

# Simulation of Hadronic Interactions with Deep Generative Models

*Tuan Minh* Pham<sup>1,\*</sup> and *Xiangyang* Ju<sup>2,\*\*</sup>

<sup>1</sup>Physics Department, University of Wisconsin-Madison, Madison, WI 53706

<sup>2</sup>Scientific Data Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

**Abstract.** Accurate simulation of detector responses to hadrons is paramount for all physics programs at the Large Hadron Collider (LHC). Central to this simulation is the modeling of hadronic interactions. Unfortunately, the absence of first-principle theoretical guidance has made this a formidable challenge. The state-of-the-art simulation tool, GEANT4, currently relies on phenomenology-inspired parametric models. Each model is designed to simulate hadronic interactions within specific energy ranges and for particular types of hadrons. Despite dedicated tuning efforts, these models sometimes fail to describe the data in certain physics processes accurately. Furthermore, fine-tuning these models with new measurements is laborious. Our research endeavors to leverage generative models to simulate hadronic interactions. While our ultimate goal is to train a generative model using experimental data, we have taken a crucial step by training conditional normalizing flow models with GEANT4 simulation data. Our work marks a significant stride toward developing a fully differentiable and data-driven model for hadronic interactions in High Energy and Nuclear Physics.

## 1 Introduction

Hadronic interactions exhibit a well-defined description in high-energy domains, typically in the hundreds of GeV range, where perturbative Quantum Chromodynamics (QCD) proves effective. However, in lower energy regimes, the perturbative theory loses its applicability. To address this challenge, various phenomenological models have been devised to characterize hadronic interactions. These models are designed for specific kinematic regions and are tailored to a limited number of hadron flavors. To provide a comprehensive description across a wide energy spectrum, spanning from MeV to hundreds of GeV, these models must be integrated into a transport code. This integration is precisely achieved in the GEANT4 framework [1]. It serves as a unified platform where these diverse models are combined to simulate hadronic interactions across an extensive energy range.

The GEANT4 framework plays a vital role in simulating detector effects for a broad spectrum of applications, including the Large Hadron Collider experiments and

---

\*e-mail: tuan.minh.pham@cern.ch

\*\*e-mail: xju@lbl.gov

beyond. The GEANT4 simulation, often referred to as “full simulation”, constitutes the most computationally intensive aspect of collider physics programs. As detectors are progressively enhanced to achieve finer resolution, the computational demands of full simulation are poised to escalate significantly.

To address this challenge, a concerted effort is emerged towards the development of fast simulation techniques employing deep generative models [2–10], despite that much improvement has been made within the GEANT4 framework [11]. Recent studies have shown that Normalizing Flow [12] can achieve state-of-the-art precision while dramatically reducing generation time, yielding computational gains of orders of magnitude compared to full simulation [13, 14]. However, these generative models are tailored to high-level detector-specific features, making them unsuitable for generalized use across different particle detectors. Our ultimate objective is to develop a generative model that learns the non-perturbative hadronic interactions from the experimental data, such as those published in Refs. [15, 16].

## 2 Methods

### 2.1 Dataset

The data is generated by the GEANT4 [1] toolkit with a physics list that includes BERTINI Cascade and FRITIOF model (a.k.a FTFP\_BERT\_ATL). The BERTINI model is applicable for incident energies below 10 GeV in most use cases, while the FRITIOF model is for incident energies from 9 GeV to 100 TeV. For simplicity, we use a uniform energy transition for the FTFP\_BERT\_ATL physics list.

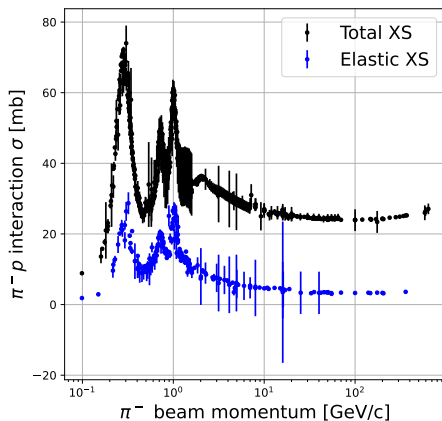
We chose the  $\pi^-p$  interactions for our studies, whose total cross sections are measured across a large range of kinetic ranges as shown in Fig. 1. There are peaks in the cross sections connected with the  $\Delta$ -isobar production in the  $s$ -channel,  $\pi^- + p \rightarrow \Delta^0$ . The main decay mode of a  $\Delta^0$  is  $\Delta^0 \rightarrow \pi^- + p$ . Thus, we define a mono-material detector that is made of hydrogen, shot pion beams to the material, and focus on simulating events that result in two outgoing hadrons.

