EPJ Web of Conferences **251**, 03049 (2021) https://doi.org/10.1051/epjconf/202125103049
CHEP 2021

Fast and Accurate Electromagnetic and Hadronic Showers from Generative Models

Erik Buhmann[1], *Sascha* Diefenbacher[1,2,*], *Engin* Eren[2], *Frank* Gaede[2], *Daniel* Hundhausen[1], *Gregor* Kasieczka[1], *William* Korcari[1], *Anatolii* Korol[4], *Katja* Krüger[2], *Peter* McKeown[2], and *Lennart* Rustige[2,3]

[1]Institut für Experimentalphysik, Universität Hamburg, Germany
[2]Deutsches Elektronen-Synchrotron, Germany
[3]Center for Data and Computing in Natural Sciences, Germany
[4]Taras Shevchenko National University of Kyiv, Ukraine

Abstract. Generative machine learning models offer a promising way to efficiently amplify classical Monte Carlo generators' statistics for event simulation and generation in particle physics. Given the already high computational cost of simulation and the expected increase in data in the high-precision era of the LHC and at future colliders, such fast surrogate simulators are urgently needed. This contribution presents a status update on simulating particle showers in high granularity calorimeters for future colliders. Building on prior work using Generative Adversarial Networks (GANs), Wasserstein-GANs, and the information-theoretically motivated Bounded Information Bottleneck Autoencoder (BIB-AE), we further improve the fidelity of generated photon showers. The key to this improvement is a detailed understanding and optimisation of the latent space. The richer structure of hadronic showers compared to electromagnetic ones makes their precise modeling an important yet challenging problem. We present initial progress towards accurately simulating the core of hadronic showers in a highly granular scintillator calorimeter.

1 Introduction

Simulation in high energy physics links our deep understanding of fundamental theories and the behaviour of detectors to experimental data. While crucial to modern data analysis, the computational cost associated with the production of these Monte Carlo-based simulations is already a limiting factor at the Large Hadron Collider (LHC) and becomes potentially prohibitive at planned higher luminosities and future experiments [1].

Generative machine learning models offer a possible way out. These models can be trained on samples from slow simulators and amplify their statistics [2]. A large number of potential network architectures such as Generative Adversarial Networks (GANs) [3], Variational Autoencoders (VAEs) [4] and models based on autoregressive flows [5] were proposed for this task. Since the initial idea to use GANs to simulate calorimeter showers in high energy physics [6], several extensions and improvements were suggested [7–14].

*e-mail: sascha.daniel.diefenbacher@uni-hamburg.de

We previously investigated the use of generative models for simulating electromagnetic showers in a $30 \times 30 \times 30$ cubic region of the SiW electromagnetic calorimeter proposed for the International Large Detector (ILD) [15]. This study achieved, for the first time, a good description of several differential distributions for high granularity calorimeters — including the minimum ionizing particle (MIP) peak in the hit energy spectrum. There, we used three generative approaches: a standard GAN; a GAN based on the Wasserstein distance [16]; and the Bounded Information Bottleneck Autoencoder (BIB-AE) architecture [17] that unifies features of GANs and VAEs.

However, there is still a long way to go before the fidelity and versatility required for real-world application is reached. In this contribution, we report progress in two directions. First, we show how tuning the generators based on an understanding of the latent space can improve the performance on distributions that are difficult to model. A more detailed look at the latent space is given in the companion contribution [18]. Second, we present initial results on simulating hadronic showers with a precision comparable to electromagnetic ones.

We first introduce the data sets in Sec. 2 and review the key generative architectures and improvements with respect to Ref. [15] in Sec. 3. An overview of results for both generative tasks is presented in Sec. 4, and Sec. 5 concludes this work.

2 Datasets

This Section introduces the two data sets used for training and evaluation of the generative models.

2.1 Photon Data

The ILD [19] detector is one of two detector concepts proposed for the ILC. It is optimized for Particle Flow, an algorithm that aims at reconstructing every individual particle in the sub-detector with the best energy resolution in order to optimize the overall detector resolution. ILD combines high-precision tracking and vertexing capabilities with very good hermiticity and highly granular electromagnetic and hadronic calorimeters.

One of the two proposed electromagnetic calorimeters for ILD, the Si-W ECal is chosen for this data set. It consists of 30 active silicon layers in a tungsten absorber stack with 20 layers of 2.1 mm followed by 10 layers of 4.2 mm thickness respectively. The silicon sensors have 5×5 mm^2 cell sizes. We project the cells with energy depositions (hits) onto a rectangular grid of $30 \times 30 \times 30$ cells. An example of a typical photon shower in this setup is shown in Fig. 1 (left). As the cells of the silicon wafers are not perfectly aligned from one layer to the next due to the geometry of the octagonal barrel calorimeter, we observe some artefacts, such as empty rows or columns in the projection. These are corrected for, such that every cell in our regular grid corresponds to exactly one sensor.

