softw. Big Sci.* **8** (2024) 7 [[2210.06204](#)].
- [14] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol et al., *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed*, *Comput. Softw. Big Sci.* **5** (2021) 13 [[2005.05334](#)].
- [15] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol et al., *Decoding Photons: Physics in the Latent Space of a BIB-AE Generative Network*, *EPJ Web Conf.* **251** (2021) 03003 [[2102.12491](#)].
- [16] E. Buhmann, S. Diefenbacher, D. Hundhausen, G. Kasieczka, W. Korcari, E. Eren et al., *Hadrons, better, faster, stronger*, *Mach. Learn. Sci. Tech.* **3** (2022) 025014 [[2112.09709](#)].
- [17] J.C. Cresswell, B.L. Ross, G. Loaiza-Ganem, H. Reyes-Gonzalez, M. Letizia and A.L. Caterini, *CaloMan: Fast generation of calorimeter showers with density estimation on learned manifolds*, in

36th Conference on Neural Information Processing Systems: Workshop on Machine Learning and the Physical Sciences, 11, 2022 [2211.15380].

- [18] S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, K. Krüger et al., *New angles on fast calorimeter shower simulation*, *Mach. Learn. Sci. Tech.* **4** (2023) 035044 [2303.18150].
- [19] B. Hashemi, *Deep generative models for ultra-high granularity particle physics detector simulation: A voyage from emulation to extrapolation*, 2024.
- [20] J. Birk, F. Gaede, A. Hallin, G. Kasieczka, M. Mozzanica and H. Rose, *OmniJet- α_C : Learning point cloud calorimeter simulations using generative transformers*, 2501.05534.
- [21] J. Sohl-Dickstein, E.A. Weiss, N. Maheswaranathan and S. Ganguli, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, 1503.03585.
- [22] Y. Song and S. Ermon, *Generative modeling by estimating gradients of the data distribution*, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, (Red Hook, NY, USA), Curran Associates Inc. (2019).
- [23] Y. Song and S. Ermon, *Improved techniques for training score-based generative models*, in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin, eds., vol. 33, pp. 12438–12448, Curran Associates, Inc., 2020, https://proceedings.neurips.cc/paper_files/paper/2020/file/92c3b916311a5517d9290576e3ea37ad-Paper.pdf.
- [24] J. Ho, A. Jain and P. Abbeel, *Denoising Diffusion Probabilistic Models*, 2006.11239.
- [25] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon and B. Poole, *Score-Based Generative Modeling through Stochastic Differential Equations*, 2011.13456.
- [26] V. Mikuni and B. Nachman, *Score-based generative models for calorimeter shower simulation*, *Phys. Rev. D* **106** (2022) 092009 [2206.11898].
- [27] F.T. Acosta, V. Mikuni, B. Nachman, M. Arratia, B. Karki, R. Milton et al., *Comparison of point cloud and image-based models for calorimeter fast simulation*, *JINST* **19** (2024) P05003 [2307.04780].
- [28] V. Mikuni and B. Nachman, *CaloScore v2: single-shot calorimeter shower simulation with diffusion models*, *JINST* **19** (2024) P02001 [2308.03847].
- [29] O. Amram and K. Pedro, *Denoising diffusion models with geometry adaptation for high fidelity calorimeter simulation*, *Phys. Rev. D* **108** (2023) 072014 [2308.03876].
- [30] C. Chen, O. Cerri, T.Q. Nguyen, J.R. Vlimant and M. Pierini, *Analysis-Specific Fast Simulation at the LHC with Deep Learning*, *Comput. Softw. Big Sci.* **5** (2021) 15.
- [31] C. Krause and D. Shih, *Fast and accurate simulations of calorimeter showers with normalizing flows*, *Phys. Rev. D* **107** (2023) 113003 [2106.05285].
- [32] C. Krause and D. Shih, *Accelerating accurate simulations of calorimeter showers with normalizing flows and probability density distillation*, *Phys. Rev. D* **107** (2023) 113004 [2110.11377].
- [33] S. Schnake, D. Krücker and K. Borras, *CaloPointFlow II Generating Calorimeter Showers as Point Clouds*, 2403.15782.
- [34] C. Krause, I. Pang and D. Shih, *CaloFlow for CaloChallenge dataset I*, *SciPost Phys.* **16** (2024) 126 [2210.14245].
- [35] A. Xu, S. Han, X. Ju and H. Wang, *Generative machine learning for detector response modeling with a conditional normalizing flow*, *JINST* **19** (2024) P02003 [2303.10148].

