CALoGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks

Michela Paganini,[1,2,*] Luke de Oliveira,[2,†] and Benjamin Nachman[2,‡]

[1]*Yale University, New Haven, Connecticut 06520, USA*
[2]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

The precise modeling of subatomic particle interactions and propagation through matter is paramount for the advancement of nuclear and particle physics searches and precision measurements. The most computationally expensive step in the simulation pipeline of a typical experiment at the Large Hadron Collider (LHC) is the detailed modeling of the full complexity of physics processes that govern the motion and evolution of particle showers inside calorimeters. We introduce CALoGAN, a new fast simulation technique based on generative adversarial networks (GANs). We apply these neural networks to the modeling of electromagnetic showers in a longitudinally segmented calorimeter and achieve speedup factors comparable to or better than existing full simulation techniques on CPU ($100\times$ –$1000\times$) and even faster on GPU (up to $\sim 10^5\times$). There are still challenges for achieving precision across the entire phase space, but our solution can reproduce a variety of geometric shower shape properties of photons, positrons, and charged pions. This represents a significant stepping stone toward a full neural network-based detector simulation that could save significant computing time and enable many analyses now and in the future.

I. INTRODUCTION

The physics programs of all experiments based at the LHC rely heavily on detailed simulation for all aspects of event reconstruction and data analysis. Simulated particle collisions, decays, and material interactions are used to interpret the results of ongoing experiments and estimate the performance of new ones, including detector upgrades.

State-of-the-art simulations are able to precisely model detector geometries and physical processes spanning distance scales as small as 10^{-20} m for the initial parton-parton scattering, all the way to the material interactions at meter length scales. These processes, which include nuclear and atomic interactions, such as ionization, as well as strong, weak, and electromagnetic processes, will alter the state of incoming particles as they propagate through and interact with layers of material in the various detector components. Detection techniques such as calorimetry exploit these physical interactions to detect the presence and measure the energy of particles such as photons, electrons and hadrons via their interactions with hundreds of thousands of detector components. Upon interaction with a calorimeter, a cascade *shower* of secondary particles is produced and their energy is collected and transformed into electric signals.

Physics-based *full simulation* modeling of particle showers in calorimeters (with GEANT4 [1] as the state of the art) is the most computationally demanding part of the whole simulation process and can take minutes per event on modern, distributed high performance platforms [2,3]. The production of physics results is often limited by the absence of adequate Monte Carlo (MC) simulation, and the increase in luminosity at the LHC will only exacerbate the problem. For example, the ATLAS and CMS experiments at the high-luminosity phase of the LHC (HL-LHC) will each see about three billion top quark pair events [4–10]; for a MC statistical uncertainty that is significantly below the data uncertainty, hundreds of billion simulated events would be required. This is not possible using full detector simulation techniques with existing computing resources. Currently, full MC simulation occupies 50%–70% of the experiments' worldwide computing resources, equivalent to billions of CPU hours per year [11–13].

The relevance of the calorimeter simulation step has sparked the development of approximate, fast simulation solutions to mitigate its computational complexity. Fast simulation techniques rely on parametrized showers [14–16] for fluctuations, and look-up tables for low energy interactions [17]. For many applications, these techniques

*michela.paganini@yale.edu
†lukedeoliveira@lbl.gov
‡bnachman@cern.ch

are sufficient. However, analyses that utilize the detailed structure of showers for particle identification as well as energy and direction calibration may not be able to rely on these simplified approaches [18].

We introduce a deep learning model to enable high-fidelity fast simulation of particle showers in electromagnetic calorimeters. Previous work [19] assessed the viability of GAN-based simulation of *jet images* [20]—sparse, structured, two-dimensional representations of jet fragmentation analogous to a single-layer, idealized calorimeter—and focused on providing architectural guidelines for this regime. Neural network-based generation, including GANs, variational auto-encoders [21], and adversarial auto-encoders [22], have also been tested in other areas of science, such as cosmology [23,24], condensed matter physics [25], and oncology [26]. The longitudinally segmented calorimeter simulation addressed in this work offers unique challenges due to the sparsity of hit cells, the nonuniform granularity among the detector layers, and their sequential structure. In addition to enabling physics analyses at the LHC, the CALOGAN may form a base for solving similar computationally intensive modeling problems in other domains of science, medicine, and technology.

The paper is organized as follows. Section II introduces the data set of calorimeter showers, and Sec. III briefly reviews the generic GAN setup. The CALOGAN is described in Sec. IV and first results of its performance are documented in Sec. V. The paper ends with conclusions and future outlook in Sec. VI.

II. DATA SET

A detector simulation begins with a list of particles with lifetimes greater than $\mathcal{O}(\text{mm}/c)$. For each particle, we are given its type (e.g. electron, pion, etc.), its energy, and its direction. The particle type determines when and how the particle interacts with the material along its trajectory. Material interactions with the detector factorize [27]: the energy deposited in a calorimeter by various particles is the sum of the energy from each shower treated independently.

There are two flavors of calorimeters: electromagnetic and hadronic. Electromagnetic calorimeters are designed to stop electrons and photons, which have shallower and narrower showers compared with protons, neutrons, and charged pions. Hadronic calorimeters are thicker and deeper in order to capture penetrating radiation that forms irregular showers from nuclear interactions. In this first application of GANs to a longitudinally segmented calorimeter, we choose to focus only on electromagnetic showers. In addition to already providing the capability to simulate electrons and photons, the electromagnetic shower contains all of the new challenges described in Sec. I.

Transverse segmentation is critical for particle identification and energy calibration in an electromagnetic calorimeter. For example, the radiation pattern can be used to distinguish prompt photons from $\pi^0 \rightarrow \gamma\gamma$, where the

distance between the two photons is $\mathcal{O}(\text{cm})$ for a 10 GeV π^0 at one meter from the interaction point. Pion rejection and an excellent resolution for photons in the Higgs boson $H \rightarrow \gamma\gamma$ discovery channel were driving factors for the design of the ATLAS Liquid Argon (LAr) electromagnetic calorimeter [28], which will serve as an inspiration for the calorimeter used in this study. In particular, the calorimeter used in this study is a cube with size 480 mm³ with no material in front of it. There are three instrumented layers in the radial (z) direction [29] with thicknesses 90 mm, 347 mm, and 43 mm. The active material is LAr and the absorber material is lead. Only the total energy per layer, that includes both the active and inactive contributions is used in what follows.

In contrast to the complex accordion geometry in the actual ATLAS calorimeter, our simplified setup (built on the GEANT4 B4 example) uses flat alternating layers of lead and LAr that are 2 mm and 4 mm thick, respectively. Each of the three layers has a different segmentation, which is also not square in the first and third layers. In particular, the cells in the first layer are 160 mm × 5 mm, the cells in the second layer are 40 mm², and the cells in the third layer are 40 × 80 mm². The short direction in the first layer (η) corresponds to what would be the pp beam direction in a full experiment. In contrast, the short direction in the third layer (ϕ) is perpendicular to η. Table I summarizes the calorimeter geometry.

