IMPERIAL

Diffusion Model For Beam Background Event Generation

Antonio Chahine, Haoyun Liu, Benedikt Maier

27/02/2026

Content

1. Motivation
2. Events Generation on Vertex Barrel Collection (aka old samples)
3. Events Generation on Monte Carlo Particles (aka new samples)
4. Discussion and Conclusion

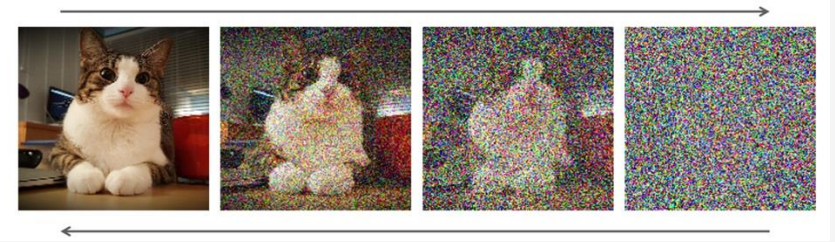
Motivation

Why do we need a generative surrogate?

- Beam-background studies require large event statistics for ML-based classification.
- Geant4/GUINEA-PIG simulation is computationally expensive → cost ~ 1 minute per event.
- Large-scale detector optimisation and systematic studies are limited by this simulation cost.

Principles of Diffusion Models:

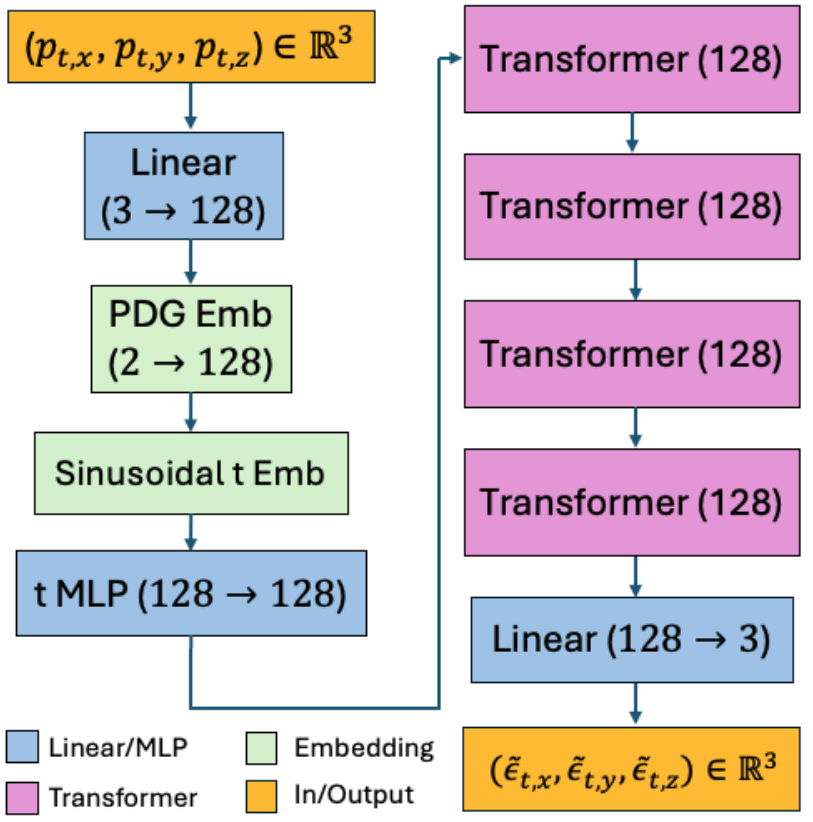
Diffusiuon models learn to model a probability distribution by reversing a gradual, predefined stochastic noising process.



