
Induced Generative Adversarial Particle Transformers

Anni Li* Venkat Krishnamohan* Raghav Kansal†
 Rounak Sen Steven Tsan Zhaoyu Zhang Javier Duarte
 University of California, San Diego
 La Jolla, CA 92093, USA

{a5li,vkrishnamohan,rkansal,r2sen,stsan,zhz057,jduarte}@ucsd.edu

Abstract

In high energy physics (HEP), machine learning methods have emerged as an effective way to accurately simulate particle collisions at the Large Hadron Collider (LHC). The message-passing generative adversarial network (MPGAN) was the first model to simulate collisions as point, or “particle”, clouds, with state-of-the-art results, but suffered from quadratic time complexity. Recently, generative adversarial particle transformers (GAPTs) were introduced to address this drawback; however, results did not surpass MPGAN. We introduce induced GAPT (iGAPT) which, by integrating “induced particle-attention blocks” and conditioning on global jet attributes, not only offers linear time complexity but is also able to capture intricate jet substructure, surpassing MPGAN in many metrics. Our experiments demonstrate the potential of iGAPT to simulate complex HEP data accurately and efficiently.

1 Introduction

In high energy physics (HEP), jets—sprays of particles produced at high energy collisions at the Large Hadron Collider (LHC), serve as a vital bridge between theoretical research and experimental observations. The intricate patterns in jets make them invaluable for understanding quantum chromodynamics (QCD), identifying rare particles like the Higgs boson, and conducting quantitative analyses of particle properties and interactions. Machine learning methods, such as generative adversarial networks (GANs), have recently emerged as effective and efficient approaches to simulating such high-energy collisions, reducing the computational burden of traditional methods. The message-passing GAN (MPGAN) [1], in particular, represented a significant leap forward, being able to capture the complex global structure of jets and handle variable-sized particle clouds using fully-connected graph neural networks. However, this fully-connected nature means its memory and time complexity scale quadratically with the number of particles per jet.

The generative adversarial particle transformer (GAPT) [2] was recently introduced to improve the efficiency of MPGAN using self-attention, but GAPT does not match MPGAN’s performance. Others have also studied normalizing flows [3], other types of transformers [4, 5], diffusion models [6, 7], and equivariant networks [8] for the same generative task. In this paper, we introduce the induced GAPT (iGAPT), featuring “induced particle attention blocks” incorporating physics-informed inductive biases of jets, to offer linear time complexity and improve output fidelity. We investigate its performance and timing compared to MPGAN and alternative architecture choices on high-momentum jets.

*Equal contribution.

†Also at Fermi National Accelerator Laboratory.

2 Architecture

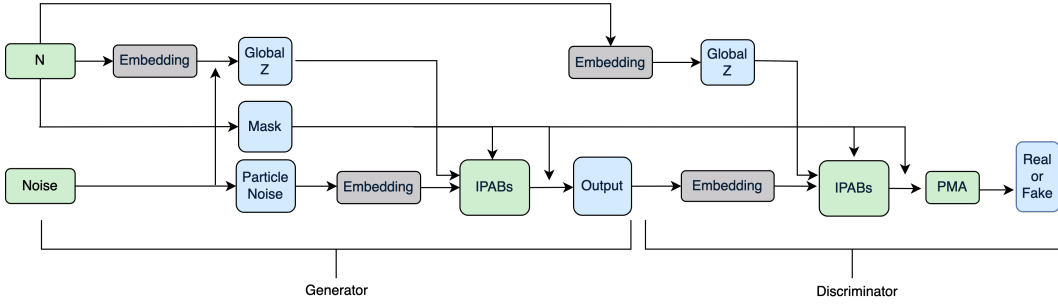


Figure 1: Illustration of the generation and discrimination process of iGAPT.