The training data set [30] is prepared as follows. GEANT4 10.2.0 [1] is used to generate particles and simulate their interaction with our calorimeter using the FTFP_BERT physics list based on the Fritiof [31–34] and Bertini intranuclear cascade [35–37] models with the standard electromagnetic physics package [38]. Positrons, photons, and charged pions with various energies are incident perpendicular on the center of the calorimeter front. Energies in the training are uniform in the range between 1 and 100 GeV. Figure 1 shows an example 10 GeV electron event with the exact energy deposits from GEANT4 [Fig. 1(a)] and after descretizing them according to our calorimeter geometry [Fig. 1(b)]. For visualization purposes, a three-dimensional particle energy signature (Fig. 2) will be displayed in the rest of this paper as a series of three two-dimensional images in η–ϕ space (Fig. 3), where the pixel intensity represents the sum of the energies of all particles incident to that cell [39]. The first

TABLE I. Dimension of a calorimeter cell. The z direction is the direction of particle propagation (radial direction in a full experiment), the η direction would be along the pp beam axis in a full experiment, and ϕ is perpendicular to z and η.

Layer	z length [mm]	η length [mm]	ϕ length [mm]
0	90	5	160
1	347	40	40
2	43	80	40
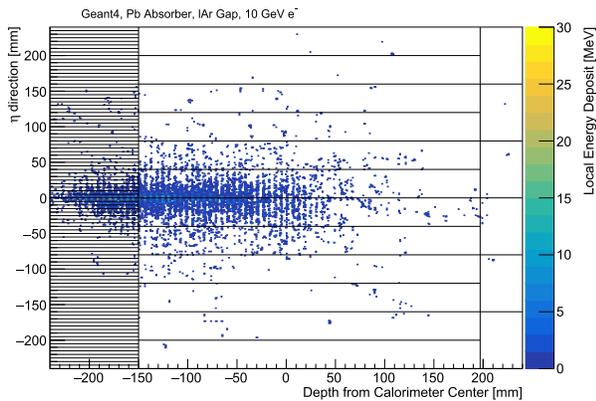
(a) Each dot represents one energy deposit from GEANT4 and the color of the dot encodes the energy. The absorber-gap structure is clearly visible, where most of the energy is lost in the absorber.
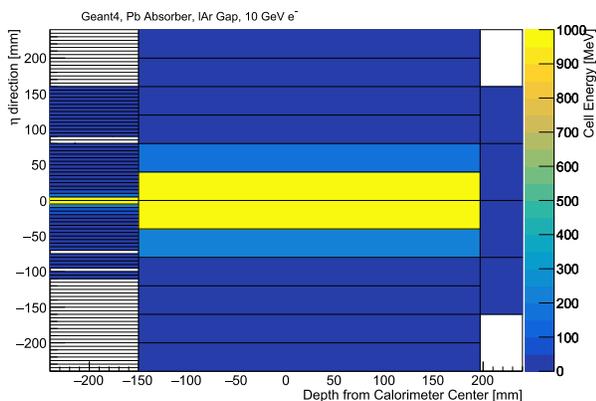
(b) Discretized version of (a), in which energy depositions are assigned to individual, discrete detector cells.

FIG. 1.   The electromagnetic shower from one 10 GeV electron event. The boundaries of the cells are shown, projecting out the ϕ segmentation.
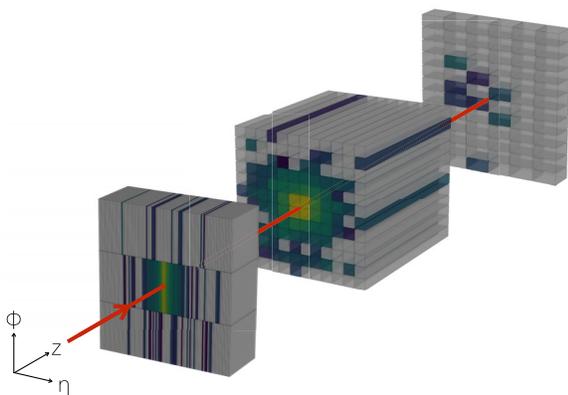

FIG. 2.   Three-dimensional representation of a 10 GeV e^+ incident perpendicular to the center of the detector. Not-to-scale separation among the longitudinal layers is added for visualization purposes.
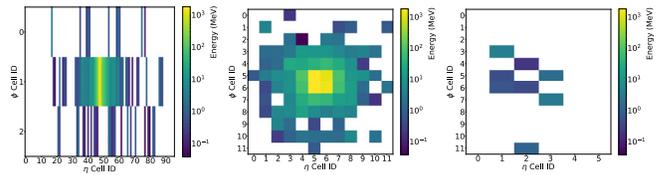

FIG. 3.   Two-dimensional, per-layer representation of the same shower as in Fig. 2.

layer can be represented as a 3×96 image, the middle layer as a 12×12 image, and the last layer as a 12×6 image.

III. GENERATIVE ADVERSARIAL NETWORKS

Since their first formulation [40], Generative adversarial networks have become a rapidly increasing area of attention in the machine learning literature with many applications in natural image processing. However, there are far fewer applications in basic science and prior to this work, no applications in high energy physics and nuclear physics.

Generative adversarial networks (GANs) cast the task of training a deep generative model as a two-player noncooperative minimax game, in which a generator network G is trained concomitantly with an Adversary, the discriminator Network D, in order to learn a target distribution f. The generator G learns a map from a latent space $z \sim p_z(z)$ (usually chosen to be $N(0, 1)$) to the space of generated samples, while D learns a map from the sample space to [0, 1], the probability that a shown sample is real. Note that the map that the generator learns implicitly defines a density g. The game-theoretical basis for this framework [40,41] ensures that if we extend the space of allowed functions that G and D can draw from to be the space of all continuous functions, then there exists some G (and, by construction, an implicit g) that exactly recovers the target distribution f, i.e., $g \rightarrow f$, while for every sample produced by the generator, the discriminator is maximally confused and admits a posterior of being real of ½. In order to train both G and D, the traditional formulation of GANs [40] utilizes the loss function $\mathcal{L}_{\text{adv}}$ shown in Eq. (1).

$$\mathcal{L}_{\text{adv}} = \underbrace{\mathbb{E}_{z \sim p_z(z)}[\log(\mathbb{P}(D(G(z)) = 0))]}_{\substack{\text{term associated with the discriminator} \\ \text{perceiving a generated sample as fake}}}$$
$$+ \underbrace{\mathbb{E}_{I \sim f}[\log(\mathbb{P}(D(I) = 1))]}_{\substack{\text{term associated with the discriminator} \\ \text{perceiving a real sample as real}}} \quad (1)$$

Though the GAN framework has shown promise, stability is still a major roadblock, and various ad-hoc and theoretical improvements have been suggested, from architectural guidelines [42–46] to reformulations of the loss specified in Eq. (1) to move away from the Jensen-Shannon divergence [47–52]. As suggested in [53], we are able to impose task-specific metrics which allow us to move away from loss level notions of quality and focus on

task-level fidelity measures. We make the conscious decision to utilize the vanilla loss formulation as we find adequate performance with this version.

IV. THE CALOGAN

Generative adversarial networks are explored as a tool to speed up full simulation of particle showers in an EM calorimeter. We identify this solution with the name CALOGAN.