To span a broad spectrum of incident  $\pi^-$  energies, we selected 29 data points between 100 MeV and 8000 MeV, spaced approximately 200 MeV apart for training. Additionally, we employed 14 distinct energy levels within the range of 200 MeV to 6500 MeV for testing. For each energy point, we generated 100,000 events with randomly sampled momentum directions. In summary, we produced 290,000 training events and 140,000 testing events.

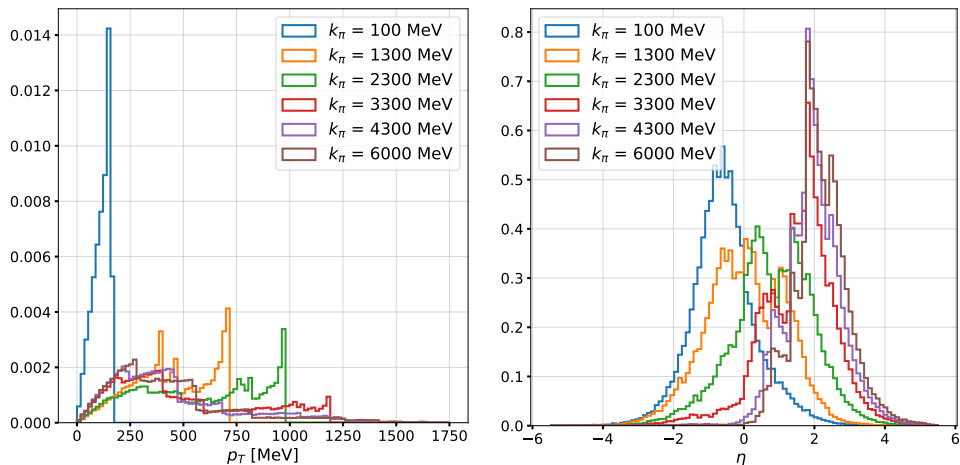
At the center-of-mass (COM) frame, the two outgoing particles are produced back-to-back, distributed uniformly across the azimuthal angle. Thanks to energy conservation, our generative model only needs to predict the properties of one of these particles. Figure 2 shows the transverse momentum and pseudorapidity of the leading particle in events featuring two outgoing particles. The kinetic energies of the leading particles change dramatically when the incident kinetic energy changes, attributed to multiple physics processes. These dependencies gradually smooth out when the pion energy exceeds 10 GeV, at which point only the FRITIOF model is employed. Accurately modeling the correlations between the leading particle energy and the incident kinetic energy poses a formidable challenge.

### 2.2 Normalizing Flow and Training

A normalizing flow (NF) uses an invertible function  $f$  (also known as a bijector) to transform a simple initial density  $\pi(\vec{z})$  to the target density distribution  $p(\vec{x})$ , and

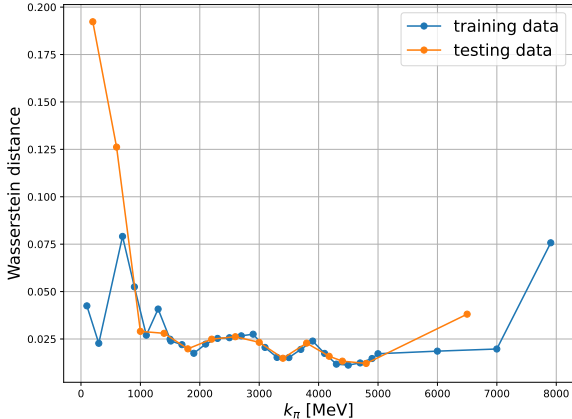


**Figure 1.** Total and elastic cross section of  $\pi^- p$  interactions as a function of incident  $\pi^-$  kinematic energy in the lab frame, taken from PDG data-base [17].



