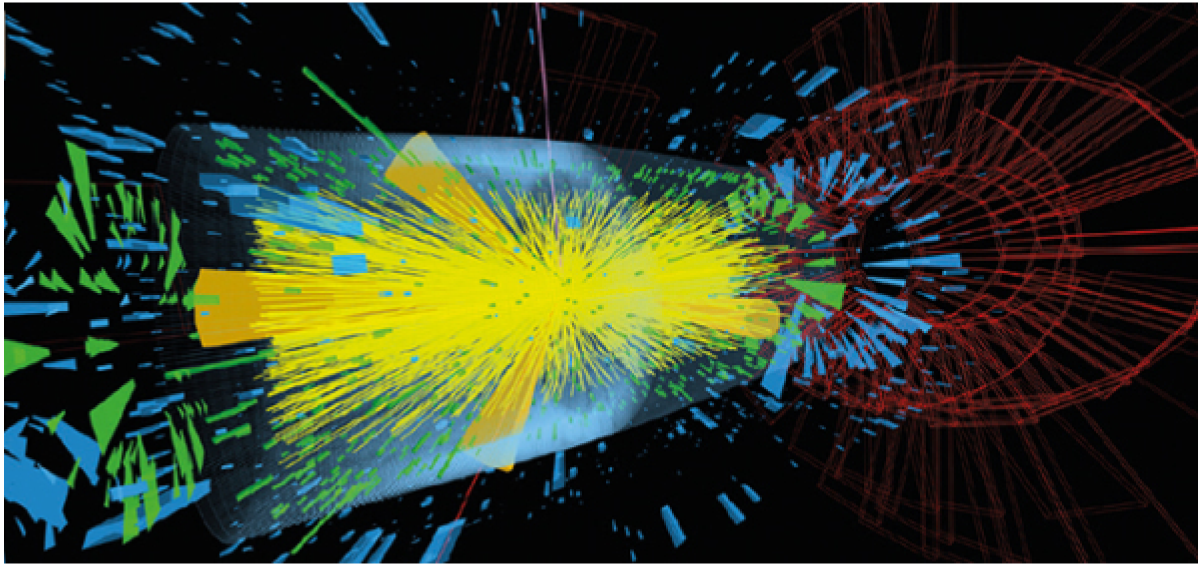
CaloDiffusion

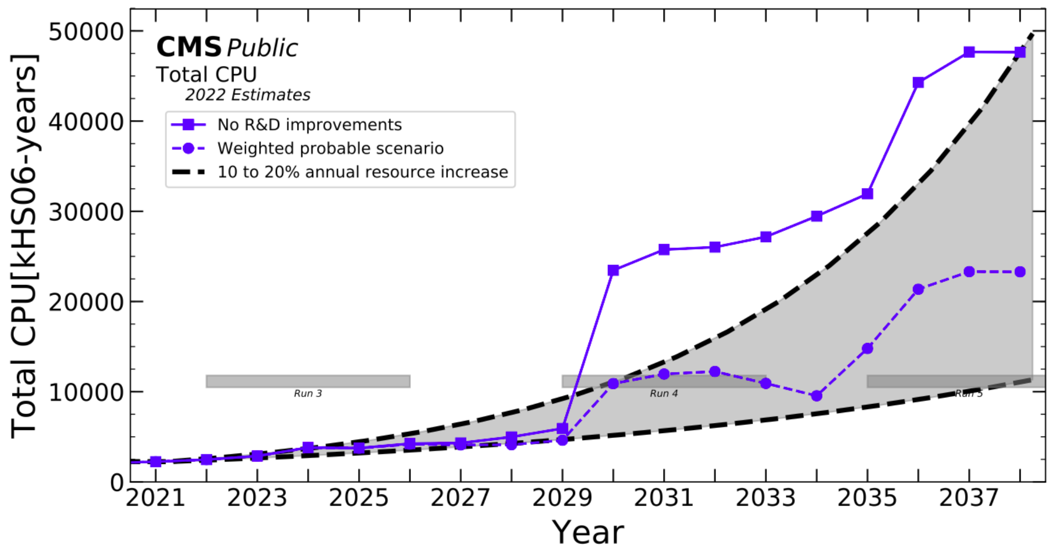


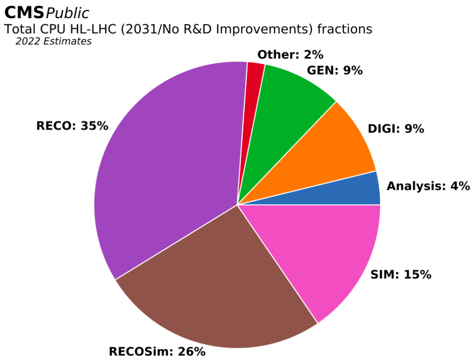
Oz Amram
On behalf of CaloDiffusion Team
HSF Simulation / IML ML WG
Dec. 1st, 2025

Fermilab

CMS

Need for Fastsim



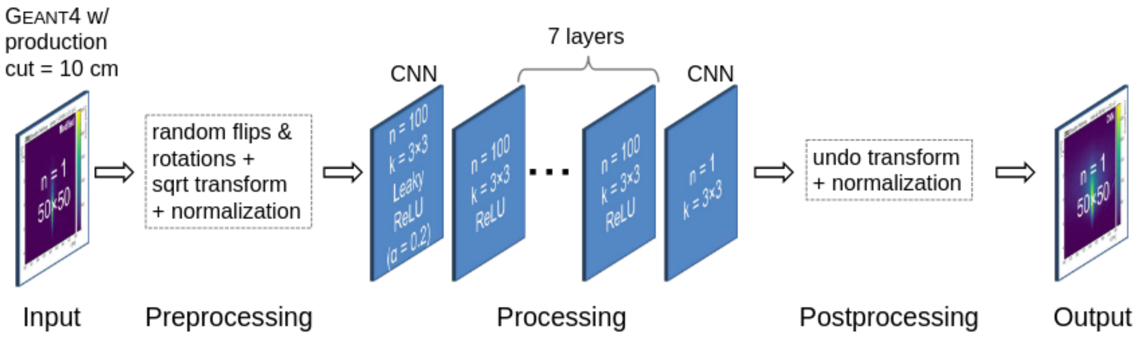


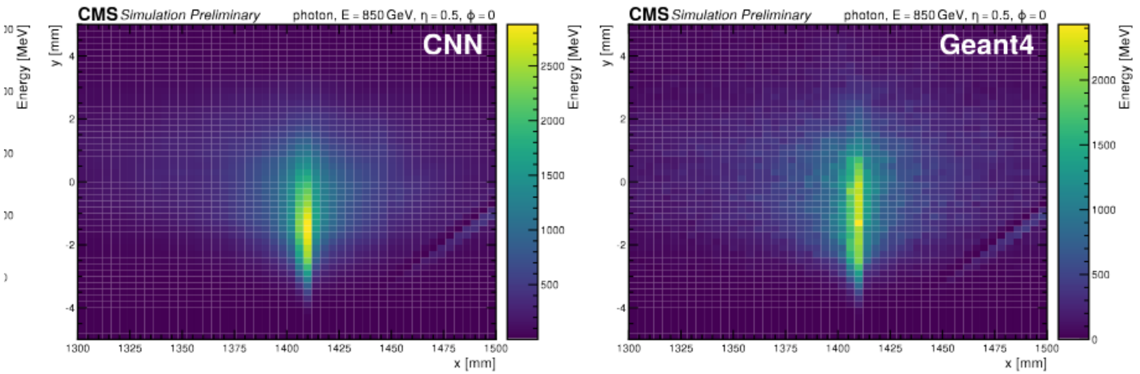
- Computing resources constrained in HL era
- Significant fraction of SIM will need to be FastSim!
  - Calo sim. bottleneck
- High quality fast simulation of calorimeters needed!

