PROCEEDINGS OF SCIENCE

Fast Simulation of Highly Granular Calorimeters with Generative Models: Towards a First Physics Application

Erik Buhmann,[a] Thorsten Buss,[a] Sascha Diefenbacher,[a,b] Engin Eren,[c] Frank Gaede,[c,d] Gregor Kasieczka,[a,d] William Korcari,[a] Anatolii Korol,[c] Katja Krüger,[c] Thomas Madlener,[c] Peter McKeown[c,*] and Lennart Rustige[c,d]

[a] *Institut für Experimentalphysik, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany*

[b] *Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA*

[c] *Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*

[d] *Center for Data and Computing in Natural Sciences CDCS, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*

E-mail: peter.mckeown@desy.de

While simulation plays a crucial role in high energy physics, it also consumes a significant fraction of the available computational resources, with these computing pressures being set to increase drastically for the upcoming high luminosity phase of the LHC and for future colliders. At the same time, the significantly higher granularity present in future detectors increases the physical accuracy required of a surrogate simulator. Machine learning methods based on deep generative models hold promise to provide a computationally efficient solution, while retaining a high degree of physical fidelity. Significant strides have already been taken towards developing these models for the generation of particle showers in highly granular calorimeters, the subdetector which constitutes the most computationally intensive part of a detector simulation. However, to apply these models to a general detector simulation, methods must be developed to cope with particles incident at various points and under varying angles in the detector. This contribution will address steps taken to tackle the challenges faced when applying these simulators in more general scenarios, as well as the effects on physics observables after interfacing with reconstruction algorithms. In particular, results achieved with bounded information bottleneck and normalising flow architectures based on regular grid geometries, as well as a more flexible diffusion model using point clouds, will be discussed. Combined with progress on integrating these surrogate simulators into existing full simulation chains, these developments bring an application to benchmark physics analyses closer.

The European Physical Society Conference on High Energy Physics (EPS-HEP2023)
21-25 August 2023
Hamburg, Germany

*Speaker

PoS(EPS-HEP2023)568

1. Introduction

The use of Monte-Carlo simulation toolkits such as Geant4 [1] provides an accurate means of detector simulation for high energy physics experiments, but will be unsustainable in the future given the expected computing budgets. Generative models have the potential to provide a fast and accurate surrogate simulation tool.

In this work, we review recent advances on three different frontiers in the development of generative models for fast simulation of showers in highly granular calorimeters. In Section 2, we address extensions made to an autoencoder architecture for use in generalised simulation scenarios, as well as the effects of interfacing with state-of-the-art reconstruction algorithms. In Section 3, we describe developments in architectures based on normalising flows that are able provide state-of-the-art generative fidelity. In Section 4 we provide an overview of a more flexible and efficient diffusion model that is able to generate calorimeter showers as a point cloud while achieving a high degree of geometry independence. Finally, in Section 5 we connect these three frontiers back to a final physics application by describing a C++ library that allows the integration of fast simulation tools based on generative models into the existing software ecosystems.

2. Generative Models in Generalised Simulation Scenarios

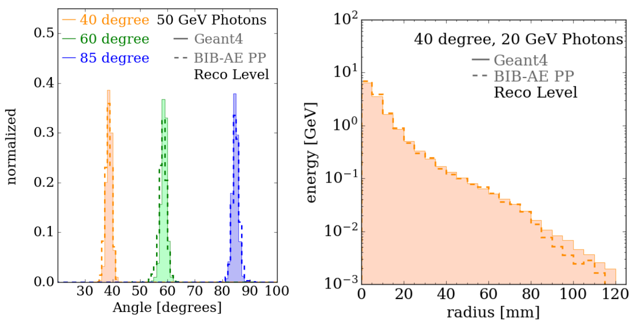

Figure 1: Exemplary distributions for Geant4 (filled histograms) and BIB-AE (dashed, unfilled histograms) after reconstruction. Left: Angular response for 50 GeV photons at various incident angles. Right: Radial profile for 20 GeV photons incident at 40 degrees. 1900 photons were used for each simulation method and for each energy and angle combination. Figures from Ref. [2]

The Bounded Information Bottleneck Autoencoder (BIB-AE) was initially applied to the problem of photon shower simulation in highly granular calorimeters in [3], and subsequently to the more challenging case of charged pion simulation in [4]. While the model demonstrated a strong performance in terms of physical fidelity for various incident energies, applicability in a generalised simulation scenario would require the model to provide the correct detector response for particles also incident under different angles and at different positions.

