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

Yan-Bang Tang

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

April 2026

Outline

CEPC and Incoherent Pair Backgrounds

Circular Electron Positron Collider (CEPC)

- e^+e^- collider at $\sqrt{s} = 240\,\mathrm{GeV}$ (Higgs factory)
- Target: $> 5\,\mathrm{ab}^{-1}$, $> 10^6$ Higgs events
- Precision measurements of Higgs couplings and EW observables

Beam–beam interaction produces pair backgrounds:

- **$\sim$2157 particles per BX**
- Energy: $0.5\,\mathrm{MeV}$ to $100\,\mathrm{GeV}$ (5 decades)
- Strongly forward-peaked angular distribution
- Critical background for detector design and physics analysis

Three QED Processes

- **Breit-Wheeler:** $\gamma\gamma \to e^+e^-$ 1.2%
- **Bethe-Heitler:** $e\gamma \to ee^+e^-$ 18.3%
- **Landau-Lifshitz:** $ee \to eee^+e^-$ 80.5%

Key beam parameters:

- $E_{\mathrm{beam}} = 120\,\mathrm{GeV}$
- $N_b = 1.3 \times 10^{11}$
- $\sigma_x/\sigma_y = 14000/36\,\mathrm{nm}$
- Crossing angle: $33\,\mathrm{mrad}$

The Simulation Bottleneck

Guinea-Pig++: the standard simulation tool

- Particle-in-cell method with full QED physics
- Provides 4-momentum + vertex for every particle
- Used for ILC, CLIC, FCC-ee, CEPC design studies

Our Goal

Train a generative model that **learns** the 7D phase-space distribution from Guinea-Pig++ data and generates new samples **instantly**

The Problem

- $\sim$**10 hours per BX** on a single CPU core
- Detector studies need 10^4–10^5 BX
- 10^5 BX $\times$ 10 h = **114 CPU-years**
- Infeasible for iterative design optimization

Guinea-Pig++
($\sim$10 h/BX)

↓ Train once

Flow Matching
($\sim$**0.13 s/BX**)

↓

$\sim$**270,000**$\times$ **faster**

From Prior Work to the Research Gap

Generative ML successes in HEP:

- **Calorimeters:** CaloGAN, CaloFlow, CaloDiffusion
- **Full detector:** FlashSim
- **Event generation:** diffusion, normalizing flows

Flow Matching — state of the art:

- Stable, simulation-free CNF training
- Optimal Transport CFM: direct linear paths
- Proven on HEP tabular data

Research Gap

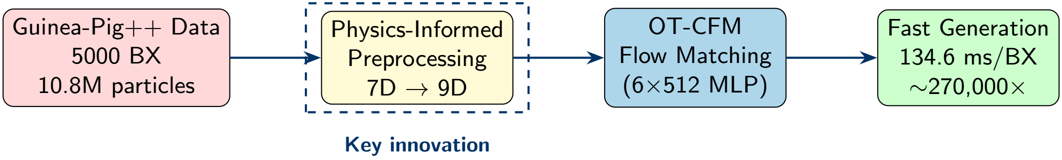
No prior work has applied generative ML to beam–beam pair background simulation

— for CEPC or any lepton collider.

This Work

First application of deep generative models to incoherent pair background fast simulation

Approach Overview



Key Innovation

Physics-informed feature engineering transforms extreme distributions into learnable representations

Model Design

Single conditional model for all 3 processes, enabling cross-process information sharing

Practical Impact

100,000 BX: from 114 CPU-years to 3.7 GPU-hours

Dataset: Guinea-Pig++ CEPC Simulation

Data generation:

- 5000 independent BX, unique random seeds
- Total: $\mathbf{10.8 \times 10^6}$ particles
- Exact charge symmetry: 50% e^-, 50% e^+

Per-BX statistics:

- Multiplicity: 2157 ± 64 particles
- Range: [1894, 2398]

Data split (by BX, no leakage):

4000 train / 500 val / 500 test

Raw features (7D + label):

Variable	Unit	Range
Energy E	GeV	$[\pm 0.5\,\text{MeV}, \pm 100\,\text{GeV}]$
β_x, β_y, β_z	—	$(-1, 1)$
x, y, z	nm	$\pm 10^7$
Process	0/1/2	BW/BH/LL

Challenges for ML

- Energy spans 5 orders of magnitude
- β_z concentrates at ± 1
- Extreme beam aspect ratio ($\sigma_y/\sigma_x \approx 1/400$)

