EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2025

Contribution ID: **186**

Type: **Poster + Flashtalk**

End-to-end Sinkhorn AutoEncoder Latent Diffusion Model for Fast Particle Physics Simulation

Simulations play a crucial role in understanding the complex dynamics of particle collisions at CERN's Large Hadron Collider (LHC). Traditionally, Monte Carlo-based simulations have been the primary tool for modeling these interactions, but their high computational cost presents significant challenges. Recently, generative machine learning models have emerged as an efficient alternative, offering the potential to drastically reduce simulation time while maintaining physical accuracy. Among these, diffusion models have demonstrated state-of-the-art performance in particle simulations, inspired by their success in computer vision. However, despite their fidelity, standard diffusion models suffer from prohibitively long generation times, limiting their practicality for High Energy Physics (HEP) applications.

To address this limitation, we turn to latent diffusion models, which accelerate generation by operating within a learned latent space rather than directly in pixel space. These models leverage powerful vision-based autoencoders to compress data, enabling significantly faster sampling without sacrificing quality. However, conventional latent diffusion models, such as Variational AutoEncoders (VAEs), impose arbitrary regularization constraints on the latent space—typically enforcing a Gaussian prior using KL-divergence. While suitable for standard computer vision tasks, such constraints can limit expressivity and accuracy when modeling complex HEP data distributions.

In this work, we propose a Sinkhorn-AutoEncoder Latent Diffusion Model (SAE-LDM), which improves upon traditional latent diffusion models by leveraging an end-to-end Sinkhorn AutoEncoder (SAE). Instead of imposing a predefined latent structure via KL-divergence, SAE directly minimizes the Wasserstein Distance between the encoded data distribution and a learned prior. This approach allows the model to better capture intricate data patterns without requiring a reparameterization trick, making it particularly well-suited for HEP simulations. By running the diffusion process within the SAE latent space, we ensure both efficient generation and high-fidelity reconstructions. Moreover, our framework jointly optimizes both the autoencoder and diffusion model using the Sinkhorn algorithm, leading to a more structured and expressive latent representation.

We evaluate our method on the task of simulating the response of the neutron Zero Degree Calorimeter (ZDC) in the ALICE experiment at CERN. Our model achieves a 50× speedup in generation time compared to Monte Carlo-based simulations, being orders of magnitude faster than pixel-space diffusion models and achieving inference speed comparable to Generative Adversarial Networks (GANs). At the same time, our approach achieves the highest simulation fidelity among all evaluated generative models on this task.

AI keywords

generative models, fast simulation, diffusion models, autoencoders, sinkhorn loss

Primary authors: Mr KITA, Mikołaj (Warsaw University of Technology); DUBIŃSKI, Jan (Warsaw University of Technology; IDEAS NCBR); ROKITA, Przemysław (Warsaw University of Technology); TRZCIŃSKI, Tomasz (Warsaw University of Technology; Tooploox); DEJA, Kamil (Warsaw University of Technology; IDEAS NCBR)

Presenter: DUBIŃSKI, Jan (Warsaw University of Technology; IDEAS NCBR)

Track Classification: Simulations & Generative Models