This conditioning regime was extended in [5] [2], where the polar angle of photons incident to the calorimeter was varied uniformly in the range 30 – 90 degrees simultaneously with the energy, which was varied in the range 10 – 100 GeV. The incident position on the calorimeter face was fixed. The BIB-AE model consisted of a mixture of 3D convolutional and fully connected layers distributed across of a number of sub-components, including an encoder-decoder pair, a number of Wasserstein-GAN-like critic networks and a post processing network targeting an improved modelling of the cell energy spectrum. Additionally, a normalising flow architecture was used for latent space sampling during inference, without which generation with the extended conditioning phase space was prohibitively slow. The performance of the model was benchmarked in comparison to Geant4 across a number of calorimetric observables. An additional comparison was made after applying calorimeter hit clustering via the state-of-the-art particle flow reconstruction algorithm PandoraPFA [6], where the BIB-AE model was shown to maintain a strong physics performance. Exemplary distributions showing the angular response for 50 GeV photons and radial profile for 20 GeV, 40 degrees photons are shown in Figure 1. This is a key step towards applying the model in a simulation framework, where physics performance after reconstruction is the ultimate target. Additional conditioning of the model on the second (azimuthal) angle is currently under study.

3. Achieving Higher Fidelity with Normalising Flows

Normalising Flow (NF) based architectures have previously been shown to achieve a high degree of fidelity in the case of the generation of calorimeter showers with relatively low-dimensionality ($d \approx 500$) [7] [8]. It has since been demonstrated that an NF-based 'Layer-to-Layer-Flows' (L2LFlows) model, operating on images of the core of photon showers with dimensionality $30 \times 10 \times 10$, was similarly able to achieve a high degree of physical fidelity. In particular, superior simulation performance in comparison to a BIB-AE architecture was demonstrated across a range of physics observables [9]. This model has since been adapted to be able to handle full showers with dimensionality $30 \times 30 \times 30$. This was achieved by using coupling layers together with convolutional elements in the NF to achieve better scaling with dimensions and significantly improved simulation speed. The simulation fidelity remains higher than that of a BIB-AE network used for comparison. One such physical observable is the voxel energy distribution, which is shown for Geant4, the BIB-AE network used for comparison and the improved NF model in Figure 2. In the future, work on this architecture will focus on attempting to further improve the simulation speed of the model to be competitive with existing architectures.

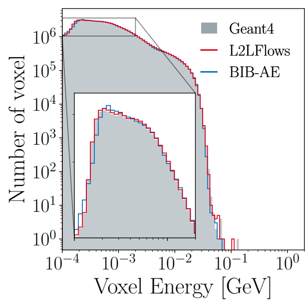

Figure 2: Distribution of voxel energies for showers produced by Geant4, a BIB-AE and the improved Normalising Flow model for 100k photons in each case, and with incident energies distribtued in the range 10 – 100 GeV.

4. Flexible Data Representations with Point Clouds

The models described in Sections 2 and 3, along with the majority of models developed for calorimeter simulation, rely on the use of regular grids. Due to the high degree of sparsity present in the data, this means a large amount of computation is wasted on empty cells. Additionally, in detectors with irregular geometries the projection to a regular grid tends to create artefacts. The irregular nature of the geometry means that the local geometry appears different at different positions in the detector. The CaloClouds model introduced in [10] was the first model to successfully generate calorimeter showers as a point cloud with high-dimensionality ($\mathcal{O}(1000)$ points). A crucial step for handling irregular geometries with this model was to use clustered Geant4 steps, essentially allowing the model to operate on a point cloud with much higher granularity than that present in the physical readout geometry.