For it to be useful in realistic physics applications, such a system needs to be able to accept requests for the generation of showers originating from an incoming particle of type P at energy E [54]. We introduce an auxiliary task of energy reconstruction to condition on E, a real valued variable. The auxiliary classifier GAN [43] formalism is tested to also condition on class P, but ultimately abandoned in favor of training a specific generative model for each particle type, as the authors expect that versioning and particle-specific improvements will be prioritized in any practical implementation.

In practice, energy is scaled by a factor of 10^2 and multiplied to the 1024-dimensional [55] latent space vector $z \in \mathbb{R}^{1024}$. The generator G then maps this input to three gray-scale image outputs with different numbers of pixels, which represent the energy patterns collected by the three calorimeter layers as the requested particle propagates through them. The discriminator D accepts the three images as inputs, along with E, the chosen value for the particle energy. The inputs are mapped to a binary output that classifies showers into real and fake, and a continuous output which calculates the total energy deposited in the three layers, then compares it with the requested energy E.

A. Model architecture

Given the sparsity levels and high dynamic range in the data described in Sec. II, we follow the LAGAN guidelines [19] to modify the DCGAN [42] architecture for this specific regime.

In the generator (shown in Fig. 4), our design combines parallel LAGAN-like processing streams with a trainable attention mechanism that encodes the sequential connection among calorimeter layers. The LAGAN submodules are composed of a 2D convolutional unit followed by two locally connected units with batch-normalization [56] layers in between. The dimensionality and granularity mismatch among the three longitudinal segmentations of the detector demand separate streams of operations with suitably sized kernels. Towards providing a readily adaptable tool, we provide an architecture construction that is simply a function of the desired output image size, as we seek a common denominator that can be readily applied to a variety of particles in order to obtain reasonable baselines in a quick R&D cycle.

Modelling the sequential nature of the relationship among the energy patterns collected by the three layer requires extra care. Drawing inspiration from Ref. [46], we choose an attention mechanism to allow dependence among layers, in which we define trainable transfer functions to optimally resize and apply knowledge of the energy pattern in previous layers to the generation of the subsequent layer readout. More specifically, in-painting takes as input a resized image from a previous layer, $\mathcal{I}'$, and the hypothesized image from the current layer, $\mathcal{I}$, and learns a per-pixel attention weight W via a weighting function $\omega(\mathcal{I}, \mathcal{I}')$ such that the pre-ReLU version of the current layer is $W \odot \mathcal{I} + (1 - W) \odot \mathcal{I}'$, where $\odot$ is the Hadamard product. This end-to-end trainable unit can utilize information about the two layers to decide what information to propagate through from the previous particle deposition. An alternative architectural choice that includes a recurrent connection will be subject of future studies.

Leaky Rectified Linear Units [57] are chosen as activation functions throughout the system, with the exception of the output layers of G, in which we prefer Rectified Linear Units [58] for the creation of sparse samples [19].

In the discriminator (shown in Fig. 5), the feature space produced by each LAGAN-style output stream is
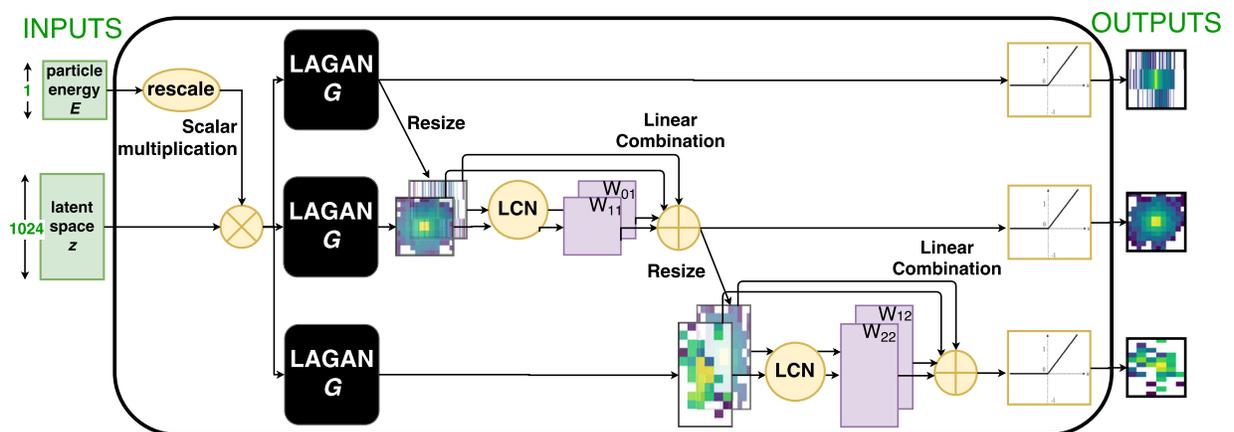

FIG. 4. Composite generator, illustrating three stream with attentional layer-to-layer dependence.
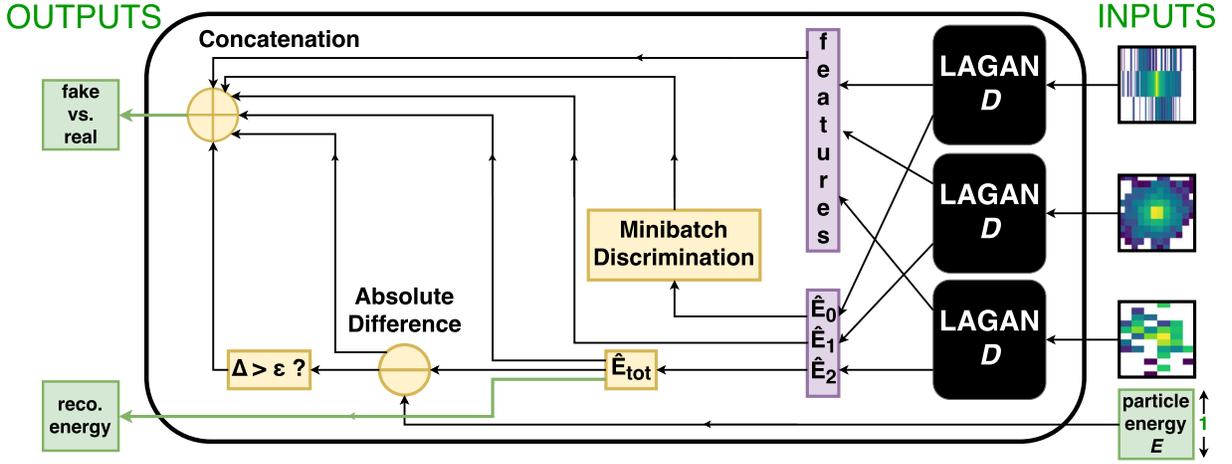
FIG. 5. Composite discriminator, depicting additional domain specific expressions included in the final feature space.

augmented with a sub-differentiable version of sparsity percentage [59], as well as minibatch discrimination [48] on both the standard locally connected network-produced features and the output sparsity itself, to ensure a well examined space of sparsities. These are represented in Fig. 5 by the "features" vector.