Physics-Informed Preprocessing

Core idea: transform raw kinematics into physically motivated, ML-friendly features

Raw variable	Transform	Rationale
Energy E	$\mathrm{sign}(E),\ \log_{10}\lvert E\rvert$	5 decades
β_x, β_y	$\log_{10}(p_T)$	Heavy tail
	$\sin\phi,\ \cos\phi$	Circular topology
β_z	$\Phi^{-1}(\hat{F}(\beta_z))$	Density at ± 1
$x,\ y,\ z$	$(\cdot - \mu)/\sigma$	Unit variance

$\Rightarrow$ 7D raw features become **9D physics-informed features**
plus global standardization

Quantile Normalization for β_z

$$\beta_z \xrightarrow{\hat{F}} [0,1] \xrightarrow{\Phi^{-1}} \mathbb{R}$$

- 50,000 quantile points from training data
- Avoids artanh divergence at $\beta_z = \pm 1$
- Rank-based $\Rightarrow$ robust to extreme values
- **Ablation:** removing QN degrades β_z KS by $4\times$

p_T/ϕ Decomposition

$(\beta_x, \beta_y) \to (\log_{10} p_T,\ \sin\phi,\ \cos\phi)$

Decouples momentum scale from direction

Ablation: removing this degrades p_T KS by $9\times$

OT-CFM Flow Matching Model

Optimal Transport Conditional Flow Matching

- Learn velocity field v_θ transporting $\mathcal{N}(0, I)$ to data
- **Linear interpolation path:**

$$x_t = (1-t)\, x_0 + t\, x_1$$

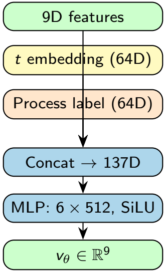
- **Training loss (MSE):**

$$\mathcal{L} = \mathbb{E}_{t, x_0, x_1, c}\left[\left\| v_\theta(x_t, t, c) - (x_1 - x_0) \right\|^2\right]$$

- **Generation:** solve ODE from $t{=}0$ to $t{=}1$ using 100 midpoint (RK2) steps

Simulation-free training; no MCMC, no adversarial loss

Architecture



Parameters	1.39M
Optimizer	AdamW (lr 10^{-3})
Epochs	200
GPU	NVIDIA A800
Training time	∼80 min

Process-Conditional Generation Strategy

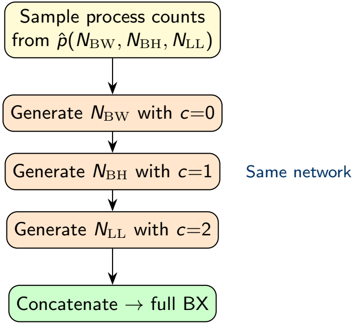
Single conditional model for all 3 processes

- Process label → 64D learned embedding
- Network learns shared & process-specific features
- **Advantages:**
  - Parameter-efficient (one model vs. three)
  - Cross-process physics learning
  - No data starvation for rare processes (BW: 1.2%)

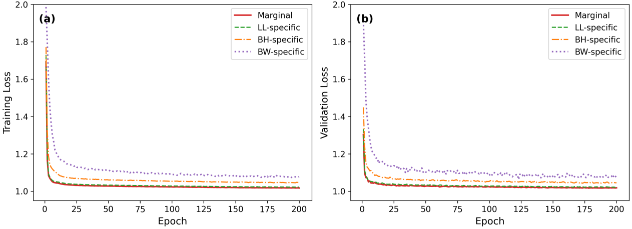
Alternative tested:

- **Combined FM**: 3 independent per-process models
- → comparison in Results

BX generation pipeline:



Training Convergence



Stable convergence, no overfitting

- All 4 model variants shown
- Validation loss tracks training loss
- Best val. loss $\approx$ 1.017 (epoch 178)
- EMA (decay 0.999) for generation

Property	Value
Architecture	6×512 MLP, SiLU
Parameters	1.39M
Batch size	4096
LR schedule	Cosine + 10-ep warmup
GPU	NVIDIA A800 80 GB
Wall time	$\sim$80 min