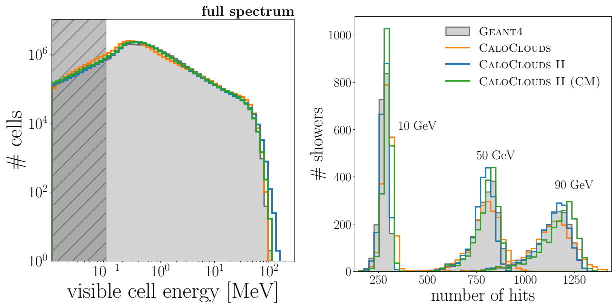

Figure 3: Left: Cell energy spectrum, with 40k photon showers in each distribution. Right: Number of hits above threshold for incident photon energies of 10, 50 and 90 GeV, with 2000 showers in each distribution. The Geant4 distribution is shown by the filled histogram, while those for CaloCLouds, CaloClouds II and its distilled consistency model CaloClouds II (CM) are shown by the unfilled, coloured histograms. Figures adapted from Ref. [11].

This enabled the model to achieve a high degree of geometry independence. The training data consisted of photon showers simulated using Geant4, with energies uniformly varying in the range 10 – 90 GeV. The model itself is a diffusion model with a number of different sub-components, including an NF for conditioning and calibration. The model has since been significantly refined with the development of CaloClouds II [11], which uses a consistency model to enable single shot generation without reducing the simulation fidelity. Notably, the model was able to simulate photon showers 46× faster than Geant4 on a single CPU core.

5. Integration into Software Ecosystems

For a final benchmarking and deployment of generative models for fast calorimeter simulation, it is essential that they be incorporated into the existing simulation chains. It is possible to interface a fast simulation model with Geant4 via hooks that allow the termination of detailed physics-based simulation once a particle enters a designated volume and fulfils the assigned criteria (particle type, energy etc.). An appropriate fast simulation method can then be used to emulate the detector response, and the output handed back to Geant4 for placement back into the geometry. These fast simulation hooks can additionally be interfaced through DD4hep [12], the geometry description toolkit used by ILD [13] and numerous other detector concepts and experiments.

A prototype C++ library called *DDFastShowerML* has been developed to allow the usage of generative models for fast simulation of calorimeter subsystems of detectors implemented in DD4hep. At the core of this library lies a class template consisting of five key components that are separated as far as possible:

- *Model*: A model specific implementation for a given architecture. This handles the preparation of the input in the correct form for the model, as well as the interpretation of the model output such that it can be converted into space points.
- *Inference*: The implementation of actually calling the inference library for the model. Currently supported libraries include LibTorch and OnnxRuntime.
- *Geometry*: The implementation of the detector geometry that allows the placement of space points produced by the model. This includes for example the positioning of calorimeter layers, and the conversion between global (envelope) and local coordinates. A local direction with respect to the calorimeter face is also computed, to provide a consistent basis for model conditioning.

- *Hit Maker*: A helper class provided by GEANT4 for placement of energy depositions produced as space points, given that their position lies within a sensitive region of the detector.
- *Trigger*: Sitting on top of the particle type and energy triggers that already exist in GEANT4 and DD4HEP, this trigger interface allows for the definition of more conditions for running the fast shower model, e.g. the exclusion of certain regions of a given geometry from fast simulation. This includes for example the intersection between modules in a polyhedral barrel or the transition between sub-detector elements, where full simulation is called instead.

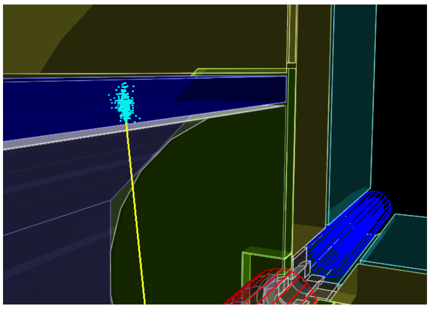

Figure 4: 20 GeV photon shower simulated in ILD with a BIB-AE.