Our total photon data set consists of 950k showers with photon energies ranging from 10-100 GeV. Additionally, we use statistically independent data sets for evaluation, consisting of one set with 40k showers with energies uniformly distributed in the 10-100 GeV range and 7 more sets each containing 4k showers from photons with discrete energies from 20 GeV to 90 GeV in steps of 10 GeV. This is the same dataset as used for Ref. [15] and we refer to that publication for additional details.

2.2 Pion Data

The simulated pion data set is generated by shooting positively-charged pions (π^+) directly into the Analogue Hadron Calorimeter (AHCal) of the ILD detector. The AHCal consists of

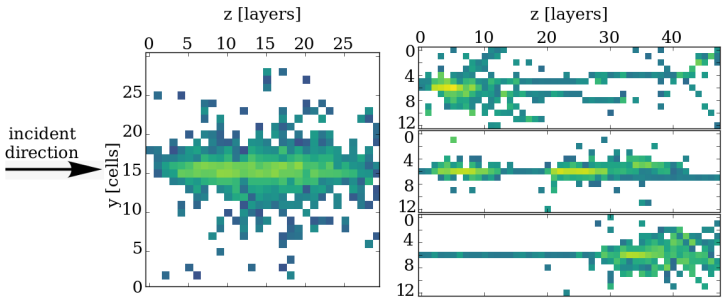

Figure 1. Projections of one photon shower (left) and three pion showers (right), indicating the large diversity among pion shower shapes. Note also the different shape of the $30 \times 30 \times 30$ photon and $48 \times 13 \times 13$ pion data.

48 layers, with scintillators as active material in a steel absorber stack. The scintillator active layers consist of $30 \times 30 \times 3$ mm^3 polystyrene tiles. For this study, we project the sensors onto a rectangular grid of $48 \times 13 \times 13$ cells. While this selection does not contain the full hadronic shower, it typically contains 98% of the energy and allows for much faster turn around times for optimizing the network architecture and training parameters. As in the case of the photon data set, we correct for artefacts, such that each cell in this grid corresponds to exactly one sensor. Figure 1 (right) shows three examples of pion showers in our data structure, indicating the large diversity of shapes among the pion showers.

To ease the networks task of learning shower structures we remove the showers from our data set where the pions simply pass through the calorimeter without showering, so called punch-through particles. We achieve this by keeping only showers that have more than 70 hits with an energy above 0.25 MeV.

Both, photon and pion showers in the ILD detector are simulated with Geant4 version 10.4 (with QGSP_BERT physics list) and a detailed and realistic detector model of ILD implemented in DD4hep [20]. The photons and pions are shot at perpendicular incident angle into the ECal (photons) and HCal (pions) barrel with energies uniformly distributed between 10-100 GeV. For pions, the ECal part of the detector, which typically lies in front of the HCal, is removed in order to avoid the complexity of handling different geometries and materials in our very first attempts for pion showers. This will be considered in future studies.

For pions, our training set is made up of 500k showers with energies uniformly sampled between 10-100 GeV. For evaluation we once again use 40k independent showers with uniformly sampled energies between 10-100 GeV and data sets with discrete pion energies from 20 GeV to 90 GeV in steps of 10 GeV, each containing 8k showers.

3 Generative Models

Generative models are designed to learn a data distribution in a way that allows subsequent sampling and thereby production of new examples. In the following, we introduce the three generative approaches used in this study: GANs, GANs optimising a Wasserstein loss, and BIB-AEs. All three models are built in PyTorch [21] using convolutional and transpose-convolutional operations. Additional details of the network architectures are provided in Ref. [15].

3.1 Generative Adversarial Network

The GAN architecture was proposed in 2014 [3] and has seen remarkable success in a number of generative tasks. It introduces generative models trained by an adversarial process, in which a generator G competes with an adversary (or discriminator) D. The goal of this

framework is to train G in order to generate samples $\widetilde{x} = G(z)$ out of noise z, that are indistinguishable from real samples x. The adversary network D is trained to maximize the probability of correctly classifying whether a given sample represents real data using a binary cross-entropy loss. The generator, on the other hand, is trained to fool the adversary D. This is represented by the loss function as

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]. \tag{1}$$

For practical applications, the GAN needs to simulate showers of a specific energy. In order to do that, we parametrize the generator and discriminator as functions of the photon energy E [22]. To achieve this, we multiply the random noise the generator receives by the energy label. This method significantly improves energy conditioning for the GAN model.