The iGAPT model, illustrated in Fig. 1, is a GAN with the generator and discriminator comprising “induced particle attention blocks” (IPABs). On top of permutation invariance and point-cloud representations, as in MPGAN and GAPT, the key inductive bias we experiment with in iGAPT is maintaining a global vector through the generation and discrimination processes, \mathbf{z} , which implicitly represents global jet features. IPABs and different ways of incorporating \mathbf{z} into the attention process are described below.

The generation process starts with sampling random Gaussian noise and a particle multiplicity N from the true distribution, which is transformed via a learned linear embedding layer. The noise has two components: a set of N_{\max} vectors representing initial particle features, and a single vector representing initial jet features. N_{\max} is the maximum number of particles per jet we want to simulate, and the number of initial particle and jet features is a hyperparameter we tune. The jet noise is added to the embedded N to produce \mathbf{z} , which is then transformed along with the particle noise via multiple IPABs to output a generated jet.

The discrimination process starts with a generated or real jet, and the sampled or true jet multiplicity N . This is again transformed via a learned embedding layer to produce the \mathbf{z} conditioning vector for the discriminator, which along with the input jet are processed through IPABs producing an intermediate jet representation. The constituents of this jet are aggregated in a permutation-invariant manner using a pooling by multihead attention (PMA) layer, as introduced in [9], the output of which is finally fed into a linear layer to classify the jet as real or fake.

Attention blocks require fixed multiplicity, however, jets naturally have a variable number of particle constituents. To handle this, we zero-pad all jets to N_{\max} particles and use masked attention in every block to ignore these.

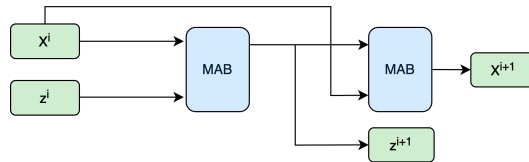


Figure 2: Illustration of an induced particle attention block (IPAB).

Attention Blocks The original GAPT model uses self-attention blocks (SABs) introduced for set transformers [9], a form of multi-head attention blocks (MABs) with the input attending itself — i.e., each particle attends to every other in the jet to update its features. This has the benefit of global, fully connected interactions between particles; however, it also means quadratic scaling with the particle multiplicity, which is undesirable.

To counter this, Ref. [9] also proposed “induced” SABs (ISABs), where the input set is first attended to by M learnt inducing vectors via a MAB, outputting an intermediate, compressed representation. This representation is then attended to by the original set to produce the new output. This retains the global interactions between particles while achieving $\mathcal{O}(NM)$ scaling—linear in the particle

multiplicity. ISABs have been used in generative adversarial set transformers (GAST) [10], yielding high quality results on computer vision datasets. GAST incorporates similar global set features \mathbf{z} by concatenating them to each particle before every ISAB operation.

In iGAPT, we experiment with an alternative version of conditioning, which we call “induced particle attention”. Here, \mathbf{z} is directly used as the inducing vector in an ISAB, and is continuously updated through the intermediate, induced attention outputs. Explicitly, as illustrated in Fig. 2, the i th IPAB receives as input the jet \mathbf{x}^i and global features \mathbf{z}^i from the previous block, after which \mathbf{x}^i is first attended to by \mathbf{z}^i to output updated features \mathbf{z}^{i+1} , and then conversely \mathbf{z}^{i+1} is attended to by \mathbf{x}^i to output the updated jet \mathbf{x}^{i+1} . This is interpreted as a way to update and learn the global jet features, such as mass and p_T , along with the individual particle features, and allow both to interact in each attention layer. An additional and significant advantage is that this induced attention operation involves only one inducing vector — \mathbf{z} , hence $M = 1$ and we achieve $\mathcal{O}(N)$ computational complexity.