NB Parallel 'FlashSim' effort is complementary

Early Work : Refining FastSim with CNN’s

- Studied whether a faster simulation (modified Geant) could be upscaled to higher resolution





- Some success → output matched Geant qualitatively
- However scaling to better quality seemed difficult

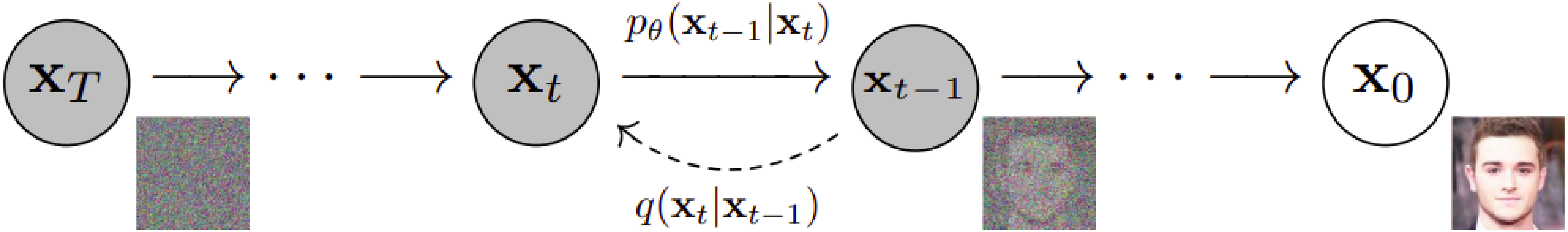
2202.05320

Enter CaloChallenge

- Grant received for further CMS R&D
- CaloChallenge announced just as I was starting my postdoc
  - Great opportunity to develop methods & compare with community
- Also diffusion models taking over gen AI world
  - Similar ‘denoising’ strategy to prev. approach

Diffusion Models

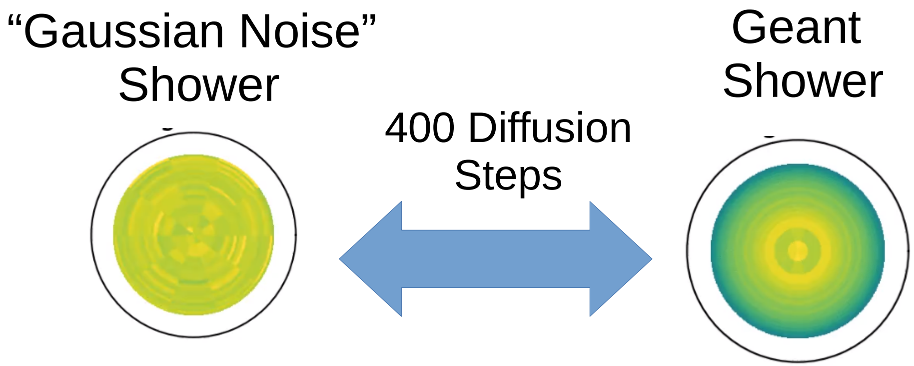
2006.11239



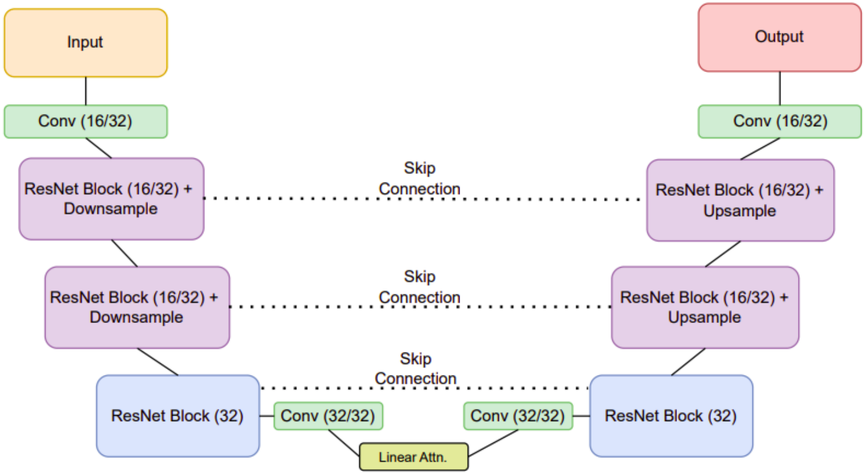
- Diffusion process: Starting with some image, **iteratively add Gaussian noise**, eventually reaching pure noise
- Train a model to **invert the diffusion process**
- Generate by starting from noise image, **iteratively denoise** using trained model
- Can condition on additional input information
  - Eg. text prompt or incident particle energy

'CaloDiffusion'

- Generative diffusion model for Calo simulation
- We use diffusion **400 steps for baseline model**
  - Quality first, speed later
- Denoising network is has 'U-net' architecture based on 3D convolutions
  - Primary input: Noisy shower
  - Conditioning inputs: incident particle energy & diffusion step
- Training objective normalized noise component of the shower
  - Denoising → subtracting noise off
- Several novel optimizations utilized



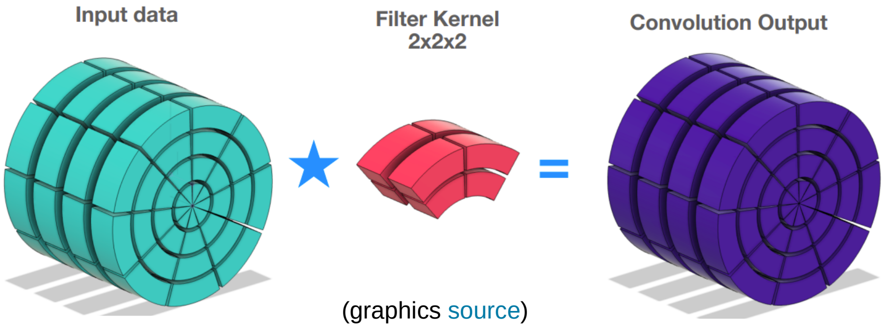
Denoising Model



Optimizing for Cylindrical Data

- Regular convolutions assume pure translation symmetry
- Our data : phi is **periodic**, and R & Z **not translation invariant**