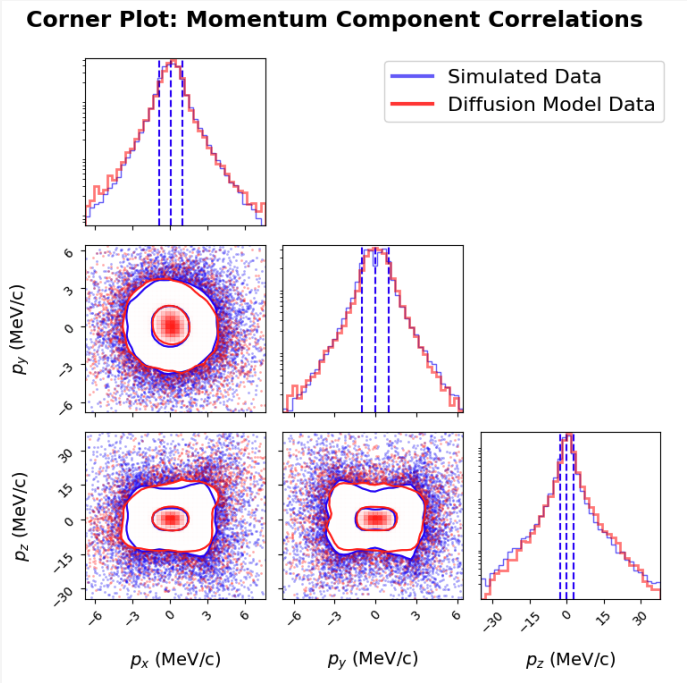
- Forward process corrupts clean data with Gaussian noise into pure noise with noise schedule: $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon$
- The denoiser is trained to approximate and predict the noise added at each time step t.
- Iterative reverse steps transform pure noise back into realistic samples by progreesively removing the predicted noise.

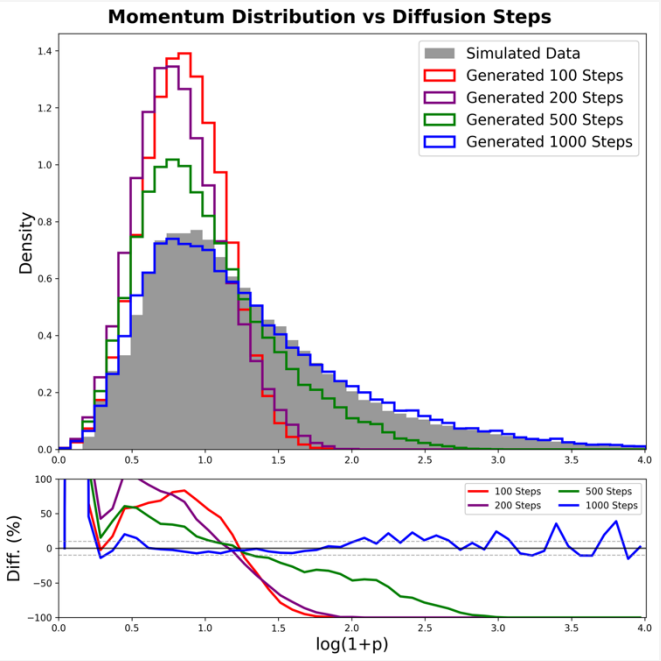
Vertex Barrel Collection

Momentum Denoiser Architecture



Metrics	Momentum Components		
	p_x	p_y	p_z
$D_{KL}(p_{SIM} \| p_{DM})$	0.0071	0.0319	0.0104
Wasserstein	0.0579	0.0878	0.4558





Metrics	Diffusion Steps			
	100	200	500	1000
ROC	0.7640	0.7277	0.6508	0.6063
ROC std	0.0052	0.0039	0.0015	0.0023