3 Experiments

Dataset We perform tests on the JetNet dataset [11] using the associated library [1, 12]. The JetNet dataset contains simulated high transverse-momentum (p_T) jets, originating from gluons, light quarks, top quarks, incorporating detector effects emulating those of the CMS detector at the LHC. We test the iGAPT model on these different classes, simulating only the 30 as well as the 150 highest p_T particles per jet. This complex task involves simulating both the initial parton showering and hadronization from an individual quark or gluon, as well as the proceeding detector response. Each jet is represented as a set, or point cloud, of its constituent particles, which are described by three features: their angular coordinates ($\eta^{\text{rel}}, \phi^{\text{rel}}$), and p_T , relative to the jet.

Training and Hyperparameters The iGAPT and GAST models were trained for a maximum of 6000 epochs on a single NVIDIA RTX1080 GPU using the RMSProp optimizer. The training time and batch size for 30- and 150-particle gluon jets is shown in Table 2. We use a two-time update rule [13], with learning rates of 0.5×10^{-4} and 1.5×10^{-4} for the generator and discriminator, respectively. The discriminator is regularized with dropout with a probability of 0.5 and layer normalization. We use a LeakyReLU activation after every linear layer, except the final generator (tanh) and discriminator (sigmoid) outputs. After hyperparameter tuning, we settle on 3 and 6 ISAB layers for the generator and discriminator respectively, and 4 and 8 IPAB layers for the iGAPT model. We use 16 and 128 initial particle and jet features, respectively, in iGAPT. GAST uses 20 inducing vectors for its ISABs.

Evaluation We evaluate the performance of each model both visually, comparing feature distributions between the real and generated samples, and using quantitative evaluation metrics. Reference [2] found the Fréchet physics distance (FPD), kernel physics distance (KPD), and 1-Wasserstein distances (W_1) between individual particle and jet features to be sensitive metrics to comprehensively evaluate the physics performance of generative models. FPD and KPD, in particular, consider a set of 36 energy flow polynomials (EFPs) [14], which collectively form a basis for all infrared- and collinear-safe observables, and hence capture significant information about the jet substructure. We report the FPD, KPD, and W_1 scores for the particle p_T ($W_{1p}^{p_T^{\text{rel}}}$) and jet mass (W_1^M) and their errors following the procedures outlined in Ref. [2] using the JetNet library for each generative model, as well as between independent samples of real jets to measure a baseline, “truth” score.

Results We test the iGAPT model on 30-particle gluon, light quark, and top quark jets, and on 150-particle gluon jets. We compare this with the MPGAN and GAPT models provided in their respective publications, as well as as our own trainings of GAST models. Out of all our trainings, we select the iGAPT and GAST models with the lowest FPD score, as advocated in Ref. [2] due to its high sensitivity to common types of mismodelling. Comparisons of real and generated feature distributions are shown in Fig. 3 and Appendix A for 30 and 150 particles, respectively, demonstrating high fidelity results by iGAPT. The truth and model scores for the aforementioned evaluation metrics are listed in Table 1. We note that MPGAN models for 150-particle jets are not provided because of their poor time complexity. We observe first that iGAPT largely outperforms the GAPT and GAST models, validating our physics-informed approach to its architecture. It also outperforms MPGAN