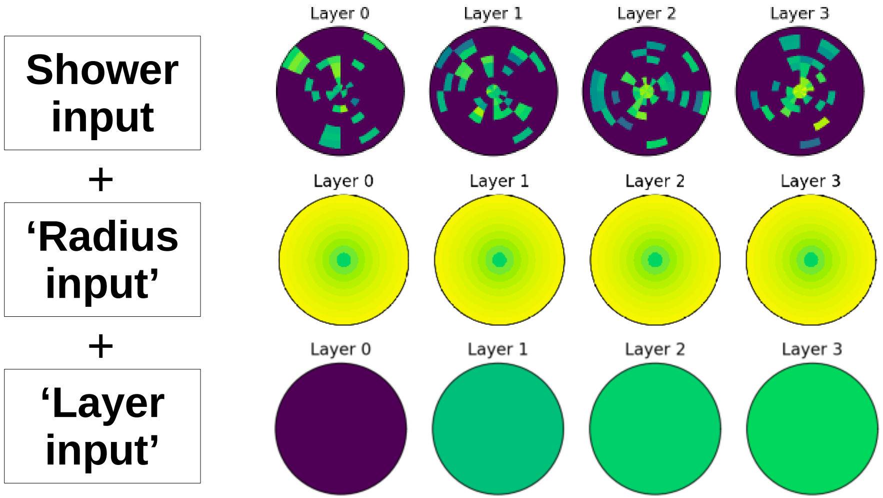
Implement **cylindrical convolutions** to respect periodic boundary of phi



(graphics source)

'Circularly' pad phi dimension before 3D conv

Allow convolutions to be **conditional on R & Z** by using additional channels

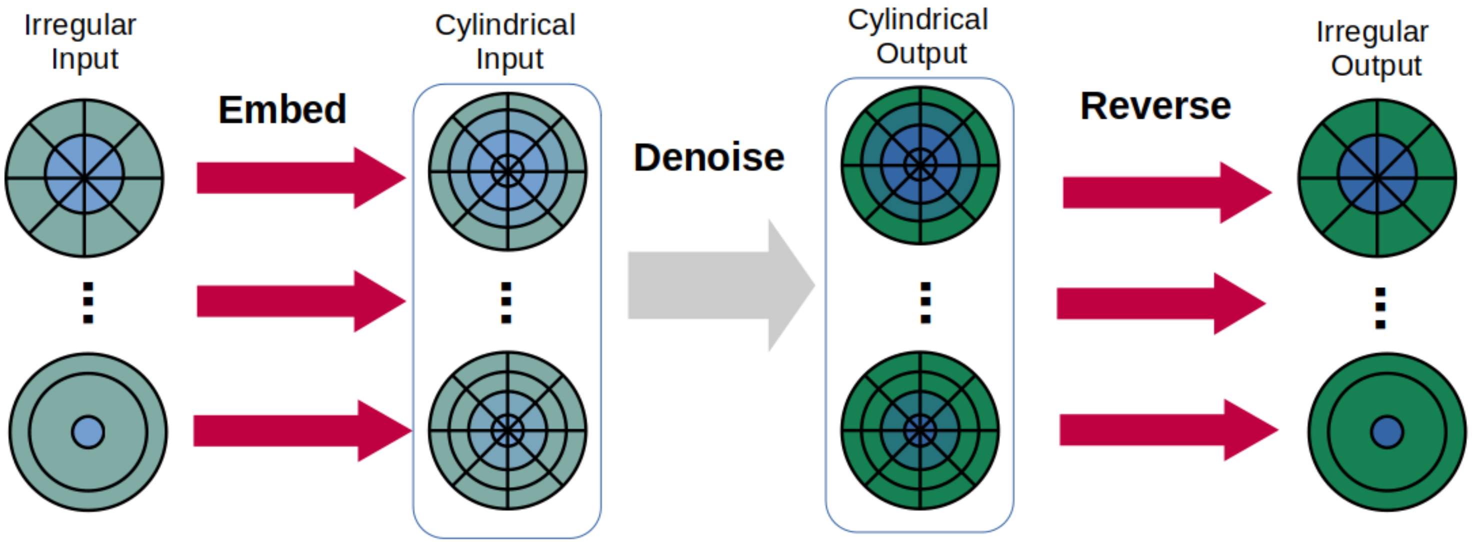


Additional input channels

One Innovation : Embedding Irregular Geometries

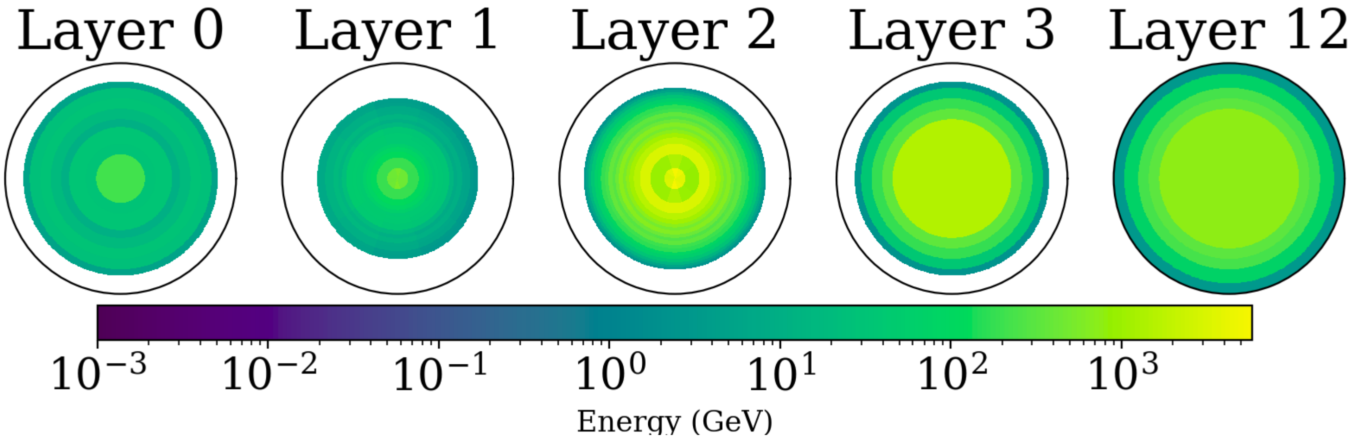
- Real detectors (ie HGCal) have **irregular structure** in different layers
  - Different radial / angular bins in each layer → can't apply your favorite ML tools (like convolutions)
- Learn an **GLaM-orous embedding** that maps input into **regular cylindrical structure**

