Modeling the CMS High-Granularity Calorimeter with Generative AI

Sitian Qian (Northwestern University and Fermilab)

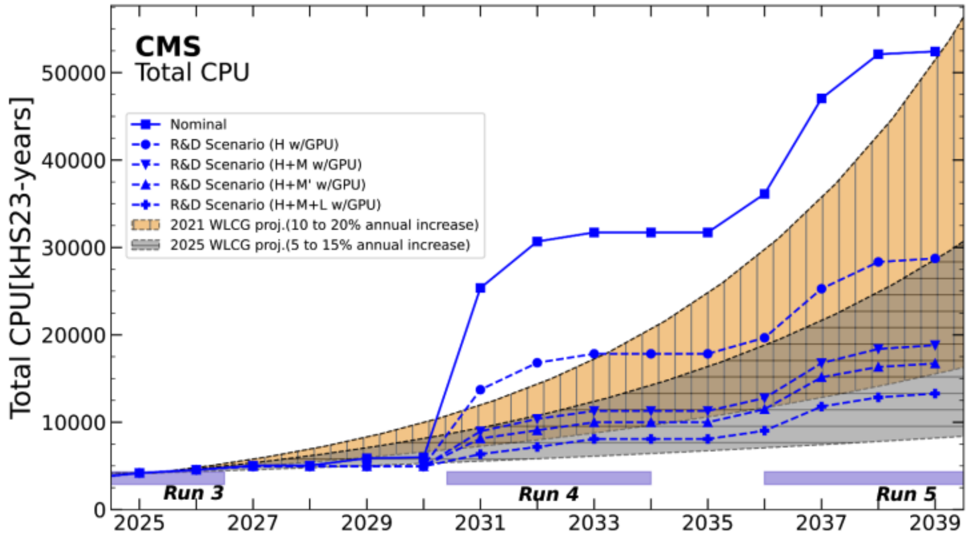
On behalf of the CMS collaboration

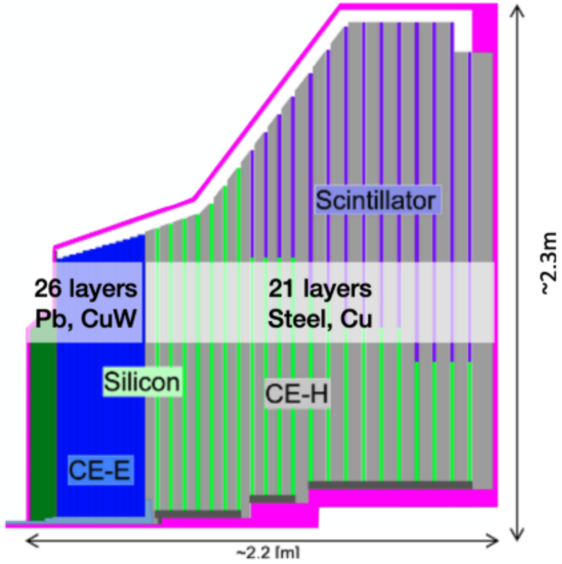
ML4Jets 2026

Based on CMS-DP-2026-049

Why simulate HGCal Fast with Generative Models?

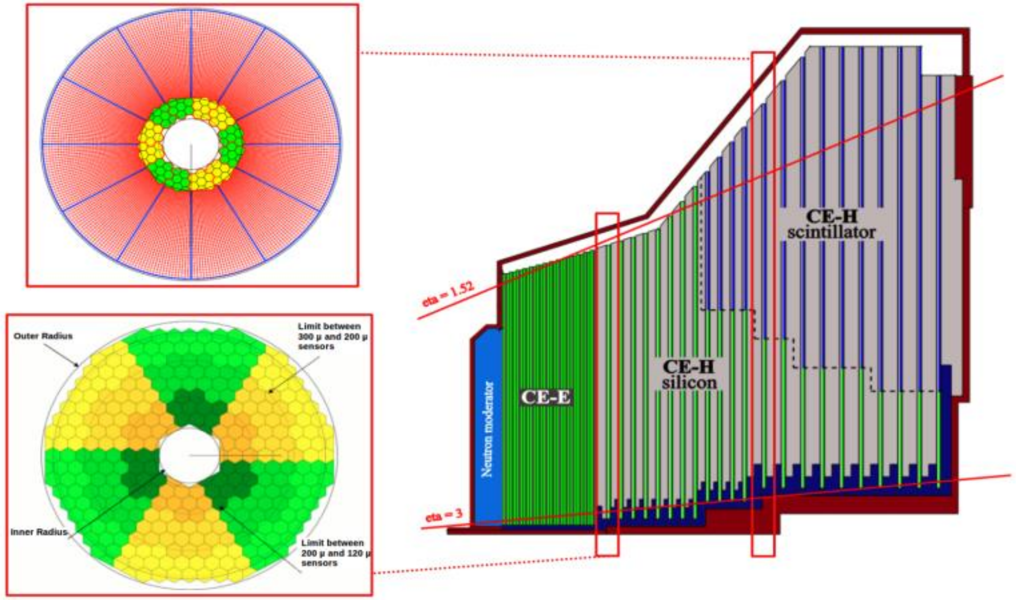
- **HL-LHC raises major computing challenge**
  - <PU> ~ 200 post great challenge
- **For CMS the challenge is unprecedented**
  - End Cap calorimetry done with a high granularity calorimeter (HGCal):
    - 47 layers, ~6 M Si channels, irregular hex geometry as per very recent design geometry
    - Full simulation with Geant4 will have a sharp growth of cost for HGCal due to such design choice
- **The dimensionality of HGCal posts a need of sophisticated generative models**
  - And for CMS a careful choice must be made
  - This talk: *benchmark generative AI as a fast-sim surrogate for CMS HGCal simulation*





