Fast Simulation of the CEPC Long-Bar Crystal Electromagnetic Calorimeter

Li Zhihao [1 2] Lin Tao [1 2] Li Weidong [1 2]

[1]Institute of High Energy Physics [2]University of Chinese Academy of Sciences

Introduction

- **CEPC** pursues staged Higgs, Z, and top programs that demand extremely high electroweak precision and sensitivity to rare decays, driving unprecedented detector-model fidelity and simulated data volumes.[1]
- **ECAL** The long-bar crystal ECAL forms an interleaved three-dimensional mesh that delivers about 3% energy resolution with fine imaging. [1]
- **Simulation Framework** The simulation workflow includes physics event generation, Geant4 based detector simulation, and digitization. The framework adopts a Gaudi-based modular architecture that integrates Geant4 . It is designed with pluggability, allowing the fast-simulation (FastSim) module to be selectively activated when specific particles enter the ECAL.

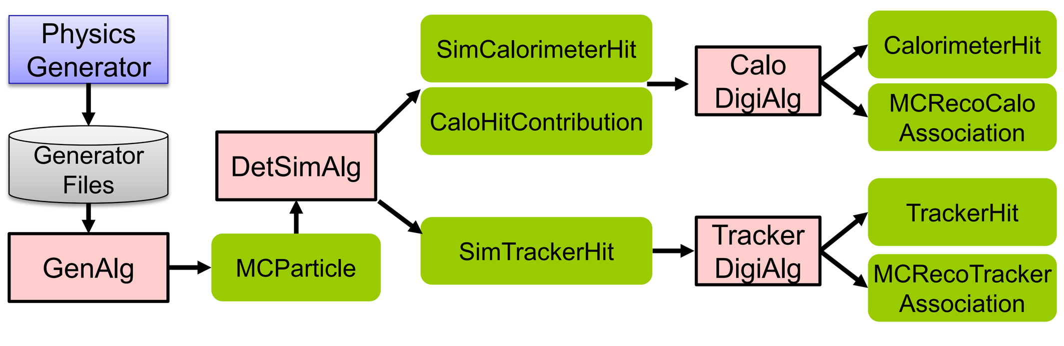

Figure 1. Overview of the CEPC Full Simulation Workflow.

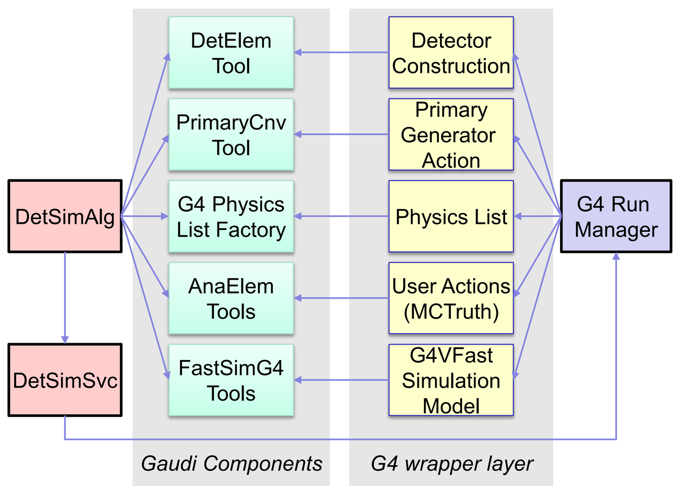

Figure 2. Architecture of the CEPC Simulation Framework and Fast Simulation Integration Points.

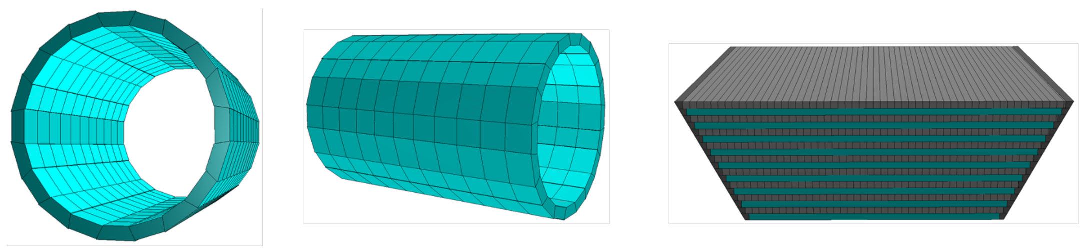

Figure 3. Geometry of the CEPC long-bar crystal ECAL Barrel.

