Generative models for hadron shower simulation

Erik Buhmann, Sascha Diefenbacher, Daniel Hundhausen, Gregor Kasieczka, William Korcari
Institut für Experimentalphysik, Universität Hamburg
Luruper Chaussee 149, 22761 Hamburg, Germany
sascha.daniel.diefenbacher@uni-hamburg.de

Engin Eren, Frank Gaede, Katja Krüger, Peter McKeown
Deutsches Elektronensynchrotron
Notkestrasse 85, 22607 Hamburg, Germany
engin.eren@desy.de

Lennart Rustige
CDCS, Universität Hamburg
Notkestrasse 11, 22607 Hamburg, Germany

Abstract

Simulations provide the crucial link between theoretical descriptions and experimental observations in the physical sciences. In experimental particle physics, a complex ecosystem of tools exists to describe fundamental processes or the interactions of particles with detectors. The high computational cost associated with producing precise simulations in sufficient quantities — e.g. for the upcoming data-taking phase of the Large Hadron Collider (LHC) or future colliders — motivates research into more computationally efficient solutions. Using generative machine learning models to amplify the statistics of a given dataset is an especially promising direction. However, the simulation of realistic showers in a highly granular detector remains a daunting problem due to the large number of cells, values spanning many orders of magnitude, and the overall sparsity of data. This contribution advances the state of the art in two key directions: Firstly, we present a precise generative model for the fast simulation of hadronic showers in a highly granular hadronic calorimeter. Secondly, we compare the achieved simulation quality before and after interfacing with a so-called particle-flow-based reconstruction algorithm. Together, these bring generative models one step closer to practical applications.

1 Introduction

Particle physics investigates the laws of nature at length scales of 10^{-18} meters. Collider experiments accelerate beams of particles (e.g., protons or electrons) close to the speed of light and bring them into collision. The resulting interactions are recorded by complex and highly granular detectors and can be described by the so-called Standard Model (SM). Highly accurate simulations of physics processes and particle-detector interactions are needed to measure the properties of the SM and look for potential deviations from it. These simulations are traditionally generated using Monte Carlo methods. They are significant consumers of computing resources in particle physics, as billions of examples need to be simulated to match the data produced by experiments.

Generative machine learning models offer a promising way to amplify available statistics [1, 2]. Methods using Generative Adversarial Network (GANs) [3], Variational Autoencoders (VAE) [4],

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

and autoregressive flows [5] have been investigated for different aspects of this challenge such as event generation [6–9], parton showering [10–12] and detector simulation [13–21].

This contribution shows progress on a particular simulation challenge: particle showers caused by hadrons in a highly-granular hadronic calorimeter. Due to the richer observed structure compared to particles which interact purely electromagnetically, these offer a larger challenge for generative models. Furthermore, we interface the showers generated by two different machine learning models to the standard reconstruction software and obtain a more realistic estimate of the quality of generated showers. The remainder of this work is structured as follows: Sec. 2 introduces the physics challenges and training datasets; Sec. 3 discusses the used generative architectures; Sec. 4 presents the obtained results; and Sec. 5 summarises our findings and shows an outlook on future work.

2 Data sets

When a highly energetic particle hits a block of heavy detector material it interacts with this material, creating a cascade of secondary particles. These secondary particles will themselves interact with the detector and create further particles. This avalanche continues until all participating particles have lost their energy. The chain of particle interactions is called a shower. In high energy physics we record these showers using calorimeters. Current state-of-the-art calorimeters consist of a sandwich of passive absorber and active sensor. These sensor layers are themselves made up of highly granular pixel detectors. During a shower the dense absorbers mediate most of the interactions taking place, while the sensors record highly resolved slices of the developing shower structure. The resulting measurements consist of the energy values deposited in these individual pixel detectors. As each pixel has a fixed position in the calorimeter we can project these pixel values into a regular 3D grid, called a calorimeter image in the following.

While previous work successfully applied generative models to photon showers [20], here we focus on so-called pions. Compared to photon showers, pion showers feature a greater range of possible interactions, resulting in an increased complexity of the shower structure. Specifically, we simulate the *Analogue Hadron Calorimeter* (AHCal) of the proposed International Linear Detector (ILD) [22] prototype as this is where the majority of the pion interactions will take place. Within the open source framework iLCSoft [23] we simulate 500k pion showers with energies uniformly distributed between 10 and 100 GeV. This simulation is performed using GEANT4 [24], the state-of-the-art software package for shower simulations. The resulting showers are projected onto a $25 \times 25 \times 48$ grid, while simultaneously being corrected for any potential artefacts arising form this projection.