Datasets

- **One sentence summary: *Single particle showers simulated with Geant4***
  - Shower types:
    - EM shower: Photons (***γ***)
    - Hadronic shower: Charged pions (***π***)
    - No pileup considered
  - Angular position:
    - Fixed incident angle: $\eta = 2.0$, $\varphi = \pi/2$
  - Energy distribution:
    - Sampled from [1, 1000] GeV
    - Subject to log-uniform distribution
  - Statistics: 1M showers per particle specifies
    - 700K for training, 300K for evaluation
  - Showers are with huge dimensionality!
    - Truncate each shower with ***Region of Interest***



- **Region of Interest (ROI) per shower**
  - Per lay definition a radius, such that > 99.5% shower energy being retained
  - ***γ***: r = 25 cm (fixed)
    - \# of cells per layer: min: 206 vs max: 2,076
  - ***π***: r grows 42 → 127 cm with depth
    - \# of cells per layer: min: 4,016 vs max: 22,268
  - **O(M) cells per shower → 20x increase w.r.t. Calochallenge**

Evaluation Metrics

- **Per shower features:**
  - Pre-process: E_{min} = 10 keV cut-off
  - Per layer: **fractional E (a measure of longitudinal profile)**, ⟨x⟩ , ⟨y⟩ , widths, occupancy
  - Transverse profile in hex-neighbour rings
  - Global:
    - E_{tot} / E_{inc}
  - Feature counts:
    - ***γ***: 231 features / shower
    - ***π***: 358 features / shower

- **Metrics:**
  - 1D separation power ***s***:
    - $s := \Sigma_i ((d_i - d_i')^2 / (d_i + d_i'))$
  - 1D Kolmogorov–Smirnov (unbinned)
  - Multivariate FPD, KPD
  - Classifier AUC (2 × 2048 MLP)
  - All the metrics are supplied with GEANT–GEANT bootstrap:
    - Perfect-model floor

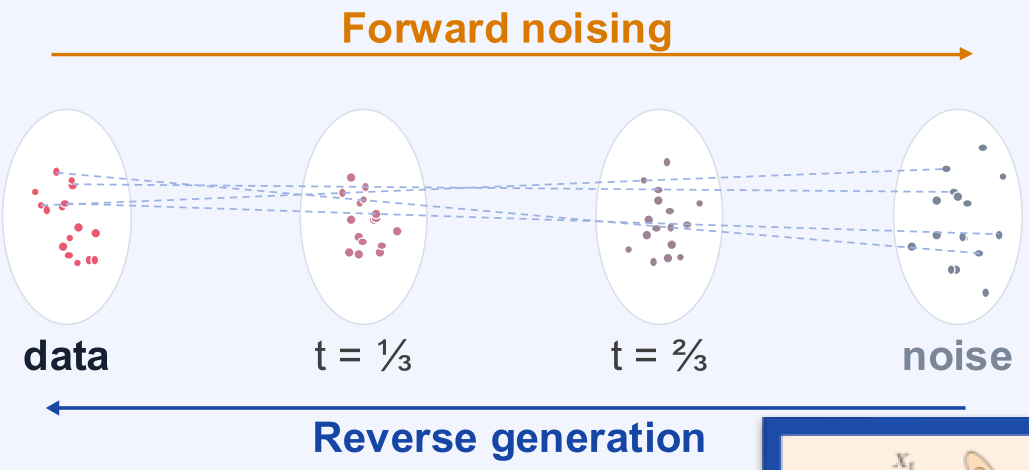
Largely following the CaloChallenge setup, with features fitting HGCal physics target

No models currently reach GEANT–GEANT indistinguishability
hence only report metric values, not p-values

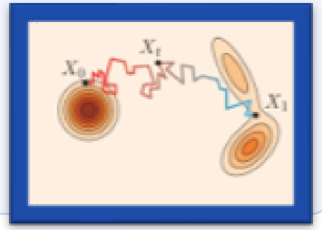
Generative Paradigms

- **Major participating models choose Diffusion Models or Flow Matching**
  - Fits the overall trend in ML community for continuous data generation