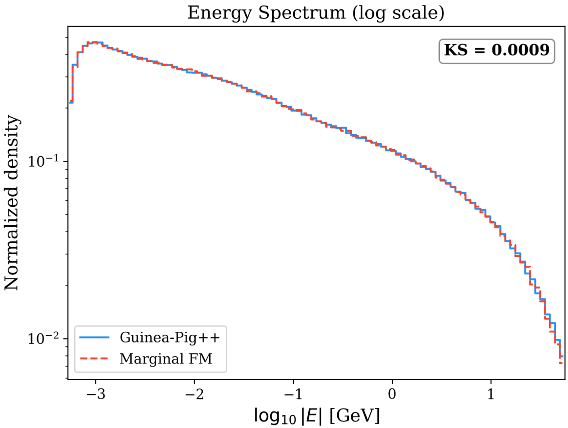
Comprehensive Comparison: Three Surrogate Methods

Variable	KDE Baseline	**Marginal FM**	Combined FM
$\|E\|$	0.030	**0.0009**	0.0017
β_x	0.139	**0.0009**	0.0011
β_y	0.170	**0.0017**	0.0018
β_z	0.024	**0.0138**	0.0145
x	0.094	**0.0009**	0.0009
y	0.156	**0.0075**	0.0084
z	0.055	**0.0014**	0.0015
p_T	0.297	**0.0016**	0.0017
θ	0.308	**0.0021**	0.0025
ϕ	—	**0.0011**	0.0015
r	—	**0.0009**	0.0014
MMD	2.82×10^{-3}	5.5×10^{-5}	1.46×10^{-4}
Speed (ms/BX)	9.3	134.6	219.1

Marginal FM: MMD 51× better than KDE

All KS values < 0.014 (Marginal FM)

Distribution Fidelity: Energy Spectrum



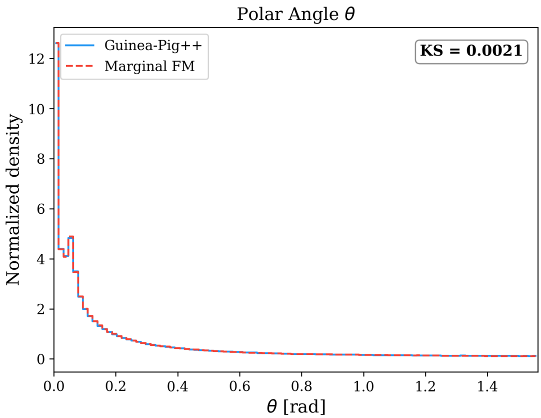
Log-scale energy spectrum

- Covers **5 orders of magnitude** (0.5 MeV–100 GeV)
- Generated and real distributions nearly indistinguishable
- KS statistic: $|E|$ KS $= 0.0009$

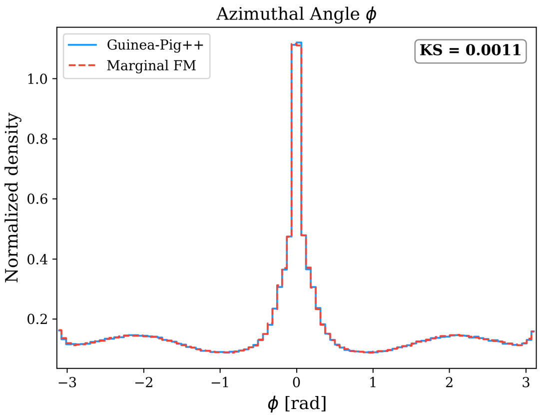
Why it works

$\log_{10}|E|$ transform compresses 5 decades into a compact range, making the distribution learnable

Distribution Fidelity: Angular Structure



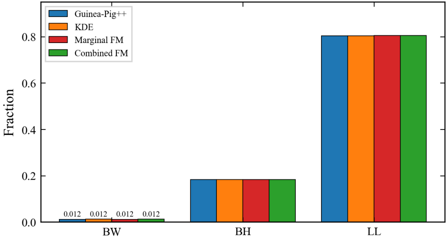
Polar angle θ: KS $= 0.0021$



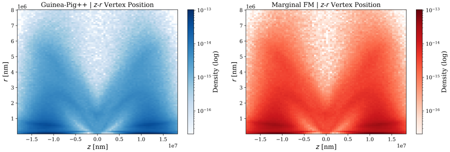
Azimuthal angle ϕ: KS $= 0.0011$

- θ: strong forward peak near beam axis
- Model precisely captures the peaked structure
- ϕ: central peak + flat wings (beam geometry)
- $(\sin\phi, \cos\phi)$ encoding avoids $\pm\pi$ discontinuity