3.2 Wasserstein-GAN

One alternative to classical GAN training is to use the Wasserstein-1 distance, also known as the earth mover's distance, as a loss function. This distance evaluates the dissimilarity between two multi-dimensional distributions and informally gives the cost expectation for moving a mass of probability along optimal transportation paths [16]. Using the Kantorovich-Rubinstein duality, the Wasserstein loss can be calculated as

$$L = \sup_{f\in \mathrm{Lip}_1}\{\mathbb{E}_{x\sim p_{data}(x)}[f(x)] - \mathbb{E}_{z\sim p_z(z)}[f(G(z))]\}. \tag{2}$$

The supremum is over all 1-Lipschitz functions f, and is approximated by a discriminator network D, termed *critic*, during the adversarial training. This critic estimates the Wasserstein distance between real and generated images. In order to enforce the 1-Lipschitz constraint on the critic [23], we add a gradient penalty term to (2).

Again, we need to ensure that the generated showers accurately resemble real showers of the requested energy. This is achieved by parametrizing the generator and critic networks in E and by adding a constrainer [10] network. The constrainer network is trained solely on the Geant4 showers; its weights are fixed during the generator training.

3.3 Bounded Information Bottleneck Autoencoder

Originally introduced as an overarching generative model [17], the Bounded Information Bottleneck Autoencoder (BIB-AE) is fundamentally an Autoencoder setup. It maps the high dimensional input to a low dimensional latent space and then reconstructs the original input form this latent representation. Most GAN and AE architectures can be understood as subsections of the BIB-AE setup.

The core of our specific BIB-AE architecture consists of an encoder (E) and a decoder (D) network. The remaining components help with training these two networks and fall into two categories: Those that facilitate reconstruction and those that regularize the latent space. The reconstruction part consists of two Wasserstein-GAN like critics: one judging whether the reconstructed output looks like a realistic shower; and one comparing the output and input image, ensuring that they look alike, denoted C and C_D, respectively.

Our latent regularization is done using a KL-Divergence term — similar to the KLD loss often used in Variational Autoencoders (VAEs) [4] — and aided by an additional critic (C_L) and a Maximum Mean Discrepancy (MMD) [24] term. The individual loss terms are combined using hyperparameter weights β.

$$
\begin{aligned}
L_{\text{BIB-AE}} = &-\beta_{C_L} \cdot \mathbb{E}_{x\sim p_{data(x)}}[C_L(E(x))] \\
&-\beta_{C} \cdot \mathbb{E}_{x\sim p_{data(x)}}[C(D(E(x)))] \\
&-\beta_{C_D} \cdot \mathbb{E}_{x\sim p_{data(x)}}[C_D(D(E(x)) - x)] \\
&+\beta_{\text{KLD}} \cdot \text{KLD}(E(x)) \\
&+\beta_{\text{MMD}} \cdot \text{MMD}(E(x), \mathcal{N}(0,1)).
\end{aligned} \tag{3}
$$

Similar to Ref. [15], our BIB-AE setup makes use of a dedicated Post-Processor network, trained with an MMD loss term, to improve the hit-energy spectrum. However, it is not trained alongside the main BIB-AE networks, but separately after the BIB-AE training is finished. We also include an additional loss term that takes a sorted vector of hit energies from the input shower and the Post-Processor output and tries to minimize their element-wise distance. This added term improves the energy conditioning of the Post-Processor network.

3.4 Improvements

Following the first implementation of these techniques in [15], improvements on the generator performance are observed by introducing several optimisations as discussed in this section.

BIB-AE with KDE latent space sampling

In a VAE-based architecture, the latent space is regularized towards a known distribution (i.e. Standard Normal) for efficient sampling of newly generated data. However, an overly regularized latent space limits the amount of encoded information during training. Traditionally, hyperparameter optimization is necessary to balance an expressive latent space encoding for accurate sample reconstruction and a well regularized latent space for high generation-fidelity. This fine-tuning can be omitted by sampling directly from the encoded latent space instead of the usual Standard Normal distributions. By sampling from the non-Gaussian encoding, the models' generation fidelity is essentially the same as its reconstruction accuracy.

For an already trained model with a well regularized latent space, this sampling can, for example, be implemented by learning a multi-dimensional Kernel Density Estimation (KDE) [25] of the encoded latent space. This approach mirrors one of the sampling methods outlined in Ref. [26]. By sampling from a KDE, all encoded correlations between shower physics and latent space variables are preserved and consequently reconstructed by the decoder network. Such sampling can be applied to our already trained BIB-AE model from Ref. [15] which improves generation performance of this model for photon samples. Here, the model is conditioned on the incident particle energy, which is included as a latent space variable in the KDE. A more in-depth discussion on the latent space sampling for the BIB-AE is provided in Ref. [18].

