

UC Irvine

UC Irvine Previously Published Works

Title

Generalizing to new geometries with Geometry-Aware Autoregressive Models (GAAMs) for fast calorimeter simulation

Permalink

<https://escholarship.org/uc/item/5fh21376>

Journal

Journal of Instrumentation, 18(11)

ISSN

1748-0221

Authors

Liu, Junze

Ghosh, Aishik

Smith, Dylan

et al.

Publication Date

2023-11-01

DOI

10.1088/1748-0221/18/11/p11003

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Generalizing to new geometries with Geometry-Aware Autoregressive Models (GAAMs) for fast calorimeter simulation

Junze Liu,^a Aishik Ghosh,^{b,c} Dylan Smith,^b Pierre Baldi,^a Daniel Whiteson^b

^aDepartment of Computer Science, University of California, Irvine, CA 92697, USA

^bDepartment of Physics and Astronomy, University of California, Irvine, CA 92697, USA

^cPhysics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

E-mail: junzel1@uci.edu

ABSTRACT: Generation of simulated detector response to collision products is crucial to data analysis in particle physics, but computationally very expensive. One subdetector, the calorimeter, dominates the computational time due to the high granularity of its cells and complexity of the interactions. Generative models can provide more rapid sample production, but currently require significant effort to optimize performance for specific detector geometries, often requiring many models to describe the varying cell sizes and arrangements, without the ability to generalize to other geometries. We develop a *geometry-aware* autoregressive model, which learns how the calorimeter response varies with geometry, and is capable of generating simulated responses to unseen geometries without additional training. The geometry-aware model outperforms a baseline unaware model by over 50% in several metrics such as the Wasserstein distance between the generated and the true distributions of key quantities which summarize the simulated response. A single geometry-aware model could replace the hundreds of generative models currently designed for calorimeter simulation by physicists analyzing data collected at the Large Hadron Collider. This proof-of-concept study motivates the design of a foundational model that will be a crucial tool for the study of future detectors, dramatically reducing the large upfront investment usually needed to develop generative calorimeter models.

Contents

1	Introduction	1
2	Context and Related Work	3
3	Dataset	3
4	Generative Models	4
4.1	Geometry-Aware Autoregressive Model (GAAM)	5
4.2	Baseline: Unconditional ARM	6
4.3	Preprocessing	6
5	Results	6
5.1	Performance in uniform segmentation	7
5.2	Performance in non-uniform segmentation	7
5.3	Performance in unseen geometries	10
6	Conclusion and Outlook	11
A	Spiral Ordering	16
B	Individual Shower Images	16

1 Introduction

Collision of particles at high-energy, such as those at the Large Hadron Collider (LHC), provide clues about the fundamental nature of matter and its interactions. A crucial tool in experimental design and data analysis is the production of vast samples of high-fidelity simulations of the detectors which observe those collisions. Samples of simulated collisions allow for characterization of the expected response of future detectors as well as simulation-based-inference in data analysis, which often require billions of samples. The highest-fidelity simulations, via Geant4 [1–3], are produced by detailed time-evolution of individual particles and the microphysics of their interactions with detector elements, which can produce showers of additional particles. The exponentially increasing number of particles results in large computational expense, limiting the production of samples of simulated collisions. Efficiency of sample generation will become increasingly vital as the LHC enters an era of high-intensity beams which will produce higher volumes of data. However, the full time evolution of the shower is not observed or recorded, and strategies which directly generate the final cell responses can be significantly cheaper. Deep generative models (DGM) are a promising approach to automatically learning such a response function.

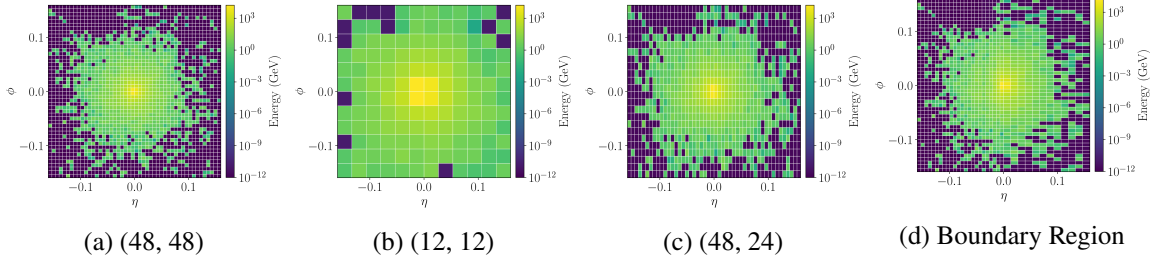


Figure 1: Simulated calorimeter response to a 65 GeV photon for four distinct calorimeter cell segmentations. Shown is the energy deposited per cell, for various segmentations in (η, ϕ) . The Boundary Region is (24, 48) for $\eta < 0$ and (12, 48) for $\eta > 0$.

Machine learning has been increasingly useful to solve challenging particle physics problems [4–8]. Many experiments, including those at the LHC, have already invested resources in developing DGMs designed to produce shower images for their own calorimeters [9–12]. However, the calorimeter geometries (i.e., the shape, position, and size of cells and their relative arrangements) not only differ greatly between experiments but can also vary widely from region to region within the same detector. As a result, experiments have to spend significant human resources developing DGM architectures for their specific calorimeter geometries. For example, the ATLAS experiment developed hundreds of generative networks to span different regions of their calorimeter [10]. Further, the geometry-specific generative model architectures developed by each experiment vary considerably. While there has recently been significant innovation aimed at improving the accuracy of shower simulations using DGMs for fixed calorimeter geometries [13–18], the key challenge remains the development of models capable of generating samples for varying calorimeter geometries, which is essential to scale these proof-of-concept demonstrations to a deployed product. This learning task has no analogy in natural images, where pixel sizes are uniform, and no off-the-shelf machine learning solutions are available. This problem requires a novel approach.

This paper addresses this crucial computational challenge, presenting a generative model which learns to interpolate across a training set with varying calorimeter geometries and is capable of zero-shot generalization to new geometries on which it has not been trained. This is achieved by training an autoregressive model that is aware of the properties of the calorimeter cell, adjusting the energy deposition based on the size and position of that cell. Such a model could reflect a generalized understanding of calorimeter response, to be shared across experiments with distinct geometries. This would enable individual experiments to fine-tune the model to their specific calorimeter, requiring smaller training samples and fewer computing resources to rapidly generate the samples of simulated collisions needed for data analysis. In addition, such a model would allow to rapidly characterize the scientific potential of future detectors. Such general-purpose, pre-trained models are often called ‘foundational models’.

The rest of this paper is organized as follows. Sec. 2 discusses related work, Sec. 3 describes the dataset used for training and evaluation, Sec. 4 details the ML methods compared, and Sec. 5 discusses the results, including success with interpolation and failure with extrapolation on novel geometries. The paper ends with a conclusion and future outlook in Sec. 6.

2 Context and Related Work

The ATLAS electromagnetic calorimeter has a complex geometry, and the simulation of its response is a computational bottleneck, motivating physicists to develop faster simulations. Early versions approximated the calorimeter response with parameterized response functions [19]. There has been significant innovation in developing DGMs for particle physics calorimeters using fixed output geometries, including with Generative Adversarial Networks (GAN) [13, 14], Variational Auto-Encoders (VAE) [15], Normalizing Flows (NF) [16, 17], and most recently diffusion models [18].

Conditional generative models have been explored, such as an early effort by the ATLAS collaboration [9], which attempts to condition a GAN on the entire calorimeter geometry at once. This ambitious effort faces the challenge of the enormous number of possible three-dimensional geometries, which requires generating a vast training dataset in Geant4, and presents difficulties in conditioning the GAN. On the other hand, training hundreds of GANs [10] for small sections of the calorimeter with fixed geometries raises questions about the sustainability of validating and maintaining so many models. Experiments like CMS and LHCb have also studied DGMs for their calorimeters [11, 12], and related studies have been made for non-LHC experiments [20].