Computing Bottlenecks

- **Classical Computing Resource Bottleneck:** The CEPC full-simulation program is expected to produce about 10^{11} events per year, yielding 284 PB of full-simulation data and consuming roughly 1570 kHS23-yr of CPU computing time annually.[1]
- **Geant4 and ECAL Simulation Bottlenecks** For $e^+e^- \to q\bar{q}$ at 240 GeV the Geant4 step dominates 46.4% of the wall time; ECAL barrel and endcaps alone consume 34.4% and 13.8% of the total simulation budget, respectively.
- **Digitization and Optical Photon Challenge** Digitization currently relies on semi-empirical parameterizations that already take 25% of the workflow; introducing optical photon tracking based on Geant4 would be prohibitive, so Opticks-based ground truth is needed for training alternative models.[2, 4]

Run stage	MC events	CPU time (kHS23-yr)	MC data (PB)
Higgs	2.0×10^9	105	4.00
Low Z	1.4×10^{11}	1465	280.00

Table 1. Projected CEPC **annual** simulation throughput and storage requirements by run stage.

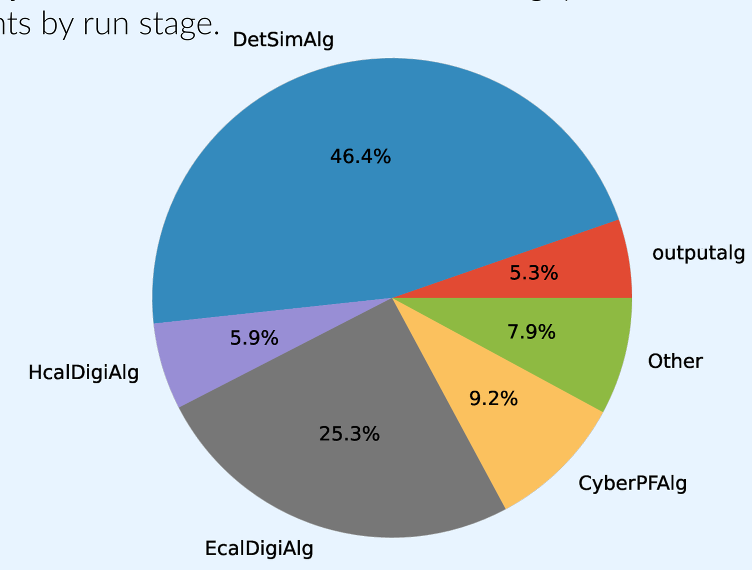

Figure 4. Breakdown of CPU Time Consumption in CEPC Full Simulation

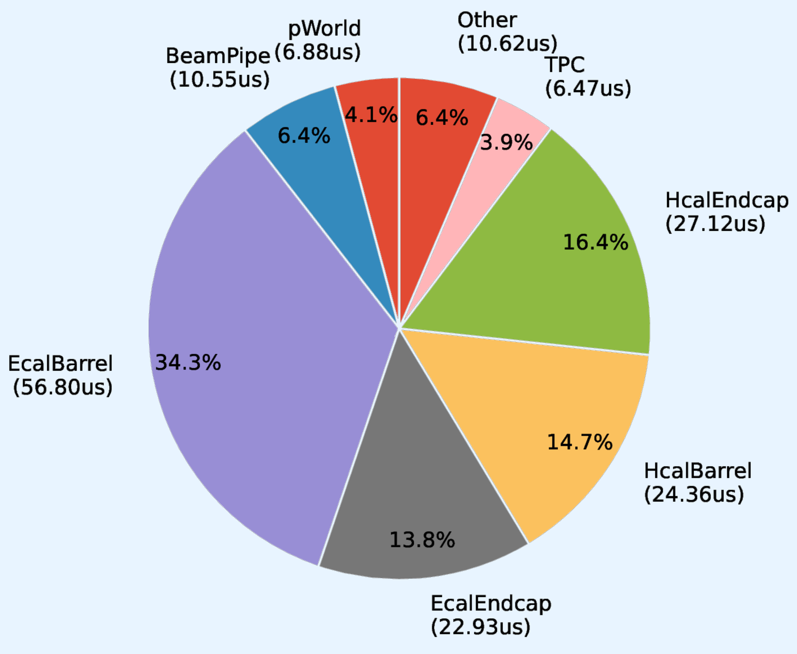

Figure 5. CEPC Geant4 Subdetector Time Profile for $e^+e^- \to q\bar{q}$ at 240 GeV