The discriminator is further customized with domain-specific features to ensure fidelity of samples. Given the importance of matching the requested energy E, D directly calculates the empirical energy per layer $\hat{E}_i$, $i \in \{0, 1, 2\}$, as well as the total energy $\hat{E}_{\text{tot}}$. Minibatch discrimination is performed on this vector of per-layer energies to ensure a proper distributional understanding. We also add $|E - \hat{E}_{\text{tot}}|$ as a feature, as well as $\mathbb{1}_{\{|E-\hat{E}_{\text{tot}}|>\varepsilon\}}$ with $\varepsilon = 5$ GeV—a binary, sub-differentiable feature which encodes the tolerance for GAN-produced scatterings to be incorrect in their reconstructed energy.

Further specifications of the exact hyper-parameter and architectural choices as well as software versioning constraints are available in the source code [60].

Two additional architectural modifications were tested in order to build a particle-type conditioning system directly into the learning process. Neither the AC-GAN [43] nor the conditional GAN [44] frameworks were able to handle the substantial differences among the three particle types. We suspect that both a significantly richer model and a larger latent space could alleviate some problems associated with conditioning using the investigated approaches. Although building a fully joint model is an interesting machine learning challenge, the practicality and flexibility of this application may suffer from having one single model for all particle showers.

B. Loss formulation

In this work, we augment the classical adversarial loss term $\mathcal{L}_{\text{adv}}$ [Eq. (1)]—which penalizes the system whenever D fails to classify samples originating from generated or target distributions—with a mean absolute error term,

$$\mathcal{L}_E = \mathbb{E}_{z \sim p_z(z)}[\delta(E, \hat{E}(G(z)))] + \mathbb{E}_{I \sim f}[\delta(E, \hat{E}(I))], \quad (2)$$

where $\delta(e, e') = |e - e'|$, E is the requested energy, and $\hat{E}$ is the reconstructed energy. This allows us to penalize instances of too little or too much deposited energy. This solution not only helps ensuring the confinement of the generated energy to a desirable range, but it also allows us to encode a "soft" physical notion of conservation of energy, according to which no more energy than the initial E of the incoming particle can be physically collected by the detector.

Note, however, that this formulation discourages, but does not forbid, a deposition of more energy than was requested. We can remedy this unphysical result by sampling from a conditional distribution until energy preservation is met. This issue is further addressed in Sec. V B.

During training, the generator will maximize Eq. (3), and the discriminator will maximize Eq. (4):

$$\mathcal{L}_{\text{generator}} = \lambda_E \mathcal{L}_E - \mathcal{L}_{\text{adv}} \quad (3)$$

$$\mathcal{L}_{\text{discriminator}} = \lambda_E \mathcal{L}_E + \mathcal{L}_{\text{adv}} \quad (4)$$

C. Training strategy

λ_E is set to 0.05 to down-weight the importance of λ_E compared to $\mathcal{L}_{\text{adv}}$ and rescale the absolute error, which is measured in GeV, to a comparable range with respect to $\mathcal{L}_{\text{adv}}$. This hyper-parameter can be tuned in a systematic way, but with minimal tuning, we were able to find a reasonable value.

The weights in the generator and discriminator are optimized in an alternating fashion over a set of 100 000 GEANT4-simulated events for each particle type in batches

of 256, using the ADAM optimizer [61]. The discriminator has a learning rate of 2×10^{-5}, and the generator has a learning rate of 2×10^{-4}. We note that outside of initial rough hyper-parameter tuning, we perform no dedicated optimization per particle type, and simply apply the same training parameters to all three networks. We expect significant performance improvements (especially for pions) with dedicated training.

Each system is trained for 50 epochs. Sixteen NVIDIA K80 graphics cards are used for initial hyper-parameter sweeps, with two Titan X Pascal Architecture cards used for final training. KERAS v2.0.3 [62] is used to construct all models, with the TENSORFLOW v1.1.0 backend [63].

V. PERFORMANCE

As discussed in [53], there exist several methods to qualitatively and quantitatively assess the performance of generative networks, but not all evaluation criteria are equally suitable and reliable for all applications. In this paper, we choose application-driven methods focused on sample quality. A first qualitative assessment will be accompanied by a quantitative evaluation based on physics-driven similarity metrics. The choice reflects the domain specific procedure for data-simulation comparison. These similarity metrics are based on one-dimensional statistics of the shower probability distribution. Visualizing and verifying the performance in higher dimensions is a challenge. One way to probe study the modeling of higher dimensions is to study ability to classify showers from different particles. This is studied in Sec. V C.

A. Qualitative assessment

We first examine the average calorimeter deposition per voxel (a volumetric pixel). On average, the systems learn a complete picture of the underlying physical processes governing the cascades of e^+, γ, and π^+ with uniform energy between 1 and 100 GeV (Figs. 6, 7, and 8).

Diversity and overtraining concerns can be investigated by considering the nearest neighbors among the training and generated data sets. Figs. 9, 10, and 11 shows five randomly selected events and their GAN-generated nearest neighbors for all three calorimeter layers for e^+, γ and π^+ showers, respectively. Good qualitative agreement can be found between the two distributions across all layers, without obvious signs of mode collapse: a failure mode in which the generator learns to produce a small subset of samples from the distribution. Compared to the other two particle types explored in this application, at the individual image level, charged pions clearly display a higher degree of complexity and diversity in their showers. Some π^+ deposit energy in all cells of a given layer, some only hitting a handful of them. This is because charged pions undergo nuclear interactions in addition to electromagnetic interactions.
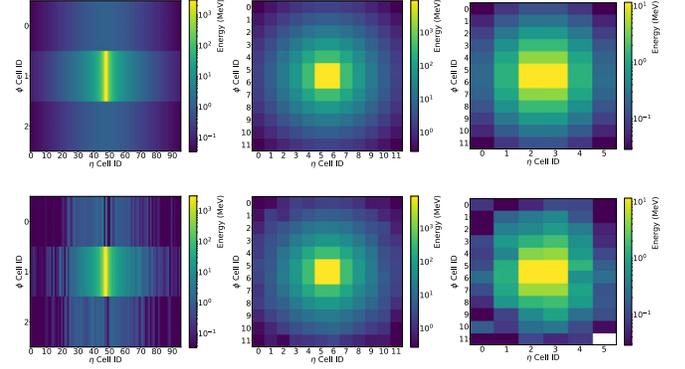

FIG. 6. Average e^+ GEANT4 shower (top), and average e^+ CALOGAN shower (bottom), with progressive calorimeter depth (left to right).
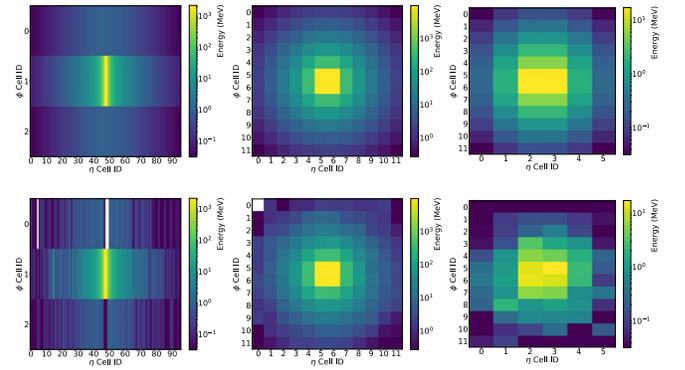