Instead of conditioning a DGM on all possible unique geometries, it is more natural to condition it on the properties of individual calorimeter cells. In this way, the model can be expected to interpolate between different cell sizes and positions that it has encountered in training, and learn from multiple cells in the same image. An autoregressive algorithm, which generates cells sequentially, provides an efficient solution by conditioning the model solely on the properties of the cell being generated and its neighbors. This cell-by-cell approach enables the model to learn more efficiently how the properties of each cell affect its energy distribution. Recently, Sparse Autoregressive Models were studied for simulating sparse calorimeter images of a fixed geometry, in which the model learns the sparsity of the model with a tractable likelihood [21]. This model serves as the inspiration for our work. The details of the construction of the model are described in Sec. 4.

3 Dataset

Samples of simulated electromagnetic calorimeter response are generated using Geant4 [1–3], in an ATLAS-like configuration modified from that used to generate earlier datasets [22]. The detector transverse segmentation (coordinates η and ϕ) was modified, to significantly increase the granularity of the calorimeter, up to 16 fold, and approximate the granularity of potential future detectors. Several different configurations were used to generate responses from a wide set of hypothetical geometries. These configurations were inspired by the most challenging geometry configurations seen in the ATLAS electromagnetic calorimeter [23, 24].

The calorimeter comprises three longitudinal layers referred to as ‘inner,’ ‘middle,’ and ‘outer’ layers and denoted by $z = 1, 2, 3$, each with area $48 \times 48 \text{ cm}^2$. The layers have a length of 5mm, 40mm, and 80mm in η , 160mm, 40mm and 40mm in ϕ and a depth of 90mm, 347mm and 43mm respectively. For further details about the detector design, refer to Ref [22]. The calorimeter layers are segmented into various rectangular cells using a grid in (n_η, n_ϕ) , the number of cells along the η and ϕ directions, respectively; see Table 3. In one particular configuration, the middle layer is structured with two different cell sizes to represent a boundary between detector regions which

presents a difficult challenge for any generative model. In this non-uniform¹ (36, 48)* geometry, the right half of the image is segmented into a (12, 48) grid and the left half into (24, 48) cells in (n_η, n_ϕ) . The outermost layer had a consistent shape of (24, 24).

A sample of 10,000 simulated photons with energy of 65 GeV are directed at the center of the calorimeter. The photon interaction with the material generates a shower of electrons and photons which deposit energy in each layer. The middle layer receives the largest fraction of the photon energy, while the outer receives the least. The simulated response to a photon for four example calorimeter geometries are shown in Fig. 1, demonstrating the dramatic impact of the geometry.

Layer	Calorimeter Segmentation
Inner	(48,4), (48,12), (48,24), (48,48), (192,4), (192,12), (192,24), (192,48)
Middle	(12,12), (48, 24), (48,48), (36,48)*
Outer	(24,24)

Table 1: Segmentation of each layer of the simulated calorimeters used in training, indicated by (n_η, n_ϕ) , the number of cells in η and ϕ , respectively. Several configurations are used, including one (marked with a *) in which the cell division is not uniform. See text for details.

In addition, a second set of configurations is prepared and photon events simulated, but which is held out of the training set in order to test the generalization to new geometries. Configurations with a middle layer segmentation of (24, 24) or (24, 12) are generated to test the ability of the model to interpolate. Configurations with a middle layer segmentation of (96,24), (6,6) or (9, 12)* are generated to test the ability of the model to extrapolate beyond its training range (summarised in Table 3).

Interpolation	Extrapolation
(24,24), (24,12)	(96,24), (6,6), (9, 12)*

Table 2: Segmentation of interpolation and extrapolation middle layer geometries of the simulated calorimeters used in evaluation, indicated by (n_η, n_ϕ) , the number of cells in η and ϕ , respectively. Several configurations are used, including one (marked with a *) in which the cell division is not uniform. See text for details.

4 Generative Models

Generative models, both conditioned on the geometry and unconditioned, are developed. Both are described below.

¹All non-uniform geometries are marked with a *.

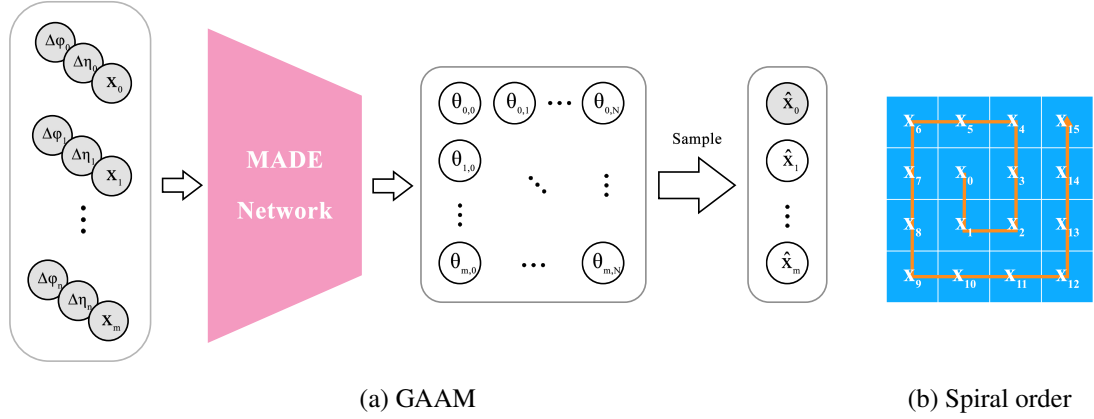


Figure 2: (a) The ARM architecture in GAAM takes a starting value x_0 and cell size $(\Delta\eta, \Delta\phi)$ as inputs, which are aggregated to the MADE network. The parameters θ_i represent the categorical distribution model (with $N + 1$ discrete energy categories) that has been learned, and the $m - 1$ output calorimeter cell energies are sampled from this distribution. (b) The two-dimensional calorimeter layer matrix is flattened along a spiral path.

4.1 Geometry-Aware Autoregressive Model (GAAM)

The GAAM is a framework based on autoregressive models (ARM) that can efficiently learn to generate 3D calorimeter images. It comprises three ARMs, each generating an individual 2D layer of the calorimeter (see Figure 2). Each ARM is constructed with Masked Autoencoder for Distribution Estimation (MADE) [25]. Unlike traditional autoregressive models that sequentially generate output cells depending on preceding cells, MADE generates all desired parameters in a single pass through the regular autoencoder. This enables parallel computation on GPUs for faster training.

To generate the inner calorimeter layer, an ARM is employed consisting of one masked fully-connected layer and one 1D convolutional layer. For the middle and outer calorimeter layers, ARMs are used comprising five masked fully-connected layers and one 1D convolutional layer. Each of them uses GELU [26] activation functions. The output of the model is discrete energy (with $N + 1$ possible values) of m cells, and each generated cell energy deposit is represented as a categorical distribution (see Sec. 4.3 for the details about the preprocessing of cell energies). A softmax layer is applied to generate these discrete outputs into $N + 1$ categories, where N is the closest integer greater than the maximum cell energy in the training data.

In addition to the previous energy deposits, the ARM takes into account the cell sizes $(\Delta\eta, \Delta\phi)$ in the η and ϕ dimensions, respectively, as conditioning features. This enables the model to dynamically generate the energy distribution to accommodate different geometries (Figure 2a). By learning to generate cell-by-cell, the model gains insight into how the spatial properties of each cell impact the energy deposition. For instance, the model should learn to deposit more energy per cell for a less granular geometry with bigger cells compared to a finer geometry with smaller cells, as well as take into account where the cell is in the image. Consequently, the model learns from multiple cells within each image, enhancing the learning efficiency. During the generation, each ARM receives a starting value, the energy of a central cell. The energy of the central cell is sampled from a known prior distribution. The generation process is initiated from this cell since the majority

of the energy tends to be deposited near the center, providing a reliable starting point for the ARMs. In addition to the central cell energy, the networks also receive two matrices describing the size of the cells to be generated. The trained GAAM then generates the energy deposits in the remaining cells by sampling from the learned categorical distribution.