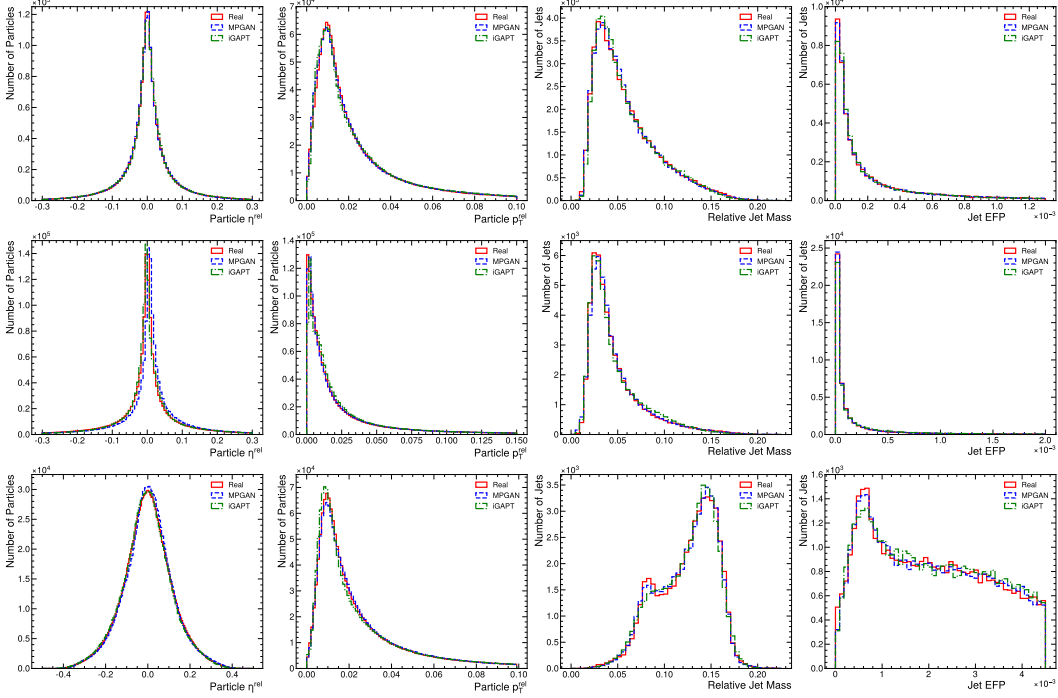


Figure 3: Low-level particle feature distributions (far left and center left) and high-level jet feature distributions (center right and far right) for the real data (red), MPGAN-generated data (blue), and iGAPT-generated data (green), for 30-particle gluon (top row), light quark (middle), and top quark jets (bottom). A sample $d = 4$ energy flow polynomial [14] is plotted in the rightmost column.

on 30-particle jets on many jet types and metrics, notably on FPD for all jet types. Important we recognize that no model is perfectly consistent with the true, baseline distributions, indicating room for improvement.

Timing A key benefit of iGAPT is its improved time complexity over MPGAN. This is demonstrated in Table 2, which shows the training and generation times for each model for 30 particle jets using the largest batch size possible on an NVIDIA 1080 GPU, with iGAPT outperforming MPGAN by a factor of 3.5. MPGAN is computationally challenging to extend to 150 particles, hence timing information is not provided; in contrast, iGAPT’s training and generation times scale well with the number of particles. Finally, we note that the “true” generation time per jet is approximately 50 ms [1], thus iGAPT represents more than a factor of 100 speed up.

4 Conclusion

We presented the induced generative adversarial particle transformer (iGAPT), which utilizes a new induced particle attention mechanism to achieve high-fidelity simulations of jets in high energy physics. By incorporating physics-informed inductive biases and conditioning on global jet features, iGAPT demonstrates not only competitive performance with previous state-of-the-art models on 30-particle gluon, light quark, and top quark jets, but also significantly better computational efficiency and time complexity. Furthermore, iGAPT exhibits strong potential to simulate larger numbers of particles, establishing its viability in applications to realistic event collisions at the LHC. However, we note room for improvement in iGAPT’s performance with respect to the baseline scores, and future work may explore further optimizations of its training and performance.

Table 1: Evaluation metrics for different jet types and models. The best-performing model on each metric and jet type is highlighted in bold.