FIG. 7. Average γ GEANT4 shower (top), and average γ CALOGAN shower (bottom), with progressive calorimeter depth (left to right).
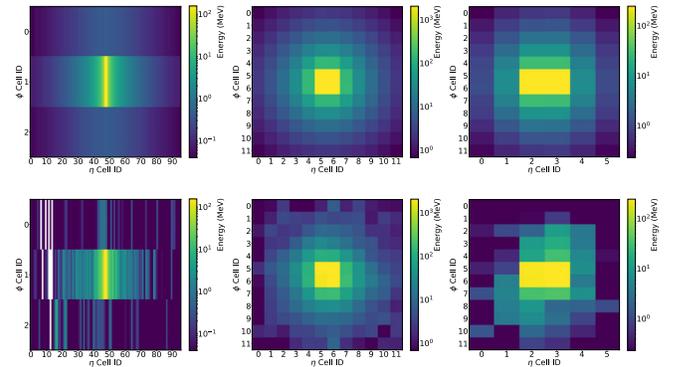

FIG. 8. Average π^+ GEANT4 shower (top), and average π^+ CALOGAN shower (bottom), with progressive calorimeter depth (left to right).

B. Shower shapes

Electron and photon classification and energy calibration use properties of the calorimeter shower [64–67]. These same features can be used to quantitatively assess the quality of the GAN samples. The list of features used for
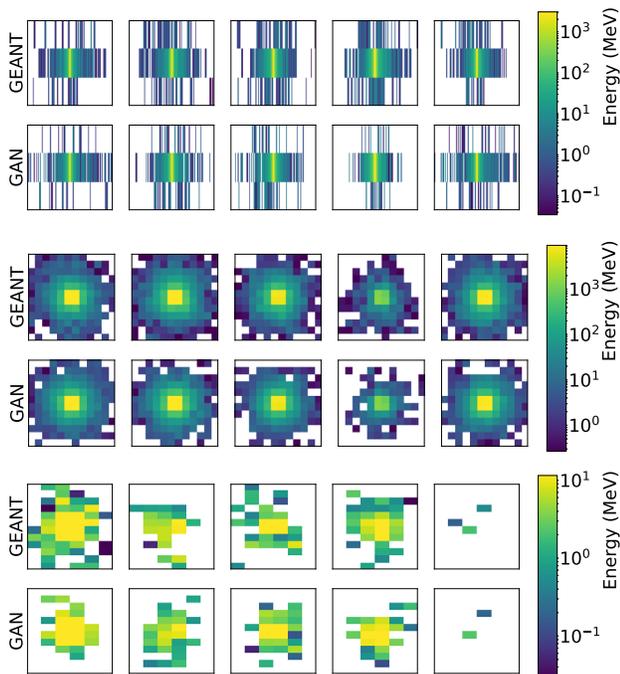
FIG. 9. Five randomly selected e^+ showers per calorimeter layer from the training set (top) and the five nearest neighbors (by euclidean distance) from a set of CALOGAN candidates.

FIG. 10. Five randomly selected γ showers per calorimeter layer from the training set (top) and the five nearest neighbors (by euclidean distance) from a set of CALOGAN candidates.

FIG. 11. Five randomly selected π^+ showers per calorimeter layer from the training set (top) and the five nearest neighbors (by euclidean distance) from a set of CALOGAN candidates.
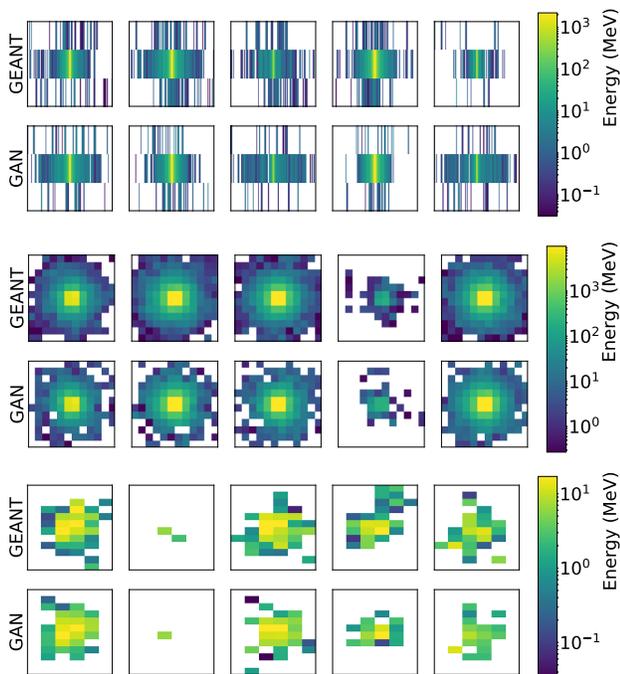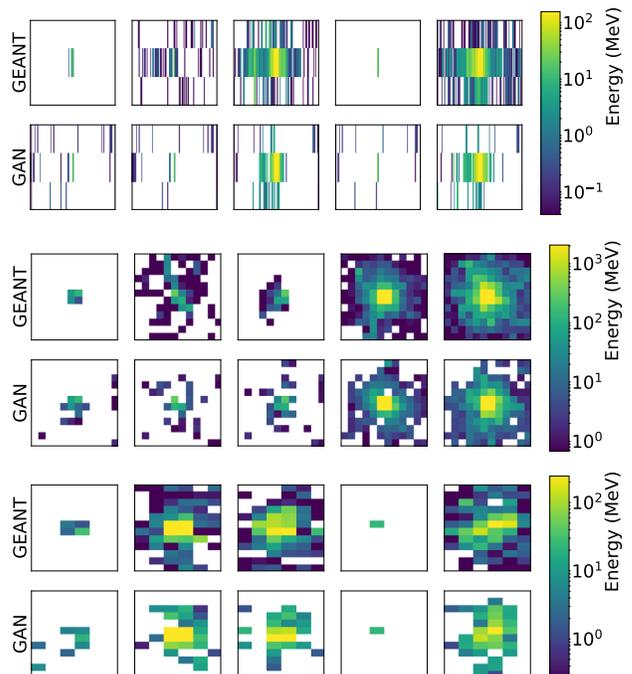
evaluation is provided in Table IV in the Appendix. The key physical quantity that governs the shapes of these distributions is the number of radiation lengths X_0 that are traversed by the particle. By definition, X_0 is the distance an electron will travel before its energy is reduced to $1/e$ on average. The equivalent distance for photons is slightly larger (by a factor of 9/7 [68]) and is set by the mean free path for pair production. The transverse shower size is also proportional to X_0. For a brief review, see e.g. [68].

The one-dimensional distributions for GEANT4- and GAN-generated samples are available in Fig. 12. Although the sparsity levels per layer are only roughly matched, note that, for the majority of the remaining variables, the GAN picks up on complex features in the distributions across several orders of magnitude and all particles types. The unique features that pions exhibit, compared to the other particles, make it unfavorable to train a single model for multiple particle types.

Note that shower shape variables were not explicitly part of the training, which is based only on the distribution of pixel intensities and energy. In the future, one can integrate the shower shape distributions into the loss-function itself. For now, we have left them out for a comprehensive validation assessment.