A sample of the dataset containing 5k showers is provided at `https://doi.org/10.5281/zenodo.5529677`.

3 Generative Models

Two generative models are trained on the same data: a Wasserstein-GAN (WGAN) [25] and a Bounded Information Bottleneck Autoencoder (BIB-AE [26]) architecture, both based on the architectures proposed in [20].

For the WGAN architecture, convolutional layers previously used in the critic network are replaced by 3D-residual blocks [27], and a fully-connected network is used (instead of convolutional layers) in the energy constrainer.

Several changes are applied to the BIB-AE to improve its generative performance:

- **Minibatch Discrimination and Resetting Critics** In addition to the individual samples we also pass information about the makeup of individual batches to the critics [28]. This reduces overfitting and teaches the network global features of the data. Further, critics can become *blind* to certain features during training leading to artefacts in the generated data. Therefore all BIB-AE critics are actually a set of two networks, one trained continuously and one that is reset every epoch.
- **Accurate Latent Space Sampling** Ideally, the latent space contains sufficient information for reconstruction but is regular enough for sampling. In a BIB-AE setup, however, the adversarial reconstruction makes balancing regularization and reconstruction problematic.

Therefore, a different latent sampling approach based on a Buffer-VAE [7, 29] is employed. We encode our training data set into latent space examples. From this we can draw new samples using a Kernel Density Estimator (KDE) [30].

- **Improved Post Processing** A Post Processor network is used to fine-tune the generated per-cell energy distribution. The Post Processor is now trained on a fixed BIB-AE model, instead of in-parallel, and several additional loss terms are added.

Both models were implemented using PyTorch [31]. The full code for both models alongside the hyperparamter settings used in the training can be found on `https://github.com/FLC-QU-hep/neurIPS2021_hadron`

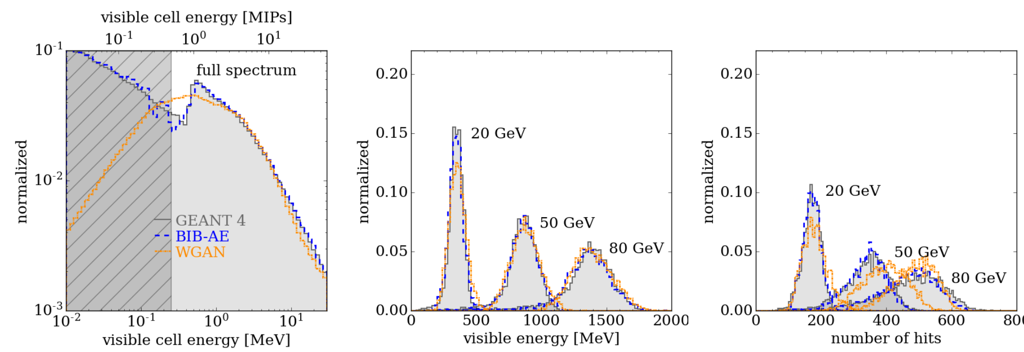

Figure 1: Differential distributions comparing physics quantities between GEANT4 (ground truth) and the different generative models at generator level. The energy per-cell is measured in MeV for the bottom axis and in multiples of the expected energy deposit of a minimum ionizing particle (MIP) for the top axis.