4.2 Baseline: Unconditional ARM

As a comparison to the geometry-aware approach described above, a similar model is trained without conditioning it to the properties of the calorimeter cells. This will serve as a baseline to understand the improvement that comes purely from being able to condition the model on cell properties.

4.3 Preprocessing

Since an ARM requires an ordering of the calorimeter cells, the two-dimensional cell matrix is flattened using a spiral counter-clockwise pattern (Figure 2b) so that cells remain close after flattening. The ordering starts from a central cell, where most energy is deposited, and moves outwards. Given the range in the distribution of energy depositions across cells spans several orders of magnitude, applying a power transformation, $\hat{x} = x^{1/p}$, with $p = 2$ during preprocessing is found to improve the training. Furthermore, the value is then discretized by rounding it to the nearest integer.

The dataset is divided into training, validation, and test sets, with a ratio of 0.7:0.1:0.2, respectively. Training the models is carried out in PyTorch [27] for up to 200 epochs, using the Adam optimizer [28]. The negative log likelihood serves as the loss function. The training is performed on two NVIDIA RTX 3090 GPUs.

5 Results

Performance of the conditional GAAM as well as the baseline unconditional model are presented both qualitatively as well as quantitatively, for geometries with uniform or non-uniform cell segmentation, and for geometries within the training sample as well as unseen geometries.

Qualitative analysis is via visualizations and inspection of generated samples. Quantitative analysis is performed by measuring the distance between histograms of true and generated samples in quantities which are important in a typical data analysis: energy weighted means, shower widths, and distances between shower means across layers. The energy weighted means, $\bar{\eta}$ and $\bar{\phi}$ are computed over all cells in an image:

$$\bar{\eta} = \frac{\sum_i \eta_i E_i}{\sum_i E_i}, \bar{\phi} = \frac{\sum_i \phi_i E_i}{\sum_i E_i} \quad (5.1)$$

where E_i is the energy deposited in the i^{th} cell. The shower widths, σ_η and σ_ϕ are calculated as:

$$\sigma_\eta = \sqrt{\frac{\sum_i E_i (\eta_i - \bar{\eta})^2}{\frac{(M-1)}{M} \sum_i E_i}}, \sigma_\phi = \sqrt{\frac{\sum_i E_i (\phi_i - \bar{\phi})^2}{\frac{(M-1)}{M} \sum_i E_i}} \quad (5.2)$$

where M is the number of cells with non-zero energy. The inter-layer distance between mean energy-weighted shower locations in the inner and middle calorimeter layers are computed as:

$$C_\eta = \bar{\eta}_{\text{inner}} - \bar{\eta}_{\text{middle}}, \quad C_\phi = \bar{\phi}_{\text{inner}} - \bar{\phi}_{\text{middle}} \quad (5.3)$$

Distances between histograms are measured using the Wasserstein distance [29, 30]. The speed of the GAAM is a function of the number of cells. Generation takes 2040.7 ms/shower for inner layers,

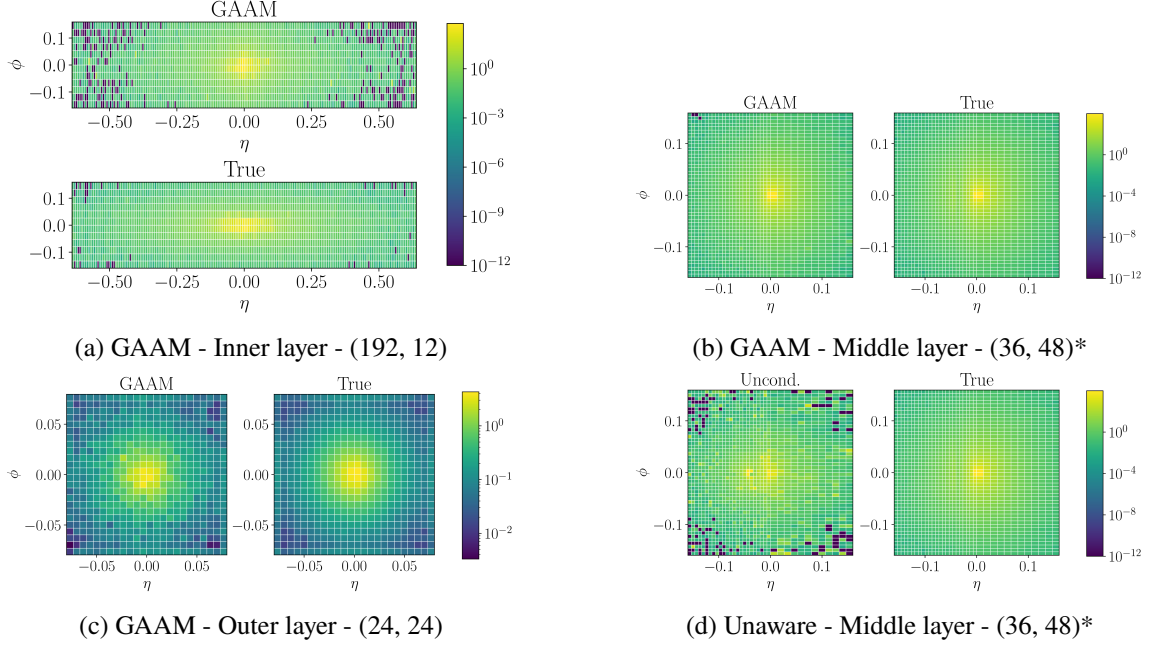


Figure 3: Average generated calorimeter images from GAAM in (a) inner (b) middle or (c) outer layers, or (d) from the geometry-unaware model middle layer, each compared the true images from Geant4.

140.5 ms/shower for the middle and 9.6 for outer layers, compared to 2058 ms/shower with Geant4 for 65 GeV photons. However, this model had not yet been optimized for speed.

5.1 Performance in uniform segmentation

Qualitative comparison of the generated images from the GAAM and the geometry-unaware model are shown in Fig. 3 in each layer, for several example segmentations, and compared to the truth generated by Geant4. The GAAM reproduces the features of the true images, while the geometry-unaware model fails to do so.

Quantitative comparison are made between histograms of energy-weighted means (Fig. 4), shower widths (Fig. 5), and inter-layer distances (Fig. 6). The Wasserstein distances between the distributions are given in Table 3. In each case, the GAAM outperforms the geometry-unaware model², and achieves a small inter-layer distance. The inner layer is particularly challenging for GAAM due to its high granularity.

5.2 Performance in non-uniform segmentation

A major challenge for implementing generative models is the non-uniform nature of the detector segmentation, which becomes more coarse or irregular in some regions. Performance under non-uniform cell divisions is studied in the configuration where the middle layer has segmentation (36, 48)*, but non-uniform cell sizes along η . Demonstration of the GAAM’s ability to learn to generate data with larger energy deposition in larger cells is shown in Fig 3b. The slight off-set of

²The outer layer has only a single geometry so no comparisons are made to a geometry-unaware model.