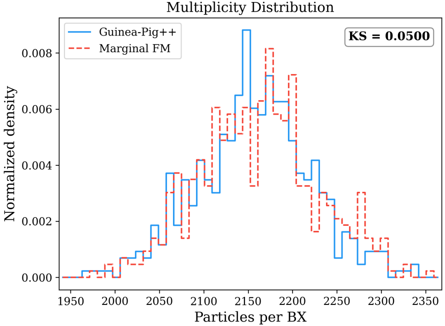
Structural Fidelity: Process Fractions & Multiplicity



Process fraction errors $< 0.1\%$ (abs.)



z–r vertex structure reproduced



Multiplicity reproduction:

- Real: 2156.9 ± 63.9
- Generated: 2161.2 ± 68.2
- KS $= 0.050$, $p = 0.56$

z–r: beam interaction geometry correctly captured

Marginal FM vs. Combined FM: Why One Model Wins

Two strategies tested:

	Marginal FM	Combined FM
Models	1 shared	3 independent
Total params	1.39M	variable
Conditioning	Process label	None

Marginal FM wins across the board:

- MMD: 5.5×10^{-5} vs. 1.46×10^{-4}
  $\Rightarrow$ 2.7$\times$ better
- Speed: 134.6 vs. 219.1 ms/BX
  $\Rightarrow$ 1.6$\times$ faster
- Simpler training and deployment

Key Insight

The single conditional model benefits from **cross-process information sharing**:

- Shared spatial distributions
- Shared energy–velocity correlations
- Process-specific kinematics captured by conditional embedding

Practical Advantage

One model to train, validate, and deploy
vs. three separate pipelines

Preprocessing Ablation: Every Component Matters

2×2 **factorial design:**
Quantile Normalization $\times$ p_T/ϕ decomposition

	Full	$-$QN	$-p_T/\phi$	Naive
QN(β_z)	✓		✓	
p_T/ϕ	✓	✓		
KS(β_z)	0.014	0.056 (4×)	0.014	0.056 (4×)
KS(β_x)	0.001	0.001	0.004 (4×)	0.004 (3×)
KS(p_T)	0.001	0.001	0.008 (9×)	0.006 (7×)
KS(θ)	0.002	0.008 (4×)	0.005 (3×)	0.015 (8×)

Findings:

1. **QN is essential for β_z**
   Without it: β_z KS degrades $4\times$
   artanh alternative diverges at ± 1
2. **p_T/ϕ decomposition is essential**
   Without it: p_T KS degrades $9\times$
   Decoupling magnitude from direction helps
3. **Both removed (Naive):**
   Severe degradation on all variables
   θ KS: $0.002 \to 0.015$ ($8\times$)
   MMD becomes undefined (NaN)

$\Rightarrow$ **Physics-informed preprocessing is the key enabler, not merely helpful**

Generation Speed: The Headline Result

Guinea-Pig++		Flow Matching
$\sim$**10 hours / BX**	$\sim$**270,000**$\times$ →	$\sim$**134.6 ms / BX**
(single CPU core)		(GPU inference)

For 100,000 BX dataset:

Guinea-Pig++	$\sim$114 CPU-years
Flow Matching	$\sim$**3.7 GPU-hours**

What this enables:

- Rapid detector geometry optimization
- Large-scale background overlay studies
- Interactive design–evaluate cycles

Conclusions

1. **First deep generative model for beam–beam pair backgrounds**
   Novel application domain: no prior work for CEPC or any lepton collider
2. **Physics-informed preprocessing is the key enabler**
   Quantile normalization for β_z and p_T/ϕ decomposition
   are essential — validated by ablation study
3. **High fidelity across all distributions**
   $\mathrm{MMD} = 5.5 \times 10^{-5}$ ($51\times$ better than KDE baseline)
   $|E|$ KS $= 0.0009$, θ KS $= 0.0021$, ϕ KS $= 0.0011$
4. **$\sim$270,000$\times$ speedup over Guinea-Pig++**
   100,000 BX: 114 CPU-years $\rightarrow$ 3.7 GPU-hours

Flow matching provides a practical, high-fidelity surrogate for beam–beam pair background simulation at future e^+e^- colliders.