	Model	$W_{1p}^{p_T^{\text{rel}}} (10^{-3})$	$W_1^M (10^{-3})$	FPD (10^{-3})	KPD (10^{-6})
Gluon (30)	Truth	0.14 ± 0.06	0.46 ± 0.08	0.14 ± 0.04	1.8 ± 11.9
	MPGAN	0.27 ± 0.02	0.7 ± 0.3	0.41 ± 0.09	0 ± 8
	GAPT	0.25 ± 0.07	1.0 ± 0.2	0.46 ± 0.06	5 ± 3
	GAST	0.8 ± 0.1	0.7 ± 0.2	0.40 ± 0.05	7.0 ± 10.3
	iGAPT	0.76 ± 0.07	0.7 ± 0.1	0.29 ± 0.04	3 ± 5
Light quark (30)	Truth	0.21 ± 0.05	0.5 ± 0.2	0.09 ± 0.03	-3 ± 3
	MPGAN	0.41 ± 0.07	0.5 ± 0.1	1.9 ± 0.2	1.7 ± 15.1
	GAPT	2.74 ± 0.09	2.54 ± 0.05	4.03 ± 0.06	96 ± 9
	GAST	1.2 ± 0.1	1.8 ± 0.2	0.65 ± 0.07	27.0 ± 11.7
	iGAPT	1.89 ± 0.04	1.2 ± 0.3	0.51 ± 0.07	12 ± 7
Top quark (30)	Truth	0.20 ± 0.05	0.7 ± 0.2	0.07 ± 0.03	-16 ± 2
	MPGAN	0.44 ± 0.08	0.5 ± 0.1	2.8 ± 0.2	14.7 ± 12.9
	GAPT	0.34 ± 0.02	1.9 ± 0.2	0.43 ± 0.03	25.4 ± 28.8
	GAST	1.16 ± 0.08	1.5 ± 0.2	0.30 ± 0.05	-2.4 ± 17.2
	iGAPT	0.54 ± 0.04	0.9 ± 0.3	0.25 ± 0.03	-0.6 ± 14.1
Gluon (150)	Truth	0.09 ± 0.03	0.7 ± 0.2	0.10 ± 0.03	0.5 ± 10.5
	GAPT	0.77 ± 0.03	1.1 ± 0.3	22.0 ± 0.1	62.5 ± 11.1
	GAST	0.68 ± 0.05	3.7 ± 0.3	3.60 ± 0.06	47.7 ± 13.8
	iGAPT	0.66 ± 0.03	4.4 ± 0.7	2.99 ± 0.06	158.1 ± 37.9

Table 2: Timing measurements for MPGAN and iGAPT, measured on an NVIDIA 1080 GPU.

Jet type	Model	Training time (s/epoch)	Generation time (μs / jet)	Batch size
Gluon, $n = 30$	MPGAN	193	142	512
	iGAPT	31	40	4096
Gluon, $n = 150$	iGAPT	267	315	512

Acknowledgments and Disclosure of Funding

This work was supported by the U.S. Department of Energy (DOE) Award Nos. DE-SC0021187 and DE-SC0021396 (FAIR4HEP) and the U.S. National Science Foundation (NSF) Cooperative Agreement OAC-2117997 (A3D3). A. L. was partially supported by an IRIS-HEP fellowship through the U.S. National Science Foundation (NSF) under Cooperative Agreement OAC-1836650. R. K. was partially supported by the LHC Physics Center at Fermi National Accelerator Laboratory, managed and operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the DOE.

References

- [1] R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, D. Gunopulos, et al., “Particle cloud generation with message passing generative adversarial networks”, in Advances in Neural Information Processing Systems, Vol. 34, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (2021), p. 23858, arXiv:2106.11535 [cs.LG].
- [2] R. Kansal, A. Li, J. Duarte, N. Chernyavskaya, M. Pierini, B. Orzari, and T. Tomei, “Evaluating generative models in high energy physics”, Phys. Rev. D **107**, 076017 (2023), arXiv:2211.10295 [hep-ex].
- [3] B. Orzari, N. Chernyavskaya, R. Cobe, J. Duarte, J. Fialho, D. Gunopulos, R. Kansal, M. Pierini, T. Tomei, and M. Touranakou, “LHC hadronic jet generation using convolutional