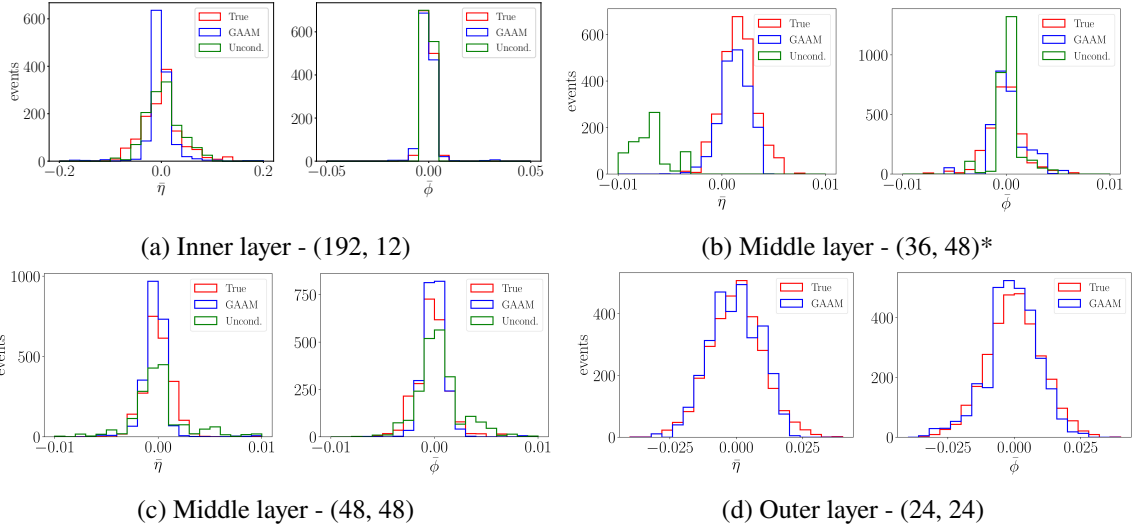


Figure 4: Distribution of the energy weighted means, $\bar{\eta}$ and $\bar{\phi}$, in samples generated by the GAAM, the unconditional model and Geant4 (True), for several calorimeter layers with varying segmentation in (η, ϕ) . The GAAM outperforms the unconditional baseline model consistently.

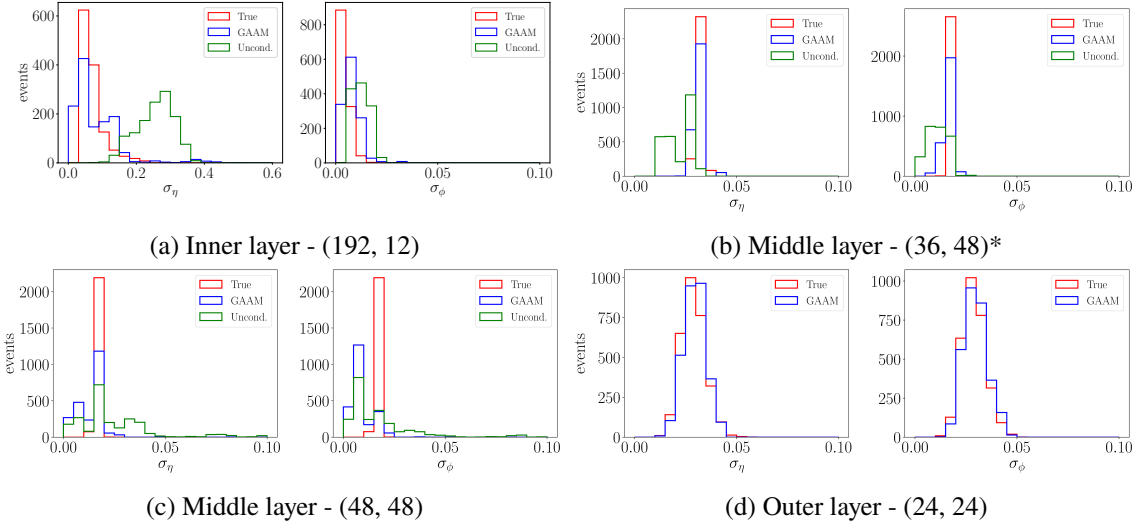


Figure 5: Distribution of the shower widths, σ_{η} and σ_{ϕ} , in samples generated by the GAAM, the unconditional model and Geant4 (True), for several calorimeter layers with varying segmentation in (η, ϕ) . The GAAM outperforms the unconditional baseline model consistently.

the peak to the right in the η distribution in Fig. 4b is a feature of boundary regions of calorimeters and therefore expected for the geometry (36, 48)* of the middle layer.

Such a task has never been attempted before with generative models, due to the challenges it poses, as illustrated by the failure of the geometry-unaware model to accurately simulate the calorimeter response in this scenario; see Figure 3d.

Layer; segmentation	Quantity	Wasserstein distance	
		GAAM (Conditional)	Baseline (Unconditional)
Inner; (192, 12)	$\bar{\eta}$	1.80×10^{-2}	1.30×10^{-2}
	$\bar{\phi}$	8.50×10^{-4}	1.36×10^{-3}
	σ_{η}	2.44×10^{-2}	1.63×10^{-1}
	σ_{ϕ}	2.71×10^{-3}	6.45×10^{-3}
Middle; (48,48)	$\bar{\eta}$	5.95×10^{-4}	6.98×10^{-3}
	$\bar{\phi}$	4.05×10^{-4}	2.61×10^{-3}
	σ_{η}	4.35×10^{-3}	1.25×10^{-2}
	σ_{ϕ}	7.07×10^{-3}	1.00×10^{-2}
Middle; (36,48)	$\bar{\eta}$	1.25×10^{-2}	2.45×10^{-2}
	$\bar{\phi}$	4.54×10^{-4}	5.58×10^{-4}
	σ_{η}	1.18×10^{-3}	1.00×10^{-2}
	σ_{ϕ}	1.24×10^{-3}	5.60×10^{-3}
Outer; (24,24)	$\bar{\eta}$	8.31×10^{-4}	N/A
	$\bar{\phi}$	1.12×10^{-3}	N/A
	σ_{η}	9.08×10^{-4}	N/A
	σ_{ϕ}	9.71×10^{-4}	N/A

Table 3: Quantitative evaluation of performance of the two generative models, the conditional GAAM and the unconditional baseline model in generating simulated calorimeter response to a photon. Shown is the Wasserstein distance between histograms of quantities important in data analysis, displayed in Figs. 4-6. The testing geometry appears in the training set for these models.

Layer; segmentation	Quantity	Wasserstein distance	
		GAAM (Conditional)	Baseline (Unconditional)
Middle; (24,24)	$\bar{\eta}$	2.90×10^{-4}	7.74×10^{-3}
	$\bar{\phi}$	6.21×10^{-4}	1.06×10^{-3}
	σ_{η}	8.98×10^{-3}	1.63×10^{-2}
	σ_{ϕ}	4.22×10^{-3}	7.39×10^{-3}
Middle; (24,12)	$\bar{\eta}$	4.77×10^{-4}	7.74×10^{-3}
	$\bar{\phi}$	1.02×10^{-3}	1.06×10^{-3}
	σ_{η}	8.98×10^{-3}	1.63×10^{-2}
	σ_{ϕ}	4.22×10^{-3}	7.39×10^{-3}

Table 4: Quantitative evaluation of performance of the two generative models, the conditional GAAM and the unconditional baseline model in generating simulated calorimeter response to a photon. Shown is the Wasserstein distance between histograms of quantities important in data analysis, displayed in Fig. 8. The testing geometry does not appear in the training set for these models, but requires interpolation.

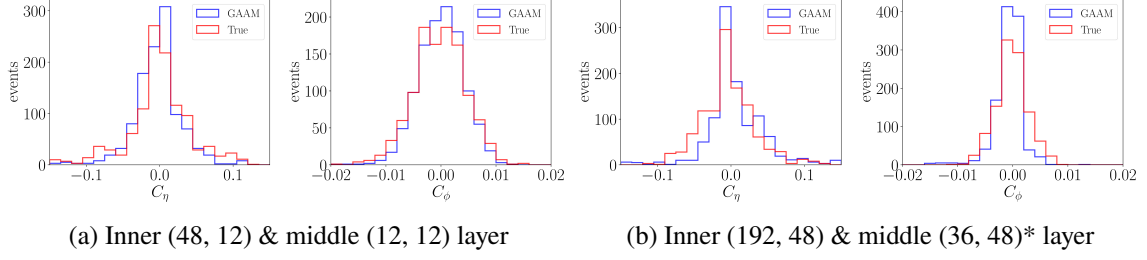


Figure 6: Distribution of the inter-layer distance of energy weighted means, C_η and C_ϕ , in samples generated by the GAAM and Geant4 (True), for several calorimeter layers with varying segmentation in (η, ϕ) . The GAAM reproduces the truth reasonably well.