Adapting Diffusion to Monte Carlo Particle Data

Raw GUINEA-PIG output: $[E, \beta_x, \beta_y, \beta_z, x, y, z]$

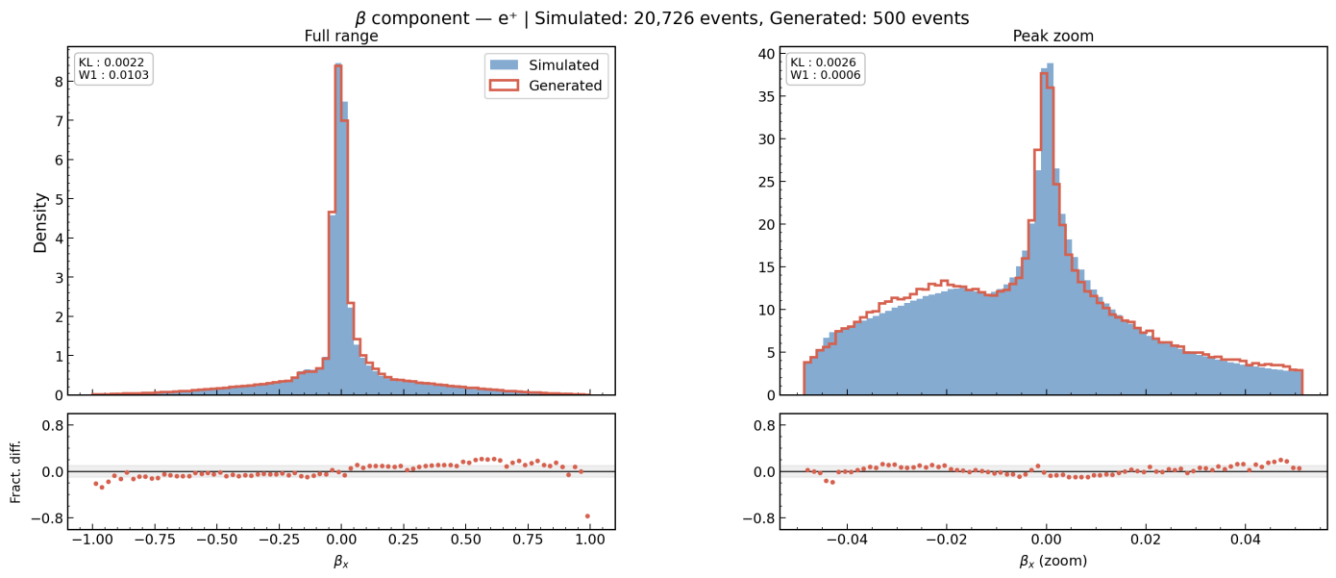
Model Representation: $[\log(|E|), u_x, u_y, u_z, x, y, z]$ + [PDG] conditioned on total multiplicity