Latent Optimization for GANs

While no encoded latent space is available for GANs, it is still possible to exploit knowledge of D to refine the latent source z. Instead of using the randomly sampled z, Ref. [27] proposes to use the optimized latent source z' such that:

$$
\Delta z = \alpha \frac{\partial f(z)}{\partial z} \qquad z' = z + \Delta z, \tag{4}
$$

where $f(z) = D(G(z))$. Intuitively, the gradient of f points in the direction that maximises the discriminator D output, which can imply better sampling [28].

As suggested by the authors of Ref. [28], we utilize the so-called natural gradient descent (NGD [29]) for latent optimization. It employs the positive semi-definite Gauss-Newton matrix to approximate the Hessian. Given the gradient of z, namely $g = \frac{\partial f(z)}{\partial z}$, in a simplified closed form, the NGD update can be written as

$$\Delta z = \frac{\alpha}{\beta + \| g \|^2} g \tag{5}$$

where β is a damping factor that regularizes the step size α. The NDG adjusts the step size via the curvature estimate $c = \frac{1}{\beta + \|g\|^2}$ and automatically smooths the scale of updates by down-scaling the gradients as their norm grows.

By including this modification of the latent space into our Wasserstein-GAN implementation we achieve a smoother and more stable training for pion showers.

Minibatch

One method to potentially avoid mode collapse, increase the variety of generated showers, and obtain a better stability in the training process, is minibatch discrimination [30]. The concept of minibatch discrimination is quite general: a discriminator model that looks at multiple examples at the same time is better equipped to avoid mode-collapse of the generator than one that considers data points in isolation.

We introduce minibatch discrimination for the GAN-photon and the BIB-AE-pion setup, while an application to the WGAN is still under investigation. In particular, our GAN setup uses a simplified version of minibatch discrimination suggested in Ref. [31] which neither adds trainable parameters nor new hyperparameters. First, we compute the standard deviation of each feature in every spatial location over the minibatch and then average these estimates over all features and spatial locations to arrive at a single value. This mean value is then broadcast to match the size of the data and subsequently used as additional (constant) feature map in the discriminator input.

In the BIB-AE implementation of minibatch discrimination, we calculate the sum and standard deviation of each critic input and then calculate the difference matrix between the sums and the difference matrix between the standard deviations in each point in a batch. These matrices are then passed through an embedding function implemented as a series of convolutions with a kernel size of one. We then aggregate the result by calculating the mean and standard deviation and pass it to the critic network. This process is done both for the normal critic inputs as well as a log-scaled version of the inputs.

4 Results

In this section, we show the performance of our generative setups by demonstrating their ability to closely reproduce the physical distributions of Geant4 showers. We compare several distributions, specifically the visible cell energy spectrum, the center of gravity along the z-axis, the radial and longitudinal energy profile, and the numbers of hits and energy sums for various particle energies.

We split our results into two parts. In Sec. 4.1 we cover photon generation and compare a GAN with minibatch discrimination, a WGAN, and a BIB-AE with KDE sampling and Post-Processor. Section 4.2 then presents initial results for generating pion showers. For this, we compare a latent optimized (LO) WGAN and a BIB-AE with KDE sampling and Post-Processor.

For all comparisons shown here, we apply an energy cutoff at 0.5 MIP, both for generative models and Geant4 data. As the ECal and HCal have different MIP values, these cutoff values differ between pions and photons, with 0.1 MeV for photons and 0.25 MeV for pions.