GLaM : Geometry Latent Mapping

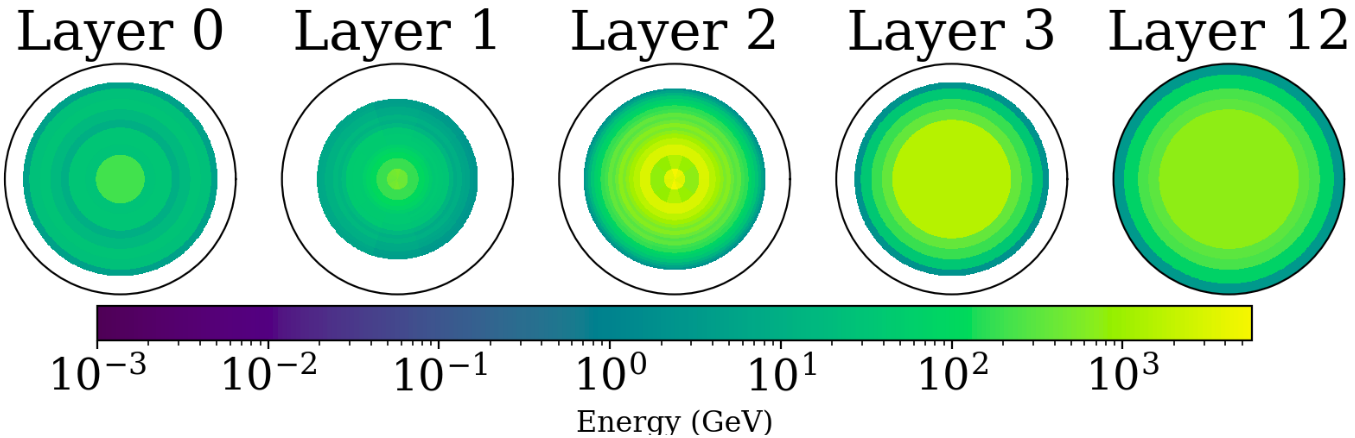


Average Showers

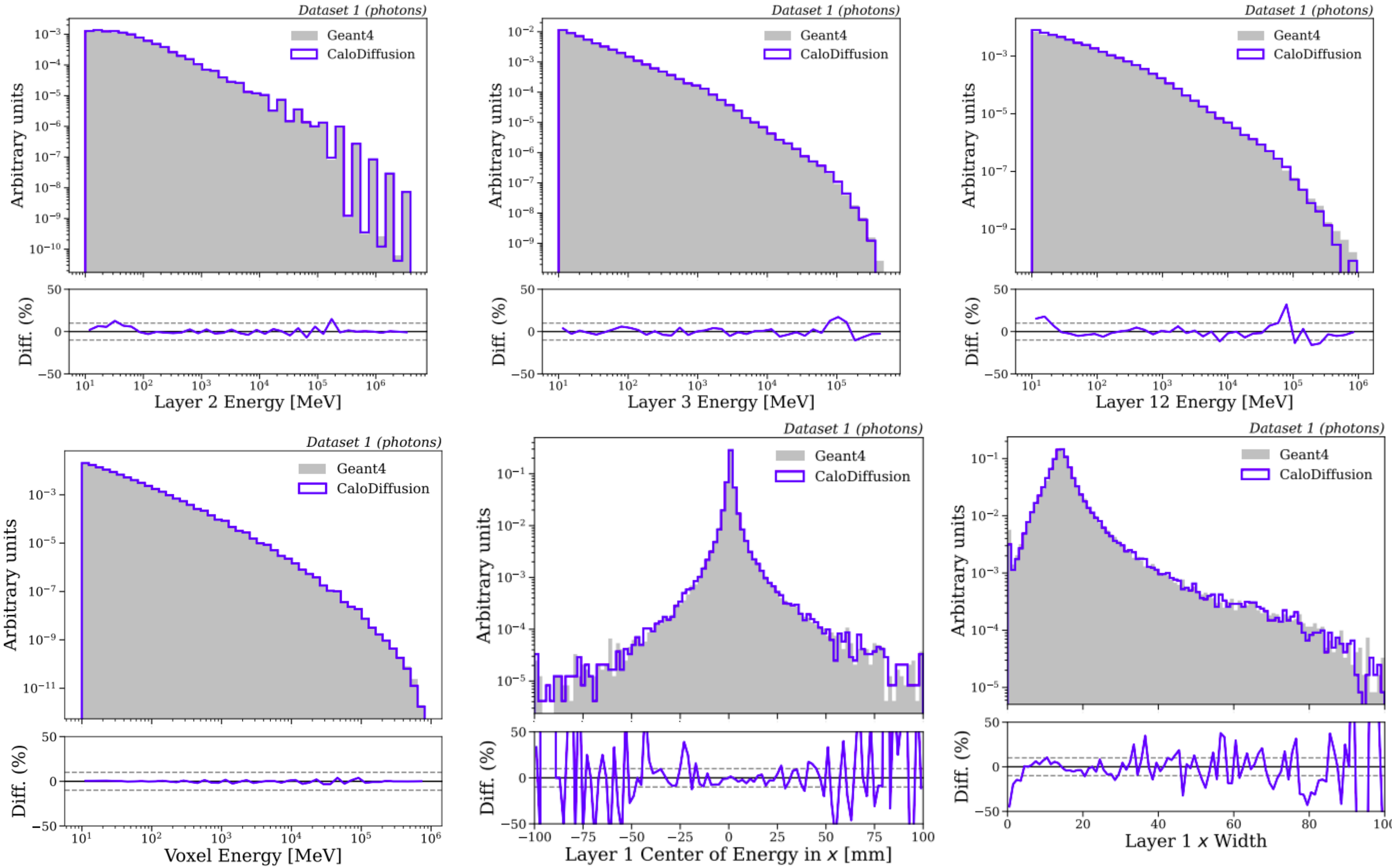
Geant



CaloDiffusion

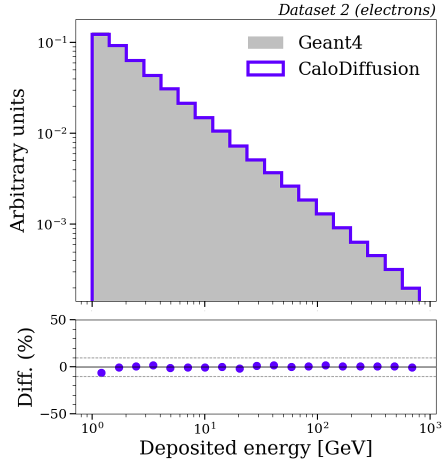


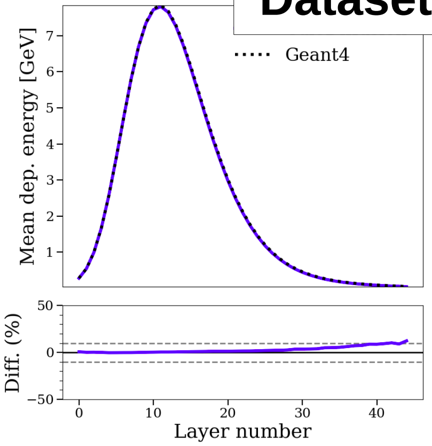
CaloChallenge Dataset 1 Results

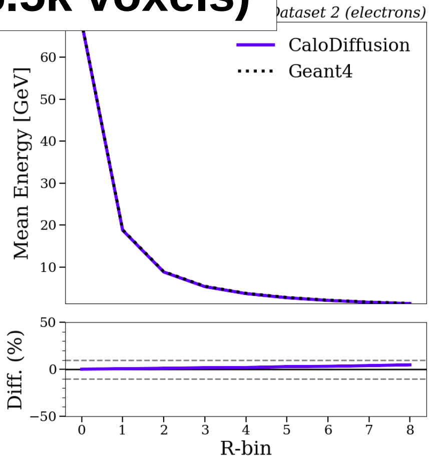


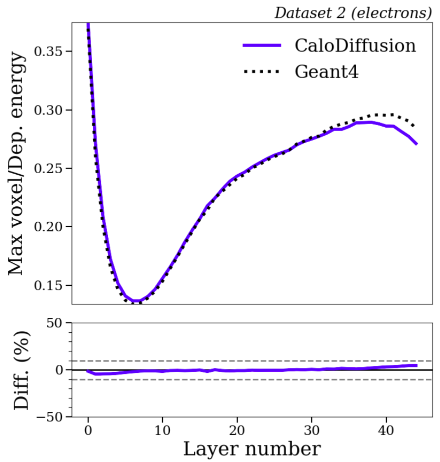
Results: Datasets 2 & 3

Dataset 2 (6.5k voxels)