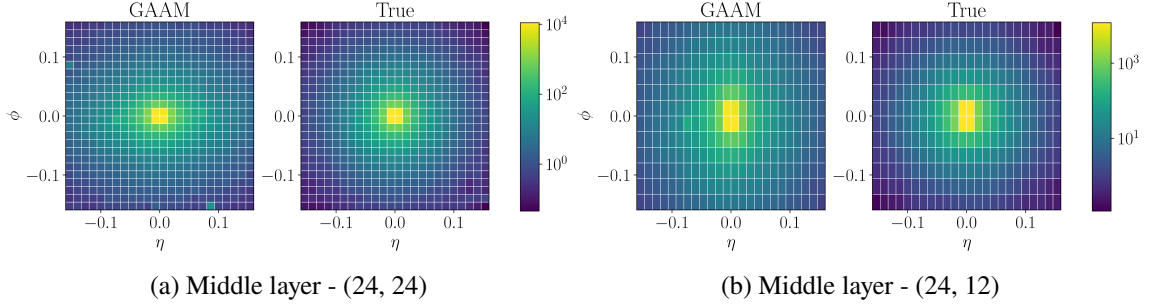


Figure 7: Average generated calorimeter images from GAAM for two geometries, (a) (24, 24) and (b) (24, 12), which do not appear in the training set and require interpolation. Each are compared to the Geant4 truth.

5.3 Performance in unseen geometries

A crucial target of geometry-aware models is the ability to learn the dependence of the energy deposition on the nature of the geometry and thus be able to generate simulated response for calorimeters whose geometry do not appear in its training set. To explore this ability of the GAAM, we evaluate its performance on unseen geometries which require interpolation within the training set, as well as those which require extrapolation out of the training set.

Figure 7 shows the average generated and true images for two unseen geometries which are bracketed by examples in the training set. Qualitatively, the GAAM has succeeded at interpolation. A quantitative comparison is provided by examining the important physical quantities defined above; see Figs. 8 and 9 for histograms and Table 4 for Wasserstein distances. The GAAM is able to reasonably interpolate to unseen geometries, although it struggles to produce very narrow shower widths.

Extrapolation beyond the training examples given is much more challenging. We present three extrapolation geometries of the middle layer, which have cell sizes outside the range of the training data: (6, 6), (96, 24), and (9, 12)*. The (6, 6) geometry is composed of larger cell sizes than the largest cells in training data in both η -dimension and ϕ -dimension. The (96, 24) geometry has a similar cell segmentation, but with cells half the size of the smallest cells in training data in the η -dimension and twice the size of the smallest cells in training data in the ϕ -dimension. Details about the spiral ordering in such cases are provided in Appendix A. The (9, 12)* geometry is composed of

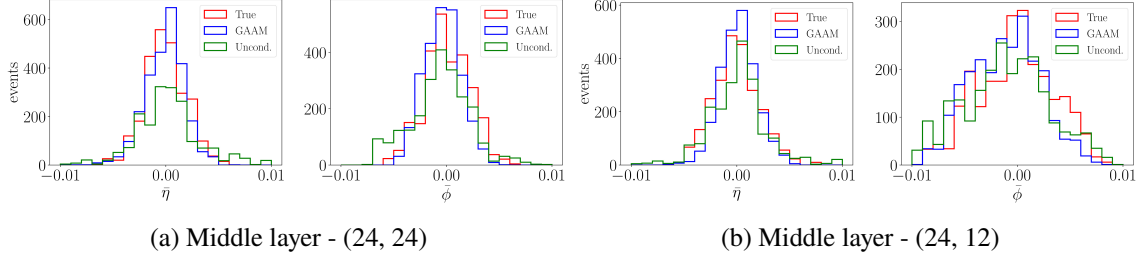


Figure 8: Distribution of the energy weighted means, $\bar{\eta}$ and $\bar{\phi}$, in samples generated by the GAAM, the unconditional model and Geant4 (True), for several calorimeter layers with unseen varying segmentation in (η, ϕ) , geometries which do not appear in the training set and require interpolation. The GAAM outperforms the unconditional baseline model consistently.

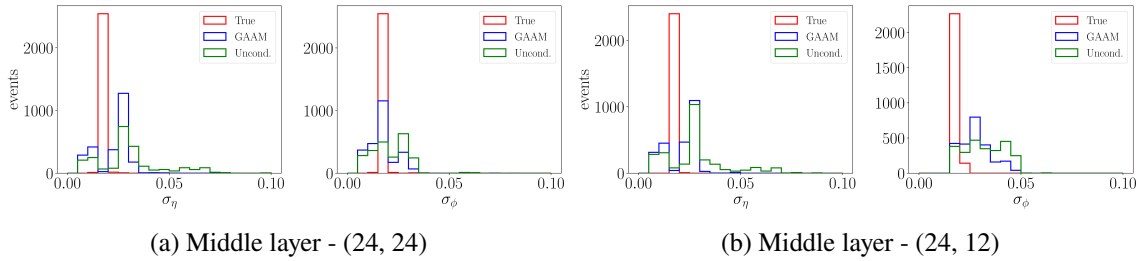


Figure 9: Distribution of the shower widths, σ_η and σ_ϕ , in samples generated by the GAAM, the unconditional model and Geant4 (True), for several calorimeter layers with unseen varying segmentation in (η, ϕ) , geometries which do not appear in the training set and require interpolation. The GAAM outperforms the unconditional baseline model but it also struggles to produce narrow distributions.

two different cell sizes in the η -dimension, similar to the feature of boundary regions. The first six cells from the left are of the same size of the largest cells seen in training, and the other three cells are twice as large as the cells on the left. In the ϕ -dimension, the cells are uniformly segmented. In contrast to interpolation, samples for unseen geometries which require extrapolation show significant differences between the generated and true images (Figure 10), with the GAAM failing to place the maximum energy at the centre of the image in the third case. For the (6, 6) geometry, the GAAM deposits insufficient energy in the central cells and too much in the exterior cells, and for (96, 24) too there are visible artifacts. Quantitative comparisons, not shown, confirm what is clear in the visualizations.

6 Conclusion and Outlook

The Geometry-Aware Autoregressive Model is capable of learning how calorimeter cell response varies with the detector geometry, allowing it to rapidly generate simulated responses to a wide array of detector geometries, including those which do not appear in the training set. This capacity for generalization eliminates the need to train and validate hundreds of models to describe the response of a typical calorimeter. By conditioning the model on the properties of individual cells rather than