In addition to comparing shower shapes to reference distribution, we want to measure the quality of conditioning on energy. As outlined in Sec. IV B, we cannot explicitly impose conservation of energy, but one can devise a simple sampling system to only keep simulated showers that obey this constraint.
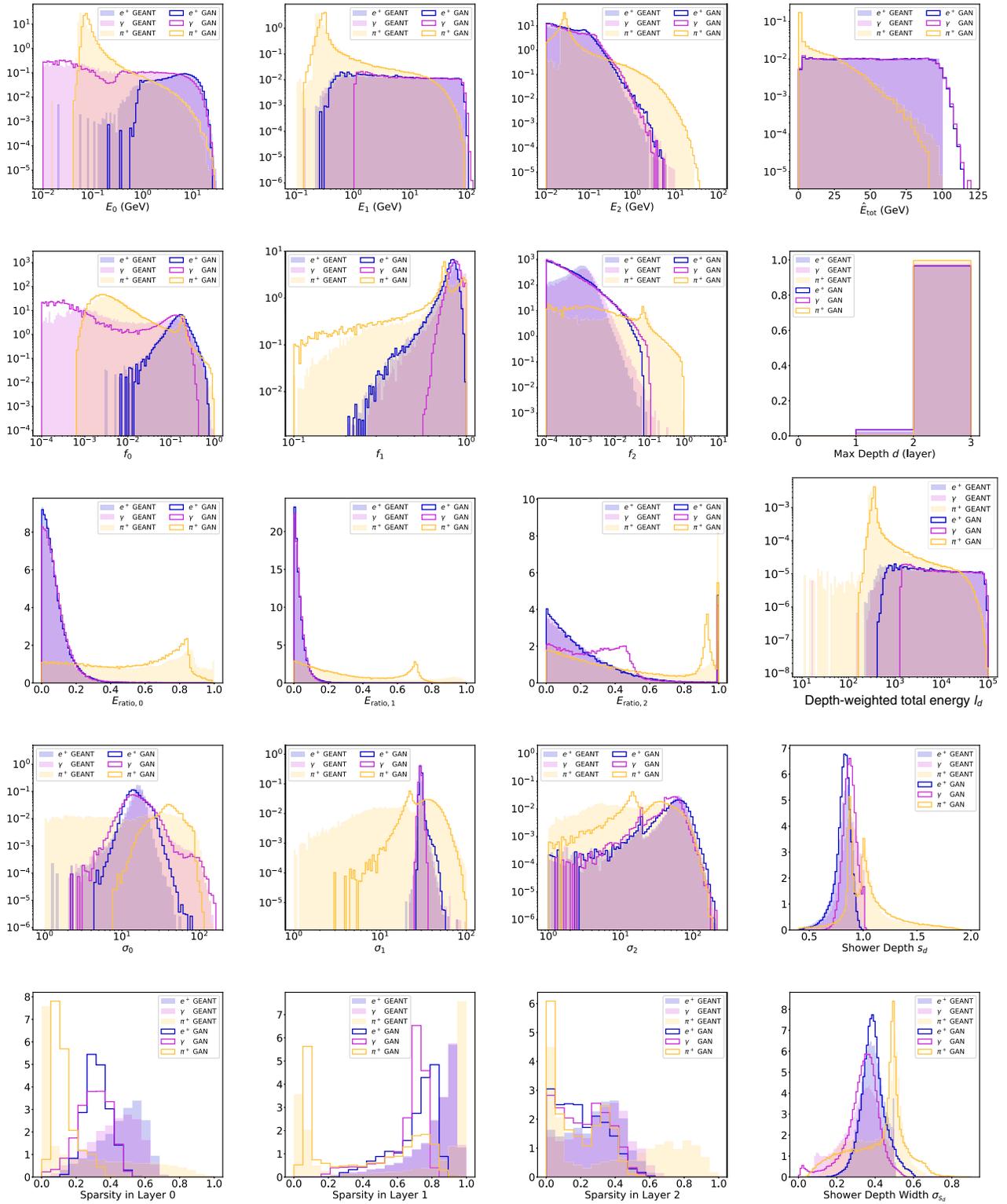
FIG. 12.   Comparison of shower shape variables, introduced in Table IV, and other variables of interest, such as the sparsity level per layer, for the GEANT4 and CALOGAN data sets for e^+, γ and π^+.
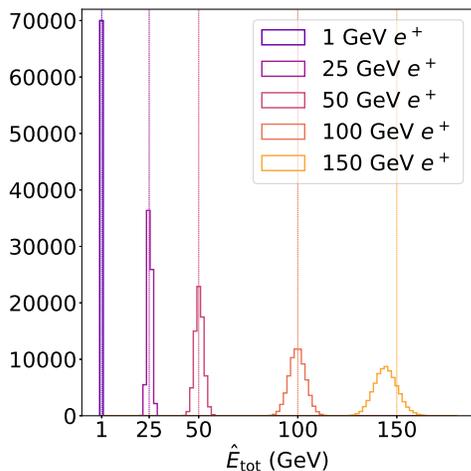
FIG. 13.   Post energy-conditioned empirical energy response for e^+ incident at 1, 25, 50, 100, and 150 GeV. Though our model is only trained on the unform range between 1 and 100 GeV, it still admits a compelling peak at 150 GeV.

TABLE II.   Mean and standard deviation over 10 particle classification trials using a six-layer fully connected network with dropout. The networks are trained using a data set from the domain specified in the first column, and tested on an independent data set from the domain specified in the header.

		e^+ vs π^+	
		Test on	
		GEANT4	CALOGAN
Train on	GEANT4	$99.6\% \pm 0.1\%$	$96.5\% \pm 1.1\%$
	CALOGAN	$98.2\% \pm 0.9\%$	$99.9\% \pm 0.2\%$
		e^+ vs γ	
		Test on	
		GEANT4	CALOGAN
Train on	GEANT4	$66.1\% \pm 1.2\%$	$70.6\% \pm 2.6\%$
	CALOGAN	$54.3\% \pm 0.8\%$	$100.0\% \pm 0.0\%$

As can be noted in the e^+ example in Fig. 13, our loss formulation coupled with the uniform training distribution admits an approximately symmetric conditional output energy distribution. In Fig. 13, note that the vertical lines that approximately coincide with the mode of each distribution represent the requested energy, and could easily be used as a threshold on selecting physical events. A noteworthy feature of this system is that one can request energies that lie outside the trained region (capped at 100 GeV in this application), to which a trained CALOGAN will return samples around the requested energy level—though with broader width, and mode shifted towards the training domain. Whether or not these extrapolated samples obey shower shape distributions and other metrics is left as future work.

C. Classification as a performance proxy

Transferability of classification performance from GAN-generated samples to GEANT4-generated samples can be used as a proxy both for CALOGAN image quality and potential utility in a practical fast simulation setting.

We perform ten identical trainings of simple six-layer fully-connected $e - \gamma$ and $e - \pi$ classifiers, and we report the accuracies for in-domain and out-of-domain testing (Table II) along with the following observations:

(i)   when training on GEANT4, testing on the generated CALOGAN data set yields similar results to testing on a separate GEANT4 data set, leading us to believe that the GAN has learned most of the discriminating physics between the classes of particles. Note that percent-level differences in accuracy may however be relevant for particular applications;

(ii)   the significantly higher performance obtained on the CALOGAN-generated test set when training on a

separate data set of CALOGAN-generated images highlights a greater inter-class differentiation in the GAN synthetic data set than originally present in the target GEANT4 distribution.