**Figure 2.** Comparison of the transverse momentum  $p_T$  and pseudorapidity  $\eta$  distributions of the leading final state particle for different kinematic energy of the incoming  $\pi^-$ . Final state particles are sorted by their  $p_T$ . Only the Bertini regions are shown.

the autoregressive density estimator models any joint density  $p(\vec{x})$  as a product of one-dimensional conditional distributions. We use a simple invertible function  $f$  in the MAF:  $x_i = f(z_i) = z_i \exp(\alpha_i) + \beta_i$  where  $\alpha_i$  and  $\beta_i$  are parameterized by neural networks, often by the MultiLayer Perceptrons (MLPs). The Adam optimizer then optimizes the learnable weights in the neural network by minimizing the negative log-likelihood function. A normalizing flow can be extended to a conditional normalizing flow by concatenating the conditional vector  $\vec{c}$  with the input vector  $\vec{x}$  and using the combined vector to estimate the target density distribution.



**Figure 3.** Wasserstein distance between the NF-generated events and the GEANT4-simulated events for different pion energies for the training and testing dataset.

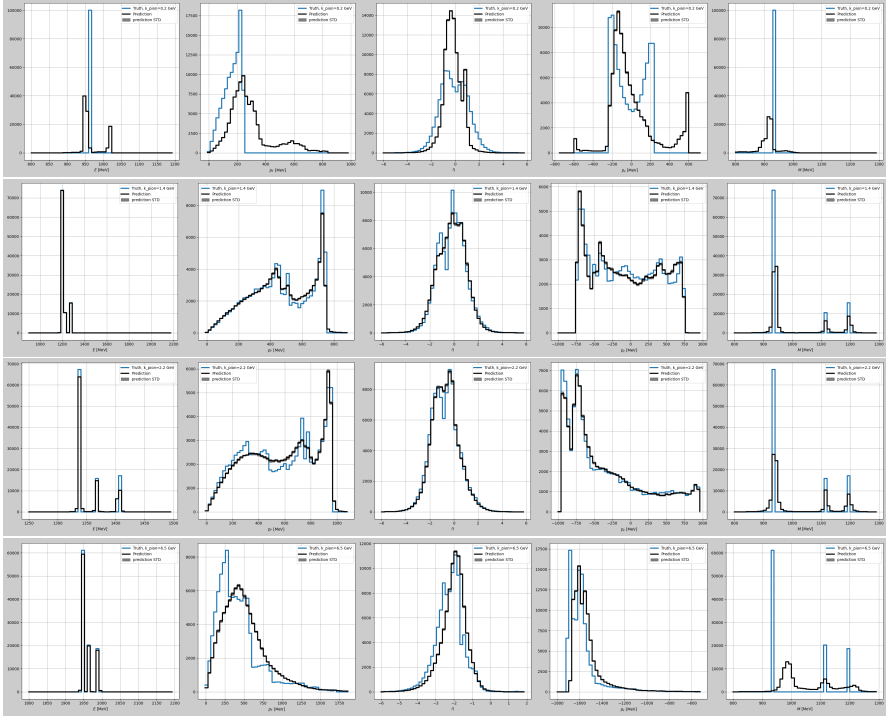
Our study employs a specific variant of normalizing flows called Masked Autoregressive Flow (MAF) [18]. MAF constructs multi-dimensional distributions by sequentially modeling each dimension based on previously modeled ones. While this autoregressive approach enhances modeling accuracy, it can introduce sampling bottlenecks and sensitivity to the input vector’s order. To mitigate the ordering impact, we introduced a permutation bijection to each MAF block. Additionally, in the final block, we appended a tanh bijector layer to ensure that the outputs fall within the target distribution’s range, namely  $[-1, 1]$ .