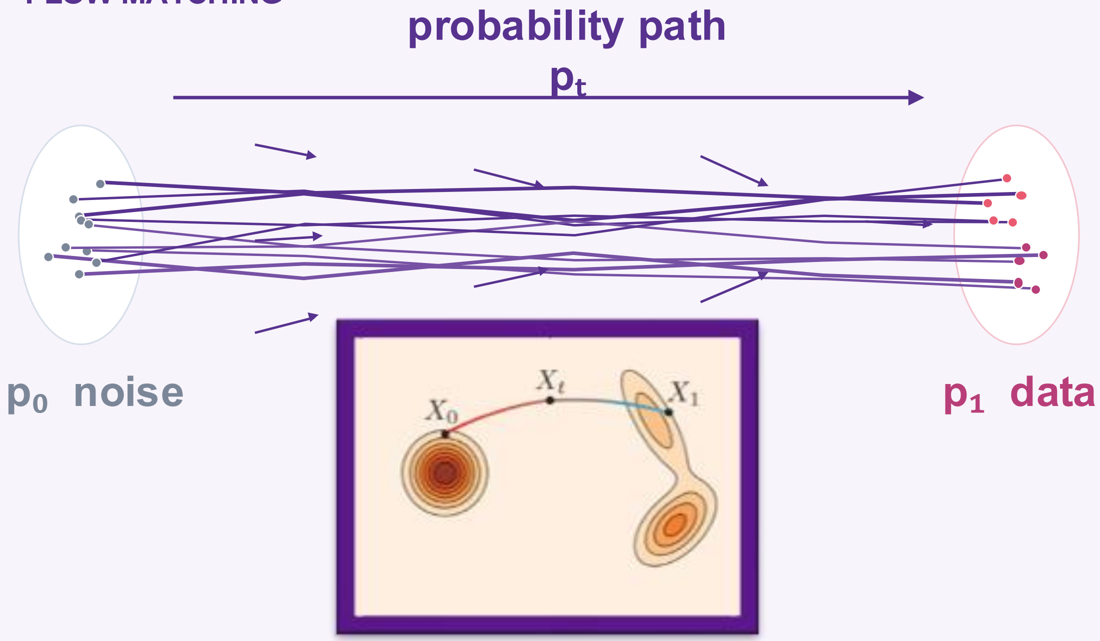
DIFFUSION MODEL



Modernized diffusion models leverage score matching for continuous time



FLOW MATCHING



Shared formulation
A time dependent vector field carries a simple prior into the data distribution

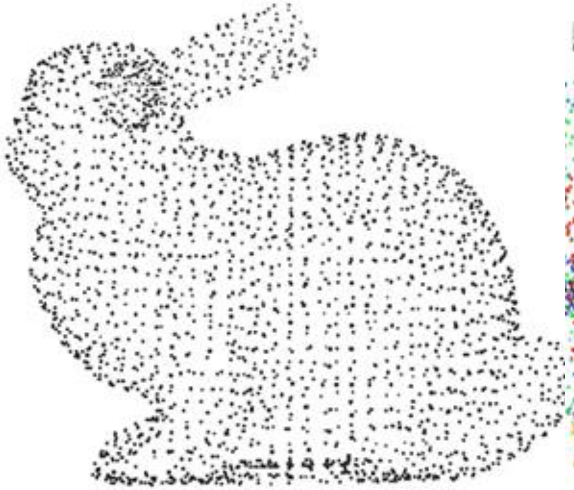
$$\partial_t p_t + \nabla \cdot (p_t v_t) = 0$$

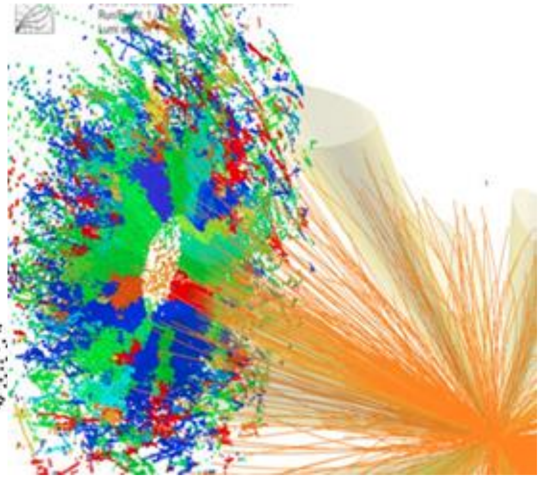
score → velocity
path → velocity

Backbones and Representations

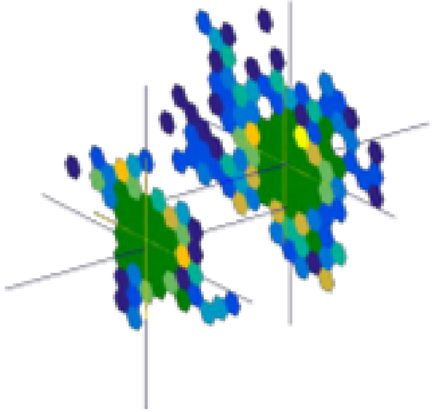
- Participating models also span over different backbones motivated by different shower representation choices

Point Cloud/*"Hits"*





Voxels/*"Clusters"*



Can be further represented as a graph

The clustering could be learnable, or modernly termed as "Patchification"

Overview of Participants

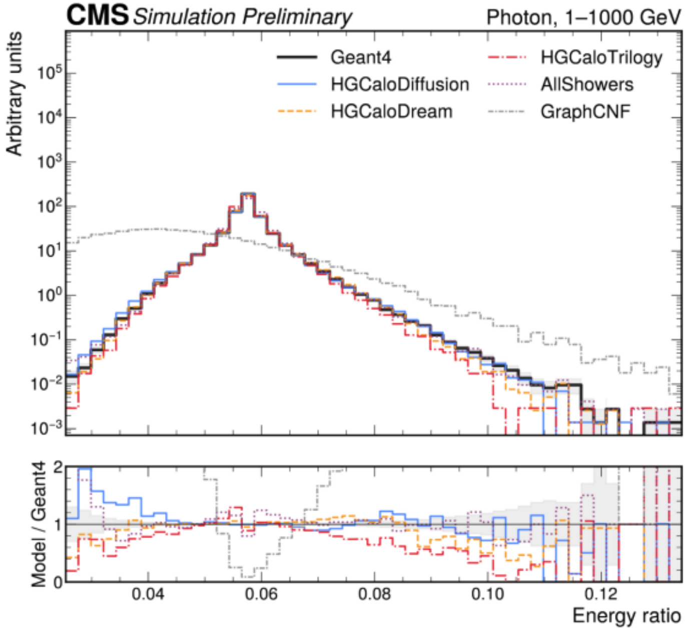
GLaM = Geometry Latent Mapping
embedding method developed for CaloDiffusion