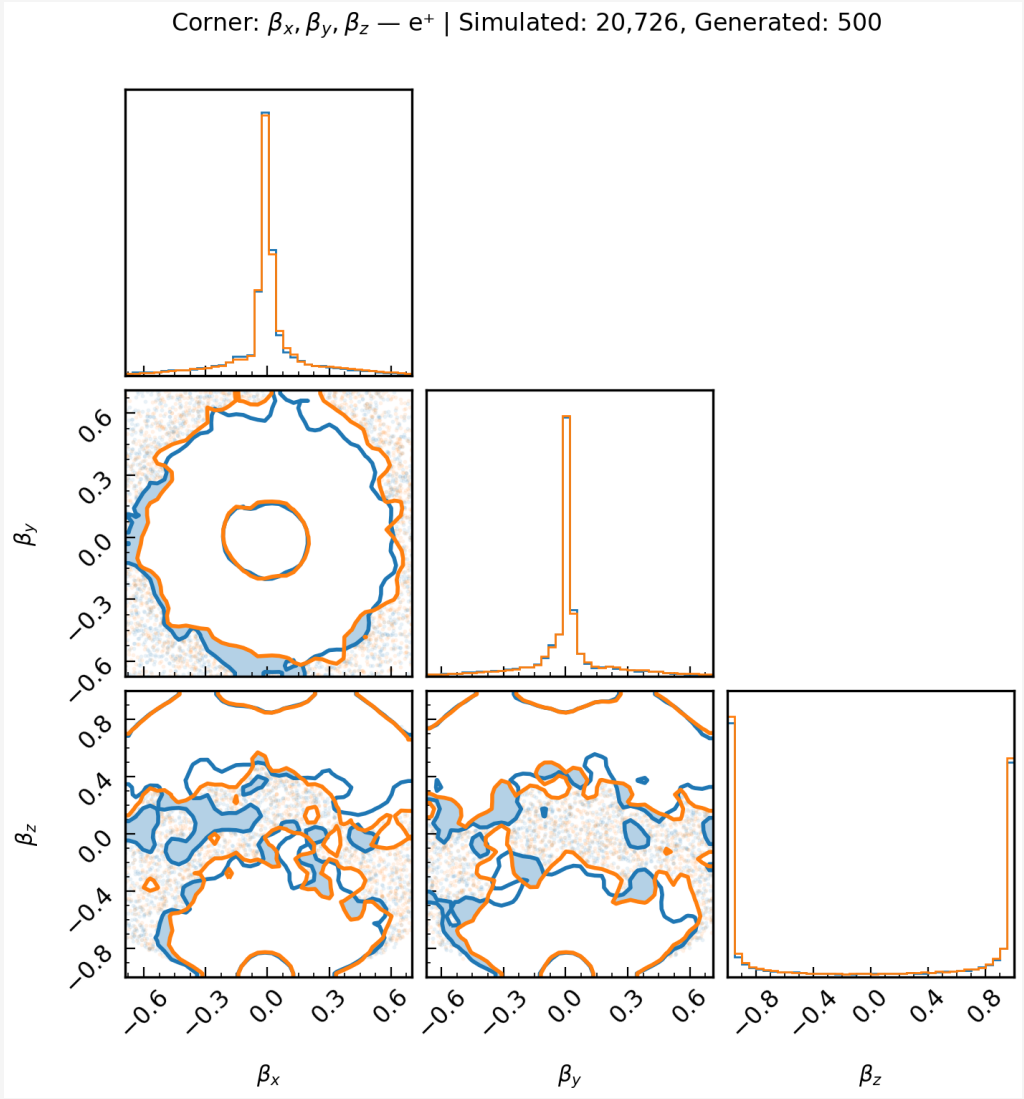
Physics-aware parameterisation:
$\beta = \tanh(|u|)$
- Guarantees $|\beta| < 1 \rightarrow$ physical outputs

Discrete PDG Diffusion:
- Gaussian noise unsuitable for labels
- Use categorical corruption schedule

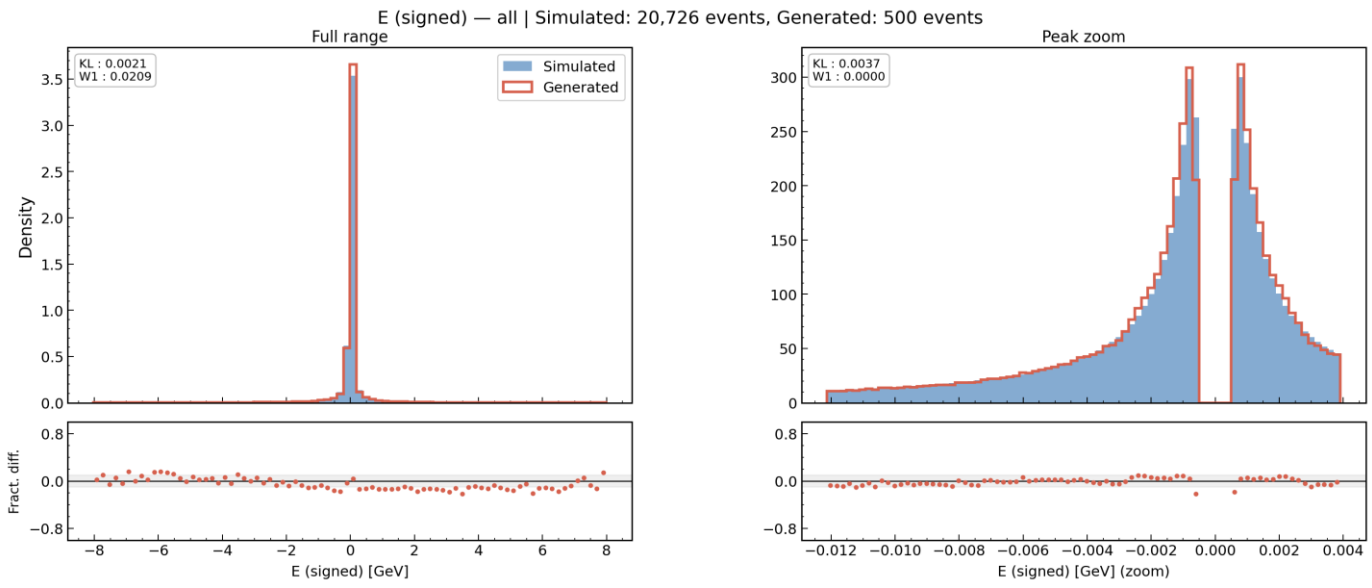
Event Generation: Sample from total multiplicity and apply reverse process for continuous + discrete variables