Fast Simulation and Digitization Flow

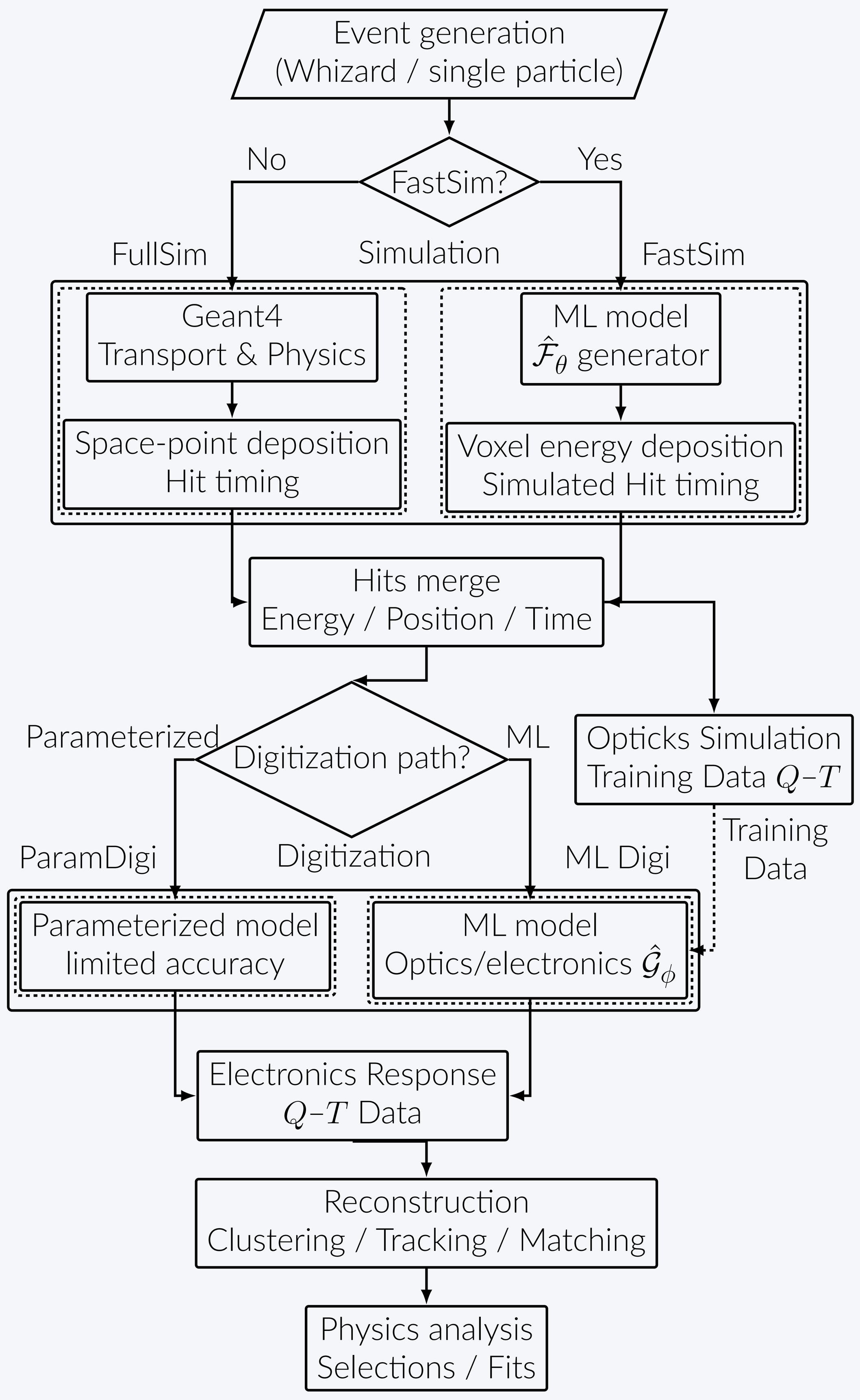

Figure 6. Workflow of CEPC Fast Simulation and Digitization.

