Uncertainties associated with GAN-generated datasets in high energy physics

Konstantin T. Matchev, Prasanth Shyamsundar*

Institute for Fundamental Theory, Physics Department, University of Florida, Gainesville, FL 32611, USA
* prasanths@ufl.edu

May 29, 2020

Abstract

Recently, Generative Adversarial Networks (GANs) trained on samples of traditionally simulated collider events have been proposed as a way of generating larger simulated datasets at a reduced computational cost. In this paper we present an argument cautioning against the usage of this method to meet the simulation requirements of an experiment, namely that data generated by a GAN cannot statistically be better than the data it was trained on.

Contents

1 Introduction **1**

2 The argument **3**
2.1 Information-theoretic outlook 4
2.2 Evaluating the robustness of GAN-generated datasets 5

3 Caveats **6**

References **8**

---

1 Introduction

Statistical inference in modern high energy experiments is typically done by comparing experimentally recorded data against simulated data corresponding to different theory models and parameters. In order to tap into the full statistical sensitivity offered by the experimental data, the simulated datasets need to keep up in volume to the real datasets, at least up to the same order of magnitude, and preferably beyond. This, coupled with the fact that Monte Carlo (MC) event generation, at least presently, is computationally expensive, poses a difficult challenge to the high energy community in terms of computational needs [1–4]. This problem will only be exacerbated at upcoming big-data experiments like the High Luminosity Large Hadron Collider (HL-LHC) [5].

Generative Adversarial Networks (GANs) [6] represent a class of machine learning (ML) models under which a generative network is trained to produce data that is statistically similar to a sample of training data. In the last few years, GANs and a few of their variants have been explored as tools to accelerate event generation[1] in high energy physics [8–30]. The idea is to

1. Train a GAN on data simulated using the available, slow, MC event generator. Alternatively, the GAN can be trained on real calibration data samples as done in [8].

2. Use the trained GAN to generate additional, statistically similar, data.

GANs can be several orders of magnitude faster than the event generators they are trying to mimic, and this speed-up can potentially be leveraged to bring down the overall computational cost of the data generation process [8–30]. Additionally, by combining several event generation stages together and directly generating high-level features on which analyses will be performed, GANs can also potentially reduce the MC disk space requirements [27]. Note that the training data still needs to be generated using the non-GAN MC generator. Consequently, in order to actually exploit the advantages offered by a GAN, its training dataset needs to be much smaller than the number of simulated events required by the analysis.

There are several works in the literature that propose to replace specific steps of the high energy MC generation pipeline with GANs. For example, the usage of GANs to *only* perform detector/calorimeter simulation was explored in [8–21]. The use of GANs for simulating the underlying hard process was considered in [22–24], while the use of GANs for pileup description was explored in [25, 26]. Recent works have also explored the idea of replacing the *entire* reconstructed-event generation pipeline with a GAN [21, 22, 25–30]. Each of these applications of GANs would differ in the nature of the training data used.

In this paper, we present an argument cautioning against the usage of GANs to meet the simulation requirements of an analysis. The basic idea is that any feature that cannot be resolved using the training data cannot be extrapolated from the training data. As an analogy, if a grainy, low-resolution CCTV image is sufficient to identify a person on the image, then one can trust a software enhancement [31] that reduces the noise and increases the image-resolution to reveal the person's face more clearly. On the other hand, if the low-res image is consistent with two or more people, even after taking all the available information in the raw image file and noise models into account, then one cannot resolve the ambiguity through software enhancement. The very fact that there are multiple potential matches at the low-res level implies that one can create plausible image-enhancers which can make the synthetic high-res image point to any of the potential matches.

Since a) the justification for leveraging a GAN for speed-up is that the training dataset, by itself, is not large enough to meet the simulation requirements, and b) GANs only extrapolate from the training dataset, we claim that GANs cannot be used to speed-up any sequence of steps of the event generation pipeline which is a bottleneck on the overall *uncertainty* from the MC dataset. In the rest of this paper, we present this argument carefully, and provide some caveats to the argument for the sake of precision and clarity.

[1] Unless specified otherwise, in what follows we shall take "event generation" to mean the complete simulation chain in high energy physics [7], including the stages of parton-level event generation, fragmentation and hadronization, detector simulation and object reconstruction.