Monte Carlo Particles – Results

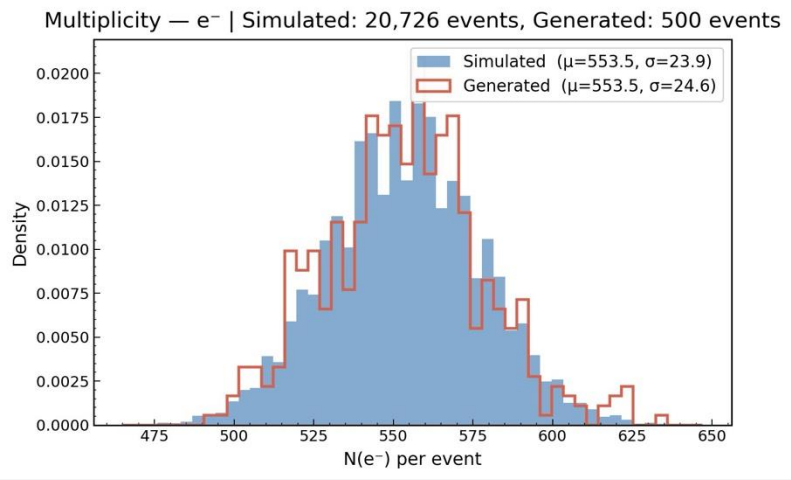


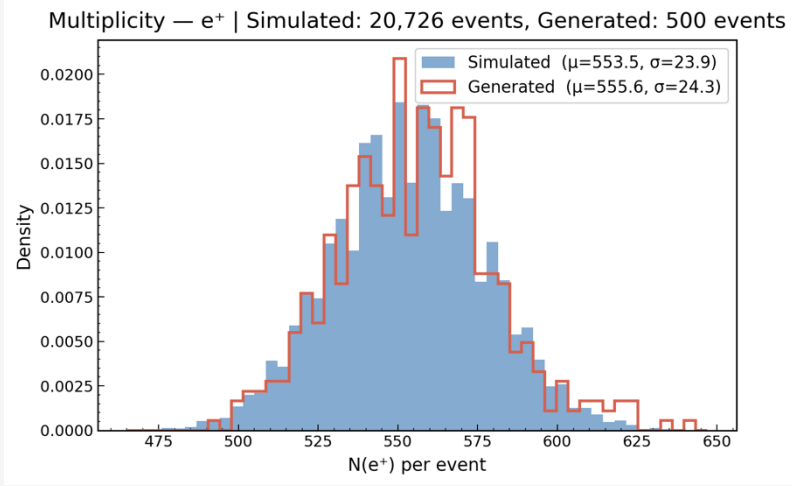


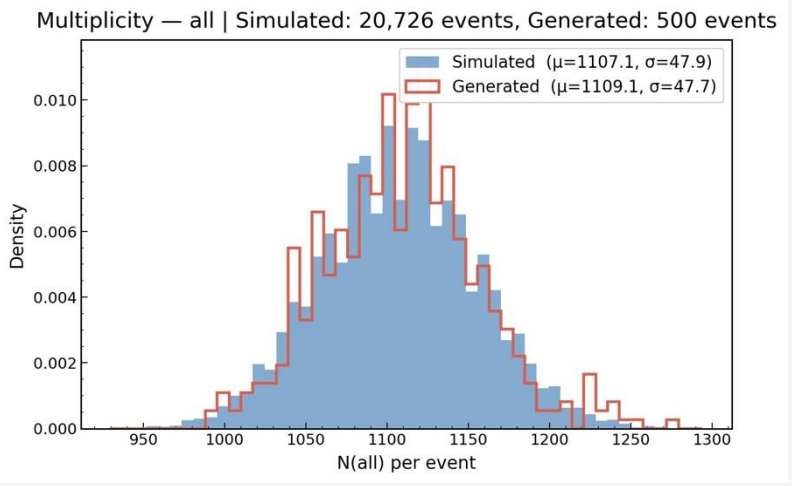