Model	Paradigm	Backbone	Representation
HGCaloDiffusion	**Diffusion (EDM/DDIM)**	Cyl.-conv U-Net + attention	Voxel (GLaM)
HGCaloTrilogy	**MeanFlow + GMM prior**	Shared with HGCaloDiffusion	Voxel (GLaM)
HGCaloDREAM	**Flow matching**	Vision Transformer (DiT)	Voxel (patched)
AllShowers (*)	**Flow matching**	PointCountFM + CNF-Transf.	Point cloud of hits
GraphCNF (**)	**Riemannian flow**	Graph NF + Riemannian FM	Graph (per cell)

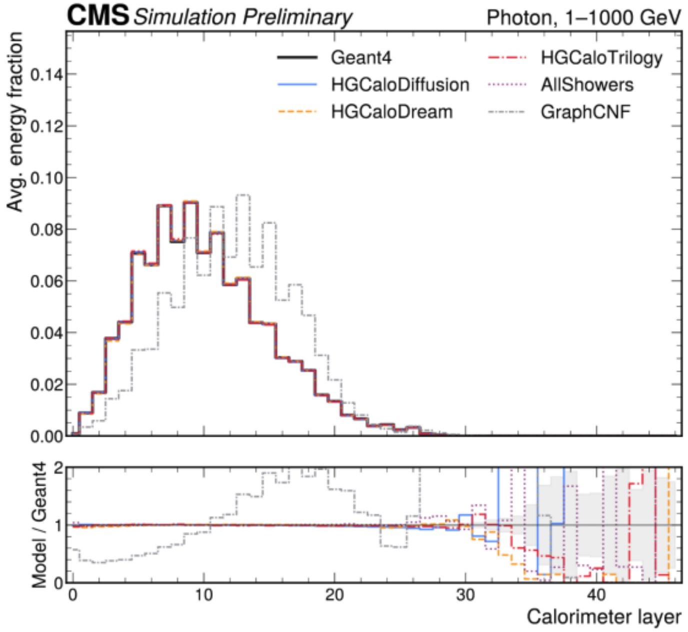
(*) AllShowers introduce particle species as a condtion → one trained model handle both showers!
(**) GraphCNF has photon (EM shower) only

Performance: Photon (EM Shower)