2 The argument

Here we lay out the argument against using GAN-generated events to meet the simulation requirements of an analysis. Throughout this paper, the term "true simulator/generator" will refer to the method used to produce the data the GAN will be trained on, regardless of the event generation stage under consideration.

In this section, we will consider the case where a GAN is used to circumvent the entire event generation pipeline (either up to the full event description or the reconstructed event description in terms of high-level objects). In Section 3, we will discuss the applicability of GANs to replace specific steps of the pipeline. To avoid possible misinterpretation of the argument, let us set the stage as follows:

- To be maximally conservative, let us assume that there are no deficiencies in the training of a GAN, and that data generated by the GAN agrees with the data it was trained on, on every conceivable distribution, up to statistical fluctuations. Note that this is an idealized situation which actually bolsters the case for GANs.

- We will not make any assumptions about the region of phase-space in which GANs will be used to provide additional simulated data. In other words, our argument holds regardless of whether the GAN will be used to sample events from the "bulk" of a distribution (well represented in the training data) or the "tail" of a distribution (poorly represented in the training data).

- We will assume that the volume of simulated data the GAN was trained on is, by itself, not sufficient for the desired sensitivity of an experiment, say, either a) to the presence of a signal and/or b) to the value of a parameter. This is fair, considering that if the training data was, by itself, sufficient, one would not need any additional GAN-generated data in the first place.

The key observation is that a GAN does not learn to mimic the true event generator, but rather learns to mimic the data it was trained on. In other words, a larger GAN-generated dataset is not guaranteed to agree with the underlying "true distribution" from which the training data was sampled, and could potentially correspond to any of the different underlying models (parameter values) compatible with the training data. This implies that **the model (parameter) discriminating power of analyses cannot be improved by augmenting the true simulated data with additional GAN-generated data**.

A useful picture here is the following: Let us say an analysis calls for 10,000 random numbers drawn from some true distribution. One cannot simply sample 1,000 data points from this distribution, estimate the underlying density from those 1,000 points (analogous to fitting a curve to the histogram), and sample the remaining 9,000 data points from the estimated density (the fitted curve). The 10,000 data points thus produced are not an adequate substitute for 10,000 data points sampled from the true distribution, despite leading to comparably smooth looking plots.

In technical terms, the statistical uncertainties in the true-simulated training data (depicted by the green dashed line in figure 1) are inherited by GAN-generated data as systematic uncertainties (the blue dotted line in figure 1). This is because the specific realization of the training data can be thought of as a probabilistic nuisance parameter in the generative model of the trained GAN. Additionally, the systematic uncertainties of the GAN-generated data

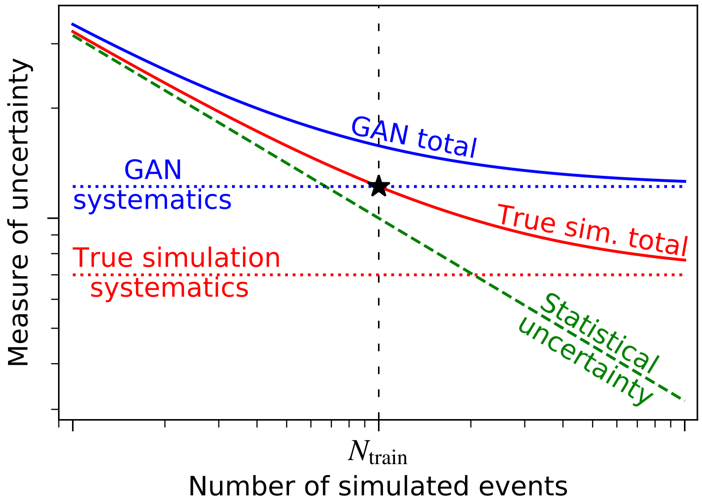

Figure 1: Typical evolution of various uncertainties with the number of simulated events. The green dashed line corresponds to the statistical uncertainty from simulation, while the red dotted and blue dotted lines represent the systematic uncertainty of the true simulation and the GAN simulation, respectively. The red (blue) solid line gives the total uncertainty of the true simulation (the GAN simulation), with the corresponding statistical and systematic uncertainties added in quadrature. We assume that the GAN is trained on a sample of N_{train} events, and as a result, the GAN-generated data inherits the total uncertainty from the training sample as depicted by the ★ symbol. For concreteness, the blue curves correspond to purely GAN-generated data, and not GAN-augmented data.