Monte Carlo Particles – Results



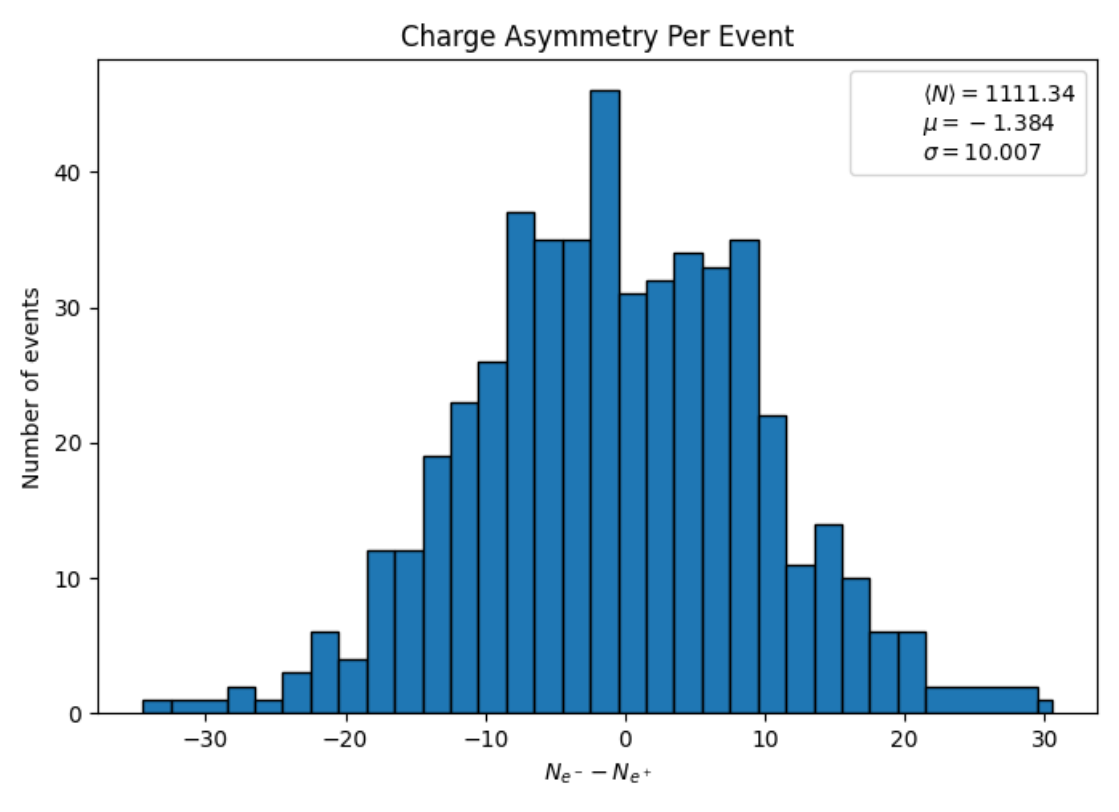
Metrics	Diffusion Steps			
	25	100	150	200
ROC	0.8441	0.6635	0.5979	0.5973
ROC std	0.0050	0.0055	0.0029	0.0033