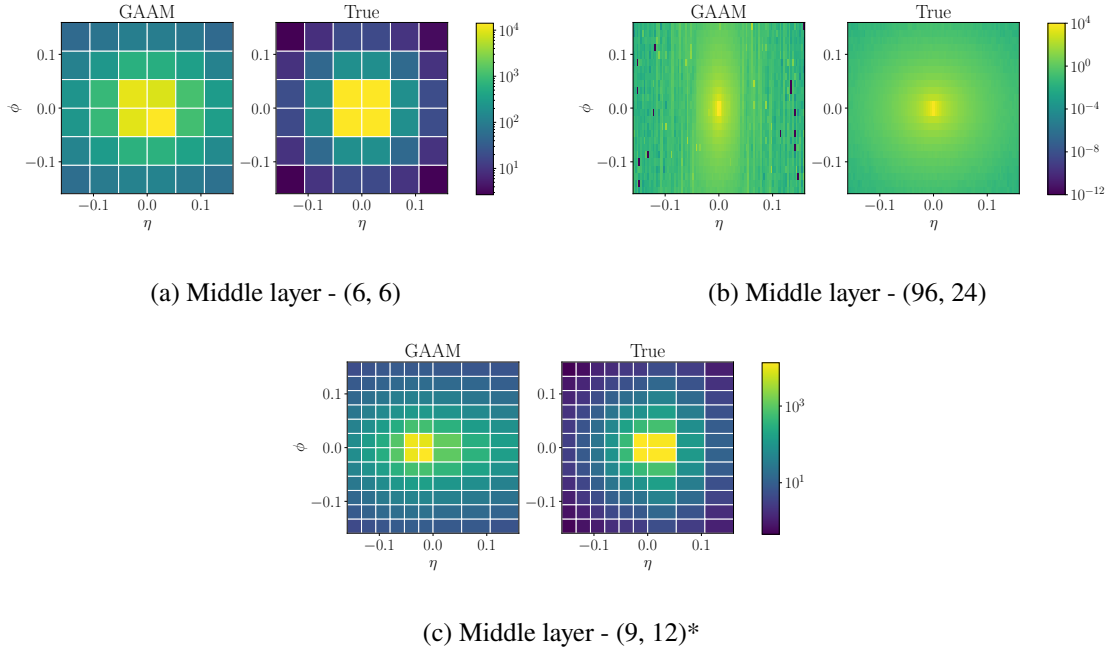


Figure 10: Average generated calorimeter images from GAAM for two (η, ϕ) segmentations, (a) (6, 6), (b) (96, 24), and (c) (9, 12)*, which do not appear in the training set and require extrapolation. Each is compared to the Geant4 truth. GAAM does not extrapolate beyond the range of cell sizes seen in training.

the complete geometry [9], the GAAM has the capacity to scale to many geometry configurations.

Further work is needed to improve the overall fidelity of the simulation, particularly for the high dimensional inner calorimeter layer. It may need to also be trained on data with a variable position of the incident particle and boundary regions that are not always at the centre of the image for real world applications. Additionally, one may compare such a geometry-aware approach with a *geometry-agnostic* approach to shower simulation with point-cloud models [31–34] to determine the suitability of each approach on a range of detector geometries. While the geometries investigated were motivated by the ATLAS calorimeter, further work would be required to generalize to even more diverse geometries such as hexagonal cells or varying number of calorimeter layers relevant to other calorimeters such as in CMS. While not studied in this work, GAAM could be trained on non-uniform segmentations in the ϕ direction if needed. As the first attempt at designing generative models that can handle variable calorimeter geometries, this work did not focus on optimization of the speed of the GAAM, which is left for future work. As with previous work such as CaloFlow [16] and the faster CaloFlow II [17] or caloScore [18] and the faster caloScore v2[35], future work can be expected to speed up geometry-aware models, for example with the use of ML-based dimensionality reduction methods.

This study is also the first step towards building ‘foundational models’ in HEP, general purpose machine learning models trained on large and diverse data, intended to be fine-tuned to individual tasks. The GAAM demonstrates the feasibility of general purpose generative models for calorimeter

simulation. These would act as foundational models, trained on various calorimeters and geometries, and then shared between experiments. The experiments would only need to fine-tune the model to their specific calorimeter, requiring far less training data, computing resources and human resources compared to developing a generative model from scratch for their specific calorimeter.

Acknowledgments

DW, AG and DS are supported by The Department of Energy Office of Science grant DE-SC0009920. AG is also supported under contract DE-AC02-05CH11231, and DS is also supported by the HEP-CAT Fellowship under contract DE-SC0022313. We are grateful to Benjamin Nachman for help with the Geant4 setup. We also thank Tobias Golling and Johnny Raine for fruitful discussions.

Note added: As this manuscript was being finalized, we became aware of the recent work in Ref. [36], which uses point-cloud generative model to simulate the future International Large Detector (ILD) calorimeter. It does not make comparisons on multiple challenging geometries as we do.

References

- [1] GEANT4 collaboration, *GEANT4—a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.
- [2] J. Allison et al., *Recent developments in Geant4*, *Nucl. Instrum. Meth. A* **835** (2016) 186.
- [3] J. Allison et al., *Geant4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.
- [4] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, *Nature communications* **5** (2014) 4308.
- [5] P. Baldi, *Deep learning in science*, Cambridge University Press (2021).
- [6] P. Calafiura, D. Rousseau and K. Terao, *Artificial Intelligence for High Energy Physics*, World Scientific (2022).
- [7] Y. Lu, A. Romero, M.J. Fenton, D. Whiteson and P. Baldi, *Resolving extreme jet substructure*, *Journal of High Energy Physics* **2022** (2022) 1.
- [8] A. Shmakov, K. Greif, M. Fenton, A. Ghosh, P. Baldi and D. Whiteson, *End-to-end latent variational diffusion models for inverse problems in high energy physics*, *arXiv preprint arXiv:2305.10399* (2023) .
- [9] ATLAS collaboration, *Deep generative models for fast photon shower simulation in ATLAS*, [2210.06204](#).
- [10] ATLAS collaboration, *AtlFast3: the next generation of fast simulation in ATLAS*, *Comput. Softw. Big Sci.* **6** (2022) 7 [[2109.02551](#)].
- [11] 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](#)].
- [12] F. Ratnikov, *Generative Adversarial Networks for LHCb Fast Simulation*, *EPJ Web Conf.* **245** (2020) 02026 [[2003.09762](#)].
- [13] 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.* **D97** (2018) 014021 [[1712.10321](#)].

- [14] Vallecorsa, Sofia, Carminati, Federico and Khattak, Gulrukh, *3d convolutional gan for fast simulation*, *EPJ Web Conf.* **214** (2019) 02010.
- [15] 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].
- [16] C. Krause and D. Shih, *CaloFlow: Fast and Accurate Generation of Calorimeter Showers with Normalizing Flows*, **2106.05285**.
- [17] C. Krause and D. Shih, *CaloFlow II: Even Faster and Still Accurate Generation of Calorimeter Showers with Normalizing Flows*, **2110.11377**.
- [18] V. Mikuni and B. Nachman, *Score-based Generative Models for Calorimeter Shower Simulation*, **2206.11898**.
- [19] ATLAS collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, .
- [20] H. Hashemi, N. Hartmann, S. Sharifzadeh, J. Kahn and T. Kuhr, *Ultra-High-Resolution Detector Simulation with Intra-Event Aware GAN and Self-Supervised Relational Reasoning*, **2303.08046**.
- [21] Y. Lu, J. Collado, D. Whiteson and P. Baldi, *Sparse autoregressive models for scalable generation of sparse images in particle physics*, *Physical Review D* **103** (2021) 036012.
- [22] M. Paganini, L. de Oliveira and B. Nachman, *CaloGAN: Simulating 3d high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks*, *Physical Review D* **97** (2018) .
- [23] ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [24] M. Aleksa and M. Diemoz, *Discussion on the electromagnetic calorimeters of atlas and cms*, *CERN Document Server* (2013) .
- [25] M. Germain, K. Gregor, I. Murray and H. Larochelle, *Made: Masked autoencoder for distribution estimation*, in *International conference on machine learning*, pp. 881–889, PMLR, 2015.
- [26] D. Hendrycks and K. Gimpel, *Gaussian error linear units (gelus)*, *arXiv preprint arXiv:1606.08415* (2016) .
- [27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds., pp. 8024–8035, Curran Associates, Inc. (2019), <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [28] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *arXiv preprint arXiv:1412.6980* (2014) .
- [29] L.V. Kantorovich, *Mathematical methods of organizing and planning production*, *Manage. Sci.* **6** (1960) 366.
- [30] C. Villani, *Optimal Transport: Old and New*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*, Springer Berlin (2008), [10.1007/978-3-540-71050-9](https://doi.org/10.1007/978-3-540-71050-9).
- [31] R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini et al., *Particle Cloud Generation with Message Passing Generative Adversarial Networks*, in *35th Conference on Neural Information Processing Systems*, 6, 2021 [2106.11535].