This could either be due to new unphysical, class-dependent features produced by the GAN, or to the inability of the GAN to cover the entire feature space for at least one of the particle classes. It is likely that both of these contribute. To some extent, unphysical features are mitigated by the discriminator network of the GAN training itself, but both physical and unphysical features that are not very useful for distinguishing real from fake could turn into very useful features for the two-particle classification case. Such information would therefore appear discriminative in GAN images but not in GEANT4. While classification is a useful metric for probing the high-dimensional feature space and shows promising results, there are still challenges for interpreting and improving upon the outcome.

D. Computational performance

In addition to the promise of being a high-fidelity fast simulation paradigm and respecting many shower shape variables, the CALOGAN affords many orders of magnitude in computational speedups [69]. We benchmark generation time on e^+ with incident energy drawn uniformly between 1 GeV and 100 GeV. GEANT4 and CALOGAN on CPU are benchmarked on nearly identical compute-nodes on the PDSF distributed cluster at the National Energy Research Scientific Computing Center (NERSC), and numerical results are obtained over an average of 100 runs. CALOGAN on GPU hardware is benchmarked on an Amazon Web Service (AWS) p2.8xlarge instance, where a single NVIDIA® K80 is used for the purposes of benchmarking.

TABLE III. Total expected time (in milliseconds) required to generate a single shower under various algorithm-hardware combinations.

Simulator	Hardware	Batch size	ms/shower
GEANT4	CPU	N/A	1772
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012

In Table III, we show the time-to-generate a single particle shower in milliseconds. We provide different batch sizes for CALOGAN, as we expect different use-cases will have different demands around batching computation. We note that a batch can accept any number of different requested energies. With the largest batch sizes on GPU, our method admits a speedup of 5 orders of magnitude compared to the single-threaded GEANT4 benchmark. In addition, generation time with GEANT4 scales with incident energy, whereas computational time is flat as a function of incident energy for the CALOGAN.

1. Implementation notes

As noted previously in Sec. IV A, separating per-particle-type CALOGAN architectures and implementations affords many benefits. It is easy to imagine a situation where the life cycles surrounding models for different particle types are very different. In addition, this allows for total independence of versioning, framework, or language.

When possible, any GAN should maximally employ batching—we imagine most applications can request all showers from one event simultaneously, maximally taking advantage of CPU/GPU while minimizing data transfer overhead.

VI. CONCLUSIONS AND FUTURE OUTLOOK

Using modern generative deep neural network techniques, we have generated three-dimensional electromagnetic showers in a multilayer sampling LAr calorimeter with uneven spatial segmentation, while attempting to preserve spatio-temporal relation among layers. Our approach infused physics domain knowledge and reproduced many aspects of key shower shape properties comparable to the ones in the GEANT4 full simulation. We showed the possibility of up to five orders of magnitude decrease in computing time.

Future work will focus on improving performance by drawing from the recent machine learning developments in GAN training procedures, as well as testing the direct inclusion of important shower shape variables as constraints at training time. Further developments will build on this result and continue expanding the complexity of the training data set to include incoming particles at different locations and angles within the detector, as well as the hadronic calorimeter. Concurrent plans include contributing to testing the computational performance on high performance computing (HPC) clusters, and porting these solutions into the simulation packages used by the nuclear and particle physics communities, in order for the various experiments to be able to maximally benefit from this new technology.

APPENDIX: SHOWER SHAPE VARIABLES

Table IV contains the description and mathematical definition of the shower shape variables used to compare the generated and target distributions. These are defined as functions of $\mathcal{I}_i$, the vector of pixel intensities for an image in layer i, where $i \in \{0, 1, 2\}$.

TABLE IV. One-dimensional observables used to assess the quality of the GAN samples.

Shower shape variable	Formula	Notes
E_i	$E_i = \sum_{\text{pixels}} \mathcal{I}_i$	Energy deposited in the ith layer of calorimeter
E_{tot}	$E_{\text{tot}} = \sum_{i=0}^{2} E_i$	Total energy deposited in the electromagnetic calorimeter
f_i	$f_i = E_i/E_{\text{tot}}$	Fraction of measured energy deposited in the ith layer of calorimeter

(Table continued)

TABLE IV. *(Continued)*

Shower shape variable	Formula	Notes
$E_{\mathrm{ratio},i}$	$\frac{\mathcal{I}_{i,(1)}-\mathcal{I}_{i,(2)}}{\mathcal{I}_{i,(1)}+\mathcal{I}_{i,(2)}}$	Difference in energy between the highest and second highest energy deposit in the cells of the ith layer, divided by the sum
d	$d = \max\{i\colon \max(\mathcal{I}_i) > 0\}$	Deepest calorimeter layer that registers nonzero energy
Depth-weighted total energy, l_d	$l_d = \sum_{i=0}^{2} i \cdot E_i$	The sum of the energy per layer, weighted by layer number
Shower depth, s_d	$s_d = l_d / E_{\mathrm{tot}}$	The energy-weighted depth in units of layer number
Shower depth width, σ_{s_d}	$\sigma_{s_d} = \sqrt{\frac{\sum_{i=0}^{2} i^2 \cdot \mathcal{I}_i}{E_{\mathrm{tot}}} - \left(\frac{\sum_{i=0}^{2} i \cdot \mathcal{I}_i}{E_{\mathrm{tot}}}\right)^2}$	The standard deviation of s_d in units of layer number
ith Layer lateral width, σ_i	$\sigma_i = \sqrt{\frac{\mathcal{I}_i \odot H^2}{E_i} - \left(\frac{\mathcal{I}_i \odot H}{E_i}\right)^2}$	The standard deviation of the transverse energy profile per layer, in units of cell numbers

[1] GEANT4 Collaboration, Nuclear Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).

[2] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **70**, 823 (2010).

[3] R. Rahmat, R. Kroeger, and A. Giammanco, J. Phys. Conf. Ser. **396**, 062016 (2012).

[4] M. Czakon and A. Mitov, Comput. Phys. Commun. **185**, 2930 (2014).

[5] M. Botje *et al.* (2011), arXiv:1101.0538.

[6] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **64**, 653 (2009).

[7] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C. P. Yuan, Phys. Rev. D **89**, 033009 (2014).

[8] R. D. Ball *et al.*, Nucl. Phys. **B867**, 244 (2013).

[9] V. Khachatryan *et al.* (CMS Collaboration), Eur. Phys. J. C **77**, 172 (2017).

[10] M. Aaboud *et al.* (ATLAS Collaboration), Eur. Phys. J. C **77**, 292 (2017).

[11] J. Flynn, CERN, Technical Report No. CERN-RRB-2015-014, 2015.

[12] E. Karavakis, J. Andreeva, S. Campana, S. Gayazov, S. Jezequel, P. Saiz, L. Sargsyan, J. Schovancova, I. Ueda (ATLAS Collaboration), J. Phys. Conf. Ser. **513**, 062024 (2014).