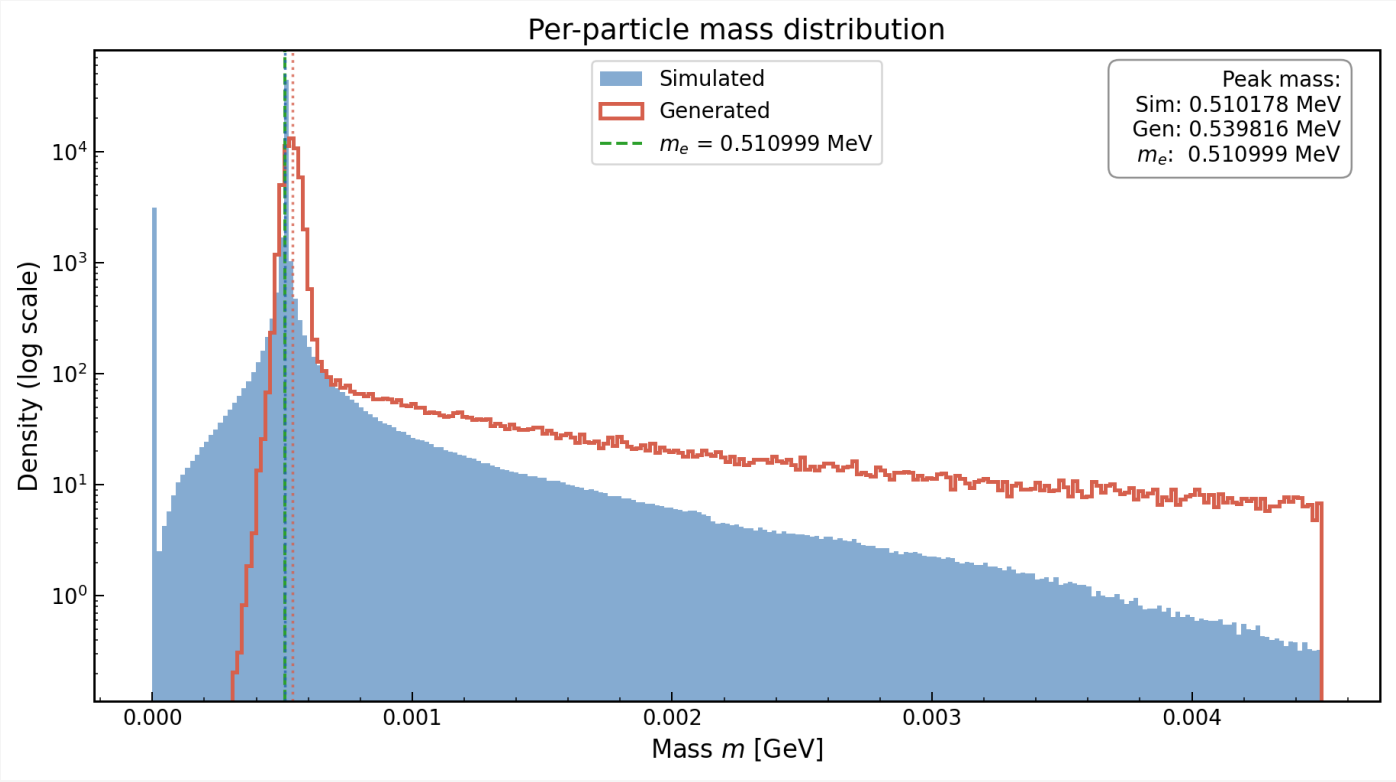






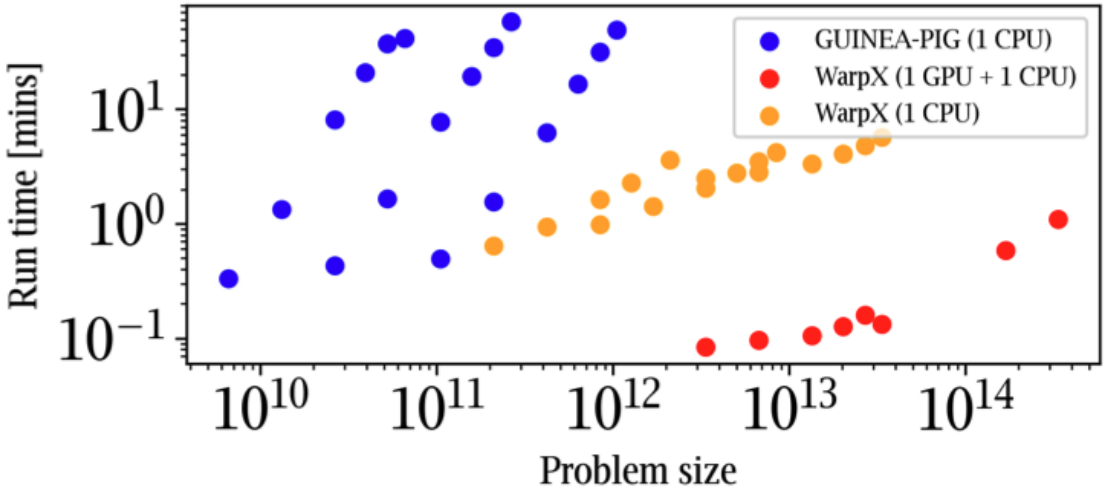
Checking the Physics



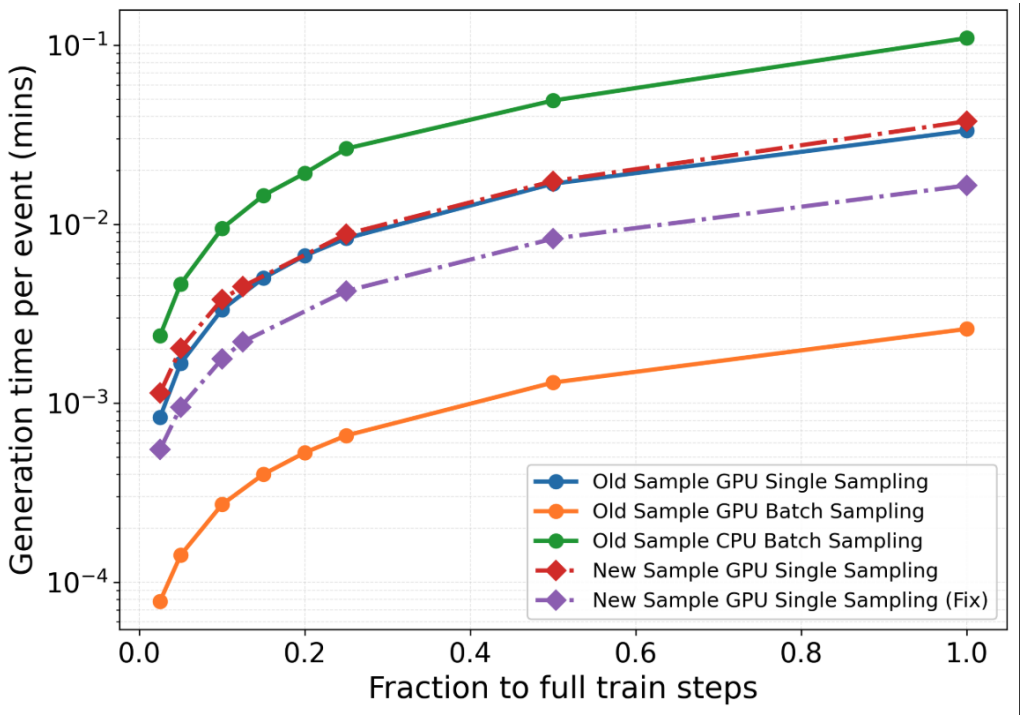


- The beam background should be symmetric in the number of e^+ and e^- produced and each particle should have an invariant mass of m_e
- Consider adding these constraints to the loss function

Time Comparison



Nguyen et al., *Comparison of WarpX and GUINEA-PIG for electron positron collisions*, arXiv:2405.09583 (2024)



The diffusion model generates background events on the SubMIT cluster using one NVIDIA A30 GPU, four CPU cores, and 16 GB of memory.

IMPERIAL

Thank you!

Antonio Chahine, Haoyun Liu, Benedikt Maier

27/02/2026