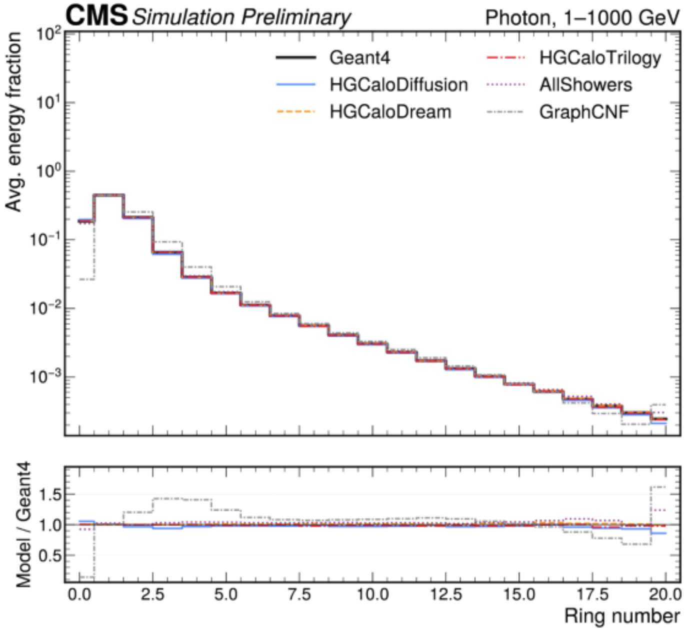
E_{tot} / E_{inc}



Longitudinal profile

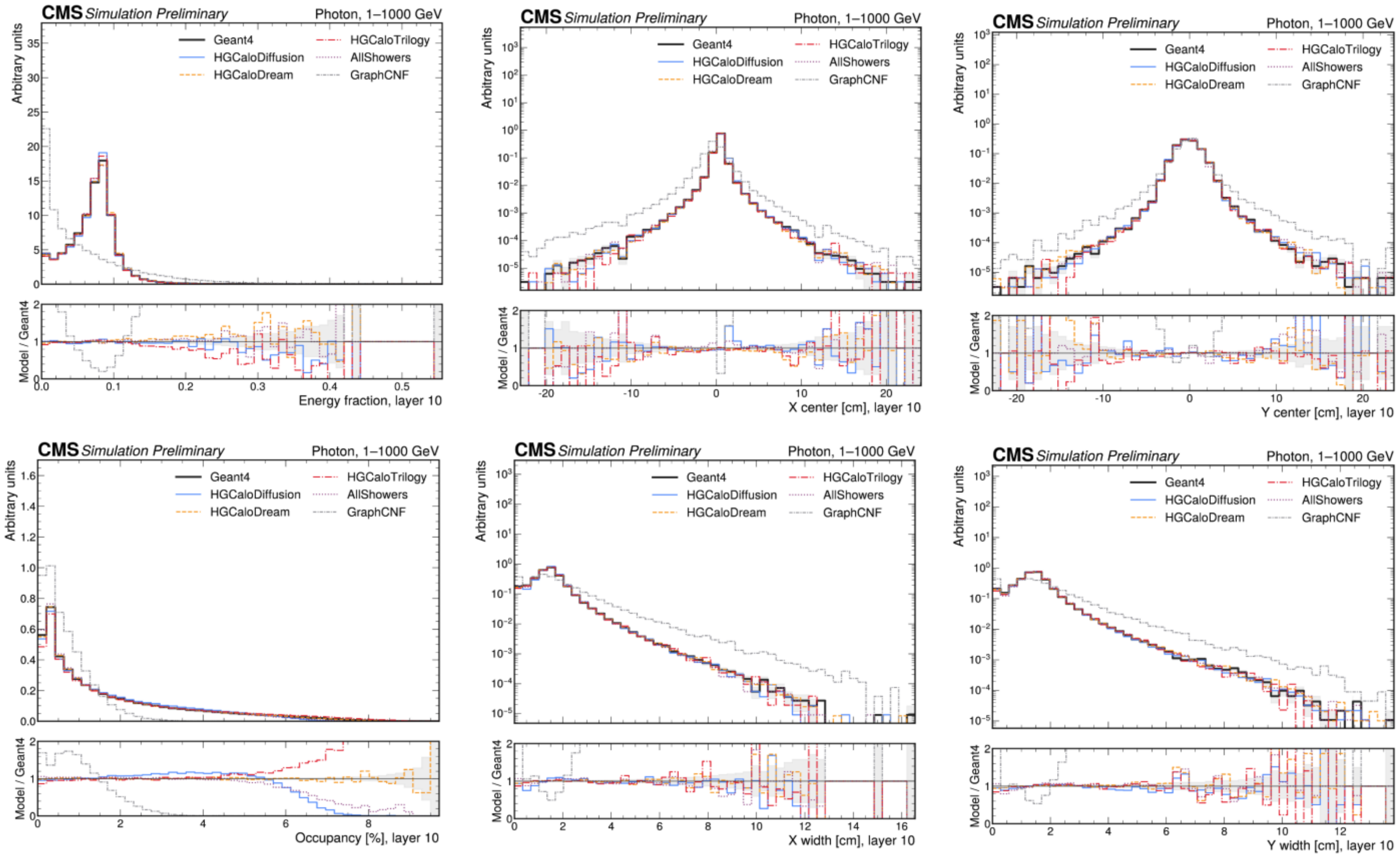


Transverse profile



Most models reproduce these global features over several orders of magnitude; transverse tails are the hardest.

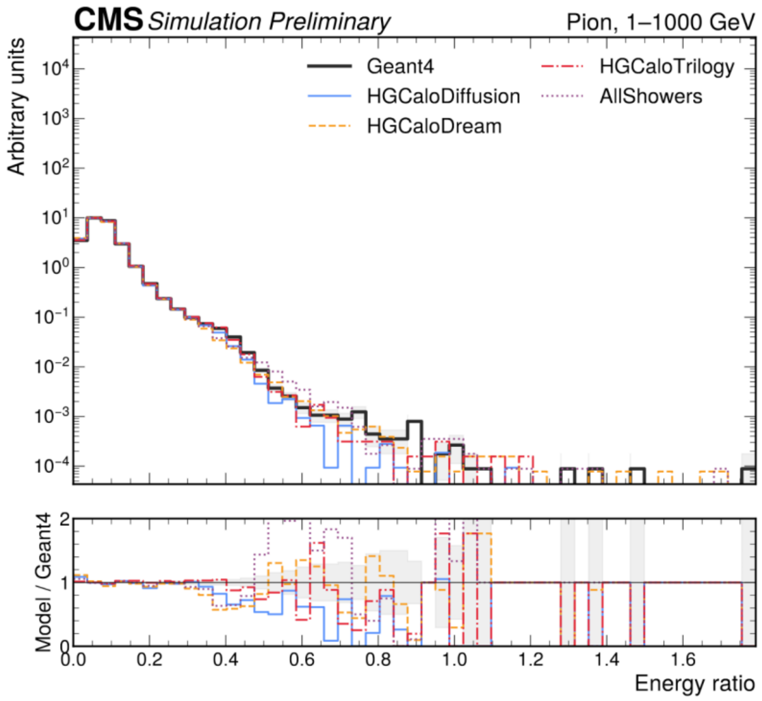
Performance: Photon (EM Shower), Layer 10



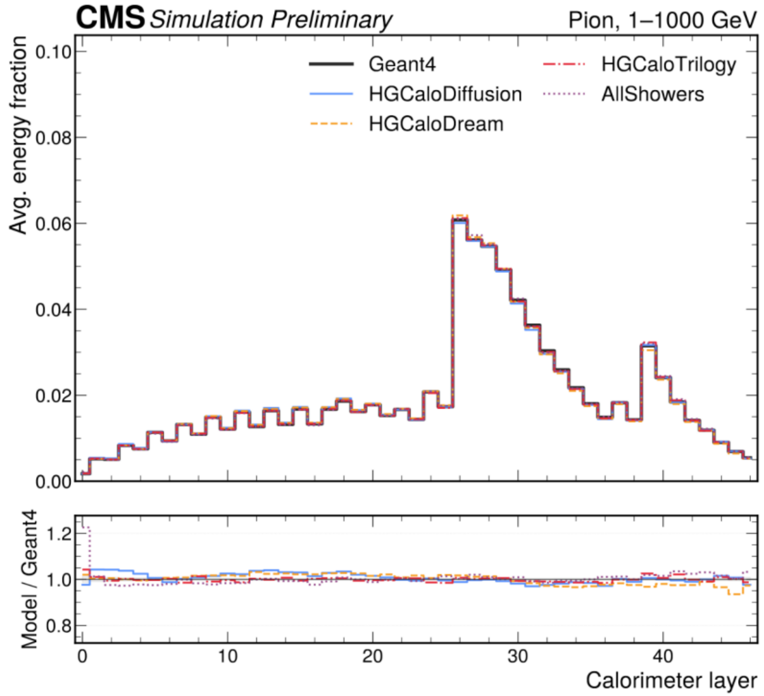
- Good modeling of key features, including most tails
- Occupancy is hardest, only HGCaloDream captures it well

Performance: Pion (Hadronic Shower)