will contain the systematic uncertainties in the training data (the red dotted line) and the ML training related uncertainties (which we are neglecting here). Since systematic uncertainties do not reduce with increasing sample size, a GAN cannot be used to beat down the statistical uncertainties in the data it was trained on—note how the solid blue line in figure 1 representing the total GAN uncertainty asymptotes *not* to the asymptotic value of the total true simulation uncertainty, but only to the value corresponding to the finite fixed number N_{train} of training events.

By statistical and systematic "uncertainties in/from/of simulated data", we mean the uncertainties in a particular analysis caused, respectively, by the limited statistics of the simulated data and the uncertainties in the generative model. Note that experiments usually include *all* uncertainties from simulated data under systematic uncertainties, with the label "statistical uncertainties" in the experimental results being reserved for statistical uncertainties from real experimental data.

2.1 Information-theoretic outlook

It is instructive to review the primary role played by simulated data in collider analyses. The goal of the collider experiments is to probe the underlying true theory of nature and figure out which of the competing hypotheses (models/model parameter values) best matches it[2].

[2] In the case of unmodeled searches, the goal is to identify deviations from the null hypothesis, but this is a moot detail.

The only handle we have on the true theory of nature is the real experimental data. On the other hand, since our theoretical models do not provide closed form expressions for the distributions of experimental data (especially once we account for the cuts, efficiencies and detector resolution), simulated data serve as our respective handle on the theoretical model candidates. The more real experimental events we have, the better our handle on the true theory. Similarly, the more simulated data we have, the better our handle on the competing theory models. The quality of the hypothesis test or parameter measurement depends on the quality of both these handles, and hence the need for simulated data to keep up in volume with real data. Note that this inference paradigm, under which experimental high energy physics works, requires that the simulated data be faithful to the assumed theoretical hypotheses[3].

To explicitly see that GAN generated data cannot contain any more information about the assumed candidate theory models/model parameters than the data it was trained on, note that a GAN generated dataset can be fully described by providing

1. The prescription for how the GAN will be trained, including the prescription for choosing architectures and tuning any hyperparameters.

2. The training dataset.

3. The (pseudo) random numbers that were used in the training process.

4. The (pseudo) random numbers that were used in the event generation process using the trained GAN.

Out of these, only the training dataset contains any information regarding the hypothesized theoretical candidate models—this information in the training dataset places an upper-limit on the information contained in the full GAN-generated dataset regarding (the differences between) the theoretical candidate models, as a consequence of the data processing inequality.

2.2 Evaluating the robustness of GAN-generated datasets

To further clarify our position, let us discuss what it should mean for a GAN to be an adequate substitute for the true simulator it is trying to mimic. Let a GAN be trained on N_{train} training events generated by the true simulator, and subsequently used to generate N_{required} GAN-events required for the analysis, with $N_{\text{required}} > N_{\text{train}}$.

The N_{required} GAN-events would be an adequate substitute for an equal number of events generated by the true simulator only if the histograms constructed out of the GAN-events (for the purposes of the relevant analysis) are consistent with the true-simulator histograms within the error-bars. Note that the relative/percent error-bars should correspond to a dataset with N_{required} events and not a dataset with only N_{train} events. **Visual similarity or any other measure of robustness that does not become stricter as N_{required} increases is not an adequate measure of the robustness of GAN-generated datasets.** Establishing this robustness could be as simple as showing histograms (using N_{required} events) with error-bars and quoting, say, the χ^2 goodness-of-fit value relative to the true simulation, demonstrating good fits. Unfortunately, this simple cross-check is rarely being done in the majority of the

[3] A moot point here is that there are different aspects of the generative models that we are uncertain about, some of which we want to be sensitive to (e.g., parton level theory), and others we want to minimize impact from (e.g., detector effects, or non-perturbative QCD in BSM physics searches). These are treated on different footings in Monte Carlo research.

current GAN literature in high energy physics with proposals to circumvent the entire event generation pipeline using GANs.