4.1 Photon Generation

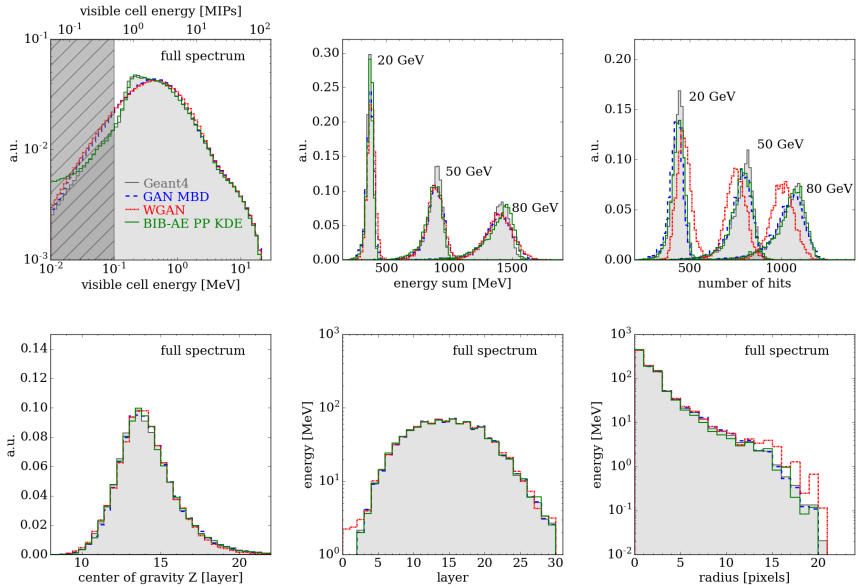

Figure 2. Differential distributions comparing physics quantities between GEANT4 and the different generative models for photon showers. In the top left the energy per-cell is measured in MeV for the bottom axis and in multiples of the expected energy deposit of a minimum ionizing particle (MIP) for the top axis. The greyed out area indicates the 0.5 MIP cutoff. The distributions labeled full spectrum are shown for showers generated with energies in the 10-100 GeV range, the remaining ones are shown with varying discrete energies.

In Fig. 2 we show the comparison between GEANT4 and our three generative models using six distributions. Starting at the top left, we see the visible cell energy spectrum. While the main body of the distribution is accurately modeled by all approaches, only the BIB-AE reproduces the bump around 0.2 MeV, thanks to the Post-Processor. In the top center, we show the visible energy distributions for 20 GeV, 50 GeV, and 80 GeV photons. Both GANs model the mean of this quite well but have a slightly more symmetric peak shape than GEANT4. This asymmetry is better captured by the BIB-AE. Next, we show the numbers of hits for the same three photon energies on the top right. While the GAN and BIB-AE properly learn this distribution, the WGAN either significantly over- or under-estimates it, depending on the photon energy.

In the bottom row, we present three distributions describing the geometrical shower properties, starting with the center of gravity in Z on the bottom left. Here we can see that all three models replicate this distribution almost perfectly. Notably, this is a significant improvement compared to the BIB-AE results shown in Ref. [15] and is due to the newly introduced KDE sampling. On the bottom center and bottom right, we show the radial and longitudinal energy profiles, respectively. Both the GANs and BIB-AE reproduce these profiles very accurately, and while the WGAN shows some discrepancies, these are mostly located in the outer regions of the shower.

In Fig. 3, we compare the visible energy deposited in the calorimeter for photons with energies ranging from 20 to 90 GeV in steps of 10 GeV. In order to easily compare this

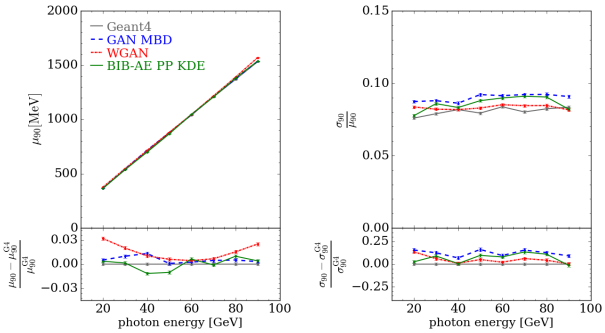

Figure 3. Mean (μ_{90}, left) and relative width (σ_{90}/μ_{90}, right) of the energy deposited in the calorimeter for various incident photon energies. In order to avoid edge effects, the phase space boundary regions of 10 and 100 GeV are removed for the response and resolution studies. In the bottom panels, the relative offset of these quantities with respect to the Geant4 simulation is shown.

we calculate the μ_{90} and σ_{90} for the energy peaks shown in Fig. 2 (top center). Note that we updated our method of calculating the μ_{90} and σ_{90} to be more in line with the common definition in Ref. [32], meaning numeric results are not directly comparable to Ref. [15]. We see that all methods reproduce the mean (left) of the visible energy quite well, showing deviations of, at most, 3%. The width (right) is not modeled quite as well but the generators still approximate Geant4 reasonably well, deviating by at most 20%.

4.2 Pion Generation

Our pion generation evaluation closely mirrors the previous section on photons. In Fig. 4 we again compare the BIB-AE and WGAN-LO results to Geant4 for the chosen six physically relevant distributions. Starting with the visible cell energy on the top right, we see that, just as in the photon case, only the BIB-AE with Post-Processor reproduces the complex peak structure around 1 MIP (0.5 MeV). For the visible energy sum (top center), both the BIB-AE and WGAN model the calorimeter response reasonably well. However, the BIB-AE seems to capture the shape of the peak better, while the WGAN peaks are more smeared out. The number of hits distributions (top right) are similar, and both models show some discrepancies compared to Geant4, with the WGAN peaks being too broad and the BIB-AE peaking at the wrong position. Conversely, the geometrical shower properties in the lower section of Fig. 4 show a high level of agreement with Geant4. For the center of gravity, both the BIB-AE and WGAN distributions almost exactly overlap with Geant4, and we only see small differences in the tail regions of the energy profiles.