- [32] R. Kansal, A. Li, J. Duarte, N. Chernyavskaya, M. Pierini, B. Orzari et al., *Evaluating generative models in high energy physics*, *Phys. Rev. D* **107** (2023) 076017 [2211.10295].
- [33] M. Leigh, D. Sengupta, G. Quétant, J.A. Raine, K. Zoch and T. Golling, *PC-JeDi: Diffusion for Particle Cloud Generation in High Energy Physics*, [2303.05376](#).
- [34] V. Mikuni, B. Nachman and M. Pettee, *Fast Point Cloud Generation with Diffusion Models in High Energy Physics*, [2304.01266](#).
- [35] V. Mikuni and B. Nachman, *CaloScore v2: Single-shot Calorimeter Shower Simulation with Diffusion Models*, [2308.03847](#).
- [36] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol et al., *CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation*, [2305.04847](#).

Appendix

A Spiral Ordering

There are special cases where the geometry does not fit perfectly into our data preprocessing approach. For example, when the events have non-square geometries, such as in the inner layer, we have a modified version of the spiral path shown in Figure 11. The spiral path will look the same for the center square until it reaches the boundary in one dimension. Then, the spiral path will go left and right in turn to cover the peripheral cells.

In another case, with an even number of cells in η and ϕ direction, there are four cells closest to the position of the shower center. We choose the top-left cell in the central four cells. The known prior distribution is obtained by sampling from the training data.

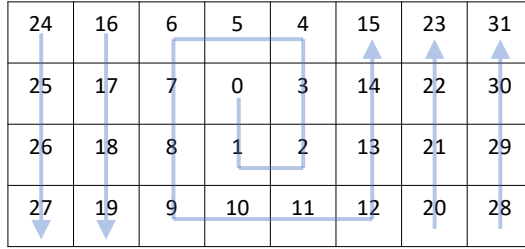


Figure 11: Illustration of how the 2D image is flattened in a spiral order when the image is non-square.

B Individual Shower Images

This section shows individual shower images in different geometries and for multiple calorimeter layers. GAAM and Geant4 generated images are shown for the inner layer in Figure 12 and Figure 13, middle in Figure 14 and Figure 15 and the outer layer in Figure 16 and Figure 17 respectively.

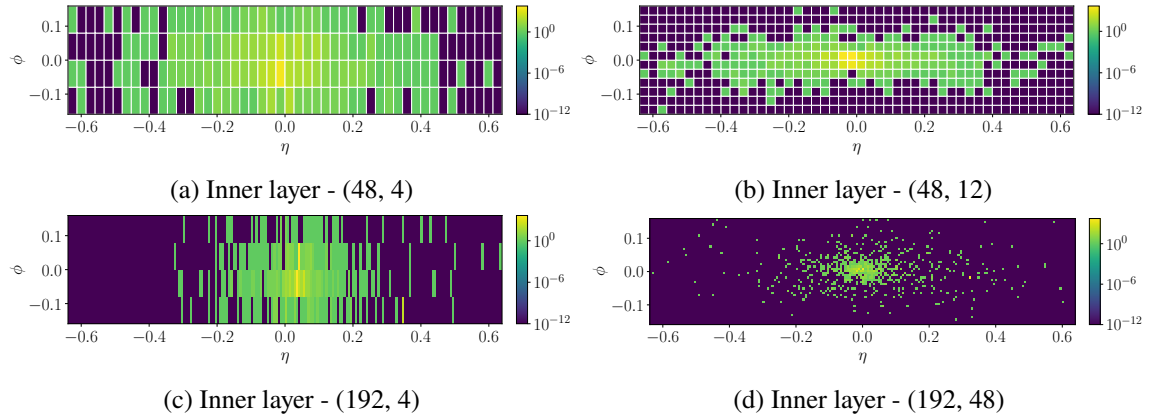


Figure 12: Examples of individual events of the inner layer generated by GAAM

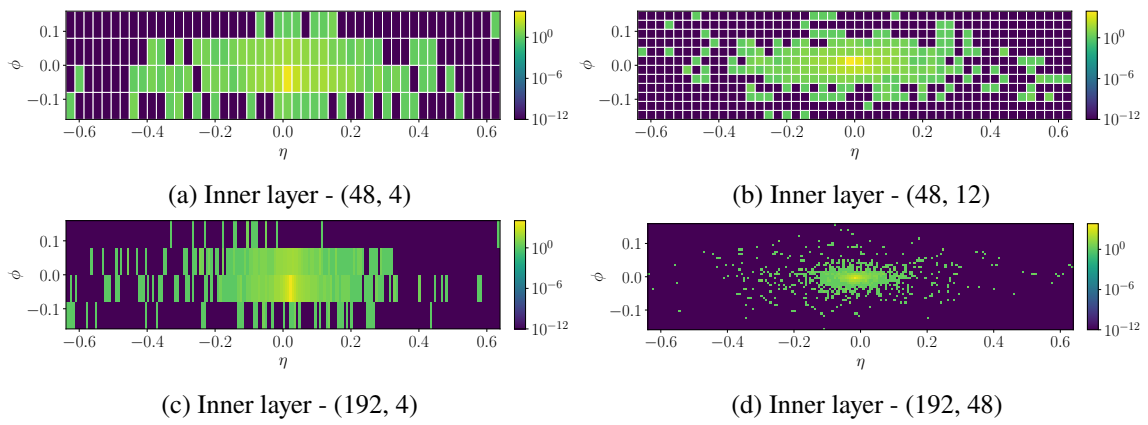
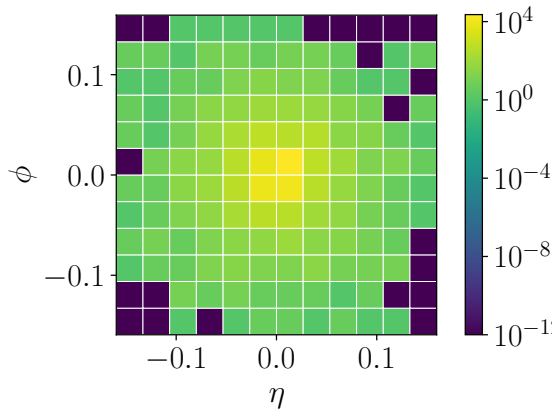
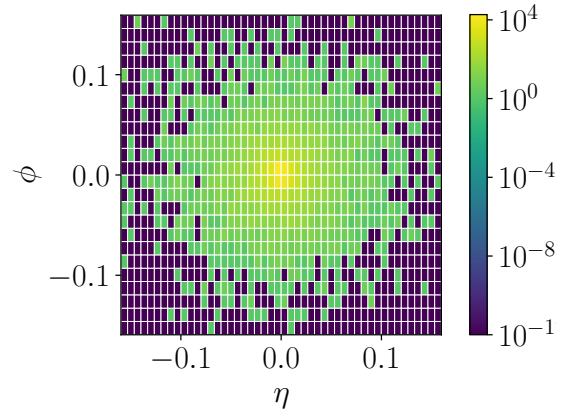


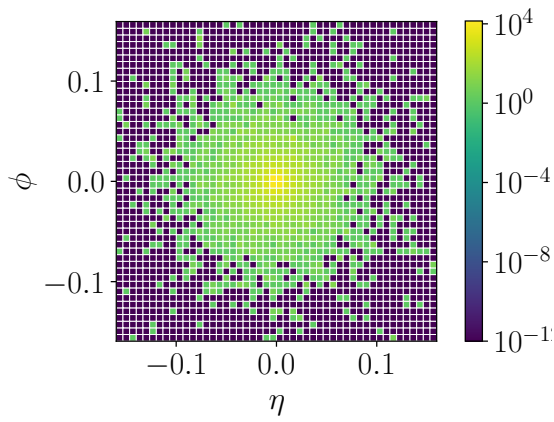
Figure 13: Examples of individual events of the inner layer generated by Geant4