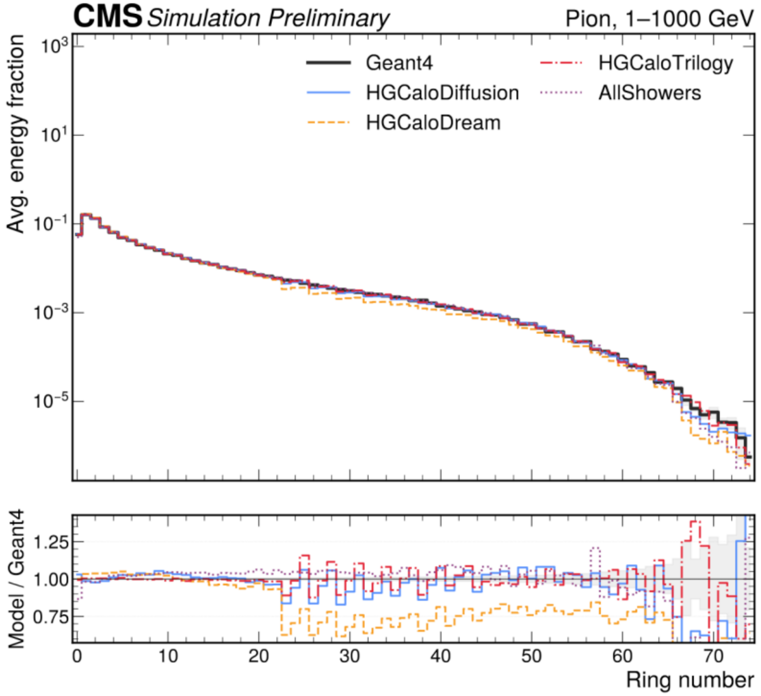
E_{tot} / E_{inc}



Longitudinal profile

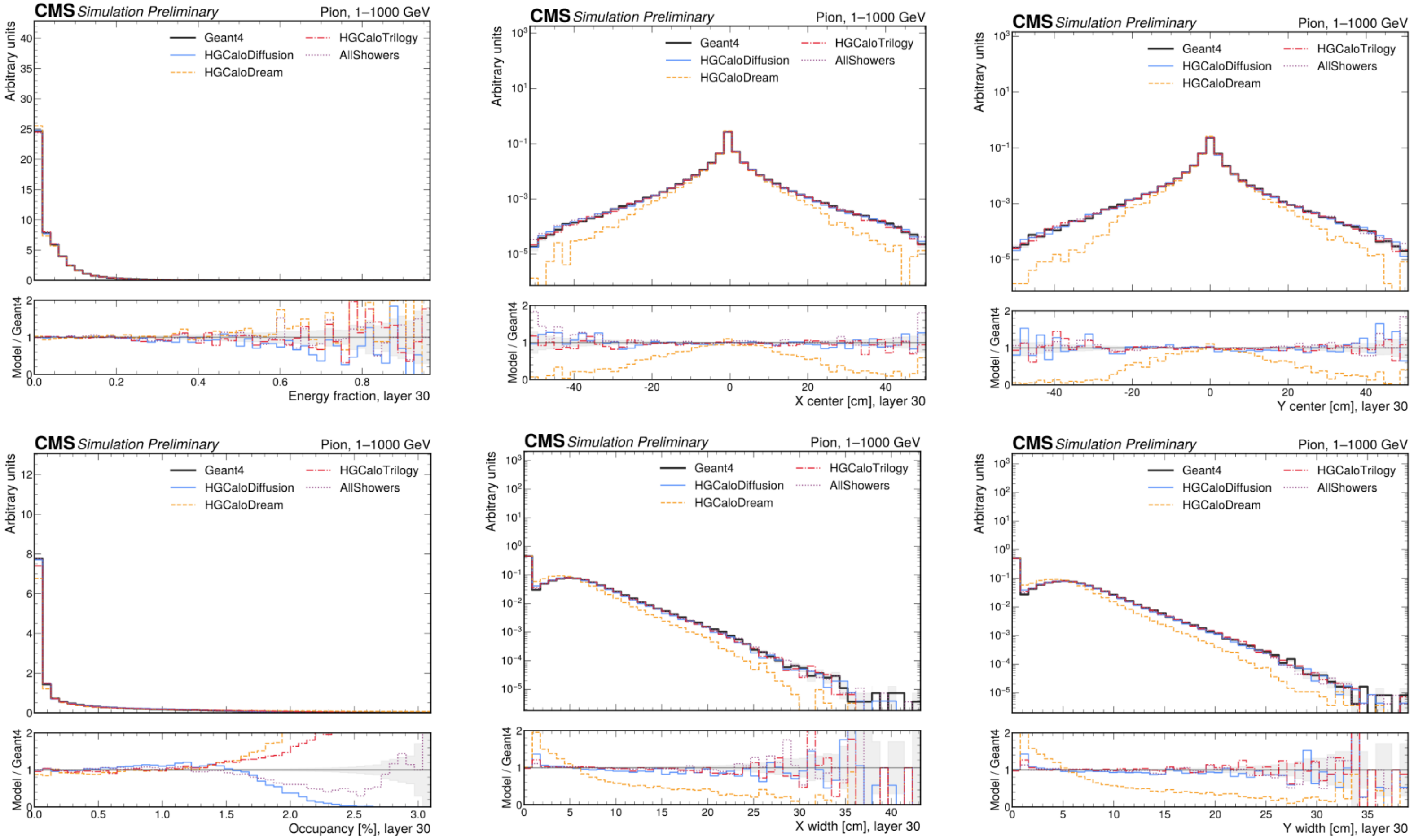


Transverse profile



Hadronic showers spread over many more cells per layer; transverse profile and tails are noticeably harder.