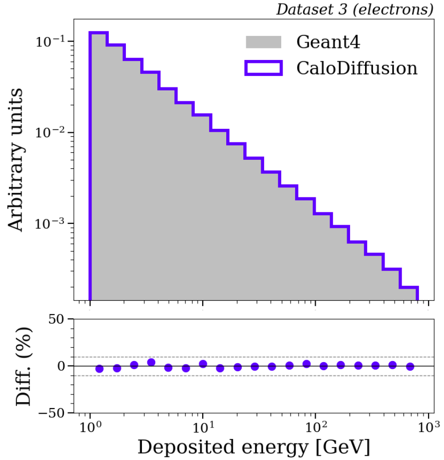


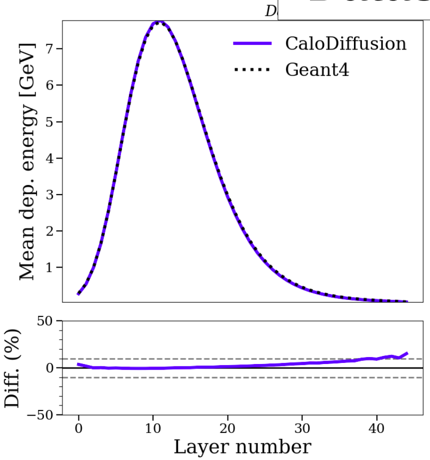


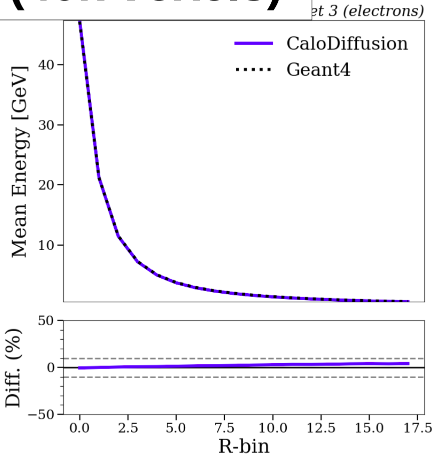


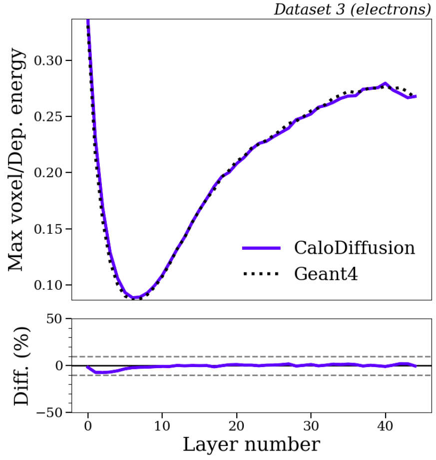


Dataset 3 (40k voxels)









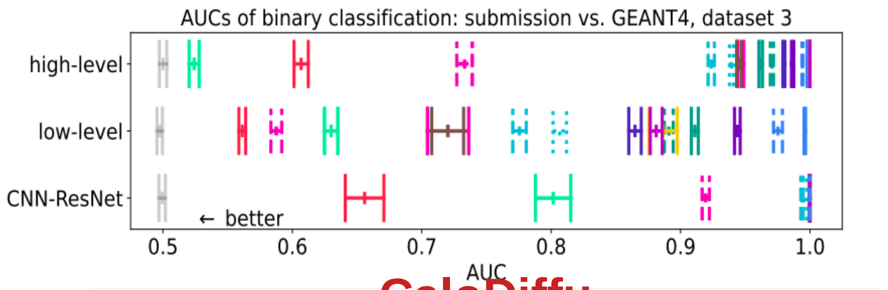
Performance Comparison

- Calodiffusion performed very well in CaloChallenge
- New benchmark in quality at time of publication
- Top-2 in quality on all datasets in final evaluation

Comparisons at time of publication

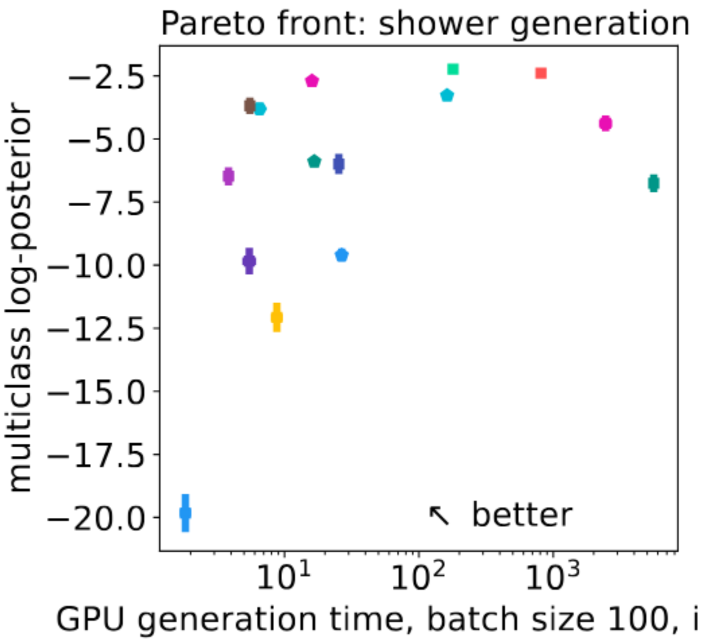
	Classifier AUC (low / high)		
Dataset	CaloDiffusion	CaloFlow	CaloScore v2
1 (photons)	**0.62** / 0.62	0.70 / **0.55**	0.76 / 0.59
1 (pions)	**0.65** / **0.65**	0.78 / 0.70	- / -
2 (electrons)	**0.56** / **0.56**	0.80 / 0.80	0.60 / 0.62
3 (electrons)	**0.56** / **0.57**	0.91 / 0.95	0.67 / 0.85