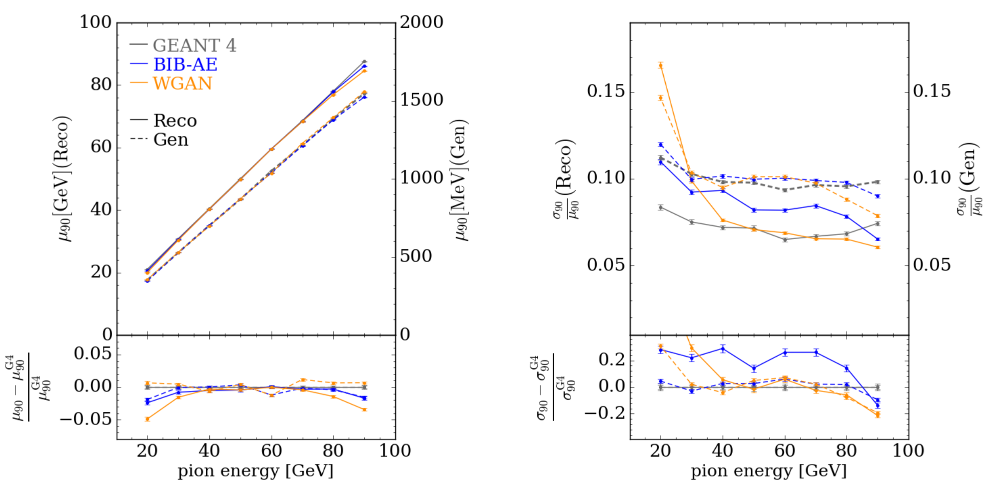

Figure 2: Mean (μ_{90}, left) and relative width (σ_{90}/μ_{90}, right) at the generator and reconstruction level for pions with various incident energies. In order to avoid edge effects, the phase space boundary regions of 10 and 100 GeV are removed for the response and resolution studies. In the bottom panels, the relative offset of these quantities with respect to the Geant4 simulation is shown.

4 Results

When applying generative networks to particle physics simulations, a correct description of differential distributions is needed in addition to accurate individual images. This can be done at two different stages: by directly investigating properties of showers coming form the generators (Generator-level) or after processing by a dedicated reconstruction software (Reconstruction level).

In Fig. 1 we show three of the relevant distributions at generator level. The first plot shows the energy contained in a single sensor (visible cell energy, left). In order to reduce electronic noise we only consider cell-energies above half the energy deposited by a minimal ionizing particle (MIP). This is indicated by the shaded area in the plot. For high energies both models match this distribution nicely. In addition, the BIB-AE also manages to capture the feature around 1 MIP, thanks to the post processing. The second plot shows the total energy sum over all pixels in a shower (center). This is accurately captured by both models. The final plot shows the distribution of the number of non-zero pixels (right). For low pion energies, both models describe this property well, however for higher pion energies the WGAN produces slightly too many hits.

Calorimeters at future $e^+ \, e^-$ colliders provide unprecedented details of particle interactions. The state-of-the-art pattern recognition algorithm used by ILD is PandoraPFA [32]. It aims at reconstructing all individual particles created in the event by exploiting the high granularity of the calorimeters such as the AHCal. The output of this reconstruction algorithm is directly used in all physics analyses. The accurate description of the distribution of visible and reconstructed energy for a given true incident pion energy is therefore of great importance.

For evaluation, we use samples of pion showers at discrete energies ranging from 20 to 90 GeV in 10 GeV steps, simulated with GEANT4 and generated with our models. For these sets of showers we calculate the mean and root-mean-square of the central 90% of the distributions, labeled μ_{90} and σ_{90} respectively. The results are shown in Fig. 2 as a function of the incident pion energy. For both models the mean (left) is correctly modelled up to five-percent deviations w.r.t GEANT4 at the reconstruction and generation level. Deviations in the relative resolution (σ_{90} / μ_{90}) are more pronounced for both models at the different levels. Note that a calibration factor has been applied to the WGAN-generated single-energy showers to improve the linearity.

The main motivation for using generative models in particle physics is to reduce the time and cost per simulated sample. Table 1 shows the time to generate a single shower using GEANT4, the WGAN and the BIB-AE. Both models offer significant speedups compared to classical generation methods. Furthermore we also see trade-offs between the models illustrated. While the BIB-AE produces overall better quality showers than the WGAN, it also offers one order of magnitude less speedup.

The BIB-AE and PostProcessor model were trained using four parallel NVIDIA® V100 GPUs for a total time of roughly 10 days. The WGAN was trained on two V100 GPUs for 13 days.

Table 1: Computational performance of WGAN and BIB-AE generators on a single core of an Intel® Xeon® CPU E5-2640 v4 (CPU) and NVIDIA® A100 with 40 GB of memory (GPU) compared to GEANT4. For the generative models, the best performing batch size is shown and given by the mean and standard deviation obtained for sets of 10000 showers.

Hardware	Simulator	Time / Shower [ms]	Speed-up
CPU	GEANT4	2684 ± 125	$\times 1$
	WGAN	47.923 ± 0.089	$\times 56$
	BIB-AE	350.824 ± 0.574	$\times 8$
GPU	WGAN	0.264 ± 0.002	$\times 10167$
	BIB-AE	2.051 ± 0.005	$\times 1309$

5 Conclusions and Outlook

Fundamental physics is facing growing difficulties from the expansion of computing resources using slow Monte-Carlo-based simulations. While these simulations encode valuable physical knowledge and are challenging to replace, better use of generated statistics is possible by using generative models. In this contribution, we show the performance of a Wasserstein-GAN and BIB-AE architecture on the challenging task of simulating hadronic showers in a highly granular calorimeter. We observe accurate modeling of physically relevant quantities over several orders of magnitude and a speed-up over three orders of magnitude. Furthermore, we also find a good description of key shower properties after processing the generated samples with a standard reconstruction software.

