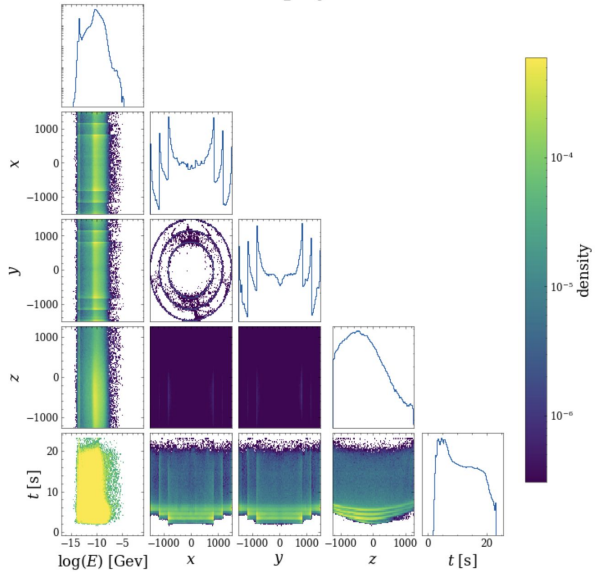
Update: GenAI for BIB Team

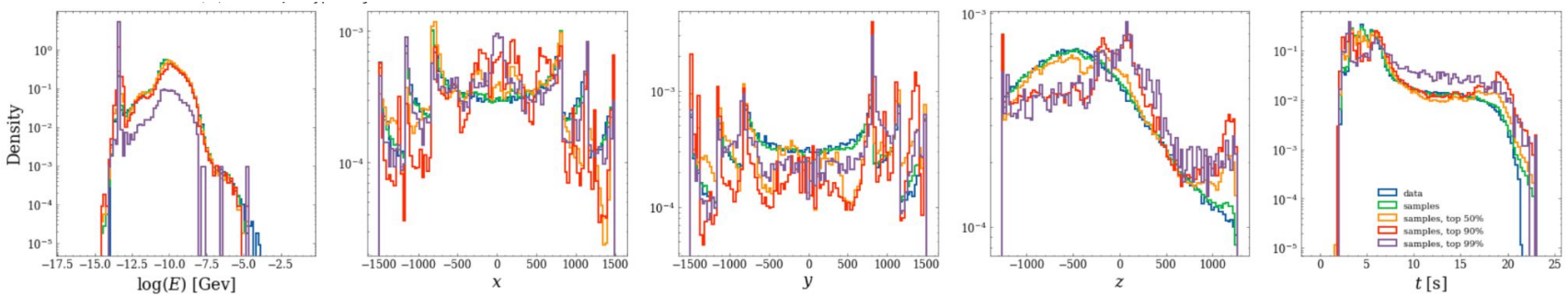
03/13/2026

Radha Mastandrea, Rose Powers, Max Zhao

Progress so far: normalizing flow-based model (UCHicago)

- We have a pipeline for BIB simulation at the SimTrackerHit (pre-digi) level (GitHub)
- We do not yet meet the criteria for indistinguishability of samples from simulation (binary classifier ROC AUC ≈ 0.6 > 0.5)



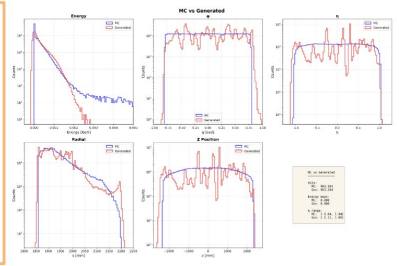


Progress so far: diffusion model (Princeton)

- Preliminary results from Rose and Max from the February 10th MAIA meeting

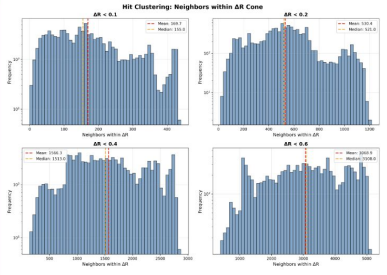
Results: Basic Kinematics

- Model trained on 700 events, avg 9k hits
- **Plotted**: Comparison of 100 MC test events and 100 generated events
  - Log scale



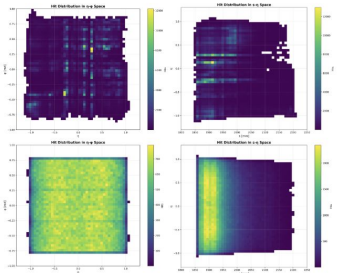
Results: Cluster Structure - Generated

- **Important goal:** capture local structure of ECAL clusters
- Note: we do not condition on any gen particle information. Ideally ECAL clusters should spontaneously appear.