Methodology

Fast Simulation

- **Geant4 Data (Baseline):** Collect hits from full `DD4hep` [6] geometry and Geant4 physics, including ID, position, energy, and time.
- **Voxelization:** Convert hits into a $15 \times 15 \times 18$ energy grid and map between global and local coordinates.
- **Dataset:** Combine voxel energies and conditions with normalized phase-space coverage.
- **Model:** Generate $p(V|c)$ under inductive priors (non-negativity, energy conservation, symmetry); export via ONNX [5].
- **Evaluation:** Validate consistency (energy spectra, profiles, EMD, closure) and usability via end-to-end reconstruction.

Figure 7. Fast-simulation data preparation pipeline from Geant4 hits to evaluation metrics.

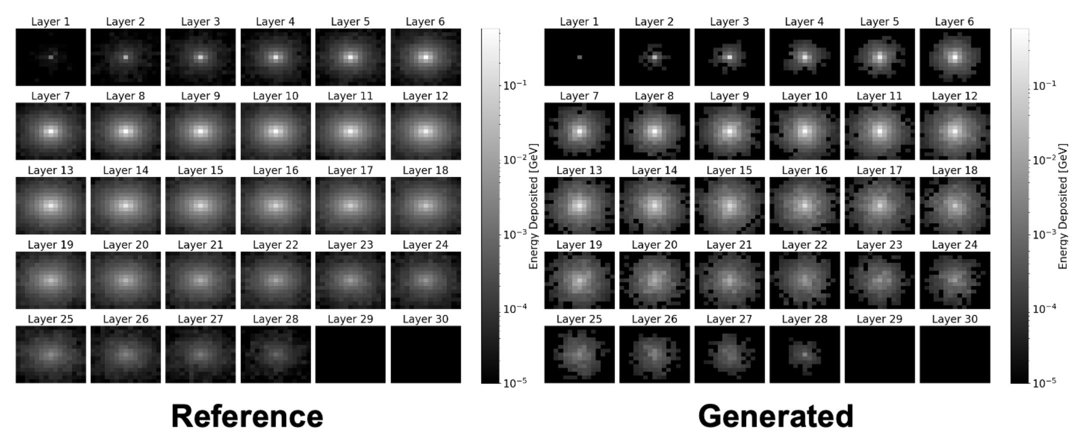

Figure 8. Generative-model shower samples contrasted with reference voxelized energy deposits.

Opticks-Driven Digitization

- **Opticks Data (High-fidelity optical model):** Use `Opticks` [3] + `ElecSim` for per-crystal optical and electronics simulation, generating dual-end readouts $(Q_1, T_1; Q_2, T_2)$ as reference.
- **Feature engineering:** Derive spatio-temporal, geometric, and electronics features (e.g. crystal length, refractive index, noise RMS) with symmetry and normalization preprocessing.
- **Dataset:** Pair features X with optical outputs $R = (Q_1, T_1; Q_2, T_2)$; apply stratified sampling over energy, geometry, and noise conditions.
- **Model:** Train supervised model $\hat{\mathcal{G}}_\phi$ to approximate $\mathcal{G}_{\text{Opticks+elec}}$ using lightweight Transformer or conditional flow networks.
- **Evaluation:** Benchmark against full simulation; check Q, T resolution, $T_1 - T_2$ vs. position linearity, energy resolution, and robustness to noise or drift.

Figure 9. Opticks-driven digitization pipeline from optical truth to evaluation metrics.

Research Objectives

- **Fast ECAL simulation:** Develop ML-based fast simulation reproducing ECAL shower response and morphology with 10–$100\times$ speedup.
- **ML digitization based on Opticks data** Build fast digitization model based on Opticks-generated data which predicting $(Q_{1,2}, T_{1,2})$ with $< 5\%$ residuals and $> 20\times$ acceleration.
- **CEPCSW integration:** Integrate models into CEPCSW via ONNX Runtime for seamless fast/full mode switching.[5]
- **Quantum ML exploration:** Study quantum ML prototypes for improved sample efficiency and lower latency.