[13] C. Bozzi, CERN, Technical Reports No. LHCb-PUB-2015-004 and No. CERN-LHCb-PUB-2015-004, 2015.

[14] G. Grindhammer and S. Peters, in *International Conference on Monte Carlo Simulation in High-Energy and Nuclear Physics—MC 93 Tallahassee, Florida, 1993* (1993) arXiv:hep-ex/0001020.

[15] M. Beckingham, M. Duehrssen, E. Schmidt, M. Shapiro, M. Venturi, J. Virzi, I. Vivarelli, M. Werner, S. Yamamoto, and T. Yamanaka (ATLAS Collaboration), CERN, Technical Report No. ATL-PHYS-PUB-2010-013, 2010.

[16] G. Grindhammer, M. Rudowicz, and S. Peters, in *Proceedings of SSC Workshop on Calorimetry for the Superconducting Super Collider Tuscaloosa, Alabama*, 1989; Nucl. Instrum. Methods Phys. Res., Sect. A **290**, 469 (1990).

[17] E. Barberio *et al.*, J. Phys. Conf. Ser. **160**, 012082 (2009).

[18] The ATLAS Collaboration, CERN Technical Report No. ATL-SOFT-PUB-2014-001, 2014.

[19] L. de Oliveira, M. Paganini, and B. Nachman, Comput. Softw. Big Sci. **1**, 4 (2017).

[20] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman, J. High Energy Phys. 02 (2015) 118.

[21] D. P. Kingma and M. Welling, arXiv:1312.6114.

[22] A. Makhzani, J. Shlens, N. Jaitly, and I. J. Goodfellow, arXiv:1511.05644.

[23] S. Ravanbakhsh, F. Lanusse, R. Mandelbaum, J. Schneider, and B. Poczos, arXiv:1609.05796.

[24] K. Schawinski, C. Zhang, H. Zhang, L. Fowler, and G. K. Santhanam, Mon. Not. R. Astron. Soc. **467**, L110 (2017).

[25] L. Mosser, O. Dubrule, and M. J. Blunt, Phys. Rev. E **96**, 043309 (2017).

[26] A. Kadurin, A. Aliper, A. Kazennov, P. Mamoshina, Q. Vanhaelen, K. Khrabrov, and A. Zhavoronkov, Oncotarget **8** 10883 (2017).

[27] Energy losses factorize, but detector readout does not. Due to threshold and digitization effects, the energy readout from two energy deposits in different detector elements need not be the same as the recorded energy from the two deposits in the same element. In detector simulations, these nonlinear effects are treated after accounting for the material interactions and are therefore beyond the scope of the CALOGAN. It may be interesting in future work to consider an end-to-end generator that includes these effects, but it may not save a lot of time since simulation is much more costly than reconstruction.

[28] The ATLAS Collaboration, CERN Technical Design Report no. ATLAS CERN-LHCC-96-041, 1996.

[29] This is the direction that prompt neutral particles at $\eta = 0$ would enter the calorimeter without any prior material interactions.

[30] B. Nachman, L. de Oliveira, and M. Paganini, Electromagnetic calorimeter shower images, 10.17632/pvn3xc3wy5.1.

[31] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, Nucl. Phys. **B281**, 289 (1987).

[32] B. Andersson, A. Tai, and B.-H. Sa, Zeitschrift für Physik C Particles and Fields **70**, 499 (1996).

[33] B. Nilsson-Almqvist and E. Stenlund, Comput. Phys. Commun. **43**, 387 (1987).

[34] B. Ganhuyag and V. Uzhinsky, Czech. J. Phys. **47**, 913 (1997).

[35] M. P. Guthrie, R. G. Alsmiller, and H. W. Bertini, Nucl. Instrum. Methods **66**, 29 (1968).

[36] H. W. Bertini and M. P. Guthrie, Nucl. Phys. **A169**, 670 (1971).

[37] V. A. Karmanov, Nucl. Phys. **B166**, 378 (1980).

[38] H. Burkhardt, V. M. Grichine, P. Gumplinger, V. N. Ivanchenko, R. P. Kokoulin, M. Maire, and L. Urban, in *IEEE Symposium Conference Record Nuclear Sciencem, 2004* (IEEE, Rome, Italy, 2004), Vol. 3, p. 1907.

[39] For the purposes of this study, the cell energy is the sum of the energy deposited in the lead and the argon; in practice, only the LAr energy deposits are measured. Dividing out these two components is left for future work (see Sec. VI).

[40] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, arXiv:1406.2661.

[41] I. J. Goodfellow, arXiv:1412.6515.

[42] A. Radford, L. Metz, and S. Chintala, arXiv:1511.06434.

[43] A. Odena, C. Olah, and J. Shlens, arXiv:1610.09585.

[44] M. Mirza and S. Osindero, arXiv:1411.1784.

[45] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, arXiv:1606.03657.

[46] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas, arXiv:1612.03242.

[47] M. Arjovsky, S. Chintala, and L. Bottou, arXiv:1701.07875.

[48] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, arXiv:1606.03498.

[49] S. Nowozin, B. Cseke, and R. Tomioka, arXiv:1606.00709.

[50] G. Qi, arXiv:1701.06264.

[51] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, arXiv:1611.04076.

[52] M. Lin, arXiv:1704.06191.

[53] L. Theis, A. van den Oord, and M. Bethge, arXiv:1511.01844.

[54] This covers a significant portion of the challenge; in practice, the fast simulation must also take the $\eta - \phi$ position of incidence and the incidence angles. We leave this to future work (see Sec. VI).

[55] The dimensionality of the latent space is a hyper-parameter that need not be the same as the dimensionality of the target space.

[56] S. Ioffe and C. Szegedy, arXiv:1502.03167.

[57] A. L. Maas, A. Y. Hannun, and A. Y. Ng, in *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing* (2013).

[58] X. Glorot, A. Bordes, and Y. Bengio, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, edited by G. Gordon, D. Dunson, and M. Dudák, Proceedings of Machine Learning Research Vol. 15 (PMLR, Fort Lauderdale, FL, USA, 2011), pp. 315–323.

[59] This is the fraction of hit pixels (occupancy). Subderivatives are a generalization of derivatives to cases with kinks.

[60] M. Paganini, L. de Oliveira, and B. Nachman, hep-lbdl/calogan: CaloGAN generation, training, and analysis code.

[61] D. P. Kingma and J. Ba, arXiv:1412.6980.

[62] F. Chollet and Keras, https://github.com/fchollet/keras (2017).

[63] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from https://www.tensorflow.org/.

[64] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **72**, 1909 (2012).

[65] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **74**, 2941 (2014).

[66] M. Aaboud *et al.* (ATLAS Collaboration), Eur. Phys. J. C **76**, 666 (2016).

[67] M. Aaboud *et al.* (ATLAS Collaboration), Eur. Phys. J. C **77**, 195 (2017).

[68] K. A. Olive *et al.* (Particle Data Group), Chin. Phys. C **38**, 090001 (2014).

[69] Note that nondistributed training can take day(s), depending on the total number of training epochs, but it always executes in constant time regardless of the number of showers requested at generation time, so it is a fixed cost that is not relevant for the majority of the total computing budget.