In Fig. 5 we compare μ_{90} and σ_{90} between the generative models and Geant4. For the mean (left) the BIB-AE matches the linearity better than the WGAN, although both diverge for higher pion energies. The WGAN indeed overestimates the width of the energy peaks (right), as was already visible in Fig. 4. The BIB-AE learns a width closer to the Geant4 values, although it still shows discrepancies of around 10% for large sections.

Finally, we compare how well the correlations between selected shower properties match up for Geant4 and our generative models. The chosen properties are: The first and second moments in the x-, y- and z-directions, labeled as $m_{1,x}$ through $m_{2,z}$, the visible energy E_{vis}, the energy of the incoming pion E_{inc}, the number of hits n_{hit}, the ratio between the energy deposited in layers 1-16/17-32/33-48 of the calorimeter and the total visible energy, labeled E_1/E_{vis}, E_2/E_{vis} and E_3/E_{vis} respectively. We calculate the matrices of Pearson correlation coefficients for Geant4 and the generative models individually and then obtain the element-wise difference between these matrices. These difference matrices are shown in fig. 6 for the WGAN (left) and BIB-AE (right). As a general rule, the smaller the differences, the better

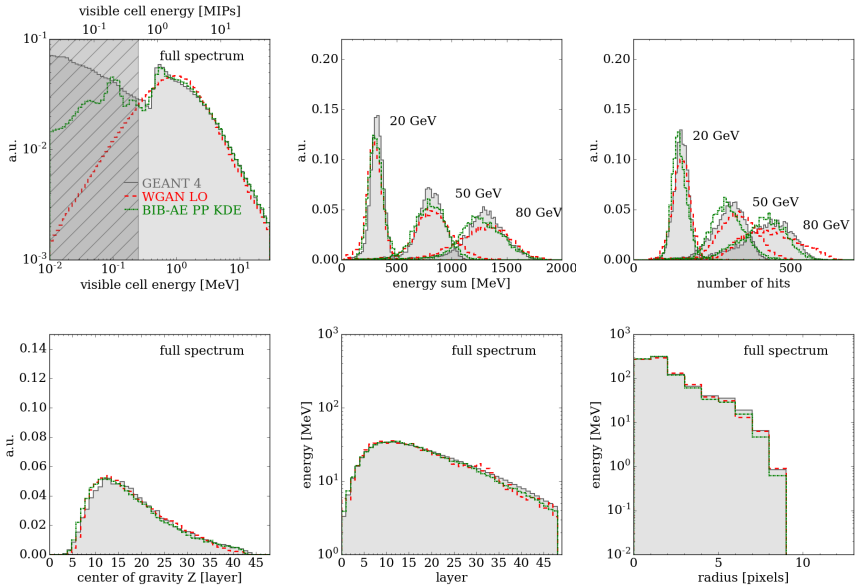

Figure 4. Differential distributions comparing physics quantities between GEANT4 and the different generative models for pion showers. In the top left the energy per-cell is measured in MeV for the bottom axis and in multiples of the expected energy deposit of a minimum ionizing particle (MIP) for the top axis. The greyed out area indicates the 0.5 MIP cutoff. The distributions labeled full spectrum are shown for showers generated with energies in the 10-100 GeV range, the remaining ones are shown with varying discrete energies.

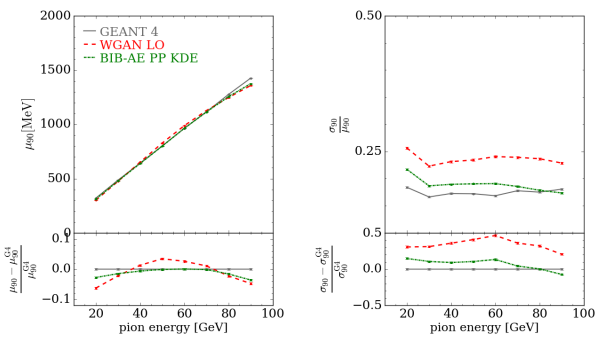

Figure 5. Plot of mean (μ_{90}, left) and relative width (σ_{90}/μ_{90}, right) of the energy deposited in the calorimeter for various incident particle energies. In order to avoid edge effects, the phase space boundary regions of 10 and 100 GeV are removed for the response and resolution studies. In the bottom panels, the relative offset of these quantities with respect to the Geant4 simulation is shown.

the agreement of the correlations between GEANT4 and generated showers. We observe that the WGAN only shows small discrepancies in the correlations, and the BIB-AE correlations match nearly perfectly.