Research Progress

Completed investigations

- **Reversible voxelization:** Implemented the bidirectional mapping between DD4hep cellIDs and local (l_x, l_y, l_z) grids with $\delta_\phi = \delta_z = 15$ mm and parity-aware $\delta_r \in \{15, 15.6\}$ mm. The logic swaps axes per `slayer`, wraps modules/staves, clones GAP hits, and enforces energy-sum conservation so ML tensors remain losslessly reversible to native hits.
- **Curriculum photon dataset:** Built the γ dataset (module 0, $\theta = \phi = 90^\circ$) that mirrors the ECAL incident composition/time correlations: $E_k = 256 \cdot 2^{k/2}\,\mathrm{MeV}$ spans 256 MeV–131 GeV with curriculum counts $N(E) = 10^4$ (low-E) down to 10^3 at the tail, totaling $\sim 1.21 \times 10^5$ samples and packaging inputs $(x_0, y_0, z_0, E_{\text{inc}}, \theta, \phi)$ with voxelized outputs.
- **Conditional WGAN-GP baselines:** Deployed the gradient-penalized Wasserstein GAN as the first $p(V \mid c)$ generator.

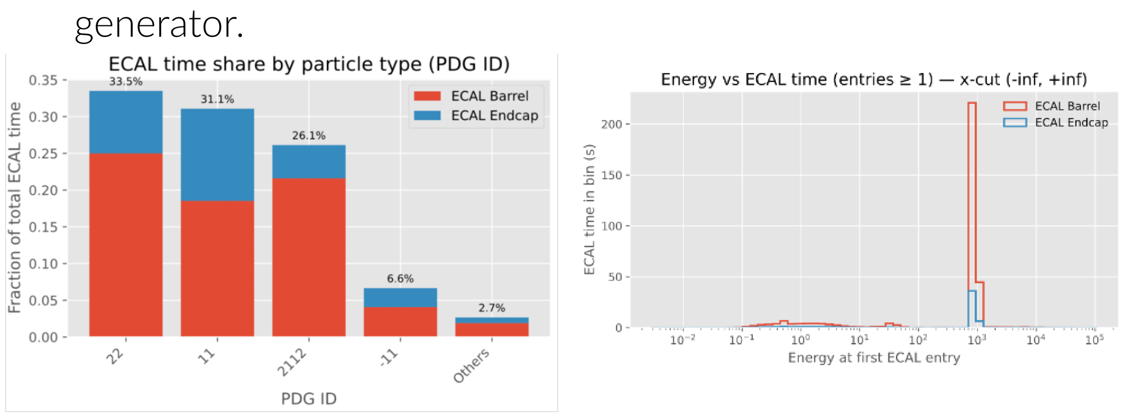

Figure 10. FastSim timing diagnostics: ECAL hit time versus PDGID (left) and incident energy versus ECAL time (right).