Results: Basic Kinematics

- Model trained on 700 events, avg 9k hits
- Top: Generated events
- Bottom: MC events



Progress so far: evaluating flow samples

- We have the start of a pipeline for samples → LCIO → digitization → reconstruction

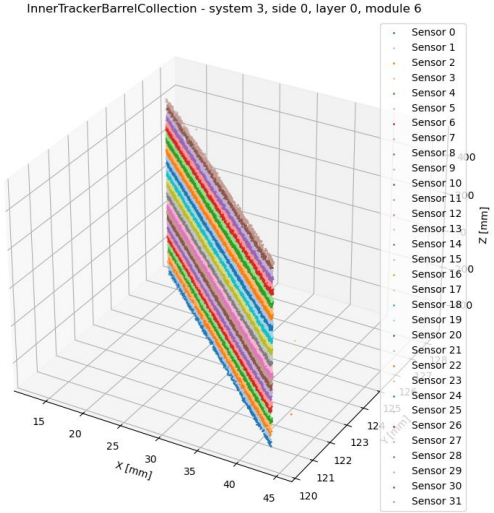
  Digitization output: (still many things to debug…)

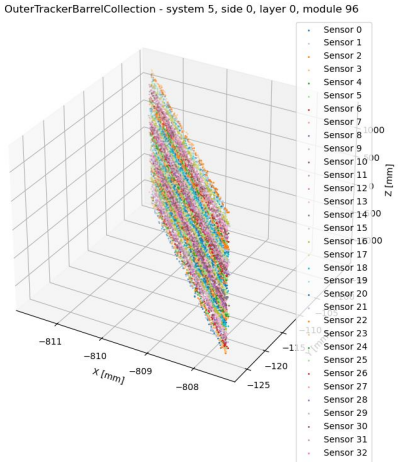
RUN: 1 EVENT: 0 DETECTOR: unknown

COLLECTION NAME	COLLECTION TYPE	# OF ELEMENTS
AllTracks	Track	0
ITBarrelHits	TrackerHitPlane	2125
ITBarrelHitsRelations	LCRelation	2125
InnerTrackerBarrelCollection	SimTrackerHit	663943
JetOut	ReconstructedParticle	0
MuonHits	CalorimeterHit	0
MuonHitsRelations	LCRelation	0
OTBarrelHits	TrackerHitPlane	128
OTBarrelHitsRelations	LCRelation	128
OuterTrackerBarrelCollection	SimTrackerHit	678750
PandoraClusters	Cluster	0
PandoraPFOs	ReconstructedParticle	0
PandoraStartVertices	Vertex	0
SeedTracks	Track	0
SelectedPandoraPFOs	ReconstructedParticle	0
SiTracks	Track	0
VXDBarrelHits	TrackerHitPlane	755
VXDBarrelHitsRelations	LCRelation	755
VertexBarrelCollection	SimTrackerHit	264185

Blocks so far

- Pipeline for samples → LCIO needs further debugging
  - In particular, in order to set `CellID0`, we need a more robust mapping from (x, y, z) →cell ID (currently using a lookup table)
  - More broadly, having someone else on board who is more familiar with the LCIO pipeline would be very useful
- In general, we don't have that much personpower





Open tasks

- Generative model optimization
  - Finding a good baseline architecture and doing a hyperparameter sweep
  - Training on multiple `SimTrackerHit` collections, potentially simultaneously
  - Smarter sampling from the models (e.g. to get crispy tracker layer boundaries)
- Downstream task optimization
  - Indistinguishability criteria for simulation vs. flow samples
  - Spontaneous track generation
  - What else?
- Possible collaboration / competition with Brown + LBL