The BIB-AE model described in section 2 currently has the most complete implementation in *DDFastShowerML* — a 20 GeV photon shower in the barrel of the ILD detector that has been simulated with the BIB-AE is shown in Figure 4. In the future, the integration of the other models described in Sections 3 and 4 is planned. Currently, the library only supports single shower simulation with a generative model (i.e. inference with a batch size of one) on a CPU. Development of the library in the future will seek to extend the functionality of the library in two key directions. Firstly, batching of shower generation within an event will be targeted. This will allow the number of sequential passes through the model to be reduced, and hence improve the simulation speed of the generative models. Secondly, support of GPUs for model inference is planned, which when combined with the aforementioned batching is expected to provide significant speed ups via parallelisation.

6. Conclusions

This work has demonstrated progress on the development of generative models for fast calorimeter simulation in three key directions: applications in generalised simulation scenarios; improved physics fidelity and simulation with more flexible and efficient data representations. An ultimate target is to develop a simulator which combines all three. With the prototype *DDFastShowerML* library it is now possible to run full event simulations in DD4HEP geometries, with a user-controlled seamless transition between full GEANT4 and generative model based fast simulation. The library is sufficiently flexible so as to make the integration of different types of architectures as straight forward as possible. These developments constitute major advances towards a final physics application of generative models for fast calorimeter shower simulation.

Acknowledgments

We thank Jenny List for valuable comments on the manuscript. This research was supported in part by the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 101004761. We acknowledge support by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC

2121 Quantum Universe – 390833306 and via the KISS consortium (05D23GU4, 13D22CH5) funded by the German Federal Ministry of Education and Research BMBF in the ErUM-Data action plan. E.B. is partially funded by a scholarship from the Friedrich Naumann Foundation for Freedom. E.B. and W.K. acknowledge funding 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*. A.K. has received support from the Helmholtz Initiative and Networking Fund's initiative for refugees as a refugee of the war in Ukraine.

References

[1] S. Agostinelli et al., *Geant4—a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 .

[2] S. Diefenbacher et al., *New Angles on Fast Calorimeter Shower Simulation*, *Mach. Learn. Sci. Tech.* **4** (2023) 035044 [2303.18150].

[3] E. Buhmann et al., *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed*, *Comput. Softw. Big Sci.* **5** (2021) 13 [2005.05334].

[4] E. Buhmann et al., *Hadrons, Better, Faster, Stronger*, *Mach. Learn. Sci. Tech.* **3** (2022) 025014 [2112.09709].

[5] S. Bieringer et al., *Generative Models for Fast Simulation of Electromagnetic and Hadronic Showers in Highly Granular Calorimeters*, *PoS* **ICHEP2022** (2022) 236.

[6] J.S. Marshall and M.A. Thomson, *The Pandora Software Development Kit for Pattern Recognition*, *Eur. Phys. J. C* **75** (2015) .

[7] C. Krause and D. Shih, *CaloFlow: Fast and Accurate Generation of Calorimeter Showers with Normalizing Flows*, 2106.05285.

[8] C. Krause and D. Shih, *CaloFlow II: Even Faster and Still Accurate Generation of Calorimeter Showers with Normalizing Flows*, 2110.11377.

[9] S. Diefenbacher et al., *L2LFlows: Generating High-Fidelity 3D Calorimeter Images*, 2302.11594.

[10] E. Buhmann et al., *CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation*, 2305.04847.

[11] E. Buhmann et al., *CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation*, 2309.05704.

[12] M. Frank et al., *DD4hep: A Detector Description Toolkit for High Energy Physics Experiments*, *J. Phys. Conf. Ser.* **513** (2014) 022010.

[13] ILD collaboration, *International Large Detector: Interim Design Report*, 2003.01116.