Outlook & Future Directions

Near-term improvements:

- **Beam parameter conditioning**
  One model covering multiple $\sqrt{s}$ and luminosity settings
- **Flow matching distillation**
  Reduce ODE steps from 100 to $\sim$10
  $\Rightarrow$ additional 10$\times$ speedup
- **Energy–momentum conservation**
  Enforce per-BX balance constraints

Deployment & generalization:

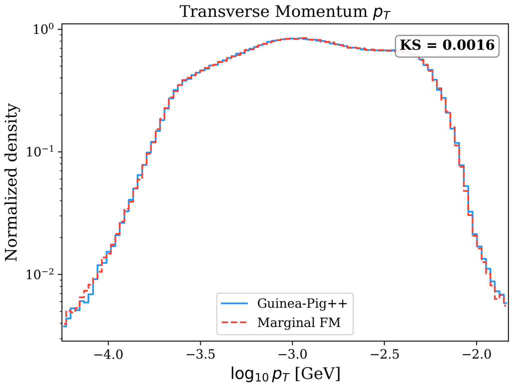
- **Key4HEP / DD4hep integration**
  Plugin for seamless Guinea-Pig++ replacement in the detector simulation framework
- **Other colliders**
  FCC-ee, CLIC, ILC — same methodology, different beam parameters
- **Uncertainty quantification**
  Ensemble or Bayesian approaches for per-sample confidence estimates

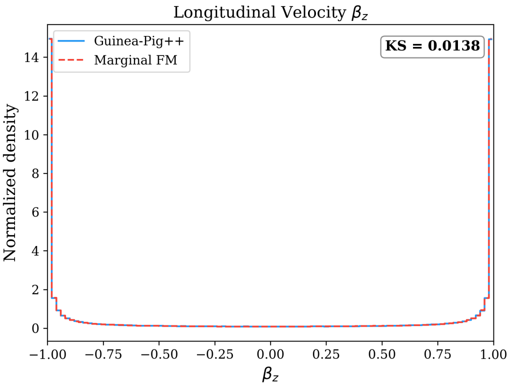
Thank You

Questions & Discussion

Yan-Bang Tang · IHEP, Chinese Academy of Sciences

Backup: Momentum & Velocity Distributions



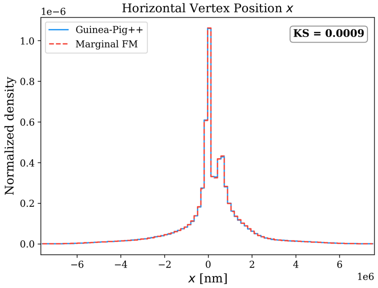


p_T distribution (log-scale): KS $= 0.0016$

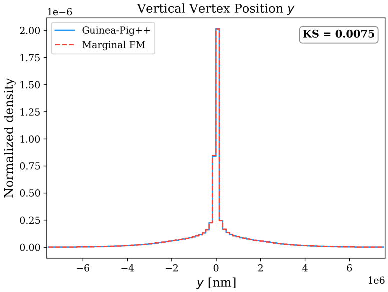
β_z distribution: KS $= 0.0138$

- p_T spans several orders of magnitude; $\log_{10} p_T$ transform enables accurate learning
- β_z concentrates near ± 1; quantile normalization maps to standard normal

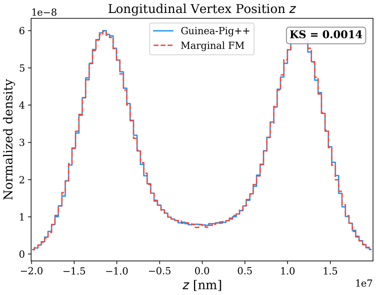
Backup: Vertex Position Distributions



x: KS $= 0.0009$



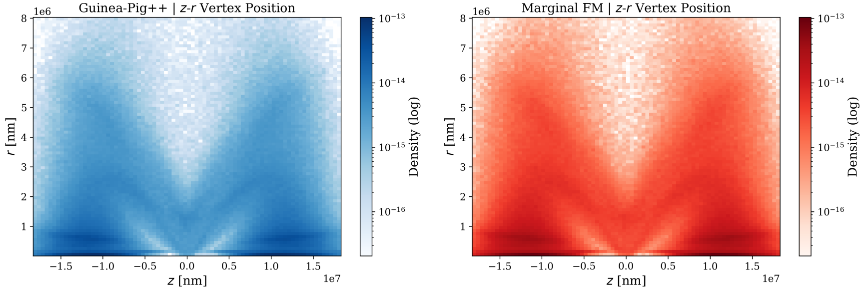
y: KS $= 0.0075$



z: KS $= 0.0014$

- y distribution is $\sim$400$\times$ narrower than x ($\sigma_y/\sigma_x \approx 1/400$)
- z shows double-peaked structure from beam interaction geometry

Backup: z–r Vertex Structure



Two-dimensional vertex distribution in (z, r) with logarithmic density scale.
The spatial extent and density structure of the beam interaction region are reproduced by the Marginal FM model.