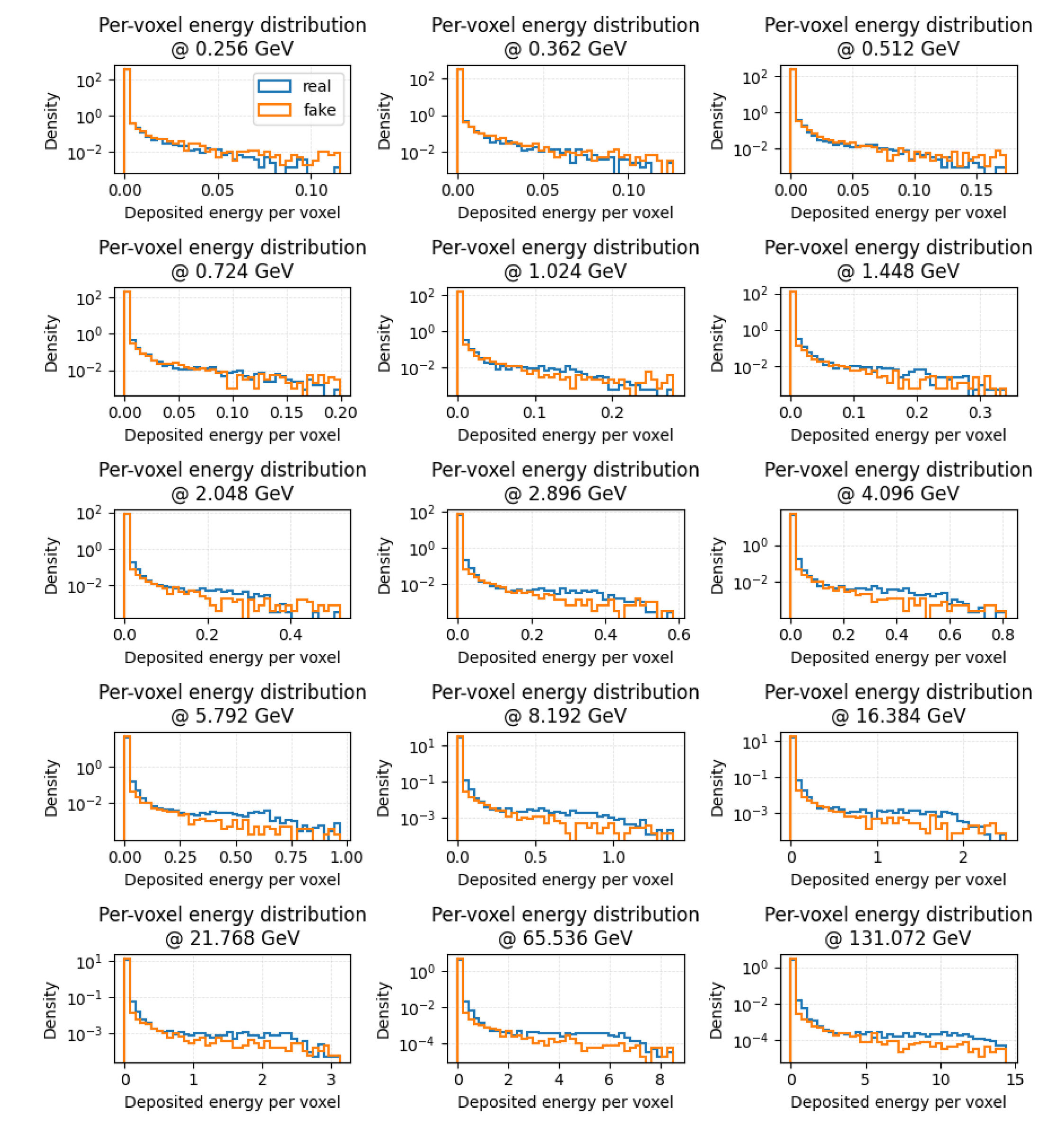

Figure 11. Voxel-level energy spectra from FastSim showers versus Geant4 ground truth for the γ dataset.

Ongoing developments

- **Unified conditioning & capacity control:** Current GAN models can already align the relative voxel-wise energy distributions, but are still under active optimization to reproduce absolute energy distributions and maintain total-energy consistency. We are preparing CNN and flow-based models with stronger energy constraints.
- **Opticks-driven digitization:** Opticks GPU photon transport supplies dual-end Q–T ground truth, and turn it into an FastDigi ML model.[2, 4]
- **CEPCSW integration:** The FastSim/FastDigi pairs are being wrapped in ONNX Runtime execution providers so mixed fast/full pipelines can be profiled inside CEPCSW, targeting $10\times$ ECAL speedup while keeping Higgs recoil mass shifts below 0.1%.
- **Quantum-ML prototypes:** QGAN/QVAE kernels are under study on small photon patches to check if hybrid quantum-classical discriminators can cut the sample count needed for rare angles without compromising latency.

References

[1] Souvik Priyam Adhya et al. "CEPC Technical Design Report - Reference Detector". In: (Oct. 2025). arXiv: 2510.05260 [hep-ex].
[2] Simon C. Blyth. "Opticks: GPU Optical Photon Simulation for Particle Physics using NVIDIA® OptiX™". In: *Proceedings of the International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018)*. Vol. 214. Open Access, CC BY 4.0. EDP Sciences, 2019, p. 02027. doi: 10.1051/epjconf/201921402027.
[3] Simon C. Blyth. *Opticks*. https://bitbucket.org/simoncblyth/opticks/src/master/. Accessed: 2025-09-16.
[4] Simon C. Blyth and JUNO Collaboration. "Meeting the challenge of JUNO simulation with Opticks: GPU optical photon acceleration via NVIDIA® OptiX™". In: *EPJ Web of Conferences*. Vol. 245. Open Access, CHEP 2020. EDP Sciences, 2020, p. 11003. doi: 10.1051/epjconf/202124511003.
[5] ONNX Runtime developers. *ONNX Runtime*. https://onnxruntime.ai/. Version: x.y.z. 2021.
[6] Markus Frank et al. *AIDASoft/DD4hep*. webpage: http://dd4hep.cern.ch/. Oct. 2018. doi: 10.5281/zenodo.592244. url: https://doi.org/10.5281/zenodo.592244.