Acknowledgments and Disclosure of Funding

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

[2] M. Paganini, L. de Oliveira, B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters*, Phys. Rev. Lett. **120**, 042003 (2018). 1705.02355

[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*, CoRR (2018). 1804.00779

[6] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat, S. Palazzo, *DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC* (2019). 1903.02433

[7] 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

[8] A. Butter, T. Plehn, R. Winterhalder, *How to GAN LHC Events*, SciPost Phys. **7**, 075 (2019). 1907.03764

[9] Y. Alanazi, N. Sato, T. Liu, W. Melnitchouk, M.P. Kuchera, E. Pritchard, M. Robertson, R. Strauss, L. Velasco, Y. Li, *Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN)* (2020). 2001.11103

[10] E. Bothmann, L. Debbio, *Reweighting a parton shower using a neural network: the final-state case*, JHEP **01**, 033 (2019). 1808.07802

[11] L. de Oliveira, M. Paganini, B. Nachman, *Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis*, Comput. Softw. Big Sci. **1**, 4 (2017). 1701.05927

[12] J.W. Monk, *Deep Learning as a Parton Shower*, JHEP **12**, 021 (2018). 1807.03685

[13] M. Paganini, L. de Oliveira, B. Nachman, *CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks*, Phys. Rev. **D97**, 014021 (2018). 1712.10321

[14] M. Erdmann, L. Geiger, J. Glombitza, D. Schmidt, *Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks*, Comput. Softw. Big Sci. **2**, 4 (2018). 1802.03325

[15] M. Erdmann, J. Glombitza, T. Quast, *Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network*, Comput. Softw. Big Sci. **3**, 4 (2019). 1807.01954

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

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

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

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

[20] 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

[21] G.R. Khattak, S. Vallecorsa, F. Carminati, G.M. Khan, *Fast simulation of a high granularity calorimeter by generative adversarial networks* (2021). 2109.07388

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

[23] *iLCSoft Project Page*, https://github.com/iLCSoft (2016)

[24] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003)

[25] 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

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

[27] K. Hara, H. Kataoka, Y. Satoh, *Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, Los Alamitos, CA, USA, 2018), pp. 6546–6555, https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00685

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

[29] 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

[30] 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

[31] A. Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, Advances in Neural Information Processing Systems 32 pp. 8024–8035 (2019), Advances in Neural Information Processing Systems 32 pp. 8024–8035

[32] J.S. Marshall, M.A. Thomson, *The Pandora software development kit for pattern recognition* (2015), http://dx.doi.org/10.1140/epjc/s10052-015-3659-3

Checklist

1. For all authors...
    (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
    (b) Did you describe the limitations of your work? [Yes] Limitations of our work are described in section 4.
    (c) Did you discuss any potential negative societal impacts of your work? [N/A] As this work is directly tied to fundamental physics we felt there to be no directly measurable societal impact.
    (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
    (a) Did you state the full set of assumptions of all theoretical results? [N/A] This work contains no theoretical results.
    (b) Did you include complete proofs of all theoretical results? [N/A] This work contains no theoretical results.
3. If you ran experiments...
    (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Link to the code is provided in Section 3. Link to sample data is provided in Section 2. Providing the full set was not possible due to its size.
    (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] The link to the code with the used hyperparameters is provided in Section 3.
    (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] Because of the extensive training times we were unable to run the experiments multiple times. We did however include statistical error bars for the training and generated data set where applicable.
    (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Compute times are specified in Section 4.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
    (a) If your work uses existing assets, did you cite the creators? [Yes] Used software packages (PyTorch, ilcsoft) are cited the first time they are mentioned.
    (b) Did you mention the license of the assets? [Yes] We mentioned a general open source (there are several different licences in ilcsoft).
    (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] URL to a data sample is given in section 2, providing the full set was not possible due to its size.
    (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] We created our own dataset.
    (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] This was not a concern for high energy physics data.
5. If you used crowdsourcing or conducted research with human subjects...
    (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] This work contains no crowdsourced data.
    (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] This work contains no crowdsourced data.
    (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] This work contains no crowdsourced data.