5 Conclusions and Outlook

Motivated by the high computational cost of shower simulation in particle physics, we presented new results on modeling electromagnetic photon showers and hadronic pion showers

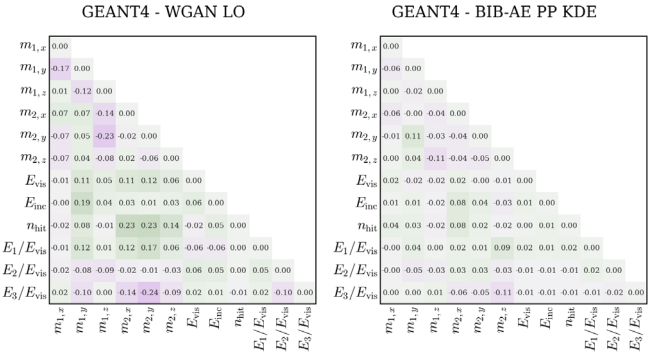

Figure 6. Difference between Linear correlation coefficients of various quantities described in the text. On the left, the difference between GEANT4 and the WGAN is shown, while the difference between GEANT4 and the BIB-AE with Post-Processor is depicted on the right.

in two high granularity calorimeter prototypes. This work builds on the generative setup introduced in Ref. [15] and i) improves the fidelity of electromagnetic showers and ii) presents first steps towards simulating hadronic showers.

The two significant improvements for photon showers are minibatch discrimination in the GAN model training and better latent space sampling via KDE-based density estimation for the BIB-AE. Minibatch discrimination improves results of the simple GAN setup across the board and. The main issue for the GAN remains the MIP peak, which the BIB-AE is able to model thanks to the Post-Processor network. Unfortunately, this Post-Processor loss function makes it significantly easier to train for an Autoencoder based architecture than for a purely generative one, limiting its applicability to the GAN or WGAN. The most notable improvement for the BIB-AE from sampling the latent space is the improved modeling of the center-of-gravity along the shower direction. This strategy is limited to latent-space-based models and discussed in more detail in the companion submission [18].

In comparison to the rather compact photon showers, pion showers exhibit a significantly larger topological variety due to the mixture of electro-magentic and nuclear interactions present in hadronic showers. Therefore they pose a much greater challenge to high fidelity generation, in particular in a high granularity calorimeter that is revealing the detailed shower sub-structure.

However, the initial agreement for WGAN and BIB-AE is encouraging, but more work is needed to reach a fidelity similar to the one achieved for electromagnetic showers. Due to the inaccuracies in the energy width, no GAN results are included for pions. Again, only the BIB-AE with Post-Processor correctly describes the MIP peak. The observed correlations are particularly promising: The differences between Pearson correlation coefficients between the BIB-AE model and GEANT4 do not exceed an absolute value of 0.11 for any combination. This improves upon the correlations for the GAN and BIB-AE models for electromagnetic showers by almost a factor of two and three, respectively.

While we see clear progress in teaching generative models increasingly complex physical processes with increasing fidelity, a long and stony road lies ahead. Future problems include simulating hadronic showers beyond the hard core, consistently handling different particle types, extending conditioning to geometrical arguments, dealing with complex geometries [1], and — of course — putting all this into production. Nevertheless, the consistently observed speed-ups by several orders of magnitude relative to classical simulations, the potential to train generative models directly on data with some control over latent space distributions [34], the promising initial results on amplification of statistics, and finally, the large gap between projected and needed computing resources strongly motivate further work in this direction.

[1] Here graph-based approaches are especially promising [33]

Acknowledgements

We would like to thank the Maxwell and National Analysis Facility (NAF) computing centers at DESY for the smooth operation and technical support. E. Buhmann is funded by a scholarship of the Friedrich Naumann Foundation for Freedom. E. Buhmann and W. Korcari are funded by the German Federal Ministry of Science and Research (BMBF) via *Verbundprojekts 05H2018 - R&D COMPUTING (Pilotmaßnahme ErUM-Data) Innovative Digitale Technologien für die Erforschung von Universum und Materie*. S. Diefenbacher is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2121 "Quantum Universe" – 390833306. E. Eren was funded through the Helmholtz Innovation Pool project AMALEA that provided a stimulating scientific environment for parts of the research done here. L. Rustige was supported by DESY and HamburgX grant LFF-HHX-03 to the Center for Data and Computing in Natural Sciences (CDCS) from the Hamburg Ministry of Science, Research, Equalities and Districts.