The claim in Section 2 is the following: It is not possible to establish that the GAN-generated dataset will be robust for a given value of N_{required} without actually validating the GAN-data against a comparible number of true-simulated data[4], which would negate any speed-up offered by GANs. Without such a validation, one will be forced to retain the relative/percent error-bars corresponding to a dataset of N_{train} events as systematic uncertainties (assuming perfect training of the GAN, otherwise the errors would be even larger).

In concluding this section, we point out that if one is willing to entertain the idea that the robustness of GAN-datasets larger than the corresponding training datasets can be established without validation, then one should also be willing to accept the absurd idea that GANs can be used to replace future experiments that seek to improve on the luminosity of current experiments (say, HL-LHC for LHC) by training on the current experimental data and producing large quantities of new GAN-generated "experimental data" (in the entire phase-space or parts of it). This task would arguably be easier than training GANs on simulated data, since real datasets do not need to be generated at different values of model parameters.

3 Caveats

We present the following caveats to our argument, both for the sake of precision and to clarify our position on the usage on GANs in high energy physics:

- The event generation and reconstruction pipeline employed in high energy physics is a complex one, and addressing the challenge from limited computational resources will require a multifaceted effort. We believe that ML-based generation approaches [32], including GANs, could play a role in the eventual solution.

  A specific sequence of steps in the event simulation process could very well be accelerated using GANs, as long as a) the sequence in question is not a major bottleneck on the overall uncertainty from the simulated dataset and b) **the trade-off between accuracy and speed is acknowledged**.

  - For example, conditional-GANs [33] could potentially be used to model the response of a detector to an incident particle of, say, a given identity, energy, and momentum, as shown in [8–21]. But this would be inappropriate if, for instance, one of the reasons for the analysis to need more simulated events is to estimate the frequency of *rare* detector-induced fakes.
  - Similarly, GANs could potentially be used to replace the underlying hard-scattering simulation, unless the reason for needing more events is to pickup subtle higher-order effects or improve the resolution of an underlying theory parameter (at the parton level), as is often the case.
  - Another example in the literature is the usage of GANs for pileup description [25,26], where a GAN is to be trained to generate minimum-bias events. This could

[4]Note that it is technically possible for a GAN to accidentally be robust up to N_{required} events. Likewise, even without any training, a lucky random initialization of weights could technically lead to a GAN which perfectly mimics the true-simulator. But neither of these can be assumed without validation.

be appropriate in situations when simply resampling[5] from an existing library of minimum-bias events (used as GAN training data) is sufficient for the purposes of the analysis.

In all these cases, while the GAN will not be a statistically equivalent substitute for the corresponding true generator, it could potentially offer a better trade-off between accuracy and computational/storage resources than other available fast-simulators. It goes without saying that even when a particular application of GAN is valid under this consideration, one still needs to estimate the systematic uncertainties in the GAN-generated data (including ML training related uncertainties) in the context of the particular analysis at hand [34].

The goal of this paper is not to discourage research on the application of GANs, but to present a consideration that could affect the applicability of GANs for different purposes within high energy physics. An evaluation of all the proposed applications of GANs in the high energy literature from this point of view is, however, beyond the scope of this work.

- Note that it is likely that a given analysis is not optimal in its usage of the simulated events. In that case, a GAN could *potentially* tap into the high-dimensional information contained in the simulated training sample in a more optimal way, and provide a robust larger dataset to the suboptimal analysis.

  However, without an estimate of this robustness, one will be forced to include the statistical uncertainty of the training sample as part of the systematic uncertainty of the GAN-generated sample, as explained in the previous section. Note that the goal of analyses is to reduce the *reported* uncertainty on measurements or the *reported* upper-limits on signal cross-sections, and not some hypothetical and intangible measures of sensitivity.

  Even for a suboptimal analysis, we presently do not see any way of estimating the robustness of GAN-generated data, without actually producing a larger validation dataset (which, of course, would defeat the purpose of using GANs for speedup). Nevertheless, we leave to the individual practitioners the onus of forecasting the robustness of a GAN-generated dataset larger than the dataset it was trained on, for use in a particular analysis.

