Fast Simulation of Incoherent Pair Backgrounds at the CEPC Using Conditional Flow Matching

Yanbang Tang · Haoyu Shi · Zhongjian Ma

Institute of High Energy Physics, Chinese Academy of Sciences, Beijing

135 ms/BX
A800 GPU generation

$\sim$270,000$\times$
vs. Guinea-Pig++ (1 CPU core)

2.8×10^{-5}
MMD on the test set

Physics-informed features: fast generation with close agreement to reference distributions.

01 Why a fast surrogate?

At $\sqrt{s} = 240$ GeV, CEPC beam–beam interactions produce $\sim$2,200 particles per bunch crossing (BX), with broad energy spectra and sharp forward peaks.

Bottleneck: Guinea-Pig++ takes $\sim$10 h/BX on one CPU core; detector studies need 10^4–10^5 BX.

Training data 5,000 BX; 10.8 million particles.
4,000 / 500 / 500 train / validation / test BX; split by BX to avoid leakage.

Processes: Breit–Wheeler 1.2%, Bethe–Heitler 18.3%, Landau–Lifshitz 80.5%.

02 Physics-informed features

Map 7 raw variables to 9 features, then standardize.

Input	Transform and purpose
Signed energy E	$\mathrm{sign}(E),\ \log_{10}\lvert E\rvert$ Compress five energy decades.
β_x, β_y	$\log_{10} p_T,\ \sin\phi,\ \cos\phi$ Separate scale and direction.
β_z	$\Phi^{-1}\left(\widehat{F}(\beta_z)\right)$ Resolve sharp peaks near ± 1.
Vertex x, y, z	Standardize each coordinate.

$\widehat{F}$: 50,000 training-data quantiles.
Φ^{-1}: standard-normal inverse CDF.

03 Process-conditional flow

Process c conditions the flow from noise u_0 to data u_1:

$$u_t = (1-t)u_0 + tu_1, \qquad t \sim \mathcal{U}(0,1)$$

$$\mathcal{L} = \mathbb{E}\left[\left\|v_\theta(u_t, t, c) - (u_1 - u_0)\right\|^2\right].$$

Network: 6×512 MLP with SiLU; 64D time and 64D process embeddings; 1.39M parameters.

Training: AdamW; 200 epochs; batch 4,096; $\sim$80 min on A800.

Generation: 100 midpoint ODE steps; sample joint process counts, generate particles, then invert preprocessing.

04 Speed and fidelity

Surrogate	MMD ↓	ms/BX
KDE	2.82×10^{-3}	6
Conditional FM	$\mathbf{2.8 \times 10^{-5}}$	**135**
Independent FMs	1.17×10^{-4}	219

Shared model: $\sim$101$\times$ lower MMD than KDE; 4.2$\times$ lower MMD and 1.6$\times$ faster than independent flows.

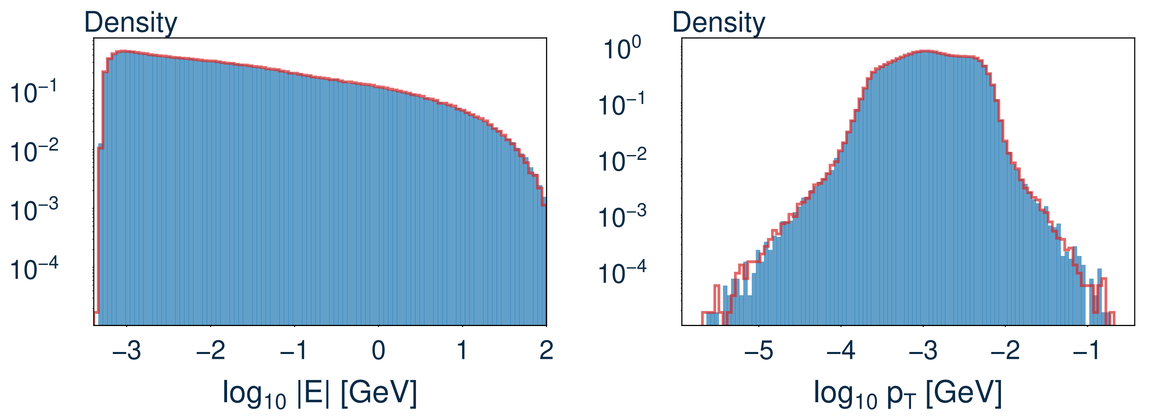
For 10^5 BX: $\sim$114 CPU-years of reference simulation $\rightarrow$ $\sim$3.75 GPU-hours.

Speedup: single-core CPU vs. A800 GPU; training excluded. FM on CPU: $\sim$4 s/BX.

05 Distribution fidelity

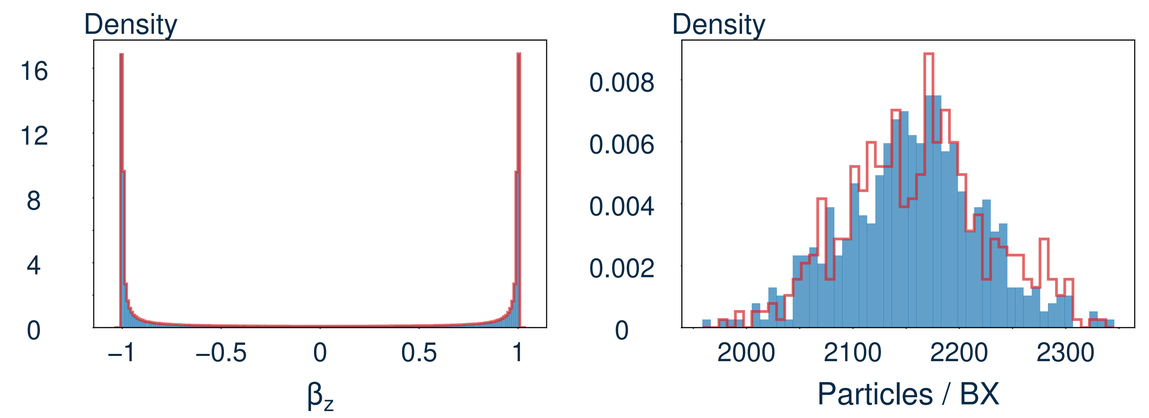
Guinea-Pig++ Conditional FM

Energy and transverse momentum



$|E|$: KS = 0.0009 p_T: KS = 0.0017

Longitudinal velocity and multiplicity



β_z: KS = 0.0132 Counts/BX: KS = 0.0500

Single-particle marginals: all nine reported KS values ≤ 0.0132; largest deviations are β_z (0.0132) and y (0.0117).

Particles/BX: reference 2156.9 ± 63.9; generated 2161.2 ± 68.2.
Multiplicity KS test: $p = 0.56$.

06 Preprocessing ablation

A 2×2 test: quantile normalization (QN) and p_T/ϕ decomposition.

KS	Full	No QN	No p_T/ϕ	Neither
β_z	0.013	0.056	0.014	0.056
p_T	0.002	0.001	0.008	0.006
θ	0.002	0.008	0.005	0.015

Both matter: QN resolves longitudinal peaks; p_T/ϕ preserves transverse structure.

07 Takeaway and next steps

Physics-informed features + one conditional flow: rapid CEPC pair-background generation with strong distribution fidelity.

Next validation targets

Beam-parameter conditioning; per-BX energy–momentum constraints; uncertainty quantification; Key4HEP / DD4hep integration.

Validated at one CEPC Higgs setting; broader beam settings and detector-level tests remain future work.

BX: bunch crossing FM: flow matching KS: Kolmogorov–Smirnov statistic MMD: maximum mean discrepancy