References

[1] R. Jansky, *The ATLAS Fast Monte Carlo Production Chain Project* (2015), J. Phys. Conf. Ser. **664**, 072024

[2] A. Butter, S. Diefenbacher, G. Kasieczka, B. Nachman, T. Plehn, *GANplifying Event Samples* (2020). 2008.06545

[3] I.J. Goodfellow et al., *Generative Adversarial Nets*, in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (Cambridge, MA, USA, 2014), NIPS'14, p. 2672–2680. 1406.2661, https://dl.acm.org/doi/10.5555/2969033.2969125

[4] D.P. Kingma, M. Welling, *Auto-Encoding Variational Bayes* (2014). 1312.6114

[5] C. Huang, D. Krueger, A. Lacoste, A.C. Courville, *Neural Autoregressive Flows* (2018), CoRR. 1804.00779

[6] M. Paganini, L. de Oliveira, B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters* (2018). 1705.02355

[7] L. de Oliveira, M. Paganini, B. Nachman, *Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis* (2017). 1701.05927

[8] M. Paganini, L. de Oliveira, B. Nachman, *CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks* (2018). 1712.10321

[9] M. Erdmann, L. Geiger, J. Glombitza, D. Schmidt, *Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks* (2018). 1802.03325

[10] M. Erdmann, J. Glombitza, T. Quast, *Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network* (2019). 1807.01954

[11] ATLAS Collaboration, Tech. Rep. ATL-SOFT-PUB-2018-001, CERN, Geneva (2018), http://cds.cern.ch/record/2630433

[12] ATLAS Collaboration, Tech. Rep. ATL-SOFT-SIM-2019-007, CERN (2019), https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2019-007/

[13] A. Ghosh (ATLAS Collaboration), Tech. Rep. ATL-SOFT-PROC-2019-007, CERN, Geneva (2019), https://cds.cern.ch/record/2680531

[14] D. Belayneh et al., *Calorimetry with Deep Learning: Particle Simulation and Reconstruction for Collider Physics* (2019). 1912.06794

[15] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, K. Krüger, *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed* (2021). 2005.05334

[16] V. Cédric, *Optimal Transport: Old and New* (Springer, Berlin, 2009)

[17] S. Voloshynovskiy, M. Kondah, S. Rezaeifar, O. Taran, T. Holotyak, D.J. Rezende, *Information bottleneck through variational glasses* (2019). 1912.00830

[18] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, K. Krüger, *Decoding Photons: Physics in the Latent Space of a BIB-AE Generative Network* (2021), submitted to vCHEP 2021. 2102.12491

[19] H. Abramowicz et al. (ILD Concept Group), *International Large Detector: Interim Design Report* (2020). 2003.01116

[20] M. Frank, F. Gaede, C. Grefe, P. Mato, *DD4hep: A Detector Description Toolkit for High Energy Physics Experiments* (2014), J. Phys. Conf. Ser. **513**, 022010

[21] A. Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library* (2019), Advances in Neural Information Processing Systems 32 pp. 8024–8035, http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[22] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, D. Whiteson, *Parameterized Machine Learning for High-Energy Physics* (2016), Eur. Phys. J. C **76**, 235. 1601.07913

[23] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, *Improved Training of Wasserstein GANs*, in *Advances in Neural Information Processing Systems 30* (2017), pp. 5767–5777. 1704.00028, http://papers.nips.cc/paper/7159-improved-training-of-wasserstein-gans.pdf

[24] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A.J. Smola, *A Kernel Method for the Two-Sample Problem* (2008), CoRR. 0805.2368

[25] E. Parzen, *On Estimation of a Probability Density Function and Mode* (1962), The Annals of Mathematical Statistics **33**, pp. 1065, http://www.jstor.org/stable/2237880

[26] S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen, D. Podareanu, R.R. de Austri, R. Verheyen, *Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer* (2019). 1901.00875

[27] Y. Wu, M. Rosca, T. Lillicrap, *Deep Compressed Sensing* (2019). 1905.06723

[28] Y. Wu, J. Donahue, D. Balduzzi, K. Simonyan, T. Lillicrap, *LOGAN: Latent Optimisation for Generative Adversarial Networks* (2020). 1912.00953

[29] S. Amari, *Natural Gradient Works Efficiently in Learning* (1998), Neural Computation **10**, 251

[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, *Improved Techniques for Training GANs* (2016). 1909.10578

[31] T. Karras, T. Aila, S. Laine, J. Lehtinen, *Progressive Growing of GANs for Improved Quality, Stability, and Variation* (2017). 1710.10196

[32] M. Thomson, *Particle flow calorimetry and the PandoraPFA algorithm* (2009), Nuclear Instruments and Methods in Physics Research Section A: Accel-erators, Spectrometers, Detectors and Associated Equipment **611**, 25–40, http://dx.doi.org/10.1016/j.nima.2009.09.009

[33] R. Kansal, J. Duarte, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.R. Vlimant, D. Gunopulos, *Graph Generative Adversarial Networks for Sparse Data Generation in*

High Energy Physics (2021). 2012.00173

[34] J.N. Howard, S. Mandt, D. Whiteson, Y. Yang, *Foundations of a Fast, Data-Driven, Machine-Learned Simulator* (2021). 2101.08944