- We would like to point out that our argument is not pertinent to parton-level ML-based generative approaches which learn directly from an oracle which can be queried for the underlying true distribution (matrix-element), instead of learning from data generated under that distribution [35–41].

  Our argument also does not affect the usage of GANs in situations where a GAN-generated dataset will not be used to yield additional sensitivity in a given analysis. For example, where appropriate, a GAN could be used as a substitute for event resampling techniques like bootstrapping and jackknifing, albeit a more complicated one. Other possible applications include the use of GANs for unfolding [42, 43], denoising [44], event subtraction [45], and model building [46].

[5]Note that "resampling" implies "with replacement", i.e., recycling the events over and over, without generating new ones.

Finally, we would like to mention that our argument could potentially affect the application of GANs for simulations outside high energy physics as well [47–50].

Acknowledgements

The authors would like to thank D. Acosta for useful comments.

Funding information The work of KM and PS is supported in part by the United States Department of Energy under Grant No. DE-SC0010296. The work of PS is supported by the University of Florida CLAS Dissertation Fellowship funded by the Charles Vincent and Heidi Cole McLaughlin Endowment.

References

[1] L. A. T. Bauerdick *et al.*, "Planning the Future of U.S. Particle Physics (Snowmass 2013): Chapter 9: Computing Frontier," arXiv:1401.6117 [hep-ex].

[2] C. Group *et al.*, "Fermilab Computing at the Intensity Frontier," J. Phys. Conf. Ser. **664**, no. 3, 032012 (2015) doi:10.1088/1742-6596/664/3/032012.

[3] J. Albrecht *et al.* [HEP Software Foundation Collaboration], "A Roadmap for HEP Software and Computing R&D for the 2020s," Comput. Softw. Big Sci. **3**, no. 1, 7 (2019) doi:10.1007/s41781-018-0018-8 [arXiv:1712.06982 [physics.comp-ph]].

[4] A. Buckley, "Computational challenges for MC event generation," arXiv:1908.00167 [hep-ph].

[5] D. Adamova and M. Litmaath, "New strategies of the LHC experiments to meet the computing requirements of the HL-LHC era," PoS BORMIO **2017**, 053 (2017) doi:10.22323/1.302.0053.

[6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Networks," arXiv:1406.2661 [stat.ML].

[7] S. Ask *et al.*, "From Lagrangians to Events: Computer Tutorial at the MC4BSM-2012 Workshop," arXiv:1209.0297 [hep-ph].

[8] A. Maevskiy *et al.* [LHCb Collaboration], "Fast Data-Driven Simulation of Cherenkov Detectors Using Generative Adversarial Networks," arXiv:1905.11825 [physics.ins-det].

[9] D. Belayneh *et al.*, "Calorimetry with Deep Learning: Particle Simulation and Reconstruction for Collider Physics," arXiv:1912.06794 [physics.ins-det].

[10] J. R. Vlimant, F. Pantaleo, M. Pierini, V. Loncar, S. Vallecorsa, D. Anderson, T. Nguyen and A. Zlokapa, "Large-Scale Distributed Training Applied to Generative Adversarial Networks for Calorimeter Simulation," EPJ Web Conf. **214**, 06025 (2019) doi:10.1051/epjconf/201921406025.

[11] D. Lancierini, P. Owen and N. Serra, "Simulating the LHCb hadron calorimeter with generative adversarial networks," Nuovo Cim. C **42**, no. 4, 197 (2019) doi:10.1393/ncc/i2019-19197-3.

[12] L. de Oliveira, M. Paganini and B. Nachman, "Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis," Comput. Softw. Big Sci. **1**, no. 1, 4 (2017) doi:10.1007/s41781-017-0004-6 [arXiv:1701.05927 [stat.ML]].

[13] S. Carrazza and F. A. Dreyer, "Lund jet images from generative and cycle-consistent adversarial networks," Eur. Phys. J. C **79**, no. 11, 979 (2019) doi:10.1140/epjc/s10052-019-7501-1 [arXiv:1909.01359 [hep-ph]].