- variational autoencoders with normalizing flows”, *Mach. Learn. Sci. Tech.* **4**, 045023 (2023), arXiv:2310.13138 [physics.comp-ph].
- [4] B. Kach, D. Krücker, and I. Melzer-Pellmann, “Point Cloud Generation using Transformer Encoders and Normalising Flows”, (2022), arXiv:2211.13623 [hep-ex].
 - [5] B. Kach and I. Melzer-Pellmann, “Attention to Mean-Fields for Particle Cloud Generation”, (2023), arXiv:2305.15254 [hep-ex].
 - [6] M. Leigh, D. Sengupta, G. Quétant, J. A. Raine, K. Zoch, and T. Golling, “PC-JeDi: Diffusion for Particle Cloud Generation in High Energy Physics”, (2023), arXiv:2303.05376 [hep-ph].
 - [7] V. Mikuni, B. Nachman, and M. Pettee, “Fast point cloud generation with diffusion models in high energy physics”, *Phys. Rev. D* **108**, 036025 (2023), arXiv:2304.01266 [hep-ph].
 - [8] E. Buhmann, G. Kasieczka, and J. Thaler, “EPiC-GAN: Equivariant point cloud generation for particle jets”, *SciPost Phys.* **15**, 130 (2023), arXiv:2301.08128 [hep-ph].
 - [9] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teh, “Set transformer: a framework for attention-based permutation-invariant neural networks”, in *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97, edited by K. Chaudhuri and R. Salakhutdinov (2019), p. 3744, arXiv:1810.00825 [cs.LG].
 - [10] K. Stelzner, K. Kersting, and A. R. Kosiorek, “Generative adversarial set transformers”, in *Workshop on Object-Oriented Learning at ICML*, Vol. 3 (2020), p. 1.
 - [11] R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, *JetNet*, version 2, <https://doi.org/10.5281/zenodo.6975118> (Zenodo, 2022).
 - [12] R. Kansal, C. Pareja, Z. Hao, and J. Duarte, “JetNet: A Python package for accessing open datasets and benchmarking machine learning methods in high energy physics”, *J. Open Source Softw.* **8**, 5789 (2023).
 - [13] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local nash equilibrium”, in *Advances in neural information processing systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (2017), p. 6626, arXiv:1706.08500 [cs.LG].
 - [14] P. T. Komiske, E. M. Metodiev, and J. Thaler, “Energy flow polynomials: A complete linear basis for jet substructure”, *JHEP* **04**, 013 (2018), arXiv:1712.07124 [hep-ph].

A Feature distributions of 150-particle jets

The distributions of real and generated particle and jet features for 150-particle gluon jets by MPGAN and iGAPT are shown in Fig. 4.

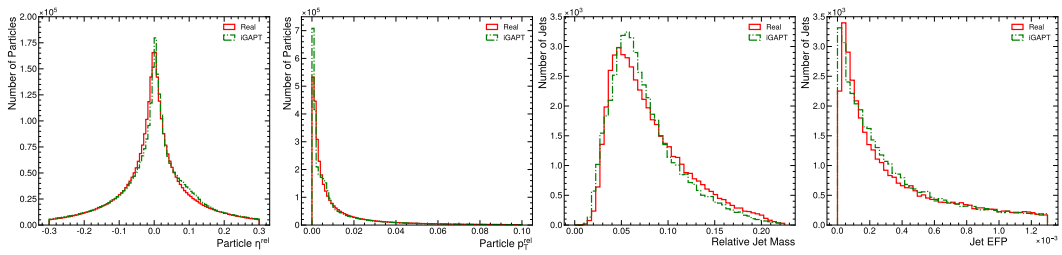


Figure 4: Low-level particle feature distributions (far left and center left) and high-level jet feature distributions (center right and far right) for the real data (red), MPGAN-generated data (blue), and iGAPT-generated data (green), for 150-particle gluon jets. A sample $d = 4$ energy flow polynomial [14] is chosen in the rightmost column.