- [36] M.R. Buckley, C. Krause, I. Pang and D. Shih, *Inductive simulation of calorimeter showers with normalizing flows*, *Phys. Rev. D* **109** (2024) 033006 [2305.11934].
- [37] F. Ernst, L. Favaro, C. Krause, T. Plehn and D. Shih, *Normalizing Flows for High-Dimensional Detector Simulations*, *SciPost Phys.* **18** (2025) 081 [2312.09290].
- [38] M. Omana Kuttan, K. Zhou, J. Steinheimer and H. Stöcker, *Towards a foundation model for heavy-ion collision experiments through point cloud diffusion*, **2412.10352**.
- [39] L. Favaro, A. Ore, S.P. Schweitzer and T. Plehn, *CaloDREAM – Detector Response Emulation via Attentive flow Matching*, *SciPost Phys.* **18** (2025) 088 [2405.09629].
- [40] J. Brehmer, V. Bresó, P. de Haan, T. Plehn, H. Qu, J. Spinner et al., *A Lorentz-Equivariant Transformer for All of the LHC*, **2411.00446**.
- [41] T. Buss, F. Gaede, G. Kasieczka, A. Korol, K. Krüger, P. McKeown et al., *CaloHadronic: a diffusion model for the generation of hadronic showers*, **2506.21720**.
- [42] P. Raikwar, A. Zaborowska, P. McKeown, R. Cardoso, M. Piorczynski and K. Yeo, *A Generalisable Generative Model for Multi-Detector Calorimeter Simulation*, **2509.07700**.
- [43] B. Hashemi and C. Krause, *Deep generative models for detector signature simulation: A taxonomic review*, *Reviews in Physics* **12** (2024) 100092.
- [44] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce et al., *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250.
- [45] M. Batič, G. Hoff and M.G. Pia, *Precision analysis of geant4 condensed transport effects on energy deposition in detectors*, *Journal of Physics: Conference Series* **396** (2012) 022004.
- [46] S. Lim, E. Yoon, T. Byun, T. Kang, S. Kim, K. Lee et al., *Score-based generative modeling through stochastic evolution equations in hilbert spaces*, in *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23*, (Red Hook, NY, USA), Curran Associates Inc., 2023.
- [47] E.G. Tabak and C.V. Turner, *A family of nonparametric density estimation algorithms*, *Communications on Pure and Applied Mathematics* **66** (2013) 145 [https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21423].
- [48] D.J. Rezende and S. Mohamed, *Variational inference with normalizing flows*, in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, p. 1530–1538, JMLR.org, 2015, DOI.
- [49] L. Dinh, D. Krueger and Y. Bengio, *Nice: Non-linear independent components estimation*, .
- [50] The ILD Collaboration, *International Large Detector: Interim Design Report*, Tech. Rep. AIDA-2020-NOTE-2020-004, CERN, Geneva (2020).
- [51] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol et al., *Caloclouds: fast geometry-independent highly-granular calorimeter simulation*, *Journal of Instrumentation* **18** (2023) P11025.
- [52] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *CoRR* **abs/1412.6980** (2014) .
- [53] M. Frank, F. Gaede, C. Grefe and P. Mato, *Dd4hep: A detector description toolkit for high energy physics experiments*, *Journal of Physics: Conference Series* **513** (2014) 022010.
- [54] Key4hep collaboration, “DDML: Ddfastshowerml – fast simulation using ml inference.” <https://github.com/key4hep/DDML>, 2025.

- [55] PyTorch Team, *Torchscript: An intermediate representation of a pytorch model*, 2023.
- [56] E. Buhmann, F. Gaede, G. Kasieczka, A. Korol, W. Korcari, K. Krüger et al., *Caloclouds ii: ultra-fast geometry-independent highly-granular calorimeter simulation*, *Journal of Instrumentation* **19** (2024) [P04020](#).
- [57] KEY4HEP collaboration, *The Key4hep turnkey software stack*, *PoS ICHEP2022* (2022) 234.
- [58] Deutsches Elektronen-Synchrotron (DESY), *Maxwell compute cluster: documentation*, 2025.
- [59] C. Joram et al., *Extension of the R&D Programme on Technologies for Future Experiments*, Tech. Rep. [CERN-EP-RDET-2023-001](#), CERN, <https://cds.cern.ch/record/2850809> (2023).