Final CaloChallenge Results



CaloDiffu
CaloDream

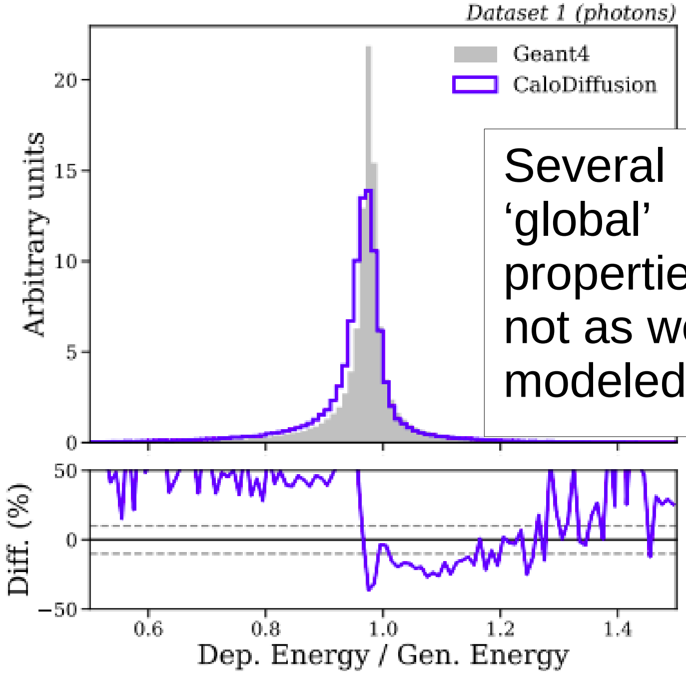
Areas for Improvement



Generation speed

CaloDiffusion slower than other models

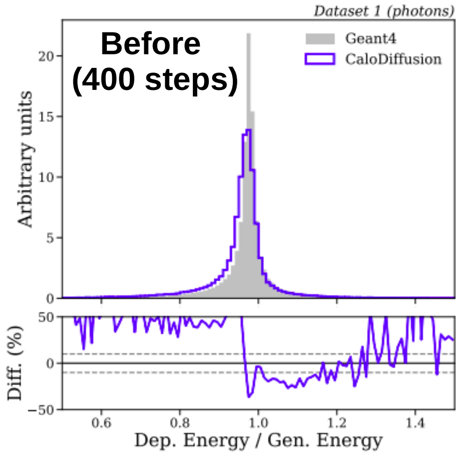
→ 400 diffusion iters is a lot

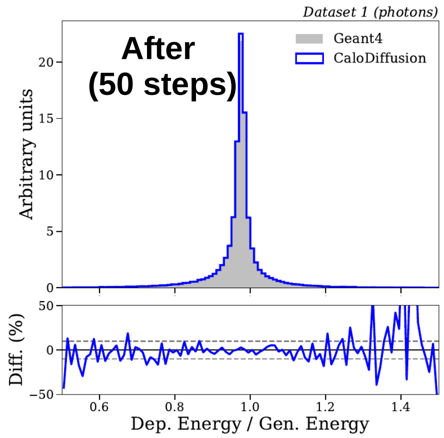


Several 'global' properties not as well modeled

Post-Challenge Improvements

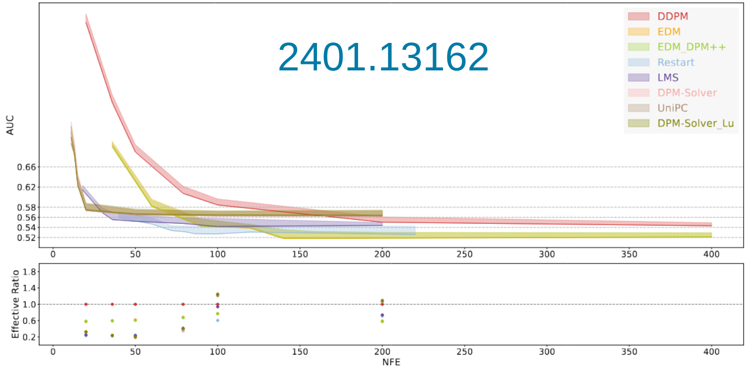
- Learn from other methods!
- Factorize model: separate network for ‘global’ properties (energy per layer), and ‘shape’
- Speedups from better sampling algo’s
  - Distillation techniques
  - Ongoing R&D
- Time to move to CMS datasets !



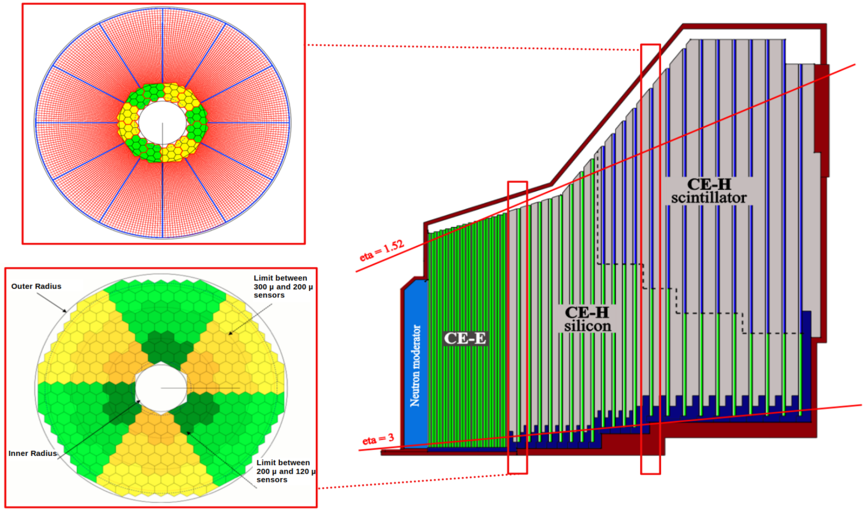


Choose Your Diffusion: Efficient and flexible ways to accelerate the diffusion model in fast high energy physics simulation

Cheng Jiang[1⋆], Sitian Qian[2†] and Huilin Qu[3‡]