[14] M. Paganini, L. de Oliveira and B. Nachman, "Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters," Phys. Rev. Lett. **120**, no. 4, 042003 (2018) doi:10.1103/PhysRevLett.120.042003 [arXiv:1705.02355 [hep-ex]].

[15] L. de Oliveira, M. Paganini and B. Nachman, "Controlling Physical Attributes in GAN-Accelerated Simulation of Electromagnetic Calorimeters," J. Phys. Conf. Ser. **1085**, no. 4, 042017 (2018) doi:10.1088/1742-6596/1085/4/042017 [arXiv:1711.08813 [hep-ex]].

[16] M. Paganini, L. de Oliveira and B. Nachman, "CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks," Phys. Rev. D **97**, no. 1, 014021 (2018) doi:10.1103/PhysRevD.97.014021 [arXiv:1712.10321 [hep-ex]].

[17] F. Carminati, A. Gheata, G. Khattak, P. Mendez Lorenzo, S. Sharan and S. Vallecorsa, "Three dimensional Generative Adversarial Networks for fast simulation," J. Phys. Conf. Ser. **1085**, no. 3, 032016 (2018) doi:10.1088/1742-6596/1085/3/032016.

[18] M. Erdmann, L. Geiger, J. Glombitza and D. Schmidt, "Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks," Comput. Softw. Big Sci. **2**, no. 1, 4 (2018) doi:10.1007/s41781-018-0008-x [arXiv:1802.03325 [astro-ph.IM]].

[19] P. Musella and F. Pandolfi, "Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks," Comput. Softw. Big Sci. **2**, no. 1, 8 (2018) doi:10.1007/s41781-018-0015-y [arXiv:1805.00850 [hep-ex]].

[20] M. Erdmann, J. Glombitza and T. Quast, "Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network," Comput. Softw. Big Sci. **3**, no. 1, 4 (2019) doi:10.1007/s41781-018-0019-7 [arXiv:1807.01954 [physics.ins-det]].

[21] S. Vallecorsa, F. Carminati and G. Khattak, "3D convolutional GAN for fast simulation," EPJ Web Conf. **214**, 02010 (2019) doi:10.1051/epjconf/201921402010.

[22] S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen, D. Podareanu, R. R. de Austri and R. Verheyen, "Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer," arXiv:1901.00875 [hep-ph].

[23] A. Butter, T. Plehn and R. Winterhalder, "How to GAN LHC Events," SciPost Phys. **7**, no. 6, 075 (2019) doi:10.21468/SciPostPhys.7.6.075 [arXiv:1907.03764 [hep-ph]].

[24] C. Ahdida *et al.* [SHiP Collaboration], "Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks," JINST **14**, P11028 (2019) doi:10.1088/1748-0221/14/11/P11028 [arXiv:1909.04451 [physics.ins-det]].

[25] S. Farrell, W. Bhimji, T. Kurth, M. Mustafa, D. Bard, Z. Lukic, B. Nachman and H. Patton, "Next Generation Generative Neural Networks for HEP," EPJ Web Conf. **214**, 09005 (2019) doi:10.1051/epjconf/201921409005.

[26] J. Arjona Martínez, T. Q. Nguyen, M. Pierini, M. Spiropulu and J. R. Vlimant, "Particle Generative Adversarial Networks for full-event simulation at the LHC and their application to pileup description," arXiv:1912.02748 [hep-ex].

[27] B. Hashemi, N. Amin, K. Datta, D. Olivito and M. Pierini, "LHC analysis-specific datasets with Generative Adversarial Networks," arXiv:1901.05282 [hep-ex].

[28] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat and S. Palazzo, "A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC," PoS LeptonPhoton **2019**, 050 (2019) doi:10.22323/1.367.0050.

[29] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat and S. Palazzo, "DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC," JHEP **1908**, 110 (2020) doi:10.1007/JHEP08(2019)110 [arXiv:1903.02433 [hep-ex]].

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

[31] M. M. P. Petrou and C. Petrou, "Image Processing: The Fundamentals (2nd ed.)," (2010) John Wiley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, UK.

[32] K. Albertsson *et al.*, "Machine Learning in High Energy Physics Community White Paper," J. Phys. Conf. Ser. **1085**, no. 2, 022008 (2018) doi:10.1088/1742-6596/1085/2/022008 [arXiv:1807.02876 [physics.comp-ph]].