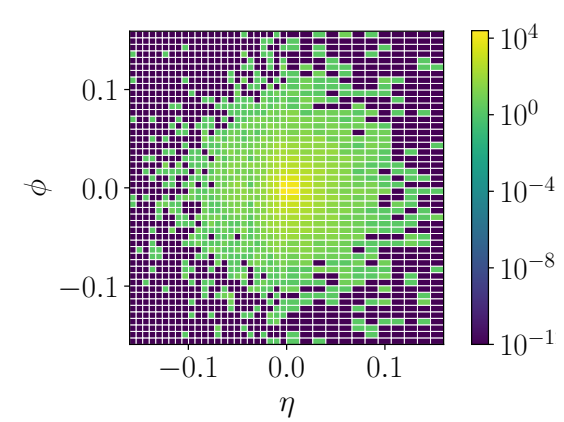
(a) Middle layer - (12, 12)



(b) Middle layer - (48, 24)

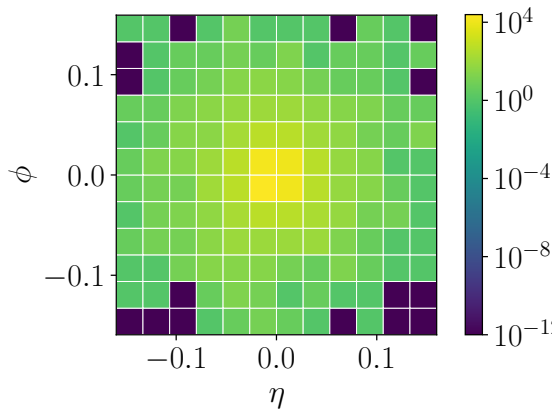


(c) Middle layer - (48, 48)

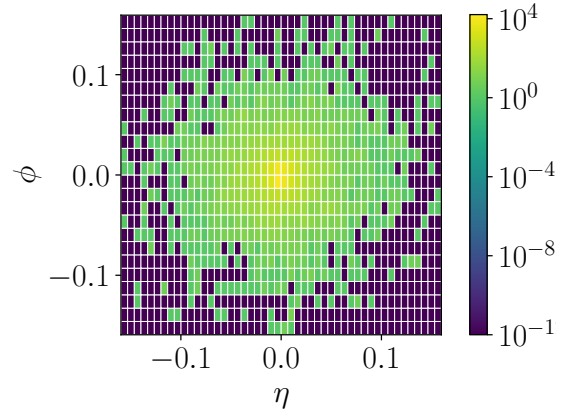


(d) Middle layer - (36, 48)*

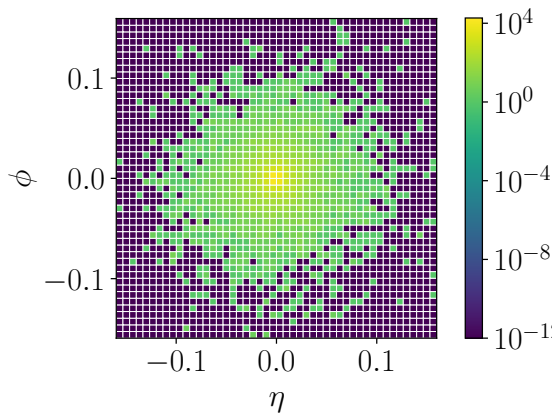
Figure 14: Examples of individual events of the middle layer generated by GAAM



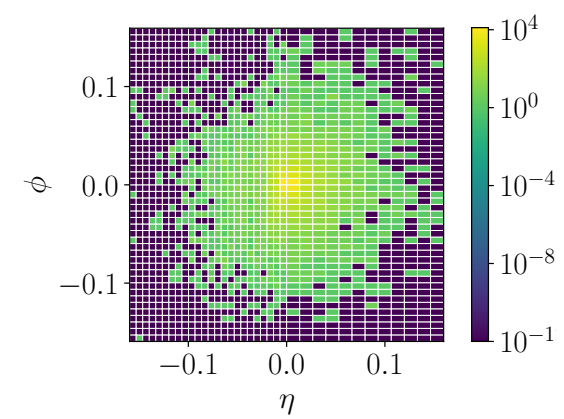
(a) Middle layer - (12, 12)



(b) Middle layer - (48, 24)

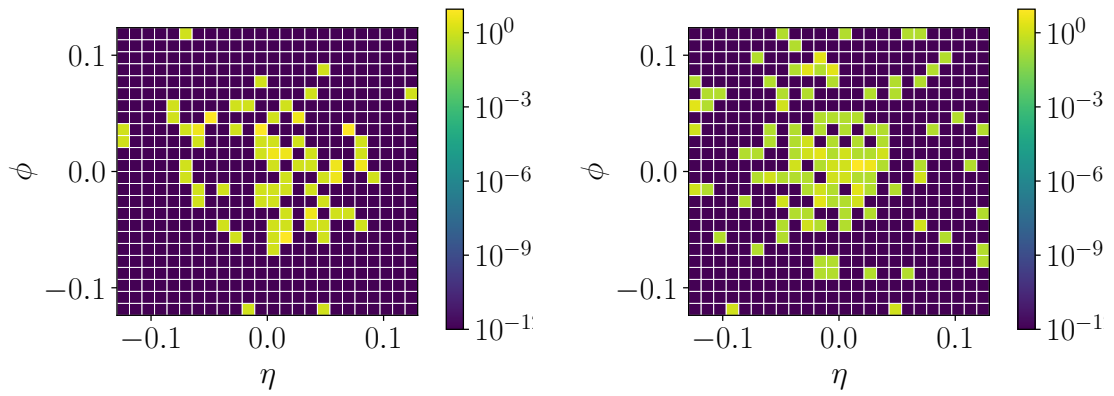


(c) Middle layer - (48, 48)



(d) Middle layer - (36, 48)*

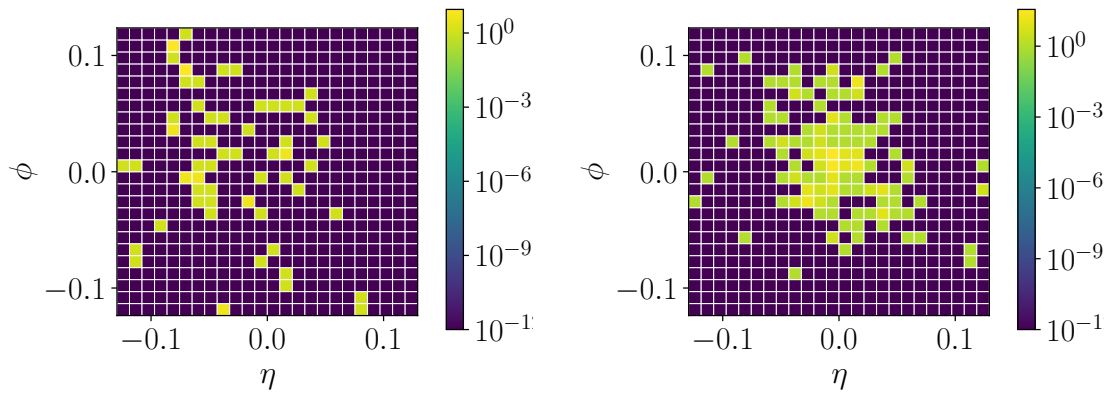
Figure 15: Examples of individual events of the middle layer generated by Geant4



(a) Outer layer - (24, 24)

(b) Outer layer - (24, 24)

Figure 16: Examples of individual events of the outer layer generated by GAAM



(a) Outer layer - (24, 24)

(b) Outer layer - (24, 24)

Figure 17: Examples of individual events of the outer layer generated by Geant4