Performance: Pion (Hadronic Shower), Layer 30



- Good modeling of key features, including most tails
- Pion showers found to be more difficult than photons

Quantitative metrics : Photons

First 5 columns displaying both separation power and KS (x 10^{-3}), FPD and KPD (x 10^{-3})

Model	Energy	Transv.	Center	Width	Occ.	AUC	FPD	KPD
GEANT	0.06 / 1.73	0.07 / 2.15	0.07 / 1.60	0.07 / 1.62	0.06 / 1.71	0.497	0.3 ± 0.1	−0.003
HGCaloDiffusion	0.17 / 7.77	0.55 / 17.9	0.71 / 11.3	0.54 / 27.0	6.54 / 22.4	0.757	18 ± 2	0.030
HGCaloDREAM	**0.15 / 5.26**	**0.12 / 3.36**	**0.22 / 5.21**	**0.21 / 5.97**	**0.10 / 3.46**	**0.562**	30 ± 28	**0.004**
HGCaloTrilogy	0.43 / 7.54	0.18 / 6.09	0.34 / 6.49	0.60 / 12.8	2.10 / 19.8	0.735	13 ± 0.3	0.099
GraphCNF	120 / 238	63.7 / 186	63.5 / 151	69.4 / 180	102 / 204	0.997	1067 ± 3	5.76
AllShowers	0.40 / 6.67	7.02 / 33.1	1.05 / 18.3	0.58 / 12.4	2.54 / 15.8	0.839	20 ± 4	0.059

- **GEANT row = bootstrap floor a perfect model could reach**
- **HGCaloDREAM best on most photon metrics (highlighted)**
- **Leading 1D separations reach < 10^{-3} on most features**

Quantitative metrics : Pions

First 5 columns displaying both separation power and KS (x 10^{-3}), FPD and KPD (x 10^{-3})

Model	Energy	Transv.	Center	Width	Occ.	AUC	FPD	KPD
GEANT	0.07 / 1.84	0.06 / 1.43	0.08 / 1.82	0.07 / 2.88	0.07 / 1.18	0.499	1.0 ± 0.1	−0.002
HGCaloDiffusion	0.15 / 10.8	0.49 / 19.7	0.88 / 20.5	1.03 / 79.4	1.90 / 13.8	0.756	16.1 ± 0.4	0.019
HGCaloDREAM	0.27 / 28.6	0.51 / 64.8	7.40 / 35.7	17.6 / 82.0	29.9 / 113	0.944	2519 ± 11	7.10
HGCaloTrilogy	**0.12 / 4.73**	**0.13 / 11.3**	**0.21 / 10.3**	0.46 / 23.7	1.87 / 31.3	0.660	15.6 ± 0.5	0.060
AllShowers	0.13 / 5.30	0.30 / 6.22	0.39 / 12.1	**0.14 / 5.94**	**0.95 / 9.33**	**0.649**	**9.0 ± 0.2**	**0.011**

- **HGCaloTrilogy leads bulk shape; AllShowers leads width / occupancy**
  - Best π model ≠ best γ model
- **Hadronic showers uniformly harder than electromagnetic**

Summary and Outlook

- **Convergence and Diversity shown in participant models:**
  - Five models — diffusion, flow matching, continuous NFs
  - Voxel, patch, point-cloud, and graph representations
  - × 20 complexity over CaloChallenge : ~1 M cells, irregular hex, 47 layers
  - Many shower features reproduced with high fidelity
    - *Leading metrics reach CaloChallenge-level performance*
  - No single model dominates:
    - Hardest observable: cell occupancy, worsens with energy
    - Best $\boldsymbol{\gamma}$ model ≠ best $\boldsymbol{\pi}$ model; universal surrogate is open work
- **More results/datasets/models to be published shortly, stay tuned!**

First comprehensive study of generative AI on the CMS HGCal

Thank you!

Reference

CMS Collaboration, "Phase-2 Upgrade of the CMS Endcap Calorimeter," CERN-LHCC-2017-023, CMS-TDR-019 (2017).

CMS Collaboration, "CMS Offline Software and Computing for HL-LHC Conceptual Design Report," CERN-LHCC-2026-003, LHCC-G-186 (2026).

Apollinari et al., HL-LHC, CERN Yellow Reports, Vol. 10 (2020).

Agostinelli et al. (Geant4), Nucl. Instrum. Meth. A 506 (2003) 250.

Krause et al., "CaloChallenge 2022," Rep. Prog. Phys. 88 (2025) 116201.

Ahmed et al., "Evaluation metrics for generative calorimeter models," arXiv:2406.12898 (2024).

Kansal et al., "Evaluating generative models in HEP," Phys. Rev. D 107 (2023) 076017.

Ho, Jain, and Abbeel, "Denoising Diffusion Probabilistic Models (DDPM)," NeurIPS 33 (2020).

Song, Meng, and Ermon, "Denoising Diffusion Implicit Models (DDIM)," ICLR 2021.

Karras et al., "Elucidating the Design Space of Diffusion-Based Generative Models (EDM)," NeurIPS 2022.

Lipman et al., "Flow Matching," ICLR 2023.

Chen et al., "Neural Ordinary Differential Equations," NeurIPS 2018.

Peebles and Xie, "Scalable Diffusion Models with Transformers (DiT)," ICCV 2023.

Chen and Lipman, "Flow Matching on General Geometries," ICLR 2024.

Amram and Pedro, "CaloDiffusion," Phys. Rev. D 108 (2023) 072014. Improved HGCal implementation: arXiv:2308.03876.