[33] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," arXiv:1411.1784 [cs.LG]

[34] K. T. Matchev, A. Roman and P. Shyamsundar, "Evaluating GAN-generated datasets from a physics analysis perspective," in preparation.

[35] J. Bendavid, "Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks," arXiv:1707.00028 [hep-ph].

[36] M. D. Klimek and M. Perelstein, "Neural Network-Based Approach to Phase Space Integration," arXiv:1810.11509 [hep-ph].

[37] C. Gao, J. Isaacson and C. Krause, "i-flow: High-Dimensional Integration and Sampling with Normalizing Flows," arXiv:2001.05486 [physics.comp-ph].

[38] C. Gao, S. Höche, J. Isaacson, C. Krause and H. Schulz, "Event Generation with Normalizing Flows," Phys. Rev. D **101**, no.7, 076002 (2020) doi:10.1103/PhysRevD.101.076002 [arXiv:2001.10028 [hep-ph]].

[39] F. Bishara and M. Montull, "(Machine) Learning Amplitudes for Faster Event Generation," arXiv:1912.11055 [hep-ph].

[40] E. Bothmann, T. Janßen, M. Knobbe, T. Schmale and S. Schumann, "Exploring phase space with Neural Importance Sampling," SciPost Phys. **8**, no.4, 069 (2020) doi:10.21468/SciPostPhys.8.4.069 [arXiv:2001.05478 [hep-ph]].

[41] S. Otten, K. Rolbiecki, S. Caron, J. S. Kim, R. Ruiz De Austri and J. Tattersall, "DeepXS: Fast approximation of MSSM electroweak cross sections at NLO," Eur. Phys. J. C **80**, no. 1, 12 (2020) doi:10.1140/epjc/s10052-019-7562-1 [arXiv:1810.08312 [hep-ph]].

[42] K. Datta, D. Kar and D. Roy, "Unfolding with Generative Adversarial Networks," arXiv:1806.00433 [physics.data-an].

[43] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn and R. Winterhalder, "How to GAN away Detector Effects," SciPost Phys. **8**, no.4, 070 (2020) doi:10.21468/SciPostPhys.8.4.070 [arXiv:1912.00477 [hep-ph]].

[44] M. Shirasaki, N. Yoshida, S. Ikeda, T. Oogi and T. Nishimichi, "Decoding Cosmological Information in Weak-Lensing Mass Maps with Generative Adversarial Networks," arXiv:1911.12890 [astro-ph.CO].

[45] A. Butter, T. Plehn and R. Winterhalder, "How to GAN Event Subtraction," arXiv:1912.08824 [hep-ph].

[46] H. Erbin and S. Krippendorf, "GANs for generating EFT models," arXiv:1809.02612 [cs.LG].

[47] A. C. Rodriguez, T. Kacprzak, A. Lucchi, A. Amara, R. Sgier, J. Fluri, T. Hofmann and A. Réfrégier, "Fast cosmic web simulations with generative adversarial networks," Comput. Astrophys. Cosmol. **5**, 4 (2018) doi:10.1186/s40668-018-0026-4 [arXiv:1801.09070 [astro-ph.CO]].

[48] J. Zamudio-Fernandez, A. Okan, F. Villaescusa-Navarro, S. Bilaloglu, A. D. Cengiz, S. He, L. Perreault Levasseur and S. Ho, "HIGAN: Cosmic Neutral Hydrogen with Generative Adversarial Networks," arXiv:1904.12846 [astro-ph.CO].

[49] A. Mishra, P. Reddy and R. Nigam, "CMB-GAN: Fast Simulations of Cosmic Microwave background anisotropy maps using Deep Learning," arXiv:1908.04682 [astro-ph.CO].

[50] F. List, I. Bhat and G. F. Lewis, "A black box for dark sector physics: Predicting dark matter annihilation feedback with conditional GANs," Mon. Not. Roy. Astron. Soc. **490**, no. 3, 3134 (2019) doi:10.1093/mnras/stz2759 [arXiv:1910.00291 [astro-ph.CO]].