Our final normalizing flow model uses the incident pion energy as the conditional variable and employs Gaussian distributions as the base distributions. Its target distributions are the four-vector of leading outgoing particles ( $p_x$ ,  $p_y$ ,  $p_z$ , and  $E$ ). The normalizing flow consists of 30 Masked Autoregressive Flow (MAF) blocks, with each MAF block using two-layer MLPs with a layer size of 128. The implementation is based on TENSORFLOW [19] (TF) and TF PROBABILITY. The model undergoes training for over 2000 epochs with a learning rate that decreases from  $10^{-4}$  to  $10^{-6}$  following a polynomial function.

### 3 Results

We evaluate the performance of the trained normalizing flows using the "Wasserstein distance" (WD) [20, 21]. This metric is calculated for each variable, comparing the NF-generated events with events simulated using GEANT4. Smaller WD values indicate better performance. Figure 3 presents the WD distances for various pion energies in both the training and testing datasets. The WD spectrum demonstrates that the NF performs exceptionally well within the intermediate range of pion incident energies. However, the NF model struggles to extrapolate the distributions to lower and higher bounds, which could be attributed to fewer sampled incident energies in these regions.

Figure 4 compares the distributions generated by the normalizing flow (NF) with GEANT4-simulated ones for incident pion energies that were not part of the training



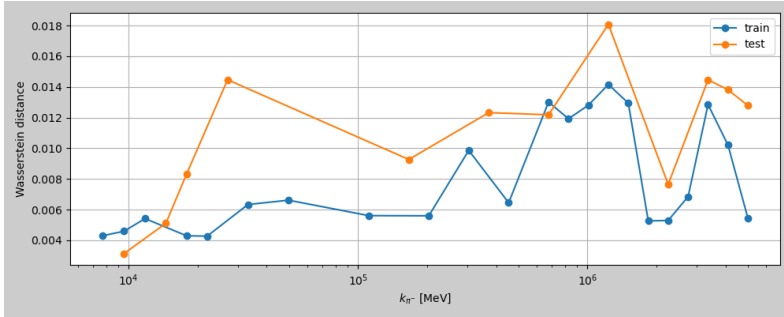
**Figure 4.** Comparison of energy  $E$ , transverse momentum  $p_T$ , and pseudorapidity  $\eta$ ,  $p_z$ , and four-vector mass of the leading outgoing particle between GEANT4-simulated events (“Truth”, blue lines) and Normalizing Flow-generated events (“Prediction”, black lines) for incidence pion kinetic energy of 200 MeV, 1.4 GeV, 2.2 GeV, and 6.5 GeV listed from top to bottom. We use the NF model to generate the same amount of events ten times with different seeds. The mean and standard deviations (“Prediction STD”) of the ten distributions are shown.

data. It’s important to note that the model is specifically trained to predict the four-vector components ( $p_x, p_y, p_z$ , and  $E$ ) of the particle. To accurately predict the transverse momentum ( $p_T$ ), the model must capture the correlations between  $p_x$  and  $p_y$ . For incident energies of  $k_\pi = 1.4$  GeV and 1.2 GeV, we observe a close agreement between the NF-generated and the GEANT4-generated events. However, it becomes evident that more incident energies are necessary for further enhancing the model’s performance, particularly in the lower and higher energy regions.

In our initial exploration of the data generated at  $k_\pi > 9$  GeV, we’ve noticed that the kinematic distributions exhibit a more gradual and continuous variation when only one model, namely the FRITIOF model, is used. Consequently, the normalizing flow can conditionally simulate the data with significantly improved agreement with the true distribution in this region. This improvement is evidenced by the Wasserstein distance observed in Fig. 5.

## 4 Conclusions

Non-perturbative hadronic interactions pose significant physics challenges and formidable computational challenges. Without first-principle theoretical guidance,



**Figure 5.** Wasserstein distance between the NF-generated events and the GEANT4-simulated events for different pion energies for the training and testing dataset.

the primary approach for simulating such interactions is to derive knowledge from experimental measurements, as seen in the implementation of numerous parametric models within the GEANT4 framework.

This study represents a promising avenue where Conditional Normalizing Flow (CNF) is employed to learn and simulate non-perturbative hadronic interactions based on simulated events. The conditional NF can predict outgoing particle properties with reasonable accuracy in specific energy regions for incident pions.