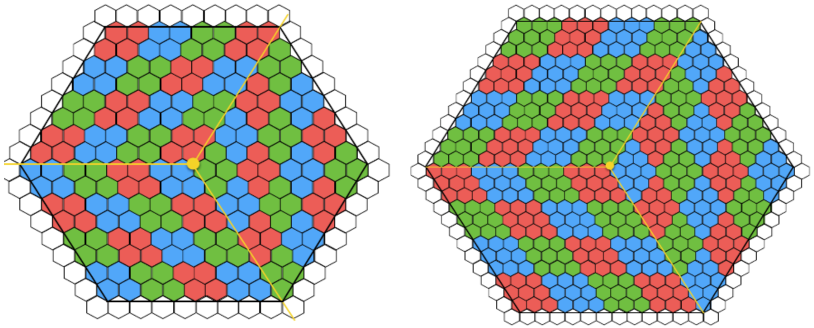
CMS HGCal



Individual Hexagon Wafers

Large Cells

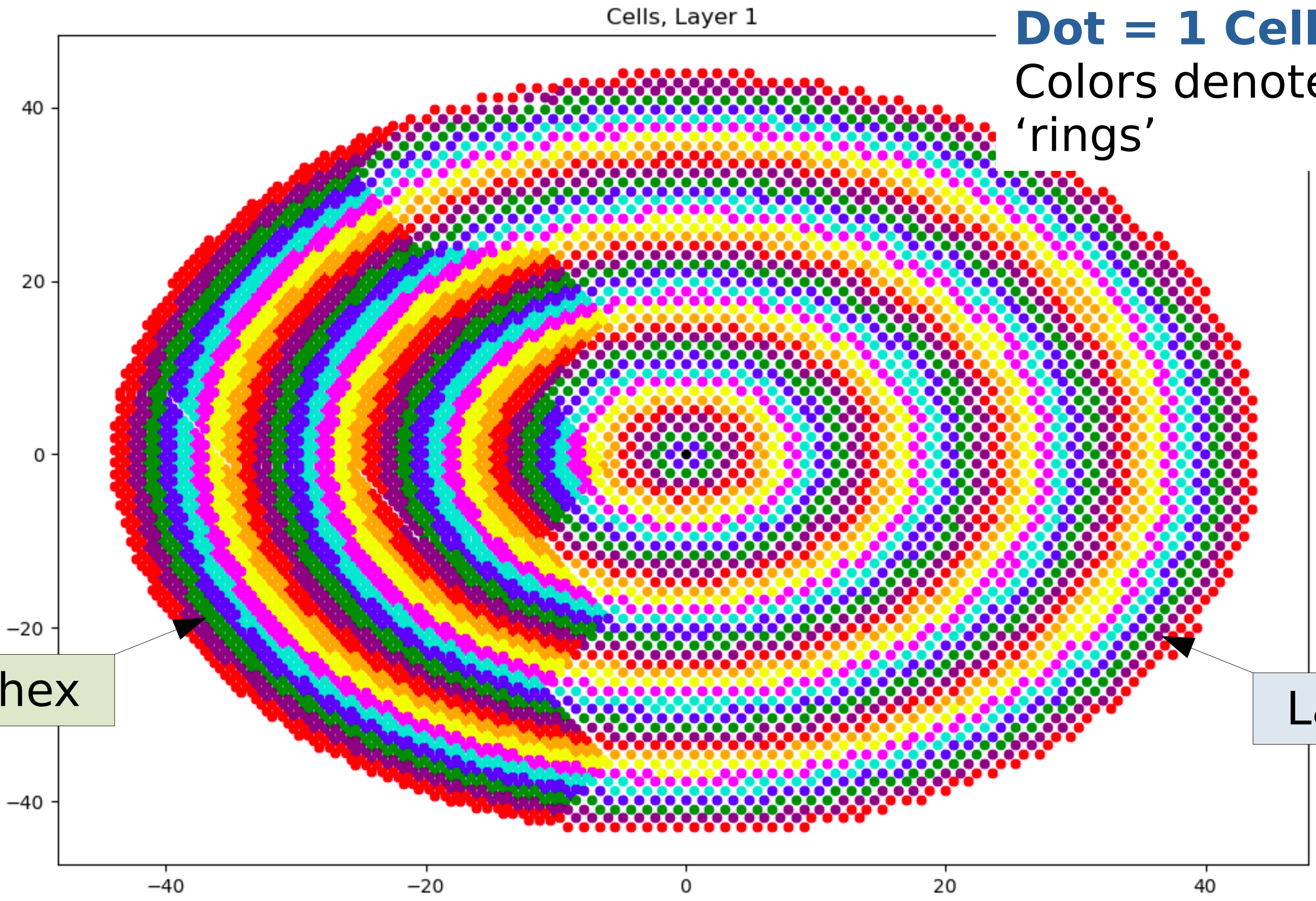
Small Cells



+ boundary conditions!

- Simulating HGCal significantly more difficult than CaloChallenge
  - Significantly higher granularity (10x or more)
  - Very complicated geometry!