Favaro et al., "CaloDREAM," SciPost Phys. 16 (2024) 134. Updated HGCal implementation: arXiv:2601.05289 (2026).

Buss et al., "AllShowers", arXiv:2601.11716 (2026).

Jiang et al., "CaloTrilogy", arXiv: 2606.06145 (2026)

Backup

Everything else

HGCaloDiffusion

Backbone

- EDM diffusion, DDIM sampling (200 steps)
- U-Net: 3 ResNet blocks down + 3 up
- Cylindrical convolutions + linear attention
- Channels 32 / 32 / 64 / 96 with skip connections
- Adam, lr = 5×10^{-4}, up to 1000 epochs (ES patience 20)

Geometry & decoding

- Voxelised cylindrical embedding (fixed embedding, not learnable GLaM)
- 12 angular × 21 (γ) / 48 (π) radial bins
- LayerDiffusion MLP : total E + per-layer fractions
- Sparsity-preserving decoding (cell p ∝ $1/N^{cells}$/voxel)
  - Improves occupancy mismodelling

HGCaloTrilogy

- Built as a branch of HGCaloDiffusion
  - Shares preprocessing, GLaM cylindrical embedding, sparsity-preserving decoding
  - The generative core is what changes

Three modifications

- MeanFlow transport — average-velocity formulation over coarser time intervals
  - → robust 1- and few-step sampling without fidelity collapse
- Learned conditional GMM prior instead of isotropic Gaussian
  - → initialises closer to the shower manifold
- Physics-guided loss on high-level observables (layer-wise E, …)
  - → no post-processing needed at inference

HGCaloDream

Two-stage factorisation

- u ~ p(u | E_inc) → layer-energy ratios
- x ~ p(x | E_inc, u) → normalised voxels
- Trained independently, combined at generation

Energy network

- Transformer encoder–decoder, MLP head
- ~ 800 k params, 250 k iters

Shape network (ViT / DiT)

- Voxels grouped into non-overlapping patches
- Flexible patch shape (P_x, P_y, P_z) for irregular geometry
- 8 blocks × 6 heads, embed 480, ~35 M params
- Adaptive conditioning (a, b, c) = g(E_inc, u, t)
- γ: 1199 patches of 60 voxels, 3 M iters, batch 16
- π: GLaM → (47, 12, 48), patches (2, 4, 4), 1 M iters

AllShowers

Point-cloud representation

- Each point = active cell (layer, x, y, E)
- Variable-length, permutation-invariant
- Post-hoc: each point → nearest cell, E summed

Two-stage CNFs

- PointCountFM: $p(n_i \mid$ l_inc, e_inc$)$ — points per layer
- CNF-Transformer: positions + E
  - OT pairing
  - Layer-index attention mask $|\Delta\ell| \leq 2$

Single model for both species

- Conditioned on particle type
- No retraining to switch γ ↔ π

Energy calibration (γ)

- Monotonic cubic spline (256 knots) on E_tot / E_inc
- Narrows the too-wide energy-ratio peak

AllShowers

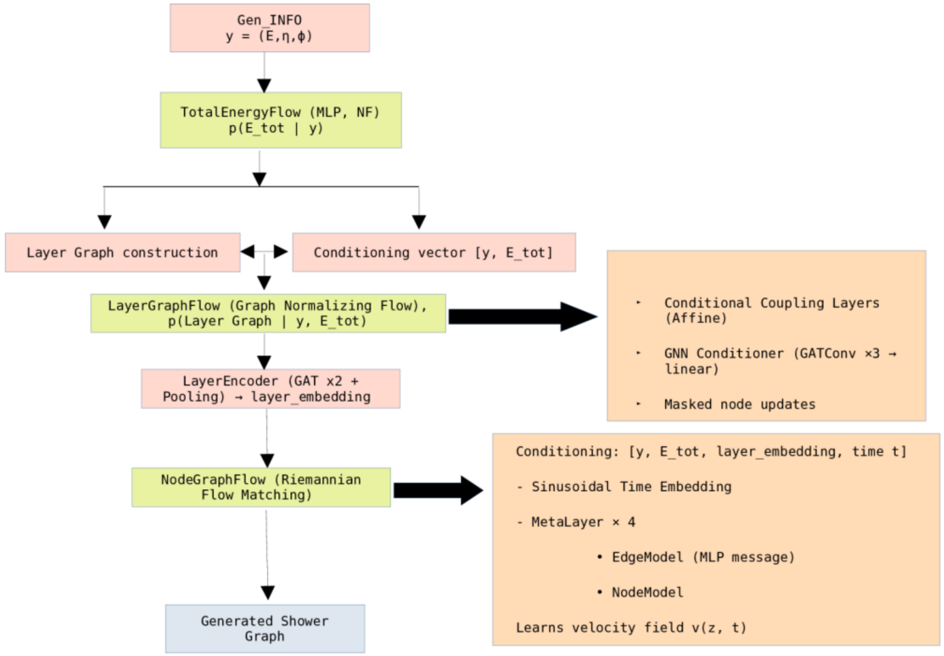
CNF — ODE-defined transport

$$dz(t) / dt = f(z(t), t; \theta), \quad z(t_0) \sim \mathcal{N}(0, I)$$

Trained by maximum likelihood (trace-Jacobian)

Three chained flow blocks

- TotalEnergyFlow: $(E, \eta, \varphi) \rightarrow$ E_tot
- LayerGraphFlow: per-layer Graph NF,
  - conditioned on particle + E_tot
- NodeGraphFlow: Riemannian FM → per-cell node energies



Standard FM (energy, layer) + Riemannian FM (nodes)

Photons only in this round — pions = future work