Preliminary studies shows that the CNF is not faster than Bertini or Fritiof models in CPUs. However, the CNF can be easily accelerated with GPUs.

Nonetheless, there’s still more work ahead to achieve a comprehensive simulation of hadronic interactions. The primary challenge lies in generating a variable number of outgoing particles and discrete particle types. The multiplicity of outgoing particles is not solely determined by the incident energy but also by the underlying physics processes. It could be intriguing to develop a physics-informed generative model to address this complexity.

Another significant challenge is constructing a single model covering the entire energy spectrum. In GEANT4 simulations, particularly in the FTFP\_BERT\_ATL physics list, different models are applied to describe hadronic interactions for incident energies below 9 GeV and those above 10 GeV, with an ad-hoc mixture of the two models for energies in between. However, there should be no such transition region in natural as in the GEANT4 simulation. It’s worth exploring whether a single generative model can simulate hadronic interactions consistently across the entire energy spectrum. Answering this question will require learning from real experimental data rather than solely relying on data generated by GEANT4.

This work was supported through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357.

## References

- [1] S. Agostinelli et al. (GEANT4), Nucl. Instrum. Meth. A **506**, 250 (2003)
- [2] M.P. L. de Olivera, B. Nachman, Computing and Software for Big Science (2017) **1**, 4 (2017), 1701.05927
- [3] L.d.O. M. Paganini, B. Nachman, Phys. Rev. Lett. **120**, 042003 (2018), 1705.02355

- [4] L.d.O. M. Paganini, B. Nachman, Phys. Rev. D **97**, 014021 (2018), 1712.10321
- [5] G. Aad et al. (ATLAS) (2021), 2109.02551
- [6] ATLAS Collaboration, *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks* (Geneva, 2020), all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2020-006>, <http://cds.cern.ch/record/2746032>
- [7] *Deep generative models for fast shower simulation in ATLAS* (Geneva, 2018), <http://cds.cern.ch/record/2630433>
- [8] V. Chekalina, E. Orlova, F. Ratnikov, D. Ulyanov, A. Ustyuzhanin, E. Zakharov, CHEP 2018 (2018), 1812.01319
- [9] K. Deja, T. Trzcinski, L. Graczykowski, EPJ Web Conf. **214**, 06003 (2019)
- [10] V. Barin Pacela, *HGCAL Fast Simulation with Deep Learning* (2018), <https://doi.org/10.5281/zenodo.1470512>
- [11] A. Dotti, J. Apostolakis, G. Folger, V. Grichine, V. Ivanchenko, M. Kosov, A. Ribon, V. Uzhinsky, D.H. Wright, J. Phys. Conf. Ser. **293**, 012022 (2011)
- [12] D. Rezende, S. Mohamed, *Variational Inference with Normalizing Flows*, in *Proceedings of the 32nd International Conference on Machine Learning*, edited by F. Bach, D. Blei (PMLR, Lille, France, 2015), Vol. 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538, <https://proceedings.mlr.press/v37/rezende15.html>
- [13] C. Krause, D. Shih (2021), 2110.11377
- [14] C. Krause, D. Shih (2021), 2106.05285
- [15] T. Abbott et al. (E-802), Phys. Rev. D **45**, 3906 (1992)
- [16] T.S. Nigmanov et al. (MIPP), Phys. Rev. D **83**, 012002 (2011), 1010.6291
- [17] P.D. Group, Progress of Theoretical and Experimental Physics **2022**, 083C01 (2022), <https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac09>
- [18] G. Papamakarios, T. Pavlakou, I. Murray (2017), 1705.07057
- [19] M. Abadi et al., *TensorFlow: Large-scale machine learning on heterogeneous systems* (2015), software available from tensorflow.org, <http://tensorflow.org/>
- [20] P.T. Komiske, E.M. Metodiev, J. Thaler, Phys. Rev. Lett. **123**, 041801 (2019), 1902.02346
- [21] T. Cai, J. Cheng, K. Craig, N. Craig, Phys. Rev. D **105**, 076003 (2022)