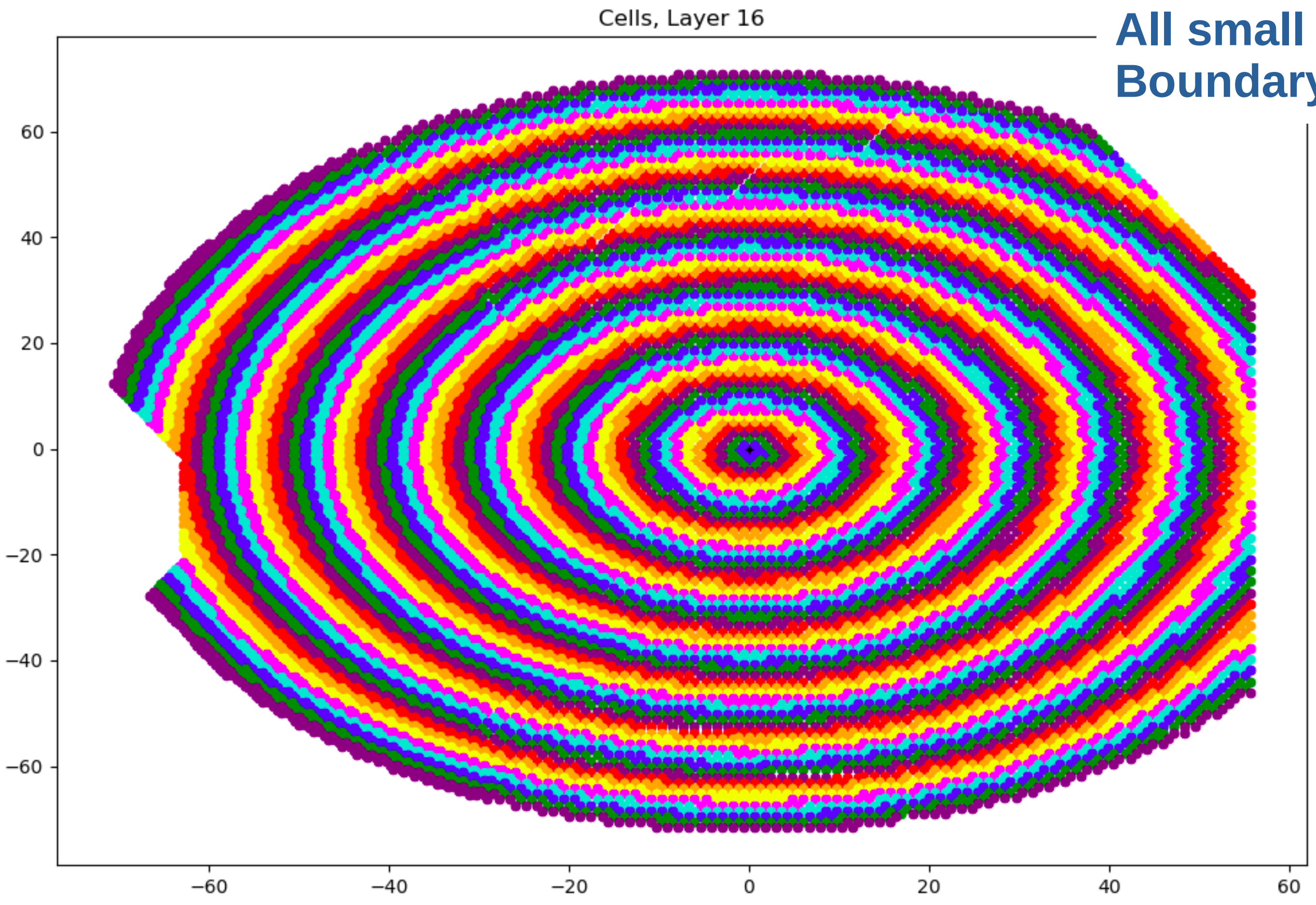
Example HGCal Region : Layer 1



Dot = 1 Cell

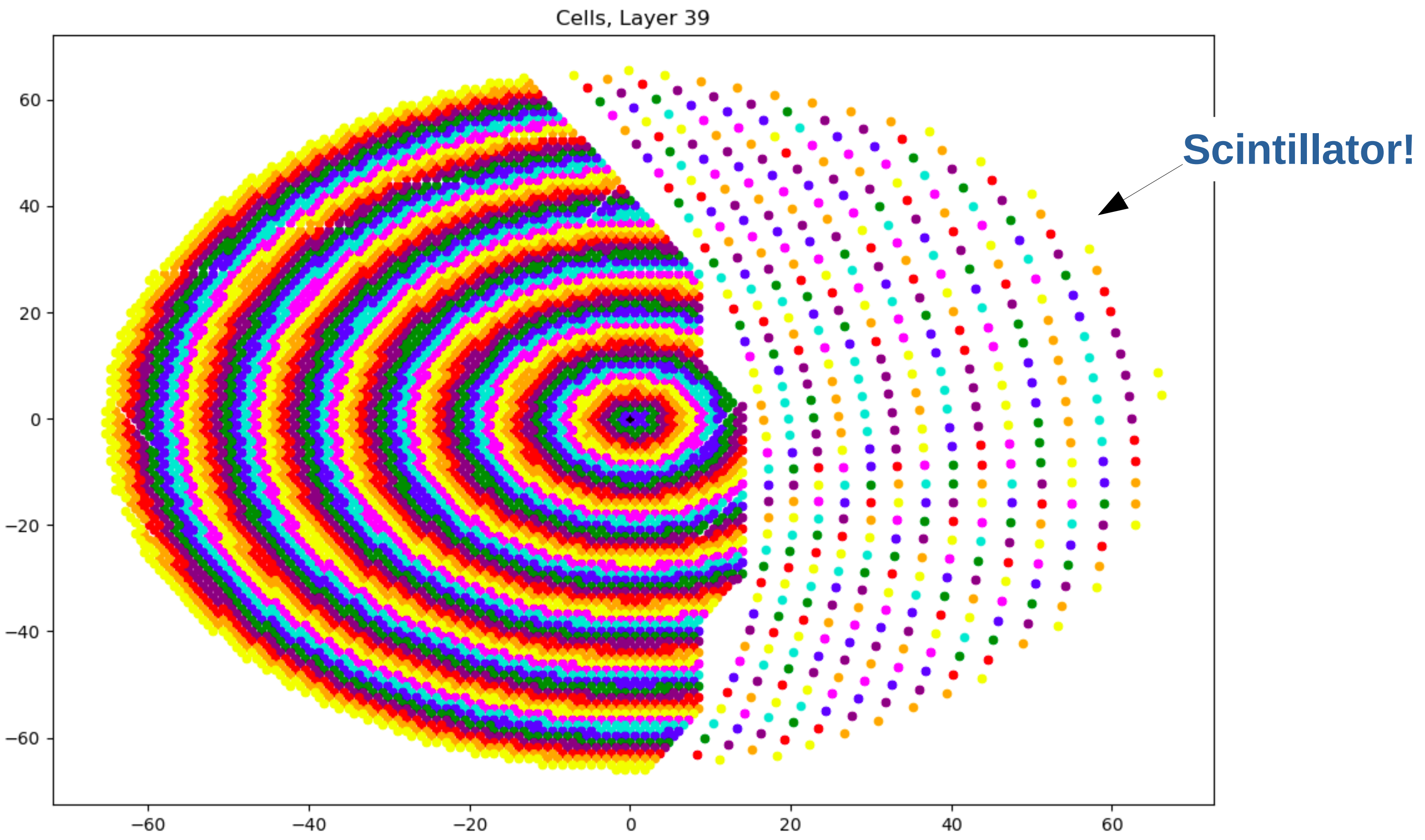
Colors denote different 'rings'

Example HGCal Region : Layer 16

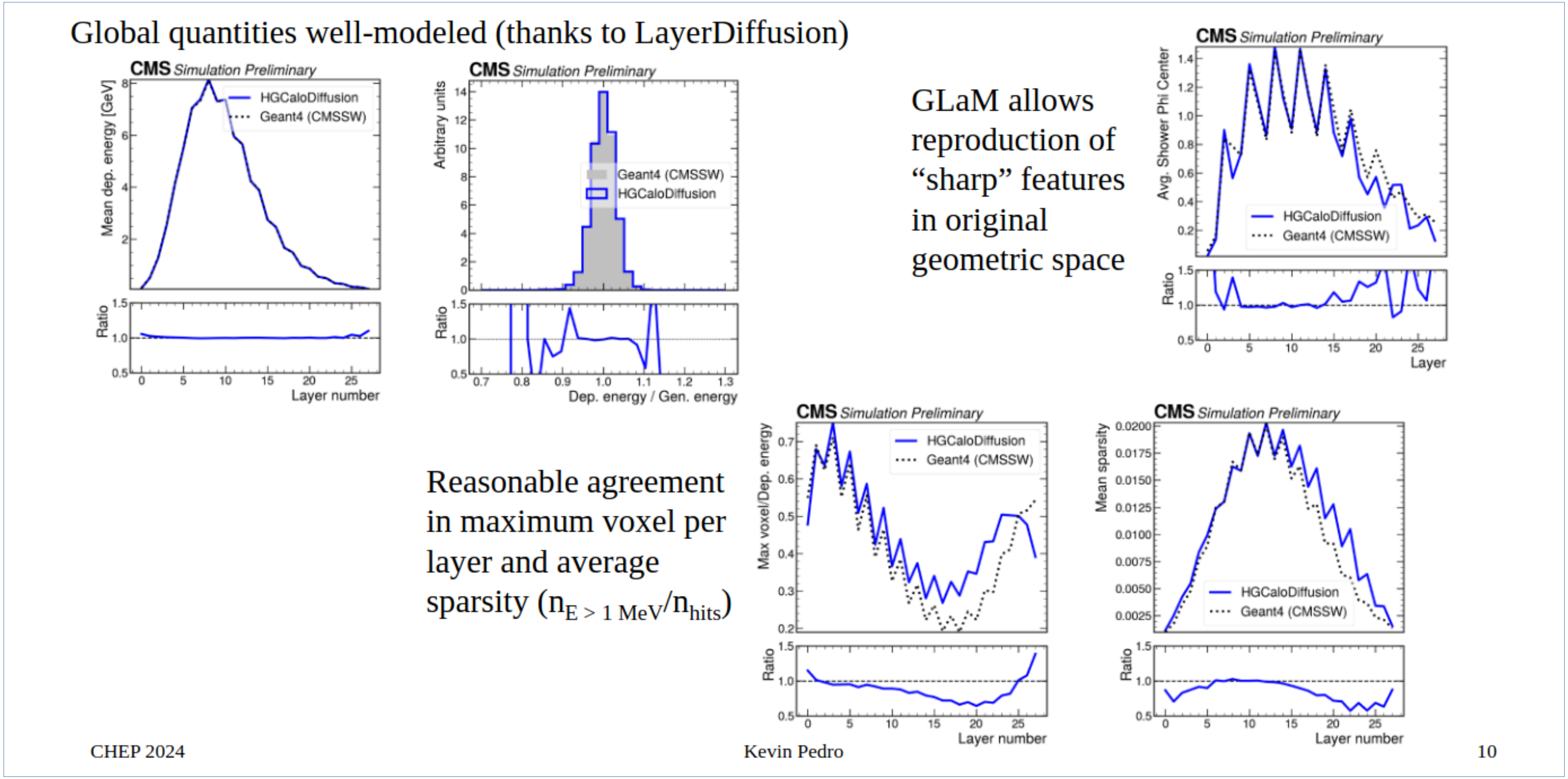


All small hex!
Boundary Conditions!

Example HGCal Region : Layer 39



Performance (CHEP '24)



[Link to slides](#)

Many further improvements in the last year!

Outlook

- CaloChallenge was very helpful for initial development of CaloDiffusion
- Now focus is on CMS application : HGCal
  - Significant step up in granularity & complexity
- Continue to improve CaloDiffusion to handle these challenges
  - Promising results so far
- Will employ SONIC (link) to enable GPU + multi-event batching
- Other models in CMS also being developed
- Look out for public results soon!